2015-10-01

Unveiling Molecular Mechanisms of piRNA Pathway from Small Signals in Big Data: A Dissertation

Wei Wang
University of Massachusetts Medical School

UNVEILING MOLECULAR MECHANISMS OF piRNA PATHWAY FROM SMALL SIGNALS IN BIG DATA

A Dissertation Presented

By

Wei Wang

Submitted to the Faculty of the

University of Massachusetts Medical School

Graduate School of Biomedical Sciences

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

October 1, 2015

Bioinformatics and Integrative Biology
&
RNA Therapeutics Institute

UNVEILING MOLECULAR MECHANISMS OF piRNA PATHWAY

FROM SMALL SIGNALS IN BIG DATA

A Dissertation Presented
By
Wei Wang

The signatures of the Dissertation Committee signify completion and approval as to style and content of the Dissertation

_____
Phillip Zamore, Ph.D. & Zhiping Weng Thesis Advisors

_____
Donal O'Carroll, Ph.D., Member of Committee

_____
Manuel Garber, Ph.D., Member of Committee

_____
Andrei Korostelev, Ph.D., Member of Committee

_____
Wen Xue, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

_____
William E. Theurkauf, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Science

Bioinformatics and Computational Biology

Oct. 1, 2015

DEDICATION

To my family and my dear grandfather!

To become a wise woman.

ACKNOWLEDGEMENTS

First of all, I would like to give my deepest gratitude to my two PhD mentors: Zhiping and Phil. From Zhiping, I was greatly influenced by her eager and endeavor for contacting and learning the most advanced technology and skills in different disciplines, and sharing what she has learned with us. As a female scientist she is a good role model for her passion and profession on science. As a PI leading a bioinformatics lab, she sets a vivid example on how to establish close and reliable collaborations with other groups. I very much appreciate her understanding and support when I was struggling with my core courses and qualifying exam. I am very lucky to have such a positive mentor during my PhD training. By recruiting a group of smart, geeky, and friendly people, Weng lab becomes a lovely working place. I would like to extend my thanks to former and current Weng lab members: Jia, Jie, Micheal, Shikui, Thom, Brian, Xianjun, Junko, Hao, Jill, Yu, Arjan, Tyler, Jiali, Eugenio, and Sowmya for valuable discussions on my projects and for their kind technical support. Barbara is a great assistant for all our work.

From Phil, I saw his endeavor to be perfect on all the work on our hands. It does not allow any time to take a breath, but it leads us to top-level science. He is dedicated to be a good educator as well as an excellent scientist. Phil gave me tremendous help on professional development. The lab he leads provides a great platform for learning by gathering great people with all sorts of strengths. In addition, he generously supports students to be exposed to great science outside

of the lab. During my time in the Zamore lab, I specifically thank Chengjian for his great mentoring during my rotation, Ryuya for his guidance and discussion that leads to an accidental and tremendous discovery, thank Bo for his seamless collaboration on multiple projects, thank Cindy for her reliable technique on fly work and thank Daniel for his trust on my skills. I have also received a lot of care and help from Jen, Chasan, Wes, Tim, Sam, Wee, Keith, Pei-Hsuan, Cansu, Amena, Yongjin, Kaycee, Paul, Yujing, Alicia, Katharine, Stefan, Elif, Chris, Traycey, Zhao, Desiree, Fabian and Xin. As a lab manager, Gwen is very helpful and helps to foster the lab a family-like atmosphere. Tiffanie is very prompt and reliable, and she helped me organizing practice talks, celebration and thesis defense (together with Barbara, Amy). I would also like to thank my committee members, Craig C. Mello, Manuel Garber, William E. Theurkauf, Donal O'Carrol and my TRAC committee member Melissa Moore. Craig offered his greatest insight on science and encouraged me to pursue for interesting science instead of dull ones. Manuel's expertise on bioinformatics inspired me greatly. Bill's high standard on molecular biology and genetics since my qualifying exam helped me grow rapidly in biology. Melissa's famous question "why do I care" always encourages deeper thought on the significance of what I am working on. I would also like to thank Donal, Wen and Andrei for agreeing to sacrifice his valuable time and serve in my defense committee. Thank you for all other people I don not have space to mention here. It is because of you, I then have such memorable five years of PhD life.

ABSTRACT

PIWI-interacting RNAs (piRNA) are a group of 23–35 nucleotide (nt) short RNAs that protect animal gonads from transposon activities. In *Drosophila* germ line, piRNAs can be categorized into two different categories— primary and secondary piRNAs— based on their origins. Primary piRNAs, generated from transcripts of specific genomic regions called piRNA clusters, which are enriched in transposon fragments that are unlikely to retain transposition activity. The transcription and maturation of primary piRNAs from those cluster transcripts are poorly understood. After being produced, a group of primary piRNAs associates Piwi proteins and directs them to repress transposons at the transcriptional level in the nucleus. Other than their direct role in repressing transposons, primary piRNAs can also initiate the production of secondary piRNA. piRNAs with such function are loaded in a second PIWI protein named Aubergine (Aub). Similar to Piwi, Aub is guided by piRNAs to identify its targets through base-pairing. Differently, Aub functions in the cytoplasm by cleaving transposon mRNAs. The 5′ cleavage products are not degraded but loaded into the third PIWI protein Argonaute3 (Ago3). It is believed that an unidentified nuclease trims the 3′ ends of those cleavage products to 23–29 nt, becoming mature piRNAs remained in Ago3. Such piRNAs whose 5′ ends are generated by another PIWI protein are named secondary piRNAs. Intriguingly, secondary piRNAs loaded into Ago3 also cleave transposon mRNA or piRNA cluster transcripts and produce more

secondary piRNAs loaded into Aub. This reciprocal feed-forward loop, named the “Ping-Pong cycle”, amplified piRNA abundance.

By dissecting and analyzing data from large-scale deep sequencing of piRNAs and transposon transcripts, my dissertation research elucidates the biogenesis of germline piRNAs in *Drosophila*.

How primary piRNAs are processed into mature piRNAs remains enigmatic. I discover that primary piRNA signal on the genome display a fixed periodicity of ~26 nt. Such phasing depends on Zucchini, Armitage and some other primary piRNA pathway components. Further analysis suggests that secondary piRNAs bound to Ago3 can initiate phased primary piRNA production from cleaved transposon RNAs. The first ~26 nt becomes a secondary piRNA that bind Aub while the subsequent piRNAs bind Piwi, allowing piRNAs to spread beyond the site of RNA cleavage. This discovery adds sequence diversity to the piRNA pool, allowing adaptation to changes in transposon sequence. We further find that most Piwi-associated piRNAs are generated from the cleavage products of Ago3, instead of being processed from piRNA cluster transcripts as the previous model suggests. The cardinal function of Ago3 is to produce antisense piRNAs that direct transcriptional silencing by Piwi, rather to make piRNAs that guide post-transcriptional silencing by Aub. Although Ago3 slicing is required to efficiently trigger phased piRNA production, an alternative, slicing-independent pathway suffices to generate Piwi-bound piRNAs that repress transcription of a subset of transposon families. The alternative pathway may help flies silence

newly acquired transposons for which they lack extensively complementary piRNAs.

The Ping-Pong model depicts that first ten nucleotides of Aub-bound piRNAs are complementary to the first ten nt of Ago3-bound piRNAs. Supporting this view, piRNAs bound to Aub typically begin with Uridine (1U), while piRNAs bound to Ago3 often have adenine at position 10 (10A). Furthermore, the majority of Ping-Pong piRNAs form this 1U:10A pair. The Ping-Pong model proposes that the 10A is a consequence of 1U. By statistically quantifying those target piRNAs not paired to g1U, we discover that 10A is not directly caused by 1U. Instead, fly Aub as well as its homologs, Siwi in silkmoth and MILI in mice, have an intrinsic preference for adenine at the t1 position of their target RNAs. On the other hand, this t1A (and g10A after loading) piRNA directly give rise to 1U piRNA in the next Ping-Pong cycle, maximizing the affinity between piRNAs and PIWI proteins.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

COPYRIGHT NOTICE

CHAPTER I Introduction

piRNA and PIWI-clade Argonautes, targeting rule

PIWI-interacting RNAs (piRNAs) are the third class of small non-coding RNAs identified after microRNAs (miRNA) and small interfering RNAs (siRNA). Central to all small RNA pathways are the Argonaute (Ago) proteins. Throughout evolution, Ago proteins have diverged into specialized clades: AGO clade and PIWI clade. AGO clade proteins bind miRNAs and siRNAs whereas PIWI clade Ago proteins bind piRNAs. AGO proteins and miRNAs are ubiquitously expressed in both animals and plants at different developmental stages and tissues, while PIWI proteins and piRNAs are predominantly expressed in animal gonads. Additionally, both miRNAs and siRNAs are processed from double-stranded precursors in a Dicer-dependent pathway, while piRNAs are processed from single-stranded precursors in a Dicer-independent manner (Vagin et al., 2006).

Structural studies have revealed that Argonaute proteins share common modules that are composed of an amino-terminal (N) domain, PIWI-ARGONAUTE-ZWILLE (PAZ), middle (MID) and PIWI domains, highly specialized for small RNA binding and target RNA slicing. Earlier crystal structural studies on bacterial and archaeal Argonaute proteins revealed that Argonaute proteins consist of two lobes connected by a hinge. One of the lobes

is made up of the MID and PIWI domains and the other one is composed of the N-terminal and PAZ domains (Parker et al., 2004; Yuan et al., 2005). A phosphate binding pocket in the MID domain anchors the 5′ end of the small RNA with the 5′ terminal base stacking on a conserved tyrosine (Ma et al., 2005; Wang et al., 2008a). The PIWI domain, which adopts a fold resembling to RNase H (Parker et al., 2004; Yuan et al., 2005), cleaves the phosphosdiester bond between the 10^{th} and 11^{th} position in the target RNA, counting from the 5′ end of guide RNA (Elbashir et al., 2001c). The side chains of conserved aspartic acid and histidine residues in the PIWI domain form a DDH catalytic triad or DEDD tetrad that coordinates a magnesium ion proposed to activate a nucleophilic water molecule for hydrolysis of the phosphosdiester bond (Tolia and Joshua-Tor, 2007; Jinek and Doudna, 2008; Sheng et al., 2014). The N domain supports small RNA loading and engages in unwinding the small RNA duplex (Wang et al., 2009; Kwak and Tomari, 2012). The PAZ domain anchors the 3′ end of the guide RNA. A hydrophobic cavity within this domain harbors the 2′-*O*-CH_3 group on the 3′ end of siRNA and piRNA guides (Simon et al., 2011).

A common theme for all three classes of small RNAs is that they guide Argonautes to find targets through sequence complementarity. The accommodation of Argonaute structure domains to the guide:target pairing affects target binding and determines the fate of targets. The extent of complementarity differs among three classes of small RNAs. For miRNAs, the “seed region” binding–from guide position 2 to positions 7 or 8 (g2–g7 or g2–g8)

primarily determines target specificity (Lewis et al., 2003; Lewis et al., 2005). The structure of Piwi-RNA complex from *A. fulgidus* show that the bases of nucleotides 2-6 of the guide strand are exposed and free to base-pair with a target mRNA (Ma et al., 2005; Parker et al., 2005). The pre-helical structure formed between a small portion of the guide and target reduces the entropic barrier to initiate target binding (Wee et al., 2012). A recent structural study of human Ago2 with and without target RNAs suggest a similar stepwise mechanism— Ago2 primarily exposes guide g2-g5 for initial target paring, which then promotes conformational changes that expose g2 to g8 and g13 to g16 for further target recognition (Schirle et al., 2014). Biochemical studies indicate that Argonaute divides its guide into different domains–the anchor, seed, central, 3′ supplementary and tail– with distinct biochemical properties that explain the differences between how animal miRNAs and siRNAs bind their target mRNAs (Wee et al., 2012). siRNAs generally bind more tightly to their target RNAs than miRNAs do. When siRNAs bind to their targets extensively— which enables a conformational change and brings the target closer to the active site— it catalyzes endonucleolytic cleavage of the targets (Elbashir et al., 2001a; Elbashir et al., 2001c; Haley and Zamore, 2004; Rivas et al., 2005; Ameres et al., 2007; Wang et al., 2008b; Wang et al., 2009). piRNAs also guide their PIWI-Argonautes to cleave their targets. It has been biochemically shown that the least amount of complementarity between a piRNA guide and target required for MIWI to slice is g2-g21 (Reuter et al., 2011), though non-cleavage regulation might

require less stringent piRNA-target pairing. Since piRNA sequences are diverse, extensive and strict pairing set specificity on the target repertoire for cleavage. An exception for this common theme is a model proposed by Vourekas et al. in which pachytene piRNAs in adult mice, rather than serving as active guides for target repression, are generated as part of a germ cell degradation mechanism for long non-coding RNAs in spermiogenesis (Vourekas et al., 2012). In their model, Miwi bind mRNA directly without using a piRNA as guide.

Transposon and piRNA clusters

Transposable elements (TE) are prevalent in the genome of all plants and animals. They are selfish DNA elements that exploit the genome and replicative machinery of host cells in order to proliferate. TE sequences constitute nearly half of the mammalian genome and nearly one-third of the fruit fly *Drosophila* genome (Kaminker et al., 2002). This number can increase to 90% in some plants. Movement of transposons within the genome induces mutations at their excision and/or insertion sites, which can result in genomic instability. Transposons can be divided into two broad classes: DNA transposons and retrotransposons. DNA transposons move via a cut and paste mechanism without copying themselves, whereas retrotransposons propagate their copies via a “copy-and-paste” mechanism. A DNA transposon is excised from its original site, creating dsDNA break, which might generate mutations. Retrotransposons are transcribed from the genome, followed by reverse transcription, a process mediated by a transposon-encoded reverse transcriptase, and then integrated

into a new location, which suggests that it only affects the target sites. Retrotransposons can be further categorized into those that possess long terminal repeats (LTRs) and those that do not (non-LTRs). LTR retrotransposons and retroviruses are quite similar in structure, while non-LTR retrotransposons are typified by long intersperse nuclear element 1 (LINE-1) elements of mammals. Non-LTR retrotransposons reshape the genome through insertional mutagenesis with 5′ truncated or 5′ inverted (cis) insertions, trans-driven insertion of processed pseudogenes (trans), deletions and duplications due to unequal homologous recombination, rearrangements and duplication at new chromosomal sites (Kazazian, 2004).

The host genome places controls on those mobile elements via multiple mechanisms. One of the defense mechanisms involves the packaging of transposon-rich regions into heterochromatin. Additionally, TEs can be regulated by siRNAs and piRNAs because a significant fraction of endo-siRNAs and piRNAs correspond to transposon sequences with perfect matches or few mismatches, implying a sequence-based recognition and repression mechanism (Vagin et al., 2006; Ghildiyal et al., 2008). In *Drosophila*, around 70% of piRNAs are derived from transposons and other repetitive elements (Brennecke et al., 2007). Endo-siRNAs regulate transposons mainly in somatic cells (Ghildiyal et al., 2008) while piRNAs guard the genome from transposon perturbations specifically in germ line cells (Vagin et al., 2006).

It has been proposed that an evolutionary arms race occurs constantly between transposons and the host genome. From the perspective of evolution driving force, host genomes might benefit from transposition, which, in certain instances, can lead to beneficial mutations (Kazazian, 2004). Transposition may also cultivate new methods to defend against newly evolved transposons. In this scenario, germ line cells are of vital importance in transposon regulation and genome integrity as they vertically transmit genetic material to future progeny.

To battle the threats from mobile elements in germ line cells, the *Drosophila* genome evolved an adaptive immune system— piRNA clusters—, to subject transposons to a silencing program via homologous elements. piRNA clusters are discrete genomic loci, computationally defined as piRNA dense regions in the fly genome. While they produce around 90 percent of the total piRNAs, 142 piRNA clusters account for only 3.5% of the fly genome sequences (Brennecke et al., 2007; Senti and Brennecke, 2010). Among all piRNA clusters, only 7 are in presumed euchromatic regions. The rest are located at pericentromeric and subtelomeric chromatin domains. These piRNA clusters are composed of a large number of randomly oriented transposons, arrays of nested transposon fragments, and records of previous invasion. Therefore, cluster-derived piRNAs are complementary to active transposon copies located outside of clusters, acting as guide to regulate transposons *in trans*.

piRNA clusters are categorized into uni-strand and dual-strand clusters based on the direction of piRNAs within each cluster. Uni-strand clusters produce

piRNAs from only one of the two strands. One of the representative uni-strand cluster is *flamenco* (*flam*) locus, which spans about 150 kb. This locus produces a major fraction of piRNAs in somatic cells. It was originally discovered as a repressor of transposons *gypsy*, *idefix*, and *ZAM* genetically*,* while efforts searching for protein coding genes in *flam* were unsuccessful (Pélisson et al., 1994). Most piRNAs from this cluster are oriented in the antisense direction to TEs, which provides a molecular basis for Piwi bound piRNAs in soma since most of them are also anti-sense to TEs. Dual-strand clusters, on the other hand, produce piRNAs from both strands. The 42AB piRNA cluster is the largest dual-strand cluster in *Drosophila* ovaries and accounts for 30% of germline piRNAs. piRNAs from this cluster are both sense and antisense to certain transposons.

piRNA cluster transcription, splicing inhibition, nuclear exporting

The canonical features of transcription had been explored around piRNA clusters since they were defined as genetic elements for piRNA sources, though the effort is futile. In flies, only *flam* cluster seems to have putative promoter and defined transcriptional start site. A *p*-element insertion in this putative promoter abolished the piRNA production from the entire cluster downstream, suggesting that a long, uni-strand transcript is produced from this region (Brennecke et al., 2007). A recent study shows that the transcription of *flam* cluster is under the control of a transcription factor, Ci (Goriaux et al., 2014). *flam* cluster in ovarian somatic cells

display canonical RNA Polymerase II regulated transcription, including defined promoter and termination sequences (Mohn et al., 2014). However, the *flam* cluster is expressed mainly in follicle cells. Whether germline dual strand clusters have two promoters at both ends, or they exploit some promoters from neighboring genes or cryptic promoters within the clusters remains enigmatic. Dual-strand clusters not only lack of clear promoters, but also a 5′ cap and transcription termination.

piRNA cluster transcription associates with heterochromatin formation or repressive chromatin marks. Deposition of Histone 3 Lysine 9 trimethyl marks (H3K9me3) by the methyltransferase dSETDB1 (egg) was proposed to be a requisite for piRNA cluster transcription in both germ cells and somatic cells (Rangan et al., 2011). Rhino, a Heterochromatin Protien 1 (HP1) paralog, Cutoff (Cuff), a yeast Rai1-like nuclease that participates in RNA quality control and transcription termination, Deadlock (Del), a protein with no particular domains and no specific molecular function, physically interact, and together they bind specifically to dual strand clusters in the germline to promote their transcription (Klattenhoff et al., 2009; Pane et al., 2011; Le Thomas et al., 2014b; Mohn et al., 2014). Since HP1a is recruited to chromatin via H3K9me3, Rhino likely binds to piRNA clusters using the same mechanism. However, H3K9me3 coated piRNA clusters are a subset of genomic regions which harbor this repressive histone mark. How then can the cell tell which loci are processed into piRNA precursors? One possibility is that the cell stores certain molecular memories, possibly from

their parents. One candidate of such molecular memory is maternally inherited piRNAs —likely homologous to piRNA clusters—, which initiate the installment of H3K9me3 mark on genomic cluster sequences (Le Thomas et al., 2014b). Further studies suggest that Rhino, Deadlock, and Cuff form a complex that anchor on H3K9me3-marked chromatin on dual-strand piRNA clusters (Mohn et al., 2014; Zhang et al., 2014). Mohn et al. and colleagues proposed that piRNA precursor transcripts from dual-strand clusters are noncanonical byproducts of convergent transcription of neighboring genes. The model proposes that Cuff binds to the uncapped 5′ end of piRNA precursors, which is generated by 3′ end cleavage and polyadenylation of the upstream genic transcript, and protects the 5′ ends of transcripts from the cap-binding complex. This results in inhibition of alternative splicing as well as polyadenylation and transcription termination, leading to continuous transcription of the piRNA cluster transcription (Mohn et al., 2014). In a different model proposed by Zhang et al., Rhino can function together with Cuff and UAP56 to suppress alternative splicing of piRNA clusters (Zhang et al., 2014). This suppression of alternative splicing can explain why piRNA machinery is able to distinguish piRNA clusters from mRNAs, though the underlying mechanism is not yet clear.

In mice testes, a germline specific MYB transcription factor protein, A-MYB, drives piRNA precursor production bi-directionally in the pachytene stage. A-MYB binding sites are enriched in a motif. The cluster transcripts in mice have both 5′ caps and 3′ poly (A) tails, features of Pol II regulated mRNA transcripts.

Therefore, piRNA clusters in mice are also called piRNA genes (Li et al., 2013). In the *Bombyx mori* tissue cultured cell line BmN4, piRNA clusters have been shown to have a high level of H3K4me3 mark, which is a hallmark of active transcription, suggesting a different mechanism of piRNA cluster transcription in this organism (Kawaoka et al., 2008).

Following cluster transcription in nuclei, piRNA precursors are exported into nuage, a perinuclear, membraneless, and electron-dense structure under the electron microscope. In the fruit fly ovary, most germline piRNA pathway components localize to nuage, including the RNA helicases Vasa and Armitage, the PIWI proteins Aub and Ago3, as well as the Tudor proteins (Brennecke et al., 2007; Li et al., 2009; Malone et al., 2009; Zhang et al., 2011). Nuage is believed to be one of the major compartments where piRNAs get produced from piRNA precursors and transposons are silenced post-transcriptionally. In a proposed model, UAP56, a putative helicase, forms a complex with Rhino, likely escorts the primary transcript through the nuclear pore to nuage because Rhino-UAP56 foci are positioned next to the nuclear pore. It is also speculated that the transcript is handed over to Vasa and funneled into the biogenesis machinery. The importance of UAP56 in the piRNA pathway comes from the observation that mutation of UAP56 leads to germline transposon up-regulation, decrease of piRNAs mapping to germline clusters, and dislocalization of Aub, Ago3 and Vasa from nuage (Zhang et al., 2012a).

piRNA biogenesis: linear primary pathway and secondary “Ping-Pong” model

The *Drosophila* ovary consists of germline cells and somatic follicle cells. Germline cells include nurse cells and oocytes, which are surrounded by somatic follicle cells supporting the development of the germ line. In both germ line cells and somatic cells the piRNA pathway safeguards transposons. However, of the three PIWI family members —Piwi, Aubergine (Aub) and Argonaute3 (Ago3)— express in germ line cells, only Piwi expresses in somatic cells. Therefore, germ line cells harbor more complex piRNA biogenesis and function pathways than somatic cells do. In somatic cells, piRNAs seem to be exclusively generated via a linear primary pathway. Long, single-stranded piRNA precursors are transcribed from piRNA clusters in the genome. Primary piRNA transcripts are then processed into intermediates by an unknown mechanism. Zucchini (Zuc), normally anchored on the mitochondria surface, processes piRNA intermediates into piRNA fragments, and is the top candidate for producing the 5′ ends of piRNAs via its endonuclease activity (Ipsaro et al., 2012; Nishimasu et al., 2012). Since recombinant Zuc fails to show any 5′ preference when it cleaves single-stranded RNA in vitro, those piRNA intermediates with 5′ Uridine (U) are likely preferentially loaded into Piwi, given that Piwi bound piRNAs typically start with a U (Frank et al., 2010). If the length of loaded piRNA intermediates is longer than that of mature piRNAs, the 3′ end of those intermediates likely undergoes trimming by an unknown enzyme (Kawaoka et al., 2011), though a direct

evidence for the existence of such a mechanism in fly has not been shown. piRNAs are 2′-*O*-methylated at their 3′ termini by a methyl-transferase Hen1 (Horwich et al., 2007). Such modification has been shown to be coupled with 3′ trimming in BmN4 cells (Kawaoka et al., 2011). It is believed that only loaded Piwi can enter into the nucleus.

In germ lines, the primary piRNA pathway is also believed to generate piRNAs for Piwi and Aub (Li et al., 2009; Malone et al., 2009). Essential primary pathway components, Zuc, Armitage (Armi) and Yb, which are expressed in follicle cells or ovarian somatic cells are also present in germline cells (Brother of Yb and Sister of Yb in germline cells for Yb) (Olivieri et al., 2010; Handler et al., 2011). The cluster transcript, in full length or fragmented intermediates, is exported into nuage, where they are further processed by primary machineries including Zuc, Armi, Shutdown, Vreteno and Yb (Handler et al., 2011). This process might occur at the surface of the outer membrane of mitochondria, where Zuc localizes. A regular shuttle might transport piRNA intermediates and mature piRNAs between nuage and mitochondria (Aravin and Chan, 2011).

In addition to Piwi-loaded primary piRNAs, a proportion of primary piRNAs, as well as maternally deposited piRNAs, can bind to Aub and initiate an amplification loop— Ping-Pong cycle — to produce abundant antisense piRNAs complementary to active transposons (Brennecke et al., 2007; Gunawardane et al., 2007b). This is called the secondary piRNA pathway in germ line cells. In this model, cluster-derived piRNAs guide Aub to recognize active transposon

Figure 1.1

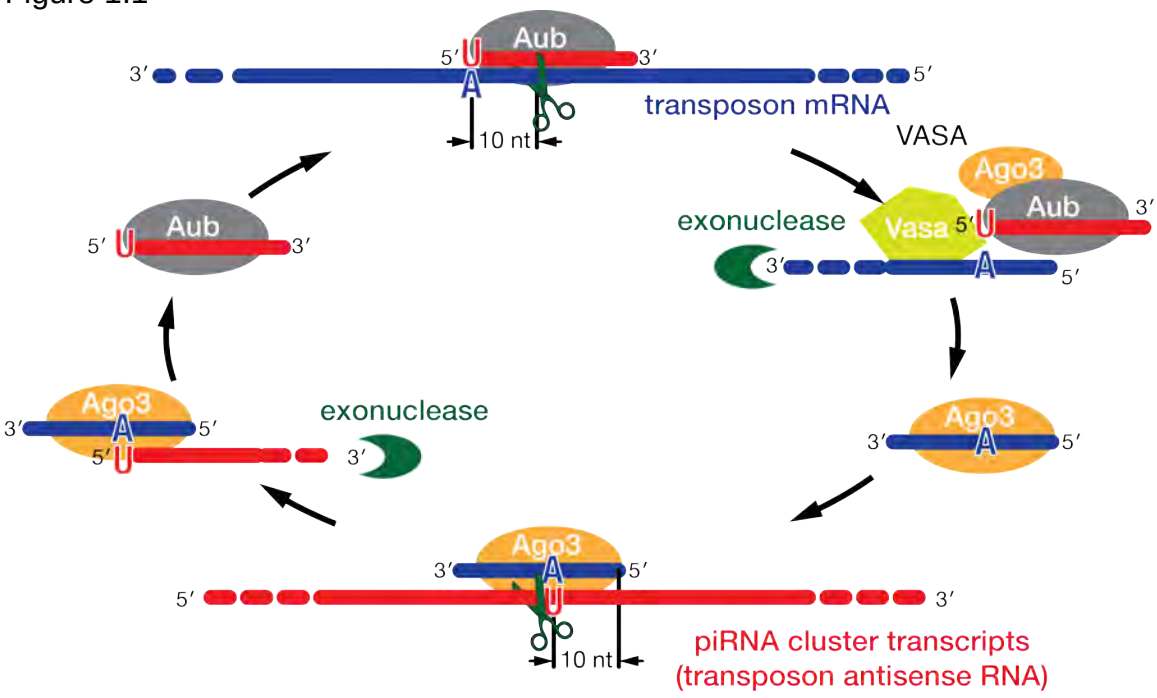

Aub
5′
U
A
3′
3′
5′
transposon mRNA
10 nt
VASA
Ago3
Aub
Vasa
exonuclease
3′
5′
Aub
5′
U
3′
Ago3
A
3′
5′
Ago3
A
3′
5′
exonuclease
5′
U
3′
Ago3
A
3′
5′
5′
3′
10 nt
piRNA cluster transcripts
(transposon antisense RNA)

Figure 1.1 Ping-Pong model.

transcripts. Upon extensive complementarity, Aub cleaves the target transcript at the phosphosdiester bond on target strand between g10 and g11, generating the 5′ end of a new, sense piRNA intermediate, which is transferred into Ago3. After the 3′ end of loaded intermediates are trimmed to a length ruled by Ago3, this mature sense piRNA subsequently directs Ago3 to bind to its homologous sequences in piRNA cluster transcripts, and its slicer activity can produce a new antisense Aub bound piRNA, completing the loop (Figure 1.1).

My dissertation mainly focuses on this sophisticated but intriguing Ping-Pong model. A peak at 10 nt overlap between 5′ end of Aub and Ago3 bound piRNAs indicates traces of PIWI protein cleavage activity (Brennecke et al., 2007; Gunawardane et al., 2007b). Aub bound piRNAs show a strong bias to uracil at their 5′ ends and Ago3 bound piRNAs are enriched with adenines at their 10^{th} positions. Those features corroborate the Ping-Pong signatures, though which one is the cause is not clear. The first part of my dissertation work attempted to answer this question. In addition, in vivo evidence of catalytic activity of Aub and Ago3 and their corresponding mutant phenotype have not been described. I will show that endonuclease activity of Aub and Ago3 is required for piRNA accumulation and Ping-Pong cycles in chapter IV. In the Ping-Pong process from silkworm cell line BmN4, the DEAD-box RNA helicase Vasa functioning as a RNA clamp assembles a complex on transposon transcript, with Siwi (Aub homolog in silkworm), Ago3 and Tudor protein Qin. After Siwi cleaves the transposon, Vasa facilitates the transfer of the 3′ cleavage product to Ago3

(Xiol et al., 2014). Without Qin, the interaction between Aub and Ago3 is weakened, and Ping-Pong occurs mainly between Aub:Aub, which fails to generate enough antisense piRNAs to silence transposons (Zhang et al., 2011; Xiol et al., 2014; Zhang et al., 2014). Qin might promote heterotypic Ago3:Aub Ping-Pong or suppress homotypic Ping-Pong. In chapter IV, we answer this question by comparing Aub:Aub Ping-Pong in *qin*$^{1/Df}$ single mutant versus *qin*$^{1/Df}$*,ago3*$^{t2/t3}$ double mutants. If extensive complementarity is required for PIWI's target cleavage, and if a transposon gains mutations in the piRNA target site to escape the scrutiny of piRNAs, how does silencing system control those transposons with newly acquired mutations? A newly discovered mechanism to generate phased Piwi piRNAs is elaborated in Chapter III.

piRNA functions: transposon silencing

How do piRNAs guide Piwi, Ago3 and Aub to regulate transposon mechanistically? Ping-Pong model proposed that Aub and Ago3 silence transposons through post-transcriptional gene silencing in germ cells. The cleavage activity of Ago3 produces antisense piRNAs from cluster transcripts. Those antisense piRNAs then fuel into Aub and direct Aub to silence transposon mRNAs post-transcriptionally. In this view, Ago3 silences transposons through Aub (Brennecke et al., 2007; Li et al., 2009; Huang et al., 2014).

Different from Aub, Piwi silences transposons at the transcriptional level. Piwi localizes to the nucleus and a mutant Piwi lacking its nuclear localization signal was trapped in the cytoplasm, leading to transposon desilencing (Saito et

al., 2009). Furthermore, slicer activity of Piwi is dispensable for transposon silencing, as a catalytically inactive Piwi rescues the null mutant phenotype (Saito et al., 2009; Sienski et al., 2012). Several studies have explored the role of Piwi in transcriptional silencing. Genome-wide studies were conducted in somatic follicle cells, ovarian germ cells or cultured, immortalized ovarian somatic cells (OSC) (Klenov et al., 2007; Klenov et al., 2011; Sienski et al., 2012; Huang et al., 2013; Le Thomas et al., 2013; Rozhkov et al., 2013). By comparing the RNA Pol II occupancy, euchromatic H3K9me3 marks on transposons, nascent transcribed RNA and steady state RNAs at a global level upon Piwi knockdown in OSC cell line, the Brennecke lab demonstrated that Piwi silences transposons at the transcriptional level (Sienski et al., 2012). The Aravin lab showed that knockdown of Piwi in fly germline lead to transposon derepression, which correlates with increased occupancy of Pol II on their promoters (Le Thomas et al., 2013). Many attempts to immunoprecipitate chromatin bound to Piwi failed (Huang et al., 2013; Lin et al., 2015; Marinov et al., 2015). Most likely, piRNAs guide Piwi to its target RNA loci, where it recruits enzymes that deposit repressive heterochromatin. The current model proposes that Piwi likely binds to nascent RNA via its antisense piRNA guide which appears to recruit Su(var)3-9, a histone methyltransferase that methylates histone H3 on lysine 9. These H3K9me3 marks bind heterochromatin protein 1 (HP1), generating chromatin that is seldom transcribed, as reflected by reduced occupancy with RNA polymerase II (Pol II). Piwi guided transcriptional gene silencing also requires the zinc-finger protein

GTSF-1/Asterix, which likely direct interacts with Piwi and is required for establishment of H3K9me3 (Dönertas et al., 2013; Muerdter et al., 2013).

piRNA functions beyond transposon silencing

A subset of piRNAs is also derived from the 3′ untranslated regions (3′ UTRs) of protein coding mRNAs (Robine et al., 2009). A specific example is that the 3′ UTR of the *traffic jam* (*tj*) gene in *Drosophila* produces sense piRNAs, excluding the possibility that those piRNAs target parental *tj* gene. Instead, via sequence complementarity, *tj*-derived piRNAs potentially target *piwi* and *Fasciclin 3 (Fas3) in trans*. Loss of TJ in gonadal somatic cells abolished Piwi expression. Without TJ or *tj*-derived piRNAs, somatic cells fail to intermingle with germ cells and Fasciclin 3 is ectopically overexpressed in testes (Saito et al., 2009). A study examining maternal *Nanos* (*nos*) mRNA decay in the *Drosophila* embryo showed that piRNAs from transposable elements could target protein-coding genes *in trans* with partial complementarity. CCR4-mediated deadenylation of *nos* depends on piRNA pathway components, Aub and Ago3, and piRNAs complementary to the 3′ UTR of *nos*. However, the silencing mechanism might be distinct from slicing and transcriptional gene silencing (Rouget et al., 2010). A recent study shows that a single W-chromosome-derived, female-specific piRNA is responsible for primary sex determination in the WZ sex determination system. This piRNA is produced from *Fem* piRNA precursor and can target a CCCH-type zinc finger protein *Masc*. In *Bombyx mori*, a gene named *doublesex* (*Bmdsx*) has four splicing isoforms: one male-specific splice variant and three female-specific

splice variants. Inhibition of *Fem*-derived piRNA-mediated signaling in female embryos leads to the production of the male-specific splice variants of *Bmdsx*. On the other hand, the silencing of *Masc* mRNA by this piRNA is required for the production of female-specific isoforms of *Bmdsx* in female embryos (Kiuchi et al., 2014).

In soma, piRNA pathway exhibits diversified functions. A study in *Aplysia* shows that piRNAs play roles in regulating memory storage in the brain by silencing CREB2, the major inhibitory constraint of memory formation (Rajasethupathy et al., 2012). *Oxytricha* piRNAs from the maternal nucleus map primarily to the somatic genome, specifying 5% of germline genomic regions for retention to in next generation (Fang et al., 2012). *Drosophila* Piwi forms a complex with heat-shock protein Hsp90 and the heat-shock organizing protein, Hop, *in vivo* to suppress phenotypic variation (Specchia et al., 2010).

mouse piRNAs

PIWI proteins and piRNAs are conserved in a wide range of eukaryotes, from sponges to humans, and they are expressed mainly in the gonads (Table 1.1). However, piRNA biogenesis pathways in different organisms also appear to be diverse. The mouse genome encodes three PIWI-clade Argonautes, MILI, MIWI2 and MIWI. All three PIWI proteins are temporally and spatially regulated during male germ cell differentiation and are essential for spermatogenesis. MILI expression begins in the embryonic testis (12.5 days post coitum, dpc) and

Table 1.1 Homologs of Piwi proteins

Drosophila melanogaster	Piwi	Aub	Ago3
mus musculus	MIWI	MILI	MIWI2
Homo sampiens	HIWI	HILI	HIWI2
Bombyx mori	N/A	Siwi	Ago3

persisting into adulthood — the round spermatid stage (Kuramochi-Miyagawa et al., 2004). MIWI2 begins to accumulate around 14.5-15.5 dpc, and starts to decline after birth. It eventually becomes undetectable around 4 days post partum (dpp), a time window correlated with cell cycle arrest and *de novo* methylation in primordial germ cells (Kuramochi-Miyagawa et al., 2008). MIWI is expressed in adult testes from 14 dpp, coinciding with the onset of the pachytene stage of meiosis (Deng and Lin, 2002). Deficiency in MILI or MIWI2 leads to activation of LINE and LTR retrotransposons including L1 and IAP elements, and spermatogenic stem cell arrest at early prophase of the first meiosis (Kuramochi-Miyagawa et al., 2004; Aravin et al., 2006; Kuramochi-Miyagawa et al., 2008). MIWI depletion also leads to L1 dysregulation and spermatogenesis is arrested at the early spermatid stage (Deng and Lin, 2002).

Three types of mouse piRNAs regulate cellular functions during spermatogenesis. Prenatal piRNAs originate from repetitive sequences; mRNA-derived piRNAs, expressed after birth and before the pachytene stage, are mainly derived from 3′ UTRs; pachytene piRNAs are produced from intergenic long non-coding transcripts (Li et al., 2013; Han and Zamore, 2014). Prenatal piRNAs and prepachytene piRNAs bind both MILI and MIWI2. The slicing activity of MILI is required to maintain L1 transposon silencing while MIWI2 transcriptionally silences transposons independent of its slicing acitivity (De Fazio et al., 2011). In embryonic testes, significant Ping-Pong occurs between MILI and MILI to feed piRNAs to MIWI2 (De Fazio et al., 2011). MIWI2, then, translocates

into the nucleus and recruits the DNA methyltransferase DNMT3L to methylate and repress transposon loci guided by piRNA in a sequence specific manner (Aravin et al., 2008; Watanabe et al., 2011). Pachytene piRNAs originate from piRNA clusters and bind to both MILI and MIWI in germ cells from pachytene spermatocytes to the round spermatid stage. The transcription of pachytene piRNA genes and other piRNA pathway components are regulated by an ancient transcription factor A-Myb. It initiates the piRNA production during early meiosis and lead to a rapid increase of their abundance (Li et al., 2013). The slicing activity of MIWI is required for LINE1 transposon silencing after birth while the Ping-Pong amplification of MIWI-associated piRNAs in adults does not depend on its slicing activity (Reuter et al., 2011; Beyret et al., 2012).

Chapter II the Study on the Preference of Tenth Adenine, Hallmark of Secondary piRNAs

PREFACE

This chapter was a product of a collaborative effort: Mayu Yoshikawa performed biochemical experiments and analyzed the results. Bo W. Han made small RNA libraries from immunoprecipitated Ago3 and Aub. He also made the degradome libraries and performed the shuffling analysis in the computational part. Natsuko Izumi performed the western blotting. I wrote the codes and did the rest of the analysis.

Summary

PIWI-interacting RNAs (piRNAs) silence transposons in animal germ cells. PIWI proteins bind and amplify piRNAs via the “Ping-Pong” pathway. Because PIWI proteins cleave RNAs between target nucleotides t10 and t11—the nucleotides paired to piRNA guide positions g10 and g11—the first ten nucleotides of piRNAs participating in the Ping-Pong amplification cycle are complementary. *Drosophila* piRNAs bound to the PIWI protein Aubergine typically begin with uridine (1U), while piRNAs bound to Argonaute3, which are produced by Ping-Pong amplification, often have adenine at position 10 (10A). The Ping-Pong model proposes that the 10A is a consequence of 1U. We find that 10A is not caused by 1U. Instead, fly Aubergine as well as its homologs, Siwi in silkworms and MILI in mice, has an intrinsic preference for adenine at the t1 position of their target RNAs; during Ping-Pong amplification; this t1A subsequently becomes the g10A of a piRNA bound to Argonaute3.

Introduction

PIWI-interacting RNAs (piRNAs) protect the genome of animal germ cells by silencing transposons and repetitive sequences (Aravin et al., 2007; Grimson et al., 2008; Houwing et al., 2008; Armisen et al., 2009; Friedlander et al., 2009; Kawaoka et al., 2009; Lau et al., 2009a). Ranging from 23–35 nt, piRNAs were first discovered in the *Drosophila melanogaster* testis, where they suppress expression of the repetitive *Stellate* locus, whose protein-product crystallizes in developing spermatocytes (Aravin et al., 2001), and in the ovary, where they thwart transposon expression (Aravin et al., 2003). PIWI proteins, which are an animal-specific clade of the Argonaute family of small RNA- or DNA-guided proteins found in all three domains of life (Cenik and Zamore, 2011), use piRNAs as guides that specify their RNA or DNA targets.

Both the PIWI and AGO clades of Argonaute proteins share a common domain architecture that allows them to bind target RNAs or DNAs via sequence complementarity to a 6–8 nt subsequence of the guide RNA or DNA, called the seed sequence. Argonaute proteins prepay a portion of the entropic cost by holding the seed sequence in a conformation that facilitates target binding (Lewis et al., 2003; Ma et al., 2005; Parker et al., 2005; Bartel, 2009; Parker et al., 2009; Wang et al., 2009; Frank et al., 2010; Boland et al., 2011; Lambert et al., 2011; Frank et al., 2012; Cora et al., 2014). Many Argonaute proteins catalyze endonucleolytic cleavage of their targets. Target cleavage requires far more extensive complementarity than binding between the guide and the target, in part

because extensive pairing enables a conformational change that brings the target closer to the Argonaute active site (Wang et al., 2008a; Wang et al., 2009). Anchoring of the 5′ phosphate of the guide in a phosphate-binding pocket ensures that the guide remains tightly bound to the Argonaute through many rounds of target cleavage (Parker et al., 2005; Yuan et al., 2005). Anchoring the 5′ phosphate in the pocket also helps position the target in the cleavage site: the scissile phosphate always lies between positions t10 and t11 of the target, which base-pair with positions g10 and g11 of the guide (Elbashir et al., 2001b; Elbashir et al., 2001c; Rivas et al., 2005).

In flies, repetitive, transposon-rich genomic regions, “piRNA clusters”, produce precursor transcripts that are subsequently processed into *primary* piRNAs (Brennecke et al., 2007). piRNA clusters reveal numerous, often nested, transposition events that keep a record of the past transposon exposure of the animal. In the fly germline, both strands of most piRNA clusters are transcribed. The transposon orientations in most piRNA clusters appear to be random, yet most piRNAs are antisense to transposon mRNAs.

The Ping-Pong model of piRNA biogenesis (Brennecke et al., 2007; Gunawardane et al., 2007a) attempts to explain how the initial piRNA population is amplified and how the antisense bias is acquired and thereby increases the number of piRNAs that can silence transposon mRNAs. The model proposes that primary or maternally deposited antisense piRNAs are loaded into the PIWI protein Aubergine (Aub), directing it to bind and cleave complementary targets,

including both piRNA cluster transcripts and transposon mRNAs (Brennecke et al., 2008). The model further proposes that the PIWI protein Argonaute3 (Ago3) binds the monophosphorylated 5′ end of the 3′ cleavage product made by Aub; the 3′ end of this product is further processed to generate an Ago3-bound, *secondary* piRNA. The piRNA can then guide Ago3 to bind and cleave another piRNA cluster transcript; the resulting 3′ cleavage product is then loaded into Aub and processed into a new Aub-bound, secondary piRNA. In turn, the Aub-bound piRNA can initiate another cycle of Ping-Pong amplification.

The Ping-Pong model seeks to explain three remarkable features of piRNAs: (1) Aub-bound piRNAs are typically antisense to transposon mRNAs, while Ago3-bound piRNAs are usually sense; (2) the first 10 nucleotides of Aub-bound piRNAs are often complementary to the first 10 nucleotides of Ago3-bound piRNAs, a relationship termed the “Ping-Pong signature”; and (3) Aub-bound piRNAs often begin with a uracil (g1U), whereas Ago3-bound piRNAs show no first nucleotide bias (g1N), but tend to have an adenosine as their tenth nucleotide (g10A). The Ping-Pong model proposes that the g1U of an Aub-bound piRNA selects a target RNA with a corresponding t1A nucleotide. When the 3′ cleavage product generated by Aub is loaded into Ago3, its former t1 position becomes the g10 position, generating the g10A signature characteristic of Ago3-bound piRNAs.

Here, we report that Aub, as well as its homologs Siwi in *Bombyx mori* and MILI in *Mus musculus*, selects targets bearing a t1A nucleotide irrespective of the

identity of the g1 nucleotide of its piRNA guide. Consequently, the g10A of Ago3-bound piRNAs arises not through g1U:t1A base pairing, but rather reflects an intrinsic physical property of the Aub protein. A similar preference for t1A was noted earlier for mammalian AGO proteins guided by microRNAs (miRNAs) (Lewis et al., 2005). In contrast, fly and silkworm Ago3 proteins do not have the nucleotide preference at the t1 position. We propose that both mammalian AGO proteins and a subset of mammalian and insect PIWI proteins contain a binding pocket that best accommodates adenine at the t1 position of their targets. In this view, the tendency of Aub-bound piRNAs to begin with a U is, at least in part, a consequence, not a cause, of the g10A of Ago3-bound piRNAs.

Results

Cause and Effect in the Ping-Pong Model

The Ping-Pong model proposes that, guided by an antisense primary piRNA, Aub binds a transposon mRNA target and cuts it between nucleotides t10 and t11, the target nucleotides that pair with nucleotides g10 and g11 of the piRNA guide, and that the resulting 5′ monophosphorylated, 3′ cleavage product is loaded into Ago3. The model states that the 5′ g1U of the guide piRNA bound to Aub dictates the t1A of the cleaved target. After the target is converted to a secondary piRNA loaded into Ago3, the t1A becomes a g10A which is the hallmark of Ago3-bound piRNAs (Figure 2.1, Model I) (Brennecke et al., 2007; Gunawardane et al., 2007a).

Despite the model's appeal, it conflicts with a property of Argonaute proteins that has been conserved through several billion years of evolution: the first nucleotide of a small RNA or DNA guide bound to a eubacterial, archael, or eukaryotic Argonaute protein is anchored in a 5′ phosphate-binding pocket that precludes the first base of the guide (g1) from pairing with a complementary base on a target (t1) (Ma et al., 2005; Parker et al., 2005; Wang et al., 2009; Frank et al., 2010; Boland et al., 2011; Elkayam et al., 2012; Frank et al., 2012; Cora et al., 2014). Biochemical studies show that a first position mismatch does not impair and can even enhance Argonaute-directed target cleavage for both AGO and PIWI

Figure 2.1

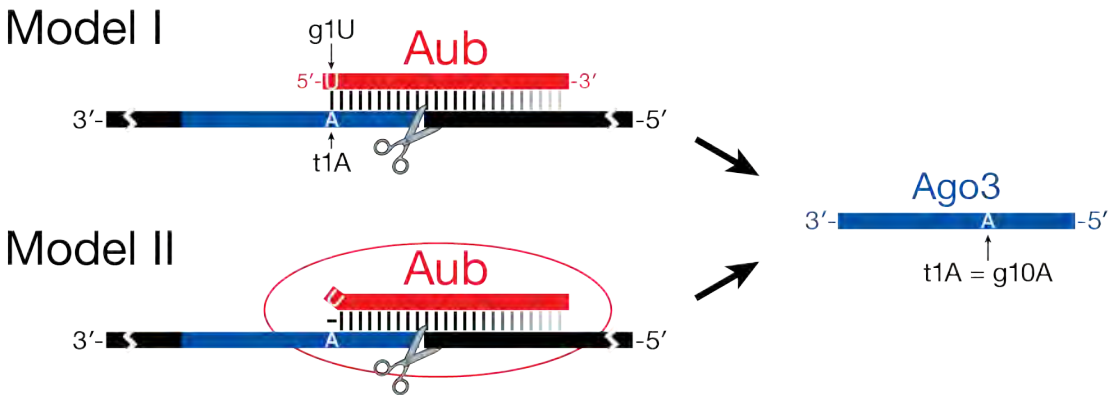


Figure 2.1. Two Models for the Enrichment of Adenine at the Tenth Position of Ago3 piRNAs

Model I posits that g1U causes t1A by Watson and Crick pairing. This t1A becomes g10A after the 3′ cleavage product is loaded into Ago3. Model II accommodates the biophysical interactions between the 5′ phosphate of guide piRNAs and the phosphate-binding pocket in Aub. A g1U is not the cause of t1A in this scenario.

proteins (Haley and Zamore, 2004; Haley and Zamore, 2004; Reuter et al., 2011; Wee et al., 2012).

In fact, the piRNA Ping-Pong model is not the first to grapple with the origin of a target t1A across from a g1U in an Argonaute-bound guide. Most mammalian miRNAs begin with uridine. Computational efforts to predict conserved miRNA targets in vertebrates reveal that high-confidence miRNA-binding sites typically bear a t1A across from the g1U position. However, the t1A characteristic of high-confidence miRNA-binding sites persists even when the miRNA begins with adenosine, cytosine, or guanosine, rather than changing to the complementary nucleotide (Lewis et al., 2005; Grimson et al., 2007). This unexpected observation likely reflects the presence of a binding site in Argonautes for the t1 base, a binding site that best accommodates adenine. Thus, the g10A of Ago3-bound piRNAs might reflect a propensity of Aub to bind targets bearing adenosine at the t1 position, rather than a consequence of the g1U of Aub-bound piRNAs (Figure 2.1, Model II).

Cis- and *trans*-Targets

Assuming that Ago3-bound piRNAs result from the target cleavage by Aub-bound piRNAs, we can infer the targets and target cleavage sites of Aub-bound piRNAs by examining Ago3-bound piRNAs. We immunoprecipitated Ago3 and Aub from wild-type (w^1) fly ovaries and sequenced the piRNAs bound to each protein. We used piRNAs uniquely bound to Ago3 (76% of species and 22% of reads of all Ago3 bound piRNAs), i.e., excluding piRNAs that were detected in

both Aub and Ago3 immunoprecipitates, to infer the original targets of the guide piRNAs uniquely bound to Aub (60% of species and 17% of reads of all Aub bound piRNAs). Such targets fall into two categories: *cis*-targets, which overlap their targets in genomic coordinates and *trans*-targets that do not. A *cis* target and its guide, for example, could correspond to the two precursor transcripts from the opposite genomic strands of a dual-strand piRNA cluster. In contrast, *trans*-targets could correspond to mRNAs transcribed from euchromatic transposon insertions.

For Aub-bound piRNAs and Their *cis*-Targets, only g1U:t1A and not Other Base Pairs Generates a Ping-Pong Signature

In its simplest form, Model I predicts that the g1 nucleotide of an Aub-bound piRNA should always pair with the t1 nucleotide of its *cis* target. Nearly half of Aub-bound piRNAs begin with a nucleotide other than U (Figure 2.2A). To test whether Aub-bound piRNAs beginning with A, C, or G participate in the Ping-Pong amplification cycle, we analyzed the 5′-to-5′ distance between each Aub-bound piRNA and its overlapping Ago3-bound piRNA from the opposite genomic strand (Figure 2.2B). piRNA:target overlaps of 1–9 and 11–16, i.e., non-Ping-Pong overlaps, were used as the background distribution to compute Ping-Pong *Z*-scores. Given that a small number of piRNA species could dominate the Z-score, we performed the analysis on both paired species and paired reads. As anticipated, the number of pairs of Aub-bound g1U piRNAs with t1A target partners was significantly greater than the background distribution attributable to

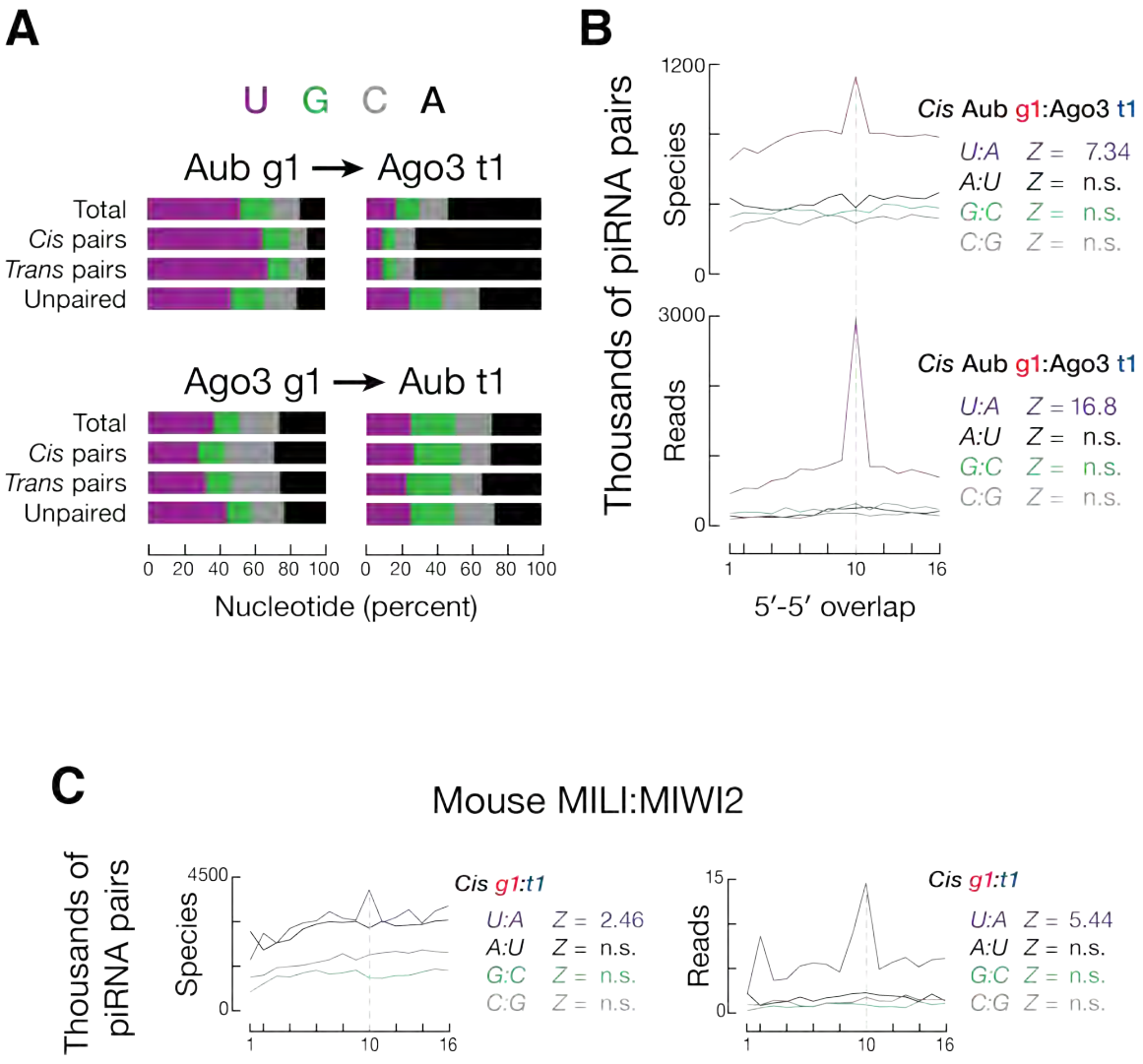
chance (*Z*-score =16.8 for reads and 7.34 for species). Yet the other three complementary pairs g1A:t1U, g1C:t1G, and g1G:t1C were not significantly different from their background distributions (Figure 2.1B).

Similarly, mouse MILI also showed significant g1U:t1A Ping-Pong with its *cis*-targets inferred from the piRNAs bound to MIWI2, but not for the three other complementary g1:t1 pairs (Figure 2.2C). In other words, g1V (i.e., not U) piRNAs bound to fly Aub or mouse MILI are not significantly amplified by the Ping-Pong cycle. We can imagine two explanations for this surprising observation: (1) efficient endonucleolytic cleavage catalyzed by Aub or MILI requires a g1U (Figure 2.1; Model I) or (2) Aub and MILI prefer to bind and cleave target RNAs that bear a t1A (Figure 2.1; Model II).

Trans-Targets of Aub Reveal Its Intrinsic Preference for t1A Targets

To distinguish between these two explanations for the paucity of piRNA:*cis* target interactions other than g1U:t1A, we analyzed the *trans*-targets of Aub-bound piRNAs. Standard Ping-Pong analyses measure the distance between the 5′ ends of piRNAs that overlap in genomic space (Brennecke et al., 2007; Klattenhoff et al., 2009; Li et al., 2009). Consequently, such analyses only detect piRNA:*cis*-target pairs. *Cis*-targets are uninformative for our

Figure 2.2


A
U G C A
Aub g1 → Ago3 t1
Total
Cis pairs
Trans pairs
Unpaired
Ago3 g1 → Aub t1
Total
Cis pairs
Trans pairs
Unpaired
0 20 40 60 80 100
0 20 40 60 80 100
Nucleotide (percent)
B
Thousands of piRNA pairs
Species
1200
0
Cis Aub g1:Ago3 t1
U:A Z = 7.34
A:U Z = n.s.
G:C Z = n.s.
C:G Z = n.s.
Reads
3000
0
Cis Aub g1:Ago3 t1
U:A Z = 16.8
A:U Z = n.s.
G:C Z = n.s.
C:G Z = n.s.
1
10
16
5′-5′ overlap

C
Mouse MILI:MIWI2
Thousands of piRNA pairs
Species
4500
0
Cis g1:t1
U:A Z = 2.46
A:U Z = n.s.
G:C Z = n.s.
C:G Z = n.s.
1
10
16
Reads
15
0
Cis g1:t1
U:A Z = 5.44
A:U Z = n.s.
G:C Z = n.s.
C:G Z = n.s.
1
10
16

Figure 2.2. For *Cis* pairs, only g1U:t1A pair generates a Ping-Pong signature

(A) Nucleotide composition of g1 from guide piRNAs and t1 (g10) from their inferred targets for Ping-Pong analysis between Aub and Ago3 and between Ago3 and Aub in w^1 ovaries.

(B) Numbers of g1U:t1A, g1A:t1U, g1C:t1G and g1G:t1C Aub:Ago3 *cis* Ping-Pong pairs.

(C) Numbers of g1U:t1A, g1A:t1U, g1C:t1G and g1G:t1C MILI:MIWI2 *cis* Ping-Pong pairs.

purposes: the g1 nucleotide of a piRNA is always complementary to the t1 of a *cis*-target, because the two strands of genomic DNA are always complementary. The biochemical properties of Argonaute proteins suggest that piRNA-directed cleavage requires extensive, but not complete complementarity between the piRNA and its target (Ameres et al., 2007; Reuter et al., 2011; Wee et al., 2012). In the crystal structure of *Thermus thermophilus* Argonaute bound to a DNA guide and a 19-nt RNA target, guide base g16 stacks over a tyrosine and target base t16 stacks over a proline, blocking the propagation of the guide:target duplex beyond position 16 (Wang et al., 2009). To test the extent of complementarity required for *trans* pairing, we examined the length of complementarity between Aub-bound guide piRNAs and their potential targets inferred from Ago3-bound piRNAs minimally with complementary bases at positions 2–10. We compared the observed frequency of piRNA:target pairing at each position between 11–23 to the frequency expected by chance (see Experimental Procedures). To estimate the minimum complementarity required by *trans* Ping-Pong, we also computed the frequency of contiguous piRNA:target base pairing from position 2 to each position between 11–23 but not beyond, and compared it to the frequency expected by chance. These data suggest that the authentic *trans*-targets of Aub-bound piRNA are generally complementary to their guides from g2 until at least position g16 (Figure 2.4A). Similarly, our analysis suggests that Ago3 can slice a target when piRNA:target complementarity extends from position g2 to at least position g14 (Figure 2.4B). Although our

results for contiguous pairing are consistent with Aub and Ago3 tolerating one or more mismatches within their required regions of complementarity, we restricted our analysis of predicted *trans*-targets to those with complete complementarity from positions g2–g16 (Figure 2.3A and Figure 2.3B). Using this metric, we could infer *cis* or *trans* targets from the Ago3 piRNA population for 35% of piRNAs uniquely bound to Aub. Conversely, we could identify *cis* or *trans* targets (inferred from Aub-bound piRNAs) for 52% of piRNAs uniquely bound to Ago3. Among all possible Aub-bound piRNA:target pairs, 33% were unambiguouly identified as *trans*-targets.

To test whether Aub requires a g1U to cleave its *trans*-targets, we analyzed the *trans*-targets of Aub-bound piRNA beginning with A, C, or G. Among piRNAs uniquely bound to Aub, nearly 48% begin with a nucleotide other than U: A, 14.3%; C, 15.9%; G, 17.6% (Figure 2.2A). For those Aub-bound piRNAs with Ping-Pong partners (i.e., having a 10-nt 5′-to-5′ overlap), ~32% begin with A, C, or G. Thus, g1V piRNAs bound to Aub appear to function as guides for *trans*-targets.

To test whether Aub prefers to bind and cleave target RNAs that bear a t1A, we asked whether these g1V, Aub-bound piRNAs were complementary to their *trans*-targets at the t1 position. Among g1V piRNA the most significant t1 nucleotide for *trans*-targets is t1A: mismatches such as g1A:t1A, g1C:t1A, g1G:t1A are far more likely than expected by chance, while the frequency of the

Figure 2.3

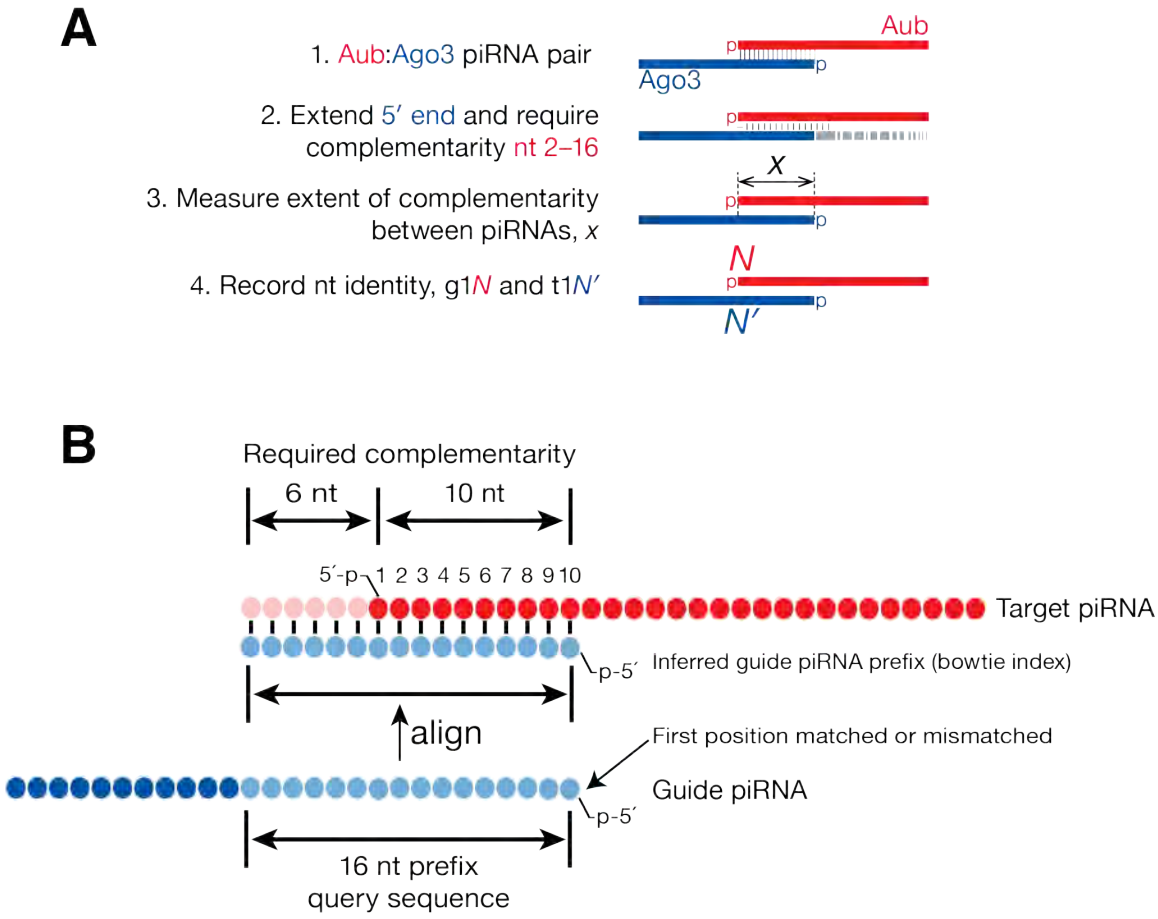

A
1. Aub:Ago3 piRNA pair
2. Extend 5′ end and require complementarity nt 2–16
3. Measure extent of complementarity between piRNAs, x
4. Record nt identity, g1N and t1N′
Aub
Ago3
p
x
N
N′
B
Required complementarity
6 nt
10 nt
5′-p
1 2 3 4 5 6 7 8 9 10
Target piRNA
p-5′
Inferred guide piRNA prefix (bowtie index)
align
First position matched or mismatched
Guide piRNA
16 nt prefix
query sequence

Figure 2.3. Schema for detecting *trans*-targets

(A) Pipeline for detecting *trans* pairs with 15 nt complementarity and allow at most 1 mismatch at g1.

(B) First, the 5′ end of target piRNA (in red) is extended to its upstream 6 nt (pink) if the 5′-5′ overlap between guide and target is 10 nt and the required complementarity is 16 nt. The extended 5′ end is the new start site and 16 nt sequence is extracted starting from the new start site. By reverse complementarity, an inferred guide piRNA prefix (light blue) is then Bowtie indexed. All the collapsed real guide piRNA prefixes are then aligned to the inferred guide piRNA indexes by Bowtie, allowing up to one mismatch at the first position.

complementary pairs, g1A:t1U, g1C:t1G, and g1G:t1C, are nearly indistinguishable from background (Figure 2.5A). Other mismatched g1V:t1B (B = not A) pairs are non-significant (g1A:t1G and g1G:t1G; a *Z*-score = 2.96 corresponds to a Bonferroni corrected, two-tailed *p*-value of 0.05) or significant but much closer to the background distribution than g1V:t1A (g1A:t1C, g1C:t1C, g1C:t1U and g1G:t1U). All the abundant piRNA species are removed from our analysis since piRNA species shared by both Ago3 and Aub tend to be more abundant. This raised a possibility that a variation at g1 position of those super abundant piRNA species (likely due to a sequencing error) passed the unique assignment cutoff, which may bias our analysis. However, all the analyzed piRNAs are perfectly mapped to the genome, this maximally reduced the likelihood that a piRNA species is from sequencing errors at certain position. Among 2,050,844 Aub bound piRNA species, read by 20,938,987 times in total, only 11,577 (0.6%) species have alterations at g1 position, corresponding to 1,645,988 reads (8%). Among those 0.6% of the species from Aub piRNAs sharing their g2-3′ end, 52.8% of them in species (60.4% in reads) are either with all the siblings shared by both Ago3 and Aub, or the unique one does not involve in Ping-Pong. After remove those species with g1 siblings and one of them (mostly the less abundant one) was in Ping-Pong with piRNAs uniquely bound to one PIWI protein, we repeated our *cis* and *trans* analysis. We observed the same results. These data suggest that an A at the t1 position of a target RNA

Figure 2.4

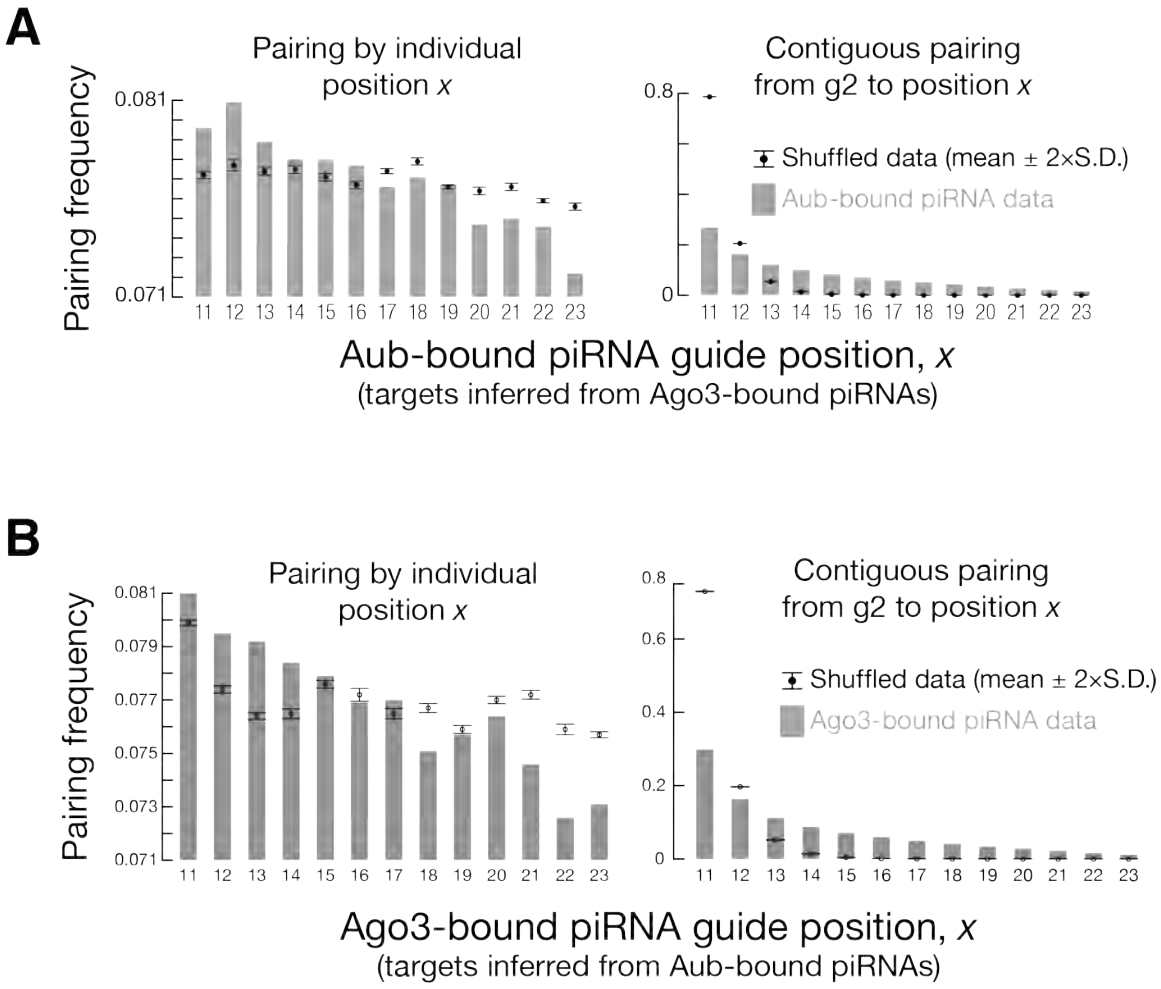

A
Pairing by individual position x
Contiguous pairing from g2 to position x
Pairing frequency
0.081
0.071
0.8
0
Shuffled data (mean ± 2×S.D.)
Aub-bound piRNA data
11 12 13 14 15 16 17 18 19 20 21 22 23
Aub-bound piRNA guide position, x
(targets inferred from Ago3-bound piRNAs)
B
Pairing by individual position x
Contiguous pairing from g2 to position x
Pairing frequency
0.081
0.079
0.077
0.075
0.073
0.071
0.8
0.6
0.4
0.2
0
Shuffled data (mean ± 2×S.D.)
Ago3-bound piRNA data
11 12 13 14 15 16 17 18 19 20 21 22 23
Ago3-bound piRNA guide position, x
(targets inferred from Aub-bound piRNAs)

Figure 2.4. Computational evaluation of the extent of complementarity required for Ping-Pong *trans*-pairing

(A) Aub requires g2–g16 complementarity to cleave a target RNA. The frequency of pairing for g11–g16 is higher in biological data than when the randomization for individual positions (left). When examining the extent of contiguous pairing rather than the frequency of pairing at individual positions, the pairing frequency for the biological data was only higher than the pairing frequency in the shuffled controls when base pairing extended from position g2 to at least position g13. Error bars (mean ± 2 × S.D., n = 10) indicate the paired frequency for the shuffled controls.

(B) Ago3 requires contiguous complementarity from position g2 to at least position g14 of its targets. The frequency of base pairing of individual positions (left) or contiguous pairing (right) for piRNA guides bound to Ago3 and targets inferred from piRNAs bound to Aub was compared to the frequency of base pairing between a piRNA guide bound to Ago3 and targets inferred from piRNAs bound to Aub for which the sequence of positions g11–g23 were shuffled among the piRNAs. Contiguous pairing from position g2 to at least position g13 was required for the frequency of pairing in the biological data to be greater than that of the shuffled controls. Black error bars (mean ± 2 × S.D., n = 10) indicate the paired frequency for the shuffled controls.

is an intrinsic property of Aub itself and does not reflect pairing between the g1 position of a guide and the corresponding t1 position of its target.

We also detected a preference for t1A, irrespective of the identity of the g1 nucleotide, for piRNAs uniquely bound to MILI and targets inferred from piRNAs uniquely bound to MIWI2 in the embryonic mouse testis (Figure 2.7B, left). We also detected a weaker preference for t1G, suggesting that MILI may select targets with a purine (A or G) at the t1 position. When we required 19-nt (g2–g20) complementarity, the t1A preference for fly Aub and the purine preference for mouse MILI persisted (Figure 2.8).

Catalytically Inactive Ago3^{ADH} Accumulates t1A Secondary piRNAs

To further test whether Aub prefers t1A targets irrespective of the identity of the g1 nucleotide of its piRNA guide, we engineered flies in which the Ping-Pong amplification cycle was abridged by disabling the catalytic activity of Ago3. These flies express mRNA encoding a catalytically inactive Ago3 protein (Ago3^{ADH}) from two transgenes, one controlled by the *aub* promoter and the other controlled by the *vasa* promoter, in an *ago3^{t2}/ago3^{t3}* mutant background; the *aub* and *vasa* promoters are active only in the germline and the combination of these two promoters drive Ago3^{WT} in *ago3* mutant background rescue fertility better than just single type of promoter (see chapter IV). Ago3^{ADH} is predicted not to cleave RNA (Liu et al., 2004). Therefore, these flies are expected not to sustain the Ping-Pong amplification cycle: Aub-bound piRNAs correspond only to primary piRNAs and maternal piRNAs and Ago3^{ADH} bound piRNAs correspond to the first

cycle of secondary piRNAs produced by Aub-catalyzed cleavage; no Aub-bound secondary piRNAs can be produced by Ago3.

In flies lacking Ago3, Aub:Aub homotypic Ping-Pong replaces normal, productive heterotypic Aub:Ago3 Ping-Pong (Li et al., 2009; Zhang et al., 2011). In contrast, the presence of Ago3^{ADH} not only blocks multiple cycles of heterotypic Ping-Pong, but also serves to suppress homotypic Aub:Aub Ping-Pong, because Ago3^{ADH} can accept cleavage products from Aub. Indeed, we observed much weaker overall Ping-Pong in the *ago3^{ADH}* ovaries (P_{vas}::*ago3^{ADH}* /P_{aub}::*ago3^{ADH}*; *ago3^{t2}*/*ago3^{t3}*, *Z*-score = 11.7) than in ovaries lacking Ago3 altogether (*ago3^{t2}*/*ago3^{t3}*, *Z*-score = 50.2). Specifically, *cis* Ping-Pong between piRNAs uniquely bound to Aub and piRNAs uniquely bound to Ago3^{ADH} (*Z*-score = 3.30) was far weaker than *cis* Ping-Pong between piRNAs uniquely bound to Aub and piRNAs uniquely bound to Ago3 (*Z*-score = 16.8; Figure 2.2B). Using piRNAs uniquely bound to Ago3^{ADH} to infer *trans*-targets, g1U:t1A, g1A:t1A and g1C:t1A were the only piRNA:*trans*-target pairs significantly different from the background (Figure 2.6). These data further support the view that Aub has an intrinsic preference for t1A irrespective of the identity of the g1 nucleotide of its piRNA guide.

In all of these analyses, *trans*-targets were inferred from the piRNAs

Figure 2.5

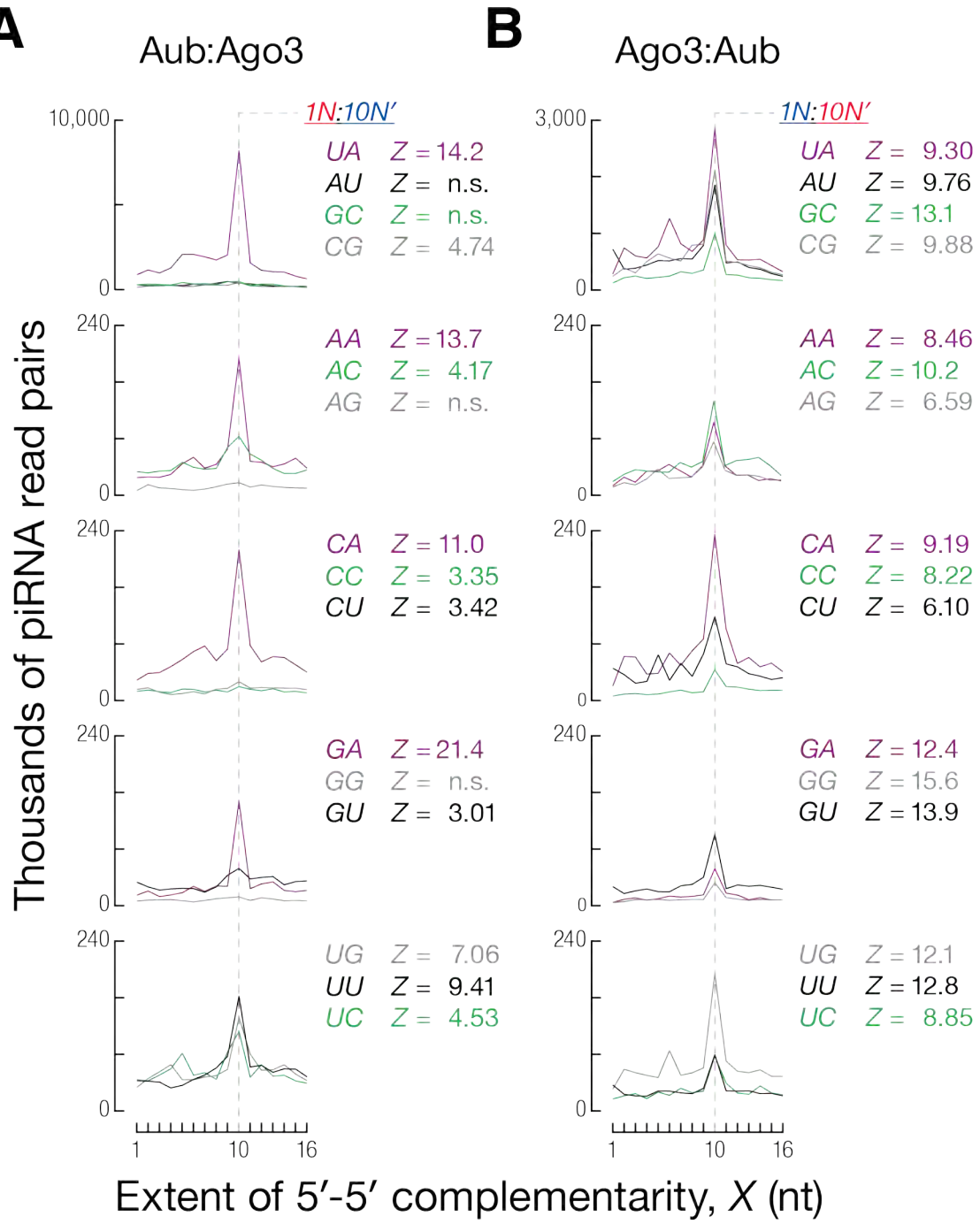

A
Aub:Ago3
B
Ago3:Aub
1N:10N′
10,000
3,000
0
240
UA Z = 14.2
AU Z = n.s.
GC Z = n.s.
CG Z = 4.74
AA Z = 13.7
AC Z = 4.17
AG Z = n.s.
CA Z = 11.0
CC Z = 3.35
CU Z = 3.42
GA Z = 21.4
GG Z = n.s.
GU Z = 3.01
UG Z = 7.06
UU Z = 9.41
UC Z = 4.53
UA Z = 9.30
AU Z = 9.76
GC Z = 13.1
CG Z = 9.88
AA Z = 8.46
AC Z = 10.2
AG Z = 6.59
CA Z = 9.19
CC Z = 8.22
CU Z = 6.10
GA Z = 12.4
GG Z = 15.6
GU Z = 13.9
UG Z = 12.1
UU Z = 12.8
UC Z = 8.85
1
10
16
Thousands of piRNA read pairs
Extent of 5′-5′ complementarity, X (nt)

Figure 2.5. *Trans* Ping-Pong Analysis Reveals t1A Preference for Aub

(A) Ping-Pong profile for each guide: target *trans* pair with at least 16 nt complementarity. Aub piRNAs as guides and Ago3 piRNAs as targets.

(B) Ago3 piRNAs as guides and Aub piRNAs as targets.

bound to wild-type or mutant Ago3. Of course, not all *trans*-targets cleaved by Aub are expected to escape destruction and generate an Ago3-bound secondary piRNA. To more accurately capture the targets cleaved by Aub, we sequenced the 5′ monophosphorylated RNAs that are longer than 200 nt in w^1 ovaries. We obtained 4,019,873 such “degradome” reads matching the fly genome; most (63.9%) corresponded to fragments of mRNAs, presumably decay intermediates from mRNA turnover pathways. An additional 2.3% mapped to transposons (Extended Experimental Procedure). Among the transposon-mapping degradome RNAs, 24.8% showed the 10-nt 5′-5′ overlap with Aub-bound piRNAs expected for Aub-catalyzed, piRNA-directed 3′ cleavage fragments. Of these, 34% were *trans*-targets.

Among these putative Aub-catalyzed, piRNA-directed 3′ cleavage products, g1U:t1A (*Z*-score = 22.5), g1A:t1A (*Z*-score = 15.5), g1C:t1A (*Z*-score = 14.7) and g1G:t1A (*Z*-score = 5.61) piRNA:*trans*-target pairs were both significantly different from background and more abundant than other types of pairings. One additional pair, g1C:t1U (*Z*-score = 5.45) was also significantly different from background, but was only 23% as abundant as the g1C:t1A pairs. Thus, both when we inferred 1) the *trans*-targets of Aub-bound piRNAs from the secondary piRNAs-bound to wild-type or catalytically inactive Ago3, or 2) identified targets directly from degradome sequencing, we find that Aub selects targets bearing a t1A nucleotide irrespective of the identity of the g1 position of its piRNA guide (Figure 2.6).

Figure 2.6

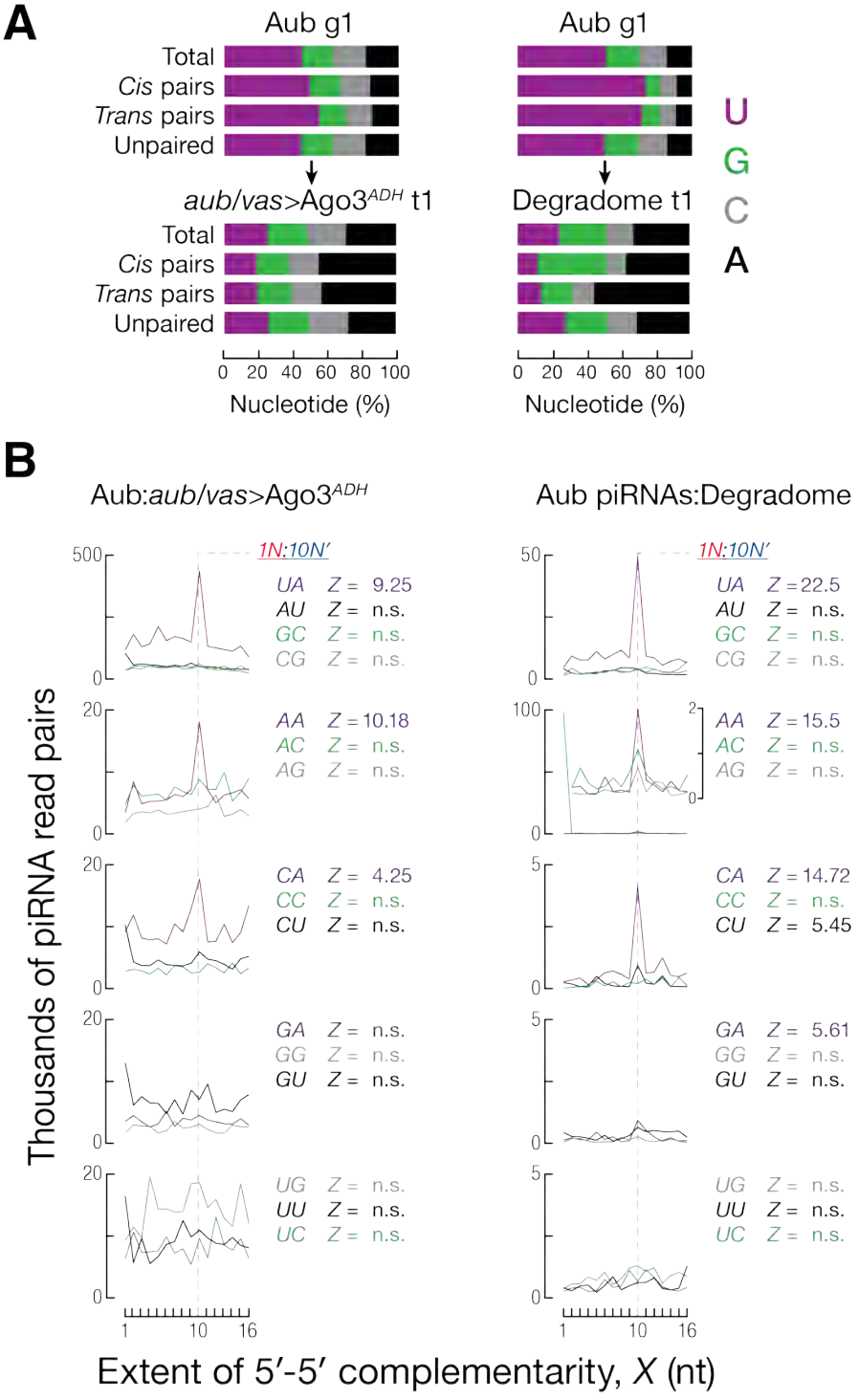

A
Aub g1
Total
Cis pairs
Trans pairs
Unpaired
aub/vas>Ago3ADH t1
Degradome t1
0 20 40 60 80 100
Nucleotide (%)
U
G
C
A
B
Aub:aub/vas>Ago3ADH
Aub piRNAs:Degradome
Thousands of piRNA read pairs
1N:10N′
UA Z = 9.25
AU Z = n.s.
GC Z = n.s.
CG Z = n.s.
AA Z = 10.18
AC Z = n.s.
AG Z = n.s.
CA Z = 4.25
CC Z = n.s.
CU Z = n.s.
GA Z = n.s.
GG Z = n.s.
GU Z = n.s.
UG Z = n.s.
UU Z = n.s.
UC Z = n.s.
UA Z = 22.5
AU Z = n.s.
GC Z = n.s.
CG Z = n.s.
AA Z = 15.5
AC Z = n.s.
AG Z = n.s.
CA Z = 14.72
CC Z = n.s.
CU Z = 5.45
GA Z = 5.61
GG Z = n.s.
GU Z = n.s.
UG Z = n.s.
UU Z = n.s.
UC Z = n.s.
1 10 16
Extent of 5′-5′ complementarity, X (nt)

Figure 2.6. The t1A Preference for Aub is Reinforced in Aub:Ago3^{ADH} and Aub:degradome guide:target pairs

(A) Nucleotide composition of position g1 of guide piRNAs and position t1 of their targets as inferred from piRNAs bound to Ago3^{ADH} or detected directly in the RNA degradome of *w*^{1} ovaries.

(B) Guide piRNAs and their inferred targets were determined by sequencing the piRNAs in immunoprecipitates of Aub or Ago3^{ADH} from P_{vas}::Ago3^{ADH}/P_{aub}::Ago3^{ADH};*ago3*^{t2}/*ago3*^{t3} ovaries. Degradome reads were from *w*^{1} ovaries.

nucleotide. We conclude that Ago3 requires only sufficient complementarity from piRNA position g2 to g14–g16 or beyond to bind and cleave its RNA targets.

MIWI2, unlike Ago3, Prefers a Purine

In both insects and mammals, the Ping-Pong pathway serves to amplify transposon-silencing piRNAs. In mice, homotypic MILI:MILI Ping-Pong appears to play the major role in piRNA amplification, perhaps explaining why mouse transposon-silencing piRNAs are sense biased (Aravin et al., 2008; De Fazio et al., 2011). Heterotypic MILI:MIWI2 Ping-Pong does take place, but in vivo data suggest that the process is not reciprocal, with MIWI2:MILI Ping-Pong dispensable for transposon-silencing (De Fazio et al., 2011).

We used publicly available data to infer the *trans*-targets of MIWI2 piRNAs from the piRNAs bound to MILI. Unlike Ago3 but like MILI, MIWI2 favors a purine at the t1A position (Figure 2.7A). Consistent with this preference, MILI-bound piRNAs favor a purine at their g10 position, the nucleotide that serves as t1 for MIWI2 (Figure 2.7B, right); no such t1 (g10) bias was detected for Aub-bound piRNAs in fly ovaries (Figure 2.2A and 2.5B).

Ago3 has no t1A Preference

Unlike Aub, Ago3 appears to have no t1A preference (Figure 2.2A). We inferred the *trans*-targets of Ago3 from the piRNAs bound to Aub. Unlike Aub-bound piRNAs, Ago3-bound piRNAs showed no t1 nucleotide preference among their inferred targets: all 16 g1:t1 Ago3:*trans*-target pairs showed significant *Z*-scores

Figure 2.7

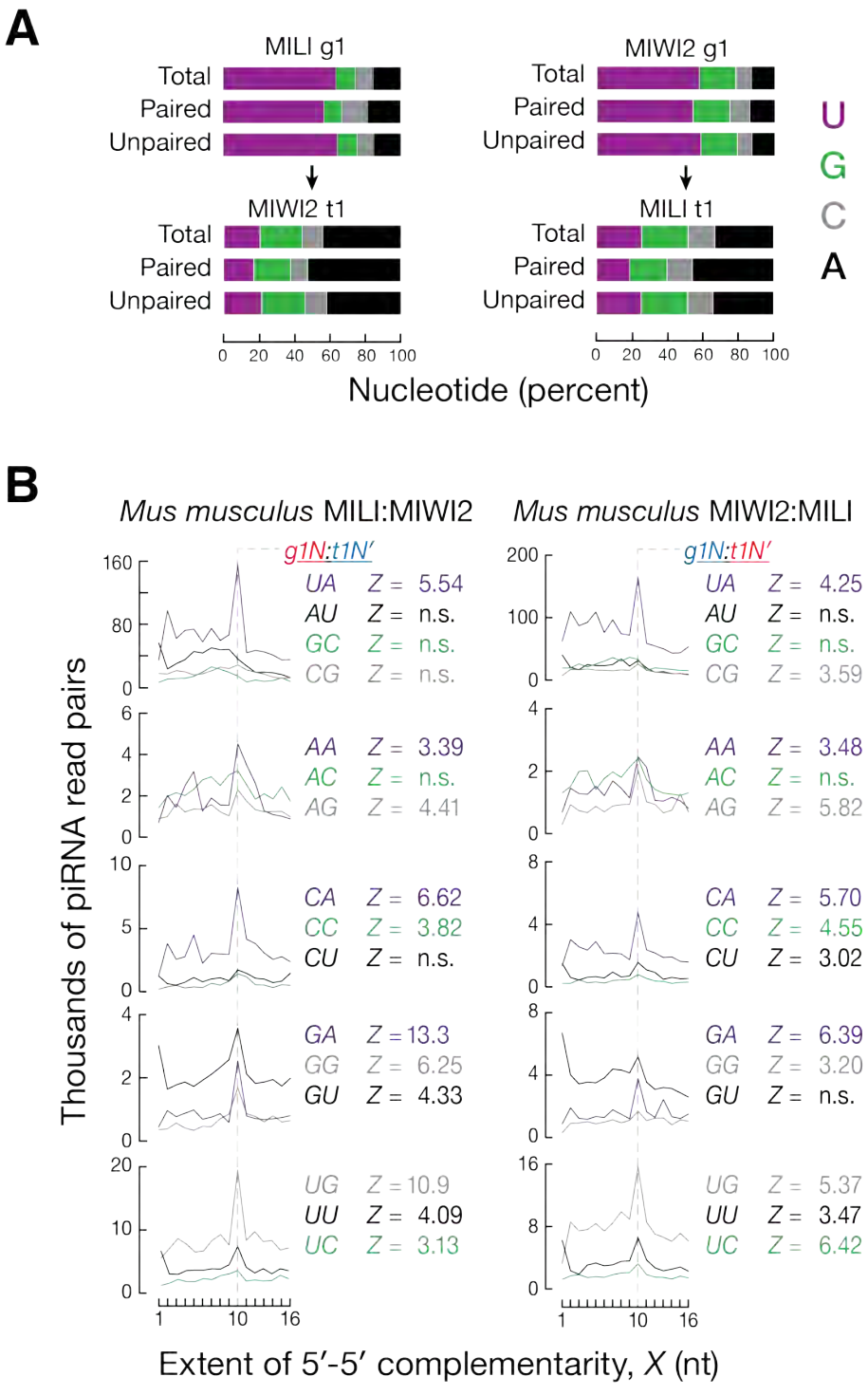

A
MILI g1
Total
Paired
Unpaired
MIWI2 t1
Total
Paired
Unpaired
0 20 40 60 80 100
MIWI2 g1
Total
Paired
Unpaired
MILI t1
Total
Paired
Unpaired
0 20 40 60 80 100
U
G
C
A
Nucleotide (percent)
B
Mus musculus MILI:MIWI2
Mus musculus MIWI2:MILI
g1N:t1N′
UA Z = 5.54
AU Z = n.s.
GC Z = n.s.
CG Z = n.s.
AA Z = 3.39
AC Z = n.s.
AG Z = 4.41
CA Z = 6.62
CC Z = 3.82
CU Z = n.s.
GA Z = 13.3
GG Z = 6.25
GU Z = 4.33
UG Z = 10.9
UU Z = 4.09
UC Z = 3.13
UA Z = 4.25
AU Z = n.s.
GC Z = n.s.
CG Z = 3.59
AA Z = 3.48
AC Z = n.s.
AG Z = 5.82
CA Z = 5.70
CC Z = 4.55
CU Z = 3.02
GA Z = 6.39
GG Z = 3.20
GU Z = n.s.
UG Z = 5.37
UU Z = 3.47
UC Z = 6.42
Thousands of piRNA read pairs
Extent of 5′-5′ complementarity, X (nt)

Figure 2.7. Mouse MILI and MIWI2 select a purine at the t1 position of their targets

(A) Nucleotide composition of g1 from guide piRNAs and t1 of inferred targets for Ping-Pong analysis between MILI and MIWI2 and between MIWI2 and MILI.

(B) MILI:MIWI2 and MIWI2:MILI Ping-Pong profiles for each *trans* pair, requiring at least 16 nt contiguous complementarity.

Figure 2.8

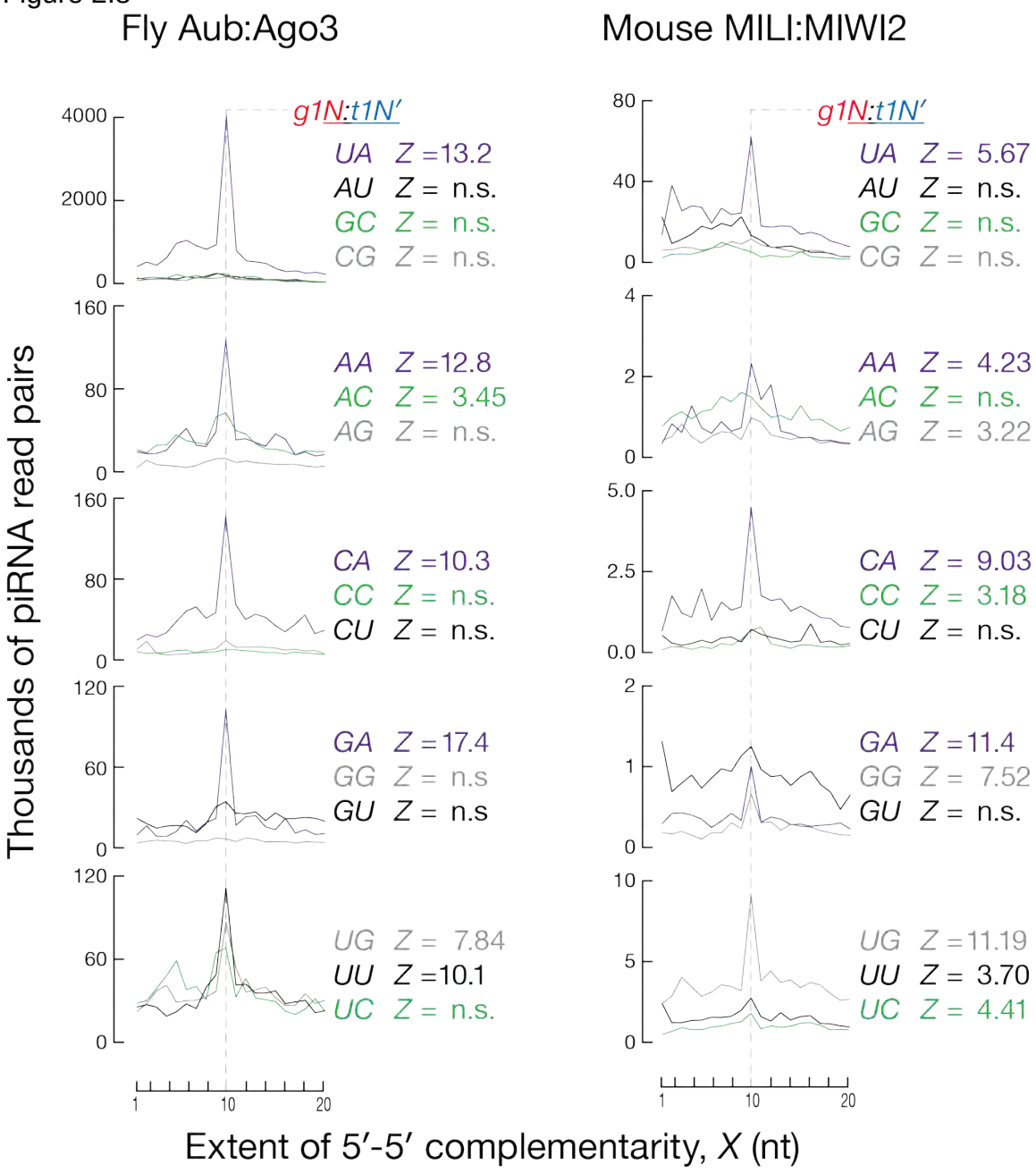

Fly Aub:Ago3
Mouse MILI:MIWI2
g1N:t1N′
UA Z =13.2
AU Z = n.s.
GC Z = n.s.
CG Z = n.s.
AA Z =12.8
AC Z = 3.45
AG Z = n.s.
CA Z =10.3
CC Z = n.s.
CU Z = n.s.
GA Z =17.4
GG Z = n.s
GU Z = n.s
UG Z = 7.84
UU Z =10.1
UC Z = n.s.
UA Z = 5.67
AU Z = n.s.
GC Z = n.s.
CG Z = n.s.
AA Z = 4.23
AC Z = n.s.
AG Z = 3.22
CA Z = 9.03
CC Z = 3.18
CU Z = n.s.
GA Z =11.4
GG Z = 7.52
GU Z = n.s.
UG Z =11.19
UU Z = 3.70
UC Z = 4.41
Thousands of piRNA read pairs
Extent of 5′-5′ complementarity, X (nt)

Figure 2.8. t1A preference persists when 20 nt complementarity (g2–g20) is required between guides and targets

Fly Aub:Ago3 and Mouse MILI:MIWI2 Ping-Pong profiles for each *trans* pair requiring at least 20 nt contiguous complementarity.

compared to background (Figure 2.5B). These data suggest that Ago3 is impartial to the identity of the t1 target nucleotide that faces the g1 piRNA.

Bombyx mori Siwi, like *Drosophila* Aub, Prefers to Cleave a t1A Target

Our analyses suggest that fly Aub and mouse MILI, but not fly Ago3, have an intrinsic preference for t1A targets, regardless of the identity of the facing g1 nucleotide. Extracts of immortalized *Bombyx mori* BmN4 cells currently provide the only system for analyzing the germline piRNA pathway in insects. We used these extracts to examine the t1 preference of *Bombyx mori* Siwi, the silkmoth ortholog of fly Aub. We measured the cleavage efficiency of PIWI proteins immunopurified from BmN4 cells expressing FLAG-tagged Siwi or FLAG-tagged BmAgo3, the ortholog of fly Ago3 (Figure 2.9A). Cross-contamination of one PIWI protein by the other was small: co-purifying Siwi in FLAG-BmAgo3 IP was ~4.8% of immunopurified Siwi in the FLAG-Siwi IP, and co-purifying BmAgo3 in FLAG-Siwi IP was ~1.4% of immunopurified BmAgo3 in FLAG-BmAgo3 IP (Figure 2.10). We used high-throughput sequencing data to identify abundant g1U and g1C piRNAs bound to FLAG-tagged Siwi or BmAgo3 (Izumi et al., 2013). For each piRNA, we constructed a set of fully complementary 5′ ^{32}P-radiolabeled target RNAs bearing A, U, G or C at the t1 position. When loaded with the g1U or the g1C piRNA, Siwi cleaved the t1A target significantly better than the t1U, t1G or t1C (g1U:t1A vs. g1U:t1U, *p*-value = 2×10^{-7} by Tukey's multiple comparison test, the same hereinafter; g1U:t1A vs. g1U:t1G, *p*-value = 1×10^{-7}; g1U:t1A vs. g1U:t1C, *p*-value = 2×10^{-7}; g1C:t1A vs. g1C:t1U, *p*-value = 2

× 10^{-5}; g1C:t1A vs. g1C:t1G, *p*-value = 1 × 10^{-3}; g1C:t1A vs. g1C:t1C, *p*-value = 1 × 10^{-4}; Figure 2.9B). In contrast, BmAgo3 showed little or no t1 nucleotide preference among the four targets (*p*-value > 0.03; Figure 2.9B). We conclude that, like *Drosophila* Aub, *Bombyx mori* Siwi selects targets bearing a t1A nucleotide, regardless of the base-pairing status between the g1 and t1 nucleotides. Conversely, BmAgo3, like fly Ago3, has little or no t1 nucleotide preference.

Figure 2.9

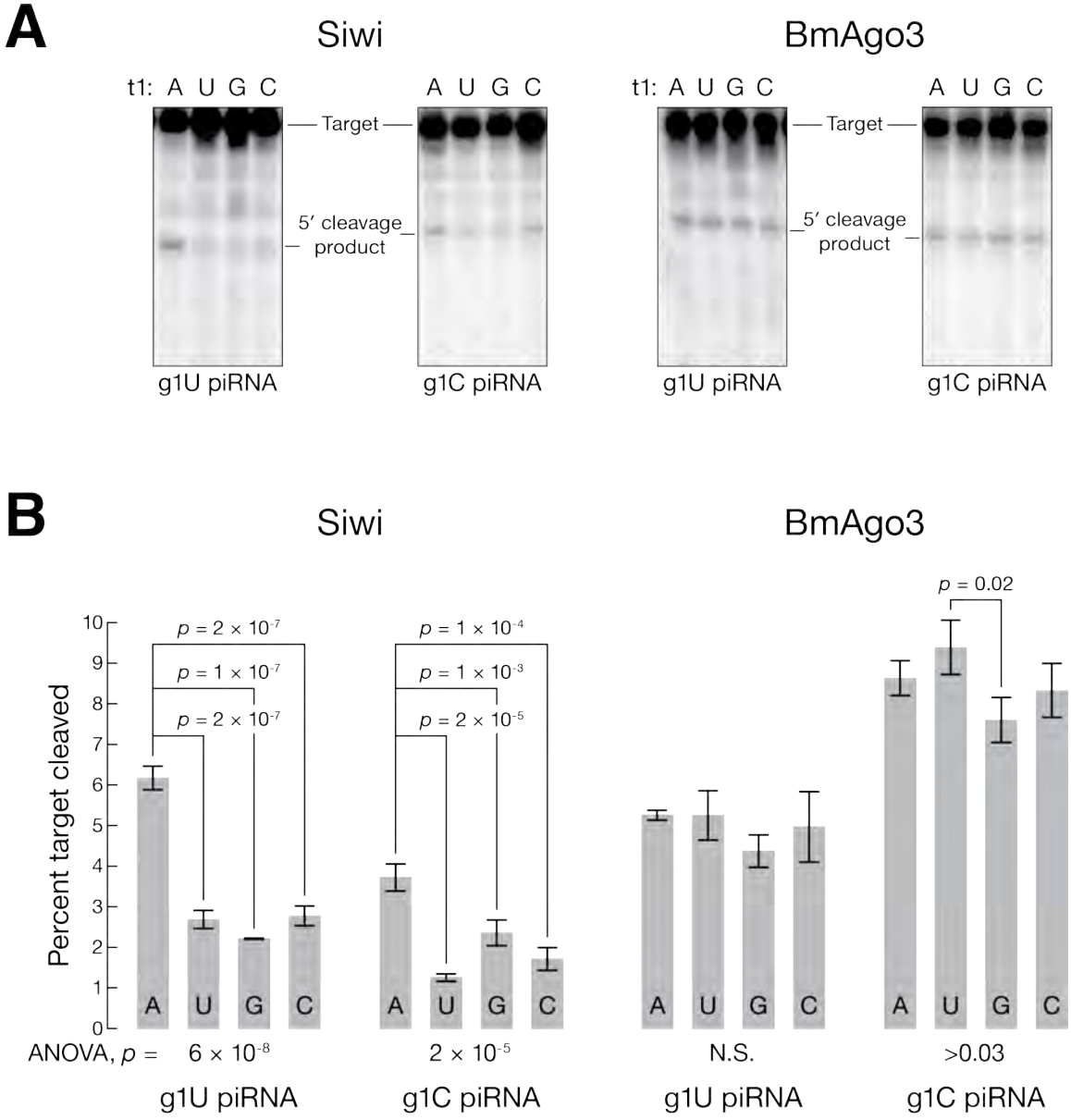

A
Siwi
BmAgo3
t1: A U G C
A U G C
Target
5′ cleavage product
g1U piRNA
g1C piRNA
B
Percent target cleaved
p = 2 × 10-7
p = 1 × 10-7
p = 2 × 10-7
p = 1 × 10-4
p = 1 × 10-3
p = 2 × 10-5
p = 0.02
ANOVA, p = 6 × 10-8
2 × 10-5
N.S.
>0.03

Figure 2.9. Siwi, the Silkmoth Ortholog of Aub, Prefers to Cleave t1A Targets

(A) FLAG-Siwi or FLAG-BmAgo3 were immunopurified from BmN4 lysates and cleavage of a set of target RNAs with t1A, U, G or C complementary to abundant g1U or g1C piRNAs was measured. Representative data is shown.

(B) Quantification of the experiments in (A). *p* values of one-way ANOVA and Tukey's multiple comparison test are shown (*n* = 3).

Figure 2.10


immunoprecipitate:
FLAG-Siwi
FLAG-BmAgo3
Percent loaded:
100 30 10 3 1 0.3
100 30 10 3 1 0.3
150 kDa
Siwi
100
4.8%
150
1.4%
BmAgo3
100

Figure 2.10. Estimation of reciprocal cross-contamination in Siwi and BmAgo3 immunoprecipitates

Immunopurified FLAG-Siwi and FLAG-BmAgo3 were analyzed by Western blotting using anti-Siwi or anti-BmAgo3 antibodies (Kawaoka et al., 2009).

Discussion

During the Ping-Pong amplification cycle, piRNAs bound to Aub or Siwi typically start with uridine (g1U), whereas Ago3- and BmAgo3-bound secondary piRNAs bear adenine at their g10 position, the nucleotide derived from the t1 position of the Aub- or Siwi-cleaved RNA (Brennecke et al., 2007; Gunawardane et al., 2007a). To date, models for secondary piRNA biogenesis have assumed that the g10A characteristic of Ago3-bound secondary piRNAs is generated by a requirement for the g1U of an Aub-bound piRNA to pair with the t1 position of its cleavage targets. Our analyses demonstrate that the t1A preference is maintained even when g1:t1 is unpaired. Taken together, our computational and biochemical experiments suggest a revision to the standard Ping-Pong model for piRNA amplification: the g10A hallmark of secondary piRNAs reflects a preference of the Aub protein for a t1A and is not the consequence of g1U:t1A base pairing. This revision to the secondary piRNA biogenesis pathway helps to unify the piRNA and miRNA target selection mechanisms: mammalian Argonaute proteins, and likely those in other animals and plants, display a strong preference for t1A that is determined by the protein and not g1:t1 base pairing. We note that for both pathways, an intrinsic protein preference for t1A, rather than g1:t1 base pairing, eliminates the paradox that the g1U base was proposed to base pair despite its appearing unavailable for base pairing in multiple three-dimensional structures.

What purpose does the preference of Aub, Siwi, and mouse MILI for an adenine at the t1 position serve in the amplification of piRNAs by the Ping-Pong cycle? One consequence of the t1A preference of, for example, Aub, is that Ago3-bound piRNAs will bear a g10A even when the Aub-bound piRNA does not begin with U. Subsequently, this Ago3-bound, g10A piRNA will cleave only targets that bear a t10U, because base pairing of the central region of a small RNA guide is required for target cleavage by Argonaute proteins generally. Because the t10 nucleotide becomes position g1 of the piRNA generated by Ago3-catalyzed target cleavage, the process ensures that secondary piRNAs loaded into Aub (or Siwi or MILI) begin with uracil. Biochemical experiments demonstrate that Siwi prefers g1U guides (Kawaoka et al., 2011), and the structure of the MID domain of mouse MIWI bound to nucleotide monophosphates explains how MIWI recognizes g1U (Cora et al., 2014). Current evidence is consistent with Aub also having an intrinsic preference for g1U. Thus, the t1A preference of Aub would serve to produce optimal piRNA substrates for loading into Aub and Piwi, the PIWI proteins that mediate transposon silencing in flies. In contrast, the absence of a t1 nucleotide preference for Ago3 expands the universe of sequences that can become piRNAs: if Ago3 also preferred a t1A, piRNA would be made largely from genomic regions whose sequences possessing an A exactly 10 nucleotides from a U. Understanding how Aub, Siwi, and MILI select for t1A remains a challenge for future structural studies of these proteins.

Experimental Procedures

Defining *Cis* and *Trans* Ping-Pong Pairs

To detect Aub:Ago3 Ping-Pong pairs, we identified Aub-bound guides immunoprecipitated with anti-Aub and inferred targets from the Ago3-bound piRNAs immunoprecipitated with anti-Ago3. We removed those piRNAs that present in both the Aub and Ago3 immunoprecipitates to avoid ambiguity in assignment and designated the remaining piRNAs as uniquely Aub-bound piRNAs and uniquely Ago3-bound piRNAs. This treatment was used for all pairs of Argonaute proteins throughout the study.

We assumed that g2–g10 pairing was required for target cleavage directed by piRNAs, and identified putative Aub:Ago3 Ping-Pong pairs by first identifying all pairs of Aub-bound piRNAs and Ago3-bound piRNAs with a 10-nt 5′-5′ overlap with perfect complementarity. To identify *cis*-pairs, we mapped all piRNAs to the reference fly genome without allowing any mismatches. A piRNA could map to multiple locations in the genome due to the repetitive nature of their transposon targets. For a putative pair, if any combinations of the genome mapping locations of the two piRNAs lead to a 10-nt 5′-5′ overlap, the pair was classified as a *cis*-pair. The remaining pairs need to satisfy the complementarity requirement detailed in the next section to qualify as *trans*-pairs.

Estimation of the Extent of Complementarity Required for *Trans* Ping-Pong

To estimate the extent of complementarity required for an authentic target, we analyzed all putative *trans* pairs for their pairing status at positions g11–g23. The t11–t23 sequence of a putative *trans*-target was inferred from the genomic sequence immediately upstream of the mapped piRNA (Figure 2.3B). The read counts of the guide and target were multiplied to yield the abundance of the pair. Then, for each position from g11–g23, we summed the abundance of pairs that exhibited complementarity. Normalization across the 13 positions resulted in the pairing frequency reported in Figure 2.4A, left panel.

To assess the significance of the pairing frequency of putative *trans*-pairs, we randomly shuffled the t11–t23 subsequences among putative *trans*-targets while maintaining their t2–t10 sequences and the pairing relationship with their respective guides. We generated 10 sets of shuffled sequences and determined their pairing frequencies shown as controls in Figure 2.4A, left panel.

In addition to assessing the pairing status of each position *X* from g11–g23, we also determined the pair abundance of contiguous complementarity from g2 until *X*, i.e., the guide and the target were not complementary at the *X* + 1 position. The pair abundances were normalized across g11–23 and shown as pairing frequency in Figure 2.4B, right panel, along with controls using shuffled data and normalized separately.

Ping-Pong Analyses of *Trans* pairs

Based on our results shown in Figure 2.4A, we designated *trans*-pairs that were perfectly complementary at positions g2–g16 as *bona fide trans* Ping-Pong pairs. We then split these *trans*-pairs into 16 categories according to the nucleotide combination at g1:t1 and tested the significance of the Ping-Pong signal in each category. As described above, we defined the abundance of each *trans*-pair as the product of guide abundance and target abundance. When multiple guides were paired with the same *trans*-target, we apportioned the reads of the target evenly among the pairs to compute pair abundance. On the other hand, when one guide was paired with multiple *trans*-targets, we did not apportion the reads of the guide. (Apportioning the guide but not the target would have reached the same result.) This way, we yield results compatible with our previous *cis* Ping-Pong analysis (Li et al., 2009).

Like the *cis* Ping-Pong analysis we developed earlier (Li et al., 2009), the significance of a *trans* Ping-Pong signature is measured as the *Z*-score of the pair abundance at the 10-nt 5′-5′ overlap, defined as the number of standard deviations away from the mean of the background distribution composed of overlap distances of 1–9 nt and 11–16 nt. The pair abundances for the non-10-nt overlapping distances were obtained using a similar procedure as the pair abundance for the 10-nt overlapping distance.

General Methods

w^{1} control female flies (1–2 day-old) were fed yeast for three days before their ovaries were dissected. BmN4 cell lines stably expressing FLAG-Siwi or FLAG-BmAgo3 (Kawaoka et al., 2009) and preparation of lysis buffer and BmN4 cell lysate by 1,000 × *g* centrifugation were as described (Haley et al., 2003; Kawaoka et al., 2011). Small RNA was isolated and prepared for 50 nt long high-throughput sequencing as described (Zhang et al., 2011). Figures were generated using Excel (Microsoft, Redmond, WA, USA), IgorPro (WaveMetrics, Lake Oswego, OR, USA), and Photoshop and Illustrator (Adobe systems, San Jose, CA, USA).

Transgenic Flies Expressing Catalytically Inactive Ago3

Flies expressed mRNA encoding a catalytically inactive Ago3 protein (Ago3^{ADH}) from two transgenes, one with an *aub* promoter and the with a *vasa* promoter (Sano et al., 2002), in an *ago3*^{t2}*/ago3*^{t3} (Li et al., 2009) mutant background.

Degradome Sequencing

Fresh total RNA from *w*^{1} female flies was depleted of rRNA by two rounds of Ribo-Zero depletion (Epicentre, WI, USA). Turbo DNase (Ambion, NY, USA) treatment was performed after the first round. Ligation of the 5′ adapter was performed at 25°C for 3 h using T4 RNA ligase (Ambion), followed by reverse transcription using a primer with a degenerate 3′ end. cDNA was purified using Agencourt RNAClean XP beads (Beckman Coulter). PCR was performed using

NEBNext mixture (New England Biolabs, MA, USA), followed by size selection of 200–400 bp DNA from a native 6% polyacrylamide gel. A second round of PCR was used to introduce barcodes. PCR products of 200–500 bp were size selected as before. Products (100 nt paired-end) were sequenced on a HiSeq 2000 (Illumina, CA, USA).

Analysis of Immunoprecipitated Small RNAs

Barcodes (6 nt) were sorted allowing no mismatches, and the 3′ adaptor was removed allowing one mismatch. Small RNAs were aligned to fly genome using Bowtie (Langmead et al., 2009) allowing no mismatches. Subsequent analyses, including removal of rRNAs, known noncoding RNAs, and miRNA hairpin-matching sequences, were as described (Li et al., 2009). All piRNA species shared by Ago3 and Aub were removed for Ping-Pong analysis.

Degradome Analysis and *Trans* Ping-Pong between Mature piRNAs and Degradome Reads

Degradome reads were aligned to *Drosophila melanogaster* genome version dm3 using STAR (Dobin et al., 2013b). Any mapper with 5′ end clipped during mapping process was removed from further analysis. BEDTools was used to annotate degradome reads (Quinlan and Hall, 2010). Transposon mapping degradome reads were retained for further analysis.

To maximally differentiate the t1 preference of Aub and Ago3 for in the degradome, all piRNAs with shared 16 nt prefixes between Aub guide piRNAs

and Ago3 guide piRNAs were removed before the Ping-Pong analysis between piRNA and degradome reads.

Datasets

We analyzed published small RNA deep sequencing data for mouse MILI and MIWI2 associated piRNAs (ERP000778; De Fazio et al., 2011).

Target RNA Preparation

Transcription templates containing the T7 promoter and the target site for corresponding piRNA were generated by PCR using the forward primer (5′-GCG TAA TAC GAC TCA CTA TAG TCA CAT CTC ATC TAC CTC C-3′) and one of the reverse primers in Table 2.1 with pGL3-Basic (Promega, Madison, WI, USA) as template. Target RNA was transcribed from the PCR products using T7-Scribe Standard RNA IVT kit (Cell Script, Madison, WI, USA). Gel-purified transcript was cap-radiolabeled with [α-^{32}P] GTP using the ScriptCap m^{7}G Capping System (Cell Script) and then gel-purified.

In Vitro Target Cleavage

For one reaction, 40 µl Dynabeads Protein G (Thermo Fisher Scientific, Waltham, MA, USA] were equilibrated with lysis buffer (Haley et al., 2003), incubated with 4 µl of 1 mg/ml anti-FLAG M2 antibody (Sigma-Aldrich, St. Louis, MO, USA) at 4°C for 30 min, washed twice with lysis buffer, and incubated at 4°C for 1 h with 120 µl lysate from BmN4 cells expressing FLAG-Siwi or FLAG-

BmAgo3. The beads were washed three times with lysis buffer, and then divided among four tubes; each was then incubated with 5′ ^{32}P-radiolabeled target RNA (1 nM). Target cleavage was analyzed by denaturing PAGE as described (Haley et al., 2003). Images were acquired using an FLA-7000 phosphorimager (Fujifilm, Minato-ku, Tokyo, Japan).

Table 2.1. Primers Used in PCR to Generate Transcription Templates for piRNA Targets

Primer	Sequence (5´-to-3´)	
Forward primer (common)	GCG TAA TAC GAC TCA CTA TAG TCA CAT CTC ATC TAC CTC C	
Reverse primer to make the targets for Siwi-bound and BmAgo3-bound g1U piRNA	t1A	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG GTT CTT CGG TAG TAT AGT GGT CAG TAT CGA AGA GAG GAG TTC ATG
	t1U	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG GTA CTT CGG TAG TAT AGT GGT CAG TAT CGA AGA GAG GAG TTC ATG
	t1G	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG GTC CTT CGG TAG TAT AGT GGT CAG TAT CGA AGA GAG GAG TTC ATG
	t1C	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG GTG CTT CGG TAG TAT AGT GGT CAG TAT CGA AGA GAG GAG TTC ATG
Reverse primer to make the targets for Siwi-bound g1C piRNA	t1A	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG TAA TCA CCA TAG AAT TAA CCC ACT GAG TGA AGA GAG GAG TTC ATG
	t1U	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG AAA TCA CCA TAG AAT TAA CCC ACT GAG TGA AGA GAG GAG TTC ATG
	t1G	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG CAA TCA CCA TAG AAT TAA CCC ACT GAG TGA AGA GAG GAG TTC ATG
	t1C	TTA GGT GAC ACT ATA GAT TTA CAT CGC GTT GAG TGT AGA ACG GTT GTA TAA AAG GAA TCA CCA TAG AAT TAA CCC ACT GAG TGA AGA GAG GAG TTC ATG
Reverse primer to make the targets for BmAgo3-bound g1C piRNA	t1A	TTA GTG ACA CTA TAG ATT TAC ATC GCG TTG AGT GTA GAA CGG TTG TAT AAA AGG TAT AGA AAG ATG CAC CAC GCC GGA ACC GAA GAG AGG AGT TCA TG
	t1U	TTA GTG ACA CTA TAG ATT TAC ATC GCG TTG AGT GTA GAA CGG TTG TAT AAA AGG AAT AGA AAG ATG CAC CAC GCC GGA ACC GAA GAG AGG AGT TCA TG
	t1G	TTA GTG ACA CTA TAG ATT TAC ATC GCG TTG AGT GTA GAA CGG TTG TAT AAA AGG CAT AGA AAG ATG CAC CAC GCC GGA ACC GAA GAG AGG AGT TCA TG
	t1C	TTA GTG ACA CTA TAG ATT TAC ATC GCG TTG AGT GTA GAA CGG TTG TAT AAA AGG GAT AGA AAG ATG CAC CAC GCC GGA ACC GAA GAG AGG AGT TCA TG

Acknowledgments

We thank members of the Zamore, the Weng and the Tomari laboratories for helpful discussions; Cindy Tipping for help with fly husbandry; Jia Xu for her pioneering efforts constructing the Ping-Pong algorithm; and Chengjian Li for advice on small RNA library construction. This work was supported in part by National Institutes of Health grants HG007000 to Z.W. and GM62862 and GM65236 to P.D.Z., and a Grant-in-Aid for Scientific Research on Innovative Areas (21115002) from the Japan Ministry of Education, Culture, Sports, Science and Technology to Y.T.

Chapter III Small Phasing Signals Reveal New Primary piRNA Biogenesis Mechanism

Preface

This chapter was a product of a collaborative effort: I discovered piRNA phasing in 2012. With the initial large-scale analysis, I pinpointed that phasing is related to primary piRNA pathway. Bo W Han found the +1U signature, which complemented the findings of phasing pattern. Bo W Han later performed fly genetics experiments and constructed the next generation sequencing libraries. I proposed the cleavage-site initiated phasing analysis on degradome reads. Bo W Han and I performed the computational analyses. Chengjian Li provided the sequencing data for the section: contribution of maternal piRNAs to phasing.

Summary

In animal gonads, PIWI-interacting RNAs (piRNAs) protect genome integrity by suppressing transposable elements. The current view is that primary piRNAs generated by the endonuclease Zucchini produce secondary piRNAs via the "Ping-Pong" pathway—reciprocal cycles of Aubergine- and Argonaute3-mediated cleavage of transposon mRNAs and piRNA precursor transcripts. Here, we show that secondary piRNAs also initiate the production of primary piRNAs, by feeding Aubergine- and Argonaute3-cleaved RNAs to Zucchini. The first ~26 nt of these cleaved RNAs become secondary piRNAs, while the next ~26 nt become the first in a series of phased primary piRNAs that bind Piwi and Aub, allowing piRNAs to spread beyond the initial site of RNA cleavage. While the Ping-Pong pathway only amplifies the abundance of inherited and *de novo* piRNAs, the production of phased primary piRNAs from adjacent sequences further introduces novel sequence diversity into the piRNA pool.

Introduction

Transposable elements occupy a substantial fraction of the typical eukaryotic genome; active transposons can cause mutations, genome rearrangements, and chromosomal breaks (Slotkin and Martienssen, 2007). Animals suppress transposon expression in the germline with an adaptive immune system composed of PIWI proteins and their single-stranded, 23–36 nucleotide (nt) small RNA guides, PIWI-interacting RNAs (piRNAs). In the *Drosophila* germline, piRNAs direct the PIWI proteins Piwi, Aubergine (Aub) and Argonaute3 (Ago3) to silence transposons (Siomi et al., 2011; Luteijn and Ketting, 2013). In flies, *maternal piRNAs* are deposited in the oocyte, ensuring their transmission to the germ cells of the next generation (Harris and Macdonald, 2001; Williams and Rubin, 2002; Megosh et al., 2006; Brennecke et al., 2008). *Primary piRNAs* derive from long, single-stranded precursor transcripts from piRNA clusters—discrete genomic loci comprising transposon fragments acquired over the last ~10 million years (Malone et al., 2009). Endonuclease Zucchini (Zuc) is thought to process cluster transcripts into fragments whose 5′ ends correspond to the 5′ ends of piRNAs, but whose length is considerably longer than piRNAs; these piRNAs precursors are proposed to be loaded into Piwi and Aub and then trimmed from their 3′ ends into mature primary piRNAs (Ipsaro et al., 2012; Nishimasu et al., 2012; Voigt et al., 2012; Luteijn and Ketting, 2013). In *zuc* mutant ovaries, piRNAs decrease, while piRNA cluster transcripts accumulate (Malone et al., 2009; Haase et al., 2010).

Figure 3.1

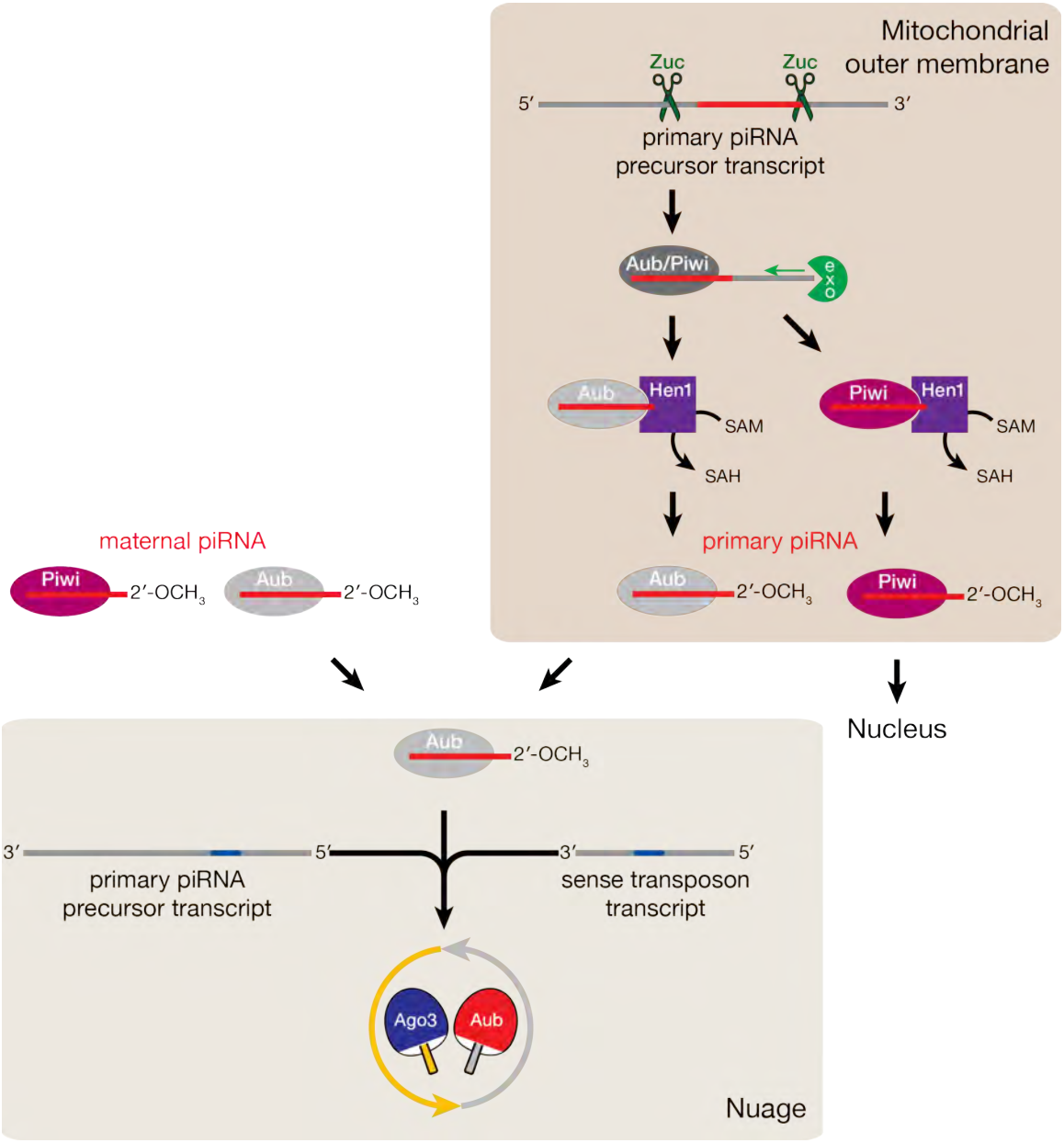

Mitochondrial outer membrane
Zuc
Zuc
5′
3′
primary piRNA precursor transcript
Aub/Piwi
exo
Aub
Hen1
SAM
SAH
Piwi
Hen1
SAM
SAH
maternal piRNA
Piwi
2′-OCH3
Aub
2′-OCH3
primary piRNA
Aub
2′-OCH3
Piwi
2′-OCH3
Nucleus
Aub
2′-OCH3
3′
5′
primary piRNA precursor transcript
3′
5′
sense transposon transcript
Ago3
Aub
Nuage

Figure 3.1. The current model for piRNA biogenesis.

Unlike other small silencing RNAs such as small interfering RNAs (siRNAs) or microRNAs, piRNAs participate in their own production, a process currently explained by the Ping-Pong model of piRNA biogenesis (Figure 3.1) (Brennecke et al., 2007; Gunawardane et al., 2007b). In this model, maternal and primary piRNAs are proposed to initiate the production of secondary piRNAs; secondary piRNAs subsequently self-amplify via reciprocal cycles of Aub- and Ago3-catalyzed cleavage of transposon mRNAs and cluster transcripts. Maternal and primary piRNAs complementary to transposons (i.e., in the antisense orientation) direct Aub to cleave transposon mRNAs. Biochemical and genetic data suggest that when Aub cleaves an RNA, e.g., a transposon mRNA, Ago3 can bind the resulting 5′ monophosphorylated cleavage product, which is then converted into a secondary piRNA (Brennecke et al., 2007; Gunawardane et al., 2007b; Kawaoka et al., 2011; Xiol et al., 2014). In turn, the Ago3-bound secondary piRNA can cleave a piRNA cluster transcript at a complementary sequence to generate a cleavage product that can bind Aub. Again, 3′ trimming generates a new secondary piRNA, essentially identical to the original Aub-bound piRNA. Additional rounds of such Ping-Pong amplification increase the abundance of the original, antisense, Aub-bound piRNA.

The Ping-Pong pathway increases piRNA abundance, but cannot create novel piRNA species. Yet piRNA populations are stunningly diverse, with most individual species of low abundance, and even genetically identical cells can have distinct piRNA profiles (Lau et al., 2009a). Here, we report that secondary piRNAs initiate the production of primary piRNAs by Zuc. Aub- or Ago3-catalyzed cleavage of long RNAs, such as transposon mRNAs, triggers the production of both primary and secondary piRNAs. The first ~26 nt of an Aub- or Ago3-cleaved

RNA becomes a secondary piRNA, while the next ~26 nt becomes the first in a series of phased primary piRNAs that bind Piwi and Aub, allowing piRNAs to spread beyond the initial site of secondary piRNA-directed cleavage. While the Ping-Pong pathway amplifies the abundance of inherited and *de novo* piRNAs, the production of phased primary piRNAs from adjacent sequences introduces novel sequence diversity into the piRNA pool.

Results

Phasing of primary piRNAs

Primary, maternal, and secondary piRNAs cannot be distinguished by their sequences. We used genetic mutants to separate the three types of piRNAs. The ovaries of *aub*$^{HN2/QC42}$*; ago3*$^{t2/t3}$ double-mutants lack the Ping-Pong pathway and contain only maternal and primary piRNAs. Accordingly, no significant Ping-Pong signal was detected in *aub*$^{HN2/QC42}$*; ago3*$^{t2/t3}$ ovaries (Z_{10} = 0.4; *Z*-score ≥ 2.81 corresponds to *p*-value ≤ 0.005, Figure 3.2A). Small RNAs from *zuc* mutant ovaries, which lack primary piRNAs, comprise maternal and secondary piRNAs. Total piRNAs, as well as piRNAs derived from the largest germline cluster, the *42AB* locus (Brennecke et al., 2007), declined in both *zuc*$^{HM27/Df}$ and *aub*$^{HN2/QC42}$*; ago3*$^{t2/t3}$, suggesting that both the primary and secondary pathways are needed to make germline piRNAs (Figure 3.3A). Although piRNAs from the *42AB* cluster decreased ~43-fold in *zuc*$^{HM27/Df}$ compared to genetically matched *w*^{1} ovaries, the piRNAs from *zuc*$^{HM27/Df}$ displayed a significant Ping-Pong signal (Z_{10} = 42), consistent with maternal piRNAs being amplified into secondary piRNAs. Remarkably, we observed that the maternal and primary piRNAs remaining in *aub*$^{HN2/QC42}$*; ago3*$^{t2/t3}$ double-mutants exhibited characteristics of phasing: the distance between the 5′ ends of piRNAs on the same genomic strand peaked at 25–28 nt—the lengths of piRNAs (Figure 3.3B). In contrast, this peak was absent in *zuc*$^{HM27/Df}$. The production of phased piRNAs suggests that a piRNA-generating nuclease initiates the production of piRNAs from a single end of a piRNA precursor, moving processively 5′-to-3′ to clip off successive piRNAs, much as Dicer makes siRNAs processively from the end of a double-stranded

RNA (Allen et al., 2005; Yoshikawa et al., 2005; Vagin et al., 2006; Cenik et al., 2011; Welker et al., 2011). In plants, phased siRNAs are produced by the nuclease *DCL4*, which initiates from a unique RNA end established by microRNA-directed cleavage (Allen et al., 2005; Gasciolli et al., 2005; Xie et al., 2005).

Unlike siRNAs, which are predominantly 21-nt long, piRNAs have a broad length distribution, and their 3′-ends are often trimmed during their maturation. Consequently, phased piRNAs are more difficult to detect than phased siRNAs. We devised two strategies to detect piRNA phasing. First, we measured the distance from the 3′ end of each piRNA to the 5′ end of the next downstream piRNA (Figure 3.2B). The most common distance was 1 nt (Z_1 score for w^1 = 6.5), implying that the probability of a single cleavage event producing both the 3′-end of an upstream piRNA and the 5′-end of the adjacent, downstream piRNA is significantly higher than expected by chance. Production of phased piRNAs required Zuc but not Ping-Pong amplification: the 1 nt peak was more prominent in $aub^{HN2/QC42}$; $ago3^{t2/t3}$ mutant ovaries (Z_1 = 22) than in wild-type (w^1), but undetectable in $zuc^{HM27/Df}$ (Z_1 = 1.5). Second, we evaluated the sequence composition of the nucleotides that immediately follow the 3′ ends of piRNAs (the “+1 percentage”). This strategy has the additional advantage of being unaffected by the number of piRNA species sequenced. In both wild-type (w^1) and $aub^{HN2/QC42}$; $ago3^{t2/t3}$, the +1 nucleotide was enriched for uridine; the uridine enrichment was lost in $zuc^{HM27/Df}$ (Figure 3.2C). Since the majority of piRNAs begin with uridine, these data suggest that piRNAs are made tail-to-head

Figure 3.2

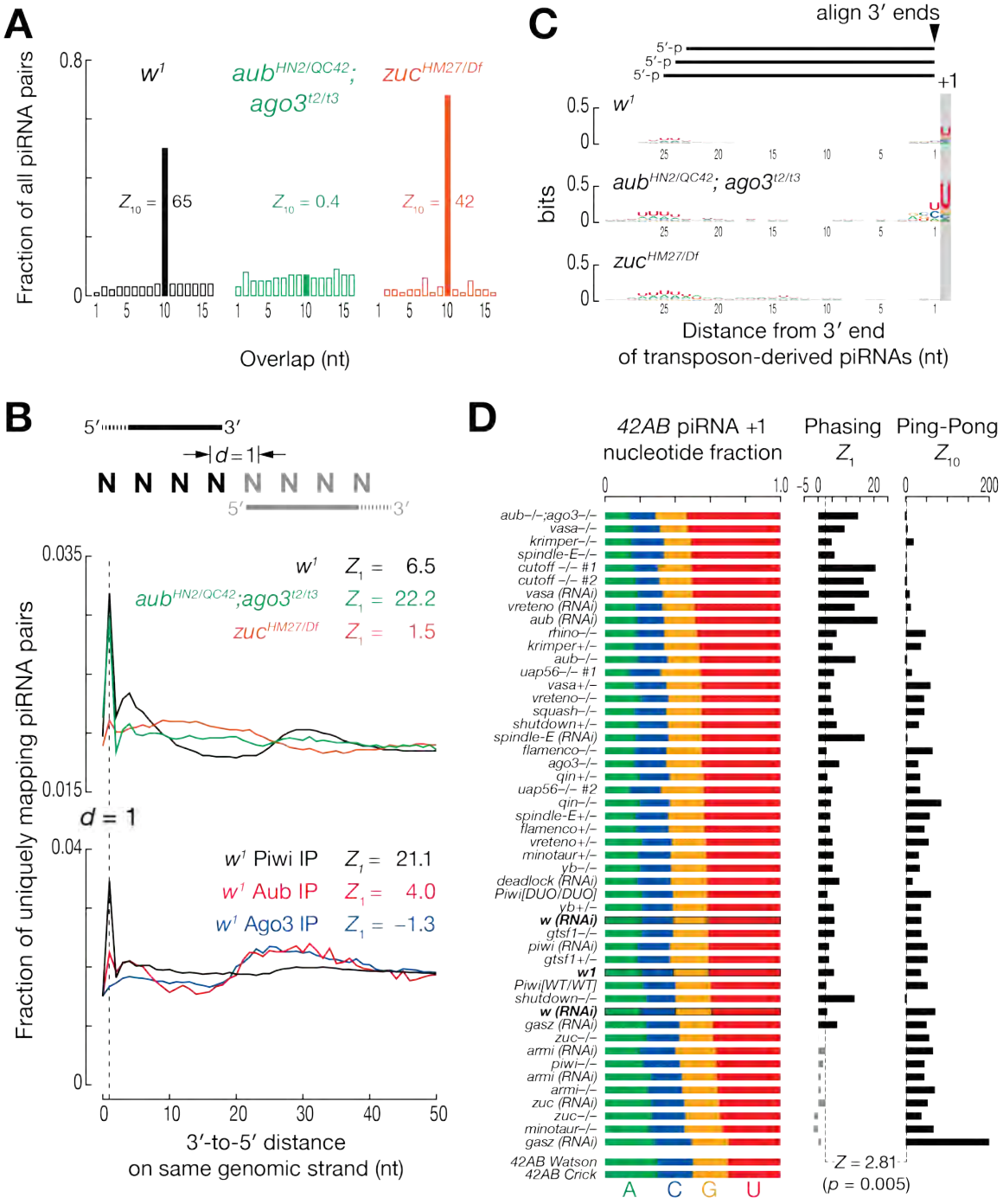

A
Fraction of all piRNA pairs
w1
aubHN2/QC42; ago3t2/t3
zucHM27/Df
Z10 = 65
Z10 = 0.4
Z10 = 42
Overlap (nt)
C
align 3′ ends
5′-p
+1
w1
aubHN2/QC42; ago3t2/t3
zucHM27/Df
bits
Distance from 3′ end
of transposon-derived piRNAs (nt)
B
5′
3′
d = 1
N N N N N N N N
Fraction of uniquely mapping piRNA pairs
w1 Z1 = 6.5
aubHN2/QC42;ago3t2/t3 Z1 = 22.2
zucHM27/Df Z1 = 1.5
d = 1
w1 Piwi IP Z1 = 21.1
w1 Aub IP Z1 = 4.0
w1 Ago3 IP Z1 = −1.3
3′-to-5′ distance
on same genomic strand (nt)
D
42AB piRNA +1
nucleotide fraction
Phasing
Z1
Ping-Pong
Z10
aub−/−;ago3−/−
vasa−/−
krimper−/−
spindle-E−/−
cutoff −/− #1
cutoff −/− #2
vasa (RNAi)
vreteno (RNAi)
aub (RNAi)
rhino−/−
krimper+/−
aub−/−
uap56−/− #1
vasa+/−
vreteno−/−
squash−/−
shutdown+/−
spindle-E (RNAi)
flamenco−/−
ago3−/−
qin+/−
uap56−/− #2
qin−/−
spindle-E+/−
flamenco+/−
vreteno+/−
minotaur+/−
yb−/−
deadlock (RNAi)
Piwi[DUO/DUO]
yb+/−
w (RNAi)
gtsf1−/−
piwi (RNAi)
gtsf1+/−
w1
Piwi[WT/WT]
shutdown−/−
w (RNAi)
gasz (RNAi)
zuc−/−
armi (RNAi)
piwi−/−
armi (RNAi)
armi−/−
zuc (RNAi)
zuc−/−
minotaur−/−
gasz (RNAi)
42AB Watson
42AB Crick
A C G U
Z = 2.81
(p = 0.005)

Figure 3.2. Zucchini-dependent phasing of primary piRNAs.

(A) Ping-Pong analysis of all piRNAs from *w*^{1}, *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ and *zuc*$^{HM27/Df}$.

(B) Distance from the 3′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs on the same genomic strand.

(C) Nucleotide composition of piRNA species (i.e., distinct sequences irrespective of abundance) 29 nt upstream and 1 nt downstream of the 3′ ends of piRNAs.

(D) Nucleotide composition of piRNA species immediately downstream of the 3′ ends of piRNAs that are uniquely mapped and derived from *42AB* cluster. *Z*-scores for Ping-Pong and phasing are shown. RNAi, germline RNA interference with double-stranded RNA or short hairpin RNA.

Figure 3.3

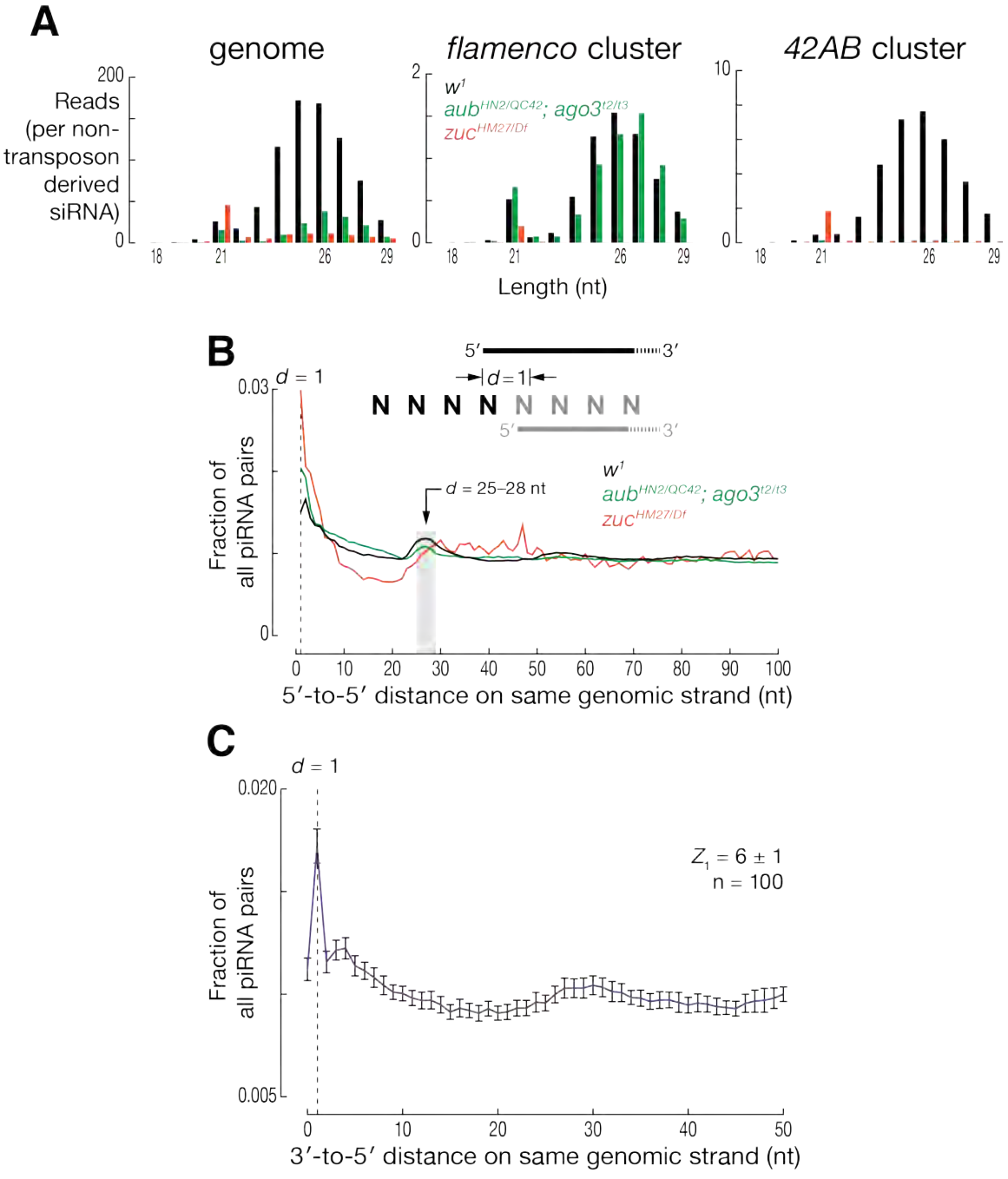

A
genome
flamenco cluster
42AB cluster
Reads (per non-transposon derived siRNA)
w1
aubHN2/QC42; ago3t2/t3
zucHM27/Df
Length (nt)
B
5′
3′
d = 1
N N N N N N N N
d = 25–28 nt
Fraction of all piRNA pairs
5′-to-5′ distance on same genomic strand (nt)
C
d = 1
Z1 = 6 ± 1
n = 100
Fraction of all piRNA pairs
3′-to-5′ distance on same genomic strand (nt)

Figure 3.3. Primary piRNAs display Zucchini-dependent phasing.

(A) Length distribution of genome-, *flamenco*- and *42AB* cluster-derived, uniquely mapping piRNAs from w^{1}, $aub^{HN2/QC42}$; $ago3^{t2/t3}$, and $zuc^{HM27/Df}$ ovaries. Reads were normalized to non-transposon-derived siRNAs, including *cis*-natural antisense transcripts and structured loci.

(B) Distance from 5′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs for uniquely mapping piRNAs from w1, aubHN2/QC42; ago3t2/t3, and zucHM27/Df ovaries. The data are reported as fraction of all piRNA pairs.

(C) Uniquely mapping, 42AB cluster-derived piRNAs from w1 were randomly down-sampled 100× to the number of 42AB cluster-derived, uniquely mapping piRNA species in zucHM27/Df. Then, distance from 3′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs were calculated for each sample. Error bars report standard deviation.

beginning at a unique site on each molecule of precursor RNA. The data further suggest that many piRNAs are not 3′-trimmed after they are clipped off their precursors.

Phased piRNAs are more readily detected when piRNAs are abundant, ensuring good genomic coverage. In theory, piRNAs might be phased in *zuc* mutants, but concealed by the low level of piRNA abundance in this mutant. To exclude this possibility, we randomly down-sampled the *42AB*-derived, uniquely mapping, piRNA species from w^{1} to the level of *42AB*-derived piRNAs in $zuc^{HM27/Df}$. The reduced set of wild-type piRNAs gave a Z_1 score (6 ± 1) very close to that obtained when using all wild-type piRNAs (Figure 3.3C).

Phasing of piRNAs differed among the three *Drosophila* PIWI proteins (Figure 3.2B). The 3′-to-5′ distance between piRNAs revealed that Piwi-bound piRNAs displayed the most significant phasing (Z_1 = 21); Aub-bound piRNAs displayed reduced, but still significant phasing relative to background (Z_1 = 4.0). In contrast, Ago3-bound piRNAs were not phased (Z_1 = −1.3). We conclude that Piwi- and Aub-bound primary piRNAs are produced by a processive mechanism that requires Zuc. Aub-piRNAs may be less phased, because they are a mixture of primary and more abundant secondary piRNAs. Ago3 binds only secondary piRNAs (Olivieri et al., 2012) and thus its piRNAs are not phased.

Genetic requirements for piRNA phasing

Analysis of the phasing of piRNAs derived from the *42AB* cluster in 21 different piRNA pathway mutants or germline RNA interference (RNAi) strains revealed significant piRNA phasing in all mutants except those with defects in the primary piRNA pathway, including *piwi, zucchini*, *armitage* (*armi*), *minotaur* and *gasz*

(Figure 3.2D) (Vagin et al., 2006; Pane et al., 2007; Malone et al., 2009; Haase et al., 2010; Olivieri et al., 2010; Czech et al., 2013; Handler et al., 2013; Vagin et al., 2013). The data comprised 49 small RNA sequencing libraries constructed in six different laboratories with five different 3′ adaptors, including one with a degenerate tetramer on its 5′ end. Therefore phasing is unlikely to be caused by technical artifacts such as library preparation. Mutants defective in piRNA Ping-Pong, including *vasa, krimper*, *spindle-E* and *aub*, all displayed more pronounced phasing, likely because the loss of secondary piRNAs reduces the background signal. The presence or absence of piRNA phasing in a mutant accurately predicted the previously defined role of the gene in primary versus secondary piRNA production. We also detected Zuc-dependent phasing in the piRNA cluster *flamenco* and the piRNA-producing 3′ UTR of the protein-coding *traffic jam* mRNA, two loci that produce piRNAs only in the somatic follicle cells that support oocyte development (Figure 3.4). Cultured, somatic ovarian sheet cells (OSCs), which possess only the primary piRNA biogenesis pathway, also display piRNA phasing (Figure 3.4). Neither somatic follicle cells nor cultured OSC cells express Aub or Ago3, and both lack a secondary piRNA pathway (Brennecke et al., 2007; Lau et al., 2009b). Thus, we conclude that phasing is an inherent feature of primary piRNA production.

Contribution of maternal piRNAs to phasing

To distinguish between maternal and primary piRNAs, we used a fly strain bearing a ~7 kb *P{GSV6}* transgene inserted into the *42AB* cluster. The transgene provided a set of unique sequences in an otherwise highly repetitive

Figure 3.4

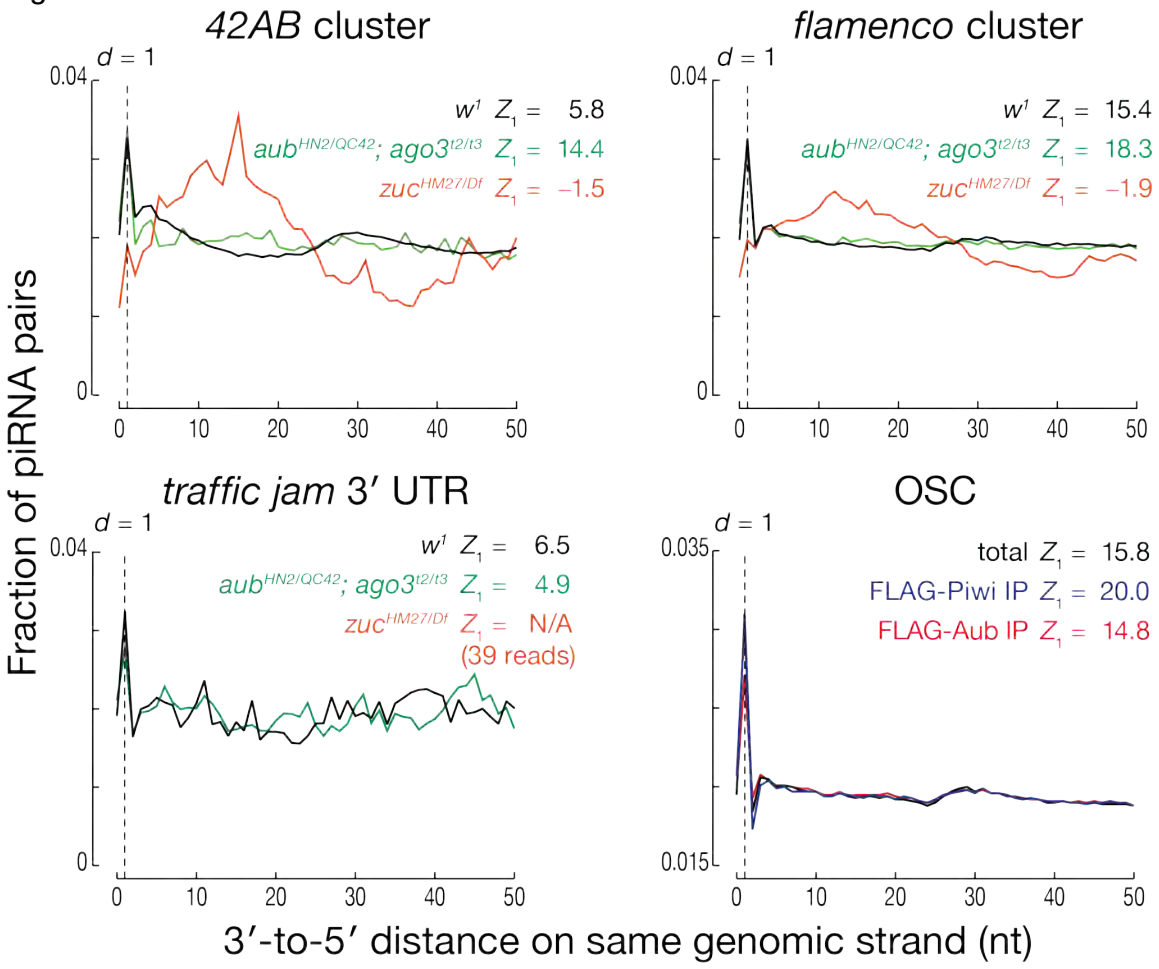

42AB cluster
flamenco cluster
traffic jam 3′ UTR
OSC
d = 1
w¹ Z1 = 5.8
aubHN2/QC42; ago3t2/t3 Z1 = 14.4
zucHM27/Df Z1 = –1.5
w¹ Z1 = 15.4
aubHN2/QC42; ago3t2/t3 Z1 = 18.3
zucHM27/Df Z1 = –1.9
w¹ Z1 = 6.5
aubHN2/QC42; ago3t2/t3 Z1 = 4.9
zucHM27/Df Z1 = N/A
(39 reads)
total Z1 = 15.8
FLAG-Piwi IP Z1 = 20.0
FLAG-Aub IP Z1 = 14.8
Fraction of piRNA pairs
3′-to-5′ distance on same genomic strand (nt)

Figure 3.4. Primary piRNAs from clusters, 3′ UTR and somatic cells display phasing

Distance from 3′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs for uniquely mapping piRNAs derived from 42AB, flamenco, the 3′ untranslated region (UTR) of traffic jam from w1, *aub*$^{HN2/QC42}$*; ago3*$^{t2/t3}$, and *zuc*$^{HM27/Df}$ ovaries. Few traffic jam-mapping piRNAs were detected for *zuc*$^{HM27/Df}$ and were not analyzed. Bottom-right: distance from 3′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs for all uniquely mapping piRNAs from cultured ovarian somatic cells (OSC), as well as those piRNAs co-purified with FLAG-HA-Piwi and FLAG-HA-Aub expressed in these cells .

locus. *P{GSV6}* carries both *gfp* and w^{+mC} and produces both sense and antisense *gfp* and w^{+mC} piRNAs (Figs. S3A and S3B). For piRNAs derived from this transgene, both piRNA Ping-Pong and overall piRNA abundance was greater when *P{GSV6}42A18* was inherited maternally rather than paternally (Figures 3.3A and 3.6A), as predicted by previous reports that maternal piRNAs enhance secondary piRNA production via the Ping-Pong pathway (Brennecke et al., 2008; Khurana et al., 2011; de Vanssay et al., 2012; Grentzinger et al., 2012; Le Thomas et al., 2014a; Le Thomas et al., 2014b). In contrast, primary piRNA phasing was essentially unaltered by the parental source of the transgene (Z_1 maternal = 13; Z_1 paternal = 13).

To further test the idea that primary piRNAs are phased, we sequenced piRNAs from $vasa^{D5/PH165}$ ovaries that had inherited the *P{GSV6}* transgene maternally or paternally (Figure 3.5A). *vasa* mutants lack the Ping-Pong pathway, so secondary piRNAs are unlikely to be present in these ovaries (Malone et al., 2009; Zhang et al., 2011; Xiol et al., 2014). w^{1}; $vasa^{D5}$ / *P{GSV6}42A18*, $vasa^{PH165}$ females obtain the transgene from their fathers and are therefore unlikely to inherit piRNAs. Conversely, w^{1}; *P{GSV6}42A18*, $vasa^{PH165}$ / $vasa^{D5}$ flies inherit the transgene from their mothers, who deposit maternal piRNAs in the developing oocyte. Nonetheless, the *P{GSV6}*-derived piRNAs in both genotypes displayed significant phasing (paternal transgene, Z_1 = 12; maternal transgene, Z_1 = 9.0; wild-type, Z_1 = 13), consistent with the idea that phasing is a primary piRNA signature that requires neither maternal piRNAs nor Ping-Pong amplification.

Loss of Vasa had no effect on either the significance of piRNA phasing (i.e., Z_1-score) or the percentage of uridine at the nucleotide immediately after

Figure 3.5

A

Maternal genotype	Paternal genotype	F1 genotype	Maternal piRNA	Primary piRNA	Secondary piRNA	Phasing Z_3	+1 U percentage	Ping-Pong Z_{10}
+/+	+/*P{GSV6}42A18*	+/*P{GSV6}42A18*	No	Yes	Yes	13	41%	3.5
P{GSV6}42A18/+	+/+	*P{GSV6}42A18*/+	Yes	Yes	Yes	13	47%	6.2
vasa/+	+/*P{GSV6}42A18, vasa*	*vasa*/*P{GSV6}42A18, vasa*	No	Yes	No	12	48%	0.8
P{GSV6}42A18, vasa/+	+/*vasa*	*P{GSV6}42A18, vasa*/*vasa*	Yes	Yes	No	9.0	48%	0.4

B

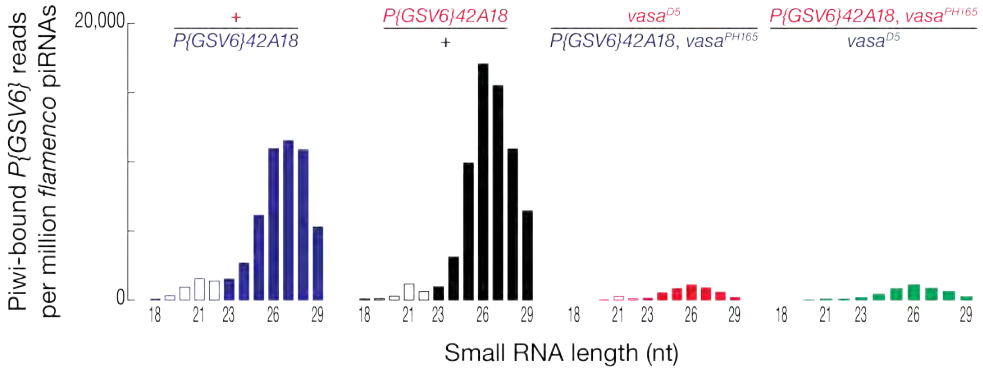
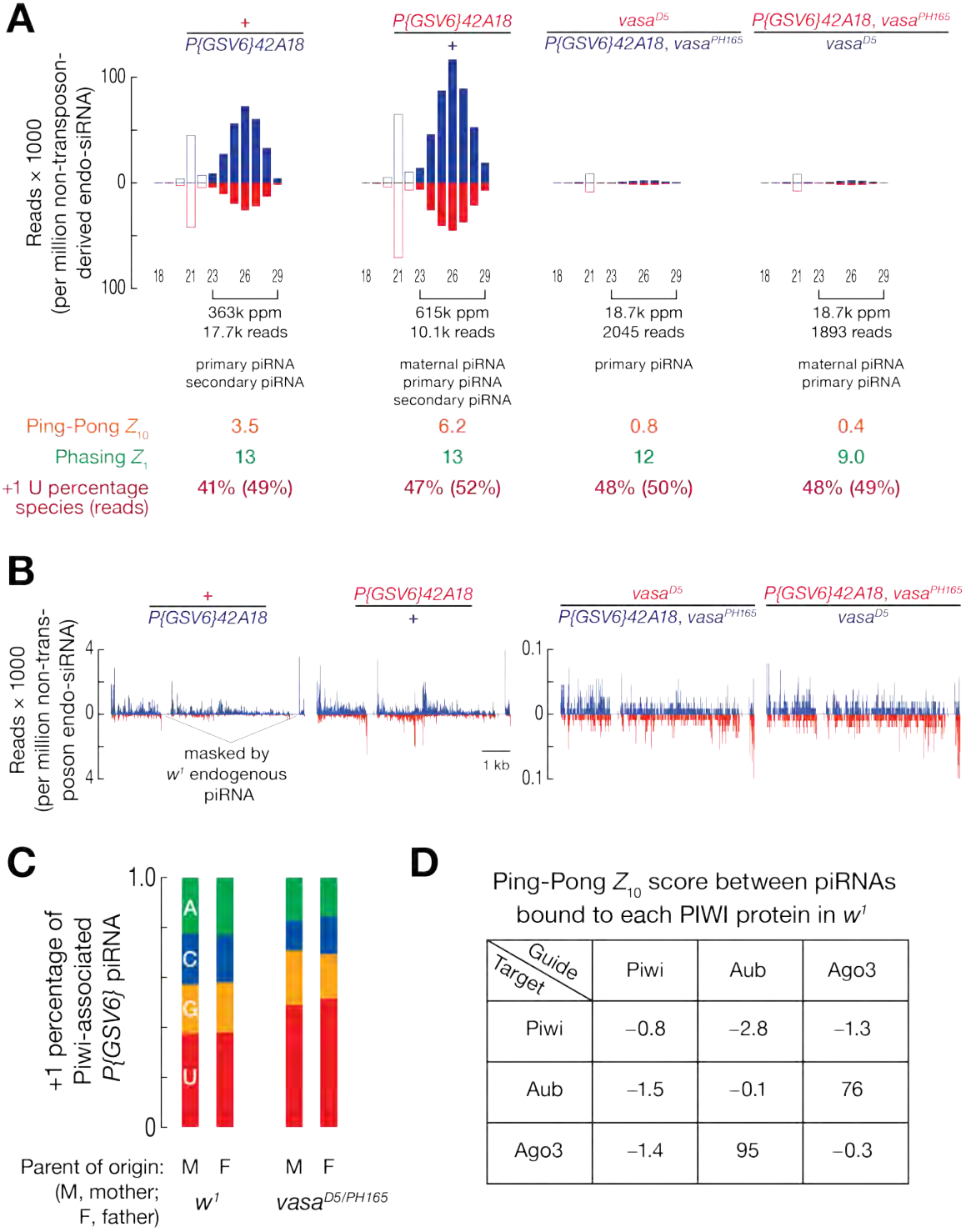


Figure 3.5. Contribution of maternal and secondary piRNAs to phasing.

(A) *Z*-scores for Ping-Pong and phasing and +1 U percentage for *P{GSV6}42A18*-derived piRNAs with the transgene inherited paternally or maternally, with or without *vasa*.

(B) Length distribution of Piwi-bound, uniquely mapping piRNAs derived from *P{GSV6}42A18* in wild-type and *vasa* mutants with the transgene inherited either maternally or paternally. Reads were normalized to *flamenco*-derived, uniquely mapping piRNAs in the same library.

Figure 3.6



Target \ Guide	Piwi	Aub	Ago3
Piwi	−0.8	−2.8	−1.3
Aub	−1.5	−0.1	76
Ago3	−1.4	95	−0.3

Figure 3.6. Phasing is a feature of primary but not maternal piRNAs.

(A) Length distribution, Ping-Pong analysis, phasing *Z*-score, and +1 U percentage are shown for piRNAs (23–29 nt) from *P{GSV6}42A18* in different genotypes. Red: maternally inherited allele; blue: paternally inherited allele. Reads were normalized to non-transposon-derived siRNAs from *cis*-NATs and structural loci.

(B) piRNA reads from the *P{GSV6}42A18* transgene in wild-type (w^1) or *vasa* mutant ovaries shown according to whether the transgene was inherited maternally or paternally.

(C) The +1 nucleotide percentage of Piwi-associated, uniquely mapping piRNA species from the *P{GSV6}42A18* transgene.

(D) Ping-Pong *Z*-scores between piRNAs associated with each PIWI protein in w^1 ovaries. Complementarity at positions 2–16 nt on the guide piRNAs were required.

the 3′ ends of the piRNAs (+1 U percentage), yet, without Vasa, the abundance of Piwi-bound piRNAs was less than one-tenth of wild type (w^1; Figures 3.5B and 3.6C). Current evidence suggests that Piwi is loaded only with primary piRNAs (Zhang et al., 2011; Sienski et al., 2012; Darricarrere et al., 2013; Le Thomas et al., 2014b) and our data support this notion (Figure 3.6D). Why then, should Vasa, a central component of the secondary piRNA pathway, affect the abundance of Piwi-bound primary piRNAs?

Phasing from 5′ monophosphorylated RNAs

One explanation for a role for Vasa in primary piRNA production is that Zuc converts secondary piRNA-directed cleavage products into phased primary piRNAs, which are subsequently loaded in Piwi. Because *vasa* mutants lack the secondary piRNA-producing Ping-Pong pathway, secondary piRNA-directed cleavage products—the substrates for primary piRNA production—are not made. To test the idea that Aub and Ago3 initiate Zuc-dependent primary piRNA production, we sequenced RNAs longer than 200 nt and bearing 5′ monophosphates (“degradome reads”) (Addo-Quaye et al., 2008; Karginov et al., 2010). Argonaute proteins produce this terminal structure when they cleave nucleic acids.

Degradome sequencing of wild-type (w^1) ovaries detected the RNA cleavage products of Aub and Ago3, but not Piwi: analysis of the distance between the 5′ ends of piRNAs bound to Aub, Ago3, or Piwi and the 5′ ends of degradome reads on the opposite genomic strand readily identified long RNAs whose 5′ ends were generated by Aub- or Ago3-catalyzed, piRNA-directed

Figure 3.7

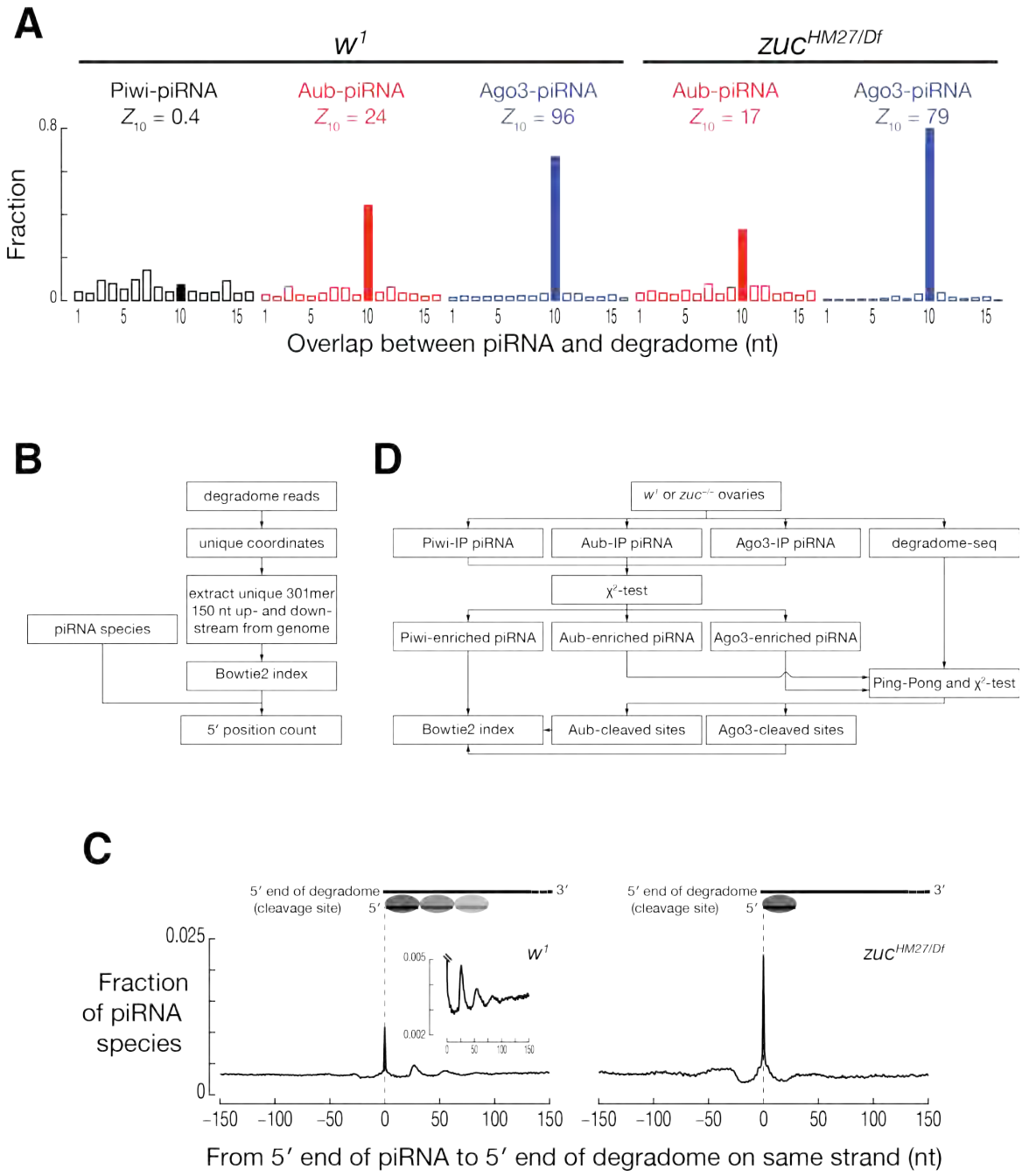

A
w1
zucHM27/Df
Piwi-piRNA
Z10 = 0.4
Aub-piRNA
Z10 = 24
Ago3-piRNA
Z10 = 96
Aub-piRNA
Z10 = 17
Ago3-piRNA
Z10 = 79
Fraction
0.8
0
1 5 10 15
Overlap between piRNA and degradome (nt)
B
degradome reads
unique coordinates
extract unique 301mer 150 nt up- and down-stream from genome
piRNA species
Bowtie2 index
5′ position count
D
w1 or zuc−/− ovaries
Piwi-IP piRNA
Aub-IP piRNA
Ago3-IP piRNA
degradome-seq
χ2-test
Piwi-enriched piRNA
Aub-enriched piRNA
Ago3-enriched piRNA
Ping-Pong and χ2-test
Bowtie2 index
Aub-cleaved sites
Ago3-cleaved sites
C
5′ end of degradome (cleavage site)
5′
3′
w1
zucHM27/Df
Fraction of piRNA species
0.025
0
0.005
0.002
0 50 100 150
−150 −100 −50 0 50 100 150
From 5′ end of piRNA to 5′ end of degradome on same strand (nt)

Figure 3.7. Piwi-associated piRNAs display phasing 3′ to the cleavage sites of Aub and Ago3.

(A) Ping-Pong analysis between PIWI protein-associated piRNAs and degradome reads.

(B) Computational strategy to measure the distance from the 5′ ends of piRNAs to the 5′ ends of degradome reads.

(C) Distance from 5′ ends of transposon-derived piRNAs to the 5′ ends of degradome reads in w^{1} (left) and $zuc^{HM27/Df}$ (right).

(D) Computational strategy to identify sites cleaved by Aub or Ago3 in degradome-seq data. These sites were then used to calculate the distance to the 5′ ends of nearby PIWI-associated piRNAs.

cleavage (Figure 3.7A) (Wang et al., 2014); in contrast, degradome reads consistent with Piwi-catalyzed cleavage were indistinguishable from background (Z_{10} = 0.4). As expected from the persistence of secondary piRNA production in the absence of Zuc (Figure 3.2A), long RNAs whose 5′ ends were generated by Aub and, especially, by Ago3 were readily detected in degradome sequences from *zuc*$^{HM27/Df}$ ovaries (Z_{10} = 17 for Aub; Z_{10} = 79 for Ago3; Figure 3.7A). Such degradome reads were not present in *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ (Z_{10} = 0.8) mutants, suggesting that our degradome sequencing approach accurately identifies the sites of Aub and Ago3 cleavage on their RNA substrates.

To test whether the 3′ cleavage products of Aub- or Ago3-catalyzed slicing are subsequently used to produce phased primary piRNAs, we analyzed the positions of piRNA 5′ ends within the sequences (on the *same* genomic strand) of the cleavage products from transposon transcripts. To accomplish this, we determined the fraction of piRNA 5′ ends at each position 150 nt upstream and 150 nt downstream from the cleavage sites (Figure 3.7B). In both wild-type (*w*^{1}) and *zuc* mutant ovaries, the 5′ ends of piRNAs were far more likely to map to the cleavage site than expected by chance (*w*^{1}, Z_0 = 27; *zuc*$^{HM27/Df}$, Z_0 = 34; Figure 3.7C). The Ping-Pong model predicts this result: it posits that the 5′ termini of Aub- and Ago3-cleaved RNAs subsequently become the 5′ ends of secondary piRNAs. However, two additional peaks of piRNA 5′ ends were present ~26 nt and ~53 nt downstream of the cleavage sites. That is, the 5′ end of a piRNA lies immediately after the 3′ end of the secondary piRNA (i.e., ~26 nt from the cleavage site), and the 5′ end of another piRNA follows the 3′ end of that piRNA (i.e., ~53 nt from the cleavage site). The ~26 and ~53 nt peaks were readily detected in wild-type, but not in *zuc*$^{HM27/Df}$ ovaries. The requirement for Zuc

Figure 3.8

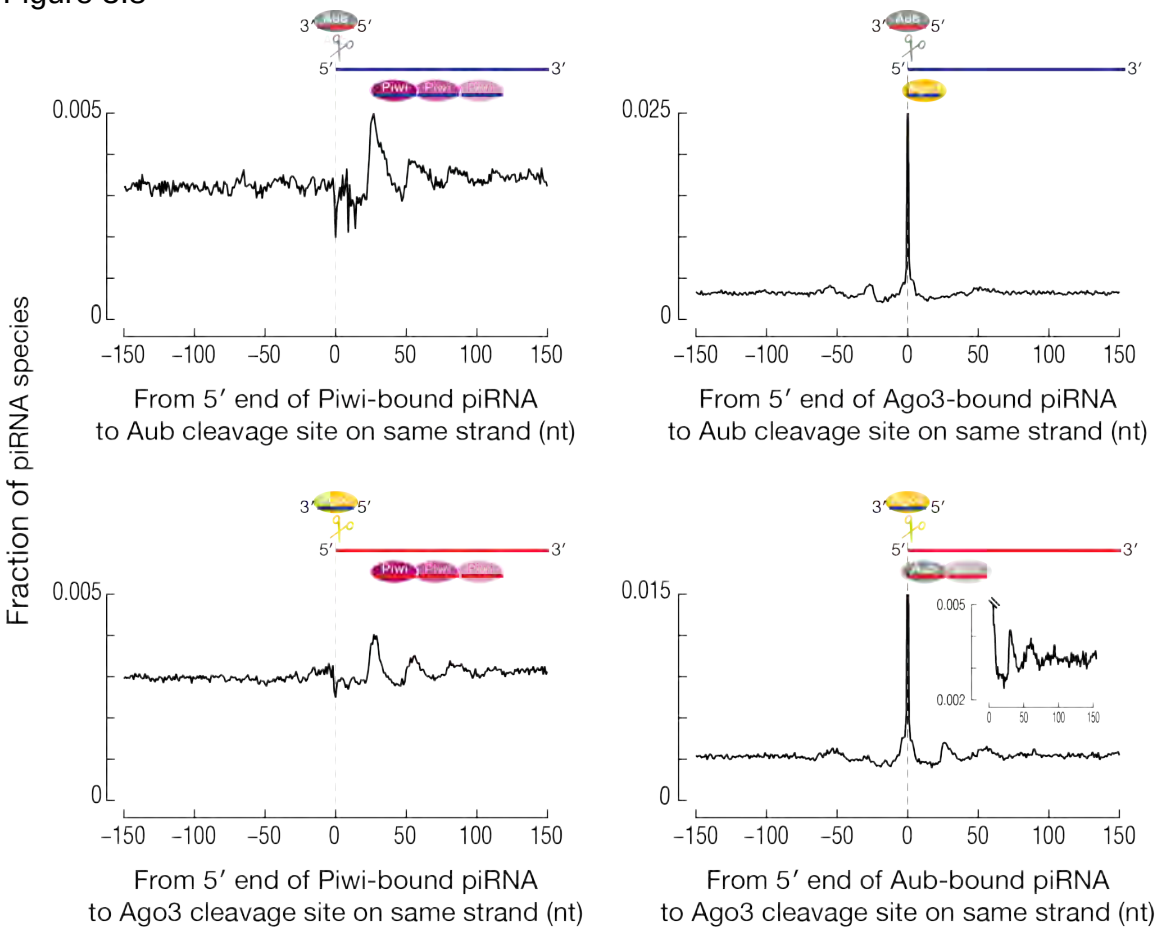

Fraction of piRNA species
From 5′ end of Piwi-bound piRNA to Aub cleavage site on same strand (nt)
From 5′ end of Ago3-bound piRNA to Aub cleavage site on same strand (nt)
From 5′ end of Piwi-bound piRNA to Ago3 cleavage site on same strand (nt)
From 5′ end of Aub-bound piRNA to Ago3 cleavage site on same strand (nt)

Figure 3.8. Phasing of Piwi-piRNAs downstream of the cleavage sites of Aub and Ago3 in *w*^{1}.

The distance between the 5′ ends of Piwi- (left), Ago3- (top-right), and Aub- (bottom-right) bound piRNAs and the cleavage sites of Aub (top) and Ago3 (bottom) on the same genomic strand in *w*^{1}.

suggests that the production of a single secondary piRNA from the 5′ end of an RNA cleaved by Aub or Ago3 is followed by the processing of the downstream sequence into phased primary piRNAs.

Phased piRNAs from Aub- and Ago3-cleaved RNAs

To specifically test whether phased primary piRNAs follow a secondary piRNA, we examined those degradome reads likely to have been produced by Aub- and Ago3-catalyzed RNA cleavage. First, we applied a statistical strategy to separate degradome reads based on the likelihood (*p*-value ≤ 0.005, χ^2 test) that they were produced by piRNAs bound to Aub or Ago3 (Figure 3.7D) (Wang et al., 2014). Then, we analyzed the distance between the 5′ ends of Piwi-bound piRNAs and the sites of Aub- or Ago3-catalyzed RNA cleavage. As expected, the 3′ cleavage products of Aub shared their 5′ ends with those of Ago3-bound piRNAs, and the cleavage products of Ago3 shared their 5′ ends with those of Aub-bound piRNAs. Unexpectedly, the 5′ ends of Piwi-bound piRNAs coincided with the Zuc-dependent peaks ~26 and ~53 nt downstream (Figure 3.8). Unlike Ago3-bound piRNAs, a smaller but still significant fraction of Aub-bound piRNAs also began ~26 and ~53 nt after the Ago3-cleaved sites.

To further distinguish RNAs cleaved by Aub- or Ago3 from those cleaved by Zuc (or a Zuc-dependent process), we used small RNA and degradome sequencing data from *zuc* mutant ovaries to identify the sites cleaved by Aub or Ago3. Next, we measured the distance between the sites of Aub- or Ago3-catalyzed cleavage and the Piwi-bound piRNAs in wild-type (w^1) ovaries. This analysis showed that the 5′ ends of Piwi-bound piRNAs typically lie ~26 and ~53 nt downstream from sites at which Aub or Ago3 has cleaved an RNA (Figure

Figure 3.9

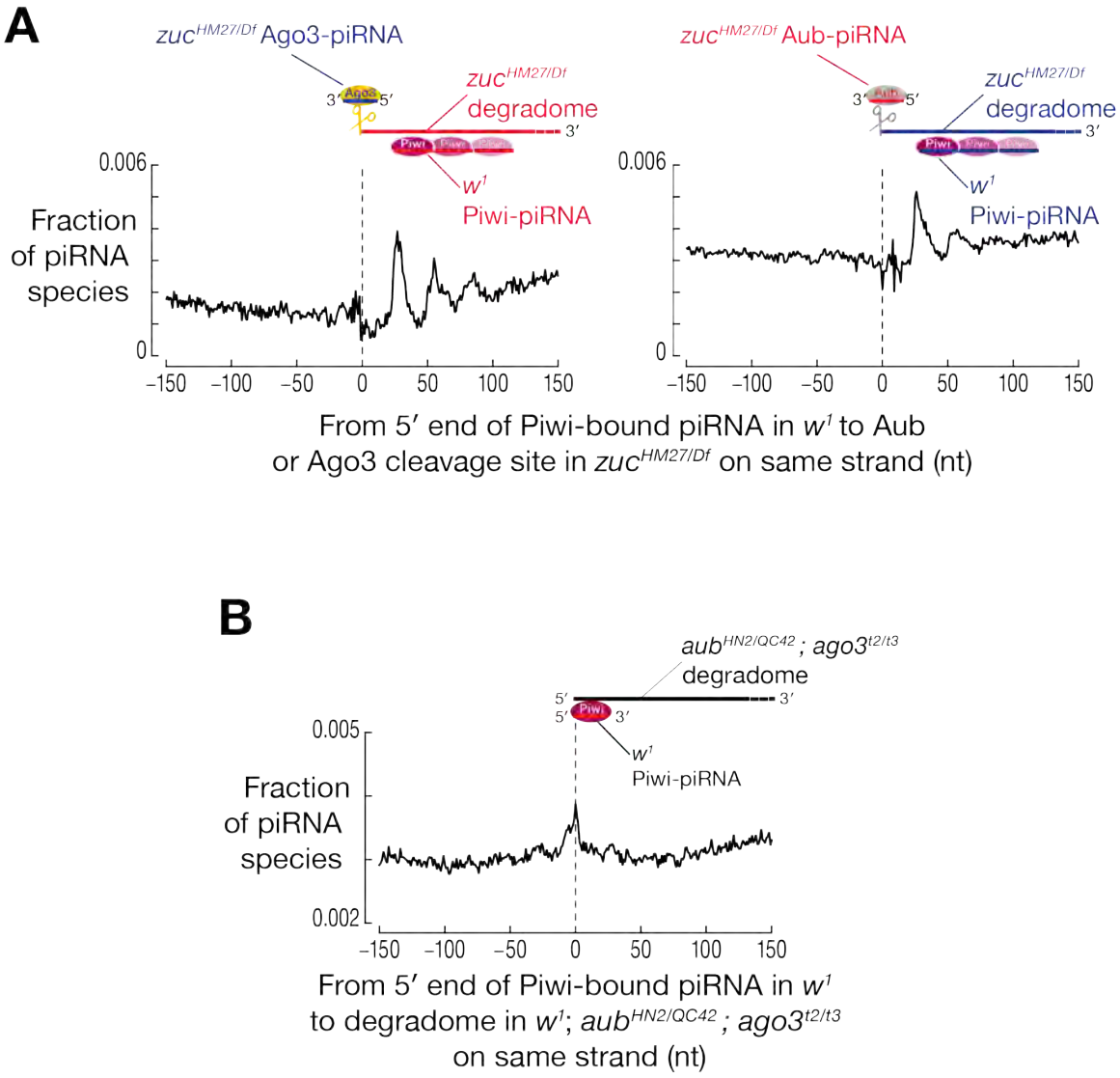

A
zuc^HM27/Df Ago3-piRNA
zuc^HM27/Df degradome
w^1 Piwi-piRNA
zuc^HM27/Df Aub-piRNA
zuc^HM27/Df degradome
w^1 Piwi-piRNA
Fraction of piRNA species
0.006
0
−150 −100 −50 0 50 100 150
From 5′ end of Piwi-bound piRNA in w^1 to Aub or Ago3 cleavage site in zuc^HM27/Df on same strand (nt)
B
aub^HN2/QC42 ; ago3^t2/t3 degradome
w^1 Piwi-piRNA
Fraction of piRNA species
0.005
0.002
−150 −100 −50 0 50 100 150
From 5′ end of Piwi-bound piRNA in w^1 to degradome in w^1; aub^HN2/QC42 ; ago3^t2/t3 on same strand (nt)

Figure 3.9. Slicing activity of Aub and Ago3 is required for phased Piwi piRNA production.

(A) Distance from the 5′ ends of Piwi-associated piRNAs in w^{1} to the cleavage sites of Ago3 (left) and Aub (right) identified in $zuc^{HM27/Df}$.

(B) Distance from 5′ ends of Piwi-associated piRNAs in w^{1} to the 5′ ends of degradome reads in w^{1}; $aub^{HN2/QC42}$; $ago3^{t2/t3}$.

3.9A). This relationship between cleavage sites and the 5′ ends of Piwi-bound piRNAs from wild-type ovaries was not detected when we repeated the analysis using degradome sequence data from *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ mutants (Figure 3.9B).

To further test the idea that cleavage by Aub and Ago3 both generates a secondary piRNA and initiates the phased production of Piwi-bound primary piRNAs, we measured the distance from the 5′ ends of Aub- and Ago3-bound piRNAs to the 5′ ends of Piwi-bound piRNAs on the same genomic strand (Figure 3.10A). As noted before, this strategy for detecting phasing is more direct than analyzing Z_1 score and +1 U percentage, although its sensitivity is hampered by the varying lengths of piRNAs. Consistent with our analysis of degradome data, the 5′ ends of Piwi-bound piRNAs were typically 26 nt downstream from the 5′ ends of Ago3-piRNAs and 27–29 nt downstream of the 5′ ends of Aub-piRNAs.

Similarly, the 5′ ends of Aub-bound piRNAs lay ~26 nt downstream from the 5′ ends of Ago3-bound piRNAs. In contrast, the 5′ ends of Ago3-bound piRNAs were no more likely to be ~26 nt downstream from the 5′ ends of Aub-bound piRNAs than would be expected by chance. Thus, RNAs cut by Ago3 produce phased, Aub-bound piRNAs, but RNAs cut by Aub do not make phased, Ago3-bound piRNAs.

We also measured the distance between the 5′ ends of Piwi-bound piRNAs and the 5′ ends of Aub- or Ago3-bound piRNAs on the opposite genomic strand (i.e., Ping-Pong analysis; Figure 3.10B). Piwi does not directly participate in Ping-Pong, and accordingly the 5′ ends of Piwi-bound piRNAs did not significantly map 10 nt from the 5′ ends of Aub- or Ago3-bound piRNAs. Instead, Piwi-bound piRNAs were found 15–19 nt downstream from the 5′ end of the Aub-

or Ago3-bound piRNAs. These distances suggest that Piwi-bound piRNAs are produced immediately downstream the 3′ end of the secondary piRNA derived from the 5′ terminal sequences of Aub- or Ago3-cleaved RNA (Figure 3.10B). That is, the 3′ cleavage product generated by Aub- or Ago3-catalyzed RNA cleavage is initially processed to generate a secondary piRNA, and thereafter is used for the Zuc-dependent production of phased primary piRNAs loaded into Piwi.

Contributions of Aub and Ago3 to phased primary piRNAs

To evaluate the contribution of Aub and Ago3 to phased primary piRNA production in the germline, we sequenced Piwi-bound piRNAs from *aub*, *ago3*, and *vasa* mutant ovaries. In the absence of Ago3 or Vasa, *42AB*-derived, Piwi-bound piRNAs decreased to ~10% of the wild-type level (Figure 3.10C). Loss of Aub had a more modest effect: *42AB*-derived, Piwi-bound piRNAs decreased to ~47% of the wild-type level. To remove potential somatic piRNAs, we excluded any reads with genomic coordinates that overlapped with the coordinates of small RNAs from OSC cells. With this restriction, Piwi-bound, uniquely mapping piRNAs were 76% of wild type in *aub* mutant ovaries, but just ~25% of wild type in *ago3* and *vasa* mutants. These data suggest that Ago3 makes a greater contribution than Aub to initiating the production of phased, Piwi-bound primary piRNAs. Supporting this view, less Piwi is present in the nurse cell nuclei of *ago3* than *aub* mutants (Li et al., 2009).

The impact of 3′ trimming on detection of phased piRNAs

We envision that cleavage by Aub or Ago3 generates the 5′ end of a secondary piRNA, while Zuc cleavage makes its 3′ end. The mean distance between the 5′

end of an Ago3-bound secondary piRNA and the 5′ end of the subsequent Piwi-bound primary piRNA was 26.1 nt; for an Aub-bound piRNA the mean distance to the first downstream Piwi-bound primary piRNA was 26.2 nt (Figure 3.10A). Yet the mean length of wild-type (w^1) Ago3-bound piRNAs was 23.6 nt; Aub-bound piRNAs were 24.9 nt. If Piwi- and Aub-bound primary piRNAs immediately follow the 3′ ends of Ago3-bound secondary piRNAs, what explains the discrepancy between the mean 5′-to-5′ distances and the shorter mean piRNA lengths? We propose that Zuc first cleaves ~26 nt downstream from the 5′ end of a secondary piRNA, generating the 5′ end of a Piwi-bound primary piRNA. This cleavage event also liberates an Aub- or Ago3-bound secondary piRNA with an immature 3′ end. Subsequently, 3′ trimming and 2′-*O*-methylation establishes the mature secondary piRNA length.

In *Bombyx mori*, Siwi- and Ago3-bound piRNAs are 3′ trimmed by a process that requires the Tudor protein Papi: when Papi is depleted by RNAi, mean piRNA length increases by 0.4 nt (Honda et al., 2013). Similarly, loss of TDRD2, the mouse ortholog of Papi, causes piRNAs in pre-pachytene spermatocytes to become 1–10 nt longer (Saxe et al., 2013).

Figure 3.10

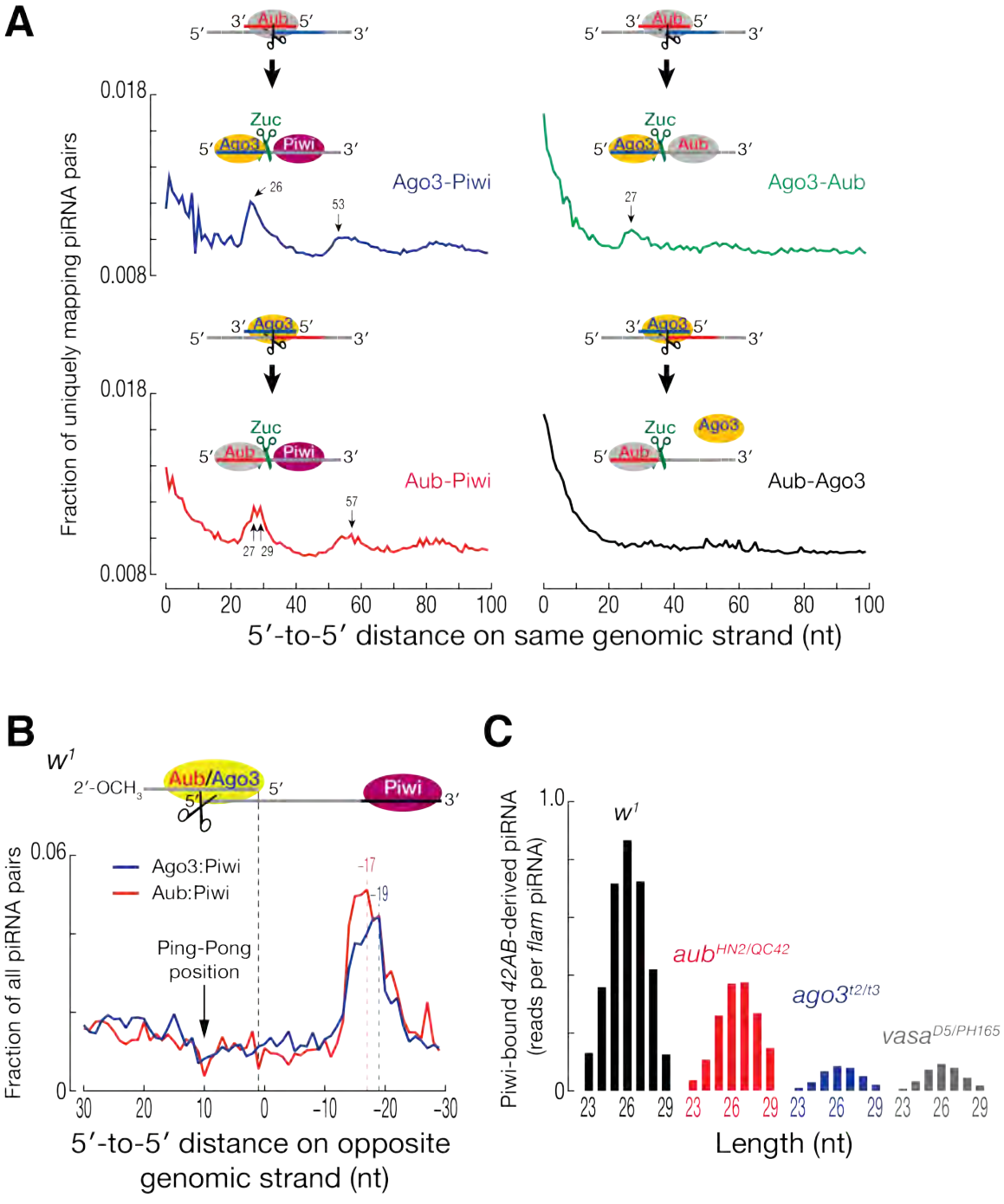

A
Aub
Zuc
Ago3
Piwi
Ago3-Piwi
Ago3-Aub
Aub-Piwi
Aub-Ago3
Fraction of uniquely mapping piRNA pairs
0.018
0.008
26
53
27
57
27 29
0 20 40 60 80 100
5′-to-5′ distance on same genomic strand (nt)
B
w1
2′-OCH3
Aub/Ago3
Piwi
Ago3:Piwi
Aub:Piwi
Ping-Pong position
-17
-19
Fraction of all piRNA pairs
0.06
0
30 20 10 0 -10 -20 -30
5′-to-5′ distance on opposite genomic strand (nt)
C
w1
aubHN2/QC42
ago3t2/t3
vasaD5/PH165
Piwi-bound 42AB-derived piRNA (reads per flam piRNA)
1.0
0
23 26 29 23 26 29 23 26 29 23 26 29
Length (nt)

Figure 3.10. Piwi-bound piRNAs produced 3′ to Aub- and Ago3-bound piRNAs.

(A) Distance from 5′ ends of upstream piRNAs to 5′ ends of downstream piRNAs on the same genomic strand for piRNAs bound to each PIWI protein in w^{1} ovaries.

(B) Distance from 5′ ends of Aub- or Ago3-bound piRNAs to 5′ ends of Piwi-bound piRNAs on the opposite genomic strand in w^{1} ovaries.

(C) The length distribution of Piwi-bound, uniquely mapping piRNAs derived from *42AB* cluster in wild-type, $aub^{HN2/QC42}$, $ago3^{t2/t3}$ and $vasa^{D5/PH165}$. Reads were normalized to *flamenco*-derived, uniquely mapping piRNAs in the same library.

To test whether 3′ trimming plays a role in the biogenesis of phased primary piRNAs, we sequenced small RNAs from *papi*$^{Df(2L)D125}$/*papi*$^{Df(2L)Exel7010}$ trans-heterozygous mutant fly ovaries (Figure 3.11A). Indeed, the median length of piRNAs from nearly all transposon families increased by 0.35 nt (*p*-value < 2.2 × 10^{-16}, Wilcoxon signed-rank test; Figures 3.11B and 3.11C). piRNA abundance was unchanged (Figure 3.11D), in agreement with the finding that *Het-A* and *blood* transposons remain silenced when Papi is depleted by RNAi (Handler et al., 2011). Piwi-bound primary piRNAs may also undergo some 3′ trimming, as the phasing of both somatic and germline piRNAs became more detectable in the absence of Papi: for *flamenco*-derived piRNAs the +1 U percentage increased from 46% to 51%, while the Z_1 score for 3′-to-5′ distance increased by 33% (Figures 3.12A and 3.12B). We propose that 3′ trimming of Piwi-bound piRNAs accommodates the use as processing sites of uridines >26 nt downstream from the 5′ end of the pre-piRNA bound to Piwi.

Phasing of mammalian piRNAs

To investigate whether the phased primary piRNA pathway is conserved in mammals, we performed 3′-to-5′ and 5′-to-5′ distance analyses for piRNAs from mouse testis (Figures 3.13A–C) (Li et al., 2013). We found no evidence for the 5′ end of a piRNA immediately after the 3′ end of an upstream piRNA (for example, 3′-to-5′ distance Z_1 = −0.5 at 26.5 day post partum, dpp). However, 5′-to-5′ distance analysis showed a broad peak at 30–40 nt (summit, 35 nt) for piRNAs derived from transposons, mRNAs, and long non-coding RNAs (pachytene piRNAs). One potential explanation for the non-significant Z_1 score is

Figure 3.11

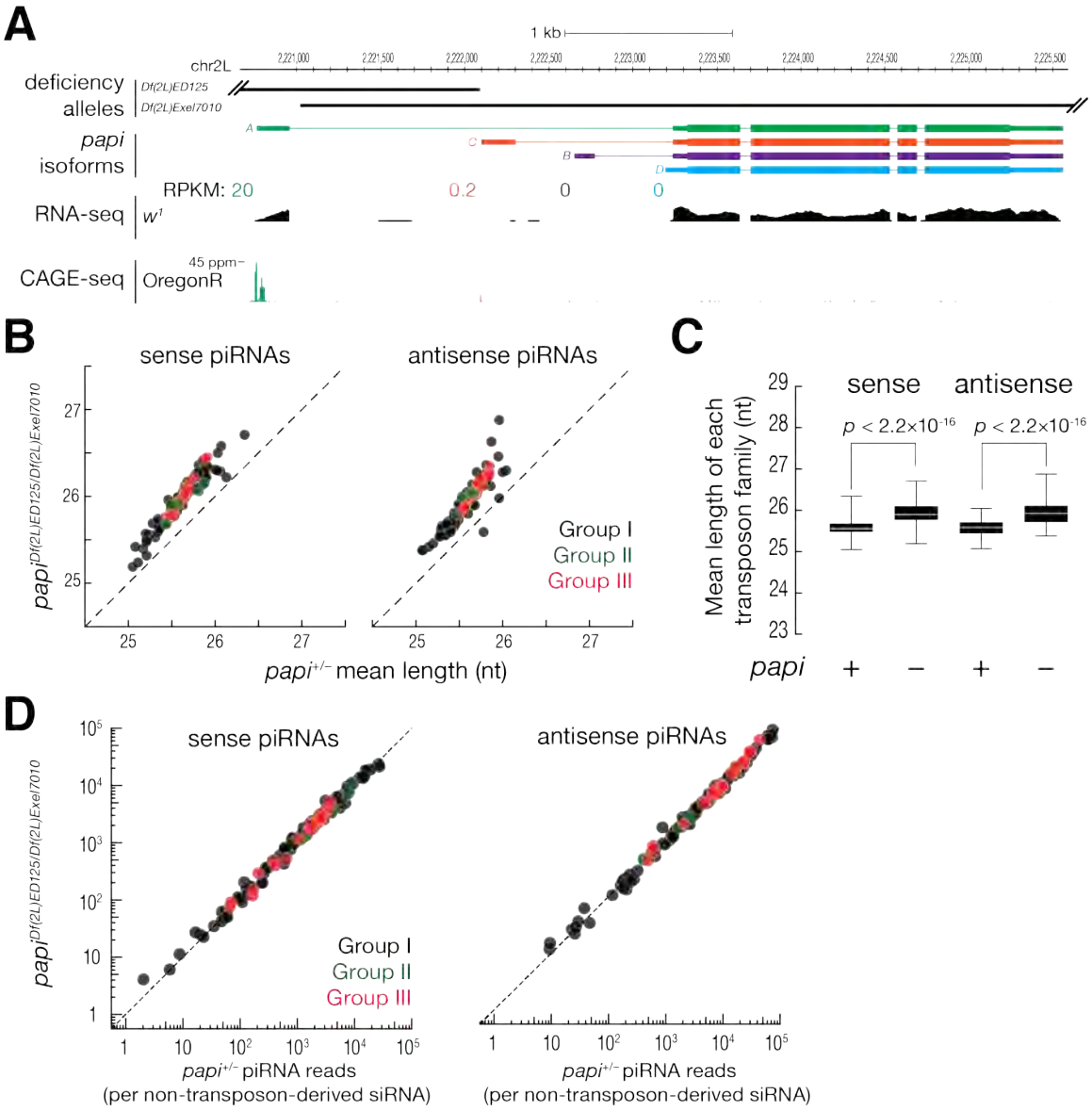

A
1 kb
chr2L
deficiency alleles
Df(2L)ED125
Df(2L)Exel7010
papi isoforms
RPKM: 20
0.2
0
0
RNA-seq
w1
CAGE-seq
45 ppm
OregonR
B
sense piRNAs
antisense piRNAs
papiDf(2L)ED125/Df(2L)Exel7010
papi+/− mean length (nt)
Group I
Group II
Group III
C
sense
antisense
p < 2.2×10-16
p < 2.2×10-16
Mean length of each transposon family (nt)
papi
D
sense piRNAs
antisense piRNAs
papiDf(2L)ED125/Df(2L)Exel7010
papi+/− piRNA reads
(per non-transposon-derived siRNA)
Group I
Group II
Group III

Figure 3.11. Papi and 3′ trimming in piRNA biogenesis

(A) Gene model for fly *papi* with RPKM values shown for each mRNA isoform calculated using RNA-seq data from w^1 ovaries. CAGE-seq data from Oregon R ovaries is also shown.

(B-C) Scatterplots (B) and boxplots (C) compare the mean lengths of piRNAs from each transposon family. *p*-values were calculated using a paired Wilcoxon test.

(D) Scatterplots compare the abundance of piRNAs for each transposon family.

Figure 3.12

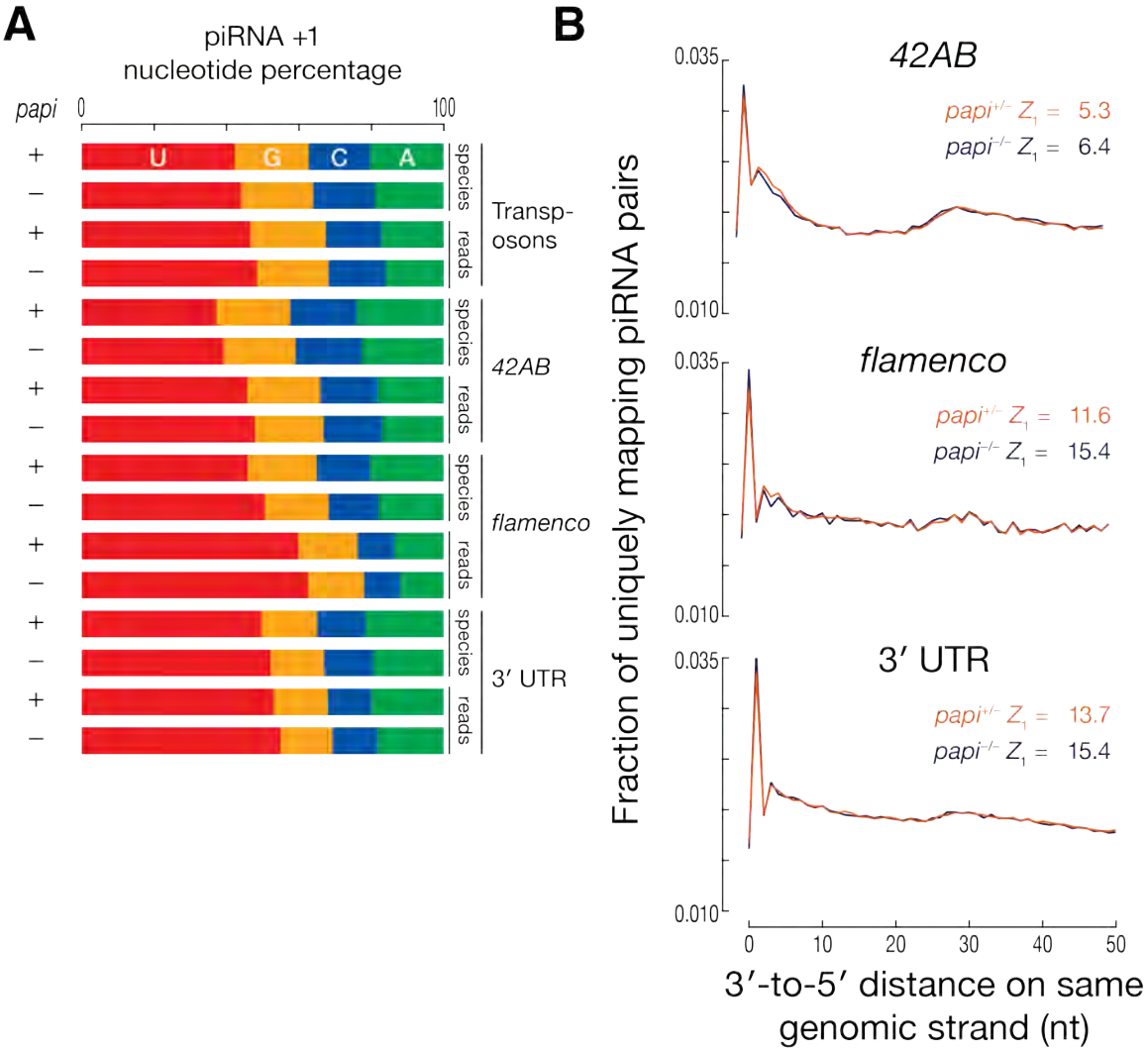

A
piRNA +1
nucleotide percentage
papi
0
100
U
G
C
A
species
reads
Transp-
osons
42AB
flamenco
3′ UTR
B
42AB
papi+/– Z1 = 5.3
papi–/– Z1 = 6.4
flamenco
papi+/– Z1 = 11.6
papi–/– Z1 = 15.4
3′ UTR
papi+/– Z1 = 13.7
papi–/– Z1 = 15.4
0.035
0.010
Fraction of uniquely mapping piRNA pairs
0
10
20
30
40
50
3′-to-5′ distance on same
genomic strand (nt)

Figure 3.12. Phasing in *papi* mutants.

(A) The +1 nucleotide percentage was determined for piRNAs derived from transposons, the *42AB* cluster, the *flamenco* locus, and the 3′ UTRs of mRNAs. (B) Distance from 3′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs for piRNAs derived from *42AB*, *flamenco*, and 3′ UTRs.

that in mice, ~3–10 nt are removed from the 3′ ends of the immediate precursors of piRNAs.

Supporting this idea, *tdrd2*$^{-/-}$ mutant testes accumulate 31–37 nt RNAs instead of 26–30 nt piRNAs; most of these longer species share their 5′ ends with mature piRNAs from *tdrd2*$^{+/-}$ heterozygotes (Saxe et al., 2013). Analysis of 3′-to-5′ piRNA distance in *tdrd2*$^{-/-}$ testes from mice at 11 dpp showed clear evidence for phasing (Figure 3.14A). Moreover, the 3′ ends of the longer RNAs present in *tdrd2*$^{-/-}$ mutant testes were generally followed by a uridine (Figures 3.14B and 3.13D); as in flies, piRNAs in mice typically begin with uridine. For both *tdrd2*$^{+/-}$ and *tdrd2*$^{-/-}$, 5′-to-5′ piRNA distance analysis showed broad peaks at 35–43 nt—the same length as the longer piRNA precursor RNAs detected in the *tdrd2*$^{-/-}$ mutant (Saxe et al., 2013) (Figure 3.14C).

Taken together, the data suggest a simple model for the production of piRNAs in the mouse testis. First, cleavage of piRNA cluster transcripts generates RNAs with 5′ monophosphorylated ends. These RNAs then bind MIWI (PIWIL1) or MILI (PIWIL2), which can next recruit PLD6, the mouse ortholog of fly Zuc. Subsequently, PLD6 travels down the RNA, perhaps propelled by the helicase protein MOV10L1, the mouse ortholog of fly Armitage, cleaving on the 5′ side of the first uridine that lies 31–37 nt downstream. Subsequent 3′ trimming and methylation of the 3′ terminal, 2′ hydroxyl by HEN1 completes piRNA maturation (Horwich et al., 2007; Kirino and Mourelatos, 2007; Saito et al., 2007; Kawaoka et al., 2011).

Figure 3.13

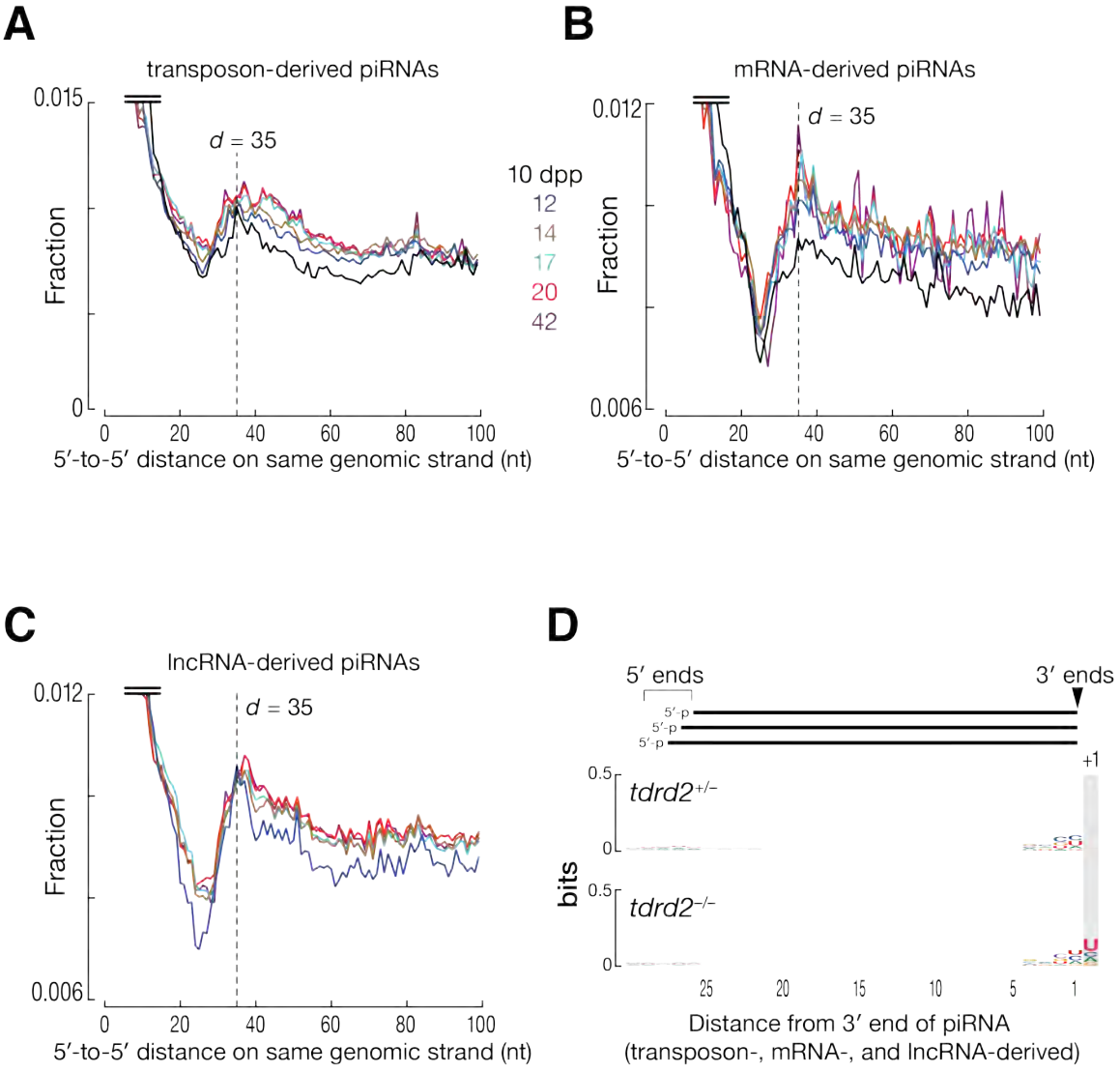

A
transposon-derived piRNAs
0.015
0
Fraction
d = 35
0
20
40
60
80
100
5′-to-5′ distance on same genomic strand (nt)
10 dpp
12
14
17
20
42
B
mRNA-derived piRNAs
0.012
0.006
Fraction
d = 35
5′-to-5′ distance on same genomic strand (nt)
C
lncRNA-derived piRNAs
0.012
0.006
Fraction
d = 35
5′-to-5′ distance on same genomic strand (nt)
D
5′ ends
3′ ends
5′-p
+1
0.5
0
bits
tdrd2+/−
tdrd2−/−
25
20
15
10
5
1
Distance from 3′ end of piRNA
(transposon-, mRNA-, and lncRNA-derived)

Figure 3.13. Piwi-associated piRNAs map immediately after the 3′ ends of Aub- and Ago3-associated piRNAs.

(A–C) Distance from 5′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs for uniquely mapping, transposon- (A), mRNA- (B), or lncRNA- (C) derived piRNAs from wild-type mouse testes at times after birth was calculated using published data.

(D) Sequence logo shows nucleotide composition of piRNA species lying 29 nt upstream and 1 nt downstream of the 3′ ends of uniquely mapping piRNAs.

Figure 3.14

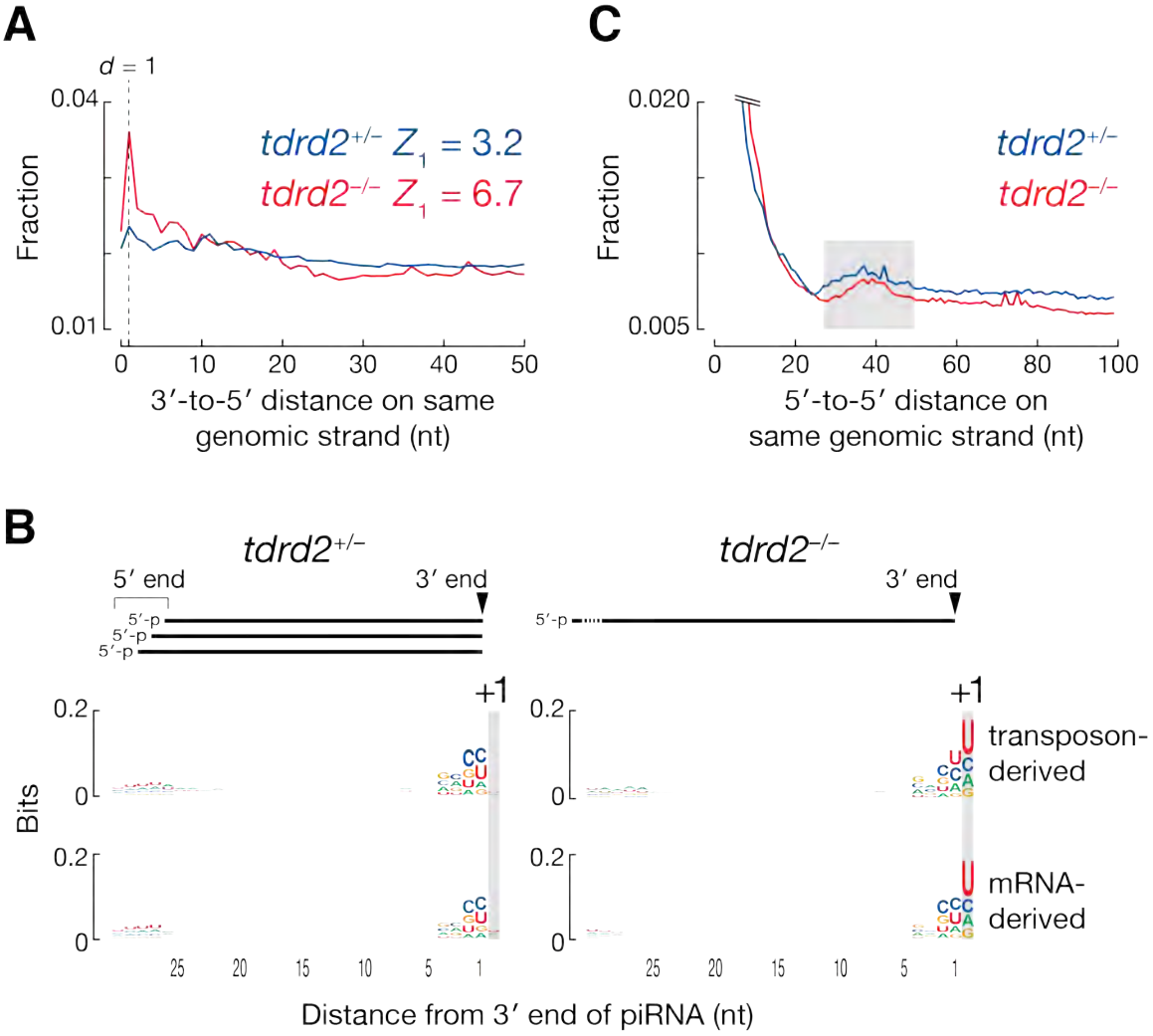

A
d = 1
0.04
0.01
Fraction
tdrd2+/– Z1 = 3.2
tdrd2–/– Z1 = 6.7
0
10
20
30
40
50
3′-to-5′ distance on same genomic strand (nt)
C
0.020
0.005
Fraction
tdrd2+/–
tdrd2–/–
0
20
40
60
80
100
5′-to-5′ distance on same genomic strand (nt)
B
tdrd2+/–
tdrd2–/–
5′ end
3′ end
5′-p
+1
0.2
0
Bits
transposon-derived
mRNA-derived
25
20
15
10
5
1
Distance from 3′ end of piRNA (nt)

Figure 3.14. Mouse piRNAs display phasing.

(A) Distance from 3′ ends of upstream piRNAs to 5′ ends of downstream piRNAs on the same genomic strand for uniquely mapping piRNAs in *tdrd2*$^{+/-}$ and *tdrd2*$^{-/-}$ in mouse testes at 11 dpp.

(B) The nucleotide composition, in species, of sequences 29 nt upstream and 1 nt downstream of the 3′ ends of uniquely mapping piRNAs. Pachytene piRNAs are not included because spermatogenesis arrests before the pachytene stage in *tdrd2*$^{-/-}$.

(C) Distance from 5′ ends of upstream piRNAs to 5′ ends of downstream piRNAs on the same genomic strand for uniquely mapping piRNAs in *tdrd2*$^{+/-}$ and *tdrd2*$^{-/-}$ mouse testes at 11 dpp. Data are from Saxe *et al*. (GSE47151).

Discussion

Our findings suggest a substantially revised model for primary piRNA biogenesis (Figure 3.15). The model proposes that each cycle of Ping-Pong amplification can generate one secondary piRNA and multiple primary piRNAs. For example, a secondary piRNA bound to Ago3 can direct cleavage of a fully or partially complementary target RNA (Wang et al., 2014). The resulting 3′ cleavage product then binds Aub. An unknown factor, possibly Armi, recruits Zuc, which makes a second cut 26–29 nt away from the 5′ monophosphate, likely at the first uridine not occluded by Aub. The two cleavage products from this reaction follow decidedly different fates. The 5′ fragment matures into a secondary piRNA bound to Aub. We envision that some but not all of such Aub-bound RNA fragments will require 3′ trimming to achieve their characteristic length (Honda et al., 2013). The 3′ fragment becomes a substrate for the production of phased primary piRNAs by Zuc. With the aid of Armi, Zuc travels 5′-to-3′ cleaving every ~26 nt. The piRNAs released by this process load mainly into Piwi. Although as much as 90% of Piwi-associated piRNAs are generated by this mechanism in the germline, piRNAs in the soma, which lacks Aub and Ago3, must deploy a different mechanism to initiate Zuc-dependent processing (Figs. 2, 4 and 6).

Our data also help explain why effective transposon silencing requires heterotypic Aub:Ago3 Ping-Pong amplification. In *ago3* mutant ovaries, homotypic Aub:Aub Ping-Pong replaces heterotypic Ping-Pong. Although antisense piRNAs are produced by homotypic Ping-Pong, they fail to silence transposon expression (Li et al., 2009; Zhang et al., 2011). We

Figure 3.15

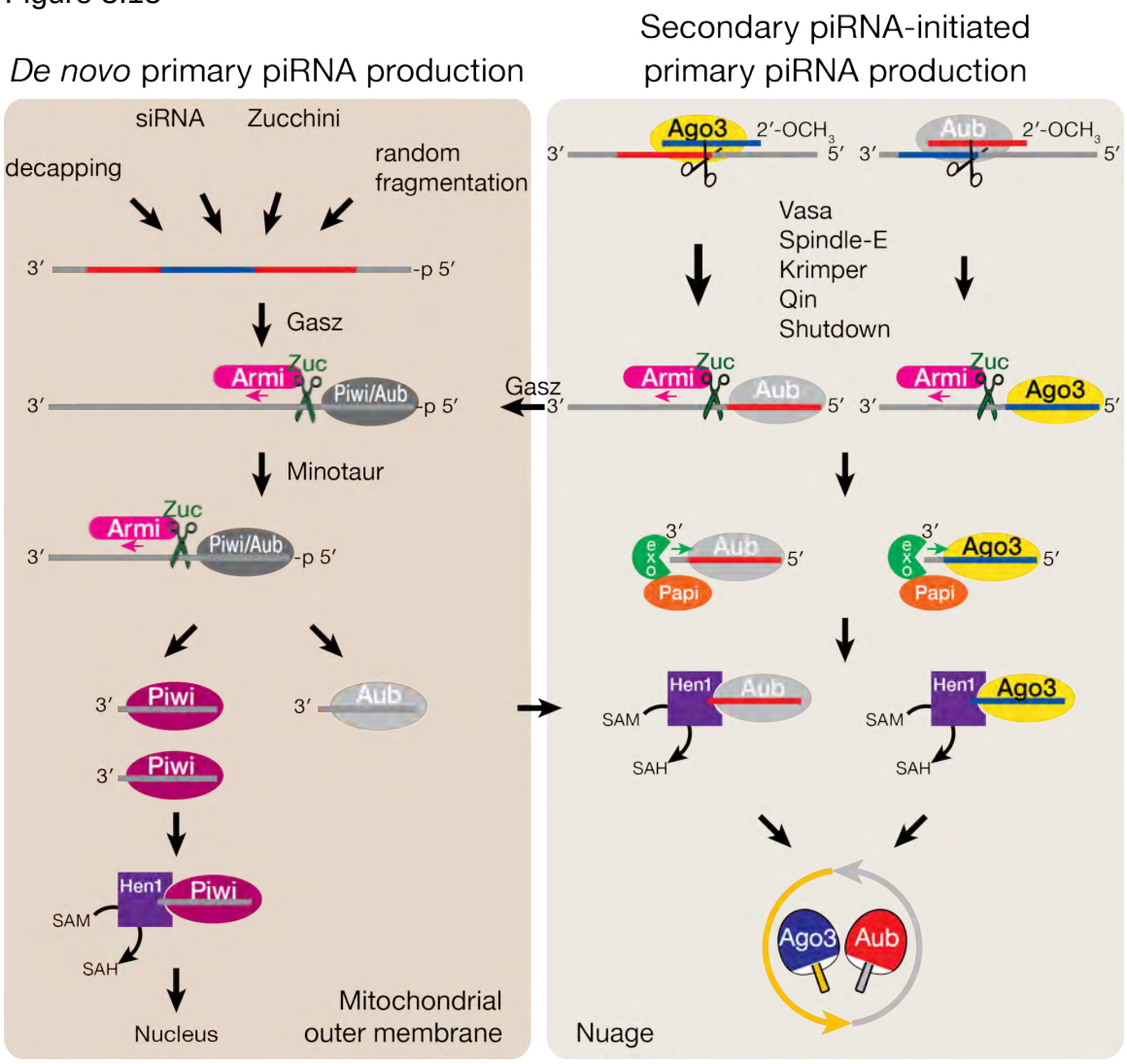

De novo primary piRNA production
Secondary piRNA-initiated primary piRNA production
siRNA
Zucchini
decapping
random fragmentation
3′
-p 5′
Gasz
Armi
Zuc
Piwi/Aub
Minotaur
Piwi
Aub
Hen1
SAM
SAH
Nucleus
Mitochondrial outer membrane
Ago3
2′-OCH3
5′
Vasa
Spindle-E
Krimper
Qin
Shutdown
exo
Papi
Nuage

Figure 3.15. A revised model for piRNA biogenesis.

(Left) The de novo primary piRNA pathway starts with piRNA intermediates released from piRNA cluster transcripts in an Aub- and Ago3-independent manner. Zuc slices them consecutively every ~26 nt, aided by Armi and other factors in the primary pathway (e.g., Minotaur and Gasz). Those primary piRNAs are loaded into Piwi and Aub, but not Ago3.

(Right) In nuage, cleavage by Ago3 or Aub produces piRNA intermediates with a 5′ monophosphate. The 3′ cleavage products are loaded into Aub and Ago3, followed by Zuc-dependent cleavage ~26 nt from their 5′ ends. This cleavage produces the 3′ ends of the “Ping-Pong partner” secondary piRNA and the 5′ ends of long RNAs that become substrates for Zuc, which processively cleaves the RNA to generate phased piRNAs loaded into Piwi and, to a lesser extent, Aub. We propose that the Zuc machinery chooses as its cleavage site the first uridine that is not protected by a PIWI protein. Consequently, some pre-piRNAs require Papi-dependent 3′ trimming before their 3′ ends are methylated by Hen1.

propose that homotypic Aub:Aub Ping-Pong is unable to replace heterotypic Ping-Pong, because it cannot generate enough Piwi-bound primary piRNAs.

Finally, the Ping-Pong model does not explain the stunning diversity of piRNA sequences: each cycle of Ping-Pong increases the abundance of a pair of piRNAs, but cannot generate piRNAs with novel sequence specificity (piRNA nucleotides 2–16) (Wang et al., 2014). Our data show that each RNA cleaved by Aub or Ago3 not only produces a secondary, Ping-Pong piRNA partner, but also produces primary piRNAs from the sequences immediately 3′ to the secondary piRNA. Such a spreading mechanism calls to mind features of siRNA production in *Caenorhabditis elegans* and *Arabidopsis thaliana* (Xie et al., 2005; Yoshikawa et al., 2005; Bagijn et al., 2012; Lee et al., 2012), and primed CRISPR adaptation in *Escherichia coli* (Datsenko et al., 2012; Swarts et al., 2012; Heler et al., 2014). Although the detailed mechanisms differ (e.g., slicing activity is dispensable in *C. elegans,* and an RNA-dependent RNA polymerase is required in *A. thaliana*), signal amplification and sequence diversification is clearly a recurrent theme for RNA-guided silencing in animals, plants, and bacteria.

Experimental Procedures

General methods

Stocks and crosses were grown at 25°C. All flies were in the w^1 background, except w^+; *Df(2L)Prl* (zuc^{Df}) (Pane et al., 2007) and the *papi* strains w^{1118}; *Df(2L)ED125* and w^{1118}; *Df(2L)Exel7010*. Ovaries were dissected in modified Robb's Buffer (55 mM CH_3COONa, 40 mM CH_3COOK, 100 mM sucrose, 10 mM glucose, 1.2 mM $MgCl_2$, 1 mM $CaCl_2$, 100 mM HEPES, pH 7.4). RNAs were purified using mirVana (Ambion, Life technologies, CA, USA).

Small RNA library construction

Total RNA (100 µg) or RNA co-immunoprecipitated with Aub, Ago3, or Piwi was purified by 15% urea polyacrylamide gel electrophoresis (PAGE), selecting for 18–30 nt long RNAs. Oxidization of RNA with $NaIO_4$ was used to deplete miRNAs and enrich for siRNAs and piRNAs (Li et al., 2009). Ligation of the 3′ adaptor (5′-rApp NNN TGG AAT TCT CGG GTG CCA AGG /ddC/-3′ or 5′-rApp TGG AAT TCT CGG GTG CCA AGG /ddC/-3′.) using truncated, K227Q mutant T4 RNA Ligase 2 at 25°C for ≥16 h and subsequent size selection by 15% PAGE was as described (Li et al., 2009). To exclude 2S rRNA from sequencing libraries, 10 pmol 2S blocker oligo was added before 5′ adaptor ligation (Wickersheim and Blumenstiel, 2013); 5′ adaptor was added using T4 RNA ligase (Ambion) at 25°C for ≥ 2 h, followed by reverse-transcription using AMV reverse transcriptase (New England Biolabs, MA, USA) and PCR using Q5 polymerase (NEB). An Illumina HiSeq 2000 was used for high-throughput, single-end 50 nt or 100 nt sequencing.

Degradome-seq library construction

Fresh RNA (4 µg) was subjected to two rounds of rRNA depletion (Ribo-Zero; Epicentre, WI, USA), treated with turbo DNase (Ambion), and then size-selected to isolate RNA ≥ 200 nt (DNA Clean & Concentrator™-5, ZYMO RESEARCH, CA, USA). T4 RNA ligase (Ambion) was used at 25°C for 2–4 hours for 5′ ligation. Reverse transcription with SuperScript III (Life Technologies) employed a primer containing degenerate sequence at its 3′ end (5′-GCA CCC GAG AAT TCC ANN NNN NNN-3′). cDNA was amplified by PCR using Q5 polymerase (NEB), and 200–400 nt dsDNA was isolated from using 6% native PAGE. An Illumina HiSeq 2000 was used to perform paired-end, 100 nt sequencing of the dsDNA products.

Small RNA immunoprecipitation

Anti-Piwi, Aub, and Ago3 (Brennecke et al., 2007) antibodies (~10 µg) were incubated with Protein A and G Dynabeads (15 µl each; Life Technologies) in lysis buffer (30 mM HEPES-KOH, pH 7.4, 100 mM CH_3COOK, 2 mM $(CH_3COO)_2Mg$, 5 mM dithiothreitol, 0.5% [v/v] NP-40, 1 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride, 0.3 µM Aprotinin, 40 µM Bestatin, 10 µM E-64, 10 µM Leupeptin) at 4°C for 4 h with rotation, then washed twice with lysis buffer. Next, 400–800 µl freshly prepared ovary lysate (5 µg/µl) was added and incubated at 4°C for 4 h with rotation. After washing the beads four times with ice-cold lysis buffer, RNA was purified using Trizol (Life Technologies).

General bioinformatics analyses

Analyses were performed using piPipes v1.4 (Han et al., 2015b). Briefly, all small RNA sequencing libraries were filtered using PHRED score ≥ 5. Genome mapping using Bowtie v1.1.0 allowed no mismatches for fly and one mismatch for mouse data (Langmead et al., 2009). Degradome mapping was performed with Bowtie2 v2.2.3 (to rRNA) and STAR v2.3.0 (to genome) (Dobin et al., 2013a). Reads whose 5′ ends could not be determined precisely (soft-clipped) during alignment were removed computationally. Alignments were categorized by genomic feature using BEDTools v2.17.0 (Quinlan and Hall, 2010). For transgene mapping, including *P{GSV6}*, *P{lArB}* and *P{A92}*, we first aligned an oxidized small RNA-seq library from w^1 (23,712,713 genome-mapping reads) to the transgene sequence, masking (turning into Ns) positions that could be mapped to piRNAs more abundant than 1 parts per million (ppm). Statistical analysis in R 3.0.2 (Team, 2013) required p value < 0.005. To compare piRNA abundance between two small RNA libraries, we normalized to non-transposon-derived siRNAs, rather than uniquely mapping reads of the genome, in order to avoid biasing genotypes such as *zuc*, in which piRNA abundance was decreased globally. To compare the abundance of piRNAs associated with Piwi, we normalized to *flamenco*-derived reads, which are unaffected by defects in the germline piRNA pathway.

Phasing analysis

Reads were mapped to genome, alignments that overlapped with rRNAs, tRNAs and snoRNAs were removed, and the remaining 23–29 nt RNAs (fly piRNAs) or 23–35 nt (mouse piRNAs) were analyzed. To analyze small RNAs in *tdrd2*, all

reads ≥ 23 nt were used. The score for a distance of *x* nt was calculated by ∑minimal (M_i, N_{i+x}) where M_i is the number of reads whose 3′ ends are located at position *i* and N_{i+x} is the number of reads whose 5′ ends are located at position *i+x*. When *x* equals 0, the 3′ and 5′ ends overlap. When *x* equals to 1, the 5′ end is immediately downstream of the 3′ end (phasing). For analyses including multi-mappers, reads were partitioned by the number of times they can be aligned to the genome. To calculate Z_1, overlaps at position 2–20 nt were used as background to calculate *Z* scores. In Ping-Pong analyses, the product, instead of the smaller value, of M and N was used (Oey et al., 2011).

Assigning immunopurified small RNA reads to Piwi, Aub, or Ago3

We used a χ^2 test with a *p*-value cutoff < 0.005 to test whether a sequence was enriched in one of the three PIWI proteins. A sequence could be unambiguously assigned only when one of two conditions was met: (1) the sequence was uniquely sequenced in only one of the three libraries (two for mutants lacking one PIWI protein) or (2) the sequence passed the χ^2 test ($p < 0.005$) and was at least five-fold more abundant in one sample than the other two.

Acknowledgments

We thank Alicia Boucher, Cindy Tipping, Gwen Farley and Ellen Kittler for technical assistance; Ryuya Fukunaga and Erik Sontheimer for insightful discussions; Julius Brennecke for reagents, helpful discussions and sharing unpublished data; and members of the Weng and the Zamore laboratories for advice and critical comments on the manuscript. This work was supported in part by National Institutes of Health grants HG007000 to Z.W. and by GM62862 and GM65236 to P.D.Z.

Chapter IV Slicing and Binding by Ago3 or Aub Trigger Piwi-bound piRNA Production by Distinct Mechanisms

Preface

This chapter was a product of a collaborative effort: Bo W Han designed the new constructs for transgenes Ago3 and Aub. I performed fly genetics for catalytically inactive Aub and constructed deep sequencing libraries correspondingly. Bo W Han performed fly genetics for catalytically inactive Ago3 and constructed deep sequencing libraries correspondingly. Bo W Han performed all the other experiments including GRO-seq and ChIP-Seq. Bo W Han and I performed the computational analyses. Bo W Han discovered the alternative pathway. Daniel Tianfang Ge offered a protocol for high-resolution immunofluorescence experiments. Zhao Zhang provided the sequencing data for *piwi* mutant. I reconceived the framework of this project after a similar story (similar to my qualifying exam proposal) was published in the middle of 2014.

Summary

In *Drosophila* ovarian germ cells, PIWI-interacting RNAs (piRNAs) direct Aubergine and Argonaute3 to cleave transposon transcripts and instruct Piwi to repress transposon transcription, thereby safeguarding the germline genome. Here, we report that RNA cleavage by Argonaute3 initiates production of most Piwi-bound piRNAs. We find that the cardinal function of Argonaute3, whose piRNA guides predominantly correspond to sense transposon sequences, is to produce antisense piRNAs that direct transcriptional silencing by Piwi, rather than to make piRNAs that guide post-transcriptional silencing by Aubergine. We also find that the Tudor domain protein Qin prevents Aubergine's cleavage products from becoming Piwi-bound piRNAs, ensuring that antisense piRNAs guide Piwi. Although Argonaute3 slicing is required to efficiently trigger phased piRNA production, an alternative, slicing-independent pathway suffices to generate Piwi-bound piRNAs that repress transcription of a subset of transposon families. This alternative pathway may help flies silence newly acquired transposons for which they lack extensively complementary piRNAs.

Introduction

In animals, PIWI-interacting RNAs (piRNAs), a diverse class of 23−36 nucleotide (nt) small silencing RNAs, repress transposons in the germline, ensuring the faithful transfer of genomic information from generation to generation (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Vagin et al., 2006). Disrupting the piRNA pathway activates transposon transcription, arresting germ cell development (Wilson et al., 1996; Lin and Spradling, 1997; Deng and Lin, 2002; Klattenhoff et al., 2007) and making one or both sexes infertile.

Drosophila germline nurse cells produce piRNAs by at least two mechanisms: the *de novo* pathway and the Ping-Pong cycle. *De novo* piRNA production begins with the fragmentation of RNA derived from piRNA clusters, discrete genomic loci comprising transposons and repetitive sequences (Brennecke et al., 2007). An endonuclease localized to the outer membrane of mitochondria, Zucchini (Zuc), assisted by proteins such as Armitage and Gasz, is thought to process those RNA fragments into primary piRNAs (Pane et al., 2007; Malone et al., 2009; Czech et al., 2013; Handler et al., 2013). The resulting piRNAs load into the Argonaute proteins Piwi and Aubergine (Aub), but not Argonaute3 (Ago3; Olivieri et al., 2012; Sato et al., 2015).

piRNAs enable Piwi to move from the cytoplasm to the nucleus, where it suppresses transposon transcription (Klenov et al., 2011; Wang and Elgin, 2011; Sienski et al., 2012; Le Thomas et al., 2013; Rozhkov et al., 2013). In contrast, Aub remains in the cytoplasm where it cleaves transposon transcripts. Like other slicing-competent Argonaute proteins, Aub hydrolyzes the phosphodiester bond between the target nucleotides paired with the tenth and eleventh nucleotides of its piRNA guide (Elbashir et al., 2001b). The 3′ target cleavage product, whose 5′

end bears a monophosphate, is proposed to be transferred to Ago3 by the DEAD-box RNA helicase Vasa (Xiol et al., 2014). Further processing converts the first ~25 nt of an Ago3-bound 3′ cleavage product into a mature secondary piRNA that can direct Ago3 to cleave complementary RNAs (Kawaoka et al., 2011).

Ago3 piRNAs correspond largely to sense transposon sequences. Cleavage by Ago3 of antisense transposon sequences transcribed from piRNA clusters generates 5′ monophosphorylated RNAs that can be transferred to Aub to generate new, antisense, secondary piRNAs. These can initiate another round of transposon RNA slicing and piRNA production. This “Ping-Pong” cycle of reciprocal cleavages amplifies the abundance of sense and antisense piRNAs, generating a characteristic 10 bp overlap between the 5′ ends of sense (Ago3): antisense (Aub) piRNA pairs (Brennecke et al., 2007; Gunawardane et al., 2007b).

The production of secondary piRNAs by the Ping-Pong cycle was long believed to be initiated by the *de novo* primary pathway (Senti and Brennecke, 2010; Siomi et al., 2011; Guzzardo et al., 2013). However, recent evidence suggests that the cleavage products of Ago3 and Aub not only yield secondary Ping-Pong piRNAs, but also trigger the Zuc-dependent production of phased, primary piRNAs that are loaded into Piwi (Han et al., 2015a; Mohn et al., 2015). Production of such Ping-Pong-dependent, primary piRNAs spreads piRNA production beyond the sites of cleavage by Ago3 and, to a lesser extent, Aub, thereby introducing new sequence diversity into the piRNA pool.

Here, we report that a majority of Piwi-bound piRNAs are made from the cleavage products of Ago3 and Aub. Without Aub and Ago3, the abundance of Piwi-bound piRNAs declines, and Piwi fails to enter the nucleus. Our data suggest that the cardinal function of Ago3 is to produce antisense piRNAs that

direct transcriptional silencing by Piwi, rather than to generate piRNAs that guide post-transcriptional silencing by Aub. Piwi-bound piRNAs are mainly initiated from the cleavage products of Ago3, not Aub, because the Tudor domain protein, Qin, prevents the products of Aub cleavage from becoming Piwi-bound piRNAs. Thus, Qin ensures that antisense piRNAs guide Piwi. Although endonucleolytic RNA cleavage by Ago3 is required for efficient production of Piwi-bound, phased piRNAs, an alternative, cleavage-independent pathway suffices to generate Piwi-bound piRNAs that repress transcription of a subset of transposon families.

Results

Ago3 and Aub Initiate Production of Most Germline Piwi-Bound piRNAs

The current model for piRNA-directed silencing posits that transposons are repressed in the *Drosophila* female germline both transcriptionally by nuclear Piwi (Klenov et al., 2011; Sienski et al., 2012; Le Thomas et al., 2013; Rozhkov et al., 2013) and post-transcriptionally by cytoplasmic Aub (Brennecke et al., 2007; Gunawardane et al., 2007b; Li et al., 2009; Malone et al., 2009; Zhang et al., 2011). Because the Ping-Pong cycle uses primary piRNAs to generate antisense, Aub-bound, secondary piRNAs, Ping-Pong amplification was assumed to feed piRNAs to Aub, fueling post-transcriptional silencing. Recently, sense secondary piRNAs were found to direct Ago3 to initiate the production of phased, Piwi-bound, antisense piRNAs from the same RNAs that produce Aub-bound, secondary piRNAs (Han et al., 2015a; Mohn et al., 2015). The discovery of this secondary piRNA-initiated primary piRNA pathway explains why Piwi-bound piRNAs dwindle in mutants defective in Ping-Pong amplification. However, the relative contributions of Piwi-directed transcriptional silencing and Aub-directed post-transcriptional silencing are unknown.

The requirement for Ping-Pong amplification in the production of piRNA guides for Piwi suggests that (1) the majority of Piwi-bound piRNAs are made by the secondary piRNA-initiated primary piRNA pathway and (2) the *de novo* piRNA pathway on its own cannot produce sufficient Piwi-bound piRNAs to silence transposons. To test these ideas, we measured the abundance of piRNAs in *aub*; *ago3* double-mutants—which contain only Piwi-bound piRNAs—and of Piwi-bound piRNAs in w^1 ovaries. In w^1, Piwi-bound piRNAs comprise

maternally deposited, *de novo*, and secondary piRNA-initiated primary piRNAs (Brennecke et al., 2008; Malone et al., 2009; Han et al., 2015a; Mohn et al., 2015). In contrast, *aub*; *ago3* double-mutant ovaries contain Piwi-bound maternal and *de novo* piRNAs but lack secondary piRNA-initiated primary piRNAs. Normalizing to the abundance of somatic *flamenco*-derived piRNAs allowed us to compare the two genotypes. We limited our analysis to 47 germline-specific transposon families whose transcript abundance increases in ovaries from *aub* or *ago3* single mutants (see below) or when *piwi* is depleted from the germ cells but does not increase when *piwi* is depleted from the somatic follicle cells by RNAi (Rozhkov et al., 2013). For these 47 families, median Piwi-bound piRNA abundance in $aub^{HN2/QC42}$; $ago3^{t2/t3}$ ovaries was 2% of the w^{1} control (Figure 4.1A). An additional 25 transposon families are expressed in both the soma and the germline—their transcript abundance increases when *piwi* is depleted by tissue-specific RNAi in either the soma or the germline. For these 25 "intermediate" transposons, median Piwi-bound piRNA abundance in $aub^{HN2/QC42}$; $ago3^{t2/t3}$ ovaries was 3% of w^{1}. Thus, Aub and Ago3 generate most Piwi-bound piRNAs in the germline.

If secondary piRNAs initiate production of most Piwi-bound, germline piRNAs, then mutants defective for Ping-Pong amplification should impair transcriptional silencing of transposons. To test this prediction, we measured sense transposon transcription using global run-on sequencing (GRO-seq; Core et al., 2008). Among the 72 families (47 germline-specific and 25

Figure 4.1

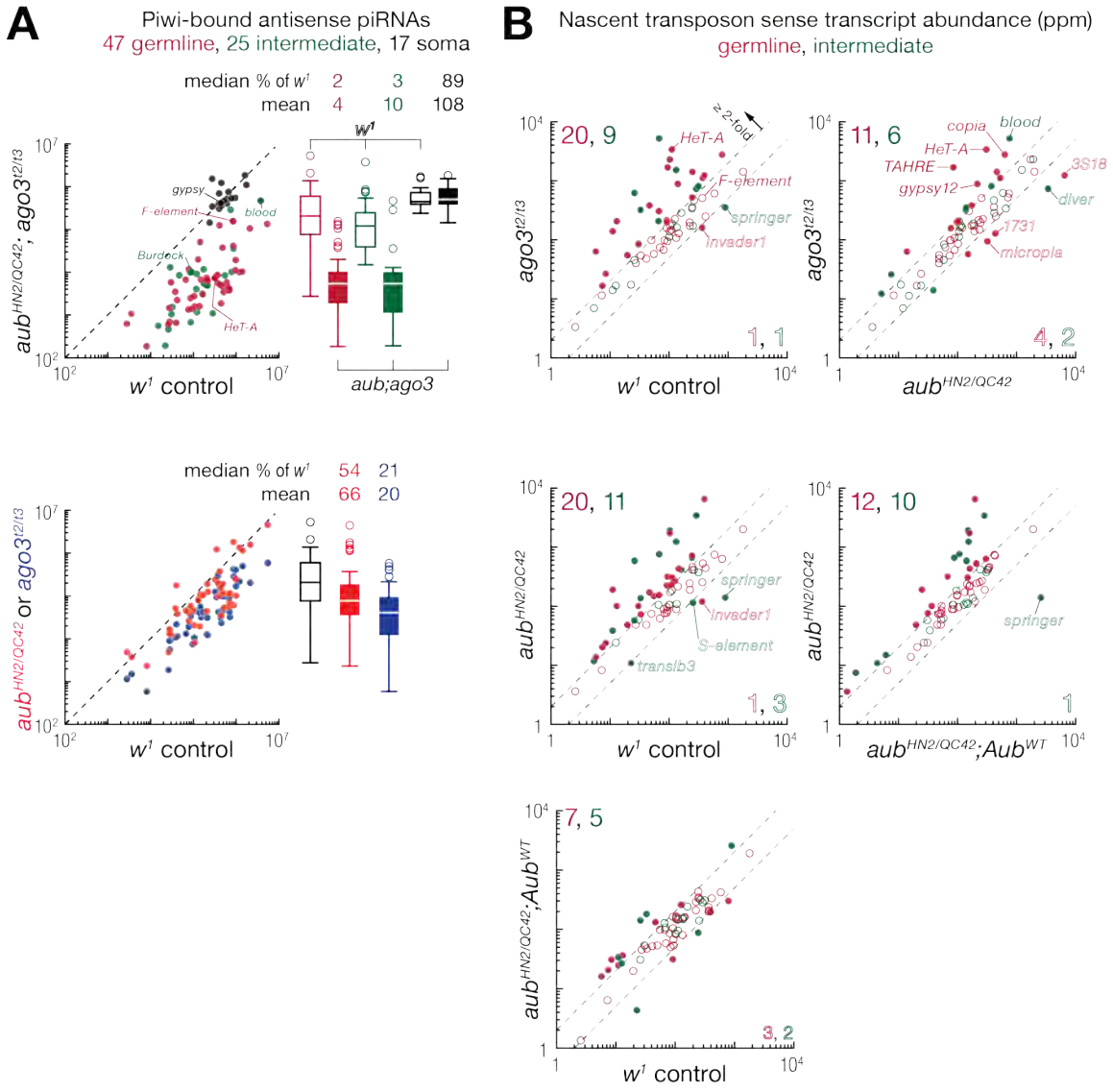

A
Piwi-bound antisense piRNAs
47 germline, 25 intermediate, 17 soma
median % of w1 2 3 89
mean 4 10 108
aubHN2/QC42; ago3t2/t3
w1 control
aub;ago3
gypsy
F-element
blood
Burdock
HeT-A
median % of w1 54 21
mean 66 20
aubHN2/QC42 or ago3t2/t3
B
Nascent transposon sense transcript abundance (ppm)
germline, intermediate
20, 9
≥ 2-fold
HeT-A
F-element
springer
invader1
ago3t2/t3
1, 1
11, 6
copia
blood
HeT-A
TAHRE
3S18
gypsy12
diver
1731
micropia
4, 2
aubHN2/QC42
20, 11
springer
invader1
S-element
transib3
1, 3
12, 10
springer
1
aubHN2/QC42;AubWT
7, 5
aubHN2/QC42;AubWT
3, 2

Figure 4.1. Most Piwi-bound piRNAs are Generated by Aub and Ago3

(A) Scatter and box plots on a common *y*-axis comparing antisense piRNA abundance, measured by small RNA-seq. Transposons were classified as germline-specific, soma-specific, and “intermediate” according to the change in abundance of their sense transcripts in *aub* and *ago3* and soma- and germline-specific depletion of *piwi* by RNAi. piRNA abundance was normalized to uniquely mapping reads from the somatic *flamenco* cluster and are reported in parts per million (ppm).

(B) Scatter plot comparing nascent, sense transposon transcript abundance, measured by GRO-seq.

intermediate) examined, sense transcription for 20 germline-specific and 9 intermediate transposons more than doubled without Ago3 (Figure 4.1B). Without Aub, sense transcription was similarly affected for 20 germline and 11 intermediate transposon families. Of the transposons whose transcription increased in *ago3*$^{t2/t3}$, 14 germline and 6 intermediate transposons also increased in *aub*$^{HN2/QC42}$. Overall, expression of 26 germline and 14 intermediate transposons—more than half of the 72 transposons analyzed—increased in *aub* or *ago3* mutant ovaries. Loss of Ago3 decreased Piwi-bound, antisense piRNAs to a greater extent than loss of Aub, and nascent transposon transcripts increased more in *ago3*$^{t2/t3}$ than *aub*$^{HN2/QC42}$ (Figures 4.1A and 4.1B). The decrease in Piwi-bound piRNA abundance and the increase in transposon transcription were a direct consequence of the loss of Ago3 or Aub, as introduction of the corresponding wild-type transgenic PIWI proteins rescued transposon silencing (Figure 4.1B; see below). We conclude that Ago3 plays a central role in the production of the Piwi-bound, antisense piRNAs that silence transposon transcription in the fly germline.

Secondary piRNA-Dependent Primary piRNAs License Piwi to Transit to the Nucleus

Without piRNAs, Piwi is depleted from the nucleus (Le Thomas et al., 2013), and in *zuc*$^{HM27/Df}$ mutant ovaries, which are depleted of piRNAs, the abundance of both somatic and germline Piwi protein declines (Figure 4.2A and 4.2B; Olivieri et al., 2010; Olivieri et al., 2012; Sienski et al., 2012). In *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ mutant ovaries, Piwi protein levels were reduced in germline nurse cells, but not in somatic follicle cells, and germline Piwi

Figure 4.2

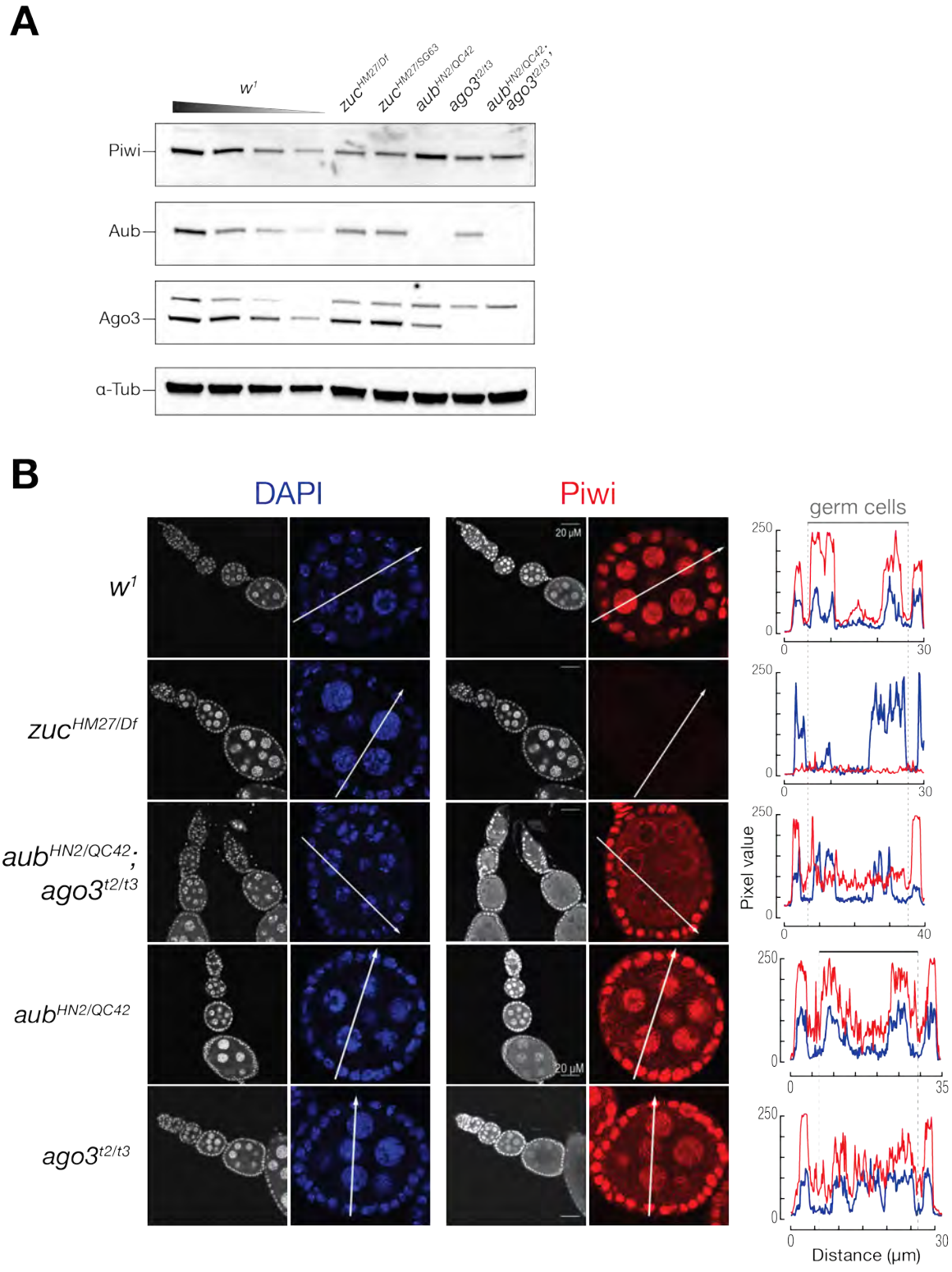

A
w1
zucHM27/Df
zucHM27/SG63
aubHN2/QC42
ago3t2/t3
aubHN2/QC42; ago3t2/t3
Piwi
Aub
Ago3
α-Tub
B
DAPI
Piwi
germ cells
w1
zucHM27/Df
aubHN2/QC42; ago3t2/t3
aubHN2/QC42
ago3t2/t3
20 μM
Pixel value
Distance (μm)

Figure 4.2. Ago3 and Aub are required for Piwi protein stability and nucleus localization

(A) Quantitative Western blot for Piwi, Aub, Ago3, and α-Tubulin in w^{1}, $zuc^{HM27/Df}$, $zuc^{HM27/SG63}$, $aub^{HN2/QC42}$, $ago3^{t2/t3}$ single-mutants, and $aub^{HN2/QC42}$; $ago3^{t2/t3}$ double-mutant.

(B) Immuno-detection of Piwi protein in control and mutant ovaries. Images for $zuc^{HM27/Df}$ were acquired at a higher gain. White line marks the region quantified. DAPI, 4′,6-diamidino-2-phenylindole.

accumulated outside of the nucleus, a pattern observed previously when the Piwi nuclear localization signal was deleted (Figure 4.2B; Klenov et al., 2011; Sienski et al., 2012). In *ago3*$^{t2/t3}$ ovaries, less Piwi accumulated in the nucleus and Piwi abundance was halved (Figure 4.2A and 4.2B and Li et al., 2009), as observed in ovaries defective for *vasa* or *spindle-E*, proteins also required for Ping-Pong (Malone et al., 2009). Piwi abundance and localization in both the germline and the soma were normal in *aub*$^{HN2/QC42}$ mutants.

Mutations that decrease germline Piwi stability and nuclear localization similarly reduce the abundance of Piwi-bound piRNAs: median antisense piRNA abundance by transposon family was 21% of *w*^{1} in *ago3*$^{t2/t3}$, 22% in *vas*$^{D5/PH165}$, 54% in *aub*$^{HN2/QC42}$, and 2% in *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ double mutants (Figure 4.1A). We conclude that the Ping-Pong pathway is required to generate sufficient primary piRNAs to license cytoplasmic Piwi to transit to the nucleus in the germline.

Ago3 and Aub Can Each Trigger Piwi-bound Primary piRNA Production

Piwi-bound piRNAs are more plentiful in *ago3* and *aub* single-mutants compared to double-mutants, suggesting that Ago3 and Aub can each generate some Piwi-bound piRNAs. To test directly whether Ago3 can generate Piwi-bound piRNAs in the absence of Aub, we analyzed the 5′-to-5′ distance on the same genomic strand between Piwi-bound piRNAs and the sites of RNA cleavage catalyzed by Ago3 in *aub*$^{HN2/QC42}$, an analysis that evaluates piRNA phasing (Han et al., 2015a). The presence of Piwi-bound piRNAs beginning ~27 nt after the site of Ago3-catalyzed cleavage of transposon RNA indicates that even without Aub, Ago3 can initiate the production of some Piwi-bound piRNAs (Figure 4.3A).

Figure 4.3

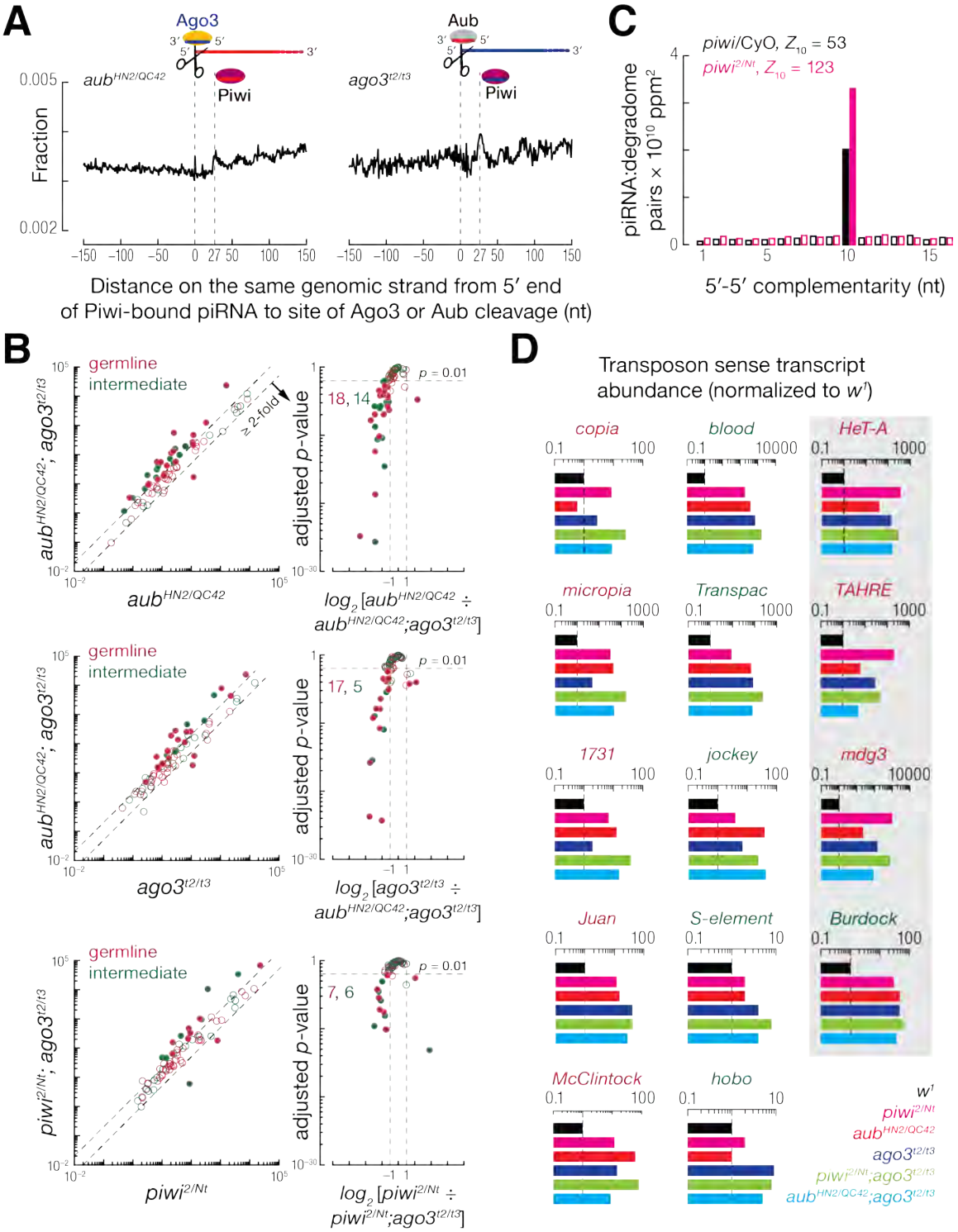

A
Ago3
Aub
Piwi
Fraction
Distance on the same genomic strand from 5′ end of Piwi-bound piRNA to site of Ago3 or Aub cleavage (nt)
C
piwi/CyO, Z10 = 53
piwi2/Nt, Z10 = 123
piRNA:degradome pairs × 10^10 ppm^2
5′-5′ complementarity (nt)
B
germline
intermediate
≥ 2-fold
p = 0.01
18, 14
17, 5
7, 6
adjusted p-value
D
Transposon sense transcript abundance (normalized to w1)
copia
blood
HeT-A
micropia
Transpac
TAHRE
1731
jockey
mdg3
Juan
S-element
Burdock
McClintock
hobo
w1
piwi2/Nt
aub HN2/QC42
ago3 t2/t3
piwi2/Nt;ago3 t2/t3
aub HN2/QC42;ago3 t2/t3

Figure 4.3. Ago3 and Aub Can Each Produce Piwi-bound piRNAs and Silence Transposons

(A) Distance on the same genomic strand from the cleavage sites of Ago3 or Aub to the 5′ ends of Piwi-bound piRNAs.

(B) Scatter and volcano plots comparing transposon sense transcript abundance. Adjusted *p*-value of 0.01 and a change greater than twofold (dotted lines) were used as the cutoff for differential analysis. Differentially expressed transposon families are represented as solid circles. Values are the mean of two biological replicates.

(C) 5′-to-5′ complementarity between transposon-derived piRNAs and degradome reads in $piwi^{+/-}$ (mixed $piwi^{Nt}$/*CyO* and $piwi^{2}$/*CyO*) and $piwi^{2/Nt}$. piRNAs were normalized to non-transposon-derived siRNAs, including cis-natural antisense transcripts and structured loci. Degradome reads were normalized to the total number of transposon-derived reads.

(D) Bar plots of the change in transposon sense transcript abundance compared to w^{1}. The names of germline and intermediate transposons are highlighted in red and green. Transposons with no significant increase in their mRNA levels in $piwi^{2/Nt}$; $ago3^{t2/t3}$ compared to $piwi^{2/Nt}$ are highlighted in grey shadow. Data are mean values of two biological replicates.

Supporting this idea, the expression of 18 germline and 14 intermediate transposon families further increased when Ago3 was removed from *aub* null ovaries (Figure 4.3B). In wild-type ovaries, Ago3 is thought to rely on Aub to generate its secondary piRNA guides. Our data suggest that in the absence of Aub, Ago3 gains access to maternally inherited piRNAs or primary piRNAs produced *de novo*. We presume that Aub, perhaps assisted by Krimper and other proteins, normally prevents primary piRNAs from binding Ago3 (Sato et al., 2015).

In *ago3*$^{t2/t3}$ ovaries, Piwi-bound piRNA 5′ ends also peaked ~27 nt downstream of the sites of Aub-catalyzed cleavage of sense and antisense transposon RNA (Figure 4.3A). Among the 72 transposon families examined, the expression of 17 germline and 5 intermediate transposons further increased when Aub was removed from *ago3*$^{t2/t3}$ ovaries (Figure 4.3B). We conclude that Ago3 and Aub can each independently initiate the production of Piwi-bound piRNAs.

The Major Function of Ago3 is to Generate Piwi-Bound Primary piRNAs

The Ping-Pong model proposes that Ago3 generates antisense piRNAs bound to Aub, which can then cleave cytoplasmic, sense transposon transcripts (Brennecke et al., 2007; Gunawardane et al., 2007b). Because Ago3 also initiates production of Piwi-bound piRNAs, we sought to measure the relative contribution of Ago3 to post-transcriptional transposon silencing by Aub and transcriptional silencing by Piwi.

In *piwi*$^{2/Nt}$ ovaries, piRNA-loaded Piwi is trapped in the cytoplasm (Klenov et al., 2011). *piwi*$^{2/Nt}$ blocks transcriptional silencing of transposons, but cleavage of transposon RNAs by Aub is predicted to be unaffected. To measure RNA

cleavage by Aub in vivo, we sequenced long (>200 nt) RNAs bearing 5′ monophosphate termini, the end structure left by Argonaute-mediated cleavage, and paired these “degradome” sequences with complementary piRNAs to identify 3′ cleavage products. To qualify as a 3′ cleavage product, an RNA was required to have 15 nucleotides of 5′ genomic complementarity with a guide piRNA (g2-g16; Wang et al., 2014) and to have a 5′ end corresponding to the position of cleavage by an Argonaute protein. The Z_{10} score measures the significance of such a match relative to the background of other patterns of piRNA:target RNA pairing. Cleavage of transposon RNA by Aub continues unabated in *piwi*$^{2/Nt}$ ovaries: in fact, the number of piRNA:degradome pairs with a significant Z_{10} score was >50% greater in *piwi*$^{2/Nt}$ ovaries than in *piwi* heterozygotes (Figure 4.3C).

Next, we examined transposon silencing in *piwi*$^{2/Nt}$ ovaries lacking Ago3. If the main role of Ago3 is to generate secondary piRNAs that direct Aub to cleave transposon RNA, transposon expression should increase in *ago3*$^{t2/t3}$; *piwi*$^{2/Nt}$ double-mutant ovaries compared to *piwi*$^{2/Nt}$ alone. Conversely, if Ago3 mainly feeds primary piRNAs to Piwi, then loss of Ago3 from *piwi*$^{2/Nt}$ should have little impact on transposon silencing. Although the overall abundance of antisense piRNAs in *piwi*$^{2/Nt}$*; ago3*$^{t2/t3}$ ovaries was 41% of *piwi*$^{2/Nt}$ (Figure 4.4A), sense transcript abundance increased significantly for only 7 of the 47 germline-specific and 6 of the 25 intermediate transposon families in *piwi*$^{2/Nt}$; *ago3*$^{t2/t3}$ versus *piwi*$^{2/Nt}$ (q-value ≤ 0.01 and ≥ 2-fold increased abundance; n = 2). In contrast, the expression of 18 germline-specific and 14 intermediate transposon families was greater in *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ than *aub*$^{HN2/QC42}$ ovaries, which lack both Aub and

Ping-Pong amplification but produce some Piwi-bound antisense piRNAs (Figure 4.3B). We conclude that Ago3 functions mainly to generate Piwi-bound piRNAs.

Sense transcript abundance for 43 of the 72 transposon families increased significantly in *ago3*$^{t2/t3}$ compared to *w*^{1} (Figure 4.4B). For nearly three-quarters of these families, the sole function of Ago3 appears to be production of piRNAs for Piwi: 32 of the 43 showed no further increase when Piwi was removed from *ago3* mutant ovaries.

Thirteen transposon families were desilenced more in *piwi*$^{2/Nt}$; *ago3*$^{t2/t3}$ than in *piwi*$^{2/Nt}$, likely because for these, Ago3 generates piRNAs for Aub via the Ping-Pong cycle. Among the 13—*copia*, *Circe*, *297*, and *invader2*—were expressed more in *piwi*$^{2/Nt}$ than in *aub*$^{HN2/QC42}$ or *ago3*$^{t2/t3}$ mutant ovaries. For these 4 families, RNA cleavage by Aub likely augments transcriptional silencing by Piwi. For the remaining nine transposon families, particularly *Juan*, *Transpac*, and *blood*, transcriptional repression by Piwi and RNA cleavage by Aub appear to be coequal (Figure 4.3D).

Thus, for a majority of germline transposon families, Ago3-bound piRNAs function mainly to generate Piwi-bound antisense piRNAs that repress transposon transcription.

Figure 4.4

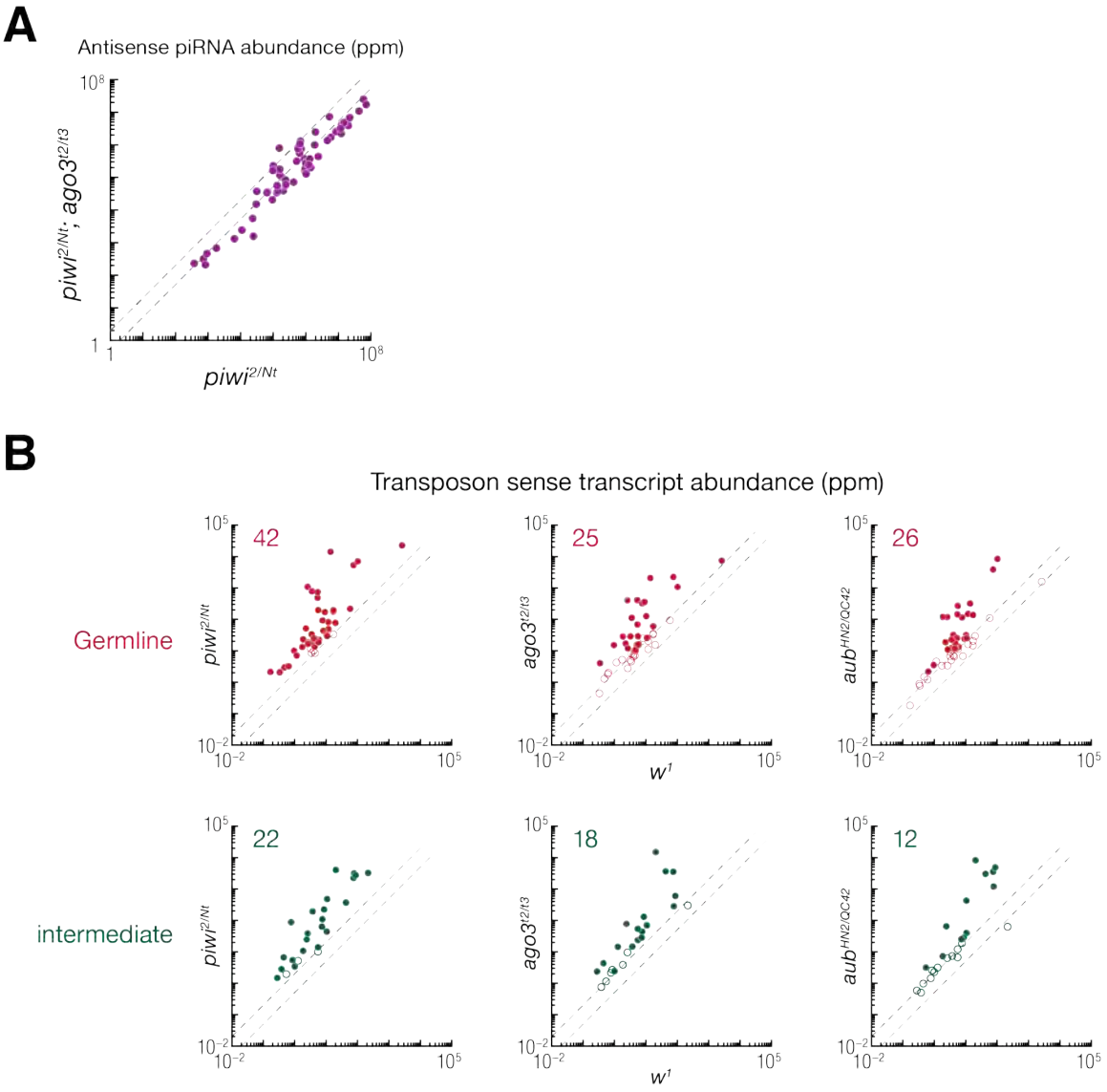

A
Antisense piRNA abundance (ppm)
piwi2/Nt; ago32/3
piwi2/Nt
B
Transposon sense transcript abundance (ppm)
Germline
intermediate
42
25
26
22
18
12
piwi2/Nt
ago32/3
aubHN2/QC42
w1

Figure 4.4. The major function of Ago3 is to generate most Piwi-bound piRNAs

(A) Scatter plot comparing antisense piRNA abundance for germline-specific and intermediate transposon families, between *piwi*$^{2/Nt}$ and *piwi*$^{2/Nt}$; *ago3*$^{t2/t3}$. piRNAs were normalized to non-transposon-derived siRNAs. Values are from a single biological sample.

(B) Scatter plots comparing transposon sense transcript abundance between *w*^{1} and *piwi*$^{2/Nt}$, *ago3*$^{t2/t3}$, *aub*$^{HN2/QC42}$ single-mutants. Adjusted *p*-value of 0.01 and change greater than twofold were used as cutoff for differential analysis. Values are the mean of two biological replicates.

Qin Blocks Conversion of Aub Cleavage Products into Piwi-bound piRNAs

Ago3 and Aub can each initiate the production of Piwi-bound piRNAs, yet Ago3 predominates in wild-type ovaries. Why are the cleavage products of Ago3 favored as substrates for making Piwi-bound piRNAs?

The germline Tudor protein Qin enforces heterotypic Ago3:Aub Ping-Pong, ensuring that Piwi-bound piRNAs are mainly antisense; without Qin, futile Aub:Aub Ping-Pong predominates and transposon silencing fails (Anand and Kai, 2011; Zhang et al., 2011; Zhang et al., 2014). In theory, Qin might promote heterotypic Ago3:Aub Ping-Pong or suppress homotypic Aub:Aub Ping-Pong (Zhang et al., 2011). To differentiate between these mechanisms, we sequenced Aub-bound piRNAs from *ago3*$^{t2/t3}$ mutant and *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ double-mutant ovaries. Without Qin, the abundance of homotypic Aub:Aub Ping-Pong piRNA pairs in *ago3*$^{t2/t3}$ ovaries increased ~200-fold (Figures 4.5A and 4.5B), consistent with Qin acting to suppress Aub:Aub Ping-Pong. In the absence of Ago3, loss of Qin increased the abundance of Piwi-bound sense and antisense piRNAs >3-fold (Figures 4.5C), but failed to restore the wild-type antisense bias of Piwi-bound piRNAs: *ago3*$^{t2/t3}$ and *ago3*$^{t2/t3}$, *qin*$^{1/Df}$ were similarly biased toward sense transposon piRNAs (Figure 4.6A). Thus, Qin normally prevents Aub cleavage products from becoming the precursors of Piwi-bound piRNAs. We detected no other role for Qin: sense transposon RNA abundance

Figure 4.5

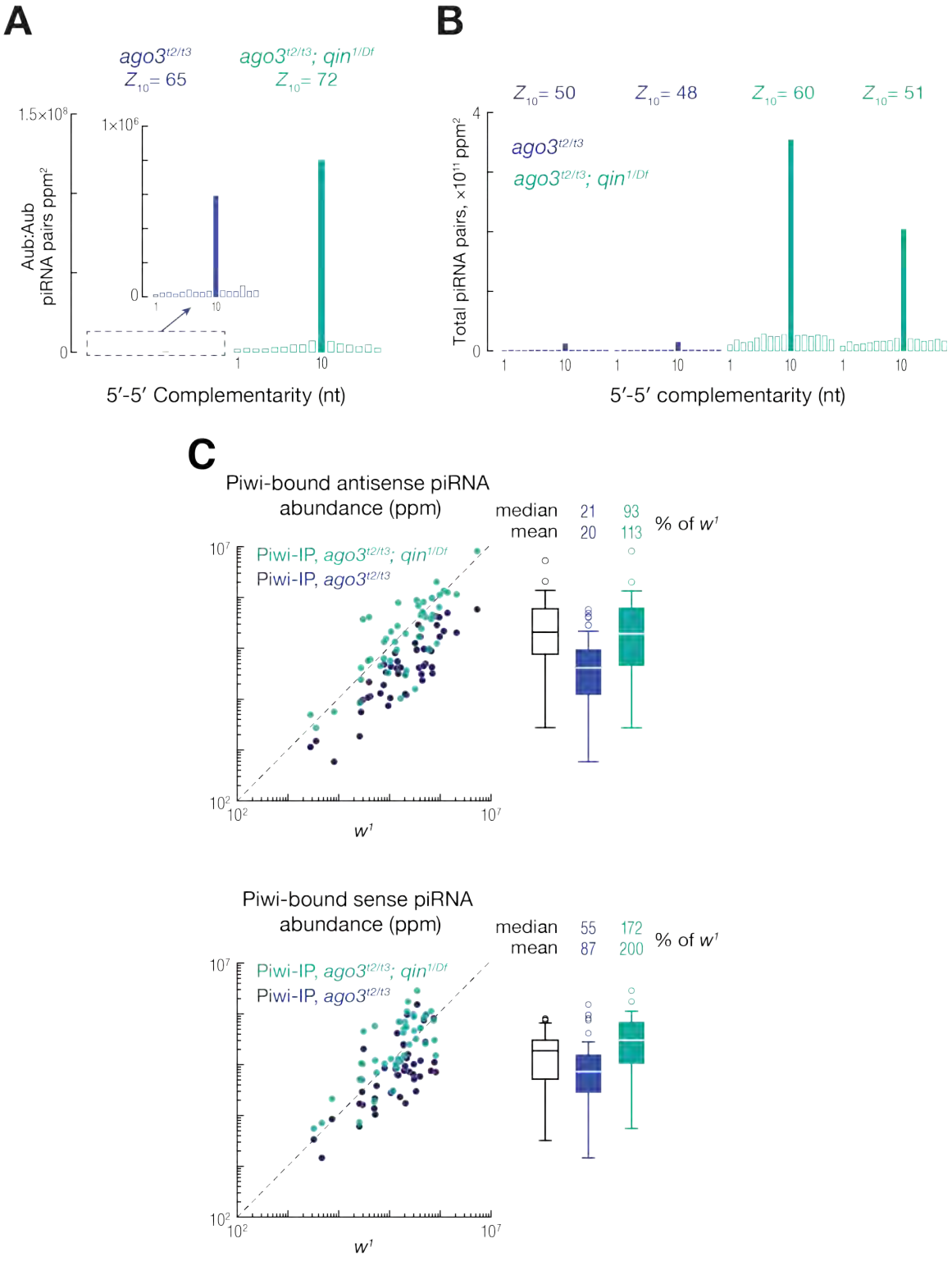

A
ago3t2/t3
Z10= 65
ago3t2/t3; qin1/Df
Z10= 72
1.5×10^8
1×10^6
0
Aub:Aub piRNA pairs ppm2
1
10
5′-5′ Complementarity (nt)
B
Z10= 50
Z10= 48
Z10= 60
Z10= 51
ago3t2/t3
ago3t2/t3; qin1/Df
Total piRNA pairs, ×10^11 ppm2
4
0
5′-5′ complementarity (nt)
C
Piwi-bound antisense piRNA abundance (ppm)
Piwi-IP, ago3t2/t3; qin1/Df
Piwi-IP, ago3t2/t3
w1
median 21 93
mean 20 113
% of w1
Piwi-bound sense piRNA abundance (ppm)
median 55 172
mean 87 200
% of w1

Figure 4.5. Qin Blocks Conversion of Aub Cleavage Products into Piwi-bound piRNAs

(A) Ping-Pong analysis for Aub-bound piRNAs in *ago3*$^{t2/t3}$ and *ago3*$^{t2/t3}$*; qin*$^{1/Df}$. piRNAs were normalized to total genome-mapping reads.

(B) Ping-Pong analysis of total piRNAs in *ago3*$^{t2/t3}$ single-mutant and *ago3*$^{t2/t3}$*; qin*$^{1/Df}$ double-mutant ovaries. piRNAs were normalized to *flamenco*-derived piRNAs. Two biological replicates for each genotype are shown separately.

(C) Scatter and box plots on a common *y*-axis comparing Piwi-bound antisense and sense piRNA abundance for 47 germline-specific transposon families in *w*^{1} control, *ago3*$^{t2/t3}$, and *ago3*$^{t2/t3}$, *qin*$^{1/Df}$ ovaries.

was essentially the same in *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$ double- and *aub*$^{HN2/QC42}$; *ago3*$^{t2/t3}$, *qin*$^{1/Df}$ triple-mutants (Figure 4.6D). Finally, Piwi protein abundance, which in *ago3*$^{t2/t3}$ was half that of the *w*^{1} control, was restored to wild-type levels in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ (110% ± 20% of *w*^{1}; Figure 4.6B), suggesting that supplying more piRNAs to Piwi stabilizes the protein.

Although removing Qin from *ago3*$^{t2/t3}$ mutant ovaries restored the abundance of both Piwi and Piwi-bound piRNAs, it failed to rescue transposon silencing. For all 47 germline-specific transposon families we examined, removing Qin did little to fix the silencing defect caused by loss of Ago3. The sense transcript abundance of just six families—*mdg3*, *Juan*, *gpysy12*, *GATE*, *3S18*, and *Doc3*—declined significantly in *ago3*$^{t2/t3}$ when Qin was removed from *ago3*$^{t2/t3}$ (Figure 4.6C). To test whether transcriptional silencing was improved in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$, we used GRO-seq to measure nascent sense transposon RNA in nuclei from *ago3*$^{t2/t3}$ and *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ ovaries. Transcription of 20 germline and nine intermediate transposon families more than doubled in *ago3*$^{t2/t3}$ compared to *w*^{1} (Figure 4.1B). Of the 20, transcription of just four germline transposons—*HMS-Beagle*, *3S18*, *GATE*, and *gypsy12*—decreased more than twofold in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ compared to *ago3*$^{t2/t3}$ ovaries (Figure 4.6E), and only *3S18* and *GATE* were rescued to the level of the *w*^{1} control. The reduced transcript levels of *Juan*, *mdg3*, and

Figure 4.6

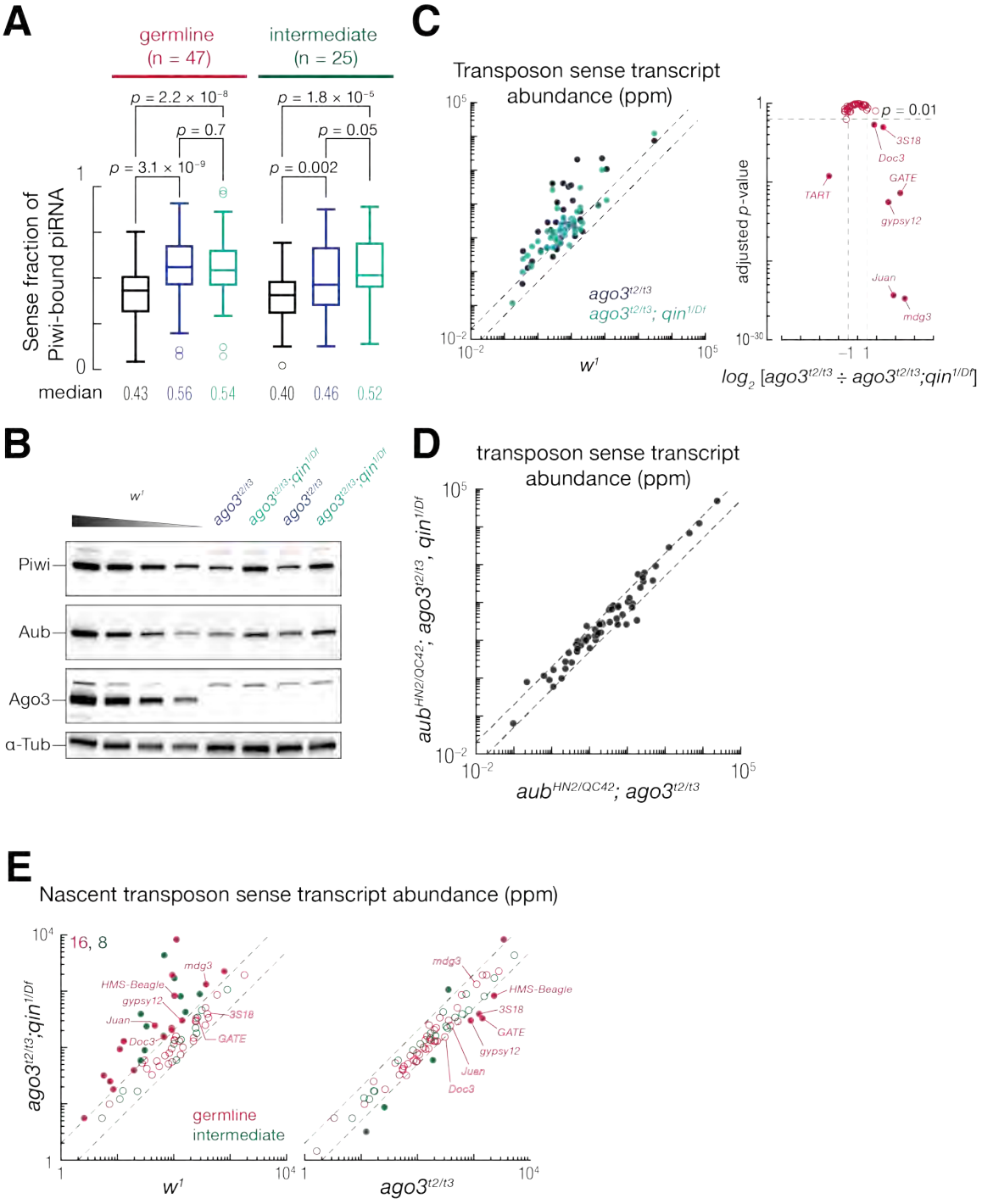

A
germline (n = 47)
intermediate (n = 25)
p = 2.2 × 10⁻⁸
p = 0.7
p = 3.1 × 10⁻⁹
p = 1.8 × 10⁻⁶
p = 0.05
p = 0.002
Sense fraction of Piwi-bound piRNA
median 0.43 0.56 0.54 0.40 0.46 0.52
B
w¹
ago3t2/t3
ago3t2/t3;qin1/Df
Piwi
Aub
Ago3
α-Tub
C
Transposon sense transcript abundance (ppm)
ago3t2/t3
ago3t2/t3; qin1/Df
w¹
adjusted p-value
p = 0.01
3S18
Doc3
GATE
TART
gypsy12
Juan
mdg3
log2 [ago3t2/t3 ÷ ago3t2/t3;qin1/Df]
D
transposon sense transcript abundance (ppm)
aubHN2/QC42; ago3t2/t3; qin1/Df
aubHN2/QC42; ago3t2/t3
E
Nascent transposon sense transcript abundance (ppm)
16, 8
mdg3
HMS-Beagle
gypsy12
Juan
Doc3
3S18
GATE
germline
intermediate
ago3t2/t3;qin1/Df
w¹
ago3t2/t3

Figure 4.6. Antisense Piwi-bound piRNAs aboundance does not determine the efficiency of Piwi-mediated transcriptional silencing

(A) Fraction of Piwi-bound sense piRNAs for different transposon families in w^{1}, $ago3^{t2/t3}$, and $ago3^{t2/t3}$; $qin^{1/Df}$. *p*-values were calculated using Wilcoxon signed-rank test.

(B) Western blots for Piwi, Aub, Ago3, and α-Tubulin in w^{1}, $ago3^{t2/t3}$, and $ago3^{t2/t3}$; $qin^{1/Df}$.

(C) Scatter and volcano plots comparing transposon sense transcript abundance as in Figure 4.3B.

(D) Scatter plot comparing transposon sense transcripts in $aub^{HN2/QC42}$; $ago3^{t2/t3}$ double-mutant and $aub^{HN2/QC42}$; $ago3^{t2/t3}$, $qin^{1/Df}$ triple-mutant ovaries. Values are the mean of two biological replicates.

(E) Scatter plot comparing nascent sense transcript abundance for transposons, measured by GRO-seq, in $ago3^{t2/t3}$ and $ago3^{t2/t3}$; $qin^{1/Df}$ ovaries.

Doc3 in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ likely result from their cleavage by Aub loaded with piRNAs produced by homotypic Aub:Aub Ping-Pong in the absence of Qin.

If the primary function of Ago3 is to generate Piwi-bound piRNAs, why do most transposons escape silencing in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ ovaries, despite restoration of Piwi-bound piRNAs to normal levels? Perhaps the steady-state abundance of Piwi-bound, antisense piRNAs alone does not determine the efficiency of Piwi-mediated transcriptional silencing, and other factors such as the ratio of sense to antisense piRNAs (Figure 4.6A) determine the efficiency of silencing.

Ago3 and Aub Endonuclease Mutants Disrupt Transposon RNA Cleavage

Sequencing data suggest that Ago3- or Aub-catalyzed cleavage of a piRNA precursor initiates Zuc-dependent production of Piwi-bound piRNAs (Han et al., 2015a; Mohn et al., 2015). We generated flies bearing either a wild-type (DDH) or mutant (ADH) Ago3 or Aub transgene in an *ago3* or *aub* mutant background (Figure 4.7A). Changing the wild-type catalytic triad of an Argonaute protein from DDH to ADH or DAH is predicted to inactivate its endonuclease activity (Liu et al., 2004; O'Carroll et al., 2007; De Fazio et al., 2011; Reuter et al., 2011; Sienski et al., 2012). The Aub transgene carried the *aub* genomic sequence, including the endogenous *aub* promoter. The large introns in *ago3* made this strategy impractical. Because a *nanos*-GAL4-VP16-driven UASp-*Ago3*^{WT} transgene failed to restore fertility to an *ago3*$^{t2/t3}$ mutant (Li et al., 2009 and data not shown), possibly because of its low expression at stages 2–6 of oogenesis (Rorth, 1998; Dufourt et al., 2014), we used a pair of transgenes with the *aub* and *vasa* promoters to express *ago3* mRNA (Figure 4.7A).

ago3 and *aub* mutant females are infertile (Schupbach and Wieschaus, 1991; Li et al., 2009). The corresponding wild-type transgenic protein partially (*ago3*) or fully (*aub*) restored fertility (Figure 4.7B). The steady-state abundance of transgenic, wild-type Ago3 in *ago3* null ovaries (w^1; $P\{w^{+mC}, vas::Ago3^{WT}\}^{53B2}$ / $P\{w^{+mC}, aub::Ago3^{WT}\}^{53B2}$; $ago3^{t2}$ / $ago3^{t3}$, hereinafter $Ago3^{WT}$) was 15 ± 1% of w^1 (Figure 4.7C), indicating that wild-type expression of Ago3 may require introns, additional regulatory elements, or the heterochromatic context of endogenous *ago3*. Despite its low expression, transgenic $Ago3^{WT}$ suppressed the dorsal appendage defects present in the offspring of *ago3* mutant mothers and increased the number of embryos that hatched from none to 35% (Figure 4.7B). The Aub^{WT} genomic fragment in the *aub* null background (w^1; aub^{HN2} / aub^{QC42}; $P\{w^{+mC}, Aub^{WT}\}^{89E11}$ / +, hereinafter Aub^{WT}) produced 1.4 ± 0.6 times more Aub than w^1. The Aub^{WT} transgene fully rescued dorsal appendage defects and hatching rates.

Neither $Ago3^{ADH}$ nor Aub^{ADH} rescued the phenotypes of the corresponding mutant (Figure 4.7B). $Ago3^{ADH}$ (7 ± 3% of w^1) and Aub^{ADH} (49 ± 4% of w^1) protein levels were expressed at about half that of the corresponding wild-type transgene. Notably, the abundance of endogenous Ago3 and Piwi in Aub^{ADH} were also half that in w^1, suggesting that a diminished piRNA pool (see below) limits the accumulation of PIWI proteins.

$Ago3^{WT}$, but not $Ago3^{ADH}$, rescued the accumulation of Piwi in the nucleus (Figure 4.7D). In $Ago3^{ADH}$, both transgenic $Ago3^{ADH}$ and endogenous Aub localized to abnormal cytoplasmic puncta, similar to those that form when *aub*

Figure 4.7 (A)

A

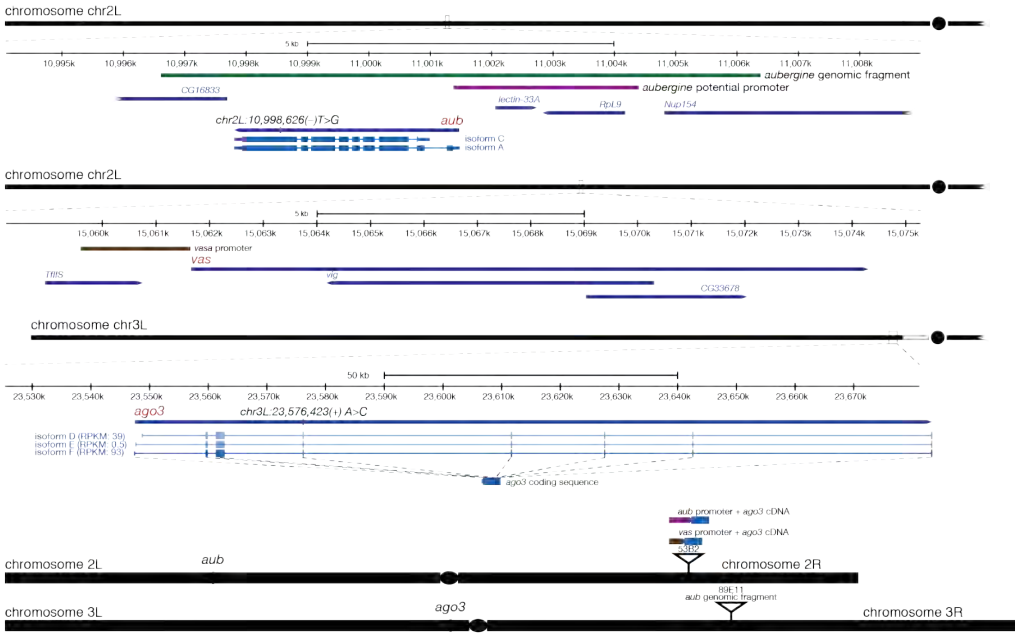
```
                                                                  D→A
sp|Q9UKV8|AGO2_HUMAN      --QTLSNLCLKINVKLGGVNNILLPQGRPPVFQQPVIFLGADVTHPPAGDGKKPSIAAVV   615
sp|O76922|AUB_DROME       LMSIATKVVIQMNAKLMGAPWQV------VIPLHGLMTVGFDVCHSPKNKN--KSYGAFV   652
sp|Q7PLK0|AGO3_DROME      IRSVVQKIVLQMNCKLGGSLWTV------KIPFKNVMICGIDSYHDPSNRG--NSVAAFV   659
sp|Q8CDG1|PIWL2_MOUSE     LRSVAQKILLQMNCKLGGELWGV------DIPLKQLMVIGMDVYHDPSRGM--RSVVGFV   759
sp|Q9JMB7|PIWL1_MOUSE     VMAIATKIALQMNCKMGGELWRV------DMPLKLAMIVGIDCYHDTTAGR--RSIAGFV   649
                                :: :::* *: *    :       :  :  :  * *  *          *  ..*

sp|Q9UKV8|AGO2_HUMAN      GSMDAHP-NRYCATVRVQQHRQEIIQDLAAMVRELLIQFYKSTRFKPTRIIFYRDGVSEG   674
sp|O76922|AUB_DROME       ATMDQKESFRYFSTVNEHIKGQELSEQMSVNMACALRSYQEQHRSLPERILFFRDGVGDG   712
sp|Q7PLK0|AGO3_DROME      ASINSSY-SQWYSKAVVQTKREEIVNGLSASFEIALKMYRKRNGKLPTNIIIYRDGIGDG   718
sp|Q8CDG1|PIWL2_MOUSE     ASINLTL-TKWYSRVVVFQMPHQEIVDSLKLCLVGSLKKYYEVNHCLPEKIVVYRDGVSDG   818
sp|Q9JMB7|PIWL1_MOUSE     ASINEGM-TRWFSRCVFQDRGQELVDGLKVCLQAALRAWSGCNEYMPSRVIVYRDGVGDG   708
                          .:::     :: :    :  :*: : :   .   *  :        * .::.:***:.:*

sp|Q9UKV8|AGO2_HUMAN      QFQQVLHHELLAIRE----ACIKLEKDYQPGITFIVVQKRHHTRLFCTDKNERVGKSGNI   730
sp|O76922|AUB_DROME       QLYQVVNSEVNTLKDRLDEIYKSAGKQEGCRMTFIIVSKRINSRYFTGH--------RNP   764
sp|Q7PLK0|AGO3_DROME      QLYTCLNYEIPQFEM----V-----CGNRIKISYIVVQKRINTRIFSGSGI----HLENP   765
sp|Q8CDG1|PIWL2_MOUSE     QLKTVANYEIPQLQK----CFEAF-DNYHPKMVVFVVQKKISTNLYLAAPD----HFVTP   869
sp|Q9JMB7|PIWL1_MOUSE     QLKTLVNYEVPQFLD----CLKSVGRGYNPRLTVIVVKKRVNARFFAQSGG----RLQNP   760
                          *:    . *:  :                    :  ::*.*:  :. :           .

sp|Q9UKV8|AGO2_HUMAN      PAGTTVDTKITHPTEFDFYLCSHAGIQGTSRPSHYHVLWDDNRFSSDELQILTYQLCHTY   790
sp|O76922|AUB_DROME       VPGTVVDDVITLPERYDFFLVSQAVRIGTVSPTSYNVISDNMGLNADKLQMLSYKMTHMY   824
sp|Q7PLK0|AGO3_DROME      LPGTVVDQHITKSNMYDFFLVSQLVRQGTVTPTHYVVLRDDCNYGPDIIQKLSYKLCFLY   825
sp|Q8CDG1|PIWL2_MOUSE     SPGTVVDHTITSCEWVDFYLLAHHVRQGCGIPTHYICVLNTANLSPDHMQRLTFKLCHMY   929
sp|Q9JMB7|PIWL1_MOUSE     LPGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHYNVIYDSSGLKPDHIQRLTYKLCHVY   820
                            **.:*  :*     **:: ::    *   *: *  : :      * :* *:::: . *

sp|Q9UKV8|AGO2_HUMAN      VRCTRSVSIPAPAYYAHLVAFRARYHLVDKEHDSAEGSHTSGQSNGRDHQALAKAVQVHQ   850
sp|O76922|AUB_DROME       YNYSGTIRVPAVCHYAHKLAFLVAESINRAPSAGL-------------------------   859
sp|Q7PLK0|AGO3_DROME      YNWAGTVRIPACCMYAHKLAYLIGQSIQRDVAEAL-------------------------   860
sp|Q8CDG1|PIWL2_MOUSE     WNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQL-------------------------   964
sp|Q9JMB7|PIWL1_MOUSE     YNWPGVIRVPAPCQYAHKLAFLVGQSIHREPNLSL-------------------------   855
                            .    : :** . *** :*:     :
```



$$w^1; \frac{aub^{QC42}}{CyO}; \frac{+}{+} \times \frac{w^1}{\rightarrow}; \frac{aub^{HN2}}{CyO}; \frac{Aub^{WT}}{TM3,Sb} \rightarrow w^1; \frac{aub^{QC42}}{aub^{HN2}}; \frac{Aub^{WT}}{+}$$

$$w^1; \frac{aub^{QC42}}{CyO}; \frac{+}{+} \times \frac{w^1}{\rightarrow}; \frac{aub^{HN2}}{CyO}; \frac{Aub^{ADH}}{TM3,Sb} \rightarrow w^1; \frac{aub^{QC42}}{aub^{HN2}}; \frac{Aub^{ADH}}{+}$$

$$w^1; \frac{vas::Ago3^{WT}}{CyO}; \frac{ago3^{t2}}{TM3,Sb} \times \frac{w^1}{\rightarrow}; \frac{aub::Ago3^{WT}}{CyO}; \frac{ago3^{t3}}{TM3,Sb} \rightarrow w^1; \frac{vas::Ago3^{WT}}{aub::Ago3^{WT}}; \frac{ago3^{t2}}{ago3^{t3}}$$

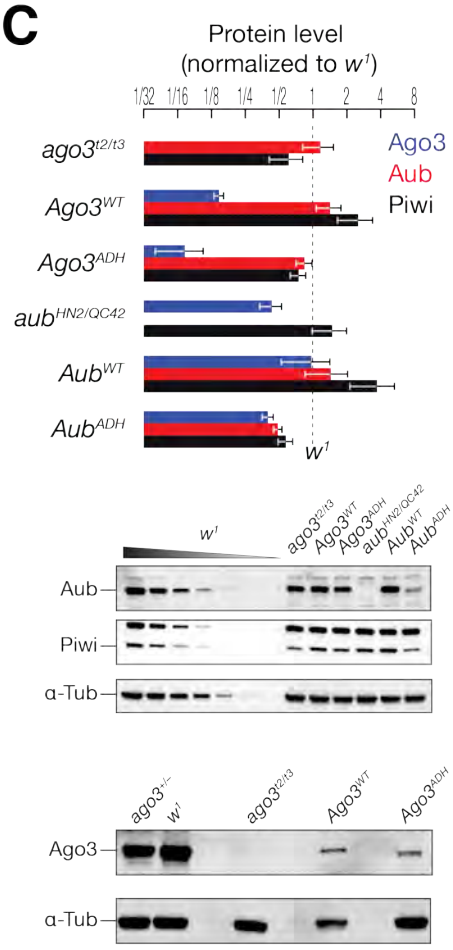
$$w^1; \frac{vas::Ago3^{ADH}}{CyO}; \frac{ago3^{t2}}{TM3,Sb} \times \frac{w^1}{\rightarrow}; \frac{aub::Ago3^{ADH}}{CyO}; \frac{ago3^{t3}}{TM3,Sb} \rightarrow w^1; \frac{vas::Ago3^{ADH}}{aub::Ago3^{ADH}}; \frac{ago3^{t2}}{ago3^{t3}}$$

Figure 4.7 (B-D)

B

Maternal genotype	Eggs per female per day	Dorsal appendages		Hatching rate (%)
		N	Wild type (%)	
ago3$^{+/-}$	33.9	2375	98.4	58.4
ago3$^{t2/t3}$	4.7	397	2.0	0
Ago3^{WT}	19.3	2218	91.4	35.2
Ago3^{ADH}	1.6	98	0	0
aub$^{+/-}$	14.5	2401	99.4	66.9
aub$^{HN2/QC42}$	2.2	159	10.7	0
Aub^{WT}	36.5	3794	99.7	92.6
Aub^{ADH}	0.6	121	0	0

C



D

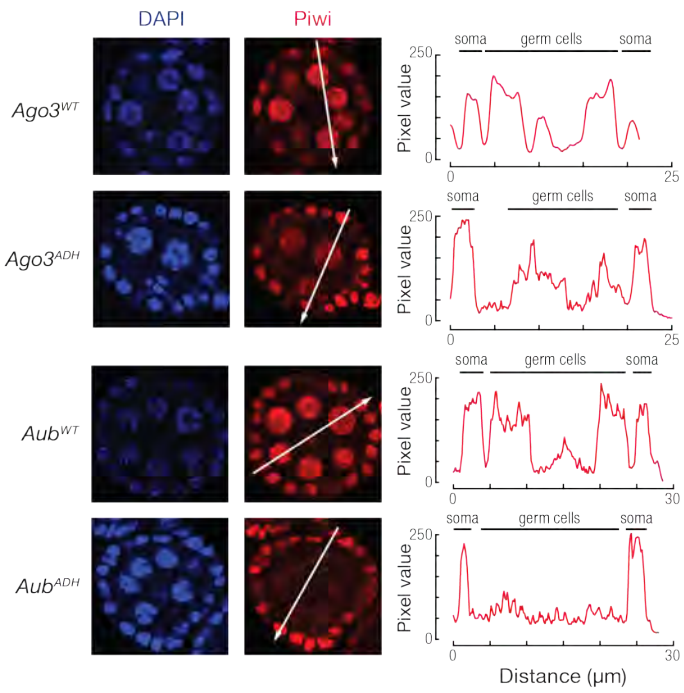


Figure 4.7(E-G)

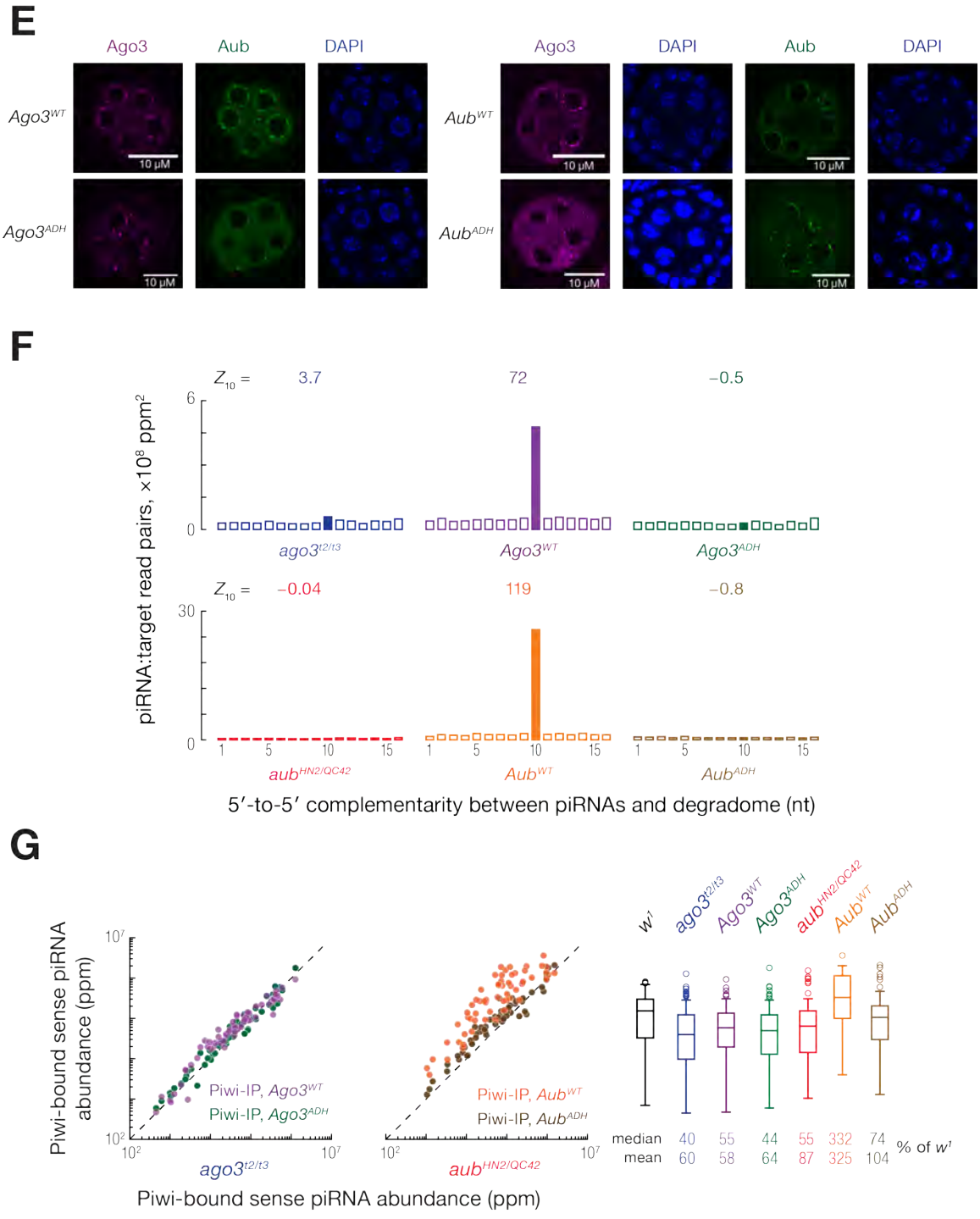

E
Ago3
Aub
DAPI
Ago3WT
Ago3ADH
AubWT
AubADH
10 μM
F
Z10 = 3.7
72
−0.5
piRNA:target read pairs, ×10^8 ppm^2
ago3t2/t3
Ago3WT
Ago3ADH
Z10 = −0.04
119
−0.8
aubHN2/QC42
AubWT
AubADH
5′-to-5′ complementarity between piRNAs and degradome (nt)
G
Piwi-bound sense piRNA abundance (ppm)
Piwi-IP, Ago3WT
Piwi-IP, Ago3ADH
Piwi-IP, AubWT
Piwi-IP, AubADH
w1
ago3t2/t3
Ago3WT
Ago3ADH
aubHN2/QC42
AubWT
AubADH
median 40 55 44 55 332 74 % of w1
mean 60 58 64 87 325 104
Piwi-bound sense piRNA abundance (ppm)

Figure 4.7. Generating catalytically inactive Ago3 and Aub mutants

(A) Top: multiple sequence alignment among human Ago2, fly Aub and Ago3, mouse PIWIL1 (MIWI) and PIWIL2 (MILI). Catalytic triad (DDH) residues are highlighted in red. Bottom: structure of *aub*, *vas*, and *ago3*. The *aub* genomic fragment (highlighted in green) was used to express transgenic Aub^{WT} and Aub^{ADH}. The *vas* (brown) and putative *aub* (purple) promoters were used to drive expression of $Ago3^{WT}$ and $Ago3^{ADH}$ cDNAs.

(B) Female fertility assays.

(C) Quantification of Aub, Ago3, and Piwi protein abundance. Values are mean ± S.D., $n = 3$.

(D) Immunostaining for Piwi in $Ago3^{WT}$, $Ago3^{ADH}$, Aub^{WT}, and Aub^{ADH}. White arrow indicates the region quantified. Due to the high expression of Piwi in Aub^{WT}, a lower gain was used.

(E) Immunofluorescence of Ago3 and Aub in $Ago3^{WT}$, $Ago3^{ADH}$, Aub^{WT}, and Aub^{ADH}.

(F) Distance between the 5′ ends of piRNAs and the 5′ ends of transposon-derived degradome reads, on the opposite strands. piRNAs were normalized to uniquely mapping reads from the *flamenco* cluster. Degradome reads were normalized to all transposon-derived reads. Values are from a single biological sample.

(G) Scatter and box plots on a common *y*-axis comparing the abundance of Piwi-bound, sense piRNAs in $ago3^{t2/t3}$, $Ago3^{WT}$, $Ago3^{ADH}$, $aub^{HN2/QC42}$, Aub^{WT}, and

Aub^{ADH}. piRNAs were normalized to uniquely mapping reads derived from the *flamenco* cluster and presented as ppm. Values are from a single biological sample.

and *shutdown* are depleted by RNAi (Figure S4E; Olivieri et al., 2012). Curiously, the sub-cellular distribution of Aub^{ADH} was normal in *Aub^{ADH}*, but endogenous Ago3 accumulated in abnormal cytoplasmic foci, and nuclear Piwi levels were reduced compared to *aub$^{HN2/QC42}$* ovaries (Figures 4.7D and 4.7E). Ago3 and Piwi protein levels were also reduced in *Aub^{ADH}* (Figure 4.7C).

To test whether the transgenic Ago3^{ADH} and Aub^{ADH} proteins were catalytically inactive in vivo, we used piRNA sequences and degradome-sequencing from mutant and transgenic ovaries to identify putative Ago3 or Aub 3′ cleavage products. *ago3$^{t2/t3}$* ovaries had more piRNA-directed target cleavage than expected by chance, consistent with our finding that Aub can cleave transposon transcripts in the absence of Ago3 (Figure 4.7F). Expressing Ago3^{WT} in *ago3$^{t2/t3}$* increased the abundance of putative Aub and Ago3 cleavage products. As noted for an *Ago3^{AAH}* transgene (Huang et al., 2014), *Ago3^{ADH}* not only failed to increase the abundance of cleaved transposon transcripts, but also prevented Aub cleavage of transposon transcripts and Ping-Pong amplification of piRNAs. *Aub^{WT}*, but not *Aub^{ADH}*, rescued target cleavage in *aub$^{HN2/QC42}$* (Figure 4.7F). We conclude that the ADH mutation disrupts RNA cleavage by Aub and Ago3 in vivo.

Ago3 and Aub Target Cleavage Triggers Production of Phased Primary piRNAs

To test whether RNA cleavage by Ago3 or Aub is required to generate substrates for the production of Piwi-bound piRNAs, we measured the 5′-to-5′ distance between Piwi-bound piRNAs and the 5′ ends of the putative 3′ cleavage products of Ago3 and Aub in *Ago3^{WT}*, *Ago3^{ADH}*, *Aub^{WT}*, and *Aub^{ADH}* (Figure 4.8A). As in *w^{1}*

Figure 4.8

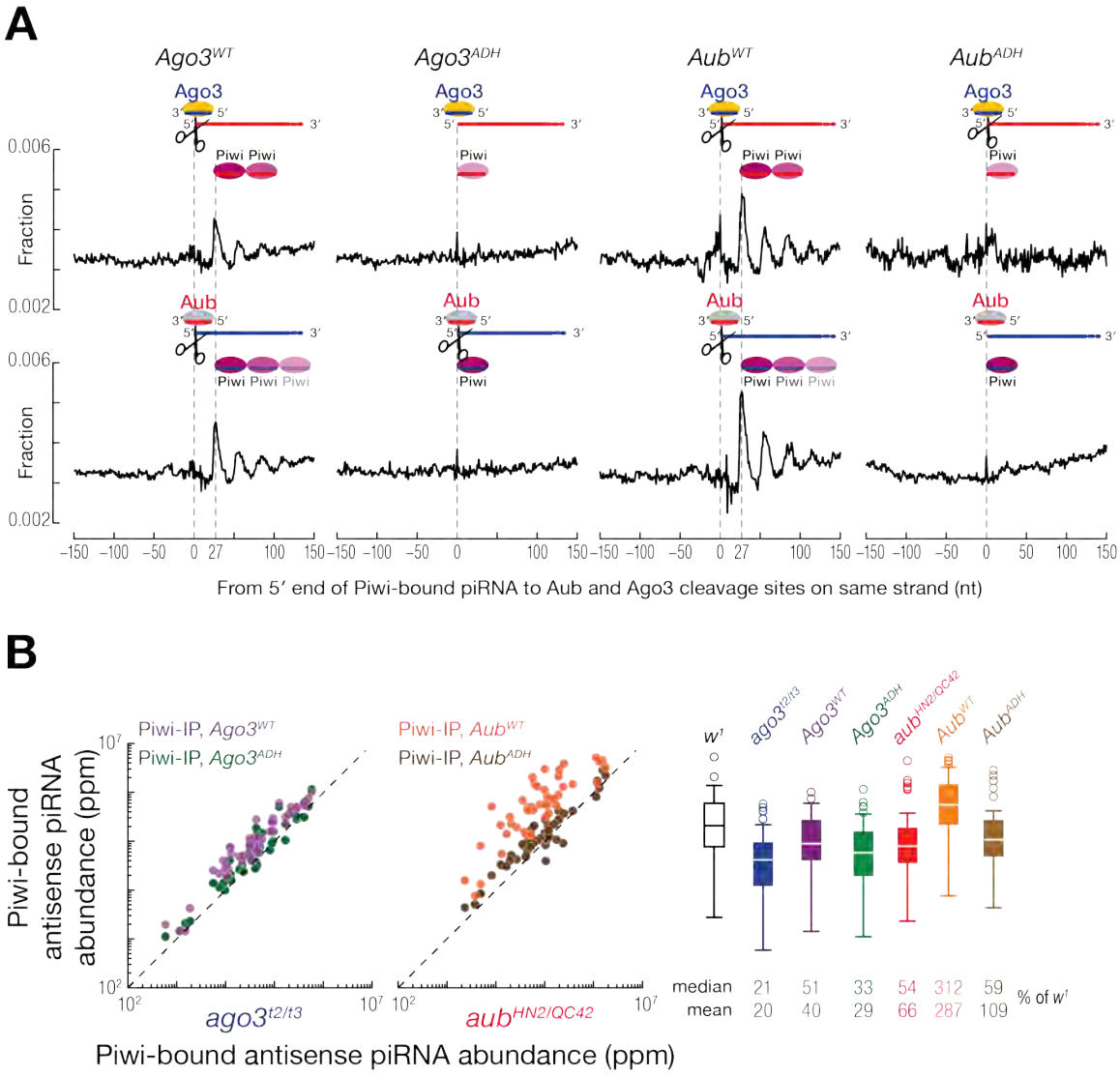

A
Ago3WT
Ago3ADH
AubWT
AubADH
Ago3
Aub
Piwi
Fraction
0.006
0.002
From 5′ end of Piwi-bound piRNA to Aub and Ago3 cleavage sites on same strand (nt)
B
Piwi-IP, Ago3WT
Piwi-IP, Ago3ADH
Piwi-IP, AubWT
Piwi-IP, AubADH
Piwi-bound antisense piRNA abundance (ppm)
ago3t2/t3
aubHN2/QC42
w1
ago3t2/t3
Ago3WT
Ago3ADH
aubHN2/QC42
AubWT
AubADH
median 21 51 33 54 312 59
mean 20 40 29 66 287 109
% of w1
Piwi-bound antisense piRNA abundance (ppm)

Figure 4.8. Ago3 or Aub Cleavage Triggers Phased Piwi-bound piRNA Production.

(A) Distance from the cleavage sites of Ago3 or Aub to the 5′ ends of Piwi-bound, transposon-derived piRNAs on the same genomic strand, in *Ago3*^{WT}, *Ago3*^{ADH}, *Aub*^{WT}, and *Aub*^{ADH}.

(B) Scatter and box plots on a common *y*-axis comparing the abundance of Piwi-bound, antisense piRNAs in *ago3*$^{t2/t3}$, *Ago3*^{WT}, *Ago3*^{ADH}, *aub*$^{HN2/QC42}$, *Aub*^{WT}, and *Aub*^{ADH}. piRNAs were normalized to uniquely mapping piRNAs from the *flamenco* cluster.

control ovaries, the 5′ ends of Piwi-bound piRNAs typically mapped ~27 and ~53 nt downstream of the Ago3 and Aub cleavage sites in *Ago3*^{WT} and *Aub*^{WT} (Han et al., 2015a). We did not detect phasing of Piwi-bound piRNAs in *Ago3*^{ADH} or *Aub*^{ADH} ovaries.

Piwi-bound antisense piRNA abundance increased from 21% of *w*^{1} in *ago3*$^{t2/t3}$ to 51% in *Ago3*^{WT} and from 54% of *w*^{1} in *aub*$^{HN2/QC42}$ to 312% in *Aub*^{WT} (Figure 4.8B). We expected that Piwi-bound piRNA abundance would be low in *Ago3*^{ADH} and *Aub*^{ADH} transgenic mutants, but Piwi-bound antisense piRNA abundance was 33% of the *w*^{1} control in *Ago3*^{ADH} and 59% of *w*^{1} in *Aub*^{ADH}. A similar trend was observed for Piwi-bound sense piRNAs (Figure 4.7G). These data suggest the existence of an alternative pathway that bypasses the need for Ago3 and Aub to cleave their targets in order to produce Piwi-bound piRNAs.

An Alternative Pathway Generates Piwi-bound piRNAs in Ago3^{ADH} and Aub^{ADH}

When Ago3 cleaves an RNA, it generates an Aub-bound secondary piRNA followed immediately by multiple, phased, Piwi-bound piRNAs. Surprisingly, Ago3^{ADH}, which does not support Ping-Pong piRNA amplification, can still initiate production of some Aub-bound piRNAs. In *Ago3*^{ADH}, the 5′-to-5′ distance between transgenic Ago3^{ADH}-bound piRNAs and endogenous Aub-bound piRNAs peaked at 10 and 26 nt, revealing an alternative mechanism for Aub-bound piRNA production (Figure 4.9A, row 3, column 2). The 10-nt peak corresponds to Ago3^{ADH}-bound piRNAs made by cleavage of transposon transcripts by Aub—the first step in normal Ping-Pong amplification and the only heterotypic Ping-Pong step remaining in *Ago3*^{ADH}. The 26-nt peak represents the

production of Aub-bound piRNAs by Ago3^{ADH} through a mechanism that bypasses the need for Ago3-catalyzed cleavage (Figure 4.10A). Analysis of the 3′-to-5′ distance between Ago3^{ADH}- and Aub-bound piRNA suggests that the 3′ end of an Ago3^{ADH}-bound piRNA establishes the 5′ end of the complementary pseudo-secondary Aub-bound piRNA (i.e., a peak ~0; Figure 4.10B). A similar phenomenon occurs in mice expressing only catalytically inactive MIWI: the 5′-to-5′ distance between piRNAs and degradome reads on opposite genomic strands peaks at 29 nt, the size of MIWI-bound piRNAs (Figure 4.10C; Reuter et al., 2011).

Although the alternative Aub-bound, pseudo-secondary piRNAs were readily detected when unique piRNA sequences irrespective of their abundance were analyzed, they were not evident when piRNA reads were analyzed (Figure 5A, blue lines). An explanation for this discrepancy is that the alternative pathway, unlike the Ping-Pong cycle, does not amplify piRNAs, so the most abundant Aub-bound piRNAs in *Ago3*^{ADH} mutant ovaries are generated by homotypic Aub:Aub Ping-Pong, obscuring the production by Ago3^{ADH} of pseudo-secondary, Aub-bound piRNAs. Supporting this view, a broad peak of ~26 nt was readily detected among piRNA reads when piRNA sequences in the top 10% by abundance were removed (data not shown).

In contrast, the production of Ago3^{ADH}-bound piRNAs by Aub, as evidenced by a 10-nt 5′-to-5′ distance between Aub and Ago3 piRNAs, was readily detected by analyzing either piRNA sequences or reads. piRNAs loaded into Ago3^{ADH} cannot direct production of a secondary copy of the original, Aub-bound primary piRNA. Instead, the alternative pathway generates a new Aub-bound piRNA that differs from the original Aub-bound primary piRNA.

Figure 4.9

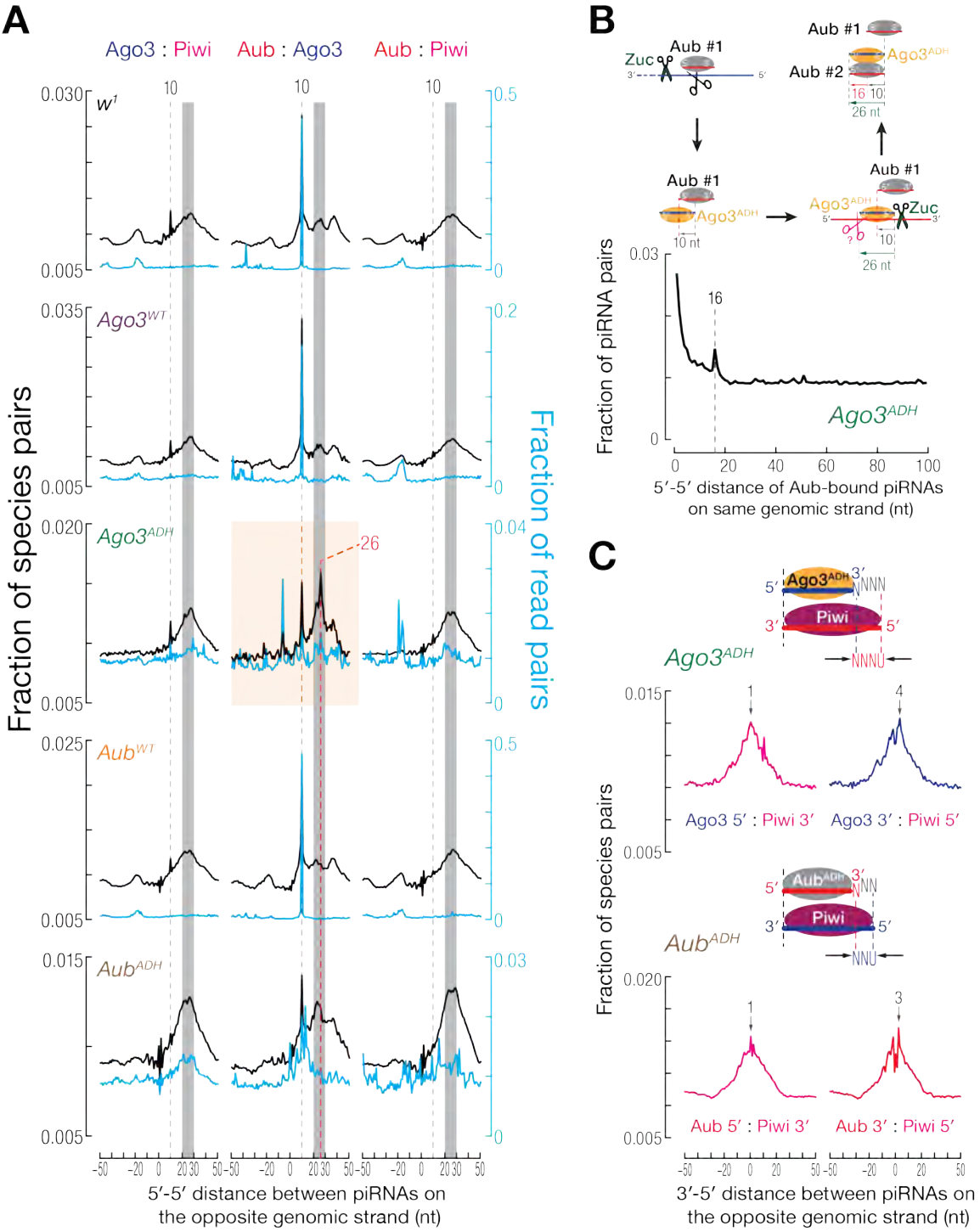

A
Ago3 : Piwi
Aub : Ago3
Aub : Piwi
w1
Ago3WT
Ago3ADH
AubWT
AubADH
10
26
Fraction of species pairs
Fraction of read pairs
5′-5′ distance between piRNAs on the opposite genomic strand (nt)
B
Aub #1
Zuc
Aub #2
Ago3ADH
16 10
26 nt
10 nt
Fraction of piRNA pairs
16
5′-5′ distance of Aub-bound piRNAs on same genomic strand (nt)
C
Piwi
NNNN
NNNU
NNU
Ago3 5′ : Piwi 3′
Ago3 3′ : Piwi 5′
Aub 5′ : Piwi 3′
Aub 3′ : Piwi 5′
3′-5′ distance between piRNAs on the opposite genomic strand (nt)

Figure 4.9. An Alternative Pathway Generates piRNAs when Ago3 is Catalytically Inactive.

(A) 5′-to-5′ distance between transposon-derived piRNAs bound to Piwi, Aub, or Ago3 on opposite genomic strands in different genetic backgrounds. piRNAs in species (black) or reads (blue) are presented on different *y*-axes.

(B) 5′-to-5′ distances between Aub-bound piRNA reads on the same genomic strand in *Ago3*^{ADH}.

(C) 3′-to-5′ distances between Piwi-, Ago3-, or Aub-bound piRNA species on the opposite genomic strand in *Ago3*^{ADH} and *Aub*^{ADH}.

Consequently, the alternative pathway cannot amplify individual piRNA sequences. Instead, this aberrant pathway produces a new Aub-bound piRNA whose 5′ end lies on the same genomic strand, ~16 nt upstream of the Aub-bound piRNA initiating the abortive Ping-Pong cycle (Figure 4.9B). The result is a pseudo-Ping-Pong relationship: the two Aub-bound piRNAs overlap by ten identical rather than ten complementary nucleotides. The alternative pathway appears to use piRNA 3′ ends to define the 5′ ends and piRNA 5′ ends to define the 3′ ends of the pseudo-secondary piRNAs it produces.

Analysis of 5′-to-5′ distance also suggests that the alternative pathway generates Piwi-bound piRNAs (Figure 4.9A, columns 1 and 3). Pairwise analysis of the 3′-to-5′ distances between Aub^{ADH}, $Ago3^{ADH}$, and Piwi-bound piRNAs supports this idea and reinforces the view that the alternative pathway uses piRNA 3′ ends to define the 5′ ends of new piRNAs (Figure 4.9C). However, we cannot exclude the possibility that Piwi-bound piRNAs might guide production of pseudo-secondary piRNAs bound by Aub or Ago3.

Though the alternative pathway was most readily seen in ovaries bearing catalytically inactive Ago3 or Aub, it was detectable in mutants that disrupt Ping-Pong amplification and in cultured somatic OSC cells, which lack the Ping-Pong pathway. The alternative pathway was also weakly detected in w^1 control ovaries (Figure 4.10D), suggesting it functions in wild-type ovaries. The alternative pathway was not detected in *zuc*, *gasz*, or *armitage* mutants, which are all defective for primary piRNA biogenesis. A requirement for these primary piRNA pathway genes is consistent with a role for Zuc in generating piRNA 3′ ends in the alternative pathway.

Figure 4.10 (A-D)

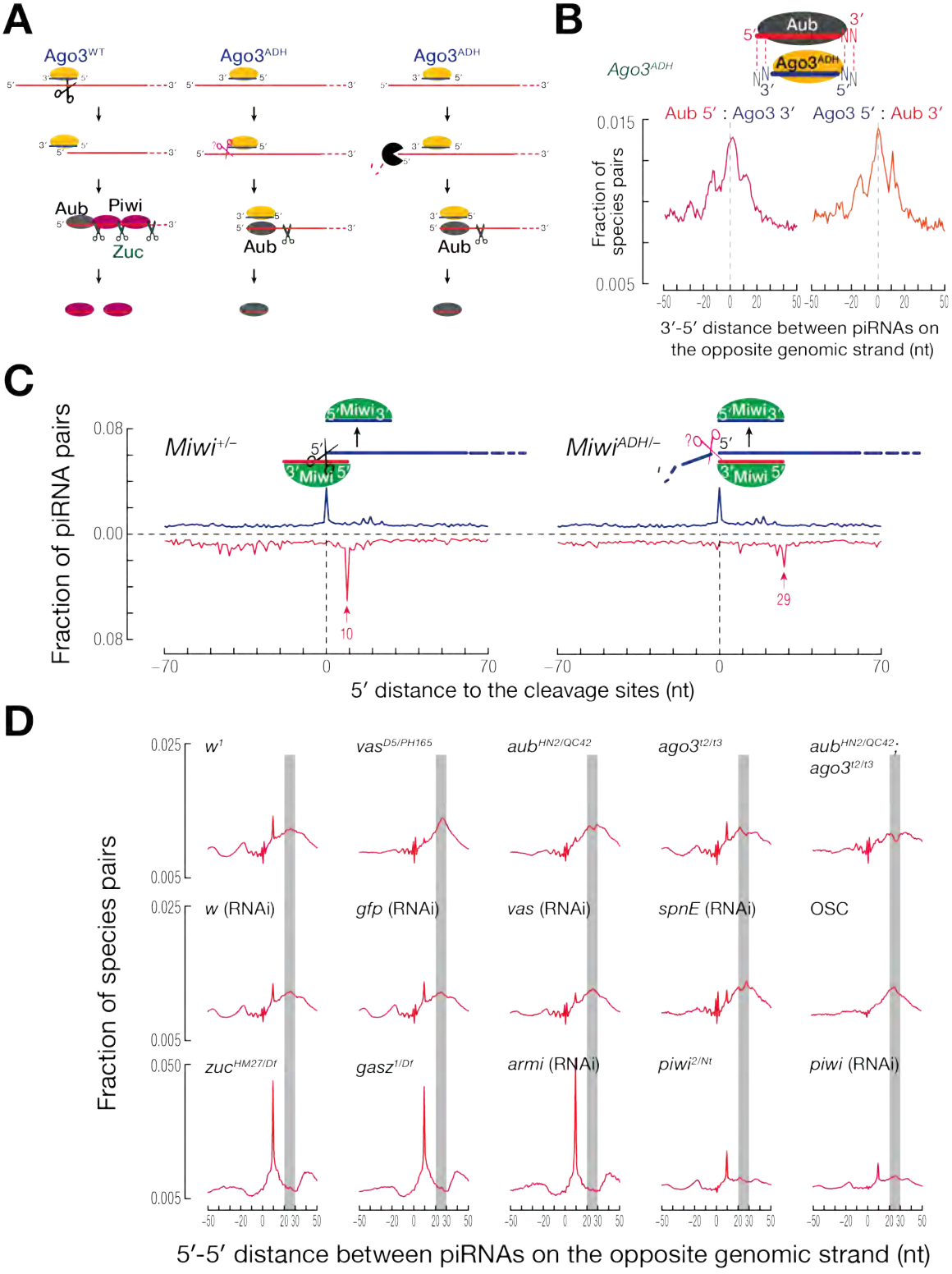


Figure 4.10 (E-F)

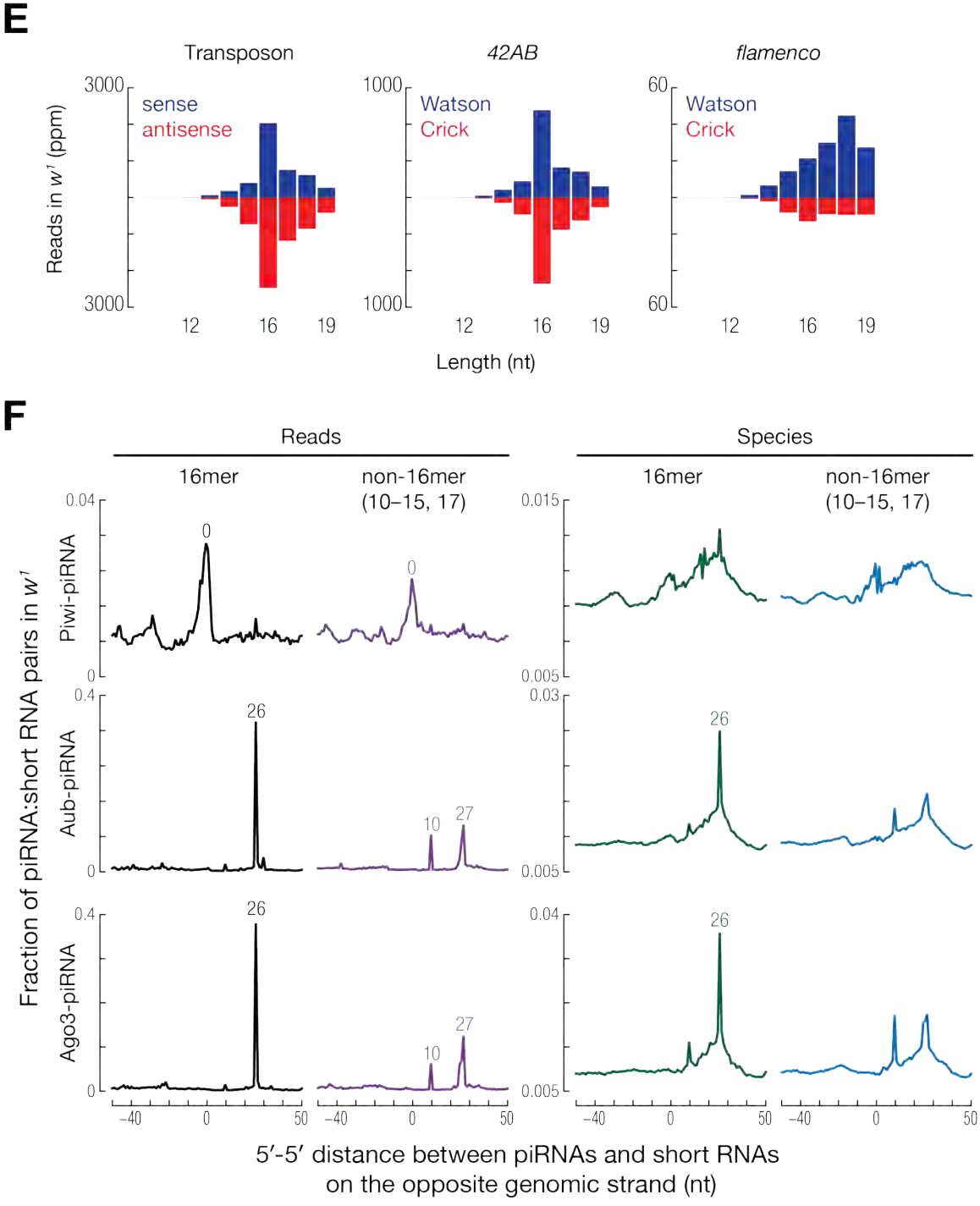


Figure 4.10 The Alternative Pathway is conserved and cleavage byproducts from alternative pathway

(A) A model for the initiation of Aub-bound piRNA production by wild-type and catalytically inactive Ago3. Endonuclease cleavage (middle) or a 5′-to-3′ exonuclease trimming (right) of an Ago3^{ADH}-bound RNA is proposed to generate the 5′ ends of piRNA intermediates, which are then further processed by the nuclease Zuc to produce new piRNAs.

(B) 3′-to-5′ distance analysis between Aub- and Ago3^{ADH}-bound piRNA species on opposite genomic strands in *Ago3^{ADH}* mutant ovaries. Values are from a single biological sample.

(C) 5′-to-5′ distance analysis between MIWI-associated piRNAs and degradome reads on the same genomic strand (blue) or the opposite genomic strand (red) in *Miwi$^{+/–}$* and *Miwi$^{ADH/–}$*.

(D) 5′-to-5′ distance analysis for transposon-derived piRNA species on opposite genomic strands. Values are from a single biological sample.

(E) Length distribution of uniquely mapped short RNAs derived from transposons, *42AB* piRNA cluster, and *flamenco* piRNA cluster. Reads are normalized to total genomic mapping reads. Values are from a single biological sample.

(F) 5′-to-5′ distance analysis between transposon-derived, Piwi-/Aub-/Ago3-bound piRNA and 10–17 nt short RNAs on the opposite genomic strands. Values are from a single biological sample.

The alternative pathway in the *Drosophila* ovary calls to mind the 16-nt piRNA biogenesis byproducts of silkmoth BmAgo3 and the 19-nt byproducts of mouse MIWI (Berninger et al., 2011; Xiol et al., 2012). When BmAgo3 or MIWI cleaves a target RNA, the resulting 3′ fragments become substrates for secondary piRNA production. In contrast, the 5′ cleavage fragments are converted into 16- nt and 19-nt RNAs by a process in which the 3′ end of the BmAgo3- or MIWI-bound piRNA directly defines the 5′ end of the byproduct RNA (Xiol et al., 2012). To examine whether such byproducts also exist in flies, we sequenced 10–17 nt RNA from w^{1} ovaries. The lengths of transposon and *42AB* piRNA cluster-derived short RNAs in this dataset peaked at 16 nt (Figure 4.10E). On the other hand, short RNAs derived from the *flamenco* cluster were rare, suggesting that the 16 nt RNAs are germline specific. We analyzed the 5′-to-5′ distance between Piwi, Aub, and Ago3-bound piRNAs and 10–17 nt short RNAs on the opposite genomic strand (Berninger et al., 2011; Xiol et al., 2012). We observed a 26 nt peak between Aub and Ago3-associated piRNAs and 16 nt byproduct RNAs, but not between piRNAs and short RNAs of other lengths, suggesting that the 3′ ends of the 16 nt RNAs are produced by Ago3 and Aub slicing, while the 5′ ends are generated by another nuclease (Figure 4.10F).

Piwi-Bound piRNAs Made by the Alternative Pathway Direct Transcriptional Silencing

To examine whether Piwi-bound piRNAs made by the alternative pathway can silence transposons, we measured transposon sense transcripts in $ago3^{t2/t3}$, $Ago3^{WT}$, and $Ago3^{ADH}$ ovaries. $Ago3^{ADH}$ supports the alternative piRNA pathway but not normal Ping-Pong. Compared to $ago3^{t2/t3}$, $Ago3^{ADH}$ restored silencing of

10 germline-specific transposons, including *HeT-A*, *TAHRE,* and *mdg3* (Figures 4.11A and 4.11D). These transposons are likely silenced by Piwi-bound piRNAs, since their reduced transcript levels were accompanied by an increase in Piwi-bound antisense piRNAs in $Ago3^{ADH}$ compared with $ago3^{t2/t3}$ (Spearman correlation coefficient, $\rho = -0.8$; p-value = 0.007). In contrast, antisense Aub-piRNA abundance correlated poorly with transposon transcript abundance ($\rho = -0.35$). For the 37 germline transposon families not rescued by $Ago3^{ADH}$, transcript abundance was uncorrelated with either Piwi- ($\rho = -0.07$; Figure 4.11B) or Aub-bound antisense piRNA levels ($\rho = 0.03$). Finally, GRO-seq revealed that nascent sense transcripts of 12 germline and 2 intermediate transposon families decreased more than twofold in $Ago3^{ADH}$ compared to $ago3^{t2/t3}$ (Figure 4.11C). The data support the idea that $Ago3^{ADH}$ initiates the alternative production of Piwi-bound piRNAs that can silence germline transposon transcription.

Figure 4.11 (A-C)

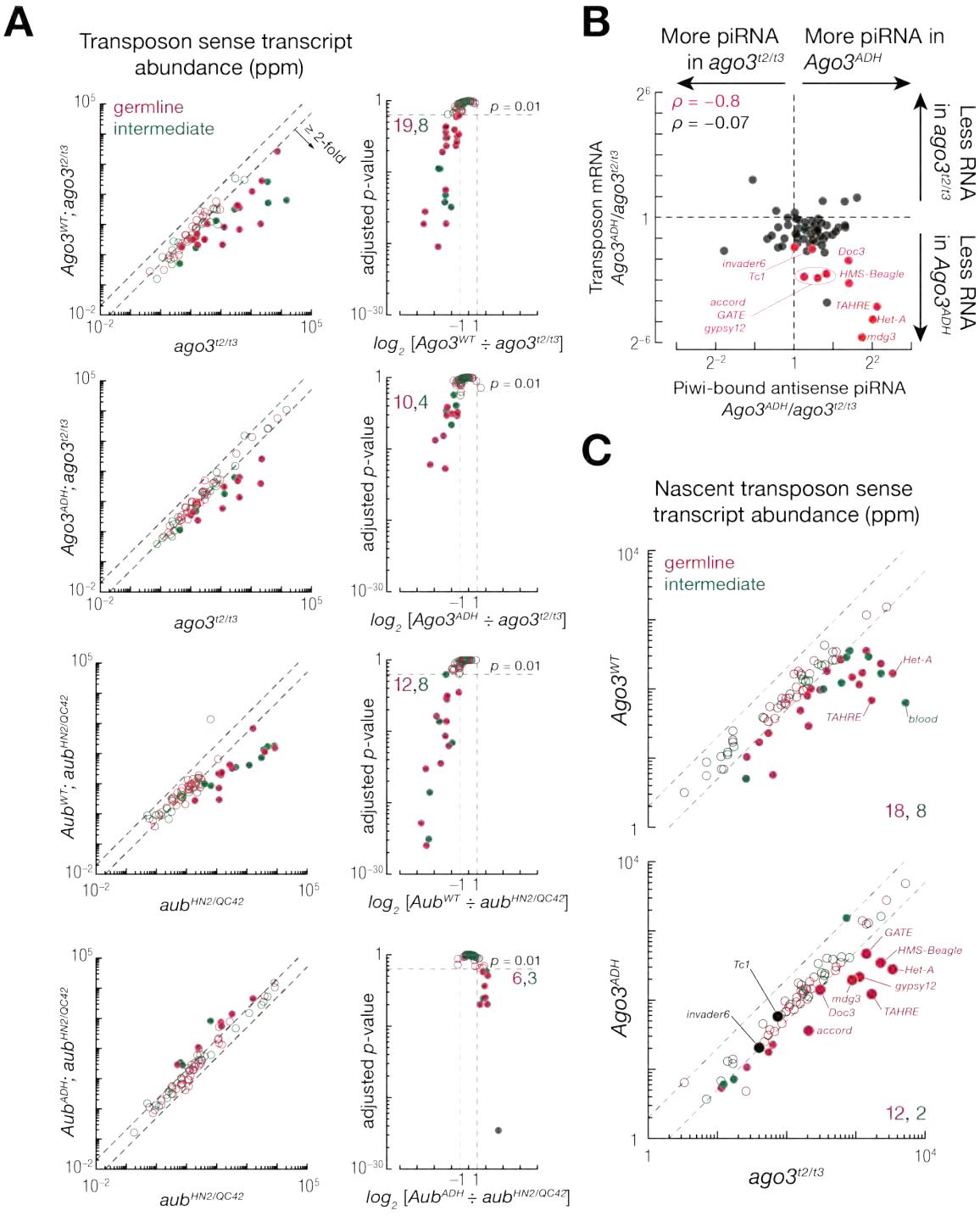


Figure 4.11 (D)

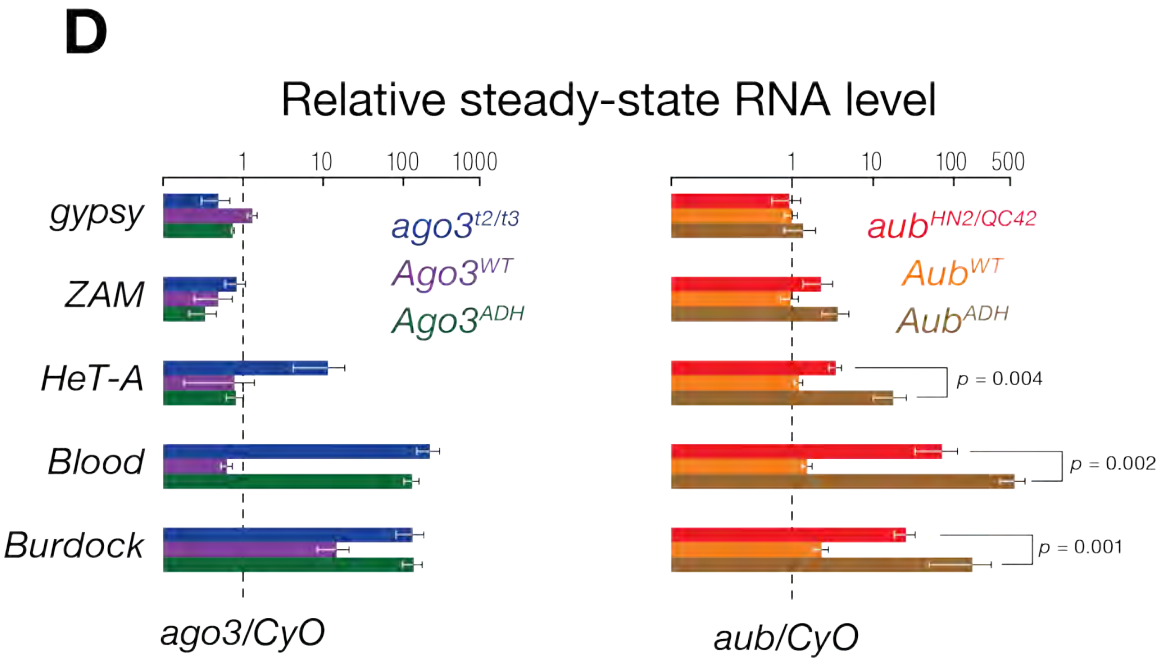

D
Relative steady-state RNA level
1
10
100
1000
gypsy
ZAM
HeT-A
Blood
Burdock
ago3t2/t3
Ago3WT
Ago3ADH
ago3/CyO
1
10
100
500
aubHN2/QC42
AubWT
AubADH
p = 0.004
p = 0.002
p = 0.001
aub/CyO

Figure 4.11 Catalytically Inactive Ago3^{ADH} Silences Transposons by Generating Piwi-bound piRNAs

(A) Scatter and volcano plots comparing transposon sense transcript abundance as in Figure 2B. Values are the mean of two biological replicates.

(B) Scatter plot comparing the change in Piwi-bound antisense piRNAs and transposon sense transcripts in *Ago3^{ADH}* compared to *ago3$^{t2/t3}$* for germline-specific transposon families. Red, transposon silencing restored by Ago3^{ADH}.

(C) Scatter plot comparing nascent, sense transposon transcript abundance in *ago3$^{t2/t3}$*, *Ago3^{WT}*, and *Ago3^{ADH}*.

(D) Quantitative RT-PCR for five transposon families in different genotypes. Values are mean ± S.D., *n* = 3. *p*-values were calculated using Student's *t*-test test.

Discussion

Our data suggest that RNA cleavage catalyzed by Ago3 or Aub initiates the production of most Piwi-bound germline piRNAs. We estimate that as few as 2% of Piwi-bound piRNAs are contributed maternally or made by other mechanisms, and that *de novo* production of Piwi-bound, primary piRNAs plays a minor role in generating guides for Piwi. Perhaps the *de novo* pathway only operates in the soma, which lacks Ago3 and Aub (Siomi et al., 2011; Guzzardo et al., 2013). In this view, Aub-bound, maternally deposited antisense piRNAs must initiate Ping-Pong amplification, leading to the accumulation of Ago3-bound sense piRNAs (Brennecke et al., 2008; Le Thomas et al., 2014b). Ping-Pong amplification of Ago3-bound piRNAs would then serves to ensure a sufficient supply of guides for Piwi.

Ago3 Drives Primary piRNA Production

Piwi-bound piRNAs are believed to recognize nascent transcripts (Le Thomas et al., 2013), explaining why most of these piRNAs are antisense to transposon mRNAs (Brennecke et al., 2007). This antisense bias reflects the initiation of primary piRNA production by cleavage of long antisense RNAs by Ago3-associated, sense piRNAs. Aub, whose guides are largely antisense, plays only an indirect role in primary piRNA biogenesis—by amplifying Ago3-bound sense piRNAs. The Tudor-E3 ligase domain protein, Qin, restricts Ping-Pong to heterotypic Aub:Ago3 cycles by suppressing futile homotypic Aub Ping-Pong, thereby enabling sense piRNAs to accumulate in Ago3 and antisense piRNAs to accumulate in Piwi (Figure 7).

Figure 4.12

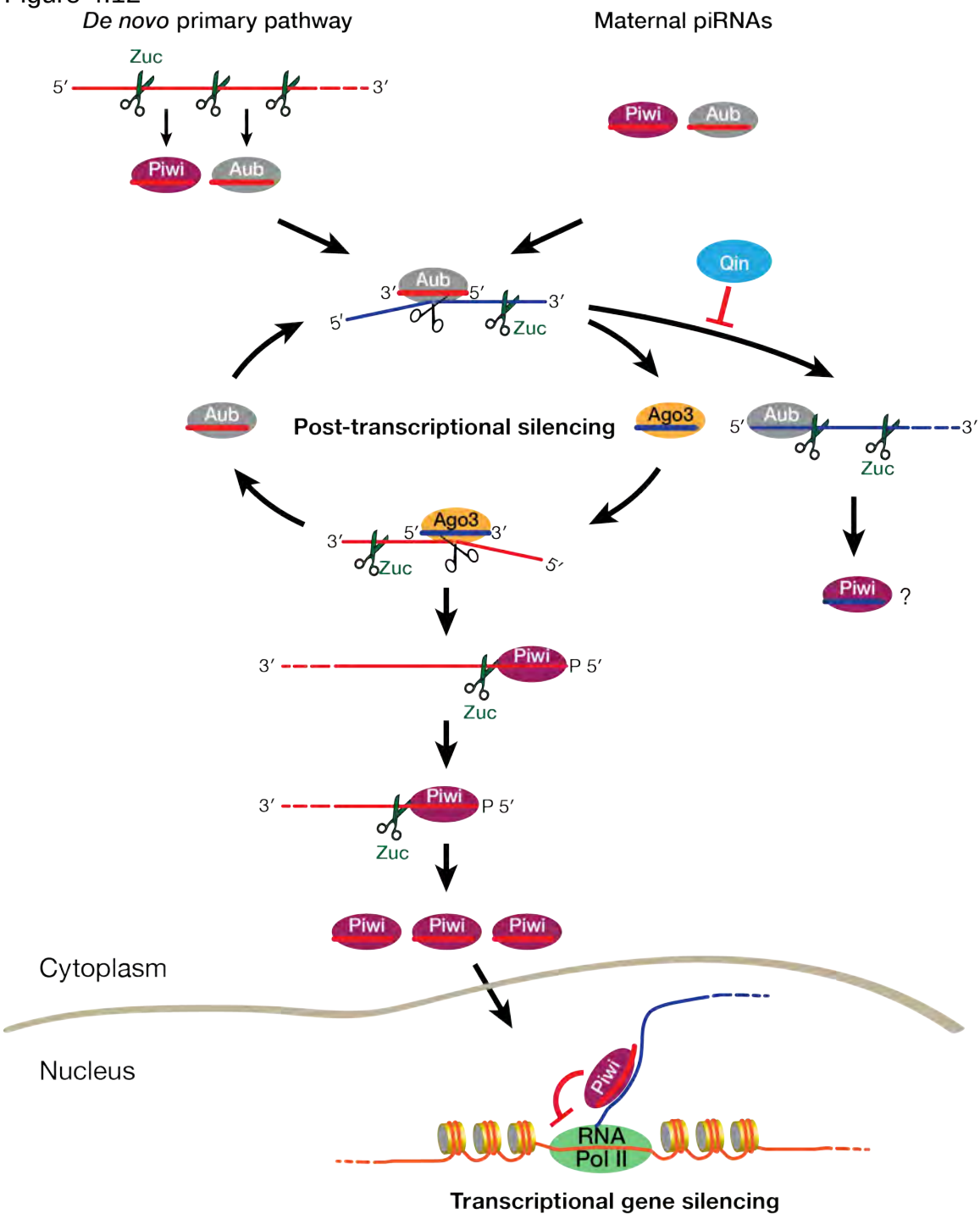

De novo primary pathway
Maternal piRNAs
Zuc
5′
3′
Piwi
Aub
Qin
Post-transcriptional silencing
Ago3
P 5′
?
Cytoplasm
Nucleus
RNA Pol II
Transcriptional gene silencing

Figure 4.12. A Model for piRNA Biogenesis and Function in the Fly Ovarian Germline

Both piRNAs made *de novo* and maternally deposited piRNAs can kick-start in germ cell Ping-Pong. Primary piRNAs, mostly antisense (red), guide Aub to cleave transposon mRNAs, generating the 5′ end of a secondary, sense piRNA (blue), which is loaded into Ago3. Zucchini cuts the Ago3-bound piRNA precursor 27 nt downstream, defining both the 3′ end of the secondary, Ago3-bound piRNA and the 5′ end of next piRNA. Qin prevents the similar cleavage product of made by Aub from binding to a second Aub protein, thereby ensuring Ping-Pong is heterotypic. Sense piRNAs guide Ago3 to bind antisense transcripts from piRNA cluster. After Ago3 cleaves, the RNA sequence following the resulting Aub-bound secondary piRNA is cut processively by Zuc every ~27 nt.

Unexpectedly, most Ago3-dependent transposon families escape silencing in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ ovaries, despite wild-type levels of Piwi-bound, antisense piRNAs. Perhaps small RNA sequencing from whole ovaries fails to faithfully report piRNA levels for individual stages of oogenesis. piRNAs from early stages of oogenesis compose just a small fraction of piRNAs sequenced, but might play an important role in establishing transcriptional silencing. Too few Piwi-bound piRNAs early in *ago3*$^{t2/t3}$*, qin*$^{1/Df}$ oogenesis might impair transcriptional silencing, despite normal levels later. Alternatively, Ago3 may have additional functions, beyond initiating piRNA production, that support Piwi-mediated transcriptional silencing. Understanding why normal levels of Piwi-bound, antisense piRNAs do not suffice to silence transposon transcription will likely require the development of methods to isolate specific stages of ovarian germ cells and measure their piRNA and transcript abundance.

Catalytically Inactive Ago3 Reveals an Alternative Pathway for Primary piRNA Production

For most transposon families, cleavage by Ago3 triggers the production of Piwi-bound piRNAs. Nonetheless, for a few transposons, catalytically inactive Ago3 can produce Piwi-bound piRNAs. In this alternative pathway, an unidentified nuclease appears to generate the 5′ ends of the precursors destined to make Piwi-bound piRNAs.

Why does the Ping-Pong pathway and not the alternative pathway predominate during oogenesis? One difference between the two pathways is that only the Ping-Pong pathway can generate piRNAs that begin with U, which are hypothesized to bind more tightly to Aub. Aub prefers targets that place an A

across from the first nucleotide of the piRNA, irrespective of its identity (Wang et al., 2014). Consequently, Ago3-bound piRNAs typically have an A at their tenth position. When Ago3 cleaves a complementary RNA, it selects for 3′ cleavage products starting with U. In contrast, the alternative pathway does not provide Aub with piRNA precursors beginning with U.

Why does the alternative pathway exist? Target cleavage is unnecessary in the alternative pathway, so it is predicted to require less complementarity between an Ago3- or Aub-bound piRNA and its target RNA than the canonical Ping-Pong pathway. Thus, the alternative pathway might play a central role in the adaptation to new transposons, for which extensively complementary piRNAs are unlikely to exist.

Experimental procedures

Fly Strains

All fly strains were maintained at 25°C. *aub* and *ago3* strains were backcrossed to w^1 for 3–5 generations before use to minimize genetic background effects.

Protein Quantification by Western Blotting

Ovary lysate was prepared as described (Li et al., 2009). Anti-Ago3 was a gift from Julius Brennecke. Quantification was performed using ImageGauge v4.22 (Fuji, Japan).

Immunofluorescence

Immunofluorescence images were quantified using ImageJ 1.49m (Schindelin et al., 2012). Images presented in the same figure were acquired at the same settings unless otherwise noted.

Construction and Analysis of High-Throughput Sequencing Libraries

Small RNA (Han et al., 2015a), RNA-seq (Zhang et al., 2012b), and degradome-seq (Wang et al., 2014) libraries were constructed as described. Computational analysis employed piPipes (Han et al., 2015b).

Transgenes and fly strains

A 9,821 bp genomic fragment of *aubergine* was amplified by PCR from w^1 genomic DNA using oligonucleotides 5′-TCC CTT AGG AGC CTC AGC AT-3′ and 5′-TGC ATA AGG TTG CCC TAC CA-3′. A 3,084 bp genomic fragment of the putative *aubergine* promoter was amplified from w^1 genomic DNA using oligonucleotides 5′-GAG CGC AAC AGG TGT GTT ATT CTC-3′ and 5′-AAT TAC ACA AAT TAC AGT TGC ACT T-3′. Complementary DNA (cDNA), including the 5′ UTR, coding sequence and 3′ UTR of *ago3* isoform F (NM_001043163) was cloned previously (Li et al., 2009). *vas* promoter was amplified from *P{vas-egfp::vas}* (Sano et al., 2002) using oligonucleotide 5′-CTG CAG CTG GTT GTA GGT GC-3′ and 5′-GTG GAA TTT CCC ATT GTG CTA TCG-3′. The PCR products were inserted into plasmid pattB using restriction sites. To generate catalytically inactive mutant Ago3 and Aub, point mutations were introduced by PCR and confirmed by Sanger sequencing (GENEWIZ, Inc., MA, USA). Site-specific integration into the fly genome was performed using ΦC31 integrase (BestGene Inc., CA, USA). Flies carrying Aub^{WT}, Aub^{ADH}, $Ago3^{WT}$, $Ago3^{ADH}$ transgenes were crossed in to w^1 background. Flies carrying transgenic Aub^{ADH} were kept as males in w^1 or *aub* heterozygous background.

Protein Quantification

Ovary lysate was prepared as described (Li et al., 2009). Lysate from w^1 ovaries was serially diluted to make standard curve. Anti-Aub and anti-Piwi were used for immuno-detection as described (Zhang et al., 2011). Anti-Ago3 was a gift from Julius Brennecke. Quantification was performed using ImageGauge v4.22 (Fuji, Japan).

Construction and Analysis of High-Throughput Sequencing Libraries

Small RNA-seq

One-to-two-day-old female flies were fed yeast for three days before their ovaries were dissected. Total RNA was extracted using the mirVana kit (Ambion, NY, USA). Small RNA libraries were prepared as described (Han et al., 2015a) and sequenced using a HiSeq 2000 (Illumina, CA, USA). Barcodes were sorted allowing one mismatch, and 3′ adaptors were identified and removed using the first ten nucleotides, allowing one mismatch. After adaptor removal, reads containing one or more nucleotides with Phred score < 5 were discarded. Analysis employed the small RNA pipeline of piPipes (git commit 85d3357b38; Han et al., 2015b).

RNA-seq

Total RNA was purified using mirVana (Ambion), processed as described (Zhang et al., 2012b), and sequenced used a HiSeq 2000 (Illumina). RNA-seq and degradome-seq analysis was performed with piPipes. Briefly, RNAs were first aligned to ribosomal RNA (rRNA) sequences using Bowtie2 (v2.2.0; Langmead and Salzberg, 2012). Unaligned reads were then mapped using STAR to fly genome dm3 with TAS and without chrUextra (v2.3.1; Dobin et al., 2013a). Sequencing depth and gene quantification was calculated with Cufflinks (v2.1.1; Trapnell et al., 2010). In parallel, non-rRNA reads are aligned to a transcriptome composed of mRNA sequences and transposon consensus sequences using Bowtie2 (v2.2.0). Next, eXpress was used to assign multi-mappers and to count the number of reads for each transcript (v1.5.1; Roberts and Pachter, 2013). Differential expression analysis was performed using DESeq2 (v1.6.3; Anders

and Huber, 2010). To examine the derepression of transposons in mutants, transcript level of genes and transposons from w^1 were used as reference to estimate the normalization factors and to perform statistical tests. To determine the rescue of *aub* and *ago3* mutants with the corresponding transgenic wild-type or catalytically inactive protein, the corresponding mutants were used as reference. Differentially expressed transposons were defined only if the abundance change was more than twofold and the Benjamini-Hochberg adjusted *p*-value < 0.01.

Global Run-On-seq

GRO-seq was performed as described with slight modifications (Rozhkov et al., 2013) and sequenced using a NextSeq 500 (Illumina). Briefly, one-to-two-day-old female flies were fed in yeast for three days before their ovaries were dissected. One hundred pairs of ovaries were homogenized in HB35 buffer (15 mM HEPES-KOH, pH 7.5, 10 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 0.05% [v/v] NP-40, 0.35 M sucrose, 1 mM dithiothreitol, 1 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride, 0.3 µM Aprotinin, 40 µM Bestatin, 10 µM E-64, 10 µM Leupeptin). Nuclei were purified by passing twice through sucrose cushions that contain 800 µL HB80 buffer (15 mM HEPES-KOH, pH 7.5, 10 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM EGTA, 0.05% [v/v] NP-40, 0.80 M sucrose, 1 mM dithiothreitol, 1 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride, 0.3 µM Aprotinin, 40 µM Bestatin, 10 µM E-64, 10 µM Leupeptin) on the bottom phase and 350 µL HB35 buffer on the top. Nuclei were washed once with 500 µL freezing buffer (50 mM Tris-HCl, pH 8.0, 40% [v/v] glycerol, 5 mM MgCl2, 0.1 mM EDTA, 1 mM

dithiothreitol, 1 mM 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride, 0.3 μM Aprotinin, 40 μM Bestatin, 10 μM E-64, 10 μM Leupeptin) and frozen in liquid nitrogen with 100 μL freezing buffer. To carry out nuclear run-on assay, 100 μL freshly prepared reaction buffer (10 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 300 mM KCl, 1% [w/v] sarkosyl, 500 μM ATP, 500 μM GTP, 500 μM Br-UTP, 2.3 μM CTP, 1 mM dithiothreitol, 20 U Promega RNasin Plus RNase Inhibitor) was added to nuclei and incubated at 30°C for 5 min. RNA was extracted using Trizol (Invitrogen). Nascent RNAs with Br-UTP incorporated were enriched by immunoprecipitation using anti-5-bromo-2′-deoxyuridine antibody (clone PRB-1; EMD Millipore, Billerica, MA) as described (Shpiz and Kalmykova, 2014), followed by rRNA depletion using RNase H (Adiconis et al., 2013), fragmentation, and library construction (Zhang et al., 2012b). Analysis was carried out using the RNA-seq pipeline in piPipes.

Acknowledgments

We thank Julius Brennecke for Ago3 antibody; Alicia Boucher for fly husbandry; and members of the Weng and Zamore laboratories for discussions and critical comments on the manuscript. This work was supported in part by National Institutes of Health grants GM62862 and GM65236 to P.D.Z. and P01HD078253 to P.D.Z. and Z.W.

Chapter V Open questions and future directions

piRNA:target pairing rule

How do piRNAs, when incorporated in different PIWI proteins, pair to their targets and mediate target regulation transcriptionally or post transcriptionally? Knowing the specific targeting rule identifies authentic targets and elucidate the specific function of a piRNA. By examining *trans*-paring between Aub-bound piRNAs and Ago3-bound piRNAs, my work suggests that cleavable trans-targets of Aub-bound piRNA are generally complementary to their guides from g2 to at least position g16. Ago3 slices a target when piRNA-target complementarity extends from g2 to g14. Agreeing with this result, applying a perfect 6 nt seed shift analysis with *Z*-score as readout through all piRNA alignments, Brennecke and colleagues conclude that pairing to g2-g10/11 stands out as being the most important factor for target slicing. The complementarity from g12 to g15/16 is less important yet still more pronounced than background (Mohn et al., 2015). The new complementarity rule has great potentials in piRNA targets identification.

Our work suggest that the major function of Ago3 is to generate Piwi-bound primary piRNAs in germline, therefore, the transcriptional silencing of transposable elements and mRNAs would be determined by the target pool of the germline Piwi-bound piRNAs. Our conventional analysis which leads to the conclusion that piRNA mainly targets transposons was carried out by allowing 1 or fewer mismatches. Lau's lab reports that, if 2 mismatches are allowed, 5% of the piRNAs can target 50% of mRNAs in *Drosophila* OSS cell line. Furthermore, allowing 3 base mismatches results in a conclusion that nearly every mRNA

transcript paired with at least a piRNA, though it does not explicitly suggest that the paring would essentially lead to a substantial silencing. Using reports bearing different level of bulges in the seed region suggest that piRNA-guided target silencing does not tolerate seed mismatches (Post et al., 2014). Whether a mRNA transcript is silenced by Piwi-piRNA complex may depend on the nature of the target region (i.e., promoter region or gene body), the position of mismatches relative to the piRNA guide, the abundance of this piRNA and how many piRNAs cooperatively bind on the targets.

Based on the expression level of several endogenous mouse genes whose expression are altered by an exogenous human piRNA cluster, Goh et al. estimate several parameters for piRNA targeting: a primary seed for prefect matching in nucleotides 2-11 and a secondary seed with a maximum of four mismatches tolerable in nucleotides 12-21 (Goh et al., 2015). As the training set of targets is sparse, whether it can really represent Miwi-piRNA:target pairing rule is unclear. Nevertheless, different from miRNAs, “seed region” for piRNA:target pairing has not been structurally and biochemically identified.

The missing of Ping-Pong signature in adult mouse testes argues against the existence of MILI-MIWI Ping-Ping cycle in the adult mouse testes. (Beyret et al., 2012). Both intergenic cluster derived piRNAs and transposon derived piRNAs display no 10-nt 5′-5′ complementarity if full complementarity is required (Beyret et al., 2012). However, if certain mismatches are allowed during target scanning, MIWI-piRNAs appear to have the capacity to cleave diverse mRNAs.

The 10th position of those mRNA targets display an enrichment of Adenine (A) (Goh et al., 2015; Zhang et al., 2015), supporting the existence of mature piRNAs that are produced from the cleavage products of other piRNAs. However, whether this “Ping-Pong” cycle amplifies piRNAs is unknown. Whether mismatches between piRNAs and their targets at certain positions affect MIWI to cleave the targets remains enigmatic since for Argonautes slicing requires perfect complementarity between the guide strand and the target around the cleavage site. Although several studies attempted to elucidate the function of pachytene piRNAs in spermatogenesis, it needs to pinpoint which one, piRNA or MIWI or a complex with both, is the genuine regulator if removal of piRNAs does not affect the fertility of the mice (Goh et al., 2015; Zhang et al., 2015).

What maintains piRNA antisense bias

piRNAs from ovarian somatic cells are mainly antisense. The general fraction of antisense piRNAs in piRNA population from ovaries including germ cells is close to 0.7, suggesting that Ping-Pong cycles is biased to produce antisense piRNAs. Our research provides an explanation for that— within Ping-Pong cycle, Ago3 cleavage can also feed antisense piRNAs for Piwi to increase the total amount of antisense piRNAs (Han et al., 2015a). We also find that the amount of primary piRNAs or maternal piRNAs to kickstart the Ping-Pong is in small proportion of the piRNA pool in germ cells, thus, the antisense bias could not mainly come from the initial bias inherited from parents. Our piRNA sequencing data in *qin*$^{1/Df}$ and *ago3*$^{t2/t3}$*; qin*$^{1/Df}$ suggests that Aub:Aub Ping-Pong produce piRNAs in sense

and antisense orientation in similar level. Two other possibilities could explain the observation: (1) Ago3 is enzymatically more active than Aub so it is able to generate antisense piRNAs with a faster pace; (2) Because both Aub and Piwi bind to antisense piRNAs, the antisense bias might simply comes from the fact the total amount of Aub and Piwi is more than that of Ago3. The examination of the two possibilities would add to our understanding to the sources of antisense bias of germline piRNAs.

The mechanism of Qin

We demonstrated that the function of Qin is to suppress the transfer of Aub cleavage products to Aub itself. The existence of such mechanism ensures that futile Aub:Aub homotypic Ping-Pong does not dominate since they cannot produce sufficient antisense piRNAs loaded into Piwi. Nonetheless, the underneath mechanism of how Qin works remain elusive and calls for an in vitro biochemical system that can recapitulate at least part of the Ping-Pong pathway. Since this process happens in nuage with many components involved, such system might not be possible with current strategies.

The unknown of Piwi-mediated silencing

Another question remained is that, why Piwi-piRNAs fail to repress transposon activity in *ago3*$^{t2/t3}$, *qin*$^{1/Df}$ when their abundance has been recovered to the wild-type level? It suggests that factors other than the abundance of piRNAs are also indispensable for transposon repression. One candidate is the piRNA sense

fraction. Although removing Qin in *ago3*$^{t2/t3}$ background recovered the abundance of Piwi-bound piRNAs, it failed to recover the sense fraction—sense and antisense piRNAs in *ago3*$^{t2/t3}$, *qin*$^{1/Df}$ double-mutants are in similar level since they are both produced by homotypic Aub:Aub Ping-Pong. However, whether sense piRNA contribute negatively to transposon silencing remains to be tested. Another possibility is that our whole ovary sequencing failed to capture the details of piRNA abundance in different developmental stages. It is possible that the abundance of Piwi-bound piRNAs is only recovered in the late stages, which take a larger portion of the sequencing space. And it is possible that the level of Piwi-associated piRNAs remain low in the early stages, where transposons escape silencing. Testing this possibility requires isolation of germline nurse cells in different developmental stages and remains an interesting project in the future.

Phasing – a new adaptive perspective on transposon silencing

The phasing model might provide an explanation for the adaptation to the new transposon invasion during hybrid-dysgenesis. Maternally inherited transposons can be silenced in the offspring germline and they are therefore fertile. However, if the transposons are invaded paternally, the offspring are sterile due to the outbreak of transposon activities in their gonads. piRNAs was identified as the maternally inherited epigenetic information that silence transposons and explains the hybrid dysgenesis phenomenon (Brennecke et al., 2008). However, hybrid dysgenesis is recovered as dysgenic hybrids age — the fertility of female hybrid-dysgenesis progenies increases when they become older. Silencing of the

defects caused by invading *P* element suggest a different piRNA production pathway in which the piRNAs targeting *P* element are produced de novo from paternally inherited clusters. However, the efficiency of the production of those piRNAs is low– only a subset of the ovarioles within 26% of ovaries regained the fertility. The authors proposed that DNA damage signaling is overcome through a mechanism related to checkpoint adaptation. Further transposition studies suggest that silencing of transposition of the resident transposable elements, activated by the *P* element invasion, is critical in the adaptation step (Khurana et al., 2011).

P element adaptation might also benefit from Ago3 and Aub cleavage initiated, phased primary piRNA production. When maternal piRNAs or de novo primary piRNAs guide Aub to search for targets, non-homologous-derived piRNA might bind a *P* element transposon mRNA by tolerating limited number of mismatches outside of the seed regions. The slicing activity by Aub generates an Ago3-bound sense *P* element transposon piRNA, as well as triggers the downstream processing of the *P* elements via the primary machinery. The alternative pathway independent of the slicing activity might also contribute to new piRNAs loaded into Ago3. Those piRNAs are derived from the *P* element transcripts and thus have fully complementarity against their targets. The further processing of *P* element transcripts produce Aub- and Piwi-bound piRNAs that are fully complementary to the *P* elements and thus conduct efficienct repression. The accumulation of such piRNAs gradually establishes the defense system

against the invading P element. This might be the reason that the hybrids start to regain fertility only after 21 days. However, In order to avoid mis-targeting endogenous mRNAs, there must be some co-factors to differentiate the foreign invaders from endogenous self.

Figure 5.1

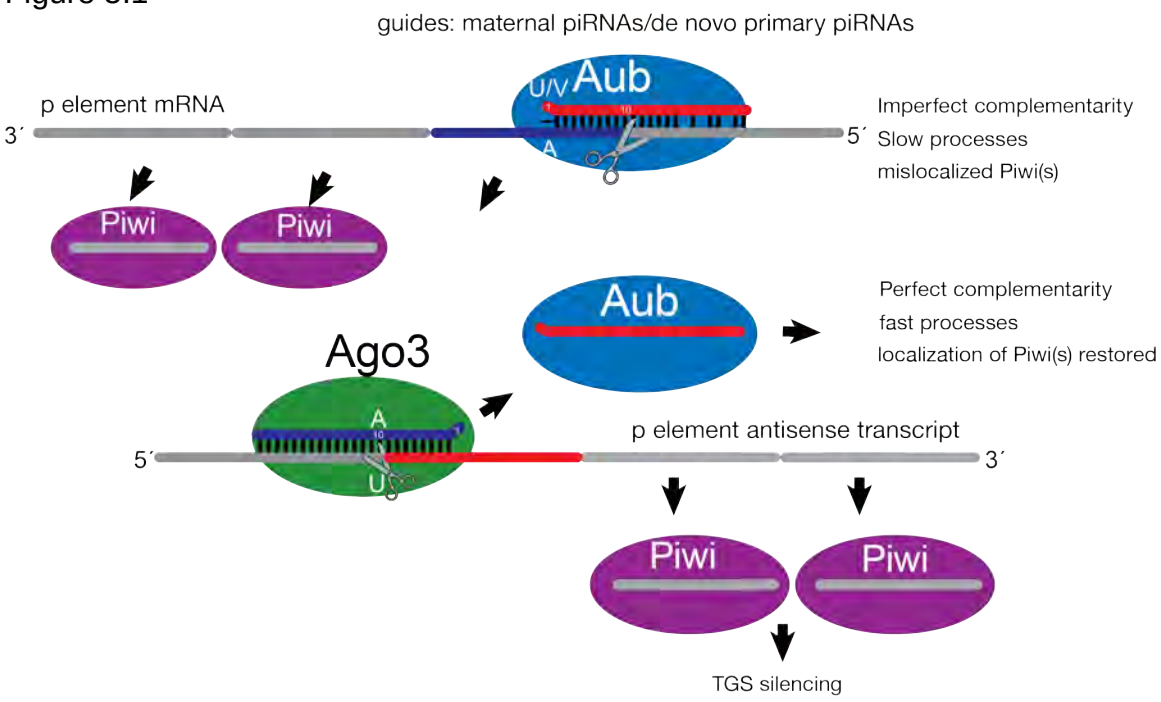

guides: maternal piRNAs/de novo primary piRNAs
p element mRNA
3´
5´
U/V
Aub
10
A
Imperfect complementarity
Slow processes
mislocalized Piwi(s)
Piwi
Piwi
Aub
Perfect complementarity
fast processes
localization of Piwi(s) restored
Ago3
A
10
U
p element antisense transcript
5´
3´
Piwi
Piwi
TGS silencing

Figure 5.1. The adaptive model to a newly invaded *P* element

BIBLIOGRAPHY

Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA Targets Identified by Sequencing of the *Arabidopsis* Degradome. Curr Biol *18*, 758-762.

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., Sivachenko, A., Thompson, D. A., Wysoker, A., Fennell, T., Gnirke, A., Pochet, N., Regev, A., and Levin, J. Z. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods *10*, 623-629.

Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207-221.

Ameres, S. L., Martinez, J., and Schroeder, R. (2007). Molecular basis for target RNA recognition and cleavage by human RISC. Cell *130*, 101-112.

Anand, A., and Kai, T. (2011). The tudor domain protein Kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. EMBO J 1593-1607.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., and Tuschl, T.

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

Aravin, A. A., and Chan, D. C. (2011). piRNAs meet mitochondria. Dev Cell *20*, 287-288.

Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. Dev Cell *5*, 337-350.

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

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

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

Armisen, J., Gilchrist, M. J., Wilczynska, A., Standart, N., and Miska, E. A. (2009). Abundant and dynamically expressed miRNAs, piRNAs, and other small RNAs in the vertebrate Xenopus tropicalis. Genome Res *19*, 1766-1775.

Bagijn, M. P., Goldstein, L. D., Sapetschnig, A., Weick, E. M., Bouasker, S., Lehrbach, N. J., Simard, M. J., and Miska, E. A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Berninger, P., Jaskiewicz, L., Khorshid, M., and Zavolan, M. (2011). Conserved generation of short products at piRNA loci. BMC Genomics *12*, 46.

Beyret, E., Liu, N., and Lin, H. (2012). piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism. Cell Res *22*, 1429-1439.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

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

Cenik, E. S., Fukunaga, R., Lu, G., Dutcher, R., Wang, Y., Tanaka Hall, T. M., and Zamore, P. D. (2011). Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease. Mol Cell *42*, 172-184.

Cenik, E. S., and Zamore, P. D. (2011). Argonaute proteins. Curr Biol *21*, R446-9.

Cora, E., Pandey, R. R., Xiol, J., Taylor, J., Sachidanandam, R., McCarthy, A. A., and Pillai, R. S. (2014). The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA *20*, 773-781.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Czech, B., Preall, J. B., McGinn, J., and Hannon, G. J. (2013). A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell *50*, 749-761.

Darricarrere, N., Liu, N., Watanabe, T., and Lin, H. (2013). Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proc Natl Acad Sci U S A *110*, 1297-1302.

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun *3*, 945.

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

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

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

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013a). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013b). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

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

Dufourt, J., Dennis, C., Boivin, A., Gueguen, N., Theron, E., Goriaux, C., Pouchin, P., Ronsseray, S., Brasset, E., and Vaury, C. (2014). Spatio-temporal requirements for transposable element piRNA-mediated silencing during Drosophila oogenesis. Nucleic Acids Res *42*, 2512-2524.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

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

Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001c). Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J *20*, 6877-6888.

Elkayam, E., Kuhn, C. D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., and Joshua-Tor, L. (2012). The Structure of Human Argonaute-2 in Complex with miR-20a. Cell *150*, 100-110.

Fang, W., Wang, X., Bracht, J. R., Nowacki, M., and Landweber, L. F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell *151*, 1243-1255.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. EMBO J *31*, 3588-3595.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5’-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Friedlander, M. R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T. A., Hirst, M., Marra, M., Nusbaum, C., Lee, W. L., Jenkin, J. C., Sanchez Alvarado, A., Kim, J. K., and Rajewsky, N. (2009). High-resolution profiling and discovery of planarian small RNAs. Proc Natl Acad Sci U S A *106*, 11546-11551.

Gasciolli, V., Mallory, A. C., Bartel, D. P., and Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr Biol *15*, 1494-1500.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L., Zapp, M. L., Weng, Z., and Zamore, P. D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

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

Goh, W. S., Falciatori, I., Tam, O. H., Burgess, R., Meikar, O., Kotaja, N., Hammell, M., and Hannon, G. J. (2015). piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. Genes Dev *29*, 1032-1044.

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep *15*, 411-418.

Grentzinger, T., Armenise, C., Brun, C., Mugat, B., Serrano, V., Pelisson, A., and Chambeyron, S. (2012). piRNA-mediated transgenerational inheritance of an acquired trait. Genome Res *22*, 1877-1888.

Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

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

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

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

Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M. C. (2007b). A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5′ End Formation *in Drosophi*la. Science *315*, 1587-1590.

Guzzardo, P. M., Muerdter, F., and Hannon, G. J. (2013). The piRNA pathway in flies: highlights and future directions. Curr Opin Genet Dev *23*, 44-52.

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

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

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

Han, B. W., Wang, W., Li, C., Weng, Z., and Zamore, P. D. (2015a). piRNA--Guided Transposon Cleavage Initiates Zucchini--Dependent, Phased piRNA Production. Science *in press*,

Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2015b). piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. Bioinformatics *31*, 593-595.

Han, B. W., and Zamore, P. D. (2014). piRNAs. Curr Biol *24*, R730-3.

Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., and Brennecke, J. (2013). The genetic makeup of the Drosophila piRNA pathway. Mol Cell *50*, 762-777.

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

Harris, A. N., and Macdonald, P. M. (2001). *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. Development *128*, 2823-2832.

Heler, R., Marraffini, L. A., and Bikard, D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. Mol Microbiol *93*, 1-9.

Honda, S., Kirino, Y., Maragkakis, M., Alexiou, P., Ohtaki, A., Murali, R., Mourelatos, Z., and Kirino, Y. (2013). Mitochondrial protein BmPAPI modulates the length of mature piRNAs. RNA *19*, 1405-1418.

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

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

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

Huang, X. A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A Major Epigenetic Programming Mechanism Guided by piRNAs. Dev Cell *24*, 502-516.

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

Izumi, N., Kawaoka, S., Yasuhara, S., Suzuki, Y., Sugano, S., Katsuma, S., and Tomari, Y. (2013). Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. RNA *19*, 896-901.

Jinek, M., and Doudna, J. A. (2008). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405-412.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M., and Celniker, S. E. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol *3*, RESEARCH0084.

Karginov, F. V., Cheloufi, S., Chong, M. M., Stark, A., Smith, A. D., and Hannon, G. J. (2010). Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. Mol Cell *38*, 781-788.

Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. RNA *15*, 1258-1264.

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

Kawaoka, S., Minami, K., Katsuma, S., Mita, K., and Shimada, T. (2008). Developmentally synchronized expression of two Bombyx mori Piwi subfamily

genes, SIWI and BmAGO3 in germ-line cells. Biochem Biophys Res Commun *367*, 755-760.

Kazazian, H. H. J. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Khurana, J. S., Wang, J., Xu, J., Koppetsch, B. S., Thomson, T. C., Nowosielska, A., Li, C., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell *147*, 1551-1563.

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

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

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

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., Zhang, F., Schultz, N., Koppetsch, B. S., Nowosielska, A., Seitz, H., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2009). The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Klenov, M. S., Lavrov, S. A., Stolyarenko, A. D., Ryazansky, S. S., Aravin, A. A., Tuschl, T., and Gvozdev, V. A. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. Nucleic Acids Res *35*, 5430-5438.

Klenov, M. S., Sokolova, O. A., Yakushev, E. Y., Stolyarenko, A. D., Mikhaleva, E. A., Lavrov, S. A., and Gvozdev, V. A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc Natl Acad Sci U S A *108*, 18760-18765.

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

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

Kwak, P. B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol *19*, 145-151.

Lambert, N. J., Gu, S. G., and Zahler, A. M. (2011). The conformation of microRNA seed regions in native microRNPs is prearranged for presentation to mRNA targets. Nucleic Acids Res *39*, 4827-4835.

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

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

Lau, N. C., Ohsumi, T., Borowsky, M., Kingston, R. E., and Blower, M. D. (2009a). Systematic and single cell analysis of Xenopus Piwi-interacting RNAs and Xiwi. EMBO J *28*, 2945-2958.

Lau, N. C., Robine, N., Martin, R., Chung, W. J., Niki, Y., Berezikov, E., and Lai, E. C. (2009b). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. Genome Res *19*, 1776-1785.

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

Le Thomas, A., Marinov, G. K., and Aravin, A. A. (2014a). A Transgenerational Process Defines piRNA Biogenesis in Drosophila virilis. Cell Rep *8*, 1617-1623.

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

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y. C., Luo, Y., Sachidanandam, R., Toth, K. F., Patel, D., and Aravin, A. A. (2014b).

Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev *28*, 1667-1680.

Lee, H. C., Gu, W., Shirayama, M., Youngman, E., Conte, D. J., and Mello, C. C. (2012). *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell *150*, 78-87.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

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

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

Lin, H., Chen, M., Kundaje, A., Valouev, A., Yin, H., Liu, N., Neuenkirchen, N., Zhong, M., and Snyder, M. (2015). Reassessment of Piwi Binding to the Genome and Piwi Impact on RNA Polymerase II Distribution. Dev Cell *32*, 772-774.

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

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Luteijn, M. J., and Ketting, R. F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat Rev Genet *14*, 523-534.

Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., and Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

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

Marinov, G. K., Wang, J., Handler, D., Wold, B. J., Weng, Z., Hannon, G. J., Aravin, A. A., Zamore, P. D., Brennecke, J., and Toth, K. F. (2015). Pitfalls of Mapping High-Throughput Sequencing Data to Repetitive Sequences: Piwi's Genomic Targets Still Not Identified. Dev Cell *32*, 765-771.

Megosh, H. B., Cox, D. N., Campbell, C., and Lin, H. (2006). The role of PIWI and the miRNA machinery in *Drosophila* germline determination. Curr Biol *16*, 1884-1894.

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

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

Muerdter, F., Guzzardo, P. M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G. J. (2013). A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. Mol Cell *50*, 736-748.

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

O'Carroll, D., Mecklenbrauker, I., Das, P. P., Santana, A., Koenig, U., Enright, A. J., Miska, E. A., and Tarakhovsky, A. (2007). A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. Genes Dev *21*, 1999-2004.

Oey, H. M., Youngson, N. A., and Whitelaw, E. (2011). The characterisation of piRNA-related 19mers in the mouse. BMC Genomics *12*, 315.

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

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

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

Pane, A., Wehr, K., and Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell *12*, 851-862.

Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M., and Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol Cell *33*, 204-214.

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

Parker, J. S., Roe, S. M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Pélisson, A., Song, S. U., Prud'homme, N., Smith, P. A., Bucheton, A., and Corces, V. G. (1994). 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.

Post, C., Clark, J. P., Sytnikova, Y. A., Chirn, G. W., and Lau, N. C. (2014). The capacity of target silencing by Drosophila PIWI and piRNAs. RNA *20*, 1977-1986.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Team, R. D. C. (2013). R: A language and environment for statistical computing. . {ISBN} 3-900051-07-0

Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., and Kandel, E. R. (2012). A Role for Neuronal piRNAs in the Epigenetic Control of Memory-Related Synaptic Plasticity. Cell *149*, 693-707.

Rangan, P., Malone, C. D., Navarro, C., Newbold, S. P., Hayes, P. S., Sachidanandam, R., Hannon, G. J., and Lehmann, R. (2011). piRNA Production Requires Heterochromatin Formation in *Drosophila*. Curr Biol *21*, 1373-1379.

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

Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J., Hannon, G. J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol *12*, 340-349.

Roberts, A., and Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods *10*, 71-73.

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

Rorth, P. (1998). Gal4 in the Drosophila female germline. Mech Dev *78*, 113-118.

Rouget, C., Papin, C., Boureux, A., Meunier, A. C., Franco, B., Robine, N., Lai, E. C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. Nature *467*, 1128-1132.

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

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

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M. C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates *2′*-O-methylation of Piwi- interacting RNAs at their 3′ ends. Genes Dev *21*, 1603-1608.

Sano, H., Nakamura, A., and Kobayashi, S. (2002). Identification of a transcriptional regulatory region for germline-specific expression of vasa gene in Drosophila melanogaster. Mech Dev *112*, 129-139.

Sato, K., Iwasaki, Y. W., Shibuya, A., Carninci, P., Tsuchizawa, Y., Ishizu, H., Siomi, M. C., and Siomi, H. (2015). Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Mol Cell

Saxe, J. P., Chen, M., Zhao, H., and Lin, H. (2013). Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. EMBO J *32*, 1869-1885.

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

Schirle, N. T., Sheu-Gruttadauria, J., and MacRae, I. J. (2014). Structural basis for microRNA targeting. Science *346*, 608-613.

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

Senti, K. A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. Trends Genet *26*, 499-509.

Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D. C., van der Oost, J., Patel, D. J., and Wang, Y. (2014). Structure-based cleavage mechanism of

Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. Proc Natl Acad Sci U S A *111*, 652-657.

Shpiz, S., and Kalmykova, A. (2014). Analyses of piRNA-mediated transcriptional transposon silencing in Drosophila: nuclear run-on assay on ovaries. Methods Mol Biol *1093*, 149-159.

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

Simon, B., Kirkpatrick, J. P., Eckhardt, S., Reuter, M., Rocha, E. A., Andrade-Navarro, M. A., Sehr, P., Pillai, R. S., and Carlomagno, T. (2011). Recognition of 2'-O-Methylated 3'-End of piRNA by the PAZ Domain of a Piwi Protein. Structure *19*, 172-180.

Siomi, M. C., Sato, K., Pezic, D., and Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.

Specchia, V., Piacentini, L., Tritto, P., Fanti, L., D'Alessandro, R., Palumbo, G., Pimpinelli, S., and Bozzetti, M. P. (2010). Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. Nature *463*, 662-665.

Swarts, D. C., Mosterd, C., van Passel, M. W., and Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. PLoS One *7*, e35888.

Tolia, N. H., and Joshua-Tor, L. (2007). Slicer and the Argonautes. Nat Chem Biol *3*, 36-43.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

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

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

Voigt, F., Reuter, M., Kasaruho, A., Schulz, E. C., Pillai, R. S., and Barabas, O. (2012). Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. RNA *18*, 2128-2134.

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

Wang, S. H., and Elgin, S. C. (2011). *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.

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

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D. J. (2008a). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature *456*, 921-926.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T., and Patel, D. J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754-761.

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

Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P. J., Toyoda, A., Gotoh, K., Hiura, H., Arima, T., Fujiyama, A., Sado, T., Shibata, T., Nakano, T., Lin, H., Ichiyanagi, K., Soloway, P. D., and Sasaki, H. (2011). Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. Science *332*, 848-852.

Wee, L. M., Flores-Jasso, C. F., Salomon, W. E., and Zamore, P. D. (2012). Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell *151*, 1055-1067.

Welker, N. C., Maity, T. S., Ye, X., Aruscavage, P. J., Krauchuk, A. A., Liu, Q., and Bass, B. L. (2011). Dicer's Helicase Domain Discriminates dsRNA Termini to Promote an Altered Reaction Mode. Mol Cell *41*, 589-599.

Wickersheim, M. L., and Blumenstiel, J. P. (2013). Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. Biotechniques *55*, 269-272.

Williams, R. W., and Rubin, G. M. (2002). ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. Proc Natl Acad Sci U S A *99*, 6889-6894.

Wilson, J. E., Connell, J. E., and Macdonald, P. M. (1996). aubergine enhances oskar translation in the Drosophila ovary. Development *122*, 1631-1639.

Xie, Z., Allen, E., Wilken, A., and Carrington, J. C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A *102*, 12984-12989.

Xiol, J., Cora, E., Koglgruber, R., Chuma, S., Subramanian, S., Hosokawa, M., Reuter, M., Yang, Z., Berninger, P., Palencia, A., Benes, V., Penninger, J., Sachidanandam, R., and Pillai, R. S. (2012). A Role for Fkbp6 and the Chaperone Machinery in piRNA Amplification and Transposon Silencing. Mol Cell

Xiol, J., Spinelli, P., Laussmann, M. A., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., Kadlec, J., Sachidanandam, R., Kaksonen, M., Cusack, S., Ephrussi, A., and Pillai, R. S. (2014). RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. Cell *157*, 1698-1711.

Yoshikawa, M., Peragine, A., Park, M. Y., and Poethig, R. S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev *19*, 2164-2175.

Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., and Patel, D. J. (2005). Crystal structure of *A. aeolicus* Argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell *19*, 405-419.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B. S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2012a). UAP56 Couples piRNA Clusters to the Perinuclear Transposon Silencing Machinery. Cell *151*, 871-884.

Zhang, P., Kang, J. Y., Gou, L. T., Wang, J., Xue, Y., Skogerboe, G., Dai, P., Huang, D. W., Chen, R., Fu, X. D., Liu, M. F., and He, S. (2015). MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. Cell Res *25*, 193-207.

Zhang, Z., Koppetsch, B. S., Wang, J., Tipping, C., Weng, Z., Theurkauf, W. E., and Zamore, P. D. (2014). Antisense piRNA amplification, but not piRNA

production or nuage assembly, requires the Tudor-domain protein Qin. EMBO J *33*, 536-539.

Zhang, Z., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2012b). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence *3*, 9.

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

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