2013-11-06

piRNA Biogenesis and Transposon Silencing in Drosophila: A Dissertation

Zhao Zhang
University of Massachusetts Medical School

piRNA BIOGENESIS AND TRANSPOSON SILENCING IN *DROSOPHILA*

A Dissertation Presented

By

Zhao Zhang

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 6, 2013

RNA Therapeutics Institute

piRNA PRODUCTION AND TRANSPOSON SILENCING IN *DROSOPHILA*

Dissertation Presented By

Zhao Zhang

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

_____
Phillip Zamore, Ph.D., Thesis Advisor

_____
William Theurkauf, Ph.D., Thesis Advisor

_____
Eric Baehrecke, Ph.D., Member of Committee

_____
Eric Lai, Ph.D., Member of Committee

_____
Sean Ryder, Ph.D., Member of Committee

_____
Zhiping Weng, 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

_____
Victor Ambros, 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 Sciences

Interdisciplinary Graduate Program

November 6, 2013

DEDICATION

TO MYSELF, NEVER STOP FIGHTING FOR A BETTER SELF.

"Stay hungry, stay foolish."

ACKNOWLEDGEMENTS

If I could redo my graduate school, there are many things I can certainly do much better. But it's absolutely impossible for me to find better advisors than my two fantastic Ph.D. mentors, Phil and Bill. Both of them are the best mentors and scientists in the world, but have beauties in different ways, just like two parents of one child. Their mentoring certainly changed me beyond the scientific research; they are still changing who I am, and more importantly, who I can be. In my life, I always burden with a short list of names, to whom I feel I owe so much that I don't even know how to pay them back. Fortunately and unfortunately, I'll always see Phil and Bill's names on the top of that list.

I also owe much gratitude to Zhiping, who has constantly supported my research in the "dry" way and provided me a great opportunity to learn bioinformatics; to Chengjian, who patiently corrected every single mistake I made; to Jie, who taught me all of my computational skills and shared with me her great smartness. For my research, I was lucky to have wonderful and trustworthy teammates: Jie, Jia, Nadine, Birgit, Cindy and Swapnil. Many thanks to my TRAC members (Victor Ambros, Eric Baehrecke, Sean Ryder, Zhiping Weng) for their insightful suggestions, not only to my research, but also to my career development!

During my graduate school life, there have been two groups of people who try their best to make fun of each other, but "somehow" still constantly love each other, help each other and learn from each other. I call them my wonderful labmates: one group is the Zamore family (Desi, Tracey, Cindy, Jen, Bo, Megha, Stefan, Gwen, Daniel, Tim, Sam, Wes, Cha San, Xuan, Cansu, Tiff, Amena, Wei, Xin, Chengjian, Herve, Wee, Ryuya, Ira, Shengmei, Alicia, Fabian, Jogi, Keith,

Chris and Elif,) and the other one is the Theurkauf clan (Travis, Birgit, Swapnil, Nadine, Fan, Jaspreet, Alfred, Gen and Anetta). They made my research and life easier, my English communicable, and my heart stronger. Most importantly, it is because of them, I was willing to come to the lab every morning when I opened my eyes. I also want to extend my gratitude to my friends in China and the US, who always help and support me along with my growing up; to Darla and Dr. Eric Lai for their help so that I can finally claim my degree; to the Baehrecke lab, who gave our family so much help and happy memories.

With love, I'd like to thank my family in China. Without the immeasurable love and sacrifice from my parents and sister in the past 30 years, I wouldn't be here today. Because of their love, I can freely chase my dream. Because of their love, I'm not afraid of the dark. I hope I made them proud! I'll always deeply appreciate the tremendous help and love from my parents-in-law, who take care of me like their own son. My brother-in-law, who also gave our family so much support and makes my life easier. Lastly, I need to mention one more name. She not only has beauty, but also owns purity and integrity. She is the one to remind me to give my mom a birthday phone call; to perk me up when I am down; to tell me slow down when I am in a rush. She is my wife, Lin, the boss of my life. Thank you for your company at every single second; for being the best wife and mom (Leo may say something else later…); for giving me the chance to win "the best husband" trophy (probably "the best daddy" trophy too)…

ABSTRACT

piRNAs guide PIWI proteins to silence transposons in animal germ cells. In *Drosophila*, the heterochromatic piRNA clusters transcribe piRNA precursors to be transported into nuage, a perinuclear structure for piRNA production and transposon silencing. At nuage, reciprocal cycles of piRNA-directed RNA cleavage—catalyzed by the PIWI proteins Aubergine (Aub) and Argonaute3 (Ago3) in *Drosophila*—destroy the sense transposon mRNA and expand the population of antisense piRNAs in response to transposon expression, a process called the Ping-Pong cycle. Heterotypic Ping-Pong between Aub and Ago3 ensures that antisense piRNAs predominate.

My thesis research mainly focuses on two fundamental questions about the piRNA production: How does the germ cell differentiate piRNA precursor from mRNAs for piRNA biogenesis? And what is the mechanism to impose Aub Ping-Pong with Ago3? For the first question, we show that the HP1 homolog protein Rhino marks the piRNA cluster regions in the genome for piRNA biogenesis. Rhino seems to anchor a nuclear complex that suppresses cluster transcript splicing, which may differentiate piRNA precursors from mature mRNAs. Moreover, LacI::Rhino fusion protein binding suppresses splicing of a reporter transgene and is sufficient to trigger *de novo* piRNA production from a trans combination of sense and antisense transgenes. For the second question, we show that Qin, a new piRNA pathway factor contains both E3 ligase and Tudor domains, colocalizes with Aub and Ago3 in nuage, enforces heterotypic Ping-Pong between Aub and Ago3. Loss of *qin* leads to less Ago3 binding to Aub, futile Aub:Aub homotypic Ping-Pong prevails, antisense piRNAs decrease, many

families of mobile genetic elements are reactivated, DNA damage accumulates in the germ cells and flies are sterile.

LIST OF FIGURES

LIST OF TABLES

COPYRIGHT NOTICE

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION

TRANSPOSON, piRNA AND GENOME STABILITY

The genome faces external challenges, including radiation damage and virus integration, and internal "shocks" from transposon element jumping, which was first discovered by McClintock six decades ago (McClintock, 1950). Transposable elements occupy chromosomes from prokaryotes to eukaryotes, and in certain species make up most of the genome (Burns and Boeke, 2012; Finnegan, 2012; Slotkin and Martienssen, 2007). Although they could be useful driving forces during evolution, they can also lead to genome damage by creating DNA breaks, genome rearrangements, and insertional mutations (Demerec, 1926; Demerec, 1927; McClintock, 1950; Klattenhoff et al., 2007). The genome must respond, especially for germline cells, which are dedicated to faithfully passing the genetic information from generation to generation. How does the germline address this challenge?

In 2006, researchers identified a highly conserved small RNA based pathway that functions in the animal gonad to suppress transposon activity. Because these small RNAs associated with PIWI clade Argonaute proteins, they were named PIWI interacting RNAs (piRNAs) (Vagin et al., 2006; Girard et al., 2006; Aravin et al., 2006; Grivna et al., 2006; Lau et al., 2006; Saito et al., 2006). Although this pathway was identified over seven years ago, we still largely do not know how piRNAs are produced or how they silence transposons. The piRNA pathway has been most extensively studied in *Drosophila*, largely due to the combination of powerful genetics and advances in high throughput sequencing technology. Here I will summarize our current understanding of piRNA biogenesis, and how these small RNAs silence transposons.

THE DISCOVERY OF THE piRNA PATHWAY

Studies from *Drosophila* have played a leading role in defining the piRNA pathway. Numerous piRNA pathway genes were first identified in EMS screens for genome loci required for the fly fertility. These screens, which were largely performed in the late 1980s and early 1990s, identified several key components of the pathway, including *spindle-E*, *vasa*, *aubergine* and *rhino*. Individually mutating any of these genes disrupts egg patterning and results in the failure of embryos to hatch (Schupbach and Wieschaus, 1989; Schupbach and Wieschaus, 1991). In 1997, Lin and colleagues identified one similar locus on the 2nd chromosome, which was indispensable for germline development and fertility for both male and female flies. Based on the gonad morphology defect, it was named as *piwi*, short for P-element induced wimpy testes (Lin and Spradling, 1997). It encodes the founding member of a sub-family of Argonaute proteins, PIWI-clade Argonaute, which are highly enriched in the animal germline. At the time, it was clear that these genes were essential for germline function, but nothing was known about their molecular function.

The first piRNAs were discovered in 2001 (Aravin et al., 2001). The *Drosophila* X chromosome contains multiple copies of the "selfish" gene *Stellate*, which encodes a protein with homology to the β subunit of protein kinase CK2. Overexpression of this gene causes Stellate protein accumulated and assembly into crystals during spermatogenesis, compromising male fertility (Livak, 1984; Livak, 1990; Balakireva et al., 1992). It had been documented that the fly Y chromosome contains the *Suppressor of Stellate* (*Su[Ste])* loci, which silence *Stellate* genes on the X chromosome (Livak, 1984; Livak, 1990; Balakireva et al., 1992). But

how *Su[Ste]* loci suppress *Stellate* expression was not known. In 2001, Gvozdev and colleagues elegantly showed that both strands of the *Su[Ste]* repeats are transcribed, and the transcripts were processed into small RNAs about 25 nt long. These small RNAs conveyed Stellate silencing in the *spindle-E* dependent manner (Aravin et al., 2001). Subsequent small RNA cloning revealed that these *Su[Ste]* small RNAs represent a new class of small RNAs mainly/solely expressed in fly ovaries and testes. Since these small RNAs mainly mapped to transposons and other repeat regions, they were initially named rasiRNAs, short for repeat associated small interfering RNAs (Aravin et al., 2003). In 2006, several groups independently reported that these small RNAs associate with PIWI clade Argonaute, and that they are produced by a mechanism distinct from the well studied miRNAs and siRNAs (Vagin et al., 2006; Girard et al., 2006; Aravin et al., 2006; Grivna et al., 2006; Lau et al., 2006; Saito et al., 2006). Moreover, mammalian PIWI proteins bind to small endogenous RNAs with similar length, but unlike flies, these species mainly arise from intergenic regions devoid of transposons. Therefore, based on their PIWI protein binding pattern, we uniformly refer these animal germline specific small RNA as piRNAs (PIWI interacting RNAs) (Vagin et al., 2006; Girard et al., 2006; Aravin et al., 2006; Grivna et al., 2006; Lau et al., 2006; Saito et al., 2006). The piRNAs, PIWI proteins, and other factors needed for piRNA biogenesis and function that were previously identified as essential proteins to maintain fly fertility, form the piRNA pathway, which defends the germline genome from transposon mobilization.

PIWI PROTEINS

To execute their functions, piRNAs must associate with PIWI clade Argonaute proteins, a larger family includes the proteins that mediate RNA interference and microRNA directed silencing. Like other Argonautes, PIWI proteins consist of three conserved domains: PAZ, Mid and PIWI. Though none of the PIWI proteins have been crystallized, the structures of related Argonautes from the prokaryote *Thermus thermophiles*, yeast, and humans predict that the Mid domain of PIWI proteins anchors the 5′ end phosphate of the piRNA, while the PAZ domain locks the 3′ end. The PIWI domain is structurally similar to RNase H, and contains the "DDH" catalytic triad associated with cleavage/slicer activity (Nakanishi et al., 2012; Elkayam et al., 2012).

The *Drosophila* genome encodes three PIWI proteins: Argonuate3 (Ago3), Aubergine and the family founder protein, Piwi. In fly ovaries, Ago3 and Aub are only expressed in the germline nurse cells and reside in an electron-dense perinuclear structure, nuage (Figure 1.1). By contrast, Piwi protein localizes in the nuclei of both germline nurse cells and the somatic lineage follicle cells that surround the germ cells (Figure 1.1). Loss of any of the three fly PIWI proteins leads to transposon activation in the germline and female sterility (Wilson et al., 1996; Cox et al., 1998; Cox et al., 2000; Saito et al., 2006; Gunawardane et al., 2007; Brennecke et al., 2007). In mice, the PIWI clade also comprises three PIWI proteins: Mili, Miwi and Miwi2. In contrast to flies, where the females are more sensitive to piRNA pathway mutations, the mouse PIWI proteins are essential for spermatogenesis and mutations do not appear to disrupt female fertility (Aravin et al., 2007; Aravin et al., 2008). Similarly, the PIWI proteins are required to maintain fecundity in both worms and zebrafish (Batista et al., 2008; Houwing et al., 2007).

Figure 1.1

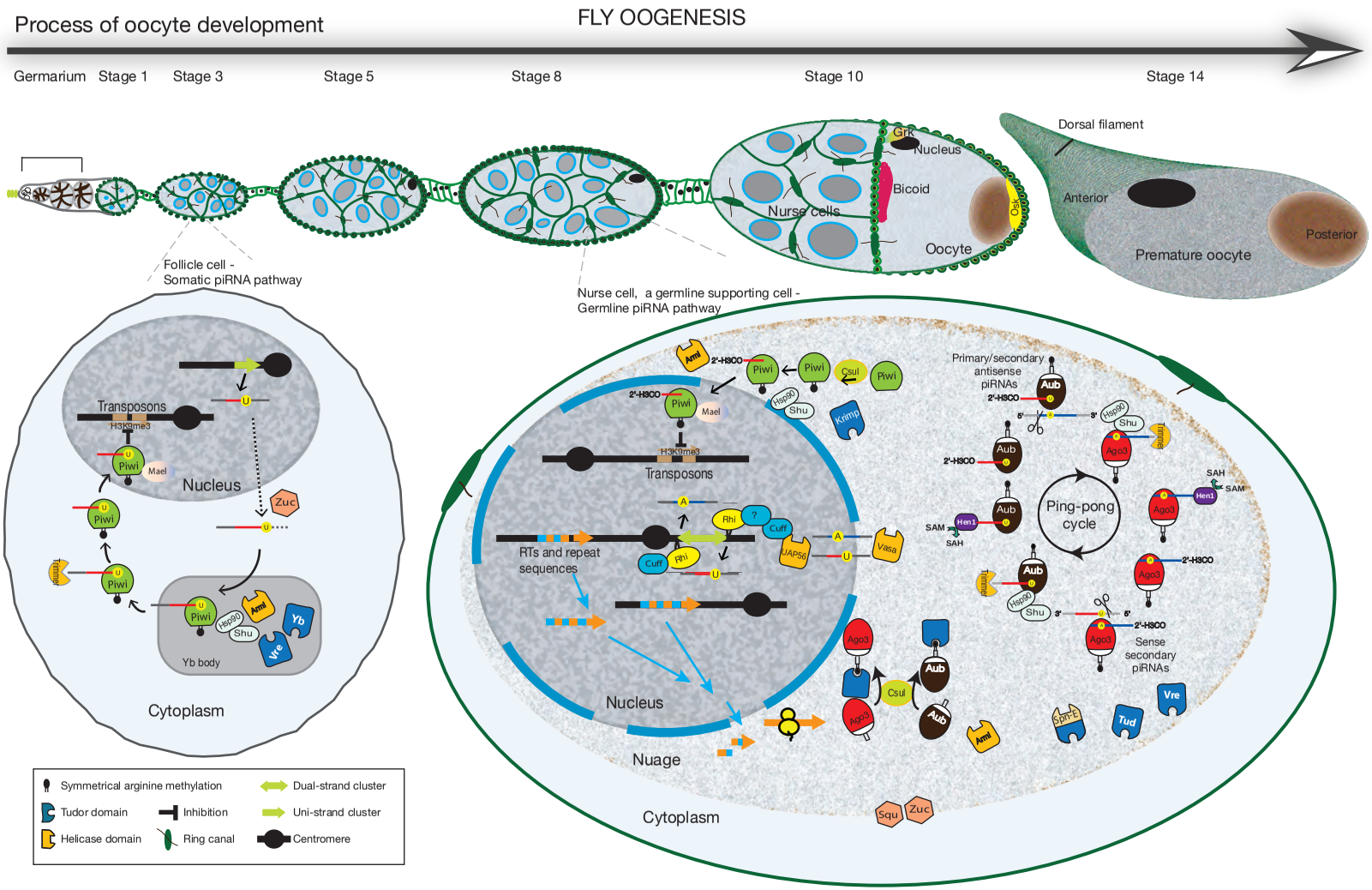
Process of oocyte development
FLY OOGENESIS
Germarium
Stage 1
Stage 3
Stage 5
Stage 8
Stage 10
Stage 14
Grk
Nucleus
Bicoid
Nurse cells
Oocyte
Osk
Dorsal filament
Anterior
Posterior
Premature oocyte
Follicle cell -
Somatic piRNA pathway
Nurse cell, a germline supporting cell -
Germline piRNA pathway
Transposons
H3K9me3
Piwi
Mael
Nucleus
Zuc
Hsp90
Shu
Ami
Yb
Vre
Yb body
Cytoplasm
Csul
Krmp
2'-H3CO
RTs and repeat
sequences
Rhi
Cuff
UAP56
Vasa
Nuage
Primary/secondary
antisense
piRNAs
Aub
Ago3
Ping-pong
cycle
Hen1
SAM
SAH
Sense
secondary
piRNAs
Spn-E
Tud
Squ
Symmetrical arginine methylation
Tudor domain
Helicase domain
Dual-strand cluster
Inhibition
Uni-strand cluster
Ring canal
Centromere

Figure 1.1. piRNA biogenesis in *Drosophila* ovaries.

From flies to mice, the PIWI proteins have a conserved post-translational modification, arginine symmetric di-methylation (Kirino et al., 2009; Nishida et al., 2009; Wang et al., 2009a; Vagin et al., 2009; Siomi et al., 2010). This modification is catalyzed by the methyltransferase Capsuleen (Csul, also known as dPRMT5 or Dart5), aided by its cofactor Valois (Vls) (Figure 1.1 and 1.2). Interestingly, mutating *csul* locus in the genome disrupts fly germline formation (Gonsalvez et al., 2006). The methylated arginines can serve as docking sites for the Tudor domain, and many components of the piRNA pathway contain either one or more Tudor domains, such as Spn-E, Krimper, Tejas, PAPI, Vreteno, Yb, Brother of Yb, Sister of Yb and the founder protein Tudor. Intriguingly, although they all share related Tudor domains, they do not appear to function redundantly, since loss any single protein leads to the typical piRNA pathway mutation phenotypes: activation of transposons and fly sterility (Gillespie and Berg, 1995; Lim and Kai, 2007; Patil and Kai, 2010; Handler et al., 2011). Similar observations have been reported in mice (Hosokawa et al., 2007; Chen et al., 2009; Lachke et al., 2011; Mathioudakis et al., 2012). We do not understand the molecular roles for most of these proteins, but studies in flies and zebrafish suggest that some may serve as bridges that bring two PIWI proteins together, or as docking sites for piRNA loading (Nishida et al., 2009; Huang et al., 2011a). In chapters 2 and 3 of this thesis, I will describe identification of a new piRNA pathway factor, Qin (also known as Kumo), which contains both E3 ligase domain and 5 Tudor domains, and its function in mediating piRNA production and transposon silencing.

piRNA CLUSTERS: THE HOME OF piRNA IN THE GENOME

Figure 1.2

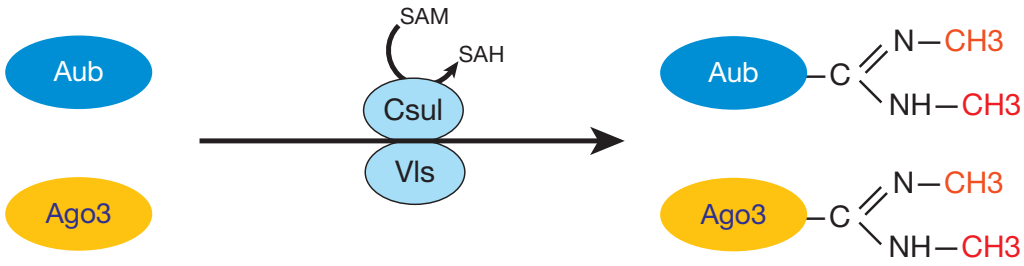

Figure 1.2. Both Aub and Ago3 contain arginine symmetric di-methylation.

In *Drosophila* ovaries, nearly 80% of the piRNAs map to transposon consensus sequences, making it hard to determine their genomic source. The other ~20% of the piRNAs that uniquely map to the genome locus are mainly derived from large discrete genomic loci, termed piRNA clusters (Brennecke et al., 2007). The top 142 piRNA clusters account for only 3.5% of the assembled fly genome, but produce 81% of the unique mapping piRNAs and 92% of the total piRNA population (Brennecke et al., 2007). These cluster regions are enriched in transposon remnants and ancient fragmented repeats, suggesting that they are the genomic memory of invading elements, likely introduced by horizontal transposon transfer or ancient virus infection. In flies, most of these piRNA clusters are located at the peri-centromeric and sub-telomeric heterochromatin regions of the chromosome (Brennecke et al., 2007).

Based on the pattern of unique piRNA production, clusters fall into two groups: uni-strand and dual-strand. While dual-strand piRNA clusters convergently produce piRNAs from both genomic strands, the uni-strand clusters are transcribed from one strand. The 42AB piRNA cluster, named for its chromosomal cytological position, is the largest dual-strand cluster in *Drosophila* ovaries and produces about 30% of ovary piRNAs (Brennecke et al., 2007). The *flamenco* locus, which has been shown to actively suppress the gypsy transposon family in somatic follicle cells, appears to be transcribed on one genomic strand and produces piRNAs that are complementary to gypsy transposon family RNAs (Brennecke et al., 2007).

In addition to clusters, it has been reported that a small portion of fly ovary piRNAs map to the 3′ UTRs of canonical mRNAs. These genic piRNAs have been mainly studied in the somatic follicle cells (Lau et al., 2009b; Lau et al.,

2009a; Robine et al., 2009; Saito et al., 2009). It is unclear whether they are present in germline cells, and their biological functions remain to be determined.

What distinguishes the piRNA clusters from the other non-piRNA-producing regions is not well understood. This specificity might be due to the modification of histones that associate with piRNA clusters. This hypothesis is supported by the finding that a rapidly evolving HP1 homologue, Rhino (Rhi, also known as HP1D), is required for dual-strand cluster piRNA production. Furthermore, Rhi also appears to localize at these loci (Figure 1.1; Klattenhoff et al., 2009). Rhi was first discovered in 1991 as a gene that is required for fly fertility and egg dorsal-ventral patterning. Disrupting this gene leads to fusion of two dorsal appendages on the egg, a phenotype shared by almost all other piRNA pathway component mutations and resembles the horn of the rhinoceros (Volpe et al., 2001). Like other HP1 genes, *rhi* encodes a protein that contains both Chromo and Chromoshadow domains (Volpe et al., 2001). Since HP1a is recruited to chromatin by H3K9me3 (Danzer and Wallrath, 2004), a histone modification that usually marks DNA regions for gene silencing, Rhi presumably coats the piRNA clusters by a similar mechanism. Supporting this view, Lehmann and colleagues reported that *Drosophila* piRNA clusters are coated by this histone modification (Rangan et al., 2011). The same is true for the silkworm (*Bombyx mori*) piRNA clusters (Kawaoka et al., 2013). However, H3K9me3 is present in the genome in many more regions than piRNA clusters, and these non-cluster-H3K9me3-loci do not appear to bind Rhi. Therefore, H3K9me3 itself is not sufficient to recruit Rhi.

piRNA BIOGENESIS: TRANSCRIPTION

While it appears that different organisms employ distinct strategies to initiate the piRNA precursor transcription, all appear to use RNA Polymerase II (Pol-II). In flies, each piRNA cluster is assumed to have one promoter at the end(s), which produces one long transcript that is processed into multiple piRNAs. This idea is primarily supported by the finding that mutating the putative *flamenco* cluster promoter drastically reduces piRNA production from the 168 kb downstream region (Brennecke et al., 2007). However, *flam* is a uni-strand cluster that appears to be expressed in the follicle cells, and it is unclear if germline specific dual-strand clusters behave in the same manner. In mice, a germline specific MYB transcription factor family protein, A-MYB, drives piRNA production in pachytene stage testes (Li et al., 2013). These cluster transcripts have 5′ caps and 3′ Poly (A) tails, typical features of Pol-II transcribed RNAs (Li et al., 2013). Similarly, the cluster transcript in the silkworm shares these Pol-II production features (Kawaoka et al., 2013).

In contrast, the nematode *C. elegans* produces piRNAs, called as 21U-RNA, in a different way. Unlike fly or mouse piRNAs, most worm piRNA species appear to be derived from distinct gene-like loci (Gu et al., 2012; Cecere et al., 2012). For each transcription unit, there is a conserved 8-mer motif at ~40 bp upstream of the piRNA producing site: CTGTTTCA, which is specifically recognized by the Forkhead family (FKH) transcription factor to initiate the downstream transcription. Transcription precisely starts 2 nt upstream of the mature piRNA, and produces a 5′ capped piRNA precursor with the length of ~ 26 nt (Gu et al., 2012; Cecere et al., 2012). How this longer precursor is processed into a mature 21U-RNA is still not clear.

piRNA BIOGENESIS: PRECURSOR DELIVERY

piRNA precursors are produced in the nucleus, yet most of the processing machinery is enriched in the nuage, a perinuclear structure conserved from *C. elegans* to mammals. How are precursors delivered to the nuage? Studies from our labs showed that the DEAD box protein UAP56 is important at this delivering step (Figure 1.1; Zhang et al., 2012a). This protein was separately identified in 1997 as a component of the spliceosome in human, and named as 56-KD U2AF65 associated protein (UAP56) (Fleckner et al., 1997), and as an enhancer for position effect variegation in *Drosophila*, and named as Hel (Eberl et al., 1997). Due to its essential function, loss of UAP56 leads to animal lethality. However, a recent study reported that a point mutation of *uap56* affects only fly fertility and shows typical piRNA pathway defects in *gurken* and *oskar* mRNA localization and egg patterning (Meignin and Davis, 2008). These observations suggested that UAP56 could be a piRNA pathway factor. Indeed, our studies showed that this allele blocks piRNA production and leads to transposon over-expression (Zhang et al., 2012a). Moreover, UAP56 physically associates with the piRNA cluster transcripts and forms distinct nuclear speckles that co-localize with the HP1 family protein Rhi. Intriguingly, UAP56-Rhi foci at the nuclear envelope are often juxtaposed with Vasa foci at the nuclear periphery (Zhang et al., 2012a). Vasa is another DEAD box protein that localizes to nuage and is required for germline development and piRNA production. The UAP56 point mutation that disrupts fertility and piRNA production also disrupts nuage localization of Vasa and other piRNA proteins (Zhang et al., 2012a). We have therefore proposed that UAP56 couples cluster transcription to the perinuclear

piRNA processing machinery. We currently do not know whether this protein conveys the same precursor delivery function in other organisms.

Rhi localization is also dependent on Cutoff (Cuff) (Pane et al., 2011), the fly homolog of yeast Rai1, which interacts with the nuclear 5'–3' exoribonuclease Rat1 (Kim et al., 2004). Together with the data that Rhi co-immunoprecipitated with Cuff, Rhi, Cuff and UAP56 may function in a common step in the nucleus (Figure 1.1).

What is the mechanism that distinguishes the piRNA precursors from mRNAs, and directs them to the piRNA biogenesis machinery? In chapter 4, I will present findings suggest that Rhi anchors a nuclear complex, containing at least Cuff and UAP56, which actively suppresses cluster transcript splicing. We speculate that this block of splicing differentiates piRNA precursors from mature mRNAs.

piRNA BIOGENESIS: PROCESSING

We know little about how the germ cells process precursors into mature piRNAs, though working models propose that two distinct mechanisms generate the piRNAs by primary biogenesis and secondary Ping-Pong amplification (Brennecke et al., 2007; Gunawardane et al., 2007). The primary pathway exists in both somatic follicle cells, where only one PIWI protein, Piwi is expressed; and germline nurse cells, where all three *Drosophila* PIWI proteins are produced. Since secondary piRNA production requires two germline specific PIWI proteins, Aub and Ago3, which are not detectable in the follicle cells, it seems that the secondary pathway only functions in the germline nurse cells (Brennecke et al., 2007; Li et al., 2009; Malone et al., 2009). Due to the simplicity and existing cell

culture system, the primary pathway has been more extensively studied in the follicle cells. It is unclear if mechanisms uncovered in these cells can be applied to the germline primary pathway, since the germline cells are dominated by dual-strand piRNA clusters while somatic lineage follicle cells and cultured cells derived from them are dominated by uni-strand clusters.

The primary pathway initiates piRNA production de novo, without pre-existing piRNAs. In the somatic follicle cells, the piRNA precursors appear to be delivered to the perinuclear Yb body, which contains the Tudor domain protein Yb and Vret, the putative RNA helicase Armitage (Armi) and the co-chaperone protein Shutdown (Shu) (Saito et al., 2010; Handler et al., 2011; Olivieri et al., 2012; Zamparini et al., 2011). Based on their genetic dependency, the following hierarchy order has been proposed: Yb -> Armi -> Vret -> Shu (Figure 1.1 and 1.3). Loss of any of these components results in Piwi mis-localization and blocking of piRNA production (Saito et al., 2010; Handler et al., 2011; Olivieri et al., 2012; Zamparini et al., 2011). We do not know which nuclease generates the 5′ end of the piRNAs. A candidate is Zucchini, a mitochondrial protein expressed in both germline and somatic cells (Pane et al., 2007). This protein comprises both a putative nuclease and phospholipase domains. Recent studies indicate that both mammalian Zuc (also known as PLD6 in mice) and *Drosophila* Zuc have single-stranded RNA or DNA cleavage activity in vitro (Nishimasu et al., 2012; Ipsaro et al., 2012). The crystallized Zuc structure illustrates that Zuc monomers dimerized to form a long, narrow, positivity charged catalytic groove, which potentially binds and cleaves precursor to generate the 5′ end of the piRNAs (Nishimasu et al., 2012; Ipsaro et al., 2012). Whether Zuc is the bona fide piRNA 5′ end cleavage enzyme needs further testing. It is also unclear how Zuc,

Figure 1.3

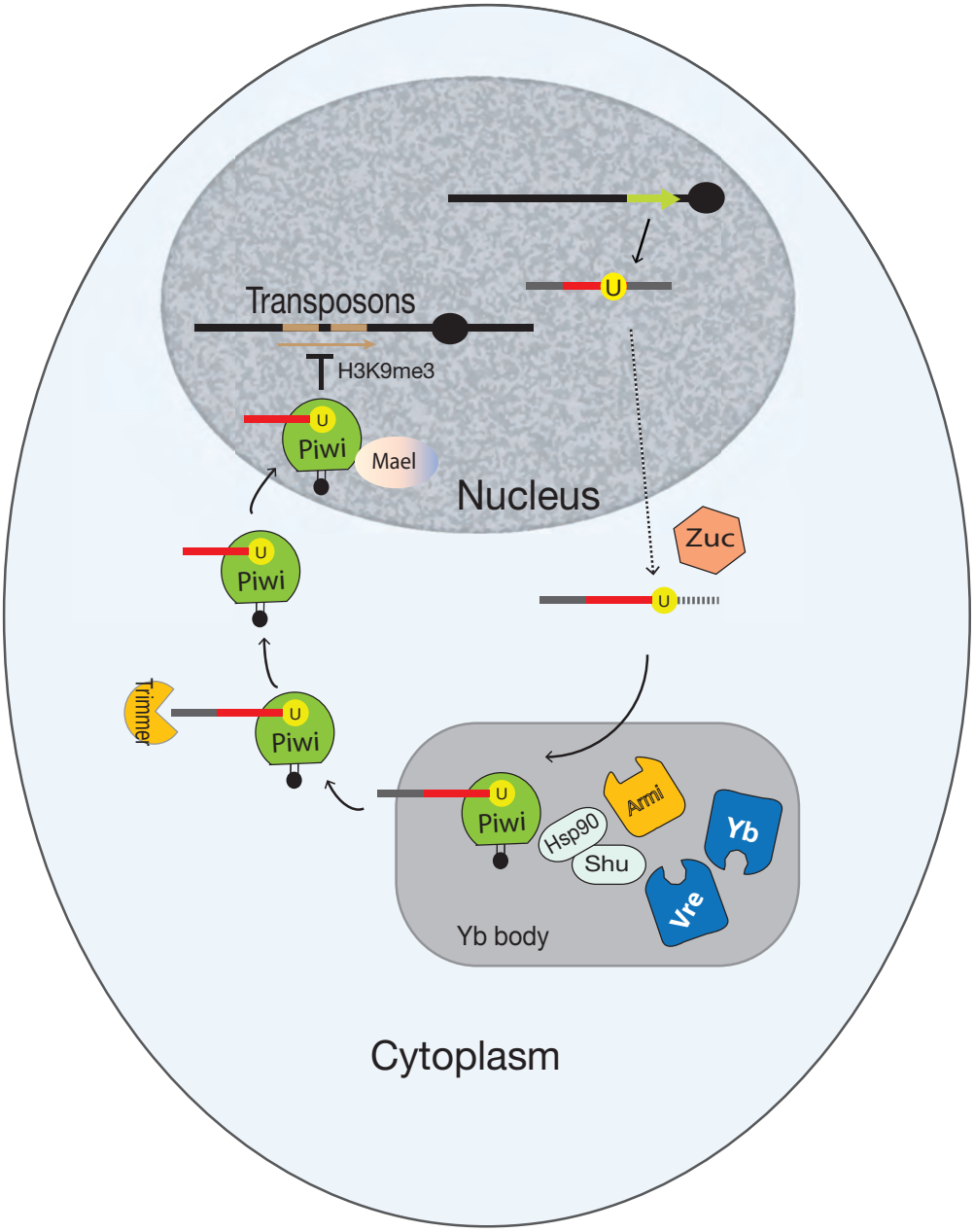
Transposons
H3K9me3
U
Piwi
Mael
Nucleus
Zuc
Trimmer
Hsp90
Shu
Ami
Yb
Vre
Yb body
Cytoplasm

Figure 1.3. piRNA production and transposon silencing in follicle cell.

localized to mitochondria, is coupled with the piRNA precursor processing machinery, either the Yb body in the follicle cell or the germline cell nuage.

After the 5′ cleavage, the resulting piRNA intermediates appear to be loaded into PIWI proteins. This loading step is highly selective, for both primary and secondary piRNA biogenesis. In *Drosophila*, Aub and Piwi enriched piRNAs begin with uridine and are often complementary to transposon mRNAs. In contrast, Ago3 preferentially associates with the sense piRNAs and often bear adenosine at the 10th position (Figure 1.1; Brennecke et al., 2007). How this loading specificity is generated remains unknown. Like other small RNAs that are loaded into Argonaute proteins, charging PIWI proteins with piRNAs is facilitated by chaperone proteins. In the somatic follicle cells, the loading is likely performed in the Yb bodies (Olivieri et al., 2010; Saito et al., 2010). One of the Yb body components, Shu, is a co-chaperone protein that interacts with the Hsp90 (encoded by the gene *hsp83* in *Drosophila*) through its tetratricopeptide repeat (TPR) domain. Breaking this interaction by mutating the conserved key amino acids in the TPR region leads to the transposon hyper-expression and fly sterility (Preall et al., 2012; Olivieri et al., 2012). As a conserved protein, its mouse homologue, FKBP6, has been shown interact with PIWI proteins and is required for transposon silencing and mouse fertility. Mouse FKBP6 appears to specifically facilitate Miwi2 piRNA loading (Xiol et al., 2012). It is unclear if other co-chaperones are required to load other PIWI proteins. Interestingly, in the silkworm ovary-derived BmN4 cells, adding Hsp90 specific inhibitor 17-AGG decreases piRNA production. Moreover, in vitro studies showed that blocking Hsp90 function with this drug compromises PIWI protein loading specificity (Izumi et al., 2013).

The 3′ end formation completes piRNA biogenesis, yet we still do not know which enzyme performs this function. This may be done by a "sloppy" nuclease, since piRNAs often show higher heterogeneity at 3′ ends than 5′ ends (Brennecke et al., 2007). In vitro studies showed that loading a 50 nt long RNA into silkworm PIWI protein Siwi produces a ~27 nt mature piRNA (Kawaoka et al., 2011b). This step appears to be catalyzed by a 3′ to 5′ exonuclease, "trimmer", in a Mg^{2+} dependent manner (Figure 1.1 and 1.4; Kawaoka et al., 2011b). Interestingly, it seems that the Tudor domain protein, PAPI, also assists this step. Depleting PAPI in the BmN4 cells leads to 3′ terminal extensions of mature piRNAs (Honda et al., 2013). The trimming of the 3′ end is coupled with 2′-*O*-methylation, executed by the methyltransferase Hua (花, the Chinese character for "flower") enhancer 1, Hen1 (Figure 1.1 and 1.4; Saito et al., 2007; Horwich et al., 2007). This end modification appears to be present in essentially all animal species, and appears to protect mature piRNA from further trimming or non-templated tailing.

In the germline of flies, zebrafish, silkworms and mice, primary piRNAs can engage in an amplification cycle to enlarge the piRNA population by generating secondary piRNAs (Brennecke et al., 2007; Houwing et al., 2007; Aravin et al, 2007; Kawaoka et al, 2009). We call this secondary piRNA pathway "Ping-Pong amplification" (Figure 1.1 and 1.4). In *Drosophila* nurse cells, this Ping-Pong cycle is mainly mediated by Aub and Ago3, the two germline specific PIWI proteins that reside in nuage (Gunawardane et al., 2007; Brennecke et al., 2007; Li et al., 2009). The Ping-Pong model proposes that Aub, guided by a tightly associated antisense piRNA, pairs and cleaves a complementary

transposon mRNA, simultaneously destroying the transposon transcript and generating a new sense piRNA intermediate loaded into Ago3, possibly by the same loading machinery as primary piRNA. Next, the trimmer shortens the 3' end and Hen1 finalizes maturation by 2'-*O*-methylation of the 3' end (Saito et al., 2007; Horwich et al., 2007). In turn, this newly generated sense piRNA guides Ago3 to bind and cleave an antisense piRNA precursor transcript, generating a new Aub bound antisense piRNA (Figure 1.1 and 1.4). This model is strongly supported by the high probability of a 10 nt overlap between the Ago3-bound sense piRNA and Aub-bound antisense piRNA (Brennecke et al., 2007). As predicted by this model, in vivo studies showed that removing Ago3 leads to the collapse of the overall piRNA pool (Li et al., 2009). This heterotypic Ping-Pong between Aub and Ago3 generates more antisense piRNAs than sense piRNAs. Each of these proteins can amplify piRNAs on its own, yet this type of single player Ping-Pong produces equal amounts of sense and antisense piRNAs (Brennecke et al., 2007; Li et al., 2009). A key question is how Ping-Pong is largely restricted to a two-player game in which Ago3 partners with Aub. We have demonstrated that a Tudor domain protein Qin, which has both an E3 ligase domain and 5 Tudor domains, is required to maintain this heterotypic Ping-Pong. I will detail this story in chapters 2 and 3.

piRNA MEDIATED TRANSPOSON SILENCING

Transposons and repetitive elements occupy nearly half of the human genome and ~30% of the fly genome. Disrupting the piRNA pathway in the animal gonad results in transposon activation, DNA breaks and compromised fertility. In animals like worm, fly, silkworm, zebrafish and mouse, either most or some

Figure 1.4

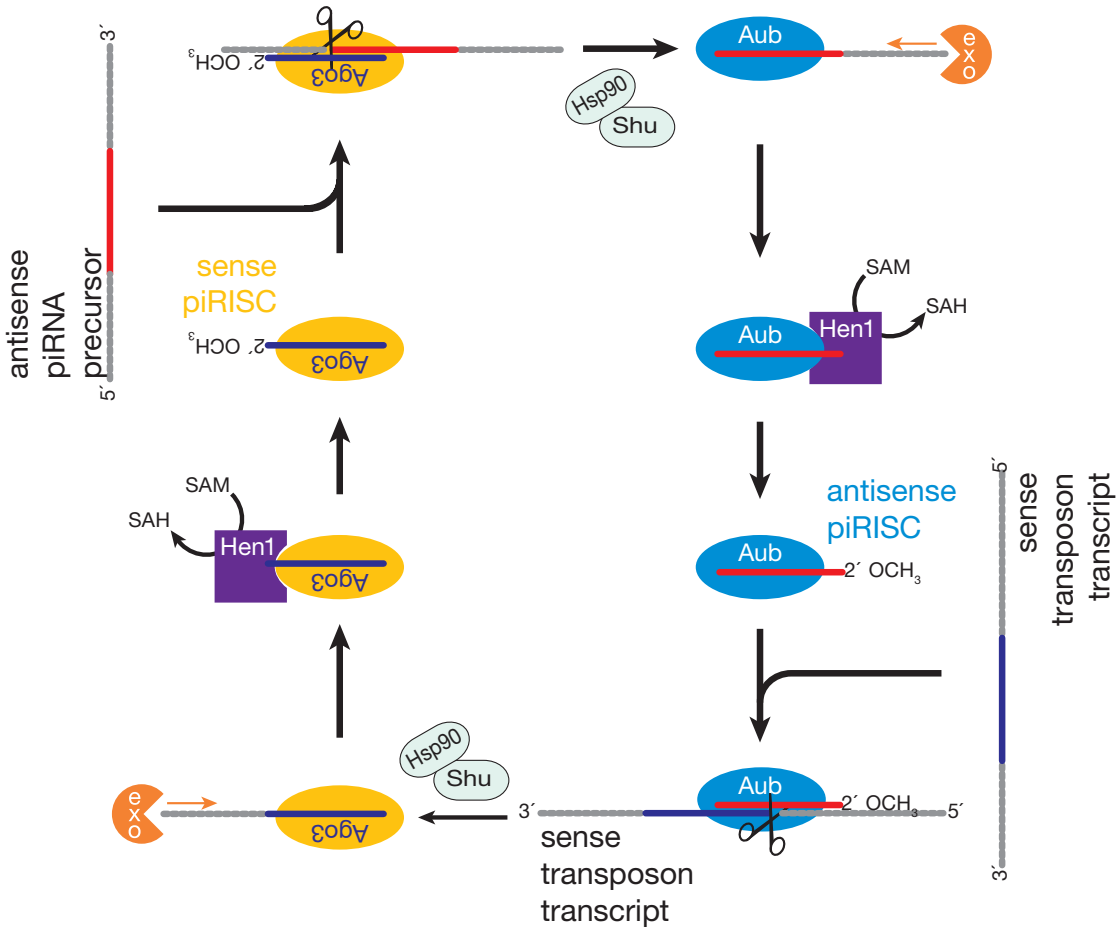
antisense
piRNA
precursor
3´
5´
Ago3
2´ OCH3
sense
piRISC
Hsp90
Shu
Aub
exo
SAM
SAH
Hen1
antisense
piRISC
2´ OCH3
sense
transposon
transcript
5´
3´

Figure 1.4. Ping-Pong model.

piRNAs share sequence similarity with transposon transcripts (Batista et al., 2008; Houwing et al., 2007; Brennecke et al., 2007; Kawaoka et al., 2009; Kawaoka et al., 2011a). Therefore, the transposon silencing function of piRNA pathway appears to be conserved. It seems evident that the piRNA pathway silences transposons at two levels: transcriptional and post-transcriptional (Figure 1.1).

Based on the Ping-Pong model, both Aub and Ago3 posses the cleavage/slicer activity to amplify the piRNA pool and simultaneously post-transcriptionally destroy transposon transcripts (Gunawardane et al., 2007; Brennecke et al., 2007). Specifically, antisense piRNAs guides Aub to recognize and cleave transposon mRNAs and antisense piRNAs that could mediate post-transcriptional silencing are more abundant than sense piRNAs (Figure 1.4). Animals may therefore employ a specific mechanism to make sure the real "silencers", antisense piRNAs, dominate. Interestingly, in chapters 2 and 3, I will show that mutations that distort this dominance are associated with transposon activation. However, there are several key questions that still need to be addressed before we can really understand this model. First, what is the mechanism to ensure that Ago3 is loaded with sense piRNAs while Aub is enriched for the antisense ones? Second, how do piRNAs associated with PIWI protein pair with targets? siRNAs and microRNAs use the first 2-9 nucleotides, the "seed sequence", to recognize their targets. Do piRNAs have seed sequences? Can mismatches be tolerated? It is worth noting that in vitro studies have shown that one of the mouse PIWI protein, Miwi, can tolerate mismatches at the 3' end starting from the 22nd nt, but not mismatches at the 5' end, except the first nucleotide (Reuter et al., 2011). Whether the slicer activity of Aub and Ago3 is indeed needed for transposon silencing in vivo in flies is still not clear, but we do

know that the catalytically dead version of two mouse PIWI proteins, Mili and Miwi, fail to suppress the non-LTR transposon LINE-1 in mouse testes (Reuter et al., 2011; De Fazio et al., 2011).

Piwi is the only PIWI protein that resides in the nucleus in *Drosophila* (Cox et al., 2000; Brennecke et al., 2007). Although it exhibits slicer activity in vitro, it seems this activity is dispensable in vivo, at least for transposon suppression and fly fertility (Sienski et al., 2012; Le Thomas et al., 2013). However, the nuclear localization is essential for Piwi function. A truncated Piwi protein, missing only the nuclear localization signal, could not be imported into nucleus. This trauncated Piwi can still bind piRNAs, however fails to silence the transposons and leads to fly sterility (Klenov et al., 2011; Sienski et al., 2012; Le Thomas et al., 2013). Hence, Piwi may silence transposons in the nucleus at the transcriptional level, possibly similar to how yeast small RNAs induce heterochromatin formation. Indeed, a number of recent studies have demonstrated that Piwi suppresses several transposons in the nucleus through triggering H3K9me3 formation, repressing transposons epigenetically. First, it has been proposed that Piwi may directly interact with the heterochromatin marker protein HP1a (Yin and Lin, 2007). Second, germline RNAi against Piwi or mutating the Piwi nuclear localization signal decreases H3K9me3 and HP1 occupancy at genomic transposon regions, as assayed by ChIP-qPCR (Wang and Elgin, 2011; Klenov et al., 2011). Finally, the recent work from Brennecke and colleagues showed on a genomic scale that depleting Piwi in the somatic follicle cells globally increases Pol-II occupancy and nascent mRNA transcription of transposons (Sienski et al., 2012). Accordingly, the H3K9me3 coating at these regions drops. Interestingly, they also reported that another germline piRNA factor, Maelstrom (Mael),

actively promotes this process (Sienski et al., 2012). Upon Mael knockdown, there is no change in piRNA production, but a handful of transposons become more active. Closer examination revealed that Pol-II increases its occupancy on these elements. However, the H3K9me3 is still maintained at similar levels as the control knockdown, but becomes more broadly spread at downstream regions (Sienski et al., 2012). This suggests that Mael is downstream of Piwi induced heterochromatin formation, possibly more engaged in the silencing step. Soon after this study, several other labs reported that Piwi also triggers H3K9me3 formation on a subgroup of transposon families in the germline nurse cells (Huang et al., 2013; Rozhkov et al., 2013; Le Thomas et al., 2013). So far, it is not clear whether the mammalian piRNA pathway also silences transposons by employing H3K9me3 coating. However, the piRNA pathway does transcriptionally inhibit transposon activity by mediating the DNA methylation at target loci in prenatal stage mouse testes.

THE DIVERSIFIED FUNCTIONS OF piRNA PATHWAY

Other than the transposon silencing function, the piRNA pathway/PIWI protein also plays other roles in both germline and somatic tissues. The *Drosophila* Piwi protein was originally identified as a factor to maintain the germline stem cells (Cox et al., 1998; Cox et al., 2000). Fly oogenesis starts from germline stem cells that localize to the anterior end of the germarium, surrounded by the somatic lineage niche cells. After each cell cycle, the asymmetric cell division generates two daughter cells: the one closer to the niche maintains the stemness; the other one undergoes differentiation and finally forms the oocyte. Because the loss of *piwi* gives rudimentary ovaries lacking developing oocytes, it was originally

reported as a gene required for stem cell maintenance (Cox et al., 1998; Cox et al., 2000). However, Lai and colleagues recently re-evaluated the *piwi* mutants and based on their quantitative and careful examination, it seems that mutating *piwi* leads to a germline stem cell differentiation defect (Jin et al., 2013). Instead of losing germline stem cells, too many undifferentiated stem cells accumulate in the germarium, displaying the "germline-stem-cell-like tumor" phenotype in *piwi* mutants (Jin et al., 2013). Since disrupting other piRNA pathway factors does not result in a similar oogenesis defect, it suggests that Piwi uniquely "has gained" this extra function besides transposon silencing. It is unclear whether this function requires Piwi to be loaded by piRNAs.

In *Drosophila*, although around 80% of piRNAs map to transposons and piRNA clusters, a subset of the piRNAs is derived from protein coding genes, mainly from 3' UTR regions (Robine et al., 2009; Lau et al., 2009b; Saito et al., 2009). Therefore, piRNAs could target mRNAs for silencing. For examples, the first discovered piRNAs, *Su[Ste]* piRNAs, suppress *Ste* mRNAs in fly testes (Aravin et al., 2001); piRNAs generated from the 3'UTR of *traffic jam* (*tj*) likely silence *FasIII* mRNAs (Saito et al., 2009); in fly embryo, Aub might mediate *nanos* mRNA deadenylation by binding with its 3' UTR through imperfectly paired piRNAs (Rouget et al., 2010). Interestingly, these 3'UTR derived genic piRNAs are also found in mouse testes and *Xenopus* eggs, suggesting they may be produced by a conserved mechanism (Robine et al., 2009; Lau et al., 2009a). The targets for the majority of these genic piRNAs, however, are still unkown.

In worms, the piRNA pathway also serves as a memory source to induce the epigenetic shutdown of nonself RNA (Lee et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). The mismatch(es) between piRNAs (21U RNA) and their

nonself targets recruits the RNA dependent RNA polymerase (RdRP). Using piRNA targets as the template, RdRP synthesizes another class of small RNAs (22G RNAs). These 22G RNAs are then loaded into the worm-specific protein, WAGO-9/HRDE-1, which then enters nucleus (Lee et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). Assisted by other worm nuclear RNAi factors and guided by 22G RNAs, WAGO-9 epigenetically silences target transcription, potentially by trigging the formation of H3K9me3. Interestingly, this silencing state can be stably inherited across multiple generations (Lee et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). This phenomenon is named RNA-induced epigenetic silencing (RNAe).

Besides its functions in the germline, the piRNA pathway has been reported to be active in neurons of at least *Drosophila* and the sea hare, *Aplysia*. In *Aplysia*, Piwi/piRNA complex likely promotes CpG methylation at the CREB2 promoter region in neurons (Rajasethupathy et al., 2012). It seems Aub and Ago3 are also expressed in fly brain. Loss of either protein in neurons leads to transposon overexpression (Perrat et al., 2013). In addition, it has been reported that some of the piRNA pathway genes are ectopically expressed in certain types of somatic tissues derived solid tumors. Interestingly, mutating these genes can inhibit tumor growth (Janic et al., 2010). How these genes drive cell over-proliferation in tumors has not been determined. Finally, in the unicellular organism, *Oxytricha*, the piRNAs can protect DNA against deletions during the genome rearrangement (Fang et al., 2012).

CHAPTER II

HETEROTYPIC piRNA PING-PONG REQUIRES QIN, A PROTEIN WITH BOTH E3 LIGASE AND TUDOR DOMAINS

PREFACE

The work presented in this chapter was a collaborative effort: Jia Xu performed computational analysis of small RNA sequencing data. Birgit Koppetsch did northern blot and confocal imaging. Jie Wang analyzed the tiling microarray data. Cindy performed part of the fertility test. I performed the rest of the experiments.

SUMMARY

piRNAs guide PIWI proteins to silence transposons in animal germ cells. Reciprocal cycles of piRNA-directed RNA cleavage—catalyzed by the PIWI proteins Aubergine (Aub) and Argonaute3 (Ago3) in *Drosophila melanogaster*—expand the population of antisense piRNAs in response to transposon expression, a process called the Ping-Pong cycle. Heterotypic Ping-Pong between Aub and Ago3 ensures that antisense piRNAs predominate. We show that *qin*, a piRNA pathway gene whose protein product contains both E3 ligase and Tudor domains, co-localizes with Aub and Ago3 in nuage, a perinuclear structure implicated in transposon silencing. In *qin* mutants, less Ago3 binds Aub, futile Aub:Aub homotypic Ping-Pong prevails, antisense piRNAs decrease, many families of mobile genetic elements are reactivated, and DNA damage accumulates in nurse cells and oocytes. We propose that Qin enforces heterotypic Ping-Pong between Aub and Ago3, ensuring that transposons are silenced and maintaining the integrity of the germline genome.

INTRODUCTION

In animals, PIWI-interacting RNAs (piRNAs) silence transposons and other repetitive elements (Klattenhoff and Theurkauf, 2008; Ghildiyal and Zamore, 2009; Siomi et al., 2011). For flies, mammals, and other bilateral animals, the piRNA pathway protects the germline genome from DNA damage and mutation, ensuring that genetic information passes faithfully from generation to generation.

piRNAs are typically 23–30 nucleotides (nt) long and bind to members of the PIWI clade of Argonaute proteins, a family that includes the proteins that mediate RNA interference and microRNA-directed gene regulation. The *Drosophila* PIWI clade comprises three proteins: Argonaute3 (Ago3), Aubergine (Aub) and Piwi. In the germline nurse cells, which support development of the oocyte, Ago3 and Aub reside in a perinuclear, cytoplasmic structure called "nuage"; Piwi resides in the nuclei of ovary germ cells and the somatic follicle cells that surround the germ cells (Wilson et al., 1996; Cox et al., 1998; Cox et al., 2000; Saito et al., 2006; Gunawardane et al., 2007; Brennecke et al., 2007).

In fly germ cells, primary piRNAs are thought to derive from discrete genomic loci, "piRNA clusters," that contain complex arrays of transposon sequences (Brennecke et al., 2007). Reciprocal cycles of Aub- and Ago3-mediated RNA cleavage are thought to increase piRNA abundance by producing secondary piRNAs. Secondary piRNAs are disproportionately antisense to germline-expressed transposons (Brennecke et al., 2007; Houwing et al., 2007; Lau et al., 2006; Grivna et al., 2006; Aravin et al., 2006; Girard et al., 2006; Vagin et al., 2006). Amplification of antisense piRNAs requires both Aub and Ago3, and piRNA production in the germline collapses in *ago3* mutants (Brennecke et al., 2007; Li et al., 2009; Malone et al., 2009). The "Ping-Pong" model for piRNA amplification (Brennecke et al., 2007; Gunawardane et al., 2007) proposes that

Aub, guided by an antisense primary piRNA, binds to and cleaves a complementary transposon mRNA, simultaneously destroying the transposon transcript and generating a new, sense piRNA that is loaded into Ago3. The model envisions that sense secondary piRNA direct Ago3 to cleave an antisense sequence in the original piRNA precursor transcript to generate another Aub-bound antisense piRNA (Brennecke et al., 2007; Li et al., 2009; Malone et al., 2009). Heterotypic Ping-Pong between Aub and Ago3 produces more antisense piRNAs than sense, but the molecular mechanisms that bias the Ping-Pong cycle towards antisense remain unknown. Strikingly, the absence of Ago3 results in a futile, homotypic Ping-Pong cycle that generates more sense than antisense piRNAs, both of which bind Aub. Homotypic Aub:Aub Ping-Pong fails to silence all transposons, and *ago3* mutant females are infertile (Li et al., 2009).

The fly genome encodes 23 predicted Tudor domain proteins, seven of which have been shown to act in the piRNA pathway (Lim and Kai, 2007; Siomi et al., 2010; Patil and Kai, 2010; Liu et al., 2011; Zamparini et al., 2011). The binding of Tudor-domain proteins to di-methylarginine-modified PIWI proteins is conserved from flies to mammals, but its molecular function remains unknown (Kirino et al., 2009; Nishida et al., 2009; Wang et al., 2009a; Vagin et al., 2009; Siomi et al., 2010). Here, we identify a *Drosophila* Tudor-domain protein, Qin, which is required for piRNA production in the germline. The *qin* gene encodes an unusual protein with an amino-terminal E3 ligase domain and five carboxy-terminal Tudor domains. In *qin* mutants, futile homotypic Aub:Aub Ping-Pong replaces heterotypic Ping-Pong between Aub and Ago3, activating transposon expression, and triggering DNA damage in the nurse cells and the oocyte. Thus, *qin* mutants phenocopy *ago3* mutants, although the abundance of Ago3 is greater than that in *ago3* heterozygotes. Qin localizes to nuage and appears to be required for the physical interaction of Aub with Ago3.

RESULTS

qin Encodes a Protein with Five Tudor and One E3 Ligase Domain

FlyBase (Drysdale et al., 2005) predicts that the putative transcription unit *CG14303* encodes a protein with five Tudor domains (Figures 2.1A and 2.S1). Disruption of *CG14303* by insertion of a piggyBac transposon ((RB)PBacCG14303e03728; (Thibault et al., 2004)) in the largest predicted exon caused female sterility and produced embryos with fused or missing dorsal appendages (Table S1), phenotypes associated with a failure of the piRNA pathway (Schupbach and Wieschaus, 1989; Schupbach and Wieschaus, 1991; Wilson et al., 1996; Chen et al., 2007; Klattenhoff et al., 2007). Because *CG14303* mutant females produce offspring that fail to develop into adults, we named the gene *qin*, after the ancient Chinese dynasty (秦) that ended after just two generations.

PBac(RB)CG14303e03728 (qin^1) homozygotes produced few eggs, all of which display dorsal appendage defects (Table 2.S1). Similarly, only 4.3% of embryos from qin^1/*Df(3R)Excel6180* females had wild-type appendages (Table 2.S1 and Figures 2.1A and 2.S2), suggesting that mutation of *qin* causes the observed phenotypes. The deficiency *Df(3R)Excel6180*, henceforth *Df*, removes *qin* as well as other genes (Figure 2.S2; (Parks et al., 2004)). Although qin^1 is predicted to encode a truncated protein, qin^1 behaved like a genetic null allele: the dorsal appendage phenotype, egg hatching, and eggs laid per day per female were as severe or worse in qin^1 homozygotes compared to qin^1 / *Df* (Table 2.S1).

To further verify that these defects resulted from mutation of *qin*, we attempted to rescue qin^1 / *Df* by creating a transgenic fly expressing genomic fragment CH322-81J04 (henceforth, gf-1; (Venken et al., 2009)), which encompasses *CG14303* (Figure 2.1A). The gf-1 transgene failed to rescue the

Figure 2.1

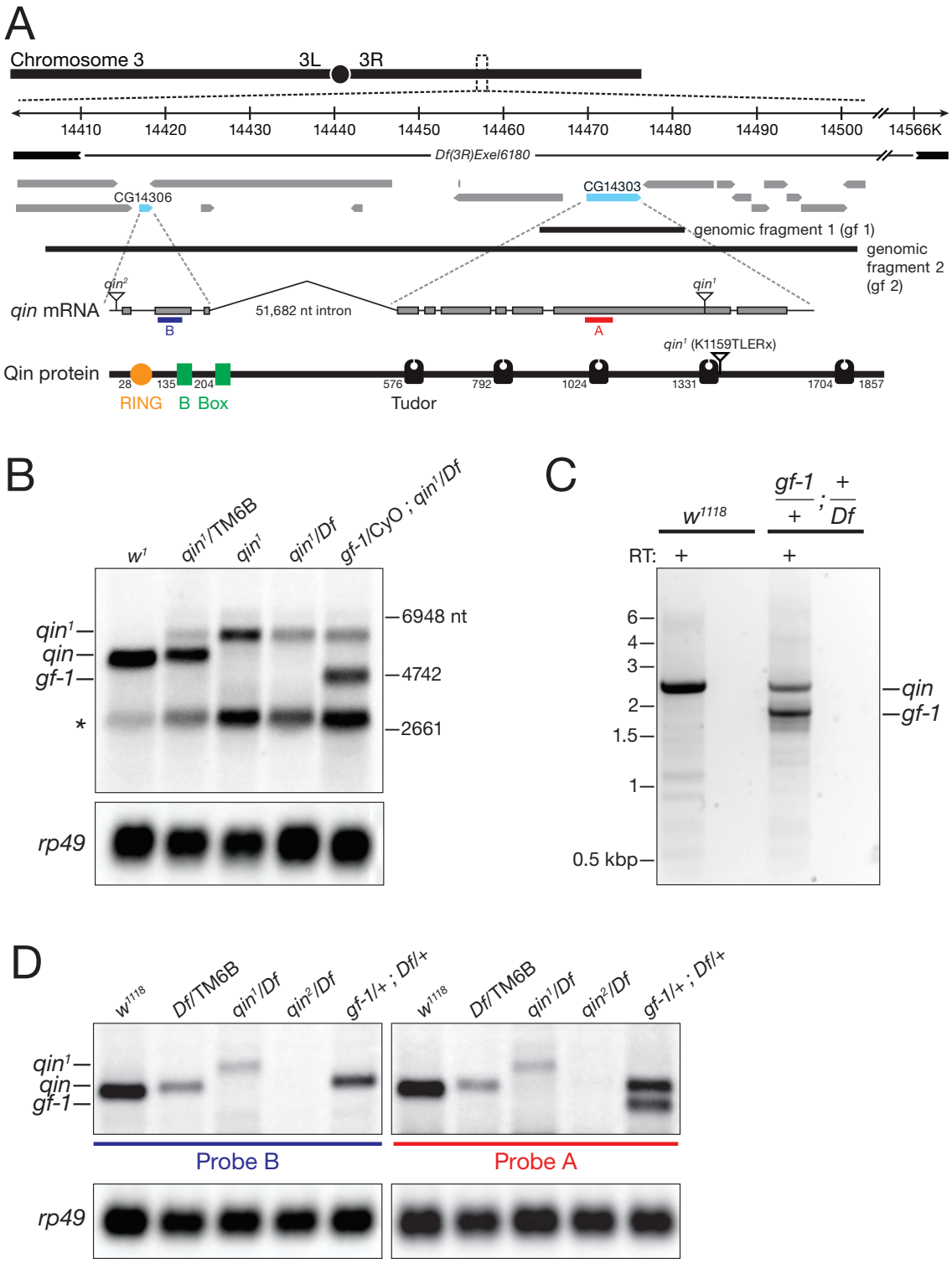
A
Chromosome 3
3L
3R
14410
14420
14430
14440
14450
14460
14470
14480
14490
14500
14566K
Df(3R)Exel6180
CG14306
CG14303
genomic fragment 1 (gf 1)
genomic fragment 2 (gf 2)
qin2
qin1
qin mRNA
51,682 nt intron
B
A
qin1 (K1159TLERx)
Qin protein
28
135
204
576
792
1024
1331
1704
1857
RING
B Box
Tudor
B
w1
qin1/TM6B
qin1
qin1/Df
gf-1/CyO ; qin1/Df
qin1
qin
gf-1
*
6948 nt
4742
2661
rp49
C
w1118
gf-1 / + ; + / Df
RT: + +
6
4
3
2
1.5
1
0.5 kbp
qin
gf-1
D
w1118
Df/TM6B
qin1/Df
qin2/Df
gf-1/+ ; Df/+
qin1
qin
gf-1
Probe B
Probe A
rp49

Figure 2.1. The *qin* Gene and Qin protein.

(A) *qin* resides on the right arm of chromosome 3, comprising two predicted genes separated by a 51,682 bp intron. *qin* encodes a protein with E3 ligase and Tudor domains. The blue and red bars indicate the location of probes used in Northern hybridization analyses.

(B) Northern analysis of *qin* transcripts in ovary RNA using the red probe. Asterisk marks an unidentified transcript.

(C) Agarose gel analysis of the 5′ rapid amplification of cDNA ends (RACE) products from control and *gf-1*/+ ; *Df*/+ ovaries.

(D) Ovary RNA analyzed in parallel by Northern hybridization using the red and blue probes.

defects in *qin*^{1}/*Df* ovaries, eggs, and embryos (Table 2.S1 and Figure 2.2A). Moreover, the transgene carrying gf-1 generated a transcript that was shorter than wild-type *qin* mRNA (Figure 2.1B). To understand why gf-1 failed to rescue, we mapped the 5′ end of the *qin* mRNA by rapid amplification of cDNA ends (5′ RACE). The 5′ end of the *qin* mRNA mapped 51,682 bp upstream of *CG14303*, in the putative gene *CG14306*, which is predicted to encode a protein with a RING domain and two B-Box domains, the hallmarks of a Ubiquitin or SUMO E3 ligase (Figures 2.1A, 2.1C, and 2.S1).

In control (*w*^{1118}) ovaries, a Northern hybridization probe for *CG14306* detected the same size mRNA as a probe for *CG14303* (Figure 2.1D). Probes for *CG14306* and *CG14303* both detected a longer RNA in *qin*^{1} whose size is consistent with transcription of three exons in *CG14306* and five exons plus part of a sixth in *CG14303* fused to PBac(RB)*CG14303e03728*, the piggyBac insertion that creates the *qin*^{1} allele. Neither hybridization probe detected the *qin* mRNA in ovaries from females bearing both the deficiency and PBac(RB)*e01936* (*qin*^{2}), a piggyBac insertion 5′ to the first predicted exon of *CG14306* (Figures 2.1D and 2.S3A). Consistent with *qin* encompassing both *CG14306* and *CG14303*, *qin*^{2} failed to complement *qin*^{1} (Figure 2.S3B). Moreover, a cDNA comprising the predicted exons of *CG14306* and *CG14303* (UASp-(FLAG)$_3$-(Myc)$_6$-Qin, henceforth FM-Qin) and driven from a UASp promoter by *nanos*-Gal4-VP16, partially rescued the female sterility and dorsal appendage (Table 2.S1) and transposon silencing defects associated with loss of Qin (Figures 2.2A and 2.2B). The incomplete rescue may reflect the poor expression of *nanos*-Gal4-VP16 in stage 2–6 egg chambers (Figure 2.S3A; Ni et al., 2011) or an effect of the amino-terminal FLAG-Myc tag on Qin function or stability.

Figure 2.2

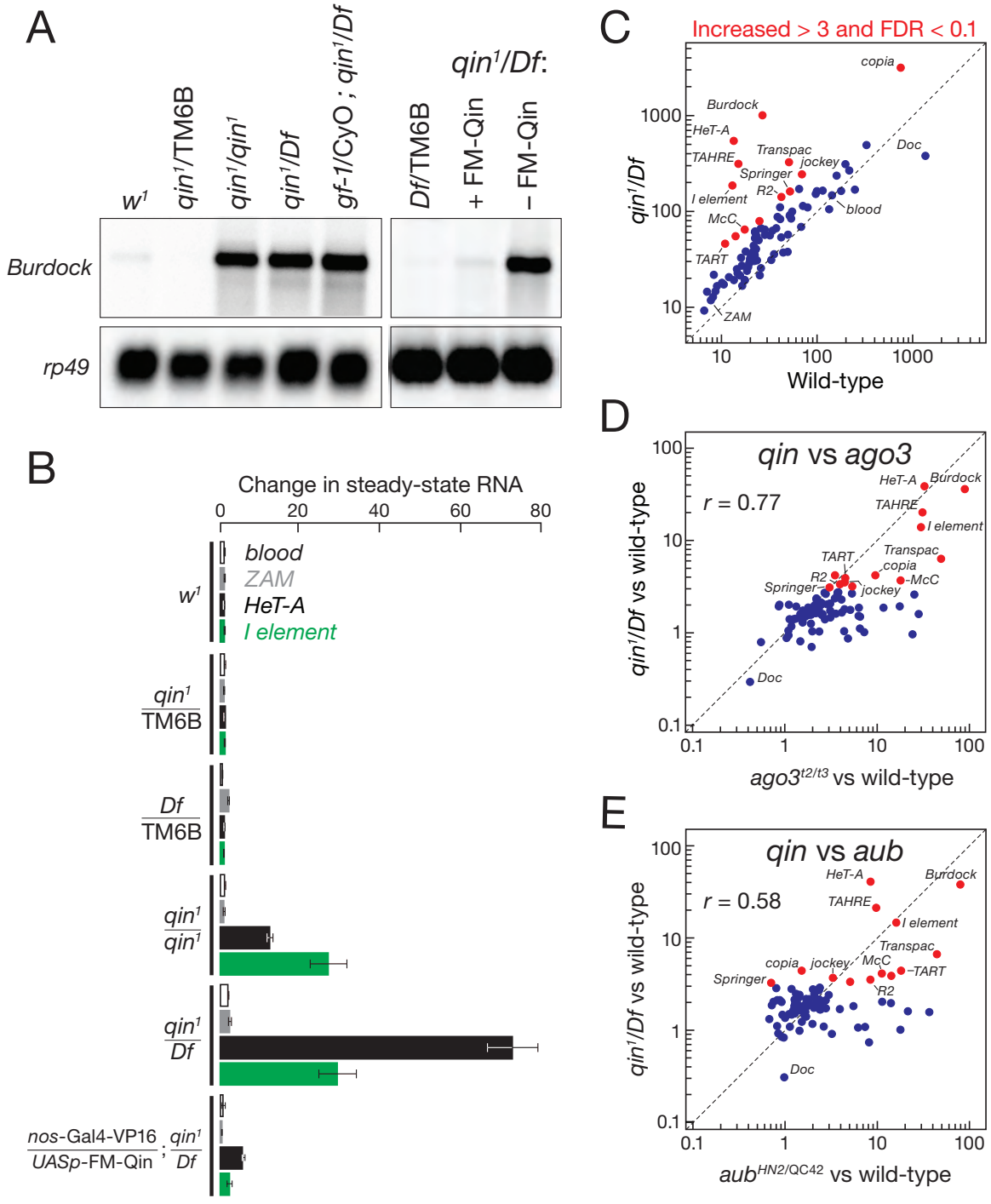
A
qin1/Df:
w1
qin1/TM6B
qin1/qin1
qin1/Df
gf-1/CyO ; qin1/Df
Df/TM6B
+ FM-Qin
– FM-Qin
Burdock
rp49
B
Change in steady-state RNA
0
20
40
60
80
blood
ZAM
HeT-A
I element
w1
qin1/TM6B
Df/TM6B
qin1/qin1
qin1/Df
nos-Gal4-VP16/UASp-FM-Qin ; qin1/Df
C
Increased > 3 and FDR < 0.1
copia
Burdock
HeT-A
TAHRE
Transpac
jockey
Springer
R2
I element
blood
McC
TART
ZAM
Doc
1000
100
10
qin1/Df
Wild-type
D
qin vs ago3
r = 0.77
HeT-A
Burdock
TAHRE
I element
TART
Transpac
copia
R2
McC
Springer
jockey
Doc
100
10
1
0.1
qin1/Df vs wild-type
ago3t2/t3 vs wild-type
E
qin vs aub
r = 0.58
HeT-A
Burdock
TAHRE
I element
Transpac
McC
copia
jockey
TART
Springer
R2
Doc
100
10
1
0.1
qin1/Df vs wild-type
aubHN2/QC42 vs wild-type

Figure 2.2. Qin Silences Transposons in *Drosophila* Ovaries.

(A) Northern hybridization for *Burdock* mRNA.

(B) Quantitative RT-PCR was used to measure transposon expression. Data are normalized to *rp49* (*RpL32*) expression. The bars report mean ± standard deviation for three biological replicates.

(C) Expression of transposons measured using whole genome tiling arrays for qin^1/Df ovaries, relative to control (w^1). The dashed line indicates equivalent expression in the two genotypes. The 13 transposon families whose expression increased > 3-fold among 3 biological replicates (false discovery rate, FDR, < 0.1) in *qin* mutants, compared to control, are marked in red.

(D) The change in transposon expression, relative to control, was compared for *qin* and *ago3* mutant ovaries.

(E) The change in transposon expression, relative to control, was compared for *qin* and *aub* mutant ovaries.

As a final test of the idea that the *qin* locus comprises both *CG14303* and *CG14306*, we constructed a transgenic fly bearing a 97,532 bp genomic fragment (gf-2, CH321-88L14) (Venken et al., 2009) encompassing all of the *qin* exons, as well as 6,602 bp 5′ to exon 1 and 32,343 bp 3′ to exon 10 (Figures 2.1A and 2.S3B). To facilitate detection of the predicted Qin protein encoded by gf-2, we inserted the EGFP coding sequence before the *qin* stop codon of the genomic fragment. Gf-2 includes part or all of the *qin* promoter, because transgenic flies bearing gf-2 expressed the Qin::EGFP fusion in the germline throughout oogenesis (Figure 2.S3C). A single copy of gf-2 partially rescued the dorsal appendage defects, low egg hatch rate, and the low number of eggs laid per day per fly observed for *w*; *Sp*/+; qin^1/*Df* females (Table 2.S1). We conclude that the *qin* locus comprises the three exons of *CG14306* and the seven exons of *CG14303* joined by the removal of a 51,682 nt intron and that the Qin protein contains an amino-terminal E3 ligase domain and five carboxy-terminal Tudor domains (Figures 2.1A and 2.S1).

Qin is Required to Protect the Germline Genome and Silence Transposons

Defects in the piRNA pathway cause double-stranded breaks in the germline genome, and γH2Av, a phosphorylated histone variant, accumulates at the sites of DNA damage (Chen et al., 2007; Klattenhoff et al., 2007). Homozygous qin^1, qin^1/*Df*, and qin^2/*Df* mutant ovaries showed many nuclear γH2Av foci (Figures 2.3A and 2.S4), unlike control ovaries, which displayed only the expected small number of γH2Av foci likely to be associated with normal endoreduplication (Figure 2.3A) or meiotic recombination (Figure 2.S4). The increased number and intensity of γH2Av foci in *qin* mutants suggest that Qin is required for transposon silencing. We used whole-genome tiling microarrays to survey the

expression of transposons and mRNAs in *qin* mutant ovaries. In *qin* mutants, 56 of 93 transposon families showed a 1.5- to 3-fold increase in steady-state transcript abundance (Figure 2.2C). The expression of another 13 transposon families increased more than 3-fold (False Discovery Rate, FDR, < 0.1). In contrast, the expression of none of the 19,987 protein coding genes detected by the microarray, including 613 annotated heterochromatic genes (Smith et al., 2007), increased or decreased >3-fold at FDR < 0.1 (Figure 2.S5). We used quantitative RT-PCR to measure the expression of those elements that increased more than 3-fold with an FDR < 0.1 in the microarray data. Of the 13 such transposon families, 11 showed a statistically significant increase in RNA expression when measured by qRT-PCR (Table 2.S2). For example, expression of the transposon *HeT-A* increased 72-fold (*p*-value = 0.004), *Springer* increased 55-fold (*p*-value = 0.002), and *Burdock* increased 41-fold (*p*-value = 0.006).

The transposon families whose expression increased in *qin* mutants correspond to a subset of those transposons whose silencing requires Ago3 (Figure 2.2D; Pearson correlation, $r = 0.77$, p-value $< 10^{-16}$). In contrast, the transposons hyper-expressed in *qin* mutants and those activated in *aub* mutants were more weakly correlated (Figure 2.2E; $r = 0.58$, p-value $= 1.7 \times 10^{9}$). A single copy transgene expressing full length *qin* cDNA in the germline via a UASp promoter driven by a *nanos*-Gal4-VP16 transgene restored transposon silencing (Figures 2.2A and 2.2B). *ZAM*, a transposon silenced by Piwi in the somatic follicle cells (Brennecke et al., 2007; Mevel-Ninio et al., 2007; Desset et al., 2008; Malone et al., 2009), was unaffected in qin^1/qin^1 and qin^1/Df ovaries (Figure 2.2B). We also did not detect a change in the expression of *blood*, a transposon thought to be silenced in both the germline and the follicle cells (Li et al., 2009; Malone et al., 2009).

In testes, piRNAs silence the multi-copy gene *Stellate*, a gene whose hyper-expression causes Stellate protein crystals to form during spermatogenesis, reducing male fertility (Hardy et al., 1981; Hardy et al., 1984; Balakireva et al., 1992; Aravin et al., 2001; Vagin et al., 2006). Anti-*Stellate* piRNAs derive from transcripts from the many copies of *Suppressor of Stellate* (*Su(Ste)*) (Hardy et al., 1984; Livak, 1984; Livak, 1990; Balakireva et al., 1992). Accumulation of *Su(Ste)*-derived piRNAs requires Aub (Aravin et al., 2004) and Ago3 (Li et al., 2009), but not Piwi (Vagin et al., 2006). *Stellate* silencing also requires Qin. Stellate crystals accumulated in the testes of qin^1 / *Df* mutant males but not their heterozygous siblings (Figure 2.3B). Expression of FM-Qin cDNA in testes restored *Stellate* silencing (Figure 2.3B).

Qin Co-localizes with Ago3 and Aub in Nuage

To define the subcellular distribution of Qin, we examined the localization of the Qin::EGFP fusion encoded by the 97,532 bp genomic fragment gf-2. We detected EGFP fluorescence in the germline throughout oogenesis, but not in the somatic follicle cells (Figure 2.S3C). Together with the finding that a FM-Qin cDNA expressed in the germline rescued the dorsal appendage defects and sterility of *qin* mutant females and rescued *Stellate* silencing in males, these data suggest that Qin acts mainly or solely in the germline. The Qin::EGFP fusion protein produced by the gf-2 transgene and the epitope tagged FM-Qin produced by the full-length *qin* cDNA transgene expressed from a UASp promoter driven by *nanos*-Gal4-VP16 were both predominately cytoplasmic and localized in perinuclear foci typical for nuage (Figures 2.S3A and 2.S3C). Moreover, in all nurse cells examined (n = 30, two measurements for each nurse cell), quantitative fluorescence microscopy showed that Qin::EGFP co-localized with Aub and

Figure 2.3

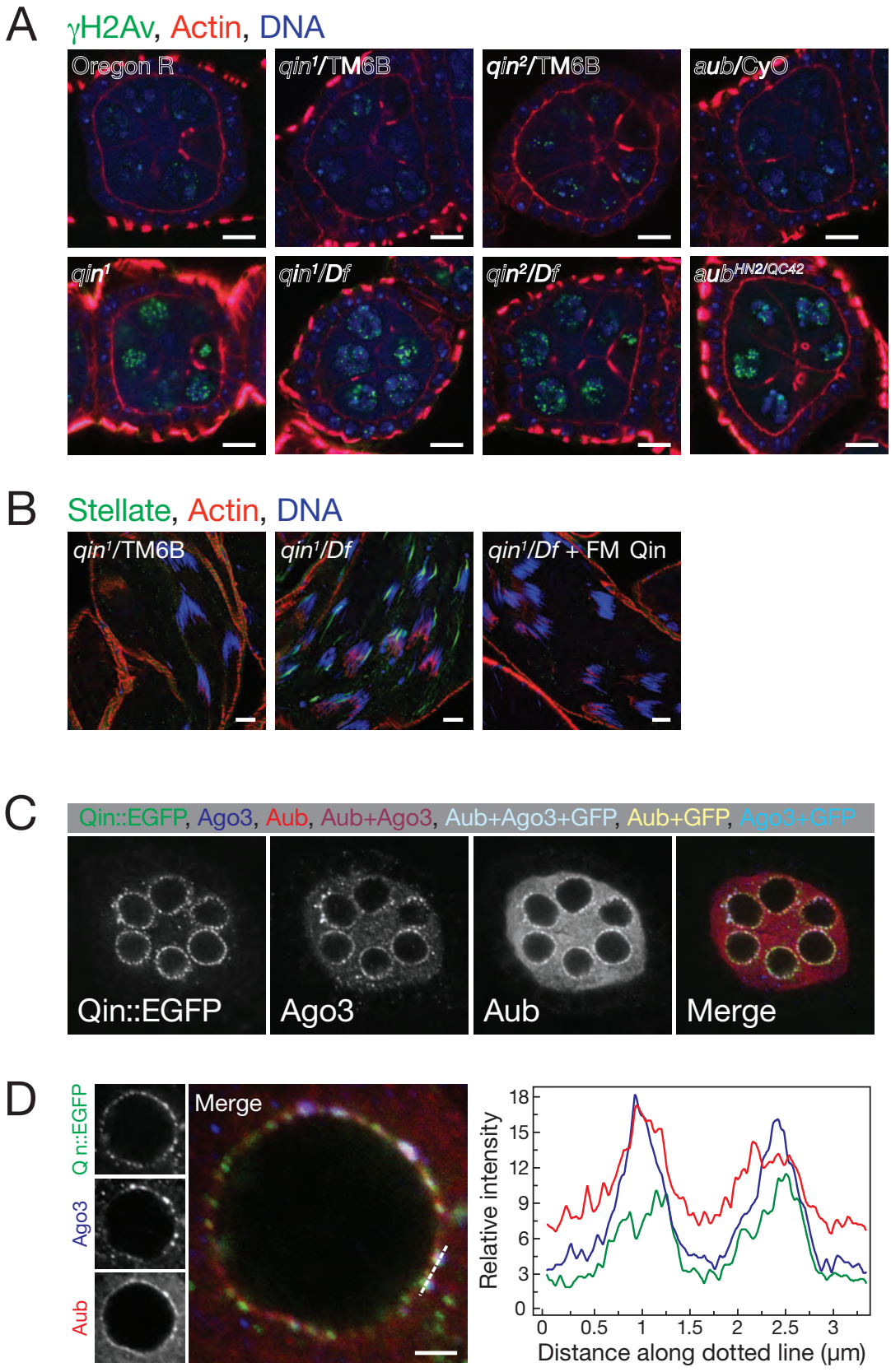
A
γH2Av, Actin, DNA
Oregon R
qin1/TM6B
qin2/TM6B
aub/CyO
qin1
qin1/Df
qin2/Df
aubHN2/QC42
B
Stellate, Actin, DNA
qin1/TM6B
qin1/Df
qin1/Df + FM Qin
C
Qin::EGFP, Ago3, Aub, Aub+Ago3, Aub+Ago3+GFP, Aub+GFP, Ago3+GFP
Qin::EGFP
Ago3
Aub
Merge
D
Q n::EGFP
Ago3
Aub
Merge
Relative intensity
0
3
6
9
12
15
18
0
0.5
1
1.5
2
2.5
3
Distance along dotted line (μm)

Figure 2.3. Qin Colocalizes with Aub and Ago3 in Nuage.

(A) Germline γH2Av foci, a hallmark of DNA damage, accumulated in *qin* and *aub* mutant ovaries. Stage 2–3 egg chambers are shown.

(B) *Stellate* silencing in testes requires Qin.

(C) Qin co-localizes with Ago3 and Aub in the nuage.

(D) Enlarged view of the nuage of a single germline nurse cell. At right, quantification of the fluorescence corresponding to Qin::EGFP, Ago3, and Aub along the dash line drawn on the left panel. (A–C) Scale bars, 10 μm. (D) Scale bar, 2 μm.

Ago3 (Figure 2.3C and 2.3D), which have been shown previously to localize to nuage (Harris and Macdonald, 2001; Snee and Macdonald, 2004; Brennecke et al., 2007; Lim and Kai, 2007). However, mutation of *qin* did not appear to affect the assembly or stability of nuage, as Aub and Ago3 all retained their perinuclear, punctate distribution in both qin^1/Df and qin^2/Df ovaries (Figure 2.S6). Neither a truncated Qin protein comprising the Qin E3 ligase-like domain alone nor a truncated Qin protein comprising the five Tudor domains but missing the E3 ligase domain localized to nuage (Figure 2.S3A) or rescued the phenotypic defects observed in *qin* mutant ovaries (data not shown).

Compared with *qin* heterozygotes, the abundance of Ago3, Aub, and Piwi was not significantly changed in *qin* mutants (Figures 2.4A and 2.S7). We did observe a small decrease in Ago3 abundance in the absence of Qin, but this 27% reduction is unlikely to explain the defects observed in *qin* mutants for two reasons. First, over-expression of Ago3 in the germline of *qin* mutant ovaries failed to rescue female sterility (hatch rate = 1.1%, n = 1,031). Second, the abundance of Ago3 was lower in *ago3* heterozygotes than in qin^1/Df ovaries, yet *ago3*/TM6B females are fertile, lay eggs with wild-type dorsal appendages, and silence their transposons (Li et al., 2009).

Compared with the control strain (w^{1118}), the abundance of Vasa increased in both *qin* heterozygous and mutant ovaries (Figures 2.4A and 2.S7), although its localization to nuage was not detectably altered (Figure 2.S6). Both *qin* heterozygous and mutant ovaries contained an additional Vasa isoform, likely phosphorylated Vasa, which has been associated with loss of transposon silencing and DNA damage in the germline (Ghabrial and Schupbach, 1999; Abdu et al., 2002; Chen et al., 2007; Klattenhoff et al., 2007).

The Ratio of Sense to Antisense piRNAs Increases in *qin* Mutants

qin mutants show little change in the total abundance of piRNAs: the percent of genome-mapping, small RNA sequence reads corresponding to transposon-derived, 23–29 nt small RNAs was essentially indistinguishable among control (cn^1; ry^{506}; 360,387/949,504 = 38%), *qin*/TM6B heterozygous (5,558,981/14,619,724 = 38%), and qin^1/Df mutant (6,064,405/16,096,174 = 38%) ovaries (Figure 2.4B and Table 2.S3A and 2.S3B). However, the effects of mutation of *qin* on the structure of piRNA populations can be readily detected by analyzing the fraction of piRNAs bearing the sense orientation (sense fraction = sense/[sense + antisense]) for each of the 93 transposon families for which we sequenced ≥ 100 parts per million (ppm) piRNA reads in both control and *qin*/TM6B ovaries. Mutation of *qin* increased the median sense fraction for the 93 transposon families from 0.24 for the control ovaries to 0.33 for *qin* heterozygotes and 0.41 for qin^1/Df mutants. We note that an increase in sense fraction implies a decrease in the *proportion*—but not necessarily the abundance—of antisense piRNAs, those piRNAs believed to direct transposon silencing.

One potential explanation for an increase in sense piRNAs in *qin* mutant ovaries is that the increased abundance of transposon transcripts directly feeds the production of sense piRNAs. In this view, the observed increase in the fraction of sense piRNAs would be a consequence, rather than the cause, of the loss of transposon silencing in *qin* mutant ovaries. To test this idea, we compared the change in transposon transcript abundance to the change in sense piRNA abundance for the 93 transposon families we analyzed (Figure 2.S8A). Transposon expression and sense piRNA abundance were uncorrelated, even when we restricted our analysis to only those 11 transposon families whose expression increased significantly. We conclude that increased transposon

Figure 2.4

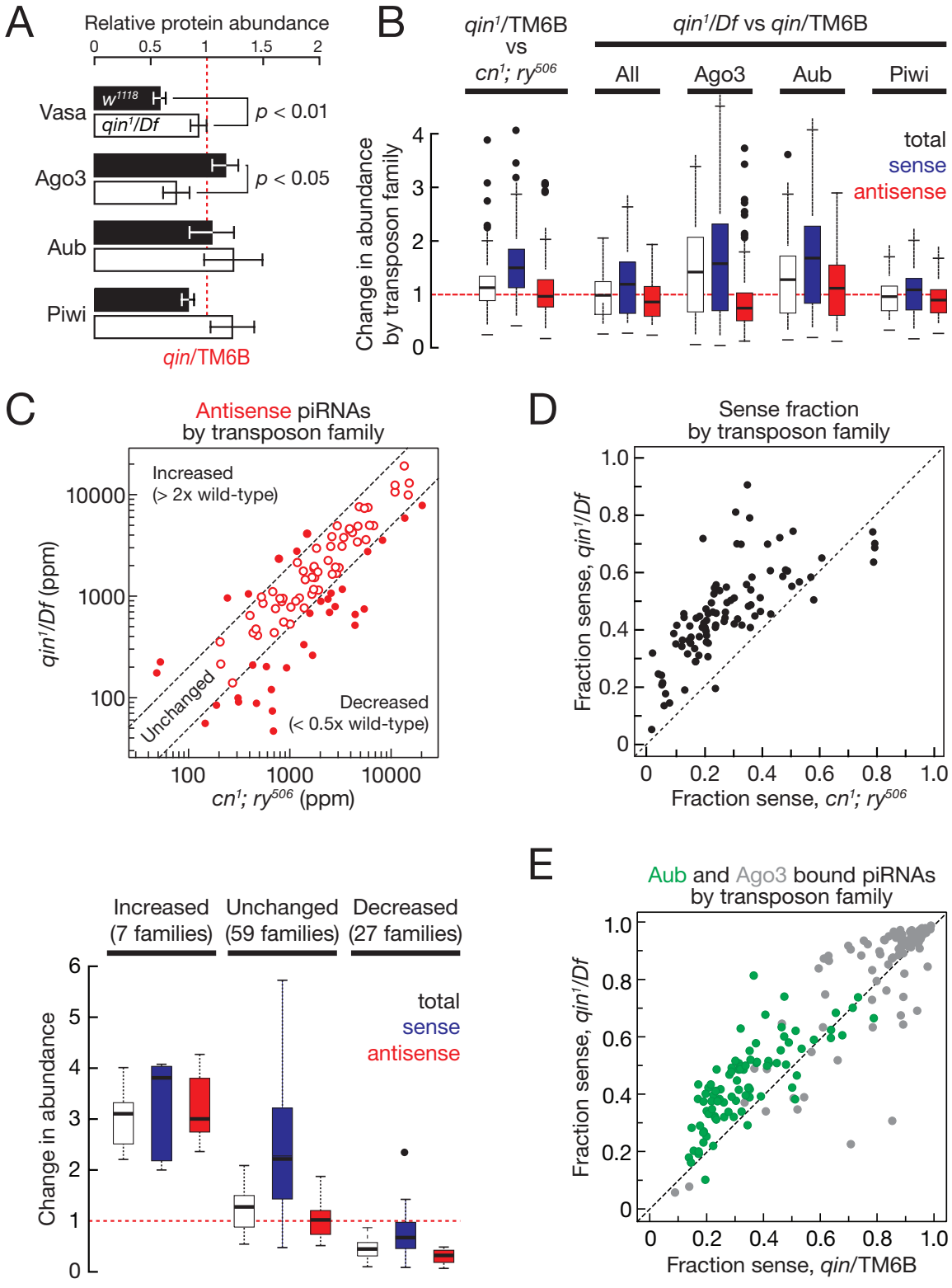
A
Relative protein abundance
0
0.5
1
1.5
2
Vasa
Ago3
Aub
Piwi
w1118
qin1/Df
p < 0.01
p < 0.05
qin/TM6B
B
qin1/TM6B
vs
cn1; ry506
qin1/Df vs qin/TM6B
All
Ago3
Aub
Piwi
total
sense
antisense
Change in abundance
by transposon family
C
Antisense piRNAs
by transposon family
Increased
(> 2x wild-type)
Unchanged
Decreased
(< 0.5x wild-type)
qin1/Df (ppm)
cn1; ry506 (ppm)
D
Sense fraction
by transposon family
Fraction sense, qin1/Df
Fraction sense, cn1; ry506
E
Aub and Ago3 bound piRNAs
by transposon family
Fraction sense, qin1/Df
Fraction sense, qin/TM6B
Increased
(7 families)
Unchanged
(59 families)
Decreased
(27 families)
total
sense
antisense
Change in abundance

Figure 2.4. Mutation of *qin* Decreases the Antisense Character of piRNA Populations.

(A) Steady-state protein abundance in control and *qin*^{1}/*Df* ovaries, relative to *qin*/TM6B (red dashed line). Data are mean ± standard deviation for 3 independent biological samples.

(B) Box plots report the change in abundance of all piRNAs mapping to transposons in control (*cn*^{1}; *ry*^{506}), *qin*/TM6B and *qin*^{1}/*Df* ovaries and of the piRNAs associated with immunoprecipitates of each PIWI protein in *qin*/TM6B and *qin*^{1}/*Df* ovaries.

(C) Top: antisense piRNA abundance for each transposon family in *qin*^{1}/*Df* ovaries was compared to control. Bottom: box plots report the change in piRNA abundance by family for each of the three groups of transposon families defined in the upper panel.

(D) The piRNA sense fraction for each transposon family was compared between control and *qin* mutant ovaries.

(E) The sense fraction for each transposon family for Aub- and Ago3-bound piRNAs was compared between *qin* heterozygous and *qin*^{1}/*Df* mutant ovaries.

expression cannot explain the increased sense piRNA abundance in *qin* mutant ovaries.

Among the 93 transposon families analyzed, 27 lost more than half their antisense piRNAs in qin^1/Df ovaries (Figures 2.4C; analyses of each transposon family are available at http://www.umassmed.edu/zamore/datasets.aspx?linkidentifier=id&itemid=66736). The median abundance of antisense piRNAs for these 27 transposon families in qin^1/Df ovaries was 32% of their median in the control, whereas the median sense piRNA abundance in qin^1/Df ovaries was 67% of the control median. Consequently, the piRNA populations from these 27 transposon families became less antisense and more sense biased. That is, loss of antisense piRNAs and not gain of sense piRNAs underlies the increase in sense fraction for these transposon families. Of the 27 transposon families, eight were among the 11 families that showed significantly elevated steady-state mRNA expression in qin^1/Df compared with control ovaries (Figure 2.2C). Overall, antisense piRNAs from the 11 derepressed transposon families decreased more than those from the other 82 transposon families (p-value = 0.018, Wilcoxon rank-sum test), suggesting that loss of antisense piRNAs caused the transposon desilencing. This correlation was particularly striking for *HeT-A*, *Burdock*, *TAHRE,* and *I element*, whose expression increased ~40-, 38-, 21-, and 14-fold in *qin* mutants and whose antisense piRNA pools declined to only 14%, 18%, 35%, and 21% of control levels (Table 2.S2A).

Of the remaining 66 transposon families, antisense piRNA abundance either increased or decreased by less than a factor of two for 59 families, and more than doubled for 7 families. While the median antisense piRNA abundance for these 66 families was unchanged from control, median sense piRNA abundance increased 2.3-fold (p-value = 5.7×10^{-10}, two-tailed, paired t-test;

Figure 2.4C). Even for these transposons, the normal antisense bias of piRNAs was reduced in *qin* mutant ovaries. Notably, *copia* expression increased 23-fold and *Transpac* expression increased 14-fold in *qin* mutant ovaries, yet *copia* antisense piRNA abundance increased 3-fold and *Transpac* antisense piRNA abundance was unchanged, compared to cn^{1}; ry^{506} control ovaries (Table 2.S2A). Thus, our data suggest that the ratio of sense to antisense piRNAs, not simply the abundance of antisense piRNAs, determines the efficacy of piRNAs in silencing transposons. The data also suggest that Qin acts to maintain the antisense bias of transposon piRNA populations.

Indeed, 74 of the 93 transposon families we analyzed had a greater piRNA sense fraction (*qin* mutants – control > 0.05) in the ovaries of the *qin* mutants than in the control (Figure 2.4D). The proportion of sense piRNAs declined in *qin* mutants compared with control (control – *qin* heterozygotes > 0.05) for only 6 transposon families (*accord2*, *diver2*, *hopper*, *hopper2*, *gypsy*, and *gyspsy12*). Among these, four families (*accord2*, *diver2*, *hopper*, *hopper2*) correspond to transposons whose *sense* piRNAs have been previously shown to be loaded into Aub, causing *antisense* piRNAs to accumulate in Ago3 (Brennecke et al., 2007; Li et al., 2009; Malone et al., 2009). It is not known if the Ago3-bound antisense piRNAs act to repress these elements. A fifth transposon family, *gypsy*, is thought to be active in the somatic follicle cells surrounding the oocyte where it is silenced by a mechanism that requires Piwi but not Aub or Ago3 (Li et al., 2009; Malone et al., 2009). We do not currently understand the mechanism that causes the sixth transposon family, *gypsy12*, to gain antisense piRNAs in qin^{1}/Df ovaries.

The overall increase in the fraction of piRNAs corresponding to sense sequences was reflected in the sense fraction of piRNAs bound to Aub in *qin* mutant ovaries: among the 93 transposon families analyzed, the median sense fraction for the piRNAs co-immunoprecipitated with Aub increased from 0.31 in

qin heterozygotes to 0.42 in *qin* mutants (Figure 2.4E); the median sense fraction of Aub-bound piRNAs in wild-type Oregon R ovaries is 0.16 (Brennecke et al., 2007). In contrast, the sense fraction of piRNAs bound to Piwi was essentially unaltered in *qin* mutant ovaries (Figure 2.S8B). Perhaps the excess sense piRNA bound to Aub in *qin* mutant ovaries sequesters factors needed for antisense piRNAs to direct transposon silencing.

Qin is Required to Maintain Aub:Ago3 Ping-Pong

To understand why in *qin* mutants the proportion of sense piRNAs increased for so many transposon families, we examined the frequency of Ping-Pong piRNA pairs among the piRNAs bound to Aub and Ago3. The first 10 nt of Ping-Pong piRNA pairs are complementary, evidence of piRNA amplification by reciprocal cycles of cleavage by PIWI proteins, typically Aub and Ago3. Our preliminary analysis, in which we used only piRNA sequence reads unique to Ago3, Aub, or Piwi immunoprecipitates detected no significant Ping-Pong between Piwi and Ago3, Piwi and Aub, or Piwi and itself (z-scores ≤ 2.8, i.e., p-value ≥ 0.05). Because so many Aub-bound species are also bound to Piwi, but Piwi participates so little in Ping-Pong, we analyzed the piRNAs bound to Ago3 or Aub, excluding only species bound to both Ago3 and Aub. In contrast to our previously published Ping-Pong analyses, we computed a z-score for the occurrence of Ping-Pong piRNA pairs using a method that is uninfluenced by sequencing depth (JX and ZW, manuscript in preparation).

The Ping-Pong model for piRNA amplification suggests a simple explanation for the effect of mutation of *qin*: in *qin* mutants homotypic Ping-Pong between Aub and itself replaces heterotypic Ping-Pong between Aub and Ago3. Homotypic Ping-Pong, which occurs at low levels in control ovaries, is predicted to generate equal amounts of sense and antisense piRNAs. Thus, homotypic

Ping-Pong is predicted to diminish the antisense bias of the piRNA population (Gunawardane et al., 2007; Brennecke et al., 2007; Malone et al., 2009). Aub:Aub homotypic Ping-Pong dominates in *ago3* mutants, which not only disrupt piRNA amplification and produce fewer piRNAs overall, but also show increased piRNA sense fraction (Li et al., 2009).

To test if inappropriate Aub:Aub Ping-Pong replaced productive Aub:Ago3 Ping-Pong in *qin* mutant ovaries, we calculated Ping-Pong *z*-scores for Aub:Aub, Aub:Ago3, and Ago3:Ago3 pairs, using only those piRNAs that could be assigned uniquely to Ago3 or Aub. In *qin* heterozygotes, Ping-Pong between Aub and Ago3 predominated (Figures 2.5A), with far more Aub:Ago3 Ping-Pong pairs (*z*-score = 25) detected than Aub:Aub (*z*-score = 13) or Ago3:Ago3 (*z*-score = 3.9). In *qin* mutants, far fewer Aub-bound piRNAs showed the characteristic 10 nt 5′ complementarity to Ago3-bound piRNAs (Aub:Ago3 *z*-score = 12), and Aub:Aub Ping-Pong emerged as the main pairing (Aub:Aub *z*-score = 26); Ago3:Ago3 Ping-Pong was lost (*z*-score = 1.8; Figures 2.5A and 2.S8C).

HeT-A, *Springer*, *Burdock*, *I element*, *Transpac*, and *TAHRE*, six of the seven transposons that had the greatest increase in transposon expression, all switched from Aub:Ago3 Ping-Pong to Aub:Aub Ping-Pong in *qin* mutants (Table 2.S2B). Intriguingly, a single transposon family, the non-long-terminal-repeat retroelement *Doc*, was hyper-silenced in both *qin* and *ago3* mutant ovaries. *Doc* expression decreased 2.6-fold in *ago3* (Li et al., 2009) and 3.7-fold in *qin*, even though seemingly inappropriate Aub:Aub Ping-Pong increased (the Aub:Aub *z*-score for *Doc* increased from 7.3 to 21), normally productive Aub:Ago3 Ping-Pong decreased (the *z*-score decreased from 12 to 4.5), the proportion of piRNAs antisense to *Doc* declined, the abundance of antisense piRNAs for *Doc* was unchanged, and the abundance of sense piRNAs more than doubled. We do not currently understand why *qin* and *ago3* mutants show enhanced *Doc* silencing.

We envision that in wild-type ovaries, piRNAs shared between Aub and Ago3 correspond to secondary piRNAs generated by Aub:Aub and by Aub:Ago3 Ping-Pong, respectively. In *qin* mutants, Aub:Aub Ping-Pong occurs more often and Aub:Ago3 Ping-Pong occurs less frequently. Therefore, mutation of *qin* should decrease the number of secondary piRNAs common to Aub and Ago3. Consistent with this idea, the fraction of piRNAs shared between Ago3 and Aub declined from 37% in *qin*/TM6B to 13% in qin^1/Df ovaries (p-value < 2.2×10^{-16}, Fisher's exact test).

When Aub participates in heterotypic Ping-Pong, Aub-bound sense piRNAs often begin with uridine (U) but rarely contain an adenosine (A) at position 10. Primary piRNAs derived from cluster transcripts are believed to begin with U, whereas a position 10 A is the signature of a secondary piRNA produced by the Ping-Pong mechanism. (Aub-bound secondary piRNAs generated by cleavage of longer RNAs by Ago3-bound secondary piRNAs also typically begin with U, but cannot be distinguished from primary piRNAs on that basis.) We computed the nucleotide composition for all Aub- and Ago3-bound piRNAs (Figure 2.5B and 2.S9A) and for the subset of piRNAs that had a Ping-Pong partner piRNA (Figure 2.5C). Aub-bound sense piRNAs were more likely to begin with U in *qin* heterozygous ovaries than in *qin* mutant ovaries (Figure 2.5B, 2.5C and 2.S9A). Conversely, Aub-bound sense piRNAs were more likely to have an A at position 10 in *qin* mutants than in heterozygotes, consistent with the emergence in the mutant ovaries of inappropriate Aub:Aub Ping-Pong that produces Aub-bound, sense secondary piRNAs. The Aub-bound putative secondary piRNAs did not, however, favor an initial U (Figure 2.5C). These data Ago3 correspond to secondary piRNAs generated by Aub:Aub and by Aub:Ago3

Figure 2.5

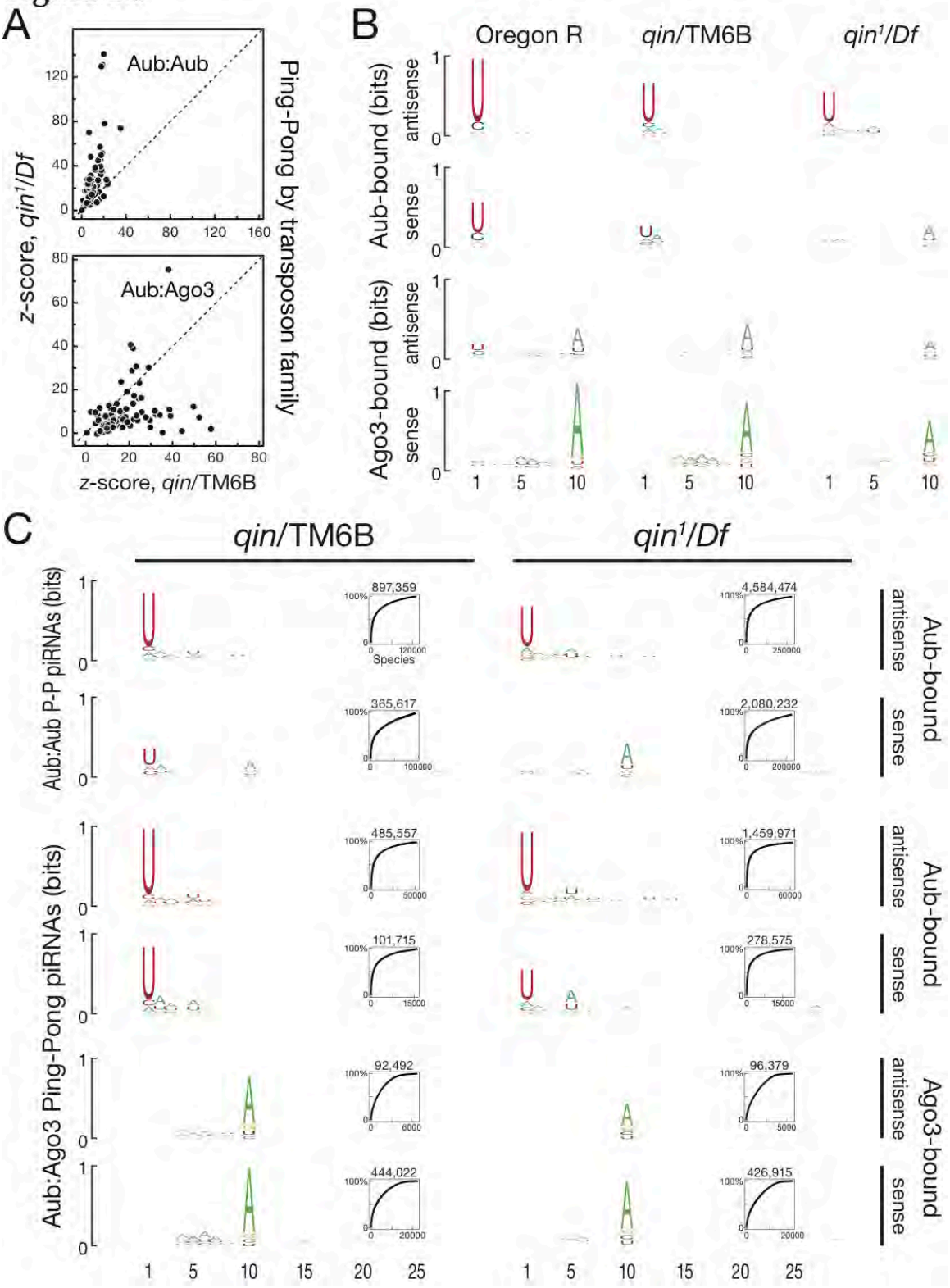
A
Aub:Aub
Aub:Ago3
z-score, qin1/Df
z-score, qin/TM6B
Ping-Pong by transposon family
B
Oregon R
qin/TM6B
qin1/Df
Aub-bound (bits)
Ago3-bound (bits)
antisense
sense
C
qin/TM6B
qin1/Df
Aub:Aub P-P piRNAs (bits)
Aub:Ago3 Ping-Pong piRNAs (bits)
Species
897,359
365,617
4,584,474
2,080,232
485,557
101,715
1,459,971
278,575
92,492
444,022
96,379
426,915
Aub-bound
Ago3-bound

Figure 2.5. Futile Aub:Aub Homotypic Ping-Pong Prevails in *qin* Mutants.

(A) Ping-Pong *z*-score, measuring the significance of 5′, 10 nt complementarity between piRNAs bound to Aub or Ago3, was calculated for Aub:Aub and Aub:Ago3 pairs for each transposon family for *qin*/TM6B and qin^{1}/*Df* ovaries.

(B) Sequence logos (Schneider and Stephens, 1990) for the first 10 nt of Aub- and Ago3-bound total piRNAs; Figure S9A illustrates all 29 nt.

(C) Sequence logos for sense and antisense piRNAs co-immunoprecipitated with Aub and uniquely participating in Aub:Aub Ping-Pong and for piRNAs co-immunoprecipitated with Aub or Ago3 and uniquely participating in Aub:Ago3 Ping-Pong. Insets report the cumulative number of species required to account for a given percent of reads; the total number of reads composing each logo appears above each inset.

Ping-Pong, respectively. In *qin* mutants, Aub:Aub Ping-Pong occurs more often and Aub:Ago3 Ping-Pong occurs less frequently. Therefore, mutation of *qin* should decrease the number of secondary piRNAs common to Aub and Ago3. Consistent with this idea, the fraction of piRNAs shared between Ago3 and Aub declined from 37% in *qin*/TM6B to 13% in qin^{1}/Df ovaries (p-value < 2.2×10^{-16}, Fisher's exact test).

When Aub participates in heterotypic Ping-Pong, Aub-bound sense piRNAs often begin with uridine (U) but rarely contain an adenosine (A) at position 10. Primary piRNAs derived from cluster transcripts are believed to begin with U, whereas a position 10 A is the signature of a secondary piRNA produced by the Ping-Pong mechanism. (Aub-bound secondary piRNAs generated by cleavage of longer RNAs by Ago3-bound secondary piRNAs also typically begin with U, but cannot be distinguished from primary piRNAs on that basis.) We computed the nucleotide composition for all Aub- and Ago3-bound piRNAs (Figure 2.5B and 2.S9A) and for the subset of piRNAs that had a Ping-Pong partner piRNA (Figure 2.5C). Aub-bound sense piRNAs were more likely to begin with U in *qin* heterozygous ovaries than in *qin* mutant ovaries (Figure 2.5B, 2.5C and 2.S9A). Conversely, Aub-bound sense piRNAs were more likely to have an A at position 10 in *qin* mutants than in heterozygotes, consistent with the emergence in the mutant ovaries of inappropriate Aub:Aub Ping-Pong that produces Aub-bound, sense secondary piRNAs. The Aub-bound putative secondary piRNAs did not, however, favor an initial U (Figure 2.5C). These data suggest that the strong tendency for Aub-bound primary piRNAs to begin with U does not reflect an intrinsic property of Aub, but rather derives from a step in piRNA production before Aub loading, such as fragmentation of primary piRNA transcripts.

Aub:Ago3 Ping-Pong continues in *qin* mutants, but it no longer comprises the majority of Ping-Pong interactions. Ping-Pong between Aub and Ago3 declined both for Aub-bound antisense piRNAs that Ping-Pong with Ago3-bound sense piRNAs (*z*-score = 20.2 in *qin* heterozygotes; 12.5 in qin^1 / *Df*) and for Aub-bound sense piRNAs that Ping-Pong with Ago3-bound antisense piRNAs (*z*-score = 16.9 *qin* heterozygotes; 11.4 in qin^1 / *Df*). Despite the replacement in *qin* mutants of Aub:Ago3 Ping-Pong with Aub:Aub Ping-Pong as the main mechanism for piRNA amplification, Ago3-bound piRNAs that participate in Ping-Pong in qin^1 / *Df* ovaries retained their 10A signature (Figure 2.5C). That is, those Ago3 piRNAs—all of which are expected to be secondary piRNAs—that are still made in qin^1/*Df* ovaries appear to be mainly generated by Aub-catalyzed RNA cleavage directed by piRNAs bearing a U at position 1, rather than by the Aub-bound secondary piRNAs produced by Aub:Aub Ping-Pong that bear an A at position 10.

To understand why, we set out to quantify Ago3-bound tertiary piRNAs generated by Aub:Aub:Ago3 Ping-Pong. We readily detected Aub-bound piRNAs that bear an A at position 10 and Ping-Pong with Ago3-bound piRNAs. These Ago3-bound piRNAs would be expected to begin with U, but might also fortuitously contain an A at position 10, making it difficult to determine whether they represent the product of an Aub-bound primary piRNA (i.e., U at position 1) or an Aub-bound secondary piRNA generated by Aub:Aub Ping-Pong (i.e., an A at position 10). To avoid this ambiguity, we restricted our analysis to Ago3-bound piRNAs that begin with U but bear a C, G, or U but not an A at position 10. Such "1U, non-10A" piRNAs comprised ~4.9% of all Ago3-bound piRNAs possessing Ping-Pong partners in *qin* heterozygotes, but encompassed ~8.2% of all Ago3-bound piRNAs with Ping-Pong partners in qin^1 / *Df* mutant ovaries. We conclude that the increase in Aub-bound secondary piRNAs generated by

inappropriate Aub:Aub Ping-Pong in *qin* mutants leads to a corresponding increase in Ago3-bound tertiary piRNAs generated by Aub:Aub:Ago3 Ping-Pong. These tertiary piRNAs do not generate a 1U signature in the Ago3 piRNA sequence logos, likely because most Ago3-bound piRNAs are still produced by Aub-bound primary piRNAs even in qin^1 / *Df* mutant ovaries.

Association of Ago3 with Aub Requires Qin

Ago3 co-immunoprecipitates with Aub from ovary lysate, suggesting that at least a fraction of Ago3 is bound to or present in a common complex with Aub (Nishida et al., 2009). We immunoprecipitated Aub from ovary lysate prepared from flies bearing a transgene expressing FLAG-Myc–tagged Ago3 (FM-Ago3), then measured the amount of co-immunoprecipitated FM-Ago3 by Western blotting using anti-FLAG antibody (Figure 2.6A). The association of Ago3 with Aub was not bridged by RNA: pre-incubation of the lysate with RNase A did not affect the co-association of FM-Ago3 with Aub (Figure 2.6B); control experiments demonstrated that the RNase treatment reduced miR-317 levels to background and reduced 2S rRNA, a highly structured RNA component of the ribosome, by 15-fold (Figure 2.6C). Qin was required for the association of FM-Ago3 with Aub. Compared to *qin* heterozygotes, > 6-fold less FM-Ago3 co-immunoprecipitated with Aub in lysate from qin^1 / *Df* ovaries (mean ± standard deviation of co-immunoprecipitated FM-Ago3 in *qin* mutant ovaries was 16 ± 12% of the heterozygous control, $n = 3$; Figures 2.6D, 2.S9B, 2.S9C, and 2.S9D).

Figure 2.6

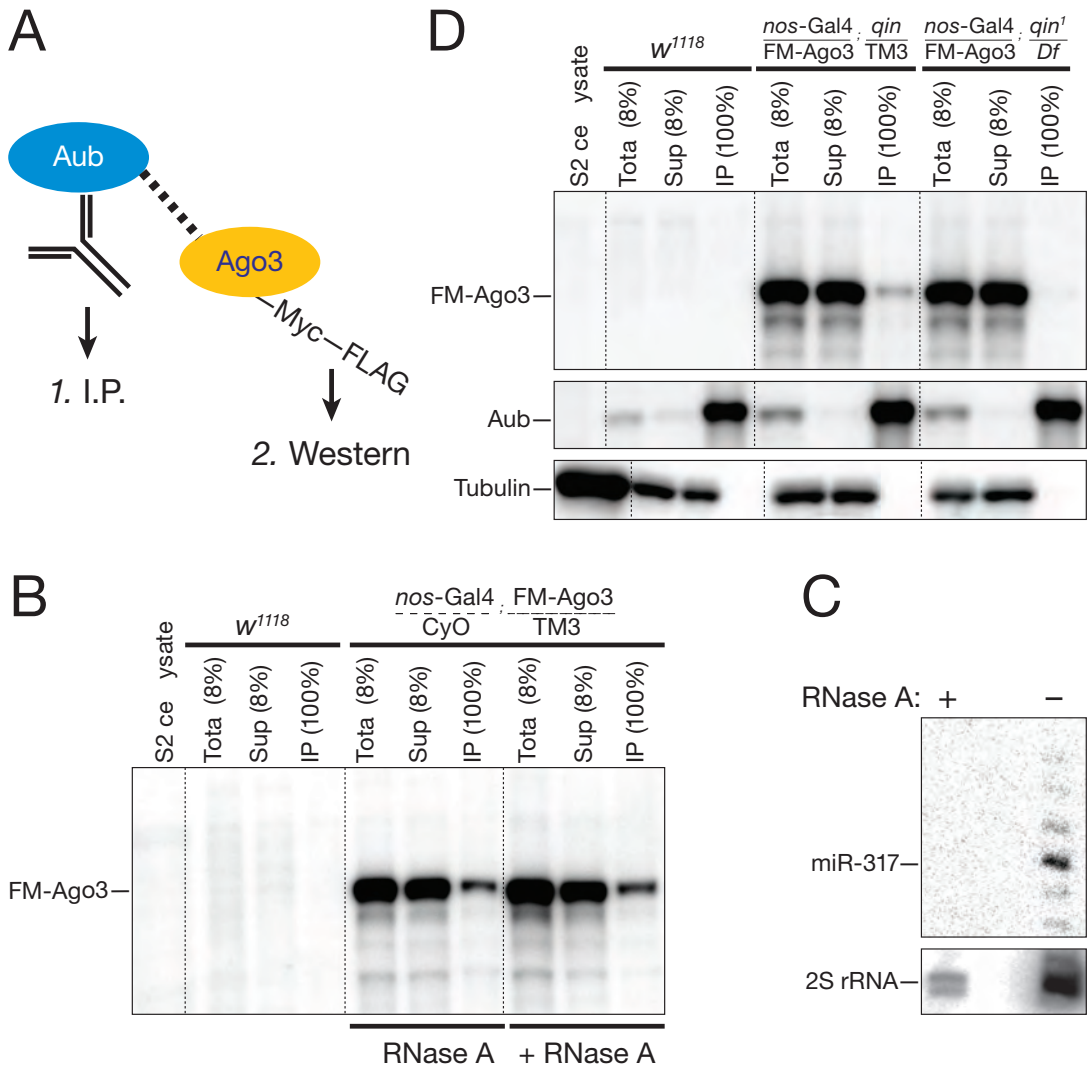
A
Aub
Ago3
Myc
FLAG
1. I.P.
2. Western
D
S2 cells ysate
w1118
nos-Gal4 ; qin / FM-Ago3 TM3
nos-Gal4 ; qin1 / Df
Tota (8%)
Sup (8%)
IP (100%)
FM-Ago3
Aub
Tubulin
B
w1118
nos-Gal4 ; FM-Ago3 / CyO TM3
FM-Ago3
RNase A
+ RNase A
C
RNase A: +
–
miR-317
2S rRNA

Figure 2.6. The Association of Ago3 with Aub Requires Qin but not RNA.

(A) Strategy for the experiments in (B–D).

(B) Binding of Ago3 to Aub does not require RNA.

(C) Treatment of ovary lysate with RNase A degrades both miRNAs and rRNAs as evidenced by the loss of miR-317 and 2S rRNA.

(D) FM-Ago3 associates with Aub in *qin*/TM6B ovaries; the interaction was reduced in qin^{1}/*Df*. After probing with anti-FLAG to detect Ago3, anti-Aub was used to measure the efficiency of immunoprecipitation. Figure 2.S9B presents the entire membrane and Figures 2.S9C and 2.S9D present two additional biological replicates.

DISCUSSION

Compared with the 11 *Drosophila* piRNA pathway mutants previously subjected to high throughput sequencing analysis, *qin* is unique (Figure 2.7). Like *qin*, mutations in *piwi* or *zucchini* increase overall Ping-Pong *z*-score and decrease antisense piRNA abundance, but neither *piwi* nor *zucchini* mutants increase the abundance of sense piRNAs. Moreover, Piwi and Zucchini appear to function mainly in the somatic follicle cells, where piRNAs are not amplified by the Ping-Pong cycle, so the increase in Ping-Pong observed in *piwi* and *zucchini* mutants largely reflects the loss of somatic piRNAs, rather than a direct effect on germline piRNAs.

Mutation of *ago3*, *aub*, *armitage*, *vasa*, *krimper*, *vret*, *rhino*, *spindle-E*, or *squash* disrupts the Ping-Pong mechanism, with mutations in *ago3*, *aub*, *vasa*, *krimper*, *vret*, *rhino*, and *spindle-E* eliminating Ping-Pong piRNA amplification for most or all transposon families. We note that *krimper*, a Tudor-domain protein, localizes to nuage like Qin, but plays a very different role in piRNA biogenesis from Qin: Ping-Pong amplification collapses in *krimper* mutants, whereas mutation of *qin* dramatically increases piRNA Ping-Pong by triggering non-productive Aub:Aub Ping-Pong. While the overall abundance of piRNAs is preserved in *qin* mutant ovaries, piRNA antisense bias declines, largely because of an increase in the abundance of sense piRNAs.

The replacement of heterotypic Aub:Ago3 Ping-Pong with homotypic Aub:Aub Ping-Pong in *qin* mutants suggests either that Qin acts to suppress homotypic Ping-Pong or that Qin promotes heterotypic Aub:Ago3 Ping-Pong. Our data cannot distinguish between these two models. Mutation of *qin* disrupts the interaction of Ago3 with Aub in ovary lysate. If Qin acts to suppress Aub self-association, the decrease in Ago3 bound to Aub might reflect the redirection of Aub from Ago3 to Aub itself. Alternatively, if Qin acts—directly or

Figure 2.7

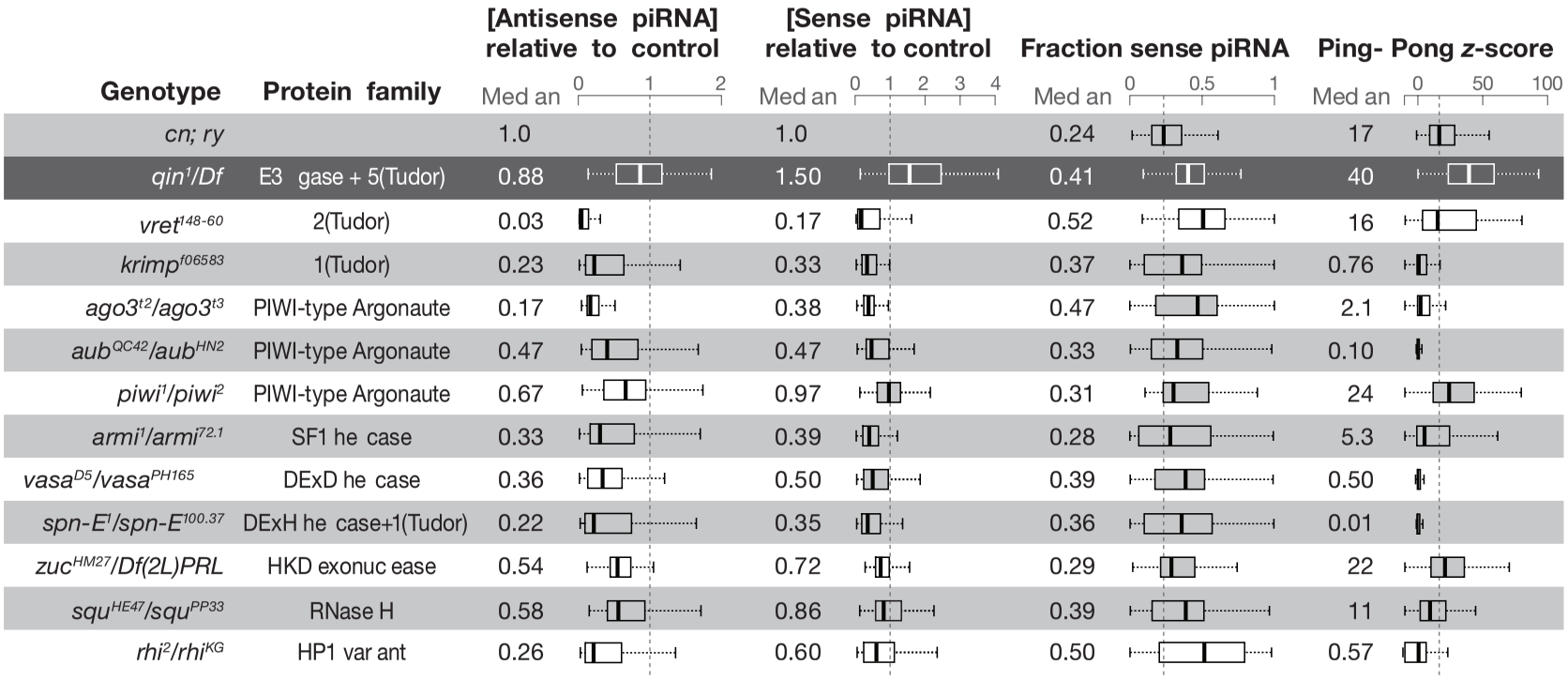
[Antisense piRNA] relative to control
[Sense piRNA] relative to control
Fraction sense piRNA
Ping- Pong z-score
Genotype
Protein family
Med an
0 1 2
0 1 2 3 4
0 0.5 1
0 50 100
cn; ry 1.0 1.0 0.24 17
qin1/Df E3 gase + 5(Tudor) 0.88 1.50 0.41 40
vret148-60 2(Tudor) 0.03 0.17 0.52 16
krimpf06583 1(Tudor) 0.23 0.33 0.37 0.76
ago3t2/ago3t3 PIWI-type Argonaute 0.17 0.38 0.47 2.1
aubQC42/aubHN2 PIWI-type Argonaute 0.47 0.47 0.33 0.10
piwi1/piwi2 PIWI-type Argonaute 0.67 0.97 0.31 24
armi1/armi72.1 SF1 he case 0.33 0.39 0.28 5.3
vasaD5/vasaPH165 DExD he case 0.36 0.50 0.39 0.50
spn-E1/spn-E100.37 DExH he case+1(Tudor) 0.22 0.35 0.36 0.01
zucHM27/Df(2L)PRL HKD exonuc ease 0.54 0.72 0.29 22
squHE47/squPP33 RNase H 0.58 0.86 0.39 11
rhi2/rhiKG HP1 var ant 0.26 0.60 0.50 0.57

Figure 2.7. Among piRNA Pathway Mutants, *qin* Has a Unique Effect on the Structure of the Ovarian piRNA Population.

For each mutant the figure summaries the change in the abundance of antisense and of sense piRNAs, relative to *cn*; *ry* control ovaries; the fraction of all piRNAs with the sense orientation; and the Ping-Pong *z*-score. Box plots present the distribution for the 93 transposon families analyzed; outliers are not shown. Numbers report the median value corresponding to the thick vertical line on each box plot.

indirectly—to promote the binding of Aub to Ago3, a decrease in Qin function would lead to futile homotypic Ping-Pong by default, especially if the intracellular concentration of Aub were greater than that of Ago3.

In contrast to the germline piRNAs of flies, which require Aub:Ago3 Ping-Pong to silence transposons during gamete development, piRNA production in the somatic follicle cells of the ovaries of Drosophilidae and of piRNAs targeting non-repetitive sequences during the pachytene phase of mammalian spermatogenesis require just a single PIWI protein. These piRNA pathways proceed without Ping-Pong amplification (Aravin et al., 2007; Robine et al., 2009; Saito et al., 2009; Lau et al., 2009b; Li et al., 2009; Malone et al., 2009). We do not know if Qin in flies or Qin-like proteins in other species play a role in the production of this second type of piRNA. It is intriguing that the mouse protein Tdrd4 (also called RNF17) resembles Qin. Like Qin, Tdrd4 contains amino-terminal E3 ligase motifs and five carboxy-terminal Tudor domains. While Tdrd4 has yet to be implicated in piRNA biogenesis or function, it is required for normal mouse spermatogenesis and localizes to nuage (Pan et al., 2005).

Heterotypic Ping-Pong between pairs of PIWI proteins drives the amplification of antisense piRNAs, the species believed to silence transposon expression during gamete formation in insects and likely many other invertebrates and during the prenatal development of the testes in mammals (Gunawardane et al., 2007; Brennecke et al., 2007; Aravin et al., 2008; Li et al., 2009; Malone et al., 2009). Our data, however, suggest that the proposal that antisense piRNA abundance alone determines the extent of repression of transposon families is too simple to explain how piRNAs silence transposons. We find that neither normal amounts of the PIWI proteins Aub and Ago3 nor the organization of these proteins into nuage suffices to promote productive amplification of silencing-competent antisense piRNAs. Nor are near wild-type

levels of antisense piRNAs sufficient to ensure adequate silencing of active transposons during *Drosophila* oogenesis. Instead, effective transposon silencing requires that Aub partner with Ago3, rather than itself, generating more Aub-bound antisense piRNAs than sense. We have shown here that this heterotypic partnership requires Qin. Key challenges for the future will be to determine how Qin promotes heterotypic Ping-Pong or represses homotypic Ping-Pong and what role its E3 ligase and Tudor domains play in this process.

EXPERIMENTAL PROCEDURES

General Methods. Female fertility tests, tiling arrays, immunoprecipitation, immunoblotting, and quantitative RT-PCR were as described (Li et al., 2009). Small RNA data for total small RNAs in cn^{1}; ry^{506} ovaries (Klattenhoff et al., 2009) was previously deposited in the NCBI trace archives (accession number SRP002060); tiling microarray data (Fig. 2, C–D) for total RNA in w^{1}, $aub^{HN2/QC42}$ and $ago3^{t2/t3}$ ovaries (Li et al., 2009; Klattenhoff et al., 2009) was previously deposited in the NCBI gene expression omnibus (accession number GSE14370). Small RNA data for Figure 2.7 was previously deposited in the NCBI trace archives SRP000458 (Li et al., 2009), GEO GSE15186 (Malone et al., 2009), and GSE30088 (Zamparini et al., 2011). Aub- and Ago3-associated small RNA data from Oregon R was published previously (Brennecke et al., 2007). Unless otherwise specified, *p*-values were calculated from at least three independent biological replicates using a two-tailed, two-sample unequal variance t-test (Excel, Microsoft).

***Drosophila* stocks.** All flies were raised at 25°C. PBac(RB)CG14303^{e03728} (qin^{1}) and *Df* flies were from the Bloomington Stock Center at Indiana University; PBac(RB)e01936 (qin^{2}) was from the Harvard Medical School stock center. P(w+mc, UASp-FM-Ago3-C2) (Chengjian Li and PDZ, unpublished) was used to express Ago3.

Transgenic flies. *qin* was isolated from Oregon R ovary cDNA by PCR using primers containing *attB* sites (Table S4), inserted into plasmid pDONR (Invitrogen, Carlsbad, CA, USA), and the cDNA sequenced (GENEWIZ, Cambridge, MA, USA) to confirm its structure. The *qin* cDNA was moved from pDONR to pPFMW vector by recombination using Clonase II Enzyme Mix

(Invitrogen). Transgene constructs were established by injection (BestGene, Chino Hills, CA, USA) using PhiC31 integrase-mediated genomic insertion and the 28E7 (BDSC 9723; gf-1), attP2 (BDSC 8622; EGFP-modified gf-2), and 53B2 (BDSC9736; UASp-FM-Qin cDNA, UASp-FM-$(Tud)_5$ and UASp-FM-E3) fly strains.

Genomic DNA fragments CH322-81J04 (gf-1) and CH321-88L14 (gf-2) were from the BACPAC Resources Center (Oakland, California, USA). EGFP was inserted into gf-2 by recombineering (Venken et al., 2009). Briefly, EGFP coding sequence and the kanamycin resistance gene (NPT II) were amplified by PCR from the plasmid pLAP (Poser et al., 2008) (gift of A. Hyman) using KOD Hot Start polymerase (EMD, Darmstadt, Germany). The P[acman] BAC clone for *qin*, CH321-88L14, was purified (NucleoBond BAC100 kit, Clontech, Mountain View, CA, USA), electroporated into SW102 cells, and positive colonies were selected using LB plates containing 12.5 μg/ml chloramphenicol and 10 μg/ml tetracycline at 32°C for 18–24 h. Positive colonies were verified by PCR, then recombineering was performed using the purified PCR product. Positive recombinants were selected on LB plates containing 25 μg/ml kanamycin, 12.5 μg/ml chloramphenicol and 10 μg/ml tetracycline at 32°C for 24–36 h. The resulting Qin::EGFP BAC was identified by PCR, and then sequenced to confirm that the EGFP tag was in-frame with Qin.

5′ RACE. 5′ RACE was performed with the SMARTer RACE Kit (Clontech) using 1 μg total ovary RNA according to the manufacturer's instructions.

Small RNA library construction and high-throughput sequencing. RNA libraries were prepared as described (Li et al., 2009) except that 2S rRNA was depleted by hybridization using a complementary, immobilized DNA

oligonucleotide (5′-biotin-TCA ATG TCG ATA CAA CCC TCA ACC ATA TGT AGT CCA AGC A-3′). Briefly, for each library 400 μl MyOne Streptavidin C1 Dynabeads (Invitrogen) were washed twice in 0.5× SSC (1× SSC: 150 mM sodium chloride, 15 mM sodium citrate, pH 7.0) at 4°C, then re-suspended in 400 μl 0.5× SSC. The beads were then loaded with 200 pmole DNA oligonucleotide at 4°C for 30 min, washed once in 0.5× SSC, re-suspended in 1 ml 0.5× SSC, and then warmed to 65°C for 5 min. Size-selected RNA (50 μg) was incubated at 80°C for 5 min and then added to the pre-warmed beads and incubated at 50°C for 1 h. The beads were removed by magnetic capture and the supernatant mixed with 3 volumes of ethanol, 0.3 M (f.c.) sodium acetate, pH 5.2, and 1 μl (20 μg/μl) glycogen (Roche, Indianapolis, IN, USA).

Co-immunoprecipitation assays. Ovaries were manually dissected from 2–4 day old females and homogenized in lysis buffer (100 mM potassium acetate, 6.4% glycerol, 30 mM HEPES-KOH, pH 7.4, 2 mM magnesium acetate, 0.1 mM EDTA, pH 8.0, 2 mM dithiothreitol (freshly added) containing 1 Mini, Complete, EDTA-free Protease Inhibitor Cocktail (Roche) tablet per 10 ml. GammaBind G Sepharose (GE Healthcare, Piscataway, NJ, USA) was loaded with rabbit anti-Aub polyclonal antibody (15 μl antibody per 30 μl Sepharose) at room temperature for 2 h, washed 5 times with lysis buffer, and then incubated with 400 μl freshly prepared ovary lysate (5 μg/μl) rotating at 4°C overnight. The supernatant was analyzed by immunoblotting to confirm Aub depletion. The beads were then washed twice with lysis buffer, 3 times with RIPA buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1.0% [v/v] NP-40), and then once with lysis buffer. Finally, the beads were boiled in 40 μl 1× SDS-sample buffer (50 mM Tris-HCl [pH 6.8], 2% [w/v] sodium dodecyl sulfate, 10% [v/v] glycerol, 1% [v/v] β-mercaptoethanol, 12.5 mM EDTA, 0.02% [w/v] bromophenol blue) and resolved

by electrophoresis through a 7.5% SDS-polyacrylamide gel until the 50 kDa protein standard was nearly at the bottom of the gel. The proteins were next transferred to 0.45 μm PVDF membrane (Millipore, Billerica, MA, USA), and the blot was probed with anti-FLAG mouse monoclonal antibody (Sigma) diluted 1:2,500 in TBST-milk (25 mM Tris-HCl pH 7.4, 3 mM KCl, 140 mM NaCl, 0.05% [v/v] Tween-20, 3% [w/v] non-fat dry milk), incubated overnight at 4°C. Next, the membrane was washed 4 times with TBST, 5 min each, and incubated 1 h at room temperature with sheep anti-mouse IgG-HRP (GE Healthcare) at 1:10,000 in TBST-milk. Then the membrane was washed four times with TBST at room temperature for 5 min and developed with SuperSignal West Dura Extended Duration Substrate (Pierce, Rockford, IL, USA). Image data were captured with an LAS-3000 image reader (Fujifilm, Tokyo, Japan). Afterwards, the membrane was re-probed with rabbit anti-Aub polyclonal antibody diluted 1:2,500 to determine the efficiency of immunoprecipitation. Rabbit anti-Aub polyclonal antibody was detected using sheep anti-mouse IgG-HRP (GE Healthcare). Quantitative analysis was performed using ImageGauge 4.22 (Fujifilm).

To determine whether the interaction between Aub and FM-Ago3 was RNA dependent, RNase A (Sigma) (f.c. 300 ng/μl) was added to the lysate during the immunoprecipitation.

Northern hybridization. Total ovary RNA (7 μg per sample) was dissolved in 25% (v/v) deionized formamide, 3% (v/v) formaldehyde, 5% (v/v) glycerol; 0.025% (w/v) bromophenol blue; 10 mM MOPS (pH 7.0), 2.5 mM sodium acetate, 1 mM EDTA, resolved by electrophoresis through a 1% (w/v) agarose, 2% (v/v) formaldehyde gel using 20 mM MOPS; 5 mM sodium acetate; 2 mM EDTA, pH 7.0, and transferred to a positively charged nylon membrane (Roche) by capillary transfer in 20× SSC. RNA was cross-linked to the membrane using 254 nm

ultraviolet light in a Stratalinker UV Crosslinker 2400 (Stratagene, Agilent, Santa Clara, CA, USA) by running the "auto crosslink" program once. The membrane was pre-hybridized in blocking buffer (Roche), and then incubated with DIG-labeled antisense probes prepared and used according to the manufacturer's instruction (Roche). The blot was developed using CDP-Star (Roche), and the image acquired using a 4000MM Image Station (Carestream, Rochester, NY, USA).

Immunohistochemistry. Egg chamber fixation and antibody labeling were as described (Li et al., 2009) except that to detect Ago3 and Aub simultaneously, mouse anti-Aub monoclonal antibody (gift of M. Siomi) was used diluted 1:1,000 and to detect FM-Qin, anti-FLAG mouse monoclonal antibody (Sigma, St. Louis, MO, USA) was used diluted 1:2,000. Images were processed using Leica Confocal Software 2.61 (Leica, Buffalo Grove, IL, USA).

Bioinformatic analyses. Analysis was as described (Li et al., 2009) except for the computation of Ping-Pong *z*-scores. For two piRNAs that were sufficiently complementary to each other at a particular 5′-to-5′ distance, a score was defined as the product of their abundances. The Ping-Pong *z*-score was then the difference of the score at the 5′-to-5′ distance of 10 nt and the mean scores of background distances, divided by the standard deviation of the scores of background distances, defined as distances of 0–9 and 11–20 nt. Two piRNAs were sufficiently complementary to each other when the nucleotides 2–10 of the first piRNA were perfectly paired with the second piRNA and there was at most one mismatch among positions 1 and 11–16 of the first piRNA. Genomic sequence adjacent to the second piRNA was used to determine complementarity when the 5′-to-5′ distance was less than 16 nt.

Accession Numbers. Sequence data generated in this study are available via the NCBI trace archives (http://www.ncbi.nlm.nih.gov/Traces/) with accession number SRP007101. Microarray data generated in this study are available via the NCBI gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo/) as GSE30061.

ACKNOWLEDGMENTS

We thank A. Boucher and G. Farley for technical assistance, J. Brennecke, G. Hannon, A. Hyman, and M. Siomi for reagents, and members of our laboratories for advice and critical comments on the manuscript. This work was supported in part by National Institutes of Health grant HD049116 to W.E.T., Z.W., and P.D.Z. and GM62862 and GM65236 to P.D.Z.

Figure 2.S1

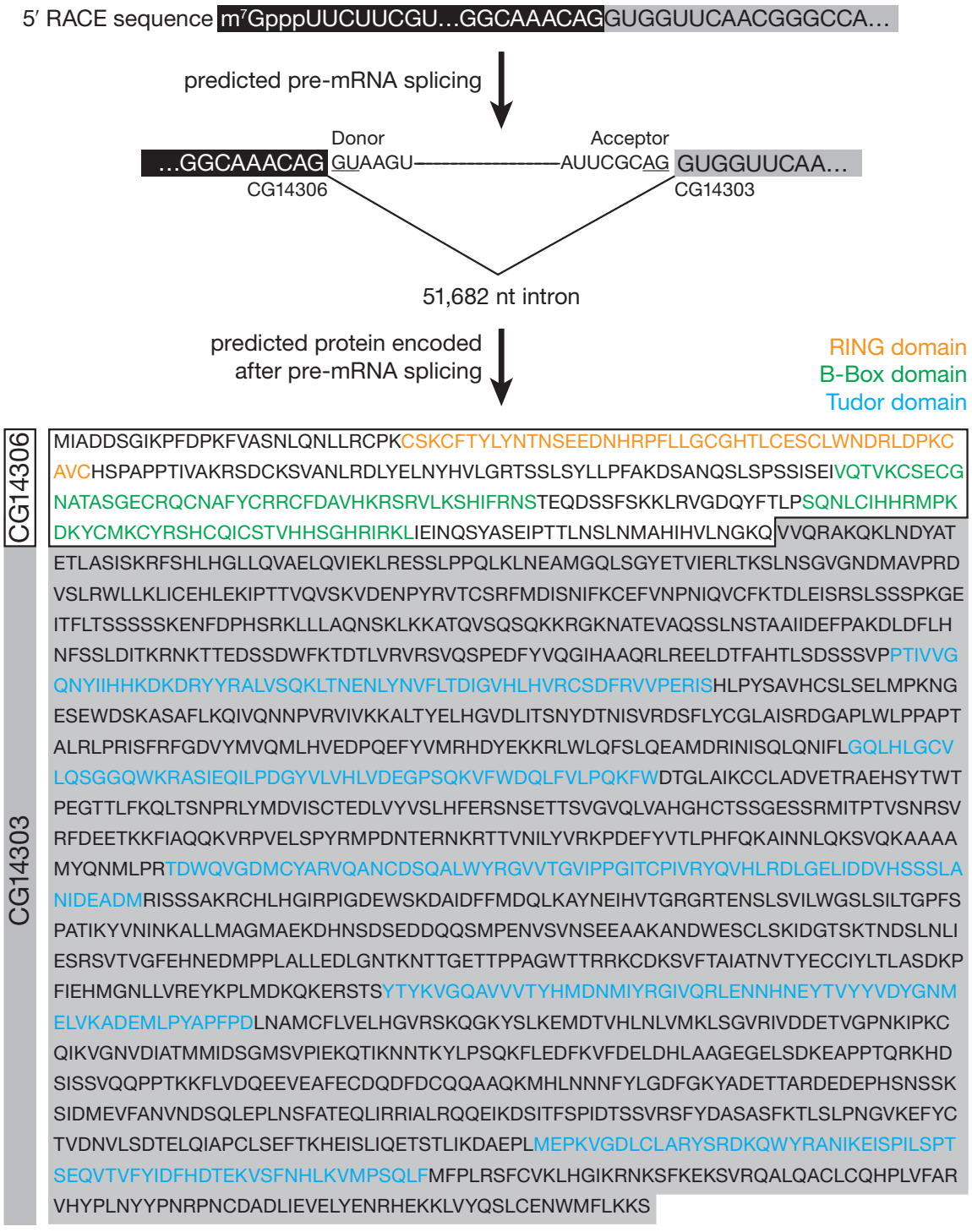
5′ RACE sequence m7GpppUUCUUCGU…GGCAAACAGGUGGUUCAACGGGCCA…
predicted pre-mRNA splicing
Donor
Acceptor
…GGCAAACAG GUAAGU AUUCGCAG GUGGUUCAA…
CG14306
CG14303
51,682 nt intron
predicted protein encoded after pre-mRNA splicing
RING domain
B-Box domain
Tudor domain
CG14306
CG14303
MIADDSGIKPFDPKFVASNLQNLLRCPKCSKCFTYLYNTNSEEDNHRPFLLGCGHTLCESCLWNDRLDPKC
AVCHSPAPPTIVAKRSDCKSVANLRDLYELNYHVLGRTSSLSYLLPFAKDSANQSLSPSSISEIVQTVKCSECG
NATASGECRQCNAFYCRRCFDAVHKRSRVLKSHIFRNSTEQDSSFSKKLRVGDQYFTLPSQNLCIHHRMPK
DKYCMKCYRSHCQICSTVHHSGHRIRKLIEINQSYASEIPTTLNSLNMAHIHVLNGKQVVQRAKQKLNDYAT
ETLASISKRFSHLHGLLQVAELQVIEKLRESSLPPQLKLNEAMGQLSGYETVIERLTKSLNSGVGNDMAVPRD
VSLRWLLKLICEHLEKIPTTVQVSKVDENPYRVTCSRFMDISNIFKCEFVNPNIQVCFKTDLEISRSLSSSPKGE
ITFLTSSSSSKENFDPHSRKLLLAQNSKLKKATQVSQSQKKRGKNATEVAQSSLNSTAAIIDEFPAKDLDFLH
NFSSLDITKRNKTTEDSSDWFKTDTLVRVRSVQSPEDFYVQGIHAAQRLREELDTFAHTLSDSSSVPPTIVVG
QNYIIHHKDKDRYYRALVSQKLTNENLYNVFLTDIGVHLHVRCSDFRVVPERISHLPYSAVHCSLSELMPKNG
ESEWDSKASAFLKQIVQNNPVRVIVKKALTYELHGVDLITSNYDTNISVRDSFLYCGLAISRDGAPLWLPPAPT
ALRLPRISFRFGDVYMVQMLHVEDPQEFYVMRHDYEKKRLWLQFSLQEAMDRINISQLQNIFLGQLHLGCV
LQSGGQWKRASIEQILPDGYVLVHLVDEGPSQKVFWDQLFVLPQKFWDTGLAIKCCLADVETRAEHSYTWT
PEGTTLFKQLTSNPRLYMDVISCTEDLVYVSLHFERSNSETTSVGVQLVAHGHCTSSGESSRMITPTVSNRSV
RFDEETKKFIAQQKVRPVELSPYRMPDNTERNKRTTVNILYVRKPDEFYVTLPHFQKAINNLQKSVQKAAAA
MYQNMLPRTDWQVGDMCYARVQANCDSQALWYRGVVTGVIPPGITCPIVRYQVHLRDLGELIDDVHSSSLA
NIDEADMRISSSAKRCHLHGIRPIGDEWSKDAIDFFMDQLKAYNEIHVTGRGRTENSLSVILWGSLSILTGPFS
PATIKYVNINKALLMAGMAEKDHNSDSEDDQQSMPENVSVNSEEAAKANDWESCLSKIDGTSKTNDSLNLI
ESRSVTVGFEHNEDMPPLALLEDLGNTKNTTGETTPPAGWTTRRKCDKSVFTAIATNVTYECCIYLTLASDKP
FIEHMGNLLVREYKPLMDKQKERSTSYTYKVGQAVVVTYHMDNMIYRGIVQRLENNHNEYTVYYVDYGNM
ELVKADEMLPYAPFPDLNAMCFLVELHGVRSKQGKYSLKEMDTVHLNLVMKLSGVRIVDDETVGPNKIPKC
QIKVGNVDIATMMIDSGMSVPIEKQTIKNNTKYLPSQKFLEDFKVFDELDHLAAGEGELSDKEAPPTQRKHD
SISSVQQPPTKKFLVDQEEVEAFECDQDFDCQQAAQKMHLNNNFYLGDFGKYADETTARDEDEPHSNSSK
SIDMEVFANVNDSQLEPLNSFATEQLIRRIALRQQEIKDSITFSPIDTSSVRSFYDASASFKTLSLPNGVKEFYC
TVDNVLSDTELQIAPCLSEFTKHEISLIQETSTLIKDAEPLMEPKVGDLCLARYSRDKQWYRANIKEISPILSPT
SEQVTVFYIDFHDTEKVSFNHLKVMPSQLFMFPLRSFCVKLHGIKRNKSFKEKSVRQALQACLCQHPLVFAR
VHYPLNYYPNRPNCDADLIEVELYENRHEKKLVYQSLCENWMFLKKS

Figure 2.S1. Sequences surrounding the 5′ and 3′ splice sites of the 51,682 nt *qin* intron and the inferred sequence of the Qin protein.

Figure 2.S2

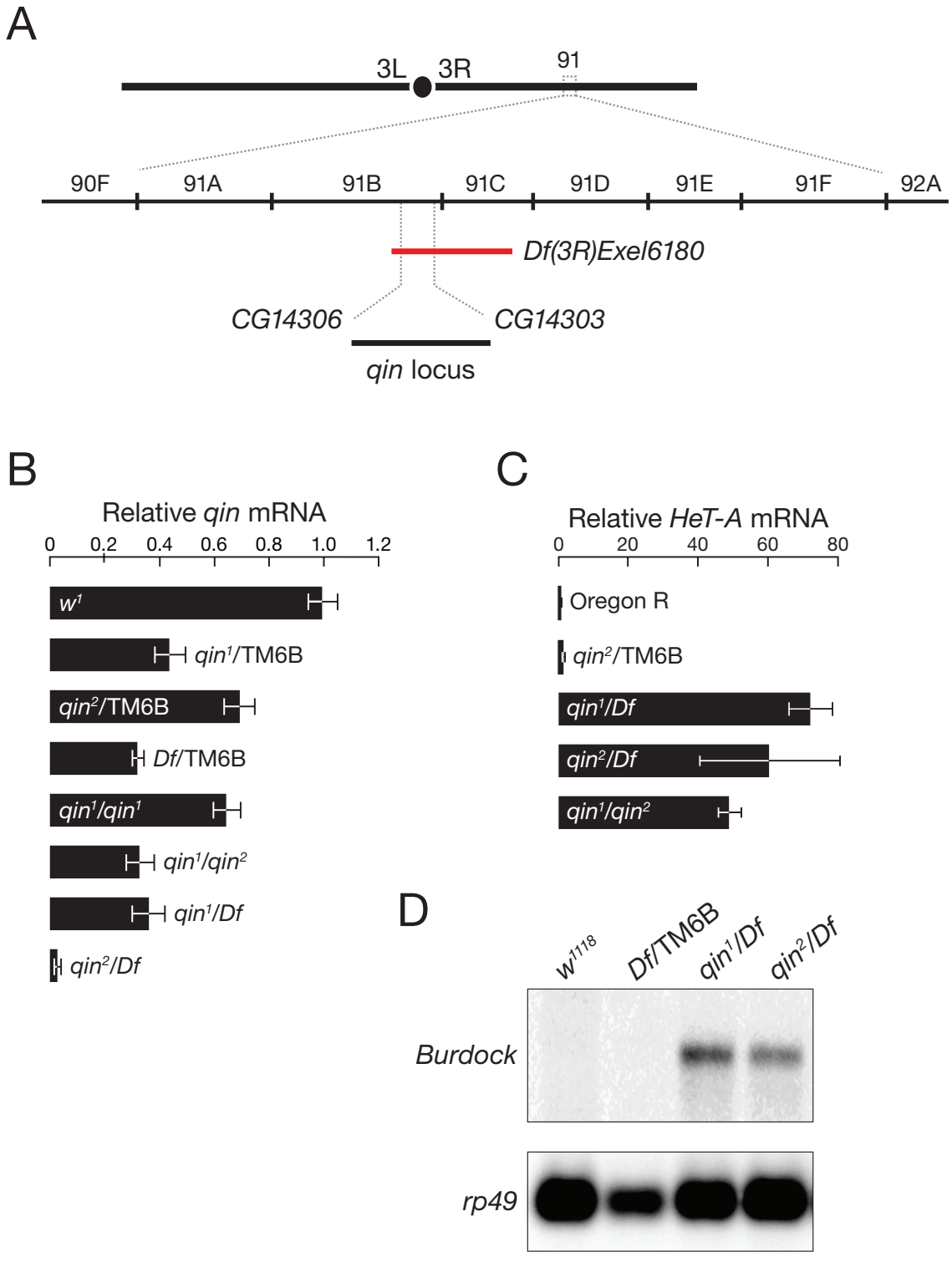
A
3L
3R
91
90F
91A
91B
91C
91D
91E
91F
92A
Df(3R)Exel6180
CG14306
CG14303
qin locus
B
Relative qin mRNA
0
0.2
0.4
0.6
0.8
1.0
1.2
w1
qin1/TM6B
qin2/TM6B
Df/TM6B
qin1/qin1
qin1/qin2
qin1/Df
qin2/Df
C
Relative HeT-A mRNA
0
20
40
60
80
Oregon R
qin2/TM6B
qin1/Df
qin2/Df
qin1/qin2
D
w1118
Df/TM6B
qin1/Df
qin2/Df
Burdock
rp49

Figure 2.S2. Qin silences *HeT-A* and *Burdock* in *Drosophila* ovaries.

(A) *Df* deletes the *qin* locus.

(B) Quantitative-RT-PCR analysis measuring *qin* mRNA expression in ovaries of the indicated genotypes. The data were normalized to *rp49* (*RpL32*) expression. The bars report mean ± standard deviation for three biological replicates.

(C) Quantitative RT-PCR measuring *HeT-A* abundance in *qin* mutant ovaries. Data were normalized to *rp49* (*RpL32*) expression. The bars report mean ± standard deviation for three biological replicates.

(D) Northern hybridization measuring *Burdock* abundance in *qin* heterozygous (*Df*/TM3) and mutant ovaries.

Figure 2.S3

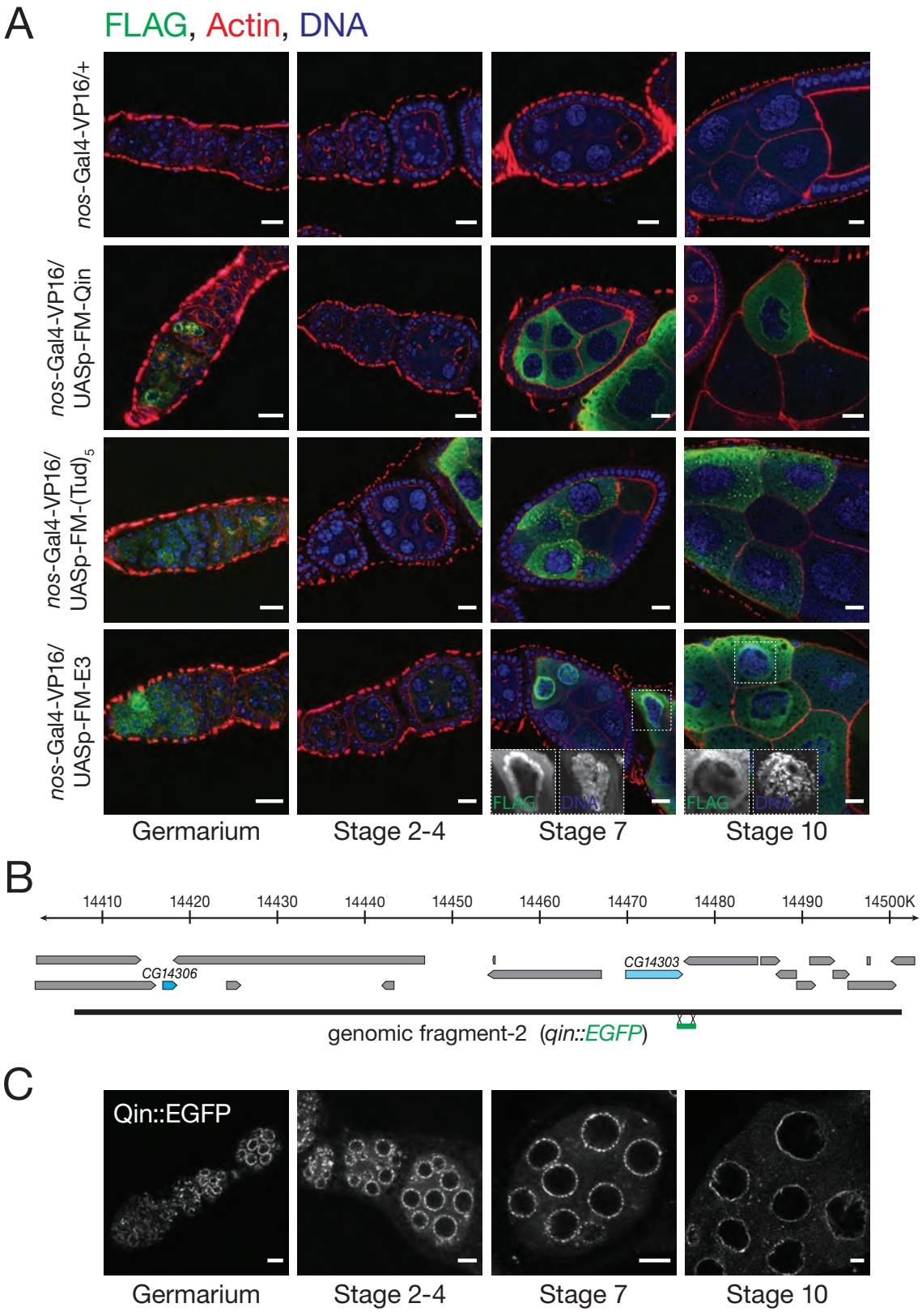
A
FLAG, Actin, DNA
nos-Gal4-VP16/+
nos-Gal4-VP16/ UASp-FM-Qin
nos-Gal4-VP16/ UASp-FM-(Tud)5
nos-Gal4-VP16/ UASp-FM-E3
FLAG
DNA
Germarium
Stage 2-4
Stage 7
Stage 10
B
14410
14420
14430
14440
14450
14460
14470
14480
14490
14500K
CG14306
CG14303
genomic fragment-2 (qin::EGFP)
C
Qin::EGFP
Germarium
Stage 2-4
Stage 7
Stage 10

Figure 2.S3. Qin localization.

(A) A *nos*-Gal4-VP16 transgene was used to drive expression of *UASp* FLAG-Myc-tagged full-length Qin, FLAG-Myc-tagged truncated $(Tudor)_5$ Qin lacking the putative E3 ligase domain, or FLAG-Myc-tagged Qin containing the E3 ligase domain but lacking the Tudor domains. *nos*-Gal4-VP16 is expressed in the germarium and in late, but not early, stages of oogenesis.

(B) Diagram of the Qin::EGFP transgene containing a genomic fragment encompassing *qin*, with EGFP inserted at the Qin carboxy-terminus to allow detection of the intracellular localization of Qin.

(C) Live-cell EGFP images of the intracellular distribution of Qin during *Drosophila* oogenesis. Scale bars, 10 μm.

Figure 2.S4

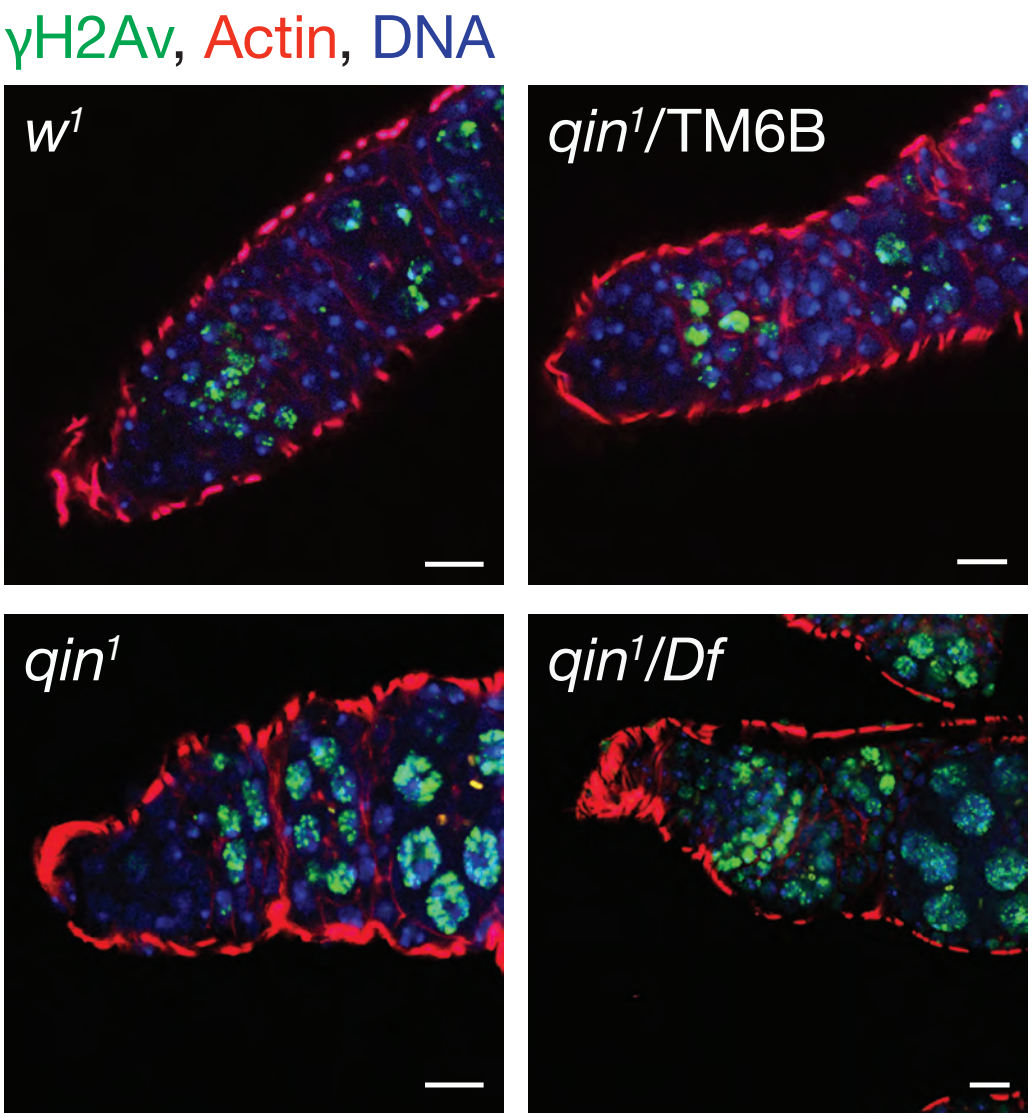

Figure 2.S4. Germline γH2Av foci in *qin* mutant ovaries.

In control (w^1) and qin^1 / TM6B ovaries, γH2Av foci in the stage 2a and 2b region of the germarium likely correspond to the normal DNA strand breaks induced during meiotic recombination. In both qin^1 and qin^1/Df ovaries, γH2Av foci inappropriately persist during subsequent stages of oogenesis. Scale bars, 10 μm.

Figure 2.S5

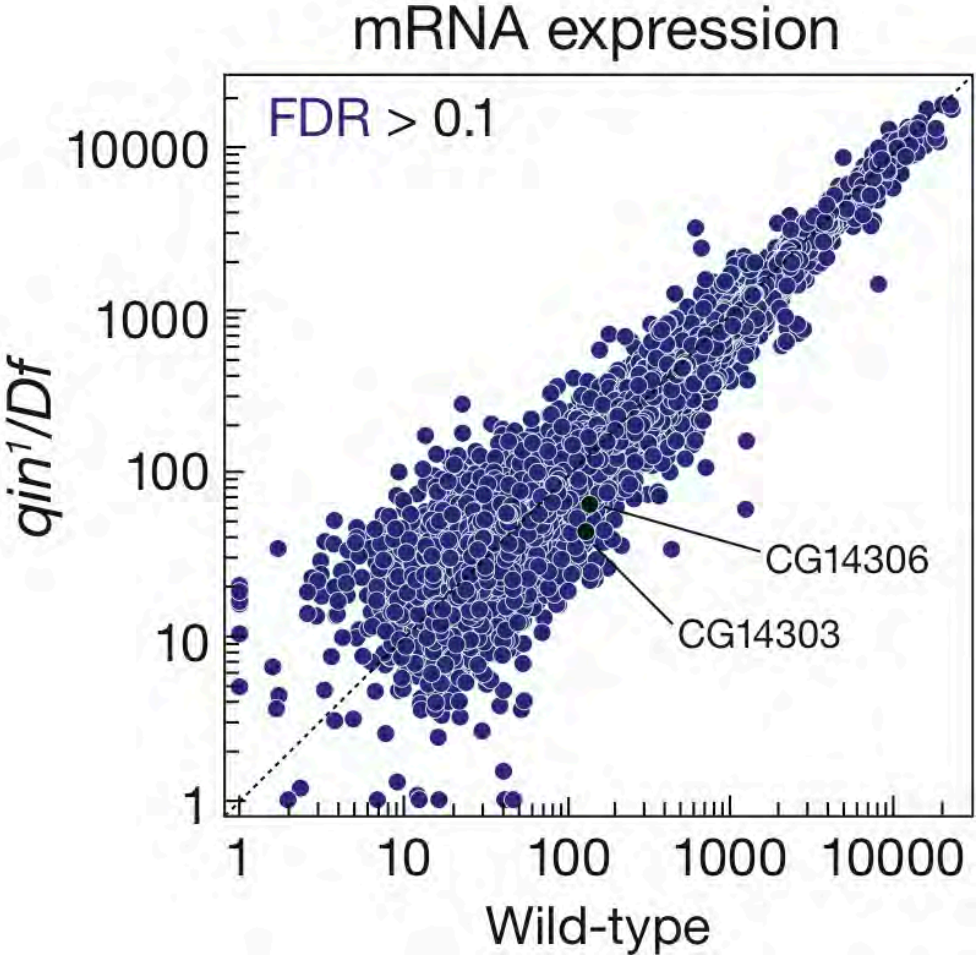
mRNA expression
FDR > 0.1
10000
1000
100
10
1
qin¹/Df
CG14306
CG14303
1
10
100
1000
10000
Wild-type

Figure 2.S5. Tiling microarray analysis of mRNA expression in control (w^1) and qin^1/Df ovaries.

All of the changes detected for mRNA corresponded to FDR ≥ 0.1, suggesting that loss of *qin* has no significant effect on mRNA expression.

Figure 2.S6

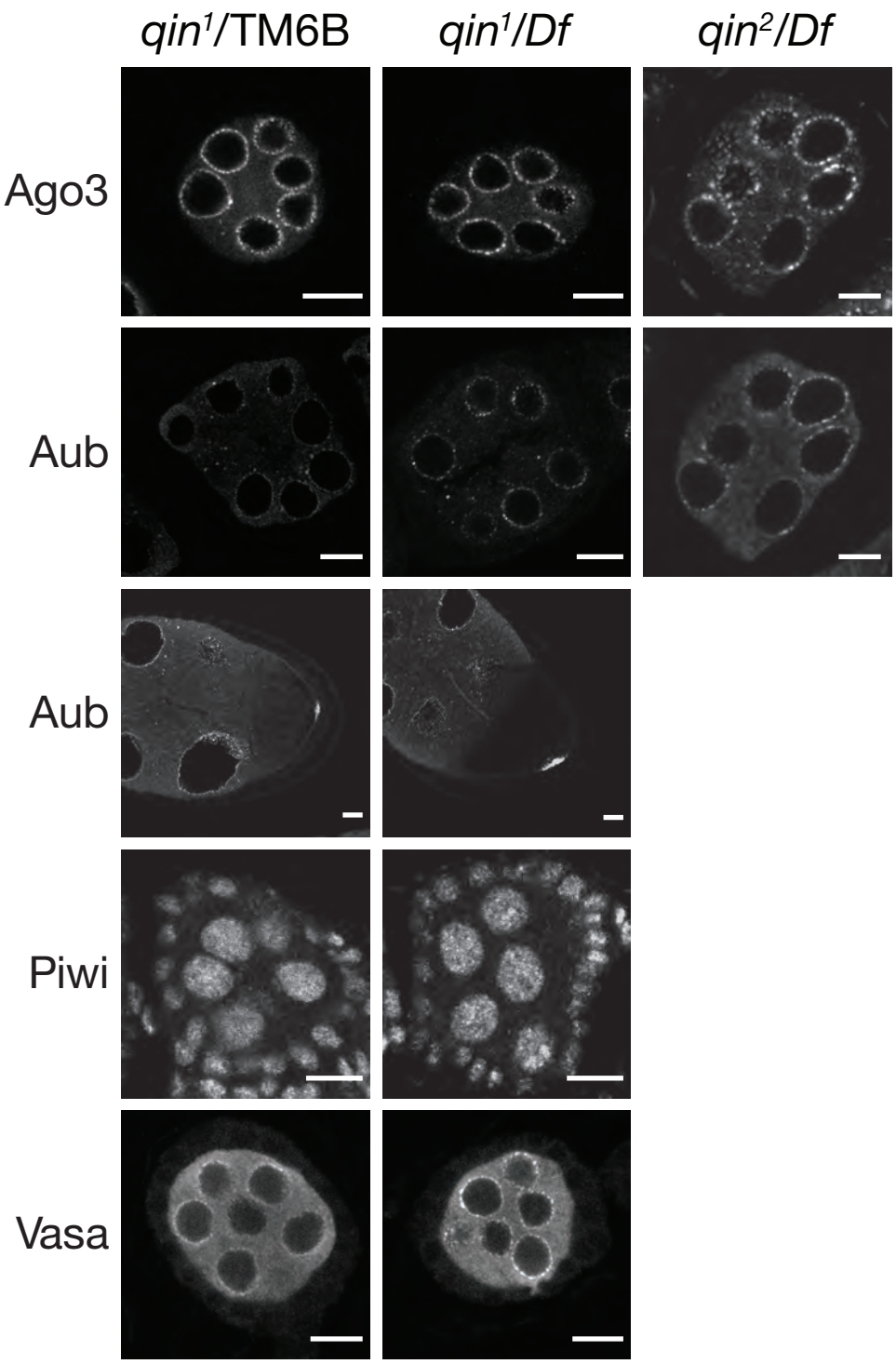

Figure 2.S6. Localization of PIWI and Vasa proteins in *qin* heterozygous and mutant ovaries.

Scale bars, 10 μm.

Figure 2.S7

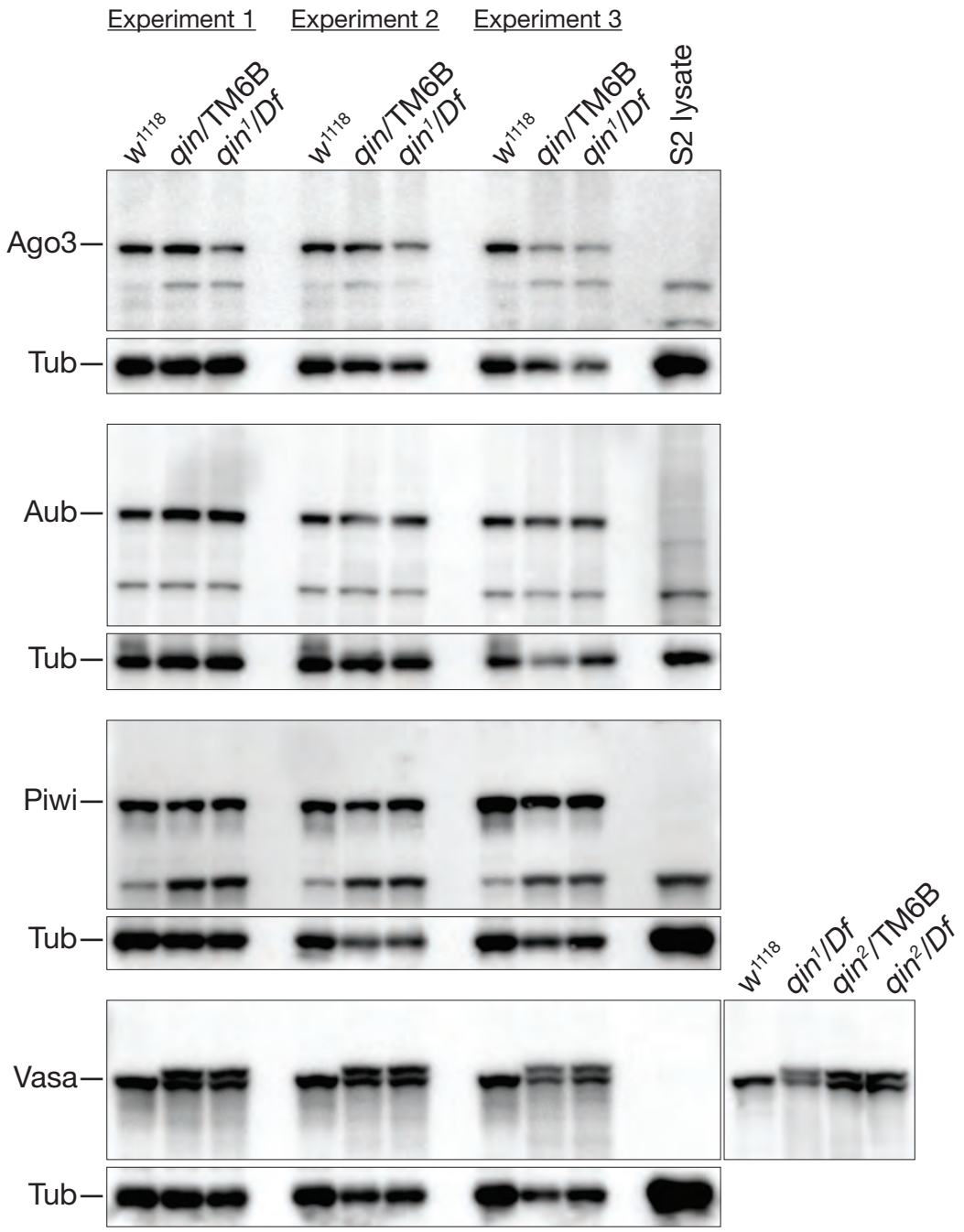
Experiment 1
Experiment 2
Experiment 3
w1118
qin/TM6B
qin1/Df
w1118
qin/TM6B
qin1/Df
w1118
qin/TM6B
qin1/Df
S2 lysate
Ago3
Tub
Aub
Tub
Piwi
Tub
w1118
qin1/Df
qin2/TM6B
qin2/Df
Vasa
Tub

Figure 2.S7. Abundance of PIWI and Vasa proteins in *qin* heterozygous and *qin* mutant ovaries.

These data were used to generate Fig. 2.4A.

Figure 2.S8

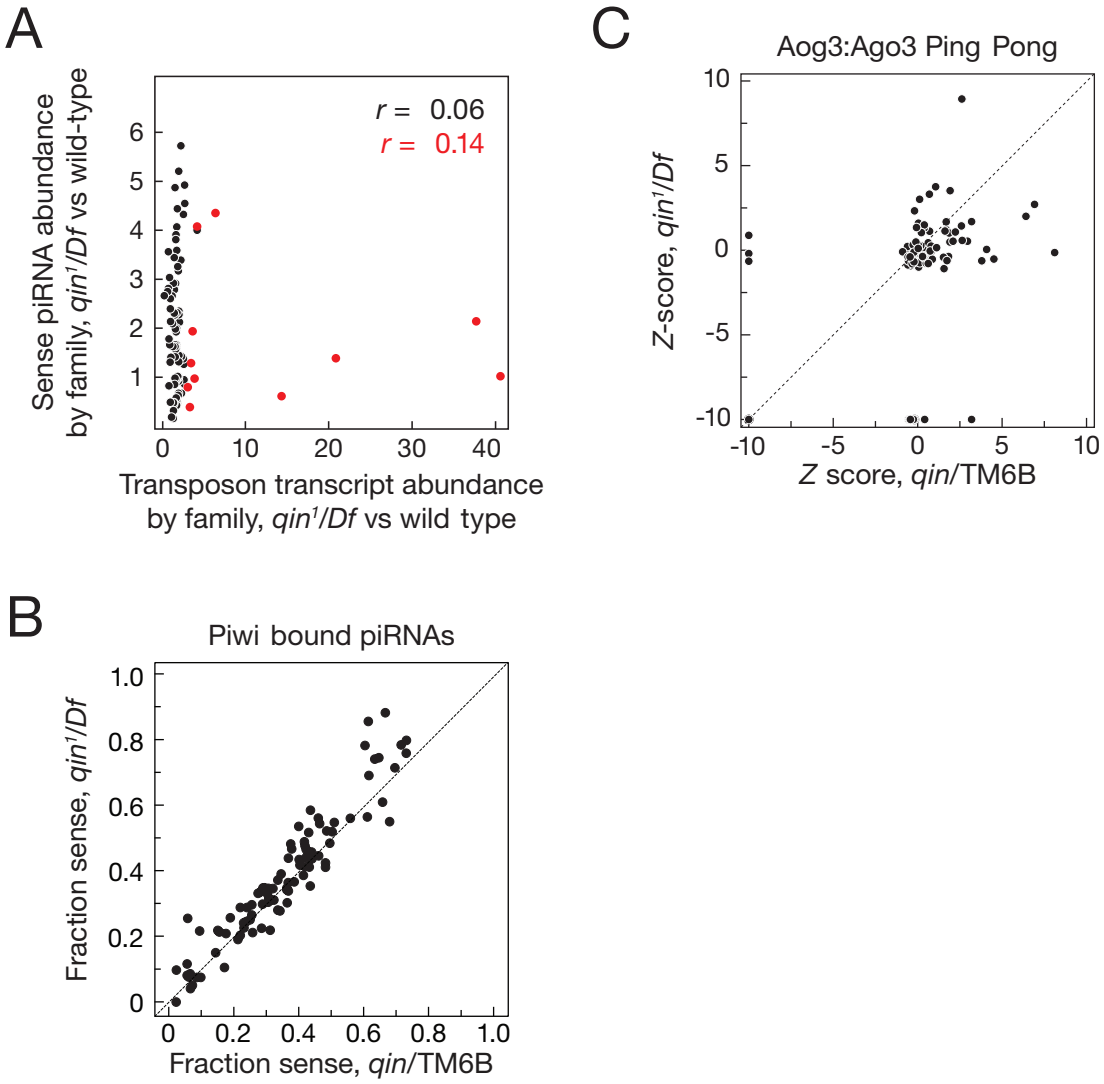
A
Sense piRNA abundance by family, qin1/Df vs wild-type
r = 0.06
r = 0.14
Transposon transcript abundance by family, qin1/Df vs wild type
B
Piwi bound piRNAs
Fraction sense, qin1/Df
Fraction sense, qin/TM6B
C
Aog3:Ago3 Ping Pong
Z-score, qin1/Df
Z score, qin/TM6B

Figure 2.S8.

(A) Scatter plots presenting the change in sense piRNA abundance versus transposon expression, analyzed by transposon family. The 11 significantly over-expressed transposon families are marked in red.

(B) The sense fraction of the Piwi-bound piRNAs for each transposon family was compared between *qin* heterozygotes and qin^1/*Df* mutant ovaries.

(C) Ping-Pong *z*-scores were calculated for Ago3:Ago3 pairs for *qin*/TM6B and qin^1/*Df* ovaries.

Figure 2.S9

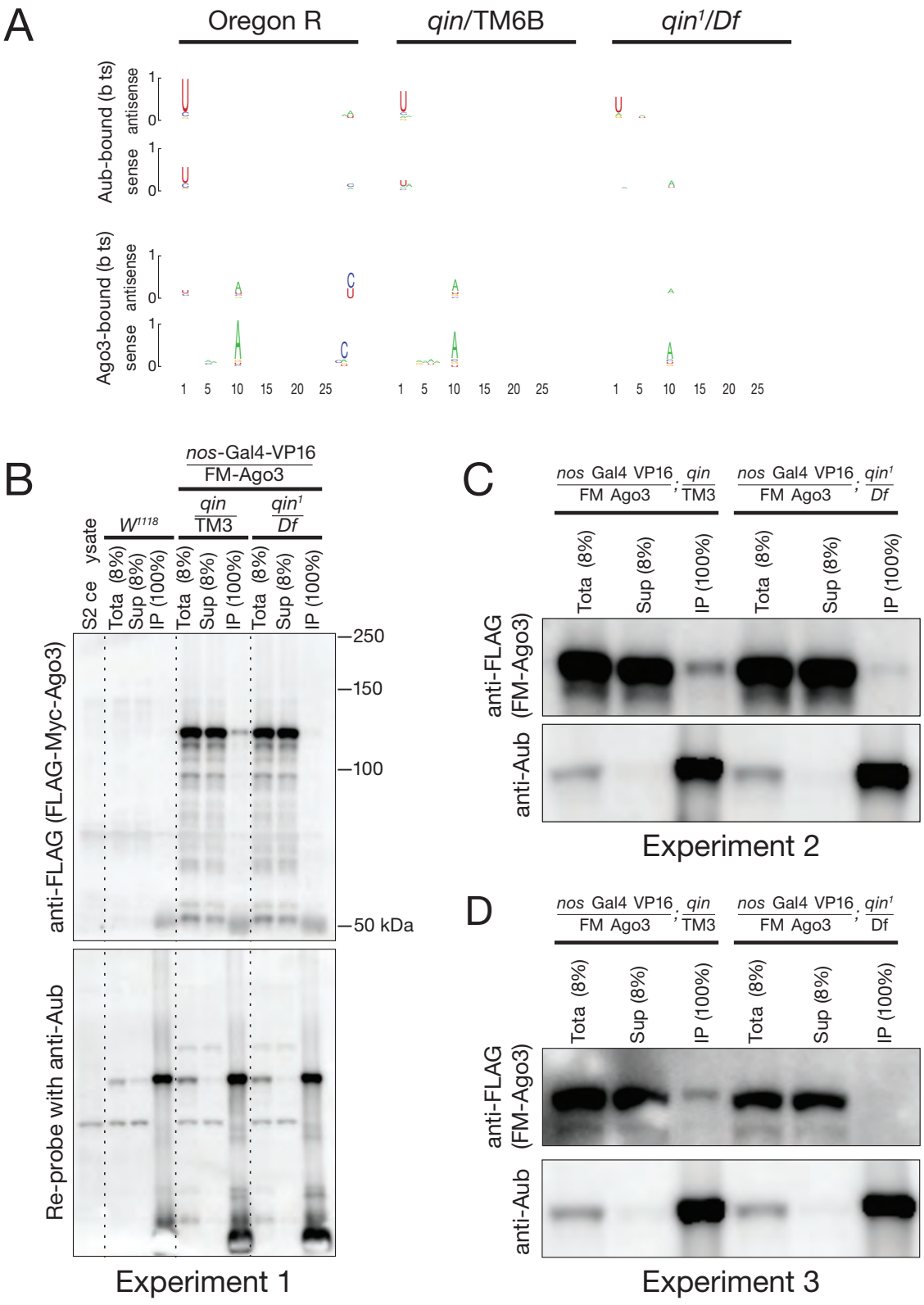
A
Oregon R
qin/TM6B
qin1/Df
Aub-bound (b ts)
sense
antisense
Ago3-bound (b ts)
1 5 10 15 20 25
B
nos-Gal4-VP16
FM-Ago3
W1118
qin/TM3
qin1/Df
S2 cell lysate
Tota (8%)
Sup (8%)
IP (100%)
anti-FLAG (FLAG-Myc-Ago3)
250
150
100
50 kDa
Re-probe with anti-Aub
Experiment 1
C
nos Gal4 VP16 FM Ago3 ; qin/TM3
nos Gal4 VP16 FM Ago3 ; qin1/Df
anti-FLAG (FM-Ago3)
anti-Aub
Experiment 2
D
nos Gal4 VP16 FM Ago3 ; qin/TM3
nos Gal4 VP16 FM Ago3 ; qin1 Df
anti-FLAG (FM-Ago3)
anti-Aub
Experiment 3

Figure 2.S9. Qin maintains Aub and Ago3 co-association.

(A) Sequence logos of Aub- and Ago3-bound total piRNAs.

(B) Complete images of the Western blots presented in Fig. 6D.

(C and D) Additional independent replicates of the experiment in Fig. 6D.

Table 2.S1. Quantitative Phenotypes Observed for *qin* Mutant Females and Their Embryos.

		Dorsal appendages (percent)			Egg laying	
Maternal genotype	***N***	**Wild-type** 	**Fused** 	**Absent** 	**Eggs hatching (percent)**	**Eggs per day per female**
$\frac{qin^1}{\text{TM6B}}$	699	75	21	4.2	68	14
$\frac{qin^1}{qin^1}$	46	0	70	30	0	1
$\frac{qin^1}{Df}$	516	4.3	85	11	0.19	10
$w\,;\frac{\text{P}\{w^{+mC},\ \text{gf-1}\}}{+};\frac{qin^1}{Df}$	349	14	66	20	1.2	7
$w\,;\frac{Sp}{\text{CyO}};\frac{qin^1}{Df}$	482	18	59	23	0.34	13
$w\,;\frac{\text{P}\{w^{+mC},\ nos\text{-Gal4-VP16}\}}{\text{P}\{w^{+mC},\ \text{UASp-FM-Qin}\}};\frac{qin^1}{Df}$	598	69	23	8.7	29	26
$w\,;\frac{\text{P}\{w^{+mC},\ nos\text{-Gal4-VP16}\}}{+};\frac{qin^1}{Df}$	402	6.7	73	20	0.25	16
$w\,;\frac{Sp}{+};\frac{qin^1}{Df}$	618	15	44	42	6.1	9
$w\,;\frac{Sp}{+};\frac{qin^1}{\text{P}\{w^{+mC},\ \text{gf-2}\},Df}$	1025	52	41	6.3	21	29

Table 2.S2. Mutation of *qin* Causes Inappropriate Transposon Expression and Alters piRNA Ping-Pong.

The change in transposon mRNA abundance (mean ± standard deviation for 3 independent experiments) between qin^1/Df and w^1 ovaries was measured by quantitative RT-PCR (A), and Ping-Pong *z*-score (B) was calculated from high throughput sequencing of small RNAs from Aub- or Ago3 immunoprecipitates. Antisense piRNA abundance was calculated from normalized high throughput sequencing data. *P*-values were calculated using Student's unpaired, two-tailed t-test.

A

Transposon family	Change in mRNA abundance	*p*-value	Antisense piRNA abundance relative to control
HeT-A	72 ± 6.2	0.004	14%
Springer	55 ± 4.5	0.002	42%
Burdock	41 ± 5.4	0.006	18%
I element	29 ± 4.6	0.01	21%
copia	23 ± 3.0	0.006	300%
Transpac	14 ± 1.3	0.003	100%
TAHRE	6.3 ± 1.4	0.02	35%
McClintock	3.1 ± 0.08	0.004	64%
jockey	2.7 ± 0.36	0.006	24%
TART	1.7 ± 0.15	0.005	47%
R2 element	1.6 ± 0.06	0.004	29%
micropia	1.2 ± 0.18	N.S.	278%
Max element	0.72 ± 0.01	0.05	67%

B

Transposon family	Ago3:Aub Ping-Pong		Aub:Aub Ping-Pong	
	qin / TM6B	qin^1 / *Df*	*qin* / TM6B	qin^1 / *Df*
I element	38	11	6.7	70
HeT-A	12	8.1	6.6	16
Springer	17	6.2	18	130
Burdock	14	4.9	7.9	28
Transpac	22	14	7.1	8.6
TAHRE	14	5.4	9.9	21

Table 2.S3A. Sequencing statistics: analysis of genome-matching sequences by reads. "Reads excluding ncRNA" correspond to genome-matching reads after excluding annotated non-coding RNAs (ncRNAs) such as rRNA, snRNA, snoRNA, or tRNA. "Transposon-matching reads" correspond to small RNAs mapped to *Drosophila melanogaster* transposons. Because a single read can map to both the sense and to the antisense orientation of a transposon, the sum of sense and antisense transposon-matching reads can be greater than the total number of transposon-matching reads. The numbers in parentheses avoid this discrepancy by apportioning a value of 0.5 to sense and antisense for each read that maps to both orientations. IP, Immunoprecipitate.

Ovary genotype	Sample type	Total reads	Reads perfectly matching genome	ncRNA reads	Reads excluding ncRNA	miRNA-matching reads	Reads excluding ncRNA & miRNA	23–29 nt small RNA reads			
								Total	Transposon-matching reads		
									Total	sense	antisense
qin/TM6B	total	27,315,040	18,703,360	4,083,636	14,619,724	5,833,986	8,785,737	7,524,798	5,558,981 (5,553,421)	1,834,654 (1,786,085)	3,852,821 (3,767,336)
qin^{1}/*Df*	total	29,081,496	18,535,404	2,439,230	16,096,174	6,858,624	9,237,550	7,993,876	6,064,405 (6,054,773)	2,361,385 (2,290,950)	3,868,735 (3,763,823)
qin/TM6B	Ago3 IP	17,530,969	8,814,361	2,640,427	6,173,934	114,424	6,059,510	5,376,206	4,392,867 (4,372,622)	3,372,343 (3,317,483)	1,155,786 (1,055,139)
	Aub IP	7,614,103	5,561,574	220,453	5,341,121	63,787	5,277,334	4,969,621	3,817,326 (3,812,952)	1,246,987 (1,186,581)	2,718,355 (2,626,371)
	Piwi IP	8,145,909	6,556,897	213,633	6,343,264	54,025	6,289,239	6,124,695	4,452,161 (4,431,094)	1,306,256 (1,277,630)	3,198,436 (3,153,463)
qin^{1}/*Df*	Ago3 IP	20,670,615	13,576,418	2,327,968	11,248,450	280,299	10,968,151	9,906,549	8,238,924 (8,212,667)	6,805,366 (6,727,862)	1,632,017 (1,484,806)
	Aub IP	30,714,553	21,565,387	697,030	20,868,357	189,545	20,678,812	19,824,137	15,285,770 (15,257,765)	6,198,355 (5,943,238)	9,711,498 (9,314,526)
	Piwi IP	16,581,308	10,715,339	358,527	10,356,812	113,174	10,243,638	9,952,288	7,543,708 (7,510,426)	2,379,092 (2,317,197)	5,278,014 (5,193,228)

Table 2.S3B. Sequencing statistics: analysis of genome matching sequences by species. "Species excluding ncRNA" correspond to genome-matching species after excluding annotated non-coding RNAs (ncRNAs) such as rRNA, snRNA, snoRNA, or tRNA. "Transposon-matching species" correspond to small RNAs mapped to *Drosophila melanogaster* transposons. Because a single species can map to both the sense and to the antisense orientation of a transposon, the sum of sense and antisense transposon-matching reads can be greater than the total number of transposon-matching reads. The numbers in parentheses avoid this discrepancy by apportioning a value of 0.5 to sense and antisense for each species that maps to both orientations. IP, immunoprecipitate.

Ovary genotype	Sample type	Total species	Species perfectly matching genome	ncRNA species	Species excluding ncRNA	miRNA-matching species	Species excluding ncRNA & miRNA	23–29 nt small RNA species			
								Total	Transposon-matching species		
									Total	sense	antisense
qin/TM6B	total	8,149,381	2,587854	105,322	2,482,532	2,580	2,479,952	1,948,429	1,247,862 (1,237,374)	494,183 (476,902)	779,334 (760,472)
qin/*Df*	total	8,949,604	2,049,973	81,995	1,967,978	2,475	1,965,503	1,520,645	1,056,239 (1,048,669)	454,562 (438,754)	626,525 (609,915)
qin/TM6B	Ago3 IP	6,236,108	189,191	29,854	159,337	254	159,803	126,156	92,757 (91,860)	62,918 (60,645)	33,171 (31,214)
	Aub IP	2,564,509	1,087,079	25,358	1,061,721	601	1,061,120	953,592	668,611 (662,404)	284,252 (272,790)	401,574 (389,614)
	Piwi IP	3,278,136	2,107,167	28,717	2,078,450	620	2,077,830	1,980,141	1,162,524 (1,152,541)	451,234 (437,945)	729,172 (714,596)
qin/*Df*	Ago3 IP	3,786,386	187,064	19,783	167,281	345	166,936	134,817	101,735 (101,025)	74070 (71,683)	31,594 (29,343)
	Aub IP	8,098,256	1,590,856	33,552	1,557,304	777	1,556,527	1,399,273	1,034,143 (1,026,431)	480,253 (461,553)	583,647 (564,878)
	Piwi I.P.	6,499,081	1,999,395	33,013	1,966,382	678	1,965,704	1,847,207	1,147,558 (1,138,388)	465,649 (450,514)	704,124 (687,874)

Table 2.S4. Synthetic DNA oligonucleotides used in this study (5′-to-3′).

For quantitative RT-PCR

Detects	Forward primer; reverse primer
qin	CTG CTG GTT GGA CTA CAC GA; GGA GGT AGA GCG CTC CTT TT
HeT-A	CGC GCG GAA CCC ATC TTC AGA; CGC CGC AGT CGT TTG GTG AGT
blood	TGC CAC AGT ACC TG ATT TCG; GAT TCG CCT TTT ACG TTT GC
ZAM	ACT TGA CCT GGA TAC ACT CAC AAC; GAG TAT TAC GGC GAC TAG GGA TAC
I element	TGA AAT ACG GCA TAC TGC CCC CA; GCT GAT AGG GAG TCG GAG CAG ATA
Rp49	CCG CTT CAA GGG ACA GTA TCT G; ATC TCG CCG CAG TAA ACG C
Burdock	AGG GAA ATA TTT GGC CAT CC; TTT TGG CCC TGT AAA CCT TG
TAHRE	CTG TTG CAC AAA GCC AAG AA; GTT GGT AAT GTT CGC GTC CT
Transpac	GGA ACG CAC CTT CAA CAT TT; GCA AAC TCG CAT TTG TCT GA
Copia	AGC AAA CAA CCC CTC ATG TC; GCA AAC CCA ATT TGT CTC GT
TART	ACC AGG GAA AAG TGT GAA CG; GGT GCA GTG GTA TGG CTT TT
McClintock	CCC TAA TCC GTT TTC CCA AT; CTG GTC GGT TCT GGT CAA AT
Jockey	TCT GCG GTC TCC AGC TTA AT; GTT GGG CAA ATG CTA GTG GT
R2 element	ATG CTC CCG AAA CAA CAA AC; GCA CTG CAG ACT TGG TTC AA
Springer	TGA AGA GCA AGA ACC GGA GT; TCC TCC AGC AAA GCT TGT TT
micropia	CGA ATG TTA CGC GGT GTA TG; CTG GTC AGG TCC AAG GTT GT
Max	ATC TAG CCA GTC GAG GCG TA; TGG AAG AGT GTC GCT TTG TG

For qin[1] genotyping

piggyBac-5R1	TGA CAC TTA CCG CAT TGA CA

piggyBac-3F1	CAA CAT GAC TGT TTT TAA AGT ACA AA
qin^{1}-gt-F	CTT TGA GCA CAA CGA GGA CA
qin^{1}-gt-R	AGG CGC TTC CTT GTC TGA TA

For qin cDNA amplification

qin-attB-F	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GAT TGC CGA CGA CAG TGG AAT AAA G
qin-attB-R	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TTA TGA TTT TTT CAG GAA CAT CCA ATT TTC
$(Tud)_5$-attB-F	GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CGT GGT TCA ACG GGC CAA AC
E3-attB-R	GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CTA CTG TTT GCC ATT GAG GAC ATG AAT AT

For qin 5′ RACE

qin 5′ RACE-1488	CAC GTG CAG CAT CTG AAC CAT GTA G

For Qin::EGFP recombineering

qin-EGFP-F	AGC TAG TGT ATC AAT CTT TAT GTG AAA ATT GGA TGT TCC TGA AAA AAT CAG ATT ATG ATA TTC CAA CTA CTG
qin-EGFP-R	AAA TAG AAA CTG AAA ATG GCT TAC ATC ATA TGC GGC TTT AGT CGT AGC CTC AGA AGA ACT CGT CAA GAA G

To generate Northern hybridization probes

Detects	**Forward primer; reverse primer**
Tudor domains	GAT TTA GGT GAC ACT ATA GAA GTG CAT TTG CGG GAT TTG GGT; TAA TAC GAC TCA CTA TAG GGT CGC ACT TTC GTC GTG TAG TCC AA

E3-ligase domains	ATT TAG GTG ACA CTA TAG AAC TCT CTG CGA GAG CTG CTT GT; TAA TAC GAC TCA CTA TAG GTT GGG CAT TCG GTG ATG GAT GC
Burdock	GAT TTA GGT GAC ACT ATA GAT CGT GAT GTG GTT AAG CCG GAT GT; TAA TAC GAC TCA CTA TAG GGA GGT GTT CTC CCG AGG ATT TGC TT
Rp49	CCA AGC ACT TCA TCC GCC ACC AGT C; TAA TAC GAC TCA CTA TAG GTC CGA CCA CGT TAC AAG AAC TCT CA

CHAPTER III

ANTISENSE PIRNA AMPLIFICATION, BUT NOT piRNA PRODUCTION OR NUAGE ASSEMBLY, REQUIRES THE TUDOR-DOMAIN PROTEIN QIN

PREFACE

The work presented in this chapter was a collaborative effort: Birgit Koppetsch performed the immuno-staining. Jie Wang mapped the sequence reads. I conducted the experiments and analyzed the sequencing data with guidance from Jie Wang and Zhingping Weng.

SUMMARY

PIWI-interacting RNAs (piRNAs) silence transposons and maintain germline genome integrity in the *Drosophila melanogaster* ovary. Wild-type piRNA production requires Qin (also known as Kumo), a protein comprising a RING domain, two B-Box domains and five Tudor domains. Two conflicting roles for Qin in piRNA function have been described. One model proposes that Qin is required to assemble nuage, a perinuclear structure containing many piRNA pathway proteins, and in qin^{kumo} mutant ovaries, germline piRNAs are reported to be lost. An alternative view proposes that Qin coordinates piRNA amplification via reciprocal cycles of RNA cleavage directed by Aubergine- and Ago3-bound piRNAs. Here, we report that in multiple loss-of-function and null mutant *qin* allelic combinations, nuage remains intact, sense piRNAs increase, and antisense piRNAs decrease, reflecting the replacement of wild-type heterotypic with homotypic piRNA "Ping-Pong" cycles. We conclude that Qin acts to ensure the antisense bias of the piRNA amplification machinery.

INTRODUCTION

Qin is required for transposon silencing by the PIWI-interacting RNA (piRNA) pathway (Zhang et al., 2011; Anand and Kai, 2012). Initial descriptions of *qin* mutants led to conflicting explanations for the role of Qin in piRNA biogenesis. One study suggested that loss of Qin causes the accumulation of sense piRNAs instead of antisense without altering total piRNA levels or perturbing the localization of Aub and Ago3 to the perinuclear nuage (Zhang et al., 2011). A second report concluded that both piRNAs and nuage were lost from the germline in *qin* mutants, leading to a complete failure of the piRNA pathway (Anand and Kai, 2012). We re-analyzed the *qin* alleles used in the two studies: qin^{1}, qin^{2} (Zhang et al., 2011) and qin^{kumo} (Anand and Kai, 2012). We conclude that the fundamental defect in *qin* mutants is not a loss of piRNAs, but rather the replacement of heterotypic Aub:Ago3 Ping-Pong with non-productive, homotypic Aub:Aub Ping-Pong. Our data suggest that the phenotypes reported for qin^{kumo} homozygotes are caused by a secondary mutation unlinked to *qin*.

RESULTS AND DISCUSSION

qin^{kumo} likely contains a secondary mutation

Compared with genotypically matched *w*^{1118} and *qin*^{kumo}/TM3 controls, homozygous *qin*^{kumo} mutant ovaries are small, with few egg chambers beyond stage 10 (Supplementary Figure 3.S1). In contrast, *qin*^{1}, *qin*^{2}, *qin*^{kumo} in *trans* to a complete deletion of the *qin* locus (*Df(3R)Excel6180*; henceforth, *Df*), as well as *qin*^{1}/*qin*^{kumo} and *qin*^{2}/*qin*^{kumo}, all had normal ovary size and shape.

qin^{kumo}/*qin*^{kumo} females laid almost no eggs (1 egg per female on day 2, and none thereafter), yet *qin*^{kumo}/*Df* females each laid ~50 eggs per day. Typically, the phenotype of a strong mutant allele remains the same or worsens in *trans* to a deficiency, but *qin*^{kumo}/*qin*^{kumo} was more severe than *qin*^{kumo}/*Df*. Potential explanations include (1) *qin*^{kumo} is a neomorph; (2) the *Df(3R)Excel6180* deficiency fails to uncover the entire *qin* gene; and (3) *qin*^{kumo} contains a secondary mutation unlinked to *qin*.

Our data support the idea that *qin*^{kumo} is a null mutation and that *Df(3R)Excel6180* removes all of *qin*: RNA-seq detected no *qin* mRNA in *qin*^{kumo}/*Df* ovaries (Figure 3.1A). The *qin*^{1} allele results from a *piggyBac* transposon insertion and produces a truncated mRNA 4,432 nt long. As anticipated, *qin*^{1}/*Df* ovaries produced a ~4400 nt RNA less than half as abundant as *qin* mRNA in *w*^{1118} (12 rpkm vs. 32 rpkm). The *qin* deficiency extends beyond the 5′ end of *qin*, disrupting the upstream gene *CG7694*: *CG7694* mRNA abundance was 12 rpkm in *w*^{1118} but only 4.2 rpkm in *qin*^{1}/*Df* and 3.6 rpkm in *qin*^{kumo}/*Df*. We conclude that both *qin*^{kumo} and *Df(3R)Excel6180* are null alleles of *qin* and that a secondary mutation unlinked to *qin* is present on the *qin*^{kumo} chromosome.

Normal nuage in *qin* mutants

Figure 3.1

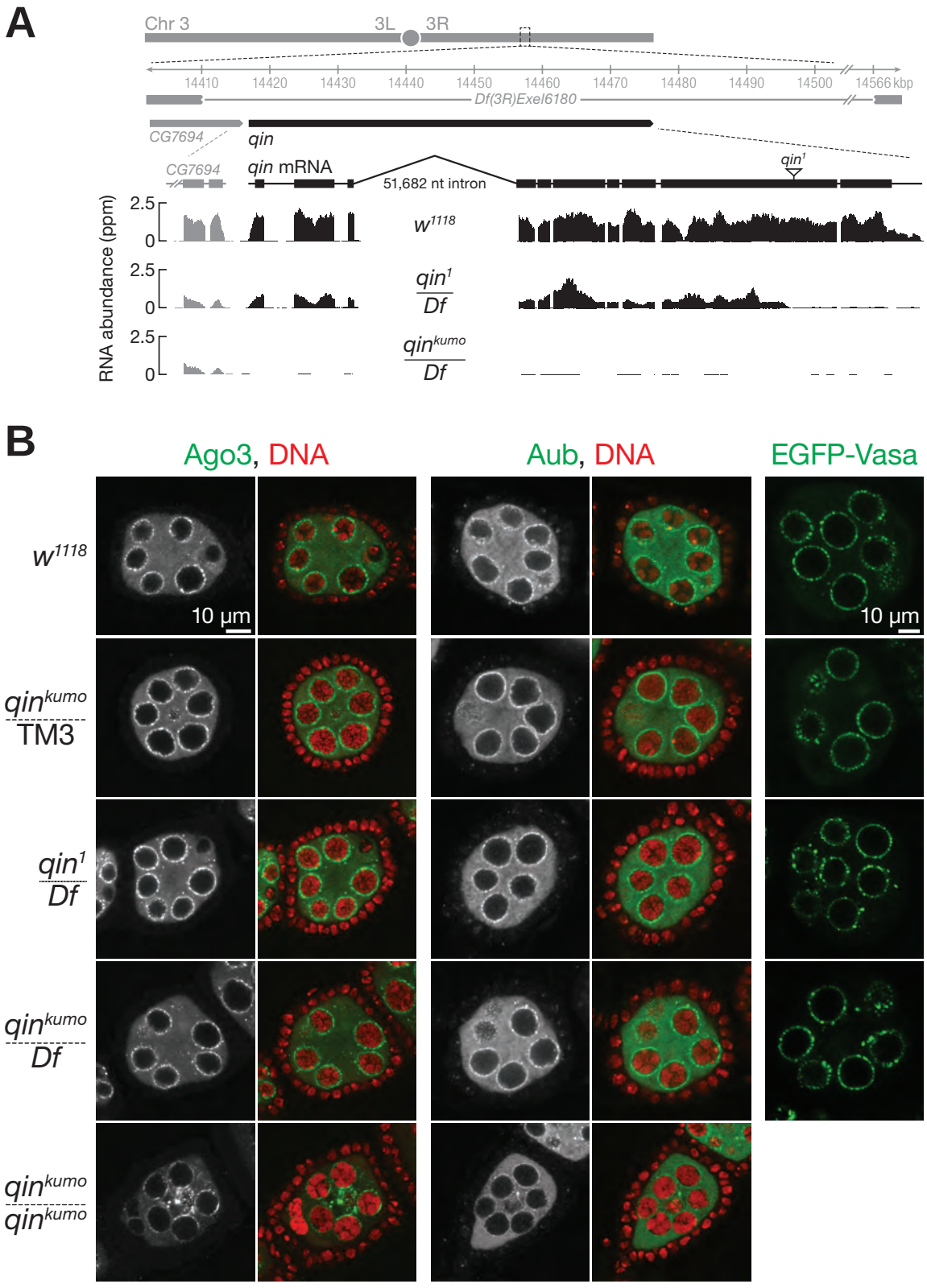
A
Chr 3
3L
3R
14410
14420
14430
14440
14450
14460
14470
14480
14490
14500
14566 kbp
Df(3R)Exel6180
CG7694
qin
CG7694
qin mRNA
51,682 nt intron
qin1
RNA abundance (ppm)
2.5
0
w1118
qin1/Df
qinkumo/Df
B
Ago3, DNA
Aub, DNA
EGFP-Vasa
w1118
qinkumo/TM3
qin1/Df
qinkumo/Df
qinkumo/qinkumo
10 µm
10 µm

Figure 3.1. Without Qin, Ago3, Aub and Vasa still reside in nurse cell nuage.

(A) RNA-seq data for wild-type and *qin* mutant ovaries.

(B) Ago3 and Aub immunostaining or live EGFP-Vasa image in *qin* mutants. EGFP-Vasa fusion protein was expressed from a transgene using the *vasa* promoter.

By immunofluorescence antibody staining, Ago3 and Aub were present in nuage in all genotypes tested except qin^{kumo}/qin^{kumo} (Figure 3.1B). For example, qin^{kumo}/Df, Ago3 was correctly localized to perinuclear foci in 62 of 67 nurse cells among 12 separate egg chambers, compared with 66 of 70 in 16 separate egg chambers from qin^{kumo}/TM3 flies and 96 of 102 in 20 separate egg chambers from w^{1118} ovaries; Aub was present in perinuclear puncta typical of nuage in 50 of 57 nurse cells from nine qin^{kumo}/Df egg chambers compared to 61 of 68 nurse cells from nine qin^{kumo}/TM3 egg chambers and 90 of 98 from 20 w^{1118} egg chambers. Aub localizes to the posterior end of the oocyte in late-stage egg chambers, and this localization was preserved in *qin* mutants (Supplementary Figure 3.S2). In contrast, qin^{kumo} homozygotes showed mislocalized Ago3 and Aub as previously reported (Figure 3.1B; (Anand and Kai, 2012)).

To provide an independent test of whether qin^{kumo}/Df disrupts nuage, we monitored the localization of GFP-Vasa, a nuage marker, in live nurse cells (Figure 3.1B). We detected no disruption of the localization of GFP-Vasa in *qin* mutants. We conclude that loss of Qin does not affect nuage structure in unfixed, living nurse cells.

Loss of Qin has little effect on the total abundance of piRNAs but increases the fraction of sense piRNAs

piRNA levels in qin^{1}/Df ovaries are indistinguishable from controls (Zhang et al., 2011). We used small RNA sequencing to measure piRNA abundance in qin^{2}/Df and qin^{kumo}/Df ovaries. Compared to heterozygotes, the abundance of total transposon-derived, 23–29 nt small RNAs in qin^{2}/Df and qin^{kumo}/Df changed < 4% (Figure 3.2A; Supplementary Tables SI and SII). Among the 93 transposon families with >100 ppm piRNA reads in qin^{1}/TM6B ovaries (Zhang et al., 2011),

there was no significant change in median piRNA abundance relative to w^{1118} controls for qin^{1}/Df ($p = 0.57$, Wilcoxon test), qin^{2}/Df ($p = 0.13$), or qin^{kumo}/Df ($p = 0.33$; Figure 3.2B; for additional analyses by transposon families see http://www.umassmed.edu/uploadedFiles/zamore/Transposon_buckets.zip). However, the fraction of piRNAs with the same orientation as the corresponding transposon sense mRNA increased: the median sense fraction (i.e., sense piRNAs/all piRNAs) among 93 transposon families was 0.25 for w^{1118} ovaries but 0.41 for qin^{1}/Df mutants ($p = 5.2 \times 10^{-7}$, Wilcoxon test), 0.37 for qin^{2}/Df ($p = 2.4 \times 10^{-4}$), and 0.38 for qin^{kumo}/Df ($p = 1.1 \times 10^{-5}$; Figure 2B).

We also measured piRNA abundance in qin^{kumo}/qin^{kumo} and qin^{kumo}/TM3 ovaries (Supplementary Tables 3.SI and 3.SII). qin^{kumo} homozygous mutant flies, like other *qin* loss-of-function mutations, produced amounts of piRNAs similar to their qin^{kumo}/TM3 siblings (Figure 3.2A). Our analysis of previously published deep sequencing data from homozygous qin^{kumo} ovaries (Anand and Kai, 2012) also led us to conclude that there was no change in total piRNA production (Figure 3.2A). Moreover, the effects of qin^{1}/Df and qin^{kumo}/Df on piRNA production were highly correlated ($r = 0.94$), but less well correlated with qin^{kumo}/qin^{kumo} ($r = 0.85$, p-value $< 2.2 \times 10^{-16}$; Figure 3.2C and Supplementary Figure 3.S3A)

All *qin* allelic combinations showed significant ($Z > 46$; p-value $< 2.2 \times 10^{-16}$) Ping-Pong amplification as measured by comparing piRNA pairs overlapping by 10 bp to other lengths of overlap (Figure 3.2D and Supplementary Figure 3.S3B). Finally, we reached these same conclusions when normalizing the data by two alternative strategies—microRNA abundance and non-coding RNA abundance (Supplementary Figure 3.S4, Figure 3.S5 and Figure 3.S6). We conclude that Qin is not required to maintain overall piRNA levels or

Figure 3.2

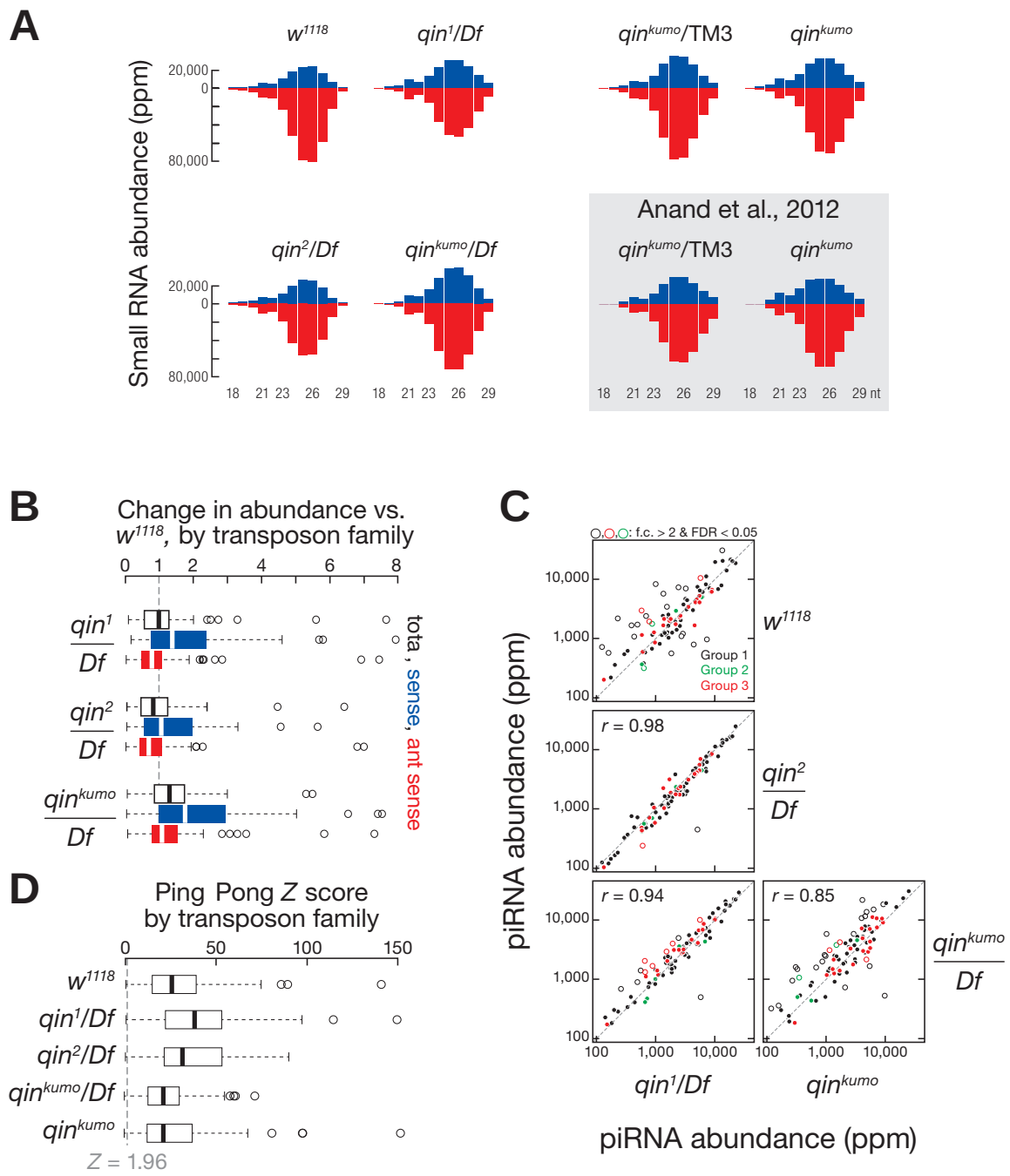
A
w1118
qin1/Df
qinkumo/TM3
qinkumo
Small RNA abundance (ppm)
20,000
0
80,000
qin2/Df
qinkumo/Df
Anand et al., 2012
18 21 23 26 29 nt
B
Change in abundance vs. w1118, by transposon family
0 1 2 4 6 8
qin1/Df
qin2/Df
qinkumo/Df
total, sense, antisense
C
○,○,○: f.c. > 2 & FDR < 0.05
Group 1
Group 2
Group 3
r = 0.98
r = 0.94
r = 0.85
piRNA abundance (ppm)
100
1,000
10,000
D
Ping Pong Z score by transposon family
0 50 100 150
Z = 1.96

Figure 3.2. piRNA abundance and Ping-Pong efficiency are unaltered in *qin* mutant ovaries.

(A) piRNA length distribution. Blue, sense piRNAs; red, antisense.

(B) Box plots reporting the change in abundance of all piRNAs mapping to transposons.

(C) qin^{1}/Df, qin^{2}/Df and qin^{kumo}/Df, but not qin^{kumo}/qin^{kumo}, affect piRNA production similarly. Group 1: transposon families with piRNAs amplified by the Ping-Pong pathway and more antisense piRNAs bound to Aub and more sense piRNAs bound to Ago3. Group 2: transposon families with piRNAs amplified by the Ping-Pong pathway and more sense piRNAs bound to Aub and more antisense piRNAs bound to Ago3. Group 3: transposon families expressed in the somatic follicle cells, predominantly antisense primary piRNAs, and little Ping-Pong amplification.

(D) Box plots reporting Ping-Pong *Z*-score by transposon family. *Z*-score = 1.96 corresponds to *p*-value = 0.05.

for Ping-Pong amplification.

Loss of *qin* leads to accumulation of piRNA cluster transcripts

We used RNA-seq to measure transcript abundance in w^{1118}, qin^{1}/Df, and qin^{kumo}/Df ovaries. Without Qin, RNA sequences mapping uniquely to the 42AB cluster, which is the longest piRNA cluster in flies and produces ~30% of all ovary piRNAs (Brennecke et al., 2007), increased from 1.5 rpkm in w^{1118} flies to 2.0 rpkm in qin^{kumo}/Df and 2.5 in qin^{1}/Df flies (Supplementary Figure 3.S7A). We note that our result disagrees with the finding that transcripts from the 42AB cluster declined in qin^{kumo} homozygous ovaries as measured by qRT-PCR (Anand and Kai, 2012). Among the 142 previously defined piRNA clusters (Brennecke et al., 2007), the steady-state abundance of transcripts from six clusters increased significantly in qin^{1}/Df ovaries (>5-fold; $q < 0.05$); just one decreased significantly (>2-fold; $q < 0.05$; Supplementary Figure 3.S7B). In qin^{kumo}/Df ovaries, the transcript abundance for 11 clusters increased significantly (>5-fold; $q < 0.05$); none decreased significantly (Supplementary Figure 3.S7B).

Increased transposon expression in *qin* mutants

Both qin^{1}/Df and qin^{kumo}/Df ovaries suffered increased transposon expression, as measured by RNA-seq (Supplementary Figure 3.S7B). Of the 93 transposon families we examined, the steady-state RNA abundance of 13 families increased significantly (>6-fold; $q < 0.05$) in qin^{1}/Df, compared with w^{1118}. Similarly, in qin^{kumo}/Df ovaries the steady-state RNA abundance of 12 transposon families increased significantly (>4-fold; $q < 0.05$). Expression of ten transposon families increased significantly in both qin^{1}/Df and qin^{kumo}/Df ovaries ($q < 0.05$), including eight of the 11 transposons whose abundance was reported to increase significantly when measured using both whole-genome tiling microarrays and qRT-PCR (Zhang et al., 2011). Transposon expression in qin^{1}/Df and qin^{kumo}/Df

were highly correlated ($r = 0.95$; p-value $< 2.2 \times 10^{-16}$).

We conclude that loss of *qin* in the fly ovary does not affect nuage assembly or overall piRNA abundance. Instead, loss of Qin leads to an increase in sense piRNAs and a decrease in antisense piRNAs. The result presented here, together with those reported previously (Zhang et al., 2011)are consistent with the loss of heterotypic Aub:Ago3 Ping-Pong in *qin* mutants. Without Qin, piRNA cluster transcripts accumulate, rather than decline. Thus, when Aub:Aub Ping-Pong predominates (Zhang et al., 2011), Ping-Pong amplification appears to consume cluster transcripts less efficiently, consistent with a role for Qin in piRNA precursor processing. Understanding how Qin couples Aub with Ago3 to efficiently generate piRNAs and silence transposons remains a challenge for future studies.

EXPERIMENTAL PROCEDURES

General Methods

RNA isolation, small RNA library construction, sequencing data analysis and Immunofluorescent antibody staining were as described (Zhang et al., 2011). Transposon families were grouped as described (Li et al., 2009). Figures were generated using R, Excel (Microsoft, Redmond, WA, USA), IgorPro (WaveMetrics, Lake Oswego, OR, USA), Adobe Photoshop and Illustrator (Adobe systems, San Jose CA, USA). Ovary small RNA sequencing data sets previously deposited in the NCBI trace archives were GSM872307 (w^{1118} ovaries; (Zhang et al., 2012a)), SRP007101 (qin^{1}/TM3 and qin^{1}/*Df* ovaries; (Zhang et al., 2011)), and GSE34728 (qin^{kumo}/TM3 and qin^{kumo}; (Anand and Kai, 2012)). Flies were reared at 25°C. PBac(RB)CG14303^{e03728} (qin^{1}) and *Df* flies were from the Bloomington Stock Center (Indiana University); PBac(RB)e01936 (qin^{2}) was from the Harvard Medical School Stock Center. qin^{kumo} was previously described as $kumo^{M41\text{-}13}$ (Anand and Kai, 2012).

RNA-seq

Strand-specific RNA-Seq libraries were prepared as described (Zhang et al., 2012b) and sequenced using the 100-nt paired-end protocol on a HiSeq 2000 (Illumina). RNA-seq reads were aligned to the *Drosophila melanogaster* genome (FlyBase r5.45/dm3) using TopHat 2.0.4 (Trapnell et al., 2009), using the options "--bowtie1 --transcriptome-mismatches 2 --genome-read-mismatches 2 --segment-length 50 --segment-mismatches 1 -r 800 -i 50 --solexa1.3-quals --coverage-search." BEDTools were used to tally reads mapped to genes,

transposons, or piRNA cluster transcripts (Quinlan and Hall, 2010). Data was normalized to the sum of the reads in the top quartile of expressed genes. DESeq (Anders and Huber, 2010) was used to detect changes in transcript abundance and calculate *q*-values. Rpkm calculations used a pseudo count of 0.001.

Accession numbers

Sequence data generated in this study are available via the NCBI trace archives (http://www.ncbi.nlm.nih.gov/Traces/) using accession number SRP024291.

ACKNOWLEDGEMENTS

We thank Toshie Kai for sharing *qin*^{kumo} flies, and Paul Lasko for sharing EGFP-Vasa flies. We thank Shikui Tu, Xiaopeng Zhu and members of our laboratories for advice, suggestions, and critical comments on the manuscript. This work was supported in part by National Institutes of Health grant HD049116 to WET, ZW, and PDZ and GM62862 and GM65236 to PDZ.

Figure 3.S1

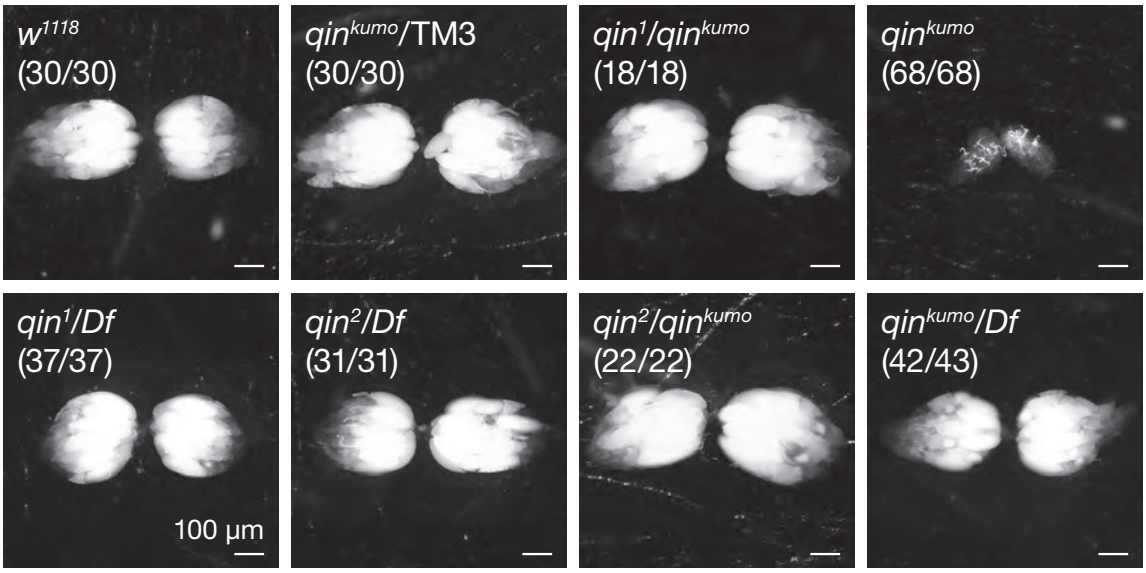
w^{1118}
(30/30)
qin^{kumo}/TM3
(30/30)
qin^{1}/qin^{kumo}
(18/18)
qin^{kumo}
(68/68)
qin^{1}/Df
(37/37)
100 µm
qin^{2}/Df
(31/31)
qin^{2}/qin^{kumo}
(22/22)
qin^{kumo}/Df
(42/43)

Figure 3.S1. All combinations of *qin* mutations, except qin^{kumo}/qin^{kumo}, have normal ovary size and shape.

Figure 3.S2

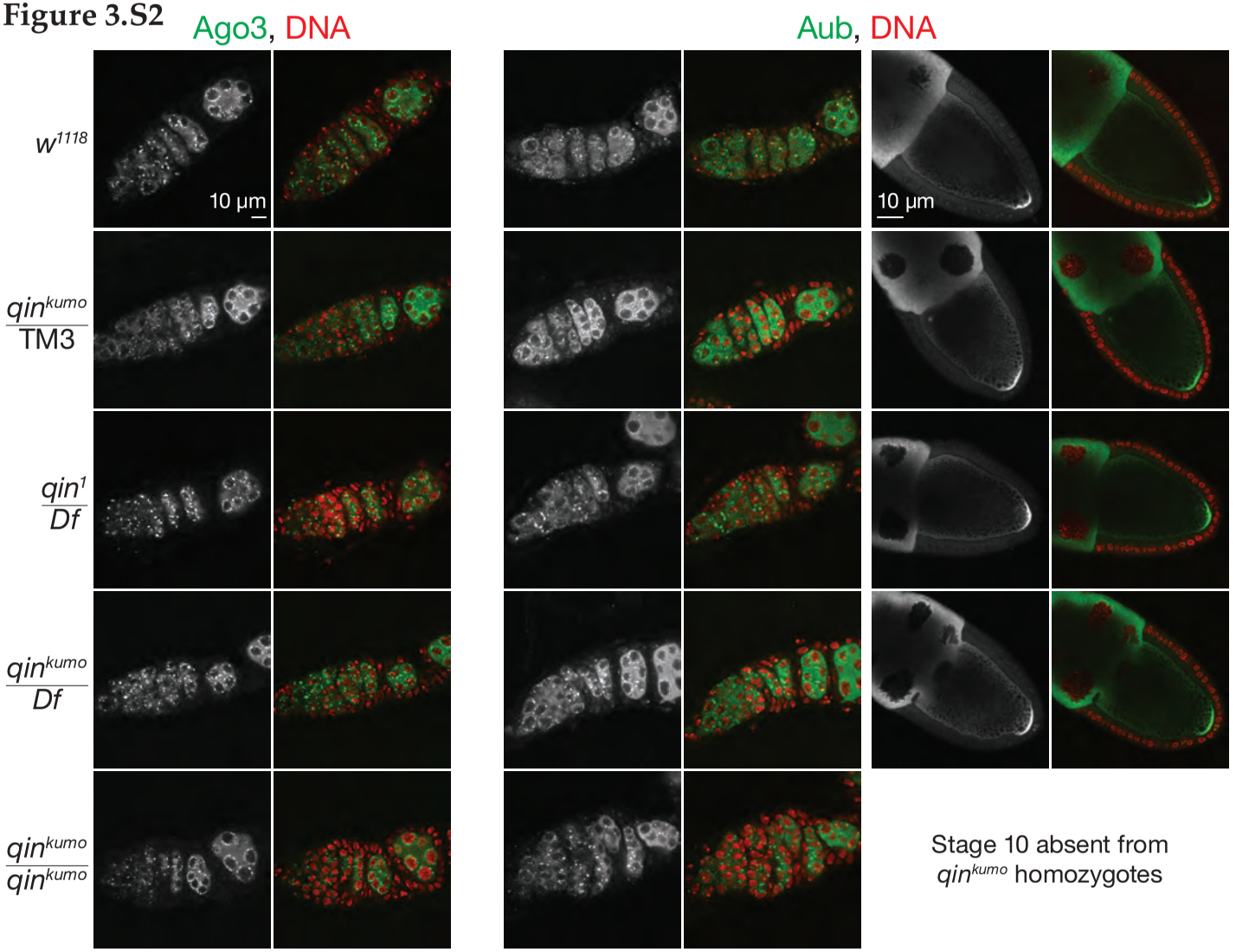
Ago3, DNA
Aub, DNA
w1118
qinkumo / TM3
qin1 / Df
qinkumo / Df
qinkumo / qinkumo
10 µm
10 µm
Stage 10 absent from qinkumo homozygotes

Figure 3.S2. Mutations in *qin* do not disrupt Ago3 and Aub localization.

Figure 3.S3

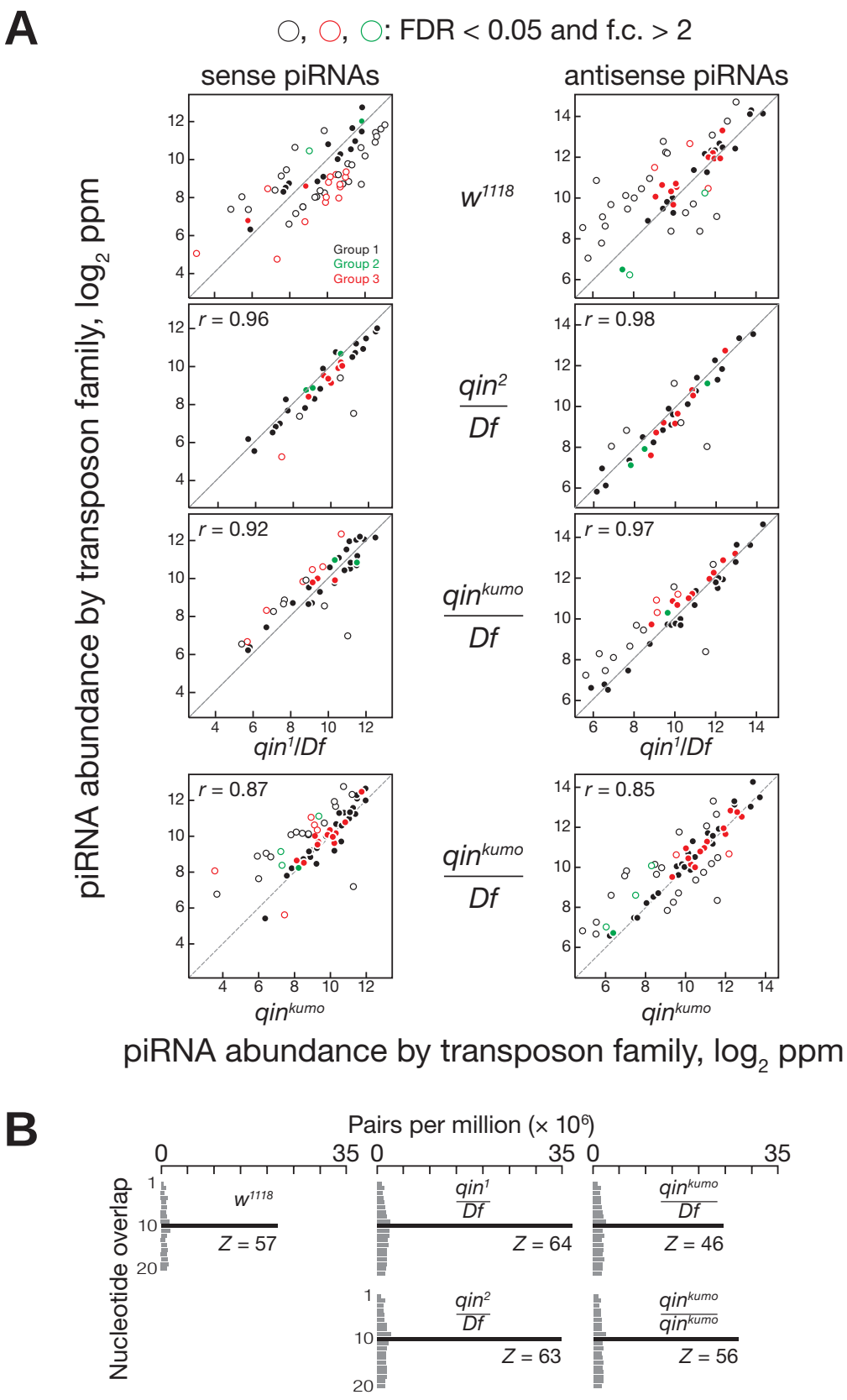
A
○, ○, ○: FDR < 0.05 and f.c. > 2
sense piRNAs
antisense piRNAs
piRNA abundance by transposon family, log2 ppm
w1118
qin2/Df
qin1/Df
qinkumo/Df
qinkumo/qinkumo
Group 1
Group 2
Group 3
r = 0.96
r = 0.98
r = 0.92
r = 0.97
r = 0.87
r = 0.85
piRNA abundance by transposon family, log2 ppm
B
Pairs per million (× 10^6)
Nucleotide overlap
Z = 57
Z = 64
Z = 46
Z = 63
Z = 56

Figure 3.S3.

(A) Scatter plots analyzing sense or antisense piRNA abundance by transposon family. Group 1: transposon families with piRNAs amplified by the Ping-Pong pathway and more antisense piRNAs bound to Aub and more sense piRNAs bound to Ago3. Group 2: transposon families with piRNAs amplified by the Ping-Pong pathway and more sense piRNAs bound to Aub and more antisense piRNAs bound to Ago3. Group 3: transposon families expressed in the somatic follicle cells, predominantly antisense primary piRNAs, and little Ping-Pong amplification.

(B) Normalized number of piRNA Ping-Pong pairs (pairs per million piRNAs) and *Z*-score for 10 nt overlap between piRNAs. *Z*-score = 1.96 corresponds to *p*-value = 0.05.

Figure 3.S4

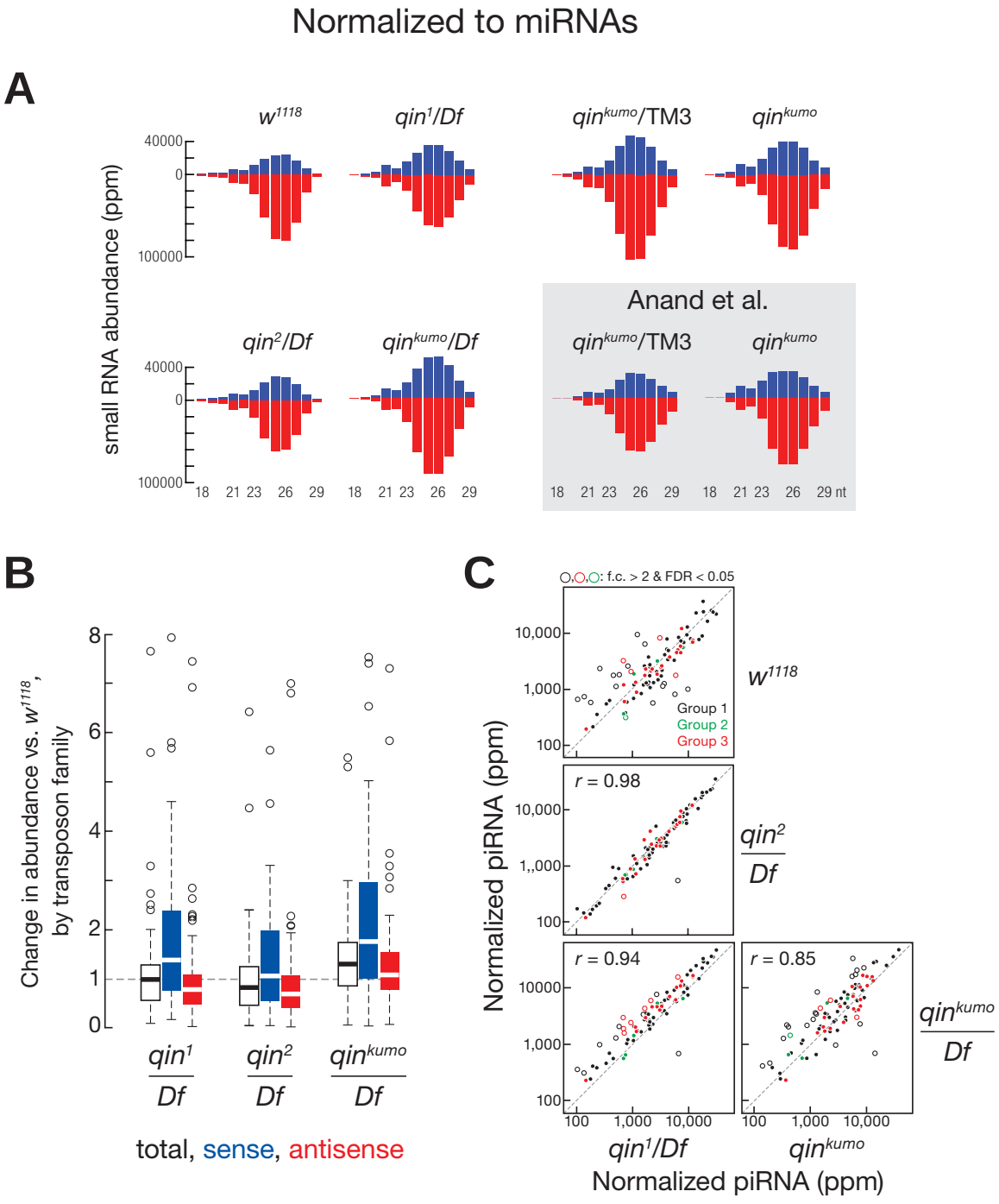
Normalized to miRNAs
A
w1118
qin1/Df
qinkumo/TM3
qinkumo
qin2/Df
qinkumo/Df
Anand et al.
qinkumo/TM3
qinkumo
small RNA abundance (ppm)
40000
0
100000
18 21 23 26 29 nt
B
Change in abundance vs. w1118, by transposon family
qin1/Df
qin2/Df
qinkumo/Df
total, sense, antisense
C
○,○,○: f.c. > 2 & FDR < 0.05
Group 1
Group 2
Group 3
r = 0.98
r = 0.94
r = 0.85
Normalized piRNA (ppm)
qin1/Df
qinkumo
Normalized piRNA (ppm)

Figure 3.S4. Relative piRNA abundance, normalized to microRNA abundance.

(A) piRNA length distribution. Blue, sense piRNAs; red, antisense.

(B) Box plots reporting the change in abundance of all piRNAs mapping to transposons.

(C) qin^{1}/Df, qin^{2}/Df and qin^{kumo}/Df, but not qin^{kumo}/qin^{kumo}, affect piRNA production similarly. Transposons were grouped as Figure S3A.

Figure 3.S5

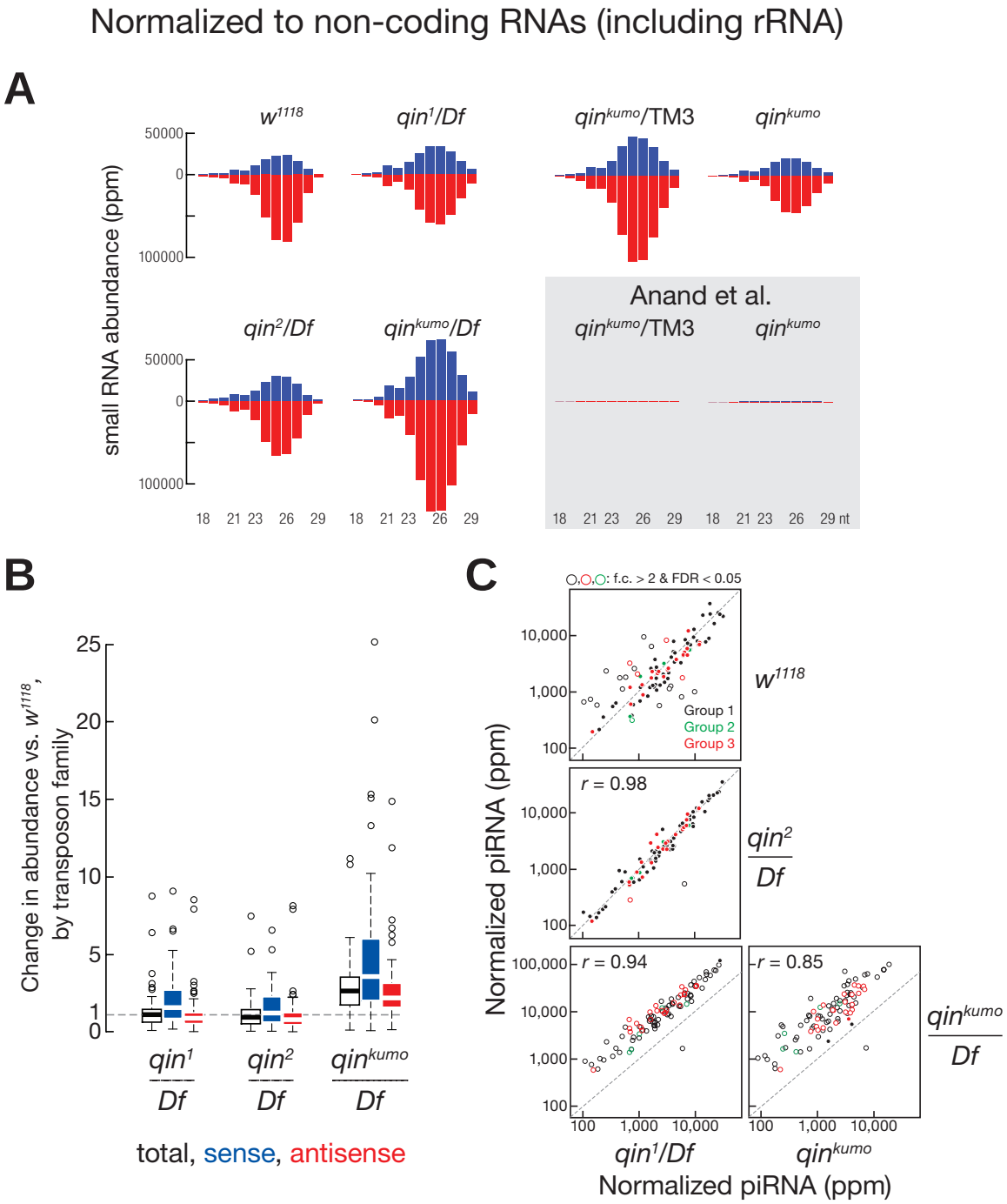
Normalized to non-coding RNAs (including rRNA)
A
small RNA abundance (ppm)
w1118
qin1/Df
qinkumo/TM3
qinkumo
qin2/Df
qinkumo/Df
Anand et al.
qinkumo/TM3
qinkumo
50000
0
100000
18 21 23 26 29 nt
B
Change in abundance vs. w1118, by transposon family
qin1/Df
qin2/Df
qinkumo/Df
total, sense, antisense
C
○,○,○: f.c. > 2 & FDR < 0.05
Group 1
Group 2
Group 3
r = 0.98
r = 0.94
r = 0.85
Normalized piRNA (ppm)
qin1/Df
qinkumo/Df
Normalized piRNA (ppm)

Figure 3.S5. Relative piRNA abundance, normalized to non-coding RNA (ncRNA) abundance, including 2S ribosomal RNA.

(A) piRNA length distribution. The normalization procedure failed for the Anand et al., 2012 dataset because > 89% of its sequence reads are 2S rRNA. Blue, sense piRNAs; red, antisense.

(B) Box plots reporting the change in abundance of all piRNAs mapping to transposons.

(C) qin^{1} / *Df*, qin^{2} / *Df* and qin^{kumo} / *Df,* but not qin^{kumo} / qin^{kumo}, affect piRNA production similarly. Transposons were grouped as Figure S3A.

Figure 3.S6

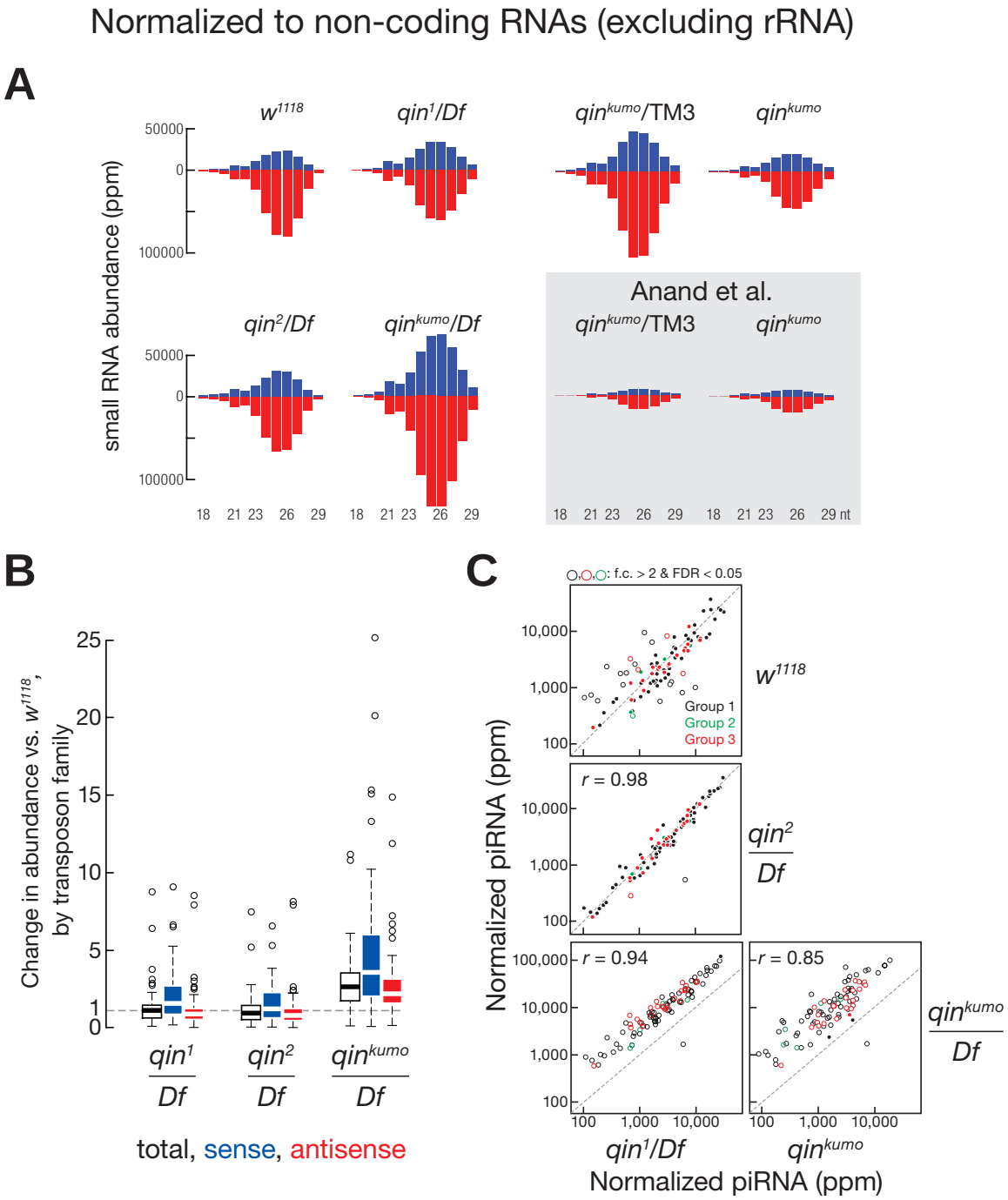
Normalized to non-coding RNAs (excluding rRNA)
A
small RNA abundance (ppm)
w1118
qin1/Df
qinkumo/TM3
qinkumo
qin2/Df
qinkumo/Df
Anand et al.
18 21 23 26 29 nt
B
Change in abundance vs. w1118, by transposon family
qin1/Df
qin2/Df
qinkumo/Df
total, sense, antisense
C
○,○,○: f.c. > 2 & FDR < 0.05
Group 1
Group 2
Group 3
r = 0.98
r = 0.94
r = 0.85
Normalized piRNA (ppm)
qin1/Df
qinkumo/Df

Figure 3.S6. Relative piRNA abundance, normalized to non-coding RNA (ncRNA) abundance, excluding 2S ribosomal RNA.

(A) piRNA length distribution. Blue, sense piRNAs; red, antisense.

(B) Box plots reporting the change in abundance of all piRNAs mapping to transposons.

(C) qin^{1}/Df, qin^{2}/Df and qin^{kumo}/Df, but not qin^{kumo}/qin^{kumo}, affect piRNA production similarly. Transposons were grouped as Figure S3A.

Figure 3.S7

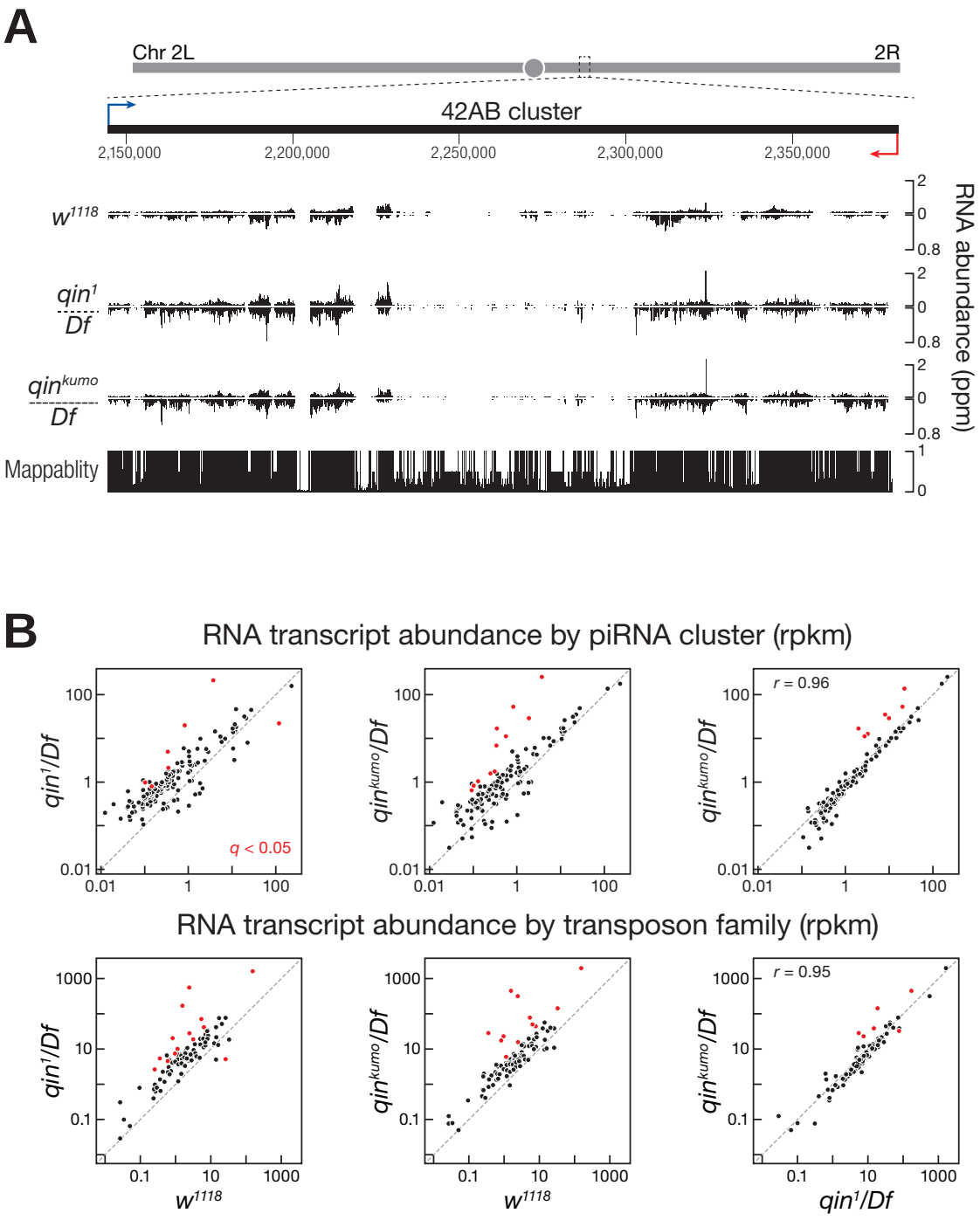
A
Chr 2L
2R
42AB cluster
2,150,000
2,200,000
2,250,000
2,300,000
2,350,000
w1118
qin1/Df
qinkumo/Df
Mappablity
RNA abundance (ppm)
B
RNA transcript abundance by piRNA cluster (rpkm)
qin1/Df
qinkumo/Df
q < 0.05
r = 0.96
RNA transcript abundance by transposon family (rpkm)
r = 0.95
w1118

Figure 3.S7. Loss of Qin leads to piRNA cluster transcript accumulation and increased transposon expression.

(A) RNA-seq data for transcripts from the 42AB piRNA cluster.

(B) piRNA cluster transcript and transposon RNA abundance measured by RNA-seq for *qin* mutants and wild-type ovaries.

Table 3.S1. Small RNA sequencing statistics: analysis of genome matching sequences by reads. "Reads excluding ncRNA" correspond to genome-matching reads after excluding annotated non-coding RNAs (ncRNAs) such as rRNA, snRNA, snoRNA, or tRNA. "Transposon-matching reads" correspond to small RNAs mapped to *Drosophila melanogaster* transposons.

Ovary genotype	Sample source	Total reads	Reads perfectly matching genome	ncRNA reads	Reads excluding ncRNA	miRNA-matching reads	Reads excluding ncRNA & miRNA	23–29 nt small RNA reads			
								Total	Transposon-matching reads		
									Total	sense	antisense
$\frac{qin^2}{TM6B}$	This work	19,940,733	17,280,375	2,053,821	1,5226,554	7,182,278	8,044,276	6,971,096	5,323,826	1,504,426	3,957,938
$\frac{qin^2}{Df}$	This work	19,104,937	16,682,386	2,164,562	14,517,824	7,196,288	7,321,536	6,242,381	4,911,725	1,654,680	3,410,405
$\frac{qin^{kumo}}{Df}$	This work	39,195,465	32,468,674	2,552,780	29,915,894	9,726,714	20,189,180	18,152,468	14,592,848	5,440,035	9,526,468
qin^{kumo}	This work	29,783,737	24,652,714	5,333,312	19,319,402	6,023,849	13,295,553	11,659,819	9,033,362	3,098,773	6,175,719
$\frac{qin^{kumo}}{TM3}$	This work	34,312,243	28,178,297	3,181,630	24,996,667	7,338,462	17,658,205	15,761,440	12,030,827	3,984,677	8,304,302
qin^{kumo}	Anand et al., 2011	19,646,773	18,177,993	16,872,510	1,305,483	433,951	871,532	753,915	578,408	187,591	406,744
$\frac{qin^{kumo}}{TM3}$	Anand et al., 2011	19,257,310	17,419,162	16,690,006	729,156	280,889	448,267	396,484	308,896	102,496	212,096
w^{1118}	Zhang et al., 2012	18,327,610	15,285,481	2,262,297	13,023,184	5,129,288	7,893,896	6,968,295	5,457,206	1,361,432	4,211,072

Table 3.S2. Small RNA sequencing statistics: analysis of genome matching sequences by species. "Species excluding ncRNA" correspond to genome-matching species after excluding annotated non-coding RNAs (ncRNAs) such as rRNA, snRNA, snoRNA, or tRNA. "Transposon-matching species" correspond to small RNAs mapped to *Drosophila melanogaster* transposons.

Ovary genotype	Source	Total species	Species perfectly matching genome	ncRNA species	Species excluding ncRNA	miRNA-matching species	Species excluding ncRNA & miRNA	23–29 nt small RNA species			
								Total	Transposon-matching species		
									Total	sense	antisense
$\frac{qin^2}{TM6B}$	This work	2,586,830	1,630,181	68,236	1,561,945	2,233	1,559,712	1,263,308	852,522	315,083	556,029
$\frac{qin^2}{Df}$	This work	2,111,801	1,306,140	65,374	1,240,766	2,211	1,238,555	962,855	689,011	289,042	416,680
$\frac{qin^{kumo}}{Df}$	This work	5,911,322	3,213,542	81,002	3,132,540	2,739	3,129,801	2,641,238	1,767,540	782,655	1,023,736
qin^{kumo}	This work	3,400,976	1,687,991	84,030	1,603,961	2,328	1,601,633	1,298,229	856,706	344,771	530,663
$\frac{qin^{kumo}}{TM3}$	This work	5,016,587	2,724,858	76,646	2,648,212	2,448	2,645,764	2,248,510	1,435,861	589,379	875,570
qin^{kumo}	Anand et al., 2012	584,778	425,024	30,791	394,233	947	393,286	317,933	217,364	81,627	140,273
$\frac{qin^{kumo}}{TM3}$	Anand et al., 2012	379,455	260,852	15,424	245,428	701	244,727	209,031	148,679	57,054	94,201
w^{1118}	Zhang et al., 2012	1,896,565	842,586	50,736	791,850	1,672	790,178	650,946	448,232	153,152	304,312

Table 3.S3. RNA sequencing statistics.

Ovary genotype	Total fragments	Genome-mapping fragments	Uniquely mapping fragments	Gene-mapping fragments	Transposon-mapping fragments	piRNA cluster-mapping fragments
W^{1118}	101,751,370	94,970,465	86,796,266	84,763,014	227,148	99,350
$\frac{qin^{1}}{Df}$	84,783,884	76,157,690	68,096,515	62,451,246	974,764	256,528
$\frac{qin^{kumo}}{Df}$	85,225,286	77,584,319	66,031,800	60,919,968	1,390,000	352,267

CHAPTER IV

RHINO ANCHORS A NUCLEAR COMPLEX THAT SUPPRESSES piRNA PRECURSOR SPLICING

PREFACE

The work presented in this chapter was a collaborative effort: Nadine Schultz constructed the plasmids for fly transformation and balanced the corresponding flies. Jie Wang mapped the high throughput sequencing reads and performed the splicing analysis. Swapnil Parhad cloned two small RNA libraries for Figure 4.S6. I did all of the rest of the experiments and analysis.

SUMMARY

piRNAs guide an adaptive genome defense system that specifically silences transposons in the germline. Differentiating piRNA precursors from mRNAs is critical to this specificity. The *Drospohila* HP1 homolog Rhino is required for piRNA production. We show that Rhino binds to the dual strand clusters that produce germline piRNA precursors, and that binding directly correlates with piRNA production. Paired-end RNA sequencing indicates that most piRNA precursors are not spliced. However, UAP56 and Cuff encode a DEAD box protein and Rai1 homolog that co localize with Rhi, and *rhino, cuff* and *uap56* mutations that block piRNA production lead to efficient cluster transcript splicing at novel donor and acceptor sites. Moreover, LacI::Rhino fusion protein binding suppresses splicing of a reporter transgene, and is sufficient to trigger *de novo* piRNA production from a trans combination of sense and antisense transgenes. Rhino thus anchors a nuclear complex that suppresses cluster transcript splicing, which may differentiate piRNA precursors from mature mRNAs.

INTRODUCTION

Transposons and other repetitive elements are major genome constituents that can mobilize and induce DNA breaks and insertional mutations (McClintock, 1950; Bennetzen, 2000; Beck et al., 2010). In the germline, which transmits the inherited genetic complement, PIWI-interacting RNAs (piRNAs) suppress transposon expression and maintain genome integrity (Khurana and Theurkauf, 2010; Siomi et al., 2011; Guzzardo et al., 2013). The 23-30 nt long piRNAs bear a 5´ monophosphate, a 3' terminal 2´-O-methyl group, and bind to PIWI clade Argonautes (Aubergine, Piwi and Ago3 in *Drosophila*) (Grivna et al., 2006; Girard et al., 2006; Aravin et al., 2006; Vagin et al., 2006; Lau et al., 2006; Horwich et al., 2007; Saito et al., 2007). piRNAs bound to PIWI proteins can guide sequence specific cleavage of complementary targets, which contributes to transposon silencing and generates the precursors of sense strand piRNAs that direct cleavage of antisense precursors and drive a "ping-pong amplification" cycle (Brennecke et al., 2007; Gunawardane et al., 2007).

The primary piRNAs that initiate this cycle are derived from large "piRNA clusters" composed of nested transposon fragments that generally reside in subtelomeric or pericentromeric heterochromatin (Brennecke et al., 2007). The majority of *Drosophila* clusters produce piRNAs from both genomic strands, and piRNAs mapping uniquely to these dual-strand clusters are the dominant species in germline cells. However, a subset of clusters produce unique piRNAs from only one genomic strand (uni-strand clusters). piRNAs in the somatic follicle cells that surround the germline are derived from uni-strand clusters (Brennecke et al., 2007; Malone et al., 2009). How transcripts from dual-strand

and uni-strand clusters are distinguished from gene transcripts is not understood.

The rapidly evolving Heterochromatin Protein 1 (HP1) homolog Rhino (Rhi) is specifically required for production of primary piRNAs from germline specific dual-strand clusters (Klattenhoff et al., 2009). Here we show that Rhi binds specifically to clusters and that binding correlates with piRNA production. Significantly, we also present evidence that Rhi functions with the Rai1 related protein Cutoff (Cuff) and the DEAD box protein UAP56 to suppress piRNA precursor splicing at specific donor and acceptor sites, and that tethering a LacI::Rhi fusion to an intron containing transgene suppresses splicing and leads to Rhi spreading through the transcription unit. Tethering LacI::Rhi to a construct producing a single strand does not trigger piRNA production, but Rhino binding to a trans combination of sense and antisense transgenes leads to piRNA production from both genomic strands. We therefore propose that Rhino anchors a nuclear complex that suppresses splicing and directs the resulting unspliced RNAs to the piRNA biogenesis machinery. Recent studies indicate that stalled splicing intermediates are the precursors of transposon-silencing siRNAs in the pathogenic yeast *Cryptococcus* (Dumesic et al., 2013). Suppressed splicing may therefore have a conserved function in differentiating potentially deleterious transcripts from mature mRNAs, and direct these transcripts to the small silencing RNA biogenesis machinery.

RESULTS

Rhino marks dual-strand piRNA clusters

The HP1 family protein Rhino (Rhi), also referred to as HP1D, is required for piRNA production from dual-strand clusters, and earlier ChIP-qPCR experiments showed that Rhi associates with two regions in the major 42AB piRNA cluster (Klattenhoff et al., 2009). To determine the specificity of Rhi binding across the genome, we performed chromatin immunoprecipitation sequencing (ChIP-Seq) using anti-Rhi antibody and preimmune control antibody. A comparison of the unique ChIP-Seq signal and unique piRNA profiles revealed a striking correlation (Figure 4.1A and 4.1B), and confirmed that Rhi shows only background-level binding to uni-strand clusters, which dominate piRNA production in somatic follicle cells (Figure 4.1C and 4.S1A). To quantify the relationship between Rhi binding and piRNA production, we plotted fold enrichment for Rhi binding to chromatin by ChIP-Seq against fold reduction in piRNA production in *rhi* mutants for 142 piRNA clusters (Brennecke et al., 2007). Eleven piRNA clusters, including the two major uni-strand clusters (Cluster 2 and *flamenco*), show no decrease in piRNA expression in $rhi^{2/KG}$ mutants and show only background Rhi binding by ChIP-Seq (Figure 4.1C and 4.S1). Over the remaining 131 piRNA clusters, by contrast, Rhi binding was significantly correlated with fold reduction in piRNA expression in *rhi* mutants (Pearson correlation coefficient $r = 0.74$; $P < 2.2\text{x}10^{-16}$). ChIP-Seq signal of a matched pre-immune serum control was not correlated with piRNA expression ($r = 0.19$, $P = 0.03$) (Figure 4.1C).

The correlation between Rhi binding and piRNA signal raised the

Figure 4.1

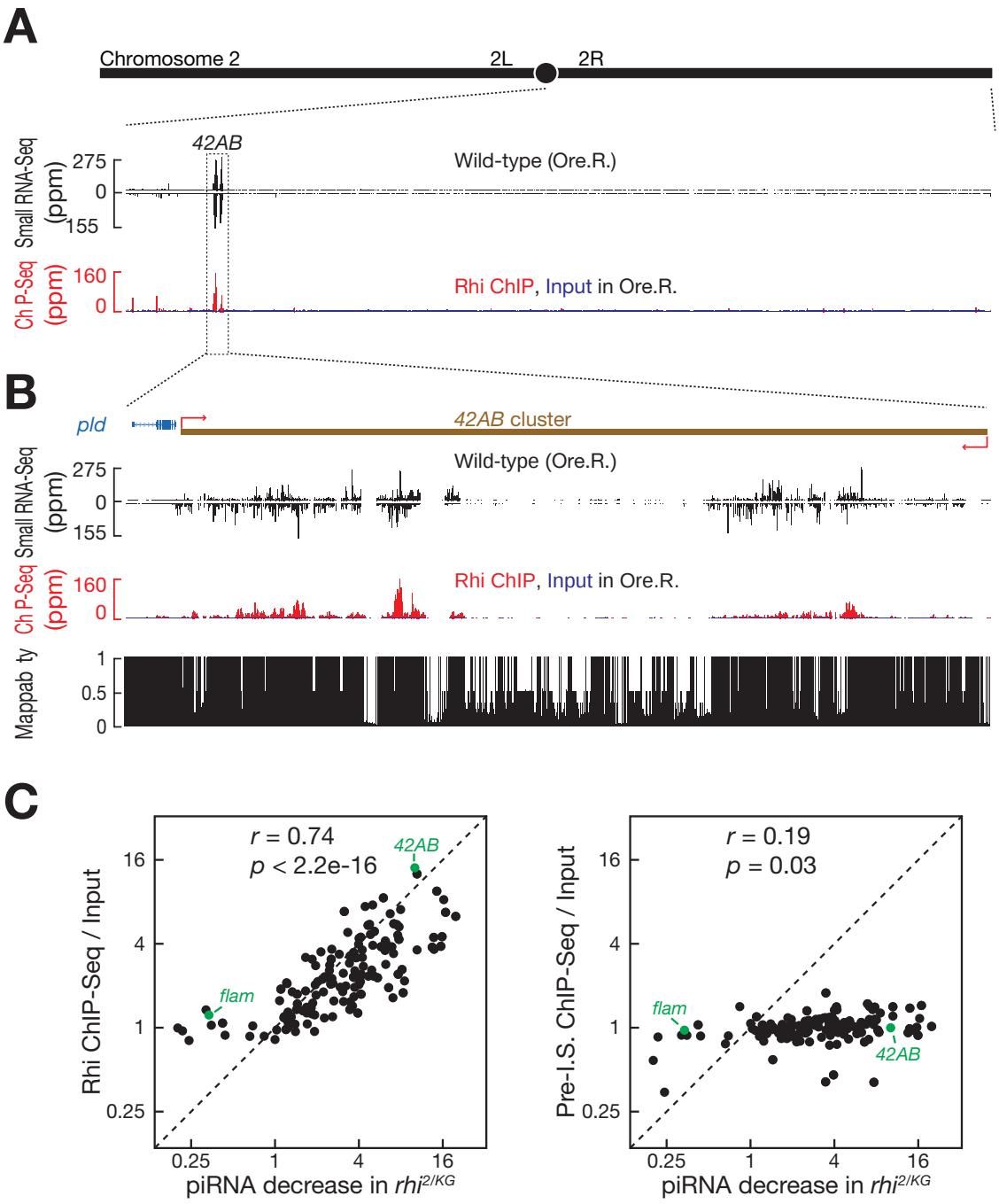
A
Chromosome 2
2L
2R
42AB
Small RNA-Seq (ppm)
275
0
155
Wild-type (Ore.R.)
Ch P-Seq (ppm)
160
0
Rhi ChIP, Input in Ore.R.
B
pld
42AB cluster
Wild-type (Ore.R.)
Rhi ChIP, Input in Ore.R.
Mappab ty
1
0.5
0
C
r = 0.74
p < 2.2e-16
42AB
flam
Rhi ChIP-Seq / Input
piRNA decrease in rhi2/KG
r = 0.19
p = 0.03
Pre-I.S. ChIP-Seq / Input
0.25
1
4
16

Figure 4.1. Rhi binding correlates with piRNA production.

(A) piRNA and Rhi ChIP-Seq signal across the right arm of chromosome 2.

(B) piRNA and Rhi ChIP-Seq signal across the 42AB piRNA cluster, which produces ~ 30% of fly ovary piRNAs.

(C) Rhi binding correlates with Rhi-dependent piRNA production. Scatter plots showing Rhi binding enrichment (*y-axis*) relative to reduction in piRNA production in $rhi^{2/KG}$ mutants (*x-axis*). Each point represents a distinct piRNA cluster. Pre-I.S. is Pre-Immune Serum. The major germline piRNA cluster at 42AB and the major somatic cluster (*flam*) are indicated in green.

possibility that piRNAs direct Rhi to chromatin, perhaps through a process analogous to siRNA guided centromeric heterochromatin assembly in *S. pombe* (Iida et al., 2008). We therefore performed Rhi ChIP-Seq on ovaries isolated from females mutant for *armi* ($armi^{1}/armi^{72.1}$), which is required for production of piRNAs from most dual-strand piRNA clusters (Malone et al., 2009). Rhi binding in *armi* mutant ovaries was significantly correlated with Rhi binding in the matched w^{1} background strain ovaries ($r = 0.84$, $p < 2.2 \times 10^{-16}$, Figure 4.2 and 4.S1, shown for the 42AB cluster in Figure 4.2A). Rhi localizes to distinct foci in germline nurse cell nuclei (Klattenhoff et al., 2009), and our ChIP-Seq data suggest that the majority of these foci represent dual-strand clusters. We previously showed that Rhi localizes to germline nuclear foci in *armi* and *aub* mutants (Klattenhoff et al., 2009). We extended these localization studies to *ago3*, *piwi*, *qin*, *krimp*, *zuc*, *squ* and *flam* mutants, which encode two PIWI proteins, two Tudor-domain proteins, two nucleases that may generate piRNA precurosors, and the major uni-strand cluster in somatic follicle cells (Cox et al., 1998; Sarot et al., 2004; Lim and Kai, 2007; Brennecke et al., 2007; Malone et al., 2009; Li et al., 2009; Zhang et al., 2011; Ipsaro et al., 2012). Rhi localized to distinct foci in germline nuclei in each of these mutant backgrounds (Figure 4.S2). These findings, with the ChIP-Seq data in *armi* mutants, suggest that Rhi localization to major germline clusters is independent of piRNA production.

Rhi, Cuff and UAP56 suppress cluster transcript splicing

Rhi co-localizes to nuclear foci with the piRNA pathway proteins Cuff and UAP 56, mutations in *rhi* disrupt localization of both proteins to nuclear foci, and mutations in *cuff* and *uap56* disrupt Rhi localization (Pane et al., 2011; Zhang

Figure 4.2

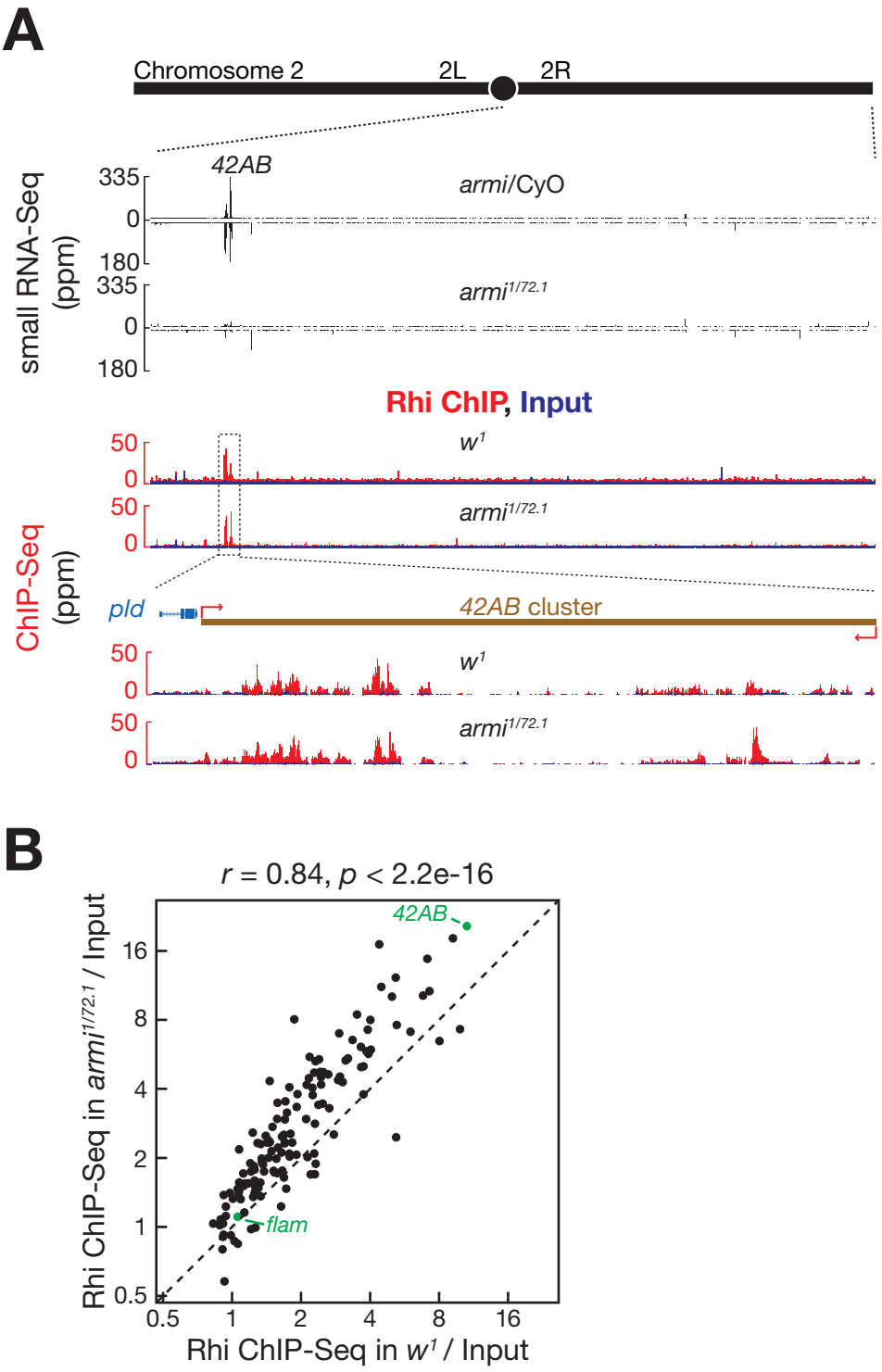
A
Chromosome 2
2L
2R
42AB
armi/CyO
armi1/72.1
small RNA-Seq (ppm)
335
0
180
Rhi ChIP, Input
w1
armi1/72.1
ChIP-Seq (ppm)
50
0
pld
42AB cluster
B
r = 0.84, p < 2.2e-16
42AB
flam
Rhi ChIP-Seq in armi1/72.1 / Input
Rhi ChIP-Seq in w1 / Input
0.5
1
2
4
8
16

Figure 4.2. Rhi binding does not depend on the piRNA production.

(A) Small RNA-Seq and Rhi ChIP-Seq signal across the right arm of chromosome 2 in control (*armi*/*CyO* and w^1) and *armi* mutants. The 42AB cluster is indicated. Mutations in *armi* nearly eliminate piRNAs mapping to 42AB (Malone et al., 2009), but do not reduce Rhi binding.

(B) Scatter plot showing Rhi ChIP-Seq enrichment in $armi^{1/72\,1}$ mutants relative to w^1 controls. Each dot represents a distinct piRNA cluster.

et al., 2012a). Cuff is related to yeast Rai1, which has been implicated in mRNA de-capping and degradation, and UAP56 is a nuclear DEAD box protein previously shown to function in splicing and mRNA export. UAP56 associates with cluster transcripts, and both UAP56 and Cuff are required for germline piRNA production (Pane et al., 2011; Zhang et al., 2012a). These observations suggest that Rhi, Cuff and UAP56 function at a related nuclear step in the piRNA biogenesis pathway.

Our previous qPCR studies showed that *rhi* mutations reduce the steady state level of precursor transcripts at a limited number of sites in two piRNA clusters, suggesting a role for Rhi in cluster transcription (Klattenhoff et al., 2009). We therefore used strand-specific paired-end RNA sequencing (RNA-Seq) to characterize cluster and gene transcription in *rhi, cuff* and *uap56* mutant ovaries, and in appropriate background control strains. By sequencing 100 nts from both ends of cloned RNA fragments, we were also able to use relatively rare polymorphisms in repeated sequences to map reads over a significant fraction of most clusters (Figure 4.3A). In apparent conflict with our earlier qPCR data (Klattenhoff et al., 2009), we did not see a consistent reduction in reads mapping to clusters (Figure 4.3S). However, visual inspection of RNA-Seq signal associated with several major germline clusters, included 42AB cluster, showed that the reads often mapped to a small number of well-defined peaks (Figure 4.3A), and the regions previously assayed by qPCR fall between these peaks.

RNA-Seq reads that cross mature splice junctions map to two genomic locations separated by the intron length. In genome browser views, these split reads produce signal profiles that are interrupted by very sharply defined gaps.

Close inspection of the cluster peaks in *rhi, cuff* and *uap56* mutants showed that they are often interrupted by sharply defined gaps that correspond to consensus splice donor and acceptor sites. Figure 4.3B shows an example in the 42AB cluster. qPCR studies confirmed that RNAs crossing this unique donor-acceptor site junction increases over 100 fold in both *cuff* and *rhi* mutants (Figure 4.3C). The spliced peak in 42AB is in the sense orientation of a *gypsy12* element that could be activated in the mutant strains, and increased splicing could be linked to activation. We therefore used qPCR to assay splicing at a putative intron in a chromosome 4 cluster composed of telomeric transposons, which is anti-sense to the transposon fragment. These studies confirmed an increase in splicing in *rhi, cuff* and *uap* mutants, which cannot be explained by sense strand transcription of the active element (not shown). These initial studies raised the intriguing possibility that Rhi functions with Cuff and Uap56 to actively suppress cluster transcript splicing.

To extend these observations across the transcriptome, we identified all split reads mapping to consensus splice donor and acceptor sites in Oregon R, w^1 and *cn,bw* control strains, and in *rhi, cuff* and *uap56* mutants. We then identified introns that were shared between mutant and control strains and determined their splicing efficiency, which we defined as split reads (defining spliced RNAs) divided by the sum of split reads and reads crossing the corresponding splice sites (defining unspliced RNAs). The scatter plots in Figure 4.4 show that *rhi, cuff* and *uap56* mutants do not lead to global changes in splicing efficiency at introns shared with control strains, including the rare shared introns mapping to piRNA clusters (Figure 4.4 A, 4.4B, 4.4C, 4.4D, red points).

Figure 4.3

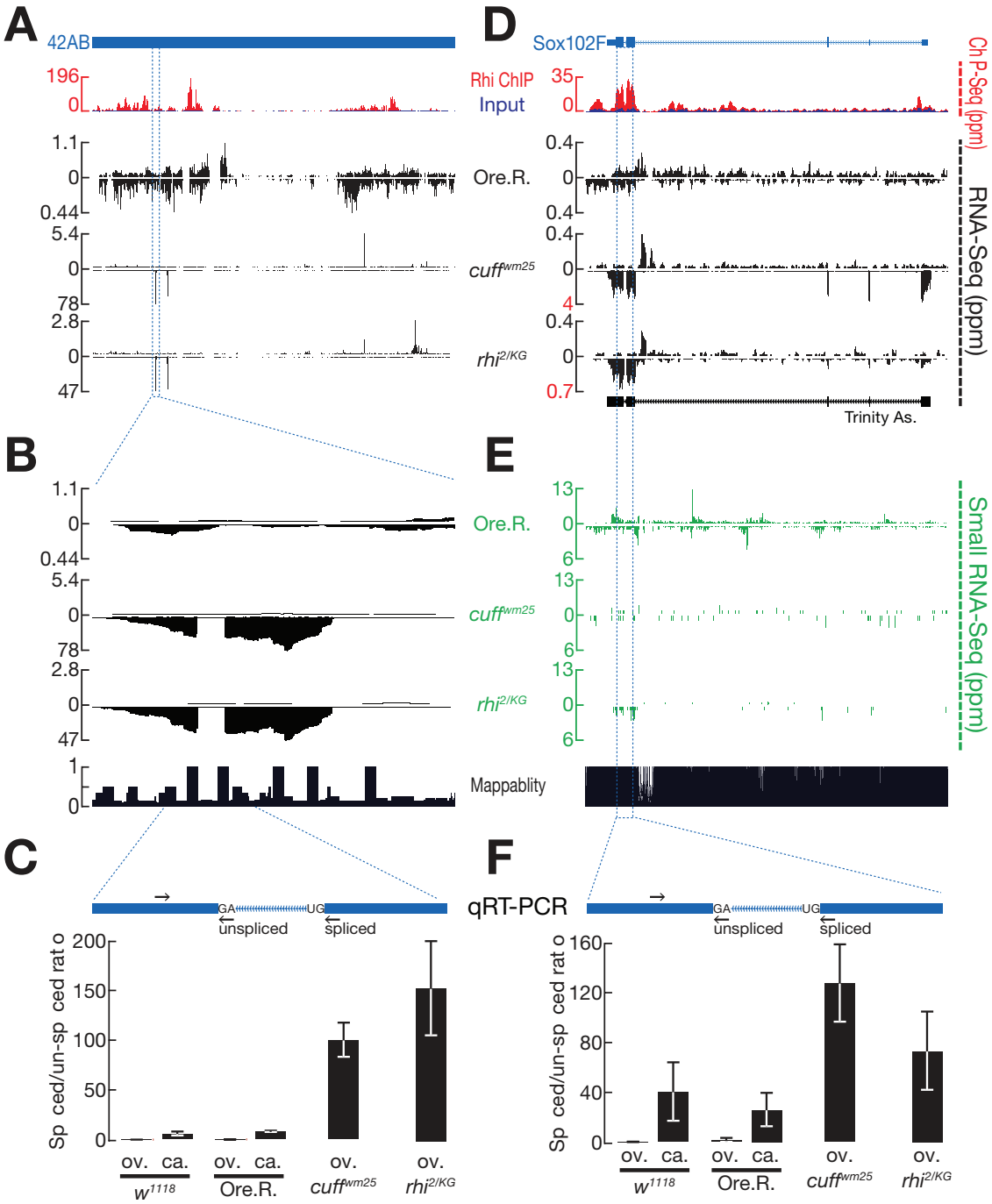
A
42AB
196
0
1.1
0
0.44
5.4
0
78
2.8
0
47
Rhi ChIP
Input
Ore.R.
cuff^wm25
rhi^2/KG
B
1.1
0
0.44
5.4
0
78
2.8
0
47
1
0
Ore.R.
cuff^wm25
rhi^2/KG
Mappablity
C
GA
UG
unspliced
spliced
Sp ced/un-sp ced rat o
200
150
100
50
0
ov. ca. ov. ca. ov. ov.
w^1118
Ore.R.
cuff^wm25
rhi^2/KG
D
Sox102F
35
0
0.4
0
0.4
0.4
0
4
0.4
0
0.7
Trinity As.
Ch P-Seq (ppm)
RNA-Seq (ppm)
E
13
0
6
13
0
6
13
0
6
Small RNA-Seq (ppm)
F
qRT-PCR
GA
UG
unspliced
spliced
Sp ced/un-sp ced rat o
160
120
80
40
0
ov. ca. ov. ca. ov. ov.
w^1118
Ore.R.
cuff^wm25
rhi^2/KG

Figure 4.3. Rhi and Cuff suppress splicing at the 42AB and *sox102F* clusters.
(A, D). Rhi ChIP-Seq and RNA-Seq profiles for the 42AB cluster (A) and *sox102F* locus (D) are shown. For ChIP-Seq, Rhi signal (red) is superimposed on input (blue). RNA-Seq is shown for Oregon R (Ore. R.) control, *cuff* and *rhi* mutants. In Ore. R. controls, signal is spread over both 42AB and *sox102F*. In *rhi* and *cuff* mutants, by contrast, signal shifts to distinct peaks (Note that data are scaled to avoid peak clipping). At the *sox102F* locus, the boundaries correspond to annotated splice sites in the mature somatic transcript, and *de novo* transcript assembly from these data yields the annotated gene structure (Trinity As.)
(B) A high resolution expansion of the indicated region of 42AB shows that the peak is interrupted by a region with little signal, defined by very sharp boundaries characteristic of intron removal.
(C, F) qRT-PCR quantify the splicing efficiency at 42AB (C) and *sox102F* (F) loci. The diagrams show the putative introns (blue) and the position of primers (arrow) used to assay unspliced and spliced transcripts at 42AB (C) and *sox102F* (F). Both spliced and unspliced transcripts are amplified using the same forward primers. Reverse Primers for unspliced transcripts span the splice site and for spliced transcripts span the mature junction. Bar graphs show the ratio of spliced to unspliced RNAs in ovary (ov.) and carcass (ca.) in two different control strains (w^{1118} and Ore. R.) and in *cuff* and *rhi* mutants. In ovaries, splicing at 42AB and *sox102F* increases over 80 fold in both *cuff* and *rhi* mutants. The *sox102F* locus is expressed in somatic tissue present in the carcass, and the transcripts are spliced.
(E) piRNA production from *sox102F* locus in Ore.R., *cuff* and *rhi* mutants.

Many of the prominent cluster-mapping introns in mutant ovaries do not appear to be utilized in control stains. Suppressed splicing of these cluster specific introns could be critical to piRNA biogenesis, and they would not be included in the analysis of splicing efficiency at shared introns. We therefore identified introns in each mutant that were not used in the *cn,bw, w*^{1} or Oregon R control strains. The three control strains share 32,273 introns, which overwhelmingly map to annotated genes (Figure 4.4E, open bar). Each of the control strains utilizes 140 to 200 genic introns that are not used in the other control strains. By contrast, *rhi, cuff* and *uap56* each utilizes approximately 600 genic introns that are not used in any of the control strains. Only 81 of these introns are used in all three mutants, strongly suggesting that most of the strain specific introns are not directly controlled by the piRNA machinery. However, Rhi, Cuff and UAP56 could directly or indirectly regulate splicing of the 81 shared introns.

The control strains share 32,273 introns mapping to 6500 annotated genes. By contrast, these strains share only 5 introns that uniquely map to 142 piRNA clusters, and only 1 to 3 introns are specific to each of the control strains. By contrast, *rhi, cuff* and *uap56* mutants utilize 27 to 81 cluster specific introns that are not spliced in any of the control strains, and 11 of these introns are utilized in all three mutants (Figure 4.4F). In addition, a significant fraction of these introns fall within the top 20 piRNA producing clusters (Figure 4.4G). Rhi, Cuff and UAP56 thus appear to suppress splicing of a subset of novel cluster mapping introns. Most clusters are composed of repeats and produce a significant fraction of transcript specific RNAseq reads that cannot be uniquely mapped to the

genome. Our analysis of cluster processing is therefore likely to underestimate the number of introns that are utilized in the mutant ovaries. Splice sites imbedded in repeats are also very difficult to verify by qPCR. To more rigorously analyze the role of Rhi and associated protein in cluster transcript processes, we therefore turned to a unique sequence cluster with experimentally verified splice sites, and to transgenic reporters.

Rhino, Cuff and UAP56 convert a somatic protein-coding gene to a germline piRNA cluster

The *sox102F* locus on chromosome 4 is largely composed of unique sequences. In somatic cells, the locus produces an experimentally verified mRNA from a spliced primary transcript, and the mature transcript encodes a putative transcription factor (Figure 4.3D). In the germline, by contrast, this locus produces piRNAs from both genomic strands, and mutations in *rhi, cuff* and *uap56* disrupt production of these uniquely mapping piRNAs. Our ChIP-Seq studies also show that Rhi binds to this region (Figure 4.3D and 4.S4), and RNA-Seq reveals RNAs from both strands, with roughly equal RNA reads in introns and exons, and very few split reads characteristic of splicing (Figure 4.3D and 4.S4). The *sox102F* locus is therefore a protein coding gene in the soma, and a dual-strand piRNA cluster in the ovary. We note that piRNA abundance and long RNA-Seq reads in the ovary data do not obviously correlate with to the intron/exon structure of the somatic *sox102F* transcript. However, Rhi ChIP-Seq peaks correspond to somatic exons. These observations raise the intriguing possibility that Rhi recruitment to chromatin is linked to splicing signals in the nascent transcript.

Figure 4.4

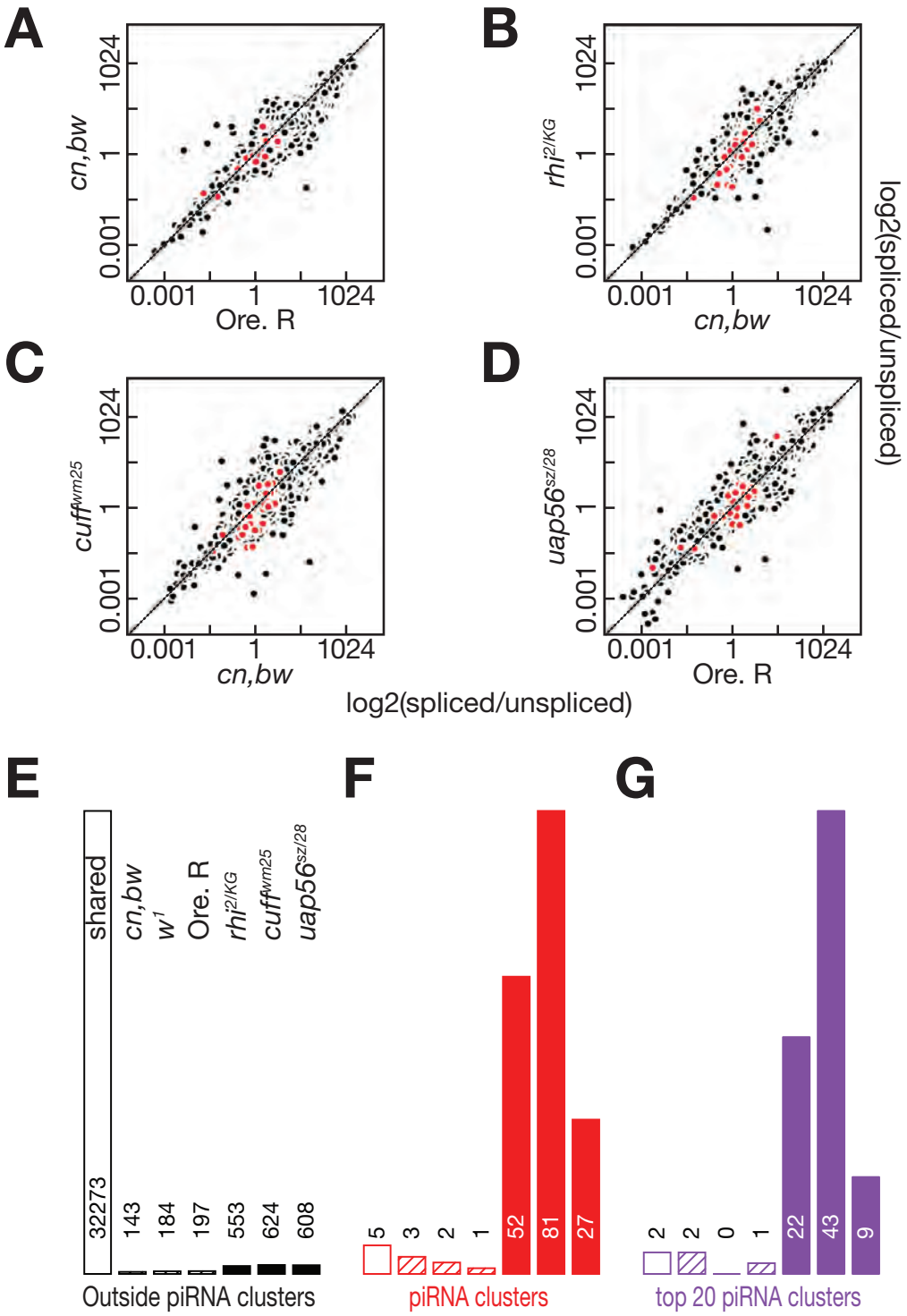
A
B
C
D
cn,bw
Ore. R
rhi2/KG
cuff wm25
uap56sz/28
0.001
1
1024
log2(spliced/unspliced)
E
F
G
shared
32273
143
184
197
553
624
608
Outside piRNA clusters
5
3
2
1
52
81
27
piRNA clusters
2
2
0
1
22
43
9
top 20 piRNA clusters

Figure 4.4. *rhi, cuff* and *uap56* mutations do not alter global splicing efficiency, but lead to splicing of novel cluster introns.

(A-D). Scatter plots showing splicing efficiency at introns shared between the two indicated genotypes. Each point is one intron. Cluster mapping introns are in red and introns mapping outside clusters in black.

(E-G). Bar graphs quantifying shared and genotype specific introns. Introns outside of piRNA clusters are in black, introns mapping to all clusters are in red, and introns mapping to the top 20 clusters are in purple. Introns shared by the *cn,bw, w*^{1} and Oregon R control strains are indicated by the open bars. Mutant specific introns are in solid bars. For each set of bar graphs, the genotypes are ordered as in E, and the number of introns detected is above or within the bar.

Strikingly, our RNA-Seq data show that *rhi, cuff* and *uap56* mutations, which block piRNA production at *sox102*, also lead to a significant increase in splicing at the annotated somatic donor and acceptor sites (Figure 4.3D and Figure 4.S4). The increase in split reads is particularly pronounced in *cuff* mutants. In this background, all of the annotated introns are efficiently excised, and *de novo* transcript assembly generates an mRNA that precisely matches the somatic *sox102F* transcript (Figure 4.3D). qPCR confirmed that *sox102F* splicing efficiency increases by 80 and 200 fold increases in *rhi* and *cuff* mutants (Figure 4.3F). *uap56* mutants produce a similar, but more modest, increase in splicing (Figure 4.S4). The *qin* locus encodes a component of the cytoplasmic nuage that is required for piRNA amplification and transposon silencing (Zhang et al., 2011). In contrast to mutations in *rhi, cuff* and *uap56, qin* mutations do not increase splicing efficiency at *sox102F* (Figure 4.S4). Therefore, increased splicing is not a secondary consequence of transposon activation and genome destabilization. Significantly, the *sox102F* locus produces both spliced and un-spliced transcripts in the w^{1118} background strain, but no piRNAs map to the mature splice junction shown in Figure 4.3F, while 21 reads map to the corresponding 3´ splice site. Rhi, Cuff and UAP56 thus suppress splicing at *sox102F,* and the resulting unspliced transcripts appear to be preferentially processed into piRNAs.

Rhi tethering suppresses splicing and directs complementary transcripts to the piRNA-processing machinery

To determine if Rhi binding is sufficient to suppress splicing and induce piRNA production, we used a transgenic LacO-LacI DNA-protein binding system to "tether" Rhino to an ectopic locus. For these experiments, we generated

Figure 4.5

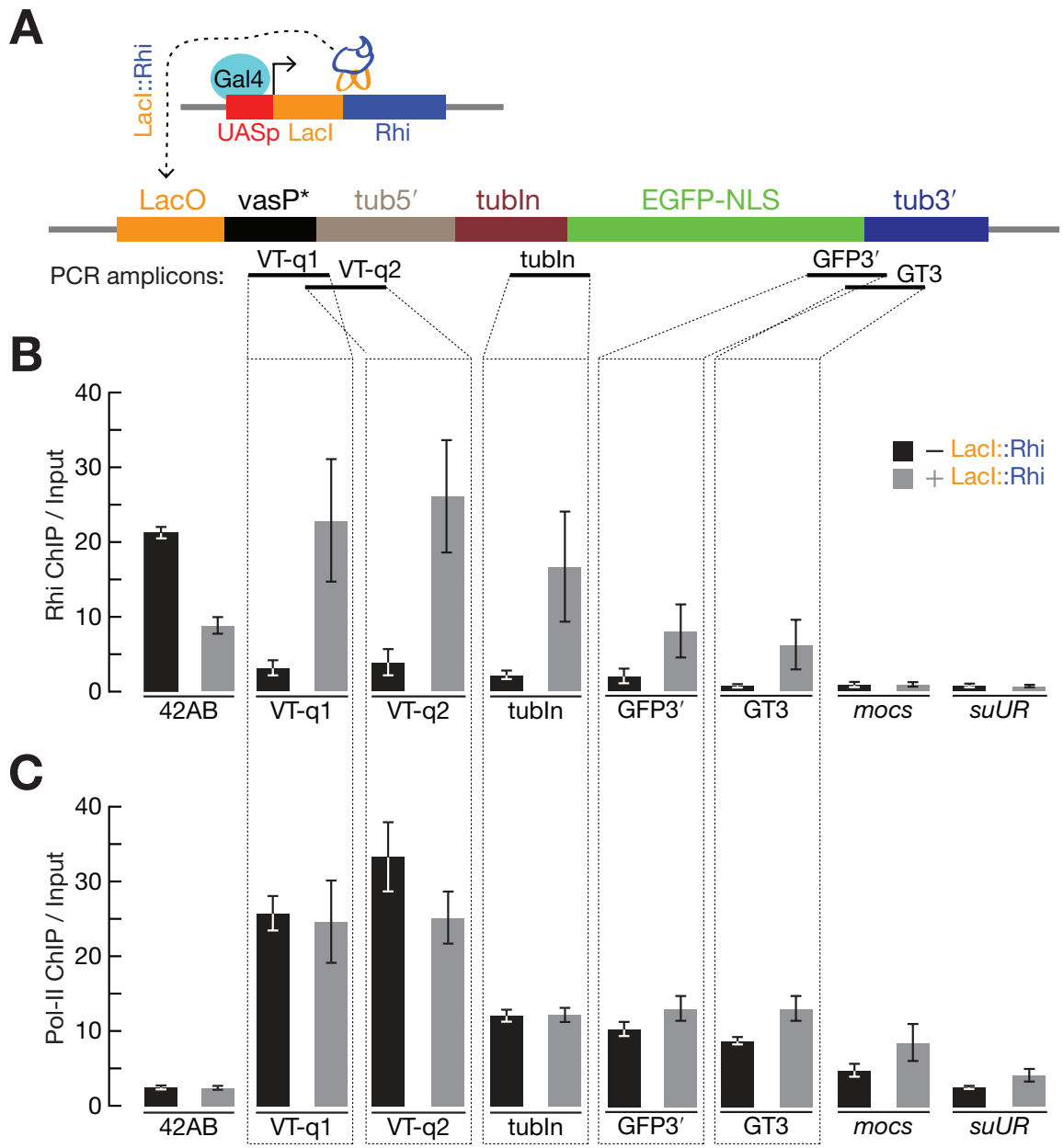
A
LacI::Rhi
Gal4
UASp LacI
Rhi
LacO
vasP*
tub5′
tubIn
EGFP-NLS
tub3′
PCR amplicons:
VT-q1
VT-q2
tubIn
GFP3′
GT3
B
Rhi ChIP / Input
40
30
20
10
0
– LacI::Rhi
+ LacI::Rhi
42AB
VT-q1
VT-q2
tubIn
GFP3′
GT3
mocs
suUR
C
Pol-II ChIP / Input
40
30
20
10
0
42AB
VT-q1
VT-q2
tubIn
GFP3′
GT3
mocs
suUR

Figure 4.5. Rhi "tethering" leads to spreading through the target transcription unit, but does not reduce Pol-II occupancy.

(A) Schematic diagram of the experimental design. An inducible UASp promoter was used to drive expression of the DNA binding protein LacI fused to full length Rhi, in the presence of a reporter gene containing LacI binding sites (LacO) upstream of a promoter (truncated *vasa* promoter) that drives expression of the 84B *alpha tubulin* 5´UTR and first intron fused to EGFP with nuclear localization signal (NLS). PCR amplicons indicates positions assayed for Rhi and Pol-II binding in panel B and C.

(B) Fold enrichment by Rhi ChIP across the reporter in the absence (grey) or presence of LacI::Rhi. The LacI::Rhi lead to Rhi binding through the transcription unit.

(C) Fold enrichment by RNA pol-II ChIP across the reporter, bars as indicated for panel B. RNA polymerase binding across the transcription unit is not altered in the presence of the LacI::Rhi. The 42AB locus is used as a positive control for Rhi binding. The *mocs* and *suUR* loci are located downstream of reporter construct in the genome. We do not detect Rhi spreading or changes in RNA Pol-II at these loci.

transgenic flies harboring a reporter transgene containing 36 LacO DNA-binding sites upstream a truncated *vasa* promoter, which drives EGFP expression in the germline and in somatic follicle cells of the ovary. The transcription unit contains the 84B *α-Tubulin* 5´ UTR and first intron, followed by an exon encoding EGFP (Figure 4.5A). We then introduced transgenes with inducible Gal4 promoters controlling expression of either a LacI control or a LacI::Rhi fusion protein, and the *nanos*-Gal4-VP16 driver to induce germline expression. Target EGFP expression was assayed by laser scanning confocal microscopy and western blotting (Figure 4.6A and 4.S5). In the presence of the LacI control, strong EGFP signal was observed in germline nurse cells and somatic follicle cells (Figure 4.6A). Immunolabeling confirmed that LacI was expressed only in the germline. Germline expression of the LacI::Rhi fusion, by contrast, silenced EGFP expression in the germline, but not in the follicle cells (Figure 4.6A). Rhi mediated EGFP silencing was confirmed by western blotting (Figure 4.S5).

HP1a recruits the methyl transferase that modifies Histone H3 to generate HP1a binding sites, which induces heterochromatin spreading (Danzer and Wallrath, 2004). Because Rhi is an HP1a homolog, we speculated that Rhi tethering may lead to Rhi spreading from the LacO sites. We therefore assayed Rhi binding at sites through the transgene reporter using ChIP and qPCR. In the absence of LacI::Rhi fusion protein induction, we observed background binding of Rhi through the transgene. By contrast, expression of the LacI::Rhi fusion was linked to significant Rhi binding through the transcription unit, with the highest levels near the LacO binding sites (Figure 4.5B). Adaptation by the piRNA pathway appears to involve insertion of invading elements into clusters. We

speculate that Rhi spreading into these new sequences, from the surrounding cluster, may be required to produce new silencing piRNAs during adaptation.

To determine if LacI::Rhi binding and Rhi spreading silences transcription, we used ChIP-qPCR to measure RNA Polymerase II (Pol II) binding at the target transgene (Figure 4.5C). Essentially identical levels of Pol II binding were observed in the absence or presence the LacI::Rhi fusion. Rhi tethering thus appears to block protein expression through a post-transcriptional mechanism. The increase in cluster splicing in *rhi* mutants led us to speculate that Rhi binding may suppress splicing. We therefore used qPCR to quantify splicing efficiency at the Tubulin intron in the target transgene. We observed a 6 fold reduction in splicing efficiency in the presence of the LacI::Rhi fusion (Figure 4.6B and 4.6C). Our analyses of piRNA cluster transcripts in *rhi* mutants and these transgenic studies, taken together, indicate that Rhi binding suppresses splicing.

To determine if LacI::Rhi tethering is sufficient to induce piRNA production, we sequenced small RNAs from ovaries carrying the target transgene and expressing either the LacI control or the LacI::Rhi fusion. With both combinations, however, we detected only very low levels of 23 to 30nt putative piRNAs mapping to the transgene (Figure 4.7 and 4.S6). Rhi is specifically required for piRNA production from dual-strand clusters, and the target transgene is transcribed from only one strand. We therefore contructed a second transgene with a promoter driving antisense expression of the target sequences, integrated this gene into the same chromosomal locus as the sense strand reporter, and generated females carrying a trans combination of sense and antisense reporters. Small RNA sequencing showed that expression of the

Figure 4.6

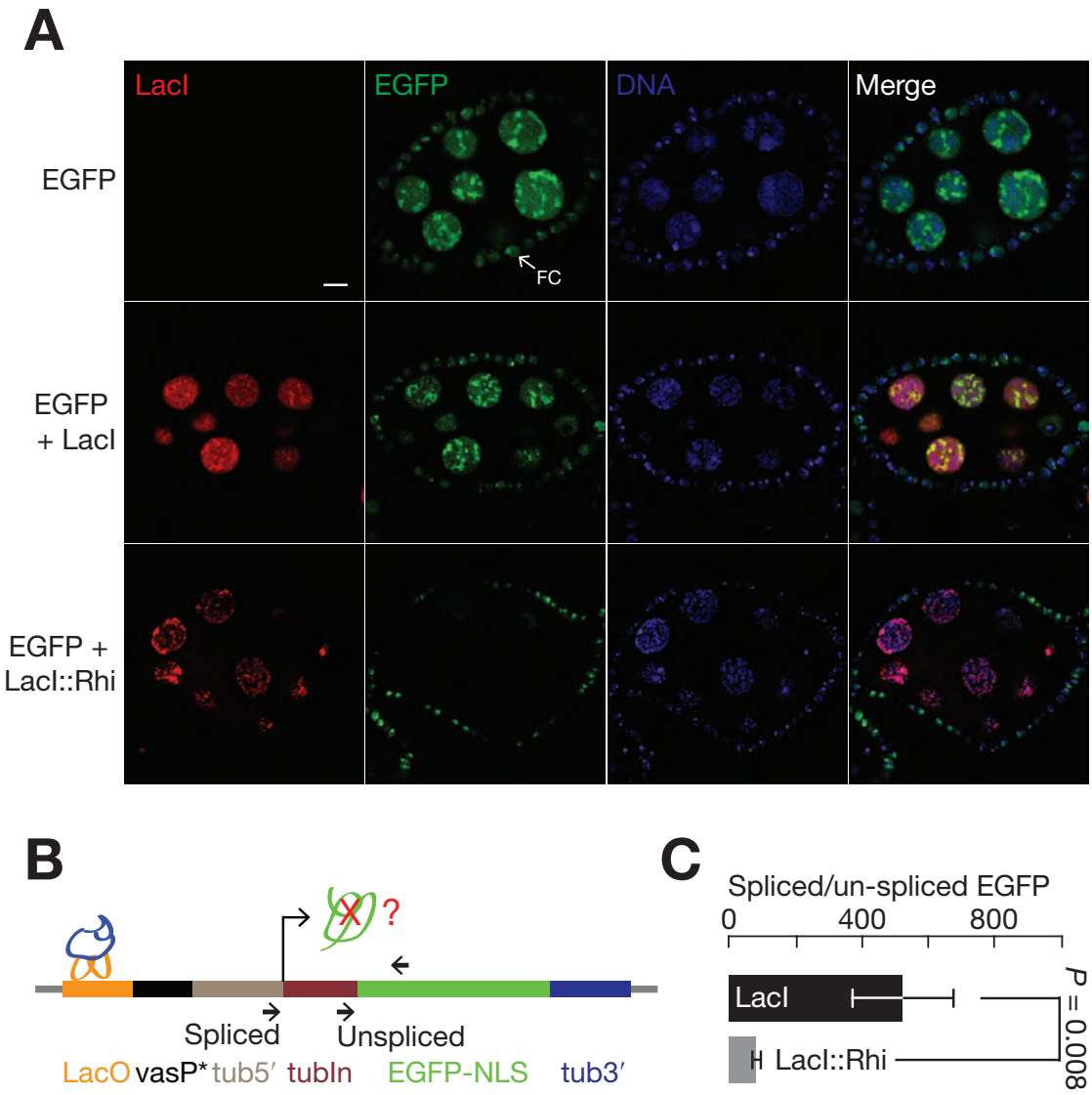
A
LacI
EGFP
DNA
Merge
EGFP
FC
EGFP
+ LacI
EGFP +
LacI::Rhi
B
?
Spliced
Unspliced
LacO
vasP*
tub5′
tubIn
EGFP-NLS
tub3′
C
Spliced/un-spliced EGFP
0
400
800
LacI
LacI::Rhi
P = 0.008

Figure 4.6. Tethering Rhi suppresses EGFP expression and splicing.

(A) Germline expression of LacI (red) does not suppress EGFP expression (green). By contrast, expression of LacI::Rhi (red) suppresses EGFP accumulation in the nurse cell nuclei (green). Note that the EGFP reporter is expressed in both the germline and surrounding somatic follicle cells (arrows, FC). The fusion does not expressed in the follicle cells, and EGFP expression in these cells is not reduced. The bar in the up right panel is 10 μm, and applies to all panels.

(B) Splicing at the target locus. The diagram shows the target transgene and indicates that position of LacI or LacI::Rhi binding (LacO) and the primers used to assay both spliced and unspliced transcripts by qRT-PCR.

(C) Bar graph showing the ratio of spliced to unspliced target in the presence of LacI (black) or LacI::Rhi (grey). LacI::Rhi binding lead to a significant reduction in splicing efficiency ($p = 0.008$).

LacI::Rhi fusion, but not LacI, triggered production of 23 to 30nt small RNAs from both strands of the reporter construct, including the intron, EGFP and LacO binding sites (Figure 4.S6). piRNAs and endo-siRNAs bear a 2´-O-methyl group at their 3' termini, which protects them from oxidation (Horwich et al., 2007; Saito et al., 2007). miRNAs and non-specific RNA degradation products, by contrast, do not carry this modification and oxidation renders them unclonable. piRNAs and endo-siRNAs are therefore selectively retained in sequencing libraries prepared after sample oxidation. The siRNAs and piRNAs in the oxidation resistant pool can be distinguished by length, as the endo-siRNAs are 21nt and the piRNAs are between 23 and 30 nt and show a characteristic normal distribution. As shown in Figure 4.7, LacI::Rhi tethering to the trans-combination of reporter genes triggered production of oxidation resistant small RNA with a size distribution typical of endogenous piRNAs. In the P-M hybrid dysgenesis system, *de novo* piRNAs complementary to the *P-element* transposon increases with adult female age (Khurana et al., 2011). Intriguingly, reporter specific piRNAs increased by two folds when flies carrying LacI::Rhi were aged from 2-4 days to 12-14 days (Figure 4.7). Total endogenous ovary piRNAs, by contrast, did not change (Table 4.S1b). The primary piRNAs are produced *de novo*—independent of pre-existing piRNAs and ping-pong amplification. The reporter specific piRNAs from opposite strands show a very weak bias toward a 10nt overlap, indicating that they are likely to be produced by a ping-pong independent mechanism. Based on these findings we conclude that LacI::Rhi binding and expression of complementary RNAs are sufficient to drive *de novo* production of primary piRNAs from a transgenic locus.

Figure 4.7

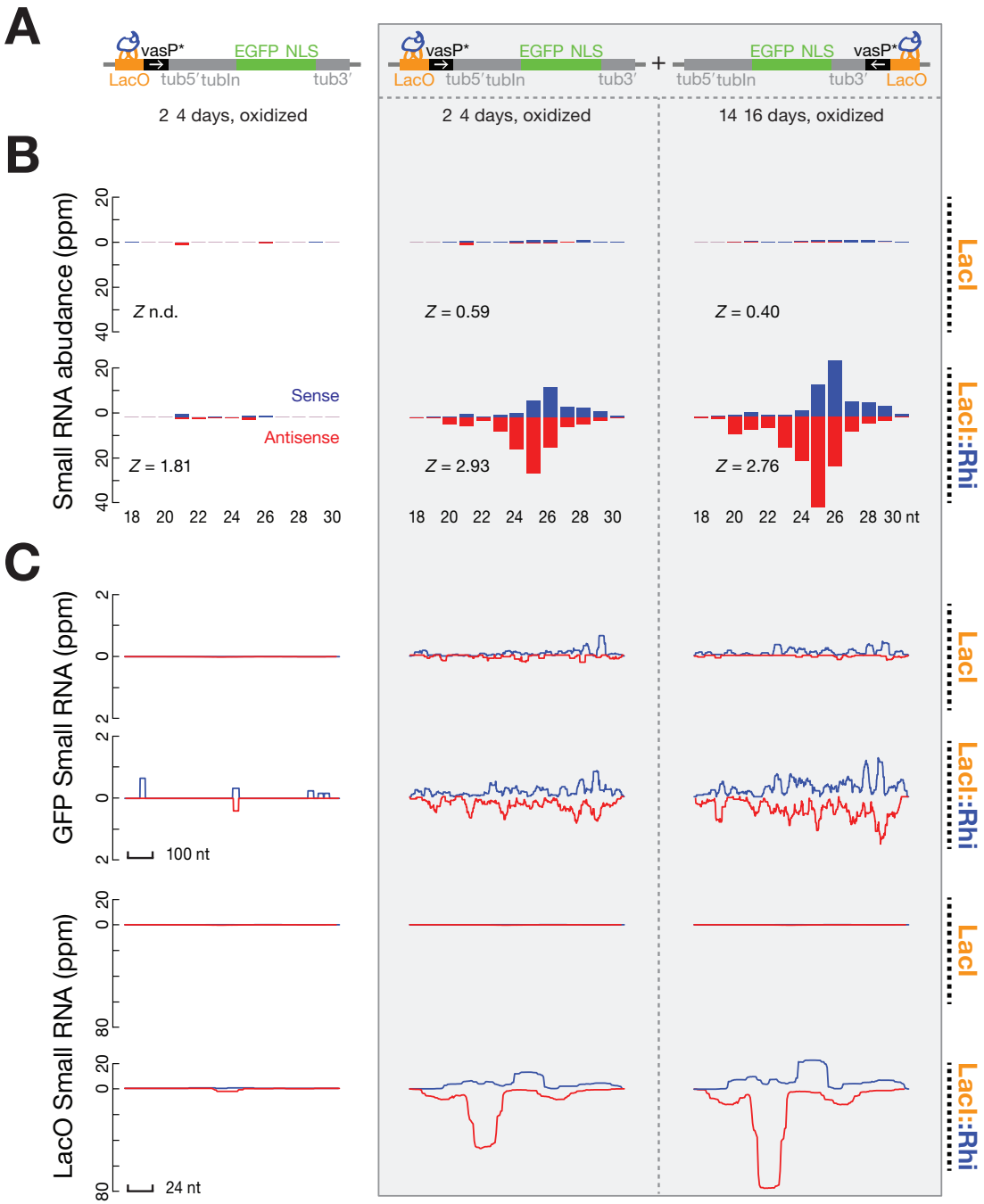
A
vasP*
EGFP NLS
LacO tub5′tubIn tub3′
2 4 days, oxidized
14 16 days, oxidized
B
Small RNA abudance (ppm)
Z n.d.
Z = 0.59
Z = 0.40
Sense
Antisense
Z = 1.81
Z = 2.93
Z = 2.76
18 20 22 24 26 28 30 nt
LacI
LacI::Rhi
C
GFP Small RNA (ppm)
100 nt
LacO Small RNA (ppm)
24 nt

Figure 4.7. Rhi binding to complementary transcription units triggers piRNA production.

(A) Diagrams show sense strand reporter and combination of sense and anti-sense reporters, indicating positions of LacI or LacI::Rhi binding and position and orientation of promoters (vasP).

(B) Length distribution of the small RNAs mapping to the reporter constructs. Blue indicates sense strand species and red indicates antisense species. *Z* scores indicate the significance of the 10 nt overlap between sense and antisense piRNAs (Ping-Pong signature). *Z*-score = 1.96 corresponds to *p*-value = 0.05. Too few piRNAs were detected with the LacI control for the Z score to be determined (indicated as n.d.). piRNAs carry a 3' end modified and therefore are resistant to oxidation. Both un-oxidized (Figure S6) and oxidized RNAs (shown) were used to prepare libraries for sequencing. LacI::Rhi binding to the sense strand reporter did not lead to significant production of oxidation resistant species between 23 and 30 nt. By contrast, LacI::Rhi binding to the combination of sense and anti-sense reporters trigger production of oxidation resistant species showing length distributions characteristic of mature piRNAs. piRNAs from opposite strands showed a weak bias toward a 10 nt overlap, which is typical of primary piRNAs produced by a ping-pong independent mechanism.

(C) Distribution of Small RNA reads over EGFP and the LacO repeats in the presence of LacI or LacI::Rhi. Sense signal is in blue and anti-sense signal is in red.

DISCUSSION

Primary piRNA production in dual-strand vs. uni-strand piRNA clusters

The piRNA pathway has an evolutionarily conserved role in transposon control during germline development, and is therefore essential for transmission of the inherited genetic complement. In the *Drosophila* ovary, the majority of piRNAs map to transposons and cannot be assigned to specific chromosomal locations, but unique piRNAs are concentrated in "piRNA clusters" composed of complex arrays of transposon fragments that are generally localized to pericentromeric or subtelomeric heterochromatin (Brennecke et al., 2007). These loci fall into two classes, based on strand bias. Clusters that produce piRNAs from both genomic strands (dual-strand clusters) are dominant in the germline, while clusters that are expressed on only one genomic strand (uni-strand clusters) produce most of the piRNAs in somatic follicle cells that surround the germline (Brennecke et al., 2007; Malone et al., 2009). Primary piRNAs from dual-strand clusters, bound to PIWI proteins, appear to drive a ping-pong cycle that amplifies the silencing RNA pool (Brennecke et al., 2007; Gunawardane et al., 2007). These primary piRNAs must be produced by a ping-pong independent mechanism. Similarly, the piRNA derived from uni-strand clusters can't be produced by ping-pong amplification. The simplest model would be that the primary piRNAs from uni-strand and dual-strand clusters are produced by a common mechanism, and that dual-strand clusters represent convergently transcribed uni-strand clusters. However, uni-strand cluster piRNA production is independent of *rhi, uap56* and *cuff,* which are essential for production of piRNAs that map uniquely to dual-strand clusters (Klattenhoff et al., 2009; Pane et al., 2011; Zhang et al., 2012a). In addition, here we provide data to show that Rhi-dependent piRNA production

from an ectopic locus requires a combination of transgenes expressing complementary transcripts (Figure 4.7 and 4.S6). The piRNAs produced from this combination of complementary target loci show a very weak bias toward a 10-nt overlap (Figure 4.7), indicating that they are produced through a ping-pong independent mechanism. These findings, along with previous genetic data, indicate that primary piRNA production by dual-strand clusters and uni-strand clusters proceed by distinct mechanisms, and that dual-strand primary piRNA biogenesis may require complementary precursor RNAs. The role of complementary RNAs in the germline piRNA biogenesis pathway remains to be determined.

Distinguishing piRNA cluster transcripts from mRNAs

Mutations that disrupt piRNA production increase expression of some target transposons by over 200 fold and destabilize the germline genome. By contrast, these mutations do not alter germline gene expression and mRNAs are not efficiently processed into piRNAs. This remarkable specificity is presumably essential to both germline gene expression and efficient transposon silencing, but it has been unclear how the precursors of trans-silencing piRNAs are differentiated from the mRNAs, which must escape silencing and direct protein synthesis. Previous studies indicate that splicing is suppressed at transgenes inserted into the *Drosohila* X-TAS cluster, and piRNAs mapping to introns are produced (Muerdter et al., 2012). However, the mechanisms that suppress splicing and the role of this process in piRNA biogenesis have not been determined. Here we show that the rapidly evolving HP1 homolog Rhi, with the Rai1 related protein Cuff and the DEAD box protein UAP56, suppress splicing of piRNA precursors in the germline. This is most clearly illustrated at the *sox102F*

locus, which produces an experimentally defined spliced mRNA from one genomic strand in the soma, but is the source of unspliced transcripts and piRNAs from both genomic strands in the ovary. We show that Rhi binds to *sox102F*, is required for piRNA production by this locus, and that mutations in *rhi*, *cuff* and *uap56* lead to efficient splicing and production of mature *sox102F* mRNAs in the germline. Significantly, we also show that tethering a LacI::Rhi fusion to a intron-containing reporter transgene suppresses its splicing, and that Rhi tethering to a trans combination of sense and antisense transgenes is sufficient to trigger *de novo* piRNA production (Figure 4.7 and 4.S6). We therefore propose that Rhino functions with Cuff and UAP56 to actively suppress cluster transcript splicing, and that the block to splicing directs cluster transcripts to the piRNA biogenesis machinery.

Defects in splicing have been linked to silencing RNA production in a number of systems. In *Arabidopsis,* mutations in splicing factors reduce siRNA directed DNA methylation, and splicing factors have been identified in *C. elegans* screens for RNAi components (Herr et al., 2006; Wypijewski et al., 2009; Christie et al., 2011; Warf et al., 2012; Zhang et al., 2013). However, in these systems, a specific role for splicing in silencing RNA production has not been demonstrated. By contrast, recent studies in the pathogenic yeast *Crypotoccous* provide compelling evidence for a direct link between stalled splicing and transposon silencing siRNA biogenesis (Dumesic et al., 2013). Dumesic et al. (2013) showed that siRNAs are produced from unspliced transposon transcripts and that splicing factors associate with the siRNA biogenesis machinery. Futhermore, intron removal reduces siRNA production and splice site mutations that reduce splicing efficiency increase siRNA production (Dumesic et al., 2013).

While the germline specific piRNAs in *Drosophila* and transposon silencing siRNA in *Cryptococcus* are produced by different mechanisms, these findings raise the intriguing possibility that suppressed splicing generates a conserved molecular signature that differentiates small silencing RNA precursors from mature mRNAs. Retrotransposons and retroviruses encode spliced transcripts that produce proteins needed for replication and packaging, but splicing must be suppressed to produce full length genomic RNAs. This novel feature of the retroviral life cycle could have driven the evolution of silencing systems that exploit stalled splicing as a molecular signature of potentially pathogenic RNAs.

EXPERIMENTAL PROCEDURES

General Methods.

RNA isolation, small RNA library construction and sequencing data analysis, immunoblotting, immunostaining and quantitative RT-PCR were performed as described (Zhang et al., 2011). Figures were generated using Excel (Microsoft, Redmond, WA, USA), IgorPro (WaveMetrics, Lake Oswego, OR, USA), Adobe Photoshop and Illustrator (Adobe systems, San Jose CA, USA). Table 4.S1 reports the statistics for the ChIP-Seq, RNA-Seq and small RNA-Seq data generated in this study. Table 4.S2 reports primer sequences for ChIP-qPCR and qRT-PCR. PCR primers used to clone the LacI binding domain, Rhi open reading frame and LacO repeats are detailed along with the supplemental text. The sources of the published deep sequencing data used in this study are summarized in Table 4.S4. Table 4.S5 lists the antibody information. Unless otherwise specified, *p*-values were calculated from at least three independent biological replicates using a two-tailed, two-sample unequal variance t-test (Excel, Microsoft).

Drosophila stocks.

All flies were raised at 25°C. Table 4.S3 summarizes the published fly alleles used in this study. Transgenic flies for tethering Rhi to EGFP locus were made as described in supplemental information.

Transgenic flies for tethering Rhi to the GFP locus.

Transgenes expressing LacI or LacI::Rhi fusion were made as follows: the 1.1 kb lacI binding domain from lacI-HP1 in pCas-hs-act, provided by L. L. Wallrath (Li et al., 2003), was PCR amplified (Forward primer: AAA GAA TTC GCC ATG GTG AAA CCA GTA ACT; Reverse primer: AAA GGA TCC AAC CTT CCT CTT CAT C), and the 1.4 kb *rhi* coding sequence from the full length rhino cDNA

clone RE36324 (Klattenhoff et al., 2009) was PCR amplified (Forward primer: AAA GGA TCC GTT ATG TCT CGC AAC CAT CAG; Reverse primer: AAA CGC GCC GCT TTA CTT GGG CAC AAT GAT). Rhi PCR product was digested with BamHI/NotI and cloned into pBstII KS+ to generate pBst-Rhi. The LacI PCR product was digested with EcoRI and BamHI and cloned into pBst-Rhi to generate pBst-lacI::Rhi. The entire insert in pBst-lacI::Rhi was cut out with KpnI/NotI and cloned into the transformation vector pUASp to generate pUASp-lacI::Rhi. pUASp-lacI::Rhi was cut with BamHI to remove the *rhi* cDNA and re-circularized to yield pUASp-lacI. All intermediates and final plasmid clones were verified by direct sequencing. The final pUASp-lacI and pUASp-lacI::Rhi constructs were used to make germline transgenes using standard protocols.

To generate transgenic flies carrying the LacO-EGFP, XbaI was used to partially digest pSV2-dhfr-8.32 provided by A. S. Belmont (Robinett et al., 1996). A 1.2 kb XbaI fragment corresponding to 32 repeats of the 36bp Lac operon was cloned into pBstII KS+ to produce pBst-32mer. This clone was used to provide the LacO repeats that were subsequently cloned into unique restriction sites upstream of a truncated *vasa* promoter driving expression of the 84B alpha tublin 5'UTR and first intron followed by EGFP-NLS and the tublin-3'UTR. The repeats were excised from pBst with SmaI/NotI, the ends were polished with Klenow Large Fragment and the DNA was cloned into either the 5' Not/blunted site or the 3' Bam/blunted site of the EGFP reporter construct. The following primers were used to amplify the partial *vasa* promoter from vasp-EGFP (Forward: AAA GGA TCC ATA TGA ATG AAT CAC TTA GG; Reverse: AAA GGA TCC GTG GAA TTT CCC ATT GTG C). This product was cut with BamHI and cloned into the unique BamHI site at the 3' end of the reporter construct in the anti-sense

direction to get the GFP-vaspAS-LacO construct. Due to the instability of the lacO repeats, all clones containing these repeat sequences were transformed into Max Efficiency Stbl2 chemically competent cells (Invitrogen cat#10268-019). Otherwise, DH5α electrocompetent cells (home made) were used. All reporter constructs contain the attB site and were integratred onto the attP2 site located at chromosome 3L-68A4.

ChIP-Seq and data analysis.

For Rhi ChIP-Seq, ovaries ovaries were first crosslinked with 2% formaldehyde for 10 minutes in Robb's medium (100 mM HEPES pH 7.4, 55 mM sodium acetate, 40 mM potassium acetate, 100 mM sucrose, 10 mM glucose, 1.2 mM $MgCl_2$, 1 mM $CaCl_2$). Then the reaction was quenched by adding Glycine to 120 mM and incubating for 5 minutes with rotation. The ovaries were then washed twice with TBS (50 mM Tris-HCl pH 7.5, 150 mM NaCl), and twice with ChIP lysis buffer (50 mM Hepes/KOH pH 7.5, 140 mM NaCl, 1% [v/v] Triton X-100, 0.1% [w/v] Na-Deoxycholate, 0.1% [w/v] SDS). Washed ovaries were sonicated for 4 ×15 minutes in sonication buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, 1 Proteinase Inhibitor tablet freshly added) with a Bioruptor Standard (diagenode cat# B01010001). The ovary lysate was centrifuged at 14000 rpm at 4˚C for 15 minutes, 200 μl supernatant was saved as the input control and the remaining supernatant was diluted 7 fold with dilution buffer (20 mM Tris-HCl, 167 mM NaCl, 1.2 mM EDTA, 0.01% [w/v] SDS, 1% [v/v] Triton X-100, 1 Proteinase Inhibitor tablet freshly added). For each ChIP-Seq library, 25 μl anti-Rhi antibody or Pre-Immune Serum(custom made, guinea pig 1943) was conjugated to 400 μl Dynabeads Protein A (Life technologies, cat # 10001D). The diluted supernatant added to the conjugated beads and incubated was at 4˚C overnight. The beads were then wash two time each with 1 ml Wash buffer A (20

mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% [w/v] SDS, 1%[v/v] Triton X-100, 150 mM NaCl), 1 ml Wash buffer B (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% [w/v] SDS, 1%[v/v] Triton X-100, 500 mM NaCl), 1 ml Wash buffer C (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% [v/v] NP-40, 1% [w/v] Na-deoxycolate, 0.25 M LiCl) and 1 ml Wash buffer D (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The beads were then resuspended with 200 μl sonication buffer, an the saved input sample was thawed and processed in parallel with the the ChIP sample, as follows: Crosslinking was reversed by adding 2 μl 5 M NaCl to the beads/input and incubating at 65˚C for 6 hours. Then 200 μl Tris-HCl buffer (10 mM Tris-HCl pH 8.5) was added to each sample. To remove RNA, 6 μl 30 mg/ml RNaseA was added and incubated at 37˚C for 2 hours. To digest protein, 20 μl 20 mg/ml Proteinase K was added and the mixture was incubated at 55 ˚C for 2 hours. Finally, phenol:chloroform extraction was used to purify the immunoprecipitated DNA, which was dissolved in 34 μl water. The sequencing library was constructed by sequentially performing end-repair, A-tailing, Y-shaped adapter ligation and PCR amplification as described (Zhang et al., 2012b). The libraries from Oregon R. ovaries were prepared with illumina Paired End DNA oligos and sequenced by illumina GAII. The libraries for w^{1}-rep1 and $armi^{1/72.1}$-rep1 were made in parallel with illumina Paired End DNA oligos and sequenced by illumina HiSeq. The libraries for w^{1}-rep2 and $armi^{1/72\,1}$-rep2 were made in parallel with illumina Multiplexing oligos and sequenced by HiSeq.

The sequencing reads were mapped to the *Drosophila melanogaster* genome (FlyBase r5.45/dm3) using bwa-0.6.1 (Li and Durbin, 2009). All libraries were normalized to sequencing depth, using total genome mapping reads. For each library, bigwig files were generated for UCSC browser visualization. To calculate the Rhi binding enrichment for the piRNA cluster regions, only the reads that

uniquely mapped to one genome position were used. The mean ppm value over each cluster was calculated by bigWigAverageOverBed, with a pseudo count of 0.01.

Bioinformatics analysis of splicing

Strand-specific RNA-Seq libraries were made as described (Zhang et al., 2012b). RNA-seq reads were aligned to the genome and the transcriptome (Flybase r5.50) using TopHat 2.0.8 (Trapnell et al., 2009) with the parameters "-x 1000 -g 1000 --read-mismatches 2 --read-edit-dist 2 --read-realign-edit-dist 0 --segment-length 50 --segment-mismatches 2". Only reads mapping uniquely were considered in the downstream analysis. BEDTools (Quinlan and Hall, 2010)was used to count the fragments within a transcript or piRNA cluster, and the number of reads per transcript were normalized by the sequencing depth and transcript length. We collapsed introns detected by TopHat from six libraries (three control strains: Ore. R, *cn,bw*, w^{1}; three mutants: $rhi^{2/KG}$, $cuff^{wm25}$, $uap56^{sz/28}$), then we counted the spliced reads and the unspliced reads across the donor/acceptor sites. The introns with fewer than 10 spliced reads in all six libraries were discarded in the analysis. Splicing efficacy was calculated as the ratio reads mapping to mature splice junctions multiplied by two over the sum of reads to the corresponding donor and acceptor sites, a pseudo count 10 was used.

Accession number.

Sequence data generated in this study are available via the NCBI trace archives (http://www.ncbi.nlm.nih.gov/Traces/) using accession number SRPXXXXXX.

ACKNOWLEDGMENTS

We thank members of the Theurkauf and Weng labs and the UMass RNA biology community for critical feedback and encouragement. This work was supported by a grant R01HD049116 from the National Institute of Child Health and Human Development, NIH.

Figure 4.S1

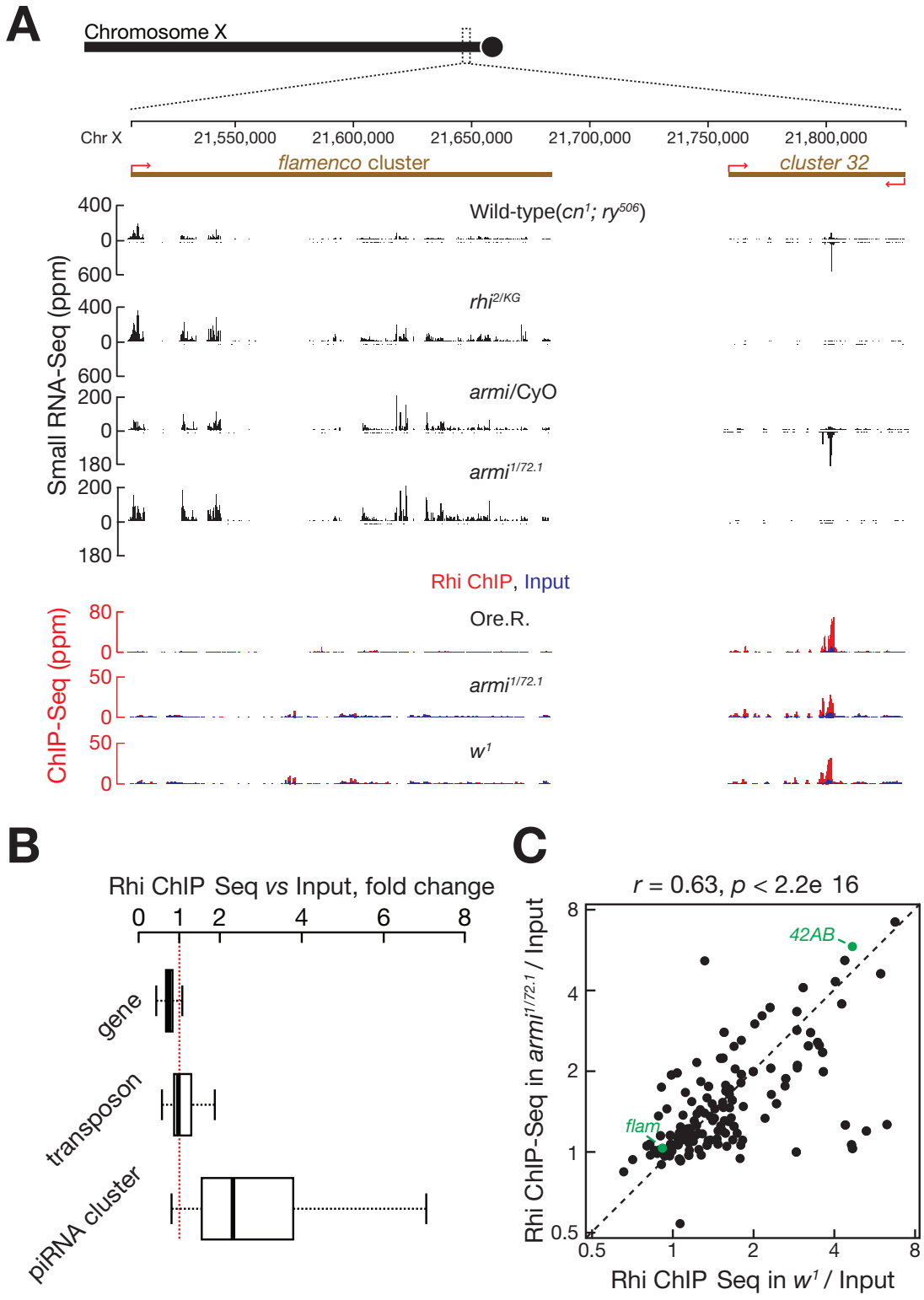
A
Chromosome X
Chr X
21,550,000
21,600,000
21,650,000
21,700,000
21,750,000
21,800,000
flamenco cluster
cluster 32
Small RNA-Seq (ppm)
Wild-type(cn1; ry506)
rhi2/KG
armi/CyO
armi1/72.1
Rhi ChIP, Input
ChIP-Seq (ppm)
Ore.R.
armi1/72.1
w1
B
Rhi ChIP Seq vs Input, fold change
gene
transposon
piRNA cluster
C
r = 0.63, p < 2.2e 16
42AB
flam
Rhi ChIP-Seq in armi1/72.1 / Input
Rhi ChIP Seq in w1 / Input

Figure 4.S1. Rhi does not bind with flamenco cluster, protein coding genes and transposons.

(A) piRNA and Rhi ChIP-Seq signal across the *flamenco* (*flam*) cluster and cluster 32, a dual-strand piRNA cluster located ~60 kb downstream of flamenco. Rhi binds to dual-strand cluster 32, but does not associate with *flam*.
(B) Boxplots showing Rhi enrichment at protein coding genes and transposons. Outliers are not shown.
(C) Scatter plot showing the correlation between Rhi ChIP-Seq in $armi^{1/72.1}$ and w^{1} fly ovaries. Each dot represents a piRNA cluster. A biological replicate is shown as Figure 2B. The *armi* mutations does not reduce Rhi binding.

Figure 4.S2

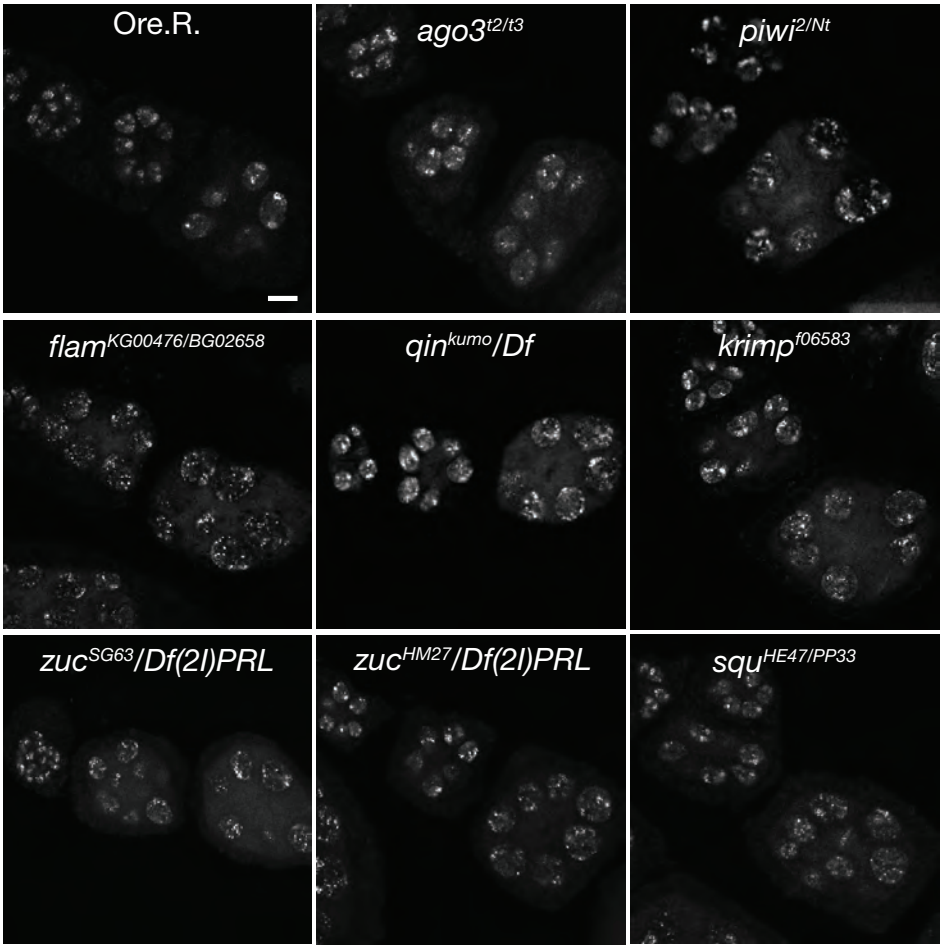
Ore.R.
ago3t2/t3
piwi2/Nt
flamKG00476/BG02658
qinkumo/Df
krimpf06583
zucSG63/Df(2l)PRL
zucHM27/Df(2l)PRL
squHE47/PP33

Figure 4.S2. Most piRNA pathway mutations do not alter Rhi localization.

The bar in the up right panel is 10 μm, and applies to all panels. Ovaries were dissected from females mutant at the indicated loci, fixed and immunolabeled for Rhi. Samples were imaged by laser scanning confocal microscopy.

Figure 4.S3

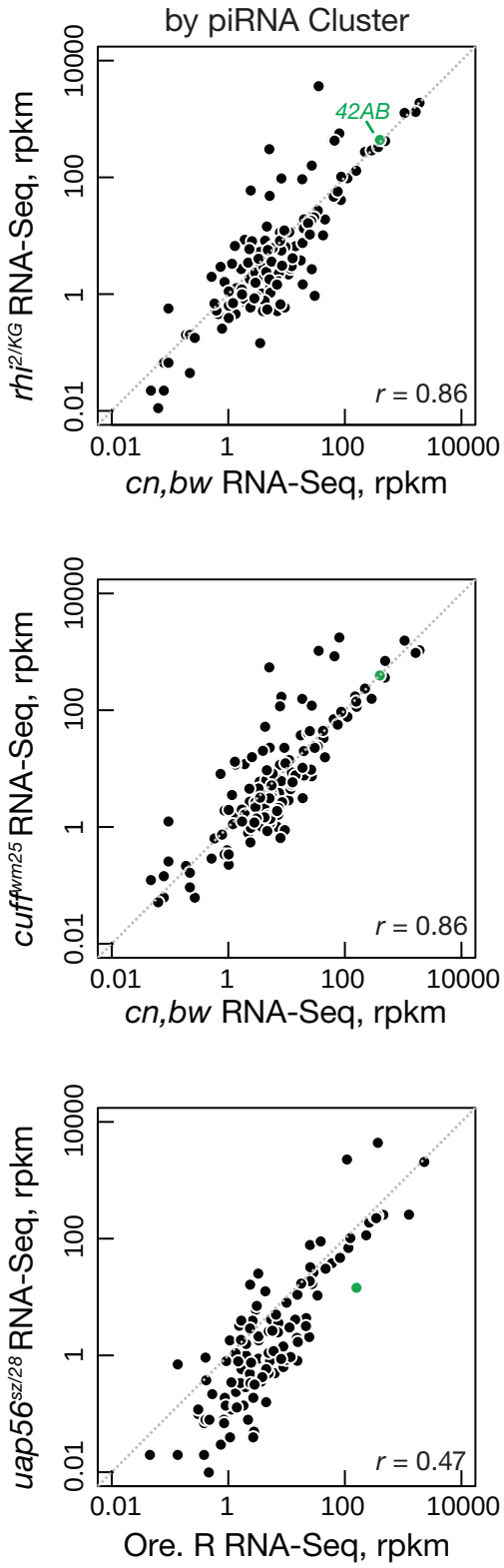
by piRNA Cluster
42AB
rhi2/KG RNA-Seq, rpkm
cn,bw RNA-Seq, rpkm
r = 0.86
cuffwm25 RNA-Seq, rpkm
cn,bw RNA-Seq, rpkm
r = 0.86
uap56sz/28 RNA-Seq, rpkm
Ore. R RNA-Seq, rpkm
r = 0.47
0.01
1
100
10000

Figure 4.S3. Loss of Rhi, Cuff or UAP56 does not affect total cluster transcript steady state level.

Scatter plots comparing normalized RNA-Seq reads uniquely mapping to clusters in *rhino*$^{2/KG}$, *cuff*^{wm25} and *uap56*$^{sz/28}$ ovaries and corresponding controls.

Figure 4.S4

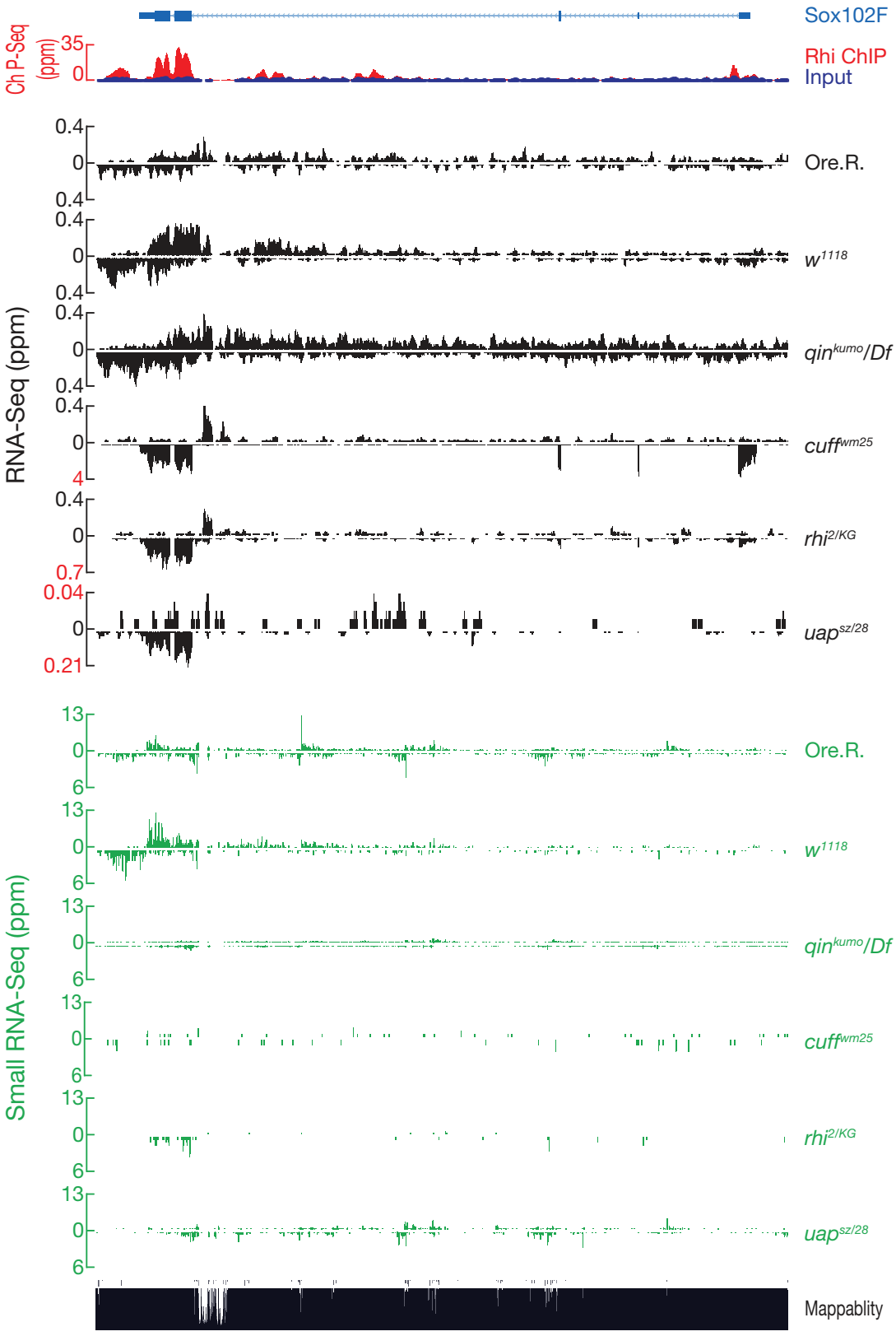
Sox102F
Ch P-Seq (ppm)
35
0
Rhi ChIP
Input
RNA-Seq (ppm)
0.4
0
0.4
Ore.R.
w1118
qinkumo/Df
cuffwm25
4
rhi2/KG
0.7
0.04
uapsz/28
0.21
Small RNA-Seq (ppm)
13
6
Mappablity

Figure 4.S4. Mutations in *rhi, cuff* and *uap56* increase splicing at the *sox102F* locus.

The experimenatlly determined intron-exon structure of the somatic *sox102F* transcript is shown in blue (top). Note that the gene is transcribed on the minus strand. Next, Rhi ChIP-Seq signal (red) is superimposed on the input control (dark blue). The following 6 tracks (black) show RNA-Seq signal in Ore. R and w^{1118} controls, and in the indicated *qin, cuff, rhi* and *uap56* mutants. The *qin* mutation, which disrupts expression of a Tudor domain protein the localizes to nuage (Zhang et al., 2012), does not increase splicing. By contrast, mutations in *cuff, rhi* and *uap56,* which encode nuclear proteins that localize to clusters, increase in splicing. piRNA expression for each genotype is indicated in the green tracks. The "mappability" of the locus, reflecting the extent of unique sequence, is shown at the bottom of the figure.

Figure 4.S5

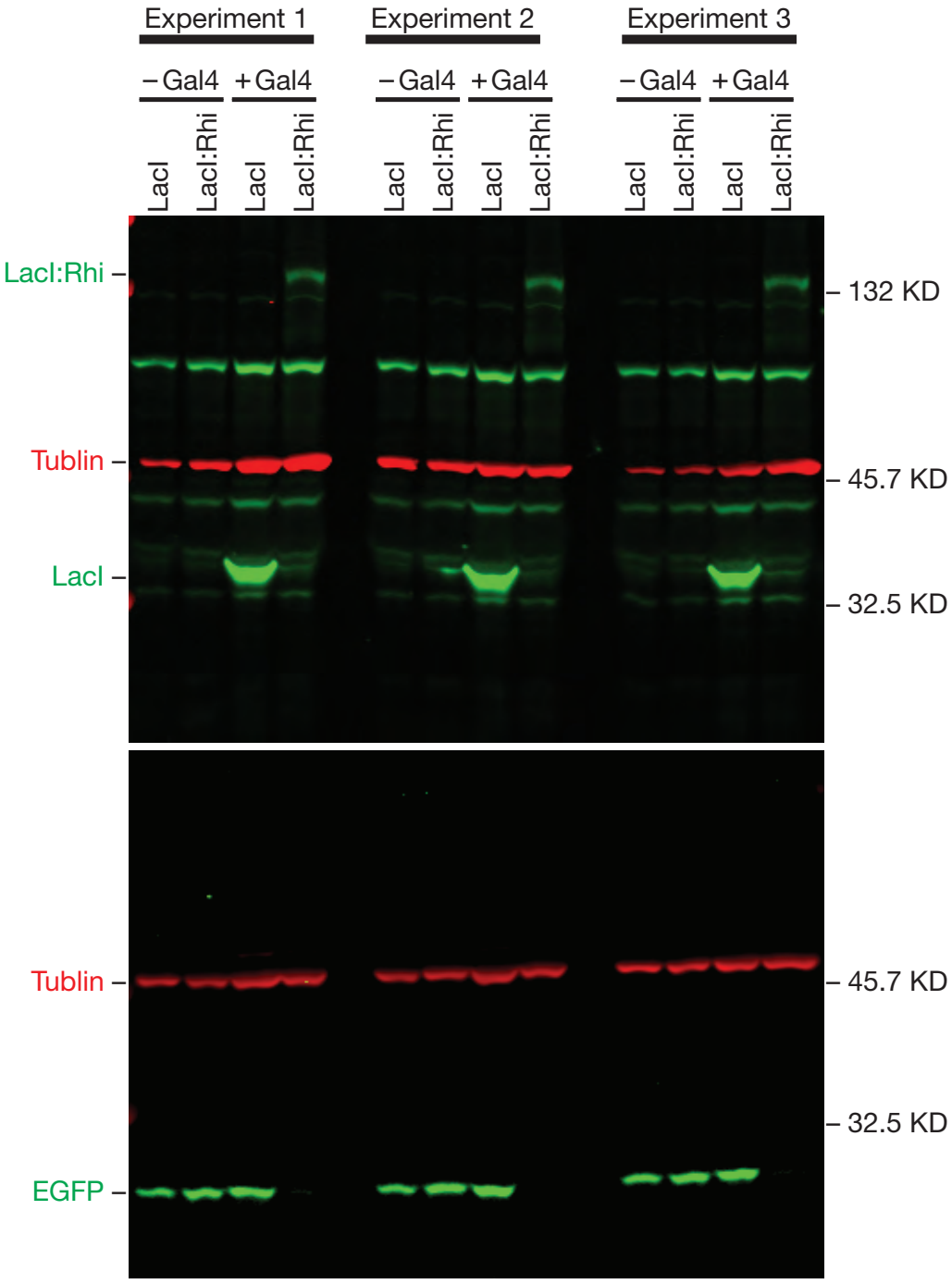
Experiment 1
Experiment 2
Experiment 3
– Gal4
+ Gal4
LacI
LacI:Rhi
LacI:Rhi
Tublin
LacI
– 132 KD
– 45.7 KD
– 32.5 KD
Tublin
EGFP
– 45.7 KD
– 32.5 KD

Figure 4.S5. Tethering LacI::Rhi, but not LacI , silences GFP protein expression. Top. Western blot showing LacI (green) and a tubulin control (red) expression in ovaries carrying inducible LacI or LacI::Rhi fusion protein genes and an EGFP reporter, in the absence of Gal4 induction (-Gal4) or the presence of Gal4 induction (+Gal4). Bottom. Parallel blot for EGFP (green) and tubulin control (red). Biological triplicate data are shown. EGFP expression is only silenced when LacI::Rhi is expressed.

Figure 4.S6

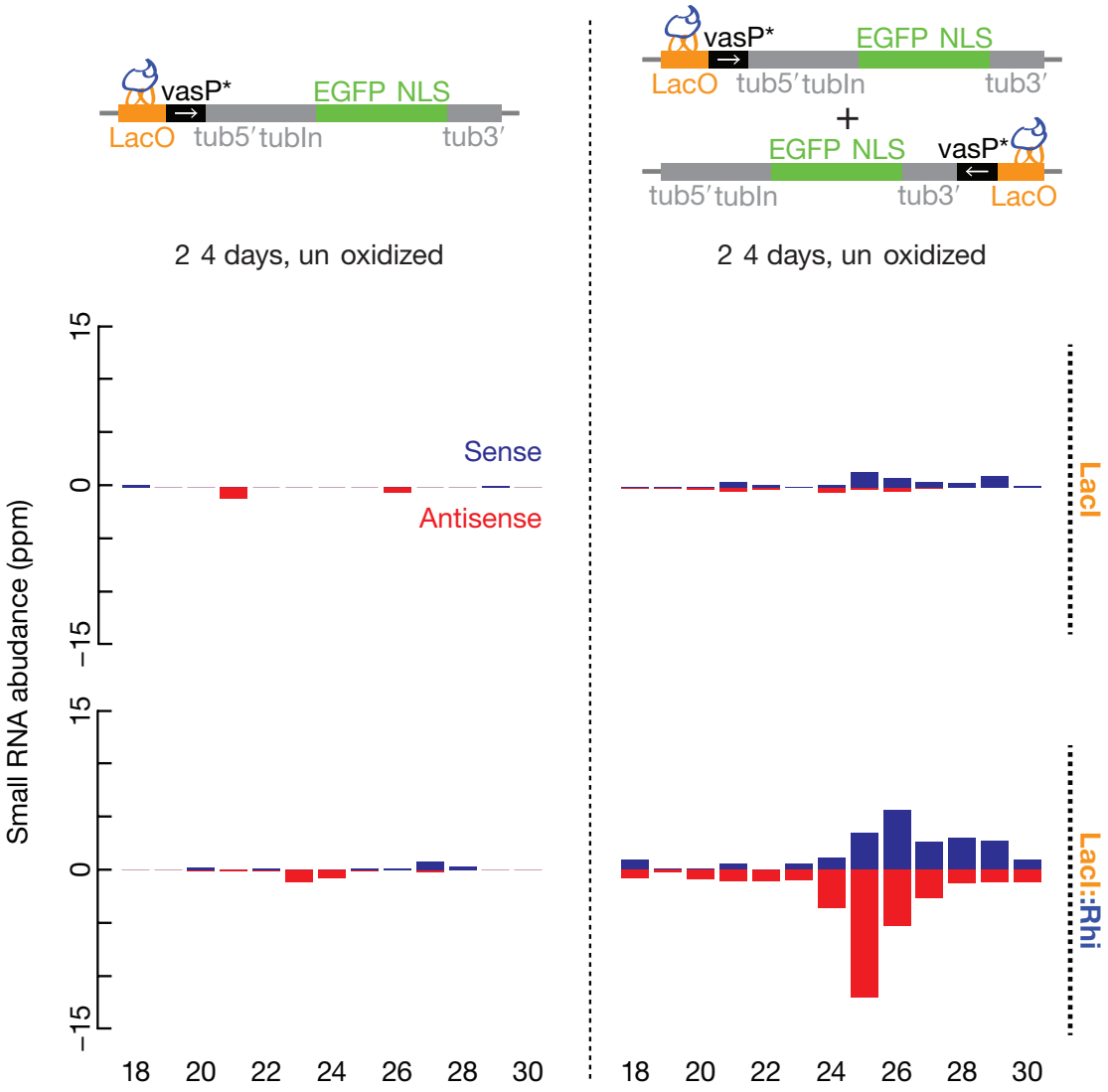
vasP*
EGFP NLS
LacO
tub5′tubIn
tub3′
+
2 4 days, un oxidized
Small RNA abudance (ppm)
15
0
−15
Sense
Antisense
18
20
22
24
26
28
30
LacI
LacI::Rhi

Figure 4.S6. Rhi binding and expression of complementary transcripts are needed to induce piRNA production.

Length distribution of the small RNAs mapping to the GFP constructs. Un-oxidized RNAs were used to generate libraries for sequencing. Blue, sense piRNAs; red, antisense.

Table 4.S1. High throughput sequencing statistics.

Table 4.S1a. ChIP-Seq samples.

Ovary genotype	Sample type	Total reads	Genome mapping reads	Uniq genome mapping reads	Uniq piRNA cluster-mapping reads	Sequencing type
Oregon R	Input	19,690,683	19,207,681	16,953,468	185,114	Single-end 36
	Rhi-ChIP	19,188,196	16,157,556	12,789,291	648,108	Single-end 36
	Pre-immune-serum ChIP	17,631,445	3,743,201	3,350,336	36,292	Single-end 36
w^{1}	Input-rep1	45,732,160	42,864,288	35,176,654	559,429	Single-end 50
	Rhi-ChIP-rep1	152,966,819	34,182,612	28,787,037	730,674	Single-end 50
	Input-rep2	21,118,300	19,687,920	17,038,394	222,400	Single-end 50
	Rhi-ChIP-rep2	30,298,852	25,665,877	20,886,439	757,626	Single-end 50
armi$^{1/72.1}$	Input-rep1	130,342,282	125,148,797	106,350,666	1,732,302	Single-end 50
	Rhi-ChIP-rep1	111,639,082	64,610,367	53,827,732	1,667,665	Single-end 50
	Input-rep2	22,516,588	21,506,254	18,476,720	322,285	Single-end 50
	Rhi-ChIP-rep2	13,795,364	12,220,648	8,839,898	753,164	Single-end 50

Table 4.S1b. Small RNA-Seq samples.

Small RNA sequencing statistics: analysis of genome matching sequences by reads. “Reads excluding ncRNA” correspond to genome-matching reads after excluding annotated non-coding RNAs (ncRNAs) such as rRNA, snRNA, snoRNA, or tRNA. “Transposon-matching reads” correspond to small RNAs mapped to *Drosophila melanogaster* transposons.

EGFP Construct Age Oxidization?	LacI/ LacI::Rhi	Total reads	Reads perfectly matching genome	ncRNA reads	Reads excluding ncRNA	miRNA-matching reads	Reads excluding ncRNA & miRNA	23–29 nt small RNA reads			
								Total	Transposon-matching reads		
									Total	Sense	Antisense
LacO-Vasp-EGFP (OVG) 2-4 days Oxidized	LacI-rep1	16,249,348	14,119,765	359,519	13,760,246	121,789	13,638,457	12,358,000	9,594,293	2,682,118	7,169,470
	LacI::Rhi-rep1	11,068,854	9,353,884	266,473	9,087,411	122,909	8,964,502	8,282,858	6,029,337	1,739,683	4,486,354
	LacI-rep2	153,157,842	127,489,611	3,255,606	124,234,005	1,083,941	123,150,064	111,624,430	86,889,772	24,302,289	64,877,410
	LacI::Rhi-rep2	177,409,973	151,678,815	4,367,112	147,311,703	1,929,109	145,382,594	134,368,774	97,914,490	28,401,266	72,641,185
LacO-Vasp-EGFP/EGFP-Vasp-LacO (OVG/GVO) 2-4 days Oxidized	LacI-rep1	46,380,606	40,244,084	822,685	39,421,399	156,822	39,264,577	35,219,730	27,366,611	8,755,773	19,149,566
	LacI::Rhi-rep1	48,735,470	41,682,162	708,211	40,973,951	180,312	40,793,639	37,315,761	28,477,289	9,319,330	19,801,667
	LacI-rep2	38,513,048	33,520,351	679,720	32,840,631	124,096	32,716,535	29,414,978	22,867,873	7,356,176	15,959,693
	LacI::Rhi-rep2	35,355,856	30,303,103	519,668	29,783,435	143,046	29,640,389	26,982,099	20,612,548	6,767,095	14,310,916
LacO-Vasp-EGFP/EGFP-	LacI-rep1	45,272,747	38,325,461	778,436	37,547,025	92,432	37,454,593	35,312,835	27,500,652	8,911,593	19,178,922

Vasp-LacO (OVG/GVO) 14-16 days Oxidized	LacI::Rhi-rep1	43,792,021	37,701,250	800,990	36,900,260	181,314	36,718,946	33,148,612	25,666,838	8,417,116	17,785,189
	LacI-rep2	32,665,847	27,627,049	552,691	27,074,358	63,953	27,010,405	25,459,224	19,823,005	6,466,333	13,782,482
	LacI::Rhi-rep2	41,996,269	35,966,320	747,541	35,218,779	166,390	35,052,389	31,613,148	24,477,416	8,064,835	16,927,858
OVG 2-4 days Un-oxidized	LacI	16,698,738	14,572,404	957,746	13,614,658	7,480,048	6,134,610	5,258,640	4,007,781	1,240,435	2,876,079
	LacI::Rhi	19,658,683	17,218,302	1,471,495	15,746,807	9,037,230	6,709,577	5,979,746	4,371,854	1,352,805	3,164,875
OVG/GVO 2-4 days Un-oxidized	LacI	27,945,613	22,347,290	1,418,311	20,928,979	7,108,237	13,820,742	12,355,771	9,507,077	3,147,024	6,582,276
	LacI::Rhi	25,523,627	20,291,554	2,269,452	18,022,102	5,693,260	12,328,842	11,056,947	8,338,308	2,778,530	5,798,249

Table 4.S1c. RNA-Seq samples.

Ovary genotype	Total reads	Genome-mapping reads	Uniquely mapping reads	Gene mapping reads	Transposon mapping reads	piRNA cluster-mapping reads	Sequencing type
Ore.R.	97,529,610	89,452,521	83,347,362	76,141,253	261,801	128,970	Paired-end 100
w^{1}-rep1	106,023,094	100,624,999	89,461,051	81,431,340	313,153	89,316	Paired-end 100
$rhi^{2/KG}$-rep1	101,764,128	90,224,659	78,162,381	66,926,117	5,206,641	387,375	Paired-end 100
w^{1}-rep2	71,051,534	67,979,397	64,136,391	58,685,981	979,334	129,752	Paired-end 100
$rhi^{2/KG}$-rep2	78,918,984	75,340,847	67,572,190	57,609,542	5,485,277	282,868	Paired-end 100
$cuff^{wm25}$	106,448,028	97,728,390	84,346,886	70,421,237	8,392,499	631,809	Paired-end 100
$uap56^{28/sz}$	108,230,898	100,104,694	78,541,631	70,768,111	2,724,098	141,180	Paired-end 100

Table 4.S2. Synthetic DNA oligonucleotides used in this study (5′-to-3′).

For quantitative RT-PCR to measure splicing ratio

sox102F-RT	TGT CCA TGA CCA TTT CCT TG
sox102F-Left	TGC AGG TAC AGG GCC TAG TT
sox102F-Right-Spliced	CTT CTA AAA AGT CAT GGG AGA GTG
sox102F-Right-Unspliced	CTT TAA TTT GTT CTA GGG GAG AGT G
42AB-RT	CTG GAA AGG CGC TCC ACT AC
42AB-Left	GCA GTT GCC GTC TCT CCT T
42AB-Right-Spliced	TGG GTC AAA GTG CAG CAG TTT T
42AB-Right-Unspliced	CGG GAA TAT AAT CGC AGC AGTT TT
rp49-RT	CGG GAA TAT AAT CGC AGC AGT TTT
rp49-Left	CCG CTT CAA GGG ACA GTA TCT G
rp49-Right	ATC TCG CCG CAG TAA ACG C
EGFP-RT	TGC TCA GGT AGT GGT TGT CG
EGFP-Right	GAA CTT CAG GGT CAG CTT GC
EGFP-Left-Spliced	ATA TGG TGA GCA AGG GCG A
EGFP-Left-Unspliced	CTC ATC CAC AGG TGA GCA AG

For ChIP-quantitative PCR

42AB	CGT CCC AGC CTA CCT AGT CA; ACT TCC CGG TGA AGA CTC

	CT
VT-q1	GCG ATA GCA CAA TGG GAA AT; GGC TTG ACA AAC GTA AAA CGA
VT-q2	CAT TTG ATG TGT TAG TGG AAA ACG; GGC AAG CTG TCG ACT TGT G
tubIn	GGC AAG CTG TCG ACT TGT G; AAC AGC TCC TCG CCC TTG
GFP3'	CGA CAA CCA CTA CCT GAG CA; ATC AGC TCG GGA TCT GAG TC
GT3	AAC AGC TCC TCG CCC TTG; CCC ATC GAG CGT TGA AGT
mocs	TCA CTG CGG ATG GAA ACA TA; GGG GAG AGA GTG TGG TGT GT
suUR	TAG CTC GTT GTC CTC GGA GT; CAC CTC AGA ATC GTT GAG CA

Table 4.S3. Published fly alleles used in this study.

$rhi^{2/KG}$	(Klattenhoff et al., 2009)
$cuff^{wm25}$	(Pane et al., 2011)
$uap56^{28/sz}$	(Zhang et al., 2012a)
qin^{kumo}/Df	Zhang et al., under revision; (Anand and Kai, 2012)
$armi^{1/72.1}$	(Cook et al., 2004)

Table 4.S4. Published high-throughput sequencing data used in this study.

$rhi^{2/KG}$-small RNA	SRP002060; (Klattenhoff et al., 2009)
$armi^{1/72.1}$-small RNA	GSE15186; (Malone et al., 2009)
$uap56^{28/sz}$-small RNA	GSE35638; (Zhang et al., 2012a)
$cuff^{wm25}$-small RNA	(Pane et al., 2011)

Ore.R.-small RNA	SRP000458; (Li et al., 2009)
W^{1118}-RNA-Seq	SRP024291; Zhang et al.; under revision
qin^{kumo}/*Df*-RNA-Seq	SRP024291; Zhang et al.; under revision

Table 4.S5. Antibody information.

anti-Rhi	Custom made, guinea pig 1943; 25 µl/ChIP
anti-GFP	Clontech, cat# 632460; Western blotting: 1:2000 dilution
anti-LacI	Rockland, cat# 600-401-B04; Western blotting: 1:2000 dilution
anti-LacI	US Biological, cat# L0899; Immunostaining: 1:500 dilution
anti-Tubulin	DSHB, cat# 12G10; Western blotting: 1:5000 dilution

SECTION V

STRAND-SPECIFIC LIBRARIES FOR HIGH THROUGHPUT SEQUENCING OF RNA (RNA-SEQ) PREPARED WITHOUT POLY(A) SELECTION

SUMMARY

High throughput DNA sequencing technology has enabled quantification of all the RNAs in a cell or tissue, a method widely known as RNA-Seq. However, non-coding RNAs such as rRNA are highly abundant and can consume >70% of sequencing reads. A common approach is to extract only polyadenylated mRNA; however, such approaches are blind to RNAs with short or no poly(A) tails, leading to an incomplete view of the transcriptome. Another challenge of preparing RNA-Seq libraries is to preserve the strand information of the RNAs. Here, we describe a procedure for preparing RNA-Seq libraries from 1–4 μg total RNA without poly(A) selection. Our method combines the dUTP/uracil-DNA glycosylase strategy to achieve strand specificity with AMPure XP magnetic beads to perform size selection. Together, these steps eliminate gel purification, allowing a library to be made in less than two days. We barcode each library during the final PCR amplification step, allowing several samples to be sequenced in a single lane without sacrificing read length. Libraries prepared using this protocol are compatible with Illumina GAII, GAIIx and HiSeq 2000 platforms. The RNA-Seq protocol described here yields strand-specific transcriptome libraries without poly(A) selection that provide ~90% mappable sequences. Typically, more than 85% of mapped reads corresponded to protein coding genes, and only ~6% derive from non-coding RNAs. The protocol has been used to measure RNA transcript identity and abundance in tissues from flies, mice, rats, chickens, and frogs, demonstrating its general applicability.

INTRODUCTION

Strand-specific RNA-Seq provides a powerful tool for transcriptome analysis. Besides measuring transcript abundance across the entire transcriptome, RNA-Seq facilitates de novo transcript annotation and assembly, quantification of splice site usage, and identification of mutations or polymorphisms between samples (Wang et al., 2009b; Metzker, 2010; Ozsolak and Milos, 2011a). Ribosomal RNAs compose an overwhelming fraction of the total RNA population (>70%) and can occupy most of the sequencing space, leaving little room for investigating other transcripts (Armour et al., 2009). The most widely used strategy employs poly(A) selection to enrich for RNA polymerase II transcripts, but this strategy cannot be used to study RNAs lacking poly(A) tails or precursor transcripts processed into fragments that have lost their poly(A) tails, e.g., 7SL RNA, 7SK RNA, the 5′ fragment of Argonaute cleavage products, processed products of piRNA precursors, and long non-coding RNAs such as Kcnq1ot1 in mammals (Huang et al., 2011b). Another strategy removes rRNA by hybridization while retaining other non-adenylated RNAs for sequencing (Huang et al., 2011b).

Although RNA can be sequenced directly, without conversion to complementary DNA (cDNA), current high throughput technologies for direct RNA sequencing have short read lengths (25–55 nt; median ~33 nt) and high error rates (~4%) (Ozsolak et al., 2009; Ozsolak and Milos, 2011b). Thus current strategies for transcriptome analysis all typically convert RNA to cDNA before sequencing, notwithstanding the artifacts that may result from template switching or structural RNA self-priming (Mader et al., 2001; Cocquet et al., 2006;

Chabot et al., 2008). The dUTP method, one of the leading cDNA-based strategies, provides excellent library complexity, strand specificity, coverage evenness, agreement with known annotation, and accuracy for expression profiling (Levin et al., 2010). In this method, the RNA is first reverse transcribed into cDNA:RNA using random primer. To synthesize the second cDNA strand, dUTP instead of dTTP is used, marking the second cDNA strand for subsequent degradation with uracil-DNA glycosylase (UDG) to preserve strand information (Parkhomchuk et al., 2009; Wang et al., 2011; Sultan et al., 2012).

Here, we describe a protocol for preparing strand-specific RNA-Seq libraries that combines rRNA removal using the Ribo-Zero kit and the dUTP method for ensuring strand specificity (Figure 5.1). Our protocol showed advantage in time saved, cost and performance (Table 5.1). We replace laborious, time-consuming gel purification steps with AMPure XP beads, whose size-selectivity and efficiency of DNA recovery allow the use of small amounts of starting RNA (Lott et al., 2011; Hawkins et al., 1994; Lis, 1980). The high sequencing depth of the Illumina HiSeq 2000 platform can easily generate >170 million reads per lane, allowing multiple barcoded samples to be pooled and sequenced in a single lane. One common method to index a library is adding barcodes during adapter ligation, so that the first 5 or 6 nucleotides of each read is the barcode. However, this strategy sacrifices read length, can increase the error rates at the 5′ or 3′ ends of reads (Wang et al., 2011), can perturb the calibration of the Illumina base calling algorithm (the HiSeq2000 platform uses the first 5 nucleotides for calibration), and may lead to differential ligation efficiency and specificity among barcoded samples. Introducing barcodes during

the final PCR amplification (Figure 5.1) bypasses these problems. The barcodes are then read using a separate primer and additional sequencing cycles after the insert has been sequenced (Figure 5.2). We modified the Illumina Multiplexing Sample Prep oligonucleotides and used 12 barcoded primers to index 12 libraries at the final PCR step (Figure 5.1). Our protocol requires only 1–4 μg total RNA as starting material and takes no more than two days to complete.

Figure 5.1

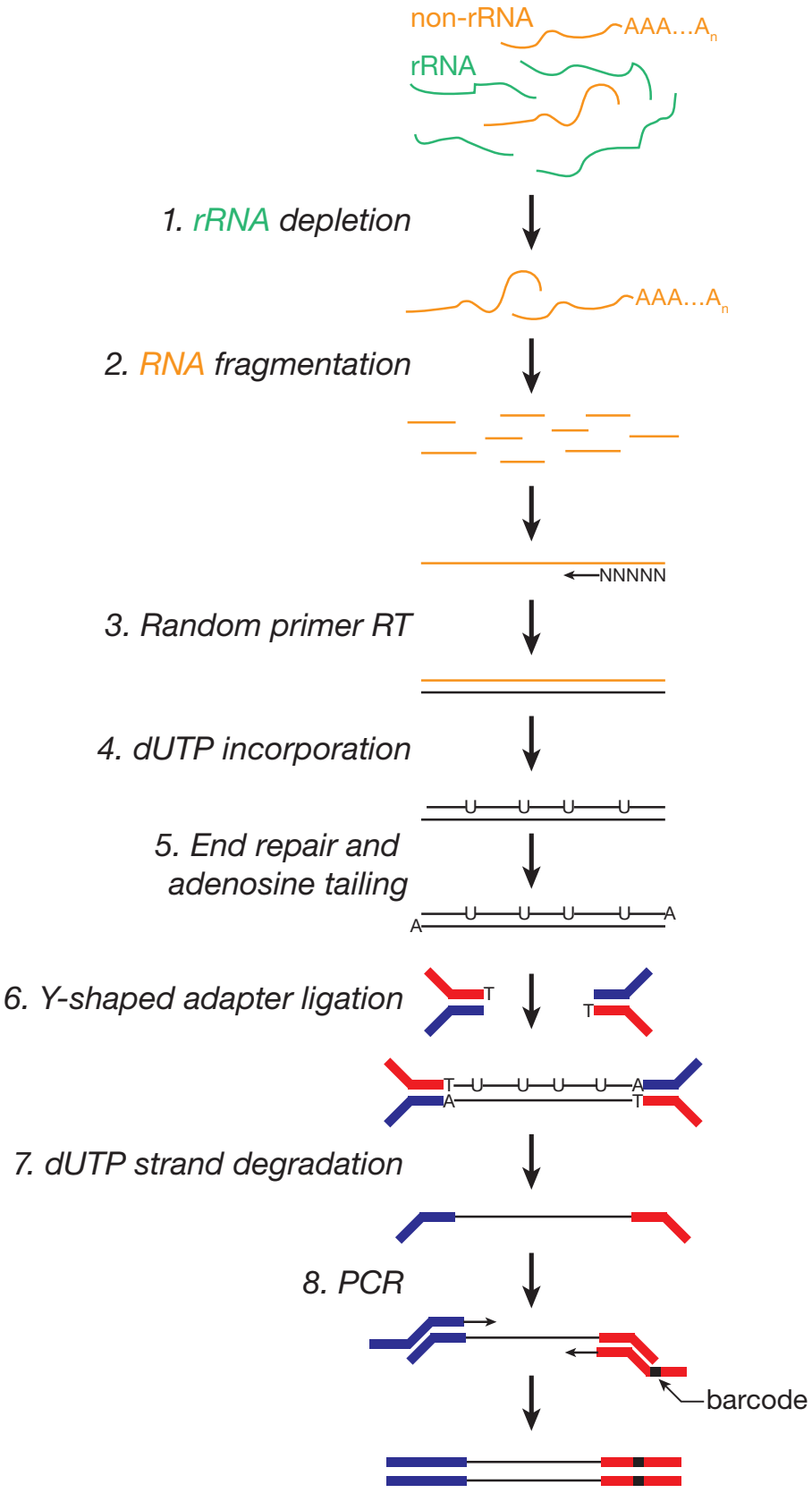
non-rRNA
AAA...An
rRNA
1. rRNA depletion
AAA...An
2. RNA fragmentation
NNNNN
3. Random primer RT
4. dUTP incorporation
U U U U
5. End repair and
adenosine tailing
A
U U U U A
6. Y-shaped adapter ligation
T
T
T U U U U A
A T
7. dUTP strand degradation
8. PCR
barcode

Figure 5.1. Protocol workflow.

Figure 5.2


Read 2 Sequenc ng Pr mer
TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTG 5′
5′AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT---cDNA---AGATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNATCTCGTATGCCGTCTTCTGCTTG 3′
3′TTACTATGCCGCTGGTGGCTCTAGATGTGAGAAAGGGATGTGCTGCGAGAAGGCTAGA---cDNA---TCTAGCCTTCTCGTGTGCAGACTTGAGGTCAGTGNNNNNNTAGAGCATACGGCAGAAGACGAAC 5′
5′ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Read 1 Sequenc ng Pr mer
5′GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Barcode Sequenc ng Pr mer

Figure 5.2. Library and sequencing primer sequences.

METHODS

Enzymes

TURBO DNase (2 U/μl, Ambion, Cat. No. AM2239)

SuperScript III Reverse Transcriptase (200 U/μl, Invitrogen, Cat. No. 18080-093)

RNase H (2 U/μl, Invitrogen, Cat. No. 18021-014)

DNA polymerase I (10 U/μl, New England Biolabs [NEB], Cat. No. M0209L)

T4 DNA polymerase (3 U/μl, NEB, Cat. No. M0203)

Klenow DNA polymerase (5 U/μl, NEB, Cat. No. M0210L)

T4 PNK (10U/μl, NEB, Cat. No. M0236L)

Klenow 3′ to 5′ exo$^-$ (5 U/μl, NEB, Cat. No. M0212L)

T4 DNA ligase (600 U/μl, Enzymatics, Cat. No. L603-HC-L)

E. coli Uracil-DNA glycosylase (5 U/μl, NEB, Cat. No. M0280S)

Phusion High-Fidelity DNA Polymerase (2 U/μl, NEB, Cat. No. M0530L)

AccuPrime Pfx DNA Polymerase (optional, 2.5 U/μl, Invitrogen, Cat. No. 12344-024)

Buffers and Reagents

Ribo-Zero magnetic kit (Epicentre, Cat. No. MRZH116)

RNA Clean & Concentrator-5 (Zymo Research, Cat. No. R1015)

Random primers (Hexamers, 3 μg/μl, Invitrogen, Cat. No. 48190011)

dATP (Bio Basic, Cat. No. DD0058)

dCTP (Bio Basic, Cat. No. DD0058)

dGTP (Bio Basic, Cat. No. DD0058)

dTTP (Bio Basic, Cat. No. DD0058)

dUTP (Bio Basic, Cat. No. DM1244)

AMPure XP beads (Beckman Coulter, Cat. No. A63880)

Elution Buffer for AMPure XP beads: 10 mM Tris-HCl (pH 8.5)

Zero Blunt® TOPO® PCR cloning kit (Invitrogen, Cat. No. K2800-20)

LB agar kanamycin plates: 1% [w/v] tryptone, 0.5% [w/v] yeast extract, 1% [w/v] NaCl, 1.5% [w/v] agar and 50 μg/ml kanamycin

GoTaq Green Master Mix (Promega, Cat. No. M7122)

TOP10 *E. Coli* competent cells (home-made (Chung and Miller, 1988, Nucleic acids research, 16, 3580-3580) or Invitrogen, Cat. No. C4040-10)

S.O.C. medium (super optimal broth with catabolite repression): 0.5% [w/v] yeast extract, 2% [w/v] tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10mM $MgSO_4$ and 20 mM glucose (sterilize the glucose stock separately using a 0.2 μm filter, and then add it into the rest ingredients, which should be sterilized by autoclaving).

70% [v/v] ethanol

2nd strand buffer/10× NEB Buffer 2: 500 mM NaCl, 100 mM Tris-HCl (pH 7.9), 100 mM $MgCl_2$ and 10 mM DTT.

10× T4 DNA ligase buffer: 500 mM Tris-HCl (pH 7.5), 100 mM $MgCl_2$, and 100 mM DTT, 10 mM ATP (note, ATP freshly added before use).

100 bp DNA ladder (e.g., Fermentas, Cat. No. SM0241)

5× First strand buffer: 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$, 50 mM DTT.

2× Rapid ligation buffer: 132 mM Tris-HCl (pH 7.6), 20 mM $MgCl_2$, 2 mM DTT, 15% PEG6000, 2 mM ATP (note, ATP freshly added before use)

Actinomycin D (optional, Sigma, Cat. No. A1410)

Equipment

Water bath or heat block

Magnetic stand for 1.5 ml centrifuge tubes

Bench top centrifuge for 1.5 ml centrifuge tubes (17,000 × *g* required)

PCR thermal cycler

Nanodrop or comparable low-volume spectrophotometer

Bioanalyzer (Optional)

37 °C Incubator

Bench top vortexer

DNA oligonucleotides

Multiplexing adapters

Adapter oligo 1: 5′-pGAT CGG AAG AGC ACA CGT CT-3′

Adapter oligo 2: 5′-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT-3′

PCR primers (barcode)

Primer 1: 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3′

Primer 2 (primer with barcode, designed by combining Illumina Multiplexing PCR Primer 2.0 and PCR Index primer into a single primer):

Primer 2-1: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **CGT GAT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-2: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **ACA TCG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-3: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **GCC TAA** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-4: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **TGG TCA** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-5: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **CAC TGT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-6: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **ATT GGC** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-7: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **GAT CTG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-8: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **TCA AGT** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-9: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **CTG ATC** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-10: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **AAG CTA** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-11: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **GTA GCC** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

Primer 2-12: 5′-CAA GCA GAA GAC GGC ATA CGA GAT **TAC AAG** GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′

M13 Forward: 5′-GTA AAA CGA CGG CCA G-3′

M13 Reverse: 5′-CAG GAA ACA GCT ATG AC-3′

Procedure

Ribosomal RNA (rRNA) depletion

High quality total RNA is essential for efficient rRNA removal. For example, in our hands, *Drosophila* RNA subjected to repeated freeze-thawing or treated with DNase cannot be efficiently depleted of ribosomal RNA.

1. Mix the Ribo-Zero magnetic beads by gently pipetting. For each total RNA sample, dispense 225 μl Ribo-Zero magnetic beads into an RNase-free 1.5 ml centrifuge tube.
2. Place the tube in the magnetic stand until the supernatant becomes clear, ~1 min.
3. With the tube still in the stand, discard the supernatant, which contains 0.1% sodium azide (chemical hazard: dispose of according to local regulations).
4. Add 225 μl RNase-free water to the tube, remove the tube from the magnetic stand, and mix the beads by gently pipetting.
5. Return the tube to the magnetic stand, wait until the solution becomes clear, and discard the water.
6. Resuspend the beads in 65 μl Ribo-Zero magnetic bead suspension solution and 1 μl RiboGuard RNase Inhibitor. Mix well by gently pipetting. Store the tube at room temperature until step 9.
7. In a 1.5 ml centrifuge tube, prepare the following mix

   4 μg fresh total RNA

   4 μl Ribo-Zero "Reaction" buffer

   10 μl Ribo-Zero rRNA removal solution

Add water to make a 40 μl total volume.

Store the unused Ribo-Zero rRNA removal solution and "Reaction" buffer at −80 °C.

8. Gently mix the solution by pipetting and incubate at 68 °C for 10 min, then incubate the tube at room temperature for 5 min.
9. Gently mix the magnetic beads from step 6 by pipetting and add the RNA solution from step 8 to the mixed beads. Using the same pipet tip, immediately mix the beads with the RNA by pipetting 10 times. Next, vortex the tube for 10 sec at medium speed. Finally, incubate the mixture at room temperature for 5 min.
10. Vortex the tube at medium speed for 5 sec and then incubate it at 50 °C for 5 min.
11. After the 5 min incubation, immediately place the tube in the magnetic stand for 2 min.
12. Carefully remove the supernatant, about 84 μl, to a new 1.5 ml centrifuge tube and place it in the magnetic stand for 1 min to get rid of the trace amount of leftover beads from last step.
13. Pipette the supernatant into a new 1.5 ml centrifuge tube and add 16 μl water.

Size selection and DNase treatment

RNA Clean & Concentrator-5 is used to enrich for RNAs >200 nt, which also removes 5S rRNA and tRNA.

14. Mix 100 μl RNA binding buffer with 100 μl 100% ethanol. Add this 200 μl mixture to the 100 μl RNA from step 13.

15. Transfer the buffer/ethanol/RNA mixture into a Zymo-Spin IC column in a collection tube. Centrifuge at 17,000 × *g* for 1 min. Discard the flow-through.
16. Add 400 μl RNA Wash Buffer to the column, spin at 17,000 × *g* for 1 min. Discard the flow-through.
17. To degrade contaminating DNA, mix the following reagents (to handle multiple samples at one time, we prefer to prepare the pre-mix for easy operation and less pipetting variance among samples)

    3 μl TURBO DNase (2 U/μl)

    3 μl 10× TURBO DNase Buffer

    24 μl RNA Wash Buffer

    and add the 30 μl mixture to the column. Incubate the column at 37 °C for 30 min. Centrifuge the column at 17,000 × *g* for 1 min, and discard the flow-through.
18. Add 400 μl RNA Prep Buffer to the column, centrifuge at 17,000 × *g* for 1 min, and discard the flow-through.
19. Add 800 μl RNA Wash Buffer to the column, centrifuge at 17,000 × *g* for 1 min, and discard the flow-through.
20. Add 400 μl RNA Wash Buffer to the column, centrifuge at 17,000 × *g* for 1 min, and discard the flow-through.
21. Centrifuge the column at 17,000 × *g* for 2 min.
22. To elute the RNA, replace the collection tube with a new 1.5 ml centrifuge tube, and then add 10 μl water to the column. Incubate at room temperature for 1 min and centrifuge at 17,000 × *g* for 1 min to collect the RNA/flow-through.

23. Take 1 μl RNA to measure the concentration using a Nanodrop spectrophotometer or comparable low volume instrument. Typically, the procedure yields 10–20 ng/μl (A260/A280 = 1.96 ~ 2.17).

Library preparation

1. Fragment the RNA. At high temperature, the metal ions within the 5× First strand buffer will hydrolyze the RNA into short fragments. In a 0.2 ml tube, mix 4 μl rRNA-depleted total RNA with 4 μl of 5× First strand buffer. Place the tube into a PCR thermal cycler pre-heated to 94 °C. Incubate for precisely 4 min and 50 sec. Then quickly chill the tube on ice for at least 1 min.
2. To reverse transcribe the RNA into first-strand cDNA, add to the PCR tube:

    1.5 μl 100 mM DTT

    1 μl Random primer (Hexamers, 3 μg/μl)

    7 μl water

    Incubate at 65 °C for 3 min, and then quickly chill the tube on ice for 1 min. Next, add:

    1 μl dNTP mixture (dATP, dCTP, dGTP, dTTP, 10 mM each)

    0.5 μl 100 mM DTT

    1 μl SuperScript III Reverse Transcriptase (200 U/μl)

    4 μg Actinomycin D (optional, may enhance strand specificity, but decrease uniformity of strand coverage (Levin et al., 2010, #51637))

Incubate at 25 °C for 5 min, then at 50 °C for 1 h. Heat at 70 °C for 15 min to inactive the reverse transcriptase. Finally, use 36 μl AMPure XP beads to purify the cDNA, eluting with 22 μl elution buffer. (A detailed protocol for purifying with AMPure XP beads follows this protocol.)

3. To convert the first-strand cDNA to double-stranded cDNA incorporating dUTP instead of dTTP, add the following to the cDNA from step 2:

    3 μl 2nd strand buffer/10× NEB Buffer 2

    2 μl dUTP mixture (20 mM dUTP, 10 mM dATP, dCTP, dGTP)

    1 μl RNase H (2 U/μl)

    2 μl DNA polymerase I (10 U/μl)

    0.5 μl 100 mM DTT

    Incubate at 16 °C for 2.5 h.

    After incubation, purify the double-stranded cDNA using 45 μl AMPure XP beads and elute the cDNA into a 1.5 ml centrifuge tube using 33 μl elution buffer. After this, one can continue or store the sample at −20 °C.

4. Repair the ends of the double-stranded cDNA. DNA polymerase I, which is used to for second strand cDNA synthesis, uses as primers the RNA leftover from RNase H digestion. Consequently, the double-stranded cDNAs generated in step 3 have 3′ overhanging ends. Step 4 converts the sticky ends into blunt ends. Add the following mixture to the DNA from step 3:

    5 μl 10× T4 DNA ligase buffer

    2 μl dNTP mixture (10 mM each)

    5 μl T4 DNA polymerase (3 U/μl)

1 μl Klenow DNA polymerase (5 U/μl)

5 μl T4 PNK (10 U/μl)

Incubate at 20 °C for 30 min.

5. To establish a library with the narrow size range (200 bp to 350 bp) required for successful high throughput sequencing, the cDNA is purified using AMPure XP beads, which exploit the finding that carboxyl coated magnetic beads bind distinct DNA size ranges depending on Polyethylene Glycol (PEG) and salt concentration (Hawkins et al., 1994; Lis, 1980, Methods in enzymology, 65, 347; http://epicentral.blogspot.com/2010/06/get-rid-of-small-stuff.html, #4688). After end repair, mix the reaction with 35 μl AMPure XP beads and incubate at room temperature for 5 min. Then place the tube in the magnetic stand for 3 min. Transfer the supernatant into a new tube and discard the beads. To the new tube with the supernatant, add an additional 40 μl of AMPure XP beads and then follow the standard AMPure XP bead purification protocol, eluting the DNA with 33 μl elution buffer.
6. Tail the PCR products with adenosine to facilitate adapter ligation. We use Klenow fragment with D355A and E357A mutations (Klenow 3′-to-5′ exo$^-$, 5 U/μl), a DNA polymerase lacking both 3′-to-5′ and 5′-to-3′ exonuclease activities, to add a single adenosine to the 3′ ends of the DNA. To the DNA from step 5, add:

5 μl 2nd strand buffer/10× NEB Buffer 2

1 μl dATP (10 mM)

3 μl Klenow 3′-to-5′ exo⁻ (5 U/μl)

9 μl water

Incubate at 37 °C for 30 min. After incubation, purify the DNA using 60 μl AMPure XP beads; elute with 24 μl elution buffer.

7. Add the Y-shaped adapters (Oligonucleotide sequences © 2007-2011 Illumina, Inc. All rights reserved.). To prepare the Y-shaped adapter, mix 25 μl Adapter oligo 1 and oligo 2 (each at 50 μM stock concentration). Heat at 95 °C for 2 min, then ramp down slowly to room temperature. We usually heat the oligo mixture in an aluminum heat block for 2 min. Then remove the block from the heater and let it cool down to room temperature, ~30 min. Ligate the adapters to the purified double-stranded cDNA by adding:

   25 μl 2× Rapid ligation buffer

   1 μl Adapter (10 μM)

   1.5 μl T4 DNA ligase (600 U/μl)

   Incubate at room temperature for 15 min. After incubation, use 50 μl AMPure XP beads to purify the DNA. Elute the DNA with 30 μl elution buffer.

8. Treat with uracil-DNA glycosylase (UDG, 5 U/μl). Add 2 μl UDG to the DNA from step 7, and incubate at 37 °C for 30 min.

9. PCR amplify the cDNA. Add the following mixture to the DNA from step 8:

   10 μl 5× HF buffer (Phusion Polymerase)

   1 μl 10 μM PCR Primer 2 (one of the 12 to provide the barcode)

1.5 μl dNTP (10 mM each)

0.5 μl Phusion High-Fidelity DNA polymerase (2 U/μl)

5 μl water

Incubate the tube at 98 ˚C for 40 sec, 65 ˚C for 30 sec and 72 ˚C for 30 sec. After the incubation, pause the PCR machine, and then add 1 μl 10 μM PCR Primer 1. Continue the PCR with 10 cycles of 98 °C for 10 sec, 65 °C for 30 sec, 72 °C for 30 sec, followed by incubation at 72 °C for 3 min. Then purify the library with 50 μl AMPure XP beads. Finally, elute the DNA with 20 μl elution buffer.

Alternatively, the PCR can be performed using AccuPrime Pfx DNA Polymerase:

5 μl 10× AccuPrime Pfx Reaction mix

1 μl 10 μM PCR Primer 2 (one of the 12 to provide the barcode)

1 μl AccuPrime Pfx DNA Polymerase (2.5 U/μl)

11 μl water

Incubate the tube at 95 ˚C for 40 sec, 65 ˚C for 30 sec and 68 ˚C for 30 sec. After the incubation, pause the PCR machine, and then add 1 μl 10 μM PCR primer 1. Continue the PCR with 10 cycles of 95 °C for 15 sec, 65 °C for 30 sec, 68 °C for 30 sec, followed by incubation at 68 °C for 3 min. Then purify the library with 50 μl AMPure XP beads. Finally, elute the DNA with 20 μl elution buffer.

Quality control (optional)

A good RNA-Seq library should have a narrow size range, retain strand information, and contain little RNA contamination from other species and few

inserts from ribosomal RNA. Because high throughput sequencing remains expensive and time consuming, assaying the quality of libraries before submitting a sample for sequencing is worthwhile, particularly when first establishing this workflow. Here, we provide two ways to test library quality: Bioanalyzer analysis and small-scale colony sequencing. Using only 1 μl of the library, Bioanalyzer analysis provides fast, sensitive and accurate information on insert size distribution. Small-scale colony sequencing provides information on insert identity.

1. Bioanalyzer analysis. Run 1 μl of RNA-Seq library on the Agilent High Sensitivity DNA Chip to check the size distribution of the library.
2. Small-scale colony sequencing. Mix:

   0.5 μl RNA-Seq library

   0.5 μl salt solution

   0.5 μl pCRII-Blunt-TOPO

   1.5 μl water

   Incubate at room temperature for 5 min. Pipette the reaction into 50 μl TOP10 *E. Coli* competent cells. To mix, gently tap the tube 5 times (do not mix by pipetting to avoid breaking the fragile cells), then incubate on ice for 20 min. Heat shock at 42 °C for 40 sec. Immediately chill on ice for 2 min, and then add 200 μl room temperature S.O.C. medium. Recover the cells at 37 °C for 1 h with orbital shaking (200 rpm).
3. Spread 60 μl of the bacterial suspension on an LB plate containing 50 μg/ml kanamycin, and incubate at 37 °C overnight.

4. Pick 10–20 colonies from each library. Gently touch a 10 μl pipette tip to a single colony, and then pipet with 12.5 μl GoTaq Green Master Mix, 1 μl of M13 Forward and Reverse Primer mix (5 μM each) and 11.5 μl water. Run the following PCR program:

    94 °C 5 min

    94 °C 30 sec

    55 °C 30 sec

    72 °C 40 sec (go to step 2 for an additional 30 cycles)

    72 °C 7 min

    4 °C hold

5. Run 5 μl of the PCR product on a 1.5% agarose gel to check the size and quality of the PCR product. Use a 100 bp DNA ladder as size markers.
6. Sanger-sequence each PCR reaction that produces a product of a single length, using the M13 Reverse primer.

Additional protocol: using AMPure XP beads to purify DNA

1. Warm the AMPure XP beads to room temperature, mix well, and pipette the required volume from the stock bottle to the sample tube.
2. Mix the beads and DNA sample by gently pipetting, and then incubate at room temperature for 5 min.
3. Place the tube in the magnet stand for 5 min until the supernatant appears clear.
4. Discard the supernatant.
5. Keep the tube in the stand and add 180 μl of 70% (v/v) ethanol into the tube without disturbing the beads.

6. Wait for 30 sec and discard the ethanol supernatant.
7. Repeat steps 5 and 6.
8. To remove any ethanol remaining on the sides of the tube, centrifuge the tube at 1000 × *g* for 1 min.
9. Place the tube in the magnetic stand for 30 sec, and then remove any residual ethanol using a 10 μl pipette.
10. Add the specified volume of elution buffer to the beads and pipette to mix.
11. Wait 3 min, and then place the tube in the magnetic stand for 3 min.
12. Use a 10 μl pipette to carefully transfer the eluted DNA to a new tube (sacrifice 1–2 μl to avoid carrying over any beads).

ANTICIPATED RESULTS

Quality control

Based on our experience, a good library will range in size from 300 to 500 bp, including 122 bp from the PCR primers plus 200–350 bp from the RNA inserts. Bioanalyzer analysis should show a peak at 320–330 bp (Figure 5.3A).

Small-scale colony sequencing can reveal the size and sequence of the inserts and barcodes for a small but representative sample of the library. When preparing libraries for the first time, sequencing ~20 colonies per library serves to validate successful library construction. The PCR amplification products should be ~600 bp (244 bp from the pCRII-Blunt-TOPO vector, 122 bp from the PCR primers, plus the RNA insert; Figure 3B). Expect one or two colonies to lack inserts, giving a 366 bp PCR product (Figure 5.3B). Of the remaining 15–18 successfully sequenced colony PCR products, one or two may derive from rRNA.

High throughput sequencing

The number of samples mixed in one sequencing lane depends on the genome size of the organism and the purpose of the research. To study low abundance RNAs from the repetitive region of the *Drosophila* genome, we usually pool four barcoded samples in a single lane of the HiSeq 2000 instrument, and sequence the libraries as 100 nt paired-end reads. We typically obtain >170,000,000 fragments per lane. For example, in one experiment, we obtained 175,991,972 fragments (for paired-end sequencing, each fragment has two reads, a total of 351,983,944 reads). Among them, 349,247,868 (99.2%) reads were successfully sorted by the barcodes.

Using TopHat (Langmead and Salzberg, 2012; Trapnell et al., 2009; Trapnell et al., 2012) to map reads to the fly genome (parameters: tophat –i 50 –p 24 –library-type fr-firststrand –G gene.gtf --coverage-search --segment-length 25 –o output_directory_name), we typically achieve 90% mappability. For example, for a typical library, 91.7% of reads mapped to the fly genome. Among the mapped reads, only 4.03% were singletons (i.e., only one of the paired fragments in the read mapped); both fragments mapped for the rest. Finally, more than 85% of mapped reads corresponded to protein coding genes, and only 6.20% derived from non-coding RNAs such as rRNA, tRNA, snRNA or snoRNA. We have used this protocol to produce libraries of similar quality from wild-type and mutant mouse tissues, as well tissues from wild-type rat, chicken, and frog, demonstrating its general applicability.

Figure 5.3

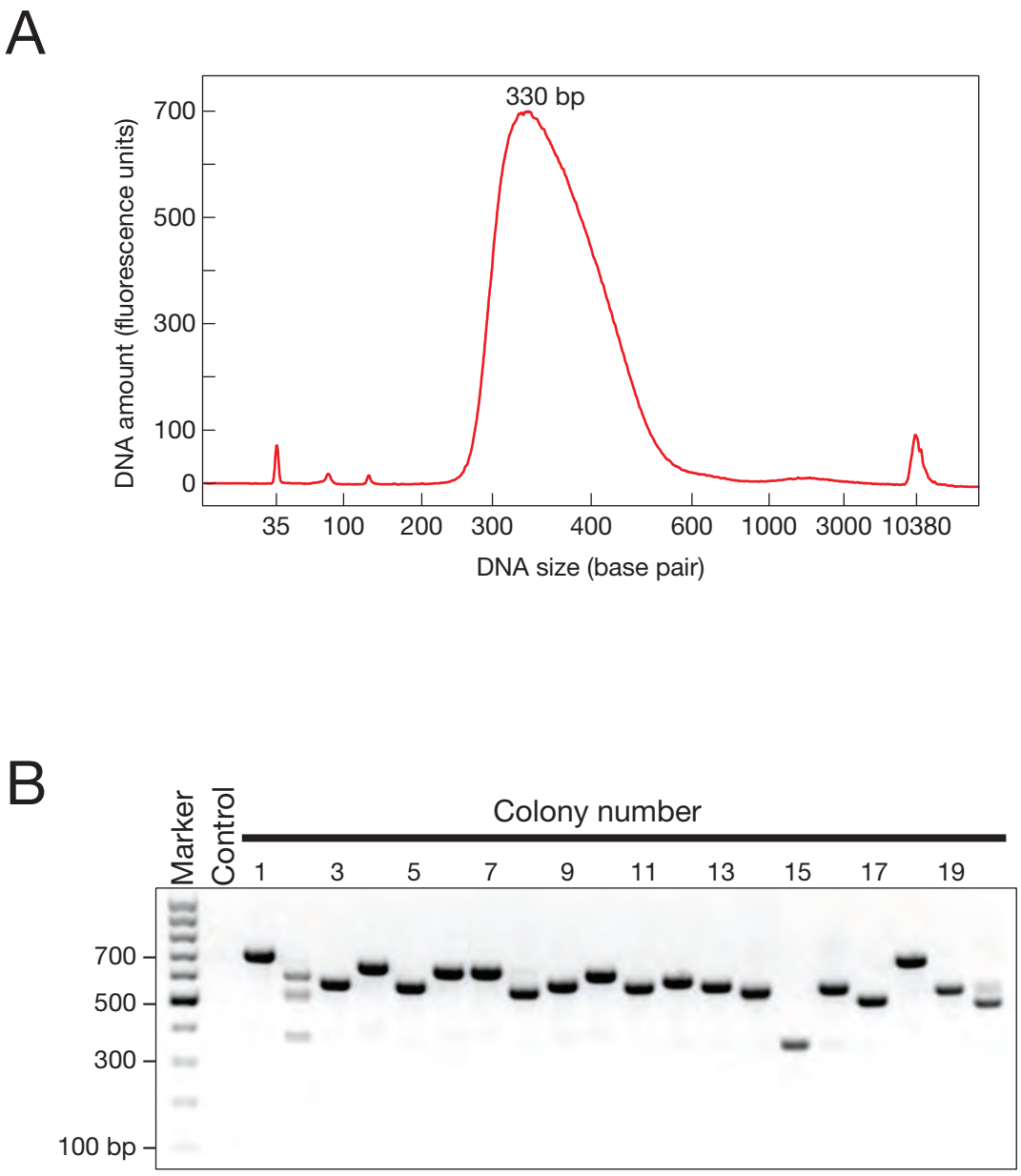
A
330 bp
700
500
300
100
0
DNA amount (fluorescence units)
35
100
200
300
400
600
1000
3000
10380
DNA size (base pair)
B
Marker
Control
Colony number
1
3
5
7
9
11
13
15
17
19
700
500
300
100 bp

Figure 5.3. Anticipated results.

(A) Bioanalyzer plot.

(B) Agarose gel for small-scale colony sequencing. Control, a PCR reaction with no bacterial colony added.

ACKNOWLEDGMENTS

We thank Ryuya Fukunaga, Chengjian Li, Liang Meng Wee for helpful discussions, Jie Wang for advice on TopHat mapping, and members of our laboratories for comments on the manuscript. This work was supported in part by National Institutes of Health grants HD049116 (W.E.T., Z.W., and P.D.Z.) and GM62862 and GM65236 (P.D.Z.).

CHAPTER VI: OPEN QUESTIONS

The piRNA pathway silences transposons in the germline, and is therefore essential to maintain germline genome stability and integrity. Due to its significance, tremendous effort has been made to dissect this pathway in the past seven years, yet still little is known about piRNA biogenesis and function. Moreover, for the models that have currently been proposed, most of the details remain to be confirmed. The research presented in this dissertation contributes to our understanding this pathway, but it also raises even more questions that need to be addressed.

piRNA biogenesis

We showed that in fly ovaries, the HP1 family protein Rhi marks piRNA clusters in germ cells for piRNA production. However, how Rhi is recruited to these regions is still largely unknown. Since HP1 coats DNA through its Chromo domain and the modified histone protein, H3K9me3 (Danzer and Wallrath, 2004), Rhi may find cluster loci in a similar manner. Indeed, based on sequence similarity between Rhi and HP1, Rhi preserves the same amino acid residues in its Chromo domain as have been shown are required for HP1 and H3K9me3 interaction. Moreover, Lehmann and colleagues have shown that piRNA cluster regions are enriched for this modified histone (Rangan et al., 2011). However, whether H3K9me3 itself is sufficient to recruit Rhi still needs to be determined.

In Chapter 4, we showed data that suggests Rhi binds to clusters independent of piRNA production. This set of experiments was performed in *armi* mutants, in which the unique mapping piRNA species are eliminated. However, the piRNAs shared by clusters and transposon consensus, which ambiguously map to multiple locations on the genome, are still partially

preserved in this mutant (Malone et al., 2009). Therefore, we cannot rule out the possibility that these multiple mapping piRNAs are necessary for Rhi recruitment.

While we showed that Rhi binding suppresses cluster transcript splicing, we still do not know how it performs this role. Most likely this is achieved by recruiting other factors. As an HP1 family protein, Rhi has a Chromo Shadow domain, which has been shown to be able to create a binding pocket by inter-domain dimerization (Huang et al., 2006). One Rhi binding partner is Cuff, but whether these two proteins directly interact is not clear (Pane et al., 2011). The molecular function of Cuff also needs to be determined.

How many other proteins are included in this Rhi complex? Deadlock may be one candidate. Recently, Hannon and colleagues have discovered that Deadlock is another piRNA pathway factor and forms Rhi-like foci in the nucleus (Czech et al., 2013). How Rhi complex couples cluster transcript with the potential precursor delivery protein, UAP56, is also an open question.

To drive de novo piRNA production from GFP construct, Rhi was recruited upstream of partial *vasa* promoter by LacI-LacO Protein DNA binding. To our surprise, Rhi appears to spread downstream of the GFP construct. What drives Rhi spreading? Is transcription required? Can Rhi still initiate piRNA production if we "tether" it to the intron or 3'UTR region? Our data suggest that the GFP antisense transcription is necessary for de novo piRNA generation. To achieve this, we inserted the construct that transcribes antisense GFP RNA into the same genomic locus as the sense GFP RNA producing vector, but on different chromosome alleles. To what extent this design scheme can mimic the natural dual-strand cluster transcription is still unknown.

The Ping-Pong model

Ping-Pong cycle is an attractive model to explain piRNA amplification and transposon posttranscriptional silencing. This model is largely supported by the observation of the 10 nt overlap between Aub bound antisense piRNAs, which often begin with uridine, and Ago3 associated sense piRNAs, that often harbor adenosine at their 10^{th} position (Brennecke et al., 2007). The mechanism that charges Aub with primary antisense piRNAs is still not clear. But once the Aub loading preference is established, Qin appears to be important to maintain this loading bias by ensuring Aub and Ago3 heterotypic Ping-Pong. This is reflected by the observation that we see more sense piRNAs are loaded into Aub in *qin* mutants. It would be interesting to swap protein domains between Aub and Ago3 to test the idea that the original loading preference is due to the protein structure.

Does the 1U bias from primary antisense piRNA drive secondary sense piRNA 10A preference by U:A paring? This is unlikely due to the general scheme of Argonaute-small RNA association, in which the first nucleotide is buried into the Mid domain to lock the small RNA into the protein, and barely contributes to target pairing (Nakanishi et al., 2012; Elkayam et al., 2012; Wee et al., 2012). Therefore, these nucleotide biases may also be generated by PIWI protein loading preferences.

Although we found that Qin is the key protein to ensure Aub Ping-Pongs with Ago3, it is still obscure how Qin promotes this heterotypic Ping-Pong. A simple hypothesis is that Qin uses its Tudor domains as docking sites to bridge these two PIWI proteins together. Indeed, we observed less Ago3 associated with Aub when *qin* is mutated. However, the bridge model is unlikely to be true,

because a truncated Qin protein still has all five Tudor domains, but lacks the N-terminal E3 ligase domain, fails to rescue *qin* mutation phenotype. Therefore, the E3 ligase domain of Qin is essential for its function, but its molecular function is still unknown.

Why does Aub:Ago3 heterotypic Ping-Pong generate more antisense piRNAs than sense? This may due to one protein being a more efficient "slicer" than the other. Since piRNAs need to be loaded into PIWI proteins to be stabilized, an alternative explanation would be that Aub is more abundant at the steady state level than Ago3, therefore more antisense piRNAs are preserved in the total piRNA population.

Transposon silencing

piRNAs silence transposons in the animal germline. How this small RNA pathway achieves transposon suppression is still not clear. There are more than 150 transposon families that reside in the fly genome (Kaminker et al., 2002). It appears they are silenced by different schemes: transcriptional or posttranscriptional (Ping-Pong cleavage). Are there other strategies that the piRNA pathway employs to put these jumping element under custody? Do piRNAs inhibit transposon mRNA translation? While these questions need to be addressed, we also need to rethink the readout for monitoring transposon activity. Based on the assumption that more transposase mRNA accumulation would lead to higher transposon jumping frequency, the transposon activity has been mainly gauged by measuring the transposase mRNA steady state level. However, this assumption is flawed. So far, there is no report to measure how much percent of the transposase mRNAs can be translated. Nor do we know

whether every transposase protein would contribute to the final transposition. Therefore, to decode how piRNAs silence transposons, it would be important to accurately measure the transposon jumping events in future studies.

"Bigger" picture

The animal germ cell employs the piRNA pathway to suppress transposon activity. What is the arsenal for somatic tissues to fight these "selfish" elements? In *Drosophila*, the somatic cells produce a KH-type RNA binding protein PSI, which can lead to mRNA alternative splicing to produce an inactive version of transposase for *P-element*, a DNA element that is silenced by the piRNA pathway in germline (Siebel et al., 1994; Siebel et al., 1995). However, little is known about how and even if other elements are silenced. It is hard to imagine that transposons in soma are massively active, since this would result in DNA breaks that would harm somatic cells in such ways as hampering the regular cell cycle (Goodarzi and Jeggo, 2013). Therefore, there may be a different mechanism to control transposons in somatic tissues. If this is true, why did animals evolve different schemes to cope with the same task? Is one strategy better than the other?

Transposons make up a large portion of genomes in animals and plants. During this long co-evolving process, the host genome may have benefited from these "selfish parasites". Are transposons only useful in this long-term view? For a single cell, the jumping events created by active transposons are most likely harmful. However, we certainly do not know whether all transposase mRNAs being transcribed or transposase proteins being translated can only contribute to transposon jumping. I definitely would not be surprised if the host cell has

already tamed or “borrowed” these transposon products for its own purposes. Would you?

BIBLIOGRAPHY

Abdu, U., Brodsky, M., and Schupbach, T. (2002). Activation of a meiotic checkpoint during Drosophila oogenesis regulates the translation of Gurken through Chk2/Mnk. Curr Biol *12*, 1645-1651.

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

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., Klenov, M. S., Vagin, V. V., Bantignies, F., Cavalli, G., and Gvozdev, V. A. (2004). Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. Mol Cell Biol *24*, 6742-6750.

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., 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.

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.

Armour, C. D., Castle, J. C., Chen, R., Babak, T., Loerch, P., Jackson, S., Shah, J. K., Dey, J., Rohl, C. A., Johnson, J. M., and Raymond, C. K. (2009). Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nat Methods *6*, 647-649.

Balakireva, M. D., Shevelyov, Y., Nurminsky, D. I., Livak, K. J., and Gvozdev, V. A. (1992). Structural organization and diversification of Y-linked sequences comprising Su(Ste) genes in Drosophila melanogaster. Nucleic Acids Res *20*, 3731-3736.

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

Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., Badge, R. M., and Moran, J. V. (2010). LINE-1 retrotransposition activity in human genomes. Cell *141*, 1159-1170.

Bennetzen, J. L. (2000). Transposable element contributions to plant gene and genome evolution. Plant Mol Biol *42*, 251-269.

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.

Burns, K. H., and Boeke, J. D. (2012). Human transposon tectonics. Cell *149*, 740-752.

Cecere, G., Zheng, G. X., Mansisidor, A. R., Klymko, K. E., and Grishok, A. (2012). Promoters recognized by forkhead proteins exist for individual 21U-RNAs. Mol Cell *47*, 734-745.

Chabot, B., Elela, S. A., and Zhuo, D. (2008). Comment on "When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis". Bioessays *30*, 1256; author reply 1257-1256; author reply 1258.

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

Chen, Y., Pane, A., and Schüpbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. Current biology : CB *17*, 637-642.

Christie, M., Croft, L. J., and Carroll, B. J. (2011). Intron splicing suppresses RNA silencing in Arabidopsis. Plant J *68*, 159-167.

Cocquet, J., Chong, A., Zhang, G., and Veitia, R. A. (2006). Reverse transcriptase template switching and false alternative transcripts. Genomics *88*, 127-131.

Cook, H. A., Koppetsch, B. S., Wu, J., and Theurkauf, W. E. (2004). The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. Cell *116*, 817-829.

Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

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

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. Molecular cell *50*, 749-761.

Danzer, J. R., and Wallrath, L. L. (2004). Mechanisms of HP1-mediated gene silencing in Drosophila. Development *131*, 3571-3580.

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.

Demerec, M. (1926). Miniature-alpha-A Second Frequently Mutating Character in Drosophila Virilis. Proc Natl Acad Sci U S A *12*, 687-690.

Demerec, M. (1927). Magenta-Alpha-A Third Frequently Mutating Character in Drosophila Virilis. Proc Natl Acad Sci U S A *13*, 249-253.

Desset, S., Buchon, N., Meignin, C., Coiffet, M., and Vaury, C. (2008). In Drosophila melanogaster the COM locus directs the somatic silencing of two

retrotransposons through both Piwi-dependent and -independent pathways. PLoS One *3*, e1526.

Dumesic, P. A., Natarajan, P., Chen, C., Drinnenberg, I. A., Schiller, B. J., Thompson, J., Moresco, J. J., Yates, J. R., Bartel, D. P., and Madhani, H. D. (2013). Stalled spliceosomes are a signal for RNAi-mediated genome defense. Cell *152*, 957-968.

Eberl, D. F., Lorenz, L. J., Melnick, M. B., Sood, V., Lasko, P., and Perrimon, N. (1997). A new enhancer of position-effect variegation in Drosophila melanogaster encodes a putative RNA helicase that binds chromosomes and is regulated by the cell cycle. Genetics *146*, 951-963.

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.

Finnegan, D. J. (2012). Retrotransposons. Current biology : CB *22*, R432-R437.

Fleckner, J., Zhang, M., Valcarcel, J., and Green, M. R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev *11*, 1864-1872.

Ghabrial, A., and Schupbach, T. (1999). Activation of a meiotic checkpoint regulates translation of Gurken during Drosophila oogenesis. Nat Cell Biol *1*, 354-357.

Ghildiyal, M., and Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.

Gillespie, D. E., and Berg, C. A. (1995). Homeless is required for RNA localization in Drosophila oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. Genes Dev *9*, 2495-2508.

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.

Gonsalvez, G. B., Rajendra, T. K., Tian, L., and Matera, A. G. (2006). The Sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. Curr Biol *16*, 1077-1089.

Goodarzi, A. A., and Jeggo, P. A. (2013). The repair and signaling responses to DNA double-strand breaks. Adv Genet *82*, 1-45.

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.

Gu, W., Lee, H.-C., Chaves, D., Youngman, E. M., Pazour, G. J., Conte, D., and Mello, C. C. (2012). CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C. elegans piRNA Precursors. Cell *151*, 1488-1500.

Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. 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.

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

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

Hardy, R. W., Tokuyasu, K. T., and Lindsley, D. L. (1981). Analysis of spermatogenesis in Drosophila melanogaster bearing deletions for Y-chromosome fertility genes. Chromosoma *83*, 593-617.

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.

Hawkins, T. L., O'Connor-Morin, T., Roy, A., and Santillan, C. (1994). DNA purification and isolation using a solid-phase. Nucleic Acids Res *22*, 4543-4544.

Herr, A. J., Molnar, A., Jones, A., and Baulcombe, D. C. (2006). Defective RNA processing enhances RNA silencing and influences flowering of Arabidopsis. Proc Natl Acad Sci U S A *103*, 14994-15001.

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. Current biology : CB *17*, 1265-1272.

Hosokawa, M., Shoji, M., Kitamura, K., Tanaka, T., Noce, T., Chuma, S., and Nakatsuji, N. (2007). Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: domain composition, intracellular localization, and function in male germ cells in mice. Dev Biol *301*, 38-52.

Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H., Hannon, G. J., Draper, B. W., and Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

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

Huang, R., Jaritz, M., Guenzl, P., Vlatkovic, I., Sommer, A., Tamir, I. M., Marks, H., Klampfl, T., Kralovics, R., Stunnenberg, H. G., Barlow, D. P., and Pauler, F. M. (2011b). An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. PLoS One *6*, e27288.

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

Huang, Y., Myers, M. P., and Xu, R. M. (2006). Crystal structure of the HP1-EMSY complex reveals an unusual mode of HP1 binding. Structure *14*, 703-712.

Iida, T., Nakayama, J., and Moazed, D. (2008). siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. Mol Cell *31*, 178-189.

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 (New York, N.Y.)

Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic expression of germline genes drives malignant brain tumor growth in Drosophila. Science *330*, 1824-1827.

Jin, Z., Flynt, A. S., and Lai, E. C. (2013). Drosophila piwi mutants exhibit germline stem cell tumors that are sustained by elevated Dpp signaling. Curr Biol *23*, 1442-1448.

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.

Kawaoka, S., Arai, Y., Kadota, K., Suzuki, Y., Hara, K., Sugano, S., Shimizu, K., Tomari, Y., Shimada, T., and Katsuma, S. (2011a). Zygotic

amplification of secondary piRNAs during silkworm embryogenesis. RNA *17*, 1401-1407.

Kawaoka, S., Hara, K., Shoji, K., Kobayashi, M., Shimada, T., Sugano, S., Tomari, Y., Suzuki, Y., and Katsuma, S. (2013). The comprehensive epigenome map of piRNA clusters. Nucleic Acids Res *41*, 1581-1590.

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. (2011b). 3′ end formation of PIWI-interacting RNAs in vitro. Mol Cell *43*, 1015-1022.

Khurana, J. S., and Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. J Cell Biol *191*, 9.

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.

Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517-522.

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

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., and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Development *135*, 3-9.

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., 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.

Lachke, S. A., Alkuraya, F. S., Kneeland, S. C., Ohn, T., Aboukhalil, A., Howell, G. R., Saadi, I., Cavallesco, R., Yue, Y., Tsai, A. C.-H., Nair, K. S., Cosma, M. I., Smith, R. S., Hodges, E., Alfadhli, S. M., Al-Hajeri, A., Shamseldin, H. E., Behbehani, A., Hannon, G. J., Bulyk, M. L., Drack, A. V., Anderson, P. J., John, S. W. M., and Maas, R. L. (2011). Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. Science *331*, 1571-1576.

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

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., 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 & development

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.

Levin, J. Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D. A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods *7*, 709-715.

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. W., 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, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

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.

Li, Y., Danzer, J. R., Alvarez, P., Belmont, A. S., and Wallrath, L. L. (2003). Effects of tethering HP1 to euchromatic regions of the Drosophila genome. Development *130*, 1817-1824.

Lim, A. K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A *104*, 6714-6719.

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.

Lis, J. T. (1980). Fractionation of DNA fragments by polyethylene glycol induced precipitation. Methods Enzymol *65*, 347-353.

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

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

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

Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A., and Eisen, M. B. (2011). Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol *9*, e1000590.

Luteijn, M. J., van Bergeijk, P., Kaaij, L. J., Almeida, M. V., Roovers, E. F., Berezikov, E., and Ketting, R. F. (2012). Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. EMBO J *31*, 3422-3430.

Mader, R. M., Schmidt, W. M., Sedivy, R., Rizovski, B., Braun, J., Kalipciyan, M., Exner, M., Steger, G. G., and Mueller, M. W. (2001). Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA. J Lab Clin Med *137*, 422-428.

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.

Mathioudakis, N., Palencia, A., Kadlec, J., Round, A., Tripsianes, K., Sattler, M., Pillai, R. S., and Cusack, S. (2012). The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. RNA *18*, 2056-2072.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

Meignin, C., and Davis, I. (2008). UAP56 RNA helicase is required for axis specification and cytoplasmic mRNA localization in Drosophila. Dev Biol *315*, 89-98.

Metzker, M. L. (2010). Sequencing technologies - the next generation. Nat Rev Genet *11*, 31-46.

Mevel-Ninio, M., Pelisson, A., Kinder, J., Campos, A. R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615-1624.

Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N. V., Czech, B., Gordon, A., Hannon, G. J., and Aravin, A. A. (2012). Production of artificial piRNAs in flies and mice. RNA (New York, N.Y.) *18*, 42-52.

Nakanishi, K., Weinberg, D. E., Bartel, D. P., and Patel, D. J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368-374.

Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., Holderbaum, L., Yang-Zhou, D., Shim, H.-S., Tao, R., Handler, D., Karpowicz, P., Binari, R., Booker, M., Brennecke, J., Perkins, L. A., Hannon, G. J., and Perrimon, N. (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nature methods *8*, 405-407.

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

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.

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. The EMBO journal *29*, 3301-3317.

Ozsolak, F., and Milos, P. M. (2011a). RNA sequencing: advances, challenges and opportunities. Nat Rev Genet *12*, 87-98.

Ozsolak, F., and Milos, P. M. (2011b). Single-molecule direct RNA sequencing without cDNA synthesis. Wiley Interdiscip Rev RNA *2*, 565-570.

Ozsolak, F., Platt, A. R., Jones, D. R., Reifenberger, J. G., Sass, L. E., McInerney, P., Thompson, J. F., Bowers, J., Jarosz, M., and Milos, P. M. (2009). Direct RNA sequencing. Nature *461*, 814-818.

Pan, J., Goodheart, M., Chuma, S., Nakatsuji, N., Page, D. C., and Wang, P. J. (2005). RNF17, a component of the mammalian germ cell nuage, is essential for spermiogenesis. Development *132*, 4029-4039.

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.

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. The EMBO journal *30*, 4601-4615.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res *37*, e123.

Parks, A. L., Cook, K. R., Belvin, M., Dompe, N. A., Fawcett, R., Huppert, K., Tan, L. R., Winter, C. G., Bogart, K. P., Deal, J. E., Deal-Herr, M. E., Grant, D., Marcinko, M., Miyazaki, W. Y., Robertson, S., Shaw, K. J., Tabios, M., Vysotskaia, V., Zhao, L., Andrade, R. S., Edgar, K. A., Howie, E., Killpack, K., Milash, B., Norton, A., Thao, D., Whittaker, K., Winner, M. A., Friedman, L., Margolis, J., Singer, M. A., Kopczynski, C., Curtis, D., Kaufman, T. C., Plowman, G. D., Duyk, G., and Francis-Lang, H. L. (2004). Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. Nat Genet *36*, 288-292.

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

Perrat, P. N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., and Waddell, S. (2013). Transposition-driven genomic heterogeneity in the Drosophila brain. Science *340*, 91-95.

Preall, J. B., Czech, B., Guzzardo, P. M., Muerdter, F., and Hannon, G. J. (2012). shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA *18*, 1446-1457.

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

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.

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. Current biology : CB *19*, 2066-2076.

Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A. S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol *135*, 1685-1700.

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 & development

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., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H., and Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev *24*, 2493-2498.

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

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.

Sarot, E., Payen-Groschene, G., Bucheton, A., and Pelisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. Genetics *166*, 1313-1321.

Schneider, T. D., and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res *18*, 6097-6100.

Schupbach, T., and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. Genetics *121*, 101-117.

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.

Shirayama, M., Seth, M., Lee, H. C., Gu, W., Ishidate, T., Conte, D. J., and Mello, C. C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65-77.

Siebel, C. W., Admon, A., and Rio, D. C. (1995). Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. Genes Dev *9*, 269-283.

Siebel, C. W., Kanaar, R., and Rio, D. C. (1994). Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. Genes Dev *8*, 1713-1725.

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.

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

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. Nature reviews. Genetics *8*, 272-285.

Smith, C. D., Shu, S., Mungall, C. J., and Karpen, G. H. (2007). The Release 5.1 annotation of Drosophila melanogaster heterochromatin. Science *316*, 1586-1591.

Snee, M. J., and Macdonald, P. M. (2004). Live imaging of nuage and polar granules: evidence against a precursor-product relationship and a novel role for Oskar in stabilization of polar granule components. J Cell Sci *117*, 2109-2120.

Sultan, M., Dokel, S., Amstislavskiy, V., Wuttig, D., Sultmann, H., Lehrach, H., and Yaspo, M. L. (2012). A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. Biochem Biophys Res Commun *422*, 643-646.

Thibault, S. T., Singer, M. A., Miyazaki, W. Y., Milash, B., Dompe, N. A., Singh, C. M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H. L., Ryner, L., Cheung, L. M., Chong, A., Erickson, C., Fisher, W. W., Greer, K., Hartouni, S. R., Howie, E., Jakkula, L., Joo, D., Killpack, K., Laufer, A., Mazzotta, J., Smith, R. D., Stevens, L. M., Stuber, C., Tan, L. R., Ventura, R., Woo, A., Zakrajsek, I., Zhao, L., Chen, F., Swimmer, C., Kopczynski, C., Duyk, G., Winberg, M. L., and Margolis, J. (2004). A complementary transposon tool kit for Drosophila melanogaster using P and piggyBac. Nat Genet *36*, 283-287.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

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., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G. J., and Aravin, A. A. (2009). Proteomic analysis of murine Piwi proteins reveals a role for arginine

methylation in specifying interaction with Tudor family members. Genes Dev *23*, 1749-1762.

Venken, K. J., Carlson, J. W., Schulze, K. L., Pan, H., He, Y., Spokony, R., Wan, K. H., Koriabine, M., de Jong, P. J., White, K. P., Bellen, H. J., and Hoskins, R. A. (2009). Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nat Methods *6*, 431-434.

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

Wang, J., Saxe, J. P., Tanaka, T., Chuma, S., and Lin, H. (2009a). Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. Curr Biol *19*, 640-644.

Wang, L., Si, Y., Dedow, L. K., Shao, Y., Liu, P., and Brutnell, T. P. (2011). A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. PLoS One *6*, e26426.

Wang, S. H., and Elgin, S. C. R. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proceedings of the National Academy of Sciences of the United States of America *108*, 21164-21169.

Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews. Genetics *10*, 57-63.

Warf, M. B., Shepherd, B. A., Johnson, W. E., and Bass, B. L. (2012). Effects of ADARs on small RNA processing pathways in C. elegans. Genome Res *22*, 1488-1498.

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.

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

Wypijewski, K., Hornyik, C., Shaw, J. A., Stephens, J., Goraczniak, R., Gunderson, S. I., and Lacomme, C. (2009). Ectopic 5' splice sites inhibit gene expression by engaging RNA surveillance and silencing pathways in plants. Plant Physiol *151*, 955-965.

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 *47*, 970-979.

Yin, H., and Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. Nature *450*, 304-308.

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

Zhang, C. J., Zhou, J. X., Liu, J., Ma, Z. Y., Zhang, S. W., Dou, K., Huang, H. W., Cai, T., Liu, R., Zhu, J. K., and He, X. J. (2013). The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in Arabidopsis. EMBO J *32*, 1128-1140.

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, 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., 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. Molecular cell *44*, 572-584.