2011-04-05

Molecular Mechanisms of piRNA Biogenesis and Function in Drosophila: A Dissertation

Chengjian Li
University of Massachusetts Medical School

MOLECULAR MECHANISMS OF PIRNA BIOGENESIS AND FUNCTION IN *DROSOPHILA*

A Dissertation Presented

By

CHENGJIAN LI

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

APRIL 5, 2011

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

MOLECULAR MECHANISMS OF PIRNA BIOGENESIS AND FUNCTION IN *DROSOPHILA*

A Dissertation Presented
By

CHENGJIAN LI

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

Phillip D. Zamore, Ph.D., Thesis Advisor

Gregory J. Hannon, Ph.D., Member of Committee

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

William E. Theurkauf, Ph.D., Member of Committee

Scot A. Wolfe, 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

Melissa J. Moore, 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

Biochemistry and Molecular Pharmacology
April 5, 2011

COPYRIGHT INFORMATION

DEDICATION

To my parents, my brothers, my wife, and my son

with love

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor, Phillip Zamore, for giving me the opportunity to work in his lab, for his mentoring and confidence in my research, and for his support to my family. I feel extremely lucky to have worked with Phil. His enthusiasm and attitudes for science as well as life exert invaluable influence on me. I deeply appreciate every minute he has spent with me. I would also like to thank all the alumni (Alla, Ben, Brad, Chris, Dianne, Hervè, Irena, Klaus, Megha, Mike, Shengmei, Tingting, Vasia, Yonatan, and Yuki), and current members (Alicia, Bo, Cindy, Desi, Elif, Fabian, Gwen, Jen, Jogi, Keith, Ryuya, Stefan, Tiffanie, Tim, Tracey, Wee, Xin, and Zhao) in Phil's lab for their friendship, for their efforts in creating such a great scientific environment, for the joyful time in the past few years.

I want to express my gratitude to Bill Theurkauf and his Lab (particularly Nadine and Carla) as well as Zhiping Weng and her Lab (particularly Jia, Soo, Jie and Jui-Hung) for the wonderful collaborations.

I would also like to acknowledge the members of my committee, Melissa Moore, Sean Ryder, Bill Theurkauf, and Scot Wolfe, for their insightful comments and suggestions on my research as well as on my career. I would also like to thank Gregory Hannon for being the external member of my committee. I want to extend my appreciation to people in UMASS Medical School and University of Pittsburgh for their support, friendship, and advice, especially, Marc Freeman, Guangping Gao, Bill Kobertz, Oliver Rando, and Michael Tsang.

For all of their support and help, I would like to thank my friends, especially, Daorong Guo, Kai Li, Randong Shan, Guoming Sun, and Lei Wang.

Most importantly, I would like to thank my parents and my brother’s families for their tremendous love and support despite the geographic distance. Most of all, I want to thank my wife, Xiaolan, for her constant love and support, as well as bringing our son, Alan, to the world, who really helped me learn how to be a good father, and at the same time deeply understand the love from my parents.

ABSTRACT

In the *Drosophila* germ line, PIWI-interacting RNAs (piRNAs) ensure genomic stability by silencing endogenous selfish genetic elements such as retrotransposons and repetitive sequences.

We examined the genetic requirements for the biogenesis and function of piRNAs in both female and male germ line. We found that piRNAs function through the PIWI, rather than the AGO, family Argonaute proteins, and the production of piRNAs requires neither microRNA (miRNA) nor small interfering RNA (siRNA) pathway machinery. These findings allowed the discovery of the third conserved small RNA silencing pathway, which is distinct from both the miRNA and RNAi pathways in its mechanisms of biogenesis and function.

We also found piRNAs in flies are modified. We determined that the chemical structure of the 3´-terminal modification is a 2´-*O*-methyl group, and also demonstrated that the same modification occurs on the 3´ termini of siRNAs in flies. Furthermore, we identified the RNA methyltransferase *Drosophila* Hen1, which catalyzes 2´-*O*-methylation on both siRNAs and piRNAs. Our data suggest that 2´-*O*-methylation by Hen1 is the final step of biogenesis of both the siRNA pathway and piRNA pathway.

Studies from the Hannon Lab and the Siomi Lab suggest a ping-pong amplification loop for piRNA biogenesis and function in the *Drosophila* germline. In this model, an antisense piRNA, bound to Aubergine or Piwi, triggers production of a sense piRNA bound to the PIWI protein Argonaute3 (Ago3). In turn, the new piRNA is envisioned to produce a second antisense piRNA. We isolated the loss-of-function mutations in *ago3*, allowing a direct genetic test of

this model. We found that Ago3 acts to amplify piRNA pools and to enforce on them an antisense bias, increasing the number of piRNAs that can act to silence transposons. Moreover, we also discovered a second Ago3-independent piRNA pathway in somatic ovarian follicle cells, suggesting a role for piRNAs beyond the germ line.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

The core of small RNA pathways

The spatiotemporal regulation of gene expression shapes all life on this planet. The discovery of RNA interference (RNAi) unveiled a hidden layer of regulation of gene expression by a variety of small RNA pathways, which are capable of regulating gene expression, defending against viral invasions, and repressing selfish genetic elements throughout evolution (Ghildiyal and Zamore, 2009; Malone and Hannon, 2009). Despite their diverse mechanisms of biogenesis and function, all these small RNA pathways share a conserved core, the effector complex, which contains an Argonaute protein and a small RNA. The small RNA confers sequence specificity via base pairing to its RNA target. The Argonaute protein executes silencing through either directly cleaving the RNA target or recruiting accessory factors that induce RNA destabilization, translational repression or heterochromatin formation.

Argonaute proteins

Eukaryotic Argonaute protein consists of three conserved domains: PAZ, Mid, and PIWI. The PAZ domain binds the 3´ end of the single-stranded RNA (Song et al., 2003; Ma et al., 2004). The Mid domain folds into a binding pocket that anchors the 5´ phosphate of the small RNA. Recent structural studies suggest that the Mid domain also contributes to the recognition of the 5´ nucleotide (Frank et al., 2010; Wang et al., 2008; Ma et al., 2005; Parker et al., 2005; Wang et al., 2008; Wang et al., 2009). The PIWI domain, which is similar in structure to RNase H, usually contains an apparent catalytic triad (Asp–Asp–His) required for endonucleolytic activity (also known as slicer activity). The slicer-competent Argonaute proteins cleave their target RNAs 5′ to the base that pairs

with the tenth nucleotide of the small RNA guide to generate products with 3′ hydroxyl and 5′ phosphate termini (Ma et al., 2005; Song et al., 2004; Parker et al., 2004; Yuan et al., 2005; Schwarz et al., 2004; Martinez and Tuschl, 2004).

Based on the similarity in amino acid sequence, animal Argonaute proteins can be classified into two major families: AGO and PIWI (Figure 1.1; Carmell et al., 2002; Parker and Barford, 2006). Five Argonaute proteins are expressed in *Drosophila*. Argonaute1 (Ago1) and Argonaute2 (Ago2) belong to the AGO family. Piwi, Aubergine (Aub), and Argonaute3 (Ago3) constitute the PIWI family. These two families have distinct expression patterns. The members of AGO family are ubiquitously expressed throughout development, whereas the expression of PIWI proteins is restricted to the gonads (Williams and Rubin, 2002). In ovaries, Piwi is localized in the nuclei of both germ cells and gonadal somatic cells (Cox et al., 2000; Saito et al., 2006; Brennecke et al., 2007). Unlike Piwi, Aub and Ago3 can only be detected in the germ cells, where both of them concentrate in the perinuclear nuage (Nishida et al., 2007; Gunawardane et al., 2007; Brennecke et al., 2007; Li et al., 2009).

Drosophila Argonaute proteins bind distinct classes of small RNAs and are segregated into three functional pathways. In the miRNA pathway, Ago1 associates with endogenous miRNAs to conduct translational repression or mRNA destabilization. In the siRNA pathway, Ago2 binds endogenous siRNAs to silence transposons and mRNAs. The PIWI proteins and their bound PIWI-interacting RNAs (piRNAs) comprise the piRNA pathway in gonads that safeguards the genome of germ cells against the deleterious influence of selfish genetic elements (Ghildiyal and Zamore, 2009; Malone and Hannon, 2009; Senti and Brennecke, 2010).

Figure 1.1.

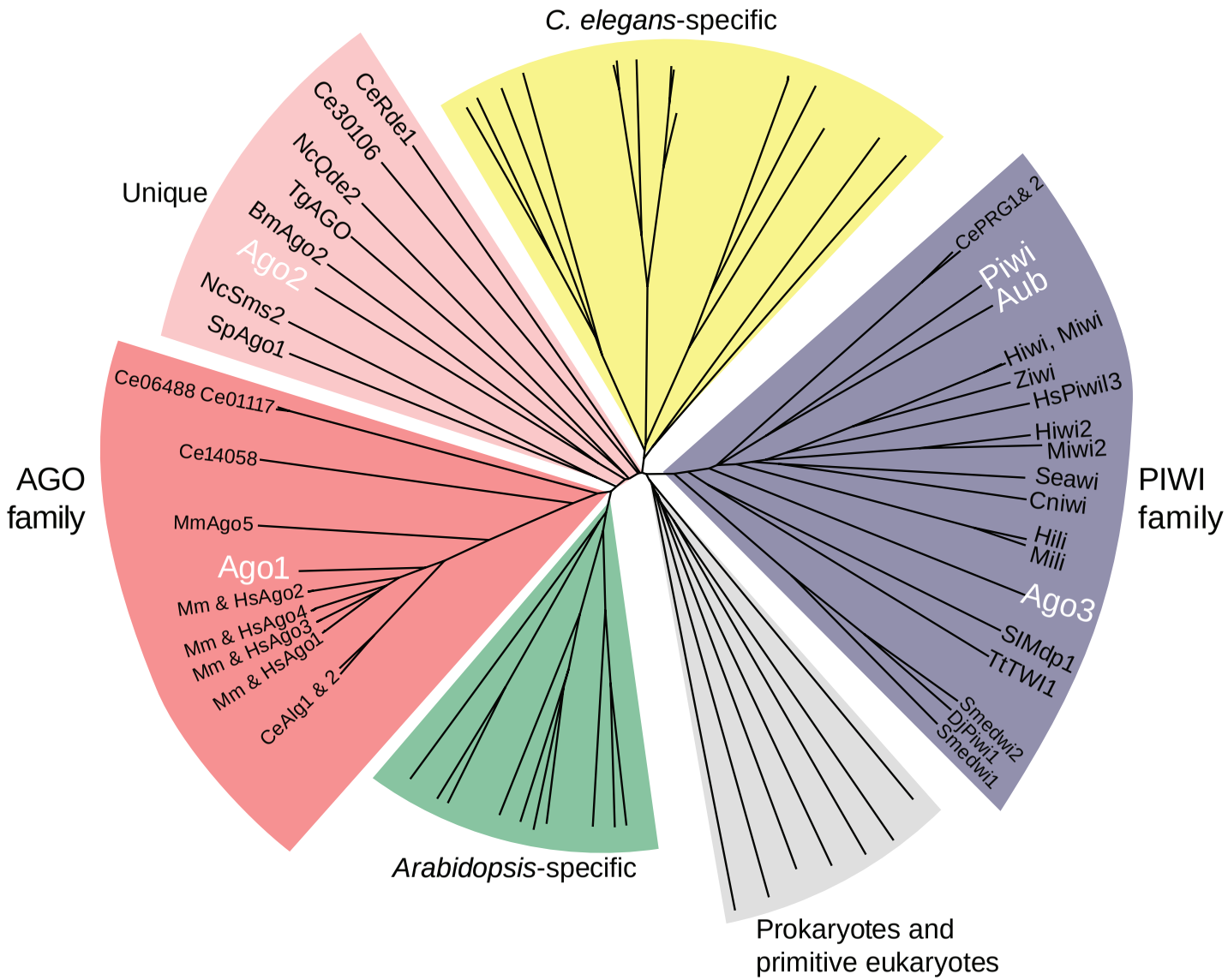

C. elegans-specific
Unique
CeRde1
Ce30106
NcQde2
TgAGO
BmAgo2
Ago2
NcSms2
SpAgo1
AGO
family
Ce06488 Ce01117
Ce14058
MmAgo5
Ago1
Mm & HsAgo2
Mm & HsAgo4
Mm & HsAgo3
Mm & HsAgo1
CeAlg1 & 2
Arabidopsis-specific
Prokaryotes and
primitive eukaryotes
CePRG1& 2
Piwi
Aub
Hiwi, Miwi
Ziwi
HsPiwil3
Hiwi2
Miwi2
Seawi
Cniwi
Hili
Mili
Ago3
SlMdp1
TtTWI1
Smedwi2
DjPiwi1
Smedwi1
PIWI
family

Figure 1.1 Phylogenetic analysis of Argonaute proteins.

Adapted from Parker and Barford, 2006.

The discovery of piRNAs

In male flies, small RNAs derived from the *Suppressor of Stellate* (*Su(Ste)*) locus on the Y chromosome silence the X-linked *Stellate* locus during spermatogenesis (Aravin et al., 2001). In *Su(Ste)* mutants, small RNAs targeting *Stellate* are lost causing *Stellate* protein crystals to form in primary spermatocytes and reduced fertility (Livak, 1990; Livak, 1984; Pimpinelli et al., 1985). Subsequent small RNA profiling studies in *Drosophila* reveal that *Su(Ste)* small RNAs only represent a small subset of a class of small RNAs with a size range, 24-26 nt, larger than miRNAs. Moreover, these small RNAs exclusively map to transposons and repetitive elements and thus were termed repeat-associated small interfering RNAs (rasiRNAs) (Aravin et al., 2003). *Drosophila* rasiRNAs are enriched in both male and female gonads and ensure genomic stability by silencing endogenous selfish genetic elements such as transposons and repetitive elements (Aravin et al., 2003; Vagin et al., 2006; Saito et al., 2006). Remarkably, they specifically interact with the PIWI family of Argonaute proteins (Vagin et al., 2006; Saito et al., 2006; Brennecke et al., 2007; Gunawardane et al., 2007; Yin and Lin, 2007). Mammalian PIWI proteins also associate with a class of endogenous RNAs termed piRNAs (Lau et al., 2006; Grivna et al., 2006; Girard et al., 2006; Aravin et al., 2006). However, unlike rasiRNAs in flies, the majority of piRNAs in mammals arise from unannotated regions of the genome devoid of transposons. *Drosophila* rasiRNAs were renamed as piRNAs based on their PIWI protein interaction patterns (Brennecke et al., 2007; Gunawardane et al., 2007; Yin and Lin, 2007).

The biogenesis of piRNAs does not require Dicer

The biogenesis of both miRNAs and siRNAs requires the RNase III family of Dicer proteins, which specifically act on double-stranded RNA (dsRNA) precursors (Figure 1.2; Hutvagner et al., 2001; Bernstein et al., 2001; Lee et al., 2004). The cleavage of Dicer yields dsRNA intermediates—siRNA or miRNA/miRNA* duplexes—that are loaded into the AGO family of Argonaute proteins. However, the production of piRNAs does not require any Dicer in either flies or zebrafish (Vagin et al., 2006; Houwing et al., 2007). These observations imply that piRNAs may arise from single-stranded RNA precursors.

The 3′ terminal modification is the last step of the biogenesis of piRNAs

Unlike animal miRNAs, which bear 2′ and 3′-hydroxyl groups on their 3′ termini, piRNAs carry 3′ terminal modification, which was initially characterized by the resistance to periodate oxidization and beta-elimination (Vagin et al., 2006). In *Drosophila*, the 3′ termini of siRNAs are also modified (Pelisson et al., 2007). Subsequent studies identified the modification as 2′-*O*-methylation (Houwing et al., 2007; Ohara et al., 2007; Kirino and Mourelatos, 2007b; Horwich et al., 2007; Saito et al., 2007).

Interestingly, the same modification has been observed on all small RNAs in *Arabidopsis* (Li et al., 2005). In plants, small RNA methylation is catalyzed by a S-adenosylmethionine-dependent RNA methyltransferase, HUA ENHANCER 1 (HEN1), which contains two dsRNA-binding domains and a methyltransferase domain on the C-terminus (Li et al., 2005; Yang et al., 2006; Yu et al., 2005). Yet, animal HEN1 homologues lack dsRNA-binding domains, suggesting that dsRNAs are not their substrates (Park et al., 2002; Tkaczuk et al., 2006). Consistent with this view, recombinant *Drosophila* Hen1 only acts on the

Figure 1.2.

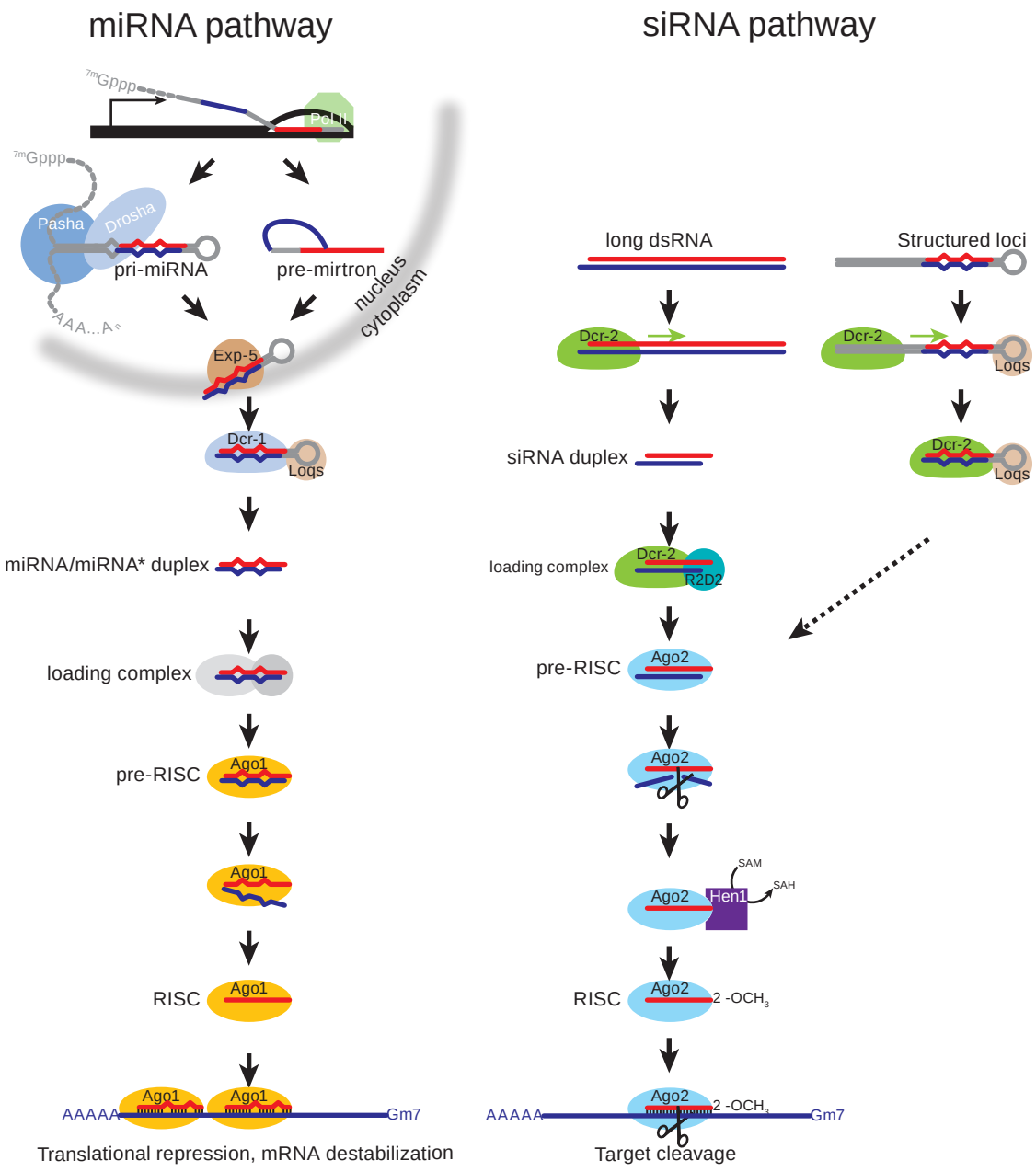

miRNA pathway
siRNA pathway
7mGppp
PolII
Pasha
Drosha
pri-miRNA
pre-mirtron
nucleus
cytoplasm
AAA...A
Exp-5
Dcr-1
Logs
miRNA/miRNA* duplex
loading complex
pre-RISC
Ago1
RISC
AAAAA
Gm7
Translational repression, mRNA destabilization
long dsRNA
Structured loci
Dcr-2
siRNA duplex
R2D2
Ago2
SAM
SAH
Hen1
2 -OCH3
Target cleavage

Figure 1.2. The biogenesis of miRNAs and siRNAs in *Drosophila*.

single-stranded RNAs (Horwich et al., 2007; Saito et al., 2007). Moreover, genetic studies indicate that Hen1 is required for piRNA modification in vivo in flies and zebrafish (Horwich et al., 2007; Kamminga et al., 2010). Although *Drosophila* Hen1 also methylates siRNAs, which are derived form siRNA duplexes, 2´-*O*-methylation specifically occurs on the guide strand after the release of the passenger strand (Horwich et al., 2007). These observations suggest that 2´-*O*-methylation by Hen1 is the last step in the biogenesis of siRNAs and piRNAs.

In flies lacking Hen1, *Su(Ste)* and *roo* piRNAs slightly decrease in both length and abundance (Horwich et al., 2007). Recent studies indicate that Hen1 prevents both uridylation and adenylation on the 3´ ends of siRNAs in flies (Ameres et al., 2010) and piRNAs in zebrafish (Kamminga et al., 2010). However, loss of Hen1 only leads to mild increase of transposons and *hen1* mutant animals are still fertile (Horwich et al., 2007; Saito et al., 2007; Kamminga et al., 2010). The function of 2´-*O*-methylation of piRNAs remains unclear.

The piRNA clusters are the source of piRNAs

Although the majority of piRNAs map to multiple sites in the genome, tracking piRNAs uniquely mapping in the genome allows the finding of the genomic origins of piRNAs, which show extraordinary clustering at discrete genomic loci (Brennecke et al., 2007). These genomic loci have been termed piRNA clusters, ranging from several to hundreds of kilobases (kb). The top 142 piRNA clusters give rise to more than 81% uniquely mapping piRNAs and 92% of total piRNAs, but occupy only 3.5% of the assembled fly genome (Brennecke et al., 2007). 95% of these clusters are present in pericentromeric and telomeric

heterochromatin regions (Brennecke et al., 2007). Pericentromeric piRNA clusters enrich defective fragments of 70-90% of annotated transposons, whereas telomeric piRNA clusters contain tandem repeats of *HeT-A*, *TART*, and *TAHRE* as well as the subtelomeric terminal associated sequences (TASs).

The piRNA clusters fall into two categories: dual-strand and uni-strand clusters. Dual-strand clusters produce uniquely mapping piRNAs from both genomic strands and comprise most piRNAs in flies. The largest dual-strand cluster is 42AB in the pericentromeric heterochromatin (cytological position 42AB) of chromosome 2R. The 42AB cluster covers 21% of uniquely mapping piRNAs and 30% of total piRNAs in the fly genome and appears to be active only in the germ cells (Brennecke et al., 2007; Malone et al., 2009). Uni-strand piRNA clusters produce piRNAs predominantly from one genomic strand. Two major uni-strand piRNA clusters were discovered on the pericentromeric heterochromatin of chromosome X. One turns out to be the *flamenco* locus (*flam*, also known as *COM*), which has been genetically identified as the suppressor of retrotransposons *gypsy*, *Idefix* and *ZAM* (Prud'homme et al., 1995; Pelisson et al., 1994). In fact, 87% of the *flam* cluster consists of nested transposon fragments including *gypsy*, *Idefix* and *ZAM*. These defective transposon fragments are arranged in the same direction, and *flam* piRNAs are entirely antisense (Brennecke et al., 2007). A single *P*-element insertion at the 5´ end of the *flam* cluster substantially eliminates all piRNAs across entire *flam*, suggesting that *flam* piRNAs are produced from a long single-stranded RNA precursor (Brennecke et al., 2007; Malone et al., 2009).

The transcriptional regulation of piRNA clusters is still a mystery. Interestingly recent studies suggest that Rhino is required for the production of

transcripts from the dual-strand piRNA clusters (Klattenhoff et al., 2009). Rhino is homologous to heterochromatin protein 1 (HP1) (Volpe et al., 2001). Chromosomal immunoprecipitations indicate that Rhino specifically binds on the 42AB but not the *flam* locus. Loss of Rhino eliminates the piRNAs derived from dual-strand cluster (Klattenhoff et al., 2009).

The secondary piRNAs are generated by ping-pong amplification

Each PIWI protein binds a distinct subset of piRNAs with a characteristic length distribution, strand bias, and nucleotide bias. The piRNAs bound to Aub or Piwi are largely antisense and typically begin with U, while the piRNAs bound to Ago3 are largely sense and typically bear an A at position 10 (Vagin et al., 2006; Saito et al., 2006; Brennecke et al., 2007; Gunawardane et al., 2007; Yin and Lin, 2007). The first ten nucleotides of piRNAs bound to Aub or Piwi are often complementary to the first ten nucleotides of piRNAs bound to Ago3 (Brennecke et al., 2007; Gunawardane et al., 2007). In addition, in vitro experiments indicate that all three *Drosophila* PIWI proteins retain their slicer activity to cut their target RNAs 5′ to the base that pairs with the tenth nucleotide of the small RNA guide (Gunawardane et al., 2007; Saito et al., 2006; Nishida et al., 2007). Taken together, these observations suggest an intriguing model, in which antisense piRNAs guide Aub or Piwi to cleave the transposon mRNAs and trigger the production of sense piRNAs bound to Ago3, and reciprocally, sense piRNAs guide Ago3 to cleave the piRNA cluster transcript and trigger the production of new antisense piRNAs (Figure 1.3). This so-called ping-pong model proposes a target-dependent amplification of antisense piRNAs, which heavily depends on Ago3. Consistent with this view, loss of Ago3 dramatically decreases antisense

Figure 1.3.

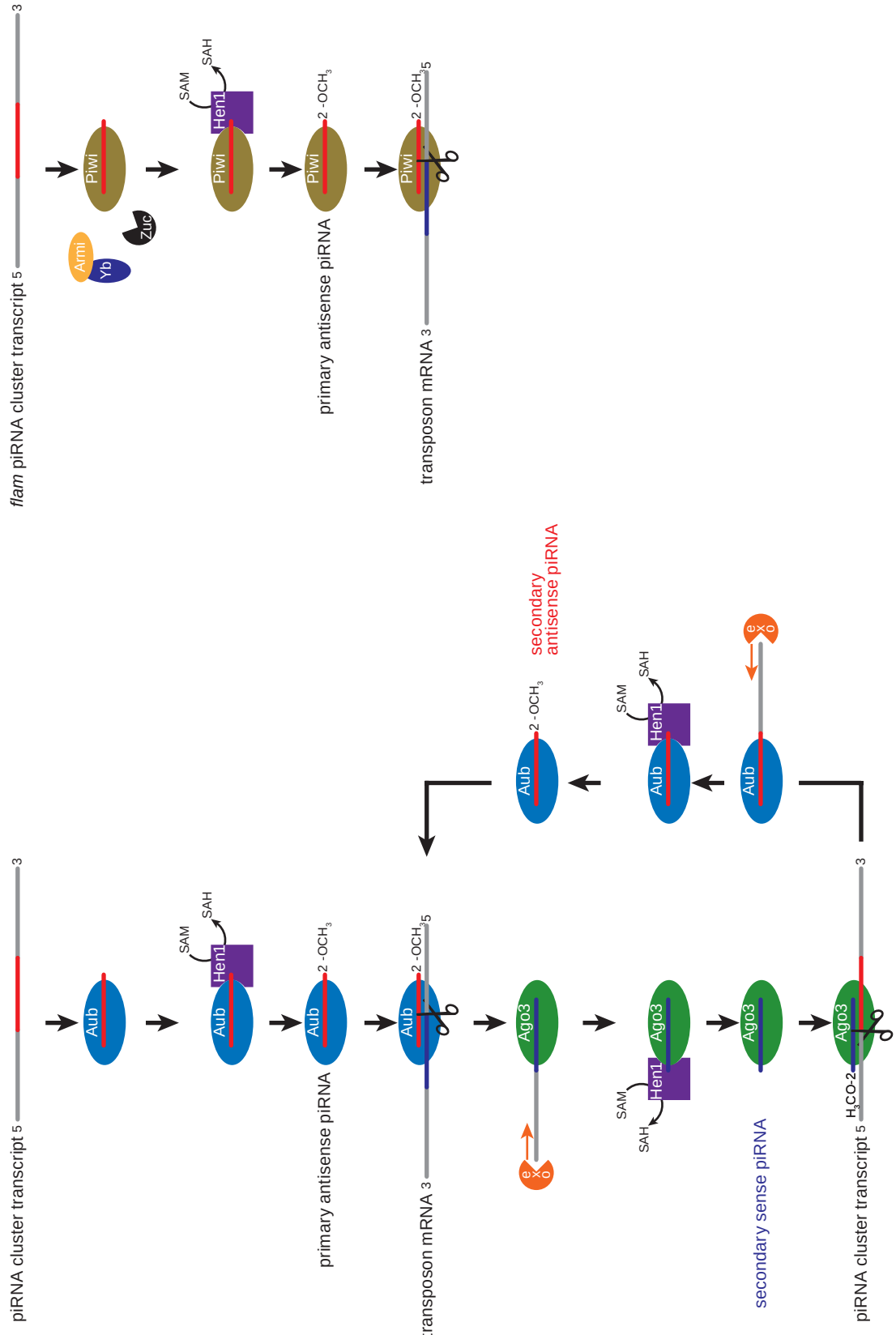

Somatic follicle cells
flam piRNA cluster transcript 5 3
Piwi
Armi
Yb
Zuc
SAM
SAH
Hen1
2 -OCH3
primary antisense piRNA
transposon mRNA 3
Germ cells
piRNA cluster transcript 5 3
Aub
Ago3
exo
secondary antisense piRNA
secondary sense piRNA
H3CO-2

Figure 1.3. The biogenesis of piRNAs in *Drosophila*.

piRNAs associated with Aub and flips the strand bias of group I and II piRNAs whose production essentially depends on the ping-pong amplification cycle (Li et al., 2009). The ping-pong amplification cycle appears to be a conserved mechanism for piRNA biogenesis and function as it has been reported in other animals including sponges, sea anemones, planarians, silkworms, zebrafish, frogs, and mice (Grimson et al., 2008; Palakodeti et al., 2008; Friedlander et al., 2009; Kawaoka et al., 2009; Houwing et al., 2007; Lau et al., 2009a; Aravin et al., 2007).

A hypothetical model for the biogenesis of primary piRNAs

The initiation of the feed forward ping-pong amplification requires Aub-bound or Piwi-bound antisense primary piRNAs as triggers, which are derived from the direct processing of the piRNA cluster transcripts. Currently, the biogenesis of primary piRNAs is still largely unknown. Previous studies reveal that somatic follicle cells harbor a ping-pong-independent piRNA pathway, which provides a simple paradigm to study the biogenesis of primary piRNAs (Figure 1.3; Li et al., 2009; Malone et al., 2009). The somatic piRNA pathway also presents in ovarian somatic sheet cells (OSSs) and ovarian somatic cells (OSCs) that derived from somatic gonadal cells (Saito et al., 2009; Lau et al., 2009b). In these cells, the uni-strand piRNA cluster, *flam*, appears to be the only source of *gypsy* and *ZAM* piRNAs. The accumulation of *gypsy* piRNAs depends on the presence of the *flam* cluster rather than active *gypsy* (Pelisson et al., 2007). *ZAM* piRNAs are only derived from the corresponding fragments located in *flam* cluster (Malone et al., 2009). The flam piRNAs are loaded into Piwi and their accumulation depends on Piwi rather than Aub and Ago3 (Brennecke et al.,

2007; Li et al., 2009; Malone et al., 2009). These observations indicate that somatic piRNA production follows a linear pathway: piRNA intermediates or piRNAs are directly loaded into Piwi after processing of the single-stranded transcripts from the *flam* cluster.

Primary piRNAs from the cluster do not show an obvious phasing pattern. Both Aub-bound and Piwi-bound piRNAs display a strong 5´U bias, which is unlikely inherited from Ago3 (Li et al., 2009). Many Argonaute proteins are able to recognize the identity of the 5´ nucleotides of their bound small RNAs (Mi et al., 2008; Ghildiyal et al., 2010; Frank et al., 2010). In vitro-translated Piwi preferentially binds small RNAs with 5´U (Yin and Lin, 2007). In addition, piRNAs associated with each PIWI protein show a distinct length distribution (Brennecke et al., 2007). In OSC cells the production of primary piRNAs appears to be independent of the catalytic activity of Piwi (Saito et al., 2009). Taken together, these observations suggest a random-processing model, in which the piRNA cluster transcripts are randomly cleaved to generate a repertoire of RNA fragments; Piwi selectively incorporates and stabilizes the RNAs that start with a 5´U (pre-piRNAs); in the following steps, the 3´ ends of the pre-piRNAs are trimmed by a putative nuclease, resulting in the characteristic length distribution of piRNAs.

In OSCs, nuclear localization of Piwi depends on the nuclear localization signal (NLS) on its N-terminus (Saito et al., 2009). Deletion of this NLS retains Piwi in the cytoplasm, but does not influence the loading of piRNAs into Piwi, suggesting that the loading of primary piRNAs occurs in the cytoplasm (Saito et al., 2009). Recent studies in flies, OSCs, and OSSs suggest that Armitage (Armi), Zucchini (Zuc), and Yb are required for primary piRNA production and

transposon silencing in somatic follicle cells (Olivieri et al., 2010; Saito et al., 2010; Haase et al., 2010; Qi et al., 2011). Yb is only expressed in somatic follicle cells and concentrates in cytoplasmic foci, termed Yb bodies, which also contain RNAs (Szakmary et al., 2009). Armi turns out to be a component of the Yb body and its localization to the Yb body completely depends on Yb. Interestingly, in the absence of Zuc, Piwi also accumulates in the Yb body instead of in the nucleus, but Piwi is still loaded with piRNAs. Consistently, immunoprecipitation experiments indicate that both Piwi and Yb interact with Armi (Saito et al., 2010). However, in the absence of Arimi or Yb, Piwi is devoid of piRNAs and is no longer localized to either the Yb body or the nucleus. These results suggest that the loading of piRNAs into Piwi occurs in the Yb body, and also requires both Armi and Yb, while Zuc is required for translocation of Piw-piRNA complex into the nucleus, presumably by acting on the piRNA or a component in loading complex.

Nuage is a cellular compartment for piRNA production

The nuage is a germ cell specific organelle that is remarkably conserved across the animal kingdom (Eddy, 1975; Saffman and Lasko, 1999). Under electron microscopy, it displays an amorphous and fibrous structure localized to the cytoplasmic face of the nuclear envelope. Nuage is often associated with mitochondria suggesting that signaling from mitochondria might regulate its assembly or function. In *Drosophila*, nuage can be observed in all germ cells from the onset of pole cell formation to determination of the oocyte. During the remaining stages of oogenesis the nuage is prominent in all nurse cells but not in the oocyte. In the late stages of oogenesis, the recruitment of particles that

contain nuage components to the posterior pole of oocyte forms germline granules (also known as polar granules), which determine the formation of primordial germ cells, pole cells in early embryos (Lim and Kai, 2007; Snee and Macdonald, 2004).

Nuage enriches almost all components of the germline piRNA pathway except Piwi and Rhino (Table 1.1). Accumulating evidence indicates that the assembly of piRNA components to the nuage appears to be a complex network (Lim and Kai, 2007; Lim et al., 2009; Patil and Kai, 2010; Findley et al., 2003; Klattenhoff et al., 2009; Li et al., 2009; Malone et al., 2009; Anne, 2010). The common concept is that mutations that disrupt the ping-pong amplification cycle usually correlate to the dislocalization of Aub and Ago3 from the nuage. This is even prominent for Rhino, a nuclear protein, which is required for the production of transcripts from the dual-strand piRNA clusters (Klattenhoff et al., 2009). In addition, Aub and Ago3 depend on each other to be localized in the nuage (Li et al., 2009; Malone et al., 2009). Taken together, these observations indicate that nuage is a cellular compartment for germline piRNA production and function.

Methylated arginines are the molecular linkers of the piRNA pathway

Many components of *Drosophila* piRNA pathway such as Krimper (Krimp), Spindle-E (Spn-E), Yb, Tejas, and Tudor (Tud) contain one or multiple Tudor domains (Siomi et al., 2010). The tudor domain has been shown to mediate protein-protein interaction via binding to methylated arginines on the partner (Cote and Richard, 2005). Eukaryotes contain three main forms of methylated arginines: monomethylated arginines (MMAs), asymmetric dimethylated arginines (aDMAs), and symmetric dimethylated arginines (sDMAs).

Table 1.1. Components of *Drosophila* piRNA pathway.

Name	Domain	Post-translational modification	Localization In germ cells	Localization In follicle cells	Protein-protein interaction
Ago3	PAZ, Mid, PIWI	sDMAs	nuage		Tudor, Hen1, Aub
Armi	Helicase		nuage	Yb body	Yb, Piwi, Armi
Aub	PAZ, Mid, PIWI	sDMAs	Nuage pole plasm		Tudor, Hen1, Ago3
cutoff	RAI1		Nuage		
dPRMT5	Methyltransferase				Vls
Hen1	Methyltransferase				Ago3, Aub, Piwi
Krimp	Tudor		nuage		
Maelstrom	HMG		Nuleus nuage		
Piwi	PAZ, Mid, PIWI	sDMAs Phosphorylation	Nucleus	Nucleus	Hen1, HP1
Rhi	Chromo box		nucleus		
Spn-E	Tudor, DExDc		nuage		
Squ	Nuclease?		Nuage		
Tej	Tudor		nuage		Vas, Aub, Spindle-E
Tud	Tudor		Nuage pole plasm		Ago3, Aub, Vls
Vasa	DEAD	sDMAs aDMAs Phosphorylation	nuage		
Vls			Nuage pole plasm		dPRMT5, Tudor
Yb	Tudor			Yb body	
Zuc	Nuclease?		Nuage	Cytoplasm	

The methylation of arginines is catalyzed post-translationally by two types of enzymes termed protein arginine methyltransferases (PRMTs). Type I PRMTs generate aDMAs, type II enzymes produce sDMAs, and both types catalyze the formation of MMAs as intermediates (Figure 1.4; Bedford and Richard, 2005). Tudor domains mainly interact with sDMAs, which are often found in the repeated sequence motifs composed of an arginine flanked by either two glycines (GRG) or by a glycine and an alanine (GRA or ARG) (Cote and Richard, 2005). In flies, symmetric dimethylation of arginines is catalyzed by dPRMT5 (also known as Capsuleenn (Csul) or Dart5) with the aid of its cofactor Valois (Vls) (Gonsalvez et al., 2006; Anne and Mechler, 2005). Interestingly, genetic disruption of *dprmt5* results in a similar phenotype as *aub* mutation (Gonsalvez et al., 2006). Like Aub and Tud, Vls is also localized to the nuage and pole plasm in ovaries. Moreover, it has been shown that Vls interacts with both dPRMT5 and Tud and is required for the localization of Tud in both the nuage and pole plasm. Taken together, these observations hint at potential roles of sDMA modification and Tudor domain proteins in piRNA biogenesis and function.

Indeed, recent studies found that all PIWI proteins in *Drosophila* contain sDMAs, which were recognized by an sDMA-specific antibody, SYM11 (Kirino et al., 2009). sDMAs on Aub and Ago3 were also mapped by mass spectrometry. All sDMAs are in the N terminal regions: R11, R13, and R15 in Aub and R4, R68, and R70 in Ago3 (Nishida et al., 2009). The sDMA sites on Piwi remain to be identified. Moreover, these sDMAs were not detectable in *dprmt5* mutant ovaries suggest that dPRMT5 directly modifies Ago3, Aub, and Piwi in flies. The initial study also suggests that sDMAs are essential for Piwi protein stabilization because expression of PIWI proteins was reduced in *dprmt5* mutant ovaries

Figure 1.4.

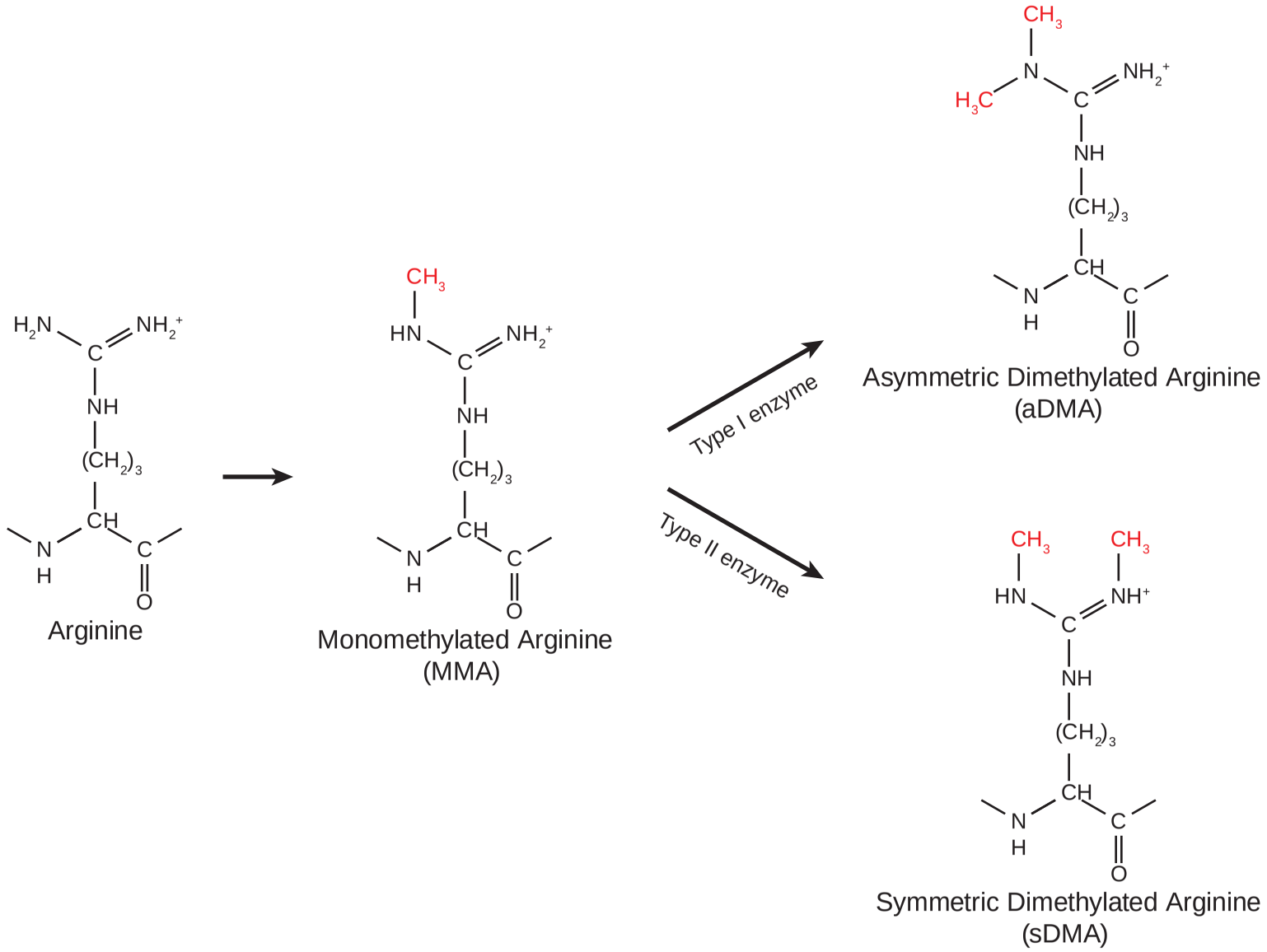

Arginine
Monomethylated Arginine
(MMA)
Type I enzyme
Type II enzyme
Asymmetric Dimethylated Arginine
(aDMA)
Symmetric Dimethylated Arginine
(sDMA)

Figure 1.4. Arginine methylation in eukaryotes.

compared to *dprmt5* heterozygous as well as wild type ovaries (Kirino et al., 2009). However, this result is contested, because subsequent studies indicated that loss of dPRMT5 does not affect the protein level of Aub (Nishida et al., 2009; Kirino et al., 2010b).

Drosophila Tud is a key component of nuage and pole plasm and is indispensable for gametogenesis (Boswell and Mahowald, 1985; Thomson and Lasko, 2004; Arkov et al., 2006). Tud contains eleven Tudor domains. As expected, Aub was found to associate with Tud in wild type ovaries and their interaction was disrupted in *dprmt5* mutant ovaries, suggesting that the binding of Tud to Aub depends on sDMAs (Nishida et al., 2009; Kirino et al., 2010b). Like Aub, Ago3 was also co-immunoprecipitated with Tud from wild type ovaries. Pull-down assays revealed that Tud is able to bind an Ago3 peptide (Thr 58 to His 82) with sDMAs but not the one without sDMAs. These observations suggest that sDMAs might also be required for the interaction of Tud and Ago3 in vivo (Nishida et al., 2009). As Ago3 and Aub are the major contribution to the ping-pong amplification cycle, these postulations reveal an interesting hypthesis, in which Tud serves as platform to accommodate both Aub and Ago3. The close proximity might facilitate the loading of 3´ cleavage products—precursors of secondary piRNAs—to Aub and/or Ago3. Supporting this view, less mature piRNAs were detected in anti-Tud immunoprecipitation as compared to anti-Aub immunoprecipitation, even if both contain similar amount of Aub. Moreover large (80-1000 nt) piRNA precursor-like molecules, in both sense and antisense orientations, were only found in anti-Tud immunoprecipitation.

However, there are still controversies. Previous immunoprecipitation experiments also indicate that Aub interacts with Ago3 regardless of the

presence of Tud, suggesting that Tud is dispensable for the interaction of Ago3 and Aub. Nevertheless, this does not rule out participation of other Tudor domain proteins to bridge the interaction between Aub and Ago3. Loss of Tud does not alter the protein levels of Piwi, Aub, or Ago3, whereas piRNAs from different transposon families were significantly varied in *tud* mutant ovaries (Nishida et al., 2009; Kirino et al., 2010b). Moreover, loss of Tud seems to increase the total piRNAs loaded into Aub and Ago3. These are in striking contrast to *dprmt5* mutants, in which the Aub-bound piRNAs are and *roo* piRNAs are substantially decreased. Furthermore, the RNA levels of *Het-A*, *accord2*, and *blood* increase in *dprmt5* mutants, whereas only *blood* increases in *tud* mutants. These observations clearly indicate that other Tudor domain proteins may participate in piRNA biogenesis and function via interacting with sDMAs on their partners. Interestingly, recent studies indicate that Vasa contains both sDMAs and aDMAs (Kirino et al., 2010a).

Drosophila **piRNAs are born to combat genomic parasites**

Transposons and repetitive sequences occupy about 30 % of the fly genome and comprise about 150 families (Kapitonov and Jurka, 2003; Biemont and Vieira, 2006). Propagation of these genomic parasites leads to accumulation of deleterious mutations, ectopic chromosomal recombination and inevitably destruction of genome integrity. In *Drosophila* gonads, two piRNA pathways were elaborated to repress these selfish genetic elements. About 80% of piRNAs map to annotated transposons and repetitive sequences with coverage of all categories of selfish genetic elements (Brennecke et al., 2007). Mutation in any member of the piRNA pathway causes derepression of transposons and leads to

reduction of fertility (Aravin et al., 2001; Vagin et al., 2006; Li et al., 2009; Brennecke et al., 2007; Kalmykova et al., 2005; Savitsky et al., 2006; Sarot et al., 2004). In germ cells, transposon silencing is achieved mainly through a posttranscriptional slicer-dependent amplification mechanism. However, emerging evidence also implies that a transcriptional gene silencing mechanism may also be involved. Piwi is localized in the nuclei of both germ cells and gonadal somatic cells and its nuclear localization is required for transposon silencing (Saito et al., 2010). In addition, transposon silencing by Piwi is dispensable of its slicer activity (Saito et al., 2010). Piwi interacts with HP1 and binds to heterochromatin in somatic cells (Brower-Toland et al., 2007). Mutations in *piwi*, *aub*, and *spn-E* suppress position effect variegation (PEV), a phenomenon caused heterochromatin assembly (Brower-Toland et al., 2007; Pal-Bhadra et al., 2004; Pal-Bhadra et al., 2002). These observations suggest that piRNAs might guide Piwi to silence transposons by directing assembly of heterochromatin.

The regulation of protein-coding genes by piRNAs

Although the majority of *Drosophila* piRNA target transposons, a few cases suggest that piRNAs also regulate the expression of protein coding genes. In testes, the *Su(Ste)* piRNAs repress the expression of *Stellate* that encodes a casein kinase II beta-subunit like protein (Aravin et al., 2001; Vagin et al., 2006; Aravin et al., 2003; Bozzetti et al., 1995; Palumbo et al., 1994; Balakireva et al., 1992). In *Su(Ste)* mutants, small RNAs targeting *Stellate* are lost, causing *Stellate* protein crystals to form in primary spermatocytes and reduced fertility (Livak, 1990; Livak, 1984; Pimpinelli et al., 1985). Moreover, Aub-bound AT-chX

piRNAs derived from the AT-chX locus show high complementarity to the *vasa* gene and downregulate the expression of *vasa* in testes (Nishida et al., 2007; Li et al., 2009). In somatic follicle cells, Piwi and piRNAs derived from the 3´ UTR of *traffic jam* (*tj*) mRNAs, which encodes a basic-leucine zipper transcription factor (Saito et al., 2009) might also negatively regulate the expression of *fasciclin 3* (*fas3*) that encodes an immunoglobulin-like cell adhesion molecule. Mutation in *piwi* or *tj* causes the over expression of Fas3, resulting in failure of intermingling of germline and somatic cells in the ovaries (Li et al., 2003; Saito et al., 2009). During maternal-zygotic transition of embryogenesis, certain species of *roo* and *412* piRNAs may induce deadenylation of *nanos* mRNA through targeting the partial complementary sequences in the 3´UTR (Rouget et al., 2010). The mechanism of regulation of these protein-coding genes by piRNAs remains unknown. It might be different from posttranscriptional RNA cleavage, as basically piRNAs targeting mRNAs show less complementary than those targeting transposons.

CHAPTER II: A DISTINCT SMALL RNA PATHWAY SILENCES SELFISH GENETIC ELEMENTS IN THE GERM LINE

Preface

The work presented in the following chapter was a collaborative effort. Vasily V. Vagin performed the Northern hybridization analyses in Figure 2.2B, 2.4C, 2.5, 2.6A, 2.6B, 2.S4, and 2.S5. Vasily V. Vagin also prepared RNA samples for tiling microarray analysis and Dr. Hervé Seitz analyzed the tiling microarray data in Figure 2.1, 2.2A, 2.4A, 2.4B, 2.S1, 2.S2, and 2.S3. Alla Sigova performed qRT-PCR analyses in Figure 2.3A, 2.3B, 2.3C, and 2.3D. Experiments in Figure 2.3E were collaborations between Alla Sigova and Vasily V. Vagin. I carried out immunopreciptation, Northern hybridization, and Western blotting analyses in Figure 2.6C, and 2.6D, and 2.S6 demonstrating Piwi and Aub specifically bind rasiRNAs.

Summary

In the *Drosophila* germ line, repeat-associated small interfering RNAs (rasiRNAs) ensure genomic stability by silencing endogenous selfish genetic elements such as retrotransposons and repetitive sequences. While small interfering RNAs (siRNAs) derive from both the sense and antisense strands of their double-stranded RNA precursors, rasiRNAs arise mainly from the antisense strand. rasiRNA production appears not to require Dicer-1, which makes microRNAs, or Dicer-2, which makes siRNAs, and rasiRNAs lack the 2´,3´ hydroxy termini characteristic of animal siRNA and miRNA. Unlike siRNAs and miRNAs, rasiRNAs function through the Piwi, rather than the Ago, Argonaute protein subfamily. Our data suggest that rasiRNAs protect the fly germ line through a silencing mechanism distinct from both the miRNA and RNAi pathways.

Results and Discussion

In plants and animals, RNA silencing pathways defend against viruses (Galiana-Arnoux et al., 2006; Wang et al., 2006; Zambon et al., 2006), regulate endogenous gene expression (Du and Zamore, 2005), and protect the genome against selfish genetic elements such as retrotransposons and repetitive sequences (Meister and Tuschl, 2004).

Common to all RNA silencing pathways are RNAs 19–30 nt long that specify the target RNAs to be repressed. In RNA interference (RNAi) (Fire et al., 1998), small interfering RNAs (siRNAs) are produced from long exogenous double-stranded RNA (dsRNA). In contrast, ~22 nt microRNAs (miRNAs) are endonucleolytically processed from endogenous RNA Pol II transcripts. Dicer RNase III enzymes produce both siRNAs and miRNAs. In flies, Dicer-2 (Dcr-2) generates siRNAs, whereas the Dicer-1 (Dcr-1)/Loquacious (Loqs) complex produces miRNAs (Lee et al., 2004; Hoa et al., 2003; Forstemann et al., 2005; Saito et al., 2005; Jiang et al., 2005). After their production, small silencing RNAs bind Argonaute proteins to form the functional RNA silencing effector complexes believed to mediate all RNA silencing processes.

Phased sense and antisense siRNAs in vivo

In *Drosophila*, processive dicing of long dsRNA and the accumulation of sense and antisense siRNAs without reference to the orientation of the target mRNA are hallmarks of RNAi in vitro (Schwarz et al., 2003; Zamore et al., 2000) and in vivo (Figure 2.1). We prepared total small RNA from the heads of adult males expressing a dsRNA hairpin (Figure 2.S1A) that silences the *white* gene

Figure 2.1.

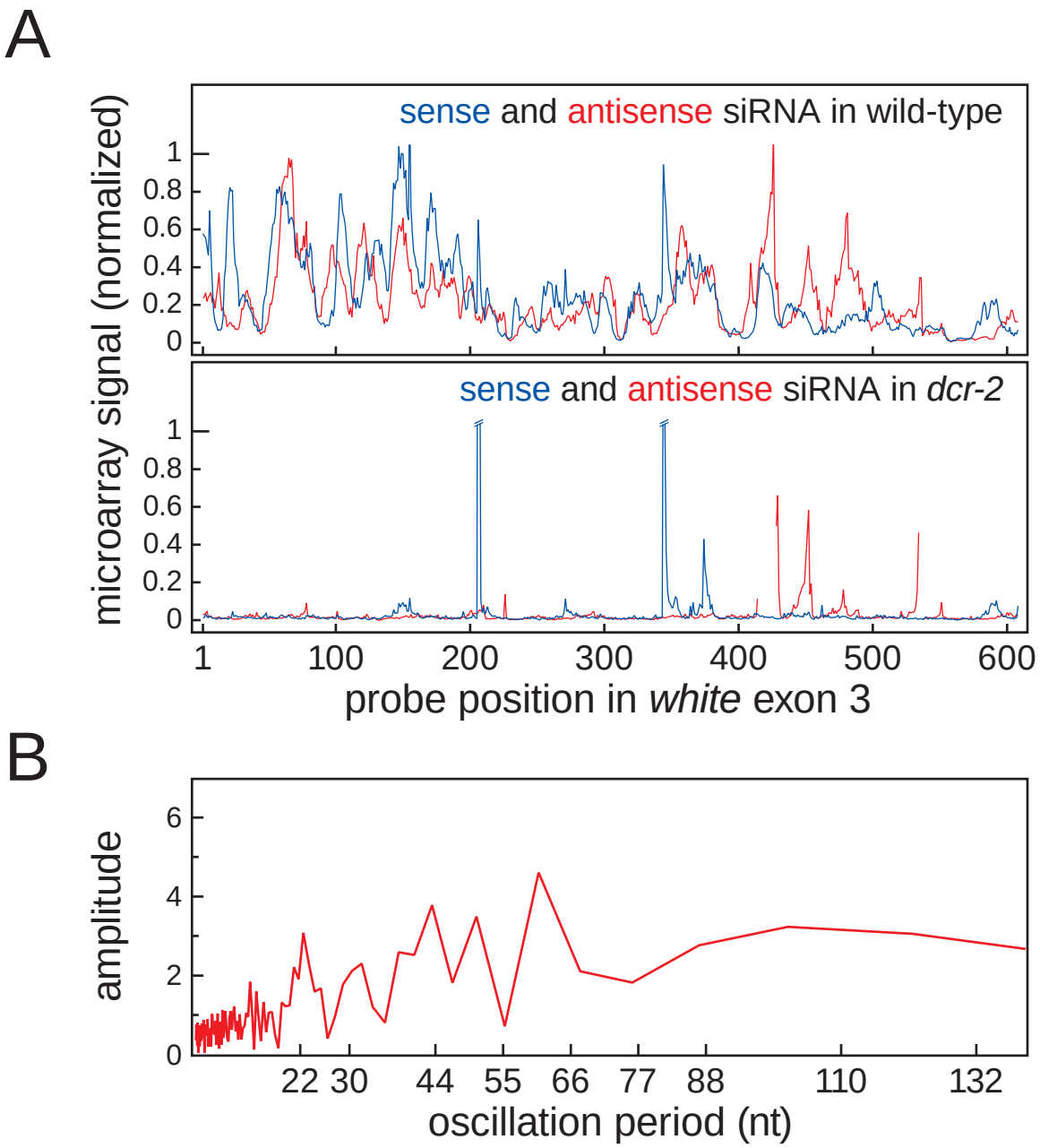

A
sense and antisense siRNA in wild-type
sense and antisense siRNA in dcr-2
microarray signal (normalized)
0
0.2
0.4
0.6
0.8
1
1
100
200
300
400
500
600
probe position in white exon 3
B
amplitude
0
2
4
6
22 30
44
55
66
77
88
110
132
oscillation period (nt)

Figure 2.1. *white* siRNAs accumulate during RNAi in vivo.

(A) Microarray analysis of the siRNAs derived from the *white* exon 3 hairpin RNAi trigger. (B) Fourier-transform analysis of the Dcr-2-dependent siRNAs in (A).

via the RNAi pathway (Lee and Carthew, 2003). *white* silencing requires Dcr-2 (Lee et al., 2004), R2D2 (Forstemann et al., 2005), and Ago2. siRNAs were detected using a microarray containing T_M-normalized, 22-nt long probes for all sense and antisense siRNAs that theoretically can be produced by dicing the *white* exon 3 hairpin (Figure 2.1A). Both sense and antisense *white* siRNAs were detected in wild-type, but not *dcr-2*^{L811fsX} homozygous mutant flies. The Dcr-2-dependent siRNAs were produced with a ~22 nt periodicity (Figure 2.1B), consistent with the phased processing of the dsRNA hairpin from the end formed by the six nt loop predicted to remain after splicing of its intron-containing primary transcript (Figure 2.S1B).

Su(Ste) rasiRNAs

Drosophila rasiRNAs can be distinguished from siRNAs by their longer length, 24–29 nts (Aravin et al., 2003; Aravin et al., 2001). rasiRNAs have been proposed to be diced from long dsRNA triggers (Aravin et al., 2001; Meister and Tuschl, 2004), such as the ~50 copies of the bi-directionally transcribed *Suppressor of Stellate* (*Su(Ste)*) locus on the Y-chromosome (Tritto et al., 2003) (Figure 2.S2B) that in testes silence the ~200 copies of the protein-coding gene, *Stellate* (*Ste*), found on the X-chromosome.

Microarray analysis of total small RNA isolated from fly testes revealed that *Su(Ste)* rasiRNAs detectably accumulate only from the antisense strand (Figure 2.2A), with little or no phasing (Figure 2.S2A). As expected, *Su(Ste)* rasiRNAs were not detected in testes from males lacking the *Su(Ste)* loci (*cry*^{1}Y) (Figure 2.2A). *Su(Ste)* rasiRNA were also absent from *armitage* (*armi*) mutant testes (Figure 2.2A), which fail to silence *Ste* and do not support RNAi in vitro

Figure 2.2.

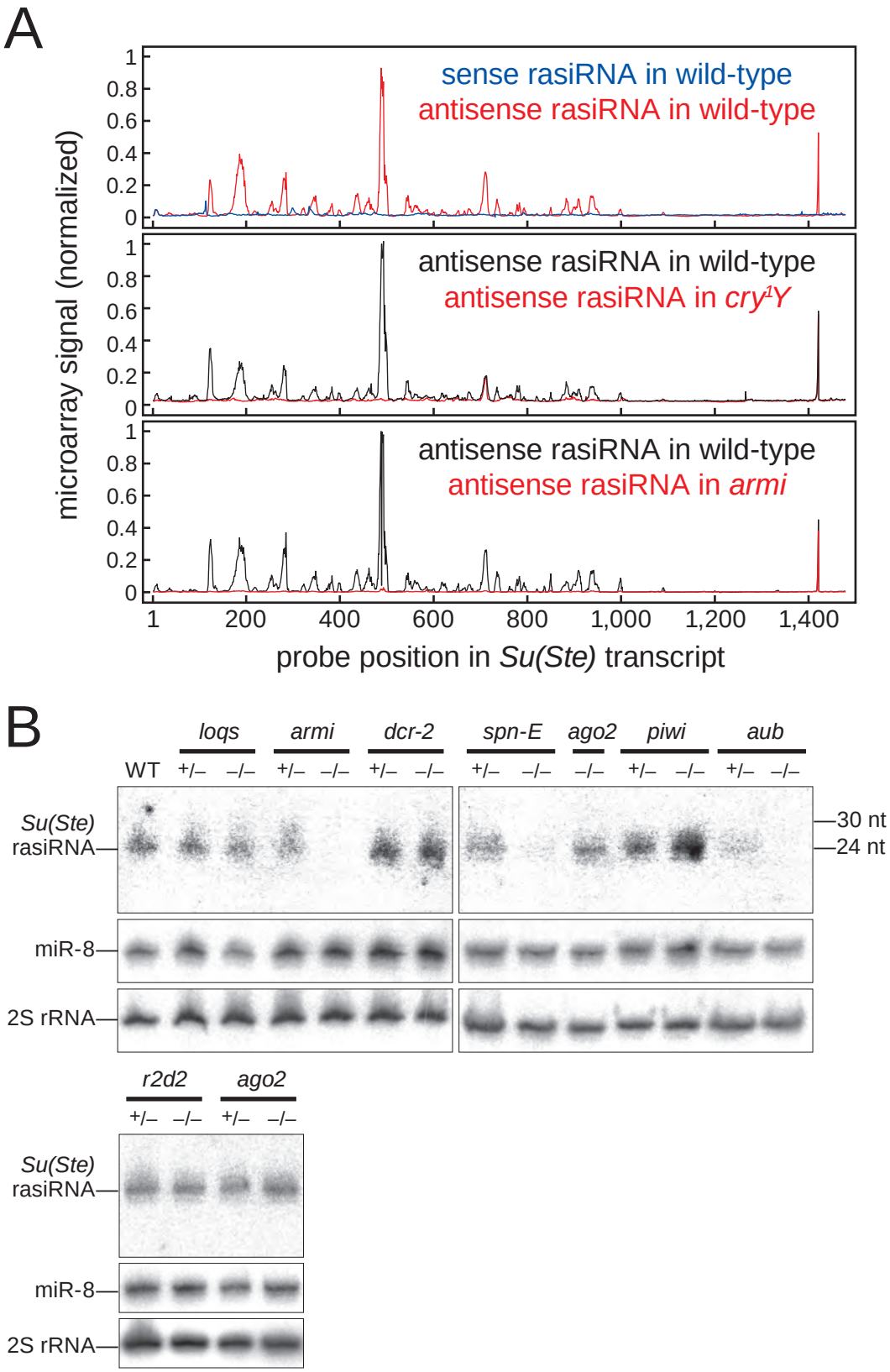

A
sense rasiRNA in wild-type
antisense rasiRNA in wild-type
antisense rasiRNA in wild-type
antisense rasiRNA in cry1Y
antisense rasiRNA in wild-type
antisense rasiRNA in armi
microarray signal (normalized)
probe position in Su(Ste) transcript
B
loqs
armi
dcr-2
spn-E
ago2
piwi
aub
WT
+/–
–/–
Su(Ste) rasiRNA
miR-8
2S rRNA
30 nt
24 nt
r2d2
ago2

Figure 2.2. *Su(Ste)* rasiRNAs in testes.

(A) Microarray analysis of *Su(Ste)* small RNAs in wild-type, *cry*^{1}Y, and *armi* mutant testes. (B) Northern analysis of the most abundant *Su(Ste)* rasiRNA in testes mutant for RNA silencing genes.

(Tomari et al., 2004). *armi* encodes a non-DEAD box helicase (Cook et al., 2004) homologous to the *Arabidopsis thaliana* protein SDE3, which is required for RNA silencing triggered by transgenes and some viruses (Dalmay et al., 2001), and depletion by RNAi of the mammalian Armi homolog, Mov10, blocks siRNA-directed RNAi in cultured human cells (Meister et al., 2005). Normal accumulation of *Su(Ste)* rasiRNA and robust *Ste* silencing also requires the putative helicase, Spindle-E (Spn-E), a member of the DExH family of adenosine triphosphatases (ATPases) (Aravin et al., 2001; Stapleton et al., 2001) (Figure 2.2B and Figure 2.S2B).

The accumulation in vivo of only antisense rasiRNAs from *Su(Ste)* implies that either sense *Su(Ste)* rasiRNA are not produced or they are selectively destroyed. Either process would make *Ste* silencing mechanistically different from RNAi. Supporting this view, mutations in the central components of the *Drosophila* RNAi pathway—*dcr-2*, *r2d2*, and *ago2*—did not diminish *Su(Ste)* rasiRNA accumulation (Figure 2.2B). Deletion of the *Su(Ste)* silencing trigger (cry^{1}Y) caused a ~65-fold increase in *Ste* mRNA (Figure 2.3A), but null or strong hypomorphic mutations in the three key RNAi proteins did not (Figure 2.3B).

Fly Argonaute proteins can be subdivided into Ago (Ago1 and Ago2) and Piwi (Aubergine (Aub), Piwi, and Ago3) subfamilies. Unlike *ago1* and *ago2*, *aub*, *piwi*, and *ago3* mRNAs are enriched at the germ line (Williams and Rubin, 2002; Cox et al., 1998; Harris and Macdonald, 2001). Aub is required for *Ste* silencing (Aravin et al., 2001) and *Su(Ste)* rasiRNA accumulation (Aravin et al., 2004). In aub^{HN2}/aub^{QC42} trans-heterozygous mutants, *Su(Ste)* rasiRNA were not detected by microarray (Figure 2.S2B) or Northern analysis (Figure 2.2B), and *Su(Ste)*-triggered silencing of *Ste* mRNA was lost completely (Figure 2.3B). Even

Figure 2.3.

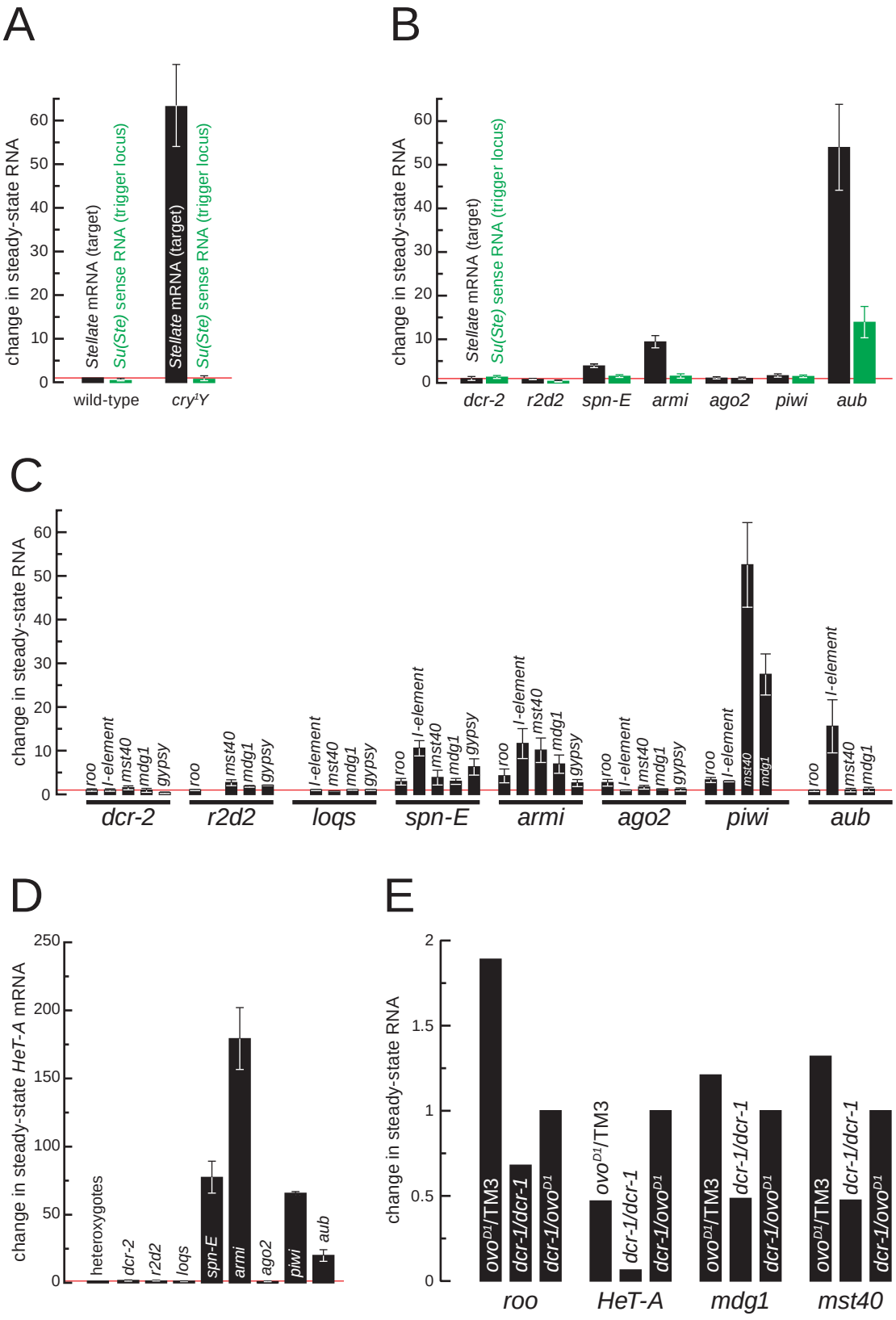

A
change in steady-state RNA
Stellate mRNA (target)
Su(Ste) sense RNA (trigger locus)
wild-type
cry1Y
B
dcr-2
r2d2
spn-E
armi
ago2
piwi
aub
C
roo
I-element
mst40
mdg1
gypsy
loqs
D
change in steady-state HeT-A mRNA
heteroxygotes
E
change in steady-state RNA
ovoD1/TM3
dcr-1/dcr-1
dcr-1/ovoD1
HeT-A
mdg1
mst40

Figure 2.3. qRT-PCR analysis of RNAs from selfish genetic elements.
RNA expression from selfish genetic elements was measured in homozygous mutants relative to heterozygotes for *Ste* silencing in testes (A and B); and for the repeated locus *mst40,* the LTR-retrotransposons *roo*, *mdg1*, and *gypsy*, and the non-LTR-retrotransposons *I-element*, and *HeT-A* in ovaries (C and D) . (E) RNA expression from selfish genetic elements in *dcr-1*^{Q1147X} null mutant clones generated by mitotic recombination in the ovary.

aub^{HN2}/+ heterozygotes accumulated less of the most abundant *Su(Ste)* rasiRNA than wild-type (Figure 2.2B). That the Ago subfamily protein Ago2 is not required, whereas the Piwi subfamily protein Aub is essential for *Ste* silencing supports the view that *Ste* is silenced by a pathway distinct from RNAi. Intriguingly *Su(Ste)* rasiRNAs hyper-accumulated in *piwi* mutant testes, where *Ste* is silenced normally (Figure 2.2B, 2.3B and Figure 2.S2B).

Mutations in *aub* also cause an increase in sense, but not antisense *Su(Ste)* RNA (Aravin et al., 2001); our results suggest that antisense *Su(Ste)* rasiRNAs can silence both *Ste* mRNA and sense *Su(Ste)* RNA, but that no *Su(Ste)* rasiRNA exist that can target the antisense *Su(Ste)* transcript. Our finding that *Su(Ste)* rasiRNAs are predominantly or exclusively antisense essentially agrees with small RNA cloning experiments, in which four of five *Su(Ste)* rasiRNAs sequenced were in the antisense orientation (Aravin et al., 2003), but is at odds with earlier reports detecting both sense and antisense *Su(Ste)* rasiRNAs by non-quantitative Northern hybridization (Aravin et al., 2001).

A third RNA silencing pathway in flies

Is germ-line RNA silencing of selfish genetic elements generally distinct from the RNAi and miRNA pathways? We examined the expression of a panel of germ-line expressed selfish genetic elements—three LTR-containing retrotransposons (*roo*, *mdg1*, and *gypsy*), two non-LTR retrotransposons (*I-element* and *HeT-A*, a component of the *Drosophila* telomere), and a repetitive locus (*mst40*)—in mutants defective for eight RNA silencing proteins. All selfish genetic elements tested behaved like *Ste*: loss of the RNAi proteins Dcr-2, R2D2, or Ago2 had little or no effect on retrotransposon or repetitive element silencing

(Figure 2.3C and 2.3D). Instead, silencing required the putative helicases Spn-E and Armi, and one or both of the Piwi subfamily Argonaute proteins, Aub or Piwi. Silencing did not require Loqs, the dsRNA-binding protein required to produce miRNAs (Figure 2.3C and 2.3D).

The null allele *dcr-1*^{Q1147X} is homozygous lethal, making it impossible to procure *dcr-1* mutant ovaries from *dcr-1*^{Q1147X}/*dcr-1*^{Q1147X} adult females (Lee et al., 2004). Therefore, we generated clones of *dcr-1*^{Q1147X}/*dcr-1*^{Q1147X} cells in the ovary by mitotic recombination in flies heterozygous for the dominant female-sterile mutation, *ovo*^{D1} (Sahut-Barnola and Pauli, 1999). We measured RNA levels, relative to *rp49* mRNA, for three retrotransposons (*roo*, *HeT-A*, and *mdg1*) and one repetitive sequence (*mst40*) in *dcr-1*/*dcr-1* recombinant ovary clones and in *ovo*^{D1}/TM3 and *dcr-1*/*ovo*^{D1} non-recombinant ovaries. The *ovo*^{D1} mutation blocks oogenesis at stage 4, after the onset of *HeT-A* (Vagin et al., 2004) and *roo* rasiRNA production. Retrotransposon or repetitive sequence transcript abundance was unaltered or decreased in *dcr-1*/*dcr-1* relative to *ovo*^{D1}/TM3 and *dcr-1*/*ovo*^{D1} controls (Figure 2.3E). We conclude that Dcr-1 is dispensable for silencing these selfish genetic elements in the *Drosophila* female germ line.

roo is the most abundant LTR-retrotransposon in flies. We analyzed *roo* silencing in the female germ line using microarrays containing 30 nt probes, tiled at 5 nt resolution, for all ~18,000 possible *roo* rasiRNAs (Figure 2.4A and 2.4B); we corroborated the data at 1 nt resolution for those rasiRNAs derived from LTR sequences (Figure 2.S3A). As observed for *Su(Ste)*, but not for *white* RNAi, *roo* rasiRNAs were non-homogeneously distributed along the *roo* sequence and accumulated primarily from the antisense strand (Figure 2.4A and Figure 2.S3A). In fact, the most abundant sense rasiRNA peak (marked with an asterisk in

Figure 2.4.

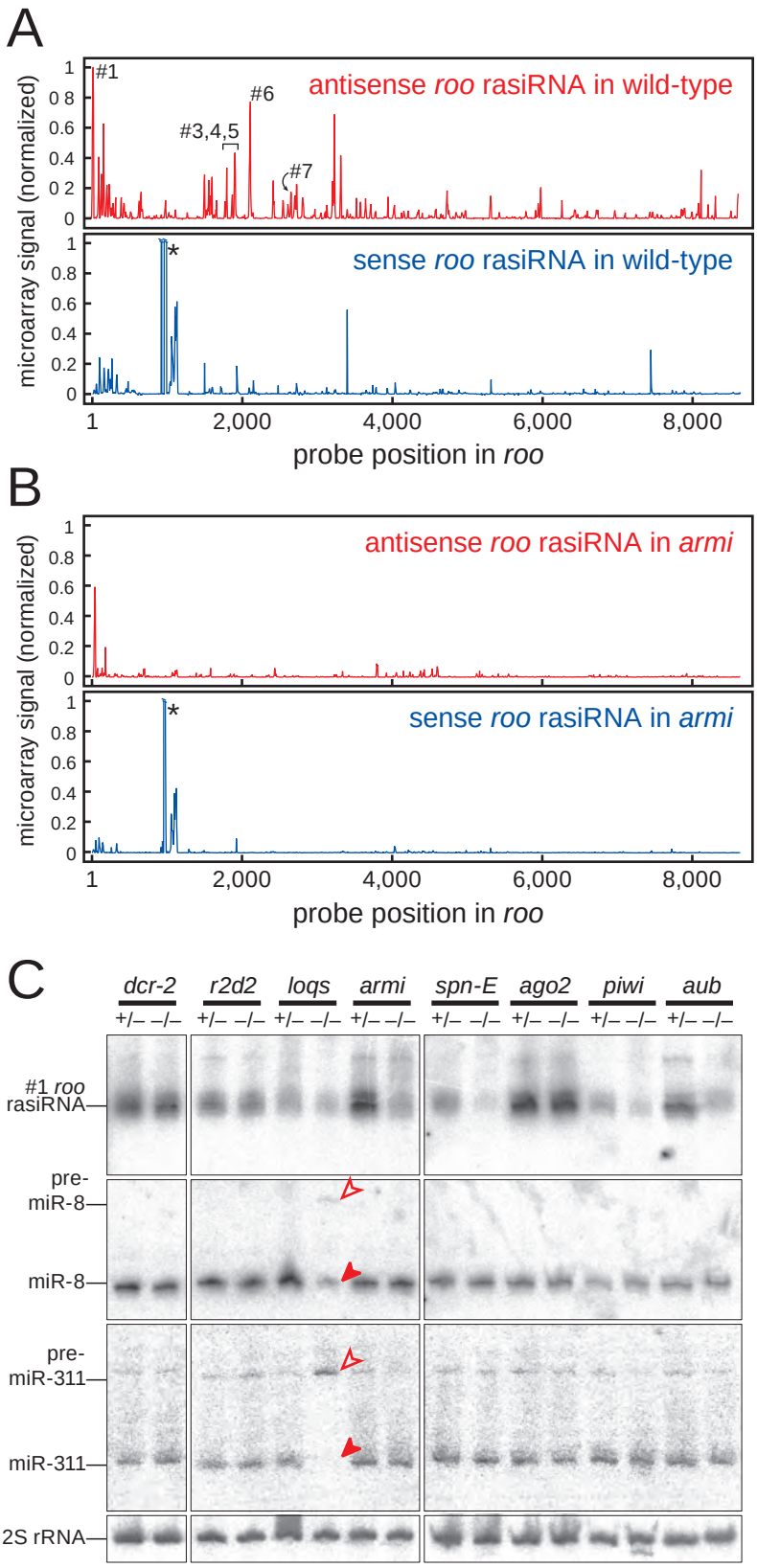

A
antisense roo rasiRNA in wild-type
sense roo rasiRNA in wild-type
#1
#6
#3,4,5
#7
*
microarray signal (normalized)
probe position in roo
B
antisense roo rasiRNA in armi
sense roo rasiRNA in armi
C
dcr-2
r2d2
loqs
armi
spn-E
ago2
piwi
aub
+/– –/–
#1 roo rasiRNA
pre-miR-8
miR-8
pre-miR-311
miR-311
2S rRNA

Figure 2.4. *roo* rasiRNAs in ovaries.

roo rasiRNAs in ovaries arise mainly from the antisense strand and require the putative helicase Armi for their accumulation. (A) Microarray analysis of *roo* rasiRNAs in wild-type and (B) *armi* mutant ovaries. The asterisk marks a peak arising from non-specific hybridization to poly(A)-containing RNAs. (C) Northern analysis for *roo* rasiRNA (peak #1 in (A)) in ovaries heterozygous (+/–) or homozygous mutant (–/–) for components of the RNA silencing machinery. The open red arrowhead highlights pre-miR-311 accumulation, and the filled red arrowhead, loss of mature miR-311, in *loqs*^{f00791}; miR-311 is expressed predominantly in the germ line.

Figure 2.4A and 2.4B) corresponded to a set of probes containing 16 contiguous uracil residues, suggesting these probes non-specifically detected fragments of the mRNA poly(A) tail; most of the remaining sense peaks were unaltered in *armi* mutant ovaries, in which *roo* expression is increased, suggesting that they do not contribute to *roo* silencing (Figure 2.3C and 2.4A). We detected no phasing in the distribution of *roo* rasiRNAs (Figure 2.S3B).

As for *Su(Ste)*, wild-type accumulation of antisense *roo* rasiRNA required the putative helicases Armi and Spn-E and the Piwi subfamily Argonaute proteins Piwi and Aub, but not the RNAi proteins Dcr-2, R2D2, and Ago2 (Figure 2.4B and 4C). Moreover, accumulation of *roo* rasiRNA was not measurably altered in *loqs*^{f00791}, an allele that strongly disrupts miRNA production in the female germ line (Figure 2.4C).

Are *roo* rasiRNAs not made by dicing?

Loss of Dcr-2 or Dcr-1 did not increase retrotransposon or repetitive element expression, suggesting that neither enzyme acts in rasiRNA-directed silencing. Moreover, loss of Dcr-2 had no detectable effect on *Su(Ste)* rasiRNA in testes or *roo* rasiRNA in ovaries (Figure 2.2B and 2.4C). We measured the amount of *roo* rasiRNA and miR-311 in *dcr-1*/*dcr-1* ovary clones generated by mitotic recombination. Comparison of recombinant (*dcr-1*/*dcr-1*) and non-recombinant (*ovo*^{D1}/TM3 and *dcr-1*/*ovo*^{D1}) ovaries by Northern analysis revealed that *roo* rasiRNA accumulation was unperturbed by the null *dcr-1*^{Q1147X} mutation (Figure 2.5A and Figure 2.S4). Pre-miR-311 increased and miR-311 declined ~3-fold in the *dcr-1*/*dcr-1* clones (Figure 2.5B and Figure 2.S4), consistent with about two-thirds of the tissue corresponding to mitotic *dcr-1*/*dcr-1* recombinant

Figure 2.5.

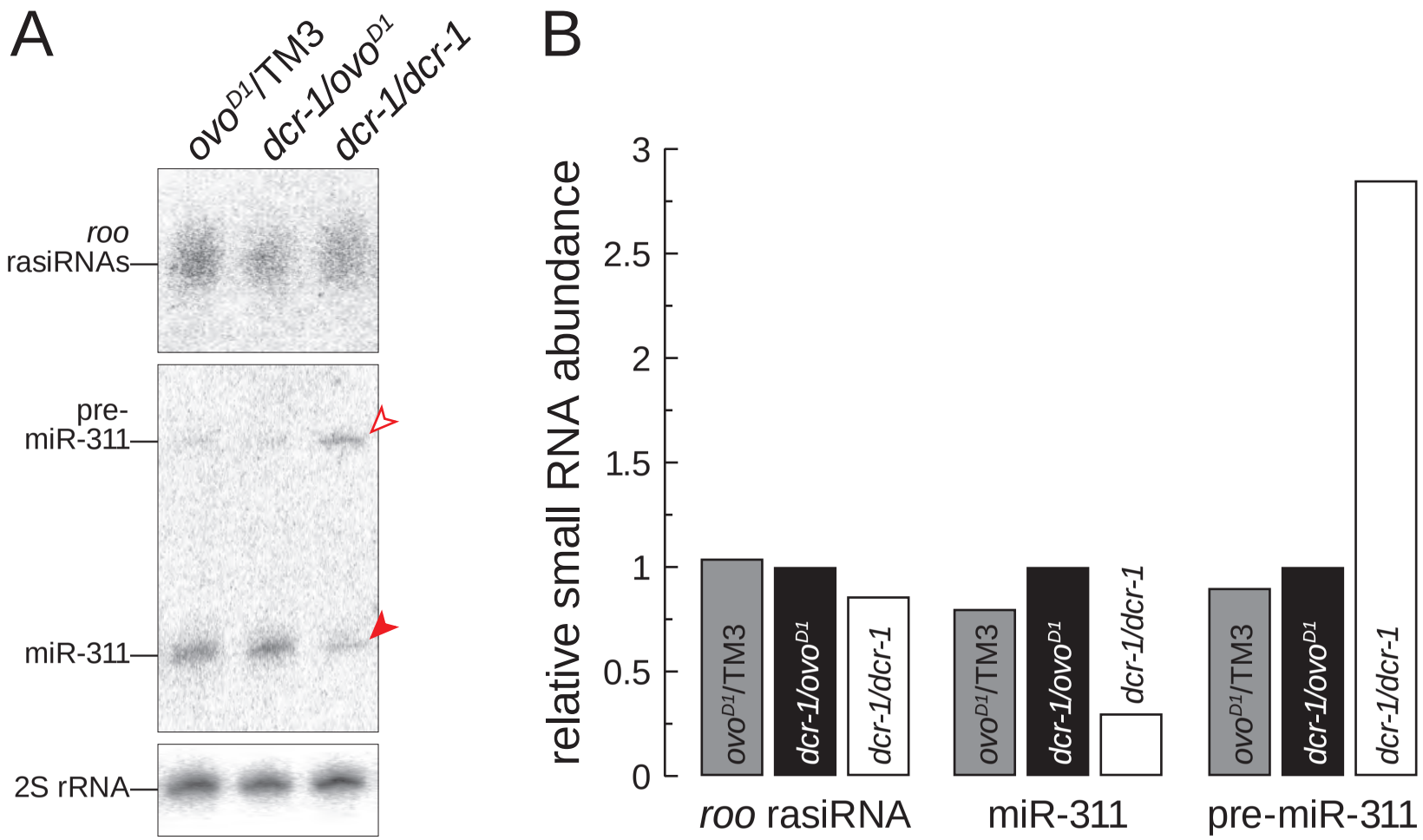

A
ovoD1/TM3
dcr-1/ovoD1
dcr-1/dcr-1
roo rasiRNAs
pre-miR-311
miR-311
2S rRNA
B
relative small RNA abundance
0
0.5
1
1.5
2
2.5
3
ovoD1/TM3
dcr-1/ovoD1
dcr-1/dcr-1
roo rasiRNA
miR-311
pre-miR-311

Figure 2.5. Northern analysis of *roo* rasiRNAs in *dcr-1* mutant ovaries. (A) Northern analysis of mitotic recombinant *dcr-1*^{Q1147X} homozygous mutant ovaries and non-recombinant controls. *roo* rasiRNA were detected using a mixture of five hybridization probes. Arrowheads as in Figure 2.4C. (B) Quantification of the data in (A), normalized to the 2S rRNA loading control.

cells. Yet, although most of the tissue lacked *dcr-1* function, we observed improved, rather than diminished, silencing for the four selfish genetic elements examined (Figure 2.3E). Moreover, the dsRNA-binding protein, Loqs, which acts with Dcr-1 to produce miRNAs, was also dispensable for *roo* rasiRNA production and selfish genetic element silencing (Figure 2.3C, 2.3D, and 2.4C). While we cannot exclude the possibility that *dcr-1* and *dcr-2* can fully substitute for each other in the production of rasiRNA in the ovary, previous biochemical evidence suggests that none of the three RNase III enzymes in flies—Dcr-1, Dcr-2, and Drosha—can cleave long dsRNA into small RNAs 24–30 nt long (Bernstein et al., 2001; Saito et al., 2005; Liu et al., 2003).

Animal siRNA and miRNA contain 5′ phosphate and 2′,3′ hydroxy termini (Elbashir et al., 2001; Hutvagner et al., 2001). We used enzymatic and chemical probing to infer the terminal structure of *roo* and *Su(Ste)* rasiRNAs. RNA from ovaries or testes was treated with calf intestinal phosphatase (CIP) or CIP followed by polynucleotide kinase plus ATP. CIP-treatment caused *roo* (Figure 2.6A) and *Su(Ste)* (Figure 2.S5) rasiRNA to migrate more slowly in polyacrylamide gel electrophoresis, consistent with the loss of one or more terminal phosphate groups. Subsequent incubation with polynucleotide kinase and ATP restored the original gel mobility of the rasiRNAs, indicating that they contained a single 5′ or 3′ phosphate before CIP treatment. The *roo* rasiRNA served as a substrate for ligation of a 23 nt long 5′ RNA adapter by T4 RNA ligase, a process that requires a 5′ phosphate; pre-treatment with CIP blocked ligation (Figure 2.6B), establishing that the monophosphate lies at the 5′ end. The rasiRNA must also contain at least one terminal hydroxyl group, because it could be joined by T4 RNA ligase to a pre-adenylated, 17 nt long 3′ RNA adapter

Figure 2.6.

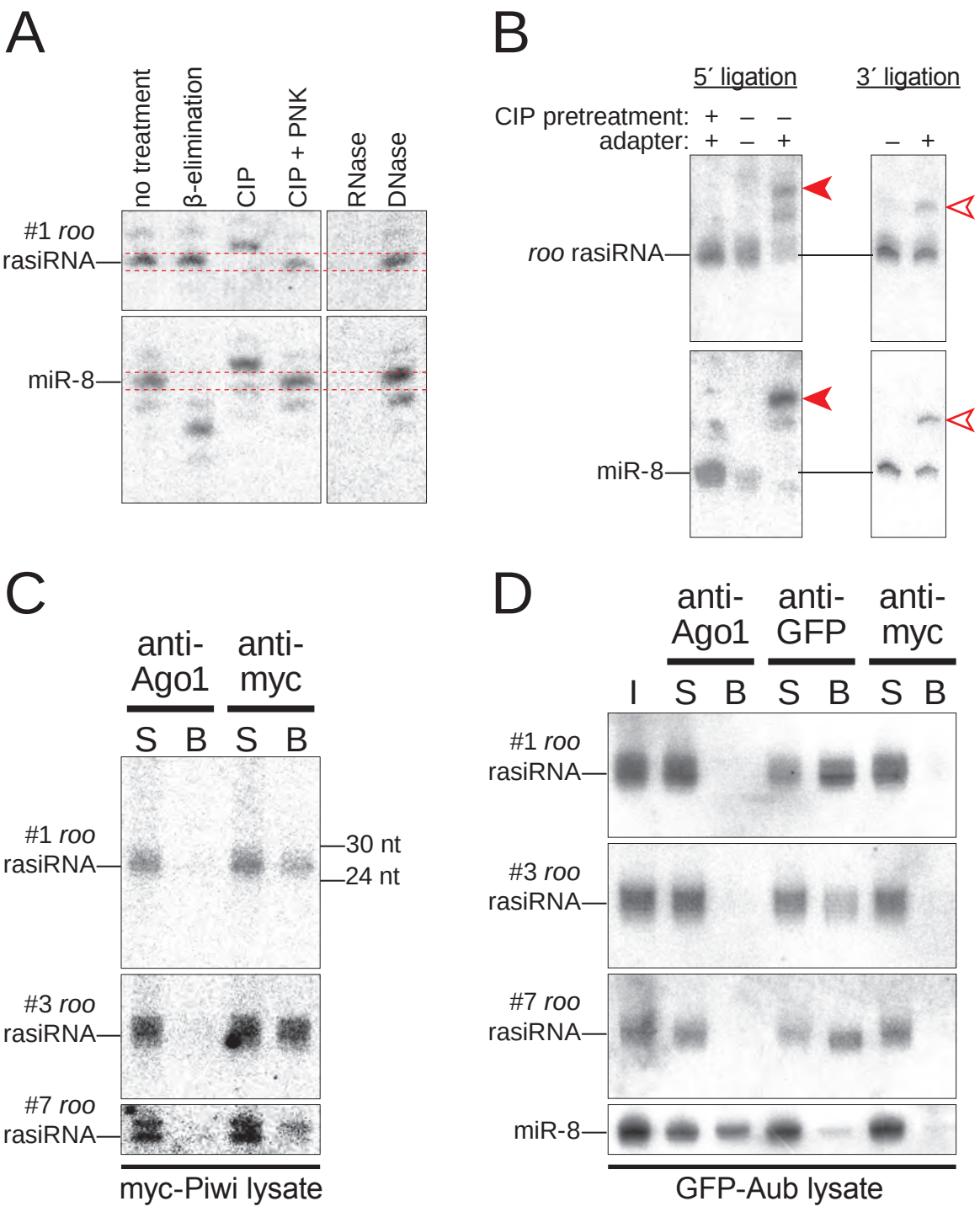

A
no treatment
β-elimination
CIP
CIP + PNK
RNase
DNase
#1 roo rasiRNA
miR-8
B
5′ ligation
3′ ligation
CIP pretreatment: + – –
adapter: + – +
– +
roo rasiRNA
miR-8
C
anti-Ago1
anti-myc
S B S B
#1 roo rasiRNA
30 nt
24 nt
#3 roo rasiRNA
#7 roo rasiRNA
myc-Piwi lysate
D
anti-Ago1
anti-GFP
anti-myc
I S B S B S B
#1 roo rasiRNA
#3 roo rasiRNA
#7 roo rasiRNA
miR-8
GFP-Aub lysate

Figure 2.6. *roo* rasiRNAs are modified at their 3´ terminus and bind Piwi and Aub.

(A) Chemical and enzymatic probing of *roo* rasiRNA structure. *roo* rasiRNA (peak #1 in Figure 2.4A) were detected by Northern hybridization. The membrane was then stripped and reprobed for miR-8. (B) *roo* rasiRNA can serve as a 3´ or a 5´ substrate for T4 RNA ligase. Filled arrowheads, 5´ ligation products; open arrowheads, 3´ ligation products. (C) *roo* rasiRNA associate with myc-tagged Piwi and (D) GFP-tagged Aub protein, but not with Ago1. I.P., immunoprecipitation; I, input; S, supernatant; B, bound.

(Figure 2.6B). Notably, the 3´ ligation reaction was less efficient for the *roo* rasiRNA than for a miRNA in the same reaction (22% versus 50% conversion to ligated product).

To determine if the rasiRNA had either a single 2´ or 3´ terminal hydroxy group, or terminal hydroxy groups at both the 2´ and 3´ positions, as do animal siRNA and miRNA, RNA from ovaries or testes was reacted with $NaIO_4$, then subject to beta-elimination. Only RNAs containing both 2´ and 3´ hydroxy groups react with $NaIO_4$; beta-elimination shortens $NaIO_4$-reacted RNA by one nucleotide, leaving a 3´ monophosphate terminus, which adds one negative charge. Consequently $NaIO_4$-reacted, b-eliminated RNAs migrate faster in polyacrylamide gel electrophoresis than the original unreacted RNA. Both *roo* (Figure 2.6A) and *Su(Ste)* (Figure 2.S5) rasiRNA lack either a 2´ or a 3´ hydroxyl groups, because they failed to react with $NaIO_4$; miRNAs in the same samples reacted with $NaIO_4$. Together, our results show that rasiRNA contain one modified and one unmodified hydroxyl. Because T4 RNA ligase can make both 3´–5´ and 2´–5´ bonds (Yang et al., 2006), we cannot currently determine the blocked position. Some plant small silencing RNAs contain a 2´-*O*-methyl modification at their 3´ terminus (Yang et al., 2006).

rasiRNA bind Piwi and Aub

Drosophila and mammalian siRNA and miRNA function through members of the Ago subfamily of Argonaute proteins, but *Su(Ste)* and *roo* rasiRNAs require at least one member of the Piwi subfamily for their function and accumulation. To ask if *roo* rasiRNAs physically associate with Piwi and Aub, we prepared ovary lysate from wild-type flies or transgenic flies expressing either

myc-tagged Piwi or GFP-tagged Aub protein (Cox et al., 2000; Harris and Macdonald, 2001), immunoprecipitated with monoclonal anti-myc, anti-GFP, or anti-Ago1 antibodies, then analyzed the supernatant and antibody-bound small RNAs by Northern blotting (Figure 2.6C, 2.6D, and Figure 2.S6). We analyzed six different *roo* rasiRNAs; all were associated with Piwi, but not Ago1, the *Drosophila* Argonaute protein typically associated with miRNAs (Okamura et al., 2004); miR-8 (Figure 2.6D and Figure 2.S6C), miR-311 and *bantam* immunoprecipitated with anti-Ago1. No rasiRNAs immunoprecipitated with the anti-myc antibody using flies lacking the myc-Piwi transgene (Figure 2.S6B).

Although *aub* mutant ovaries silence *roo* mRNA normally, they showed reduced accumulation of *roo* rasiRNA relative to *aub*/+ heterozygotes (Figure 2.4C), suggesting that *roo* rasiRNAs associate with both Piwi and Aub. We analyzed the supernatant and antibody-bound small RNAs after anti-GFP immunoprecipitation of ovary lysate from GFP-Aub transgenic flies and flies lacking the transgene. *roo* rasiRNA was recovered only when the immunoprecipitation was performed with the anti-GFP antibody in ovary lysate from GFP-Aub transgenic flies (Figure 2.6D and Figure 2.S6D). The simplest interpretation of our data is that *roo* rasiRNA physically associate with both Piwi and Aub, although it remains possible that the *roo* rasiRNAs are loaded only into Piwi, and that Aub associates with Piwi in a stable complex. The association of *roo* rasiRNA with both Piwi and Aub, and the modest reduction in *roo* silencing—by a factor of 3.3 ± 0.6 (average ± standard deviation, $n = 3$)—in *piwi* but not in *aub* mutants (Figure 2.3C), suggest that *piwi* and *aub* are partially redundant. Alternatively, *roo* silencing might proceed through Piwi alone, but the two proteins could function in the same pathway to silence selfish genetic elements.

Our data suggest that in flies rasiRNAs are produced by a mechanism that requires neither Dcr-1 nor Dcr-2, yet the patterns of rasiRNAs that direct *roo* and *Ste* silencing are as stereotyped as the distinctive siRNA population generated from the *white* hairpin by Dcr-2 (Figure 2.1A) or the unique miRNA species made from each pre-miRNA by Dcr-1. A key challenge for the future will be to determine what enzyme makes rasiRNAs and what sequence or structural features of the as-yet-unknown rasiRNA precursor lead to the accumulation of a stereotyped pattern of predominantly antisense rasiRNAs.

Materials and Methods

Fly stocks

The following fly stocks were used: *ey*Flp;;FRT82B *dcr-1*^{Q1147X}/TM3 *Sb* (Hatfield et al., 2005; Lee et al., 2004), hsFl*p*;;P[neoFRT]82B P[*ovo*$^{D1-18}$] 3R/TM3 *Sb*, *w*^{1118};*loqs*^{00791}/CyO (Forstemann et al., 2005), *y w ey*Flp;FRT42D *dcr-2*^{L811fsX} (Lee et al., 2004; Pham et al., 2004), *y w*;*r2d2*/CyO (Liu et al., 2003), *y w*;*ago2*^{414}/TM6B *Tb* (Okamura et al., 2004), *w*;*ago2*^{51B}/TM6B *Tb* (Xu et al., 2004), *w*;*piwi*^{2}/CyO (Cox et al., 1998), *aub*^{HN2} *cn bw*/CyO , and *aub*^{QC42} *cn bw*/CyO (Schupbach and Wieschaus, 1991), *w*;;*spnE*$^{100.37}$/TM3 *Sb* (kind gift of R. Lehmann), *armi*$^{72.1}$/TM3 *Sb* P[hs-*hid*] (Cook et al., 2004) and P[*nos*-Gal4-VP16]/TM3 *Sb* (Van Doren et al., 1998), P[UAS-GFP-*aub*]/TM3 *Sb* (Harris and Macdonald, 2001) and P[hs-myc-*piwi*]/CyO (Cox et al., 2000), FRT42D *dcr-2*^{L811fsX}/CyO;P[*w*-IR]/P[*w*-IR] and P[*w*-IR]/TM6B *Tb* (Kim et al., 2005; Lee and Carthew, 2003) and *cry*^{1}Y (Belloni et al., 2002; Schmidt et al., 1999).

Tiling microarrays

The expression profile for all known miRNAs was analyzed together with all possible rasiRNAs derived from *Su(Ste)* or *roo* sequence using custom synthesized microarrays containing 22 nt, T_M-normalized probes every 1 nt or 30 nt, T_M-normalized probes every 5 nt. Microarray fabrication, hybridization and data acquisition were performed at LC Sciences (Houston, TX USA). 20 µg ovary, testis, head total small RNA was isolated from wild-type or mutant flies using the mirVana kit (Ambion, Austin, TX, USA). The presence of antisense and the absence of sense rasiRNA peaks for *Su(Ste)* and *roo* were confirmed by

Northern hybridization. All microarray data is available at http://www.ncbi.nlm.nih.gov/geo/ using Gene Expression Omnibus accession number GSE4932. Figure 2.1A displays the average of three independent samples for wild-type); Figure 2.2A shows the average of seven independent samples.

Microarray data analysis

Raw data was processed as follows: background (defined as the average of signal readings on "BKG0" spots, which were spotted with chemical linker, but no probe) was subtracted from each intensity value. Intensities above exp(5) (~150) were considered significant, as described in reference (Lu et al., 2005), and intensities below 30,000 were considered non-saturated, as established with a titration of synthetic 21 nt RNAs added to the sample. Control RNAs (ribosomal RNAs, snRNAs, snoRNAs, tRNAs and synthetic RNAs added to the sample) were used for normalization, dividing each ‘mutant’ reading by the geometric average of all mutant/wt ratios for significantly, non-saturated, control RNAs. Irreproducible peaks, i.e. those that appeared in only one of a series of replicate experiments, and single-probe peaks, i.e. those peaks where the probe signal was both higher than the average signal and more than five times greater than the signal from its immediate neighbors, were removed from the datasets. For 5 nt tiling resolution, we imposed an additional criterion before removing peaks from the data set: the difference in signal intensity between the single peak and its neighbors was required to be greater than 500. When available, color-reversed datasets were averaged. Data were scaled to the highest peak in the wild-type genotype. Fourier-transform analysis parses a signal into its periodic

components and thus allows comparison of the contribution of each frequency of oscillation within a data set. To permit Fourier transform analysis, deleted peaks were replaced by the average signal intensity of the rest of the dataset to ensure a continuous wave function. Where required, the last data point was removed, so that the number of points was always an even number, a requirement of the fast Fourier transformation algorithm we used. Datasets were centered by subtracting the average from each data point, then fast Fourier transform and its modulus were computed using Igor Pro version 4.06A Carbon (WaveMetrics, Portland, OR, USA). We normalized Fourier-transform data by dividing each value by the average of the values in the interval from 5 nt to 140 nt interval, the range of values displayed in the figures, to allow direct comparison of different experiments.

RNA isolation and detection by Northern blot

Ovaries and testes were manually dissected with forceps into *Drosophila* Ringer's solution (182 mM KCl, 46 mM NaCl, 3 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5). RNA was isolated from testes or ovaries using the mirVana kit (Ambion), and then treated with RQ1 DNase (Promega, Madison, WI, USA). Total RNA was quantified by absorbance at 260 nm, and 1-10 µg of total RNA was resolved by 20% denaturing polyacrylamide/urea gel electrophoresis (National Diagnostics, Atlanta, GA, USA). 5´-^{32}P-radiolabeled RNA oligonucleotides were used as size markers. After electrophoresis, the polyacrylamide gel was transferred to Hybond N+ (Amersham-Pharmacia, Little Chalfont, UK) in 0.5x TBE by semi-dry transfer (Transblot SD, Bio-Rad) at 20 V for 1–2 h. The RNA was cross-linked to the membrane by UV irradiation (1200 µjoules/cm; Stratalinker, Stratagene, La Jolla,

CA, USA) and pre-hybridized in Church buffer (Church and Gilbert, 1984) for 1 h at 37 or 65°C. 25 pmol of RNA (Dharmacon, Lafayette, CO, USA) or DNA (IDT, Coralville, IA, USA) probe was 5´-^{32}P-radiolabeled with polynucleotide kinase (New England Biolabs, Beverly, MA, USA) and 330 µCi γ-^{32}P-ATP (7,000 µCi/mmol; New England Nuclear, Boston, MA, USA) and purified using a Sephadex G-25 spin column (Roche, Basel, Switzerland). To detect 2S rRNA, 1/50th of the ^{32}P-radiolabeled probe was diluted with unlabeled 2S rRNA probe. The ^{32}P-radiolabeled probes were hybridized in Church buffer for 4–12 h. For RNA probes, hybridization was at 65°C; for DNA probes, hybridization was at 37°C. After hybridization, membranes were washed twice with 2x SSC/0.1% (w/v) sodium dodecyl sulfate (SDS) and once with 1x SSC/0.1% (w/v) SDS for 30 min. Membranes were analyzed by phosophorimagery (Fuji, Tokyo, Japan). To strip probes, membranes were boiled in 0.1% (w/v) SDS for 5 min, then re-exposed to confirm probe removal. Probe sequences are listed in Table 2.S1.

Generation of mutant germ line clones

dcr-1 mutant female germ line clones were generated by Flp/FRT recombination as described previously (Xu and Rubin, 1993; Chou et al., 1993; Chou and Perrimon, 1992; Chou and Perrimon, 1996). hsFlp; ; P[neoFRT]82B P[*ovo*^{D1}] 3R/TM3 *Sb* males were crossed with eyFlp;FRT82B *dcr-1*^{Q1147X}/TM3 *Sb* females. After 3 days, parental flies were removed and the larvae heat shocked at 37°C for 2 h daily until eclosion. Two to three days after eclosion, females were collected and fed on yeasted media for 3 days, then ovaries manually dissected. Only ovaries containing late stage oocytes were used as *dcr-1*/*dcr-1* mutant clones for analysis.

Analysis of rasiRNA and miRNA chemical structure

15 µg of wild-type total ovary or testis small RNA was used for each reaction. RNase treatment was in 0.2 M NaCl with 1 µl 32 mg/ml RNase A (Sigma-Aldrich, St. Louis, MO, USA) for 2 h at 30°C. DNase treatment was in 1 x RQ1 DNase buffer with 5 U RQ1 DNase (Promega, Madison, WI, USA). Nucleases were inactivated by phenol/ chloroform extraction, followed by precipitation of the RNA in the presence of 20 µg glycogen carrier (Roche Diagnostics GmbH, Mannheim, Germany) at –70°C for 1 h, the precipitate was collected by centrifugation and dissolved in denaturing gel loading buffer. $NaIO_4$ reaction and β-elimination (Alefelder et al., 1998) was performed by adding 13.5 µl (20 µg) total RNA in water to 4 µl 5x borate buffer (148 mM borax,148 mM boric acid, pH 8.6) and 2.5 µl freshly dissolved 200 mM $NaIO_4$ and incubating for 10 min at room temperature. 2 µl glycerol was added to quench unreacted $NaIO_4$ and incubated for an additional 10 min at room temperature. Samples were dried by centrifugation under vacuum for 1 h at room temperature, then dissolved in 50 µl 1x borax buffer (30 mM borax and 30mM boric acid, 50 mM NaOH, pH 9.5) and incubated for 90 min at 45°C. 20 µg of glycogen was added, the RNA was precipitated at –70°C for 1 hour, the precipitate collected by centrifugation and then dissolved in the denaturing gel loading buffer. Enzymatic probing of RNA 5´ ends was performed by incubating the RNA with 2 U calf intestinal phosphatase in 1x restriction buffer 3 (CIP; New England Nuclear) for 2 h at 37°C. CIP was inactivated by phenol/ chloroform extraction, precipitated in the presence of 20 µg of glycogen carrier for 1 h at –70°C, the pellet collected by centrifugation, and then dissolved in water. Part of the CIP-treated RNA was then treated with T4

polynucleotide kinase in 10x PNK buffer (New England Nuclear) and 10 μM ATP. The reaction was inactivated by phenol/ chloroform extraction, the RNA precipitated in the presence of 20 μg of glycogen carrier for 1 h at –70°C, the precipitate collected by centrifugation, and then dissolved in denaturing gel loading buffer. The treated total small RNA was resolved by high-resolution polyacrylamide gel electrophoresis (PAGE), then analyzed by Northern blotting. A single nylon membrane was probed first to detect the *roo* or *Su(Ste)* rasiRNA, then the probe stripped and the membrane re-probed to detect miR-8 or miR-311 as an internal control.

Quantitative RT-PCR analysis

0.5 μg total RNA were used to reverse transcribe target sequences using oligo(dT) primer and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's directions. The resulting cDNA was analyzed by quantitative RT-PCR performed in a DNA Engine OPTICON 2 (MJ Research, Bio-Rad, Hercules, CA, USA) or iQ5 (Bio-Rad, Hercules, CA, USA) instrument using the QuantiTect SYBR Green PCR kit (QIAGEN) according to manufacturer's instructions. Relative steady-state mRNA levels were determined from the threshold cycle for amplification using the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen, 2001). Table 2.S1 lists the PCR primer sequences, and Figure 2.S7 indicates their position on each selfish genetic element. Control experiments measuring the change in ΔC_T with template dilution demonstrated that the efficiencies of amplification of the target genes and the control (*act5C* or *rp49*) were approximately the same.

Immunoprecipitation

myc-*piwi* transgenic flies (Cox et al., 2000) were heat shocked at 37°C for 1 h every 12 h for 3 days before dissecting the ovaries. *aub*/CyO; UAS-GFP-*aub*/*nos*-Gal4-VP16 and *aub*/*aub*; UAS-GFP-*aub*/*nos*-Gal4-VP16 transgenic flies (Harris and Macdonald, 2001) were generated by crossing *aub*/CyO; UAS-GFP-*aub* with *aub*/CyO; *nos*-Gal4-VP16 flies. *aub*/Cyo; UAS-GFP-*aub*/*nos*-Gal4-VP16 ovaries composed the majority of the tissue isolated. The *nanos* (*nos*) promoter drives expression of GFP-Aub early in spermatogenesis, in the tip of the testes, and is sufficient to rescue the *Ste* silencing defect in *aub* mutant males. Quantitative western blotting (Figure 2.S6A) established that the ratio of myc-Piwi to endogenous Piwi in the myc-*piwi*, *piwi*$^{+}$/myc-*piwi*, *piwi*$^{+}$ flies used was 0.95 ± 0.15 to 1; the ratio of GFP-Aub to endogenous Aub in the *aub*/CyO; UAS-GFP-*aub*/*nos*-Gal4-VP16 flies was 3.1 ± 0.8 to 1. Since the myc-Piwi-expressing ovaries contained two wild-type and two transgenic myc-*piwi* genes and the ovaries expressing GFP-Aub contained one mutant *aub* gene and no more than one wild-type *aub* gene, each genotype expressed about twice as much total Piwi or Aub protein as wild-type. For immunoprecipitation, 30 µl Protein A/G agarose (Calbiochem, San Diego, CA, USA) were incubated with 20 µl anti-Ago1 (1B8)(Okamura et al., 2004), 5 µl anti-GFP (B-2, Santa Cruz Biotechnology, Santa Cruz, CA, USA, or 3E6, Qbiogene, Montreal, Quebec, Canada), or 5 µl anti-myc (9E10, Sigma, St. Louis, Missouri, USA) monoclonal antibody for 2 h at 4°C. Next, the agarose beads were washed twice with lysis buffer (30 mM HEPES-KOH, pH 7.4, 100 mM KOAc, 2 mM $Mg(OAc)_2$, 5 mM DTT, 0.5% [v/v] NP-40). Subsequently, 50 µl of ovary lysate (20 µg/µl total protein) was added and the mixture agitated gently for 4 h at 4°C. The supernatant (50 µl) and

antibody bound material were separated by centrifugation at 13,000 rpm for 2 min at 4°C, and the antibody-bound beads washed four times then resuspended in 50 µl lysis buffer. 2 µl of Input (I), Supernatant (S), or Bound (B) samples were subject to Western Blotting. The remaining samples were treated with Proteinase K (1 mg/ml proteinase K, 100 mM Tris–Cl, pH 7.5, 12.5 mM EDTA pH 8.0, 150 mM NaCl, 1% [w/v] SDS) for 1 h at 65°C, extracted with an equal volume of phenol/chloroform, and then precipitated with 3 volumes of 100% ethanol in the presence of 20 µg glycogen carrier. The precipitate was washed once with 80% ethanol and then dissolved in nuclease-free water. 15% of each RNA sample was resolved in a 20% denaturing polyacrylamide gel and analyzed by Northern hybridization.

Western blotting

Samples reserved for Western blotting were added to 20 µl SDS sample buffer (10 mM Tris-Cl, pH 6.8, 2% [w/v] SDS, 100 mM DTT, 10% [v/v] glycerol), boiled for 10 min, and then resolved by SDS-PAGE and transferred to a PVDF membrane (Immobilon-P, Millipore). To detect Ago1, Aubergine, or Piwi, the membrane was incubated for 2 hours at room temperature with anti-Ago1 (1B8), anti-Aubergine (ab17724, Abcam PLC, Cambridge, UK), or anti-Piwi (ab5207, Abcam PLC, Cambridge, UK) antibody diluted 1:1000 into TBST-milk (25 mM Tris-Cl, pH 7.4, 3.0 mM KCl, 140 mM NaCl, 0.05% [v/v] Tween-20, 5% [w/v] non-fat dry milk). Subsequently, the membrane was washed 3 times with TBST and incubated 1 h at room temperature with either goat anti-mouse (Ago1) or goat anti-rabbit (Aubergine and Piwi) HRP-conjugated secondary antibody diluted

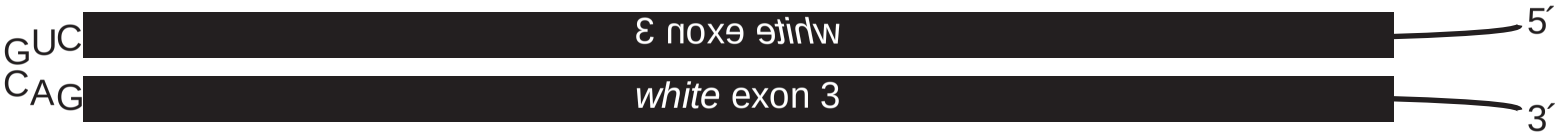
1:1000 into TBST-milk, and then developed with SuperSignal West Dura Extended Duration Substrate (Pierce, Rockford, IL, USA).

Acknowledgments

We thank Tingting Du and Alicia Boucher for assistance with fly husbandry, and J. Birchler, R. Carthew, F. Gao, R. Lehmann, H. Lin, P. MacDonald, M. and H. Siomi, D. Smith, E. Sontheimer, H. Ruohola-Baker, and M. Van Doren for fly stocks. PDZ is a W.M. Keck Foundation Young Scholar in Medical Research. This work was supported in part by grants from the National Institutes of Health to PDZ (GM62862 and GM65236) and post-doctoral fellowships from the EMBO and the Human Frontier Science Program to HS.

Figure 2.S1.

A



B

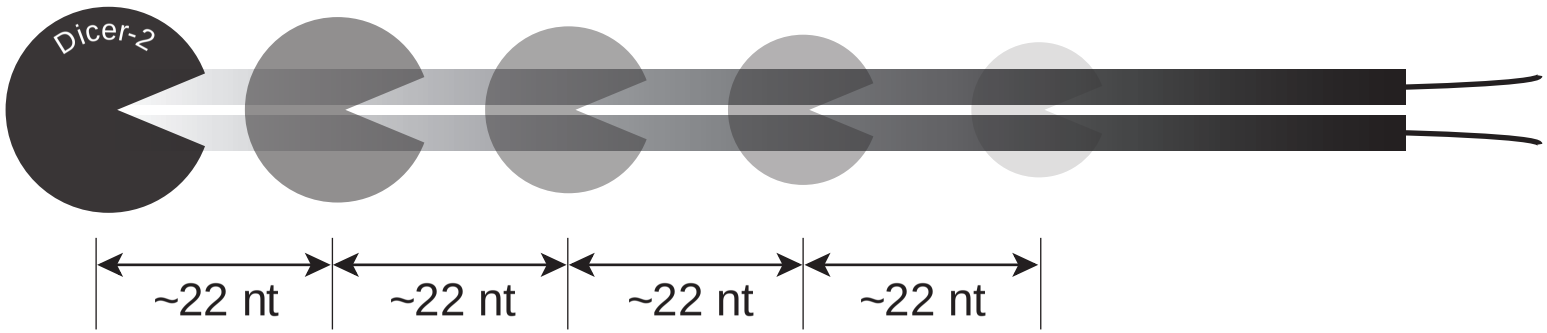


Figure 2.S1. A model of phasing *white* exon 3 hairpin RNA by Dicer-2.
(A) The structure of the *white* exon 3 hairpin trigger RNA. (B) A model describing the data in Figure 2.1, which suggest that Dicer-2 initiates dsRNA processing from the loop end of the *white* exon 3 hairpin, then moves in ~22 nt steps to create a phased series of siRNAs.

Figure 2.S2.

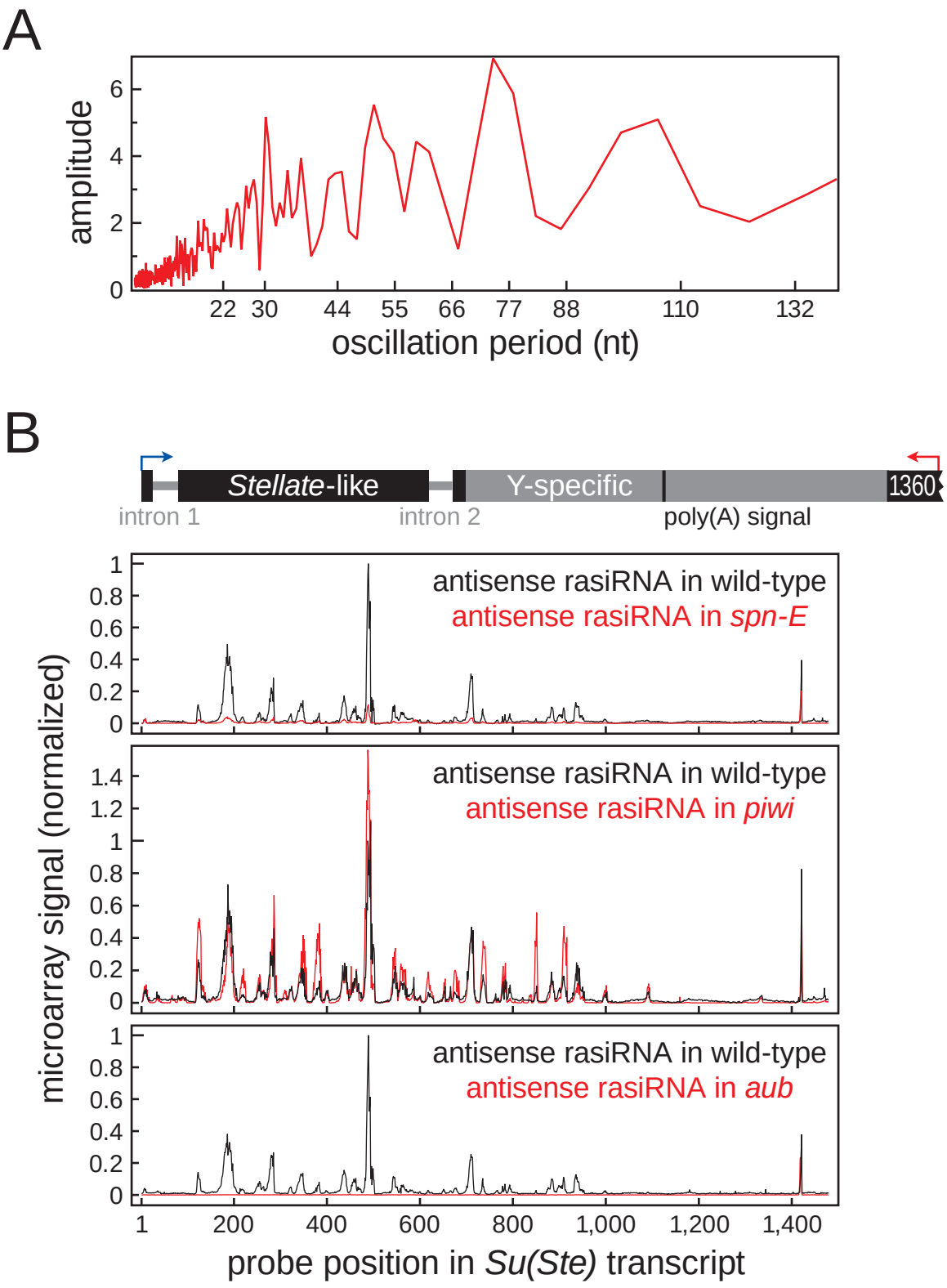

A
amplitude
0
2
4
6
22
30
44
55
66
77
88
110
132
oscillation period (nt)
B
Stellate-like
Y-specific
1360
intron 1
intron 2
poly(A) signal
antisense rasiRNA in wild-type
antisense rasiRNA in spn-E
antisense rasiRNA in wild-type
antisense rasiRNA in piwi
antisense rasiRNA in wild-type
antisense rasiRNA in aub
microarray signal (normalized)
1
200
400
600
800
1,000
1,200
1,400
probe position in Su(Ste) transcript

Figure 2.S2. Characterization of *Su(Ste)* rasiRNAs from testes.

(A) The microarray data for sense and antisense *Su(Ste)* rasiRNAs in wild-type were examined for phasing using Fourier-transform analysis. No oscillations corresponding to the length of either a typical siRNA (21–23 nt) or rasiRNA (24–30 nt) were detected. The amplitude scale here is the same as that in Figure 2.1B. (B) Microarray analysis of *Su(Ste)* antisense rasiRNAs in testes from wild-type and *spn-E*, *piwi*, or *aub* mutant flies. Above the array data, the schematic indicates the posistion of the rasiRNAs within a single *Su(Ste)* locus.

Figure 2.S3.

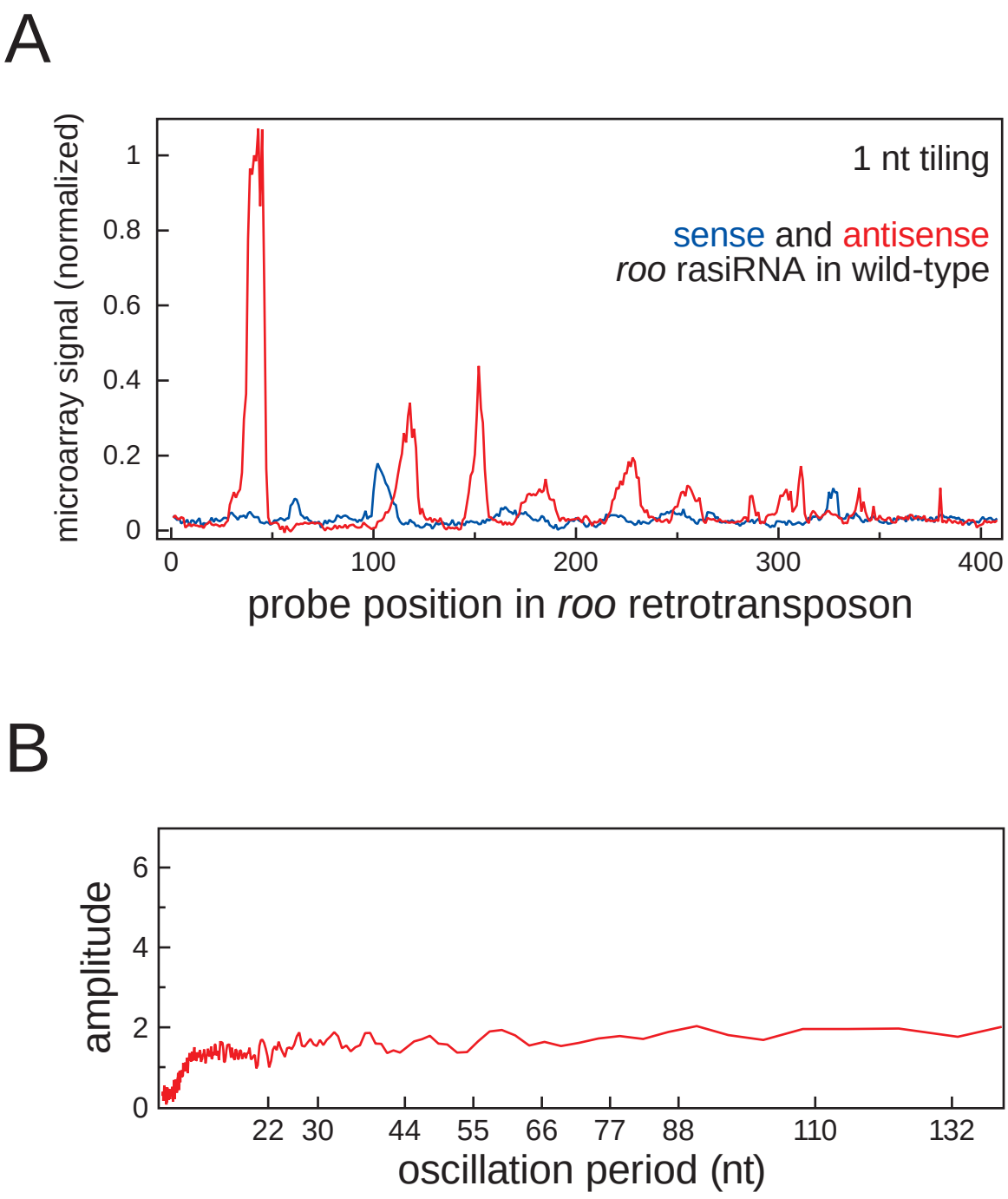

A
microarray signal (normalized)
1
0.8
0.6
0.4
0.2
0
1 nt tiling
sense and antisense
roo rasiRNA in wild-type
0
100
200
300
400
probe position in roo retrotransposon
B
amplitude
6
4
2
0
22
30
44
55
66
77
88
110
132
oscillation period (nt)

Figure 2.S3. Characterization of *roo* rasiRNAs from ovaries.

(A) High resolution tiling microarray analysis across the *roo* LTR. The absolute peak heights in this experiment have been normalized so that they can be compared directly with those in Figure 2.4A and 2.4B. (B) Fourier-transform analysis of the combined sense and antisense *roo* rasiRNA data for wild-type ovaries. No oscillations corresponding to the length of either a typical siRNA (21-23 nt) or rasiRNA (24-30 nt) were detected. The amplitude scale here is the same as that in Figure 2.1B and Figure 2.S2.

Figure 2.S4.

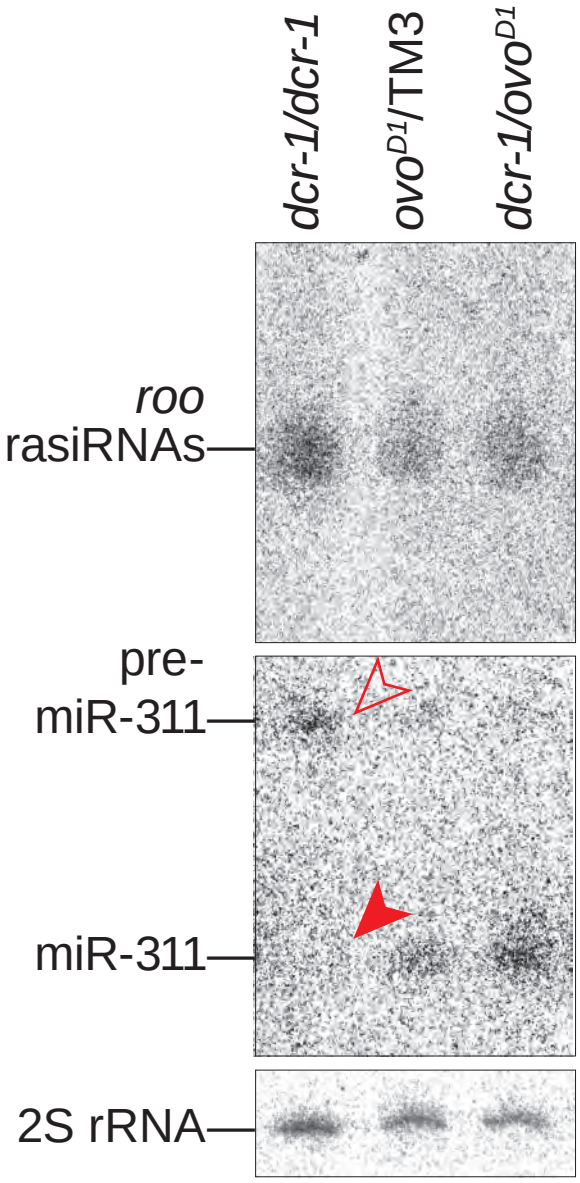

dcr-1/dcr-1
ovo^D1/TM3
dcr-1/ovo^D1
roo rasiRNAs
pre-miR-311
miR-311
2S rRNA

Figure 2.S4. Northern analysis of *roo* rasiRNAs in *dcr-1* mutant ovaries. Northern analysis of mitotic recombinant *dcr-1*^{Q1147X} homozygous mutant ovaries and non-recombinant controls. *roo* rasiRNA were detected using a mixture of hybridization probes for five different rasiRNA peaks. The open red arrowhead highlights pre-miR-311 accumulation in the *dcr-1* mutant tissue; the filled red arrowhead highlights loss of mature miR-311. The data here correspond to an experiment independent from that in Figure 2.5. A third independent experiment that produced essentially identical results is not shown.

Figure 2.S5.

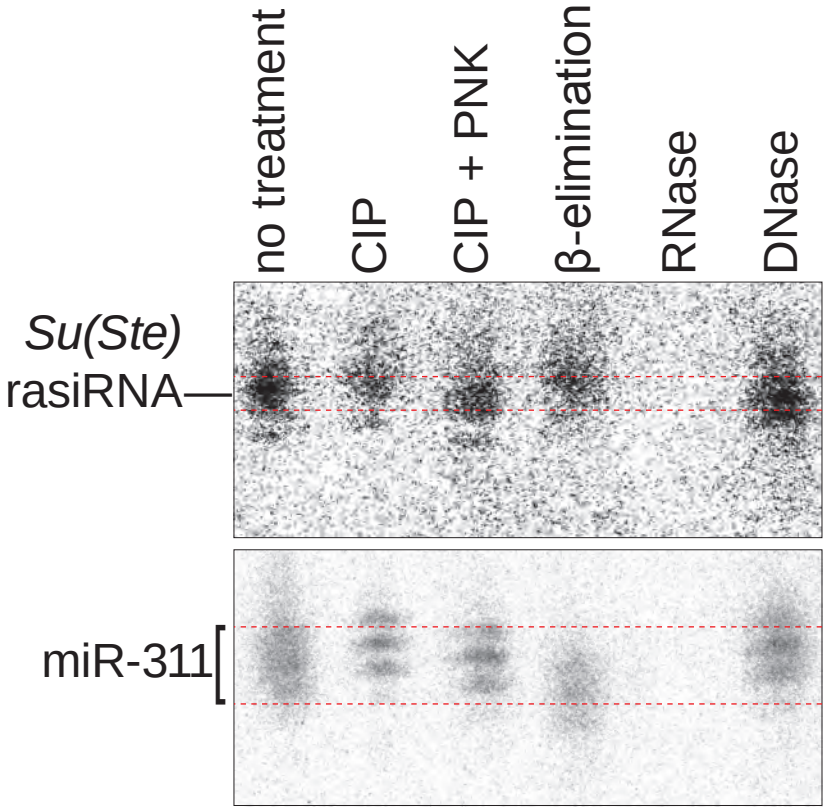

no treatment
CIP
CIP + PNK
β-elimination
RNase
DNase
Su(Ste)
rasiRNA
miR-311

Figure 2.S5. *Su(Ste)* rasiRNAs are modified at their 3´ terminus.

The major *Su(Ste)* rasiRNA contains a 5´ monophosphate terminus, but lacks the 2´,3´ hydroxy terminus characteristic of siRNA and miRNA. The major *Su(Ste)* rasiRNA was detected by Northern hybridization. The membrane was then stripped and reprobed for miR-311. As expected, miR-311 contains a 5´ phosphate and a 2´,3´ hydroxy terminus.

Figure 2.S6.

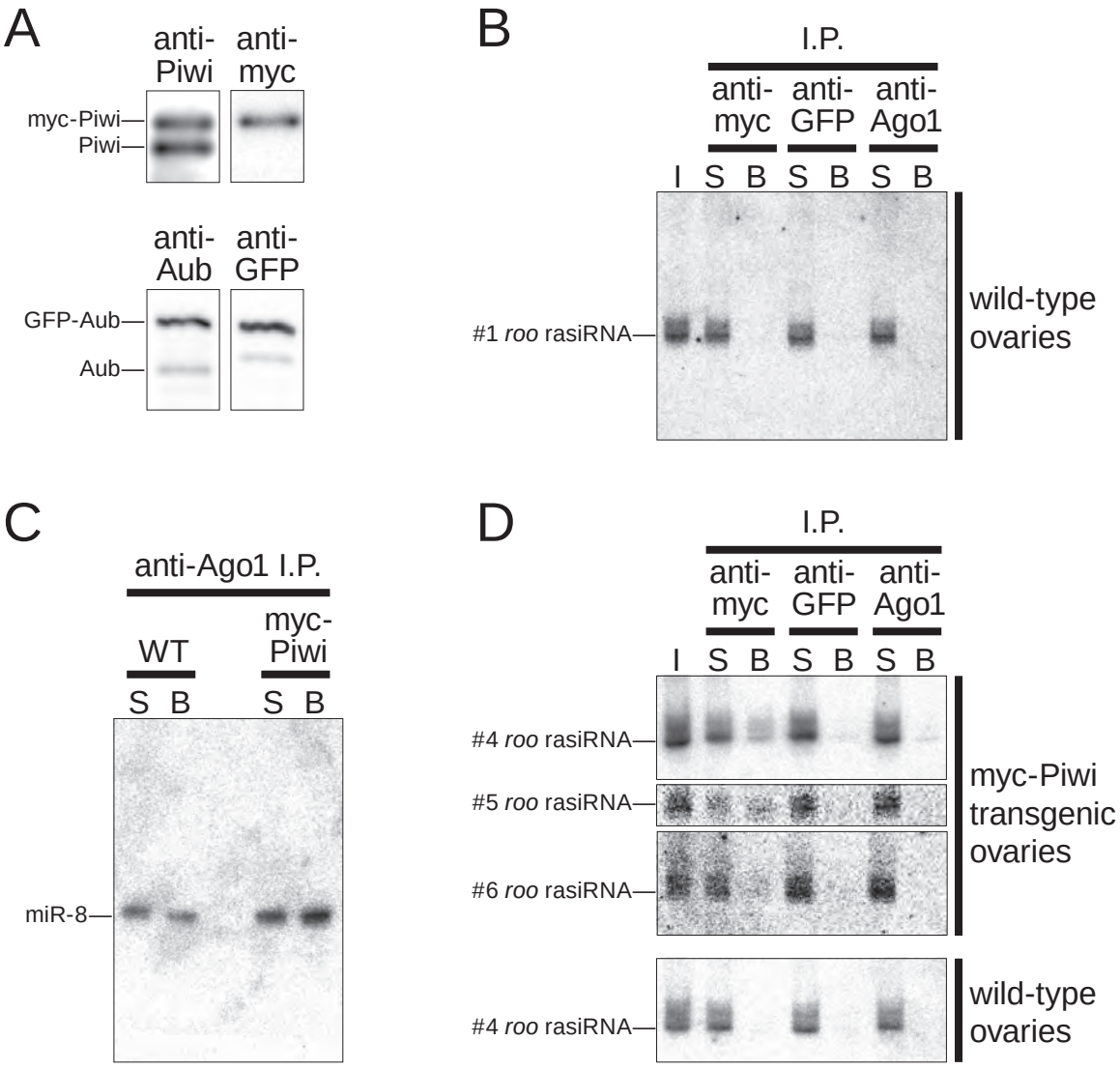

A
anti-Piwi
anti-myc
myc-Piwi
Piwi
anti-Aub
anti-GFP
GFP-Aub
Aub
B
I.P.
anti-myc
anti-GFP
anti-Ago1
I S B S B S B
#1 roo rasiRNA
wild-type ovaries
C
anti-Ago1 I.P.
WT
myc-Piwi
S B S B
miR-8
D
I.P.
anti-myc
anti-GFP
anti-Ago1
I S B S B S B
#4 roo rasiRNA
#5 roo rasiRNA
#6 roo rasiRNA
myc-Piwi transgenic ovaries
#4 roo rasiRNA
wild-type ovaries

Figure 2.S6. *roo* rasiRNAs specifically associate with Piwi and Aub.
(A) Transgenic myc-tagged and endogenous Piwi or transgenic GFP-tagged and endogenous Aub proteins were resolved by high resolution SDS-PAGE and measured by quantitative Western blotting. Representative data from the analysis is shown and expression values are given in the text. The blots were probed first with anti-Piwi or anti-Aub, then stripped and reprobed with anti-myc or anti-GFP antibodies. (B) *roo* rasiRNA were not detected by Northern analysis in the immunoprecipitates when anti-myc, anti-GFP, or anti-Ago1 monoclonal antibodies were used with ovary lysate prepared from wild-type flies carrying neither a myc-Piwi- nor a GFP-Aub-expressing transgene. (C) miR-8 was detected associated with Ago1 in both wild-type (WT) and myc-Piwi-expressing transgenic ovary lysate. (D) Three *roo* rasiRNAs identified by microarray analysis (Figure 2.4A) co-immunoprecipitated with myc-Piwi in ovary lysate isolated from myc-Piwi-expressing transgenic flies but not wild-type ovary lysate. I.P., immunoprecipitation; I, input; S, supernatant; B, bound (i.e., the immunoprecipitate).

Figure 2.S7.

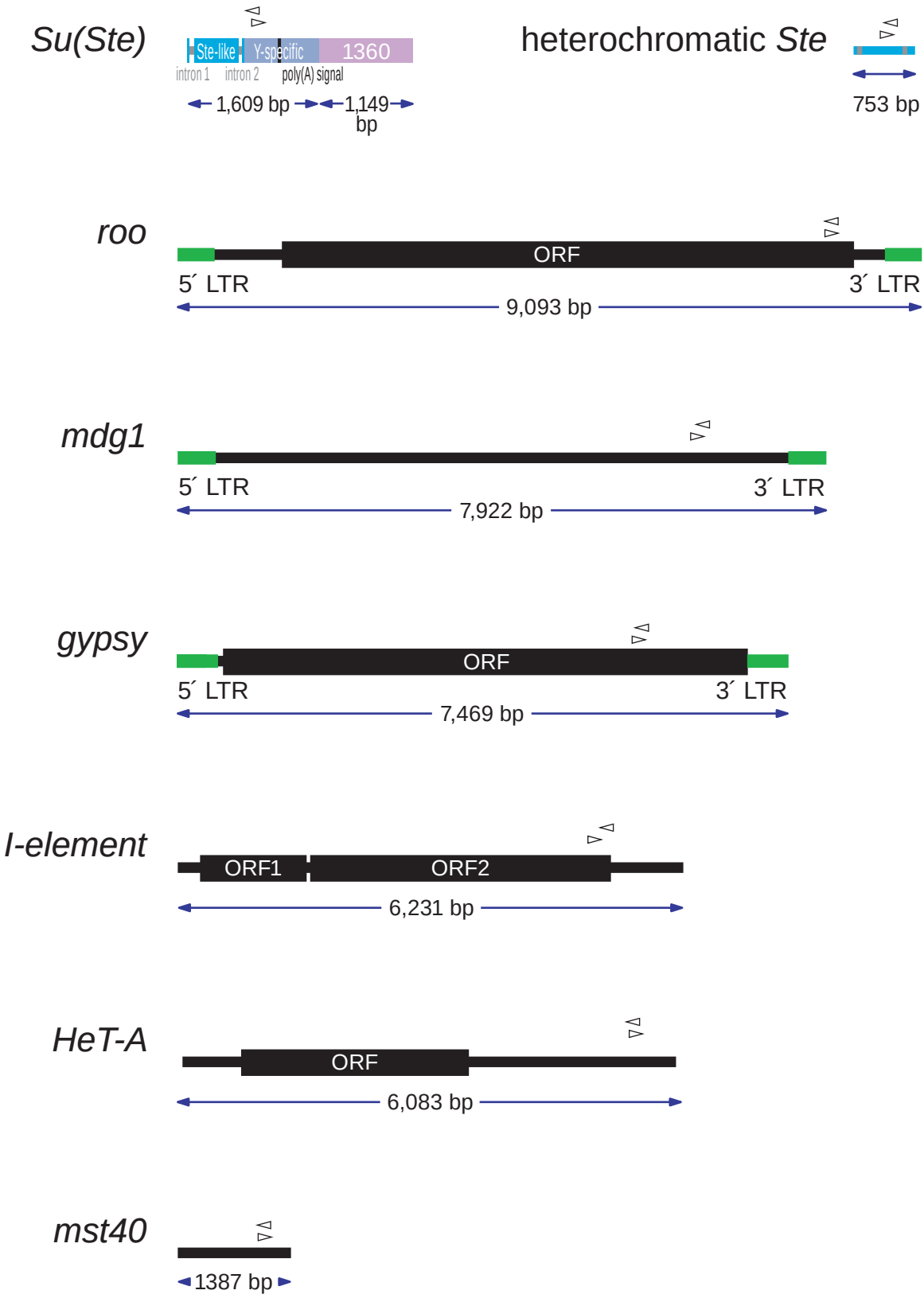

Su(Ste)
Ste-like
Y-specific
1360
intron 1
intron 2
poly(A) signal
1,609 bp
1,149 bp
heterochromatic Ste
753 bp
roo
ORF
5´ LTR
3´ LTR
9,093 bp
mdg1
5´ LTR
3´ LTR
7,922 bp
gypsy
ORF
5´ LTR
3´ LTR
7,469 bp
I-element
ORF1
ORF2
6,231 bp
HeT-A
ORF
6,083 bp
mst40
1387 bp

Figure 2.S7. Selfish genetic elements examined in this study.

Open arrowheads show the position of the 5´ ends of the quantitative RT-PCR primers. All elements are drawn at the same scale. The primers for heterochromatic Ste also amplify RNA from the euchromatic Ste loci and thus monitor total Ste expression.

Table 2.S1. Probes and Primers.

Probes for Northern hybridization

Small RNAs	DNA or *RNA* sequences
Su(Ste) antisense rasiRNA	5´-*CCACGCUGUAAAAACUACUUC*-3´
#1 *roo* antisense rasiRNA	5´-TGGGCTCCGTTCATATCTTATG-3´
#1 *roo* sense rasiRNA	5´-CATAAGATATGAACGGAGCCCA-3´
#3 *roo* antisense rasiRNA	5´-TGAGAGTTCGCTATTCGAAGAA-3´
#3 *roo* sense rasiRNA	5´-TTCTTCGAATAGCGAACTCTCA-3´
#4 *roo* antisense rasiRNA	5´-TCGACTCAGTGGCACAATAAAT-3´
#4 *roo* sense rasiRNA	5´-ATTTATTGTGCCACTGAGTCGA-3´
#5 *roo* antisense rasiRNA	5´-CGACGGTTGTTAAGGCGGAGGA-3´
#5 *roo* sense rasiRNA	5´-TCCTCCGCCTTAACAACCGTCG-3´
#6 *roo* antisense rasiRNA	5´-GCTCGGCTCTGGAGAAAATTAT-3´
#6 *roo* sense rasiRNA	5´-ATAATTTTCTCCAGAGCCGAGC-3´
#7 roo antisense rasiRNA	5´-TCTGAGGCATCCGTTTGGTAAA-3´
#7 *roo* sense rasiRNA	5´-TTTACCAAACGGATGCCTCAGA-3´
miR-8	5´-GACATCTTTACCTGACAGTATTA-3´
miR-311	5´-*UCAGGCCGGUGAAUGUGCAAUA*-3´
bantam	5´-CAGCTTTCAAAATGATCTCACT-3´
2S rRNA	5´-TACAACCCTCAACCATATGTAGTCCAAGCA-3´

Primers for quantitative PCR

Selfish genetic elements	Forward, Reverse
Ste	5´-GTCAGAGCGTGGCGTGATTGC-3´, 5´-GGCCCGAACATCGCTCCGTCC-3´
Su(Ste)	5´-TTCCGAAGTCAAGCGCTTCAATG-3´, 5´-GGAATCTGTTTAATTGCAACAAC-3´
roo	5´-CGTCTGCAATGTACTGGCTCT-3´, 5´-CGGCACTCCACTAACTTCTCC-3´
HeT-A	5´-CGCGCGGAACCCATCTTCAGA-3´, 5´-CGCCGCAGTCGTTTGGTGAGT-3´
I-element	5´-GACCAAATAAAAATAATACGACTTC-3´, 5´-AACTAATTGCTGGCTTGTTATG-3´
mdg1	5´-AACAGAAACGCCAGCAACAGC-3´, 5´-CGTTCCCATGTCCGTTGTGAT-3´
mst40	5´-TTCCCTAAGTCCCTCGCAATC-3´, 5´-GCACAGTTTCCGCAGCCATCA-3´

CHAPTER III: THE *DROSOPHILA* RNA METHYLTRANSFERASE, DMHEN1, MODIFIES BOTH PIRNAS AND SINGLE-STRANDED SIRNAS IN RISC AND IS REQUIRED FOR SILENCING SELFISH GENETIC ELEMENTS IN THE GERM LINE

Preface

The work presented in the following chapter was a collaborative effort. I carried out experiments in Figure 3.1, 3.3B, 3.3C, 3.S1, and 3.S2. Experiments in Figure 3.2 were joint efforts of Christian Matranga and me. Experiments in Figure 3.3A, 3.3D, and 3.S3 were collaborations between Michael D. Horwich and me. Experiments in the remaining figures were carried out by Christian Matranga (Figure 3.4C, 3.5, and 3.S7), Michael D. Horwich (Figure 3.4A, 3.4B, and 3.S6), Peng Wang (Figure 3.S4), and Gwen Farley (Figure 3.S5). Vasily V. Vagin provided critical help for reagents and discussion.

Summary

In plants and animals, small silencing RNAs guide Argonaute proteins to repress gene expression by a set of related mechanisms collectively called RNA silencing pathways (Zamore and Haley, 2005; Meister and Tuschl, 2004). In the RNA interference (RNAi) pathway (Fire et al., 1998), small interfering RNAs (siRNAs) defend cells from invasion by foreign nucleic acids, such as those produced by viruses. In contrast, microRNAs (miRNAs) sculpt expression of endogenous mRNAs (Bartel and Chen, 2004). In animals, a third class of small RNAs, Piwi-interacting RNAs (piRNAs), defends the genome from molecular parasites such as transposons (Aravin et al., 2003; Aravin et al., 2007; Aravin et al., 2001; Vagin et al., 2006; Brennecke et al., 2007). Here, we report that piRNAs in flies contain a 2´-*O*-methyl group on their 3´ termini, a modification previously reported for miRNAs and siRNAs in plants (Yang et al., 2006) and piRNAs in mice (Kirino and Mourelatos, 2007a; Ohara et al., 2007) and rats (Houwing et al., 2007). In plants, small RNA methylation is catalyzed by the HEN1 protein and serves to protect miRNAs and siRNAs from degradation (Li et al., 2005; Yang et al., 2006; Yu et al., 2005). We find that the *Drosophila* homolog of HEN1, DmHen1, methylates the termini of both siRNAs and piRNAs. In the absence of DmHen1, both the length and abundance of piRNAs are decreased, and piRNA function is perturbed. Unlike plant HEN1, the *Drosophila* enzyme acts on single-stranded rather than duplex small RNAs, explaining how it can use as substrates both siRNAs—which derive from double-stranded precursors—and piRNAs—which do not (Houwing et al., 2007; Vagin et al., 2006). 2´-*O*-methylation of siRNAs may be the final step in assembly of the

RNAi-enzyme complex, RISC, occurring after an Argonaute-bound siRNA duplex is converted to single-stranded RNA.

Results and Discussion

Drosophila piRNAs are 2´-*O*-methylated at their 3´ termini

In flies, both piRNAs (also known as repeat-associated siRNAs, rasiRNAs) and siRNAs, but not miRNAs, are modified at their 3´ termini (Pelisson et al., 2007; Vagin et al., 2006). We selectively labeled (Figure 3.S1) the terminal nucleotide of *Drosophila melanogaster* 0–2 h embryo and mouse and bull testicular piRNAs. The resulting ^{32}P-radiolabeled nucleoside 2´ or 3´-monophosphates were resolved by two-dimensional thin-layer chromatography (2D TLC) using a solvent system that can resolve nucleoside 2´ monophosphates, nucleoside 3´ monophosphates, and 2´-*O*-methyl nucleoside 3´ monophosphates (Figure 3.S2). Modified nucleoside monophosphates derived from the 3´ termini of piRNAs were identified by comparison to modified and unmodified nucleoside 2´ and 3´ monophosphate standards (Figure 3.1A). The terminal nucleotide of the piRNAs of all three animals co-migrated with 2´-*O*-methyl nucleoside 3´ monophosphate standards, but not with any unmodified nucleoside monophosphate standard. Since mouse piRNAs were previously shown to contain 2´-*O*-methyl modified 3´ termini using both mass spectrometry (Ohara et al., 2007) and a 2D TLC system (Kirino and Mourelatos, 2007a) distinct from ours, we conclude that *Drosophila* and bull piRNAs also contain a 2´-*O*-methyl group at their 3´ termini.

Drosophila siRNAs are 2´-*O*-methylated at their 3´ termini

siRNAs, like piRNAs, are modified at their 3´termini (Pelisson et al., 2007; Vagin et al., 2006); do siRNAs have a 2´-*O*-methyl group at the 3´ end? To test if

Figure 3.1.

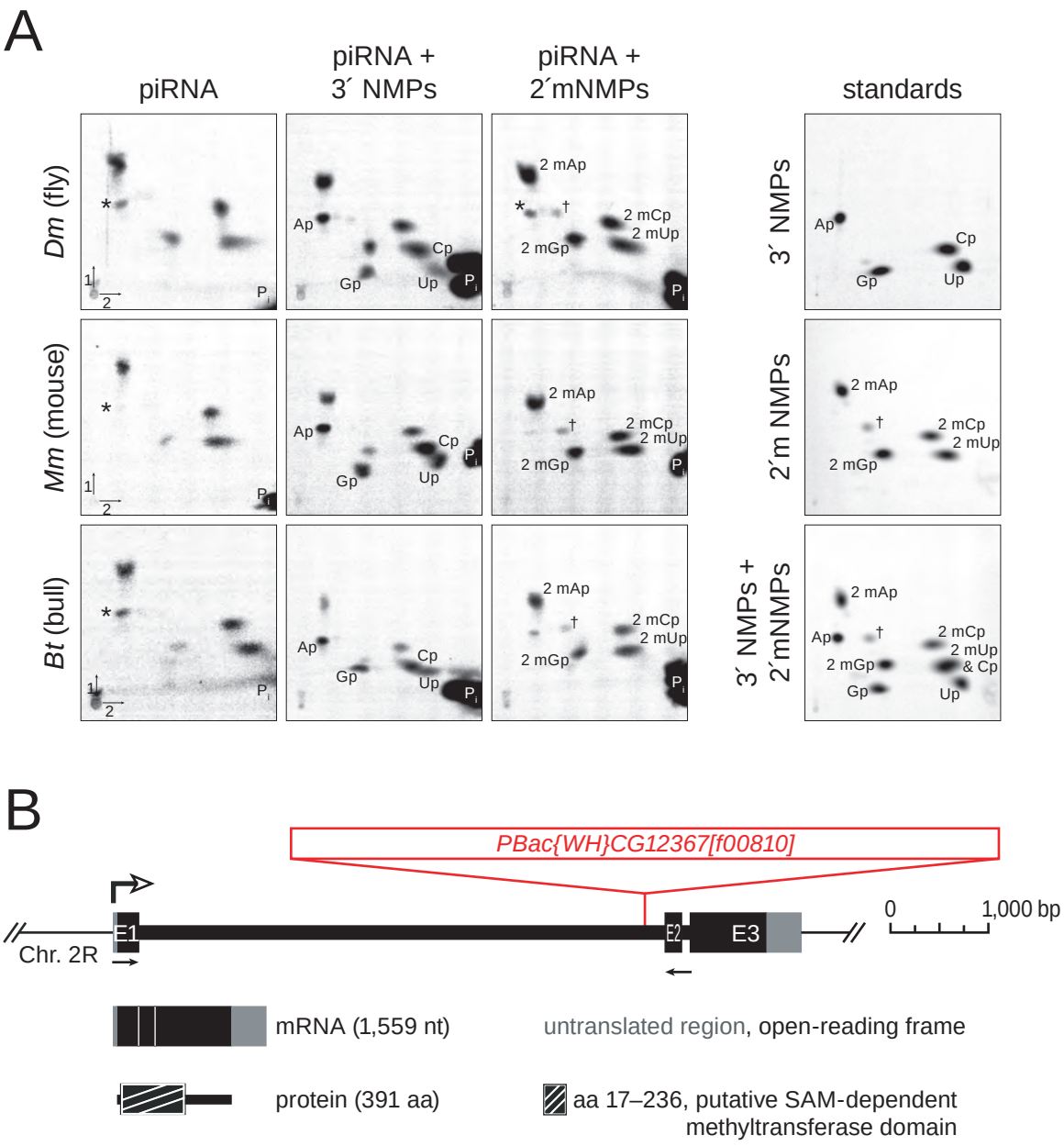

A
piRNA
piRNA + 3´ NMPs
piRNA + 2´mNMPs
standards
Dm (fly)
Mm (mouse)
Bt (bull)
3´ NMPs
2´m NMPs
3´ NMPs + 2´mNMPs
Ap
Cp
Gp
Up
2 mAp
2 mCp
2 mUp
2 mGp
& Cp
B
PBac{WH}CG12367[f00810]
Chr. 2R
E1
E2
E3
0
1,000 bp
mRNA (1,559 nt)
untranslated region, open-reading frame
protein (391 aa)
aa 17–236, putative SAM-dependent methyltransferase domain

Figure 3.1. 2´-*O*-methylation of piRNAs in *Drosophila*.

(A) *Drosophila* piRNAs are 2´-*O*-methylated at their 3´ termini. The modified nucleotides on the 3' termini of piRNAs from 0–2 h fly embryos, and mouse and bull testes were selectively ^{32}P radiolabeled. The radiolabeled modified mononucleotides from each species were resolved by 2D-TLC individually (piRNA), with ^{32}P-radiolabeled 3' mononucleotide standards (piRNA+3'NMPs), and with ^{32}P-radiolabeled 2'-*O*-methyl, 3' phosphate mononucleotide standards (piRNA+2'mNMPs). The modified nucleotides from the piRNA from all three animals co-migrated with 2'mNMPs standards, but not with 3'NMPs standards. Ap, 3'AMP; Gp, 3´GMP; Cp, 3´CMP; Up, 3'UMP; 2´mAp, 2'-*O*-methyl AMP; 2´mGp, 2'-*O*-methyl GMP; 2´mCp, 2'-*O*-methyl CMP; 2´mUp, 2'-*O*-methyl UMP; and P_i, phosphate. The asterisk marks a contaminant, likely 3' AMP, present in the [5´ ^{32}P] cytidine 5´,3´ bis-phosphate used to radiolabel the piRNA. † marks a contaminant present in the 2'-*O*-methyl, 3' phosphate mononucleotide standards. (B) *Drosophila hen1* gene (CG12367), mRNA, and protein. The piggyBac transposon, PBAC{WH}CG12367[f00810], is inserted 207 bp upstream of the second exon. The open arrow indicates the predicted start of transcription. The closed arrows denote the position of the qRT-PCR primers used in Figure 3.3A. The first intron of *hen1* contains another gene, CG8878, transcribed in the opposite direction, whose expression is unaltered by the piggyBac insertion (Figure 3.S3).

the ultimate nucleotide of siRNAs is 2´-*O*-methylated, we selectively labeled synthetic *luciferase* (*luc*) siRNAs that were assembled in *Drosophila* 0-2 hour embryo lysate for 4 hours. The guide strand of the *luc* siRNA was 5´radiolabeled and contained an ultimate 3´ adenosine. After preparation of total RNA and extraction of the radiolabeled RNA from a polyacrylamide gel, the guide strand of the siRNA was subjected to our selective labeling procedure and subsequent TLC analysis (Figure 3.2). In a mock experiment, total RNA was prepared from embryo lysate without exogenous siRNA**.** We extracted RNA that migrated to the same position as the radiolabeled guide strand of the luc siRNA in our polyacrylamide gel system. We found that the terminal nucleotide of the siRNAs co- migrated with 2´-*O*-methyl adenosine 3´ monophosphate standards, but not with any other modified or unmodified nucleoside monophosphate standard. The signal for 2´-*O*-methyl adenosine 3´ monophosphate was significantly high above background found in the mock TLC. We conclude that, like piRNAs, siRNAs are modified with a 3´terminal 2´-*O*-methyl in *Drosophila*.

DmHen1 is required for piRNA modification in vivo

In *Arabidopsis*, the RNA methyltransferase, HEN1, modifies the terminal 2´ hydroxyl group of small silencing RNAs. In *Drosophila*, predicted gene CG12367, whose 1559 nucleotide mRNA encodes a 391 amino acid protein with a 220 amino acid evolutionarily conserved methyltransferase domain (Tkaczuk et al., 2006), most closely resembles *Arabidopsis* HEN1 (Figure 3.1B) (Park et al., 2002). For simplicity, we call this gene *Drosophila melanogaster* (Dm) *hen1*. When homozygous, a piggyBac transposon insertion (PBAC{WH}CG12367[f00810]) within the first intron of the fly *hen1* gene reduces

Figure 3.2.

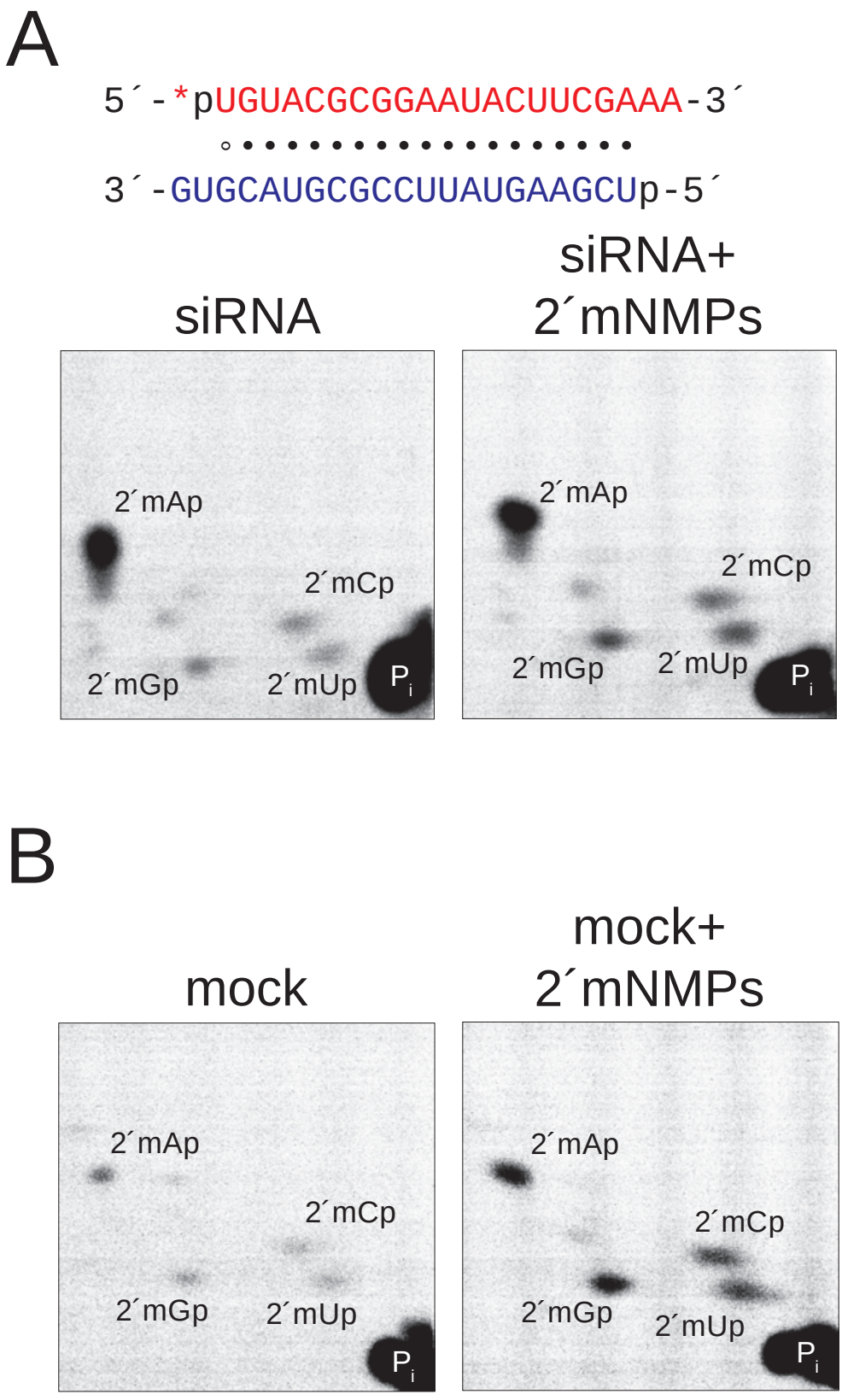

A
5´ - *pUGUACGCGGAAUACUUCGAAA - 3´
3´ - GUGCAUGCGCCUUAUGAAGCUp - 5´
siRNA
siRNA+
2´mNMPs
2´mAp
2´mCp
2´mGp
2´mUp
Pi
2´mAp
2´mCp
2´mGp
2´mUp
Pi
B
mock
mock+
2´mNMPs
2´mAp
2´mCp
2´mGp
2´mUp
Pi
2´mAp
2´mCp
2´mGp
2´mUp
Pi

Figure 3.2. 2´-*O*-methylation of siRNAs in *Drosophila*.

(A) *Drosophila* siRNAs are 2´-*O*-methylated at their 3´ termini. The modified nucleotides on the 3’ termini of the guide strand of the *Luciferase* siRNAs assembled in 0–2 h fly embryo were selectively 32P radiolabeled. The radiolabeled modified mononucleotides were resolved by 2D-TLC (siRNA), with 32P-radiolabeled 2’-*O*-methyl, 3’ phosphate mononucleotide standards (siRNA+2’mNMPs). The modified 3´ ultimate nucleotide from the siRNA—a 2´-*O*-methyl adenosine 3´ monophosphate—co-migrated with 2’mAMPs standard. (B) Total RNA from 0-2 hr embryo lysate was resolved on a polyacrylamide gel. RNA that comigrated with the 21nt guide strand used in our experimental sample (in A) was collected, subjected to selective radiolabeling, and 2D-TLC.

the accumulation of *hen1* mRNA by 1,000-fold in testes and by more than 40,000-fold in ovaries (Figure 3.3A) and can therefore be considered a null mutation, which we refer to as *hen1*^{f00810}.

We examined the 3´ termini of two types of highly abundant piRNAs in the germ line of flies heterozygous or homozygous for *hen1*^{f00810}. In testes, the *Suppressor of Stellate* (*Su(Ste)*) locus produces 24–27 nt rasiRNAs, a subclass of piRNAs that directs silencing of the selfish genetic element *Stellate*. *Su(Ste)* rasiRNAs, like other *Drosophila* piRNAs, are modified at their 3´ termini and therefore do not react with $NaIO_4$ (Vagin et al., 2006). In contrast, *Su(Ste)* rasiRNAs from *hen1*^{f00810}/*hen1*^{f00810} mutant testes reacted with $NaIO_4$ and could therefore be beta-eliminated to remove the last nucleotide of the RNA, increasing their gel mobility (Figure 3.3B) and indicating that in the absence of DmHen1 protein, they are not modified. Similarly, rasiRNAs that guide silencing of *roo*, the most abundant retrotransposon in *Drosophila melanogaster*, were not modified in *hen1*^{f00810} homozygous ovaries (Figure 3.3C). The *Su(Ste)* and *roo* rasiRNAs were also shorter in the *hen1*^{f00810} homozygotes. In contrast, the length and amount of miR-8, which is expressed in both the male and female germ line, was unaltered in *hen1*^{f00810} homozygotes. For both *Su(Ste)* and *roo*, rasiRNAs were on average shorter and less modified even in *hen1*^{f00810} heterozygotes, compared to wild-type, suggesting that the abundance of DmHen1 protein limits the stability or production of piRNAs in flies.

DmHen1 is required for piRNA function in vivo

Modification of the termini of *Drosophila* piRNAs plays a role in their function: mRNA expression from *HeT*-A, the element whose expression is most

Figure 3.3.

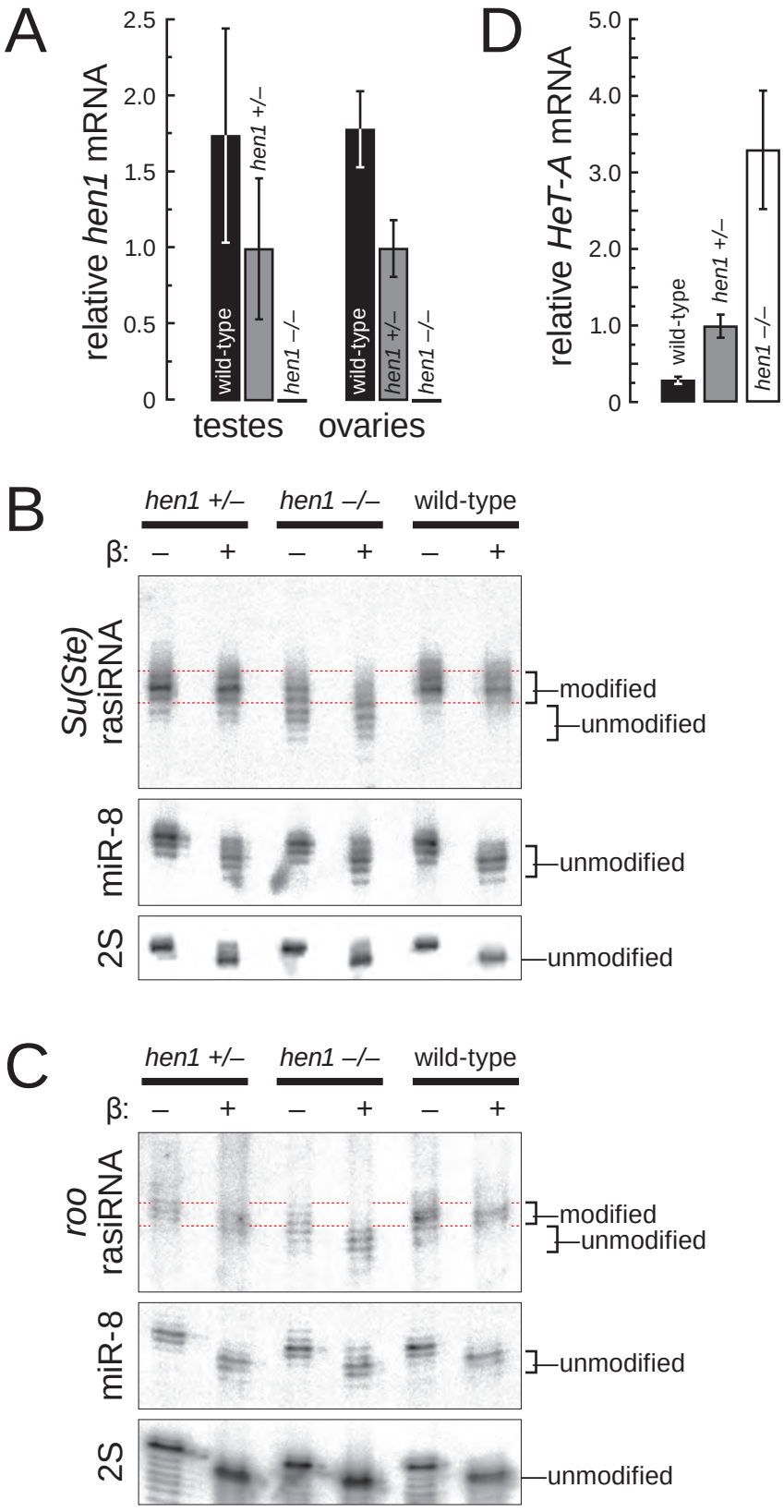

A
relative hen1 mRNA
2.5
2.0
1.5
1.0
0.5
0
wild-type
hen1 +/–
hen1 –/–
testes
ovaries
D
relative HeT-A mRNA
5.0
4.0
3.0
2.0
1.0
0
B
hen1 +/–
hen1 –/–
wild-type
β: – + – + – +
Su(Ste) rasiRNA
miR-8
2S
modified
unmodified
C
roo rasiRNA

Figure 3.3. DmHen1 is required for normal piRNA biogenesis and complete silencing of the *HeT-A*.

(A) Quantitative RT-PCR of *hen1* mRNA in testes and ovaries from wild-type or *hen1*^{f00810} flies, relative to *rp49*. Bars report the average ± standard deviation for at least four independent experiments. (B, C) Northern hybridization to detect modification of small RNAs in wild-type and *hen1*^{f00810} testes (B) or ovaries (C). The same blot was probed sequentially to detect the most abundant *Su(Ste)* rasiRNA (B) or the three most abundant *roo* rasiRNAs (C), miR-8, and 2S ribosomal RNA. (D) Quantitative RT-PCR of *HeT-A* transposon mRNA in ovaries. *HeT-A* mRNA levels were measured relative to *rp49* using total RNA prepared from wild-type, *hen1*^{f00810} heterozygous or homozygous ovaries. Bars report the average ± standard deviation for at least four independently prepared samples.

sensitive to mutations that disrupt piRNA-directed silencing in the female germ line (Vagin et al., 2006; Vagin et al., 2004; Savitsky et al., 2006), quadrupled in *hen1*^{f00810} heterozygotes and was increased more than 11-fold in homozygotes, relative to wild-type tissue (Figure 3.3D). We conclude that Hen1 protein is required for *HeT-A* silencing in the *Drosophila* germ line.

DmHen1 is required for siRNA modification

To test if DmHen1 is required for modification of the 3´ termini of siRNAs, we depleted Hen1 by RNAi in cultured *Drosophila* S2 cells. We transfected the cells with long double-stranded RNA (dsRNA) targeting *hen1* on day 1 and day 5, then co-transfected them with both GFP dsRNA and *hen1* dsRNA on day 8. Total RNA was harvested on day 9, probed for modification with $NaIO_4$/beta-elimination, and analyzed by Northern hybridization using a 5´ ^{32}P-radiolabeled DNA probe complementary to the most abundant GFP-derived siRNA (MDH, Megha Ghildiyal, and PDZ, unpublished data). DsRNAs targeting two different regions of the fly *hen1* mRNA both reduced the amount of GFP siRNA modified at its 3´ terminus, whereas all the GFP siRNA remained modified when a control dsRNA was used (Figure 3.4A).

Surprisingly, RNAi-mediated depletion of Ago2, but not Ago1, prevented the GFP siRNA from being modified. This result suggests that Ago2, but not Ago1, plays a role in the modification of siRNAs by DmHen1. To test this idea, we examined the modification status of the 3´ terminus of miR-277, an siRNA that partitions into both Ago1 and Ago2 complexes in vivo (Forstemann et al., 2007). *Drosophila* miRNAs associate predominantly or exclusively with Ago1 (Okamura et al., 2004) and have unmodified 3´ termini

Figure 3.4.

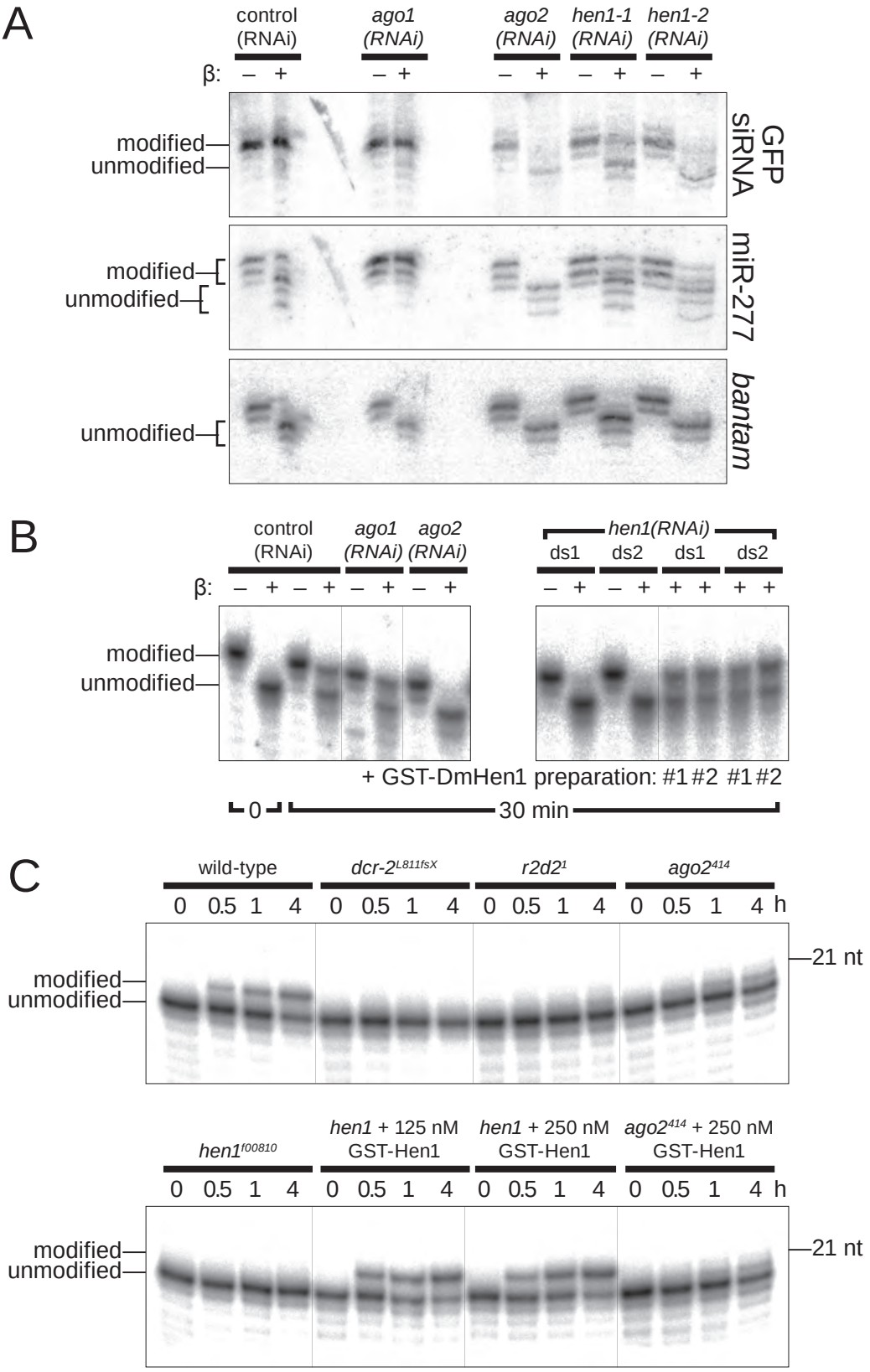

A
control (RNAi)
ago1 (RNAi)
ago2 (RNAi)
hen1-1 (RNAi)
hen1-2 (RNAi)
β: – + – + – + – + – +
modified
unmodified
GFP siRNA
modified
unmodified
miR-277
unmodified
bantam
B
control (RNAi)
ago1 (RNAi)
ago2 (RNAi)
hen1(RNAi)
ds1 ds2 ds1 ds2
β: – + – + – + – + – + – + + +
modified
unmodified
+ GST-DmHen1 preparation: #1 #2 #1 #2
0
30 min
C
wild-type
dcr-2L811fsX
r2d21
ago2414
0 0.5 1 4 0 0.5 1 4 0 0.5 1 4 0 0.5 1 4 h
modified
unmodified
21 nt
hen1f00810
hen1 + 125 nM GST-Hen1
hen1 + 250 nM GST-Hen1
ago2414 + 250 nM GST-Hen1
0 0.5 1 4 0 0.5 1 4 0 0.5 1 4 0 0.5 1 4 h
modified
unmodified
21 nt

Figure 3.4. DmHen1 modifies Ago2-associated small RNAs.

(A) Modification of siRNAs and miRNAs in *Drosophila* S2 Cells. A stable S2 cell line expressing GFP was treated with the indicated dsRNA alone (day 1 and 5), then together with GFP dsRNA (day 8). Two non-overlapping dsRNAs were used to target *hen1*. Total RNA was collected on day 9, treated with $NaIO_4$/beta-elimination, then dsGFP-derived siRNA, miR-277 and *bantam* detected by sequential Northern hybridization of the same blot. (B) In vitro siRNA modification in dsRNA-treated S2 cell lysates. (C) In vitro siRNA modification in *Drosophila* mutant ovary lysates.

(Hutvagner et al., 2001; Pelisson et al., 2007; Vagin et al., 2006). In contrast, about half the miR-277 in cultured S2 cells failed to react with $NaIO_4$ (Figure 3.4A), suggesting that about half of miR-277 is modified at its 3´ terminus. The fraction of miR-277 that was modified was reduced when two different dsRNAs were used to deplete DmHen1 by RNAi. When the cells were treated with dsRNA targeting *ago1*, all detectable miR-277 was modified, whereas all miR-277 became unmodified when dsRNA targeting *ago2* was used. In contrast, *bantam*, a miRNA that associates nearly exclusively with Ago1 (Okamura et al., 2004), was unmodified under all conditions (Figure 3.4A).

siRNA modification correlates with Ago2-RISC assembly in vitro

siRNA modification can be recapitulated in lysates of embryos, ovaries, or cultured S2 cells. Modification of siRNA in vitro was inhibited by *S*-adenosyl homocysteine, but not by *S*-adenosyl methionine, consistent with DmHen1 transferring a methyl group from *S*-adenosyl methionine to the terminal 2´ hydroxyl group of the RNA, thereby generating *S*-adenosyl homocysteine as a product (Figure 3.S4).

Our data from cultured S2 cells suggested that DmHen1 modifies that portion of miR-277 that enters the Ago2-RISC assembly pathway, but not the population of miR-277 that assembles into Ago1-RISC. To further test the idea that small RNA modification requires both Hen1 and the Ago2-RISC assembly pathway, we prepared cytoplasmic lysates from dsRNA-treated cultured S2 cells. Lysate from control-treated cells modified the 3´ terminus of a 5´ ^{32}P-radiolabeled synthetic siRNA duplex, but not lysate from *hen1*-depleted cells (Figure 3.4B). The addition of either of two different preparations of purified, recombinant

DmHen1, expressed in *E. coli* as a ~74 kDa glutathione S-transferase fusion protein (GST-DmHen1; Figure 3.S5), restored the ability of the lysates to modify the siRNA, indicating that loss of DmHen1 caused the loss of siRNA modification. Moreover, lysates depleted for Ago2, but not Ago1, could not modify the ^{32}P-siRNA in vitro (Figure 3.4B). These in vitro data, together with our S2 cell experiments, suggest that modification of the 3´ terminus of siRNAs and miRNAs is coupled to assembly into Ago2-RISC.

Dcr-2 and R2D2 act to load double-stranded siRNAs into Ago2. We prepared lysates from ovaries homozygous mutant for *hen1*, *dcr-2*, *r2d2*, and *ago2* using alleles unable to produce the corresponding protein (Lee et al., 2004; Liu et al., 2003; Okamura et al., 2004). A 5´ ^{32}P-radiolabeled siRNA duplex was incubated in each lysate to assemble RISC. At each time point, we determined if the siRNA was 3´ terminally modified by assessing its reactivity with $NaIO_4$ (Figure 3.4C). No modified siRNA accumulated when the duplex was incubated in *hen1*^{f00810}, *dcr-2*^{L811fsX}, *r2d2*^{1}, or *ago2*^{414} mutant lysate. Adding 250 nM purified, recombinant GST-DmHen1 restored siRNA modification to the *hen1*^{f00810} but not the *ago2*^{414} lysate. We conclude that the defect in *ago2*^{414} reflects a requirement for Ago2 in small RNA modification by DmHen1, rather than an indirect effect, such as destabilization of DmHen1 in the absence of Ago2. GST-DmHen1 similarly rescued lysate from *hen1(RNAi)* but not *ago2(RNAi)* treated S2 cells (Figure 3.S6). Together, the results of our experiments using cultured S2 cells—a somatic cell line—and ovaries, which comprise mainly germ line tissue, suggest that a functional Ago2-RISC assembly pathway is required for siRNA modification in *Drosophila*.

siRNAs are modified only after Ago2-RISC maturation

To test at which step in the Ago2-RISC assembly pathway siRNAs become modified, we determined if siRNAs are 2´-*O*-methylated by DmHen1 as single-strands or as duplexes. In vitro, assembly of siRNAs into Ago2-RISC follows an ordered pathway in which the siRNA duplex first binds the Dicer-2/R2D2 heterodimer to form the RISC-loading complex (RLC). The RLC determines which of the two siRNA strands will become the guide for Ago2 and which will be destroyed (the passenger strand). The siRNA is then loaded into Ago2 as a duplex, with the passenger strand occupying the same position as future target RNAs, generating pre-RISC (Kim et al., 2007). Cleavage of the passenger strand by the Ago2 endonuclease domain converts pre-RISC to mature RISC (Leuschner et al., 2006; Matranga et al., 2005; Rand et al., 2005; Miyoshi et al., 2005). No single-stranded guide or passenger RNA is produced prior to this maturation step. Thus, all single-stranded siRNA produced in vitro or in vivo (Kim et al., 2007) corresponds to mature RISC.

We assembled Ago2-RISC in vitro using an siRNA designed to load only one of its two strands into Ago2 (Schwarz et al., 2003). We then sampled the reaction over time, isolating the 5´ ^{32}P-radiolabeled siRNA under conditions previously demonstrated to preserve its structure (Nykanen et al., 2001), and separated single- from double-stranded siRNA by native gel electrophoresis (Figure 3.S7). (Full-length siRNA duplexes and siRNA heteroduplexes comprising a full-length guide paired to a cleaved passenger strand co-migrate as double-stranded siRNA in these gel conditions (Matranga et al., 2005).) The RNAs were then isolated from the gel and tested for reactivity with $NaIO_4$ to determine the presence of modification at their 3´ termini (Figure 3.5A and 3.5B). At each time,

Figure 3.5.

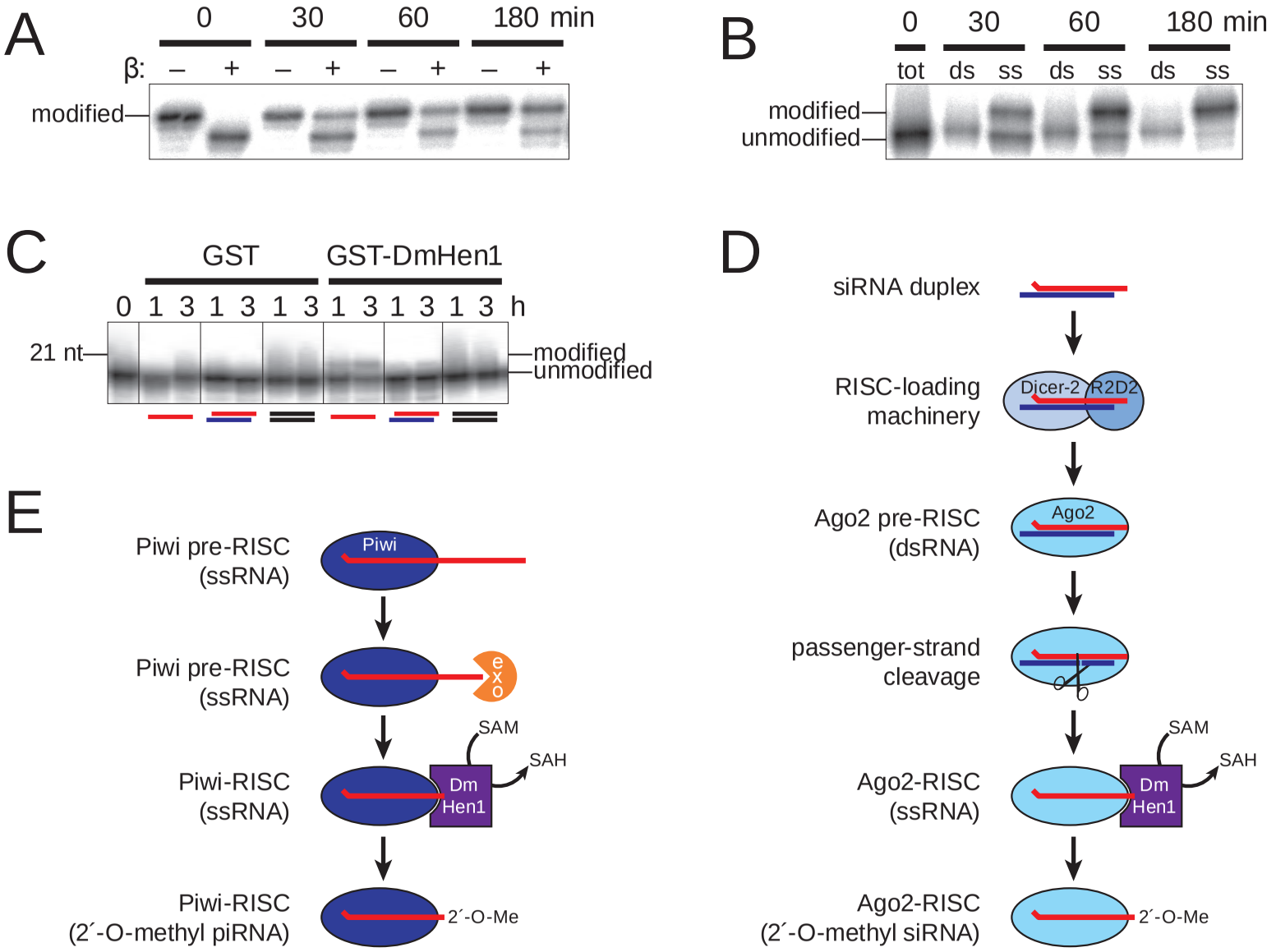

A
0
30
60
180 min
β: – + – + – + – +
modified
B
0
30
60
180 min
tot ds ss ds ss ds ss
modified
unmodified
C
GST
GST-DmHen1
0 1 3 1 3 1 3 1 3 1 3 1 3 h
21 nt
modified
unmodified
D
siRNA duplex
RISC-loading machinery
Dicer-2
R2D2
Ago2 pre-RISC (dsRNA)
Ago2
passenger-strand cleavage
Ago2-RISC (ssRNA)
SAM
SAH
Dm Hen1
Ago2-RISC (2´-O-methyl siRNA)
2´-O-Me
E
Piwi pre-RISC (ssRNA)
Piwi
Piwi pre-RISC (ssRNA)
exo
Piwi-RISC (ssRNA)
SAM
SAH
Dm Hen1
Piwi-RISC (2´-O-methyl piRNA)
2´-O-Me

Figure 3.5. siRNAs are modified in the mature RISC.

siRNA duplexes with a 5´ ^{32}P-radiolabeled guide strand were incubated in *Drosophila* embryo lysate, and then tested for the presence of a 3´ terminal modification. (A) Total RNA from each time point in RISC assembly, without (–) and with (+) reaction with $NaIO_4$ and beta-elimination. (B) At each times in (A), single- and double-stranded siRNA were resolved and purified by first native gel electrophoresis (Figure 3.S7), then analyzed by denaturing electrophoresis separately for the presence of a 3´ terminal modification on the siRNA guide strand. (A) and (B) are the left and right halves of a single gel. (C) Recombinant, purified GST-DmHen1, but not GST alone, can modify single-stranded 21 nt RNA, but not double-stranded siRNAs or blunt 21 nt RNA duplexes. However, in contrast to GST-DmHen1 incubated with *hen1*^{f00810} mutant ovary lysate (Figure 3.4C), the enzyme alone is inefficient. All samples were oxidized with $NaIO_4$, then beta-eliminated. (D) A model for 2´-*O*-methylation of siRNAs. (E) A proposed role for 2´-*O*-methylation in piRNA biogenesis.

total siRNA was analyzed in parallel. 3′ terminal modification increased over the course of RISC assembly and, at all times, was restricted to single-stranded siRNA: within our limits of detection, all double-stranded siRNA was unmodified, even after 3 h. We conclude that siRNA modification is coupled to RISC assembly and occurs only after the conversion of pre-RISC to mature RISC.

Recombinant DmHen1 modifies single-stranded small RNA

While *Arabidopsis* HEN1 contains an N-terminal double-stranded RNA-binding motif (Yu et al., 2005), DmHen1 does not. To test if DmHen1 modifies double-stranded small RNAs, we incubated purified, recombinant, GST-DmHen1 with either single-stranded or double-stranded siRNAs. We detected modification, evidenced by loss of reactivity with $NaIO_4$, only for the single-stranded RNA, suggesting that DmHen1 modifies single-stranded substrates, but not siRNAs or blunt RNA duplexes (Figure 3.5C). A preference for single-stranded RNA would explain how DmHen1 could act on both siRNAs, which are born double-stranded, and piRNAs, which are not. We note that the purified, recombinant GST-DmHen1 protein was more than 50-fold less active on its own than when supplemented with ovary lysate from *hen1*^{f00810} homozygous flies. We speculate that the Ago2-RISC machinery is required for Hen1 function in flies, although we cannot yet exclude the possibility that the lysate contains a factor (e.g., a kinase) required to activate Hen1. Modification of single-stranded siRNAs—that is, those loaded in fully mature Ago2-RISC, but not double-stranded siRNAs (Figure 3.5D) might allow cells to distinguish siRNAs loaded successfully into functional complexes from those that fail to assemble. For example, if a 3′-to-5′ nuclease acts to degrade single-stranded siRNAs, 2′-*O*-

methylation of single-stranded siRNAs in Ago2-RISC may protect them from destruction. Moreover, such a nuclease might trim the 3′ end of piRNAs. 2′-*O*-methylation of the piRNA 3′ terminus may occur only when the length of RNA extending beyond the Piwi family protein is short enough to permit the simultaneous binding of the final ribose sugar to the active site of DmHen1 and the interaction of DmHen1 with the Piwi protein itself (Figure 3.5E). Modification of the terminus of the trimmed piRNA would then block further 3′-to-5′ trimming of the small RNA, generating its Piwi-, Aubergine-, or Ago3-specific length. Our observation that piRNAs are shorter in *hen1*^{f00810} mutants supports this model.

We note that all 2′-*O*-methyl modified small RNAs identified thus far are associated with RISC complexes that efficiently cleave their RNA targets—i.e., Ago1-associated plant miRNAs (Baumberger and Baulcombe, 2005; Qi et al., 2005), animal piRNAs (Lau et al., 2006), and Ago2-associated siRNAs in flies (Hammond et al., 2001)—whereas *Drosophila* miRNAs are typically both unmodified and associated with Ago1-RISC, which does not catalyze mRNA target cleavage in vivo (Forstemann et al., 2007). We speculate that DmHen1 is recruited to RISC complexes containing single-stranded small silencing RNAs according to the identity of their Argonaute protein. This model predicts that DmHen1 will bind only to complexes containing fly Ago2 or the three fly Piwi proteins, Piwi, Aubergine, and Ago3, but not Ago1. Clearly, future experiments will need to test this hypothesis.

Experimental Procedures

General Methods

Preparation of 0–2 h embryo, ovary, and S2 cell lysates and in vitro RISC assembly and RNAi reactions, and Northern hybridization were as described (Haley et al., 2003; Forstemann et al., 2005; Vagin et al., 2006). Sequences of synthetic RNA and DNA oligonucleotides are available online (Table 3.S1).

^{32}P-radiolabeled 3´ mononucleotide standards

Synthetic RNA oligonucleotides (Table 3.S1) were radiolabeled in a 20 µl reaction containing 50 mM Tris-HCl (pH 7.8), 10 mM $MgCl_2$, 1 mM ATP, 10 mM dithiothreitol, 10% (v/v) DMSO, 10 µg/ml BSA, 2 units/µl RNasin (Promega, Madison, WI), 1.5 µCi/µl [5´ ^{32}P] cytidine 5´,3´ bis-phosphate ([5´ ^{32}P]-pCp; Perkin-Elmer, Waltham, MA), 1 unit/µl T4 RNA Ligase 1 (New England Biolabs, Ipswich, MA) at 4°C, overnight. The radiolabeled small RNAs were purified from a 15% denaturing urea-polyacrylamide sequencing gel, and then digested with 1.5 U/µl micrococcal nuclease (Takara Mirus Bio, Madison, WI) in a 40µl reaction containing 20 mM Tris-HCl pH 8.0, 5 mM NaCl, 2.5 mM $CaCl_2$. 3´ ^{32}P-mononucleotides were further purified from a 22.5% denaturing urea-polyacrylamide sequencing gel.

2D-TLC

Small RNAs (21-29 nt, containing both modified piRNAs and unmodified small RNAs) from 0–2 h wild-type (Oregon R) fly embryos and small RNAs (26-31 nt, containing mostly modified piRNAs) from mouse and bull testes were

purified from a 10% denaturing urea-polyacrylamide gel stained with SYBR® Gold (Invitrogen). About 100 pmol purified small RNAs were radiolabeled as described above, except in a 40 µl reaction using 3µCi/µl [5´ ^{32}P]-pCp and 1 unit/µl T4 RNA Ligase 1, and then gel purified. The purified, ^{32}P-radiolabeled RNA was hydrolyzed in 200mM Na_2CO_3 at 100°C for 1h, then neutralized with an equal volume of 200 mM HCl, dephosphorylated with 0.5 units/µl calf intestinal alkaline phosphatase (New England Biolabs) in a 200 µl reaction containing 50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol. Alkaline phosphatase was inactivated by extraction with phenol /chloroform, RNA in the aqueous phase was oxidized with 80 mM $NaIO_4$ in borax/boric acid buffer (60 mM borax, 60 mM boric acid, pH 8.6) at room temperature for 30 min, and then beta-eliminated with 200 mM NaOH at 45°C for 90 min. 5 µl of this reaction was mixed with an equal volume of formamide loading buffer (98% deionized formamide, 10 mM EDTA (pH 8.0), 0.025% (w/v) xylene cyanol, 0.025% (w/v) bromophenol blue), and resolved on a 22.5% denaturing urea-polyacrylamide sequencing gel. Equal intensities of modified mononucleotides and standards were spotted on 20 x 20 cm PEI-cellulose F glass TLC plates (EMD Chemicals, Gibbstown, NJ) and separated first with isobutyric acid/25%ammonia/water (66:1:33, v:v:v) and then 0.1 M sodium phosphate pH6.8/ammonium sulfate/1-propanol (100:60:2, v:w:v).

Analysis of RNA 3´ termini

RNA was incubated for 30 min at room temperature with 25 mM $NaIO_4$ in borax/boric acid buffer (60 mM borax, 60 mM boric acid, pH 8.6), then 45.2 mM (f.c.) NaOH added, and incubation continued for 90 min at 45°C (beta-

elimination). The reaction was stopped by the addition of 300 mM (f.c.) NaCl, 1 µg glycogen, and three volumes absolute ethanol. After 30 min on ice, the precipitated RNA was collected by centrifugation.

Recombinant *Drosophila* Hen1 Protein

DmHen1 coding sequence was amplified from *Drosophila* ovary cDNA and inserted into pEnt-D-Topo (Invitrogen, Carlsbad, CA). The entire *hen1* sequence was confirmed by sequencing. The entry plasmid was recombined with the N-terminal GST expression vector, pDest-15, using LR Clonase (Invitrogen). GST-Hen1 was expressed in BL21 Star DE3 cells (Invitrogen) grown at 37°C in LB broth containing 100 ug/ml ampicillin until to OD_{600} reached 0.50. The culture was then cooled to 25°C and 0.4 mM IPTG added to induce protein production. The culture was incubated at 25°C with vigorous shaking for three hours. The cells were harvested by centrifugation at 7,300 x g for 20 min, washed with PBS, centrifuged again, and then the cell paste frozen in liquid nitrogen and stored at -80°C. Hen1 fusion protein was purified using the GST Purification Kit (Clontech, Mountainview, CA). Cells were resuspended in 40 ml of Extraction/Loading buffer and lysed by sonication (duty 30% for 6 minutes; Branson Sonificator II, Danbury, CT), with cooling in an ice bath. Two ml of clarified lysate was added to the column resin, and the column inverted several times to disperse the resin. The resin was then allowed to pack, and the remaining 38 ml passed through the column by gravity flow. Subsequent steps were according to the manufacturer's directions.

Analysis of double- and single-stranded siRNA

Double and single-stranded, 5´ ^{32}P-radiolabeled siRNA guide strands (10 nM; Figure 3.5 and Figure 3.S7) were separated as described (Nykanen et al., 2001). Briefly, RISC assembly reactions were stopped with 2x Proteinase K buffer, 2 mg/ml Proteinase K, 1 µg glycogen, and 250 nM unlabeled siRNA guide strand to prevent reannealing. After incubation for 30 min at 25˚C, 3 volumes absolute ethanol were added, and the RNA precipitated for 30 min on ice. The precipitates were collected by centrifugation, washed with 80% (v/v) ethanol, then dissolved in 2 mM Tris-Cl (pH 7.5), 3% (w/v) Ficoll-400, 0.04% (w/v) bromophenol blue, 100 mM KOAc, 30 mM HEPES-KOH, 2 mM $Mg(CH_3CO_2)_2$, and resolved by electrophoresis through a 15% native polyacrylamide gel (19:1 acrylamide:bis; 89 mM Tris-Borate, pH 8.3, 2 mM Na-EDTA, 2.5 mM $Mg(CH_3CO_2)_2$). The region of the native gel corresponding to double-or single-stranded siRNA was excised, and the RNA eluted overnight in 1 M NaCl. 1 µg glycogen and ethanol (60% final volume) was added to the eluate, the RNA collected using MegaClear filter cartridges (Ambion), eluted with H_2O, and then precipitated for 30 min on ice by adding 500 mM (f.c.) $NH_4CH_3CO_2$ and 2.5 volumes absolute ethanol. The precipitate was collected by centrifugation, washed with 80% (v/v) ethanol, and the samples reacted with $NaIO_4$ and subsequent beta-elimination (see above). The precipitated RNA was dissolved in 98% deionized formamide, 10 mM EDTA (pH 8.0), 0.025% (w/v) xylene cyanol, 0.025% (w/v) bromophenol blue, and then resolved on a 15% denaturing urea-polyacrylamide sequencing gel.

Acknowledgments

We thank Alicia Boucher for assistance with fly husbandry, Marian Walhout for the GST destination vector, Steven Miller for GST protein, and members of the Zamore lab for advice, suggestions, and critical comments on the text. We are especially grateful to Alla Sigova for assistance preparing germ line tissue and with qPCR and Tingting Du for help preparing ovary lysates. PDZ is a W.M. Keck Foundation Young Scholar in Medical Research. This work was supported in part by grants from the National Institutes of Health to PDZ (GM62862 and GM65236), an NRSA pre-doctoral MD/PhD fellowship from National Institute on Aging (F30AG030283) to MDH, and a predoctoral fellowship from the American Heart Association to CM.

Figure 3.S1.

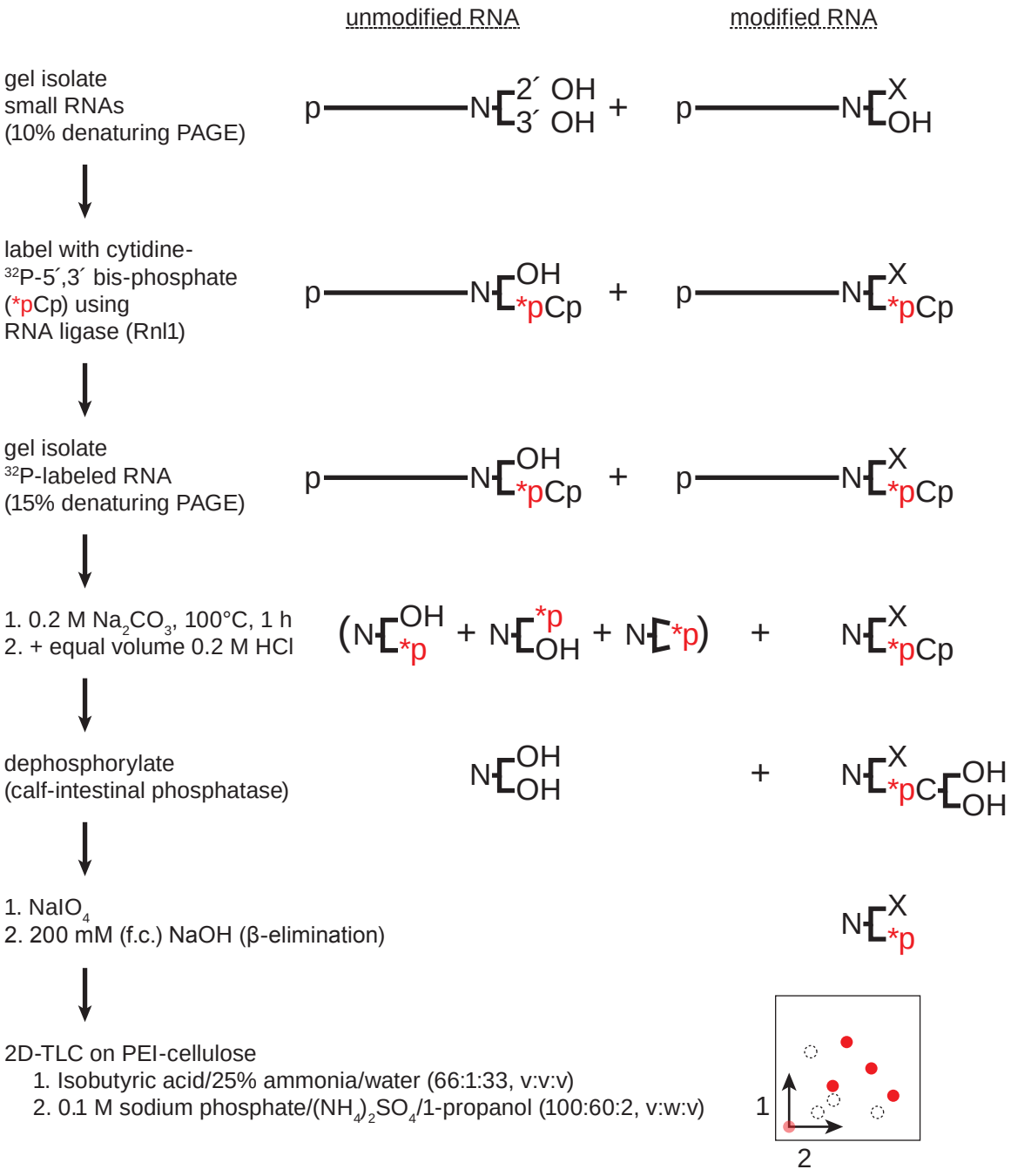

unmodified RNA
modified RNA
gel isolate
small RNAs
(10% denaturing PAGE)
label with cytidine-
32P-5´,3´ bis-phosphate
(*pCp) using
RNA ligase (Rnl1)
gel isolate
32P-labeled RNA
(15% denaturing PAGE)
1. 0.2 M Na2CO3, 100°C, 1 h
2. + equal volume 0.2 M HCl
dephosphorylate
(calf-intestinal phosphatase)
1. NaIO4
2. 200 mM (f.c.) NaOH (β-elimination)
2D-TLC on PEI-cellulose
1. Isobutyric acid/25% ammonia/water (66:1:33, v:v:v)
2. 0.1 M sodium phosphate/(NH4)2SO4/1-propanol (100:60:2, v:w:v)
1
2

Figure 3.S1. Scheme for selectively labeling the 3´ terminal nucleotide of modified small RNAs.

Because Rnl1 will use either a 2´ or a 3´ hydroxyl as a ligation donor, the scheme will also label 3´ modified RNAs. Only phosphodiester linkages flanked by a 2´ hydroxyl are subject to base hydrolysis and only adjacent 2´ and 3´ hydroxyls react with $NaIO_4$, a prerequisite for beta-elimination. X, 2´ or 3´ modification; *p, ^{32}P-radiolabeled phosphate group.

Figure 3.S2.

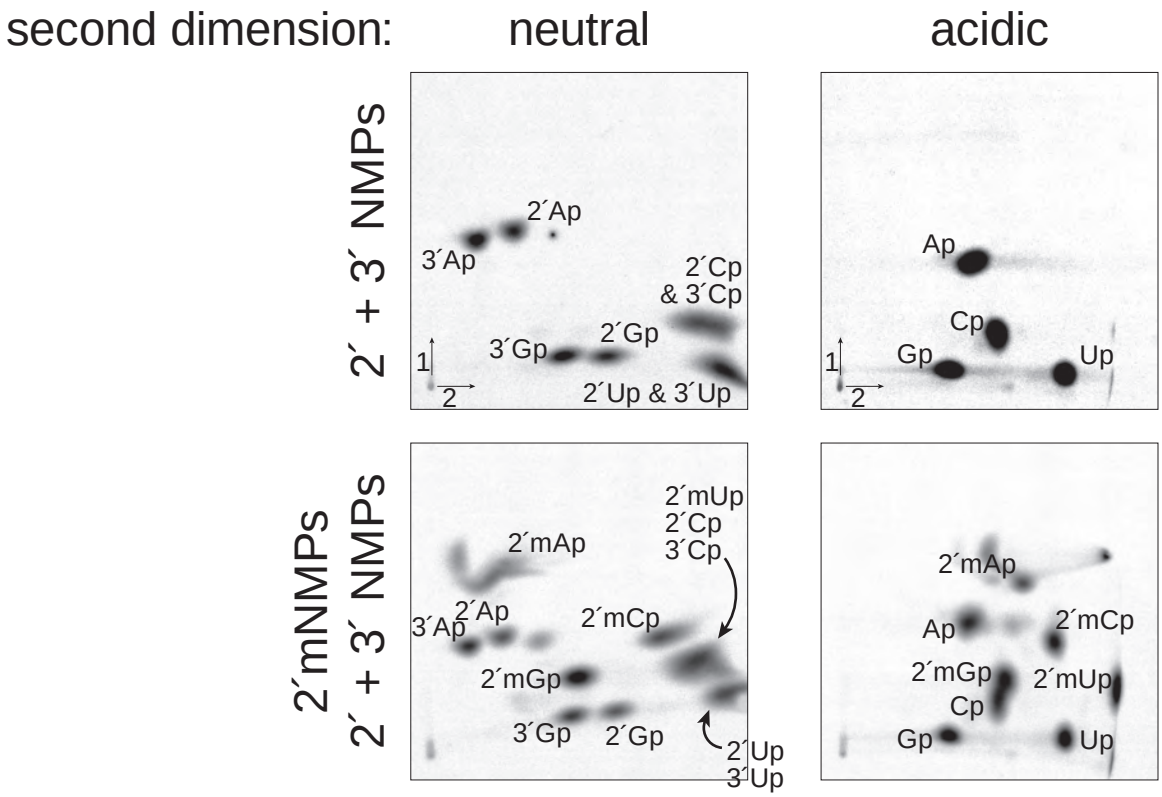

second dimension:
neutral
acidic
2´ + 3´ NMPs
2´mNMPs 2´ + 3´ NMPs
2´Ap
3´Ap
2´Cp
& 3´Cp
3´Gp
2´Gp
2´Up & 3´Up
1
2
Ap
Cp
Gp
Up
2´mUp
2´Cp
3´Cp
2´mAp
2´Ap
3´Ap
2´mCp
2´mGp
3´Gp
2´Gp
2´Up
3´Up
2´mAp
Ap
2´mCp
2´mGp
2´mUp
Cp
Gp
Up

Figure 3.S2. Comparison of 2D TLC systems.

For both, the first dimension was Isobutyric acid/25% ammonia/water (66:1:33, v:v:v). Neutral second dimension (0.1 M sodium phosphate/$(NH_4)_2SO_4$/1-propanol [100:60:2, v:w:v]) (this paper); acidic second dimension (2-propanol/ HCl/H_2O (70:15:15, v:v:v) (Kirino and Mourelatos, 2007a). 2´ and 3´ NMPs were prepared by base hydrolysis, and 3´ NMP spots identified by their comigration with 3´ NMPs generated by complete digestion of RNA with micrococcal nuclease.

Figure 3.S3.

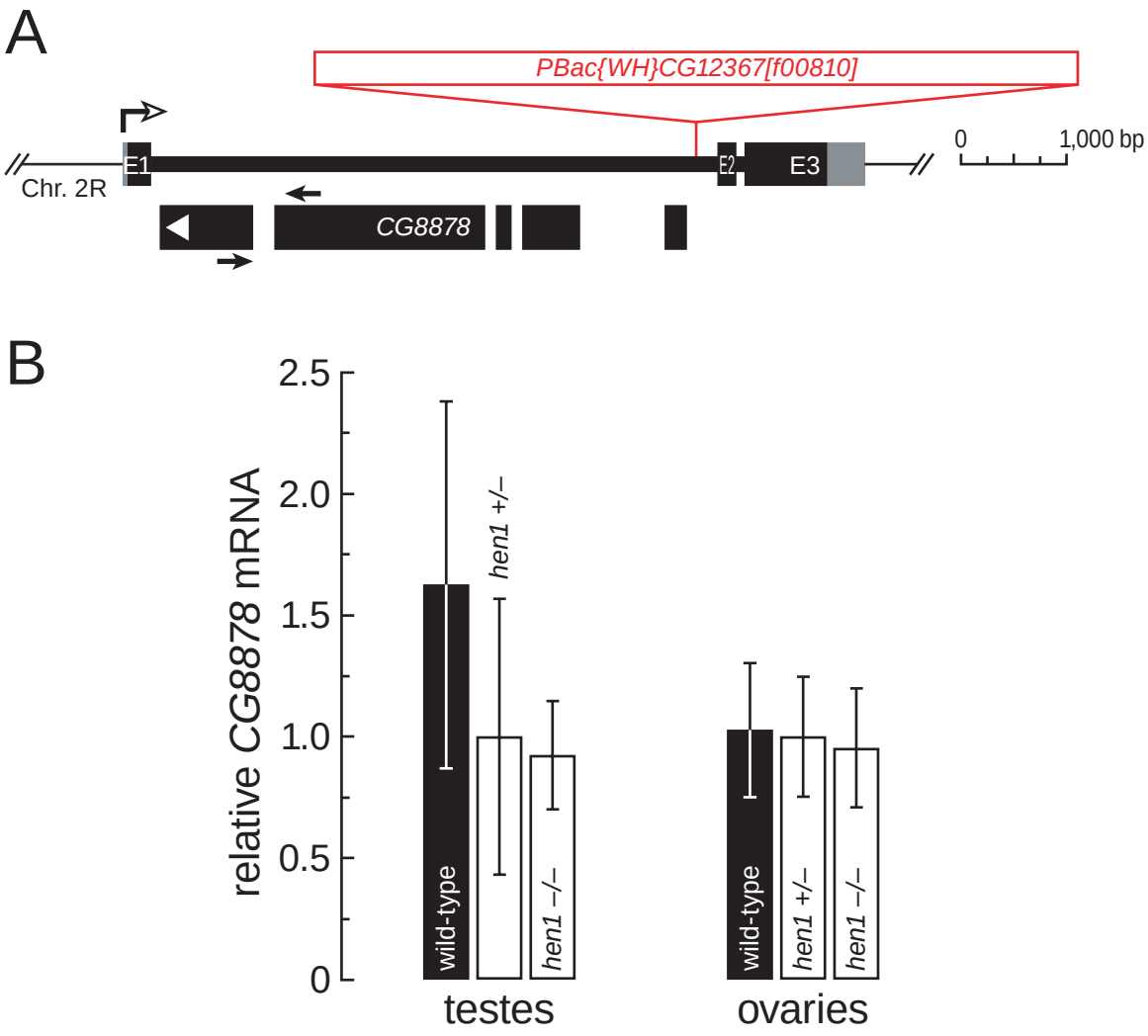

A
PBac{WH}CG12367[f00810]
E1
E2
E3
Chr. 2R
CG8878
0
1,000 bp
B
relative CG8878 mRNA
2.5
2.0
1.5
1.0
0.5
0
wild-type
hen1 +/–
hen1 –/–
testes
ovaries

Figure 3.S3. PBAC{WH}CG12367[f00810] disrupts *hen1* but not *CG8878*. (A) The gene *CG8878* resides in the first intron of *Drosophila hen1* and is transcribed in the opposite orientation. The location of the qRT-PCR primers are shown as closed arrows. (B) We performed qRT-PCR to determine the effect on *CG8878* expression of the PBAC{WH}CG12367[f00810] transposon insertion, which disrupts *hen1* (Figure 3.3A). In both testes and ovaries, we can detect no statistically meaningful effect of this piggyBac transposon insertion on the expression of *CG8878*. We conclude that the PBAC{WH}CG12367[f00810] insertion disrupts only *hen1*. We therefore rename this insertion $hen1^{f00810}$. Bars show average ± standard deviation for four or five independent trials.

Figure 3.S4.

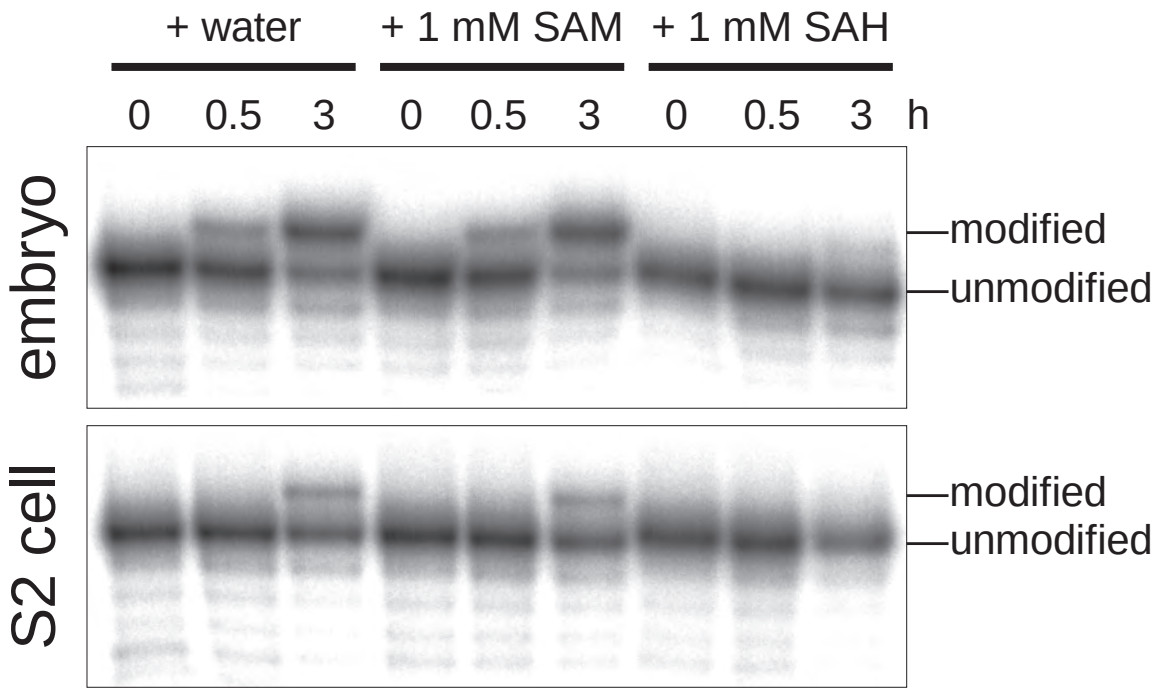
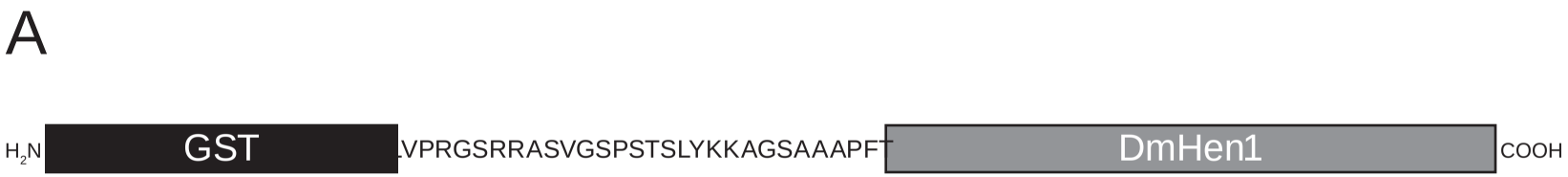

+ water
+ 1 mM SAM
+ 1 mM SAH
0 0.5 3 0 0.5 3 0 0.5 3 h
embryo
modified
unmodified
S2 cell
modified
unmodified

Figure 3.S4. S-adenosyl homocysteine inhibits DmHen1.

S-adenosyl homocysteine (SAH), an inhibitor of S-adenosyl methionine (SAM) dependent methyltransferases, inhibits modification of single-stranded siRNA in lysate prepared from either 0–2 h embryos or cultured S2 cells. RNA was purified from the reactions at the indicated times and then tested for modification by reaction with $NaIO_4$ followed by beta-elimination. The RNA was resolved by denaturing electrophoresis.

Figure 3.S5.


A
H2N
GST
LVPRGSRRASVGSPSTSLYKKAGSAAAPFT
DmHen1
COOH

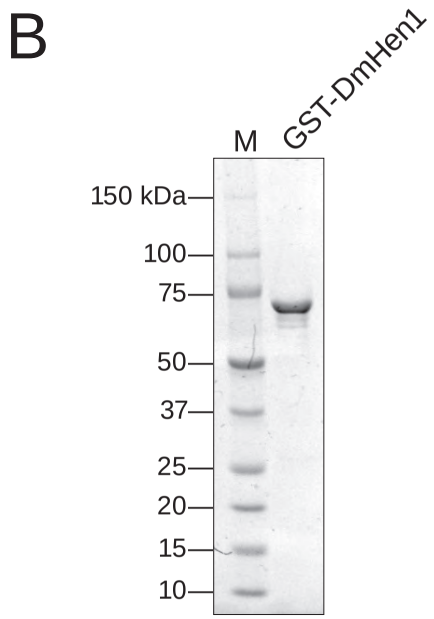

B
M
GST-DmHen1
150 kDa
100
75
50
37
25
20
15
10

Figure 3.S5. GST-DmHen1.

(A) Schematic of the 74,254 Da fusion protein, comprising the 220 amino acid glutathione S-transferase (GST) module, a 30 amino acid linker, and the entire 391 amino acid DmHen1 protein, terminated at its native stop codon. (B) Purified, recombinant GST-DmHen1 protein (3 µg) was resolved by electrophoresis through a 4–20% polyacrylamide gradient SDS-gel, then stained with colloidal Coomassie G-250. The apparent masses of molecular weight markers (M) are indicated.

Figure 3.S6.

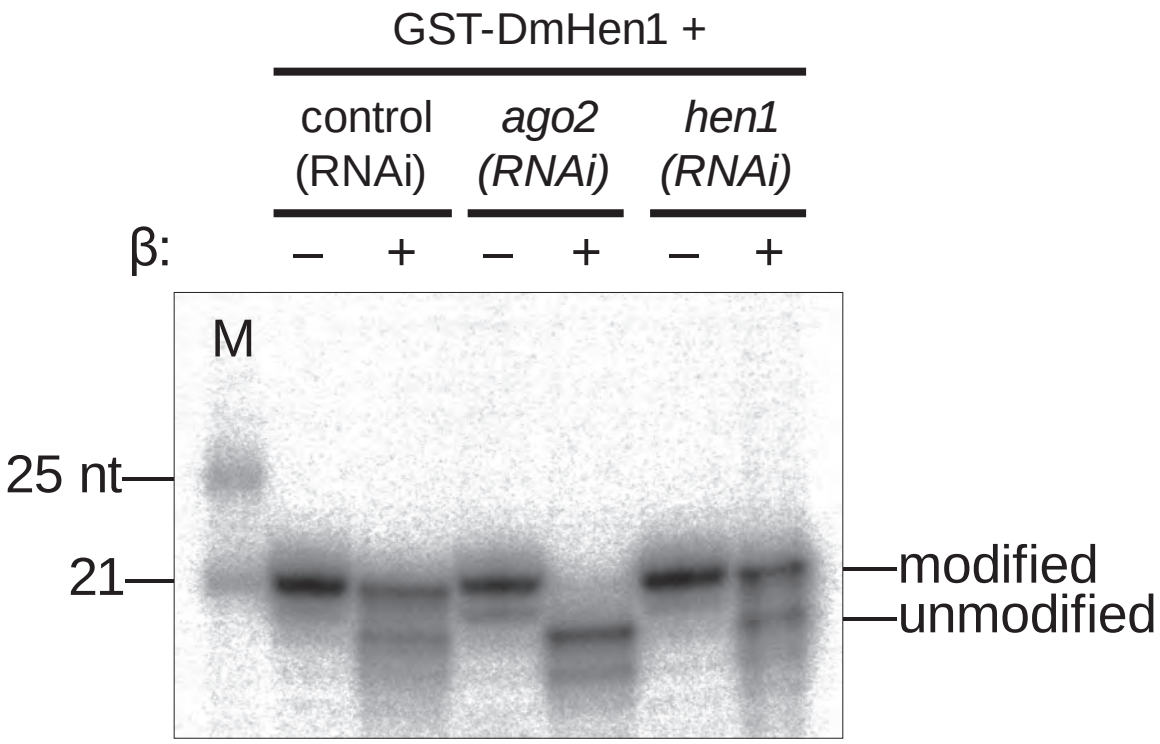

GST-DmHen1 +
control (RNAi)
ago2 (RNAi)
hen1 (RNAi)
β:
– + – + – +
M
25 nt
21
modified
unmodified

Figure 3.S6. GST-DmHen1 protein rescues S2 cells depleted by RNAi of Hen1, but not Ago2.

S2 cells were transfected as described with dsRNA targeting *ago2*, *hen1*, or a control sequence, then lysate prepared. siRNA was incubated in the S2 cell lysate to allow RISC assembly, then siRNA guide strand modification was assayed by reaction with $NaIO_4$ followed by beta-elimination (b). Without the addition of GST-DmHen1 protein, *hen1*(RNAi) and *ago2*(RNAi) lysates are deficient in Hen1 activity, relative to the control dsRNA treated cells (Figure 3.4B). M, 5´-phosphorylated synthetic RNA size markers.

Figure 3.S7.

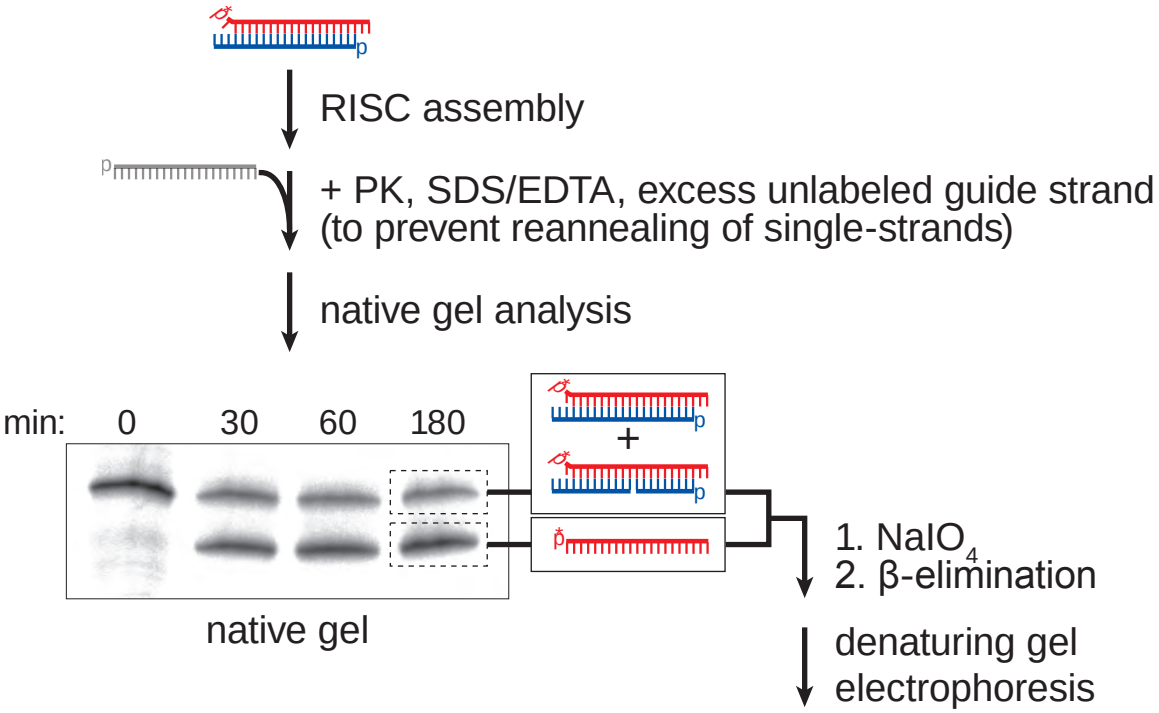


Figure 3.5B

Figure 3.S7. Strategy for testing when during RISC assembly siRNAs are modified at their 3´ termini.

Double-stranded siRNA was incubated in embryo lysate to assemble RISC, then the siRNA purified away from protein using a procedure that preserves the single- or double-stranded structure of the siRNA at the time assembly was stopped. The siRNAs were then resolved on a native gel, and the single- and double-stranded siRNA was isolated from the gel, reacted with $NaIO_4$ followed by beta-elimination and analyzed by denaturing gel electrophoresis (shown in Figure 3.5B).

Table 3.S1. Synthetic oligonucleotides

siRNA duplex sequences

siRNA duplexes	Guide, Passenger
luc siRNA	5´-UGUACGCGGAAUACUUCGAAA-3´, 5´-UCGAAGUAUUCCGCGUACGUG-3´
luc siRNA	5´-UCGAAGUAUUCCGCGUACGUG-3´, 5´-CGUACGCGGAAUACUUCGAAA-3´
let-7 siRNA	5´-UGAGGUAGUAGGUUGUAUAGU-3´, 5´-UAUACAACCUACUACCUCCUU-3´

Oligonucleotides used to generate mononucleotide standards for 2D-TLC

5´-CCUUAUCAUUCUCUCGCCCCG-3´
5´-CCUUAUCAUUCUCUCGCCCCC-3´
5´-CCUUAUCAUUCUCUCGCCCCU-3´
5´-CCUUAUCAUUCUCUCGCCCCA-3´
5´-UUACAUAAGAUAUGAACGGAGCCC-A-2´-O-Me,3´-OH
5´-UUACAUAAGAUAUGAACGGAGCCC-C-2´-O-Me,3´-OH
5´-UUACAUAAGAUAUGAACGGAGCCC-U-2´-O-Me,3´-OH
5´-UUACAUAAGAUAUGAACGGAGCCC-G-2´-O-Me,3´-OH´

Primers used to generate dsRNAs for *Dmhen1* knockdown

#	Forward, Reverse
1	5´-GCGTAATACGACTCACTATAGGGTTCGGAAGAATTGAGGGACAC-3´, 5´-GCGTAATACGACTCACTATAGGGGGTGCGCGTATCCACATAAA-3´
2	5´-GCGTAATACGACTCACTATAGGGAAATAGAGCTTCAGCGTTTCA-3´, 5´-GCGTAATACGACTCACTATAGGGCGATCTTCGGTTTGTCTGG-3´

Primers for *Dmhen1* cloning

Forward	5´-CACCATGTTTTCGCACAAGTTTATTTGC-3´
Reverse	5´-TTATGATTCGGGGCCTTGATC-3´

Primers for quantitative PCR

Gene	Forward, Reverse
Dmhen1	5´-GCGTCGCATTGAGACCATAG-3´, 5´-GCGATTACAGCGTCAGTGTCC-3´
CG8878	5´-TTGTCCAGGAGCATATTGAGG-3´, 5´-TGCCCTCCATTGATCTATCC-3´
rp49	5´-CCGCTTCAAGGGACAGTATCTG-3´, 5´-ATCTCGCCGCAGTAAACGC-3´

Probes for Northern hybridization

Small RNAs	Sequences
#4 *Su(Ste)* rasiRNA	5´-TCGGGCTTGTTCTACGACGATGAGA-3´
#1 *roo* rasiRNA	5´-TGGGCTCCGTTCATATCTTATG-3´
#3 *roo* rasiRNA	5´-TGAGAGTTCGCTATTCGAAGAA-3´
#7 *roo* rasiRNA	5´-TCTGAGGCATCCGTTTGGTAAA-3´
miR-8	5´-GACATCTTTACCTGACAGTATTA-3´
bantam	5´-CAGCTTTCAAAATGATCTCACT-3´
2S rRNA	5´-TACAACCCTCAACCATATGTAGTCCAAGCA-3´
miR-277	5´-TCGTACCAGATAGTGCATTTA-3´
GFP siRNA	5´-GGCCGCGACTCTAGATCATAAT-3´

CHAPTER IV: COLLAPSE OF GERM-LINE PIRNAS IN THE ABSENCE OF ARGONAUTE3 REVEALS SOMATIC PIRNAS IN FLIES

Preface

The work presented in the following chapter was a collaborative effort among several lab members. Dr. Zhiping Weng's lab (Soohyun Lee, Jia Xu, Hualin Xi, and Dr. Zhiping Weng) performed computational analysis of high-throughput sequencing data and tiling microarray data. Dr. Barry M. Honda's lab (Monika Syrzycka, Dr. Barry M. Honda) provided 3 potential ago3 deficiency strains. The UMass Medical School Deep sequencing core facility (Drs. Ellen L.W. Kittler and Maria L. Zapp) performed high-throughput sequencing. Dr. William E. Theurkauf's lab (Carla Klattenhoff, Nadine Schulz, and Dr. William E. Theurkauf) performed immunostaining in Figure 4.2. Dr. Vasily V. Vagin did northern hybridization analysis in Figure 4.3A and 4.3B. Dr. Vasily V. Vagin and Shengmei Ma carried out the initial screen of *ago3* tilling mutations. Dr. Hervé Seitz performed some preliminary computational analysis of high-throughput sequencing data. Dr. Michael D. Horwich made the Flag-Myc-tagged Ago3 construct. Dr. Vasily V. Vagin and I characterized $ago3^{t1}$, $ago3^{t2}$, and $ago3^{t3}$ EMS alleles as well as *ago3* deficiency allele, $ago3^{Df(3L)TTT}$. Shengmei Ma and I performed fertility analysis. I cloned small RNA libraries and performed the rest of the experiments in the figures.

Summary

Piwi-interacting RNAs (piRNAs) silence transposons in animal germ cells. piRNAs are thought to derive from long transcripts spanning transposon-rich genomic loci and to direct an auto-amplification loop in which an antisense piRNA, bound to Aubergine or Piwi protein, triggers production of a sense piRNA bound to the PIWI protein Argonaute3 (Ago3). In turn, the new piRNA is envisioned to produce a second antisense piRNA. Here, we describe strong loss-of-function mutations in *ago3*, allowing a direct genetic test of this model. We find that Ago3 acts to amplify piRNA pools and to enforce on them an antisense bias, increasing the number of piRNAs that can act to silence transposons. We also detect a second, Ago3-independent piRNA pathway centered on Piwi. Transposons targeted by this second pathway often reside in the *flamenco* locus, which is expressed in somatic ovarian follicle cells, suggesting a role for piRNAs beyond the germ line.

Introduction

The ability to tame transposons while retaining them in the genome is a particular specialty of eukaryotes. Transposons, repetitive sequences, and other forms of "selfish" DNA comprise as much as 42% of the human genome and nearly 30% of the genome of *Drosophila melanogaster*. In metazoa, transposons are silenced by the piRNA pathway, which is guided by 23–30 nt RNAs (Vagin et al., 2006; Brennecke et al., 2007).

The piRNA pathway is distinct from other RNA silencing pathways in that its small RNA guides are not produced by dicing. In contrast, both small interfering RNAs (siRNAs) and microRNAs (miRNAs) are cleaved by double-stranded RNA-specific endonucleases, Dicers, to yield double-stranded intermediates—siRNA or miRNA/miRNA* duplexes—that are loaded into members of the Argonaute family of proteins (reviewed in Ghildiyal and Zamore, 2009). piRNAs, too, act as guides for Argonaute proteins, but they appear not to exist as stable double-stranded intermediates at any point in their biogenesis (Vagin et al., 2006; Houwing et al., 2007). piRNAs bind PIWI proteins, a sub-family of Argonaute proteins that are expressed in germ-line cells (Das et al., 2008; Wang and Reinke, 2008; Vagin et al., 2006; Brennecke et al., 2007; Gunawardane et al., 2007; Lau et al., 2006; Girard et al., 2006; Aravin et al., 2006; Batista et al., 2008). PIWI proteins were first identified by their roles in maintaining (Cox et al., 1998; Cox et al., 2000) and patterning (Harris and Macdonald, 2001; Wilson et al., 1996) *Drosophila* germ cells. The defects in the organization of embryonic pattern in piRNA pathway mutations are likely an indirect consequence of their larger role in maintaining genomic stability

(Klattenhoff et al., 2007). The three *Drosophila* PIWI proteins, Piwi, Aubergine (Aub), and Argonaute3 (Ago3), are expressed in the male and female germ line cells (Nishida et al., 2007; Cox et al., 2000; Cox et al., 1998; Wilson et al., 1996; Gunawardane et al., 2007; Brennecke et al., 2007; Williams and Rubin, 2002).

The prevailing model for piRNA biogenesis—the “ping-pong” model—reflects the discovery that the first 10 nt of piRNAs bound to Aub or Piwi, which are largely antisense and typically begin with uridine, are often complementary to the first 10 nt of piRNAs bound to Ago3, which are largely sense and typically bear an adenosine at position 10 (Brennecke et al., 2007; Gunawardane et al., 2007). Many Argonaute proteins can act as RNA-guided RNA endonucleases, and all such Argonautes cut their target RNAs 5′ to the base that pairs with the tenth nucleotide of the small RNA guide; all three fly PIWI proteins retain their endonuclease activity (Gunawardane et al., 2007; Saito et al., 2006). Thus, the observed 10 nt 5′ complementarity between piRNAs suggests that the 5′ ends of piRNAs bound to Aub or Piwi are defined by Ago3-catalyzed cleavage, and, reciprocally, that the 5′ ends of piRNAs bound to Ago3 are defined by piRNAs bound to Aub or Piwi. The ping-pong model seeks to explain these observations, as well as the role of piRNA cluster transcripts in piRNA biogenesis, the function of piRNAs in silencing transposons, and the extraordinary antisense bias of piRNAs generally. At its core, the model proposes that piRNAs participate in an amplification loop in which transposon sense transcripts (e.g., transposon mRNAs) trigger the production of new, antisense piRNAs. Ago3, guided by sense piRNAs, lies at the heart of the amplification loop.

To test for the ping-pong model, we isolated strong loss-of-function mutations in *ago3*. Here, we report the detailed analysis of *ago3* and *aub* mutant

flies. Our data provides strong support for an amplification cycle in which Ago3 amplifies piRNA pools and enforces on them a strong antisense bias, increasing the number of piRNAs that can act to destroy transposon mRNAs. Moreover, we detect a second, perhaps somatic, piRNA pathway centered on Piwi and functioning without benefit of Ago3-catalyzed amplification. Most of the transposons targeted by this second pathway reside in the *flamenco* piRNA cluster, which was first identified as a repressor of transposon expression in somatic follicle cells (Pelisson et al., 1994).

Results

Loss-of-function *ago3* alleles

The *Drosophila ago3* gene resides in pericentromeric heterochromatin on the left arm of chromosome 3 (Figure 4.1A). Heterochromatin is refractory to *P*-element or *piggyBac* transposon insertion and to targeted recombination, complicating isolation of *ago3* mutants. We used Targeting Induced Local Lesions in Genomes (TILLING) (Cooper et al., 2008) to identify three mutant alleles (*t1*, *t2*, and *t3*) that create premature stop codons in *ago3* (Figure 4.1B) and obtained one chromosomal deficiency (*Df(3L)TTT*) that deletes at least six genes, including *ago3* (Figure 4.S1A). One homozygous ($ago3^{t3}/ago3^{t3}$) and three trans-heterozygous ($ago3^{t1}/ago3^{t2}$, $ago3^{t2}/ago3^{t3}$, and $ago3^{t1}/ago3^{t3}$) combinations, as well as each mutant allele in trans to *Df(3L)TTT*, were viable. Full-length Ago3 protein was essentially undetectable in all seven allelic combinations (Figure 4.1C and Figure 4.S1B), and $ago3^{t2}$/TM6B or $ago3^{t3}$/TM6B heterozygous ovaries contained half as much Ago3 protein as Oregon R wild-type ovaries (Figure 4.1D), suggesting that the *ago3* alleles correspond to strong loss-of-function mutations. We refer to $ago3^{t2}/ago3^{t3}$ trans-heterozygotes as *ago3* for brevity.

Mutually interdependent localization of PIWI proteins

The intracellular localization of each PIWI protein appears to require other PIWI proteins. In the *Drosophila* female germ line, Piwi is largely in the nucleus, whereas Aub and Ago3 are cytoplasmic and concentrated in “nuage,” a ring

Figure 4.1.

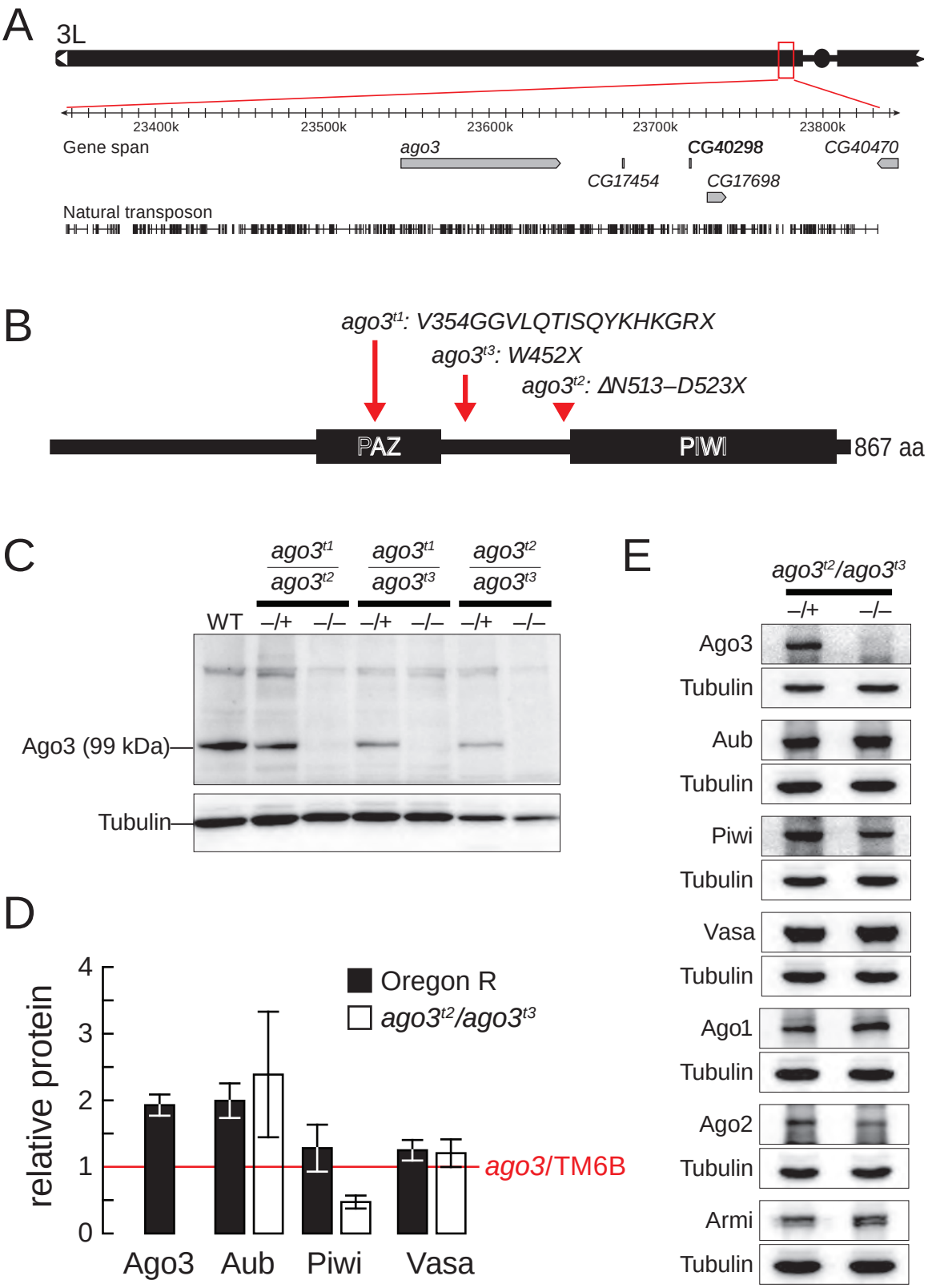

A
3L
23400k
23500k
23600k
23700k
23800k
Gene span
ago3
CG40298
CG40470
CG17454
CG17698
Natural transposon
B
ago3t1: V354GGVLQTISQYKHKGRX
ago3t3: W452X
ago3t2: ΔN513–D523X
PAZ
PIWI
867 aa
C
ago3t1/ago3t2
ago3t1/ago3t3
ago3t2/ago3t3
WT
–/+
–/–
–/+
–/–
–/+
–/–
Ago3 (99 kDa)
Tubulin
D
relative protein
4
3
2
1
0
Oregon R
ago3t2/ago3t3
ago3/TM6B
Ago3
Aub
Piwi
Vasa
E
ago3t2/ago3t3
–/+
–/–
Ago3
Tubulin
Aub
Tubulin
Piwi
Tubulin
Vasa
Tubulin
Ago1
Tubulin
Ago2
Tubulin
Armi
Tubulin

Figure 4.1. *ago3* mutants.

(A) The *ago3* gene resides in pericentromeric heterochromatin on the left arm of chromosome 3. (B) *ago3* mutant alleles were identified by TILLING. (C) Full-length Ago3 protein was not detected in trans-heterozygous *ago3* ovaries, but was readily detected in heterozygotes and wild-type. (D) Protein levels in wild-type and *ago3* ovaries, relative to *ago3/*TM6B (red line). The average ± standard deviation for at least three independent biological samples is shown. (E) Representative data for Aub, Piwi, Vasa, Argonaute1 (Ago1), Argonaute2 (Ago2), and Armitage (Armi) in *ago3* ovaries. Tubulin served as a loading control.

around the cytoplasmic face of the nuclei of the transcriptionally active germ-line nurse cells (Findley et al., 2003; Pane et al., 2007; Brennecke et al., 2007). The putative RNA helicase Vasa, which marks the germ-line in most animals, is also in nuage (Liang et al., 1994). Mutations that disrupt piRNA biogenesis, but not those that block siRNA or miRNA production, disrupt localization of Aub to nuage (Lim and Kai, 2007).

Ago3 is also required for Aub to localize to nuage (Figure 4.2A). However, Ago3 is not required for Aub expression or stability: the concentration of Aub protein and the amount of Aub in the cytoplasm increased in *ago3* ovaries relative to their heterozygous siblings (Figure 4.1D and 4.1E). Reciprocally, Aub plays a role in the localization of Ago3 to nuage, although some Ago3 persisted in nuage in an *aub* hypomorphic mutant allelic combination (aub^{HN2}/aub^{QC42}). The function of nuage is unknown, but it may have a complex substructure, because the localization of Vasa within nuage only partially coincides with that of Ago3 (Figure 4.2B) and Aub (Figure 4.2C).

The interdependence of Ago3 and Aub for their localization to nuage likely reflects an underlying requirement for these two proteins for nuage assembly: the peri-nuclear localization of Vasa, a marker for nuage, was absent in both *ago3* and *aub* mutants, although the abundance of Vasa was unaltered in *ago3* (Figure 4.1D and 1E). The abundance of Armitage, Argonaute1, and Argonaute2, proteins required for small silencing RNA biogenesis or function, were similarly unaltered in *ago3* mutants (Figure 4.1E). In *ago3*/TM6B heterozygotes, which produce half as much Ago3 protein as wild-type, and in *aub*/CyO heterozygotes, less Aub, Ago3, or Vasa was present in nuage than in wild-type, suggesting that

Figure 4.2.

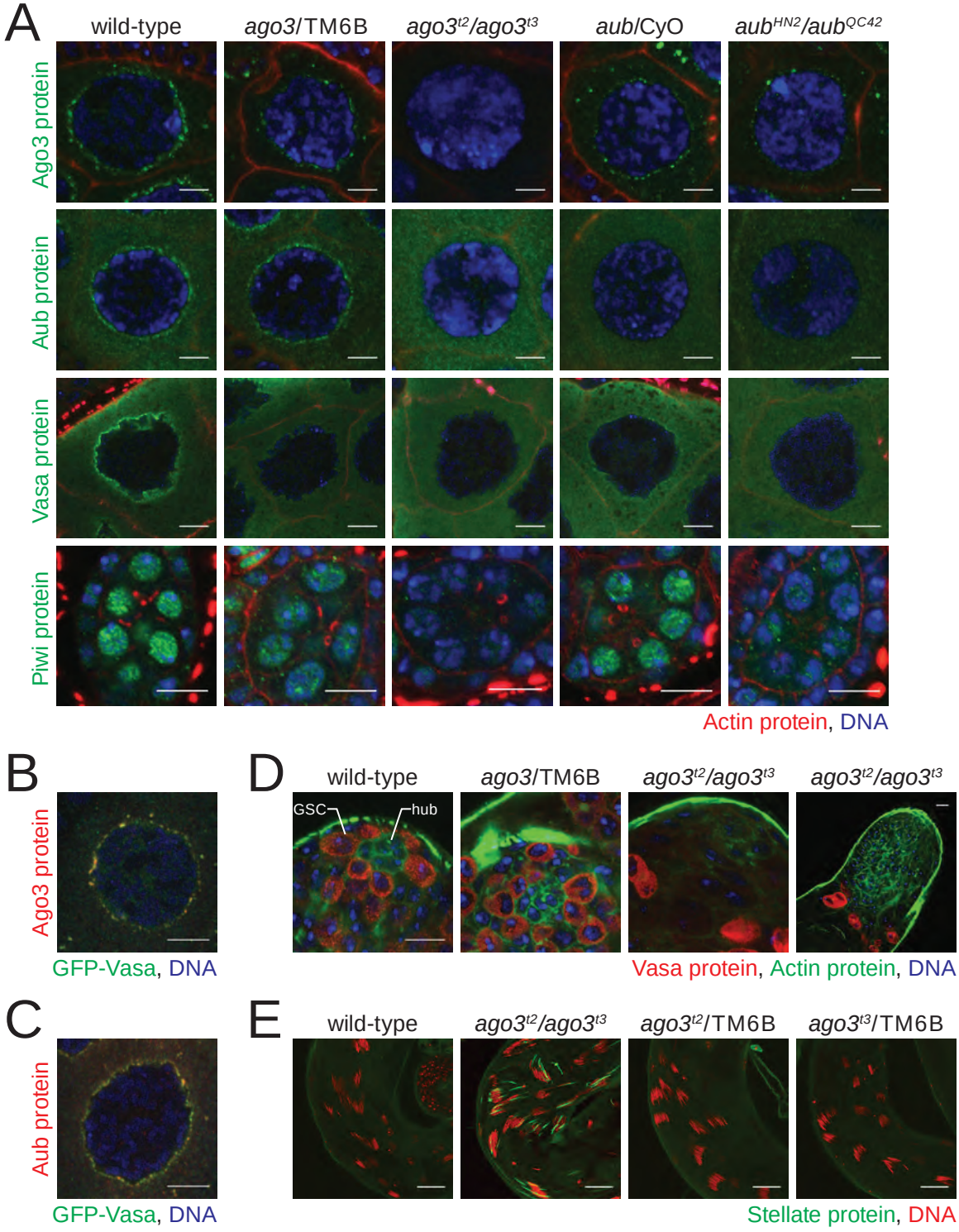

A
wild-type
ago3/TM6B
ago3t2/ago3t3
aub/CyO
aubHN2/aubQC42
Ago3 protein
Aub protein
Vasa protein
Piwi protein
Actin protein, DNA
B
Ago3 protein
GFP-Vasa, DNA
C
Aub protein
GFP-Vasa, DNA
D
wild-type
ago3/TM6B
ago3t2/ago3t3
ago3t2/ago3t3
GSC
hub
Vasa protein, Actin protein, DNA
E
wild-type
ago3t2/ago3t3
ago3t2/TM6B
ago3t3/TM6B
Stellate protein, DNA

Figure 4.2. Consequences of loss of Ago3 in the *Drosophila* germ line. (A) Mutually interdependent incorporation of Ago3 and Aub, into nuage in ovaries. Images correspond to single confocal sections (63x magnification); scale bar: 5 μm (Ago3, Aub, and Vasa panels) or 10 μm (Piwi). (B) and (C) Vasa localization within nuage is only partially congruent with that of Ago3 and Aub. Images correspond to single confocal sections of stage 4/5 egg chambers (63x lens, 4x zoom). Scale bars: 10 μm. (D) In 5–7 day old male adults, Vasa-expressing germ-line stem cells (GSC), which normally surround the somatic hub cells (hub), were not detected in *ago3* testes. Since Vasa expression increases in *ago3*, the two images from *ago3* mutants were acquired using reduced gain relative to wild-type and heterozygous testes. Scale bar: 10 μm (first three images) or 20 μm (fourth). (E) *Stellate* silencing in testes requires Ago3. Scale bars: 20 μm.

assembly of nuage is exquisitely sensitive to the amount of Ago3 and Aub (Figure 4.2A).

The peak of nuclear-localized Piwi protein expression occurs earlier in oogenesis than when Ago3 and Aub are maximally perinuclear (Cox et al., 2000). Less Piwi is found in the nucleoplasm of *ago3*/TM6B or *aub*/CyO heterozygotes than in wild-type, and little or no Piwi is present in the nuclei of the *ago3* or *aub* homozygotes (Figure 4.2A); *ago3* ovaries contain about half as much Piwi as their heterozygous siblings (Figure 4.1E). (The localization of Aub and Ago3 cannot be studied in *piwi*, because *piwi* mutants arrest early in oogenesis.) We conclude that the correct intracellular localization of Ago3, Aub, and Piwi is inter-dependent.

As reported previously for *aub* testes, Vasa protein expression increased in male germ cells in *ago3* testes, relative to wild-type or *ago3*/TM6B, consistent with the proposal that the AT-chX-1 and AT-chX-2 piRNAs down-regulate *vasa* expression (Nishida et al., 2007); AT-chX-1 piRNA levels were dramatically reduced in *ago3* testes (Figure 4.3A).

ago3 mutations affect fecundity

ago3 females are sterile. *ago3* females initially laid far fewer eggs than their heterozygous siblings and by day 10 stopped laying altogether (Figure 4.S2A). The egg-laying deficit was rescued by a single-copy transgene expressing Ago3 in the germ line or in both the germ line and the soma. Of those eggs laid, none of the embryos hatched, compared with ~90% for wild-type (Table 4.S1). For 29%-67% of the embryos produced by *ago3* mothers, the dorsal appendages were fused or absent (Table 4.S1), indicating a maternally

Figure 4.3.

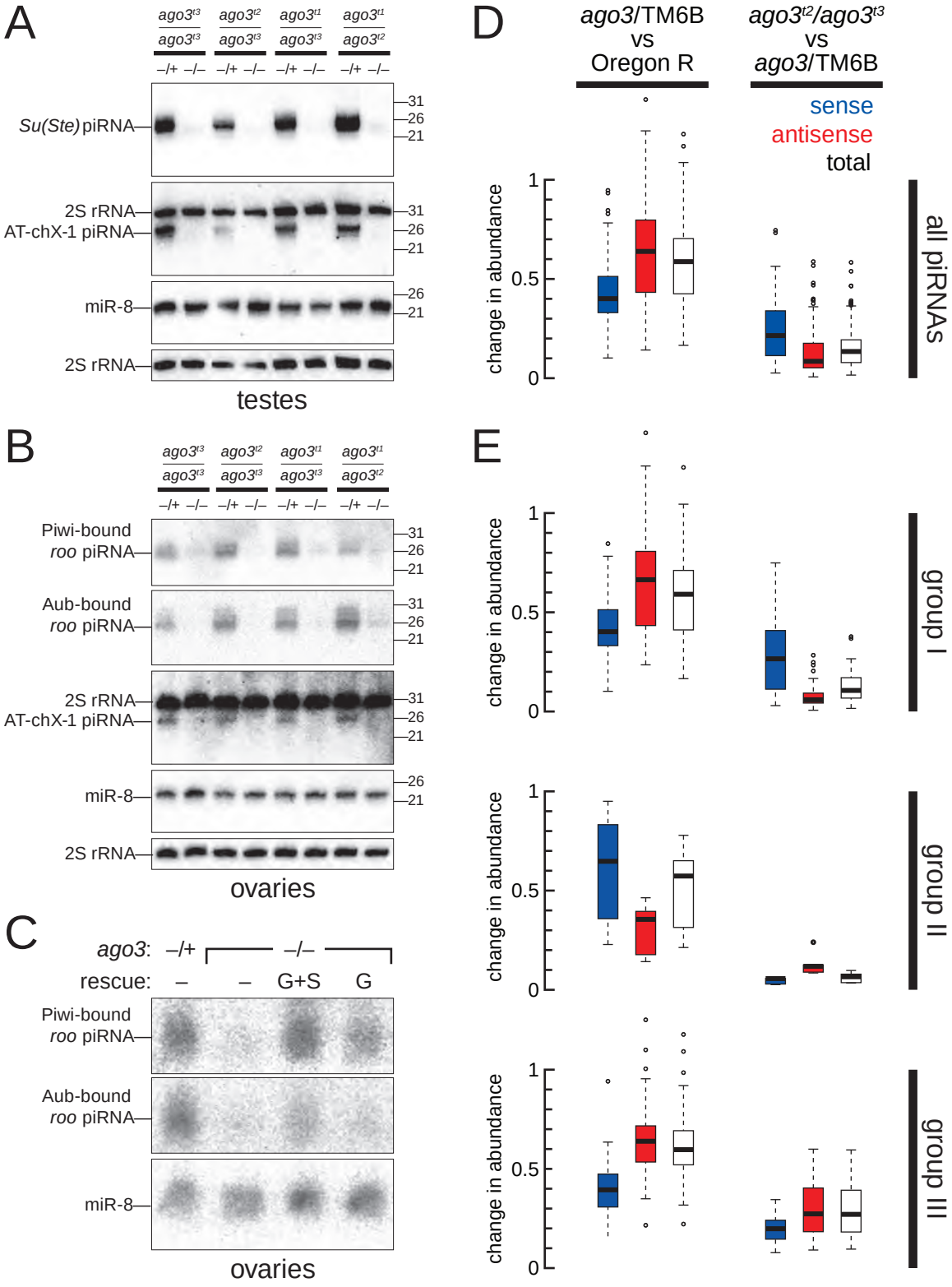

A
ago3t3/ago3t3
ago3t2/ago3t3
ago3t1/ago3t3
ago3t1/ago3t2
–/+ –/–
Su(Ste) piRNA
2S rRNA
AT-chX-1 piRNA
miR-8
2S rRNA
31
26
21
testes
B
Piwi-bound roo piRNA
Aub-bound roo piRNA
2S rRNA
AT-chX-1 piRNA
miR-8
2S rRNA
ovaries
C
ago3: –/+ –/–
rescue: – – G+S G
Piwi-bound roo piRNA
Aub-bound roo piRNA
miR-8
ovaries
D
ago3/TM6B vs Oregon R
ago3t2/ago3t3 vs ago3/TM6B
sense
antisense
total
change in abundance
1
0.5
0
all piRNAs
E
group I
group II
group III

Figure 4.3. Aub- and Piwi-bound piRNAs disappear without Ago3.
(A) Accumulation of *Su(Ste)* and AT-chX-1 piRNAs in testes required Ago3. (B) Accumulation of *roo* antisense piRNAs, as well as the AT-chX-1 piRNA, in ovaries required Ago3. (C) Accumulation of Piwi- and Aub-bound *roo* antisense piRNAs in the ovary was rescued in *ago3* mutants by a single-copy transgene expressing Ago3 in the germ line (G) or in both the germ line and the soma (G+S). (D) Box plots illustrating the change in abundance of piRNAs, analyzed by transposon family, in *ago3*/TM6B versus Oregon R (left panel) and *ago3* versus *ago3*/TM6B (right panel) ovaries. (E) Box plots illustrating the change in abundance for piRNAs for each of 95 transposon families, separated by group.

inherited patterning defect, as has been reported for other piRNA pathway mutations (Schupbach and Wieschaus, 1991; Wilson et al., 1996; Cook et al., 2004; Klattenhoff et al., 2007; Pane et al., 2007) Loss of Ago3 in the germ line, rather than a secondary mutation present in the *ago3* mutants, caused the patterning defects, as a single-copy transgene expressing Ago3 from the germ line-specific *nanos* promoter rescued the dorsal appendage defects to essentially wild-type rates (Table 4.S1).

In males, Ago3 is required to maintain germ-line stem cells (Figure 4.2D). We examined testes from 5–7 day old wild-type, *ago3/*TM6B, and *ago3* males, using Vasa protein expression to identify germ-line cells. In wild-type and *ago3/*TM6B, Vasa-expressing cells were present at the tip of the testes, in a ring around the hub cells, a group of somatic cells that support the adjacent germ-line stem cells (Fuller, 1993). In contrast, the corresponding region in *ago3* was devoid of Vasa-expressing cells, suggesting that germ-line stem cells are not properly maintained in the absence of Ago3. Consistent with a failure to maintain germ-line stem cells, *ago3* males are semi-fertile, siring fewer offspring than males with the same genetic background (Figure 4.S2B).

Silencing selfish genetic elements in germ line requires Ago3

In males, antisense piRNAs derived from the *Suppressor of Stellate* (*Su(Ste)*) locus on the Y chromosome silence the X-linked *Stellate* locus during spermatogenesis (Aravin et al., 2001; Vagin et al., 2006; Aravin et al., 2003; Bozzetti et al., 1995; Palumbo et al., 1994; Balakireva et al., 1992). *Su(Ste)* piRNAs were the first piRNAs to be identified (Aravin et al., 2001), and, in *Su(Ste)* mutants, piRNAs targeting *Stellate* are lost causing Stellate protein

crystals to form in primary spermatocytes (Livak, 1990; Livak, 1984; Pimpinelli et al., 1985). *Stellate* silencing requires the piRNA pathway genes, *aub*, *spindle* E, *armitage*, *squash*, and *zucchini* (Schmidt et al., 1999; Stapleton et al., 2001; Tomari et al., 2004; Pane et al., 2007). Stellate protein crystals form in primary spermatocytes in *ago3* testes (Figure 4.2E).

Virtually all *Su(Ste)* piRNAs are antisense (Vagin et al., 2006) and bound to Aub, not Piwi (Nishida et al., 2007). *Su(Ste)* piRNAs fail to accumulate in *aub* mutants, but accumulate to higher than normal levels in *piwi* testes (Vagin et al., 2006). We used Northern hybridization to examine *Su(Ste)* piRNA production in the testes (Figure 4.3A). *Su(Ste)* piRNAs, as well as AT-chX-1 piRNA, were strongly reduced in *ago3* males, relative to their heterozygous siblings (Figure 4.3A). Thus, both Aub and Ago3 are required to silence *Stellate*, and both proteins are required to produce or stabilize those *Su(Ste)* piRNAs normally bound to Aub. We speculate that *Stellate*-derived, Ago3-bound sense piRNAs, while rare, amplify Aub-bound, antisense *Su(Ste)* piRNAs in testes.

piRNAs have been best characterized in the *Drosophila* ovary. piRNAs derived from *roo* LTR retrotransposons are among the most abundant ovarian piRNAs and are disproportionately antisense to *roo* coding sequences; these antisense piRNAs are bound to Aub and Piwi (Vagin et al., 2006; Gunawardane et al., 2007), but not to Ago3 (Brennecke et al., 2007). *roo* antisense piRNAs failed to accumulate in *ago3* ovaries, but were readily detectable in *ago3*/TM6B ovaries (Figure 4.3B). Production of *roo*-derived piRNAs was rescued by a single-copy transgene expressing Ago3 from the *actin5c* promoter and more weakly rescued by a single-copy transgene expressing Ago3 from the *nanos* promoter (Figure 4.3C).

Genome-wide piRNA analysis

To obtain a broader view of the function of Ago3 in piRNA biogenesis, we sequenced piRNAs from wild-type (Oregon R), *ago3*/TM6B, $ago3^{t2}/ago3^{t3}$, *aub*/CyO, and aub^{HN2}/aub^{QC42} ovaries. Libraries were prepared from oxidized small RNAs to permit greater effective sequencing depth. In all, 3,282,391 genome-matching, non-ncRNA, non-miRNA, 23–29 nt long small RNAs were sequenced (Table 4.S2). Despite the large number of sequences obtained from each genotype, most piRNA species were sequenced only once. This remained true even when all data sets were pooled, suggesting that piRNAs comprise the most diverse class of regulatory molecules in the fly. Together, our piRNA sequences from wild-type, *ago3*/TM6B and *ago3* ovaries cover 10% of the fly genome.

In parallel, we immunoprecipitated Ago3, Aub, and Piwi from *ago3*/TM6B ovaries and Aub and Piwi from *ago3* ovaries, and then constructed and sequenced libraries of the small RNAs bound to each protein (Table 4.S2). We do not know the extent to which PIWI proteins co-purify, e.g., because they are bound to a common RNA or present in a common complex. To avoid potential mis-assignment of piRNAs to a specific PIWI protein, we analyzed only those piRNAs that associated uniquely with Ago3, Aub, or Piwi.

We calibrated the abundance of the piRNAs uniquely associated with Ago3, Aub, or Piwi such that the aggregate abundance of a subset of piRNAs in an immunoprecipitation dataset equaled that in the total piRNA data set from the corresponding genotype. This subset of piRNAs was defined as those that mapped only once to the fly genome and were sequenced at least once in both

of the two data sets. This strategy allows direct comparison of the relative abundance of piRNAs uniquely bound to one PIWI protein with those bound to another, as well as comparison of the uniquely bound piRNAs between *ago3* heterozygotes and mutants.

Ago3 limits sense piRNA accumulation and amplifies antisense piRNAs

Both *ago3* and *ago3/*TM6B ovaries had fewer piRNAs than wild-type (Figure 4.3D). Ago3 was previously found to bind mainly piRNAs corresponding to the sense, mRNA strand of transposons. Our data suggest that the intracellular concentration of Ago3 limits the accumulation of sense piRNAs: the median abundance sense piRNAs, analyzed by transposon family, in *ago3/*TM6B heterozygotes was ~40% of wild-type. Antisense piRNAs were less affected by halving the amount of Ago3: the median abundance of antisense, transposon-mapping piRNAs in *ago3/*TM6B ovaries was ~64% of wild-type. While complete loss of Ago3 further depressed the abundance of transposon-mapping, sense piRNAs, antisense piRNA accumulation in *ago3* ovaries collapsed. The median abundance by transposon family for antisense piRNAs in *ago3* ovaries was less than a tenth of the median abundance in the heterozygotes, and less than one-twentieth the median abundance in wild-type. The data support the view that for most selfish genetic element families, Ago3, presumably guided by sense piRNAs, acts to amplify antisense piRNAs bound to Aub.

Immunoprecipitation data confirmed this idea. In *ago3/*TM6B ovaries, 71% of small RNAs uniquely bound to Aub were antisense, but in *ago3* mutants, only 41% of the remaining Aub-bound piRNAs were antisense (Table 4.S3). Thus, the characteristic antisense bias of Aub-bound piRNAs is enforced by Ago3. The

strand bias of Piwi-bound piRNAs was less affected by loss of Ago3: in the *ago3* heterozygotes, 73% of the piRNAs uniquely associated with Piwi were antisense; in the mutants, 62% of the remaining piRNAs uniquely bound to Piwi were antisense. That the strandedness of Piwi-bound piRNAs changes at all suggests that Ago3 plays a role in the production of at least some antisense, Piwi-bound piRNAs.

Three piRNA groups

While antisense piRNAs are generally more abundant than sense piRNAs, individual transposon families have distinct ratios of sense and antisense piRNAs. To determine the role of Ago3 in establishing these ratios, we analyzed the fraction of sense piRNAs (sense/(sense + antisense)) for each of the 95 transposon families for which we obtained ≥ 500 reads in *ago3*/TM6B ovaries. We compared the fraction of sense piRNAs in the heterozygotes to the fraction of sense piRNAs in *ago3* mutants (Figure 4.4A and 4.4B). We also compared the fraction of sense piRNAs in the heterozygotes to both the fraction of sense piRNAs in *ago3* mutants and the fold-change in antisense piRNAs between the two genotypes (Figure 4.S3).

We detected three groups of transposons. For 63 transposon families, the abundance of antisense piRNAs declined dramatically, causing an increase in the fraction of sense piRNAs (Figure 4.S3). Among these 63 group I transposon families, 61 also had a decreased antisense bias comparing *ago3* to Oregon R (Figure 4.4C). (For convenience, the enigmatic transposon family *1360* was included in group I in our subsequent analyses.)

Figure 4.4.

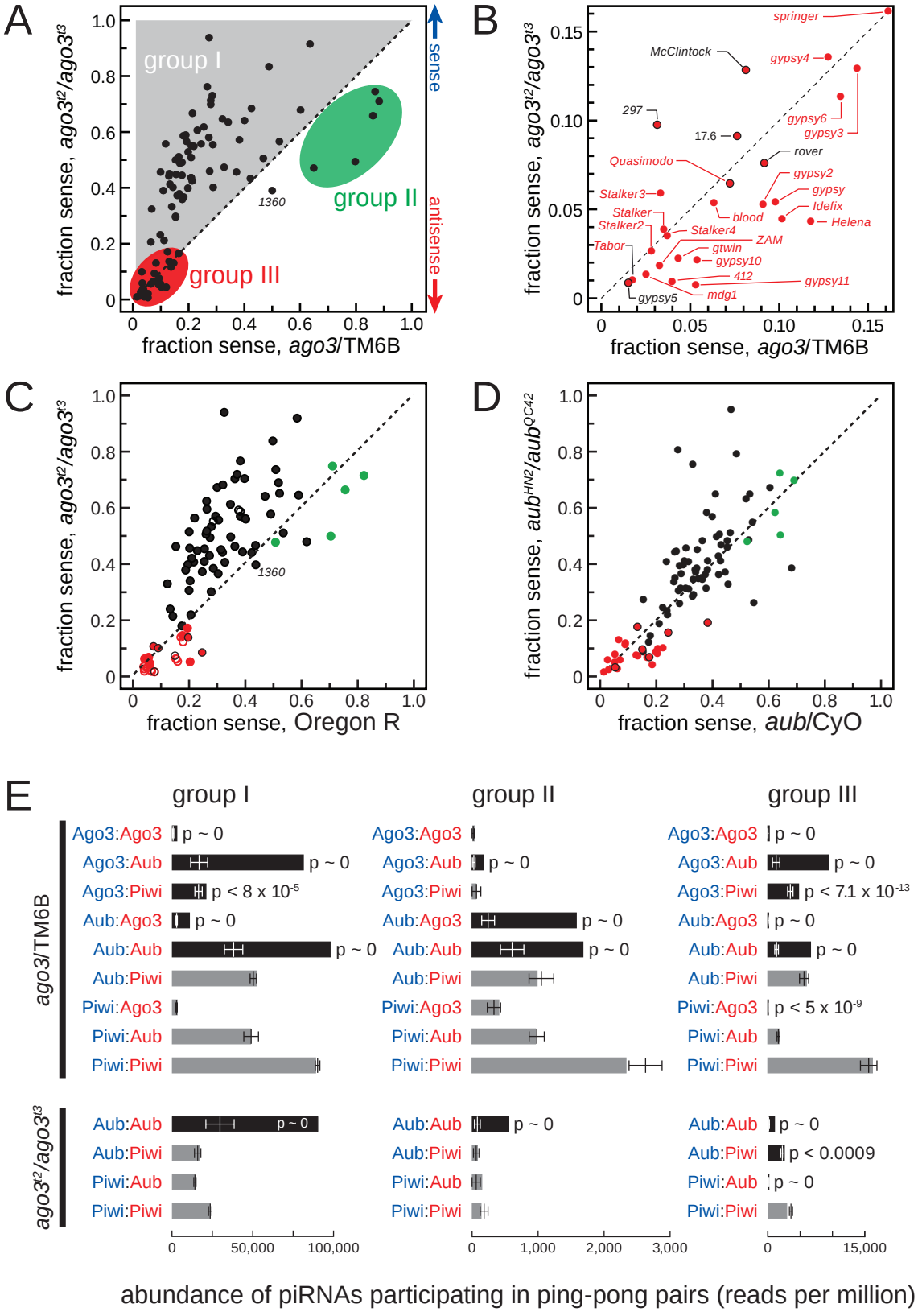

A
fraction sense, ago3t2/ago3t3
fraction sense, ago3/TM6B
group I
group II
group III
sense
antisense
1360
B
fraction sense, ago3t2/ago3t3
fraction sense, ago3/TM6B
springer
McClintock
gypsy4
297
gypsy6
gypsy3
17.6
rover
Quasimodo
gypsy2
Stalker3
gypsy
Idefix
Stalker
blood
Helena
Stalker2
Stalker4
ZAM
Tabor
gtwin
gypsy10
412
gypsy11
gypsy5
mdg1
C
fraction sense, ago3t2/ago3t3
fraction sense, Oregon R
1360
D
fraction sense, aubHN2/aubQC42
fraction sense, aub/CyO
E
group I
group II
group III
ago3/TM6B
Ago3:Ago3
Ago3:Aub
Ago3:Piwi
Aub:Ago3
Aub:Aub
Aub:Piwi
Piwi:Ago3
Piwi:Aub
Piwi:Piwi
p ~ 0
p < 8 x 10-5
p < 7.1 x 10-13
p < 5 x 10-9
ago3t2/ago3t3
Aub:Aub
Aub:Piwi
Piwi:Aub
Piwi:Piwi
p ~ 0
p < 0.0009
0
50,000
100,000
1,000
2,000
3,000
15,000
abundance of piRNAs participating in ping-pong pairs (reads per million)

Figure 4.4. Three distinct groups of piRNAs.

(A) The sense fraction of the piRNAs was compared for each transposon family between *ago3* and *ago3/*TM6B ovaries. (B) An enlargement of the lower left corner of (A). Transposons labeled in red are present in the *flamenco* locus. Here and in (C) and (D), group III transposon families not present in *flamenco* are denoted with a red-filled black circle. (C) The sense fraction of piRNAs was compared between *ago3* and Oregon R ovaries. (D) The sense fraction of the piRNAs was compared between *aub* and *aub*/CyO ovaries. (E) The normalized abundance of ping-pong pairs detected for all nine (*ago3/*TM6B) or four (*ago3*) possible PIWI protein pairings for the piRNA species uniquely associated with a single PIWI protein. Black bars, Bonferroni-corrected p-value < 0.005; gray bars, p-value > 0.005. p-values < 10^{-30} are reported as $p \sim 0$. Values are average ± two standard deviations.

Five transposon families compose group II. These all had more sense than antisense piRNAs in *ago3* heterozygotes, and four of the five had more sense than antisense piRNAs in wild-type ovaries, as previously noted (Brennecke et al., 2007). In *ago3* ovaries, both the fraction (Figure 4.4A, 4.4C, and 4.S3) and the absolute amount (Figure 4.3E) of sense piRNAs declined for group II transposons, relative to heterozygous or wild-type ovaries.

piRNAs mapping to the 26 group III transposons were disproportionately antisense, ranging from 80 to nearly 100 percent antisense in *ago3* heterozygotes (Figure 4.4A and 4.4B). In *ago3* ovaries, group III transposons generally retained their antisense piRNAs to a greater extent than those in groups I and II (Figure 4.3E and 4.S3).

Paradigmatic examples of each group—*HeT-A* for group I, *accord2* for group II, and *ZAM* for group III—are analyzed in greater detail in Figure 4.5. Figure 4.S4, 4.S5, and 4.S6 present the corresponding data for each of the 95 transposons, as well as the tandem repeat *mst40*.

Group I transposons require Ago3 for antisense piRNA amplification

HeT-A is the quintessential group I transposon (Figure 4.5). The piRNAs that map to group I transposons show a strong “ping-pong” signature that derives from the 10 nt overlap between antisense piRNAs bound to Aub and sense piRNAs bound to Ago3 and Aub (Figure 4.4E). For group I transposons, although Ago3 associates almost exclusively with sense piRNAs, it is the Aub-bound antisense piRNAs that disappear in its absence (Figure 4.S6). A central postulate of the ping-pong model is that the action of Ago3, rather than Aub, drives the production of more antisense than sense piRNAs. Consistent with this proposal,

Figure 4.5.

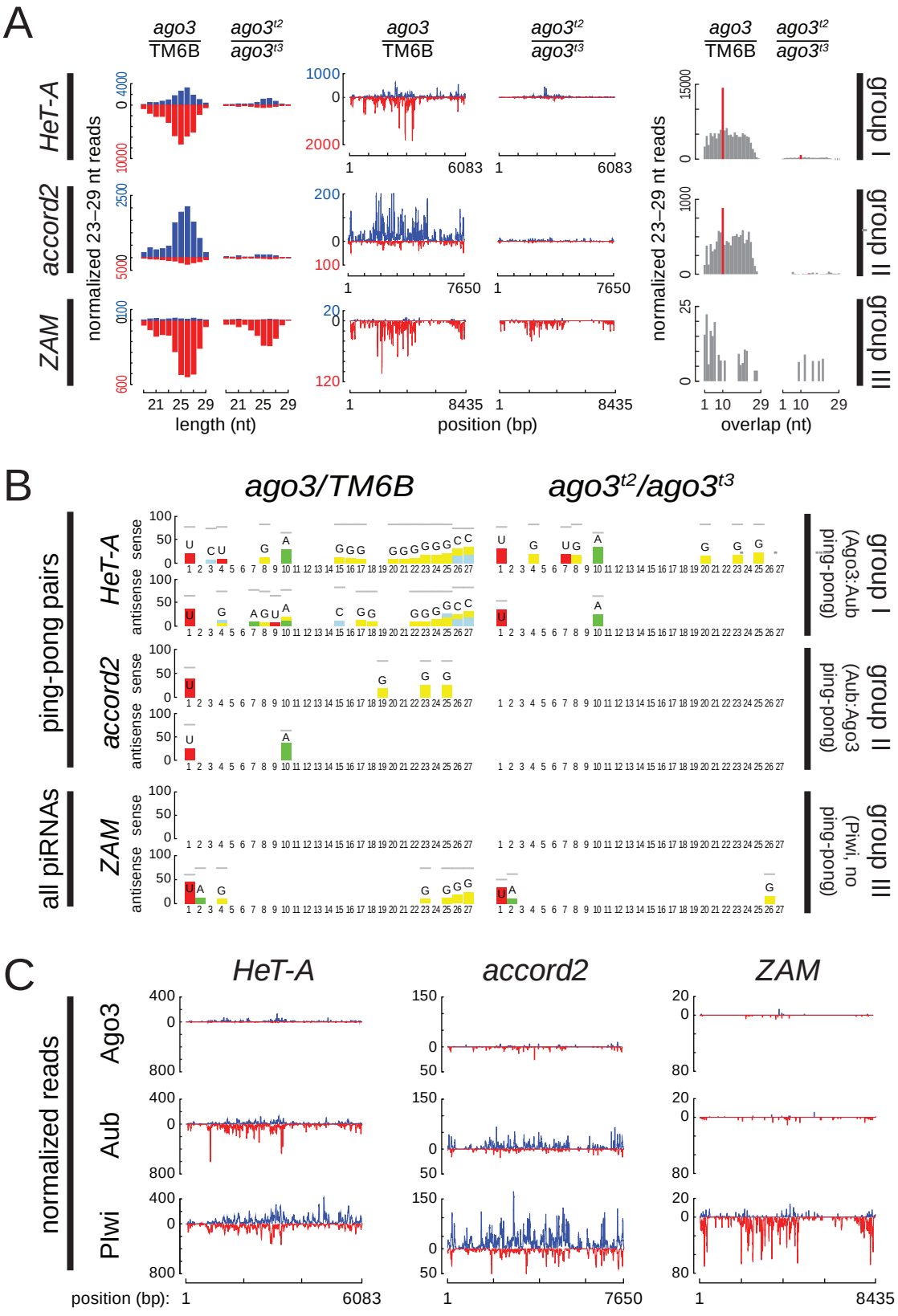

A
ago3 / TM6B
ago3t2 / ago3t3
HeT-A
accord2
ZAM
normalized 23–29 nt reads
length (nt)
position (bp)
overlap (nt)
group I
group II
group III
B
ago3/TM6B
ago3t2/ago3t3
ping-pong pairs
all piRNAs
sense
antisense
group I (Ago3:Aub ping-pong)
group II (Aub:Ago3 ping-pong)
group III (Piwi, no ping-pong)
C
HeT-A
accord2
ZAM
normalized reads
Ago3
Aub
Piwi
position (bp):

Figure 4.5. Paradigmatic examples of each transposon group.

(A) piRNA length distribution, abundance relative to consensus position, and ping-pong pair abundance. Blue, sense piRNAs; red, antisense. (B) Sequence-excess logos generated by subtracting the background from the relative frequency of each nucleotide at each position of the 23–29 nt RNAs in each sample. Only the nucleotide positions where foreground was significantly higher than background are shown (Bonferroni-corrected *p*-values < 0.001). Gray horizontal bars indicate the maximum possible value for each position. (C) Normalized and calibrated abundance of piRNAs uniquely associated with Ago3, Aub, or Piwi in *ago3/*TM6B or *ago3* ovaries.

the fraction of sense piRNAs for most group I transposons shows little overall change in *aub* ovaries (Figure 4.4D).

Group II transposons act “backwards”

For group II transposons, antisense piRNA levels appear to reflect the abundance of Ago3: the median abundance by transposon family of group II antisense piRNAs in *ago3* heterozygotes was ~35% of wild-type (Figure 4.3E). Sense piRNAs declined less when the abundance of Ago3 was halved. However, in the absence of Ago3, group II sense piRNAs declined ~17-fold from their levels in heterozygotes and ~27-fold from their levels in wild type. Group II antisense piRNAs also decreased, but less dramatically. The “backwards” behavior of group II piRNAs, relative to group I, suggests that the production or accumulation of Aub-bound piRNAs generally requires Ago3, irrespective of their sense or antisense identity.

piRNAs from group I and group II transposons normally partition among PIWI proteins according to their orientation and sequence bias (Figure 4.S4, 4.S5, and 4.S6). Group I antisense (or group II sense) piRNAs typically begin with U (U1) and bind Aub, whereas group I sense (or group II antisense) piRNAs show less 5′ nucleotide bias but typically bear an A at position 10 (A10) and bind Ago3 (e.g., *accord2* in Figure 4.5). In the absence of Ago3, both sense and antisense piRNAs remained associated with Aub (Figure 4.S6). Some of these showed ping-pong pairing, but without respect to orientation, so that the U1 that normally characterizes antisense piRNAs and the A10 that normally characterizes sense piRNAs became conflated (Figure 4.5B and Figure 4.S4). In the absence of Ago3, those Aub-bound piRNAs with an A at position 10 began

with U more often than expected by chance (Figure 4.S7) and began with U about as often as did those Aub-bound piRNAs without an A at position 10. These observations suggest that a 5′ U favors binding of a piRNA to Aub. Moreover, the data suggest that the A10 signature of Ago3-bound piRNAs is a consequence of the preponderance of Aub-bound piRNAs that start with U, consistent with the idea that Aub directs the production of Ago3-bound piRNAs.

piRNAs for group III transposons are produced by both Aub- and Ago3-dependent and Aub- and Ago3-independent pathways

In the absence of Ago3, the abundance of group III antisense piRNAs decreased far less than for group I (group I median = 0.06, group III median = 0.27; Figure 4.3E and S3). Antisense piRNAs from group III transposon families were mostly bound to Piwi: the median ratio between the amount of antisense piRNAs bound by Piwi and the amount bound by Aub in *ago3* ovaries was 31.3 for group III transposon families, a five-fold increase from the median Piwi/Aub ratio for group III transposons in *ago3*/TM6B ovaries. This contrasts with the median Piwi/Aub ratios for groups I (2.00) or II (2.94) transposon families (Figure 4.S8A). Furthermore, loss of Ago3 caused group III antisense piRNAs bound to Piwi to decrease far less than antisense piRNAs bound to Aub; for group I, piRNAs bound to Piwi and Aub declined to a comparable extent (Figure 4.S8B). Group III antisense piRNAs were also less affected by loss of Aub than the other two groups (Figure 4.S9 and 4.S10). Collectively, these observations suggest that group III transposons are predominantly silenced by a Piwi-dependent, Ago3-independent pathway.

While Piwi-bound piRNAs persisted in the absence of Ago3 to a greater extent than Aub-bound piRNAs, the absence of Ago3 clearly reduced the abundance of Piwi-bound piRNAs for all groups (Figure 4.S8B). Perhaps some but not all Piwi-bound piRNAs are generated in the germ line by an Ago3-directed amplification cycle. Supporting this view, we detected a statistically significant ($p < 7.1 \times 10^{-13}$) 10 nt overlap between sense Ago3-bound piRNAs and antisense Piwi-bound group III piRNAs (Figure 4.4E). Such group III Ago3 (sense):Piwi (antisense) piRNA ping-pong pairs were about half as abundant as Ago3:Aub pairs.

Most group III antisense piRNAs bound to Aub were eliminated in *ago3* mutants (Figure 4.S8), suggesting that our group III piRNA data conflates at least two distinct pathways: a Piwi-dependent, Ago3- and Aub-independent pathway in which piRNA-directed piRNA amplification plays a little if any role and an Aub-dependent pathway in which sense piRNAs bound to Ago3 act catalytically to amplify the antisense piRNAs associated with Aub. piRNAs acting in the Piwi-dependent pathway were disproportionately antisense, and this antisense bias required neither Ago3 nor Aub: production of antisense piRNAs for group III transposons remained essentially unchanged in aub^{HN2}/aub^{QC42} mutants (Figure 4.4D).

Group III transposons often reside in the *flamenco* piRNA cluster

Remarkably, of the 26 transposon families comprising group III, 20 are present in the *flamenco* locus (Pelisson et al., 1994; Brennecke et al., 2007). *flamenco* was originally identified as required for the silencing of *gypsy* (Prud'homme et al., 1995) and later shown to be required to silence *ZAM* and

Idefix (Mevel-Ninio et al., 2007; Desset et al., 2008). *gypsy*, *ZAM*, and *idefix* are all group III transposons. Reporter experiments suggest that Piwi is required to silence these transposons (Sarot et al., 2004; Desset et al., 2008). Given that Piwi, but not Aub or Ago3, is readily detected in ovarian somatic follicle cells (Nishida et al., 2007; Gunawardane et al., 2007; Saito et al., 2006; Cox et al., 2000); that *gypsy* silencing requires *piwi* but not *aub* (Pelisson et al., 2007); that the *gypsy* promoter drives expression of *gypsy* in the somatic follicle cells, which produce retrovirus-like particles that then infect the oocyte (Pelisson et al., 1994); that *ZAM* and *Idefix* are silenced in the somatic follicle cells by Piwi (Desset et al., 2008); and that follicle cell clones mutant for *piwi* or *flamenco* have remarkably similar defects(Cox et al., 1998; Mevel-Ninio et al., 2007), it is tempting to speculate that group III transposons in general are repressed by a Piwi-dependent, Ago3- and Aub-independent pathway that operates in follicle cells. In this view, antisense, Piwi-bound piRNAs would provide the primary somatic cell defense against group III transposon expression, whereas an Aub- and Ago3-dependent pathway provides a secondary defense in the germ line, the ultimate target of these transposons.

For groups I and II, loss of Ago3 switches the bias of both Aub and Piwi-bound piRNAs from strongly antisense-biased (or strongly sense-biased for group II) to slightly sense-biased (or antisense-biased for group II). That is, in wild-type flies, Ago3 acts to skew the strandedness of piRNA pools for group I and group II transposons. For group III transposons, however, loss of Ago3 slightly increased the antisense bias of piRNAs uniquely bound to Piwi: in *ago3*/TM6B ovaries, 88% of such piRNAs were antisense; in *ago3*, 90% were antisense (Table 4.S3). We conclude that most Piwi-bound group III piRNAs are

made directly from antisense transcripts such as the hypothesized precursor transcript that spans the *flamenco* locus. This explanation is consistent with the previous proposal that *flamenco* triggers silencing of transposons such as *gypsy*, *ZAM*, and *idefix* in somatic follicle cells—a cell type that expresses little if any Ago3 or Aub.

Ago3 amplifies piRNAs

Overall, piRNAs associated with Piwi were ~15-fold and piRNAs bound to Aub were ~6.4-fold more abundant than those bound to Ago3. The greater abundance of Piwi- and Aub-bound piRNAs is consistent both with the idea that Ago3 is less abundant than the other two PIWI proteins and with the proposal that Ago3 acts catalytically to amplify Piwi- and Aub-bound piRNAs. Consistent with the idea that group III piRNAs are largely Ago3-independent, the likelihood of a group III piRNA being associated with Piwi rather than Aub more than doubled in *ago3* ovaries: the Piwi-bound to Aub-bound piRNA ratio was 5.0 in *ago3*/TM6B, but 11.7 in *ago3*. This trend was not observed for group I piRNAs, where Piwi-bound piRNAs were twice as abundant as Aub-bound for *ago3*/TM6B, but only 1.6 times more abundant in *ago3* ovaries. Thus, for group III piRNAs, the absence of Ago3-catalyzed amplification shifts the piRNA pool towards the Piwi-dependent pathway.

Loss of Ago3 increases Group I transposon expression

What is the molecular consequence of piRNA loss? We used whole-genome tiling microarrays to measure the effect of loss of Ago3 on gene and transposon expression (Figure 4.6A and S11) and quantitative RT-PCR of

selected transposons to corroborate the microarray data (Figure 4.6B). The abundance of genic mRNA levels were generally unchanged in both *ago3* (Figure 4.S11A) and *aub* ovaries (Figure 4.S12A), compared with wild-type controls (w^1). In contrast, using a false-discovery rate (FDR) < 0.02, the expression of 33 of the 64 group I transposons and 14 of the 26 group III transposons increased in the absence of Ago3 (Figure 4.6A and S11B). However, the level of expression for many of these in both wild-type and *ago3* ovaries was less than the 50th percentile of expression for mRNA in wild-type, making accurate quantification of their change in expression challenging. Using this threshold for expression, 14 of the 64 group I and 4 of the 26 group III transposon families showed increased expression in the absence of Ago3 (Figure 4.6A).

Despite the loss of group II piRNAs in *ago3* mutants, expression of the five group II transposons was not significantly altered at a FDR < 0.02 (Figure 4.6A). Perhaps the production of Aub-bound sense piRNAs for group II transposons is futile and does not silence these elements, since they were likewise not activated in an *aub* mutant (Figure 4.S12B and 4.S13).

Compared with wild-type, expression of 18 of the 64 group I transposon families increased in *aub* ovaries (8 of these 18 had expression in both w^1 and *ago3* greater than the 50th percentile for mRNA expression in wild-type), including 14 that also increased in *ago3* ovaries (Figure 4.S12B and 4.S13). At a FDR < 0.02, the expression of only 2 group III transposons—*rover* and *McClintock*—was increased; both were also desilenced in *ago3* ovaries. These data support the idea that group I transposons rely on Ago3 and Aub for

Figure 4.6.

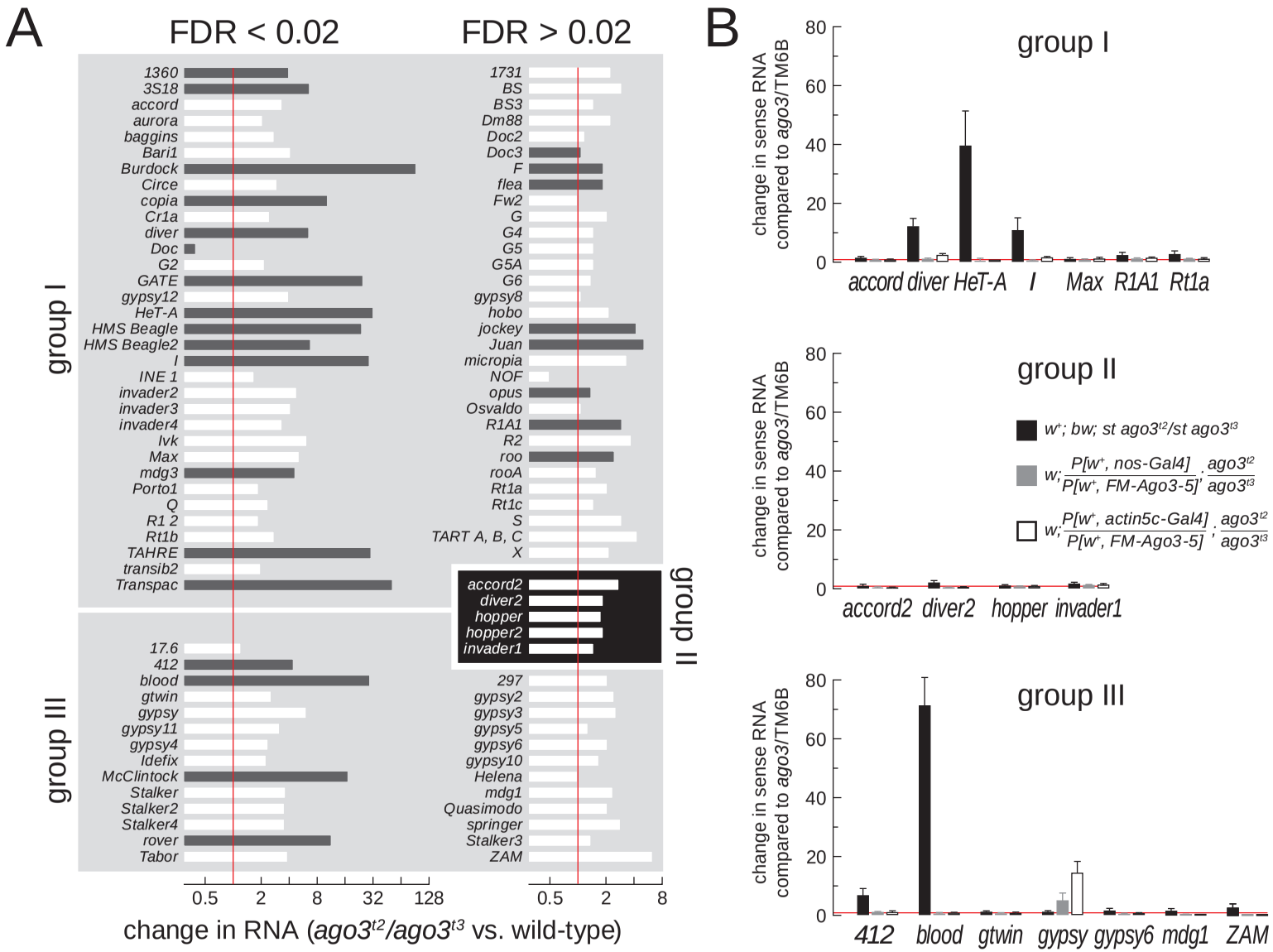

A
FDR < 0.02
FDR > 0.02
group I
group II
group III
1360
3S18
accord
aurora
baggins
Bari1
Burdock
Circe
copia
Cr1a
diver
Doc
G2
GATE
gypsy12
HeT-A
HMS Beagle
HMS Beagle2
I
INE 1
invader2
invader3
invader4
Ivk
Max
mdg3
Porto1
Q
R1 2
Rt1b
TAHRE
transib2
Transpac
1731
BS
BS3
Dm88
Doc2
Doc3
F
flea
Fw2
G
G4
G5
G5A
G6
gypsy8
hobo
jockey
Juan
micropia
NOF
opus
Osvaldo
R1A1
R2
roo
rooA
Rt1a
Rt1c
S
TART A, B, C
X
accord2
diver2
hopper
hopper2
invader1
17.6
412
blood
gtwin
gypsy
gypsy11
gypsy4
Idefix
McClintock
Stalker
Stalker2
Stalker4
rover
Tabor
297
gypsy2
gypsy3
gypsy5
gypsy6
gypsy10
Helena
mdg1
Quasimodo
springer
Stalker3
ZAM
0.5
2
8
32
128
0.5
2
8
change in RNA (ago3t2/ago3t3 vs. wild-type)
B
group I
change in sense RNA compared to ago3/TM6B
80
60
40
20
0
accord diver HeT-A I Max R1A1 Rt1a
group II
w+; bw; st ago3t2/st ago3t3
w; P[w+, nos-Gal4] / P[w+, FM-Ago3-5]' . ago3t2 / ago3t3
w; P[w+, actin5c-Gal4] / P[w+, FM-Ago3-5] ' . ago3t2 / ago3t3
accord2 diver2 hopper invader1
group III
412 blood gtwin gypsy gypsy6 mdg1 ZAM

Figure 4.6. Loss of Ago3 increases expression of some group I and group III, but not group II, transposons.

(A) Expression of group I, group II, and group III transposon families in *ago3* ovaries, relative to wild-type (w^1) ovaries, was assayed using whole-genome tiling microarrays. White bars: expression of the transposon family in both *ago3*/TM6B and *ago3* ovaries was less than that of the 50th percentile for expression of all mRNAs in wild-type, suggesting that expression change cannot be reliably quantified. Black bars: transposon families with expression greater than this threshold in one or both genotypes. Significant (false-discovery rate (FDR) < 0.02) and non-significant data (FDR > 0.02) are separated. (B) Quantitative RT-PCR was used to assess transposon expression, relative to Actin, for *ago3* ovaries (black bars) and ovaries expressing one copy of a UAS-Ago3 transgene driven by *nanos*-Gal4 (gray bars) or *actin5c*-Gal4 (white bars), relative to *ago3*/TM6B ovaries. The figure reports mean ± standard deviation for three independent biological samples.

silencing, whereas group III transposons are silenced by Piwi, with help from Ago3, likely in the germ line only, for some transposons.

Germ line expression of Ago3 rescues group I transposon silencing

Our data suggest that Ago3 is required for piRNA production and silencing of group I transposons in the female *Drosophila* germ line. As a final test of this hypothesis, we expressed Ago3 in *ago3* ovaries using the germ-line restricted *nanos* promoter. We used quantitative RT-PCR to measure the levels of mRNA for seven group I transposons, including five whose expression was increased in the absence of Ago3: *HeT-A*, *Diver*, *I-element*, *R1A1*, and *Rt1a*. For each, germ-line expression of Ago3 using the *nanos* promoter rescued silencing (Figure 4.6B).

Mis-expression of Ago3 in the soma interferes with group III transposon silencing

Argonaute proteins can compete for binding small RNAs. For example, decreasing the concentration of Ago1 in cultured S2 cells increases the loading of miRNAs into Ago2, whereas decreasing Ago2 increases miRNA loading into Ago1 (Forstemann et al., 2007; Horwich et al., 2007). Similarly, mis-expression of Aub in the soma inhibits Ago2-mediated RNAi (Specchia et al., 2008). To test the role of Ago3 in silencing transposons in somatic ovarian cells, we used the *actin5c* promoter to express Ago3 in both the germ line and the soma of *ago3* ovaries.

Surprisingly, ectopic expression of Ago3 in the soma *increased* the expression of the group III transposon *gypsy*, but enhanced silencing of the group I transposon

HeT-A (Figure 4.6B). First, these data confirm our observation that increased Ago3 expression in the female germ line increases silencing of group I transposons, perhaps because the concentration of Ago3 in the germ-line nurse cells limits piRNA amplification. Second, the data suggest that Ago3 cannot silence those transposons normally expressed in the somatic cells of the ovary. Perhaps Ago3 competes with Piwi for piRNAs, but, unlike Piwi, cannot act directly to silence transposons. Further supporting the idea that PIWI proteins compete for piRNAs, *gypsy* was silenced to a ~43-fold greater extent in *aub* ovaries than in *aub*/CyO (Figure 4.S13). We speculate that when Aub levels are low, more *gypsy* piRNAs associate with Piwi, leading to enhanced silencing of *gypsy.*

Discussion

Disentangling multiple piRNA pathways

Because ovaries contain both germ-line and somatic cells, our data conflate two distinct cell lineages. Combining our data with extensive genetic studies of *gypsy* and other transposon families represented in the *flamenco* locus, we have attempted to disentangle germ-line and somatic piRNA function (Figure 4.7). We propose that the somatic piRNA pathway is the more straightforward, involving only Piwi and not Ago3 or Aub. Hannon and co-workers (Malone et al., 2009) similarly deduce the existence of a somatic piRNA pathway by analyzing a broad panel of piRNA mutants and examining piRNAs from early embryos, which contain maternally deposited germ-line piRNAs but lack follicle cell-derived piRNAs. Thus, their study and ours infer from distinct data sets the existence of a somatic pathway in which primary piRNAs derived from *flamenco* are loaded directly into Piwi and not further amplified.

Our data suggest that Piwi cannot act alone to amplify piRNAs. We envision that Piwi-bound piRNAs in the soma are produced by a ribonuclease that randomly generates single-stranded guides that are subsequently loaded into Piwi and trimmed to length. Although Piwi-bound piRNAs generally begin with U and Piwi shows in vitro a preference for binding small RNA that begins with U (Yin and Lin, 2007), current evidence cannot distinguish between a putative piRNA-generating ribonuclease cleaving mainly at U and Piwi selecting U1 piRNAs from a set of RNAs with all possible 5′ nucleotides.

Without an amplification cycle to ensure an antisense bias, some other mechanism must operate to explain why Piwi-bound piRNAs are overwhelmingly

Figure 4.7.

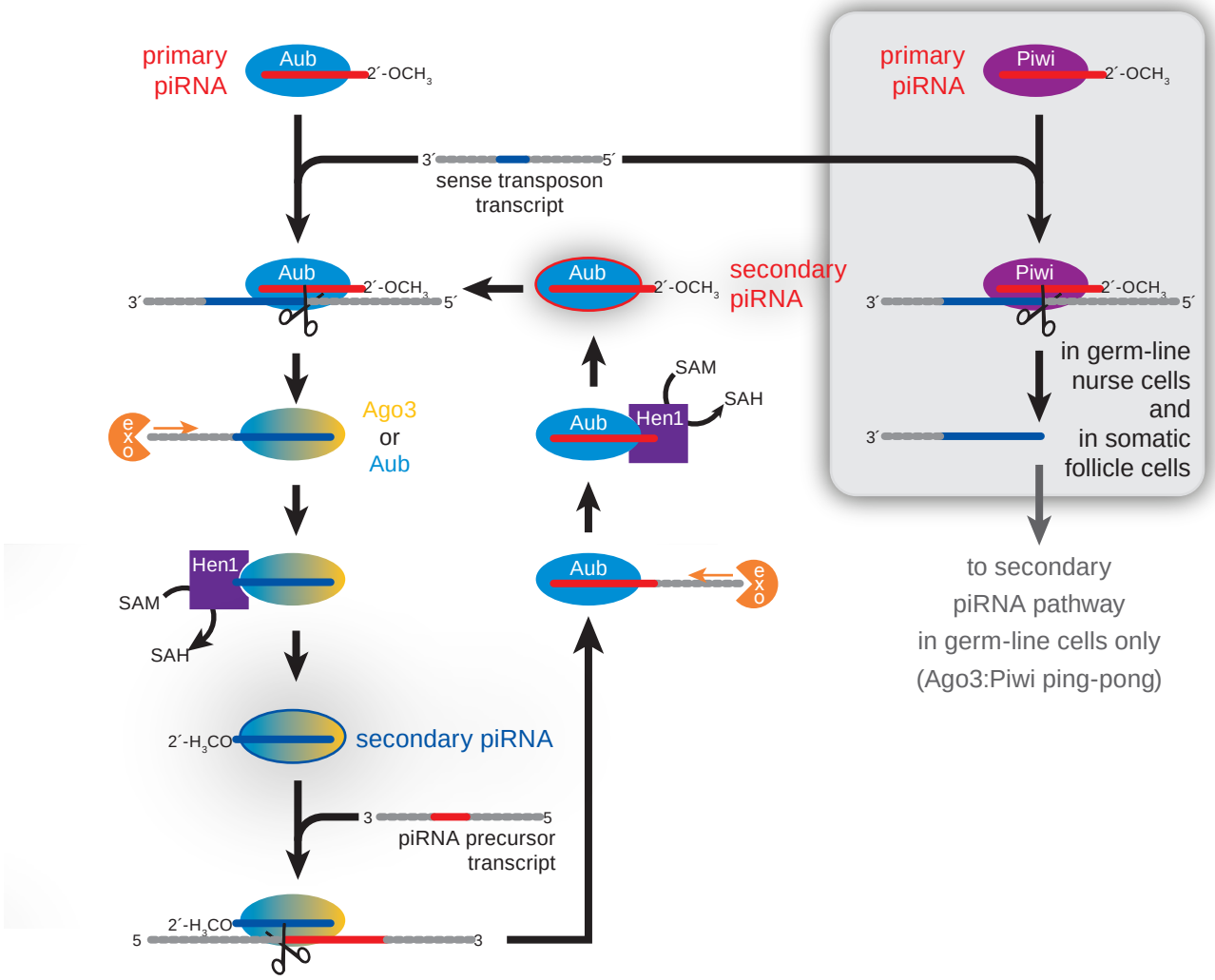

primary piRNA
Aub
2´-OCH3
3´
5´
sense transposon transcript
primary piRNA
Piwi
2´-OCH3
Aub
2´-OCH3
3´
5´
secondary piRNA
Piwi
2´-OCH3
3´
5´
in germ-line nurse cells and in somatic follicle cells
exo
Ago3 or Aub
SAM
SAH
Aub
Hen1
Hen1
SAM
SAH
Aub
exo
to secondary piRNA pathway in germ-line cells only (Ago3:Piwi ping-pong)
2´-H3CO
secondary piRNA
3
5
piRNA precursor transcript
2´-H3CO
5
3

Figure 4.7. A model for piRNA biogenesis.

The Aub- and Ago3-dependent piRNA amplification cycle is envisioned to operate only in the germ line, whereas a Piwi-dependent, Aub- and Ago3-independent pathway is shown for somatic cells. In the germ line, Piwi can also partner with Ago3 to amplify piRNAs.

antisense. A plausible but somewhat unsatisfying explanation comes from *flamenco* itself, whose constituent transposons are nearly all oriented in a single direction, so that the ~160 kb *flamenco* transcript is almost entirely antisense to the transposons. How such a non-random array of transposons could arise is unknown. Other non-randomly oriented piRNA clusters may explain the smaller number of transposons in group III that are not present in *flamenco*.

The transposons in most piRNA clusters do not show such a pronounced non-random orientation. These likely act in the germ line to produce primary piRNAs that load into Aub. The observed antisense bias of Aub-bound piRNAs arises subsequently, when Aub generates Ago3-bound secondary piRNAs and Ago3 acts, in turn, to produce Aub-bound secondary piRNAs. We propose that in the absence of Ago3, the sense/antisense ratio of Aub-bound piRNAs reverts to the inherent sense/antisense bias of the transposable element sequences present in the transcripts of piRNA clusters.

For this cycle to skew the Aub-bound piRNA population toward antisense, the substrate for cleavage by primary piRNA-bound Aub must be largely sense RNA. The best candidate for such sense RNA is mRNA derived from actively transcribed transposon copies. If such sense mRNA were largely found in the cytoplasm, it would be spatially segregated from the cluster transcripts, which we envision to be retained in the nucleus. Supporting this idea, sense transcripts from the group I transposon, *I* element, normally accumulate only in the nuclei of germ-line nurse cells (Chambeyron et al., 2008), likely because they are destroyed in the cytoplasm by Aub-bound primary piRNAs and Aub-bound piRNAs produced by Ago3-dependent amplification. In the absence of Aub, these sense transposon transcripts accumulate in the cytoplasm instead, consistent

with the strong desilencing of *I* element in *aub* and *ago3* mutants (Figure 4.6, 4.S12, and 4.S13 , and Vagin et al., 2006).

Nuage and the paradox of piRNA production

The piRNA ping-pong hypothesis predicts a role for Ago3 in the production of Aub-bound antisense piRNAs, but our finding that loss of Ago3 also reduced the abundance of Piwi-bound antisense piRNAs was unexpected. The majority of Aub and Ago3 is found in nuage and in the cytoplasm, but Piwi is predominantly nuclear. How then can Ago3 direct the production of Piwi-bound piRNAs? Perhaps Piwi transits the nuage en route from its site of synthesis, the cytoplasm, to where it accumulates, the nucleus. In this view, cytoplasmic Piwi is predicted to lack a small RNA guide. Piwi would then acquire its small RNA guide in the nuage, through a process that requires Ago3. Loading a piRNA into Piwi might then license it for entry into the nucleus, where it could act post-transcriptionally or transcriptionally to silence transposon expression. In this view, mutations in genes required for nuage assembly or stability, such as *vasa*, as well as genes required for Piwi loading would reduce the amount of nuclear Piwi. A similar mechanism may operate in mammals, where the PIWI protein MILI is found in cytoplasmic granules, whereas MIWI2 is nuclear. In the absence of MILI, MIWI2 delocalizes from the nucleus to the cytoplasm, although MIWI2 is not required for the localization of MILI (Aravin et al., 2008)

Such a model cannot explain the loading of Piwi in the somatic follicle cells, which contain little or no Ago3 or Aub and which do not contain nuage. A simple but untested hypothesis for these cells is that in the absence of nuage, empty Piwi readily enters the nucleus, where it obtains its small RNA guide. We might

reasonably expect that in germ cells the absence of nuage would impair the loading of Piwi by eliminating the Ago3-dependent, germ-line specific Piwi-loading process, but also facilitate entry of some empty Piwi into the nucleus, where it could obtain small RNA guides. Consistent with this idea, we do detect some Piwi in the nucleus in *ago3* ovaries. The simplicity of this hypothesis, of course, belies the complexity of testing it.

Why two distinct piRNA production pathways?

Retrotransposons "reproduce" by producing sense RNA encoding transposases and other proteins that allow them to jump to new locations in the germ cell genome. The conservation of the piRNA ping-pong cycle in animals (Aravin et al., 2001; Grimson et al., 2008) suggests that it is an ancient and conserved germ-line defense against retrotransposition. In flies, the *gypsy* family of retroelements appears to have moved its reproductive cycle to the somatic follicle cells adjacent to the germ line, which it infects using retrovirus-like particles. *gypsy* thus appears to avoid germ-line piRNA surveillance by transcribing and packaging its RNA in the soma. Perhaps expression of Piwi in *Drosophila* follicle cells reflects an adaptive evolutionary counter move to the *gypsy* reproductive strategy. The simplicity of the direct loading of Piwi with antisense piRNAs derived from *flamenco* may have made this counter defense more evolutionarily accessible than a strategy requiring expression of all the proteins needed for the Ago3:Aub ping-pong mechanism. In the future, more extensive analysis of the cellular and genetic requirements for ping-pong-independent and ping-pong-dependent piRNA mechanisms in *Drosophila*

melanogaster and in more ancient animal species may provide a test for these ideas.

Aub:Aub ping-pong pairs

In addition to Ago3:Aub and Aub:Ago3 ping-pong pairs, we also detected Aub:Aub ping-pong pairs in *ago3/*TM6B ovaries. These Aub:Aub ping-pong pairs have a strand bias similar to that observed for Ago3:Aub and Aub:Ago3 pairs. That is, group I, Aub-bound antisense piRNAs were 2.1 times more abundant than their Ago3-bound sense piRNA partners, and group II, Aub-bound sense piRNAs were 2.1 times more abundant than their Ago3-bound antisense piRNA partners. Similarly, Aub:Aub antisense piRNAs were twice as abundant than sense for group I and 2.3 times less abundant for group II.

These observations are hard to reconcile with the idea that a difference in the concentrations of Ago3 and Aub biases Aub-bound piRNAs towards antisense, as the protein partners in an Aub:Aub ping-pong cycle are, of course, at equal concentration. Moreover, piRNAs participating in Aub:Aub ping pong are antisense-biased in *ago3/*TM6B ovaries, but become sense biased or unbiased in $ago3^{t2}/ago3^{t3}$ ovaries: group I piRNAs participating in an Aub:Aub ping-pong cycle were sense biased (antisense/sense = 0.78), and the group II piRNAs had essentially no bias (antisense/sense = 1.06).

The simplest resolution for this paradox is that our nearly 5 million piRNA reads (derived from more than 16 million genome-matching reads) from immunoprecipitation experiments significantly underestimate the number of Ago3:Aub ping-pong pairs, likely because Ago3-bound piRNAs are so much rarer than those bound to Aub. Perhaps in vivo many Aub-bound piRNAs have a

corresponding Ago3-bound partner that we failed to sequence. Clearly, much deeper sequencing of piRNAs will be required to test this idea.

Experimental Procedures

General Methods

Preparation of 2–4 day ovary lysate (Tomari et al., 2004; Tuschl et al., 1999) and Northern hybridization and quantitative RT-PCR analysis were as described (Ghildiyal et al., 2008; Vagin et al., 2006). Tables S4 and S5 report probe and primer sequences. Table 4.S6 describes fly stocks. Figures were prepared using Excel (Microsoft, Redmond, WA, USA), IgorPro (WaveMetrics, Lake Oswego, OR, USA), and Illustrator (Adobe Systems, San Jose CA, USA). Microarray data are available via the NCBI gene expression omnibus web site (http://www.ncbi.nlm.nih.gob/geo/) using accession number GSE14370.

Isolation of *ago3*^{t1}, *ago3*^{t2}, and *ago3*^{t3} alleles

6,000 EMS-mutagenized fly lines (Koundakjian et al., 2004) were screened by tilling (Winkler et al., 2005; Till et al., 2003; Colbert et al., 2001) by the Seattle TILLING project (http://tilling.fhcrc.org/) according to Cooper et al. (2008). Candidate lines that contained mutations that induced premature stop codons in the *ago3* coding sequence were further characterized by sequencing of genomic PCR amplicons. A 1,532 bp region of *ago3* was analyzed using the primers 5′-AAC GAC GGA TGA ATC CAA GGG AGT TTT-3′ and 5′-GAA ATA CCA TTG GTT TGC CGA ATT TGA-3′. PCR resequencing of candidate mutations employed the primers 5′-ATG AAT CCA AGG GAG T-3′ and 5′-AGA GCA TTA CCA AGA ATC-3′ or 5′-GGA CGT TAA TCA ACC A-3′ and 5′-TGA CGA ATA CAA AGG T-3′.

Isolation and characterization of *Df(3L)TTT*

In a screen for deficiency (Df) mutations in 3L heterochromatin, KG03264/KG03264 males (carrying a P(Roseman et al., 1995) insert near the *nrm* locus) were treated with 4,000 rads of X-ray radiation, and mated en masse to *w*/*w*; TM3 *Sb*/TM6 *Hu Tb* virgin females. Single F1 K3264*/TM3 *Sb* males (where * indicates a mutagenized third chromosome) were then crossed to *Df(3L)FX3 e*/TM3 *Sb* females (Schulze et al., 2005), and stocks of putative Df mutations established from K3264*/TM3 *Sb* siblings. To map the newly isolated Dfs relative to individual lethal complementation groups, these putative Dfs were tested for lethality against smaller heterochromatic deficiencies on 3L, as well as against lethal mutations in individual essential genes. The molecular extent of *Df(3L)TTT* was further characterized by isolating genomic DNA from homozygous *Df(3L)TTT*/*Df(3L)TTT* embryos, and testing it by PCR: primers were used to look for the presence of exon regions of several 3L heterochromatic gene sequences (including *ago3*) which are distal to previously defined genetic loci. This analysis revealed that *Df(3L)TTT* is a large deficiency deleting a number of genes, from *ago3* through to *vtd* as the most proximal gene removed.

Generation of transgenic flies

Ago3 was cloned from Oregon R ovary cDNA using 3′ RACE. Ovary RNA was reverse transcribed using Superscript RT II (Invitrogen, Carlsbad, CA, USA) and a reverse transcription primer provided in the Gene Racer Kit (Invitrogen) (5′-GCT GTC AAC GAT ACG CTA CGT AAC GGC ATG ACA GTG $(T)_{18}$-3′). First strand cDNA was treated with RNase H. Full length Ago3 coding sequence was amplified by PCR using KOD polymerase (Takara Bio, Otsu, Japan) and the

primers 5′-CAC CAT GTC TGG AAG AGG AAA TTT GTT GAG C-3′, which spans the predicted translational initiation site (Williams and Rubin, 2002) and 5′-GCT GTC AAC GAT ACG CTA CGT AAC G-3′ (from the 3' RACE kit). The ~2,600 nt product was cloned into pEntr/D-Topo (Invitrogen). Sequencing of the full length insert revealed an 867 amino acid open reading frame and a 70 nt 3′ UTR. The open reading frame contained a single base change (A2143G) yielding I715V. pEntr/D-Topo-Ago3 was recombined into the pPFMW vector (UASp-$Flag_3$-Myc_6-ORF) from the Carnegie *Drosophila* Gateway Collection using Clonase (Invitrogen). Insert junctions were confirmed by sequencing. Genetic Services Inc. (Sudbury, MA, USA) injected plasmid DNA and identified transgenic flies. Transgenic insertions were mapped by inverse PCR (Huang et al., 2000).

Male fertility testing

To test male fertility, one virgin male was mated to five virgin Oregon R females; five individual males were tested for each genotype. After three days, the male was removed and the five females were transferred to a fresh vial, and then transferred to new vials every other day until 4 vials were obtained. The females were removed from the last vial after two additional days. The number of the adult progeny from each vial was counted.

Female fertility testing

Ten female virgins were collected and mated to five Oregon R virgin males on the day of collection in a small cage with a 60 mm diameter grape juice agar plate dabbed with yeast paste plate at 25˚C. For wild-type controls, six

Oregon R female virgins were mated to three Oregon R virgin males. After two days, the first plate was discarded and replaced with a fresh plate, and then the plate was changed and scored every subsequent day. The numbers of total eggs, eggs per female per day, and the dorsal appendage phenotype of embryos were scored every 24 h, and the numbers of eggs hatching were scored 48 h after the plate was changed. Twelve plates were scored in total for each genotype.

Immunohistochemistry and microscopy

Egg chamber fixation and whole-mount antibody labeling were performed as previously described (Theurkauf, 1994). Vasa protein was detected with rabbit polyclonal anti-Vasa antibody (Liang et al., 1994) diluted 1:1000. Piwi, Aub and Ago3 were detected with rabbit polyclonal anti-Piwi, anti-Aub and anti-Ago3 antibodies (Brennecke et al., 2007) and diluted 1:1000. Stellate antiserum (Klattenhoff et al., 2007) diluted 1:1000. Rhodamine-conjugated phalloidin (Invitrogen) was diluted 1:100 to stain F-actin; TOTO3 (Invitrogen) was used at 0.2 μM to visualize DNA. All tissues were mounted in 90% glycerol/PBS, with 1 mg/ml *p*-Phenylenediamine (Sigma-Aldrich, St. Louis, MO, USA). Samples were analyzed using a Leica TCS-SP inverted laser-scanning microscope with 63x (NA 1.32) PlanApo and 40x (NA 1.25) PlanApo oil-immersion objectives. Identical imaging conditions were used for each set of wild-type and mutant samples. Images were processed using Image J software.

Western blotting

Fifty μg total protein was mixed with an equal volume of 100 mM Tris-Cl, pH 6.8, 4% [w/v] SDS, 200 mM DTT, 20% [v/v] glycerol, and 0.2% (w/v) bromophenol blue, boiled for 10 min, and then resolved by 8% polyacrylamide/SDS gel electrophoresis. After electrophoresis, proteins were transferred to a PVDF membrane (Immobilon-P, Millipore, Billerica, MA, USA). The membrane was blocked in TBST-milk (25 mM Tris-Cl, pH 7.4, 3.0 mM KCl, 140 mM NaCl, 0.05% [v/v] Tween-20, 5% [w/v] non-fat dry milk) at room temperature for 2 h. After blocking, the membrane was cut according the molecular weight markers and incubated overnight at 4°C in TBST-milk containing primary antibody (anti-Ago3 (Brennecke et al., 2007) diluted 1:500, anti-Aub (Brennecke et al., 2007) diluted 1:1000, anti-Piwi (Saito et al., 2006) diluted 1:50, anti-Vasa (Liang et al., 1994) diluted 1:1000, anti-Ago1 (Miyoshi et al., 2005) diluted 1:1000, anti-Ago2 (Miyoshi et al., 2005) without dilution, anti-Armi at 1:1000, and anti-**a**-Tubulin (Sigma-Aldrich) diluted 1:10,000. Next, the membrane was washed three times with TBST at room temperature for 30 min and incubated 2 h at room temperature with either sheep anti-mouse IgG-HRP (GE Healthcare, Piscataway, NJ, USA) at 1:10,000 or goat anti-rabbit IgG-HRP (GE Healthcare) at 1:10,000 in TBST-milk. Then the membrane was washed three times with TBST at room temperature for 30 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). Quantitative analysis was performed using ImageGauge V4.22 (Fujifilm, Tokyo, Japan).

Immunoprecipitation

Twenty µl GammaBind G Sepharose (GE Healthcare) were incubated with 10 µl anti-Ago3, anti-Aub or anti-Piwi antibody for 2 h at 4°C. Next, the beads were washed five times with lysis buffer (30 mM HEPES-KOH, pH 7.4, 100 mM KCH_3CO_2, 2 mM $Mg(CH_3CO_2)_2$) containing 5 mM DTT, 0.5% [v/v] NP-40 and 1 tablet/50 ml complete-EDTA-free protease inhibitor cocktail tablet (Roche). Subsequently, 200 µl of ovary lysate (10 µg/µl total protein) was added and the mixture agitated gently at 4°C overnight. The supernatant (200 µl) and antibody-bound material were separated by centrifugation at 3,000 rpm for 1 min at 4°C, and the antibody-bound beads washed five times with RIPA buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1.0%[v/v] NP-40, 0.5%[w/v] DOC, 0.1%[w/v] SDS) containing 1 tablet/50 ml complete-EDTA-free protease inhibitor cocktail tablet, then resuspended in 200 µl lysis buffer. Five µl of input, supernatant, or bound samples were subject to western blotting analysis to confirm immunoprecipitation. The remaining samples were treated with 1 mg/ml proteinase K in 100 mM Tris-Cl, pH 7.5, 12.5 mM EDTA pH 8.0, 150 mM NaCl, 1% [w/v] SDS for 1 h at 65°C, extracted with an equal volume of phenol/chloroform, and then precipitated with three volumes 100% ethanol in the presence of 20 µg glycogen carrier. The precipitate was washed once with 80% ethanol and then dissolved in nuclease-free water.

Small RNA cloning and sequencing

Total RNA was isolated from manually dissected ovaries from 2–4 day flies using mirVana (Ambion, Austin, TX, USA). The RNA was quantified by

absorbance at 260 nm. To deplete 2S rRNA, 100 µg total RNA was annealed with 200 pmole of a complementary DNA oligonucleotide (5′-AGT CTT ACA ACC CTC AAC CAT ATG TAG TCC AAG CAG CAC T-3′) in 20 µl at 95°C for 2 min, then the temperature was gradually decreased over 1 h to room temperature. 2S rRNA was digested with 2 units of RNase H (Invitrogen) in a 30 µl reaction containing 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, and 10 mM DTT. After 2S rRNA depletion, 18–29 nt small RNA was purified from a 15% denaturing urea-polyacrylamide gel (National Diagnostics, Atlanta, GA, USA). Half of the purified RNA was oxidized with 25 mM $NaIO_4$ in 60 mM borax, 60 mM boric acid, pH 8.6, for 30 min at room temperature followed by ethanol precipitation. 100 pmole of 3′ pre-adenylated adapter (5′-rAppTCG TAT GCC GTC TTC TGC TTG T/ddC/-3′) was ligated to oxidized or un-oxidized small RNAs using mutant Rnl2 (amino acids 1-249, K227Q) (17-23B) (Addgene, Cambridge, MA, USA) at 4°C for 12 h in 20 µl in 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 60 µg/ml BSA, 10% (v/v) DMSO, and 40 U RNasin (Promega, Madison, WI, USA). 3′ ligated product was purified from a 15% denaturing urea-polyacrylamide gel (National Diagnostics), and then ligated to 100 pmole of 5′ RNA adapter (5′-GUU CAG AGU UCU ACA GUC CGA CGA UC-3′) using T4 RNA ligase (Ambion, Austin, TX, USA) at room temperature for 6 h in 20 µl of 50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, and 10% DMSO. The ligated product was purified from a 10% denaturing urea-polyacrylamide gel (National Diagnostics). Half the ligated product was used to synthesize cDNA using Superscript III (Invitrogen) and the reverse transcription primer (5′-CAA GCA GAA GAC GGC ATA CGA-3′), and half was used as a minus RT control. The small RNA library was amplified using forward (5′-AAT

GAT ACG GCG ACC ACC GAC AGG TTC AGA GTT CTA CAG TCC GA-3´) and reverse (5´-CAA GCA GAA GAC GGC ATA CGA-3´) primers, and then purified from a NuSieve GTG agarose gel (Lonza, Basel, Switzerland). Purified libraries were sequenced using a Solexa Genome Analyzer (Illumina, San Diego, CA, USA). Sequence and abundance data are available via the NCBI gene expression omnibus web site using accession number SRA007727.

Sequence extraction and annotation

For each sequence read, the first occurrence of the 6-mer perfectly matching the 5′ end of the 3′ linker was identified. Sequences without a match were discarded. The extracted inserts for sequences that contained the 3′ linker were then mapped to the female *Drosophila melanogaster* genome (Release R5.5, excluding chromosome YHet). Inserts that matched fully to a genomic sequence were collected using either Bowtie (Langmead et al., 2009) or in-house suffix tree-based software (Delcher et al., 1999; Delcher et al., 2002; Gusfield, 1997), and the corresponding genomic coordinates were determined for downstream functional analysis. Sequences corresponding to pre-miRNAs or non-coding RNAs (ncRNAs) were identified using the same suffix tree-based software. Genes and transposon annotations were retrieved from FlyBase (R5.5). In addition, unannotated transposons were identified using BLAST (Altschul et al., 1990) to query each transposon consensus sequence against the female genome, using an *e*-value cutoff of 10^{-10}. The number of reads for each small RNA was normalized by the number of times that RNA mapped to the genome and by the abundance of ncRNAs in the sample. For all transposon analyses, if a piRNA mapped to a location where there were two or more blast-

based transposon annotations, the number of reads were normalized by the number of annotations.

Background-corrected sequence bias detection

For a set of sequences of various lengths (e.g. all the piRNAs bound to Piwi protein), the frequency of each nucleotide at each position was computed as a foreground count matrix. The background frequencies at each position were computed by averaging over all possible *k*-mers (*k* = 23–29) in the consensus sequence of each transposon, strand-specifically (either sense or antisense), weighted by the length distribution of each foreground set. Binomial testing was performed for each nucleotide at each position in the foreground matrix, against the corresponding position in the background matrix. The significance level of testing used was a = 2.2×10^{-6}. This stringency ($p = 0.001$) corrects for multiple testing (four nucleotides at 29 positions, in two orientations (sense and antisense) for two classes (total versus ping-pong pairs). Only the nucleotides significantly above background were displayed, e.g. in Figure 4.5B, with *y*-axis values corresponding to the relative frequency of foreground minus background. This background correction helps distinguish an A or U bias that results from the inherent AU-richness of transposon sequences from an A or U bias reflecting the mechanism of piRNA biogenesis. At each significant position, the maximum possible y-value is indicated as a grey bar. Sequences were analyzed as species and not weighted by their abundance (number of reads).

Consensus mapping of total piRNA

The 23–29 nt small RNAs from the total RNA sequence reads in Oregon R, *ago3*/TM6B, $ago3^{t2}/ago3^{t3}$ *aub*/CyO, and aub^{HN2}/aub^{QC42} were mapped to the consensus sequence of each transposon to generate Figure 4.5A, 4.S4C and 4.S4D, 4.S5C and 4.S5D, and 4.S10C and 4.S10D. To account for the sequence variation among individual copies of the same transposon in the genome, we first establish coordinate conversion for individual copies and the consensus. For each BLAST-hit (the sequence of a copy), all the ungapped segments of its alignment with the consensus sequence were collected and used for coordinate conversion. The genomic coordinate of any piRNA that mapped to an ungapped segment of a BLAST hit was converted to the corresponding position on the consensus sequence. Overlaps of a piRNA to multiple segments were corrected for multiple counting, as well as multiple hits on the same genomic position caused by repetitive sequences. An advantage of coordinate conversion over direct sequence matching to the consensus sequence is that it allows consistency in the small RNAs included in different analyses, i.e., the same set of the small RNAs mapped to transposon regions in the genome are included in the set mapped on the consensus sequence. For each piRNA, all the bases that cover that piRNA were counted as having the number of reads of that piRNA, normalized by the number of times they mapped to the genome and by the total abundance (number of reads) of ncRNAs in the sample.

Overlap analysis of total RNAs

For each genotype, all the piRNAs whose 5’ ends overlapped with another piRNA on the opposite strand were collected in order to generate Figure 4.5A, 4.S4G, 4.S4H, 4.S5G, 4.S5H, 4.S10G, and 4.S10H.

Filtering and normalization of immunoprecipitation data

To avoid potential mis-assignment of piRNAs to a specific PIWI protein, we analyzed only those piRNAs that associated uniquely with Ago3, Aub, or Piwi. As 58% (corresponding to 58,265 reads) of Ago3- , 63% (301,656 reads) of Aub-, or 81% (544,102 reads) of Piwi-bound piRNAs were associated with just one of the three proteins, this restriction still permitted analysis of the majority of piRNA species. The normalizing factor was set to the ratio of the sum of reads in each immunoprecipitated data set to that in the total RNA sample, for the small RNAs that mapped to the fly genome only once and were also sequenced at least once in both the immunoprecipitation sample and the genotype-matched total RNA sample. All piRNA reads in each immunoprecipitation data set was divided by this normalizing factor. This normalization strategy allows direct comparison of the relative abundance of piRNAs uniquely bound to one PIWI protein with those bound to another, as well as comparison between *ago3* heterozygotes and mutants.

Consensus mapping of immunoprecipitation data

piRNAs from *ago3/*TM6B or *ago3*^{t2}*/ago3*^{t3} ovaries that were uniquely bound to Ago3, Aub, or Piwi were each mapped to the consensus sequence of each transposon, as described for “Consensus mapping of total piRNAs,” above. The abundance was normalized as described above.

Overlap graph by ping-pong types

For *ago3*/TM6B, there are nine possible ping-pong protein pairs: Ago3 (sense):Ago3 (antisense), Ago3 (sense):Aub (antisense), Ago3 (sense):Piwi (antisense), Aub (sense):Ago3 (antisense), Aub (sense):Aub (antisense), Aub (sense):Piwi (antisense), Piwi (sense):Ago3 (antisense), Piwi (sense):Aub (antisense), and Piwi (sense):Piwi (antisense). For $ago3^{t2}/ago3^{t3}$, there are four possible ping-pong protein pairs: Aub (sense):Aub (antisense), Aub (sense):Piwi (antisense), Piwi (sense):Aub (antisense), and Piwi (sense):Piwi (antisense). The sequences that were shared by two or more proteins were excluded. For each transposon, all the piRNA sequences that overlapped with piRNAs from the opposite strand were tallied (Figure 4.4E, 4.S6C, and 4.S6D).

Statistical testing of ping-pong pairs

To test the role of Ago3-mediated ping-pong amplification of Piwi- and Aub-bound piRNAs, we tallied the number of ping-pong pairs for each of the nine potential protein pairs for transposon-derived piRNAs (Figure 4.4E). We also tallied the number of pairs that could form in the same data set for overlaps of 8, 9, 11, and 12 nt, taking the average value as a measure of the number of ping-pong pairs that occur simply by chance, explained as follows. Given two sets of piRNAs (e.g. Ago3-bound sense piRNAs and Aub-bound antisense piRNAs for group II transposons), the observed abundance of ping-pong pairs between the two sets was defined as the sum of all the piRNA reads in the first set and those the second set that have a 10-bp overlap at their 5′ ends. Note that all the reads were normalized as described for “Normalization of immunoprecipitation data,” above. The same computation was applied to 8, 9, 11, and 12-nt overlaps, and the expected value was computed by averaging over these four numbers. Pooled

sample standard deviations from the two sets were used for background standard deviation. Estimating the expectation for the 10-bp overlap is the problem of randomly throwing *n* number of the 5′ end of antisense piRNAs and *m* number of the 10th position of sense piRNAs within some genomic scope (e.g. all transposon copies in the genome) and counting how many fall on the same position. Using overlaps of 8, 9, 11 and 12 nt would result in nearly identical distributions of random noise, given that there is no ping-pong for these overlaps, and thus can be used to estimate the expected value of the 10-nt overlap by chance. The *p*-value of each 10-nt overlap was computed from *Z*-score = (Observed - Expected)/(standard deviation), assuming that it follows a Gaussian distribution (Figure 4.4E, 4.6C, and 4.6D).

Expression Analysis

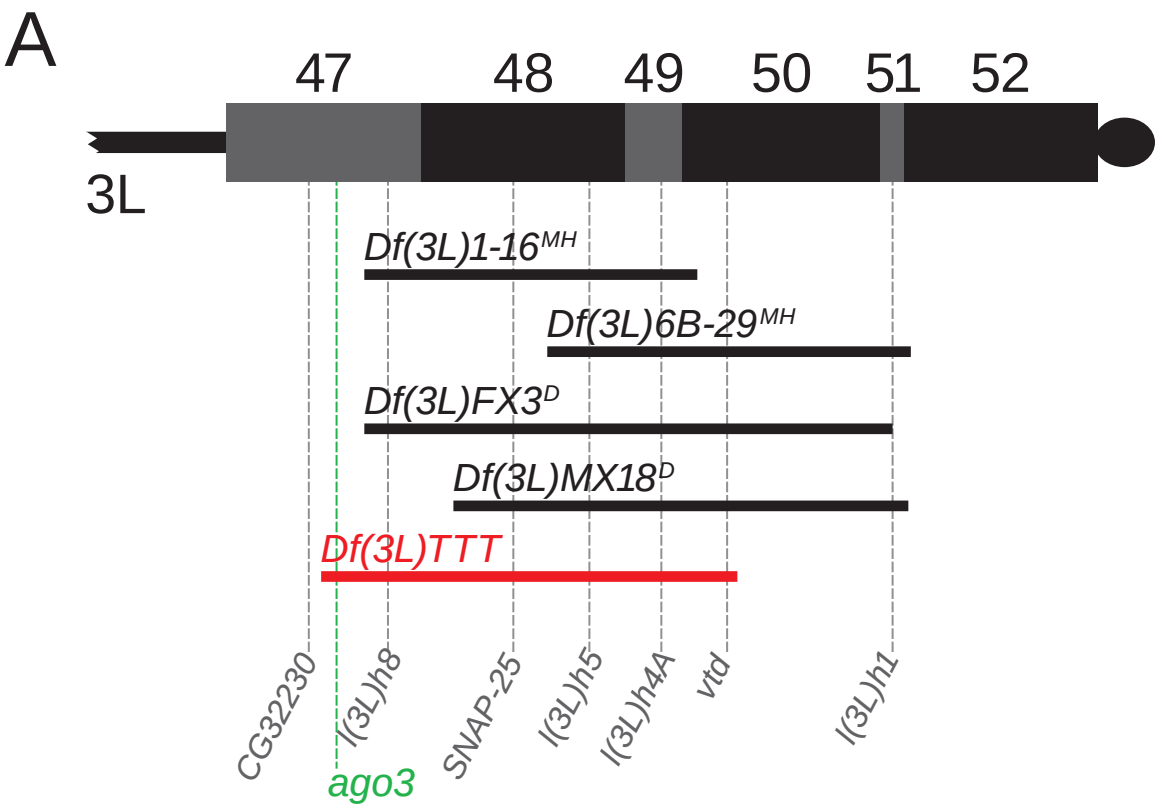
Coordinates and raw signal values for *Drosophila* tiling 2.0R Array probes were extracted from Affymetrix (Santa Clara, CA, USA) Tiling Array Software. Because transposons account for ~20% of the probes on the tiling array, a special normalization work flow was designed to avoid over-normalization of signals. All probes that mapped to transposons were identified. The remaining probes (which include intergenic and genic regions, with majority of probes corresponding to intergenic regions) were quantile-normalized across wild-type (w^1) and mutant replicates based on the assumption that the overall distribution of these probes should remain the same across the different strains. Then the signals for transposon probes were calculated using the normalized values of non-transposon probes at the same raw signal level. To summarize the signal for each transposon element, probes that mapped to any copy of the element were

grouped together and the Hodges-Lehmann estimator (Hollander and Wolfe, 1999) was used to calculate the pseudo-median of their normalized signals; pseudo-median is less sensitive to large numbers of outlier probes (probes whose values were 1, which indicate the values could not be reliably measured) in tiling array experiments. Differentially expressed transposon families were identified by contrasting their pseudo-median values in mutant replicates against the wild-type replicates. To correct for multiple testing, False Discovery Rates (FDRs) were calculated from t-test *p*-values. FDR < 0.02 was used to call significantly changed transposons. Transposons with average expression value less < 149 (the 50^{th} percentile expression value of all RefSeq mRNAs) in both wild type (w^1) and mutants were flagged, as they are at the detection limit of the tiling array. Similarly, expression values of mRNAs were summarized by calculating the pseudo-median of probes mapped to each of the RefSeq mRNA transcripts. Differentially expressed transcripts were identified using FDR < 0.02 cutoff (Figure 4.6A, 4.S11, and 4.S12).

Acknowledgments

We thank Mikiko and Haruhiko Siomi and Greg Hannon for antibodies, Erica Selva for flies expressing *nanos*-Gal4 from a transgene on the second chromosome, Paul Lasko for GFP-Vasa flies, Bradley Till for assistance with tilling primer design, Alicia Boucher for assistance with fly husbandry, Gwen Farley for technical assistance, and members of the Zamore and Weng laboratories for advice, suggestions, and critical comments on the text. This work was supported in part by grants from the National Institutes of Health to PDZ (GM62862 and GM65236) and WET (GM050898 and HD049116) and from NSERC Canada and CIHR to BMH. HX was supported in part by Pfizer, Inc.

Figure 4.S1.


A
47
48
49
50
51
52
3L
Df(3L)1-16MH
Df(3L)6B-29MH
Df(3L)FX3D
Df(3L)MX18D
Df(3L)TTT
CG32230
l(3L)h8
ago3
SNAP-25
l(3L)h5
l(3L)h4A
vtd
l(3L)h1

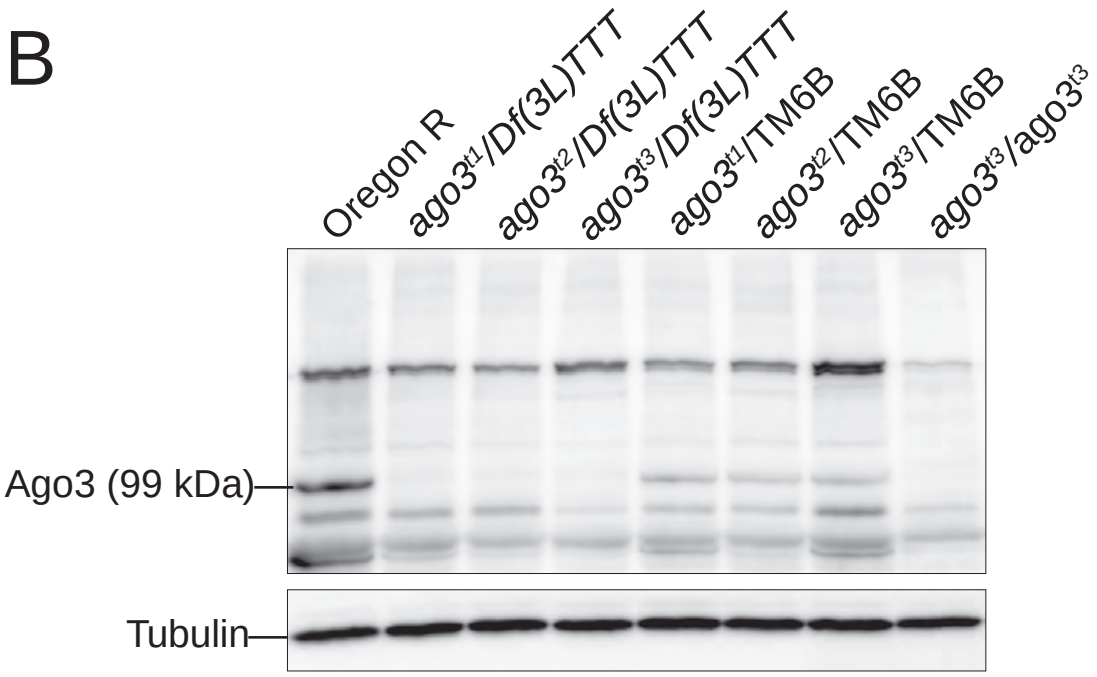

B
Oregon R
ago3t1/Df(3L)TTT
ago3t2/Df(3L)TTT
ago3t3/Df(3L)TTT
ago3t1/TM6B
ago3t2/TM6B
ago3t3/TM6B
ago3t3/ago3t3
Ago3 (99 kDa)
Tubulin

Figure 4.S1. *ago3* alleles.

(A) *Df(3L)TTT* deletes multiple genes (Schulze et al., 2005) including the *ago3* locus, creating the $ago3^{Df(3L)TTT}$ allele. (B) No full-length Ago3 protein was detected by Western blotting with anti-Ago3 antibody in any of the trans-heterozygous combinations of *Df(3L)TTT* with the three *ago3* EMS alleles isolated by tilling.

Figure 4.S2.

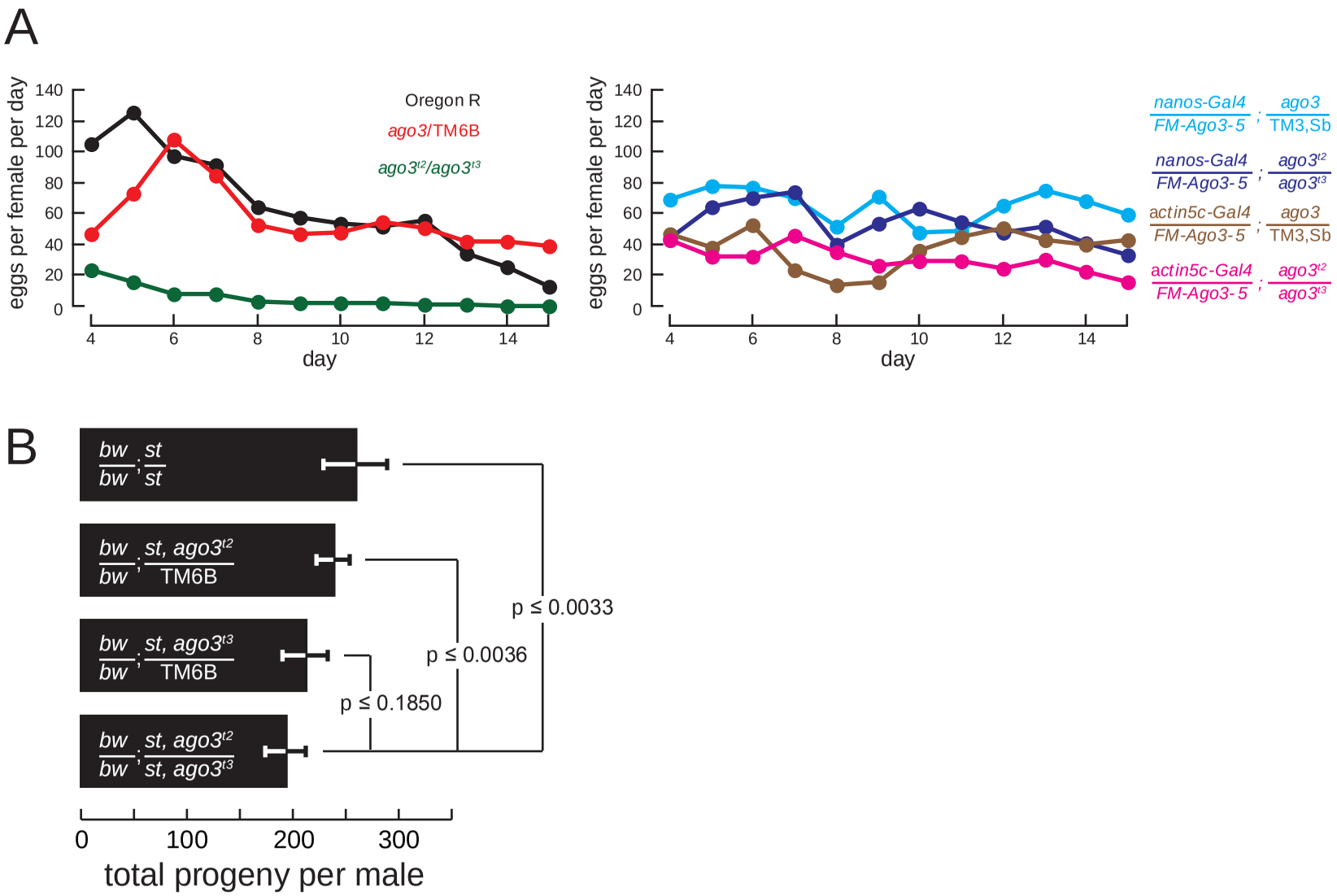

A
eggs per female per day
day
Oregon R
ago3/TM6B
ago3t2/ago3t3
nanos-Gal4 / FM-Ago3-5 ; ago3 / TM3,Sb
nanos-Gal4 / FM-Ago3-5 ; ago3t2 / ago3t3
actin5c-Gal4 / FM-Ago3-5 ; ago3 / TM3,Sb
actin5c-Gal4 / FM-Ago3-5 ; ago3t2 / ago3t3
B
bw / bw ; st / st
bw / bw ; st, ago3t2 / TM6B
bw / bw ; st, ago3t3 / TM6B
bw / bw ; st, ago3t2 / st, ago3t3
p ≤ 0.0033
p ≤ 0.0036
p ≤ 0.1850
total progeny per male

Figure 4.S2. Fertility analysis.

(A) *ago3* mutant females lay few eggs. The low egg production is rescued by an Ago3-expressing transgene, establishing that loss of Ago3 is the cause of the defect. (B) *ago3* mutant males have reduced fertility.

Figure 4.S3 is a movie that is available at the following website.

http://www.cell.com/supplemental/S0092-8674%2809%2900452-8

Figure 4.S3. Resolution of piRNAs into three distinct groups of transposon families by comparing their sense fraction in *ago3*/TM6B and *ago3*^{t2}/*ago3*^{t3} ovaries with the change in the abundance of antisense piRNAs between the two genotypes.

The piRNAs from different transposon families can be resolved into three distinct groups by comparing the fraction of sense piRNAs in *ago3*/TM6B heterozygous and *ago3*^{t2}/*ago3*^{t3} mutant ovaries with the change in antisense piRNA abundance between the two genotypes. The graph analyzes the 95 transposon families for which >= 500 sequence reads were obtained from *ago3*/TM6B ovaries. Black spheres, group I; blue, group II; red, group III.

Figure 4.S4 contains 96 pages and is available at the following website.

http://www.cell.com/supplemental/S0092-8674%2809%2900452-8

Figure 4.S4. Detailed analysis of piRNAs in Oregon R versus *ago3/*TM6B ovaries.

The figure presents the 95 transposon families for which we tallied 500 piRNA reads in *ago3/*TM6B, plus the tandem repeat *mst40*. (A, B) Sequence-bias plot ("sequence-excess logo") for all piRNAs mapping to the transposon family (A) or for piRNAs with ping-pong partners (B). The graphs report only nucleotide biases significantly different from the intrinsic bias of the transposon family itself (see Experimental Procedures). The gray bars indicate the maximum possible value for each position. (C, D) The abundance (per million) of piRNAs (23–29 nt) in Oregon R (C) and *ago3/*TM6B ovaries (D) for each base-pair position of the transposon consensus sequence. Abundance was both normalized by the number of times a piRNA mapped to the genome and calibrated to the total abundance of sequences corresponding to non-coding RNAs in the sample. For each base, the total normalized abundance of piRNAs that fall on that position is reported. Here and below, blue denotes sense strand-mapping piRNAs; red, antisense strand mapping piRNAs. (E, F) Length distribution of 19–29 nt RNAs mapping to the transposon family for Oregon R (E) and *ago3/*TM6B ovaries (F). The number of reads was normalized as in (C) and (D). (G, H) Distribution of the length of 5′ overlaps between sense and antisense piRNAs mapped to the transposon family for Oregon R (G) and *ago3/*TM6B (H) ovaries. Only pairs that overlapped at their 5′ ends were included. A peak at 10 bp indicates ping-pong pairing.

Figure 4.S5 contains 96 pages and is available at the following website.

http://www.cell.com/supplemental/S0092-8674%2809%2900452-8

Figure 4.S5. Detailed analysis of piRNAs in *ago3*/TM6B versus *ago3*^{t2}/*ago3*^{t3} ovaries.

The figure presents the 95 transposon families for which we tallied 500 piRNA reads in *ago3*/TM6B, plus the tandem repeat *mst40*. (A, B) Sequence-bias plot ("sequence-excess logo") for all piRNAs mapping to the transposon family (A) or for piRNAs with ping-pong partners (B). The graphs report only nucleotide biases significantly different from the intrinsic bias of the transposon family itself (see Experimental Procedures). The gray bars indicate the maximum possible value for each position. (C, D) The abundance (per million) of piRNAs (23–29 nt) in *ago3*/TM6B (C) and *ago3*^{t2}/*ago3*^{t3} ovaries (D) for each base-pair position of the transposon consensus sequence. Abundance was both normalized by the number of times a piRNA mapped to the genome and calibrated to the total abundance of sequences corresponding to non-coding RNAs in the sample. For each base, the total normalized abundance of piRNAs that fall on that position is reported. Here and below, blue denotes sense strand-mapping piRNAs; red, antisense strand mapping piRNAs. (E, F) Length distribution of 19–29 nt RNAs mapping to the transposon family for *ago3*/TM6B (E) and *ago3*^{t2}/*ago3*^{t3} ovaries (F). The number of reads was normalized as in (C) and (D). (G, H) Distribution of the length of 5′ overlaps between sense and antisense piRNAs mapped to the transposon family for *ago3*/TM6B (G) and *ago3*^{t2}/*ago3*^{t3} (H) ovaries. Only pairs that overlapped at their 5′ ends were included. A peak at 10 bp indicates ping-pong pairing.

Figure 4.S6 contains 96 pages and is available at the following website.

http://www.cell.com/supplemental/S0092-8674%2809%2900452-8

Figure 4.S6. Detailed analysis of piRNAs bound to Ago3, Aub, and Piwi in *ago3*/TM6B and *ago3^{t2}/ago3^{t3}* ovaries.

The figure presents the 95 transposon families for which we tallied 500 piRNA reads in *ago3*/TM6B, plus the tandem repeat *mst40*. (A, B) The abundance (per million) of piRNAs uniquely associated with Ago3, Aub, or Piwi in *ago3*/TM6B (A) and to Aub or Piwi protein for *ago3^{t2}/ago3^{t3}* ovaries (B) for each base-pair position of the transposon consensus sequence. Abundance was both normalized by the number of times a piRNA mapped to the genome and calibrated to the aggregate abundance of the same sequences in the corresponding total small RNA sample (see Experimental Procedures). For each base, the total normalized abundance of piRNAs that fall on that position is reported. Here and below, blue denotes sense strand mapping piRNAs; red, antisense-strand mapping piRNAs. (C, D) Distribution of the length of 5′ overlaps between sense and antisense piRNAs mapped to the transposon family and uniquely bound to Ago3, Aub, or Piwi protein for *ago3*/TM6B (C) and uniquely bound to Aub or Piwi protein for *ago3^{t2}/ago3^{t3}* (D) ovaries. Only pairs that overlapped at their 5′ ends were included. For each length of overlap, the percentage of the sense piRNA reads that have partners is plotted in blue at the top. Likewise, the percentage of antisense piRNA reads that have partners is plotted in red on the bottom. All reads were normalized by number of times mapped to the genome. A peak at 10 bp indicates ping-pong pairing.

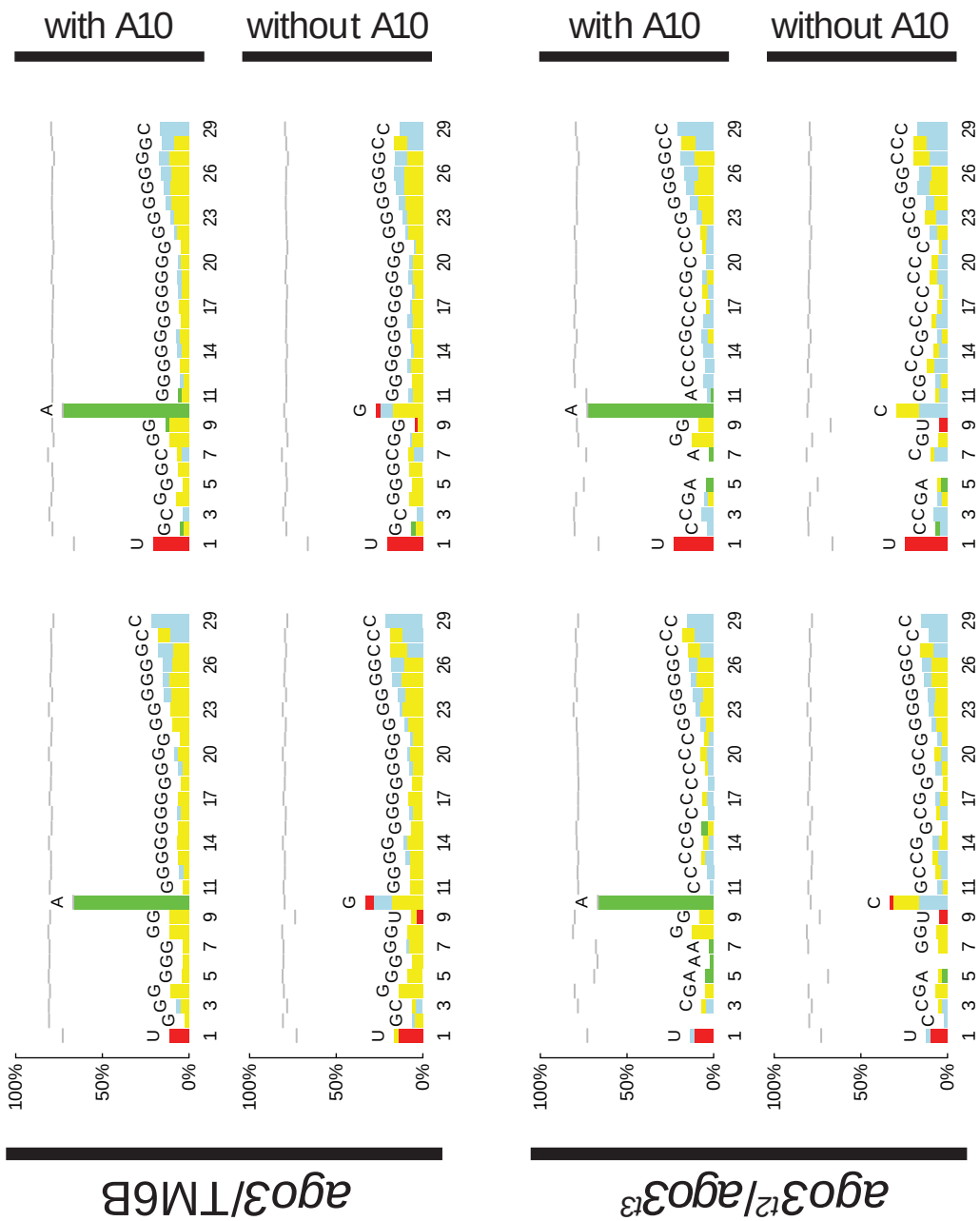

antisense
sense
with A10
without A10
with A10
without A10
ago3/TM6B
ago3t2/ago3t3
100%
50%
0%

Figure 4.S7.

Figure 4.S7. Sequence bias of Aub-bound piRNAs in *ago3*/TM6B and *ago3*^{t2}/*ago3*^{t3} ovaries.

piRNAs were separated by orientation—sense or antisense—and by the identity of their tenth nucleotide—A or not A. All four classes of piRNAs in both genotypes were more likely to begin with U than expected from the bias inherent in the underlying bias among all annotated transposon families in *Drosophila melanogaster*, irrespective of the identity of their tenth nucleotide. The analysis reports the percentage of each nucleotide in excess of the inherent bias of all transposon sequences for each piRNA position (see Experimental Procedures). Significance testing (p = 0.001) was Bonferroni corrected for four nucleotides at 29 positions, in two orientations (sense and antisense), among either three (heterozygotes) or two proteins (mutant). Gray bars indicate the maximum possible value for each position.

Figure 4.S8.

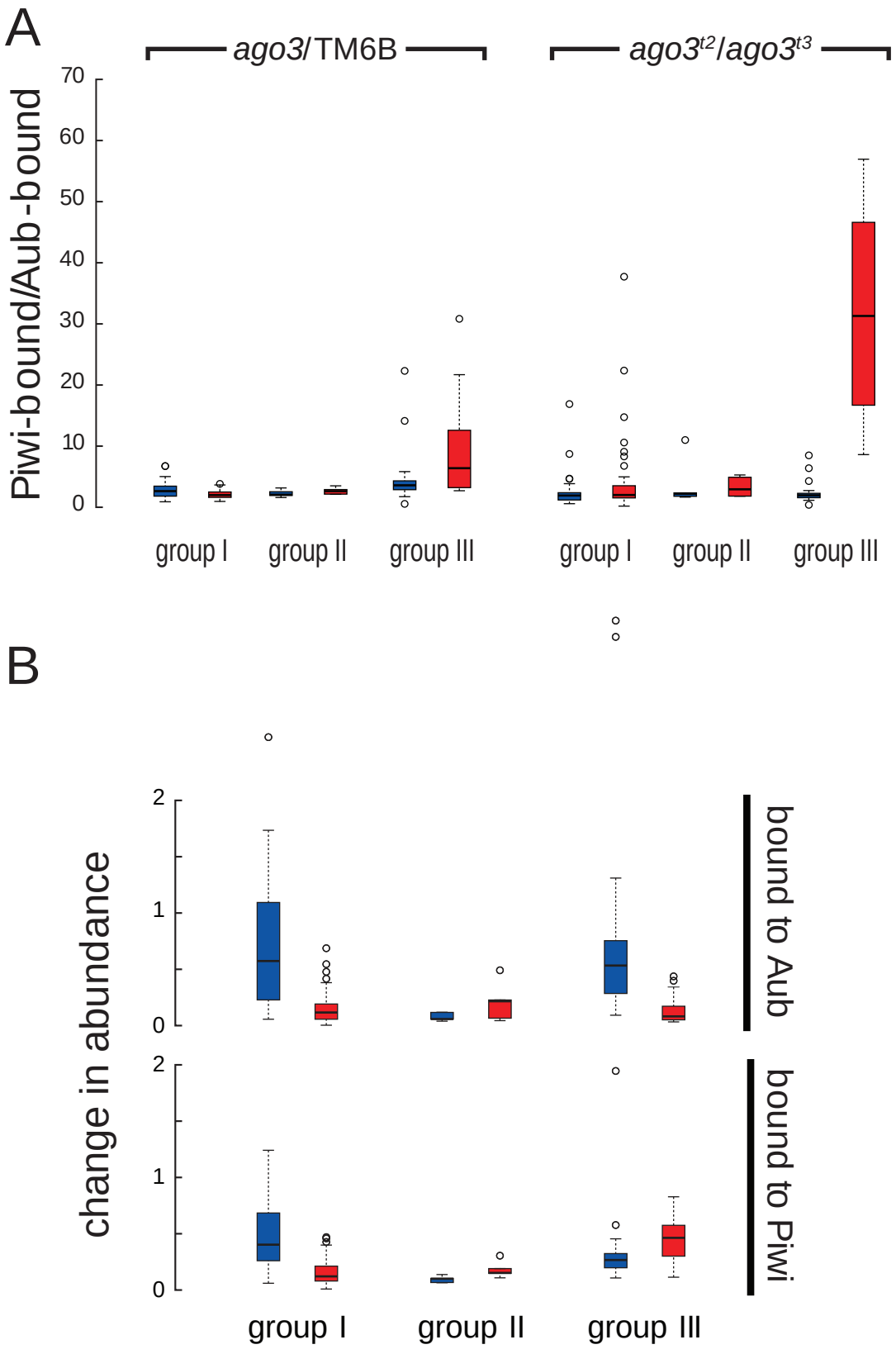

A
ago3/TM6B
ago3t2/ago3t3
Piwi-bound/Aub-bound
70
60
50
40
30
20
10
0
group I
group II
group III
group I
group II
group III
B
change in abundance
2
1
0
2
1
0
bound to Aub
bound to Piwi
group I
group II
group III

Figure 4.S8. Analysis by class of piRNAs uniquely bound to Aub or Piwi in *ago3*/TM6B and *ago3*^{t2}/*ago3*^{t3} for 95 transposon families.

(A) The ratio of uniquely Piwi-bound to uniquely Aub-bound piRNAs in ovaries. Median values for *ago3*/TM6B ovaries were 2.64 (sense) and 2.00 (antisense) for group I; 2.06 (sense) and 2.63 (antisense) for group II; and 3.59 (sense) and 6.39 (antisense) for group III. Median values for *ago3*^{t2}/*ago3*^{t3} ovaries were 1.92 (sense) and 2.00 (antisense) for group I; 2.21 (sense) and 2.94 (antisense) for group II; and 1.95 (sense) and 31.3 (antisense) for group III. (B) Box plots illustrating the change in abundance for piRNAs, separated by group and uniquely bound to Aub or Piwi in *ago3*^{t2}/*ago3*^{t3} versus *ago3*/TM6B for the 95 transposon families for which we tallied ≥ 500 piRNA reads in *ago3*/TM6B. Blue, sense; red, antisense.

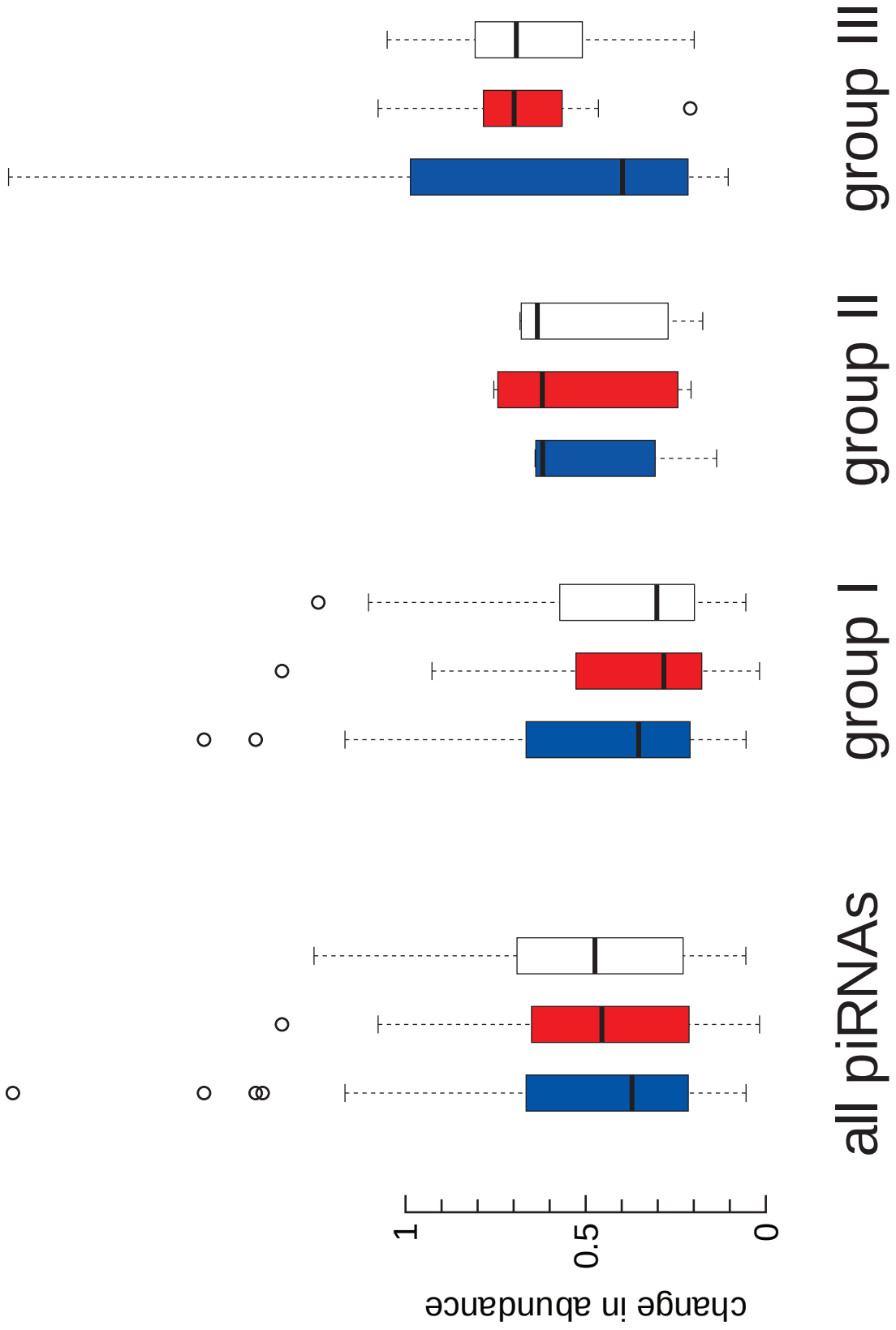

change in abundance
1
0.5
0
all piRNAs
group I
group II
group III

Figure 4.S9.

Figure 4.S9. Overview of piRNA abundance in *aub*/CyO and aub^{HN2}/aub^{QC42} for 95 transposon families

Box plots illustrating the change in abundance for piRNAs, separated by group, in aub^{HN2}/aub^{QC42} versus *aub*/CyO for the 95 transposon families for which we tallied ≥ 500 piRNA reads in *ago3*/TM6B. The median change for sense, antisense, and total for all piRNAs was 0.37, 0.45, 0.47; for group I, 0.35,. 0.28, 0.30; for group II, 0.62, 0.62, 0.63; and for group III, 0.40, 0.70, 0.69. Blue, sense; red, antisense; black, total.

Figure 4.S10 contains 96 pages and is available at the following website.

http://www.cell.com/supplemental/S0092-8674%2809%2900452-8

Figure 4.S10. Detailed analysis of piRNAs in *aub*/CyO versus *aub*^{HN2}/*aub*^{QC42} ovaries.

The figure presents the 95 transposon families for which we tallied 500 piRNA reads in *ago3*/TM6B, plus the tandem repeat *mst40*. (A, B) Sequence-bias plot ("sequence-excess logo") for all piRNAs mapping to the transposon family (A) or for piRNAs with ping-pong partners (B). The graphs report only nucleotide biases significantly different from the intrinsic bias of the transposon family itself (see Experimental Procedures). The gray bars indicate the maximum possible value for each position. (C, D) The abundance (per million) of piRNAs (23–29 nt) in *aub*/CyO (C) and *aub*^{HN2}/*aub*^{QC42} ovaries (D) for each base-pair position of the transposon consensus sequence. Abundance was both normalized by the number of times a piRNA mapped to the genome and calibrated to the total abundance of sequences corresponding to non-coding RNAs in the sample. For each base, the total normalized abundance of piRNAs that fall on that position is reported. Here and below, blue denotes sense strand-mapping piRNAs; red, antisense strand mapping piRNAs. (E, F) Length distribution of 19–29 nt RNAs mapping to the transposon family for *aub*/CyO (E) and *aub*^{HN2}/*aub*^{QC42} ovaries (F). The number of reads was normalized as in (C) and (D). (G, H) Distribution of the length of 5′ overlaps between sense and antisense piRNAs mapped to the transposon family for *aub*/CyO (G) and *aub*^{HN2}/*aub*^{QC42} (H) ovaries. Only pairs that overlapped at their 5′ ends were included. A peak at 10 bp indicates ping-pong pairing.

Figure 4.S11.

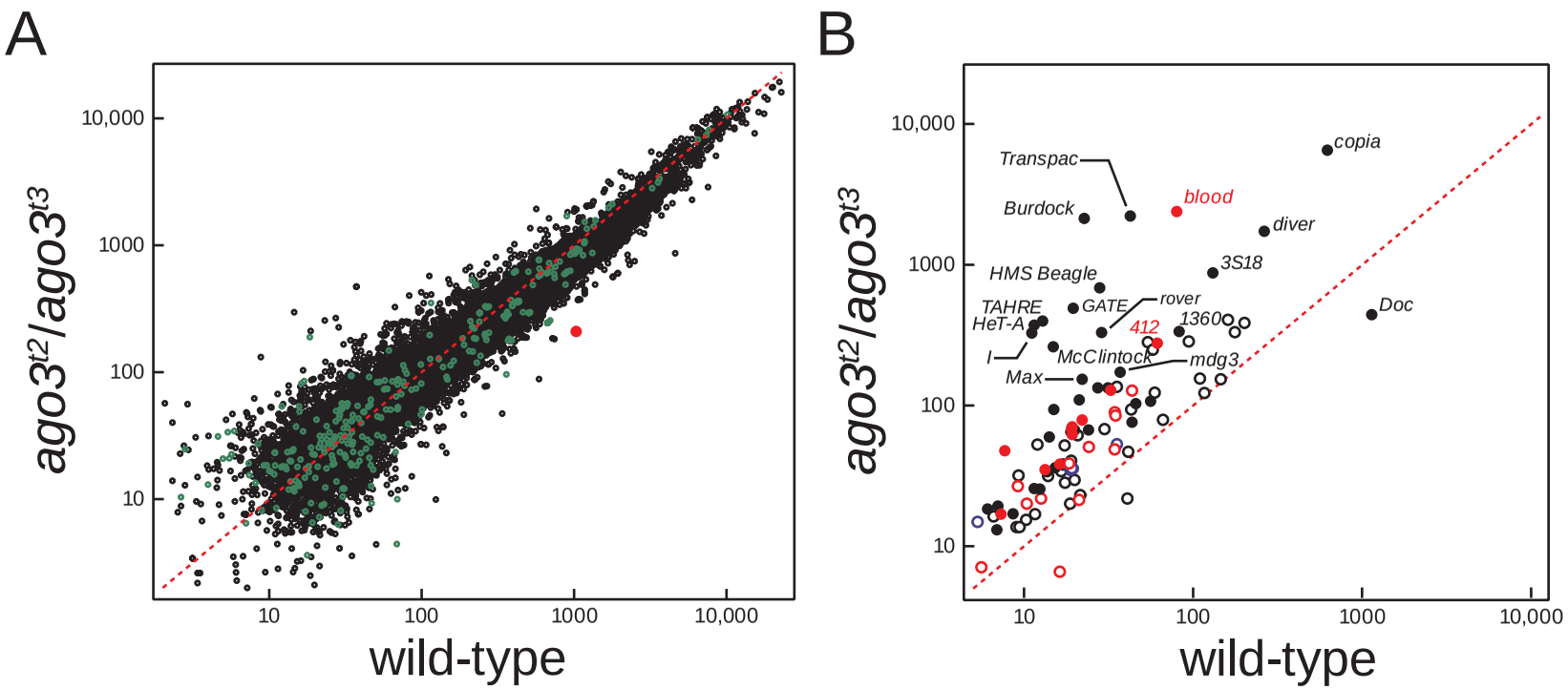

A
B
ago3t2/ago3t3
wild-type
10
100
1000
10,000
Transpac
Burdock
blood
copia
diver
3S18
HMS Beagle
TAHRE
HeT-A
GATE
rover
412
1360
Doc
I
McClintock
mdg3
Max

Figure 4.S11. Tiling microarray analysis of transposon and mRNA expression in wild-type (*w*^{1}) and *ago3* mutant ovaries.

(A) mRNA expression. Black, euchromatic genes; green, heterochromatic genes; Red, the *ago3* mRNA. All of the changes detected for mRNA corresponded to FDR > 0.02, suggesting that loss of Ago3 has no significant effect on mRNA expression. The FDR for *ago3* mRNA was 0.089. (B) Transposon expression. Group I transposons, black circles; group II transposons, blue circles; group III transposons, red circles. Open circles, FDR > 0.02; filled circles, FDR < 0.02.

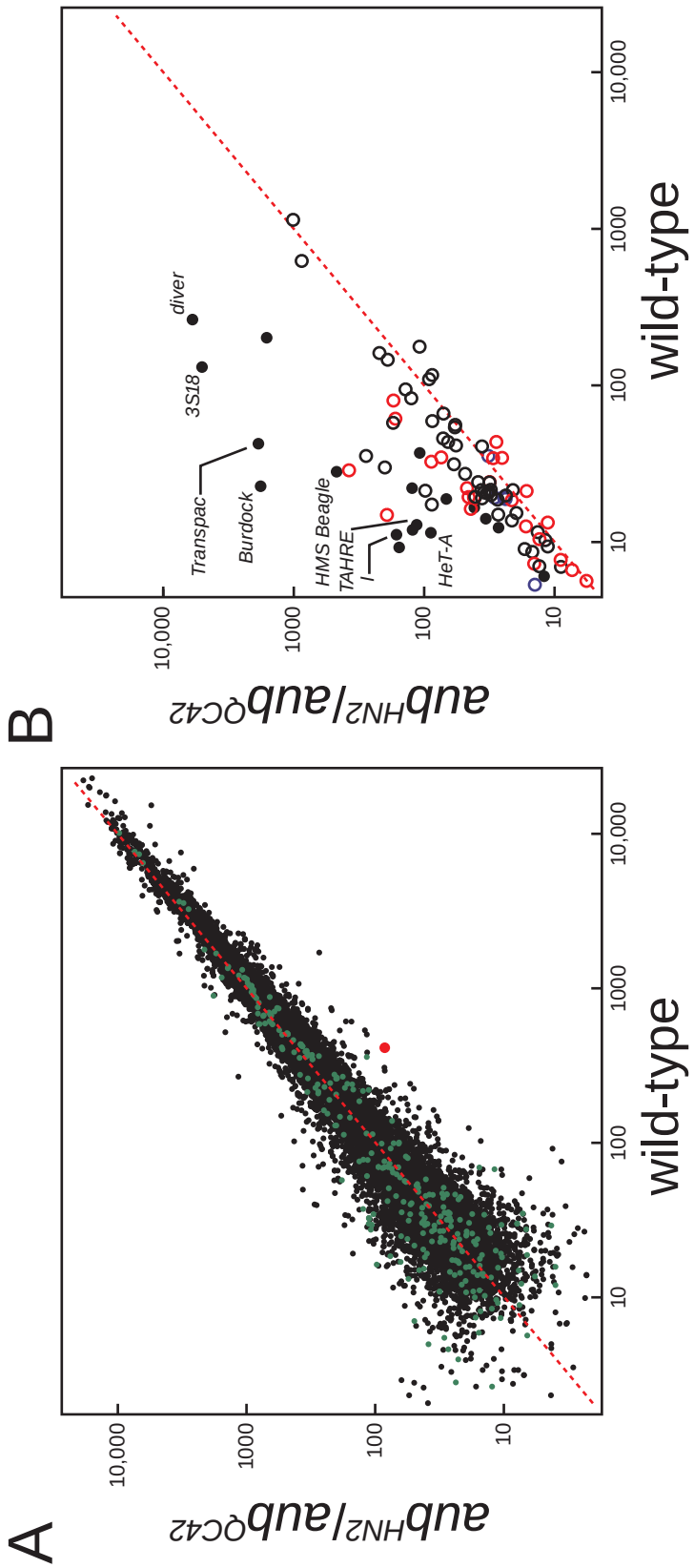

A
B
aubHN2/aubQC42
wild-type
10
100
1000
10,000
diver
3S18
Transpac
Burdock
HMS Beagle
TAHRE
I
HeT-A

Figure 4.S12.

Figure 4.S12. Tiling microarray analysis of transposon and mRNA expression in wild-type (w^1) and *aub* mutant ovaries.

(A) mRNA expression. Black, euchromatic genes; green, heterochromatic genes; Red, the *aub* mRNA. All of the changes detected for mRNA corresponded to FDR > 0.02, suggesting that loss of Aub has no significant effect on mRNA expression. (B) Transposon expression. Group I transposons, black circles; group II transposons, blue circles; group III transposons, red circles. Open circles, FDR > 0.02; filled circles, FDR < 0.02.

Figure 4.S13.

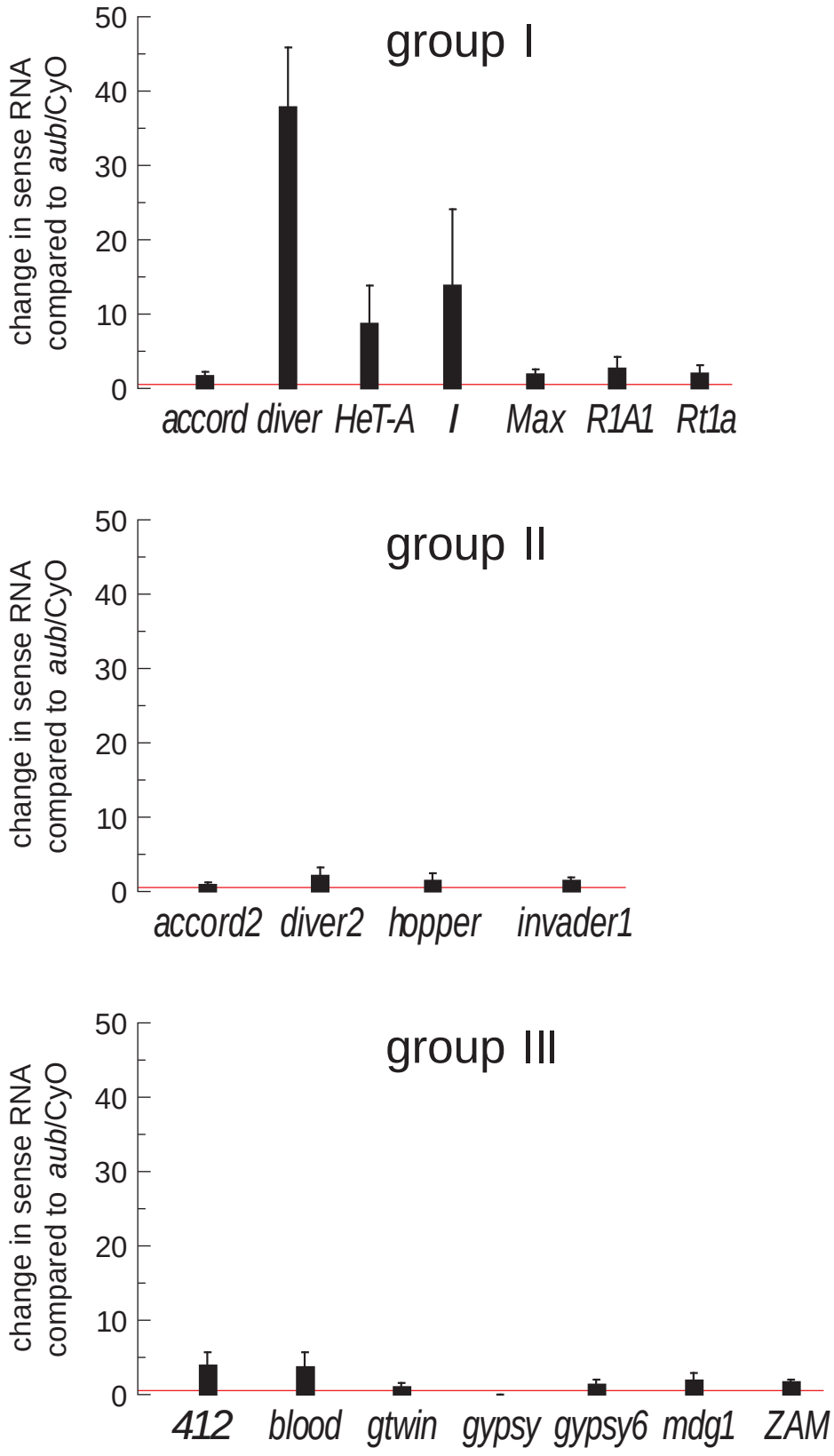

group I
change in sense RNA compared to aub/CyO
50
40
30
20
10
0
accord
diver
HeT-A
I
Max
R1A1
Rt1a
group II
change in sense RNA compared to aub/CyO
50
40
30
20
10
0
accord2
diver2
hopper
invader1
group III
change in sense RNA compared to aub/CyO
50
40
30
20
10
0
412
blood
gtwin
gypsy
gypsy6
mdg1
ZAM

Figure 4.S13. Quantitative RT-PCR analysis of transposon expression in *aub*^{HN2}/*aub*^{QC42} mutant ovaries.

Quantitative RT-PCR was used to assess the change in expression, relative to *actin*, between *aub* heterozygous and homozygous ovaries. The average ± standard deviation for at least three independent biological samples are shown.

Table 4.S1.

Maternal genotype	Total eggs	Hatch Rate (%)	Dorsal appendage phenotype (%)		
			Wild-type appendages	appendages fused	no appendages
Oregon R	3,880	96.1	100	0	0
bw; st	3,214	70.5	98.4	1.1	0.6
ago3/TM6B, Tb	3,545	76.1	94.8	2.9	2.3
$ago3^{t2}/ago3^{t3}$	398	0	33.2	30.2	36.7
$ago3^{t1}/Df(3L)TTT$	7,329	0	53.4	25.3	21.3
$ago3^{t2}/Df(3L)TTT$	5,514	0	61.1	28.6	10.2
$ago3^{t3}/Df(3L)TTT$	3,233	0	58.2	28.1	13.7
actin5c-Gal4/FM-Ago3-5; +/+	4,418	94.6	97.1	1.4	1.5
actin5c-Gal4/FM-Ago3-5; ago3/TM3, Sb	3,956	93.5	98.4	0.8	0.7
actin5c-Gal4/FM-Ago3-5; $ago3^{t2}/ago3^{t3}$	2,982	15.3	96.3	1.7	2.0
nanos-Gal4/FM-Ago3-5; +/+	5,915	86.2	96.5	2.0	1.5
nanos-Gal4/FM-Ago3-5; ago3/TM3, Sb	7,804	88.2	95.9	1.8	2.4
nanos-Gal4/FM-Ago3-5; $ago3^{t2}/ago3^{t3}$	3,922	4.0	95.9	2.1	2.0

Table 4.S1. Female fertility, hatch-rates, and patterning defects in *ago3* mutants and *ago3* mutants rescued with Ago3-expressing transgenes. FM, amino-terminal flag-myc tag.

Table 4.S2A.

Ovary genotype	Oxidized or immuno-precipitated	Total reads	Reads perfectly matching genome	Reads matching annotated ncRNAs	Small RNA reads (excluding ncRNAs)	Pre-miRNA–matching reads	Reads excluding ncRNA and pre-miRNA-matching	23–29 nt small RNA reads			
								Total	Transposon-matching		
									Total	Sense	Antisense
Oregon R	oxidized	2,605,637	1,989,207	292,579	1,696,628	16,564	1,680,064	1,314,928	1,038,522	282,861	764,617
ago3/TM6B	oxidized	1,848,746	1,388,205	326,407	1,061,798	6,660	1,055,138	871,512	688,643	135,762	559,033
$ago3^{t2}/ago3^{t3}$	oxidized	851,177	696,611	398,975	297,636	6,994	290,642	201,355	145,279	50,356	96,292
ago3/TM6B	Ago3 I.P.	5,268,201	1,536,789	768,350	768,439	89,537	678,902	395,070	146,649	103,232	44,390
	Aub I.P.	2,839,217	2,147,087	451,786	1,695,301	50,603	1,644,698	1,409,665	980,737	204,360	786,029
	Piwi I.P.	2,420,222	1,862,025	243,907	1,618,118	56,438	1,561,680	1,382,576	1,001,089	217,705	788,826
$ago3^{t2}/ago3^{t3}$	Aub I.P.	3,003,999	2,306,992	938,452	1,368,540	50,054	1,318,486	929,370	301,084	177,123	126,996
	Piwi I.P.	2,601,550	2,000,239	674,739	1,325,500	43,861	1,281,639	858,076	248,596	100,696	149,182
aub/CyO	oxidized	5,433,382	3,689,928	2,393,685	1,296,243	44,138	1,252,105	569,032	385,284	115,499	272,788
aub^{HN2}/aub^{QC42}	oxidized	4,621,929	3,110,675	2,283,336	827,339	52,258	775,081	309,848	81,526	19,163	63,031

Table 4.S2A. Sequencing statistics: small RNA abundance, as reflected in the number of reads sequenced. miRNAs were depleted from total RNA by oxidation, and piRNAs were enriched by immunoprecipitation (I.P.) using antibodies specific for Ago3, Aub, or Piwi protein. "Total small RNA reads" correspond to genome-matching reads after excluding annotated non-coding RNAs (ncRNAs), such as rRNA, snRNA, or snoRNA. "Transposon-matching reads" correspond to small RNAs mapped to *Drosophilia melanogaster* transposons.

Table 4.S2B.

Ovary genotype	Oxidized or immuno-precipitated	Total species	Species perfectly matching genome	Species matching annotated ncRNAs	Small RNA species (excluding ncRNAs)	Pre-miRNA–matching species	Small RNA species, excluding ncRNA and pre-miRNA-matching	23–29 nt small RNA species			
								Total	Transposon-matching		
									Total	Sense	Antisense
Oregon R	oxidized	1,034,974	647,975	20,558	627,417	551	626,866	447,700	314,178	103,432	212,741
ago3/TM6B	oxidized	730,995	478,968	16,686	462,282	411	461,871	359,888	255,565	66,438	190,818
$ago3^{t2}/ago3^{t3}$	oxidized	253,587	172,309	17,682	154,627	460	154,167	109,671	71,910	24,803	47,569
ago3/TM6B	Ago3 I.P.	1,142,088	207,949	32,401	175,548	656	174,892	100,597	32,764	19,495	13,528
	Aub I.P.	782,728	421,529	20,848	400,681	510	400,171	322,760	190,275	58,524	133,133
	Piwi I.P.	1,047,007	665,608	25,746	639,862	518	639,344	546,271	342,924	101,464	243,484
$ago3^{t2}/ago3^{t3}$	Aub I.P.	1,169,524	716,569	43,355	673,214	665	672,549	457,111	84,655	48,544	36,763
	Piwi I.P.	1,228,783	801,931	39,446	762,485	601	761,884	508,121	129,511	56,026	74,204
aub/CyO	oxidized	1,797,796	610,612	43,306	567,306	756	566,550	263,243	185,005	62,804	123,333
aub^{HN2}/aub^{QC42}	oxidized	1,372,955	412,058	39,926	372,132	781	371,351	160,716	32,482	7,994	24,681

Table 4.S2B. Sequencing statistics: small RNA complexity, as reflected in the number of unique species sequenced. miRNAs were depleted from total RNA by oxidation, and piRNAs were enriched by immunoprecipitation (I.P.) using antibodies specific for Ago3, Aub, or Piwi protein. "Total small RNA reads" correspond to genome-matching reads after excluding annotated non-coding RNAs (ncRNAs), such as rRNA, snRNA, or snoRNA. "Transposon-matching reads" correspond to small RNAs mapped to *Drosophilia melanogaster* transposons.

Table 4.S3. The fraction of sense and antisense piRNAs that were uniquely bound to each PIWI protein.

Overall

genotype	Ago3-bound		Aub-bound		Piwi-bound	
	sense	antisense	sense	antisense	sense	antisense
ago3/TM6B	0.76	0.24	0.29	0.71	0.27	0.73
$ago3^{t2}$/$ago3^{t3}$			0.59	0.41	0.38	0.62

Group I piRNAs

genotype	Ago3-bound		Aub-bound		Piwi-bound	
	sense	antisense	sense	antisense	sense	antisense
ago3/TM6B	0.78	0.22	0.29	0.71	0.31	0.69
$ago3^{t2}$/$ago3^{t3}$			0.59	0.41	0.52	0.48

Group II piRNAs

genotype	Ago3-bound		Aub-bound		Piwi-bound	
	sense	antisense	sense	antisense	sense	antisense
ago3/TM6B	0.26	0.74	0.73	0.27	0.70	0.30
$ago3^{t2}$/$ago3^{t3}$			0.47	0.53	0.54	0.46

Group III piRNAs

genotype	Ago3-bound		Aub-bound		Piwi-bound	
	sense	antisense	sense	antisense	sense	antisense
ago3/TM6B	0.75	0.25	0.19	0.81	0.12	0.88
$ago3^{t2}$/$ago3^{t3}$			0.64	0.36	0.10	0.90

Table 4.S4. Probes for Northern hybridization.

Small RNA detected	**Probe sequence (DNA or RNA)**
Aub-bound *roo* piRNA	5′-AAGAAATCAGTAATGCACTCTAGTA-3′
Piwi-bound *roo* piRNA	5′-GACTATTTACTTAGGCCTCTGCGTA-3′
Aub-bound *roo* piRNA	5′-AAGAAAUCAGUAAUGCACUCUAGUA-3′
Piwi-bound *roo* piRNA	5′-GACUAUUUACUUAGGCCUCUGCGUA-3′
Su(Ste) piRNA	5′-TCGGGCTTGTTCTACGACGATGAGA-3′
AT-chX-1 piRNA	5′-GCCCGAGCCGTCTAACGATGAAACA-3′
miR-8	5′-GACATCTTTACCTGACAGTATTA-3′
2S rRNA	5′-TACAACCCTCAACCATATGTAGTCCAAGCA-3′

Table 4.S5. Primers for quantitative PCR.

RNA detected	**Forward, Reverse**
412	5′-CACCGGTTTGGTCGAAAG-3′, 5′-GGACATGCCTGGTATTTTGG-3′
Accord	5′-ACAATCCACCAACAGCAACA-3′, 5′-AAAAGCCAAAATGTCGGTTG-3′
Accord2	5′-TTGCTTTCGGACTTCGTCTT-3′, 5′-TTCCACAACGAAAACAACCA-3′
Actin5c	5′-AAGTTGCTGCTCTGGTTGTCG-3′, 5′-GCCACACGCAGCTCATTGTAG-3′
Blood	5′-TGCCACAGTACCTGATTTCG-3′, 5′-GATTCGCCTTTTACGTTTGC-3′
Diver	5′-GGCACCACATAGACACATCG-3′, 5′-GTGGTTTGCATAGCCAGGAT-3′
Diver2	5′-CTTCAGCCAGCAAGGAAAAC-3′, 5′-CTGGCAGTCGGGTGTAATTT-3′
gtwin	5′-TTCGCACAAGCGATGATAAG-3′, 5′-GATTGTTGTACGGCGACCTT-3′
gypsy	5′-GTTCATACCCTTGGTAGTAGC-3′, 5′-CAACTTACGCATATGTGAGT-3′
gypsy6	5′-GACAAGGGCATAACCGATACTGTGGA-3′, 5′-AATGATTCTGTTCCGGACTTCCGTCT-3′
HeT-A	5′-CGCGCGGAACCCATCTTCAGA-3′,

	5′-CGCCGCAGTCGTTTGGTGAGT-3′
Hopper	5′-GGCTGGCTTCAACAAAAGAA-3′, 5′-GGACTCCCGAAAACGTCATA-3′
I-element	5′-GACCAAATAAAAATAATACGACTTC-3′, 5′-AACTAATTGCTGGCTTGTTATG-3′
Invader1	5′-GTACCGTTTTTGAGCCCGTA-3′, 5′-AACTACGTTGCCCATTCTGG-3′
Max	5′-TCTAGCCAGTCGAGGCGTAT-3′, 5′-TGGAAGAGTGTCGCTTTGTG-3′
mdg1	5′-AACAGAAACGCCAGCAACAGC-3′, 5′-CGTTCCCATGTCCGTTGTGAT-3′
R1A1	5′-AATTCCCGAGCTGTGCTAGA-3′, 5′-GTCTCAAGGCACCTTTCAGC-3′
rp49	5′-CCGCTTCAAGGGACAGTATCTG-3′, 5′-ATCTCGCCGCAGTAAACGC-3′
Rt1a	5′-CCACACAGACTGAGGCAGAA-3′, 5′-ACGCATAACTTTCCGGTTTG-3′
ZAM	5′-ACTTGACCTGGATACACTCACAAC-3′, 5′-GAGTATTACGGCGACTAGGGATAC-3′

Table 4.S6. Fly stocks.

Genotype	**Notes**
aub^{HN2} *cn bw*/CyO	(Schupbach and Wieschaus, 1991)
aub^{QC42} *cn bw*/CyO	(Schupbach and Wieschaus, 1991)
bw; st ago3^{t1}/TM6B *Tb*	*ago3*^{t1} allele
bw; st ago3^{t2}/TM6B *Tb*	*ago3*^{t2} allele
bw; st, ago3^{t3}/TM6B *Tb*	*ago3*^{t3} allele
Df(3L)TTT/TM3 Sb	*ago3* deficiency
w; P[w^{+}*, actin5c-Gal4]*/CyO*; st ago3*^{t2}/TM3 *Sb*	For Ago3 rescue
w; P[w^{+}*, nos-Gal4-VP16]*/CyO*; st, ago3*^{t2}/TM3 *Sb*	For Ago3 rescue
w; P[w^{+}*, UASp-FM-Ago3-5]*/CyO*; st ago3*^{t3}/TM3 *Sb*	For Ago3 rescue
Oregon R	wild-type
w^{1}	wild-type
w; Sp/CyO*; GFP-vasa*/TM3 *Sb*	(Johnstone and Lasko, 2004)

CHAPTER V: FINAL SUMMARY AND PERSPECTIVES

Drosophila piRNA pathway resides in gonads to safeguard the genome of germ cells against the deleterious influence of selfish genetic elements. Although it shows some similarities to the miRNA and siRNA pathways, the piRNA pathway is largely distinct from these two pathways in its mechanisms of the biogenesis and function.

We have identified 2´-*O*-methylation by Hen1 as the last step in the biogenesis of siRNAs and piRNAs. However, loss of Hen1 only leads to mild increase of transposons and typically *hen1* mutant animals are still fertile (Horwich et al., 2007; Saito et al., 2007; Kamminga et al., 2010). Generally, *hen1* mutation shows less severe defects as compared to mutations in other genes of the piRNA pathway. Why do piRNAs need to be modified? The function of 2´-*O*-methylation of piRNAs in *Drosophila* gonads remains elusive. Recent studies indicate that Hen1 prevents the 3´ tailing and trimming of siRNAs in flies (Ameres et al., 2010) and piRNAs in zebrafish (Kamminga et al., 2010). The 3´ tailing and trimming events might also occur on piRNAs in *hen1* mutants, as both *Su(Ste)* and *roo* piRNAs decrease in both length and abundance (Horwich et al., 2007). Animal miRNAs do not have 3´ terminal modifications and show less complementarity to their targets. Surprisingly, in the presence of fully complementary targets, miRNAs encounter 3´ tailing and trimming, as do siRNAs in *hen1* mutants (Ameres et al., 2010). These results suggest that 3´ tailing and trimming depend on the complementarity between small RNAs and their targets and only occur during pairing. In germ cells, the piRNA pathway elaborates a potent slicer-dependent ping-pong amplification mechanism to destroy

complementary mRNAs of active transposons. We envision a mild influence on the transposon silencing by the lost of 3´ terminal modification of piRNAs. Precisely, 2´-*O*-methylation of piRNAs affects the multiple turnovers of PIWI effector complex. This hypothesis explains the observed phenotype in *hen1* mutants but it remains to be tested.

Our systemic analyses of *ago3* and *aub* mutations largely validate the ping-pong amplification mechanism. However, there are still some key questions remaining to be answered.

How are 3´ ends of piRNAs generated? The ping-pong model illustrates the biogenesis of the 5´ ends of secondary piRNAs. Presently, biogenesis of the 3´ ends of piRNAs is unknown. One possibility is that 3´ ends of piRNAs are generated before loading of piRNAs into PIWI proteins. Alternatively, after the formation of 5´ ends, the 3´ cleavage products are loaded into PIWI protein and then 3´ends are generated by a putative 3´-5´ exonuclease. Given the distinct size distribution of piRNAs binding to each PIWI protein, the latter model is more attractive.

How are piRNAs and piRNA precursors loaded into PIWI proteins? Although the investigations of the somatic piRNA pathway in somatic follicle cells, OSSs, and OSCs, reveal that Armi, Yb, and Zuc are key players in the primary piRNA pathway, the molecular mechanism about how piRNAs or piRNA precursors are loaded into Piwi is still obscure. Moreover, these studies also discovered another subset of piRNAs derived from a subset of cellular mRNAs (Robine et al., 2009; Saito et al., 2009). These exonic piRNAs do not show correlation to the abundant mRNAs in the same cells and so far have no

apparent biological function. How these mRNAs and piRNA cluster transcripts are selectively fed into the piRNA pathway is still an open question.

How do other components function in the piRNA pathway? Genetic studies have identified more than a dozen of genes that are required for piRNA production and transposon silencing. Especially, *krimp*, *spn-E*, and *vasa* have been show to participate in the ping-pong amplification cycle (Malone et al., 2009). However, the molecular details of their function remain unclear. Moreover, Armi, Spn-E, and Vasa all contain helicase domains. Given evidence that supports piRNAs arise from single-stranded precursors, how do these helicase contribute to the biogenesis of piRNAs? We can envision several possibilities: RNA helicases may help unwind local dsRNA structures of the piRNA cluster to facilitate loading them into piRNA pathway; They may help to release the cleavage products during piRNA ping-pong cycles; They may also regulate the conformation of PIWI-piRNA complex. All these possibilities remain to be tested.

BIBLIOGRAPHY

Alefelder, S., Patel, B. K., and Eckstein, F. (1998). Incorporation of terminal phosphorothioates into oligonucleotides. Nucleic Acids Res *26*, 4983-4988.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Ameres, S. L., Horwich, M. D., Hung, J. H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P. D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

Anne, J. (2010). Targeting and anchoring Tudor in the pole plasm of the Drosophila oocyte. PLoS One *5*, e14362.

Anne, J., and Mechler, B. M. (2005). Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuleen. Development *132*, 2167-2177.

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.

Arkov, A. L., Wang, J. Y., Ramos, A., and Lehmann, R. (2006). The role of Tudor domains in germline development and polar granule architecture. Development *133*, 4053-4062.

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.

Bartel, D. P., and Chen, C. Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet *5*, 396-400.

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.

Baumberger, N., and Baulcombe, D. C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A *102*, 11928-11933.

Bedford, M. T., and Richard, S. (2005). Arginine methylation an emerging regulator of protein function. Mol Cell *18*, 263-272.

Belloni, M., Tritto, P., Bozzetti, M. P., Palumbo, G., and Robbins, L. G. (2002). Does Stellate cause meiotic drive in Drosophila melanogaster? Genetics *161*, 1551-1559.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Biemont, C., and Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. Nature *443*, 521-524.

Boswell, R. E., and Mahowald, A. P. (1985). tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. Cell *43*, 97-104.

Bozzetti, M. P., Massari, S., Finelli, P., Meggio, F., Pinna, L. A., Boldyreff, B., Issinger, O. G., Palumbo, G., Ciriaco, C., Bonaccorsi, S., and et, a. (1995). The Ste locus, a component of the parasitic cry-Ste system of Drosophila melanogaster, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. Proc Natl Acad Sci U S A *92*, 6067-6071.

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.

Brower-Toland, B., Findley, S. D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S. C., and Lin, H. (2007). Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev *21*, 2300-2311.

Carmell, M. A., Xuan, Z., Zhang, M. Q., and Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev *16*, 2733-2742.

Chambeyron, S., Popkova, A., Payen-Groschene, G., Brun, C., Laouini, D., Pelisson, A., and Bucheton, A. (2008). piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. Proc Natl Acad Sci U S A *105*, 14964-14969.

Chou, T. B., Noll, E., and Perrimon, N. (1993). Autosomal P[ovoD1] dominant female-sterile insertions in Drosophila and their use in generating germ-line chimeras. Development *119*, 1359-1369.

Chou, T. B., and Perrimon, N. (1992). Use of a yeast site-specific recombinase to produce female germline chimeras in Drosophila. Genetics *131*, 643-653.

Chou, T. B., and Perrimon, N. (1996). The autosomal FLP-DFS technique for generating germline mosaics in Drosophila melanogaster. Genetics *144*, 1673-1679.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. Proc Natl Acad Sci U S A *81*, 1991-1995.

Colbert, T., Till, B. J., Tompa, R., Reynolds, S., Steine, M. N., Yeung, A. T., McCallum, C. M., Comai, L., and Henikoff, S. (2001). High-throughput screening for induced point mutations. Plant Physiol *126*, 480-484.

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.

Cooper, J. L., Till, B. J., and Henikoff, S. (2008). Fly-TILL: reverse genetics using a living point mutation resource. Fly (Austin) *2*, 300-302.

Cote, J., and Richard, S. (2005). Tudor domains bind symmetrical dimethylated arginines. J Biol Chem *280*, 28476-28483.

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.

Dalmay, T., Horsefield, R., Braunstein, T. H., and Baulcombe, D. C. (2001). SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. EMBO J *20*, 2069-2078.

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

Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. (1999). Alignment of whole genomes. Nucleic Acids Res *27*, 2369-2376.

Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res *30*, 2478-2483.

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.

Du, T., and Zamore, P. D. (2005). microPrimer: the biogenesis and function of microRNA. Development *132*, 4645-4652.

Eddy, E. M. (1975). Germ plasm and the differentiation of the germ cell line. Int Rev Cytol *43*, 229-280.

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

Findley, S. D., Tamanaha, M., Clegg, N. J., and Ruohola-Baker, H. (2003). Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development *130*, 859-871.

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

Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y., and Zamore, P. D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Forstemann, K., Tomari, Y., Du, T., Vagin, V. V., Denli, A. M., Bratu, D. P., Klattenhoff, C., Theurkauf, W. E., and Zamore, P. D. (2005). Normal microRNA

maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol *3*, e236.

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

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

Fuller, M.T. (1993). Spermatogenesis. In The Development of *Drosophila melanogaster*, M. Bate and A.A. Martinez, eds. (Plainview, N.Y: Cold Spring Harbor Laboratory Press), pp. 71–148.

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A., and Imler, J. L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol *7*, 590-597.

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

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P. D. (2010). Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA *16*, 43-56.

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

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.

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

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

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

Gusfield, D. (1997). Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology (Cambridge, UK: Cambridge University Press).

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

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

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., and Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

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.

Hatfield, S. D., Shcherbata, H. R., Fischer, K. A., Nakahara, K., Carthew, R. W., and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. Nature *435*, 974-978.

Hoa, N. T., Keene, K. M., Olson, K. E., and Zheng, L. (2003). Characterization of RNA interference in an Anopheles gambiae cell line. Insect Biochem Mol Biol *33*, 949-957.

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

Houwing, S., 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, A. M., Rehm, E. J., and Rubin, G. M. (2000). Recovery of DNA Sequences Flanking P-element Insertions: Inverse PCR and Plasmid Rescue. In *Drosophila* Protocols, Sullivan, W., M. Ashburner, and R. S. Hawley, eds. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 429-437.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., and Liu, Q. (2005). Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. Genes Dev *19*, 1674-1679.

Johnstone, O., and Lasko, P. (2004). Interaction with eIF5B is essential for Vasa function during development. Development *131*, 4167-4178.

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

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

Kapitonov, V. V., and Jurka, J. (2003). Molecular paleontology of transposable elements in the Drosophila melanogaster genome. Proc Natl Acad Sci U S A *100*, 6569-6574.

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.

Kim, D. H., Behlke, M. A., Rose, S. D., Chang, M. S., Choi, S., and Rossi, J. J. (2005). Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. Nat Biotechnol *23*, 222-226.

Kim, K., Lee, Y. S., and Carthew, R. W. (2007). Conversion of pre-RISC to holo-RISC by Ago2 during assembly of RNAi complexes. RNA *13*, 22-29.

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

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

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

Kirino, Y., Vourekas, A., Kim, N., de Lima Alves, F., Rappsilber, J., Klein, P. S., Jongens, T. A., and Mourelatos, Z. (2010a). Arginine methylation of vasa protein is conserved across phyla. J Biol Chem *285*, 8148-8154.

Kirino, Y., Vourekas, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., Rappsilber, J., Jongens, T. A., and Mourelatos, Z. (2010b). Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. RNA *16*, 70-78.

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

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., Zhang, F., Schultz, N., Koppetsch, B. S., Nowosielska, A., Seitz, H., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2009). The Drosophila HP1 homolog Rhino is required for

transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Koundakjian, E. J., Cowan, D. M., Hardy, R. W., and Becker, A. H. (2004). The Zuker collection: a resource for the analysis of autosomal gene function in Drosophila melanogaster. Genetics *167*, 203-206.

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

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

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

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

Lee, Y. S., and Carthew, R. W. (2003). Making a better RNAi vector for Drosophila: use of intron spacers. Methods *30*, 322-329.

Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., and Carthew, R. W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Leuschner, P. J., Ameres, S. L., Kueng, S., and Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO Rep *7*, 314-320.

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

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol *15*, 1501-1507.

Li, M. A., Alls, J. D., Avancini, R. M., Koo, K., and Godt, D. (2003). The large Maf factor Traffic Jam controls gonad morphogenesis in Drosophila. Nat Cell Biol *5*, 994-1000.

Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. Development *120*, 1201-1211.

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.

Lim, A. K., Tao, L., and Kai, T. (2009). piRNAs mediate posttranscriptional retroelement silencing and localization to pi-bodies in the Drosophila germline. J Cell Biol *186*, 333-342.

Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H. E., Smith, D. P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

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.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., and Golub, T. R. (2005). MicroRNA expression profiles classify human cancers. Nature *435*, 834-838.

Ma, J. B., Ye, K., and Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

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

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., and Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Malone, C. D., and Hannon, G. J. (2009). Small RNAs as guardians of the genome. Cell *136*, 656-668.

Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev *18*, 975-980.

Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., and Zamore, P. D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Meister, G., Landthaler, M., Peters, L., Chen, P. Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Curr Biol *15*, 2149-2155.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343-349.

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.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J., and Qi, Y. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell *133*, 116-127.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M. C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

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.

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

Nykanen, A., Haley, B., and Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.

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

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

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

Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (2002). RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol Cell *9*, 315-327.

Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A., and Elgin, S. C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.

Palakodeti, D., Smielewska, M., Lu, Y. C., Yeo, G. W., and Graveley, B. R. (2008). The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. RNA *14*, 1174-1186.

Palumbo, G., Bonaccorsi, S., Robbins, L. G., and Pimpinelli, S. (1994). Genetic analysis of Stellate elements of Drosophila melanogaster. Genetics *138*, 1181-1197.

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.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. Curr Biol *12*, 1484-1495.

Parker, J. S., and Barford, D. (2006). Argonaute: A scaffold for the function of short regulatory RNAs. Trends Biochem Sci *31*, 622-630.

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

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

Patil, V. S., and Kai, T. (2010). Repression of Retroelements in Drosophila Germline via piRNA Pathway by the Tudor Domain Protein Tejas. Curr Biol

Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. J Virol *81*, 1951-1960.

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

Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W., and Sontheimer, E. J. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. Cell *117*, 83-94.

Pimpinelli, S., Sullivan, W., Prout, M., and Sandler, L. (1985). On biological functions mapping to the heterochromatin of Drosophila melanogaster. Genetics *109*, 701-724.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697-711.

Qi, H., Watanabe, T., Ku, H. Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb Body, a Major Site for Piwi-associated RNA Biogenesis and a Gateway for Piwi Expression and Transport to the Nucleus in Somatic Cells. J Biol Chem *286*, 3789-3797.

Qi, Y., Denli, A. M., and Hannon, G. J. (2005). Biochemical specialization within Arabidopsis RNA silencing pathways. Mol Cell *19*, 421-428.

Rand, T. A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621-629.

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

Roseman, R. R., Johnson, E. A., Rodesch, C. K., Bjerke, M., Nagoshi, R. N., and Geyer, P. K. (1995). A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in Drosophila melanogaster. Genetics *141*, 1061-1074.

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.

Saffman, E. E., and Lasko, P. (1999). Germline development in vertebrates and invertebrates. Cell Mol Life Sci *55*, 1141-1163.

Sahut-Barnola, I., and Pauli, D. (1999). The Drosophila gene stand still encodes a germline chromatin-associated protein that controls the transcription of the ovarian tumor gene. Development *126*, 1917-1926.

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., Ishizuka, A., Siomi, H., and Siomi, M. C. (2005). Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. PLoS Biol *3*, e235.

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.

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. Genes Dev *20*, 345-354.

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

Schulze, S. R., Sinclair, D. A., Fitzpatrick, K. A., and Honda, B. M. (2005). A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of Drosophila melanogaster. Genetics *169*, 2165-2177.

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.

Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Schwarz, D. S., Tomari, Y., and Zamore, P. D. (2004). The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol *14*, 787-791.

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

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.

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.

Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.

Song, J. J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Specchia, V., Benna, C., Mazzotta, G. M., Piccin, A., Zordan, M. A., Costa, R., and Bozzetti, M. P. (2008). aubergine gene overexpression in somatic tissues of aubergine(sting) mutants interferes with the RNAi pathway of a yellow hairpin dsRNA in Drosophila melanogaster. Genetics *178*, 1271-1282.

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

Szakmary, A., Reedy, M., Qi, H., and Lin, H. (2009). The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. J Cell Biol *185*, 613-627.

Theurkauf, W. E. (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. Methods Cell Biol *44*, 489-505.

Thomson, T., and Lasko, P. (2004). Drosophila tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning. Genesis *40*, 164-170.

Till, B. J., Colbert, T., Tompa, R., Enns, L. C., Codomo, C. A., Johnson, J. E., Reynolds, S. H., Henikoff, J. G., Greene, E. A., Steine, M. N., Comai, L., and Henikoff, S. (2003). High-throughput TILLING for functional genomics. Methods Mol Biol *236*, 205-220.

Tkaczuk, K. L., Obarska, A., and Bujnicki, J. M. (2006). Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. BMC Evol Biol *6*, 6.

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

Tritto, P., Specchia, V., Fanti, L., Berloco, M., D'Alessandro, R., Pimpinelli, S., Palumbo, G., and Bozzetti, M. P. (2003). Structure, regulation and evolution of the crystal-Stellate system of Drosophila. Genetica *117*, 247-257.

Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P., and Sharp, P. A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev *13*, 3191-3197.

Vagin, V. V., Klenov, M. S., Kalmykova, A. I., Stolyarenko, A. D., Kotelnikov, R. N., and Gvozdev, V. A. (2004). The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. RNA Biol *1*, 54-58.

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.

Van Doren, M., Williamson, A. L., and Lehmann, R. (1998). Regulation of zygotic gene expression in Drosophila primordial germ cells. Curr Biol *8*, 243-246.

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

Wang, X. H., Aliyari, R., Li, W. X., Li, H. W., Kim, K., Carthew, R., Atkinson, P., and Ding, S. W. (2006). RNA interference directs innate immunity against viruses in adult Drosophila. Science *312*, 452-454.

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

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

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

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

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

Winkler, S., Schwabedissen, A., Backasch, D., Bokel, C., Seidel, C., Bonisch, S., Furthauer, M., Kuhrs, A., Cobreros, L., Brand, M., and Gonzalez-Gaitan, M.

(2005). Target-selected mutant screen by TILLING in Drosophila. Genome Res *15*, 718-723.

Xu, K., Bogert, B. A., Li, W., Su, K., Lee, A., and Gao, F. B. (2004). The fragile X-related gene affects the crawling behavior of Drosophila larvae by regulating the mRNA level of the DEG/ENaC protein pickpocket1. Curr Biol *14*, 1025-1034.

Xu, T., and Rubin, G. M. (1993). Analysis of genetic mosaics in developing and adult Drosophila tissues. Development *117*, 1223-1237.

Yang, Z., Ebright, Y. W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Res *34*, 667-675.

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

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932-935.

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

Zambon, R. A., Vakharia, V. N., and Wu, L. P. (2006). RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster. Cell Microbiol *8*, 880-889.

Zamore, P. D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science *309*, 1519-1524.

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

APPENDICES: PUBLISHED MANUSCRIPTS

A distinct small RNA pathway silences selfish genetic elements in the germline.

The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC.

Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells.

Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies.

The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters.

Overlapping functions of Pea3 ETS transcription factors in FGF signaling during zebrafish development.

Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals.

MicroRNA-regulated, Systemically Delivered rAAV9: A Step Closer to CNS-restricted Transgene Expression.

A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline

Vasily V. Vagin,[1,2]* Alla Sigova,[1]* Chengjian Li,[1] Hervé Seitz,[1] Vladimir Gvozdev,[2] Phillip D. Zamore[1]†

In the *Drosophila* germline, repeat-associated small interfering RNAs (rasiRNAs) ensure genomic stability by silencing endogenous selfish genetic elements such as retrotransposons and repetitive sequences. Whereas small interfering RNAs (siRNAs) derive from both the sense and antisense strands of their double-stranded RNA precursors, rasiRNAs arise mainly from the antisense strand. rasiRNA production appears not to require Dicer-1, which makes microRNAs (miRNAs), or Dicer-2, which makes siRNAs, and rasiRNAs lack the 2′,3′ hydroxy termini characteristic of animal siRNA and miRNA. Unlike siRNAs and miRNAs, rasiRNAs function through the Piwi, rather than the Ago, Argonaute protein subfamily. Our data suggest that rasiRNAs protect the fly germline through a silencing mechanism distinct from both the miRNA and RNA interference pathways.

In plants and animals, RNA silencing pathways defend against viruses (*1*–*3*), regulate endogenous gene expression (*4*), and protect the genome against selfish genetic elements such as retrotransposons and repetitive sequences (*5*). Common to all RNA silencing pathways are RNAs 19 to 30 nucleotides (nt) long that specify the target RNAs to be repressed. In RNA interference (RNAi) (*6*), siRNAs are produced from long exogenous double stranded RNA (dsRNA). In contrast, ~22 nt miRNAs are endonucleolytically processed from endogenous RNA polymerase II transcripts. Dicer ribonuclease III (RNase III) enzymes produce both siRNAs and miRNAs. In flies, Dicer 2 (Dcr 2) generates siRNAs, whereas the Dicer 1 (Dcr 1) Loquacious (Loqs) complex produces miRNAs (*7*–*11*). After their production, small silencing RNAs bind Argonaute proteins to form the functional RNA silencing effector complexes believed to mediate all RNA silencing processes.

Phased sense and antisense siRNAs in vivo. In *Drosophila*, processive dicing of long dsRNA and the accumulation of sense and antisense siRNAs without reference to the orientation of the target mRNA are hallmarks of RNAi in vitro (*12*, *13*) and in vivo (Fig. 1). We prepared total small RNA from the heads of adult males expressing a dsRNA hairpin (fig. S1A) that silences the *white* gene via the RNAi pathway (*14*). *white* silencing requires Dcr 2 (*7*), R2D2 (*9*), and Ago2. siRNAs were detected with a microarray containing T_M (melting temperature) normalized probes, 22 nt long, for all sense and antisense siRNAs that theoretically can be produced by dicing the *white* exon 3 hairpin (Fig. 1A). Both sense and antisense *white* siRNAs were detected in wild type flies but not in *dcr 2*^{L811fsX} homozygous mutant flies. The Dcr 2 dependent siRNAs were produced with a periodicity of ~22 nt (Fig. 1B), consistent with the phased processing of the dsRNA hairpin from the end formed by the 6 nt loop predicted to remain after splicing of its intron containing primary transcript (fig. S1B).

***Su(Ste)* rasiRNAs.** *Drosophila* repeat associated small interfering RNAs (rasiRNAs) can be distinguished from siRNAs by their longer length, 24 to 29 nt (*15*, *16*). rasiRNAs have been proposed to be diced from long dsRNA triggers (*5*, *16*), such as the ~50 copies of the bidirectionally transcribed *Suppressor of Stellate* [*Su(Ste)*] locus on the Y chromosome (*17*) (fig. S2B) that in testes silence the ~200 copies of the protein coding gene *Stellate* (*Ste*) found on the X chromosome.

Microarray analysis of total small RNA isolated from fly testes revealed that *Su(Ste)* rasiRNAs detectably accumulate only from the antisense strand (Fig. 2A), with little or no phasing (fig. S2A). As expected, *Su(Ste)* rasiRNAs were not detected in testes from males lacking the *Su(Ste)* loci (*cry*^{1}Y) (Fig. 2A). *Su(Ste)* rasiRNAs were also absent from *armitage* (*armi*) mutant testes (Fig. 2A), which fail to silence *Ste* and do not support RNAi in vitro (*18*). *armi* encodes a non DEAD box helicase (*19*) homologous to the *Arabidopsis thaliana* protein SDE3, which is required for RNA silencing triggered by transgenes and some viruses (*20*), and depletion by RNAi of the mammalian Armi homolog Mov10 blocks siRNA directed RNAi in cultured human cells (*21*). Normal accumulation of *Su(Ste)* rasiRNA and robust *Ste* silencing also require the putative helicase Spindle E (Spn E), a member of the DExH family of adenosine triphosphatases (ATPases) (*16*, *22*) (Fig. 2B and fig. S2B).

The accumulation in vivo of only antisense rasiRNAs from *Su(Ste)* implies that sense *Su(Ste)* rasiRNAs either are not produced or are selectively destroyed. Either process would make *Ste* silencing mechanistically different from RNAi. In support of this view, mutations in the central components of the *Drosophila* RNAi pathway *dcr 2*, *r2d2*, and *ago2* did not diminish *Su(Ste)* rasiRNA accumulation (Fig. 2B). Deletion of the *Su(Ste)* silencing trigger (*cry*^{1}Y) caused a factor

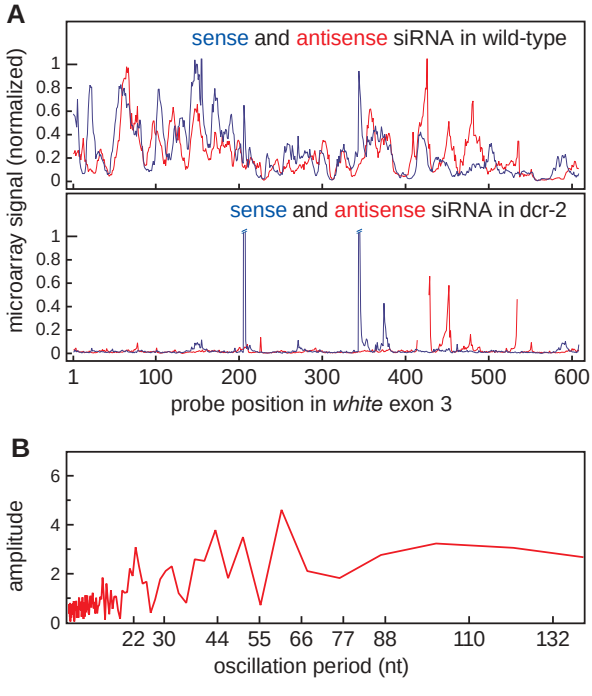


Fig. 1. Sense and antisense siRNAs accumulate during RNAi in vivo. **(A)** Microarray analysis of the siRNAs derived from the *white* exon 3 hairpin RNAi trigger. **(B)** Fourier transform analysis of the Dcr-2–dependent siRNAs in (A).

[1]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA. [2]Department of Animal Molecular Genetics, Institute of Molecular Genetics, Moscow 123182, Russia.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E mail: phillip.zamore@umassmed.edu

of ~65 increase in *Ste* mRNA (Fig. 3A), but null or strong hypomorphic mutations in the three key RNAi proteins did not (Fig. 3B).

Fly Argonaute proteins can be subdivided into the Ago (Ago1 and Ago2) and Piwi [Aubergine (Aub), Piwi, and Ago3] subfamilies. Unlike *ago1* and *ago2*, the *aub*, *piwi*, and *ago3* mRNAs are enriched in the germline (*23 25*). Aub is required for *Ste* silencing (*16*) and *Su(Ste)* rasiRNA accumulation (*26*). In aub^{HN2}/aub^{QC42} trans heterozygous mutants, *Su(Ste)* rasiRNAs were not detected by microarray (fig. S2B) or Northern analysis (Fig. 2B), and *Su(Ste)* triggered silencing of *Ste* mRNA was lost completely (Fig. 3B). Even $aub^{HN2}/+$ heterozygotes accumulated less of the most abundant *Su(Ste)* rasiRNA than did the wild type (Fig. 2B). That the Ago subfamily protein Ago2 is not required for *Ste* silencing, whereas the Piwi subfamily protein Aub is essential for it, supports the view that *Ste* is silenced by a pathway distinct from RNAi. Intriguingly, *Su(Ste)* rasiRNAs hyperaccumulated in *piwi* mutant testes, where *Ste* is silenced normally (Figs. 2B and 3B and fig. S2B).

Mutations in *aub* also cause an increase in sense, but not antisense, *Su(Ste)* RNA (*16*); our results suggest that antisense *Su(Ste)* rasiRNAs can silence both *Ste* mRNA and sense *Su(Ste)* RNA, but that no *Su(Ste)* rasiRNAs exist that can target the antisense *Su(Ste)* transcript. Our finding that *Su(Ste)* rasiRNAs are predominantly or exclusively antisense is essentially in agreement with the results of small RNA cloning experiments, in which four of five *Su(Ste)* rasiRNAs sequenced were in the antisense orientation (*15*), but is at odds with earlier reports detecting both sense and antisense *Su(Ste)* rasiRNAs by nonquantitative Northern hybridization (*16*).

A third RNA silencing pathway in flies. Is germline RNA silencing of selfish genetic elements generally distinct from the RNAi and miRNA pathways? We examined the expression of a panel of germline expressed selfish genetic elements three long terminal repeat (LTR) containing retrotransposons (*roo*, *mdg1*, and

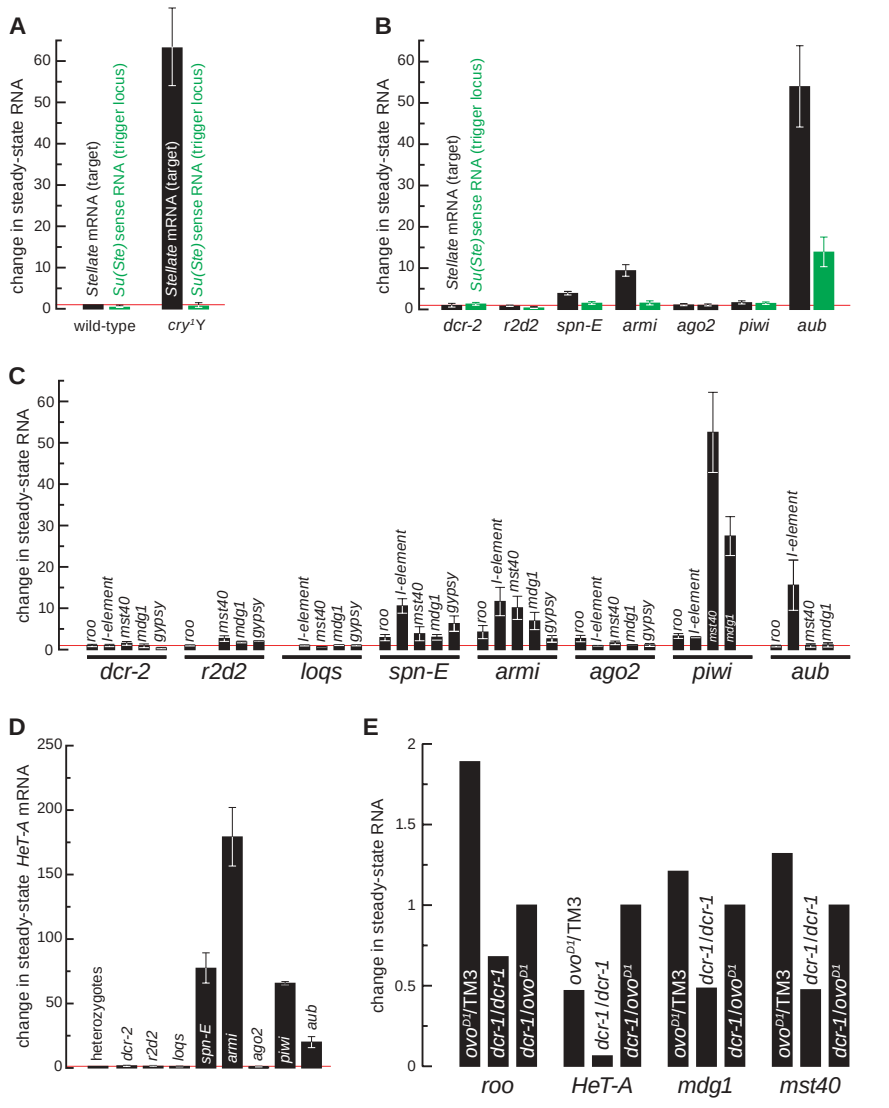


Fig. 3. (**A** to **D**) RNA expression from selfish genetic elements was measured in homozygous mutants relative to heterozygotes for *Ste* silencing in testes [(A) and (B)] and for the repeated locus *mst40*, the LTR retrotransposons *roo*, *mdg1*, and *gypsy*, and the non–LTR retrotransposons *I-element* and *HeT-A* in ovaries [(C) and (D)]. (**E**) RNA expression from selfish genetic elements in $dcr\text{-}1^{Q1147X}$ null mutant clones generated by mitotic recombination in the ovary.

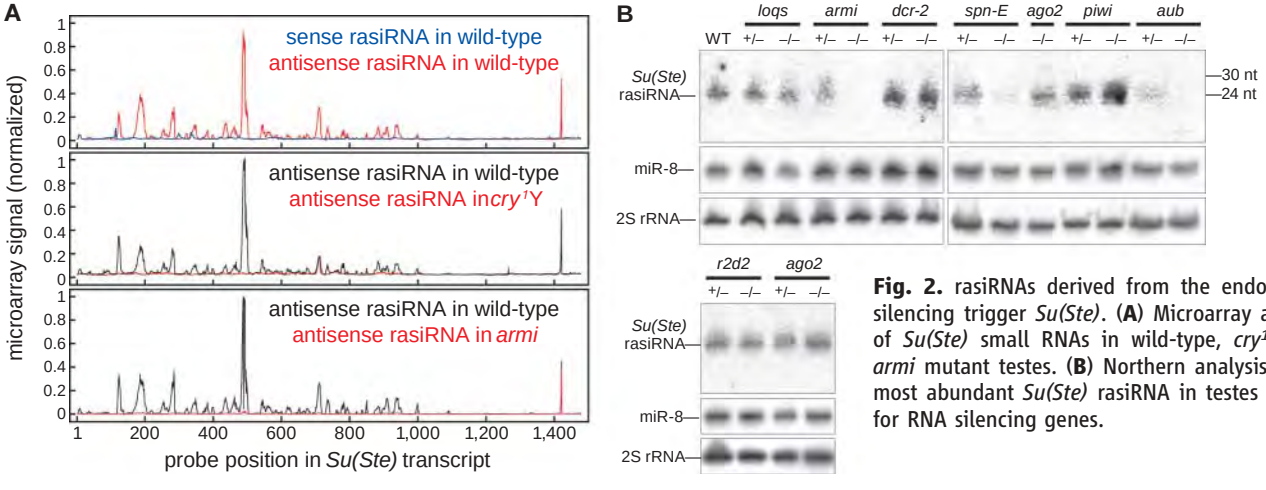


Fig. 2. rasiRNAs derived from the endogenous silencing trigger *Su(Ste)*. (**A**) Microarray analysis of *Su(Ste)* small RNAs in wild-type, cry^1Y, and *armi* mutant testes. (**B**) Northern analysis of the most abundant *Su(Ste)* rasiRNA in testes mutant for RNA silencing genes.

gypsy), two non LTR retrotransposons (*I element* and *HeT A*, a component of the *Drosophila* telomere), and a repetitive locus (*mst40*) in mutants defective for eight RNA silencing proteins. All selfish genetic elements tested behaved like *Ste*: Loss of the RNAi proteins Dcr 2, R2D2, or Ago2 had little or no effect on retrotransposon or repetitive element silencing (Fig. 3, C and D). Instead, silencing required the putative helicases Spn E and Armi plus one or both of the Piwi subfamily Argonaute proteins, Aub and Piwi. Silencing did not require Loqs, the dsRNA binding protein required to produce miRNAs (Fig. 3, C and D).

The null allele *dcr* 1^{Q1147X} is homozygous lethal, making it impossible to procure *dcr 1* mutant ovaries from *dcr* 1^{Q1147X}/*dcr* 1^{Q1147X} adult females (*7*). Therefore, we generated clones of *dcr* 1^{Q1147X}/*dcr* 1^{Q1147X} cells in the ovary by mitotic recombination in flies heterozygous for the dominant female sterile mutation ovo^{D1} (*27*). We measured RNA levels, relative to *rp49* mRNA, for three retrotransposons (*roo*, *HeT A*, and *mdg1*) and one repetitive sequence (*mst40*) in *dcr 1/dcr 1* recombinant ovary clones and in ovo^{D1}/TM3 and *dcr 1*/ovo^{D1} nonrecombinant ovaries. The ovo^{D1} mutation blocks oogenesis at stage 4, after the onset of *HeT A* (*28*, *29*) and *roo* rasiRNA production. Retrotransposon or repetitive sequence transcript abundance was unaltered or decreased in *dcr 1/dcr 1* relative to ovo^{D1}/TM3 and *dcr 1*/ovo^{D1} controls (Fig. 3E). We conclude that Dcr 1 is dispensable for silencing these selfish genetic elements in the *Drosophila* female germline.

roo is the most abundant LTR retrotransposon in flies. We analyzed *roo* silencing in the female germline with the use of microarrays containing 30 nt probes, tiled at 5 nt resolution, for all ~18,000 possible *roo* rasiRNAs (Fig. 4, A and B); we corroborated the data at 1 nt resolution for those rasiRNAs derived from LTR sequences (fig. S3A). As observed for *Su(Ste)* but not for *white* RNAi, *roo* rasiRNAs were nonhomogeneously distributed along the *roo* sequence and accumulated primarily from the antisense strand (Fig. 4A and fig. S3A). In fact, the most abundant sense rasiRNA peak (asterisk in Fig. 4, A and B) corresponded to a set of probes containing 16 contiguous uracil residues, which suggests that these probes nonspecifically detected fragments of the mRNA polyadenylate [poly(A)] tail. Most of the remaining sense peaks were unaltered in *armi* mutant ovaries, in which *roo* expression is increased; this result implies that they do not contribute to *roo* silencing (Figs. 3C and 4A). We detected no phasing in the distribution of *roo* rasiRNAs (fig. S3B).

As for *Su(Ste)*, wild type accumulation of antisense *roo* rasiRNA required the putative helicases Armi and Spn E and the Piwi subfamily Argonaute proteins Piwi and Aub, but not the RNAi proteins Dcr 2, R2D2, and Ago2 (Fig. 4, B and C). Moreover, accumulation of *roo* rasiRNA was not measurably altered in *loqs* ^{f00791}, an allele that strongly disrupts miRNA production in the female germline (Fig. 4C).

Are *roo* rasiRNAs not made by dicing? Loss of Dcr 2 or Dcr 1 did not increase retrotransposon or repetitive element expression, which suggests that neither enzyme acts in rasiRNA directed silencing. Moreover, loss of Dcr 2 had no detectable effect on *Su(Ste)* rasiRNA in testes or *roo* rasiRNA in ovaries (Figs. 2B and 4C). We measured the amount of *roo* rasiRNA and miR 311 in *dcr 1/dcr 1* ovary clones generated by mitotic recombination. Comparison of recombinant (*dcr 1/dcr 1*) and nonrecombinant (ovo^{D1}/ TM3 and *dcr 1*/ovo^{D1}) ovaries by Northern analysis revealed that *roo* rasiRNA accumulation was unperturbed by the null *dcr* 1^{Q1147X} mutation (Fig. 5A and fig. S4). Pre miR 311 increased and miR 311 declined by a factor of ~3 in the *dcr 1/dcr 1* clones (Fig. 5B and fig. S4), consistent with about two thirds of the tissue corresponding to mitotic *dcr 1/dcr 1* recombinant cells. Yet, although most of the tissue lacked *dcr 1* function, we observed improved, rather than diminished, silencing for the four selfish genetic elements examined (Fig. 3E). Moreover, the dsRNA binding protein Loqs, which acts with Dcr 1 to produce miRNAs, was also dispensable for *roo* rasiRNA production and selfish genetic element silencing (Fig. 3, C and D, and Fig. 4C). Although we cannot exclude the possibility that *dcr 1* and *dcr 2* can fully substitute for each other in the production of rasiRNA in the ovary, previous biochemical evidence suggests that none of the three RNase III enzymes in flies Dcr 1, Dcr 2, and Drosha can cleave long dsRNA into small RNAs 24 to 30 nt long (*10*, *30*, *31*).

Animal siRNA and miRNA contain 5′ phosphate and 2′,3′ hydroxy termini (*32*, *33*). We used enzymatic and chemical probing to infer the terminal structure of *roo* and *Su(Ste)* rasiRNAs. RNA from ovaries or testes was treated with calf intestinal phosphatase (CIP) or CIP followed by polynucleotide kinase plus ATP. CIP treatment caused *roo* (Fig. 6A) and *Su(Ste)* (fig. S5) rasiRNA to migrate more slowly in polyacrylamide gel electrophoresis, consistent with the loss

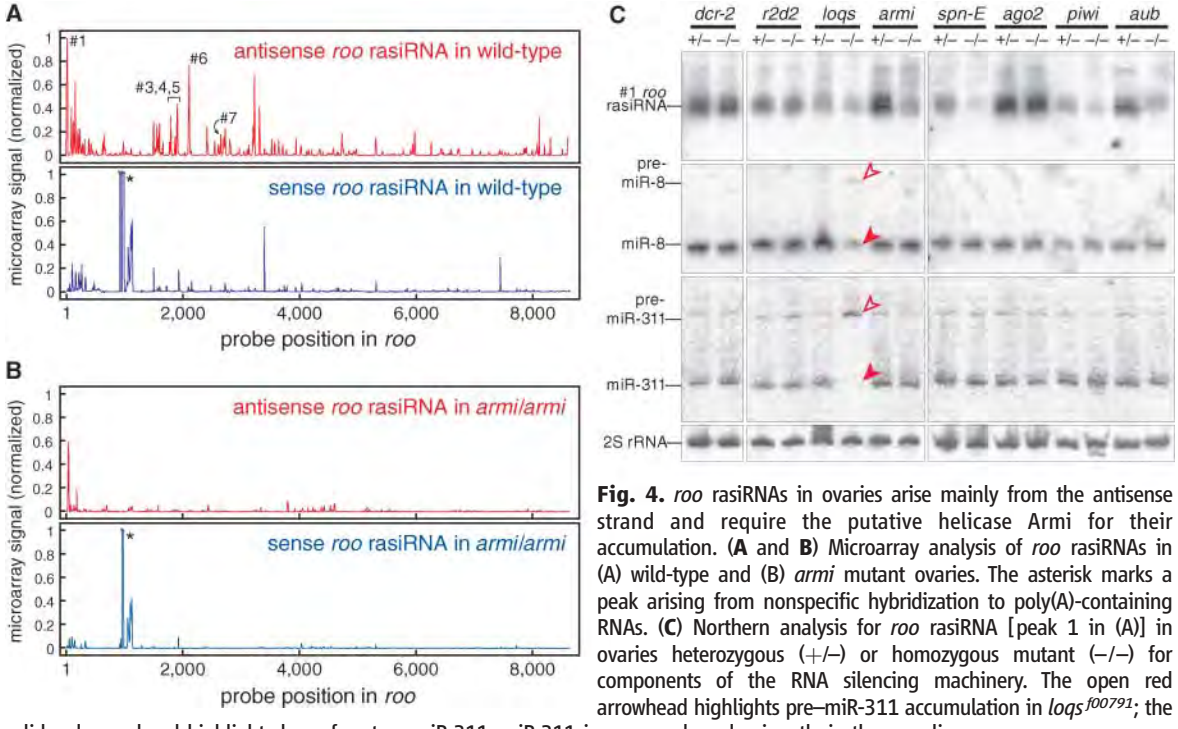


Fig. 4. *roo* rasiRNAs in ovaries arise mainly from the antisense strand and require the putative helicase Armi for their accumulation. (**A** and **B**) Microarray analysis of *roo* rasiRNAs in (A) wild-type and (B) *armi* mutant ovaries. The asterisk marks a peak arising from nonspecific hybridization to poly(A)-containing RNAs. (**C**) Northern analysis for *roo* rasiRNA [peak 1 in (A)] in ovaries heterozygous (+/–) or homozygous mutant (–/–) for components of the RNA silencing machinery. The open red arrowhead highlights pre–miR-311 accumulation in *loqs* ^{f00791}; the solid red arrowhead highlights loss of mature miR-311. miR-311 is expressed predominantly in the germline.

of one or more terminal phosphate groups. Subsequent incubation with polynucleotide kinase and ATP restored the original gel mobility of the rasiRNAs, indicating that they contained a single 5′ or 3′ phosphate before CIP treatment. The *roo* rasiRNA served as a substrate for ligation of a 23 nt 5′ RNA adapter by T4 RNA ligase, a process that requires a 5′ phosphate; pretreatment with CIP blocked ligation (Fig. 6B), thus establishing that the monophosphate lies at the 5′ end. The rasiRNA must also contain at least one terminal hydroxyl group, because it could be joined by T4 RNA ligase to a preadenylated 17 nt 3′ RNA adapter (Fig. 6B). Notably, the 3′ ligation reaction was less efficient for the *roo* rasiRNA than for a miRNA in the same reaction (22% versus 50% conversion to ligated product).

RNA from ovaries or testes was reacted with $NaIO_4$, then subjected to β elimination, to determine whether the rasiRNA had either a single 2′ or 3′ terminal hydroxy group or had terminal hydroxy groups at both the 2′ and 3′ positions, as do animal siRNA and miRNA. Only RNAs containing both 2′ and 3′ hydroxy groups react with $NaIO_4$; β elimination shortens $NaIO_4$ reacted RNA by one nucleotide, leaving a 3′ monophosphate terminus, which adds one negative charge. Consequently, $NaIO_4$ reacted, β eliminated RNAs migrate faster in polyacrylamide gel electrophoresis than does the original unreacted RNA. Both *roo* (Fig. 6A) and *Su(Ste)* (fig. S5) rasiRNA lack either a 2′ or a 3′ hydroxyl group, because they failed to react with $NaIO_4$; miRNAs in the same samples reacted with $NaIO_4$. Together, our results show that rasiRNAs contain one modified and one unmodified hydroxyl. Because T4 RNA ligase can make both 3′ 5′ and 2′ 5′ bonds (*34*), we cannot currently determine the blocked position. Some plant small silencing RNAs contain a 2′ *O* methyl modification at their 3′ terminus (*34*).

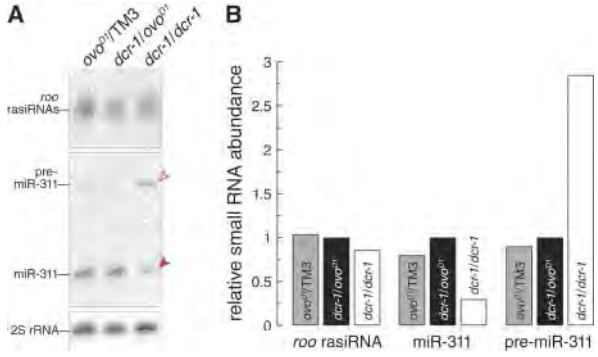


Fig. 5. (**A**) Northern analysis of mitotic recombinant *dcr-1*^{Q1147X} homozygous mutant ovaries and nonrecombinant controls. *roo* rasiRNAs were detected with a mixture of five hybridization probes. Arrowheads are as in Fig. 4C. (**B**) Quantification of the data in (A), normalized to the 2*S* rRNA loading control.

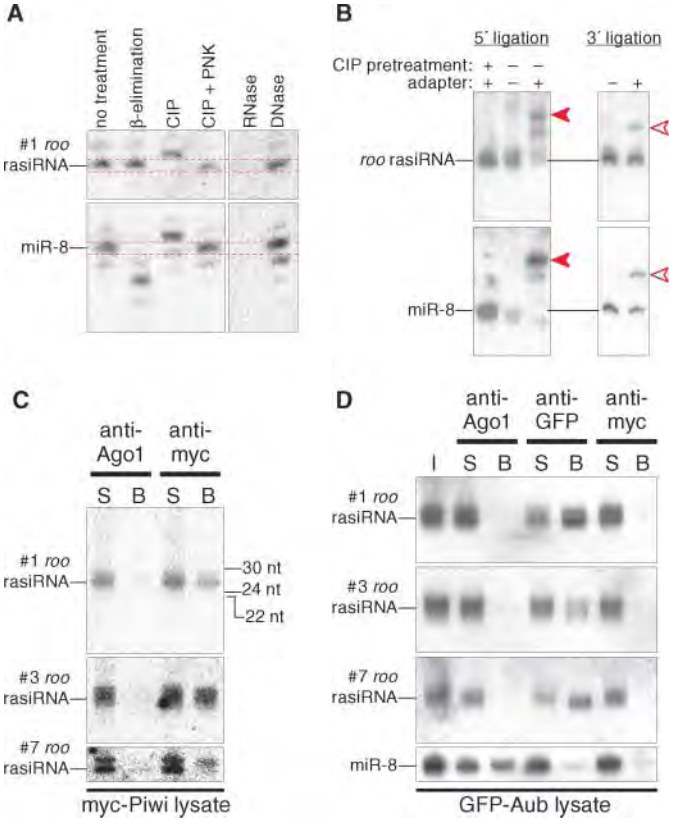


Fig. 6. *roo* rasiRNAs are modified at their 3′ terminus and associate with Piwi subfamily Argonaute proteins. (**A**) Chemical and enzymatic probing of *roo* rasiRNA structure. *roo* rasiRNAs (peak 1 in Fig. 4A) were detected by Northern hybridization. The membrane was then stripped and reprobed for miR-8. (**B**) *roo* rasiRNAs can serve as a 3′ or a 5′ substrate for T4 RNA ligase. Solid arrowheads, 5′ ligation products; open arrowheads, 3′ ligation products. (**C** and **D**) *roo* rasiRNAs associate with myc-tagged Piwi (C) and GFP-tagged Aub protein (D), but not with Ago1. I, input; S, supernatant; B, bound.

rasiRNAs bind Piwi and Aub. *Drosophila* and mammalian siRNA and miRNA function through members of the Ago subfamily of Argonaute proteins, but *Su(Ste)* and *roo* rasiRNAs require at least one member of the Piwi subfamily for their function and accumulation. To determine whether *roo* rasiRNAs physically associate with Piwi and Aub, we prepared ovary lysate from wild type flies or transgenic flies expressing either myc tagged Piwi or green fluorescent protein (GFP) tagged Aub protein (*25*, *35*); immunoprecipitated them with monoclonal antibodies (mAbs) to myc, GFP, or Ago1; and then analyzed the supernatant and antibody bound small RNAs by Northern blotting (Fig. 6, C and D, and fig. S6). We analyzed six different *roo* rasiRNAs. All were associated with Piwi but not with Ago1, the *Drosophila* Argonaute protein typically associated with miRNAs (*36*); miR 8 (Fig. 6D and fig. S6C), miR 311, and *bantam* immunoprecipitated with Ago1 mAb. No rasiRNAs immunoprecipitated with the myc mAb when we used lysate from flies lacking the myc Piwi transgene (fig. S6B).

Although *aub* mutant ovaries silenced *roo* mRNA normally, they showed reduced accumulation of *roo* rasiRNA relative to *aub*/+ heterozygotes (Fig. 4C), which suggests that *roo* rasiRNAs associate with both Piwi and Aub. We analyzed the supernatant and antibody bound small RNAs after GFP mAb immunoprecipitation of ovary lysate from GFP Aub transgenic flies and flies lacking the transgene. *roo* rasiRNA was recovered only when the immunoprecipitation was performed with the GFP mAb in ovary lysate from GFP Aub transgenic flies (Fig. 6D and fig. S6D). The simplest interpretation of our data is that *roo* rasiRNAs physically associate with both Piwi and Aub, although it remains possible that the *roo* rasiRNAs are loaded only into Piwi and that Aub associates with Piwi in a stable complex. The association of *roo* rasiRNA with both Piwi and Aub suggests that *piwi* and *aub* are partially redundant, as does the modest reduction in *roo* silencing by a factor of 3.3 ± 0.6 (average ± SD, n 3) in *piwi* but not in *aub* mutants (Fig. 3C). Alternatively, *roo* silencing might proceed through Piwi alone, but the two proteins could function in the same pathway to silence selfish genetic elements.

Our data suggest that in flies, rasiRNAs are produced by a mechanism that requires neither Dcr 1 nor Dcr 2, yet the patterns of rasiRNAs that direct *roo* and *Ste* silencing are as stereotyped as the distinctive siRNA population generated from the *white* hairpin by Dcr 2 (Fig. 1A) or the unique miRNA species made from each pre miRNA by Dcr 1. A key challenge for the future will be to determine what enzyme makes rasiRNAs and what sequence or structural features of the unknown rasiRNA precursor lead to the accumulation of a stereotyped pattern of predominantly antisense rasiRNAs.

References and Notes

1. D. Galiana Arnoux, C. Dostert, A. Schneemann, J. A. Hoffmann, J. L. Imler, *Nat. Immunol.* **7**, 590 (2006).

2. X. H. Wang *et al.*, *Science* **312**, 452 (2006); published online 22 March 2006 (10.1126/science.1125694).
3. R. A. Zambon, V. N. Vakharia, L. P. Wu, *Cell. Microbiol.* **8**, 880 (2006).
4. T. Du, P. D. Zamore, *Development* **132**, 4645 (2005).
5. A. I. Kalmykova, M. S. Klenov, V. A. Gvozdev, *Nucleic Acids Res.* **33**, 2052 (2005).
6. A. Fire *et al.*, *Nature* **391**, 806 (1998).
7. Y. S. Lee *et al.*, *Cell* **117**, 69 (2004).
8. N. T. Hoa, K. M. Keene, K. E. Olson, L. Zheng, *Insect Biochem. Mol. Biol.* **33**, 949 (2003).
9. K. Forstemann *et al.*, *PLoS Biol.* **3**, e236 (2005).
10. K. Saito, A. Ishizuka, H. Siomi, M. C. Siomi, *PLoS Biol.* **3**, e235 (2005).
11. F. Jiang *et al.*, *Genes Dev.* **19**, 1674 (2005).
12. D. S. Schwarz *et al.*, *Cell* **115**, 199 (2003).
13. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).
14. Y. S. Lee, R. W. Carthew, *Methods* **30**, 322 (2003).
15. A. A. Aravin *et al.*, *Dev. Cell* **5**, 337 (2003).
16. A. A. Aravin *et al.*, *Curr. Biol.* **11**, 1017 (2001).
17. P. Tritto *et al.*, *Genetica* **117**, 247 (2003).
18. Y. Tomari *et al.*, *Cell* **116**, 831 (2004).
19. H. A. Cook, B. S. Koppetsch, J. Wu, W. E. Theurkauf, *Cell* **116**, 817 (2004).
20. T. Dalmay, R. Horsefield, T. H. Braunstein, D. C. Baulcombe, *EMBO J.* **20**, 2069 (2001).
21. G. Meister *et al.*, *Curr. Biol.* **15**, 2149 (2005).
22. W. Stapleton, S. Das, B. D. McKee, *Chromosoma* **110**, 228 (2001).
23. R. W. Williams, G. M. Rubin, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6889 (2002).
24. D. N. Cox *et al.*, *Genes Dev.* **12**, 3715 (1998).
25. A. N. Harris, P. M. Macdonald, *Development* **128**, 2823 (2001).
26. A. A. Aravin *et al.*, *Mol. Cell. Biol.* **24**, 6742 (2004).
27. I. Sahut Barnola, D. Pauli, *Development* **126**, 1917 (1999).
28. V. V. Vagin *et al.*, *RNA Biol.* **1**, 54 (2004).
29. M. Savitsky, D. Kwon, P. Georgiev, A. Kalmykova, V. Gvozdev, *Genes Dev.* **20**, 345 (2006).
30. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
31. Q. Liu *et al.*, *Science* **301**, 1921 (2003).
32. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev.* **15**, 188 (2001).
33. G. Hutvágner *et al.*, *Science* **293**, 834 (2001); published online 12 July 2001 (10.1126/science.1062961).
34. Z. Yang, Y. W. Ebright, B. Yu, X. Chen, *Nucleic Acids Res.* **34**, 667 (2006).
35. D. N. Cox, A. Chao, H. Lin, *Development* **127**, 503 (2000).
36. K. Okamura, A. Ishizuka, H. Siomi, M. C. Siomi, *Genes Dev.* **18**, 1655 (2004).

37. We thank T. Du, M. Ghildiyal, and A. Boucher for assistance with fly husbandry, and J. Birchler, R. Carthew, F. Gao, R. Lehmann, H. Lin, P. MacDonald, M. Siomi, H. Siomi, D. Smith, E. Sontheimer, H. Ruohola Baker, and M. Van Doren for fly stocks. P.D.Z. is a W.M. Keck Foundation Young Scholar in Medical Research. Supported in part by NIH grants GM62862 and GM65236 (P.D.Z.) and postdoctoral fellowships from the European Molecular Biology Organization (EMBO) and the Human Frontier Science Program (H.S.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1129333
Materials and Methods
Figs. S1 to S7
Table S1
References

1 May 2006; accepted 19 June 2006
Published online 29 June 2006;
10.1126/science.1129333
Include this information when citing this paper.

α-Synuclein Blocks ER-Golgi Traffic and Rab1 Rescues Neuron Loss in Parkinson's Models

Antony A. Cooper,[1]*† Aaron D. Gitler,[2]* Anil Cashikar,[2]‡ Cole M. Haynes,[1]§ Kathryn J. Hill,[1]† Bhupinder Bhullar,[2,3] Kangning Liu,[4,5] Kexiang Xu,[4] Katherine E. Strathearn,[6] Fang Liu,[6] Songsong Cao,[7] Kim A. Caldwell,[7] Guy A. Caldwell,[7] Gerald Marsischky,[3] Richard D. Kolodner,[8] Joshua LaBaer,[3] Jean-Christophe Rochet,[6] Nancy M. Bonini,[4,5] Susan Lindquist[2,9]‖

Alpha-synuclein (αSyn) misfolding is associated with several devastating neurodegenerative disorders, including Parkinson's disease (PD). In yeast cells and in neurons αSyn accumulation is cytotoxic, but little is known about its normal function or pathobiology. The earliest defect following αSyn expression in yeast was a block in endoplasmic reticulum (ER)–to–Golgi vesicular trafficking. In a genomewide screen, the largest class of toxicity modifiers were proteins functioning at this same step, including the Rab guanosine triphosphatase Ypt1p, which associated with cytoplasmic αSyn inclusions. Elevated expression of Rab1, the mammalian *YPT1* homolog, protected against αSyn-induced dopaminergic neuron loss in animal models of PD. Thus, synucleinopathies may result from disruptions in basic cellular functions that interface with the unique biology of particular neurons to make them especially vulnerable.

Parkinson's disease (PD) is the second most common neurodegenerative disorder (*1*, *2*). Accruing evidence points to a causative role for the presynaptic protein alpha synuclein (αSyn) in PD pathogenesis. αSyn is a major constituent of Lewy Bodies cellular inclusions that are the hallmark pathological feature of PD and other neurodegenerative disorders collectively referred to as synucleinopathies (*3*). Moreover, missense mutations in the αSyn gene (A53T, A30P, E46K) (*4 6*) and duplication or triplication of the locus cause PD (*7 9*). In mouse, rat, fly, and nematode models of PD, increased levels of αSyn lead to neurodegeneration (*10 13*). Elucidating the mechanisms underlying the cytotoxic effects of αSyn will be essential for the development of treatments to ameliorate or prevent the synucleinopathies.

Despite extensive study, little is known about αSyn's normal function or how αSyn contributes to disease. Many cellular defects have been implicated in the etiology of synucleinopathies, including impairment of the ubiquitin proteasome system, mitochondrial dysfunction, accumulation of lipid droplets, production of reactive oxygen species (ROS), and stress within the ER (*14*). A yeast PD model, with dosage sensitivity for αSyn expression, recapitulates many of these defects (*15*). But which are cause and which effect remain unclear. Here, two independent approaches, genetic and cell biological, converged to identify inhibition of ER Golgi trafficking as a major component of synuclein dependent toxicity.

αSyn accumulation causes ER stress. An increase in αSyn gene dosage in yeast from one copy (no growth defect) to two copies results in growth arrest and cell death (*15*) (Fig. 1A). To investigate the earliest defects caused by αSyn, we took advantage of the ability to rapidly and synchronously induce its expression from a galactose inducible promoter. A slight decline in viability was observed after 4 hours of induction, and 60% of cells lost colony forming ability by 8 hours (Fig. 1, A and B). ER stress, measured by a reporter for the unfolded protein response, appeared earlier. Expression of wild type αSyn (αSyn WT) or disease associated αSyn (αSyn A53T) caused a fourfold increase in ER stress relative to control cells after 6 hours (Fig. 1C).

αSyn accumulation impairs degradation of selective ERAD substrates. ER stress typically results from the accumulation of misfolded proteins within the ER. Such malformed proteins are retrotranslocated from the ER to the cytoplasm for degradation by the proteasome through a process termed ERAD (endoplasmic reticulum associated degradation) (*16*). Misfolded cytosolic αSyn might impair the proteasome's capacity for protein degradation and so cause an accumulation of misfolded proteins in the ER and associated ER stress. To investigate this possi

[1]School of Biological Sciences, University of Missouri Kansas City, Kansas City, MO 64110, USA. [2]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [3]Harvard Institute of Proteomics, 320 Charles Street, Cambridge, MA 02141, USA. [4]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Howard Hughes Medical Institute, Philadelphia, PA 19104, USA. [6]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907 2091, USA. [7]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA. [8]Ludwig Institute for Cancer Research, University of California, San Diego, School of Medicine, La Jolla, CA 92093, USA. [9]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

*These authors contributed equally to this work.
†Present address: Garvan Institute of Medical Research, Sydney, Australia.
‡Present address: Medical College of Georgia, Augusta, GA, USA.
§Present address: New York University, New York, NY, USA.
‖To whom correspondence should be addressed. E mail: lindquist admin@wi.mit.edu

Current Biology 17, 1265 1272, July 17, 2007 DOI 10.1016/j.cub.2007.06.030

Report

The *Drosophila* RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC

Michael D. Horwich,[1,3] Chengjian Li,[1,3] Christian Matranga,[1,3] Vasily Vagin,[1,2] Gwen Farley,[1] Peng Wang,[1] and Phillip D. Zamore[1,*]

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

[2] Department of Animal Molecular Genetics
Institute of Molecular Genetics
Moscow 123182
Russia

Summary

Small silencing RNAs repress gene expression by a set of related mechanisms collectively called RNA-silencing pathways [1, 2]. In the RNA interference (RNAi) pathway [3], small interfering mRNA (siRNAs) defend cells from invasion by foreign nucleic acids, such as those produced by viruses. In contrast, microRNAs (miRNAs) sculpt endogenous mRNA expression [4]. A third class of small RNAs, Piwi-interacting RNAs (piRNAs), defends the genome from transposons [5–9]. Here, we report that *Drosophila* piRNAs contain a 2′-*O*-methyl group on their 3′ termini; this is a modification previously reported for plant miRNAs and siRNAs [10] and mouse and rat piRNAs [11, 12, 13]. Plant small-RNA methylation is catalyzed by the protein HEN1 [10, 14, 15]. We find that DmHen1, the *Drosophila* homolog of HEN1, methylates the termini of siRNAs and piRNAs. Without DmHen1, the length and abundance of piRNAs are decreased, and piRNA function is perturbed. Unlike plant HEN1, DmHen1 acts on single strands, not duplexes, explaining how it can use as substrates both siRNAs—which derive from double-stranded precursors—and piRNAs—which do not [8, 13]. 2′-*O*-methylation of siRNAs may be the final step in assembly of the RNAi-enzyme complex, RISC, occurring after an Argonaute-bound siRNA duplex is converted to single-stranded RNA.

Results and Discussion

Drosophila piRNAs Are 2′-*O*-methylated at Their 3′ Termini

In flies, both piRNAs (also known as repeat-associated siRNAs, rasiRNAs) and siRNAs, but not miRNAs, are modified at their 3′ termini [8, 16]. We selectively labeled (Figure S1 in the Supplemental Data available online) the terminal nucleotide of *Drosophila melanogaster* 0–2 hr embryo and mouse and bull testicular piRNAs. The resulting ^{32}P-radiolabeled nucleoside 2′ or 3′-monophosphates were resolved by 2D thin-layer chromatography (2D TLC) with a solvent system that can resolve nucleoside 2′ monophosphates, nucleoside 3′ monophosphates, and 2′-*O*-methyl nucleoside 3′ monophosphates (Figure S2). Modified nucleoside monophosphates derived from the 3′ termini of piRNAs were identified by comparison to modified and unmodified nucleoside 2′ and 3′ monophosphate standards (Figure 1A). The terminal nucleotide of the piRNAs of all three animals comigrated with 2′-*O*-methyl nucleoside 3′ monophosphate standards but not with any unmodified nucleoside monophosphate standard. Because mouse piRNAs were previously shown to contain 2′-*O*-methyl modified 3′ termini by both mass spectrometry [12] and a 2D TLC system [11] distinct from ours, we conclude that *Drosophila* and bull piRNAs also contain a 2′-*O*-methyl group at their 3′ termini.

DmHen1 Is Required for piRNA Modification In Vivo

In *Arabidopsis*, the RNA methyltransferase, HEN1, modifies the terminal 2′ hydroxyl group of small silencing RNAs. In *Drosophila*, predicted gene CG12367, whose 1559 nucleotide mRNA encodes a 391 amino acid protein with a 220 amino acid evolutionarily conserved methyltransferase domain [17], most closely resembles *Arabidopsis* HEN1 (Figure 1B) [18]. For simplicity, we call this gene *Drosophila melanogaster* (Dm) *hen1*. When homozygous, a piggyBac transposon insertion (PBac{WH}CG12367[f00810]) within the first intron of the fly *hen1* gene reduces the accumulation of *hen1* mRNA by 1000-fold in testes and by more than 40,000-fold in ovaries (Figure 2A) and can therefore be considered a null mutation, which we refer to as *hen1*^{f00810}.

We examined the 3′ termini of two types of highly abundant piRNAs in the germline of flies heterozygous or homozygous for *hen1*^{f00810}. In testes, the *Suppressor of Stellate* [*Su(Ste)*] locus produces 24–27 nucleotide rasiRNAs, a subclass of piRNAs that directs silencing of the selfish genetic element *Stellate*. *Su(Ste)* rasiRNAs, like other *Drosophila* piRNAs, are modified at their 3′ termini and therefore do not react with $NaIO_4$ [8]. In contrast, *Su(Ste)* rasiRNAs from *hen1*^{f00810}/*hen1*^{f00810} mutant testes reacted with $NaIO_4$ and could therefore be β-eliminated to remove the last nucleotide of the RNA, thereby increasing their gel mobility (Figure 2B) and indicating that in the absence of DmHen1 protein, they are not modified. Similarly, rasiRNAs that guide silencing of *roo*, the most abundant retrotransposon in *Drosophila melanogaster*, were not modified in *hen1*^{f00810} homozygous ovaries (Figure 2C). The *Su(Ste)* and *roo* rasiRNAs were also shorter in the *hen1*^{f00810} homozygotes. In contrast, the length and amount of miR-8, which is expressed in both the male and female germline, was unaltered in *hen1*^{f00810} homozygotes. For both *Su(Ste)* and *roo*, rasiRNAs were on average shorter and less modified even in *hen1*^{f00810} heterozygotes, compared to the wild-type, suggesting that the abundance of DmHen1 protein limits the stability or production of piRNAs in flies.

*Correspondence: phillip.zamore@umassmed.edu
[3] These authors contributed equally to this work.

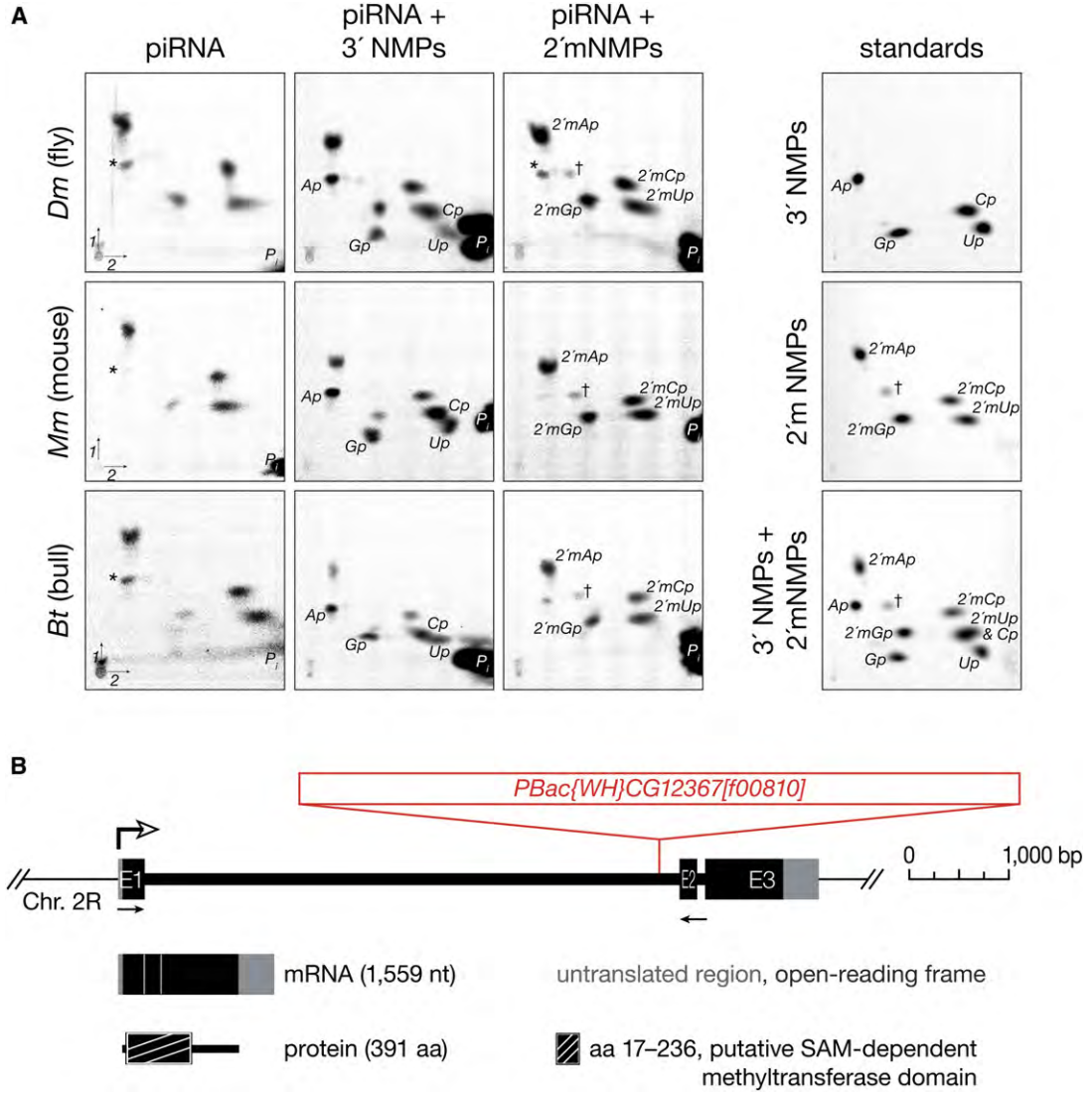


Figure 1. 2′ *O* Methylation of piRNAs in *Drosophila*

(A) *Drosophila* piRNAs are 2′ *O* methylated at their 3′ termini. The modified nucleotides on the 3′ termini of piRNAs from 0 2 hr fly embryos and mouse and bull testes were selectively ^{32}P radiolabeled. The radiolabeled modified mononucleotides from each species were resolved by 2D TLC individually (piRNA), with ^{32}P radiolabeled 3′ mononucleotide standards (piRNA + 3′NMPs), and with ^{32}P radiolabeled 2′ *O* methyl, 3′ phos phate mononucleotide standards (piRNA + 2′mNMPs). The modified nucleotides from the piRNA from all three animals comigrated with 2′mNMPs standards, but not with 3′NMPs standards. The following abbreviations are used: Ap, 3′AMP; Gp, 3′GMP; Cp, 3′CMP; Up, 3′UMP; 2′mAp, 2′ *O* methyl AMP; 2′mGp, 2′ *O* methyl GMP; 2′mCp, 2′ *O* methyl CMP; 2′mUp, 2′ *O* methyl UMP; and P_i, phosphate. The asterisk marks a contaminant, probably 3′ AMP, present in the [5′ ^{32}P] cytidine 5′,3′ bis phosphate used to radiolabel the piRNA. The cross marks a contaminant present in the 2′ *O* methyl, 3′ phosphate mononucleotide standards.

(B) *Drosophila hen1* gene (CG12367), mRNA, and protein. The piggyBac transposon, PBac{WH}CG12367[f00810], is inserted 207 bp upstream of the second exon. The open arrow indicates the predicted start of transcription. The closed arrows denote the position of the qRT PCR primers used in Figure 2A. The first intron of *hen1* contains another gene, CG8878, that is transcribed in the opposite direction and whose expression is unaltered by the piggyBac insertion (Figure S3).

DmHen1 Is Required for piRNA Function In Vivo

Modification of the termini of *Drosophila* piRNAs plays a role in their function: mRNA expression from *HeT-A*, the element whose expression is most sensitive to mutations that disrupt piRNA-directed silencing in the female germline [8, 19, 20], quadrupled in *hen1*^{f00810} heterozygotes and was increased by more than 11-fold in homozygotes, relative to wild-type tissue (Figure 2D). We conclude that Hen1 protein is required for piRNA-directed silencing in the *Drosophila* germline.

DmHen1 Is Required for siRNA Modification

To test whether DmHen1 is required for modification of the 3′ termini of siRNAs, we depleted Hen1 by RNAi in cultured *Drosophila* S2 cells. We transfected the cells with long double-stranded RNA (dsRNA) targeting *hen1* on day 1 and day 5, then cotransfected them with both GFP dsRNA and *hen1* dsRNA on day 8. Total RNA was harvested on day 9, probed for modification with $NaIO_4$/β-elimination, and analyzed by Northern hybridization with a 5′ ^{32}P-radiolabeled DNA probe

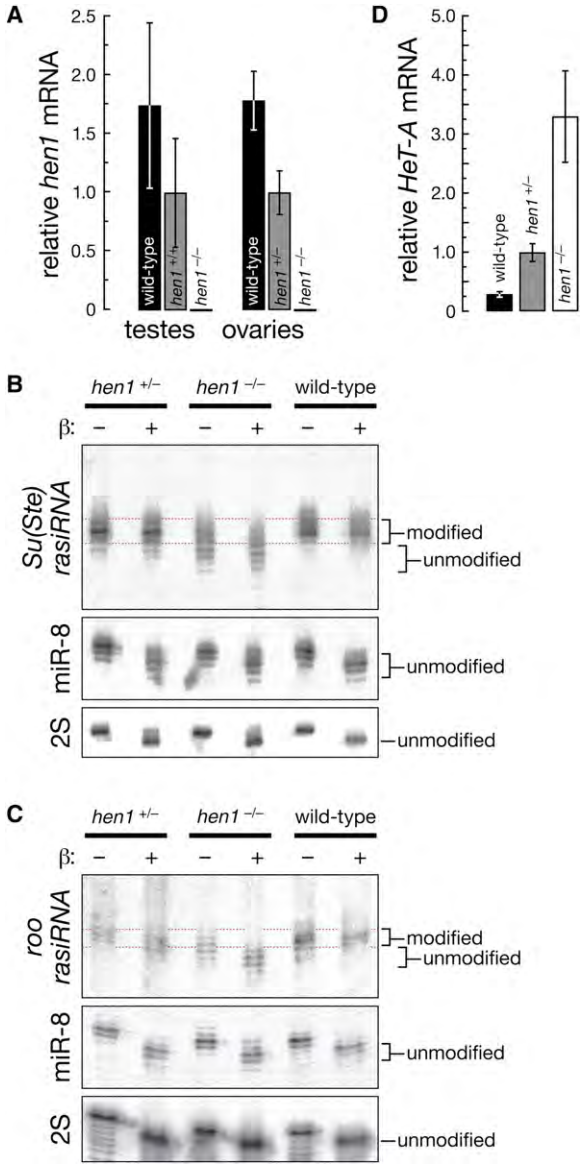


Figure 2. DmHen1 Is Required for Normal piRNA Biogenesis and Complete Silencing of the *HeT A* Transposon

(A) Quantitative RT PCR of *hen1* mRNA in testes and ovaries from wild type or *hen1*^{f00810} flies, relative to *rp49*. Error bars report the average ± SD for at least four independent experiments.

(B and C) Northern hybridization to detect modification of small RNAs in wild type and *hen1*^{f00810} testes (B) or ovaries (C). The same blot was probed sequentially to detect the most abundant *Su(Ste)* rasiRNA (B) or the three most abundant *roo* rasiRNAs (C), miR 8, and 2S ribosomal RNA.

(D) Quantitative RT PCR of *HeT A* transposon mRNA in ovaries. *HeT A* mRNA levels were measured relative to *rp49* with total RNA prepared from wild type and *hen1*^{f00810} heterozygous or homozygous ovaries. Error bars report the average ± SD for at least four independently prepared samples.

complementary to the most abundant GFP-derived siRNA (M.D.H., M. Ghildiyal, and P.D.Z., unpublished data). DsRNAs targeting two different regions of the fly *hen1* mRNA both reduced the amount of GFP siRNA modified at its 3′ terminus, whereas all the GFP siRNA remained modified when a control dsRNA was used (Figure 3A).

Surprisingly, RNAi-mediated depletion of Ago2, but not Ago1, prevented the GFP siRNA from being modified. This result suggests that Ago2, but not Ago1, plays a role in the modification of siRNAs by DmHen1. To test this idea, we examined the modification status of the 3′ terminus of miR-277, which partitions into both Ago1 and Ago2 complexes in vivo [21]. *Drosophila* miRNAs associate predominantly or exclusively with Ago1 [22] and have unmodified 3′ termini [8, 16, 23]. In contrast, approximately half the miR-277 in cultured S2 cells failed to react with $NaIO_4$ (Figure 3A), suggesting that approximately half of miR-277 is modified at its 3′ terminus. The fraction of miR-277 that was modified was reduced when two different dsRNAs were used to deplete DmHen1 by RNAi. When the cells were treated with dsRNA targeting *ago1*, all detectable miR-277 was modified, whereas all miR-277 became unmodified when dsRNA targeting *ago2* was used. In contrast, *bantam*, a miRNA that associates nearly exclusively with Ago1 [22], was unmodified under all conditions (Figure 3A).

siRNA Modification Correlates with Ago2-RISC Assembly In Vitro

siRNA modification can be recapitulated in lysates of embryos, ovaries, or cultured S2 cells. Modification of siRNA in vitro was inhibited by *S*-adenosyl homocysteine, but not by *S*-adenosyl methionine, consistent with DmHen1 transferring a methyl group from *S*-adenosyl methionine to the terminal 2′ hydroxyl group of the RNA, thereby generating *S*-adenosyl homocysteine as a product (Figure S4).

Our data from cultured S2 cells suggested that DmHen1 modifies that portion of miR-277 that enters the Ago2-RISC-assembly pathway, but not the population of miR-277 that assembles into Ago1-RISC. To further test the idea that small-RNA modification requires both Hen1 and the Ago2-RISC-assembly pathway, we prepared cytoplasmic lysates from dsRNA-treated cultured S2 cells. Lysate from control-treated cells modified the 3′ terminus of a 5′ ^{32}P-radiolabeled synthetic siRNA duplex but not lysate from *hen1*-depleted cells (Figure 3B). The addition of either of two different preparations of purified, recombinant DmHen1, expressed in *E. coli* as a ~74 kDa glutathione S-transferase fusion protein (GST-DmHen1; Figure S5), restored the ability of the lysates to modify the siRNA, indicating that loss of DmHen1 caused the loss of siRNA modification. Moreover, lysates depleted for Ago2, but not Ago1, could not modify the ^{32}P-siRNA in vitro (Figure 3B). These in vitro data, together with our S2-cell experiments, suggest that modification of the 3′ terminus of siRNAs and miRNAs is coupled to assembly into Ago2-RISC.

Dcr-2 and R2D2 act to load double-stranded siRNAs into Ago2. We prepared lysates from ovaries homozygous mutant for *hen1*, *dcr-2*, *r2d2*, and *ago2* by using alleles that were unable to produce the corresponding protein [22, 24, 25]. A 5′ ^{32}P-radiolabeled siRNA duplex was incubated in each lysate to assemble RISC. At each time point, we determined whether the siRNA was 3′ terminally modified by assessing its reactivity with $NaIO_4$ (Figure 3C). No modified siRNA accumulated when the duplex was incubated in *hen1*^{f00810}, *dcr-2*^{L811fsX}, *r2d2*^{1}, or *ago2*^{414} mutant lysate. Adding

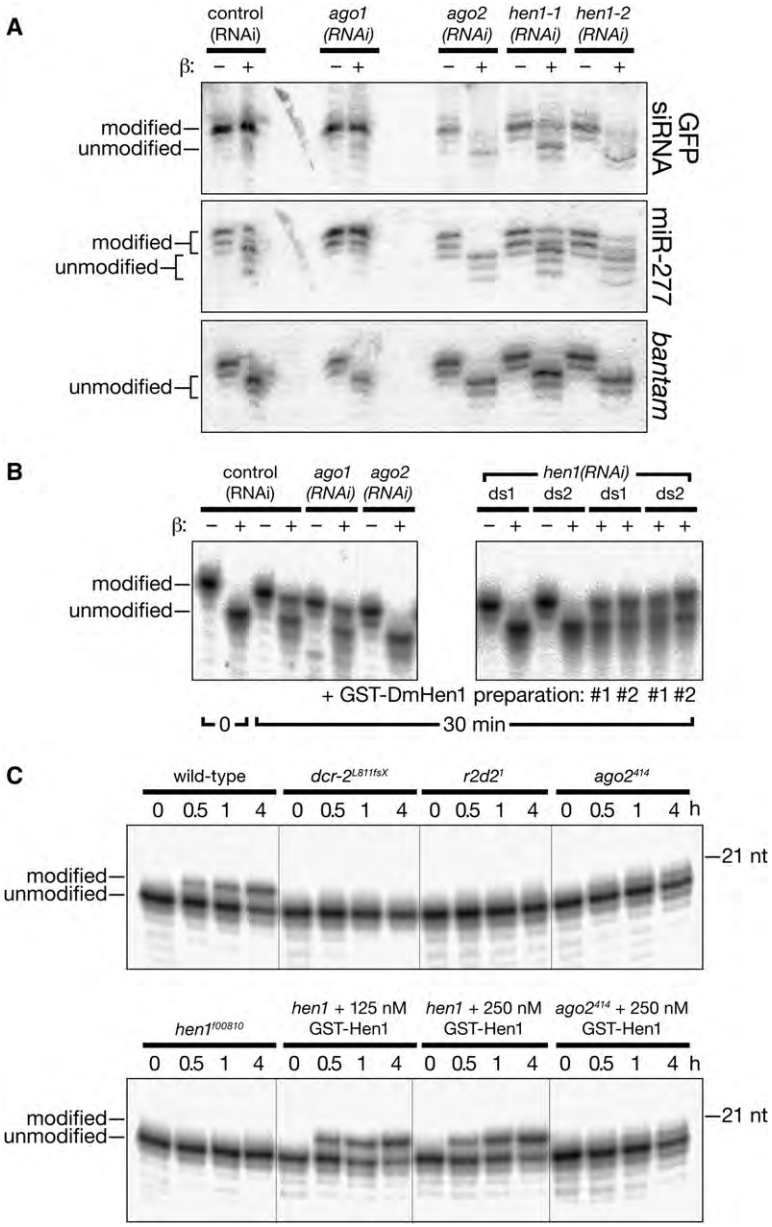


Figure 3. DmHen1 Modifies Ago2 Associated Small RNAs

(A) Modification of siRNAs and miRNAs in *Drosophila* S2 Cells. A stable S2 cell line expressing GFP was treated with the indicated dsRNA alone (day 1 and 5), then together with GFP dsRNA (day 8). Two nonoverlapping dsRNAs were used to target *hen1*. Total RNA was collected on day 9 and treated with $NaIO_4$/β elimination; then, dsGFP derived siRNA, miR 277, and *bantam* were detected by sequential Northern hybridization of the same blot.

(B) In vitro siRNA modification in dsRNA treated S2 cell lysates.

(C) In vitro siRNA modification in *Drosophila* mutant ovary lysates.

250 nM purified, recombinant GST-DmHen1 restored siRNA modification to the *hen1*^{f00810} but not the *ago2*^{414} lysate. We conclude that the defect in *ago2*^{414} reflects a requirement for Ago2 in small-RNA modification by DmHen1, rather than an indirect effect such as destabilization of DmHen1 in the absence of Ago2. GST-DmHen1 similarly rescued lysate from *hen1(RNAi)* but not *ago2(RNAi)*-treated S2 cells (Figure S6). Together, the results of our experiments using cultured S2 cells—a somatic-cell line—and ovaries, which comprise mainly germline tissue, suggest that a functional Ago2-RISC-assembly pathway is required for siRNA modification in *Drosophila*.

siRNAs Are Modified Only after Ago2-RISC Maturation

To test at which step in the Ago2-RISC-assembly pathway siRNAs become modified, we determined whether siRNAs are 2′-*O*-methylated by DmHen1 as single strands or as duplexes. In vitro, assembly of siRNAs into Ago2-RISC follows an ordered pathway in which the siRNA duplex first binds the Dicer-2/R2D2 heterodimer to form the RISC-loading complex (RLC). The RLC determines which of the two siRNA strands will become the guide for Ago2 and which will be destroyed (the passenger strand). The siRNA is then loaded into Ago2 as a duplex. In this pre-RISC complex, the passenger strand occupies the same position as future target RNAs [26]. Cleavage of the passenger strand by the Ago2 endonuclease domain converts pre-RISC to mature RISC [27–30]. No single-stranded guide or passenger RNA is produced prior to this maturation step. Thus, all single-stranded siRNA produced in vitro or in vivo [26] corresponds to mature RISC.

We assembled Ago2-RISC in vitro by using an siRNA designed to load only one of its two strands into Ago2

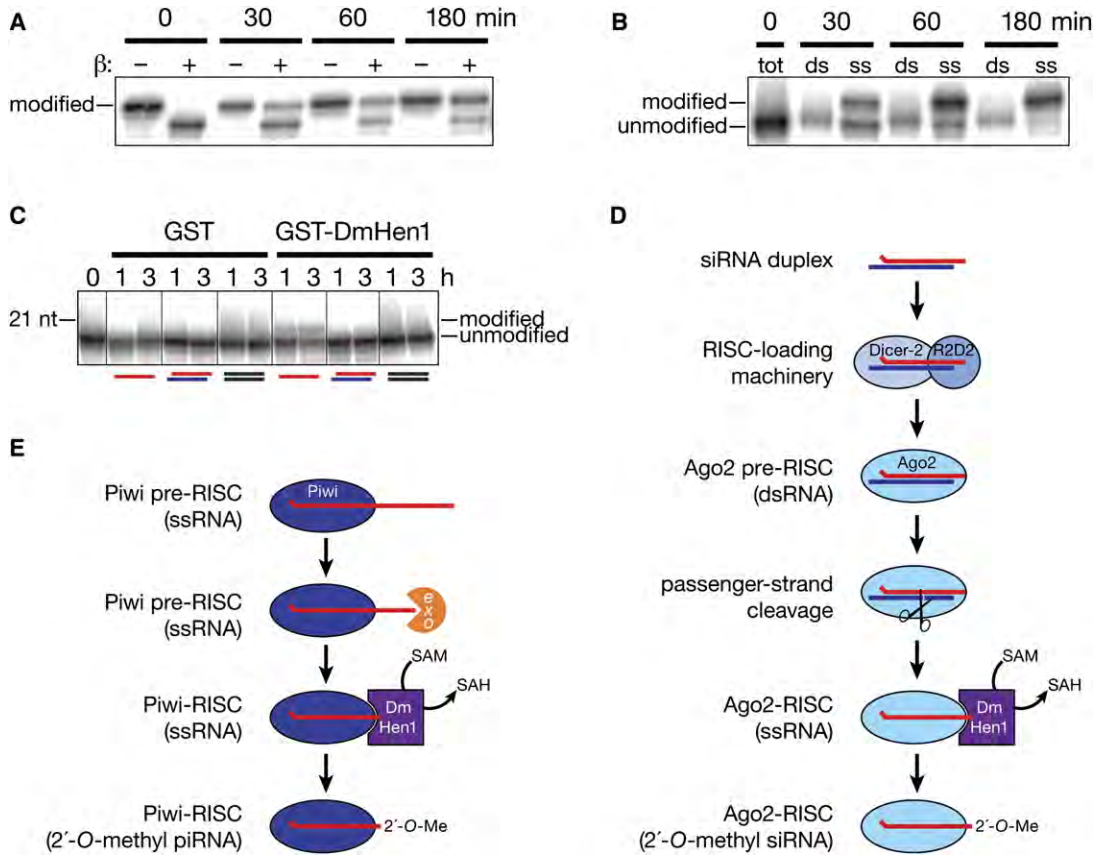


Figure 4. DmHen1 Modifies Single Stranded RNAs in RISC

siRNAs are modified after the conversion of pre RISC which contains double stranded siRNA into mature RISC, which contains only single stranded siRNA. siRNA duplexes with a 5′ ^{32}P radiolabeled guide strand were incubated in *Drosophila* embryo lysate and then tested for the presence of a 3′ terminal modification.

(A) Total RNA from each time point in RISC assembly, without ( ) and with (+) reaction with $NaIO_4$ and β elimination.

(B) At each time in (A), single and double stranded siRNA were resolved and purified by native gel electrophoresis (Figure S7), then analyzed by denaturing electrophoresis separately for the presence of a 3′ terminal modification on the siRNA guide strand. (A) and (B) show the left and right halves of a single gel.

(C) Recombinant, purified GST DmHen1, but not GST alone, can modify single stranded 21 nt RNA, but not double stranded siRNAs or blunt 21 nt RNA duplexes. However, in contrast to GST DmHen1 incubated with *hen1*^{f00810} mutant ovary lysate (Figure 3C), the enzyme alone is inefficient. All samples were oxidized with $NaIO_4$, then β eliminated.

(D) A model for 2′ *O* methylation of siRNAs.

(E) A proposed role for 2′ *O* methylation in piRNA biogenesis.

[31]. We then sampled the reaction over time, isolating the 5′ ^{32}P-radiolabeled siRNA under conditions previously demonstrated to preserve its structure [32], and separated single- from double-stranded siRNA by native gel electrophoresis (Figure S7). (Full-length siRNA duplexes and siRNA heteroduplexes comprising a full-length guide paired to a cleaved passenger strand comigrate as double-stranded siRNA in these gel conditions [28].) The RNAs were then isolated from the gel and tested for reactivity with $NaIO_4$ to determine the presence of modification at their 3′ termini (Figures 4A and 4B). At each time, total siRNA was analyzed in parallel. 3′ terminal modification increased over the course of RISC assembly and, at all times, was restricted to single-stranded siRNA: Within our limits of detection, all double-stranded siRNA was unmodified, even after 3 hr. We conclude that siRNA modification is coupled to RISC assembly and occurs only after the conversion of pre-RISC to mature RISC.

Recombinant DmHen1 Modifies Single-Stranded Small RNA

Whereas *Arabidopsis* HEN1 contains an N-terminal double-stranded RNA-binding motif [15], DmHen1 does not. To test whether DmHen1 modifies double-stranded small RNAs, we incubated purified, recombinant GST-DmHen1 with either single-stranded or double-stranded siRNAs. We detected modification, evidenced by loss of reactivity with $NaIO_4$, only for the single-stranded RNA, suggesting that DmHen1 modifies single-stranded substrates, but not siRNAs or blunt RNA duplexes (Figure 4C). A preference for single-stranded RNA would explain how DmHen1 could act on both siRNAs, which are born double stranded, and piRNAs, which are not. We note that the purified, recombinant GST-DmHen1 protein was more than 50-fold less active on its own than when supplemented with ovary lysate from *hen1*^{f00810} homozygous flies. We speculate that the Ago2-RISC machinery is required for Hen1 function in

flies, although we cannot yet exclude the possibility that the lysate contains a factor (e.g., a kinase) required for activating Hen1.

Modification of single-stranded siRNAs—that is, those loaded in fully mature Ago2-RISC but not double-stranded siRNAs (Figure 4D)—might allow cells to distinguish siRNAs loaded successfully into functional complexes from those that fail to assemble. For example, if a 3′-to-5′ nuclease acts to degrade single-stranded siRNAs, 2′-*O*-methylation of single-stranded siRNAs in Ago2 RISC may protect them from destruction. Moreover, such a nuclease might trim the 3′ end of piRNAs. 2′-*O*-methylation of the piRNA 3′ terminus may occur only when the length of RNA extending beyond the Piwi-family protein is short enough to permit the simultaneous binding of the final ribose sugar to the active site of DmHen1 and the interaction of DmHen1 with the Piwi protein itself (Figure 4E). Modification of the terminus of the trimmed piRNA would then block further 3′-to-5′ trimming of the small RNA, generating its Piwi-, Aubergine-, or Ago3-specific length. Our observation that piRNAs are shorter in *hen1*^{f00810} mutants supports this model.

We note that all 2′-*O*-methyl-modified small RNAs identified thus far are associated with RISC complexes that efficiently cleave their RNA targets—i.e., Ago1-associated plant miRNAs [33, 34], animal piRNAs [35], and Ago2-associated siRNAs in flies [36]—whereas *Drosophila* miRNAs are typically both unmodified and associated with Ago1 RISC, which does not catalyze mRNA target cleavage in vivo [21]. We speculate that DmHen1 is recruited to RISC complexes containing single-stranded small silencing RNAs according to the identity of their Argonaute protein. This model predicts that DmHen1 will bind only to complexes containing fly Ago2 or the three fly Piwi proteins, Piwi, Aubergine, and Ago3, but not Ago1. Clearly, future experiments will need to test this hypothesis.

Experimental Procedures

General Methods

Preparation of 0 2 hr embryo, ovary, and S2 cell lysates, in vitro RISC assembly and RNAi reactions, and Northern hybridization were as described [8, 37, 38]. A list of sequences of synthetic RNA and DNA oligonucleotides is available in the Supplemental Experimental Procedures section in the Supplemental Data online.

^{32}P-radiolabeled 3′ Mononucleotide Standards

Synthetic RNA oligonucleotides (Supplemental Experimental Procedures) were radiolabeled in a 20 μl reaction containing 50 mM Tris HCl (pH 7.8), 10 mM $MgCl_2$, 1 mM ATP, 10 mM dithiothreitol, 10% (v/v) DMSO, 10 μg/ml BSA, 2 units/μl RNasin (Promega), 1.5 μCi/μl [5′ ^{32}P] cytidine 5′,3′ bis phosphate ([5′ ^{32}P] pCp; Perkin Elmer), and 1 unit/μl T4 RNA Ligase 1 (New England Biolabs) at 4°C, overnight. The radiolabeled small RNAs were purified from a 15% denaturing urea polyacrylamide sequencing gel and then digested with 1.5 U/μl micrococcal nuclease (Takara Mirus Bio) in a 40 μl reaction containing 20 mM Tris HCl (pH 8.0), 5 mM NaCl, and 2.5 mM $CaCl_2$. 3′ ^{32}P mononucleotides were further purified from a 22.5% denaturing urea polyacrylamide sequencing gel.

2D-TLC

Small RNAs (21 29 nucleotides, containing both modified piRNAs and unmodified small RNAs) from 0 2 hr wild type (Oregon R) fly embryos and small RNAs (26 31 nucleotdes, containing mostly modified piRNAs) from mouse and bull testes were purified from a 10% denaturing urea polyacrylamide gel stained with SYBR Gold (Invitrogen). Approximately 100 pmol purified small RNAs were radiolabeled as described above, except in a 40 μl reaction with 3 μCi/μl [5′ ^{32}P] pCp and 1 unit/μl T4 RNA Ligase 1, and then gel purified. The purified, ^{32}P radiolabeled RNA was hydrolyzed in 200 mM Na_2CO_3 at 100°C for 1 hr, neutralized with an equal volume of 200 mM HCl, then dephosphorylated with 0.5 units/μl calf intestinal alkaline phosphatase (New England Biolabs) in a 200 μl reaction containing 50 mM Tris HCl (pH 7.9), 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol. Alkaline phosphatase was inactivated by extraction with phenol/chloroform; RNA in the aqueous phase was oxidized with 80 mM $NaIO_4$ in borax/boric acid buffer (60 mM borax and 60 mM boric acid [pH 8.6]) at room temperature for 30 min and then β eliminated with 200 mM NaOH at 45°C for 90 min. A total of 5 μl of this reaction was mixed with an equal volume of formamide loading buffer (98% deionized formamide, 10 mM EDTA [pH 8.0], 0.025% (w/v) xylene cyanol, and 0.025% (w/v) bromophenol blue) and resolved on a 22.5% denaturing urea polyacrylamide sequencing gel. Equal intensities of modified mononucleotides and standards were spotted on 20 × 20 cm PEI cellulose F glass TLC plates (EMD Chemicals) and separated first with isobutyric acid/25%ammonia/water (66:1:33, v:v:v) and then 0.1 M sodium phosphate (pH 6.8)/ammonium sulfate/1 propanol (100:60:2, v:w:v).

Analysis of RNA 3′ Termini

RNA was incubated for 30 min at room temperature with 25 mM $NaIO_4$ in borax/boric acid buffer (60 mM borax and 60 mM boric acid [pH 8.6]), then 45.2 mM (f.c.) was NaOH added, and incubation continued for 90 min at 45°C (β elimination). The reaction was stopped by the addition of 300 mM (f.c.) NaCl, 1 μg glycogen, and three volumes absolute ethanol. After 30 min on ice, the precipitated RNA was collected by centrifugation.

Recombinant *Drosophila* Hen1 Protein

DmHen1 coding sequence was amplified from *Drosophila* ovary cDNA and inserted into pEnt D Topo (Invitrogen). The entire *hen1* sequence was confirmed by sequencing. The entry plasmid was recombined with the N terminal GST expression vector, pDest 15, with LR Clonase (Invitrogen). GST Hen1 was expressed in BL21 Star DE3 cells (Invitrogen) grown at 37°C in LB broth containing 100 ug/ml ampicillin until to OD_{600} reached 0.50. The culture was then cooled to 25°C, and 0.4 mM IPTG was added to induce protein production. The culture was incubated at 25°C with vigorous shaking for 3 hr. The cells were harvested by centrifugation at 7,300 × g for 20 min, washed with PBS, and centrifuged again; then the cell paste was frozen in liquid nitrogen and stored at 80°C.

Hen1 fusion protein was purified with the GST Purification Kit (Clontech). Cells were resuspended in 40 ml of extraction/loading buffer and lysed by sonication (duty 30% for 6 min; Branson Sonifier II), with cooling in an ice bath. A total of 2 ml of clarified lysate was added to the column resin, and the column inverted several times to disperse the resin. The resin was then allowed to pack, and the remaining 38 ml passed through the column by gravity flow. Subsequent steps were in accordance with the manufacturer's directions.

Analysis of Double- and Single-Stranded siRNA

Double and single stranded 5′ ^{32}P radiolabeled siRNA guide strands (10 nM; Figure S7 and Figure 4) were separated as described [32]. In brief, RISC assembly reactions were stopped with 2× Proteinase K buffer, 2 mg/ml Proteinase K, 1 μg glycogen, and 250 nM unlabeled siRNA guide strand to prevent reannealing. After incubation for 30 min at 25°C, 3 volumes absolute ethanol were added, and the RNA precipitated for 30 min on ice. The precipitates were collected by centrifugation, washed with 80% (v/v) ethanol, then dissolved in 2 mM Tris Cl (pH 7.5), 3% (w/v) Ficoll 400, 0.04% (w/v) bromophenol blue, 100 mM KCH_3CO_2, 30 mM HEPES KOH, and 2 mM $Mg(CH_3CO_2)_2$ and resolved by electrophoresis through a 15% native polyacrylamide gel (19:1 acrylamide:bis; 89 mM Tris Borate [pH 8.3], 2 mM Na EDTA, and 2.5 mM $Mg[CH_3CO_2]_2$). The region of the native gel corresponding to double or single stranded siRNA was excised, and the RNA was eluted overnight in 1 M NaCl. A total of 1 μg glycogen and ethanol (60% final volume) was added to the eluate, the RNA collected with MegaClear filter cartridges (Ambion), eluted with H_2O, and then precipitated for 30 min on ice by

an addition of 500 mM (f.c.) $NH_4CH_3CO_2$ and 2.5 volumes absolute ethanol. The precipitate was collected by centrifugation and washed with 80% (v/v) ethanol, and the samples reacted with $NaIO_4$ and subsequent β elimination (see above). The precipitated RNA was dissolved in 98% deionized formamide, 10 mM EDTA (pH 8.0), 0.025% (w/v) xylene cyanol, and 0.025% (w/v) bromophenol blue and then resolved on a 15% denaturing urea polyacrylamide sequencing gel.

Supplemental Data

Additional Experimental Procedures and seven figures are available at http://www.current biology.com/cgi/content/full/17/14/1265/DC1/.

Acknowledgments

We thank Alicia Boucher for assistance with fly husbandry, Marian Walhout for the GST destination vector, Steven Miller for GST pro tein, and members of the Zamore lab for advice, suggestions, and critical comments on the text. We are especially grateful to Alla Sigova for assistance preparing germline tissue and with qPCR and Tingting Du for help preparing ovary lysates. P.D.Z. is a W.M. Keck Foundation Young Scholar in Medical Research. This work was supported in part by grants from the National Institutes of Health to P.D.Z. (GM62862 and GM65236), a National Research Ser vice Award pre doctoral MD/PhD fellowship from National Institute on Aging (F30AG030283) to M.D.H., and a predoctoral fellowship from the American Heart Association to C.M.

Received: May 20, 2007
Revised: June 14, 2007
Accepted: June 15, 2007
Published online: June 28, 2007

References

1. Zamore, P.D., and Haley, B. (2005). Ribo gnome: The big world of small RNAs. Science *309*, 1519 1524.
2. Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double stranded RNA. Nature *431*, 343 349.
3. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double stranded RNA in Caenorhabditis elegans. Nature *391*, 806 811.
4. Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene ex pression: The potentially widespread influence of metazoan microRNAs. Nat. Rev. Genet. *5*, 396 400.
5. 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 develop ment. Dev. Cell *5*, 337 350.
6. 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.
7. Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M., and Gvozdev, V.A. (2001). Double stranded RNA medi ated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. Curr. Biol. *11*, 1017 1027.
8. 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.
9. Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachi danandam, R., and Hannon, G.J. (2007). Discrete small RNA generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089 1103.
10. Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recog nizes 21 24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. Nucleic Acids Res. *34*, 667 675.
11. Kirino, Y., and Mourelatos, Z. (2007). Mouse Piwi interacting RNAs are 2′ O methylated at their 3′ termini. Nat. Struct. Mol. Biol. *14*, 347 348.
12. Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H., Miyauchi, K., and Suzuki, T. (2007). The 3′ termini of mouse Piwi interacting RNAs are 2′ O methylated. Nat. Struct. Mol. Biol. *14*, 349 350.
13. 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., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69 82.
14. Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′ end uridylation activity in Arabidopsis. Curr. Biol. *15*, 1501 1507.
15. Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932 935.
16. Pelisson, A., Sarot, E., Payen Groschene, G., and Bucheton, A. (2007). A novel repeat associated small interfering RNA medi ated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. J. Virol. *81*, 1951 1960.
17. Tkaczuk, K.L., Obarska, A., and Bujnicki, J.M. (2006). Molecular phylogenetics and comparative modeling of HEN1, a methyl transferase involved in plant microRNA biogenesis. BMC Evol. Biol. *6*, 6.
18. Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CAR PEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in *Arabidopsis thaliana*. Curr. Biol. *12*, 1484 1495.
19. Vagin, V.V., Klenov, M.S., Kalmykova, A.I., Stolyarenko, A.D., Kotelnikov, R.N., and Gvozdev, V.A. (2004). The RNA interfer ence proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila mela nogaster*. RNA Biol. *1*, 54 58.
20. Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvoz dev, V. (2006). Telomere elongation is under the control of the RNAi based mechanism in the Drosophila germline. Genes Dev. *20*, 345 354.
21. Förstemann, K., Horwich, M.D., Wee, L. M., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct Argonaute protein complexes after their production by Dicer 1. Cell, in press.
22. Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Dis tinct roles for Argonaute proteins in small RNA directed RNA cleavage pathways. Genes Dev. *18*, 1655 1666.
23. Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Balint, É., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA interference enzyme Dicer in the maturation of the let 7 small temporal RNA. Science *293*, 834 838.
24. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer 1 and Dicer 2 in the siRNA/miRNA silencing pathways. Cell *117*, 69 81.
25. Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effec tor steps of the Drosophila RNAi pathway. Science *301*, 1921 1925.
26. Kim, K., Lee, Y.S., and Carthew, R.W. (2006). Conversion of pre RISC to holo RISC by Ago2 during assembly of RNAi com plexes. RNA *13*, 22 29.
27. Leuschner, P.J., Ameres, S.L., Kueng, S., and Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO Rep. *7*, 314 320.
28. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger strand cleavage facilitates assembly of siRNA into Ago2 containing RNAi enzyme complexes. Cell *123*, 607 620.
29. Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti guide strand of siRNA during RISC activation. Cell *123*, 621 629.
30. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involve ment in RISC formation. Genes Dev. *19*, 2837 2848.
31. Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199 208.

32. Nykanen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309–321.
33. Baumberger, N., and Baulcombe, D.C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA *102*, 11928–11933.
34. Qi, Y., Denli, A.M., and Hannon, G.J. (2005). Biochemical specialization within Arabidopsis RNA silencing pathways. Mol. Cell *19*, 421–428.
35. 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.
36. Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146–1150.
37. Haley, B., Tang, G., and Zamore, P.D. (2003). In vitro analysis of RNA interference in Drosophila melanogaster. Methods *30*, 330–336.
38. Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ line stem cell maintenance requires Loquacious, a double stranded RNA binding domain protein. PLoS Biol. *3*, e236.

Note Added in Proof

This version differs slightly from the one published previously online in that on page 1 of the Supplemental Data, the dots have now been properly aligned beneath the sequences, whereas previously they were not.

LMA and T_b were placed into a food restriction paradigm under DD conditions, we found that they maintained the rhythm that had been en trained by light (high amplitude, free running pe riod of ~23.7 hours) and never showed an increase in LMA or T_b in anticipation of the food presen tation (fig. S3). Hence, although a BMAL1 based clock is necessary to support food entrainment, restoration of clock function in the SCN alone is not able to rescue this behavior.

To test the hypothesis that the BMAL1 based clock induced in the DMH during restricted feeding might drive circadian entrainment, we performed stereotaxic bilateral delivery of AAV BMAL1 (the same construct and vector as used in the SCN) into the DMH of *Bmal1*$^{-/-}$ mice. Mice who sustained bilateral DMH injections of the AAV BMAL1 did not demonstrate entrainment to a 12:12 LD cycle or free running rhythms of T_b or LMA in DD (Fig. 3A). By contrast, under conditions of food restriction in DD, they ex hibited a clear anticipatory increase in T_b and LMA before food presentation (Fig. 2C and Fig. 3B). Each individual mouse showed very little day to day variation in the timing of the increase in T_b and LMA under DD (i.e., the phase angle of entrainment was stable). Finally, the increase in T_b and LMA before the predicted period of food presentation persisted during a 24 hour fast at the end of restricted feeding (arrow in Fig. 3B), dem onstrating the circadian nature of the response.

In both our study and the study by Mieda *et al.* (*8*), clock gene expression in the DMH was largely restricted to cells in the compact part of the nu cleus, which consists of small, closely packed neurons that are highly reminiscent of the SCN itself. These neurons appear mainly to have local connections with the adjacent output zones of the DMH (*23*), suggesting that the timing signal from the compact part of the DMH may impinge upon the same output neurons in the remainder of the DMH as are used to control light entrained rhythms directed by the SCN. This relationship may explain how the DMH clock is able to override the SCN clock input during conditions of food entrainment in an intact animal. It is un likely that feedback from the DMH alters activity in the SCN in any major way, because the SCN remains phase locked to the LD cycle for many weeks during food entrainment (as long as the animals are not also hypocaloric). These obser vations also raise the interesting possibility that the DMH may form the neuroanatomic basis of the so called methamphetamine sensitive circa dian oscillator (MASCO), which also operates independent of the SCN and does not entrain to light [for a review, see (*24*)].

Our data indicate that there is an inducible clock in the DMH that can override the SCN and drive circadian rhythms when the animal is faced with limited food availability. Thus, under re stricted feeding conditions, the DMH clock can assume an executive role in the temporal regula tion of behavioral state. For a small mammal, finding food on a daily basis is a critical mission. Even a few days of starvation, a common threat in natural environments, may result in death. Hence, it is adaptive for animals to have a secondary "master clock" that can allow the animal to switch its behavioral patterns rapidly after a period of starvation to maximize the opportunity of finding food sources at the same time on following days.

In an intact animal, peripheral oscillators in many tissues in the body, including the stomach and liver, as well as elsewhere in the brain, may contribute to food entrainment of circadian rhythms (*25*, *26*). Consequently, it has been dif ficult to dissect this system by using lesions of individual components of the pathway (*3*, *9*, *10*). However, by starting with a genetically arrhyth mic mouse and using focal genetic rescue in the brain, we have identified the SCN molecular clock as sufficient for light but not food entrain ment of T_b and LMA rhythms in mice, and the DMH as sufficient for food but not light entrain ment of circadian rhythms of T_b and LMA. These results demonstrate the power of viral based gene replacement in the central nervous system to dissect complex neural functions.

References and Notes

1. M. H. Hastings, A. B. Reddy, E. S. Maywood, *Nat. Rev. Neurosci.* **4**, 649 (2003).
2. C. B. Saper, J. Lu, T. C. Chou, J. Gooley, *Trends Neurosci.* **28**, 152 (2005).
3. J. J. Gooley, A. Schomer, C. B. Saper, *Nat. Neurosci.* **9**, 398 (2006).
4. F. K. Stephan, J. M. Swann, C. L. Sisk, *Behav. Neural Biol.* **25**, 346 (1979).
5. P. L. Lowrey, J. S. Takahashi, *Annu. Rev. Genomics Hum. Genet.* **5**, 407 (2004).
6. A. Kramer *et al.*, *Science* **294**, 2511 (2001).
7. F. K. Stephan, *Physiol. Behav.* **46**, 489 (1989).
8. M. Mieda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12150 (2006).
9. G. J. Landry, M. M. Simons, I. C. Webb, R. E. Mistlberger, *Am. J. Physiol.* **290**, 6 (2006).
10. J. J. Gooley, C. B. Saper, *J. Biol. Rhythms* **22**, 479 (2008).
11. N. Gekakis *et al.*, *Science* **280**, 1564 (1998).
12. B. Zheng *et al.*, *Cell* **105**, 683 (2001).
13. M. K. Bunger *et al.*, *Cell* **103**, 1009 (2000).
14. E. L. McDearmon *et al.*, *Science* **314**, 1304 (2006).
15. M. K. Bunger *et al.*, *Genesis* **41**, 122 (2005).
16. A. Laposky *et al.*, *Sleep* **28**, 395 (2005).
17. S. Shimba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12071 (2005).
18. R. D. Rudic *et al.*, *PLoS Biol.* **2**, e377 (2004).
19. Materials and methods are available as supporting material on *Science* Online.
20. T. C. Chou *et al.*, *J. Neuroscience* **23**, 1069 (2003).
21. C. A. Fuller *et al.*, *Am. J. Physiol.* **241**, 5 (1981).
22. D. T. Krieger, H. Hauser, L. C. Krey, *Science* **197**, 398 (1977).
23. R. H. Thompson, N. S. Canteras, L. W. Swanson, *Brain Res. Brain Res. Rev.* **27**, 89 (1998).
24. T. Hiroshige, K. Honma, S. Honma, *Brain Res. Bull.* **27**, 441 (1991).
25. K. A. Stokkan *et al.*, *Science* **291**, 490 (2001).
26. F. Damiola *et al.*, *Genes Dev.* **14**, 2950 (2000).

27. We thank Q. Ha and M. Ha for technical work. Support was provided by U.S. Public Health Service grants HL60292, NS33987, NS051609, NS057119, and HL07901 08.

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5879/1074/DC1
Materials and Methods
Figs. S1 to S4
References

26 November 2007; accepted 14 April 2008
10.1126/science.1153277

Endogenous siRNAs Derived from Transposons and mRNAs in *Drosophila* Somatic Cells

Megha Ghildiyal,[1]* Hervé Seitz,[1]* Michael D. Horwich,[1] Chengjian Li,[1] Tingting Du,[1] Soohyun Lee,[2] Jia Xu,[3] Ellen L.W. Kittler,[4] Maria L. Zapp,[4] Zhiping Weng,[5] Phillip D. Zamore[1]†

Small interfering RNAs (siRNAs) direct RNA interference (RNAi) in eukaryotes. In flies, somatic cells produce siRNAs from exogenous double-stranded RNA (dsRNA) as a defense against viral infection. We identified endogenous siRNAs (endo-siRNAs), 21 nucleotides in length, that correspond to transposons and heterochromatic sequences in the somatic cells of *Drosophila melanogaster*. We also detected endo-siRNAs complementary to messenger RNAs (mRNAs); these siRNAs disproportionately mapped to the complementary regions of overlapping mRNAs predicted to form double-stranded RNA in vivo. Normal accumulation of somatic endo-siRNAs requires the siRNA-generating ribonuclease Dicer-2 and the RNAi effector protein Argonaute2 (Ago2). We propose that endo-siRNAs generated by the fly RNAi pathway silence selfish genetic elements in the soma, much as Piwi-interacting RNAs do in the germ line.

Three RNA silencing pathways have been identified in flies and mammals: RNA interference (RNAi), guided by small in terfering RNAs (siRNAs) derived from exogenous double stranded RNA (dsRNA); the microRNA (miRNA) pathway, in which endogenous small RNAs repress partially complementary mRNAs; and the Piwi interacting RNA (piRNA) pathway, whose small RNAs repress transposons in the germ line (*1*–*3*) and can activate transcription in hetero chromatin (*4*).

Endogenous siRNAs (endo siRNAs) silence retrotransposons in plants (*5*, *6*), and siRNAs corresponding to the L1 retrotransposon have

[1]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA. [2]Program in Bioinformatics, Boston University, Boston, MA 02215, USA. [3]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA. [4]Program in Molecular Medicine and Center for AIDS Research, University of Massachusetts Medical School, Worcester, MA 01605, USA. [5]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E mail: phillip.zamore@umassmed.edu

been detected in cultured mammalian cells (*7*). Genetic and molecular evidence suggests that in addition to suppressing viral infection, the RNAi pathway silences selfish genetic elements in the fly soma: Mutations in the RNAi gene *rm62* (*8*) suppress mutations caused by retroelement insertion (*9*); depletion of the Argonaute proteins Ago1 or Ago2 increases transposon expression in cultured *Drosophila* Schneider 2 (S2) cells (*10*); small RNAs have been detected in *Drosophila* Kc cells for the *1360* transposon (*11*) and are produced during transgene silencing in flies (*12*); and siRNAs have been proposed to repress germline expression of *suffix*, a short interspersed nuclear element (SINE) (*13*).

The defining properties of *Drosophila* siRNAs are their production from long dsRNA by Dicer 2 (Dcr 2), which generates 5′ monophosphate termini; their loading into Argonaute2 (Ago2); and their Ago2 dependent, 3′ terminal, 2′ *O* methylation by the methyltransferase Hen1 (*14 16*), unlike most miRNAs (*17*). In vivo (Fig. 1A, rightmost

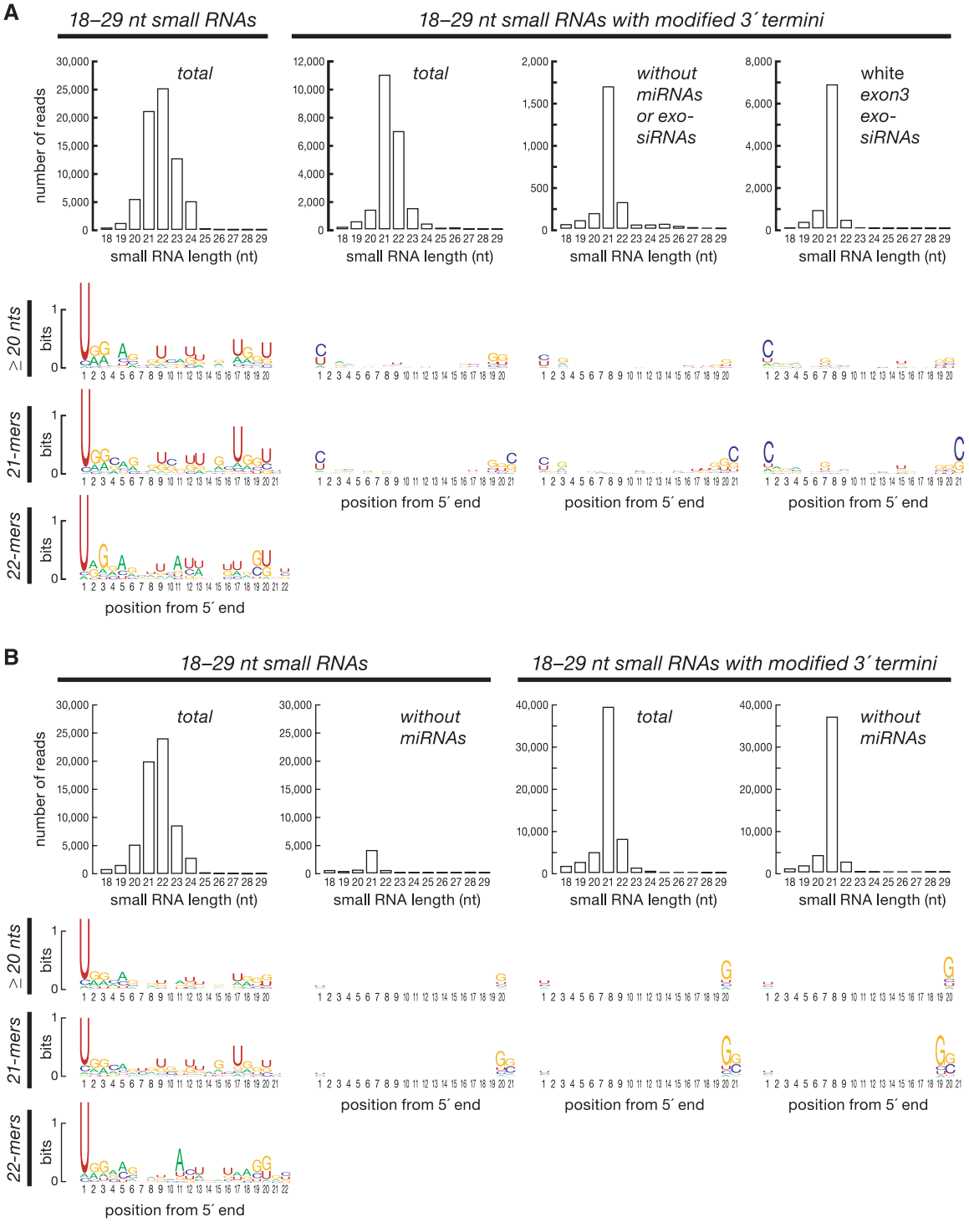


Fig. 1. High-throughput pyrosequencing revealed 3′-terminally modified 21-nt RNAs in the fly soma. (**A**) Length and sequence composition of the small RNA sequences from a library of total small RNA from the heads of flies expressing an inverted repeat (IR) silencing the *white* gene and for a parallel library enriched for RNAs modified at their 3′ ends. (**B**) Similar analysis for small RNA sequences from *Drosophila* S2 cells. For data labeled "without miRNAs," pre-miRNA–matching sequences were removed computationally.

panel) and in vitro (*18*), nearly all siRNAs produced by Dcr 2 from exogenous dsRNA are 21 nucleotides (nt) in length.

We characterized the somatic small RNA content of S2 cells (*19*) and of heads expressing an RNA hairpin silencing the *white* gene by RNAi (*20*). To identify endo siRNA candidates, we analyzed two types of RNA libraries. For total 18 to 29 nt RNA libraries, 89% (S2 cells) and 96% (heads) mapped to annotated miRNA loci. In contrast, libraries enriched for small RNAs bearing a 3′ terminal, 2′ *O* methyl modification (*21*) were depleted of miRNAs: Only 19% (S2 cells) and 49% (heads) of reads and 2.4% (S2 cells; 58,681 reads; 12,036 sequences) and 12% (heads; 22,685 reads; 2929 sequences) of unique sequences mapped to miRNA loci.

Table 1. Endo-siRNAs preferentially map to overlapping, complementary mRNAs.

Sample	Enrichment	Enrichment after randomization		*Z* score	*P*
		Mean	SD		
Fly heads	10.9	1.0	0.38	26.1	7.9×10^{-151}
S2 cells	12.3	1.1	0.42	27.0	5.2×10^{-161}

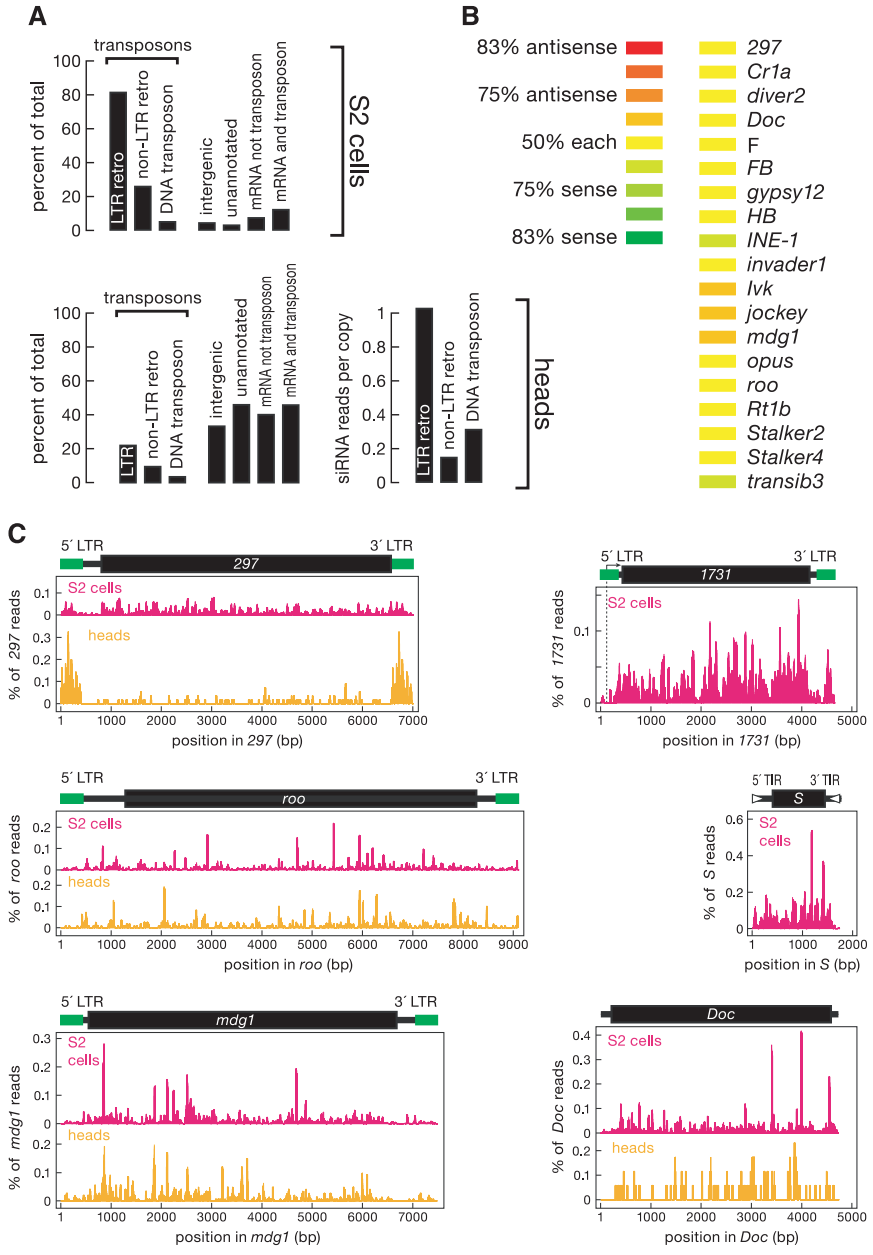


Fig. 2. Endo-siRNAs correspond to transposons. (**A**) Distribution of annotations for the genomic matches of endo-siRNA sequences. Bars total more than 100% because some siRNAs match both LTR and non-LTR retrotransposons or match both mRNA and transposons. (**B**) Transposon-derived siRNAs with more than 50 21-nt reads mapped about equally to sense and antisense orientations. (**C**) Alignment of endo-siRNA sequences to *Drosophila* transposons. The abundance of each sequence is shown as a percentage of all transposon-matching siRNA sequences. LTR, long terminal repeat; TIR, terminal inverted repeat. Here and in subsequent figures, data from high-throughput pyrosequencing and sequencing-by-synthesis were pooled for wild-type heads.

Figure 1 shows the length distribution and sequence composition of the four libraries. The total RNA samples were predominantly miRNAs, a bias reflected in their modal length (22 nt) and pronounced tendency to begin with uracil. Exclusion of miRNAs revealed a class of small RNAs with a narrow length distribution and no tendency to begin with uracil. Except for an unusual cluster of X chromosome small RNAs (fig. S1) and a miRNA like sequence with an unusual putative precursor on chromosome 2 (fig. S2), few of these small RNAs are likely to correspond to novel miRNAs: None lie in the arms of hairpins predicted to be as thermo dynamically stable as most pre miRNAs (i.e., < 15 kcal/mol).

After excluding known miRNAs, 64% (heads) (Fig. 1A) and 78% (S2 cells) (Fig. 1B) of se quences in the libraries enriched for 3′ terminally modified small RNAs that is, those likely to be Ago2 associated were 21 nt long. For fly heads, 37% (8404 reads) derived from the *white* dsRNA hairpin. The abundance of these exo siRNAs can be estimated by comparing them to the number of reads for individual miRNAs in the total small RNA library, where 1.6% (660 antisense and 491 sense reads) were 21 nt oligo mers (21 mers) and matched the *white* sequences in the dsRNA expressing transgene. The collec tive abundance of all *white* exo siRNAs was less than the individual abundance of the 10 most abundant miRNAs in this sample; the median abundance of any one exo siRNA species was two reads. The *white* inverted repeat (IR) trans gene phenocopies a nearly null mutation in *white*, yet the sequence of the most abundant exo siRNA was read just 37 times.

In heads, the sequence composition of the 21 nt, 3′ terminally modified small RNAs close ly resembled that of exo siRNAs, which tended to begin and end with cytosine. In heads and S2 cells, the 21 mers lacked the sequence features of piRNAs, which either begin with uracil (Aub and Piwi bound) or contain an adenine at po sition 10 (Ago3 bound) and are 23 to 29 nt long (*1*, *2*). These data suggest that the 21 nt small RNAs are somatic endo siRNAs.

In S2 cells, endo siRNAs mapped largely to transposons (86%); in fly heads, they mapped about equally to transposons, intergenic and unannotated sequences, and mRNAs. The finding that 41% of endo siRNAs mapped to mRNAs without mapping to transposons sug gests that endo siRNAs may regulate mRNA

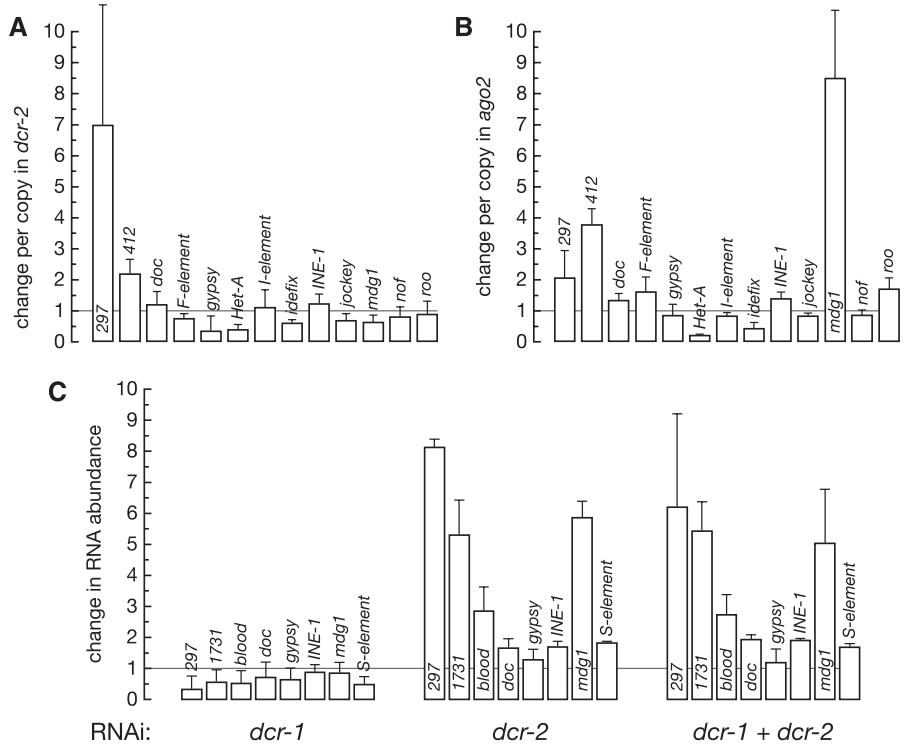


Fig. 3. Transposon silencing requires Dcr-2 and Ago2, but not Dcr-1. (**A** and **B**) The change in mRNA expression (mean ± SD, N = 3) for each transposon between *dcr-2*^{L811fsX} (A) or *ago2*^{414} (B) heterozygous and homozygous heads was measured by quantitative reverse transcription polymerase chain reaction. The data were corrected for differences in transposon copy number between the paired genotypes. (**C**) The change in transposon expression (mean ± SD, N = 3) in S2 cells was measured for the indicated RNAi depletion relative to a control dsRNA.

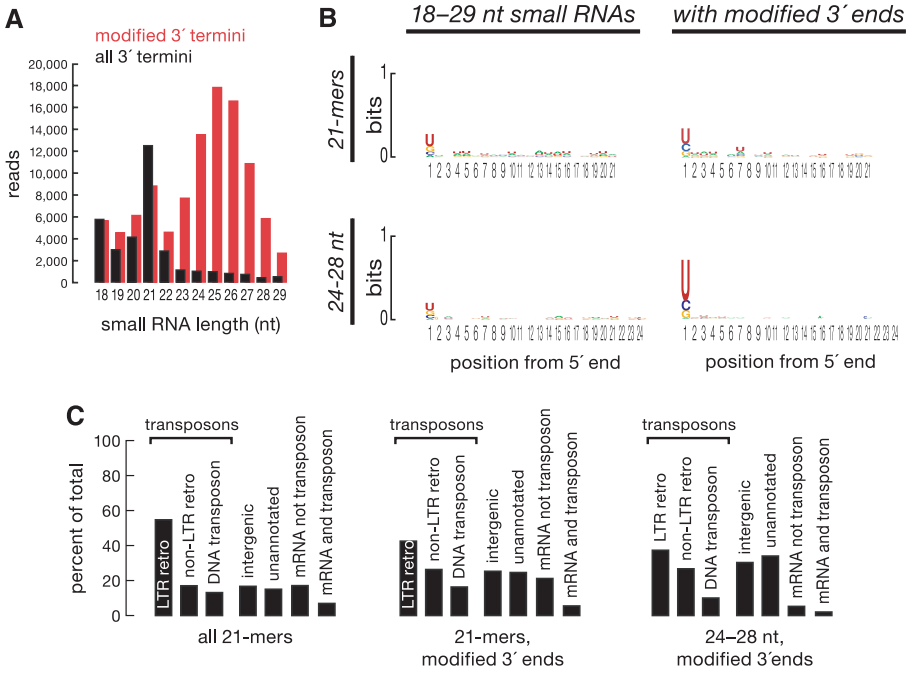


Fig. 4. The composition of somatic small RNAs is altered in the absence of Ago2. (**A** and **B**) Size distribution (A) and sequence composition (B) of sequences from a library of total 18- to 29-nt RNA from the heads of *ago2* null mutant flies or a library enriched for 3′-terminally modified RNAs. Reads matching pre-miRNA sequences were removed. (**C**) Distribution of annotations for the genomic matches of small RNA sequences from the two *ago2* libraries.

expression. Endo siRNAs mapping to mRNAs were likelier by a factor of >10 than expected by chance ($5.22 \times 10^{-161} < P < 8 \times 10^{-151}$) to derive from genomic regions annotated to produce over lapping, complementary transcripts (Table 1 and table S1). These data suggest that such over lapping, complementary transcripts anneal in vivo to form dsRNA that is diced into endo siRNAs. We note that among the mRNAs for which we detected complementary 21 mers was *ago2* itself.

Endo siRNAs mapped to all three large chro mosomes (figs. S3 to S5). siRNAs corresponding to the three transposon types in *Drosophila* were detected, but long terminal repeat (LTR) retro transposons, the dominant class of selfish genetic elements in flies, were overrepresented even after accounting for their abundance in the genome (Fig. 2A and table S2). Unlike piRNAs, which are disproportionately antisense to transpos ons, but like siRNAs derived from exogenous dsRNA, about equal numbers of sense and anti sense transposon matching endo siRNAs were detected (Fig. 2B and fig. S6) (*1 3*, *22*). Like piRNAs, endo siRNAs map to large genomic clusters (table S3). Of 172 endo siRNA clusters in S2 cells, four coincided with previously iden tified piRNA clusters (cluster 1, at 42A of chro mosome 2R; clusters 7 and 10 in unassembled genomic sequence; and cluster 15 in the chro mosome 3L heterochromatin). In heads, we de tected 17 clusters; five corresponded to clusters found in S2 cells, but only one was shared with the germline piRNAs: the *flamenco* locus, con sistent with recent genetic evidence that a Piwi independent but *flamenco* dependent pathway represses the *Idefix* and *ZAM* transposons in the soma (*23*). That both endo siRNAs and piRNAs can arise from the same region suggests either that a single transcript can be a substrate for both piRNA and siRNA production or that dis tinct classes of transcripts arise from a single locus. The abundance and distribution of endo siRNAs across the sequences of individual transposon species reflected the natural history of when the elements entered the fly genome, but not their mechanism of transposition (Fig. 2C) (*24*).

Statistically significant reductions in siRNA abundance were observed in *dcr 2*^{L811fsX} null mutant heads relative to heads from heterozygous siblings for 38 transposons (fig. S7 and table S4). Normalized for sequencing depth, sequencing results from homozygous *dcr 2* mutant heads yielded fewer 21 mers overall (by a factor of 3.1) and fewer 21 mers corresponding to transposons (by a factor of 6.3) than did their heterozygous siblings ($P < 2.2 \times 10^{-16}$; χ^2 test). In contrast, overall miRNA abundance normalized to se quencing depth was essentially unchanged be tween *dcr 2* heterozygotes and homozygotes (fig. S7 and table S5). These data suggest that endo siRNAs are produced by Dcr 2, but we do not yet know why some endo siRNAs persist in *dcr 2*^{L811fsX} mutants.

Transposon expression in the soma reflects both the silencing of transposons potentially by either or both posttranscriptional and transcrip tional mechanisms and the tissue specificity of transposon promoters. *Drosophila* somatic cells may contain siRNAs targeting transposons that would not be highly expressed even in the absence of those siRNAs, because the promoters of those transposons are not active in some or all somatic tissues or because they are repressed by additional mechanisms. We analyzed the expres sion of a panel of transposons in heads from *ago2* and *dcr 2* mutants and in S2 cells depleted of Dcr 1, Dcr 2, or Ago2 by RNAi (Fig. 3 and fig. S8). We found that the steady state abundance of RNA from the LTR retrotransposons *297* and *412* increased in heads from *dcr* $2^{L811fsX}$ null mu tants (Fig. 3A). Similarly, the steady state abun dance of RNA from the LTR retrotransposons *297*, *412*, *mdg1*, and *roo*, the non LTR retrotransposon *F element*, and the SINE like element *INE 1* increased in *ago2*^{414} mutant heads (Fig. 3B).

In S2 cells, RNA expression from the LTR retrotransposons *297*, *1731*, *mdg1*, *blood*, and *gypsy* and from the DNA transposon *S element* all increased significantly ($0.00001 < P < 0.002$) when Dcr 2 was depleted or when both Dcr 2 and Dcr 1 were depleted, but not when Dcr 1 alone was depleted (Fig. 3C). Similarly, *ago2(RNAi)* in S2 cells desilenced transposons, including nine LTR and non LTR retrotransposons and the DNA transposon *S element* (fig. S8).

Is Ago2 required for the production or accumulation of endo siRNAs? We sequenced 18 to 29 nt small RNAs from *ago2*^{414} homozy gous fly heads and from the same small RNA sample treated to enrich for 3′ terminally modified RNAs. After computationally removing miRNAs, the sequences from the untreated library con tained a prominent 21 nt peak (Fig. 4A) that pre dominantly began with uracil (Fig. 4B), much like miRNAs and unlike siRNAs in wild type heads, which often began with cytosine (Fig. 1A). Perhaps in the absence of Ago2, only a sub population of endo siRNAs that can bind Ago1 accumulates. The small RNAs from the *ago2*^{414} library enriched for 3′ terminally modified se quences were predominantly 24 to 27 nt long and often began with uracil a length distribution and sequence bias characteristic of piRNAs, which, like siRNAs, are 2′ *O* methylated at their 3′ ends. Both the 21 nt small RNAs and the piRNA like RNAs in the *ago2* mutant heads mapped to transposons, unannotated hetero chromatic and unassembled sequences, but the piRNA like sequences mapped to mRNAs far less frequently than did either the 21 mers or wild type endo siRNAs (Fig. 4C). How these piRNA like small RNAs are generated and whether they contribute to transposon silencing in the fly soma remain unknown.

Note added in proof: The loci described here in figs. S1 and S2 correspond to endo siRNA generating hairpins recently identified in (*25 27*).

References and Notes

1. L. S. Gunawardane *et al*., *Science* **315**, 1587 (2007); published online 21 February 2007 (10.1126/science. 1140494).
2. J. Brennecke *et al*., *Cell* **128**, 1089 (2007).
3. V. V. Vagin *et al*., *Science* **313**, 320 (2006); published online 28 June 2006 (10.1126/science.1129333).
4. H. Yin, H. Lin, *Nature* **450**, 304 (2007).
5. A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe, *EMBO J.* **21**, 4671 (2002).
6. R. Sunkar, T. Girke, J. K. Zhu, *Nucleic Acids Res.* **33**, 4443 (2005).
7. N. Yang, H. H. J. Kazazian, *Nat. Struct. Mol. Biol.* **13**, 763 (2006).
8. A. Ishizuka, M. C. Siomi, H. Siomi, *Genes Dev.* **16**, 2497 (2002).
9. A. K. Csink, R. Linsk, J. A. Birchler, *Genetics* **138**, 153 (1994).
10. J. Rehwinkel *et al*., *Mol. Cell. Biol.* **26**, 2965 (2006).
11. K. A. Haynes, A. A. Caudy, L. Collins, S. C. Elgin, *Curr. Biol.* **16**, 2222 (2006).
12. M. Pal Bhadra, U. Bhadra, J. A. Birchler, *Mol. Cell* **9**, 315 (2002).
13. N. A. Tchurikov, O. V. Kretova, *PLoS ONE* **2**, e476 (2007).
14. M. D. Horwich *et al*., *Curr. Biol.* **17**, 1265 (2007).
15. A. Pelisson, E. Sarot, G. Payen Groschene, A. Bucheton, *J. Virol.* **81**, 1951 (2007).
16. K. Saito *et al*., *Genes Dev.* **21**, 1603 (2007).
17. K. Okamura, A. Ishizuka, H. Siomi, M. C. Siomi, *Genes Dev.* **18**, 1655 (2004).
18. A. Nykanen, B. Haley, P. D. Zamore, *Cell* **107**, 309 (2001).
19. *Drosophila* RNAi Screening Center at Harvard Medical School (http://flyrnai.org/cgi bin/RNAi FAQ lines.pl).
20. Y. S. Lee, R. W. Carthew, *Methods* **30**, 322 (2003).
21. H. Seitz, M. Ghildiyal, P. D. Zamore, *Curr. Biol.* **18**, 147 (2008).
22. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).
23. S. Desset, N. Buchon, C. Meignin, M. Coiffet, C. Vaury, *PLoS ONE* **3**, e1526 (2008).
24. See supporting material on *Science* Online.
25. B. Czech *et al*., *Nature* 10.1038/nature07007 (2008).
26. Y. Kawamura *et al*., *Nature* 10.1038/nature06938 (2008).
27. K. Okamura *et al*., *Nature* 10.1038/nature07015 (2008).

28. We thank A. Boucher and S. Ma for technical assistance; G. Farley for encouragement, support, and technical assistance; and Roche Applied Science for high throughput sequencing. P.D.Z. is a W. M. Keck Foundation Young Scholar in Medical Research. Supported by NIH grants GM62862 and GM65236 (P.D.Z.), GM080625 (J.X. and Z.W.), and HG003367 (S.L.); EMBO long term (ALTF 910 2004) and Human Frontier Science Program (LT00575/2005 L) fellowships (H.S.); and a National Research Service Award predoctoral MD/PhD fellowship from the National Institute on Aging (F30AG030283) (M.D.H.). NCBI Gene Expression Omnibus accession numbers for sequence and abundance data are GSE9389 and GSE11019, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1157396/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S7
References

5 March 2008; accepted 31 March 2008
Published online 10 April 2008;
10.1126/science.1157396
Include this information when citing this paper.

Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton

Dana E. Hunt,[1]* Lawrence A. David,[2]* Dirk Gevers,[1,3,4] Sarah P. Preheim,[1] Eric J. Alm,[1,5]† Martin F. Polz[1]†

Identifying ecologically differentiated populations within complex microbial communities remains challenging, yet is critical for interpreting the evolution and ecology of microbes in the wild. Here we describe spatial and temporal resource partitioning among *Vibrionaceae* strains coexisting in coastal bacterioplankton. A quantitative model (AdaptML) establishes the evolutionary history of ecological differentiation, thus revealing populations specific for seasons and life-styles (combinations of free-living, particle, or zooplankton associations). These ecological population boundaries frequently occur at deep phylogenetic levels (consistent with named species); however, recent and perhaps ongoing adaptive radiation is evident in *Vibrio splendidus*, which comprises numerous ecologically distinct populations at different levels of phylogenetic differentiation. Thus, environmental specialization may be an important correlate or even trigger of speciation among sympatric microbes.

Microbes dominate biomass and control biogeochemical cycling in the ocean, but we know little about the mecha nisms and dynamics of their functional differen tiation in the environment. Culture independent analysis typically reveals vast microbial diversity, and although some taxa and gene families are differentially distributed among environments (*1*, *2*), it is not clear to what extent coexisting genotypic diversity can be divided into function ally cohesive populations (*1*, *3*). First, we lack broad surveys of nonpathogenic free living bacte ria that establish robust associations of individual strains with spatiotemporal conditions (*4*, *5*); sec ond, it remains controversial what level of genetic diversification reflects ecological differentiation. Phylogenetic clusters have been proposed to cor respond to ecological populations that arise by neutral diversification after niche specific selec tive sweeps (*6*). Clusters are indeed observed among closely related isolates (e.g., when ex amined by multilocus sequence analysis) (*7*) and in culture independent analyses of coastal bacterio plankton (*8*). Yet recent theoretical studies suggest

Cell

Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies

Chengjian Li,[1,9] Vasily V. Vagin,[1,9,10] Soohyun Lee,[2,5,9] Jia Xu,[2,6,9] Shengmei Ma,[1] Hualin Xi,[5,7] Hervé Seitz,[1] Michael D. Horwich,[1] Monika Syrzycka,[8] Barry M. Honda,[8] Ellen L.W. Kittler,[3] Maria L. Zapp,[3] Carla Klattenhoff,[4] Nadine Schulz,[4] William E. Theurkauf,[4] Zhiping Weng,[2,*] and Phillip D. Zamore[1,*]

[1]Department of Biochemistry and Molecular Pharmacology and Howard Hughes Medical Institute
[2]Program in Bioinformatics and Integrative Biology
[3]Program in Molecular Medicine and Center for AIDS Research (CFAR)
[4]Program in Molecular Medicine
University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA
[5]Bioinformatics Program
[6]Biomedical Engineering
Boston University, Boston, MA 02215, USA
[7]Computational Sciences Center of Emphasis, Pfizer, Inc., 620 Memorial Drive, Cambridge, MA 02139, USA
[8]Molecular Biology and Biochemistry, Simon Fraser University, Burnaby BC V5A 1S6, Canada
[9]These authors contributed equally to this work
[10]Present address: Cold Spring Harbor Laboratories, Cold Spring Harbor, New York 11724, USA
*Correspondence: zhiping.weng@umassmed.edu (Z.W.), phillip.zamore@umassmed.edu (P.D.Z.)
DOI 10.1016/j.cell.2009.04.027

SUMMARY

Piwi-interacting RNAs (piRNAs) silence transposons in animal germ cells. piRNAs are thought to derive from long transcripts spanning transposon-rich genomic loci and to direct an autoamplification loop in which an antisense piRNA, bound to Aubergine or Piwi protein, triggers production of a sense piRNA bound to the PIWI protein Argonaute3 (Ago3). In turn, the new piRNA is envisioned to produce a second antisense piRNA. Here, we describe strong loss-of-function mutations in *ago3*, allowing a direct genetic test of this model. We find that Ago3 acts to amplify piRNA pools and to enforce on them an antisense bias, increasing the number of piRNAs that can act to silence transposons. We also detect a second, Ago3-independent piRNA pathway centered on Piwi. Transposons targeted by this second pathway often reside in the *flamenco* locus, which is expressed in somatic ovarian follicle cells, suggesting a role for piRNAs beyond the germline.

INTRODUCTION

The ability to tame transposons while retaining them in the genome is a particular specialty of eukaryotes. Transposons, repetitive sequences, and other forms of "selfish" DNA comprise as much as 42% of the human genome and nearly 30% of the genome of *Drosophila melanogaster*. In metazoa, transposons are silenced by the piRNA pathway, which is guided by 23–30 nt RNAs (Vagin et al., 2006; Brennecke et al., 2007).

The piRNA pathway is distinct from other RNA silencing pathways in that its small RNA guides are not produced by dicing. In contrast, both small interfering RNAs (siRNAs) and microRNAs (miRNAs) are cleaved by double-stranded RNA-specific endonucleases, Dicers, to yield double-stranded intermediates—siRNA or miRNA/miRNA* duplexes—that are loaded into members of the Argonaute family of proteins (reviewed in Ghildiyal and Zamore, 2009). piRNAs, too, act as guides for Argonaute proteins, but they appear not to exist as stable double-stranded intermediates at any point in their biogenesis (Vagin et al., 2006; Houwing et al., 2007). piRNAs bind PIWI proteins, a subfamily of Argonaute proteins that are expressed in germline cells (Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006; Vagin et al., 2006; Brennecke et al., 2007; Gunawardane et al., 2007; Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). PIWI proteins were first identified by their roles in maintaining (Cox et al., 1998, 2000) and patterning (Wilson et al., 1996; Harris and Macdonald, 2001) *Drosophila* germ cells. The defects in the organization of embryonic pattern in piRNA pathway mutations are likely an indirect consequence of their larger role in maintaining genomic stability (Klattenhoff et al., 2007). The three *Drosophila* PIWI proteins, Piwi, Aubergine (Aub), and Argonaute3 (Ago3), are expressed in the male and female germline cells (Wilson et al., 1996; Cox et al., 1998, 2000; Williams and Rubin, 2002; Brennecke et al., 2007; Gunawardane et al., 2007; Nishida et al., 2007).

The prevailing model for piRNA biogenesis—the "ping-pong" model—reflects the discovery that the first 10 nt of piRNAs bound to Aub or Piwi, which are largely antisense and typically

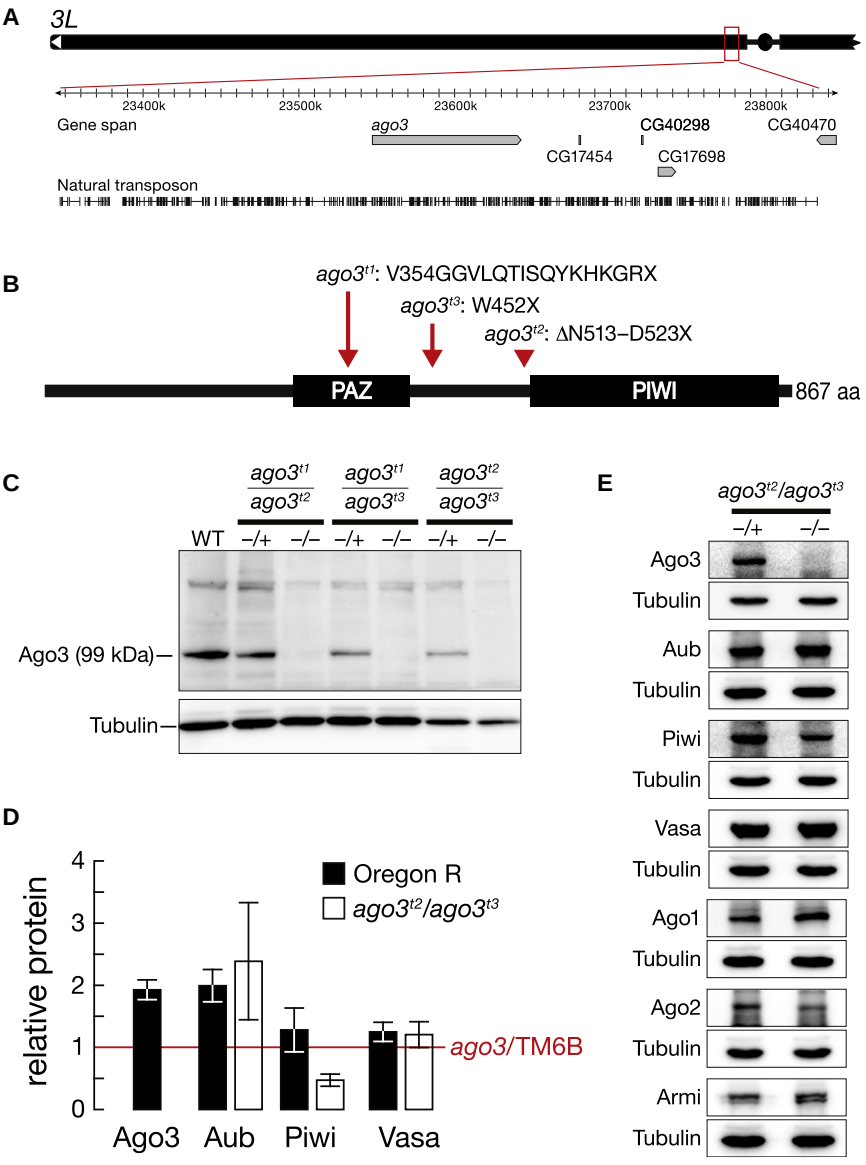


Figure 1. *ago3* Mutants

(A) The *ago3* gene resides in pericentromeric heterochromatin on the left arm of chromosome 3.

(B) *ago3* mutant alleles were identified by TILLING.

(C) Full length Ago3 protein was not detected in transheterozygous *ago3* ovaries, but was readily detected in heterozygotes and wild type ovaries.

(D) Protein levels in wild type and *ago3* ovaries, relative to *ago3*/TM6B (red line). The average ±standard deviation for at least three independent biological samples is shown.

(E) Representative data for Aub, Piwi, Vasa, Argonaute1 (Ago1), Argonaute2 (Ago2), and Armitage (Armi) in *ago3* ovaries. Tubulin served as a loading control.

begin with uridine, are often complementary to the first 10 nt of piRNAs bound to Ago3, which are largely sense and typically bear an adenosine at position 10 (Brennecke et al., 2007; Gunawardane et al., 2007). Many Argonaute proteins can act as RNA-guided RNA endonucleases, and all such Argonautes cut their target RNAs 5′ to the base that pairs with the tenth nucleotide of the small RNA guide; all three fly PIWI proteins retain their endonuclease activity (Saito et al., 2006; Gunawardane et al., 2007). Thus, the observed 10 nt 5′ complementarity between piRNAs suggests that the 5′ ends of piRNAs bound to Aub or Piwi are defined by Ago3-catalyzed cleavage, and, reciprocally, that the 5′ ends of piRNAs bound to Ago3 are defined by piRNAs bound to Aub or Piwi. The ping-pong model seeks to explain these observations, as well as the role of piRNA cluster transcripts in piRNA biogenesis, the function of piRNAs in silencing transposons, and the extraordinary antisense bias of piRNAs generally. At its core, the model proposes that piRNAs participate in an amplification loop in which transposon sense transcripts (e.g., transposon mRNAs) trigger the production of new, antisense piRNAs. Ago3, guided by sense piRNAs, lies at the heart of the amplification loop.

To test the ping-pong model, we isolated strong loss-of-function mutations in *ago3*. Here, we report the detailed analysis of *ago3* and *aub* mutant flies. Our data provide strong support for an amplification cycle in which Ago3 amplifies piRNA pools and enforces on them a strong antisense bias, increasing the number of piRNAs that can act to destroy transposon mRNAs. Moreover, we detect a second, perhaps somatic, piRNA pathway centered on Piwi and functioning without benefit of Ago3-catalyzed amplification. Most of the transposons targeted by this second pathway reside in the *flamenco* piRNA cluster, which was first identified as a repressor of transposon expression in somatic follicle cells (Pelisson et al., 1994).

RESULTS

Loss-of-Function *ago3* Alleles

The *Drosophila ago3* gene resides in pericentromeric heterochromatin on the left arm of chromosome 3 (Figure 1A). Heterochromatin is refractory to *P* element or *piggyBac* transposon insertion and to targeted recombination, complicating isolation of *ago3* mutants. We used TILLING (Cooper et al., 2008) to identify three mutant alleles (*t1*, *t2*, and *t3*) that create premature stop codons in *ago3* (Figure 1B) and obtained one chromosomal deficiency [*Df(3L)TTT*] that deletes at least six genes, including *ago3* (Figure S1A available online). One homozygous ($ago3^{t3}/ago3^{t3}$) and three transheterozygous ($ago3^{t1}/ago3^{t2}$, $ago3^{t2}/ago3^{t3}$,

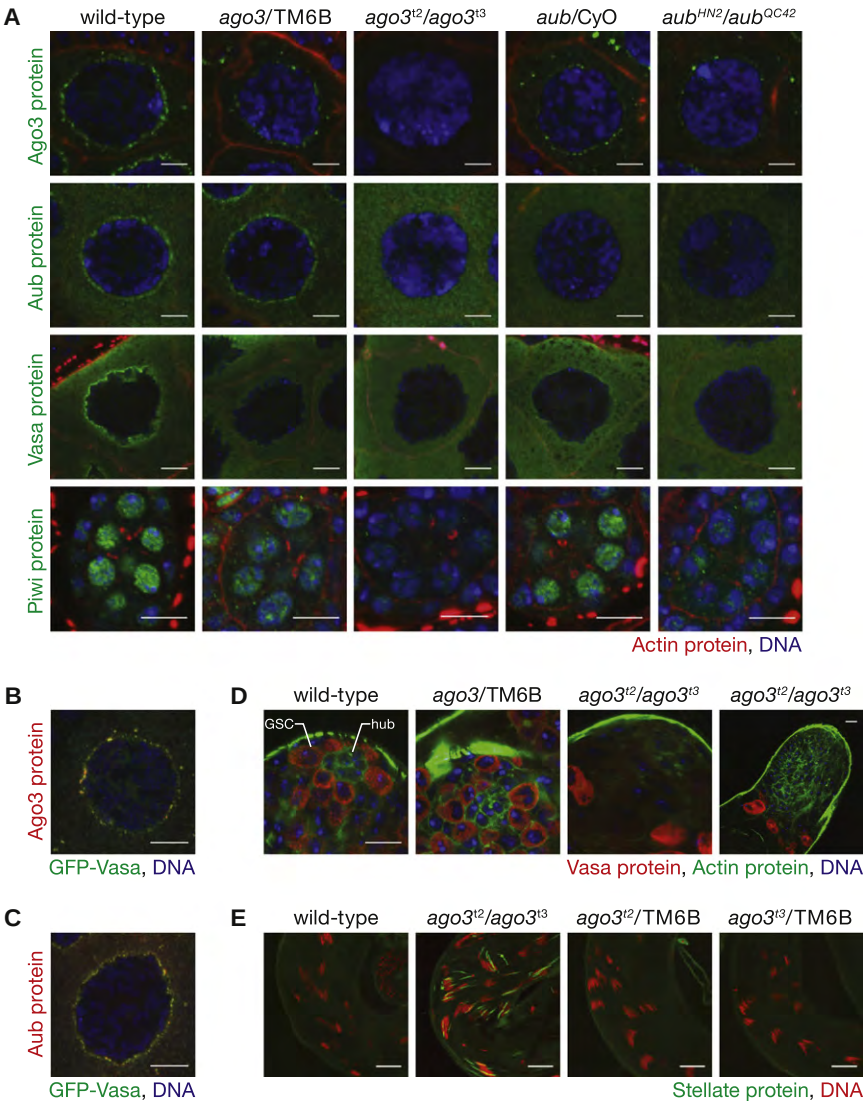


Figure 2. Consequences of Loss of Ago3 in the *Drosophila* Germline

(A) Mutually interdependent incorporation of Ago3 and Aub into nuage in ovaries. Images correspond to single confocal sections (63× magnification); scale bars represent 5 μm (Ago3, Aub, and Vasa panels) or 10 μm (Piwi).

(B and C) Vasa localization within nuage is only partially congruent with that of Ago3 and Aub. Images correspond to single confocal sections of stage 4/5 egg chambers (63× lens, 4× zoom). Scale bars represent 10 μm.

(D) In 5 to 7 day old male adults, Vasa expressing germline stem cells (GSC), which normally surround the somatic hub cells (hub), were not detected in *ago3* testes. Since Vasa expression increases in *ago3*, the two images from *ago3* mutants were acquired with reduced gain relative to wild type and heterozygous testes. Scale bars represent 10 μm (first three images) or 20 μm (fourth).

(E) *Stellate* silencing in testes requires Ago3. Scale bars represent 20 μm.

and $ago3^{t1}/ago3^{t3}$) combinations, as well as each mutant allele in *trans* to *Df(3L)TTT*, were viable. Full-length Ago3 protein was essentially undetectable in all seven allelic combinations (Figure 1C and Figure S1B), and $ago3^{t2}$/TM6B or $ago3^{t3}$/TM6B heterozygous ovaries contained half as much Ago3 protein as Oregon R wild-type ovaries (Figure 1D), suggesting that the *ago3* alleles correspond to strong loss-of-function mutations. We refer to $ago3^{t2}/ago3^{t3}$ transheterozygotes as *ago3* for brevity.

Mutually Interdependent Localization of PIWI Proteins

The intracellular localization of each PIWI protein appears to require other PIWI proteins. In the *Drosophila* female germline, Piwi is largely in the nucleus, whereas Aub and Ago3 are cytoplasmic and concentrated in "nuage," a ring around the cytoplasmic face of the nuclei of the transcriptionally active germline nurse cells (Findley et al., 2003; Brennecke et al., 2007; Pane et al., 2007). The putative RNA helicase Vasa, which marks the germline in most animals, is also in nuage (Liang et al., 1994). Mutations that disrupt piRNA biogenesis, but not those that block siRNA or miRNA production, disrupt localization of Aub to nuage (Lim and Kai, 2007).

Ago3 is also required for Aub to localize to nuage (Figure 2A). However, Ago3 is not required for Aub expression or stability: the concentration of Aub protein and the amount of Aub in the cytoplasm increased in *ago3* ovaries relative to their heterozygous siblings (Figures 1D and 1E). Reciprocally, Aub plays a role in the localization of Ago3 to nuage, although some Ago3 persisted in nuage in an *aub* hypomorphic mutant allelic combination (aub^{HN2}/aub^{QC42}). The function of nuage is unknown, but it may have a complex substructure, because the localization of Vasa within nuage only partially coincides with that of Ago3 (Figure 2B) and Aub (Figure 2C).

The interdependence of Ago3 and Aub for their localization to nuage likely reflects an underlying requirement for these two proteins for nuage assembly: the perinuclear localization of Vasa, a marker for nuage, was absent in both *ago3* and *aub* mutants, although the abundance of Vasa was unaltered in *ago3* (Figures 1D and 1E). The abundance of Armitage, Argonaute1, and Argonaute2, proteins required for small silencing RNA biogenesis or function, were similarly unaltered in *ago3* mutants (Figure 1E). In *ago3*/TM6B heterozygotes—which produce half as much Ago3 protein as the wild-type—and in *aub*/CyO heterozygotes, less Aub, Ago3, or Vasa was present in nuage than in the wild-type, suggesting that assembly of nuage is exquisitely sensitive to the amount of Ago3 and Aub (Figure 2A).

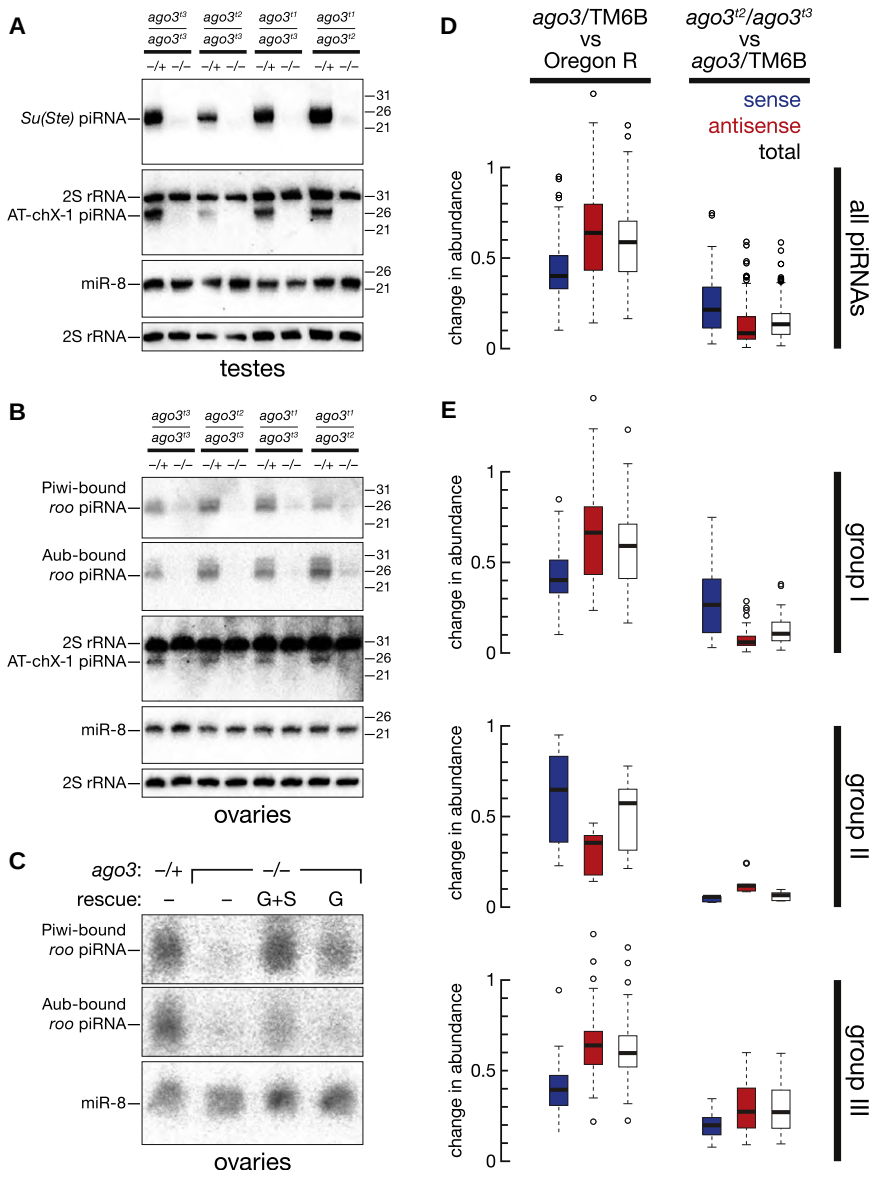


Figure 3. Aub and Piwi bound piRNAs Disappear without Ago3

(A) Accumulation of *Su(Ste)* and AT chX 1 piRNAs in testes required Ago3.

(B) Accumulation of *roo* antisense piRNAs, as well as the AT chX 1 piRNA, in ovaries required Ago3.

(C) Accumulation of Piwi and Aub bound *roo* antisense piRNAs in the ovary was rescued in *ago3* mutants by a single copy transgene expressing Ago3 in the germline (G) or in both the germline and the soma (G+S).

(D) Box plots illustrating the change in abundance of piRNAs, analyzed by transposon family, in *ago3*/TM6B versus Oregon R (left panel) and *ago3* versus *ago3*/TM6B (right panel) ovaries.

(E) Box plots illustrating the change in abundance for piRNAs for each of 95 transposon families, separated by group.

The peak of nuclear-localized Piwi protein expression occurs earlier in oogenesis than when Ago3 and Aub are maximally perinuclear (Cox et al., 2000). Less Piwi is found in the nucleoplasm of *ago3*/TM6B or *aub*/CyO heterozygotes than in the wild-type, and little or no Piwi is present in the nuclei of the *ago3* or *aub* mutants (Figure 2A); *ago3* ovaries contain about half as much Piwi as their heterozygous siblings (Figure 1E). (The localization of Aub and Ago3 cannot be studied in *piwi*, because *piwi* mutants arrest early in oogenesis.) We conclude that the correct intracellular localization of Ago3, Aub, and Piwi is interdependent.

As reported previously for *aub* testes, Vasa protein expression increased in male germ cells in *ago3* testes, relative to the wild-type or *ago3*/TM6B (data not shown), consistent with the proposal that the AT-chX-1 and AT-chX-2 piRNAs downregulate *vasa* expression (Nishida et al., 2007); AT-chX-1 piRNA levels were dramatically reduced in *ago3* testes (Figure 3A).

ago3 Mutations Affect Fecundity

ago3 females are sterile. *ago3* females initially laid far fewer eggs than their heterozygous siblings and by day 10 stopped laying altogether (Figure S2A). The egg-laying deficit was rescued by a single-copy transgene expressing Ago3 in the germline or in both the germline and the soma. Of those eggs laid, none of the embryos hatched, compared with ~96% for the wild-type (Table S1). For 39%–67% of the embryos produced by *ago3* mothers, the dorsal appendages were fused or absent (Table S1), indicating a maternally inherited patterning defect, as has been reported for other piRNA pathway mutations (Schupbach and Wieschaus, 1991; Wilson et al., 1996; Cook et al., 2004; Klattenhoff et al., 2007; Pane et al., 2007). Loss of Ago3 in the germline, rather than a secondary mutation present in the *ago3* mutants, caused the patterning defects, as a single-copy transgene expressing Ago3 from the germline-specific *nanos* promoter rescued the dorsal appendage defects to essentially wild-type rates (Table S1).

In males, Ago3 is required to maintain germline stem cells (Figure 2D). We examined testes from 5- to 7-day-old wild-type, *ago3*/TM6B, and *ago3* males, using Vasa protein expression to identify germline cells. In the wild-type and *ago3*/TM6B, Vasa-expressing cells were present at the tip of the testes, in a ring around the hub cells, a group of somatic cells that support the adjacent germline stem cells (Fuller, 1993). In contrast, the corresponding region in *ago3* was devoid of Vasa-expressing cells, suggesting that germline stem cells are not properly maintained in the absence of Ago3. Consistent with a failure to maintain germline stem cells, *ago3* males are semifertile, siring fewer offspring than do males with the same genetic background (Figure S2B).

Silencing Selfish Genetic Elements in Germline Requires Ago3

In males, antisense piRNAs derived from the *Suppressor of Stellate* [*Su(Ste)*] locus on the Y chromosome silence the X-linked *Stellate* locus during spermatogenesis (Balakireva et al., 1992; Palumbo et al., 1994; Bozzetti et al., 1995; Aravin et al., 2001, 2003; Vagin et al., 2006). *Su(Ste)* piRNAs were the first piRNAs to be identified (Aravin et al., 2001), and, in *Su(Ste)* mutants, piRNAs targeting *Stellate* are lost, causing Stellate protein crystals to form in primary spermatocytes (Livak, 1984; Pimpinelli et al., 1985; Livak, 1990). *Stellate* silencing requires the piRNA pathway genes, *aub*, *spindle* E, *armitage*, *squash*, and *zucchini* (Schmidt et al., 1999; Stapleton et al., 2001; Tomari et al., 2004; Pane et al., 2007). Stellate protein crystals form in primary spermatocytes in *ago3* testes (Figure 2E).

Virtually all *Su(Ste)* piRNAs are antisense (Vagin et al., 2006) and bound to Aub, not Piwi (Nishida et al., 2007). *Su(Ste)* piRNAs fail to accumulate in *aub* mutants, but accumulate to higher than normal levels in *piwi* testes (Vagin et al., 2006). We used northern hybridization to examine *Su(Ste)* piRNA production in the testes (Figure 3A). *Su(Ste)* piRNAs, as well as AT-chX-1 piRNA, were strongly reduced in *ago3* males, relative to their heterozygous siblings, for all seven allelic combinations examined (Figure 3A and data not shown). Thus, both Aub and Ago3 are required to silence *Stellate*, and both proteins are required to produce or stabilize those *Su(Ste)* piRNAs normally bound to Aub. We speculate that *Stellate*-derived, Ago3-bound sense piRNAs, while rare, amplify Aub-bound, antisense *Su(Ste)* piRNAs in testes.

piRNAs have been best characterized in the *Drosophila* ovary. piRNAs derived from *roo* LTR retrotransposons are among the most abundant ovarian piRNAs and are disproportionately antisense to *roo* coding sequences; these antisense piRNAs are bound to Aub and Piwi (Vagin et al., 2006; Gunawardane et al., 2007), but not to Ago3 (Brennecke et al., 2007). *roo* antisense piRNAs failed to accumulate in *ago3* ovaries, but were readily detectable in *ago3*/TM6B ovaries (Figure 3B). Production of *roo*-derived piRNAs was rescued by a single-copy transgene expressing Ago3 from the *actin5c* promoter and more weakly rescued by a single-copy transgene expressing Ago3 from the *nanos* promoter (Figure 3C).

Genome-wide piRNA Analysis

To obtain a broader view of the function of Ago3 in piRNA biogenesis, we sequenced piRNAs from wild-type (Oregon R), *ago3*/TM6B, $ago3^{t2}/ago3^{t3}$, *aub*/CyO, and aub^{HN2}/aub^{QC42} ovaries. Libraries were prepared from oxidized small RNAs to permit greater effective sequencing depth. In all, 3,282,391 genome-matching, non-ncRNA, non-miRNA, 23–29 nt long small RNAs were sequenced (Table S2). Despite the large number of sequences obtained from each genotype, most piRNA species were sequenced only once. This remained true even when all data sets were pooled, suggesting that piRNAs comprise the most diverse class of regulatory molecules in the fly. Together, our piRNA sequences from wild-type, *ago3*/TM6B, and *ago3* ovaries cover 10% of the fly genome.

In parallel, we immunoprecipitated Ago3, Aub, and Piwi from *ago3*/TM6B ovaries and Aub and Piwi from *ago3* ovaries and then constructed and sequenced libraries of the small RNAs bound to each protein (Table S2). We do not know the extent to which PIWI proteins copurify, e.g., because they are bound to a common RNA or present in a common complex. To avoid potential misassignment of piRNAs to a specific PIWI protein, we analyzed only those piRNAs that associated uniquely with Ago3, Aub, or Piwi.

We calibrated the abundance of the piRNAs uniquely associated with Ago3, Aub, or Piwi such that the aggregate abundance of a subset of piRNAs in an immunoprecipitation data set equaled that in the total piRNA data set from the corresponding genotype. This subset of piRNAs was defined as those that mapped only once to the fly genome and were sequenced at least once in both of the two data sets. This strategy allows direct comparison of the relative abundance of piRNAs uniquely bound to one PIWI protein with those bound to another, as well as comparison of the uniquely bound piRNAs between *ago3* heterozygotes and mutants.

Ago3 Limits Sense piRNA Accumulation and Amplifies Antisense piRNAs

Both *ago3* and *ago3*/TM6B ovaries had fewer piRNAs than wild-type ovaries (Figure 3D). Ago3 was previously found to bind mainly piRNAs corresponding to the sense, mRNA strand of transposons. Our data suggest that the intracellular concentration of Ago3 limits the accumulation of sense piRNAs: the median abundance sense piRNAs, analyzed by transposon family, in *ago3*/TM6B heterozygotes was ~40% that of wild-type ovaries. Antisense piRNAs were less affected by halving of the amount of Ago3: the median abundance of antisense, transposon-mapping piRNAs in *ago3*/TM6B ovaries was ~64% that of the wild-type. While complete loss of Ago3 further depressed the abundance of transposon-mapping, sense piRNAs, antisense piRNA accumulation in *ago3* ovaries collapsed. The median abundance by transposon family for antisense piRNAs in *ago3* ovaries was less than a tenth of the median abundance in the heterozygotes, and less than one-twentieth the median abundance in wild-type ovaries. The data support the view that for most selfish genetic element families, Ago3, presumably guided by sense piRNAs, acts to amplify antisense piRNAs bound to Aub.

Immunoprecipitation data confirmed this idea. In *ago3*/TM6B ovaries, 71% of small RNAs uniquely bound to Aub were antisense, but in *ago3* mutants, only 41% of the remaining Aub-bound piRNAs were antisense (Table S3). Thus, the characteristic antisense bias of Aub-bound piRNAs is enforced by Ago3. The strand bias of Piwi-bound piRNAs was less affected by loss of Ago3: in the *ago3* heterozygotes, 73% of the piRNAs uniquely associated with Piwi were antisense; in the mutants, 62% of the remaining piRNAs uniquely bound to Piwi were antisense. That the strandedness of Piwi-bound piRNAs changes at all suggests that Ago3 plays a role in the production of at least some antisense, Piwi-bound piRNAs.

Three piRNA Groups

While antisense piRNAs are generally more abundant than sense piRNAs, individual transposon families have distinct ratios of sense and antisense piRNAs. To determine the role of Ago3 in establishing these ratios, we analyzed the fraction of sense

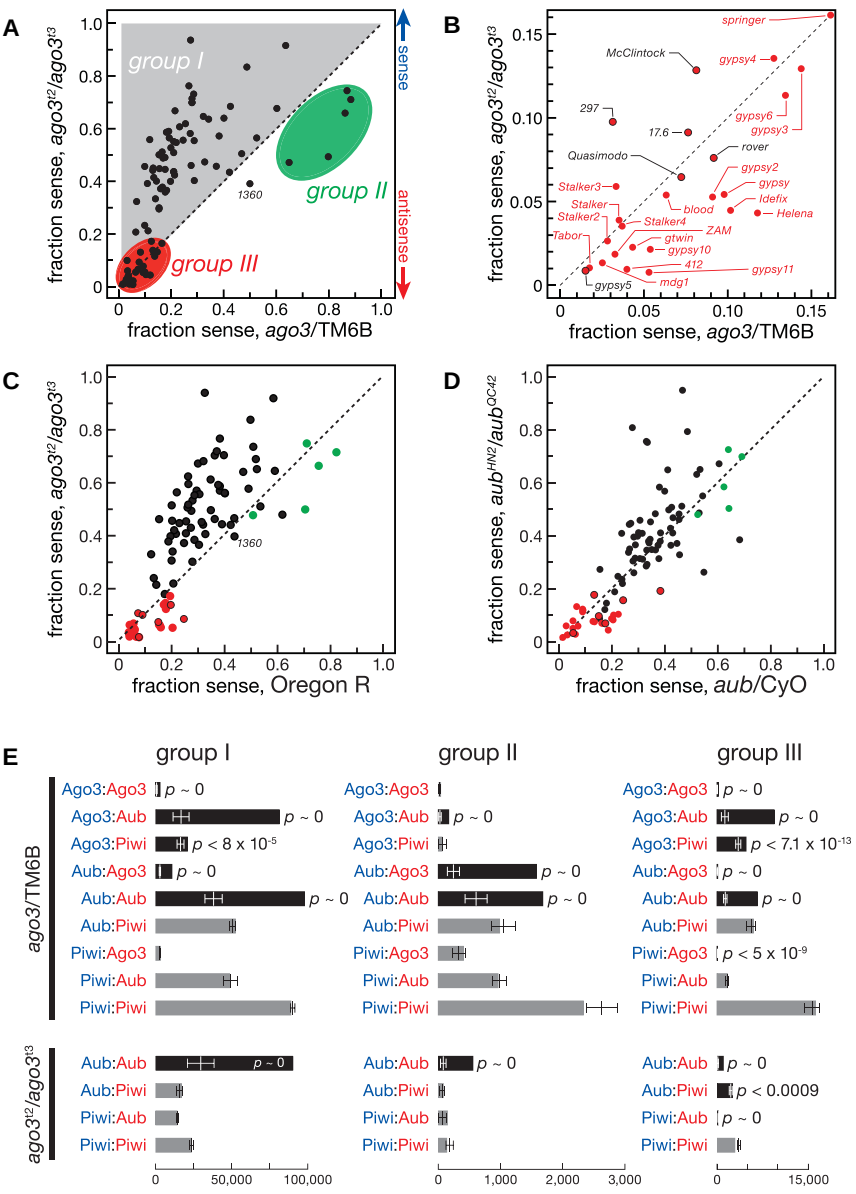


Figure 4. Three Distinct Groups of piRNAs

(A) The sense fraction of the piRNAs was compared for each transposon family between *ago3* and *ago3*/TM6B ovaries.

(B) An enlargement of the lower left corner of (A). Transposons labeled in red are present in the *flamenco* locus. Here and in (C) and (D), group III transposon families not present in *flamenco* are denoted with a red filled black circle.

(C) The sense fraction of piRNAs was compared between *ago3* and Oregon R ovaries.

(D) The sense fraction of the piRNAs was compared between *aub* and *aub*/CyO ovaries.

(E) The normalized abundance of ping pong pairs detected for all nine (*ago3*/TM6B) or four (*ago3*) possible PIWI protein pairings for the piRNA species uniquely associated with a single PIWI protein. Blue, sense; red, antisense. Black bars, Bonferroni corrected p value <0.005; gray bars, p value >0.005. p values $<10^{-30}$ are reported as "$p \sim 0$." Values are shown as average ±two standard deviations.

piRNAs (sense/[sense + antisense]) for each of the 95 transposon families for which we obtained ≥500 reads in *ago3*/TM6B ovaries. We compared the fraction of sense piRNAs in the heterozygotes to the fraction of sense piRNAs in *ago3* mutants (Figures 4A and 4B). We also compared the fraction of sense piRNAs in the heterozygotes to both the fraction of sense piRNAs in *ago3* mutants and the fold change in antisense piRNAs between the two genotypes (Figure S3).

We detected three groups of transposons. For 63 transposon families, the abundance of antisense piRNAs declined dramatically, causing an increase in the fraction of sense piRNAs (Figure S3). Among these 63 group I transposon families, 61 also had a decreased antisense bias comparing *ago3* to Oregon R (Figure 4C). (For convenience, the enigmatic transposon family *1360* was included in group I in our subsequent analyses.)

Five transposon families compose group II. These all had more sense than antisense piRNAs in *ago3* heterozygotes, and four of the five had more sense than antisense piRNAs in wild-type ovaries, as previously noted (Brennecke et al., 2007). In *ago3* ovaries, both the fraction (Figures 4A, 4C, and S3) and the absolute amount (Figure 3E) of sense piRNAs declined for group II transposons, relative to heterozygous or wild-type ovaries.

piRNAs mapping to the 26 group III transposons were disproportionately antisense, ranging from 80% to nearly 100% antisense in *ago3* heterozygotes (Figures 4A and 4B). In *ago3* ovaries, group III transposons generally retained their antisense piRNAs to a greater extent than those in groups I and II (Figures 3E and S3).

Paradigmatic examples of each group—*HeT-A* for group I, *accord2* for group II, and *ZAM* for group III—are analyzed in greater detail in Figure 5. Figures S4–S6 present the corresponding data for each of the 95 transposons, as well as the tandem repeat *mst40*.

Group I Transposons Require Ago3 for Antisense piRNA Amplification

HeT-A is the quintessential group I transposon (Figure 5). The piRNAs that map to group I transposons show a strong "ping-pong" signature that derives from the 10 nt overlap between antisense piRNAs bound to Aub and sense piRNAs bound to Ago3 and

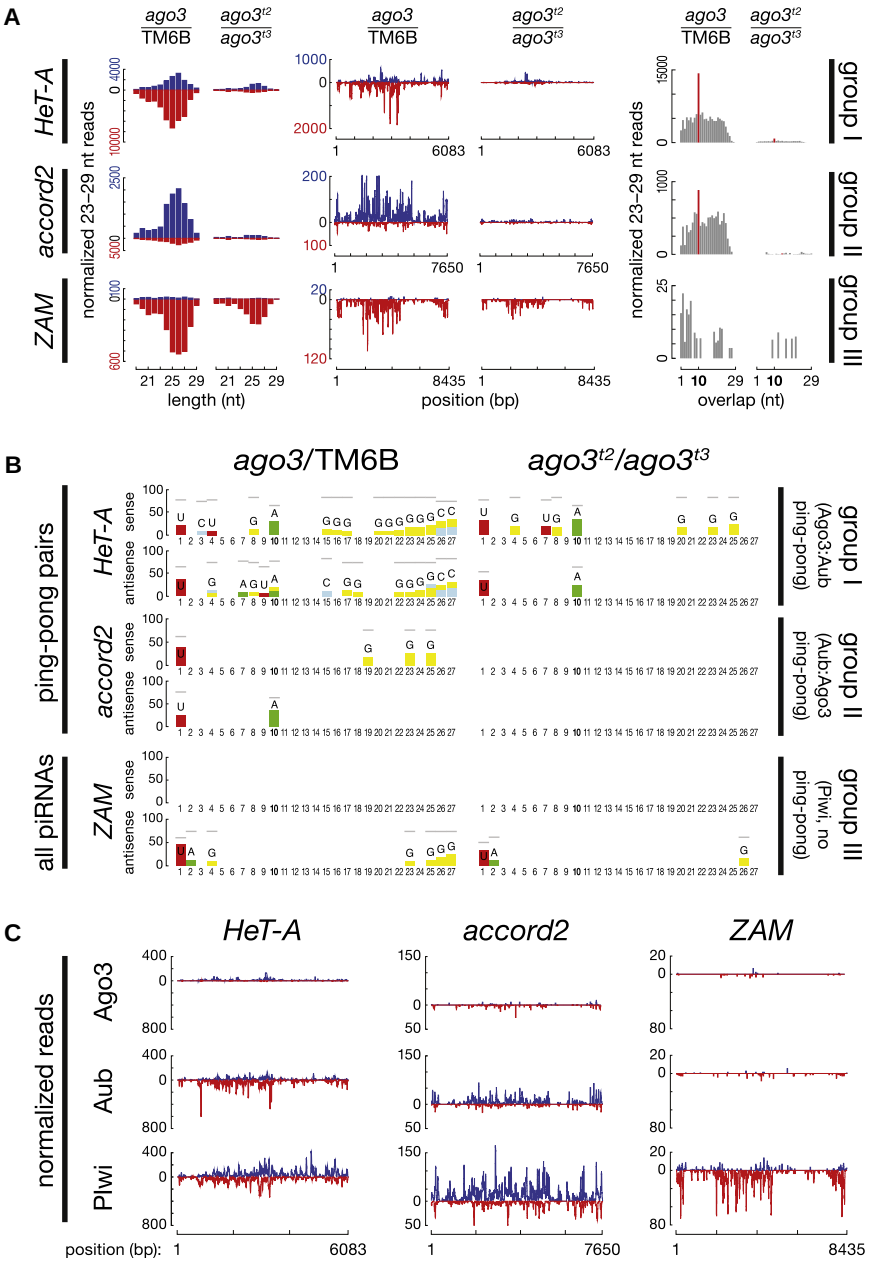


Figure 5. Paradigmatic Examples of Each Transposon Group

(A) piRNA length distribution, abundance relative to consensus position, and ping pong pair abundance. Blue, sense piRNAs; red, antisense.

(B) Sequence excess logos generated by subtracting the background from the relative frequency of each nucleotide at each position of the 23–29 nt RNAs in each sample. Only the nucleotide positions where foreground was significantly higher than background are shown (Bonferroni corrected p values <0.001). Gray horizontal bars indicate the maximum possible value for each position.

(C) Normalized and calibrated abundance of piRNAs uniquely associated with Ago3, Aub, or Piwi in *ago3*/TM6B.

Aub (Figures 4E). For group I transposons, although Ago3 associates almost exclusively with sense piRNAs, it is the Aub-bound antisense piRNAs that disappear in its absence (Figure S6). A central postulate of the ping-pong model is that the action of Ago3, rather than Aub, drives the production of more antisense than sense piRNAs (see Supplemental Discussion). Consistent with this proposal, the fraction of sense piRNAs for most group I transposons shows little overall change in *aub* ovaries (Figure 4D).

Group II Transposons Act "Backwards"

For group II transposons, antisense piRNA levels appear to reflect the abundance of Ago3: the median abundance by transposon family of group II antisense piRNAs in *ago3* heterozygotes was ~35% that of the wild-type (Figure 3E). Sense piRNAs declined less when the abundance of Ago3 was halved. However, in the absence of Ago3, group II sense piRNAs declined ~17-fold from their levels in heterozygotes and ~27-fold from their levels in the wild-type. Group II antisense piRNAs also decreased, but less dramatically. The "backwards" behavior of group II piRNAs, relative to group I, suggests that the production or accumulation of Aub-bound piRNAs generally requires Ago3, irrespective of their sense or antisense identity.

piRNAs from group I and group II transposons normally partition among PIWI proteins according to their orientation and sequence bias (Figures S4–S6). Group I antisense (or group II sense) piRNAs typically begin with U (U1) and bind Aub, whereas group I sense (or group II antisense) piRNAs show less 5′ nucleotide bias but typically bear an A at position 10 (A10) and bind Ago3 (e.g., *accord2* in Figure 5). In the absence of Ago3, both sense and antisense piRNAs remained associated with Aub (Figure S6). Some of these showed ping-pong pairing, but without respect to orientation, so that the U1 that normally characterizes antisense piRNAs and the A10 that normally characterizes sense piRNAs became conflated (Figures 5B and S4). In the absence of Ago3, those Aub-bound piRNAs with an A at position 10 began with U more often than expected by chance (Figure S7) and began with U about as often as did those Aub-bound piRNAs without an A at position 10. These observations suggest that a 5′ U favors binding of a piRNA to Aub. Moreover, the data suggest that the A10 signature of Ago3-bound piRNAs is a consequence of the preponderance of Aub-bound piRNAs that start with U, consistent with the idea that Aub directs the production of Ago3-bound piRNAs.

piRNAs for Group III Transposons Are Produced by Both Aub- and Ago3-Dependent and Aub- and Ago3-Independent Pathways

In the absence of Ago3, the abundance of group III antisense piRNAs decreased far less than for group I (group I median = 0.06, group III median = 0.27; Figures 3E and S3). Antisense piRNAs from group III transposon families were mostly bound to Piwi: the median ratio between the amount of antisense piRNAs bound by Piwi and the amount bound by Aub in *ago3* ovaries was 31.3 for group III transposon families, a 5-fold increase from the median Piwi/Aub ratio for group III transposons in *ago3*/TM6B ovaries. This contrasts with the median Piwi/Aub ratios for group I (2.00) or II (2.94) transposon families (Figure S8A). Furthermore, loss of Ago3 caused group III antisense piRNAs bound to Piwi to decrease far less than antisense piRNAs bound to Aub; for group I, piRNAs bound to Piwi and Aub declined to a comparable extent (Figure S8B). Group III antisense piRNAs were also less affected by loss of Aub than the other two groups (Figures S9 and S10). Collectively, these observations suggest that group III transposons are predominantly silenced by a Piwi-dependent, Ago3-independent pathway.

While Piwi-bound piRNAs persisted in the absence of Ago3 to a greater extent than Aub-bound piRNAs, the absence of Ago3 clearly reduced the abundance of Piwi-bound piRNAs for all groups (Figure S8B). Perhaps some but not all Piwi-bound piRNAs are generated in the germline by an Ago3-directed amplification cycle. Supporting this view, we detected a statistically significant ($p < 7.1 \times 10^{-13}$) 10 nt overlap between sense Ago3-bound piRNAs and antisense Piwi-bound group III piRNAs (Figure 4E). Such group III Ago3 (sense):Piwi (antisense) piRNA ping-pong pairs were about half as abundant as Ago3:Aub pairs.

Most group III antisense piRNAs bound to Aub were eliminated in *ago3* mutants (Figure S8), suggesting that our group III piRNA data conflate at least two distinct pathways: a Piwi-dependent, Ago3- and Aub-independent pathway in which piRNA-directed piRNA amplification plays little if any role and an Aub-dependent pathway in which sense piRNAs bound to Ago3 act catalytically to amplify the antisense piRNAs associated with Aub. piRNAs acting in the Piwi-dependent pathway were disproportionately antisense, and this antisense bias required neither Ago3 nor Aub: production of antisense piRNAs for group III transposons remained essentially unchanged in aub^{HN2}/aub^{QC42} mutants (Figure 4D).

Group III Transposons Often Reside in the *flamenco* piRNA Cluster

Remarkably, of the 26 transposon families comprising group III, 20 are present in the *flamenco* locus (Pelisson et al., 1994; Brennecke et al., 2007). *flamenco* was originally identified as required for the silencing of *gypsy* (Prud'homme et al., 1995) and later shown to be required to silence *ZAM* and *Idefix* (Mevel-Ninio et al., 2007; Desset et al., 2008). *gypsy*, *ZAM*, and *idefix* are all group III transposons. Reporter experiments suggest that Piwi is required to silence these transposons (Sarot et al., 2004; Desset et al., 2008). Given that Piwi, but not Aub or Ago3, is readily detected in ovarian somatic follicle cells (Cox et al., 2000; Saito et al., 2006; Gunawardane et al., 2007; Nishida et al., 2007); that *gypsy* silencing requires *piwi* but not *aub* (Pelisson et al., 2007); that the *gypsy* promoter drives expression of *gypsy* in the somatic follicle cells, which produce retrovirus-like particles that then infect the oocyte (Pelisson et al., 1994); that *ZAM* and *Idefix* are silenced in the somatic follicle cells by Piwi (Desset et al., 2008); and that follicle cell clones mutant for *piwi* or *flamenco* have remarkably similar defects (Cox et al., 1998; Mevel-Ninio et al., 2007), it is tempting to speculate that group III transposons in general are repressed by a Piwi-dependent, Ago3- and Aub-independent pathway that operates in follicle cells. In this view, antisense, Piwi-bound piRNAs would provide the primary somatic cell defense against group III transposon expression, whereas an Aub- and Ago3-dependent pathway provides a secondary defense in the germline, the ultimate target of these transposons.

For groups I and II, loss of Ago3 switches the bias of both Aub and Piwi-bound piRNAs from strongly antisense-biased (or strongly sense-biased for group II) to slightly sense-biased (or antisense-biased for group II). That is, in wild-type flies, Ago3 acts to skew the strandedness of piRNA pools for group I and group II transposons. For group III transposons, however, loss of Ago3 slightly increased the antisense bias of piRNAs uniquely bound to Piwi: in *ago3*/TM6B ovaries, 88% of such piRNAs were antisense; in *ago3*, 90% were antisense (Table S3). We conclude that most Piwi-bound group III piRNAs are made directly from antisense transcripts such as the hypothesized precursor transcript that spans the *flamenco* locus. This explanation is consistent with the previous proposal that *flamenco* triggers silencing of transposons such as *gypsy*, *ZAM*, and *idefix* in somatic follicle cells—a cell type that expresses little if any Ago3 or Aub.

Ago3 Amplifies piRNAs

Overall, piRNAs associated with Piwi were ~15-fold and piRNAs bound to Aub were ~6.4-fold more abundant than those bound to Ago3. The greater abundance of Piwi- and Aub-bound piRNAs is consistent both with the idea that Ago3 is less abundant than the other two PIWI proteins and with the proposal that Ago3 acts catalytically to amplify Piwi- and Aub-bound piRNAs. Consistent with the idea that group III piRNAs are largely Ago3-independent, the likelihood of a group III piRNA being associated with Piwi rather than Aub more than doubled in *ago3* ovaries: the Piwi-bound to Aub-bound piRNA ratio was 5.0 in *ago3*/TM6B, but 11.7 in *ago3*. This trend was not observed for group I piRNAs, where Piwi-bound piRNAs were twice as abundant as Aub-bound for *ago3*/TM6B, but only 1.6 times more abundant in *ago3* ovaries. Thus, for group III piRNAs, the absence of Ago3-catalyzed amplification shifts the piRNA pool toward the Piwi-dependent pathway.

Loss of Ago3 Increases Group I Transposon Expression

What is the molecular consequence of piRNA loss? We used whole-genome tiling microarrays to measure the effect of loss of Ago3 on gene and transposon expression (Figures 6A and S11) and quantitative RT-PCR of selected transposons to corroborate the microarray data (Figure 6B). The abundance of genic mRNA levels were generally unchanged in both *ago3* (Figure S11A) and *aub* (Figure S12A) ovaries, compared with wild-type controls (w^{1118}). In contrast, with a false-discovery

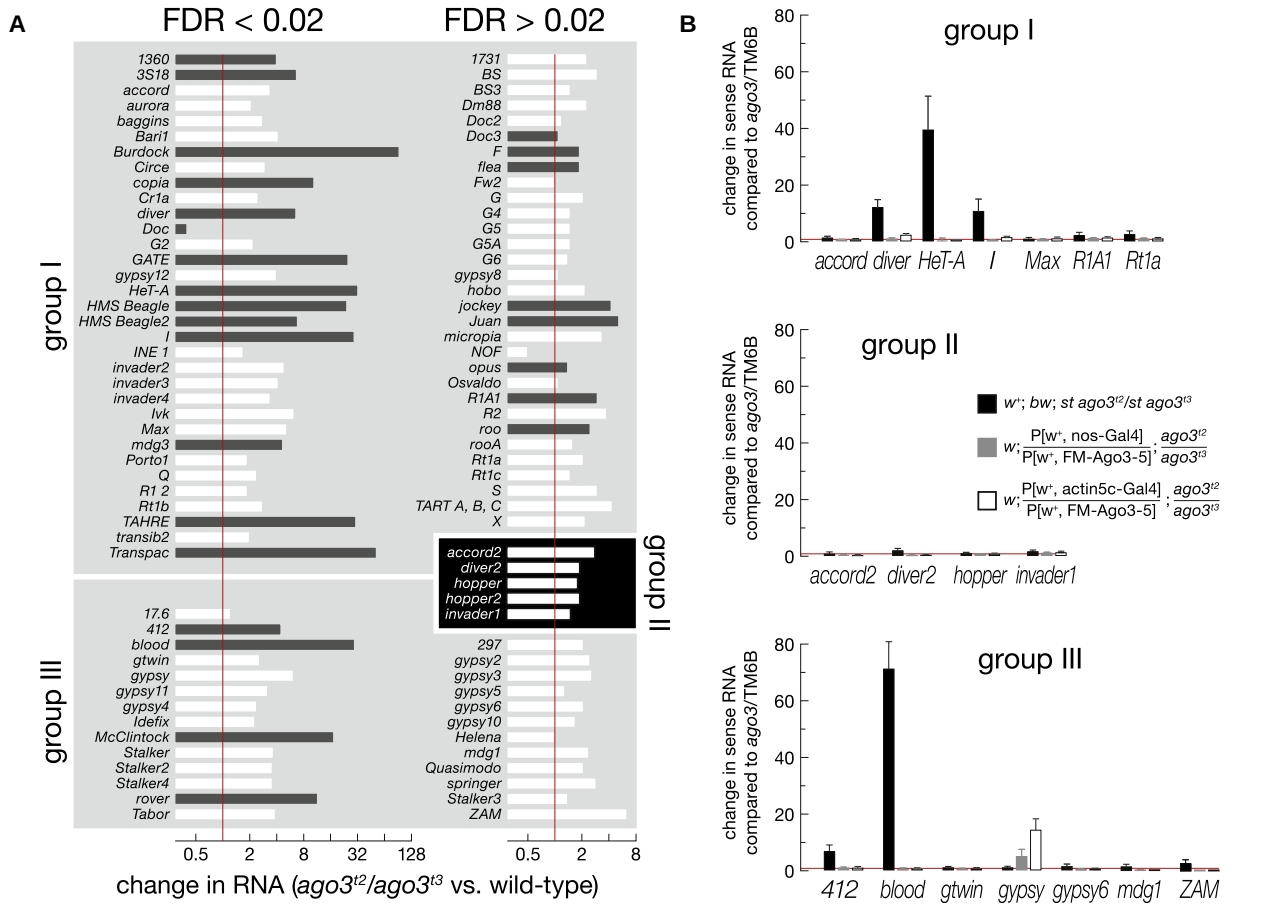


Figure 6. Loss of Ago3 Increases Expression of Some Group I and Group III, but not Group II, Transposons

(A) Expression of group I, group II, and group III transposon families in *ago3* ovaries, relative to wild type (w^{1118}) ovaries, was assayed with whole genome tiling microarrays. White bars show expression of the transposon family in both wild type and *ago3* ovaries was less than that of the 50th percentile for expression of all mRNAs in wild type, suggesting that expression change cannot be reliably quantified. Black bars show transposon families with expression greater than this threshold in one or both genotypes. Significant (false discovery rate [FDR] <0.02) and nonsignificant data (FDR >0.02) are separated.

(B) Quantitative RT PCR was used to assess transposon expression, relative to Actin, for *ago3* ovaries (black bars) and ovaries expressing one copy of a UAS Ago3 transgene driven by *nanos* Gal4 (gray bars) or *actin5c* Gal4 (white bars), relative to *ago3*/TM6B ovaries. The figure reports mean ±standard deviation for three independent biological samples.

rate (FDR) <0.02, the expression of 32 of the 64 group I transposons and 14 of the 26 group III transposons increased in the absence of Ago3 (Figures 6A and S11B). However, the level of expression for many of these in both wild-type and *ago3* ovaries was less than the 50th percentile of expression for mRNA in wild-type ovaries, making accurate quantification of their change in expression challenging. With this threshold for expression, 13 of the 64 group I and four of the 26 group III transposon families showed increased expression in the absence of Ago3 (Figure 6A).

Despite the loss of group II piRNAs in *ago3* mutants, expression of the five group II transposons was not significantly altered at a FDR <0.02 (Figure 6A). Perhaps the production of Aub-bound sense piRNAs for group II transposons is futile and does not silence these elements, since they were likewise not activated in an *aub* mutant (Figures S12B and S13).

Compared with wild-type ovaries, expression of 18 of the 64 group I transposon families increased in *aub* ovaries (eight of these 18 had expression in both w^{1118} and *ago3* greater than the 50th percentile for mRNA expression in the wild-type), including 14 that also increased in *ago3* ovaries (Figures S12B and S13). At a FDR <0.02, the expression of only two group III transposons—*rover* and *McClintock*—was increased; both were also desilenced in *ago3* ovaries. These data support the idea that group I transposons rely on Ago3 and Aub for silencing, whereas group III transposons are silenced by Piwi, with help from Ago3, likely in the germline only, for some transposons.

Germline Expression of Ago3 Rescues Group I Transposon Silencing

Our data suggest that Ago3 is required for piRNA production and silencing of group I transposons in the female *Drosophila* germline. As a final test of this hypothesis, we expressed Ago3 in *ago3* ovaries using the germline restricted *nanos* promoter. We used quantitative RT-PCR to measure the levels of mRNA for seven group I transposons, including five whose expression was

increased in the absence of Ago3: *HeT-A*, *Diver*, *I element*, *R1A1*, and *Rt1a*. For each, germline expression of Ago3 using the *nanos* promoter rescued silencing (Figure 6B).

Misexpression of Ago3 in the Soma Interferes with Group III Transposon Silencing

Argonaute proteins can compete for binding small RNAs. For example, decreasing the concentration of Ago1 in cultured S2 cells increases the loading of miRNAs into Ago2, whereas decreasing Ago2 increases miRNA loading into Ago1 (Förstemann et al., 2007; Horwich et al., 2007). Similarly, misexpression of Aub in the soma inhibits Ago2-mediated RNAi (Specchia et al., 2008). To test the role of Ago3 in silencing transposons in somatic ovarian cells, we used the *actin5c* promoter to express Ago3 in both the germline and the soma of *ago3* ovaries.

Surprisingly, ectopic expression of Ago3 in the soma *increased* the expression of the group III transposon *gypsy*, but enhanced silencing of the group I transposon *HeT-A* (Figure 6B). First, these data confirm our observation that increased Ago3 expression in the female germline increases silencing of group I transposons, perhaps because the concentration of Ago3 in the germline nurse cells limits piRNA amplification. Second, the data suggest that Ago3 cannot silence those transposons normally expressed in the somatic cells of the ovary. Perhaps Ago3 competes with Piwi for piRNAs, but, unlike Piwi, cannot act directly to silence transposons. Further supporting the idea that PIWI proteins compete for piRNAs, *gypsy* was silenced to an ~43-fold greater extent in *aub* ovaries than in *aub*/CyO (Figure S13). We speculate that when Aub levels are low, more *gypsy* piRNAs associate with Piwi, leading to enhanced silencing of *gypsy.*

DISCUSSION

Disentangling Multiple piRNA Pathways

Because ovaries contain both germline and somatic cells, our data conflate two distinct cell lineages. Combining our data with extensive genetic studies of *gypsy* and other transposon families represented in the *flamenco* locus, we have attempted to disentangle germline and somatic piRNA function (Figure 7). We propose that the somatic piRNA pathway is the more straightforward, involving only Piwi and not Ago3 or Aub. Hannon and coworkers (Malone et al., 2009) similarly deduce the existence of a somatic piRNA pathway by analyzing a broad panel of piRNA mutants and examining piRNAs from early embryos, which contain maternally deposited germline piRNAs but lack follicle cell-derived piRNAs. Thus, their study and ours infer from distinct data sets the existence of a somatic pathway in which primary piRNAs derived from *flamenco* are loaded directly into Piwi and not further amplified.

Our data suggest that Piwi cannot act alone to amplify piRNAs. We envision that Piwi-bound piRNAs in the soma are produced by a ribonuclease that randomly generates single-stranded guides that are subsequently loaded into Piwi and trimmed to length. Although Piwi-bound piRNAs generally begin with U and Piwi shows in vitro a preference for binding small RNA that begins with U (Yin and Lin, 2007), current evidence cannot distinguish between a putative piRNA-generating ribonuclease

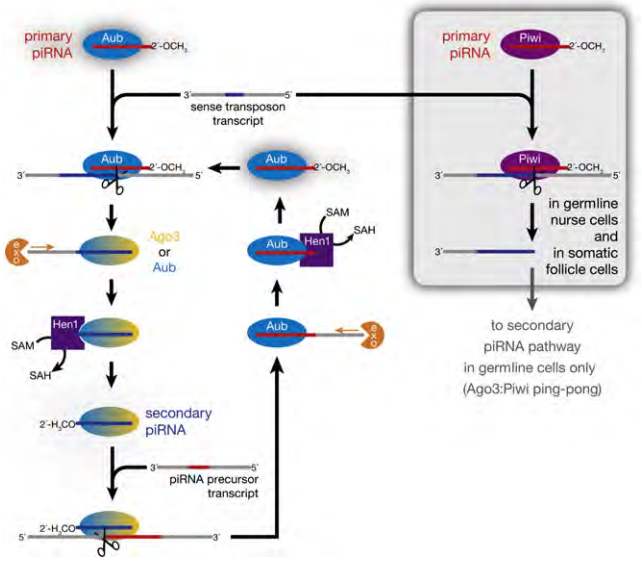


Figure 7. A Model for piRNA Biogenesis

The Aub and Ago3 dependent piRNA amplification cycle is envisioned to operate only in the germline, whereas a Piwi dependent, Aub and Ago3 independent pathway is shown for somatic cells. In the germline, Piwi can also partner with Ago3 to amplify piRNAs.

cleaving mainly at U and Piwi selecting U1 piRNAs from a set of RNAs with all possible 5′ nucleotides.

Without an amplification cycle to ensure an antisense bias, some other mechanism must operate to explain why Piwi-bound piRNAs are overwhelmingly antisense. A plausible but somewhat unsatisfying explanation comes from *flamenco* itself, whose constituent transposons are nearly all oriented in a single direction, so that the ~160 kb *flamenco* transcript is almost entirely antisense to the transposons. How such a nonrandom array of transposons could arise is unknown. Other nonrandomly oriented piRNA clusters may explain the smaller number of transposons in group III that are not present in *flamenco*.

The transposons in most piRNA clusters do not show such a pronounced nonrandom orientation. These probably act in the germline to produce primary piRNAs that load into Aub. The observed antisense bias of Aub-bound piRNAs arises subsequently, when Aub generates Ago3-bound secondary piRNAs and Ago3 acts, in turn, to produce Aub-bound secondary piRNAs. We propose that in the absence of Ago3, the sense/antisense ratio of Aub-bound piRNAs reverts to the inherent sense/antisense bias of the transposable element sequences present in the transcripts of piRNA clusters.

For this cycle to skew the Aub-bound piRNA population toward antisense, the substrate for cleavage by primary piRNA-bound Aub must be largely sense RNA. The best candidate for such sense RNA is mRNA derived from actively transcribed transposon copies. If such sense mRNA were largely found in the cytoplasm, it would be spatially segregated from the cluster transcripts, which we envision to be retained in the nucleus. Supporting this idea, sense transcripts from the group I transposon, *I* element, normally accumulate only in the nuclei of germline nurse cells (Chambeyron et al., 2008), likely because

they are destroyed in the cytoplasm by Aub-bound primary piRNAs and Aub-bound piRNAs produced by Ago3-dependent amplification. In the absence of Aub, these sense transposon transcripts accumulate in the cytoplasm instead, consistent with the strong desilencing of *I* element in *aub* and *ago3* mutants (Figures 6, S12, and S13) (Vagin et al., 2006).

Nuage and the Paradox of piRNA Production

The piRNA ping-pong hypothesis predicts a role for Ago3 in the production of Aub-bound antisense piRNAs, but our finding that loss of Ago3 also reduced the abundance of Piwi-bound antisense piRNAs was unexpected. The majority of Aub and Ago3 is found in nuage and in the cytoplasm, but Piwi is predominantly nuclear. How then can Ago3 direct the production of Piwi-bound piRNAs? Perhaps Piwi transits the nuage en route from its site of synthesis, the cytoplasm, to where it accumulates, the nucleus. In this view, cytoplasmic Piwi is predicted to lack a small RNA guide. Piwi would then acquire its small RNA guide in the nuage, through a process that requires Ago3. Loading a piRNA into Piwi might then license it for entry into the nucleus, where it could act posttranscriptionally or transcriptionally to silence transposon expression. In this view, mutations in genes required for nuage assembly or stability, such as *vasa*, as well as genes required for Piwi loading would reduce the amount of nuclear Piwi. A similar mechanism may operate in mammals, where the PIWI protein MILI is found in cytoplasmic granules, whereas MIWI2 is nuclear. In the absence of MILI, MIWI2 delocalizes from the nucleus to the cytoplasm, although MIWI2 is not required for the localization of MILI (Aravin et al., 2008).

Such a model cannot explain the loading of Piwi in the somatic follicle cells, which contain little or no Ago3 or Aub and which do not contain nuage. A simple but untested hypothesis for these cells is that in the absence of nuage, empty Piwi readily enters the nucleus, where it obtains its small RNA guide. We might reasonably expect that in germ cells the absence of nuage would impair the loading of Piwi by eliminating the Ago3-dependent, germline-specific Piwi-loading process, but also facilitate entry of some empty Piwi into the nucleus, where it could obtain small RNA guides. Consistent with this idea, we do detect some Piwi in the nucleus in *ago3* ovaries. The simplicity of this hypothesis, of course, belies the complexity of testing it.

Why Two Distinct piRNA Production Pathways?

Retrotransposons "reproduce" by producing sense RNA encoding transposases and other proteins that allow them to jump to new locations in the germ cell genome. The conservation of the piRNA ping-pong cycle in animals (Grimson et al., 2008) suggests that it is an ancient and conserved germline defense against retrotransposition. In flies, the *gypsy* family of retroelements appears to have moved its reproductive cycle to the somatic follicle cells adjacent to the germline, which it infects using retrovirus-like particles. *gypsy* thus appears to avoid germline piRNA surveillance by transcribing and packaging its RNA in the soma. Perhaps expression of Piwi in *Drosophila* follicle cells reflects an adaptive evolutionary counter move to the *gypsy* reproductive strategy. The simplicity of the direct loading of Piwi with antisense piRNAs derived from *flamenco* may have made this counter defense more evolutionarily accessible than a strategy requiring expression of all the proteins needed for the Ago3:Aub ping-pong mechanism. In the future, more extensive analysis of the cellular and genetic requirements for ping-pong-independent and ping-pong-dependent piRNA mechanisms in *Drosophila melanogaster* and in more ancient animal species may provide a test for these ideas.

EXPERIMENTAL PROCEDURES

Detailed experimental and computational procedures are described in Supplemental Data.

Isolation of *ago3^{t1}*, *ago3^{t2}*, and *ago3^{t3}* Alleles

From the 6000 EMS mutagenized fly lines (Koundakjian et al., 2004) screened by the Seattle TILLING project (http://tilling.fhcrc.org/), candidate lines that contained mutations that induced premature stop codons in the *ago3* coding sequence were characterized by sequencing of genomic PCR amplicons.

Molecular Cloning and Generation of Transgenic Flies

Ago3 was cloned from Oregon R ovary complementary DNA (cDNA) by 3′ RACE. Ovary RNA was reverse transcribed. First strand cDNA was treated with RNase H. Full length Ago3 coding sequence was amplified by PCR, cloned, and recombined into the pPFMW vector. Insert junctions were confirmed by sequencing. Plasmid DNA was injected, and transgenic flies were identified.

Immunohistochemistry and Microscopy

Egg chamber fixation and whole mount antibody labeling were performed as previously described (Theurkauf, 1994). Samples were analyzed with a Leica TCS SP inverted laser scanning microscope, with identical imaging conditions for each set of wild type and mutant.

Immunoprecipitation

Immunoprecipitation was performed with anti Ago3, anti Aub, or anti Piwi antibodies bound to protein G sepharose beads. Input, supernatant, and bound samples were subject to western blotting analysis to confirm immunoprecipitation.

Small RNA Cloning and Sequencing

Total RNA was isolated from manually dissected ovaries from 2 to 4 day old flies. After 2S rRNA depletion, 18–29 nt small RNA was purified and oxidized, followed by ethanol precipitation. 3′ ligated product was purified from a denaturing urea polyacrylamide gel and then ligated to a 5′ RNA adaptor. The small RNA library was amplified and then purified from an agarose gel. Purified libraries were sequenced with a Solexa Genome Analyzer (Illumina, San Diego, CA).

ACCESSION NUMBERS

Sequence data are available via the NCBI trace archives (http://www.ncbi.nlm.nih.gov/Traces/) with accession number SRP000458. Microarray data are available via the NCBI gene expression omnibus (http://www.ncbi.nlm.nih.gov/geo/) using accession number GSE14370.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Discussion, Supplemental Experimental Procedures, six tables, and 13 figures and can be found with this article online at http://www.cell.com/supplemental/S0092-8674(09)00452-8.

ACKNOWLEDGMENTS

We thank Mikiko and Haruhiko Siomi and Greg Hannon for antibodies, Erica Selva for flies expressing *nanos*-Gal4 from a transgene on the second chromosome, Paul Lasko for GFP-Vasa flies, Bradley Till for assistance with tilling

primer design, Alicia Boucher for assistance with fly husbandry, Gwen Farley for technical assistance, and members of the Zamore and Weng laboratories for advice, suggestions, and critical comments on the text. This work was supported in part by grants from the National Institutes of Health to P.D.Z. (GM62862 and GM65236) and W.E.T. (GM050898 and HD049116) and from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research to B.M.H. H.X. was supported in part by Pfizer, Inc.

Received: January 20, 2009
Revised: February 24, 2009
Accepted: April 10, 2009
Published online: April 23, 2009

REFERENCES

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., 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., Gaidatzis, D., Pfeffer, S., Lagos Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi Miyagawa, S., Nakano, T., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203 207.

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.

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., et al. (2008). PRG 1 and 21U RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol. Cell *31*, 67 78.

Bozzetti, M.P., Massari, S., Finelli, P., Meggio, F., Pinna, L.A., Boldyreff, B., Issinger, O.G., Palumbo, G., Ciriaco, C., Bonaccorsi, S., et al. (1995). The Ste locus, a component of the parasitic cry Ste system of *Drosophila melanogaster*, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. Proc. Natl. Acad. Sci. USA *92*, 6067 6071.

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.

Chambeyron, S., Popkova, A., Payen Groschene, G., Brun, C., Laouini, D., Pelisson, A., and Bucheton, A. (2008). piRNA mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. Proc. Natl. Acad. Sci. USA *105*, 14964 14969.

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.

Cooper, J.L., Till, B.J., and Henikoff, S. (2008). Fly TILL: Reverse genetics using a living point mutation resource. Fly (Austin) *2*, 300 302.

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.

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

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.

Findley, S.D., Tamanaha, M., Clegg, N.J., and Ruohola Baker, H. (2003). Maelstrom, a *Drosophila* spindle class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development *130*, 859 871.

Förstemann, K., Horwich, M.D., Wee, L. M., Tomari, Y., and Zamore, P.D. (2007). *Drosophila* microRNAs are sorted into functionally distinct Argonaute protein complexes after their production by Dicer 1. Cell *130*, 287 297.

Fuller, M.T. (1993). Spermatogenesis. In The Development of *Drosophila melanogaster*, M. Bate and A.A. Martinez, eds. (Plainview, N.Y: Cold Spring Harbor Laboratory Press), pp. 71 148.

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

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.

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

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.

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.

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

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69 82.

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.

Koundakjian, E.J., Cowan, D.M., Hardy, R.W., and Becker, A.H. (2004). The Zuker collection: a resource for the analysis of autosomal gene function in *Drosophila melanogaster*. Genetics *167*, 203 206.

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.

Liang, L., Diehl Jones, W., and Lasko, P. (1994). Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA binding and helicase activities. Development *120*, 1201 1211.

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. USA *104*, 6714 6719.

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.

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*, in press. 10.1016/j.cell.2009.03.040.

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.

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

Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of *Stellate* elements of *Drosophila melanogaster*. Genetics *138*, 1181 1197.

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.

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

Pelisson, A., Sarot, E., Payen Groschene, G., and Bucheton, A. (2007). A novel repeat associated small interfering RNA mediated silencing pathway downre gulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. J. Virol. *81*, 1951 1960.

Pimpinelli, S., Sullivan, W., Prout, M., and Sandler, L. (1985). On biological functions mapping to the heterochromatin of *Drosophila melanogaster*. Genetics *109*, 701 724.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. Genetics *139*, 697 711.

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.

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.

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

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.

Specchia, V., Benna, C., Mazzotta, G.M., Piccin, A., Zordan, M.A., Costa, R., and Bozzetti, M.P. (2008). *aubergine* gene overexpression in somatic tissues of *auberginesting* mutants interferes with the RNAi pathway of a *yellow* hairpin dsRNA in *Drosophila melanogaster*. Genetics *178*, 1271 1282.

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

Theurkauf, W.E. (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. Methods Cell Biol. *44*, 489 505.

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

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.

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

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

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

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

Cell

The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters

Carla Klattenhoff,[1] Hualin Xi,[3,7,8] Chengjian Li,[2] Soohyun Lee,[3,7] Jia Xu,[3,6] Jaspreet S. Khurana,[1] Fan Zhang,[1] Nadine Schultz,[1] Birgit S. Koppetsch,[1] Anetta Nowosielska,[1] Herve Seitz,[2,4,5] Phillip D. Zamore,[2,*] Zhiping Weng,[3,*] and William E. Theurkauf[1,*]

[1]Program in Molecular Medicine and Program in Cell Dynamics
[2]Department of Biochemistry and Molecular Pharmacology and Howard Hughes Medical Institute
[3]Program in Bioinformatics and Integrative Biology
University of Massachusetts Medical School, Worcester MA, 01605, USA
[4]Laboratoire de Biologie Moléculaire Eucaryote, Université de Toulouse, Université Paul Sabatier, F-31000 Toulouse, France
[5]Laboratoire de Biologie Moléculaire Eucaryote, Centre National de la Recherche Scientifique, F-31000 Toulouse, France
[6]Department of Biomedical Engineering
[7]Bioinformatics Program
Boston University, Boston, MA 02215, USA
[8]Computational Science Center of Emphasis, Pfizer Inc, 620 Memorial Drive, Cambridge, MA 02139, USA
*Correspondence: phillip.zamore@umassmed.edu (P.D.Z.), zhiping.weng@umassmed.edu (Z.W.), william.theurkauf@umassmed.edu (W.E.T.)
DOI 10.1016/j.cell.2009.07.014

SUMMARY

Piwi-interacting RNAs (piRNAs) silence transposons and maintain genome integrity during germline development. In *Drosophila*, transposon-rich heterochromatic clusters encode piRNAs either on both genomic strands (dual-strand clusters) or predominantly one genomic strand (uni-strand clusters). Primary piRNAs derived from these clusters are proposed to drive a ping-pong amplification cycle catalyzed by proteins that localize to the perinuclear nuage. We show that the HP1 homolog Rhino is required for nuage organization, transposon silencing, and ping-pong amplification of piRNAs. *rhi* mutations virtually eliminate piRNAs from the dual-strand clusters and block production of putative precursor RNAs from both strands of the major 42AB dual-strand cluster, but not of transcripts or piRNAs from the uni-strand clusters. Furthermore, Rhino protein associates with the 42AB dual-strand cluster,but does not bind to uni-strand cluster 2 or *flamenco*. Rhino thus appears to promote transcription of dual-strand clusters, leading to production of piRNAs that drive the ping-pong amplification cycle.

INTRODUCTION

Mutations in the *Drosophila* piwi-interacting RNA (piRNA) pathway disrupt transposon silencing, cause DNA break accumulation during female germline development, and lead to defects in posterior and dorsoventral axis specification (Brennecke et al., 2007; Chambeyron et al., 2008; Klattenhoff et al., 2007; Vagin et al., 2006). The axis specification defects associated with piRNA pathway mutations are dramatically suppressed by mutations in *mnk* and *mei-41*, which encode Chk2 and ATR kinase homologs that function in DNA damage signaling (Chen et al., 2007; Klattenhoff et al., 2007; Pane et al., 2007). The developmental defects linked to piRNA pathway mutations thus appear to be secondary to DNA damage, which may result from transposon mobilization. PIWI proteins bind piRNAs, and mutations in genes encoding mouse and Zebrafish *piwi* homologs lead to transposon overexpression and germline-specific apoptosis (Carmell et al., 2007; Houwing et al., 2007), which could be triggered by DNA damage. The piRNA pathway may therefore have a conserved function in transposon silencing and maintenance of germline genome integrity.

Drosophila piRNAs appear to be derived from transposon rich clusters, most of which are localized in pericentromeric and subtelomeric heterochromatin (Brennecke et al., 2007). The majority of clusters produce piRNAs from both genomic strands (dual-strand clusters). However, two major clusters on the X chromosome produce piRNAs predominantly from one genomic strand (uni-strand clusters) (Brennecke et al., 2007; Brennecke et al., 2008). One of these uni-strand clusters maps to *flamenco,* a locus required for transposon silencing in the somatic follicle cells (Brennecke et al., 2007; Mevel-Ninio et al., 2007; Pelisson et al., 2007; Pelisson et al., 1994; Prud'homme et al., 1995; Sarot et al., 2004). The *flamenco* cluster contains fragments of a number of transposons, including *Zam*, *idefix*, and *gypsy*, and *flamenco* mutations disrupt silencing of these transposons (Desset et al., 2008; Mevel-Ninio et al., 2007; Prud'homme et al., 1995). In addition, transgenes carrying fragments of transposons in this cluster show *flamenco*-dependent silencing (Sarot et al., 2004). These findings suggest that piRNAs encoded by *flamenco trans*-silence complementary transposons located outside this cluster (Brennecke et al., 2007).

The mechanism of *trans*-silencing by piRNA is not well understood. piRNA-PIWI protein complexes catalyze homology-dependent target cleavage, suggesting that target transposon mRNAs are cotranscriptionally or posttranscriptionally degraded (Gunawardane et al., 2007; Saito et al., 2006). However, several *Drosophila* piRNA pathway mutations have been reported to modify position effect variegation (PEV) (Brower-Toland et al., 2007; Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004), which is linked to spreading of transcriptionally silent heterochromatin from pericentric and telomeric regions (Girton and Johansen, 2008). Piwi protein also binds to heterochromatin in somatic cells, and interacts with Heterochromatin protein-1 (HP1) in yeast two-hybrid and immunoprecipitation assays (Brower-Toland et al., 2007). piRNA-Piwi protein complexes could therefore silence target transposons by directing assembly of heterochromatin-like domains. In fission yeast, which do not have piRNAs, small interfering RNAs (siRNAs) and Argonaute 1 (Ago1) appear to recognize nascent transcripts at the centromere, triggering both transcript destruction and HP1 recruitment and assembly of centromeric heterochromatin (Buhler et al., 2006; Verdel and Moazed, 2005). A similar combination of homology dependent cleavage and heterochromatin assembly could drive piRNA based silencing in the *Drosophila* germline.

The mechanism of piRNAs biogenesis also remains to be fully elucidated. Dicer endonucleases cleave double-stranded precursors to produce miRNAs and siRNAs (reviewed in Ghildiyal and Zamore, 2009), but piRNA production is Dicer independent (Houwing et al., 2007; Vagin et al., 2006). A subset of sense and antisense piRNAs overlap by 10 base pairs and show a strong bias toward an A at position 10 of the sense strand and a complementary U at the 5′ end of the antisense strand, suggesting that positions 1 and 10 base pair (Brennecke et al., 2007; Gunawardane et al., 2007). As Argonautes cleave their targets between positions 10 and 11 of the guide strand (Gunawardane et al., 2007; Saito et al., 2006), these finding suggest that piRNAs are produced by a ''ping-pong'' amplification cycle in which antisense strand piRNAs bound to Argonaute proteins cleave complementary RNAs to produce the 5′ end of sense piRNAs, which in turn direct a reciprocal reaction that generates the 5′ end of antisense strand piRNAs (Brennecke et al., 2007; Gunawardane et al., 2007). However, most piRNAs cannot be assigned to ping-pong pairs, some clusters produce piRNAs from only one strand (Brennecke et al., 2007), and the mechanism of 3′ end generation has not been determined. It is also unclear how ping-pong amplification is initiated, since the cycle depends on pre-existing primary piRNAs.

Here, we show that Rhino (Rhi), a member of the Heterochromatin Protein 1 (HP1) subfamily of chromo box proteins (Volpe et al., 2001), is required for transposon silencing, production of piRNAs by dual-strand heterochromatic clusters, and efficient ping-pong amplification. Significantly, Rhi protein associates with the 42AB dual-strand cluster and is required for production of longer RNAs from both strands of this cluster. Rhi thus appears to promote expression of trigger RNAs that are processed to from primary piRNAs that drive ping-pong amplification and transposon silencing. We also show that protein coding genes carrying transposons and transposon fragments within introns escape silencing, suggesting that piRNA silencing is imposed after RNA processing. Furthermore, *rhi* mutations disrupt nuage, a perinuclear structure that is enriched in piRNA pathway components. We therefore speculate that the nuage functions as a perinuclear surveillance machine that scans RNAs exiting the nucleus and destroys transcripts with piRNA complementarity.

RESULTS

Drosophila piRNA pathway mutations lead to germline DNA damage and disrupt axis specification through activation of Chk2 and ATR kinases, which function in DNA damage signaling (Chen et al., 2007; Cook et al., 2004; Pane et al., 2007). Mutations in the *rhi* locus lead to very similar patterning defects (Volpe et al., 2001). The *mei-41* and *mnk* genes encode ATR and Chk2, respectively (Brodsky et al., 2004; Hari et al., 1995). To determine whether the axis specification defects associated with *rhi* result from damage signaling, we generated double mutants with *mnk* and *mei-41* and quantified axis specification by scoring for assembly of dorsal appendages, which are egg shell structures that form in response to dorsal signaling during oocyte development (Table S1 available online). Only 17% (n = 700) of embryos from rhi^{KG}/rhi^{2} females had two wild-type appendages. However, 80% (n = 689) of embryos from $mnk;rhi^{KG}/rhi^{2}$ double-mutant females had two appendages (Table S1). In addition, 33% (n = 732) of embryos from $mei\text{-}41;rhi^{KG}/rhi^{2}$ double-mutant females had two appendages (Table S1). Consistent with these observations, *rhi* mutations disrupt dorsal localization of Gurken and posterior localization of Vasa in the oocyte, and localization of both proteins is restored in $mnk;\ rhi^{KG}/rhi^{2}$ double mutants (Figure 1).

Both ATM and ATR kinases have been reported to activate Chk2 (Wang et al., 2006). Mutations in the *Drosophila atm* gene are lethal, but caffeine inhibits ATM and to a lesser extent ATR (Sarkaria et al., 1999). Strikingly, 88% (n = 473) of embryos from *rhi* mutant mothers fed caffeine had wild-type dorsal appendages (Table S1). Similarly, only 2% (n = 277) of embryos from *armi* mutant females had two dorsal appendages, compared with 11% (n = 477) after caffeine treatment (Table S1). In addition, 56% (n = 575) of embryos from $mei^{41D3}/mei^{41D3};\ armi^{72.1}/armi^{1}$ females had wild-type appendages, but 83% (n = 226) of embryos from $mei^{41D3}/mei^{41D3};\ armi^{72.1}/armi^{1}$ double mutants fed with caffeine had two appendages (Table S1). Caffeine combined with *mei-41* mutations thus leads to levels of suppression that are similar to *mnk* single mutations, suggesting that ATM and ATR redundantly activate Chk2 in *armi* and *rhi* mutants.

The *mei-W68* locus encodes the *Drosophila* Spo11 homolog, which is required for meiotic double-strand break formation (McKim and Hayashi-Hagihara, 1998). However, *mei-W68* mutations fail to suppress the dorsal appendage defects associated with *rhi* (Table S1), indicating that DNA damage signaling in *rhi* mutants is not due to defects in meiotic break repair.

The phosphorylated form of the *Drosophila* histone H2AX (γ H2Av) accumulates near DNA double-strand break sites (Gong et al., 2005; Modesti and Kanaar, 2001; Redon et al., 2002). In wild-type ovaries, γ-H2Av foci are generally restricted to region 2 of the germarium, where meiotic double-strand breaks are formed (Figures 1D–1F) (Jang et al., 2003). As the

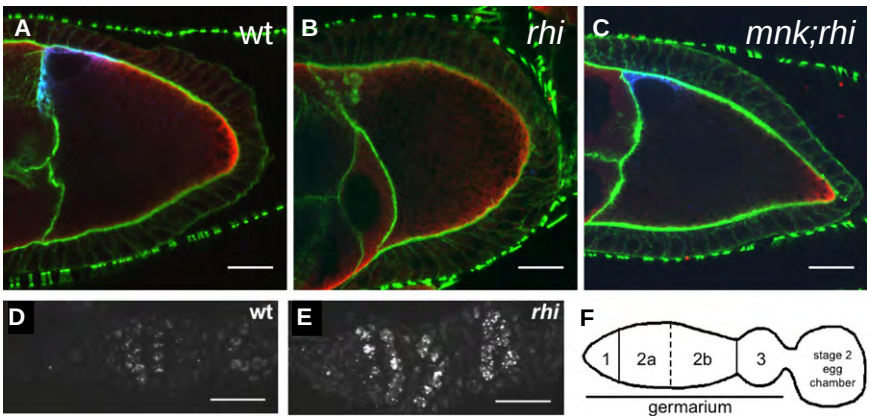


Figure 1. DNA Damage Signaling in *rhi* Mutants

(A C) Mutations in *mnk*, which encodes the DNA damage signaling kinase Chk2, suppress the Gurken and Vasa protein localization defects in *rhi* mutants.

(A) In a stage 9 wild type oocyte, Grk (blue) is localized at the dorsal anterior cortex near the oocyte nucleus and Vas (red) is localized at the posterior cortex. Actin filaments (green) mark the cell boundaries.

(B) In *rhi* egg chambers, this localization pattern is lost, with Grk and Vas dispersed throughout the oocyte.

(C) *mnk* suppresses the *rhi* phenotype, and rescues Grk and Vas localization during late oogenesis. Images were acquired under identical conditions. Projections of two serial 0.6 μm optical sections are shown. Scale bars represent 20 μm.

(D F) *rhi* mutants have increased DNA damage in the germline.

(D) Foci of γ H2Av are observed in wild type ovaries in region 2a and 2b of the germarium and correspond to the DSBs induced during meiotic recombination.

(E) In *rhi* mutants, much larger foci also appear in region 2a of the germarium but persist in region 3 and the developing egg chambers. Samples were labeled and images were acquired under identical conditions. Projections of 5 serial 1μm optical sections are shown. Posterior is oriented to the right. Scale bars represent 10 μm.

(F) A schematic representation of the regions of the germarium and a developing egg chamber.

cysts mature and pass through region 3 of the germarium, γ-H2Av labeling is reduced. Stage 2 egg chambers, which bud from the germarium, show only low levels of γ-H2Av labeling. In *rhi* mutants, prominent γ-H2Av foci are present in germline cells of the germarium, and these foci persist and increase in intensity as cysts mature and bud to form stage 2 egg chambers (Figures 1D and 1E). *rhi* mutations thus appear to trigger germline-specific DNA breaks and damage signaling through ATM, ATR, and Chk2.

Transposon Silencing and Gene Expression

The piRNA pathway is required for transposon silencing in the *Drosophila* female germline (Vagin et al., 2006) but has also been implicated in heterochromatic gene silencing in somatic cells (Brower-Toland et al., 2007; Pal-Bhadra et al., 2002; Pal-Bhadra et al., 2004). We therefore assayed both transposon and protein-coding gene expression using whole-genome tiling arrays (Figure 2). In both *rhi* and armi mutants, most transposon families show a relatively modest 1.5- to 2-fold increase in expression, which is not statistically significant (false discovery rate [FDR] > 0.02). However, a subset of transposon families are dramatically overexpressed in both *rhi* and *armi* mutants (Figures 2B and 2C; blue points indicate FDR < 0.02). For example, *HeT-A* expression increased 70-fold in *rhino* and 117-fold in *armi* (Table S2). In total, 15 of 17 transposon families that are significantly overexpressed in *rhi* are also overexpressed in *armi* (Figure S1). 11 families are overexpressed with an FDR < 0.02 in *armi* mutants, but not in *rhi* (Figure S1). Rhino thus appears to silence a subset of the transposons silenced by Armi. This could reflect a role for Armi in transposon silencing in both somatic follicle cells and the germline (Klattenhoff et al., 2007), while Rhi appears to be restricted to the germline (see below).

Both *rhi* and *armi* mutations increased expression of long terminal repeat (LTR) elements, non-LTR retrotransposons, and inverted repeat (IR) elements (Figure S2) (Vagin et al., 2006), Similar patterns of transposon overexpression are observed in *aub* and *ago3* mutants, which disrupt piRNA biogenesis (Li et al., 2009). Mutations in established piRNA pathway genes and in the *rhino* locus thus disrupt transposon silencing, independent of transposition mechanism.

piRNAs from the *suppressor of stellate* locus silence the *Stellate* gene during male germline development, and Stellate protein overexpression leads to Stellate crystal formation during spermatogenesis (Aravin et al., 2001; Bozzetti et al., 1995; Livak, 1984, 1990; Palumbo et al., 1994). However, *rhi* mutations do not lead to Stellate crystal formation or compromise male fertility (Figure S3 and data not shown).

HP1 and several genes in the piRNA pathway have been implicated in position effect variegation, which is linked to spreading of heterochromatin from centromeric and telomeric regions (Elgin and Grewal, 2003; Pal-Bhadra et al., 2004). However, neither *rhino* nor *armi* led to statistically significant changes in the expression of any protein coding genes, including the 613 annotated heterochromatic genes (Smith et al., 2007) (Figures 2D and 2E; green points indicate heterochromatic genes). piRNA pathway and *rhi* mutations thus do not produce changes in heterochromatin organization sufficient to alter protein coding gene expression during oogenesis.

Figure 2A shows a genome browser view of the region containing the heterochromatic gene *jing*. Expression of exons that comprise the mature *jing* transcript are essentially identical in w^{1118} and *rhi*, but expression of a *flea* transposon located in a major intron increases 7-fold (FDR < 0.02), and several transposons in the intergenic regions near *jing* are also overexpressed (Figure 2A, *rhino,* pink bars). The repeated nature of natural transposons and the design of the arrays makes it impossible to determine which specific transposon copy or copies are overexpressed, but we can conclude that at least one member of the

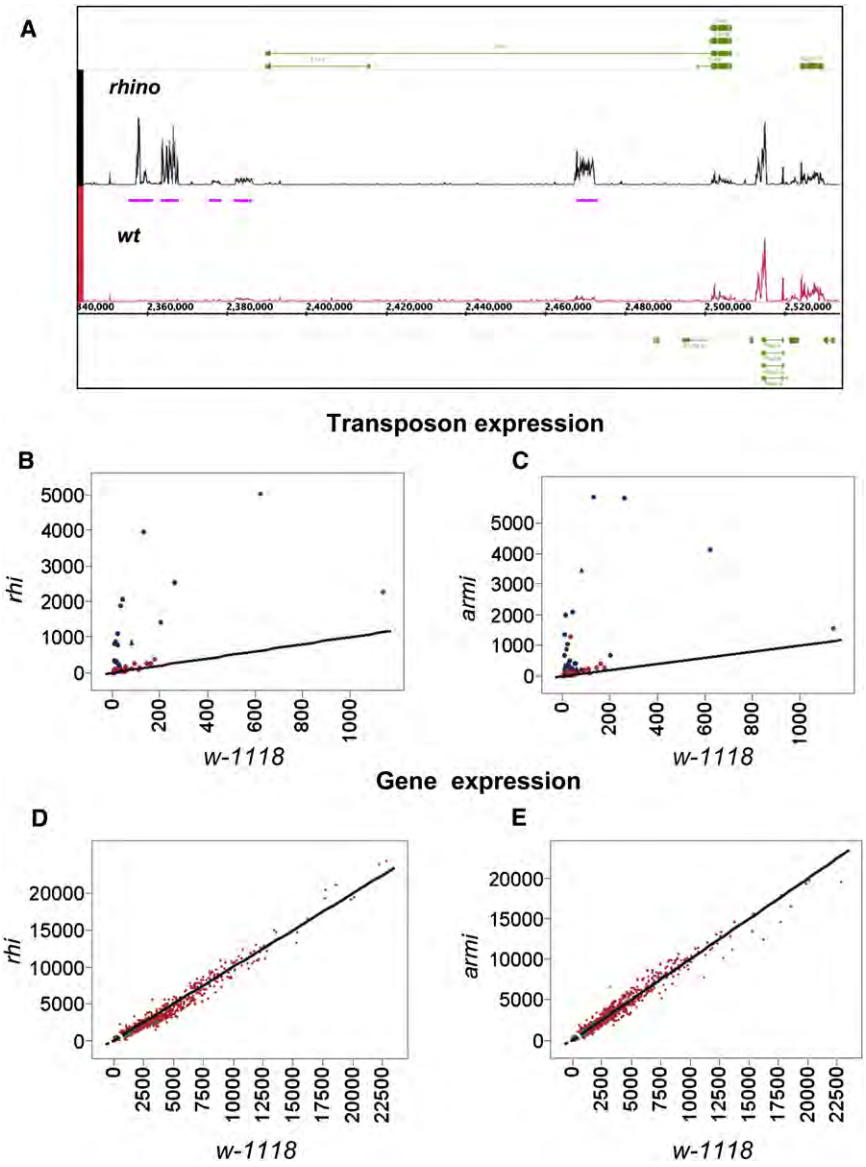


Figure 2. Gene and Transposon Expression in *rhi* Mutant Ovaries

(A) Genome browser view of tiling array data near *jing*, a protein coding gene in pericentromeric heterochromatin on chromosome 2R. Expression of *jing* exons (green bars) is unaltered by *rhi* mutants. However, several intronic and extragenic transposons are significantly overexpressed (pink bars).

(B and C) Genome wide analysis of transposon family expression in *rhi* and *armi* mutants. Tiling arrays were used to quantify expression of 95 transposon families in *rhi*, *armi*, and *w 1118* controls. Graphs show expression in *rhi* and *armi* plotted against expression in *w 1118*. The lines intercept the origin and have a slope of 1, and thus indicate equal expression in both genotypes. Significantly overexpressed transposon families are indicated by blue data points.

(D and E) Genome wide comparison of protein coding gene expression in *rhi* and *armi* mutants, plotted against expression in *w1118*. Heterochromatic genes are indicated by green data points, and euchromatic genes are indicated by red data point. Both classes cluster around the diagonal, indicating similar expression levels in mutant and controls.

transposon family is overexpressed. Over 1300 protein coding genes carry transposon insertions within introns, and thus have primary transcripts that could base pair with piRNAs. This includes *ago3,* which encodes an Arognaute protein that is expressed in the female germline and is required for ping-pong amplification of piRNAs (Li et al., 2009). Our array studies show that expression of *ago3,* and the other protein coding genes carrying intronic transposon insertions, does not significantly change in *rhi* or *armi* mutants (Figures 2D and 2E and data not shown). These observations suggest that piRNA-dependent silencing may be imposed after splicing, which removed transposon homology from protein coding genes.

Rhi Localization

To define the subcellular distribution of Rhi, we generated a GFP-*rhi* transgene and raised anti-Rhi antibodies, which were used to localize the protein in vivo and immunolabel whole-mount egg chambers. Both methods revealed germline-specific nuclear foci that are present throughout oogenesis (Figures 3A–3D). In addition, germline-specific expression of the GFP-Rhi fusion protein rescued fertility and axial patterning in *rhi* mutations (Table S1). Rhi thus appears to function specifically within the germline cells of the ovary.

To determine whether Rhi foci are associated with centromeres, we labeled for Rhi and CID, the *Drosophila* homolog of the centromere-specific, histone H3-like CENP-A (Blower and Karpen, 2001). Rhi accumulated in regions adjacent to most CID foci in germline nuclei, consistent with localization to pericentromeric heterochromatin (Figures 3E–3G). However, many Rhi foci were not obviously linked to CID. Some of these foci could be linked to telomeres or other chromatin domains. Resolving this question will require higher-resolution molecular approaches.

To determine whether Rhi localization depends on the piRNA pathway, we immunolabeled egg chambers mutant for *aub* and *armi*. Rhino localization to nuclear foci was not disrupted by either mutation (Figures 3A–3C). In striking contrast, *rhi* mutations disrupt localization of Aub and Ago3 to nuage, a perinuclear structure implicated in RNA processing (Figures 3H, 3I, 3L, and 3M). Vasa is a core component of nuage, and perinuclear localization of Vasa was also lost in *rhi* mutants (Figure S4). Piwi localizes to nuclei in both germline cells and the somatic follicle cells. In wild-type ovaries, Piwi is most abundant in

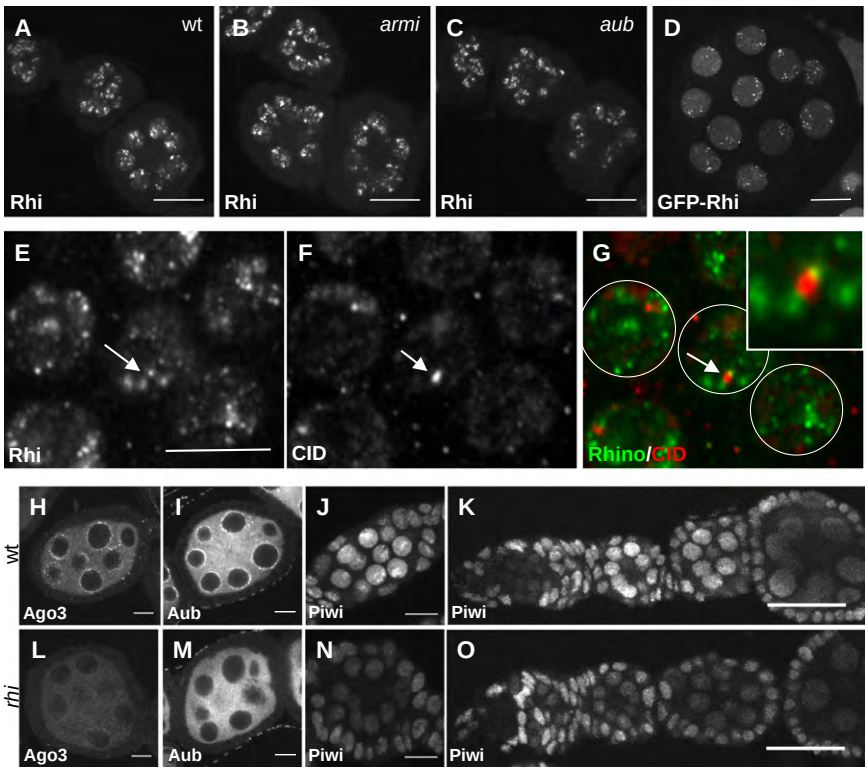


Figure 3. Rhi Localization Is Independent of the piRNA Pathway, but Localization of the PIWI Proteins Ago3 and Aub Requires Rhi

(A C) Rhi localization appears similar in wild type (A), *armi* (B), and *aub* (C) stage 2 to stage 4 egg chambers. Projections of five serial 1 μm optical sections are shown. Scale bars represent 20 μm. (D) GFP Rhi transgene shows localization pattern similar to endogenous Rhi detected with anti Rhi antiserum in the germline nuclei of stage 4 5 egg chambers. The scale bar represents 10 μm.

(E G) Wild type ovaries immunostained with (E and G) anti Rhi antiserum and (F and G) anti CID antibody show that some Rhi foci localize adjacent to CID foci (arrows) consistent with binding to peri centromeric heterochromatin in some but not all chromosomes. The scale bar represents 5 μm.

(H O) *rhi* mutation disrupts localization of PIWI class Argonautes. Stage 4 5 (H, I, L, and M), stage 2 3 (J and N), and germarium to stage 3 4 egg chambers (K and O) of wild type and *rhi* mutant ovaries were immunostained with corresponding antibodies. Projections of three serial 1 μm optical sections are shown. Scale bars represent 10 μm (H, I, J, L, M, and N) and 20 μm (J, K, and O). Wild type localization of Ago3 and Aub proteins to perinuclear nuage is disrupted in *rhi* mutants. Piwi protein localizes to the nuclei of both germline and somatic cells in wild type egg chambers. Only germline nuclear localization of Piwi in early stages is disrupted by mutations in *rhi*.

germline nuclei during early stages of oogenesis (Figure 3K). In *rhi* mutants, nuclear localization of Piwi is reduced during these early stages (Figures 3N and 3O). However, in later-stage egg chambers, which make up the bulk of the ovary, Piwi localization in *rhi* is similar to wild-type controls (Figures 3J, 3K, 3N, and 3O). These findings suggest *rhi* functions upstream of Ago3 and Aub, but may have a less critical role in Piwi-dependent processes.

piRNA Expression Is Ablated for Most Transposon Families in *rhino* Mutants

To determine whether Rhi is required for piRNA expression, we sequenced small RNAs from control and *rhi* mutant ovaries. Unlike miRNAs, piRNAs carry 2′ methoxy, 3′ hydroxy termini that render them resistant to oxidation and stabilize these RNAs in vivo (Vagin et al., 2006). To enrich for piRNAs and increase effective sequencing depth, we oxidized RNA samples prior to library construction and sequencing and normalized the data to surviving noncoding RNA fragments (Ghildiyal et al., 2008; Seitz et al., 2008) (see Table S3 for sequencing statistics). These studies indicate that *rhi* mutations reduce total piRNA abundance by approximately 80% (Figures 4A and 4B). Northern blotting for specific piRNAs and miRNAs support these findings (Figure S5). Defects in 3′ modification destabilize piRNAs and would lead to preferential loss of piRNAs in oxidized samples. We therefore deep sequenced unoxidized RNAs and normalized piRNA abundance to miRNAs. These studies confirm that *rhi* mutations reduce piRNA abundance by 80%, and indicate that this reduction does not result from a defect in end modification (data not shown).

The majority of *Drosophila* piRNAs are derived from transposons and other repeated elements (Aravin et al., 2003; Brennecke et al., 2007). We analyzed the impact of *rhi* mutations on piRNA expression from 95 families with at least 500 matching reads in control samples (Table S3) (Li et al., 2009). *rhi* mutations lead to a 50% or greater reduction in antisense piRNA abundance for 83% of these transposon families, and a 98% reduction in antisense piRNAs for approximately 30% of these elements (Figure S7). For 66 of 95 families, both sense and antisense piRNAs are reduced. For example, *rhi* mutations nearly eliminate sense and antisense piRNAs from the telomeric transposon *HeT-A* (Figure 5A). Eight transposon families continue to express at least 50% of wild-type sense strand piRNAs but show an 80% or greater reduction in antisense piRNAs. The *jockey* element falls into this class. Mutations in *rhi* reduce sense strand piRNAs linked to *jockey* by only 10%, but antisense strand piRNAs are reduced by 95% (Figure 4B, *jockey*). For all of the transposon families that show reduced antisense piRNAs, including those that retain sense strand piRNAs, there is a clear reduction in opposite strand piRNAs that overlap by 10 nt, consistent with defects in ping-pong amplification (Figures 5Ac and 5Bc). A comparison of the p values for the 10 nt overlap bias across all transposon families confirms that the loss of ping-pong pairs in *rhi* is very highly significant (Figure S6, p = 3e-10). The loss of species that overlap by 10nt is also clear from an analysis of total piRNAs (Figures 4C and 4D). The *rhi* mutations thus lead to a near collapse of the ping-pong cycle amplification cycle.

Only 10 of 95 transposon families continue to express antisense piRNAs at or above 75% of wild-type levels in *rhi* mutants

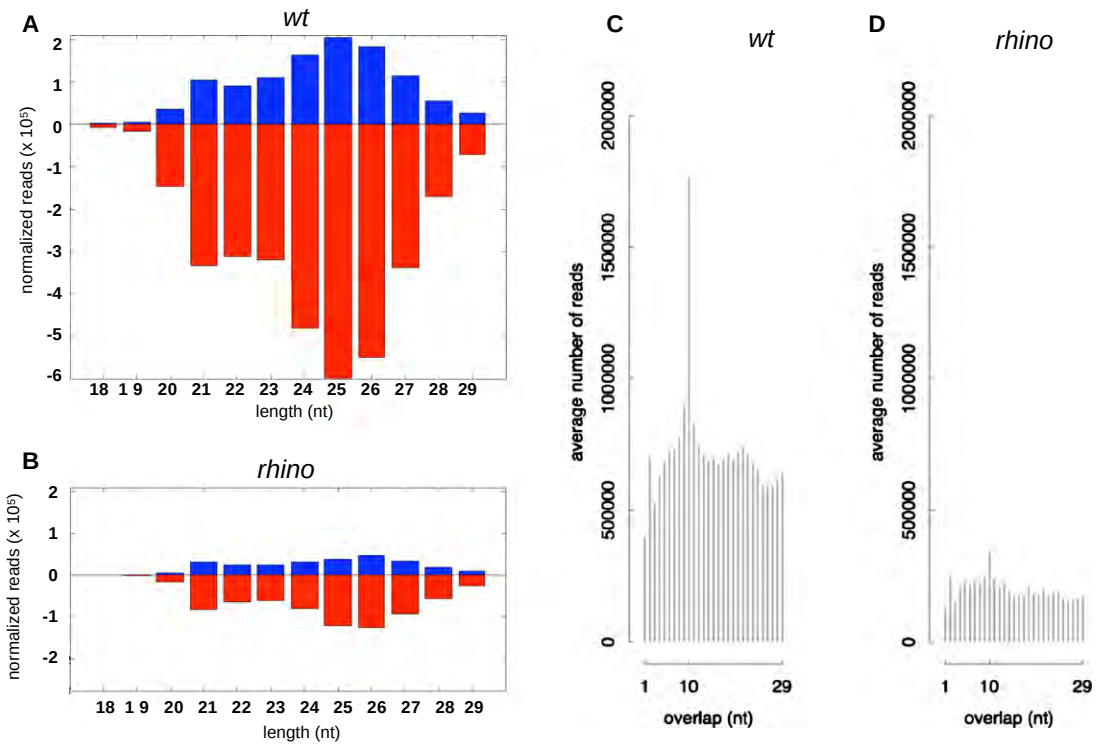


Figure 4. piRNA Production in *rhi* Mutants

(A) Length histogram of piRNAs expressed in wild type ovaries.

(B) Length histogram of piRNAs produced in *rhi* mutants. Sense and antisense piRNAs are reduced by approximately 80%, and peak length shifts from 25 nt to 26 nt.

(C) Histogram of overlapping sense and antisense piRNA in wild type ovaries, showing a pronounced peak at 10 nt, characteristic of ping pong amplification.

(D) Histogram of overlapping sense and antisense piRNA in *rhi* mutant ovaries. The 10 nt peak is nearly eliminated, suggesting a breakdown in the ping pong amplification cycle.

(*blood*, *mdg-1*, *Tabor*, *Stalker*, *Stalker 2*, *Stalker3*, *Stalker4*, *412*, *297*, *gypsy 5*; Table S4). Eight of these families (*blood*, *mdg-1*, *Tabor*, *Stalker*, *Stalker 2*, *Stalker3*, *Stalker4*, *412*) also show an increase in sense strand piRNAs (Figures 5Ca and S8). The sense strand piRNAs generally map to the same regions as peaks of antisense piRNAs (Figure 5Ca, *blood*; Figure S8). This pattern could indicate that antisense strand piRNA direct production of the sense strand piRNAs. Alternatively, specific regions within full-length elements or fragments of elements that lie within specific clusters may be preferentially utilized during piRNA production. The available data cannot distinguish between these alternatives.

An analysis of piRNAs encoded by the ten transposon families that show Rhi-independent piRNA production revealed three patterns with respect to overlapping sense and antisense species. The overlapping piRNAs encoded by *Stalker3* did not show a statistically significant ($p > 0.001$) 10 nt overlap bias in either wild-type or *rhi* mutants, indicating that their production is independent of ping-pong amplification. However, six families showed a statistically significant 10 nt overlap peak in both wild-type and *rhi* mutants, indicating that at least some of the piRNAs are produced by a ping-pong cycle that is independent of Rhi (*Tabor*, *Stalker*, *Stalker 2*, *Stalker4*, *412*, *297*; Figure S8). The final class of elements includes *blood*, *mdg1*, and *gypsy5*, which show a statistically significant ping-pong peak in wild-type, but loose the 10 nt overlap bias in *rhi* mutants (Figure 5Cc, *blood*; Figure S8). For this class, Rhi thus appears to promote production of only a subset of piRNAs through ping-pong amplification. Intriguingly, *rhi* leads to a 10-fold increase in *blood* expression, suggesting the minor ping-pong pool of piRNAs may be critical to transposon silencing (Figure S9).

Overlapping ping-pong pairs show transposon family-specific nucleotide biases at positions 1 and 10 that appear to reflect the specific PIWI proteins that participate in the amplification cycle (Brennecke et al., 2007; Gunawardane et al., 2007). For example, elements for which sense strand piRNAs are primarily bound by Ago3 and antisense strand piRNAs are primarily bound by Aub, show an A bias at position 10 of the sense strand and a U bias at position 1 of the antisense strand (Brennecke et al., 2007; Gunawardane et al., 2007). Families that retain a statistically significant ping-pong peak generally retain the pattern of nucleotide bias observed in the wild-type (Figure S8), suggesting that *rhi* reduces the efficiency of the ping-pong amplification but does not alter the specific PIWI proteins that participate in the cycle.

Antisense piRNAs can base pair with target RNAs and guide cleavage by PIWI proteins and are therefore presumed to be the effectors of transposon silencing. To determine whether loss of antisense piRNAs in *rhi* mutants correlates with loss of

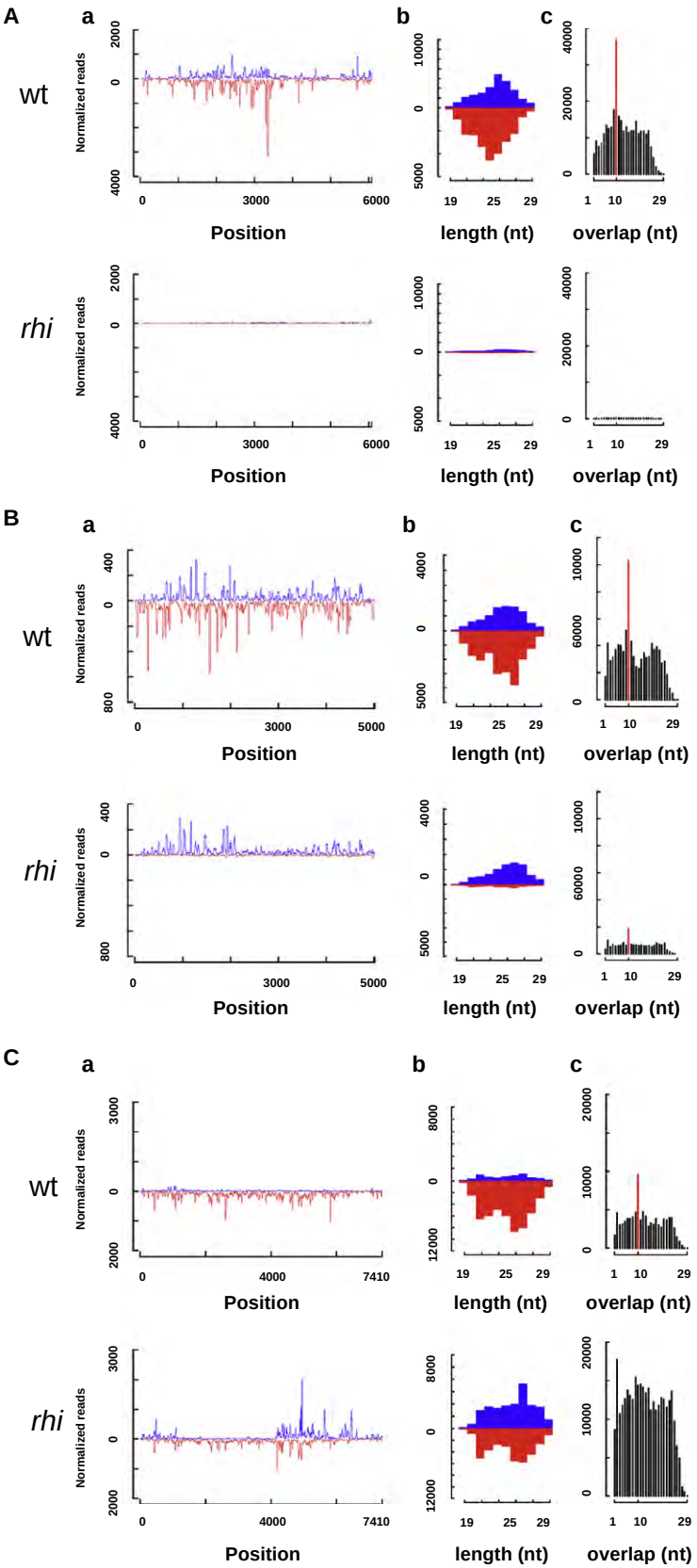


Figure 5. Transposon Specific Changes in piRNA Abundance

(A) Sense and antisense piRNA reads mapping to the consensus *Het A* sequence (Aa), length histograms for all *HetA* piRNAs (Ab), and frequency distribution of overlapping *HetA* piRNA (Ac). (B) Sense and antisense piRNAs reads mapping to the consensus *jockey* sequence (Ba), length histograms for all *jockey* piRNAs (Bb), and frequency distribution of overlapping *jockey* piRNAs (Bc). (C) Sense and antisense piRNA reads mapping to the consensus *blood* sequence (Ca), length histograms for all *blood* piRNAs (Cb), and frequency distribution of overlapping *blood* piRNA (Cc). For the majority of transposons, including *HetA*, *rhi* mutations dramatically reduce sense and antisense piRNAs and nearly eliminate piRNA that overlap by 10 nt. For a subset of elements, represented by *jockey*, *rhi* leads to a loss of antisense piRNAs but no significant reduction in sense strand piRNAs. A very limited number of transposons, including *blood*, show no change or an increase in sense strand piRNAs in *rhi* mutants. Mutations in *rhi* reduce piRNAs with a 10 nt overlap, even for elements that show an increase in piRNAs from opposite strands (Cc, *blood*). In (Ac), (Bc) and (Cc), A statistically significant 10 nt bias is indicated by a red bar. This is characteristic of the ping pong amplification cycle.

silencing, we plotted the fold change in transposon expression (*rhi*/wt) against the fold-change in antisense piRNAs (Figure S9). All of the transposon families that increased in expression by 20-fold or greater in *rhi* mutants also showed a 75% or greater reduction in antisense piRNA abundance. In addition, none of the families that retained antisense piRNA expression at 80% or higher levels were significantly overexpressed (FRD < 0.02; Figure S9). However, many transposon families that show a reduction in antisense piRNAs abundance of over 10-fold did not show a statistically significant increase in expression (Figure S9). These elements may be silenced by a piRNA independent mechanism. Alternatively, piRNAs linked to these elements could silence these elements, perhaps by inhibiting translation, without altering target transcript stability.

piRNA Clusters

The majority of piRNAs match transposons that are present in multiple copies in the genome and cannot be uniquely mapped. However, piRNAs encoded by polymorphic transposons, divergent transposon fragments, or other unique sequences can be mapped. Chromosome profiles of these "unique mappers" reveal dispersed piRNA peaks in the euchromatic chromosome arms and a limited number of prominent pericentromeric and subtelomeric clusters, which appear to be the source of the majority of piRNAs (Brennecke et al., 2007). *rhi* mutations essentially eliminate piRNAs mapping to pericentromeric heterochromatin on all of the autosomes (Figures S10 and 6). By contrast, piRNAs mapping to the pericentromeric region on the X are retained (Figures S10 and 6).

Most heterochromatic clusters produce piRNAs from both the plus and minus genomic strands (dual-strand clusters), but two major pericentromeric clusters on the X chromosome produce piRNA almost exclusively from one strand (uni-strand clusters) (Brennecke et al., 2007). We find that *rhi* reduces by 30- to 50-fold piRNAs from both strands of the top 11 dual-strand clusters (Figure S11, blue bars). For example, piRNA production from cluster 1/42AB, which is estimated to produce up to 30% of all piRNAs (Brennecke et al., 2007), is reduced by over 97% (Figures 6A–6C). In striking contrast, piRNAs encoded by the uni-strand clusters are only minimally impacted by *rhi* (Figure S11, red bars). As shown in Figures 6B and 6D, piRNAs from cluster 2 are derived almost exclusively from one strand, and production of these piRNAs is nearly unchanged in *rhi* mutants (Figures 6B and 6D). This does not appear to reflect expression of cluster 2 piRNAs exclusively in the somatic follicle cells, since Ago3 is germline specific and *ago3* mutations reduce piRNAs linked to this locus by close to 20-fold (Figure S12). In addition, unique piRNAs mapping to this cluster immunoprecipitate with the germline specific PIWI proteins Aub and Ago3 (Figure S12). Both dual-strand and uni-strand clusters thus appear to be expressed in the germline, but *rhi* mutations only disrupt piRNA production by the dual-strand clusters. Consistent with these findings, nine of the ten transposon families that continue to express high levels of antisense piRNAs in *rhi* (75% of wild-type or greater) have insertions in one or both of the major uni-strand clusters (Table S5).

To determine whether Rhino protein associates with clusters, we performed chromatin immunoprecipitation (ChIP) using anti-GFP antibodies and flies expressing a functional Rhino-GFP transgene specifically in the germline. To control for nonspecific binding, precipitation was performed using nonimmune IgG on chromatin from ovaries expressing the GFP-rhino transgene. Additionally, anti-GFP antibodies were used on chromatin isolated from wild-type flies that do not express the GFP fusion. Quantitiative PCR (qPCR) assays showed only background signal in both of these control reactions (data not shown). The anti-GFP fractions were assayed for three regions of cluster 1/42AB, two regions of cluster 2, two regions in the heterochromatic protein coding genes *jing* and *pld* that flank the 42A/B cluster, and the euchromatic protein coding genes *rp49* and *ry* (Figure 6G). GFP-Rhi binding, measured as a fraction of input chromatin, was enriched at all three sites in the dual-strand cluster relative to the euchromatic protein coding genes (Figure 6F). By contrast, the two sites in uni-strand cluster 2 showed no enrichment relative to *rp49* or *ry* controls (Figure 6F). Regions in the two heterochromatic genes immediately flanking the 42A/B cluster showed binding that was 3- to 4-fold lower than the peak region in the cluster (1A) and appoximately 3-fold higher than binding to euchromatic genes. Rhino thus appears to be enriched at dual-strand heterochromatic clusters and may spread somewhat beyond the computationally defined limits of these clusters.

Cluster Transcription

The piRNA clusters are proposed to produce long precursor RNAs that are processed to form primary piRNAs, which in turn trigger the ping-pong amplification cycle by targeting sense strand transposon transcripts. To determine whether Rhino is required for RNA production by clusters, we used quantitative reverse transcriptase (RT)-PCR to assay RNAs derived from both strands of cluster 1/42AB, cluster 2, and *flam*. Reactions without RT produced no significant signal, and the low level of signal obtained in the absence of the strand specific RT primers was subtracted from the signal obtained with the strand specific primers. Consistent with production of piRNAs from both genomic strands, we detected longer RNAs from both strands at two independent locations in cluster 1/42AB (Figure 6E, blue bars). Significantly, RNAs from both strands were nearly eliminated in *rhi* homozygous mutants (Figure 6E, red bars). At cluster 1 and *flam,* which produce piRNA almost exclusively form the plus strand (Figure 6D) (Brennecke et al., 2007), RT-PCR detected RNA from only the plus strands (Figure 6F, blue bars). In striking contrast to cluster 1/42AB, *rhi* mutations led to a slight increase in plus strand transcript from cluster 2 and only a modest decrease in plus strand RNA from *flam.* These observations suggest that Rhino promotes production of precursor RNAs from dual-strand cluster 1/42AB, and possibly all dual-strand clusters.

Brennecke et al. (2007) proposed that antisense piRNAs derived from the clusters initiate ping-pong amplification by cleaving sense strand transcripts from target transposons. However, unique piRNAs derived from opposite strands of cluster1/42AB show a strong 10 nt overlap bias (Figure S13), indicating that they are produced by ping-pong processing of precursor RNAs derived from the cluster. Antisense piRNAs derived through cluster based ping-pong amplification thus

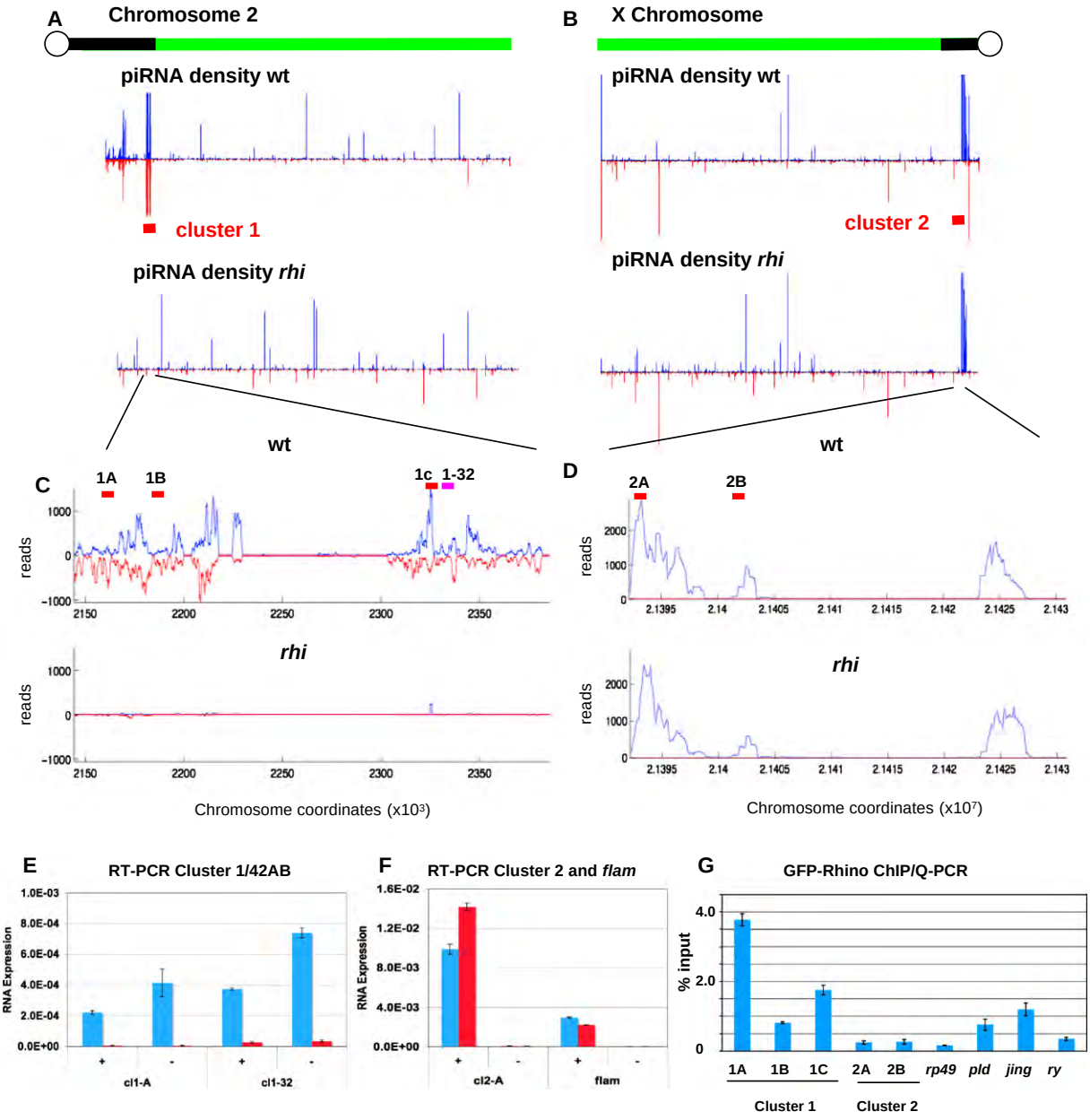


Figure 6. Rhino Is Required for piRNA Production by Dual Strand Heterochromatic Clusters

(A and B) Chromosome 2R and X density profiles of uniquely mapping plus (blue) and minus (red) strand piRNAs in wild type and *rhi* mutants. piRNAs map to dispersed loci on the chromosome arms and prominent heterochromatic clusters. Pericentromeric piRNAs from chromosome 2R (A), and all other autosomes (Figure S10) are dramatically reduced in *rhi* mutants. Pericentromeric piRNAs on the X chromosome show relatively little change.

(C and D) Higher resolution maps of clusters 1 and 2, which map to the indicated regions on 2R and X, respectively. Mutations in *rhi* nearly eliminated piRNAs encoded by cluster 1, which is the major dual strand cluster, but have little impact on piRNAs from cluster 2, which is the major uni strand cluster.

(E and F) Quantitative strand specific RT PCR for RNA derived from dual strand cluster 1 (E) and uni strand clusters 2 and *flam* (F). In wild type ovaries, RNA is detected from both the plus (+) and minus ( ) strands of cluster 1, at two independent locations (F, cl1 A and cl1 32, blue bars). RNAs from both strands of cluster 1 are dramatically reduced in *rhi* mutants (red bars). Significant levels of RNA are only detected from the plus strand of cluster 2 and *flam* (F, blue bars), and *rhi* does not block expression of these RNAs (F, red bars).

(G) Chromatin immunoprecipitation/quantitative PCR analysis of Rhino binding to cluster 1/42AB, the euchromatic genes *rp49* and *ry*, and the heterochromatic genes *pld* and *jing*, which flank cluser 1. Rhino protein is highly enriched at cluster 1 relative to cluster 2 and the euchromatic genes. The protein coding genes flanking cluster 1 show intermediate levels of binding, suggesting that Rhino may spread to regions flanking the dual strand clusters. Anti GFP antibodies were used to precipitate Rhino GFP from crosslinked ovary chromatin fractions.

The approximate positions of the qPCR primer pairs used in ChIP and RT PCR reactions are indicated by the red bars in (C) and (D). The pink bar in (C) indicates the approximate position of an additional primer pair used in RT PCR.

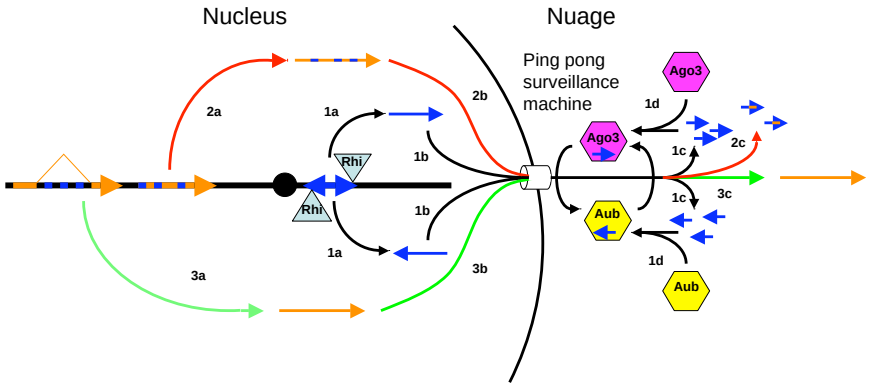


Figure 7. Model for Rhino Dependent Transposon Silencing

Step 1 (black arrows): Rhino binds to dual strand clusters and promotes production of RNAs from both genomic strands (1a), which are exported from the nucleus (1b) and processed into piRNAs by a ping pong cycle driven by Ago3 (pink hexagon) and Aub (yellow hexagon), localized to the perinuclear nuage (1c).

Step 2 (red arrows): Transposons carrying piRNA homology are transcribed (2a), exported from the nucleus (2b), and degraded as they encounter the "ping pong surveillance machine" within the nuage (2c).

Step 3 (green arrows): Protein coding genes with intronic transposon insertions are spliced, which removes piRNA homology (3a). These transcripts are exported from the nucleus (3b) escape recogni tion by the surveillance system, and are translated (3c). Sequences matching piRNAs are indicated by blue. Other transcribed regions are in orange.

appear to target sense strand RNAs derived from functional transposons located throughout the genome.

DISCUSSION

piRNAs encoded by transposon-rich heterochromatic clusters have been proposed to initiate a ping-pong cycle that amplifies the piRNA pool and mediates transposon silencing (Brennecke et al., 2007; Lin, 2007; O'Donnell and Boeke, 2007). However, the mechanisms of piRNA biogenesis and silencing are not well understood, and it is unclear how the piRNA clusters are differentiated from other chromatin domains. We show that the HP1 homolog Rhino is required for production of piRNAs from dual-strand clusters and associates with the major 42AB cluster by ChIP. Significantly, we also identify putative piRNA precursor RNAs from both strands of the 42AB cluster and show that Rhino is required for production of these RNAs. These findings lead us to propose that Rhi binding promotes transcription of dual-strand clusters, and that the resulting RNAs are processed to form primary piRNAs that drive the ping-pong amplification cycle and transposon silencing (Figure 7, black pathway).

While Rhino protein appears to be restricted to germline nuclei, *rhi* mutations disrupt perinuclear localization of Ago3 and Aub (Figure 3), which catalyze the ping-pong amplification cycle (Li et al., 2009). Mutations in *krimper*, which encodes a component of the perinuclear nuage, also disrupt transposon silencing and piRNA production (Lim and Kai, 2007). piRNA silencing and nuage assembly thus appear to be codependent processes. These observations, with our finding that protein coding genes carrying piRNA homology within introns escape silencing by the piRNA pathway (Figure 2A), suggest that transcripts are scanned for piRNA homology within the nuage, after splicing and nuclear export. Mature protein coding mRNAs thus pass through the nuage and are translated because piRNA homology has been removed by splicing. By contrast, mature transposon transcripts carry piRNA complementarity are recognized by the perinuclear ping-pong machine, leading to destruction. Interestingly, mutations in the mouse *maelstrom* gene disrupt nuage and lead to male sterility and significant overexpression of LINE-1 elements (Soper et al., 2008). Nuage may therefore have a conserved function in transposon RNA surveillance and silencing.

In *S. pombe,* siRNAs bound to Ago1 appear to recruit HP1 to centromeres through interactions with nascent transcripts, thus triggering heterochromatin assembly and transcriptional silencing. Our data indicate that the HP1 homolog Rhino is required for transposon silencing, but this process appears to be mechanistically distinct from centromeric heterochromatin silencing in yeast. For example, localization of the Rhino HP1 homolog to nuclear foci is independent of piRNA production, and Rhino binding appears to promote transcription of heterchromatic clusters. This in turn generates piRNAs that may direct silencing through posttranscriptional target cleavage. However, piRNAs bound to PIWI proteins have been implicated in heterochromatin assembly in somatic cells, and this process could be related evolutionarily to heterochromatin assembly in fission yeast.

Intriguingly, *rhi* is a rapidly evolving gene, and all three Rhi protein domains (chromo, chromo shadow, and hinge) show evidence of strong positive selection (Vermaak et al., 2005). On the basis of these observations, Vermaak et al. (2005) proposed that *rhino* is involved in a genetic conflict within the germline. The observations reported here suggest that the conflict between transposon propagation and maintenance of germline DNA integrity drives *rhi* evolution, and that the heterochromatic dual-strand clusters have a key role in this battle. Rhino appears to define heterochromatic domains that produce transposon silencing piRNAs. Rhino could therefore have evolved to bind transposon integration proteins, which would promote transposition into clusters and production of *trans*-silencing piRNAs. In this model, the transposon integration machinery would evolve to escape Rhino binding and silencing. The rapid pace of *rhino* evolution makes identification of homologs in other species difficult (Vermaak et al., 2005), but the conserved role for piRNAs in germline development suggests that HP1 variants may have critical roles in the conflict between selfish elements and genome integrity in other species, including humans.

EXPERIMENTAL PROCEDURES

Drosophila Stocks

All animals were raised at 25°C. *Oregon R*, w^{1118} and cn^{1}; ry^{506} were used as controls, as noted. The following alleles were used: mnk^{P6} (Brodsky et al., 2004; Takada et al., 2003); $rhi^{KG00910}$ (rhi^{KG}) and rhi^{02086} (rhi^{2}) (Volpe et al., 2001); $armi^{72.1}$ and $armi^{1}$ (Cook et al., 2004); $mei41^{D3}$, (Hari et al., 1995; Hawley and Tartof, 1983); and P[*lacW*]*mei W68*^{K05603}, *mei W68*^{1} (McKim and Hayashi Hagihara, 1998). The mnk^{P6} allele was kindly provided by M. Brodsky (Brodsky et al., 2004). All other stocks were obtained from the Bloomington Stock Center, Bloomington, Indiana (http://flystocks.bio.indiana.edu/). Standard genetic procedures were used to generate double mutant combinations.

Immunohistochemistry

Antibody production is described in the Supplemental Experimental Procedures. Egg chamber fixation and whole mount antibody labeling were performed as previously described (Theurkauf, 1994). Vas protein was labeled with rabbit polyclonal anti Vas antibody (Liang et al., 1994) at 1:1000. Gurken protein was labeled with mouse monoclonal anti Gurken antibody (obtained from the Developmental Studies Hybridoma Bank, University of Iowa) at 1:10. Rhi protein was labeled with a guinea pig polyclonal anti Rhi antiserum developed by our group (see above) at 1:2000. Piwi, Aub, and Ago3 were labeled with rabbit polyclonal anti Piwi, anti Aub and anti Ago3 antibodies developed for this study (see above) at 1:1000. Antibody against γ H2Av was kindly provided by K. McKim (Gong et al., 2005) and egg chambers were labeled as described previously (Belmont et al., 1989). CID was labeled with an affinity purified chicken anti CID antibody provided by G. Karpen at 1:100 (Blower and Karpen, 2001). HOAP was labeled with a polyclonal rabbit anti Hoap antibody generated by our group (see above) at 1:1000. Rhodamine conjugated phalloidin (Molecular Probes) was used at 1:100 to stain F Actin, and TOTO3 (Molecular Probes) was used at 1:500 (0.2 mM final concentration) to visualize DNA.

Labeled tissue was mounted and analyzed with a Leica TCS SP inverted laser scanning microscope as described previously (Cha et al., 2001).

GFP-Rhino Transgene

The GFP Rhi transgene was generated by recombining the Rhi DONR (see above) construct with a modified pCasper vector containing the GFP sequence and Gateway cloning cassette B (Invitrogen). The resulting vector contained GFP fused in frame to the N terminus of Rhino under the control of the Gal4 promoter. Transgenic animals were generated using standard embryo microinjection techniques at Genetic Services.

RNA Isolation and Tiling Array Hybridization

Total RNA from was isolated from manually dissected ovaries from 2 to 4 day old flies with RNeasy (QIAGEN) according to the manufacturer's instructions. The RNA was quantified by absorbance at 260 nm. Three independent RNA isolates from each genotype was then assayed as follows: Double stranded cDNA was prepared with the GeneChip WT Amplified Double Stranded cDNA Synthesis Kit (Affymetrix). DNA was labeled with the GeneChip WT Double Stranded DNA Terminal Labeling Kit (Affymetrix). Labeled DNA was hybridized to GeneChip *Drosophila* Tiling 2.0R Arrays (Affymetrix) with the GeneChip Hybridization, Wash, and Stain Kit (Affymetrix) at the University of Massachusetts Medical School genomic core facility.

To determine whether genetic background or DNA damage significantly alters gene or transposon expression, we assayed ovarian RNA isolated from two common laboratory strains (*w 1118* and *cn,bw*), the meiotic repair mutant *okra,* the DNA damage signaling mutant *mnk*, and *mnk;okra* double mutants. Pairwise comparisons show little difference in genome wide patterns of gene or transposon expression in any of these five strains (Figure S12). The background for the *rhi* heteroallelic combination used here is *cn*/+; *ry*/+, which is genetically wild type. Since the *armi* allelic combination used here is in a homozygous *w 1118* background, this genotype was used a control in our array studies.

The tiling array data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE14370 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14370).

Small RNA isolation, oxidation, and sequencing were performed as described elsewhere (Li et al., 2009). Bioinformatics methods, chromatin immunoprecipitation and strand specific RT PCR procedures are described in the Supplemental Experimental Procedures.

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, 14 figures, and six tables and can be found with this article online at http://www.cell.com/supplemental/S0092 8674(09)00853 8.

ACKNOWLEDGMENTS

We thank Celeste Berg for *rhi* mutant stocks, Maria Zapp and Ellie Kittler in the UMass Deep Sequencing Core and Phyllis Spartick in the UMass Genomics core for expert assistance with small RNA sequencing and tiling array analyses, and members of the Zamore, Weng, and Theurkauf labs for critical discussions during the course of these studies. The Genomics Core was supported by the Diabetes and Endocrinology research Center grant DK032502. This work was supported in part by grants from the National Institutes of Health to W.E.T. (HD049116) and P.D.Z. (GM62862 and GM65236). H.X. was supported in part by Pfizer.

Received: February 20, 2009
Revised: May 7, 2009
Accepted: July 2, 2009
Published online: September 3, 2009

REFERENCES

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., 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.

Belmont, A.S., Braunfeld, M.B., Sedat, J.W., and Agard, D.A. (1989). Large scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. Chromosoma *98*, 129 143.

Blower, M.D., and Karpen, G.H. (2001). The role of Drosophila CID in kinetochore formation, cell cycle progression and heterochromatin interactions. Nat. Cell Biol. *3*, 730 739.

Bozzetti, M.P., Massari, S., Finelli, P., Meggio, F., Pinna, L.A., Boldyreff, B., Issinger, O.G., Palumbo, G., Ciriaco, C., Bonaccorsi, S., et al. (1995). The Ste locus, a component of the parasitic cry Ste system of Drosophila melanogaster, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the beta subunit of casein kinase 2. Proc. Natl. Acad. Sci. USA *92*, 6067 6071.

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

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

Brodsky, M.H., Weinert, B.T., Tsang, G., Rong, Y.S., McGinnis, N.M., Golic, K.G., Rio, D.C., and Rubin, G.M. (2004). Drosophila melanogaster MNK/Chk2 and p53 regulate multiple DNA repair and apoptotic pathways following DNA damage. Mol. Cell. Biol. *24*, 1219 1231.

Brower Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C., and Lin, H. (2007). Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev. *21*, 2300 2311.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi and heterochromatin dependent gene silencing. Cell *125*, 873 886.

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

Cha, B. J., Koppetsch, B., and Theurkauf, W.E. (2001). *In vivo* analysis of *bicoid* mRNA localization reveals a novel microtubule dependent anterior axis specification pathway. Cell *106*, 35 46.

Chambeyron, S., Popkova, A., Payen Groschene, G., Brun, C., Laouini, D., Pelisson, A., and Bucheton, A. (2008). piRNA mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. Proc. Natl. Acad. Sci. USA *105*, 14964 14969.

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

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.

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 path ways. PLoS ONE *3*, e1526.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207 210.

Elgin, S.C., and Grewal, S.I. (2003). Heterochromatin: silence is golden. Curr. Biol. *13*, R895 R898.

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

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

Girton, J.R., and Johansen, K.M. (2008). Chromatin structure and the regula tion of gene expression: the lessons of PEV in Drosophila. Adv. Genet. *61*, 1 43.

Gong, W.J., McKim, K.S., and Hawley, R.S. (2005). All paired up with no place to go: pairing, synapsis, and DSB formation in a balancer heterozygote. PLoS Genet. *1*, e67.

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.

Hari, K.L., Santerre, A., Sekelsky, J.J., McKim, K.S., Boyd, J.B., and Hawley, R.S. (1995). The mei 41 gene of D. melanogaster is a structural and functional homolog of the human ataxia telangiectasia gene. Cell *82*, 815 821.

Hawley, R.S., and Tartof, K.D. (1983). The effect of mei 41 on rDNA redun dancy in Drosophila melanogaster. Genetics *104*, 63 80.

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

Jang, J.K., Sherizen, D.E., Bhagat, R., Manheim, E.A., and McKim, K.S. (2003). Relationship of DNA double strand breaks to synapsis in Drosophila. J. Cell Sci. *116*, 3069 3077.

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.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509 521.

Liang, L., Diehl Jones, W., and Lasko, P. (1994). Localization of vasa protein to the Drosophila pole plasm is independent of its RNA binding and helicase activities. Development *120*, 1201 1211.

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. USA *104*, 6714 6719.

Lin, H. (2007). piRNAs in the germ line. Science *316*, 397.

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

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

McKim, K.S., and Hayashi Hagihara, A. (1998). mei W68 in Drosophila melanogaster encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. Genes Dev. *12*, 2932 2942.

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.

Modesti, M., and Kanaar, R. (2001). DNA repair: spot(light)s on chromatin. Curr. Biol. *11*, R229 R232.

O'Donnell, K.A., and Boeke, J.D. (2007). Mighty Piwis defend the germline against genome intruders. Cell *129*, 37 44.

Pal Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mecha nisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. Mol. Cell *9*, 315 327.

Pal Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669 672.

Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of Stellate elements of Drosophila melanogaster. Genetics *138*, 1181 1197.

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.

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

Pelisson, A., Payen Groschene, G., Terzian, C., and Bucheton, A. (2007). Restrictive flamenco alleles are maintained in Drosophila melanogaster popu lation cages, despite the absence of their endogenous gypsy retroviral targets. Mol. Biol. Evol. *24*, 498 504.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697 711.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. Curr. Opin. Genet. Dev. *12*, 162 169.

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.

Sarkaria, J.N., Busby, E.C., Tibbetts, R.S., Roos, P., Taya, Y., Karnitz, L.M., and Abraham, R.T. (1999). Inhibition of ATM and ATR kinase activities by the radiosensitizing agent, caffeine. Cancer Res. *59*, 4375 4382.

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

Seitz, H., Ghildiyal, M., and Zamore, P.D. (2008). Argonaute loading improves the 5′ precision of both MicroRNAs and their miRNA strands in flies. Curr. Biol. *18*, 147 151.

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.

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

Takada, S., Kelkar, A., and Theurkauf, W.E. (2003). Drosophila checkpoint kinase 2 couples centrosome function and spindle assembly to genomic integrity. Cell *113*, 87 99.

Theurkauf, W.E. (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. Methods Cell Biol. *44*, 489 505.

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.

Verdel, A., and Moazed, D. (2005). RNAi directed assembly of heterochro matin in fission yeast. FEBS Lett. *579*, 5872 5878.

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

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, X.Q., Redpath, J.L., Fan, S.T., and Stanbridge, E.J. (2006). ATR depen dent activation of Chk2. J. Cell. Physiol. *208*, 613 619.

Developmental Biology 342 (2010) 11–25

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Overlapping functions of Pea3 ETS transcription factors in FGF signaling during zebrafish development

Wade A. Znosko [a,b], Shibin Yu [c], Kirk Thomas [d], Gabriela A. Molina [a], Chengjian Li [a], Warren Tsang [a], Igor B. Dawid [e], Anne M. Moon [c,f,g], Michael Tsang [a,h,*]

[a] Department of Microbiology and Molecular Genetics, 3501 Fifth Avenue, BST3-5062, University of Pittsburgh, Pittsburgh, PA 15213, USA
[b] Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15213, USA
[c] Department of Pediatrics, University of Utah, Salt Lake City, UT 84112, USA
[d] Department of Internal Medicine, University of Utah, Salt Lake City, UT 84112, USA
[e] Program in Genomics of Development, NICHD, NIH, Building 6B/420 9000 Rockville Pike, Bethesda, MD 20892, USA
[f] Department of Neurobiology and Anatomy, University of Utah, Salt Lake City, UT 84112, USA
[g] Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA
[h] Department of Developmental Biology, University of Pittsburgh, School of Medicine, Pittsburgh PA, USA

ARTICLE INFO

Article history:
Received for publication 24 October 2009
Revised 11 March 2010
Accepted 15 March 2010
Available online 24 March 2010

Keywords:
FGF
Pea3 ETS factors
Heart
Left/right patterning
Zebrafish
Kupffer's vesicle
Cilia
Dusp6

ABSTRACT

Fibroblast growth factors (FGFs) are secreted molecules that activate the RAS/mitogen activated protein kinase (MAPK) signaling pathway. In zebrafish development, FGF signaling is responsible for establishing dorsal polarity, maintaining the isthmic organizer, and cardiac ventricle formation. Because several ETS factors are known transcriptional mediators of MAPK signaling, we hypothesized that these factors function to mediate FGF signaling processes. In zebrafish, the simultaneous knock down of three Pea3 ETS proteins, Etv5, Erm, and Pea3, produced phenotypes reminiscent of embryos deficient in FGF signaling. Morphant embryos displayed both cardiac and left/right patterning defects as well as disruption of the isthmic organizer. Furthermore, the expression of FGF target genes was abolished in Pea3 ETS depleted embryos. To understand how FGF signaling and ETS factors control gene expression, transcriptional regulation of *dusp6* was studied in mouse and zebrafish. Conserved Pea3 ETS binding sites were identified within the *Dusp6* promoter, and reporter assays showed that one of these sites is required for *dusp6* induction by FGFs. We further demonstrated the interaction of Pea3 ETS factors with the *Dusp6* promoter both *in vitro* and *in vivo*. These results revealed the requirement of ETS factors in transducing FGF signals in developmental processes.

Introduction

Fibroblast growth factors (FGFs) are a family of 22 proteins in humans that are critical for proper development (Furthauer et al., 2004; Itoh, 2007; Ornitz and Itoh, 2001; Sekine et al., 1999; Sun et al., 1999). FGFs bind to a family of four receptor tyrosine kinases (FGFRs), which initiate several signaling cascades, including the rat sarcoma homologue (RAS)/mitogen activated protein kinase (MAPK) cascade. Within this signaling arm, activation of a MAPK protein, *extracellular signal regulated protein kinase* (ERK), eventually leads to gene regulation through the modification of transcription factors (Dailey et al., 2005; Powers et al., 2000; Tsang and Dawid, 2004).

The importance of Fgf8 signaling in multiple developmental processes is highly conserved across species as revealed by severe brain and cardiac defects in *Fgf8* deficient and tissue specific mouse mutants (Frank et al., 2002; Macatee et al., 2003; Meyers et al., 1998; Park et al., 2006). In zebrafish, mutations in *Fgf* ligand genes have shown the importance of these factors in cell survival, migration, and patterning (Draper et al., 2003; Herzog et al., 2004; Reifers et al., 1998; Yamauchi et al., 2009). Due to the wide ranging biological importance of FGFs, tight control of the signal is essential to regulate the many FGF mediated developmental processes (Thisse and Thisse, 2005; Tsang and Dawid, 2004). To limit FGF signaling, several feedback attenuators have been identified within this pathway, including *sef* (Furthauer et al., 2002; Tsang et al., 2002), *sprouty* (*spry*) (Furthauer et al., 2001; Hacohen et al., 1998; Kramer et al., 1999; Nutt et al., 2001; Tefft et al., 2002) and *dusp6* (Kawakami et al., 2003; Tsang et al., 2004). In addition, Pea3 ETS (E26 transformation specific) factors are thought to function as transcriptional regulators of the pathway that allow proper signaling levels to be reached and maintained during development (Munchberg et al., 1999; Raible and Brand, 2001; Roehl and Nusslein Volhard, 2001).

The ETS family of transcription factors is important for cell proliferation, differentiation, and migration (Kobberup et al., 2007; Wasylyk et al., 1998). All family members contain the ETS domain, an 85 amino acid winged helix loop helix domain that binds to a core

* Corresponding author. Department of Microbiology and Molecular Genetics, 3501 Fifth Avenue, BST3-5062, University of Pittsburgh, Pittsburgh, PA 15213, USA.
E-mail address: tsang@pitt.edu (M. Tsang).

0012-1606/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2010.03.011

GGAA/T DNA sequence. The Pea3 subfamily of ETS transcription factors contains three members, Polyomavirus enhancer activator 3 (PEA3/ETV4), Er81 (ETV1), and ETS related molecule (ERM), also known as ETS variant 5 (ETV5) in mouse and human. In zebrafish, due to genome duplication, two ERM/ETV5 genes exist, *erm* and *etv5*, in addition to *er81* and *pea3* (Kudoh et al., 2001; Roussigne and Blader, 2006). In resting cells these proteins are folded in an inactive state, but upon ERK phosphor ylation undergo a conformational change, allowing the ETS domain to be exposed for DNA binding and gene transcription (Laget et al., 1996; O'Hagan et al., 1996; Oikawa and Yamada, 2003; Sharrocks, 2001). With the exception of *Er81*, the FGF/RAS/MAPK pathway regulates expression of *Pea3* transcription factors in zebrafish (Roussigne and Blader, 2006).

Several members of the mouse Pea3 ETS proteins have been characterized, providing some evidence for the role for these transcrip tion factors in FGF signaling and development (Brent and Tabin, 2004; Liu et al., 2003; Lu et al., 2009; Mao et al., 2009; Zhang et al., 2009). In this study, we focused on the role of Pea3 ETS transcription factors in FGF signaling and embryogenesis in the zebrafish. Ectopic expression of activated forms of Etv5 induced transcription of *dusp6* and *sef*. Conversely, depletion of Pea3 ETS proteins by antisense morpholino injections disrupted the isthmic organiser and heart formation, and altered left/right patterning. The essential DNA sequence required for FGF mediated induction of *dusp6* was identified. Further, the ortholo gous region of the mouse *Dusp6* promoter is regulated by Etv4 and Etv5 via conserved ETS sites and is bound by ETV4 *in vivo*. These results indicate the importance of ETS factors in relaying FGF signals during development and provide insights as to how FGF target genes are regulated.

Materials and methods

Zebrafish maintenance, RNA injections, and in situ hybridization

These procedures were performed as described previously (Tsang et al., 2000) with the following modifications for RNA injections: Wildtype AB* zebrafish embryos were injected with *etv5* (100 pg), *etv5:VP16* (75 pg), *etv5:EnR* (150 pg), or mutated forms of *etv5; etv5:T135D, etv5: T139D, etv5:S142D* (50 pg) at the 1 2 cell stage. All mutations within Etv5 were generated as described in the QuickChange II Site directed Mutagenesis Kit (Stratagene) using the primers: 5′ GGGTTCAAGCC CATTGGACCCTCCCTCGGACCCC 3′ (forward *etv5:T135D*), 5′ GGGGTCCGAGGGAGGGTCCAATGGCTTGAACCC 3′ (reverse *etv5:T135D*), 5′ ACTCCTCCCTCGGACCCCGTCTCCCCATGT 3′ (forward *etv5:T139D*), 5′ ACATGGGGAGACGGGGTCCGAGGGAGGAGT 3′ (reverse *etv5:T139D*), 5′ CCCTCGACGCCCGTCGACCCATGTGTCCCCAGC 3′ (forward *etv5:S142D*), 5′ GCTGGGGACACATGGGTCGACGGGCGTCGAGGG 3′ (reverse *etv5: S142D*). *In situ* hybridizations were performed as described (Kudoh et al., 2001). The following antisense riboprobes were generated in this study; *etv5* (Kudoh et al., 2001), *erm* (Munchberg et al., 1999), *pea3* (Brown et al., 1998), *fgf3* (Kiefer et al., 1996), *fgf8* (Reifers et al., 1998), *dusp6* (Kawakami et al., 2003; Tsang et al., 2004), *sef* (Tsang et al., 2002), *bmp4* (Hwang et al., 1997), *chordin* (Schulte Merker et al., 1997), *pax2a* (Krauss et al., 1991), *her5* (Bally Cuif et al., 2000), *amhc* (Berdougo et al., 2003), *vmhc* (Yelon et al., 1999), *nkx2.5* (Chen and Fishman, 1996), *gata4* (Reiter et al., 1999), *scl* (Liao et al., 1998), *spaw* (Long et al., 2003) and *hand2* (Yelon et al., 2000).

Immunofluorescence

Whole mount immunofluorescence was performed to detect monocilia in the Kupffer's vesicle as described (Yamauchi et al., 2009). Monoclonal acetylated tubulin primary antibody (1:1000; Sigma) and Cy3 conjugated anti mouse IgG secondary antibody (1:500; Jackson ImmunoResearch Laboratories) was used in this study.

Antisense morpholino injections

Morpholino oligos (MOs) were designed and purchased from Gene Tools (www.gene tools.com):

etv5MO sequence, 5′ ATCCGTCCATGTCACCTGGGTCTTC 3′;
etv5/ermMO sequence, 5′ TGCTGGTCATAAAATCCGTCCATGT 3′;
fgf8MO sequence, 5′ GAGTCTCATGTTTATAGCCTCAGTA 3′;
and *ContMO* sequence, 5′ CCTCTTACCTCAGTTACAATTTATA 3′.

From Open Biosystems (www.openbiosystems.com):

ermMO sequence, 5′ AACCCATCCATGTCGCTTGCTTCTC 3′;
pea3MO sequence, 5′ ATCCATGCCTTAACCGTTTGTGGTC 3′;

ContMO (10 ng), *fgf8MO* (1 3 ng), single *EtsMO* (5 ng), *2EtsMO* (2.5 ng of each *EtsMO*), or *3EtsMO* (1.3 ng of each *EtsMO*) were injected into AB* or *Tg(d2EGFP)*^{pt6} embryos as described (Molina et al., 2007).

Generation of Dusp6 reporter constructs

Zebrafish *Dusp6* promoter sequences were PCR amplified and directionally cloned into the pENTR Gateway vector (Invitrogen) with the following primers:

Forward primers: 5Kb: 5′ CACCGACCGGTAGTGAATTTTGATTTGAAC 3′;
4Kb: 5′ CACCGGCCTAGTCGGCACTCAAACCAGTGA 3′;
3Kb: 5′ CACCACTGTGGCATTACAGTGACAGGCCCG 3′;
2Kb: 5′ CACCCTGCGCAGAAGTTCACTTAGACAGTG 3′;
1Kb: 5′ CACCCCACACTGAACTGAGCTAAACTGAAC 3′;
Reverse primer: *Dusp6* Rev: 5′ GGTACCGTGAGACCTTAAAACTGCGG 3′.

The promoter sequences were verified and subcloned into a Gateway modified pGL3 promoter (Promega) by the Gateway Sys tem (Invitrogen). Pea3A and Pea3B mutant reporter constructs were generated using QuickChange II with: MutA: 5′ CACTCG CACTCCTCCGGCCGTCCCGTGAAGCGCCTCTCG 3′; MutB: 5′ CCGCTGATCCGGCGCGGCCCGTCCTTTCCGTTTTTGTG 3,′ underlined se quence indicate mutation of the ETS binding site.

A 782 bp fragment of DNA 5′ to the mouse *Dusp6* gene was amplified by PCR and cloned into pGL3 Promoter. The primers were:

5′ ccctggtaccGTACCGTTGGATTAGCATTTAACACTTCGT (sense, UCSC genome browser Chr10: chr10:98,725,230)

5′ ccctagatctAGTCTAGCGGCTCTTAATCCTC (anti sense, UCSC genome browser Chr10: 98,726,011); upper case letters indicate genomic sequence. Amplification was performed with Pfx Platinum polymerase (Invitrogen) and 0.1 μg mouse DNA (C57Bl/6) under manufacturer conditions. The single reaction product was purified by Qiaquick spin column (Qiagen), digested with Acc65I and BglII (New England Biolabs), and cloned into pGL3 Promoter. Mutagenesis of the conserved putative PEA3 binding sites 1 3 was performed with a Quick Change Lightening Kit (Agilent) with mutagenic oligonucleotides designed by the manufac turer (see Supplementary Fig. 5A) and plasmids were sequenced.

Mammalian cell and Xenopus explant cultures and luciferase assays

Human 293 T cells (ATCC) were grown at 37 °C, 5% CO_2 in DMEM (GIBCO) containing 10% FBS, 100U/ml penicillin and streptomycin. Cells were seeded into 24 well plates and were 70 80% confluence at the time of transient transfection using FuGene6 (Roche). Each well received 100 ng of pGL3 Control luciferase reporter, 10 ng of pRL TK vector (Promega) encoding *Renilla* luciferase and increasing amounts of *Pea3/Etv4* (Open Biosystems, Clone ID: 3854349) or *Erm/Etv5* (Open Biosystems, Clone ID: 4036564) expression vector. The total amount of DNA in each well was adjusted to 310 ng with pcDNA3. The cells were harvested after 48 h using Passive Lysis buffer (Promega).

Xenopus laevis 2 cell stage embryos were injected with *fgf8* RNA (25 pg), *Dusp6:luc* construct (125 pg), and pCMV Renilla (50 pg) into each cell. In other experiments, *etv5* (50 pg), *etv5:VP16* (30 pg) or *etv5:T135D* (40 pg) was injected in place of *fgf8* RNA. Animal caps were dissected at stage 8.5 and cultured for 6 h. Animal cap lysates were prepared from a population of four animal caps. In both cell and frog assays, luciferase activities were determined using Dual Lucifer ase Reporter Assay System (Promega). The data were normalized by calculating the ratio of firefly luciferase to *Renilla* luciferase. A two sample equal variance T test using a two tailed distribution was applied to analyze the data for statistical significance.

Electrophoretic Mobility Shift Assays (EMSA)

5′ labeled biotinylated and standard oligonucleotides (Midland Certified) 5′ GTTTGTTTGCACTCCGCT 3′ (forward) and 5′ TTTGTCATTCA CAAAAAC 3′ (reverse) were used to generate biotinylated and competitor PCR amplified *Dusp6* promoter. For oligonucleotide competition assays, 18mer Pea3B site competitors were generated:

5′ TCCGGAGCG GAAATTCCT 3′ (forward)
5′ AGGAATTTCCGCTCCGGA 3′ (reverse).

A random sequence competitor oligonucleotide was generated using:

5′ CCCTCGACGCCCGTCGACCCATGTGTCCCCAGC 3′ (forward)
5′ GCTGGGGACACATGGGTCGACGGGCGTCGAGGG 3′ (reverse).

GST Etv5 ETS DNA binding domain protein were expressed in bacteria (BL21) cells (Invitrogen) and batch purified with GST sepharose beads (Amersham). Biotin labeled DNA was incubated with Etv5 ETS protein (10ug) and resolved on a 6% native polyacrylamide gel. The products were transferred to a nylon membrane, cross linked by a UV Stratalinker (Stratagene) and detected with streptavidin HRP (Pierce). Competition assays were performed with unlabeled oligos or PCR amplified *Dusp6* promoter at 1 to 500 fold molar excess to labeled probes.

Chromatin immunoprecipitation (ChIP) assay

Chromatin was immunoprecipitated using a variation of the protocol described at http://www.cellsignal.com. Briefly, 100 E9.5 mouse embryos pharyngeal arch regions 2 6 were dissected, minced into fine slurry, cross linked in 2% formaldehyde for 30 min, washed with ice cold PBS, pelleted, and re suspended in 1% NP 40 non denaturing lysis buffer with protease inhibitors (Roche, 1 836 153). The chromatin DNA was digested with micrococcal nuclease (New England Biolabs, M0247S) for 5 min to generate fragments ranging from 150 to 900 bp and diluted in ChIP buffer with protease inhibitors. Two percent of the diluted supernatant was kept as input control. Sample was incubated at 4 °C overnight with antibody against Pea3 (Etv4) (Santa Cruz, sc 113 X). A normal mouse IgG (Santa Cruz, sc 2025) was used as negative control. Complexes were precipitated with Dynabeads® Pan Mouse IgG (Invitrogen, 110 41). Immunoprecipitated chromatin was washed, eluted in elution buffer, incubated at 65 °C for 8 h, and treated with proteinase K. DNA was purified by phenol extraction and ethanol precipitation.

Two different *Dusp6* primer sets were employed to determine if the putative PEA3 binding sites in the *Dusp6* promoter were enriched in chromatin immunoprecipitated with anti ETV4 antibody relative to distant sequences at this locus: the "*Dusp6* prom" set amplifies the highly conserved region of the mouse *Dusp6* promoter: 5′ AGTGCCCTGGTTTATGTGC 3′ (chr10: 98,725,599 98,725,618, sense), 5′ CGGGAGGAAGGAGAAAGAA 3′ (chr10: 98,725,599 98,725,618, anti sense). The "*Dusp6* neg" negative control primer set amplifies a non conserved region 5′ to the *Dusp6* gene: 5′ AAGGCCGAGGAAAAGACTTC3′ (chr10: 98,721,292 98,721,311, sense,), 5′ ACCCGTGTTACTGGAGATCG′ (chr10: 98,721,426 98,721,445, anti sense). An additional negative control primer set was employed that amplifies a region upstream of β actin gene: 5′ GTGCTTAAGAGTCCACTATGAGGG3′ (sense), 5′ TCCACTCGCAATCATATACTTAGG3′ (anti sense). Equal quantities of input and immunoprecipitated DNA samples were subject to PCR (35 cycles), electrophoresed on a 2% agarose gel and visualized by ethidium bromide staining.

To quantitate the enrichment of the Dusp6 promoter region in the ChIPed region, quantitative PCR was performed in the Bio Rad iCycler IQ™ Multicolor Real Time PCR Detection System using 25 µl IQ™ SYBR® Green Supermix (Bio Rad, 170 8882), 250 nM of each primer and 2 µl of immunoprecipitated and input DNA samples. The amplification ramp included an initial hold of 5 min at 94 °C, followed by a three step cycle consisting of denaturation at 94 °C (30 s), annealing at 57 °C (30 s) and extension at 72 °C (30 s); the amplification fluorescence was read at the end of the cycle. The Ct values and standard deviations were analyzed using a method modi fied from that of SuperArray Biosciences: (http://www.workingth ebench.com/search/label/chromatin%immunoprecepitation). This protocol correctly propagates the standard errors in the PCR data and permits quantitation of the *Dusp6* target amplicon in the ETV4 ChIPed sample relative to that ChIPed by a nonspecific antibody (normalized to the actin negative control amplicon to control for total DNA quantity in each sample), and expressed as a percent of amplicon detected in the input sample. The enrichment was nearly 7 fold and this was highly reproducible over multiple ChIP experiments with this tissue.

Results

Pea3 ETS transcription factors can induce FGF target gene expression

Amino acid alignment between the three Pea3 ETS transcription factors shows high sequence conservation, especially within the DNA binding domain and their developmental expression overlaps with each other and with *fgf8* and *fgf3* (Supplementary Fig. 1A G) (Raible and Brand, 2001; Roehl and Nusslein Volhard, 2001; Roussigne and Blader, 2006). *Er81* expression is not limited to the domains where *fgf* ligands are expressed and suppression of FGF signaling does not affect *Er81* expression. It is likely that *Er81* has evolved in zebrafish to serve a separate function that is distinct from its orthologs in mouse and *Xenopus laevis* and therefore, *Er81* was not examined in these studies. Sequence conservation and similar expression pattern suggest that the three ETS factors may perform redundant functions. To determine if Pea3 ETS factors can regulate FGF signaling, we employed gain of function studies with *etv5*. Microinjection of *etv5* mRNA into the 1 2 cell stage embryo did not evoke discernable phenotypes or induce the FGF target genes *dusp6* and *sef* (Fig. 1A, B, F, G, U). In support of these findings, dorsal ventral patterning as determined by expression of *bmp4* (Fig. 1K, L, U) and *chordin* (Fig. 1P, Q, U) was unaffected. Similar results were observed after injection of *erm* or *pea3* mRNA, suggesting that the ectopically expressed factors were non functional (not shown). The lack of phenotype noted in *etv5* injected embryos was likely due to the auto inhibitory domain located within the amino terminus of Pea3 proteins that suppresses DNA binding (Supplemen tary Fig. 2A) (Laget et al., 1996; O'Hagan et al., 1996). To overcome this inhibition, the Etv5 ETS domain was fused to the VP16 transcriptional activation motif (*etv5:VP16*) (Supplementary Fig. 2B). Ectopic expres sion of *etv5:VP16* greatly induced *dusp6* and *sef* expression (Fig. 1C, H, U) and exhibited dorsalized phenotypes (not shown). *bmp4* expres sion was diminished while *chordin* was increased in *etv5:VP16* injected embryos (Fig. 1M, R, U). The Pea3 proteins act as transcrip tional activators and fusion of an Engrailed transcriptional repressor (EnR) domain was used to generate a dominant negative protein (Brent and Tabin, 2004; Liu et al., 2003). Injection of *etv5:EnR* resulted in ventralized phenotypes (not shown), and expression of *dusp6* and

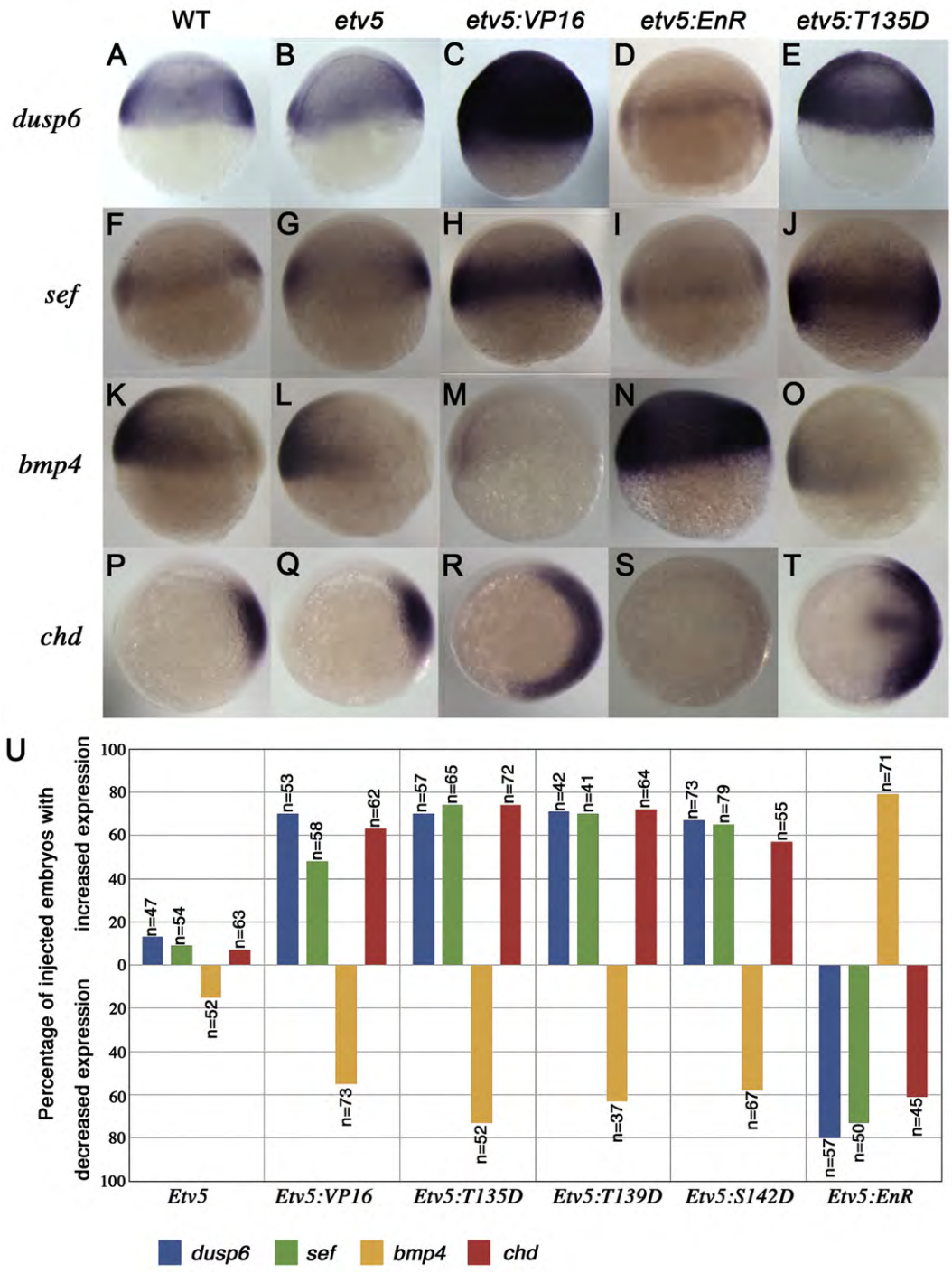


Fig. 1. Etv5 functions as a positive effector in FGF signaling. (A–O) Lateral views and (P–T) animal views at shield stage. *In situ* probes and constructs injected are indicated on the left and above, respectively. Expression of downstream targets of FGF signaling, *dusp6* and *sef*, are unaltered in *etv5*-injected embryos (B, G), but were induced by *etv5:VP16* mRNA injections (C, H). Ectopic expression of dominant negative (*etv5:EnR*) blocks *dusp6* and *sef* expression (D, I). Injection of *etv5:T135D* increased *dusp6* and *sef* expression (compare A, F to E, J). Expression of *bmp4* and *chordin* was unchanged in *etv5*-injected embryos (L, Q). A dorsalized phenotype was noted in *etv5:VP16*-injected embryos (M, R) and a ventralized phenotype upon *etv5:EnR* mRNA injection (N, S). Injection of *etv5:T135D* mRNA evoked dorsalized phenotypes (O, T). (U) Frequency of phenotypes elicited by microinjection of mRNAs indicated.

sef were both suppressed (Fig. 1D, I, U). Further, *bmp4* expression was expanded with the concomitant reduction in *chordin* transcripts (Fig. 1N, S, U). These results indicated that Etv5 was capable of regulating FGF target gene expression.

The DNA binding activity of ERM/ETV5 is regulated by ERK2 phosphorylation (Laget et al., 1996; O'Hagan et al., 1996). We identified a series of conserved putative ERK phosphorylation motifs (PXS/TP) in zebrafish Etv5 and Erm (Supplementary Fig. 2C). To determine the importance of these residues, we individually mutated T135, T139, and S142 residues to aspartic acid to mimic phosphorylation (*etv5:T135D*, *etv5:T139D*, and *etv5:S142D*). Injection of *etv5:T135D* mRNA induced both *dusp6* and *sef* expression similar to the results obtained with *etv5:VP16* injections (Fig. 1E, J, U). As predicted, *bmp4* expression was suppressed while *chd* was expanded, indicating strong dorsalization (Fig. 1O, T, U). Dorsalized phenotypes were also noted in *etv5:T139D* and *etv5:S142D* injected embryos (Fig. 1U). We

also identified a putative ERK phosphorylation site in Pea3 at serine residue 100 and ectopic expression of *pea3:S100D* resulted in dorsalized phenotypes and expanded the expression of FGF target genes (not shown) (Guo and Sharrocks, 2009). These experiments indicate that Etv5 and Pea3 are transcriptional regulators of FGF target genes and that their transcriptional activity could be controlled by ERK phosphorylation.

Pea3 ETS factors are required for FGF signaling

To determine the requirement of Pea3 ETS factors for FGF mediated developmental processes, we generated antisense morpholinos (MOs) that blocked translation of members of the Pea3 ETS genes. Each antisense MO targeted the AUG codon of *etv5*, *erm*, or *pea3*, and one MO that targeted both *etv5* and *erm* (*etv5*/*erm* MO), given the high degree of identity between these two genes. To confirm MO specificity, embryos were co injected with MOs and a construct containing the MO targeting site fused to enhanced green fluorescent protein (EGFP; Supplementary Fig. 3A). At the shield stage, each individual MO suppressed expression of its respective Ets EGFP fusion protein and not the other related constructs (Supplementary Fig. 3B and C), indicating that the MOs were effective and specific. Targeted depletion of Pea3, Erm or Etv5 individually did not result in discernable phenotype at 28hpf (not shown). In contrast the knock down of all three genes (herein referred to as *3EtsMO*) resulted in mid hindbrain (MHB) defects similar to the *fgf8* zebrafish mutant, *acerebellar* (*ace*) (Fig. 2B) (Picker et al., 1999; Reifers et al., 1998). These results suggest that Pea3 ETS transcription factors are functionally redundant. *dusp6* expression in *3EtsMO* injected embryos was reduced in the MHB at 28hpf, suggesting that Pea3 ETS factors are required to mediate FGF signaling in the MHB (Fig. 2H). We confirmed that these factors function redundantly as microinjection of MO resistant *etv5* (*HAetv5*) or *pea3* (*HApea3*) mRNA rescued the MHB phenotype and *dusp6* expression (Fig. 2D I; 73% rescued with *HAetv5*, $n=71$ and 66% rescued with *HApea3*, $n=44$). Furthermore, the co injection of *HAetv5* and *HApea3* mRNA together did not result in an additive or synergistic rescue of *3EtsMO* morphants, suggesting that either gene can compensate for the loss of the other (Fig. 2F, I; 69% rescue, $n=56$). These data suggest that Pea3 ETS factors function redundantly and are required for FGF signaling.

The depletion of *pea3*, *erm*, or *etv5* alone did not alter *dusp6* or *sef* expression at shield stage (Fig. 3A, B, E, F, Q and data not shown). However, the knock down of both *etv5* and *erm* did result in a mild suppression of these FGF target genes (Fig. 3C, G, Q). In contrast, the concerted depletion of all three factors dramatically diminished both *dusp6* and *sef* expression, even with lower concentrations of MO injected for each target (Fig. 3D, H, Q). At 28hpf, single or two ETS (*erm* and *etv5* herein referred to as *2EtsMO*) knockdown displayed minimal reduction of *her5* and *pax2a* expression in the MHB (Fig. 3J, K, N, O, Q). In contrast, depletion of all three family members strongly suppressed expression of these genes (Fig. 3L, P, Q). These loss of function studies further emphasized functional redundancy within the Pea3 ETS family. Our experiments indicated that the reduction of one Pea3 ETS family member was not sufficient to generate discernable phenotypes, but the knockdown of three Pea3 ETS genes resulted in an inhibition of FGF signaling and evoked defects similar to that of *fgf8* mutants, suggesting the importance of Pea3 ETS factors within this signaling cascade.

A role for Pea3 ETS factors maintaining cardiac progenitors

In mice, altering gene dosage of *Fgf8* or the source of Fgf8 protein demonstrated a role for *Fgf8* in cardiovascular development (Abu Issa et al., 2002; Frank et al., 2002; Macatee et al., 2003; Meyers and Martin,

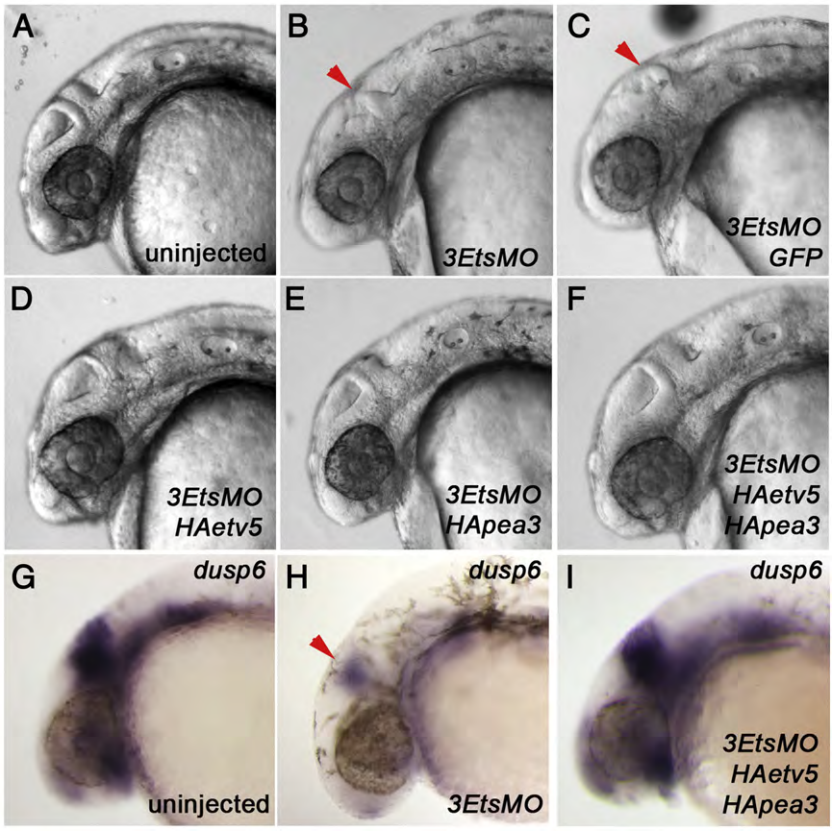


Fig. 2. The Pea3 ETS factors perform redundant functions and are required for mid–hindbrain boundary (MHB) formation. (A–I) Lateral views of 28hpf embryos. Injection of antisense morpholinos targeting *pea3*, *erm* and *etv5* (*3EtsMO*) resulted in MHB defects (B) as compared to the Control scrambled morpholino (A). *3EtsMO*-induced MHB defects can be rescued by co-injection of *HAetv5* or *HApea3* mRNA (D–F) but not by *GFP* mRNA (C). Expression of *dusp6* is diminished in *3EtsMO*-injected embryos (H). Co-injection of *HAetv5* and *HApea3* rescued *dusp6* expression (I). Red arrowhead marks loss of MHB in *MO*-injected embryos.

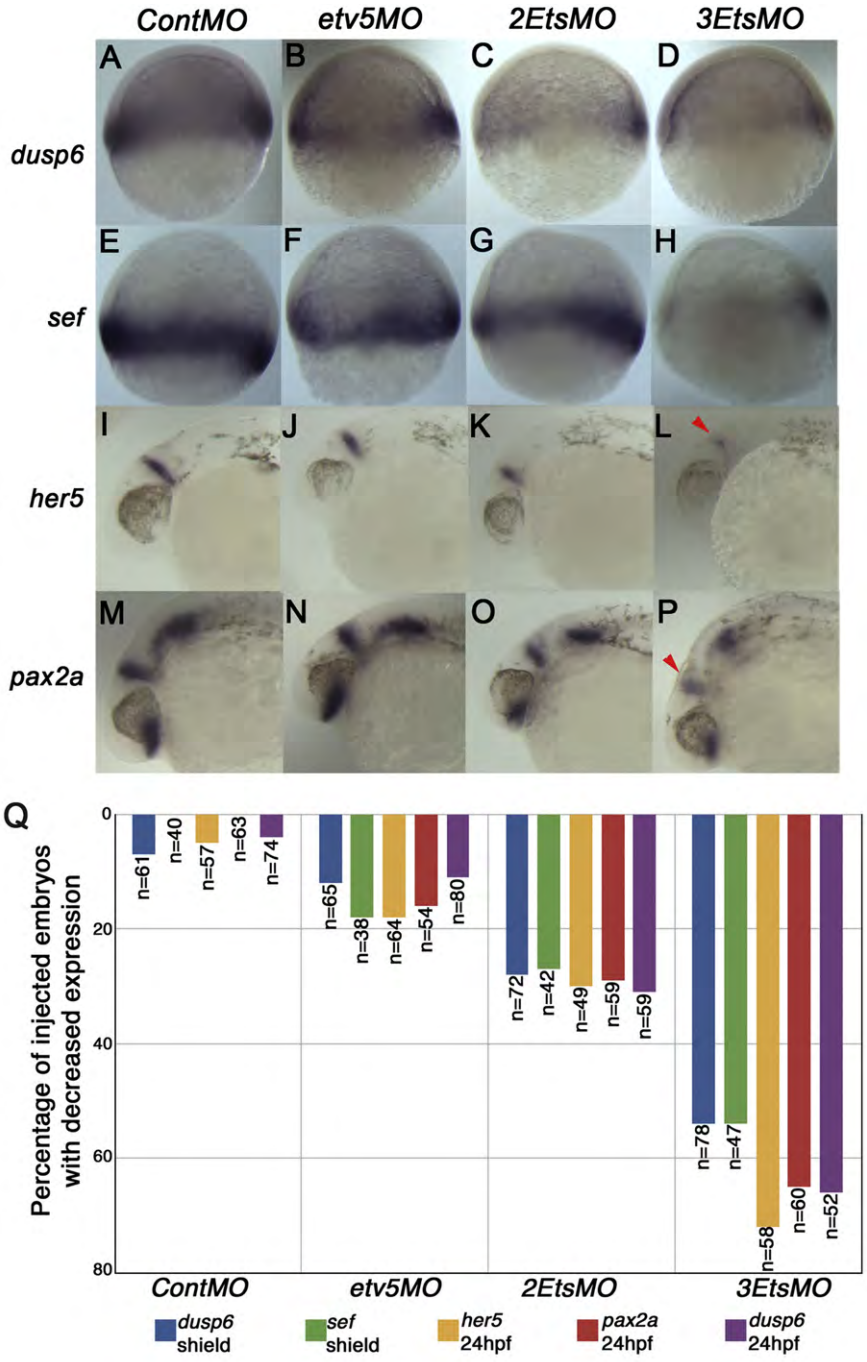


Fig. 3. Pea3 ETS factors are required for target gene responses to FGF signaling. (A–P) Lateral views of embryos injected with MOs indicated above and assayed with probes indicated on the left. Expression of FGF target genes, *dusp6* and *sef*, were unchanged in *etv5MO*-injected embryos (B, F). In contrast, the double or triple knockdown of Pea3 ETS genes elicited a drastic reduction of *dusp6* and *sef* (C, D, G, H). Knockdown of Pea3 ETS genes resulted in MHB defect as measured by *her5* (I–L) and *pax2a* (M–P) expression. Individual knock-down had minimal effects, while *3EtsMO* injections resulted in dramatic reduction of *her5* (L) and *pax2a* (P). (Q) Graph depicting the phenotypic effects of *etv5MO*, *2EtsMO* and *3EtsMO* on expression FGF target genes and MHB markers. Red arrowhead marks loss of MHB gene expression.

1999; Park et al., 2006). More recently, conditional ablation of *Fgf8* and *Fgf receptors* with different Cre driver lines revealed that Fgf8 regulates expression of *Pea3/Etv4* and *Erm/Etv5* in heart precursors and is required for outflow tract formation (Ilagan et al., 2006; Park et al., 2006; Park et al., 2008). These studies determined the temporal and spatial role for Fgf8 derived from both the endoderm and mesoderm for proper morphogenesis of the heart, and showed that Fgf8 is required for proliferation and survival of cardiac progenitors (Ilagan et al., 2006; Park et al., 2006). FGF signaling in zebrafish is necessary to specify and maintain cardiac progenitor cells during somitogenesis stages (Marques et al., 2008; Molina et al., 2009; Reifers et al., 2000). Based on previous fate mapping studies in zebrafish, we observed expression of *pea3, erm*, and *etv5* in the zebrafish anterior lateral plate mesoderm (ALPM), suggesting that these factors may play a role in establishing and maintaining cardiac progenitors during somitogenesis stages (not shown) (Lee et al., 1994; Stainier et al., 1993). Injection of *etv5MO* mildly reduced expression of a cardiac specific transcription factor, *nkx2.5* (Fig. 4A, B, Y). However, injection of *2EtsMO* or *3EtsMO* markedly reduced this cardiac population, a phenotype that was similar to *fgf8MO* injected embryos or in *ace* mutants (Fig. 4C F, Y). *Gata4*, a gene expressed throughout the ALPM, was also reduced in both *2EtsMO* and *3EtsMO* injected embryos, and was comparable to *fgf8* deficient embryos (Fig. 4G L, Y) (Draper et al., 2001; Marques et al., 2008;

Serbedzija et al., 1998). We next analyzed expression of *hand2*, a marker for lateral cardiac progenitors in the ALPM (Yelon et al., 2000). Complimentary to the observed reduction of *nkx2.5*$^{+}$ cells, *hand2* expression was also reduced with knock down of multiple Pea3 factors (Fig. 4M R, Y).

Within the ALPM, an important interplay between the hematopoi etic/vascular cells and cardiac progenitors exists to maintain the size of both populations (Keegan et al., 2004; Schoenebeck et al., 2007). We analyzed *scl* expression at the 10 somite stage in *3EtsMO* injected embryos to determine if these factors play a role in endothelial lineages within the ALPM. Knock down of Pea3 ETS factors resulted in an expansion of *scl* expression at the 10 somite stage (Fig. 4S V, Y). Most striking was that the depletion of *etv5* alone was sufficient to expand *scl* populations, indicating that this precursor population is particularly sensitive to Etv5 activity (Fig. 4T). Similar MO injections against only *erm* or *pea3* had no effect on *scl* expression, indicating the importance of a single, specific ETS family member, *etv5*, in restricting endothelial domains within the ALPM (data not shown). Injection of *2EtsMO* or *3EtsMO* elicited stronger expansion of hematopoietic and vessel lineages as also noted in *fgf8* depleted or *ace* embryos, implicating that the loss of

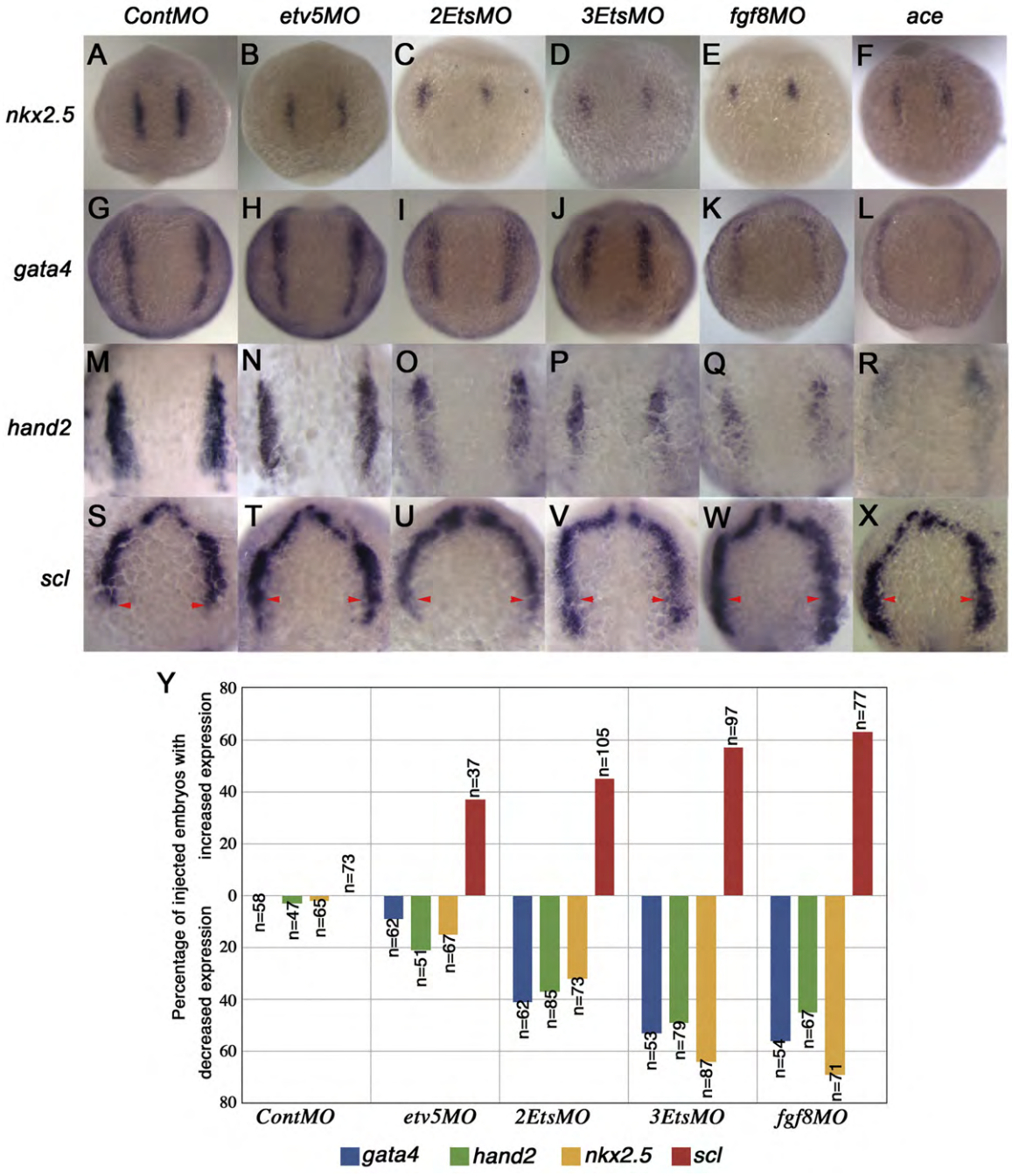


Fig. 4. Pea3 ETS factors are required to maintain cardiac progenitors. (A–X) Dorsal views at 10-somite stage. *In situ* probe listed on left and MO above. (A–D) *nkx2.5* expression was reduced as multiple Pea3 factors were knocked down (D). This was similar to *fgf8MO* knock-down and in *ace* mutants (E–F). (G–J; M–P) *gata4* expression (G–J) and *hand2* expression (M–P) were also reduced after Pea3 depletion (J, P). A similar phenotype was observed in *fgf8MO* knock-downs and in *ace* mutants (K, L;Q, R). (S–V) Dorsal view of *scl* expression indicated expansion of endothelial lineages in *EtsMO* injections (T–V) as indicated by arrowheads that mark the caudal limit in uninjected embryos (S). Similar results were noted in *fgf8MO* and *ace* embryos suggesting that FGF signaling is required to maintain cardiac progenitors and to limit endothelial lineages to the rostral ALPM (W, X). (Y) Graph providing quantitative data for MO experiments.

FGF signaling expanded endothelial lineages (Fig. 4W Y). Taken together, these data provided evidence that Pea3 ETS factors are required to maintain cardiac progenitors within the ALPM.

Pea3 ETS factors function in Left/Right (L/R) patterning of the heart

We next determined if the loss of cardiac progenitors at early somitogenesis stages resulted in later heart defects. We assayed expression of two specific cardiac differentiation markers at 24hpf, *ventricular myosin heavy chain* (*vmhc*) for ventricular tissue, and for atria, *atrial myosin heavy chain* (*amhc*) (Berdougo et al., 2003; Yelon et al., 1999). In *etv5MO* injected embryos, *amhc* was expressed in a population of cells just below the left eye (Fig. 5B), resembling observations in control scrambled morpholino (*ContMO*) injected embryos (Fig. 5A). In the *2EtsMO* and *3EtsMO* injected embryos, diffuse *amhc* staining was noted at the midline, suggesting disruption of heart tube morphogenesis (Fig. 5C, D). In *ContMO* and *etv5MO* injected embryos, *vmhc* expression outlined the heart tube as it "jogged" to the left (Fig. 5E, F). However in *2EtsMO* or *3EtsMO* injected embryos, ventricle morphology was altered and was predominantly located at the midline between the eyes, suggesting a L/R patterning defect (Fig. 5G, H). Analysis of *cardiac myosin light chain 2* (*cmlc2*) expression at 48hpf revealed disruption of cardiac looping in *3EtsMO* injected larvae, resulting in either non looping straight hearts or reversed looping hearts (Fig. 5L, M; 50% $n=97$). These data suggested that Pea3 ETS factors are required for proper heart morphogenesis and likely to relay

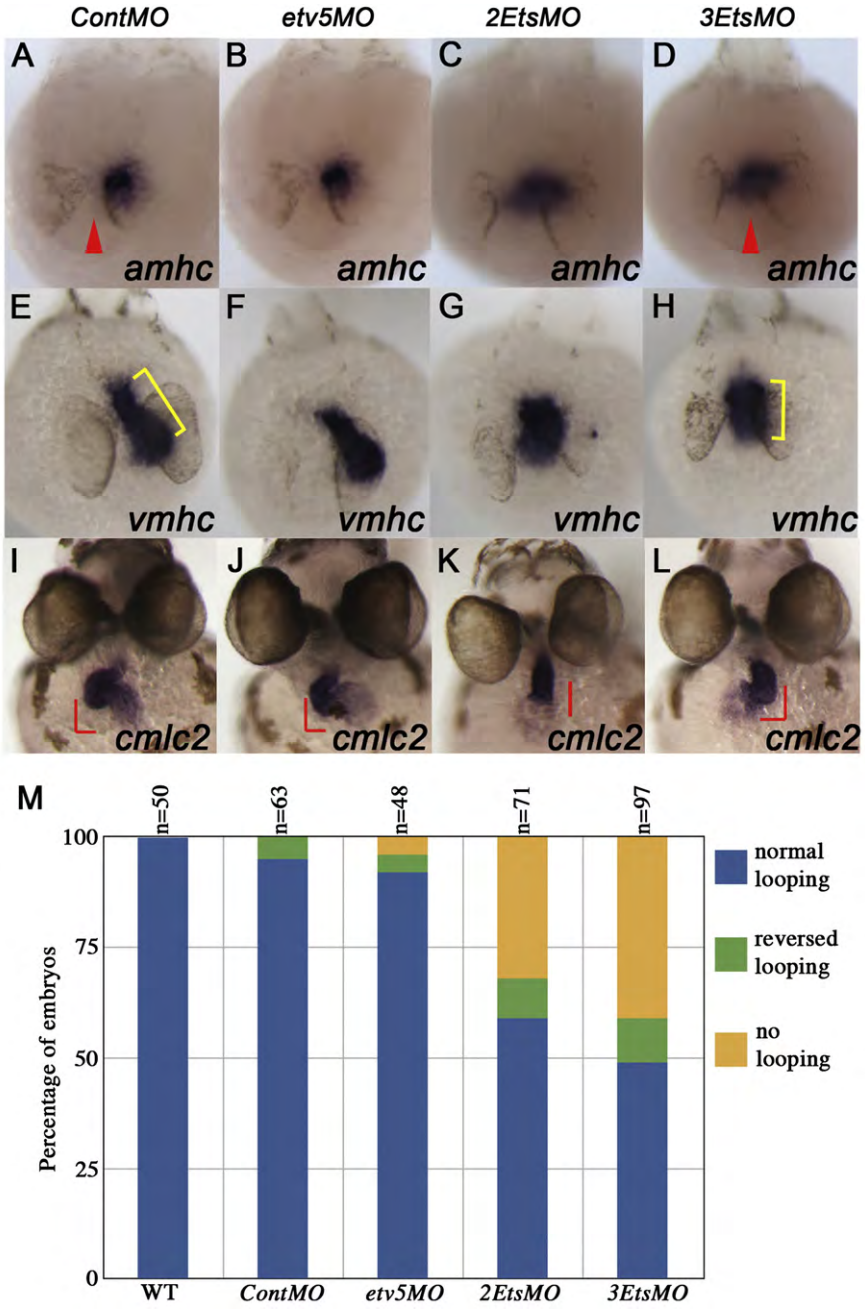


Fig. 5. Pea3 ETS factors are required for heart development. (A–L) In situ hybridization showing cardiac gene expression at 24hpf (A–H) and 48hpf (I–L). Knock-down of multiple Pea3 ETS factors altered *amhc* (D; red arrowhead indicates midline of embryo) and *vmhc* expression (H; cell population and spatial arrangement in yellow bracket). Cardiac looping was severely affected in *3EtsMO*-injected embryos as shown by expression of *cmlc2* (K, L). (M) Quantification of effects on heart looping in MO-injected embryos as exemplified in I; normal looping, K; no looping, L; reverse looping.

Fgf8, Fgf24 and Fgf4 signaling to establish L/R asymmetry (Neugebauer et al., 2009; Yamauchi et al., 2009).

Given the L/R patterning defects we observed in the knock down of Pea3 ETS factors, we next investigated their role in Kupffer's vesicle (KV) formation. The KV is a transient fluid filled organ derived from the dorsal forerunner cells that plays a critical role in establishing L/R polarity (Cooper and D'Amico, 1996; Essner et al., 2005; Kramer Zucker et al., 2005; Melby et al., 1996; Okabe et al., 2008). We have generated a transgenic reporter line *Tg(dusp6:d2EGFP)*^{pt6} that expresses GFP in response to FGF signaling (Molina et al., 2007). In transgenic embryos, GFP is detected in the dorsal forerunner cells and in KV, indicating that FGF signaling is active in these cells. Injection of *3EtsMO* into *Tg(dusp6:d2EGFP)*^{pt6} embryos resulted in decreased GFP fluorescence within the MHB, rhombomere 4 and KV, suggesting that Pea3 ETS proteins are required for GFP reporter gene expression (Fig. 6B). Although GFP expression was extinguished in the KV, the vesicle did form (Fig. 6B′). Therefore, FGF signaling is not required for KV formation. Recent studies have shown a role for FGF signaling in ciliogenesis, especially in the KV (Hong and Dawid, 2009; Neugebauer et al., 2009; Yamauchi et al., 2009). Cilia are required for proper L/R patterning and disrupting their formation or function randomizes L/R polarity. Thus one hypothesis was that cilia number could be affected in Pea3 ETS depleted embryos. Analysis of cilia in the KV of *3EtsMO* injected embryos revealed a significant decrease in cilia number (Fig. 6C, D, E; $p = 5.21 \times 10^{-8}$). To determine if L/R patterning was affected at late somitogenesis stages, we analyzed expression of *spaw*, a gene that is normally expressed in left lateral plate mesoderm (Long et al., 2003). *spaw* expression was often absent, bilateral or right sided in *3EtsMO* injected embryos, consistent with disruption of upstream initiation of the L/R cascade (Fig. 6F J). In addition, expression of *foxj1a*, a transcription factor shown to be involved in ciliogenesis was reduced or absent in *3EtsMO* injected embryos, implicating the importance of ETS transcription factors in cilia formation (Supplementary Fig. 4A, B; 33% with reduced *foxj1a* $n = 112$) (Bonnafe et al., 2004; Brody et al., 2000). A recent study shows a similar decrease in *foxj1a* expression in Fgfr1 depleted embryos (Neugebauer et al., 2009). Thus Pea3 ETS factors are required to relay FGF signaling downstream of multiple ligands to maintain *foxj1a* expression, a prerequisite of ciliogenesis.

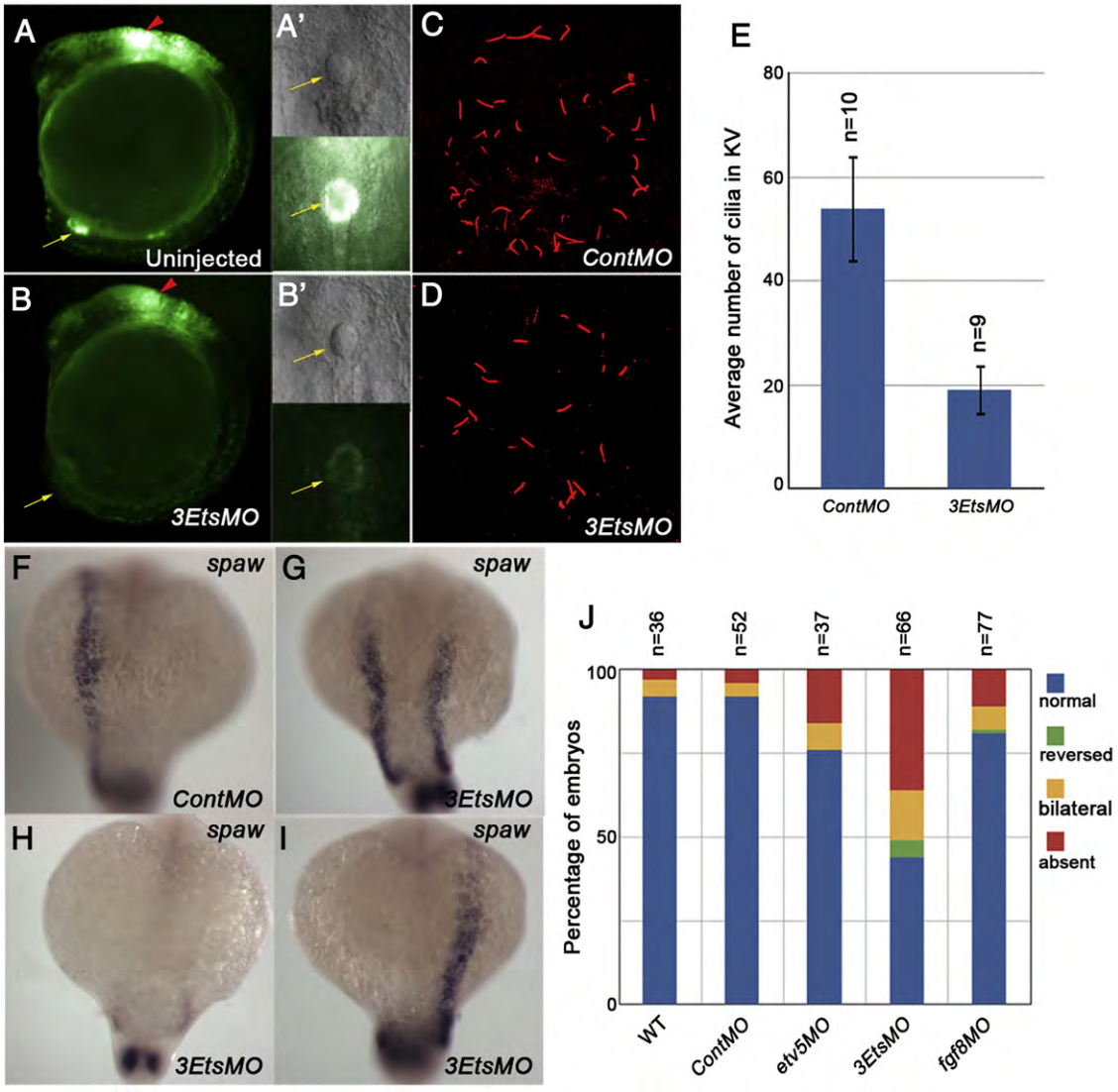


Fig. 6. Pea3 ETS factors are required for ciliogenesis in the Kupffer's vesicle and subsequent L/R patterning. (A, B) *Tg(dusp6:EGFP)*^{pt6} embryos injected with *3EtsMO* show grossly normal formation of Kupffer's vesicle, even though FGF signaling was suppressed (B, B′). (C, D) Cilia number in the KV was reduced in *3EtsMO*-injected embryos at the 12-somite stage. (E) Graph showing average number of cilia in *MO*-injected embryos. Expression of *spaw* in *3EtsMO*-injected embryos reveals L/R defects at 18-somite stage (F–I). (J) Graph depicting L/R defects in Pea3 ETS and Fgf8 depleted embryos.

Identification of FGF regulatory elements

We next determined if FGF target genes contain Pea3 ETS binding sites. For this purpose we analyzed the *Dusp6* promoter and utilized *Xenopus laevis* explant luciferase assays to identify regulatory elements within its upstream promoter sequence. A 10 kb sequence upstream of Exon 1 in *Dusp6* is sufficient to drive reporter expression in an FGF dependent manner (Molina et al., 2007). Luciferase reporter constructs containing varying lengths of upstream sequence (5Kb to 1Kb) were co injected into each cell of a 2 cell stage *Xenopus* embryo together with *fgf8* mRNA (Fig. 7A). A 3 fold increase in luciferase activity was observed in the presence of Fgf8 ($n>4$ for each construct, Fig. 7B and data not shown). Since these results indicated the presence of FGF responsive elements within 1Kb upstream of the transcription start site, we aligned this genomic region from several species. Significant homology was identified among all four vertebrate promoters studied, consistent with a recent report (Fig. 7C) (Ekerot et al., 2008). Several putative transcription factor binding sites were found, including two sites for Pea3 ETS proteins, Pea3A and Pea3B, that were also described in the mouse *Dusp6* promoter (Fig. 7C) (Ekerot et al., 2008). We mutated these putative Pea3 binding sites and found that the Pea3B site, which contains three perfectly conserved core 5′ GGAA3′ motifs, was critical for the FGF mediated induction of luciferase activity (Fig. 7B, C; MutA and MutB, sequences of the mutations in red). To determine if Etv5 can regulate the *Dusp6* promoter directly, we injected *etv5:VP16* with either a wildtype 2.2Kb *Dusp6* reporter or a Pea3 mutant reporter construct (MutA or MutB). Ectopic expression of *etv5:VP16* caused a dramatic increase in luciferase activity with the wildtype and MutA *Dusp6* reporters, whereas the MutB reporter exhibited little induction (Fig. 7D). Furthermore, *etv5:T135D* increased luciferase activity and was dependant on the Pea3B sequence (Fig. 7E). These results indicated that FGF signaling directly regulates *Dusp6* transcription through the activity of Pea3 ETS factors.

To determine if Pea3 ETS factor function is conserved between zebrafish and mouse, we co transfected mouse Etv4 or Etv5 expression plasmids with the zebrafish reporter constructs into 293 T cells. Both ETV4 and ETV5 were capable of inducing reporter gene expression in 293 T cells, and this was dependent on the Pea3B site in 293 T cells (Supplementary Fig. 5). Furthermore, we isolated the mouse *Dusp6* promoter and generated luciferase reporter constructs with 0.6Kb of upstream promoter sequence. Ectopic expression of mouse ETV4 or ETV5 transactivated the promoter, and this activity was also dependent on the exact Pea3B sequence found in the zebrafish *Dusp6* promoter (Supplementary Fig. 6). These results reveal the evolutionary conser vation of Pea3 ETS factors as proteins from either mouse or zebrafish can activate reporter gene expression driven by either promoter.

Pea3 ETS protein directly binds to the Dusp6 promoter

Using electrophoretic mobility shift assays (EMSAs) *in vitro* and by chromatin immunoprecipitation (ChIP) assays *in vivo* we asked whether Pea3 ETS factors could bind to the *Dusp6* promoter. We generated biotin labeled oligos that flank the Pea3B site (60 bp) and amplified this region by PCR (Fig. 8A). Recombinant GST Etv5:ETS domain protein interacted with the biotin labeled *Dusp6* promoter DNA (Fig. 8D; lane 2). The bound DNA was competed by increasing amounts of non biotin labeled promoter (Fig. 8D; lanes 3 7). Likewise, the Pea3B 18mer oligo (Fig. 8C) also competed Etv5 binding (Fig. 8D; lanes 8 12), while a random sequence used as control competitor did not compete with the biotin labeled *Dusp6* fragment (Fig. 8D; lanes 13 17). To verify the importance of the Pea3B site for Etv5 binding, a 60 bp biotin labeled Pea3B mutant *Dusp6* promoter was used in EMSAs (Fig. 8B). Etv5 ETS protein did not shift the Pea3B mutant DNA (Fig. 8D; lanes 18, 19).

To determine if Pea3 ETS factors can directly bind to the mouse *Dusp6* promoter *in vivo*, we performed ChIP assays on isolated pharyngeal tissue from E9.5 mouse embryos, a region with active FGF signaling that expresses high levels of *Etv4*, *Etv5* and *Dusp6* (Dickinson et al., 2002; Ilagan et al., 2006; Park et al., 2006). By ChIP and PCR, we found that the Pea3B binding site was enriched with specific ETV4 antibodies, suggesting that ETV4 directly binds to the *Dusp6* promoter *in vivo* (Fig. 9A C). qPCR quantitation revealed a 7 fold enrichment of this element in the ETV4 ChIPed sample compared to negative control (Fig. 9D). Taken together, we show that *Dusp6* is directly regulated by Pea3 ETS factors during zebrafish and mouse embryonic development.

Discussion

The role of Pea3 ETS factors in FGF mediated developmental processes

In this report we show that Pea3 ETS factors function to mediate the transcriptional response in the FGF pathway to directly activate *dusp6* expression. We have identified putative ERK phosphorylation sites in Etv5 and show that introducing phosphomimetic mutations into these sites resulted in constitutively active ETS transcription factors. A similar putative MAPK phosphorylation site was identified in Pea3, suggesting that all three factors are regulated by the FGF/RAS/MAPK pathway (Guo and Sharrocks, 2009). Antisense MO injections blocking the translation of *erm*, *etv5*, and *pea3* resulted in an MHB defect in the zebrafish embryo, as well as suppression of FGF target genes. The importance of these ETS factors in heart develop ment was revealed by the loss of cardiac progenitors in Pea3 ETS factor depleted zebrafish embryos, consistent with a critical role for these factors downstream of FGFs in mice. In zebrafish, experimental manipulations that expand cardiac progenitors negatively affected the blood and vessel lineages, and vice versa (Molina et al., 2009; Schoenebeck et al., 2007). Given that Fgf8 and Pea3 ETS factors play an important role in maintaining cardiac identity, we reasoned that the knock down of Pea3 ETS factors might expand endothelial lineages. Indeed, the concerted depletion of Pea3 ETS factors resulted in an increase in blood and vessel progenitors as marked by expanded *scl* expression. Interestingly, one ETS factor, Etv5, was shown to be critical to repress *scl* expression, indicating that, although Pea3 ETS factors appear to function redundantly in most FGF mediated developmental processes, this specific factor appears to be solely responsible to restrict endothelial expansion within the ALPM. Thus, FGF activity is required to maintain cardiac progenitor populations and suppress endothelial differentiation within the caudal domain of the ALPM.

Studies have highlighted a role for FGF signaling in L/R patterning of the zebrafish embryo (Albertson and Yelick, 2005; Hong and Dawid, 2009; Neugebauer et al., 2009; Yamauchi et al., 2009). The knock down of either Fgf ligands or two effectors of the FGF pathway did not disturb formation of the KV, but did disrupt motile cilia formation (Hong and Dawid, 2009; Neugebauer et al., 2009; Yamauchi et al., 2009). In this study, depletion of Pea3 ETS factors did not disrupt KV formation, but cilia number was markedly reduced. This is responsible for the laterality defects such as alteration of lateral *spaw* expression and both the cardiac jogging and looping defects. Interestingly, among these studies, the role of FGF signaling in proper motile cilia formation within the KV has ranged from contributing to cilia length (Neugebauer et al., 2009; Yamauchi et al., 2009) to contributing to the total number of motile cilia in the KV (Hong and Dawid, 2009). Because of this, it can be hypothesized that the specificity in patterning of KV motile cilia may involve several different Fgf ligands and/or transcription factors. In this study, we have revealed a role for Pea3 ETS transcription factors in ciliogenesis within KV.

Pea3 ETS factors directly bind to the Dusp6 promoter in vivo

How FGF signaling is relayed into a transcriptional response in development is not fully defined. Because *Dusp6* is a known target of FGF activity, we focused on identifying cis elements within the *Dusp6* promoter to determine if Pea3 ETS factors can directly regulate its

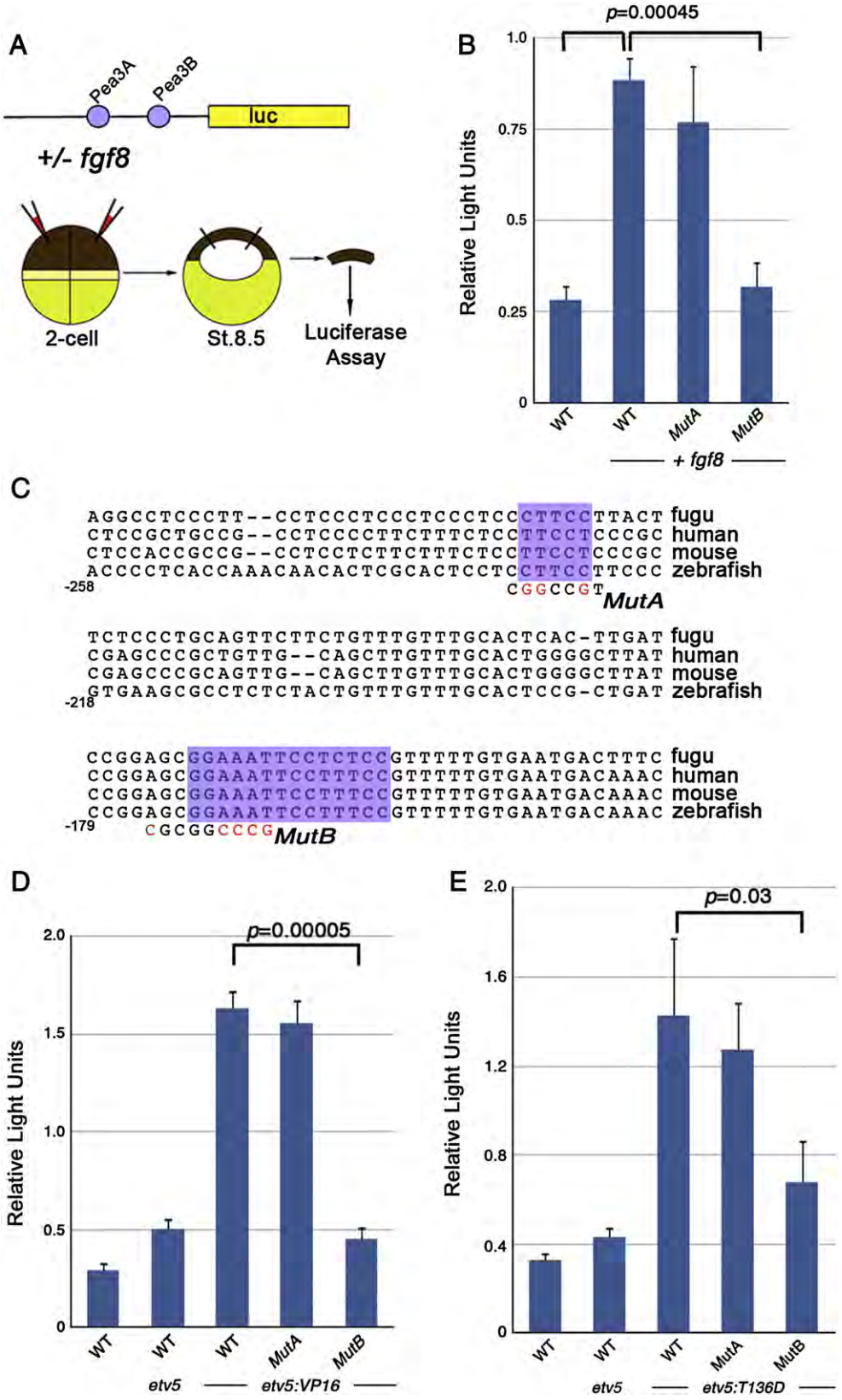


Fig. 7. *fgf8* and *etv5* activates *Dusp6* promoter. (A) Diagram showing the *Dusp6* promoter assay in *Xenopus* animal caps. (B) Luciferase activity measured with 2 kb *Dusp6* promoter (WT) constructs in the presence or absence of *fgf8*, indicating the requirement for the Pea3B site (MutB). (C) Alignment of *Dusp6* promoter from several vertebrate species showing conserved putative ETS binding sites (highlighted in purple) that were mutated as shown in red below the sequence. (D) *etv5:VP16* could activate the *Dusp6* reporter (WT), and mutation of the Pea3B (MutB) site diminished this activity. (E) Ectopic expression of *etv5:T135D* activated the *Dusp6* reporter (WT), but not the mutant Pea3B reporter (MutB).

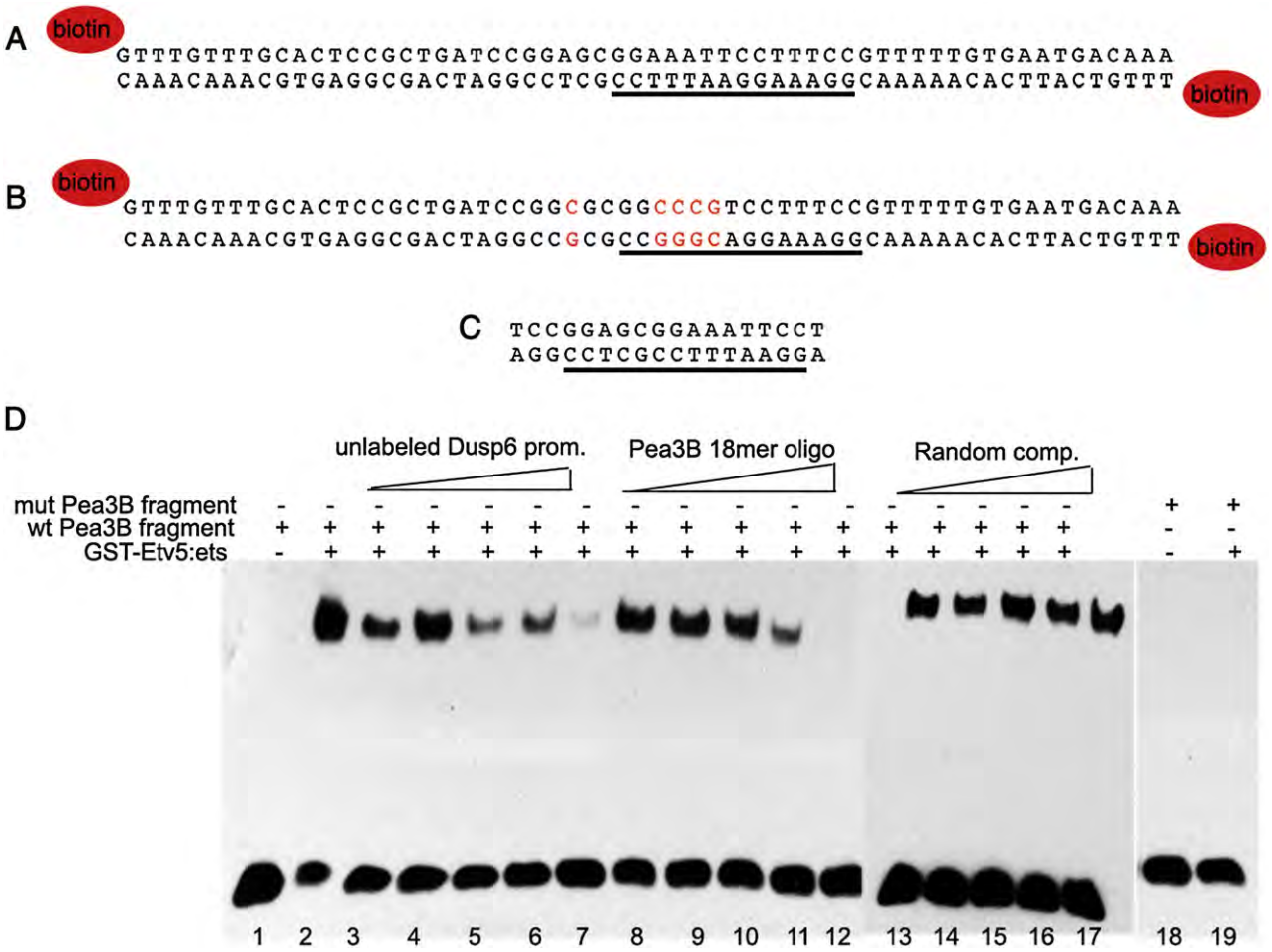


Fig. 8. Etv5-ETS domain binds to the *Dusp6* promoter. (A) Biotin labeled PCR product. (B) Biotin labeled mutant Pea3B *Dusp6* promoter fragment. (C) Pea3B 18 bp fragment used in competition assays. Underlined sequence represent the putative Pea3B binding site (D) EMSA using 60 bp *Dusp6* promoter biotin fragment and GST-Etv5:ETS protein. The Etv5-ETS domain bound to the 60 bp *Dusp6* promoter containing the Pea3B site (lane 2). This binding was competed with the non-biotin labeled 60 bp *Dusp6* promoter (lanes 3–7) and the Pea3B 18 bp fragment (lanes 8–12). A random oligonucleotide sequence did not compete this binding (lanes 13–17). Binding of Etv5-ETS to the Pea3B mutant *Dusp6* promoter was not observed (lanes 18–19).

transcription (Molina et al., 2007; Tsang et al., 2004). We identified several conserved putative ETS binding sites within the *Dusp6* promoter and through reporter and binding assays we indentified one particular site that was critical for reporter gene activity and *in vitro* binding. This site is identical to what has been recently described in mouse *Dusp6* promoter (Ekerot et al., 2008). Ekerot et al., (2008) had demonstrated by ChIP studies in NIH3T3 cells that both Ets1 and Ets2 can bind to this particular region. Our studies demonstrate for the first time that Pea3 can bind to the *Dusp6* promoter *in vivo* and provide strong evidence that FGF signaling regulates *Dusp6* transcription by Pea3 ETS factors during development. These differences may reflect varying expression levels of these ETS factors between NIH3T3 cells and mouse pharyngeal tissue. Since *Fgf8*, *Etv4*, *Etv5* and *Dusp6* are strongly expressed in pharyngeal arch, our results provide strong evidence for the direct binding of ETV4 to the *Dusp6* promoter (Dickinson et al., 2002; Ilagan et al., 2006; Park et al., 2006). In contrast, both Ets1 and Ets2 transcripts are predominantly expressed in vascular and lymphatic tissues and not co expressed with *Dusp6* (Maroulakou et al., 1994). However, in NIH3T3 cells, expression of Ets1 and Ets2 may play a greater role in regulating *Dusp6* expression *in vitro*.

Redundant function of Pea3 ETS transcription factors during development

Etv5 or *Etv4* gene knock out approaches in mice resulted in developmentally normal animals that survived to adulthood, but developed motor neuron differentiation defects, spermatogonial stem cell renewal defects, and movement disorders (Arber et al., 2000; Chen et al., 2005; Livet et al., 2002). The lack of gross developmental phenotypes in these knock out mice suggested these genes may function redundantly in mediating FGF signaling. To circumvent redundancy among Pea3 factors, engrailed repressor fusion constructs were ectopically expressed in both mouse and chick embryos (Brent and Tabin, 2004; Liu et al., 2003). The repression of multiple Pea3 ETS family members using this technique overcame redundancy to show a role for these factors in somite and lung development (Brent and Tabin, 2004; Liu et al., 2003). Further evidence of their redundant function was shown by ectopic expression of a dominant negative version of *Etv5* in the mouse limb which resulted in shortened limbs and polydactyly (Mao et al., 2009). This was confirmed by the knock out of both *Etv4* and *Etv5* in early mesodermal lineages that also resulted in limb and digit defects (Zhang et al., 2009). In our studies, antisense morpholinos allowed the targeted depletion of multiple members of the Pea3 family. In general, knocking down only one member of this family rarely resulted in developmental defects, but the knock down of two or three members resulted in distinct phenotypes.

Functional redundancy within another ETS sub family of transcription factors in zebrafish has also been described (Pham et al., 2007). This study examined four ETS family members expressed in the vasculature: *fli1*, *fli1b*, *ets1*, and *etsrp* (Pham et al., 2007). Using an

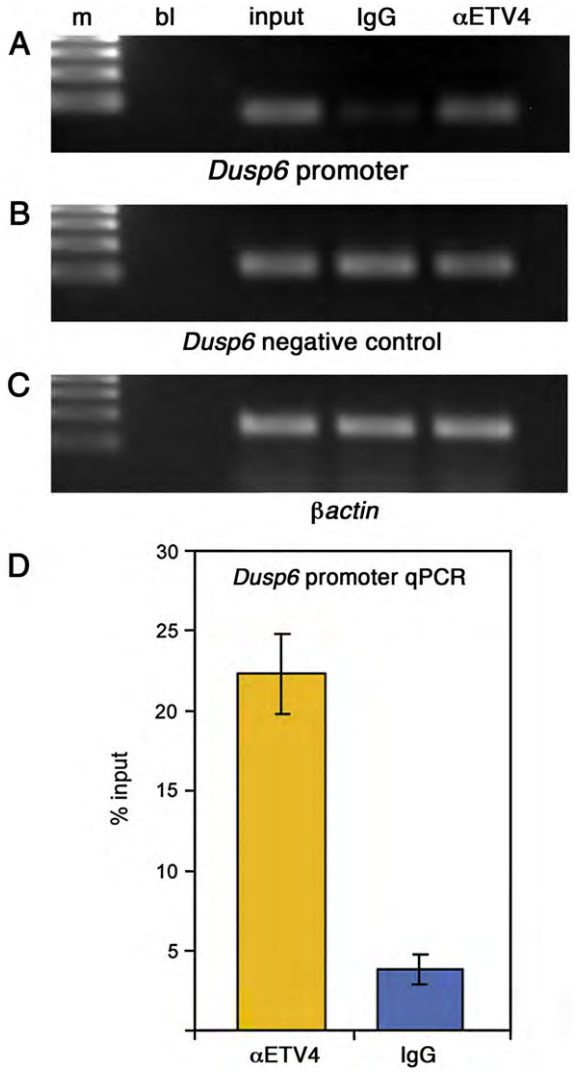


Fig. 9. ETV4 directly binds to *Dusp6* promoter *in vivo* in mouse. The mouse *Dusp6* promoter region bearing highly conserved PEA3 binding sites is enriched in chromatin immunoprecipitated from mouse embryonic tissues with anti-mouse ETV4 antibody. (A) Agarose gel showing PCR products obtained from equal amounts of input, mouse IgG ChIPed, and anti-ETV4 ChIPed DNAs with the *Dusp6* promoter ("Dusp6 prom") primers that amplify the region containing the putative PEA3 binding sites. (B). Agarose gel showing PCR products obtained on the same DNA samples as in A with *Dusp6* negative control primers. (C) Agarose gel showing PCR products obtained on the same DNA samples as in A with *β-actin* negative control primers. Only the *Dusp6* promoter was differentially precipitated by the anti-ETV4 antibody; B and C negative control primers give identical bands as they detect the non-specific, background precipitated DNA. (D) Graphical representation of quantitative PCR results show nearly 7-fold enrichment of the amplicon containing the highly conserved region of the *Dusp6* promoter in ETV4 ChIPed DNA compared to IgG negative control expressed relative to that detected in input. m, marker; bl, blank lane.

antisense morpholino approach to knock down expression of all four genes, both vascular and hematopoietic development was disrupted, showing the importance of these genes for vessel sprouting and circulation. Interestingly, a hierarchy was observed, whereby the knock down of *estrp* showed stronger phenotypes than a single knock down of *fli1*, *fli1b*, or *ets1* (Pham et al., 2007). Similarly, a reduction of *etv5* was shown in our study to have a strong effect on the *scl* expression domain, indicating a hierarchy may also exist between members of the Pea3 ETS factors. In conclusion, we have defined the importance of Pea3 ETS transcription factors in mediating FGF signaling during development.

Our data suggest that in some instances, the knockdown of two Pea3 ETS proteins was sufficient to reveal a phenotype in a majority of the injected embryos, as noted with the expression of cardiac progenitor markers, implicating an overlapping function between these factors. Given that knockdown studies can be difficult to quantitate, a more effective means is to generate genetic nulls and determine the relative contribution of each Pea3 ETS factor to FGF signaling and development. One interesting observation that demonstrates ETS overlapping roles comes from a recent study in mouse kidney formation. Compound heterozygotes of two factors (*Etv4+/−; Etv5+/−*) or *Etv4* nulls did result in a small percentage of embryos showing renal agenesis. In contrast, *Etv4−/−; Etv5+/−* mice exhibited a complete lack of kidneys, suggesting that a full compliment of Etv5 can function to restore normal kidney development, but in a reduced state, it can not (Lu et al., 2009). Thus gene dosage may play a critical role in how these factors bind to promoters and regulate transcription and development. A more detailed gene expression profiling and readout is required to correlate gene activity with developmental outcome.

Acknowledgments

We are thankful to the Hukriede lab for assistance with the *Xenopus* experiments, R. Burdine, S. Sanker, and P.K. Umasankar for help in imaging cilia. We would like to thank D. Yelon, B. Roman, N. Bahary, and S.K. Hong for reagents, and to Michael Rebagliati for critical reading of the manuscript and experimental suggestions. The project described was supported in part by Award Number R01HL088016 to M.T. from the National Heart, Lung, and Blood Institute (NHLBI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI. This work was also supported by the American Heart Association (0565400U) to MT, University of Pittsburgh CMRF, and by the intramural program of the NICHD, NIH. AMM's laboratory is supported by NICHD HD044157.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2010.03.011.

References

Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K., Meyers, E.N., 2002. Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. Development 129, 4613–4625.
Albertson, R.C., Yelick, P.C., 2005. Roles for fgf8 signaling in left–right patterning of the visceral organs and craniofacial skeleton. Dev. Biol. 283, 310–321.
Arber, S., Ladle, D.R., Lin, J.H., Frank, E., Jessell, T.M., 2000. ETS gene Er81 controls the formation of functional connections between group Ia sensory afferents and motor neurons. Cell 101, 485–498.
Bally-Cuif, L., Goutel, C., Wassef, M., Wurst, W., Rosa, F., 2000. Coregulation of anterior and posterior mesendodermal development by a hairy-related transcriptional repressor. Genes Dev. 14, 1664–1677.
Berdougo, E., Coleman, H., Lee, D.H., Stainier, D.Y., Yelon, D., 2003. Mutation of weak atrium/atrial myosin heavy chain disrupts atrial function and influences ventricular morphogenesis in zebrafish. Development 130, 6121–6129.
Bonnafe, E., Touka, M., AitLounis, A., Baas, D., Barras, E., Ucla, C., Moreau, A., Flamant, F., Dubruille, R., Couble, P., Collignon, J., Durand, B., Reith, W., 2004. The transcription factor RFX3 directs nodal cilium development and left–right asymmetry specification. Mol. Cell. Biol. 24, 4417–4427.
Brent, A.E., Tabin, C.J., 2004. FGF acts directly on the somitic tendon progenitors through the Ets transcription factors Pea3 and Erm to regulate scleraxis expression. Development 131, 3885–3896.
Brody, S.L., Yan, X.H., Wuerffel, M.K., Song, S.K., Shapiro, S.D., 2000. Ciliogenesis and left–right axis defects in forkhead factor HFH-4-null mice. Am. J. Respir. Cell Mol. Biol. 23, 45–51.
Brown, L.A., Amores, A., Schilling, T.F., Jowett, T., Baert, J.L., de Launoit, Y., Sharrocks, A. D., 1998. Molecular characterization of the zebrafish PEA3 ETS-domain transcription factor. Oncogene 17, 93–104.
Chen, J.N., Fishman, M.C., 1996. Zebrafish tinman homolog demarcates the heart field and initiates myocardial differentiation. Development 122, 3809–3816.
Chen, C., Ouyang, W., Grigura, V., Zhou, Q., Carnes, K., Lim, H., Zhao, G.Q., Arber, S., Kurpios, N., Murphy, T.L., Cheng, A.M., Hassell, J.A., Chandrashekar, V., Hofmann, M. C., Hess, R.A., Murphy, K.M., 2005. ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature 436, 1030–1034.

Cooper, M.S., D'Amico, L.A., 1996. A cluster of noninvoluting endocytic cells at the margin of the zebrafish blastoderm marks the site of embryonic shield formation. Dev. Biol. 180, 184–198.
Dailey, L., Ambrosetti, D., Mansukhani, A., Basilico, C., 2005. Mechanisms underlying differential responses to FGF signaling. Cytokine Growth Factor Rev. 16, 233–247.
Dickinson, R.J., Eblaghie, M.C., Keyse, S.M., Morriss-Kay, G.M., 2002. Expression of the ERK-specific MAP kinase phosphatase PYST1/MKP3 in mouse embryos during morphogenesis and early organogenesis. Mech. Dev. 113, 193–196.
Draper, B.W., Morcos, P.A., Kimmel, C.B., 2001. Inhibition of zebrafish fgf8 pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. Genesis 30, 154–156.
Draper, B.W., Stock, D.W., Kimmel, C.B., 2003. Zebrafish fgf24 functions with fgf8 to promote posterior mesodermal development. Development 130, 4639–4654.
Ekerot, M., Stavridis, M.P., Delavaine, L., Mitchell, M.P., Staples, C., Owens, D.M., Keenan, I.D., Dickinson, R.J., Storey, K.G., Keyse, S.M., 2008. Negative-feedback regulation of FGF signalling by DUSP6/MKP-3 is driven by ERK1/2 and mediated by Ets factor binding to a conserved site within the DUSP6/MKP-3 gene promoter. Biochem. J. 412, 287–298.
Essner, J.J., Amack, J.D., Nyholm, M.K., Harris, E.B., Yost, H.J., 2005. Kupffer's vesicle is a ciliated organ of asymmetry in the zebrafish embryo that initiates left–right development of the brain, heart and gut. Development 132, 1247–1260.
Frank, D.U., Fotheringham, L.K., Brewer, J.A., Muglia, L.J., Tristani-Firouzi, M., Capecchi, M.R., Moon, A.M., 2002. An Fgf8 mouse mutant phenocopies human 22q11 deletion syndrome. Development 129, 4591–4603.
Furthauer, M., Reifers, F., Brand, M., Thisse, B., Thisse, C., 2001. sprouty4 acts in vivo as a feedback-induced antagonist of FGF signaling in zebrafish. Development 128, 2175–2186.
Furthauer, M., Lin, W., Ang, S.L., Thisse, B., Thisse, C., 2002. Sef is a feedback-induced antagonist of Ras/MAPK-mediated FGF signalling. Nat. Cell Biol. 4, 170–174.
Furthauer, M., Van Celst, J., Thisse, C., Thisse, B., 2004. Fgf signalling controls the dorsoventral patterning of the zebrafish embryo. Development 131, 2853–2864.
Guo, B., Sharrocks, A.D., 2009. Extracellular signal-regulated kinase mitogen-activated protein kinase signaling initiates a dynamic interplay between sumoylation and ubiquitination to regulate the activity of the transcriptional activator PEA3. Mol. Cell. Biol. 29, 3204–3218.
Hacohen, N., Kramer, S., Sutherland, D., Hiromi, Y., Krasnow, M.A., 1998. sprouty encodes a novel antagonist of FGF signaling that patterns apical branching of the Drosophila airways. Cell 92, 253–263.
Herzog, W., Sonntag, C., von der Hardt, S., Roehl, H.H., Varga, Z.M., Hammerschmidt, M., 2004. Fgf3 signaling from the ventral diencephalon is required for early specification and subsequent survival of the zebrafish adenohypophysis. Development 131, 3681–3692.
Hong, S.K., Dawid, I.B., 2009. FGF-dependent left–right asymmetry patterning in zebrafish is mediated by Ier2 and Fibp1. Proc. Natl. Acad. Sci. U. S. A. 106, 2230–2235.
Hwang, S.P., Tsou, M.F., Lin, Y.C., Liu, C.H., 1997. The zebrafish BMP4 gene: sequence analysis and expression pattern during embryonic development. DNA Cell Biol. 16, 1003–1011.
Ilagan, R., Abu-Issa, R., Brown, D., Yang, Y.P., Jiao, K., Schwartz, R.J., Klingensmith, J., Meyers, E.N., 2006. Fgf8 is required for anterior heart field development. Development 133, 2435–2445.
Itoh, N., 2007. The Fgf families in humans, mice, and zebrafish: their evolutional processes and roles in development, metabolism, and disease. Biol. Pharm. Bull. 30, 1819–1825.
Kawakami, Y., Rodriguez-Leon, J., Koth, C.M., Buscher, D., Itoh, T., Raya, A., Ng, J.K., Esteban, C.R., Takahashi, S., Henrique, D., Schwarz, M.F., Asahara, H., Izpisua Belmonte, J.C., 2003. MKP3 mediates the cellular response to FGF8 signalling in the vertebrate limb. Nat. Cell Biol. 5, 513–519.
Keegan, B.R., Meyer, D., Yelon, D., 2004. Organization of cardiac chamber progenitors in the zebrafish blastula. Development 131, 3081–3091.
Kiefer, P., Mathieu, M., Mason, I., Dickson, C., 1996. Secretion and mitogenic activity of zebrafish FGF3 reveal intermediate properties relative to mouse and Xenopus homologues. Oncogene 12, 1503–1511.
Kobberup, S., Nyeng, P., Juhl, K., Hutton, J., Jensen, J., 2007. ETS-family genes in pancreatic development. Dev. Dyn. 236, 3100–3110.
Kramer, S., Okabe, M., Hacohen, N., Krasnow, M.A., Hiromi, Y., 1999. Sprouty: a common antagonist of FGF and EGF signaling pathways in Drosophila. Development 126, 2515–2525.
Kramer-Zucker, A.G., Olale, F., Haycraft, C.J., Yoder, B.K., Schier, A.F., Drummond, I.A., 2005. Cilia-driven fluid flow in the zebrafish pronephros, brain and Kupffer's vesicle is required for normal organogenesis. Development 132, 1907–1921.
Krauss, S., Johansen, T., Korzh, V., Fjose, A., 1991. Expression of the zebrafish paired box gene pax[zf-b] during early neurogenesis. Development 113, 1193–1206.
Kudoh, T., Tsang, M., Hukriede, N.A., Chen, X., Dedekian, M., Clarke, C.J., Kiang, A., Schultz, S., Epstein, J.A., Toyama, R., Dawid, I.B., 2001. A gene expression screen in zebrafish embryogenesis. Genome Res. 11, 1979–1987.
Laget, M.P., Defossez, P.A., Albagli, O., Baert, J.L., Dewitte, F., Stehelin, D., de Launoit, Y., 1996. Two functionally distinct domains responsible for transactivation by the Ets family member ERM. Oncogene 12, 1325–1336.
Lee, R.K., Stainier, D.Y., Weinstein, B.M., Fishman, M.C., 1994. Cardiovascular development in the zebrafish. II. Endocardial progenitors are sequestered within the heart field. Development 120, 3361–3366.
Liao, E.C., Paw, B.H., Oates, A.C., Pratt, S.J., Postlethwait, J.H., Zon, L.I., 1998. SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish. Genes Dev. 12, 621–626.
Liu, Y., Jiang, H., Crawford, H.C., Hogan, B.L., 2003. Role for ETS domain transcription factors Pea3/Erm in mouse lung development. Dev. Biol. 261, 10–24.
Livet, J., Sigrist, M., Stroebel, S., De Paola, V., Price, S.R., Henderson, C.E., Jessell, T.M., Arber, S., 2002. ETS gene Pea3 controls the central position and terminal arborization of specific motor neuron pools. Neuron 35, 877–892.
Long, S., Ahmad, N., Rebagliati, M., 2003. The zebrafish nodal-related gene southpaw is required for visceral and diencephalic left–right asymmetry. Development 130, 2303–2316.
Lu, B.C., Cebrian, C., Chi, X., Kuure, S., Kuo, R., Bates, C.M., Arber, S., Hassell, J., MacNeil, L., Hoshi, M., Jain, S., Asai, N., Takahashi, M., Schmidt-Ott, K.M., Barasch, J., D'Agati, V., Costantini, F., 2009. Etv4 and Etv5 are required downstream of GDNF and Ret for kidney branching morphogenesis. Nat. Genet. 41, 1295–1302.
Macatee, T.L., Hammond, B.P., Arenkiel, B.R., Francis, L., Frank, D.U., Moon, A.M., 2003. Ablation of specific expression domains reveals discrete functions of ectoderm- and endoderm-derived FGF8 during cardiovascular and pharyngeal development. Development 130, 6361–6374.
Mao, J., McGlinn, E., Huang, P., Tabin, C.J., McMahon, A.P., 2009. Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb. Dev. Cell 16, 600–606.
Maroulakou, I.G., Papas, T.S., Green, J.E., 1994. Differential expression of ets-1 and ets-2 proto-oncogenes during murine embryogenesis. Oncogene 9, 1551–1565.
Marques, S.R., Lee, Y., Poss, K.D., Yelon, D., 2008. Reiterative roles for FGF signaling in the establishment of size and proportion of the zebrafish heart. Dev. Biol. 321, 397–406.
Melby, A.E., Warga, R.M., Kimmel, C.B., 1996. Specification of cell fates at the dorsal margin of the zebrafish gastrula. Development 122, 2225–2237.
Meyers, E.N., Martin, G.R., 1999. Differences in left–right axis pathways in mouse and chick: functions of FGF8 and SHH. Science 285, 403–406.
Meyers, E.N., Lewandoski, M., Martin, G.R., 1998. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. Nat. Genet. 18, 136–141.
Molina, G.A., Watkins, S.C., Tsang, M., 2007. Generation of FGF reporter transgenic zebrafish and their utility in chemical screens. BMC Dev. Biol. 7, 62.
Molina, G., Vogt, A., Bakan, A., Dai, W., Queiroz de Oliveira, P., Znosko, W., Smithgall, T.E., Bahar, I., Lazo, J.S., Day, B.W., Tsang, M., 2009. Zebrafish chemical screening reveals an inhibitor of Dusp6 that expands cardiac cell lineages. Nat. Chem. Biol. 5, 680–687.
Munchberg, S.R., Ober, E.A., Steinbeisser, H., 1999. Expression of the Ets transcription factors erm and pea3 in early zebrafish development. Mech. Dev. 88, 233–236.
Neugebauer, J.M., Amack, J.D., Peterson, A.G., Bisgrove, B.W., Yost, H.J., 2009. FGF signalling during embryo development regulates cilia length in diverse epithelia. Nature 458, 651–654.
Nutt, S.L., Dingwell, K.S., Holt, C.E., Amaya, E., 2001. Xenopus Sprouty2 inhibits FGF-mediated gastrulation movements but does not affect mesoderm induction and patterning. Genes Dev. 15, 1152–1166.
O'Hagan, R.C., Tozer, R.G., Symons, M., McCormick, F., Hassell, J.A., 1996. The activity of the Ets transcription factor PEA3 is regulated by two distinct MAPK cascades. Oncogene 13, 1323–1333.
Oikawa, T., Yamada, T., 2003. Molecular biology of the Ets family of transcription factors. Gene 303, 11–34.
Okabe, N., Xu, B., Burdine, R.D., 2008. Fluid dynamics in zebrafish Kupffer's vesicle. Dev. Dyn. 237, 3602–3612.
Ornitz, D.M., Itoh, N., 2001. Fibroblast growth factors. Genome Biol. 2 (3), REVIEWS3005.
Park, E.J., Ogden, L.A., Talbot, A., Evans, S., Cai, C.L., Black, B.L., Frank, D.U., Moon, A.M., 2006. Required, tissue-specific roles for Fgf8 in outflow tract formation and remodeling. Development 133, 2419–2433.
Park, E.J., Watanabe, Y., Smyth, G., Miyagawa-Tomita, S., Meyers, E., Klingensmith, J., Camenisch, T., Buckingham, M., Moon, A.M., 2008. An FGF autocrine loop initiated in second heart field mesoderm regulates morphogenesis at the arterial pole of the heart. Development 135, 3599–3610.
Pham, V.N., Lawson, N.D., Mugford, J.W., Dye, L., Castranova, D., Lo, B., Weinstein, B.M., 2007. Combinatorial function of ETS transcription factors in the developing vasculature. Dev. Biol. 303, 772–783.
Picker, A., Brennan, C., Reifers, F., Clarke, J.D., Holder, N., Brand, M., 1999. Requirement for the zebrafish mid–hindbrain boundary in midbrain polarisation, mapping and confinement of the retinotectal projection. Development 126, 2967–2978.
Powers, C.J., McLeskey, S.W., Wellstein, A., 2000. Fibroblast growth factors, their receptors and signaling. Endocr. Relat. Cancer 7, 165–197.
Raible, F., Brand, M., 2001. Tight transcriptional control of the ETS domain factors Erm and Pea3 by Fgf signaling during early zebrafish development. Mech. Dev. 107, 105–117.
Reifers, F., Bohli, H., Walsh, E.C., Crossley, P.H., Stainier, D.Y., Brand, M., 1998. Fgf8 is mutated in zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. Development 125, 2381–2395.
Reifers, F., Walsh, E.C., Leger, S., Stainier, D.Y., Brand, M., 2000. Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). Development 127, 225–235.
Reiter, J.F., Alexander, J., Rodaway, A., Yelon, D., Patient, R., Holder, N., Stainier, D.Y., 1999. Gata5 is required for the development of the heart and endoderm in zebrafish. Genes Dev. 13, 2983–2995.
Roehl, H., Nusslein-Volhard, C., 2001. Zebrafish pea3 and erm are general targets of FGF8 signaling. Curr. Biol. 11, 503–507.
Roussigne, M., Blader, P., 2006. Divergence in regulation of the PEA3 family of ETS transcription factors. Gene Expr. Patterns 6, 777–782.
Schoenebeck, J.J., Keegan, B.R., Yelon, D., 2007. Vessel and blood specification override cardiac potential in anterior mesoderm. Dev. Cell 13, 254–267.
Schulte-Merker, S., Lee, K.J., McMahon, A.P., Hammerschmidt, M., 1997. The zebrafish organizer requires chordino. Nature 387, 862–863.
Sekine, K., Ohuchi, H., Fujiwara, M., Yamasaki, M., Yoshizawa, T., Sato, T., Yagishita, N., Matsui, D., Koga, Y., Itoh, N., Kato, S., 1999. Fgf10 is essential for limb and lung formation. Nat. Genet. 21, 138–141.
Serbedzija, G.N., Chen, J.N., Fishman, M.C., 1998. Regulation in the heart field of zebrafish. Development 125, 1095–1101.
Sharrocks, A.D., 2001. The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. 2, 827–837.

Stainier, D.Y., Lee, R.K., Fishman, M.C., 1993. Cardiovascular development in the zebrafish. I. Myocardial fate map and heart tube formation. Development 119, 31–40.
Sun, X., Meyers, E.N., Lewandoski, M., Martin, G.R., 1999. Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. Genes Dev. 13, 1834–1846.
Tefft, D., Lee, M., Smith, S., Crowe, D.L., Bellusci, S., Warburton, D., 2002. mSprouty2 inhibits FGF10-activated MAP kinase by differentially binding to upstream target proteins. Am. J. Physiol. Lung Cell. Mol. Physiol. 283, L700–L706.
Thisse, B., Thisse, C., 2005. Functions and regulations of fibroblast growth factor signaling during embryonic development. Dev. Biol. 287, 390–402.
Tsang, M., Dawid, I.B., 2004. Promotion and attenuation of FGF signaling through the Ras-MAPK pathway. Sci. STKE pe17.
Tsang, M., Kim, R., de Caestecker, M.P., Kudoh, T., Roberts, A.B., Dawid, I.B., 2000. Zebrafish nma is involved in TGFbeta family signaling. Genesis 28, 47–57.
Tsang, M., Friesel, R., Kudoh, T., Dawid, I.B., 2002. Identification of Sef, a novel modulator of FGF signalling. Nat. Cell Biol. 4, 165–169.
Tsang, M., Maegawa, S., Kiang, A., Habas, R., Weinberg, E., Dawid, I.B., 2004. A role for MKP3 in axial patterning of the zebrafish embryo. Development 131, 2769–2779.
Wasylyk, B., Hagman, J., Gutierrez-Hartmann, A., 1998. Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. Trends Biochem. Sci. 23, 213–216.
Yamauchi, H., Miyakawa, N., Miyake, A., Itoh, N., 2009. Fgf4 is required for left–right patterning of visceral organs in zebrafish. Dev. Biol. 332, 177–185.
Yelon, D., Horne, S.A., Stainier, D.Y., 1999. Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish. Dev. Biol. 214, 23–37.
Yelon, D., Ticho, B., Halpern, M.E., Ruvinsky, I., Ho, R.K., Silver, L.M., Stainier, D.Y., 2000. The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. Development 127, 2573–2582.
Zhang, Z., Verheyden, J.M., Hassell, J.A., Sun, X., 2009. FGF-regulated Etv genes are essential for repressing Shh expression in mouse limb buds. Dev. Cell 16, 607–613.

Cell

Paternally Induced Transgenerational Environmental Reprogramming of Metabolic Gene Expression in Mammals

Benjamin R. Carone,[1,10] Lucas Fauquier,[1,10] Naomi Habib,[4,5,10] Jeremy M. Shea,[1,10] Caroline E. Hart,[1] Ruowang Li,[2] Christoph Bock,[6,7] Chengjian Li,[1] Hongcang Gu,[6] Phillip D. Zamore,[1,3] Alexander Meissner,[6,7] Zhiping Weng,[2] Hans A. Hofmann,[8] Nir Friedman,[4,9] and Oliver J. Rando[1,*]

[1]Department of Biochemistry and Molecular Pharmacology
[2]Program in Bioinformatics and Integrative Biology
[3]Howard Hughes Medical Institute
University of Massachusetts Medical School, Worcester, MA 01605, USA
[4]School of Computer Science and Engineering, The Hebrew University, Jerusalem 91904, Israel
[5]Department of Molecular Genetics and Biotechnology, Faculty of Medicine, The Hebrew University, Jerusalem 91120, Israel
[6]Broad Institute, Cambridge, MA 02142, USA
[7]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[8]Section for Integrative Biology, Institute for Cellular & Molecular Biology, Institute for Neuroscience, University of Texas at Austin, Austin, TX 78712, USA
[9]Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel
[10]These authors contributed equally to this work
*Correspondence: oliver.rando@umassmed.edu
DOI 10.1016/j.cell.2010.12.008

SUMMARY

Epigenetic information can be inherited through the mammalian germline and represents a plausible transgenerational carrier of environmental information. To test whether transgenerational inheritance of environmental information occurs in mammals, we carried out an expression profiling screen for genes in mice that responded to *paternal* diet. Offspring of males fed a low-protein diet exhibited elevated hepatic expression of many genes involved in lipid and cholesterol biosynthesis and decreased levels of cholesterol esters, relative to the offspring of males fed a control diet. Epigenomic profiling of offspring livers revealed numerous modest (~20%) changes in cytosine methylation depending on paternal diet, including reproducible changes in methylation over a likely enhancer for the key lipid regulator *Ppara*. These results, in conjunction with recent human epidemiological data, indicate that parental diet can affect cholesterol and lipid metabolism in offspring and define a model system to study environmental reprogramming of the heritable epigenome.

INTRODUCTION

The past few decades have seen an important expansion of our understanding of inheritance, as a wide variety of epigenetically inherited traits have been described (Jablonka and Lamb, 1995, 2005; Rando and Verstrepen, 2007). One implication of epigenetic inheritance systems is that they provide a potential mechanism by which parents could transfer information to their offspring about the environment they experienced. In other words, mechanisms exist that could allow organisms to "inform" their progeny about prevailing environmental conditions. Under certain historical circumstances—for example, repeated exposure over evolutionary time to a moderately toxic environment that persists for tens of generations—such non-Mendelian information transfer could be adaptive (reviewed in Jablonka and Lamb, 1995; Rando and Verstrepen, 2007). Whether or not organisms can inherit characters induced by ancestral environments has far-reaching implications, and this type of inheritance has come to be called "Lamarckian" inheritance after the early evolutionary theorist J.B. Lamarck, although it is worth noting that both Darwin and Lamarck believed in the inheritance of acquired characters.

Despite these theoretical considerations, at present there is scant evidence for transgenerational effects of the environment, particularly in mammals. The majority of examples of transgenerational environmental effects described have been maternal effects (see Harris and Seckl, 2010; Whitelaw and Whitelaw, 2008; Youngson and Whitelaw, 2008 for review), including in utero passage of photoperiod information in various rodents (Horton, 2005), cultural inheritance of stress reactivity and maternal grooming behavior in rats (Meaney et al., 2007; Weaver et al., 2004), and metabolic and psychiatric sequelae of fetal malnutrition in humans and rodents (Hales and Barker, 2001; Harris and Seckl, 2010; Symonds et al., 2009). However, maternal effects are difficult to separate from direct effects of in utero environmental exposure on offspring.

A small number of studies have identified *heritable* epigenetic effects of environmental perturbations on offspring. Treatment of pregnant rat mothers with the endocrine disruptor vinclozolin results in decreased fertility and behavorial changes in several generations of offspring (Anway et al., 2005; Crews et al., 2007). In another study, withholding methyl donors from pregnant female mice resulted in decreased cytosine methylation across the agouti viable yellow A^{vy} reporter locus (Waterland and Jirtle, 2003), and the altered cytosine methylation profile persisted well beyond the first generation (Cropley et al., 2006).

Whereas demonstration of multigenerational changes (e.g., an F2 effect) is important when using maternal treatment protocols to rule out simple plastic responses of offspring to the in utero environment, paternal effects avoid this issue as fathers often contribute little more than sperm to offspring. A handful of paternal effects have been documented in the literature—pre-mating fasting of male mice has been reported to affect serum glucose levels in offspring (Anderson et al., 2006), and chronic exposure of male rats to high-fat diet affects pancreatic islet biology in offspring (Ng et al., 2010). Furthermore, epidemiological data from human populations link experience of famine in paternal grandfathers to obesity and cardiovascular disease two generations later (Kaati et al., 2002; Pembrey et al., 2006). These results motivate a deeper exploration of the mechanisms of pre-mating paternal diet on offspring phenotype.

It is therefore of great interest to determine what environmental conditions have transgenerational effects in mammals, and to characterize the mechanisms that mediate these effects. Here, we describe a genomic screen for transgenerational effects of paternal diet on gene expression in offspring in mice. Expression of hundreds of genes changes in the offspring of males fed a low-protein diet, with coherent upregulation of lipid and cholesterol biosynthetic pathways. Epigenomic profiling in offspring livers identified changes in cytosine methylation at a putative enhancer for the key lipid transcription factor *Ppara*, and these changes correlated with the downregulation of this gene in offspring. Interestingly, we did not find effects of paternal diet on methylation of this locus in sperm, and overall sperm cytosine methylation patterns were largely conserved under various dietary regimes. These results establish an inbred, genetically tractable model system for the study of transgenerational effects of diet and may have implications for the epidemiology of several major human diseases.

RESULTS

Experimental Paradigm

Male mice were fed control or low-protein diets (11% rather than 20% protein, with the remaining mass made up with sucrose) from weaning until sexual maturity. Note that although the relevant dietary change in this experiment could be protein content, sucrose content, fat/protein ratio, etc., for simplicity we refer to the diet as low protein throughout the text. Mice on either diet were then mated to females reared on control diet (Figure 1A and Figure S1A available online). Fathers were removed after 1 or 2 days of mating, limiting their influence on their progeny to the mating itself. All mothers were maintained on control diet throughout the course of the experiment. After birth, the offspring were reared with their mothers until 3 weeks old, at which point their livers were harvested for RNA isolation. DNA microarrays were used to profile global gene expression differences in the livers of the offspring from the two types of crosses (Table S1).

Testing for differences between 26 matched pairs of mice from the two F1 groups, we found a significant overabundance of differentially expressed genes, relative to the null hypothesis that the parental treatment does not affect offspring (1595 genes at a false discovery rate—FDR—of 0.001, Figures S1B and S1C). We also identified a more robust (t test with null hypothesis of mean change 0.2, FDR of 0.01) group of 445 genes whose expression strongly depended on the diet consumed by their fathers (Figure 1B). In our analysis we focus on this more robust group of genes; however, all the phenomena described below are true for the larger group as well. These gene expression changes were observed in 13 (7 low-protein, 6 control) litters in experiments spanning several years, carried out in three different animal facilities (Figures S2A and S2B). In principle, random factors should be distributed equally between our two groups given the numbers of offspring examined, but we directly address a number of potential artifacts nonetheless, including changes in cell populations, circadian cycle, litter size, order of sacrifice, and cage location (Figure S2, see Experimental Procedures).

We confirmed our results by q-RT-PCR (Figures 1C, Figure S1A). Squalene epoxidase (*Sqle*), which catalyzes the first oxygenation step in sterol biosynthesis, exhibited an ~3-fold increase in the low-protein cohort in our microarray data, and q-RT-PCR showed a similar average expression difference across over 25 animals, gathered in crosses carried out several years apart (Figure 1C). The differences we observe occur in both male and female progeny (Figure 1C, Figure S2C), though these dietary history-dependent differences are superimposed on a baseline of differential expression between the sexes.

Upregulation of Proliferation and Lipid Biosynthesis Genes in Low-Protein Offspring

To help define the physiological differences between our cohorts, we calculated enrichments of various Gene Ontology (GO) processes in the differentially expressed genes. Genes upregulated in our treatment group's offspring were enriched for a number of categories of genes involved in fat and cholesterol biosynthesis, including lipid biosynthesis ($p < 9 \times 10^{-26}$), steroid biosynthesis ($p < 3 \times 10^{-19}$), cholesterol biosynthesis ($p < 2 \times 10^{-12}$), and oxidation-reduction ($p < 4 \times 10^{-10}$). Another major group of upregulated genes are annotated to be involved in S phase, such as DNA replication ($p < 2 \times 10^{-9}$) and related annotations. Downregulated genes were enriched for GO annotations such as sequence-specific DNA binding ($p < 6 \times 10^{-6}$) and ligand-dependent nuclear receptor activity ($p < 6 \times 10^{-5}$), although the number of genes matching these annotations was small (14 and 5, respectively).

The increase in S phase genes likely indicates a hyperproliferative state, whereas the metabolic expression differences suggest that lipid metabolism is altered in these animals. To explore the mechanisms responsible for these altered gene expression programs, we asked whether the observed gene expression

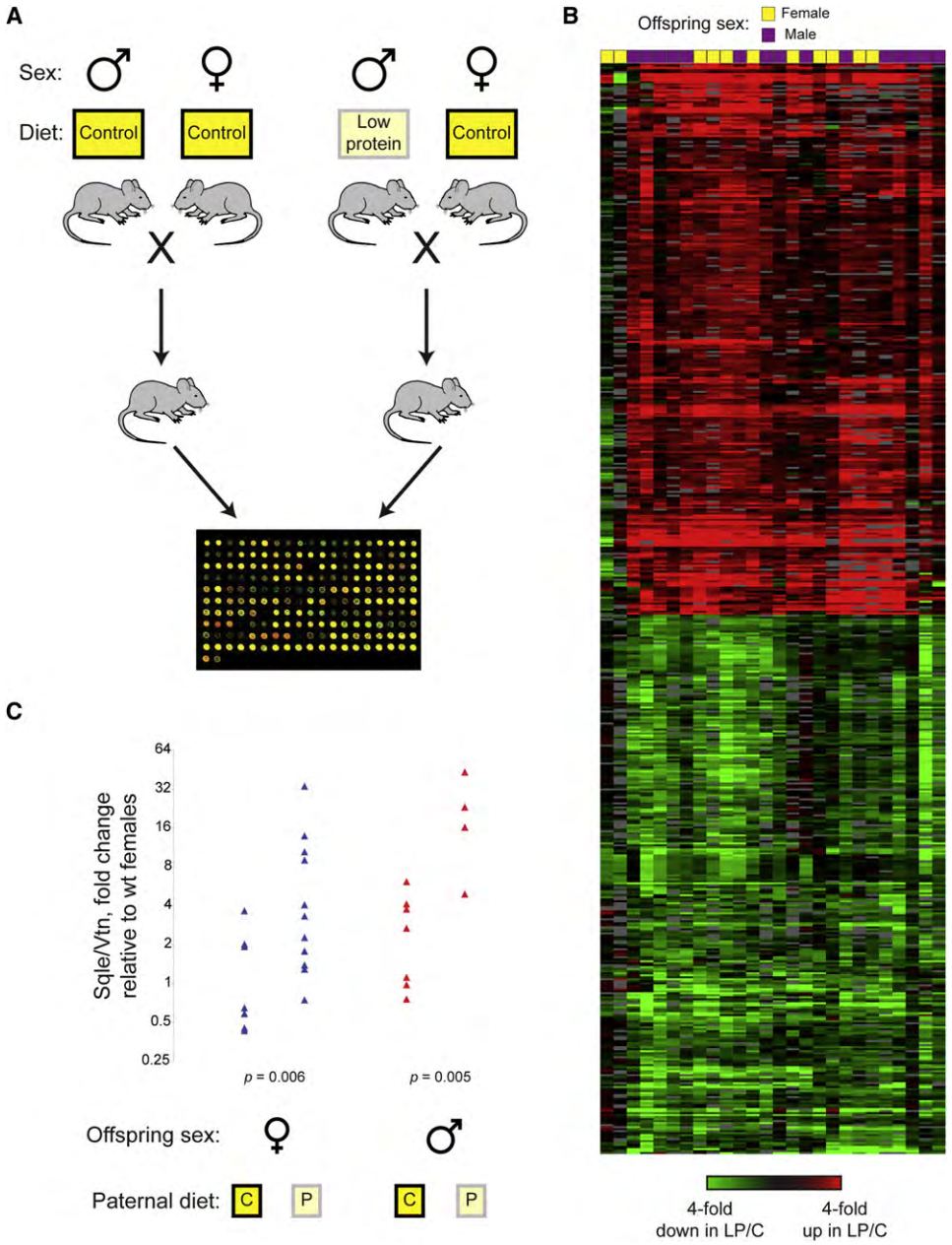


Figure 1. A Screen for Genes Regulated by Paternal Diet

(A) Experimental design. Male mice were fed control or low (11%) protein diet from weaning until sexual maturity, then were mated to females that were raised on control diet. Males were removed after 1 or 2 days of mating. Livers were harvested from offspring at 3 weeks, and RNA was prepared, labeled, and hybridized to oligonucleotide microarrays.

(B) Overview of microarray data, comparing offspring of sibling males fed different diets red boxes indicate higher RNA levels in low protein than control offspring, green indicates higher expression in controls. Boxes at the top indicate comparisons between two male (purple) or two female (yellow) offspring. Each column shows results from a comparison of a pair of offspring. Only genes passing the stringent threshold for significant change (Figure S1B) are shown. Data are clustered by experiment (columns) and by genes (rows).

(C) Validation of microarray data. Quantitative RT PCR was used to determine levels of Squalene epoxidase (*Sqle*) relative to the control gene Vitronectin (*Vtn*), which showed no change in the microarray dataset. Animals are grouped by paternal diet and by sex, and data are expressed as ΔC_T between *Sqle* and *Vtn*, normalized relative to the average of control females.

Additional validation is shown in Figure S1A. p values were calculated using t test. See also Table S1, Figure S1, and Figure S2.

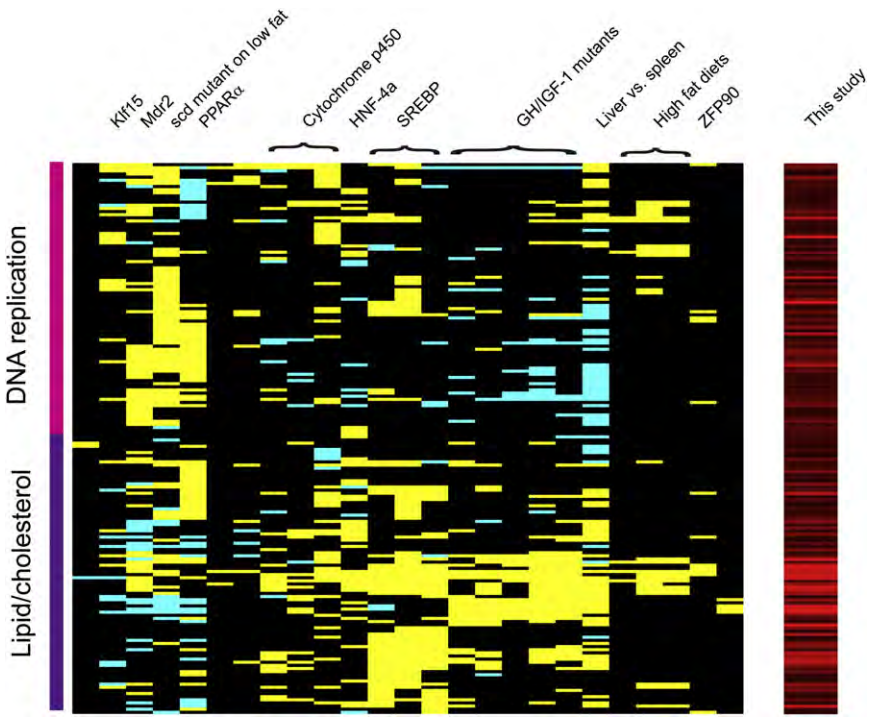


Figure 2. Multiple Pathways Are Affected by Paternal Diet

Comparison of upregulated gene expression profile with a compendium of public datasets of hepatic gene expression. A clustering of our upregulated genes according to their notation in the 28 significant ($p < 0.00025$) overlapping signatures from an assembled compendium of 120 publicly available murine liver signatures under various conditions and genetic perturbations (GEO; Horton et al., 2003; Yang et al., 2009). For each significant profile, the majority of overlapping genes are shown as yellow, whereas genes with opposite regulation (i.e., down rather than up in the dataset in question) are blue. The genes divide into two distinct clusters, one enriched in DNA replication and the other in various categories of fat and cholesterol biosynthesis. See also Table S2 and Figure S3.

differences might reflect altered regulation of a small number of pathways. We checked for significant overlaps of the gene expression profile observed in our low-protein offspring with a compendium of 120 publicly available murine liver gene expression datasets (Experimental Procedures). Our low-protein offspring gene expression profile significantly ($p < .05$ after Bonferroni correction) overlapped gene expression changes from 28 published profiles (Figure 2, Table S2), including gene expression profiles associated with perturbation of transcription factors that regulate cholesterol and lipid metabolism (SREBP [Horton et al., 2003], KLF15 [Gray et al., 2007], PPARα [Rakhshandehroo et al., 2007], and ZFP90 [Yang et al., 2009]). Our gene expression dataset also significantly matched hepatic gene expression in a variety of mice with mutations affecting growth hormone (GH) and insulin-like growth factor 1 (IGF-1) levels (Boylston et al., 2004; Madsen et al., 2004; Tsuchiya et al., 2004). Hierarchical clustering according to the enriched public profiles revealed two types of prominent gene functions in our data: DNA replication ($p < 6 \times 10^{-14}$) and lipid or cholesterol biosynthesis ($p < 2 \times 10^{-27}$) (Figure 2). The partial overlap observed with each of many different transcription factor and growth factor profiles suggests that the altered gene expression profile observed in low-protein offspring is likely related to reprogramming of multiple distinct pathways.

To assess whether the reprogrammed state in offspring reproduces the paternal response to low-protein diet, we measured global gene expression changes in the livers of pairs of animals weaned to control or low-protein diet as in Figure 1A. Genes that change in offspring are not the same as the genes induced in the parental generation by these protocols (Figure S3). Instead, males fed the low-protein diet upregulate immune response and apoptosis-related genes and downregulate genes involved in carboxylic acid metabolism (analysis not shown).

Transgenerational Effects on Lipid Metabolism

We further focused on cholesterol biosynthesis genes. Coherent upregulation of genes involved in cholesterol metabolism is observed in the offspring of low-protein fathers (Figure 3A). Figure 3B shows a more detailed comparison between our upregulated dataset and published data (Horton et al., 2003) for genes activated by a major transcriptional regulator of cholesterol metabolism, SREBP. Many of the genes upregulated in low-protein offspring have previously been shown to be upregulated by overexpression of SREBP-1a or SREBP-2 or downregulated by loss of the SREBP-activating gene *Scap*.

To explore the correspondence between hepatic gene expression and physiology, we measured lipid levels in three pairs of control and treatment livers to determine whether increased levels of lipid biosynthesis genes were associated with changes in lipid levels (Figure 3C, Experimental Procedures). Livers in the cohort with low-protein diet fathers were depleted of cholesterol and cholesterol esters (whose levels were reduced more than 2-fold). Additional differences were found in specific lipid classes, such as substantial increases in relative levels of saturated cardiolipins, saturated free fatty acids, and saturated and monounsaturated triacylglycerides in low-protein offspring (Table S3). Together, these results demonstrate that paternal diet affects metabolites of key biomedical importance in offspring.

MicroRNAs in Offspring

Small (19–35 bp) RNAs such as microRNAs (miRNAs) have recently been implicated in epigenetic inheritance in mice (Wagner et al., 2008). To determine whether altered small RNA populations might drive our reprogramming effect, we characterized by high-throughput sequencing the small (19–35 bp) RNA population from control and low-protein offspring livers (Ghildiyal et al., 2008) and mapped reads to known miRNAs (Table S4).

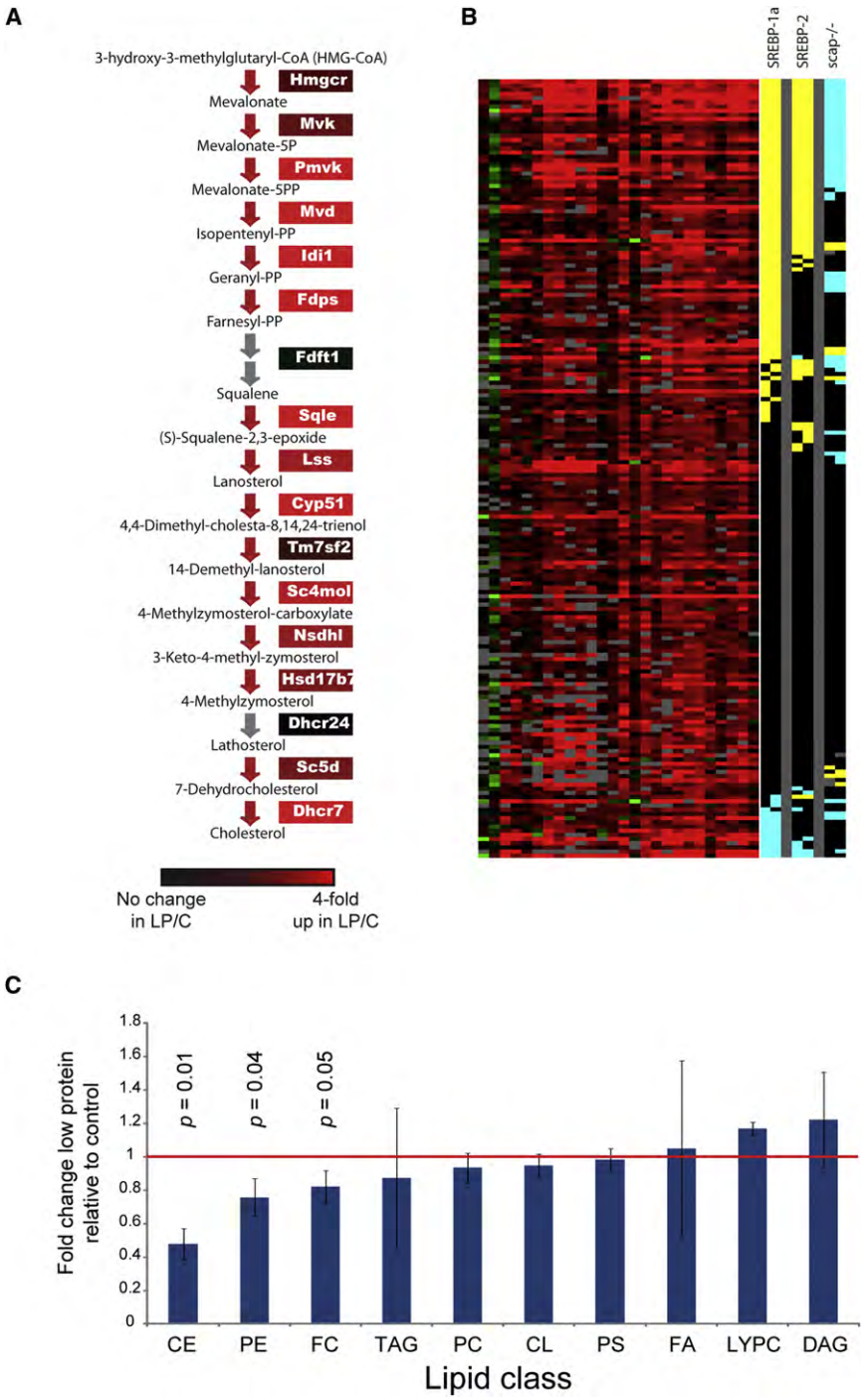


Figure 3. Altered Cholesterol Metabolism in the Low Protein Cohort

(A) Cholesterol biosynthesis. Genes annotated as cholesterol biosynthesis genes are shown, with colors indicating average difference in expression in low protein versus control comparisons.

(B) Many genes upregulated in the low protein cohort are SREBP targets. Upregulated cluster from Figure 1B is shown, along with data from Horton et al. (2003). Genes scored as up in both replicates from Horton et al. (2003) are shown as yellow, genes scored as down are blue. Columns show data from transgenic mice overexpressing SREBP 1a or SREBP 2 or from *Scap* knockout mice.

(C) Cholesterol levels are decreased in livers of low protein offspring. Data from lipidomic profiling of liver tissue from three control and three low protein animals are shown as mean ± standard deviation. Red line indicates no change. p values were calculated using a paired t test on log transformed lipid abundance data. Cholesterol esters, CE; phosphatidylethanolamine, PE; free cholesterol, FC; triacylglycerol, TAG; phopshatidylcholine, PC; cardiolipin, CL; phosphatidylserine, PS; free fatty acid, FA; lysophosphatidylcholine, LYPC; and diacylglycerol, DAG.

See also Table S3.

A number of miRNAs changed expression in the offspring from low-protein diet fathers (Figure 4). Changes were often subtle in magnitude (~50%), but were reproduced in four control versus low-protein comparisons (paired t test), and given the number of sequencing reads obtained for these RNAs this magnitude of difference is well outside of counting error (Table S4). Offspring of low-protein cohort upregulated miR-21, let-7, miR-199, and miR-98 and downregulated miR-210. Many of these upregulated miRNAs are associated with proliferation in liver, with miR-21 and miR-199 both associated with hepatocellular carcinoma (Jiang et al., 2008), whereas let-7 is well-known as a tumor suppressor (Jerome et al., 2007). The increase in growth-associated miRNAs is consistent with the hyperproliferative gene expression profile observed in the offspring of low-protein diet fathers.

We found no statistically significant overlap ($p > 0.05$) between the predicted targets of the miRNAs here and the gene expression changes we observe, though the subtle (~50%) changes in miRNA abundance we observe might be expected to have little effect on mRNA—even when specific miRNAs are artificially introduced in cells, downregulation of target mRNAs is less than 2-fold for the majority of predicted targets (Hendrickson et al., 2008). Our results therefore suggest that miRNAs are likely to be additional targets of the reprogramming pathway yet are likely not the direct upstream regulators of the entire response (but see Wagner et al., 2008).

Cytosine Methylation in Offspring

How are offspring reprogrammed by paternal diet? Cytosine methylation is a widespread DNA modification that is

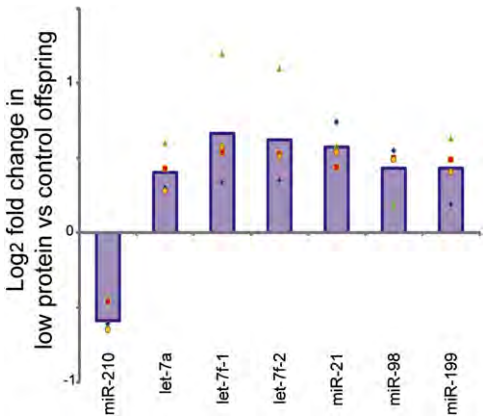


Figure 4. Proliferation Related MicroRNAs Respond to Paternal Diet
Small (<35 nt) RNAs from the livers of eight offspring (four control, four low protein) were isolated and subjected to high throughput sequencing. MicroRNAs that exhibited consistent changes in all four pairs of animals are shown, with average change shown as a bar and individual comparisons shown as points. See also Table S4.

environmentally responsive and carries at least some heritable information between generations (Bartolomei et al., 1993; Cropley et al., 2006; Holliday, 1987; Rakyan et al., 2003; Waterland and Jirtle, 2003). As imprinted loci are often involved in growth control (Moore and Haig, 1991), we first asked whether a subset of candidate imprinted loci exhibited altered cytosine methylation in low-protein offspring (Figure S4A). As these loci did not exhibit significant changes in methylation, we therefore turned to genome-scale mapping studies to search for differentially methylated loci between control and low-protein offspring.

We performed reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2008) to characterize cytosine methylation at single-nucleotide resolution across ~1% of the mouse genome (Table S5). RRBS was performed for livers from a pair of control and low-protein offspring, and fraction of methylated CpGs was calculated for a variety of features such as promoters, enhancers, and other nongenic CpG islands. In general, we found that cytosine methylation was well-correlated between control and low-protein offspring (Figures 5A and 5B). However, we did observe widespread modest (~10%–20%) changes in CpG methylation between the two samples (red and green dots in Figures 5A and 5B), consistent with many observations indicating that environmental changes tend to have small quantitative effects on cytosine methylation in the next generation (Blewitt et al., 2006; Heijmans et al., 2008; Ng et al., 2010; Weaver et al., 2004). Importantly, changes in promoter methylation did not globally correlate with changes in gene expression in offspring, indicating that the gene expression program in offspring is unlikely to be epigenetically specified at each individual gene (Figure 5C). Of course, widespread gene expression differences can be caused by changes to a small number of upstream regulators, and a number of differentially methylated regions are associated with cholesterol- or lipid-related genes (Table S5).

Most interestingly, we found a substantial (~30%) increase in methylation at an intergenic CpG island ~50 kb upstream of *Ppara* (Figure 6A). This locus is likely an enhancer for *Ppara*, as it is associated with the enhancer chromatin mark H3K4me1 (Heintzman et al., 2007) in murine liver (F. Yue and B. Ren, personal communication). *Ppara* is downregulated in the majority (but not all) of offspring livers (Table S1, Figure 6B), and the overall gene expression profile in our offspring livers significantly matches the gene expression changes observed in *Ppara* knockout mice (Figure 2), suggesting that epigenetic regulation of this single locus could drive a substantial fraction of the observed gene expression changes in offspring. Indeed, variance of *Ppara* mRNA levels alone can be used to explain ~13.7% of the variance in the entire gene expression dataset (although this of course does not determine causality).

We therefore assayed the methylation status of this locus by bisulfite sequencing in an additional 17 offspring livers (8 control and 9 low-protein), finding average differences of up to 8% methylation between low-protein and control livers at several CpGs in this locus (Figure 6C). Importantly, these pooled data underestimate the potential role of this locus in reprogramming as they include animals exhibiting a range of changes in *Ppara* gene expression—individual animal pairs with large differences in *Ppara* mRNA levels exhibit methylation differences of up to 30% at various cytosines across this locus. Figure 6D shows individual bisulfite clones for three pairs of animals with varying extents of *Ppara* downregulation (not all animals used for methylation analysis were analyzed by microarray). Taken together, these results identify a differentially methylated locus that is a strong candidate to be one of the upstream controllers of the hepatic gene expression response.

Cytosine Methylation, RNA, and Chromatin in Sperm

The link between paternal diet and offspring methylation patterns lead us to consider the hypothesis that paternal diet affects cytosine methylation patterns in sperm. We therefore isolated highly pure (>99%) sperm from the caudal epididymis of males consuming control or low-protein diet. We assayed the *Ppara* enhancer for methylation by bisulfite sequencing but found no significant changes between males consuming control or low-protein diet (Figure S4B). These results indicate either that cytosine methylation in sperm is not the relevant paternally transmitted dietary information at this locus (but changes at some point during development; Blewitt et al., 2006), or that we captured animals whose offspring would not manifest significant changes in expression of the associated genes—as seen in Figure 1B or Figure 6B, *Ppara* downregulation is variably penetrant in low-protein offspring.

To globally investigate effects of paternal diet on sperm cytosine methylation, we isolated sperm from four males—two consuming control diet, one consuming low-protein diet, and one subjected to a caloric restriction regimen. We then surveyed cytosine methylation patterns across the entire genome via MeDIP-Seq (immunoprecipitation using antibodies against 5me-C followed by deep sequencing; Jacinto et al., 2008; Weber et al., 2005) (Figure 7A, Figure S5A, and Figure S6). Notably, global cytosine methylation profiles were highly correlated between any pair of samples, indicating that the sperm "epigenome" is largely unresponsive to these differences in diet (Figures 7B–7D, Figures S5B–S5E). Indeed, littermates on different diets (Figures 7B and 7C) were better-correlated for promoter methylation than were the pairs of control animals

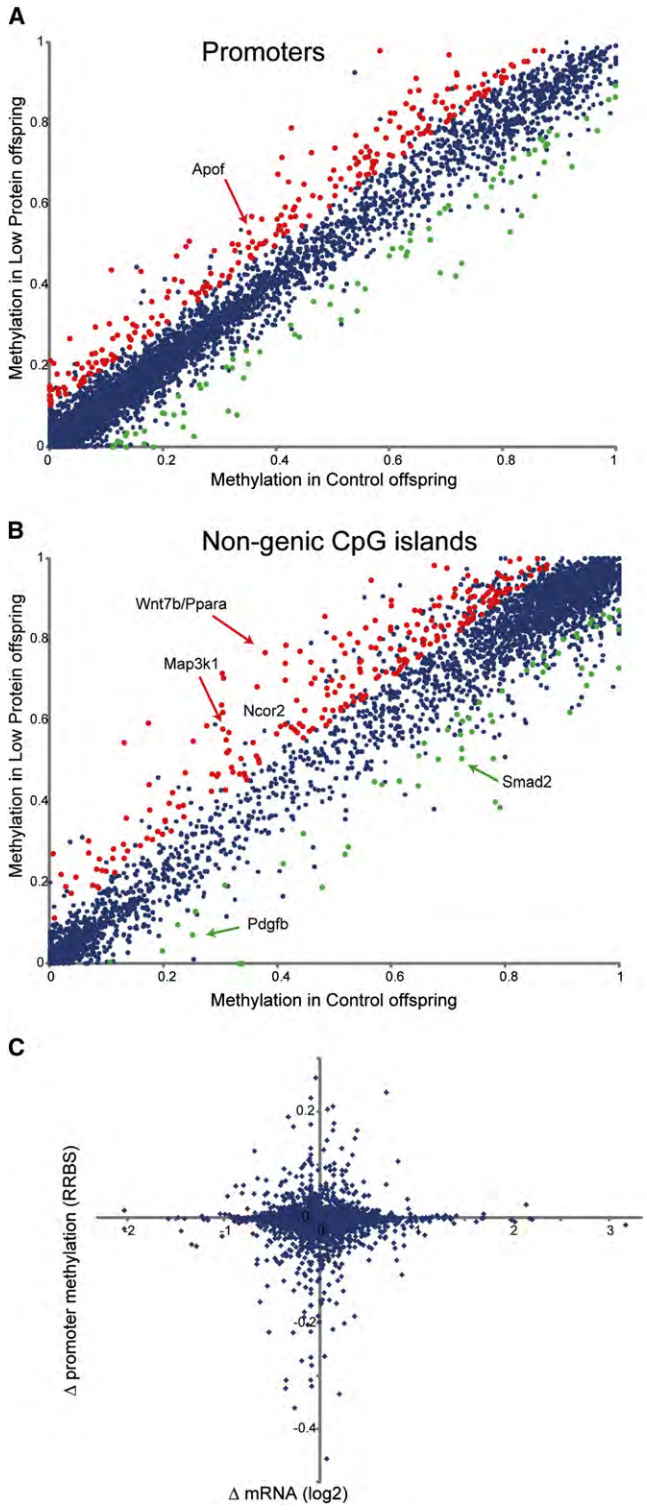


Figure 5. Transgenerational Effects of Paternal Diet on Hepatic Cytosine Methylation

(A) Genomic DNA from control and low protein offspring livers was subjected to reduced representation bisulfite sequencing (RRBS). For all annotated promoters, average fraction of CpGs that were methylated is shown for the control sample (x axis) compared to the low protein sample (y axis). Red and green dots indicate promoters with significant ($p < 0.05$) methylation changes of over 10%.

(B) As in (A), for nongenic CpG islands.

(C) Promoter cytosine methylation changes are uncorrelated with gene expression changes. For each promoter, the average change in cytosine methylation is compared to the change in mRNA abundance from Figure 1B. See also Table S5 and Figure S4.

from different litters (Figure 7D). Although these results do not rule out cytosine methylation in sperm as the relevant carrier of epigenetic information about paternal diet, the high correlation between samples, coupled with the absence of cytosine methylation changes at the *Ppara* enhancer in sperm, leads us to consider alternative epigenetic information carriers including RNA (Rassoulzadegan et al., 2006; Wagner et al., 2008) and chromatin (Arpanahi et al., 2009; Brykczynska et al., 2010; Hammoud et al., 2009; Ooi and Henikoff, 2007).

We used Affymetrix microarrays to analyze RNA levels for three pairs of males and for two matched epididymis samples (Figure S6, Figure S7A, Table S6). Curiously, low-protein and caloric restriction samples consistently exhibited more "sperm-like" RNA populations (as opposed to epididymis RNA) than did control samples (Figures S7B and S7C). Whether this reflects systematic contamination issues or biological differences in sperm maturity or quality is presently unknown, although we note that we confirmed consistently higher levels of the sperm-specific *Dnahc3* by q-RT-PCR in an additional 7/8 low-protein sperm samples (Figure S7E). We note that control sperm samples were routinely >99.5% sperm as assayed by microscopy (Figure S6), but nonetheless we cannot completely rule out systematic contamination issues. With this possibility in mind, we identified genes that were differentially packaged in control versus low-protein sperm by correcting for potential epididymal contamination (Figures S7B–S7F). Interestingly, we observed downregulation of a number of transcription factors and chromatin regulators such as *Smarcd3* and *Ppard*, although q-RT-PCR validation was not statistically significant due to high inter-animal variability (Figure S7F).

Although the downregulation of *Smarcd3* was not significantly confirmed by q-RT-PCR, this could reflect the variable penetrance of paternal diet on offspring described above. Given that heterozygous mutants in chromatin remodelers can affect offspring phenotype even when the mutant allele segregates away (Chong et al., 2007), we used an initial genome-wide mapping (not shown) of overall histone retention (pan-H3 ChIP) abundance and the key epigenetic histone modification H3K27me3 in sperm to identify targets for single locus analysis. We observed a consistent decrease in H3K27me3 in low-protein sperm at the promoter of *Maoa* (Monoamine oxidase) in 5/5 pairs of sperm samples and a decrease in H3K27me3 at *Eftud1* in 4/5 paired samples (Figures S7G and S7H). These results demonstrate proof of principle that the sperm epigenome is regulated by dietary conditions, although the biological implications of these observations are not yet clear.

DISCUSSION

Taken together, our results demonstrate that paternal diet affects lipid- and proliferation-related gene expression in the

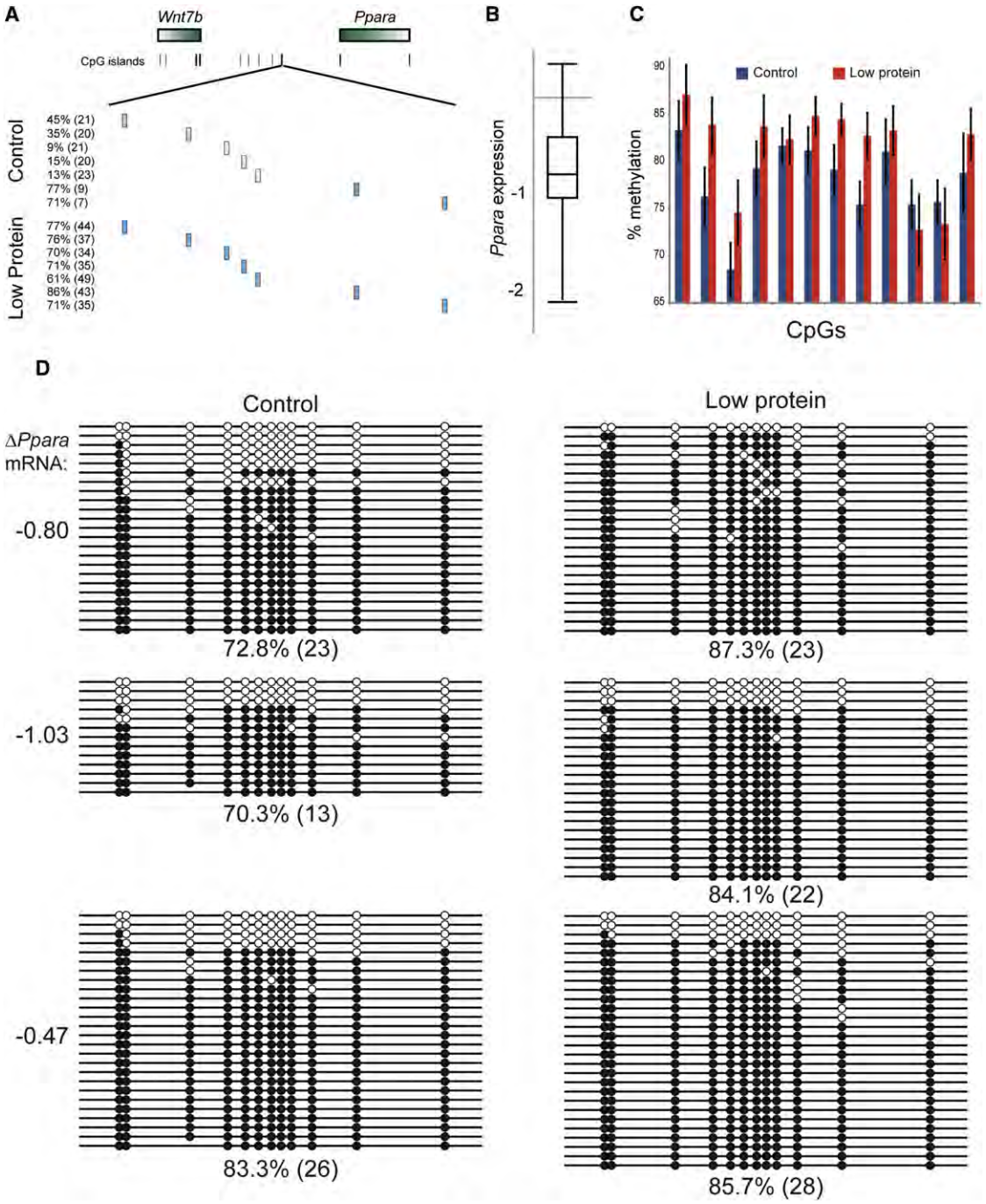


Figure 6. Effects of Paternal Diet on Methylation of a Putative *Ppara* Enhancer

(A) Differential methylation of a putative *Ppara* enhancer. Top panel shows a schematic of chromosome 15: 85,360,000 85,640,000. Zoomed in region represents chr15: 85,514,715 85,514,920. RRBS data for one control and one low protein offspring pair are shown below, with assayed CpGs represented as boxes colored to indicate % of clones methylated. Numbers to the left indicate % methylation, with number of sequence reads covering the CpG in parentheses.

(B) *Ppara* is downregulated in most low protein offspring livers. Box plot shows mean, quartiles, and highest and lowest values from Table S1.

(C) Putative enhancer methylation correlates with *Ppara* downregulation. DNA from eight control and nine low protein pairs of offspring livers was bisulfite treated, and at least 13 clones were analyzed for each animal. Percent methylation at each of the 12 CpGs in this region plotted on the y axis; data are shown as mean ± standard error of the mean (SEM).

(D) Individual bisulfite clones are shown for three control and three low protein offspring. White circles indicate unmethylated CpGs, black circles indicate methylated CpGs. Microarray data for change in *Ppara* RNA levels between the paired animals are shown to the left, in $\log_2$. Values under each bisulfite grouping indicate overall % methylation, with number of clones analyzed in parentheses.

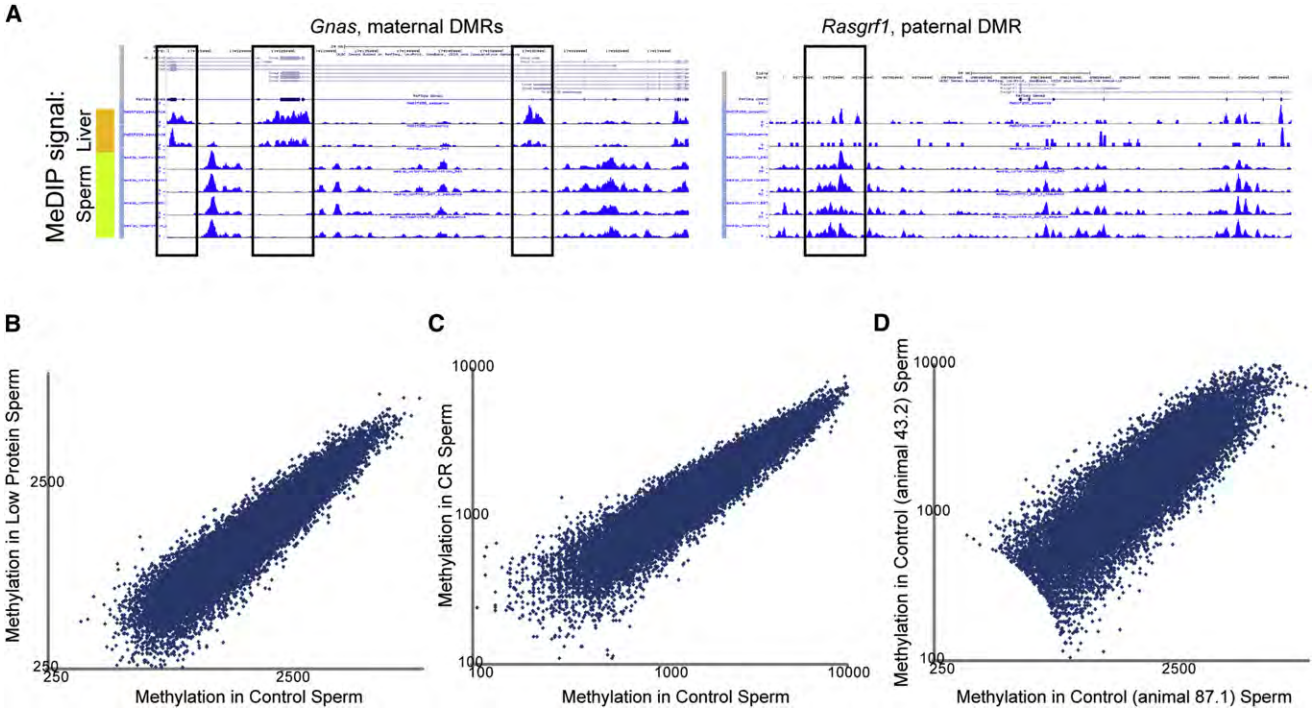


Figure 7. Modest Effects of Diet on the Sperm Epigenome

(A) MeDIP sequencing data are shown for two liver samples (top two tracks) and four sperm samples (bottom four) at a maternally methylated region (*Gnas*, left) and a paternally methylated region (*Rasgrf1*, right).

(B) Comparison of control and low protein methylation. For each promoter, methylation levels were averaged for 8 kb surrounding the TSS, and values are scat terplotted for control sperm (x axis) versus low protein sperm (y axis). x and y axes are plotted on logarithmic scales.

(C) As in (B), but for control versus caloric restriction.

(D) As in (B), but for the pair of control samples.

Similar results for (B) (D) are found when focusing on the 1 kb surrounding the TSS (not shown). See Figure S7 for analyses of consistent RNA and chromatin differences between low protein and control sperm.

offspring of inbred mice, and that epigenetic information carriers in sperm respond to environmental conditions. These results have potential implications for human health and raise numerous mechanistic questions, discussed below.

Paternal Diet Affects Metabolism in Offspring

Our results clearly identify a set of physiological pathways whose expression is sensitive to paternal diet. Specifically, we find that hepatic expression of genes involved in proliferation and cholesterol biosynthesis can be regulated by paternal diet, and these changes are reflected in levels of several lipid metabolites. Combined with data showing that offspring glucose levels are affected by paternal fasting in mice (Anderson et al., 2006), these results demonstrate that paternal diet has wide-ranging effects on the metabolism of offspring in rodents. Interestingly, a very recent study from Ng et al. (Ng et al., 2010) reported that chronic exposure of male rats to high-fat diet was associated with pancreatic beta cell dysfunction in female offspring. It will naturally be of great interest in the near future to compare the transgenerational effects of high-fat and low-protein diets, although one clear difference is that in our system a transgenerational effect is observed in both sex offspring.

Whether the effects we observe on cholesterol metabolism prove advantageous in low-protein conditions remains to be tested, but it will be important to investigate ecologically relevant diets in order to speculate more firmly about adaptive significance of any observed transgenerational effects. For example, at present we cannot say with certainty what aspect of the low-protein regimen is sensed by males—it is possible that offspring metabolism is affected by overall protein consumption, or high sucrose, or fat/protein ratio, or even levels of micronutrients, as our males consumed diets ad libitum and thus might have overconsumed the low-protein diet.

The Reprogrammed State: Liver

What is the mechanistic basis for the reprogrammed gene expression state? Genome-scale analyses of cytosine methylation in offspring livers identified several lipid-related genes that were differentially methylated depending on paternal diet. Most notably, a putative enhancer for a major lipid regulator, *Ppara*, exhibited generally higher methylation in low-protein offspring than in control offspring. Methylation at this locus was variable between animals, consistent with the partial penetrance of *Ppara* downregulation in our dataset. The overall gene expression profile observed in low-protein offspring significantly overlaps

gene expression changes observed in *Ppara* knockout mice (Rakhshandehroo et al., 2007), leading to the hypothesis that epigenetic *Ppara* downregulation via enhancer methylation is an upstream event that affects an entire downstream regulon in reprogrammed animals. Note that although the hepatic downregulation of *Ppara* suggests a liver-autonomous epigenetic change, we cannot rule out that hepatic gene expression changes result from global physiological changes resulting from downregulation of *Ppara* in some other tissue.

Interestingly, *Ppara* expression in liver is also regulated by *maternal* diet—offspring of female mice consuming a high-fat diet exhibit altered hepatic *Ppara* expression, with increased expression at birth but decreased expression at weaning (Yamaguchi et al., 2010). Together with our data, these results suggest that *Ppara* is a key nexus that integrates ancestral dietary information to control offspring metabolism.

Mechanistic Basis for Transgenerational Paternal Effects

Paternal diet could potentially affect offspring phenotype via a number of different mechanisms. Although we focus here on epigenetic inheritance systems, it is important to note that parental information can also be passed to offspring via social or cultural inheritance systems (Avital and Jablonka, 2000; Champagne and Meaney, 2001; Jablonka and Lamb, 1995; Meaney et al., 2007; Weaver et al., 2004). Although such maternally provided social inheritance is unlikely in our paternal effect system—males were typically only in females' cages for one day—it is known that in some animals females can judge mate quality and allocate resources accordingly (Pryke and Griffith, 2009), and that seminal fluid can influence female postcopulatory behavior in *Drosophila* (Fricke et al., 2008; Wolfner, 2002). These and other plausible transgenerational information carriers cannot be excluded at present—ongoing artificial insemination and in vitro fertilization experiments will determine whether sperm carry the relevant metabolic information in our system.

Here we focused on the hypothesis that paternal dietary information does indeed reside in sperm epigenetic information carriers. First, a subset of cytosine methylation patterns in sperm are known to be heritable (Chong et al., 2007; Cropley et al., 2006; Rakyan et al., 2003; Waterland and Jirtle, 2003). Second, several reports suggest that RNA molecules packaged in sperm can affect offspring phenotype (Rassoulzadegan et al., 2006; Wagner et al., 2008). Third, chromatin structure has been proposed to carry epigenetic information, as sperm are largely devoid of histone proteins but retain them at a subset of developmentally important loci (Arpanahi et al., 2009; Brykczynska et al., 2010; Chong et al., 2007; Hammoud et al., 2009). Finally, it is conceivable that additional or novel epigenetic regulators (such as prions) are packaged into sperm, or that sperm quality is affected by diet, or that genetic changes are directed by the environment (although it is important to emphasize that inbred mouse strains were used in this study).

Here, we report whole-genome characterization of cytosine methylation patterns and RNA content in sperm obtained from mice maintained on control, low-protein, and caloric restriction diets. Globally, cytosine methylation patterns are similar in all three conditions, indicating that the sperm epigenome is largely unaffected by these diets. Nonetheless, changes in relatively few loci can have profound effects in the developing animal, and our data do not rule out the possibility of inheritance through sperm cytosine methylation, especially given that MeDIP is unlikely to identify ~10%–20% of differences in methylation at a small number of cytosines. Importantly, the putative enhancer of *Ppara* (Figure 6) was not differentially methylated in sperm. It will therefore be of great interest in the future to determine when during development the differential methylation observed in liver is established and to identify the upstream events leading to differential methylation (Blewitt et al., 2006).

Interestingly, we did identify effects of diet on RNA content and chromatin packaging of sperm. For example, sperm from control animals were consistently depleted of the highly sperm-specific *Dnahc3* gene (Figure S7) relative to sperm from low-protein animals. We cannot presently determine whether this represents reproducible differences in contamination, differences in sperm maturity, or something else. Finally, based on our observation that low-protein sperm tended to be depleted of genes encoding a number of chromatin regulators, we have begun to search for dietary effects on sperm chromatin structure. Interestingly we found that the *Maoa* promoter was consistently depleted of the key Polycomb-related chromatin mark H3K27me3 (Figure S7G), demonstrating as a proof of concept that chromatin packaging of the sperm genome is responsive to the environment and motivating genome-wide investigation into dietary effects on sperm chromatin. Given the common behavorial changes observed in many transgenerational inheritance paradigms, the possibility that H3K27me3 at *Maoa* affects offspring behavior (potentially via altered offspring responses to maternal stress; Harris and Seckl, 2010) will be of great future interest.

Relevance to Human Disease

These results are likely to be relevant for human disease because not only is *maternal* starvation in humans correlated with obesity and diabetes in children (Lumey et al., 2007), but also, remarkably, limited food in paternal grandfathers has been associated with changed risk of diabetes and cardiovascular disease in grandchildren (Kaati et al., 2002; Pembrey et al., 2006). Interestingly, in these studies ancestral access to food and disease risk were not associated with disease risk in the next generation but were only associated with F2 disease risk. However, it is important to note that the transgenerational effects of food availability for paternal grandfathers depend on the exact period during childhood of exposure to rich or poor diets (Pembrey et al., 2006), whereas our experimental protocol involved continuous low-protein diet from weaning until mating. Thus, future studies are required to more precisely define when and how ancestral exposure to a low-protein diet affects epigenetic programming of offspring metabolism.

Together, these results suggest rethinking basic practices in epidemiological studies of complex diseases such as diabetes, heart disease, or alcoholism. We believe that future environmental exposure histories will need to include parental exposure histories as well as those of the patients to disentangle induced epigenetic effects from the currently sought genetic and environmental factors underlying complex diseases. Our observations

provide an inbred mammalian model for transgenerational reprogramming of metabolic phenotype that will enable dissection of the exposure history necessary for reprogramming and genetic analysis of the machinery involved in reprogramming, and they suggest a number of specific pathways likely to be the direct targets of epigenetic reprogramming.

EXPERIMENTAL PROCEDURES

Mouse Husbandry

All animal care and use procedures were in accordance with guidelines of the Institutional Animal Care and Use Committee. C57/Bl6 mice were obtained from Jackson Labs and from Charles River Laboratories (for different iterations of this experiment). All experiments were performed with mice that had been raised for at least two generations on control diet to attempt to minimize any transgenerational effects of transitioning to control diet from chow provided by animal provider. For all comparisons shown, male mice were weaned from mothers at 21 days of age, and sibling males were put into cages with low protein or control diet (moistened with water to allow mice to break the hard pellets). Females were weaned to control diet. Males were raised on diet until 9 12 weeks of age, at which point they were placed with females for 1 or 2 days. Control and low protein mating cages were always interspersed with one another. Note that we always used virgin females to avoid confounding effects of the female's litter number, although this results in many lost litters as first litters were often consumed by their mothers. After 1 to 2 days, males were removed, and pregnant females were left alone with control diet and a shepherd shack until their litters were 3 weeks of age. At 3 weeks of age offspring were sacrificed by isoflurane and cervical dislocation, and median lobe of liver was rapidly dissected out and flash frozen in liquid N_2.

Diets

Diets were obtained from Bio serv, and compositions are listed in Table S7. For most experiments only low protein diet was sterilized per standard protocol at Bio serv. For later experiments, both diets were sterilized.

RNA Extraction

Liver samples were ground with a liquid N2 cooled mortar and pestle. Total RNA for microarray analysis was extracted from liver powder using Trizol.

Microarray Hybridization

Thirty micrograms of total RNA was labeled for 2 hr at 42°C with Superscript II reverse transcriptase using 4 μg of random hexamer and 4 μg of oligo dT. Cy3 and Cy5 labeled samples were hybridized to home printed "MEEBO" microarrays. MEEBO information is available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6352. Microarrays were hybridized at 65°C for 16 hr, washed as previously described (Diehn et al., 2002), and scanned using an Axon Genepix 4000B microarray scanner.

Comparison to Public Murine Liver Microarray Data

We built a compendium of public microarray data consisting of 120 gene expression profiles in the murine liver under various conditions and genetic perturbations. Signatures of differentially expressed genes were determined using a combination of two one tailed t tests, with FDR correction of 0.1. Profiles significantly enriched with up or downregulated genes in low protein offspring were defined by a hypergeometric p value ≤ 0.05 after correction for multiple hypotheses ($p < 0.00025$).

Lipid Measurements

~50 100 mg of ground liver tissue from six animals (three paired sets) was sent to Lipomics for "Truemass" mass spectrometry characterization of 450 lipid levels (Table S4). Note that samples 73 1 and 76 1 come from PBS perfused livers, whereas the other four samples were dissected without perfusion.

Small RNA Cloning and Sequencing

Total RNA was isolated from ground liver tissue using mirVana (Ambion). 18 35 nt small RNA was purified from 100 μg of total RNA, ligated to adaptors, amplified, gel purified, and sequenced using a Solexa Genome Analyzer (Illumina) (Ghildiyal et al., 2008).

RRBS

Reduced representation bisulfite sequencing was carried out as previously described (Meissner et al., 2008). Data are available at http://thriftyepigenome.computational epigenetics.org.

Sperm Isolation

Caudal epididymis was dissected from sacrificed animals, punctured, and incubated for 30 min in M2 media (Sigma) at 37°C. Supernatant was removed, pelleted (3000 g for 5 min), washed 2× with PBS and 1× in water, and incubated in Somatic Cell Lysis buffer. Sperm preparations were used only if they were >99.5% pure as assessed by microscopy, and q RT PCR was also used to reject any sperm samples based on the ratio between epididymis specific genes *Actb* or *Myh11* and sperm specific genes *Smcp* or *Odf1* (Figure S6).

MeDIP

Methyl DNA immunoprecipitation was carried out essentially as described (Weber et al., 2005, 2007). Four micrograms of purified genomic DNA was fragmented to a mean size of 300 bp using a Covaris machine, denatured, and immunoprecipitated with 5mC antibody (Eurogentec). ChIP material was Solexa sequenced, with ~21 million uniquely mappable reads per library.

ACCESSION NUMBERS

All microarray data and deep sequencing data used in this study have been deposited to GEO (http://www.ncbi.nlm.nih.gov/projects/geo/), accession # GSE25899.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at doi:10.1016/j.cell.2010.12.008.

ACKNOWLEDGMENTS

We thank D. Haig for initial discussions motivating this experiment and K. Ahmad, P. Kaufman, C. Meiklejohn, Y. Nahmias, and members of the Rando lab for critical reading of the manuscript. O.J.R. is supported in part by a Career Award in Biomedical Sciences from the Burroughs Wellcome Fund, by grant GM088618 from NIGMS, and by the Mathers Foundation and is a member of the UMass Diabetes Endocrinology Research Center (DK32520). N.F. is supported by grants from the NIH and the US Israel Binational Science Foundation (BSF). H.A.H. is supported by NSF, an Alfred P. Sloan Foundation Fellowship and a Dwight W. and Blanche Faye Reeder Centennial Fellowship in Systematic and Evolutionary Biology. P.D.Z. is supported by grants GM62862 and GM65236 from the NIGMS. N.H. is supported by the Meidan fellowship. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

O.J.R. and H.A.H. designed the original expression experiment, and a pilot was carried out by C.E.H., O.J.R., and H.A.H. O.J.R., B.C, L.F., and J.S. carried out animal husbandry and gene expression experiments, and L.F. and C.L. carried out miRNA profiling. C.B. and A.M. carried out RRBS experiments, L.F. carried out sperm RNA profiling, and J.S. carried out MeDIP experiments. J.S. and B.C. carried out bisulfite sequencing, and B.C. carried out chromatin ChIPs. N.H., O.J.R., and N.F. analyzed the gene expression and microRNA data, R.L., Z.W., and O.J.R. analyzed the MeDIP data. O.J.R., P.D.Z., and N.F. wrote the paper.

Received: October 29, 2010
Revised: December 6, 2010
Accepted: December 8, 2010
Published: December 23, 2010

REFERENCES

Anderson, L.M., Riffle, L., Wilson, R., Travlos, G.S., Lubomirski, M.S., and Alvord, W.G. (2006). Preconceptional fasting of fathers alters serum glucose in offspring of mice. Nutrition *22*, 327 331.

Anway, M.D., Cupp, A.S., Uzumcu, M., and Skinner, M.K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science *308*, 1466 1469.

Arpanahi, A., Brinkworth, M., Iles, D., Krawetz, S.A., Paradowska, A., Platts, A.E., Saida, M., Steger, K., Tedder, P., and Miller, D. (2009). Endonuclease sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. Genome Res. *19*, 1338 1349.

Avital, E., and Jablonka, E. (2000). Animal Traditions: Behavioural Inheritance in Evolution (Cambridge, UK: Cambridge University Press).

Bartolomei, M.S., Webber, A.L., Brunkow, M.E., and Tilghman, S.M. (1993). Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. Genes Dev. *7*, 1663 1673.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet. *2*, e49.

Boylston, W.H., Gerstner, A., DeFord, J.H., Madsen, M., Flurkey, K., Harrison, D.E., and Papaconstantinou, J. (2004). Altered cholesterologenic and lipo genic transcriptional profile in livers of aging Snell dwarf (Pit1dw/dwJ) mice. Aging Cell *3*, 283 296.

Brykczynska, U., Hisano, M., Erkek, S., Ramos, L., Oakeley, E.J., Roloff, T.C., Beisel, C., Schubeler, D., Stadler, M.B., and Peters, A.H. (2010). Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat. Struct. Mol. Biol. *17*, 679 687.

Champagne, F., and Meaney, M.J. (2001). Like mother, like daughter: evidence for non genomic transmission of parental behavior and stress responsivity. Prog. Brain Res. *133*, 287 302.

Chong, S., Vickaryous, N., Ashe, A., Zamudio, N., Youngson, N., Hemley, S., Stopka, T., Skoultchi, A., Matthews, J., Scott, H.S., et al. (2007). Modifiers of epigenetic reprogramming show paternal effects in the mouse. Nat. Genet. *39*, 614 622.

Crews, D., Gore, A.C., Hsu, T.S., Dangleben, N.L., Spinetta, M., Schallert, T., Anway, M.D., and Skinner, M.K. (2007). Transgenerational epigenetic imprints on mate preference. Proc. Natl. Acad. Sci. USA *104*, 5942 5946.

Cropley, J.E., Suter, C.M., Beckman, K.B., and Martin, D.I. (2006). Germ line epigenetic modification of the murine A vy allele by nutritional supplementa tion. Proc. Natl. Acad. Sci. USA *103*, 17308 17312.

Diehn, M., Alizadeh, A.A., Rando, O.J., Liu, C.L., Stankunas, K., Botstein, D., Crabtree, G.R., and Brown, P.O. (2002). Genomic expression programs and the integration of the CD28 costimulatory signal in T cell activation. Proc. Natl. Acad. Sci. USA *99*, 11796 11801.

Fricke, C., Bretman, A., and Chapman, T. (2008). Adult male nutrition and reproductive success in Drosophila melanogaster. Evolution *62*, 3170 3177.

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

Gray, S., Wang, B., Orihuela, Y., Hong, E.G., Fisch, S., Haldar, S., Cline, G.W., Kim, J.K., Peroni, O.D., Kahn, B.B., et al. (2007). Regulation of gluconeogen esis by Kruppel like factor 15. Cell Metab. *5*, 305 312.

Hales, C.N., and Barker, D.J. (2001). The thrifty phenotype hypothesis. Br. Med. Bull. *60*, 5 20.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473 478.

Harris, A., and Seckl, J. (2010). Glucocorticoids, prenatal stress and the programming of disease. Horm. Behav. Published online June 18, 2010. 10. 1016/j.yhbeh.2010.06.007.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc. Natl. Acad. Sci. USA *105*, 17046 17049.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311 318.

Hendrickson, D.G., Hogan, D.J., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2008). Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. PLoS ONE *3*, e2126.

Holliday, R. (1987). The inheritance of epigenetic defects. Science *238*, 163 170.

Horton, J.D., Shah, N.A., Warrington, J.A., Anderson, N.N., Park, S.W., Brown, M.S., and Goldstein, J.L. (2003). Combined analysis of oligonucleotide micro array data from transgenic and knockout mice identifies direct SREBP target genes. Proc. Natl. Acad. Sci. USA *100*, 12027 12032.

Horton, T.H. (2005). Fetal origins of developmental plasticity: animal models of induced life history variation. Am. J. Hum. Biol. *17*, 34 43.

Jablonka, E., and Lamb, M.J. (1995). Epigenetic Inheritance and Evolution: The Lamarckian Dimension (Oxford, UK: Oxford University Press).

Jablonka, E., and Lamb, M.J. (2005). Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life (Cam bridge, MA: MIT Press).

Jacinto, F.V., Ballestar, E., and Esteller, M. (2008). Methyl DNA immunoprecip itation (MeDIP): hunting down the DNA methylome. Biotechniques *44*, 35, 37, 39 passim.

Jerome, T., Laurie, P., Louis, B., and Pierre, C. (2007). Enjoy the silence: the story of let 7 microRNA and cancer. Curr. Genomics *8*, 229 233.

Jiang, J., Gusev, Y., Aderca, I., Mettler, T.A., Nagorney, D.M., Brackett, D.J., Roberts, L.R., and Schmittgen, T.D. (2008). Association of microRNA expres sion in hepatocellular carcinomas with hepatitis infection, cirrhosis, and patient survival. Clin. Cancer Res. *14*, 419 427.

Kaati, G., Bygren, L.O., and Edvinsson, S. (2002). Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur. J. Hum. Genet. *10*, 682 688.

Lumey, L., Stein, A.D., Kahn, H.S., van der Pal de Bruin, K.M., Blauw, G., Zybert, P.A., and Susser, E.S. (2007). Cohort profile: the Dutch hunger winter families study. Int. J. Epidemiol. *36*, 1196 1204.

Madsen, M.A., Hsieh, C.C., Boylston, W.H., Flurkey, K., Harrison, D., and Papaconstantinou, J. (2004). Altered oxidative stress response of the long lived Snell dwarf mouse. Biochem. Biophys. Res. Commun. *318*, 998 1005.

Meaney, M.J., Szyf, M., and Seckl, J.R. (2007). Epigenetic mechanisms of peri natal programming of hypothalamic pituitary adrenal function and health. Trends Mol. Med. *13*, 269 277.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766 770.

Moore, T., and Haig, D. (1991). Genomic imprinting in mammalian develop ment: a parental tug of war. Trends Genet. *7*, 45 49.

Ng, S.F., Lin, R.C., Laybutt, D.R., Barres, R., Owens, J.A., and Morris, M.J. (2010). Chronic high fat diet in fathers programs beta cell dysfunction in female rat offspring. Nature *467*, 963 966.

Ooi, S.L., and Henikoff, S. (2007). Germline histone dynamics and epigenetics. Curr. Opin. Cell Biol. *19*, 257 265.

Pembrey, M.E., Bygren, L.O., Kaati, G., Edvinsson, S., Northstone, K., Sjos trom, M., and Golding, J. (2006). Sex specific, male line transgenerational responses in humans. Eur. J. Hum. Genet. *14*, 159 166.

Pryke, S.R., and Griffith, S.C. (2009). Genetic incompatibility drives sex alloca tion and maternal investment in a polymorphic finch. Science *323*, 1605 1607.

Rakhshandehroo, M., Sanderson, L.M., Matilainen, M., Stienstra, R., Carlberg, C., de Groot, P.J., Muller, M., and Kersten, S. (2007). Comprehensive analysis

of PPARalpha dependent regulation of hepatic lipid metabolism by expression profiling. PPAR Res. *2007*, 26839.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V., and Whitelaw, E. (2003). Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal trans mission. Proc. Natl. Acad. Sci. USA *100*, 2538 2543.

Rando, O.J., and Verstrepen, K.J. (2007). Timescales of genetic and epige netic inheritance. Cell *128*, 655 668.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA mediated non mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469 474.

Symonds, M.E., Sebert, S.P., Hyatt, M.A., and Budge, H. (2009). Nutritional programming of the metabolic syndrome. Nat. Rev. Endocrinol. *5*, 604 610.

Tsuchiya, T., Dhahbi, J.M., Cui, X., Mote, P.L., Bartke, A., and Spindler, S.R. (2004). Additive regulation of hepatic gene expression by dwarfism and caloric restriction. Physiol. Genomics *17*, 307 315.

Wagner, K.D., Wagner, N., Ghanbarian, H., Grandjean, V., Gounon, P., Cuzin, F., and Rassoulzadegan, M. (2008). RNA induction and inheritance of epige netic cardiac hypertrophy in the mouse. Dev. Cell *14*, 962 969.

Waterland, R.A., and Jirtle, R.L. (2003). Transposable elements: targets for early nutritional effects on epigenetic gene regulation. Mol. Cell. Biol. *23*, 5293 5300.

Weaver, I.C., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., and Meaney, M.J. (2004). Epigenetic programming by maternal behavior. Nat. Neurosci. *7*, 847 854.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome wide and promoter specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. *37*, 853 862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. *39*, 457 466.

Whitelaw, N.C., and Whitelaw, E. (2008). Transgenerational epigenetic inheri tance in health and disease. Curr. Opin. Genet. Dev. *18*, 273 279.

Wolfner, M.F. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in Drosophila. Heredity *88*, 85 93.

Yamaguchi, R., Nakagawa, Y., Liu, Y.J., Fujisawa, Y., Sai, S., Nagata, E., Sano, S., Satake, E., Matsushita, R., Nakanishi, T., et al. (2010). Effects of maternal high fat diet on serum lipid concentration and expression of peroxisomal pro liferator activated receptors in the early life of rat offspring. Horm. Metab. Res. *42*, 821 825.

Yang, X., Deignan, J.L., Qi, H., Zhu, J., Qian, S., Zhong, J., Torosyan, G., Majid, S., Falkard, B., Kleinhanz, R.R., et al. (2009). Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. Nat. Genet. *41*, 415 423.

Youngson, N.A., and Whitelaw, E. (2008). Transgenerational epigenetic effects. Annu. Rev. Genomics Hum. Genet. *9*, 233 257.

MTOpen

MicroRNA-regulated, Systemically Delivered rAAV9: A Step Closer to CNS-restricted Transgene Expression

Jun Xie[1,2], Qing Xie[1,3], Hongwei Zhang[1,2], Stefan L Ameres[4], Jui-Hung Hung[5], Qin Su[1], Ran He[1], Xin Mu[1,6], Seemin Seher Ahmed[1,2], Soyeon Park[1,2], Hiroki Kato[7], Chengjian Li[4], Christian Mueller[1,8], Craig C Mello[7,9], Zhiping Weng[10], Terence R Flotte[1,2,8], Phillip D Zamore[4,9] and Guangping Gao[1,2]

[1]Gene Therapy Center, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [2]Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [3]Department of Microbiology, Peking University Health Science Center, Beijing, China; [4]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [5]Department of Biomedical Engineering, Boston University, Boston, Massachusetts, USA; [6]Department of Dermatology, The First Affiliated Hospital of Medical College of Xian Jiaotong University, Xian, Shanxi, China; [7]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [8]Department of Pediatrics, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [9]Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, Massachusetts, USA; [10]Program in Bioinformatics, University of Massachusetts Medical School, Worcester, Massachusetts, USA

Recombinant adeno-associated viruses (rAAVs) that can cross the blood–brain-barrier and achieve efficient and stable transvascular gene transfer to the central nervous system (CNS) hold significant promise for treating CNS disorders. However, following intravascular delivery, these vectors also target liver, heart, skeletal muscle, and other tissues, which may cause untoward effects. To circumvent this, we used tissue-specific, endogenous microRNAs (miRNAs) to repress rAAV expression outside the CNS, by engineering perfectly complementary miRNA-binding sites into the rAAV9 genome. This approach allowed simultaneous multi-tissue regulation and CNS-directed stable transgene expression without detectably perturbing the endogenous miRNA pathway. Regulation of rAAV expression by miRNA was primarily via site-specific cleavage of the transgene mRNA, generating specific 5′ and 3′ mRNA fragments. Our findings promise to facilitate the development of miRNA-regulated rAAV for CNS-targeted gene delivery and other applications.

Received 4 September 2010; accepted 19 November 2010; published online 21 December 2010. doi:10.1038/mt.2010.279

INTRODUCTION

Gene transfer mediated by recombinant adeno-associated virus (rAAV) holds promise for treatment of a large number of neurological disorders, but the blood–brain barrier blocks systemic vector delivery to the central nervous system (CNS). Nonetheless, previous proof-of-concept clinical studies using intracranial injection of AAV serotype 2-based vectors to the brain provided sustained gene expression and therapeutic effects.[1–4] Although localized vector delivery may be effective in treating CNS disorders that map to well defined anatomic and functional regions of the brain, the process requires invasive neurosurgical procedures and, therefore, is costly and potentially risky. If AAV vectors could cross the blood–brain barrier and be specifically expressed in the CNS, intravascular delivery of rAAV would provide a more effective method for gene therapy of CNS diseases that affect large areas of the brain and spinal cord.

Recent advancements in the discovery, rational design, and directed evolution of AAV have created a collection of distinct recombinant AAVs (rAAVs) appropriate for different gene transfer applications.[5] Among these vectors, rAAV6, rAAV8, and rAAV9 stand out for their ability to deliver genes transvascularly to target tissues.[6–9] Intravascularly administered rAAV9 and some other rAAVs with capsid modifications can cross the blood–brain barrier, efficiently transferring genes to the CNS in mice, cats, and nonhuman primates.[10–14] Nonetheless, several potential challenges remain to using rAAV9 to treat CNS disorders. The first challenge is to deliver rAAV specifically to the CNS. The viral capsid is the principal determinant for AAV tissue tropism,[5] and the liver is the major target for AAV vectors. Systemic administration of some rAAV serotypes transduces the liver and other tissues, including the CNS, skeletal muscle, heart, pancreas, and antigen-presenting cells.[6,7] Such off-target transduction raises the specter of overexpression of transgenes outside the CNS, potentially eliciting toxic responses. Thus, the second challenge is to confine transgene expression to the CNS when rAAV9 is delivered systemically. Historically, CNS-specific promoters have been used to limit transgene expression to the CNS. However, tissue-specific, strong CNS promoters are often too large to be packaged into the rAAV genome.

As an alternative, we used endogenous microRNAs (miRNAs) to suppress transgene expression outside the CNS. miRNAs are small, noncoding RNAs that regulate gene expression by post-transcriptional silencing. miRNAs silence genes

Correspondence: *Guangping Gao, Gene Therapy Center, University of Massachusetts Medical School, 381 Plantation Street, Suite 250, Worcester, Massachusetts 01655, USA. E-mail: guangping.gao@umassmed.edu or Phillip D Zamore, Howard Hughes Medical Institute, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. E-mail: phillip.zamore@umassmed.edu*

by two mechanisms. When partially complementary to mRNA sequences, they typically reduce target mRNA stability and protein expression by two- to fourfold or less, a mode of regulation thought to tune mRNA expression.[15] In contrast, when miRNAs are nearly perfectly complementary to their mRNA targets, they cleave the mRNA, triggering its wholesale destruction. miRNA-binding sites were first used in lentiviral vectors to suppress transgene expression in hematopoietic cells, thereby attenuating transgene immunogenicity.[16] The strategy was subsequently used to regulate naked DNA-mediated gene transfer, detarget oncolytic viral therapeutics from noncancer tissues to reduce host toxicity, and, in another application, attenuate live, attenuated virus-based viral vaccines to reduce toxicity in vaccine recipients.[16–21] The use of miRNA detargeting poses special challenges for systemic delivery of rAAV to transduce the CNS. Intravascularly delivered rAAV for CNS-targeted transduction requires suppressing high levels of rAAV expression in multiple peripheral tissues even when transgene transcripts are successfully expressed in the brain.

Here, we report the use of miRNAs to detarget rAAV9 expression both separately and concurrently in the liver, heart, and skeletal muscle, the three tissues that are most efficiently targeted by intravenously delivered rAAV9.[7,8] Silencing of transgene expression in liver, heart, and muscle exploited the natural expression of the abundant (≥60,000 copies/cell) miRNAs, miR-122, which is expressed in hepatocytes, and miR-1, a miRNA found in the heart and skeletal muscle of virtually all animals.[22,23] miR-122-binding sites have been successfully used to prevent hepatotoxicity of a transgene from an adenovirus vector.[18] Perfectly complementary sites for miR-1, miR-122, or both were engineered into the 3′ untranslated region (UTR) of a nuclear-targeted, β-galactosidase (*nLacZ*) reporter transgene whose expression was driven by a cytomegalovirus-enhancer, chicken β-actin (CB) promoter. We present multiple independent lines of evidence to show that the miRNAs repress *nLacZ* expression by cleaving the transgene mRNA at exactly the same site as by all Argonaute-bound small RNAs in eukaryotic cells.[24] When delivered systemically *in vivo*, the miRNA-detargeted rAAV9 vector successfully expressed the reporter transgene in the CNS, but not the liver or heart or skeletal muscle.

RESULTS

miRNAs efficiently repress reporter gene expression in cultured cells

Endogenous miRNA-mediated post-transcriptional gene silencing has proven to be an effective and tissue-specific approach to regulate transgene or viral gene expression *in vivo* for plasmids, viral vectors, or live, attenuated viral vaccines.[16–21] To evaluate this strategy for rAAV-mediated transduction, we introduced one or three tandem copies of a perfectly complementary binding site for miR-1 or miR-122 into the 3′ UTR of *nLacZ* in a rAAV plasmid vector. We transfected the constructs into HuH7 cells, a human hepatoma cell line expressing ~16,000 copies of miR-122 per cell,[23] and measured the number of nLacZ-positive cells. The number of nLacZ-expressing HuH7 cells for the one-site plasmid was about half that of the no site control; three sites reduced the number of nLacZ-expressing cells more than sevenfold (**Figure 1a**).

Next, we analyzed expression of the *nLacZ* constructs in human embryonic kidney 293 cells, which naturally express low levels of both miR-122 and miR-1, when miR-1 or miR-122 was introduced as a pri-miRNA from a second plasmid. We transfected 293 cells with the *nLacZ* reporter plasmids carrying 0, 1, or 3 miR-122 or miR-1-binding sites, together with a plasmid expressing either pri-miR-122 (**Figure 1b**) or pri-miR-1 (**Figure 1c**). To vary the concentration of the miRNA, we used either a low (1:3) or a high (1:10) molar ratio of the *nLacZ*-binding site plasmid to the miRNA expression plasmid. When miR-122 or miR-1 was introduced into the cells, *nLacZ* expression was repressed only when the *nLacZ* reporter mRNA contained the corresponding miRNA-binding sites; there was no reduction of nLacZ-positive cells when miR-1 was coexpressed with *nLacZ* containing miR-122-binding sites or when miR-122 was coexpressed with *nLacZ* containing miR-1-binding sites (**Figure 1b,c**).

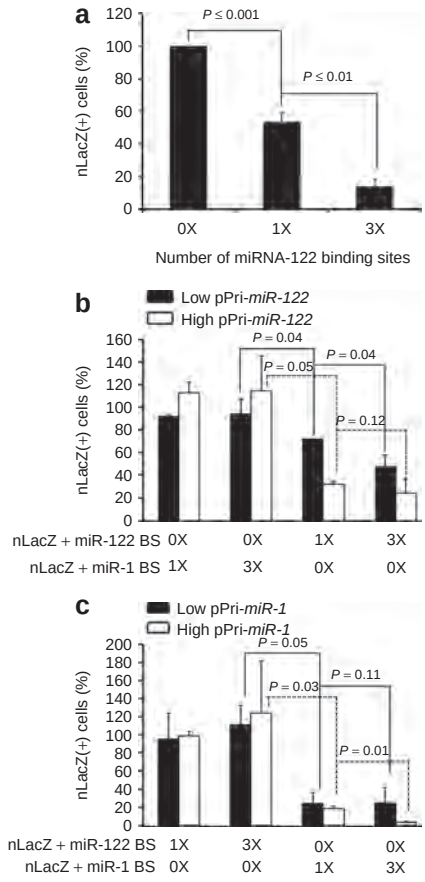


Figure 1 ***In vitro*** **validation of artificial miRNA-binding sites for reporter silencing.** Plasmids harboring the rAAVCB*nLacZ* genome with or without miR-1 or miR-122-binding sites were transfected into human hepatoma (HuH7) cells (**a**) which express miR-122 or cotransfected into 293 cells, together with a plasmid expressing either pri-miR-122 (**b**) or pri-miR-1 (**c**) at molar ratios of 1:3 (low) or 1:10 (high). 0X: no miRNA-binding site; 1X: one miRNA-binding site; 3X: three miRNA-binding sites. The cells were fixed and stained histochemically with X-gal 48 hours after transfection and blue cells counted. The percentage of nLacZ-positive cells in each transfection were compared to transfection of the control plasmid (prAAVCB*nLacZ*). CB, chicken β-actin; miR, microRNA; *nLacZ*, β-galactosidase reporter transgene; rAAV, recombinant adeno-associated viruses.

Tissue-specific endogenous miRNAs can regulate expression of rAAV9 delivered systemically in adult mice

To evaluate miRNA regulation of systemically delivered AAV9CB*nLacZ* vectors *in vivo*, we produced AAV9CB*nLacZ* vectors carrying 0, 1, or 3 miRNA-binding sites perfectly complementary to either miR-122 or miR-1. The vectors were administered by tail vein injection to adult male C56BL/6 mice at a dose of 5×10^{13} genome copies per kg (GC/kg) body weight. Four weeks later, we examined the liver and heart of the transduced animals. LacZ staining revealed that the *nLacZ* transgene was silenced by the endogenous miRNAs in the cell type and organ in which they are predominantly expressed: the transgene was specifically silenced by miR-122 in the liver and by miR-1 in the heart (**Figure 2a,b**). While nLacZ positive cells were reduced in the livers of the animals treated with rAAV9CB*nLacZ* bearing one or three miR-122-binding sites, nLacZ expression levels in the hearts of the same animals were similar to those in the animals treated with AAV9CB*nLacZ* bearing no sites (**Figure 2a**). Similarly, nLacZ expression was not detected in the hearts of the animals that received AAV9CB*nLacZ* containing one or three miR-1-binding sites, but nLacZ expression in the livers of the same animals was not affected as compared to that in the control animal (**Figure 2b**). Our data suggest that the greater the number of sites for a miRNA in rAAV, the lower the nLacZ expression in the tissue where the corresponding miRNA is expressed (**Figure 2a,b**).

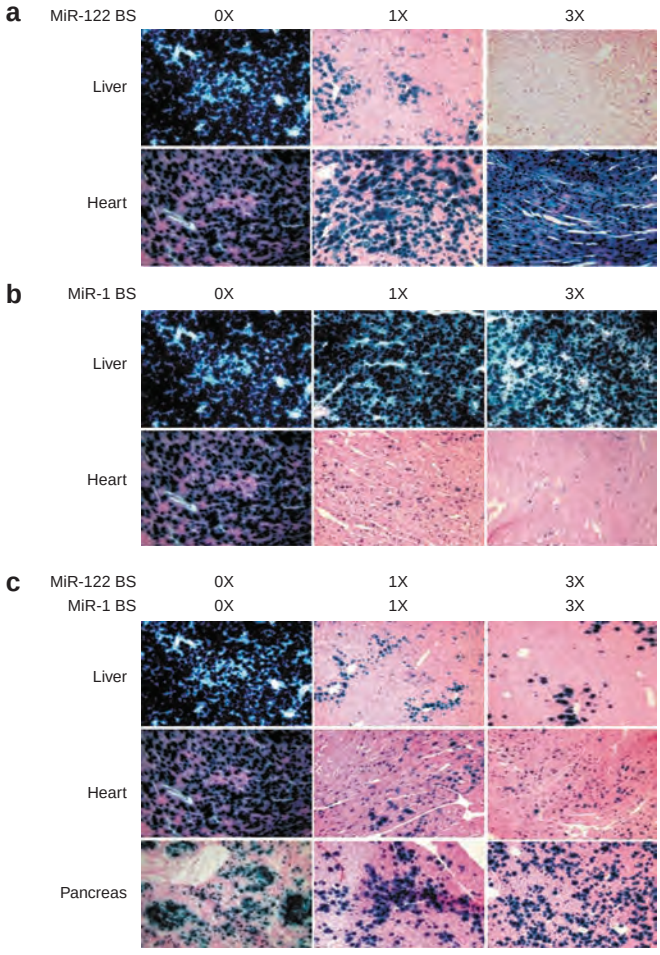


Figure 2 ***In vivo*** **evaluation of endogenous miRNA-mediated transgene silencing in rAAV9 transduction.** (**a–c**) Adult male C58BL/6 mice were injected intravenously with 5×10^{13} genome copies per kg (GC/kg) each of rAAV9CB*nLacZ* (no binding site), (**a**) rAAVCB9*nLacZ-miR-122BS* (one miR-122-binding site) and rAAV9CB*nLacZ-(miR-122BS)*$_3$ (three miR-122-binding sites), (**b**) rAAV9CB*nLacZ-miR-1BS* (one miR-1 binding site) and rAAV9CB*nLacZ-(miR-1BS)*$_3$ (three miR-1-binding sites, and (**c**) rAAV9CB*nLacZ-miR-1BS-miR-122BS* (1X each binding site) and rAAV9CB*nLacZ-(miR-1BS)*$_3$*–(miR-122BS)*$_3$ (three miR-1 and three miR-122-binding sites). The animals were necropsied 4 weeks after vector administration, and appropriate tissues were harvested for cryosectioning and X-gal histochemical staining. miR, microRNA; *nLacZ*, β-galactosidase reporter transgene; rAAV, recombinant adeno-associated viruses.

Next, we tested whether transgene silencing could be achieved simultaneously in multiple tissues. We inserted different numbers of both miR-122- and miR-1-binding sites in the 3′ UTR of the rAAV9CB*nLacZ* genome and examined their expression in rAAV9 transduced mice. Histochemical staining of tissue sections showed that nLacZ expression was suppressed in both heart and liver for rAAV9CB*nLac* containing one or three copies each of the miR-1- and miR-122-binding sites, but nLacZ was readily detectable in pancreas, where expression of both miR-122 and miR-1 is low[25] (**Figure 2c**). Quantitative, β-galactosidase assays of homogenized liver tissue similarly showed that nLacZ expression was significantly lower when the transgene contained the miRNA-binding sites (one miR-122-binding site: 7.8 ± 7.4%, *P* value = 0.005; three miR-122-binding sites: 1.6 ± 1.0%, *P* value = 0.005; one miR-1- plus one miR-122-binding site: 8.6 ± 5.7%, *P* value = 0.005; three miR-1- plus three miR-122-binding sites: 3.1 ± 1.2%, *P* value = 0.005; three miR-1-binding sites: 105.7 ± 11.6%) (**Supplementary Figure S1**).

miRNA repression of rAAV expression does not perturb endogenous miRNA pathways

Highly expressed transgenes bearing miRNA-complementary sites have been reported to promote degradation of the corresponding miRNA.[26] We examined the levels of miR-122, miR-22, miR-26a, and *let-7* in rAAV transduced liver. We detected no difference in abundance of the four miRNAs among the three study groups (**Figure 3a**). Moreover, our preliminary data from high throughput sequencing analyses of small RNA from the livers of one animal each from the three study groups show no change in miRNA levels (data not shown).

Introducing miRNA-complementary sites into a highly expressed transgene might also divert the corresponding miRNA to the transgene mRNA from its natural targets, derepressing them. We tested whether the miRNA-binding sites in the transgene transcripts would deregulate the expression of the known endogenous target mRNAs of miR-122 or miR-1. We analyzed the expression of cyclin G1, a miR-122 target in liver (**Figure 3b,c**) and calmodulin, a miR-1 target in heart (**Figure 3d**). We detected no significant alteration in cyclin G1 or calmodulin expression. miR-122 regulates cholesterol biosynthesis in the liver, and agents that block miR-122 function produce readily detectable changes in serum cholesterol levels.[27] We could detect no change in total cholesterol, high-density lipoprotein, or low-density lipoprotein levels in mice 4 weeks after transduction with either control rAAV9 or rAAV9 expressing a transgene bearing miR-122-binding sites (**Figure 3e**). We conclude that miRNA-mediated detargeting of

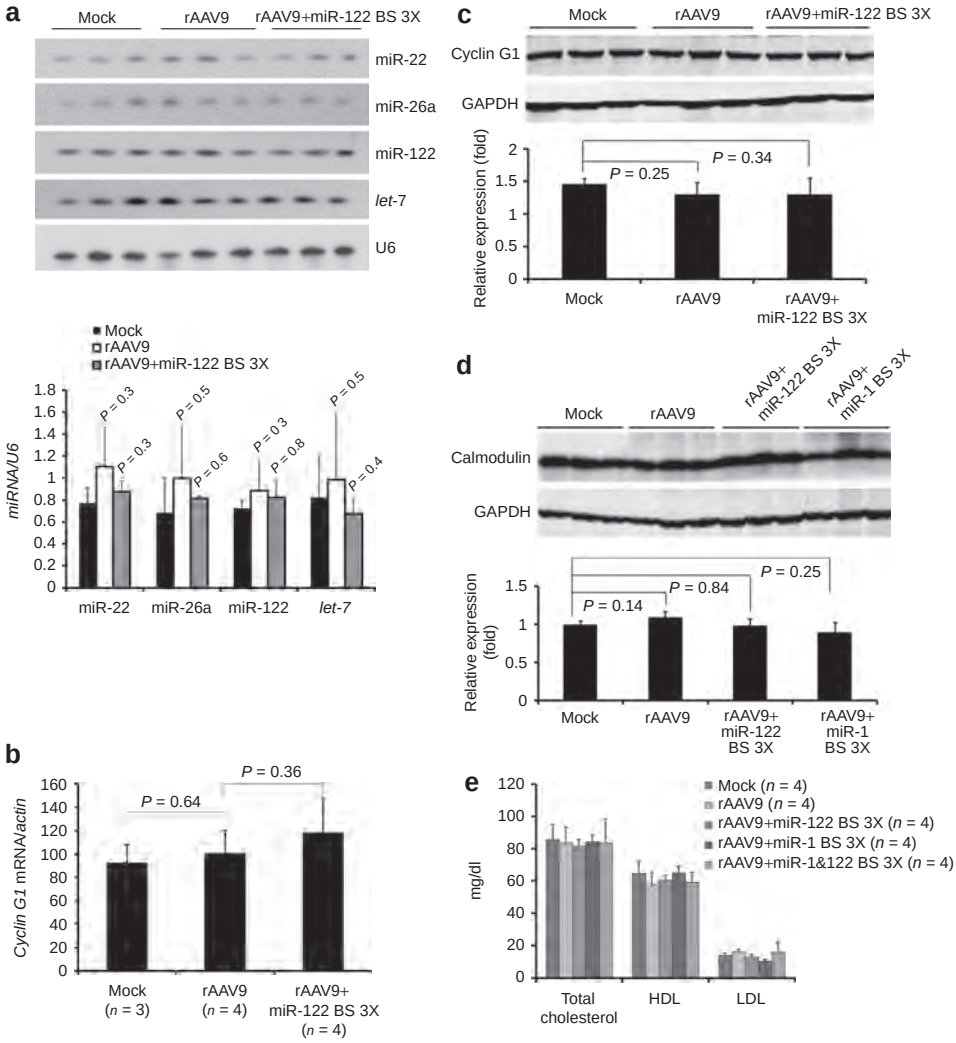


Figure 3 Analysis of expression levels of cognate miRNA, mRNA, and protein of endogenous miRNA target genes in mice transduced with rAAV9CB*nLacZ* with or without miRNA-binding sites. Total cellular RNA or protein was prepared from (**a–c**) liver or (**d**) heart. (**a**) Northern blot detection of miRNAs. U6 small nuclear RNA provides a loading control. (**b**) Quantitative reverse-transcription PCR measuring cyclin G1 mRNA. The data are presented as relative *cyclin G1* mRNA levels normalized to β-actin. (**c**,**d**) Western blot analyses of protein levels of endogenous targets of miR-122 and miR-1. Total cellular protein prepared from (**c**) liver or (**d**) heart was analyzed for cyclin G1 and calmodulin. (**e**) Serum cholesterol levels. Serum samples from mice that received rAAV9 with or without miRNA-binding sites were collected after 4 weeks and measured for total cholesterol, high-density lipoprotein (HDL) and low-density lipoprotein (LDL). miR, microRNA; *nLacZ*, β-galactosidase reporter transgene; rAAV, recombinant adeno-associated viruses.

rAAV expression has no detectable effect on endogenous miRNA expression or function.

Endogenous miRNAs silence rAAV transduction by site-specific cleavage of transgene mRNA

To determine how miRNAs suppress expression of transgenes delivered by rAAV *in vivo*, we characterized the transgene mRNA in liver by conventional PCR (**Figure 4b**), quantitative reverse-transcription PCR (qRT-PCR) (**Figure 4c**), Northern hybridization (**Figure 4d,e**), and rapid amplification of 5′ complimentary DNA (cDNA) ends (5′ RACE; **Figure 4f**). When we used primers that amplify the region between the 3′ end of *nLacZ* (A^+F primer) and the 5′ end of the poly(A) signal (A^+R primer), an amplicon that spans the miRNA-biding sites, we detected a 145 basepair (bp) product after 26 cycles of amplification for the samples that received control rAAV. An additional six cycles of amplification were required to detect a weak 220 bp band for the samples transduced by rAAV containing three miR-122-binding sites. These data are consistent with low levels of intact *nLacZ* mRNA (**Figure 4a,b**).

To quantitatively assess the extent of the miRNA-directed repression of the transgene transcripts, we performed qRT-PCR using either oligo(dT) or random hexamer primers for reverse-transcription and PCR primer pairs that span either a 5′ (nLacZ5′F/5′R), or 3′ (nLacZ 3′F/3′R) region of the *nLacZ* coding sequence (**Figure 4a**). We examined the levels of *nLacZ* mRNA with intact 5′ and 3′ ends in total liver RNA extracted from four animals that received the control rAAV9CB*nLacZ* and four that received rAAV9CB*nLacZ* containing three miR-122-binding sites in the 3′ UTR. We observed reductions ranging from 3 ± 1

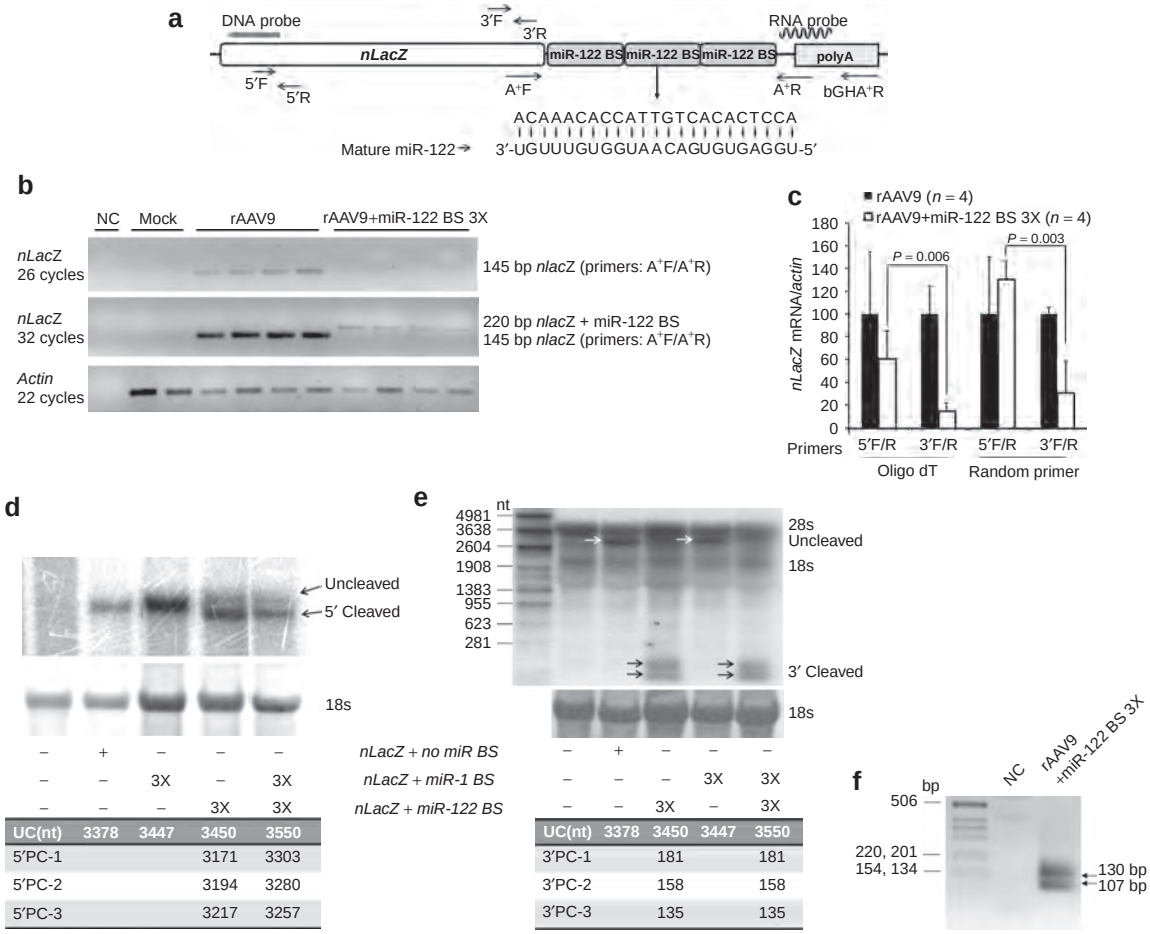


Figure 4 Molecular characterization of transgene mRNAs with or without miRNA-binding sites. (**a**) Locations of the probes and primers, the sequences of mature miR-122 and its perfectly complementary binding site in the transgene mRNA are presented. (**b**) Total cellular RNA from liver was analyzed either by conventional reverse-transcription PCR (RT-PCR) by using primers that span a region between the 3′ end of *nLacZ* and the 5′ end of poly(A) signal (**c**) or by quantitative RT-PCR; data are presented as relative nLacZ mRNA levels normalized to β-actin. (**d**) For the northern blot analysis of *nLacZ* mRNA, 18S RNA served as a loading control, and the blots were hybridized with either a transgene DNA (**e**) or RNA probe. (**f**) In addition, poly(A) bearing mRNA from the liver of an animal received rAAV containing three miR-1- and three miR-122-binding sites was analyzed by 5′ RACE; the PCR product was resolved on an ethidium bromide-stained agarose gel. miR, microRNA; *nLacZ*, β-galactosidase reporter transgene; rAAV, recombinant adeno-associated viruses.

(random hexamer) to 7 ± 1 (oligo[dT])-fold in *nLacZ* mRNA with an intact 3′ end in the animals that had received rAAV9 containing miR-122-binding sites, relative to the control. In contrast, we detected little or no decrease in *nLacZ* mRNA with an intact 5′ end for the same samples using the 5′F/5′R primer pair (**Figure 4c**). Our results suggest that the primary mode of turnover of the mRNA that has been cleaved by a miRNA is 3′-to-5′ exonucleolytic degradation.

To further characterize the fate of the transgene mRNA targeted by miR-1 or miR-122, we performed Northern blot analyses. A transgene probe binding to the 5′ end of *nLacZ* mRNA detected a ~3.4 kb RNA in an animal injected with control rAAV9Cb*nLacZ*, the expected size of the of the full-length *nLacZ* transcript; a slightly larger band was detected in the liver sample from a mouse treated with rAAV9CB*nLacZ* bearing three miR-1-binding sites (**Figure 4a,d**). In contrast to the single transcript detected for the rAAV9 expressing *nLacZ* bearing three miR-1-binding sites, two RNAs of different sizes were detected for the rAAV expressing *nLacZ* bearing three miR-122 sites (**Figure 4d**). The lengths of these transcripts suggest that the longer transcript likely represents the full-length mRNA, whereas the shorter, more abundant transcript corresponds to 5′ fragments of *nLacZ* RNA cleaved by miR-122 at the corresponding miR-122-binding sites in the 3′ UTR (**Figure 4d**).

To confirm this finding, we repeated the Northern analysis using an RNA probe spanning a portion of 3′ UTR of the transgene mRNA. In addition to detecting full-length *nLacZ* transcripts in the samples transduced by rAAV9 lacking miRNA-binding sites, two closely migrating species smaller than the 281 nucleotide RNA marker were detected. The size of these fragments is consistent with miRNA-directed 3′ cleavage products of the *nLacZ* mRNA (**Figure 4e**). These two 3′ cleavage products were also detected by gel electrophoresis of the product from the 5′ RACE experiment described below (**Figure 4f**).

To determine whether such target cleavage occurs *in vivo* when the *nLacZ* transcript contained miR-1 or miR-122-binding sites, we performed rapid amplification of 5′ cDNA ends (5′ RACE). **Figure 5** presents the sequences of 21 clones recovered using 5′ RACE from

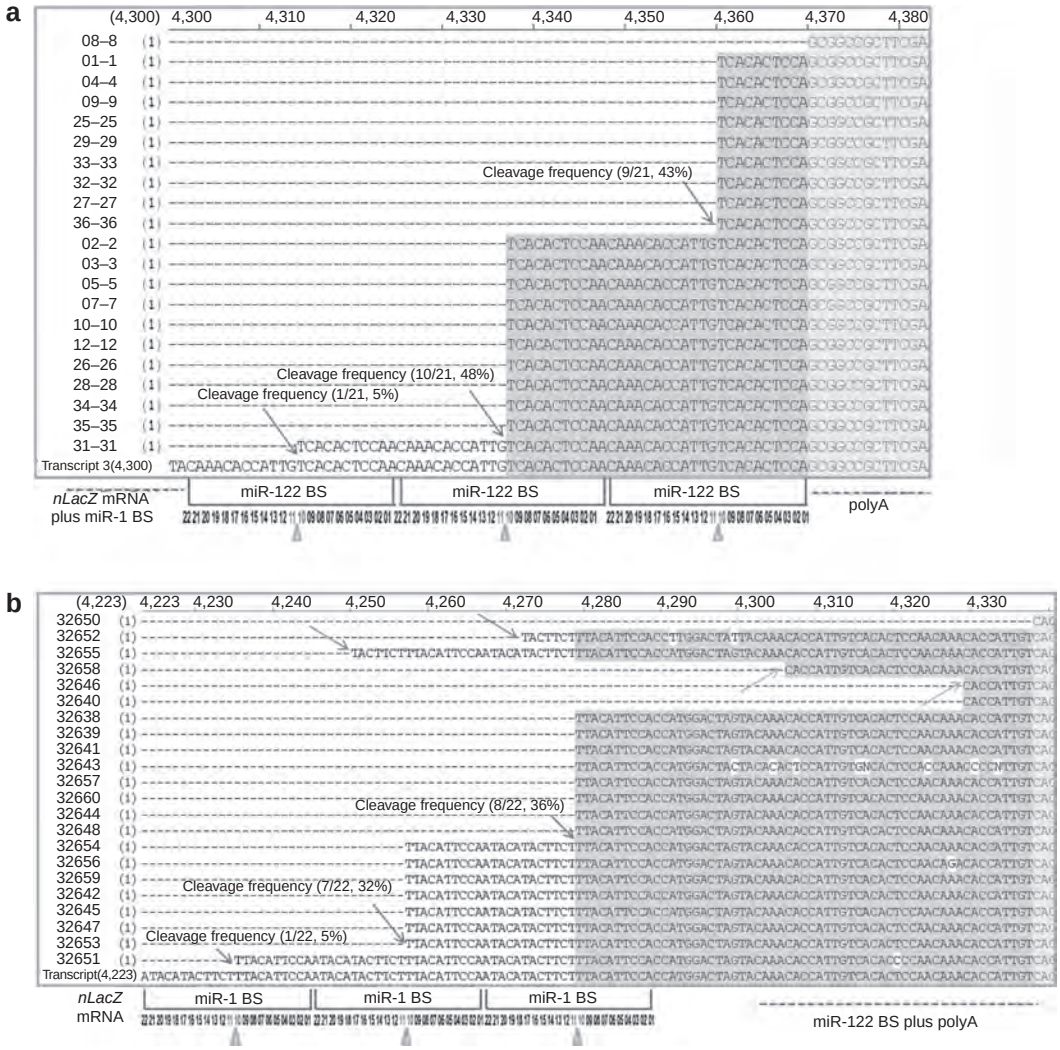


Figure 5 Alignment of sequences spanning the miRNA-binding sites and poly(A) signal regions recovered by 5′ RACE. Poly(A)-containing mRNA was isolated from the (**a**) liver and (**b**) heart of an animal injected with rAAV9CB*nLacZ-(miR-1BS)*$_3$*–(miR-122BS)*$_3$. Twenty-one liver-derived and 22 heart-derived clones were sequenced. The putative cleavage sites in each clone are identified by arrows; the frequencies of miRNA-directed, site-specific cleavage for each miRNA-binding site are reported; triangles point to the positions of the expected miRNA-directed cleavage sites (**a**,**b**). miRNA, microRNA, *nLacZ*, β-galactosidase reporter transgene; rAAV, recombinant adeno-associated viruses.

liver RNA (Figure 5a) and 22 clones isolated from heart RNA (Figure 5b) from the animals injected with rAAV9 in which the *nLacZ* 3′ UTR contained three miR-1 and three miR-122-binding sites. In liver, we detected the sequence signatures for miR-122-directed cleavage of the transgene mRNA at each miR-122-binding site: 5% for the first binding site, 48% for the second binding site, and 43% for the third binding site. All 5′ ends mapped to the phosphate that lies between the target nucleotides that pair with positions 10 and 11 of the sequence perfectly complementary to miR-122, the precise site cleaved by small RNAs bound to Argonaute proteins in all eukaryotes[24] (Figure 5a). Similar results were obtained in the heart for the miR-1 sites (Figure 5b).

Table 1 presents an expanded 5′ RACE analysis for additional vector groups. We note that none of the 5′ RACE products sequenced corresponded to miR-1-directed site-specific cleavage in liver or miR-122-directed site-specific cleavage in heart (Table 1). Although no cleavage was detected within miR-1-binding sites in the liver, some clones from heart were cleaved within the miR-122-binding sites, but not at the hallmark position for miRNA-directed cleavage.

Intravascularly delivered rAAV9 can be efficiently controlled by endogenous miRNAs

Recent studies suggest that high doses of self-complementary AAV9 (scAAV9), delivered intravascularly efficiently transduce the CNS in both mice and cats.[11,12] However, intravenous administration of such vectors also transduces liver, heart, and pancreas. We added miRNA-1 and miRNA-122-binding sites into the scAAV9CB enhanced GFP (*EGFP)* vector genome and injected 10-week-old C57BL/6 male mice with 2×10^{14} GC/kg. After 3 weeks, we prepared 40 µm sections of brain and spinal cord and 8 µm sections of liver, heart, and skeletal muscle and examined EGFP protein expression. As reported previously,[10–13] intravenously delivered scAAV9CB*EGFP* efficiently transduced the CNS; EGFP was

Table 1 Summary of microRNA-guided transgene mRNA cleavage in mouse liver and heart

miR BS cleavage			Position	Cleavage site: Between 10 and 11 nt		Between 17 and 18 nt		Between 18 and 19 nt		Random site
Liver	1 Copy of miR-122 BS (21 clones)		1	17/21	81%	ND		ND		19%
	3 Copies of miR-122 BS (11 clones)		1	ND	100%	ND		ND		0%
			2	4/11						
			3	7/11						
	3 Copies each of miR-1 and miR-122 BS in a single vector (21 clones)	miR-1 3x BS	1	ND	ND	ND		ND		0%
			2	ND						
			3	ND						
		miR-122 3x BS	1	1/21	95%	ND		ND		5%
			2	10/21						
			3	9/21						
Heart	1 Copy of miR-1BS (12 clones)		1	12/12	100%	ND		ND		0%
	3 Copies of miR-1BS (21 clones)		1	ND	80%	4/21	20%	ND		0%
			2	16/21		ND				
			3	1/21		ND				
	3 Copies each of miR-1 and miR-122 BS in a single vector (22 clones)	miR-122 3x BS	1	ND	ND	ND		1/22	14%	4%
			2	ND				1/22		
			3	ND				ND		
		miR-1 3x BS	1	1/22	73%	ND	9%	ND		0%
			2	7/22		1/22				
			3	8/22		1/22				

Abbreviations: BS, binding site; miR, microRNA; ND, not detected.

readily detectable in the thalamus region of the brain and the cervical region of the spinal cord, but also in non-CNS tissues such as liver, heart, and muscle (**Figure 6a**). In contrast, transgene expression in those non-CNS tissues was reduced when miR-1 and miR-122-binding sites were included in the transgene; EGFP expression was unaltered in the CNS, where miR-1 and miR-122 are not present (**Figure 6a**). We also used qRT-PCR to measure the differential expression of the miRNA-repressed EGFP transgene in brain (41.2 ± 7.7%), liver (3.0 ± 0.5%), heart (0.4 ± 0.1%), and muscle (1.3 ± 0.4%), relative to the EGFP transgene lacking miRNA-binding sites (**Figure 6b**). To eliminate changes caused by variations in transduction efficiency between experiments, we normalized the data to the number of vector genomes detected in the experimental and control samples. Like the microscopic analyses of native EGFP expression, the qRT-PCR data show that the presence of miR-122- or miR-1-binding sites reduced transgene expression in liver (20-fold), heart (100-fold), and muscle (50-fold), but did not detectably alter transgene expression in brain.

DISCUSSION

Discovery that rAAV9 can efficiently cross the blood–brain barrier to deliver genes to the CNS represented a major advance toward the goal of developing gene therapy for CNS disorders.[12,14] Here, we report that rAAV9 can be engineered so that endogenous miRNAs repress transgene expression outside the CNS. This strategy provides another step toward the development of systematically deliverable, CNS-targeted rAAV-based neurotherapeutics.

Recent advances in AAV vector development have substantially improved primate AAV vectors that achieve efficient and stable transvascular transduction in multiple tissues simultaneously.[6,7] For example, systemic delivery of rAAV9 achieves extensive and robust transduction of astrocytes throughout the CNS of adult mice.[11,12] This unique feature of rAAV9 promises to provide a protective therapy for the degenerating neurons associated with Parkinson's disease, Alzheimer's disease and amyotrophic lateral sclerosis, by expressing neurotrophic growth factors such as insulin-like growth factor, brain-derived neurotrophic factor or glial-derived neurotrophic factor in the transduced astrocytes.[28] Yet, non-CNS expression derived from the peripheral tissues transduced by systemically delivered rAAV9 could lead to long-term circulation of super-physiological and potentially toxic levels of those growth factors.[29–31] Thus, in some applications, CNS-directed gene transfer may also cause toxicity and immune responses if the transgene is expressed outside the CNS.

Achieving transgene expression in only the target tissues is vital for the clinical development of safe CNS gene delivery. Here, we have shown that endogenous miRNAs can be harnessed to restrict the tissue- and cell-type specificity of rAAV expression, as was initially shown for lentiviral vectors.[16] Our data demonstrate that endogenous miRNAs can effectively repress transgene expression

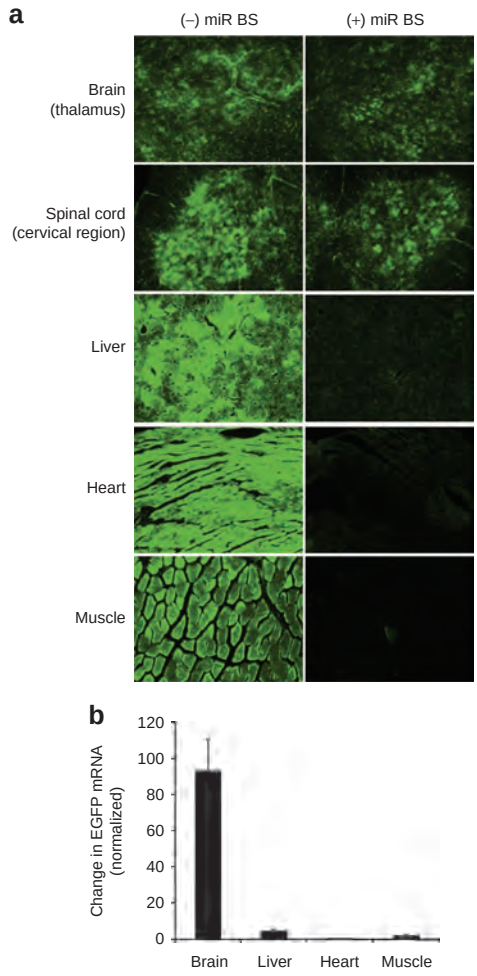


Figure 6 Endogenous miRNA-repressed, CNS-directed EGFP gene transfer by systemically delivered rAAV9. Ten-week-old male C57BL/6 mice were injected intravenously with scAAV9CBEGFP or scAAV9CB*nLacZ-(miR-1BS)*$_3$*–(miR-122BS)*$_3$ at a dose of 2×10^{14} genome copies per kg (GC/kg) body weight. The animals were necropsied 3 weeks later for whole body fixation by transcardiac perfusion. (**a**) Brain, spinal cord, liver, heart, and muscle were harvested for cryosectioning, immunofluorescent staining for EGFP (brain and cervical spinal cord), and fluorescence microscopy to detect EGFP. Total cellular DNA and RNA were extracted from brain, liver, heart and muscle to measure the amount of persistent vector genome by qPCR and EGFP mRNA by qRT-PCR. (**b**) For each tissue, the relative abundance of the EGFP mRNA containing miRNA-binding sites was compared to that of the EGFP mRNA lacking miRNA-binding sites. For each sample, mRNA abundance was normalized to the amount of vector genome detected in the tissue. EGFP, enhanced green fluorescent protein; miRNA, microRNA; *nLacZ*, β-galactosidase reporter transgene; qRT-PCR, quantitative reverse-transcription PCR; rAAV, recombinant adeno-associated viruses.

from rAAV. In both heart and liver, the miRNAs repressed transgene expression by directing endonucleolytic cleavage of the transgene mRNA (**Figure 7**). Importantly, miRNA regulation of rAAV expression did not perturb the expression or function of the corresponding endogenous miRNA, allowing transgene expression to be restricted to the CNS in mice. Our data suggest that a strategy that combines multiple binding sites for miRNAs expressed in the periphery but not the CNS may enable the development of safer, CNS-specific gene therapy vectors. In the future, validation

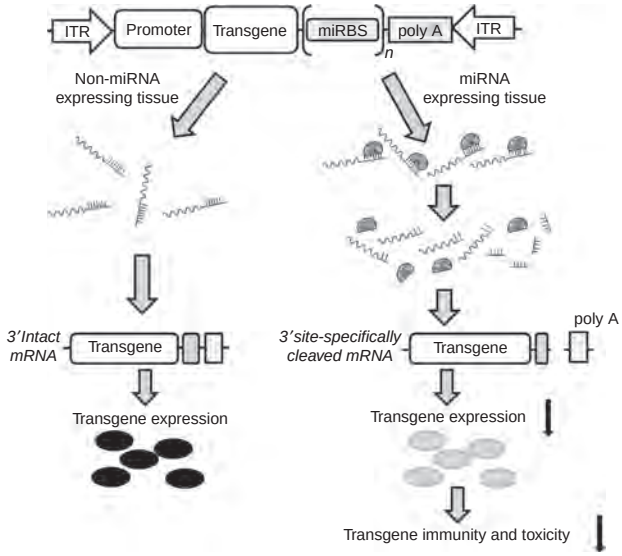


Figure 7 A molecular model for endogenous miRNA-regulated rAAV expression. miRNA, microRNA; rAAV, recombinant adeno-associated viruses.

of miRNA-mediated restriction of rAAV transgene expression in large animal models is required to pave the way for the development of intravascularly delivered rAAV-based gene therapy to treat disorders that affect the CNS.

MATERIALS AND METHODS

Vector design, construction, and production. Perfectly complementary miRNA-binding sites were designed based on the annotated miR-1 and miR-122 sequences in miRBase[32] and inserted into the *BstB*I restriction site in the 3′ UTR of the nLacZ expression cassette of the ubiquitously expressed pAAVCB nuclear-targeted β-galactosidase (nLacZ) plasmid using synthetic oligonucleotides (**Figure 4a** and **Supplementary Table S1**). This vector uses a hybrid cytomegalovirus enhancer/CB promoter cassette that is active in most cells and tissues. To express miR-122 and miR-1, pri-miR-122 and pri-miR-1 fragments were amplified by PCR from C57/B6 mouse genomic DNA (**Supplementary Table S1**) and inserted into the XbaI restriction site 3′ to a firefly luciferase cDNA in the pAAVCBF-Luc plasmid. The identity of each pri-miRNA was verified by sequencing. AAV9 vectors used in this study were generated, purified, and titered as described.[33]

Cell culture and transfection. HEK-293 and HuH7 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 100 mg/l of penicillin-streptomycin (Hyclone, South Logan, UT). Cells were maintained in a humidified incubator at 37 °C and 5% CO_2. Plasmids were transiently transfected using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

Mouse studies. Male C57BL/6 mice (Charles River Laboratories, Wilmington, MA) were obtained and maintained and all animal procedures performed according to the guidelines of the Institutional Animal Care and Use Committee of the University of Massachusetts Medical School. To monitor lipid profiles of the study animals, serum samples were collected 4 weeks after rAAV9 injection and analyzed for total cholesterol, high-density lipoprotein and low-density lipoprotein on a COBAS C 111 analyzer (Roche Diagnostics, Lewes, UK). To evaluate endogenous miRNA-mediated, CNS-restricted EGFP gene transfer, 10-week-old male C57BL/6 mice were injected intravenously (tail vein) with AAV9CBnLacZ-[miR-122-binding site $(BS)_1$], AAV9CBnLacZ-(miR-122BS)$_3$, AAV9CBnLacZ-(miR-1BS)$_1$,

AAV9CBnLacZ-(miR-1BS)$_3$, AAV9CBnLacZ-(miR-1BS)$_1$-(miR-122BS)$_1$, and AAV9CBnLacZ-(miR-1BS)$_3$-(miR-122BS)$_3$122BS)$_3$, respectively, at 5×10^{13} GC/kg body weight) or scAAV9CBEGFP at 2×10^{14} GC/kg body weight). Animals receiving nLacZ vectors were necropsied 4 weeks later; 8 µm cryosections of liver, heart, and pancreas tissues were prepared for X-gal-histochemical staining. Animals that received EGFP vectors were necropsied 3 weeks later and fixed by transcardial perfusion with 4% (wt/vol) paraformaldehyde. Brain, spinal cord, liver, heart, and muscle were harvested for cryosectioning. Brain and cervical spinal cord tissue were stained as floating sections in a 12-well plate using rabbit anti-EGFP antibody (Invitrogen) diluted 1:500, followed by goat anti-rabbit secondary antibody (Invitrogen) diluted 1:400. Outside the CNS, EGFP expression was detected directly by fluorescence. EGFP and antibody fluorescence was recorded using a Nikon TE-2000S inverted microscope at ×10 magnification and an exposure time of 3 seconds for liver, heart, and muscle, and 5 seconds for thalamus (brain) and cervical spinal cord.

Vector genome quantification by qPCR. Genome DNA was extracted from the selected tissues using QIAamp DNA Mini Kit (Qiagen, West Sussex, UK), according to the manufacturer's instructions. Quantitative PCR were carried out in triplicate using 50 ng DNA and 0.3 µmol/l EGFP-specific primers (EGFP-F and EGFP-R) using GoTaq qPCR master mix (Promega, Madison, WI) in a StepOne Plus real-time PCR instrument (Applied Biosystems, Foster City, CA).

qRT-PCR analysis. RNA was extracted using Trizol (Invitrogen), according to the manufacturer's instructions. Total RNA (0.5–1.0 µg) was primed with random hexamers or oligo(dT) and reverse-transcribed with MultiScribe Reverse Transcriptase (Applied Biosystems). Quantitative PCR were performed in triplicate with 0.3 µmol/l gene-specific primer pairs (nLacZ5′F/5′R, nLacZ 3′F/3′R, cyclinG1F/R and EGFP-F/EGFP-R) using the GoTaq qPCR master mix in a StepOne Plus Real-time PCR device. The specificity of qRT-PCR products derived from the 5′ and 3′ ends of *nLacZ* mRNA was confirmed by gel electrophoresis.

Northern blot analysis. Total RNA was extracted from mouse liver and analyzed by Northern hybridization.[34] To detect *nLacZ* mRNA, a 618 bp fragment of *nLacZ* cDNA was isolated by NcoI and PciI digestion of pAAVCB*nLacZ* and labeled with α-^{32}P dCTP by random priming (Takara, Shiga, Japan). To detect 3′ fragments of the cleaved nLacZ mRNA, an 111 bp fragment of the poly(A) sequence in the vector genome was cloned into pCR4-TOPO (Invitrogen) for preparation of antisense RNA probe labeled with α-^{32}P CTP during *in vitro* transcription using the Riboprobe System T7 kit (Promega). To detect miR-122, miR-26a, miR-22, and let-7 or U6 in total liver RNA, small RNAs were resolved by denaturing 15% polyacrylamide gels, transferred to Hybond N+ membrane (Amersham BioSciences, Pittsburgh, PA), and crosslinked with 254 nm light (Stratagene, La Jolla, CA). Synthetic oligonucleotides, 5′ end-labeled with γ-^{32}P ATP using T4 polynucleotide kinase (New England Biolabs, Beverly, MA), were used as DNA probes (**Supplementary Table S1**) and hybridized in Church buffer (0.5 mol/l $NaHPO_4$, pH 7.2, 1 mmol/l EDTA, 7% (w/v) sodium dodecyl sulphate) at 37 °C. Membranes were washed using 1× SSC (150 mM sodium chloride, 15 mM sodium citrate), 0.1% sodium dodecyl sulphate buffer, and then visualized using an FLA-5100 Imager (Fujifilm, Tokyo, Japan).

Western blot analysis. Proteins were extracted with radioimmunoprecipitation assay buffer [25 mmol/l Tris–HCl, pH 7.6, 150 mmol/l NaCl, 1% (vol/vol) NP-40, 1% (wt/vol) sodium deoxycholate, 0.1% (w/v) sodium dodecyl sulphate] containing a protease inhibitor mixture (Boston BP, Boston, MA). Protein concentration was determined using the Bradford method (Bio-Rad, Melville, NY). Protein samples, 50 µg each, were loaded onto 12% polyacrylamide gels, electrophoresed, and transferred to nitrocellulose membrane (Amersham BioSciences). Briefly, membranes were blocked with blocking buffer (LI-COR Biosciences, Lincoln, NE) at room temperature for 2 hours, followed by incubation with either anti-GAPDH (Millipore, Billerica, MA), anti-cyclin G1 (Santa Cruz Biotechnology, Santa Cruz, CA) or anti-calmodulin (Millipore) for 2 hours at room temperature. After three washes with PBS containing 0.1% (vol/vol) Tween-20, membranes were incubated with secondary antibodies conjugated to LI-COR IRDye for 1 hour at room temperature, and then antibodies detected using the Odyssey Imager (LI-COR).

β-Galactosidase assay. Proteins were extracted with radioimmunoprecipitation assay buffer and quantified as described above. Fifty micrograms of protein was used for each β-galactosidase assay using the Galacto-Star System (Applied Biosystems), according to the manufacturer's instructions.

5′ RACE. 5′ RACE was performed as described.[35] The 5′ RACE Outer Primer and the *nLacZ* gene-specific primer bGHpolyAR (**Supplementary Table S1**) were used for the first round of nested PCR. The 5′ RACE Inner Primer and the nLacZ gene-specific primer nLacZpolyR, which is located near the stop codon of nLacZ cDNA, were used for the second round of nested PCR (**Supplementary Table S1**). PCR products were TOPO-cloned into pCR-4.0 (Invitrogen) and sequenced.

Statistical analysis. All results are reported as mean ± SD and compared between groups using the two-tailed Student's *t*-test.

SUPPLEMENTARY MATERIAL

Figure S1. Quantification of β-galactosidase activities in liver tissue from animals that received rAAV*nLacZ* vectors with and without miRNA-binding sites.

Table S1. Oligonucleotide primers and probes used in the study.

ACKNOWLEDGMENTS

We greatly appreciate the contributions of the Vector Core of Gene Therapy Center and Analytic Cores of Diabetes Center of the University of Massachusetts Medical School. This work was funded by an internal grant to G.G. from the University of Massachusetts Medical School. The contributions of C.M. and T.R.F. to this work were supported with a PO1grant from the National Institute of Health (DK 58327). G.G., P.D.Z., C.C.M., and T.R.F. are members of UMASS DERC (DK 32520). This publication was made possible by grant 5P30DK 32520 from National Institute of Diabetes and Digestive and Kidney Diseases.

REFERENCES

1. Kaplitt, MG, Feigin, A, Tang, C, Fitzsimons, HL, Mattis, P, Lawlor, PA *et al.* (2007). Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. *Lancet* **369**: 2097–2105.
2. Muramatsu, S, Fujimoto, K, Kato, S, Mizukami, H, Asari, S, Ikeguchi, K *et al.* (2010). A phase I study of aromatic L-amino acid decarboxylase gene therapy for Parkinson's disease. *Mol Ther* **18**: 1731–1735.
3. Federici, T and Boulis, NM (2009). Invited review: festschrift edition of neurosurgery peripheral nervous system as a conduit for delivering therapies for diabetic neuropathy, amyotrophic lateral sclerosis, and nerve regeneration. *Neurosurgery* ***65***(4 Suppl): A87–A92.
4. Bjorklund, T and Kordower, JH (2010). Gene therapy for Parkinson's disease. *Mov Disord* **25** Suppl 1: S161–S173.
5. Vandenberghe, LH, Wilson, JM and Gao, G (2009). Tailoring the AAV vector capsid for gene therapy. *Gene Ther* **16**: 311–319.
6. Wang, Z, Zhu, T, Qiao, C, Zhou, L, Wang, B, Zhang, J *et al.* (2005). Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. *Nat Biotechnol* **23**: 321–328.
7. Zincarelli, C, Soltys, S, Rengo, G and Rabinowitz, JE (2008). Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther* **16**: 1073–1080.
8. Pacak, CA, Mah, CS, Thattaliyath, BD, Conlon, TJ, Lewis, MA, Cloutier, DE *et al.* (2006). Recombinant adeno-associated virus serotype 9 leads to preferential cardiac transduction *in vivo*. *Circ Res* **99**: e3–e9.
9. Foust, KD, Poirier, A, Pacak, CA, Mandel, RJ and Flotte, TR (2008). Neonatal intraperitoneal or intravenous injections of recombinant adeno-associated virus type 8 transduce dorsal root ganglia and lower motor neurons. *Hum Gene Ther* **19**: 61–70.
10. Chen, YH, Chang, M and Davidson, BL (2009). Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. *Nat Med* **15**: 1215–1218.
11. Duque, S, Joussemet, B, Riviere, C, Marais, T, Dubreil, L, Douar, AM *et al.* (2009). Intravenous administration of self-complementary AAV9 enables transgene delivery to adult motor neurons. *Mol Ther* **17**: 1187–1196.

12. Foust, KD, Nurre, E, Montgomery, CL, Hernandez, A, Chan, CM and Kaspar, BK (2009). Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes. *Nat Biotechnol* **27**: 59–65.
13. Gray, SJ, Blake, BL, Criswell, HE, Nicolson, SC, Samulski, RJ, McCown, TJ *et al.* (2010). Directed evolution of a novel adeno-associated virus (AAV) vector that crosses the seizure-compromised blood–brain barrier (BBB). *Mol Ther* **18**: 570–578.
14. Foust, KD, Wang, X, McGovern, VL, Braun, L, Bevan, AK, Haidet, AM *et al.* (2010). Rescue of the spinal muscular atrophy phenotype in a mouse model by early postnatal delivery of SMN. *Nat Biotechnol* **28**: 271–274.
15. Bartel, DP (2009). MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233.
16. Brown, BD, Venneri, MA, Zingale, A, Sergi Sergi, L and Naldini, L (2006). Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nat Med* **12**: 585–591.
17. Wolff, LJ, Wolff, JA and Sebestyén, MG (2009). Effect of tissue-specific promoters and microRNA recognition elements on stability of transgene expression after hydrodynamic naked plasmid DNA delivery. *Hum Gene Ther* **20**: 374–388.
18. Suzuki, T, Sakurai, F, Nakamura, S, Kouyama, E, Kawabata, K, Kondoh, M *et al.* (2008). miR-122a-regulated expression of a suicide gene prevents hepatotoxicity without altering antitumor effects in suicide gene therapy. *Mol Ther* **16**: 1719–1726.
19. Edge, RE, Falls, TJ, Brown, CW, Lichty, BD, Atkins, H and Bell, JC (2008). A let-7 MicroRNA-sensitive vesicular stomatitis virus demonstrates tumor-specific replication. *Mol Ther* **16**: 1437–1443.
20. Barnes, D, Kunitomi, M, Vignuzzi, M, Saksela, K and Andino, R (2008). Harnessing endogenous miRNAs to control virus tissue tropism as a strategy for developing attenuated virus vaccines. *Cell Host Microbe* **4**: 239–248.
21. Brown, BD and Naldini, L (2009). Exploiting and antagonizing microRNA regulation for therapeutic and experimental applications. *Nat Rev Genet* **10**: 578–585.
22. Lagos-Quintana, M, Rauhut, R, Yalcin, A, Meyer, J, Lendeckel, W and Tuschl, T (2002). Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**: 735–739.
23. Chang, J, Nicolas, E, Marks, D, Sander, C, Lerro, A, Buendia, MA *et al.* (2004). miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol* **1**: 106–113.
24. Elbashir, SM, Lendeckel, W and Tuschl, T (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**: 188–200.
25. Lee, EJ, Gusev, Y, Jiang, J, Nuovo, GJ, Lerner, MR, Frankel, WL *et al.* (2007). Expression profiling identifies microRNA signature in pancreatic cancer. *Int J Cancer* **120**: 1046–1054.
26. Ameres, SL, Horwich, MD, Hung, JH, Xu, J, Ghildiyal, M, Weng, Z *et al.* (2010). Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328**: 1534–1539.
27. Krützfeldt, J, Rajewsky, N, Braich, R, Rajeev, KG, Tuschl, T, Manoharan, M *et al.* (2005). Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**: 685–689.
28. Hester, ME, Foust, KD, Kaspar, RW and Kaspar, BK (2009). AAV as a gene transfer vector for the treatment of neurological disorders: novel treatment thoughts for ALS. *Curr Gene Ther* **9**: 428–433.
29. Bensimon, G, Lacomblez, L and Meininger, V (1994). A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. *N Engl J Med* **330**: 585–591.
30. Kordower, JH, Palfi, S, Chen, EY, Ma, SY, Sendera, T, Cochran, EJ *et al.* (1999). Clinicopathological findings following intraventricular glial-derived neurotrophic factor treatment in a patient with Parkinson's disease. *Ann Neurol* **46**: 419–424.
31. Su, X, Kells, AP, Huang, EJ, Lee, HS, Hadaczek, P, Beyer, J *et al.* (2009). Safety evaluation of AAV2-GDNF gene transfer into the dopaminergic nigrostriatal pathway in aged and parkinsonian rhesus monkeys. *Hum Gene Ther* **20**: 1627–1640.
32. Griffiths-Jones, S (2006). miRBase: the microRNA sequence database. *Methods Mol Biol* **342**: 129–138.
33. Gao, G, Alvira, MR, Somanathan, S, Lu, Y, Vandenberghe, LH, Rux, JJ *et al.* (2003). Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc Natl Acad Sci USA* **100**: 6081–6086.
34. Xu, ZF, Qi, WQ, Ouyang, XZ, Yeung, E and Chye, ML (2001). A proteinase inhibitor II of Solanum americanum is expressed in phloem. *Plant Mol Biol* **47**: 727–738.
35. Guo, HS, Xie, Q, Fei, JF and Chua, NH (2005). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *Plant Cell* **17**: 1376–1386.