



# RNA

A PUBLICATION OF THE RNA SOCIETY

## Production of artificial piRNAs in flies and mice

Felix Muerdter, Ivan Olovnikov, Antoine Molaro, et al.

*RNA* 2012 18: 42-52 originally published online November 17, 2011

Access the most recent version at doi:[10.1261/rna.029769.111](https://doi.org/10.1261/rna.029769.111)

---

**Supplemental  
Material**

<http://rnajournal.cshlp.org/content/suppl/2011/11/01/rna.029769.111.DC1.html>

**References**

This article cites 59 articles, 20 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/18/1/42.full.html#ref-list-1>

Article cited in:

<http://rnajournal.cshlp.org/content/18/1/42.full.html#related-urls>

**Email alerting  
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

---

---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>

---

# Production of artificial piRNAs in flies and mice

FELIX MUERDTER,<sup>1,2,5</sup> IVAN OLOVNIKOV,<sup>3,4,5</sup> ANTOINE MOLARO,<sup>1,5</sup> NIKOLAY V. ROZHKOVA,<sup>1,5</sup> BENJAMIN CZECH,<sup>1,2</sup> ASSAF GORDON,<sup>1</sup> GREGORY J. HANNON,<sup>1,6</sup> and ALEXEI A. ARAVIN<sup>3,6</sup>

<sup>1</sup>Howard Hughes Medical Institute, Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

<sup>2</sup>Zentrum für Molekularbiologie der Pflanzen, Entwicklungsgenetik, University of Tübingen, 72076 Tübingen, Germany

<sup>3</sup>California Institute of Technology, Division of Biology, Pasadena, California 91125, USA

<sup>4</sup>Institute of Molecular Genetics, Russian Academy of Sciences, 123182 Moscow, Russia

## ABSTRACT

In animals a discrete class of small RNAs, the piwi-interacting RNAs (piRNAs), guard germ cell genomes against the activity of mobile genetic elements. piRNAs are generated, via an unknown mechanism, from apparently single-stranded precursors that arise from discrete genomic loci, termed piRNA clusters. Presently, little is known about the signals that distinguish a locus as a source of piRNAs. It is also unknown how individual piRNAs are selected from long precursor transcripts. To address these questions, we inserted new artificial sequence information into piRNA clusters and introduced these marked clusters as transgenes into heterologous genomic positions in mice and flies. Profiling of piRNA from transgenic animals demonstrated that artificial sequences were incorporated into the piRNA repertoire. Transgenic piRNA clusters are functional in non-native genomic contexts in both mice and flies, indicating that the signals that define piRNA generative loci must lie within the clusters themselves rather than being implicit in their genomic position. Comparison of transgenic animals that carry insertions of the same artificial sequence into different ectopic piRNA-generating loci showed that both local and long-range sequence environments inform the generation of individual piRNAs from precursor transcripts.

**Keywords:** piwi; noncoding RNA; piRNA

## INTRODUCTION

In several animals, including *Drosophila* and mammals, piRNAs have been shown to form the core of a small RNA-based innate immune system that recognizes and represses mobile elements (Saito et al. 2006; Vagin et al. 2006; Aravin et al. 2007a; Brennecke et al. 2007; Gunawardane et al. 2007; Malone and Hannon 2009; Siomi et al. 2011). This function is essential for proper germ-line development, and mutations in the piRNA pathway lead to male and/or female sterility (Cox et al. 2000; Harris and Macdonald 2001; Li et al. 2009; Malone and Hannon 2009). In essence, piRNAs play a major role in defining genomic content as being transposon related; piRNAs comprise a catalog of transposon sequences that an organism has defined as targets for repression (Brennecke et al. 2007). Omission from that catalog can mean that an element escapes repression. In

the case of flies, the lack of an effective piRNA-based definition for the *I-* or *P-element* in some strains means that introduction of even this single transposon can lead to highly penetrant sterility (Pelisson 1981; Rubin et al. 1982; Brennecke et al. 2008).

Sequencing of piRNA populations has revealed their extreme diversity; literally, millions of distinct piRNA sequences can be identified in a single individual (Aravin et al. 2006, 2007b; Girard et al. 2006; Brennecke et al. 2007; Houwing et al. 2007; Lau et al. 2009). Genomic mapping indicates that piRNAs arise from three different types of loci. First, the dominant source of piRNAs can be found in so-called piRNA clusters (Aravin et al. 2006, 2007b; Brennecke et al. 2007). These loci range from a few kilobases to >200 kb in size. They are often strongly enriched in transposon sequences, in accord with a function of the piRNA pathway in transposon control (Vagin et al. 2006; Brennecke et al. 2007; Gunawardane et al. 2007). In the majority of cases, clusters generate a mixture of small RNAs, with some sense and some antisense to each targeted transposon. Second, piRNAs can be derived from protein-coding genes, with these almost invariably being sense species from 3' UTRs (Aravin et al. 2008; Robine et al.

<sup>5</sup>These authors contributed equally to this work.

<sup>6</sup>Corresponding authors.

E-mail [aaa@caltech.edu](mailto:aaa@caltech.edu).

E-mail [hannon@cshl.edu](mailto:hannon@cshl.edu).

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.029769.111>.

2009; Saito et al. 2009). It is as yet unclear whether a single transcript isoform can be either translated into protein or processed into small RNAs or whether a specific transcript variant serves as a piRNA precursor. Only a few genes give rise to piRNAs, and these do not show uniformly high expression, suggesting that some specific determinant or motif, rather than a high-transcript abundance marks specific genes for processing. Third, piRNAs can arise from dispersed, euchromatic transposon copies (Brennecke et al. 2007, 2008; Aravin et al. 2008). These are often full length and close to consensus, representing the potentially active representatives of each transposon family.

The three types of piRNA generative loci produce small RNAs through two different mechanisms. piRNA clusters and genic loci generate “primary” piRNAs, which appear to be sampled from long, single-stranded transcripts through the action of an unknown nucleolytic machinery (Aravin et al. 2006, 2007b; Brennecke et al. 2007; Malone et al. 2009). Abundant primary piRNAs share no apparent sequence or structural motifs except for the presence of a 5' terminal U residue (1U), which may reflect a binding preference of some Piwi family proteins. Secondary piRNAs are produced through a slicer-dependent mechanism, termed the ping-pong cycle and have a characteristic bias for an A at position 10 (paired with the 1U in the primary piRNA) (Brennecke et al. 2007; Gunawardane et al. 2007).

Combined analysis of piRNA sequences and animals bearing mutations in piRNA pathway components has led to a model for the role of these small RNAs (Malone and Hannon 2009; Saito and Siomi 2010; Senti and Brennecke 2010; Siomi et al. 2011). piRNA clusters produce a multitude of individual piRNAs, and the sequence content of piRNA cluster defines sequences of mature piRNAs generated from it. With the notable exception of pachytene piRNAs that are expressed during male meiosis in mouse, piRNA clusters in both flies and mice are highly enriched in transposable element sequences. The sequence content of the piRNA clusters determines the capacity of the system to respond to a given element, in essence comprising an organisms' evolving molecular definition of transposons. Inherent in this scenario is the ability of the system to adapt to colonization by new elements by incorporating their sequence into a piRNA cluster. A clear example can be found in the *P-element*, which swept through global *Drosophila melanogaster* populations after the sequestration of common laboratory strains (Rubin et al. 1982). Laboratory strains have no ability to repress the *P-element*. In retrospect, studies of strains with natural or acquired *P-element* resistance suggested that integration of the element into a piRNA cluster was key to its control (Ronsseray et al. 1991, 1996, 2003).

Here, we sought to test whether the ability to translate new genomic content into small RNAs was a general characteristic of piRNA loci in flies and mice. We find that clusters can be programmed to produce artificial piRNAs

(apiRNAs). Furthermore, we were able to separate functional piRNA clusters from their native genomic locations, indicating that the clusters themselves contain sufficient information to funnel their RNA products into the piRNA biogenesis pathway. We made use of marked transgenic clusters that carry insertions of the same sequence into different contexts to evaluate the features that lead to the production of individual piRNA species. We find that critical determinants lie both in the local and long-range sequence environments of the piRNA cluster.

## RESULTS AND DISCUSSION

The current model for acquiring piRNA-dependent resistance against new transposon invasion implies that insertion of active transposons into an existing piRNA cluster leads to the generation of new piRNA species and enables element repression. This model suggests that any sequence, if inserted into a piRNA cluster, will lead to the generation of new piRNAs. Though attractive, this model has not been rigorously tested. Acquisition of natural resistance against transposable elements by transposition into piRNA clusters is difficult to study in an experimental setting. However, this scenario can be modeled using transgenes carrying new sequence information within a piRNA-generating locus.

Over the years, large collections of *Drosophila* stocks have been produced that carry transgenes integrated randomly throughout the genome. We took advantage of these tools by searching for integration events in native piRNA clusters. The line P{IArB}A171.1F1 (also known as P-1152) has a 18.3-kb construct P{IArB} integrated into a telomeric piRNA cluster on the X-chromosome (chromosomal location 1A) (Wilson et al. 1989; Roche and Rio 1998). The P{IArB} transgene contains sequences derived from the *hsp70*, *Adh*, and *rosy* genes of *D. melanogaster* and a bacterial *lacZ* gene. Unlike P{IArB} insertions in other genomic sites, P-1152 has unusual properties. It is able to suppress the expression of other *lacZ* transgenes in germ cells, a phenomenon termed *trans*-silencing (Fig. 1A; Supplemental Fig. S1; Ronsseray et al. 1991). The P{IArB} insertion in P-1152 is mapped to the Telomere Associated Sequence (TAS) repeats that produce abundant piRNAs from both genomic strands. These piRNAs are loaded into Piwi, Aub, and Ago3 in the germ cells of *D. melanogaster* ovaries (Brennecke et al. 2007). Aub and Ago3-loaded piRNAs derived from TAS repeats display the characteristic features of the ping-pong amplification cycle, including a prevalent 10-nt 5' overlap of sense and antisense species and an enrichment for an A at position 10 of secondary piRNAs. The *trans*-silencing properties of P-1152 transgene and the association of these properties with its localization in the piRNA cluster suggested that insertion of *lacZ* into an existing piRNA cluster led to the generation of new anti-*lacZ* piRNAs that are able to suppress cognate transcripts in germ cells. Indeed, the presence of small

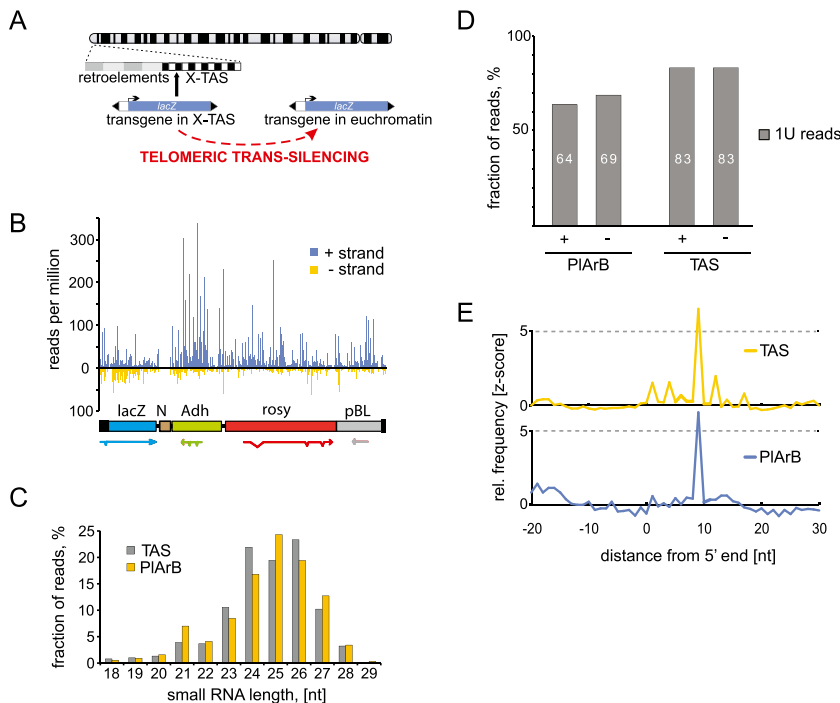
RNAs complementary to *lacZ* was recently demonstrated using RNase-protection assay in ovaries of the P-1152 line (Todeschini et al. 2010).

To analyze more deeply any artificial piRNAs derived from the P{IArB} transgene, we sequenced small RNAs from ovaries of the P-1152 line, examining a size range from 18 to 29 nt. This includes piRNAs, siRNAs, and miRNAs. P{IArB} generated abundant small RNA species that mapped to both genomic strands (Fig. 1B). Their size profile indicated that the majority were likely piRNAs, ranging from 23 to 27 nt, while a minor fraction corresponded to 21-nt endo-siRNAs that are also a product of bidirectionally transcribed piRNA loci (Fig. 1C; Czech et al. 2008; Lau et al. 2009). Further analysis confirmed that the 23- to 27-nt RNAs were genuine piRNAs that could be separated into primary (1U-biased) and secondary (10A-biased) populations (Fig. 1D; data not shown). Transgene piRNAs mapping to opposite genomic strands tended to have a 10-nt overlap between their 5' ends that is a characteristic feature of the ping-pong cycle (Fig. 1E). Notably, P{IArB} contains the only *lacZ* sequence information in the P-1152 strain. Since signatures of the ping-pong

cycle were evident for *lacZ*-derived piRNAs, this demonstrates unequivocally that cluster transcripts derived from the plus and minus genomic strands can participate in the piRNA amplification loop. Native *Adh* and *rosy* transcripts are not processed into piRNAs in wild-type flies (data not shown). Therefore, it is unlikely that any specific signals that trigger piRNA processing might be present in these genes. Moreover, bacterial sequences are unlikely to have evolved as a trigger for piRNA production. Thus, our results indicate that, when present in the context of a piRNA cluster, virtually any sequence can serve as a substrate for piRNA biogenesis. We confirmed previous observations that the P{IArB} transgene inserted in TAS is able to silence *lacZ* expression from separate, euchromatic locations (Supplemental Fig. S1), demonstrating that artificial anti-*lacZ* piRNAs are functional and able to silence transcripts that share sequence content in *trans*.

piRNAs are processed from the entire P{IArB} transgene independently of the origin of the inserted fragments; both *D. melanogaster* and bacterial sequences generate piRNAs with similar efficiency (Fig. 1B). Throughout the construct there are approximately twofold more piRNAs derived from the plus than from the minus genomic strand independently of the orientation of the genes within the construct, just as is observed for native components of the cluster. For example, *Adh* and *rosy* have different orientations, but for both fragments the majority of piRNAs are mapped to the plus genomic strand. RT-PCR shows that *rosy* transcripts are present in ovaries of P-1152 females, but absent in wild-type flies or flies that have a P{IArB} insertion outside of the piRNA cluster (Supplemental Fig. S2), indicating that *rosy* expression is dependent on insertion of P{IArB} into TAS. Overall, both the distribution of piRNAs along P{IArB} transgene and RT-PCR results suggest that transcript of both plus- and minus-strand RNAs, which are processed to piRNAs, initiates outside of the transgenic construct, likely within adjacent TAS sequences.

Mapping of piRNAs to P{IArB} revealed that intronic sequences present within *Adh* and *rosy* gave rise to piRNA from both genomic strands. Even when present in the sense orientation, where the intron could have been removed by the splicing apparatus, piRNA levels remained comparable in adjacent intronic and exonic regions. The generation of piRNA from intronic sequence is unexpected, as primary piRNA bio-



**FIGURE 1.** Production of artificial piRNAs (apiRNAs) from the *Drosophila* X-TAS cluster. (A) The P{IArB} insertion into the X-TAS cluster is shown schematically along with an illustration of *trans*-silencing. (B) Below is a schematic of the P{IArB} insert with the inferred structures of the transcripts it can produce (see text). N is an area where the sequence is unknown. Above is a plot of piRNA read frequencies along the plus and minus strands (indicated) of the element. (C) Small RNA lengths are plotted as a fraction of reads for TAS and for the inserted element. (D) Fractions of reads beginning with a 5' U are plotted for the P{IArB} and TAS plus and minus strands. (E) The degree of 5' overlap for piRNAs from the plus and minus strands for P{IArB} and TAS were quantified and plotted as relative frequencies (Z-scores). The spike at position 9 is a signature of the ping-pong amplification cycle.

genesis is thought to occur in the cytoplasm and has been linked to specific cytoplasmic bodies, e.g., nuage and Yb bodies, which concentrate components such as zucchini and armitage, which are implicated in piRNA processing (Tomari et al. 2004; Lim and Kai 2007; Pane et al. 2007; Malone et al. 2009; Haase et al. 2010; Olivieri et al. 2010; Saito et al. 2010; Qi et al. 2011). Furthermore, genic piRNAs that are processed from mRNA of protein-coding genes in *Drosophila* and mice are mapped almost exclusively to exonic sequences (Aravin et al. 2008; Robine et al. 2009; Gan et al. 2011). To reconcile these disparities, we searched explicitly for piRNAs that crossed predicted exon-exon junctions, since these must arise from spliced mRNAs. We did detect a few such small RNAs for *rosy* and *Adh*, coming only from the genomic strand with the intron in the appropriate orientation for splicing to occur. Considered together, these data suggest a model in which piRNA biogenesis normally occurs following intron removal, but that recognition of some RNA processing signals might be suppressed when they are present within a piRNA cluster. In this regard, strand-specific RT-PCR indicated that more than half of sense-oriented *rosy* transcripts are not spliced in P-1152 ovaries (Supplemental Fig. S2). Suppression of conventional RNA processing signals within piRNA clusters would make sense in many ways, since the insertion of a new element would often bring at least a polyadenylation signal, which under normal circumstances could negate the production of piRNAs downstream from that site by terminating transcription or preventing the export of piRNA precursors.

Generation of artificial piRNAs by insertion of a new sequence into a piRNA cluster provides a molecular tag that allows the monitoring of cluster function even if the native, nontagged cluster is present in the same genome. We exploited this fact to test whether the presence of piRNA clusters at precise genomic positions was important to their function.

In flies, piRNA clusters occur mainly at the boundaries between heterochromatin and euchromatin, particularly in pericentromeric regions (Brennecke et al. 2007). In mammals, piRNA clusters that are expressed in meiotic cells occur in strictly syntenic positions, even though the sequence content of these loci is not conserved (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006). These observations have strongly suggested that the genomic context of piRNA clusters might be key to their function. Precedent can be drawn from plants and fission yeast, where small RNAs are generated from loci whose function relies upon the presence of normally repressive chromatin marks (Huisinga and Elgin 2009; Lahmy et al. 2010). In turn, the repressive chromatin marks themselves are maintained by small RNA-directed complexes, closing the cycle. To determine whether specific chromatin environments, which are a property of the genomic context of piRNA clusters, are essential for piRNA production, we created

ectopic insertions of tagged piRNA clusters in non-native sites.

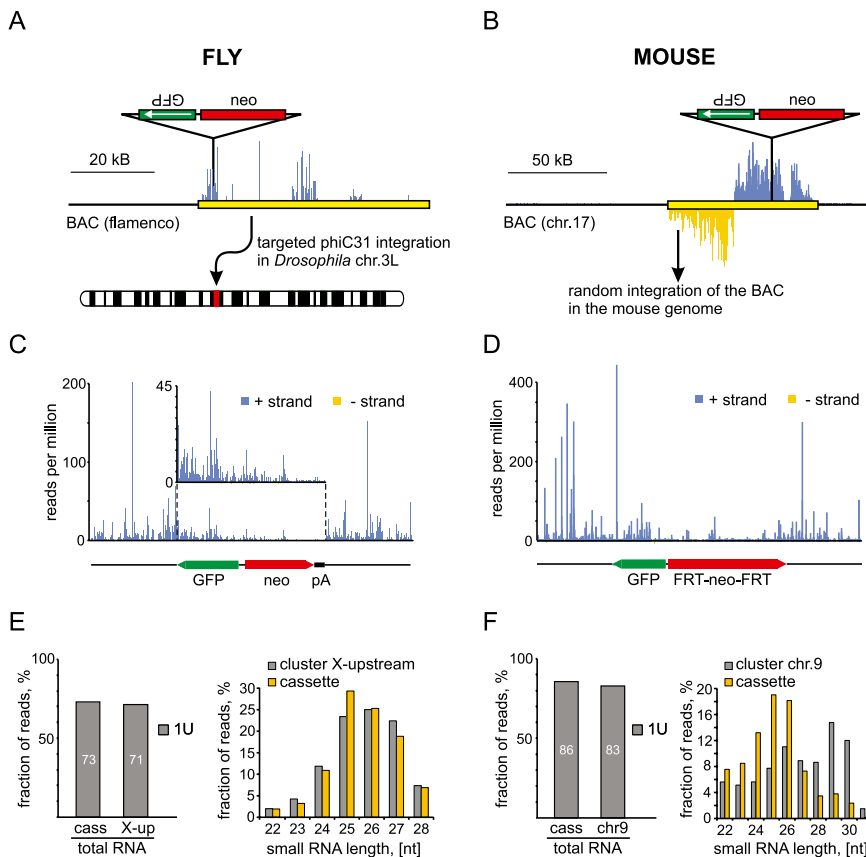
As one test of the aforementioned hypothesis, we examined the position dependence of the *flamenco* cluster in *Drosophila* (Fig. 2A). *Flamenco* is present at the boundary between euchromatin and pericentromeric heterochromatin on the *Drosophila* X chromosome, and its position proximal to the *DIP1* gene is conserved through at least 12 M years of *Drosophila* evolution (Sarot et al. 2004; Brennecke et al. 2007; Malone et al. 2009). It produces piRNAs from only one genomic strand and is exclusively expressed in the somatic follicle cells of the ovary. We selected a P[acman] BAC clone that extended from a position ~30 kb upstream of the first annotated piRNA ~86 kb toward the X chromosome centromere (Venken et al. 2009). This encompassed ~30% of the *flamenco* cluster. To distinguish any ectopic copies of *flamenco* from the native locus, we marked the BAC by recombineering, inserting a cassette comprising a nonfunctional GFP sequence and a bacterial neomycin resistance gene (Copeland et al. 2001; Venken et al. 2006; Sharan et al. 2009). Marker sequences were inserted ~4 kb downstream from the first annotated piRNA in a site, which we had previously shown to produce abundant small RNAs.

For mice, we chose to modify a piRNA cluster on mouse chromosome 17 that is a major contributor to piRNA populations in developing male germ cells from the pachytene stage through the end of meiosis (Fig. 2B; Aravin et al. 2006; Girard et al. 2006). This cluster occurs in syntenic locations in rat and in human, indicating conservation through at least 80 M years of evolution. Like *flamenco*, each region of the ch17 cluster produces piRNAs from only one genomic strand. A mouse BAC clone comprising ~187 kb of chromosome 17 carried the complete ch17 cluster and extended 60 kb upstream of and 30 kb downstream from the locus. It was similarly marked by recombineering to insert a modified GFP/neo cassette.

In flies, we took advantage of a phiC-31 attachment site in the P[acman]-BAC to insert the modified *flamenco* cluster into a known genomic locus (Venken et al. 2006, 2009). Given that *flamenco* is normally present in a location annotated as heterochromatic (X chromosome, band 20A), we chose a gene-rich, euchromatic site to insert the transgene. Specifically, we created lines with one additional copy of *flamenco* on chromosome 3L at band 62E1 (landing pad 31) (Venken et al. 2006). For mice, we used standard pronuclear injection to create two independent founder lines (R13 and R37) with ch17 transgene insertions in presumably distinct random locations.

Small RNA cloning and Illumina sequencing revealed that abundant piRNAs derived from GFP were produced from ectopic clusters in both flies and mice (Fig. 2C,D). Like the native loci, these produced small RNAs from only one genomic strand. Unlike X-TAS, neither *flamenco* nor the ch17 cluster normally participate in the ping-pong





**FIGURE 2.** Generation of apiRNAs from ectopic clusters in flies and mice. (A) A schematic representation of the GFP/Neo cassette is shown along a diagram of the *flamenco* locus (in yellow, piRNA densities in blue) in the BAC used for transgenesis. Below is a schematic indicating that the transgene is inserted into chromosome 3L. (B) The GFP/neo insertion into the mouse chromosome 17 cluster is diagrammed as in A. (C) The structure of the *flamenco* GFP/Neo insertion is diagrammed below a plot of piRNA frequencies along the insert on the plus and minus strands (indicated). For reference, piRNAs are also mapped to flanking regions, though these represent a mixture of RNAs derived from the two native and one ectopic *flamenco* clusters. (D) A scheme of the GFP/Neo insert into the mouse chromosome 17 cluster is shown below piRNAs mapping to the insert and its context as in C. Again, piRNAs that flank the insert can be derived from the two native or inserted ectopic loci. (E) The 1U bias (left) and size distributions (right) of apiRNAs from the ectopic *flamenco* cluster are compared with another piRNA cluster (X-upstream) that also produces piRNAs from one genomic strand in follicle cells. (F) As in E, apiRNAs from the ectopic ch17 cluster in mice are compared with a similarly structured cluster on chromosome 9.

amplification loop, and the ectopic insertions also lacked signatures of the cycle, namely, small RNAs with a 10A bias and sense/antisense pairs that overlap by 10 nt. Small RNAs from the ectopic clusters did show the strong 1U bias that is a signature of primary piRNA populations (Fig. 2E,F; Supplemental Fig. S3A).

It seemed likely that the transgenic clusters would generate piRNAs both from the inserted marker gene and from sequences that represent their native content; however, it is impossible to distinguish the latter from piRNAs derived from endogenous loci. The ectopic cluster is present as a single copy in the genome, as compared with two endogenous copies. We might therefore expect piRNA levels coming from shared regions to increase by 1.5-fold if all

copies were equally active. Indeed, we noted a 1.3-fold increase in piRNAs, which are derived from the portion of the *flamenco* cluster present in transgene. Similarly, the levels of MILI and MIWI piRNAs derived from the chr17 cluster in mouse increased by between 1.2- and 1.5-fold relative to a nonmodified cluster on ch9 in two independent transgenic lines. The profiles of piRNA mapped to the *flamenco* and ch17 clusters are very similar in wild-type and transgenic flies and mice (Supplemental Fig. S4). Therefore, the heterologous insertion of a marker gene does not appear to exert a strong influence on the processing of piRNAs from transgenic loci. Overall, our data indicate that transgenic piRNA clusters have similar activity to their endogenous counterparts, despite being present at non-native genomic positions.

*Flamenco*-derived piRNAs associate exclusively with Piwi, the only family member that is expressed in follicle cells (Sarot et al. 2004; Brennecke et al. 2007). Thus, they have a characteristic size profile, peaking at around 25 nt. piRNAs from the ectopic *flamenco* insertion shared this size distribution (Fig. 2E). piRNAs from the ch17 cluster (and other murine clusters expressed during meiosis) normally associate with both MILI and MIWI (Supplemental Fig. S3B). These complexes have distinct small RNA size profiles, with MILI to associate with a ~26-nt and MIWI harboring a ~30-nt species (Fig. 2F; Aravin et al. 2006; Girard et al. 2006). Overall, MIWI-bound species are substantially more abundant than MILI bound species (Aravin et al. 2006; Girard

et al. 2006). While the ectopic ch17 cluster produced small RNAs with sizes characteristic of MILI and MIWI complexes, their ratio was very different than expected based upon the behavior of the native cluster (Fig. 2F; Supplemental Fig. S3B). RNAs with the size of MILI partners greatly outnumbered those with the size of MIWI-bound species. Thus, the ectopic cluster appeared to have a strong preference for one of its two potential Piwi-family partners (Fig. 2F; Supplemental Fig. S5).

Overall, our data indicate that piRNA clusters can function even when divorced from their normal genomic locale. With *flamenco*, the ectopic insertion behaved indistinguishably from the native locus, even though it had been substantially truncated on the centromere-proximal

side. For the ch17 cluster, piRNAs were still produced in abundance from the ectopic insertions, but the behavior of the small RNAs shifted toward preferential MILI association. This could indicate that some element of chromosomal context was important for signaling an ultimate association with MIWI or perhaps that critical signals that mark the cluster as a source of MIWI piRNAs were missing from our BAC clone, despite its extending well beyond the two ends of the cluster. Our results by no means rule out chromatin structure as a contributory element in defining piRNA clusters. However, if specific chromatin structures are important, the signals for their formation must be tightly linked to the piRNA loci themselves.

The inclusion of the same artificial sequence in piRNA clusters in multiple locations and in distinct organisms afforded the opportunity to probe the determinants of piRNA selection. In contrast to miRNAs and siRNAs, whose processing from longer precursors is informed by their specific secondary structure and is well understood, no rules that explain the selection of individual piRNAs have been defined. The only bioinformatic study that addressed this question came to the conclusion that the processing of individual piRNA from precursors is quasi-random, with only weak influences of local sequence (positions  $-1$  to  $+4$  relative to the 5' end of the piRNA) (Betel et al. 2007). However, sequencing efforts from our and other groups showed that individual piRNAs are not produced uniformly along clusters. Instead, certain small RNAs appear substantially more abundant (Aravin et al. 2006, 2008; Girard et al. 2006; Brennecke et al. 2007). Characteristics underlying these inequalities could be intrinsic to the local sequence environment of each individual piRNA or could be conferred by long-distance interactions and formation of secondary structures within the precursor molecule. Alternatively, patterns could be essentially random, with the abundance of each species being determined stochastically.

As with native piRNAs, read distributions along the marker cassettes in the ectopic clusters were very uneven (Fig. 3A; Supplemental Fig. S5). Focusing on the GFP coding sequence, 1% of nucleotide residues contribute 19% of all 5' ends of GFP-mapping piRNA reads in flies, while 10% of positions account for 70% of reads (Fig. 3B). In mouse, the distribution was even more skewed with 1% of GFP residues contributing 42% of piRNA reads (Fig. 3B). To probe the causes leading to these skewed distributions, we compared GFP-derived piRNAs in the two independent mouse transgenic lines. The correlation in the abundance of individual small RNAs was remarkable ( $R^2 = 0.99$ ) (Fig. 3C), ruling out the notion that the patterns that we observe are random within each sample. Procedures for preparing small RNA libraries include steps with well-established sequence-based biases, namely, RNA adapter ligations and PCRs (Linsen et al. 2009). We therefore considered the possibility that those biases dominated apparent sequence

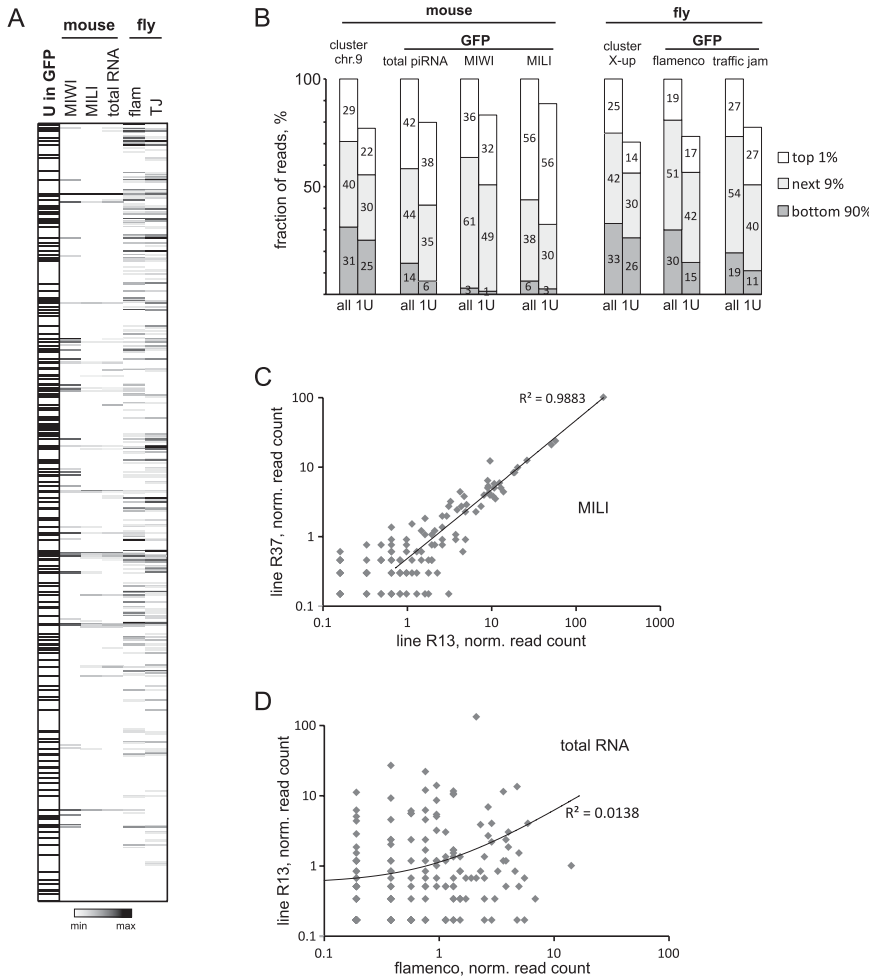
preferences in apiRNA generation. However, very little correlation was seen between GFP piRNAs in flies and mice ( $R^2 = 0.01$ ) (Fig. 3D), contrary to what one would expect if the patterns that we observed were strongly influenced by the biases of library preparation methods.

Considered as a whole, our data strongly support the existence of signals that determine the efficiency of production of individual piRNAs and raise several possibilities as to the nature of those signals. First, the biased distribution of piRNAs could be an exclusive consequence of their context within the cluster. This would imply that large-scale features, such as the structure of the transcript or preferential entry sites for the primary processing machinery determine differential piRNA production, akin to the generation of phased siRNAs from long dsRNAs in plants and animals (Zamore et al. 2000; Howell et al. 2007). Alternatively, determinants of efficient piRNA biogenesis could still be defined by the local sequence environment of each individual piRNA, with sequence determinants being interpreted differently in our two experimental models. To begin to discriminate between these possibilities, it was necessary to insert the same sequence (GFP) into different piRNA precursors that are expressed and processed in the same cell type.

The *traffic jam* (*tj*) gene encodes a basic leucine zipper transcription factor and is expressed in the follicle cells of the *Drosophila* ovary, just as is *flamenco* (Li et al. 2003; Saito et al. 2009). Importantly, *tj* generates piRNAs from a discrete segment of its 3'-UTR region (Saito et al. 2009). We created a marked, ectopic copy of *tj* by inserting a GFP coding sequence in the antisense orientation into its piRNA-producing domain and integrated this into a euchromatic site on chromosome 3L (Fig. 4A).

Sequencing of small RNAs (Fig. 4B) yielded abundant piRNAs from the inserted GFP sequence. These had the same characteristics as native *tj*-derived piRNAs, including being produced from the sense strand of the locus, having a size distribution characteristic of Piwi-associated species, and a strong bias for a 5' terminal U residue (Fig. 4C,D). Position-dependent differences in piRNA abundance were also apparent, with the most abundant 10% of possible GFP piRNAs contributing 81% of all GFP-mapping reads (Fig. 3B).

To discriminate local- from long-distance sequence effects, we compared the abundance of individual piRNAs from the *tj* and *flamenco* transgenes. As compared with the patterns derived from independent insertions of the same transgenes in mice ( $R^2 = 0.99$ ) (Fig. 3C), patterns of GFP piRNAs from *tj* and *flamenco* appeared quite different ( $R^2 = 0.24$ ) (Fig. 4D). However, they were much more similar than patterns produced in mouse versus fly ( $R^2 = 0.01$ ) (Fig. 3D). At the extremes, uridine positions in GFP that generate abundant piRNAs from the *flamenco* transgene tended also to generate abundant piRNAs from *tj* (Fig. 4E). Conversely, those that did not generate piRNAs from *flamenco* did not generate piRNAs from *tj*.



**FIGURE 3.** apiRNA production is not uniform along inserted sequences. (A) A heatmap of piRNA abundance is displayed for all positions in the GFP insert carried in ectopic piRNA clusters as indicated. Sequence measurements were from total RNAs except in mouse, where MIWI and MILI immunoprecipitates (indicated) were also analyzed. The first column simply indicates U positions relative to the heatmaps. (B) All possible positions for piRNA production from GFP sequences inserted into ectopic clusters (all sites or only U positions, indicated) were ranked by their contribution to actual piRNA populations. The fraction of piRNAs contributed by the top 1%, the next 9%, or the remaining 90% were measured and indicated. Native clusters (indicated) were similarly analyzed for reference. (C) MILI-bound piRNAs were quantified by sequencing from two independent lines carrying the ectopic ch17 cluster. Correlations between read counts for GFP-derived piRNAs are shown. Libraries were normalized as described in the Materials and Methods. (D) A similar analysis was performed for GFP-derived piRNAs in total reads, comparing the R13 mouse line carrying the ectopic ch17 cluster and the fly strain carrying the ectopic *flamenco* cluster.

Considered together, these results indicate effects of both local and long-range sequence environment on piRNA biogenesis. Small RNAs generated from GFP embedded in different piRNA precursor transcripts in the same species were more similar than expected by chance. Influences of sequence, however, seem species or cell-type specific, since these same biases did not extend from fly to mouse. Strong effects also appear to be exerted by the context within the cluster, given the near identity in GFP piRNA populations in independent mouse lines and their dissimilarity in comparisons of marked *flamenco* and *tj* transcripts. The

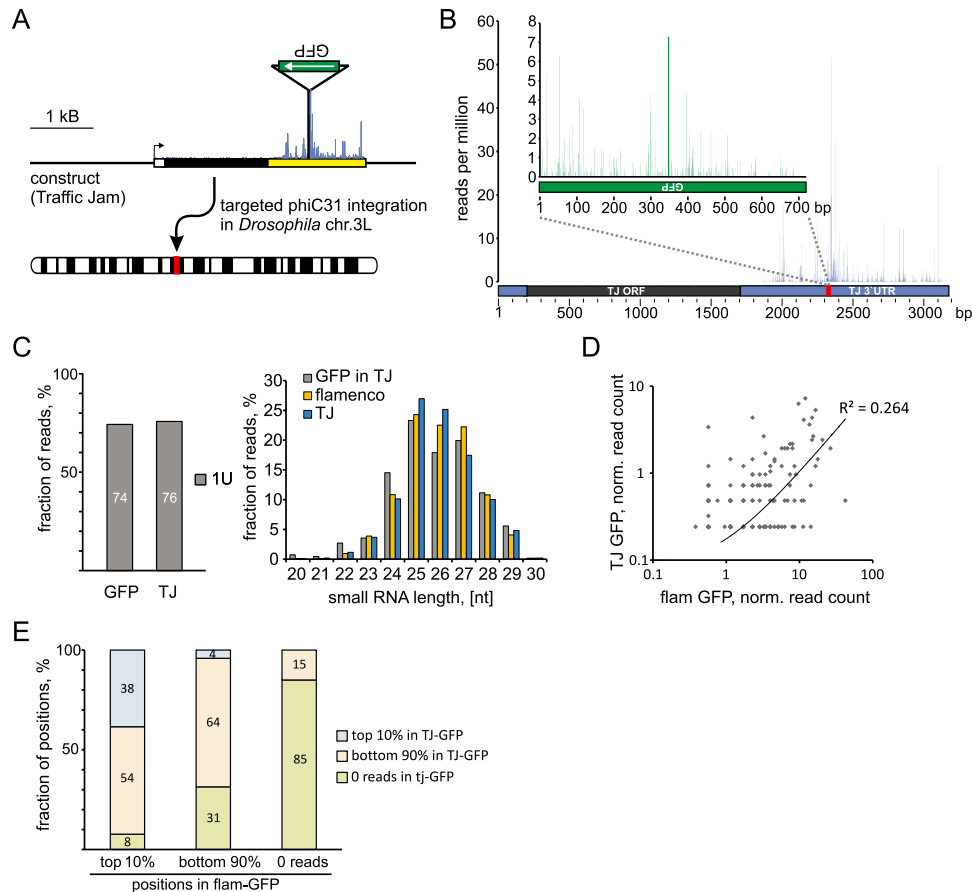
precise nature of such context-dependent effects is unclear, but could depend upon the overall secondary or tertiary structures of piRNA precursors.

Our data are consistent with the model in which new insertions of transposable elements become incorporated into the piRNA repertoire as a mechanism of acquiring resistance. Indeed, our data indicate that any sequence will probably produce piRNAs immediately upon its incorporation into a functional piRNA cluster. Furthermore, our data demonstrate that the position of the cluster in the genome is not important, and, therefore, transgenic piRNA clusters can be created in heterologous genomic locations.

Previous bioinformatic analyses described the generation of individual piRNAs from long precursor molecules as a pseudo-random process with a weak influence of the local sequence environment of individual piRNA species (Betel et al. 2007). However, the distribution of individual piRNAs within the precursor is far from being random; different Us have drastically different propensities to generate piRNAs, and some non-U positions produce substantially more piRNAs than nonprocessed U positions. Here, we showed that patterns of individual piRNAs within the precursors are highly reproducible if the sequence is present within the same context. Patterns become less reproducible if the local sequence is embedded in a different context, indicating that both local and long-range sequence environments impact processing efficiency. This result explains a failure in the identification of simple rules that would explain the production of abundant piRNAs from a given precursor molecule.

The general approach we describe here, using marked ectopic piRNA clusters to produce apiRNA species, provides a path toward further dissection of elements that discriminate piRNA clusters and marks corresponding transcripts for piRNA processing. The ability to program the piRNA pathway to produce artificial piRNAs has implications for harnessing this system for controlling gene expression. In particular, in mammals this approach may present advantages over harnessing the miRNA pathway, since piRNAs can induce epigenetic silencing of loci through the recruitment, directly or indirectly, of the de





**FIGURE 4.** piRNA production from the 3' UTR of *traffic jam*. (A) A schematic of the GFP insertion into the 3' UTR of the *traffic jam* gene indicates the transcriptional start site (arrow), the coding sequence (black box), and the 3' UTR (yellow box). Below, a diagram indicates site-specific insertion into 3L. (B) piRNA read counts are plotted along the inserted GFP sequence (green *inset*) and the surrounding areas of the *tj* 3' UTR. Note that sequences mapping outside of GFP could be produced from the ectopic insert or the two endogenous copies of *tj*. (C) The 1U bias (left) and the size distribution of piRNAs mapping to the GFP insert are shown with reference to piRNAs from the *flamenco* cluster. (D) Normalized piRNA read counts (see Materials and Methods) were compared for the GFP insertions into the ectopic *flamenco* or *tj* piRNA clusters. (E) Read counts are calculated for all possible piRNAs that start with uridine derived from the GFP insertion into *flamenco*. These were divided into the top 10%, the next 90%, and the subset that contributed no reads. For each subset, the number that were present in the top 10%, the next 90%, or the noncontributors for the GFP insertion into *tj* were plotted.

novo DNA methylation machinery (Carmell et al. 2007; Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008; Siomi et al. 2011).

## MATERIALS AND METHODS

### *D. melanogaster* strains and crosses

The line, P-1152, which carries an insertion of the P{IArB} construct in telomeric sequences of X chromosome (site 1A) is described in Roche and Rio (1998). To test *trans*-silencing with P-1152, females of this line were crossed with males that have *lacZ* expressed from a euchromatic location on chromosome 2L (line BC69, site 35B10–35C1) (Lemaitre et al. 1993).

### Cloning and recombineering—*D. melanogaster*

The *flamenco* transgene was created using P[acman] clone CH321-35A24, which contains an interval from chromosome X that includes ~20 kb of upstream sequence and the 5' portion of the

*flamenco* piRNA cluster (Venken et al. 2009). An antisense EGFP sequence was introduced into the BAC by recombineering as described in Sharan et al. (2009). The GFP-Neo insertion cassette was built by overlapping PCR based on a FRT-PGK-gb2-neo-FRT cassette (Gene Bridges). The position of the insertion within the *flamenco* cluster was selected based on uniqueness and high frequency of piRNA production from the surrounding region. The cassette was introduced into the the BAC using a pSim6 plasmid described in Datta et al. (2006). To promote recombination, *Escherichia coli* containing pSim6 were transferred to a 2-mL Eppendorf tube and induced at 42°C in an Eppendorf tabletop shaker. The linear DNA substrate was introduced by electroporation using the Gene Pulser XCell. Using exponential decay as a pulse-type, the cells were electroporated at 3000 V, 25  $\mu$ F, and 200  $\Omega$  for 5 msec. After outgrowth and selection of cells, recombinant clones were screened for by PCR, sequencing and restriction digestion, followed by pulse-field gel electrophoresis.

The *D. melanogaster traffic jam* gene with 2 kb upstream and 0.5 kb downstream genomic regions was amplified from the CH322-145O22 P[acman] clone and inserted between the BspHI

and ClaI sites of the pIZ-V5-His vector (Invitrogen). A sequence ATTATTCTGATTGCGACAATAAATTCCGAT in the *TJ* 3' UTR was substituted with the sequence CTTAAGCTGATTGCGACA TAAATACCGGT by overlap PCR to introduce unique AflII and AgeI sites, which were used to insert the inverted EGFP sequence. The modified *traffic-jam* sequence was transferred into the pCasper5-attB vector (a modified *P-element* pCaSpeR5 vector [Le et al. 2007] with a phiC31 attB site to allow site-specific integration).

### Cloning and recombineering—mouse

The transgene containing the modified chr17 piRNA cluster (Chr17: 27427600–27488899) was created using BAC clone RP23-131B16, which contains ~180 kb of genomic sequence that includes the whole chr17 piRNA cluster. We used the FRT-PGK-gb2-neo-FRT cassette (Gene Bridges) and a purified vector containing the EGFP sequence, to construct the GFP-Neo insert for recombineering. After three steps of overlapping PCR (KOD hot start DNA polymerase, Novagen), the recombineering inserts were cloned in a 2.1-TOPO vector (Invitrogen, Version U) according to the manufacturer's protocol. Homology arms for recombineering were added by PCR of purified plasmid.

Recombineering was carried using the Red/ET plasmid-expressing recombination proteins under an arabinose-inducible promoter (Counter-Selection BAC Modification Kit, Gene Bridges, 2007). We followed the manufacturer's protocol, except that recombined clones were selected on Kanamycin and the counter-selection step was skipped. The integrity of modified BAC DNAs were verified by restriction digests and sequencing.

### Transgenic animal production—*D. melanogaster*

Tagged BAC DNA was purified with a Plasmid Maxi Kit (QIAGEN). The DNA was used for PhiC31 integrase-mediated transgenesis, which was carried out by BestGene (<http://www.thebestgene.com/>). *Flamenco* and *tj* transgenes were integrated into attP docking sites on chromosome 3 (VK00031—site 62E1, and VK00033—site 65B2, respectively).

### Transgenic animal production—mouse

BAC DNAs were prepared from overnight *E. coli* cultures using Nucleobond BAC 100 columns (Clontech). DNA was eluted in Injection Buffer (10 mM TRIS, 0.1 mM EDTA, 100 mM NaCl, 1X polyamines) and linearized with PI-SceI enzyme for 4 h. Following linearization, BAC DNA was dialyzed overnight on a 25-mm, 0.025- $\mu$ m filter (Millipore) by floating on Injection Buffer. Transgenic animals were obtained by pro-nuclear injection into B6xSJL F1 hybrids oocytes. Founder animals were crossed to C57BL/6J mice. R37 and R13 transgenic lines were initiated from two independent founder mice.

### Immunoprecipitation of PIWI proteins

Immunoprecipitations from *D. melanogaster* ovaries were carried out according to previously described procedures (Brennecke et al. 2007). For mice, MILI and MIWI were immunoprecipitated from adult testis using antibodies and procedures previously described (Aravin et al. 2007b; Vagin et al. 2009). Briefly, testis were dounced in lysis buffer (10 mM Hepes at pH 7.0, 100 mM

KCl, 5 mM MgCl<sub>2</sub>, 0.5% NP-40, 1% triton X-100, 10% Glycerol, 1 mM DTT, proteinase and RNAase inhibitors). Antibodies (MILI-N2 and MIWI-N2) were then added to the cleared lysates and binding reactions were allowed to proceed overnight at 4°C. Protein A beads are then added to the solution and incubated 3–4 h at 4°C with rotation. After three to four washes in NT-2 buffer (5 mM Tris at pH 7.4, 150 mM NaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 0.05% NP-40, RNAase inhibitors, 1 mM DTT), antibody complexes were proteinase K treated and RNAs ethanol precipitated following phenol/chlorophorm extraction. A fraction of the precipitated RNAs was radiolabeled and size profiles verified on 15% urea polyacrylamide gels.

### Small RNA cloning

Small RNAs from IPs and total RNA extracts were cloned as previously described in Brennecke et al. (2007) and Aravin et al. (2008). Briefly, small RNAs within a 19–33-nt window for mouse samples or a 19–28-nt window for *D. melanogaster* were isolated from 12% polyacrylamide gels. 3' and 5' linkers were ligated, and products were reverse transcribed using Superscript III (Invitrogen). Following PCR amplification, libraries were submitted for sequencing using the Illumina GA2x platform.

### Detection of $\beta$ -galactosidase activity in *D. melanogaster* ovaries

Dissected ovaries from 3–5-d-old flies were fixed in freshly prepared 2% glutaraldehyde in PBS for 20 min, washed twice in PBS, and stained for several hours at 37°C in Fe/NaP buffer (3.1 mM K<sub>3</sub>Fe(CN)<sub>6</sub>; 3.1 mM K<sub>4</sub>Fe(CN)<sub>6</sub>; 10 mM NaH<sub>2</sub>PO<sub>4</sub>xH<sub>2</sub>O; 0.15 M NaCl; 1 mM MgCl<sub>2</sub>) with 0.25% X-Gal. Stained ovaries were mounted in 70% glycerol/PBS.

### Bioinformatic analysis of small RNA libraries

After FASTQ to FASTA conversion, the Illumina dapter (CTGTAGGCACCATCAATTC) was clipped from the 3' end of the read and sequences shorter than 16 nt were discarded from further analysis. The remaining sequences were collapsed into a nonredundant list and mapped to the *D. melanogaster* genome (*D. melanogaster* Apr. 2006 [BDGP R5/dm3]) or the mouse genome (mm9) using the short read aligner bowtie (Langmead et al. 2009). Up to two mismatches were allowed. Sequences that failed to map to the genome were mapped against the artificially introduced sequences. The multiplicity count of mapped sequences was normalized to the total number of reads that mapped to the genome. All further bioinformatic analysis on mapping sequences was done using Unix-based text utilities. Details of those scripts can be obtained upon request. Small RNA sequencing data are deposited at GEO, accession number GSE32435.

### SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

### ACKNOWLEDGMENTS

We thank members of the Hannon and Aravin labs for helpful discussion and comments on the manuscript. We thank members

of the McCombie lab (CSHL) and Igor Antoshechkin (Caltech) for help with RNA sequencing. We thank Andres Canela (CSHL) for technical assistance and Simon Knott (CSHL) and Alex Zahn (Caltech) for help with statistical analysis. Sang Yong Kim (CSHL) created the transgenic mice used in this study. F.M. was supported by the Volkswagen Foundation and B.C. by the Boehringer Ingelheim Fonds. This work was supported by grants from the National Institutes of Health (DP2 OD007371A and R00 HD057233 to A.A.A.; 5R01GM062534 to G.J.H.), by the Ellison Medical Foundation (A.A.A.), and by a kind gift from Kathryn W. Davis (G.J.H.).

Received August 8, 2011; accepted September 26, 2011.

## REFERENCES

- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, 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 AA, Hannon GJ, Brennecke J. 2007a. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**: 761–764.
- Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007b. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744–747.
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**: 785–799.
- Betel D, Sheridan R, Marks DS, Sander C. 2007. Computational analysis of mouse piRNA sequence and biogenesis. *PLoS Comput Biol* **3**: e222. doi: 10.1371/journal.p0030222.
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**: 1089–1103.
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**: 1387–1392.
- Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**: 503–514.
- Copeland NG, Jenkins NA, Court DL. 2001. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**: 769–779.
- Cox DN, Chao A, Lin H. 2000. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**: 503–514.
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**: 798–802.
- Datta S, Costantino N, Court DL. 2006. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**: 109–115.
- Gan H, Lin X, Zhang Z, Zhang W, Liao S, Wang L, Han C. 2011. piRNA profiling during specific stages of mouse spermatogenesis. *RNA* **17**: 1191–1203.
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202.
- Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi M, Siomi MC. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**: 1587–1590.
- Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev* **24**: 2499–2504.
- Harris AN, Macdonald PM. 2001. Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128**: 2823–2832.
- Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**: 69–82.
- Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, Givan SA, Kasschau KD, Carrington JC. 2007. Genome-wide analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 pathway in *Arabidopsis* reveals dependency on miRNA- and tasiRNA-directed targeting. *Plant Cell* **19**: 926–942.
- Huisinga KL, Elgin SC. 2009. Small RNA-directed heterochromatin formation in the context of development: what flies might learn from fission yeast. *Biochim Biophys Acta* **1789**: 3–16.
- Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908–917.
- Lahmy S, Bies-Etheve N, Lagrange T. 2010. Plant-specific multi-subunit RNA polymerase in gene silencing. *Epigenetics* **5**: 4–8.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. doi: 10.1186/gb-2009-10-3-r25.
- Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science* **313**: 363–367.
- Lau NC, Robine N, Martin R, Chung WJ, Niki Y, Berezikov E, Lai EC. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res* **19**: 1776–1785.
- Le T, Yu M, Williams B, Goel S, Paul SM, Beitel GJ. 2007. CaSpeR5, a family of *Drosophila* transgenesis and shuttle vectors with improved multiple cloning sites. *Biotechniques* **42**: 164–166.
- Lemaitre B, Ronsseray S, Coen D. 1993. Maternal repression of the P element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotype. *Genetics* **135**: 149–160.
- Li MA, Alls JD, Avancini RM, Koo K, Godt D. 2003. The large Maf factor Traffic Jam controls gonad morphogenesis in *Drosophila*. *Nat Cell Biol* **5**: 994–1000.
- Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509–521.
- Lim AK, Kai T. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc Natl Acad Sci* **104**: 6714–6719.
- Linsen SE, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, Fritz B, Wyman SK, de Bruijn E, Voest EE, et al. 2009. Limitations and possibilities of small RNA digital gene expression profiling. *Nat Methods* **6**: 474–476.
- Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136**: 656–668.
- Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**: 522–535.
- Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* **29**: 3301–3317.
- Pane A, Wehr K, 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. 1981. The I–R system of hybrid dysgenesis in *Drosophila melanogaster*: are I factor insertions responsible for the mutator effect of the I–R interaction? *Mol Gen Genet* **183**: 123–129.
- Qi H, Watanabe T, Ku HY, Liu N, Zhong M, 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.
- Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC. 2009. A broadly conserved pathway generates 3' UTR-directed primary piRNAs. *Curr Biol* **19**: 2066–2076.
- Roche SE, Rio DC. 1998. *Trans*-silencing by *P* elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, *Enhancer of zeste*. *Genetics* **149**: 1839–1855.
- Ronsseray S, Lehmann M, Anxolabéhère D. 1991. The maternally inherited regulation of *P* elements in *Drosophila melanogaster* can be elicited by two *P* copies at cytological site 1A on the X chromosome. *Genetics* **129**: 501–512.
- Ronsseray S, Lehmann M, Nouaud D, Anxolabéhère D. 1996. The regulatory properties of autonomous subtelomeric *P* elements are sensitive to a *Suppressor of variegation* in *Drosophila melanogaster*. *Genetics* **143**: 1663–1674.
- Ronsseray S, Josse T, Boivin A, Anxolabéhère D. 2003. Telomeric transgenes and *trans*-silencing in *Drosophila*. *Genetica* **117**: 327–335.
- Rubin GM, Kidwell MG, Bingham PM. 1982. The molecular basis of P–M hybrid dysgenesis: the nature of induced mutations. *Cell* **29**: 987–994.
- Saito K, Siomi MC. 2010. Small RNA-mediated quiescence of transposable elements in animals. *Dev Cell* **19**: 687–697.
- Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* **20**: 2214–2222.
- Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. 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 KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24**: 2493–2498.
- Sarot E, Payen-Groschene G, Bucheton A, 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.
- Senti KA, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* **26**: 499–509.
- Sharan SK, Thomason LC, Kuznetsov SG, Court DL. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4**: 206–223.
- Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246–258.
- Todeschini AL, Teyssset L, Delmarre V, Ronsseray S. 2010. The epigenetic *trans*-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PLoS ONE* **5**: e11032. doi: 10.1371/journal.pone.0011032.
- Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, Cook HA, Koppetsch BS, Theurkauf WE, Zamore PD. 2004. RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* **116**: 831–841.
- Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.
- Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23**: 1749–1762.
- Venken KJ, He Y, Hoskins RA, Bellen HJ. 2006. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**: 1747–1751.
- Venken KJ, Carlson JW, Schulze KL, Pan H, He Y, Spokony R, Wan KH, Koriabine M, de Jong PJ, White KP, et al. 2009. Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nat Methods* **6**: 431–434.
- Wilson C, Pearson RK, Bellen HJ, O'Kane CJ, Grossniklaus U, Gehring WJ. 1989. P-element-mediated enhancer detection: An efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev* **3**: 1301–1313.
- Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25–33.