Article

Endogenous RNA Interference Provides a Somatic Defense against *Drosophila* Transposons

Wei-Jen Chung,[1,2] Katsutomo Okamura,[1,2] Raquel Martin,[1,2] and Eric C. Lai[1,*]
[1]Department of Developmental Biology
Sloan-Kettering Institute
521 Rockefeller Research Laboratories
1275 York Ave, Box 252
New York, New York 10065

Summary

Background: Because of the mutagenic consequences of mobile genetic elements, elaborate defenses have evolved to restrict their activity. A major system that controls the activity of transposable elements (TEs) in flies and vertebrates is mediated by Piwi-interacting RNAs (piRNAs), which are ∼24–30 nucleotide RNAs that are bound by Piwi-class effectors. The piRNA system is thought to provide primarily a germline defense against TE activity.
Results: Here, we describe a second system that represses *Drosophila* TEs by using endogenous small interfering RNAs (siRNAs), which are 21 nucleotide, 3′-end-modified RNAs that are dependent on Dicer-2 and Argonaute-2. In contrast to piRNAs, we find that the TE-siRNA system is active in somatic tissues, and particularly so in various immortalized cell lines. Analysis of the patterns and properties of TE-derived small RNAs reveals further distinctions between TE regions and genomic loci that are converted into piRNAs and siRNAs, respectively. Finally, functional tests show that many transposon transcripts accumulate to higher levels in cells and animal tissues that are deficient for Dicer-2 or Argonaute-2.
Conclusions: *Drosophila* utilizes two small-RNA systems to restrict transposon activity in the germline (mostly via piRNAs) and in the soma (mostly via siRNAs).

Introduction

After the initial discovery of RNA interference (RNAi), a post-transcriptional silencing mechanism involving double-stranded RNA (dsRNA) [1], intense studies revealed a cornucopia of related RNA-silencing pathways mediated by Argonaute proteins [2]. The Argonaute family includes both AGO-class and Piwi-class members, all of which use a small-RNA guide to identify complementary transcripts for repression. Their best-characterized RNA guides include small interfering RNAs (siRNAs), microRNAs (miRNAs), and Piwi-interacting RNAs (piRNAs). Generally speaking, 24–30 nt piRNAs load into Piwi-class proteins, which are mostly restricted to the germline, whereas ∼21–22 nt siRNAs and miRNAs load into AGO-class proteins, which are more broadly expressed [2, 3].

Mutants in core components of the *Drosophila* miRNA pathway are lethal [4, 5], reflecting the fundamental roles of miRNAs in host-gene regulation. Mutants in the core piRNA pathway are sterile and exhibit massive deregulation of

*Correspondence: laie@mskcc.org
[2]These authors contributed equally to this work

transposon activity in the germline [6]. Mutants in the *Drosophila* RNAi pathway, such as the core siRNA-generating enzyme Dicer-2 and the core siRNA-effector AGO2, are virus hypersensitive, indicating that RNAi restricts viral replication [7]. On the other hand, the fact that RNAi mutants are otherwise homozygous viable, fertile, and morphologically quite normal [4, 8] raised the question of whether endogenous RNAi plays any significant regulatory role in this organism.

Although some studies in the literature are confounded by the fact that certain earlier-classified "RNAi" pathway components actually involve piRNAs, several reports have suggested that canonical RNAi controls the transposition of selfish genetic elements. This is most evident in *C. elegans*, for which several mutants that deregulate transposons are also RNAi defective [9]. Transcriptional read through across worm Tc1 elements was proposed to generate RNA snapbacks between terminal IR sequences, which are processed into siRNAs that silence Tc1 elements in *trans* [10]. In mammalian cells, some 21 nt RNAs were recovered from retrotransposons and shown to mediate LINE element silencing [11, 12]. The control of transposon activity in mouse oocytes by siRNAs was recently shown more explicitly [13, 14]. In *Drosophila*, both siRNAs and piRNAs were reported to regulate F element activity [15], and depletion of AGO-class proteins increases transposon transcript levels in cultured cells [16].

We now show that transposons are a substantial source of endogenous siRNAs in *Drosophila*. We can confidently define these on the basis of their precise 21 nt size and their 3′-end modification, consistent with their genetic requirements for Dicer-2 and AGO2. We find that the siRNA system, in contrast to the piRNA system, is active in somatic tissues and cultured cells. Consistent with this, we observe that Dicer-2- and AGO2-deficient cells deregulate the expression of multiple transposon classes. Similar to concurrent studies [17–19], we conclude that endogenous RNAi provides a somatic layer of defense against selfish genetic elements in *Drosophila*.

Results

Transposons Are a Major Source of Endogenous siRNAs in *Drosophila*

Using massive data sets generated by Solexa sequencing, we analyzed the size of small RNAs mapping to annotated transposons. Our data comprise nearly 15 million mapped small RNAs (including more than 3.7 million miRNA reads) cloned from 0–1 hr embryos, 2–6 hr embryos, 6–10 hr embryos, mass-isolated imaginal discs and brains, female heads, female bodies, male heads, male bodies, S2 cells, and Kc cells (Table S1 available online). In addition to piRNAs (classified here as 24–30 nt reads matching perfectly to transposon sense or antisense sequences), we also observed a considerable peak of 21 nt reads matched to transposons (Figure 1). In contrast to miRNAs, which have an average length of 21.91 nt, the precise 21 nt peak of transposable element (TE) reads corresponded to the size of Dcr-2-processed siRNAs produced from a perfectly double-stranded substrate [20]. We therefore called these TE-siRNAs.
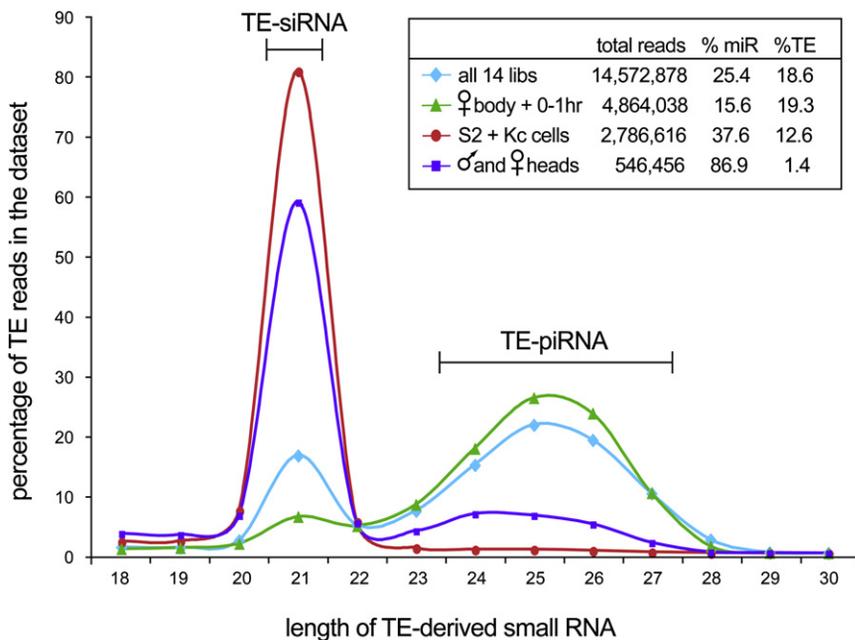
Figure 1. *Drosophila* Transposable Elements Generate Two Classes of Small RNAs with Distinct Size and Expression Characteristics

The inset tabulates the number of total reads, miRNA reads, and TE reads in each set of libraries; only the TE reads were considered in the main graph. Analysis of all transposon-derived small-RNA reads from all libraries (light blue dataset) reveals a 21 peak that is distinct from the 24–27 nt peak. The latter corresponds to piRNAs, which are especially abundant in the female body and early embryos (green dataset). However, these samples contain a distinct peak of 21 nt reads corresponding to siRNAs. Adult heads (purple dataset) and S2 + Kc cells (red dataset) generate TE-small RNAs that are nearly exclusively 21 nt in size.

The biphasic size distribution of TE-derived small RNAs clearly indicated that the 21 nt RNAs were not simply a subpopulation of degraded piRNAs but rather derived from a specific biogenesis pathway. Consistent with this, TE-derived siRNAs and piRNAs exhibited distinct patterns of spatiotemporal accumulation. TE-piRNAs were especially abundant in female bodies and early embryos, whereas TE-siRNAs were strongly enriched in some somatic tissues. For example, TE-derived reads from adult heads were more than twice as likely to be exactly 21 nt as they were to be 24–30 nt combined (Table S1). Strikingly, the TE-derived small RNAs of S2 cells and Kc cells were nearly exclusively 21 nt in length (Figure 1 and Table S1).

Relationship between TE-siRNAs and Parent TEs

The small RNAs of S2 and Kc cultured cells were composed of 12.6% TE reads, whereas male and female heads contained only 1.4% TE reads (Figure 1). The greater accumulation of TE-siRNAs in S2 and Kc cells correlated with the fact that many long terminal repeat (LTR) retrotransposons, including *copia*, *1731*, *412*, *297*, and *mdg* elements, are highly expressed and genomically amplified in cultured cells [21–24]. Indeed, these elements were among the most highly expressed TE-siRNA families and in some cases generated more overall siRNAs than piRNAs (Table 1). The complete tabulations of siRNA and piRNA reads mapped to all canonical TE sequences are available in Table S2.

We examined the distribution of siRNAs and piRNAs mapped onto TE consensus sequences. For this purpose, we used only 21 nt reads from S2 and Kc cells and adult heads as siRNA populations, and we used $\geq$ 24 nt reads from female bodies and 0–1 hr embryos as piRNA populations (Figure 1). Most *Drosophila* TEs generated piRNAs and siRNAs from both sense and antisense strands along the entire TE (Figure S1). In the case of *mdg1*, siRNAs and piRNAs were appreciably depleted in the LTRs relative to the TE body (Figure S1A). In the case of the *297* element, there were substantial hotspots of siRNAs and piRNAs mapped to its LTRs (even after we corrected for double mappings to both LTRs); still, these were not so remarkably enriched relative to the body of the TE (Figure S1B). In light of these mapping patterns, it is relevant to recall that substantial quantities of cytoplasmic, polyadenylated dsRNA comprising the entirety of *mdg1* and *mdg3* elements was reported nearly 30 years ago [25].

Despite their distinct sizes and tissue distributions, TE-derived siRNAs and piRNAs shared a bias to match the antisense strand. Analyzed over all TE families (Table S2), this bias was

Table 1. The Top Ten siRNA-Generating TE Families Identified from Our 14 Solexa Libraries

TE Name	Type	Total siRNAs	Sense siRNAs	Antisense siRNAs	Total piRNAs	Sense piRNAs	Antisense piRNAs
mdg1	LTR	85280	34643	50637	14260	1140	13120
412	LTR	65044	27677	37367	11846	3987	7859
297	LTR	53854	16192	37662	96272	2244	94028
blood	LTR	46094	18079	28015	126690	3503	123187
roo	LTR	35985	15894	20091	271160	23063	248097
1731	LTR	30585	<u>16661</u>	<u>13924</u>	12174	<u>6623</u>	<u>5551</u>
diver	LTR	27177	11032	16145	7729	3390	4339
copia	LTR	16827	<u>9967</u>	<u>6860</u>	27521	<u>14615</u>	<u>12906</u>
TART-B	non-LTR	14623	5911	8712	33980	13007	20973
3S18 (BEL)	LTR	12589	5649	6940	11927	3317	8610
Total canonical reads, all TEs		506506	206625	299881	1906336	503517	1402819

We mapped 21 nt (siRNA) and $\geq$ 24 nt (piRNA) reads to consensus TE sequences. Nine of the ten families account for 73.7% of all consensus TE-siRNAs and are LTR retrotransposons. Bulk TE-siRNAs across most TE classes collectively exhibit a slight bias for antisense reads, although less pronounced than for bulk TE-piRNAs. Note that *1731* and *copia* elements had a mild sense bias for bulk siRNAs and piRNAs (underlined). The complete TE mapping data are available in Table S2.
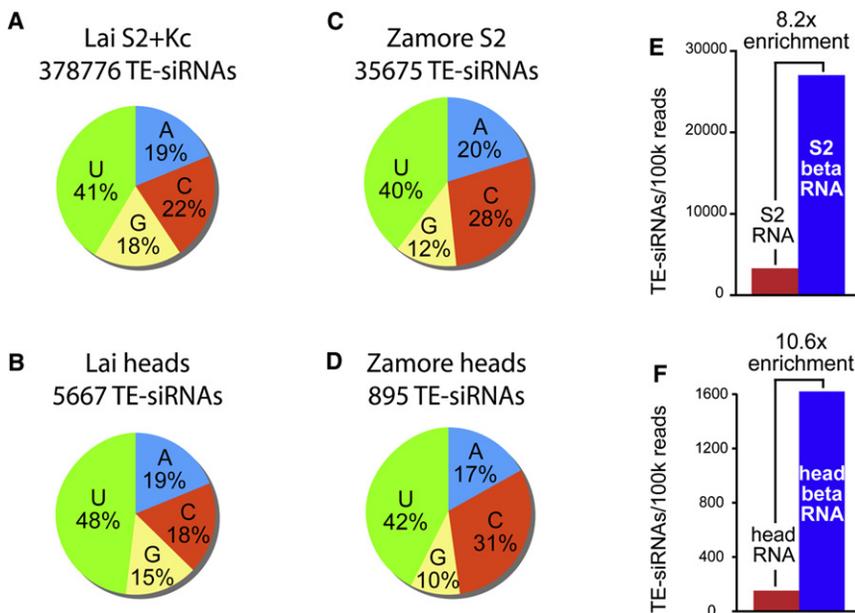
Figure 2. 5′ and 3′ Characteristics of TE-siRNAs

(A–D) 5′ nucleotide bias of TE-siRNAs; i.e., 21 nt reads matching the sense or antisense strand of TEs. (A) TE-siRNAs in our S2 and Kc cell libraries. (B) TE-siRNAs in our male and female head libraries. (C) TE-siRNAs in the S2 data of Seitz and Zamore [26]; combined reads from non-β-eliminated and β-eliminated RNA. (D) TE-siRNAs in the head data of Seitz and Zamore [26]; combined reads from non-β-eliminated and β-eliminated RNA. In all datasets, there is preference for 5′ U and mild bias against 5′ G.

(E and F) TE-siRNAs from S2 RNA and β-eliminated S2 RNA (E), and head RNA and β-eliminated head RNA (F) in the data of Seitz and Zamore [26]. Counts in each dataset were normalized per 100,000 total reads. In both S2 and head data, there is strong enrichment for TE-siRNAs after β-elimination.

less for TE-siRNAs (59%) than for TE-piRNAs (75%). However, because this bias (1) was independently observed across the vast majority of distinct TE families, and (2) was independently observed in five different cultured-cell and head libraries (whose TE reads are mostly siRNAs), this antisense bias appeared to be significant. Of the top ten families of TE-siRNAs, only two (*1731* and *copia*) did not exhibit an antisense siRNA bias; interestingly, these were also the only two elements in top ten that did not exhibit an antisense bias in bulk piRNAs. The mild preference for antisense TE-siRNAs was consistent with the possibility of their preferred usage to target sense TE transcripts.

5′ and 3′ Characteristics of TE-siRNAs

We examined the 5′ and 3′ ends of TE-siRNAs in further detail, using only the 21 nt TE reads from head and cultured-cell libraries (Figure 1). We first noticed that bulk TE-siRNAs were biased to begin with a 5′ uridine residue (Figures 2A and 2B). The 5′-U bias of TE-siRNAs was less than that of bulk piRNAs and miRNAs, but it was significant given the large number of TE-siRNA reads analyzed (nearly 400,000) and the fact that similar 5′-U biases were observed in five independently constructed libraries (ranging from 40% to 48% 5′ U, Table S3). We corroborated these results by analyzing the TE-siRNAs of S2 and head libraries reported by Seitz and colleagues [26] and similarly found them to be composed of 40%–42% 5′ U (Figures 2C and 2D and Table S3). Curiously, TE-siRNAs from all of these libraries were mildly deficient in 5′ G residues (Figures 2A–2D).

TE-piRNAs and TE-siRNAs were also similar at their 3′ ends. Unlike miRNAs, piRNAs and exogenous siRNAs are methylated at their 3′ ends by the Hen1 methyltransferase [27–29]. Such a 3′ blockage can be inferred from the enrichment of small-RNA sequences after their cloning from oxidized, β-eliminated samples; this cloning depletes for RNAs with 2′, 3′ hydroxy termini [26]. For example, in the data of Seitz of colleagues [26], bulk miRNAs were depleted from β-eliminated samples, whereas exogenously induced siRNAs were enriched. We found that TE-siRNAs of all classes—LTR retrotransposon, non-LTR retrotransposon, and inverted repeat (IR) elements—were all enriched after β-elimination (Table S4), yielding 8- to 10-fold more TE-siRNAs in treated samples

(normalized according to the number of reads per library). This was true of TE-siRNAs from both S2 cells and adult heads (Figures 2E and 2F), indicating that TE-siRNAs in cultured cells and in the animal are generally blocked at their 3′ ends.

Distinct Genomic Origins of TE-siRNAs and TE-piRNAs

We examined the genomic distribution of TE-derived small RNAs with the aim of understanding their origins. Once again, in order to consider relatively pure populations of siRNAs and piRNAs, we restricted our analysis to the 21 nt TE-derived siRNAs of S2 and Kc cells and the ≥ 24 nt TE-derived piRNAs of female bodies and 0–1 hr embryos. As is the case for TE-piRNAs, TE-siRNAs mapped broadly across the genome to the locations of most transposons (Figures 3A, 3B, and 3E). However, a very clear difference emerged when all siRNA or piRNA mappings were compared to a set of normalized siRNA or piRNA mappings, in which the raw read number at any position was divided by its number of perfect genomic hits. It was previously observed that this procedure collapses the vast majority of TE-piRNA peaks into a few predominant genomic clusters, termed "piRNA master loci" [30] (compare Figures 3E and 3F). In contrast, the genomic landscape of normalized siRNA density had a similar overall profile to the raw siRNA density, except that the read density was reduced about 30-fold because of multiple mappings (compare Figures 3B and 3C). This suggested that TE-siRNAs do not predominantly derive from master loci as inferred for TE-piRNAs.

This picture was further supported by considering only uniquely mapped TE-siRNAs and piRNAs. In the case of TE-piRNAs, this reduced their landscape to only a handful of predominant clusters (e.g., 42AB on chromosome 2R, Figure 3G). Nevertheless, the genomic densities of normalized total piRNAs and unique piRNAs were not very different, with the latter reduced by less than two-fold (compare Figures 3F and 3G). This pattern was nearly identical to a previous report of *Drosophila* piRNAs [30]. In contrast, restricting the analysis to uniquely mapped TE-siRNAs severely reduced their numbers, reducing their read density by about 30-fold relative to the normalized siRNA density (compare Figures 3C and 3D).

The apparent clusters of uniquely mapped TE-siRNAs showed partial overlap with TE-piRNA clusters (Figures 3D
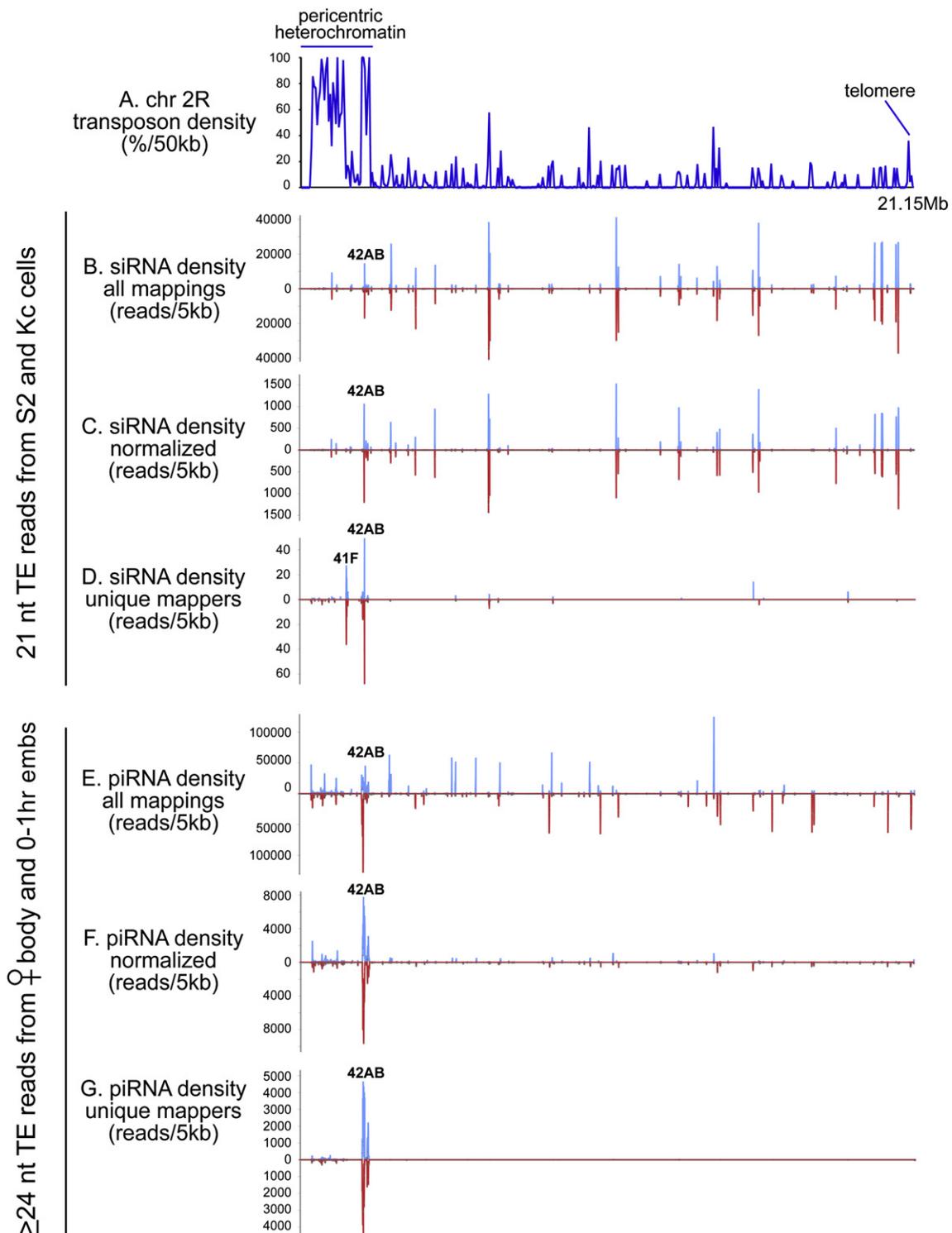
Figure 3. Genomic Origins of TE-siRNA and piRNAs

(A) Overall transposon density across the ~21 megabases of chromosome 2R.

(B–D) Density of 21 nt TE reads from S2 and Kc cells; graphs depict all such reads mapped to all locations (B), reads normalized for multiple mapping (C), or just the uniquely mapped reads (D).

(E–G) Similar analyses were performed for the ≥ 24 nt TE reads from female body and 0–1 hr embryos. siRNAs and piRNAs generally map all over the chromosome (B and E). Normalization for multiple mappings does not significantly change the TE-siRNA landscape (other than causing a ~30-fold reduction in overall numbers (C), whereas this treatment severely dampens the TE-piRNA landscape, leaving only the predominant 42AB cluster (F). Analysis of uniquely mapped siRNAs results in another ~30-fold reduction in overall numbers, leaving behind a few modest clusters in addition to 42AB (D). In contrast, the uniquely mapped piRNAs collapse to only the 42AB cluster and other regions of the pericentric heterochromatin but maintain > 50% of the density values of the normalized, total piRNA population (G).
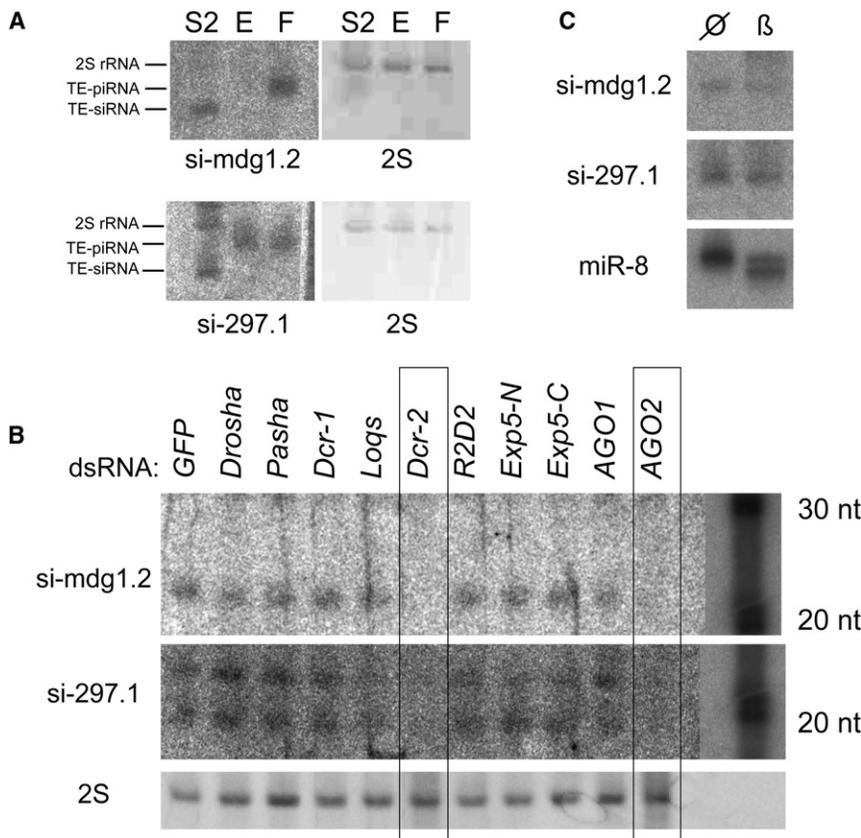
Figure 4. Chemical Properties and Biogenesis Requirements of TE-siRNAs

(A) TE probes that detect mdg1- and 297-derived piRNAs in early embryos (E) and female bodies (F) instead detect siRNAs in S2 cells (S2); 2S indicates ethidium staining of 2S rRNA.

(B) S2 cells were treated with the designated dsRNAs and tested for the accumulation of 21 nt TE-siRNAs. The accumulation of both TE-siRNAs is highly dependent on Dcr-2 and AGO2, and 297.1 is mildly dependent upon Loqs. Because individual siRNAs are present at low levels, the exposure of the main portion of the blot was adjusted separately from the RNA ladders. Longer gels were used in (B) compared to (A) and resolved a ~19 nt band that hybridized to the 297.1 probe. Although its identity is unclear, it exhibited similar sensitivity to the 21 nt band.

(C) TE-siRNAs are resistant to β-elimination (β) and show the same mobility as untreated RNA (null sign). In contrast, miRNAs such as miR-8 exhibit increased mobility after β-elimination.

These probes allowed us to examine TE-siRNA biogenesis by using dsRNA-mediated knockdowns of candidate factors. We analyzed a panel of canonical miRNA pathway components (*Drosha*, *Pasha*, *Exp-5*, *Dcr-1*, *Loqs*, *AGO1*) and canonical siRNA pathway components (*Dcr-2*, *r2d2*, *AGO2*) and used dsRNA against *GFP* as a control. We found that TE-siRNAs required Dcr-2 and AGO2 for their accumulation, consistent with their classification as siRNAs (Figure 4B). Interestingly, knockdown of the double-stranded RNA binding domain (dsRBD) factor R2D2 had little impact on TE-siRNA accumulation, in contrast to its reported role as an obligate partner of Dcr-2. Instead, we observed a mild reduction of 297.1 siRNAs in cells depleted for the dsRBD Loqs. This indicated that a canonical miRNA factor is actually required for the accumulation of at least some TE-siRNAs (Figure 4B), adding complexity to the proposed partnerships of dsRBD and Dicer proteins [2].

We sought experimental evidence for the 3′ modification of TE-siRNAs, as was suggested by the read-count analysis of β-eliminated samples (Figures 2E and 2F). We tested this directly by probing for TE-siRNAs in β-eliminated RNAs. As seen in Figure 4C, both mdg1.2 and 297.1 siRNAs were resistant to this treatment, whereas a majority of miR-8 was sensitive and migrated more quickly after β-elimination. These data demonstrate that TE-siRNAs are indeed modified at their 3′ termini.

and 3G). However, their relatively small numbers of reads suggested that they were not ultimately the origin of TE-siRNAs (as with TE-piRNA clusters). Rather, it appears that these genomic regions, as transposon "graveyards," incidentally happen to have an especially high density of mutated TEs and therefore more uniquely encoded TE sequences. Instead, our observations are consistent with the view that bulk TE-siRNAs are produced from TE transcripts that derive from large numbers of genomic locations. The distinct profiles of TE-siRNAs and TE-piRNAs in the normalized and unique mappings were reproduced when the S2 and Kc data were considered separately, providing further evidence for their distinct genomic origin. A full analysis of all TE-siRNAs and TE-piRNAs across all the chromosome arms is reported in Tables S5 and S6, respectively.

A possible caveat to this analysis lies in the fact that the bulk of TE-siRNAs available for analysis derived from S2 and Kc cells, which have experienced genomic amplification of various TEs. Therefore, we cannot rule out the possibility that the genomes of S2 and Kc cells might contain specialized TE-siRNA master loci that are not apparent from our mappings to the reference *y; cn bw sp Drosophila melanogaster* genome. However, our data suggest that the previously identified piRNA master loci are not likely to be the major source of TE-siRNAs.

Genetic Requirements for TE-siRNA Biogenesis

We identified individual TE-siRNAs that were abundant enough to be detected with conventional northern analysis. Interestingly, the same probes that detected 21 nt *mdg1*- and *297*-derived siRNAs in S2 cells instead hybridized to ~26 nt TE-piRNAs in female bodies (Figure 4A). This served as explicit evidence for separate mechanisms for TE-siRNA and piRNA production.

RNAi Restricts the Level of Transposon Transcripts

We asked whether components of TE-siRNA biogenesis were required to restrain transposon activity. We treated S2 cells with dsRNA against core RNAi factors and then monitored the levels of various TE transcripts by using qRT-PCR (normalizing these data to their level after treatment with *GFP* dsRNA). We found that the RNA levels of many different TEs and repetitive elements, including *297*, *mdg1*, *1731*, *BEL*, *F* and *Blood*, were increased several-fold after knockdown of either *Dcr-2* or *AGO2* (Figure 5A). These data corroborate a previous study that noted increased TE levels after *AGO1* and/or *AGO2* knockdown [16]. Our biogenesis data refine this observation
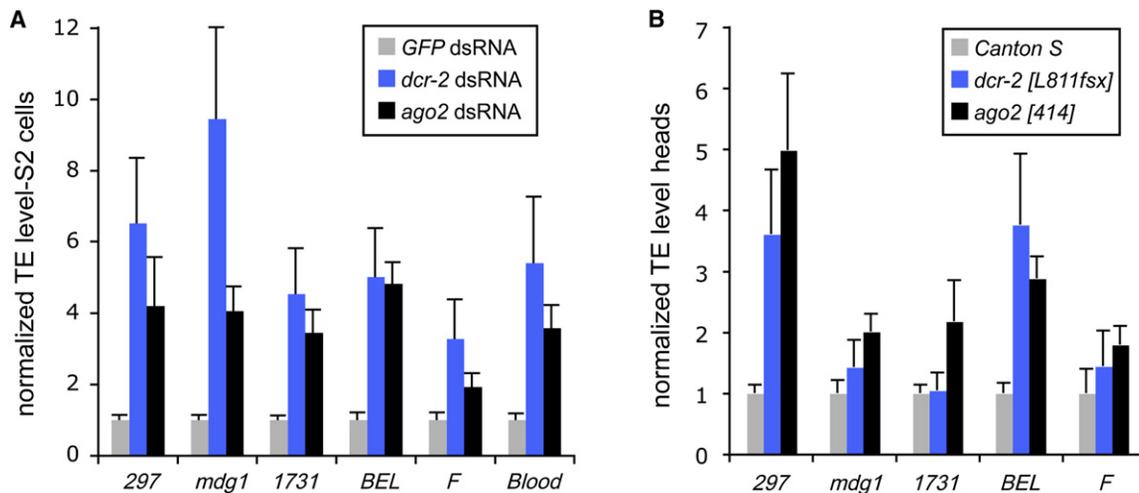
Figure 5. Dcr-2- and AGO2-Deficient Cells Exhibit Increased Levels of TE Transcripts

Relative RNA levels (mean ± SD) are shown.

(A) TE transcripts were measured by quantitative RT-PCR of S2 cells treated with dsRNA against *Dcr-2* or *AGO2*, normalized to GFP dsRNA-treated cells.

(B) The heads of *dcr-2* or *ago2* homozygous flies similarly exhibit increased levels of several TE transcripts relative to Canton S heads.

by demonstrating that Dcr-2 mediates the production of TE-siRNAs, which in turn repress TE accumulation via AGO2. The mild requirement previously reported for AGO1 might potentially reflect that a subset of TE-siRNAs are sorted to AGO1, just as some miRNAs are reciprocally sorted to AGO2 [31, 32].

We wondered whether this phenomenon was limited to cultured cells. Homozygous *dcr-2* or *ago2* mutant animals are viable, fertile, and of relatively normal morphology; nevertheless, qRT-PCR tests revealed that several TEs exhibited elevated levels in the heads of one or both mutants (Figure 5B). The derepression of TE transcript levels in these null mutant heads was less than that observed for RNAi-deficient S2 cells, even though the knockdown approach induces only partial loss of function. This was probably due to the elevated levels of TE activity, and thus heightened TE-siRNA response, seen in S2 cells (Figure 1). In summary, our finding that multiple classes of TEs are deregulated in cultured cells and animals that are deficient for core RNAi components demonstrates the utilization of siRNAs to restrain TE activity in somatic cells.

Conclusions

Multiple Classes of Endogenous siRNAs in Drosophila

Although the *Drosophila* RNAi pathway produces regulatory siRNAs in response to viral invasion, exogenous dsRNA, or IR transgenes [4, 7, 33], relatively little was historically known about the endogenous usage of *Drosophila* RNAi. In this study, we described a rich set of bona fide siRNAs that derive from transposable elements in *Drosophila*. These data add to a host of concurrent studies that recently elucidated multiple classes of siRNAs that derive from the host genome, not only from TEs, but also from 3′ *cis*-natural antisense gene pairs, long IR transcripts, and two unique intronic and exonic clusters localized to the *klarsicht* and *thickveins* genes [17–20, 34]. Although these myriad siRNAs differ in origin, with some derived from bidirectional transcription and others from intramolecular dsRNA, they are united by their dependence on Dcr-2 and Ago2, their 3′-end modification, and, for at least some members of each class, an appreciable dependence on Loqs.

piRNAs and siRNAs Mediate Parallel Restriction of Drosophila Transposons

We may confidently distinguish TE-siRNAs from previously described TE-piRNAs on the basis of their characteristic sizes, genomic origins, tissue distribution, and origin from within a given TE. Both types of small-RNA pathways are demonstrably required to restrict TE transcript accumulation, and their separable roles correlate with their distinct tissue requirements. The germline is highly active in TE-piRNA production and uses piRNA components to restrict TE accumulation [6], whereas somatic tissues such as adult heads specifically produce TE-siRNAs and use RNAi components to restrict TE levels. Similar conclusions on TE-siRNA biogenesis and function were recently reached in the concurrent studies of other groups [17–19]. Curiously, whereas the mouse male germline depends strongly on piRNAs to restrict transposon activity, the mouse female germline appears to use both piRNAs and siRNAs to control TE activity [13, 14, 35]. Therefore, there has been evolutionary flux in how these conserved small-RNA pathways are used to control TEs in animals.

Curiously, we found that independently derived lines of cultured cells, namely S2 and Kc, exhibit pronounced siRNA responses to a subclass of LTR retrotransposons. This can be directly correlated with the fact of deregulation and genomic amplification of these particular TEs [21–23]. It is possible that transposon deregulation was a direct consequence of the process of cell immortalization. However, one could speculate that their deregulation was a gradual consequence of divorcing these cells from piRNA control, which in the animal occurs mostly in the germline and is transmitted from generation to generation via maternal deposition of piRNA complexes into the embryo. In either case, the stronger TE-siRNA response in cultured cells may be viewed as an adaptive response to deregulated transposons, as proposed for the piRNA pathway [30].

Experimental Procedures

Molecular Analyses of Small RNAs and Their Function

We isolated total RNA from staged preparations of Canton S animals, S2 cells, or Kc cells by using Trizol, and ~18–26 nt libraries were generated via the

method of Lau and colleagues [36]. Adult flies were frozen, vortexed, and sieved to yield head and body fractions; legs were mostly lost in this procedure. High-throughput sequencing was performed with the Illumina 1G Genome analyzer. We used standard methods for small-RNA northerns [37] with locked nucleic acid probes (Exiqon): 297.1, 5′-AAGAACCCAAGAGCGAGGCTCTCC-3′; mdg1.2, 5′-CAAGTGCACTCGTAAACACTCAGAA-3′. dsRNA treatment of S2 cells was performed with the soaking strategy and previously described templates [34, 38]; β-elimination of RNA was performed as described [28]. Quantitative RT-PCR was performed as described [39] with SYBR Green (ABI) and the IQ5 Real-Time PCR System (BioRad). The expression levels were normalized to corresponding *rp49* values, and then each of the knockdown samples was normalized to the value from *GFP* knockdown (for S2 samples) or *Canton S* (for adult head samples). qPCR primers were as follows: Blood_637, 5′-GAC CAAAGCCCTTGACCATA-3′; Blood_717, 5′-GGCCACCCCTCTTCTTTTTA-3′; 1731_746+, 5′-CTGAGCAAACGTCTGTTGGA-3′; 1731_825−, 5′-GCAT CAAGGGCATCAAAGAT-3′; mdg1_F, 5′-AACAGAAACGCCAGCAACAGC-3′; mdg1_R2, 5′-TTTCTGATCTTGGCAGTGGA-3′; 297_884+, 5′-GGTGATCCA GAAACCCTTCA-3′; 297_993−, 5′-CTTTCGATGGCTCCCAGTAG-3′; Felement_2296+, 5′-TCATCTTCCATCGTTGTGGA-3′; Felement_2376−, 5′-CA CATTCTGCAGTTCGCTTC-3′; BEL_1229+, 5′-GGGATCCCTGGCTAATTTTC -3′; BEL_1336−, 5′-ATCGGTTGATGGTCACACCT-3′; rp49 A2, 5′-ATCGGT TACGGATCGAACA-3′; and rp49 B2, 5′-ACAATCTCCTTGCGCTTCTT-3′.

Computational Analyses of Small-RNA Sequences

We clipped the raw sequences by demanding that at least six our of seven nucleotides match to the 3′ proximal linker sequence (5′-CTGTAGG-3′). Because of imprecision in size selection during cloning and/or RNA degradation, we recovered clipped reads that were shorter or longer than the selected 18–26 nt window. Given that short sequences can be of ambiguous genomic origin, we kept only reads ≥ 18 nt in length; all other reads were kept, although we note that the content of > 26 nt reads is probably not completely representative of total RNA. Even so, our libraries probably retained the bulk of the piRNA population, whose average sizes are 24–26 nt [30]. The clipped reads from the nine libraries were deposited at NCBI-GEO (see Accession Numbers).

We mapped the clipped reads to the dm3 (April 2006) *Drosophila* assembly by using Release 5.3 annotations and the UCSC genome browser [40]. We extracted all reads that mapped perfectly to either strand of an annotated transposable element and binned these into the 21 nt (siRNA) and ≥ 24 nt (piRNA) classes. The genomic mappings were used to derive overall TE and miRNA read counts (Figure 1 and Table S1), the length distribution of TE reads (Figure 1), the nucleotide composition of TE-siRNAs (Figure 2 and Table S3), and the genomic origin of TE-siRNAs and TE-piRNAs (Figure 3 and Tables S5 and S6). In general, we tracked reads nonredundantly, meaning that an individual read with more than one perfect genomic hit was still considered only once. The exception to this was in the genomic maps in Figure 3, Table S5, and Table S6, where we considered redundant mappings (where each read was counted in each perfectly matching genomic location), normalized mappings (where the read count at a given genomic locations was divided by the number of perfectly matching genomic locations), and unique mappings (only those reads with a single genomic match).

To generate the mappings to consensus TEs (Figure S1), we collected all reads that matched perfectly to a given TE consensus sequence [40] and divided the read number by the number of hits to the consensus TE (to account for double mappings to LTRs). We also used the consensus mappings to derive the sense or antisense strand bias of TE reads (Table 1 and Table S2). The number of reads that mapped perfectly to TE consensus sequences was lower than the total number of reads that mapped to genomic regions annotated as TEs, because a substantial number of TEs are mutated or damaged. However, when total genomic mappings were used to analyze strand bias, the sum of sense and antisense reads for a given element was often substantially greater than the total number of nonredundant reads. This appeared to be result of TE sequences arranged in complex nested patterns, for which segments arranged in antisense orientation were genomically annotated as sense orientation, and vice versa. We avoided this ambiguity by considering only consensus TE mappings, which should reflect active transposons, to derive information on their strand bias.

Accession Numbers

The clipped reads from the nine libraries were deposited at NCBI-GEO under the following accession numbers: GSM286601 (male head), GSM286602 (male body), GSM286603 (female body), GSM286604 (0–1 hr embryo, first biological replicate), GSM286605 (2–6 hr embryo, first biological replicate), GSM286606 (2–6 hr embryo, second biological replicate), GSM286607 (6–10 hr embryo, first biological replicate), GSM286611 (6–10 hr embryo, second biological replicate), and GSM286613 (0–1 hr embryo, second biological replicate). Other libraries analyzed included GSM240749 (female heads), GSM272651 (S2 and Kc cells), GSM272652 (S2 cells), GSM272653 (Kc cells), and GSM275691 (mixed imaginal discs and brains) [20, 34].

Supplemental Data

One figure and six tables are available at http://www.current-biology.com/cgi/content/full/18/11/795/DC1/.

References

1. 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.

2. Farazi, T.A., Juranek, S.A., and Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development *135*, 1201–1214.

3. 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.

4. 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.

5. Kataoka, Y., Takeichi, M., and Uemura, T. (2001). Developmental roles and molecular characterization of a Drosophila homologue of Arabidopsis Argonaute1, the founder of a novel gene superfamily. Genes Cells *6*, 313–325.

6. Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761–764.

7. Ding, S.W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell *130*, 413–426.

8. 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.

9. Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.

10. Sijen, T., and Plasterk, R.H. (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. Nature *426*, 310–314.

11. Yang, N., and Kazazian, H.H., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat. Struct. Mol. Biol. *13*, 763–771.

12. Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev. *20*, 1732–1743.

13. Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature, in press. Published online April 10, 2008. 10.1038/nature06904.

14. Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al.

(2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature, in press. Published online April 10, 2008. 10.1038/nature06908.

15. Tchurikov, N.A., and Kretova, O.V. (2007). Suffix-specific RNAi leads to silencing of F element in Drosophila melanogaster. PLoS ONE 2, e476.

16. Rehwinkel, J., Natalin, P., Stark, A., Brennecke, J., Cohen, S.M., and Izaurralde, E. (2006). Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster. Mol. Cell. Biol. 26, 2965–2975.

17. 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, in press. Published online April 10, 2008. 10.1126/science.1157396.

18. Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute2 in somatic cells. Nature, in press. Published online May 7, 2008. 10.1038/nature06938.

19. Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J., Sachidanandam, R., et al. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature, in press. Published online May 7, 2008. 10.1038/nature07007.

20. Okamura, K., Balla, S., Martin, R., Liu, N., and Lai, E.C. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila. Nat. Struct. Mol. Biol., in press. 10.1038/nsmb1438.

21. Maisonhaute, C., Ogereau, D., Hua-Van, A., and Capy, P. (2007). Amplification of the 1731 LTR retrotransposon in Drosophila melanogaster cultured cells: Origin of neocopies and impact on the genome. Gene 393, 116–126.

22. Potter, S.S., Brorein, W.J., Jr., Dunsmuir, P., and Rubin, G.M. (1979). Transposition of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 17, 415–427.

23. Di Franco, C., Pisano, C., Fourcade-Peronnet, F., Echalier, G., and Junakovic, N. (1992). Evidence for de novo rearrangements of Drosophila transposable elements induced by the passage to the cell culture. Genetica 87, 65–73.

24. Finnegan, D.J., Rubin, G.M., Young, M.W., and Hogness, D.S. (1978). Repeated gene families in Drosophila melanogaster. Cold Spring Harb. Symp. Quant. Biol. 42, 1053–1063.

25. Ilyin, Y.V., Chmeliauskaite, V.G., and Georgiev, G.P. (1980). Double-stranded sequences in RNA of Drosophila melanogaster: relation to mobile dispersed genes. Nucleic Acids Res. 8, 3439–3457.

26. 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.

27. 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.

28. 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.

29. 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.

30. 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.

31. Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell 130, 299–308.

32. 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.

33. 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.

34. Okamura, K., Chung, W.-J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature, in press. Published online May 7, 2008. 10.1038/nature07015.

35. Murchison, E.P., Stein, P., Xuan, Z., Pan, H., Zhang, M.Q., Schultz, R.M., and Hannon, G.J. (2007). Critical roles for Dicer in the female germline. Genes Dev. 21, 682–693.

36. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294, 858–862.

37. Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 130, 89–100.

38. 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.

39. Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nat. Struct. Mol. Biol. 15, 354–363.

40. Karolchik, D., Kuhn, R.M., Baertsch, R., Barber, G.P., Clawson, H., Diekhans, M., Giardine, B., Harte, R.A., Hinrichs, A.S., Hsu, F., et al. (2008). The UCSC Genome Browser Database: 2008 update. Nucleic Acids Res. 36, D773–D779.