LETTERS

An endogenous small interfering RNA pathway in *Drosophila*

Benjamin Czech[1]*, Colin D. Malone[1]*, Rui Zhou[2], Alexander Stark[3,4], Catherine Schlingeheyde[1], Monica Dus[1], Norbert Perrimon[2], Manolis Kellis[3], James A. Wohlschlegel[5], Ravi Sachidanandam[1], Gregory J. Hannon[1] & Julius Brennecke[1]

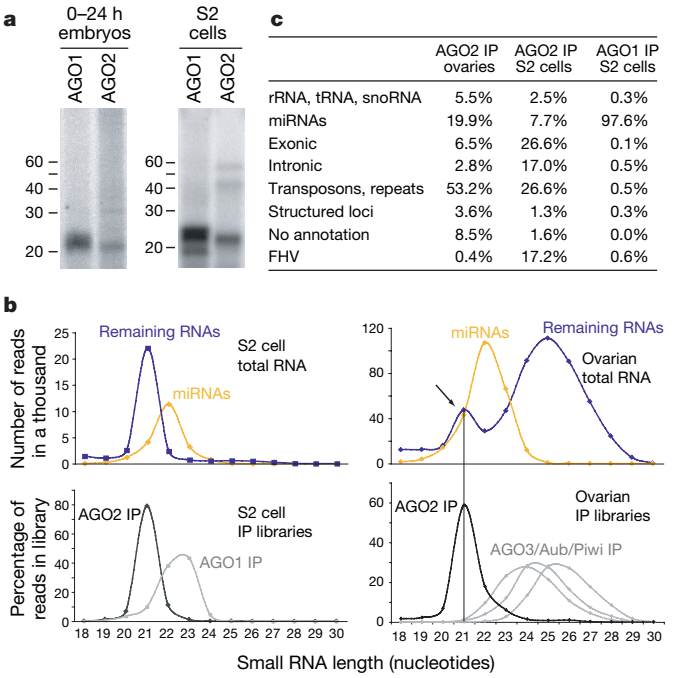
***Drosophila* endogenous small RNAs are categorized according to their mechanisms of biogenesis and the Argonaute protein to which they bind. MicroRNAs are a class of ubiquitously expressed RNAs of ~22 nucleotides in length, which arise from structured precursors through the action of Drosha–Pasha and Dicer-1–Loquacious complexes[1–7]. These join Argonaute-1 to regulate gene expression[8,9]. A second endogenous small RNA class, the Piwi-interacting RNAs, bind Piwi proteins and suppress transposons[10,11]. Piwi-interacting RNAs are restricted to the gonad, and at least a subset of these arises by Piwi-catalysed cleavage of single-stranded RNAs[12,13]. Here we show that *Drosophila* generates a third small RNA class, endogenous small interfering RNAs, in both gonadal and somatic tissues. Production of these RNAs requires Dicer-2, but a subset depends preferentially on Loquacious[1,4,5] rather than the canonical Dicer-2 partner, R2D2 (ref. 14). Endogenous small interfering RNAs arise both from convergent transcription units and from structured genomic loci in a tissue-specific fashion. They predominantly join Argonaute-2 and have the capacity, as a class, to target both protein-coding genes and mobile elements. These observations expand the repertoire of small RNAs in *Drosophila*, adding a class that blurs distinctions based on known biogenesis mechanisms and functional roles.**

Drosophila melanogaster expresses five Argonaute proteins, which segregate into two classes. The Piwi proteins (Piwi, Aubergine and AGO3) are expressed in gonadal tissues and act with Piwi-interacting RNAs (piRNAs) to suppress mobile genetic elements[10,11]. The Argonaute class contains AGO1 and AGO2. AGO1 binds microRNAs (miRNAs) and regulates gene expression[8,9]. The endogenous binding partners of AGO2 have remained enigmatic.

We generated transgenic flies expressing epitope-tagged AGO2 under the control of its endogenous promoter. Tagged AGO2 localized to the cytoplasm of germline and somatic cells of the ovary (Supplementary Fig. 1). Immunoprecipitated AGO2-associated RNAs differed in their mobility from those bound to AGO1 (Fig. 1a). Deep sequencing of small RNAs from AGO1 and AGO2 complexes yielded 2,094,408 AGO1-associated RNAs and 916,834 AGO2-associated RNAs from Schneider (S2) cells, and 455,227 AGO2-associated RNAs from ovaries that matched perfectly to the *Drosophila* genome. We also sequenced three libraries derived from 18–29-nucleotide RNAs (936,833 sequences from wild-type ovaries, 1,042,617 sequences from *Dicer-2* (*Dcr-2*) mutant ovaries, and 1,946,339 sequences from *loquacious* (*loqs*) mutant ovaries) and an 18–24-nucleotide library from wild-type testes (522,848 sequences). Finally, we added to our analysis 92,363 published sequences derived from 19–26-nucleotide RNAs from S2 cells[15].

We noted that among the ~50% of AGO2-associated RNAs from S2 cells that did not match the genome, ~17% matched the flock house virus (FHV), a pathogenic RNA virus and reported target for RNAi in flies[16,17]. These probably arose because of persistent infection of our S2 cultures.

After excluding presumed degradation products of abundant cellular RNAs, we divided each of the total RNA libraries into two categories: annotated miRNAs and the remainder (Fig. 1b). For the S2 cell library, the size distribution of these populations formed two



c

	AGO2 IP ovaries	AGO2 IP S2 cells	AGO1 IP S2 cells
rRNA, tRNA, snoRNA	5.5%	2.5%	0.3%
miRNAs	19.9%	7.7%	97.6%
Exonic	6.5%	26.6%	0.1%
Intronic	2.8%	17.0%	0.5%
Transposons, repeats	53.2%	26.6%	0.5%
Structured loci	3.6%	1.3%	0.3%
No annotation	8.5%	1.6%	0.0%
FHV	0.4%	17.2%	0.6%

Figure 1 | AGO2 binds endogenous small RNAs. a, RNA was isolated from AGO1 and AGO2 immunoprecipitates (IP) from embryos and S2 cells. **b**, Length profiles for small RNAs isolated from S2 cells and ovaries are shown. Species are split into miRNAs and remainder, and are compared to those obtained from individual Argonaute complexes (as indicated). **c**, Annotation of AGO1- and AGO2-associated small RNAs from ovaries and S2 cells. rRNA, ribosomal RNA; snoRNA, small nucleolar RNA; tRNA, transfer RNA.

[1]Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. [2]Harvard Medical School, Department of Genetics, Howard Hughes Medical Institute, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA. [4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Biological Chemistry, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California 90095, USA.
*These authors contributed equally to this work.

peaks, with non-miRNAs lying at 21 nucleotides and miRNAs exhibiting a broader peak from 21 to 23 nucleotides. Libraries derived from AGO1 and AGO2 complexes almost precisely mirrored these two size classes. In the ovary library, this approach revealed three size classes. Whereas two reflected those seen in S2 cells, a third class comprised piRNAs. Again, RNA size profiles from AGO2 or Piwi family immunoprecipitates[12] mirrored those within the total ovary library. These data demonstrate that AGO2 is complexed with a previously uncharacterized population of small RNAs.

Whereas known miRNAs comprised more than 97% of AGO1-associated RNAs in S2 cells, they made up only 8% or 20% of the AGO2-bound species in S2 cells or ovaries, respectively. The remaining small RNAs in AGO2 complexes formed a complex mixture of endogenous siRNAs (endo-siRNAs; Fig. 1c). Among these, transposons and satellite repeats contributed substantially to AGO2-associated small RNAs in S2 cells (27%) and ovaries (53%). The nature of the transposons giving rise to abundant siRNAs in ovaries and S2 cells differed substantially (Fig. 2a), probably reflecting differential expression of specific transposons in these tissues. Unlike piRNAs[12,13,18,19], neither somatic nor germline siRNAs exhibited a pronounced enrichment for sense or antisense species (Supplementary Fig. 2a).

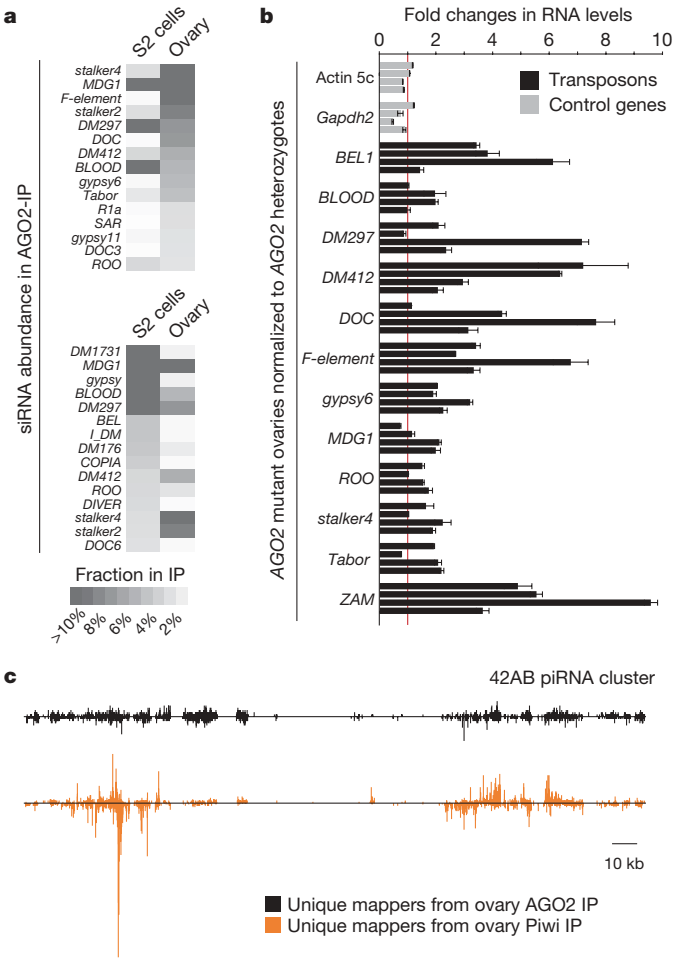


Figure 2 | A subset of endo-siRNAs originates from transposons. **a**, Indicated are the cloning frequencies of AGO2-bound siRNAs in ovaries and S2 cells that match individual transposons. **b**, RNA levels of twelve transposons and two control genes in ovaries mutant for *AGO2* as compared to *AGO2* heterozygotes (four biological replicates; error bars indicate technical variation). **c**, Distributions of AGO2-bound siRNAs (black) and Piwi-bound piRNAs (orange) from ovaries on the piRNA cluster at cytological position 42AB (ref. 12; relative abundances of both populations can be estimated from Supplementary Fig. 2d).

In accord with these findings, knockdown of *AGO2* in S2 cells leads to increased expression of several mobile elements[20]. In the germ line, the Piwi–piRNA system has been reported as the dominant transposon-silencing pathway[19]. Nevertheless, we found that several transposons, with a potential to be targeted by siRNAs, were substantially derepressed in *AGO2* mutant or *Dcr-2* mutant ovaries (Fig. 2b and Supplementary Fig. 2c). Although comparisons of relative abundance were difficult, both piRNAs and siRNAs mapped to piRNA clusters, with the regions that generate uniquely mapping species generally overlapping (Fig. 2c and Supplementary Fig. 2d). Thus, piRNA loci are a possible source for antisense RNAs matching transposons and might serve a dual function in small RNA generation. Considered together, these data suggest that endo-siRNAs repress the expression of mobile elements, in some tissues acting alongside piRNA pathways.

To probe the nature of the remaining endo-siRNAs, we computationally extracted genomic sites, which give rise to multiple uniquely mapping RNAs that do not fall into heterochromatic regions. These generally segregated into two categories, which we term structured loci and convergently transcribed loci.

Transcripts from structured loci can fold to form extensive double-stranded RNA directly. The two major loci, termed *esi-1* and *esi-2* (Fig. 3a and Supplementary Fig. 3), gave rise to half of the 20 most abundant endo-siRNAs in ovaries and also generated siRNAs in embryos, larvae and adults (not shown). *esi-1*, annotated as CG18854, can produce an ~400-base pair (bp) dsRNA through interaction of its 5′ and 3′ untranslated regions (UTRs; Supplementary Fig. 3). *esi-2* overlaps with CG4068 and consists of 20 palindromic ~260-nucleotide repeats (Fig. 3a). All siRNAs derived from these two loci arise from one genomic strand. In some previously characterized instances (for example, *Arabidopsis trans*-acting-siRNAs[21]) Dicer generates 'phased' siRNAs with 5′ ends showing a 21-nucleotide periodicity. In all tissues examined, *esi-1* and *esi-2* produced phased siRNAs, consistent with a defined initiation site for Dicer processing (Fig. 3a and Supplementary Fig. 3). Phasing was not observed for viral or repeat-derived siRNAs. Finally, siRNAs from both loci also joined AGO1 in proportions greater than siRNAs produced from transposons and repeats, perhaps owing to the imperfect nature of the dsRNA that they produce[22,23] (Fig. 1c).

AGO2 regulates gene expression by cleavage of complementary sites rather than by recognition of seed sites typical of AGO1–miRNA-mediated regulation[23]. We searched for possible targets of endo-siRNAs by identifying transcripts with substantial complementarity. A highly abundant siRNA from *esi-2* is highly complementary to the coding sequence of the DNA-damage-response gene *mutagen-sensitive 308* (*mus308*). Using a modified rapid amplification of cDNA ends (RACE) protocol, we detected *mus308* fragments with 5′ ends corresponding precisely to predicted endo-siRNA cleavage sites (Fig. 3b). Moreover, *AGO2* and *Dcr-2* loss consistently increased *mus308* expression in testis and to a lesser extent in ovaries, consistent with the relative abundance of *esi-2* siRNAs in these tissues (Fig. 3b, c). Finally, a reporter gene containing two *mus308* target sites was significantly derepressed in S2 cells on depletion of Dcr-2 or AGO2 but not of Dcr-1 or AGO1 (Fig. 3c). Although extensive complementarity between other endo-siRNAs and messenger RNAs was rare, we found several *esi-1*-derived siRNAs complementary to CG8289 (Supplementary Fig. 3), suggesting a potential regulatory interaction *in vivo*.

A second group of siRNA-generating loci contained regions in which dsRNAs can arise from convergent transcription. If sorted for siRNA density, most of the top 50 ovarian and S2 cell siRNA loci lay in regions where annotated 3′ UTRs or expressed-sequence-tags corresponding to convergently transcribed protein-coding genes overlap (Supplementary Tables 1 and 2). Typically, siRNAs arise on both genomic strands but only from overlapping portions of convergent transcripts (Fig. 3d). Examining all 998 convergently transcribed gene pairs in the *Drosophila* genome with annotated

overlapping transcripts, we found the peak abundance of ovarian siRNAs to be at the centre of the overlap, with sharp declines away from this region (Supplementary Fig. 4). In an alternative arrangement, *Pgant35A* produces sense and antisense siRNAs across its entire annotated transcript, consistent with expressed-sequence-tag support for antisense transcription traversing this locus (Supplementary Fig. 5).

Thus, a large number of *Drosophila* genes generate endogenous siRNAs, with most having perfect complementarity to the 3′ UTRs of neighbouring genes. Relative levels of endo-siRNAs generated from each convergent transcription unit were low (not shown), and we found no or little change (up to a ~1.3-fold increase) in the expression of such genes in *AGO2* mutant ovaries. Possibly, the level of small RNAs produced by this genomic arrangement is

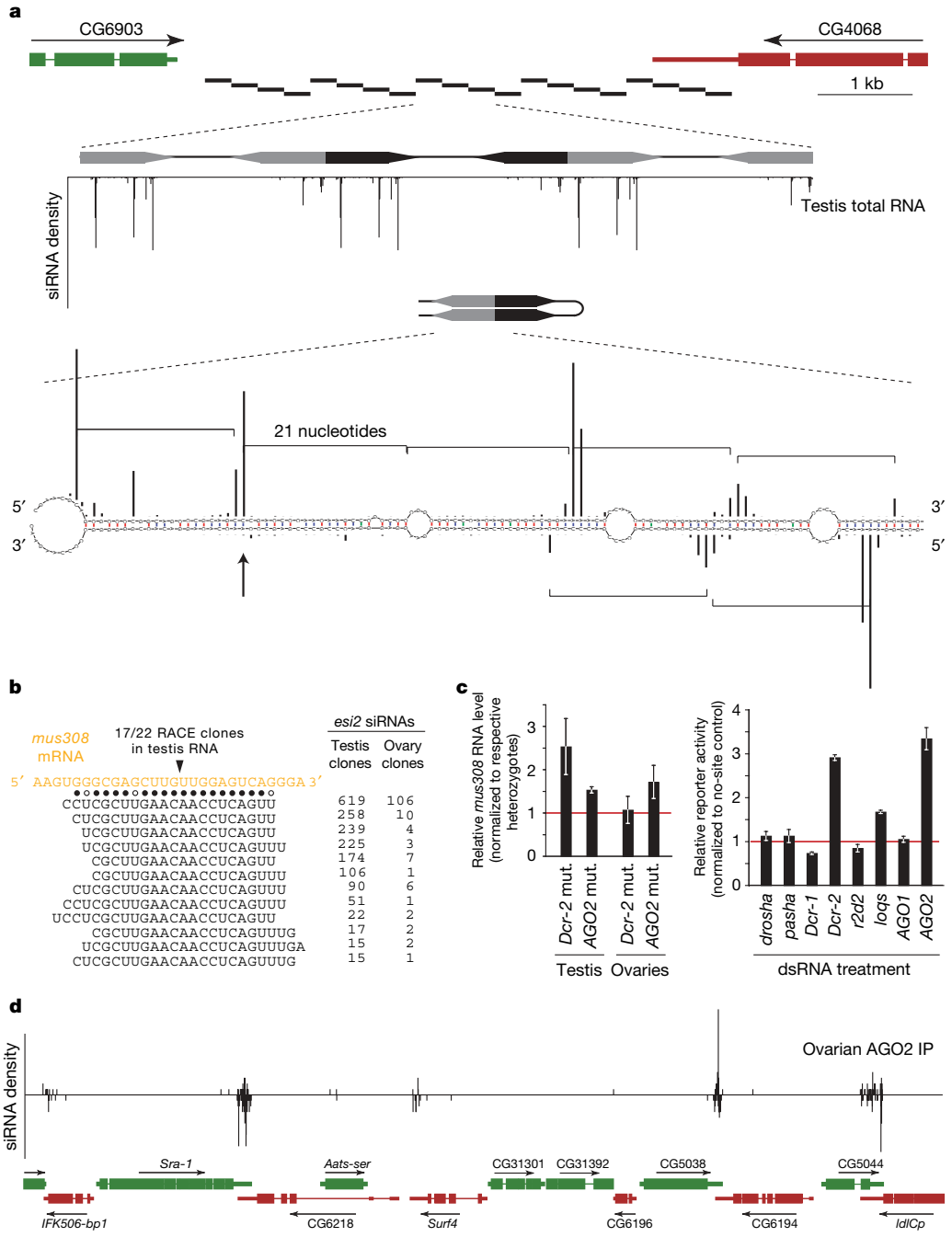


Figure 3 | Two types of genic endo-siRNA loci in *Drosophila*. **a**, FlyBase gene structure at the *esi-2* locus indicating the twenty ~260-bp repeats of *esi-2* (black bars). Below, three repeats are magnified and the small RNA density is depicted. Most siRNAs match up to 20 times and cannot be unambiguously assigned to a chromosomal location. Each repeat is a palindromic sequence, offering a multitude of possible structures. One structure is shown with the 5′ ends of cloned siRNAs as black bars (the height correlates with cloning frequency; phasing is indicated by brackets). **b**, An abundant siRNA from *esi-2* (indicated by an arrow in **a**) is highly complementary to the *mus308* mRNA. Shown is the *mus308* target site and cloned siRNAs (solid dots indicate canonical base pairs; open dots indicate GU base pairs). The only detected cleavage site within this duplex is indicated above. **c**, Shown are *mus308* transcript levels from *AGO2* and *Dcr-2* mutant (mut.) flies compared to their respective heterozygotes (error bars indicate standard deviation). To the right, average reporter levels (error bars indicate standard deviation) of a construct containing two *mus308* target sites in S2 cells depleted of the indicated genes by RNAi are shown. **d**, An example of siRNAs arising from convergent transcription units. A 30-kb region containing multiple instances of convergently transcribed genes is displayed, with the density of AGO2-associated RNAs in ovaries shown above.

inconsequential, amounting to noise within silencing pathways. However, there are probably circumstances wherein regulation by such arrangements might substantially impact expression.

In S2 cells, two neighbouring loci encoded nearly 16% of AGO2-associated RNAs (Supplementary Table 2). These reside within a large intron of *klarsicht* (Supplementary Fig. 6) and did not generate siRNAs in any other tissue. A similar locus, corresponding to CG14033, was found within an intron of *thickveins* (Supplementary Fig. 7) and gave rise to testis-specific siRNAs. Although the function of both siRNA clusters is unclear, the *thickveins* cluster shares considerable complementarity to CG9203, and loss of *AGO2* and *Dcr-2* mildly increased CG9203 mRNA levels in testis but not in ovaries (Supplementary Fig. 7).

Dcr-2 has been implicated in the production of siRNAs from viral replication intermediates or exogenously introduced dsRNAs, whereas Dcr-1 has been linked to miRNA biogenesis[6,16,17]. In agreement with these observations, all endo-siRNA classes were lost in *Dcr-2* mutant ovaries (Fig. 4a). To obtain more insight into the genetic requirements for endo-siRNA biogenesis and stability, we depleted components of siRNA and miRNA pathways in S2 cells and analysed levels of abundant siRNAs derived from structured loci (Fig. 4b and Supplementary Fig. 8). Although depletion of Dcr-2 and AGO2 resulted in substantial reductions in siRNA levels, little or no changes were observed on Drosha, Pasha, Dcr-1 or AGO1 depletion. Unexpectedly, we found virtually no requirement for the Dcr-2 partner R2D2 (ref. 14) but a strong requirement for the Dcr-1 partner Loquacious[1,4,5]. Only one analysed siRNA exhibited partial dependence on R2D2, potentially correlating with the extensive dsRNA character of its precursor duplex (Supplementary Fig. 9). Artificial sensors for endo-siRNAs from *esi-1* and *esi-2* in S2 cells gave patterns of de-repression that matched our analysis of endo-siRNA levels (Figs 3c and 4c).

Analysis of the most abundant siRNA from *esi-2* in flies mutant for *Dcr-2*, *AGO2*, *r2d2* or *loqs* extended our findings from cell culture (Supplementary Fig. 10). To examine the unexpected requirement for *loqs* more broadly, we sequenced small RNAs from *loqs*-mutant ovaries and observed a near complete loss of endo-siRNAs from structured loci (Fig. 4a). A much smaller impact of *loqs* was seen on endo-siRNAs derived from repeats and convergent transcription units. However, an involvement of Loqs and not R2D2 in the function of siRNAs derived from perfect dsRNA precursors was supported by analysing the impact of depleting siRNA/miRNA pathway components on the ability to suppress FHV replication in our infected S2 cell cultures (Supplementary Fig. 11).

Our results uncover an unanticipated role for Loqs in siRNA biogenesis and suggest that R2D2 has a lesser impact on at least two types of endogenous siRNAs. It is well established that Loqs partners with Dcr-1 for miRNA processing. To probe a molecular interaction with Dcr-2, we catalogued Loqs binding partners using quantitative proteomics. Dcr-1 and Dcr-2 were both abundant in Loqs immunoprecipitates from cultured cells and flies (Supplementary Fig. 12), supporting a physical interaction between Dcr-2 and Loqs.

Among animals, endo-siRNA pathways have so far been restricted to *Caenorhabditis elegans*[24–27]. Our results extend the prevalence of

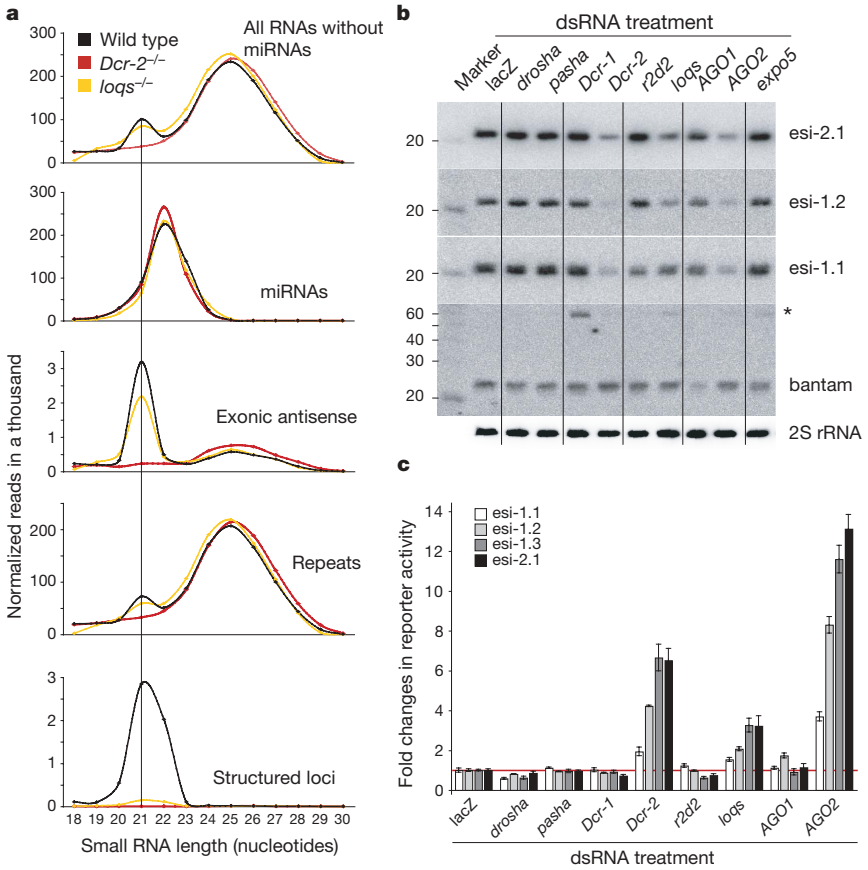


Figure 4 | Genetic requirements for siRNA biogenesis. **a**, Length distributions of small RNAs from total RNA libraries obtained from wild-type (black), *Dcr-2* mutant (red) and *loqs* mutant (yellow) ovaries. Excluding miRNAs and piRNAs (23–29 nucleotides), the population of 21-nucleotide siRNAs mapping to structured loci, genes and repeats is lost from *Dcr-2* mutants and those mapping to structured loci are strongly reduced in *loqs* mutants (all libraries are adjusted to the same total small RNA count). **b**, Northern blots showing levels of three siRNAs encoded from structured loci *esi-1* and *esi-2* (esi-2.1, esi-1.1 and esi-1.2) in S2 cells treated with dsRNA against the genes indicated. As controls, northern blots of *bantam* (pre-miRNA indicated by the asterisk) and 2S rRNA are shown below. **c**, *Renilla* luciferase reporter assays are shown for the siRNAs examined in **b** and an additional esi-1-derived species (esi-1.3) in S2 cells treated with dsRNAs against the indicated genes (error bars indicate standard deviation; $n = 3$).

such systems to *Drosophila* and parallel recent discoveries of an endo-siRNA pathway in mouse oocytes[28,29]. These systems have many common features but also key differences. In both, siRNAs collaborate with piRNAs to repress transposons. Also, mouse and *Drosophila* both generate endo-siRNAs from structured loci. In mouse, dsRNAs can form by pairing of sense protein-coding transcripts with antisense transcripts from pseudogenes. Whether or not transcripts from unlinked sites lead to siRNA production in *Drosophila* is unclear. However, transposon sense transcripts may hybridize to antisense sequences transcribed from piRNA clusters to form endo-siRNA precursors. In flies, a much larger number of genic loci enter the pathway as compared to mice because convergent transcription of neighbouring genes frequently creates overlapping transcripts. Overall, annotation of the *Drosophila* genome indicates that a significant proportion is transcribed in both orientations, providing widespread potential for dsRNA formation. This property is shared by many other annotated genomes, raising the possibility that the RNAi pathway has broad impacts on gene regulation. Viewed in combination, our studies suggest an evolutionarily widespread adoption of dsRNAs as regulatory molecules, a property previously ascribed only to miRNAs.

METHODS SUMMARY

The fly stocks used were *Dcr-2*^{L811Fsx} (ref. 6), *AGO2*^{414} (ref. 30), *loqs*^{f00791} (ref. 1) and *r2d2*^{1} (ref. 14). Recombineering was used to insert a Flag–haemagglutinin (HA) tag at the amino terminus of the *AGO2* coding sequence in the context of the genomic *AGO2* locus including flanking regulatory regions (for details, see Methods). Polyclonal anti-AGO1 antibody was obtained from Abcam (lot number 113754). Small RNAs for library production were isolated from ovarian total RNA or from Argonaute immunoprecipitates. Libraries were produced as described[12] and sequenced using the Illumina platform (protocol available on request). A description of the bioinformatics methods can be found online. For quantitative real-time PCR (qRT–PCR) analyses, we used total RNA preparations and random hexamer primers. Details and all primer sequences are given in the Supplementary Information. S2 cell knockdown treatments were for eight days with two sequential dsRNA soakings. For the reporter experiments, inducible expression plasmids for *Renilla* and firefly were transfected into S2 cells together with dsRNA for the desired knockdown target. Renilla constructs contained two target sites for endogenous siRNAs, whereas the firefly construct was used for normalization. For details on plasmids, dsRNAs and target sites, see Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 January; accepted 18 April 2008.
Published online 7 May 2008.

1. Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3,** e236 (2005).
2. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425,** 415–419 (2003).
3. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432,** 231–235 (2004).
4. Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3,** e235 (2005).
5. Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19,** 1674–1679 (2005).
6. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117,** 69–81 (2004).
7. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366 (2001).
8. Eulalio, A., Huntzinger, E. & Izaurralde, E. Getting to the root of miRNA-mediated gene silencing. *Cell* **132,** 9–14 (2008).
9. Bushati, N. & Cohen, S. M. microRNA functions. *Annu. Rev. Cell Dev. Biol.* **23,** 175–205 (2007).
10. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318,** 761–764 (2007).
11. Klattenhoff, C. & Theurkauf, W. Biogenesis and germline functions of piRNAs. *Development* **135,** 3–9 (2008).
12. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128,** 1089–1103 (2007).
13. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315,** 1587–1590 (2007).
14. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301,** 1921–1925 (2003).
15. Ruby, J. G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* **17,** 1850–1864 (2007).
16. Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J. A. & Imler, J. L. Essential function *in vivo* for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nature Immunol.* **7,** 590–597 (2006).
17. Wang, X. H. *et al.* RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312,** 452–454 (2006).
18. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20,** 2214–2222 (2006).
19. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313,** 320–324 (2006).
20. Rehwinkel, J. *et al.* Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* **26,** 2965–2975 (2006).
21. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* **121,** 207–221 (2005).
22. Tomari, Y., Du, T. & Zamore, P. D. Sorting of *Drosophila* small silencing RNAs. *Cell* **130,** 299–308 (2007).
23. Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by Dicer-1. *Cell* **130,** 287–297 (2007).
24. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127,** 1193–1207 (2006).
25. Sijen, T., Steiner, F. A., Thijssen, K. L. & Plasterk, R. H. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* **315,** 244–247 (2007).
26. Pak, J. & Fire, A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* **315,** 241–244 (2007).
27. Yigit, E. *et al.* Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127,** 747–757 (2006).
28. Tam, O. H. *et al.* Pseudogene-derived siRNAs regulate gene expression in mouse oocytes. *Nature* advance online publication, doi:10.1038/nature06904 (10 April 2008).
29. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* advance online publication, doi:10.1038/nature06908 (10 April 2008).
30. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* **18,** 1655–1666 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Carthew, H. Siomi, P. Zamore and D. Smith for reagents. We are grateful to M. Rooks, E. Hodges and D. McCombie for help with deep sequencing. B.C. was supported by the German Academic Exchange Service. C.D.M. is a Beckman fellow of the Watson School of Biological Sciences and is supported by an NSF Graduate Research Fellowship. R.Z. is a Special Fellow of the Leukemia and Lymphoma Society. M.D. is an Engelhorn fellow of the Watson School of Biological Sciences. J.B. is supported by the Ernst Schering foundation. A.S. is supported by an HFSP fellowship. This work was supported in part from grants from the NIH to G.J.H. and N.P. and a gift from K. W. Davis (G.J.H.).

Author Information Small RNA sequences were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE11086. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.edu) or J.B. (brenneck@cshl.edu).

METHODS

Fly stocks. A 3× Flag–HA tag was inserted at the N terminus of the Flybase RB transcript of *AGO2* into BAC RP98-21A13 by means of bacterial red/ET recombination (Gene Bridges GmbH). A 13.9-kb AvrII/XhoI fragment of the modified BAC encompassing the *AGO2* locus including parts of flanking genes (chromosome 3L coordinates: 15,544,405–15,558,309) was cloned into pCasper4 (XbaI/XhoI). Transgenic flies were generated at Bestgene Inc. Expression of tagged AGO2 in embryos, ovaries and whole flies was verified in multiple lines by western blotting using a monoclonal anti-HA-peroxidase antibody (1:500; catalogue number 12013819001, Roche). *Dcr-2*^{L811Fsx} flies were a gift of R. Carthew[6], *AGO2*^{414} flies were a gift of H. Siomi[30], *loqs*^{f00791} flies were a gift of P. Zamore[1] and *r2d2*^{1} flies were a gift of D. Smith[14]. For wild-type fly stocks, stock number 2057 from Bloomington (Celera sequencing strain) and OregonR flies were used.

Small RNA libraries. Twenty-six 10-cm plates with 50–70% confluent Schneider cells were transfected with pCasper_Flag-HA-AGO2 using calcium phosphate, harvested after 36 h and lysed in buffer A (20 mM HEPES, pH 7.0, 150 mM NaCl, 2.5 mM $MgCl_2$, 0.3% Triton, 30% glycerol) supplemented with 1 mM PMSF and protease inhibitors (Complete, Roche). Cleared extract was split and incubated with rabbit polyclonal AGO1 antibody (1:20; lot number 113754, Abcam) or mouse anti-Flag M2-agarose (1:25, Sigma) for 4 h at 4 °C. AGO1 antibodies were isolated by adding protein G beads (1:10; Roche) for 1 h. Beads were washed six times each for 10 min in buffer B (30 mM HEPES, pH 7.4, 800 mM NaCl, 2 mM $MgCl_2$, 0.1% NP-40), which contained equal supplements as buffer A. The immunoprecipitation was analysed by western blotting using anti-AGO1 (1:2,000; Abcam) and anti-HA-peroxidase (1:500; catalogue number 12013819001, Roche). AGO1- and AGO2-associated RNAs were isolated with phenol/chloroform and ethanol precipitated. For the ovarian AGO2 IP library, ~500 mg ovaries from transgenic Flag-HA-AGO2 flies were dissected and lysed mechanically in buffer A. AGO2 complexes and associated RNAs were purified as above.

AGO1- and AGO2-bound small RNAs as well as small RNAs from total RNA were cloned as described[12] (the detailed protocol is available on request). The following small RNA libraries from total RNA were prepared for this study: 18–28 nucleotides from ovaries of the Celera sequenced strain (Bloomington, number 2057); 18–28 nucleotides from ovaries of *Dcr-2*^{L811Fsx} homozygous flies; 18–28 nucleotides from *loqs*^{f00791} homozygous flies; and 18–24 nucleotides from the testis of OregonR flies.

Libraries were sequenced in house using the Illumina platform. Published libraries used in this study were a 16–26-nucleotide S2 cell total RNA library[15] and libraries from ovarian Piwi/Aub/AGO3 immunoprecipitates[12].

Bioinformatic analysis of small RNA libraries. Small RNA sequences were matched to the *Drosophila* release 5 genome and to genomes of *Drosophila* C virus, FHV and cricket paralysis virus. Only reads matching the fly genome 100% and viral genomes with up to three mismatches were used for further analysis. For annotations, we used Flybase for protein-coding genes, UCSC for non-coding RNAs and transposons/repeats (http://genome.ucsc.edu/) and the most recent miRNA catalogue[15,31].

siRNA clusters were extracted by mapping all 20–22-nucleotide-long RNAs from the AGO2-IP libraries to the genome (only uniquely mapping RNAs were used) and retaining 200-nucleotide windows that contained at least three distinct small RNAs. Windows separated by a maximum of 200 nucleotides were fused and those with more than 40 unique reads were sorted after the density of siRNAs per base pair.

For the transposon analysis, 20–22-nucleotide AGO2-bound RNAs from ovaries and S2 cells were mapped onto the Repbase collection of transposons[32] with up to three mismatches to construct heatmaps indicating cloning frequency and strand bias of siRNAs. For the latter analysis, only siRNAs unambiguously mapped to one strand were considered.

Cleavage site mapping for endo-siRNA targets. Wild-type testes were dissected on ice into 1× PBS. Total RNA was isolated using Trizol (Invitrogen) according to the manufacturer's protocol. Total RNA (5 μg) was used as starting material. Ligation of an RNA adaptor, reverse transcription using the GeneRacer oligo(dT) primer and 5′ RACE–PCR were performed according to the manufacturer's instructions (GeneRacer kit, Invitrogen). 5′ RACE–PCR was carried out using the GeneRacer 5′ primer (5′-CGACTGGAGCACGAGGACACTGA-3′) and a *mus308* gene-specific reverse primer (5′-TGCTTTGCAGAGTCGAA-GCTGATTG-3′), and followed by one round of nested PCR using the GeneRacer 5′ nested primer (5′-GGACACTGACATGGACTGAAGGAGTA-3′) and a nested primer specific to *mus308* (5′-CCGCTAGCTCTACCAAACTGGTGAT-3′). PCR products were gel purified and cloned into pCR4Blunt-TOPO (Invitrogen). Twenty-two clones were sequenced with T7 (5′-GTAATACGA-CTCACTATAGGGC-3′) and T3 (5′-AATTAACCCTCACTAAAGGG-3′) primers, and subjected to further analysis.

dsRNA treatment of Schneider cells. Approximately 3×10^6 S2-NP cells were soaked in 1.5 ml serum-free Schneider's medium containing 10 μg dsRNAs in 6-well plates, and 3 ml serum-containing medium was added 45 min later. After 4 days of initial dsRNA treatment, cells were treated with a second round of dsRNAs using the same procedure, and were harvested another 4 days later. Total RNA was extracted with Trizol (Invitrogen). Sequences of the primers for generating dsRNAs are listed below.

siRNA reporter constructs. A SalI/BglII fragment from pGL3-Basic (Promega) was cloned to pRmHa-3 using SalI/BamHI (pMT-Firefly-long). The coding region of the *Renilla* luciferase gene was amplified by PCR and cloned into pRmHa-3 using BamHI/EcoRI sites (pMT-Renilla). A pair of oligonucleotides containing two perfect binding sites for si1_1, si1_2 or si2 were annealed and cloned into pMT-Renilla (BamHI/SalI) to generate sensor constructs (Supplementary Fig. 13).

Transfection was performed in a 384-well plate format. For each well, ~100 ng plasmid DNA (5 ng pMT-Renilla, 20 ng pMT-Renilla-si1_1, 50 ng pMT-si1_2 or 100 ng pMT-si2, 5 ng pMT-Firefly-long, and corresponding amounts of pRmHa-3 serving as carrier DNA) and ~80 ng dsRNA were mixed with 0.8 μl enhancer in 15 μl EC (Qiagen) and the mixture was incubated at room temperature (23 °C) for 5 min. After this, 0.35 μl of Effectene reagent was added and the mixture was immediately dispensed into each well containing dsRNA. After incubation at room temperature for 10 min, 40 μl S2-NP cells (10^6 cells ml^{-1}) were dispensed into the well. Cells were induced with 200 μM $CuSO_4$ 132 h post transfection, and luciferase assays were performed 36 h later using DualGlo reagents (Promega). For each well, the reporter activity was calculated as the ratio of *Renilla* luciferase to firefly luciferase. Each data point was normalized against the data points where dsRNA against LacZ was transfected. Presented are average results with standard deviation ($n = 3$).

Northern blotting. Total RNA was isolated using Trizol (Invitrogen). RNA (30 μg) was separated on a 15% denaturing polyacrylamide gel and transferred onto a Hybond-N+ membrane (Amersham Biosciences) in 1× TBE (Tris-Borate-EDTA) buffer. The RNA was crosslinked using ultraviolet light (Stratalinker) to the membrane and pre-hybridized in ULTRAhyb buffer (Ambion) for 1 h. DNA probes complementary to the indicated endo-siRNAs, bantam and 2S rRNA were 5′ radiolabelled and added to the hybridization buffer (hybridization overnight at 37 °C). Membranes were washed 4–6 times in 1× SSC with 0.1% SDS at 37 °C and exposed to PhosphoImager screens. Probes were stripped by boiling the membrane twice in 0.2× SSC containing 0.1% SDS in a microwave.

Quantitative real-time PCR. Ovaries and testis from homozygous or heterozygous flies were dissected on ice into 1× PBS. Total RNA of dissected tissues or S2 cells was extracted using Trizol (Invitrogen). RNA was treated with DNase I Amplification Grade (Invitrogen) according to the manufacturer's instructions. Complementary DNA was prepared by reverse transcription using SuperScript III Reverse Transcriptase (Invitrogen) and random hexamer primer. qRT–PCR was carried out using SYBR GREEN PCR Master Mix (Applied Biosystems) and a Chromo4 Real-Time PCR Detector (BioRad). C_t values were calculated within the log-linear phase of the amplification curve using the Opticon Monitor 3.1.32 software (BioRad). Quantification was normalized to the mRNA coding for the endogenous ribosomal protein rp49, and relative expression levels were calculated using the following equation: $A = 1.8^{[Ct(ref) - Ct(ref\text{-}control)] - [Ct(sample) - Ct(sample\text{-}control)]}$. Transposon analysis was carried out with four biological replicates (individually shown; error bars indicate technical replicates); siRNA target analysis was carried out with three biological replicates. Oligonucleotide primers used in this study are listed below.

DNA oligonucleotides. PCR primers for the generation of dsRNAs were as follows: T7-Dicer-1-F-14 TAATACGACTCACTATAGGGTGCGACAACAA-TCTGC; T7-Dicer-1-R-565 TAATACGACTCACTATAGGGTCAGTTGCTG-CAGCTCAC; T7-Dicer-2-F-5 TAATACGACTCACTATAGGGAAGATGTGG-AAATCAAGCC; T7-Dicer-2-R-555 TAATACGACTCACTATAGGGCCACG-TTCGTAATTTC; T7-Drosha-F-3356 TAATACGACTCACTATAGGGTGAA-TCAGGACTGGAACG; T7-Drosha-R-3910 TAATACGACTCACTATAGGG-AGCCATCGCTATCACTGC; T7-Exportin5-F-55 TAATACGACTCACTATA-GGGATCTAGTCATGAACCCG; T7-Exportin5-R-623 TAATACGACTCACT-ATAGGGAACGCAGTCACATGCTGC; T7-AGO1-F1225 TAATACGACTC-ACTATAGGGAACGGACAGACCGTAGAG; T7-AGO1-R1858 TAATACGAC-TCACTATAGGGTGGCGTACTTACAGAAGC; T7-AGO2-F2211 TAATACG-ACTCACTATAGGGAGCCACATCGACGAACG; T7-AGO2-R2855 TAATAC-GACTCACTATAGGGCGAGGATCATCCTTGATC; T7-R2D2-PZ-F TAAT-ACGACTCACTATAGGGCATACACGGCTTGATGAAGGATTC; T7-R2D2-PZ-R TAATACGACTCACTATAGGGTTGCTTGTGCTCGCTACTTGC; T7-Pasha-F452 TAATACGACTCACTATAGGGACTTTGAAGTCCTACCCG; T7-Pasha-R1177 TAATACGACTCACTATAGGGCTCCTTGAACTCATAGG; T7-Loqs-F-1 TAATACGACTCACTATAGGGATGGACCAGGAGAATTTCC; T7-

Loqs-R-540 TAATACGACTCACTATAGGGAAGGGCGTATCCTTGTC; T7-LacZ-F TAATACGACTCACTATAGGGCATTATCCGAACCATCC; T7-LacZ-R TAATACGACTCACTATAGGGCAGAACTGGCGATCGTTCG. endo-siRNA sensor oligonucleotides were as follows: BamHI-esi-1_2-S2-2P-F, GATCCCAACAGTTTATTTACTTGGAGGCAACATAATCAAATGAACTGAGGGTTACTTGGAGGCAACATAATCAG; SalI-esi-1_2-S2-2P-R, TCGACTGATTATGTTGCCTCCAAGTAACCCTCAGTTCATTTGATTATGTTGCCTCCAAGTAAATAAACTGTTGG; BamHI-esi-2_1-S2-2P-F, GATCCCAACAGTTTATTGGAGCGAACTTGTTGGAGTCAAAATGAACTGAGGGTGGAGCGAACTTGTTGGAGTCAAG; SalI-esi-2_1-S2-2P-R, TCGACTTGACTCCAACAAGTTCGCTCCACCCTCAGTTCATTTTGACTCCAACAAGTTCGCTCCAATAAACTGTTGG; BamHI-esi-1_3-S2-2P-F, GATCCCAACAGTTTATTCATTTGATCCATAGTTTCCCGAATGAACTGAGGGTCATTTGATCCATAGTTTCCCGG; SalI-esi-1_3-S2-2P-R, TCGACCGGGAAACTATGGATCAAATGACCCTCAGTTCATTCGGGAAACTATGGATCAAATGAATAAACTGTTGG; BamHI-esi-1_1_ -S2-2P-F, GATCCCAACAGTTTATTGCCAAGGTACGTGGTCGACCGAAATGAACTGAGGGTGCCAAGGTACGTGGTCGACCGAG; SalI-esi-1_1_BC36-S2-2P-R, TCGACTCGGTCGACCACGTACCTTGGCACCCTCAGTTCATTTCGGTCGACCACGTACCTTGGCAATAAACTGTTGG; BamHI-mus308_target-S2-2P-F, GATCCCAACAGTTTATTGGGCGAGCTTGTTGGAGTCAGAATGAACTGAGGGTGGGCGAGCTTGTTGGAGTCAGG; SalI-mus308_target-S2-2P-R, TCGACCTGACTCCAACAAGCTCGCCCACCCTCAGTTCATTCTGACTCCAACAAGCTCGCCCAATAAACTGTTGG.

Northern probes were as follows: esi-2.1, GGAGCGAACTTGTTGGAGTCAA; esi-1.1, GCCAAGGTACGTGGTCGACCGA; esi-1.2, CATTTGATCCATAGTTTCCCG; miR-bantam, AATCAGCTTTCAAAATGATCTCA; 2S rRNA, TACAACCCTCAACCATATGTAGTCCAAGCA.

Quantitative real-time PCR. For transposon analysis: rp49_F, ATGACCATCCGCCCAGCATAC; rp49_R, CTGCATGAGCAGGACCTCCAG; GAPDH2_F, TGATGAAATTAAGGCCAAGGTTCAGGA; GAPDH2_R, TCGTTGTCGTACCAAGAGATCAGCTTC; actin5c_F, AAGTTGCTGCTCTGGTTGTCG; actin5c_R, GCCACACGCAGCTCATTGTAG; BEL1_F, ATTATACAAACGCCCAATTGCCAAAA; BEL1_R, TCCGATGAAGCTGCAGACAAATAAGA; BLOOD_F, AGACGTTCATTACAGATCAAGGTACGGA; BLOOD_R, AGTTCGTATGGGCAATAGTCATGGACT; DM412_F, AAAGTACGGTCCAATGAAGACG; DM412_R, GTGGTGATGAGCTGTTGATGTT; F-element_F, TTGTTGAACAGCATACCACTCC; F-element_R, CCAGAGTTGATGAGCCAGTGTA; gypsy6_F, GACAAGGGCATAACCGATACTGTGGA; gypsy6_R, AATGATTCTGTTCCGGACTTCCGTCT; MDG1_F, AACAGAAACGCCAGCAACAGC; MDG1_R, CGTTCCCATGTCCGTTGTGAT; ROO_F, CGTCTGCAATGTACTGGCTCT; ROO_R, CGGCACTCCACTAACTTCTCC; stalker4_F, TTTGGAAGATTACCAAGGCAGTTCGC; stalker4_R, GGATCTAACTTATGACCCGATTCGTTCC; ZAM_F, ACTTGACCTGGATACACTCACAAC; ZAM_R, GAGTATTACGGCGACTAGGGATAC; FHV_F, CCCTGGAGTCGCTTACTTGAGTGCT; FHV_R, ATGGAAGCGTACCTGAAGGAGGACA; DOC_F, TACCTTAAACAAACAAACATGCCCACC; DOC_R, TTTGTATGGGTGGTCAGCTTTTCGT; DM297_F, GCCAGTACACACGAACGAAATA; DM297_R, AATTGAATTTTGGCAATTTTGG; Tabor_F, GAGCAAGAATTATGCTCGAAGAA; Tabor_R, AATTTATGTCCGGTTTCGTTTTT.

For endo-siRNA target analysis: *mus308*_F, AAGGATTAGCGCCAAGCTGGAGGAT; *mus308*_R, ACCACGACCACTGCCACAGAGATTC; CG9203_F, AGCTGGCAGAAAAACCATGACCAGT; CG9203_R, CAATTCTTTTGGCGTAGCTTGAGCA.

For S2 cell knockdown analysis: S2-Dcr-1_F, ACGCCTTCCATCTCCCAGTTTTACC; S2-Dcr-1_R, GCCACCCTGCTTATTCTGACTGCTC; S2-Dcr-2_F, AAACGAGAGATTCGTGCCCAAAACA; S2-Dcr-2_R, CTGTCCTTGCTCTTATCGGCCTTGT; S2-Drosha_F, AGATGCCAGAGAACTTCACCATCCA; S2-Drosha_R, GAAAGAAGTGAAAAGCTGGGCAGGA; S2-Pasha_F, TGTCAAGGACAAGATAACGGGCAACA; S2-Pasha_R, GTTGGGAGATGGCTCCGTCGTCT; S2-AGO1_F, ACTACCACGTTCTGTGGGACGACAA; S2-AGO1_R, GAATCGTGCTCCTTCTCCACCAGAT; S2-AGO2_F, AACCCTCAAAAGTAAATCATGGGAAA; S2-AGO2_R, ATTTTTGCTGTTGGCCTCCTTG; S2-Loqs_F, GTGTGTGCGTCTGGATTTTGCTGTA; S2-Loqs_R, GTTTTCGGGAGGATTCGGTGTGTAT; S2-R2D2_F, GCGAAGACGGAGGGTACGTCTGTAA; S2-R2D2_R, AGTCGAATCCTTCATCAAGCCGTGT.

31. Stark, A. *et al.* Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res.* **17,** 1865–1879 (2007).
32. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110,** 462–467 (2005).