Current Biology 18, 147–151, January 22, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2007.12.049

Report

Argonaute Loading Improves the 5′ Precision of Both MicroRNAs and Their miRNA* Strands in Flies

Hervé Seitz,[1] Megha Ghildiyal,[1] and Phillip D. Zamore[1,*]
[1]Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, Massachusetts 01605

Summary

MicroRNAs (miRNAs) are short regulatory RNAs that direct repression of their mRNA targets. The miRNA "seed"—nucleotides 2–7—establishes target specificity by mediating target binding [1–5]. Accurate processing of the miRNA 5′ end is thought to be under strong selective pressure [6, 7] because a shift by just one nucleotide in the 5′ end of a miRNA alters its seed sequence, redefining its repertoire of targets (Figure 1). Animal miRNAs are produced by the sequential cleavage of partially double-stranded precursors by the RNase III endonucleases Drosha and Dicer, thereby generating a transitory double-stranded intermediate comprising the miRNA paired to its partially complementary miRNA* strand [8, 9]. Here, we report that in flies, the 5′ ends of miRNAs and miRNA* strands are typically more precisely defined than their 3′ ends. Surprisingly, the precision of the 5′ ends of both miRNA and miRNA* sequences increases after Argonaute2 (Ago2) loading. Our data imply that either many miRNA* sequences are under evolutionary pressure to maintain their seed sequences—that is, they have targets—or that secondary constraints, such as the sequence requirements for loading small RNAs into functional Argonaute complexes, narrow the range of miRNA and miRNA* 5′ ends that accumulate in flies.

Results and Discussion

We used high-throughput pyrosequencing of 18–30 nt RNAs to identify miRNAs expressed in *Drosophila melanogaster* heads and in cultured *Drosophila* S2 cells. Among the 120,896 miRNA reads (66,377 from fly heads; 54,519 from S2 cells), we observed two sources of heterogeneity for the ends of fly miRNAs: the addition of nucleotides not present in the gene from which the miRNA is transcribed (nontemplated nucleotides) and inaccurate or alternative cleavage by Drosha or Dicer. Approximately 5% of the reads for a typical miRNA contained nontemplated nucleotides on at least one end (Figure 2A and Figure S1 available online), most frequently the addition of a single uridine or adenosine to the 3′ end, but longer extensions were also observed, both on the 5′ and the 3′ ends (Table S1). Interestingly, longer extensions were also U- and A-rich at the 3′ end, whereas at the 5′ end, the 3′-most nontemplated nucleotide was frequently a cytidine, and other added nucleotides were typically uridines. This observation could prove to be useful for the identification of the 5′-elongating enzymatic activity. The nontemplated addition of nucleotides, especially uridines, to the 3′ ends of miRNAs has been reported previously in wild-type *Caenorhabditis elegans* [6] and *hen1* mutant *Arabidopsis thaliana* [10]. Overall, the addition of nontemplated nucleotides to the 5′ end of miRNAs was rarer (~1%; Figure 2A and Table S1).

We also observed a second, more frequent type of heterogeneity: variability in the position of the miRNA 5′ and 3′ ends within the sequence of the miRNA precursors (Figure 2B). Nontemplated nucleotides fortuitously matching the templated sequence are predicted to occur much less often than the heterogeneity we observe (Table S2). Similar terminal heterogeneity has been noted for the 3′ ends of *C. elegans* [6] and the 5′ and 3′ ends of mouse [11] miRNAs. The aberrant miRNA termini we observe are likely to reflect imprecision in precursor cleavage by Drosha and Dicer. They are unlikely to correspond to degradation products because we recorded nearly as many miRNA reads that were longer than the dominant species as were shorter (Figure S2) and because 93% (S2 cells) and 99% (fly heads) of sequences of the fly-specific 30 nt 2S ribosomal RNA (rRNA)—whose termini are expected to be single-stranded—were full length (Supplemental Discussion). 3′ degradation was slightly more common than 5′ degradation: We detected 3′ degradation for 1010 reads versus 5′ degradation for 201 reads among the 33,505 total 2S rRNA reads from S2 cells and fly heads combined; five reads corresponded to 2S rRNA trimmed from both ends.

The 5′ ends of miRNAs were more precisely defined than their 3′ ends, irrespective of whether the miRNA originated from the 5′ or 3′ arm of the pre-miRNA (Figure 3A). Thus, the difference in cleavage accuracy between the 5′ and 3′ ends cannot be attributed to an intrinsic difference in fidelity between Drosha and Dcr-1. We expected that the 3′ ends of miRNA* strands would be precisely defined because they are created by the pair of cuts that generates the 5′ ends of miRNA and that the 5′ ends of miRNA* strands would be imprecisely determined because they are created by the pair of cleavages that generates the highly heterogeneous 3′ ends of miRNA. Instead, we found that the 5′ end of a strand (for example, the miRNA) was more accurate than the 3′ end of the adjacent strand (in this example, the miRNA*; Figure 3B); these two extremities are produced by a pair of cuts catalyzed by the same enzyme.

Current dogma holds that the local sequence or structure of miRNA precursors is under strong selective pressure to generate accurate 5′ ends because a precise miRNA 5′ end directly establishes the seed sequence and hence the targets of the miRNA. Because we observe that, in flies, the 5′ ends of both the miRNA and the miRNA* are more precisely determined than the 3′ ends of either strand, this explanation implies that miRNA* sequences are under selective pressure to establish a unique seed sequence, implying that they, too, have regulatory targets.

It is also possible that both Drosha and Dcr-1—whose active sites are homologous—may also be intrinsically more precise in 5′ cleavage than in 3′ cutting. A third alternative is that 5′ and 3′ ends might be generated with similar, imperfect accuracy, but subsequent constraints in RISC loading or stability select for those small RNAs that begin with

*Correspondence: phillip.zamore@umassmed.edu

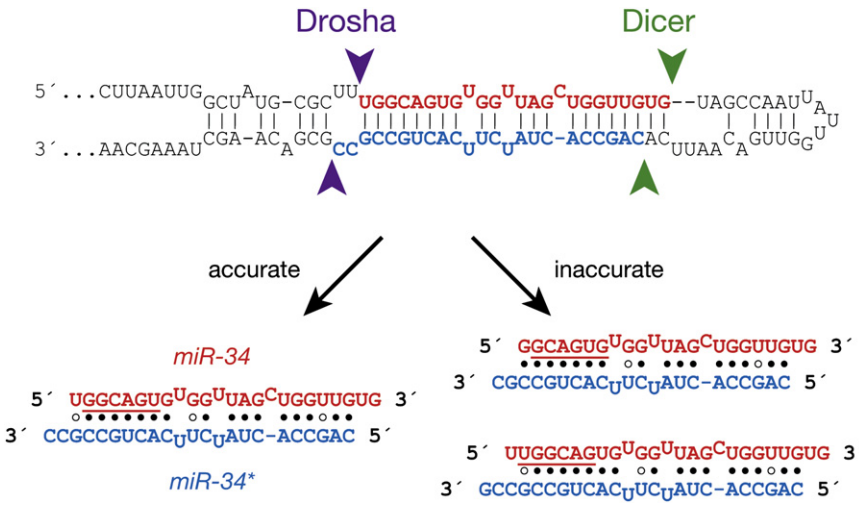

Figure 1. Inaccurate Processing of the 5′ End of a miRNA Alters its Seed Sequence

miRNA precursors are cleaved by two RNase III enzymes, Drosha and Dicer, liberating a short duplex: In this duplex, the mature miRNA (red) is paired to a partially complementary small RNA, the miRNA* (blue), derived from the opposite arm of the pre-miRNA stem. Inaccurate cleavage of the miRNA 5′ end changes its seed sequence (underlined).

a particular nucleotide or sequence. The subsequent destruction of miRNAs without these 5′ features would increase the apparent accuracy of miRNA 5′ ends and retain miRNA 3′ heterogeneity. To test this idea, we separately sequenced small RNAs containing modified 3′ termini (Table S3). In flies, the 3′ termini of small RNAs that are loaded into Ago2 [12], but not those bound to Argonaute1 [13], are 2′-*O*-methylated by *Drosophila* Hen1 as the last step in Ago2-RISC maturation [14]. To sequence small RNAs bearing 2′-*O*-methylated 3′ ends, we treated the total small RNA with $NaIO_4$ followed by β-elimination; this method blocks ligation of adapters to small RNAs bearing 2′,3′ hydroxy termini, preventing them from being sequenced.

To determine whether the greater accuracy of miRNA and miRNA* 5′ versus 3′ ends reflects the constraints of RISC assembly or stability, rather than more accurate 5′ versus 3′ cleavage by Drosha and Dicer, we compared the terminal heterogeneity of miRNA and miRNA* reads from the 3′ modified population to the heterogeneity of the total miRNA and miRNA* population. As a control, we compared the 3′ heterogeneity between the two populations. For both analyses, we only considered miRNA or miRNA* strands displaying some heterogeneity in the total population. For both fly heads and S2 cells, we observed a dramatic increase in the precision of the 5′—but not the 3′—ends of miRNAs and miRNA* strands upon loading into Ago2 (Figure 4). We also performed the analysis for those small RNAs that both had heterogeneous 5′ termini and were specifically enriched in the β-eliminated sequences relative to the non-β-eliminated set. For the 13 small RNAs (four miRNAs and nine miRNA*s) meeting these criteria, the 5′ ends in the subpopulation of miRNA and miRNA* sequences loaded into Ago2—i.e., those that were 2′-*O*-methylated—were again more precisely defined than the 5′ ends of the

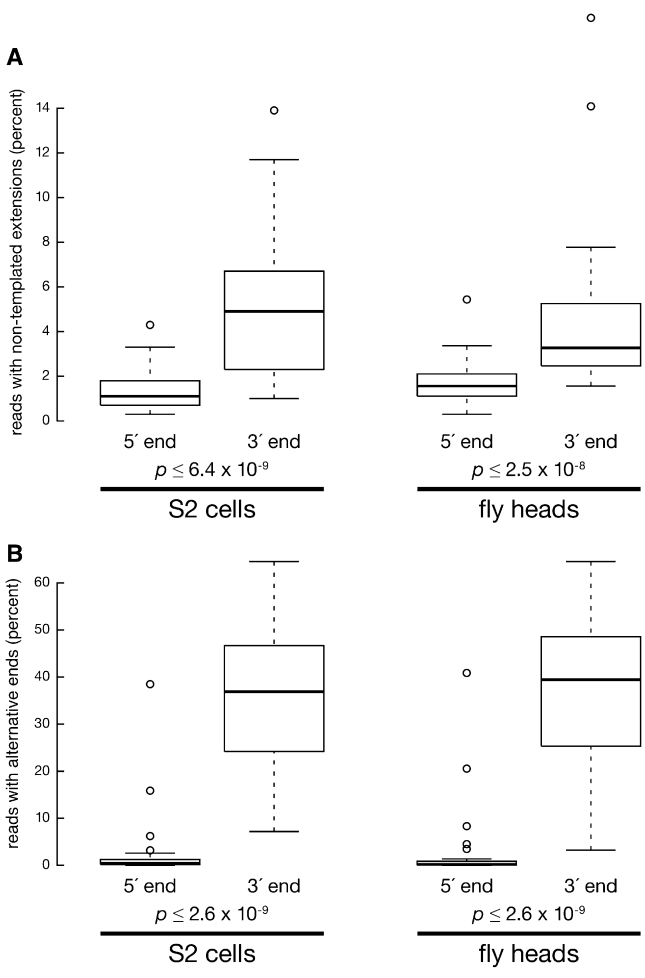

Figure 2. Cleavage Inaccuracies Are More Frequent than Nontemplated Additions

(A) The percentage of reads with nontemplated 5′ or 3′ extensions was evaluated for each miRNA whose sequence was read at least 100 times.

(B) The most abundant 5′ and 3′ ends were identified for each miRNA, and all other ends corresponding to the sequence of the primary miRNA transcript were flagged as "alternative." The percentage of reads with alternative ends was then determined for each miRNA read at least 100 times. Note the difference in the y axis scales in (A) and (B). Box plots follow Tukey's standard conventions: A rectangle encloses all data from the first to the third quartiles, a bold horizontal line reports the median, whiskers connected to the rectangle indicate the largest and smallest nonoutlier data, and outliers (values distant from the box by more than 1.5× the interquartile range) are displayed as open circles.

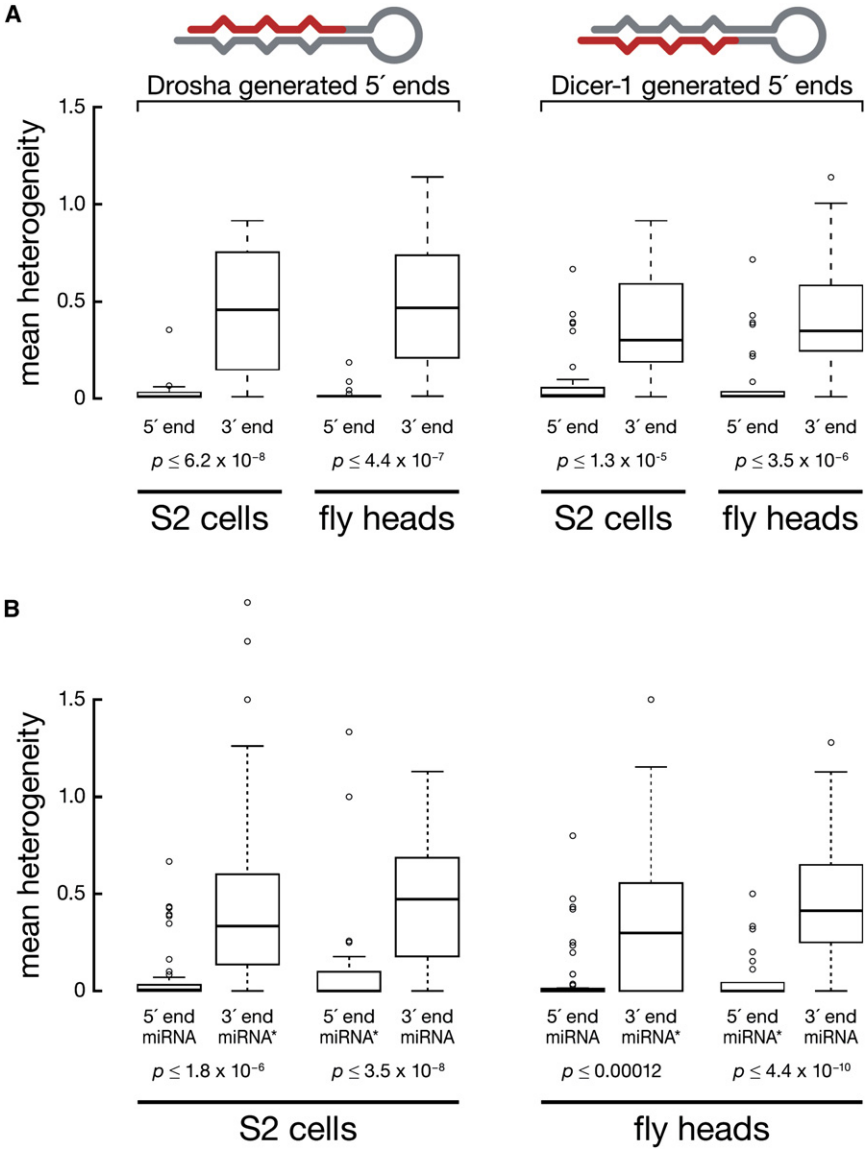

Figure 3. miRNA and miRNA* 5′ Ends Are More Precisely Defined than Their 3′ Ends

(A) miRNAs originating from the 5′ (left panels) or 3′ (right panels) arm of their pre-miRNAs were analyzed separately. For each miRNA, the heterogeneity of its termini was calculated as the mean of the absolute values of the distance between the 5′ or 3′ extremity of an individual templated read and the most abundant 5′ or 3′ end for that miRNA. Sequences read from RNA isolated from fly heads and cultured S2 cells were analyzed separately.

(B) Box plots show the distribution of mean heterogeneity for the 5′ and 3′ ends of miRNA and miRNA* sequences.

same small RNA sequences in the total small RNA population (Figure S3). We conclude that loading or stabilization of miRNAs in Ago2, and perhaps Argonaute proteins in general, imposes a purifying selection on their 5′ ends.

The mechanism responsible for the homogenization of 5′ ends remains to be determined. We can imagine that the efficiency of Argonaute loading is affected by the nature of the 5′ end of a small RNA, much as the stability of its pairing to the other strand influences this process [15]. The 5′ sequence itself may also play a role in RISC assembly, with some miRNA variants loaded more efficiently than others, according to the identity of their 5′ nucleotide(s). Alternatively, some Argonaute complexes might be selectively stabilized after their assembly, for example, by the presence of a target RNA whose binding stabilizes those RISCs containing miRNA isoforms with a complementary seed sequence.

Experimental Procedures

Biological Sources

We isolated fly heads by vigorously shaking liquid nitrogen-frozen flies expressing a long double-stranded hairpin RNA corresponding to *white* [16, 17] in nested, prechilled sieves (U.S.A. standard sieve, Humboldt MFG, Chicago, IL), allowing the heads to pass through the top sieve (No. 25), and collecting them on the bottom sieve (No. 40). S2 cell RNA was prepared from a clonal line containing the stably integrated GFP transgene (pKF63) and was transiently transfected with a double-stranded RNA against GFP [18].

RNA Preparation

A amount of 400 μg total RNA was extracted with the mirVana kit (Ambion), then 18-to-30 nt long RNAs were gel purified. 2S rRNA was depleted by hybridization to immobilized DNA oligonucleotide (5′-biotin-TCA ATG TCG ATA CAA CCC TCA ACC ATA TGT AGT CCA AGC A-3′). A total of 1.6 nmol of the biotinylated oligonucleotide was bound to 32 mg M270 Streptabeads (Dynal, Norway) in 3.2 ml 0.5× SSC for 30 min on ice, and then the beads were washed with ice-cold 0.5× SSC, resuspended in 8 ml 0.5× SSC, and incubated 5 min at 65°C. Gel-purified RNAs were diluted with 7 volumes 0.5× SSC to a final volume of 160 μl and denatured at 80°C for 5 min; then they were added to the bead suspension and incubated 1 hr at 50°C. Beads were magnetically captured for 1 min at room temperature, and then the 2S rRNA-depleted supernatant collected and precipitated with absolute ethanol. Greater than 99% of the 2S rRNA was routinely removed without measurably altering miRNA concentration; without the depletion step, nearly all the small RNA reads would correspond to 2S rRNA. Half the sample was then β-eliminated as described [19], and half was subject to the same treatment, except that sodium periodate was omitted.

Amplification and Pyrosequencing

Adapters were ligated to the small RNA sample, and the resulting library was amplified by PCR as described [20], except that a truncation mutant of RNA ligase 2 [Rnl2(1-249) (see [21])] was used for the 3′ ligation step; T4 RNA ligase (Ambion) was used for 5′ ligation. The 5′ adaptor was 5′-dAdTdC dGdTrA rGrGrC rArCrC rUrGrA rArA-3′ (Dharmacon, Lafayette, CO); 3′ "preadenylated" adapters were 5′-rAppdCdA dCdTdC dGdGdG dCdAdC dCdAdA dGdGdA ddC-3′ for fly head and 5′-rAppdTdT dTdAdA dCdCdG dCdGdA dAdTdT dCdCdA dGddC-3′ for S2 cell RNA (IDT DNA, Coralville, IA). After adaptor addition, the RNA was amplified by PCR with DNA primers corresponding to the adapters. This PCR pool was gel purified (4% Metaphor Agarose, Cambrex, East Rutherford, NJ) with Qiaex II (QIAGEN, Valencia, CA), then reamplified by PCR (common 5′ primer, 5′-GCC TCC CTC GCG CCA TCA GAT CGT AGG CAC CTG AAA-3′; 3′ primer for fly heads, 5′-GCC TTG CCA GCC CGC TCA GTC CTT GGT GCC CGA GTG-3′; 3′-primer for S2 cells, 5′-GCC TTG CCA GCC CGC TCA GCT GGA ATT CGC GGT TAA A-3′). The PCR-amplified libraries were pyrosequenced by Roche Applied Science (Branford, CT).

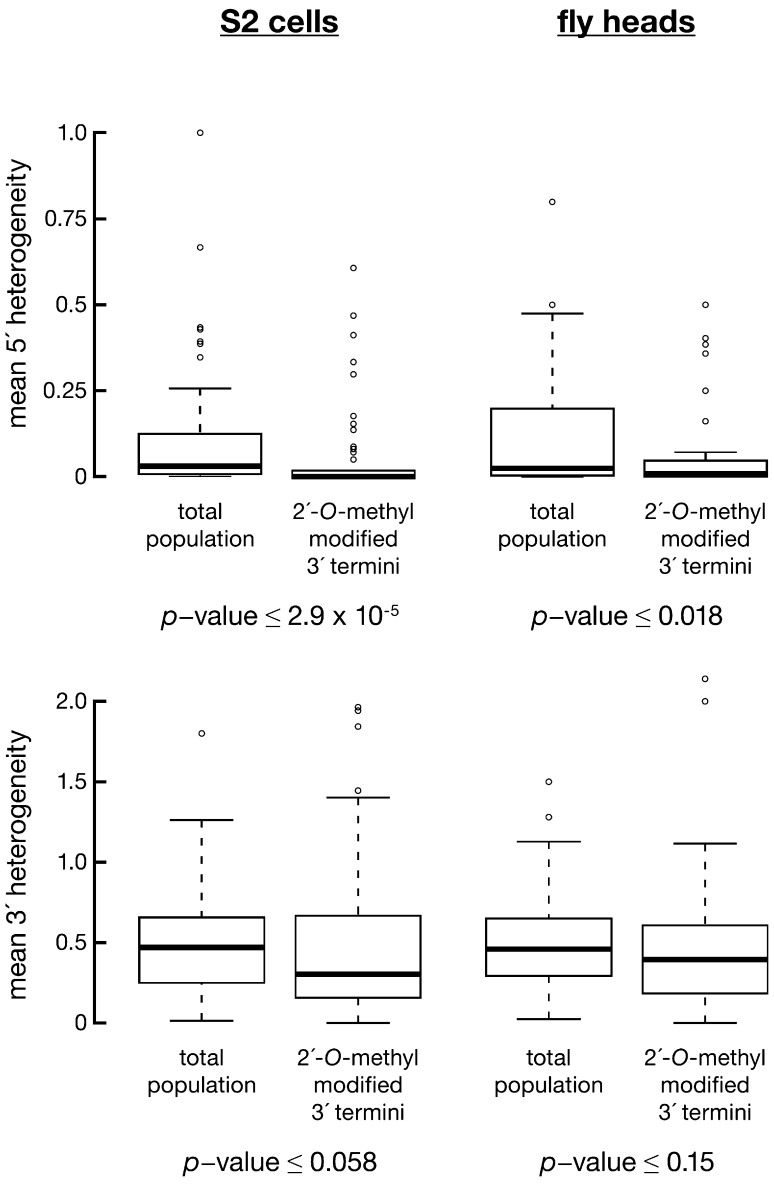

Figure 4. Ago2 Loading, as Evidenced by 3′ Terminal 2′-*O*-Methylation, Refines miRNA and miRNA* 5′ Ends

On average, the 5′ ends of the miRNAs and miRNA* strands in the 2′-*O*-methylated populations from both fly heads and S2 cells were more precisely defined than in the total population. We observed no statistically significant increase in the precision of the 3′ ends of the 3′ modified miRNAs and miRNA* strands.

Acknowledgments

We thank Stewart Shuman and Greg Hannon for providing the truncated Rnl 2 expression plasmid, Alicia Boucher for assistance with fly husbandry, Gwen Farley for preparing RNA ligase, and members of the Zamore lab for advice, suggestions, and critical comments on the text. We are especially grateful to Roche Applied Science for high-throughput sequencing. P.D.Z. is a W.M. Keck Foundation Young Scholar in Medical Research. This work was supported in part by grants from the National Institutes of Health to P.D.Z. (GM62862 and GM65236) and EMBO long-term (ALTF 910-2004) and HFSP (LT00575/2005-L) fellowships to H.S. P.D.Z. is a founder of Alnylam Pharmaceuticals and a member of its scientific advisory board. He is also a member of the scientific advisory board of Regulus Therapeutics.

Received: November 1, 2007
Revised: December 18, 2007
Accepted: December 19, 2007
Published online: January 17, 2008

References

1. Lai, E.C. (2002). MicroRNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. *30*, 363–364.
2. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20.
3. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.
4. Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. Nat. Struct. Mol. Biol. *11*, 599–606.
5. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. Nat. Genet. *37*, 495–500.
6. Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193–1207.
7. Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. (2005). The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science *310*, 1817–1821.
8. Du, T., and Zamore, P.D. (2005). microPrimer: The biogenesis and function of microRNA. Development *132*, 4645–4652.
9. Kim, V.N. (2005). MicroRNA biogenesis: Coordinated cropping and dicing. Nat. Rev. Mol. Cell Biol. *6*, 376–385.
10. Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. Curr. Biol. *15*, 1501–1507.
11. Wu, H., Neilson, J.R., Kumar, P., Manocha, M., Shankar, P., Sharp, P.A., and Manjunath, N. (2007). miRNA profiling of naive, effector and memory CD8 T cells. PLoS ONE. *2*, e1020.
12. Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing

Computational Methods

Eighteen-to-thirty-nucleotide-long reads were mapped to the *Drosophila melanogaster* genome (FlyBase assemblyR5.1; http://flybase.org/) and to the *D. melanogaster* "stem-loops" (which include the pre-miRNA sequences, usually extended by a few nucleotides) listed in miRBase (http://microrna.sanger.ac.uk/sequences/; version 10.0, August 2007). To identify nontemplated microRNA additions, we iteratively trimmed nongenome matching sequences by one to three nucleotides on either the 5′ or the 3′ end and mapped them to stem-loops.

Among stem-loop-matching reads, miRNA-matching and miRNA*-matching reads were identified, with either the experimentally detected miRNA* sequence (when it was available in the miRBase records) or the product of conceptual dicing of the hairpin [15]. So that reads that showed extremities different from those annotated in miRBase could be included, a distance of as many as nine nucleotides 5′ or 3′ from the annotated miRNA or miRNA* sequence was tolerated. Statistical calculations were made with the R statistical package; p values were calculated with the Wilcoxon test.

Supplemental Data

Additional Discussion, three figures, and three tables are available at http://www.current-biology.com/cgi/content/full/18/2/147/DC1/.

pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. J. Virol. *81*, 1951–1960.
13. 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.
14. 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.
15. Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.
16. 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.
17. Lee, Y.S., and Carthew, R.W. (2003). Making a better RNAi vector for Drosophila: Use of intron spacers. Methods *30*, 322–329.
18. Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol. *3*, e236.
19. 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.
20. 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.
21. Ho, C.K., Wang, L.K., Lima, C.D., and Shuman, S. (2004). Structure and mechanism of RNA ligase. Structure *12*, 327–339.

Accession Numbers

The GEO (http://www.ncbi.nlm.nih.gob/geo/) accession number for sequence and abundance data reported in this paper is GSE9389.