The long and short of inverted repeat genes in animals:

microRNAs, mirtrons and hairpin RNAs

Katsutomo Okamura, **Wei-Jen Chung**, and **Eric C. Lai**[*]
Sloan-Kettering Institute; Department of Developmental Biology; New York, New York USA

Abstract

MicroRNAs (miRNAs) are endogenous transcripts that contain intramolecular double stranded RNA (dsRNA) and are processed by Dicer. Their mature products are ~21-24 nucleotides in length, and they collectively regulate a broad network of endogenous transcripts. A subset of animal miRNAs are produced from mirtrons, short hairpin introns whose splicing bypasses the normal nuclear processing of canonical miRNAs. Recent studies revealed novel, extended intramolecular dsRNA produced by defined transcription units in flies and mammals, termed hairpin RNAs (hpRNAs). Detailed biogenesis studies in Drosophila showed that hpRNAs are not merely "long" miRNAs, but are actually processed by a distinct biogenesis pathway that is related to the canonical RNA interference pathway. We compare and contrast the miRNA and hpRNA pathways in this review, and describe some of the key questions that the recognition of this novel pathway raises.

Keywords

microRNA; mirtron; hairpin RNA; siRNA; Drosophila

Introduction

Three well-conserved classes of small RNAs associate with Argonaute (Ago) proteins in animals.[1] Small interfering RNAs (siRNAs) are ~21 nt, 3' methylated RNAs that reside in a "Slicer" Ago protein that is specialized for target cleavage. These species mediate classically-defined RNA interference (RNAi), in which exogenous long double-strand RNA is processed into siRNAs that cleave perfectly complementary transcripts.[2,3] microRNAs (miRNAs) are endogenous ~22 nt RNAs with 2', 3' hydroxy termini that derive from endogenous transcripts bearing inverted repeats.[4-8] Most miRNAs reside in Ago-class proteins that have modest or no cleavage activity; instead, they exert their main effects via translational inhibition and/or target deadenylation.[9] miRNAs collectively regulate thousands of endogenous target transcripts via as little as 7 nt "seed" pairing with nucleotides 2-8 of the miRNA. Piwi-interacting RNAs (piRNAs) are ~24-30 nt, 3' methylated RNAs that reside in Piwi-class Argonautes in flies and vertebrates, whose tested members are all Slicers.[10] piRNAs are found predominantly in the germline, and one of their essential functions is to restrict the transposition of selfish genetic elements.

Until recently, the dogma was that siRNAs and piRNAs are produced from intermolecular double strand RNA (dsRNA), whereas miRNAs are produced from intramolecular dsRNA. In particular, siRNAs are produced by Dicer-mediated cleavage of perfectly double stranded

*Correspondence to: Eric C. Lai; Sloan-Kettering Institute; 1275 York Ave., Box 252; New York, New York 10065 USA; Tel.: 212.639.5578; Fax: 212.717.3604; Email: E-mail: laie@mskcc.org.

RNA;[11] a specialized nematode pathway also generates secondary siRNAs via unprimed synthesis off the 3' end of cleaved transcripts by an RNA dependent RNA polymerase.[12,13] piRNA production involves a feed-forward loop in which piRNAs anneal to complementary transposon-related transcripts.[14-16] "Antisense" piRNAs derived from piRNA master transcripts guide Piwi-class complexes to cleave active transposon transcripts and generate "sense" piRNAs; these reciprocally generate more antisense piRNAs through the cleavage of piRNA master transcripts. (Note that the biogenesis mechanism for abundant non-transposon-derived piRNAs present in pachytene-stage mammalian sperm is unclear. Their generation involves long single stranded precursors, but their antisense counterparts, if any, remain to be found). In contrast, miRNAs are produced from intramolecular dsRNA in the form of an inverted repeat transcript. As such, miRNAs were generally believed to be the only Argonaute-bound small RNA that could be generated by conventional transcription units without the need for a trans-annealing mechanism (i.e., via complementary transcripts) or a copying mechanism (i.e., via an RNA dependent RNA polymerase that could synthesize a complementary strand).

Recently, a new pathway was found to generate siRNAs via long inverted repeat transcripts in flies[17-19] and mice,[20-21] termed hairpin RNAs (hpRNAs). Best characterized in Drosophila, these longer cousins of miRNA transcripts are processed by a distinct pathway that includes many canonical RNAi factors but also a canonical miRNA factor. The realization of this pathway blurs the distinction between the miRNA and RNAi pathways, and raises new questions about the genesis and function of inverted repeat RNA genes in animals.

Biogenesis and Function of Canonical miRNAs and mirtrons

Animal miRNAs are processed from short hairpins termed pre-miRNAs, which themselves are the products of much longer precursors known as pri-miRNAs.[22] The nuclear RNAse III enzyme Drosha is responsible for "cropping" pre-miRNA hairpins out of pri-miRNAs (Figs. 1 and 2A).[23] Most pri-miRNAs are non-coding transcripts, and the pre-miRNA can derive from either the introns or exons of such transcripts. However, about a third of pre-miRNAs are located in the introns of protein-coding genes.[24]

A subset of miRNAs derive from short hairpin introns termed "mirtrons" (Fig. 1).[25,26] The splicing machinery directly defines the mirtron ends, and following their linearization by lariat debranching enzyme they adopt hairpin structures typical of pre-miRNAs (Fig. 2B). Mirtrons constitute 5-10% of miRNA genes in invertebrates[25,26] and vertebrates,[27,28] but as typical mirtrons are expressed at much lower levels than typical canonical miRNAs, their regulatory influence is presumed to be comparatively more narrow.

Pre-miRNA hairpins produced from canonical precursors and mirtrons are both exported from the nucleus via Exportin-5. In the cytoplasm, pre-miRNAs are cleaved by the RNase III enzyme Dicer to yield a ~21-22 nt duplex. One strand of the duplex, termed the miRNA, is preferentially selected for entry into a silencing complex that includes an Argonaute (AGO) protein.[29,30] The miRNA guides the AGO complex to complementary sites on target transcripts, which are mostly (although not exclusively) located in 3' UTRs. The partner strand of the duplex, known as the miRNA* species, is preferentially degraded; however, it was recently recognized that many miRNA* species can also access AGO complexes and regulate targets.[31]

If one introduces an artificial target that is perfectly complementary to an endogenous miRNA, it will be usually be readily sliced at a position that lies opposite nucleotides 10 and 11 measured from the 5' end of the miRNA,[32] just as mRNAs perfectly complementary to artificial siRNAs are.[33,34] In Drosophila, miRNAs are preferentially sorted into the poor slicer AGO1, with a small portion of miRNAs sorted to the main Slicer AGO2.[17,19,35-37] It is presently unclear what endogenous regulatory purpose the slicing capability of miRNAs serves, since with few exceptions,[38,39] animal miRNAs pair imperfectly to endogenous mRNAs. In fact, as little as

7 nucleotides of Watson-Crick basepairing to positions 2-8 from their 5' ends (the miRNA "seed") can be sufficient to direct substantial repression.[40-43] Transcripts bearing highly conserved seed matches are inferred to maintain functional regulatory interactions with cognate miRNAs, and comprise at least 30% of mammalian transcripts.[44] Additional functional miRNA targets may have imperfect seed matching[40,45,46] or be poorly conserved.[47-49] The mechanism(s) by which imperfectly-paired targets are repressed by miRNAs is controversial at present, with experimental support for many models published in recent years. These include sequestration from ribosomes (by relocalizing into P bodies), blockage of translational initiation, translational repression after initiation, co-translational degradation, and target deadenylation coupled to transcript degradation.[9]

Strategies for miRNA Genefinding

Although a few miRNAs were found purely by genetic means, most were found by cloning and/or by computational means. In the cloning method, short RNAs are prepared either by size fractionation by polyacrylamide gel electrophoresis,[4,6,7] or by co-immunoprecipitation with Argonaute-containing complex.[50-52] Following reverse transcription and sequencing, one attempts to map cloned ~21-24 nt RNAs to a local genomic inverted repeat. In the computational method, one identifies evolutionarily conserved hairpins that exhibit characteristic patterns of nucleotide divergence, namely candidates for which the terminal loop evolves more quickly than either hairpin arm.[53,54] Consideration of several structural and/ or sequence details can further enrich for likely miRNAs, although the improvement in specificity always comes at some tradeoff in sensitivity, since some bona fide miRNA hairpins inevitably fail to pass certain criteria.[55-59]

Current computational approaches exploit the powerful comparative analysis now possible with tens of assembled animal genomes, and cloning approaches benefit tremendously from the capacity of next-generation sequencing strategies that produce millions of short RNA sequences in a single run. These combined approaches have now yielded many thousands of miRNAs.[60] Does this depth of miRNA analysis mean that we have a good handle on what a miRNA gene looks like? In fact, we do not. In contrast to protein-coding genefinders, there is presently no effective method to perform miRNA genefinding across a complex genome without taking into account comparative genomics. Another serious issue concerns the fact that most computational efforts depend on a fairly narrow set of criteria based on the initial set of cloned miRNAs.[61] For instance, a recent study used deep-sequencing data to confidently some unexpectedly long miRNA hairpins in Drosophila that were twice the length of typical pre-miRNAs,[58] a range that is all but ignored by most miRNA annotators. In addition, the recent discovery that miRNAs can derive from mirtrons, whose hairpins are too short to qualify as canonical miRNAs,[25-27] further suggests that miRNA genes of atypical structure may remain to be identified.

Drosophila hairpin RNAs generate siRNAs from long inverted repeat transcripts

A few years ago, transgenic RNAi was induced in flies using inverted repeat snapback constructs,[62-64] although they were difficult to clone and the knockdowns in some cases were variable. It was subsequently realized that the inclusion of an spacer between the hairpin stems[65] could improve the stability of such inverted repeats in bacteria, as well as potentially in transgenic animals,[66] although the knockdowns remained variable in some cases. A third innovation was the inclusion of an intron[67] or the incorporation of genomic fragments spanning introns;[68] it was suggested that these promoted their knockdown activity. Interestingly, an early study aimed at annotating putative non-coding mRNA-like (pncr) molecules in Drosophila identified a spliced, polyadenylated transcript named pncr009.[69] Although not

reported at the time, in some respects it very much resembled the design of typical RNAi transgenes: it has ~400 bp of stem and a long, spliced, terminal loop. The only difference is that its stem is imperfectly paired, akin to a long miRNA hairpin.

Computational methods can be employed to identify other potential examples of long inverted transcripts. This strategy was used effectively in Arabidopsis to identify miRNA genes.[70,71] However, in the absence of supporting experimental evidence, lists of genomic inverted repeats are of limited utility. Thousands of long inverted repeats are easily found, but they are not necessarily contained within a defined transcription unit. The vast majority involve the long terminal repeats of transposons, and a substantial number of the remaining loci comprise tandem duplicated tRNA or protein-coding genes arranged in convergent or divergent orientation.[18] However, a handful of these are contained within sequenced, spliced, cDNAs, indicating that they may also belong to the pncr009 class. In fact, several of them are clustered in the genome with pncr009.[18]

Recently, the availability of large catalogs of cloned Drosophila small RNAs allowed lists of genomic inverted repeats to be rationally evaluated. This revealed a limited number of inverted repeats that generated small RNAs and did not map to transposable elements or satellite sequences.[18] Of these, the possibility that some of these loci were degraded into cloned breakdown fragments, as opposed to generating processed small RNAs, was assessed by examining the size distribution of RNAs that mapped to each locus. By demanding that the majority of cloned small RNAs were 21-22 nt in length, relative to all other sizes, this analysis boiled down to 7 genomic loci that were predicted as long inverted repeats and generated primarily 21-22 nt RNAs (e.g., Fig. 2C); one of the loci (hp-CG4068) consisted of 20 tandem repeats.[18] Several of these loci were represented by full length polyadenylated cDNAs, indicating that they derive from conventional RNA polymerase II transcripts; some of these were also spliced. These were collectively designated as hairpin RNAs (hpRNAs).

Despite the fact that miRNA precursors and hpRNAs are both imperfect inverted repeats, these RNAs are channeled into distinct biogenesis pathways. Similar to exogenous siRNAs generated from artificial substrates, hpRNAs are processed by Dcr-2 instead of Dcr-1, are loaded into AGO2, and are modified at their 3' ends by Hen1 methyltransferase (Fig. 1).[17-19] Like siRNAs from artificial long inverted repeats,[72] the siRNA duplexes derived from hpRNAs are also phased (Fig. 2C). Tests of artificial and endogenous substrates demonstrated that hpRNAs could repress and directly cleave targets.[17,18] Thus, hpRNAs generate siRNAs. Curiously then, the accumulation of hpRNA-derived siRNAs requires Loquacious (Loqs), which is dsRBD protein that assists Dcr-1 in the cleavage of pre-miRNA hairpins (Fig. 1).[1]

Endogenous hpRNA Targets in Drosophila

In plants, most miRNAs exhibit extended complementarity to one or a few mRNAs. Quite a few have now been validated to direct the cleavage of these targets,[70] although it should be noted that plant miRNAs can also mediate translational repression of highly complementary targets.[73] miR-196 is a rare example of an animal miRNA that is essentially perfectly complementary to a target (*HoxB8*) and directs its cleavage.[38,39] Several Drosophila hpRNA-derived siRNAs are nearly perfectly complementary to endogenous transcripts. For example, one of the siRNA products of hp-CG4068 is antisense to the coding region of *mus308*, a DNA polymerase involved in the DNA damage response. This is the best-characterized hpRNA target thus far, having been validated by sensor tests in cultured cells and by cleavage assays that tested both endogenous hp-CG4068 and endogenous *mus308*.[17,18] Another hpRNA, hp-CG18854, is a pseudogene of *CG8289*, which encodes a chromodomain protein.[17,18] The endogenous regulatory activity of hp-CG18854-derived siRNAs was detected in cultured cells, and elevated hp-CG18854 could repress a *CG8289-GFP* fusion target in trans.[18]

These observations indicate that hpRNA-derived siRNAs can target endogenous transcripts by cleaving them. However, it is unclear whether hpRNA-derived siRNAs might target endogenous targets with limited complementarity. In one test performed thus far, the *mus308*-targeting hp-CG4068 siRNA did not appear to substantially repress a seed-matched target.[18] However, as Drosophila siRNAs are able to cleave targets that are less-than-perfectly complementary,[74] it is conceivable that additional hpRNA targets might exist.

Analogous siRNA Systems for Mammalian Pseudogenes

Concurrent studies of mouse oocytes revealed four loci that were designated as hairpin RNAs. In fact, one of the most abundant siRNA loci was an hpRNA comprising a long-inverted repeat pseudogene of the *Rangap1* gene,[20,21] similar to the case of hp-CG18854:CG8289 in *Drosophila*. As with Drosophila hpRNAS, the accumulation of siRNAs from *hp-Rangap1* (also known as *Au76*) is Dicer-dependent, and *Dicer*-mutant oocytes exhibit substantial upregulation of cognate *Rangap1* transcripts. It is unclear whether this regulatory strategy is conserved or convergent, but these findings suggest that analogous systems to generate siRNAs from long inverted repeats operate in both Drosophila and mammals.

Curiously, mammalian oocytes exhibit a much more complex network of pseudogene-derived siRNAs that involve at least 20-30 antisense transcribed pseudogenes and their cognate progenitors. It was proposed that the siRNAs are generated from dsRNA formed from annealed gene:pseudogene complementary pairs, and their regulatory impact was evidenced by the upregulation of many the siRNA-complementary transcripts in *Dicer*-mutant oocytes.[20,21] Thus, while pseudogenes have generally been assumed to be nonfunctional loci, both intramolecular and intermolecular mechanisms can produce regulatory siRNAs from pseudogenes.

Future Questions for hpRNA Research

The discovery of the hpRNA pathway raises many questions that will undoubtedly keep small RNA researchers busy for many years. One key question is regards how hpRNA and miRNA precursors are segregated. Just as the shortest giants and the tallest dwarves overlap in height, there are "short" hpRNAs[18] and "long" miRNAs[58] that appear indistinguishable. Yet, these are sorted into distinct biogenesis pathways, since long miRNA hairpins still generate only a single small RNA duplex, whereas short hpRNAs generate multiple small RNA duplexes. In the absence of merely a size-dependent mechanism, there is presumably a more sophisticated strategy that either prevents hpRNAs from being processed by Dcr-1, and/or prevents pre-miRNAs from being cleaved by Dcr-2. The nature of either putative sorting mechanism remains to be better understood.

More typically, though, hpRNA stemloops are much longer than miRNA stemloops. What is the upper limit on hpRNA stem length? Could there be read-through transcripts across duplicated inverted repeat genes that might generate hpRNA-like foldbacks many kilobases in length? Might this sort of pathway generate some fraction of transposon-derived siRNAs?[17,19,36,75] This might conceivably be akin to the generation of nematode siRNAs from transcriptional readthrough across Tc1 elements, which form intramolecular dsRNA across the terminal inverted repeats.[76] Finally, several of the hpRNAs are derived from spliced primary transcripts.[17,18] Is there any limit to the size of their introns? If not, might there exist hpRNAs whose stems are separated by long introns perhaps tens of kilobases in length? The answers to these remain for future studies.

The finding that the canonical miRNA factor Loquacious (Loqs) plays a substantial role in hpRNA processing raises another mystery. Loqs is a dsRBD protein that is usually portrayed as a core miRNA biogenesis factor that aids Dicer-1 in cleaving pre-miRNA hairpins.[77-79]

However, careful scrutiny of *loqs* null mutant animals and biochemical reconstitution experiments recently suggest that Loqs has only an auxiliary role.[80] Although it is certainly involved in miRNA biogenesis, the maturation of a substantial number of miRNA loci is reasonably normal in the absence of Loqs. Surprisingly then, Loqs proved to be very strongly required for the accumulation of many hpRNA-derived siRNAs (Fig. 1).[17,18] Does Loqs partner directly with Dicer-2 to generate some kinds of siRNAs? Addressing this will likely require in vitro reconstitution of hpRNA processing with defined factors.

The biogenesis of siRNAs from perfectly double-stranded substrates was previously shown to require the dsRBD protein R2D2, which directs Dicer-2 to load siRNAs into AGO2.[81,82] R2D2binds Dicer-2 directly, just as Loqs binds Dicer-1 directly. One might have thought that the imperfect dsRNA found in pre-miRNAs and hpRNAs might explain their shared requirement for Loqs. However, this does not seem to be the case, since several transposonderived siRNAs and cis-NAT-siRNAs similarly require Loqs for their accumulation.[17,75,83] Although not revealed by earlier protein interaction studies, proteomic analysis of Dicer-2 complexes suggest that it Dicer-2/Loqs complexes may exist.[17] Do these proteins directly work as a team to process endo-siRNAs, and is so, is their requirement in dicing or loading of siRNAs? Clearly, there is much to be learned about the functional partnerships of RNAse III enzymes and their dsRBD partners.

Perhaps most importantly, what are the endogenous functions of hpRNAs? The best characterized targets of hpRNAs encode a chromodomain protein (*CG8289*) and a nuclease (*mus308*).[17,18] The generation and analysis of hpRNA deletion mutants will help to address the endogenous significance of these regulatory interactions. Curiously, Drosophila cis-natural antisense transcripts that generate siRNAs are also strongly enriched for various nuclease and transcription cofactors.[83] Perhaps these findings suggest a shared molecular axis for the Drosophila endo-siRNA network.

Acknowledgements

References

1. Farazi TA, Juranek SA, Tuschl T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. Development 2008;135:1201–14. [PubMed: 18287206]

2. Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA. Targeted mRNA degradation by double-stranded RNA in vitro. Genes and Development 1999;13:3191–7. [PubMed: 10617568]

3. Zamore PD, Tuschl T, Sharp PA, Bartel DP. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 2000;101:25–33. [PubMed: 10778853]

4. Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 2001;294:862–4. [PubMed: 11679672]

5. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell 1993;75:843–54. [PubMed: 8252621]

6. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001;294:853–8. [PubMed: 11679670]

7. Lau N, Lim L, Weinstein E, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 2001;294:858–62. [PubMed: 11679671]

8. Reinhart BJ, Slack F, Basson M, Pasquinelli A, Bettinger J, Rougvie A, Horvitz HR, Ruvkun G. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 2000;403:901–6. [PubMed: 10706289]

9. Liu J. Control of protein synthesis and mRNA degradation by microRNAs. Curr Opin Cell Biol 2008;20:214–21. [PubMed: 18329869]

10. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 2007;318:761–4. [PubMed: 17975059]

11. Bernstein E, Caudy A, Hammond S, Hannon G. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001;409:363–6. [PubMed: 11201747]

12. Sijen T, Steiner FA, Thijssen KL, Plasterk RH. Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science 2007;315:244–7. [PubMed: 17158288]

13. Pak J, Fire A. Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. Science 2007;315:241–4. [PubMed: 17124291]

14. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell 2007;128:1089–103. [PubMed: 17346786]

15. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science 2007;315:1587–90. [PubMed: 17322028]

16. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 2007;316:744–7. [PubMed: 17446352]

17. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel J, Sachidanandam R, Hannon G, Brennecke J. An endogenous siRNA pathway in Drosophila. Nature 2008;453:798–802. [PubMed: 18463631]

18. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 2008;453:803–6. [PubMed: 18463630]

19. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada T, Siomi MC, Siomi H. Drosophila endogenous small RNAs bind to Argonaute2 in somatic cells. Nature 2008;453:793–7. [PubMed: 18463636]

20. Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 2008;453:534–8. [PubMed: 18404147]

21. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 2008;453:539–43. [PubMed: 18404146]

22. Du T, Zamore PD. microPrimer: the biogenesis and function of microRNA. Development 2005;132:4645–52. [PubMed: 16224044]

23. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing. Nature 2003;425:415–9. [PubMed: 14508493]

24. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Res 2004;14:1902–10. [PubMed: 15364901]

25. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. Cell 2007;130:89–100. [PubMed: 17599402]

26. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature 2007;448:83–6. [PubMed: 17589500]

27. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. Mol Cell 2007;28:328–36. [PubMed: 17964270]

28. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res 2008;18:957–64. [PubMed: 18469162]

29. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell 2003;115:209–16. [PubMed: 14567918]

30. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell 2003;115:199–208. [PubMed: 14567917]

31. Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. Nat Struct Mol Biol 2008;15:354–63. [PubMed: 18376413]

32. Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. Science 2002;297:2056–60. [PubMed: 12154197]

33. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 2001;15:188–200. [PubMed: 11157775]

34. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. Embo J 2001;20:6877–88. [PubMed: 11726523]

35. Seitz H, Ghildiyal M, Zamore PD. Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA strands in flies. Curr Biol 2008;18:147–51. [PubMed: 18207740]

36. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD. Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. Science 2008;320:1077–81. [PubMed: 18403677]

37. Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell 2007;130:287–97. [PubMed: 17662943]

38. Yekta S, Shih IH, Bartel DP. MicroRNA-directed cleavage of HOXB8 mRNA. Science 2004;304:594–6. [PubMed: 15105502]

39. Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J, Chaudhuri A, Farzan-Kashani R, Zuker M, Pasquinelli AE, Ruvkun G, Sharp PA, Tabin CJ, McManus MT. MicroRNA-responsive `sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat Genet 2004;36:1079–83. [PubMed: 15361871]

40. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of MicroRNA-Target Recognition. PLoS Biol 2005;3:85.

41. Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. Genes Dev 2003;17:438–42. [PubMed: 12600936]

42. Lai EC. microRNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet 2002;30:363–4. [PubMed: 11896390]

43. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003;115:787–98. [PubMed: 14697198]

44. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20. [PubMed: 15652477]

45. Ha I, Wightman B, Ruvkun G. A bulged *lin-4/lin-14* RNA duplex is sufficient for *Caenorhabditis elegans lin-14* temporal gradient formation. Genes & Dev 1996;10:3041–50. [PubMed: 8957004]

46. Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. Genes Dev 2004;18:132–7. [PubMed: 14729570]

47. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 2006;38:813–8. [PubMed: 16751773]

48. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 2006;312:75–9. [PubMed: 16484454]

49. Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. Science 2005;310:1817–21. [PubMed: 16308420]

50. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev 2002;16:720–8. [PubMed: 11914277]

51. Ding L, Spencer A, Morita K, Han M. The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in *C. elegans*. Mol Cell 2005;19:437–47. [PubMed: 16109369]

52. Nelson PT, De Planell-Saguer M, Lamprinaki S, Kiriakidou M, Zhang P, O'Doherty U, Mourelatos Z. A novel monoclonal antibody against human Argonaute proteins reveals unexpected characteristics of miRNAs in human blood cells. Rna 2007;13:1787–92. [PubMed: 17720879]

53. Lai EC, Tomancak P, Williams RW, Rubin GM. Computational identification of Drosophila microRNA genes. Genome Biol 2003;4:1–20.

54. Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 2005;120:21–4. [PubMed: 15652478]

55. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP. Vertebrate microRNA genes. Science 2003;299:1540. [PubMed: 12624257]

56. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. The microRNAs of *Caenorhabditis elegans*. Genes Dev 2003;17:991–1008. [PubMed: 12672692]

57. Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ, Kellis M. Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. Genome Res 2007;17:1865–79. [PubMed: 17989255]

58. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res 2007;17:1850–64. [PubMed: 17989254]

59. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 2006;125:887–901. [PubMed: 16751099]

60. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008;36:154–8.

61. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T. A uniform system for microRNA annotation. RNA 2003;9:277–9. [PubMed: 12592000]

62. Kennerdell JR, Carthew RW. Heritable gene silencing in Drosophila using double-stranded RNA. Nat Biotechnol 2000;18:896–8. [PubMed: 10932163]

63. Fortier E, Belote JM. Temperature-dependent gene silencing by an expressed inverted repeat in Drosophila. Genesis 2000;26:240–4. [PubMed: 10748461]

64. Lam G, Thummel CS. Inducible expression of double-stranded RNA directs specific genetic interference in Drosophila. Curr Biol 2000;10:957–63. [PubMed: 10985382]

65. Piccin A, Salameh A, Benna C, Sandrelli F, Mazzotta G, Zordan M, Rosato E, Kyriacou CP, Costa R. Efficient and heritable functional knock-out of an adult phenotype in Drosophila using a GAL4-driven hairpin RNA incorporating a heterologous spacer. Nucleic Acids Res 2001;29:55–5. [PubMed: 11125048]

66. Giordano E, Rendina R, Peluso I, Furia M. RNAi triggered by symmetrically transcribed transgenes in *Drosophila melanogaster*. Genetics 2002;160:637–48. [PubMed: 11861567]

67. Lee YS, Carthew RW. Making a better RNAi vector for Drosophila: use of intron spacers. Methods 2003;30:322–9. [PubMed: 12828946]

68. Kalidas S, Smith DP. Novel genomic cDNA hybrids produce effective RNA interference in adult Drosophila. Neuron 2002;33:177–84. [PubMed: 11804566]

69. Tupy JL, Bailey AM, Dailey G, Evans-Holm M, Siebel CW, Misra S, Celniker SE, Rubin GM. Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. Proc Natl Acad Sci USA 2005;102:5495–500. [PubMed: 15809421]

70. Jones-Rhoades MW, Bartel DP. Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. Mol Cell 2004;14:787–99. [PubMed: 15200956]

71. Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, Carrington JC. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet 2004;36:1282–90. [PubMed: 15565108]

72. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. Science 2006;313:320–4. [PubMed: 16809489]

73. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O. Widespread translational inhibition by plant miRNAs and siRNAs. Science 2008;320:1185–90. [PubMed: 18483398]

74. Haley B, Zamore PD. Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol 2004;11:599–606. [PubMed: 15170178]

75. Chung WJ, Okamura K, Martin R, Lai EC. Endogenous RNA interference provides a somatic defense against Drosophila transposons. Current Biology 2008;18:795–802. [PubMed: 18501606]

76. Sijen T, Plasterk RH. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature 2003;426:310–4. [PubMed: 14628056]

77. Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol 2005;3:236.

78. Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. Genes Dev 2005;19:1674–9. [PubMed: 15985611]

79. Saito K, Ishizuka A, Siomi H, Siomi MC. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. PLoS Biol 2005;3:235.

80. Liu X, Park JK, Jiang F, Liu Y, McKearin D, Liu Q. Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. Rna 2007;13:2324–9. [PubMed: 17928574]

81. Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science 2003;301:1921–5. [PubMed: 14512631]

82. Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. A protein sensor for siRNA asymmetry. Science 2004;306:1377–80. [PubMed: 15550672]

83. Okamura K, Balla S, Martin R, Liu N, Lai EC. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila. Nat Struct Mol Biol 2008;15:581–90. [PubMed: 18500351]
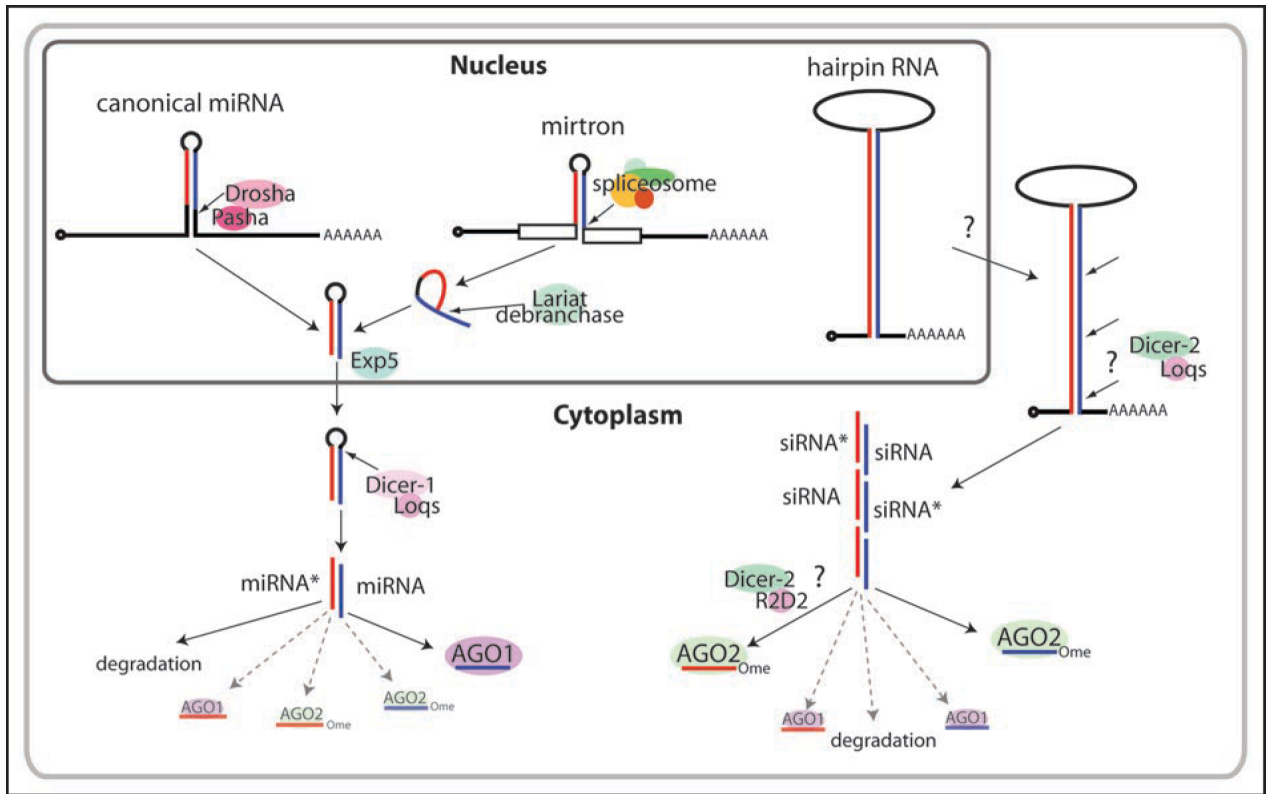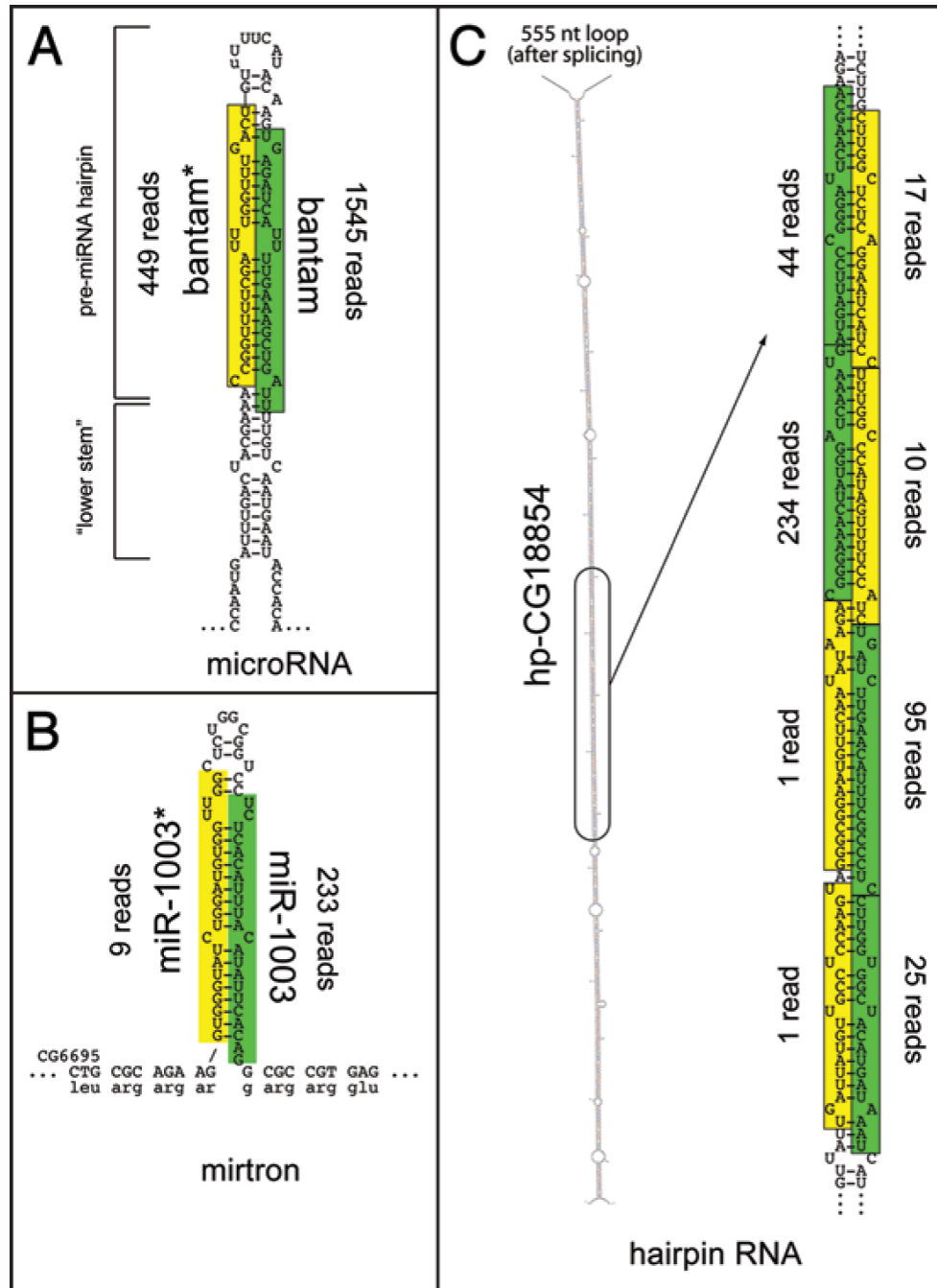
Figure 1.
Models for the biogenesis of canonical miRNAs, mirtrons, and hpRNAs in Drosophila. All three originate in the nucleus as RNA polymerase II transcripts containing inverted repeats, and are presumably capped and polyadenylated species. Canonical miRNAs are first cleaved by the Pasha/Drosha complex to release the pre-miRNA hairpin. Mirtrons are short intronic hairpins that are spliced and debranched to yield pre-miRNA mimics. Both types of pre-miRNAs are exported to the cytoplasm via Exportin-5, where they are cleaved by the Dicer-1/Loqs complex. The usual fate of the resulting small RNA duplex is for one strand to be preferentially loaded into AGO1 as the mature miRNA, while its partner miRNA* strand is preferentially degraded. However, a fraction of miRNA species are transferred into AGO2 and methylated at their 3' ends; a fraction of miRNA* strands are also transferred into AGO proteins. Hairpin RNA processing is not dependent on Exportin-5, and so the relevant factor is not known (?); possibly it transits the conventional mRNA export pathway. In the cytoplasm, Dicer-2 and Loqs are required for hpRNA maturation, but it is not known if these function as a discrete complex or not (?). The typical fate of the resulting small RNA duplexes is their preferential loading into AGO2 and 3' methylation as mature siRNAs. However, some hpRNA products associate with AGO1. A Dicer-2/R2D2 complex loads exogenously-derived siRNAs into AGO2, but it is not known if this is true for hpRNA-derived siRNAs (?).

Figure 2.
Examples of canonical microRNA, mirtron, and hairpin RNA precursors from Drosophila. (A) A canonical miRNA precursor includes a lower duplex stem that is bound by Pasha/DGCR8 and cleaved by the Drosha RNAse III enzyme. The resultant pre-miRNA hairpin is cleaved on the terminal loop side by the Dicer RNAse III enzyme (specifically Dicer-1 in Drosophila) to yield a miRNA/miRNA* duplex. One duplex strand (in this case the bantam species, shaded in green) is preferentially loaded into an Argonaute protein, resulting in its preferred stability and higher cloning frequency. The numbers of cloned bantam/bantam* reads are from Ruby et al.[58] (B) A mirtron is a short hairpin intron; shown is an example from the CG6695 gene. The hairpin ends coincide with the spliced intron, and thus the left RNA begins with a splice

donor (here, GUGGGU) and the right RNA ends with a splice acceptor (here, CAG). Both canonical miRNAs and mirtrons are processed on their loop sides by Dicer-1. However, while mature miRNAs collectively derive from both left and right hairpin arms, the mature miRNA of Drosophila mirtrons almost always derives from the right arm (shaded in green). Reads are from Ruby et al.[26] (C) A hairpin RNA is an inverted repeat transcript whose duplex region is usually much larger than that of canonical miRNAs; in this case, it is ~400 bp. hp-CG18854, like some other hpRNA loci, contains a large unstructured loops and is also spliced. Thus, the duplex regions are separated by substantial intervals of genomic sequence. hpRNAs generate small RNA duplexes with 3' overhangs, as is the case for canonical miRNAs and mirtrons. However, hpRNAs alone generate multiple cloned duplexes that are phased, likely because they are processed by Dicer-2. The dominant species of a given duplex can derive from either the left or right arm (as shaded in green); reads are from Okamura et al.[18]