Published in final edited form as:
Biochimie. 2011 November ; 93(11): 1897–1904. doi:10.1016/j.biochi.2011.06.017.

Mirtrons: microRNA biogenesis via splicing

Jakub O. Westholm[1] and **Eric C. Lai**[1,2]
[1]Department of Developmental Biology, Sloan-Kettering Institute, 1275 York Ave, Box 252, New York NY 10065

Abstract

A well-defined mechanism governs the maturation of most microRNAs (miRNAs) in animals, via stepwise cleavage of precursor hairpin transcripts by the Drosha and Dicer RNase III enzymes. Recently, several alternative miRNA biogenesis pathways were elucidated, the most prominent of which substitutes Drosha cleavage with splicing. Such short hairpin introns are known as mirtrons, and their study has uncovered related pathways that combine splicing with other ribonucleolytic machinery to yield Dicer substrates for miRNA biogenesis. In this review, we consider the mechanisms of splicing-mediated miRNA biogenesis, computational strategies for mirtron discovery, and the evolutionary implications of the existence of multiple miRNA biogenesis pathways. Altogether, the features of mirtron pathways illustrate unexpected flexibility in combining RNA processing pathways, and highlight how multiple functions can be encoded by individual transcripts.

Keywords

mirtron; microRNA; small RNA biogenesis; splicing

1. Introduction

The past decade has seen an explosion in the diversity of processing pathways that generate ~20–30 nucleotide (nt) regulatory RNAs, including the microRNA (miRNA), small interfering RNA (siRNA), and Piwi-interacting RNA (piRNA) pathways [1]. All of these pathways have intimate connections with double-stranded RNA (dsRNA) for biogenesis and/or functional activity. They are further linked in that the resultant mature small RNAs are loaded into Argonaute family proteins, and guide them to target transcripts. The Argonaute proteins can be broadly grouped into the AGO and Piwi clades; miRNAs and siRNAs associate with the former and piRNAs with the latter [2].

The most widely-studied class of Argonaute cargoes are the ~22 nt miRNAs, which typically number in the hundreds amongst well-studied model systems [3]. Plant miRNAs typically exhibit extensive or perfect complementarity to targets, with individual miRNAs usually regulating one or a few targets, with strong bias for transcription factors [4, 5]. However, animal miRNAs can regulate transcripts bearing as little as 7 nt complements to their 5' ends [6–8], with the result that many animal miRNAs have captured hundreds of

[2]Corresponding author. tel: (212) 639-5578, fax: (212) 717-3604, laie@mskcc.org.

conserved targets [9]. Consequently, almost all biological processes in animal cells are conceivably under miRNA control, a fact that has provoked a torrent of miRNA research. Indeed, the deregulation or dysfunction of miRNAs has been extensively linked to developmental aberrations, physiological and behavioral abnormalities, and cancer [10]. We have recently compared and contrasted plant and animal miRNA pathways [5], and focus here on animal systems.

2. Biogenesis of canonical miRNAs

The biogenesis of canonical animal miRNAs involves stepwise cleavage of longer primary miRNA (pri-miRNA) transcripts (Figure 1A). Pri-miRNAs are typically (although not exclusively) transcribed by RNA polymerase II, and they contain one or more inverted repeats that are substrates for miRNA production. About one third of miRNA genes are located in the introns of protein-coding genes; these are overwhelming found on the sense strand, implying some linkage of miRNA and host mRNA transcription [11, 12]. It is often assumed that intronic miRNAs are processed from spliced introns. However, in cases where it has been examined closely the miRNA hairpin were actually cleaved first, followed by splicing of the severed mRNA [13].

Pri-miRNA cleavage occurs near the base of the hairpin stem, and is executed in the nucleus by the Drosha RNase III enzyme and its dsRNA binding partner DGCR8 (also known as Pasha in invertebrates). One model posits that DGCR8 identifies the junction between single-stranded and double-stranded regions at the hairpin base, thereby positioning Drosha to cleave ~1 helical turn into the stem [14]; the influence of the terminal loop in regulating Drosha cleavage has also been raised [15]. Drosha cleavage releases a pre-miRNA hairpin that is most frequently ~55–70 nt in length, although some pre-miRNAs, particularly in *Drosophila*, are more than twice as long [16]. Cleavage by RNase III enzymes leaves behind a ~2 nt 3' overhang, and this feature of pre-miRNAs is recognized by the export factor Exportin-5 [17]. Once in the cytoplasm, pre-miRNA hairpins are recognized and cleaved within their stems about two helical turns from the base by the Dicer RNase III enzyme and its dsRNA binding partner TRBP (also known as Loquacious in Drosophila). This yields paired ~22 nt RNAs, known as the miRNA/miRNA* duplex, which exhibit characteristic 3' overhangs at either end [1].

Strand terminology is an operational definition: the side that accumulates to a higher level is known as the "mature" miRNA, while its less abundant partner is referred to as the miRNA* or "star" strand. Generally speaking, the asymmetric accumulation of miRNA/miRNA* strands reflects the preferential loading of one strand into functional Ago complex, and the preferential degradation of its partner strand [18, 19]. Amongst conserved miRNA genes, mature miRNAs are generally embedded into much larger regulatory networks than their companion star strands. Nevertheless, this does not mean that miRNA* strands are merely carriers that solely promote miRNA biogenesis. In fact, many miRNA* strands are well-conserved, exhibit regulatory capacity [20–22], and/or have distinct Ago sorting properties [23–25].

3. mirtrons: miRNA biogenesis via splicing

Following the elucidation of the core aspects of canonical miRNA biogenesis, a number of non-canonical pathways have emerged [26]. The first alternative miRNA biogenesis pathway to be characterized, the "mirtron" pathway, marries intron splicing with dicing. Mirtrons were first recognized in the fly and worm, by virtue of cloned small RNAs that mapped precisely to the termini of short intronic hairpins [27, 28]. These loci gave rise to pre-miRNA hairpins with 3' overhangs and subsequently to mature ~22 nt species that function as typical miRNA-class regulatory RNAs, features that suggested their maturation

by Dicer. However, their precursor stem lengths were shorter than with canonical pri-miRNAs, since they comprised only the miRNA/miRNA* duplex and lacked the lower stem of ~1 helical turn that typically recruits and mediates cleavage by Drosha/DGCR8 complex [14]. Instead, the fact that the resultant short RNAs directly abutted intron-exon boundaries suggested that splicing might substitute for Drosha cleavage.

Indeed, structure-function and knockdown studies verified that the mirtron pathway generates pre-miRNA mimics by splicing of short introns with hairpin potential. The initial spliced intron product is not linear, as with all spliced introns, but instead a lariat in which the 3' branchpoint is ligated to the 5' end of the intron. However, following resolution of this structure by lariat debranching enzyme, the intron can adopt a pre-miRNA fold and be transferred to the cytoplasm via Exportin-5, cleaved by Dicer, and loaded into Ago for target regulation [27, 28] (Figure 1B).

Mirtrons were originally recognized in flies and worms, but similar loci (i.e., short hairpin introns associated with small RNA reads extending to intronic termini) were later recognized in rodents and primates [29, 30], chicken [31], cow [32], and even rice [33]. The independence of mirtrons from Drosha and DGCR8 has now been validated by true genetic tests, since mirtron-derived small RNAs persist in Northern blots or libraries prepared from the corresponding *Drosophila* [34] and mouse [30, 35, 36] mutants. In fact, small RNA libraries from *drosha* and *dgcr8* mutants appear to be enriched in mirtron-derived reads. It remains to be clearly established whether dicing of mirtrons is actually enhanced in these conditions, perhaps due to loss of abundant competing substrates from canonical miRNA loci. An alternative possibility is that it represents a normalization effect owing to the loss of the abundant canonical miRNAs reads in *drosha/dgcr8* mutants. After all, something has to be sequenced in these libraries, so heterogeneous degradation products may also tend to be over-represented in these conditions. This may be controlled by careful selection of reference short RNAs for normalization, beyond simply normalizing to total mapped reads [30]. Such an approach has also been useful for the discovery of other small RNA derived from non-canonical pathways.

A particularly strict set of functional genetic tests was conducted in *Drosophila*, using a transgenic system in which repression of a *GFP* sensor bearing complementary sites to canonical or mirtron-derived miRNAs could be monitored *in vivo* [34]. In homozygous clones of cells bearing a null allele for the miRNA-generating enzyme encoded by *dicer-1*, neither a canonical miRNA or a mirtron-derived miRNA could repress their targets, reflecting their shared requirements for dicing. In contrast, cell clones homozygous for a null allele of *pasha*, encoding the obligate Drosha cofactor, failed to repress via a canonical miRNA but maintained strong activity of a mirtron [34]. These assays provide stringent evidence, *in vivo*, that mirtrons generate functional regulatory RNAs in cells that completely lack the canonical miRNA pathway.

4. "Add-on" mirtron pathways: 5' and 3' tailed mirtron loci

With conventional mirtron loci, the resultant small RNAs begin and end precisely with splice donor and splice acceptor sites. In other words, both ends of the pre-miRNA are generated by the splicing reaction. However, certain mirtron-like loci have been annotated where the small RNA-generating hairpin resides at one end of the intron. Two flavors of "tailed" mirtrons have been found, in which the unstructured extensions are found either 5' or 3' to the hairpin [27, 30]. The existence of such loci implies that the splicing machinery generates an RNA intermediate that must undergo additional nucleolytic processing, prior to eventual dicing.

The biogenesis of 3' tailed mirtrons was elucidated with respect to *mir-1017*, a locus that is highly conserved amongst the Drosophilids. The intronic sequence following the *mir-1017* hairpin comprises a ~100 nt tail extending to the splice acceptor site [37]. As with canonical miRNAs and conventional mirtrons, *mir-1017* can repress seed-matched target transcripts, indicating that this locus generates a miRNA-class regulatory RNA. After splicing and debranching, the 3' tail following the hairpin is then trimmed by the RNA exosome, the major eukaryotic 3'–>5' exonuclease complex (Figure 1C). Knockdowns of four different exosome subunits revealed a common accumulation of the linear, untrimmed *pri-mir-1017* intron [37]. However, *in vitro* reconstitution experiments indicated dependence of this reaction on the Rrp6 exonuclease, which is a specific component of the nuclear RNA exosome. This suggests that the trimming reaction occurs in the nucleus and is prerequisite to shorten the tail sufficiently to serve as an Exportin-5 substrate, as opposed to occurring in the cytoplasm to directly generate the Dicer substrate. Although it may seem dangerous to have an essential biogenesis step be carried out by a "professional" exonuclease complex, which normally degrades substrates entirely, the RNA exosome is known to be inhibited by stable secondary structures [38]. Indeed, *in vitro* processing reactions using reconstituted exosomes indicated their conversion of linear primary tailed mirtron substrates into stable pre-miRNAs [37].

mir-1017 (Figure 2B) is the only well-conserved 3' tailed mirtron currently known, but bioinformatic searches from deeply sequenced short RNA data revealed a family of other less-conserved 3' tailed mirtrons in *D. melanogaster* [37]. Curiously, no 3' tailed mirtrons have yet been identified in vertebrate species. Instead, a number of 5' tailed mirtrons (Figure 2C) have been found in chicken and various mammals [30, 32, 39]. Thus far, their biogenesis has not been studied in biochemical detail, except that analysis of appropriate mutant celltypes has established that at least some of these are Dicer-dependent, but Drosha/DGCR8-independent [30, 36]. Presumably the biogenesis of 5' tailed mirtrons involves a different pathway than their 3' tailed brethren. A potential candidate to remove the 5' tails might be XRN1/2, the major 5'->3' exonucleases in eukaryotes, although this remains to be tested.

5. Detection of mirtrons by experimental and computational methods

As discussed, the first mirtrons were discovered by virtue of cloned small RNAs that mapped precisely to the ends of short hairpin introns. In fact, the founding *C. elegans* mirtron *mir-62* was originally annotated as a canonical miRNA [40], and only later recognized as a mirtron [27] (Figure 2A). Initial scans of the fly and worm genomes using then-available short RNA data yielded 14 fly mirtrons and 4 worm mirtrons [27]. In *Drosophila*, where the large number of sequenced fly genomes facilitates recognition of characteristic evolutionary signatures, a strong similarity in patterns of constraint for conserved canonical miRNAs and mirtrons is apparent. In particular, initial computational analysis of canonical *Drosophila* miRNAs revealed that their terminal loop regions evolve much more quickly than their stems [41]. The same was subsequently observed for vertebrate canonical miRNAs [42], then for conventional mirtrons [28], and eventually with 3' tailed mirtrons [37]. Such evolutionary similarities provided impetus to infer that mirtrons generate miRNA-class regulatory RNAs, even before the formal experimental proof was in hand.

The observation of an evolutionary signature for conserved mirtrons suggested that their discovery using comparative genomics might be possible. A computational scan for intronic hairpins that were conserved between primates and rodents, and exhibited accelerated divergence in the terminal loop relative to the stems, yielded 13 candidates. Three of these loci were validated as genuine mirtrons by corresponding intron-terminal small RNAs

cloned in various primate and rodent species, indicating conservation over ~80 million years of evolution [29]. However, it has commonly been observed that substantial numbers of canonical miRNAs are not sufficiently conserved to be detected using phylogenetic analysis, even though they can be detected using deep sequencing [39, 43, 44]. Indeed, analysis of small RNAs from a panel of dissected portions of primate brains provided evidence for some 16 additional species-restricted mirtrons [29]. With the recent advent of deep sequencing, analysis of small RNA libraries has become the preferred method for small RNA discovery, including of mirtrons. In most of these cases, mirtrons were not specifically sought, but happened to be amongst the loci that could be annotated from large sequence data [39, 43–45].

In addition to elucidating the mature small RNA bounds and the capacity to detect even weakly-expressed loci with sufficiently large datasets, a key advantage of the sequencing approach is that it does not rely upon species conservation. While computational prediction of miRNAs has been a lively research area for the past decade, reliable prediction of canonical miRNAs is currently possible only when incorporating comparative genomics. To our knowledge, none of the many available methods achieves a reasonable balance of sensitivity and specificity that can obviate the need for sequence-based validation of predicted loci. In other words, it is simply not possible to obtain a working rough draft of the canonical miRNA content of a given genome using a *de novo* genefinder, the way one might run a protein-coding genefinder and expect to be able to make credible estimates of gene content.

The difficulty of canonical miRNA prediction is in large part due to the substantial search space of genome-encoded hairpins with plausible similarity to known, high-confidence miRNAs. Folding of invertebrate genomes can yield on the order of 100,000 candidate hairpins [41] and folding of mammalian genomes yields millions of such candidates [46]. It may seem patently unreasonable to suggest that there are millions of mammalian miRNAs, especially when extensive cloning and sequencing efforts provide support for substantially fewer than one thousand [39]. Nevertheless, it is not yet clear what specific features permit only certain hairpins to serve as miRNA substrates. Some of the best canonical miRNA prediction programs [47, 48] show reasonably high (~60–95%) sensitivity and specificity. However, since the number of hairpin candidates is so large, these levels of specificity translate into tens or hundreds of thousands false positive predictions genomewide.

The difficulty in purely bioinformatic assessment of candidate miRNAs is further highlighted by the study of editing events or single nucleotide polymorphisms (SNPs). In some cases, SNPs are associated with seemingly substantial structural changes, but have little effect on *in vivo* processing [49]. Other cases of SNPs or editing may induce modest changes in hairpin quality, but can abolish miRNA biogenesis [50, 51]. Until we can reliably predict how single nucleotide changes in characterized miRNA genes might affect processing, it seems that assessing the competence of hairpins of arbitrary sequence and structure to transit the Drosha-Dicer pathway will continue to be a challenge.

With this in mind, it is notable that mirtrons were recently shown be amenable to relatively effective prediction from genomic sequence alone [52]. This effort was aided by data gained from a panel of structure-function tests that defined the key features of functional mirtrons in *Drosophila* cells. In particular, the 3' overhang was shown to promote effective processing, with 2 nt-3' overhangs seemingly optimal, consistent with known conformation of Exportin-5 substrates. In addition, extension of the intron, either by introducing additional stem sequence or unstructured nucleotides, was inhibitory to mirtron processing. This implied that shorter hairpin introns were preferentially processed. Beyond this, mirtron biogenesis was tolerant of broad changes to loop nucleotides and overall structure, beyond

the need to maintain a decent level of hairpin structure. Interestingly, inspection of known mirtrons showed that a disproportionately high number of validated mirtrons had large internal loops (4–5 nt) that were comparatively rare amongst validated canonical miRNAs. This suggested that a highly progressive scheme for hairpin evaluation, shown to be effective for classification of canonical miRNAs [41, 53], might inappropriately penalize functional mirtrons.

This knowledge was applied in a supervised learning approach to rank *D. melanogaster* introns on the basis of their likelihood to be Dicer substrates and generate short RNAs [52]. Three general sets of features were used to train a support vector machine (SVM): overhang conformation, stem loop features (bulges, base composition, etc.) and structural similarity to known mirtrons vs. non-mirtrons. The SVM classifier performed well in genomewide rankings of ~27,600 introns 50–120 nt in length, identifying 23 mirtrons amongst the top 52 predictions validated by small RNA sequencing data; 8 additional candidates had some support from small RNA reads.

Despite the fact that many novel mirtrons were found amongst high-scoring loci, a potential concern was that a substantial portion of the training set comprised the known mirtrons, raising the possibility of over-fitting. This was addressed by running the classifier anew on the *C. elegans* genome, for which only 4 mirtrons had been previously annotated. In fact, 13 confident and 2 candidate mirtrons amongst the top 27 predictions (from ~30,600 total 50–120 nt introns) were validated from deep sequence data, thus substantially increasing the catalog of nematode mirtrons [52]. Therefore, computational modeling is a feasible strategy for identifying the mirtron subset of miRNAs, and characteristic features distinguish mirtrons from bulk introns even in the absence of comparative genomic data.

It is clear that computational prediction of mirtrons has certain advantages compared to canonical miRNAs. First, restricting the search for mirtrons to short (50–120 nt) introns of protein-coding genes dramatically reduces the search space, compared to canonical miRNAs for which candidate hairpins are spread throughout the genome. Second, with mirtrons the exact splice site and nature of the hairpin overhang are known; thus only cleavage by Dicer has to be modeled. In contrast, models for canonical miRNAs must assess the likelihood of being both a Drosha and a Dicer substrate, resulting in more complex models with a larger parameter space. A corollary of this is that if the capacity of hairpins to serve as Drosha substrates could be predicted accurately, assessment of canonical miRNAs using solely computational evaluation may become feasible. This concept was proposed earlier [47]; however, practical implementation remains to be demonstrated. Such efforts will undoubtedly be dependent on additional mechanistic knowledge for how the Drosha/DGCR8 complex selects its preferred substrates.

6. Evolutionary origin of mirtrons

It is clear that the molecular coupling of Dicer and Argonaute proteins for small RNA-mediated gene regulation was an ancient event during eukaryotic evolution. One notion is that this pathway may have its origins as a defense against invasive nucleic acids, such as transposable elements and viruses, which frequently have a signature dsRNA phase of their lifecycle. Perhaps subsequent to the assembly of the Dicer/Argonaute/small RNA pathway for genome defense, species may have recycled the capacity for endogenous gene regulation triggered by genome-encoded intramolecular dsRNAs encoded by inverted repeats. Although the small RNA products of endogenous inverted repeat transcripts comprise a wide variety of species produced by sundry biogenesis pathways, at the heart of it, these do include canonical and non-canonical miRNA genes [5].

However, while there exist miRNAs that are well-conserved amongst plants, miRNAs that are well-conserved amongst animals, and even miRNAs in certain fungi (e.g. *Neurospora*), there are no miRNAs that can be confidently assessed to be common across these branches of life [5]. Moreover, there are clear differences between plant and animal miRNA biogenesis, and fungal miRNA biogenesis is even stranger still. Therefore, it is seems reasonable to infer that the capacity for miRNA biogenesis has emerged several times, taking advantage of a pre-existing pathway that processes dsRNA via Dicer and loads the resulting short RNAs into Argonaute.

With only a single exception (rice *MIR1429*) [33], virtually all mirtrons have been annotated from animals. Certain mirtrons are well-conserved amongst Drosophilids, amongst nematodes, or amongst mammals; however, none are conserved across these animal clades. This is in clear contrast to canonical miRNAs, a number of which are perfectly constrained between invertebrates and vertebrates, and a few of which date back to early-branching bilaterians [54]. In addition, there appear to be substantial differences between *Drosophila* and mammalian mirtrons. In mammalian mirtrons the dominant ("mature") RNA species tend to originate from the 5' of the stem, whereas invertebrate mirtrons mostly produce 3' dominant species. Mirtrons in mammalian genomes are also much more GC-rich than invertebrate mirtrons, and thus form more stable hairpins. Finally, the overhangs differ between mammalian and invertebrate mirtrons: mirtrons in invertebrates most typically have 0:2 overhangs, whereas the proportion of mammalian mirtrons with 1:1 overhangs is higher [29].

These observations may suggest that the mirtron pathway has itself evolved independently in different animal clades, building upon the backbone of a pathway that exports short hairpins bearing 3' overhangs to the cytoplasm, where they are specifically recognized and processed by Dicer. In this way of thinking, the existence of mirtrons reflects parasitization of a pre-existing canonical miRNA pathway. The notion of independent birth of mirtron pathways in invertebrates and vertebrates is supported by the fact that they have apparently been further parasitized by different tailing pathways; 3' tailed in invertebates and 5' tailed in vertebrates. Taken together, these diverse pathways seem to reflect remarkable flexibility to mix and match RNA processing pathways that may otherwise seem rigid and obligate, especially when drawn as model figures that imply dogma.

Fly and worm genes happen to have a high fraction of short introns, which coincide in length with typical pre-miRNAs [27, 55]. This may suggest that these species have an innate capacity to generate pre-miRNA mimics via splicing, since random evolution of the population of short introns (numbering 27–30,000 in flies and worms) will eventually generate hairpins. This certainly seems plausible, and is supported by the observation that many computationally predicted mirtrons in *D. melanogaster* and *C. elegans* could be validated as being endogenously diced into short RNAs [52]. On the other hand, since mammalian genomes have many more introns than fly and worm, their smaller fraction of short introns still results in a large absolute number of mirtron candidates. Furthermore, the proportion of tailed mirtrons is greater in vertebrates than in invertebrates. Since splicing does not need to generate both ends of the pre-miRNA, this opens up the potential pool for substrate evolution further.

7. Evolutionary flux of mirtrons

If we consider that mirtrons and canonical miRNAs evolve according to their own paths, then their respective evolutionary flux need not be expected to be identical. Indeed, the data currently available suggests that mirtrons evolve at a higher rate than canonical miRNAs; thus, there is not a universal rate of miRNA flux. As mentioned, no mirtrons have been

found that are shared between mammals, flies and worms, in contrast to the many canonical miRNAs that are deeply conserved across the animal kingdom. Furthermore, detailed annotations of canonical miRNAs and mirtrons from deeply sequenced small RNAs from three species of Drosophilds clearly indicated that mirtrons evolve faster than miRNAs [43]. In particular, while mirtrons comprised only a small fraction of the deeply conserved miRNAs shared between all the sequenced flies, the proportion of mirtrons that were progressively more species-restricted progressively increased [43].

A possible explanation for the more rapid evolutionary flux of mirtrons, compared to miRNAs, lies in their different biogenesis strategies. For a sequence to enter the canonical miRNA biogenesis pathway, it has to simultaneously adapt its conformation to become a substrate of Drosha as well as of Dicer. Since there are at most a few hundred known Drosha substrates in most animal species, becoming an efficient Drosha substrate may be a substantial hurdle. Indeed, the model that DGCR8 recruits Drosha to cleave near the junctions of single-stranded and double-stranded RNA [14] does not seem sufficient to explain their apparent specificity for bona fide miRNA substrates from amongst a forest of structure junctions across the transcriptome. There may exist additional necessary features, perhaps that may only become apparent at the level of three-dimensional hairpin structures. On the other hand, there are substantial pools of 10,000s of constitutively spliced short introns, from which Dicer substrates might emerge by random mutational processes.

Ultimately, for either a canonical miRNA or a mirtron to become evolutionarily fixed, it must acquire beneficial target interactions that outweigh their detrimental regulatory effects. It has been proposed that newly-evolved miRNAs likely have mostly detrimental effects, assuming that overall networks of gene regulation are honed to an optimal state by evolution, and introduction of a novel regulatory RNA is likely to induce inappropriate repression [56, 57]. However, under the right environmental or stress conditions, one can imagine that beneficial regulatory interactions might emerge, and eventually be subject to positive selection for increased biogenesis capacity of the miRNA, concomitant with purging of deleterious target sites. Such considerations apply equally to both canonical miRNAs and mirtrons, the difference being that mirtrons may have a tendency to arise more quickly. However, it must evidently be more difficult for mirtrons to be successfully stabilized into regulatory circuits, because mirtrons are also extinguished more quickly than canonical miRNAs [43]. The evolutionary underpinning to this difference remains to be understood.

8. Regulation by mirtrons *in trans* and *in cis*?

Biochemical studies have shown that the function of mirtrons is indistinguishable from that of miRNAs, with regard to repression of target genes *in trans*. To date, only a few endogenous targets of mirtrons have been validated *in vivo*. One potentially interesting example is that the tailed mirtron product miR-1017, which is specifically expressed from the intron of a neural-specific acetycholine receptor, directly represses *yan*, which encodes a repressor of neurogenesis [37]. This is reminiscent of the mutually repressive relationship between yan and the canonical miRNA miR-7 that is expressed in differentiating neural territories [58]. Otherwise, conserved targets of miRNAs derived from conserved mirtrons can be easily predicted as with canonical miRNAs. Although one must be cautious in interpreting the phenotypic relevance of lists of predicted targets, the observation of specifically conserved seed matches across a large cohort of genomes is a sensitive indicator for regulatory sites that are subject to functional purifying selection. For instance, the TargetScan rubric [59], which focuses upon conserved seed matches in mRNAs, predicts on average 87 target genes for the fly mirtrons (ranging from 10 to 388 genes). Still, *in vivo* evidence of mirtron functions from knockouts remain to be studied, as no mirtrons have yet

been specifically deleted. Care will need to be taken to avoid disruption of host mRNA processing, as with intronic canonical miRNAs [60]; this may prove even more of an issue considering how short mirtron introns are.

It is worth noting that the average mirtron yields substantially fewer mature reads in small RNA libraries, relative to canonical miRNAs. Tissue specificity and/or regulated processing might confound the comparison of expression levels of any individual loci. However, aggregate analysis of several hundred million mapped *D. melanogaster* reads across a wide diversity of developmental stages and tissue/cell types indicated that the highest expressed canonical miRNAs generate ~1000-fold as many reads as the highest expressed mirtrons [61]. By this metric, it is expected that mirtrons will in general have more modest roles than canonical miRNAs. Nevertheless, it is clear that a number of Drosophilid and mammalian mirtrons have been strictly conserved over tens of millions of years of evolution [28, 29], and these presumably have been selected for some biologically relevant functions. Reciprocally, the relatively high proportion of recently-evolved mirtrons, compared to canonical miRNAs, suggests that they may potentially influence species-specific characteristics.

In addition to *trans*-regulatory interactions, one may wonder whether mirtron biogenesis has any consequences *in cis*. Thus far, most annotated mirtrons seem to derive from constitutive splicing events. An exception is the *Drosophila* mirtron *mir-2494*, which is produced from an alternatively spliced intron of *CG17560*. It remains to be tested what effect, if any, the biogenesis of miR-2494 has on its host mRNA. Curiously, though, usage of the alternative splice site giving rise to the mirtron is predicted to place the remainder of *CG17560* out of frame [52]. It is plausible that the act of *mir-2494* biogenesis is used to regulate the translation of *CG17560*, either to inhibit the production of the full-length protein or alternatively to engage a shorter polypeptide product. Conversely, one may hypothesize that splicing regulation might conceivably serve to regulate mirtron biogenesis. Finally, there are many documented cases in which intronic basepaired structures serve to modulate splicing efficiency [62–64]. It is not known whether such regulatory possibilities exist for mirtrons, but they deserve experimental investigation.

9. Conclusions and Prospects

The elucidation of the mirtron pathway not only uncovered a strategy for miRNA biogenesis broadly applicable across animal (and even one plant) species, it provided a precedent for searches of additional non-canonical miRNA pathways. This has indeed proven to be a rich field of study, having produced a diversity of Drosha-independent pathways including other strategies that exploit aspects of snoRNA or tRNA processing for miRNA biogenesis [26]. Even more unexpectedly, a Dicer-independent pathway for miRNA biogenesis involving direct cleavage of short hairpins by Ago2 was elucidated most recently [65–67]. Certainly additional pathways remain to be characterized, including the nucleases responsible for 5' tailed mirtron biogenesis.

The demonstration that mirtrons are the first class of metazoan miRNA to be amenable to reasonably accurate computational discovery without consideration for evolutionary conservation is a promising advance [52]. One can already imagine that improvements may be had over the initial efforts. For example, retraining with the larger set of validated mirtrons now available may improve recall performance. There are certainly many additional Drosophild and nematode species that could serve as new testbeds to rule out overfitting. Certainly, a comprehensive analysis of species-restricted mirtrons across these invertebrate genomes could provide great insights into the dynamics of mirtron evolution. In addition, some mirtrons appear to derive from non-coding RNAs and unannotated introns

[44, 52]. Therefore, it may be desirable to explore whether mirtrons can be discovered without relying upon precomputed locations of splice sites. This may provide a more fair assessment of "genomewide" predictions and the efficacy of mirtron discovery. Finally, it remains to be seen whether appropriate models can be developed for the computational discovery of vertebrate mirtrons.

On a final note, it is worth reflecting on how small RNA biogenesis interfaces with longer aspects of the transcriptome, beyond the facts that pri-miRNA transcripts are "long" and the fact that many canonical miRNAs are embedded in introns of protein-coding genes. Unlike the latter, mirtron biogenesis is directly linked to mRNA maturation, providing an unexpected precedent of the intersection of these RNA worlds. There is now also a growing appreciation of the possibility of Drosha cleavage of mRNAs [36, 68–70], and even the production of miRNAs from coding sequences of mRNAs [61]. A class of endogenous siRNAs derives from the double-stranded regions of hundreds of convergently-transcribed genomic loci, especially from overlapping 3' UTRs [71–75]. Reciprocally, endogenous siRNAs can be generated along the length of gene bodies in mammalian oocytes, when subject to pairing to antisense transcribed pseudogenes [76, 77]. Finally, piRNA production from the 3' UTRs of thousands of *Drosophila* and mammalian protein-coding transcripts in certain gonadal contexts has been shown [78, 79]. Altogether, these myriad intersections between the short and long transcriptomes provide glimpses into a complex and interleaved organization of the genome, and diverse networks of *cis*- and *trans*-regulation that we are only beginning to understand.

Acknowledgments

The authors would like to thank Jaaved Mohammed for helpful discussions. J.O.W. was supported by a fellowship from the Swedish Research Council. Work in E.C.L.'s group was supported by the Burroughs Wellcome Fund, the Starr Cancer Consortium (I3-A139) and the NIH (R01-GM083300).

References

1. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009; 10:126–139. [PubMed: 19165215]
2. Czech B, Hannon GJ. Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet. 2010; 12:19–31. [PubMed: 21116305]
3. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011; 39:D152–D157. [PubMed: 21037258]
4. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP. Prediction of plant microRNA targets. Cell. 2002; 110:513–520. [PubMed: 12202040]
5. Axtell MJ, Westholm JO, Lai EC. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 2011; 12:221–221. 213. [PubMed: 21554756]
6. Lai EC. microRNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet. 2002; 30:363–364. [PubMed: 11896390]
7. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of MicroRNA-Target Recognition. PLoS Biol 3. 2005:e85.
8. Doench JG, Sharp PA. Specificity of microRNA target selection in translational repression. Genes Dev. 2004; 18:504–511. [PubMed: 15014042]
9. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136:215–233. [PubMed: 19167326]
10. Flynt AS, Lai EC. Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet. 2008; 9:831–842. [PubMed: 18852696]
11. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Res. 2004; 14:1902–1910. [PubMed: 15364901]

12. Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA. 2005; 11:241–247. [PubMed: 15701730]
13. Kim YK, Kim VN. Processing of intronic microRNAs. Embo J. 2007; 26:775–783. [PubMed: 17255951]
14. 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]
15. Zhang X, Zeng Y. The terminal loop region controls microRNA processing by Drosha and Dicer. Nucleic Acids Res. 2011; 38:7689–7697. [PubMed: 20660014]
16. 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; 17:1850–1864. [PubMed: 17989254]
17. Okada C, Yamashita E, Lee SJ, Shibata S, Katahira J, Nakagawa A, Yoneda Y, Tsukihara T. A high-resolution structure of the pre-microRNA nuclear export machinery. Science. 2009; 326:1275–1279. [PubMed: 19965479]
18. 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]
19. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003; 115:209–216. [PubMed: 14567918]
20. 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; 15:354–363. [PubMed: 18376413]
21. Ro S, Park C, Young D, Sanders KM, Yan W. Tissue-dependent paired expression of miRNAs. Nucleic Acids Res. 2007; 35:5944–5953. [PubMed: 17726050]
22. Yang JS, Phillips MD, Betel D, Mu P, Ventura A, Siepel AC, Chen KC, Lai EC. Widespread regulatory activity of vertebrate microRNA* species. RNA. 2011; 17:312–326. [PubMed: 21177881]
23. Okamura K, Liu N, Lai EC. Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. Mol Cell. 2009; 36:431–444. [PubMed: 19917251]
24. Ghildiyal M, Xu J, Seitz H, Weng Z, Zamore PD. Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA. 2010; 16:43–56. [PubMed: 19917635]
25. Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ. Hierarchical rules for Argonaute loading in Drosophila. Mol Cell. 2009; 36:445–456. [PubMed: 19917252]
26. Miyoshi K, Miyoshi T, Siomi H. Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. Mol Genet Genomics. 2010; 284:95–103. [PubMed: 20596726]
27. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature. 2007; 448:83–86. [PubMed: 17589500]
28. 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]
29. Berezikov E, Chung W-J, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. Mol Cell. 2007; 28:328–336. [PubMed: 17964270]
30. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. 2008; 22:2773–2785. [PubMed: 18923076]
31. 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; 18:957–964. [PubMed: 18469162]
32. Glazov EA, Kongsuwan K, Assavalapsakul W, Horwood PF, Mitter N, Mahony TJ. Repertoire of bovine miRNA and miRNA-like small regulatory RNAs expressed upon viral infection. PLoS ONE. 2009; 4:e6349. [PubMed: 19633723]

33. Zhu QH, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, Helliwell C. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. 2008; 18:1456–1465. [PubMed: 18687877]
34. Martin R, Smibert P, Yalcin A, Tyler DM, Schaefer U, Tuschl T, Lai EC. A Drosophila pasha mutant distinguishes the canonical miRNA and mirtron pathways. Mol Cell Biol. 2009; 29:861–870. [PubMed: 19047376]
35. Yi R, Pasolli HA, Landthaler M, Hafner M, Ojo T, Sheridan R, Sander C, O'Carroll D, Stoffel M, Tuschl T, Fuchs E. DGCR8-dependent microRNA biogenesis is essential for skin development. Proc Natl Acad Sci U S A. 2009; 106:498–502. [PubMed: 19114655]
36. Chong MM, Zhang G, Cheloufi S, Neubert TA, Hannon GJ, Littman DR. Canonical and alternate functions of the microRNA biogenesis machinery. Genes Dev. 2010; 24:1951–1960. [PubMed: 20713509]
37. Flynt AS, Chung WJ, Greimann JC, Lima CD, Lai EC. microRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol Cell. 2010; 38:900–907. [PubMed: 20620959]
38. Liu Q, Greimann JC, Lima CD. Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell. 2006; 127:1223–1237. [PubMed: 17174896]
39. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010
40. Lau N, Lim L, Weinstein E, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science. 2001; 294:858–862. [PubMed: 11679671]
41. Lai EC, Tomancak P, Williams RW, Rubin GM. Computational identification of *Drosophila* microRNA genes. Genome Biol. 2003; 4:R42.41–R42.20. [PubMed: 12844358]
42. 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; 120:21–24. [PubMed: 15652478]
43. Berezikov E, Liu N, Flynt AS, Hodges E, Rooks M, Hannon GJ, Lai EC. Evolutionary flux of canonical microRNAs and mirtrons in Drosophila. Nat Genet. 2010; 42:6–9. [PubMed: 20037610]
44. Jan CH, Friedman RC, Ruby JG, Bartel DP. Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature. 2010
45. Yao J, Wang Y, Wang W, Wang N, Li H. Solexa sequencing analysis of chicken pre-adipocyte microRNAs. Biosci Biotechnol Biochem. 2011; 75:54–61. [PubMed: 21228487]
46. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet. 2005; 37:766–770. [PubMed: 15965474]
47. Helvik SA, Snove O Jr, Saetrom P. Reliable prediction of Drosha processing sites improves microRNA gene prediction. Bioinformatics. 2007; 23:142–149. [PubMed: 17105718]
48. van der Burgt A, Fiers MW, Nap JP, van Ham RC. In silico miRNA prediction in metazoan genomes: balancing between sensitivity and specificity. BMC Genomics. 2009; 10:204. [PubMed: 19405940]
49. Diederichs S, Haber DA. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. Cancer Res. 2006; 66:6097–6104. [PubMed: 16778182]
50. Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep. 2007; 8:763–769. [PubMed: 17599088]
51. Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Hum Mol Genet. 2007; 16:1124–1131. [PubMed: 17400653]
52. Chung WJ, Agius P, Westholm JO, Chen M, Okamura K, Robine N, Leslie CS, Lai EC. Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. Genome Res. 2011; 21:286–300. [PubMed: 21177960]
53. Lim L, Lau N, Weinstein E, Abdelhakim A, Yekta S, Rhoades M, Burge C, Bartel D. The microRNAs of Caenorhabditis elegans. Genes Dev. 2003; 17:991–1008. [PubMed: 12672692]

54. Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D. Ancient animal microRNAs and the evolution of tissue identity. Nature. 2010; 463:1084–1088. [PubMed: 20118916]
55. Lim LP, Burge CB. A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci U S A. 2001; 98:11193–11198. [PubMed: 11572975]
56. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Genet. 2004 May.:396–400.
57. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet. 2007; 8:93–103. [PubMed: 17230196]
58. Li X, Carthew RW. A microRNA Mediates EGF Receptor Signaling and Promotes Photoreceptor Differentiation in the Drosophila Eye. Cell. 2005; 123:1267–1277. [PubMed: 16377567]
59. 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]
60. Osokine I, Hsu R, Loeb GB, McManus MT. Unintentional miRNA ablation is a risk factor in gene knockout studies: a short report. PLoS Genet. 2008; 4:e34. [PubMed: 18282110]
61. Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung JH, Okamura K, Dai Q, Bortolamiol-Becet D, Martin R, Zhao Y, Zamore PD, Hannon GJ, Marra MA, Weng Z, Perrimon N, Lai EC. Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. Genome Res. 2011; 21:203–215. [PubMed: 21177969]
62. Warf MB, Berglund JA. Role of RNA structure in regulating pre-mRNA splicing. Trends Biochem Sci. 2010; 35:169–178. [PubMed: 19959365]
63. Kreahling JM, Graveley BR. The iStem, a long-range RNA secondary structure element required for efficient exon inclusion in the Drosophila Dscam pre-mRNA. Mol Cell Biol. 2005; 25:10251–10260. [PubMed: 16287842]
64. Graveley BR. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell. 2005; 123:65–73. [PubMed: 16213213]
65. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 2010; 465:584–589. [PubMed: 20424607]
66. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson N, Wolfe S, Giraldez AJ. A Novel miRNA Processing Pathway Independent of Dicer Requires Argonaute2 Catalytic Activity. Science. 2010; 328:1694–1698. [PubMed: 20448148]
67. Yang JS, Maurin T, Robine N, Rasmussen KD, Jeffrey KL, Chandwani R, Papapetrou EP, Sadelain M, O'Carroll D, Lai EC. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A. 2010; 107:15163–15168. [PubMed: 20699384]
68. Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, Yang WY, Haussler D, Blelloch R, Kim VN. Posttranscriptional crossregulation between Drosha and DGCR8. Cell. 2009; 136:75–84. [PubMed: 19135890]
69. Triboulet R, Chang HM, Lapierre RJ, Gregory RI. Post-transcriptional control of DGCR8 expression by the Microprocessor. RNA. 2009; 15:1005–1011. [PubMed: 19383765]
70. Kadener S, Rodriguez J, Abruzzi KC, Khodor YL, Sugino K, Marr MT 2nd, Nelson S, Rosbash M. Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. RNA. 2009; 15:537–545. [PubMed: 19223442]
71. 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; 15:581–590. [PubMed: 18500351]
72. 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]
73. 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; 453:793–797. [PubMed: 18463636]

74. 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; 320:1077–1081. [PubMed: 18403677]

75. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, Di Stefano L, Berezikov E, Brown CD, Candeias R, Carlson JW, Carr A, Jungreis I, Marbach D, Sealfon R, Tolstorukov MY, Will S, Alekseyenko AA, Artieri C, Booth BW, Brooks AN, Dai Q, Davis CA, Duff MO, Feng X, Gorchakov AA, Gu T, Henikoff JG, Kapranov P, Li R, Macalpine HK, Malone J, Minoda A, Nordman J, Okamura K, Perry M, Powell SK, Riddle NC, Sakai A, Samsonova A, Sandler JE, Schwartz YB, Sher N, Spokony R, Sturgill D, van Baren M, Wan KH, Yang L, Yu C, Feingold E, Good P, Guyer M, Lowdon R, Ahmad K, Andrews J, Berger B, Brenner SE, Brent MR, Cherbas L, Elgin SC, Gingeras TR, Grossman R, Hoskins RA, Kaufman TC, Kent W, Kuroda MI, Orr-Weaver T, Perrimon N, Pirrotta V, Posakony JW, Ren B, Russell S, Cherbas P, Graveley BR, Lewis S, Micklem G, Oliver B, Park PJ, Celniker SE, Henikoff S, Karpen GH, Lai EC, Macalpine DM, Stein LD, White KP, Kellis M. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science. 2010

76. 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; 453:534–538. [PubMed: 18404147]

77. 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; 453:539–543. [PubMed: 18404146]

78. Robine N, Lau NC, Balla S, Jin Z, Okamura k, Kuramochi-Miyagawa S, Blower MD, Lai EC. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr Biol. 2009; 19:2066–2076. [PubMed: 20022248]

79. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature. 2009; 461:1296–1299. [PubMed: 19812547]

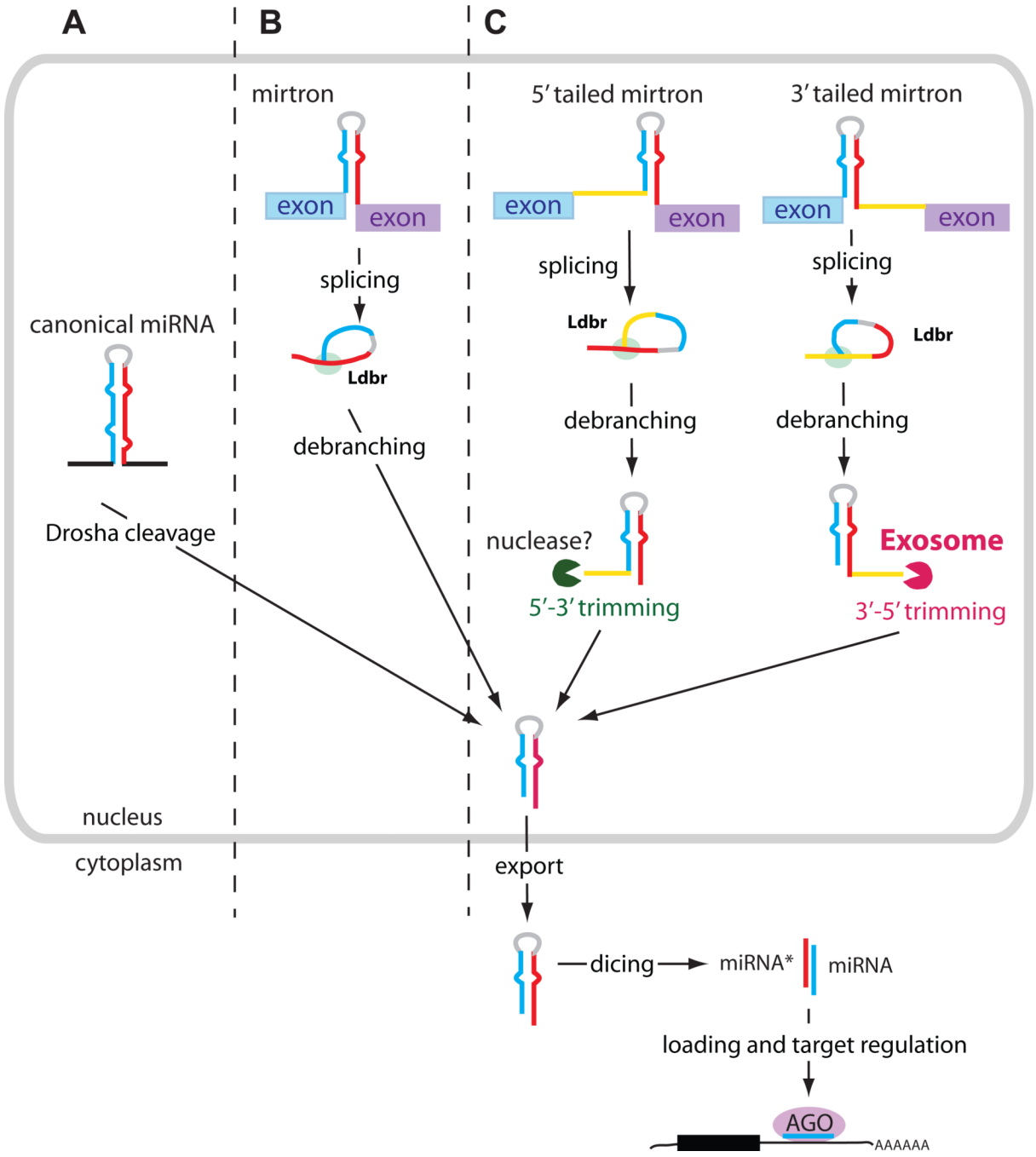

Figure 1.

Schematic overview of canonical miRNA and mirtron biogenesis. (A) The canonical miRNA pathway produces pre-miRNAs by Drosha cleavage of pri-miRNA transcripts. (B) Introns entering the mirtron pathway are spliced and debranched by lariat debranching enzyme (Ldbr), after which they fold into pre-miRNA hairpins. (C) Tailed mirtrons also undergo splicing and debranching, after which the tails on the resulting hairpins are trimmed back. 3’ tails are trimmed by the RNA exosome, while the enzymes responsible for 5’ trimming are not known. All of these pathways generate pre-miRNA hairpins, whose subsequent steps of nuclear export, Dicer cleavage and loading into Argonaute complexes are shared.

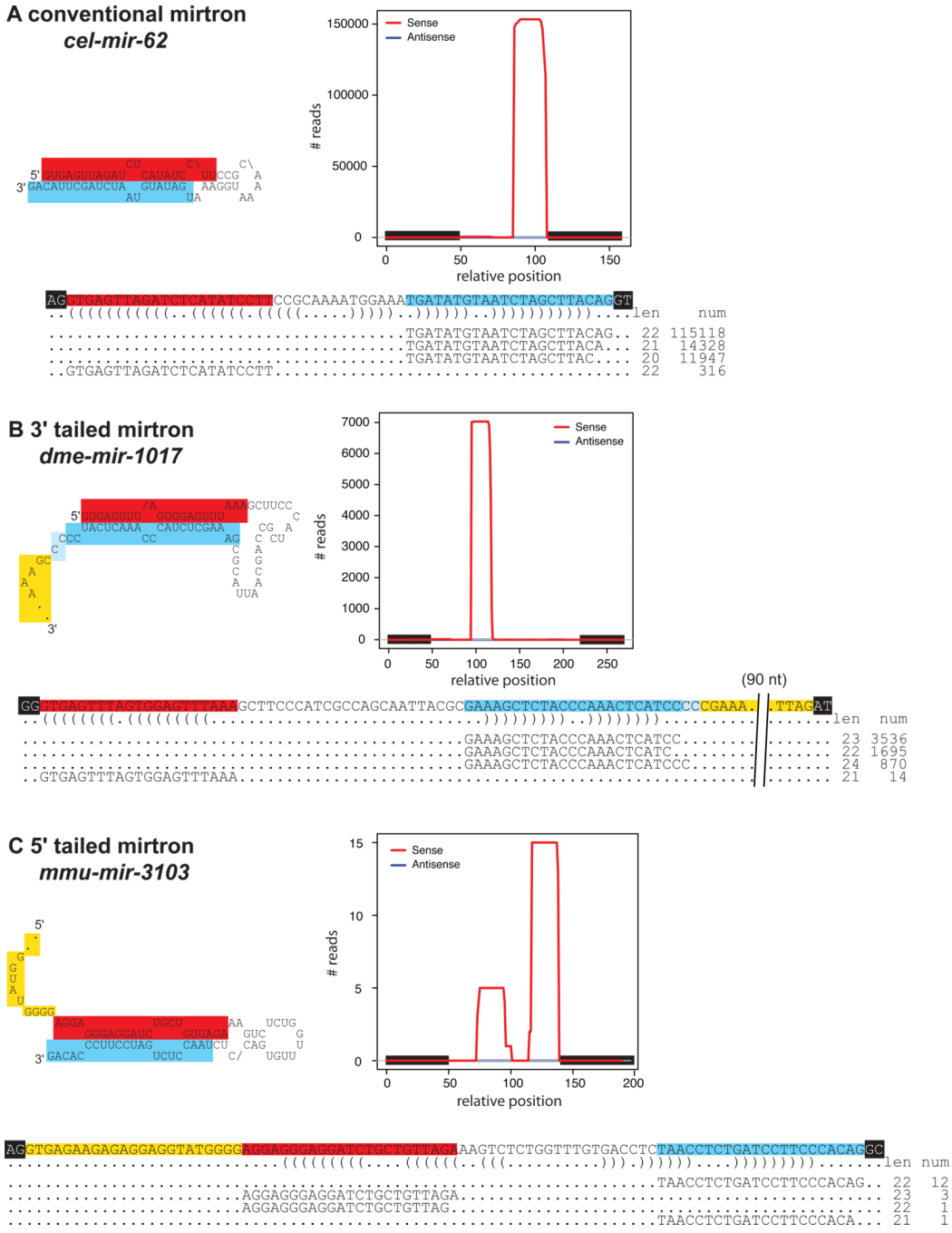

Figure 2.

Examples of conventional and tailed mirtrons. (A) *C. elegans mir-62* generates small RNA reads extending to both splice donor and acceptor sites, and the miRNA/star duplex exhibits a 3' overhang on the terminal loop side indicating Dicer cleavage of its precursor hairpin. The pre-miRNA hairpin displays the 0:2 overhang characteristic of invertebrate mirtrons. (B) *Drosophila mir-1017* is a 3' tailed mirtron whose precursor intron exhibits a pre-miRNA hairpin initiating with the GUGAGU splice donor site, followed by a ~100 nucleotide tail on the 3' end. Otherwise, the properties of its cloned short RNAs are similar to conventional mirtrons. (C) Mouse *mir-3103* is a 5' tailed mirtron, with a 23 nt 5' tail prior to a pre-miRNA hairpin that extends to the AG splice acceptor. In all schematics, the mature RNA

species are highlighted in blue, the miRNA* in red, the terminal loops in gray, the tailed regions in yellow, and the flanking exons in black. Cloned reads for *cel-mir-62* were compiled earlier [52], cloned reads for *dme-mir-1017* and *mmu-mir-3103* are from miRBase v16 (http://www.mirbase.org). To highlight the specificity of Dicer processing, only the most abundant mature miRNA and miRNA* reads are shown in the alignments. Since less abundant reads are not shown, the total read numbers in the graphs are greater than those tallied in the alignments. The graphs also reflect that almost all the short RNA reads derive from the same strand as host mRNA transcription.