MicroRNAs: deviants no longer

Amy E. Pasquinelli

Almost ten years ago, the Ambros laboratory made the extraordinary discovery that a gene essential for development in *Caenorhabditis elegans* encoded a 22-nucleotide, untranslated RNA. Further genetic studies in this nematode revealed the existence of a second tiny RNA gene that turned out to be conserved in animals as diverse as flies and humans. Now, the Ambros, Bartel and Tuschl laboratories have proven that those odd RNAs were just the first examples of a large family of RNAs, termed microRNAs (miRNAs). Although untranslated RNA genes, such as transfer RNAs and ribosomal RNAs, perform essential housekeeping roles in all living organisms, growing numbers of other RNAs, some widely conserved across phyla and others limited to certain species, are being uncovered and shown to fulfill specific duties. The discovery of miRNAs establishes a new class of regulatory RNAs and highlights the existence of unexpected RNA genes that, although ancient, are not extinct.

Deciphering the human, *Drosophila* and *Caenorhabditis elegans* genomes not only revealed a wealth of protein-coding genes, but also exposed long stretches of seemingly barren sequence. Now, however, we can reconsider the contents of such 'junk DNA' sequences in the light of recent reports that a new class of non-coding RNA genes are scattered, perhaps densely, throughout these animal genomes. Although the tiny size of the RNA products of these genes, aptly called microRNAs (miRNAs), probably contributed to their previous obscurity, it also motivated their present discovery by three different groups [1–3]. Before these reports, two *C. elegans* genes, *lin-4* and *let-7*, had been uncovered through genetic screens and found to encode ~22-nucleotide (nt) regulatory RNAs [4,5]. The subsequent discovery that *let-7* is also expressed as a ~22-nt RNA in flies, humans and apparently all bilaterally symmetric animals [6] came on the heels of reports that similarly sized RNAs were implicated in post-transcriptional gene silencing (PTGS) mechanisms in plants and dsRNA-mediated interference (RNAi) in animals [7–9]. So, following the hint that ~22 nt is a popular size for functional RNAs, the trio of labs successfully employed biochemical and computational approaches to mine nearly 100 new RNA genes from human, *Drosophila* and *C. elegans*. We now are faced with the task of understanding how these unusual miRNAs are generated and how they function.

How are miRNAs regulated?

Although predictions can be made on the basis of what we know about *lin-4* and *let-7*, there are bound to be exceptions and surprises. Fitting with their roles in regulating the timing of developmental events, expression of the *lin-4* and *let-7* RNAs is temporally controlled [4–6]. However, *cis*-acting elements that dictate this expression, and the polymerase responsible for synthesizing these RNAs, remain to be determined. Now that we have expanded the dataset from two to almost 100 miRNA genes, common elements could emerge that help address these questions. Although expression of some of the new miRNAs is constitutive, that of others appears to be under developmental and tissue-specific control. Indeed, other yet-to-be-discovered miRNAs might only surface under certain environmental conditions.

One approach for defining *cis*-elements that control expression specificity is to compare miRNA genes that share conserved sequence and expression patterns. Likewise, by comparing cross-species developmental events that bracket miRNA expression we could gain insight into the mechanisms that control this specificity. For example, *let-7* RNA is first expressed at late larval stages in *C. elegans* and *Drosophila*, both members of the ecdysozoan clade [6]. In *Drosophila*, this expression correlates with the ecdysone spikes that drive metamorphosis. Thus, *let-7* expression might also be induced by a hormone circuit in *C. elegans*.

Although miRNA genes might harbor different elements for transcriptional regulation, many, if not all, encode longer RNAs that are precursors of the ~22-nt forms [1–3] (Fig. 1). Both *lin-4* and *let-7* are initially transcribed as ~70-nt, partially double-stranded RNAs that undergo processing by Dicer [10–12], an RNase with orthologs in species ranging

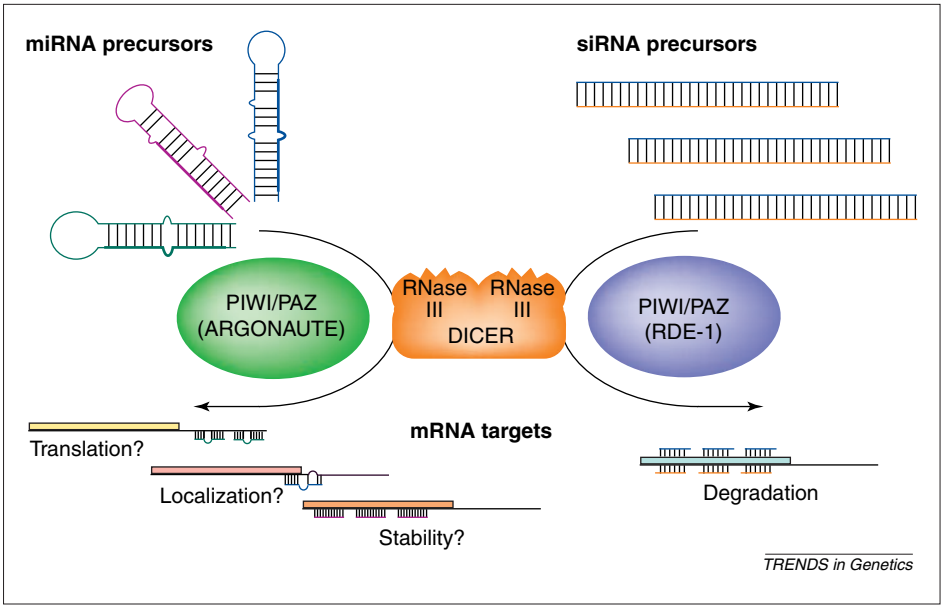

Fig. 1. Encoded ~70-nucleotide (nt) miRNA precursors are recruited to the RNase Dicer for processing to the mature, single-stranded ~22-nt forms. Likewise, long dsRNAs that initiate dsRNA-mediated interference (RNAi) undergo processing by Dicer to generate ~22-nt short interfering RNA (siRNA) duplexes. Distinct members of the PIWI/PAZ-domain protein family could facilitate processing and downstream functions of miRNAs and siRNAs. Protein-coding mRNAs that contain regions complementary to these tiny RNAs are targeted for post-transcriptional regulation. The siRNA duplexes cause degradation of their targets, but miRNAs could conceivably direct positive or negative regulation at a variety of levels depending upon the specific miRNA and target base-pair interactions and the co-factors that recognize them.

http://tig.trends.com 0168-9525/02/$ – see front matter PII: S0168-9525(01)02624-5

from *Arabidopsis* to humans [13]. Dicer first gained fame for its role in processing exogenous dsRNAs to the ~22-nt, short-interfering RNAs (siRNAs) that direct RNAi [13]. Thus, the potential stem-loop structure that can be formed by longer versions of each miRNA is probably critical for maturation to the ~22-nt form.

Some of the miRNAs cluster in genomic regions and could be transcribed as a single RNA that undergoes multiple processing steps. For example, seven miRNAs, miRNA-35 to miRNA-41, reside within a sequence of 800 nt on *C. elegans* chromosome 2 and exhibit expression restricted to eggs [1]. All seven are predicted to fold into non-overlapping stem-loop structures, and corresponding ~70-nt precursors can be detected, but it has yet to be shown whether these actually derive from a single transcript.

Curiously, none of the miRNA precursors exhibits perfect complementarity in the stem that forms the mature RNA, and typically only a half of the stem appears retained as the mature miRNA. This contrasts with the perfectly double-stranded RNAs that are delivered to Dicer for generation of the ~22-nt siRNAs, which are composed of both sense and antisense strands [14] (Fig. 1). Thus, additional factors are likely to be involved in recruiting superficially double-stranded RNAs to Dicer and in maintaining and channeling the Dicer products to specific pathways. Dicer shares a PAZ domain with members of a large family of proteins [15], some of which might fulfill these accessory roles. For example, in *C. elegans* RDE-1 (standing for 'RNAi defective') is essential for RNAi [16], whereas its paralogs ALG-1 and ALG-2 (standing for 'argonaute-like genes') function in generation of the *lin-4* and *let-7* RNAs that regulate developmental timing [10]. There are at least 20 additional PIWI/PAZ proteins in *C. elegans* that are candidates for operating with specific miRNAs.

What are the functions of miRNAs?

The *lin-4* and *let-7* RNAs negatively regulate the expression of mRNAs containing 3′ untranslated regions (UTRs) with sites complementary to the ~22-nt RNAs [4,5,17,18]. The *lin-4* target, *lin-14*, was uncovered by clever genetic analyses. Mutations in *lin-4* resulted in inappropriate reiterations of early larval cell division patterns and loss-of-function mutations in *lin-14* produced the opposite phenotype – the absence of those same early larval division patterns. But, gain-of-function mutations in *lin-14* caused phenotypes that mimic those in *lin-4* mutants, hinting that *lin-4* negatively regulates *lin-14*. The peculiar activating mutations mapped to deletions in the *lin-14* 3′UTR, and a closer look at those sequences revealed multiple sites with limited complementary to *lin-4* RNA.

The target of *let-7* negative regulation, *lin-41*, also came out of genetic analyses, and was then found to contain sites with imperfect complementarity to *let-7* in its 3′UTR. This time, the target was identified as a suppressor of mutations in *let-7* RNA – insufficient *let-7* RNA can be partially compensated for by decreased expression of the protein-coding gene it negatively regulates. Thus, determining targets of miRNAs is much more complicated than just searching for antisense complementarity in mRNAs, because bulges and loops are not only tolerated, but seem to be the rule.

Instead, we could need to return to genetics to uncover the pathways regulated by each of the miRNAs. It might be particularly fruitful to scan for existing mutations that map near miRNAs, because the absence of open reading frames could have thwarted previous attempts to understand the molecular lesion of such mutants. Additionally, deletions that remove these tiny genes could be targeted in *Drosophila* or screened for in *C. elegans* to obtain mutants defective in expression of specific miRNAs. Unfortunately, the power of RNAi might not help in analyzing the function of miRNAs, as this approach has not succeeded in reproducing *lin-4* or *let-7* phenotypes. Although phenotypes resulting from deletions of specific miRNAs will illuminate which processes or pathways they regulate, subsequent suppression or enhancement of those phenotypes could be necessary to uncover the genes controlled by each miRNA. Genetic approaches to determine the targets of miRNAs are advantageous in that they avoid bias towards how the miRNA is controlling gene expression. At this point, it would be imprudent to predict that all miRNAs will negatively regulate expression of target mRNAs with complementary sequences in their 3′UTRs – the type of regulation could vary depending on the base-paired structure formed between the miRNA and its target, and this site of interaction could potentially be harbored anywhere in the mRNA transcript (Fig. 1).

In fact, the elegance of gene regulation that is dependent upon RNAs only just big enough to exert specificity is in its versatility. The type of control directed by each miRNA could relate to the base-paired structure it forms with its target, which could serve as a tag for positively or negatively regulating stability, processing, localization or translation of the mRNA (Fig. 1). For example, *lin-4* RNA causes decreased LIN-14 protein expression without affecting quantity or polysome occupance of *lin-14* mRNA [19]. Thus, *lin-4* apparently exerts negative regulation over its target at the translational level. But, this mechanism begs the question of why a cell would squander valuable translational machinery on an mRNA whose expression is forbidden. This problem is particularly disconcerting in light of the fact that the mechanism must be set up very early in development and apparently maintained for the duration of the nematode's life. Perhaps the advantage of this type of repression is its potential for rapid reversibility.

Conclusion

The ~100 miRNAs discovered in humans, *Drosophila* and *C. elegans* are probably the inaugural members of an extensive family of tiny RNA genes. Hundreds, perhaps thousands, more miRNA genes lurk in these genomes, and it would not be surprising to uncover them in the plant *Arabidopsis thaliana* or even the fungus *Schizosaccharomyces pombe*, given that genes designed to process RNAs to ~22 nt appear to be conserved in these species [15]. Although biochemical and genomic approaches could continue to be the most proficient ways to unearth new miRNAs, subsequent genetic studies will be imperative to tackle the question of what they do. After the founder of the miRNA gene family, *lin-4*, was discovered, the question was posed: was this unusual gene an emissary portending the existence of additional tiny RNA genes, or was it an isolated deviant borne of some genetic accident and restricted to nematodes [20]? Now the answer is pleasingly clear.

Acknowledgements

I thank Gary Ruvkun and the Ruvkun lab for stimulating discussions and insightful comments.

References

1 Lau, N.C. *et al.* (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862
2 Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294, 862–864
3 Lagos-Quintana, M. *et al.* (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858
4 Lee, R.C. *et al.* (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854
5 Reinhart, B.J. *et al.* (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906
6 Pasquinelli, A.E. *et al.* (2000) Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature*. 408, 86–89
7 Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952
8 Hammond, S.M. *et al.* (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293–296
9 Zamore, P.D. *et al.* (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101, 25–33
10 Grishok, A. *et al.* (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106, 23–34
11 Hutvagner, G. *et al.* (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* 293, 834–838
12 Ketting, R.F. *et al.* (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15, 2654–2659
13 Bernstein, E. *et al.* (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366
14 Elbashir, S.M. *et al.* (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15, 188–200
15 Cerutti, L. *et al.* (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem. Sci.* 25, 481–482
16 Tabara, H. *et al.* (1999) The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99, 123–132
17 Wightman, B. *et al.* (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862
18 Slack, F.J. *et al.* (2000) The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell.* 5, 659–669
19 Olsen, P.H. and Ambros, V. (1999) The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* 216, 671–680
20 Wickens, M. and Takayama, K. (1994) Deviants or emissaries. *Nature* 367, 17–18

Amy E. Pasquinelli
Dept of Molecular Biology, Wellman 8,
Massachusetts General Hospital,
50 Blossom St, Boston, MA 02114, USA.
e-mail: pasquina@frodo.mgh.harvard.edu

Modulating heteroplasmy

Patrick F. Chinnery

Patients with mitochondrial DNA (mtDNA) disease usually harbor a mixture of mutant and wild-type mtDNA (a state termed heteroplasmy), and the clinical features of the disease depend on the percentage of mutant mtDNA (the 'mutation load') in vulnerable tissues. Factors that modulate the mutation load are poorly understood, but recent work has started to unravel the mechanisms. In certain circumstances heteroplasmy might be regulated at the level of the individual mitochondrial genome.

Mitochondrial DNA (mtDNA) mutations are an important cause of human disease. More than 100 different mtDNA point mutations, and an even larger number of different mtDNA deletions have been catalogued over the last decade [1]. Together these mutations are found in at least one in 8000 adults [2], making mtDNA disorders among the most common inherited metabolic diseases. Clinical variability is the hallmark of mtDNA disease [3,4], and the mechanisms behind this variability are only just starting to become clear. In their recent work, Brendan Battersby and Eric Shoubridge [5] have taken an important step towards a complete understanding of how these common genetic defects cause such devastating clinical phenotypes.

The 16.5-kb human mitochondrial genome encodes 13 essential respiratory chain polypeptides and 24 RNAs that are required for intramitochondrial protein synthesis [6]. It therefore comes as no surprise that pathogenic mtDNA mutations impair mitochondrial respiratory chain activity. When the pathology of mtDNA disease is approached from the molecular end, it all seems very simple. The mutations affect respiratory chain function, this leads to a defect in cellular energy metabolism, and this subsequently causes a disease phenotype. Unfortunately, nothing could be further from the truth.

Different mtDNA mutations can cause markedly different disorders – some have a characteristic 'mitochondrial' phenotype (such as the progressive loss of eye muscle function in chronic progressive external ophthalmoplegia), but some can present with a common disorder (such as diabetes mellitus, or heart failure). To make matters more confusing, the same genetic defect can present with different clinical features, even within the same family [1]. Although the primary mtDNA defect is an essential requirement for the disease to occur, understanding the precise genetic defect is not sufficient, and other factors are clearly important.

The transmission of mtDNA heteroplasmy

Early studies of families transmitting heteroplasmic mtDNA point mutations showed major differences in mutation load between individual siblings within the same family, broadly corresponding to the severity of the clinical phenotype [7,8]. Our understanding of this process was greatly advanced by the generation of heteroplasmic mice by karyoplast (cytoplast) transfer [9–11]. In the mid 1990s, Eric Shoubridge and colleagues used this technique to develop mouse lines transmitting variable proportions of the NZB and BALB mtDNA sequence variants [9]. (N2B and BALB are inbred strains of laboratory mouse that have different mitochondrial genotypes.) These studies showed that the variation in mutation load between offspring of single female was established during early oocyte development, before the formation of the primary oocyte. It is currently thought that a restriction in the cellular copy number of mitochondrial genomes during early oocyte development (the mitochondrial genetic bottleneck) leads to rapid random genetic drift through the unequal partitioning of mitochondrial

 0168-9525/02/$ – see front matter PII: S0168-9525(01)02636-1