Non-micro-short RNAs: the new kids on the block

Bijan K. Dey, Adam C. Mueller, and Anindya Dutta
Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA 22908

ABSTRACT The advent of ultra–high-throughput sequencing has led to the discovery of a large group of small, noncoding RNAs that are not microRNAs. The functional relevance of microRNAs has been well established over the last decade. In this Perspective, we focus on the non-micro-short RNAs that comprise a variety of functional classes and range from 16–40 nucleotides in size. We will highlight how some of these non-micro-short RNAs were discovered, as well as their biogenesis, potential mechanisms of action, and role in diverse biological processes, development, and disease. Finally, we will describe what must be done to further our understanding of these enigmatic molecules.

INTRODUCTION

The study of RNA-mediated regulation of gene expression has greatly expanded our understanding of metazoan development and cell biology. In 1993, the first small, noncoding RNA, lin-4, was discovered to be a regulator of developmental timing by genetic screens in Caenorhabditis elegans (Lee et al., 1993; Wightman et al., 1993). Within 5 years, the discovery of RNA interference in the same system by Fire and Mello resulted in a paradigm shift in the roles of small noncoding RNAs within cells (Fire et al., 1998). Following these discoveries, several more classes of non-micro-short RNAs were identified in plants, suggesting additional roles for small RNAs in regulating cellular processes. In recent years, as a result of high-throughput sequencing and bioinformatics analyses, a vast array of non-micro-short RNAs has been identified in multiple organisms, including mammals, and classified by their biogenesis and function (see Figure 1 and Table 1 for details; see also Ghildiyal and Zamore, 2009). Some of these classes of non-micro-short RNAs are generated by Dicer and its homologues, but several others are not. Similarly, many but not all of these short RNAs associate with members of the Argonaute family of proteins to regulate target gene expression. The rapidly growing classes of non-micro-short RNAs have roles in diverse biological processes in eukaryotes ranging from fungi to humans (reviewed by Ghildiyal and Zamore, 2009). These molecules control gene expression through regulation of chromatin and DNA modifications, transcription, and RNA splicing, editing, translation, and turnover. It is becoming clear that non-micro-short RNAs play important roles in the regulation of development in higher eukaryotes. If non-micro-short RNAs become implicated in diseases, they and their synthesis pathways could be attractive targets for new therapies and diagnostic biomarkers. This Perspective will summarize our current knowledge of a few of these intriguing molecules from plants and animals and will describe how they are involved in surveillance and regulation of the genome and transcriptome.

siRNA-RELATED NONCODING RNAs

In plants, there are many categories of silencing RNAs that regulate endogenous genes, some of which share the canonical small interfering RNA (siRNA) and microRNA machinery, and others with unique genetic origins and processing enzymes. Beyond microRNAs and traditional siRNAs, trans-acting siRNAs (tasiRNAs), cis-acting siRNAs (casiRNAs), and natural antisense siRNAs (natsiRNAs) have been identified. Similar to microRNAs, casiRNAs and tasiRNAs are excised from PollI-transcribed precursor RNAs by Dicer homologues and form an RNA-induced silencing complex (RISC) and silence precursor RNAs in the cytoplasm (Chapman and Carrington, 2007). NatalsiRNAs, on the other hand, are transcribed by PollIV and converted into double-stranded RNA precursors by RNA-dependent RNA polymerase RDR6, after which they undergo cleavage into numerous short siRNA isoforms. NatalsiRNAs target nascent RNA at sites of active transcription during particular developmental stages or during cellular stress response (Borsani et al., 2005). Short RNAs have also been identified that can direct heterochromatin formation
While most of these RNA silencing mechanisms have thus far been observed in plants, the plethora of noncoding transcripts of unknown function being discovered in other eukaryotes suggests siRNA-based regulation of gene expression exists in other higher eukaryotes. One more recently observed parallel between plants and animals is that of endogenous siRNAs, which have now been identified in Drosophila (Czech et al., 2008) and mice (Watanabe et al., 2008). These siRNAs can be produced from duplex RNA formed by bidirectional transcription or from inverted repeat sequences, and regulate the expression of retrotransposons and protein-coding genes.

![Classification of small noncoding RNAs.](image)

**FIGURE 1:** Classification of small noncoding RNAs.

| Biogenesis | Name       | Source                | Length  | Function                                           | References                                      |
|------------|------------|-----------------------|---------|----------------------------------------------------|------------------------------------------------|
| Drosha, Dicer | miRNA    | Plants, algae, viruses, protists, animals | 17–25 nt | mRNA stability, translational repression of endogenous transcripts | Lee et al., 1993; Wightman et al., 1993; Ghildiyal and Zamore, 2009 |
| Dicer       | endosiRNA  | Plants, fungi, animals | 21–24 nt | Posttranscriptional regulation of mRNA and transposons during development | Borsani et al., 2005; Czech et al., 2008 |
|             | exosiRNA   | Plants, fungi, animals | >–24 nt  | Transcriptional, posttranscriptional regulation of mRNA and transposons, defense against viral genes | Chapman and Carrington, 2007 |
| DCL1, 2     | natsiRNA   | Plants               | 21–24 nt | Posttranscriptional gene silencing during development, salt tolerance, pathogen resistance | Borsani et al., 2005; Chapman and Carrington, 2007 |
| DCL3        | casiRNA    | Plants               | 24 nt    | Methylation of transposon loci in genome           | Chapman and Carrington, 2007 |
| DCL4        | tasiRNA    | Plants               | 21 nt    | Amplification of microRNA-initiated posttranscriptional gene silencing | Chapman and Carrington, 2007 |
| Non-Dicer   | piRNA      | Germ cells           | 26–31 nt | Transposon regulation, germ line development       | Aravin et al., 2006, 2007; Vagin et al., 2006; Brennecke et al., 2007; Houwing et al., 2007; Cheng et al., 2011; Banisch et al., 2012; Rajasethupathy et al., 2012 |
|             | piRNA-like | Drosophila, C. elegans | 24–30 nt | Unknown                                           | Ghildiyal and Zamore 2009 |
|             | rasiRNA    | Plants, Drosophila, animals | 26–31 nt | Genomic stability by regulating transposons, germ cell development | Aravin et al., 2006; Houwing et al., 2007 |
|             | tiRNAs     | Yeast, Drosophila, animals | 30–40 nt | Inhibition of translational initiation             | Ivanov et al., 2011; Li and Hu, 2012 |
|             | tRFs       | Drosophila, animals   | 17–26 nt | Cellular proliferation                             | Cole et al., 2009; Lee et al., 2009; Haussecker et al., 2010 |

**TABLE 1:** Types of small, noncoding RNAs and their cellular functions.
Piwi-INTERACTING RNAs (piRNAs)

piRNAs comprise the largest class of small, noncoding RNA molecules in animals, and as the name indicates, they bind to the piwi group of Argonaute proteins. The piRNAs are primarily involved in suppression of transposon expression and regulation of epigenetic modifications during germ cell development. piRNAs are 26–31 nucleotides (nt) long, map to distinct clusters in the genome, and are frequently encoded by repeat sequences, in which case they are alternatively designated repeat-associated siRNA (ra-siRNA; Aravin et al., 2006; Houwing et al., 2007). In addition to these repeat sequence-derived piRNAs, a class of piRNAs in vertebrates derived from nonrepetitive sequences has been revealed by high-throughput sequencing (Vagin et al., 2006). Mammalian piRNAs are divided into two subclasses: 26- to 28-nucleotide-long, prepauchytype piRNAs that associate with mammalian Piwi family members MILI or MIWI-2, and 29- to 31-nucleotide-long, pauchytype piRNAs that preferentially bind to MIWI proteins (Aravin et al., 2007; Brennecke et al., 2007). Pauchytype piRNAs are expressed in spermatogonia before meiosis, with a reduction in expression beginning during the midpauchyte stage, whereas pauchytype piRNAs are expressed in spermatogonia during the pauchyte stage of meiosis, later declining at the spermatid stage.

Primary piRNAs are produced from nonoverlapping genomic clusters that generate long, single-stranded RNA transcripts. The best-studied of these clusters is the flamenco locus in Drosophila, which acts as a repressor of gypsy, ZAM, and Idefix transposons. Although the mechanism of primary piRNA production is not well understood, the Drosophila RNA helicase Armitage, the putative nuclease Zucchini, and Yb protein are thought to be involved in piRNA production (reviewed by Banisch et al., 2012). After they are generated, piRNAs become 2′-O-methylated at their 3′ ends by HEN1 methyltransferase and are loaded onto Piwi/Aubergine, guiding those proteins to their transposon RNA targets to cleave them. The small cleavage products of the transposon RNAs are then recruited by Ago3 protein, bind to the piRNA primary transcripts, and subsequently generate more piRNAs. The means by which the Piwi-protein–mediated cleavage of transposon RNA targets amplifies the generation of secondary piRNAs is called the “ping pong” mechanism (Brennecke et al., 2007).

piRNAs had initially been shown to maintain germ line stem cells, in which they repress transposons and other repetitive elements by heterochromatin formation, as well as transcriptional and posttranscriptional silencing (reviewed by Banisch et al., 2012). Recently, specific piRNAs have been shown to be up-regulated in various cancers and to play a role in memory and learning (Cheng et al., 2011; Rajasethupathy et al., 2012). These findings beg the question of whether piRNAs might play a much more widespread role in normal development and various diseases, and it will be of much interest to study their function in a wider array of tissues and developmental contexts.

tRNA-DERIVED STRESS-INDUCED FRAGMENTS (tRFs)

A number of studies in the last few years have characterized a class of non-micro-short RNAs known as tRNA halves or tRFs (reviewed by Li and Hu, 2012). tRFs, produced in a variety of organisms ranging from yeast to humans during nutrient deprivation and oxidative or xenotoxic stress, inhibit translation (Ivanov et al., 2011). These ~30- and ~40-nucleotide non-micro-short RNAs (5′- and 3′-tRFs, respectively) are produced by cleavage of the anticodon loops of mature tRNAs by secreted ribonucleases, such as RNY1 in yeast and angiogenin in humans. The detailed mechanisms by which tRFs exert this inhibition are not well understood. However, their ability to induce the formation of stress granules or cytoplasmic foci suggests they may inhibit translational initiation. Intriguingly, transfection of 5′- but not 3′-tRF fragments results in global translational inhibition in mammalian cell lines. Interestingly, angiogenin is a neuroprotective factor, and may promote cell survival in part by producing these non-micro-short RNAs during times of cellular stress. Loss of function of angiogenin has been linked to amyotrophic lateral sclerosis, suggesting that these tRFs may play a crucial role in preventing neurological disorders.

tRNA-DErIVED FRAGMENTS (tRFs)

Recently, another novel class of small, noncoding RNAs of 17–26 nucleotides have been identified by ultra–high-throughput sequencing. Termed tRFs (Lee et al., 2009), these RNAs are specifically derived from either 5′ or 3′ ends of mature tRNAs (tRF-5 and tRF-3 series, respectively) and from the 3′ trailer region of precursor tRNAs (tRF-1 series). tRFs were the second most abundant small RNAs after microRNAs that appeared in short-RNA libraries and have been described by several other groups (Cole et al., 2009; Haussecker et al., 2010). We have analyzed publicly available high-throughput sequencing data sets of more than 50 short-RNA libraries and uncovered tRFs across species, from yeast to humans, including immortalized and primary embryonic cell lines (unpublished data).

The biogenesis of tRFs has remained largely unknown. At least one tRF-1, tRF-1001 SerGA, is generated in the cytoplasm by the tRNA endonuclease ELAC2 (Lee et al., 2009). Another group has shown that Dicer is required for the generation of the tRF-3 series in HeLa cells (Cole et al., 2009). However, a careful analysis of the yield of tRFs in short-RNA libraries from mice or flies with mutations in Dicer or DGCR8 is necessary to definitively determine whether the biogenesis of tRFs is as dependent as microRNAs on Dicer or DGCR8 function. Thus, open questions remain regarding how these tRFs are generated, whether each series is generated by shared or distinct cellular machinery, and how their biogenesis is regulated during development and disease.

Most importantly, while these tRFs appear to be present ubiquitously, their biological functions remain mostly unknown. We have shown that at least one tRF-1, tRF-1001 SerGA, is essential for the proliferation of prostate cancer cells, suggesting that they play a specific role in cellular function. Argonaute protein has been shown to be associated with a subset of tRFs (Haussecker et al., 2010), indicating the presence of some mechanistic overlap with the various microRNA-related pathways. Indeed, provision of the antisense strand to a tRF can drive the tRFs into a RISC complex and silence an mRNA with complementary sequence (Haussecker et al., 2010). Phenotypic screens, in conjunction with studies to understand the precise modes of action of these tRFs should help us uncover the biological function of tRFs. In addition, if future analysis of short-RNA sequence data shows an up-regulation of a class of tRFs (e.g., tRF-1 relative to tRF-5 or tRF-3) or of specific tRFs (e.g., tRF-1001), then these short molecules may also be useful as a biomarker of that cancer.

FUTURE PERSPECTIVES

The discovery of non-micro-short RNAs has greatly broadened the horizon of genetic regulation of diverse biological processes. However, more questions have been raised about their function than have been answered. Despite their presence in multiple organisms, the biogenesis of many of these non-micro-short RNAs is not completely understood, yet these molecules contribute a large portion of the non-micro-short RNA pool in cells. The mechanisms of action of many of these non-micro-short RNAs have remained elusive.
Therefore, it will be very important to elucidate these mechanisms by identifying their interacting partners and regulatory targets. Although many bioinformatics algorithms are available to predict microRNA targets, and microarrays have been designed to quantitatively screen levels of many microRNAs, such tools are not available for the vast majority of non-micro-short RNAs. Another obvious step is to determine the functional roles of these short RNAs using in vitro and in vivo model systems by examining the effects of their loss or overexpression on the phenotypes of cells and organisms. Because many of these molecules cannot be knocked out using conventional approaches, modified antisense oligonucleotides or “tough-decoy” constructs containing antisense sequences that sponge up endogenous small RNA molecules (Haraguchi et al., 2009; Dey et al., 2012) should be used to knock down non-micro-short RNAs, which would enable phenotypic screens in cells and model organisms. It will also be of great interest to determine whether these small RNA pathways can cross-talk within a genetic network by competing with, collaborating with, or regulating the expression of other classes and isoforms of short RNAs. In addition, discovery of mutations or dysregulation of non-micro-short RNAs in human diseases will be of great importance, and may serve to explain some known disease loci or disease-predisposing single-nucleotide polymorphisms (SNPs) not directly affecting coding-gene expression. Finally, as we have observed that short-RNA molecules such as tRFs are up-regulated exponentially in some cancers, such as B-cell lymphomas, these tRFs and other small RNA molecules have the potential to be used as biomarkers for specific cancers and genetic diseases.

REFERENCES
Aravin A et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203–207.
Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316, 744–747.
Banisch TU, Goudarzi M, Raz E (2012). Small RNAs in germ cell development. Curr Top Dev Biol 99, 79–113.
Borsani O, Zhu J, Verslues PE, Sunark R, Zhu J-K (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123, 1279–1291.
Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128, 1089–1103.
Chapman EJ, Carrington JC (2007). Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet 8, 884–896.
Cheng J, Guo JM, Xiao BX, Mao Y, Jiang Z, Zhou H, Li QN (2011). piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. Clin Chim Acta 412, 1621–1625.
Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, Green PJ, Barton GJ, Huttunen M (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA 15, 2147–2160.
Czech B et al. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature 453, 798–802.
Dey BK, Gagan J, Yan Z, Dutta A (2012). mir-26a is required for skeletal muscle differentiation and regeneration in mice. Genes Dev 26, 2180–2191.
Fire A, Xu S, Montgomery MK, Costas SA, Driver SE, Mello CC (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806–811.
Ghilardi M, Zamore PD (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet 10, 94–108.
Haraguchi T, Ozaki Y, Iba H (2009). Vectors expressing efficient RNA decoys achieve the long-term suppression of specific microRNA activity in mammalian cells. Nucleic Acids Res 37, e43.
Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA (2010). Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA 16, 673–695.
Houwing S et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129, 69–82.
Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P (2011). Angiogenin-induced tRNA fragments inhibit translation initiation. Mol Cell 43, 613–623.
Lee RC, Feinbaum RL, Ambros V (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843–854.
Lee YS, Shibata Y, Malhotra A, Dutta A (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 23, 2639–2649.
Li S, Hu GF (2012). Emerging role of angiogenin in stress response and cell survival under adverse conditions. J Cell Physiol 227, 2822–2826.
Matzke M, Kanno T, Daxinger L, Huettel B, Matzke AJ (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol 21, 367–376.
Rajaseethupathy P, Antinov I, Sheridan R, Frey S, Sander C, Tuschl T, Kandel ER (2012). A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell 149, 693–707.
Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD (2006). A novel class of small RNAs: tRFs. RNA 15, 2147–2160.
Wightman B, Ha I, Ruvkun G (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell 75, 855–862.