Small silencing RNAs: an expanding universe

Megha Ghildiyal and Phillip D. Zamore

Abstract | Since the discovery in 1993 of the first small silencing RNA, a dizzying number of small RNA classes have been identified, including microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). These classes differ in their biogenesis, their modes of target regulation and in the biological pathways they regulate. There is a growing realization that, despite their differences, these distinct small RNA pathways are interconnected, and that small RNA pathways compete and collaborate as they regulate genes and protect the genome from external and internal threats.

The defining features of small silencing RNAs are their short length (~20–30 nucleotides), and their association with members of the Argonaute family of proteins, which they guide to their regulatory targets, typically resulting in reduced expression of target genes. Beyond these defining features, different small RNA classes guide diverse and complex schemes of gene regulation. Some small silencing RNAs, such as small interfering RNAs (siRNAs), derive from dsRNA, whereas others, such as Piwi-interacting RNAs (piRNAs), do not. These different classes of regulatory RNAs also differ in the proteins required for their biogenesis, the constitution of the Argonaute-containing complexes that execute their regulatory functions, their modes of gene regulation and the biological functions in which they participate. New small RNA classes and new examples of existing classes continue to be discovered, such as the recent identification of endogenous siRNAs (endo-siRNAs) in flies and mammals. Here, we provide an overview of small silencing RNAs from plants and metazoan animals. For each class, we describe its biogenesis, function and mode of target regulation, providing examples of the regulatory networks in which each participates (TABLE 1). Finally, we highlight several examples of unexpected, and often unexplained, complexity in the interactions between distinct small RNA pathways.

The discovery of RNAi

In 1998, Fire and Mello established dsRNA as the silencing trigger in Caenorhabditis elegans. Their experiments overturned the contemporary view that antisense RNA induced silencing by base pairing to its mRNA counterpart, thereby preventing its translation into protein. In worms and other animals, siRNA-mediated silencing is known as RNAi. Remarkably, RNAi is systemic in both plants and nematodes, spreading from cell to cell. In C. elegans, RNAi is also heritable: silencing can be transferred to the progeny of the worm that was originally injected with the trigger dsRNA. Viral infection, inverted-repeat transgenes or aberrant transcription products all lead to the production of dsRNA. dsRNA is converted to siRNAs that direct RNAi. siRNAs were discovered in plants and were later shown in animal extracts to serve as guides that direct endonucleolytic cleavage of their target RNAs; siRNAs can be classified according to the proteins involved in their biogenesis, their mode of regulation or their size. In this Review, we differentiate the major types of siRNAs according to the molecules that trigger their production — a classification scheme that best captures the biological distinctions among small silencing RNAs.

siRNAs derived from exogenous agents. Early examples of RNAi were triggered by exogenous dsRNA. In these cases, long exogenous dsRNA is cleaved into double-stranded siRNAs by Dicer, a dsRNA-specific RNase III family ribonuclease. siRNA duplexes produced by Dicer comprise two ~21 nucleotide strands, each bearing a 5′ phosphate and 3′ hydroxyl group, paired in a way that leaves two-nucleotide overhangs at the 3′ ends. The strand that directs silencing is called the guide, whereas the other strand, which is ultimately destroyed, is the passenger. Target regulation by siRNAs is mediated by the RNA-induced silencing complex (RISC), which is the generic name for an Argonaute–small RNA complex. In addition to an Argonaute protein and a small RNA guide, the RISC might also contain auxiliary proteins that extend or modify its function; for example, proteins that redirect the target mRNA to a site of general mRNA degradation.
Table 1  Types of small silencing RNAs

<table>
<thead>
<tr>
<th>Name</th>
<th>Organism</th>
<th>Length (nt)</th>
<th>Proteins</th>
<th>Source of trigger</th>
<th>Function</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>miRNA</td>
<td>Plants, algae, animals, viruses, protists</td>
<td>20–25</td>
<td>Drospha (animals only) and Dicer</td>
<td>Pol II transcription (pri-miRNAs)</td>
<td>Regulation of mRNA stability, translation</td>
<td>93–95, 200–202,226</td>
</tr>
<tr>
<td>casiRNA</td>
<td>Plants</td>
<td>24</td>
<td>DCL3</td>
<td>Transposons, repeats</td>
<td>Chromatin modification</td>
<td>38,44,51, 52,61–63</td>
</tr>
<tr>
<td>tasiRNA</td>
<td>Plants</td>
<td>21</td>
<td>DCL4</td>
<td>miRNA-cleaved RNAs from the TAS loci</td>
<td>Post-transcriptional regulation</td>
<td>64–68</td>
</tr>
<tr>
<td>natsiRNA</td>
<td>Plants</td>
<td>22</td>
<td>DCL1</td>
<td>Bidirectional transcripts induced by stress</td>
<td>Regulation of stress-response genes</td>
<td>71,72</td>
</tr>
<tr>
<td>Exo-siRNA</td>
<td>Animals, fungi, protists</td>
<td>~21</td>
<td>Dicer</td>
<td>Transgenic, viral or other exogenous dsRNA</td>
<td>Post-transcriptional regulation, antiviral defense</td>
<td>4,5,8,227</td>
</tr>
<tr>
<td>Endo-siRNA</td>
<td>Plants, algae, animals, fungi, protists</td>
<td>~21</td>
<td>Dicer (except secondary siRNAs in C. elegans, which are products of RdRP transcription, and are therefore not technically siRNAs)</td>
<td>Structured loci, convergent and bidirectional transcription, mRNAs paired to antisense pseudogene transcripts</td>
<td>Post-transcriptional regulation of transcripts and transposons; transcriptional gene silencing</td>
<td>75–79,82, 83,86,87, 200,201, 228</td>
</tr>
<tr>
<td>piRNA-like</td>
<td>Drosophila melanogaster</td>
<td>24–30</td>
<td>Dicer-independent</td>
<td>In ago2 mutants in Drosophila</td>
<td>Unknown</td>
<td>76</td>
</tr>
<tr>
<td>21U RNA</td>
<td>Caenorhabditis elegans</td>
<td>21</td>
<td>Dicer-independent</td>
<td>Individual transcription of each piRNA?</td>
<td>Transposon regulation, unknown functions</td>
<td>114, 173–175</td>
</tr>
<tr>
<td>26G RNA</td>
<td>Caenorhabditis elegans</td>
<td>26</td>
<td>RdRP?</td>
<td>Enriched in sperm</td>
<td>Unknown</td>
<td>114</td>
</tr>
</tbody>
</table>

ago2, Argonaute2; casiRNA, cis-acting siRNA; DCL, Dicer-like; endo-siRNA, endogenous small interfering RNA; exo-siRNA, exogenous small interfering RNA; miRNA, microRNA; natsiRNA, natural antisense transcript-derived siRNA; piRNA, Piwi-interacting RNA; Pol II, RNA polymerase II; pri-miRNA, primary microRNA; RdRP, RNA-dependant RNA polymerase; tasiRNA, trans-acting siRNA.

Mammals and C. elegans each have a single Dicer that makes both microRNAs (miRNAs) and siRNAs, whereas Drosophila species have two Dicers: DCR-1 makes miRNAs, whereas DCR-2 is specialized for siRNA production. The fly RNAi pathway defends against viral infection, and Dicer specialization might reduce competition for Dicer between precursor miRNAs (pre-miRNAs) and viral dsRNAs. Alternatively, DCR-2 and Argonaute2 (AGO2) specialization might reflect the evolutionary pressure on the siRNA pathway to counter rapidly evolving viral strategies to escape RNAi. In fact, dcr-2 and ago2 are among the most rapidly evolving Drosophila genes. C. elegans might achieve similar specialization with a single Dicer by using the dsRNA-binding protein RNAi defective (RDE-4) as the gatekeeper for entry into the RNAi pathway. However, no natural virus infection has been documented in C. elegans. By contrast, mammals might not use the RNAi pathway to respond to viral infection, having evolved an elaborate, protein-based immune system.

The relative thermodynamic stabilities of the 5’ ends of the two siRNA strands in the duplex determines the identity of the guide and passenger strands. In flies, this thermodynamic difference is sensed by the dsRNA-binding protein R2D2, the partner of DCR-2 and a component of the RISC loading complex (RLC). The RLC recruits AGO2, to which it transfers the siRNA duplex. AGO2 can then cleave the passenger strand as if it were a target RNA. AGO2 always cleaves its RNA target at the phosphodiester bond between the nucleotides that are paired to nucleotides 10 and 11 of the guide strand. Release of the passenger strand after its cleavage converts pre-RISC to mature RISC, which contains only single-stranded guide RNA. In flies, the guide strand is 2’O-methylated at its 3’ end by the S-adenosyl methionine-dependent methyltransferase HEN1 (also known as piRNA methyltransferase, PIMET), completing RISC assembly. In plants, both miRNAs and siRNAs are terminally methylated, a modification that is crucial for their stability.

Plants exhibit a surprising diversity of small RNA types and the proteins that generate them. The diversification of RNA silencing pathways in plants might reflect the need of a sessile organism to cope with biotic and abiotic stress. The number of RNA silencing proteins can vary enormously among animals too, with C. elegans producing 27 distinct Argonaute proteins compared with 5 in flies. Phylogenetic data suggest that nearly all of these ‘extra’ C. elegans Argonautes act in the secondary siRNA pathway, perhaps because endogenous secondary siRNAs are so plentiful in worms. Arabidopsis thaliana has four Dicer-like (DCL) proteins and 10 Argonautes, with both unique and redundant functions. In plants, inverted-repeat transgenes or co-expressed sense and antisense transcripts produce two
The three small RNA silencing pathways in flies are the small interfering RNA (siRNA), microRNA (miRNA) and Piwi-interacting RNA (piRNA) pathways. These pathways differ in their substrates, biogenesis, effector proteins and modes of target regulation. a | dsRNA precursors are processed by Dicer-2 (DCR-2) to generate siRNA duplexes containing guide and passenger strands. DCR-2 and the dsRNA-binding protein R2D2 (which together form the RISC-loading complex, RLC) load the duplex into Argonaute2 (AGO2). A subset of endogenous siRNAs (endo-siRNAs) exhibits dependence on dsRNA-binding protein R2D2 (which together form the RISC-loading complex, RLC) load the duplex into Argonaute2 (AGO2). b | miRNAs are encoded in the genome and are transcribed to yield a primary miRNA (pri-miRNA). Alternatively, miRNAs can be present in introns (termed mirtrons) that are liberated following splicing to yield authentic pre-miRNAs. Pre-miRNAs are exported from the nucleus to the cytoplasm, where they are further processed by Dicer-2 (DCR-2) to generate siRNA duplexes containing guide and passenger strands. DCR-2 and the dsRNA-binding protein R2D2 (which together form the RISC-loading complex, RLC) load the duplex into Argonaute2 (AGO2). c | piRNAs are thought to derive from ssRNA precursors and are made without a dicing step. piRNAs are mostly antisense, but a small fraction is in the sense orientation. Antisense piRNAs are preferentially loaded into Piwi or Aubergine (AUB), whereas sense piRNAs associate with AGO3. The methyltransferase HEN1 adds the 2′-O-methyl modification at the 3′ end. Piwi and AUB collaborate with AGO3 to mediate an interdependent amplification cycle that generates additional piRNAs, preserving the bias towards antisense. The antisense piRNAs probably direct cleavage of transposon mRNA or chromatin modification at transposon loci. SAH, S-adenosyl homocysteine; SAM, S-adenosyl methionine.
resulting dsRNA is cleaved by Dicer into siRNAs that are terminally 2′-O-methylated byHEN1 (REF. 36). It is poorly understood why plant RNAs that are expressed from transgenes are converted by RDR6 into dsRNA but abundant endogenous mRNAs are not. Recent evidence that some housekeeping exonucleases compete with plant RNA silencing pathways for aberrant RNAs suggests that substandard RNA transcripts — for
example, those lacking a 5′ cap or 3′ poly(A) tail — act as substrates for RdRPs. Highly expressed transgenes might overwhelm normal RNA quality-control pathways, escape destruction and be converted to dsRNA by RdRPs.

endo-siRNAs

The first endo-siRNAs were detected in plants and *C. elegans*, and the recent discovery of endo-siRNAs in flies and mammals suggests that endo-siRNAs are ubiquitous among higher eukaryotes.

**Plant endo-siRNAs.** In plants, cis-acting siRNAs (casiRNAs) originate from transposons, repetitive elements and tandem repeats such as 5S ribosomal RNA genes, and comprise the bulk of endo-siRNAs (FIG. 2). casiRNAs are predominantly 24 nucleotides and are methylated by HEN1. Their accumulation requires DCL3 and the RNA polymerases RDR2 and POL IV, and either AGO6 (primarily) or AGO4, which acts redundantly. casiRNAs promote heterochromatin formation by directing DNA methylation and histone modification at the loci from which they originate.

Another class of plant endo-siRNAs illustrates how distinct small RNA pathways interact. Trans-acting siRNAs (tasiRNAs) are endo-siRNAs generated by the convergence of the miRNA and siRNA pathways in plants. miRNA-directed cleavage of certain transcripts recruits the RdRP enzyme RDR6. RDR6 then copies the cleaved transcript into dsRNA, which DCL4 dices into tasiRNAs that are phased. This phasing suggests that DCL4 begins dicing precisely at the miRNA cleavage site, making a tasiRNA every 21 nucleotides.

The site of miRNA cleavage is crucial, because in determining the entry point for Dicer it establishes the target specificity of the tasiRNA produced. One of the determinants that seems to predispose a transcript to produce tasiRNAs after its cleavage by a miRNA is the presence of a second miRNA- or siRNA-complementary site on the transcript. Of particular note is the *TAS3* locus, the RNA transcript of which has two binding sites for miR-390. Only one of these sites is efficiently cleaved

---

**Figure 2 | Plant endogenous small interfering RNA (endo-siRNA) biogenesis.** Cis-acting siRNAs (casiRNAs), trans-acting siRNAs (tasiRNAs) and natural antisense transcript-derived siRNAs (natsiRNAs) are derived from distinct loci. Several of the proteins involved in their biogenesis are genetically redundant, whereas others have specialized roles. a | casiRNAs are the most abundant endogenously produced siRNAs in plants. The RNA polymerases POL IV and RDR2 are proposed to generate dsRNA precursors, which are then diced by DICER-LIKE 3 (DCL3) to generate 24-nucleotide casiRNAs. The methyltransferase HEN1 adds the 2′-O-methyl modification at the 3′ end. These small RNAs load into ARGONAUTE 4 (AGO4) and perhaps AGO6, and promote heterochromatin assembly by targeting DNA methylation and histone modification at the corresponding loci. b | tasiRNA biogenesis requires microRNA (miRNA)-mediated cleavage of transcripts from the *TAS* loci (pre-tasiRNA), which triggers the production of dsRNA by RDR6. The dsRNA is diced into 21-nucleotide tasiRNAs by DCL4 and acts through either AGO1 or AGO7. c | natsiRNAs are derived from overlapping regions of convergent transcripts and require DCL1 or DCL2, POL IV, RDR6 and SUPPRESSOR OF GENE SILENCING 3 (SGS3) for their biogenesis.
by miR-390, but binding of the miRNA to both sites seems to be required to initiate conversion of the TAS3 transcript to dsRNA by RDR6 (REFS 69, 70).

Figure 3 | Genomic sources of dsRNA triggers for endogenous small interfering RNAs (endo-siRNAs) in flies and mammals. siRNAs are derived from dsRNA precursors. endo-siRNAs can arise from intramolecularly produced long dsRNA, complementary overlapping transcripts and bidirectionally transcribed loci. endo-siRNAs might also originate from protein-coding genes that can pair with their cognate pseudogenes and from regions of pseudogenes that can form inverted-repeat structures. Solid grey arrows indicate orientation, whereas black arrows indicate transcription.

The first mammalian endo-siRNAs to be reported corresponded to the long interspersed nuclear element (L1) retrotransposon and were detected in cultured human cells73. Full-length L1 contains both sense and antisense promoters in its 5' UTR that could, in principle, drive bidirectional transcription of L1, producing overlapping complementary transcripts to be processed into siRNAs by Dicer. However, the precise mechanism by which transposons trigger siRNA production in mammals remains unknown.

More recently, endo-siRNAs have been detected in somatic and germ cells of Drosophila species and in mouse oocytes. High-throughput sequencing of small RNAs from germline and somatic tissues of Drosophila and of AGO2 immunoprecipitates revealed a population of small RNAs that could readily be distinguished from miRNAs and piRNAs74–81. These small RNAs are nearly always exactly 21 nucleotides, are present in both sense and antisense orientations, have modified 3' ends and, unlike miRNAs and piRNAs, are not biased towards beginning with uracil. Production of the 21-mers requires DCR-2, although in the absence of DCR-2 a remnant of the endo-siRNA population inexplicably persists.

Fly endo-siRNAs derive from transposons, heterochromatic sequences, intergenic regions, long RNA transcripts with extensive structure and, most interestingly, from mRNAs (FIG. 5). Expression of transposon mRNAs increases in both dcr-2 and ago-2 mutants,
implicating an endogenous RNAi pathway in the silencing of transposons in flies, as reported previously for *C. elegans*[^62][^63]. siRNAs derived from mRNAs are over ten times more likely to come from regions that are predicted to produce overlapping convergent transcripts than expected by chance[^8], suggesting that endo-siRNAs originate from the endogenous dsRNA that is formed when these complementary transcripts pair.

A subset of fly endo-siRNAs derives from ‘structured loci’, RNA transcripts of which can fold into long intramolecularly paired hairpins[^7][^8][^9]. Accumulation of these siRNAs requires DCR-2 and the dsRNA-binding protein Lociquacious (LOQS) — which is typically considered the partner of DCR-1, the Dicer that produces miRNA — rather than R2D2 [REF 84], the usual partner of DCR-2 (FIG. 1). Although surprising, a role for LOQS in the biogenesis of endo-siRNAs from structured loci was anticipated by the earlier finding that LOQS has a role in the production of siRNAs from transgenes that are designed to produce long intramolecularly paired inverted-repeat transcripts that trigger RNAi in flies[^9].

endo-siRNAs have also been identified in mouse oocytes[^46][^47]. As in flies, mouse endo-siRNAs are 21 nucleotides, Dicer-dependent and derived from a variety of genomic sources (FIG. 3). The mouse endo-siRNAs were bound to AGO2, the only mammalian Argonaute protein thought to mediate target cleavage, although it is not known if they also associate with any of the other three mouse Argonaute proteins. Mammalian AGO2 is not, however, the orthologue of fly AGO2, the sequence of which is considerably diverged from other Argonaute proteins.

A subset of mouse oocyte endo-siRNAs maps to regions of protein-coding genes that are capable of pairing to their cognate pseudogenes, and to regions of pseudogenes that are capable of forming inverted-repeat structures (FIG. 3). Pseudogenes can no longer encode proteins, but they drift from their ancestral sequence more slowly than would be expected if they were simply junk DNA. Perhaps some pseudogene sequences are under evolutionary selection to retain the ability to produce antisense transcripts that can pair with their cognate genes to produce endo-siRNAs[^9].

A key challenge for the future will be to understand the biological function of endo-siRNAs, especially those that can pair with protein-coding mRNAs. Do they regulate mRNA expression? And can endo-siRNAs act like miRNAs, tuning the expression of large numbers of genes?

**miRNAs**

The first miRNA to be discovered, *lin-4*, was identified in a screen for genes that are required for postembryonic development in *C. elegans*[^9]. The *lin-4* locus produces a 22-nucleotide RNA that is partially complementary to sequences in the 3’ UTR of its regulatory target, the *lin-14* mRNA[^90][^92]. miRNA binding to partially complementary sites in mRNA 3’ UTRs is now considered to be a hallmark of animal miRNA regulation. In 2001, tens of miRNAs were identified in humans, flies and worms by small RNA cloning and sequencing, thereby establishing miRNAs as a new class of small silencing RNAs[^91][^99]. miRBase (release 12.0), the registry that coordinates miRNA naming, now lists 1,638 distinct miRNA genes in plants and 6,930 in animals and their viruses[^9].

**miRNA biogenesis.** miRNAs derive from precursor transcripts called primary miRNAs (pri-miRNAs), which are typically transcribed by RNA polymerase II (RNA Pol II)[^97][^100]. Several miRNA genes are present as clusters in the genome and probably derive from a common pri-miRNA transcript. Liberating a 20–24-nucleotide miRNA from its pri-miRNA requires the sequential action of two RNase III endonucleases, assisted by their dsRNA-binding domain (dsRBD) partner proteins [FIG. 1]. First, the pri-miRNA is processed in the nucleus into a 60–70-nucleotide pre-miRNA by Drosha, acting with its dsRBD partner — DGCR8 in mammals and Pasha in flies[^97][^101][^102]. The resulting pre-miRNA has a hairpin structure: a loop flanked by base-paired arms that form a stem. Pre-miRNAs have a two-nucleotide overhang at their 3’ ends and a 5’ phosphate group, which are indicative of their production by an RNase III. The nuclear export protein Exportin 5 carries the pre-miRNA to the cytoplasm bound to Ran, a GTPase that moves RNA and proteins through the nuclear pore[^106][^109].

In the cytoplasm, Dicer and its dsRBD partner protein, TRBP in mammals and LOQS in flies, cleaves the pre-miRNA[^7][^11][^85][^103][^109]. Dicer — like Argonaute proteins but unlike Drosha — contains a PAZ domain, presumably allowing it to bind the two-nucleotide 3’-overhanging end left by Drosha. Dicer cleavage generates a duplex containing two strands, termed miRNA and miRNA*, corresponding to the two sides of the base of the stem. These correspond to the guide and passenger strands of an siRNA, and similar thermodynamic criteria influence the choice of miRNA versus miRNA[^22][^23]. miRNAs can arise from either arm of the pre-miRNA stem, and some pre-miRNAs produce mature miRNAs from both arms, whereas others show such pronounced asymmetry that the miRNA* is rarely detected even in high-throughput sequencing experiments (BOX 2).

In flies, worms and mammals, a few pre-miRNAs are produced by the nuclear pre-mRNA splicing pathway instead of through processing by Drosha[^15][^16][^17]. These pre-miRNA-like introns, termed mirtrons, are spliced out of mRNA precursors. The spliced introns first accumulate as lariat products that require 2’–5’ debranching by a lariat-debranching enzyme. Debranching yields an authentic pre-miRNA, which can then enter the standard miRNA biogenesis pathway.

In plants, DCL1 fills the roles of both Drosha and Dicer, converting pri-miRNAs to miRNA–miRNA* duplexes[^44][^120][^122]. DCL1, assisted by its dsRBD partner HYLI, converts pri-miRNAs to miRNA–miRNA* duplexes in the nucleus, after which the miRNA–miRNA* duplex is thought to be exported to the
cytoplasm by HASTY, an Exportin 5 homologue (hasty mutants develop precociously, hence their name)\textsuperscript{5,122–124}. Unlike animal miRNAs, plant miRNAs are 2’-O-methylated at their 3’ ends by HEN1 (REFS \textsuperscript{34, 120, 125}). HEN1 protects plant miRNAs from 3’ uridylation, which is thought to be a signal for degradation\textsuperscript{34}. HEN1 probably acts before miRNAs are loaded into AGO1, because both miRNA\textsuperscript{*} and miRNA strands are modified in plants\textsuperscript{34}.

### Target regulation by miRNAs.

The mechanism by which a miRNA regulates its mRNA target reflects both the specific Argonaute protein into which the small RNA is loaded and the extent of complementarity between the miRNA and the mRNA\textsuperscript{126–128}. A few miRNAs in flies and mammals are nearly fully complementary to their mRNA targets; these direct endonuclease cleavage of the mRNA\textsuperscript{129–133}. Such extensive complementarity is considered the norm in plants, as target cleavage was thought to be the main mode of target regulation in plants\textsuperscript{30, 62, 134}. However, in flies and mammals, most miRNAs pair with their targets through only a limited region of sequence at the 5’ end of the miRNA called the ‘seed region’; these miRNAs repress translation and direct degradation of their mRNA targets\textsuperscript{135–140}. The seed region of all small silencing RNAs contributes most of the energy for target binding\textsuperscript{141, 142}. Thus, the seed is the primary specificity determinant for target selection. The small size of the seed means that a single miRNA can regulate many, even hundreds, of different genes\textsuperscript{143, 144}.

Intriguingly, recent data suggest that the nuclear transcriptional history of an mRNA influences whether a miRNA represses its translation at the initiation or the elongation step\textsuperscript{145}.

As plant miRNAs are highly complementary to their mRNA targets, they can direct mRNA target cleavage. Nonetheless, AGO1-loaded plant miRNAs can also block translation, suggesting a common mechanism between plant and animal miRNAs, despite the absence of specific miRNAs shared between the two kingdoms\textsuperscript{146}.

### Functions of miRNAs.

Like transcription factors, miRNAs regulate diverse cellular pathways and are widely believed to regulate most biological processes in plants and animals, ranging from housekeeping functions to responses to environmental stress. Covering this vast body of work is beyond the scope of this article; the cited reviews provide valuable insight\textsuperscript{147–149}.

The study of miRNA pathway mutants provided early evidence for the influence of miRNAs on biological processes in both plants and animals. Loss of Dicer or miRNA-associated Argonaute proteins is nearly always lethal in animals, and such mutants show severe developmental defects in both plants and animals. In Drosophila species, dcr-1 mutant germ-line stem-cell clones divide slowly; in Arabidopsis species, embryoabnormal is abnormal in dcl1 mutants; in C. elegans, dcr-1 mutants display defects in germ-line development and embryonic morphogenesis; zebrafish lacking both maternal and zygotic Dicer are similarly defective in embryogenesis; and mice lacking Dicer die as early embryos, apparently devoid of stem cells\textsuperscript{143, 120, 150–153}. Loss of Dicer in mouse embryonic fibroblasts causes increased DNA damage and, consequently, the upregulation of signalling by the tumour suppressor proteins p19ARF and p53, which induces premature senescence\textsuperscript{154}.

Many miRNAs function in specific biological processes, in specific tissues and at specific times\textsuperscript{155}. The importance of small silencing RNAs goes far beyond the RNA silencing field: long-standing questions about the molecular basis of pluripotency, tumorigenesis, apoptosis, cell identity and so on are finding answers in small RNA\textsuperscript{147, 156}.

#### piRNAs: the longest small RNAs

piRNAs function in the germ line. piRNAs are the most recently discovered class of small RNAs and, as their name suggests, they bind to the Piwi clade of Argonaute proteins. (Animal Argonaute proteins can be subdivided by sequence relatedness into Argonaute and Piwi subfamilies.) The Piwi clade comprises Piwi, Aubergine (AUB) and AGO3 in flies, MILI, MIWI and MIW12 in mice (also called PIWIL1, PIWIL2 and PIWIL4, respectively), and HILL1, HIW11, HIW12 and HIW13 in humans (also called PIWIL2, PWIL, PIWIL4 and PIWIL3, respectively).
piRNAs were first proposed to ensure germline stability by repressing transposons when Aravin and colleagues discovered in flies a class of longer small RNAs (~25–30 nucleotides) associated with silencing of repetitive elements\(^b\). Later, these ‘repeat associated small interfering RNAs’ — subsequently renamed piRNAs — were found to be distinct from siRNAs: they bind Piwi proteins and do not require DCR-1 or DCR-2 for their production, unlike miRNAs and siRNAs\(^a\). Moreover, they are 2’-O-methylated at their 3’ termini, unlike miRNAs but similar to siRNAs in flies\(^b\).

High-throughput sequencing of vertebrate piRNAs revealed a class of piRNAs unrelated to repetitive sequences\(^b\). Mammalian piRNAs can be divided into pre-pachytene and pachytene piRNAs, according to the stage of meiosis at which they are expressed in developing spermatocytes. Like piRNAs in flies, pre-pachytene piRNAs predominantly correspond to repetitive sequences and are implicated in silencing transposons, such as L1 and intracisternal A-particle (IAP)\(^b\). In male mice, gametic methylation patterns are established when germ cells arrest their cell cycle 14.5 days post-coitum, resuming cell division 2–3 days after birth\(^b\). Both MILI and MIW12 are expressed during this period, and MIW12-deficient mice lose DNA methylation marks on transposons\(^b\). The pre-pachytene piRNAs, which bind MIW12 and MILI, might serve as guides to direct DNA methylation of transposons. In contrast to pre-pachytene piRNAs, the pachytene piRNAs mainly arise from unannotated regions of the genome, not transposons, and their function remains unknown\(^b\).

Three recent studies report that the previously discovered germline ‘21U’ RNAs in *C. elegans* are piRNAs\(^b\). These small RNAs were initially identified by high-throughput sequencing\(^b\). They are precisely 21 nucleotides, begin with a uridine 5’-monophosphate and are 3’ modified. They bind Piwi-related gene (PRG-1), a *C. elegans* Piwi protein. Each 21U-RNA might be transcribed separately, as they are all flanked by a common upstream motif. Like piRNAs in *Drosophila* species, the 21U-RNAs are required for maintenance of the germ line and fertility, and like *Drosophila* AUB and other piRNA pathway components, PRG-1 is found in specialized ‘P granules’, which are associated with germline function, in a perinuclear ring called nuage. Worm piRNAs resemble pachytene piRNAs in mammals: their targets and functions are largely unknown.

**Biogenesis of piRNAs.** PiRNA sequences are stunningly diverse, with more than 1.5 million distinct piRNAs identified thus far in flies, but collectively they map to a few hundred genomic clusters\(^b\). The best studied cluster is the *flamenco* locus. *flamenco* was identified genetically as a repressor of the gypsy, *ZAM* and *Idex* transposons\(^a\). Unlike siRNAs, *flamenco* piRNAs are mainly antisense, suggesting that piRNAs arise from long, single-stranded precursors. In fact, disruption of *flamenco* by insertion of a P-element near the 5’ end of the locus blocks the production of distal piRNAs up to 168 kb away. Thus, an enormously long, ssRNA transcript seems to be the source of the piRNAs that derive from the *flamenco* locus\(^b\).

The current model for piRNA biogenesis was inferred from the sequences of piRNAs that are bound to Piwi, AUB and AGO3 (REFS 177, 184). piRNAs bound to Piwi and AUB are typically antisense to transposon mRNAs, whereas AGO3 is loaded with piRNAs corresponding to the transposon mRNAs themselves (FIG. 1). Moreover, the first 10 nucleotides of antisense piRNAs are frequently complementary to the sense piRNAs found in AGO3. This unexpected sequence complementarity has been proposed to reflect a feed-forward amplification mechanism — ‘piRNA ping-pong’ — that is activated only after transcription of transposon mRNA\(^b\). A similar amplification loop has been inferred from high-throughput piRNA sequencing in vertebrates, implying that it has been conserved through evolution\(^b\). Many aspects of the ping-pong model remain speculative. Why AGO3 seems to bind only sense piRNAs derived from transposon mRNAs is unknown. An untested idea is that different forms of RNA Pol II transcribe primary piRNA transcripts and transposon mRNAs, and that the specialized RNA Pol II that transcribes the primary piRNA precursor recruits Piwi and AUB, but not AGO3. How the 3’ ends of piRNAs are made is also not known.

**PiRNA function and regulation.** Piwi family proteins are indispensable for germline development in many, perhaps all, animals; but they have thus far been most extensively studied in *Drosophila* species. Piwi is restricted to the nucleoplasm of *Drosophila* germ cells and adjacent somatic cells. Piwi is required to maintain germline stem cells and to promote their division; the protein is required in both the somatic niche cells that support germline stem cells and in the stem cells themselves\(^b\). In the male germ line, AUB is required for the silencing of the repetitive *Stellate* locus, which would otherwise cause male sterility. Expression of *Stellate* is controlled by the related repetitive *Suppressor of Stellate* locus, the source of antisense piRNAs that act through AUB to repress *Stellate*\(^b\).

*aubergine* was originally identified because it is required for specification of the embryonic axes\(^b\). The loss of anterior–posterior and dorsal–ventral patterning in embryos from mothers lacking AUB is an indirect consequence of the dsDNA breaks that occur in the oocyte in its absence\(^b\). The breaks seem to activate a DNA-damage checkpoint that disrupts patterning of the oocyte and, consequently, of the embryo. The defects in patterning, but not in silencing repetitive elements, are rescued by mutations that bypass the DNA damage signalling pathway, suggesting that the breaks are caused by transposition. That activation of a DNA damage checkpoint should inappropriately reorganize embryonic polarity was unexpected, but further underscores the vital part piRNAs play in germline development.
piRNAs outside the germ line? The role of piRNAs in the fly soma is hotly debated. Piwi and AUB are required to silence tandem arrays of white, a gene required to produce red eye pigment. It is not understood if piRNAs are produced in the soma as well as in the germ line, or if piRNAs that are present during germline development deposit long-lived chromatin marks that exert their effects days later.

Both piRNAs and endo-siRNAs repress transposons in the germ line, where mutations caused by transposition would propagate to the next generation. siRNAs, which are produced by the RNAi pathway, probably provide a rapid response to the introduction of a new transposon into the germ line, a challenge not dissimilar to a viral infection. By contrast, the piRNA system seems to provide a more robust, permanent solution to the acquisition of a transposon. In the soma, however, endo-siRNAs are the predominant transposon-derived small RNA class, and their loss in dcr-2 and ago2 mutants increases transposon expression. Somatic piRNA-like small RNAs have been observed in ago2 mutant flies. Perhaps, in the absence of endo-siRNAs, piRNAs are produced somatically and resume transposon surveillance. Such a model implies significant cross-talk between the piRNA and endo-siRNA–generating machineries.

Interwoven pathways
The RNAi, miRNA and piRNA pathways were initially believed to be independent and distinct. However, the lines distinguishing them continue to fade. These pathways interact and rely on each other at several levels, competing for and sharing substrates, effector proteins and cross-regulating each other.
Competition for substrates during loading. Both the siRNA and miRNA pathways load dsRNA duplexes containing an ~19 bp double-stranded core flanked by 3′ overhangs of two nucleotides. An siRNA duplex contains guide and passenger strands and is complementary throughout its core; a miRNA–miRNA* duplex contains mismatches, bulges and GU wobble pairs. In Drosophila species, biogenesis of small RNA duplexes is uncoupled from its loading into AGO1 or AGO2 (REFS 191, 192). Instead, loading is governed by the structure of the duplex: duplexes bearing bulges and mismatches are sorted into the miRNA pathway and hence loaded into AGO1; duplexes with greater double-stranded character partition into AGO2, the Argonaute protein that is associated with RNAi.

The partitioning of small RNAs between AGO1 and AGO2 also has implications for target regulation. AGO1 primarily represses translation, whereas AGO2 represses by target cleavage, reflecting the faster rate of target cleavage by AGO2 compared with AGO1 [REF. 192]. Sorting creates competition between the two pathways for substrates. In Drosophila, loading of a small RNA duplex into one pathway decreases its association with the other pathway.

Different dsRNA precursors require distinct combinations of proteins to produce small silencing RNAs. For example, Drosophila endo-siRNAs derived from structured loci require LOQS rather than R2D2 (REFS 77–79). We presume that under some circumstances the endo-siRNA and miRNA pathways might therefore compete for LOQS. The endo-siRNA and RNAi pathways probably also compete for shared components.

In contrast to Drosophila, plants load small RNAs into Argonautes according to the identity of the 5′ nucleotide of the small RNA. AGO1 is the major effector Argonaute for miRNAs, and the majority of miRNAs begin with uridine; AGO4 is the major effector of the heterochromatic pathway and is predominantly loaded with small RNAs beginning with an adenosine. AGO2 and AGO5, however, have no characterized function in plants. Changing the 5′ nucleotide from adenosine to uracil shifts the loading bias of plant small RNAs from AGO2 to AGO1, and vice versa. Similarly, Arabidopsis AGO4 binds small RNAs that begin with adenosine, whereas AGO5 prefers cytidine.

piRNAs bound to AUB and Piwi typically begin with uracil, whereas those bound to AGO3 show no 5′ nucleotide bias. It remains to be determined if this reflects a 5′-nucleotide preference like the situation for the plant Argonautes or some feature of an as yet undiscovered piRNA-loading machinery that sorts piRNAs between Piwi proteins.

Cross talk. Small RNA pathways are often entangled. tasiRNA biogenesis in Arabidopsis is a classic example of such cross talk between pathways. miRNA-directed cleavage of tasiRNA-generating transcripts initiates tasiRNA production and subsequent regulation of tasiRNA targets. In C. elegans at least one piRNA has been implicated in initiating endo-siRNA production, and in flies the endo-siRNA pathway might repress expression of piRNAs in the soma. Moreover, small RNA levels might be buffered by negative feedback loops in which small RNAs from one pathway alter the expression levels of RNA silencing proteins that act in the same or in other RNA silencing pathways.

Conclusions

Despite our growing understanding of the mechanism and function of small RNAs, their evolutionary origins remain obscure. siRNAs are present in all three eukaryotic kingdoms — plants, animals and fungi — and provide antiviral defence in at least plants and animals. Thus, the siRNA machinery was present in the last common ancestor of plants, animals and fungi. By contrast, miRNAs have only been found in land plants, in the unicellular green alga Chlamydomonas reinhardtii and in metazoan animals, but not in unicellular choanoflagellates or fungi. Deep-sequencing experiments have found no miRNAs shared by plants and animals, suggesting that miRNA genes, unlike the miRNA protein machinery, arose independently at least twice in evolution. Finally, piRNAs seem to be the youngest major small RNA family, having been found only in metazoan animals. Although Dicer proteins have been identified only in eukaryotes, Argonaute proteins can also be found in eubacteria and archaea, raising the prospect that small nucleic acids might have served as guides for proteins at the dawn of cellular life, and even though the machinery might be ancient, the small RNA guides have diversified over time to acquire specialized roles.

The history of small silencing RNAs makes predicting the future particularly daunting, as new discoveries have come at a breakneck pace, with each new small RNA mechanism or function forcing a re-evaluation of cherished models and ‘facts’. Several longstanding but unanswered questions, however, are worth highlighting. First, does RNAi — in the sense of an siRNA-guided defence against external nucleic acid threats, such as viruses — exist in mammals? Second, how do miRNAs repress gene expression? Do several parallel mechanisms coexist in vivo, or will the current, apparently contradictory models for mRNA-directed translational repression and mRNA decay ultimately be unified in a larger mechanistic scheme? Third, can miRNA-regulated genes ever be identified by computation alone, or will computational predictions ultimately give way to high-throughput experimental methods for associating individual miRNA species with their regulatory targets? Will network analysis uncover themes in the relationships between miRNAs and their targets that reveal why miRNA regulation is so widespread in animals? Fourth, how are piRNAs made? The feed-forward amplification ‘ping-pong’ model is appealing, but probably underestimates the complexity of piRNA biogenesis mechanisms. We do not yet know how piRNA 3′ ends are generated. Nor do we have a coherent model for how long, antisense transcripts from piRNA clusters are fragmented into piRNAs. Finally, will the increasing number of examples of small RNAs carrying epigenetic information across generations ultimately force us to re-examine our Mendelian view of inheritance?
This paper and reference 9 deduced the structure of siRNAs and discovered that they became the target that lies across from bases 10 and 11 sequence of the target RNA. This paper and reference 6 showed that siRNAs are active and functional in the Dicer-dependent cleavage of mRNA at 21 to 23 nucleotide motifs. Zamore et al., 2001, deduced the mechanism for inheritance of RNAi in plants. Vilcek, J., Fifty years of interferon research: aiming at a cure for viral infections and cancer. Immunity, 2005, 28(6), 885–894.

This study, together with references 12 and 13, identifies a class of short interfering RNAs (siRNAs) as the guides for post-transcriptional gene silencing in plants. Xu, Y., Australia. Functional siRNAs and miRNAs exhibit strand bias. Science, 2005, 310, 191–194.

This paper identified small RNAs—sirRNAs and miRNAs— that protected developing self-organized tissues from viral infections. It was the first report of small interfering RNAs and their role in antiviral defense. Takeda et al., 2001, identified the role of the RNA-directed RNA polymerase (RDR2) in the generation of long antisense RNAs that control virus infection and related processes. Xu, Y., Australia. A cellular function for the RNA-directed RNA polymerase (RDR2). Science, 2005, 310, 191–194.

In this work, Dicer was shown to be the dsRNA-specific ribonuclease that converts long dsRNA into siRNAs. Elbashir et al., 2001, demonstrated that Dicer is required for RNAi in plants. Kurimura et al., 2004, identified the role of the RNA-directed RNA polymerase (RDR2) in the generation of long antisense RNAs that control virus infection and related processes. Xu, Y., Australia. A cellular function for the RNA-directed RNA polymerase (RDR2). Science, 2005, 310, 191–194.

This paper and reference 9 showed that the siRNAs are present in all tissues and organs, and that they are essential for RNAi. Elbashir et al., 2001, demonstrated that Dicer is required for RNAi in plants. Kurimura et al., 2004, identified the role of the RNA-directed RNA polymerase (RDR2) in the generation of long antisense RNAs that control virus infection and related processes. Xu, Y., Australia. A cellular function for the RNA-directed RNA polymerase (RDR2). Science, 2005, 310, 191–194.

This paper and reference 9 showed that the siRNAs are present in all tissues and organs, and that they are essential for RNAi. Elbashir et al., 2001, demonstrated that Dicer is required for RNAi in plants. Kurimura et al., 2004, identified the role of the RNA-directed RNA polymerase (RDR2) in the generation of long antisense RNAs that control virus infection and related processes. Xu, Y., Australia. A cellular function for the RNA-directed RNA polymerase (RDR2). Science, 2005, 310, 191–194.

This paper reports the first computational evidence that microRNA targets determined by the promoter of the target gene. Nature 460, 741–745 (2009).


This paper reports the first computational evidence for a seed sequence and uses seed pairing to identify miRNA targets. Nature 460, 741–745 (2009).


This paper reports the first computational evidence for a seed sequence and uses seed pairing to identify miRNA targets. Nature 460, 741–745 (2009).


A novel class of small RNAs bind to Piwi proteins, and that they are largely made from single-stranded precursors without the need for 3′-end modification and are bound to Piwi proteins, and that they are largely antisense to selfish genetic elements.


Ohara, T. et al. The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. Nucleic Acids Res. 34, 547–553 (2006).


Acknowledgements
We thank H. Seitz for bioinformatic assistance and X. Chen, I. Pekker and S. Ameres for helpful discussions. This work was supported, in part, by grants from the National Institutes of Health to P.D.Z. (GM065236 and GM062862).

DATABASES
UniProtKB: http://www.uniprot.org
AGO2 | AUB | DCR-1 | DCR-2 | Drosha | Piwi | R2D2

FURTHER INFORMATION
Phillip D. Zamore's homepage: http://www.umassmed.edu/bmp/faculty/zamore.cfm
miRBase: http://microrna.sanger.ac.uk
ALL LINKS ARE ACTIVE IN THE ONLINE PDF