POST-TRANSCRIPTIONAL CONTROL

Biogenesis of small RNAs in animals

V. Narry Kim, Jinju Han* and Mikiko C. Siomi‡*

Abstract | Small RNAs of 20–30 nucleotides can target both chromatin and transcripts, and thereby keep both the genome and the transcriptome under extensive surveillance. Recent progress in high-throughput sequencing has uncovered an astounding landscape of small RNAs in eukaryotic cells. Various small RNAs of distinctive characteristics have been found and can be classified into three classes based on their biogenesis mechanism and the type of Argonaute protein that they are associated with: microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs or esiRNAs) and Piwi-interacting RNAs (piRNAs). This Review summarizes our current knowledge of how these intriguing molecules are generated in animal cells.

Heterochromatin
Highly condensed regions of the genome in which transcription is generally limited.

RNase III-type protein
An endonuclease that cleaves double-stranded RNAs and creates 5′-phosphate and 3′-hydroxyl termini, leaving 2-nucleotide 3′ overhangs.

**School of Biological Sciences and Center for National Creative Research, Seoul National University, Seoul, 151-742, Korea. ‡Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, and the Japan Science and Technology Agency (JST), Core Research for Evolutionary Science and Technology (CREST), Saitama 332-0012, Japan. Correspondence to V.N.K. and M.C.S.*
e-mails: narrykim@snu.ac.kr; siomim@sc.itc.keio.ac.jp
doi:10.1038/nrm2632

The first small RNA, lin-4, was discovered in 1993 by genetic screens in nematode worms[1,2]. The number of known small RNAs has since expanded substantially, mainly as a result of the cloning and sequencing of size-fractionated RNAs[3–5]. The recent development of deep-sequencing technologies[6,7] and computational prediction methods[8–11] has accelerated the discovery of less abundant small RNAs. The functions of small RNAs range from heterochromatin formation to mRNA destabilization and translational control[12,13]. Through such extensive patrolling across the genome and transcriptome, small RNAs are involved in almost every biological process, including developmental timing, cell differentiation, cell proliferation, cell death, metabolic control, transposon silencing and antiviral defence.

'Small RNA' is a rather arbitrary term, because it was previously used for other non-coding RNAs, such as small nuclear RNAs (snRNAs) and transfer RNAs (tRNAs). Bacterial short regulatory RNAs have also been referred to as small RNAs, but they are not related to eukaryotic small RNAs. What distinguishes and defines eukaryotic small RNAs in the RNA silencing pathway is their limited size (~20–30 nucleotides (nt)) and their association with Argonaute (Ago)-family proteins (BOX 1; TABLE 1). The Ago family can be grouped further into two clades: the Ago subfamily and the Piwi subfamily. At least three classes of small RNAs are encoded in our genome, based on their biogenesis mechanism and the type of Ago protein that they are associated with: microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs or esiRNAs) and Piwi-interacting RNAs (piRNAs). It should be noted, however, that the recent discoveries of numerous non-canonical small RNAs have blurred the boundaries between the classes.

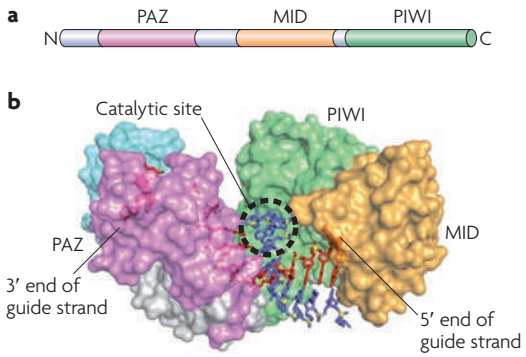
The best understood among the three classes, miRNAs are generated from local hairpin structures by the action of two RNase III-type proteins, Drosha and Dicer (BOX 2). Mature miRNAs of ~22 nt are then bound by Ago-subfamily proteins. miRNAs target mRNAs and thereby function as post-transcriptional regulators. The longest of the three classes, piRNAs (24–31 nt in length) are associated with Piwi-subfamily proteins. Intriguingly, the biogenesis of piRNAs does not depend on Dicer[14]. piRNAs are highly abundant in germ cells and at least some of them are involved in transposon silencing through heterochromatin formation or RNA destabilization. endo-siRNAs have been studied mostly in *Drosophila melanogaster*, although they have also been found in mouse oocytes and embryonic stem (ES) cells[15–17]. Despite their similarity with miRNAs in terms of their association with Ago-subfamily proteins, endo-siRNAs differ from miRNAs in that they are derived from long double-stranded RNAs (dsRNAs) and are dependent only on Dicer and not on Drosha[18–20]. They are also slightly shorter (~21 nt) than miRNAs. At least some of the endo-siRNAs have been shown to function as post-transcriptional regulators that target RNAs.

This Review summarizes our current knowledge of the biogenesis pathways of small RNAs. The focus of the Review will remain on animal small RNAs, mainly in mammals and flies. Several excellent reviews on small RNAs in plants and yeast are available elsewhere[21–24].

microRNA biogenesis

miRNAs are single-stranded RNAs (ssRNAs) of ~22 nt in length that are generated from endogenous hairpin-shaped transcripts[25]. miRNAs function as guide molecules in post-transcriptional gene regulation by

Box 1 | **Argonaute proteins and their associated small RNAs**



The Argonaute (Ago) family can be classified into two subclades: the Ago subfamily and the Piwi subfamily (TABLE 1). The Ago proteins are expressed ubiquitously, interact with microRNAs (miRNAs) or small interfering RNAs (siRNAs), and function as post-transcriptional regulators. The Piwi proteins are abundantly expressed in germ cells and function in transposon silencing, together with Piwi-interacting RNAs (piRNAs).

Ago-family proteins are composed of three characteristic domains: the PAZ, MID and PIWI domains (see the figure, part **a**). The PAZ domain serves as a docking site for the 3′ end of small RNA[181–184], whereas the MID domain anchors the 5′ terminal nucleotide[183–187] (see the figure, part **b**). Recent studies have determined the structure of *Thermus thermophilus* Ago with a guide strand and target strand duplex[183,184]. The PIWI domain has a structure that is similar to RNase H, which cuts the RNA strand of an RNA–DNA hybrid. Indeed, the PIWI domain of some Ago proteins can cleave the target RNA bound to small RNA: this is called slicer activity. Of the four human Ago proteins (AGO1–4; also known as EIF2C1–4), only AGO2 has slicer activity, whereas in *Drosophila melanogaster* all Ago and Piwi proteins possess slicer activity. Apart from the endonucleolytic cleavage that is mediated by the PIWI domain, the Ago proteins can induce translational repression and exonucleolytic mRNA decay through interaction with other protein factors[13].

The fly Piwi proteins Aubergine (AUB) and AGO3 can cleave target mRNAs, resulting in the silencing of retrotransposons, other intergenic repetitive elements and protein-coding genes, such as *Stellate* (also known as *SteXh*)[145,148]. PIWI might function differently from AUB and AGO3, as it associates with chromatin and interacts with HP1a (heterochromatin protein 1a), which is involved in heterochromatin formation. Thus, PIWI might contribute to the epigenetic control of the fly genome[188]. Consistently, AUB and AGO3 are localized in the cytoplasm, whereas PIWI is found in the nucleus[145–148]. The expression patterns are also different; PIWI, but not AUB and AGO3, was detected in the soma of ovaries, indicating that PIWI has a function outside of germ cells. Epigenetic changes were observed in *Mili* (also known as *Piwil2*) and *Miwi2* (also known as *Piwil4*) knockout mice, indicating that mammalian Piwi homologues function in a similar way to fly PIWI in heterochromatin control[156].

Derepression of transposons was shown in flies that were deficient in *Dicer 2* or *Ago2* and in S2 cells that lack Dicer 2 (REFS 19,20,170,171), which suggests that endogenous siRNAs (endo-siRNAs) as well as piRNAs can defend against nucleic acid-based parasites. Some piRNAs and endo-siRNAs are complementary to protein-coding genes and they might function in *trans* to suppress protein-coding gene expression, probably through mRNA cleavage[18,20]. Figure part **b** is reproduced, with permission, from *Nature* REF. 183 © (2008) Macmillan Publishers Ltd. All rights reserved.

base-pairing with the target mRNAs, usually in the 3′ untranslated region (UTR). Binding of a miRNA to the target mRNA typically leads to translational repression and exonucleolytic mRNA decay, although highly complementary targets can be cleaved endonucleolytically. Other types of regulation, such as translational activation[13] and heterochromatin formation[26], have also been described. Over one-third of human genes are predicted to be directly targeted by miRNAs. Consequently, the unique combination of miRNAs in each cell type determines the use of thousands of mRNAs.

Bilaterian
An animal that has a front, a back, an upside and a downside (bilateral symmetry).

Paralogue
A gene or protein with a highly similar sequence to another that is encoded in the same genome.

miRNA genes and their transcription. At present, the miRNA database contains 154 *Caenorhabditis elegans*, 152 *D. melanogaster*, 337 *Danio rerio* (zebrafish), 475 *Gallus gallus* (chicken), 695 human and 187 *Arabidopsis thaliana* miRNAs. miRNAs are even present in simple multicellular organisms, such as poriferans (sponges) and cnidarians (starlet sea anemone)[27]. Many of the bilaterian animal miRNAs are phylogenetically conserved; ~55% of *C. elegans* miRNAs have homologues in humans, which indicates that miRNAs have had important roles throughout animal evolution[28]. Animal miRNAs seem to have evolved separately from those in plants because their sequences, precursor structure and biogenesis mechanisms are distinct from those in plants[29,30].

Most mammalian miRNA genes have multiple isoforms (paralogues) that are probably the result of gene duplications. For instance, the human genome has 12 loci for let-7-family miRNAs. Paralogues often have identical sequences at nucleotide positions 2–7 relative to the 5′ end of the miRNA. Because these six nucleotides (called seed) are crucial in base pairing with the target mRNA, the paralogues are thought to act redundantly. However, because the 3′ sequences of miRNAs also contribute to target binding and because the expression patterns of these sister miRNAs are often different from each other, members of the same seed family might have distinct roles *in vivo*[31].

Approximately 50% of mammalian miRNA loci are found in close proximity to other miRNAs. These clustered miRNAs are transcribed from a single polycistronic transcription unit (TU)[32], although there may be exceptional cases in which individual miRNAs are derived from separate gene promoters. Some miRNAs are generated from non-coding TUs, whereas others are encoded in protein-coding TUs (FIG. 1). Approximately 40% of miRNA loci are located in the intronic region of non-coding transcripts, whereas ~10% are placed in the exonic region of non-coding TUs. miRNAs in protein-coding TUs are usually found in intronic regions, which account for ~40% of all miRNA loci. Some 'mixed' miRNA genes can be assigned to either intronic or exonic miRNA groups depending on the alternative splicing patterns.

The transcription of most miRNA genes is mediated by RNA polymerase II (Pol II)[33,34], although a minor group of miRNAs that are associated with Alu repeats can be transcribed by Pol III[35]. A range of Pol II-associated transcription factors control miRNA gene transcription[36]. Thus, Pol II-dependent transcription allows miRNA genes to be elaborately regulated in specific conditions and cell types.

Nuclear processing by Drosha. The primary transcripts (pri-miRNAs) that are generated by Pol II are usually several kilobases long and contain local stem-loop structures (FIG. 2a). The first step of miRNA maturation is cleavage at the stem of the hairpin structure, which releases a small hairpin that is termed a pre-miRNA[32]. This reaction takes place in the nucleus by the nuclear RNase III-type protein Drosha[37]. Drosha requires a cofactor, the DiGeorge syndrome critical region

Table 1 | **Types of small RNA-associated Argonaute proteins, and the origin and mechanism of action of small RNAs**

Subfamily	Ago-family protein	Class of small RNA*	Length of small RNA	Origin of small RNA‡	Mechanism of action
Mammals					
Ago	AGO1–4	miRNA	21–23 nt	miRNA genes	Translational repression, mRNA degradation, mRNA cleavage and heterochromatin formation?
		endo-siRNA§	21–22 nt	Intergenic repetitive elements, pseudogenes and endo-siRNA clusters	mRNA cleavage?
Piwi	MILI (PIWIL2 in humans)	Pre-pachytene piRNA and pachytene piRNA	24–28 nt	Transposons and piRNA clusters	Heterochromatin formation (DNA methylation)
	MIWI (PIWIL1 in humans)	Pachytene piRNA	29–31 nt	piRNA clusters	?
	MIWI2 (PIWIL4 in humans)	Pre-pachytene piRNA	27–29 nt	Transposons and piRNA clusters	Heterochromatin formation (DNA methylation)
	(PIWIL3 in humans)	?	?	?	?
Drosophila melanogaster					
Ago	AGO1	miRNA	21–23 nt	miRNA genes	Translational repression and mRNA degradation
	AGO2	endo-siRNA	~21 nt	Transposons, mRNAs and repeats	RNA cleavage
		exo-siRNA	~21 nt	Viral genome	Viral RNA cleavage
Piwi	AUB	piRNA	23–27 nt	Transposons, repeats, piRNA clusters and *Su(Ste)* locus	RNA cleavage
	AGO3	piRNA	24–27 nt	Transposons and repeats (unknown in testis)	RNA cleavage
	PIWI	piRNA	24–29 nt	Transposons, repeats and piRNA clusters	Heterochromatin formation?
Schizosaccharomyces pombe					
Ago	Ago1	endo-siRNA	~21 nt	Outer centromeric repeats, mating-type locus and subtelomeric regions	Heterochromatin formation
Arabidopsis thaliana‖					
Ago	AGO1	miRNA	20–24 nt	miRNA genes	mRNA cleavage and translational repression
		endo-siRNA (tasiRNA including TAS3)	21 nt	TAS genes	mRNA cleavage
		exo-siRNA	20–22 nt	Viral genome	Viral RNA cleavage
	AGO4 and AGO6	rasiRNA	24 nt	Transposons and repetitive elements	Heterochromatin formation
	AGO7	miR-390	21 nt	miRNA gene	Cleavage of TAS3 RNA

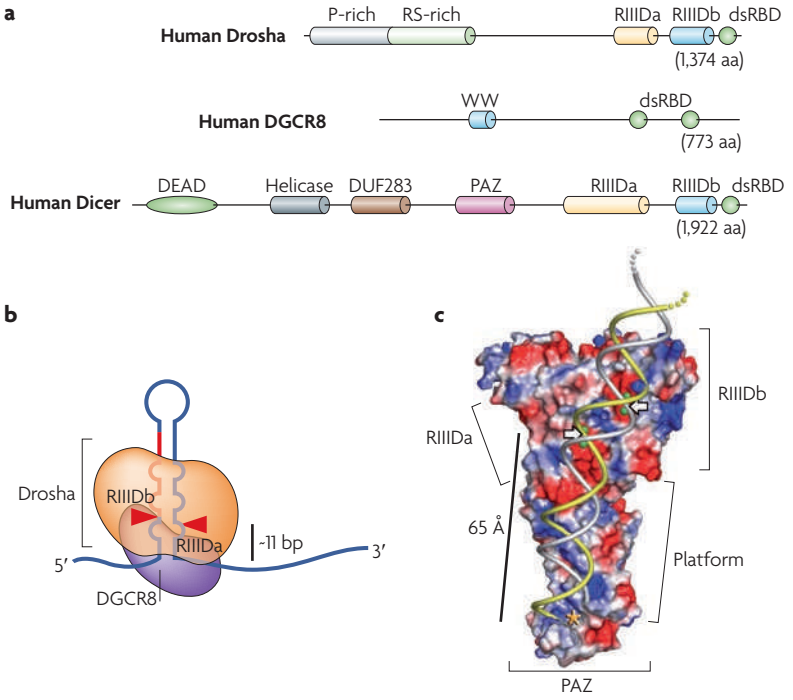
*Small RNAs that are the main partners of a given Ago protein are listed. ‡miRNAs, as a class, are expressed in all cell types, whereas endo-siRNAs and piRNAs are expressed abundantly in germ cells and contribute to germline development. §So far, only AGO2 has been shown to be required for endo-siRNAs. ‖Plants have ten Ago proteins, but only those with known small RNA partners are shown. Ago, Argonaute; AUB, Aubergine; endo-siRNA, endogenous small interfering RNA; exo-siRNA, exogenous small interfering RNA; miRNA, microRNA; nt, nucleotide; piRNA, Piwi-interacting RNA; rasiRNA, repeat-associated siRNA; *Su(Ste)*, *Suppressor of Stellate*; TAS, tasi gene; tasiRNA, *trans*-acting siRNA.

gene 8 (DGCR8) protein in humans (Pasha in *D. melanogaster* and *C. elegans*)[38–41]. Together with DGCR8 (or Pasha), Drosha forms a large complex known as the Microprocessor complex, which is ~500 kDa in *D. melanogaster*[39] and ~650 kDa in humans[38,40]. Mouse ES cells that are deficient in the *Dgcr8* gene fail to produce miRNAs and manifest defects in proliferation and differentiation, thereby confirming the essential role of DGCR8 in the miRNA pathway and the importance of miRNAs in ES-cell function[42]. Drosha and DGCR8 are conserved only in animals[43–45].

How does the Microprocessor complex recognize its substrates? A typical metazoan pri-miRNA consists of a stem of ~33 base pairs (bp), a terminal loop and flanking

Polycistronic transcription unit
An RNA transcript that includes regions that represent multiple gene products.

Box 2 | **RNase III proteins and their mechanism of action**



Two types of RNase III are found in animals: Drosha and Dicer. Both proteins have two tandem RNase III domains (RIIIDs) and a double-stranded RNA-binding domain (dsRBD; see the figure, part **a**). Two RIIIDs interact with each other to make an intramolecular dimer in which the two catalytic sites are located closely to each other. The first RIIID cuts the 3′ strand of dsRNA, whereas the second RIIID cleaves the 5′ strand, generating a 2-nucleotide (nt) 3′ overhang[38,189,190] (see the figure, parts **b**,**c**).

Drosha is a nuclear protein of 130–160 kDa. The dsRBD is necessary for activity, although it is not sufficient to bind a primary transcript (pri-miRNA). To provide the RNA-binding activity, Drosha interacts with a dsRNA-binding protein, DiGeorge syndrome critical region gene 8 (DGCR8; Pasha in *Drosophila melanogaster* and *Caenorhabditis elegans*), through its middle region (this Drosha–DGCR8 complex is known as the Microprocessor; see the figure, part **b**). The amino-terminal region contains the nuclear localization signal, which is dispensable for cleavage activity *in vitro*[38,191].

DGCR8 and Pasha are ~120 kDa and localize to the nucleoplasm and the nucleolus[177,191]. They contain two dsRBDs that recognize the ssRNA–dsRNA junction of a substrate[46]. DGCR8 and Pasha interact with Drosha through their C-terminal domain[191]. A structure of the partial human DGCR8 protein that contains the two dsRBDs and a part of the carboxy-terminal domain showed that the two dsRBDs are arranged with pseudo two-fold symmetry and that the C-terminal helix is closely packed against the two dsRBDs[192]. The middle domain of DGCR8 is required for haem binding, oligomerization and pri-miRNA processing[191,193].

Dicer homologues are cytoplasmic RNase III proteins of ~200 kDa. The middle region of Dicer contains a PAZ domain, which binds to the 3′ protruding end of RNAs[181,182,194,195]. Structural studies on the *Giardia intestinalis* Dicer protein show that the processing centre made of the two RIIIDs is connected to the PAZ domain by a long, positively charged helix[196]. Based on the structure (see the figure, part **c**), it is predicted that the 3′ end of a dsRNA is docked to the PAZ domain and the RNA stem interacts with the flat, positively charged extension to reach the catalytic centre. The distance between the PAZ domain and the catalytic site (65 Å) approximately matches the length of the product of *G. intestinalis* Dicer (25 nt). The DEAD-box RNA helicase domain is not necessary for Dicer activity *in vitro*[78], and its role remains unknown. Figure part **c** is reproduced, with permission, from REF. 196 © (2006) American Association for the Advancement of Science.

ssRNA segments. DGCR8 interacts with pri-miRNAs through the ssRNA segments and the stem of ~33 bp, and assists Drosha to cleave the substrate ~11 bp away from the ssRNA–dsRNA junction[46,47].

Recent studies show that pri-miRNA processing might be a co-transcriptional process[48–50]. The initial model was based on the finding that Drosha processing of intronic miRNA precedes the splicing of a host intron[48] (FIG. 2b). Interestingly, the cleavage of the intron by Drosha does not impair splicing[48]. This is consistent with a previous 'exon-tethering' model[51], which suggested that the exons of Pol II transcripts are co-transcriptionally assembled into the spliceosome. Thus, Drosha processing might take place after the transcript is tied to the splicing commitment complex (also known as the early spliceosome complex), but before the intron is excised. Thus, cropping and splicing might be highly coordinated co-transcriptional processes. Supporting this, pri-miRNA and Drosha localize to the transcription sites[50] and pri-miRNAs are enriched in the chromatin-associated nuclear fraction[49,50]. Chromatin immunoprecipitation and nuclear run-on assays have provided further evidence that pri-miRNA processing is a co-transcriptional process[49,50]. The nuclear exosome (which comprises 3′→5′ exonucleases) and XRN2 (a 5′→3′ exonuclease) also associate with the intronic miRNA hairpin region and promote the degradation of the fragments[49,50].

When the miRNA hairpin is located in the exonic region, Drosha processing can destabilize the transcript and reduces protein synthesis from it[52]. It was recently shown that Drosha can cleave not only pri-miRNAs but also mRNAs that contain long hairpins[52]. Drosha negatively regulates its own cofactor, DGCR8, by cleaving the hairpins in the second exon of the *DGCR8* mRNA[52]. It would be interesting to establish how widespread this type of mRNA stability control is.

Apart from canonical intronic miRNAs, small groups of miRNA-like RNAs have been discovered in introns in flies and mammals[53–55]. These small RNAs are embedded in short introns, and their biogenesis does not require Drosha processing (FIG. 2c). Following the completion of splicing, the branch point of the lariat-shaped intron is resolved and the debranched intron forms a hairpin structure that resembles pre-miRNA. Some precursors (mirtrons) contain extended tails at either the 5′ or 3′ end, which therefore require exonucleolytic trimming in order to become a substrate for nuclear export (FIG. 2c). Small RNAs that are derived from other non-coding RNAs, such as tRNA[17] or small nucleolar RNA (snoRNA)[56], have also been described. Therefore, it is becoming clear that multiple non-canonical pathways can feed pre-miRNAs into the miRNA pathway through Drosha-independent processes.

Nuclear export by exportin 5. Following nuclear processing, pre-miRNAs are exported to the cytoplasm[57]. Export of pre-miRNAs is mediated by exportin 5 (EXP5), which is a member of the nuclear transport receptor family[58–60]. Although EXP5 was originally known as a minor export factor for tRNAs[61,62], the major cargos of EXP5 turned out to be pre-miRNAs[58–60]. As with other nuclear transport receptors, EXP5 binds cooperatively to its cargo and the GTP-bound form of the cofactor

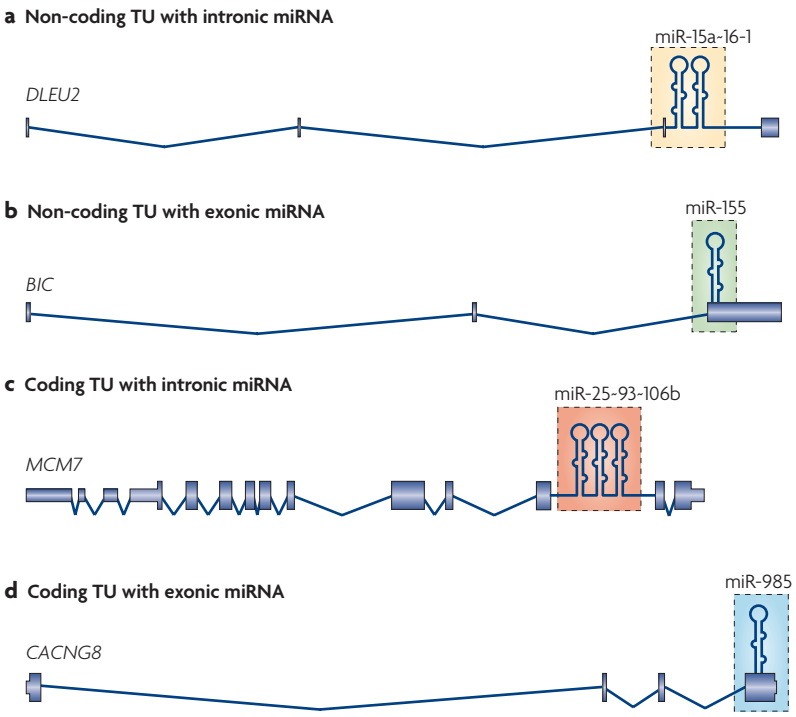


Figure 1 | **Genomic location and gene structure of miRNAs.** MicroRNAs (miRNAs) can be categorized into four groups according to their genomic locations relative to exon and intron positions. **a** | Intronic miRNAs in non-coding transcripts, such as the miR-15a~16-1 cluster. The miR-15a~16-1 cluster is found in the intron of a well-defined non-coding RNA gene, *DLEU2* (REF. 197). **b** | Exonic miRNAs in non-coding transcripts. miR-155 was found in a previously defined non-coding RNA gene, *BIC*[198]. **c** | Intronic miRNAs in protein-coding transcripts. Shown here as an example is the miR-25~93~106b cluster, which is embedded in the intron of the DNA replication licensing factor *MCM7* transcript. **d** | Exonic miRNAs in protein-coding transcripts. The miR-985 hairpin is found in the last exon of *CACNG8* mRNA. The hairpins represent miRNA stem-loops. Blue boxes indicate the protein-coding regions. This figure is roughly to scale. TU, transcription unit.

Stem-loop structure
A lollipop-shaped structure that is formed when a single-stranded nucleic acid molecule loops back on itself to form a complementary double helix (stem) topped by a loop.

Mirtron
A microRNA that is generated from a short spliced intron without Drosha-mediated cleavage.

dsRBD
(Double-stranded-RNA-binding domain). A protein domain that binds to the A-form double-stranded RNA helix. Proteins that contain a dsRBD function in RNA localization, editing, translational repression and post-transcriptional gene silencing.

Ran in the nucleus, and releases the cargo following the hydrolysis of GTP in the cytoplasm. EXP5 recognizes the >14-bp dsRNA stem along with a short 3′ overhang (1–8 nt)[58,63–65].

Cytoplasmic processing by Dicer. Following export from the nucleus, pre-miRNAs are cleaved near the terminal loop by Dicer, releasing ~22-nt miRNA duplexes[66–70] (FIG. 2a). Thus, Drosha predetermines mature miRNA sequences by generating one end of the mature miRNA, whereas the other end is created by Dicer, which measures ~22 nt from the pre-existing terminus of the pre-miRNA (BOX 2). Dicer is a highly conserved protein that is found in almost all eukaryotic organisms, including *Schizosaccharomyces pombe*, plants and animals. Some organisms contain multiple Dicer homologues, whereby different Dicer isotypes have distinct roles[71,72]. For instance, *D. melanogaster* Dicer 1 is required for miRNA biogenesis, whereas Dicer 2 functions in siRNA production[71].

Dicer associates with dsRNA-binding proteins. *D. melanogaster* Dicer 1 requires Loquacious (LOQS; also known as R3D1), which contains three dsRNA-binding domains (dsRBDs) for pre-miRNA processing[73–75]. Human Dicer interacts with two closely related proteins, TRBP (TAR RNA-binding protein; also known as TARBP2)[76,77] and PACT (also known as PRKRA)[78]. Although neither TRBP nor PACT are required for processing activity itself, they seem to contribute to formation of the RNA-induced silencing complex (RISC; see below)[76–78]. The detailed biochemical roles of these proteins remain to be determined. Earlier studies reported that TRBP functions as a negative regulator of dsRNA-dependent protein kinase (PKR; also known as EIF2AK2), whereas PACT stimulates PKR[79]. At present, it is unclear whether there is any functional connection between the miRNA pathway and PKR signalling.

Argonaute loading. Following Dicer cleavage, the resulting ~22-nt RNA duplex is loaded onto an Ago protein so as to generate the effector complex, RISC. One strand of the ~22-nt RNA duplex remains in Ago as a mature miRNA (the guide strand or miRNA), whereas the other strand (the passenger strand or miRNA*) is degraded. Studies on siRNA duplexes indicate that the relative thermodynamic stability of the two ends of the duplex determines which strand is to be selected[80–82]. The strand with relatively unstable base pairs at the 5′ end typically survives (for example, a GU pair compared with a GC pair)[81,82]. Thermodynamic stability profiling studies on miRNA precursors suggested that the same rule might apply to most, although not all, miRNAs[46,82]. Because strand selection is often not a stringent process, some hairpins produce miRNAs from both strands at comparable frequencies.

Dicer, TRBP (and/or PACT) and Ago proteins contribute to RISC assembly by forming a RISC loading complex (RLC) in humans (whereas in flies, RLC comprises Dicer 1, LOQS and AGO1)[76,83–86]. Although it is currently unknown how RLC binds to RNA and facilitates Ago loading, evidence suggests that the miRNA duplex is released from Dicer after cleavage and that the stable end of the RNA duplex is bound to TRBP in the RLC, whereas the other end interacts with the Ago protein[86,87] The mechanistic details of strand selection and RISC assembly have been best delineated with synthetic RNA duplexes in *D. melanogaster* extracts[86,88]. R2D2, a protein with two dsRBDs, senses the thermodynamic inequalities. R2D2 forms a stable heterodimeric complex with Dicer 2, binds to the more stable end of the RNA duplex, and orientates AGO2, on the RNA duplex. It was shown that the endonucleolytic enzymatic activity (slicer activity) of Ago protein is responsible for the removal of the passenger strand of siRNA duplexes and some miRNA duplexes[89–92]. However, most miRNA duplexes (unlike siRNA duplexes) contain mismatches in the middle, and some Ago proteins (for example, AGO1, AGO3 and AGO4 in humans; also known as EIF2C1, EIF2C3 and EIF2C4, respectively) lack slicer activity, which prevents the cleavage of the passenger strand. An RNA helicase activity is thought to mediate the unwinding and removal of the unselected strand of the miRNA duplex.

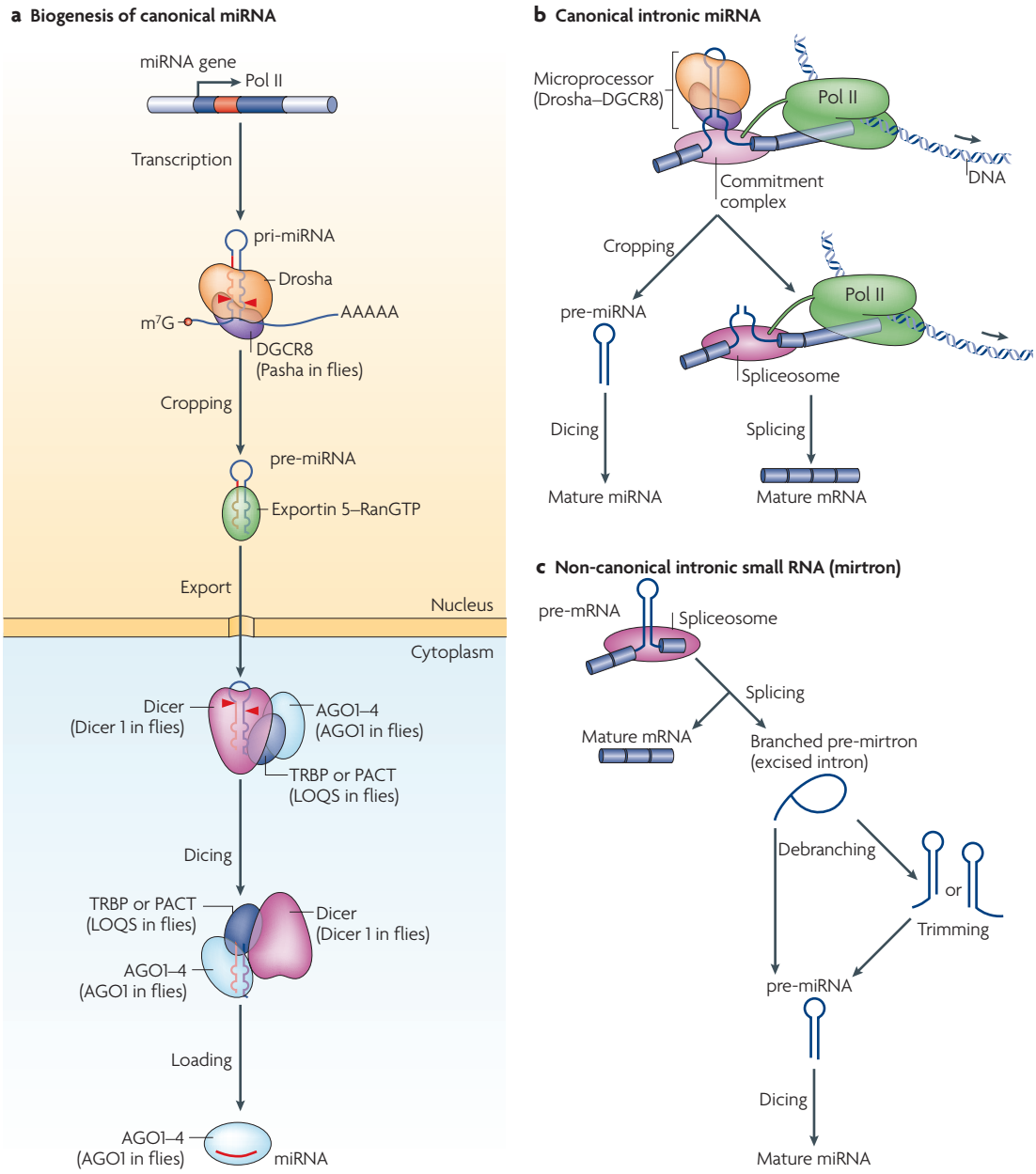


Figure 2 | **miRNA biogenesis pathway. a** | Canonical microRNA (miRNA) genes are transcribed by RNA polymerase II (Pol II) to generate the primary transcripts (pri-miRNAs). The initiation step (cropping) is mediated by the Drosha–DiGeorge syndrome critical region gene 8 (DGCR8; Pasha in *Drosophila melanogaster* and *Caenorhabditis elegans*) complex (also known as the Microprocessor complex) that generates ~65 nucleotide (nt) pre-miRNAs. Pre-miRNA has a short stem plus a ~2-nt 3′ overhang, which is recognized by the nuclear export factor exportin 5 (EXP5). On export from the nucleus, the cytoplasmic RNase III Dicer catalyses the second processing (dicing) step to produce miRNA duplexes. Dicer, TRBP (TAR RNA-binding protein; also known as TARBP2) or PACT (also known as PRKRA), and Argonaute (AGO)1–4 (also known as EIF2C1–4) mediate the processing of pre-miRNA and the assembly of the RISC (RNA-induced silencing complex) in humans. One strand of the duplex remains on the Ago protein as the mature miRNA, whereas the other strand is degraded. Ago is thought to be associated with Dicer in the dicing step as well as in the RISC assembly step. In *D. melanogaster*, Dicer 1, Loquacious (LOQS; also known as R3D1) and AGO1 are responsible for the same process. In flies, most miRNAs are loaded onto AGO1, whereas miRNAs from highly base-paired precursors are sorted into AGO2. The figure shows the mammalian processing pathways with fly components in brackets. **b** | Canonical intronic miRNAs are processed co-transcriptionally before splicing. The miRNA-containing introns are spliced more slowly than the adjacent introns for unknown reasons. The splicing commitment complex is thought to tether the introns while Drosha cleaves the miRNA hairpin. The pre-miRNA enters the miRNA pathway, whereas the rest of the transcript undergoes pre-mRNA splicing and produces mature mRNA for protein synthesis. **c** | Non-canonical intronic small RNAs are produced from spliced introns and debranching. Because such small RNAs (called mirtrons) can derive from small introns that resemble pre-miRNAs, they bypass the Drosha-processing step. Some introns have tails at either the 5′ end or 3′ end, so they need to be trimmed before pre-miRNA export. m^7G, 7-methylguanosine.

Because there are several different Ago proteins in a cell, small RNAs have multiple choices of Ago proteins during RISC assembly. In *D. melanogaster,* the major factor that determines small RNA sorting is the structure of the precursor[93,94]. miRNA duplexes with central mismatches are preferentially sorted into AGO1, whereas perfectly matching siRNA duplexes are incorporated into AGO2. A similar sorting mechanism functions in *C. elegans*, in which the Ago-like proteins ALG-1 and RDE-1 associate with small RNAs from mismatched precursors and perfect dsRNA precursors, respectively[95]. The Dicer 2–R2D2 complex seems to function as a gatekeeper in *D. melanogaster* by favouring perfectly matching siRNA duplexes and discriminating against miRNA duplexes with mismatches, although the molecular mechanism for this is not fully understood. In humans, all four Ago proteins, AGO1–4, bind to miRNAs with only marginal differences in miRNA repertoire[96–98], which suggests that human AGO1–4 might not have significantly differentiated functions.

The ends of miRNAs are often heterogeneous owing to either the addition or deletion of 1–2 nt[98,99]. Sequence variations are found at both 5′ and 3′ ends, although the 3′ ends tend to be much more variable than the 5′ ends. The mechanisms of the variations of the 5′ end are unknown, but they might be explained by imprecise or alternative processing by RNase III. Changes in the 5′ terminus result in shifts of the seed sequences (2–7 nt from the 5′ end), which alter the target specificity of the miRNA. The 3′ end often contains untemplated nucleotides (mostly uracil and adenine), which must be added after processing by unknown terminal uridyl/adenyl transferases. Deletions of the 5′- and 3′-end nucleotides are also often observed, which are probably due to exonucleolytic activities.

Regulation of miRNA biogenesis

Expression profiling studies indicate that most miRNAs are under the control of developmental and/or tissue-specific signalling[100]. Precise control of miRNA levels is crucial to maintain normal cellular functions, and dysregulation of miRNA is often associated with human diseases, such as cancer[101].

Transcriptional control. Transcription is a major point of regulation in miRNA biogenesis. Numerous Pol II-associated transcription factors are involved in transcriptional control of miRNA genes. For instance, myogenic transcription factors, such as myogenin and myoblast determination 1 (MYOD1), bind upstream of miR-1 and miR-133 loci and induce the transcription of these miRNAs during myogenesis[102–104]. Some miRNAs are under the control of tumour-suppressive or oncogenic transcription factors. The tumour suppressor p53 activates the miR-34 family of miRNAs (for a review, see REF. 105), whereas the oncogenic protein MYC transactivates or represses a number of miRNAs that are involved in the cell cycle and apoptosis[106,107]. Epigenetic control also contributes to miRNA gene regulation; the miR-203 locus frequently undergoes DNA methylation in T-cell lymphoma but not in normal T lymphocytes[108].

Post-transcriptional regulation. Drosha processing confers another important point of regulation. miR-21 is induced in response to bone morphogenetic protein (BMP)/transforming growth factor-β (TGFβ) signalling without transcriptional activation[109]. It was proposed that SMAD proteins activated by BMP/TGFβ interact with Drosha and DDX5 (also known as p68) to stimulate Drosha processing, although the detailed mechanism for this remains unclear. Drosha processing of pri-miR-18 is dependent on the heterogeneous ribonucleoprotein particle A1 (REFS 110,111). How many of such regulatory factors exist is unclear, but it is plausible that nuclear RNA-binding proteins influence miRNA processing through specific interactions with a subset of pri-miRNAs.

The let-7 miRNAs show interesting expression patterns[112]. The primary transcript of let-7 (pri-let-7) is expressed in both undifferentiated and differentiated ES cells, whereas mature let-7 is detected only in differentiated cells, indicating that let-7a might be post-transcriptionally controlled[113–115]. Similar post-transcriptional inhibition of let-7 also takes place in tumour cells[114]. Recent studies show that an RNA-binding protein, LIN28, is responsible for the suppression of let-7 biogenesis[116–119]. Several different mechanisms of LIN28 action have been proposed: blockage of Drosha processing[116,117], interference with Dicer processing[118,119] and terminal uridylation of pre-let-7 (REF. 119). Given the cytoplasmic localization of LIN28 (REF. 120) and its strong interaction with pre-let-7 (but not with pri-let-7)[119], LIN28 is likely to function mainly in the cytoplasm by interfering with pre-let-7 processing and/or by inducing terminal uridylation of pre-let-7. The U tail (~14 nt) that is added to the 3′ end of pre-let-7 blocks Dicer processing and facilitates the decay of pre-let-7 (REF. 119). It is unknown how widespread this type of regulation is and which enzyme is responsible for pre-miRNA uridylation.

Turnover of miRNA is a largely unexplored area. RNA decay enzymes might target not only mature miRNAs but also the precursors (pri-miRNAs and pre-miRNAs). Once bound to Ago proteins, mature miRNAs seem to be more stable than average mRNAs; the half-life of most miRNAs is greater than 14 hours[121]. However, certain miRNAs (for example, miR-29b) might be degraded much more rapidly than other miRNAs[121], which suggests a specific recognition of miRNA sequences by nucleases. The 3′→5′ exonuclease ERI1 (also known as THEX1) was previously shown to be responsible for the degradation of siRNAs in *C. elegans*[122]. A group of exoribonucleases named small RNA degrading nuclease (SDN) proteins were recently reported to affect the stability of miRNAs in plants[123]. However, it remains unclear which nucleases are responsible for miRNA degradation in animals.

RNA editing is another possible way of regulating miRNA biogenesis. The alteration of adenines to inosines, a process that is mediated by adenine deaminases (ADARs), has been observed in miR-142 (REF. 124) and miR-151 (REF. 125). Because the modified pri-miRNAs or pre-miRNAs become poor substrates of RNase III proteins, editing of the precursor can interfere with miRNA processing. Editing can also change the target specificity of the miRNA if it occurs in miRNA sequences[126].

An increasing number of miRNAs are controlled at the post-transcriptional level, yet the mechanisms of regulation remain poorly understood for most miRNAs. miR-138 is specifically expressed in neuronal cells, whereas its expression is suppressed at the Dicer-mediated processing step in non-neuronal cells[127]. Human miR-31, miR-128 and miR-105, however, might be controlled at the nuclear export step because the precursors are retained in the nucleus without producing mature miRNA in certain cell types[128]. Mature miR-7, miR-143 and miR-145 show reduced expression in cancer cells compared with normal tissue, although the precursor levels are similar between the tumour and normal tissues, which suggests that post-transcriptional misregulation occurs in cancer cells[129,130].

Feedback circuits in miRNA networks. miRNA biogenesis is controlled by multiple layers of feedback loops that involve the biogenesis factors, the miRNAs themselves and their targets[131]. Two types of feedback circuits are frequently observed: single-negative feedback and double-negative feedback. Single-negative feedback usually results in stable or oscillatory expression of both components, whereas double-negative feedback serves as a bistable switch that results in mutually exclusive expression.

Levels of Drosha and Dicer are controlled by single-negative feedback to maintain the homeostasis of miRNA production[52,132,133]. Drosha constitutes a regulatory circuit together with DGCR8; Drosha downregulates DGCR8 by cleaving *DGCR8* mRNA, whereas DGCR8 upregulates Drosha through protein stabilization[52,191]. This loop seems to be highly effective because even when the DGCR8 gene copy number is reduced by one-half in *Dgcr8* heterozygous cells, an almost normal level of DGCR8 protein is produced and miRNA levels are also unaffected. Human Dicer is controlled by its own product, let-7, which binds to the 3′ UTR and coding region of the *Dicer* mRNA[132,133]. This might explain why Dicer knockdown is often more transient and moderate than knockdown of other genes. Furthermore, numerous examples of single-negative feedback have been described in worms, flies and mammals, in which a transcription factor that transactivates miRNA is itself repressed by that same miRNA[131].

Double-negative feedback control is also often used as an effective genetic switch of specific miRNAs during differentiation. One interesting example is the conserved loop that involves let-7 and LIN28 (REFS 116,117,119,134). let-7 suppresses LIN28 protein synthesis, whereas LIN28 blocks let-7 maturation. The miR-200 family and the transcriptional repressors ZEB1 and ZEB2 also constitute a double-negative feedback loop that functions in epithelial–mesenchymal transition[135].

Piwi-interacting RNAs

piRNAs were originally discovered during small RNA profiling studies of *D. melanogaster* development[136,137]. These studies uncovered a subset of endogenous, germ-cell-specific small RNAs (24–29 nt) that were clearly distinct in their size from miRNAs. Most of this longer species corresponded to intergenic repetitive elements, including retrotransposons. Therefore, they were termed repeat-associated small interfering RNAs (rasiRNAs)[137]. Earlier studies in *D. melanogaster* suggested that small RNAs that correspond to retrotransposons might be involved in the silencing of transposable elements[136]. In addition, PIWI protein was shown to be essential for self-renewal of germline stem cells[138–140] and for transposon mobility control[141,142]. Aubergine (AUB), another Piwi-subfamily protein, is required for pole-cell formation[143] and for repression of retrotransposons[144]. Later, Vagin *et al.* showed that *piwi* and *aub* mutations resulted in a loss of rasiRNA accumulation in ovaries[14], strengthening the connection between rasiRNAs and Piwi proteins. The physical interaction of rasiRNAs with Piwi proteins was revealed by immunopurification of Piwi complexes, mostly by using specific antibodies against each of the Piwi proteins, including the third member of the subfamily, AGO3 (REFS 14,145–148) (BOX 1; TABLE 1).

The Piwi-subfamily proteins in mice (MIWI, MILI and MIWI2; also known as PIWIL1, PIWIL2 and PIWIL4, respectively)[149–152], as well as those of zebrafish (Ziwi and Zili; also known as Piwil1 and Piwil2, respectively)[153], were also shown to be associated with small RNAs that resemble rasiRNAs. These small RNAs (24–31 nt) were termed piRNAs. Mutations in *Mili* and *Miwi2* derepressed transposable elements, suggesting a function for mouse piRNAs in transposon control[154–156]. On the basis of their common features, *D. melanogaster* rasiRNAs are now also referred to as piRNAs[155,157].

piRNA biogenesis in flies. piRNAs are produced from intergenic repetitive elements on chromosomes. Most piRNAs can be mapped to particular loci, which are termed piRNA clusters. Interestingly, Piwi-associated piRNAs include those that are derived from a particular piRNA cluster locus, *flamenco*, on the X chromosome[146]. The regulatory system that leads to such locus discrimination remains unclear. Furthermore, piRNAs that are associated with each of the *D. melanogaster* Piwis differ in size from each other[146,147]. The differences in subcellular localization, the expression patterns and the associated proteins might influence the selective piRNA association (BOX 1; TABLE 1).

AUB- and PIWI-associated piRNAs are mainly derived from antisense transcripts of retrotransposons, whereas AGO3-associated piRNAs arise mostly from sense transcripts[14,145–148]. Nucleotide bias is also observed: AUB- and PIWI-associated piRNAs show strong preferences for uracil at their 5′ ends, whereas AGO3-associated piRNAs mostly have adenine at nucleotide 10 but show no bias at the 5′ ends[146,147]. AUB-associated piRNAs frequently show complementarity to AGO3-associated piRNAs in their first 10 nt[146,147]. Judging from the fact that piRNAs accumulate in *Dicer* mutant ovaries[14], piRNA production in the fly ovary does not depend on Dicer. The requirement for Drosha has not been formally tested. So, how are piRNAs produced without Dicer? One hypothesis is that the Piwi proteins themselves might be involved in piRNA biogenesis, given that they possess slicer nuclease activity; Piwi cleaves target RNA at between positions 10 and 11 relative to the 5′ end of the associated small RNAs[145]. Based on this and the characteristics of piRNAs

Retrotransposon
A transposon that mobilizes through RNA intermediates; the DNA elements are transcribed into RNA and then reverse-transcribed into DNA, which is inserted at new sites in the genome.

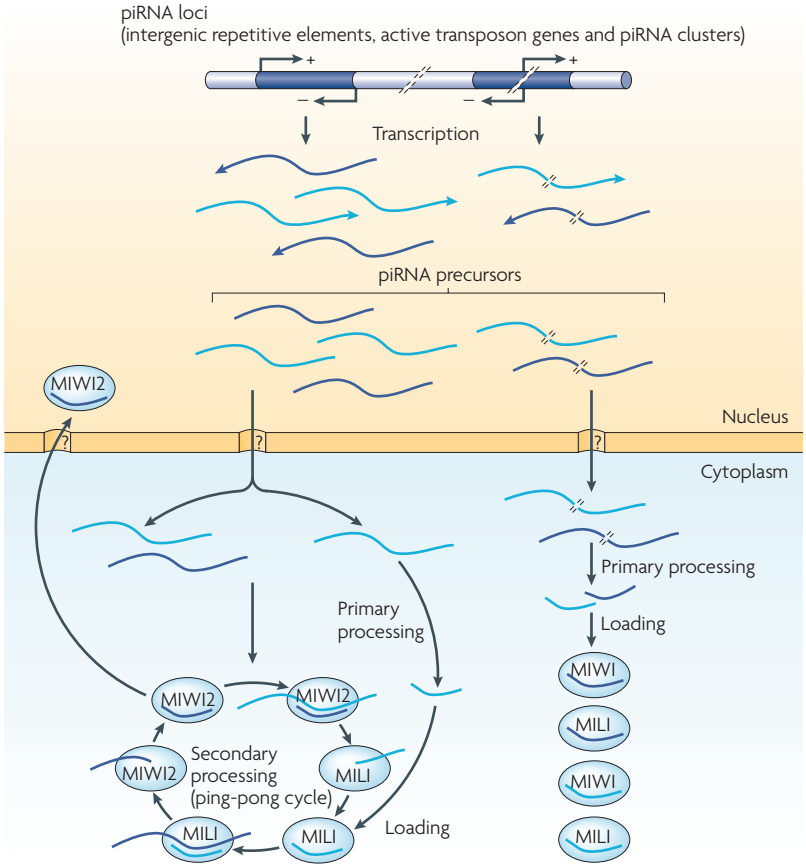


Figure 3 | **piRNA biogenesis pathway.** Piwi-interacting RNAs (piRNAs) of 24–31 nucleotides (nt) are processed from single-stranded RNA precursors that are transcribed from intergenic repetitive elements, transposons or large piRNA clusters. piRNAs associate with Piwi-subfamily proteins; they are not dependent on either Drosha or Dicer, although the requirement for Drosha has not been formally tested. They might instead use the nuclease activity of the Piwi proteins themselves for their processing. piRNA biogenesis involves primary and secondary processing mechanisms. The strand bias of piRNAs in secondary processing differs between mice and flies (for this reason, fly components are not shown). Primary processing and loading might occur in the cytoplasm because Piwi proteins (MIWI and MILI; also known as PIWIL1 and PIWIL2, respectively) are localized in the cytoplasm. Factors that are needed for primary processing are unknown. In the secondary processing step (ping-pong cycle), MILI introduces a cleavage in the precursor and thereby defines the 5′ end of piRNA, which is subsequently accepted by MIWI2 (also known as PIWIL4). MIWI2 also cleaves the opposite strand precursor through its slicer activity, generating the 5′ end of the piRNA that subsequently binds to MILI. The nuclease that creates the 3′ end of piRNA is unknown. MIWI2 seems to be transported into the nucleus.

described above, a model for piRNA biogenesis has been proposed[146,147]. AUB or PIWI that is associated with antisense piRNA cleaves sense retrotransposon transcripts, and this cleavage creates the 5′ ends of sense piRNA that in turn associates with AGO3. Newly made AGO3 that is associated with sense piRNA subsequently cleaves antisense retrotransposon transcripts and thereby makes the 5′ end of antisense piRNAs that subsequently bind to AUB or PIWI. Factors that are necessary for the formation of the 3′ end of piRNAs have yet to be identified. These reciprocal reactions, known as the ping-pong cycle, would occur continuously *in vivo*, thereby amplifying the piRNA population[146,147]. Concomitantly, retrotransposon silencing can be maintained.

Mutations in several genes, such as those that encode the putative nucleases Squash and Zucchini[158], and others such as *Spindle-E*[14], *Krimper*[159] and *Maelstrom*[160], cause the depletion of piRNAs in fly ovaries, suggesting that they are involved in piRNA biogenesis. However, the precise roles of these proteins remain elusive. piRNAs derived from *flamenco* (*flam*-piRNAs) are apparently independent of the ping-pong pathway because they are found only in PIWI. This separate pathway for *flam*-piRNAs is now referred to as the primary processing pathway[161].

How does such a piRNA biogenesis cycle initiate during development? In *D. melanogaster*, at least AUB and possibly PIWI are deposited for the next generation by germline transmission[146,148]. The maternal loading of Piwi proteins into embryos was also observed in fish[153]. A recent study shows that piRNAs that are maternally inherited to embryos have an epigenetic regulatory role in transposon silencing[162]. This study explains how a phenomenon called hybrid dysgenesis is related to the loss of piRNA inheritance. Transmitted piRNAs, presumably in a form that is associated with Piwi proteins, most likely act as primary 'seeds' that initiate an amplification cycle of piRNA biogenesis in embryos, and thus, the cycle could operate between generations.

piRNA biogenesis in mice. In mammals, two classes of Piwi-interacting RNAs (piRNAs) have been identified[149–152,163]. One class of mouse piRNAs (pre-pachytene piRNAs) is expressed before meiotic pachytene and is derived from repeat- and transposon-rich clusters[149]. In this regard, the pre-pachytene piRNAs are similar to *D. melanogaster* piRNAs. Pre-pachytene piRNAs interact with two members of the mouse Piwi proteins, MILI and MIWI2 (REFS 155,156,164). The other class of piRNAs becomes abundant during the pachytene stage and associates with MILI and MIWI[150,155,164]. These pachytene piRNAs remain an enigma, as their sequences give no clues as to their possible targets. The pachytene piRNAs are extremely abundant in spermatocytes: >80,000 distinct species are derived from large genomic clusters of up to 200 kilobases. These piRNA clusters exhibit a marked strand asymmetry, as if the piRNAs are processed from one or a few huge transcripts. Intriguingly, piRNA sequences are not conserved among mammals. However, they are produced from syntenic regions of the mouse, rat and human genomes.

Data from studies of mouse piRNAs support the idea that the ping-pong mechanism originally proposed in *D. melanogaster* might also apply to mouse pre-pachytene piRNAs[164] (FIG. 3). However, the different compositions of the embryonic, neonatal and adult piRNAs suggest that the biogenesis cycle does not continue throughout male germ-cell development. In addition, pachytene piRNAs do not show the obvious characteristics that are observed for piRNAs produced by the ping-pong pathway. Thus, the primary processing pathway is likely to be at work as in the case of *D. melanogaster*[164], although it is currently unclear whether or not the primary processing pathways in flies and mice are mechanically similar (FIG. 3).

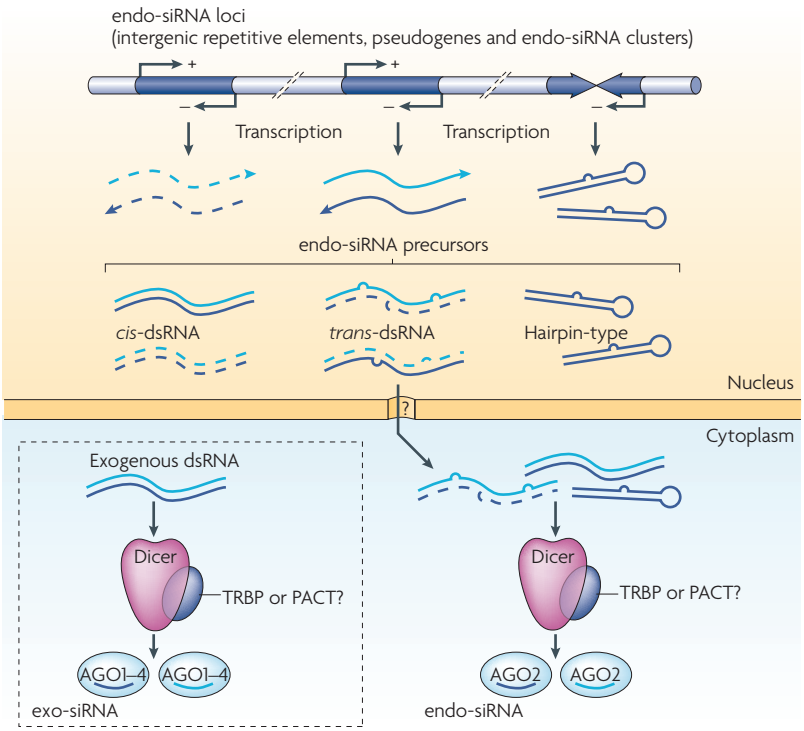


Figure 4 | **Exo- and endo-siRNA biogenesis pathway.** Exogenous small interfering RNAs (exo-siRNAs; see inset) are derived from experimentally introduced double-stranded RNAs (dsRNAs) or viral RNAs. Human exo-siRNAs are loaded onto Argonaute (AGO)1–4 (also known as EIF2C1–4), but only AGO2 has slicer activity. In flies, a complex that comprises Dicer 2 and R2D2 cleaves the exo-siRNA precursor, and AGO2 accept the RNAs. Endogenous siRNA (endo-siRNA) precursors are derived from repetitive sequences, sense–antisense pairs or long stem-loop structures. In humans, Dicer and AGO2 function in the endo-siRNA pathway; it has not been determined whether human endo-siRNAs associate with other Ago proteins, or whether TRBP (TAR RNA-binding protein; also known as TARBP2) and/or PACT (also known as PRKRA) are involved in this pathway (see question mark). In flies, Dicer 2, together with Loquacious (LOQS; also known as R3D1), processes the precursor into ~21-nt endo-siRNAs, which are accepted by AGO2.

As in flies, some mouse piRNAs correspond to transposon sequences, implying that they also function in silencing selfish DNA elements. However, even though the functional outcome is similar, whether they use the same strategy for silencing is questionable. In mouse testes, the promoter regions of particular retrotransposons are methylated. Mutations in *Mili* and *Miwi2* eliminated DNA methylation of long interspersed nuclear element 1 (LINE1) and intracisternal A particle (IAP) and led to male infertility[154,155]. MILI and MIWI2 have essential roles in establishing *de novo* DNA methylation of transposons in fetal male germ cells[156,164]. These results suggest that mouse piRNAs are involved in transcriptional silencing of target genes through DNA methylation.

Unlike animal miRNAs but similar to plant miRNAs, piRNAs have a 2′-*O*-methyl modification at their 3′ ends[14,165–168]. This modification is carried out by homologues of *A. thaliana* HEN1 methyltransferase[169] (known as HEN1 or PIMET in flies)[165,166]. It was also shown that HEN1 associates with Piwi proteins in ovaries. However, the biological significance of the 2′-*O*-methyl modification remains unknown because *D. melanogaster* mutants of HEN1 show no obvious malfunctions[165].

RNA-dependent RNA polymerase
An RNA polymerase that transcribes RNAs from RNA templates.

Endogenous siRNAs

Deep sequencing of small RNAs in *D. melanogaster* somatic tissue, cultured cells and ovaries has identified a novel class of ~21-nt RNAs[17–20,170–172]. These RNAs are derived from transposon transcripts, sense–antisense transcript pairs and long stem-loop structures[17–20,170–172]. Specific association of these endo-siRNAs with AGO2 was shown biochemically in flies[19,20]. Extensive small-RNA profiling studies in mice also revealed that numerous types of endo-siRNAs are present in oocytes[15,16] and less abundantly in ES cells[17]. Mouse endo-siRNAs are also dependent on the Ago-family protein (mouse AGO2; also known as EIF2C2)[15].

Apart from the endo-siRNA pathway, *D. melanogaster* AGO2 functions in the RNA interference pathway by associating with exogenous siRNAs (exo-siRNAs; which are ~21 nt in length) that arise from ectopically introduced long dsRNAs or siRNA duplexes[96]. For example, flies use the RNA interference mechanism to defend against viruses that produce dsRNAs during infection[173,174].

Numerous endogenous siRNAs have been described in other species, such as plants and *C. elegans*[29]. Endo-siRNA pathways in plants and worms are more complex than those in flies and mammals, and they possess RNA-dependent RNA polymerases (RdRPs)[29] for the generation of endo-siRNAs. Because flies and mammals lack RdRP, fly and mammalian endo-siRNAs should be categorized as a new group of endo-siRNAs that are produced in an RdRP-independent manner.

Endo-siRNA processing by Dicer 2 in flies. The size distribution of endo-siRNAs is narrow and concise (~21 nt) compared with, for example, the 21–23-nt lengths of miRNAs. The biogenesis system probably holds the key to this difference. Similar to exo-siRNAs, but different from miRNAs, the processing of endo-siRNAs is dependent on Dicer 2 rather than Dicer 1 (REFS 18–20,170–172). Interestingly, unlike exo-siRNAs, which require R2D2 (REF. 88), endo-siRNAs, particulary those that originate from long stem-loops, rely on LOQS, which functions in the miRNA pathway[18–20,73,75,170–172]. Dicer 2 seems to be able to associate with LOQS in *D. melanogaster* cells[146]. How and why LOQS selects Dicer 2 over Dicer 1 as a partner in endo-siRNA processing is unknown. It would also be interesting to understand how the Dicer 2–LOQS complex recognizes its substrates and determines the cleavage sites.

Endo-siRNA precursors are mainly produced from sense–antisense pairs derived from transposons. They can also arise from convergent transcription of protein-coding genes and from unannotated regions of the genome[19,20,170,171]. These transcripts are not necessarily transcribed from the same loci; thus, the dsRNA precursors tend to have natural mismatches and bulges. The second type of endo-siRNA precursor is a single-stranded, but self-hybridizing, transcript that forms a long stem-loop structure[19,20,170,171]. They are distinguished from miRNA precursors by the extended length of the stems. Interestingly, a significant number of endo-siRNAs (~20% of AGO2-associated endo-siRNAs

in S2 cells) show sequence substitutions[19]. As most of the mutations found were A–G substitutions, this is probably due to RNA editing by ADAR[175]. The editing activity of ADAR is restricted in the nucleus and accepts only dsRNAs as the substrate. Therefore, the dsRNA precursors must have been formed in the nucleus, in which they serve as the substrates for ADAR. This post-transcriptional nucleotide modification causes further tiny bulges in the precursors. These two structural features of endo-siRNA precursors (long dsRNA structures and bulges) might be recognized selectively by the processing factors Dicer 2 and LOQS. Dicer 2 might be required for long dsRNA processing and for interaction with AGO2. LOQS might contribute to the binding of Dicer 2 to mismatched dsRNAs and to their assembly into RISC.

Some piRNA-generating loci can be a source of endo-siRNAs[19], although Czech *et al.*[20] reported a controversial observation that only a few piRNA loci produce endo-siRNAs. Like piRNAs, endo-siRNAs have 2′-*O*-methyl groups at their 3′ ends in flies, which might be added by the methyltransferase HEN1 (REF. 19). However, piRNAs and endo-siRNAs are clearly different with regard to their size (24–29-nt compared with 21-nt), their protein partners (Piwi proteins compared with AGO2) and the cells in which they are expressed (mostly germ cells compared with ubiquitous expression). Their editing status is also unequal; endo-siRNAs can be edited by ADAR, but this is unlikely to be the case for piRNAs, which implies that endo-siRNAs are generated from dsRNAs, whereas piRNAs originate from ssRNAs. Retrotransposons can be transcribed in both directions. However, the amounts of the sense and antisense products may not be equal. This would result in a mixture of dsRNAs and ssRNAs. Regardless of editing, the dsRNAs are transported to the cytoplasm by an unknown mechanism, and serve as substrates for Dicer 2 activity to generate endo-siRNAs. The single-stranded transcripts are also transported to the cytoplasm, probably through the mRNA export pathway, and become substrates for the piRNA-producing machineries.

Endo-siRNA biogenesis in mouse oocytes. Like *D. melanogaster* endo-siRNAs, mouse endo-siRNAs are ~21 nt long and are derived from a range of sources, including transposable elements[15,16]. Common features can also be found in their biogenesis. Dicer is required for mouse endo-siRNA production, as is Dicer 2 in *D. melanogaster* (FIG. 4). The precursors of mouse endo-siRNAs are transcripts that contain long hairpin structures or dsRNAs that are derived from sense–antisense pairs. Some of these sense–antisense pairs (*cis*-endo-siRNA precursors) are transcribed from convergent transcription of the same loci, whereas others (*trans*-endo-siRNA precursors) are derived from two similar but separate loci. It was shown that pseudogene transcripts can generate siRNAs by annealing to their cognate functional transcripts, and that the functional genes are upregulated in the *Dicer* and *Ago2* mutants[15,16]. Thus, pseudogenes might have a regulatory role by suppressing their functional counterparts through an RNA silencing mechanism.

Interestingly, mouse oocytes produce both endo-siRNAs and piRNAs, whereas mouse testes express only piRNAs[15,16]. Thus, in ovaries, transposon silencing is ensured by two pathways: piRNA and endo-siRNA pathways. During evolution, transposons in male and female mouse germ lines may have been regulated in different ways. Although some of them still remain active in females (in which they might confer some advantage in oocyte development), the endo-siRNA pathway might have been lost in males.

Conclusions

Tens of thousands of small RNAs have been discovered, and more small RNAs are expected to be discovered owing to rapidly expanding sequencing power. This is an exciting era in RNA biology, but at the same time we are facing multiple challenges. Most of all, it is difficult to distinguish between functional small RNAs and non-functional 'noises'. In fact, there are numerous misannotated small RNA entries in databases, which are apparently degradation products from other longer RNA species. Evolutionary conservation confers compelling evidence for the functionality of the cloned RNAs, but the small RNAs will eventually need to be experimentally validated to prove their functionality. Confirmation of their expression and their dependence on the Ago proteins will first be needed to validate individual small RNAs. Dependence on other biogenesis factors, such as Drosha, DGCR8 and Dicer, will also be useful in validating small RNAs as well as in classifying them[17].

The miRNA biogenesis pathway is well studied in comparison to piRNA and endo-siRNA pathways, although many questions remain unanswered. A more detailed understanding of the mechanism awaits the structures of the complexes, including Microprocessor, EXP5 and Dicer–RISC in association with the substrate RNAs. Many factors are implicated in miRNA biogenesis, such as DDX5, DDX17 (also known as p72), NFAR (also known as ILF3)[40,176,177], SNIP1 (REF. 178), PACT[78] and Mei-P26 (REF. 179), the biochemical roles of which are unknown. Furthermore, additional protein factors in the pathway need to be identified. For instance, we do not know the identities of the factors that are involved in miRNA turnover in animals. We also need to understand the significance and enzymology of the modifications of small RNAs, such as uridylation and adenylation of miRNAs. Moreover, we have yet to learn about the auxiliary factors that regulate miRNA maturation. It would be interesting to understand how miRNA biogenesis is controlled by these factors in response to various cellular signals. Developmental and environmental cues should be interpreted accurately by these factors to correctly operate the miRNA network.

The piRNA pathway is still largely elusive. piRNAs are highly diverse in sequence, and many of them are derived from non-repetitive sequences, especially in mice. Because the ping-pong model does not explain all characteristics of piRNAs, it is likely that an additional biogenesis pathway exists for certain piRNAs, particularly those that are associated with Piwi in *D. melanogaster* and pachytene piRNAs in mice. It would also be interesting to

understand the mechanistic details of how piRNAs relate to heterochromatin formation. Identification and analyses of the proteins associated with the Piwi proteins might provide clues.

There are numerous other small RNAs, like mirtrons, that are generated through non-canonical pathways. Many of these are difficult to classify and their biogenesis pathways remain poorly understood. Although most of these non-canonical small RNAs are less abundant and less conserved than major small RNAs, such as miRNAs, they might have species-specific functions that are not yet fully appreciated.

Another interesting issue is how small RNA pathways are connected to other aspects of RNA metabolism, including transcription, pre-mRNA splicing and mRNA decay. Given that intronic miRNAs are processed co-transcriptionally, it is tempting to speculate that these processes (transcription, splicing and Drosha processing) are coupled by specific factors. It would be interesting to unveil the mechanistic link between these nuclear events. It is also of note that animal RNase III, Drosha and Dicer are involved not only in small RNA pathways but also mRNA stability control of hairpin-containing mRNAs[52,180]. Further studies are needed to identify additional substrates of RNase III proteins.

Small RNAs are integral components of immensely complicated gene networks. Although individual miRNAs suppress their targets only moderately, miRNAs can exert broad and strong effects. This is possible because they target multiple genes, and because they are often engaged in feedback loops along with other regulatory factors. Thus, miRNAs seem to serve as hubs of gene networks that are rich in information flow. In-depth knowledge of small RNAs will help us to better understand gene networks, and to acquire safer and more effective strategies for genetic manipulation.

1. Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
2. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
3. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
4. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
5. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
6. Lu, C. *et al.* Elucidation of the small RNA component of the transcriptome. *Science* **309**, 1567–1569 (2005).
7. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
8. Lai, E. C., Tomancak, P., Williams, R. W. & Rubin, G. M. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4**, R42 (2003).
9. Nam, J. W. *et al.* Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Res.* **33**, 3570–3581 (2005).
10. Li, S. C., Pan, C. Y. & Lin, W. C. Bioinformatic discovery of microRNA precursors from human ESTs and introns. *BMC Genomics* **7**, 164 (2006).
11. Huang, T. H. *et al.* MiRFinder: an improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinformatics* **8**, 341 (2007).
12. Chu, C. Y. & Rana, T. M. Small RNAs: regulators and guardians of the genome. *J. Cell Physiol.* **213**, 412–419 (2007).
13. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).
14. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
The first paper to describe that piRNAs are produced in a Dicer-independent manner. The authors also reported that *D. melanogaster* piRNAs are modified at their 3′ ends.
15. Watanabe, T. *et al.* Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**, 539–543 (2008).
16. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**, 534–538 (2008).
References 15 and 16 contributed to the identification of many endo-siRNAs in mouse oocytes and also to the proposal of the functions of pseudogenes in silencing the parental genes.
17. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).
18. Okamura, K. *et al.* The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453**, 803–806 (2008).
19. Kawamura, Y. *et al. Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793–797 (2008).
20. Czech, B. *et al.* An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453**, 798–802 (2008).
21. Jones-Rhoades, M. W., Bartel, D. P. & Bartel, B. MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53 (2006).
22. Mallory, A. C. & Vaucheret, H. Functions of microRNAs and related small RNAs in plants. *Nature Genet.* **38**, S31–S36 (2006).
23. Zhang, B., Pan, X., Cobb, G. P. & Anderson, T. A. Plant microRNA: a small regulatory molecule with big impact. *Dev. Biol.* **289**, 3–16 (2006).
24. Verdel, A. & Moazed, D. RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Lett.* **579**, 5872–5878 (2005).
25. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
26. Kim, D. H., Saetrom, P., Snove, O. Jr. & Rossi, J. J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl Acad. Sci. USA* **105**, 16230–16235 (2008).
27. Grimson, A. *et al.* Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197 (2008).
28. Ibanez-Ventoso, C., Vora, M. & Driscoll, M. Sequence relationships among *C. elegans*, *D. melanogaster* and human microRNAs highlight the extensive conservation of microRNAs in biology. *PLoS ONE* **3**, e2818 (2008).
29. Chapman, E. J. & Carrington, J. C. Specialization and evolution of endogenous small RNA pathways. *Nature Rev. Genet.* **8**, 884–896 (2007).
30. Millar, A. A. & Waterhouse, P. M. Plant and animal microRNAs: similarities and differences. *Funct. Integr. Genomics* **5**, 129–135 (2005).
31. Ventura, A. *et al.* Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* **132**, 875–886 (2008).
32. Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).
33. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
34. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957–1966 (2004).
35. Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nature Struct. Mol. Biol.* **13**, 1097–1101 (2006).
36. Lee, Y. S. & Dutta, A. MicroRNAs in cancer. *Annu. Rev. Pathol.* 25 Sep 2008 (doi:10.1146/annurev.pathol.4.110807.092222).
37. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
38. Han, J. *et al.* The Drosha–DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
39. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).
40. Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
41. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* **14**, 2162–2167 (2004).
42. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genet.* **39**, 380–385 (2007).
43. Filippov, V., Solovyev, V., Filippova, M. & Gill, S. S. A novel type of RNase III family proteins in eukaryotes. *Gene* **245**, 213–221 (2000).
44. Wu, H., Xu, H., Miraglia, L. J. & Crooke, S. T. Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. *J. Biol. Chem.* **275**, 36957–36965 (2000).
45. Fortin, K. R., Nicholson, R. H. & Nicholson, A. W. Mouse ribonuclease III. cDNA structure, expression analysis, and chromosomal location. *BMC Genomics* **3**, 26 (2002).
46. Han, J. *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha–DGCR8 complex. *Cell* **125**, 887–901 (2006).
47. Zeng, Y. & Cullen, B. R. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J. Biol. Chem.* **280**, 27595–27603 (2005).
48. Kim, Y. K. & Kim, V. N. Processing of intronic microRNAs. *EMBO J.* **26**, 775–783 (2007).
49. Morlando, M. *et al.* Primary microRNA transcripts are processed co-transcriptionally. *Nature Struct. Mol. Biol.* **15**, 902–909 (2008).
50. Pawlicki, J. M. & Steitz, J. A. Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *J. Cell Biol.* **182**, 61–76 (2008).
51. Dye, M. J., Gromak, N. & Proudfoot, N. J. Exon tethering in transcription by RNA polymerase II. *Mol. Cell* **21**, 849–859 (2006).
52. Han, J. *et al.* Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* **136**, 75–84 (2009).

53. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448**, 83–86 (2007).
54. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. & Lai, E. C. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* **130**, 89–100 (2007).
55. Berezikov, E., Chung, W. J., Willis, J., Cuppen, E. & Lai, E. C. Mammalian mirtron genes. *Mol. Cell* **28**, 328–336 (2007).
56. Ender, C. *et al.* A human snoRNA with microRNA-like functions. *Mol. Cell* **32**, 519–528 (2008).
57. Kim, V. N. MicroRNA precursors in motion: exportin-5 mediates their nuclear export. *Trends Cell Biol.* **14**, 156–159 (2004).
58. Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
59. Yi, R., Doehle, B. P., Qin, Y., Macara, I. G. & Cullen, B. R. Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA* **11**, 220–226 (2005).
60. Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).
61. Bohnsack, M. T. *et al.* Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. *EMBO J.* **21**, 6205–6215 (2002).
62. Calado, A., Treichel, N., Muller, E. C., Otto, A. & Kutay, U. Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *EMBO J.* **21**, 6216–6224 (2002).
63. Gwizdek, C. *et al.* Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J. Biol. Chem.* **278**, 5505–5508 (2003).
64. Basyuk, E., Suavet, F., Doglio, A., Bordonne, R. & Bertrand, E. Human let-7 stem-loop precursors harbor features of RNase III cleavage products. *Nucleic Acids Res.* **31**, 6593–6597 (2003).
65. Zeng, Y. & Cullen, B. R. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.* **32**, 4776–4785 (2004).
66. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
67. Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
68. Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
69. Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
70. Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
71. Lee, Y. S. *et al.* Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**, 69–81 (2004).
72. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
73. Forstemann, K. *et al.* Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* **3**, e236 (2005).
74. Jiang, F. *et al.* Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* **19**, 1674–1679 (2005).
75. Saito, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Processing of pre-microRNAs by the Dicer-1–Loquacious complex in *Drosophila* cells. *PLoS Biol.* **3**, e235 (2005).
76. Chendrimada, T. P. *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740–744 (2005).
77. Haase, A. D. *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* **6**, 961–967 (2005).
78. Lee, Y. *et al.* The role of PACT in the RNA silencing pathway. *EMBO J.* **25**, 522–532 (2006).
79. Garcia, M. A., Meurs, E. F. & Esteban, M. The dsRNA protein kinase PKR: virus and cell control. *Biochimie* **89**, 799–811 (2007).
80. Aza-Blanc, P. *et al.* Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12**, 627–637 (2003).
81. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
82. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).
83. Maniataki, E. & Mourelatos, Z. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* **19**, 2979–2990 (2005).
84. MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V. & Doudna, J. A. *In vitro* reconstitution of the human RISC-loading complex. *Proc. Natl Acad. Sci. USA* **105**, 512–517 (2008).
85. Gregory, R. I., Chendrimada, T. P., Cooch, N. & Shiekhattar, R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123**, 631–640 (2005).
86. Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
87. Preall, J. B. & Sontheimer, E. J. RNAi: RISC gets loaded. *Cell* **123**, 543–545 (2005).
88. Liu, Q. *et al.* R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
89. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P. & Zamore, P. D. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123**, 607–620 (2005).
90. Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H. & Siomi, M. C. Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev.* **19**, 2837–2848 (2005).
91. Leuschner, P. J., Ameres, S. L., Kueng, S. & Martinez, J. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep.* **7**, 314–320 (2006).
92. Diederichs, S. & Haber, D. A. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* **131**, 1097–1108 (2007).
93. Tomari, Y., Du, T. & Zamore, P. D. Sorting of *Drosophila* small silencing RNAs. *Cell* **130**, 299–308 (2007).
94. Forstemann, K., Horwich, M. D., Wee, L., Tomari, Y. & Zamore, P. D. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**, 287–297 (2007).
95. Steiner, F. A. *et al.* Structural features of small RNA precursors determine Argonaute loading in *Caenorhabditis elegans*. *Nature Struct. Mol. Biol.* **14**, 927–933 (2007).
96. Liu, J. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
97. Meister, G. *et al.* Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
98. Azuma-Mukai, A. *et al.* Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc. Natl Acad. Sci. USA* **105**, 7964–7969 (2008).
99. Seitz, H., Ghildiyal, M. & Zamore, P. D. Argonaute loading improves the 5′ precision of both microRNAs and their miRNA strands in flies. *Curr. Biol.* **18**, 147–151 (2008).
100. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
101. Jiang, Q. *et al.* miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **37**, D98–D104 (2008).
102. Chen, J. F. *et al.* The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genet.* **38**, 228–233 (2006).
103. Kim, H. K., Lee, Y. S., Sivaprasad, U., Malhotra, A. & Dutta, A. Muscle-specific microRNA miR-206 promotes muscle differentiation. *J. Cell Biol.* **174**, 677–687 (2006).
104. Rao, P. K., Kumar, R. M., Farkhondeh, M., Baskerville, S. & Lodish, H. F. Myogenic factors that regulate expression of muscle-specific microRNAs. *Proc. Natl Acad. Sci. USA* **103**, 8721–8726 (2006).
105. He, L., He, X., Lowe, S. W. & Hannon, G. J. microRNAs join the p53 network — another piece in the tumour-suppression puzzle. *Nature Rev. Cancer* **7**, 819–822 (2007).
106. He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828–833 (2005).
107. Chang, T. C. *et al.* Widespread microRNA repression by Myc contributes to tumorigenesis. *Nature Genet.* **40**, 43–50 (2008).
108. Bueno, M. J. *et al.* Genetic and epigenetic silencing of microRNA-203 enhances *ABL1* and *BCR-ABL1* oncogene expression. *Cancer Cell* **13**, 496–506 (2008).
109. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56–61 (2008).
110. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14**, 591–596 (2007).
111. Michlewski, G., Guil, S., Semple, C. A. & Caceres, J. F. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol. Cell* **32**, 383–393 (2008).
112. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
113. Suh, M. R. *et al.* Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* **270**, 488–498 (2004).
114. Thomson, J. M. *et al.* Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
115. Wulczyn, F. G. *et al.* Post-transcriptional regulation of the *let-7* microRNA during neural cell specification. *FASEB J.* **21**, 415–426 (2007).
116. Newman, M. A., Thomson, J. M. & Hammond, S. M. Lin-28 interaction with the let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**, 1539–1549 (2008).
117. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science* **320**, 97–100 (2008).
118. Rybak, A. *et al.* A feedback loop comprising *lin-28* and *let-7* controls pre-*let-7* maturation during neural stem-cell commitment. *Nature Cell Biol.* **10**, 987–993 (2008).
119. Heo, I. *et al.* Lin28 mediates the terminal uridylation of let-7 precursor microRNA. *Mol. Cell* **32**, 276–284 (2008).
120. Balzer, E. & Moss, E. G. Localization of the developmental timing regulator Lin28 to mRNP complexes, P-bodies and stress granules. *RNA Biol.* **4**, 16–25 (2007).
121. Hwang, H. W., Wentzel, E. A. & Mendell, J. T. A hexanucleotide element directs microRNA nuclear import. *Science* **315**, 97–100 (2007).
122. Kennedy, S., Wang, D. & Ruvkun, G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* **427**, 645–649 (2004).
123. Ramachandran, V. & Chen, X. Degradation of microRNAs by a family of exoribonucleases in *Arabidopsis*. *Science* **321**, 1490–1492 (2008).
124. Yang, W. *et al.* Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nature Struct. Mol. Biol.* **13**, 13–21 (2006).
125. Kawahara, Y., Zinshteyn, B., Chendrimada, T. P., Shiekhattar, R. & Nishikura, K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer–TRBP complex. *EMBO Rep.* **8**, 763–769 (2007).
126. Kawahara, Y. *et al.* Frequency and fate of microRNA editing in human brain. *Nucleic Acids Res.* **36**, 5270–5280 (2008).
127. Obernosterer, G., Leuschner, P. J., Alenius, M. & Martinez, J. Post-transcriptional regulation of microRNA expression. *RNA* **12**, 1161–1167 (2006).
128. Lee, E. J. *et al.* Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA* **14**, 35–42 (2008).
129. Michael, M. Z., O'Connor, S. M., van Holst Pellekaan, N. G., Young, G. P. & James, R. J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* **1**, 882–891 (2003).
130. Kefas, B. *et al.* microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. *Cancer Res.* **68**, 3566–3572 (2008).
131. Martinez, N. J. *et al.* A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev.* **22**, 2535–2549 (2008).
132. Tokumaru, S., Suzuki, M., Yamada, H., Nagino, M. & Takahashi, T. *let-7* regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis* **29**, 2073–2077 (2008).
133. Forman, J. J., Legesse-Miller, A. & Coller, H. A. A search for conserved sequences in coding regions reveals that the *let-7* microRNA targets Dicer within its coding sequence. *Proc. Natl Acad. Sci. USA* **105**, 14879–14884 (2008).

134. Piskounova, E. *et al.* Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. *J. Biol. Chem.* **283**, 21310–21314 (2008).
135. Bracken, C. P. *et al.* A double-negative feedback loop between ZEB1–SIP1 and the microRNA-200 family regulates epithelial–mesenchymal transition. *Cancer Res.* **68**, 7846–7854 (2008).
136. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11**, 1017–1027 (2001).
137. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
This was the herald of studies for a novel class of small RNAs of 24–29 nt (rasiRNAs or piRNAs) that are specifically expressed in germ lines.
138. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
139. Cox, D. N., Chao, A. & Lin, H. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514 (2000).
140. Szakmary, A., Cox, D. N., Wang, Z. & Lin, H. Regulatory relationship among *piwi*, *pumilio*, and *bag-of-marbles* in *Drosophila* germline stem cell self-renewal and differentiation. *Curr. Biol.* **15**, 171–178 (2005).
141. Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* **33**, 2052–2059 (2005).
142. Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. Evidence for a *piwi*-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* **166**, 1313–1321 (2004).
143. Harris, A. N. & Macdonald, P. M. *aubergine* encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development* **128**, 2823–2832 (2001).
144. Vagin, V. V. *et al.* The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol.* **1**, 54–58 (2004).
145. Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* **20**, 2214–2222 (2006).
146. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
147. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
The ping-pong pathway for piRNA biogenesis was proposed on the basis of these *D. melanogaster* studies.
148. Nishida, K. M. *et al.* Gene silencing mechanisms mediated by Aubergine–piRNA complexes in *Drosophila* male gonad. *RNA* **13**, 1911–1922 (2007).
Shows that *Suppressor of Stellate* piRNAs are directly associated with AUB in *D. melanogaster* testes and that the complex can slice the *Stellate* transcripts.
149. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).
150. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
151. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709–1714 (2006).
152. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **20**, 1732–1743 (2006).
153. Houwing, S. *et al.* A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell* **129**, 69–82 (2007).
154. Carmell, M. A. *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* **12**, 503–514 (2007).
155. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744–747 (2007).
156. Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917 (2008).
157. Zamore, P. D. RNA silencing: genomic defence with a slice of pi. *Nature* **446**, 864–865 (2007).
158. Pane, A., Wehr, K. & Schupbach, T. *zucchini* and *squash* encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev. Cell* **12**, 851–862 (2007).
159. Lim, A. K. & Kai, T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **104**, 6714–6719 (2007).
160. Findley, S. D., Tamanaha, M., Clegg, N. J. & Ruohola-Baker, H. *Maelstrom*, a *Drosophila spindle*-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development* **130**, 859–871 (2003).
161. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).
162. Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**, 1387–1392 (2008).
163. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367 (2006).
164. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* **31**, 785–799 (2008).
165. Saito, K., Sakaguchi, Y., Suzuki, T., Siomi, H. & Siomi, M. C. Pimet, the *Drosophila* homolog of HEN1, mediates 2′-*O*-methylation of Piwi-interacting RNAs at their 3′ ends. *Genes Dev.* **21**, 1603–1608 (2007).
166. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272 (2007).
167. Kirino, Y. & Mourelatos, Z. Mouse Piwi-interacting RNAs are 2′-*O*-methylated at their 3′ termini. *Nature Struct. Mol. Biol.* **14**, 347–348 (2007).
168. Ohara, T., Sakaguchi, Y., Suzuki, T., Ueda, H. & Miyauchi, K. The 3′ termini of mouse Piwi-interacting RNAs are 2′-*O*-methylated. *Nature Struct. Mol. Biol.* **14**, 349–350 (2007).
169. Yu, B. *et al.* Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
170. Ghildiyal, M. *et al.* Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320**, 1077–1081 (2008).
171. Chung, W. J., Okamura, K., Martin, R. & Lai, E. C. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* **18**, 795–802 (2008).
172. Okamura, K., Balla, S., Martin, R., Liu, N. & Lai, E. C. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*. *Nature Struct. Mol. Biol.* **15**, 581–590 (2008).
173. van Rij, R. P. *et al.* The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* **20**, 2985–2995 (2006).
174. Wang, X. H. *et al.* RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312**, 452–454 (2006).
175. Nishikura, K. Editor meets silencer: crosstalk between RNA editing and RNA interference. *Nature Rev. Mol. Cell Biol.* **7**, 919–931 (2006).
176. Fukuda, T. *et al.* DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nature Cell Biol.* **9**, 604–611 (2007).
177. Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. & Shimizu, N. Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Exp. Cell Res.* **313**, 4196–4207 (2007).
178. Yu, B. *et al.* The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. *Proc. Natl Acad. Sci. USA* **105**, 10073–10078 (2008).
179. Neumuller, R. A. *et al.* Mei-P26 regulates microRNAs and cell growth in the *Drosophila* ovarian stem cell lineage. *Nature* **454**, 241–245 (2008).
180. Krol, J. *et al.* Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol. Cell* **25**, 575–586 (2007).
181. Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**, 318–322 (2004).
182. Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nature Struct. Mol. Biol.* **11**, 576–577 (2004).
183. Wang, Y. *et al.* Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* **456**, 921–926 (2008).
184. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213 (2008).
185. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**, 1434–1437 (2004).
186. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex. *Nature* **434**, 663–666 (2005).
187. Ma, J. B. *et al.* Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666–670 (2005).
188. Brower-Toland, B. *et al.* *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev.* **21**, 2300–2311 (2007).
189. Blaszczyk, J. *et al.* Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure* **9**, 1225–1236 (2001).
190. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).
191. Yeom, K. H., Lee, Y., Han, J., Suh, M. R. & Kim, V. N. Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res.* **34**, 4622–4629 (2006).
192. Sohn, S. Y. *et al.* Crystal structure of human DGCR8 core. *Nature Struct. Mol. Biol.* **14**, 847–853 (2007).
193. Faller, M., Matsunaga, M., Yin, S., Loo, J. A. & Guo, F. Heme is involved in microRNA processing. *Nature Struct. Mol. Biol.* **14**, 23–29 (2007).
194. Song, J. J. *et al.* The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Mol. Biol.* **10**, 1026–1032 (2003).
195. Yan, K. S. *et al.* Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).
196. Macrae, I. J. *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195–198 (2006).
197. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro- RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **99**, 15524–15529 (2002).
198. Tam, W. Identification and characterization of human *BIC*, a gene on chromosome 21 that encodes a noncoding RNA. *Gene* **274**, 157–167 (2001).

Acknowledgements

We are grateful to H. Siomi, T. Watanabe, Y. Watanabe and S. Kuramochi-Miyagawa for helpful discussions and comments. This work was supported by the Creative Research Initiatives Program (V.N.K.), the BK21 Research Funds from the Ministry of Education (J.H.) from Science and Technology of Korea, New Energy and Industrial Technology Development Organization (NEDO) grants (M.C.S.) and Core Research for Evolutionary Science and Technology (CREST) from the Japan Science and Technology Agency (JST) (M.C.S.). M.C.S. is Associate Professor of Global COE for Human Metabolomics Systems Biology by the Ministry of Education, Culture, Science and Technology, Japan.

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
flamenco
The miRNA Registry: http://microrna.sanger.ac.uk/sequences
let-7 | lin-4 | miR-1 | miR-7 | miR-29b | miR-31 | miR-34 | miR-105 | miR-128 | miR-133 | miR-138 | miR-142 | miR-143 | miR-145 | miR-151 | miR-200 | miR-203 | pri-miR-18
UniProtKB: http://www.uniprot.org
AGO1 | AGO2 | AGO3 | AGO4 | ALG-1 | AUB | DGCR8 | Dicer 1 | Dicer 2 | Drosha | ERI1 | EXP5 | LIN28 | LOQS | MILI | MIWI | MIWI2 | MYC | MYOD1 | myogenin | p53 | PACT | PIWI | R2D2 | RDE-1 | TRBP | XRN2 | ZEB1 | ZEB2 | Zili | Ziwi

FURTHER INFORMATION

V. Narry Kim's homepage: http://www.narrykim.org
Mikiko C. Siomi's homepage: http://web.sc.itc.keio.ac.jp/dmb/sindex.html
miRNA database: http://microrna.sanger.ac.uk

ALL LINKS ARE ACTIVE IN THE ONLINE PDF