MicroRNA BIOGENESIS: COORDINATED CROPPING AND DICING

V. Narry Kim

Abstract | The recent discovery of microRNAs (miRNAs) took many by surprise because of their unorthodox features and widespread functions. These tiny, ~22-nucleotide, RNAs control several pathways including developmental timing, haematopoiesis, organogenesis, apoptosis, cell proliferation and possibly even tumorigenesis. Among the most pressing questions regarding this unusual class of regulatory miRNA-encoding genes is how miRNAs are produced in cells and how the genes themselves are controlled by various regulatory networks.

MicroRNAs (miRNAs) are single-stranded RNAs (ssRNAs) of 19–25 nucleotides in length that are generated from endogenous hairpin-shaped transcripts[1–3]. miRNAs function as guide molecules in post-transcriptional gene silencing by base pairing with target mRNAs, which leads to mRNA cleavage or translational repression. With >200 members per species in higher eukaryotes, miRNAs are one of the largest gene families, accounting for ~1% of the genome[2]. Recent studies have revealed that miRNAs have key roles in diverse regulatory pathways, including control of developmental timing, haematopoietic cell differentiation, apoptosis, cell proliferation and organ development (BOX 1). miRNAs and their targets seem to form complex regulatory networks. For example, a single miRNA can bind to and regulate many different mRNA targets and, conversely, several different miRNAs can bind to and cooperatively control a single mRNA target[4]. Recent work by the Bartel and Burge laboratories predicted that over one third of all human genes are targeted by miRNAs[5]. Consequently, the unique combination of miRNAs that are expressed in each cell type might affect or 'dampen' the utilization of thousands of mRNAs[6]. Although the complexity of this regulatory circuitry is currently overwhelming, one of the first key steps towards dissecting the network would be to understand how miRNA genes themselves are regulated. In this review, I will discuss recent findings on the gene structure of the miRNA gene family, and on miRNA biogenesis, which comprises miRNA transcription and the coordinated processes that are involved in the subsequent processing and maturation of miRNAs.

What is a microRNA?

According to the current convention, a miRNA is defined as a ssRNA of ~22 nucleotides in length, which is generated by the RNase-III-type enzyme Dicer from an endogenous transcript that contains a local hairpin structure[1]. In practice, when a small RNA is discovered by cDNA cloning, it must meet the following criteria to be classified as a miRNA. First, its expression should be confirmed by hybridization to a size-fractionated RNA sample, usually by northern blotting. Other detection methods are available, including PCR after reverse transcription of RNA (RT-PCR), primer extension analysis, RNase protection assay and microarray. However, northern blotting is still the method of choice for the confirmation of miRNAs, because the blot usually shows both the mature form (a ~22-nucleotide band) and the hairpin precursor (a ~70-nucleotide band). Second, the small RNA sequence should be present in one arm of the hairpin precursor, which lacks large internal loops or bulges. The precursors are usually ~60–80 nucleotides in animals, but the lengths are more variable in plants. Third, the small RNA sequences should be phylogenetically conserved. The sequence conservation should also be seen in the precursor hairpin, usually to a lesser extent

Department of Biological Sciences and Institute of Molecular Biology and Genetics, Seoul National University, Seoul, 151-742, Korea.
e-mail: narrykim@snu.ac.kr
doi:10.1038/nrm1644
Published online 15 April 2005

Box 1 | MicroRNA function

MicroRNAs (miRNAs) have important roles in diverse regulatory pathways[2,100]. The paradigm for the function of miRNAs has been provided by nematode *lin-4* and *let-7* RNAs. They function as post-transcriptional repressors of their target genes when bound to the specific sites in the 3′ untranslated region (UTR) of the target mRNA[19,101,102]. The level of target mRNA does not change significantly, which indicates that the inhibition occurs at the level of translation, although the mechanism of inhibition remains obscure. To date, only a small number of miRNAs are known for their biological functions (see table). For example, *bantam* RNA from *Drosophila melanogaster* suppresses apoptosis and stimulates cell proliferation[20]. Being expressed in a temporal and tissue-specific manner, *bantam* RNA regulates tissue formation during development. Another nematode miRNA, *lsy-6* RNA, was identified in a genetic screen for left–right asymmetry of neuronal chemoreceptor expression[103,104]. *Lsy-6* RNA targets *cog-1*, which encodes a transcription factor. In mammals, *miR-181* is involved in the control of haematopoiesis through as-yet-unknown target(s)[105]. More recently, *miR-196* miRNAs were shown to repress the expression of the *HoxB8* gene, which encodes a transcription factor that is important in developmental regulation[106]. miR-196 family RNAs are the first examples of animal miRNAs that cause target mRNA cleavage rather than translational repression.

Plant miRNAs generally show a higher degree of complementarity to the target mRNAs and this results in target cleavage[2]. Interestingly, most of the known targets of plant miRNAs are transcription factors, particularly those involved in developmental regulation and cell differentiation.

miRNA	Function	Known targets	Species	Refs
lin-4	Developmental timing	*lin-14*, *lin-28*	*Ce*	19,101
let-7	Developmental timing	*lin-41*, *hbl-1*	*Ce*	102
lsy-6	Neuronal patterning	*cog-1*	*Ce*	104
miR-273	Neuronal patterning	*die-1*	*Ce*	103
bantam	Cell death, proliferation	*hid*	*Dm*	20
miR-14	Cell death, fat storage	Caspase?	*Dm*	107
miR-181	Haematopoiesis	?	*Mm*	105
miR-196	Development?	*HoxB8*, *HoxC8*, *HoxD8*, *HoxA7*	*Mm*	106
miR-143	Adipocyte differentiation	?	*Hs*	108
miR-375	Insulin secretion	Myotrophin	*Mm*	128

Ce, *Caenorhabditis elegans*; *Dm*, *Drosophila melanogaster*; *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*.

than in the mature miRNA segment. Last, the evidence can be strengthened if the precursor accumulates in the presence of reduced Dicer function. However, this criterion is not used as often as the others, because of the technical difficulties that are involved in depleting Dicer in certain cell types.

None of the above criteria on its own is sufficient for a candidate gene to be annotated as a novel miRNA. Typically, the first criterion (expression) plus the second criterion (structure), or the first criterion (expression) plus the third criterion (conservation), are regarded as adequate. When the expression is too low for detection, a small RNA that meets the third criterion alone (conservation), together with cDNA cloning, can be annotated as a miRNA. However, if the small RNA is found by methods other than cDNA cloning, expression of the small RNA must be demonstrated along with the precursor structure and conservation. If a small RNA is predicted but cannot be detected, the phylogenetically conserved hairpin precursor should instead be shown to accumulate in the absence of Dicer (the third and fourth criteria). Close homologues in other species can be annotated as miRNA orthologues without experimental validation, provided that they fulfil the third (conserved precursor) criterion. Small RNAs that do not meet these requirements can be classified as either SMALL INTERFERING RNAS (siRNAs) or other provisionary classes[7] (BOX 2).

At the time of writing, The miRNA Registry (see online links box) contained 116 *Caenorhabditis elegans* miRNAs, 78 *Drosophila melanogaster* miRNAs, 30 *Danio rerio* miRNAs, 121 *Gallus gallus* miRNAs, 222 *Homo sapiens* miRNAs, 112 *Arabidopsis thaliana* miRNAs and 5 Epstein–Barr virus miRNAs. The list is still expanding as a result of both intensive cloning and computational prediction approaches.

Gene structure and microRNA gene transcription

Early annotation of the genomic position of miRNAs indicated that most miRNAs are located in intergenic regions (>1 kb away from annotated/predicted genes), although a sizeable minority was found in the intronic regions of known genes in the sense or antisense orientation[8,9]. This led to the postulation that most miRNA genes are transcribed as autonomous transcription units. Another interesting observation was that ~50% of known miRNAs are found in close proximity to other miRNAs[8–10], which raised the possibility that these clustered miRNAs might be transcribed from a single POLYCISTRONIC TRANSCRIPTION UNIT. A detailed analysis of miRNA gene expression showed that miRNA genes can be transcribed from their own promoters[11,12], and that the clustered miRNAs are generated as polycistronic primary transcripts (pri-miRNAs)[13].

SMALL INTERFERING RNA (siRNA; also known as short interfering RNA). A small (21–24 nucleotide), non-coding RNA that is generated from long double-stranded RNA. siRNAs function as guide molecules in small-RNA-mediated gene silencing.

POLYCISTRONIC TRANSCRIPTION UNIT
A transcript that includes regions representing multiple, non-overlapping gene products.

CAP STRUCTURE
A structure, which consists of m^7GpppN (where m^7G represents 7-methylguanylate, p represents a phosphate group and N represents any base), that is located at the 5′ end of eukaryotic mRNAs.

POLY(A) TAIL
A homopolymeric stretch of usually 25–200 adenine nucleotides that is present at the 3′ end of most eukaryotic mRNAs.

SMALL NUCLEOLAR RNA
(snoRNAs). A small RNA molecule that functions in ribosome biogenesis in the nucleolus by guiding the assembly of macromolecular complexes on the target RNA to allow site-specific modifications or processing reactions to occur.

TRANS-SPLICING
Intermolecular splicing that occurs in trypanosomes and worms where a short sequence (SL RNA) is linked to the 5′ end of many pre-mRNAs.

Transcription of miRNA genes is mediated by RNA polymerase II (pol II)[11,12], although the possibility that a small number of miRNA genes might be transcribed by other RNA polymerases cannot be excluded. Pol III was initially believed to mediate miRNA transcription because it transcribes most small RNAs, such as tRNAs. However, pri-miRNAs are sometimes several kilobases long and contain stretches of more than four uracils, which would have terminated transcription by pol III (REF. 13). Several poly(A)-containing transcripts comprising miRNA sequences as well as pieces of adjacent mRNAs have been identified by expressed sequence tag (EST) analyses and gene structure analyses[14–18]. The expression profiles of miRNAs also indicated that miRNAs are under elaborate control during development and in various tissues, similar to other genes that are transcribed by pol II (REFS 19–33). In addition, a fully functional mature miRNA can be generated from a plasmid that contains a pri-miRNA segment under the control of a heterologous pol II promoter[12,34].

Three lines of more direct evidence were recently reported. First, pri-miRNAs were shown to contain both CAP STRUCTURES and POLY(A) TAILS[11,12]. Second, miRNA transcription activity is sensitive to α-amanitin at a concentration that specifically inhibits pol II, but not pol I or pol III (REF. 11). Last, the physical association of pol II with the promoter of *miR-23a~27a~24-2* was demonstrated by chromatin immunoprecipitation analyses[11]. Pol II-dependent transcription probably confers several advantages. For example, miRNA gene transcription can be controlled by various pol-II-associated regulatory factors, so that a specific set of miRNAs can be expressed during development as well as under specific conditions and in certain cell types. In addition, the expression of miRNA and protein-coding genes might be coordinated, especially when an miRNA and a protein-coding region both reside in a single transcript.

A recent analysis of miRNA gene locations relative to known transcription units brought to light new issues regarding miRNA biogenesis[15]. By combining up-to-date genome assemblies and transcription unit databases, Bradley and colleagues showed that ~70% of mammalian miRNA genes (161 out of 232) are located in defined transcription units (FIG. 1). Interestingly, 117 miRNA genes were found in the introns in the sense orientation, which is more than previously expected. Of these 117 intronic miRNAs, 90 miRNAs are in the introns of protein-coding genes, whereas 27 miRNAs are in the introns of non-coding RNAs (ncRNAs). A further 30 overlap with the exons of ncRNAs. And 14 miRNAs overlap with either an exon or an intron ('mixed') depending on the alternative splicing pattern. So, miRNA genes can be grouped on the basis of their genomic locations (FIG. 1): first, exonic miRNA in non-coding transcription units; second, intronic miRNA in non-coding transcription units; and third, intronic miRNA in protein-coding transcription units. Mixed miRNA genes can be assigned to one of the above groups depending on the given splicing pattern.

An obvious question is whether the intronic or exonic location affects miRNA biogenesis or, more precisely, is there any connection between splicing and pri-miRNA processing? It is now generally believed that splicing precedes pri-miRNA processing and that the resulting intron lariats get processed to release pre-miRNAs. This would be analogous to SMALL NUCLEOLAR RNAS (snoRNAs) that are also generated from intron lariats. In *C. elegans*, *TRANS*-SPLICING upstream of *let-7* is required for pri-miRNA processing[18]. *Trans*-splicing reaction replaces the sequences upstream of the *let-7* stem-loop with spliced leader 1 (SL1) sequences, altering the secondary structure around the stem-loop, so that the spliced transcript can be more efficiently processed. However, it is unlikely that this example represents the general mode of miRNA maturation, as miRNA sequences are usually located far from the splice junctions.

Another interesting incident is when a single transcript contains both a protein-coding region and miRNA sequences (miRNAs in protein-coding transcription unit)[14,15]. It was recently shown that a single transcription unit that includes both luciferase cDNA and *miR-21* precursor sequences can produce luciferase protein as well as *miR-21* RNA[12]. However, it remains unclear whether a single primary transcript molecule can generate miRNA as well as protein, or whether each primary transcript has to choose between two synthesis pathways — either the miRNA pathway or the mRNA pathway.

The 'host' transcript and miRNAs usually portray similar expression profiles, which indicates that these miRNAs are transcribed as part of the long host transcription units[8,9,15]. Currently, only a few miRNA promoters have been identified experimentally[11,12,16–18,33].

Box 2 | Different types of small RNAs

Apart from microRNAs (miRNAs), several other groups of small RNAs have been described[7]: small interfering RNAs (siRNAs), tiny non-coding RNAs (tncRNAs) and a small modulatory RNA (smRNA). siRNAs are defined as 21–28-nucleotide RNAs, which are produced by Dicer from long double-stranded RNAs (dsRNAs). Small RNAs used in RNA-interference-mediated experiments are traditionally referred to as siRNAs, regardless of their origin. Endogenous siRNAs have also been discovered in various organisms. Endogenous siRNAs can be further categorized into at least three distinct sub-classes: *trans*-acting siRNAs (tasiRNAs), repeat-associated siRNAs (rasiRNAs) and small-scan RNAs (scnRNAs). tasiRNAs found in plants are ~21-nucleotide RNAs that are encoded in intergenic regions, and they act on mRNAs *in trans* to induce mRNA cleavage[88,109,110]. Some endogenous siRNAs match repetitive elements in the sense or antisense orientation and are therefore known as rasiRNAs[22,111–120]. Evidence from *Schizosaccharomyces pombe* and *Arabidopsis thaliana* indicates that rasiRNAs might function in transcriptional gene silencing through methylation of histones and/or DNA[88,118–121]. siRNAs of ~28 nucleotides found in *Tetrahymena thermophila* are known as scnRNA because they might scan for regions in the DNA for genome rearrangement[122–124]. The Argonaute family protein TWI1 mediates histone methylation, which is, in turn, required for DNA elimination[122].

There are additional small RNAs of which the biogenesis and function remain less clear, and which hinders their classification. tncRNAs described in nematode worms are similar to tasiRNAs[96], although their origin from the long dsRNA precursor needs to be verified. A smRNA of ~20 nucleotides that has been cloned from neural stem cells modulates transcription possibly by interacting with a transcription factor[125]. The biogenesis of smRNAs is as yet unknown.

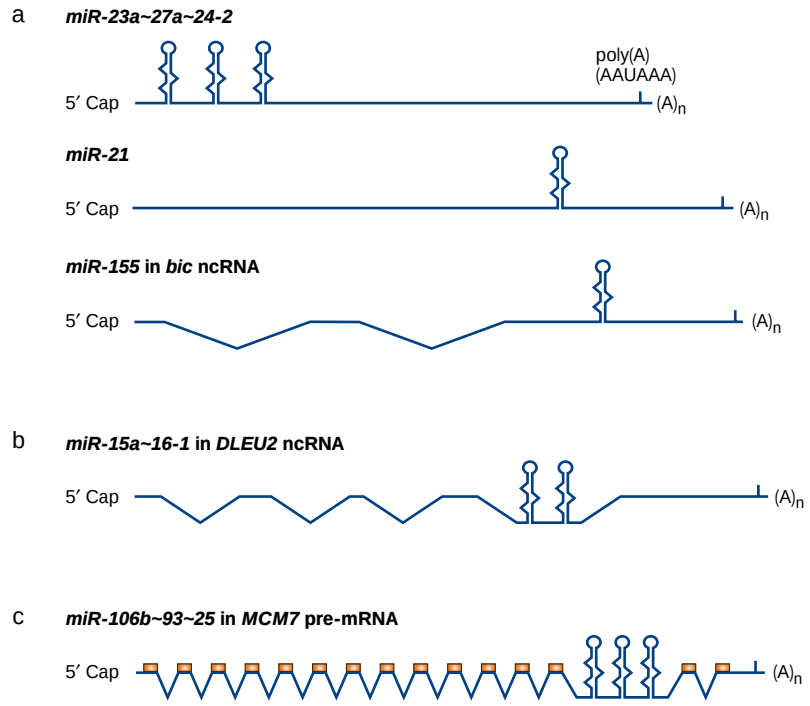

Figure 1 | **The structure of five pri-miRNAs.** Primary transcripts that encode miRNAs, pri-miRNAs, contain 5′ cap structures as well as 3′ poly(A) tails. miRNAs can be categorized into three groups according to their genomic locations relative to their positions in an exon or intron. **a** | Exonic miRNAs in non-coding transcripts such as an *miR-23a~27a~24-2* cluster, *miR-21* and *miR-155*. *miR-155* was found in a previously defined non-coding RNA (ncRNA) gene, *bic*[17]. **b** | Intronic miRNAs in non-coding transcripts. For example, an *miR-15a~16-1* cluster was found in the fourth intron of a previously defined non-coding RNA gene, *DLEU2* (REF. 126). **c** | Intronic miRNAs in protein-coding transcripts. For example, an *miR-106b~93~25* cluster is embedded in the thirteenth intron of DNA replication licensing factor *MCM7* transcript (variant 1, which encodes isoform 1). The mouse *miR-106b~93~25* homologue is also found in the thirteenth intron of the mouse *MCM7* homologue gene[15]. The hairpins indicate the miRNA stem-loops. Orange boxes indicate the protein-coding region. This figure is not to scale.

Further analysis of the promoters of various miRNA genes will be necessary to reveal the general features of miRNA promoters. Information on miRNA-specific element(s) would be invaluable both for the prediction of miRNA genes and for studies on miRNA biogenesis. Interestingly, the sequence alignment of nematode miRNA loci revealed a highly conserved sequence motif that is present upstream of almost all nematode miRNA genes[35], although the importance of this motif in transcriptional control remains to be determined. No common sequence motifs have so far been identified in the miRNA gene loci of other organisms[35].

MicroRNA maturation

The current model for miRNA maturation was formulated on the basis of two simple observations[13]. First, miRNAs are transcribed as long primary transcripts that are first trimmed into the hairpin intermediates (pre-miRNAs) and subsequently cleaved into mature miRNAs. Second, the catalytic activities for the first and the second processing are compartmentalized into the nucleus and the cytoplasm, respectively. So the nuclear export of pre-miRNA is necessary for the cytoplasmic processing to occur.

WW DOMAIN
A protein domain that binds to proline-rich regions.

Nuclear processing by Drosha. Transcription of miRNA genes yields primary transcripts, pri-miRNAs, that are usually several kilobases long and that contain a local hairpin structure (FIG. 1). The stem-loop structure is cleaved by the nuclear RNase III Drosha to release the precursor of miRNA (pre-miRNA)[36] (FIG. 2). The remnants (the flanking fragments) are thought to be degraded in the nucleus, although the ~1.7-kb 3′ flanking fragment of *miR-23a~27a~24-2* has been cloned multiple times in EST studies and is currently annotated as an mRNA[11]. Accordingly, it remains to be seen whether the 5′ and 3′ fragments that surround the stem-loop have their own functions.

Drosha is a large protein of ~160 kDa, and is conserved in animals[37–39] (FIG. 3). It contains two tandem RNase III domains (RIIIDs) and a double-stranded RNA-binding domain (dsRBD) that are crucial for catalysis[40]. The central region of the protein, adjacent to the RIIIDs, is also essential for pri-miRNA processing[40]. Drosha forms a large complex of ~500 kDa in *D. melanogaster*[41] or ~650 kDa in humans[40,42]. In this complex, which is known as the Microprocessor complex[41,42], Drosha interacts with its cofactor, the DiGeorge syndrome critical region gene 8 (DGCR8) protein in humans (also known as Pasha in *D. melanogaster* and *C. elegans*)[40–43]. DGCR8/Pasha is a ~120 kDa protein that contains two dsRBDs. It also contains a putative WW DOMAIN, which is known to be an interaction module for specific proline-rich sequences. It remains to be determined if this domain interacts with the proline-rich region of Drosha. Although the biochemical role of DGCR8/Pasha is currently unclear, it is believed to assist Drosha in substrate recognition[40–43].

There are hundreds of different pri-miRNAs in animal cells, but they do not seem to share any common sequence motifs. So how does the Drosha complex recognize its substrates? This question cannot be fully answered yet, but mutagenesis analyses indicate that the tertiary structure of pri-miRNAs is the primary determinant for substrate specificity[36,44,45]. Both the double-stranded stem structure around the cleavage site[36,44] and the large terminal loop (consisting of >10 nucleotides)[44,45] are vital. Interestingly, it seems that the Drosha complex can measure the length of the stem, because the cleavage site is located approximately two helical turns (~22 nucleotides) from the terminal loop[45] (FIG. 4). To comprehend how the Drosha complex measures the length, the biochemical roles of the individual domains of Drosha and DGCR8 need to be delineated.

It remains unknown as to whether Drosha participates in the processing of other types of RNA. Drosha was initially implicated in pre-ribosomal RNA processing[39], although this effect might have been indirect. As a myriad of double-stranded RNAs (dsRNAs) have been discovered in recent years, it would be intriguing to explore the role of Drosha and other dsRBD-containing proteins in ancillary dsRNA pathways.

Nuclear export by exportin-5. Following nuclear processing by Drosha, pre-miRNAs are exported to the cytoplasm. Once there, they are subjected to the second

processing step by Dicer (another RNase III enzyme) to generate the final ~22-nucleotide product (see below). Owing to compartmentalization of the two processing events, nuclear export of pre-miRNAs is a crucial step in miRNA biogenesis[46,47]. Nuclear transport occurs through nuclear pore complexes, which are large proteinaceous channels embedded in the nuclear membrane[48]. Members of the nuclear export receptor family bind cooperatively to a cargo as well as to the GTP-bound form of the cofactor Ran in the nucleus. Following export, hydrolysis of GTP to GDP results in release of the cargo from the export complex.

Export of pre-miRNA is mediated by one of the nuclear transport receptors, exportin-5 (REFS 49–51). When the cells were depleted of exportin-5, the pre-miRNA level and the mature miRNA level were reduced in the cytoplasm[49,50]. Notably, pre-miRNA does not accumulate in the nucleus subsequent to the depletion of exportin-5. This indicates that pre-miRNA might be relatively unstable and also that pre-miRNA might be stabilized through its interaction with exportin-5 (REF. 50). Exportin-5 was originally known as a minor export factor for tRNAs, because it can transport tRNAs when the primary export factor, exportin-t, is depleted or overloaded[52,53]. Considering that the affinity of exportin-5 to pre-miRNA is much higher than to tRNA, and that miRNAs are as abundant as 50,000 copies per cell[54], pre-miRNAs are probably the main cargo for exportin-5.

Exportin-5 can also export adenoviral RNA VA1, a 160-nucleotide ncRNA[55]. Analysis of *cis*-acting elements for nuclear export in VA1 revealed a structural motif known as the 'minihelix motif', which consists of a >14-bp stem and a 3–8-nucleotide 3′ overhang[55]. A similar structural motif can be found in pre-miRNA stem-loops, which typically comprise a stem of ~22 bp, a terminal loop and a 3′ overhang of ~2 nucleotides[36,49,56,57]. By introducing mutations in the *pre-miR-30a*, Cullen and colleagues confirmed that an RNA stem of >16 bp and a short 3′ overhang are significant structural requirements for pre-miRNA export[57].

Cytoplasmic processing by Dicer. Following their export from the nucleus, pre-miRNAs are subsequently processed into ~22-nucleotide miRNA duplexes by the cytoplasmic RNase III Dicer[58–62]. Because Dicer was originally found to function in generating siRNAs, which are similar in size (21–25 nucleotides) to miRNAs[58,63], it was predicted that Dicer also functions in the processing of ~70-nucleotide stem-loop RNAs into mature miRNAs. Indeed, it was later proven that immunoprecipitated Dicer generates ~22-nucleotide miRNAs from *in vitro* synthesized ~70-nucleotide *let-7* stem-loop RNAs[61,62]. Moreover, when the human homologue and *C. elegans* homologue of Dicer were knocked down or knocked out, the ~70-nucleotide pre-miRNAs accumulated, whereas the ~22-nucleotide mature miRNAs diminished[59–62].

Dicer is a highly conserved protein that is found in almost all eukaryotic organisms such as *Schizosaccharomyces pombe*, plants and animals. Some organisms contain multiple Dicer homologues, in

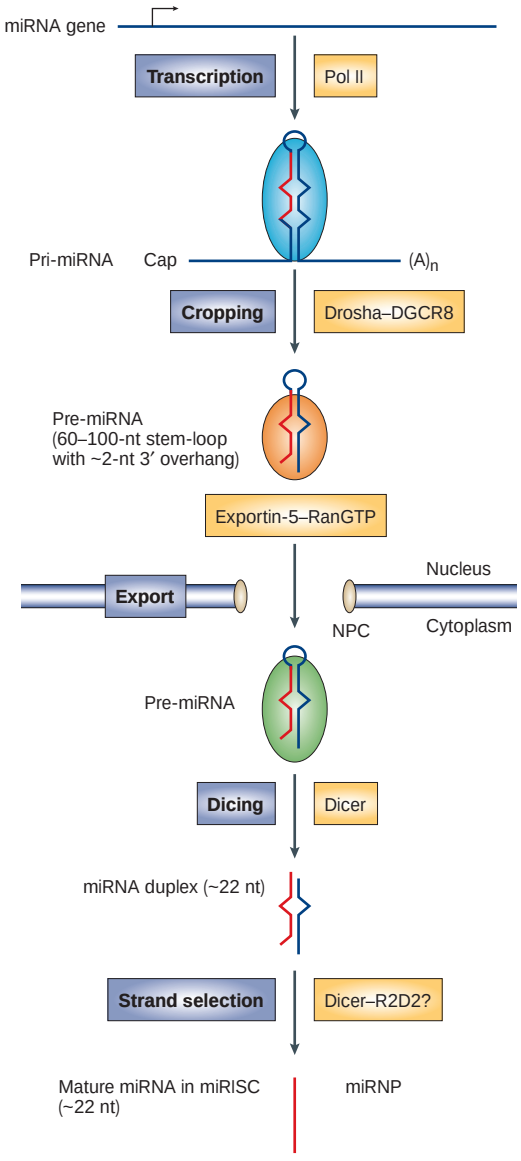

Figure 2 | **Model for microRNA biogenesis.** MicroRNA (miRNA) genes are transcribed by RNA polymerase II (pol II) to generate the primary transcripts (pri-miRNAs). The initiation step ('cropping') is mediated by the Drosha–DGCR8 complex (also known as the Microprocessor complex). Drosha and DGCR8 are both located mainly in the nucleus. The product of this nuclear processing step is a ~70-nucleotide (nt) pre-miRNA, which possesses a short stem plus a ~2-nucleotide 3′ overhang. This structure might serve as a signature motif that is recognized by the nuclear export factor exportin-5. Pre-miRNA constitutes a transport complex together with exportin-5 and its cofactor Ran (the GTP-bound form). Following export, the cytoplasmic RNase III Dicer participates in the second processing step ('dicing') to produce miRNA duplexes. The duplex is separated and usually one strand is selected as the mature miRNA, whereas the other strand is degraded. In *Drosophila melanogaster*, R2D2 forms a heterodimeric complex with Dicer and binds to one end of the small interfering RNA duplex. It thereby selects one strand of the duplex. It is not known if miRNAs use the same machinery for strand selection (see question mark). It is also unclear as to whether an R2D2 homologue functions in other animals apart from *D. melanogaster*. miRISC, miRNA-containing RNA-induced silencing complex; NPC, nuclear pore complex.

DEAD-BOX RNA HELICASE DOMAIN
An evolutionarily conserved domain in a family of enzymes that use ATP hydrolysis to unwind RNA duplexes. The domain is named after the DEAD (Asp-Glu-Ala-Asp) motif.

PAZ DOMAIN
A conserved RNA-binding domain found in members of the Dicer and Argonaute protein families, and that preferentially interacts with the 3′ end of RNA.

ARGONAUTE PROTEINS
(also known as PPD proteins). A family of proteins that are characterized by the presence of two homology domains, PAZ and PIWI. These proteins are essential for diverse small-RNA pathways.

which different Dicer isotypes are often assigned to take on distinct roles. For example, *D. melanogaster* Dicer-1 is required for pre-miRNA cleavage, whereas Dicer-2 is needed for siRNA generation[64]. Dicer homologues are multi-domain proteins of ~200 kDa (FIG. 3). Apart from two RIIIDs and a dsRBD, Dicer has a long N-terminal segment that contains a DEAD-BOX RNA HELICASE DOMAIN, as well as a DUF283 domain and a PAZ DOMAIN. The PAZ domain is also found in a group of highly conserved proteins known as ARGONAUTE PROTEINS. Structural and biochemical studies of the PAZ domain from *D. melanogaster* AGO1 and AGO2 indicate that the PAZ domain binds to the 3′ protruding end of small RNAs[65–68]. The roles of the other domains in Dicer remain unknown.

Dicer associates with several other proteins including RDE-4 in *C. elegans*[69], R2D2 (REF. 70) and FMR1 (REFS 70–73) in *D. melanogaster*[70], and Argonaute family proteins in various organisms[74,75]. These Dicer-interacting proteins do not seem to be required for the cleavage reaction itself as purified human Dicer and *D. melanogaster* Dicer-2 can catalyse the cleavage reaction[70,76,77]. Instead, Dicer-interacting proteins have various roles in miRNA stability and effector complex formation and action. For example, one of the Argonaute family proteins, human AGO2, was recently shown to function as the 'slicer' enzyme that cleaves target mRNA[78–80].

Mature miRNAs are incorporated into effector complexes that are known as 'miRNP' (miRNA-containing ribonucleoprotein complex), 'mirgonaute' or 'miRISC' (miRNA-containing RNA-induced silencing complex). On the other hand, the effector complex that contains siRNA is known as 'RISC', 'sirgonaute' or 'siRISC'.

The cleavage products (~22-nucleotide miRNA duplexes) do not persist in the cell for long. Usually, one strand of this short-lived duplex disappears, whereas the other strand remains as a mature miRNA. Studies on siRNA duplexes indicate that the relative thermodynamic stability of the two ends of the duplex determines which strand is to be selected[81,82]. The strand with relatively unstable base pairs at the 5′ end typically remains (for example, G:U pair versus G:C pair)[81,82]. The same rule is thought to be applicable to miRNA.

The mechanistic details of strand selection have been delineated in *D. melanogaster*[70,83]. R2D2, which has two dsRBDs, was shown to sense the differences in thermodynamic inequalities. R2D2 forms a stable heterodimeric complex with Dicer-2 and binds to the more stable end of the siRNA duplex, thereby orienting the complex on the siRNA duplex. It is not known how the unselected strand is removed and degraded. It also remains to be determined if the same machinery acts on the miRNA duplex and how conserved this machinery is.

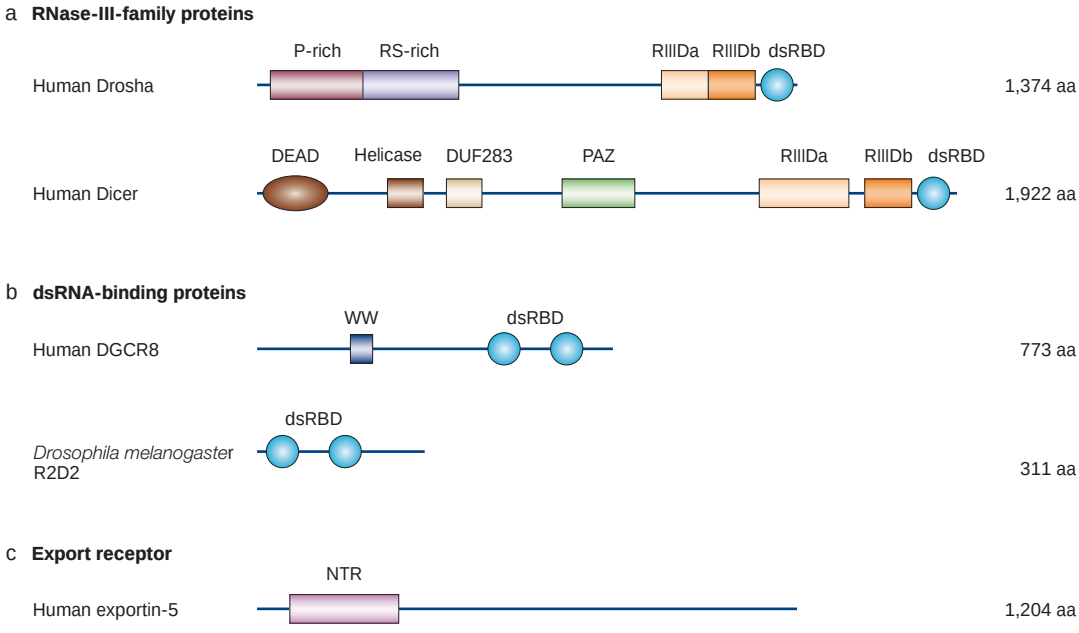

Figure 3 | **Domain organization of microRNA biogenesis factors. a** | The RNase III domain (RIIID) is the catalytic domain that is responsible for the endonucleolytic reaction of RNase III enzymes such as Dicer and Drosha. The RIIIDs (shown as RIIIDa and RIIIDb) are well conserved motifs found in RNase-III-type proteins of eubacterial, archaeal and eukaryotic origin. The double-stranded RNA-binding domain (dsRBD) is also a well conserved motif in many double-stranded RNA (dsRNA)-binding proteins of diverse functions, including Drosha and Dicer. The biological significance of the proline-rich (P-rich) region is unknown. The RS-rich region is abundant in arginine and serine residues. The function of this region is also unclear, although the C terminus of this region was shown to be important for the activity of Drosha[40]. The PAZ domain binds to the 3′ end of small RNAs[65–68]. The DEAD-box RNA helicase domain is typical of enzymes that hydrolyse ATP and unwind an RNA duplex. The DUF283 domain has no known function. **b** | The WW motif is known as a protein interaction module that binds to the P-rich domain, although the role of this domain in DGCR8 remains unclear. **c** | The nuclear transport receptor (NTR) domain is found in many Ran-dependent nuclear transport factors[48]. aa, amino acids.

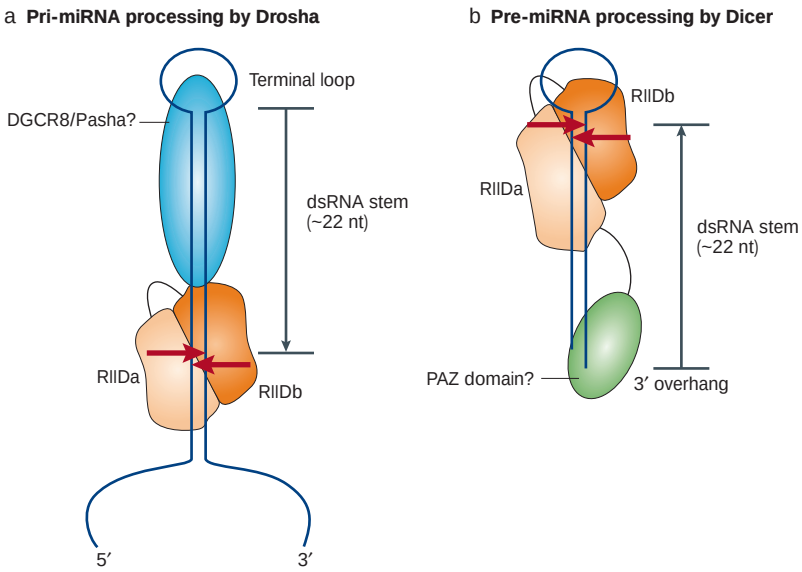

Figure 4 | **Possible mechanisms of actions for Drosha and Dicer.** The 'single processing centre' model[29] applies to both Drosha[40] and Dicer[29]. According to this model, one processing centre is formed in the interface between two RNase III domains (RIIIDa and b). The processing centre contains two closely located catalytic sites that cleave two nearby phosphodiester bonds on opposite RNA strands. The catalytic sites are composed of residues that correspond to E40, D44, D107 and E110 of *Aquifex aeolicus* RNase III. This 'single processing centre model' differs from the previous model, which assumed the existence of two processing centres with the second catalytic site being composed of residues E37 and E64 in *A. aeolicus* RNase III (REF. 127). The double-stranded RNA (dsRNA) substrate is laid on the cleft between RIIIDa (light orange) and RIIIDb (dark orange). The catalytic site on the RIIIDa side cleaves the 3′ strand, whereas another catalytic site on the RIIIDb cleaves the 5′ strand. **a** | Drosha binds to a primary transcript (pri-miRNA) and introduces a cut at approximately two helical turns, ~22 nucleotides (nt), from the terminal loop. For simplicity, the domains of Drosha other than the RIIIDs are not shown. It was proposed that DGCR8/Pasha might help the binding of the complex to RNA and/or orienting the complex on pri-miRNA. **b** | Dicer cleaves at approximately two helical turns (~22 nucleotides) from the 3′ terminus. The PAZ domain of Dicer is believed to hold the 3′ end of pre-miRNA. The domains of Dicer other than the RIIIDs and the PAZ domain are not shown.

Collaborative actions of biogenesis factors

The multiple steps in miRNA biogenesis seem to be remarkably well coordinated. Drosha initiates miRNA processing by the specific cropping of the stem-loop precursor in the nucleus[36]. Like other RNase-III-type endonucleases, Drosha cuts dsRNA to create a short ~2-nucleotide 3′ overhang[36,56] (FIG. 4). The resulting structure (a short stem plus a 3′ overhang) seems to be a signature motif for all dsRNAs that are involved in small-RNA pathways. Exportin-5 recognizes this signature motif to export pre-miRNA to the cytoplasm[49,50]. Following export, pre-miRNA is handed over to another RNase III Dicer, which has a preference for the terminus of dsRNAs containing the short 3′ overhang[77]. Therefore, Drosha confers specificity to the biogenesis pathway and facilitates the overall production rate by generating the 3′ protruding ends, which are recognized efficiently by the downstream biogenesis factors[36,49,77]. In addition, Drosha pre-determines mature miRNA sequences by precisely generating one end of the mature miRNA[36,49]. The other end is created by Dicer and measures ~22 nucleotides from the pre-existing terminus of the pre-miRNA (FIG. 4).

RNA interference (RNAi)-mediated knock-down of Drosha, exportin-5 or Dicer results in a significant reduction, but not the full loss, of mature miRNA[36,49,50,60,61]. This is probably due to the long half-life of mature miRNA. miRNAs seem to be quite stable once they enter into the effector complex. But the possibility that there could be minor alternative biogenesis pathway(s) cannot be excluded. For example, a small portion of pri-miRNAs might be exported by mRNA export pathway and become processed in the cytoplasm.

MicroRNA biogenesis in plants

Homologues of Drosha and DGCR8/Pasha have not been found in plants, which indicates that the Drosha-dependent stepwise processing model applies only to animal cells. Neither miRNA nor the Drosha homologue has been found in yeasts. Plant miRNA precursors are quite diverse in structure, and the stem-loops are usually longer than animal pri-miRNAs. Genetic studies in *A. thaliana* showed that DCL1, one of the four Dicer-like proteins in *A. thaliana* (also known as CARPEL FACTORY (CAF)), is important for miRNA accumulation[84–88]. DCL1 is a nuclear protein, which indicates that mature ~22-nucleotide miRNA might be generated in the nucleus in plants[86]. Although detection of miRNA precursors has been difficult, presumably owing to efficient processing and/or rapid turnover of precursors, at least one study showed that plant miRNAs, similar to animal miRNAs, might be processed in a stepwise manner[87], in which DCL1 might be solely responsible for all processing steps.

HASTY (HST) is a plant homologue of exportin-5 on the basis of its amino-acid sequence. An HST mutant showed pleiotropic phenotypes, which indicates that this protein might function in miRNA biogenesis too[89,90]. A recent study showed that loss-of-function mutants in HST reduced the accumulation of most miRNAs, indicating that HST functions as a nuclear export receptor[91]. The identity of the actual cargo for HST — whether it is precursor, miRNA duplex or single-stranded miRNA — is unclear.

Additional miRNA biogenesis factors include HYL1, a two dsRBD-containing nuclear protein of unknown biochemical function[92,93], and HEN1, a protein with a dsRBD and a methyltransferase domain[84,94,95]. It was recently shown that HEN1 methylates miRNA duplex at the 2′ hydroxyl groups of the 3′-end nucleotides[95]. The biochemical role of this 2′-*O*-methyl group awaits further investigation.

Regulation of microRNA biogenesis

Expression profiling studies indicate that most miRNAs are under the control of developmental and/or tissue-specific signalling pathways[21–32]. miRNA expression might be regulated at multiple steps of RNA biogenesis, although it remains to be determined which step is controlled and how this control is achieved. Transcriptional regulation is probably the main control mechanism. For example, the temporal regulation of *let-7* RNA in *C. elegans* is dependent on a transcriptional enhancer element, known as the temporal regulatory

element (TRE)[33]. This element is situated ~1200 bp upstream of the mature *let-7* RNA. Electrophoretic mobility shift assays using nuclear extract indicated that a ~22-nucleotide inverted repeat in the TRE might be a binding site for a transcription factor that is yet to be identified[33].

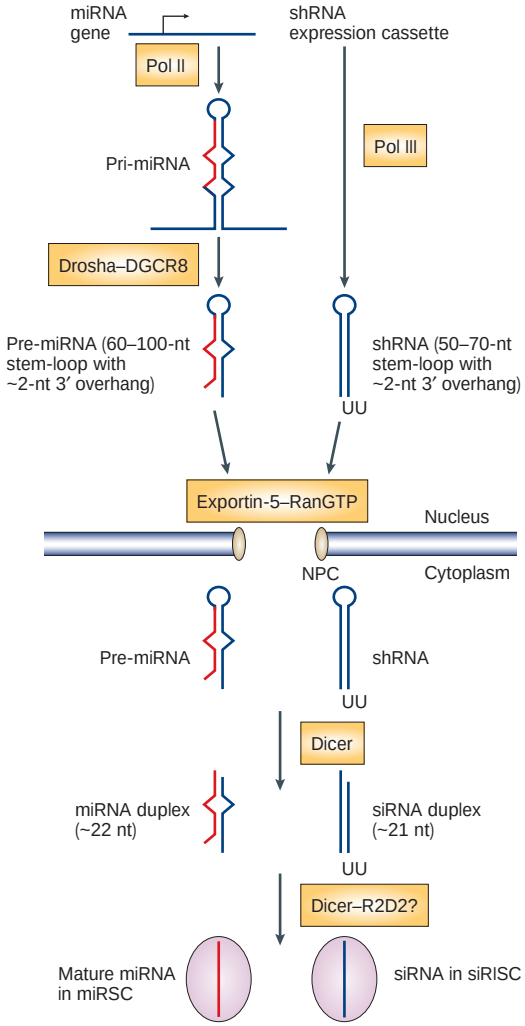

Figure 5 | **Generation of small hairpin RNAs.** MicroRNA (miRNA) genes are transcribed by RNA polymerase II (pol II), whereas short hairpin RNAs (shRNAs) are usually designed to be transcribed from pol III promoters such as the H1 promoter or the U6 promoter. The initiation step ('cropping') by Drosha results in ~70-nucleotide (nt) pre-miRNAs. shRNAs resemble pre-miRNAs in their structure and, as such, bypass the cropping step. Both pre-miRNAs and shRNAs are exported by exportin-5. From this step on, shRNAs get processed by the miRNA maturation machinery. Following export, Dicer cleaves pre-miRNA or shRNA to produce miRNA or small interfering RNA (siRNA) duplexes, respectively. The duplex is separated and usually one strand is selected as mature miRNA or siRNA, respectively, by Dicer–R2D2. The human homologue of R2D2 has not been identified so far. The final products are incorporated into miRISC or siRISC complexes, respectively, to function as guide molecules in translational control or cleavage of certain mRNAs. miRISC, miRNA-containing RNA-induced silencing complex; NPC, nuclear pore complex; pri-miRNA, primary transcript miRNA; siRISC, siRNA-containing RNA-induced silencing complex.

Some miRNAs seem to be controlled at the post-transcriptional level. For example, *C. elegans miR-38* is expressed only in the embryo, whereas the precursor (*pre-miR-38*) is ubiquitously detected, which indicates that the maturation of *pre-miR-38* might be temporally regulated[96]. It is possible that *pre-miR-38* is retained in the nucleus until export is triggered by a certain developmental signal. Alternatively, it might be bound to negative regulatory factor(s) in the cytoplasm, so that they remain inaccessible to Dicer until a certain stage.

Conclusions and what next?

We are only beginning to grasp the intricacies of the gene regulatory networks operated by miRNAs. But one thing is certain, understanding the basic mechanism of miRNA biogenesis is central to delineating these gene regulatory networks. It is imperative to identify any additional factors that are involved in the biogenesis pathway. For example, the Drosha complex (Microprocessor) might contain other components apart from DGCR8/Pasha, which might regulate the activity and/or specificity of Drosha. The activity of exportin-5 and Dicer might also be controlled by regulatory cofactors. Further work is also needed to characterize each biogenesis step in greater detail. How is a given pri-miRNA recognized by the Drosha complex? Which component is responsible for specific recognition of the substrate? Bioinformatics approaches could prove useful because they will allow the identification of common motifs in miRNA precursors, so that commonalities can be recognized. It is also important to solve the structure of biogenesis factors such as Drosha, DGCR8/Pasha, exportin-5 and Dicer, to understand the molecular basis of their activities.

It will also be interesting to understand how miRNA maturation interfaces with other aspects of RNA metabolism including transcription, pre-mRNA splicing, capping and 3′-end processing. Although each maturation step can be uncoupled from other steps in experimental settings, the coupling might exist *in vivo* to increase the efficiency and accuracy of miRNA biogenesis. Another interesting issue relates to how the miRNA pathway is related to other small-RNA pathways. Mammals and nematode worms have somewhat converged pathways for small-RNA biogenesis: for example, a single Dicer homologue seems to function in all small-RNA pathways. By contrast, insects and plants possess multiple Dicer homologues that are involved in distinct small-RNA pathways, such as miRNA, virus-induced siRNA and repeat-associated siRNA pathways[64,88]. Both biochemical and genetic approaches will be necessary to understand the differences in each pathway at the molecular level.

In-depth knowledge on miRNA biogenesis will also be crucial to improve RNAi technology[97]. miRNA biogenesis factors are required for the efficient production of siRNAs from SHORT HAIRPIN RNAS (shRNAs)[98] (FIG. 5). shRNAs are usually transcribed from a pol III promoter and contain a 19–29-bp stem and a ~2-nucleotide 3′ overhang (UU) (FIG. 5). Because of the structural similarities, shRNA might be seen as an artificial pre-miRNA. shRNAs can also be generated from pri-miRNA-like

transcripts and can be transcribed from pol II promoters[34,99]. In this case, the expression cassette includes not only the hairpin but also the flanking sequences, so that the primary transcript can be processed like an endogenous pri-miRNA in the cell. It was recently shown that the expression of exportin-5 is low in most cell types, so exportin-5 might be a key limiting factor for both miRNA biogenesis and shRNA expression[98]. As the overexpression of shRNA inhibits miRNA biogenesis[98], this might be problematic in some RNAi applications, especially in clinical settings. As such, an ideal strategy would be not to simply maximize the shRNA expression level, but to optimize the shRNA design, so that it can be processed and incorporated into RISC at the highest possible efficiency without competing with endogenous miRNAs. Further studies on miRNA biogenesis will help us to acquire a safer and more efficient strategy in applications of small RNAs.

1. Ambros, V. *et al*. A uniform system for microRNA annotation. *RNA* **9**, 277–279 (2003).
2. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
3. Cullen, B. R. Transcription and processing of human microRNA precursors. *Mol. Cell* **16**, 861–865 (2004).
4. Lewis, B. P., Shih, I. H., Jones Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
5. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
6. Bartel, D. P. & Chen, C. Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature Rev. Genet.* **5**, 396–400 (2004).
7. Kim, V. N. Small RNAs: classification, biogenesis, and function. *Mol. Cells* **19**, 1–15 (2005).
8. 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).
9. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
10. Mourelatos, Z. *et al*. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
11. Lee, Y. *et al*. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
12. 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).
This paper, together with reference 11, presents direct evidence for the pol-II-dependent transcription of miRNA genes and delineates the structure of miRNA genes.
13. 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).
14. Smalheiser, N. R. EST analyses predict the existence of a population of chimeric microRNA precursor–mRNA transcripts expressed in normal human and mouse tissues. *Genome Biol.* **4**, 403 (2003).
15. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. Identification of mammalian microRNA host genes and transcription units. *Genome Res.* **14**, 1902–1910 (2004).
Analyses the genomic locations of miRNA genes relative to defined transcription units.
16. Aukerman, M. J. & Sakai, H. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *Plant Cell* **15**, 2730–2741 (2003).
17. Tam, W. Identification and characterization of human *BIC*, a gene on chromosome 21 that encodes a noncoding RNA. *Gene* **274**, 157–167 (2001).
18. Bracht, J., Hunter, S., Eachus, R., Weeks, P. & Pasquinelli, A. E. *Trans*-splicing and polyadenylation of *let-7* microRNA primary transcripts. *RNA* **10**, 1586–1594 (2004).
19. 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).
This article reports the discovery of the first miRNA.
20. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. *M. bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
21. Lagos-Quintana, M. *et al*. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
22. Aravin, A. A. *et al*. The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* **5**, 337–350 (2003).
23. Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K. & Kosik, K. S. A microRNA array reveals extensive regulation of microRNAs during brain development. *RNA* **9**, 1274–1281 (2003).
24. Sempere, L. F., Sokol, N. S., Dubrovsky, E. B., Berger, E. M. & Ambros, V. Temporal regulation of microRNA expression in *Drosophila melanogaster* mediated by hormonal signals and broad-Complex gene activity. *Dev. Biol.* **259**, 9–18 (2003).
25. Sempere, L. F. *et al*. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* **5**, R13 (2004).
26. Calin, G. A. *et al*. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl Acad. Sci. USA* **101**, 11755–11760 (2004).
27. Liu, C. G. *et al*. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl Acad. Sci. USA* **101**, 9740–9744 (2004).
28. Schmittgen, T. D., Jiang, J., Liu, Q. & Yang, L. A high-throughput method to monitor the expression of microRNA precursors. *Nucleic Acids Res.* **32**, e43 (2004).
29. Babak, T., Zhang, W., Morris, Q., Blencowe, B. J. & Hughes, T. R. Probing microRNAs with microarrays: tissue specificity and functional inference. *RNA* **10**, 1813–1819 (2004).
30. Barad, O. *et al*. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.* **14**, 2486–2494 (2004).
31. Miska, E. A. *et al*. Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.* **5**, R68 (2004).
32. Sun, Y. *et al*. Development of a micro-array to detect human and mouse microRNAs and characterization of expression in human organs. *Nucleic Acids Res.* **32**, e188 (2004).
33. Johnson, S. M., Lin, S. Y. & Slack, F. J. The time of appearance of the *C. elegans let-7* microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. *Dev. Biol.* **259**, 364–379 (2003).
34. Zeng, Y., Wagner, E. J. & Cullen, B. R. Both natural and designed micro RNAs can inhibit the eExpression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
35. Ohler, U., Yekta, S., Lim, L. P., Bartel, D. P. & Burge, C. B. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* **10**, 1309–1322 (2004).
36. Lee, Y. *et al*. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
This paper demonstrates the function of Drosha in primary miRNA processing.
37. Filippov, V., Solovyev, V., Filippova, M. & Gill, S. S. A novel type of RNase III family proteins in eukaryotes. *Gene* **245**, 213–221 (2000).
38. Fortin, K. R., Nicholson, R. H. & Nicholson, A. W. Mouse ribonuclease III. cDNA structure, expression analysis, and chromosomal location. *BMC Genomics* **3**, 26 (2002).
39. 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).
40. Han, J. *et al*. The Drosha–DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
The authors delineate the domain structure of Drosha, a class-II RNase III protein, and confirm the 'single processing centre' model.
41. 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).
42. Gregory, R. I. *et al*. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).
43. 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).
References 40–43 report DGCR8/Pasha as the essential cofactor for Drosha.
44. Zeng, Y. & Cullen, B. R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
45. Zeng, Y., Yi, R. & Cullen, B. R. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* **24**, 138–148 (2005).
46. Kim, V. N. MicroRNA precursors in motion: exportin-5 mediates their nuclear export. *Trends Cell Biol.* **14**, 156–159 (2004).
47. Murchison, E. P. & Hannon, G. J. miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Curr. Opin. Cell Biol.* **16**, 223–229 (2004).
48. Nakielny, S. & Dreyfuss, G. Transport of proteins and RNAs in and out of the nucleus. *Cell* **99**, 677–690 (1999).
49. Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
50. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
51. 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).
Together with references 49 and 50, this paper shows that exportin-5 mediates the nuclear export of pre-miRNAs.
52. 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).
53. 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).
54. Lim, L. P. *et al*. The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **2**, 991–1008 (2003).
55. Gwizdek, C. *et al*. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J. Biol. Chem.* **278**, 5505–5508 (2003).
56. 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).
57. Zeng, Y. & Cullen, B. R. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.* **32**, 4776–4785 (2004).
58. 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).
59. 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).
60. 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).
61. 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).
62. 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).
References 58–62 reveal the key role of Dicer in small-RNA pathways.
63. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
64. 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).

65. 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).
66. 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).
67. Song, J. J. *et al*. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Struct. Biol.* **10**, 1026–1032 (2003).
68. Yan, K. S. *et al*. Structure and conserved RNA binding of the PAZ domain. *Nature* **426**, 468–474 (2003).
69. Tabara, H., Yigit, E., Siomi, H. & Mello, C. C. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell* **109**, 861–871 (2002).
70. Liu, Q. *et al*. R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* **301**, 1921–1925 (2003).
71. Ishizuka, A., Siomi, M. C. & Siomi, H. A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.* **16**, 2497–2508 (2002).
72. Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16**, 2491–2496 (2002).
73. Jin, P. *et al*. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nature Neurosci.* **7**, 113–117 (2004).
74. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
75. Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
76. Zhang, H., Kolb, F. A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
This paper presents a new model for the mechanism of action for RNase III proteins.
77. 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).
78. 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).
79. Liu, J. *et al*. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437–1441 (2004).
80. Meister, G. *et al*. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* **15**, 185–197 (2004).
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. Tomari, Y., Matranga, C., Haley, B., Martinez, N. & Zamore, P. D. A protein sensor for siRNA asymmetry. *Science* **306**, 1377–1380 (2004).
84. Park, W. *et al*. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
85. Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
86. Papp, I. *et al*. Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol.* **132**, 1382–1390 (2003).
87. Kurihara, Y. & Watanabe, Y. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl Acad. Sci. USA* **101**, 12753–12758 (2004).
88. Xie, Z. *et al*. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, E104 (2004).
89. Bollman, K. M. *et al*. HASTY, the *Arabidopsis* ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* **130**, 1493–1504 (2003).
90. Telfer, A. & Poethig, R. S. HASTY: a gene that regulates the timing of shoot maturation in *Arabidopsis thaliana*. *Development* **125**, 1889–1898 (1998).
91. Park, M. Y. *et al*. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **102**, 3691–3696 (2005).
92. Vazquez, F., Gasciolli, V., Crete, P. & Vaucheret, H. The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.* **14**, 346–351 (2004).
93. Han, M. H., Goud, S., Song, L. & Fedoroff, N. The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl Acad. Sci. USA* **101**, 1093–1098 (2004).
94. Boutet, S. *et al*. *Arabidopsis* HEN1. A genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.* **13**, 843–848 (2003).
95. Yu, B. *et al*. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932–935 (2005).
96. Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T. & Jewell, D. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* **13**, 807–818 (2003).
97. Hannon, G. J. & Rossi, J. J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371–378 (2004).
98. 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).
99. Zeng, Y., Cai, X. & Cullen, B. R. Use of RNA polymerase II to transcribe artificial microRNAs. *Methods Enzymol.* **392**, 371–380 (2005).
100. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
101. 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).
102. Reinhart, B. J. *et al*. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
103. Chang, S., Johnston, R. J., Jr., Frokjaer-Jensen, C., Lockery, S. & Hobert, O. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* **430**, 785–789 (2004).
104. Johnston, R. J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
105. Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
106. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* **304**, 594–596 (2004).
107. Xu, P., Vernooy, S. Y., Guo, M. & Hay, B. A. The *Drosophila* microRNA *Mir-14* suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**, 790–795 (2003).
108. Esau, C. *et al*. MicroRNA-143 regulates adipocyte differentiation. *J. Biol. Chem.* **279**, 52361–52365 (2004).
109. Vazquez, F. *et al*. Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* **16**, 69–79 (2004).
110. Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H. L. & Poethig, R. S. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* **18**, 2368–2379 (2004).
111. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679 (2002).
112. Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**, 1605–1619 (2002).
113. Mette, M. F., van der Winden, J., Matzke, M. & Matzke, A. J. Short RNAs can identify new candidate transposable element families in *Arabidopsis*. *Plant Physiol.* **130**, 6–9 (2002).
114. Djikeng, A., Shi, H., Tschudi, C. & Ullu, E. RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24–26-nucleotide RNAs. *RNA* **7**, 1522–1530 (2001).
115. 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).
116. Aravin, A. A. *et al*. Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Mol. Cell. Biol.* **24**, 6742–6750 (2004).
117. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**, 315–327 (2002).
118. Hall, I. M. *et al*. Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
119. Reinhart, B. J. & Bartel, D. P. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**, 1831 (2002).
120. Volpe, T. A. *et al*. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
121. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
122. Mochizuki, K., Fine, N. A., Fujisawa, T. & Gorovsky, M. A. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* **110**, 689–699 (2002).
123. Liu, Y., Mochizuki, K. & Gorovsky, M. A. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. *Proc. Natl Acad. Sci. USA* **101**, 1679–1684 (2004).
124. Taverna, S. D., Coyne, R. S. & Allis, C. D. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. *Cell* **110**, 701–711 (2002).
125. Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K. & Gage, F. H. A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* **116**, 779–793 (2004).
126. 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).
127. Blaszczyk, J. *et al*. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure (Camb.)* **9**, 1225–1236 (2001).
128. Poy, M.N. *et al.* A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**, 226–230 (2004).

Acknowledgements

I am grateful to members of my laboratory, especially to Young Kook Kim for the bioinformatics analysis of the miRNA gene structure. This work was supported by a Molecular and Cellular BioDiscovery Research Program grant from the Ministry of Science and Technology and a Research Fellowship from the Ministry of Education and Human Resources Development of Korea.

Competing interests statement

The author declares no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez
let-7
Swiss-Prot: http://us.expasy.org/sprot/
CAF | DGCR8 | Dicer | Drosha | exportin-5 | HEN1 | RDE-4
The miRNA Registry:
http://www.sanger.ac.uk/Software/Rfam/mirna
miR-21 | *miR-38*
FlyBase: http://flybase.bio.indiana.edu
AGO1 | AGO2 | Dicer-1 | Dicer-2 | FMR1 | R2D2
TAIR: http://www.arabidopsis.org
HST | HYL1

FURTHER INFORMATION

V. Narry Kim's laboratory: http://plaza.snu.ac.kr/~narrykim
Noncoding RNAs Database:
http://biobases.ibch.poznan.pl/ncRNA
RNA & DNA Folding Applications:
http://www.bioinfo.rpi.edu/applications/mfold
The miRNA Registry:
http://www.sanger.ac.uk/Software/Rfam/mirna

Access to this interactive links box is free online.