DRUG DISCOVERY TODAY BIOSILICO

A computational view of microRNAs and their targets

James R. Brown and Philippe Sanseau

Small non-coding RNAs called microRNAs have been shown to play important roles in gene regulation across a broad range of metazoans from plants to humans. In this review, the nature and function of microRNAs will be discussed, with special emphasis on the computational tools and databases available to predict microRNAs and the genes they target.

One of the most important developments in molecular biology over the past two decades is the emerging picture of a new layer of gene regulation under the control of small yet versatile RNAs [1]. In 1993, two papers suggested that, in *Caenorhabditis elegans*, a small RNA, called *lin-4*, was responsible for regulating the expression of the *lin-14* gene through direct interaction with its mRNA [2,3]. Over the following years, it became evident that short RNAs are much more versatile than originally thought. It has now been clearly demonstrated that they play a major biological role by altering the expression levels of a diverse repertoire of genes in a sequence-dependent manner. This is achieved at either the transcriptional or post-transcriptional level.

Small RNAs can be divided into four subfamilies: microRNAs (miRNAs) [4–6], short interfering RNAs (siRNAs) [7,8], tiny non-coding RNAs (tncRNAs) [9] and unique double-stranded RNAs called small modulatory RNAs (smRNAs) [10]. The characteristics of these families have been reviewed elsewhere [11]. Briefly, miRNAs are single-stranded RNAs 21–25 nucleotides in length that silence cellular target genes at the post-transcriptional level. siRNAs are a class of double-stranded short RNAs 21–22 nucleotides in length that have the exact complementary sequence to the mRNA of their respective target genes. Both tncRNAs and smRNAs have been discovered recently. The function of tncRNAs is unknown. smRNA appears to play a major role in neuronal differentiation by regulating the expression of neuron-specific genes, possibly via interactions at the protein level. Overall, differences between some of these families seem slightly superficial. For example, it has been recently shown that one mammalian miRNA, *miR-196*, binds to *Hoxb8* mRNA to direct its degradation, an effect traditionally associated with siRNAs [12]. Conversely, in mammalian tissue culture, imperfect binding of siRNAs with their targets will lead to translational repression, similar to the mode of action of miRNAs [13].

Since the identification of *lin-4* in *C. elegans* [2,3], hundreds of miRNAs have been identified in a wide range of organisms from plants to humans. Furthermore, the general activity of short non-coding RNAs is now exploited as a major experimental genomics platform to analyse gene function by silencing target genes in a process named RNA interference (RNAi) [11,14]. This review will focus on a particular naturally occurring family of short RNAs, the miRNAs, for which computational approaches have been important in not only the discovery of new members but also the identification of target genes.

Functional characteristics and biogenesis of miRNAs

The functional characteristics and biogenesis of miRNAs have been thoroughly reviewed previously, but a brief review here is warranted [15–19]. The

James R. Brown
GlaxoSmithKline,
Bioinformatics Discovery and Analysis,
Upper Providence,
1250 South Collegeville Road,
UP1345,
PO Box 5089,
Collegeville, PA 19426-0989,
USA
Philippe Sanseau
GlaxoSmithKline,
Bioinformatics Discovery and Analysis,
Stevenage,
Gunnels Wood Road,
Hertfordshire,
SG1 2NY,
UK
e-mail:
philippe.x.sanseau@gsk.com

1359-6446/05/$ – see front matter PII: S1359-6446(05)03398-2

mature miRNA, usually 21–25 nucleotides in length, is originally derived from a larger precursor, ~60–70 nucleotides long, that folds into an imperfect stem-loop structure (Figure 1). In animals, these miRNA precursors are themselves derived from cleavage of the primary miRNA transcript by a multiprotein complex, the microprocessor. Known microprocessor components are Drosha RNase III [20] and Pasha (Partner of Drosha), a double-stranded (ds)RNA-binding protein [21,22]. After cleavage, miRNA precursors are transported into the cytoplasm by a cargo transporter called Exportin 5 [23,24]. Subsequently, the miRNA precursors are cleaved into an imperfect dsRNA duplex by another endonuclease RNase III enzyme, called Dicer [20,25–27]. This duplex is composed of the mature miRNA strand and its complementary strand, commonly noted as miRNA*. The mature miRNA strand of the duplex is then incorporated into an effector complex, the RISC (RNA-induced silencing complex). The other strand is apparently rapidly degraded. Interestingly, it appears that the strand that enters the RISC is almost always the one with the less stable 5′ end [28,29].

Purification from cells of the RISC has shown that it contains at least one member of the Argonaute (AGO) protein family [30,31]. Recent structural and mutagenesis studies suggest that Ago2 is specifically responsible for RISC cleavage activity [32–34]. Several other proteins have also been co-purified from the RISC, including the RNA-binding proteins VIG, fragile X-related protein [35] and nuclease Tudor-SN [36], and the helicases Gemin-3 and Gemin-4 [37].

Within the RISC, miRNAs have been shown to regulate gene expression by either inducing mRNA cleavage or translational repression. In plants, the majority of miRNAs induce cleavage by being perfect or nearly perfect complements to the coding region of their target mRNA [19,38,39]. In animals, most miRNAs bind with imperfect complementarity to multiple sites in the 3′ untranslated region (UTR) of their target mRNAs and cause translational repression [40]. The regulatory mechanism of translational repression is not fully understood. Experiments in *C. elegans* suggest that protein synthesis is affected at the translational elongation or termination steps, while the amount of mRNA remains unchanged.

From the perspective of genome organisation, a majority of mammalian miRNA genes overlap with transcription units in introns of protein-coding genes, and in both introns and exons of mRNA-like non-coding genes [41,42]. In addition, it appears that miRNAs are transcribed in parallel with these host genes. Intronic miRNAs have also been identified in *C. elegans* [43].

Animal miRNAs seem to regulate several diverse biological processes, such as development [44], apoptosis [45], haematopoietic differentiation [46] and fat metabolism [47]. Plant miRNAs appear to target particularly transcription factors [39]. Some miRNAs display higher expression levels in particular tissues [48–50] or differentiated expression levels in tumour tissues [51,52]. Recently, microarrays have been used to show other instances of tissue-specific expression of miRNAs [53–55].

The capability of animal miRNAs to be active through imperfect complementarity has profound implications for determining the spectrum of target genes, as well as for the use of RNAi as an experimental tool to silence gene expression. There are several reports of synthesized siRNAs affecting non-targeted mRNA transcripts with only partial sequence complementarity [56]. In

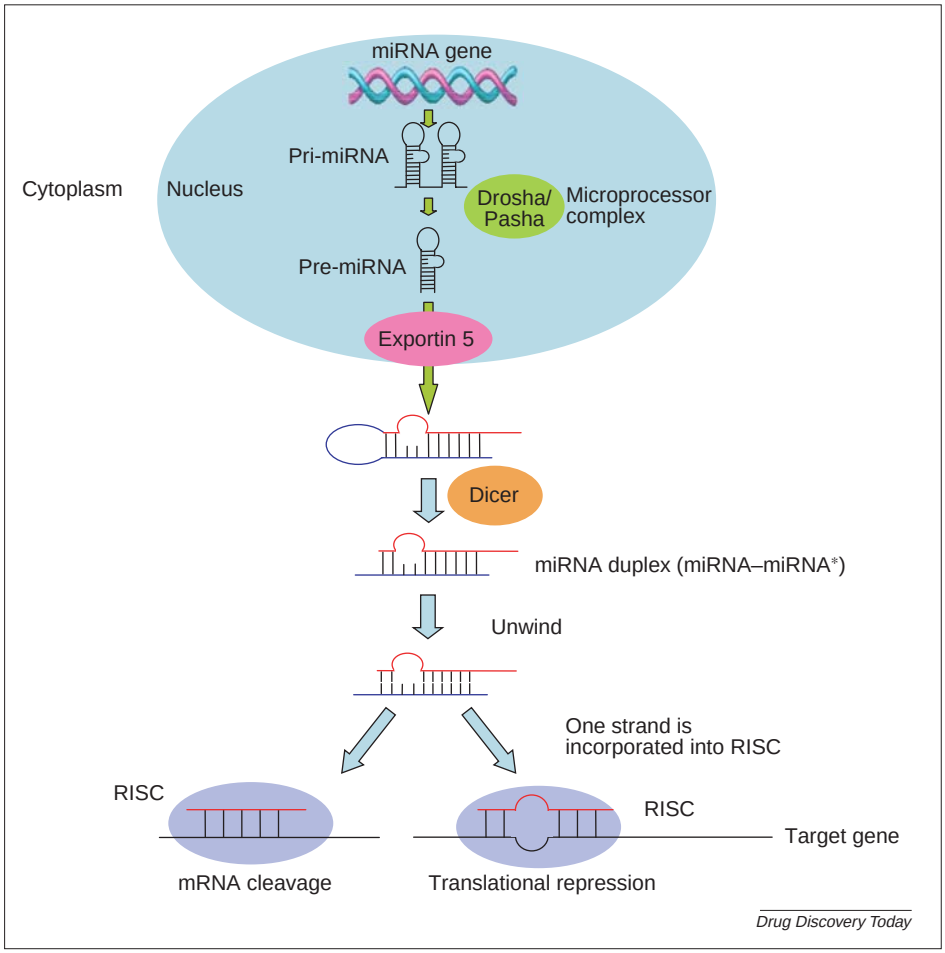

FIGURE 1

Gene expression regulation by miRNAs. Primary miRNAs (pri-miRNAs) are initially processed by the Drosha/Pasha complex ('microprocessor') into ~60–70 nucleotide precursor miRNAs (pre-miRNAs) in the nucleus. These pre-miRNAs are transported into the cytoplasm by Exportin 5. In a next step, they are cleaved by Dicer into an imperfect double-stranded duplex. One strand of this duplex is incorporated into the RISC (RNA-induced silencing complex). This complex binds to the target gene and will lead to translational repression. However, in some cases, miRNAs regulate gene expression by mRNA cleavage [12] rather than by translational repression.

TABLE 1

Some major computational approaches used to identify animal miRNA genes

Program/approach	URL	Species	Number of genes estimated	References
MiRseeker		*Drosophila*	110	[69]
MiRscan	http://genes.mit.edu/mirscan	*C. elegans*	120	[70]
		Human	180–255	[71]
Phylogenetic shadowing		Human	976	[75]

addition, it has been shown that siRNAs can affect the translation of unintended targets [57,58] and induce a non-specific interferon response [59,60]. One should be aware of these potential liabilities when designing RNAi experiments [61]. However, the potential of RNAi remains enormous, not only as an experimental tool but also for the discovery of RNA-based drugs, as recently demonstrated by the silencing of a gene involved in cholesterol metabolism, apolipoprotein B [62]. Several companies now offer siRNA libraries to silence genes of pharmaceutical importance. For example, both Qiagen (Venlo, Netherlands; http://www.qiagen.com) and Ambion (Austin, Texas; http://www.ambion.com) have developed siRNA libraries that focus on genes of therapeutic interest, such as G-protein-coupled receptors, ion channels, kinases or nuclear receptors.

Computational identification of miRNAs

Table 1 summarizes the principal computational approaches that have been used to identify miRNA genes in animals and plants. It was the discovery of the *C. elegans* miRNA *let-7* [63] that led to computational approaches for discovering other novel miRNAs. Through simple sequence homology searches using BLASTN [64], orthologues of *let-7* were identified in numerous species [65]. Orthologues of the first identified *C. elegans* miRNA, *lin-4*, escaped computational identification because of cross-species sequence divergence. However, in many other cases, simple sequence searches did identify homologues and orthologues of miRNAs isolated by cloning [5,66]. Alternative approaches used RNA fold prediction to identify sequences likely to form stem-loop structures. For example, the program Mfold, which predicts RNA secondary structure by free energy minimisation [67], was used to identify novel miRNAs in *C. elegans* [9]. The sequences analysed were initially derived from intergenic regions conserved between the closely related nematode species *Caenorhabditis briggsae* and *C. elegans*, as identified by WABA, an algorithm developed for large-scale genomic alignment [68]. Mfold was also used to select potential miRNAs obtained from cloning screens [4].

The MiRseeker procedure [69] also examines the folding of RNA sequences conserved between two *Drosophila* species using Mfold. The MiRseeker procedure uses predictions of stem-loop structure formation as key criteria. But it also takes into account the nucleotide divergence of miRNA candidates, as the authors detected less selective pressure in the loop sequences of orthologous precursor miRNAs. Their miRNaseeker program identified 48 conserved novel miRNAs in *Drosophila*.

MiRscan (http://genes.mit.edu/mirscan) is yet another computational approach that has been used to scan the genomes of *C. elegans* and humans [70,71]. It relies on the secondary structure prediction program RNAfold [72], as an alternative to Mfold, to computationally identify potential stem-loops. After first predicting >35 000 stem-loops in sequences conserved between *C. briggsae* and *C. elegans*, 50 published miRNAs from these two species were used as a training set to develop the MiRscan program. The principle of MiRscan is to slide a 21-nucleotide window along candidate stem-loops and assign a log-likelihood score. The scoring method was developed using criteria such as the number of base pairs from the loop to the closest end of the candidate miRNA, variations in 5′ and 3′ sequence conservation, miRNA base pairing, the extension of base pairing, bulges, sequence biases in the first five bases of the miRNA (such as a U as the first base), and 2 to 9 consensus base pairs between the miRNA and the terminal loop region. Using this method in combination with sequencing and validation techniques, 30 additional miRNAs were identified. MiRscan was also used to extrapolate that the *C. elegans* genome probably has 120 miRNA genes. Another study using a similar search strategy but different selection criteria and fewer experimentally validated candidates yielded a comparable prediction of 120–300 miRNAs in the *C. elegans* genome [73].

MiRscan has been used to estimate the number of miRNA genes in other species. Human stem-loop sequences were predicted from intergenic regions with some conservation between humans, mouse and pufferfish (*Takifugu rubripes*). For these sequences, MiRscan estimated an upper bound of ~255 human miRNA genes, and a lower bound between 180 and 200 genes. However, some miRNAs appear to be poorly conserved between humans and pufferfish [71]. Therefore, the estimated upper bound of predicted human miRNA genes might be low. MiRseeker analysis estimated that around 110 miRNA genes are present in the *Drosophila* genome [69]. These various estimates all correspond roughly to 1% of the total number of predicted genes in the respective genomes. The MiRscan program

has recently been improved by adding criteria based on the presence of a consensus sequence upstream of almost all transcribed miRNAs in *C. elegans* [50].

Very recently, a particular phylogenetic approach has been used to identify potential novel human miRNAs [74]. The strategy, known as phylogenetic shadowing [75], is based on the sequence comparison of closely related species and provides an accurate method to identify conserved regions at the nucleotide level. The authors compared the sequences of over 100 miRNA regions in 10 different primates. A characteristics profile was identified: variation in loop sequences, conservation in stems of hairpins and a significant decrease in conservation of sequences flanking the hairpins. This pattern was used to identify potential new miRNAs in pairwise alignments of more divergent species, such as humans and mouse or humans and rat. After additional filtering, such as looking at the folding free energy of candidate sequences [76], 976 potential human RNAs were identified; this set contains over 80% of all known human miRNAs in release version 3.1 of the miRNA Registry (see below) [77]. Northern blot analyses combined with database searches reached a conservative estimate of 200–300 novel human miRNAs, a twofold increase over previous studies [71].

In a recent study, the entire set of human and mouse precursor and mature miRNAs in the miRNA Registry, version 2.2 [77] was compared to the human genome using the sequence comparison tool BLAT [78]. After further filtering using tools (Mfold) and criteria (G•U base pairings), 35 human and 45 mouse new potential miRNAs were identified [79].

Similar computational strategies have been used to predict miRNAs in plants. However, these methods had to be modified to take into account some specific characteristics of plant miRNAs, such as the variability in length of miRNA precursors (invalidating a fixed-window approach), differences in G+C content and lower sequence conservation of precursors. In an initial analysis, the computational pipeline known as MIRFINDER predicted 91 miRNA genes in the genome of *Arabidopsis thaliana* [80]. Another study predicted a similar number, 95 miRNA genes in *A. thaliana* [81]. A different computational strategy is based on the observation that miRNA genes appear to be organized in clusters and scans for stem-loops near known miRNAs [82].

Computational prediction of miRNA targets

The identification of miRNA targets is an essential step towards understanding their regulatory function. Recently, different computational approaches have been developed and used to uncover such targets [83]; these are summarized in Table 2. The large majority of these methods are based on knowledge of the nature of the pairing between the miRNA and the target gene in animals and plants.

In animals, imperfect complementarity with miRNAs makes computational prediction of their targets particularly challenging. However, over the past few years, progress has been made in the prediction of *Drosophila* and mammalian miRNA targets [84–90]. The principles behind these approaches are relatively similar and are based on previously acquired knowledge of the pairing between animal miRNAs, such as *lin-4* [2] and *let-7* in *C. elegans* [63,65], and *bantam* [45] in *Drosophila*, and their targets. In general, the criteria for developing these computational pipelines are the complementarity between the 3′ UTRs of the potential targets and miRNAs, with an emphasis on the critical pairing at the 5′ end of the miRNA [91,92]; the conservation of target 3′ UTR sequences in orthologous genes; and the kinetics and thermodynamics of the association between the miRNA and its target, as determined by RNA folding programs [93]. Some of these criteria, such as the importance of the pairing at the 5′ end of the miRNA, were strongly suggested by computational approaches before experimental evidence [94]. In combination, computational searches for known pairings and experimental confirmation are commonly used to measure the validity of the approaches. Although the general principles are similar, some technical details are not. Key differences exist in the methods used to measure conservation and to predict single [88] or multiple binding sites [85] in miRNA targets, and in the statistical approaches chosen. One illustration of this is the different conservation rules applied to select 3′ UTRs for *Drosophila* [84,86]. A review of all published methods used to predict mammalian targets found that the overlap of identical predictions from the different computational approaches varied between 10% and nearly 50% for a common set of 79 miRNAs [89].

TargetScan [85] is a widely used program for predicting miRNA targets. In its latest version, TargetScanS, certain criteria have been simplified, such as the omission of both the thermodynamic stability of pairings and multiple target sites per UTR [95]. With the availability of new genomes (i.e. chicken and dog) and updated annotations for others (i.e. humans, mouse and rat), TargetScanS looks for target site sequence conservation across species, thereby reducing the number of false positive predictions. Some original parameters of TargetScan, such as the requirement for a seven-nucleotide match in the region of complementarity (the seed) between the miRNA and its target gene, can also be relaxed, to a six-nucleotide match. In addition, the presence of conserved adenosines flanking the seed miRNA sequence has been integrated into TargetScanS. Using this refined algorithm on four genomes (humans, mouse, rat and dog), the authors predicted that the 3′ UTRs of 30% (5300 genes) of the gene set used in the analysis (17 850 orthologous mammalian genes) were potential miRNA targets. This work reinforces the possibility that a larger proportion of mammalian genes than initially thought are under the potential control of miRNAs.

Interestingly, miRNAs have also been identified in viruses such as the Epstein–Barr virus (EBV) [96]. The authors used a computational approach tested first on *Drosophila* [85]

TABLE 2

Summary of major programs for the identification of animal miRNA targets

Name	URL	Species	References
TargetScan and TargetScanS	http://genes.mit.edu/targetscan	Vertebrates	[86,96]
DIANA MicroT	http://diana.pcbi.upenn.edu/DIANA-microT	Human/mouse	[89]
miRanda	http://www.microrna.org//miranda.html	*Drosophila*/human	[85,90]
miRNA - Target prediction	http://www.russell.embl.de/miRNAs/home.html	*Drosophila*	[84]
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/welcome.html	*Drosophila*	[88]

to identify potential miRNA targets. Several targets were predicted, including B-cell-specific chemokines and cytokines, transcriptional regulators, and genes involved in signal transduction pathways, or cell proliferation or death. Therefore, miRNA silencing is possibly a mechanism used by EBV to regulate host gene expression.

The computational identification of miRNA targets in plants is relatively more straightforward than in animals, as a large majority of miRNAs bind to their gene targets with near to perfect complementarity. In a first approach, PatScan [97] was used to identify complementary RNAs for 16 miRNAs from *A. thaliana* [39]. Hits selected had four or less mismatches and no gaps, with non-canonical and G•U pairs considered mismatches. This analysis led to the prediction of 49 targets for 16 miRNAs. The same authors have recently improved this computational approach by relaxing the criteria, allowing for gaps and more mismatched nucleotides. Additionally, similar to the comparative genomic analyses used to predict animal miRNAs, they looked at sequence conservation between *A. thaliana* and *Oryza sativa*, which revealed 19 additional plant miRNA targets [98].

With the growing number of verified and potential miRNA targets, animal miRNA targets appear to be more functionally diverse than their plant counterparts [15]. Plant miRNAs appear to target in particular transcription factors involved in development [99]. Although animal miRNAs control similar target genes, they also affect a large spectrum of diverse biological functions, such as protein metabolism, transport, intracellular signalling cascades, cell organization and response to external stimuli [85,88,89,95]. Some of these miRNA targets have particular relevance in a disease context, such as genes of the TGF-β signalling pathway [95].

Computational prediction of miRNA targets is bound to improve in accuracy over time. This will be achieved by the thorough refinement of training sets, rules for defining complementarity and statistical approaches. Recently, it has been suggested to examine not only 3′ UTR sequences but also 5′ UTRs [18]. As always, experimental validation, in an iterative process with *in silico* methods, will be essential to benchmark any new predictive algorithms.

miRNA data resources

The increasing number of identified miRNAs has led to the establishment of a searchable database maintained at the Sanger Centre in the UK, called the miRNA Registry (http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml); it can also be downloaded by FTP (fast transfer protocol) [77]. The database can be searched by BLAST [64] for both the hairpin and mature sequences, and by keywords, names, references and annotations. At the time of writing, the latest release (version 5.1) of the miRNA Registry contains 1420 entries from 2 plant and 7 metazoan species, the majority of which are vertebrates. The current number of human entries, 222, is probably an underestimate if novel miRNAs recently identified by phylogenetic shadowing are confirmed experimentally [74].

The criteria for the inclusion of miRNA sequences in the database are not identical across species. For example, all *C. elegans* entries are supported by experimental evidence by cloning or tissue distribution. However, the 79 miRNAs from *C. briggsae* were identified by close sequence homology to *C. elegans* miRNAs and for their capacity to form stem-loop structures. Human, mouse and rat sequences are mixtures of experimentally verified miRNAs and computationally determined putative orthologues. Of the 222 miRNAs listed for humans, only 126 have been experimentally verified. In support of cataloguing miRNAs, the miRNA Registry provides names for novel miRNAs before publication, as well as guidelines to promote greater consistency in the annotation of miRNAs [99].

A specific *Arabidopsis* database for small RNAs has also been developed and is available at http://asrp.cgrb.oregonstate.edu [100]. It contains miRNAs and also siRNAs. Version 3.0 contains a total of 1920 unique sequences.

Conclusions

Over the past few years, the complex and subtle roles of miRNAs in gene regulation have been increasingly appreciated [101]. Computational approaches have played a key role in identifying miRNAs from plants to animals, as well as in predicting their putative gene targets. Studies that verify *in silico* predictions with wet-bench experiments will be crucial to the testing and refinement of miRNA computational databases and algorithms. Sequences of miRNAs are now being tracked in a comprehensive database using agreed standards. Many challenges remain in understanding miRNAs and dissecting the affected pathways. One hopes that iterative interactions between *in silico* and experimental methods will continue to push forward future developments in this exciting field.

References

1 Ruvkun, G. *et al.* (2004) The 20 years it took to recognize the importance of tiny RNAs. *Cell* 116(suppl. 2), S93–S96
2 Lee, R.C. *et al.* (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854
3 Wightman, B. *et al.* (1993) Post-transcriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862
4 Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294, 862–864
5 Lagos-Quintana, M. *et al.* (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858
6 Lau, N.C. *et al.* (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862
7 Elbashir, S.M. *et al.* (2001) Duplexes of 21 nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498
8 Elbashir, S.M. *et al.* (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15, 188–200
9 Ambros, V. *et al.* (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* 13, 807–818
10 Kuwabara, T. *et al.* (2004) A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* 116, 779–793
11 Novina, C.D. and Sharp, P.A. (2004) The RNAi revolution. *Nature* 430, 161–164
12 Yekta, D. *et al.* (2004) MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* 304, 594–596
13 Doench, J.G. *et al.* (2003) siRNAs can function as miRNAs. *Genes Dev.* 17, 438–442
14 Fire, A. *et al.* (1998) Potent and specific genetic interference by double stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811
15 Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297
16 He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531
17 Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 343–349
18 Ambros, V. (2004) The functions of animal microRNAs. *Nature* 431, 350–355
19 Baulcombe, D. (2004) RNA silencing in plants. *Nature* 431, 356–363
20 Lee, Y. *et al.* (2003) RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419
21 Denli, A.M. *et al.* (2004) Processing of primary microRNAs by the microprocessor complex. *Nature* 432, 231–235
22 Gregory, R.I. *et al.* (2004) The microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240
23 Yi, R. *et al.* (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17, 3011–3016
24 Lund, E. *et al.* (2004) Nuclear export of microRNA precursors. *Science* 303, 95–98
25 Hutvagner, G. *et al.* (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* 293, 834–838
26 Grishok, A. *et al.* (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106, 23–34
27 Ketting, R.F. *et al.* (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15, 2654–2659
28 Schwartz, D.S. *et al.* (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199–203
29 Khvorova, A. *et al.* (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115, 209–216
30 Hammond, S.M. *et al.* (2001) Argonaute 2, a link between genetic and biochemical analyses of RNAi. *Science* 293, 1146–1150
31 Carmell, M.A. *et al.* (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* 16, 2733–2742
32 Song, J-J. *et al.* (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305, 1434–1437
33 Meister, G. *et al.* (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15, 185–197
34 Liu, J. *et al.* (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441
35 Caudy, A.A. *et al.* (2002) Fragile X-related protein and VIG associate with RNA interference machinery. *Genes Dev.* 16, 2491–2496
36 Caudy, A.A. *et al.* (2003) A micrococcal nuclease homologue in RNAi effector complexes. *Nature* 425, 411–414
37 Hutvagner, G. and Zamore, P.D. (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297, 2056–2060
38 Llave, C. *et al.* (2002) Cleavage of *Scarecrow-like* mRNA targets directed by a class *Arabidopsis* miRNA. *Science* 297, 2053–2056
39 Rhoades, M.W. *et al.* (2002) Prediction of plant microRNA targets. *Cell* 110, 513–520
40 Olsen, P.H. and Ambros, V. (1999) The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* 216, 671–680
41 Rodriguez, A. *et al.* (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res.* 14, 1902–1910
42 Ying, S-Y. and Lin, S-L. (2005) Intronic microRNAs. *Biochem. Biophys. Res. Commun.* 326, 515–520
43 Ohler, U. *et al.* (2004) Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* 10, 1309–1322
44 Olsen, P.H. and Ambros, V. (1999) The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* 216, 671–680
45 Brennecke, J. *et al.* (2003) *bantam* encodes a developmentally regulate microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* 113, 25–36
46 Chen, C.Z. *et al.* (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303, 83–86
47 Xu, P. *et al.* (2003) The *Drosophila* microRNA *mir-14* suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* 13, 784–789
48 Houbaviy, H.B. *et al.* (2003) Embryonic stem cell-specific microRNAs. *Dev. Cell* 5, 351–358
49 Sempere, L.F. *et al.* (2004) 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
50 Suh, M-R. *et al.* (2004) Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* 270, 488–498
51 Calin, G.A. *et al.* (2002) Frequent deletions and down-regulation of microRNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukaemia. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15524–15529
52 Michael, M.Z. *et al.* (2003) Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol. Cancer Res.* 1, 882–891
53 Liu, C-G. *et al.* (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9740–9744
54 Babak, T. *et al.* (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. *RNA* 10, 1813–1819
55 Sun, Y. *et al.* (2004) Development of a micro-array to detect human and mouse microRNAs and chracterization of expression in human organs. *Nucleic Acids Res.* 32, e188
56 Jackson, A.L. *et al.* (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* 21, 635–637
57 Saxena, S. *et al.* (2003) Small RNAs with imperfect match to endogenous mRNA repress translation: implications for off-target activity of siRNA in mammalian cells. *J. Biol. Chem.* 278, 44312–44319
58 Scacheri, P.C. *et al.* (2004) Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1892–1897
59 Sledz, C.A. *et al.* (2003) Activation of the interferon system by short-interfering RNAs. *Nat. Cell Biol.* 5, 834–839
60 Bridge, A.J. *et al.* (2003) Induction of an interferon response by RNAi vectors in mammalian cells. *Nat. Genet.* 34, 263–264
61 Jackson, A.L. and Linsley, P.S. (2004) Noise amidst the silence: off-target effects of siRNAs. *Trends Genet.* 20, 521–524
62 Soutschek, J. *et al.* (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* 432, 173–178
63 Reinhardt, B.J. *et al.* (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906
64 Altschul, S.F. *et al.* (1990) Basic alignment search tool. *J. Mol. Biol.* 215, 403–410
65 Pasquinelli, A.E. *et al.* (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408, 86–89
66 Lau, N.C. *et al.* (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862
67 Mathews, D.H. *et al.* (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary

structure. *J. Mol. Biol.* 288, 911–940
68 Kent, W.J. and Zahler, A.M. (2000) Conservation, regulation, synteny, and introns in a large-scale *C. briggsae-C. elegans* genomic alignment. *Genome Res.* 10, 1115–1125
69 Lai, E.C. *et al.* (2003) Computational identification of *Drosophila* microRNA genes. *Genome Biol.* 4, R42
70 Lim, L.P. *et al.* (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* 17, 991–1008
71 Lim, L.P. *et al.* (2003) Vertebrate microRNA genes. *Science* 299, 1540
72 Hofacker, I.L. *et al.* (1994) Fast folding and comparison of RNA secondary structures. *Monatsheftef. Chemie* 125, 167–188
73 Grad, Y. *et al.* (2003) Computational and experimental identification of *C. elegans* microRNAs. *Mol. Cell* 11, 1253–1263
74 Berezikov, E. *et al.* (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* 120, 21–24
75 Bofelli, D. *et al.* (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* 299, 1391–1394
76 Bonnet, E. *et al.* (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20, 2911–2917
77 Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic Acids Res.* 32, D109–D111
78 Kent, W.J. (2002) BLAT: the BLAST-like alignment tool. *Genome Res.* 12, 656–664
79 Weber, M.J. (2005) New human and mouse microRNA genes found by homology search. *FEBS J.* 272, 59–73
80 Bonnet, E. *et al.* (2004) Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11511–11516
81 Wang, X-J. *et al.* (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. *Genome Biol.* 5, R65
82 Seitz, H. *et al.* (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* 34, 261–262
83 Lai, E.C. (2004) Predicting and validating microRNA targets. *Genome Biol.* 5, 115
84 Stark, A. *et al.* (2003) Identification of *Drosophila* MicroRNA targets. *PLoS Biol* 1, E60
85 Enright, A.J. *et al.* (2003) MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1
86 Lewis, B.P. *et al.* (2003) Prediction of mammalian microRNA targets. *Cell* 115, 787–798
87 Rajewsky, N. and Socci, N.D. (2004) Computational identification of microRNA targets. *Dev. Biol.* 267, 529–535
88 Rehmsmeier, M. *et al.* (2004) Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507–1517
89 Kiriakidou, M. *et al.* (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* 18, 1165–1178
90 John, B. *et al.* (2004) Human microRNA targets. *PLoS Biology* 2, e363
91 Vella, M.C. *et al.* (2004) The *C. elegans* microRNA *let-7* binds to imperfect *let-7* complementarity sites from the *lin-41* 3′ UTR. *Genes Dev.* 18, 32–37
92 Doench, J.G. and Sharp, P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.* 18, 504–511
93 Wutchy, S. *et al.* (1999) Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* 49, 145–165
94 Lai, E.C. (2002) Micro RNAs are complementary to 3′UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* 30, 363–364
95 Lewis, B.P. *et al.* (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20
96 Pfeffer, S. *et al.* (2004) Identification of virus-encoded microRNAs. *Science* 304, 734–736
97 Dsouza, M. *et al.* (1997) Searching for patterns in genomic data. *Trends Genet.* 13, 497–498
98 Jones-Rhoades, M.W. and Bartel, D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14, 787–799
99 Ambros, V. *et al.* (2003) A uniform system for microRNA annotation. *RNA* 9, 277–279
100 Gustafson, A.M. *et al.* (2005) ASRP: the *Arabidopsis* small RNA project database. *Nucleic Acids Res.* 33 Database Issue, D637–D640
101 Bartel, D.P. and Chen, C.Z. (2004) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.* 5, 396–401