REVIEW

Computational Identification of MicroRNAs and Their Targets

Sungroh Yoon and Giovanni De Micheli

One of the most important advances in biology in recent years may be the discovery of RNAs that can regulate gene expression. As one kind of such functional noncoding RNAs, microRNAs (miRNAs) form a class of endogenous 19–23-nucleotide RNAs that can have important regulatory roles in animals and plants by targeting transcripts for cleavage or translational repression. Since the discovery of the very first miRNAs, computational methods have been an invaluable tool that can complement experimental approaches to understand the biology of miRNAs. Most computational approaches associated with miRNA research can be classified into two broad categories, namely miRNA gene identification and miRNA target prediction. In this review, we summarize the principles of in silico prediction of miRNAs and their targets, and provide a comprehensive survey of specific methods that have been proposed in the field. **Birth Defects Research (Part C) 78: 118–128, 2006.** © 2006 Wiley-Liss, Inc.

> **Since the discovery of the very first miRNAs, computational methods have been an invaluable tool that can complement experimental approaches to understand the biology of miRNAs.**

INTRODUCTION

MicroRNAs (miRNAs) are endogenous small noncoding RNAs that can have important roles in the regulation of genes in animals and plants. The mature miRNA is transcribed from the miRNA gene followed by successive cleavage by enzymes Drosha and Dicer and is typically 19–23 nucleotides long. So far, more than 300 human miRNAs have been identified (Griffiths-Jones et al., 2006; Hsu et al., 2006) (Fig. 1), and it has been suggested that the total number of human miRNAs is at least 800 (Bentwich et al., 2005). An example of human miRNAs is shown in Figure 2. miRNAs can target mRNAs for posttranscriptional regulation, such as translational repression (mostly animals) and mRNA cleavage (mostly plants), as illustrated in Figure 3. It has been estimated that miRNAs target more than 5300 human genes (Lewis et al., 2005). Many plant miRNAs are mostly involved in transcriptional regulation, whereas the regulatory impact of animal miRNAs are more pervasive, with roles identified in developmental timing, cell death, cell proliferation, hematopoiesis, and patterning of the nervous system (Ambros, 2004). A further review of the biology of miRNAs can be found in other articles in this issue or in Ambros (2004), Bartel (2004), and Bartel and Chen (2004).

Since the very first miRNAs *lin-4* and *let-7* were identified by genetic analysis of *C. elegans* developmental timing (Lee et al., 1993; Reinhart et al., 2000), researchers have developed numerous computational models and tools that can complement biological experiments to understand the diverse regulatory roles of miRNAs. These bioinformatics approaches have been invaluable in coping with the complexity of finding putative miRNA genes and their targets, as well as in deciphering their functions. Most computational methods that have been used for miRNA studies can be classified into two broad categories, namely miRNA identification and miRNA target identification. Methods in each category can further be divided depending on whether a method is for animals, plants, or viruses, since the biology of miRNAs is somewhat different in each case. The organization of this review follows this categorization.

Regardless of these categories, the basic principle of many computational methods is simple: learn from known examples to find new ones. For instance, an miRNA target detection algorithm can be trained using the properties of known miRNA-mRNA duplexes and can then be used for finding new miRNA-mRNA duplexes. From a computational perspective, this corresponds to the problem of machine learning, an area of artificial intelligence used to develop tech-

Sungroh Yoon is from the Computer Systems Laboratory, Stanford University, Stanford, California.
Giovanni De Micheli is from the Integrated Systems Center, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

*Correspondence to: Sungroh Yoon, Room 334 William Gates Hall, 353 Serra Mall, Stanford, CA 94305. E-mail: sryoon@stanford.edu
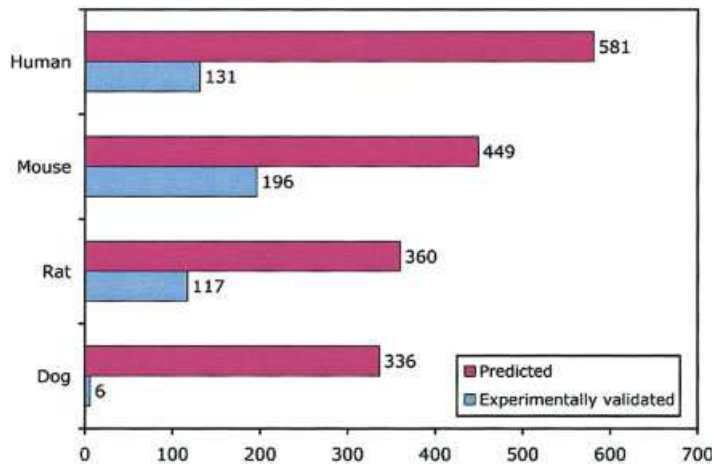
Figure 1. The total numbers of predicted and experimentally validated miRNAs for four different species (http://mirnamap.mbc.nctu.edu.tw/php/statistics.php).

niques that allow computers to learn from examples (Mitchell, 1997; Hastie et al., 2001; Alpaydin, 2004). Since all the mechanisms behind miRNAs and their actions are not completely revealed, computational tasks associated with miRNA studies are often posed as a challenging machine learning problem with limited prior information.

> **Since all the mechanisms behind miRNAs and their actions are not completely revealed, computational tasks associated with miRNA studies are often posed as a challenging machine learning problem with limited prior information.**

COMPUTATIONAL IDENTIFICATION OF miRNA GENES

Challenges

As previously stated, the task of computationally identifying miRNA genes can be formulated as a machine learning problem. As such, the main difficulty is to increase the specificity of a prediction algorithm or to reduce the number of false-positive predictions of miRNA genes. Primary sequence information alone or secondary structure information alone is not specific enough. This is because the length of a miRNA sequence is small, and the number of stem-loop or hairpin structures that exist along a genome is huge. Typically, it is not a good idea to simply scan the genome for hairpins, hoping to identify miRNA genes. Additional filtering is required to increase specificity, as is described in the next section.

Another computational challenge arises due to the lack of information on the level of miRNA expressions (Aravin and Tuschl, 2005). Given the many stages of development and many different cell and tissue types as well as environmental conditions, it is often difficult to validate whether a prediction is a false positive or not. Posttranscriptional modifications of miRNAs may only be identified using a cloning approach (Bartel, 2004; Pfeffer et al., 2004), and direct cloning and sequencing coupled with computational methods may be most effective. For a review of miRNA gene validation methods, the reader is directed to Bentwich (2005).
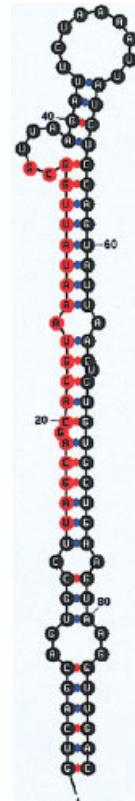

Figure 2. An example of human miRNA (hsa-mir-16-1). Shown are stem loops involving the mature miRNAs (red) and flanking sequences (black). The plot was generated from the miRNAMap server (Hsu et al., 2006) using the Mfold package (Mathews et al., 1999; Zuker, 2003). Human miR-16 has been cloned by independent groups (Lagos-Quintana et al., 2001; Mourelatos et al., 2002). Two identical chromosome 13 loci were reported, and they appear to map to the same locus in subsequent genome assemblies (Lim et al., 2003a). This gene and miR-15a are clustered within 0.5 kb at 13q14. This region has been shown to be deleted in more than half of B cell chronic lymphocytic leukemias (CLL). Both miR-15a and miR-16 are deleted or downregulated in more than two-thirds of CLL cases (Lim et al., 2003a). A second putative mir-16 hairpin precursor is located on chromosome 3.

Prediction Principles

miRNA genes can be searched by motif searches combining sequence, structure, and conservation information. More specifically, algorithms to detect new miRNA genes use the following properties of the known miRNA genes as prediction principles:

- miRNA genes are small noncoding genes (<150 bp). They are typically cloned multiple times,

and their clone length distribution sharply peaks between 21 and 23 nt (Bartel, 2004; Pfeffer et al., 2004).

- A mature miRNA is derived from its precursor miRNA called a primary RNA, and its structure can be used as a search template. A primary RNA forms a short (~60-nt) stable extended stem-loop structure (or a hairpin structure), with continuous helical pairing and a few internal bulges (Fig. 1). The genomic sequence flanking an miRNA contains a highly complementary (20- to 30-nt) segment, which is required to form the characteristic pre-miRNA hairpin structure (Bartel, 2004; Pfeffer et al., 2004). To identify a precursor structure and the miRNA embedded in the stem (and not the loop) of the precursor, an RNA folding program such as the Vienna package (Hofacker, 2003) or the Mfold package (Mathews et al., 1999; Zuker, 2003) is used. A precursor should have a low free energy level to be stable (>30 kcal/mol), and the precursor structure should be unbranched (Bengert and Dandekar, 2005). To avoid repetitive DNAs and other uninformative predictions, we can also check if each of the four nucleotides is sufficiently represented in the structure (Bengert and Dandekar, 2005).
- miRNAs are normally highly conserved in the genomes of related species, although a small number of miRNAs may be universally conserved in all animals (Bartel, 2004; Pfeffer et al., 2004). Most prediction methods, therefore, search for the conservation of miRNA gene sequences (Bonnet et al., 2004) and stem-loop structures across species (Lai et al., 2003; Lim et al., 2003a, 2003b). This approach is helpful for screening out many false positives but finds only conserved miRNAs, so some recent approaches do not rely on sequence conservation (Bentwich et al., 2005). The conservation profile often resembles a saddle-like structure because the loop sequence and sequences flanking

Figure 3. miRNAs and targets (Lai, 2004). **A:** Plant miRNAs exhibit extensive complementarity to their targets, but animal miRNAs generally do not. **B:** Various configurations for miRNA-target duplexes: one near-perfect binding site for one miRNA (upper left), one strong site for one miRNA (lower left), multiple modest sites for one miRNA (upper right), and multiple modest sites for multiple miRNAs (lower right).

the stem structure are much more variable than the miRNA and its complementary miRNA* sequence (Lai et al., 2003; Altuvia et al., 2005; Berezikov et al., 2005). The evolutionary divergence between orthologous miRNAs shows a characteristic pattern: the terminal loops usually have more mutations than the arms of the stem-loops, and the miRNA-coding arms are more conserved than the non-miRNA-coding arms (Kong and Han, 2005).

- Many miRNAs occur in clusters, usually separated by a few kilobases (Seitz et al., 2003; Sachidanandam, 2005; Sewer et al., 2005). These clusters can be detected by mapping the miRNA genes to the genome. The miRNA genes in each cluster can be transcribed together and may control mRNAs involved in related functions (Sachidanandam, 2005).
- An miRNA has to be complementary to the 3′ untranslated regions (UTRs) of a target mRNA (Bartel, 2004).
- miRNAs can occur in intergenic regions, in introns of protein coding regions, or in exons and introns of noncoding genes (Rodriguez et al., 2004).
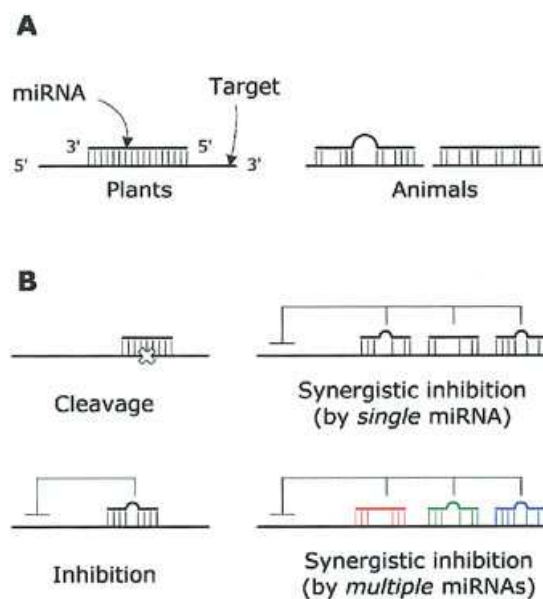
In addition, the prediction of miRNAs in plants or viruses needs some special considerations:

- Plant miRNAs tend to have characteristics somewhat different from animal counterparts. For instance, the level of sequence conservation of miRNA precursors is lower in plants. The length of hairpin structures is also more variable in plants, and some algorithms that use a fixed-sized sequence window need to be modified appropriately. In addition, the G + C content often differs in plants and animals (Jones-Rhoades and Bartel, 2004; Sachidanandam, 2005).
- Viral miRNAs are rarely evolutionarily conserved according to the cloning of small RNAs from virus-infected cells (Aravin and Tuschl, 2005). Thus, viral miRNA prediction algorithms often do not consider miRNA primary sequence conservation (Pfeffer et al., 2004).

Methods to Detect Animal miRNAs

PalGrade

Bentwich et al. (2005) developed an integrative approach combining

computational predictions with microarray analysis and sequence-directed cloning to prove their hypothesis that the total number of miRNAs may be much larger and that several have emerged only in primates. Their algorithm, called PalGrade, does not rely on sequence conservation, unlike many other techniques described in this section. Thus, PalGrade could identify a large number of miRNAs that are unique to primates and are undetected by other prediction algorithms. They reported 89 novel human miRNAs, 53 of which are not conserved beyond primates, and suggested that the total number of miRNAs in human could be at least 800.

MiRscan

Lim et al. (2003a, 2003b) developed a computational method called MiRscan to find miRNA genes conserved in more than one genome. They identified 30 novel miRNAs in *C. elegans* and 38 novel human miRNAs. The program MiRscan used the secondary structure prediction program RNAfold (Hofacker, 2003) to identify evolutionarily conserved hairpin structures, each of which was considered as a potential miRNA precursor. To further assess the location of the miRNA within each hairpin structure, MiRscan passed a 21-nt window along the hairpin and assigned a log-likelihood score to each position for its similarity to known miRNAs. A total of 50 published miRNAs from *C. briggsae* and *C. elegans* were used as a training set for their algorithm, which successfully identified conserved miRNAs within many conserved hairpins found in the genome (~35,000 hairpins conserved between *C. briggsae* and *C. elegans* and ~15,000 hairpins conserved between man, mouse, and puffer fish).

Ohler et al. (2004) further improved the accuracy of MiRscan by considering sequence conservation about 200 bp upstream of the miRNA fold-back and a highly significant sequence motif (with consensus CTCCGCCC) that is present upstream of almost all independently transcribed nematode miRNA genes. They observed that sequence features outside of the RNA secondary structure can therefore be very useful for the computational identification of eukaryotic noncoding RNA genes. They estimated that the total number of confidently identified nematode miRNAs could be as many as 100.

Phylogenetic shadowing

Berezikov et al. (2005) identified 16 novel human miRNAs using phylogenetic shadowing, a powerful technique that can assess the degree of conservation of each nucleotide in a sequence (Boffelli et al., 2003). They sequenced 122 miRNAs in 10 primate species to reveal conservation characteristics of miRNA genes. They observed that nucleotides in the stem of miRNA hairpin precursors are significantly more conserved than in sequences flanking the hairpin structure and in the loops of the hairpins. Their results suggested the presence of significantly higher numbers of miRNAs in the human genome than previously estimated. They estimated that there could be at least 200–300 new putative human miRNAs, a two-fold increase over previous studies such as (Lim et al., 2003a).

miRseeker

Lai et al. (2003), using a detection program called miRseeker, estimated that *Drosophila* genomes contain around 110 miRNA genes and identified 48 miRNA candidates. They examined the folding of RNA sequences conserved between two *Drosophila* species using the Mfold package (Mathews et al., 1999; Zuker, 2003) in order to detect extended stem-loop structures having a nucleotide divergence characteristic of known miRNAs.

ProMiR

Nam et al. (2005) proposed ProMiR, a genetic programming approach to learn common structures of miRNAs from known miRNA precursors. Unlike previous approaches that could detect only abundantly expressed miRNAs or close homologs of previously identified miRNAs, their method is a probabilistic colearning model for miRNA gene finding, which simultaneously considers the structure and sequence of miRNA precursors. Their study suggested that the miRNA gene family may be more abundant than previously anticipated and confer highly extensive regulatory networks on eukaryotic cells.

Cluster approaches

Sewer et al. (2005) exploited the property that miRNAs are occasionally found in clusters and focused on genomic regions around already known miRNAs. Starting with the known human, mouse, and rat miRNAs, they analyzed 20 kb of flanking genomic regions for the presence of putative precursor miRNAs. Their result is available at http://www.mirz.unibas.ch.

Seitz et al. (2003) scanned for stem-loops near known miRNAs, based upon the observation that miRNA genes appear to be organized in clusters.

Comparative methods

Xie et al. (2005) presented a comparative analysis of the human, mouse, rat, and dog genomes to create a systematic catalog of common regulatory motifs in promoters and 3′ UTRs. In this approach, conserved sequence motifs within 3′ UTRs of mRNAs were first identified, and then conserved candidate miRNAs were predicted. These motifs were about 7-nt-long and were mostly complementary to the 5′ end of known miRNA sequences as well as candidate sequences with fold-back structure. Their results suggested that previous estimates of the number of human miRNA genes were low.

Weber (2005) reported the results of a systematic search for interspecies orthologs of miRNA precursors. These authors first compared the entire human and mouse precursors and mature miR-

NAs in the miRNA Registry (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006) with the human genome using a sequence comparison tool. They further used RNA folding and G-U base pairing criteria and identified 35 human and 45 mouse novel miRNA genes.

Other methods

Ambros et al. (2003) employed cDNA sequencing and comparative genomics to identify *C. elegans* small RNAs with properties similar to miRNAs. The program Mfold (Mathews et al., 1999; Zuker, 2003), which predicts RNA secondary structure by free energy minimization, was used to identify novel miRNAs.

Grad et al. (2003) developed a method to predict miRNAs in the *C. elegans* genome using sequence conservation and structural similarity to known miRNAs and generated 214 candidates. They estimated that the *C. elegans* genome may encode between 140 and 300 miRNAs and potentially many more. Their search strategy was similar to MiRscan (Lim et al., 2003a, 2003b), but used different selection criteria and reported fewer (14 miRNAs in total) experimentally validated candidates.

Legendre et al. (2005) estimated how many new miRNAs could be recovered using a profile-based strategy and produced 265 new miRNA candidates that were not previously found in miRNA databases.

Rodriguez et al. (2004) annotated the genomic position and context of the class of noncoding RNAs (ncRNAs) in the human and mouse genomes, in order to derive a global perspective on the transcription of miRNAs in mammals.

Methods to Identify miRNA Genes in Plants or Viruses

Computational methods for predicting plant miRNAs have been focused on the *Arabidopsis thaliana* and *Oryza sativa* genomes.

Wang et al. (2004a) described the identification of 20 miRNAs from a cDNA library for *O. sativa*.

Bonnet et al. (2004) presented a genome-wide computational pipeline known as MIRFINDER and predicted 91 miRNA genes in the *A. thaliana* genome. Their method was based on the conservation of short sequences between two genomes, *A. thaliana* and *O. sativa*, as well as on properties of the secondary structure of miRNA precursors. This method was fine-tuned to take into account plant-specific properties, such as the variable length of the miRNA precursor sequences.

Wang et al. (2004b) predicted 95 miRNA genes in *A. thaliana* using a computational method for genome-wide prediction of miRNAs and their target mRNAs. This method used characteristic features of known plant miRNAs as criteria to search for miRNAs conserved between *A. thaliana* and *O. sativa*. Their results suggested that at least some miRNA precursors are polyadenylated at certain stages.

Jones-Rhoades and Bartel (2004) developed a comparative genomic approach to systematically identify both miRNAs and their targets that are conserved in *A. thaliana* and *O. sativa*.

Adai et al. (2005) presented a single genome approach for the detection of miRNAs in *A. thaliana*. Their tool, called findMiRNA, predicted potential miRNAs within candidate precursor sequences that have corresponding target sites within transcripts.

Recently, methods that use expressed sequence tag (EST) analysis have been proposed. Williams et al. (2005) developed a protocol to mine a nonannotated, noncoding EST database in order to discover new *A. thaliana* small RNA. Zhang et al. (2005) identified and characterized new plant miRNAs using EST analysis. A total of 338 new potential miRNAs were identified in 60 plant species.

miRNAs have also been identified in viruses such as the Epstein-Barr virus (EBV), herpes virus, and HIV. Pfeffer et al. (2005) combined a new miRNA gene prediction method with small-RNA cloning from several virus-infected cell types, and identified miRNA genes in herpes virus, a pathogenic virus. Pfeffer et

al. (2004) showed that EBV expresses several miRNA genes. Couturier and Root-Bernstein (2005) hypothesized that viral-encoded miRNA from HIV-1 may directly alter T cell, macrophage, and dendritic cell activity. To investigate a potential correlation between the genomic complementarity of HIV-1 and host cell protein expression, a local alignment search was performed to assess for regions of complementarity between the HIV-1 proviral genome and the mRNA coding sequence of various proteins expressed by CD+ T cells and macrophages.

COMPUTATIONAL PREDICTION OF miRNA TARGETS

Prediction of Plant miRNA Targets

Because of the near perfect complementarity of miRNAs to their targets, the prediction of miRNA targets in plants is straightforward (Rhoades et al., 2002), and automated plant miRNA target prediction can now be performed online (Zhang, 2005). Transcription factors represent only about 5% of *A. thaliana* protein-coding genes (Rhoades et al., 2002; Jones-Rhoades and Bartel, 2004), but more than 50% of the putative plant miRNA targets are transcription factors. These miRNA-regulated transcription factors regulate developmental patterning, cell proliferation, and environmental and hormonal responses (Kidner and Martienssen, 2005). Interestingly, miRNAs can tune their own expression, suggesting a negative feedback mechanism (Rhoades et al., 2002; Xie et al., 2003; Vaucheret et al., 2004).

Principles of Predicting miRNA Targets in Animals

Computational prediction of miRNA targets is more challenging in animals because of the imperfect complementarity of miRNAs to their targets. The detection principles used by most approaches are relatively similar, and are based on pre-

vious knowledge on the pairing of mRNAs and miRNAs such as lin-4 (Lee et al., 1993) and let-7 in *C. elegans* (Pasquinelli et al., 2000; Reinhart et al., 2000), as well as bantam in *Drosophila* (Brennecke et al., 2003). Prediction criteria include the following:

- The miRNA sequence is complementary to the 3′ UTR sequence of potential target mRNAs. Especially, the strong binding of the 5′ end (the first eight base pairs) of the mature miRNA to the 3′ UTR sequence is very important for targeting, whereas the G:U wobble pairing reduces the silencing efficiency (Doench and Sharp, 2004). In addition to the 3′ UTR regions, Ambros (2004) suggested that 5′ UTR regions of a potential target mRNA should also be checked.
- The kinetics and thermodynamics of RNA-RNA duplexes can be determined by RNA folding programs, and have been considered important by most algorithms. However, a recent study by Lewis et al. (2005) showed that this condition can be omitted without lowering the specificity of a detection algorithm by incorporating other conserved sequence information.
- The conservation of target 3′ UTR sites in related genomes is critical. TargetScanS (Lewis et al., 2005) also considered the presence of conserved adenosines surrounding the seed miRNA sequence.
- More than one miRNA typically regulates one message, indicative of cooperative translational control, whereas one miRNA may have several target genes, reflecting target multiplicity (Enright et al., 2003). That is, combinatorial control of a single target by multiple miRNAs may be an important feature of miRNA targeting, very similar to the mode of transcription factor control of genes (Doench and Sharp, 2004; Hobert, 2004; Xie et al., 2005), and multiple binding sites for an miRNA on the 3′ UTR can increase the efficiency of RNA si-

lencing (Doench and Sharp, 2004).
- Du and Zamore (2005) observed that lack of a strong secondary structure at the miRNA-binding site on the target may be an important feature.
- miRNAs could also target other miRNAs for silencing (Lai et al., 2004).

Methods for Animal miRNA Target Detection

TargetScan and TargetScanS

Lewis et al. (2003) initially developed TargetScan, a computational method to predict the targets of conserved vertebrate miRNAs. They selected 79 pan-mammalian miRNAs with homologs in human, mouse, and puffer fish and identical sequence in human and mouse (not necessarily in puffer fish), as well as 55 pan-vertebrate miRNAs that had identical sequences in all three genomes. TargetScan was applied using these two sets of miRNAs, and the authors predicted 451 regulatory target genes by identifying mRNAs with conserved pairing to the 5′ region of the miRNA and evaluating the number and quality of these complementary sites. More specifically, TargetScan searched for a strong 7-nt seed, starting from the second nucleotide from the 5′ end of the miRNA. The algorithm then used the RNAFold package (Hofacker, 2003) to calculate the thermodynamic free energy of the binding and assigned scores to both single binding sites and multiple binding sites. The functions of the predicted mRNA targets were distributed over a broad range, with particular enrichment in transcriptional regulation.

Lewis et al. (2005) further improved TargetScan and proposed TargetScanS by relaxing some of the parameters of TargetScan and incorporating new criteria. A seed was originally defined to be a 7-nt match (nucleotide positions 2–8 of the miRNA) in the region of complementarity between the miRNA and its target gene. TargetScanS relaxed this condition and used a 6-nt match (nucleotide positions 2–7) as

a seed. In addition, TargetScanS did not consider the thermodynamic stability of pairings and multiple sites in each target. On the other hand, TargetScanS scanned target site sequence conservation across two more species (chicken and dog) in addition to the three genomes (mouse, rat, and human) examined in TargetScan. This helped TargetScanS reduce the number of false positives. Moreover, TargetScanS succeeded in further increasing sensitivity by considering the presence of conserved adenosines surrounding the seed miRNA sequence. (It was found that the immediate downstream position of the seed match is highly conserved and is often an adenosine, which mostly base-pairs with the U of the first nucleotide of the miRNA.) The algorithm was estimated to have a 22% false-positive rate for targets conserved in mammals.

TargetScanS specifically recovered all known miRNA targets, and Lewis et al. (2005) predicted that miRNAs target more than 5300 human genes, which represented 30% of the gene set used in the analysis (17,850 orthologous mammalian genes). This work suggested that a larger number of mammalian genes than initially thought are potentially controlled by miRNAs.

PicTar

PicTar (http://pictar.bio.nyu.edu) is a computational method for identifying common targets of miRNAs in vertebrates, *C. elegans*, and *Drosophila*. Krek et al. (2005) initially presented PicTar, and found that each vertebrate miRNA targets roughly 200 transcripts on average. Using PicTar, Grun et al. (2005) exploited cross-species comparisons to predict that each miRNA targets 54 genes on average in *Drosophila*.

The algorithm used in PicTar can identify binding sites that are co-regulated by multiple miRNAs in a coordinated manner, in addition to binding sites targeted by a single miRNA. To filter out false positives, PicTar used statistical tests based on genome-wide alignments of

eight vertebrate genomes, and considered clustering coexpressed miRNAs and matching miRNAs with putative targets that are expressed in the same context. The algorithm could specifically recover published miRNA targets and was estimated to have an approximately 30% false-positive rate.

miRanda

To identify miRNA targets in *Drosophila* and human, Enright et al. (2003) and John et al. (2004) developed an algorithm called miRanda (http://www.microrna.org). For each miRNA, miRanda selected target genes on the basis of three properties: sequence complementarity using a position-weighted local alignment algorithm (to emphasize binding of the 5′-end segment more than the 3′-end segment in the miRNA), free energies of RNA-RNA duplexes, and conservation of target sites in related genomes. The algorithm correctly recovered nine out of 10 published miRNA targets and was estimated to have a 24% false-positive rate. The functions of the predicted target genes were enriched in development, transcription factors, translational regulation, and cell fate specification, including the nervous system.

Using miRanda, Enright et al. (2003) identified several hundred target genes potentially regulated by one or more known miRNAs, and John et al. (2004) predicted that miRNA genes (about 1% of human genes) regulate protein production for 10% or more of all human genes.

MovingTargets

MovingTargets is a software program that allows a researcher to predict a set of miRNA targets that satisfy an adjustable set of biological constraints (Burgler and Macdonald, 2005). The method consisted of two steps: the creation of a database of potential targets and the screening of all possible miRNA–target pairs for adherence to constraints suggested by analysis of the known miRNA–target interac-

tions. Biological miRNA target constraints they suggested include the number of target sites in the mRNA, the strength of miRNA–mRNA hybridization, the number of consecutive base pairs involving the 5′ part of the miRNA, the total number of miRNA 5′ nucleotides involved in base pairing to the target, and the number of nucleotides in the miRNA 5′ region involved in G:U base pairs. Using MovingTargets, the authors identified a high-likelihood set of 83 miRNA targets in Drosophila, all of which adhere to strict biological constraints.

RNAhybrid

Rehmsmeier et al. (2004) presented RNAhybrid, a program that predicts multiple potential binding sites of miRNAs in large target RNAs (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid). This program used an RNA folding algorithm improved over Mfold (Mathews et al., 1999; Zuker, 2003) or RNAFold (Hofacker, 2003), and could find the hybridization sites that were energetically most favorable. RNAhybrid predicted *Drosophila* miRNA targets by matching a 6-nt seed starting from the second nucleotide from the 5′ end of the miRNA.

DIANA-microT

Kiriakidou et al. (2004) used a combined bioinformatics and experimental approach to predict human miRNA targets. The authors described a computational program called DIANA-microT that can computationally identify miRNA targets by focusing on single binding site targets and by scanning binding sites with a typical central bulge and 3′ binding. DIANA-microT successfully recovered all published *C. elegans* miRNA targets.

TargetBoost

Saetrom et al. (2005) proposed a machine learning algorithm called TargetBoost that works on a set of validated miRNA targets in lower organisms to create weighted se-

quence motifs that capture the binding characteristics between miRNAs and their targets (http://www.interagon.com/demo). The authors showed that TargetBoost's weighted sequence motif approach is favorable to using both the duplex stability and the sequence complementarity steps.

Binary classification of target sites

Brennecke et al. (2005) evaluated the minimal requirements for functional miRNA-target duplexes in vivo and classified target sites with different functional properties into two categories, namely 5′ dominant target sites that have high complementarity to the 5′ end of miRNA, and 3′ compensatory sites that have weak 5′ complementarity and depend on strong base-pairing to the 3′ end of miRNA. The authors further classified the 5′ dominant target sites into canonical sites and seed sites depending upon 3′ base-pairing. They found that high complementarity is sufficient for the miRNA-mRNA duplex to be functional, and that certain sites with as little as seven base-pairings between the 5′ end of miRNA and the target are sufficient to regulate the target in vivo. The authors estimated that an average miRNA has approximately 100 target sites, indicating that miRNAs regulate a large fraction of protein-coding genes and that miRNA 3′ ends are key determinants of target specificity within miRNA families.

miRNA module prediction

Yoon and De Micheli (2005) proposed an unsupervised machine learning method to predict miRNA regulatory modules or groups of miRNAs and target genes that are believed to participate cooperatively in posttranscriptional gene regulation. Their method was based upon the observation that more than one miRNA typically regulates one message, and that one miRNA may have several target genes (Enright et al., 2003). The authors represented the interac-
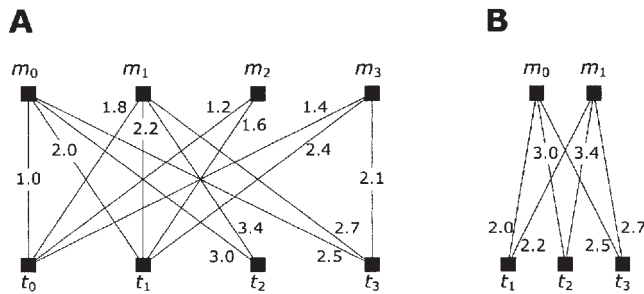
Figure 4. Graphical modeling of miRNA-target interactions (Yoon and De Micheli, 2005). **A:** A weighted bipartite graph representation. The vertices in the first row (m_0, m_1, m_2, m_3) represent miRNAs, and the vertices in the second row (t_0, t_1, t_2, t_3) correspond to mRNAs. An edge represents binding of an miRNA to a target and is assigned a weight according to the binding strength. **B:** A miRNA regulatory module is defined as a complete subgraph (with similar weights for the edges incident on target vertices) in the weighted bipartite graph. The example module shown consists of two miRNAs (m_1, m_2) and three targets (t_1, t_2, t_3).

tions between miRNAs and their targets by a weighted bipartite graph, and modeled a miRNA regulatory module as a complete subgraph (biclique) in the graph (Fig. 4). A total of 431 miRNA regulatory modules were predicted from the human genome.

Use of mRNA folding

Robins et al. (2005) incorporated the folded mRNA structure information to target prediction but did not consider evolutionary conservation. Their result suggested that miRNAs have fewer targets than previously reported, which is contradictory to other recent studies (Berezikov et al., 2005; Lewis et al., 2005; Xie et al., 2005).

Other methods

Xie et al. (2005) used a computational method to identify a large class of 7-nt conserved regulatory motifs. Many of these motifs are likely to be miRNA target sites, according to the observation that these motifs have strong directional bias with respect to the DNA strand, and they end with an adenosine complementary to the 5′-end of a binding miRNA. The authors suggested that miRNAs regulate at least 20% of human genes.

Smalheiser and Torvik (2004) presented a computational study to show that human miRNA–mRNA target interactions follow different rules than have been previously

characterized in *Drosophila* and *C. elegans*. The authors performed a population-wide statistical analysis of how human miRNAs interact complementarily with human mRNAs, searching for characteristics that are significantly different from scrambled control sequences. These characteristics were used to identify 71 statistically significant outlier mRNAs. It was suggested that many human miRNAs exhibited long exact matches of 10 nt or more, unlike *C. elegans* and *Drosophila*.

Rajewsky and Socci (2004) proposed a target detection program that incorporated both kinetic and thermodynamic components of target recognition and applied the program to identifying evolutionary conserved sequences.

Stark et al. (2003) used a target prediction algorithm to detect *Drosophila* miRNA targets by screening conserved 3′ UTR sequences from the Drosophila genome for potential miRNA targets. The screening procedure combines detecting conserved complementary sequences of the 5′-end 8-nt seed of the miRNA and calculating the thermodynamic stability of the binding using the Mfold package (Mathews et al., 1999; Zuker, 2003). This approach revealed striking clusters of functionally related targets among the top predictions for specific miRNAs (for instance, notch target genes for miR-7, proapoptotic genes for the miR-2 family, and enzymes from a metabolic pathway

for miR-277). Multiple binding sites were required in order to achieve significant predictive power, but the authors suggested that valid targets can be identified from sequence alone.

miRNA Target Detection for Plants and Viruses

Similar to plant miRNA gene prediction studies, miRNA target identification algorithms have been focused on two species, *A. thaliana* and *O. sativa*.

Rhoades et al. (2002) predicted regulatory targets for 14 *A. thaliana* miRNAs by identifying mRNAs with near complementarity. The PatScan program (Dsouza et al., 1997) was used to identify 49 unique targets complementary to the miRNAs. Having four or less mismatches and zero gaps was considered as a match, and noncanonical and G:U pairs were called a mismatch. Many of these 49 plant miRNA target sites have since been confirmed experimentally (Llave et al., 2002; Emery et al., 2003; Kasschau et al., 2003; Tang et al., 2003). Complementary sites within predicted targets are conserved in rice, and the target sites were often found in transcription factors involved in developmental patterning or stem cell maintenance and identity (Bengert and Dandekar, 2005).

Jones-Rhoades and Bartel (2004) developed a comparative genomic approach to identify both miRNAs and their targets that are conserved in *A. thaliana* and *O. sativa*. This method relaxed the criteria used by Rhoades et al. (2002) and allowed gaps and more mismatches. Their algorithm searched for sequence conservation between *A. thaliana* and *O. sativa*, just like animal miRNA prediction methods. The authors revealed 19 additional plant miRNA targets.

Li and Zhang (2005) detected 96 candidate Arabidopsis miRNAs by searching short complementary sequences between transcription factor open-reading frames and intergenic region sequences, and considering RNA secondary structures and the sequence conversa-

TABLE 1. Online Resources for miRNA Research

Name	URL	Main feature	References
miRNA registry/ miRBase	http://microrna.sanger.ac.uk	miRNA sequences, annotations, and predicted targets	Griffiths-Jones (2004, 2006)
miRNAMap	http://mirnamap.mbc.nctu.edu.tw	Genomic maps for miRNA genes and targets	Hsu et al. (2006)
MiRscan	http://genes.mit.edu/mirscan	miRNA gene scan	Lim et al. (2003a, b); Ohler et al. (2004)
RNA regulatory networks	http://www.mirz.unibas.ch	Putative miRNA gene and target scan	Sewer et al. (2005)
TargetScan/ TargetScanS	http://genes.mit.edu/targetscan	Prediction of miRNA targets	Lewis et al. (2005, 2003)
PicTar	http://pictar.bio.nyu.edu	miRNA target prediction for vertebrates and flies	Grun et al. (2005); Krek et al. (2005)
miRanda	http://www.microma.org	Human, flies, and zebrafish miRNA target search	Enright et al. (2003); John et al. (2004)
DIANA-microT	http://www.diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi	Human, mouse, rat miRNA target scan	Kiriakidou et al. (2004)
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid	Prediction of miRNA binding sites	Rehmsmeier et al. (2004)
Tarbase	http://www.diana.pcbi.upenn.edu	List of experimentally supported miRNA targets	Sethupathy et al. (2006)
miRU	http://bioinfo3.noble.org/miRNA/miRU.htm	Plant miRNA target finder	Zhang (2005)
TargetBoost	https://demo1.interagon.com/demo	miRNA-target binding characterization	Saetrom et al. (2005)
Vienna package	http://www.tbi.univie.ac.at/~ivo/RNA	RNA secondary structure prediction and comparison	Hofacker (2003)
Mfold package	http://www.bioinfo.rpi.edu/~zukerm/rna	RNA folding and hybridization prediction	Mathews et al. (1999); Zuker (2003)

tion between the genomes of *A. thaliana* and *O. sativa*.

Wang et al. (2004b) used characteristic features of known plant miRNAs as criteria to search for miRNAs conserved between *A. thaliana* and *O. sativa*. Extensive sequence complementarity between miRNAs and their target mRNAs was used to predict miRNA-regulated transcripts.

The method of Bonnet et al. (2004) was based upon the conservation of short sequences between the genomes of *A. thaliana* and *O. sativa* as well as on properties of the secondary structure of the miRNA precursor. The authors fine-tuned their method to consider plant-specific properties, such as the variable length of the miRNA precursor sequences.

Pfeffer et al. (2004) recorded the small RNA profile of cells infected by EBV in order to probe for function of RNA silencing during infection of human cells by a DNA virus. A computational approach similar to Enright et al. (2003) was used to identify potential miRNA targets and predicted several target genes such as B-cell-specific chemokines and cytokines, transcriptional regulators, and genes involved in signal transduction pathways, cell proliferation, and death. It was suggested that miRNA silencing may be a mechanism used by EBV to control the expression of host genes.

SUMMARY AND OUTLOOK

In this review, we summarize current research efforts in computational methods for miRNA research by classifying them into two categories (miRNA gene identification and miRNA target prediction) and providing principles of computational algorithms as well as specific examples in each category. Most algorithms combine sequence, structure, and/or conservation information in order to maximize the specificity of the algorithm designed, and in silico methods for miRNA research have already become an invaluable tool that can complement biological experiments.

A summary of the online resources mentioned in this review is listed in Table 1. Alternative reviews on computational methods in miRNA research can be found in Aravin and Tuschl (2005), Bengert and Dandekar (2005), Bentwich (2005), Brown and Sanseau

(2005), Kong and Han (2005), and Sachidanandam (2005).

Clearly, as the understanding of in vivo mechanisms of miRNAs deepens, more advanced in silico modeling and discovery will be possible. For example, having more examples of miRNA-mRNA duplexes may allow us to build a generative model (e.g., hidden Markov models) (Rabiner, 1989) for miRNA target transcripts. Furthermore, if we use "negative" examples (non-functional or invalid miRNA-mRNA duplexes) in addition to "positive" examples (functional or valid miRNA-mRNA duplexes), we can use a discriminative model (e.g., support vector machines) (Vapnik, 1998) to find novel miRNA targets.

As is often the case with other biological discoveries, it is critical to biologically validate any result obtained by in silico methods for miRNA studies. Conversely, such biological validation techniques often rely on computational algorithms (Bentwich et al., 2005). Hence, iterative integration of computational and experimental methods is expected to produce an optimal framework for further deciphering biogenesis, functions, and mechanisms of miRNAs that have been unappreciated.

---

iterative integration of computational and experimental methods is expected to produce an optimal framework for further deciphering biogenesis, functions, and mechanisms of miRNAs that have been unappreciated.

REFERENCES

Adai A, Johnson C, Mlotshwa S, et al. 2005. Computational prediction of miRNAs in *Arabidopsis thaliana*. Genome Res 15:78–91.

Alpaydin E. 2004. Introduction to machine learning. Cambridge: MIT Press. pp. 17–38.

Altuvia Y, Landgraf P, Lithwick G, et al. 2005. Clustering and conservation patterns of human microRNAs. Nucleic Acids Res 33:2697–2706.

Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D. 2003. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr Biol 13:807–818.

Ambros V. 2004. The functions of animal microRNAs. Nature 431:350–355.

Aravin A, Tuschl T. 2005. Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett 579: 5830–5840.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297.

Bartel DP, Chen CZ. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet 5:396–400.

Bengert P, Dandekar T. 2005. Current efforts in the analysis of RNAi and RNAi target genes. Brief Bioinform 6:72–85.

Bentwich I. 2005. Prediction and validation of microRNAs and their targets. FEBS Lett 579:5904–5910.

Bentwich I, Avniel A, Karov Y, et al. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37:766–770.

Berezikov E, Guryev V, van de Belt J, et al. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120:21–24.

Boffelli D, McAuliffe J, Ovcharenko D, et al. 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 299:1391–1394.

Bonnet E, Wuyts J, Rouze P, Van de Peer Y. 2004. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proc Natl Acad Sci USA 101:11511–11516.

Brennecke J, Hipfner DR, Stark A, et al. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the pro-apoptotic gene hid in Drosophila. Cell 113:25–36.

Brennecke J, Stark A, Russell RB, Cohen SM. 2005. Principles of microRNA-target recognition. PLoS Biol 3:e85.

Brown JR, Sanseau P. 2005. A computational view of microRNAs and their targets. Drug Discov Today 10:595–601.

Burgler C, Macdonald PM. 2005. Prediction and verification of microRNA targets by MovingTargets, a highly adaptable prediction method. BMC Genomics 6:88.

Couturier JP, Root-Bernstein RS. 2005. HIV may produce inhibitory microRNAs (miRNAs) that block production of CD28, CD4 and some interleukins. J Theor Biol 235:169–184.

Doench JG, Sharp PA. 2004. Specificity of microRNA target selection in translational repression. Genes Dev 18:504–511.

Dsouza M, Larsen N, Overbeek R. 1997. Searching for patterns in genomic data. Trends Genet 13:497–498.

Du T, Zamore PD. 2005. microPrimer: the biogenesis and function of microRNA. Development 132:4645–4652.

Emery JF, Floyd SK, Alvarez J, et al. 2003. Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes. Curr Biol 13:1768–1774.

Enright AJ, John B, Gaul U, et al. 2003. MicroRNA targets in Drosophila. Genome Biol 5:R1.

Grad Y, Aach J, Hayes GD, et al. 2003. Computational and experimental identification of *C. elegans* microRNAs. Mol Cell 11:1253–1263.

Griffiths-Jones S. 2004. The microRNA Registry. Nucleic Acids Res 32(Database issue):D109–D111.

Griffiths-Jones S, Grocock RJ, van Dongen S, et al. 2006. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34(Database issue):D140–D144.

Grun D, Wang YL, Langenberger D, et al. 2005. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. PLoS Comput Biol 1:e13.

Hastie T, Tibshirani R, Friedman J. 2001. The elements of statistical learning. New York: Springer-Verlag. pp. 9–40.

Hobert O. 2004. Common logic of transcription factor and microRNA action. Trends Biochem Sci 29:462–468.

Hofacker IL. 2003. Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431.

Hsu PW, Huang HD, Hsu SD, et al. 2006. miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. Nucleic Acids Res 34(Database issue):D135–D139.

John B, Enright AJ, Aravin A, et al. 2004. Human microRNA targets. PLoS Biol 2:e363.

Jones-Rhoades MW, Bartel DP. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14:787–799.

Kasschau KD, Xie Z, Allen E, et al. 2003. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. Dev Cell 4:205–217.

Kidner CA, Martienssen RA. 2005. The developmental role of microRNA in plants. Curr Opin Plant Biol 8:38–44.

Kiriakidou M, Nelson PT, Kouranov A, et al. 2004. A combined computational-experimental approach predicts human microRNA targets. Genes Dev 18:1165–1178.

Kong Y, Han JH. 2005. MicroRNA: biological and computational perspective. Genomics Proteomics Bioinformatics 3:62–72.

Krek A, Grun D, Poy MN, et al. 2005. Combinatorial microRNA target predictions. Nat Genet 37:495–500.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. Science 294:853–858.

Lai EC, Tomancak P, Williams RW, Rubin GM. 2003. Computational identification of Drosophila microRNA genes. Genome Biol 4:R42.

Lai EC. 2004. Predicting and validating microRNA targets. Genome Biol 5:115.

Lai EC, Wiel C, Rubin GM. 2004. Complementary miRNA pairs suggest a regulatory role for miRNA:miRNA duplexes. RNA 10:171–175.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75:843–854.

Legendre M, Lambert A, Gautheret D. 2005. Profile-based detection of microRNA precursors in animal genomes. Bioinformatics 21:841–845.

Lewis BP, Shih IH, Jones-Rhoades MW, et al. 2003. Prediction of mammalian microRNA targets. Cell 115:787–798.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20.

Li X, Zhang YZ. 2005. Computational detection of microRNAs targeting transcription factor genes in Arabidopsis thaliana. Comput Biol Chem 29:360–367.

Lim LP, Glasner ME, Yekta S, et al. 2003a. Vertebrate microRNA genes. Science 299:1540.

Lim LP, Lau NC, Weinstein EG, et al. 2003b. The microRNAs of Caenorhabditis elegans. Genes Dev 17:991–1008.

Llave C, Xie Z, Kasschau KD, Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 297:2053–2056.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288:911–940.

Mitchell TM. 1997. Machine learning. New York: McGraw Hill. pp. 1–19.

Mourelatos Z, Dostie J, Paushkin S, et al. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev 16:720–728.

Nam JW, Shin KR, Han J, et al. 2005. Human microRNA prediction through a probabilistic co-learning model of sequence and structure. Nucleic Acids Res 33:3570–3581.

Ohler U, Yekta S, Lim LP, et al. 2004. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. RNA 10:1309–1322.

Pasquinelli AE, Reinhart BJ, Slack F, et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408:86–89.

Pfeffer S, Zavolan M, Grasser FA, et al. 2004. Identification of virus-encoded microRNAs. Science 304:734–736.

Pfeffer S, Sewer A, Lagos-Quintana M, et al. 2005. Identification of microRNAs of the herpesvirus family. Nat Methods 2:269–276.

Rabiner LR. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE Inst Electr Electron Eng 77:257–286.

Rajewsky N, Socci ND. 2004. Computational identification of microRNA targets. Dev Biol 267:529–535.

Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. Rna 10:1507–1517.

Reinhart BJ, Slack FJ, Basson M, et al. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403:901–906.

Rhoades MW, Reinhart BJ, Lim LP, et al. 2002. Prediction of plant microRNA targets. Cell 110:513–520.

Robins H, Li Y, Padgett RW. 2005. Incorporating structure to predict microRNA targets. Proc Natl Acad Sci USA 102:4006–4009.

Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. 2004. Identification of mammalian microRNA host genes and transcription units. Genome Res 14:1902–1910.

Sachidanandam R. 2005. RNAi as a bioinformatics consumer. Brief Bioinform 6:146–162.

Saetrom O, Snove O Jr, Saetrom P. 2005. Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms. RNA 11:995–1003.

Seitz H, Youngson N, Lin SP, et al. 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat Genet 34:261–262.

Sethupathy P, Corda B, Hatzigeorgiou AG. 2006. TarBase: a comprehensive database of experimentally supported animal microRNA targets. RNA 12:192–197.

Sewer A, Paul N, Landgraf P, et al. 2005. Identification of clustered microRNAs using an ab initio prediction method. BMC Bioinformatics 6:267.

Smalheiser NR, Torvik VI. 2004. A population-based statistical approach identifies parameters characteristic of human microRNA-mRNA interactions. BMC Bioinformatics 5:139.

Stark A, Brennecke J, Russell RB, Cohen SM. 2003. Identification of Drosophila microRNA targets. PLoS Biol 1:E60.

Tang G, Reinhart BJ, Bartel DP, Zamore PD. 2003. A biochemical framework for RNA silencing in plants. Genes Dev 17:49–63.

Vapnik VN. 1998. Statistical learning theory. New York: Wiley. pp. 401–570.

Vaucheret H, Vazquez F, Crete P, Bartel DP. 2004. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev 18:1187–1197.

Wang JF, Zhou H, Chen YQ, Luo QJ, Qu LH. 2004a. Identification of 20 microRNAs from Oryza sativa. Nucleic Acids Res 32:1688–1695.

Wang XJ, Reyes JL, Chua NH, Gaasterland T. 2004b. Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. Genome Biol 5:R65.

Weber MJ. 2005. New human and mouse microRNA genes found by homology search. Febs J 272:59–73.

Williams L, Carles CC, Osmont KS, Fletcher JC. 2005. A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and ARF4 genes. Proc Natl Acad Sci USA 102:9703–9708.

Xie Z, Kasschau KD, Carrington JC. 2003. Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. Curr Biol 13:784–789.

Xie X, Lu J, Kulbokas EJ, et al. 2005. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434:338–345.

Yoon S, De Micheli G. 2005. Prediction of regulatory modules comprising microRNAs and target genes. Bioinformatics 21(Suppl 2):ii93–ii100.

Zhang Y. 2005. miRU: an automated plant miRNA target prediction server. Nucleic Acids Res 33(Web Server issue):W701–W704.

Zhang BH, Pan XP, Wang QL, et al. 2005. Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15:336–360.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415.