CHAPTER FOUR

COMPUTATIONAL METHODS FOR MICRORNA TARGET PREDICTION

Yuka Watanabe,*† Masaru Tomita,*†‡ and Akio Kanai,*†‡

Contents
1. Introduction 66
2. Principles of miRNA Target Recognition 69
3. Resources for Analysis of miRNA Target Genes 73
4. Software Useful for miRNA Target Prediction 74
5. Original Strategies for Prediction of miRNA Target Genes 79
6. Validation of Computational Predictions 81
7. Concluding Remarks 82
Acknowledgments 83
References 83

Abstract
The discovery of microRNAs (miRNAs) has introduced a new paradigm into gene regulatory systems. Large numbers of miRNAs have been identified in a wide range of species, and most of them are known to downregulate translation of messenger RNAs (mRNAs) via imperfect binding of the miRNA to a specific site or sites in the 3′ untranslated region (UTR) of the mRNA. Identification of genes targeted by miRNAs is widely believed to be an important step toward understanding the role of miRNAs in gene regulatory networks. As part of the effort to understand interactions between miRNAs and their targets, computational algorithms have been developed based on observed rules for features such as the degree of hybridization between the two RNA molecules. These in silico approaches provide important tools for miRNA target detection, and together with experimental validation, help to reveal regulated targets of miRNAs. Here, we summarize the knowledge that has been accumulated about the principles of target recognition by miRNAs and the currently available computational methodologies for prediction of miRNA target genes.

* Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan
† Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa, Japan
‡ Faculty of Environment and Information Studies, Keio University, Fujisawa, Japan
1. Introduction

MicroRNAs (miRNAs) comprise a small class of noncoding RNA (ncRNA) genes identified in the genomes of various species. Discovery of novel gene regulatory systems under the control of small ncRNAs has had a significant impact on molecular biology. These RNAs are known to recognize and bind to partially complementary sites in 3′ untranslated regions (3′UTRs) of mRNAs to regulate translation of the mRNA (Bartel, 2004). Over 4000 miRNAs have been identified from a number of different species (Griffiths-Jones, 2004). In addition, it is thought that functional characterization of miRNAs will depend heavily on identification of their specific target mRNAs. However, experimental studies have touched on only a handful of the possible ranges of function of miRNAs, and numerous bioinformatic methods were developed to allow high-throughput prediction of miRNA target genes (Bentwich, 2005; Brown and Sanseau, 2005; Rajewsky, 2006; Yoon and De Micheli, 2006).

Computational prediction of miRNA target sites consists of four main steps: (1) extraction of rules related to formation of miRNA–mRNA duplexes; (2) incorporation of those rules in computational algorithms; (3) prediction of novel miRNA target sites using those algorithms; and (4) validation of the results, and thus the algorithm itself, using computational and experimental approaches. In the case of animal miRNA targets, the fact that miRNAs are very short and miRNA–mRNA duplexes are not entirely complementary made the elucidation of hybridization patterns rather challenging. Computational and experimental approaches have revealed that not only the binding pattern but also the relationship between the targeting miRNAs play an important role within miRNA target recognition (Bentwich, 2005; Brown and Sanseau, 2005; Rajewsky, 2006; Yoon and De Micheli, 2006). There are numerous useful resources and software tools available for analysis of miRNA targets, as summarized in Tables 4.1 and 4.2. Results derived using these computational algorithms have been validated biologically, and feedback from validation results have greatly improved performance of in silico miRNA target prediction algorithms.

Here, we will focus on the efforts made to predict miRNA target genes and understand the biology of miRNAs. We summarize principles of miRNA target recognition, available resources for computational prediction of miRNA target sites, and validation strategies for computational prediction. This comprehensive survey on analysis of miRNA targets is expected to provide an overall view of the knowledge that has been accumulated in the field.
<table>
<thead>
<tr>
<th>Name of the database</th>
<th>URL</th>
<th>Main features</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mirBase</td>
<td><a href="http://microrna.sanger.ac.uk/">http://microrna.sanger.ac.uk/</a></td>
<td>miRNA sequences, annotations, and computationally predicted targets</td>
<td>Griffiths-Jones et al., 2004, 2006</td>
</tr>
<tr>
<td>Argonaute</td>
<td><a href="http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface/">http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface/</a></td>
<td>Detailed information about known miRNAs and their targets</td>
<td>Shahi et al., 2006</td>
</tr>
<tr>
<td>miRNAMap</td>
<td><a href="http://mirnamap.mbc.nctu.edu.tw/">http://mirnamap.mbc.nctu.edu.tw/</a></td>
<td>Known and computationally predicted miRNAs and their targets</td>
<td>Hsu et al., 2006</td>
</tr>
<tr>
<td>Tarbase</td>
<td><a href="http://www.diana.pcbi.upenn.edu/tarbase.html">http://www.diana.pcbi.upenn.edu/tarbase.html</a></td>
<td>Experimentally verified miRNA targets</td>
<td>Sethupathy et al., 2006</td>
</tr>
<tr>
<td>Arabidopsis Small RNA Project Database</td>
<td><a href="http://asrp.cgrb.oregonstate.edu/">http://asrp.cgrb.oregonstate.edu/</a></td>
<td>Arabidopsis miRNA sequences and corresponding target genes in addition to other small RNAs</td>
<td>Gustafson et al., 2005</td>
</tr>
<tr>
<td>Name of the software</td>
<td>URL or availability</td>
<td>Supported organism(s)</td>
<td>Reference(s)</td>
</tr>
<tr>
<td>----------------------</td>
<td>--------------------</td>
<td>-----------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>TargetScan, TargetScanS</td>
<td><a href="http://genes.mit.edu/targetscan/">http://genes.mit.edu/targetscan/</a></td>
<td>Vertebrates</td>
<td>Lewis et al., 2003, 2005</td>
</tr>
<tr>
<td>DIANA-microT</td>
<td><a href="http://diana.pcbi.upenn.edu/DIANA-microT/">http://diana.pcbi.upenn.edu/DIANA-microT/</a></td>
<td>Vertebrates</td>
<td>Kiriakidou et al., 2004</td>
</tr>
<tr>
<td>RNAhybrid</td>
<td><a href="http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/">http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/</a></td>
<td>Flies</td>
<td>Rehmsmeier et al., 2004</td>
</tr>
<tr>
<td>GUUGle</td>
<td><a href="http://bibiserv.techfak.uni-bielefeld.de/guugle/">http://bibiserv.techfak.uni-bielefeld.de/guugle/</a></td>
<td>Flies</td>
<td>Gerlach et al., 2006</td>
</tr>
<tr>
<td>MovingTargets</td>
<td>Available by request on DVD</td>
<td>Flies</td>
<td>Burgler et al., 2005</td>
</tr>
<tr>
<td>FastCompare</td>
<td><a href="http://tavazoielab.princeton.edu/mirnas/">http://tavazoielab.princeton.edu/mirnas/</a></td>
<td>Nematodes, flies</td>
<td>Chan et al., 2005</td>
</tr>
<tr>
<td>miRU</td>
<td><a href="http://bioinfo3.noble.org/miRNA/miRU.htm">http://bioinfo3.noble.org/miRNA/miRU.htm</a></td>
<td>Plants</td>
<td>Zhang 2005</td>
</tr>
<tr>
<td>TargetBoost</td>
<td><a href="https://demo1.interagon.com/demo/">https://demo1.interagon.com/demo/</a></td>
<td>Nematodes, flies</td>
<td>Saetrom et al., 2006</td>
</tr>
<tr>
<td>miTarget</td>
<td><a href="http://cbit.snu.ac.kr/~miTarget/">http://cbit.snu.ac.kr/~miTarget/</a></td>
<td>Any</td>
<td>Kim et al., 2006</td>
</tr>
</tbody>
</table>
2. PRINCIPLES OF miRNA TARGET RECOGNITION

It has been challenging to predict target genes for miRNAs in animals because of the complexity of miRNA target recognition. This complexity includes the fact that miRNAs are short and that miRNA–mRNA duplexes are not entirely complementary to one another. To develop computational algorithms that identify miRNA target genes, empirical evidence is examined carefully, and principles of miRNA target recognition are extracted. For example, base pairing between miRNAs and their targets can be analyzed and checked for features, such as stable binding at 5’ end of the miRNA. Next, thermodynamic analysis of miRNA–mRNA duplexes is performed via calculation of the free energy of duplex formation and evaluation of the thermodynamics of binding. Then, cross-species sequence comparison is used to ask whether the target sequence has been evolutionarily conserved between related species. Finally, the number of target sites for the miRNA is counted, as suggested by previous studies that mRNAs are likely to be regulated by miRNA binding at more than one target site. These criteria can be applied stepwise, such as in the pipeline shown in Fig. 4.1 (Bentwich, 2005; Brown and Sanseau, 2005; Rajewsky, 2006; Yoon and De Micheli, 2006).

There is a specific base pairing pattern within miRNA–mRNA duplexes (Fig. 4.2), and it has been suggested that there is some importance to the pattern. In particular, the 5’ region of the miRNA is perfectly complementary to the corresponding sequence of target mRNA in most cases, and is generally very well conserved among paralogous miRNAs. Analysis of patterns of complementarity has been performed for vertebrates (Lewis et al., 2003), humans (Kiriakidou et al., 2004), flies (Brennecke et al., 2005; Stark et al., 2003), and worms (Lall et al., 2006; Watanabe et al., 2006). The 7 to 8 base pair sequence starting from either the first or second base of the 5’ end of the miRNA is referred to as the “seed” region of the miRNA–mRNA duplex. G:U wobble base pairing is thought to be kept at a minimum within seed regions, as weak pairing reduces silencing efficiency (Doench and Sharp, 2004). At the same time, base pairing at the 3’ region of miRNA is thought to be weaker and less important (Brennecke et al., 2005; Kiriakidou et al., 2004; Stark et al., 2003; Vella et al., 2004; Watanabe et al., 2006). However, some reports indicate that stronger binding at the 3’ region can compensate for weaker binding in the 5’ seed (Brennecke et al., 2005; Doench and Sharp, 2004). Also, there is at least one case in which the target gene is regulated independently of base pairing in the seed region (Didiano and Hobert, 2006).

In contrast to the strong binding at seed region, weaker binding is observed at the central region of the miRNA–mRNA duplex (Brennecke et al., 2005; Kiriakidou et al., 2004; Stark et al., 2003; Vella et al., 2004; Watanabe et al., 2006). Mismatches in the central region of miRNA–mRNA duplex form a bulge structure, which may be significant for regulation of target gene
Figure 4.1 A pipeline for miRNA target prediction. The main steps in identifying miRNA target genes are shown. When miRNA and mRNA (3’UTR) sequences are provided as input data sets, similar data sets from related species are constructed using data on putative orthologs. After preparation of the data sets, miRNA binding sites are identified by determining the base pairing pattern of miRNAs and mRNAs according to the complementarity within specific regions (Step 1); determining the strength of the resulting miRNA–mRNA duplex by calculation of the free energy (Step 2); comparative sequence analysis (Step 3); and checking for the presence of multiple target sites per transcript (Step 4).
expression. Taken together, these facts suggest the possibility that the mechanism of miRNA recognition differs between species or cellular conditions and also that commonalities between recognition mechanisms are not fully understood.

In most of miRNA target prediction algorithms, the thermodynamic properties of miRNA–mRNA duplex formation are analyzed by calculation of free energy ($\Delta G$), which is considered an important aspect for evaluation. The Vienna package, which can be used to estimate free energy and secondary structure, is the most commonly used software tool for thermodynamic analysis (Wuchty et al., 1999). With this tool, the free energy threshold is calculated based on both specificity and sensitivity, and as such, the values are likely to differ for different organisms. For example, free energy thresholds were set to $-14$ kcal/mol for Drosophila (D.) melanogaster (Enright et al., 2003), $-17$ kcal/mol for Homo sapiens (John et al., 2006), and $-20$ kcal/mol for Caenorhabditis (C.) elegans, (Watanabe et al., 2006). However, it is very difficult to determine the appropriate thresholds of free energy, as data sets of known miRNA–mRNA duplexes are very limited, and a lower free energy (stable binding) does not always result in reliable prediction of miRNA target genes. Therefore, it is necessary to take other characteristics into account. Toward that end, one report demonstrates a method to effectively predict miRNA targets without using thermodynamics and by using conservation analysis instead (Lewis et al., 2005).

The degree of sequence conservation is another criterion commonly used to filter miRNA targets from genome sequence. Many of the target prediction algorithms identify orthologous $3\prime$UTR sequences and then check whether the potential miRNA target site is conserved in other species, such as via comparison of relatively closely related species, including C. elegans and Caenorhabditis (C.) briggsae or D. melanogaster and Drosophila.

**Figure 4.2** A typical pattern of base pairing between miRNAs and target mRNAs. Typically, the miRNA binds to a specific site or sites within the $3\prime$UTR region of the mRNA sequence. According to thermodynamic analysis, some degree of complex formation occurs along the entire miRNA–mRNA duplexed region (A). Base pairing is particularly weak in the central region (B) and particularly strong at the $5\prime$ end (seed region) of the miRNA (C). These aspects are commonly used to identify putative novel binding sites. Base pairing between let-7 miRNA and hbl-1 mRNA in C. elegans is shown as an example (Lin et al., 2003).
pseudoobscura (Brennecke et al., 2005; Burgler and Macdonald, 2005; Enright et al., 2003; Kiriakidou et al., 2004; Rajewsky and Socci, 2004; Rehmsmeier, 2006; Stark et al., 2003; Watanabe et al., 2006). Generally, 3′ UTR sequences of different species are aligned using software tools such as BLASTn (Waterman and Eggert, 1987) or AVID (Bray et al., 2003; Couronne et al., 2003), and conservation of the sequence and/or position within 3′ UTR sequence are checked. Moreover, the University of California Santa Cruz (UCSC) genome database (Karolchik et al., 2003) provides genome-wide alignment for a wide range of species and also is commonly used for conservation analysis.

Another way of incorporating phylogeny into the extraction scheme is by comparing sequences from a wide range of species (Chan et al., 2005; Grun et al., 2005; Krek et al., 2005; Lall et al., 2006; Wang and Wang, 2006). This is rather challenging, as noncoding RNAs and 3′ UTR regions are generally less conserved than protein coding regions, and thus, there is a risk of a substantial rate of false negatives (i.e., cases in which true miRNA targets are not recognized as such). At the same time, it has also been suggested that incorporating a wide range of evolutionary conservation into the target extraction pipeline improves the signal-to-noise ratio of the extracted miRNA target candidates and allows for high accuracy of prediction of miRNA target genes (Rajewsky, 2006). Therefore, users must determine appropriate methods and thresholds according to what they wish to gain from a given analysis.

A number of studies have shown that more than one miRNA can potentially bind to a single targeted gene and that together, multiple miRNAs may cooperatively control the expression of target genes (Enright et al., 2003; John et al., 2004; Watanabe et al., 2006). Computational evidence suggests cooperation in miRNA target interactions occurs in a manner different than what can be expected by chance interactions (Stark et al., 2005). Also, some experimental evidence indicates that the degree of miRNA target regulation differs according to the number of miRNAs that bind to a target gene (Krek et al., 2005; Vella et al., 2004). This feature of multiple target sites may contribute to fine adjustment of target gene expression or may act as a back-up for essential gene regulation, possibilities that remain to be proven.

One report suggests that prediction tools can be improved by incorporating the folded structure of the mRNA into the prediction algorithm (Robins et al., 2005). The way that this group took secondary structure into account was based on their hypothesis that single-stranded miRNAs can only reach potential target sites when stretches of the target mRNA do not form base pairs with another part of mRNA. Thus, they first checked whether the mRNA is folded in a way that the site of interest is base paired with another part of mRNA by predicting the secondary structure of the 3′ UTR sequence. By taking the structure of mRNA into account, the
algorithm was reportedly able to locate target sites without relying on evolutionary conservation.

The principles previously described are useful for prediction of miRNA targets in animals. For plants, miRNA target prediction is rather straightforward, as plant miRNAs are believed to base pair to their targets with perfect or nearly perfect antisense complementarity (Llave et al., 2002). Attempts have been made to detect miRNA target genes in plants, primarily for Arabidopsis (A.) thaliana (Bonnet et al., 2004; Li and Zhang, 2005; Rhoades et al., 2002; Wang et al., 2004). Target search in A. thaliana is carried out in two main steps: (1) detection of mRNA regions that are complementary or nearly complementary to miRNA sequences; and (2) identification of orthologous sites of miRNA complementarity in related species. Because plant miRNA bind to their targets with perfect or nearly perfect base pairing, some have speculated that plant miRNAs may act similarly to small interfering RNAs (siRNAs), contributing to cleavage of target mRNA (Rhoades et al., 2002).

3. Resources for Analysis of miRNA Target Genes

As our understanding of miRNAs and their target genes has improved, some useful databases have been constructed and serve as a registry of this information (Table 4.1). In 2004, a comprehensive database of miRNA sequences was constructed, and it has been updated as new miRNAs have been identified (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006). The authors of this database also proposed a uniform definition for miRNAs (Ambros et al., 2003). As increasing numbers of miRNAs have been revealed, several databases that house information about miRNA targets have also been constructed (Gustafson et al., 2005; Hsu et al., 2006; Sethupathy et al., 2006; Shahi et al., 2006). In this section, resources which can be very useful to initiate study of miRNA targeted genes are described.

The miRNA registry, or miRBase, is a well-known and widely used database of miRNA sequences (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006). This database was initially developed to assign uniform names to miRNAs. A web interface (http://microrna.sanger.ac.uk/) allows users to search and browse for miRNAs in a wide range of species, and provides a downloadable flat file containing miRNA sequence information via the FTP site. The database is also open to submission of newly identified miRNAs. Since its initial release, the database has been updated to include novel miRNA sequences. The current release 8.2 contains 4039 entries from 45 different species. The miRBase has been expanded to include not only miRNA sequence data but also information about the potential genomic targets of miRNAs. The miRBase Target database provides predicted
miRNA target genes for various species, relying on the miRBase Sequences database as a data source for miRNA sequences and on the miRanda software as a miRNA target prediction algorithm (John et al., 2004; Stark et al., 2003).

The Argonaute database provides comprehensive and precise information about miRNAs and their target genes (Shahi et al., 2006). For example, information about expression patterns, proposed or known functions, and target genes are provided via a web interface (http://www.ma.uni-heidelberg.de/apps/zmf/argonaute/interface/). A comprehensive list of miRNA families and proteins involved in processing of miRNA are also provided. This database is currently composed of 893 miRNAs from H. sapiens, Mus (M.) musculus, and Rattus (R.) norvegicus.

The miRNAMap database includes known and putative miRNAs and their known and putative target genes, and is available at http://mirnamap.mbc.nctu.edu.tw/ (Hsu et al., 2006). This database integrates experimental evidence pertaining to miRNAs and their targets culled from existing databases and published literature. In addition, putative miRNAs are predicted using RNAz software (Washietl et al., 2005), and putative miRNA targets are predicted using miRanda software (John et al., 2004; Stark et al., 2003). Information is available at miRNAMap for four mammalian species, H. sapiens, M. musculus, R. norvegicus, and Canis familiaris.

TarBase is a manually curated collection of experimentally supported miRNA targets and can be accessed at http://www.diana.pcbi.upenn.edu/tarbase.html. This database covers a wide range of species, from plants to human, and provides information on a total of 763 target sites at 570 target genes (Fig. 4.3). Also, information about false targets (predicted miRNA-target interaction with negative experimental results) can be found in this database.

Most of the aforementioned databases have been constructed using information about animal miRNA targets. By contrast, the Arabidopsis Small RNA Project (ASRP) database provides detailed data about A. thaliana miRNAs and their targets (Gustafson et al., 2005). This database was originally constructed as a repository for small RNA sequences cloned by the ASRP but now also serves as a comprehensive database of Arabidopsis miRNAs and their experimentally validated and computationally predicted target genes. This database is available through a web interface at http://asrp.cgrb.oregonstate.edu/.

4. Software Useful for miRNA Target Prediction

As an increasing number of miRNAs from various species are identified, demand for prediction of new miRNA target sites using optimal miRNA and mRNA sequences has increased. Software with different
characteristics has been developed and applied for prediction of a large number of target genes (Brown and Sanseau, 2005; Yoon and De Micheli, 2006). In this section, we describe 11 software tools useful for miRNA target prediction and prediction results derived from use of those software tools. The algorithms we describe are listed in Table 4.2, including their sources, suitable species, and references.

TargetScan is an algorithm developed by Lewis et al. (2003) for prediction of miRNA targets in vertebrates. This software combines thermodynamics-based miRNA–mRNA duplex prediction and comparative sequence analysis. Lewis et al. observed perfect Watson–Crick complementarity at bases 2 to 8 in the 5′ end of miRNAs (the seed sequence). They incorporate this feature into their prediction algorithm by searching for seed

![Image](https://via.placeholder.com/150)

**Figure 4.3** miRNA target sites supported by experimental evidence. The number of miRNA-mRNA pairs identified either experimentally (open bars) or informatically and validated experimentally (solid bars) for *Homo sapiens* (Hsa), *Mus musculus* (Mmu), *Drosophila melanogaster* (Dme), *Caenorhabditis elegans* (Cel), and *Arabidopsis thaliana* (Ath). **A**, miRNAs that translationally repress target mRNAs. Original counts were obtained from TarBase version 2 (Sethupathy et al., 2006). **B**, miRNAs that cleave target mRNAs.
matches against mRNA sequences and extending the seed matches to predict the remaining extent of miRNA–mRNA binding. The thermodynamic properties of binding between the putative miRNA target and the extended seed sequence is calculated using the RNAfold program (Hofacker, 2003). Phylogenic analysis was carried out using genomic sequence from human, mouse, rat, and pufferfish. Eleven of 15 predicted targets generated using this software were experimentally validated. Moreover, the false-positive rate was estimated at 31% for mammalian miRNA targets, and the software was used to predict 451 potential miRNA targets. Although functions of the predicted target genes encompassed a broad range, the group was enriched for genes involved in transcriptional regulation.

Later, the authors improved the TargetScan algorithm and proposed TargetScanS, which requires a shorter seed match (six nucleotides), is independent of thermodynamic stability or multiple target sites, and requires the presence of conserved adenosine around the seed sequence (Lewis et al., 2005). The authors also added two more species to their cross-species conservation analysis, dog and chicken. These changes reduced the estimated false-positive rate of the algorithm to 22% in mammals. The algorithm successfully predicted all of the known miRNA–target interactions and in total, results in prediction of over 5300 human genes as potential targets of miRNAs. Thus, these analyses suggest that over one third of human genes are controlled by miRNAs.

The miRanda software was initially designed to predict miRNA target genes in *D. melanogaster* (Enright et al., 2003; John et al., 2004). This algorithm consists of three basic phases: (1) identification of a sequence that may be bound to miRNA; (2) calculation of the free energy for predicted miRNA–mRNA duplexes; and (3) detection of evolutionary conservation among *D. melanogaster, D. pseudoobscura, and Anopheles gambiae*. This method correctly identified 9 of 10 published miRNA-target interactions. The false-positive rate was estimated at 24%. As previously reported, the set of predicted target gene functions was enriched for transcription factors, reaffirming the possible importance of miRNA in development. This algorithm was also applied to prediction of human miRNA targets (Enright et al., 2003; John et al., 2004). About 2000 putative human miRNA target genes were identified, suggesting that 10% or more of human genes are regulated by miRNAs.

Kiriakidou et al. (2004) searched for miRNA binding properties using an experimental approach and incorporated the properties they identified into a computational algorithm, DIANA-microT. In contrast to previously discussed approaches, with this approach the authors addressed the necessity for a central bulge and strong binding at 3' end of miRNA when 5' seed pairing is rather weak. Also, in contrast of the previous works, this method uncovers predominant miRNA targets that contain only single target sites.
This algorithm successfully identified all of the documented *C. elegans* miRNA-target pairs. Moreover, seven predicted mammalian miRNA target genes were validated experimentally.

Rehmsmeier et al. (2004) presented the RNAhybrid program, which is an extension of classical RNA secondary structure prediction software tools, such as RNAfold (Hofacker, 2003) and Mfold (Mathews et al., 1999). This software searches for energetically optimal binding sites for a small RNA within a large RNA sequence. This algorithm was applied to miRNA target search in *D. melanogaster* using a six nucleotide seed match starting from the second base of the 5′ end of the miRNA. A currently released utility program called GUUGle (Gerlach and Giegerich, 2006), which can locate potential helical regions in RNA sequence, may be combined with RNAhybrid to improve the speed of search for miRNA target genes.

Cross-species comparisons provide powerful criteria for identifying miRNA target genes, and PicTar software fully relies on comparative data from several species to identify common targets for miRNAs (Krek et al., 2005). Moreover, PicTar computes the maximum likelihood that a given sequence is bound by one or more miRNAs. Target genes are first predicted using common criteria, such as optimal binding free energy, and are then tested statistically using genome-wide alignment of eight vertebrate genomes to filter out false positives. The false-positive rate for PicTar has been estimated to be about 30%, and known miRNA target sequences were identified correctly using this software. Krek et al. used this algorithm to predict vertebrate miRNA targets and suggested that on average, approximately 200 transcripts are regulated by a single miRNA. They were able to experimentally validate 7 of 13 mouse miRNA target candidates.

Using the PicTar algorithm with cross-species comparison of seven *Drosophila* species, Grun et al. (2005) were able to predict miRNA targets in *D. melanogaster*. The results suggest that on average, 54 genes are regulated by a given miRNA and that *D. melanogaster* miRNAs regulate expression of target genes coordinately. PicTar was also applied to genome-wide search of miRNA targets in *C. elegans* (Lall et al., 2006). For this study, the researchers analyzed cross-species conservation for three nematodes. The results suggested at least 10% of *C. elegans* genes are predicted miRNA targets. The authors also speculate that miRNAs regulate biological processes by targeting genes that are functionally related to one another.

The web tool MicroInspector was developed for identification of miRNA binding sites within targets from a variety of species (Rusinov et al., 2005). The simple web server allows a user to assign mRNA sequences, miRNA dataset (categorized by organism), hybridization temperature, and a free energy cut-off for calculation of possible miRNA target sites. This web tool successfully located all of the known miRNA–mRNA interactions.
MovingTargets software can be used to predict a set of miRNA targets that fulfill a set of biological constraints (Burgler and Macdonald, 2005). This algorithm applies five biological constraints: (1) the number of target sites in one mRNA; (2) thermodynamic strength of miRNA–mRNA binding; (3) the number of consecutive seed pairings at the 5′ end of the miRNA; (4) the total number of 5′ nucleotides in the miRNA involved in target base pairing; and (5) the number of G:U wobbles at the 5′ region of the miRNA. Applying this software to D. melanogaster revealed a set of 83 high-likelihood miRNA targets. Three of these candidates were tested and all three could be experimentally verified.

Chan et al. (2005) proposed a rather unique way of detecting miRNA target sites. In contrast to previous approaches, the approach used by this group relies on searching for target sequences before using miRNA sequences. Highly conserved mRNA motifs were detected via network-level conservation using FastCompare software, and these were considered potential miRNA target sites. Next, the researchers searched for preexisting miRNAs with complementarity to this set of potential targets. If existing miRNAs could not be identified, the researchers went on to search for novel miRNAs in genomic sequence. In this way, target sites were searched within D. melanogaster and D. pseudoobscura and C. elegans and C. briggsae genome pairs, and the results suggested there is likely a large number of miRNAs that have not yet been identified and validated.

As recognition mechanism seems to differ between animal and plant miRNAs, software for detecting miRNA targets from animals cannot be used to detect plant miRNAs. miRU is a web server developed specifically for plant miRNA target gene prediction (Zhang, 2005). Because plant miRNAs recognize their target mRNAs via perfect or nearly perfect base pairing, this algorithm searches for potential complementary target sites that approach perfect complementarity. This software is capable of detecting any plant miRNA target site and can search for evolutionarily conserved miRNA target sites when genome sequence is available for an appropriate comparison species. True positives and false positives can also be estimated based on the characteristics observed within predicted miRNA–mRNA duplexes.

The TargetBoost algorithm relies on training with a set of known miRNA targets, which are used to create weighted sequence motifs that capture characteristics common to validated miRNA–mRNA binding sites (Saetrom et al., 2005). The authors indicated that the weighted sequence motif approach used in TargetBoost incorporates both duplex stability and complementarity. Using a set of 36 experimentally verified target sites as the training set, the authors of TargetBoost were able to predict potential miRNA target sites for genes involved in D. melanogaster body patterning.

Miranda et al. (2006) introduced a pattern-based method for prediction of miRNA target sites. This software, rna22, first searches for putative
miRNA target sites and then identifies possible targeting miRNAs. The rna22 program was used to successfully predict miRNA target sites without the use of cross-species validation. Subsequently, 226 predicted target genes were tested in a luciferase reporter gene assay, and for 168 of them, miRNA-dependent repression was observed. They suggest that some miRNAs may regulate as many as a few thousand target genes and up to 74 to 92% of all transcripts in some species.

A support vector machine (SVM) classifier has also been applied to miRNA target prediction in the form of the miTarget software tool (Kim et al., 2006). miTarget incorporates structural, thermodynamic, and positional information as inputs for SVM-based analysis. This method was reported to predict a biologically relevant set of miRNA targets with higher performance than previously published tools according to the receiver operating characteristic (ROC) curve analysis. The authors also indicate that base pairing at positions four, five, and six in the 5′ region of the miRNA is particularly important.

5. Original Strategies for Prediction of miRNA Target Genes

A number of reports present specific pipelines for extracting miRNA target sites. Those methods are not incorporated as software, but their methodologies and prediction results give a good idea of how and what kinds of genes are regulated by miRNAs. Also, in some of these reports, miRNA target prediction is performed for species such as plants and viruses, whose miRNA target recognition mechanisms may differ greatly from that of animals.

Stark et al. (2003) first characterized hybridization patterns of known miRNA–mRNA duplexes in D. melanogaster and used the information to extract target sites with conserved 8mer nucleotides at the 5′ end of the miRNAs in addition to taking into account evolutionary conservation with a closely related species, D. pseudoobscura. This group used the MFold package (Mathews et al., 1999) to calculate the thermodynamic stability of miRNA–mRNA hybridization. Known miRNA target genes scored high with their approach. In addition, the importance of multiple miRNA target sites per target gene was also suggested by their results. Six of the predicted miRNA target candidates were tested and validated.

Brennecke et al. (2003) used a reporter gene assay to extract hybridization characteristics of miRNAs and their targets for further computational analysis. From their experimental analysis, these researchers found that target sites can be grouped into two types: target sites with high complementarity at the 5′ end of the miRNA, and target sites with 3′ base pairing and a low degree of complementarity at the 5′ end. These characteristics
were incorporated in a computational algorithm and applied to genome-wide extraction of *D. melanogaster* miRNA targets. The analysis provides evidence that an average miRNA has approximately 100 target sites and that a large number of cellular genes are under the control of miRNAs.

Systematic identification of human miRNA target genes was carried out by combining an *in silico* miRNA target prediction method with microarray analysis (Wang and Wang, 2006). This group first predicted miRNA targets in the human genome using an algorithm that combines relevant parameters for miRNA target recognition and calculates a score according to the different weights assigned to each of these parameters. Predicted target genes for miR-124 were systematically verified using microarray analysis of target genes in the context of overexpression of miR-124. From the analysis, significant down regulation of cell cycle–related genes was observed.

To elucidate target sites within the *C. elegans* genome, a target extraction pipeline was constructed according to the characteristics of known miRNA–mRNA duplexes of *C. elegans* (Watanabe et al., 2006). Extraction of miRNA target candidates was performed in three steps: overall binding of miRNA and mRNA was predicted using the RNAhybrid software (Rehmsmeier et al., 2004); hybridization patterns were selected and conservation between *C. elegans*, and a related nematode, *C. briggsae*, was taken into account; and finally, the number of target sites within a single target gene was determined. Free energy was calculated for every dinucleotide of the miRNA–mRNA duplex to analyze the hybridization pattern more precisely than that achieved via reliance on analysis of simple complementarity. This method predicted 687 potential target genes in *C. elegans*, many of which are classified as genes involved in development.

Most of the miRNA target prediction algorithms have been constructed for finding miRNA targets in animals. This is likely to be the results of the observation that base pairing and other attributes of miRNA-target interactions are complicated and not uniform, such that different rules can be applied within a species, and in addition, different rules may apply to different species. In contrast, miRNA–target interactions in plants appear to more heavily rely on base pairing. Although there are fewer algorithms available, a number of plant miRNA targets have been predicted as they can rely on straightforward base pairing searches and conservation analyses. For example, Rhoades et al. (2002) predicted regulatory targets for 14 *A. thaliana* miRNAs, and extracted 49 candidate miRNA targets. Thirty-four of these miRNA target candidates are members of transcription factor gene families, suggesting the possible contribution of miRNA to transcription control mechanisms, similar to what was observed in animals. Also, Jones-Rhoades and Bartel (2004) developed a comparative genomic algorithm that can systematically identify miRNAs and their targets. Nineteen novel miRNA target candidates, conserved in both *A. thaliana* and
*Oryza sativa*, are biologically relevant. Targets predicted using this algorithm also tend to be genes involved in development.

Virus-encoded miRNA has been identified ([Pfeffer et al., 2004](#)), and possible targets for viral miRNAs were predicted using the miRanda algorithm ([Enright et al., 2003](#)). Another report suggested the existence of a novel type of miRNA which functions in mammalian antiviral immunity ([Lecellier et al., 2005](#)), indicating that there may be a complicated miRNA-mediated interaction between viruses and their hosts. Computational methods for identification of viral miRNA targets have yet to be developed, likely due to a lack of experimentally analyzed examples of interaction between viral miRNAs and their targets. However, there is a report of computational search for cellular targets carried out using predicted HIV-I encoded miRNAs ([Bennasser et al., 2004](#)). Additionally, a search for human miRNA targets in HIV-1 genes has also been reported ([Hariharan et al., 2005](#)). In that study, the authors used four well-established target-finding software tools in their analysis, and some known features (e.g., seed matching and cross-species comparison) were also taken into account in their extraction of possible target sites. Crucial HIV-1 genes, including the *nef* gene, were predicted to be possible miRNA target genes. There remains much to be learned about the mechanisms of miRNA regulation as it relates to viruses, and it is very difficult to carry out target prediction based solely on the existing knowledge. Nonetheless, computational target prediction may be able to contribute greatly to the understanding of miRNA-mediated interactions between viruses and their hosts in the future.

6. **Validation of Computational Predictions**

Validation of computational prediction algorithms and predicted miRNA targets is crucial for understanding the biological significance of prediction results. Moreover, feedback of validation results should be useful for the further analysis of predicted targets and evaluation of computational algorithms. There are two main strategies used for validation of miRNA target prediction software: evaluation of known miRNA target genes, and calculation of signal-to-noise ratio or false-positive rates using negative control tests (i.e., evaluation using artificial miRNA-like sequences).

Most miRNA prediction algorithms report the significance of their method by showing how well the algorithm performs in terms of correctly identifying known miRNA–mRNA interactions. Thus, a limitation of this method of reporting the accuracy of predictions would be the lack of an available positive data set. Only a small number of miRNA target interactions have been experimentally analyzed at this point (see Fig. 4.3); therefore, it is difficult to determine statistical significance. Moreover, training
sets used in the development of algorithms are usually extracted from this small dataset of validated targets, making the dataset that can subsequently be used for evaluation even smaller. Saetrom et al. (2005) used a method similar to cross-validation to help overcome these problems. With their method, accuracy of the algorithm was determined using a dataset of known miRNA–mRNA duplexes.

Another strategy commonly used for evaluation of miRNA target prediction is comparing prediction results calculated from real and artificial miRNA data inputs. Artificial miRNAs are shuffled sequences that resemble real miRNAs in terms of features such as base composition, sequence length, and frequency of appearance in the genome. By comparing the prediction results derived from real and artificial miRNAs and determining how the results differ, a signal-to-noise ratio or false-positive rate can be estimated.

Experimental validation of miRNA target interactions is crucial to detecting novel miRNA targets, as computational methods are not perfect, and there is a risk of false-positive prediction. Although experimental validation of miRNA target genes is challenging compared to computational validation, there are more and more examples of miRNA target genes from various species that were identified using combined computational and biological approaches (see Fig. 4.3). Experimental validation is performed against two types of predicted miRNA targets that have different regulatory mechanisms: translational repression of target mRNAs (see Fig. 4.3A) and cleavage of target mRNAs (see Fig. 4.3B). Methods such as reporter-gene assays (Burgler and Macdonald, 2005; Kiriakidou et al., 2004; Krek et al., 2005; Miranda et al., 2006; Robins et al., 2005), gene mutation (Brennecke et al., 2005; Stark et al., 2003, 2005), and 5’ RACE (Jones-Rhoades and Bartel, 2004; Wang et al., 2004) have commonly been used for experimental validation of computationally predicated miRNA targets that are translationally regulated. In addition, microarray analysis provides a powerful and high-throughput method for observing cleaved target mRNAs (Wang and Wang, 2006). For example, this experimental approach was used to identify a large number of human miRNA targets that appear to be cleaved by miRNAs (see Fig. 4.3B; Lim et al., 2005).

7. Concluding Remarks

Prediction and identification of miRNA target genes should be the first steps toward understanding the biology of miRNAs. Computational and experimental approaches have revealed numerous characteristics of miRNA-target interactions and have led to effective prediction of target sites. The accumulated knowledge about miRNA target recognition has revealed the principles of miRNA–mRNA duplex formation, such as
strong base pairing at the 5’ seed region of the miRNA. However, there are some exceptions to these generalized rules, and it is also true that target selection mechanisms differ among species. To improve the accuracy of miRNA target prediction, phylogenetic analysis has been applied and has been used to successfully predict a number of target genes. Moreover, experimental validation, including feedback of validation data to computational tools, has also played an important role in identification of computationally predicted miRNA target genes.

Combining experimental analysis with computational approaches should be very effective in developing methods to predict miRNA target genes with even more accuracy than what is currently possible. A number of groups have reported expression profile data for miRNAs in different tissues or developmental stages (Aravin et al., 2003; Barad et al., 2004; Baskerville and Bartel, 2005; Chen et al., 2005). These data may be combined with expression profile data from targeted mRNAs to detect potential in vivo interactions between miRNAs and their target genes. High-throughput experimental strategies such as microarray analysis have been developed and applied to miRNA target analysis, raising expectations for large-scale analysis of miRNA target sites (Barad et al., 2004; Baskerville and Bartel, 2005; Wang and Wang, 2006). Additionally, “knockdown” or “knockout” genetic tests of miRNAs or their target genes may be crucial for gaining a better understanding of the correlation between the function of miRNAs and specific phenomenon.

Once miRNA targets can be predicted with a fair degree of accuracy, the next step may be to work to understand the roles of miRNAs in living systems. To do this, gaining a comprehensive view of gene expression systems may be crucial. Computational approaches, which excel in the handling of genomic, transcriptomic, and proteomic data, should provide invaluable tools for identification of the relative position of miRNAs in various biological networks. Analysis of miRNAs and their target genes in the context of additional functional genomic data is expected to shed light on the potentially diverse and important biological functions of miRNAs within living systems.

ACKNOWLEDGMENTS

The authors acknowledge Nozomu Yachie, Koji Numata, and Dr. Rintaro Saito of Keio University for their helpful discussions.

REFERENCES


