REVIEW

Computational methods for the identification of microRNA targets

Yang Dai[1]
Xiaofeng Zhou[2]

[1]Department of Bioengineering, Department of Computer Science, College of Engineering, [2]Center for Molecular Biology of Oral Diseases, College of Dentistry, and Graduate College, UIC Cancer Center, University of Illinois at Chicago, Chicago, IL, USA

Correspondence: Yang Dai
Department of Bioengineering (MC063),
University of Illinois at Chicago,
851 South Morgan Street, Chicago,
IL 60607, USA
Tel +1 312 413 1487
Fax +1 312 996 5921
Email yangdai@uic.edu
or
Xiaofeng Zhou
Center for Molecular Biology of Oral
Diseases, College of Dentistry, University
of Illinois at Chicago, 801 South Paulina
Street, Chicago, IL, USA
Tel +1 312 355 0825
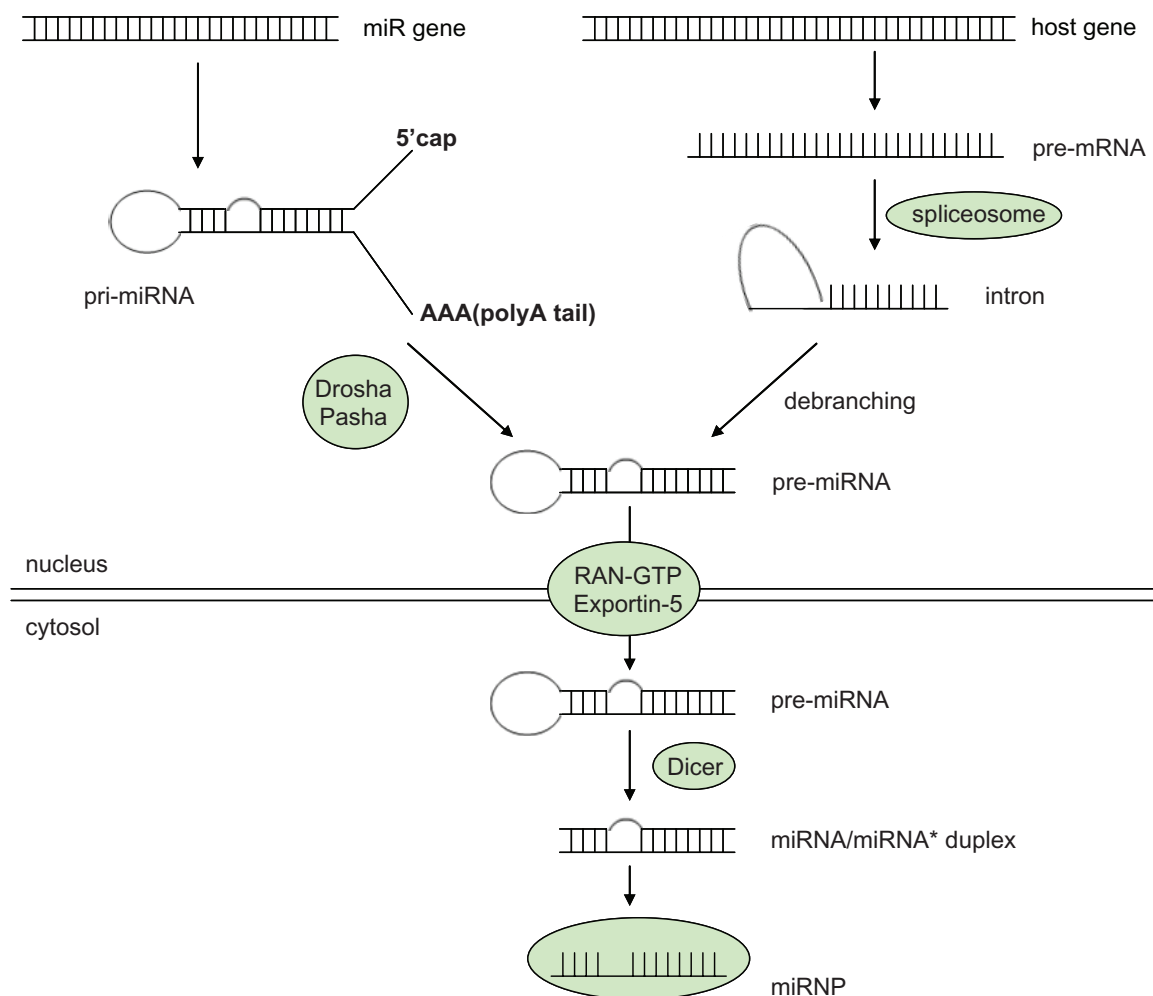Fax +1 312 413 1604
Email xfzhou@uic.edu

Abstract: MicroRNAs are pivotal regulators of development and cellular homeostasis. They act as post-transcriptional regulators, which control the stability and translation efficiency of their target mRNAs. The prediction of microRNA targets and detection of microRNA-mRNA regulatory modules (MRMs) are crucial components for understanding of microRNA functions. Numerous computational methods for microRNA target prediction have been developed. Computationally-predicted targets have been recently used in the integrative analysis of microRNA and mRNA expression analysis to identify microRNA targets and MRMs. In this article we review these recent developments in the integrative analysis methods. We also discuss the remaining challenges and our insights on future directions.

Keywords: microRNA target prediction, integrative analysis, microRNA regulatory mechanism, microRNA profiling, mRNA expression profiles

MicroRNAs (or miRNAs) are small endogenously expressed noncoding RNA molecules of 20–24 nucleotides in length, which regulate target gene expression at post-transcriptional levels (eg, regulation of the mRNA stability and the translation efficiency). The microRNA biogenesis has been well characterized (Figure 1). The genes encoding microRNAs are much longer than the processed mature microRNA molecule. MicroRNAs are first transcribed as primary transcripts (pri-miRNA) with a cap and poly-A tail. They are then processed into short, 70-nucleotide stem-loop structures known as pre-miRNA in the cell nucleus. In animals, this processing is performed by a protein complex consisting of the nuclease Drosha and the double-stranded RNA binding protein Pasha. These pre-miRNAs are then processed to mature microRNAs in the cytoplasm by interaction with the endonuclease Dicer, which also initiates the formation of the RNA-induced silencing complex (RISC). This complex is responsible for the gene silencing that is induced by both microRNA and RNA interference. In plants, the pathway varies slightly due to their lack of Drosha homologs; instead, Dicer homologs alone are responsible for both pri-miRNA and pre-miRNA processing steps. The pathway is also different for microRNAs derived from intronic stem-loops; these are processed by Dicer. No Drosha activity is needed for their maturation. For detailed discussion on microRNA biogenesis, we refer you to earlier reviews.[1,2]

MicroRNAs represent approximately 1%–2% of the known eukaryotic genomes. Although microRNAs account for a minor fraction of the expressed genome, microRNAs are pivotal regulators of development and cellular homeostasis through their control of diverse cellular processes. MicroRNAs constitute an important class

29

Figure 1 MicroRNA biogenesis.

of fine-tuning regulators of gene expression[3] and have been referred to as "dimmer switches" because of their ability to repress gene expression without completely silencing it. MicroRNAs are related to, but distinct from, short interfering RNAs (siRNAs). A key difference between siRNA and microRNA is that siRNA requires almost complete complementary to its targeting sequence for it to exert the silencing function, whereas a microRNA usually binds to its target genes through partial complementary binding. Because of this unique feature, a single microRNA has multiple target genes and, thus, could regulate a large number of protein-coding genes. This may explain why microRNAs play a fundamental role in regulation of diverse cellular processes. Increasing efforts to identify specific targets of microRNAs have led to the speculation that microRNAs may regulate at least 30% of human protein encoding genes.[4–6]

In a number of studies attempts have been made to analyze microRNA function by validating microRNA-mRNA interactions with direct and indirect experimental methods. These experimental endeavors have drastically increased the number of validated microRNA target genes; which are now collected in microRNA databases, such as TarBase,[7,8] and the component of miRecords called validated targets.[9] For example, the currently version of TarBase includes more than 1300 microRNA–mRNA interactions in several species.

Indirect methods for microRNA target discovery are based on high-throughput experimental technologies, such as microarrays and protein quantification experiments. The microarray technology has reached a certain level of maturity, therefore systematic identification of the functional microRNA targets through large scale mRNA expression profiling has been widely used in recent studies.[10–16] Two types of experiments are the most commonly used: 1) differential mRNA profiling on a microRNA transfected cell line and its negative control, and 2) simultaneous microRNA and mRNA expression profiling on samples of

different phenotypes using microarrays. The first approach involves the investigation of functional microRNA targets when a specific microRNA is over- or underexpressed.[12,17–19] The second approach is aimed at discovering microRNAs with altered expression related to different phenotypes and uncovering their targets mRNAs.

Despite the significant increase in the number of experimentally validated microRNA-mRNA regulatory relationships, the majority of the microRNA targeted genes remains unknown. Computational prediction provides a rapid method to identify putative microRNA targets, and, complemented by biological validation techniques, this has been the major approach to microRNA target identification. For indirect experimental methods the intelligent use of predicted targets is a vital step to differentiate the putative microRNA targets from those differently expressed mRNAs that are caused by biological regulations other than microRNAs. A variety of computational algorithms for the prediction of microRNAs and target genes have been proposed and implemented based on the sequence and structure information of microRNAs and target genes, for example, miRBase,[20,21] PicTar,[22] RNAHybrid,[23] TargetScan,[4,24,25] miRanda,[26] and DIANA-microT.[27] In several recent articles excellent surveys on computational and experimental approaches to the identification of microRNA genes and their targets have been provided.[5,6,25,28–33] Computational analysis indicates that each microRNA can target more than 200 transcripts and that a single mRNA may be regulated by multiple miRNAs.[5,34] This implies that microRNAs and their targets are part of the complex regulatory network. The identification of a set of microRNAs and their target mRNAs that work in a collaborative manner is important for the understanding of microRNA cellular functions.

In this article, we will review recent developments in this area. We will focus on integrative computational approaches to identify microRNA targets and microRNA–mRNA regulatory modules (MRMs) based on the use of available information on predicted targets, microRNA and mRNA expression profiles, and functional annotations on targets.

Integrated resources on experimentally validated and predicted microRNA targets

Since the number of experimentally-validated microRNA targets is still small, the investigation of novel microRNA targets requires computational predictions. Recently released microRNA-related database resources offer not only the confirmed targets but also the predicted targets collectively from multiple programs. TarBase[7,8] is a comprehensive database exclusively for experimentally-validated animal microRNA targets as described previously. miRecords,[9] on the other hand, hosts two components. The validated targets component in this database provides a large, high-quality database of experimentally-validated microRNA targets produced from meticulous literature curation. This component of the database contains 1,597 records of interactions between 429 microRNAs and 1,143 target genes in nine animal species as of September 4, 2009. Among these records, 916 were curated from low throughput experiments. These databases are an indispensible resource for guiding function studies on microRNAs and their target genes.

The existence of multiple prediction tools presents a challenge for choosing one method over the other, with each method having different levels of coverage and false positive prediction. The evaluation of precision and coverage of the developed tools has proven difficult due to the small number of validated microRNA targets. In an early study by Sethupathy et al[35] the authors carried out an analysis of the performance of five individual programs (DIANA-microT, miRanda, TargetScan, TargetScanS, and PicTar) and various combinations of these programs. They compiled several benchmark data sets of experimentally-supported microRNA-target gene interactions. It was observed that the earliest developed programs (TargetScan and DIANA-microT) achieve a relatively low sensitivity on the benchmark data sets (less than 7% on conserved unbiased) and also predicted a relatively small number of total microRNA-target gene interactions. However, the next program developed (miRanda) picks up nearly 65% of conserved unbiased experimentally supported interactions. It also exhibits a substantial increase in the number of total predictions. It was also shown that the most recently developed programs (TargetScanS and PicTar) demonstrate almost identical sensitivity to miRanda but predict several thousand fewer microRNA-target gene interactions. Different levels of reduction in the sensitivity were observed when intersections of different programs were used. Based on these observations, Sethupathy et al[35] provided a guideline for microRNA target prediction using a stepwise approach and the available bioinformatic tools (DIANA-microT, miRanda, TargetScan, TargetScanS, and PicTar) at the time of their analysis (2006).

In a recent study published by Alexiou et al[32] an excellent assessment of the performance of the latest versions of various tools for target prediction is provided. It was demonstrated that a group of methods (DIANA-MicroT 3.0, TargetScan 5.0, TargetScanS, PicTar, and EIMMo[36]) has a precision of approximately 50% with a sensitivity that ranges

from 6%–12%. Interestingly, these methods are designed based on the concept of evolutionary conservation of the seed region or some small extension of this region. These findings imply that taking the intersection of all predicted targets may result in a set of mRNA targets with a higher level of specificity, but with a limited degree of sensitivity. On the other hand, taking the union of all predicted targets can produce a set which includes many false positives despite increasing the level of sensitivity.[35] In the following discussion, however, we will list a number of web resources that collectively store the predictions from various programs. The users should be aware of the above issue and carefully choose which prediction method to use.

Below, we list a number of web resources that collectively store the predictions from various programs. The first web resource, MiRGen,[37] is an integrated database of 1) positional relationships between animal microRNAs and genomic annotation sets, and 2) animal microRNA targets according to combinations of four widely used target prediction programs (DIANA-microT, miRanda, PicTar, and TargetScanS). The target prediction offers various options on the choice of prediction programs. For example, one could choose the use of any single program or unions or intersections of targets predicted from the tools and experimentally validated targets included in TarBase.

Two other web resources are miRNAMap 2.0[38] and miRecords.[9] These resources consolidate experimentally confirmed and predicted microRNA targets for multiple species. miRNAMap 2.0 provides information on microRNAs collected in miRBase,[21] and literature as well as mRNA targets predicted from miRanda, RNAHybrid, and TargetScan with additional filtering criteria for false positive reduction. These criteria are: 1) it was predicted by at least two tools, 2) target genes were found to contain multiple sites, and 3) the target site was found to be accessible. The database currently includes information on 542 human microRNAs (445 experimentally confirmed and 97 predicted) and 6,750 predicted targets that satisfy the filtering criteria. miRecords includes a component of validated targets, as mentioned above, and a component of predicted targets. The latter component provides a collection of predicted microRNA targets obtained from 11 established microRNA target prediction programs: DIANA-microT, MicroInspector,[39] miRanda, miRDB,[14,40] miTarget,[41] NBmiRTar,[42] PicTar, PITA,[43] RNA22,[44] RNAhybrid, and TargertScanS.

Some other web resources also provide additional features for functional analysis, such as enrichment analysis of gene ontology[45] (GO) annotations and pathways, on the predicted microRNA targets. miRGator[46] integrates target prediction, functional analysis, expression data, and genome annotation. It produces the predicted targets by combining information on the validated targets from miRBase and the predicted targets from miRanda, PicTar, and TargetScanS. The predicted targets can be further analyzed for functional enrichment with terms in GO and pathways in *Kyoto Encyclopedia of Genes and Genomes* (KEGG),[47] ENMAPP,[45] BioCarta.[48] A unique feature provided by this tool is a compendium of microRNA-related expression profiles that currently covers 12 datasets, 106 comparison studies, and 566 samples, most of those being human cases based on the data from the Gene Expression Omnibus (GEO) database.[49]

Using public databases of the predicted microRNA targets, SigTerms[50] provides a convenient pipeline supporting the integrative analysis on a large scale. This tool produced a list of genes by retrieving all microRNA–mRNA functional pairs in a given database. An enrichment statistic for over-representation of predicted targets in the gene set is also provided. GO annotation term enrichment analysis is provided in order to evaluate whether a given microRNA targets a higher number of genes within the selected gene set than would be expected in a random gene set. The tool, which is a set of excel macros, currently takes the entire set of microRNA predictions from the databases of PicTar, TargetScan, and miRanda. The tool, however, does not take the mRNA expression into account.

Another tool, CORNA,[51] implemented as a R package, offers slightly different methods for the prediction of microRNA-target mRNA module. The unique features of CORNA includes: 1) a query function if a list of mRNAs is enriched by the targets of particular microRNAs using a hypergeometric test and 2) the ability to identify statistically associated GO terms and KEGG pathways based on the predicted targets of microRNAs. In addition, CORNA has a function that directly reads microRNA-target data from various resources such as miRBase and microRNA. org.[26]

A different option for prediction programs was presented by Peng et al.[52] These authors proposed a strategy for computational identification of hepatitis C virus-associated MRMs in human livers using microRNA and mRNA expression profiles obtained from the same samples. It was suggested that the use of predicted targets based on the seed matches alone may be sufficient. They demonstrated that filtering with inverse-expressed relationships can largely reduce the number of targets (approximately 17,000 genes). After the filter was applied, the expression profiles of the remaining targets showed strong correlations with that of microRNAs.

Another interesting way to generated a more reliable list of targets from several ranked lists of the predicted targets was suggested using a global optimization technique: the cross entropy Monte Carlo method.[53]

Computational approaches to identify functional microRNA targets and MRMs using mRNA expression profiles

In order to minimize false positives and to detect the functional microRNA targets under a specific biological condition, recently developed methods have integrated analysis of expression profiles of microRNA and mRNA in conjunction with the predicted microRNA targets. Most of the integrative methods are based on the simple principle that inverse relationships in their expression profiles should be held between a specific microRNA and its functional mRNA targets. However, this approach is only sensitive to the mRNA targets that are regulated by microRNA mediated degradation, and not sensitive to the targets that are regulated by microRNA mediated translational inhibition. Therefore, this approach will supplement the sequence level analysis, not replace it. Researchers have revealed that microRNAs directly repress translation of hundreds of genes after over-expression or under-expression of a microRNA.[54,55]

MMIA[56] integrates microRNA and mRNA expression data with predicted microRNA target information (obtained from TargetScan, PicTar, PITA) to analyze microRNA-associated phenotypes and biological functions by Gene Set Enrichment Analysis (GSEA). To assign biological relevance to the integrated microRNA/mRNA profiles, MMIA uses exhaustive human genome coverage (5,782 gene sets), including various disease-associated genes as well as conventional canonical pathways and GO. This web server provides users with combined analysis tools to identify the differentially expressed microRNAs, mRNAs, and significantly inversely correlated microRNA-mRNA pairs based on the microRNA and mRNA expression data. The biological pathways that are associated with the subset of targets (intersection of the predicted targets and inversely correlated targets) are provided through the GSEA using various predefined gene set databases, including the KEGG, MIT MSigDB,[57] and G2D databases.[58] The tool also reports information on the transcription binding sites in the primary transcript regions of the upregulated microRNAs. In addition, it also reports diseases associated with the upregulated microRNAs based on the miR2Disease[59] database.

Joung and Fei[60] proposed a two-step strategy to identify microRNA targets from mRNA profiling datasets. For those mRNAs with differential expression, the method makes an initial classification of putative target mRNAs and nontarget mRNAs based on a classification model, learned from a support vector machine (SVM).[61] For their study, these authors used the microRNA and mRNA expression profiling datasets of *Arabidopsis* across different conditions. The final decision on the putative target mRNAs is made using the SVM prediction and the alignment scores derived from the sequences of the pair of predicted microRNA and the mRNA. More specifically, using the collection of *Arabidopsis* microRNAs in the miRBase[21] and mRNAs sequences from the TAIR database,[62] a microRNA is aligned against all mRNA sequences, and the weighted summation of the number of mismatches is defined as the score between a microRNA and mRNA sequences.

While early methods for prediction of MRMs did not effectively incorporate the mRNA information, Yoon and De Micheli[63] proposed the first method that explicitly searches for the "many-to-many" relationships among microRNAs and the target genes. They model the relationships of microRNAs and their mRNA targets by a weighted bipartite graph. Candidate MRMs are represented as bicliques on the bipartite graph. In a biclique, every node in the first set (microRNAs) is connected to every node in the second set (mRNAs). Their novel algorithm looks for a microRNA seeding sequence or a set of microRNAs that bind a common target with similar binding strength and identifies MRMs by merging the seeds and the statistically significant analysis. In particular, they used the prediction methods described by Lewis et al[64] and John et al[65] in order to quantify the strength of microRNA-target binding, although other methods that produce predicted targets along with the binding strength (such as the local alignment score and free energy of a microRNA-target duplexes) are also applicable.

GeneSet2miRNA[66] identifies a set of microRNA regulators for a list of mRNAs in which the predicted targets of the set of microRNAs are mostly over-represented. The target mRNAs predicted from the predicted targets component of miRecords (from 11 prediction tools) were arranged based on confidence levels: the more the tools predicted the interaction, the better the confidence. They used two specific cut-off values; the genes included in the final candidate list have to be predicted by at least four or five tools, respectively. The tool accepts a query list of genes and, for each regulatory model, the number of genes in the list is counted. Then the significance of this number is tested using the following null hypothesis: the

gene list is independent of the candidate target list by using hypergeometric tests with further adjustment of multiple testing. This method, however, can only detect modules that include up to four microRNAs, since it exhaustively considers all modules consisting of three or four microRNAs. A greedy heuristic is proposed for the reduction of search time without guaranteeing the optimal solution.

Joung et al improved the method developed by Yoon and De Micheli[63] by including additional information on microRNA and mRNA expression profiles.[67] They proposed a population-based probabilistic learning model to identify synergistic microRNAs involved in the regulation of their condition-specific targets. More specifically, the improved method employs a framework of the genetic algorithm searching for a subset of microRNAs and a subset from their predicted target mRNAs with a best fitness score. The fitness score is a balanced aggregation of binding strength of microRNAs and its targets, expression coherence score of microRNAs in the microRNA subset, as well as that of mRNAs in the selected mRNA subset. The coherence score is defined as the mean of the Pearson correlation coefficient (PCC) between all microRNA–mRNA pairs. A unique technique based on co-evolutionary learning is used to maintain two respective populations of microRNAs and mRNAs and to ensure that they evolve cooperatively in the context of each other. The new populations are generated by estimation-of-distribution algorithms.[68,69] The method was evaluated using a data set found in the study by Lu et al.[70] This method, however, encounters a problem when identifying functional MRMs over specific conditions, as it is based on the expression correlation over all available conditions instead of a subset of conditions. Moreover, the method requires the expression profiles of both microRNAs and their targets under a set of common conditions, which have not always been met in various studies.

GenMiR++[71] uses a Bayesian learning algorithm to obtain a set of functional microRNA targets with high confidence based on the use of predicted microRNA targets and expression profiles of microRNA and mRNAs. The Bayesian learning method not only models mRNA expression using microRNA expression data, but also infers the complex nature of the microRNA regulation. The method searches for functional interactions by detecting negative correlations from the predicted target mRNA set while accounting for differences in regulatory power, hybridization conditions, and normalizations among datasets. The method detects not only functional targets but also groups of microRNAs and their target mRNAs that involve in similar biological processes, ie, microRNA-mRNA regulatory modules.

The rule-based learning method for identification of MRMs proposed by Tran et al[72] is based on the assumption that genes regulated by the same microRNAs should show similar expression profiles. Their method first uses PicTar for microRNA target prediction. The results are organized in a table with 0 and 1 elements in a row indicating an mRNA and a column indicating a microRNA. The value 1 implies that the mRNA is the predicted target of the microRNA and 0 otherwise. The PCCs are then calculated between a microRNA and all mRNAs using their expression data and an mRNA is labeled as either *similarity* or *dissimilarity* according to the PCC value. This information is appended as a class-column to the table. Finally, the CNS-SD rule induction system,[73] a subgroup discovery method, is used to produce a set of microRNA-mRNA regulatory modules. Further filtering procedures are needed to select those rules that only include microRNAs with highly correlated expression profiles. They evaluated the method using the same datasets as in the study by Joung et al.[67] They analyzed human genes and microRNAs and identified 79 MRMs, in which high correlation in expression patterns of microRNAs as well as mRNAs are present. They also revealed that the mRNAs included in the same modules shared similar biological functions.

To infer MRMs, Peng et al[52] used simultaneously microarray profiled expression of cellular microRNAs and mRNAs across 30 hepatitis C virus (HCV) positive or negative human liver biopsy samples to construct a microRNA-mRNA regulatory network. This network is a bipartite graph where microRNAs and mRNA targets are disjointed sets of nodes. A connection is made between a pair of microRNA and predicted target mRNA if there is a significant inverse correlation between the expression profiles of the pair. A biclique is considered as a candidate regulatory module. The method systematically assesses the statistical significance (the probability of finding it by chance and differential expression of target mRNA between HCV positive and HCV negative samples) of all maximal bicliques. These significant maximal bicliques are further evaluated by the functional enrichment analysis of GO biological category and KEGG pathways. As we mentioned before, the authors proposed that the use of the seed match between a microRNA and its target is sufficient for target prediction, as the reverse correlation between microRNAs and mRNAs can filter out a large number of mRNAs. This finding indicates that the integrated use (eg, intersection) of predicted mRNA targets from multiple programs may limit valid, potential microRNA targets.

Similar to the previous approach, Liu et al[74] proposed a putative microRNA regulatory network based on a bipartite graph, where a connection is made between a specific microRNA and its predicted target without the requirement of the inverse relationship between their expression profiles. The functional MRMs are defined as maximal bicliques where the involved microRNAs and mRNAs are in reverse correlation for their expression profiles. To discover those functional MRMs, a data mining technique called association rule mining is used to identify frequent microRNAs and mRNAs with reverse patterns on the expression profiles.

An approach that attempts to identify an extended regulatory mechanism was proposed in Joung and Fei.[75] They specifically defined the microRNA regulatory modules as functional clusters of a set of microRNAs with their target mRNAs that are involved in the same biological processes. In addition, the microRNA regulatory modules can contain the regulators of microRNAs. Their method integrates data from four diverse resources: 1) large-scale mRNA expression dataset, 2) microRNAs and their target mRNAs, 3) known transcription factor-binding elements, and 4) promoter sequences of microRNAs. An elaborate probabilistic graphical model-based method for prediction of such functional MRMs was proposed. The method was evaluated for microRNAs in *Abrabidopsis*. Specifically, the 637 microRNA-target pairs for 137 microRNAs and 382 mRNAs in *Abrabidopsis* were collected from several genome-wide screens, computational prediction, and curated microRNA target databases. A total of 95 promoter sequences which are 300–350 bp upstream of *Abrabidopsis* microRNAs were used to screen for known transcriptional binding sites. The GO term enrichment test was used for functional analysis of target genes in each identified microRNA regulatory module.

Another interesting method for identification of MRMs using a matched microRNA-mRNA time-course data was published recently.[76] In this study, a cancer dataset corresponding to a drug study involving a multiple myeloma cell line was used. The data consist of microarray microRNA and mRNA expression profiles from the same samples over six time points. An odds-ratio statistic was proposed for identification of microRNAs with regulatory potential, microRNA-target pairs, and time lags between changes in microRNA expression and those of its targets.

With a somewhat different goal from the identification of MRMs, Cheng and Li[77] examined how regulatory activity changes of microRNAs effect the expression changes of their target transcripts measured by microarray. The predicted binding affinity score for a microRNA and target gene pair is defined as the sum of absolute values of binding energies for all the target sites of the microRNA within the 3'-UTR of the gene. The binding energy can be obtained from any target prediction tool (eg, miRanda) that assigns each microRNA-target duplex a binding energy. Then the information of binding scores is combined with the gene expression to calculate an activity change (AC) score for each microRNA. This score is a measure of the inferred relative activities of microRNAs between two different conditions. A positive AC score indicates activity enhancement of the corresponding microRNA, while a negative AC score indicates activity reduction. The method was evaluated using microarray data sets that measure gene expression changes in cell lines after transfection or inhibition of several specific microRNAs. The results indicate that the method can detect activity enhancement of the transfected microRNAs as well as activity reduction of the inhibited microRNAs with high sensitivity and specificity. It was also demonstrated that the inference is robust with respect to false positives of target prediction.

We provide a list of available online resources for integrative databases and analysis methods for identification of functional microRNA targets and MRMs using microRNA and mRNA expression data (Table 1). It can be seen in Table 1 that the number of publically available tools on functional MRMs analysis is limited.

Conclusions and perspectives

As described above, a variety of methods have recently been published which are aimed at identifying the functional (or condition-specific) MRMs involved in complex interactions between microRNAs and their target genes. However, these methods have usually been aimed at identifying the direct interaction (*cis*-regulatory mechanism) even through the general term, functional MRM, was used.

The regulatory mechanism of microRNAs in complex cellular systems can be multifactorial. Each microRNA can regulate the mRNA transcripts of hundreds of target genes. One microRNA can have multiple target sites in the mRNA transcript of a downstream gene. On the other hand, one mRNA can be targeted by multiple microRNAs. Two potential mechanisms that may be involved are proposed (Figure 2).[78] In the *cis*-regulation mechanism, a microRNA directly targets the mRNA and regulates the expression of the target gene at post-transcriptional levels (eg, enhance mRNA degradation and inhibit translation). In the trans-regulation mechanism, following the expression changes of specific microRNA targeted genes (eg, genes coded


Table I Online resources for integrative prediction of mRNA targets and microRNA-mRNA regulatory modules

Resources	Remarks	URL
TarBase[7]	Experimentally confirmed mRNA targets	http://diana.cslab.ece.ntua.gr/tarbase/
miRecords[9]	Validated targets component provides a large, high-quality database of experimentally validated microRNA targets from meticulous literature curation; predicted targets component provides a collection of predicted microRNA targets obtained from 11 established microRNA target prediction programs (DIANA-microT, MicroInspector, miRanda, miRDB, miTarget, NBmiRTar, PicTar, PITA, RNA22, RNAhybrid, and TargertScanS).	http://mirecords.biolead.org/
miRMap 2.0[38]	Information on microRNAs collected in miRBase,[21] and literature as well as mRNA targets predicted from miRanda, RNAHybrid, and TargetScan with additional filtering criteria for false positive reduction. These criteria are: (1) it must have been predicted by at least two tools, (2) target genes must contain multiple sites, and (3) the target site must be accessible.	http://mirnamap.mbc.nctu.edu.tw/
MiRGen[37]	Provide (1) positional relationships between animal microRNAs and genomic annotation sets and (2) animal microRNA targets according to combinations of four widely used target prediction programs (DIANA-microT, miRanda, PicTar, and TargetScanS).	http://www.diana.pcbi.upenn.edu/miRGen.html
miRGator[46]	Integrates target prediction, functional analysis, expression data, and genome annotation. Use microRNAs from (miRBase) and the predicted targets from different programs (miRanda, PicTar, and TargetScanS). The predicted targets are further analyzed for functional enrichment with terms in GO and KEGG,[47] ENMAPP,[45] BioCarta.[48] A compendium of microRNA expression is also provided.	http://genome.ewha.ac.kr/miRGator/
SigTerms[50]	Retrieves all microRNA-mRNA functional pairs in a given database for a list of genes. GO term enrichment statistic for overrepresentation of predicted targets in the gene set is also provided. Use the entire set of microRNA predictions from (PicTar, TargetScan, and miRanda). Free Excel macros available.	http://sigterms.sourceforge.net/
MMIA[56]	Integrate microRNA and mRNA expression data with predicted microRNA target information (obtained from TargetScan, PicTar, PITA) for analyzing microRNA-associated phenotypes and biological functions by Gene Set Enrichment Analysis (GSEA). Requires microRNA and mRNA expression profiles.	http://129.79.233.81/~MMIA/mmia_main.html
CORNA[51]	Allow users to test for over-representation of microRNA-target associations, GO terms, and KEGG pathways enrichment analysis in a list of predicted mRNA target. The package is written in R.	http://corna.sourceforge.net/
GeneSet2miRNA[66]	Identify regulatory modules including up to five microRNA regulators for a list of mRNAs in which the predicted targets of the set of microRNAs are mostly over-represented. The mRNA targets are obtained from the predicted targets component of miRecords. Requires microRNA and mRNA expression profiles.	http://mips.helmholtz-muenchen.de/proj/gene2mir/
GenMIR++[80]	Use a Bayesian learning algorithm to obtain a set of functional microRNA targets of high confidence based on the use of predicted microRNA targets and expression profiles of microRNA and mRNAs. Requires microRNA and mRNA expression profiles. Free Matlab source code available.	http://www.psi.toronto.edu/genmir/

for transcription factors, genes coded for RNA regulating proteins, and genes coded for proteins that will interact with the target protein), subsequent effects may alter the transcription of other genes, levels of other mRNAs, or interactions among proteins. Thus microRNA may exert its functional effects through *trans*-regulatory mechanism(s). An interesting example in a recent study on tongue cancer suggests that hsa-miR-222 regulates the MMP1 expression through both direct *cis*-regulatory mechanism (targeting MMP1 mRNA) and indirect *trans*-regulatory mechanism

cis-regulation

trans-regulation
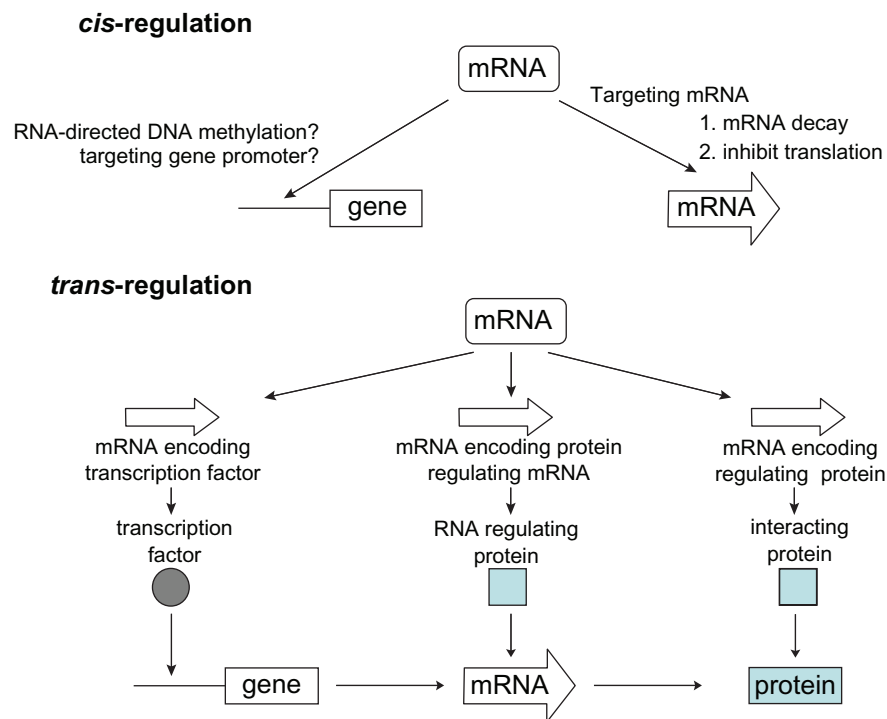
Figure 2 Potential microRNA regulation mechanisms. Copyright © 2009. Adapted with permission from Lin Liu X, Chen Z, Yu J, Xia J, Zhou X. MicroRNA profiling and head and neck cancer. *Comp Funct Genomics*. 2009:837514. [78]

(indirect controlling of MMP1 gene expression by regulating SOD2 expression and H_2O_2 production, which inference the activity of MMP1 gene promoter).[17]

The basic principle of integrative analysis for identifying functional target mRNAs is the reverse correlation of the microRNA and mRNA expression profiles. Therefore, most of the methods surveyed can only link microRNAs with their potential direct target mRNAs. The relationship between microRNAs and other secondary targets cannot be clarified, but it may be possible to infer these from the expression data. There are some differentially expressed mRNAs which are not direct targets of microRNAs, but are down-stream genes of some transcription factors (TFs) that are direct targets of microRNAs. The mRNA expression levels of these TFs may or may not be affected by the deregulation of the microRNAs; in the latter case it could be that the mRNAs of the TFs are inhibited for translation by their microRNA regulators. In addition, if the TFs are suppressors, the expression level of the downstream genes of the TFs can be positively correlated with that of the microRNAs. Accordingly, the current tools are not capable of identifying this type of microRNA's indirect regulatory targets, even though this relationship can be uncovered through further analysis of microRNA and mRNA expression data. Therefore, analysis methods that can predict the so-called *trans*-regulatory mechanism indicated in Figure 2 will be

needed. The only approach aimed explicitly at examining this type of mechanism is the method used by Joung and Fei.[75] New strategies on integrative analysis for the detection of MRMs using various information sources, including mRNA expression profile, predicted microRNA targets, transcription factors, and their DNA-binding motifs are also clearly needed. Furthermore, the identification of MRMs on an underlying network which includes microRNA, mRNA/TFs, and downstream TF target genes may provide better understanding of the functional roles of microRNAs. The performance of the methods should be systematically evaluated as the experimental evidence on MRMs grows.

Recently, significant differences between microRNA and nonmicroRNA target genes have been observed in terms of the length of the gene, the rate of mRNA expression, the rate of mRNA decay, and the protein half-life.[79] These observations imply that the suppression of the output of pre-existing protein-coding gene transcripts by microRNAs facilitates a more rapid and robust transition to new expression program(s). This could be achieved by enhancing mRNA degradation through an additive effect from multiple microRNA targeting. Accordingly, quantitative analysis of target mRNA expression levels of those with multiple microRNA-binding sites and those with single site may provide insights on the microRNA regulatory mechanism.

Acknowledgments/disclosures

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281–297.
2. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell*. 2009;136(4):642–655.
3. Wu L, Fan J, Belasco JG. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A*. 2006;103(11):4034–4039.
4. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120(1):15–20.
5. Gusev Y. Computational methods for analysis of cellular functions and pathways collectively targeted by differentially expressed microRNA. *Methods*. 2008;44(1):61.
6. Barbato C, Arisi I, Frizzo ME, Brandi R, Sacco LD, Masotti A. Computational challenges in miRNA target predictions: To be or not to be a true target? *J Biomed Biotechnol*. 2009;2009:803069.
7. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. The database of experimentally supported targets: a functional update of TarBase. *Nucl Acids Res*. 2009;37(Suppl 1):D155–D158.
8. Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*. 2006;12(2):192–197.
9. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucl Acids Res*. 2009;37(Suppl 1):D105–D110.
10. Huang JC, Babak T, Corson TW, et al. Using expression profiling data to identify human microRNA targets. *Nat Meth*. 2007;4(12):1045–1049.
11. Herrera B, Lockstone H, Taylor J, et al. MicroRNA-125a is over-expressed in insulin target tissues in a spontaneous rat model of Type 2 Diabetes. *BMC Med Genomics*. 2009;2(1):54.
12. Lim LP, Lau NC, Garrett-Engele P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005;433(7027):769–773.
13. Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci U S A*. 2006;103(8):2746–2751.
14. Wang X, Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res*. 2006;34(5):1646–1652.
15. Gennarino VA, Sardiello M, Avellino R, et al. MicroRNA target prediction by expression analysis of host genes. *Genome Res*. 2009;19:481–490.
16. Ambs S, Prueitt RL, Yi M, et al. Genomic profiling of microRNA and messenger RNA reveals deregulated microRNA expression in prostate cancer. *Cancer Res*. 2008;68(15):6162–6170.
17. Liu X, Yu J, Jiang LU, et al. MicroRNA-222 regulates cell invasion by targeting matrix metalloproteinase 1 (MMP1) and manganese superoxide dismutase 2 (SOD2) in tongue squamous cell carcinoma cell lines. *Cancer Genomics Proteomics*. 2009;6(3):131–139.
18. Liu X, Jiang L, Wang A, Yu J, Shi F, Zhou X. MicroRNA-138 suppresses invasion and promotes apoptosis in head and neck squamous cell carcinoma cell lines. *Cancer Lett*. 2009;286(2):217–222.
19. Webster RJ, Giles KM, Price KJ, Zhang PM, Mattick JS, Leedman PJ. Regulation of epidermal growth factor receptor signaling in human cancer cells by MicroRNA-7. *J Biol Chem*. 2009;284(9):5731–5741.
20. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucl Acids Res*. 2006;34(Suppl 1):D140–144.
21. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008(36 Database):D154–D158.
22. Krek A, Grun D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005;37(5):495–500.
23. Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucl Acids Res*. 2006;34(Suppl 2):W451–W454.
24. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009;19:92–105.
25. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*. 2007;27(1):91–105.
26. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucl Acids Res*. 2008;36(Suppl 1):D149–D153.
27. Maragkakis M, Reczko M, Simossis VA, et al. DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res*. 2009(37 Web Server):W273–W276.
28. Chaudhuri K, Chatterjee R. MicroRNA detection and target prediction: Integration of computational and experimental approaches. *DNA Cell Biol*. 2007;26(5):321–337.
29. Maziere P, Enright AJ. Prediction of microRNA targets. *Drug Discov Today*. 2007;12(11–12):452–458.
30. Watanabe Y, Tomita M, Kanai A, John JRaGJH. Computational methods for MicroRNA target prediction. *Methods Enzymol*. 2007;427:65–86.
31. Mendes ND, Freitas AT, Sagot MF. Current tools for the identification of miRNA genes and their targets. *Nucl Acids Res*. 2009;37(8):2419–2433.
32. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*. 2009;25(23):3049–3055.
33. Zhang B, Pan X, Wang Q, Cobb GP, Anderson TA. Computational identification of microRNAs and their targets. *Comput Biol Chem*. 2006;30(6):395–407.
34. Lindow M, Gorodkin J. Principles and limitations of computational microRNA gene and target finding. *DNA Cell Biol*. 2007;26(5):339–351.
35. Sethupathy P, Megraw M, Hatzigeorgiou AG. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat Methods*. 2006;3(11):881–886.
36. Swiss Institute of Bioinformatics. ElMMo miRNA target prediction server. 2009. Available from: http://www.mirz.unibas.ch/ElMMo3/. Accessed February 10, 2010.
37. Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG. miRGen: a database for the study of animal microRNA genomic organization and function. *Nucl Acids Res*. 2007;35(Suppl 1):D149–D155.
38. Hsu PWC, Huang HD, Hsu SD, et al. miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucl Acids Res*. 2006;34(Suppl 1):D135–D139.
39. Rusinov V, Baev V, Minkov IN, Tabler M. MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucl Acids Res*. 2005;33(Suppl_2):W696–W700.
40. Wang X. miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA*. 2008;14(6):1012–1017.
41. Kim SK, Nam JW, Rhee JK, Lee WJ, Zhang BT. miTarget: microRNA target gene prediction using a support vector machine. *BMC Bioinformatics*. 2006;7:411.
42. Yousef M, Jung S, Kossenkov AV, Showe LC, Showe MK. Naive Bayes for microRNA target predictions machine learning for microRNA targets. *Bioinformatics*. 2007;23(22):2987–2992.
43. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007;39(10):1278–1284.
44. Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. 2006;126(6):1203–1217.

45. Gladstone Institutes, University of California at San Francisco. GenMAPP. 2009. Available from: http://www.genmapp.org/about.html. Accessed February 10, 2010.

46. Nam S, Kim B, Shin S, Lee S. miRGator: an integrated system for functional annotation of microRNAs. *Nucl Acids Res*. 2008;36(Database issue):D159–D164.

47. Kanehisa M, Goto S, Hattori M, et al. From genomics to chemical genomics: new developments in KEGG. *Nucl Acids Res*. 2006;34(Suppl 1):D354–D357.

48. BioCarta. 2010. Available from: http://www.biocarta.com/. Accessed February 10, 2010.

49. Barrett T, Troup DB, Wilhite SE, et al. NCBI GEO: archive for high-throughput functional genomic data. *Nucl Acids Res*. 2009;37(Suppl 1):D885–D890.

50. Creighton CJ, Nagaraja AK, Hanash SM, Matzuk MM, Gunaratne PH. A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *RNA*. 2008;14(11):2290–2296.

51. Wu X, Watson M. CORNA: testing gene lists for regulation by microRNAs. *Bioinformatics*. 2009;25(6):832–833.

52. Peng X, Li Y, Walters KA, et al. Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics*. 2009;10(1):373.

53. Lin S, Ding J. Integration of ranked lists via cross entropy Monte Carlo with applications to mRNA and microRNA studies. *Biometrics*. 2009;65:9–18.

54. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature*. 2008;455(7209):64–71.

55. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. *Nature*. 2008;455(7209):58–63.

56. Nam S, Li M, Choi K, Balch C, Kim S, Nephew KP. MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucl Acids Res*. 2009;37(Web Server issue):W356–W362.

57. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.

58. Perez-Iratxeta C, Bork P, Andrade-Navarro MA. Update of the G2D tool for prioritization of gene candidates to inherited diseases. *Nucl Acids Res*. 2007;35(Suppl 2):W212–W216.

59. Jiang Q, Wang Y, Hao Y, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucl Acids Res*. 2009;37(Suppl 1):D98–D104.

60. Joung JG, Fei Z. Computational identification of condition-specific miRNA targets based on gene expression profiles and sequence information. *BMC Bioinformatics*. 2009;10(Suppl 1):S34.

61. Vapnik VN. *Statistical Learning Theory*. Philadelphia, PA: Wiley Interscience; 1998.

62. TAIR. The Arabidopsis Information Resource (TAIR). 2010. Available from: http://www.arabidopsis.org/. Accessed January 10, 2010.

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

64. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell*. 2003;115(7):787–798.

65. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. *PLoS Biol*. 2004;2(11):e363.

66. Antonov AV, Dietmann S, Wong P, Lutter D, Mewes HW. GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucl Acids Res*. 2009;37(Web Server issue):W323–W328.

67. Joung JG, Hwang KB, Nam JW, Kim SJ, Zhang BT. Discovery of microRNA mRNA modules via population-based probabilistic learning. *Bioinformatics*. 2007;23(9):1141–1147.

68. Baluja S. *Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning*. Pittsburgh, PA: Carnegie Mellon University; 1994.

69. Baluja S. Genetic algorithms and explicit search statistics. *Adv Neural Inf Proc Syst*. 1997;9:319–325.

70. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005;435(7043):834.

71. Huang JC, Babak T, Corson TW, et al. Using expression profiling data to identify human microRNA targets. *Nat Methods*. 2007;4(12):1045–1049.

72. Tran D, Satou K, Ho T. Finding microRNA regulatory modules in human genome using rule induction. *BMC Bioinformatics*. 2008;9(Suppl 12):S5.

73. Nada L, Branko K, ek, Peter F, Ljup, o T. Subgroup discovery with CN2-SD. *J Mach Learn Res*. 2004;5:153–188.

74. Liu B, Li JY, Tsykin A. Discovery of functional miRNA-mRNA regulatory modules with computational methods. *J Biomed Inform*. 2009;42(4):685–691.

75. Joung JG, Fei Z. Identification of microRNA regulatory modules in Arabidopsis via a probabilistic graphical model. *Bioinformatics*. 2009;25(3):387–393.

76. Jayaswal V, Lutherborrow M, Ma DDF, Yang YH. Identification of microRNAs with regulatory potential using a matched microRNA-mRNA time-course data. *Nucl Acids Res*. 2009;37(8):e60.

77. Cheng C, Li LM. Inferring microRNA activities by combining gene expression with microRNA target prediction. *PLoS ONE*. 2008;3(4):e1989.

78. Liu X, Chen Z, Yu J, Xia J, Zhou X. MicroRNA Profiling and head and neck cancer. *Comp Funct Genomics*. 2009;837514.

79. Hu Z. Insight into microRNA regulation by analyzing the characteristics of their targets in humans. *BMC Genomics*. 2009;10(1):594.

80. Huang JC, Morris QD, Frey BJ. Bayesian inference of microRNA targets from sequence and expression data. *J Comp Biol*. 2007;14(5):550–563.