

Data integration in functional analysis of microRNAs

Hasan Oğul¹, Mahinur S. Akkaya²

¹*Department of Computer Engineering, Baskent University, Ankara, Turkey.*

²*Department of Chemistry, Middle East Technical University, Ankara, Turkey.*

Abstract: The discovery of microRNAs (miRNAs), about a decade ago, has completely changed our understanding of the complexity of gene regulatory networks. It has already been shown that they are abundantly found in many organisms and can regulate hundreds of genes in post-transcriptional level. To elucidate the individual or co-operative effects of miRNAs, it is required to place them in the overall network of gene regulation and link them to other pathways and systems-level processes. One key step in this effort is predicting targets of individual miRNAs. Although current tools are helpful in predicting miRNA-mRNA binding to a considerable extent, they are not able to model many-to-many relationships between miRNAs and their targets using solely sequence information. Therefore, other types of information sources have been employed for better prediction of these functional relationships. This report focuses on the state-of-the-art solutions and current challenges on mining miRNA-related data to discover the systems-level role of miRNAs, with an emphasis on the integration of different information sources. We aim to provide new insights for fusion of different types of biochemical and experimental information sources which may facilitate functional analysis of miRNAs.

INTRODUCTION

MicroRNAs are small non-coding RNAs which regulate gene expression at the post-transcriptional level. They are about 22 nucleotides long and they predominantly bind to the 3' untranslated region (3'UTR) of messenger RNAs (mRNAs) to inhibit translation or to induce cleavage. miRNAs can have hundreds of targets and take on unique properties that enable them to regulate several processes such as cellular differentiation, proliferation, and apoptosis [8, 22, 85]. It is currently known that miRNAs are linked to several diseases, such as cancer, due to the defects in the regulation of mRNA translation, which lead to an abnormal production of a protein [121].

The first miRNA lin-4 was discovered more than a decade ago [60], but the term of microRNA was first coined in 2001 when tens of small RNAs with regulatory potential were discovered in *C. elegans* [57, 59, 98]. The discovery of miRNAs has completely shifted our understanding in gene regulation process and changed the direction of research in life sciences. Along this direction, several computational problems have arisen and bioinformatics research in the last decade has strongly affected by these developments.

Early works in the area of miRNA informatics focused on recognizing miRNA genes along long strands [79]. The problem is defined as finding a region which represents a putative miRNA from given sequence. The experimental identification of miRNAs is still cumbersome since some miRNAs are difficult to isolate by cloning or other conventional techniques due to their low expression values, short lengths and tissue specificity. Thus, computational identification provides a valuable complement to cloning and other techniques. The problem has been approached by several methods: using filters based on sequence content or conserved motifs over an initial candidate set [26, 84], searching sequence or structure based homologs [125, 127], employing targets to evaluate complementarities for potential binding [130], and utilizing machine learning algorithms based on predefined feature sets [13, 33]. Recent experiments have demonstrated that a significant improvement has been achieved in miRNA prediction accuracy [79].

A more challenging problem in the field is the recognition of target genes for specific miRNAs. The difficulties in experimental verification of miRNA target genes have motivated the development of computational methods to produce potential targets. The problem has been widely studied in the last decade [30]. Many of available methods use sequence information coming from miRNA and potential

mRNA target. While the problem is easily tractable in plants due to almost perfect complementarity between miRNA sequences and mRNA targets, the process cannot be easily described in animals, where most miRNAs are imprecisely complementary to their targets and inhibition mechanism has not been completely understood yet. Some recent works attempted to use gene expression data, when available, to assess the miRNA-target relationships [24, 25]. Fusion of gene expression and sequence information has been shown to improve the accuracy of predictive and exploratory analyses.

While miRNA-target relationships can provide invaluable information about the role of individual miRNAs and corresponding target gene's regulation, to get a deeper insight into their systems-level functions, it is desired to place miRNAs in the overall network of gene regulation. This information is essential to describe the many-to-many relationships between miRNAs, their targets and potential other factors. Current science is weak in this respect; only a couple of papers which introduce methods to analyze available data to identify systems-level processes can be found in the literature.

To shed light on future research in miRNA informatics, an attempt has been made in this survey to compile some of the existing literature pertaining to the fusion of different heterogeneous data sources in analysis of miRNA related processes. We start with a brief description of the state-of-the-art in sequence-based prediction of individual miRNA targets. After introducing the alternative information sources which have high potential to help understanding miRNA functionalities, we present the recent attempts for integrating several information sources for miRNA target prediction, activity prediction, and regulatory network inference. We finally discuss the remaining challenges and opportunities, from both computational and biological aspects, regarding better prediction of information flows originated, catalyzed or redirected by miRNAs, with a special emphasis on data fusion.

PREDICTION OF miRNA TARGETS

Which genes are regulated by an individual miRNA is a high potential biological knowledge. It serves a preliminary step in understanding the interaction of a miRNA with other genes, then its functional position in regulatory mechanisms. On the other hand, experimental identification of this relationship is long and cumbersome. Therefore, many studies have been devoted to predict the binding relationship between given miRNA and mRNA in the last decade. Available methods usually take the miRNA sequence and 3'UTR sequence of mRNA and output a binary decision whether given miRNA can recognize the mRNA as its target.

In 5' end of mature miRNA, there is a region called seed, of 7-8 nucleotides length, which is commonly believed to govern the recognition of binding site in 3'UTR of target mRNA. Several experiments have revealed that a strong base pairing is usually observed in validated miRNA-target pairs [108]. Early methods in target prediction widely used this knowledge to develop specific base-pairing rules between miRNA and target sequences [63]. It was later found that the other regions of mature miRNA sequence could also make a contribution in target selection. Therefore, some recent works used whole sequence information while searching a sufficient level of complementarity between miRNAs and candidate targets [27, 96] or the existence of some binding motifs which are previously detected to be enriched in validated pairs [80]. However, it has been observed that the predictions based solely on imperfect sequence complementarity produced a large number of false positives due to random appearance of base-pairings [91]. This fact encouraged the researchers in the field to use additional information in developing specific binding rules. As a consequence, many of the recent methods have relied on a combination of several factors such as sequence complementarity, minimization of free energy in miRNA-target duplex [44, 93], cross-species conservation [23, 75], and site accessibility defined over predicted secondary structures [34, 49]. Recent research suggests that incorporation of various machine learning methods by using the same factors as classifier's feature space has improved overall target prediction accuracy [6, 17, 53, 99, 134].

While many successful attempts for computational miRNA target prediction have been reported in the last years, the lack of a large consensus in their prediction results [2, 91, 104] has suggested that the molecular

mechanism behind target recognition is far beyond the empirical assumptions made over sequence and structure information by current tools. It is likely that other important functional parameters which mediate miRNA-target interactions remain to be identified. Recent experiments have already established that miRNAs have a certain effect on the expression level of the mRNAs recognized as targets [65, 87, 104, 107]. Therefore, it is very reasonable to argue that a strong correlation should exist between miRNA and mRNA expression profiles if there is a binding relationship between them. This motivated the use of gene expression data, when available, in target prediction algorithms [37, 94]. It has been possible to develop improved algorithms which integrate the output of traditional prediction tools with gene expression data to identify more reliable targets [20, 38, 44, 81]. Table 1 delivers a timely and exhaustive list of available miRNA target prediction tools by highlighting the input data integrated in their algorithms. Since this paper seeks to give a perspective on one step beyond, the reader is referred to recent surveys for implementation details and performance comparison of these methods [2, 7, 8, 30, 100, 111, 135, 136].

PREDICTION OF miRNA ACTIVITY

While the appearance of target sites on mRNA sequences provides useful information for identifying putative miRNA-target duplexes, many target sites are not functional *in vivo*, thus, target predictions may not yield reliable information to define effective miRNA functions. With an additional goal of inferring effective miRNA activity, Cheng and Li combined miRanda target predictions with expression profiles over a statistical model [19]. The effective regulatory activity change of a miRNA is defined using the expression changes of its target genes instead of its own expression change. A binding affinity score is computed using the sum of absolute values of binding energies, which were inferred from miRanda predictions, for all target sites. Final affinity is integrated into an activity score with corresponding expression changes in targets. They demonstrated the performance of their method on three independent miRNA transfection data sets having six miRNAs. They reported two limitations for their method. First, relating miRNA activity change directly to target mRNA expression change is limited because the target gene suppression might be mediated through translation repression. In this case, gene expression data does not have the ability to reflect the activity change in miRNA. Second, their method considers different miRNAs independently during activity changes whereas multiple miRNAs may involve interactively in the regulation of target genes.

Another method with a similar goal was introduced by Madden et al. with a slight difference in the definition of miRNA activity [74]. They suggested that miRNA activity could be better explained by its association with a particular disease or condition. In a similar manner, their method extracts miRNA-gene interactions using other target prediction tools (TargetScan, PicTar and miRanda) and integrates the predictions with expression data using a multivariate coupling approach, called co-inertia analysis. Between-group analysis is combined with correspondence analysis for supervised classification of miRNAs into particular disease or conditions. This procedure does not provide solutions to the limitations of Cheng and Li's approach, but it gives a different perspective to miRNA activity analysis.

Tsang et al. introduced mirBridge [116], a method which considers cooperative effects of miRNAs on gene sets to dissect miRNA functions. The algorithm takes a gene set of known functions to assess the enriched sequence motifs which may introduce functional sites to be recognized by a given miRNA. In evaluation of functional significance of target sites found, mirBridge uses a combination of the results of three different statistical tests. The tests consider the gene set properties and biological characteristics of target sites to compute a score which defines the proportion of likely functional targets. They reported that this approach would offer better sensitivity and specificity in comparison with a standard procedure like Fisher's exact test which evaluates the number of predicted target overlaps. The miRNAs grouped into same families based on shared seed sequences are assumed to function together by regulating common targets. Resulting associations between the modules of miRNA families and input gene sets constitute a kind of cotargeting networks. MirBridge is example of integrating sequence data with signaling pathway and protein complex data to infer miRNA functions. Although, the model can produce a better view of many-to-

many relationships between miRNA activities and target genes, the results might be functionally irrelevant since the interactions are based only on seed matches. While genome-wide behaviors of target genes are implicitly modeled upon gene sets, miRNA expression changes are completely ignored.

Table 1. miRNA target prediction tools.

Name	Input Data	Method	Date	Ref
TargetScan	Sequence, conservation, energy	Scoring and ranking	2003	[64]
RNAHybrid	Sequence, energy	Dynamic programming and statistical analyses	2004	[93]
miRanda	Sequence, conservation, energy	Dynamic programming and filtering	2004	[44]
PicTar	Sequence, conservation, energy	Hidden Markov model	2005	[53]
TargetBoost	Sequence	Boosted genetic algorithm	2005	[99]
Moving Target	Sequence, energy	Scoring and ranking	2005	[15]
TargetScanS	Sequence, conservation	Scoring and ranking	2005	[63]
MicroInspector	Sequence, energy	Scoring and ranking	2005	[97]
RNA22	Sequence, energy	Markov chains	2006	[80]
miTarget	Sequence, energy, structure	Support vector machine	2006	[51]
MicroTar	Sequence, energy	Scoring and ranking	2006	[113]
MirTarget	Sequence, conservation, energy	Scoring and ranking	2006	[124]
EIMMo	Sequence, conservation	Bayesian model	2007	[23]
Pita	Sequence, conservation, energy, structure	Scoring and ranking	2007	[49]
TargetRank	Sequence, conservation	Scoring and ranking	2007	[83]
NBmirTar	Sequence, energy	Naive Bayes model	2007	[134]
GenMiR++	Other target predictions (TargetScanS), miRNA expression, mRNA expression	Variational Bayesian learning	2007	[38]
mirWIP	Sequence, conservation, energy, structure	Scoring and ranking	2008	[29]
miRTif	Sequence	Support vector machine	2008	[131]
FindTar	Sequence, structure	Scoring and ranking	2008	[132]
HuMiTar	Sequence	Scoring and ranking	2008	[96]
miRGator	Other target predictions (miRanda, PicTar, TargetScanS), miRNA expression, mRNA expression	Gene set enrichment analysis, scoring and ranking	2008	[81]
MirTarget2	Sequence, conservation, energy, miRNA expression, mRNA expression	Support vector machine	2008	[123]
SigTerms	Other target predictions (miRanda, PicTar, TargetScanS), mRNA expression	Enrichment statistics, scoring and ranking	2008	[20]
DIANA-Micro-T	Sequence, conservation	Scoring and ranking	2008	[75]
TargetMiner	Sequence	Support vector machine	2009	[6]
FastH	Sequence, energy	Scoring and ranking	2009	[90]
SVN+SC	Sequence, mRNA expression	Sequence alignment, support vector machine	2009	[45]
HocTar	Other target predictions (miRanda, PicTar, TargetScanS), miRNA expression, mRNA expression	Scoring and ranking	2009	[25]
Mtar	Sequence, energy	Artificial neural network	2010	[17]
PACMIT	Sequence, conservation, energy, structure	Markov model	2010	[77]
SVMicro	Sequence, conservation, energy, structure	Support vector machine	2010	[70]
TargetSpy	Sequence, energy, structure	Support vector machine	2010	[110]
ExprTarget	Other target predictions (miRanda, PicTar, TargetScanS), miRNA expression, mRNA expression	Multivariate logistic regression	2010	[24]
ExpMicro	Other target predictions (SVMicro), miRNA expression	Gaussian mixture model	2010	[71]

CORNA [128] and GeneSet2miRNA [3] are similar tools which explore the link between an individual miRNA and a gene set. They both take a list of genes and a miRNA as input and examine if there is a regulatory association between the miRNA and the gene set. Having predicted targets in miRbase database, CORNA employs three standard statistical procedures (HyperGeometric test, Fisher's exact test and Chi-square test) for enrichment analysis and can report a list of associated gene sets for a given miRNA, or alternatively a list of miRNAs which potentially regulate a given gene set. GeneSet2miRNA uses predicted targets in miRecords database [129] and perform HyperGeometric test to analyze gene set enrichments. It may evaluate the cooperation of up to four miRNAs by searching more complex regulatory relationships using greedy heuristics. The performance of these tools strongly depends on the availability and reliability of gene-sets used. Therefore, their generalization ability is limited by the source where these gene-sets are obtained. Furthermore, available or predicted gene-sets may not necessarily explain the behavior of all individual genes.

Time-course expression data was further included in identification of functional miRNA-target mRNA pairs with the time lags between changes in miRNA expression and those of its target mRNAs [43]. An odd-ratio statistics was used with computational target predictions to identify miRNA-mRNA regulatory relationships over a cancer data set comprising miRNA and mRNA expression profiles from the same samples in different time points. This work was an initial attempt to consider the effects of time-lags, however, some advanced probabilistic methods, instead of simple odd-ratio statistics, are needed to handle noisy and potentially missing data at different time points.

Several web-servers have been released in recent years to analyze miRNA activities. MMIA is one of these tools for integrative analysis of miRNA functions [82]. In a 4-step analysis, it may provide information about miRNA-associated phenotypes and biological functions. Further analysis can provide output for miRNA associated diseases, relevant pathways and gene ontology annotations. This tool cannot take into account whole expression profiles of miRNAs and mRNAs, instead, consider the variables as up-regulation or down-regulation. This may lead to loss of the degree of correlations between them. Ulitsky et al. introduced another web-based framework [118], called FAME (functional assignment of miRNAs via enrichment), based on weighted graphs to identify miRNA activities with two alternative applications (1) Using targeted gene clusters with common annotations to relate miRNA functions, and (2) Matching miRNA and mRNA expression profiles to predict miRNA-based regulation. A recent online tool by Sales et al. [101], called MAGIA, can provide a combination of several traditional statistical inference methods to analyze whole expression profiles together with target predictions by other tools. It can offer a visual analysis of combinatorial miRNA activity by a regulatory network. These web-based tools do not face the computational challenges of data integration; therefore do not propose new models for statistical analyses. However, they consider the problem in user-level and provide online solutions over known statistical methods with high functionality. Although they provide sample guidelines about how to use them, the pipeline to be used in specific problems is usually left to their users.

REGULATORY NETWORKS COMPRISING miRNAS

A significant progress has been made toward identifying targets for an individual miRNA using various types of information sources during the last several years. However, the action of miRNAs may not be merely effecting the expression of one or a few prominent genes, but also coordinately regulating many genes by influencing the structure of the network where these regulators have a central role. Therefore, available tools or methods for predicting target genes or individual activities of a single miRNA may not be sufficient to answer several fundamental questions regarding the functional and systems-level role of miRNAs, e.g. how multiple miRNAs can work in cooperation to regulate a group of genes or a specific disease related pathway, or which miRNAs are expressed under a specific condition or in a certain time period? It has been therefore an emerging need to develop methods which can model more complex interactions and relationships between miRNAs, their targets and their regulators.

The first attempt in this respect was the study of Yoon and De Micheli [133]. They introduced the term of miRNA regulatory modules (MRMs) to define the coordinated activity of miRNAs with their targets. To identify MRMs, they proposed a graph-theoretical approach using miRNA and mRNA sequence data as input. Their method has three steps. First, miRNA-mRNA pairs are selected based on other target prediction tools. Second, the pairs are represented in a weighted bipartite graph, which is also called as relation graph. Each pair is weighted using a function determined by Principal Component Analysis on a feature space obtained from miRNA-target duplex formation. The feature space is built over local sequence alignment scores and free energy of putative duplex. In the final step, starting from an initial set of miRNAs that have similar binding patterns on created graph, the algorithm iteratively updates the graph and collects a set of regulatory modules, i.e. groups of miRNA-mRNA pairs based on refined duplex scores. Although their model produced hundreds of modules for 7886 predicted pairs of 2888 genes and 156 mRNAs in human genome, they demonstrated the performance of their method on only one of the modules using enriched GO annotations. Therefore, the model requires further biological validation to put into practice.

The model proposed by Yoon and De Micheli uses solely the sequence information. This approach has two major limitations: (1) The model quality mainly depends on the reliability of target predictions, and (2) Sequence information cannot afford to explain changing behaviors of miRNAs and their targets due to time and other environmental factors. In a later work by Joung et al., gene expression data is further integrated to capture instant behavior of miRNAs and target genes [47]. They defined a fitness function which evaluates the weighted sum of an expression coherence score and sequence-based binding scores for predicted miRNA-mRNA duplexes in current module. Their evolutionary learning algorithm attempts to optimize the fitness function with a prior setting of weight parameters from a random initial configuration of modules. The algorithm produces new configurations in each generation and is finalized with a solution which attains best fitness value. The validity of the algorithm was evaluated on a data set of human cancer samples containing 217 miRNAs and 16063 mRNAs. However, the framework produced only two modules which may have incidentally meaningful relations due to the abundance of genes and miRNAs in each module. The main drawback of this approach is the fact that it relies on an evolutionary algorithm which largely depends on manually curated and highly sensitive parameters. Therefore, it can easily be stuck on a local optima and produce results which are useful only in a certain context. A second drawback the method is its limitation of applications due to its dependency to some data obtained in a set of common conditions, which are usually unavailable. Another limitation of the method is that it can evaluate the expression coherence among miRNAs or mRNAs but not between them.

Tran et al. used the assumption that the genes regulated by same miRNA or miRNA groups should behave in a similar manner in terms of their expression[114]. By combining sequence based target predictions and expression profiles of miRNA and putative target genes, they built a rule-base to obtain many-to-many relations between miRNAs and mRNA. The algorithm starts with calculating correlation coefficient between target genes' expression profiles and converting them to a binary representation by a predefined threshold. Using the intersection of predicted binding matrix with binary correlations, the gene sets are divided into two classes: "similar" or "dissimilar". From this initial set, the rule-base is constructed by a rule induction system, called CN2-SD [58], which iteratively learns a rule set by several rule editing schemes in each step until a stopping criteria is reached. Uninformative rules are filtered out using a manually curated filtering policy. Each rule is considered to be a regulatory module and evaluated by GO annotations. On the same data set as used by Joung et al. [47], they produced 79 modules. In contrast to the work of Joung et al., they were able to unravel more specific relationships in small modules having only a couple of related genes and regulator miRNAs.

A probabilistic graphical model, similar to author-topic models previously used for text collections [109] was introduced by Joung and Fei [45], which combines mRNA expression profiles with targeting information to cluster simultaneously the miRNAs and the conditions at which the predicted targets are expressed. The model represents each gene targeted by specific miRNAs by a list of counts indicating expression levels in specific conditions. Each miRNA is associated with a multinomial distribution over a latent variable

representing regulatory modules. Another latent variable is used to characterize which miRNA is associated with given gene expression activities. The model parameters are inferred using an approximation method based on collapsed Gibbs sampling. Their approach is first in a sense that the roles of miRNAs are modeled over a specific subset of conditions instead of considering all available conditions. In the experiments performed on 637 miRNA-target pairs of 137 miRNAs and 382 genes from Arabidopsis, they revealed 10 regulatory modules which are highly correlated to known development processes. They also hypothesized, based on the enriched sequence motifs in promoters of clustered miRNAs, that some transcription factors exist to regulate miRNA groups in several developmental processes in Arabidopsis. One limitation of this method is that designed probabilistic framework does not allow the inclusion of miRNA expression profiles in parameter estimation but only considers the expression profiles of target genes. Therefore, the results are strongly depend on target mRNA information. Since the list of targets is usually predicted using sequence information, the algorithm ignores the own behavior and ability of miRNA in question.

Incorporation of condition specificity in regulatory networks may provide better insights in some cases, for example, when the local analysis of cell-specific or tissue-specific regulatory mechanisms is required or the correlation between some diseased samples and gene activities are investigated. It was also recently shown that expression profiles for miRNAs are distinct in different cancers [72]. In this respect, another condition-specific approach was introduced by Liu et al. to identify miRNA groups with their targets for normal and cancer samples [68]. Instead of unsupervised construction of unknown number of modules, they attempted to cluster miRNAs into two groups to understand the correlation between miRNA activities and corresponding disease. Their algorithm uses two different information sources in two independent steps. First, a set of putative networks, i.e. group of miRNA-target pairs, are identified using sequence based target predictions. Based on the output of the first step, condition-specific modules, which they called as functional miRNA regulatory modules (FMRMs), are built using miRNA and gene expression profiles. The model represents miRNA-target relationships by a bipartite graph and employs gene expression data to get a subgraph by pruning the initial one using association rule mining techniques [1]. In 12 prostate samples from normal and cancer tissues, they recorded high correlations between expression changes and cancer using differentiated modules.

It was recently found that miRNA can serve up-regulation in some cases on the contrary to early belief that miRNAs down-regulate target genes [119]. Liu et al. used this information in their later work to develop a new framework which can model more complex miRNA-target relationships such as down-regulation, up-regulation or mix-regulation, in which a miRNA may up-regulate its target in one condition, but repress translation in other condition [67]. In addition to sequence and gene expression data obtained from both miRNAs and mRNAs, they utilized sample categories of comparative design of experiments. Their algorithm is built upon a Bayesian network structure, where pairwise relations between miRNAs and their targets are defined as dependencies of their states encoded in a graphical representation. Similar to their previous work, different physiological conditions are taken into account and embedded into the Bayesian structure. Starting from an initial structure space, the algorithm applies a learning procedure to optimize a Bayesian score defined over expression profiles. The same approach was previously used by several gene regulatory network discovery models [32, 61], and now shown to be useful for the networks comprising miRNAs. In their work, different networks constructed for mRNAs and miRNAs are combined to generate significant pairwise interactions for a specific condition using predicted binding information.

Inverse expression relationships were taken into account in the work of Peng et al. [85] while inferring miRNA-mRNA regulatory modules associated with hepatitis C virus. They used a similar way in Yoon and Michelli's work, where an initial bipartite graph is pruned and post-processed, to represent multiple relationships between miRNAs and their targets. However, instead of using only sequence based target predictions in initial configuration, they proposed a couple of integrative evaluation steps before constructing relation graph. First, a miRNA-mRNA correlation matrix is calculated using the similarities in the expressions across samples. Then, a binary relation matrix is deduced based on a series of thresholds which minimizes false detection rates. Final matrix is combined with predicted binding matrix to infer an

initial relation graph. The algorithm has an advantage that miRNA and target expressions are evaluated using simultaneous transcriptional profiling information across the common set of samples.

A new probabilistic graphical model was proposed by Liu et al. [69]. They adopted correspondence latent Dirichlet allocation [9] to derive solution for functional miRNA regulatory module discovery. They started with the assumption that functional modules are reflected by expression data and they represented each functional module as a latent variable. The graphical model could give a solution by approximation of parameters defined over latent variables. To avoid biases sourced by sequence based target predictions, they tried to identify modules with and without using putative binding information to make a comparative analysis and they showed that the networks inferred without using sequence information had surprisingly produced results which largely overlap with predicted target relationships. They reported that several modules could be identified related to two breast cancer subtypes on experiments over mouse samples for 1112 miRNAs and 19223 mRNAs.

Table 2. Algorithms and tools for miRNA functional analysis

Input	Objective	Method	Date	Reference
miRNA and mRNA sequence	Identifying miRNA regulatory modules	Relation graphs (weighted bipartite graphs)	2005	[133]
Sequence-based target predictions, mRNA expression	Identifying miRNA regulatory modules	Population-based probabilistic learning	2007	[47]
Sequence-based target predictions, mRNA expression	Inferring miRNA activity on putative targets	Statistical analysis	2007	[19]
Sequence-based target predictions, miRNA and mRNA expression	Identifying miRNA regulatory modules	CN2-SD Rule induction	2008	[115]
Sequence-based target predictions, gene sets	Inferring miRNA activity on gene sets	Chi-square, HyperGeometric and Fisher's exact tests	2008	[128]
Sequence-based target predictions, gene sets	Inferring miRNA activity on gene sets	HyperGeometric test	2009	[3]
Sequence-based target predictions, miRNA and mRNA expression	Inferring miRNA activity on putative targets	Statistical tests and gene set analysis	2009	[82]
Sequence-based target predictions, miRNA and mRNA expression	Identifying condition-dependent miRNA regulatory modules	Association rule mining on bipartite graphs	2009	[68]
Sequence-based target predictions, miRNA and mRNA expression, sample categories	Identifying condition-dependent miRNA regulatory modules	Bayesian networks	2009	[67]
Sequence-based target predictions, mRNA expression	Identifying condition-dependent miRNA regulatory modules	Author-Topic Model	2009	[46]
Time-course miRNA and mRNA expression	Prediction of miRNA activity on putative targets	Odd-ratio statistics	2009	[43]
Sequence-based target predictions, miRNA and mRNA expression	Identifying miRNA regulatory modules	Bipartite graphs	2010	[85]
Sequence-based target predictions, miRNA and mRNA expression	Prediction of miRNA activity on putative targets	Weighted graphs and statistical tests	2010	[118]

Sequence-based target predictions, miRNA expression	Inferring miRNA activity on a disease or condition	Co-inertia and between-group analysis	2010	[74]
miRNA and mRNA sequence, gene sets	Inferring miRNA activity on gene sets	A newly defined statistical test	2010	[116]
miRNA and mRNA expression	Identifying condition-dependent miRNA regulatory modules	Latent Dirichlet Allocation	2010	[69]
miRNA and mRNA expression, a clinical parameter	Identifying condition-dependent miRNA regulatory modules	Module networks	2010	[11,12]
Sequence-based target predictions, miRNA and mRNA expression	Prediction of combinatorial miRNA activity on putative targets	Combination of correlation indexes, a variational Bayesian model and meta-analysis	2010	[101]

To explore the regulatory networks associated with prostate cancer, Bonnet et al. [11, 12] used a modified version of module networks, first introduced by Segal et al. [102]. In module network approach, genes are clustered into modules in terms of their similarity in expression profiles and the relations between modules are represented by a Bayesian network. Thus, a representation of many-to-many relationships can be achieved. They did not actually propose a new model for miRNA-included networks, instead, they employed a known algorithm to understand the difference between normal and cancer samples by first selecting a list of potential regulators, including transcription factors and miRNAs and then inferring condition specific module structure. They showed that assignment of a cancer-related clinical parameter, called Gleason grade system, as if another regulator, along with transcription factors, miRNAs and possible other signal transducers could improve the explanatory power of predicted network and provide a better comprehension of results. One possible critic to this application is the fact that binding information between regulators and their targets are completely ignored in the framework.

A summary of these algorithms and tools developed for miRNA activity prediction or regulatory network identification is given in Table 2 for a brief view of the state-of-the-art in the field.

FUTURE PERSPECTIVES

New findings in target recognition

Target prediction is still a primary step in miRNA functional analysis. Many of the available methods build their framework over a list of predicted targets for a given miRNA set. Therefore, the reliability of final result is strongly affected by initial predictions. Without heeding to enormous number of current predictors, it is expected that the need to develop new algorithms for miRNA target prediction will continue to increase due to new findings to explain structural miRNA-mRNA interactions. The location of miRNA-mRNA binding site is still elusive. Although there is a wide assumption that the binding occurs at 3'UTR region of putative target, recent reports have shown evidence that binding sites might be functional in other places; 5'UTR [73] or even coding regions [66, 112]. A recent investigation of 3'UTR motifs that couple to miRNA-based regulation suggests that miRNA regulation may be modulated by other sequence elements such as binding sites for RNA binding proteins [41]. Similarly, Wang et al. claim that a set of arginines concentrated in binding site of Argonaute (Ago) protein is more effective than seed pairing in stabilizing the miRNA-target interaction [126]. To the best of authors' knowledge, no systematic attempt has been done so far to utilize RNA-binding protein motifs in a miRNA target prediction algorithm.

More complex interactions

It is well established that the regulatory interactions are far beyond the pairwise interactions of individual genes with their regulators; miRNAs and transcription factors (TFs). A complete regulatory network comprises more complex interactions, such as between miRNAs and TFs [122], miRNAs and other non-

coding RNAs [16], and miRNA and their host genes [95]. Now that several types of these interactions are possible to retrieve from publicly available databases [21] or to predict using computational methods, it is further possible to define enhanced functional networks or modules such that miRNAs, their host genes and targets, TFs and other non-coding RNAs are also included.

The combinatory regulatory roles of miRNA and TFs have been discussed in several contexts. For example, Guo et al. explored the miRNA and TF regulation in schizophrenia genes and identified several feed-forward loops [28]. An analysis of TargetScan and PicTar datasets enlightened several recurrent network motifs including co-regulation of miRNAs and TFs, feed-forward-loops from miRNA to TF or from TF to miRNA, and composite loops having both directions [105]. Similar patterns were reported to appear in p53 master regulatory network [106]. Re et al. analyzed promoter motifs, in both target genes and miRNAs, to identify miRNA-TF feed-forward regulatory circuits [92]. In a computational analysis of gene expression data, Tsang et al. classified the network motifs into two types; type-I, where the transcription rate of a miRNA and its target gene are positively coregulated by a TF, and type-II, the rates are negatively coregulated [117]. They reported that type-II motifs are abundant in mammals while type-I motifs are also prevalent. The attention should be drawn into three facts in integrated network designs: (1) Loops may appear in interaction graphs; which is contradictory to traditional tree structures commonly used in modeling regulatory networks, (2) Up-regulation and down-regulation may concurrently be observed in different nodes, (3) TF and miRNA regulations differ in some aspects, therefore, when combined, one network may alter to conform to other network [40, 89].

Context sensitivity

Current methods encounter problems in identifying miRNA functions over specific conditions because they largely investigate the correlations for all available conditions instead of a specific context. A few methods can identify condition-dependent miRNA-target modules, where specified conditions are diseased and normal tissues. Complex interactions of miRNAs with other biomolecules motivate the reengineering of these methods as to include several other phenomenon. Concurrent grouping of miRNAs and their targets into independent modules, as done in available methods, cannot offer understanding of detailed interactions between them. Clusters must be associated with multi-level links between them. Importance of time-course analysis has been emphasized in several reports [43, 55]. Recent experiments have indicated that a miRNA may act as a switch such that it can be active/inactive in regulation of its targets in different time and conditions [120]. Therefore, the definition of “condition” requires an extension from a few static cases to dynamic time-dependent contexts.

Systems-level measurements

With increasing research in the field, the discoveries of new systems-level properties of miRNAs pose several challenges and create new opportunities for computational studies in modeling miRNAs functions. Larsson et al. recently analyzed the relationship between genome-wide transcript level changes and miRNA over-expression and showed that the turnover rates of predicted target mRNAs limit the activity of regulating miRNAs [56]. Their experimental findings suggest that the integration of mRNA turnover rates can improve the accuracy of target predictions tools. Another implication of the experiment is that, due to pre-existing decay rate, a target mRNA in one cell may not exhibit in a same manner in another cell although it is still targeted by same miRNA. The reverse regulatory effect of mRNAs, which was hypothesized by Seitz [103], was also confirmed by Poliseno et al. [88]. They showed that a pseudogene could bind the same miRNAs as the related protein-coding gene and perform regulatory activities with a relatively lower expression. The transfection experiments of Khan et al. showed that the competition between transfected and endogenous miRNAs for RNA-induced silencing complex caused the up-regulation of some endogenous miRNAs and depression of corresponding target genes [50]. In another transfection experiment, Arvey et al. found that a larger number of predicted target transcripts reduced the activity of miRNAs or siRNAs, resulting in down-regulation of each individual target to a lesser extent [4]. It means that a miRNA may more effectively down-regulate its targets when their count is lower even the binding

site is weak. Or conversely, a strong binding site may not be active if it has a large number of putative targets. Another genome-scale analysis revealed significant differences between miRNA and non-miRNA target genes in terms of their functional characteristics such as duration of half-life and gene structure [36]. These findings suggest that, in addition to local sequence or structural features, other global systems-level features, such as target mRNA abundance, pre-existing decay rates, and contribution of other miRNAs or pseudogenes, need to be considered in future research in developing tools for miRNA function analysis. The results also imply an obvious need that the miRNA target prediction should go through a transformation from a binary classification task to an integrative application where the level of interactions between miRNAs and their targets can be evaluated quantitatively in different contexts.

Increasing evidence suggests that, in addition to effecting transcript levels, miRNAs can also repress translation of some genes, resulting with a high impact on the identities of responsive proteins [5, 104]. Therefore, the protein output can be considered as another indicator of miRNA activities. In a recent experiment, for example, Kowarsh et al. showed that miRNA target proteins and disease-related proteins exhibit opposing patterns in their cellular localizations, signaling pathways and activatory/inhibitory effects [52]. In the case of having relevant data, this information can definitely be useful in construction of condition-specific interaction graphs including miRNAs.

Krützfeldt et al. emphasized the role of environmental and genetic factors which may affect miRNA expression in addition to associated signaling pathways in elucidation of miRNA functions [54]. They turned the attention of researchers into combinatorial effects of miRNAs, and stressed the importance of the utilization of multiple experiments. Another valuable strategy they suggested is to use chemically modified RNA analogs that are complementary to miRNAs, which can selectively diminish the abundance of some miRNAs, and in this way, inhibit their functions.

Linking to diseases

The strong association between miRNAs and cancer is obvious [10]. However, what behavioral characteristics of miRNAs lie behind this association is still elusive. It is already known that prominent miRNAs may appear as oncogenic or tumor-suppressive actors in regulation of several genes in cancer-regulatory networks [121]. Though a few attempts for modeling disease-related regulatory networks comprising miRNAs, the explanatory power of current methods are limited. They all start with a couple of assumptions to simplify the computational process by limiting the number of distinct biological processes which may involve in regulatory networks. One of the major reasons of chronic diseases, such as cancer, is dysregulation of the processes which control the specificity, timing and concentration of gene products. These controls largely depend on the cell responses to environmental changes. Recent studies have established that the decision of how cells respond to specific stress conditions are modulated by miRNA functions, which are particularly determined by the amount of miRNAs, abundance of target genes, and activity of miRNA-protein complexes [62]. Together with these functional determinants, analyzing miRNA interactions linked onto known interaction networks, such as Gene Ontology [114], Ingenuity Pathways [76] and KEGG [48], has the potential to improve both the understanding of the pathways as well as the function of the microRNAs in regulation of cancer and other diseases.

Altered expression of miRNA has been shown to promote oncogenesis [35]. Therefore, computational identification of differentially expressed miRNAs from high-throughput profiles promise a great benefit in cancer research. However, small n (number of samples), large p (number of variables or entities) paradigm of high-throughput experiments can limit the generalization ability of current computational methods. These challenges have triggered the use of multi-source, multi-view or multi-task frameworks over probabilistic inference models [39]. This recent trend promises more reliable integration of miRNA and mRNA profiles obtained from several data sources such as microarrays, methylation arrays, ChIP-and-chip experiments, genome-wide association analysis, proteomics and output of other functional genomics analyses [18]. Together with underlying computational models which try to explain biological or biochemical processes, there will be a pressing need to high-level software frameworks which will provide

all-in-pipeline functionality to analyze combinatorial miRNA functions [42]. We anticipate that the number relevant web-based or standalone software tools will rapidly increase in coming years.

CONCLUSION

Several kinds of information are important to analyze miRNA functions; however an information modality alone can be useful to only a certain degree. Therefore, a vital issue concerns the need for accurate and reliable algorithms capable of integrating or synthesizing several information sources to explain the embedded knowledge in static or experimental data, which in turn facilitates generating testable hypothesis about the role of individual microRNAs. The increase in the available expression data has already provided valuable opportunities for data integration. This has enabled the extraction of additional rules for miRNA interactions with target genes. Since miRNAs are involved in several processes in development, differentiation, growth and metabolism, the integration of other types of data, such as intracellular networks, TFs, systems-level measurements and varying environmental conditions also promises to obtain more reliable results in computational analyses.

ACKNOWLEDGEMENT

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under the Project 110E160.

REFERENCES

- [1] Agrawal R, Imielinski T, Swami A. Mining Associations between Sets of Items in Large Databases. *Proc ACM SIGMOD Int Conf Man Data* 1993; 207-216.
- [2] Alexiou P, Maragkakis M, Papadopoulos GL et al. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* 2009; 25:3049-3055.
- [3] Antonov AV, Dietmann S, Wong P et al. GeneSet2miRNA: finding the signature of cooperative miRNA activities in the gene lists. *Nucleic Acids Res* 2009; 37:W323-W328.
- [4] Arvey A, Larsson E, Sander C et al. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol* 2010; 6:363.
- [5] Baek D, Villen J, Shin C et al. The impact of microRNAs on protein output. *Nature* 2008; 455:64-71.
- [6] Bandyopadhyay S and Mitra R. TargetMiner: MicroRNA target prediction with systematic identification of tissue specific negative examples. *Bioinformatics* 2009; 25:2625-2631.
- [7] Barbato C, Arisi I, Frizzo ME et al. Computational Challenges in miRNA Target Predictions: To Be or Not to Be a True Target? *J Biomed Biotechnol* 2009; 2009:803069.
- [8] Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; 136: 215-233.
- [9] Blei D and Jordan M. Modeling annotated data. *Proc Int ACM SIGIR Conf Res Dev Info Ret* 2003; 127–134.
- [10] Blenkiron C and Miska EA. miRNAs in cancer: approaches, aetiology, diagnostics and therapyregulons. *Human Mol Genetics* 2007; 16:R106–R113.
- [11] Bonnet E, Tatari M, Joshi A et al. Module Network Inference from a Cancer Gene Expression Data Set Identifies MicroRNA Regulated Modules. *PLoS ONE* 2010; 5:e10162
- [12] Bonnet E., Michoel T, Van de Peer Y. Prediction of a gene regulatory network linked to prostate cancer from gene expression, microRNA and clinical data. *Bioinformatics* 2010; 26:i638–i644.
- [13] Brameier M. and Wiuf C. Ab initio identification of human microRNAs based on structure motifs. *BMC Bioinformatics* 2007; 8:478.
- [14] Breving K and Kerscher AE. The complexities of microRNA regulation: mirandering around the rules, *Int J Biochem Cell Biol* 2010; 42:1316–1329.
- [15] Burgler C and Macdonald PM. Prediction and verification of microRNA targets by MovingTargets, a highly adaptable prediction method. *BMC Genomics* 2005, 6:88.
- [16] Cazalla D, Yario T, Steitz JA. Down-Regulation of a Host MicroRNA by a Herpesvirus saimiri Noncoding RNA. *Science*. 2010; 328:1563-1566.

- [17] Chandra V, Girijadevi R, Nair AS et al. MTar: a computational microRNA target prediction architecture for human transcriptome. *BMC Bioinformatics* 2010; 11: S2.
- [18] Chen X, Jorgenson E, Cheung ST. New tools for functional genomic analysis. *Drug Discov Today* 2009; 14: 754–760.
- [19] Cheng C and Li LM. Inferring microRNA activities by combining gene expression with microRNA target prediction. *PLoS ONE* 2008; 3: 1–9.
- [20] Creighton CJ, Nagaraja AK, Hanash SM et al. A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *RNA* 2008; 14: 2290–2296.
- [21] Friard O, Re A, Taverna D et al. CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse, *BMC Bioinformatics* 2010; 11:435.
- [22] Friedman RC, Farh KK, Burge CB et al. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; 19: 1-11.
- [23] Gaidatzis D, van Nimwegen E, Hausser J et al.. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* 2007; 8: 69-79;
- [24] Gamazon ER, Im HK, Duan S et al. ExprTarget: An Integrative Approach to Predicting Human MicroRNA Targets. *PLoS ONE* 2010; 5: e13534.
- [25] Gennarino VA, Sardiello M, Avellino R et al. MicroRNA target prediction by expression analysis of host genes. *Genome Res* 2009; 19: 481–490.
- [26] Grad Y, Aach J, Hayes GD et al. Computational and experimental identification of *C. elegans* microRNAs. *Mol Cell* 2003; 11:1253–1263.
- [27] Grimson A, Farh KK, Johnston WK et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; 27:91-105.
- [28] Guo AY, Sun J, Jia P et al. A Novel microRNA and transcription factor mediated regulatory network in schizophrenia. *BMC Systems Biol* 2010; 4:10
- [29] Hammell M, Long D, Zhang L et al. mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Meth* 2008; 5: 813-820.
- [30] Hammell M. Computational methods to identify miRNA targets. *Semin Cell Dev Biol* 2011 (in press) doi:10.1016/j.semcdb.2010.01.004..
- [31] He L, Thomson JM, Hemann MT et al. A microRNA polycistron as a potential human oncogene. *Nature* 2005; 435:828–33.
- [32] Hecker M, Lambeck S, Toepfer S et al. Gene regulatory network inference: Data integration in dynamic models-A review. *Biosystems* 2009; 96: 86-103.
- [33] Hertel J and Stadler PF. Hairpins in a haystack: recognizing microRNA precursors in comparative genomics data. *Bioinformatics* 2006; 22:e197–e202.
- [34] Hofacker IL. How microRNAs choose their targets. *Nature Genet* 2007; 39:1191-1192.
- [35] Hu Y, Correa AM, Hoque A, et al. Prognostic significance of differentially expressed miRNAs in esophageal cancer. *Int J Cancer* 2011; 128:132-43.
- [36] Hu Z. Insight into microRNA regulation by analyzing the characteristics of their targets in humans, *BMC Genomics* 2009; 10:594.
- [37] Huang JC, Babak T, Corson TW et al. Using expression profiling data to identify human microRNA targets. *Nature Meth* 2007; 4: 1045-1049.
- [38] Huang JC, Morris QD, Frey BJ. Bayesian inference of microRNA targets from sequence and expression data. *J Comput Biol* 2007; 14: 550–563.
- [39] Huopaniemi I, Suvitaival T, Nikkilä J et al. Multivariate multi-way analysis of multi-source data. *Bioinformatics* 2010; 26: i391-i398.
- [40] Iwama H, Murao K, Imachi H et al. MicroRNA Networks Alter to Conform to Transcription Factor Networks Adding Redundancy and Reducing the Repertoire of Target Genes for Coordinated Regulation. *Mol Biol Evol* 2011; 28:639-646.
- [41] Jacobsen A, Wen J, Marks DS et al. Signatures of RNA binding proteins globally coupled to effective microRNA target sites. *Genome Res* 2010; 20: 1010–1019.
- [42] Jacobsen A, Krogh A, Kauppinen S et al. miRMaid: a unified programming interface for microRNA data resources, *BMC Bioinformatics* 2010, 11:29.

- [43] Jayaswal V, Lutherborrow M, Ma DF et al. Identification of microRNAs with regulatory potential using a matched microRNA-mRNA time-course data. *Nucl Acids Res* 2009; 37:e60
- [44] John B, Enright A.J, Aravin A et al. Human microRNA targets. *PLoS Bio.* 2004; 2: 363.
- [45] Joung JG and Fei Z. Computational identification of condition-specific miRNA targets based on gene expression profiles and sequence information. *BMC Bioinformatics* 2009; 10: S34.
- [46] Joung JG and Fei Z. Identification of microRNA regulatory modules in Arabidopsis via a probabilistic graphical model. *Bioinformatics* 2009; 25:387-393.
- [47] Joung JG, Hwang KB, Nam JW et al. Discovery of microRNA-mRNA modules via population-based probabilistic learning. *Bioinformatics* 2007; 23: 1141-1147.
- [48] Kanehisa M, Goto S, Furumichi M et al. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010; 38:D355-D360.
- [49] Kertesz M, Iovino N, Unnerstall U et al. The role of site accessibility in microRNA target recognition. *Nature Genet* 2007; 39:1278.
- [50] Khan AA, Betel D, Miller ML et al. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nature Biotech* 2009; 27:549 – 555.
- [51] Kim SK, Nam JW, Rhee JK et al. miTarget: microRNA target gene prediction using a support vector machine. *BMC Bioinformatics* 2006, 7:411.
- [52] Kowarsch A, Marr C, Schmidl D et al. Tissue-Specific Target Analysis of Disease-Associated MicroRNAs in Human Signaling Pathways. *PLoS ONE* 2010; 5:e11154
- [53] Krek A, GrÄun D, Poy MN et al. Combinatorial microRNA target predictions. *Nat Genet* 2005; 37:495-500.
- [54] Krutzfeldt J, Poy MN, Stoffel M. Strategies to determine the biological function of microRNAs. *Nat. Genet* 2006; 38: S14-S19.
- [55] Kundaje A, Miidendorf M, Gao F et al. Combining sequence and time series expression data to learn transcriptional modules. *IEEE/ACM Tr Comput Biol Bioinf* 2005; 2:194-202.
- [56] Larsson E, Sander C, Marks D. mRNA turnover rate limits siRNA and microRNA efficacy. *Mol Sys Biol* 2010; 16:433.
- [57] Lau NC, Lim LP, Weinstein EG et al. An abundant class of tiny RNAs with probable regulatory roles in *aenorhabditis elegans*. *Science* 2001; 294:858-62.
- [58] Lavrac N, Kavsek B, Flach PA et al. Subgroup Discovery with CN2-SD. *Journal of Machine Learning Research* 2004; 5:153-188.
- [59] Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001; 294:862-864.
- [60] Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993; 75: 843-854.
- [61] Lee W and Tzou WS. Computational methods for discovering gene networks from expression data. *Brief Bioinform* 2009; 10:408-423.
- [62] Leung AKL and Sharp PA. MicroRNA Functions in Stress Responses. *Mol Cell* 2010; 40:205-215.
- [63] 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: 15-20.
- [64] Lewis BP, Shih IH, Jones-Rhoades MW et al. Prediction of mammalian microRNA targets. *Cell* 2003; 115:787–798.
- [65] Lim LP, Lau NC, Garrett-Engle P et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 2005; 17:769-773.
- [66] Lindow M, Jacobsen A, Nygaard S et al. Intragenomic Matching Reveals a Huge Potential for miRNA-Mediated Regulation in Plants. *PLoS Comput Biol* 2007; 3:e238.
- [67] Liu B, Li J, Tsykin A et al. Exploring complex miRNA-mRNA interactions with bayesian networks by splitting-averaging strategy. *BMC Bioinformatics* 2009; 10:408.
- [68] Liu B, Li J, Tsykin A. Discovery of functional miRNA-mRNA regulatory modules with computational methods. *J Biomed Infor* 2009; 42:685-691.
- [69] Liu B, Liu L, Tsykin A et al. Identifying functional miRNA-mRNA regulatory modules with correspondence latent dirichlet allocation *Bioinformatics* 2010; 26:3105-3111.

- [70] Liu H, Yue D, Chen Y et al. Improving performance of mammalian microRNA target prediction. *BMC Bioinformatics* 2010b; 11:476.
- [71] Liu H, Yue D, Zhang L et al. A Bayesian approach for identifying miRNA targets by combining sequence prediction and gene expression profiling. *BMC Genomics* 2010a; 11:S12.
- [72] Lu J, Getz G, Miska EA et al. MicroRNA expression profiles classify human cancers. *Nature* 2005; 435: 834–838.
- [73] Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci USA* 2007; 104:9667–72.
- [74] Madden SF, Carpenter SB, Jeffery IB et al. Detecting microRNA activity from gene expression data, *BMC Bioinformatics* 2010; 11:257.
- [75] Maragkakis M, Reczko M, Simossis VA et al. DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acid Res* 2009; 37: W273-W276
- [76] Marín AJ, Romero MC, Boo, MR et al. Biological pathway analysis by ArrayUnlock and Ingenuity Pathway Analysis. *BMC Proceedings* 2009, 3(Suppl 4):S6
- [77] Marin RM and Vaníček J. Efficient use of accessibility in microRNA target prediction. *Nucl Acids Res* 2010; doi:10.1093/nar/gkq768.
- [78] Maziere P, Enright A. Prediction of microRNA targets. *Drug Discovery Today* 2007; 12:452-458.
- [79] Mendes ND, Freitas AT, Sagot MF. Current tools for the identification of miRNA genes and their targets. *Nucleic Acid Res* 2009; 37: 2419-2433.
- [80] Miranda KC, Huynh T, Tay Y et al. A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell* 2006; 126: 1203-1217.
- [81] Nam S, Kim B, Shin S, Lee S. miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res* 2008; 36: D159–D164.
- [82] Nam S, Li M, Choi K et al. MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucl Acids Res* 2009; 37:W356-W362.
- [83] Nielsen CB, Shomron N, Sandberg N et al. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 2007; 13:1894–1910.
- [84] Ohler U, Yekta S, Lim LP et al. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* 2004, 10:1309–1322.
- [85] 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; 10:373.
- [86] Pillai RS. MicroRNA function: Multiple mechanisms for a tiny RNA? *RNA* 2005; 11:1753–1761.
- [87] Place RF, Li LC, Pookot D et al. MicroRNA-373 induces expression of genes with complementary promoter sequences. *PNAS* 2008; 105: 1608-1613.
- [88] Poliseno L, Salmena L, Zhang J et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010; 465:1033–103
- [89] Qiu C, Wang J, Yao P et al. MicroRNA evolution in a human transcription factor and microRNA regulatory network, *BMC Systems Biology* 2010; 4:90.
- [90] Ragan C, Cloonan N, Grimmond SM et al. Transcriptome-Wide Prediction of miRNA Targets in Human and Mouse Using FASTH. *PLoS ONE* 2009; 4: e5745.
- [91] Rajewsky N. microRNA target predictions in animals. *Nat Genet* 2006; 38: S8–13.
- [92] Re A, Cora D, Taverna D et al. Genome-wide survey of microRNA-transcription factor feed-forward regulatory circuits in human. *Mol Biosyst* 2009; 5: 854–67.
- [93] Rehmsmeier M, Steffen P, Hochsmann M et al. Fast and effective prediction of microRNA-target duplexes. *RNA* 2004; 10: 1507-1517.
- [94] Ritchie W, Rajasekhar M, Flamant S et al. Conserved expression patterns predict microRNA targets. *PLOS Comput Biol* 2009; 5: e1000513.
- [95] Rodriguez A, Griffiths-Jones S, Ashurst JL et al. Identification of mammalian microRNA host genes and transcription units. *Genome Res* 2004; 14:1902-1910.
- [96] Ruan J, Chen H, Kurgan L et al. HuMiTar: A sequence-based method for prediction of human microRNA targets, *Alg Mol Biol* 2008; 3:16.

- [97] Rusinov V, Baev V, Minkov IN et al. MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence. *Nucleic Acids Res* 2005; 33:W696-W700.
- [98] Ruvkun G. Molecular biology: Glimpses of a tiny RNA world. *Science* 2001; 294: 797-799.
- [99] Saetrom O, Snøve O Jr, Saetrom P. Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms. *RNA* 2005; 11: 995-1003.
- [100] Saito T, Saetrom P. MicroRNAs - targeting and target prediction. *New Biotechnology* 2010; 27: 243-249.
- [101] Sales G, Coppe A, Bisognin A et al. MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Res.*2010; 38:W352-W359.
- [102] Segal E, Shapira M, Regev A et al.. Learning module networks, *Journal of Machine Learning Research* 2005; 6: 557-588.
- [103] Seitz H. Redefining microRNA targets. *Curr Biol* 2009; 19: 870–873.
- [104] Selbach M, Schwanhaussner B, Thierfelder N et al. Widespread changes in protein synthesis induced by microRNAs. *Nature* 2008; 455: 58-63.
- [105] Shalgi R, Lieber D, Oren M et al. Global and Local Architecture of the Mammalian microRNA–Transcription Factor Regulatory Network. *PLoS Comput Biol* 2007; 3: e131
- [106] Sinha AU, Kaimal V, Chen J et al. Dissecting microregulation of a master regulatory network. *BMC Genomics* 2008, 9:88.
- [107] Slack FJ, Big roles for small RNAs. *Nature* 2010; 463: 616.
- [108] Stark A, Brennecke J, Russell RB et al. Identification of drosophila microRNA targets. *PLOS Biol* 2003; 1: E60.
- [109] Steyvers M., Smyth P, Zvi MR et al. Probabilistic author-topic models for information discovery. *Proc ACM SIGKDD Int Conf Know Dis Data Min* 2004.
- [110] Sturm M, Hackenberg M, Langenberger D et al. TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics* 2010; 11:292
- [111] Szabo PM, Tombol Z, Molnar V et al. MicroRNA Target Prediction: Problems and Possible Solutions. *Current Bioinformatics* 2010; 5:81-88.
- [112] Tay Y, Zhang J, Thomson AM et al. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 2008; 455:1124–8.
- [113] Thadani R and Tammi,MT. MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics* 2006; 7:S20.
- [114] The Gene Ontology Consortium. The Gene Ontology project in 2008. *Nucl Acids Res* 2008;b36: D440–4.
- [115] Tran DH, Satou K, Ho TB. Finding microRNA regulatory modules in human genome using rule induction. *BMC Bioinformatics* 2008, 9:S5.
- [116] Tsang J, Ebert MS, van Oudenaarden A. Genome-wide Dissection of MicroRNA Functions and Cotargeting Networks Using Gene Set Signatures. *Mol Cell* 2010; 38:140-153.
- [117] Tsang J, Zhu J, van Oudenaarden A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals *Mol Cell* 2007; 26:753–767.
- [118] Ulitsky I, Laurent LC, Shamir R. Towards computational prediction of microRNA function and activity. *Nucl Acids Res* 2010; doi:10.1093/nar/gkq570.
- [119] Vasudevan S, Tong Y, Steitz JA. Switching from Repression to Activation: MicroRNAs Can Up-Regulate Translation. *Science* 2007; 318:1931-1934.
- [120] Vohradsky J, Panek J, Vomastek T. Numerical modelling of microRNA-mediated mRNA decay identifies novel mechanism of microRNA controlled mRNA downregulation. *Nucl Acids Res* 2010; 38: 4579-4585
- [121] Voorhoeve PM. MicroRNAs: Oncogenes, tumor suppressors or master regulators of cancer heterogeneity. *Bioc Bioph Acta* 2010; 1805: 72-86.
- [122] Wang J, Lu M, Qiu C et al. TransmiR: a transcription factor–microRNA regulation database. *Nucl Acids Res* 2010a; 38: D119-D122.

- [123] Wang X and Naga IM. Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics* 2008; 24: 325-332.
- [124] Wang X and Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acid Res* 2006; 34: 1646-1652.
- [125] Wang X, Zhang J, Li F et al. MicroRNA identification based on sequence and structure alignment. *Bioinformatics* 2005; 21: 3610–3614.
- [126] Wang Y, Li Y, Ma Z et al. Mechanism of MicroRNA-Target Interaction: Molecular Dynamics Simulations and Thermodynamics Analysis. *PLoS Comput Biol* 2010b; 6: e1000866.
- [127] Weaver D, Anzola J, Evans J et al. Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol* 2007; 8:R97.
- [128] Wu X, Watson M. CORNA: testing gene lists for regulation by microRNAs. *Bioinformatics* 2009; 25: 832-833.
- [129] Xiao F, Zuo Z, Cai G et al. miRecords: an integrated resource for microRNA-target interactions, *Nucleic Acid Res* 2009; 37: D105-D110.
- [130] Xie X, Lu J, Kulbokas E.J. et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 2005; 434: 338–345.
- [131] Yang Y, Wang YP, Li KB. MiRTif: a support vector machine-based microRNA target interaction filter. *BMC Bioinformatics* 2008; 9:S4.
- [132] Ye W, Lv Q, Wong CK et al. The effect of central loops in miRNA:MRE duplexes on the efficiency of miRNA-mediated gene regulation. *PLoS ONE* 2008; 3:e1719.
- [133] Yoon S and Micheli G. Prediction of regulatory modules comprising microRNAs and target genes, *Bioinformatics* 2005; 21: i93-i100.
- [134] Yousef M, Jung S, Kossenkova AV et al. Naïve Bayes for microRNA target predictions—machine learning for microRNA targets. *Bioinformatics* 2007; 23: 2987-2992.
- [135] Yousef M, Showe L, Showe M. A study of microRNAs in silico and in vivo: bioinformatics approaches to microRNA discovery and target identification. *FEBS Journal* 2009; 276:2150–2156.
- [136] Yue D., H. Liu, Y. Huan. Survey of Computational Algorithms for MicroRNA Target Prediction. *Current Genomics* 2009; 10: 478-492.
- [137] Zhang Y and Verbeek FJ. Comparison and Integration of Target Prediction Algorithms for microRNA Studies. *J Integrative Bioinform* 2010; 7:127.