From miRNA regulation to miRNA - TF co-regulation: computational approaches and challenges

[1,*]Thuc Duy Le, [1]Lin Liu, [2]Junpeng Zhang, [3]Bing Liu, and [1,*]Jiuyong Li

[1]School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia

[2]Faculty of Engineering, Dali University, Dali, China

[3]Children's Cancer Institute Australia for Medical Research, Lowy Cancer Research Centre, Randwick, NSW 2031, Australia

[*]Contact Author: Thuc.Le@unisa.edu.au and Jiuyong.Li@unisa.edu.au

Biographical note:

Thuc Duy Le is a research associate at the University of South Australia (UniSA). He received his BSc (2002) and MSc (2006) in pure Mathematics from the University of Pedagogy, Ho Chi Minh City, Vietnam, and BSc (2010) in Computer Science from UniSA. He completed his PhD thesis in Bioinformatics (January 2014) at UniSA. His research interests are Bioinformatics, data mining, and machine learning.

Lin Liu is a senior lecturer at the School of Information Technology and Mathematical Sciences, University of South Australia (UniSA). She received her bachelor and master degrees in Electronic Engineering from Xidian University, China in 1991 and 1994 respectively, and her PhD degree in computer systems engineering from UniSA in 2006. Dr Liu's research interests include data mining and bioinformatics, as well as Petri nets and their applications to protocol verification and network security analysis.

Junpeng Zhang is a teaching assistant at the Faculty of Engineering, Dali University. He received his BSc (2009) in Bio-medical Engineering and MSc (2012) in Control Theory and Control Engineering from Kunming University of Science and Technology, Kunming City, China. His research interests include bioinformatics and data mining.

Bing Liu is a bioinformatician at Children's Cancer Institute Australia (CCIA) for medical research. CCIA is the only independent medical research institute in Australia devoted to research into the causes, prevention, better treatment and ultimately a cure of childhood cancer. Dr. Bing Liu received his Bsc (1997) and Msc (2000) in electronics and computer engineering from the Yunnan University, China, and Phd (2010) in computer science from the University of South Australia. Before he joined CCIA, he has worked as a research fellow at the Centre for Cancer Biology, Adelaide and the University of Newcastle, Australia. His major research areas are in bioinformatics and data mining, particularly integrating heterogeneous data for biological/medical research.

Jiuyong Li is a professor at the School of Information Technology and Mathematical Sciences, University of South Australia. He received his BSc degree in physics and MPhil degree in information processing from the Yunnan University, China in 1987 and 1998, respectively, and received his PhD degree in computer science from the Griffith University, Australia (2002). His research interests are in the field of data mining, privacy preserving and Bioinformatics. His research has been supported by five prestigious Australian Research Council Discovery grants since 2005.

Key points:

- microRNAs and transcription factors are important regulators, but few feasible experimental techniques are available for exploring all of their functions. Computational methods can generate hypotheses about miRNA and TF targets, narrowing down the vast amount of possibilities to be tested by wet lab experiments.
- Integrating multiple sources of data and/or multiple types of data enhances the capability to identify miRNA functions.
- Exploring miRNA and TF co-regulation gives more insights into the causes of diseases. Some recent methods utilise both target information and gene expression profiles to construct the regulatory network with the presence of miRNAs, TFs, and genes.
- Although several computational methods are presented and made available, it is challenging to evaluate them and decide which method is better than the others as they are more often complementary to one another.

Keywords

miRNA, transcription factor, causality discovery, data integration, differential analysis, miRNA target, co-regulation, regulatory network, model evaluation, selection of models.

Abstract

microRNAs (miRNAs) are important gene regulators. They control a wide range of biological processes and are involved in several types of cancers. Thus, exploring miRNA functions is important for diagnostics and therapeutics. To date, there are few feasible experimental techniques for discovering miRNA regulatory mechanisms. Alternatively, predictions of miRNA-mRNA regulatory relationships by computational methods have increasingly achieved promising results. Computational approaches are proving their ability as effective tools in reducing the number of biological experiments that must be conducted and to assist with the design of the experiments. In this review, we categorise and review different computational approaches to identify miRNA activities and functions, including the co-regulation of miRNAs and TFs. Our main focuses are on the recent approaches that utilise multiple data types for exploring miRNA functions. We discuss the remaining challenges in the evaluation and selection of models based on the results from a case study. Finally, we analyse the remaining challenges of each computational approach and suggest some future research directions.

INTRODUCTION

The human genome contains more than twenty thousand genes. Most of these genes are expressed differentially to create proteins, the main actors in a living cell. The gene expression program is an extremely complex and well-organised procedure. While the expression program is made up of individual genes expressing on their own mechanisms, together they create a unified big picture to ensure that cells can function properly. Any interference to the expression program of genes would result in diseases. Therefore, understanding the mechanisms of regulating gene expression is crucial for preventing and curing diseases.

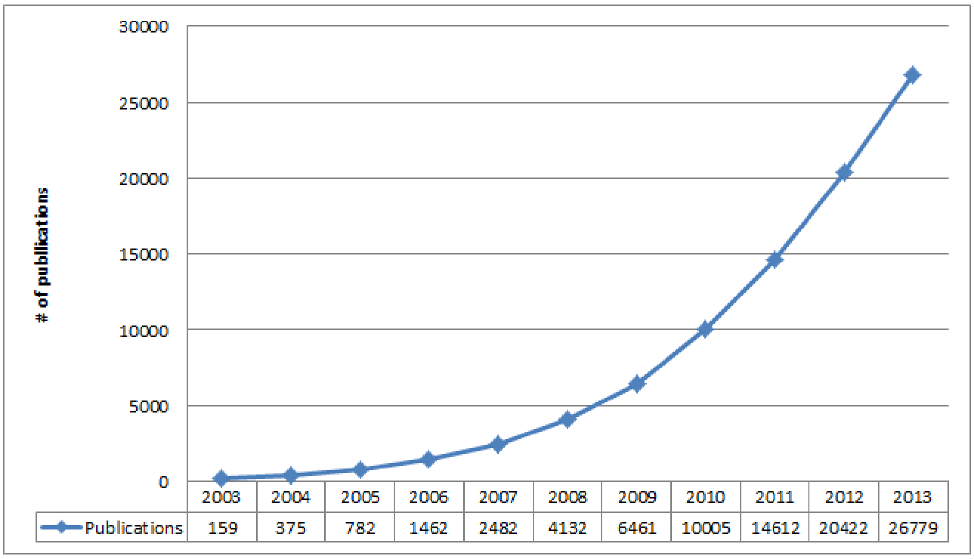
At the transcription level, transcription factors (TFs) are known to be the main gene regulators. TFs are proteins that control gene expression by directly binding to the DNA sequences of the target genes. The binding will result in an increase or decrease in the expression levels of the target genes. There has been extensive research on the regulation of TFs, and many experimental and computational methods have been proposed to elucidate the regulatory mechanisms of this type of regulators (see [1] for a review).

Recently, microRNAs (miRNAs) have been discovered to be the main regulators at post-transcriptional level. miRNAs are small nucleotide sequences of 21-25 bases [2] which are transcribed from the non-coding parts of the DNA. They recognise target genes by base pairing to complementary sequences in the 3'-untranslated region (3'UTR), 5’UTR, and sometimes in the open reading frames of the target mRNA.

More and more evidence has emerged on miRNA involvements in biological processes and diseases. By regulating their target genes, miRNAs control a wide range of biological processes such as proliferation [3, 4], metabolism [5], differentiation [6], development [7], apoptosis [8], cellular signaling [9] and even cancer development and progression [2, 10]. Therefore, it is not surprising that miRNAs have been identified to be involved in several types of cancers including breast cancer [11], prostate cancer [12], lung cancer [13], colon cancer [14], ovarian cancer [15], and many other diseases [10, 16-19].

One of the challenges in miRNA research is that miRNAs have their own characteristics, making it difficult or impossible to apply the experimental and computational methods used for other gene regulators such as TFs, to miRNA research. Like TFs, one miRNA can regulate multiple target genes simultaneously and multiple miRNAs may regulate their target genes cooperatively. A single miRNA can control up to hundreds of target genes [20], and most target genes in the genome can be regulated by multiple miRNAs [21]. Unlike TFs, however, miRNAs are very short in length (21-24 nucleotides), and the gene regulatory region of a miRNA is small in size, less than 1kb compared to dozens of kbs in the case of a TF [22]. The small target regulatory region poses challenges for computational methods that search for the base pairing between miRNAs and the binding sites, as the complementary region is too short to be statistically significant. Therefore novel and appropriate methods especially designed for miRNAs are required.

The significant roles of miRNAs have given rise to a fast growing body of research over the last decade (see Figure 1 for an illustration). In the early years of miRNA research, great effort was made to discover new miRNAs as well as to explore miRNA functions with wet experiments (see [23] for a review). These studies have significantly improved our understanding of miRNA functions. On the other hand, they also pose new challenges to scientists by producing a tremendous amount of data that needs to be analysed. Given the increasing number of novel miRNAs discovered (coupled with the large number of their regulated targets), it is infeasible to validate all the possible regulatory interactions by means of biological experiments. In response to the calls for discovering new knowledge from the vast amount of data available, computational methods have been developed and many of them have proved to be effective tools in assisting with the design of wet experiments, short-listing statistically significant regulatory interactions, thus making it feasible to conduct validation experiments [23, 24]. Although computational methods may never replace wet-lab experiments, they will assist with the design of the experiments.



	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Publications	159	375	782	1462	2482	4132	6461	10005	14612	20422	26779

Figure 1. The number of miRNA-related publications accumulated in the past decade. The number of miRNA-related publications from PubMed library with the keyword “miRNA".

A variety of computational methods have been proposed to explore miRNA functions, ranging from sequence based methods to methods incorporating expression data or combining multiple sources of data; correlation based methods to causality discovery based methods; methods for discovering single interactions to those finding modules of interacting molecules; analyses of interactions in a specific condition to differential analyses involving multiple conditions; and methods of studying miRNA regulation alone to those considering cooperative effects of miRNAs and TFs. These methods provide complementary views and approaches to exploring miRNA functions, and have their own advantages and limitations.

Table 1. Summary of computational approaches for inferring miRNA functions. The grey areas are the topics have been reviewed elsewhere. The areas with ticks are the topics will be covered in this review.

	Sequence data	**Expression data**	**Multiple data sources**
miRNA-mRNA regulatory relationships	See [23, 24] for reviews	• Classical approaches: See [23] for a review • Emerging approaches ✓	✓
miRNA-mRNA regulatory modules	See [24] for a review	See [24] for a review	✓
miRNA-TF co-regulation	✓	✓	✓

In this paper, we review the new work for exploring (human) miRNA functions complementary to the previous reviews and discuss challenging issues in model evaluation. We categorise the methods into three main approaches: inferring miRNA-mRNA regulatory relationships, identifying miRNA-mRNA regulatory modules, and discovering miRNA-TF co-regulatory relationships. Table 1 shows the overall picture of the computational approaches, including the topics that have been reviewed elsewhere [23, 24], and the topics that will be covered in this review. We discuss the remaining challenges of the evaluation and selection of models with a case study on three real world cancer datasets. Finally, we discuss some future research directions.

INFERRING miRNA-mRNA REGULATORY RELATIONSHIPS

Identifying miRNA targets is the first and foremost task to understand miRNA functions and regulatory mechanisms. Therefore, there have been several methods developed to predict putative target mRNAs at the sequence level [20, 25, 26]. The methods have contributed to the understanding of the regulatory functions of miRNAs, and have helped to narrow down the otherwise immeasurable amount of experimental work to be conducted. However, these approaches have a high rate of false positive predictions, and do not take biological context into account. These approaches have been reviewed in [23, 24], and we do not cover them in this review.

In recent years, with the advances in experimental technology, gene expression data has emerged as important and promising resources for exploring miRNA functions. Various computational methods have been devised to incorporate gene expression profiles into the study of miRNA-mRNA regulatory relationships. Figure 2 shows a common framework used by these existing methods. In this framework, matched samples of miRNA and mRNA expression data are integrated into a dataset. To reduce the computational complexity, researchers apply some constraints such as differentially expressed gene analysis to reduce the number variables (genes) in the dataset, or use sequence based target predictions to reduce the number of miRNA-mRNA interactions considered by the method. The input expression dataset is then analysed by statistical/machine learning models to predict miRNA targets. Recent advances focus on integrating multiple expression datasets, or using heterogeneous data such as over-expression data and sequence data in building a model. There are also some newly emerged methods for inferring miRNA targets, such as the methods based on causality discovery approach. In this section, we review these new works (summarised in Table 2) for inferring miRNA-mRNA regulatory relationships.

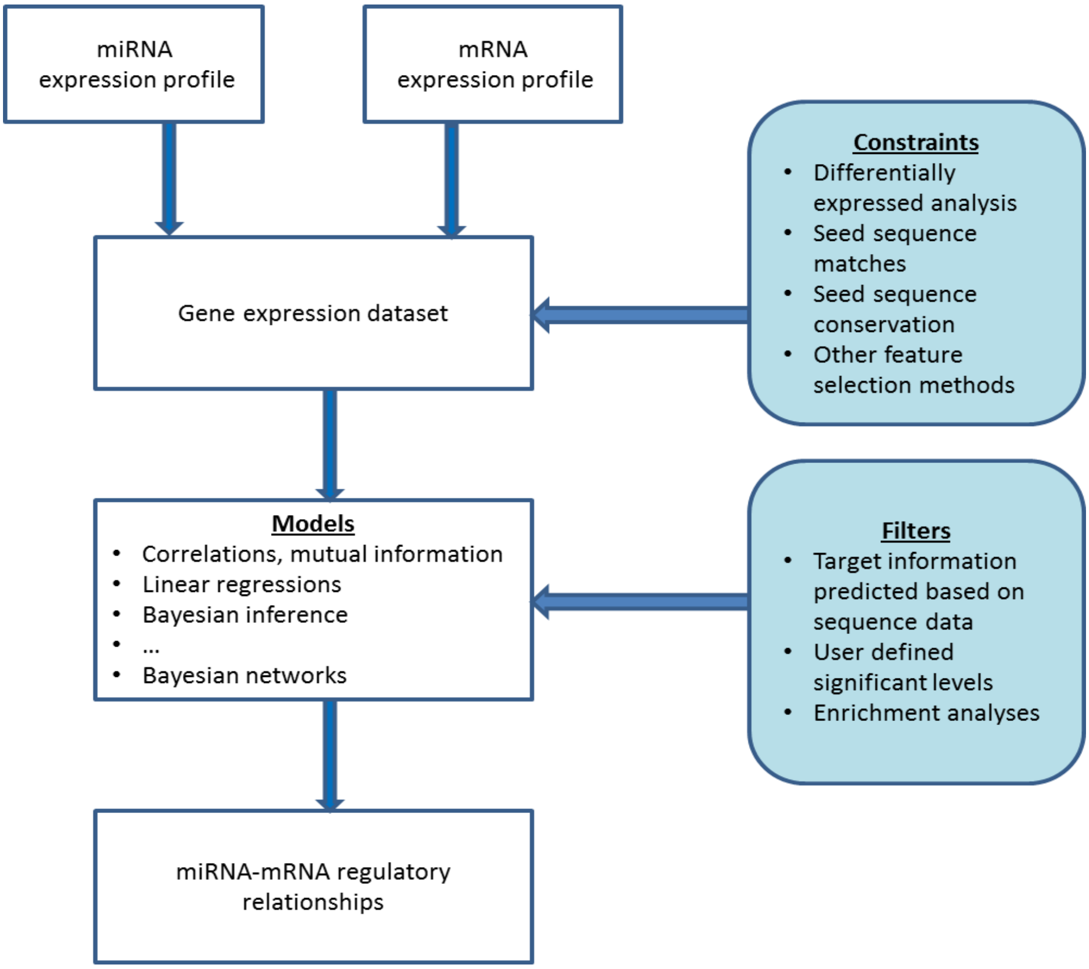

miRNA expression profile
mRNA expression profile
Constraints
• Differentially expressed analysis
• Seed sequence matches
• Seed sequence conservation
• Other feature selection methods
Gene expression dataset
Models
• Correlations, mutual information
• Linear regressions
• Bayesian inference
• …
• Bayesian networks
Filters
• Target information predicted based on sequence data
• User defined significant levels
• Enrichment analyses
miRNA-mRNA regulatory relationships

Figure 2. A common framework of existing computational methods. The matched miRNA and mRNA expression profiles are integrated into a dataset. A set of constraints can be firstly applied to this dataset for feature selection purpose. These constraints include differentially expressed gene analysis, sequence based techniques such as seed sequence matching and seed sequence conservation. The gene expression dataset is then input into a statistical model to predict the miRNA-mRNA regulatory relationships. The predictions can be further filtered using target information, statistically significant thresholds or enrichment analyses such as pathway analysis.

Classical approaches with gene expression data and beyond

Classical approaches such as correlation, regression analysis, and Bayesian parameter learning, have achieved significant results in inferring miRNA-mRNA regulatory relationships. The principle of these methods is that if a gene is regulated by a miRNA, a correlation should show between the expression levels of the gene and the miRNA. Some of these methods also take into consideration the available miRNA target information previously predicted using sequence data, which has proved to reduce the false discoveries compared to sequence based approaches. Details of these approaches have been reviewed elsewhere [23]. Recent advances in these approaches concern the problem of integrating heterogeneous datasets.

Table 2. Summary of methods to infer miRNA-mRNA regulatory relationships

Method	**Brief description**	**Data sources**	**Notes**	**Software tool**
Chen et al. [27]	Combine different gene expression datasets into a model to infer miRNA-mRNA interactions. • Calculate the correlation in expression levels between miRNA and mRNA in each individual dataset. • Use Empirical Bayes model for integrating the correlation coefficients to output the ranked interactions	• Multiple matched miRNA and mRNA expression datasets	• Suitable for analysing multiple expression datasets of the same condition, e.g. cancer.	http://bioinformatics.med.yale.edu/group/
Jacobsen et al. [28]	Use regression analysis to analyse miRNA-target interactions across diverse cancer types. • Utilise heterogeneous data including DNA copy-number, promoter methylation, and expression data from TCGA to generate a good resource for exploring the recurrent miRNA-mRNA interactions.	• DNA copy-number, • promoter methylation • miRNA and mRNA expression data	• A good resource for biologists to explore the miRNA-mRNA interactions that occur in multiple cancer types. • Given the cancer types of interest, users can query the top miRNA-mRNA recurrent interactions.	http://cancerminer.org
Li et al. [29]	Integrate miRNA-overexpression data and target information to predict miRNA-mRNA interactions. • Curate the miRNA overexpression data from literature • Use Variational Bayesian-Gaussian Mixture Model to integrate the score from overexpression data and from sequence based prediction methods	• miRNA-overexpression, • target information based on sequence data	• A good resource of miRNA overexpression data • When new overexpression data is available, users can apply the method to predict miRNA targets.	http://www.bioconductor.org/packages/devel/bioc/html/TargetScore.html
Liu et al. [30]	Use Bayesian network learning in split samples to learn miRNA-mRNA interactions. • Apply Bayesian network learning for each condition of the dataset • Integrate the results from different conditions	• Target information based on sequence data, • matched miRNA and mRNA expression data	• Suitable for datasets with multiple conditions	Upon request
Le et al. [31]	• Use a causality discovery based method to infer the causal effect that a miRNA has on a mRNA. • Learn the regulatory network from expression data • Simulate the intervention procedure to estimate the causal effect that a miRNA has on a mRNA	• Matched miRNA and mRNA expression data	• A good method of simulating the gene knockdown experiments • Computational complexity is high	http://bioinformatics.oxfordjournals.org/content/29/6/765/suppl/DC1
Liang et al. [32]	Explore miRNA activities in different conditions of the datasets • Infer the activity of miRNA in a sample and then analyse the overall behaviour of the miRNA activity in samples with different biological conditions	• miRNA and mRNA expression data	• Suitable for finding active miRNAs in expression datasets with multiple conditions	http://sysbio.ustc.edu.cn/software/mirAct
Amar et al. [33]	Identify a group of differential co-expression genes • Identify groups of genes that are co-express in all samples, but express differently between two conditions of interest. • Assume that these groups may be the targets of a miRNA family	• miRNA and mRNA expression data	• Can be used to find a group of biomarkers for a condition of interest.	http://acgt.cs.tau.ac.il/dicer/

Chen et al. [27] proposed a method to combine different gene expression datasets into a model for exploring miRNA-mRNA regulatory relationships. The main argument of the method is that analysing the relationships from multiple datasets at the same time may improve the identification of miRNA-target interactions. The method firstly used Pearson correlation coefficient method to calculate the correlation of miRNA-mRNA pairs in each individual dataset. It then used the empirical Bayes approach to incorporate shared information between datasets for the identification of miRNA-mRNA interactions. Experimental results on both simulation studies and real world cancer datasets proved that the method is better than the ones that use only one dataset or simple aggregated dataset.

Meanwhile, Jacobsen et al. [28] used regression analysis to analyse miRNA-target interactions across diverse cancer types. The method firstly inferred the miRNA targets in individual cancer types using a multivariate linear model. Apart from miRNA and mRNA expression profiles, the multivariate linear model takes into consideration the effects of DNA copy-number and promoter methylation at the mRNA gene locus. Specifically, the mRNA expression is represented as a linear function of miRNA expression, DNA copy number changes (log2 tumor/normal ratio), and promoter methylation, i.e. methylation beta-value of the selected methylation probe. The selected methylation probe is the one that shows the strongest negative correlation between its beta value and the gene mRNA expression across all samples in a cancer type. It is found that the miRNA-mRNA pairs predicted by the method largely overlap with those predicted by miRanda [34] and TargetScan [20]. To explore the recurrence of target association across cancer types, the authors proposed a statistical score (REC score) to rank the miRNA-mRNA expression associations. The top ranked miRNA-mRNA interactions are those that have strong association across different cancer types. They applied the method to 11 different cancer types in The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp), and reported all the results as an online resource at http://cancerminer.org.

In contrast to the above methods which used expression data, Li et al. [29] developed a method called TargetScore to integrate miRNA-overexpression data and target information based on sequence data. The method firstly compiled 113 miRNA transfected experiment data from 84 Gene Expression Omnibus (GEO) datasets and calculated the gene expression fold-changes in each experiment. The authors then used the Variational Bayesian-Gaussian Mixture Model to integrate the scores from sequence-based prediction methods and the miRNA-overexpression data. The prediction results were validated using mirTarBase [35], proteomic data in Baek et al. [36], and gene functional enrichment analysis. The transfection data and source codes are available at Bioconductor: http://www.bioconductor.org/packages/devel/bioc/html/TargetScore.html.

These methods effectively apply classical approaches to different types of data, which increasingly become available, to explore miRNA-target relationships. In another direction, researchers have also explored novel approaches for elucidating miRNA functions by designing and applying the emerging data mining and machine learning techniques.

Emerging approaches

The first emerging approach is to infer miRNA-target relationships using causality discovery based methods. The main argument of this approach over the classical association approach is that correlations or associations are not necessarily causality. For instance, the expression values of a miRNA and a mRNA may be strongly correlated across a set of samples, but it is not sufficient to conclude that the miRNA regulates the mRNA. The strong correlation between the miRNA and the mRNA may be a result of the mRNA regulating the miRNA, or a third molecule such as a TF regulating both the miRNA and the mRNA.

The gold standard for tackling the causality discovery problem is randomised control experiments. For example, we can use gene knockdown experiments to knock down miRNAs one by one whilst measuring the changes (i.e. causal effects) in the expression level of mRNAs. However, such experiments are time consuming, expensive, and not necessarily definitive. In computer science, the main theory of discovering causality from data is the theory of causal Bayesian networks. However, the computational complexity of Bayesian network learning algorithms is very high, and thus learning the causal network becomes impractical in high dimensional datasets like gene expression datasets. In application, researchers usually integrate some biological knowledge into the Bayesian network learning process to reduce the complexity.

For instance, Bayesian network structure learning was used in [30] to discover the regulatory relationships between miRNAs and mRNAs. Bayesian network learning algorithm searches for all possible networks in the form of directed acyclic graphs and uses a scoring function to score each graph based on observational data. In this work, the authors assumed that there was a bipartite of interactions between the group of miRNAs and group of mRNAs (miRNAs regulate mRNAs) and there were no interactions between molecules in the same group. Target information predicted with sequence-based methods was used to initialise the network (bipartite). The gene expression profiles of miRNAs and mRNAs were then discretised and input into the Bayesian network learning algorithm [37-39]. The innovation of this work is that the authors used target information to restrict the search space for the computational expensive Bayesian network learning algorithm. Additionally, they split the samples according to their conditions (i.e. normal and cancer) and infer the top interactions under each condition. The final results are the merge of the interactions in all conditions.

However, the method assumes the bipartite of interactions between miRNAs and mRNAs. This assumption may limit the ability to interpret the results and may not necessarily hold in reality. For example, a transcription factor may regulate other mRNAs and even may regulate miRNAs [40]. Therefore, it is necessary to consider different kinds of interactions, e.g. TF regulates miRNA. Ideally, we should assume that every molecule could interact with each other when building a computation model.

Recently, Le et al. [31] adapted a causal discovery approach, IDA (intervention-calculus when the DAG is absent) [41, 42], to uncover the causal regulatory relationship between miRNAs and mRNAs. The method firstly learnt the causal structure from the expression profiles of miRNAs and mRNAs using PC (Peter & Clark) algorithm [43, 44]. *do-calculus* [38, 42] was then used to estimate the causal effect that a miRNA has on a mRNA. The estimated causal effects simulate the effects of randomised controlled experiments. The method tackles two drawbacks of current miRNA regulatory relationships research. Firstly, the method discovers causal relationships between miRNAs and mRNAs, not just the statistical relationships. Secondly, the method assumes that miRNAs and mRNAs interact with each other in a complex system; for instance, a miRNA can causally regulate mRNAs as well as other miRNAs. This assumption is more reasonable than the assumption from commonly used approaches that considers only the bipartite of interactions between miRNAs and mRNAs. For example, Zisoulis et al. [45] shows that let-7 can regulate other non-coding RNAs including miRNAs. However, the method has high computational complexity and can only infer the lower bounds of the real causal effects.

Another emerging approach is utilising differential analyses. To understand the causes of a disease, it often requires analysing the differences between normal and disease samples. For example, differentially expressed analysis identifies genes that express differently in different conditions, and thus reveal possible biomarkers for a disease. Recent advances have seen several forms of differential analyses for different purposes, such as differential co-expression and differential networking. Differential co-expression analysis identifies the groups of co-expressed genes that differ markedly between disease and control samples [33, 46]. Meanwhile, differential networking goes a step further to identify the differences in the gene regulatory networks between diseased and healthy conditions.

In this research stream, researchers infer miRNA activity changes in two biological conditions. Highlights in this direction are miReduce [47], DIANA-mirExTra [48], Sylamer [49], MIR [50]. The common feature of these methods is that they firstly infer the differences in gene expression levels in the two biological conditions, then correlate those alterations with the miRNA binding motifs predicted based on sequence data.

Contrastingly, Liang et al. [32] proposed a method (mirAct) to explore the miRNA activity in a sample and then analyze the overall behavior of the miRNA activity in samples with different biological conditions. Meanwhile, Amar et al. [33] proposed a method called DICER to detect differential co-expression in disease and control samples. They hypothesized that changes in co-expression may be the result of changes in regulatory patterns, and thus the discovered differential co-expression may be the targets of specific miRNA families. To test the approach, they identified miRNA families whose targets are enriched in the gene groups detected by their method, and tested whether those miRNAs are associated with the relevant diseases.

As suggested in [46], the differential networking may be a promising approach to elucidate the causes of diseases. An example of such approach is a work in protein-protein interaction networks [51]. The method identifies the differences of the hubs of protein-protein interactions between two conditions of interest. They hypothesized that the hubs of interactions play an important role in the regulation networks and the differences in the interaction behaviors may provide some clues for the causes of diseases. However, these differential networking techniques are rare in miRNA research.

DISCOVERING miRNA-mRNA MODULES BY INTEGRATING HETEROGENEOUS DATA SOURCES

It is important to know the modular organisation of the regulatory networks, as the recognition of these structures advances our understanding of the complex systems [52-54]. However, this is a challenge in the case of miRNA-gene interactions, as each miRNA can regulate a large number of genes, and multiple miRNAs can regulate the same gene. Therefore, researchers aim to search for a set of miRNAs and their co-regulated genes [55].

In the first stream of research, researchers identify the groups of co-expressed miRNAs and mRNAs using sequence data, or integrate sequence and expression data with or without taking the biological conditions of the datasets into consideration (see [24] for a review). However, these methods utilise only one or two data resources (sequence and expression data), and are thus sensitive to the data noise as pointed out in [55]. As each data type provides different but complementary information, recent studies (Table 3) integrate multiple sources of data into inferring miRNA-gene regulatory modules.

Zhang et al. [55] proposed a framework to identify miRNA-gene regulatory modules by integrating miRNA target predictions based on sequence data, miRNA and gene expression profiles, protein-protein interaction and DNA-protein interaction networks. In their work, sequence-based miRNA target predictions were considered as a static prior network, and expression profiles were used subsequently to identify the active miRNA-gene interactions. These active interactions were further refined by the gene-gene interaction networks (protein-protein and DNA-protein networks). Applied to the human ovarian cancer samples from TCGA, the method discovered several miRNA-gene regulatory modules. The results are then validated against miRNA cluster from miRBase (http://www.mirbase.org/), and the mRNAs are validated using gene functional enrichment analysis.

Similarly, Le et al. [56] proposed a regression-based method to integrate sequence, expression, and protein interactions data for identifying modules of miRNAs and mRNAs for a specific condition. The authors firstly used a regression model to link the expression profiles of miRNAs and mRNAs. The assumption was that the expression level of a mRNA can be represented as a linear function of expression profiles of all predicted miRNAs. The predicted miRNA regulators for a mRNA were taken from sequence-based prediction databases. To assign miRNAs and mRNAs into a module, they designed a function to measure the strength of the predicted miRNA-mRNA interactions based on the information from miRNA target information and protein-protein interactions. Specifically, the assigning function is based on the logistic-sigmoid function with parameters to adjust the contributions of the two types of interaction data. The higher the probabilities of interactions are, the more chances the interacting entities are assigned into the same module. The method was applied to multiple cancer datasets from TCGA to explore the regulators (miRNAs) that are common for all cancer types and the specific active regulators for each cancer type. The results were then validated against knowledge from literature and by gene functional enrichment analysis.

Zhang et al. [57] proposed a method to integrate cancer genomic data from different platforms, including DNA methylation, gene expression, and miRNA expression data. The method adopted a joint matrix factorization technique to integrate different data sources for ovarian cancer samples from TCGA. The method firstly identified the subsets of miRNAs, mRNAs and methylation markers that show the similar patterns across a subset or all of the samples. The aims were to reduce the complexity of the data and provide a global overview of the data from different platforms. The authors applied the method to the ovarian cancer datasets to find the so-called multi-dimensional modules. They validated the method by investigating the genes in the modules that have been confirmed as ovarian cancer related genes and by gene functional enrichment analyses.

In a similar fashion but targeting different data types, Li et al. [58] proposed a method to integrate multiple data sources, including copy number variation (CNV), DNA methylation (DM), gene expression data (GE), and miRNA expression data (ME) for inferring multi-layer gene regulatory modules. The proposed method is called sparse Multi-Block Partial Least Squares regression method (MBPLS) and is employed to identify multi-dimensional regulatory modules from the data. The assumption was that CNV, DM, and ME all regulate the gene expression. The method projected each data type into a summary vector, and maximises the covariance between the summary vectors of source data (CNV, DM, ME) and the response data (GE). Finally, it used the weighted sum of the summary vectors of source data to represent the unique input source data, and again maximises the covariance between the input data and the response data. The method was tested on simulated data as well as the ovarian cancer datasets from TCGA and was capable of identifying the modules that have significant functional and transcriptional enrichments. The results predicted from this method were proved to be better than the results from those methods that use only one type of data.

The results from the above studies encourage future work of integrating more types of genomic data for elucidating the causes of diseases.

Table 3. Summary of methods to discover miRNA-mRNA modules

Method	Brief description	Data sources	Notes	Software tool
Zhang et al. [55]	A framework to identify miRNA-gene regulatory modules by integrating multiple data sources • Use sequence based predictions to build the prior network. • Use expression data, and known gene-gene interactions to refine the findings	• Sequence based miRNA target predictions, • miRNA and gene expression profiles, • protein-protein interactions, • DNA-protein interactions	• Can be applied when we have matched datasets for multiple data types. • The outputs are groups of miRNAs and mRNAs that are important in the biological condition of the dataset	http://zhoulab.usc.edu/SNMNMF/
Le et al. [56]	A regression based method for identifying modules of miRNAs and mRNAs for a specific condition by integrating multiple data sources • Use sequence based predictions to assign regulator-target relationships. • Use expression data to build a regression model between regulators and targets • Use protein-protein interactions to design a function for measuring the strength of the predicted miRNA-mRNA relationships.	• Sequence based miRNA target predictions, • miRNA and mRNA expression data, • protein-protein interactions	• Provide a list of miRNA-mRNA regulatory modules from multiple cancer dataset from TCGA. This could be a good resource for further exploration	NA
Zhang et al. [57]	A joint matrix factorisation method to integrate cancer genomic data from different platform • Integrate multiple data types using matrix factorization technique • Apply to Ovarian datasets from TCGA	• DNA methylation, • miRNA and gene expression data	• Suitable when matched samples of DNA methylation data available • May exclude important genes	http://nar.oxfordjournals.org/content/40/19/9379/suppl/DC1
Li et al. [58]	The sparse Multi-Block Partial Least Squares method to integrate multiple data sources • Reduce the dimension of each data type and represent as a vector • Use regression to integrate the summarised vectors each data type	• Copy number variation, • DNA methylation, • miRNA and gene expression data	• Prove that using multiple data types generates better results than using single data type • May miss important genes in the modules	http://zhoulab.usc.edu/sMBPLS/

DISCOVERING miRNA AND TF CO-REGULATORY RELATIONSHIPS

TFs and miRNAs are primary gene regulators, and identifying their functions is a challenging and important research topic. Currently, there are still no feasible experimental techniques to discover miRNA and TF co-regulatory mechanisms. Meanwhile, computational methods have mainly focused on exploring the functions of miRNA and TF separately in the past decade.

A unified picture of regulatory relationships of the two main regulators and target genes would provide useful insights into the causes of diseases. The combined regulations of miRNAs and TFs are important but difficult to explore, as miRNAs and TFs can regulate each other in addition to regulating target genes. Recently, there are some studies constructing the gene regulatory networks with the presence of both TFs and miRNAs based on sequence data. Few other works utilise both sequence based target predictions of miRNAs and TFs and expression profiles to learn the complex regulatory network and inferring network motifs. Table 4 shows the basic features of the methods in this category.

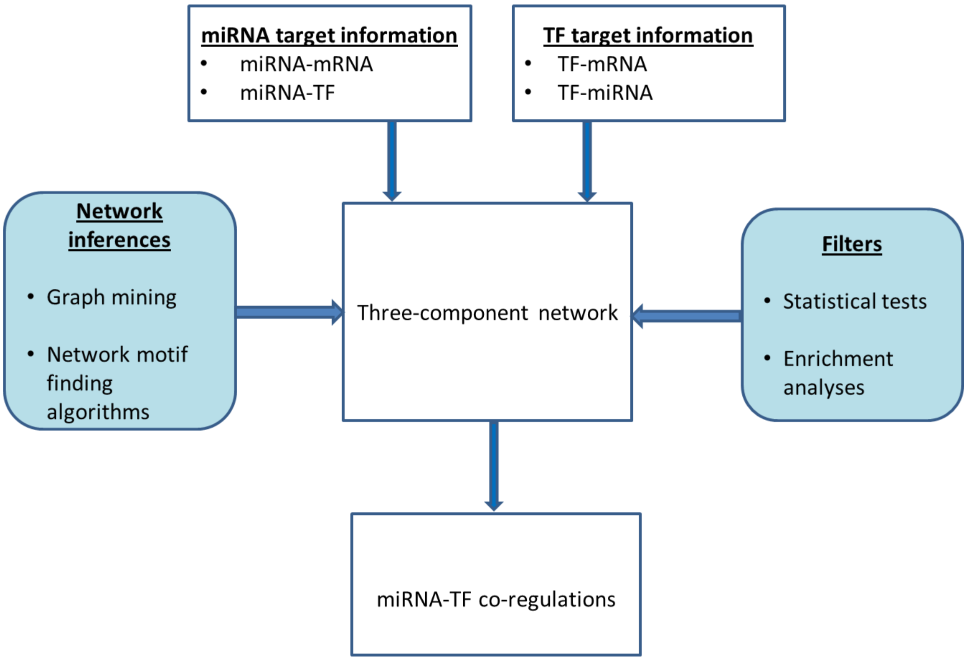


Figure 3. A common framework for exploring miRNA-TF co-regulations. The target information of miRNA and TF is integrated into a network. This network contains three types of molecules, which are miRNAs, TFs, and mRNAs, and it may contain loop of interactions. At this stage, network inferences such as network motif finding algorithms and other filters can be applied to infer miRNA-TF co-regulation knowledge.

Table 4. Summary of methods to infer miRNA-TF co-regulatory relationships

Method	**Brief description**	**Data sources**	**Notes**	**Software tool**
Shalgi et al. [59]	Build the regulatory network that involves miRNAs, TFs, and mRNAs • Use sequence based predictions to build the network • Identify the shared targets of a pair of regulators	• Sequence based target predictions	• The hubs of interactions are usually TFs. • The results may involve a high rate of false discoveries	• No software tool available. • The procedure is repeatable. • Users can replace the target information with recent prediction programs.
Zhou et al. [60]	Build the miRNA-TF-mRNA regulatory network • Use sequence based predictions to build the network • Use Fisher's exact test to identify the significant shared targets of regulators	• Sequence based target predictions	• Shared targets of TFs are much more abundant than that of TF-miRNA • Shared targets in FFLs motifs with TFs as master regulators are statistically significant • Involve high false discovery rate	As above
Chen et al. [61]	Explore the miRNA-TF co-regulatory relationships • Use sequence based predictions to build the network • Use gene functional enrichment analysis to find functional profiles of shared target genes • Identify significant shared targets of regulators	• Sequence based target predictions • Gene ontology	• Some biological processes emerged only in co-regulation • Did not consider the relationships between the regulators.	As above
Tran et al. [62]	Discover gene regulatory modules that consist of miRNAs, TFs, and mRNAs using a rule based method • Build the regulatory network based on target information • Refine the network based on a set of rules, e.g. binding score is significant	• Sequence based target predictions	• Output a set of modules that involve miRNAs and mRNAs • Results are sensitive to the defined rules	As above
Le Bechec et al. [63]	Provide a web tool for miRNA-TF co regulation analysis • Use available target prediction databases to build the network • Infer network motifs based on the built network	• Sequence based target predictions	• A good resource for exploring network motifs that involve miRNA and TF • Did not use expression data, so the network may be static	http://mironton.uni.lu
Roqueiro et al. [64]	Identify the key regulators (miRNAs or TFs) of pathways • Curate known pathways that involve miRNAs and TFs • Use Bayesian inference to integrate sequence based target predictions	• KEGG pathways, • Sequence based target predictions	• Can be used to identify key regulators of a disease • Good resources for validating computational predictions	NA

Sun et al. [65]	Uncover miRNA-TF regulatory network in Glioblastoma (GBM) • Curate a list of miRNAs, TFs and genes related to GBM • Build a regulatory network based on sequence data • Use gene expression to gene-gene interactions, and network motifs	• Sequence based target predictions, • gene expression profiles	• Specific for GBM • Good resource for exploring network motifs involved in GBM	NA
Jiang et al. [66]	Identify active miRNA-TF regulatory pathways in Alzheimer's disease • Use curated databases to build the regulatory network • Use gene expression to identify differentially expressed miRNAs and genes between disease and control samples	• Curated miRNA and TF target databases, • gene expression data	• Specific for Alzheimer's disease • A good resource for validating computational predictions related to Alzheimer's disease	NA
Huang et al. [67]	Develop a web tool (mirConnX) for constructing the regulatory networks that include miRNAs, TFs, and mRNAs • Use target prediction and experimentally confirmed targets to build the static prior regulatory network • Expression data is used to refine the findings for a specific condition	• Sequence based target predictions, • Experimentally validated target databases • expression profiles of miRNAs, TFs, and mRNAs	• A good web tool for exploring the regulatory relationships between miRNAs, TFs and genes • Users can input gene expression data and receive the regulatory relationships for the input dataset.	http://www.benoslab.pitt.edu/mirconnx
Zacher et al. [68]	Explain the activities of miRNAs and TFs in different biological conditions • Use Bayesian inference on expression data to identify the switch in the states of regulators (active, inactive) between two conditions	• Gene expression profiles	• A good tool to identify marker regulator for a condition of interest • Suitable for expression datasets with multiple conditions	http://www.bioconductor.org/packages/release/bioc/html/birta.html
Le et al. [69]	Learn the regulatory networks that include miRNAs, TFs, and mRNAs from heterogeneous data • Split the samples based on the biological conditions • Use Bayesian network to learn the gene regulatory network in each condition • Integrate the results from different conditions, and identify network motifs	• Sequence based target predictions, • expression data of miRNAs, TFs, • mRNAs, sample categories	• Suitable for datasets with multiple conditions • Can be used to explore the interplay between regulators (miRNA and TF) • Long running time with big datasets.	Upon request

Sequence based approaches

A common framework of exploring miRNA-TF co-regulatory relationships is to integrate the putative target information of both TFs and miRNAs to obtain an interaction network with the three components, miRNAs, TFs, and mRNAs. Statistical tests, network inference algorithms, or gene functional enrichment analyses are then used to infer gene regulation knowledge from the combined network. Figure 3 shows the procedure of this framework.

In the first stream of research, researchers studied the co-regulation of TFs and miRNAs by discovering their shared downstream targets [59, 60]. Shalgi et al. [59] built the network that involves miRNAs, TFs and mRNAs using sequence data. They used evolutionary conserved binding sites of miRNA targets to construct the interactions between miRNAs and genes (including TFs). Meanwhile, conserved binding sites of TFs in promoters were used to uncover the interactions between TFs and mRNAs and the interactions between TFs and miRNAs.   The combined network was then analysed to identify the shared targets of the regulators. It was found that the hub of interactions is usually TFs and also discovered some network motifs that involve miRNAs, TFs and mRNAs. Similar to Shalgi's approach, Zhou et al. [60] used PicTar [26] as the miRNA putative targets and Transfac [70] as TF target information to build the network of miRNAs, mRNAs, and TFs. They then used Fisher's Exact Test to measure the significance of the shared targets between the regulators, and to remove the insignificant co-regulating interactions that occurred by chance. They found that the shared targets of TF pairs and miRNA pairs are much more abundant than that of TF-miRNA pairs, and that the shared targets in feed-forward-loops with TF playing as a master regulator are more statistically significant than other types and feed-forward-loops.

Tran et al. [62] proposed a rule based method to discover the gene regulatory modules that consist of miRNAs, TFs, and their target genes based on the available predicted target binding information. Sequence based target prediction database was used to construct putative targets of miRNAs and TFs. The obtained network was refined by retaining only the genes regulated by at least two miRNAs and two TFs with a significant binding score (p-value<0.05). The method then searched for the module that consists of genes, miRNAs and TFs such that the miRNAs and TFs regulate the genes with a stringent p-value. The validity of the modules was assessed using gene functional enrichment analysis for the target genes.

Contrastingly from the above works that use statistical tests as the filter (Figure 3), Chen el al. [61] utilised gene functional enrichment analysis (second filter in Figure 3) to explore the co-regulatory relationships. They also used target information to construct the co-regulation network as the first step. The authors then applied Gene Ontology (GO) for gene functional enrichment analysis for the shared target genes to find the functional profiles for these co-regulation pairs. To calculate the significant levels of the shared targets, they compare their method with the randomly pick method and use the hypergeometric distribution to calculate the p-values of the findings. It was found that some biological processes emerged only in co-regulation and that the disruption of co-regulation might be closely related to cancers, suggesting the importance of the co-regulation of miRNAs and TFs.

Apart from the methods for inferring miRNA-TF co-reguations, there are some works focusing on providing resources for exploring miRNA and TF shared targets, network motifs involving miRNAs and TFs, and known pathways related to miRNAs and TFs. For instances, Le Bechec et al [63] provided a web tool for miRNA-TF co-regulation analysis, MIR@NT@N, which is available at http://mironton.uni.lu. They integrated available target prediction databases for both TFs and miRNAs to construct a regulatory network that involves miRNAs, TFs, and mRNAs. This work provides a web resource for facilitating the retrieval of regulatory relationships and network motifs. Users can explore the shared targets of miRNAs and TFs, and query a list of Feed-Forward-Loops (FFLs) and Feed-Back-Loops (FBLs) that involve miRNAs and TFs. Differently, Roqueiro et al. [64] proposed a method to identify the key regulators (miRNAs or TFs) of pathways. The method used Bayesian inference on known pathway structures to infer a set of regulators in the pathway network. The Bayesian network in this method was constructed manually using the known KEGG pathways by removing the cycles in the pathways and applying some filtering criteria. The method drew findings based on existing knowledge and provided a good resource for other methods to validate their results. However, it is not fit for application in further exploratory studies.

The common feature standing out from the above methods is their employment of sequence-based putative target information. However, the networks constructed from sequence-based predictions involve a high rate of false negatives and false positives [71]. Therefore these methods are only the first step of exploring the complex relationships between the three components, miRNAs, TFs and mRNAs. It would be ideal if expression data can be incorporated to refine the discoveries.

Expression data and other data based approaches

Recently, there has been some work in the second stream of research into miRNA and TF co-regulations that incorporates gene expression data into the studies. These methods also use sequence based target prediction programs to initialise the network at the first step, after which the expression data is used to refine the findings. This procedure is similar to the framework in Figure 2 of the previous section. For instance, Sun et al. [65] proposed a method to uncover miRNA and TF regulatory networks in Glioblastoma GBM). They firstly filtered the miRNAs, TF, and genes related to GBM based on existing knowledge from literature. They then integrated the target prediction of miRNA and TF, which are based on sequence data for constructing the regulatory network. Only the gene expression was utilised to infer the gene-gene interactions by assuming that the interaction occurs when they are co-expressed. The authors then infer the 3-node FFL and 4-node motifs, which involve miRNA-TF interactions. These motifs were integrated into a so-called GBM miRNA-TF mediated network. The authors then conducted the signalling pathway and gene functional enrichment analyses to validate the results.

Similarly, Le et al. [69] proposed a framework to learn from heterogeneous data the three-component regulatory networks, with the presence of miRNAs, TFs, and mRNAs. They firstly used target information of miRNA and TF to define the bipartite of interactions between regulators and target genes. They then utilised Bayesian network structure learning to construct a regulatory network from gene expression profiles of miRNAs, TFs and mRNAs. Then, in order to produce more meaningful results for further biological experimentation and research, the method searched the learnt network to identify the interplay between miRNAs and TFs, and the FFLs that involve miRNAs and TFs from the learnt network.

In another direction, researchers used target information to build the TF, miRNA, and mRNA regulatory networks as the first step. The expression data was then used to identify active pathways that involve miRNAs and TFs [66], or to identify active regulators in different biological conditions of the datasets [68]. For example, Jiang et al. [66] proposed a method to identify active miRNA-TF regulatory pathways in Alzheimer's disease. In the first instance, the curated databases, including TransmiR [40], TRANSFAC [70], miRecords [72], TarBase [73], miRTarBase [35] were used to create the network that included miRNAs, TFs and genes. The authors then integrated miRNA and gene expression data from different sources and identified the differently expressed genes and miRNAs between disease and control samples. They defined these genes as active seeds (nodes) in the curated network, and found the sub-network by connecting all the active nodes with their immediate neighbours. Furthermore, the authors searched for the active pathway in the sub-network by searching for the directed acyclic paths from nodes without parents to nodes without children. The results were validated using gene functional enrichment analysis.

Differently, to create a tool for clinical scientists to generate hypotheses and for explorations, Huang et al. [67] develops a web tool (mirConnX) for constructing the regulatory networks that include miRNAs, TFs, and mRNAs. They firstly created the prior network based on sequence-based miRNA and TF target prediction programs as well as the experimentally confirmed target databases. The expression data were then used to build the association network where the edges represent the significant correlation between the expression levels of the two genes (nodes). They then integrated the association network based on expression data and the prior network based on sequence data into the final network. Using this tool, users can input the expression data and receive the regulatory network that includes miRNA, TFs, and mRNAs. The built networks can be further analysed to identify network motifs. However, an edge in this network simply shows the statistical association in expression levels between a regulator and a target gene, which may not indicate a regulatory relationship.

It is challenging and interesting to design a new class of methods that integrate multiples types of data for exploring miRNA-TF-mRNA regulatory complex relationships.

CHALLENGES IN EVALUATION AND SELECTION OF MODELS

Evaluation of models

A question raised even prior to designing any computational methods of identifying miRNA-mRNA interactions is how to validate the predictions. Although there has been some progress in tackling the validation problem, the challenge remains due to the sparse number of experimentally confirmed miRNA-mRNA interactions. Therefore, there is no complete ground-truth for evaluating and comparing different computational methods. In this section, we review the current methods of validating or enriching miRNA target predictions.

Currently, the common methods for validating computational results about miRNA targets are:

- Using experimentally validated target databases such as TarBase [73], miRecords [72], miRTarbase [35]
- Using miRNA transfection experiment data
- Performing enrichment analysis using proteomic data
- Calculating the number of target genes appearing in known pathways
- Applying functional and/or pathway enrichment analyses to investigate the relevance of the target genes to the biological conditions of the dataset

Tarbase and miRecords are manually curated experimental interaction databases and are commonly referred to when validating miRNA target predictions. These databases contain the collection of the experimentally confirmed interactions, and provide interfaces for facilitating the information retrieval process. mirTarBase stores miRNA targets validated by high-confidence low-throughput assays such as Western blot. Meanwhile, Transmir [40] stores the curated experimentally confirmed interactions between miRNAs and TFs. These databases provide good tools to validate miRNA-mRNA regulatory predictions, although the number of confirmed targets is still small.

A common approach to validating miRNA-mRNA regulatory predictions is to examine the overlap between the predictions and the experimentally confirmed databases. However, the small number of experimentally validated targets would make the validation of a new method difficult, as we do not know whether a predictive target that is not in those databases is a false discovery or a novel true target.

Another approach of validating miRNA-mRNA regulatory predictions is to use miRNA transfection data. Transfection data presents the changes in gene expression between the control and miRNA transfected conditions. The differentially expressed genes identified from the control and miRNA transfected samples are considered as targets of the miRNA. Although this approach cannot differentiate between direct and indirect miRNA-mRNA interactions, it provides a tool to evaluate the real effect that a miRNA has on mRNAs.

Khan et al. [74] collected the miRNA transfection data from 151 published transfection experiments in seven different human cell types. Luo et al [75] provided the fold-change in the gene expression levels between control and transfected samples in MDA-MB-231 cells for miR-200c, miR-375, and miR-205. Le et al [31] presented the transfection data for miR-200 family in MDA-MB-231 cells. Li et at. [29] compiled 84 datasets from GEO for 113 transfected miRNAs. Table 5 shows the available miRNA transfection data sources.

Table 5. Summary of miRNA transfection data collected from literature by Khan et al [74], Luo et al. [75], Li et al. [29], and Le et al. [31].

References	**Cell types**	**Transfected miRNAs**
Luo et al. [75]	MDA-MB-231	miR-200c, miR-375, miR-205
Lim et al. [76]	HeLa	miR-124, miR-1, miR-373, miR-124mut5-6, miR-124mut9-10, chimiR-1-124, chimiR-124-1
Linsley et al. [77]	HeLa, HCT116, HCT116 Dicer	miR-106b, miR-200a/b, miR-141, miR-15a/b, miR-16, miR-103, miR-20, let-7c, miR-195, miR-107, miR-192, miR-215, miR-17-5p
Grimson et al. [78]	HeLa	miR-7, miR-9, miR-122a, miR-128a, miR-132, miR-133a, miR-142, miR-148, miR-181a
He et al. [79]	HeLa, A549, TOV21G, HCT116 Dicer	miR-34a, miR-34b, miR-34c
Selbach et al. [80]	HeLa	miR-1, miR-155, let-7b, miR-30
Baek et al. [36]	HeLa	miR-181a, miR-124, miR-1
Chang et al. [81]	HeLa	miR-34a
Wang et al. [82]	HepG2	miR-124
Li et al. [29]	77 human tissue or cells from GEO	113 miRNAs
Le et al. [31]	MDA-MB-231	miR-200a, miR-200b

In the similar fashion, protein expression data in miRNA transfection experiments can be used to validate computational predictions. Although measuring protein expression data generates higher cost compared to gene expression data, protein expression data better reflect the effect of miRNA regulations. For instance, Baek et al [36] reported both gene and protein expression data for transfection experiments of hsa-miR-1, 124, and 181a. They then use the data to validate and compare different target prediction methods.

In another direction, researchers use pathway and/or gene functional enrichment analyses to investigate the relevance of the functions of predicted targets to the biological conditions of the datasets used for the predication [23, 30, 83]. The assumption is that a good prediction method will generate a set of miRNA target genes which are involved in the pathways and processes relevant to the biology behind the used dataset. There are several software tools designed for this purpose. The common ones are GO [84] enrichment analysis, KEGG pathway analysis [85], GeneCodis [86], Ingenuity Pathway Analysis (IPA, Ingenuity Systems, www.ingenuity.com), GeneGo Metacore from GeneGo Inc.. The first three are free for research use, while the last two are commercial software tools with a limited trial time.

To date, the validation tools for evaluating miRNA-mRNA regulatory predictions are still limited and there is no unique ground-truth for assessing the performance of different computational methods. A possible remedy for the problem is to combine different validating methods. For example, a good computational method should have a significant number miRNA-mRNA interactions confirmed by transfection experiments and have the target genes highly relevant to the biological condition of the datasets. However, the differences in biological conditions between the training datasets and the transfection experiments may result in biased validation results. It is desirable to have follow-up experiments with the same biological conditions of the dataset used for validating the results of a computational prediction method.

Selection of models

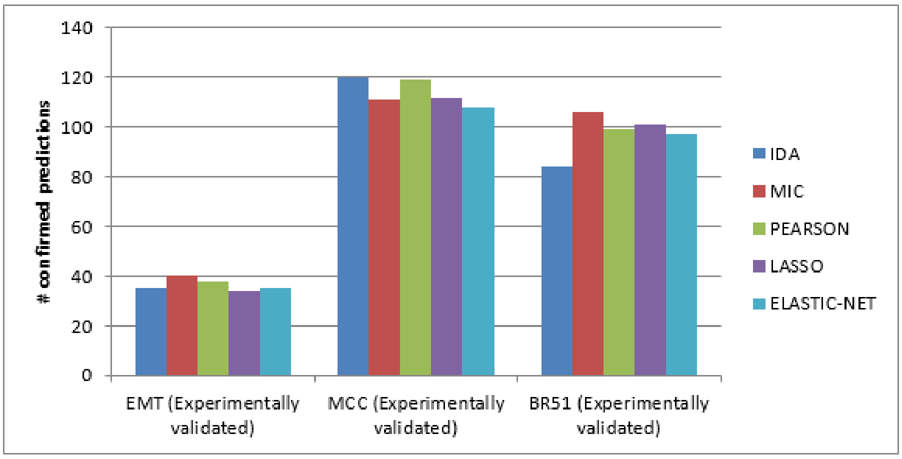
As we do not have the ground-truth for evaluating the models, it is impossible to conclude which method is better than the other. Therefore, selecting a model for assisting with the experiment design is a difficult task. To investigate whether we can select one model over another based on its relative performance validated by the above-mentioned evaluation methods, in this section we conduct a case study in three different cancer datasets.

In this study, we choose Pearson correlation [87], maximal information coefficient (MIC) [88], Lasso [89], Elastic-net [90], and the method which is based on IDA in [31] for the comparison. We apply them to the matched miRNA and mRNA expression profiles from epithelial-mesenchymal transition (EMT) [91, 92], multi-class cancer (MCC) [93, 94], and 51 cell lines breast cancer (BR51) [95] datasets. We then use experimentally validated miRNA targets and miRNA transfection data for validating the predictions.

For EMT datasets, the miRNA expression profiles are from [92] (data available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26375). They were profiled from the 60 cancer cell lines of the drug screening panel of human cancer cell lines at the National Cancer Institute (NCI-60). The mRNA expression profiles of EMT for NCI-60 were obtained from ArrayExpress http://www.ebi.ac.uk/arrayexpress accession number E-GEOD-5720. There are 11 samples of epithelial, and 36 samples of mesenchymal. For MCC datasets, the miRNA expression profiles were obtained from [94] (data available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2564). The mRNA expression profiles of MCC are from [93]. They can be downloaded at http://www.broad.mit.edu/cancer/pub/migcm. Samples of the MCC data classified as normal (21 samples) and tumor (67 samples). Finally, for BR51 datasets, the miRNA data are from [95] and the mRNA gene expression data are available at: http://www.ncbi.nlm.nih.govgeoqueryacc.cgiacc=GSE41313. There are 27 samples in the luminal group and 23 samples in the basal group for this dataset.

We perform differential gene expression analysis on the gene expression profiles to identify differentially expressed miRNAs and mRNAs between the two conditions in each data set. In this work, we use *limma* package [96] of Bioconductor for the analysis. As a result of the analysis, 46 probes of miRNAs and 1635 probes of mRNAs for the EMT dataset, and 62 probes of miRNAs and 1363 probes of mRNAs for the MCC data set have been identified to be differentially expressed at significant level (adjusted p-value <0.05, adjusted by BH method). Meanwhile, 92 miRNAs (adjusted p-value < 0.2) and 2354 mRNAs (adjusted p-value<0.0001) are identified to be differentially expressed in the BR51 dataset. To cover the important miRNAs mentioned in the analysis of [95] and to have a manageable number of mRNAs for the computational method, we choose different thresholds of adjusted p-values in the differential gene expression analysis for this dataset.

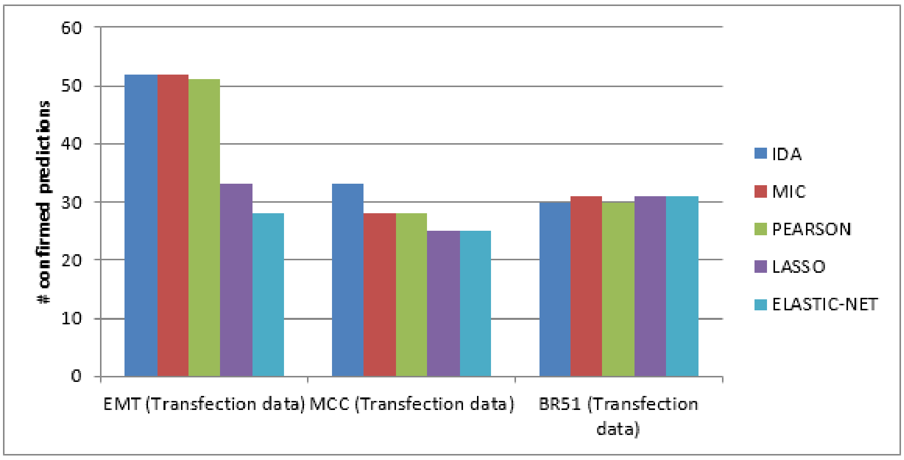
Figure 4 shows the performance of the methods validated by experimentally confirmed target databases. We firstly use the union of three databases, Tarbase 6.0 [97], miRecords [72], and miRWalk [98] as the ground-truth. For each miRNA in a dataset, we extract the top 100 target genes predicted by each of the methods and validate them against the experimentally confirmed interactions. We then compare the number of confirmed interactions for the methods in each of the three datasets.



p-values	IDA	MIC	PEARSON	LASSO	ELASTIC-NET
EMT (Experimentally validated)	4.00E-03	2.60E-04	8.50E-04	6.20E-03	4.00E-03
MCC (Experimentally validated)	1.90E-01	5.10E-01	2.20E-01	4.60E-01	6.20E-01
BR51 (Experimentally validated)	7.20E-06	2.10E-12	4.50E-10	1.00E-10	1.90E-09

Figure 4. Number of interactions validated by experimentally confirmed target databases for IDA, Pearson correlation coefficient, MIC, Lasso, and Elastic-net in EMT, MCC and BR51 datasets.

Meanwhile, for transfection data, we use the miR-200a transfection data from [35] to validate the predictions of the methods. The comparison results are shown in Figure 5.



p-values	IDA	MIC	PEARSON	LASSO	ELASTIC-NET
EMT (Transfection data)	1.00E-12	1.00E-12	4.90E-12	1.00E-03	2.00E-02
MCC (Transfection data)	6.60E-04	1.00E-02	1.00E-02	4.60E-02	4.60E-02
BR51 (Transfection data)	1.10E-04	4.50E-05	1.10E-04	4.50E-05	4.50E-05

Figure 5. Number of interactions validated by miR-200a transfection data for IDA, Pearson correlation coefficient, MIC, Lasso, and Elastic-net in EMT, MCC and BR51 datasets.

In general, there is no superior method that outperforms all other methods in any validation method. Association methods such as MIC and Pearson perform well when validating against experimentally confirmed target databases. Meanwhile, the causality discovery based method, IDA, performs well when we use miRNA transfection data as the ground-truth.

It is important to note that the validation results, however, do not imply that all the methods are of equal merits. Figure 6A shows the overlap in miR-200a predicted targets by IDA, Pearson, and MIC. There is a significant number of targets predicted by a method, but other methods fail to discover. This suggests different methods may infer different sets of the miRNA targets. Moreover, Figure 6B, 6C, and 6D show a large number of confirmed genes predicted by a method but not by the other. There is a strong implication that these methods discover results which are complementary to each other. Therefore, it is not simple to claim that a method is outstanding over the others using existing validation methods.

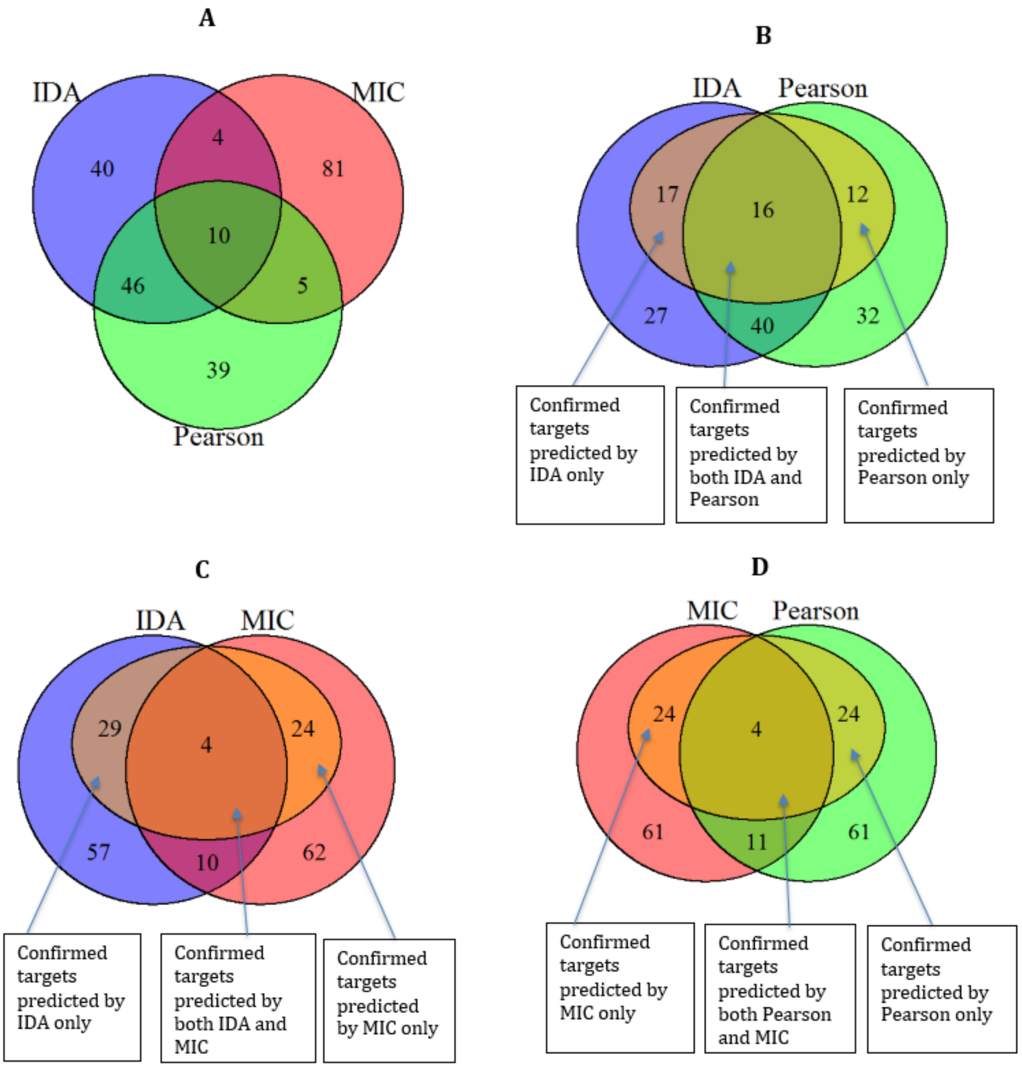

A
IDA
MIC
Pearson
40
4
81
10
46
5
39
B
IDA
Pearson
17
16
12
27
40
32
Confirmed targets predicted by IDA only
Confirmed targets predicted by both IDA and Pearson
Confirmed targets predicted by Pearson only
C
IDA
MIC
29
4
24
57
10
62
Confirmed targets predicted by IDA only
Confirmed targets predicted by both IDA and MIC
Confirmed targets predicted by MIC only
D
MIC
Pearson
24
4
24
61
11
61
Confirmed targets predicted by MIC only
Confirmed targets predicted by both Pearson and MIC
Confirmed targets predicted by Pearson only

Figure 6. A) Overlapping between predicted targets from IDA, Pearson correlation coefficient, and MIC. B) Overlapping in targets confirmed by miR-200a transfection data between IDA and Pearson correlation coefficient. There are 17 confirmed genes predicted by IDA only, and 12 confirmed genes predicted by Pearson correlation coefficient only. C) Overlapping in targets confirmed by miR-200a transfection data between IDA and MIC. There are 29 confirmed genes predicted by IDA only, and 24 confirmed genes predicted by Pearson correlation coefficient only. D) Overlapping in targets confirmed by miR-200a transfection data between MIC and Pearson correlation coefficient. There are 24 confirmed genes predicted by IDA only, and 24 confirmed genes predicted by Pearson correlation coefficient only.

CONCLUSIONS AND OUTLOOKS

As more and more evidence has suggested the important roles of miRNAs in the development of several diseases, identifying miRNA functions will reveal further insights into the causes of fatal diseases such as cancer. The huge amount of data available in different types poses opportunities and challenges for computational approaches to exploring miRNA functions. These methods assist with the designs of the wet lab experiments. In this review we have discussed different computational approaches to inferring miRNA functions. These approaches provide different views on how to elucidate the complex regulatory mechanism of miRNAs and their relationships with target genes as well as other regulators.

The approaches which are based on sequence data provide a list of potential target genes. However, the results from these approaches involve a high rate of false discoveries. There is a need to design new accurate methods for utilizing sequence data. Recently, such computational and experimental methods have started to emerge. For instances, Wang et al [99] proposed the mirTarPri method to reduce the false discovery rates of commonly used target prediction methods by ranking the predicted targets to select optimal results. In another direction, experimental methods such as HITS-CLIP [100] and PAR-CLIP [101] have achieved better accuracy rates compared to prior prediction programs. These new approaches may have potential to enhance our understanding in miRNA functions.

Complementary to sequence based approaches, which discover static miRNA targets, approaches that utilise gene expression data can infer the miRNA activities and miRNA-mRNA relationships in a specific condition. These approaches have long been studied based on the foundation of correlation or association. However, it is well known that association is not causation, and thus the correlation between a miRNA and target genes may not imply the real regulatory relationships, which are causal relationships. Recently, causality discovery based methods have emerged and provided an alternative approach to inferring miRNA-mRNA causal regulatory relationships. In saying so, these methods usually have a set of assumptions on the distribution of the data and those assumptions may be mostly violated in some real world datasets. As a result, the causality discovery methods may not perform consistently in all datasets. Therefore, it is crucial to design a new class of methods that can infer causal relationships as well as produce stable results across different datasets.

On another note, the approaches of integrating multiple sources of data have been shown to be effective in improving the predictive power of a computational method. Each data type provides complement information, and thus integrating more sources of data would provide a clearer picture of gene regulations. However, different data sources often come from different labs with different experiment settings and different biological conditions. The question raised here is how to uniformly transfer the knowledge learnt from one lab to another.

As the ultimate research goal is to elucidate the causes of a disease, differential analysis approaches are promising. We have seen differential analysis techniques in identifying gene differentially expressed analysis, and recent methods of identifying differential co-expression. As suggested in [46], the next step would be identifying differential networking, i.e. identifying the differences between gene regulatory networks in different conditions. However, it is still unclear how to achieve the goal.

As discussed in the above section, it is still challenging to evaluate and select a model for assisting with the experimental design. It is necessary to create tools for systematically evaluating a model based on current knowledge, and visualising prediction results from different methods. As different methods may infer complementary results and have their own advantages, ensemble approaches such as in [102] can be promising in producing stable and high accurate results. Although we used human cancer datasets in the case study, the methods discussed in this review generally are not limited to human cancer datasets. The methods can be applied to different organisms and biological conditions.

Gene regulatory networks involve several classes of regulators, and thus constructing the unified picture of the network with the presence of important regulators and genes is crucial. Recent methods have constructed the gene regulatory networks to present the relationships between miRNAs, TFs and genes. However, most of the methods utilise the sequence based target prediction to build the network, which may involve a high rate of false discoveries. There is a potential to integrate multiple sources and/or types of data to learn such complex networks. Furthermore, differential analyses can be utilised to explore the changes in regulatory patterns, e.g. the changes in Feed-Forward Loops (FFLs) and Feed-Back-Loops (FBLs) that involve miRNAs, TFs, and/or genes. The FBLs and FFLs have been found to play important roles in cancers and other diseases [103]. Please refer to [103] for a recent review of the TF-miRNA motif roles in biological processes and diseases. Understanding the roles of such motifs in disease development would better assist with the diagnosis process and the design of pharmaceutical products for treatment.

Together with the two major gene regulators, miRNAs and TFs, long non-coding RNAs (lncRNAs) play important roles in biological processes and diseases [104]. More emerging evidence has shown that lncRNAs, miRNAs, and TFs are important nodes in the signaling networks that regulate vital biological processes and diseases, including cancer [104, 105]. Recent evidence also reveals the interactions between miRNAs, TFs and lncRNAs [106]. With more and more data available, there is a strong possibility for computational methods to tap into the area of learning gene regulatory networks with the existence of miRNAs, TFs, lncRNAs, and genes. It would pose interesting implications for future work as these methods can help elucidate the complex gene regulatory relationships and the causes of diseases.

Funding

This work has been partially supported by Australian Research Council Discovery Project DP130104090 and DP140103617.

References

1. Vaquerizas MJ, Kummerfeld KS, Teichmann AS et al. A census of human transcription factors: function, expression and evolution., Nature Review Genetics 2009;10.
2. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 2004;116:281-297.
3. Chen J-F, Mandel EM, Thomson JM et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation., Nature Genetics 2006;38:228-233.

4. Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis., Nature 2005;436:214-220.

5. Poy MN, Eliasson L, Krutzfeldt J et al. A pancreatic islet-specific microRNA regulates insulin secretion., Nature 2004;432:226-230.

6. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer., Nature Reviews. Cancer 2006;6:259-269.

7. Jin P, Zarnescu DC, Ceman S et al. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway., Nature Neuroscience 2004;7:113-117.

8. Xu C, Lu Y, Pan Z et al. The muscle-specific microRNAs miR-1 and miR-133 produce opposing effects on apoptosis by targeting HSP60, HSP70 and caspase-9 in cardiomyocytes., Journal of Cell Science 2007;120:3045-3052.

9. Cui Q, Yu Z, Purisima EO et al. Principles of microRNA regulation of a human cellular signaling network., Molecular Systems Biology 2006;2:1-7.

10. Beveridge N, Gardiner E, Carroll A et al. Schizophrenia is associated with an increase in cortical microRNA biogenesis, Molecular Psychiatry 2009;15:1176-1189.

11. Iorio MV, Ferracin M, Liu C-G et al. MicroRNA gene expression deregulation in human breast cancer, Cancer Research 2005;65:7065-7070.

12. Porkka KP, Pfeiffer MJ, Waltering KK et al. MicroRNA expression profiling in prostate cancer, Cancer Research 2007;67:6130-6135.

13. Yanaihara N, Caplen N, Bowman E et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis, Cancer Cell 2006;9:189-198.

14. Akao Y, Nakagawa Y, Naoe T. MicroRNA-143 and-145 in colon cancer, DNA and Cell Biology 2007;26:311-320.

15. Yang H, Kong W, He L et al. MicroRNA expression profiling in human ovarian cancer: miR-214 induces cell survival and cisplatin resistance by targeting PTEN, Cancer Research 2008;68:425-433.

16. Zhang X, Cairns M, Rose B et al. Alterations in miRNA processing and expression in pleomorphic adenomas of the salivary gland, International Journal of Cancer 2009;124:2855-2863.

17. Hébert SS, Horré K, Nicolaï L et al. MicroRNA regulation of Alzheimer's Amyloid precursor protein expression, Neurobiology of Disease 2009;33:422-428.

18. Cox MB, Cairns MJ, Gandhi KS et al. MicroRNAs miR-17 and miR-20a inhibit T cell activation genes and are under-expressed in MS whole blood, PLoS One 2010;5:e12132.

19. Croce CM. Causes and consequences of microRNA dysregulation in cancer, Nature Reviews Genetics 2009;10:704-714.

20. 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.

21. Hobert O. Gene regulation by transcription factors and microRNAs, Science 2008;319:1785-1786.

22. Davidson EH. Genomic regulatory systems: in development and evolution. Access Online via Elsevier, 2001.

23. Muniategui A, Pey J, Planes FJ et al. Joint analysis of miRNA and mRNA expression data, Briefings in Bioinformatics 2012;14:263-278.

24. Liu B, Li J, Cairns MJ. Identifying miRNAs, targets and functions, Briefings in Bioinformatics 2012.

25. Enright AJ, John B, Gaul U et al. microRNA targets in Drosophila, Genome Biology 2004;5:R1-R1.

26. Krek A, Grun D, Poy MN et al. Combinatorial microRNA target predictions, Nature Genetics 2005;37:495-500.

27. Chen X, Slack FJ, Zhao H. Joint analysis of expression profiles from multiple cancers improves the identification of microRNA–gene interactions, Bioinformatics 2013;29:2137-2145.

28. Jacobsen A, Silber J, Harinath G et al. Analysis of microRNA-target interactions across diverse cancer types, Nature Structural & Molecular Biology 2013.

29. Li Y, Goldenberg A, Wong K-C et al. A probabilistic approach to explore human miRNA targetome by integrating miRNA-overexpression data and sequence information, Bioinformatics 2013:btt599.
30. Liu B, Li J, Tsykin A et al. Exploring complex miRNA-mRNA regulatory networks by splitting-averaging strategy, BMC Bioinformatics 2009;19:1-19.
31. Le TD, Liu L, Tsykin A et al. Inferring microRNA-mRNA causal regulatory relationships from expression data, Bioinformatics 2013;29:765-771.
32. Liang Z, Zhou H, He Z et al. mirAct: a web tool for evaluating microRNA activity based on gene expression data, Nucleic Acids Research 2011;39:W139-W144.
33. Amar D, Safer H, Shamir R. Dissection of regulatory networks that are altered in disease via differential co-expression, PLoS Computational Biology 2013;9:e1002955.
34. John B, Enright AJ, Aravin A et al. Human microRNA targets, PLoS Biology 2004;2:e363.
35. Hsu S-D, Tseng Y-T, Shrestha S et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions, Nucleic Acids Research 2014;42:D78-D85.
36. Baek D, Villén J, Shin C et al. The impact of microRNAs on protein output, Nature 2008;455:64-71.
37. Neapplitan R. Learning Bayesian networks. Prentice Hall, 2003.
38. Pearl J. Causality: models, reasoning, and inference. Cambridge University Press, 2000.
39. Heckerman D, Chickering DM. Learning Bayesian networks : the combination of knowledge and statistical data metrics for belief networks :, Machine Learning 1995;20:197-243.
40. Wang J, Lu M, Qiu C et al. TransmiR: a transcription factor-microRNA regulation database., Nucleic Acids Research 2010;38:D119-122.
41. Maathuis HM, Colombo D, Kalisch M et al. Predicting causal effects in large-scale systems from observational data, Nature Methods 2010;7:247-249.
42. Maathuis HM, Kalisch M, Buhlmann P. Estimating high-dimensional intervention effects from observational data, Annals of Statistics 2009;37:3133-3164.
43. Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. Cambridge, MA: MIT Press, 2000.
44. Kalisch M, Buhlmann P. Estimating high-dimensional directed acyclic graphs with the PC-algorithm, Journal of Machine Learning Research 2007;8:613-636.
45. Zisoulis DG, Kai ZS, Chang RK et al. Autoregulation of microRNA biogenesis by let-7 and Argonaute, Nature 2012;486:541-544.
46. de la Fuente A. From 'differential expression'to 'differential networking'-identification of dysfunctional regulatory networks in diseases, Trends in genetics 2010;26:326-333.
47. Sood P, Krek A, Zavolan M et al. Cell-type-specific signatures of microRNAs on target mRNA expression, Proceedings of the National Academy of Sciences of the United States of America 2006;103:2746-2751.
48. Alexiou P, Maragkakis M, Papadopoulos GL et al. The DIANA-mirExTra web server: from gene expression data to microRNA function, PLoS One 2010;5:e9171.
49. van Dongen S, Abreu-Goodger C, Enright AJ. Detecting microRNA binding and siRNA off-target effects from expression data, Nature Methods 2008;5:1023-1025.
50. Cheng C, Li LM. Inferring microRNA activities by combining gene expression with microRNA target prediction, PLoS One 2008;3:e1989.
51. Jayaswal V, Schramm S-J, Mann GJ et al. VAN: an R package for identifying biologically perturbed networks via differential variability analysis, BMC research notes 2013;6:430.
52. Hartwell LH, Hopfield JJ, Leibler S et al. From molecular to modular cell biology, Nature 1999;402:C47-C52.
53. Ihmels J, Friedlander G, Bergmann S et al. Revealing modular organization in the yeast transcriptional network, Nature Genetics 2002;31:370-377.
54. Qi Y, Ge H. Modularity and dynamics of cellular networks, PLoS Computational Biology 2006;2:e174.

55. Zhang S, Li Q, Liu J et al. A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules, Bioinformatics 2011;27:i401-i409.

56. Le H-S, Bar-Joseph Z. Integrating sequence, expression and interaction data to determine condition-specific miRNA regulation, Bioinformatics 2013;29:i89-i97.

57. Zhang S, Liu C-C, Li W et al. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data, Nucleic Acids Research 2012;40:9379-9391.

58. Li W, Zhang S, Liu C-C et al. Identifying multi-layer gene regulatory modules from multi-dimensional genomic data, Bioinformatics 2012;28:2458-2466.

59. Shalgi R, Lieber D, Oren M et al. Global and local architecture of the mammalian microRNA-transcription factor regulatory network, PLoS Computational Biology 2007;3:e131.

60. Zhou Y, Ferguson J, Chang JT et al. Inter- and intra-combinatorial regulation by transcription factors and microRNAs, BMC Genomics 2007;8:396.

61. Chen C-y, Chen S-t, Fuh C-s et al. Coregulation of transcription factors and microRNAs in human transcriptional regulatory network, BMC Bioinformatics 2011;12:S41.

62. Tran DH, Satou K, Ho TB et al. Computational discovery of miR-TF regulatory modules in human genome, Bioinformation 2010;2063:371-377.

63. Bechec AL, Portales-casamar E, Vetter G et al. MIR @ NT @ N : a framework integrating transcription factors , microRNAs and their targets to identify sub-network motifs in a meta-regulation network model, BMC Bioinformatics 2011;12:67.

64. Roqueiro D, Huang L, Dai Y. Identifying transcription factors and microRNAs as key regulators of pathways using Bayesian inference on known pathway structures., Proteome Science 2012;10 Suppl 1:S15.

65. Sun J, Gong X, Purow B et al. Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma, PLoS Computational Biology 2012;8:e1002488.

66. Jiang W, Zhang Y, Meng F et al. Identification of active transcription factor and miRNA regulatory pathways in Alzheimer's disease, Bioinformatics 2013;29:2596-2602.

67. Huang GT, Athanassiou C, Benos PV. mirConnX: condition-specific mRNA-microRNA network integrator., Nucleic Acids Research 2011;39:W416-423.

68. Zacher B, Abnaof K, Gade S et al. Joint Bayesian inference of condition-specific miRNA and transcription factor activities from combined gene and microRNA expression data., Bioinformatics (Oxford, England) 2012;28:1714-1720.

69. Le TD, Liu L, Liu B et al. Inferring microRNA and transcription factor regulatory networks in heterogeneous data, BMC Bioinformatics 2013;14:92.

70. Matys V. TRANSFAC(R): transcriptional regulation, from patterns to profiles, Nucleic Acids Research 2003;31:374-378.

71. Sethupathy P, Megraw M, Hatzigeorgiou AG. A guide through present computational approaches for the identification of mammalian microRNA targets, Nature Methods 2006;3:881-886.

72. Xiao F, Zuo Z, Cai G et al. miRecords: an integrated resource for microRNA-target interactions, Nucleic Acids Research 2009;37:D105-D110.

73. Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: A comprehensive database of experimentally supported animal microRNA targets, RNA 2006;12:192-197.

74. Khan AA, Betel D, Miller ML et al. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs, Nature Biotechnology 2009;27:549-555.

75. Luo D, Wilson JM, Harvel N et al. A systematic evaluation of miRNA: mRNA interactions involved in the migration and invasion of breast cancer cells, Journal of Translational Medicine 2013;11:57.

76. Lim LP, Lau NC, Garrett-Engele P et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs, Nature 2005;433:769-773.

77. Linsley PS, Schelter J, Burchard J et al. Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression, Molecular and Cellular Biology 2007;27:2240-2252.

78. Grimson A, Farh KK-H, Johnston WK et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing, Molecular Cell 2007;27:91-105.
79. He L, He X, Lim LP et al. A microRNA component of the p53 tumour suppressor network, Nature 2007;447:1130-1134.
80. Selbach M, Schwanhäusser B, Thierfelder N et al. Widespread changes in protein synthesis induced by microRNAs, Nature 2008;455:58-63.
81. Chang T-C, Wentzel EA, Kent OA et al. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis, Molecular Cell 2007;26:745-752.
82. Wang X, Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling, Nucleic Acids Research 2006;34:1646-1652.
83. Liu B, Liu L, Tsykin A et al. Identifying functional miRNA-mRNA regulatory modules with correspondence latent Dirichlet allocation, Bioinformatics 2010;26:3105-3111.
84. Harris M, Clark J, Ireland A et al. The Gene Ontology (GO) database and informatics resource., Nucleic Acids Research 2004;32:D258-261.
85. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Research 2000;28:27-30.
86. Nogales-Cadenas R, Carmona-Saez P, Vazquez M et al. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information, Nucleic Acids Research 2009;37:W317-W322.
87. Pearson K. Mathematical Contributions to the Theory of Evolution.-On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs, Proceedings of the Royal Society of London 1896;60:489-498.
88. Reshef DN, Reshef YA, Finucane HK et al. Detecting novel associations in large data sets, Science 2011;334:1518-1524.
89. Tibshirani R. Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) 1996:267-288.
90. Zou H, Hastie T. Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005;67:301-320.
91. Monks A, Scudiero D, Skehan P et al. Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines, Journal of the National Cancer Institute 1991;83:757-766.
92. Søkilde R, Kaczkowski B, Podolska A. Global microRNA analysis of the NCI-60 cancer cell panel, Molecular Cancer Therapeutics 2011;10:375-384.
93. Ramaswamy S, Tamayo P, Rifkin R et al. Multiclass cancer diagnosis using tumor gene expression signatures, Proceedings of the National Academy of Sciences 2001;98:15149-15154.
94. Lu J, Getz G, Miska EA et al. MicroRNA expression profiles classify human cancers, Nature 2005;435:834-838.
95. Riaz M, van Jaarsveld MT, Hollestelle A et al. miRNA expression profiling of 51 human breast cancer cell lines reveals subtype and driver mutation-specific miRNAs, Breast Cancer Research 2013;15:R33.
96. Smyth GK. Limma : Linear models for microarray data, Bioinformatics and Computational Biology Solutions using R and Bioconductor 2005:397-420.
97. Vergoulis T, Vlachos IS, Alexiou P et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support, Nucleic Acids Research 2012;40:D222-D229.
98. Dweep H, Sticht C, Pandey P et al. miRWalk-database: prediction of possible miRNA binding sites by walking the genes of three genomes, Journal of Biomedical Informatics 2011;44:839-847.
99. Wang P, Ning S, Wang Q et al. mirTarPri: Improved Prioritization of MicroRNA Targets through Incorporation of Functional Genomics Data, PLoS One 2013;8:e53685.
100. Chi SW, Zang JB, Mele A et al. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps, Nature 2009;460:479-486.

101. Hafner M, Landthaler M, Burger L et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP, Cell 2010;141:129-141.
102. Marbach D, Costello JC, Küffner R et al. Wisdom of crowds for robust gene network inference, Nature Methods 2012;9:796-804.
103. Zhang H-M, Kuang S, Xiong X et al. Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases, Briefings in Bioinformatics 2013:bbt085.
104. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development, Nature Reviews Genetics 2013.
105. Gupta RA, Shah N, Wang KC et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis, Nature 2010;464:1071-1076.
106. Li J-H, Liu S, Zhou H et al. starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Research 2014;42:D92-D97.