

Computational Methods for Identifying MicroRNA-Gene Regulatory Modules



Yin Liu

Abstract MicroRNAs (miRNAs) are small noncoding RNA molecules that participate in post-transcriptional gene regulation through mRNA degradation or protein translational repression. It has been estimated that up to 60% of mammalian genes are regulated by miRNAs. Understanding the complexity of gene regulation mediated by miRNAs requires a detailed description of active miRNA-gene regulatory modules, which are composed of a set of miRNAs and their target genes. These studies are now feasible at the genome-wide level with recent advances in high-throughput technologies. In this chapter, we briefly review current computational methods on identifying miRNA-gene regulatory modules from multiple types of functional genomic data. We also present some interesting areas for further research.

1 Introduction

As a major class of small noncoding RNAs, microRNAs (miRNAs) are essential for a variety of biological processes including development [1, 2], proliferation [3], differentiation [4], and cellular signaling [5, 6]. MiRNAs regulate post-transcriptional gene expression through protein translational repression and/or mRNA degradation [7–9]. It has been estimated that miRNAs are able to regulate approximately 60% of protein-coding genes in mammalian genomes [10, 11]. Disruption in miRNA expression affects normal cellular functions, leading to the development and progression of complex human diseases such as cancers [12, 13] and neurodegenerative [14–16] and cardiovascular diseases [17, 18]. MiRNAs have demonstrated medical significance as noninvasive biomarkers for disease diagnosis and prognosis

Y. Liu (✉)

Department of Neurobiology and Anatomy, University of Texas Health Science Center at Houston, Houston, TX, USA

University of Texas Graduate School of Biomedical Science, Houston, TX, USA

e-mail: Yin.Liu@uth.tmc.edu

© The Author(s), under exclusive license to Springer-Verlag GmbH, DE,

part of Springer Nature 2022

H. H.-S. Lu et al. (eds.), *Handbook of Statistical Bioinformatics*,

Springer Handbooks of Computational Statistics,

https://doi.org/10.1007/978-3-662-65902-1_10

[19, 20]. Furthermore, preclinical studies using miRNA-based therapeutics have been successfully tested on various disease models, suggesting novel therapeutic interventions can be developed upon the manipulation and *in vivo* delivery of miRNAs [21, 22]. Given the importance of miRNAs in gene regulation, a detailed description of the miRNA regulatory effects on protein-coding genes will be critical for us to understand their roles in normal biological processes, disease development, and therapeutic design. Many computational methods have been proposed for identifying the targets of miRNAs, a key step to reveal miRNA functions and link miRNAs to protein-coding genes [23]. Early studies in this area were based on several well-known miRNA target recognition rules of genomic sequence features, including sequence complementarity between miRNAs and target genes, thermodynamic stability, target site context, and the degree of site conservation [10, 24–27]. However, these target prediction methods based on sequence information alone could include many false positives, and more importantly, the predicted static miRNA-gene interactions cannot capture the dynamics of miRNA regulatory effects among different conditions and tissues [28]. In recent years, with advances in high-throughput technologies such as RNA-sequencing, many computational approaches have been developed to integrate heterogeneous data resources into sequence-based target predictions to obtain more reliable information on miRNA-mediated gene regulation at the genome-wide level. These methods identify a list of individual miRNA-gene interactions in the context of miRNA regulatory network. Detailed reviews of these methods can be found elsewhere [23, 29, 30].

In addition to identifying each individual miRNA-gene interaction pair, another important area in understanding the relationship between miRNAs and genes is to analyze multiple miRNAs and genes simultaneously by constructing miRNA-gene modules, each of which is composed of a group of miRNAs and their target genes collectively interacting in similar biological processes. It has been well-known that a single miRNA can target multiple genes, but the effect of a single miRNA on a given target is generally modest [31, 32]. It is often required that multiple miRNAs act cooperatively to exert significant regulation on their common target genes [33, 34]. Given the modular organization of miRNA regulatory networks, recent studies have aimed at identifying the many-to-many relationships between a set of co-expressed miRNAs and their target genes, organized as miRNA-gene regulatory modules. In this chapter, we present an overview of current computational and bioinformatics approaches that identify the regulatory modules of miRNAs and genes by integrating diverse genomic data.

2 Identifying MiRNA-Gene Modules by Integrating Heterogeneous Data Sources

2.1 Bipartite Graph-Based Methods

Bipartite graphs have been widely used for analyzing biological networks. A bipartite graph is defined as $G = (V, E)$, where V denotes two disjoint sets of nodes and E denotes a set of edges connecting the nodes. In a miRNA-gene interaction network, V consists of vertices of miRNAs and protein-coding genes, and E represents the weighted edges between the miRNA and gene vertices. Peng et al. [35] was among the first to identify miRNA-gene regulatory modules using a bipartite graph. Figure 1 depicts an overview of the proposed framework. Two complementary types of information are used in the analysis: expression data of both miRNAs and mRNAs and computational predictions of miRNA targets. The expression data allows the calculation of Pearson correlation coefficients for each miRNA-gene pair, and one large miRNA-gene correlation matrix can be created. Based on the assumption that the expression levels of miRNAs and their target genes are inversely correlated, a threshold value for Pearson correlation coefficient is chosen corresponding to an estimated false detection rate around 5%. By applying the threshold, the correlation

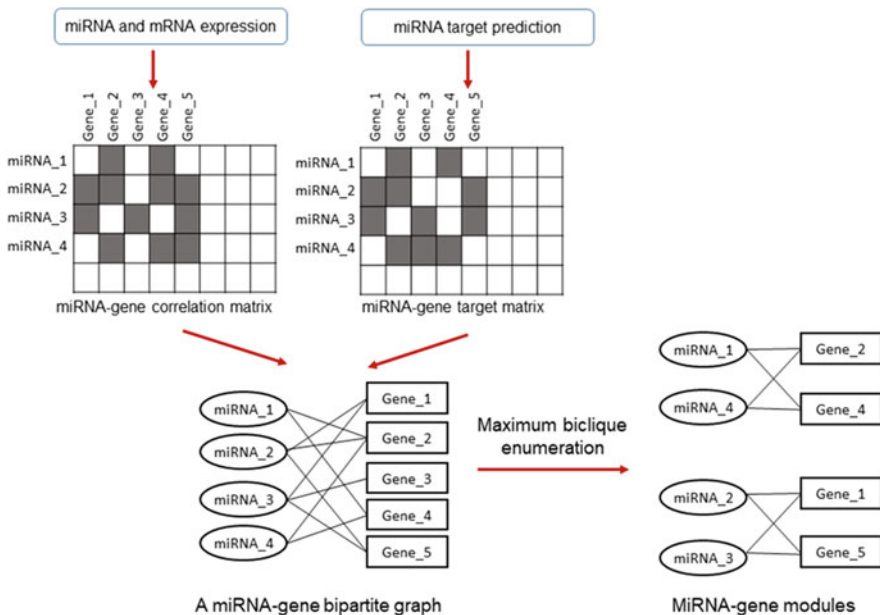


Fig. 1 Overview of a bipartite-based method. The gray shaded entries in the two matrices indicate the expression levels of the corresponding miRNA-gene pair are significantly correlated (in miRNA-gene correlation matrix) or the gene is predicted as the miRNA target based on sequence information

matrix is converted into a binary miRNA-gene correlation network. The resulting network is then combined with the miRNA-gene target matrix, which is predicted based on the seed matches, to generate an unweighted miRNA-gene bipartite graph. An edge is present between a miRNA and a gene if the expression level of the miRNA was highly correlated to that of the gene, and the gene is predicted to be the miRNA target. Within the miRNA-gene bipartite graph, a biclique corresponds to a miRNA-gene regulatory module, where every miRNA is connected to every gene in the same module. Therefore, the identification of miRNA-gene modules is transformed into a task of finding the maximal bicliques, which can be achieved with an implementation of the maximal biclique enumeration algorithm [36]. Each identified biclique is considered as a candidate miRNA-gene regulatory module, and then subject for further statistical significance assessment to filter out statistically insignificant modules.

While this graph theory-based method sets up a promising framework for discovering putative miRNA regulatory modules, it is argued that the biclique enumeration algorithm was originally proposed for general unipartite graphs and unadapted to the structure of bipartite graphs [37]. Another disadvantage of the method is that it searches for maximal bicliques, which could be too stringent because it requires that all miRNAs target all genes in each identified module [38]. However, it is well-known that some miRNA-gene interaction may be missing in the target prediction, so the all-to-all relationship between miRNAs and genes may not be present in all the modules. This restriction yields very small miRNA-gene modules, with most modules containing only one miRNA with many genes. The starlike structures of these identified modules may obscure the combinatorial regulatory effects mediated by multiple miRNAs. To add flexibility to module identification, Veksler-Lublinsky et al. [38] computed maximal quasi-bicliques, which allow some missing interactions between miRNAs and genes. More recently, Liang et al. [39] applied a biclique merging (BCM) method that iteratively merged the completely connected bipartite subgraphs based on their overlaps as well as the gene-gene interactions. To quantify the closeness between two modules, an overlapping scoring function is defined to facilitate the module merging process. The function indicated the relative edge weights gained from merging to modules. Therefore, the process generated modules with high density and functional enrichment.

As we can see, the essential step of the abovementioned bipartite graph-based methods is to construct a miRNA-gene regulatory network, which is a weighted or unweighted bipartite graph. Therefore, the performance of these methods is dependent on the accuracy and completeness of the graph and can be very sensitive to noise in the data sources. However, the input gene expression correlation and miRNA-gene target predictions used to construct the bipartite graph may contain erroneous miRNA-gene interactions and exclude false negatives as well, which adversely affects the quality of the identified miRNA regulatory modules. It is also noted that focusing on negative correlations between miRNA and gene expression profiles neglects the situation that miRNA can upregulate target genes [40]. Therefore, several studies have extended the bipartite graph by including indirect upregulating miRNA-gene interactions. For instance, miRMAP [37] takes

both negative and positive correlated miRNA-gene interactions as the input in constructing the bipartite graph and then compiles an integrated association matrix by incorporating computationally predicted miRNA target information. The miRMAP method uses the BUBBLE bi-clustering algorithm with simulated annealing search method to locate high correlated “seeds” within the integrated association matrix, and multiple seeds are expanded deterministically by adding correlated rows and columns up to a maximum threshold. The resulting submatrices correspond to different functional modules. Another example that constructs the weighted edges of bipartite graph beyond utilizing the negative expression correlation between miRNAs and mRNAs is the maximum weighted merger method (MWMM) [41]. The rationale of the method is that the expression correlation coefficients of a miRNA-mRNA pair can be changed from positive in normal to negative in tumor samples, or vice versa. The miRNA-gene pairs with inverse correlation coefficients between normal and tumor samples should be important in tumor progression. The method then computes an integrated mean value weight to quantify the correlation change of miRNA-mRNA pairs to represent the edges in the miRNA-mRNA bipartite graph. Finally, the modules are identified by applying the Hungarian and Blossom algorithms on the bipartite graph. Compared to other module identification methods, the MWMM method focuses on altered miRNA-mRNA correlations when constructing the bipartite graph, which helps identify tumor-specific miRNA-mRNA modules.

2.2 Nonnegative Matrix Factorization Methods

Nonnegative matrix factorization (NMF) technique assumes that data have an intrinsic low-dimensional nonnegative representation, with the low dimension corresponding to the number of miRNA-gene modules. Therefore, NMF method can be viewed as one of dimensional reduction techniques. It decomposes a nonnegative matrix into two lower rank matrices, a basis matrix W , and a coefficient matrix H , such that neither of these matrices contain negative elements. The matrix factorization can be achieved by minimizing the following objective function:

$$\min_{W, H \geq 0} \|X - WH\|_F^2 \quad (1)$$

Here X is a $p \times N$ observed omic matrix, W is a $p \times K$ matrix of basis vectors, and H is a $K \times N$ matrix of coefficient vectors, where K is the number of modules. The notation $\|\cdot\|_F$ indicates the Frobenius norm of a matrix.

The SNMNMf method. Using the NMF technique, Zhang et al. developed one of the earliest approaches that integrated miRNA and gene expression profiles in a multiple NMF framework, namely, the SNMNMf method [42]. An overview of the SNMNMf method is shown in Fig. 2. SNMNMf method extended the original NMF technique by simultaneously analyzing multiple matrices that represent different genomic data sources. The input are two sets of expression

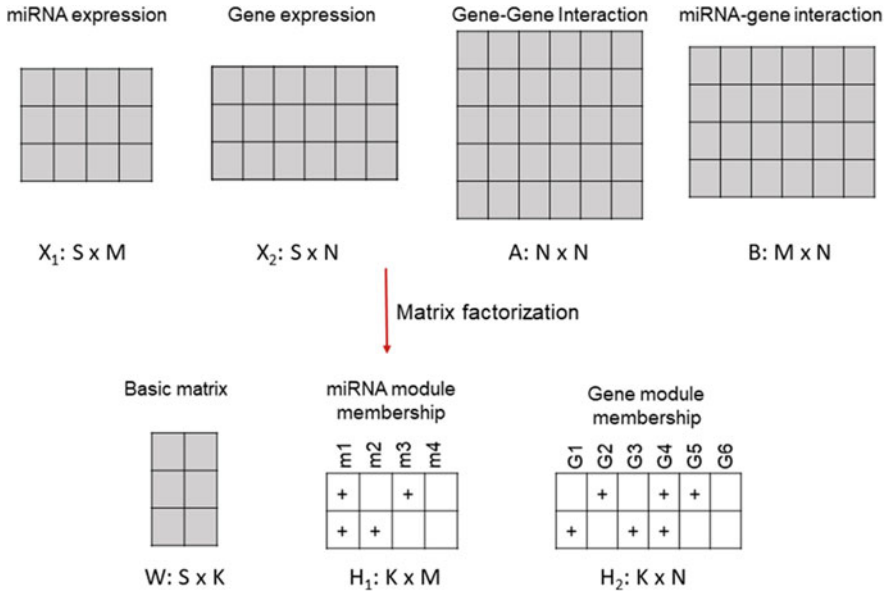


Fig. 2 Overview of a nonnegative matrix factorization method. The letters S , M , N , and K represent the number of samples, miRNAs, genes, and modules, respectively. Given the expression matrices X_1 and X_2 , matrix factorization is performed with sparsity constraints and network-regularized constraints imposed by matrices A and B , by minimizing the objective function in Eq. (2). This results in three matrices, the basic matrix W and the miRNA and gene module membership matrices H_1 and H_2 , where $X_1 \approx WH_1$ and $X_2 \approx WH_2$. If the elements in the same row on H_1 or H_2 are higher than a predefined threshold (indicated by a “+” sign), the corresponding miRNAs and genes are assigned to the same module. The example in this figure suggests there are two modules, where m_1, m_3, G_2, G_4 , and G_5 belong to one module and m_1, m_2, G_1, G_3 , and G_4 belong to the other

profiles for miRNAs and protein-coding genes, $X_1 \in \mathbb{R}^{S \times M}$ and $X_2 \in \mathbb{R}^{S \times N}$, a matrix $A \in \{0, 1\}^{N \times N}$ representing gene-gene interaction network, and a matrix $B \in \{0, 1\}^{M \times N}$ representing the list of predicted miRNA-gene regulatory interaction based on sequence information. Here S is the number of samples, and M and N represent the number of miRNAs and genes, respectively. The advantage of the SNMNMf method is, when the two expression matrices are factored into a common basis W and two coefficient matrices H_1 and H_2 , additional prior knowledge consisting of predicted miRNA-gene interactions and gene-gene interaction can be easily incorporated with network-regularized constraints. Sparsity constraints can also be imposed on this framework to make the coefficient matrices H_1 and H_2

sparse. The method is therefore formulated as minimizing the objective function as follows:

$$\begin{aligned} \mathcal{F}(W, H_1, H_2) = & \sum_{I=1,2} \|X_I - WH_I\|_F^2 - \lambda_1 \text{Tr}(H_2 A H_2^T) - \lambda_2 \text{Tr}(H_1 B H_1^T) \\ & + \gamma_1 \|W\|_F^2 + \gamma_2 \left(\sum_j \|h_j\|_1^2 + \sum_{j'} \|h_{j'}\|_1^2 \right) \end{aligned} \quad (2)$$

where $W \in \mathbb{R}^{S \times K}$ is the common basis matrix. In the specific problem of miRNA-gene module identification, K is the number of modules, which is set to 50 prior to optimization step. H_1 and H_2 are new representations of X_1 and X_2 on W . The parameters λ_1 and λ_2 are weights for the constraints defined in matrices A and B . The parameters γ_1 and γ_2 are used to constrain the growth of W and encourage the sparsity, respectively. By iteratively updating matrices W , H_1 , and H_2 in an alternating manner until the objective function converges to a local minimum, the matrices decomposition is learned. The decomposed matrices H_1 and H_2 are then used to determine miRNA-gene module membership. If the elements in the same row on H_1 or H_2 are higher than a predefined threshold, the corresponding miRNAs and genes are assigned to the same module. In this way, some miRNAs or genes can be included to multiple modules, while others may not be present in any module.

The NetNMF method. An alternative factorization approach to NMF is the tri-matrix factorization that can be used to not only identify miRNA-gene modules but also decipher the associations among identified modules. One such example of applying tri-matrix factorization technique is the NetNMF method [43]. Given the miRNA and gene expression data matrices X_1 and X_2 , three matrices R_{11} , R_{12} , and R_{22} are computed via Pearson correlation, where $R_{11} \in \mathbb{R}^{M \times M}$, $R_{22} \in \mathbb{R}^{N \times N}$ are symmetric similarity matrices corresponding to miRNAs and genes, respectively, and $R_{12} \in \mathbb{R}^{M \times N}$ corresponds to the similarities between them. Then NetNMF simultaneously decomposes R_{11} , R_{12} , and R_{22} to get the underlying modules assignment. Each similarity matrix R is factored into GSG^T . The objective function is formulated as

$$\begin{aligned} \min_{G_1, G_2, S_{11}, S_{22} \geq 0} & \|R_{11} - G_1 S_{11} G_1^T\|_F^2 + \lambda_1 \|R_{12} - G_1 G_2^T\|_F^2 \\ & + \lambda_2 \|R_{22} - G_2 S_{22} G_2^T\|_F^2 \end{aligned} \quad (3)$$

G_1 , G_2 , S_{11} , and S_{22} are the nonnegative factored matrices and provide a low-dimensional representation for input matrices. The term of $\|R_{12} - G_1 G_2^T\|$ identifies the one-to-one relationships between the miRNAs and genes, thus providing the miRNA-gene co-module membership. More specifically, the i th co-module is identified based on the i th column vector in the factored matrices G_1 and G_2 , while the association between the i th and j th module is determined by elements in matrices S_{11} or S_{22} .

The jNMF and iNMF methods. In the above matrix factorization methods, only the expression profiles of miRNAs and genes are factored into lower rank matrices. Several groups have aimed to extend the framework to include multiple types of genomic data. For example, Zhang et al. developed an extension for integrating DNA methylation data with the expression profiles of miRNAs and genes [44]. In this extension method jNMF, the sample is assumed to have the same low-dimensional representation for all three types of data. The method has successfully identified modules with significant functional associations when being applied to a TCGA ovarian cancer dataset. However, it was noted that the jNMF method is not methodologically different from standard NMF. It does not distinguish between different data sources and is thus sensitive to heterogeneous noise and confounding effects across sources [45]. To solve this issue, a new method iNMF has been developed that models heterogeneous effects among different data sources with an additional penalty term. The objective function in Eq. (1) is rewritten as

$$\min_{W, H_1, \dots, H_K, V_1, \dots, V_K} \geq 0 \sum_{k=1}^K \|X_k - (W + V_k) H_k\|_F^2 + \lambda \sum_{k=1}^K \|V_k H_k\|_F^2 \quad (4)$$

where K is the number of heterogeneous sources and $V_k H_k$ allows the model to represent heterogeneous effects differently for different data sources. Applied on a simulation study and a real ovarian cancer dataset, the iNMF method was found to be more robust to heterogeneous noise across the data sources than jNMF for module identification. Similarly, another study based on the pattern fusion analysis (PFA) framework identifies significant miRNA-gene modules from heterogeneous types of data by optimally adjusting the effects of each data type [46]. In particular, PFA first derives local sample patterns for every type of data independently. Then, it aligns these local sample patterns into a global sample pattern across multiple data types. During this process, the contributions of each data type are evaluated, and the bias can be iteratively decreased to better fit the data through an adaptive optimization strategy.

One limitation of the matrix factorization approaches is the requirement for a fixed number of modules, which may be difficult to predetermine before the matrix decomposition. In addition, the solution is often not unique and the computational complexity is often high, which makes reproducing and interpreting the prediction results difficult. Another major limitation of these methods is that the identified modules do not provide information on the regulation strength between a miRNA and a gene within a module. To address this issue, one recent study proposed the THEIA method that simultaneously learns the composition of miRNA-gene modules and the regulation strength and direction (upregulation or downregulation) of individual miRNA-gene interactions [47]. Unlike other NMF-based method that only factorizes expression matrices, THEIA factorizes both the gene-gene interaction and putative miRNA-gene interaction matrices to assemble miRNAs and genes into modules. It first obtains the lower-ranked gene membership matrix $V = (v_{jk}) \in [0, \infty)^{I \times K}$ by factorizing the gene-gene interaction matrix and then

learns the miRNA membership matrix $U = (u_{jk}) \in [0, \infty)^{J \times K}$ by factorizing the putative miRNA-gene interaction matrix, where I and J are the number of genes and miRNAs, respectively, and K is the number of modules. The matrix entries u_{ik} and v_{jk} denote the likelihood that the i th miRNA and j th gene belong to the k th module, respectively, and a greater magnitude indicates a greater chance of belonging to the module. By calculating UV^T , the regulation weight matrix W can be learned by a regression method. The value of w_{ij} estimates how strongly the i th miRNA regulates the j th gene. Further, the sign of w_{ij} defines the direction of regulation, such that negative values indicate downregulation and positive values indicate upregulation.

2.3 Statistical Modeling Approaches

The PIMiM method. A probabilistic regression-based model called protein interaction-based miRNA modules (PIMiM) was developed to identify miRNA modules [48]. Similar to other module identification methods we have discussed, the PIMiM uses miRNA and mRNA expression data as the input. In addition, it integrates the sequence-based prediction of miRNA-gene interactions and static protein-protein interaction data into the model. The overall goal of the method is to learn a regularized probabilistic regression model in which the gene expression can be written as a function of the miRNAs regulating the genes and the set of proteins the genes interact with. This module-based method assigned miRNAs and predicted genes to one of K modules, where K was a predetermined number. The assumption of the model is that the expression values of mRNAs are downregulated by a linear combination of expression profiles of all their predicted miRNA regulators. For example, mRNA j 's expression is distributed as $y_j \sim \mathcal{N}(\mu - \sum_{i \in S_j} w_{ij} X_i, \Sigma)$, where X and Y denote the expression profiles of miRNAs and mRNAs and μ is the baseline expression level without regulation. The weights associated with miRNAs i are denoted by w_{ij} , and S_j is the set of predicted miRNA regulators assigned to the modules where mRNA j belongs to. Let matrices U and V represent the entries of the miRNA and mRNA module membership, respectively. Φ and Ω are the lists of predicted miRNA-mRNA interactions and protein-protein interactions, respectively. Given these notations, the overall negative log-likelihood of the observed expression values is

$$\begin{aligned} \mathcal{L}(Y, X, \Phi, \Omega) = & -\log p(Y|U, V, X, \mu, \Sigma) \\ & - \sum_{i,j} \log p(I_{\phi i,j}|U, V) - \sum_{j \neq j'} \log p(I_{w_{jj'}} = 1|V) \end{aligned} \quad (5)$$

The first term optimizes the relationship between the observed miRNA and mRNA expression, and the second and third terms are rewards for assigning sequence-predicted miRNA-mRNA pairs and protein-protein interaction pairs to the

same module, respectively. To constrain the solutions, the method uses two sets of L1-norm to encourage sparsity leading to smaller and tighter modules. Specifically, the function is minimized under the constraints:

$$\|u_i\|_1 \leq C_1, i = 1, \dots, M \quad \text{and} \quad \|v_j\|_1 \leq C_2, j = 1, \dots, N \quad (6)$$

where C_1 and C_2 are two different regularization parameters for miRNAs and mRNAs, respectively, and chosen through an iterative line search. PMiM was found to detect modules with higher functional enrichment than the matrix factorization method using the ovarian cancer dataset as the test case, but one potential disadvantage of this supervised method is that the modules identified naturally tend toward the input data source.

The Mirsynergy method. Given the expression profiles of miRNAs and mRNAs, the Mirsynergy method [49] first infers an miRNA-mRNA interaction weights (MMIW) matrix \mathbf{W} using L1-norm regularized linear regression model (i.e., LASSO). Then the method goes through two clustering stages: In stage 1, the miRNA-miRNA synergistic scores s_{jk} between miRNA j and k are calculated as

$$s_{j,k} = \frac{\sum_{i=1}^N w_{ij} w_{ik}}{\min \left[\sum_i w_{ij} \sum_i w_{ik} \right]} \quad (7)$$

where w_{ij} is the weight for miRNA k targeting mRNA i based on the MMIW matrix. The synergy score $s(V_c)$ for any miRNA module V_c is then defined as

$$s(V_c) = \frac{w^{in}(V_c)}{w^{in}(V_c) + w^{bound}(V_c) + \alpha(V_c)} \quad (8)$$

where $w^{in}(V_c)$ and $w^{bound}(V_c)$ denote the total weights of the internal edges within a miRNA module and the total weights of the edges connecting the miRNAs within the module to those outside the module, respectively, and $\alpha(V_c)$ is the penalty scores for forming cluster V_c . Given the synergistic scores, miRNA clusters are formed with an overlapping neighborhood expansion clustering algorithm [50]. In state 2, a similar clustering algorithm is performed to assign only mRNAs to each miRNA module so that the synergy scores of the modules are maximized. In this stage, the edge weights are updated by combining the MMIW matrix and gene-gene interaction weight (GGIW) matrix that involves known transcription factor binding and protein-protein interaction information. Finally, the overlapping clustering assignments of miRNA-gene modules are identified after the modules with small density scores are filtered out. Mirsynergy was found to produce module structures that were highly dependent on initial clustering of miRNAs and the GGI data, but it has two major advantages: First, it is able to determine the module number automatically during iteration. Second, the computation is efficient, with theoretical bound reduced from $O(K(T+N+M)^2)$ per iteration to only $O(M(N+M))$ for N mRNAs and M miRNAs across T samples. Nonetheless, the

performance of Mirsynergy is sensitive to the quality of MMIW and GGIW. In this regard, other MMIW or GGIW matrices (generated from improved methods) can be easily incorporated into Mirsynergy as the function parameters.

Bayesian network method. Another approach that incorporates the GGI information with gene expression profiles is developed by Jin et al. [51]. This method combines a blustering algorithm and a Gaussian Bayesian network. First, based on the assumption that a subset of genes related to similar functions or pathways will have similar expression profiles in a subset of samples, the authors constructed the gene-sample modules using a SAMBA biclustering algorithm, which allows genes and samples to be included in multiple modules. By integrating the gene-gene interaction information, the modules are further expanded to include genes that interact directly with at least one gene in the module. This clustering step reduces the parameter space for the next step of Bayesian network modeling, where the gene-regulating miRNAs are selected to be added onto the gene-sample modules based on a Gaussian Bayesian network. Given the joint distribution of genes $X = \{X_1, X_2, \dots, X_n\}$ and miRNAs $Y = \{Y_1, Y_2, \dots, Y_m\}$, the likelihood of X and Y can be represented by

$$\mathcal{L}(X, Y) = P(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m) = \prod_{i=1}^n P(X_i | P_a^G(X_i)) \quad (9)$$

where the conditional probability of X_i , given its parents $P_a^G(X_i)$, can be represented by

$$P(X_i | P_a^G(X_i)) = p(X_i | Y_j, \dots, Y_k) \sim N\left(a_0 + \sum_{j'} a_{j'} \cdot Y_{j'}, \sigma^2\right) \quad (10)$$

The dependencies between expression values of miRNAs and genes are estimated by a Bayes information criterion (BIC), a measure that assesses the Bayesian network structure of miRNAs and genes:

$$BIC = \log(L) - \log(M)/2 + O(1) \quad (11)$$

where M is the sum of the number of miRNAs and genes. To constrain the search space, the authors only select candidate miRNAs whose average of absolute correlation coefficients for genes in a given module are in the top 7% among all miRNAs. It was found that the average number of enriched pathways in modules using this method was larger than that of the SNMNMf method when comparing the method performance on the ovarian cancer and glioblastoma datasets. However, the same research group later pointed out that using only the gene expression profiles might be limited in determining the relationships between miRNAs and genes, as mRNA expression is not sufficient to represent the gene regulation and protein

translation processes. Therefore, they improved this method by integrating protein expression data into the module identification framework [52].

RFCM³ method. An algorithm named the relevant and functionally consistent miRNA-mRNA modules (RFCM³) identifies potential miRNA-gene modules in cervical cancer based on mutual information calculation [53]. First, this method generates star-shaped modules containing only one miRNA and multiple genes by maximizing the functional similarity between the genes, as well as by maximizing relatedness between the miRNA and genes within a module. Mutual information is used to compute both the relevance and functional similarity between genes. Since the expression values are continuous, they need to be discretized to calculate the marginal and joint probabilities for further mutual information computation. Next, the star-shaped modules are merged by maximizing the similarity between their miRNAs in different modules. Because miRNAs with similar functions are most often associated with similar diseases, the relationship between miRNAs can be represented by a directed acyclic graph (DAG). Based on this DAG, a miRNA-miRNA similarity matrix can be constructed [54], which is further used to merge similar star-shaped modules. Finally, miRNA-gene modules are generated containing multiple miRNAs and genes. The authors claimed that the RFCM³ method generated more significant miRNA-gene regulatory modules highly related to cervical cancer, while the Mirsynergy and SNMNMf methods were unable to do. However, performance of this method highly relies on the miRNA similarity matrix, which may not be available on other than specific cancer types.

The three categories of computation approaches we review here may not be clearly distinguishable, as some of the algorithms presented here may fit into more than one category. For example, the method by Jin et al. is a statistic modeling approach but also uses bipartite graph to organize the miRNAs and genes into modules. Therefore, we describe these methods in the categories where we consider them to fit most. At the end of this chapter, we provide a list of major miRNA-gene module identification methods we have discussed (Table 1).

3 Evaluating the Performance of MiRNA-Gene Module Identification Methods

The availability of such a wide range of methods requires a comprehensive evaluation on their performance, simply because scientists are faced with a seemingly endless choice of methods for their data analyses. However, evaluating these module identification methods is a challenging task because there is no existing ground truth on the compositions of miRNA-gene modules. Nevertheless, one approach to validating these methods is to test their performance on simulated input datasets. In simulation studies, the parameters used to generate datasets can be controlled, and the underlying ground truth including the true module membership as well as the interaction strength between miRNAs and genes is known. Therefore, the similarity

Table 1 List of methods for identifying miRNA-gene modules

Methods/references	Data sources	Availability
<i>Bipartite graph-based methods</i>		
Peng et al. [35]	ME, GE, MT	Software upon request
miRMAP [37]	ME, GE, MT	https://www.dropbox.com/s/pnluebfg9to6iwiw/miRMAP.zip
Biclique merging [39]	ME, GE, MT	http://doi.ieeeecomputersociety.org/10.1109/TCBB.2015.2462370
NWMM [41]	ME, GE	https://github.com/BaiLab/MWMM
<i>Nonnegative matrix factorization methods</i>		
SNMMF [42]	ME, GE, MT, GGI	http://zhoulab.usc.edu/SNMMF/
NetNMF method [43]	ME, GE	http://www.zhanglab-amss.org/homepage/software.html
iNMF [45]	ME, GE, DM	https://github.com/yangzi4i/NMF
THEIA [47]	ME, GE, MT, GGI	https://bitbucket.org/cellsandmachines/theia/
<i>Statistical modeling approaches</i>		
PiMiM [48]	ME, GE, MT, GGI	Software upon request
Mirsynergy [49]	ME, GE, MT, GGI	http://www.cs.utoronto.ca/~yueli/Mirsynergy.html
Jin et al. [51]	ME, GE, GGI	Software upon request
RFCM ³ [53]	ME, GE, MS	http://home.ijit.ac.in/~sushmitapaul/CBL/software.html

ME miRNA expression, GE miRNA gene expression, MT predicted miRNA-gene interactions, GGI gene-gene interactions based on protein-protein interactions and DNA-protein interactions, DM DNA methylation, MS miRNA functional similarity

between modules predicted by computational methods and the true modules can be directly measured. The adjusted Rand index (ARI) has been used to compute the similarity between two modules by computing the percentage of element pairs that are assigned to the same module [47]. Other metrics for measuring module accuracy and quality include the normalized mutual information and topological properties such as module density and modularity [55]. However, while simulation studies provide datasets in which the ground truth is preset, these studies may oversimplify the biological systems when making assumptions to generate synthetic data. Therefore, many studies have relied on other evidence related to the biological significance of the identified modules for method evaluation. The underlying rationale is that the true miRNA-gene modules are likely biologically meaningful. As we will see, there is no such an evaluation method that can be both comprehensive and accurate for any types of input data. Therefore, studies often apply different methods in combination to provide a thorough and unbiased evaluation.

MiRNA family enrichment analysis. It is evident that members from the same miRNA family tend to be involved in the same biological functions [56]. Therefore, a miRNA family enrichment analysis can be used to verify whether the miRNAs within an identified miRNA-gene module are enriched in a miRNA family and thus participate cooperatively in gene regulation. A similar strategy to evaluate biological significance of the miRNAs within a module is by testing the spatial miRNA cluster enrichment of each module. Since most miRNAs within 50 kb tend to be co-expressed and regulate common target genes, spatially clustered miRNAs can be functionally related and assigned to the same module [57]. Both the miRNA family and spatial cluster information can be obtained from miRBase, which hosts information on miRNA sequences and family classification based on sequence similarity in the seed regions [58]. The hypergeometric test is performed to evaluate whether each module is significantly enriched in at least one miRNA family or miRNA spatial cluster after multiple testing correction. The main drawback of validation methods in this category is that they do not examine the module membership of target genes and the functional significance of genes within an identified module cannot be verified.

Functional enrichment analysis. In contrary to miRNA family enrichment analysis that focuses on miRNAs, functional enrichment analysis examines whether the target genes in each miRNA-gene module are functionally enriched in at least one Gene Ontology term, commonly in the ontology of “biological process” [59]. The GO terms often need to be preselected to exclude some terms with too many or too few associated genes. Since the analysis only focuses on target genes, the miRNA-gene relationship within each module is not assessed.

Analysis of miRNA-gene pairs within modules. One strategy to evaluate the predicted miRNA-gene interactions within each module is to examine the agreement between computational prediction and experimental results and assess the percentage of experimentally validated miRNA-gene interactions can be recovered in prediction results. The list of experimentally validated interactions can be downloaded from miRTarBase [60]. However, since the list is far from completeness, the absence of an miRNA-gene pair in miRTarBase does not necessarily

indicate the pair does not interact. In fact, some miRNA-gene interaction may have not yet been validated by experiments. Therefore, the specificity of a prediction method can be underestimated. However, the detection rate, which is the ratio of detected interactions to the total number of validated interactions, can be computed accurately and used to compare the performance among different methods. An alternative approach to verify the miRNA-gene interactions within a module is to examine the expression correlation between miRNAs and genes. The rationale of this evaluation method is that the expression levels of miRNA-gene interacting pairs are highly anticorrelated. The statistical significance of correlation between miRNAs and genes within a miRNA-gene module can be computed to evaluate the validity of the module. However, since many miRNA-gene pairs in the same module may not directly interact, and even if they interact, miRNAs could exert both positive and negative regulation on their target genes [40, 61], this evaluation approach has its limitation as well.

Implication of identified modules in cancer. Some studies have applied their computational methods on datasets that involve cancer patient samples, such as those using TCGA clinical data. Therefore, the identified modules are expected to be related to a specific type of cancer. To test this hypothesis, the miRNAs in the identified modules can be compared to a cancer-related miRNA benchmark dataset from miRCancer [62] and whether the identified miRNAs are enriched in miRCancer can be examined. In addition, whether the genes in each module are enriched in cancer-related pathways can also be analyzed by integrative pathway analysis [63]. Furthermore, the survival predictability of identified modules can be assessed. This is generally done by first dividing patients into two groups based on their expression profiles of miRNAs and genes in the module, and then performing the Kaplan-Meier survival analysis for patient samples to compare the survival characteristics between two patient groups. Using survival analysis to evaluate the module validity is only applicable on datasets with patient survival information.

4 Discussion

So far we have discussed methods for identifying miRNA-gene modules using one condition-specific expression dataset. Recent availability of miRNA and gene expression across multiple related conditions, such as different types of cancer, has motivated studies for characterizing the similarities and differences in miRNA-gene modules identified across multiple conditions [64]. For example, the PiMiM method we previously discussed was also used to integrate multiple types of cancers to learn a set of common modules for different cancer types. PiMiM uses a L1/L2 penalty of group lasso to regularize the modules over multiple conditions, so that it encourages miRNAs and genes to be assigned to the same modules across conditions [48]. While PiMiM focuses on identifying common miRNA-gene functional modules across different cancer types, the tensor sparse canonical correlation analysis (TSCCA) method aims at identifying cancer-specific modules [65]. TSCCA is a

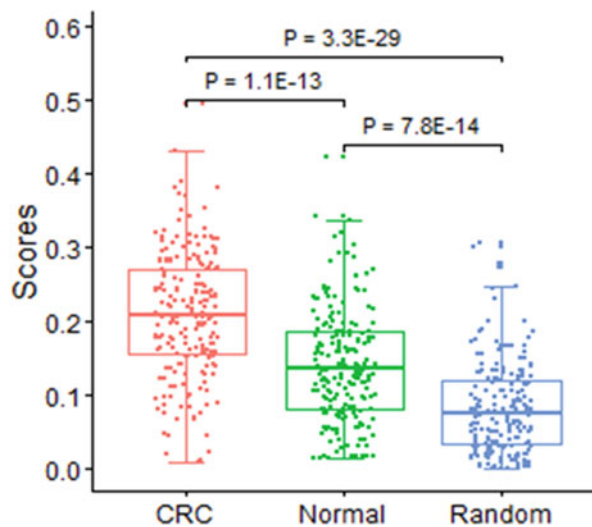
natural extension of matrix factorization method with the use of tensor, which are higher order matrices. In this framework, given the matched miRNA and gene expression matrices of multiple types of cancer, a cancer-miRNA-gene Pearson correlation tensor is computed as a “3D” array with $p \times q \times M$ dimensions, where p , q , and M represent the number of genes, miRNAs, and cancers, respectively. The goal is to decompose the correlation tensor into multiple sparse latent factors to represent the relative contribution of genes, miRNAs, and cancers. The nonzero entries on the same row in the latent factors correspond to a cancer-specific miRNA-gene module. Another recent study combines the multivariate regression model and matrix factorization technique to identify cancer-specific miRNA-gene modules [66]. The advantage of this method is that it can estimate the effective number of latent factors by incorporating the parameter into a regularized factor regression model, so that it does not need to take the number of modules as an input parameter. Nevertheless, the joint analysis of multiple conditions to identify common and divergent modules across conditions presents additional challenges, including the confounding effects due to the difference in experimental platforms and sample heterogeneity. Therefore, future improvements in module identification tools that effectively leverage information from multiple conditions are anticipated.

Since miRNAs are not the only molecules that play important roles in gene regulation, recent studies have aimed at incorporating other gene regulators, such as transcription factors (TFs) and long noncoding RNAs (lncRNAs) into miRNA-gene modules. Transcription factors play a major role in gene transcription, and they have been shown to work with miRNAs to regulate gene expression. In feed-forward loops or feed-back loops (FFLs), TF and miRNA can regulate each other so that TF may regulate the expression of a miRNA and a miRNA may repress a TF and both of them can jointly regulate target gene expression [67]. Given the gene expression profiles, there has been an increasing number of studies that incorporate the miRNA-TF regulations information into miRNA-gene regulatory network [68, 69]. However, our current knowledge on the regulation between miRNAs and TFs is very limited for understanding their cooperative effects on gene regulation in different physiological and pathological conditions. Algorithms have been proposed to predict TF-miRNA regulations by combining TF binding motifs, CHIP-Seq data, and transcriptome profiles [70]. With recent development of deepCAGE sequencing and nuclear run-on techniques that facilitate the annotation of miRNA gene transcription start sites [71, 72], resources have been established for TF-miRNA regulations by incorporating the information of the locations of cell-specific miRNA promoters [73–75], or based on manual literature curation [76]. In the future, as the annotation of miRNA transcription starting sites becomes more complete and accurate, we expect the methods for studying co-regulation of miRNAs and TFs will have better performance, which will help us refine the list of key regulators.

Another noncoding RNA class, lncRNAs, can also regulate mRNAs via diverse mechanisms [77]. In addition, miRNAs and lncRNAs can regulate each other through their binding sites. lncRNAs harbor miRNA-binding sites and act as miRNA sponges by competing with mRNAs for miRNA binding and thereby

relieving miRNA-mediated targets repression [78]. Conversely, lncRNA stability can be reduced through the interaction with specific miRNAs [79]. The interplay between them is important in modulating gene expression [80]. We have tested the hypothesis that lncRNA-miRNA-mRNA competing interactions are dynamic across different conditions. First, we identified candidate lncRNA-mRNA competing interactions by collecting a list of miRNA-mRNA and miRNA-lncRNA pairs from TargetScan 7.2 [26] and DIANA-LncBase v3 [81], and assessing whether there is a significant number of shared miRNAs for each lncRNA-mRNA competing pair with the cumulative hypergeometric test [82]. After obtaining the lncRNA-mRNA competing pairs with $FDR < 0.05$, we evaluated the strength of competition for each pair, using a dataset of RNA expression from a cohort of 635 colorectal cancer patients in TCGA data portal [83]. We defined the competing activity score as $(|corr_{lm}| + |corr_{mg}| + |corr_{lg}|)/3$ for each lncRNA-miRNA-mRNA competing triplet, where $corr_{lm}$, $corr_{mg}$, and $corr_{lg}$ represent Pearson correlations for lncRNA-miRNA, miRNA-mRNA, and lncRNA-mRNA pairs based on expression data, respectively. A higher competing activity score indicates greater competition between the lncRNA and mRNA for miRNA binding. We showed lncRNA H19-mediated lncRNA-miRNA-mRNA triplets in Fig. 3 as an example. For each triplet, five random competing scores were generated by randomly shuffling the expression profiles. Our results indicated the H19-mediated competing activity scores in samples of colorectal cancer were significantly higher than those in normal and random samples. The results suggest that the lncRNA-miRNA-mRNA competing interactions are dynamic across different conditions and could play an important role in cancer progression. This is consistent with experimental studies that have shown lncRNA H19 promotes tumor proliferation through competitively binding to a number of miRNAs [84–86]. Therefore, it will be critical to include

Fig. 3 Box plot comparison of competing scores of lncRNA-miRNA-mRNA triplets mediated by H19 among colorectal cancer, normal, and random samples. Significant p-values were calculated by the Wilcoxon signed rank test. *CRC* colorectal cancer



this competing regulatory relationship in the inference of ncRNA-mediated regulatory network, as shown in some recent studies. For example, based on joint orthogonality nonnegative matrix factorization, the CeModule method detected lncRNA-miRNA-mRNA regulatory modules on TCGA samples [87]. A graph-based method (EPLMI) was proposed to predict lncRNA-miRNA interactions using two-way diffusion [88]. These computational studies take advantage of lncRNA expression profiles, without considering how lncRNA sequences and structural features related to their regulatory effects. The integrated knowledge of these features, including the lncRNA sequences, expression, and structural organization, will increase our understanding of lncRNAs' functions and their interaction with miRNAs and protein-coding genes.

5 Conclusions

Gene regulation is dynamic and complex. MiRNAs have been recognized as one of the most important players in gene regulatory. With the availability of large amount of sequence information and high-throughput technologies, there has been a surge of computational methods for identifying miRNA-gene modules in the last decade. Meanwhile, methods have been developed to prioritize condition-specific modules, such as those related to a specific type of cancer. Undoubtedly, all of these studies provide valuable insights to characterize the combinatorial effects of miRNAs on the post-transcriptional gene regulation.

References

1. Cho KHT, Xu B, Blenkiron C, Fraser M (2019) Emerging roles of miRNAs in brain development and perinatal brain injury. *Front Physiol* 10
2. DeVeale B, Swindlehurst-Chan J, Billeloch R (2021) The roles of microRNAs in mouse development. *Nat Rev Genet* 22:307–323
3. Fu H, Zhou F, Yuan Q, Zhang W, Qiu Q, Yu X, He Z (2019) MiRNA-31-5p mediates the proliferation and apoptosis of human spermatogonial stem cells via targeting JAZF1 and cyclin A2. *Mol Ther-Nucleic Acids* 14:90–100
4. Otto T, Candido SV, Pilarz MS, Sicinska E, Bronson RT, Bowden M, Lachowicz IA, Mulry K, Fassl A, Han RC (2017) Cell cycle-targeting microRNAs promote differentiation by enforcing cell-cycle exit. *Proc Natl Acad Sci* 114:10660–10665
5. Inui M, Martello G, Piccolo S (2010) MicroRNA control of signal transduction. *Nat Rev Mol Cell Biol* 11:252–263
6. Mukherjee S, Paricio N, Sokol NS (2021) A stress-responsive miRNA regulates BMP signaling to maintain tissue homeostasis. *Proc Natl Acad Sci* 118
7. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microRNAs on protein output. *Nature* 455:64–71
8. Gebert LF, MacRae IJ (2019) Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* 20:21–37

9. Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11:597–610
10. Friedman RC, Farh KK-H, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19:92–105
11. Catalanotto C, Cogoni C, Zardo G (2016) MicroRNA in control of gene expression: an overview of nuclear functions. *Int J Mol Sci* 17:1712
12. Cui M, Wang H, Yao X, Zhang D, Xie Y, Cui R, Zhang X (2019) Circulating microRNAs in cancer: potential and challenge. *Front Genet* 10:626
13. Peng Y, Croce CM (2016) The role of MicroRNAs in human cancer. *Signal Transduct Target Ther* 1:1–9
14. Jużwik CA, Drake SS, Zhang Y, Paradis-Isler N, Sylvester A, Amar-Zifkin A, Douglas C, Morquette B, Moore CS, Fournier AE (2019) microRNA dysregulation in neurodegenerative diseases: a systematic review. *Prog Neurobiol* 182:101664
15. Moradifard S, Hoseinbeyki M, Ganji SM, Minuchehr Z (2018) Analysis of microRNA and gene expression profiles in Alzheimer’s disease: a meta-analysis approach. *Sci Rep* 8:1–17
16. Wang M, Qin L, Tang B (2019) MicroRNAs in Alzheimer’s disease. *Front Genet* 10:153
17. Wahlquist C, Jeong D, Rojas-Muñoz A, Kho C, Lee A, Mitsuyama S, van Mil A, Park WJ, Sluijter JP, Doevendans PA (2014) Inhibition of miR-25 improves cardiac contractility in the failing heart. *Nature* 508:531–535
18. Boon RA, Iekushi K, Lechner S, Seeger T, Fischer A, Heydt S, Kaluza D, Tréguer K, Carmona G, Bonauer A (2013) MicroRNA-34a regulates cardiac ageing and function. *Nature* 495:107–110
19. Hayes J, Peruzzi PP, Lawler S (2014) MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med* 20:460–469
20. Thompson AG, Gray E, Heman-Ackah SM, Mäger I, Talbot K, El Andaloussi S, Wood MJ, Turner MR (2016) Extracellular vesicles in neurodegenerative disease—pathogenesis to biomarkers. *Nat Rev Neurol* 12:346–357
21. Rupaimoole R, Slack FJ (2017) MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat Rev Drug Discov* 16:203–222
22. Gabisonia K, Prosdocimo G, Aquaro GD, Carlucci L, Zentilin L, Secco I, Ali H, Braga L, Gorgodze N, Bernini F (2019) MicroRNA therapy stimulates uncontrolled cardiac repair after myocardial infarction in pigs. *Nature* 569:418–422
23. Chen L, Heikkinen L, Wang C, Yang Y, Sun H, Wong G (2019) Trends in the development of miRNA bioinformatics tools. *Brief Bioinform* 20:1836–1852
24. Grimson A, Farh KK-H, Johnston WK, Garrett-Engle P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27:91–105
25. Krüger J, Rehmsmeier M (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* 34:W451–W454
26. Agarwal V, Bell GW, Nam J-W, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. *elife* 4:e05005
27. Xu W, San Lucas A, Wang Z, Liu Y (2014) Identifying microRNA targets in different gene regions. *BMC Bioinformatics* 15:1–11
28. Wang Z, Xu W, Liu Y (2015) Integrating full spectrum of sequence features into predicting functional microRNA–mRNA interactions. *Bioinformatics* 31:3529–3536
29. Akhtar MM, Micolucci L, Islam MS, Olivieri F, Procopio AD (2016) Bioinformatic tools for microRNA dissection. *Nucleic Acids Res* 44:24–44
30. Nazarov PV, Kreis S (2021) Integrative approaches for analysis of mRNA and microRNA high-throughput data. *Comput Struct Biotechnol J*
31. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136:215–233
32. Cherone JM, Jorgji V, Burge CB (2019) Cotargeting among microRNAs in the brain. *Genome Res* 29:1791–1804
33. Hausser J, Zavolan M (2014) Identification and consequences of miRNA-target interactions—beyond repression of gene expression. *Nat Rev Genet* 15:599–612

34. Xu J, Li C-X, Li Y-S, Lv J-Y, Ma Y, Shao T-T, Xu L-D, Wang Y-Y, Du L, Zhang Y-P (2011) MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res* 39:825-836
35. Peng X, Li Y, Walters K-A, Rosenzweig ER, Lederer SL, Aicher LD, Proll S, Katze MG (2009) Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics* 10:1-20
36. Alexe G, Alexe S, Crama Y, Foldes S, Hammer PL, Simeone B (2004) Consensus algorithms for the generation of all maximal bicliques. *Discret Appl Math* 145:11-21
37. Bryan K, Terrile M, Bray IM, Domingo-Fernandez R, Watters KM, Koster J, Versteeg R, Stallings RL (2014) Discovery and visualization of miRNA-mRNA functional modules within integrated data using bicluster analysis. *Nucleic Acids Res* 42:e17-e17
38. Veksler-Lublinsky I, Shemer-Avni Y, Meiri E, Bentwich Z, Kedem K, Ziv-Ukelson M (2012) Finding quasi-modules of human and viral miRNAs: a case study of human cytomegalovirus (HCMV). *BMC Bioinformatics* 13:1-18
39. Liang C, Li Y, Luo J (2015) A novel method to detect functional microRNA regulatory modules by bicliques merging. *IEEE/ACM Trans Comput Biol Bioinform* 13:549-556
40. Tan H, Huang S, Zhang Z, Qian X, Sun P, Zhou X (2019) Pan-cancer analysis on microRNA-associated gene activation. *EBioMedicine* 43:82-97
41. Ding L, Feng Z, Bai Y (2019) Clustering analysis of microRNA and mRNA expression data from TCGA using maximum edge-weighted matching algorithms. *BMC Med Genet* 12:1-27
42. Zhang S, Li Q, Liu J, Zhou XJ (2011) A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules. *Bioinformatics* 27:i401-i409
43. Chen J, Zhang S (2018) Discovery of two-level modular organization from matched genomic data via joint matrix tri-factorization. *Nucleic Acids Res* 46:5967-5976
44. Zhang S, Liu C-C, Li W, Shen H, Laird PW, Zhou XJ (2012) Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Res* 40:9379-9391
45. Yang Z, Michailidis G (2016) A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* 32:1-8
46. Shi Q, Zhang C, Peng M, Yu X, Zeng T, Liu J, Chen L (2017) Pattern fusion analysis by adaptive alignment of multiple heterogeneous omics data. *Bioinformatics* 33:2706-2714
47. Roth M, Jain P, Koo J, Chaterji S (2021) Simultaneous learning of individual microRNA-gene interactions and regulatory comodules. *BMC Bioinformatics* 22:1-29
48. Le H-S, Bar-Joseph Z (2013) Integrating sequence, expression and interaction data to determine condition-specific miRNA regulation. *Bioinformatics* 29:i89-i97
49. Li Y, Liang C, Wong K-C, Luo J, Zhang Z (2014) Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion. *Bioinformatics* 30:2627-2635
50. Nepusz T, Yu H, Paccanaro A (2012) Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods* 9:471-472
51. Jin D, Lee H (2015) A computational approach to identifying gene-microRNA modules in cancer. *PLoS Comput Biol* 11:e1004042
52. Seo J, Jin D, Choi C-H, Lee H (2017) Integration of microRNA, mRNA, and protein expression data for the identification of cancer-related microRNAs. *PLoS One* 12:e0168412
53. Paul S (2019) RFCM 3: computational method for identification of miRNA-mRNA regulatory modules in cervical cancer. *IEEE/ACM Trans Comput Biol Bioinform* 17:1729-1740
54. Wang D, Wang J, Lu M, Song F, Cui Q (2010) Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26:1644-1650
55. Yang Y, Song Y (2019) A stacked autoencoder-based miRNA regulatory module detection framework. *Int J Comput Intell Syst* 12:822-832
56. Wang Q, Wei L, Guan X, Wu Y, Zou Q, Ji Z (2014) Briefing in family characteristics of microRNAs and their applications in cancer research. *Biochim Biophys Acta (BBA)-Proteins Proteom* 1844:191-197

57. Wang Y, Luo J, Zhang H, Lu J (2016) microRNAs in the same clusters evolve to coordinately regulate functionally related genes. *Mol Biol Evol* 33:2232–2247
58. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42:D68–D73
59. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47:D419–D426
60. Huang H-Y, Lin Y-C-D, Li J, Huang K-Y, Shrestha S, Hong H-C, Tang Y, Chen Y-G, Jin C-N, Yu Y (2020) miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res* 48:D148–D154
61. Xu P, Wu Q, Yu J, Rao Y, Kou Z, Fang G, Shi X, Liu W, Han H (2020) A systematic way to infer the regulation relations of miRNAs on target genes and critical miRNAs in cancers. *Front Genet* 11:278
62. Xie B, Ding Q, Han H, Wu D (2013) miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics* 29:638–644
63. Paczkowska M, Barenboim J, Sintupisut N, Fox NS, Zhu H, Abd-Rabbo D, Mee MW, Boutros PC, Reimand J (2020) Integrative pathway enrichment analysis of multivariate omics data. *Nat Commun* 11:1–16
64. Walsh CJ, Hu P, Batt J, Santos CCD (2016) Discovering microRNA-regulatory modules in multi-dimensional cancer genomic data: a survey of computational methods. *Cancer Informat* 15:CIN-S39369
65. Min W, Chang T-H, Zhang S, Wan X (2021) TSCCA: a tensor sparse CCA method for detecting microRNA-gene patterns from multiple cancers. *PLoS Comput Biol* 17:e1009044
66. Mokhtardoost M, Gönen M (2020) An efficient framework to identify key miRNA–mRNA regulatory modules in cancer. *Bioinformatics* 36:i592–i600
67. Zhang H-M, Kuang S, Xiong X, Gao T, Liu C, Guo A-Y (2015) Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Brief Bioinform* 16:45–58
68. Maulik U, Sen S, Mallik S, Bandyopadhyay S (2018) Detecting TF-miRNA-gene network based modules for 5hmC and 5mC brain samples: an intra-and inter-species case-study between human and rhesus. *BMC Genet* 19:1–22
69. Qin G, Mallik S, Mitra R, Li A, Jia P, Eischen CM, Zhao Z (2020) MicroRNA and transcription factor co-regulatory networks and subtype classification of seminoma and non-seminoma in testicular germ cell tumors. *Sci Rep* 10:1–14
70. Ruffalo M, Bar-Joseph Z (2016) Genome wide predictions of miRNA regulation by transcription factors. *Bioinformatics* 32:i746–i754
71. Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT (2014) Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* 46:1311–1320
72. DGT RPC and Consortium F (2014) A promoter-level mammalian expression atlas. *Nature* 507:462–470
73. Hua X, Tang R, Xu X, Wang Z, Xu Q, Chen L, Wingender E, Li J, Zhang C, Wang J (2018) mirTrans: a resource of transcriptional regulation on microRNAs for human cell lines. *Nucleic Acids Res* 46:D168–D174
74. Zhou K-R, Liu S, Sun W-J, Zheng L-L, Zhou H, Yang J-H, Qu L-H (2016) ChIPBase v2. 0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res*:gkw965
75. Wang S, Talukder A, Cha M, Li X, Hu H (2021) Computational annotation of miRNA transcription start sites. *Brief Bioinform* 22:380–392
76. Tong Z, Cui Q, Wang J, Zhou Y (2019) TransmiR v2. 0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res* 47:D253–D258
77. Szczesniak MW, Makalowska I (2016) lncRNA-RNA interactions across the human transcriptome. *PLoS One* 11:e0150353

78. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146:353–358
79. Anastasiadou E, Jacob LS, Slack FJ (2018) Non-coding RNA networks in cancer. *Nat Rev Cancer* 18:5–18
80. Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22:96–118
81. Karagkouni D, Paraskevopoulou MD, Tastsoglou S, Skoufos G, Karavangeli A, Pierros V, Zacharopoulou E, Hatzigeorgiou AG (2020) DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res* 48:D101–D110
82. Zhang Y, Tao Y, Li Y, Zhao J, Zhang L, Zhang X, Dong C, Xie Y, Dai X, Zhang X, Liao Q (2018) The regulatory network analysis of long noncoding RNAs in human colorectal cancer. *Funct Integr Genomics*
83. Xiong Y, Wang R, Peng L, You W, Wei J, Zhang S, Wu X, Guo J, Xu J, Lv Z, Fu Z (2017) An integrated lncRNA, microRNA and mRNA signature to improve prognosis prediction of colorectal cancer. *Oncotarget* 8:85463–85478
84. Liang WC, Fu WM, Wong CW, Wang Y, Wang WM, Hu GX, Zhang L, Xiao LJ, Wan DC, Zhang JF, Waye MM (2015) The lncRNA H19 promotes epithelial to mesenchymal transition by functioning as miRNA sponges in colorectal cancer. *Oncotarget* 6:22513–22525
85. Ou L, Wang D, Zhang H, Yu Q, Hua F (2017) Decreased expression of MiR-138-5p by LncRNA H19 in cervical cancer promotes tumor proliferation. *Oncol Res*
86. Wang SH, Ma F, Tang ZH, Wu XC, Cai Q, Zhang MD, Weng MZ, Zhou D, Wang JD, Quan ZW (2016) Long non-coding RNA H19 regulates FOXM1 expression by competitively binding endogenous miR-342-3p in gallbladder cancer. *J Exp Clin Cancer Res* 35:160
87. Xiao Q, Luo J, Liang C, Cai J, Li G, Cao B (2019) CeModule: an integrative framework for discovering regulatory patterns from genomic data in cancer. *BMC Bioinformatics* 20:1–13
88. Huang Y-A, Chan KC, You Z-H (2018) Constructing prediction models from expression profiles for large scale lncRNA–miRNA interaction profiling. *Bioinformatics* 34:812–819