Inferring MicroRNA-Disease Associations by Random Walk on a Heterogeneous Network with Multiple Data Sources

Yuansheng Liu, Xiangxiang Zeng, Zengyou He, and Quan Zou

Abstract—Since the discovery of the regulatory function of microRNA (miRNA), increased attention has focused on identifying the relationship between miRNA and disease. It has been suggested that computational method is an efficient way to identify potential disease-related miRNAs for further confirmation using biological experiments. In this paper, we first highlighted three limitations commonly associated with previous computational methods. To resolve these limitations, we established disease similarity subnetwork and miRNA similarity subnetwork by integrating multiple data sources, where the disease similarity is composed of disease semantic similarity and disease functional similarity, and the miRNA similarity is calculated using the miRNA-target gene and miRNA-lncRNA (long non-coding RNA) associations. Then, a heterogeneous network was constructed by connecting the disease similarity subnetwork and the miRNA similarity subnetwork using the known miRNA-disease associations. We extended random walk with restart to predict miRNA-disease associations in the heterogeneous network. The leave-one-out cross-validation achieved an average area under the curve (AUC) of 0.8049 across 341 diseases and 476 miRNAs. For five-fold cross-validation, our method achieved an AUC from 0.7970 to 0.9249 for 15 human diseases. Case studies further demonstrated the feasibility of our method to discover potential miRNA-disease associations. An online service for prediction is freely available at http://ifmda.aliapp.com.

Index Terms—MiRNA-disease association prediction, similarity measurement, random walk with restart

1 Introduction

MicroRNA (miRNA) is a small non-coding RNA molecule (containing about 22 nucleotides). Since the first miRNA, lin-4, was discovered by Lee et al. in 1993 [1], many miRNAs have been discovered [2], [3], [4], [5], [6] and miRBase [7] has now accumulated more than 28,000 of them. However, the regulatory function of miRNAs was not recognized until the early 2000s [8]. Since then, miRNA research has revealed multiple roles for miRNAs in many biological processes [9], [10], [11], [12], [13], such as proliferation, apoptosis and cell differentiation. Just as the miRNA involves in the normal functioning of eukaryotic cells, its dysregulation has been shown to be associated with disease [14], [15], [16]. For example, an increasing number of studies have indicated that many miRNAs are associated with the development of various cancers [17], [18], [19], [20], [21]. Identifying the relationship between miRNA and disease has significant contribution to biomedical research. The knowledge on disease-related miRNAs can promote our understanding of disease pathogenesis at the molecular level [22], [23], and benefit the design of molecular tools for disease diagnosis and treatment [24]. Moreover, miRNA-based therapies are also under investigation [25], [26]. Against this background, there is growing demand to identify miRNA-disease associations, which has led to intensive research activities in the biomedical field.

Classical biological experiments have generally been used to identify miRNA-disease associations. However, they are very time-consuming and great expense. With an increasing number of miRNAs having been discovered, there are a large number of miRNA-disease interactions awaiting identification. It is necessary to select miRNA-disease interactions with the greatest potential for further biological experiments to reduce the number of resources used for such biological experiments. To achieve this objective, computational methods have been a focus of substantial research efforts since the first computational method based on hypergeometric distribution was proposed in 2010 [27]. Some methods based on biological networks have also been developed for predicting miRNA-disease associations [28], [29], [30], [31], [32], [33], [34], [35], [36]. Zou and Zeng et al. reviewed different similarity computation strategies and compared existing computational methods [37], [38]. Such methods exhibit various flaws and limitations, so it is difficult for other researchers to compare and evaluate their performance. Therefore, elaborate biological networks must be constructed to successfully employ computational methods to infer miRNA-disease associations.

This paper makes three innovative contributions. First, we report three limitations commonly associated with previous computational methods, namely, the use of only a single dataset, the inadequacy of disease semantic similarity,

- *Y. Liu is with the Advanced Analytics Institute, FEIT, University of Technology Sydney, Broadway, NSW 2007, Australia.*
- *X. Zeng is with the Department of Computer Science, Xiamen Unviersity, Xiamen 361005, Fujian, China.*
- *Z. He is with the School of Software, Dalian University of Technology, Dalian 116024, China. E-mail: zyhe@dlut.edu.cn.*
- *Q. Zou is with the School of Computer Science and Technology, Tianjin University, Tianjin 300072, China. E-mail: zouquan@nclab.net.*

Manuscript received 28 Nov. 2015; revised 11 Feb. 2016; accepted 24 Mar. 2016. Date of publication 5 Apr. 2016; date of current version 4 Aug. 2017.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2016.2550432

1545-5963 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

TABLE 1
Comprehensive Comparison of Datasets

Method	Datasets
RWRMDA [28]	MeSH, HMDD
NetCBI [31]	MeSH, HMDD, MimMiner
RLSMDA [33]	MeSH, HMDD
HDMP [32]	MeSH, HMDD, MimMiner (miRNA families and clusters)
MIDP & MIDPE [35]	MeSH, HMDD

and overestimation of the predictive accuracy, and make some suggestions to guide further design of predictive methods. Second, to resolve the problem caused by the use of a single dataset, we average the disease semantic similarity and disease functional similarity to obtain a more comprehensive measure of the disease similarity. The data on the miRNA-related target gene and long non-coding RNA (lncRNA) are also combined to improve the accuracy of miRNA similarity. The two similarity subnetworks are connected by experimentally verified miRNA-disease associations, and then a heterogeneous network is constructed. Third, we develop a random walk with restart (RWR) on the heterogeneous network to infer potential miRNA-disease associations. The area under the curve (AUC) of leave-one-out cross-validation is 0.8049, which provides significant improvement over the two methods RWRMDA [28] and RLSMDA [33]. The AUC values of five-fold cross-validation for 15 diseases range from 0.7970 to 0.9249. Our method is also evaluated by applying it to diseases with no known related miRNAs and most of the top 30 candidates identified in this way are confirmed by various databases.

2 Overview of Three Limitations Commonly Associated with Previous Computational Methods

Zou and Zeng et al. reviewed the different similarity computation models and different predictive methods [37], [38]. They also identified some issues that impeded the improvements in the performance of predictions, such as the problem of imbalanced and unlabeled data of machine learning methods, and high false-positive and false-negative rates. In this section, we further reveal three hidden limitations commonly associated with previous methods of predicting miRNA-disease associations. We also correspondingly provide three suggestions on the design of new predictive methods, which would act as a common framework to improve their performance.

For better comparisons with existing computational methods and improving future design of predictive methods, we summarize and discuss the limitations of the existing methods as follows:

- *Use of a single dataset*. Zeng et al. [38] compared predictive methods for disease-related miRNAs in terms of the type of networks, and the type of dataset, among other variables. Here, we present a comprehensive analysis of the datasets that are employed in some existing predictive methods. In accordance with Wang et al. [39], the dataset MISIM is obtained through experimentally verified miRNA-disease associations and Medical Subject Headings (MeSH) tree structures. Therefore, MISIM consists of two datasets: MeSH and human miRNA-disease database (HMDD). A comparison of the datasets used in five reported studies is shown in Table 1. It reveals that only limited biological data are employed in those computational methods.

TABLE 2
The Number of Values of the Semantic Similarity above Zero

Disease name	Numbers
Acute coronary syndrome	103
Brain injuries	56
Crohn disease	40
Heart failure	42
Hodgkin disease	156
Lung diseases	41
Lung diseases, interstitial	41
Nerve sheath neoplasms	172
Schistosomiasis	1
Wounds and injuries	5

Suggested rule 1. To provide higher predictive accuracy, more biological datasets should be used.

- *Inadequacy of disease semantic similarity*. Disease semantic similarity is widely used to measure the miRNA similarity [39] and to predict miRNA-disease associations [32], [33], [35]. The basic approach involves calculating the disease semantic similarity from MeSH, which gives a hierarchical link between diseases. However, many zeros are present in the disease semantic similarity matrix. To reflect this problem, we randomly select 10 from among 341 different diseases as used in Section 3.3. For each disease, we count the number of values of the semantic similarity with the other 340 diseases that are above zero, the results of which are listed in Table 2. The average number of values of the semantic similarity above zero for the 341 diseases is 86. As an extreme case, only one disease is related with the disease schistosomiasis. This indicates that the semantic measurement method identifies many diseases as not being related to each other, despite many of them actually are closely related. The number of disease-related genes of two diseases are given in Table 3. The Alzheimer's disease and lung neoplasms are not related in terms of MeSH's hierarchical structure. However, Table 3 demonstrates that 309 genes are related to both of these diseases, that is to say, the two diseases can be caused by some shared genes. Therefore, there are some commons in the pathogenesis of these two diseases. This illustrates the unreliability of using only disease semantic similarity.

Suggested rule 2. Disease semantic similarity should not be the only component used for disease similarity; disease-related genes can also be considered to measure the disease similarity.

- *Overestimation of the predictive accuracy*. Cross-validation has been widely used to measure the accuracy of predictive methods. The receiver-operating

TABLE 3
Illustration for Inadequacy of Disease Semantic Similarity

Disease name	MeSH tree number	Numbers of disease related genes	Numbers of shared genes
Alzheimer's disease	F03.087.400.100; C10.574.945.249; C10.228.140.380.100	1,064	309
lung neoplasms	C08.381.540; C08.785.520; C04.588.894.797.520	1,471	

characteristic (ROC) curve was drawn and AUC was calculated to evaluate the performance of predictive methods. One or more experimentally verified miRNA-disease associations are typically left out as the test dataset in the cross-validation procedure. However, we found that some existing computational methods do not correctly perform the cross-validation. The MISIM [39] is used in NetCBI [31], RWRMDA [28] and RLSMDA [33]. Our previous analysis has shown that the dataset MISIM is generated using HMDD and MeSH. The similarity between two miRNAs depends markedly on the set of known miRNA-disease associations. Therefore, it is not appropriate to employ the dataset MISIM directly in those methods. For example, assume that two miRNA-disease associations $\langle m_1, d\rangle$ and $\langle m_2, d\rangle$ were used to measure the functional similarity between m_1 and m_2. When leave-one-out cross-validation was performed in RWRMDA, the association $\langle m_1, d\rangle$ was treated as test data. However, the functional similarity between m_1 and m_2 should actually be measured again without considering the association $\langle m_1, d\rangle$, but the authors of RWRMDA did not take this into account. Therefore, the validation accuracy is overestimated by these methods. The overestimation problem caused by incorrect validation procedure in the method of lncRNA-disease association prediction was discussed [40].

To illustrate this problem more clearly, we further implement RWRMDA and RLSMDA based on their corresponding datasets, and perform a revised validation that recalculates the miRNA functional similarity network when related miRNA-disease associations are removed. The results are shown in Fig. 1. The solid line and the dashed line represent the original and revised validation results, respectively. RWRMDA and RLSMDA achieve relatively poor AUC values 0.7265 and 0.6964, respectively. The AUC values in the original literature on these methods are approximately 0.1 higher than these revised results.

Other methods for which the accuracy was overestimated are also emphasized here. For example, the validity of the AUC value of 0.8066 obtained for NetCBI [31] is also questionable as the miRNA-disease association was not removed when obtaining the miRNA functional similarity. Based on MISIM, HDMP [32] improved the miRNA functional similarity calculation method by incorporating information on disease terms and phenotype similarity between diseases, and by considering the miRNA family or cluster simultaneously. MIDP and MIDPE [35] recalculated the miRNA functional similarity using the latest released of the HMDD. These methods appeared to perform relatively well. However, the authors neglected to acknowledge that the miRNA-disease associations are also not removed in their corresponding miRNA functional similarity matrix. Therefore, we assert that the predictive accuracy of these methods has been overestimated.

Suggested rule 3. The miRNA similarity should not depend only on the experimentally verified miRNA-disease associations in the procedure of predicting such associations.

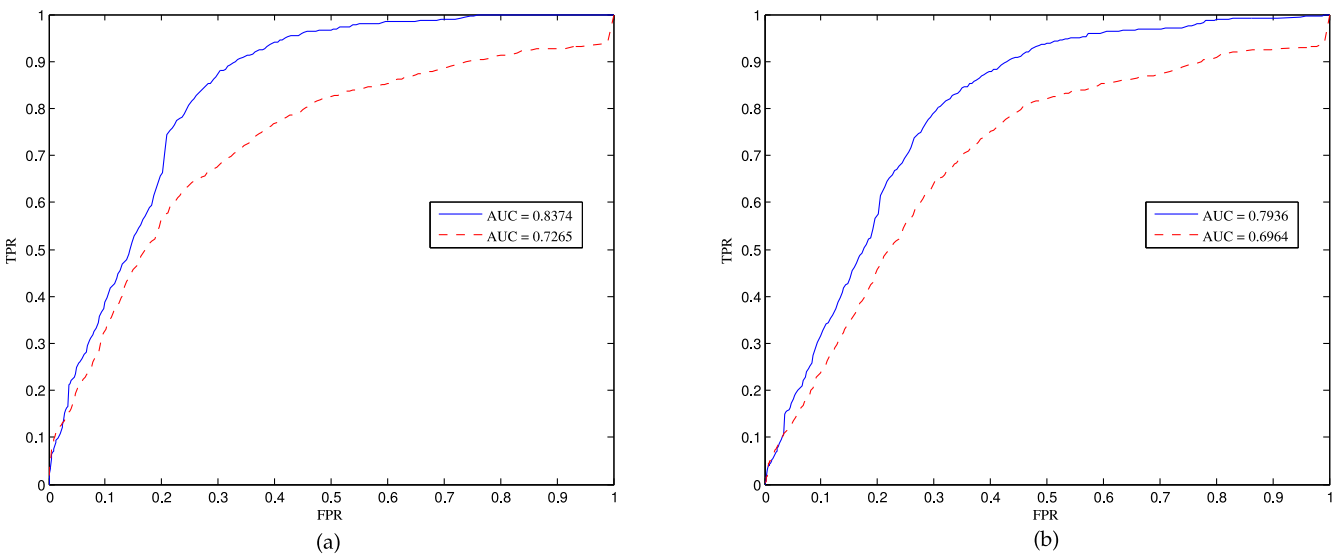

Fig. 1. Comparison between the original validation (solid line) and the revised validation (dashed line) in terms of ROC curve and AUC value based on leave-one-out cross-validation: (a) RWRMDA; (b) RLSMDA.

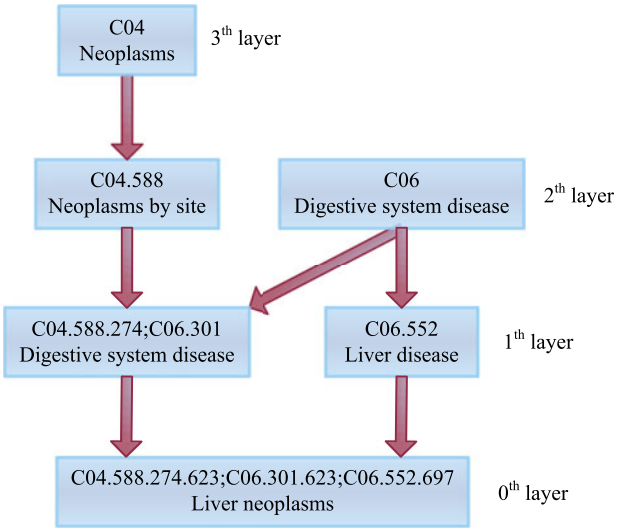

Fig. 2. The DAG of liver neoplasms.

3 Construction of a Heterogeneous Network

In this section, we first propose a novel method to measure the similarity between two diseases and then present a new method to calculate miRNA similarity. Finally, we provide a detailed description on how to construct of a heterogeneous network.

3.1 Disease Similarity Measurement

Disease similarity has been widely used in many fields, such as predicting disease-related genes [41], inferring similarity in the functions of miRNAs [39], identifying novel drug indications [42], and predicting miRNA-disease associations [32], [32], [33], [35]. Most methods of predicting miRNA-disease associations only consider the disease semantic similarity. As we argued in Suggested rule 2, disease semantic similarity should not be the only component of disease similarity.

Here, we incorporate the following two components into the calculation of disease similarity: disease semantic similarity based on MeSH and disease functional similarity referring to the genes. The disease structure of a directed acyclic graph (DAG) and disease-related genes are acquired from the MeSH descriptor and DisGeNET [43], respectively. Moreover, the probability of a functional linkage between two genes is measured using HumanNet [44]. The details are as follows:

- *Disease semantic similarity*: The disease semantic similarity is calculated according to Wang et al. [39]. Its calculation is based on the MeSH's hierarchical structure, which can be visualized as a DAG. An example of a DAG is shown in Fig. 2. Assuming that disease t is an ancestor of disease d or $t = d$, we give a recursive definition of the contribution of disease t to disease d as follows:

$$C_d(t) = \begin{cases} 1, & \text{if } t = d, \\ \max\{0.5 \cdot C_d(t') \mid t' \in \text{children of } t\}, & \text{if } t \neq d. \end{cases}$$

Finally, the semantic similarity between two disease d_1 and d_2 can be calculated by their shared ancestors, which is define as:

$$SS(d_1, d_2) = \frac{\sum_{t \in \mathbb{T}(d_1) \cap \mathbb{T}(d_2)} \left(C_{d_1}(t) + C_{d_2}(t)\right)}{\sum_{t \in \mathbb{T}(d_1)} C_{d_1}(t) + \sum_{t \in \mathbb{T}(d_2)} C_{d_2}(t)},$$

where $\mathbb{T}(d)$ is a disease set including d and all ancestors of d.

- *Disease functional similarity*: The disease functional similarity is calculated on the disease related genes [45]. Gene function networks are widely used to infer disease-gene associations [46]. HumanNet [44] is a probabilistic functional gene network, in which each interaction has an associated log-likelihood score (LLS) that measures the probability of an interaction between two genes. We normalized the associated LLS as follows:

$$LLS_N(g_i, g_j) = \frac{LLS(g_i, g_j) - LLS_{min}}{LLS_{max} - LLS_{min}},$$

where g_i and g_j are two genes, $LLS(g_i, g_j)$ represents LLS between the two genes, $LLS_N(g_i, g_j)$ represents the normalized LLS, and LLS_{max} and LLS_{min} are the maximum and minimum LLS in HumanNet, respectively. The functional similarity score between gene g_i and gene g_j is:

$$FSG(g_i, g_j) = \begin{cases} 1, & g_i = g_j, \\ 0, & e(g_i, g_j) \notin \text{HumanNet}, \\ LLS_N(g_i, g_j), & e(g_i, g_j) \in \text{HumanNet}, \end{cases}$$

where $e(g_i, g_j)$ represents the interaction between g_i and g_j. The functional similarity between two diseases d_1 and d_2 can be defined as follows:

$$FS(d_1, d_2) = \frac{\sum_{g \in \mathbb{G}(d_1)} FG(g, \mathbb{G}(d_2)) + \sum_{g \in \mathbb{G}(d_2)} FG(g, \mathbb{G}(d_1))}{|\mathbb{G}(d_1)| + |\mathbb{G}(d_2)|},$$

where $\mathbb{G}(d_1)$ and $\mathbb{G}(d_2)$ represent the gene sets related to diseases d_1 and d_2, respectively, $|\mathbb{G}|$ is the cardinality of gene set $\mathbb{G}$, and $FG(g, \mathbb{G}) = \max_{g_i \in \mathbb{G}}\{FSG(g, g_i)\}$.

Finally, the similarity between diseases d_1 and d_2 is calculated as follows:

$$DS(d_1, d_2) = \nu \cdot SS(d_1, d_2) + (1 - \nu) \cdot FS(d_1, d_2),$$

where $\nu \in [0, 1]$ is the weight coefficient of disease semantic similarity. If ν is 0.5, disease semantic similarity and disease functional similarity are equally weighted. If ν is above 0.5, disease semantic similarity is more important than disease functional similarity. In our experiments, we treat the two parts as being equally weighted, namely, $\nu = 0.5$.

3.2 miRNA Similarity Measurement

It has been reported that miRNAs with similar functions are often implicated in similar diseases [47]. The miRNA similarity networks have been widely used to infer miRNA-disease associations [28], [31], [32], [35]. As mentioned in

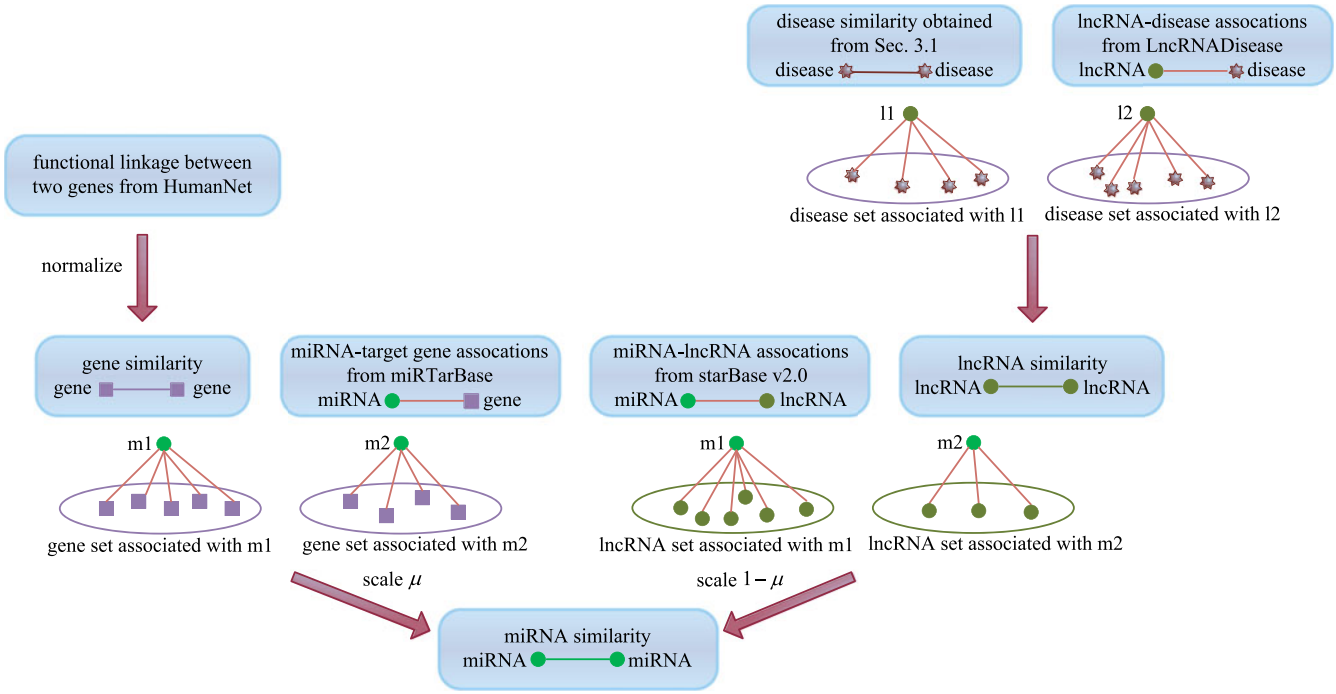

Fig. 3. The flowchart of miRNA similarity measurement.

Section 2, the usage of MISIM is associated with two problems, namely, the use of only a single dataset and the overestimation of the accuracy of prediction. In line with Suggested rule 3, we developed a novel method to quantify the association between two miRNAs, and which is completely independent of miRNA-disease associations. Instead, in our method, the miRNA-target gene association, the miRNA-lncRNA association and the lncRNA-disease association are used to determine the miRNA similarity. These three datasets are obtained from miRTarBase [48], starBase v2.0 [49] and LncRNADisease [50], respectively.

A flowchart of the miRNA similarity measurement is shown in Fig. 3. The similarity between miRNAs m_1 and m_2 is calculated as follows:

$$MS(m_1, m_2) = \mu \cdot GS(m_1, m_2) + (1 - \mu) \cdot LS(m_1, m_2),$$

where $GS(m_1, m_2)$ and $LS(m_1, m_2)$ represent the similarity contribution refer to target genes and related lncRNAs, respectively, and $\mu \in [0, 1]$ is the weight coefficient of the similarity contribution. We give equal importance of the two parts, that is, $\mu = 0.5$.

The first part, $GS(m_1, m_2)$, refers to genes functions, which is similar to the procedure of disease functional similarity measurement. Let $\mathbb{G}(m_1)$ and $\mathbb{G}(m_2)$ represent the target gene sets of miRNAs m_1 and m_2, respectively. The value of $GS(m_1, m_2)$ is given as follows:

$$GS(m_1, m_2) = \frac{\sum_{g \in \mathbb{G}(m_1)} FG(g, \mathbb{G}(m_2)) + \sum_{g \in \mathbb{G}(m_2)} FG(g, \mathbb{G}(m_1))}{|\mathbb{G}(m_1)| + |\mathbb{G}(m_2)|}.$$

For the other part, $LS(m_1, m_2)$, the similarity of the miRNA-associated lncRNA groups is considered. It has been proven that some lncRNAs act as competing endogenous RNAs in the regulation of gene expression [52]. The functional interactions between miRNAs and lncRNAs [53] drive us to measure the miRNA similarity based on miRNA-lncRNA associations. Therefore, the lncRNA similarity should be measured first. For two lncRNAs l_1 and l_2, the lncRNA similarity is calculated by the following formula:

$$FLS(l_1, l_2) = \frac{\sum_{d \in \mathbb{D}(l_1)} MD(d, \mathbb{D}(l_2)) + \sum_{d \in \mathbb{D}(l_2)} MD(d, \mathbb{D}(l_1))}{|\mathbb{D}(l_1)| + |\mathbb{D}(l_2)|},$$

where $\mathbb{D}(l_1)$ and $\mathbb{D}(l_2)$ represent the disease sets related to lncRNA l_1 and l_2, respectively, and $MD(d, \mathbb{D}) = \max_{d_i \in \mathbb{D}} \{DS(d, d_i)\}$. Subsequently, all of the known lncRNAs associated with miRNAs m_1 and m_2 are obtained, which are denoted by $\mathbb{L}(m_1)$ and $\mathbb{L}(m_2)$, and we formulate $LS(m_1, m_2)$ as follows:

$$LS(m_1, m_2) = \frac{\sum_{l \in \mathbb{L}(m_1)} FL(l, \mathbb{L}(m_2)) + \sum_{l \in \mathbb{L}(m_2)} FL(l, \mathbb{L}(m_1))}{|\mathbb{L}(m_1)| + |\mathbb{L}(m_2)|},$$

where $FL(l, \mathbb{L}) = \max_{l_i \in \mathbb{L}} \{FLS(l, l_i)\}$.

3.3 Heterogeneous Network

We connect the above two subnetworks, the disease similarity network and miRNA similarity network, using an experimentally verified miRNA-disease interaction network into a heterogeneous network of miRNAs and diseases. The experimentally verified associations between miRNAs and diseases are downloaded from the HMDD v2.0 [51]. We first filter out those miRNA-disease associations that are invalid because of an incorrect disease name or miRNA name, and then combine the different miRNA transcripts that produce the same mature miRNA into a single entity, as in previous work [32], [35], [39]. The correct disease name and miRNA name are obtained from the National Library of Medicine and the miRBase database [7], respectively. Finally, there are $5,298$ associations among 476 miRNAs and 341 diseases can be established in the heterogeneous network. We list

TABLE 4
The Details of Datasets Used in the Heterogeneous Network

Name	Website	Reference	Description
HMDD	http://www.cuilab.cn/hmdd	[51]	human miRNA-disease database
MeSH	http://www.ncbi.nlm.nih.gov/mesh		Medical Subject Headings
DisGeNET	http://www.disgenet.org/	[43]	a database of gene-disease associations
miRTarBase	http://mirtarbase.mbc.nctu.edu.tw/	[48]	experimentally validated microRNA-target gene interactions
starBase v2.0	http://starbase.sysu.edu.cn/mirLncRNA.php	[49]	miRNA-lncRNA interactions
LncRNADisease	http://www.cuilab.cn/lncrnadisease	[50]	the experimentally supported lncRNA-disease association
HumanNet	http://www.functionalnet.org/humannet/	[44]	a probabilistic functional gene network

the datasets used to construct the heterogeneous network in Table 4.

Suppose that $\mathbf{D} = \{D(i,j)\}_{i=1,j=1}^{n,n}$ is an adjacency matrix of the disease similarity network, $\mathbf{M} = \{M(i,j)\}_{i=1,j=1}^{m,m}$ is an adjacency matrix of the miRNA similarity network and $\mathbf{B} = \{B(i,j)\}_{i=1,j=1}^{n,m}$ is the miRNA-disease interaction network, where n and m represent the numbers of disease and miRNA entities, respectively. A simple example of the heterogeneous network is provided in Fig. 4. The adjacency matrix of the heterogeneous network can be represented as follows:

$$\mathbf{H} = \begin{bmatrix} \mathbf{D} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{M} \end{bmatrix},$$

where $\mathbf{B}^T$ represents the transpose of matrix $\mathbf{B}$.

4 Random Walk with Restart on a Heterogeneous Network

The RWR provides an outstanding framework to evaluate the relevance score between two nodes in a weighted graph, and it has been successfully used in numerous contexts, such as for predicting disease-associated miRNAs [28], [35]

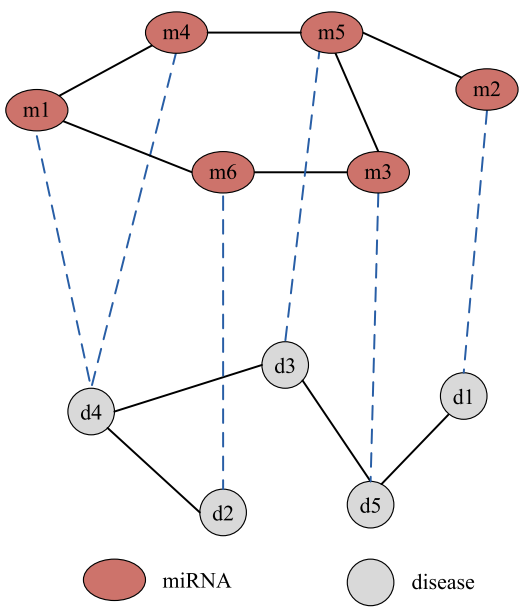

Fig. 4. Illustration of the heterogeneous network. The upper subnetwork is miRNA similarity network, and the lower subnetwork is disease similarity network. The two subnetworks are bridged by known miRNA-disease associations.

and prioritizing candidate disease genes [54]. RWR has been extended to heterogeneous networks for predicting drug-target interaction predictions [55], [56], identifying disease genes [57], [58], and prioritizing disease-related lncRNAs [59]. We implement the RWR on a heterogeneous network of miRNAs and diseases to infer potential miRNA-disease associations.

When a random walker walks in a heterogeneous network $\mathbf{H}$, the transition matrix should be decided. The transition matrix can be represented as follows:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{DD} & \mathbf{W}_{DM} \\ \mathbf{W}_{MD} & \mathbf{W}_{MM} \end{bmatrix},$$

where $\mathbf{W}_{DD} = \{W_{DD}(i,j)\}_{i=1,j=1}^{n,n}$ is the disease-to-disease transition matrix, $\mathbf{W}_{DM} = \{W_{DM}(i,j)\}_{i=1,j=1}^{n,m}$ is the transition matrix from the disease similarity network to the miRNA similarity network, $\mathbf{W}_{MD} = \{W_{MD}(i,j)\}_{i=1,j=1}^{m,n}$ is the transition matrix from the miRNA similarity network to the disease similarity network, and $\mathbf{W}_{MM} = \{W_{MM}(i,j)\}_{i=1,j=1}^{m,m}$ is the miRNA-to-miRNA transition matrix. Let λ be the jumping probability from the disease similarity network to the miRNA similarity network, and δ be the jumping probability from the miRNA similarity network to the disease similarity network. The intra-subnetwork transition probabilities from vertex d_i to d_j and vertex m_i to m_j are defined as:

$$W_{DD}(i,j) = \begin{cases} D(i,j) \Big/ \sum_{k=1}^{n} D(i,k), & \text{if } \sum_{k=1}^{m} B(i,k) = 0, \\ (1-\lambda) D(i,j) \Big/ \sum_{k=1}^{n} D(i,k), & \text{otherwise}, \end{cases}$$

and

$$W_{MM}(i,j) = \begin{cases} M(i,j) \Big/ \sum_{k=1}^{m} M(i,k), & \text{if } \sum_{k=1}^{n} B(k,i) = 0, \\ (1-\delta) M(i,j) \Big/ \sum_{k=1}^{m} M(i,k), & \text{otherwise}, \end{cases}$$

respectively. Furthermore, the inter-subnetwork transition probabilities from vertex d_i to m_j and m_i to d_j are defined as:

$$W_{DM}(i,j) = \begin{cases} 0, & \text{if } \sum_{k=1}^{m} B(i,k) = 0, \\ \lambda B(i,j) \Big/ \sum_{k=1}^{m} B(i,k), & \text{otherwise}, \end{cases}$$

and

$$W_{MD}(i,j) = \begin{cases} 0, & \text{if } \sum_{k=1}^{n} B(k,i) = 0, \\ \delta B(j,i) \Big/ \sum_{k=1}^{n} B(k,i), & \text{otherwise}, \end{cases}$$

respectively.

Let γ be the restart probability of the random walker, and $\mathbf{P}_t = \{p_t(i)\}_{i=1}^{m+n}$ be a probability vector in which $p_t(i)$ is the probability of the random walker at node i at step t. The RWR is implemented by the following iterative formula:

$$\mathbf{P}_{t+1} = (1-\gamma)\mathbf{W}^T\mathbf{P}_t + \gamma\mathbf{P}_0,$$

where $\mathbf{P}_0$ is the initial probability of the heterogeneous network, namely:

$$\mathbf{P}_0 = \begin{bmatrix} (1-\eta)\mathbf{u}_0 \\ \eta\mathbf{v}_0 \end{bmatrix}.$$

The parameter $\eta \in [0,1]$ balances the levels of importance of the disease similarity network and the miRNA similarity network. The two vectors $\mathbf{u}_0$ and $\mathbf{v}_0$ are the initial probabilities of the disease similarity network and the miRNA similarity network, respectively. If we want to predict potential miRNAs of a given disease d_i, d_i is the seed node in the disease similarity network, and the miRNAs that are known to be connected to d_i as our seed nodes in the miRNA similarity network. In the disease similarity network, $u_0(i) = 1$ and probability 0 is given to other nodes. The initial probability of the miRNA similarity network $\mathbf{v}_0$ is formed such that equal probabilities are assigned to those seed nodes in the miRNA similarity network with a sum equal to 1.

After some iteration steps, $\mathbf{P}_t$ converges to a stable status that the L_1 norm between $\mathbf{P}_t$ and $\mathbf{P}_{t+1}$ is less than 10^{-10}, and we denote the stable probability as

$$\mathbf{P}_\infty = \begin{bmatrix} (1-\eta)\mathbf{u}_\infty \\ \eta\mathbf{v}_\infty \end{bmatrix}.$$

By this time, the miRNAs can be ranked according to $\mathbf{v}_\infty$. The miRNAs that have a high probability value are more likely to be associated with disease d_i.

5 Experiments and Results

In this section, first, the effect of parameter is analyzed, and then our method is compared with other existing methods under leave-one-out cross-validation and five-fold cross-validation. Finally, some other experiments are performed to validate the feasibility of our method.

5.1 Parameter Tuning

Four parameters λ, δ, η and γ, are explored in our algorithm. The parameters λ and δ are the jumping probabilities, which control the preference between the disease similarity network and the miRNA similarity network. An investigation of the effects of these two parameters showed that greater values of λ and δ result in a better performance. The parameter η controls the impact of the two subnetworks. If the value of η is below 0.5, the random walker prefers to walk in the disease similarity matrix. The parameter γ is the restart probability.

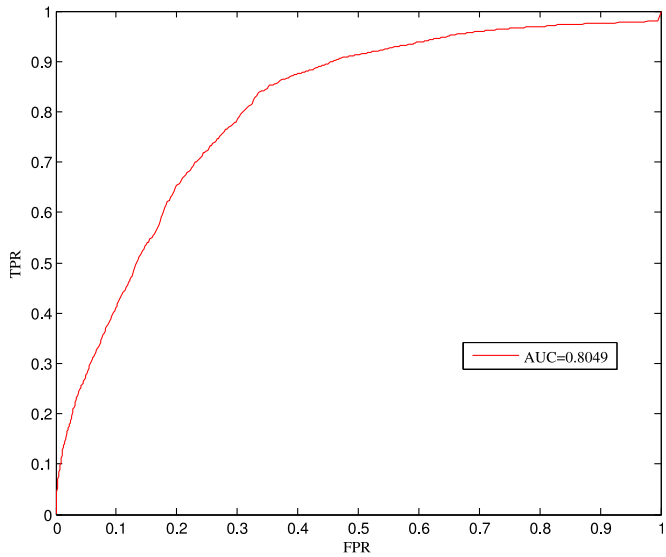

Fig. 5. The ROC curve and AUC value of our method.

Previous work [54] has shown that the restart probability has only a minor effect on the results. After investigation of these four parameters, the following values are used in our experiments: $\lambda = 0.8$, $\delta = 0.9$, $\eta = 0.1$ and $\gamma = 0.5$.

5.2 Leave-One-Out Cross-Validation

For a given disease d, each known d-related miRNA is left out in turn as the test miRNA and other known d-related miRNAs are taken as known information to predict potential candidates. The remaining miRNAs, which have no evidence to verify they are associated with the d, and the test miRNA constitute the d-related miRNA candidate pool. In the candidate pool, the test miRNA is regarded as a positive sample, and the others are negative samples. The probabilities associated with d of all miRNAs in the candidate pool are obtained by the predictive method. All miRNA candidates are ranked by their probabilities. The higher the ranking of the test miRNA, the better the predictive performance. If the ranking of the test miRNA exceeds a given threshold, it is regarded that the model successfully predicts the experimentally verified association. After each association has been tested, we calculated the true positive rate (TPR) and the false positive rate (FPR) for each threshold using the following formulas:

$$TPR = \frac{TP}{TP + FN},$$

and

$$FPR = \frac{FP}{TN + FP},$$

where TP and TN are the numbers of positive samples and negative samples that are identified under the given threshold, and FP and FN are the numbers of positive samples and negative samples that are identified incorrectly, respectively. By varying the threshold, the ROC curve is obtained, and the AUC is used to measure the predictive accuracy.

Fig. 5 shows the ROC curve and AUC value of our method. A comprehensive comparison of AUC values of different methods under leave-one-out cross-validation are shown in Table 5. The comparison results show that our method deals with more miRNAs and diseases than other methods and achieves a high AUC value.

TABLE 5
Comprehensive Comparison of Different Methods with Leave-One-Out Cross-Validation

Method	Our method	Jiang et al.'s method [27]	RWRMDA [28]	RLSMDA [33]	NetCBI [31]
Number of miRNAs	**476**	120	271	271	99
Number of diseases	**341**	53	137	137	51
AUC	**0.8049**	0.758	0.7265	0.6964	<0.8066

5.3 Five-Fold Cross-Validation

In the five-fold cross-validation, for a specific disease d, the known d-related miRNAs are randomly divided into five subsets. Each subset is left out as a test set, and other four subsets are used as the training set. All miRNAs in the test set are positive samples. After five rounds, we calculate the AUC value by the same as the method as mentioned for the leave-one-out cross-validation procedure. Because the majority of diseases are associated with only a few miRNAs, the performance five-fold cross-validation may not be sufficient for them. Therefore, we test 15 specific diseases that have been investigated in the literature [32], [35]. The results of our method and a comparison with HDMP and MIDP are shown in Table 6.

As shown in Table 6, the average AUC values of our method for all 15 diseases is 0.8476. The average AUC values of HMDP and MIDP are 0.8428 and 0.862. Our method's average AUC is slightly higher than HMDP. In contrast, and MIDP appears to achieve a higher average AUC value than our method. However, as shown in Section 2, the two methods HMDP and MIDP overestimate the predictive accuracy. In addition, for some diseases, such as lung neoplasms and squamous cell carcinoma, the results of our method are even higher than those overestimated AUC values. For the other diseases, the results of our method also approach the overestimated results obtained by MIDP.

5.4 Application to Diseases with no Known Related MiRNAs

It is very important that the predictive method has the ability to predict novel miRNAs for diseases with no known related miRNAs [33]. To estimate its performance in this regard, we implemented our method for the above mentioned 15 diseases. For a disease d, we remove all d-related miRNAs, and then our method is applied to predict d-related miRNAs. We count the numbers of confirmed d-related miRNAs under different ranking thresholds. The results are shown in Table 7. For breast neoplasms, we remove 197 known related miRNAs, and our method is implemented to uncover new related miRNAs. The result show that there are 90 confirmed associated miRNAs ranked in the predictive list of the top 100. This indicates that our method can make high-quality predictions of miRNA-disease associations for those diseases with no known related miRNAs.

TABLE 6
Comparison of our Method, HDMP, and MIDP Using Five-Fold Cross-Validation

Disease name	AUC		
	Our Method	HDMP [32]	MIDP [35]
Acute myeloid leukemia	0.8708	<0.822	<0.913
Breast neoplasms	0.8256	<0.819	<0.838
Colorectal neoplasms	0.8331	<0.785	<0.845
Glioblastoma	0.8386	<0.887	<0.786
Heart failure	0.8119	<0.797	<0.821
Hepatocellular carcinoma	0.8022	<0.785	<0.807
Lung neoplasms	**0.9249**	<0.899	<0.876
Melanoma	0.8337	<0.845	<0.837
Ovarian neoplasms	0.8958	<0.836	<0.923
Pancreatic neoplasms	0.9010	<0.922	<0.945
Prostatic neoplasms	0.8421	<0.884	<0.882
Renal cell carcinoma	0.8150	<0.828	<0.862
Squamous cell carcinoma	**0.8719**	<0.812	<0.870
Stomach neoplasms	0.7970	<0.866	<0.821
Urinary bladder neoplasms	0.8508	<0.885	<0.897

5.5 Case Studies: Lung Neoplasms and Breast Neoplasms

Case studies of two diseases, lung neoplasms and breast neoplasms, are analyzed for further evaluation of the ability of our method to predict potential miRNA-disease associations. The top 30 predictive results are then confirmed using the following three databases: dbDEMC [60], miRCancer [61] and miRdSNP [62]. The results are shown in Table 8.

For lung neoplasms, associations with the two miRNA candidates, hsa-mir-122 and hsa-mir-302a, do not appear in the above three databases. Fortunately, it has been found that oleanolic acid induces cell cycle arrest in lung cancer cells through the miR-122 pathway [63], and that miRNA hsa-mir-302a targets epidermal growth factor receptor in lung cancer [64]. Among top 30 candidates for breast

TABLE 7
The Numbers of the Confirmed MiRNA-Disease Associations Predicted by Our Method for 15 Diseases with No Known Related MiRNAs under Different Ranking Thresholds

Disease name	Number of known related miNRAs	Ranking threshold				
		20	40	60	80	100
Acute myeloid leukemia	62	15	26	30	38	42
Breast neoplasms	197	21	40	57	74	90
Colorectal neoplasms	143	17	35	47	60	74
Glioblastoma	94	14	28	34	46	51
Heart failure	120	13	28	38	50	57
Hepatocellular carcinoma	208	20	37	55	71	86
Lung neoplasms	128	17	37	53	71	84
Melanoma	136	20	36	49	62	70
Ovarian neoplasms	112	17	28	39	54	62
Pancreatic neoplasms	97	18	30	44	57	65
Prostatic neoplasms	116	18	29	40	52	61
Renal cell carcinoma	104	10	25	38	44	54
Squamous cell carcinoma	78	14	26	35	44	52
Stomach neoplasms	174	18	32	49	65	82
Urinary bladder neoplasms	90	17	30	40	46	55

TABLE 8
The Top 30 Lung Neoplasms and Breast Neoplasms Related Candidates

Lung neoplasms		
Ranking	miRNA name	Evidences
1	hsa-mir-16	dbDEMC, miRCancer
2	hsa-mir-15a	dbDEMC, miRCancer
3	hsa-mir-106b	dbDEMC, miRdSNP
4	hsa-mir-122	Unconfirmed
5	hsa-mir-195	dbDEMC, miRCancer
6	hsa-mir-15b	dbDEMC, miRCancer
7	hsa-mir-141	dbDEMC, miRCancer
8	hsa-mir-429	dbDEMC, miRCancer, miRdSNP
9	hsa-mir-20b	miRdSNP
10	hsa-mir-23b	miRdSNP
11	hsa-mir-302b	dbDEMC
12	hsa-mir-130a	dbDEMC
13	hsa-mir-204	miRCancer, miRdSNP
14	hsa-mir-193b	dbDEMC, miRCancer
15	hsa-mir-378a	miRdSNP
16	hsa-mir-373	miRCancer
17	hsa-mir-99a	dbDEMC, miRCancer
18	hsa-mir-342	dbDEMC, miRCancer
19	hsa-mir-196b	dbDEMC
20	hsa-mir-451a	dbDEMC
21	hsa-mir-302a	Unconfirmed
22	hsa-mir-320a	miRdSNP
23	hsa-mir-152	dbDEMC, miRCancer
24	hsa-mir-302c	dbDEMC
25	hsa-mir-10a	dbDEMC, miRCancer
26	hsa-mir-194	miRCancer
27	hsa-mir-149	dbDEMC, miRCancer, miRdSNP
28	hsa-mir-92b	miRCancer
29	hsa-mir-129	miRCancer
30	hsa-mir-296	dbDEMC
Breast neoplasms		
Ranking	miRNA name	Evidences
1	hsa-mir-150	dbDEMC, miRCancer, miRdSNP
2	hsa-mir-142	dbDEMC, miRdSNP
3	hsa-mir-15b	dbDEMC, miRdSNP
4	hsa-mir-106a	dbDEMC, miRdSNP
5	hsa-mir-192	dbDEMC, miRdSNP
6	hsa-mir-130a	dbDEMC, miRCancer, miRdSNP
7	hsa-mir-138	dbDEMC, miRdSNP
8	hsa-mir-378a	dbDEMC
9	hsa-mir-212	dbDEMC, miRdSNP
10	hsa-mir-99a	dbDEMC, miRCancer, miRdSNP
11	hsa-mir-196b	dbDEMC, miRCancer, miRdSNP
12	hsa-mir-372	dbDEMC, miRdSNP
13	hsa-mir-181c	dbDEMC, miRdSNP
14	hsa-mir-30e	dbDEMC, miRdSNP
15	hsa-mir-98	dbDEMC, miRCancer, miRdSNP
16	hsa-mir-92b	dbDEMC, miRdSNP
17	hsa-mir-144	dbDEMC, miRdSNP
18	hsa-mir-130b	dbDEMC, miRdSNP
19	hsa-mir-424	miRdSNP
20	hsa-mir-370	dbDEMC, miRdSNP
21	hsa-mir-181d	dbDEMC, miRdSNP
22	hsa-mir-28	Unconfirmed
23	hsa-mir-99b	miRCancer, miRdSNP
24	hsa-mir-449a	miRdSNP
25	hsa-mir-494	dbDEMC, miRCancer, miRdSNP
26	hsa-mir-134	dbDEMC
27	hsa-mir-362	miRCancer, miRdSNP
28	hsa-mir-449b	dbDEMC, miRdSNP
29	hsa-mir-32	dbDEMC, miRdSNP
30	hsa-mir-491	dbDEMC

neoplasms, only one miRNA, hsa-mir-28, for which there is no confirmation in the three databases. However, in 2011, Yang et al. identified erythroid 2-related factor 2 as a target of hsa-mir-28 in breast cancer [65], which suggests that the unconfirmed candidates such as this one have a high probability of being associated with lung cancer or breast cancer. We conclude that these case studies demonstrate that our method is powerful for predicting miRNA-disease associations with a high level of reliability.

5.6 Predicting Potential MiRNA-Disease Associations

After confirming the predictive accuracy of our method by cross-validation and case studies, all of the known miRNA-disease associations were used as training data to predict potential disease related miRNAs. For each of the 341 diseases studied, we have published the top 100 candidate miRNAs on our web server http://ifmda.aliapp.com. The potential miRNA-disease associations predicted in this manuscript are most likely real, which is supported by the case studies that demonstrated that 29 of the top 30 breast cancer-related miRNAs candidates are confirmed by at least one of the databases. The others should be validated by further biological experiments.

6 Conclusion

In this paper, we have highlighted three limitations associated with other computational methods, and presented three suggestions as design guidelines for method of predicting miRNA-disease associations. Two novel datasets were developed by integrating multiple types of information. The framework of RWR on a heterogeneous network of miRNAs and diseases was developed to predict miRNA candidates that could potentially be associated with diseases. The cross-validation and case studies demonstrated that our method has superior predictive performance. Moreover, an online service for prediction is also available at http://ifmda.aliapp.com.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Nos. 61370010, 61572094, 61202011) and the Natural Science Foundation of Fujian Province China (No. 2014J01253).

References

[1] R. C. Lee, R. L. Feinbaum, and V. Ambros, "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14," *Cell*, vol. 75, no. 5, pp. 843–854, 1993.

[2] C. Wang, L. Wei, M. Guo, and Q. Zou, "Computational approaches in detecting non-coding RNA," *Current Genomics*, vol. 14, no. 6, p. 371, 2013.

[3] L. Wei, M. Liao, Y. Gao, R. Ji, Z. He, and Q. Zou, "Improved and promising identification of human microRNAs by incorporating a high-quality negative set," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 11, no. 1, pp. 192–201, Jan./Feb. 2014.

[4] X. Zhang, D. Wu, L. Chen, X. Li, J. Yang, D. Fan, T. Dong, M. Liu, P. Tan, J. Xu, Y. Yi, Y. Wang, H. Zou, Y. Hu, K. Fan, J. Kang, Y. Huang, Z. Miao, M. Bi, N. Jin, K. Li, X. Li, J. Xu, and D. Wang, "RAID: A comprehensive resource for human RNA-associated (RNA–RNA/RNA–protein) interaction," *RNA*, vol. 20, no. 7, pp. 989–993, 2014.

[5] Y. Li, C. Wang, Z. Miao, X. Bi, D. Wu, N. Jin, L. Wang, H. Wu, K. Qian, C. Li, T. Zhang, C. Zhang, Y. Yi, H. Lai, Y. Hu, L. Cheng, K.-S. Leung, X. Li, F. Zhang, K. Li, X. Li, and D. Wang, "ViRBase: A resource for virus-host ncRNA-associated interactions," *Nucleic Acids Res.*, vol. 43, pp. D578–D582, 2014.

[6] S. F. Smith and K. C. Wiese, "Integrating thermodynamic and observed-frequency data for non-coding RNA gene search," *Trans. Comput. Syst. Biol. X.*, vol. 5410, pp. 124–142, 2008.

[7] A. Kozomara and S. Griffiths-Jones, "miRBase: Annotating high confidence microRNAs using deep sequencing data," *Nucleic Acids Res.*, vol. 42, pp. D68–D73, 2013.

[8] B. J. Reinhart, F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, and G. Ruvkun, "The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans," *Nature*, vol. 403, no. 6772, pp. 901–906, 2000.

[9] A. M. Cheng, M. W. Byrom, J. Shelton, and L. P. Ford, "Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis," *Nucleic Acids Res.*, vol. 33, no. 4, pp. 1290–1297, 2005.

[10] E. A. Miska, "How microRNAs control cell division, differentiation and death," *Current Opinion Genetics Develop.*, vol. 15, no. 5, pp. 563–568, 2005.

[11] P. Xu, M. Guo, and B. A. Hay, "MicroRNAs and the regulation of cell death," *TRENDS Genetics*, vol. 20, no. 12, pp. 617–624, 2004.

[12] D. Wu, Y. Huang, J. Kang, K. Li, X. Bi, T. Zhang, N. Jin, Y. Hu, P. Tan, L. Zhang, Y. Yi, W. Shen, J. Huang, X. Li, X. Li, J. Xu, and D. Wang, "ncRDeathDB: A comprehensive bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system," *Autophagy*, vol. 11, no. 10, pp. 1917–1926, 2015.

[13] Y. Li, L. Zhuang, Y. Wang, Y. Hu, Y. Wu, D. Wang, and J. Xu, "Connect the dots: a systems level approach for analyzing the miRNA-mediated cell death network," *Autophagy*, vol. 9, no. 3, pp. 436–439, 2013.

[14] M. Esteller, "Non-coding RNAs in human disease," *Nature Rev. Genetics*, vol. 12, no. 12, pp. 861–874, 2011.

[15] M. Mraz and S. Pospisilova, "MicroRNAs in chronic lymphocytic leukemia: From causality to associations and back," *Exp. Rev. Hematol*, vol. 5, no. 6, pp. 579–581, 2012.

[16] J. Das, S. Podder, and T. C. Ghosh, "Insights into the miRNA regulations in human disease genes," *BMC Genomics*, vol. 15, no. 1, p. 1010, 2014.

[17] J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz, and T. R. Golub, "MicroRNA expression profiles classify human cancers," *Nature*, vol. 435, no. 7043, pp. 834–838, 2005.

[18] F. Xin, M. Li, C. Balch, M. Thomson, M. Fan, Y. Liu, S. M. Hammond, S. Kim, and K. P. Nephew, "Computational analysis of microRNA profiles and their target genes suggests significant involvement in breast cancer antiestrogen resistance," *Bioinf.*, vol. 25, no. 4, pp. 430–434, 2009.

[19] S. Hua, W. Yun, Z. Zhiqiang, and Q. Zou, "A discussion of micrornas in cancers," *Current Bioinf.*, vol. 9, no. 5, pp. 453–462, 2014.

[20] Q. Wang, L. Wei, X. Guan, Y. Wu, Q. Zou, and Z. Ji, "Briefing in family characteristics of micrornas and their applications in cancer research," *Biochimica et Biophysica Acta (BBA)-Proteins Proteomics*, vol. 1844, no. 1, pp. 191–197, 2014.

[21] Y. Wang, L. Chen, B. Chen, X. Li, J. Kang, K. Fan, Y. Hu, J. Xu, L. Yi, J. Yang, Y. Huang, L. Cheng, Y. Li, C. Wang, K. Li, X. Li, J. Xu, and D. Wang, "Mammalian ncRNA-disease repository: A global view of ncRNA-mediated disease network," *Cell Death Dis*, vol. 4, p. e765, 2013.

[22] G. A. Calin and C. M. Croce, "MicroRNA signatures in human cancers," *Nature Rev. Cancer*, vol. 6, no. 11, pp. 857–866, 2006.

[23] R. F. Duisters, A. J. Tijsen, B. Schroen, J. J. Leenders, V. Lentink, I. van der Made, V. Herias, R. E. van Leeuwen, M. W. Schellings, P. Barenbrug, J. G. Maessen, S. Heymans, Y. M. Pinto, and E. E. Creemers, "miR-133 and miR-30 regulate connective tissue growth factor Implications for a role of microRNAs in myocardial matrix remodeling," *Circulation Res.*, vol. 104, no. 2, pp. 170–178, 2009.

[24] M. S. Weinberg and M. J. Wood, "Short non-coding RNA biology and neurodegenerative disorders: Novel disease targets and therapeutics," *Human Molecular Genetics*, vol. 18, no. R1, pp. R27–R39, 2009.

[25] P. Trang, J. Weidhaas, and F. Slack, "MicroRNAs as potential cancer therapeutics," *Oncogene*, vol. 27, pp. S52–S57, 2008.

[26] S. Nana-Sinkam and C. Croce, "Clinical applications for microRNAs in cancer," *Clinical Pharmacol. Therapeutics*, vol. 93, no. 1, pp. 98–104, 2013.

[27] Q. Jiang, Y. Hao, G. Wang, L. Juan, T. Zhang, M. Teng, Y. Liu, and Y. Wang, "Prioritization of disease microRNAs through a human phenome-microRNAome network," *BMC Syst. Biol.*, vol. 4, no. Suppl 1, p. S2, 2010.

[28] X. Chen, M.-X. Liu, and G.-Y. Yan, "RWRMDA: Predicting novel human microRNA–disease associations," *Molecular BioSyst.*, vol. 8, no. 10, pp. 2792–2798, 2012.

[29] H. Shi, J. Xu, G. Zhang, L. Xu, C. Li, L. Wang, Z. Zhao, W. Jiang, Z. Guo, and X. Li, "Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes," *BMC Syst. Biol.*, vol. 7, no. 1, p. 101, 2013.

[30] S. Mørk, S. Pletscher-Frankild, A. P. Caro, J. Gorodkin, and L. J. Jensen, "Protein-driven inference of miRNA-disease associations," *Bioinf.*, vol. 30, no. 3, pp. 392–397, 2013.

[31] H. Chen and Z. Zhang, "Similarity-based methods for potential human microRNA-disease association prediction," *BMC Med. Genomics*, vol. 6, no. 1, p. 12, 2013.

[32] P. Xuan, K. Han, M. Guo, Y. Guo, J. Li, J. Ding, Y. Liu, Q. Dai, J. Li, Z. Teng, and Y. Huang, "Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors," *PloS One*, vol. 8, no. 8, p. e70204, 2013.

[33] X. Chen and G.-Y. Yan, "Semi-supervised learning for potential human microRNA-disease associations inference," *Sci. Rep.*, vol. 4, 2014.

[34] C. Xu, Y. Ping, X. Li, H. Zhao, L. Wang, H. Fan, Y. Xiao, and X. Li, "Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles," *Mol. BioSyst.*, vol. 10, pp. 2800–2809, 2014.

[35] P. Xuan, K. Han, Y. Guo, J. Li, X. Li, Y. Zhong, Z. Zhang, and J. Ding, "Prediction of potential disease-associated microRNAs based on random walk," *Bioinf.*, vol. 31, pp. 1805-1815, 2015.

[36] Q. Zou, J. Li, Q. Hong, Z. Lin, Y. Wu, H. Shi, and Y. Ju, "Prediction of microrna-disease associations based on social network analysis methods," *BioMed Res. Int.*, vol. 2015, 2015.

[37] Q. Zou, J. Li, L. Song, X. Zeng, and G. Wang, "Similarity computation strategies in the microRNA-disease network: A survey," *Briefings Functional Genomics*, vol. 15, no. 1, pp. 55–64, 2015.

[38] X. Zeng, X. Zhang, and Q. Zou, "Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks," *Briefings Bioinf.*, vol. 17, pp. 193–203, 2015.

[39] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases," *Bioinf.*, vol. 26, no. 13, pp. 1644–1650, 2010.

[40] X. Chen and G.-Y. Yan, "Novel human lncrna-disease association inference based on lncrna expression profiles," *Bioinf.*, vol. 29, pp. 2617–2624, 2013.

[41] X. Wu, Q. Liu, and R. Jiang, "Align human interactome with phenome to identify causative genes and networks underlying disease families," *Bioinf.*, vol. 25, no. 1, pp. 98–104, 2009.

[42] A. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, "Predict: A method for inferring novel drug indications with application to personalized medicine," *Molecular Syst. Biol.*, vol. 7, no. 1, p. 496, 2011.

[43] J. Piñero, N. Queralt-Rosinach, À. Bravo, J. Deu-Pons, A. Bauer-Mehren, M. Baron, F. Sanz, and L. I. Furlong, "DisGeNET: A discovery platform for the dynamical exploration of human diseases and their genes," *Database*, vol. 2015, p. bav028, 2015.

[44] I. Lee, U. M. Blom, P. I. Wang, J. E. Shim, and E. M. Marcotte, "Prioritizing candidate disease genes by network-based boosting of genome-wide association data," *Genome Res.*, vol. 21, no. 7, pp. 1109–1121, 2011.

[45] L. Cheng, J. Li, P. Ju, J. Peng, and Y. Wang, "SemFunSim: A new method for measuring disease similarity by integrating semantic and gene functional association," *PloS One*, vol. 9, no. 6, p. e99415, 2014.

[46] B. Linghu, E. S. Snitkin, Z. Hu, Y. Xia, and C. DeLisi, "Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network," *Genome Biol.*, vol. 10, no. 9, p. R91, 2009.

[47] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, and Q. Cui, "An analysis of human microrna and disease associations," *PloS One*, vol. 3, no. 10, p. e3420, 2008.
[48] S.-D. Hsu, Y.-T. Tseng, S. Shrestha, Y.-L. Lin, A. Khaleel, C.-H. Chou, C.-F. Chu, H.-Y. Huang, C.-M. Lin, S.-Y. Ho, T. Y. Jian, F. M. Lin, T. H. Chang, S. L. Weng, K. W. Liao, I. E. Liao, C. C. Liu, and H. D. Huang, "miRTarBase update 2014: An information resource for experimentally validated miRNA-target interactions," *Nucleic Acids Res.*, vol. 42, no. D1, pp. D78–D85, 2014.
[49] J.-H. Li, S. Liu, H. Zhou, L.-H. Qu, and J.-H. Yang, "starBase v2. 0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data," *Nucleic Acids Res.*, vol. 42, no. D1, pp. D92–D97, 2014.
[50] G. Chen, Z. Wang, D. Wang, C. Qiu, M. Liu, X. Chen, Q. Zhang, G. Yan, and Q. Cui, "LncRNADisease: A database for long-noncoding RNA-associated diseases," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D983–D986, 2013.
[51] Y. Li, C. Qiu, J. Tu, B. Geng, J. Yang, T. Jiang, and Q. Cui, "HMDD v2. 0: A database for experimentally supported human microRNA and disease associations," *Nucleic Acids Res.*, vol. 42, no. D1, pp. D1070–D1074, 2014.
[52] T. Xia, Q. Liao, X. Jiang, Y. Shao, B. Xiao, Y. Xi, and J. Guo, "Long noncoding RNA associated-competing endogenous RNAs in gastric cancer," *Scientific Rep.*, vol. 4, p. 6088, 2014.
[53] J.-H. Yoon, K. Abdelmohsen, and M. Gorospe, "Functional interactions among microRNAs and long noncoding RNAs," *Seminars Cell Developmental Biol.*, vol. 34., pp. 9–14, 2014.
[54] S. Köhler, S. Bauer, D. Horn, and P. N. Robinson, "Walking the interactome for prioritization of candidate disease genes," *Am. J. Human Genetics*, vol. 82, no. 4, pp. 949–958, 2008.
[55] X. Chen, M.-X. Liu, and G.-Y. Yan, "Drug-target interaction prediction by random walk on the heterogeneous network," *Molecular BioSyst.*, vol. 8, no. 7, pp. 1970–1978, 2012.
[56] A. Seal, Y.-Y. Ahn, and D. J. Wild, "Optimizing drug–target interaction prediction based on random walk on heterogeneous networks," *J. Cheminformat.*, vol. 7, no. 1, pp. 1–12, 2015.
[57] Y. Li and J. C. Patra, "Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network," *Bioinf.*, vol. 26, no. 9, pp. 1219–1224, 2010.
[58] Y. Li and J. Li, "Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data," *BMC Genomics*, vol. 13, no. Suppl 7, p. S27, 2012.
[59] M. Zhou, X. Wang, J. Li, D. Hao, Z. Wang, H. Shi, L. Han, H. Zhou, and J. Sun, "Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network," *Molecular BioSyst.*, vol. 11, no. 3, pp. 760–769, 2015.
[60] Z. Yang, F. Ren, C. Liu, S. He, G. Sun, Q. Gao, L. Yao, Y. Zhang, R. Miao, Y. Cao, Y. Zhao, Y. Zhong, and H. Zhao, "dbDEMC: A database of differentially expressed miRNAs in human cancers," *BMC Genomics*, vol. 11, no. Suppl 4, p. S5, 2010.
[61] B. Xie, Q. Ding, H. Han, and D. Wu, "miRCancer: A microRNA-cancer association database constructed by text mining on literature," *Bioinf.*, vol. 29, pp. 638–644, 2013.
[62] A. E. Bruno, L. Li, J. L. Kalabus, Y. Pan, A. Yu, and Z. Hu, "miRdSNP: A database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes," *BMC Genomics*, vol. 13, no. 1, p. 44, 2012.
[63] X. Zhao, M. Liu, and D. Li, "Oleanolic acid suppresses the proliferation of lung carcinoma cells by mir-122/cyclin g1/mef2d axis," *Molecular Cellular Biochemistry*, vol. 400, no. 1–2, pp. 1–7, 2015.
[64] L. WC Chan, F. F Wang, and W. CS Cho, "Genomic sequence analysis of EGFR regulation by microRNAs in lung cancer," *Current Topics Medicinal Chemistry*, vol. 12, no. 8, pp. 920–926, 2012.
[65] M. Yang, Y. Yao, G. Eades, Y. Zhang, and Q. Zhou, "MiR-28 regulates Nrf2 expression through a Keap1-independent mechanism," *Breast Cancer Res. Treatment*, vol. 129, no. 3, pp. 983–991, 2011.

Yuansheng Liu received the BE and MSc degrees in computer science from the Xiangtan University of China in 2012 and 2015, respectively. He is currently working toward his PhD in the Advanced Analytics Institute at University of Technology Sydney. His research interests include bioinformatics and multimedia security.

Xiangxiang Zeng received the BS degree in automation from Hunan University, China, in 2005 and the PhD degree in system engineering from the Huazhong University of Science and Technology, China, in 2011. From 2010 to 2011, he spent one year working in the group of natural computing at Seville University, Spain. Currently, he is an associate professor in the Department of Computer Science, Xiamen University. His main research interests include natural computing, DNA computing, neural computing, and automaton theory. He is a member of the IEEE.

Zengyou He received the BS, MS, and PhD degrees in computer science from the Harbin Institute of Technology, China, in 2000, 2002, and 2006, respectively. He was a research associate in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology from February 2007 to February 2010. Since March 2010, he has been an associate professor in the School of Software at the Dalian University of Technology. His research interests include data mining and computational mass spectrometry.

Quan Zou received the PhD degree from the Harbin Institute of Technology, P.R. China, in 2009. He is a professor of computer science at Tianjin University. From 2009 to 2015, he was an assistant and associate professor at Xiamen University, P.R. China. His research interests include bioinformatics, machine learning, and parallel computing. Several related works have been published in *Briefings in Bioinformatics*, *Bioinformatics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, etc. Google scholar showed that his more than 100 papers have been cited more than 1,300 times. He is also a reviewer for many impacted journals, including *Bioinformatics*, *Briefings in Bioinformatics*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *BMC Bioinformatics*, etc. In February 2005, he received the Outstanding Reviewers for Computers in Biology and Medicine award. He is a member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.