



SNFIMCMDA: Similarity Network Fusion and Inductive Matrix Completion for miRNA–Disease Association Prediction

Lei Li¹, Zhen Gao¹, Chun-Hou Zheng^{1,2}, Yu Wang¹, Yu-Tian Wang^{1*} and Jian-Cheng Ni^{1*}

¹ School of Software, Qufu Normal University, Qufu, China, ² School of Computer Science and Technology, Anhui University, Hefei, China

OPEN ACCESS

Edited by:

Liang Cheng,
Harbin Medical University, China

Reviewed by:

Xiujuan Lei,
Shaanxi Normal University, China
Zhu-Hong You,
Xinjiang Technical Institute of Physics
& Chemistry, Chinese Academy
of Sciences (CAS), China

*Correspondence:

Yu-Tian Wang
wytfuture@163.com
Jian-Cheng Ni
nijch@163.com

Specialty section:

This article was submitted to
Molecular Medicine,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 15 October 2020

Accepted: 05 January 2021

Published: 09 February 2021

Citation:

Li L, Gao Z, Zheng C-H, Wang Y,
Wang Y-T and Ni J-C (2021)
SNFIMCMDA: Similarity Network
Fusion and Inductive Matrix
Completion for miRNA–Disease
Association Prediction.
Front. Cell Dev. Biol. 9:617569.
doi: 10.3389/fcell.2021.617569

MicroRNAs (miRNAs) that belong to non-coding RNAs are verified to be closely associated with several complicated biological processes and human diseases. In this study, we proposed a novel model that was Similarity Network Fusion and Inductive Matrix Completion for miRNA–Disease Association Prediction (SNFIMCMDA). We applied inductive matrix completion (IMC) method to acquire possible associations between miRNAs and diseases, which also could obtain corresponding correlation scores. IMC was performed based on the verified connections of miRNA–disease, miRNA similarity, and disease similarity. In addition, miRNA similarity and disease similarity were calculated by similarity network fusion, which could masterly integrate multiple data types to obtain target data. We integrated miRNA functional similarity and Gaussian interaction profile kernel similarity by similarity network fusion to obtain miRNA similarity. Similarly, disease similarity was integrated in this way. To indicate the utility and effectiveness of SNFIMCMDA, we both applied global leave-one-out cross-validation and five-fold cross-validation to validate our model. Furthermore, case studies on three significant human diseases were also implemented to prove the effectiveness of SNFIMCMDA. The results demonstrated that SNFIMCMDA was effective for prediction of possible associations of miRNA–disease.

Keywords: miRNA, disease, miRNA–disease association, similarity network fusion, inductive matrix completion

INTRODUCTION

MicroRNAs (miRNAs) belong to small non-coding RNAs, which effectively control the expression of their mRNA targets through RNA cleavage or translation repression (Ambros, 2004; Bartel, 2004; Ambros, 2001). In recent years, researchers have discovered various of miRNAs in many living organisms (Bruce et al., 1993; Calin and Croce, 2006). The expression of a great quantity of target genes is controlled by miRNAs, with the result that the whole miRNA pathway is an important technique for gene expression control (Xu et al., 2004; Miska, 2005; Bartel, 2009). The dysregulation of miRNAs results in progression of various diseases and conduces to developmental defects (Meola et al., 2009). Hence, identifying miRNAs that are associated with diseases is helpful in understanding the consequences of complex diseases and genetic causes. During the past few years, traditional experiments have confirmed a large number of connections of miRNA–disease

(Thomson et al., 2007; Mohammadi-Yeganeh et al., 2013). Previous experimental methods such as polymerase chain reaction can reveal the relationship between miRNA and disease, but which are time-consuming and costly. Thus, revealing the more unknown relationship between miRNAs and diseases need effective experiment methods. Researchers have made every effort to achieve effective and accurate prediction methods so that future biological experiments will reliably obtain more and more reasonable and valid relationship of miRNA–disease (Han et al., 2014).

In the past period of time, a great deal of computation-based algorithms and methods were developed to predict possible relationship between miRNAs and diseases (You et al., 2017; Chen et al., 2018a). Based on an assumption that miRNAs with similar functions are highly likely to be related to diseases that were phenotypic similar and *vice versa* (Zeng et al., 2015), Jiang et al. (2010) established a novel model that identified the feasible connections of miRNA–disease by using hypergeometric distribution. However, the model had the disadvantage that it only used local similarities between two miRNAs with a large number of shared target genes. In addition, Mørk et al. (2014) constructed an miPRD model to infer the miRNA–protein connections and disease–protein connections. Then, these connections were exploited to predict the possible relationship between miRNAs and diseases. The Jaccard similarity was first introduced by Chen et al. (2018) in the model of BLHARMDA to recognize possible miRNA–disease connections. The model of BLHARMDA also introduced the system of KNN into the bipartite local model method.

Obviously, the authenticity of global network similarity measures is superior to that of local network similarity measures (Köhler et al., 2008; Zhang et al., 2014). Considering this fact, Chen et al. (2012) constructed the novel RWRMDA model to infer unknown connections of miRNA–disease. Compared to local network similarity measures, RWRMDA discovered that global network similarity was more valid to find the potential relationship between miRNAs and diseases. Therefore, the performance of previous local network-based methods was worse than RWRMDA model. However, the RWRMDA model was unsuitable for new diseases that did not associate with miRNAs. Random walk method had been proposed by many researchers so as to effectively solve this problem. Liu et al. (2016) put forward an unused method that implemented the random walk algorithm to seek miRNAs associated with diseases. The method was to construct a heterogeneous graph by integrating various similarities of diseases and miRNAs. Then, the random walk with restart method in the heterogeneous graph is applied to seek unknown connections between miRNAs and diseases. Luo and Xiao (2017) established a heterogeneous network, which was made up of the similarity of miRNA, disease semantic similarity, and verified connections of miRNA–disease. Different from the method of Liu et al. (2016), they applied the imbalanced bi-random walk method to look for diseases that related to miRNAs. Furthermore, Chen et al. (2016b) presented the WBSMDA, which integrated the various of similarities of miRNA and disease, respectively. This model also could reliably obtain the possible relationship of miRNA–disease. Another

model HGIMDA was also presented by Chen et al. (2016a). The heterogeneous graph was generated by combining the verified miRNA–disease association network and the processed similarity networks of miRNA and disease in HGIMDA. It was important that an iterative equation was used in the model for the accurate prediction of potential miRNA–disease association. The model of HGIMDA performed better than previous methods, but the problem was the choice of parameters that was still not well resolved. For the purpose of inferring feasible and reasonable relationship of miRNA–disease, an identification medium was proposed by Yu et al. (2017). The medium changed the methods of maximizing information flow in existence, which consisted of functional similarity of miRNA, semantic similarity, and phenotypic similarity of disease. The verified connections and unknown connections of miRNA–disease were all adopted into a phenome–miRNAome network in this method. The NCMCMDA (Chen et al., 2020) model integrated neighborhood constraint with matrix completion algorithm to change the recovery task into an optimization problem. This model applied the fast iterative shrinkage-thresholding algorithm to recover missing interactions between miRNAs and diseases.

Recently, considerable amount of models that based on machine learning was gradually applied to expose the potential relationship of miRNA–disease. Xu et al. (2014) introduced a new method that prioritized novel disease-related miRNAs based on the miRNA target-dysregulated network (MTDN). In this model, the SVM classifier was constructed to extract the feature of network topologic information, which could effectively identify positive associations from negative associations of miRNA–disease. However, because negative samples were hard to obtain, the sets of negative samples were usually obtained by removing the pairs of positive sample sets from all pairs of miRNA–disease. In addition, Chen and Yan (2014) constructed the novel model of RLSMDA to infer potential miRNAs that were related to diseases. The association scores of miRNA–disease were effectively calculated by the model of RLSMDA. Therefore, RLSMDA could provide prediction score to new disease. Different from MTDN, RLSMDA could avert using negative miRNAs diseases associations, which could improve experimental efficiency and get more accurate results. The RBMMMDA (Chen et al., 2015) method was developed according to the restricted Boltzmann machine. RBMMMDA used the two-layer undirected graph to obviously represent the relationship of miRNA–disease. The two-layer undirected graph contained visible layer and hidden layer. RBMMMDA could gain new connections of miRNA–disease with the corresponding scores. Furthermore, another model named RKNNMDA (Chen et al., 2017) started to apply KNN method to deal with miRNAs and diseases. The support vector machine ranking model was also implemented in this method to handle these KNNs obtained by KNN method. The last ranking result of feasible connections between miRNAs and diseases was obtained by the weighted voting in this model. The disadvantage of this model was that miRNAs might associate with more known diseases owing to the bias. The BHCMDA (Zhu et al., 2020) model utilized biased heat conduction (BHC) algorithm to predict unknown

connections between miRNAs and diseases through combining miRNA similarity matrix, disease similarity matrix, and miRNA–disease association matrix. The probabilistic matrix factorization (PMF) algorithm was used in IMIPMF (Ha et al., 2020) model to infer potential miRNA–disease interactions. The PMF was widely used in recommender systems, so it could effectively make use of all information to recommend miRNAs, which are strongly associated with the disease.

Because there were several limitations existing in previous models, we constructed a new model that was Similarity Network Fusion and Inductive Matrix Completion for miRNA–Disease Association Prediction (SNFIMCMDA). We used the method of similarity network fusion (SNF) to obtain similarity of miRNA, which was gained by integrating function similarity and Gaussian interaction profile (GIP) kernel similarity of miRNA. And we also used the same way to obtain the disease similarity, which was gained by integrating semantic similarity and GIP kernel similarity of disease. After collecting data and integrating similarity for miRNA and disease, we used inductive matrix completion (IMC) method to efficiently obtain possible connections of miRNA–disease. The global leave-one-out cross-validation and five-fold cross-validation were carried out to evaluate the effectiveness of SNFIMCMDA. Furthermore, colon neoplasms, lung neoplasms, and breast neoplasms were performed as case studies. As a consequence, the 44, 43, and 43 of the top 50 miRNAs inferred by SNFIMCMDA, which were validated to associate with these human diseases according to the HMDD v3.2 (Huang et al., 2019) database and dbDEMC v2.0 (Zhen et al., 2017) database, respectively. Experimental results showed that our model was effective and reliable for predicting possible relationship of miRNA–disease.

MATERIALS AND METHODS

Human miRNA–Disease Associations

In this article, we downloaded the verified association data of miRNA–disease from HMDD v2.0 database (Li et al., 2013). There are 5,430 experimentally verified links of miRNA–disease in the known association data. Furthermore, we defined an adjacency matrix $A \in R^{nd \times nm}$ to describe the verified connections of miRNA–disease. There is no doubt that nd is defined as the amount of diseases, and nm is defined as the amount of miRNAs. The element $A(i, j)$ is equal to 1 if disease d_i is validated to be related to miRNA m_j , and 0 otherwise.

miRNA Functional Similarity

If functions of two miRNAs are similar, they have a high probability of being related to diseases that are similar and *vice versa* (Cui, 2010; Goh et al., 2007). Obviously, the miRNA functional similarity is obtained by this assumption. miRNA functional similarity information that we obtained was downloaded from the website of <http://www.cuilab.cn/files/images/cuilab/misim.zip>. In addition, we indicated the matrix MF to stand for the miRNA functional similarity. The value of similarity between miRNA m_i and miRNA m_j is represented by element $MF(m_i, m_j)$.

Disease Semantic Similarity

The Directed Acyclic Network (DAG) based on the Mesh descriptor (Lipscomb, 2000) can be utilized to describe diseases. The DAG of disease D includes two parts: nodes and edges. The nodes in DAG represent not only the D itself but also ancestor nodes of D . The edges in DAG are applied to connect child nodes with their parent nodes directly. Then $DAG(D) = (D, T(D), E(D))$ is utilized in our article to intuitively represent the DAG of disease D , where $T(D)$ and $E(D)$ indicated the node set and edge set, respectively. The semantic score of disease D is calculated according to the following equation:

$$DV(D) = \sum_{d \in T(D)} D_D(d) \quad (1)$$

where the contribution score of disease d is obtained by the following formula:

$$\begin{cases} D_D(d) = 1 & \text{if } d = D \\ D_D(d) = \max \{ \Delta * D_D(d') \mid d' \in \text{children of } d \} & \text{if } d \neq D \end{cases} \quad (2)$$

here, the semantic contribution factor $\Delta = 0.5$ in our article based on previous literature (Xuan et al., 2013).

The equation to calculate semantic similarity score between disease d_i and disease d_j is as follows:

$$DS(d_i, d_j) = \frac{\sum_{t \in T(d_i) \cap T(d_j)} (D_{d_i}(t) + D_{d_j}(t))}{DV(d_i) + DV(d_j)} \quad (3)$$

Gaussian Interaction Profile Kernel Similarity

If functions of two miRNAs are similar, they are likely to relate to similar or same diseases and *vice versa* (Lu et al., 2008; Sanghamitra et al., 2010). Therefore, the miRNA similarity and disease similarity can use the GIP kernel similarity to represent (Chen et al., 2016; Cheng et al., 2017). First, after observing whether there is known association between disease d_i and each miRNA or not, the interaction profile of disease d_i was represented by vector $K(d_i)$. We used vector $K(m_i)$ to represent the interaction profile of miRNA m_i in a similar way. Then, the equations to calculate GIP kernel similarity of diseases and miRNAs are as follows:

$$GKD(d_i, d_j) = \exp(-\rho_d \|K(d_i) - K(d_j)\|^2) \quad (4)$$

$$GKM(m_i, m_j) = \exp(-\rho_m \|K(m_i) - K(m_j)\|^2) \quad (5)$$

where the GKD and GKM represent GIP kernel similarity of disease and miRNA, respectively. The ρ_d and ρ_m are utilized to regulate the bandwidths of kernel. ρ_d is calculated by normalizing the original bandwidth ρ'_d . The specific formula is described as follows:

$$\rho_d = \rho'_d / \left(\frac{1}{nd} \sum_{i=1}^{nd} \|K(d_i)\|^2 \right) \quad (6)$$

The ρ_m can be obtained in a similar way:

$$\rho_m = \rho'_m / \left(\frac{1}{nm} \sum_{i=1}^{nm} \|K(m_i)\|^2 \right) \quad (7)$$

Similarity Network Fusion to Integrate Similarity

The similarity between miRNAs is calculated by functional similarity and GIP kernel similarity of miRNA, respectively. Similarly, the similarity between diseases is calculated by semantic similarity and GIP kernel similarity of disease, respectively. In this section, we introduced SNF (Wang et al., 2014) method to obtain ultimate similarity networks of disease and miRNA. The SNF method integrated similarity for disease included the following main steps. First, normalized weight matrices of disease similarity networks can be obtained by the below formulas:

$$DSP(d_i, d_j) = \begin{cases} \frac{DS(d_i, d_j)}{2 \sum_{k \neq i} DS(d_i, d_k)} & j \neq i \\ \frac{1}{2} & j = i \end{cases} \quad (8)$$

$$KDP(d_i, d_j) = \begin{cases} \frac{GKD(d_i, d_j)}{2 \sum_{k \neq i} GKD(d_i, d_k)} & j \neq i \\ \frac{1}{2} & j = i \end{cases} \quad (9)$$

where DSP denotes the normalized weight matrix of disease semantic similarity network, and KDP denotes the normalized weight matrix of GIP kernel similarity for diseases. Then, we used KNN method to calculate disease local relationship by the following two formulas:

$$DSK(d_i, d_j) = \begin{cases} \frac{DS(d_i, d_j)}{\sum_{k \in N_i} DS(d_i, d_k)} & j \in N_i \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

$$KDK(d_i, d_j) = \begin{cases} \frac{GKD(d_i, d_j)}{\sum_{k \in N_i} GKD(d_i, d_k)} & j \in N_i \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where N_i denotes the number of neighbors of disease d_i . DSK denotes the local relationship matrix of disease semantic similarity; KDK represents the local relationship matrix of GIP kernel similarity for diseases. Based on the previous literature (Wang et al., 2014), the essence of SNF method could be described as an iterative update of similar matrices. In our article, after we brought disease data into the network fusion formula of SNE, the specific process of network fusion corresponded to each

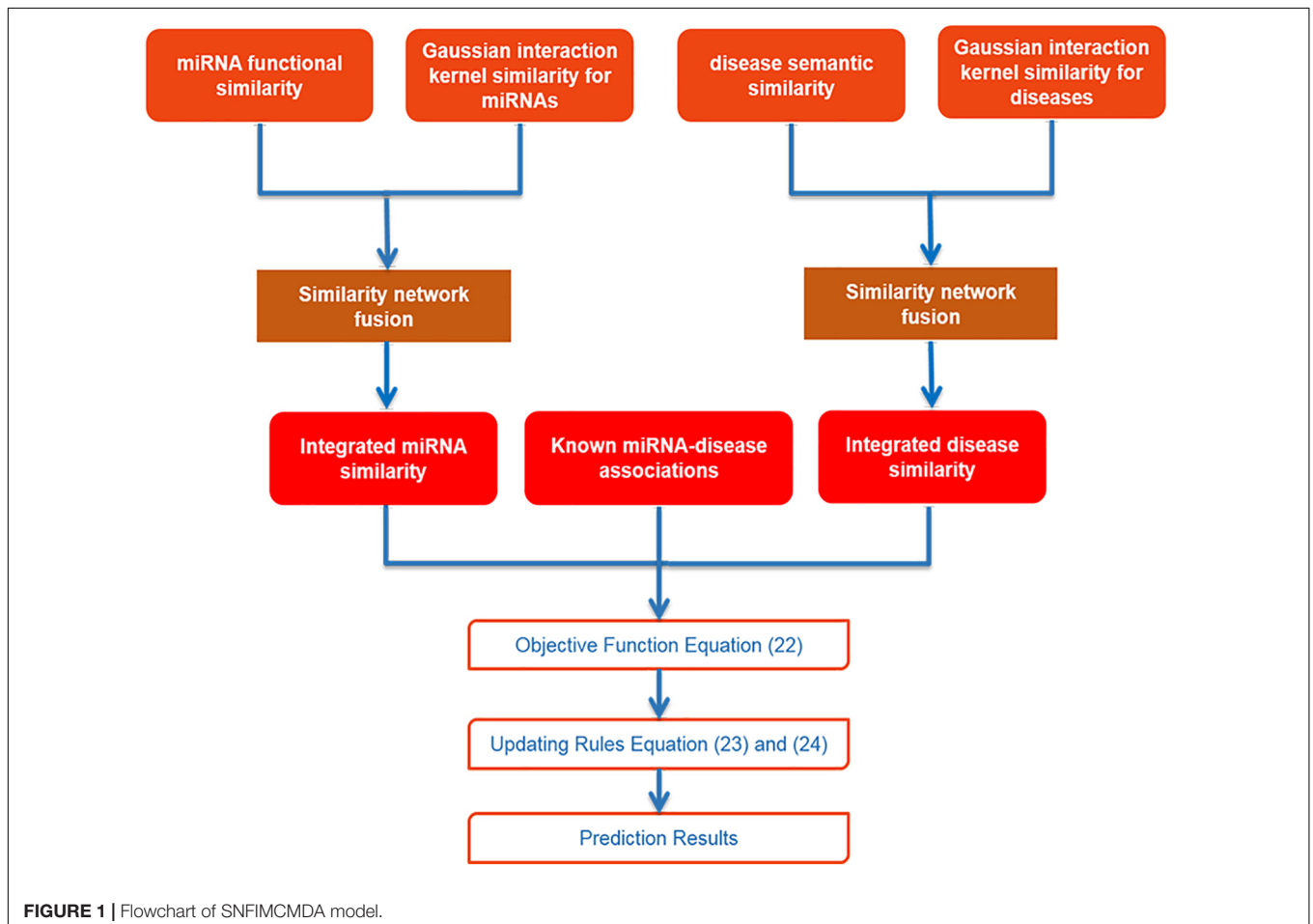


FIGURE 1 | Flowchart of SNFIMCMDA model.

data type is presented by the following equations:

$$DSP(d_i, d_j) = DSK(d_i, d_j) * KDP(d_i, d_j) * (DSK(d_i, d_j))^T \quad (12)$$

$$KDP(d_i, d_j) = KDK(d_i, d_j) * DSP(d_i, d_j) * (KDK(d_i, d_j))^T \quad (13)$$

The final similarity matrix of disease that integrated all data types is presented by the below formula:

$$S_d(d_i, d_j) = \frac{1}{2} (DSP(d_i, d_j) + KDP(d_i, d_j)) \quad (14)$$

where S_d denotes the final similarity matrix of disease.

Similarity network fusion for miRNA is defined in a similar way by the following formulas:

$$MFP(m_i, m_j) = \begin{cases} \frac{MF(m_i, m_j)}{2 \sum_{k \neq i} MF(m_i, m_k)} & j \neq i \\ \frac{1}{2} & j = i \end{cases} \quad (15)$$

$$KMP(m_i, m_j) = \begin{cases} \frac{GKM(m_i, m_j)}{2 \sum_{k \neq i} GKM(m_i, m_k)} & j \neq i \\ \frac{1}{2} & j = i \end{cases} \quad (16)$$

$$MFK(m_i, m_j) = \begin{cases} \frac{MF(m_i, m_j)}{\sum_{k \in N_i} MF(m_i, m_k)} & j \in N_i \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

$$MKM(m_i, m_j) = \begin{cases} \frac{GKM(m_i, m_j)}{\sum_{k \in N_i} GKM(m_i, m_k)} & j \in N_i \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

$$MFP(m_i, m_j) = MFK(m_i, m_j) * KMP(m_i, m_j) * (MFK(m_i, m_j))^T \quad (19)$$

$$KMP(m_i, m_j) = MKM(m_i, m_j) * MFP(m_i, m_j) * (MKM(m_i, m_j))^T \quad (20)$$

$$S_m(m_i, m_j) = \frac{1}{2} (MFP(m_i, m_j) + KMP(m_i, m_j)) \quad (21)$$

where S_m denotes the miRNA similarity matrix.

Inductive Matrix Completion

After collecting data and using SNF to integrate similarities for miRNA and disease, we utilized IMC method to obtain final prediction result. The specific flowchart of SNFIMCMDA is presented in **Figure 1**. The IMC method was employed according to the verified connection matrix of miRNA-disease $A \in R^{nd \times nm}$, miRNA similarity matrix $S_m \in R^{nm \times nm}$, and disease similarity matrix $S_d \in R^{nd \times nd}$. Here, the feature matrix of nm miRNAs was used $S_m \in R^{nm \times nm}$ to represent, and the feature matrix of nd diseases was used $S_d \in R^{nd \times nd}$ to represent. The feature vector of miRNA m_j was denoted by $S_m(j)$, and the feature vector of disease d_i was denoted by $S_d(i)$. Then we made $A = UV^T$, where $U \in R^{nd \times r}$ and $V \in R^{nm \times r}$. Here, the r is desired rank that also is the same as $\min(\text{rank}(U), \text{rank}(V))$. The convergence speed of the IMC algorithm is also affected by r . The matrices U and V

can be treated as the answers of the optimization problem as follows:

$$\min_{U, V} \phi = \frac{1}{2} \|A - S_d UV^T S_m^T\|_F^2 + \frac{\lambda_1}{2} \|U\|_F^2 + \frac{\lambda_2}{2} \|V\|_F^2 \quad (22)$$

such that, $U \geq 0, V \geq 0$

where $\|\cdot\|_F^2$ is Frobenius norm of matrix that is set to solve overfitting problems. λ_1 and λ_2 are equal to $1 \|\cdot\|_F$ that are regularization parameters.

In addition, $U \in R^{nd \times r}$ and $V \in R^{nm \times r}$ were two random dense matrices by the iterative equation to update. In our experiment, when the convergence criterion met 10^{-6} , U and V would be obtained by iterative process. The process of IMC algorithm to obtain U and V are presented by the following formulas:

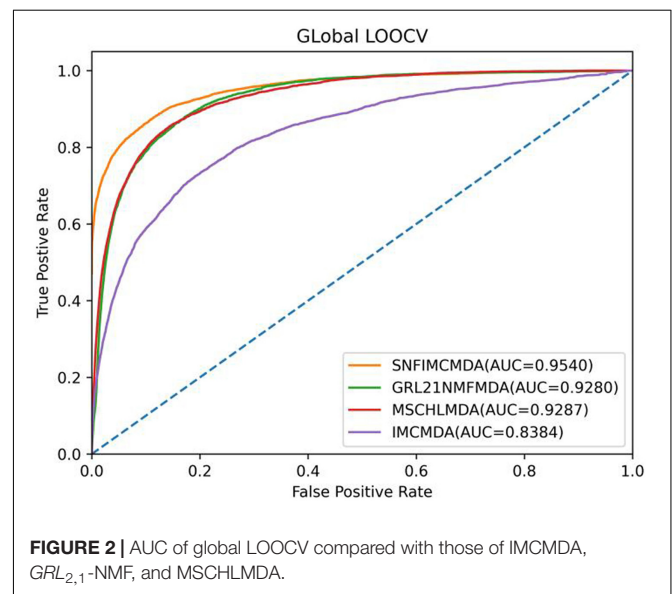
$$V_{jk} \leftarrow V_{jk} \frac{(S_m^T A^T S_d U)_{jk}}{(S_m^T S_m V U^T S_d^T S_d U + \lambda_2 V)_{jk}} \quad (23)$$

$$U_{ik} \leftarrow U_{ik} \frac{(S_d^T A S_m V)_{ik}}{(S_d^T S_d U V^T S_m^T S_m V + \lambda_1 U)_{ik}} \quad (24)$$

The $S(d_i, m_j)$ indicates the predicted association chance between d_i and m_j . $S(d_i, m_j)$ can be obtained by applying U and V to calculate:

$$S(d_i, m_j) = S_d(i) UV^T S_m(j) \quad (25)$$

Furthermore, if the feature vector of disease $newd_i$ is acquired, $S(newd_i, j)$ can be utilized to obtain association score between this disease and any miRNA. We will realize disease $newd_i$ associated with which miRNAs effectively.



RESULTS

Performance Evaluation

For the purpose of affirming the accuracy of predicted result of SNFIMCMDA, we compared our model with three previous computational models: IMCMDA (Chen et al., 2018b), *GRL_{2,1}-NMF* (Gao et al., 2020), and MSCHLMDA (Wu et al., 2020). Based on the verified connections of miRNA–disease that were downloaded from HMDD v2.0 database, global leave-one-out cross-validation (global LOOCV) and five-fold cross-validation (5-CV) were utilized to validate the actual performance of these computational models.

In the framework of global LOOCV, we applied the associations of miRNA–disease to train model. First, we selected each verified connection of miRNA–disease in turn for testing, whereas other experimentally confirmed associations were training sets. In addition to verified associations, there still were some connections between miRNAs and diseases without evidence that were treated as candidate samples. Then we calculated all association scores after implementing SNFIMCMDA, the test samples would obtain the predicting rankings by comparing with the candidate samples. If a given threshold was inferior to the ranking of each test sample, we thought SNFIMCMDA was valid. Furthermore, we could draw receiver operating characteristic (ROC) curve by plotting the true-positive rate against the false-positive rate. Finally, for the purpose of evaluating performance of SNFIMCMDA, we calculated the areas under ROC curve (AUCs) of all models. The ultimate result clearly indicated that the AUC values of SNFIMCMDA, IMCMDA, *GRL_{2,1}-NMF*,

and MSCHLMDA reached 0.9540, 0.8384, 0.9280, and 0.9287, respectively (Figure 2). Obviously, the AUC of SNFIMCMDA was higher than other methods.

In the framework of 5-CV, first, all observed connections of miRNA–disease were randomly divided into five parts; where the test set was held by each one of the five parts for each round, the training set consisted of the other four parts in turn. In addition to observed connections, there still were several connections without evidence that were treated as candidate samples. After implementing the SNFIMCMDA, we could obtain the predicted rankings of test samples compared with those of the candidate samples. Furthermore, we performed 100 times repeated segmentations on known connections, so as to avoid the possible deviations generated in the process of random sample segmentation. Finally, similar to global LOOCV, we could obtain ROC curve and AUCs of these models. The specific result indicated that the AUC values of SNFIMCMDA, IMCMDA, *GRL_{2,1}-NMF*, and MSCHLMDA were 0.9539, 0.8330, 0.9276, and 0.9263, respectively (Figure 3). Obviously, the AUC of SNFIMCMDA was also higher than other methods.

Case Study

In this article, several types of human diseases that included colon neoplasms, breast neoplasms, and lung neoplasms were applied to validate the prediction result of SNFIMCMDA. These diseases actually pose a great threat to human beings. Colon neoplasms belong to the common malignant tumor in the gastrointestinal tract (Jemal et al., 2011). There were a large amount of new cases and deaths that were caused by colon neoplasms in recent years (Thackeray et al., 2011). Several miRNAs that relate to

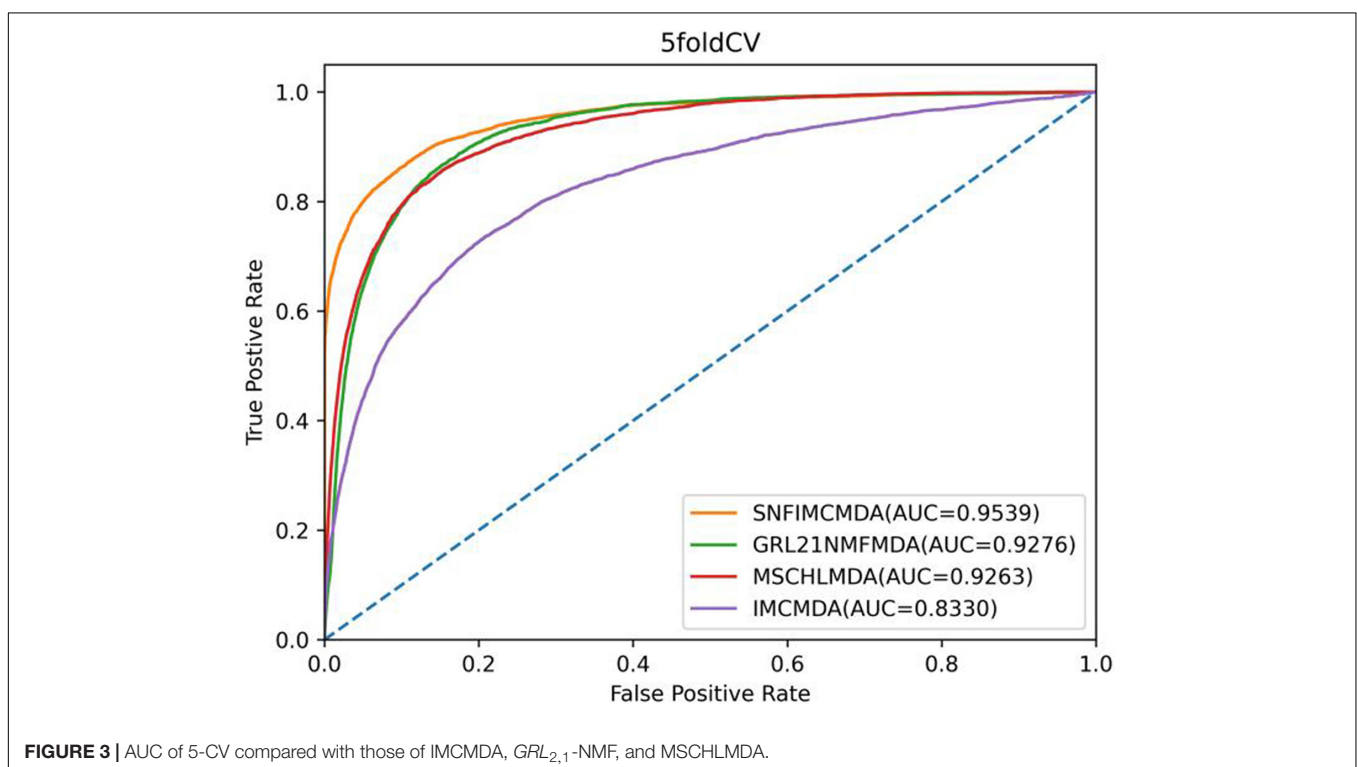


FIGURE 3 | AUC of 5-CV compared with those of IMCMDA, *GRL_{2,1}-NMF*, and MSCHLMDA.

TABLE 1 | The top 50 potential miRNAs associated with colon neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-34a	H; d	hsa-mir-133b	H; d
hsa-mir-29a	H; d	hsa-mir-206	d
hsa-mir-29b	H; d	hsa-mir-20a	H; d
hsa-mir-16	Unconfirmed	hsa-mir-30a	H; d
hsa-mir-125b	H; d	hsa-mir-31	H; d
hsa-mir-15a	H; d	hsa-mir-21	H; d
hsa-mir-221	H; d	hsa-mir-122	d
hsa-mir-181a	H; d	hsa-mir-155	H; d
hsa-mir-1	d	hsa-mir-15b	H; d
hsa-mir-223	H; d	hsa-mir-9	d
hsa-mir-29c	d	hsa-mir-7	d
hsa-mir-146a	H; d	hsa-mir-203	H; d
hsa-mir-150	H; d	hsa-mir-19b	d
hsa-mir-199a	Unconfirmed	hsa-mir-214	Unconfirmed
hsa-mir-24	H; d	hsa-mir-23a	H; d
hsa-mir-210	H; d	hsa-mir-148a	H; d
hsa-mir-181b	H; d	hsa-mir-34c	Unconfirmed
hsa-mir-222	H; d	hsa-mir-196a	H; d
hsa-mir-143	H; d	hsa-mir-132	H; d
hsa-mir-106b	H; d	hsa-mir-125a	H; d
hsa-mir-195	H; d	hsa-mir-192	H; d
hsa-mir-133a	H; d	hsa-let-7a	Unconfirmed
hsa-mir-146b	d	hsa-mir-200b	H; d
hsa-mir-142	H	hsa-mir-124	Unconfirmed
hsa-mir-92a	d	hsa-mir-18a	H; d

H: HMDD v3.2 database; d: dbDEMC v2.0 database.

TABLE 2 | The top 50 potential miRNAs associated with breast neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-106a	H; d	hsa-mir-363	H; d
hsa-mir-142	H	hsa-mir-196b	H; d
hsa-mir-192	H; d	hsa-mir-99b	d
hsa-mir-138	H; d	hsa-mir-98	H; d
hsa-mir-542	H	hsa-mir-552	d
hsa-mir-32	H; d	hsa-mir-186	d
hsa-mir-15b	H; d	hsa-mir-421	H; d
hsa-mir-449a	H; d	hsa-mir-144	H; d
hsa-mir-181d	d	hsa-mir-28	d
hsa-mir-150	H; d	hsa-mir-212	H; d
hsa-mir-92b	H; d	hsa-mir-130b	H; d
hsa-mir-498	d	hsa-mir-518b	Unconfirmed
hsa-mir-211	Unconfirmed	hsa-mir-494	H; d
hsa-mir-330	Unconfirmed	hsa-mir-410	H; d
hsa-mir-185	H; d	hsa-mir-512	Unconfirmed
hsa-mir-130a	H; d	hsa-mir-370	H; d
hsa-mir-518c	d	hsa-mir-181c	H; d
hsa-mir-622	d	hsa-mir-574	H; d
hsa-mir-615	Unconfirmed	hsa-mir-376a	H; d
hsa-mir-30e	H	hsa-mir-548d	H
hsa-mir-184	H; d	hsa-mir-539	d
hsa-mir-483	Unconfirmed	hsa-mir-372	H; d
hsa-mir-99a	H; d	hsa-mir-454	d
hsa-mir-502	H	hsa-mir-424	H; d
hsa-mir-630	H; d	hsa-mir-455	Unconfirmed

H: HMDD v3.2 database; d: dbDEMC v2.0 database.

colon neoplasms have been confirmed by a mass of biological experiments. Breast neoplasms can be regarded as a common disease in females, which has negative effects on the health of women (Koboldt et al., 2012; Liang et al., 2016). Based on clinical experiments and evidences, numerous miRNAs that are related to the breast neoplasms have been found by researchers (Fu et al., 2011; Zhu et al., 2014) in the past few years. Lung neoplasms are considered as the fastest-growing neoplasm in morbidity rate and mortality rate (Jemal et al., 2011; Carol et al., 2019). The HMDD v2.0 database was used as the training database; we applied verified associations of miRNA–disease to produce prediction results. The HMDD v3.2 database and dbDEMC v2.0 database served as validation databases utilized to validate prediction results. Furthermore, the candidate miRNAs of these diseases obtained from the SNFIMCMDA were ranked by prediction scores. Finally, we gained the top 50 miRNAs that connected to these human diseases and the 44, 43, and 43 of the top 50 miRNAs certified by HMDD v3.2 database and dbDEMC v2.0 database, respectively. The specific results are listed in **Tables 1–3**. In conclusion, we tested the predictive performance of SNFIMCMDA on the HMDD v2.0 database to observe whether the model had a well performance on it. As the validation results are shown in the tables, the effectiveness of SNFIMCMDA on predicting unknown interactions between miRNAs and diseases had been confirmed by the HMDD v3.2 database and dbDEMC v2.0 database.

TABLE 3 | The top 50 potential miRNAs associated with lung neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-99a	H; d	hsa-mir-424	d
hsa-mir-429	d	hsa-mir-299	H
hsa-mir-194	H; d	hsa-mir-23b	d
hsa-mir-296	Unconfirmed	hsa-mir-495	H
hsa-mir-16	H; d	hsa-mir-130b	H
hsa-mir-151a	Unconfirmed	hsa-mir-342	H; d
hsa-mir-196b	H; d	hsa-mir-141	H; d
hsa-mir-15b	d	hsa-mir-379	d
hsa-mir-10a	H; d	hsa-mir-324	H
hsa-mir-204	d	hsa-mir-663a	Unconfirmed
hsa-mir-708	d	hsa-mir-376c	d
hsa-mir-211	d	hsa-mir-520a	d
hsa-mir-122	H; d	hsa-mir-152	H; d
hsa-mir-15a	H; d	hsa-mir-608	H
hsa-mir-451a	H; d	hsa-mir-500a	Unconfirmed
hsa-mir-432	d	hsa-mir-190a	Unconfirmed
hsa-mir-217	H	hsa-mir-520d	d
hsa-mir-154	H	hsa-mir-215	H; d
hsa-mir-625	d	hsa-mir-144	H; d
hsa-mir-195	H; d	hsa-mir-184	H; d
hsa-mir-99b	d	hsa-mir-433	d
hsa-mir-28	d	hsa-mir-520c	Unconfirmed
hsa-mir-362	H	hsa-mir-367	d
hsa-mir-501	Unconfirmed	hsa-mir-520b	H; d
hsa-mir-423	H	hsa-mir-520e	H; d

H: HMDD v3.2 database; d: dbDEMC v2.0 database.

DISCUSSION

The researches for inferring possible relationship of miRNA–disease would provide deep insight into the pathogenesis of diseases and contribute to the treatment of diseases. Therefore, we constructed the novel model of SNFIMCMDA. The prediction score of each miRNA–disease pair was calculated by combining the known association between miRNAs and diseases and integrated similarities of both miRNA and disease in the SNFIMCMDA. Different from the model of IMCMDA that had been published in previous years, we made a change in integrating similarity for miRNAs and diseases. The method of SNF was used to integrate similarity in place of a previous method in IMCMDA. After adopting SNF, there was a significant improvement in the prediction results. In the framework of global LOOCV, the AUC of SNFIMCMDA was 0.9540, which was higher than 0.8330 calculated by IMCMDA. And in the framework of 5-CV, the AUC of SNFIMCMDA was 0.9539 that was also higher than 0.8330 obtained by IMCMDA. Moreover, the AUC of SNFIMCMDA performed better than other previous methods in both global LOOCV and 5-CV. Furthermore, three different human diseases were performed as case study that had effectively certified the reliable performance of the SNFIMCMDA. Therefore, SNFIMCMDA could be utilized as a reliable biological tool for extracting the most promising disease-related miRNAs, thereby enhancing our comprehension on the disease mechanisms of miRNAs and contributing to the prevention, discovery, and diagnosis of complex diseases in the future.

In our article, the model of SNFIMCMDA completed the missing association scores between miRNAs and diseases, which utilized the feature vector method to succinctly represent disease and miRNA, respectively. Furthermore, if we had the feature vector of the disease without any known associated miRNAs, the SNFIMCMDA could reliably predict this disease associated with which miRNAs and obtained the scores between them. In addition, our model belonged to semi-supervised model, so it had no use for negative samples. The obvious advantage of our model was that it only needs positive and unlabeled samples, which effectively lowered the level of difficulty of modeling to a large extent. In addition, the function of SNF was to combine different types of experimental data. We applied the SNF algorithm to combine different-type similarity data of miRNA and disease so that it makes the prediction result more reliable. Finally, the alternating gradient descent of IMC algorithm was used to

find the optimal solution, which ensured the reliability of the eigenvectors of miRNA and disease.

There were some limitations that influenced the performance of SNFIMCMDA. First, the materials that we used included verified connections of miRNA–disease, miRNA function similarity, and disease semantic similarity, which may obtain noise and outliers. In addition, SNFIMCMDA used the least square error function that would cause noises and outliers. Furthermore, the model of SNFIMCMDA included several parameters. It was an obvious challenge to discover optimal parameters. Therefore, with the increasing of verified biological data, we would develop optimization strategy to improve accuracy of our model.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

LL designed the experiment, performed the experiment, and wrote the manuscript. ZG and YW performed the experiment. C-HZ processed the data. Y-TW and J-CN revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. U19A2064, 61873001, 61872220, 61861146002, and 11701318). This article is recommended by the 5th Computational Bioinformatics Conference.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2021.617569/full#supplementary-material>

REFERENCES

- Ambros, V. (2001). microRNAs: tiny regulators with great potential. *Cell* 107, 823–826. doi: 10.1016/S0092-8674(01)00616-X
- Ambros, V. (2004). The function of animal MicroRNAs. *Nature* 431, 350–355. doi: 10.1038/nature02871
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Bruce, W., Ilho, R., and Gary, R. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862. doi: 10.1016/0092-8674(93)90530-4
- Calin, G. A., and Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6, 857–866. doi: 10.1038/nrc1997
- Carol, E. D., Kimberly, D. M., Ann, G. S., Ahmedin, J., and Rebecca, L. S. (2019). Cancer statistics for African Americans, 2019. *CA A Cancer J. Clin.* 69, 211–233. doi: 10.3322/caac.21555
- Chen, X., Cheng, J. Y., and Yin, J. (2018). Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. *RNA Biol.* 15, 1192–1205. doi: 10.1080/15476286.2018.1517010
- Chen, X., Gong, Y., Zhang, D. H., You, Z. H., and Li, Z. W. (2018a). DRMDA: deep representations-based miRNA–disease association prediction. *J. Cell Mol. Med.* 22, 472–485. doi: 10.1111/jcmm.13336

- Chen, X., Wang, L., Qu, J., Guan, N. N., and Li, J. Q. (2018b). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* 24, 4256–4265. doi: 10.1093/bioinformatics/bty503
- Chen, X., Huang, Y. A., You, Z. H., Yan, G. Y., and Wang, X. S. (2016). A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* 33, 733–739. doi: 10.1093/bioinformatics/btw715
- Chen, X., Liu, M. X., and Yan, G. Y. (2012). RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.* 8, 2792–2798. doi: 10.1039/c2mb25180a
- Chen, X., Sun, L. G., and Zhao, Y. (2020). NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief. Bioinform.* S2, 1–12. doi: 10.1093/bib/bbz159
- Chen, X., Wu, Q. F., and Yan, G. Y. (2017). RKNMMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol.* 14, 952–962. doi: 10.1080/15476286.2017.1312226
- Chen, X., Yan, C. C., You, Z. H., Huang, Y. A., and Yan, G. Y. (2016a). HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* 7, 65257–65269. doi: 10.18632/oncotarget.11251
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L., Liu, Y., et al. (2016b). WBSMDA: within and between score for MiRNA-disease association prediction. *Sci. Rep.* 6:21106. doi: 10.1038/srep21106
- Chen, X., Yan, C. C., Zhang, X., Li, Z., Deng, L., Zhang, Y., et al. (2015). RBMMMDA: predicting multiple types of disease-microRNA associations. *Entific Rep.* 5:13877. doi: 10.1038/srep13877
- Chen, X., and Yan, G. Y. (2014). Semi-supervised learning for potential human microRNA-disease associations inference. *Entific Rep.* 4:5501. doi: 10.1038/srep05501
- Cheng, Y., Wang, J., Lan, W., Wu, F. X., and Pan, Y. (2017). SDTRLs: predicting drug-target interactions for complex diseases based on chemical substructures. *Complexity* 2017, 1–10. doi: 10.1155/2017/2713280
- Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* 26, 1644–1650. doi: 10.1093/bioinformatics/btq241
- Fu, S. W., Chen, L., and Man, Y. G. (2011). miRNA biomarkers in breast cancer detection and management. *J. Cancer* 2, 116–122. doi: 10.7150/jca.2.116
- Gao, Z., Wang, Y., Wu, Q., Ni, J., and Zheng, C. (2020). Graph regularized L2,1-nonnegative matrix factorization for miRNA-disease association prediction. *BMC Bioinformatics* 21:61. doi: 10.1186/s12859-020-3409-x
- Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., Barabási, A. L., et al. (2007). The human disease network. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8685–8690. doi: 10.1073/pnas.0701361104
- Ha, J., Park, C. H., Park, C. Y., and Park, S. (2020). IMIPMF: inferring miRNA-disease interactions using probabilistic matrix factorization. *J. Biomed. Inform.* 102:103358. doi: 10.1016/j.jbi.2019.103358
- Han, K., Xuan, P., Ding, J., Zhao, Z. J., and Zhong, Y. L. (2014). Prediction of disease-related microRNAs by incorporating functional similarity and common association information. *Genet. Mol. Res. Gmr.* 13, 2009–2019. doi: 10.4238/2014.March.24.5
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., et al. (2019). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* 47, D1013–D1017. doi: 10.1093/nar/gky1010
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA Cancer J. Clin.* 61, 69–90. doi: 10.3322/caac.20107
- Jiang, Q., Hao, Y., Wang, G., Juan, L., and Wang, Y. (2010). Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* S2:4. doi: 10.1186/1752-0509-4-S1-S2
- Koboldt, D. C., Fulton, R. S., Schmidt, H., Dooling, D. J., and Ding, L. (2012). Comprehensive molecular portraits of human breast tumors. *Nature* 490, 61–70. doi: 10.1038/nature11412
- Köhler, S., Bauer, S., Horn, D., and Robinson, P. N., (2008). Walking the Interactome for Prioritization of Candidate Disease Genes. *Am. J. Hum. Genet.* 82, 949–958. doi: 10.1016/j.ajhg.2008.02.013
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., et al. (2013). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 42, D1070–D1074. doi: 10.1093/nar/gkt1023
- Liang, C., Zhang, H., and Mei, Q. (2016). A discriminative feature extraction approach for tumor classification using gene expression data. *Curr. Bioinform.* 11:1. doi: 10.2174/1574893611666160728114747
- Lipscomb, C. E. (2000). Medical subject headings (MeSH). *Bull. Med. Libr. Assoc.* 88, 265–266.
- Liu, Y., Zeng, X., He, Z., and Zou, Q. (2016). Inferring MicroRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 4:1. doi: 10.1109/TCBB.2016.2550432
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., et al. (2008). An analysis of human MicroRNA and disease associations. *PLoS One* 3:e3420. doi: 10.1371/journal.pone.0003420
- Luo, J., and Xiao, Q. (2017). A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J. Biomed. Inform.* 66, 194–203. doi: 10.1016/j.jbi.2017.01.008
- Meola, N., Gennarino, V., and Banfi, S. (2009). microRNAs and genetic diseases. *Pathogenetics* 2:7. doi: 10.1186/1755-8417-2-7
- Miska, E. A. (2005). How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.* 15, 563–568. doi: 10.1016/j.jgde.2005.08.005
- Mohammadi-Yeganeh, S., Paryan, M., Mirab Samiee, S., Soleimani, M., Arefian, E., Azadmanesh, K., et al. (2013). Development of a robust, low cost stem-loop real-time quantification PCR technique for miRNA expression. *Mol. Biol. Rep.* 40, 3665–3674. doi: 10.1007/s11033-012-2442-x
- Mørk, S., Sune, P. F., Albert, P. C., Jan, G., and Juhl, J. L. (2014). Protein-driven inference of miRNA-disease associations. *Bioinformatics* 30, 392–397. doi: 10.1093/bioinformatics/btt677
- Sanghamitra, B., Ramkrishna, M., Ujjwal, M., and Zhang, M. Q. (2010). Development of the human cancer microRNA network. *Silence* 1:6. doi: 10.1186/1758-907X-1-6
- Thackeray, E. W., Charatcharoenwitthaya, P., Elfaki, D., Sinakos, E., and Lindor, K. D. (2011). Colon neoplasms develop early in the course of inflammatory bowel disease and primary sclerosing cholangitis. *Clin. Gastroenterol. Hepatol.* 9, 52–56. doi: 10.1016/j.cgh.2010.09.020
- Thomson, J. M., Parker, J. S., and Hammond, S. M. (2007). Microarray analysis of miRNA gene expression. *Methods Enzymol.* 427:107. doi: 10.1016/S0076-6879(07)27006-5
- Wang, B., Mezzini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11, 333–337. doi: 10.1038/nmeth.2810
- Wu, Q., Wang, Y., Gao, Z., Ni, J., and Zheng, C. (2020). MSCHLMDA: multi-similarity based combinative hypergraph learning for predicting MiRNA-disease association. *Front. Genet.* 11:354. doi: 10.3389/fgene.2020.0354
- Xu, C., Ping, Y., Li, X., Zhao, H., Wang, L., Fan, H., et al. (2014). Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles. *Mol. Biosyst.* 10, 2800–2809. doi: 10.1039/c4mb00353e
- Xu, P., Guo, M., and Hay, B. A. (2004). MicroRNAs and the regulation of cell death. *Trends Genet.* 20, 617–624. doi: 10.1016/j.tig.2004.09.010
- Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., et al. (2013). Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One* 8:e70204. doi: 10.1371/annotation/28592478-72f5-4937-919b-b2342d6ceda0
- You, Z., Huang, Z. A., Zhu, Z. X., Yan, G. Y., and Chen, X. (2017). PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* 13:e1005455. doi: 10.1371/journal.pcbi.1005455
- Yu, H., Chen, X., and Lu, L. (2017). Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. *Entific Rep.* 7:43792. doi: 10.1038/srep43792

- Zeng, X., Zhang, X., and Zou, Q. (2015). Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* 17, 193–203. doi: 10.1093/bib/bv033
- Zhang, H., Cao, L., and Gao, S. (2014). A locality correlation preserving support vector machine. *Pattern Recogn.* 47, 3168–3178. doi: 10.1016/j.patcog.2014.04.004
- Zhen, Y., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H., et al. (2017). dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.* 45, D812–D818. doi: 10.1093/nar/gkx1079
- Zhu, J., Zheng, Z., Wang, J., Sun, J., Wang, P., Cheng, X., et al. (2014). Different miRNA expression profiles between human breast cancer tumors and serum. *Front. Genet.* 5:149. doi: 10.3389/fgene.2014.00149
- Zhu, X., Wang, X., Zhao, H., Pei, T., Kuang, L., and Wang, L. (2020). BHCMDA: a new biased heat conduction based method for potential MiRNA-disease association prediction. *Front. Genet.* 11:384. doi: 10.3389/fgene.2020.00384

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Li, Gao, Zheng, Wang, Wang and Ni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.