Incorporating Clinical, Chemical and Biological Information for Predicting Small Molecule-microRNA Associations based on Non-negative Matrix Factorization

Jiawei Luo[1*], Cong Shen[1], Zihan Lai[1], Jie Cai[1], and Pingjian Ding[2]
[1]College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410083 China
[2]School of Computer Science, University of South China, Hengyang, 421001 China

Abstract-Small molecule(SM) drugs can affect the expression of miRNAs, which plays crucial roles in many important biological processes. The chemical structure and clinical information of small molecule can simultaneously incorporate information such as anatomical distribution, therapeutic effects and structural characteristics. It is necessary to develop a novel model that incorporates small molecule chemical structure and clinical information to reveal the unknown small molecule-miRNA associations. In this study, we developed a new framework based on non-negative matrix factorization, called SMANMF, to discover the potential small molecules-miRNAs associations. First, the functional similarity of two miRNAs can be obtained by computing the overlap of the target gene sets in which the miRNAs interact together, and we integrated two types of small molecule similarities, including chemical similarity and clinical similarity. Then, we utilized a non-negative matrix factorization model to discover the unknown relationship between small molecules and miRNAs. The evaluation results indicate that our model can achieve superior prediction performance compared with previous approaches in 5-fold cross-validation. At the same time, the results of case studies also reveal that the SMANMF model has good predictive performance for predicting the potential association between small molecules and miRNAs.

***Index Terms*—Small molecule-associated miRNAs prediction, clinical similarity, chemical similarity, Non-negative Matrix Factorization.**

I. Introduction

RECENT studies have indicated that miRNAs are important regulatory molecule in many crucial biological processes [1]. Improvements in miRNA characterization and functional analysis techniques not only reveal their role in various cellular processes, but also reveal the unusual expression patterns of miRNAs in various diseases[2][3]. Several studies have revealed the multiple roles of miRNAs in a variety of important biological processes[4][5]. In recent years, relevant studies centering on miRNA, such as functional similarity calculation of miRNA[6], prediction of relationship between miRNA and target[7], prediction of associations between miRNAs and diseases[8][9][10], motif discovery in co-regulatory network[11] and module identify[12][13], have become important research directions in bioinformatics[14]. The above studies to some extent indicated the importance of miRNAs in biological processes.

Small molecule drugs are a class of complex organic compounds, about 1 nanometer in size, which can help regulate biological processes in molecular biology and pharmacology[15]. For a long time, people have studied the function of small molecules from small molecule targeted therapeutic proteins[16], but only 10-15% of these small molecule targeting proteins are directly related to disease[17]. Among all the proteins associated with the disease, many are lack of motifs that can bind directly to small molecules[17]. Therefore, from the perspective of human gene expression proteins to design small molecule drugs that account for only a small proportion[18]. However, assumed that the design of small molecule drugs can be studied from small molecules acting on non-coding RNAs, the research scope of small molecule drugs will be broadened[19].

At present, high-throughput screening methods for small molecule-miRNA modifications have been extensively studied, which provide a novel direction for miRNA-targeted treatment[20]. Small molecules can indirectly reduce or increase the expression of miRNA by binding to transcription factors or altering the miRNA promoter region[21]. Small molecules can also bind to RNA endonucleases to disrupt miRNA maturation[22]. In conclusion, it is important to study the associations between small molecules and miRNAs for the treatment of diseases and the clinical application of known drugs[20][21]. However, since most biological processes in living organisms are especially complex, it is a long-term, complicated and time-consuming work to determine the relationship of small molecules with miRNAs through experiments. Therefore, it is urgent to present novel calculational framework for discovering the unknown relationship between small molecules and miRNAs by known data and associations.

Recently, some methods have been developed to calculate the associations between small molecules and miRNAs. For instance, feature-based models are the primary types of calculational models. Wang et al.[23] developed a new calculational framework to discover unknown relationship between small molecules and miRNAs by computing functional similarity, which also revealed the associations of drugs and diseases based on integrating known relationship between small molecules and miRNAs by validated disease related miRNAs. Wang et al.[24] presented a new computational model for small molecule-miRNA correlation prediction (RFSMMA) based on random forests. Salma et al.[25] constructed

Manuscript received December 1, 2012; revised August 26, 2015. Corresponding author: M. Shell (email: http://www.michaelshell.org/contact.html).

a prediction model for mining small RNA-binding molecules form macromolecular data sets based on Random Forest and Naive Bayes. Feature-based methods can achieve good results on most data sets, but there is still not a clear definition in terms of effective feature extraction and prediction model selection.

In contrast, network-based models are another type of predictive model used to calculate the associations of small molecules and miRNAs. Lv et al.[26] proposed a new computational framework to comprehensively reveal the unknown relationship between small molecules and miRNAs by utilizing Random Walk with Restart algorithm on the bipartite network. Li et al.[27] established a network-based miRNA pharmacogenomics model based on the inference (SMiR_NBI) framework for predicting small molecule-miRNA networks to reveal the potential mechanisms of miRNA-mediated responses of anticancer drug. Qu et al.[28] constructed predicting calculation model(HSSMMA) of small molecules with miRNAs based on HeteSim inference by implementing path-based HeteSim measurement methods on heterogeneous networks. Meng et al.[29] developed a systematic computational model to build an association network of bioactive small molecules with miRNAs in Alzheimers disease(AD), in which the functional and topological analysis of the small molecules-miRNA in Alzheimers disease from multiple perspectives were performed. Guan et al.[30] proposed a small molecule-miRNA correlation prediction model(GISMMA) based on meta-pattern interaction. Yin et al.[31] developed a computational model of heterogeneous graph inference and sparse learning for small molecule-miRNA association prediction. Qu et al.[32] predicted potential small molecule-miRNA association based triple layer heterogeneous network. Wang et al.[33] developed a computational framework (CLDISMMA) for predicting small molecule-miRNA associations based on cross-layer dependency inference on multilayered networks. The experiments results show that CLDISMMA obtained great performance. In order to improve the accuracy, Zhao et al.[34] presented a novel model (SNMFSMMA), which used symmetric nonnegative matrix factorization to discover potential small molecule-associate miRNAs. Network-based models can avoid some short-comings that there are difficulties about effective feature extraction and prediction model selection in feature-based models, but there are still difficulties in how to effectively fuse multi-source information. In simple terms, regardless of the network-based calculation models or the feature-based calculation models, which often use the single small molecule similarity or simply weight a variety of similarity, which does not effectively utilize the chemical and clinical information of small molecules. However, the clinical similarity of small molecules is calculated by ATC codes can simultaneously incorporate information such as anatomical distribution, therapeutic effects and structural characteristics. Meanwhile, the chemical similarity of small molecules computed by MACCS fingerprints of each small molecule can fully consider chemical structure information of small molecules. Thus these small molecule similarity calculations incorporate a variety of pharmacological information, which plays an important role in drug-target prediction and drug combination research.

In this study, we develop a novel computational framework (SMANMF) to reveal the relationships between small molecules and miRNAs. SMANMF fully exploits the clinical information and chemical information between small molecules, the small molecule-miRNA network, and the similarities for miRNAs based on experimentally validated miRNAs-genes associations. Accumulated studies have found that matrix factorization technique is used in many areas, such as recommendation systems [35], microRNA-disease association prediction [14][36][37], long noncoding RNA-disease association prediction[38][39], and synergistic drug combinations[40], etc. Inspired by above, we integrate known small molecules similarity, miRNAs similarity, and associations between small molecules and miRNAs as a matrix factorization model, which uses two types of small molecule similarities and one type of miRNA similarity as graph regularizations, and the L2 paradigm is added to prevent overfitting. Finally, we use a 5-fold cross-validation to compare with the previous method. In order to verify the robustness of the model, we remove the association of 10%, 20%, 30% and 40% of known small molecules with miRNAs, respectively. We also design experiments to demonstrate the importance of multi-drug information. Finally, case studies are used to demonstrate the performance of this model in discovering the potential association of small molecules with miRNAs.

II. Materials and methods

A. Methods Overview

The computational model SMANMF, could discover the potential associations of small molecules with miRNAs, which can be divided into three parts. Firstly, we compute the similarities of small molecules based on the chemical and clinical information, and the similarities of miRNAs based on miRNA-gene associations. Meanwhile, we construct the interaction network between small molecules and miRNA. Secondly, in order to reveal the potential relationship of small molecules with miRNAs, the framework of non-negative matrix factorization is applied to discover the unknown associations. Finally, we use the results of nonnegative matrix factorization to predict the association between small molecules and miRNAs. The overall process is shown in Fig.1, where the notations and explanations are in Table 1.

TABLE I
NOTATIONS AND EXPLANATIONS

Notation	Explanation
d	Small molecule
m	miRNA
N_d	The count of small molecules
N_m	The count of miRNAs
k	Sub-space dimensionality
$S^c \in \mathbb{R}^{N_d \times N_d}$	Chemical similarity of small molecules
$S^a \in \mathbb{R}^{N_d \times N_d}$	Clinical similarity of small molecules
$S^m \in \mathbb{R}^{N_m \times N_m}$	The similarity of miRNAs
$Y \in \mathbb{R}^{N_d \times N_m}$	Adjacency matrix
$A \in \mathbb{R}^{N_d \times k}$	Representation matrix of small molecule
$B \in \mathbb{R}^{N_m \times k}$	Representation matrix of miRNA

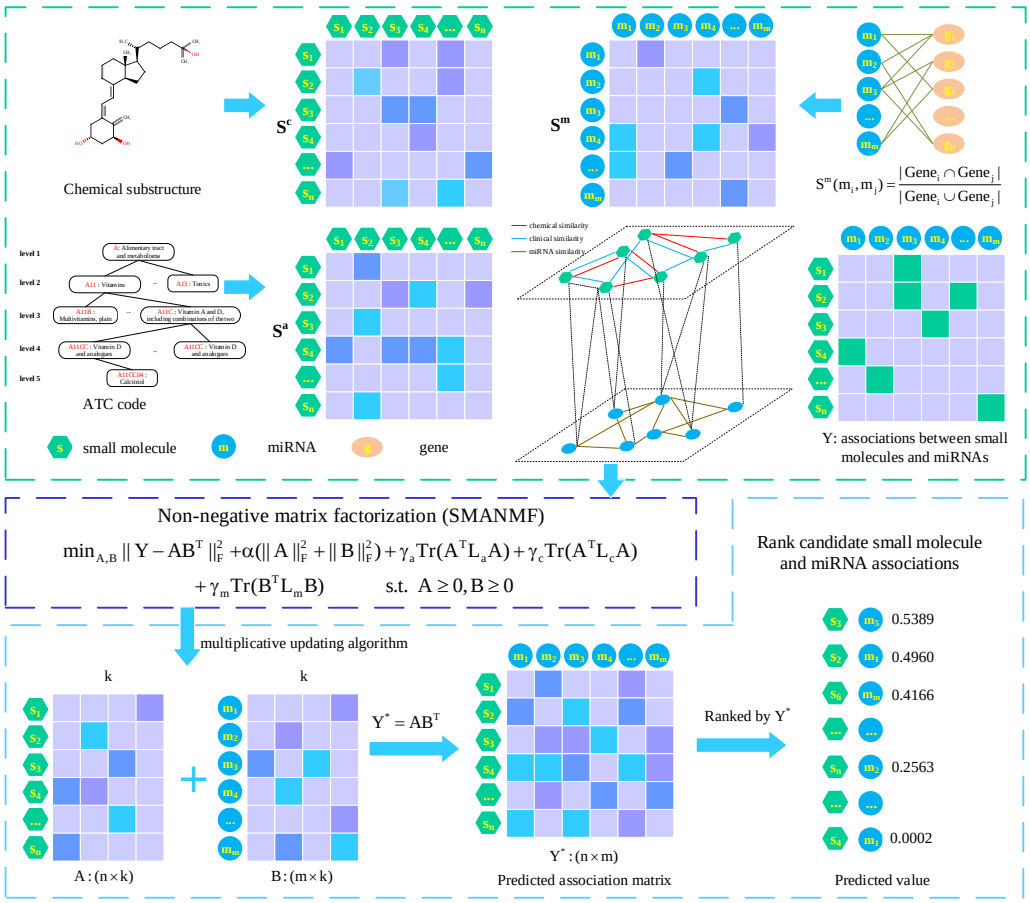


Fig. 1. Overall framework of SMANMF for predicting unknown associations between small molecules and miRNAs.

B. Construction of Small Molecule-miRNA Network

1) Small Molecule Pairs Similarity

Chemical similarity The DrugBank database[41] provides chemical structure and the Open Babel[42] is a valid programming library that has been effectively applied in computing MACCS fingerprints of each small molecule. Assumed that the counts of bits of small molecule d_i and m_j are a and b, respectively, while c is the bits set in the fingerprints of both small molecules, the chemical similarity of a small molecule-small molecule pair is represented as:

$$S^c = \frac{c}{a+b-c} \tag{1}$$

The chemical similarities are often applied to drug discovery[43] and drug combinations[44] providing a value between 0 and 1.

Clinical similarity The ATC coding systems[45] have been effectively applied in calculating the drug-drug similarities. The DrugBank database[41] provides the ATC codes for the drugs. We could define the ATC code similarity $S^l(d_i, d_j)$ between drugs d_i and d_j in l-th level based on the ATC codes as follows:

$$S^l(d_i, d_j) = \frac{\left|d_i^l \cap d_j^l\right|}{\left|d_i^l \cup d_j^l\right|} \tag{2}$$

where d_i^l and d_j^l represent total ATC codes of drugs d_i and d_j at the l-th level respectively. If both drugs d_i and d_j do not have ATC codes, then their similarity is 0. We calculate the clinical similarity of drugs d_i with d_j based on the ATC code similarity $S^l(d_i, d_j)$ as follows:

$$S^a(d_i, d_j) = \frac{\sum_{k=1}^{l} S^l(d_i, d_j)}{n} \tag{3}$$

where l represents the count of layers of the ATC codes (ranging from 1 to 3). The first three layers of the ATC code represent anatomic classification, therapeutic classification, and pharmacological classification, which cover most of the information of the ATC code[46]. Therefore, we use the first three layers of the ATC code to calculate the clinical similarity of small molecules, which is commonly used to facilitate the calculation of the small molecular drugs similarity[47]. There are multiple ATC codes in many drugs. For example, caffeine (a central nervous stimulant) has three different ATC codes: V04CG30, R03DA20, N06BC01. We compute the ATC code similarity for each ATC code in each layer while the drug has multiple ATC codes, then obtain the clinical similarity by taking the average of ATC code similarity in each layer that are used in[48].

2) MiRNA Similarity Measure

In this study, we measure the similarity based on shared genes and download the miRNA-gene association from miRTarBase[49] database. Meanwhile, we use the GSFS method[50] to compute the similarity of the miRNAs m_i with

m_j by calculating the overlap of their miRNA target gene sets. The miRNA similarity $S^m(m_i, m_j)$ of miRNAs m_i and m_j is introduced based on the target gene sets as follows:

$$S^m(m_i, m_j) = \frac{|Gene_i \cap Gene_j|}{|Gene_i \cup Gene_j|} \tag{4}$$

where sets of genes associated with miRNA m_i and m_j are represented as $Gene_i$ and $Gene_j$, respectively. The count of sets is represented as $|\cdot|$. Assuming that miRNAs m_i and m_j do not have a synergistic gene, then their similarity is 0.

3) The Associations of Small Molecules with MiRNAs

The bipartite association network involving in small molecules and miRNAs is constructed by the known relationship between small molecules and miRNAs, and the known small molecule-associated miRNAs from SM2miR[51]. The edge sets are constructed by the matrix $Y = [y_{ij}] \in \mathbb{R}^{N_d \times N_m}$, where y_{ij} indicates the association between the i-th small molecule and the j-th miRNA. The value of y_{ij} is 1 if small molecule d_i is connected with miRNA m_j and 0,otherwise.

C. The Prediction Model of Small Molecule-miRNA Association

1) Standard NMF

Non-negative matrix factorization (NMF) is a significant method which can be effectively applied to the data representation [52][53]. Its aims to obtain a better problem analysis and presentation by decomposing the original matrix into two non-negative matrices. Let matrix $Y \in \mathbb{R}^{N_d \times N_m}$ represent the relationship between small molecules and miRNAs, it is known by NMF that the matrix Y can be represented by the multiplication of two matrices, for instance, $Y \approx AB^T$, which needs to meet $A \in \mathbb{R}^{N_d \times k}$ and $B \in \mathbb{R}^{N_m \times k} (k \ll min(N_d, N_m))$. The objective function is mathematically formulated for the problem of miRNA-related small molecules prediction as follows:

$$min_{A,B} \left|\left|Y - AB^T\right|\right|_F^2 \quad s.t. \ A \geq 0, \ B \geq 0 \tag{5}$$

where $||\cdot||_F$ is the Frobenius norm of a matrix. The iterative algorithm of literature[54] can minimize the above objective function.

2) SMANMF

The standard NMF in formula (5) fails to incorporate topological information and molecular structure information in the data space, it only performs the learning in the Euclidean space [55][56]. To prevent overfitting and fully consider the chemical and clinical information of small molecule, and significantly improve the learning performance, a new objective function is presented by incorporating both graph Laplacian regularization items and Tikhonov (L2) into the NMF for discovering the associations between small molecules and miRNAs. We ensure the smoothness of A and B by the Tikhonov regularization [57] and fully exploit the topological information and molecular structure information by the graph regularization[58]. The objective function of SMANMF can be calculated as follows:

$$\begin{aligned} min_{A,B} &\left|\left|Y - AB^T\right|\right|_F^2 + \alpha(||A||_F^2 + ||B||_F^2) \\ &+ \gamma_a Tr(A^T L_a A) + \gamma_c Tr(A^T L_c A) \\ &+ \gamma_m Tr(B^T L_m B) \ s.t. \ A \geq 0, \ B \geq 0 \end{aligned} \tag{6}$$

where $\alpha, \gamma_a, \gamma_c$ and γ_m represent the regularization parameters, Tr (.) is the trace of a matrix, $L_a = D_a - S^a$, $L_c = D_c - S^c$ and $L_m = D_m - S^m$ represent the graph Laplacian matrices of S_a, S_c and S_m, respectively. S_a and S_c represent the small molecules chemical similarity matrices and clinical similarity matrices, S_m is the miRNAs similarity matrices. D_a, D_c and D_m are the diagonal matrices and the values on the diagonal are rows(or columns) sums of S_a, S_c and S_m, respectively.

3) Optimization

To obtain the minimum of formula (6), let $\Phi = [\varphi_{ik}]$ and $\Psi = [\psi_{jk}]$ represent the Lagrange multipliers[59] and meet the constrains $a_{ik} \geq 0$ and $b_{jk} \geq 0$, respectively. The corresponding optimization function L_f of formula (6) is written as follows:

$$\begin{aligned} L_f =& Tr(YY^T) - 2Tr(YBA^T) + Tr(AB^TBA^T) \\ &+ \alpha Tr(AA^T) + \alpha Tr(BB^T) + \gamma_a Tr(A^T L_a A) \\ &+ \gamma_c Tr(A^T L_c A) + \gamma_m Tr(B^T L_m B) \\ &+ Tr(\Phi A^T) + Tr(\Psi B^T) \end{aligned} \tag{7}$$

The partial derivatives of A and B can be calculated as:

$$\frac{\partial L_f}{\partial A} = -2YB + 2AB^TB + 2\alpha A + 2\gamma_a L_a A + 2\gamma_c L_c A + \Phi \tag{8}$$

$$\frac{\partial L_f}{\partial B} = -2Y^TA + 2BA^TA + 2\alpha B + 2\gamma_m L_m B + \Psi \tag{9}$$

The Karush−Kuhn−Tucker(KKT) condition[60] $\varphi_{ik} a_{ik} = 0$ and $\psi_{jk} b_{jk} = 0$ is applied into the following equations for a_{ik} and b_{ik}:

$$\begin{aligned} &-(YB)_{ik}a_{ik} + (AB^TB)_{ik}a_{ik} + (\alpha A)_{ik}a_{ik} \\ &+ [\gamma_a(D_a - S^a)A]_{ik}a_{ik} + [\gamma_c(D_c - S^c)A]_{ik}a_{ik} = 0 \end{aligned} \tag{10}$$

$$\begin{aligned} &-(Y^TA)_{jk}b_{jk} + (BA^TA)_{jk}b_{jk} + (\alpha B)_{jk}b_{jk} \\ &+ [\gamma_m(D_m - S^m)B]_{jk}b_{jk} = 0 \end{aligned} \tag{11}$$

The updating rules could be simplified as follows based on above equation:

$$a_{ik} \leftarrow a_{ik} \frac{(YB + \gamma_c S^c A + \gamma_a S^a A)_{ik}}{(AB^TB + \alpha A + \gamma_a D_a A + \gamma_c D_c A)_{ik}} \tag{12}$$

$$b_{jk} \leftarrow b_{jk} \frac{(Y^TA + \gamma_m S^m B)_{jk}}{(BA^TA + \alpha B + \gamma_m D_m B)_{jk}} \tag{13}$$

updating the formula (12) and formula (13) until the nonnegative matrices A and B converges.

The predicted score matrix of small molecule-miRNA associations can be obtained based on $Y^* = AB^T$, then we rank the correlation scores of small molecules and miRNAs in Y^*. In theory, the higher the ranking, the greater the likelihood that the corresponding small molecule and miRNA are associated. Therefore, we can use the obtained correlation score to evaluate the predicted result by conducting experiments.

III. RESULTS

A. Data Description

We obtained 5112 known connections between small molecules and miRNAs from SM2miR[51]. For one thing, as done in previous studies[61][62], we merged different small molecule-miRNA copies that produce the same mature miRNA. For another, after the removel of duplicate associations, 251 small molecules and 901 miRNAs collected from SM2miR, 4182 associations between small molecules and miRNAs were obtained for calculation of this model. To compute the similarity of miRNAs, we collected the miRNA-gene associations from miRTarBase [49] database, including 196565 associations of 22697 genes and 706 miRNAs are retained. In SMANMF, the similarity of small molecules includes chemical similarity and clinical similarity, for the chemical similarity, 9296 SMs and their chemical structure information are acquired from the DrugBank[41], for the chemical similarity, we acquired 2247 SMs and their ATC codes from the DrugBank[41]. To facilitate the calculation, we finally obtained 251 SMs and 901 miRNAs based on the small molecule-miRNA database in SM2miR[51]. The details of multi-type data are shown in Table 2.

TABLE II
THE DETAILS OF MULTI-TYPE DATA

data type	database	description
small molecule-miRNA associations	SM2miR[51]	251 small molecules, 901 miRNAs and 4182 associations
small molecule clinical similarity	Drugbank[41]	83 small molecules and their ATC codes
small molecule chemical similarity	Drugbank[41]	251 small molecules and their chemical structure
miRNA functional similarity	miRTarbase[49]	706 miRNAs, 22697 genes and 196565 associations

B. Experimental Evaluations and Discussions

1) Experimental Settings and Evaluation Metrics

To comparatively study SMANMF effect in predicting small molecule-miRNA associations, we use 5-fold cross-validation (CV) based on known relationship between small molecules and miRNAs. All known associations of small molecules with miRNAs are divided into 5 equal subsets randomly; four of them are used as training sets and the remaining one is used as a test sample. In this study, the parameters are estimated based on CV experiments. We use grid search for all parameter combinations. The mainly combinations come from the following values: $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ for k and $\{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}$ for α. Subsequently, we choose the three parameters from $\{10^0, 10^{-1}, 10^{-2}, 10^{-3}\}$ for γ_m, γ_a and γ_c.

The Receiver Operating Characteristic (ROC) curve[63] is effectively applied in studying the generalization performance of learners. The ROC curve can be obtained by taking different thresholds for the TPRs and FPRs. TPR and FPR are introduced as follows:

$$TPR = \frac{TP}{TP + FN},\ FPR = \frac{FP}{TN + FP} \tag{14}$$

where TP and TN represent true positives and true negatives, and the false negatives and false positives are represented as FN and FP, respectively. The area under the ROC curve (AUC) is effectively utilized to evaluate overall prediction effect.

Assumed that the known (positive samples) and unknown associations (negative samples) between small molecule and miRNA are serious imbalance, the recall is more effective[64], and defined as:

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

In order to evaluate the performance of the model from multiple perspectives and consider the sparsity of known small molecule and miRNA associations, we use specificity and G_mean as evaluation indicators as follows:

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$G\_mean = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{TN + FP}} \tag{17}$$

The recall indicates the completeness of the correct classification of the positive samples. The specificity indicates the completeness of the negative classification of the negative samples. G_mean also pays attention to the performance of the two categories, indicating the equilibrium value of the classification accuracy of the positive and negative categories.

2) Baseline Methods

We compare SMANMF with some previous models, including the computational model based on random forest for predicting small molecule-miRNA associations[24], the methods based on random walk with restart[26] and network-based identification model[27], to identify the superiority of the developed SMANMF. The data of SMANMF is applied to the comparison model to obtain the comparison experiment results.

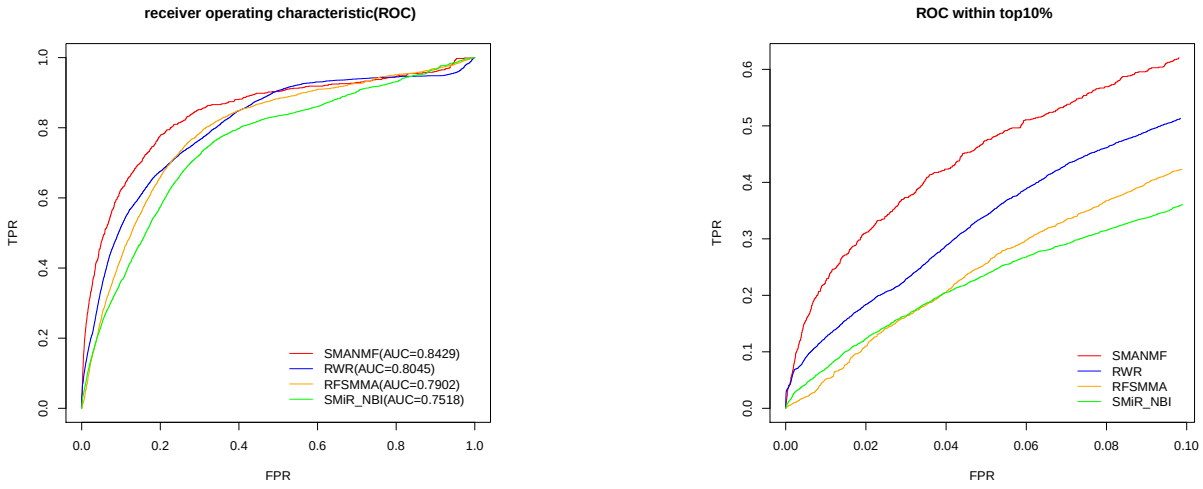
RFSMMA[24] was a calculation approach based on random forest for predicting the associations of small molecules with miRNAs (RFSMMA). The numbers of estimators, max-features and min samples leaf were set to 100, 0.2, and 10 respectively according to the original literature[24].

RWR[26] was a computational framework to comprehensively discover associations between small molecules and miRNAs based on Random Walk with Restart algorithm (RWR). The parameters of RWR were set as $\gamma = 0.7$, $\lambda = 0.5$ and $\eta = 0.5$ according to the literature [26].

SMiR_NBI[27] was constructed by network-based inference(NBI) method to discover the underlying mechanisms of anticancer drug responses mediated by miRNAs. SMiR-NBI model displayed high performance in cross-validation and experimental validation in several case studies.

C. Performance Evaluation

As shown in Fig.2, the SMANMF framework was superior compared with RWR, RFSMMA and SMiR_NBI. As we expected, among all of the 4182 known associations, the AUC values of SMANMF, RWR, RFSMMA and SMiR_NBI are 0.8429, 0.8045, 0.7902 and 0.7518 in Fig.2(a), respectively. The experiment results shows that SMANMF achieve the



(a) (b)

Fig. 2. Comparison performance of SMANMF with the previous methods. (a) ROC curves of SMANMF with the previous methods. (b) ROC curves of SMANMF with the previous methods within top 10%.

highest performance among all methods. Fig.2 (b) shows the ROC curve of top 10% to better study the top ranking of the prediction results of small molecules and miRNAs associations. The top 10% ROC curve of SMANMF is located at the top of the other three comparison methods, indicating that the SMANMF method also has better performance in predicting the relationship between small molecules and miRNA in the top 10% results.

We also use compared indicators recall, specificity and G_mean within top 5%, 10%, 15%, 20% and 25% to verify the model's performance. As shown in Table 3, SMANMF was 12% higher on recall than the second best method RWR, and SMANMF had a greater advantage in specificity than the comparison method. We compared the SMANMF with the previous method on the G_mean indicator for balancing the recall and specificity. The result shows that the G_mean values of SMANMF reach 0.6654, 0.7483, 0.7715, 0.7885 and 0.7827, from top 5% to top 25% respectively. SMANMF's G_mean values are on average 10% higher than the second best method, indicating the superiority of SMANMF in discovering the potential associations between small molecules and miRNAs. Whether AUC, recall, specificity or G_mean as evaluation indicators, SMANMF model has achieved good performance. The reason of obtaining great result may be that the two types of similarities of small molecules are considered in SMANMF model, and the non-negative matrix factorization model is great for processing multi-source data.

TABLE III
THE AVERAGE RECALL, SPECIFICITY AND G_MEAN ACROSS ALL TESTED SMALL MOLECULE BASED ON THE DATA OF HAVING REMOVED 10% KNOWN ASSOCIATIONS AT DIFFERENT TOP K CUTOFFS

	method	Ranking threshold				
		top 5%	top 10%	top 15%	top 20%	top 25%
Recall	SMANMF	0.4653	0.6208	0.6986	0.7751	0.8146
	RFSMMA	0.2526	0.4227	0.5522	0.6562	0.7294
	RWR	0.3368	0.5132	0.6045	0.6744	0.7237
	SMiR_NBI	0.2340	0.3603	0.4677	0.5722	0.6605
specificity	SMANMF	0.9516	0.9020	0.8521	0.8022	0.7521
	RFSMMA	0.9508	0.9012	0.8515	0.8017	0.7518
	RWR	0.9511	0.9016	0.8517	0.8018	0.7518
	SMiR_NBI	0.9507	0.9010	0.8512	0.8014	0.7515
G_mean	SMANMF	0.6654	0.7483	0.7715	0.7885	0.7827
	RFSMMA	0.4899	0.6171	0.6856	0.7253	0.7405
	RWR	0.5659	0.6800	0.7174	0.7352	0.7375
	SMiR_NBI	0.4715	0.5697	0.6309	0.6771	0.7045

D. Importance of Various Drug Information

In order to fully illustrate the need to consider multiple drug information, we designed three variant models based on SMANMF as follows: 1) SMANMFnoac: SMANMFnoac model did not consider the chemical similarity and clinical similarity of small molecules simultaneously, only the similarity of miRNAs was used as regularization item. 2) SMANMFc: SMANMFc model only considered the chemical similarity of small molecule and ignored the clinical similarity. 3) SMANMFa: SMANMFa model only considered the clinical similarity of small molecule and ignored the chemical similarity.

We set the parameters of variant models the same as those of our proposed model SMANMF. In order to more fairly evaluate the performance of SMANMF and SMANMFnoac, SMANMFc, SMANMFa, we used ROC curve and AUC scores for comparison. The results of the three models are shown in Fig.3. We used AUC as the performance evaluation index, the AUC values of SMANMF, SMANMFa, SMANMFc and SMANMFnoac are 0.8429, 0.8015, 0.8023 and 0.5435 respectively. The performance of SMANMF model is better than SMANMFa, SMANMFc and SMANMFnoac. As shown by the above experimental results, the SMANMF took full account into the chemical similarity and clinical similarity of small molecules at the same time, while predicting the relationship between small molecules and miRNAs. It may be on account of the limitations of only considering the chemical

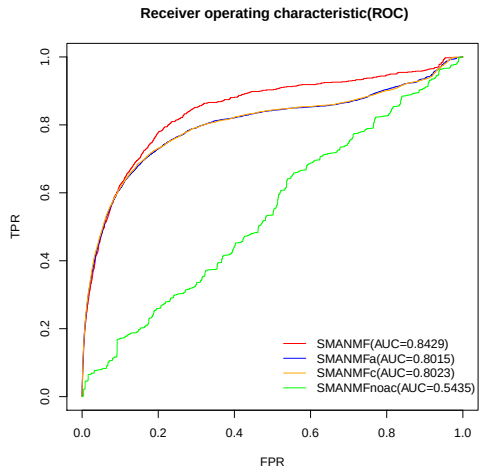
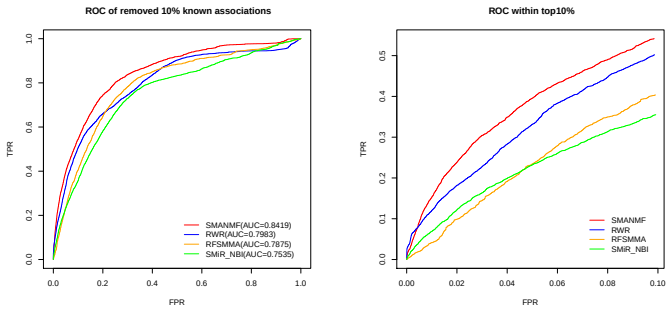


Fig. 3. The ROC curve with the performance comparison between the SMANMF and the variant model.

structure similarity or clinical similarity, and the properties of small molecules could not be fully utilized.

E. Robustness Analysis

Since there were many false positives in the calculation methods of existing small molecule miRNA relationship prediction, the error tolerance of the model played an important role in the relationship prediction. Therefore, we compared the robustness of SMANMF model with the previous methods to verify the error tolerance.



(a) (b)

Fig. 4. the ROC after removed 10% known associations, (a) the full ROC of SMANMF and previous methods, (b) the top 10% ROC of the SMANMF and previous methods.

We randomly removed 10%, 20%, 30%, and 40% of known small molecule and miRNA associations. 5-fold cross-validation was performed on the processed data, then repeated ten times and took the average as results. In order to ensure the fairness of the experiment, the parameters of the SMANMF and the comparison method were all unchanged. The AUC, recall, specificity and G_mean were still selected as verification indicators. The results of removing the 10% association were shown in Fig.4 and Table 4. The results of removing 20%, 30%, and 40% were shown in the supplemental material. In Fig.4, after removed 10% of the known small molecule-miRNA associations, the SMANMF's AUC value can still reach 0.8419 and is 5.5% higher than the AUC value of the second good model RWR of 0.7983. In terms of indicators such as recall, specificity and G_mean, SMANMF also indicates better performance in Table 4. Supplementary materials Fig.1 and Table 1 show the performance comparison after removing 20% of known associations. The AUC value of SMANMF is 0.8391, which is still higher than the comparison method and showed good results in the recall, specificity and G_mean. Fig.2 and Table 2, Fig.3 and Table 3 of the supplementary materials show the performance comparison of the 30% and 40% known associations removed, respectively. Even if the 40% known relationship was removed, the AUC value of the SMANMF model can still reach 0.8326. SMANMF's recall, specificity and G_mean values in top 25% are 0.7657, 0.7527 and 0.7591, respectively. Compared with the previous method, it still demonstrates better performance. The good robustness of the SMANMF model may be the result of the fusion of similarities of multiple small molecules. Because we fuse multiple small molecule similarities can reduce the impact of known small molecule-miRNA association data on the model.

TABLE IV
THE AVERAGE RECALL, SPECIFICITY AND G_MEAN ACROSS ALL TESTED SMALL MOLECULE BASED ON THE DATA WHICH IS REMOVED 10% KNOWN ASSOCIATIONS AT DIFFERENT TOP K CUTOFFS

	method	Ranking threshold				
		top 5%	top 10%	top 15%	top 20%	top 25%
Recall	SMANMF	0.3904	0.5415	0.6569	0.7412	0.7979
	RFSMMA	0.2311	0.4029	0.5316	0.6441	0.7260
	RWR	0.326	0.5016	0.5979	0.6585	0.6995
	SMiR_NBI	0.2298	0.3548	0.4803	0.5779	0.6625
specificity	SMANMF	0.9512	0.9016	0.8519	0.8021	0.7522
	RFSMMA	0.9507	0.9012	0.8515	0.8018	0.7520
	RWR	0.9510	0.9015	0.8517	0.8019	0.7519
	SMiR_NBI	0.9507	0.9010	0.8513	0.8016	0.7518
G_mean	SMANMF	0.6094	0.6986	0.7481	0.7710	0.7747
	RFSMMA	0.4686	0.6025	0.6728	0.7186	0.7389
	RWR	0.5568	0.6723	0.7136	0.7266	0.7251
	SMiR_NBI	0.4673	0.5654	0.6394	0.6805	0.7057

F. Parameter Sensitivity Analysis

In this section, we evaluated the scalability and how parameters influenced the performance. Especially, we evaluated the effect of the embedding dimension k. For brevity, we reported the results of AUC scores based on the datasets in section 3.1. Noted that except for the parameter being tested, we set all other parameters to default values. We have shown how the dimension of embedding dimensions k affects the performance in Fig.5. Since higher embedding dimensions can embody more information, the performance raises firstly while the number of embedding dimension increases. Then after the dimension k exceeds 70, the performance to drop slowly. The reason may be that SMANMF needs a suitable dimension to encode the information and larger dimension may introduce additional redundancies.

G. Case Studies

Case studies further confirmed the superiority of the SMANMF model in predicting the potential associations between

TABLE V
THE TOP 15 POTENTIAL MIRNA CANDIDATES DISCOVERED BY SMANMF FOR THE THREE SELECTED SMALL MOLECULES

Rank	SM	miRNA	evidence	SM	miRNA	evidence	SM	miRNA	evidence
1	CID3385	mir-451	28670496	CID 5793	mir-451	25937278	CID 8490	mir-203	unconfirmed
2	CID3385	mir-92	29849934	CID 5793	mir-151	unconfirmed	CID 8490	mir-92	unconfirmed
3	CID3385	mir-320	28255248	CID 5793	mir-92	unconfirmed	CID 8490	mir-663	25616258
4	CID3385	mir-103	unconfirmed	CID 5793	mir-320	22900199	CID 8490	mir-190	24672518
5	CID3385	mir-181a	29795190	CID 5793	mir-103	29511499	CID 8490	mir-365	unconfirmed
6	CID3385	mir-99b	25356050	CID 5793	mir-663	unconfirmed	CID 8490	mir-21	19270793
7	CID3385	mir-125b	28670496	CID 5793	mir-191	unconfirmed	CID 8490	mir-487b	unconfirmed
8	CID3385	mir-335	unconfirmed	CID 5793	mir-365	unconfirmed	CID 8490	mir-155	unconfirmed
9	CID3385	mir-487b	unconfirmed	CID 5793	mir-181a	29207650	CID 8490	mir-17	unconfirmed
10	CID3385	mir-22	29042944	CID 5793	mir-21	29207650	CID 8490	mir-31	unconfirmed
11	CID3385	mir-26a	29719405	CID 5793	mir-99b	unconfirmed	CID 8490	mir-133a	25616258
12	CID3385	mir-181b	unconfirmed	CID 5793	mir-125b	26966351	CID 8490	mir-10a	19270793
13	CID3385	mir-126	27203443	CID 5793	mir-27b	28698281	CID 8490	mir-324	unconfirmed
14	CID3385	mir-107	26636340	CID 5793	mir-24	unconfirmed	CID 8490	mir-34c	unconfirmed
15	CID3385	mir-151	unconfirmed	CID 5793	mir-335	29122960	CID 8490	mir-638	25616258

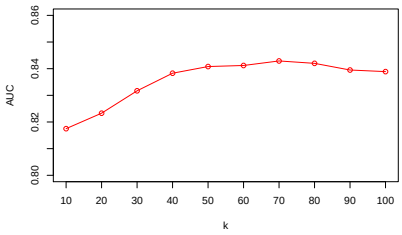


Fig. 5. Analysis of parameter sensitivity.

small molecules and miRNAs. In this part, all of the known associations were used to predict model, and the unknown association as validating. In case study, all the parameters involved were set as optimal parameters. For each small molecule, the top miRNA associated with the corresponding small molecule is obtained by predicting the score.

To further evaluate the prediction results of SMANMF, case studies were conducted based on three common small molecules, namely, 5-Fluorouracil(CID 3385), Glucose(CID 5793), and Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX)(CID 8490). These three small molecules are the most closely related to human life and health. 5-Fluorouracil play key role in treatment of treatment of colon cancer. Meanwhile, 5-FU has serious cardiac toxicity that manifested as cardiogenic shock, ventricular fibrillation and myocardial infarction[65]. Glucose is essential for life and a severe drop in the blood levels rapidly leads to coma and death[66]. RDX has been classified as a class C potential human carcinogen by the U.S. Environmental Protection Agency[67]. We verify the predicted small molecule-miRNA associations by finding literatures in Pubmed. Table 5 provided potential candidate miRNAs that may be associated with three small molecules.

As shown above, we selected 15 miRNAs where the fact that 10, 8, and 6 of them were associated with 5-Fluorouracil, Glucose and RDX respectively was verified. Taking the small molecule 5-Fluorouracil as an example, the mir-92 associated with it was verified in the literature[68]. The experiments in the literature showed that transfection of mir-92a significantly blocked the expression of caspase-3 and PARP 5-Fu-induced apoptosis. Above discovery indicated that these miRNAs were associated with small molecules. Fig.6 showed that in the association network, in which is the predicted candidates miRNA in the top 20 for the three small molecules, some of the top-ranked candidate miRNAs were associated with one or more small molecules. In general, the prediction results further verified that the availability of SMANMF in revealing unknown associations of small molecule with miRNA.

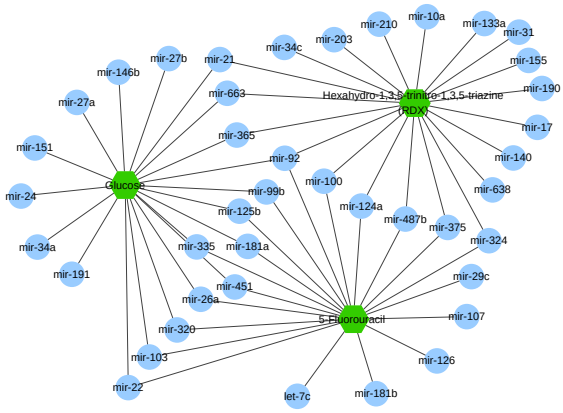


Fig. 6. The predicted candidate miRNAs top 20 for the three small molecules.

IV. CONCLUSIONS

Recently, some researches have focused on discovering potential associations of small molecules and miRNAs by computational models. In this paper, we developed a novel framework based on non-negative matrix factorization, called SMANMF, to verify the potential relationship between small molecules and miRNAs. We calculated the two types of small molecules similarity and the miRNA similarities. Meanwhile, SMANMF could also obtain the inter-relationship of small molecules with miRNAs. Moreover, the two types of small molecules similarity were considered at the same time to enhance inference on the associations of small molecules with miRNAs. The estimated association scores of small molecules with miRNAs can be obtained based on an iterative algorithm, and we ranked the candidate miRNAs by these scores for each of the small molecule. The experiment results indicated that SMANMF consistently performed better than the previous methods. Case studies on three small molecules demonstrated

the SMANMFs superior performance in discovering the potential miRNA indications. The reasons of obtaining superior performance include: First, the Non-negative Matrix Factorization model fully combines the similarity of small molecules and similarity of miRNAs, which is of great help to improve the results; Second, we consider two types of similarity of small molecules at the same time which can improve the performance and robustness of the model by fusing of more valuable data. Meanwhile, a new calculation model has been developed to predict the associations of small molecules and miRNA, which also can be reused to approximate prediction problems (e.g. small molecule-target, small molecule-disease, miRNA-disease and miRNA-target).

The evaluation result of SMANMF showed that we developed model can effectively enhance the effect compared with several state-of-the-art models. However, there were still some shortcomings that need to be improved. First, binding sequence information (such as miRNA sequence information) may contribute to the improvement of experimental performance, which is the point to be considered in this paper. Second, considering the specificity of miRNA expression in specific cancer cell lines, it may be more accurate to discover the potential associations between small molecules and miRNAs, which may be our next stage of work.

ACKNOWLEDGMENT

This work has been supported by the National Natural Science Foundation of China (Grant no. 61873089, 61572180)

REFERENCES

[1] D. P. Bartel, "Micrornas: target recognition and regulatory functions," *cell*, vol. 136, no. 2, pp. 215–233, 2009.

[2] Y. Huang, X. J. Shen, Q. Zou, S. P. Wang, S. M. Tang, and G. Z. Zhang, "Biological functions of microrna: a review," *Journal of physiology and biochemistry*, vol. 67, no. 1, pp. 129–139, 2011.

[3] X. Zeng, L. Liu, L. Lü, and Q. Zou, "Prediction of potential disease-associated microrna using structural perturbation method," *Bioinformatics*, vol. 34, no. 14, pp. 2425–2432, 2018.

[4] C.-Z. Chen, L. Li, H. F. Lodish, and D. P. Bartel, "Microrna modulate hematopoietic lineage differentiation," *science*, vol. 303, no. 5654, pp. 83–86, 2004.

[5] W. Tang, S. Wan, Z. Yang, A. E. Teschendorff, and Q. Zou, "Tumor origin detection with tissue-specific mirna and dna methylation markers," *Bioinformatics*, vol. 34, no. 3, pp. 398–406, 2017.

[6] P. Ding, J. Luo, C. Liang, J. Cai, Y. Liu, and X. Chen, "A novel group wise-based method for calculating human mirna functional similarity," *IEEE Access*, vol. 5, pp. 2364–2372, 2017.

[7] Y. Liu, J. Luo, and P. Ding, "Inferring microrna targets based on restricted boltzmann machines," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 427–436, 2019.

[8] J. Luo, P. Ding, C. Liang, and X. Chen, "Semi-supervised prediction of human mirna-disease association based on graph regularization framework in heterogeneous networks," *Neurocomputing*, vol. 294, pp. 29–38, 2018.

[9] P. Ding, J. Luo, C. Liang, Q. Xiao, and B. Cao, "Human disease mirna inference by combining target information based on heterogeneous manifolds," *Journal of biomedical informatics*, vol. 80, pp. 26–36, 2018.

[10] X. Zeng, X. Zhang, and Q. Zou, "Integrative approaches for predicting microrna function and prioritizing disease-related microrna using biological interaction networks," *Briefings in bioinformatics*, vol. 17, no. 2, pp. 193–203, 2015.

[11] C. Liang, Y. Li, J. Luo, and Z. Zhang, "A novel motif-discovery algorithm to identify co-regulatory motifs in large transcription factor and microrna co-regulatory networks in human," *Bioinformatics*, vol. 31, no. 14, pp. 2348–2355, 2015.

[12] C. Pan, J. Luo, J. Zhang, and X. Li, "Bimodule: biclique modularity strategy for identifying transcription factor and microrna co-regulatory modules," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 17, no. 1, pp. 321–326, 2019.

[13] J. Luo, C. Pan, G. Xiang, and Y. Yin, "A novel cluster-based computational method to identify mirna regulatory modules," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 2, pp. 681–687, 2018.

[14] X. Chen, L. Wang, J. Qu, N.-N. Guan, and J.-Q. Li, "Predicting mirna–disease association based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 24, pp. 4256–4265, 2018.

[15] T. J. Dougherty and M. J. Pucci, *Antibiotic discovery and development*. Springer Science & Business Media, 2011.

[16] A. P. Russ and S. Lampel, "The druggable genome: an update." *Drug discovery today*, vol. 10, no. 23-24, pp. 1607–1610, 2005.

[17] S. J. Dixon and B. R. Stockwell, "Identifying druggable disease-modifying gene products," *Current opinion in chemical biology*, vol. 13, no. 5-6, pp. 549–555, 2009.

[18] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea *et al.*, "A comprehensive map of molecular drug targets," *Nature reviews Drug discovery*, vol. 16, no. 1, p. 19, 2017.

[19] X. Chen, N.-N. Guan, Y.-Z. Sun, J.-Q. Li, and J. Qu, "Microrna-small molecule association identification: from experimental results to computational models," *Brief. Bioinform*, vol. 21, no. 1, pp. 47–61, 2018.

[20] P. d. C. Monroig, L. Chen, S. Zhang, and G. A. Calin, "Small molecule compounds targeting mirnas for cancer therapy," *Advanced drug delivery reviews*, vol. 81, pp. 104–116, 2015.

[21] A. Deiters, "Small molecule modifiers of the microrna and rna interference pathway," *The AAPS journal*, vol. 12, no. 1, pp. 51–60, 2010.

[22] G. Hutvágner, J. McLachlan, A. E. Pasquinelli, É. Bálint, T. Tuschl, and P. D. Zamore, "A cellular function for the rna-interference enzyme dicer in the maturation of the let-7 small temporal rna," *Science*, vol. 293, no. 5531, pp. 834–838, 2001.

[23] J. Wang, F. Meng, E. Dai, F. Yang, S. Wang, X. Chen, L. Yang, Y. Wang, and W. Jiang, "Identification of associations between small molecule drugs and mirnas based on functional similarity," *Oncotarget*, vol. 7, no. 25, p. 38658, 2016.

[24] C.-C. Wang, X. Chen, J. Qu, Y.-Z. Sun, and J.-Q. Li, "Rfsmma: a new computational model to identify and prioritize potential small molecule-mirna associations," *Journal of chemical information and modeling*, vol. 59, no. 4, pp. 1668–1679, 2019.

[25] S. Jamal, V. Periwal, V. Scaria, O. Consortium *et al.*, "Computational analysis and predictive modeling of small molecule modulators of microrna," *Journal of cheminformatics*, vol. 4, no. 1, p. 16, 2012.

[26] Y. Lv, S. Wang, F. Meng, L. Yang, Z. Wang, J. Wang, X. Chen, W. Jiang, Y. Li, and X. Li, "Identifying novel associations between small molecules and mirnas based on integrated molecular networks," *Bioinformatics*, vol. 31, no. 22, pp. 3638–3644, 2015.

[27] J. Li, K. Lei, Z. Wu, W. Li, G. Liu, J. Liu, F. Cheng, and Y. Tang, "Network-based identification of micrornas as potential pharmacogenomic biomarkers for anticancer drugs," *Oncotarget*, vol. 7, no. 29, p. 45584, 2016.

[28] J. Qu, X. Chen, Y.-Z. Sun, Y. Zhao, S.-B. Cai, Z. Ming, Z.-H. You, and J.-Q. Li, "In silico prediction of small molecule-mirna associations based on the hetesim algorithm," *Molecular Therapy-Nucleic Acids*, vol. 14, pp. 274–286, 2019.

[29] F. Meng, E. Dai, X. Yu, Y. Zhang, X. Chen, X. Liu, S. Wang, L. Wang, and W. Jiang, "Constructing and characterizing a bioactive small molecule and microrna association network for alzheimer's disease," *Journal of The Royal Society Interface*, vol. 11, no. 92, p. 20131057, 2014.

[30] N.-N. Guan, Y.-Z. Sun, Z. Ming, J.-Q. Li, and X. Chen, "Prediction of potential small molecule-associated micrornas using graphlet interaction," *Frontiers in pharmacology*, vol. 9, p. 1152, 2018.

[31] J. Yin, X. Chen, C.-C. Wang, Y. Zhao, and Y.-Z. Sun, "Prediction of small molecule-microrna associations by sparse learning and heterogeneous graph inference," *Molecular pharmaceutics*, vol. 16, no. 7, pp. 3157–3166, 2019.

[32] J. Qu, X. Chen, Y.-Z. Sun, J.-Q. Li, and Z. Ming, "Inferring potential small molecule–mirna association based on triple layer heterogeneous network," *Journal of cheminformatics*, vol. 10, no. 1, p. 30, 2018.

[33] C.-C. Wang and X. Chen, "A unified framework for the prediction of small molecule–microrna association based on cross-layer dependency inference on multilayered networks," *Journal of chemical information and modeling*, vol. 59, no. 12, pp. 5281–5293, 2019.

[34] Y. Zhao, X. Chen, J. Yin, and J. Qu, "Snmfsmma: using symmetric nonnegative matrix factorization and kronecker regularized least squares to predict potential small molecule-microrna association," *RNA biology*, pp. 1–11, 2019.
[35] X. Luo, M. Zhou, S. Li, Z. You, Y. Xia, and Q. Zhu, "A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 3, pp. 579–592, 2016.
[36] J. Qu, X. Chen, J. Yin, Y. Zhao, and Z.-W. Li, "Prediction of potential mirna-disease associations using matrix decomposition and label propagation," *Knowledge-Based Systems*, vol. 186, p. 104963, 2019.
[37] X. Chen, J. Yin, J. Qu, and L. Huang, "Mdhgi: Matrix decomposition and heterogeneous graph inference for mirna-disease association prediction," *PLoS computational biology*, vol. 14, no. 8, p. e1006418, 2018.
[38] X. Chen, Y.-Z. Sun, N.-N. Guan, J. Qu, Z.-A. Huang, Z.-X. Zhu, and J.-Q. Li, "Computational models for lncrna function prediction and functional similarity calculation," *Briefings in functional genomics*, vol. 18, no. 1, pp. 58–82, 2019.
[39] X. Chen, C. C. Yan, X. Zhang, and Z.-H. You, "Long non-coding rnas and complex diseases: from experimental results to computational models," *Briefings in bioinformatics*, vol. 18, no. 4, pp. 558–576, 2017.
[40] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "Nllss: predicting synergistic drug combinations based on semi-supervised learning," *PLoS computational biology*, vol. 12, no. 7, p. e1004975, 2016.
[41] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu *et al.*, "Drugbank 4.0: shedding new light on drug metabolism," *Nucleic acids research*, vol. 42, no. D1, pp. D1091–D1097, 2013.
[42] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open babel: An open chemical toolbox," *Journal of cheminformatics*, vol. 3, no. 1, p. 33, 2011.
[43] P. Willett, "Similarity-based virtual screening using 2d fingerprints," *Drug discovery today*, vol. 11, no. 23-24, pp. 1046–1053, 2006.
[44] P. Ding, R. Yin, J. Luo, and C. K. Kwoh, "Ensemble prediction of synergistic drug combinations incorporating biological, chemical, pharmacological and network knowledge," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 1336–1345, 2018.
[45] M. Pahor, E. Chrischilles, J. Guralnik, S. Brown, R. Wallace, and P. Carbonin, "Drug data coding and analysis in epidemiologic studies," *European journal of epidemiology*, vol. 10, no. 4, pp. 405–411, 1994.
[46] X. Cheng, S.-G. Zhao, X. Xiao, and K.-C. Chou, "iatc-misf: a multi-label classifier for predicting the classes of anatomical therapeutic chemicals," *Bioinformatics*, vol. 33, no. 3, pp. 341–346, 2016.
[47] F. Cheng, I. A. Kovács, and A.-L. Barabási, "Network-based prediction of drug combinations," *Nature communications*, vol. 10, no. 1, pp. 1–11, 2019.
[48] F. Cheng, W. Li, Z. Wu, X. Wang, C. Zhang, J. Li, G. Liu, and Y. Tang, "Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space," *Journal of chemical information and modeling*, vol. 53, no. 4, pp. 753–762, 2013.
[49] C.-H. Chou, N.-W. Chang, S. Shrestha, S.-D. Hsu, Y.-L. Lin, W.-H. Lee, C.-D. Yang, H.-C. Hong, T.-Y. Wei, S.-J. Tu *et al.*, "mirtarbase 2016: updates to the experimentally validated mirna-target interactions database," *Nucleic acids research*, vol. 44, no. D1, pp. D239–D247, 2015.
[50] S. Lv, Y. Li, Q. Wang, S. Ning, T. Huang, P. Wang, J. Sun, Y. Zheng, W. Liu, J. Ai *et al.*, "A novel method to quantify gene set functional association based on gene ontology," *Journal of the royal society interface*, vol. 9, no. 70, pp. 1063–1072, 2011.
[51] X. Liu, S. Wang, F. Meng, J. Wang, Y. Zhang, E. Dai, X. Yu, X. Li, and W. Jiang, "Sm2mir: a database of the experimentally validated small molecules effects on microrna expression," *Bioinformatics*, vol. 29, no. 3, pp. 409–411, 2012.
[52] C.-H. Zheng, D.-S. Huang, L. Zhang, and X.-Z. Kong, "Tumor clustering using nonnegative matrix factorization with gene selection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 599–607, 2009.
[53] D.-S. Huang and C.-H. Zheng, "Independent component analysis-based penalized discriminant method for tumor classification using gene expression data," *Bioinformatics*, vol. 22, no. 15, pp. 1855–1862, 2006.
[54] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
[55] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE transactions on cybernetics*, vol. 47, no. 11, pp. 3840–3853, 2017.
[56] L. Yuan, L. Zhu, W.-L. Guo, X. Zhou, Y. Zhang, Z. Huang, and D.-S. Huang, "Nonconvex penalty based low-rank representation and sparse regression for eqtl mapping," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 14, no. 5, pp. 1154–1164, 2017.
[57] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2030–2048, 2011.
[58] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.
[59] R. T. Rockafellar, "Lagrange multipliers and optimality," *SIAM review*, vol. 35, no. 2, pp. 183–238, 1993.
[60] F. Facchinei, C. Kanzow, and S. Sagratella, "Solving quasi-variational inequalities via their kkt conditions," *Mathematical Programming*, vol. 144, no. 1-2, pp. 369–412, 2014.
[61] 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," *Bioinformatics*, vol. 26, no. 13, pp. 1644–1650, 2010.
[62] 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," *Bioinformatics*, vol. 31, no. 11, pp. 1805–1815, 2015.
[63] K. Hajian-Tilaki, "Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation," *Caspian journal of internal medicine*, vol. 4, no. 2, p. 627, 2013.
[64] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.
[65] J. D. Allison, T. Tanavin, Y. Yang, G. Birnbaum, and U. Khalid, "Various manifestations of 5-fluorouracil cardiotoxicity: A multicenter case series and review of literature," *Cardiovascular Toxicology*, pp. 1–6, 2020.
[66] Y. Wang, W. Xu, Q. Zhang, T. Bao, H. Yang, W. Huang, and H. Tang, "Follow-up of blood glucose distribution characteristics in a health examination population in chengdu from 2010 to 2016," *Medicine*, vol. 97, no. 8, 2018.
[67] B. Zhang and X. Pan, "Rdx induces aberrant expression of micrornas in mouse brain and liver," *Environmental health perspectives*, vol. 117, no. 2, pp. 231–240, 2008.
[68] V. Y. Shin, M.-T. Siu, X. Liu, E. K. Ng, A. Kwong, and K.-M. Chu, "Mir-92 suppresses proliferation and induces apoptosis by targeting ep4/notch1 axis in gastric cancer," *Oncotarget*, vol. 9, no. 36, p. 24209, 2018.

Jiawei Luo received the bachelor´s degree, the master´s degree and Ph.D. degree in Computer Science from Hunan University (HNU), China. She is currently a Full Professor with the College of Computer Science and Electronic Engineering, Hunan University (HNU), China. Her research interests include graph theory, data mining, computational biology and bioinformatics. Prof. Luo has authored more than 120 journal and conference papers. She has often been invited as organizing member for a number of conferences including ICIC, ISBRA, ICNC, etc. She has served as reviewer for a number of journals including Bioinformatics, Briefings in Bioinformatics, Plos Computational Biology, etc.

Cong Shen is currently pursing the Ph.D. degree with the college of Computer Science and Electronic Engineering, Hunan University, China. He received the Bachelor degree in College of Information Science and Technology, Hunan Agricultural University, China, in 2018. His current research interests include pharmacogenomics, biological networks and computational systems biology.

Zihan Lai recieved the Bachelor degree in School of Computer from Central China Normal University of Computer science and technology, China, in 2019.She is working toward the Master degree in college of Information Science and Electronic Engineering, Hunan University. Her research interests include data mining and bioinformatics.

Jie Cai is currently pursing the Ph.D. degree with the college of Computer Science and Electronic Engineering, Hunan University, China. Her current research interests include construction and analysis of biological network.

Pingjian Ding received the Ph.D. degree in Computer Science from Hunan University (HNU), China. From Aug. 2017 to Sep. 2018, he was a joint Ph.D. student at Biomedical Informatics Lab (BIL), Nanyang Technological University (NTU), Singapore. In 2019, he joined University of South China (USC) as a Special-Term Professor. His research interests include construction and analysis of biological and chemical network. Dr. Ding has published more than 20 journal and conference papers. He has served as reviewer for a number of international conferences and journals including IEEE TCBB, IEEE ACCESS, BMC Cancer, etc.