Send Orders for Reprints to reprints@benthamscience.net

Current Genomics, 2019, *20*, 581-591

RESEARCH ARTICLE

IMPMD: An Integrated Method for Predicting Potential Associations Between miRNAs and Diseases

Meiqi Wu[1,#], Yingxi Yang[2,#], Hui Wang[3], Jun Ding[1], Huan Zhu[1] and Yan Xu[1,*]

[1]*Department of Information and Computer Science, University of Science and Technology Beijing, Beijing 100083, China;* [2]*Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Hong Kong, China;* [3]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China*

ARTICLE HISTORY

Received: June 25, 2019
Revised: August 07, 2019
Accepted: October 16, 2019

DOI:
10.2174/1389202920666191023090215

Abstract: ***Background*:** With the rapid development of biological research, microRNAs (miRNAs) have increasingly attracted worldwide attention. The increasing biological studies and scientific experiments have proven that miRNAs are related to the occurrence and development of a large number of key biological processes which cause complex human diseases. Thus, identifying the association between miRNAs and disease is helpful to diagnose the diseases. Although some studies have found considerable associations between miRNAs and diseases, there are still a lot of associations that need to be identified. Experimental methods to uncover miRNA-disease associations are time-consuming and expensive. Therefore, effective computational methods are urgently needed to predict new associations.

***Methodology*:** In this work, we propose an integrated method for predicting potential associations between miRNAs and diseases (IMPMD). The enhanced similarity for miRNAs is obtained by combination of functional similarity, gaussian similarity and Jaccard similarity. To diseases, it is obtained by combination of semantic similarity, gaussian similarity and Jaccard similarity. Then, we use these two enhanced similarities to construct the features and calculate cumulative score to choose robust features. Finally, the general linear regression is applied to assign weights for Support Vector Machine, K-Nearest Neighbor and Logistic Regression algorithms.

***Results*:** IMPMD obtains AUC of 0.9386 in 10-fold cross-validation, which is better than most of the previous models. To further evaluate our model, we implement IMPMD on two types of case studies for lung cancer and breast cancer. 49 (Lung Cancer) and 50 (Breast Cancer) out of the top 50 related miRNAs are validated by experimental discoveries.

***Conclusion*:** We built a software named IMPMD which can be freely downloaded from https://github.com/Sunmile/IMPMD.

Keywords: miRNA, disease, miRNA-disease associations, integrated algorithm, IMPMD, computational methods.

1. INTRODUCTION

MicroRNAs (miRNAs) are a group of short non-coding RNAs (20-25nt), which have an essential influence on the post-transcriptional level of gene expression, mainly inhibiting gene expression [1-5]. In 1993, lin-4 was the first detected miRNA in a study of the developmental time of *C. elegans* larvae [6]. Since then, thousands of miRNAs have been found in a variety of species [7, 8]. Currently, 2588 miRNAs in the human genome have been annotated [8].

More and more studies have found that miRNAs play a key role in multiple stages of biological processes, such as cell growth [9], cell death [10], cell proliferation [11], cell differentiation [12], immune reaction [13], viral infection [12], *etc*. In addition, the regulatory meaning of miRNAs is attracting more and more attention in the expression of abnormal genes. For example, many studies have confirmed that the dysregulation of the miRNAs has become a major cause of abnormal cell behavior [8]. Hence, it is no surprise that miRNAs could be associated with different kinds of diseases [14, 15]. MiR-708 affects bladder cancer progression by directly inhibiting Caspase-2 [16]. The ability of miRNA-197 and miRNA-223 to predict cardiovascular death and future burden of cardiovascular has been reported in [17]. Moreover, a novel therapeutic application of the miR-23/27/24 cluster in vascular disorders and ischemic heart disease has been illustrated [18]. It can be concluded from the above facts that the identification of disease-related miRNAs is a valuable biomedical research field, which tends to benefit the treatment, diagnosis, and prevention for a variety of clinically important diseases.

The study of predicting the potential associations between miRNAs and diseases provides a new opportunity to explore the molecular mechanism of diseases. With the development of biological technology, a large number of associations have been confirmed by traditional experiments [19, 20]. Subsequently, plenty of biological data has been collected and organized into different databases. It is feasible and necessary to develop computational models to reveal the

*Address correspondence to this author at the Department of Information and Computer Science, University of Science and Technology Beijing, Beijing 100083, China; Tel: 861062332589; E-mail: xuyan@ustb.edu.cn

[#]*These authors have contributed equally to this work*.

1389-2029/19 $58.00+.00

potential disease-miRNA associations based on these databases [21]. In recent couple of years, more and more researchers have put forward the computational methods to predict specific associations. The existing computing methods can be roughly divided into two categories, one is to build the network and apply the corresponding network-based algorithm [22-25], the other is to harness machine learning algorithm [26-29].

Based on the notion that functionally related microRNAs tend to be associated with phenotypically similar diseases, Jiang *et al.* [30] proposed a computational model by combining three similarity networks. However, the number of target genes verified by experiments is insufficient, which limits the predictor performance. By implementing random walk analysis, Shi *et al.* [31] developed a bipartite miRNA-disease network. Chen *et al.* [32] developed a method named RWRMDA by implementing random walks on miRNA functional similarity networks. However, it is unable to predict diseases without known related miRNAs. Fortunately, this shortcoming was overcome by Chen *et al.* [33], who proposed a model MIDP for miRNA-disease associations' prediction. Recently, a novel computational method Symmetric Nonnegative Matrix Factorization for MiRNA-Disease Association prediction (SNMFMDA) [34] adopted symmetric non-negative matrix factorization (SymNMF) to interpolate the integrated similarity matrix.

Furthermore, there are some models based on machine learning. For instance, utilizing semi-supervised learning, a model named Regularized Least Squares for MiRNA-Disease Association (RLSMDA) [35] was developed. By combining the advantages of similarity algorithms and machine learning methods, Chen *et al.* [36] advanced a model ELLPMDA which output the weighted combination of the ranks given by three classic similarity-based algorithms, Common Neighbors, Jaccard index and Katz index. Recently, Jaccard-similarity was implemented in Bipartite Local models and Hubness-Aware Regression for MiRNA-Disease Association prediction (BLHARMDA) [37]. Niu *et al.* [38] integrated random walk and binary regression to identify novel miRNA-disease association.

After long-term development, these classifiers solved the problem of failure to predict new disease-associated miRNAs and improve predictive performance. These predictors made it possible to predict the associations by using computational methods. However, although these models have achieved good prediction performance, the accuracy of predictors still has room for improvement. Besides, all the models are trained based on the HMDD v2.0 data. HMDD v3.0 has more than twice as much data as HMDD v2.0, which will be more conducive to the predictor performance. In this work, we utilized HMDD v3.0 data to build a predictor IMPMD. Firstly, according to the previous studies, an important hypothesis is that miRNAs with similar functions are more likely to be associated with phenotypically similar diseases [36, 39-41]. In other words, miRNAs with similar functions may be associated with the same disease. Thus, we constructed an enhanced similarity representation for miRNAs based on the functional similarity, gaussian similarity and Jaccard similarity. For diseases, the enhanced similarity representation was obtained by integrating semantic similarity, gaussian similarity and Jaccard similarity. Then, based on the two similarity representations, the feature vector for every miRNA-disease pair could be extracted. We assigned the known experimentally verified miRNA-disease associations as positive samples. To optimize the model, we extracted 200 robust features from the whole features. Finally, linear regression is integrated with Support Vector Machine (SVM), K-Nearest Neighbor (KNN) as well as Logistic Regression (LR) to predict the unknown associations (Fig. **1**). For the convenience of researches, IMPMD can be freely downloaded from https://github.com/Sunmile/IMPMD.

2. RESULTS

2.1. Performance

To evaluate the performance of our predictor, 10-fold cross validation is applied. We randomly selected the unknown miRNA-disease pairs with the same number of known miRNA-disease associations as the negative samples for SVM, KNN and LR. In total, there are 33974 samples in the training dataset. And all the known miRNA-disease associations are regarded as positive samples. Afterwards, we divide the training set into ten subsets on average, and each subset take turns as the test data. The relevance score of each sample in the testing dataset is calculated. If a positive sample has higher score than a given threshold, it is regarded as a successfully identified positive sample. If a negative sample has a score lower than threshold, it is regarded as a correctly identified negative sample. By calculating True Positive Rate (TPR) and False Positive Rate (FPR), we get the Receiver Operating Characteristic (ROC) curve and its Area Under Curve (AUC) value. Besides, the common performance measures accuracy (ACC), sensitivity (Sen), specificity (Spe) and Matthews Correlation Coefficient (MCC) are carried out to evaluate the performance of our predictor.

As shown in Figs. (**2** and **3**), the predictor using SVM algorithm has the best prediction performance, followed by LR, and finally KNN. Their AUC values all exceed 0.9. Moreover, the value of ACC, Sen and Spe are all over 0.8, indicating the efficiency of the predictors. In order to get a more efficient predictor, we used the outputs of the three predictors as new features, and assigned weights for the three features by linear regression to generate the predictor IMPMD. It could be seen from Figs. (**2D** and **3**) that the prediction performance of the comprehensive predictor IMPMD has improved. The value of AUC is 0.9386, which demonstrates the reliable and efficient performance of IMPMD. To illustrate the performance of IMPMD, we compare it with several existing methods: RWBRMDA [38], SNMFMDA [34], NSEMDA [42] and MCMDA [43]. It is obvious that the AUC value of IMPMD is higher than the other models (Table **1**), which proves the superiority of IMPMD. In addition, the potential associations predicted by IMPMD are listed in the Supplementary Table **1** for conducting future experimental validation. The predicted miRNA-disease associations help guide researchers to selectively identify disease-associated miRNAs without blindly identifying whether all miRNAs are associated with a certain disease.

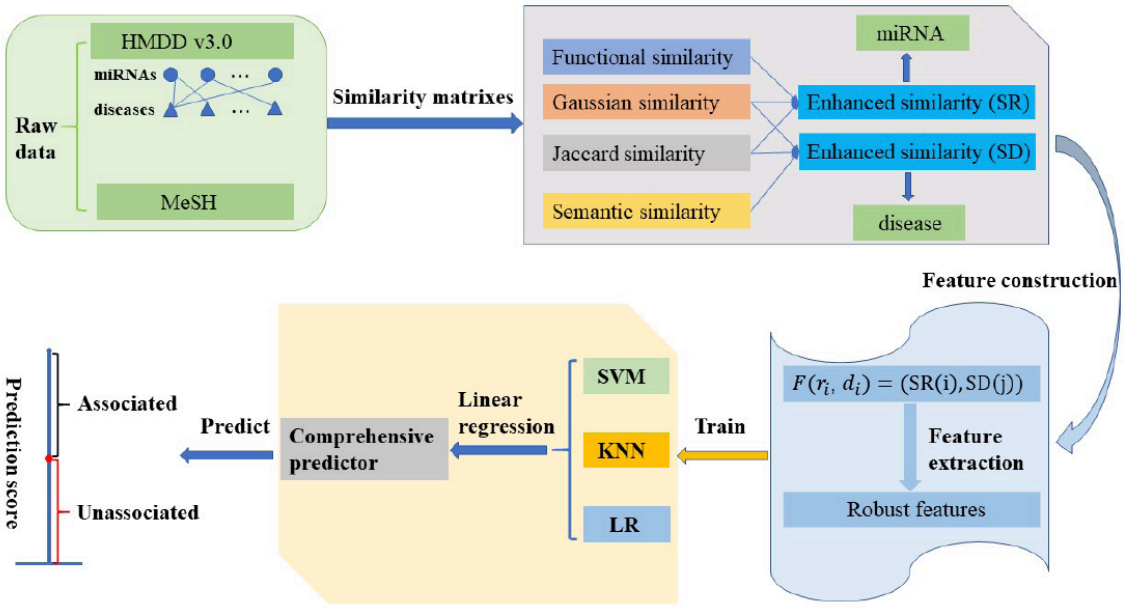

Fig. (1). The computational framework of the predictor. Step 1, enhanced similarity matrixes of disease and miRNA are obtained based on HMDD v3.0 database and MeSH. Step 2, feature vectors are constructed and extracted for miRNA-disease associations. Step 3, train the predictors obtained by SVM, KNN and LR and integrated them by linear regression. Step 4, the comprehensive predictor IMPMD is applied to predict the potential miRNA-disease associations. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

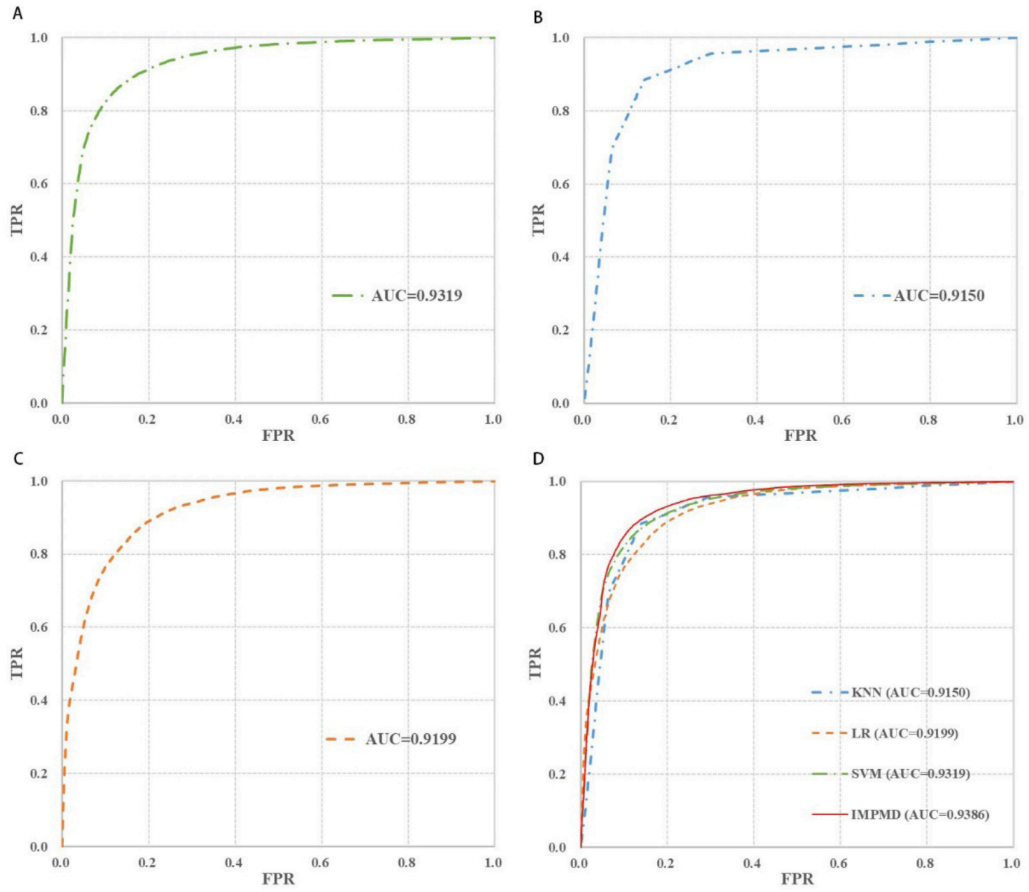

Fig. (2). The ROC curves of different algorithms based on 10-fold cross validation. (**A**) ROC curve of SVM. (**B**) ROC curve of KNN. (**C**) ROC curve of LR. (**D**) ROC curves of comprehensive predictor. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

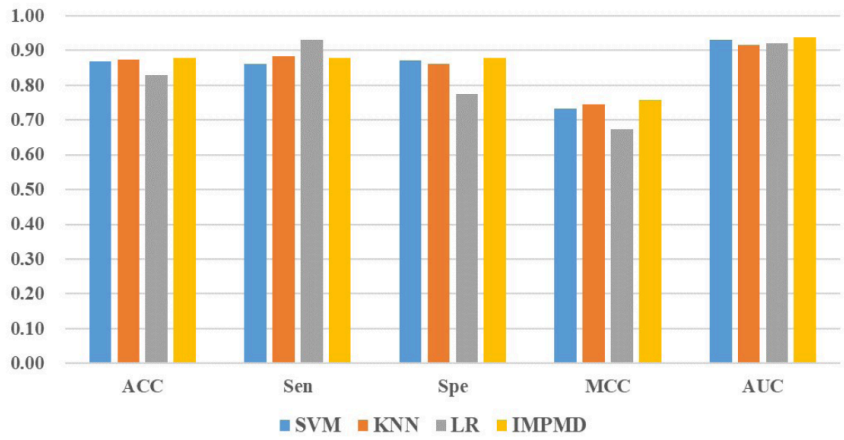

Fig. (3). Performance measures of the predictors trained by different algorithms. Four different algorithms are used to build models, SVM, KNN, LR and comprehensive algorithm, respectively. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table 1. The comparison result of IMPMD with other predictors.

Predictor	RWBRMDA	SNMFMDA	NSEMDA	MCMDA	IMPMD
AUC	0.8076	0.9007	0.8899	0.8749	**0.9386**

2.2. Case Studies

To further show the effectiveness of predictor IMPMD, two types of case study on two common diseases (lung cancer and breast cancer) are implemented. Three databases dbDEMC 2.0 [44], PhenomiR [45] and miRwayDB [46] are used to validate the predicted results. The dbDEMC 2.0 database records a large number of differentially expressed miRNAs in various cancers through high-throughput methods. PhenomiR database and miRwayDB database utilize the data collected from published literatures to provide differentially regulated miRNA expression in diseases and other biological processes or pathways in various pathophysiological conditions.

Two types of case study are implemented. One is to predict the investigated disease related miRNAs based on all known miRNA-disease associations, and the other is based on removing all known miRNA-disease associations that contain the disease under investigation. For lung cancer, we adopt the first method. Specifically, the lung cancer-miRNA associations are predicted by IMPMD based on the known disease-miRNA associations recorded in HMDD v3.0 database, and then verified by the records in the other three databases. Lung cancer is the leading cause of cancer deaths with an estimated 1.4 million deaths each year [47]. And there are numerous reports of significant increases in lung cancer morbidity and mortality [48]. The gold standard for lung cancer grading is routine histopathology, but it has limitations [49, 50]. Such as, when a small number of diagnostic sample is available, it is difficult to obtain adequate number of tumor cells with desired tissue architecture [51, 52]. Therefore, a more reliable lung cancer screening and diagnosis tool is needed. The miRNA expression profiles are critical to understand the regulation of gene expression in lung cancer cells, which in turn will contribute to the identification of homologous therapeutic targets and therapeutic measures [53]. The identification of miRNAs that is associated with lung cancer is significant. In this work, we predict the potential lung cancer-miRNA associations by IMPMD, and the results show that 49 out of the top 50 potential associated miRNAs are validated by dbDEMC 2.0, PhenomiR and miRwayDB (Table **2**). The prognostic value of hsa-mir-200b (1^{st}) in squamous cell lung cancer has been reported [54]. Lu *et al.* [55] performed a systematic expression analysis of 217 mammalian miRNAs from 334 samples (including multiple human cancers) by a new, bead-based flow cytometry miRNA expression profiling method. A variety of miRNAs have been confirmed to have associations with lung cancer, including hsa-mir-200a (2^{nd}), hsa-let-7a (4^{th}), *etc*.

We implement another type of case study for breast cancer. The training data still come from HMDD v3.0, but all known associations containing breast cancer are regarded as unlabeled pairs. Then training the model, ranking all candidate miRNAs for breast cancer and verifying the predicted results based on HMDD v3.0, dbDEMC 2.0, PhenomiR and miRwayDB database are executed. Breast cancer remains the second leading cause of cancer death after lung cancer [56]. It is the most frequently diagnosed cancer and the leading cause of cancer death among females, accounting for 23% of the total cancer cases and 14% of the cancer deaths [57]. Some studies on miRNA have shown that loss of the tumor suppressor miRNA or overexpression of the oncogenic miRNA can lead to the onset or metastasis of breast cancer [58-60]. MiRNA has a promising prospect in guiding the diagnosis of breast cancer. So, we put forward to predict the potential related miRNAs for breast cancer. As shown in Table **3**, all the top 50 predicted miRNAs are verified by HMDD v3.0, dbDEMC, PhenomiR, and miRwayDB. In the

Table 2. The first 50 potential miRNAs associated with lung cancer predicted by IMPMD.

miRNAs	Databases	miRNAs	Databases
hsa-mir-200b	dbDEMC, PhenomiR	hsa-mir-122	dbDEMC, PhenomiR, miRwayDB
hsa-mir-200a	dbDEMC, PhenomiR	hsa-mir-132	dbDEMC, PhenomiR
hsa-mir-29b	dbDEMC	hsa-mir-29b-1	PhenomiR
hsa-let-7a	dbDEMC	hsa-mir-486	dbDEMC, PhenomiR
hsa-mir-148a	dbDEMC, PhenomiR	hsa-let-7i	dbDEMC, PhenomiR
hsa-mir-25	dbDEMC, PhenomiR	hsa-mir-139	dbDEMC, PhenomiR
hsa-mir-145	dbDEMC, PhenomiR, miRwayDB	hsa-let-7e	dbDEMC, PhenomiR
hsa-mir-106b	dbDEMC, PhenomiR	hsa-mir-503	dbDEMC
hsa-mir-222	dbDEMC, PhenomiR	hsa-mir-29b-2	PhenomiR
hsa-mir-34a	dbDEMC, PhenomiR, miRwayDB	hsa-mir-127	PhenomiR
hsa-mir-125b	dbDEMC	hsa-mir-574	PhenomiR
hsa-mir-15a	dbDEMC, PhenomiR, miRwayDB	hsa-mir-99b	dbDEMC, PhenomiR
hsa-mir-9	dbDEMC	hsa-mir-328	dbDEMC, PhenomiR
hsa-mir-126	dbDEMC, PhenomiR, miRwayDB	hsa-mir-326	dbDEMC, PhenomiR
hsa-mir-96	dbDEMC, PhenomiR	hsa-mir-181c	dbDEMC, PhenomiR
hsa-mir-494	dbDEMC, miRwayDB	hsa-mir-367	dbDEMC, PhenomiR
hsa-mir-182	dbDEMC, PhenomiR	hsa-mir-590	PhenomiR
hsa-mir-199a	dbDEMC	hsa-mir-187	dbDEMC, PhenomiR
hsa-mir-181a	dbDEMC	hsa-mir-331	dbDEMC, PhenomiR
hsa-mir-150	dbDEMC, PhenomiR	hsa-mir-378a	unconfirmed
hsa-mir-223	dbDEMC, PhenomiR, miRwayDB	hsa-mir-372	dbDEMC, PhenomiR
hsa-let-7d	dbDEMC, PhenomiR	hsa-mir-198	dbDEMC, PhenomiR
hsa-mir-92a	dbDEMC	hsa-mir-217	dbDEMC, PhenomiR
hsa-mir-206	dbDEMC, PhenomiR, miRwayDB	hsa-mir-129-2	PhenomiR
hsa-let-7c	dbDEMC, PhenomiR	hsa-mir-488	dbDEMC

reference [61], nine dysregulated miRNAs including seven up-regulated and two down-regulated miRNAs served as candidate diagnostic markers for breast cancer. And seven upregulated miRNAs contain hsa-mir-200a (2^{nd}) and hsa-mir-141 (50^{th}), and two downregulated miRNAs contain hsa-mir-145 (3^{rd}).

3. DISCUSSION

As evidenced by numerous literatures, miRNAs are involved in many diseases and play an important role in the prevention, diagnosis and treatment of diseases. The identification of new miRNA-disease associations has significant biological implications and plays a key role in understanding the pathogenesis of diseases at the miRNA level. However, experimental methods are time-consuming and expensive, computational methods are good choice to solve the issue and predict potential miRNA-disease associations with high accuracy. In this work, we combine three different similarities of diseases and miRNAs to construct features of miRNA-disease associations and select 200 robust features by feature extraction. In addition, we utilize linear regression to integrate three different classification algorithms, namely SVM, KNN, and LR to obtain the comprehensive predictor IMPMD which performs better than the classifiers obtained by a single algorithm. The experimental results show that IMPMD with AUC of 0.9386 in 10-fold cross-validation which is superior to other predictors. To further validate the effectiveness, case studies of lung cancer and breast cancer are conducted. In the case study, two different methods are implemented. One is to predict disease-related miRNA for investigated diseases based on all known miRNA-disease associations, the other is to hide investigated diseases related miRNA. And the majority of the top 50 potential miRNAs inferred by IMPMD are verified by several databases.

Table 3. The first 50 potential miRNAs associated with breast cancer predicted by IMPMD.

miRNAs	Databases	miRNAs	Databases
hsa-mir-200b	dbDEMC, PhenomiR, miRwayDB, HMDD	hsa-mir-205	dbDEMC, PhenomiR, HMDD
hsa-mir-200a	dbDEMC, PhenomiR, miRwayDB, HMDD	hsa-mir-375	dbDEMC, PhenomiR, HMDD
hsa-mir-145	dbDEMC, PhenomiR, HMDD	hsa-mir-29a	dbDEMC, PhenomiR, miRwayDB, HMDD
hsa-mir-143	dbDEMC, PhenomiR, HMDD	hsa-mir-218	dbDEMC, HMDD
hsa-mir-29b	dbDEMC, HMDD	hsa-mir-101	dbDEMC, miRwayDB, HMDD
hsa-let-7a	dbDEMC, HMDD	hsa-mir-30a	dbDEMC, PhenomiR, HMDD
hsa-mir-224	dbDEMC, PhenomiR, HMDD	hsa-mir-27b	dbDEMC, PhenomiR, HMDD
hsa-mir-203	dbDEMC, PhenomiR, HMDD	hsa-mir-23b	dbDEMC, PhenomiR, HMDD
hsa-mir-29c	dbDEMC, PhenomiR, HMDD	hsa-mir-100	dbDEMC, PhenomiR, miRwayDB, HMDD
hsa-mir-34c	dbDEMC, HMDD	hsa-mir-222	dbDEMC, PhenomiR, HMDD
hsa-mir-195	dbDEMC, PhenomiR, HMDD	hsa-mir-107	dbDEMC, PhenomiR, HMDD
hsa-mir-18a	dbDEMC, PhenomiR, HMDD	hsa-mir-23a	dbDEMC, PhenomiR, HMDD
hsa-mir-25	dbDEMC, PhenomiR, HMDD	hsa-mir-378	dbDEMC, PhenomiR, miRwayDB, HMDD
hsa-mir-16	dbDEMC, HMDD	hsa-mir-106b	dbDEMC, PhenomiR, HMDD
hsa-mir-138	dbDEMC, HMDD	hsa-mir-133b	dbDEMC, PhenomiR, HMDD
hsa-mir-200c	dbDEMC, PhenomiR, HMDD	hsa-mir-127	dbDEMC, PhenomiR, HMDD
hsa-mir-221	dbDEMC, PhenomiR, HMDD	hsa-mir-99a	dbDEMC, PhenomiR, miRwayDB, HMDD
hsa-mir-22	dbDEMC, PhenomiR, miRwayDB, HMDD	hsa-mir-218-1	PhenomiR, HMDD
hsa-mir-31	dbDEMC, PhenomiR, HMDD	hsa-mir-218-2	PhenomiR, HMDD
hsa-mir-17	dbDEMC, PhenomiR, HMDD	hsa-mir-497	dbDEMC, PhenomiR, miRwayDB, HMDD
hsa-mir-148a	dbDEMC, PhenomiR, miRwayDB, HMDD	hsa-mir-125b	dbDEMC, miRwayDB, HMDD
hsa-mir-26a	dbDEMC, HMDD	hsa-let-7g	dbDEMC, PhenomiR, HMDD
hsa-mir-34a	dbDEMC, PhenomiR, miRwayDB, HMDD	hsa-mir-106a	dbDEMC, PhenomiR, HMDD
hsa-mir-7	dbDEMC, miRwayDB, HMDD	hsa-let-7b	dbDEMC, PhenomiR, HMDD
hsa-mir-20a	dbDEMC, PhenomiR, HMDD	hsa-mir-141	dbDEMC, PhenomiR, HMDD

The performance and reliability of IMPMD can be attributed to three main factors. First, we adopt HMDD v3.0 database, which has more than twice as much data as HMDD2. Second, IMPMD generalizes various similarity networks, including miRNA functional similarity, disease semantic similarity, Gaussian interaction profile kernel similarity of miRNAs and diseases, and Jaccard similarity of miRNAs and diseases. In addition, the feature selection method is implemented to reduce the dimension of the features and select the robust features. These various features could reflect different intrinsic aspects of miRNA-disease associations. Last, IMPMD synthesizes different classification algorithms by linear regression to improve the prediction performance. Nevertheless, there are some disadvantage of IMPMD, such as, for a new disease, it takes a complex calculation to convert it into a vector. Thus, there is still a room for improvement. For instance, better calculation methods and more biological information can be expected to represent the association between miRNAs and diseases. Besides, algorithms that enable deeply mine data features can be used to build models. In future work, we will consider deep learning framework to study the association of miRNA-disease. There are only 16987 known associations between 850 diseases and 1102 miRNAs, and the number of unknown miRNA-disease pairs is more than 50 times than that of known associations. We believe the performance will be improved with the existence of more and more experimental verified miRNA-disease associations.

4. MATERIALS AND METHODS

4.1. Human miRNA-disease Associations

Through continuous accumulation of biological experiments, HMDD (Human MicroRNA Disease Database) has

evolved into a database of considerable data. We extract 16,987 miRNA-disease associations between 1102 miRNAs and 850 diseases without duplicates from HMDD v3.0 database (http://www.cuilab.cn/hmdd). For the convenience of readers, all the diseases, miRNAs and miRNA-disease associations are shown in Supplementary Table **2**. According to these data we establish an adjacency matrix $A_{1102\times850}$ (Supplementary Table **3**). Element $A_{(i,j)}$ is set to be 1 if miRNA r_i and diseases d_j has association. Otherwise, it is equal to 0.

4.2. MiRNA Functional Similarity

Based on the assumption that functionally similar miRNAs tend to be associated with phenotypically similar diseases, Wang *et al.* [40] advanced MISIM method to calculate the functional similarity between miRNA pairs. As shown in Fig. (**4**), the calculation of miRNA functional similarity is divided into four steps. The detailed calculation procedure is explained in the Supplementary **S4**. By calculating the functional similarity of miRNAs, we obtain a matrix FR.

4.3. Disease Semantic Similarity

According to [36, 40], we downloaded the MeSH descriptors from the National Library of Medicine (http://www.nlm.nih.gov/) and the information of descriptors is shown in Supplementary Table **5**. We construct DAG (Directed Acyclic Graph) to describe the diseases (Fig. **5**). Disease D can be denoted as DAG(D)= (D, T(D), E(D)), where T(D) represents a set of the node D itself and its ancestor nodes, E(D) stands for the edges between parent and child nodes. Afterwards, we use two methods to calculate the semantic similarity of diseases, the detailed process is shown in the Supplementary file **S4**.

4.4. Gaussian Interaction Profile Kernel Similarity for miRNAs and Diseases

As explanation by van Laarhoven *et al.* [62], similar diseases tend to be related to miRNAs with similar functions, the Gaussian interaction profile kernel similarity between miRNA r_i and miRNA r_j is constructed. We denote the i-th and the j-th row vector of the matrix A as $VR(r_i)$ and $VR(r_j)$, respectively. Then, the Gaussian interaction profile kernel similarity of miRNAs could be computed as (Eq. 1):

$$GR(r_i, r_j) = \exp(-\beta_r \, ||VR(r_i) - VR(r_j)||^2) \quad (1)$$

where β_r is used to control the kernel bandwidth which is defined by normalizing a bandwidth parameter β_r' divided by the average number of diseases associated with each miRNA (Eq. 2).

$$\beta_r = \frac{\beta_r'}{\frac{1}{nr}\sum_{u=1}^{nr} ||VR(r_u)||} \quad (2)$$

where nr is the number of miRNAs, β_r' is defined as 1 (Chen and Yan (2013) [63].

For the diseases, the Gaussian interaction profile kernel similarity between disease d_i and disease d_j is constructed as the same as the miRNAs (Eq. 3).

$$GD(d_i, d_j) = \exp(-\beta_d \, ||VD(d_i) - VD(d_j)||^2) \quad (3)$$

where binary interaction profile vectors $VD(d_i)$ and $VD(d_j)$ are defined as the i-th and the j-th column vector of the matrix A.

$$\beta_d = \frac{\beta_d'}{\frac{1}{nd}\sum_{u=1}^{nd} ||VD(d_u)||} \quad (4)$$

where nd is the number of diseases, β_d' is defined as 1 (Eq. 4) [63].

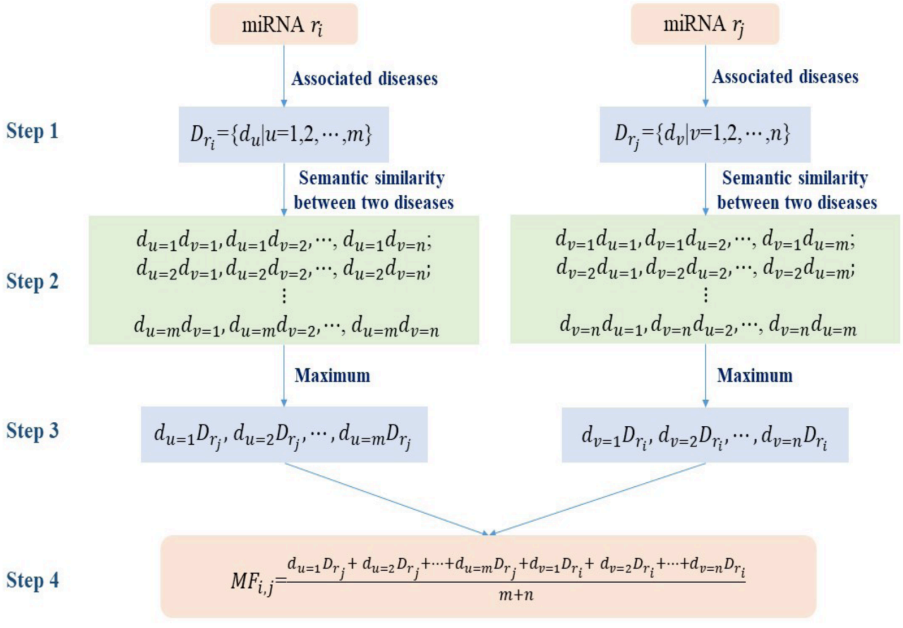

Fig. (4). Computational flow chart of miRNA functional similarity. Step 1, find out the diseases set (diseases associated with miRNA r_i and diseases associated with miRNA r_j, respectively). Step 2, calculate the semantic similarity between each disease in one disease set and each disease in another disease set. Step 3, find out the max semantic similarity for every disease. Step 4, calculate the functional similarity. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

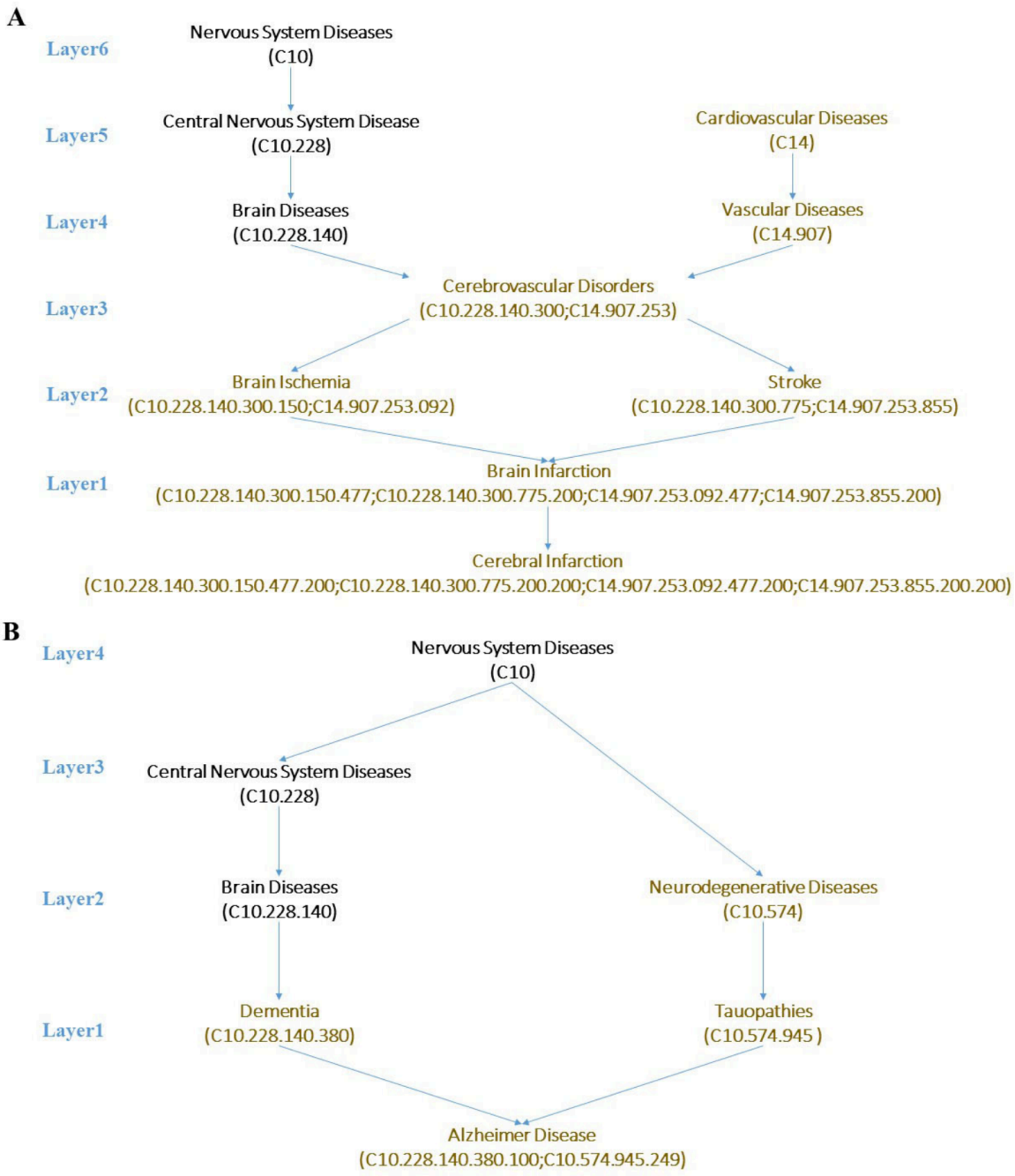

Fig. (5). The disease DAGs of Cerebral Infarction and Alzheimer Disease. (**A**) The addresses of Cerebral Infarction and its ancestor nodes. (**B**) The addresses of Alzheimer Disease and its ancestor nodes. The nodes with bold black font represent the common nodes of the two DAGs. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

4.5. Integrated Similarity for miRNAs and Diseases

As mentioned above, disease semantic similarity is calculated *via* DAG. However, some diseases are calculated without DAG. We combine the disease functional similarity with disease Gaussian interaction profile kernel similarity to calculate the integrated disease similarity. The formulations are shown as follows (Eq. 5):

$$SD(d_i,d_j) = \begin{cases} \dfrac{SD1(d_i,d_j)+SD2(d_i,d_j)}{2}, & d_i \text{ and } d_j \textit{ have their own DAG} \\ GD(d_i,d_j), & \textit{otherwise} \end{cases} \quad (5)$$

Furthermore, in the same way, we obtain the integrated miRNA similarity as follows (Eq. 6):

$$SR(r_i,r_j) = \begin{cases} FR(r_i,r_j), & r_i \text{ and } r_j \textit{ have functional similarity} \\ GR(r_i,r_j), & \textit{otherwise} \end{cases}. \quad (6)$$

4.6. Jaccard-similarity for miRNAs and Diseases

The Gaussian interaction profile kernel similarity is calculated based on the known miRNA-disease associations. In order to further exploit the information of known associa-

tions, we adopt Jaccard-similarity (Chen *et al.* 2018b). The Jaccard-similarity between miRNA r_i and miRNA r_j can be computed as following (7):

$$JR(r_i, r_j) = \frac{|AD(r_i) \cap AD(r_j)|}{|AD(r_i) \cup AD(r_j)|} \quad (7)$$

where $AD(r_i)$ and $AD(r_i)$ represent the number of diseases set associated with miRNA r_i and miRNA r_j, respectively.

Similar to the miRNAs, the Jaccard-similarity between disease d_i and disease d_j can be computed by the following formula (8).

$$JD(d_i, d_j) = \frac{|AR(d_i) \cap AR(d_j)|}{|AR(d_i) \cup AR(d_j)|} \quad (8)$$

where $AR(d_i)$ denotes the number of miRNAs set that is associated with disease d_i.

4.7. Enhanced Similarity-based Representation for miRNAs and Diseases

After the above steps, we obtain the integrated similarity and the Jaccard-similarity for miRNAs and diseases. By combining the integrated similarity matrix and Jaccard similarity matrix, an enhanced matrix can be obtained. Finally, we expand SR (Supplementary Table **6**) into a $nr \times 2nr$ (nr is the number of miRNAs) matrix, and SD (Supplementary Table **7**) into a matrix of $nd \times 2nd$ (nd is the number of diseases). The left $nr \times nr$ (or $nd \times nd$) square matrix is the integral similarity matrix, while the right $nr \times nr$ (or $nd \times nd$) square matrix is the Jaccard-similarity matrix.

4.8. Integrated Method

Based on the two matrices *SR* and *SD*, the features of miRNA-disease associations are constructed. Firstly, we randomly select the unknown miRNA-disease pairs with the same number as the known associations. Meanwhile, all the known miRNA-disease associations are regarded as positive samples. Then, each miRNA-disease pair can be represented by a vector *via* enhanced similarity representation for miRNAs and diseases. We define the feature of a miRNA-disease pair (r_i, d_j) as (Eq. 9):

$$F(r_i, d_j) = \left(SR(r_i), SD(d_j)\right) \quad (9)$$

where $SR(r_i)$ represents the *i*-th row of SR, $SD(d_i)$ is the *j*-th row of SD.

To reduce the training time and efficiently distinguish related miRNA-disease pairs from unrelated miRNA-disease pairs, we choose 200 robust features among 3904 features. For each feature f, we calculate its cumulative score in positive dataset (*P*) and negative dataset (*N*) as follows (Eq. 10):

$$cssum(f, P) = \sum_{P_i \in P} cs(P_i, f) \quad (10)$$

where $cs(P_i, f)$ is the feature value of the *i*-th sample in *P*. For the negative dataset, the $cssum(f, N)$ could be calculated in the same way. Then, we calculate the specificity in *P* and *N* as follows (11):

$$ss(f) = \left(cssum(f, P) + cssum(f, N)\right) \times log\left(\frac{|P|}{cssum(f, P)} + \frac{|N|}{cssum(f, N)}\right). \quad (11)$$

When one of $cssum(f, P)$ and $cssum(f, N)$ is small and the other is large, then the value of $ss(f)$ is large. Conversely, when they are both large or small, the value of $ss(f)$ will be lower. We choose 100 features from *SR* and 100 features from *SD* through sorting the ss values.

CONCLUSION

Afterwards, we train three models based on SVM, KNN and LR, respectively by implementing 10-fold cross validation. Finally, to improve prediction performance, we adopt linear regression method to assign weights for these three models to obtain a comprehensive predictor.

AUTHOR'S CONTRIBUTIONS

Y.X and Y.Y conceived and designed the experiments. M.W and H.W performed the experiments and data analysis. M.W and Y.X wrote the paper. J.D and Y.Y developed the webserver. Y.X, H.Z and Y.Y revised the manuscript. All the authors read and agreed with the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the Human MicroRNA Disease Database at http://www.cuilab.cn/hmdd, and the National Library of Medicine at http://www.nlm.nih.gov/, reference number [51].

FUNDING

This work was supported by grants from the Natural Science Foundation of China (11671032).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Ambros, V. The functions of animal microRNAs. *Nature,* **2004**, *431*(7006), 350-355. http://dx.doi.org/10.1038/nature02871 PMID: 15372042

[2] Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell,* **2004**, *116*(2), 281-297. http://dx.doi.org/10.1016/S0092-8674(04)00045-5 PMID: 14744438

[3] Meister, G.; Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature,* **2004**, *431*(7006), 343-349. http://dx.doi.org/10.1038/nature02873 PMID: 15372041

[4] Ambros, V. MicroRNAs: tiny regulators with great potential. *Cell,* **2001**, *107*(7), 823-826. http://dx.doi.org/10.1016/S0092-8674(01)00616-X PMID: 11779458

[5] Chen, X.; Gong, Y.; Zhang, D.H.; You, Z.H.; Li, Z.W. DRMDA: deep representations-based miRNA-disease association prediction. *J. Cell. Mol. Med.,* **2018**, *22*(1), 472-485. http://dx.doi.org/10.1111/jcmm.13336 PMID: 28857494

[6] Lee, R.C.; Feinbaum, R.L.; Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell,* **1993**, *75*(5), 843-854. http://dx.doi.org/10.1016/0092-8674(93)90529-Y PMID: 8252621

[7] Jopling, C.L.; Yi, M.; Lancaster, A.M.; Lemon, S.M.; Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science,* **2005**, *309*(5740), 1577-1581. http://dx.doi.org/10.1126/science.1113329 PMID: 16141076

[8] Kozomara, A.; Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.,* **2011**, *39*(Database issue), D152-D157. http://dx.doi.org/10.1093/nar/gkq1027 PMID: 21037258

[9] Ambros, V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell,* **2003**, *113*(6), 673-676. http://dx.doi.org/10.1016/S0092-8674(03)00428-8 PMID: 12809598

[10] Xu, P.; Guo, M.; Hay, B.A. MicroRNAs and the regulation of cell death. *Trends Genet.,* **2004**, *20*(12), 617-624. http://dx.doi.org/10.1016/j.tig.2004.09.010 PMID: 15522457

[11] Cheng, A.M.; Byrom, M.W.; Shelton, J.; Ford, L.P. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.,* **2005**, *33*(4), 1290-1297. http://dx.doi.org/10.1093/nar/gki200 PMID: 15741182

[12] Miska, E.A. How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.,* **2005**, *15*(5), 563-568. http://dx.doi.org/10.1016/j.gde.2005.08.005 PMID: 16099643

[13] Taganov, K.D.; Boldin, M.P.; Chang, K.J.; Baltimore, D. NF-kappaB-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc. Natl. Acad. Sci. USA,* **2006**, *103*(33), 12481-12486. http://dx.doi.org/10.1073/pnas.0605298103 PMID: 16885212

[14] Calin, G.A.; Dumitru, C.D.; Shimizu, M.; Bichi, R.; Zupo, S.; Noch, E.; Aldler, H.; Rattan, S.; Keating, M.; Rai, K.; Rassenti, L.; Kipps, T.; Negrini, M.; Bullrich, F.; Croce, C.M. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA,* **2002**, *99*(24), 15524-15529. http://dx.doi.org/10.1073/pnas.242606799 PMID: 12434020

[15] Chen, X.; Xie, D.; Zhao, Q.; You, Z.H. MicroRNAs and complex diseases: from experimental results to computational models. *Brief. Bioinform.,* **2019**, *20*(2), 515-539. http://dx.doi.org/10.1093/bib/bbx130 PMID: 29045685

[16] Song, T.; Zhang, X.; Zhang, L.; Dong, J.; Cai, W.; Gao, J.; Hong, B. miR-708 promotes the development of bladder carcinoma *via* direct repression of Caspase-2. *J. Cancer Res. Clin. Oncol.,* **2013**, *139*(7), 1189-1198. http://dx.doi.org/10.1007/s00432-013-1392-6 PMID: 23568547

[17] Schulte, C.; Molz, S.; Appelbaum, S.; Karakas, M.; Ojeda, F.; Lau, D.M.; Hartmann, T.; Lackner, K.J.; Westermann, D.; Schnabel, R.B.; Blankenberg, S.; Zeller, T. miRNA-197 and miRNA-223 predict cardiovascular death in a cohort of patients with symptomatic coronary artery disease. *PLoS One,* **2015**, *10*(12), e0145930. http://dx.doi.org/10.1371/journal.pone.0145930 PMID: 26720041

[18] Bang, C.; Fiedler, J.; Thum, T. Cardiovascular importance of the microRNA-23/27/24 family. *Microcirculation,* **2012**, *19*(3), 208-214. http://dx.doi.org/10.1111/j.1549-8719.2011.00153.x PMID: 22136461

[19] Mohammadi-Yeganeh, S.; Paryan, M.; Mirab Samiee, S.; Soleimani, M.; Arefian, E.; Azadmanesh, K.; Mostafavi, E.; Mahdian, R.; Karimipoor, M. Development of a robust, low cost stem-loop real-time quantification PCR technique for miRNA expression analysis. *Mol. Biol. Rep.,* **2013**, *40*(5), 3665-3674. http://dx.doi.org/10.1007/s11033-012-2442-x PMID: 23307300

[20] Thomson, J.M.; Parker, J.S.; Hammond, S.M. Microarray analysis of miRNA gene expression. *Methods Enzymol.,* **2007**, *427*, 107-122. http://dx.doi.org/10.1016/S0076-6879(07)27006-5 PMID: 17720481

[21] Chen, X. Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci. Rep.,* **2015**, *5*, 13186. http://dx.doi.org/10.1038/srep13186 PMID: 26278472

[22] Chen, X.; Wang, L.; Qu, J.; Guan, N.N.; Li, J.Q. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics,* **2018**, *34*(24), 4256-4265. http://dx.doi.org/10.1093/bioinformatics/bty503 PMID: 29939227

[23] Chen, X.; Yin, J.; Qu, J.; Huang, L. MDHGI: Matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLOS Comput. Biol.,* **2018**, *14*(8), e1006418. http://dx.doi.org/10.1371/journal.pcbi.1006418 PMID: 30142158

[24] Chen, X.; Huang, L. LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for MiRNA-Disease Association prediction. *PLOS Comput. Biol.,* **2017**, *13*(12), e1005912. http://dx.doi.org/10.1371/journal.pcbi.1005912 PMID: 29253885

[25] You, Z.H.; Huang, Z.A.; Zhu, Z.; Yan, G.Y.; Li, Z.W.; Wen, Z.; Chen, X. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *PLOS Comput. Biol.,* **2017**, *13*(3), e1005455. http://dx.doi.org/10.1371/journal.pcbi.1005455 PMID: 28339468

[26] Chen, X.; Wang, L.Y.; Huang, L. NDAMDA: Network distance analysis for MiRNA-disease association prediction. *J. Cell. Mol. Med.,* **2018**, *22*(5), 2884-2895. http://dx.doi.org/10.1111/jcmm.13583 PMID: 29532987

[27] Chen, X.; Xie, D.; Wang, L.; Zhao, Q.; You, Z.H.; Liu, H. BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics,* **2018**, *34*(18), 3178-3186. http://dx.doi.org/10.1093/bioinformatics/bty333 PMID: 29701758

[28] Chen, X.; Huang, L.; Xie, D.; Zhao, Q. EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. *Cell Death Dis.,* **2018**, *9*(1), 3. http://dx.doi.org/10.1038/s41419-017-0003-x PMID: 29305594

[29] Zhao, Y.; Chen, X.; Yin, J. Adaptive boosting-based computational model for predicting potential miRNA-disease associations. *Bioinformatics,* **2019**, *35*(22), 4730-4738. http://dx.doi.org/10.1093/bioinformatics/btz297 PMID: 31038664

[30] Jiang, Q.; Hao, Y.; Wang, G.; Juan, L.; Zhang, T.; Teng, M.; Liu, Y.; Wang, Y. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.,* **2010**, *4*(Suppl. 1), S2. http://dx.doi.org/10.1186/1752-0509-4-S1-S2 PMID: 20522252

[31] Shi, H.; Xu, J.; Zhang, G.; Xu, L.; Li, C.; Wang, L.; Zhao, Z.; Jiang, W.; Guo, Z.; Li, X. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst. Biol.,* **2013**, *7*, 101. http://dx.doi.org/10.1186/1752-0509-7-101 PMID: 24103777

[32] Chen, X.; Liu, M.X.; Yan, G.Y. RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.,* **2012**, *8*(10), 2792-2798. http://dx.doi.org/10.1039/c2mb25180a PMID: 22875290

[33] Xuan, P.; Han, K.; Guo, Y.; Li, J.; Li, X.; Zhong, Y.; Zhang, Z.; Ding, J. Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics,* **2015**, *31*(11), 1805-1815. http://dx.doi.org/10.1093/bioinformatics/btv039 PMID: 25618864

[34] Zhao, Y.; Chen, X.; Yin, J. A novel computational method for the identification of potential miRNA-disease association based on symmetric non-negative matrix factorization and kronecker regularized least square. *Front. Genet.,* **2018**, *9*, 324. http://dx.doi.org/10.3389/fgene.2018.00324 PMID: 30186308

[35] Chen, X.; Yan, G.Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.,* **2014**, *4*, 5501. http://dx.doi.org/10.1038/srep05501 PMID: 24975600

[36] Chen, X.; Zhou, Z.; Zhao, Y. ELLPMDA: Ensemble learning and link prediction for miRNA-disease association prediction. *RNA Biol.,* **2018**, *15*(6), 807-818. http://dx.doi.org/10.1080/15476286.2018.1460016 PMID: 29619882

[37] Chen, X.; Cheng, J.Y.; Yin, J. Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. *RNA Biol.,* **2018**, *15*(9), 1192-1205. http://dx.doi.org/10.1080/15476286.2018.1517010 PMID: 30196756

[38] Niu, Y.W.; Wang, G.H.; Yan, G.Y.; Chen, X. Integrating random walk and binary regression to identify novel miRNA-disease association. *BMC Bioinformatics,* **2019**, *20*(1), 59. http://dx.doi.org/10.1186/s12859-019-2640-9 PMID: 30691413

[39] Pasquier, C.; Gardès, J. Prediction of miRNA-disease associations with a vector space model. *Sci. Rep.,* **2016**, *6*, 27036. http://dx.doi.org/10.1038/srep27036 PMID: 27246786

[40] Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics,* **2010**, *26*(13), 1644-1650. http://dx.doi.org/10.1093/bioinformatics/btq241 PMID: 20439255

[41] Xuan, P.; Han, K.; Guo, M.; Guo, Y.; Li, J.; Ding, J.; Liu, Y.; Dai, Q.; Li, J.; Teng, Z.; Huang, Y. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One,* **2013**, *8*(8), e70204. http://dx.doi.org/10.1371/journal.pone.0070204 PMID: 23950912

[42] Wang, C.C.; Chen, X.; Yin, J.; Qu, J. An integrated framework for the identification of potential miRNA-disease association based on novel negative samples extraction strategy. *RNA Biol.,* **2019**, *16*(3), 257-269. http://dx.doi.org/10.1080/15476286.2019.1568820 PMID: 30646823

[43] Li, J.Q.; Rong, Z.H.; Chen, X.; Yan, G.Y.; You, Z.H. MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget,* **2017**, *8*(13), 21187-21199. http://dx.doi.org/10.18632/oncotarget.15061 PMID: 28177900

[44] Yang, Z.; Wu, L.; Wang, A.; Tang, W.; Zhao, Y.; Zhao, H.; Teschendorff, A.E. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.,* **2017**, *45*(D1), D812-D818. http://dx.doi.org/10.1093/nar/gkw1079 PMID: 27899556

[45] Ruepp, A.; Kowarsch, A.; Schmidl, D.; Buggenthin, F.; Brauner, B.; Dunger, I.; Fobo, G.; Frishman, G.; Montrone, C.; Theis, F.J.; Phenomi, R. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.,* **2010**, *11*(1), R6. http://dx.doi.org/10.1186/gb-2010-11-1-r6 PMID: 20089154

[46] Das, S. S.; Saha, P.; Chakravorty, N. miRwayDB: a database for experimentally validated mi-croRNA-pathway associations in pathophysiological conditions. *Database (Oxford),* **2018**.

[47] Xue, Z.; Wen, J.; Chu, X.; Xue, X. A microRNA gene signature for identification of lung cancer. *Surg. Oncol.,* **2014**, *23*(3), 126-131. http://dx.doi.org/10.1016/j.suronc.2014.04.003 PMID: 25031224

[48] Cho, W.C. Role of miRNAs in lung cancer. *Expert Rev. Mol. Diagn.,* **2009**, *9*(8), 773-776. http://dx.doi.org/10.1586/erm.09.57 PMID: 19895222

[49] Landi, M.T.; Chatterjee, N.; Yu, K.; Goldin, L.R.; Goldstein, A.M.; Rotunno, M.; Mirabello, L.; Jacobs, K.; Wheeler, W.; Yeager, M.; Bergen, A.W.; Li, Q.; Consonni, D.; Pesatori, A.C.; Wacholder, S.; Thun, M.; Diver, R.; Oken, M.; Virtamo, J.; Albanes, D.; Wang, Z.; Burdette, L.; Doheny, K.F.; Pugh, E.W.; Laurie, C.; Brennan, P.; Hung, R.; Gaborieau, V.; McKay, J.D.; Lathrop, M.; McLaughlin, J.; Wang, Y.; Tsao, M.S.; Spitz, M.R.; Wang, Y.; Krokan, H.; Vatten, L.; Skorpen, F.; Arnesen, E.; Benhamou, S.; Bouchard, C.; Metspalu, A.; Vooder, T.; Nelis, M.; Välk, K.; Field, J.K.; Chen, C.; Goodman, G.; Sulem, P.; Thorleifsson, G.; Rafnar, T.; Eisen, T.; Sauter, W.; Rosenberger, A.; Bickeböller, H.; Risch, A.; Chang-Claude, J.; Wichmann, H.E.; Stefansson, K.; Houlston, R.; Amos, C.I.; Fraumeni, J.F.; Savage, S.A.; Bertazzi, P.A.; Tucker, M.A.; Chanock, S.; Caporaso, N.E. A Genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.,* **2011**, *88*(6), 861. http://dx.doi.org/10.1016/j.ajhg.2011.05.003 PMID: 28472664

[50] Rodenhuis, S.; Slebos, R.J. Clinical significance of ras oncogene activation in human lung cancer. *Cancer Res.,* **1992**, *52*(9 Suppl.), 2665s-2669s. PMID: 1562997

[51] Marchetti, A.; Martella, C.; Felicioni, L.; Barassi, F.; Salvatore, S.; Chella, A.; Camplese, P.P.; Iarussi, T.; Mucilli, F.; Mezzetti, A.; Cuccurullo, F.; Sacco, R.; Buttitta, F. EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *J. Clin. Oncol.,* **2005**, *23*(4), 857-865. http://dx.doi.org/10.1200/JCO.2005.08.043 PMID: 15681531

[52] Shigematsu, H.; Lin, L.; Takahashi, T.; Nomura, M.; Suzuki, M.; Wistuba, I.I.; Fong, K.M.; Lee, H.; Toyooka, S.; Shimizu, N.; Fujisawa, T.; Feng, Z.; Roth, J.A.; Herz, J.; Minna, J.D.; Gazdar, A.F. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J. Natl. Cancer Inst.,* **2005**, *97*(5), 339-346. http://dx.doi.org/10.1093/jnci/dji055 PMID: 15741570

[53] Iorio, M.V.; Ferracin, M.; Liu, C.G.; Veronese, A.; Spizzo, R.; Sabbioni, S.; Magri, E.; Pedriali, M.; Fabbri, M.; Campiglio, M.; Ménard, S.; Palazzo, J.P.; Rosenberg, A.; Musiani, P.; Volinia, S.; Nenci, I.; Calin, G.A.; Querzoli, P.; Negrini, M.; Croce, C.M. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.,* **2005**, *65*(16), 7065-7070. http://dx.doi.org/10.1158/0008-5472.CAN-05-1783 PMID: 16103053

[54] Raponi, M.; Dossey, L.; Jatkoe, T.; Wu, X.; Chen, G.; Fan, H.; Beer, D.G. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res.,* **2009**, *69*(14), 5776-5783. http://dx.doi.org/10.1158/0008-5472.CAN-09-0587 PMID: 19584273

[55] Lu, J.; Getz, G.; Miska, E.A.; Alvarez-Saavedra, E.; Lamb, J.; Peck, D.; Sweet-Cordero, A.; Ebert, B.L.; Mak, R.H.; Ferrando, A.A.; Downing, J.R.; Jacks, T.; Horvitz, H.R.; Golub, T.R. MicroRNA expression profiles classify human cancers. *Nature,* **2005**, *435*(7043), 834-838. http://dx.doi.org/10.1038/nature03702 PMID: 15944708

[56] Eisemann, N.; Waldmann, A.; Katalinic, A. Epidemiology of breast cancer - current figures and trends. *Geburtshilfe Frauenheilkd.,* **2013**, *73*(2), 130-135. http://dx.doi.org/10.1055/s-0032-1328075 PMID: 24771909

[57] Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. Global cancer statistics. *CA Cancer J. Clin.,* **2011**, *61*(2), 69-90. http://dx.doi.org/10.3322/caac.20107 PMID: 21296855

[58] Tang, J.; Ahmad, A.; Sarkar, F.H. MicroRNAs in breast cancer therapy. *Curr. Pharm. Des.,* **2014**, *20*(33), 5268-5274. http://dx.doi.org/10.2174/1381612820666140128205239 PMID: 24479805

[59] Rask, L.; Balslev, E.; Søkilde, R.; Høgdall, E.; Flyger, H.; Eriksen, J.; Litman, T. Differential expression of miR-139, miR-486 and miR-21 in breast cancer patients sub-classified according to lymph node status. *Cell Oncol. (Dordr.),* **2014**, *37*(3), 215-227. http://dx.doi.org/10.1007/s13402-014-0176-6 PMID: 25027758

[60] Shen, S.; Sun, Q.; Liang, Z.; Cui, X.; Ren, X.; Chen, H.; Zhang, X.; Zhou, Y. A prognostic model of triple-negative breast cancer based on miR-27b-3p and node status. *PLoS One,* **2014**, *9*(6), e100664. http://dx.doi.org/10.1371/journal.pone.0100664 PMID: 24945253

[61] Xiong, D.D.; Lv, J.; Wei, K.L.; Feng, Z.B.; Chen, J.T.; Liu, K.C.; Chen, G.; Luo, D.Z. A nine-miRNA signature as a potential diagnostic marker for breast carcinoma: An integrated study of 1,110 cases. *Oncol. Rep.,* **2017**, *37*(6), 3297-3304. http://dx.doi.org/10.3892/or.2017.5600 PMID: 28440475

[62] van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics,* **2011**, *27*(21), 3036-3043. http://dx.doi.org/10.1093/bioinformatics/btr500 PMID: 21893517

[63] Chen, X.; Yan, G.Y. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics,* **2013**, *29*(20), 2617-2624. http://dx.doi.org/10.1093/bioinformatics/btt426 PMID: 24002109