RNA BIOLOGY
2018, VOL. 15, NO. 9, 1215–1227
https://doi.org/10.1080/15476286.2018.1521210

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

GLNMDA: a novel method for miRNA-disease association prediction based on global linear neighborhoods

Sheng-Peng Yu[a], Cheng Liang[a], Qiu Xiao[b], Guang-Hui Li[c], Ping-Jian Ding[d], and Jia-Wei Luo[d]

[a]School of Information Science and Engineering, Shandong Normal University, Jinan, China; [b]College of Information Science and Engineering, Hunan Normal University, Changsha, China; [c]School of Information Engineering, East China Jiaotong University, Nanchang, China; [d]College of Information Science and Engineering, Hunan University, Changsha, China

ABSTRACT

Recently, increasing studies have shown that miRNAs are involved in the development and progression of various complex diseases. Consequently, predicting potential miRNA-disease associations makes an important contribution to understanding the pathogenesis of diseases, developing new drugs as well as designing individualized diagnostic and therapeutic approaches for different human diseases. Nonetheless, the inherent noise and incompleteness in the existing biological datasets have limited the prediction accuracy of current computational models. To solve this issue, in this paper, we propose a novel method for miRNA-disease association prediction based on global linear neighborhoods (GLNMDA). Specifically, our method obtains a new miRNA/disease similarity matrix by linearly reconstructing each miRNA/disease according to the known experimentally verified miRNA-disease associations. We then adopt label propagation to infer the potential associations between miRNAs and diseases. As a result, GLNMDA achieved reliable performance in the frameworks of both local and global LOOCV (AUCs of 0.867 and 0.929, respectively) and 5-fold cross validation (average AUC of 0.926). Case studies on five common human diseases further confirmed the utility of our method in discovering latent miRNA-disease pairs. Taken together, GLNMDA could serve as a reliable computational tool for miRNA-disease association prediction.

ARTICLE HISTORY
Received 21 May 2018
Revised 22 August 2018
Accepted 24 August 2018

KEYWORDS
Mirna-disease associations; global linear neighborhoods; label propagation

Introduction

MicroRNAs(miRNAs) are highly enriched small non-coding RNAs of approximately 22 nucleotides that normally regulate gene expression at the post-transcriptional level by targeting mRNA for cleavage or translational inhibition[1–3]. Since the discovery of the first two mammalian microRNAs, mounting evidences have shown that miRNAs are involved in a variety of physiological and pathological processes[4]. Many major cellular functions such as development, differentiation, growth and metabolism are known to be regulated by miRNAs[5]. In addition, it has been suggested that miRNAs play vital roles in the pathogenesis of human diseases. For instance, by digging into the miRNA expression profiles of 93 primary human breast tumors, Blenkiron et al. identified a number of miRNAs that were differentially expressed between different molecular tumor subtypes[6]. Recently, Zhang et al. identified miRNA-26a as a key regulon that inhibits progression and metastasis of c-Myc/EZH2 double high advanced hepatocellular carcinoma[7]. Consequently, many studies aim at identifying key miRNAs as diagnostic and therapeutic biomarkers for human diseases. It is thus of great significance to uncover the potential associations between miRNAs and various diseases.

Many efforts made to predict potential disease-related miRNAs using experimental approaches have been proven successful, such as qRT-PCR and microarray profiling. Although reliable, experimental based methods are generally time-consuming and labor-intensive[8]. With the increasing amount of available biological data, a great number of computational models have been developed by taking advantage of multiple data sources to effectively and efficiently predict associations between miRNAs and diseases[9–11]. Under the assumption that miRNAs with similar functions tend to be associated to phenotypically similar diseases[12,13]. Jiang et al. proposed the first computational model based on hypergeometric distribution to predict new miRNA-disease associations[14], in which they integrated the phenotypic similarity network of diseases, the miRNA functional similarity network as well as the known human disease-miRNA association networks. Xu et al. introduced a network-centric approach to prioritize candidate disease miRNAs by constructing four topological features that are distinguishable between prostate cancer (PC) and non-PC miRNAs[15]. Xuan et al. proposed a model named HMDP which calculated miRNA-disease associations based on the functional similarities of k most similar neighbors of disease-associated miRNAs[16]. Specifically, miRNAs within the same clusters or families were assigned higher weights since they were more likely to be related to similar diseases when calculating the miRNA functional similarity matrix. Nevertheless, HDMP cannot be applied to diseases without any known related miRNAs since it is based on local similarity measures. To solve this issue, Chen et al. developed a novel computational approach called HGIMDA

CONTACT Cheng Liang alcs417@sdnu.edu.cn

which integrates miRNA functional similarity, disease semantic similarity, kernel similarity of Gaussian interaction profile, and experimentally validated miRNA-disease associations to predict potential miRNA-disease associations[17]. They further constructed a heterogeneous graph to iteratively update the association scores between unconfirmed miRNAs and diseases. Based on the assumption that miRNAs with targets related to a given disease were also likely to be associated with that disease, Shi et al. developed a computational framework to identify the miRNA-disease associations by conducting random walk with restart (RWR) algorithm on protein-protein interaction (PPI) networks[18]. Chen et al. proposed a method named WBSMDA to uncover the potential miRNAs related with multiple complex diseases by calculating a within score and a between score to obtain the final relevance scores for the unconfirmed miRNA-disease associations. Besides, WBSMDA could also be applied to diseases without any known related miRNAs[19].

Recently, several studies taking advantage of network topological structures have been proposed to prioritize disease-related miRNAs. Sun et al. developed NTSMDA to predict potential disease-miRNA associations by calculating the network topological similarity for both miRNAs and diseases. Nevertheless, since NTSMDA only utilized the known miRNA-disease association network to compute the network topological similarities, it is quite sensitive to the quality of the input data and cannot be applied to diseases without any known associated miRNAs[20]. You et al. developed a path-based model named PBMDA for miRNA-disease association prediction by integrating various biological data. Concretely, PBMDA adopted a depth-first search algorithm to search paths of certain lengths for given miRNA-disease pairs on a heterogeneous graph and obtained comparable performance. However, the computational complexity of PBMDA could be extremely high in large networks[21]. Chen et al. proposed NDAMDA to predict miRNA-disease associations based on network distance analysis. The highlight of their method lies in that two types of distances were considered, i.e. the direct distance and average distance. The direct distance represented a distance between two miRNAs (diseases) and the average distance represented the mean network distances of all miRNAs (diseases)[22].

In addition, several machine learning-based models were proposed to predict the potential miRNA-disease associations. Jiang et al. adopted the support vector machine (SVM) to predict the associations between miRNAs and diseases. They first extracted a set of features for each positive and negative miRNA-disease association, and then trained the SVM classifier with the constructed features to classify candidate disease-related miRNAs[23]. Chen et al. developed RBMMMDA which can not only predict the new associations between miRNAs and diseases, but also obtain the type of corresponding association[24]. Zou et al. introduced a biased SVM which was trained by a bagging algorithm to classify miRNA-disease pairs[25]. Liu et al. first constructed a heterogeneous network by connecting disease similarity network, miRNA similarity network as well as known miRNA-disease associations. They then extended random walk with restart to predict miRNA-disease associations in the heterogeneous network[26]. Li et al. utilized the matrix completion algorithm to update the adjacency matrix of known miRNA-disease associations and then predicted the potential miRNA-disease associations [27]. Chen et al. proposed another computational model called RKNNMDA which utilized the SVM ranking model to obtain reliable *k*-nearest-neighbors for each miRNA and disease. Specifically, it can be used to predict potential miRNAs for diseases without any known miRNAs[28]. They further proposed another model named MKRMDA to discover the potential miRNA-disease associations[29]. The innovation of MKRMDA was that it could automatically optimize the multiple kernel combinations of disease and miRNA. Chen et al. presented a computational model named LRSSLMDA, which projected miRNAs/diseases' statistical feature profile and graph theoretical feature profile to a common subspace. It used Laplacian regularization to preserve the local structures of the training data and a L_1-norm constraint to select important miRNA/disease features for prediction[30]. Xiao et al. proposed a graph regularized non-negative matrix factorization method for identifying miRNA-disease associations and their method was robust to the noises existing in the current datasets[31]. Zeng et al. derived a structural perturbation method to predict potential associations between miRNAs and diseases by using structural consistency as an indicator to estimate the link predictability of related networks[32]. Chen et al. developed the first decision tree learning-based model named EGBMMDA by employing Extreme Gradient Boosting Machine[33]. They constructed an informative feature vector by incorporating statistical measures, graph theoretical measures as well as matrix factorization results. Generally, a limitation of the machine learning-based algorithms is that there are no validated negative samples for miRNA-disease associations. They further used ensemble learning to combine rank results obtained by three classic similarity-based algorithms to predict miRNA-disease associations[34]. Recently, Chen et al. proposed a novel computational model to predict miRNA-disease associations based on bipartite network projection, which achieved comparable results in different cross-validation frameworks[35].

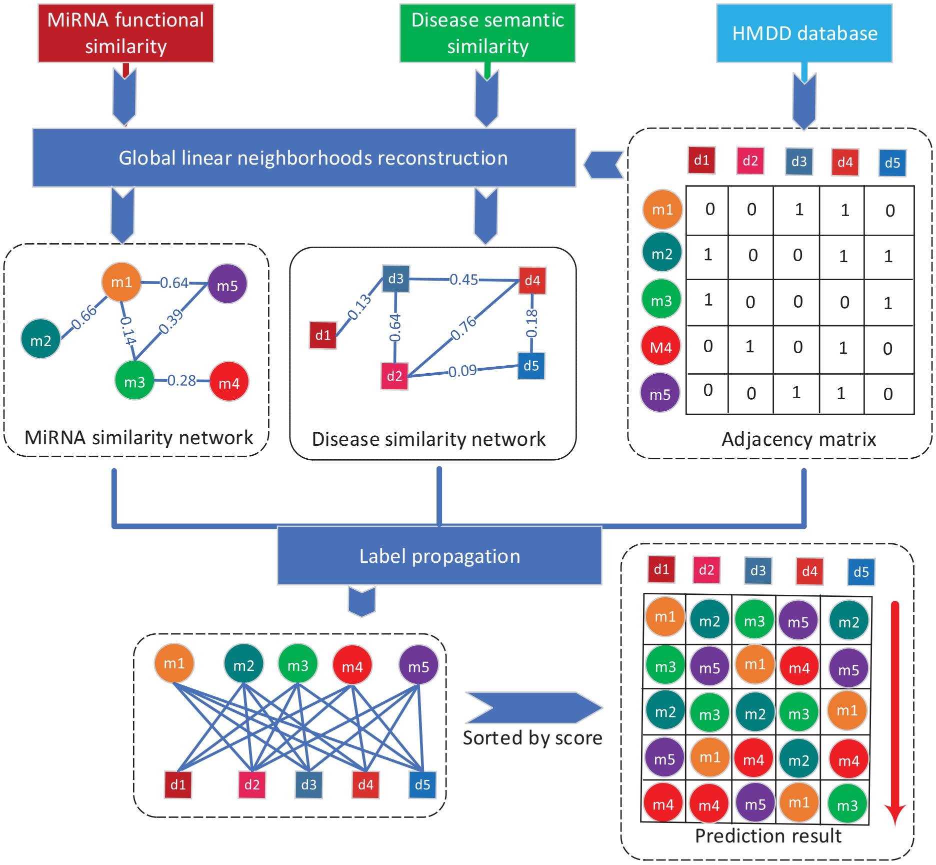
Although existing computational methods have been greatly improved in many details, they still have limitations. Therefore, developing novel methods to efficiently and reliably excavate the potential miRNA-disease associations is significant for human health and medical advance. In this study, we propose a novel method for MiRNA-Disease Association prediction based on Global Linear Neighborhoods (GLNMDA). Specifically, GLNMDA linearly reconstructs each miRNA (disease) by weighted combinations of its direct neighbors and indirect neighbors that can be reached by any steps of random walks. To demonstrate the effectiveness of our method, we implement leave-one-out cross-validation (LOOCV) and five-fold cross-validation for GLNMDA. As a result, GLNMDA obtained global AUC value of 0.929, local AUC value of 0.867 and 5-fold cross validation value of 0.926, respectively. Moreover, we compared our method with four state-of-the-art methods and the results indicated that our method consistently outperformed the other methods. In addition, three types of case studies were performed on five common cancers to verify the reliability and robustness of GLNMDA. Together, GLNMDA is an effective method for predicting potential miRNA-disease associations.

Results

Performance evaluation

In this section, we applied LOOCV and 5-fold cross-validation to test the prediction performance of our method based on known miRNA-disease associations from HMDD v2.0 databases[36]. LOOCV could be carried out in two manners: global and local LOOCV. In both frameworks, each known miRNA-disease association was left in turn as a test sample and other known miRNA-disease associations were regarded as training samples[37]. The only difference between global LOOCV and local LOOCV was that whether all the diseases were investigated simultaneously. In the global LOOCV, the test sample was compared and ranked with all candidate miRNAs, whereas in the local LOOCV, the test sample is compared and ranked with the miRNAs only associated with the specific disease. We also implemented 5-fold cross validation to evaluate the performance of GLNMDA. In the framework of 5-fold cross validation, all the known miRNA-disease associations were randomly divided into five disjoint parts, where each part was picked out as test samples in turn and the other four parts were treated as training samples. In addition, Receiver Operating Characteristics (ROC) curves were plotted by calculating the true positive rate (TPR) and the false positive rate (FPR) at varying thresholds[38]. The prediction performance of GLNMDA can be quantitatively evaluated by calculating the Area Under the ROC Curve (AUC). Specifically, the value of AUC is from 0 to 1 and the larger the AUC values, the better the predicted results. As shown in Figure 2, GLNMDA achieved AUC values of 0.929, 0.867 and 0.926 in global LOOCV, local LOOCV and 5-fold cross-validation, respectively, which clearly demonstrated the superior performance of our method.

We further compared GLNMDA with four state-of-the-art methods (i.e. HGIMDA[17], EGBMMDA[33], PBMDA[21], MKRMD[29]), all of which have also achieved excellent performances in predicting potential miRNA-disease associations. As mentioned above, HGIMDA was an efficient prediction framework based on heterogeneous graph inference. Both EGBMMDA and MKRMDA were machine learning-based approaches with different feature extraction schemas. PBMDA was a depth-first model which took network topology into account. As shown in Figure 2, HGIMDA, EGBMMDA, PBMDA and MKRMDA obtained AUCs of 0.875, 0.912, 0.922 and 0.904 in global LOOCV, respectively. Similarly, they obtained AUCs of 0.823, 0.807, 0.853 and 0.827 in the local LOOCV framework, respectively (Figure 3). For 5-fold cross-validation, they achieved AUCs of 0.867, 0.904, 0.916 and 0.884,



	d1	d2	d3	d4	d5
m1	0	0	1	1	0
m2	1	0	0	1	1
m3	1	0	0	0	1
M4	0	1	0	1	0
m5	0	0	1	1	0

d1	d2	d3	d4	d5
m1	m2	m3	m5	m2
m3	m5	m1	m4	m5
m2	m3	m2	m3	m1
m5	m1	m4	m2	m4
m4	m4	m5	m1	m3

Figure 1. Flowchart of potential disease-miRNA association prediction based on the computational model of GLNMDA.

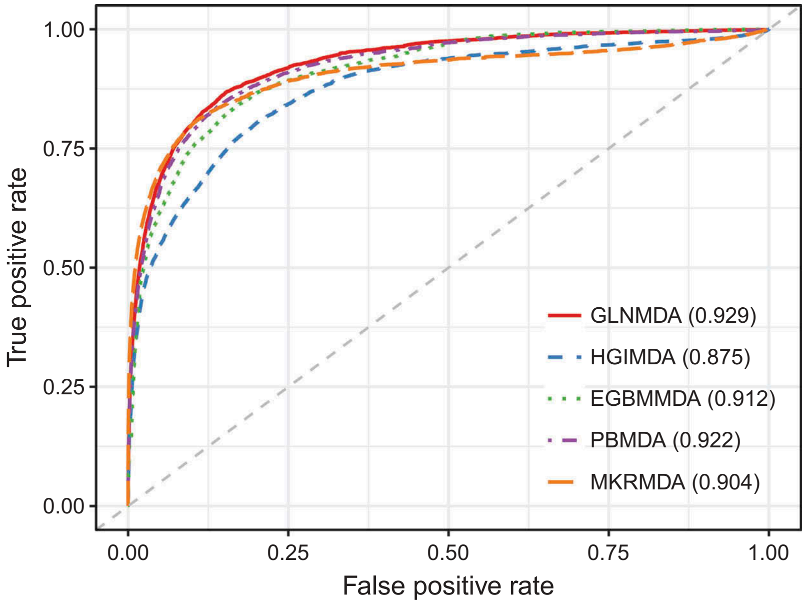


Figure 2. The comparison results between GLNMDA and the othe1`r four computational models in the framework of global LOOCV.

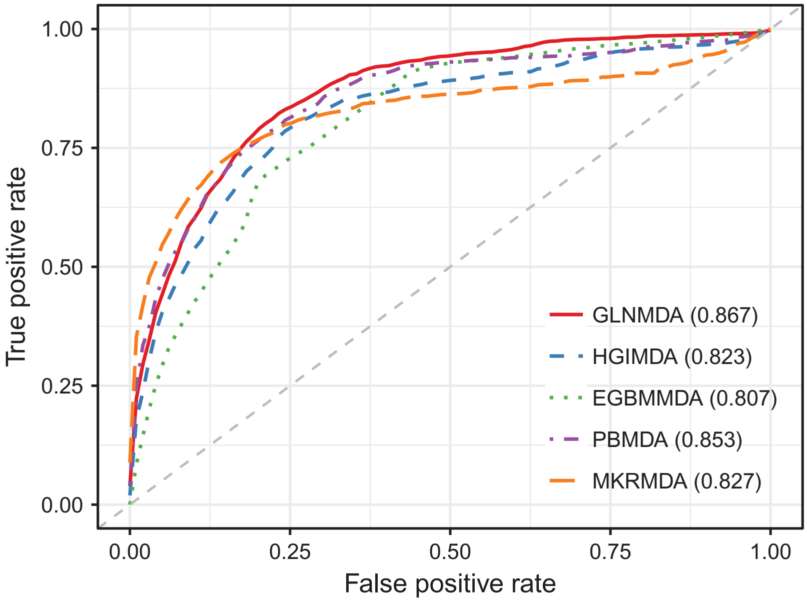


Figure 3. The comparison results between GLNMDA and the other four computational models in terms of local LOOCV.

respectively (Figure 4). Obviously, GLNMDA consistently outperformed the four methods in all three cross-validation frameworks. In conclusion, GLNMDA could serve as a reliable tool to predict the potential associations between miRNAs and diseases.

Parameter analysis

One important step in GLNMDA is to learn a rank-k non-negative symmetric matrix to reconstruct the miRNA similarity network and disease similarity network from miRNA space and disease space, respectively. To test whether different values of k would affect the final prediction results, we selected eleven values of k ranging from 20 to 120 with an interval of 10 and then compared the prediction accuracy in all three cross-validation frameworks. As illustrated in Figure 5–7, GLNMDA obtained the worst performance in all the cross validations when $k = 20$ while the performance remains relatively stable when $k > 20$. Therefore, we can conclude that different values of k only have minor effects on the final results.

Case studies

To further demonstrate the predictive power of GLNMDA, we conducted three types of case studies on five common human diseases. Specifically, we selected 16 common diseases among the four databases (i.e. dbDEMC[39], miR2Disease[40], miRwayDB[41] and PhenomiR[42]) for the subsequent case studies and validated the prediction

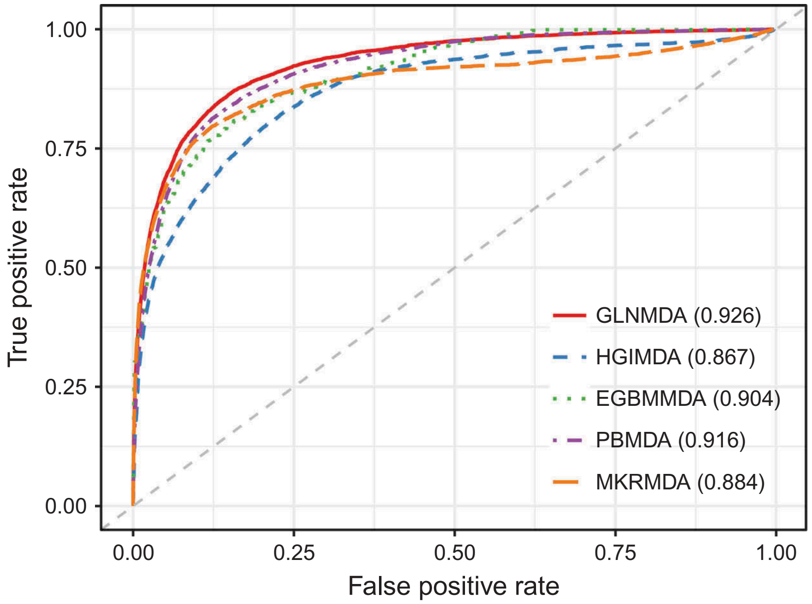


Figure 4. The comparison results between GLNMDA and the other four computational models in the framework of 5-fold cross-validation.

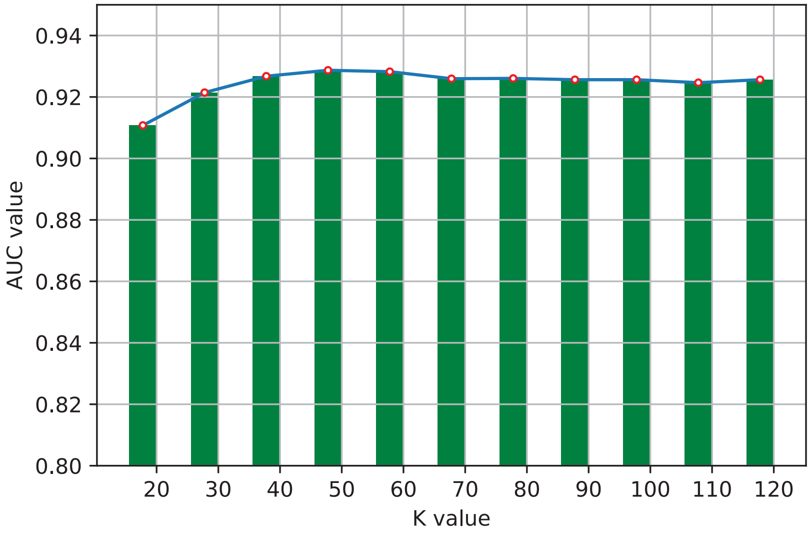


Figure 5. The effects of different values of *k* in global cross validation.

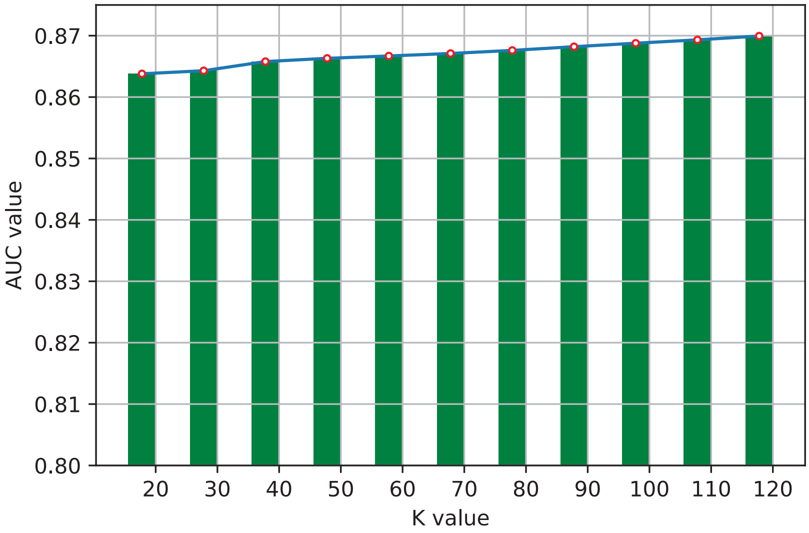


Figure 6. The effects of different values of *k* in local cross validation.

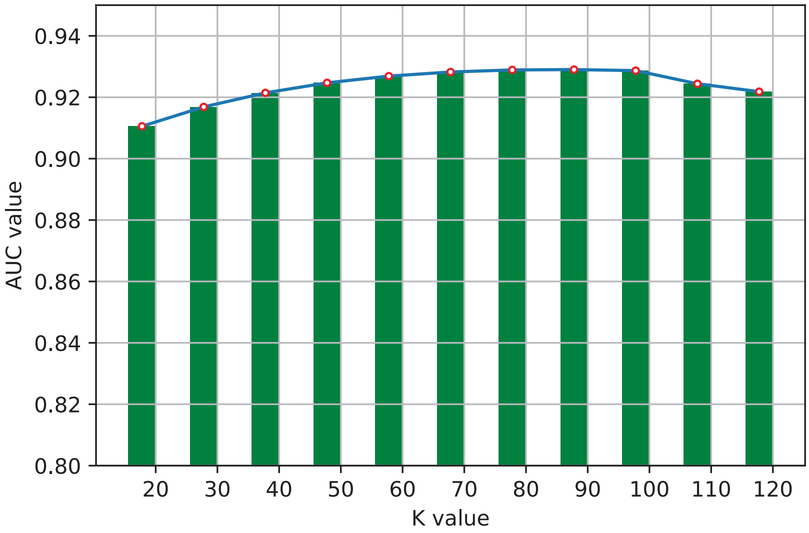


Figure 7. The effects of different values of *k* in 5-fold cross validation.

results across all the databases. The 16 common diseases are Breast Neoplasms, Cervical Intraepithelial Neoplasia, Colorectal Neoplasms, Hepatocellular Carcinoma, Lymphoma, Lung Neoplasms, Leukemia, Nasopharyngeal Neoplasms, Liver Neoplasms, Ovarian Neoplasms, Pancreatic Neoplasms, Prostatic Neoplasms, Stomach Neoplasms, Thyroid Neoplasms, Urinary Bladder Neoplasms and Uterine Cervical Neoplasms. Due to space limitations, we provided the validation results of 5 diseases in the main text and put the results of the other diseases on Github(https://github.com/ShengPengYu/GLNMDA/tree/master/CaseStudy). The first type of case study was implemented for Lung Neoplasm (LN), Hepatocellular Carcinoma (HC) and Breast Neoplasms (BN), in which we prioritized the top 50 predicted miRNAs for the given diseases based on the known disease-miRNA associations from HMDD v2.0. The prediction results were then verified by another four databases recording experimentally validated disease-related miRNAs.

Lung Neoplasms (LN) characterized by high mortality and high concurrency is one of the most common cancers and have caused a serious threat to human health especially in male[43]. It has been reported that untreated patients with small cell lung cancer will quickly deteriorate and eventually die in 12 weeks [44,45]. Increasing evidence has suggested that miRNAs can not only be utilized to classify LNs, but also have the potential to be biomarkers for early diagnosis and clinical treatment of LN[46–49]. As shown in Table 1, 45 out of the top 50 candidate miRNAs were confirmed to be associated with LN. For instance, the hsa-let-7 family which regulates the cell cycle and the hsa-mir-200 family that induces cell death and cell proliferation were all differentially expressed in LN tumor samples[50]. Among the five unconfirmed miRNAs, hsa-miR-499 has been found that the rs3746444T> C polymorphism in its mature sequence could contribute to poor prognosis by modulating cancer-related gene expression and thus involve in the tumorigenesis of LN[51]. Besides, studies have shown that miR-103 was able to promote proliferation of small cell lung cancer cells through targeting MED26 mRNA 3′-UTR[52].

Hepatocellular Carcinoma (HC) is a primary malignancy of the liver and occurs predominantly in patients with underlying chronic liver disease and cirrhosis. Accumulating evidences have shown that the expression patterns of certain miRNAs were significantly different between HC and normal tissues, which might serve as a diagnostic tool for HC[53]. For instance, the ectopic expression of hsa-mir-101 could dramatically suppress the ability of hepatoma cells to form colonies in vitro and to develop tumors in nude mice[54]. The top 50 HC-related miRNAs predicted by our method was listed in Table 2. As a result, 46 of the top 50 predicted miRNAs were confirmed to be associated with the given disease by at least one database from dbDEMC, miR2Disease, miRwayDB and PhenomiR. As a matter of fact, one of the unconfirmed miRNAs, hsa-mir-34a, has been to shown to inhibit migration and invasion by down-regulation of c-Met expression in human hepatocellular carcinoma cells[55].

Breast Neoplasms (BN) is one of the most common female cancers that threatens women’s physical and mental health, accounting for 22% of female cancers[56]. Recent research on miRNAs has implicated that the loss of tumor suppressor miRNAs or overexpression of oncogenic miRNAs can lead to breast cancer tumorigenesis or metastasis. Our prediction results showed that 47 of top 50 candidate miRNAs were confirmed by experimental findings recorded in at least one of the four databases dbDEMC, miR2Disease, miRwayDB and PhenomiR (Table 3). For example, the overexpression of hsa-mir-21 (ranked 1st in the prediction list) in human breast cancer is associated with advanced clinical stage, lymhp node metastasis and patient poor prognosis. Moreover, solid evidence has been provided that the C allele of hsa-mir-146a (ranked 2nd in the prediction list) is associated with early familial breast tumor development[57].

In addition, to test the ability of GLNMDA in predicting for diseases without any known associated miRNAs, we conducted the second type of case study on Colorectal Neoplasms (CN). It is reported that more than 1 million individuals will develop colorectal cancer every year worldwide and the disease-specific mortality rate is nearly 33% in the developed

Table 1. Top 50 predicted miRNAs associated with Lung Neoplasms based on known associations in HMDD. I, II, III and IV represent dbDEMC, miR2Disease, miRwayDB and PhenomiR, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	I;II;III;IV;	hsa-mir-29b	I;II;
hsa-mir-155	I;II;IV;	hsa-mir-143	I;II;IV;
hsa-mir-146a	I;II;III;IV;	hsa-mir-574	IV;
hsa-mir-17	I;III;IV;	hsa-mir-223	I;III;IV;
hsa-mir-125b	I;II;	hsa-mir-199a	I;
hsa-mir-34a	I;II;III;IV;	hsa-let-7c	I;II;IV;
hsa-mir-145	I;II;III;IV;	hsa-mir-34c	I;II;III;IV;
hsa-mir-20a	I;II;IV;	hsa-mir-200b	I;II;IV;
hsa-mir-126	I;II;III;IV;	hsa-mir-365a	IV;
hsa-mir-1297	I;	hsa-let-7e	I;II;IV;
hsa-mir-511	I;II;	hsa-mir-326	I;IV;
hsa-mir-499a	unconfirmed;	hsa-mir-200c	I;II;III;IV;
hsa-mir-221	I;II;IV;	hsa-mir-103b	unconfirmed;
hsa-let-7a	I;II;	hsa-mir-365b	unconfirmed;
hsa-mir-92a	I;	hsa-mir-513a	unconfirmed;
hsa-mir-18a	I;II;IV;	hsa-let-7d	I;II;IV;
hsa-mir-138	I;	hsa-mir-222	I;II;IV;
hsa-mir-103a	unconfirmed;	hsa-mir-146b	I;II;IV;
hsa-mir-19b	I;	hsa-mir-29c	I;II;IV;
hsa-mir-193a	I;IV;	hsa-mir-210	I;II;IV;
hsa-mir-128	I;III;	hsa-mir-9	I;II;
hsa-mir-19a	I;II;IV;	hsa-mir-31	I;II;III;IV;
hsa-mir-1	I;II;	hsa-let-7g	I;II;IV;
hsa-mir-29a	I;II;IV;	hsa-mir-34b	I;II;IV;
hsa-let-7b	I;II;IV;	hsa-let-7f	I;II;

Table 2. Top 50 predicted miRNAs associated with Hepatocellular Carcinoma based on known associations in HMDD. I, II, III and IV represent dbDEMC, miR2Disease, miRwayDB and PhenomiR, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	I;II;III;IV;	hsa-mir-143	unconfirmed;
hsa-mir-1322	I;II;III;IV;	hsa-mir-200c	I;II;IV;
hsa-mir-200c	I;	hsa-mir-223	I;II;
hsa-mir-617	I;II;IV;	hsa-mir-19a	I;II;
hsa-mir-766	I;II;IV;	hsa-let-7b	I;IV;
hsa-mir-192	I;II;III;IV;	hsa-mir-200a	I;II;III;IV;
hsa-mir-155	I;II;III;	hsa-let-7c	I;IV;
hsa-mir-205	I;II;	hsa-mir-199a	unconfirmed;
hsa-mir-203	I;II;IV;	hsa-mir-31	unconfirmed;
hsa-mir-1246	I;II;	hsa-mir-34c	I;II;IV;
hsa-mir-548d	I;II;IV;	hsa-mir-210	I;
hsa-mir-302f	I;II;	hsa-mir-15a	I;II;IV;
hsa-mir-451a	I;II;IV;	hsa-mir-29c	I;II;IV;
hsa-mir-145	I;II;III;	hsa-mir-203	I;
hsa-mir-141	I;II;	hsa-mir-141	I;IV;
hsa-mir-146a	I;II;III;IV;	hsa-mir-34b	I;II;IV;
hsa-mir-499a	I;II;IV;	hsa-mir-196a	I;
hsa-mir-193a	I;	hsa-mir-101	I;
hsa-mir-574	I;IV;	hsa-mir-148a	I;II;IV;
hsa-mir-425	I;II;IV;	hsa-mir-205	I;IV;
hsa-mir-20a	I;III;	hsa-mir-150	I;II;IV;
hsa-mir-34a	unconfirmed;	hsa-mir-100	I;IV;
hsa-mir-126	I;II;IV;	hsa-mir-133a	I;
hsa-let-7a	I;	hsa-mir-214	I;
hsa-mir-92a	I;	hsa-mir-375	I;II;III;IV;

Table 3. Top 50 predicted miRNAs associated with Breast Neoplasms based on known associations in HMDD. I, II, III and IV represent dbDEMC, miR2Disease, miRwayDB and PhenomiR, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	I;II;III;IV;	hsa-let-7c	I;IV;
hsa-mir-155	I;II;III;IV;	hsa-mir-223	I;III;IV;
hsa-mir-17	I;IV;	hsa-mir-451a	I;III;
hsa-mir-146a	I;II;III;IV;	hsa-mir-10b	I;II;III;IV;
hsa-mir-145	I;II;IV;	hsa-mir-222	I;II;IV;
hsa-mir-125b	I;II;III;	hsa-mir-151a	unconfirmed;
hsa-mir-20a	I;II;IV;	hsa-mir-200a	I;II;III;IV;
hsa-mir-34a	I;III;IV;	hsa-mir-205	I;II;IV;
hsa-mir-126	I;II;IV;	hsa-mir-499a	unconfirmed;
hsa-let-7a	I;II;	hsa-mir-34c	I;
hsa-mir-221	I;II;IV;	hsa-let-7d	I;II;IV;
hsa-mir-18a	I;II;IV;	hsa-let-7e	I;IV;
hsa-mir-92a	I;	hsa-mir-146b	II;
hsa-mir-19b	I;III;	hsa-mir-9	I;III;
hsa-mir-16	I;	hsa-mir-15a	I;IV;
hsa-mir-29a	I;III;IV;	hsa-mir-708	I;
hsa-mir-19a	I;III;IV;	hsa-mir-181a	I;II;
hsa-mir-200b	I;II;III;IV;	hsa-mir-128	I;
hsa-let-7b	I;IV;	hsa-mir-29c	I;II;IV;
hsa-mir-1	I;III;	hsa-mir-320a	I;III;IV;
hsa-mir-200c	I;II;IV;	hsa-mir-31	I;II;IV;
hsa-mir-29b	I;II;	hsa-let-7f	I;II;
hsa-mir-143	I;II;IV;	hsa-mir-106b	I;IV;
hsa-mir-103a	I;	hsa-mir-34b	I;IV;
hsa-mir-199a	I;III;	hsa-mir-218	I;

Table 4. Top 50 predicted miRNAs associated with Colorectal Neoplasms based on known associations in HMDD. I, II, III and IV represent dbDEMC, miR2Disease, miRwayDB and PhenomiR, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	HMDD;I;II;IV;	hsa-mir-222	HMDD;I;IV;
hsa-mir-155	HMDD;I;II;IV;	hsa-let-7b	HMDD;I;II;IV;
hsa-mir-146a	HMDD;I;IV;	hsa-mir-199a	unconfirmed;
hsa-mir-145	HMDD;I;II;IV;	hsa-let-7c	HMDD;I;IV;
hsa-mir-17	HMDD;I;IV;	hsa-mir-29c	I;III;IV;
hsa-mir-125b	I;	hsa-mir-142	HMDD;IV;
hsa-mir-126	HMDD;I;II;IV;	hsa-mir-200a	HMDD;I;IV;
hsa-mir-20a	HMDD;I;II;IV;	hsa-mir-210	HMDD;I;IV;
hsa-mir-34a	HMDD;I;II;IV;	hsa-mir-181a	I;II;
hsa-mir-16	I;	hsa-let-7e	HMDD;I;IV;
hsa-mir-221	HMDD;I;II;IV;	hsa-mir-133b	HMDD;I;II;IV;
hsa-mir-29a	HMDD;I;II;IV;	hsa-let-7f	I;
hsa-mir-92a	I;	hsa-mir-34c	HMDD;II;IV;
hsa-let-7a	I;II;	hsa-let-7d	I;IV;
hsa-mir-143	HMDD;I;II;IV;	hsa-mir-9	I;
hsa-mir-18a	HMDD;I;II;IV;	hsa-mir-146b	HMDD;IV;
hsa-mir-19b	I;II;	hsa-mir-106b	I;II;IV;
hsa-mir-1	I;II;	hsa-mir-150	HMDD;I;IV;
hsa-mir-29b	I;II;	hsa-let-7g	I;II;IV;
hsa-mir-223	I;II;IV;	hsa-let-7i	I;IV;
hsa-mir-200c	HMDD;I;II;IV;	hsa-mir-181b	I;II;
hsa-mir-200b	HMDD;I;IV;	hsa-mir-133a	I;II;
hsa-mir-19a	HMDD;I;II;IV;	hsa-mir-101	I;
hsa-mir-15a	I;IV;	hsa-mir-30a	HMDD;I;IV;
hsa-mir-31	HMDD;I;II;IV;	hsa-mir-182	HMDD;I;II;IV;

world[58]. Firstly, we removed all known associations related with CN and we then used GLNMDA to predict the potential associations between miRNAs and diseases. As a result, 49 of top 50 predicted candidate miRNAs have been confirmed by at least one database from dbDEMC, miR2Disease, miRwayDB and PhenomiR or HMDD (Table 4). The only unconfirmed miRNA was hsa-mir-199a. As a matter of fact, evidences have demonstrated that hsa-mir-199a plays a critical role in the cell biological behaviors of colorectal cancer through its target genes[59]. Our prediction results were consistent with existing findings and provided computational evidence for its association with CN.

Lastly, we conducted the third type of case studies for Lymphoma where the older version of HMDD was used to prioritize miRNAs with the given disease and the latest version of HMDD v2.0 was adopted to evaluate the prediction results. Due to the distribution characteristics of the lymphatic system, lymphoma is a systemic disease which can invade almost any tissue and organ in the body[60]. miRNAs have also been shown to act as potential biomarkers for the diagnosis of Lymphoma. For example, the under-expression of hsa-mir-150 will increase the incidence of apoptosis and reduced cell proliferation in normal cells[61]. Here, we implemented GLNMDA based on the older version of HMDD which included 1395 associations between 271 miRNAs and 137 diseases. As a result, 49 out of the top 50 predicted miRNAs were confirmed by the HMDD v2.0 and/or the other four databases (Table 5). Only hsa-mir-199a was not confirmed. The results showed that GLNMDA is a reliable method to predict the potential miRNA-disease associations.

Discussion

The identification of novel associations between miRNAs and diseases plays a crucial role in understanding the disease pathogenesis at the miRNA level. In this study, considering the sparsity and incompleteness of disease semantic similarity matrix and miRNA functional similarity matrix, we presented a novel method for miRNA-disease association prediction based on global linear neighborhoods. To demonstrate the effectiveness of the proposed method, we applied global LOOCV, local LOOCV and 5-fold cross-validation to evaluate the prediction performance. GLNMDA achieved AUCs of 0.929, 0.867 and 0.926 in the three frameworks, respectively. We further compared GLNMDA with four state-of-the-art methods and the results confirmed the superior performance of GLNMDA over the other methods. Besides, three types of case studies were implemented on five common human diseases to further validate the utility of GLNMDA. As a result, GLNMDA could uncover novel miRNA-disease associations as expected.

The success of GLNMDA could be largely attributed to the following factors. Firstly, we used the global neighborhoods information to reconstruct the miRNA similarity matrix and disease similarity matrix, which alleviated the sparsity and incompleteness problem existing in the current datasets. Secondly, known experimentally verified miRNA-disease information were used as the benchmark dataset in the cross-validation schema and the initial dataset for predicting latent human miRNA-disease association. Lastly, the known information was propagated by label propagation algorithm iteratively to the whole network according to the similarities reconstructed by GLNMDA.

Table 5. Top 50 predicted miRNAs associated with Lymphoma based on known associations in the older version of HMDD. I, II, III and IV represent dbDEMC, miR2Disease, miRwayDB and PhenomiR, respectively. The first and third columns record the 1–25 and 26–50 related miRNAs, respectively.

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	HMDD;I;II;III;IV;		HMDD;I;IV;
hsa-mir-155	HMDD;I;II;III;IV;	hsa-mir-181a	I;III;
hsa-mir-146a	HMDD;I;IV;	hsa-mir-34b	I;II;IV;
hsa-mir-221	HMDD;I;II;	hsa-mir-195	HMDD;I;IV;
hsa-let-7a	I;II;	hsa-mir-200a	HMDD;I;II;IV;
hsa-mir-223	HMDD;I;	hsa-mir-30a	I;IV;
hsa-mir-29a	HMDD;I;	hsa-mir-205	HMDD;I;
hsa-mir-29b	I;	hsa-mir-183	I;III;
hsa-mir-143	HMDD;I;II;IV;	hsa-mir-7	I;
hsa-mir-222	HMDD;I;	hsa-mir-133a	I;
hsa-let-7b	HMDD;I;IV;	hsa-mir-27a	HMDD;I;IV;
hsa-mir-1	I;	hsa-mir-141	I;
hsa-mir-31	I;	hsa-mir-10b	I;
hsa-let-7c	HMDD;I;II;IV;	hsa-mir-106a	I;II;IV;
hsa-mir-199a	unconfirmed;	hsa-mir-148a	I;
hsa-mir-9	I;II;III;	hsa-mir-93	HMDD;I;
hsa-mir-18a	HMDD;I;II;III;IV;	hsa-mir-100	HMDD;I;
hsa-mir-19a	HMDD;I;II;III;IV;	hsa-mir-150	HMDD;I;II;III;IV;
hsa-mir-34c	IV;	hsa-mir-15b	I;IV;
hsa-let-7d	HMDD;I;	hsa-mir-25	I;IV;
hsa-mir-106b	I;III;IV;	hsa-mir-199b	I;IV;
hsa-mir-182	HMDD;I;III;IV;	hsa-mir-203	HMDD;I;
hsa-mir-126	HMDD;I;IV;	hsa-mir-224	I;
hsa-let-7e	HMDD;I;II;IV;	hsa-mir-22	HMDD;I;IV;
hsa-let-7f	I;	hsa-mir-133b	HMDD;I;IV;

Nevertheless, there are still limitations in the current version of GLNMDA. Our approach can be improved in the following directions. Firstly, the performance of GLNMDA can be further improved by integrating more available experimentally-verified human miRNA-disease associations. Secondly, multiple information sources can be integrated properly to measure the functional similarity between miRNAs, such as the information of their target genes. In essence, construction for reliable miRNA similarity matrix as well as the disease similarity matrix would help improve the accuracy of GLNMDA.

Materials and methods

Human mirna-disease associations

The human microRNA disease database (HMDD), which contains 5340 experimentally verified links between 495 miRNAs and 383 diseases, is a reliable database[36]. We downloaded miRNA-disease associations information from HMDD database directly. Furthermore, we constructed an adjacent matrix R, of which the element was defined as follows: $R_{ij} = 1$ if disease $d(i)$ have an interaction with miRNA $m(j)$, and 0 otherwise. Our goal is to confirm the uncertain associations between miRNAs and diseases.

miRNA functional similarity

The miRNA functional similarity used in this paper was calculated by Wang et al. and can be downloaded directly at (http://www.cuilab.cn/files/images/cuilab/misim.zip) [62]. We used M to denote the miRNA functional similarity network, where each element M_{ij} represents the functional similarity score between miRNA $m(i)$ and $m(j)$.

Disease semantic similarity model

Mesh database (http://www.ncbi.nlm.nih.gov/) is a strict system for disease classification and is a credible dataset for effectively researching the relationship between different diseases[62]. The relationship between different diseases can be described through a structure of Directed Acyclic Graph (DAG). A disease A can be described as $DAG(A) = (A, T(A), E(A))$, where $T(A)$ represents all its ancestors and itself, and $E(A)$ contains edge information including the direct edges linking parent nodes to child nodes. The contribution of disease d_i in $DAG(A)$ to the semantic value of disease A was defined as follows:

$$\begin{cases} D_A(d_i) = 1\, if\ d = A \\ D_A(d_i) = \max\{\Delta * D_A(d_i')|d_i' \in childen\ of\ d_i\} if \mathrm{d} \neq A \end{cases} \tag{1}$$

Here, Δ is the semantic contribution factor and we set $\Delta = 0.5$ in this paper. For disease d_i, the contribution of itself is 1, while the contribution of another disease d_j decreases as the distance between d_i and d_j increases. Hence, the semantic value of disease A can be calculated according to the contribution of ancestor diseases and disease A itself [63]:

$$DV(A) = \sum\nolimits_{d_i \in T(A)} D_A(d_i) \tag{2}$$

Taken together, the semantic similarity of disease d_i and disease d_j can be calculated as follows:

$$S(d_i, d_j) = \frac{\sum_{t \in T(j) \cap T(i)} (D_i(t) + D_j(t))}{DV(i) + DV(j)} \tag{3}$$

According to Equation (3), we can construct an overall disease semantic similarity matrix D where D_{ij} represents the semantic similarity between disease d_i and disease d_j.

GLNMDA

In this work, we present a novel framework named GLNMDA to predict potential disease-related miRNAs based on global linear neighborhoods reconstruction. The key assumption of GLNMDA is that each miRNA (disease) can be linearly reconstructed by weighted combinations of its direct neighbors and indirect neighbors which can be reached by any steps of random walk. GLNMDA mainly consists of three steps: Firstly, we reconstruct the miRNA similarity network and disease similarity network based on the known miRNA-disease associations. Secondly, we utilize label propagation algorithm to prioritize novel interactions based on the reconstructed networks, respectively. Lastly, we obtain the final prediction results by combining the output from both miRNA space and disease space. An overall workflow is illustrated in Figure 1.

Feature representation for miRNAs and diseases

Generally, the reconstruction algorithm is conducted on feature vectors. Therefore, the first step of our algorithm is to construct the feature vectors for both diseases and miRNAs. As presented in the previous work[64], we adopted 'interaction profile' to build the features for miRNAs and diseases. Specifically, suppose the miRNA-disease interaction network consists of m RNAs and n diseases, where (M_1, M_2, M_3, …, M_m) and (D_1, D_2, D_3, …, D_n) represent the miRNA set and disease set, respectively. As stated above, if miRNA M_i is related with disease D_j, the entry in the corresponding adjacency matrix $R_{m \times n}$ is 1 and 0 otherwise. As a result, we could take each column as the feather vector for a given disease and each row as the feature vector for a given miRNA. Obviously, the adjacency matrix R is the disease feature matrix and the transpose of R represents the miRNA feature matrix.

Reconstruction of similarity matrix for diseases and miRNAs

With the rapid development of bio-technology, an increasing amount of biological data is now available for miRNA-disease association studies, including various similarity datasets for diseases and miRNAs. However, due to the limitation of current experimental conditions as well as the inherent noises in these datasets, the miRNA functional similarity matrix M and disease semantic similarity matrix D obtained were in general sparse and incomplete, which might greatly affect the accuracy of prediction results. To address this problem, we here use global linear neighborhoods reconstruction (GLNR) to rebuild the miRNA similarity network and disease similarity network. We assume that each miRNA (disease) can be linearly reconstructed by weighted combinations of its direct neighbors and indirect neighbors which can be reached by any steps of random walk[65]. Let X be the $n \times m$ data matrix where $x_i (i = 1,2,\ldots,n)$ is the i-th data point in X. According to GLNR, x_i can be reconstructed as follows:

$$x_i = \sum\nolimits_{j:x_j \in g(x_i)} W_{ij} x_j \text{ s.t. } W_{ij} > 0, \sum\nolimits_{j:x_j \in g(x_i)} W_{ij} = 1 \tag{4}$$

where $g(x_i)$ is the global neighborhood of x_i. Let W be the symmetric $n \times n$ similarity matrix between the data points to be learned. Instead of explicitly selecting k neighbors to make W sparse[66], we propose to learn a rank-k non-negative symmetric matrix $W = UU^T$ by the following objective function:

$$\min Q(U) = ||X - UU^T X||^2, \; s.t. \; U_{ij} \geq 0 \tag{5}$$

where U is a $n \times k$ feature matrix. In this paper, for a more general description, X could be either miRNA feature matrix R^T or disease feature matrix R. To solve the optimization problem, we first calculated the derivative of Equation (5) with respect to U and we have:

$$\frac{\partial Q}{\partial U} = -2(X - UU^T X)X^T U - 2X(X^T - X^T UU^T)U \tag{6}$$

Since X contains only non-negative data, we could obtain the multiplicative update rule as follows:

$$U_{ij} \leftarrow U_{ij} \times \sqrt{\frac{(2XX^T U)_{ij}}{(UU^T XX^T U + XX^T UU^T U)_{ij}}} \tag{7}$$

It is worth noting that to guarantee the convergence of the iterative update rule, we need to normalize our training data in advance[67,68]. Besides, to get an informative value of k for matrix factorization, we employed the clusterONE algorithm accordingly[69], a method for detecting potentially overlapping protein complexes from protein-protein interaction networks. Specifically, clusterONE builds on the concept of the cohesiveness score and uses a greedy growth process to find groups in protein-protein interaction networks that are likely to correspond to protein complexes. It has also been widely adopted to identify cohesive clusters in other types of biological networks due to its simplicity and efficiency[70]. By substituting M into Equation (7), a miRNA clustering matrix $\tilde{U}$ was learned as follows:

$$\tilde{U}_{ij} \leftarrow \tilde{U}_{ij} \times \sqrt{\frac{(2R^T R\tilde{U})_{ij}}{(\tilde{U}\tilde{U}^T R^T R\,\tilde{U} + R^T R\,\tilde{U}\tilde{U}^T\,\tilde{U})_{ij}}} \tag{8}$$

We then reconstructed the miRNA similarity matrix $\tilde{M}$ based on the learned clustering matrix $\tilde{U}$:

$$\tilde{M} = \tilde{M}_P^{-1/2}(\tilde{U}\tilde{U}^T)\,\tilde{M}_P^{-1/2} \tag{9}$$

Where $\tilde{M}_P$ is a diagonal matrix with its (*i*,*i*)-th element equal to the sum of the *ith* row of $\tilde{U}\tilde{U}^T$. Similarly, we could get the disease clustering matrix $\hat{U}$ by substituting D into Equation (7) as follows:

$$\hat{U}_{ij} \leftarrow \hat{U}_{ij} \times \sqrt{\frac{(2RR^T\hat{U})_{ij}}{(\hat{U}\hat{U}^T RR^T\hat{U} + RR^T\hat{U}\hat{U}^T\hat{U})_{ij}}} \tag{10}$$

The reconstructed disease similarity matrix $\tilde{D}$ was then obtained by:

$$\tilde{D} = \hat{D}_P^{-1/2}(\hat{U}\hat{U}^T)\hat{D}_P^{-1/2} \tag{11}$$

Where $\hat{D}_P$ is a diagonal matrix with its (*i*,*i*)-th element equal to the sum of the *ith* row of $\hat{U}\hat{U}^T$.

After $\tilde{M}$ and $\tilde{D}$ were learned, we combined them with existing similarity matrices as follows:

$$SD(i,j) = \begin{cases} \tilde{D}(i,j),\ if D(i,j) = 0 \\ \frac{D(i,j)+\tilde{D}(i,j)}{2},\ otherwise \end{cases} \tag{12}$$

$$SM(i,j) = \begin{cases} \tilde{M}(i,j),\ M(i,j) = 0 \\ \frac{M(i,j)+\tilde{M}(i,j)}{2},\ otherwise \end{cases} \tag{13}$$

Eventually, we obtained the final disease similarity matrix *SD* and miRNA similarity matrix *SM* according to Equation (12) and Equation (13).

Label propagation

After the reconstructed miRNA similarity matrix and disease similarity matrix were obtained, we applied label propagation to predict miRNA-disease associations in miRNA space and disease space, respectively. Generally, a traditional label propagation problem can be presented as follows:

$$Z^{t+1} = \alpha W Z^{t-1} + (1-\alpha)Y \tag{14}$$

where t is the time step and Z^{t+1} represents the iteration results after $t + 1$ steps of label propagation. $\alpha \in (0,1)$ is a hyper-parameter, Y is a binary matrix encoding the initial label information of data points against each class[65]. The label information of the vertices propagates iteratively between adjacent vertices and the propagation process will eventually converge to a unique global optimization quadratic criterion. Equation (14) has a closed-form solution: $Z = (1\text{-}\alpha)(I\text{-}\alpha L)^{-1}Y$, where I is an identity matrix, $L = D^{-1/2}WD^{-1/2}$ is the *Laplacian* matrix of W and D is a diagonal matrix with its (*i*, *i*)-th element equal to the sum of the *i*-th row of W, i.e. $D_{ii} = \sum_j (W_{ij} + W_{ji})/2$.

We will use Equation (14) to update the label of each data object until convergence since the closed-form solution to Equation (14) has high computational complexity due to the matrix inversion operation. Here, 'convergence' means that the predicted labels of unlabeled data does not change in successive iterations. Therefore, we can predict miRNA-disease association from both disease space and miRNA space:

$$FD^{t+1} = \alpha \times SD \times FD^t + (1-\alpha) \times R \tag{15}$$

$$FM^{t+1} = \alpha \times SM \times FM^t + (1-\alpha) \times R^T \tag{16}$$

where *FD* and *FM* represent the prediction results from disease space and miRNA space, respectively. Parameter $\alpha \in (0,\ 1)$ was used to allocate the weight rate of its neighbors while (1-α) represents the probability of receiving its initial label information. The final prediction result *F* was obtained by combining the results from both miRNA space and disease space:

$$F = \beta(FD) + (1-\beta)(FM)^T \tag{17}$$

Parameter β was used to balance the prediction results from disease space and miRNA space, and we simply set β = 0.5. The procedure of GLNMDA is summarized in Algorithm 1. In addition, the source code of GLNMDA could be freely downloaded at https://github.com/ShengPengYu/GLNMDA .

```
Algorithm 1: GLNMDA
Input: Matrices $M \in \mathbb{R}^{m*m}$,$D \in \mathbb{R}^{n*n}$,$R \in \mathbb{R}^{m*n}$, parameter $\alpha$ and $\beta$.
Output: Predicted association matrix $F$.
1. Obtain the $k$ value for miRNAs and diseases by ClusterONE algorithm.
2. Repeat:
     Update $\tilde{U}$and $\hat{U}$ by the following rules:
       $\tilde{U}_{ij} \leftarrow \tilde{U}_{ij} \times \sqrt{\frac{(2R^T R\tilde{U})_{ij}}{(\tilde{U}\tilde{U}^T R^T R\tilde{U}+R^T R\tilde{U}\tilde{U}^T\tilde{U})_{ij}}}$
       $\hat{U}_{ij} \leftarrow \hat{U}_{ij} \times \sqrt{\frac{(2RR^T\hat{U})_{ij}}{(\hat{U}\hat{U}^T RR^T\hat{U}+RR^T\hat{U}\hat{U}^T\hat{U})_{ij}}}$
   Until convergence
3. Obtain the reconstructed similarity matrix $\tilde{M}$and $\tilde{D}$:
     $\tilde{M} = \tilde{M}_P^{-1/2}(\tilde{U}\tilde{U}^T)\tilde{M}_P^{-1/2}$
     $\tilde{D} = \hat{D}_P^{-1/2}(\hat{U}\hat{U}^T)\hat{D}_P^{-1/2}$
4. Integrate similarity information to get $SD$ and $SM$ according to Equation
   (12) and Equation (13).
5. Predict from miRNA space and disease space:
     Repeat:
       $FD^{t+1} = \alpha \times SD \times FD^t + (1-\alpha) \times R$
       $FM^{t+1} = \alpha \times SM \times FM^t + (1-\alpha) \times R^T$
     Until convergence
6. Integrate the results
     $F = \beta(FD) + (1-\beta) \times (FM)^T$
7. Return $F$
```

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

CL was supported by the National Natural Science Foundation of China under Grant No. 61602283 and the Natural Science Foundation of Shandong, China, under Grant No.ZR2016FB10. JWL was supported by the National Natural Science Foundation of China under Grant No. 61572180.

ORCID

Cheng Liang http://orcid.org/0000-0003-3832-0969
Guang-Hui Li http://orcid.org/0000-0001-6531-1166

References

[1] Ardekani AM, Naeini MM. The role of micrornas in human diseases. Avicenna J Med Biotechnol. 2010 Oct;2(4):161–179. PubMed PMID: 23407304; PubMed Central PMCID: PMC3558168.

[2] Miska EA. How microRNAs control cell division, differentiation and death. Curr Opin Genet Dev. 2005 Oct;15(5):563–568. . PubMed PMID: 16099643.

[3] Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res. 2005 Aug 15; 65(16):7065–7070. PubMed PMID: 16103053.

[4] Ambros V. The functions of animal microRNAs. Nature. 2004 Sep 16;431(7006):350–355. . PubMed PMID: 15372042.

[5] Tang W, Wan SX, Yang Z, et al. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics. 2018 Feb 1;34(3):398–406. PubMed PMID: WOS:000423978700006; English.

[6] Blenkiron C, Goldstein LD, Thorne NP, et al. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol. 2007;8(10):R214. . PubMed PMID: 17922911; PubMed Central PMCID: PMC2246288

[7] Zhang X, Zhang X, Wang T, et al. MicroRNA-26a is a key regulon that inhibits progression and metastasis of c-Myc/EZH2 double high advanced hepatocellular carcinoma. Cancer Lett. 2018;426:98–108. .PubMed PMID: WOS:000432877900011; English

[8] Zeng X, Zhang X, Zou Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. Brief Bioinform. 2016 Mar;17(2):193–203. . PubMed PMID: 26059461.

[9] Tang W, Liao ZJ, Zou Q. Which statistical significance test best detects oncomiRNAs in cancer tissues? An exploratory analysis. Oncotarget. 2016 Dec 20;7(51):85613–85623. . PubMed PMID: WOS:000391353200147; English.

[10] Chen X, Yan CC, Zhang X, et al. Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief Bioinform. 2017 Jul;18(4):558–576. PubMed PMID: WOS:000405717400002; English.

[11] Liao ZJ, Li DP, Wang XR, et al. Cancer diagnosis through isomir expression with machine learning method. Curr Bioinf. 2018; 13(1):57–63. PubMed PMID: WOS:000425531200008; English.

[12] Lu M, Zhang Q, Deng M, et al. An analysis of human microRNA and disease associations. Plos One. 2008;3(10):e3420. . PubMed PMID: 18923704; PubMed Central PMCID: PMCPMC2559869.

[13] Zou Q, Li JJ, Song L, et al. Similarity computation strategies in the microRNA-disease network: a survey. Brief Funct Genomics. 2016 Jan;15(1):55–64. PubMed PMID: WOS:000370155900008; English.

[14] Jiang Q, Hao Y, Wang G, et al. Prioritization of disease microRNAs through a human phenome-microRNAome network. Bmc Syst Biol. 2010 May 28;4(Suppl 1):S2. PubMed PMID: 20522252; PubMed Central PMCID: PMCPMC2880408. English.

[15] Xu J, Li CX, Lv JY, et al. Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer. Mol Cancer Ther. 2011 Oct; 10(10):1857–1866. PubMed PMID: 21768329.

[16] Xuan P, Han K, Guo M, et al. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. Plos One. 2013 Aug 8;8(8):e70204. PubMed PMID: 23950912; PubMed Central PMCID: PMCPMC3738541. English.

[17] Chen X, Yan CC, Zhang X, et al. HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. Oncotarget. 2016 Oct 4;7(40):65257–65269. PubMed PMID: WOS:000387281000057; English.

[18] Shi H, Xu J, Zhang G, et al. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. Bmc Syst Biol. 2013 Oct;8(7):101. . PubMed PMID: 24103777; PubMed Central PMCID: PMCPMC4124764. English.

[19] Chen X, Yan CC, Zhang X, et al. WBSMDA: within and between score for miRNA-disease association prediction. Sci Rep. 2016 Feb;16(6):21106. . PubMed PMID: 26880032; PubMed Central PMCID: PMCPMC4754743. English.

[20] Sun DD, Li A, Feng HQ, et al. NTSMDA: prediction of miRNA-disease associations by integrating network topological similarity. Mol Biosyst. 2016;12(7):2224–2232. . PubMed PMID: WOS:000378395000020; English.

[21] You ZH, Huang ZA, Zhu Z, et al. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. Plos Comput Biol. 2017 Mar;13(3):e1005455. PubMed PMID: 28339468; PubMed Central PMCID: PMCPMC5384769. English.

[22] Chen X, Wang LY, Huang L. NDAMDA: network distance analysis for MiRNA-disease association prediction. J Cell Mol Med. 2018 May;22(5):2884–2895. . PubMed PMID: WOS:000430392700032; English.

[23] Jiang QH, Wang GH, Jin SL, et al. Predicting human microRNA-disease associations based on support vector machine. Int J Data Min Bioin. 2013;8(3):282–293. . PubMed PMID: WOS:000324166600002; English.

[24] Chen X, Yan CC, Zhang X, et al. RBMMMDA: predicting multiple types of disease-microRNA associations. Sci Rep. 2015 Sep; 8(5):13877. . PubMed PMID: 26347258; PubMed Central PMCID: PMCPMC4561957. English.

[25] Zou Q, Li J, Hong Q, et al. Prediction of MicroRNA-disease associations based on social network analysis methods. Biomed Res Int. 2015;2015:810514. .PubMed PMID: 26273645; PubMed Central PMCID: PMCPMC4529919. English

[26] Liu YS, Zeng XX, He ZY, et al. Inferring MicroRNA-disease associations by random walk on a heterogeneous network with multiple data sources. Ieee Acm T Comput Bi. 2017 Jul-Aug; 14(4):905–915. PubMed PMID: WOS:000407464700014; English.

[27] Li JQ, Rong ZH, Chen X, et al. MCMDA: matrix completion for MiRNA-disease association prediction. Oncotarget. 2017 Mar 28; 8(13):21187–21199. PubMed PMID: WOS:000397642400057; English.

[28] Chen X, Wu QF, Yan GY. RKNNMDA: ranking-based KNN for MiRNA-disease association prediction. Rna Biology. 2017; 14(7):952–962. . PubMed PMID: WOS:000407258600015; English.

[29] Chen X, Niu YW, Wang GH, et al. MKRMDA: multiple kernel learning-based Kronecker regularized least squares for MiRNA-disease association prediction. J Transl Med. 2017 Dec 12; 15(1):251. PubMed PMID: 29233191; PubMed Central PMCID: PMCPMC5727873. English.

[30] Chen X, Huang L. LRSSLMDA: laplacian regularized sparse subspace learning for mirna-disease association prediction. Plos Comput Biol. 2017 Dec;13(12):e1005912. . PubMed PMID: 29253885; PubMed Central PMCID: PMCPMC5749861. English.

[31] Xiao Q, Luo JW, Liang C, et al. A graph regularized non-negative matrix factorization method for identifying microRNA-disease

associations. Bioinformatics. 2018 Jan 15;34(2):239–248. PubMed PMID: WOS:000419593000008; English.

[32] Zeng XX, Liu L, Lu LY, et al. Prediction of potential disease-associated microRNAs using structural perturbation method. Bioinformatics. 2018 Jul 15;34(14):2425–2432. PubMed PMID: WOS:000438248700012; English.

[33] Chen X, Huang L, Xie D, et al. EGBMMDA: extreme gradient boosting machine for mirna-disease association prediction. Cell Death Dis. 2018 Jan 5;9(1):3. PubMed PMID: 29305594; PubMed Central PMCID: PMCPMC5849212. English.

[34] Chen X, Zhou Z, Zhao Y, ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction. RNA Biol. 2018 May 25:1–12. DOI:10.1080/15476286.2018.1460016. PubMed PMID: 29619882.

[35] Chen X, Xie D, Wang L, et al. BNPMDA: bipartite network projection for miRNA-disease association prediction. Bioinformatics. 2018 Apr 25. DOI:10.1093/bioinformatics/bty333. PubMed PMID: 29701758.

[36] Li Y, Qiu CX, Tu J, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res. 2014 Jan;42(D1):D1070–D1074. PubMed PMID: WOS:000331139800157; English.

[37] Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recogn. 2015 Sep;48(9):2839–2846. . PubMed PMID: WOS:000356112400007; English.

[38] Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J Eval Clin Pract. 2006 Apr;12(2):132–139. . PubMed PMID: 16579821.

[39] Yang Z, Wu LC, Wang AQ, et al. dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. Nucleic Acids Res. 2017 Jan 4;45(D1):D812–D818. PubMed PMID: WOS:000396575500113; English.

[40] Jiang Q, Wang Y, Hao Y, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009 Jan;37(Database issue):D98–104. PubMed PMID: 18927107; PubMed Central PMCID: PMCPMC2686559.

[41] Das SS, Saha P, Chakravorty N. miRwayDB: a database for experimentally validated microRNA-pathway associations in pathophysiological conditions. Database (Oxford). 2018 Jan 1;2018. doi:10.1093/database/bay023. PubMed PMID: 29688364; PubMed Central PMCID: PMCPMC5829561. English.

[42] Ruepp A, Kowarsch A, Schmidl D, et al. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. Genome Biol. 2010 Jan 20;11(1):R6. PubMed PMID: 20089154; PubMed Central PMCID: PMCPMC2847718. English.

[43] Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006 Mar;9(3):189–198. PubMed PMID: 16530703.

[44] Yu SL, Chen HY, Chang GC, et al. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell. 2008 Jan;13(1): 48–57. PubMed PMID: 18167339.

[45] Walker S. Updates in small cell lung cancer treatment. Clin J Oncol Nurs. 2003 Sep-Oct;7(5):563–568. . PubMed PMID: 14603554.

[46] Raponi M, Dossey L, Jatkoe T, et al. MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. Cancer Res. 2009 Jul 15;69(14):5776–5783. PubMed PMID: 19584273.

[47] Seike M, Goto A, Okano T, et al. MiR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. Proc Natl Acad Sci U S A. 2009 Jul 21;106(29):12085–12090. PubMed PMID: 19597153; PubMed Central PMCID: PMCPMC2715493.

[48] Patnaik SK, Kannisto E, Knudsen S, et al. Evaluation of microRNA expression profiles that may predict recurrence of localized stage i non-small cell lung cancer after surgical resection. Cancer Res. 2010 Jan 1;70(1):36–45. PubMed PMID: WOS:000278404300007; English.

[49] Croce CM. Causes and consequences of microRNA dysregulation in cancer. Eur J Cancer. 2012 Jul;48:S8–S9. PubMed PMID: WOS:000313036500033; English.

[50] Inamura K, Ishikawa Y. MicroRNA In Lung Cancer: novel biomarkers and potential tools for treatment. J Clin Med. 2016 Mar 9;5(3). DOI:10.3390/jcm5030036 PubMed PMID: 27005669; PubMed Central PMCID: PMCPMC4810107. English.

[51] Fm Q, Yang L, Xx L, et al. Sequence variation in mature microRNA-499 confers unfavorable prognosis of lung cancer patients treated with platinum-based chemotherapy. Clin Cancer Res. 2015 Apr 1;21(7):1602–1613. PubMed PMID: WOS:000352076700015; English.

[52] Zhu XX, Zhang X, Wang HF, et al. MTA1 gene silencing inhibits invasion and alters the microRNA expression profile of human lung cancer cells. Oncol Rep. 2012 Jul;28(1):218–224. PubMed PMID: WOS:000304638900031; English.

[53] Murakami Y, Yasuda T, Saigo K, et al. Comprehensive analysis of microRNA expression patterns in hepatocellular carcinoma and non-tumorous tissues. Oncogene. 2006 Apr 20;25(17):2537–2545. PubMed PMID: 16331254.

[54] Su H, Yang JR, Xu T, et al. MicroRNA-101, down-regulated in hepatocellular carcinoma, promotes apoptosis and suppresses tumorigenicity. Cancer Res. 2009 Feb 1;69(3):1135–1142. PubMed PMID: 19155302.

[55] Li N, Fu H, Tie Y, et al. miR-34a inhibits migration and invasion by down-regulation of c-Met expression in human hepatocellular carcinoma cells. Cancer Lett. 2009 Mar 8;275(1):44–53. PubMed PMID: 19006648.

[56] Al-Hajj M, Wicha MS, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003 Apr 1;100(7):3983–3988. PubMed PMID: 12629218; PubMed Central PMCID: PMCPMC153034.

[57] Pastrello C, Polesel J, Della Puppa L, et al. Association between hsa-mir-146a genotype and tumor age-of-onset in BRCA1/BRCA2-negative familial breast and ovarian cancer patients. Carcinogenesis. 2010 Dec;31(12):2124–2126. PubMed PMID: WOS:000284953900013; English.

[58] Ogino S, Giannakis M. Immunoscore for (colorectal) cancer precision medicine. Lancet. 2018 May 26;391(10135):2084–2086. . PubMed PMID: WOS:000433257200007; English.

[59] Han Y, Kuang YT, Xue XF, et al. NLK, a novel target of miR-199a-3p, functions as a tumor suppressor in colorectal cancer. Biomed Pharmacother. 2014 Jun;68(5):497–505. PubMed PMID: WOS:000342667800001; English.

[60] Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. Ca-Cancer J Clin. 2017 Jan-Feb;67(1):7–30. . PubMed PMID: WOS:000393807800003; English.

[61] Watanabe A, Tagawa H, Yamashita J, et al. The role of microRNA-150 as a tumor suppressor in malignant lymphoma. Leukemia. 2011 Aug;25(8):1324–1334. PubMed PMID: WOS:000293778900012; English.

[62] Wang D, Wang JA, Lu M, et al. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010 Jul 1;26(13):1644–1650. PubMed PMID: WOS:000278967500010; English.

[63] Qu Y, Zhang HX, Liang C, et al. KATZMDA: prediction of miRNA-disease associations based on katz model. Ieee Access. 2018;6:3943–3950. .PubMed PMID: WOS:000426286900001; English

[64] Zhang W, Qu QL, Zhang YQ, et al. The linear neighborhood propagation method for predicting long non-coding RNA - protein interactions. Neurocomputing. 2018 Jan;17(273):526–534. . PubMed PMID: WOS:000414762100049; English.

[65] Zhang W, Chen Y, Li D. Drug-target interaction prediction through label propagation with linear neighborhood information. Molecules. 2017 Nov 25;22(12). DOI:10.3390/molecules22122056 PubMed PMID: 29186828; English.

[66] Zhu L, Shen JL, Xie L, et al. Unsupervised topic hypergraph hashing for efficient mobile image retrieval. Ieee T Cybernetics. 2017 Nov; 47(11):3941–3954. PubMed PMID: WOS:000413003100037; English.

[67] Zhu L, Shen JL, Jin H, et al. Landmark classification with hierarchical multi-modal exemplar feature. Ieee T Multimedia. 2015 Jul; 17(7):981–993. PubMed PMID: WOS:000356522300006; English.

[68] Wong KC. MotifHyades: expectation maximization for de novo DNA motif pair discovery on paired sequences. Bioinformatics. 2017 Oct 1;33(19):3028–3035. PubMedPMID: WOS:000411514100008; English.

[69] Nepusz T, Yu HY, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods. 2012 May 9;5:471–U81. PubMed PMID: WOS:000303544800024; English.

[70] Li Y, Liang C, Wong KC, et al. Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion. Bioinformatics. 2014 Sep 15;30(18):2627–2635. PubMed PMID: WOS:000342913000012; English.