

RESEARCH ARTICLE

Open Access



# Integrating random walk and binary regression to identify novel miRNA-disease association

Ya-Wei Niu<sup>1</sup>, Guang-Hui Wang<sup>1\*</sup>, Gui-Ying Yan<sup>2</sup> and Xing Chen<sup>3\*</sup> 

## Abstract

**Background:** In the last few decades, cumulative experimental researches have witnessed and verified the important roles of microRNAs (miRNAs) in the development of human complex diseases. Benefitting from the rapid growth both in the availability of miRNA-related data and the development of various analysis methodologies, up until recently, some computational models have been developed to predict human disease related miRNAs, efficiently and quickly.

**Results:** In this work, we proposed a computational model of Random Walk and Binary Regression-based MiRNA-Disease Association prediction (RWBRMDA). RWBRMDA extracted features for each miRNA from random walk with restart on the integrated miRNA similarity network for binary logistic regression to predict potential miRNA-disease associations. RWBRMDA obtained AUC of 0.8076 in the leave-one-out cross validation. Additionally, we carried out three different patterns of case studies on four human complex diseases. Specifically, Esophageal cancer and Prostate cancer were conducted as one kind of case study based on known miRNA-disease associations in HMDD v2.0 database. Out of the top 50 predicted miRNAs, 94 and 90% were respectively confirmed by recent experimental reports. To simulate new disease without known related miRNAs, the information of known Breast cancer related miRNAs was removed. As a result, 98% of the top 50 predicted miRNAs for Breast cancer were confirmed. Lymphoma, the verified ratio of which was 88%, was used to assess the prediction robustness of RWBRMDA based on the association records in HMDD v1.0 database.

**Conclusions:** We anticipated that RWBRMDA could benefit the future experimental investigations about the relation between human disease and miRNAs by generating promising and testable top-ranked miRNAs, and significantly reducing the effort and cost of identification works.

**Keywords:** microRNA, Disease, miRNA-disease association, Random walk, Binary regression

## Background

MicroRNAs (miRNAs) are one category of small single-stranded non-coding RNA molecule (containing 20~25 nucleotides), which function in regulation of gene expression at the posttranscriptional level [1, 2]. Generally, miRNAs could cause mRNAs degradation by binding to the 3' untranslated regions (UTRs) of their target mRNAs [1–5]. Since the first discovery of miRNA about 20 years ago, a plenty variety of miRNAs

have been discovered so far, ranging from nematodes to humans [6–10]. With the in-depth biology research about miRNAs, a vast amount of studies have explicitly shown that miRNAs played important roles in many fundamental biological processes, such as cell growth, proliferation, metabolism, differentiation, apoptosis, signal transduction and so forth [11–15]. In last decades, it was found that the dysregulation of miRNAs could lead to many maladjusted cell behaviors [16], which made miRNAs increasingly be recognized as key regulatory players in gene expression process. Therefore, it's interpretable that many miRNAs have been reported to be related with the development of enormous complex human diseases, including cancers,

\* Correspondence: [ghwang@sdu.edu.cn](mailto:ghwang@sdu.edu.cn); [xingchen@amss.ac.cn](mailto:xingchen@amss.ac.cn)

<sup>1</sup>School of Mathematics, Shandong University, Jinan 250100, China

<sup>3</sup>School of Information and Control Engineering, China University of Mining and Technology, No.1, Daxue Road, Xuzhou 221116, Jiangsu, China

Full list of author information is available at the end of the article



neurological disorders and so on [17–19]. For example, Kliese et al. found that miRNA-145 was downregulated in atypical meningiomas and negatively functioned by regulating the proliferation and motility of meningioma cells [20]. Besides, it was found that in the breast cancer patient, the expression level of miRNA-141 was significantly higher than normal group [21]. What's more, Zhao et al. discovered that miRNA-106a could be seen as an independent biomarker in glioblastoma patients [22]. In addition, compared with normal lymph cells, the expression level of miRNA-19a in canine lymphoid cell lines was obviously increased [23]. Therefore, it's meaningful and uncontroversial to regard disease-related miRNAs as potential biomarkers, which could not only significantly contribute to comprehending the diseases mechanisms, but also benefit the detection, prognosis, diagnosis, treatment and prevention of human complex diseases [24–27]. Nevertheless, the intrinsic disadvantage of traditional experimental method made the identification process of disease-miRNA associations costly and time consuming. Considering the massive increases in the reliability and volume of miRNA-related data based on the accumulated researches about miRNAs, it became necessary and doable to develop effective computational models for predicting potential miRNA-disease associations, which could further enhance the understanding of disease development in miRNA level. More importantly, the promising prediction results of computational approaches could also offer convenience for the follow-up validation experiment by biologic or biomedical researchers [28, 29].

Indeed, having the verified miRNA related data in one hand and the assumption that functionally similar miRNAs are more likely to be associated with phenotypically similar diseases and vice versa in the other, many computational methods have been proposed to predict the underlying miRNA-disease associations in aspect of network science, combinatorial optimization, machine learning, system biology and so on [9, 30–38]. For example, Jiang et al. [24] proposed a computational model based on hypergeometric distribution to predict novel miRNA-disease associations. They firstly constructed some classic network models, such as disease phenotypical similarity network, miRNA functional similarity network and known phenome-miRNAome network according to multi-source biological data. Then they integrated all the networks to finally prioritize the human miRNAs for diseases of interest. However, the strong dependence on the miRNA-target interactions resulted in a high rate of false positive result of the method. Xu et al. [39] investigated the expression profiles of miRNAs and proposed a computational model, in which they constructed miRNA target-dysregulated network to extract pivotal feature

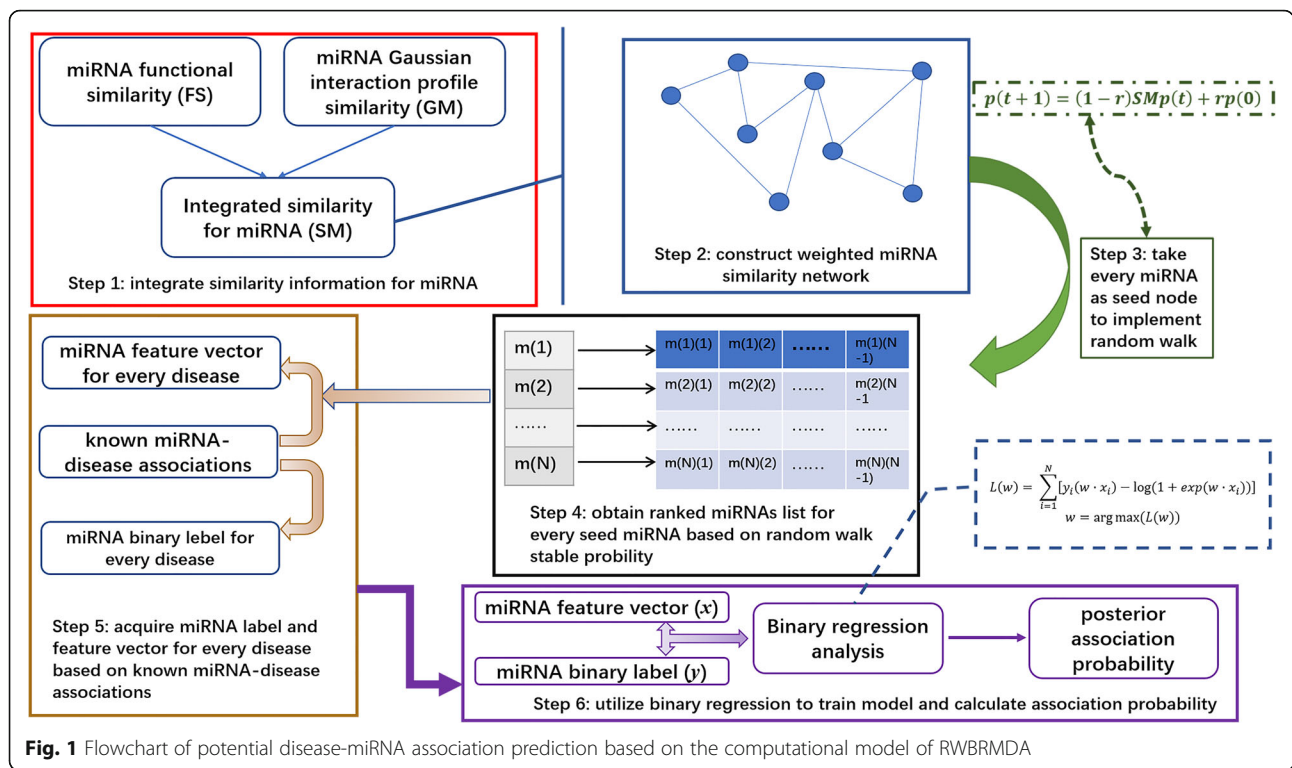
vectors for miRNAs. Support vector machine (SVM) was then conducted in their model to distinguish positive disease-related miRNAs from negative ones. However, the difficulty of obtaining negative disease-related miRNAs made the model have very narrow applications. Differing from traditional local network similarity measures, Chen et al. [40] utilized the global network similarity measures and proposed the Random Walk with Restart for MiRNA–Disease Association (RWRMDA) model. In this model, they constructed the global miRNA functional similarity network, on which they further implemented random walk with restart. Based on the stationary state of the random walk dynamic process, namely the association probability of each disease-miRNA pair, authors finally prioritized candidate miRNAs for diseases investigated. Likewise, focusing on the functional connections between miRNA targets and disease genes in protein-protein interaction (PPI) networks, Shi et al. [41] identified potential miRNA-disease associations by performing random walk on the PPI network. Meanwhile, Mørk et al. [42] proposed the computational model of miRPD (miRNA-Protein-Disease), in which they did network analysis on both of the known protein-miRNA associations and the text mined disease-protein associations to infer miRNA–disease associations. However, these models also strongly relied on the interactions of miRNA and target with a high rate of false-positive results. MirAI model was proposed by Pasquier et al. [43] in which they represented different types of miRNA-related data, such as miRNA-disease associations information, miRNA-neighbor associations information, miRNA-target associations information, miRNA-word associations information and miRNA-family associations information, into a high-dimensionality vector space to further predicted the potential disease-miRNA association information. Obviously, the suitable choice of dimensionality was of great importance for the prediction performance. However, in their model there was no optimal dimension given. Recently, Chen et al. [44] proposed a Bipartite Network Projection for MiRNA–Disease Association prediction (BNPMDA) model based on integrated miRNA and disease similarity and the known miRNA–disease associations. They firstly defined the preference degree for miRNAs and diseases with the bias ratings. Then, bipartite network-based recommendation algorithm was implemented based on resource allocation process between miRNAs and diseases to predict the potential miRNA–disease associations.

Meanwhile, there also some other machine learning-based models be successively put forward later. For example, Chen et al. [45] developed the model of Regularized Least Squares for MiRNA-Disease Association (RLSMDA), which needed no negative samples resulting from the characteristic of semi-supervised learning. It's worth pointing out that RLSMDA could be conducted for

diseases without any known miRNA associations. Additionally, Xuan et al. [46] proposed a HDMP method to predict potential disease-miRNA associations based on weighted  $k$  most similar neighbors. In this model, they figured out the miRNA family and cluster information and recalculated miRNA functional similarity by endowing higher weight to miRNAs in the same family or cluster. However, the chosen number of neighbors would influence the prediction performance of the computational model to some extent. Considering that the traditional similarity-based  $k$ -nearest-neighbors (KNN) method was lazy learning and not reliable enough, Chen et al. [47] proposed a model of Ranking-based KNN for MiRNA-Disease Association prediction (RKNNMDA). In this model, to solve the limitation of normal ranking method, they firstly took use of SVM method via learning features from training data. Then, based on Hamming loss metric, they reranked the similarity-based sorted neighbors to obtain better prediction results. Furthermore, Chen et al. [48] proposed the first model that could infer the association types of disease-miRNA associations, namely the computational model of Restricted Boltzmann Machine for Multiple types of MiRNA-Disease Association prediction (RBMMDA). It's no doubt that the biology information about the different types of disease-miRNA associations obtained from RBMMDA could benefit the understanding about the mechanism of diseases in the level of miRNAs. To further enhance the prediction performance, Chen et al. [49] then developed the model of Within and Between Score for MiRNA-Disease Association prediction (WBSMDA). This model was aimed to predict potential miRNAs related with plethora of human complex diseases by integrating the miRNA and disease Gaussian interaction profile kernel similarity, miRNA functional similarity, disease semantic similarity and also the known miRNA-disease associations. WBSMDA could also be utilized for new diseases and new miRNAs without any known relation information. Soon after, by integrating the biological dataset involved in WBSMDA into a heterogeneous graph, Chen et al. [50] further proposed another method named Heterogeneous Graph Inference for MiRNA-Disease Association prediction (HGIMDA). HGIMDA calculated the disease-miRNA association possibility by investigating all the 3-length paths in the constructed heterogeneous graph. HGIMDA obtained better prediction performance in terms of cross validation compared with most of previously mentioned models. Recently, Li et al. [51] presented a model of Matrix Completion for MiRNA-Disease Association prediction (MCMDA) using matrix completion algorithm to predict the potential miRNA-disease associations. In this model, they constructed initial matrix according to known miRNA-disease associations. Singular value threshold (SVT) algorithm was then implemented in the matrix

completion process. The prediction scores were immediately calculated after they finished the matrix completion. By maximizing network information flow of the phenome-microRNAome network, Yu et al. [52] designed a combinatorial prioritization algorithm and proposed an computational model named MaxFlow to discover new disease-miRNA associations. Nowadays, Chen et al. [53] presented a model named Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction (EGBMMDA), which was the first decision tree learning-based model for predicting novel miRNA-disease association. In this model, they constructed informative feature vector to train a regression tree under the gradient boosting framework built on the graph theoretical measures, statistical measures and matrix factorization outcomes for all the miRNA-disease pairs. Lately, in the literature review by Chen et al. [54] about miRNA-disease association prediction, 20 state-of-the-art in silico models were introduced from different perspectives. The authors summarized the existing difficulties in potential disease-miRNA association prediction task and pointed out five feasible and meaningful research schemas for further development of computational model designment in this field.

In this work, we presented a Random Walk and Binary Regression-based MiRNA-Disease Association prediction (RWBRMDA) method to predict underlying miRNA-disease associations. Specifically, we constructed an integrated miRNA similarity network based on miRNA functional similarity and miRNA Gaussian similarity. Then we implemented random walk with restart on the integrated miRNA similarity network for every miRNA in turn. Thirdly, we extracted feature vector for every miRNA according to the results of the random walk and the known miRNA-disease associations. Next, considering the field information about known disease-miRNA associations, we labelled 1 to those miRNAs with known associations with currently investigated disease, otherwise 0. Finally, we employed binary logistic regression method based on the feature vectors and label information to predict miRNAs for diseases of interest (See Fig. 1). Furthermore, we implemented Leave-one-out cross validation (LOOCV) for RWBRMDA. As a result, RWBRMDA obtained AUC value of 0.8076. What's more, we carried out three different patterns of case studies in this work. Generally, in three types of case studies, we respectively evaluated the prediction performance of RWBRMDA for complex human diseases with miRNA associations recorded in HMDD v2.0 database [55], new diseases without any known related miRNAs and known diseases with miRNA associations recorded in HMDD v1.0 database [19]. By validating the prediction results based on other two important databases, miR2Disease [56] and dbDEMC [57], RWBRMDA obtained high confirmation ratios of the predicted



miRNAs in all three ways of case studies. Therefore, it showed the effectivity of RWBRMDA in predicting potential miRNA-disease associations for various categories of diseases.

## Results

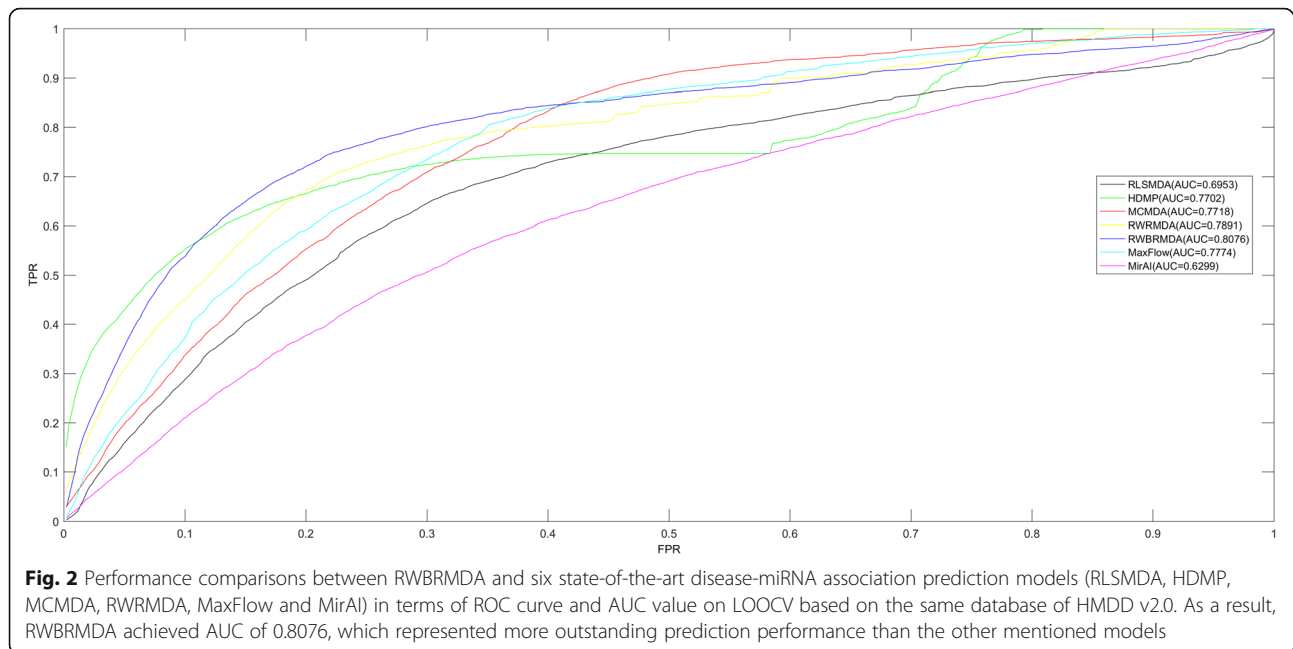
### Performance evaluation

Leave-one-out cross validation (LOOCV) is often utilized to evaluate the prediction performance of computational model. In this work, LOOCV was implemented as follows: for an investigated disease, based on the records in HMDD v2.0 [55] database, each known disease-related miRNA was left out in turn as test sample and the other known disease-related miRNAs were regarded as seed samples. Then, current test sample and candidate samples, namely the miRNAs without known association with the investigated disease would be ranked according to the prediction score of the model. If the test sample was ranked above the given threshold, the model would be considered to successfully predict this miRNA-disease association. Further, Receiver operating characteristics (ROC) curve could be drawn by plotting the true positive rate (TPR) versus the false positive rate (FPR) at different thresholds. Generally, the area under the ROC curve (AUC) is calculated and utilized to evaluate the prediction performance. Specifically,  $AUC = 1$  means the best prediction performance and  $AUC = 0.5$  indicates a random performance. As a result, RWBRMDA obtained the AUC value of 0.8076, which

was higher than some previously mentioned computational models (RLSMDA: 0.6953 [45], HDMP:0.7702 [46], MCMMDA:0.7718 [51], RWRMDA:0.7891 [40], MaxFlow:0.7774 [52], MirAI:0.6299 [43]) as shown in Fig. 2. It should be mentioned that we repeated all the 6 comparison methods based on the same HMDD v2.0 database, drew the corresponding ROC curves and compared the AUC values. In particular, the AUC value of MirAI seemed relatively small because the collaborative filtering technology utilized in this model was influenced by the sparsity problem of the biological data. Therefore, to some extent, RWBRMDA obtained better performance in the prediction of potential miRNA-disease associations.

### Case studies

As mentioned before, we carried out three different patterns of case studies in this work. Specifically, one approach was that we implemented RWBRMDA for disease-miRNA associations prediction based on the known diseases-miRNAs associations recorded in HMDD v2.0 database [55], then we verified the prediction results based on another two important miRNA-disease association databases, miR2Disease [56] and dbDEMC database [57]. The second approach was that we removed all the original miRNA associations information of the investigated disease, and then we verified the prediction results of the disease based on HMDD v2.0 database, miR2Disease and dbDEMC database. This method aimed to test the prediction performance for a new disease without any known



associations. The third approach was that we used the diseases-miRNAs associations recorded in HMDD v1.0 database [19], then we verified the prediction results of some complex diseases based on HMDD v2.0 database, miR2Disease, and dbDEMC database. This approach aimed to assess the prediction robustness on different datasets of the computational model.

Case studies on Esophageal cancer and Prostate cancer were implemented in the first way. Esophageal cancer is a kind of cancer arising from the esophagus and it was reported as the sixth deadly cancers and the eighth most common cancer worldwide [58]. Statistical analysis showed that it was three to four times more common in male than female [59]. The treatment on esophageal cancer is strongly dependent on the cancer's stage. There was clinic research showing that the survival rate could increase to 90% if the tumors could be diagnosed at an early stage [60]. Therefore, it's obvious that the early detection of esophageal cancer is vital to cancer treatment [61, 62]. Some miRNAs have been confirmed to be related with esophageal cancer. For example, the relative expressions of miRNA-155, miRNA-183, and miRNA-20a in esophageal tissue were found to be significantly associated with increased risk for esophageal cancer [63]. In the case study for esophageal cancer, candidate miRNAs, namely miRNAs without known association with esophageal cancer in HMDD v2.0 database, were prioritized according to the scores obtained from RWBRMDA. As a result, 10 out of top 10, 47 out of top 50 were confirmed by recent experimental results recorded in miR2Disease and dbDEMC (See Table 1).

Prostate cancer develops in the epithelial cells of prostate, the cancer cells of which might spread from the prostate to other parts of the body, particularly the bones and lymph nodes [64]. Prostate cancer was reported to be the second leading cause of cancer-related death among men in developed countries [65]. Up to now, lots of miRNAs have been confirmed to be related to prostate cancer. For instance, it was reported that miRNA-183 expression was significantly higher in prostate cancer cells and tissues, compared with that in matched normal prostate cells and tissues [66]. It meant that the inhibition of miRNA-183 expression might be therapeutically beneficial for prostate cancer treatment [66]. Taking prostate cancer as a case study to implement RWBRMDA for potential miRNA-disease association prediction, for the top 10 and top 50 potential prostate cancer associated miRNAs, 10 and 45 of them were respectively confirmed to have experimental literature evidences recorded in miR2Disease and dbDEMC database (See Table 2). For example, miRNA-29b was ranked the second by RWBRMDA and it was the highest ranked miRNA, simultaneously confirmed by both miR2Disease and dbDEMC database. In fact, miRNA-29b was down-regulated from research about miRNA expression profiling of prostate cancer cell lines [67].

We conducted case study on Breast cancer by way of the second case study method, in which we removed all the related miRNAs information of breast cancer to model the situation where a new disease without known miRNA associations was investigated. Breast cancer is known as the most leading type of cancer in women worldwide, accounting for about 25% of all the female's

**Table 1** We implemented RWBRMDA on esophageal cancer for potential disease-related miRNA prediction and conducted the first pattern of case study, in which the disease-miRNA associations recorded in HMDD v2.0 were used as training samples and miRNAs without known associations with currently considered diseases were regarded as test samples. According to the prediction results, among the top 10 and 50 potential esophageal cancer related miRNAs, 10 and 47 were confirmed by miR2Disease and dbDEMC databases

miRNA	Evidence	miRNA	Evidence
hsa-mir-29b	dbDEMC	hsa-mir-18a	dbDEMC
hsa-mir-1	dbDEMC	hsa-mir-221	dbDEMC
hsa-mir-19b	dbDEMC	hsa-mir-7	dbDEMC
hsa-mir-16	dbDEMC	hsa-mir-15b	dbDEMC
hsa-let-7e	dbDEMC	hsa-mir-106a	dbDEMC
hsa-mir-29a	dbDEMC	hsa-mir-218	unconfirmed
hsa-let-7d	dbDEMC	hsa-mir-10b	dbDEMC
hsa-mir-106b	dbDEMC	hsa-mir-132	dbDEMC
hsa-mir-146b	dbDEMC	hsa-mir-30c	dbDEMC
hsa-mir-222	dbDEMC	hsa-mir-429	dbDEMC
hsa-mir-24	dbDEMC	hsa-mir-93	dbDEMC
hsa-mir-200b	dbDEMC	hsa-mir-199b	dbDEMC
hsa-mir-181b	dbDEMC	hsa-mir-124	dbDEMC
hsa-let-7f	unconfirmed	hsa-mir-107	dbDEMC, miR2Disease
hsa-mir-181a	dbDEMC	hsa-mir-133b	dbDEMC
hsa-let-7i	dbDEMC	hsa-mir-23b	dbDEMC
hsa-mir-9	dbDEMC	hsa-mir-127	dbDEMC
hsa-let-7 g	dbDEMC	hsa-mir-206	dbDEMC
hsa-mir-125b	dbDEMC	hsa-mir-20b	dbDEMC
hsa-mir-17	dbDEMC	hsa-mir-122	unconfirmed
hsa-mir-195	dbDEMC	hsa-mir-224	dbDEMC
hsa-mir-142	dbDEMC	hsa-mir-18b	dbDEMC
hsa-mir-182	dbDEMC	hsa-mir-27b	dbDEMC
hsa-mir-30a	dbDEMC	hsa-mir-373	dbDEMC, miR2Disease
hsa-mir-125a	dbDEMC	hsa-mir-302b	dbDEMC

death cases all over the world [68]. Some researches on breast cancer have confirmed that many miRNAs were associated with breast cancer. For example, microarray-based miRNA profiling on whole blood of early stage breast cancer patients showed that miRNA-106b was up-regulated in whole blood of breast cancer patients [69]. What's more, it was found that downregulation of miRNA-140 promoted cancer stem cell formation in basal-like early stage breast cancer [70]. We verified the predicted underlying breast cancer related miRNAs obtained by RWBRMDA. Consequently, 10 out of the top 10 and 49 out of the top 50 predicted miRNAs were experimentally confirmed by HMDD v2.0, miR2Disease and dbDEMC database (See Table 3).

Lymphoma often refers to a group of cancerous blood cell tumors that developed from lymphocytes [71]. Worldwide, lymphoma was reported to be the seventh-most common cancer and also be the third-most common cancer in children [72]. Benefitting from the development of deep sequencing technology, several miRNAs have been discovered to be related with lymphomas. For example, miRNA-155, miRNA-21 and miRNA-221 were observed over-expressed in lymphoma cell lines [73]. In order to test the prediction robustness of RWBRMDA in different datasets, we conducted the third way of case study on lymphoma, in which we only used the known disease-related miRNAs recorded in HMDD v1.0 database as training samples and used associations in HMDD v2.0 database, miR2Disease, and dbDEMC database as test datasets. As a result, 10 out of the top 10 and 44 out of the top 50 predicted miRNAs were confirmed based on the three test datasets (See Table 4). For instance, miRNA-29c, which was the highest ranked miRNA confirmed by dbDEMC and HMDD v2.0 databases, was reported to show down-regulation in lymphoma cells [74].

In conclusion, the promising results obtained from LOOCV and case studies in three different ways have demonstrated the reliable prediction performance of RWBRMDA. Therefore, we further prioritized candidate miRNAs for all the diseases recorded in HMDD v2.0 database. The predicted ranks of miRNAs for each disease were publicly released for further experimental validation (Additional file 1). The potential disease-miRNA associations with relatively high ranks were expected to be confirmed by clinical observation or biological experiments in the future.

## Discussion

Several important factors contributed to the excellent performance of RWBRMDA. Firstly, benefitting from the valid and updated disease-miRNA association data by abundant biology researches, RWBRMDA could have more chance to obtain higher prediction accuracy. Secondly, RWBRMDA took full advantage of the similarity information of the miRNA functional similarity and Gaussian interaction profile kernel similarity to obtain integrated global similarity network for miRNAs. Generally, the more similarity information was utilized, the better prediction performance would be. Thirdly, based on the previously mentioned similarity information, RWBRMDA further implemented random walk with restart, an effective and widely used method, to investigate global reachability between any pair of miRNAs. A higher stable probability meant a higher reachability between two miRNAs or meant a higher association probability with the same disease of these two miRNAs. Then according to the random walk

**Table 2** We also conducted the first pattern of case study on prostate cancer by RWBRMDA. As a result, among the top 10 and 50 potential prostate cancer related miRNAs, 10 and 45 were confirmed by miR2Disease and dbDEMC databases

miRNA	Evidence	miRNA	Evidence
hsa-mir-146a	miR2Disease	hsa-mir-34a	dbDEMC, miR2Disease
hsa-mir-29b	dbDEMC, miR2Disease	hsa-mir-34c	dbDEMC
hsa-mir-1	dbDEMC	hsa-mir-200b	unconfirmed
hsa-mir-223	dbDEMC, miR2Disease	hsa-mir-155	dbDEMC
hsa-mir-21	dbDEMC, miR2Disease	hsa-mir-181b	dbDEMC, miR2Disease
hsa-mir-126	dbDEMC, miR2Disease	hsa-let-7f	dbDEMC, miR2Disease
hsa-let-7a	dbDEMC, miR2Disease	hsa-mir-133a	dbDEMC
hsa-let-7b	dbDEMC, miR2Disease	hsa-mir-181a	dbDEMC, miR2Disease
hsa-mir-19b	dbDEMC, miR2Disease	hsa-mir-196a	dbDEMC
hsa-mir-29c	dbDEMC	hsa-let-7i	dbDEMC
hsa-let-7c	dbDEMC, miR2Disease	hsa-mir-9	dbDEMC
hsa-mir-16	dbDEMC, miR2Disease	hsa-let-7 g	dbDEMC, miR2Disease
hsa-mir-15a	dbDEMC, miR2Disease	hsa-mir-34b	dbDEMC
hsa-mir-143	dbDEMC, miR2Disease	hsa-mir-17	miR2Disease
hsa-let-7e	dbDEMC	hsa-mir-203	unconfirmed
hsa-mir-29a	dbDEMC, miR2Disease	hsa-mir-195	dbDEMC, miR2Disease
hsa-mir-199a	dbDEMC, miR2Disease	hsa-mir-205	dbDEMC, miR2Disease
hsa-mir-150	dbDEMC	hsa-mir-142	unconfirmed
hsa-let-7d	dbDEMC, miR2Disease	hsa-mir-182	dbDEMC, miR2Disease
hsa-mir-106b	dbDEMC	hsa-mir-30a	miR2Disease
hsa-mir-210	miR2Disease	hsa-mir-101	dbDEMC, miR2Disease
hsa-mir-146b	unconfirmed	hsa-mir-200a	dbDEMC
hsa-mir-222	dbDEMC, miR2Disease	hsa-mir-19a	dbDEMC
hsa-mir-24	dbDEMC, miR2Disease	hsa-mir-125a	dbDEMC, miR2Disease
hsa-mir-200c	dbDEMC	hsa-mir-18a	unconfirmed

result we could extract more reliable feature vector for every miRNA as the input of next binary logistic regression. More reliable and valuable feature vector would be help for a better output of the binary logistic regression. In other words, integrating random walk and binary logistic regression was an innovative and efficient research practice.

There were also some limitations in RWBRMDA. Firstly, because we took use of binary logistic regression in the model, we needed prior association label information for investigated miRNAs. If the known association information was too little or none, the AUC value of RWBRMDA might be a little lower. Secondly, RWBRMDA partly depended on the parameters used in our model, such as the restart probability in random walk and the length of the feature vector of miRNA. Hence, a technical analysis for selecting appropriate and optimized parameter values was necessary when RWBRMDA was conducted based on other biology dataset.

## Conclusions

Identifying potential miRNA-disease associations was vitally important for investigating the biomarker of disease diagnosis at the miRNA level. Based on the fundamental hypothesis that functionally similar miRNAs greatly tended to be relevant to phenotypically similar diseases and vice versa, in this work, we introduced a computational model named RWBRMDA to predict underlying miRNA-disease associations. RWBRMDA was developed mainly based on random walk with restart and binary logistic regression. The known miRNA-disease association information in HMDD v2.0 database was utilized to assign prior label to miRNAs for any disease we investigated. Considering that the network modeling was a primitive and intuitive way for modeling biological data, we also took use of miRNA functional similarity, Gaussian interaction profile kernel similarity for miRNAs and integrated similarity for miRNAs to map miRNAs to a weighted network. We complemented random walk with restart on the constructed network for every

**Table 3** We conducted case study of breast cancer in the second way by RWBRMDA, in which we removed all the known breast cancer related miRNAs to simulate a new disease without any known associations. Then we verified the prediction results based on HMDD v2.0 database, miR2Disease and dbDEMOC database. As a result, among the top 10 and 50 potential miRNAs, 10 and 49 were confirmed

miRNA	Evidence	miRNA	Evidence
hsa-mir-146a	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-146b	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-21	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-221	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-1	dbDEMOC,HMDD v2.0	hsa-mir-210	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-29b	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-24	dbDEMOC,HMDD v2.0
hsa-let-7a	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-200b	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-155	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-106b	dbDEMOC,HMDD v2.0
hsa-mir-145	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-18a	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-223	dbDEMOC,HMDD v2.0	hsa-mir-181a	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-126	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-34c	dbDEMOC,HMDD v2.0
hsa-mir-19b	dbDEMOC,HMDD v2.0	hsa-mir-19a	dbDEMOC,HMDD v2.0
hsa-let-7b	dbDEMOC,HMDD v2.0	hsa-mir-181b	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-29a	dbDEMOC,HMDD v2.0	hsa-mir-200c	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-16	dbDEMOC,HMDD v2.0	hsa-let-7f	dbDEMOC,miR2Disease,HMDD v2.0
hsa-let-7c	dbDEMOC,HMDD v2.0	hsa-mir-133a	dbDEMOC,HMDD v2.0
hsa-mir-17	miR2Disease,HMDD v2.0	hsa-mir-196a	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-15a	dbDEMOC,HMDD v2.0	hsa-let-7i	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-29c	dbDEMOC,miR2Disease,HMDD v2.0	hsa-let-7 g	dbDEMOC,HMDD v2.0
hsa-mir-34a	dbDEMOC,HMDD v2.0	hsa-mir-203	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-143	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-142	unconfirmed
hsa-let-7e	dbDEMOC,HMDD v2.0	hsa-mir-9	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-150	dbDEMOC	hsa-mir-195	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-199a	dbDEMOC,HMDD v2.0	hsa-mir-205	dbDEMOC,miR2Disease,HMDD v2.0
hsa-mir-125b	miR2Disease,HMDD v2.0	hsa-mir-92a	HMDD v2.0
hsa-mir-222	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-34b	dbDEMOC,HMDD v2.0
hsa-let-7d	dbDEMOC,miR2Disease,HMDD v2.0	hsa-mir-182	dbDEMOC,miR2Disease,HMDD v2.0

miRNA to obtain the global feature vector of miRNA, which was used for binary logistic regression with the known prior label information to calculate the posterior association probability of investigated disease-miRNA pairs (See Fig. 1). Both cross validation result (AUC = 0.8076) and three different kinds of case study on esophageal cancer (94%), prostate cancer (90%), breast cancer (98%) and lymphoma (88%) have demonstrated the reliable prediction ability of RWBRMDA. Therefore, RWBRMDA was anticipated to be valuable for further research on miRNA-disease associations and be beneficial to human disease diagnosis, treatment, prevention and prognosis.

## Methods

### Human miRNA-disease associations

In this study, we take use of human disease-miRNA associations in HMDD v2.0 database [55], which records 5430 known miRNA-disease associations with respect to 383 human diseases and 495 miRNAs. Technically, we

could construct the adjacent matrix  $A$  to clearly describe the relation of each disease-miRNA pairs. Specifically, if miRNA  $m(i)$  is confirmed to be related to disease  $d(j)$  in the database, the entry  $A(i,j)$  is defined as 1, otherwise 0. Finally, 5430 entries of matrix  $A$  are assigned 1, the rest ones are assigned 0.

### MiRNA functional similarity

Based on the basic assumption that miRNAs with similar function are more likely to be related to semantically similar diseases and vice versa, miRNA functional similarity have been calculated by Wang et al [32]. In our study, owing to their relevant researches, we obtain the miRNA functional similarity information from <http://www.cuilab.cn/files/images/cuilab/misim.zip>. Furthermore, we construct the miRNA functional similarity matrix  $FS$  to store the data, where the entry  $FS(i,j)$  describes the functional similarity between miRNA  $m(i)$  and miRNA  $m(j)$ .



**Table 4** Lymphoma was conducted as a case study in the third way, in which we only used known disease-miRNA association based on HMDD v1.0 database as test samples to assess the robustness of the prediction model in the different datasets, and then we verified the prediction results according to HMDD v2.0 database, miR2Disease, and dbDEMC database. As a result, among the top 10 and 50 potential lymphoma related miRNAs, 10 and 44 were confirmed

miRNA	Evidence	miRNA	Evidence
hsa-mir-223	dbDEMC	hsa-mir-200b	dbDEMC,HMDD v2.0
hsa-mir-15b	dbDEMC	hsa-mir-199a	dbDEMC
hsa-mir-29c	dbDEMC,HMDD v2.0	hsa-mir-95	dbDEMC
hsa-mir-106a	dbDEMC	hsa-mir-183	dbDEMC
hsa-mir-146a	dbDEMC,HMDD v2.0	hsa-let-7e	dbDEMC
hsa-mir-99b	dbDEMC	hsa-mir-141	dbDEMC
hsa-mir-100	dbDEMC	hsa-let-7c	dbDEMC
hsa-mir-145	dbDEMC	hsa-let-7a	dbDEMC
hsa-mir-143	dbDEMC	hsa-mir-128b	unconfirmed
hsa-mir-155	dbDEMC,HMDD v2.0	hsa-mir-21	dbDEMC,HMDD v2.0
hsa-mir-222	dbDEMC	hsa-mir-29b	dbDEMC
hsa-let-7 g	dbDEMC	hsa-mir-34c	unconfirmed
hsa-mir-101	dbDEMC,HMDD v2.0	hsa-mir-214	dbDEMC
hsa-mir-224	dbDEMC	hsa-mir-127	dbDEMC
hsa-mir-34a	dbDEMC	hsa-let-7b	dbDEMC
hsa-mir-146b	unconfirmed	hsa-mir-132	dbDEMC
hsa-mir-221	dbDEMC	hsa-mir-137	dbDEMC
hsa-mir-125b	unconfirmed	hsa-mir-376c	unconfirmed
hsa-let-7i	dbDEMC	hsa-mir-181b	dbDEMC
hsa-mir-203	dbDEMC,HMDD v2.0	hsa-mir-139	dbDEMC,HMDD v2.0
hsa-mir-126	dbDEMC,HMDD v2.0	hsa-mir-122	dbDEMC,HMDD v2.0
hsa-mir-335	dbDEMC	hsa-mir-31	dbDEMC
hsa-mir-196b	unconfirmed	hsa-mir-9	dbDEMC
hsa-mir-140	dbDEMC	hsa-mir-181d	dbDEMC
hsa-mir-191	dbDEMC	hsa-mir-206	dbDEMC

**Gaussian interaction profile kernel similarity**

Thanks to a kind of widely used Gaussian kernel function, which named Radial Basis Function (RBF), Gaussian interaction profile kernel similarity could be calculated and put into use for prediction task [75]. The interaction profile of miRNA  $m(i)$  could be expressed built on the adjacency matrix  $A$ . Specifically, based on the binary vector  $IP(m(i))$ , namely the  $i$ th row of the adjacency matrix  $A$ , Gaussian kernel similarity between miRNA  $m(i)$  and  $m(j)$  could be obtained:

$$GM(m(i), m(j)) = \exp(-\gamma_m \|IP(m(i)) - IP(m(j))\|^2) \tag{1}$$

where  $r_m$  is used to control bandwidth of the kernel,

$GM$  is denoted as the Gaussian interaction profile kernel similarity matrix for miRNAs. What's more,  $r_m$  could be calculated by normalizing a new bandwidth parameter  $r'_m$  by the average number of known associations with diseases per miRNA as follows:

$$\gamma_m = \frac{\gamma'_m}{\left(\frac{1}{n_m} \sum_{i=1}^{n_m} \|IP(m(i))\|^2\right)} \tag{2}$$

where  $n_m$  is the number of all the miRNAs investigated. In this article,  $r'_m$  is set 1 based on previous studies [76, 77].

**Integrated similarity for miRNAs**

Integrated miRNA similarity between miRNAs  $m(i)$  and  $m(j)$  is calculated based on the miRNA functional similarity and Gaussian interaction profile kernel similarity for miRNA [49] as follows, and  $SM$  is defined as the integrated miRNA similarity matrix:

$$SM(m(i), m(j)) = \begin{cases} FS(m(i), m(j)) & m(i) \text{ and } m(j) \text{ has functional similarity} \\ GM(m(i), m(j)) & \text{otherwise} \end{cases} \tag{3}$$

**RWBRMDA**

In this work, we propose a computational model of RWBRMDA by integrating known miRNA-disease associations, miRNA functional similarity and Gaussian interaction profile kernel similarity for miRNAs (See Fig. 1) motivated by study in [78, 79]. It's known that random walk could be used to rank the relation probability for the nodes in a network. Binary regression could be used for classification problems or prediction problems. We implement random walk with restart for every miRNA on the integrated miRNA similarity network to obtain corresponding feature vector of the investigated miRNA. Based on the feature vectors of miRNAs and the known miRNA-disease associations, we could assign the binary label 0 or 1 to every miRNA for the given disease. Then we utilize binary regression to predict the association probability between the miRNA with label 0 and the corresponding disease of interest.

Technically, based on the known miRNA-disease associations in HMDD v2.0, we have constructed the adjacent matrix  $A$ . According to the integrated similarity matrix  $SM$  for miRNAs, we construct a weighted miRNAs relation network, which consists of 495 miRNA nodes. The weight of pairwise miRNAs in the network is assigned their integrated similarity value in the  $SM$ . Random walk with restart is then implemented on the weighted network, taking every miRNA node as start node in turn. Specifically, every miRNA node is considered as seed node for one time of random walk with restart. Other miRNA nodes are considered as candidate

nodes. For a seed miRNA  $m(i)$ , the initial probability  $p(0)$  is the normalized  $i$ th row of matrix  $SM$ . Here we define the restart probability of random walk at source nodes as  $r$  ( $0 < r < 1$ ) in every time step. Then a vector  $p(t)$  could be defined in which the  $j$ th element meant the probability of finding the walker at node  $j$  at step  $t$ . Finally, the random walk process could be defined as follows:

$$p(t + 1) = (1-r)SMp(t) + rp(0) \tag{4}$$

Random walk would finally reach the stable state after some steps. We call the random walk reach the stationary stage if the change between  $p(t)$  and  $p(t + 1)$  is less than a cutoff (here we chose  $10^{-6}$  as the cutoff) measured by  $L1$  norm. When the random walk reaches the stable state, the candidate miRNAs for the seed miRNA  $m(i)$  could be ranked built on the stable probability of  $p_{\infty}(m(i))$ . Generally, after 495 times random walk on the weighted miRNAs relation network, we could obtain the corresponding ranked candidate miRNAs sequence or list for every seed miRNA, which we call the global relationship information of every miRNA. In previous model of RWRMDA [40], the author also uses random walk with restart. While, there exists many differences between the implementation progress of our model and the implementation progress of RWRMDA model. First, the motivation of random walk is different. In RWRMDA, random walk is used directly to predict disease-related miRNAs, which means they aim to mine the pair-wise relationship of miRNA and disease. In current work of RWBRMDA, we utilize random walk to seek the relationship between miRNAs, which is more suitable because the random walk process is conducted on the miRNA similarity network. Second, the choices of seed nodes are different. In RWRMDA, they choose known disease-related miRNAs as seed nodes, while in this work we take every miRNA in turn as seed node, which is more practical in cases where the field knowledge is short. In principle, the aims of random walk in these two works were different.

Next, for an arbitrary disease  $d(j)$ , the  $j$ th column of adjacency matrix  $A$  is regarded as the binary label vector of all the miRNAs with respect to disease  $d(j)$ . Binary logistic regression is then conducted to calculate the posterior association probability of those miRNAs with label 0 to  $d(j)$  as follows:

$$P(y = 1|x) = \frac{\exp(w \cdot x)}{1 + \exp(w \cdot x)} \tag{5}$$

$$P(y = 0|x) = \frac{1}{1 + \exp(w \cdot x)} \tag{6}$$

where  $y$  is the binary label,  $w$  is the weight vector, which

needs to be trained,  $w \cdot x$  is the inner product of vector  $w$  and vector  $x$ .

Given a training set of  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ , where  $x_i \in R^n$ ,  $y_i \in \{0, 1\}$  and  $N$  is the number of samples, we could train the parameter  $w$  by maximum likelihood estimator. Likelihood function is calculated as follows:

$$\prod_{i=1}^N [\pi(x_i)]^{y_i} [1-\pi(x_i)]^{1-y_i} \tag{7}$$

where  $\pi(x_i) = P(y = 1|x_i) = \frac{\exp(w \cdot x_i)}{1 + \exp(w \cdot x_i)}$ , maximum likelihood function means maximizing the following logarithm function, namely

$$L(w) = \sum_{i=1}^N [y_i \log \pi(x_i) + (1-y_i) \log(1-\pi(x_i))] \tag{8}$$

then we could obtain:

$$L(w) = \sum_{i=1}^N [y_i(w \cdot x_i) - \log(1 + \exp(w \cdot x_i))] \tag{9}$$

$$w = \arg \max(L(w)) \tag{10}$$

Suppose the maximum likelihood estimation for  $w$  is  $w^*$ , then the binary logistic regression model finally becomes:

$$P(y = 1|x) = \frac{\exp(w^* \cdot x)}{1 + \exp(w^* \cdot x)} \tag{11}$$

$$P(y = 0|x) = \frac{1}{1 + \exp(w^* \cdot x)} \tag{12}$$

Back to our prediction task for the novel miRNA-disease associations, the  $j$ th column of matrix  $A$  is regarded as the binary label vector of all the miRNAs with respect to disease  $d(j)$ . If we could find feature vector for every miRNA with respect to disease  $d(j)$ , we could then utilize binary logistic regression model to calculate the association probability for miRNAs with label 0 to disease  $d(j)$ . Certainly, previously described random walk strategy is prepared for extracting feature vector for miRNAs. Assume we have already performed random walk with restart for miRNA  $m(i)$  on the weighted integrated miRNAs network and gotten the global relationship information for miRNA  $m(i)$ , namely the candidate miRNA ranks for seed miRNA  $m(i)$ . Here to extract feature vector of  $m(i)$ , we consider the top  $K$  ranked candidate miRNAs according to the random walk result. In this work,  $K$  is considered as 50, namely about 10% of the total number of miRNAs. These top  $K$  ranked candidate miRNAs would be used to build feature vector of  $m(i)$ . For a disease  $d$ , the feature vector of  $m(i)$  with respect to  $d$  is regarded as  $Vec(m(i))$ , as follows:

$$Vec(m(i)) = (1, \emptyset_{i1}, \emptyset_{i0}) \tag{13}$$

where the stable random walk probability of the top  $K$

ranked candidate miRNAs with label 1 respect to disease  $d$  were added up as  $\varnothing_{i1}$ . Similarly, we added up the stable random walk probability of the top  $K$  ranked miRNAs with label 0 respect to disease  $d$  as  $\varnothing_{i0}$ . Specially, the element '1' in the feature vector represents the constant term. Then for disease  $d$ , we get the ternary feature vector of every miRNA. Together with the binary label information and feature vector of miRNA, we could easily take use of binary logistic regression model to calculate the posterior association probability for the given disease.

## Additional file

**Additional file 1:** We prioritized corresponding candidate miRNAs for all the diseases recorded in HMDD v2.0 database. The predicted results for each disease were publicly released for further experimental validation. The potential and promising disease-miRNA associations with relatively high ranks were expected to be confirmed by biological experiments and clinical observation in the future. (XLSX 4198 kb)

## Abbreviations

RWR: random walk with restart; ROC: receiver operating characteristics; AUC: the area under the ROC curve; LOOCV: leave-one-out cross validation; RBF: radial basis function; TPR: true positive rate; FPR: false positive rate; HMDD: human microRNA disease database; dbDEMC: database of differentially expressed miRNAs in human cancers

## Acknowledgements

Not applicable.

## Funding

XC was supported by National Natural Science Foundation of China under Grant Nos. 61772531 and 11631014. GHW was supported by National Natural Science Foundation of China under Grant Nos. 11471193 and 11631014, the Foundation for Distinguished Young Scholars of Shandong Province under Grant No. JQ201501 and the Qilu Scholar Award of Shandong University. GYY was supported by National Natural Science Foundation of China under Grant No. 11631014.

## Availability of data and materials

The miRNA functional similarity matrix was obtained from <http://www.cuilab.cn/files/images/cuilab/misim.zip>. The known miRNA-disease associations were downloaded from HMDD v2.0 database.

## Authors' contributions

YWN developed the prediction method, implemented the experiments, analyzed the result, and wrote the paper. XC conceived the project, developed the prediction method, analyzed the result and wrote the paper. GHW and GYY analyzed the result and wrote the paper. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>School of Mathematics, Shandong University, Jinan 250100, China.

<sup>2</sup>Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. <sup>3</sup>School of Information and Control Engineering, China University of Mining and Technology, No.1, Daxue Road, Xuzhou 221116, Jiangsu, China.

Received: 26 July 2018 Accepted: 18 January 2019

Published online: 28 January 2019

## References

- Ambros V. The functions of animal microRNAs. *Nature*. 2004;431(7006):350–5.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281–97.
- Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. *Nature*. 2004;431(7006):343–9.
- Ambros V. microRNAs: tiny regulators with great potential. *Cell*. 2001;107(7):823–6.
- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136(2):215–33.
- Lee RC, Feinbaum RL, Ambros V. The *C. Elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. 1993;75(5):843–54.
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*. 2000;403(6772):901–6.
- Pasquinelli AE, Ruvkun G. Control of developmental timing by microRNAs and their targets. *Annu Rev Cell Dev Biol*. 2002;18:495–513.
- Bandyopadhyay S, Mitra R, Maulik U, Zhang MQ. Development of the human cancer microRNA network. *Silence*. 2010;1(1):6.
- Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(Database issue):D68–73.
- Cheng AM, Byrom MW, Shelton J, Ford LP. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res*. 2005;33(4):1290–7.
- Karp X, Ambros V. Developmental biology. Encountering microRNAs in cell fate signaling. *Science (New York, NY)*. 2005;310(5752):1288–9.
- Miska EA. How microRNAs control cell division, differentiation and death. *Curr Opin Genet Dev*. 2005;15(5):563–8.
- Xu P, Guo M, Hay BA. MicroRNAs and the regulation of cell death. *Trends Genet*. 2004;20(12):617–624.
- Alshalalfa M, Alhaji R. Using context-specific effect of miRNAs to identify functional associations between miRNAs and gene signatures. *BMC bioinformatics*. 2013;14(Suppl 12):S1.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34(Database issue):D140–D144.
- Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*. 2006;6(4):259–69.
- Latronico MV, Catalucci D, Condorelli G. Emerging role of microRNAs in cardiovascular biology. *Circ Res*. 2007;101(12):1225–36.
- Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. An analysis of human microRNA and disease associations. *PLoS One*. 2008;3(10):e3420.
- Kliese N, Gobrecht P, Pachow D, Andrae N, Wilisch-Neumann A, Kirches E, Riek-Burchardt M, Angenstein F, Reifenberger G, Riemenschneider MJ, et al. miRNA-145 is downregulated in atypical and anaplastic meningiomas and negatively regulates motility and proliferation of meningioma cells. *Oncogene*. 2013;32(39):4712–20.
- Madhavan D, Zucknick M, Wallwiener M, Cuk K, Modugno C, Scharpf M, Schott S, Heil J, Turchinovich A, Yang R, et al. Circulating miRNAs as surrogate markers for circulating tumor cells and prognostic markers in metastatic breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(21):5972–82.
- Zhao S, Yang G, Mu Y, Han D, Shi C, Chen X, Deng Y, Zhang D, Wang L, Liu Y, et al. MiR-106a is an independent prognostic marker in patients with glioblastoma. *Neuro-Oncology*. 2013;15(6):707–17.
- Uhl E, Krimer P, Schliekelman P, Tompkins SM, Suter S. Identification of altered MicroRNA expression in canine lymphoid cell lines and cases of B- and T-cell lymphomas. *Genes, chromosomes & cancer*. 2011;50(11):950–67.

24. 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.
25. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6(11):857–66.
26. Cho WC. MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int J Biochem Cell Biol*. 2010;42(8):1273–81.
27. Tricoli JV, Jacobson JW. MicroRNA: potential for Cancer detection, diagnosis, and prognosis. *Cancer Res*. 2007;67(10):4553–5.
28. Chen X, Yan CC, Zhang X, You ZH. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform*. 2016;bbw060.
29. Chen X, Yan CC, Zhang X, Dai F, Yin J, Zhang Y. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform*. 2016;17(4):696–712.
30. Perez-Iratxeta C, Wjst M, Bork P, Andrade MA. G2D: a tool for mining genes associated with disease. *BMC Genet*. 2005;6:45.
31. Perez-Iratxeta C, Bork P, Andrade MA. Association of genes to genetically inherited diseases using data mining. *Nat Genet*. 2002;31(3):316–9.
32. 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 (Oxford, England)*. 2010;26(13):1644–50.
33. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A*. 2007;104(21):8685–90.
34. Chen X, Wang L, Qu J, Guan NN, Li JQ. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics (Oxford, England)*. 2018.
35. Chen X, Yin J, Qu J, Huang L. MDHG: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput Biol*. 2018;14(8):e1006418.
36. You ZH, Huang ZA, Zhu Z, Yan GY, Li ZW, 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.
37. Chen X, Zhou Z, Zhao Y. ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction. *RNA Biol*. 2018;15(6):807–18.
38. Chen X, Huang L. LRSSLMDA: Laplacian regularized sparse subspace learning for miRNA-disease association prediction. *PLoS Comput Biol*. 2017;13(12):e1005912.
39. Xu J, Li CX, Lv JY, Li YS, Xiao Y, Shao TT, Huo X, Li X, Zou Y, Han QL, et al. Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer. *Mol Cancer Ther*. 2011;10(10):1857–66.
40. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. *Mol BioSyst*. 2012;8(10):2792–8.
41. 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.
42. Mork S, Pletscher-Frankild S, Palleja Caro A, Gorodkin J, Jensen LJ. Protein-driven inference of miRNA-disease associations. *Bioinformatics (Oxford, England)*. 2014;30(3):392–7.
43. Pasquier C, Gardes J. Prediction of miRNA-disease associations with a vector space model. *Sci Rep*. 2016;6:27036.
44. Chen X, Xie D, Wang L, Zhao Q, You ZH, Liu H. BNPMDA: bipartite network projection for MiRNA-disease association prediction. *Bioinformatics (Oxford, England)*. 2018.
45. Chen X, Yan GY. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci Rep*. 2014;4:5501.
46. Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, Liu Y, Dai Q, Li J, Teng Z, et al. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One*. 2013;8(8):e70204.
47. Chen X, Wu QF, Yan GY. RKNMMDA: ranking-based KNN for MiRNA-disease association prediction. *RNA Biol*. 2017;14(7):952–62.
48. Chen X, Yan CC, Zhang X, Li Z, Deng L, Zhang Y, Dai Q. RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci Rep*. 2015;5:13877.
49. Chen X, Yan CC, Zhang X, You ZH, Deng L, Liu Y, Zhang Y, Dai Q. WBSMDA: within and between score for MiRNA-disease association prediction. *Sci Rep*. 2016;6:21106.
50. Chen X, Clarence Yan C, Zhang X, You ZH, Huang YA, Yan GY. HGMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget*. 2016;7(40):65257–69.
51. Li JQ, Rong ZH, Chen X, Yan GY, You ZH. MCMMA: matrix completion for MiRNA-disease association prediction. *Oncotarget*. 2017.
52. Yu H, Chen X, Lu L. Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. *Sci Rep*. 2017;7:43792.
53. 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.
54. Chen X, Xie D, Zhao Q, You ZH. MicroRNAs and complex diseases: from experimental results to computational models. *Brief Bioinform*. 2017.
55. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*. 2014;42(Database issue):D1070–4.
56. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009;37(Database):D98–104.
57. Yang Z, Ren F, Liu C, He S, Sun G, Gao Q, Yao L, Zhang Y, Miao R, Cao Y, et al. dbDEM: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*. 2010;11(Suppl 4):S5.
58. He B, Yin B, Wang B, Xia Z, Chen C, Tang J. MicroRNAs in esophageal cancer (review). *Mol Med Rep*. 2012;6(3):459–65.
59. Enzinger PC, Mayer RJ. Esophageal cancer. *N Engl J Med*. 2003;349(23):2241–52.
60. Daly JM, Fry WA, Little AG, Winchester DP, McKee RF, Stewart AK, Fremgen AM. Esophageal cancer: results of an American College of Surgeons patient care evaluation study. *J Am Coll Surg*. 2000;190(5):562–72 discussion 572–563.
61. Wan J, Wu W, Che Y, Kang N, Zhang R. Insights into the potential use of microRNAs as a novel class of biomarkers in esophageal cancer. *Diseases of the esophagus : official journal of the International Society for. Dis Esophagus*. 2016;29(5):412–20.
62. Xie Z, Chen G, Zhang X, Li D, Huang J, Yang C, Zhang P, Qin Y, Duan Y, Gong B, et al. Salivary microRNAs as promising biomarkers for detection of esophageal cancer. *PLoS One*. 2013;8(4):e57502.
63. Liu R, Liao J, Yang M, Shi Y, Peng Y, Wang Y, Pan E, Guo W, Pu Y, Yin L. Circulating miR-155 expression in plasma: a potential biomarker for early diagnosis of esophageal cancer in humans. *Journal of toxicology and environmental health Part A*. 2012;75(18):1154–62.
64. Gmyrek GA, Walburg M, Webb CP, Yu HM, You X, Vaughan ED, Vande Woude GF, Knudsen BS. Normal and malignant prostate epithelial cells differ in their response to hepatocyte growth factor/scatter factor. *Am J Pathol*. 2001;159(2):579–90.
65. Goto Y, Kurozumi A, Enokida H, Ichikawa T, Seki N. Functional significance of aberrantly expressed microRNAs in prostate cancer. *International journal of urology : official journal of the Japanese Urological Association*. 2015;22(3):242–52.
66. Ueno K, Hirata H, Shahryari V, Deng G, Tanaka Y, Tabatabai ZL, Hinoda Y, Dahiya R. microRNA-183 is an oncogene targeting Dkk-3 and SMAD4 in prostate cancer. *Br J Cancer*. 2013;108(8):1659–67.
67. Porkka KP, Pfeiffer MJ, Waltering KK, Vessella RL, Tammela TL, Vissakorpi T. MicroRNA expression profiling in prostate cancer. *Cancer Res*. 2007;67(13):6130–5.
68. Van Asten K, Neven P, Lintermans A, Wildiers H, Paridaens R. Aromatase inhibitors in the breast cancer clinic: focus on exemestane. *Endocr Relat Cancer*. 2014;21(1):R31–49.
69. Schrauder MG, Strick R, Schulz-Wendtland R, Strissel PL, Kahmann L, Loehberg CR, Lux MP, Jud SM, Hartmann A, Hein A, et al. Circulating microRNAs as potential blood-based markers for early stage breast cancer detection. *PLoS One*. 2012;7(1):e29770.
70. Li Q, Yao Y, Eades G, Liu Z, Zhang Y, Zhou Q. Downregulation of miR-140 promotes cancer stem cell formation in basal-like early stage breast cancer. *Oncogene*. 2014;33(20):2589–600.
71. Gibcus JH, Tan LP, Harms G, Schakel RN, de Jong D, Blokzijl T, Moller P, Poppema S, Kroesens BJ, van den Berg A. Hodgkin lymphoma cell lines are characterized by a specific miRNA expression profile. *Neoplasia (New York, NY)*. 2009;11(2):167–76.
72. Sandlund JT, Onciu M. Childhood lymphoma - Abeloff's clinical oncology (fifth edition) - 97. *Abeloffs Clinical Oncology* 2014. 1873–1889:e1876.
73. Lawrie CH, Soneji S, Marafioti T, Cooper CD, Palazzo S, Paterson JC, Cattani H, Enver T, Mager R, Boulwood J, et al. MicroRNA expression distinguishes between germinal center B cell-like and activated B cell-like subtypes of diffuse large B cell lymphoma. *Int J Cancer*. 2007;121(5):1156–61.
74. Leich E, Zamo A, Horn H, Haralambieva E, Puppe B, Gascoyne RD, Chan WC, Brazier RM, Rimsza LM, Weisenburger DD, et al. MicroRNA profiles of t(14;18)-negative follicular lymphoma support a late germinal center B-cell phenotype. *Blood*. 2011;118(20):5550–8.

75. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* (Oxford, England). 2011;27(21):3036–43.
76. Chen X, Ren B, Chen M, Wang Q, Zhang L, Yan G. NLLSS: predicting synergistic drug combinations based on semi-supervised learning. *PLoS Comput Biol*. 2016;12(7):e1004975.
77. Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics* (Oxford, England). 2013;29(20):2617–24.
78. Chen B, Li M, Wang J, Shang X, Wu FX. A fast and high performance multiple data integration algorithm for identifying human disease genes. *BMC Med Genet*. 2015;8(Suppl 3):S2.
79. Wei Z. Research on disease gene prediction algorithm based on gene network[D]. Harbin institute of. Technology. 2016.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

