BMC Genomics

BLNIMDA: identifying miRNA-disease associations based on weighted bi-level network

Junliang Shang, Yi Yang, Feng Li, Boxin Guan, Jin-Xing Liu and Yan Sun[*]

Abstract

Background: MicroRNAs (miRNAs) have been confirmed to be inextricably linked to the emergence of human complex diseases. The identification of the disease-related miRNAs has gradually become a routine way to unveil the genetic mechanisms of examined disorders.

Methods: In this study, a method BLNIMDA based on a weighted bi-level network was proposed for predicting hidden associations between miRNAs and diseases. For this purpose, the known associations between miRNAs and diseases as well as integrated similarities between miRNAs and diseases are mapped into a bi-level network. Based on the developed bi-level network, the miRNA-disease associations (MDAs) are defined as strong associations, potential associations and no associations. Then, each miRNA-disease pair (MDP) is assigned two information properties according to the bidirectional information distribution strategy, i.e., associations of miRNA towards disease and vice-versa. Finally, two affinity weights for each MDP obtained from the information properties and the association type are then averaged as the final association score of the MDP. Highlights of the BLNIMDA lie in the definition of MDA types, and the introduction of affinity weights evaluation from the bidirectional information distribution strategy and defined association types, which ensure the comprehensiveness and accuracy of the final prediction score of MDAs.

Results: Five-fold cross-validation and leave-one-out cross-validation are used to evaluate the performance of the BLNIMDA. The results of the Area Under Curve show that the BLNIMDA has many advantages over the other seven selected computational methods. Furthermore, the case studies based on four common diseases and miRNAs prove that the BLNIMDA has good predictive performance.

Conclusions: Therefore, the BLNIMDA is an effective method for predicting hidden MDAs.

Keywords: miRNA similarity, Disease similarity, Association type, Bi-level network

Background

MicroRNAs (MiRNAs) are a type of non-protein-coding ribonucleic acids with a length of approximately 22 nucleotides [1]. It modulates the biological activities of proteins by promoting and inhibiting the expression of respective genes, thereby being able to initiate the emergence of diseases [2]. In recent years, miRNAs have been confirmed

to be inextricably linked to the emergence of complex human diseases. For example, upregulation of miR-132 can contribute to the development of AIDS disease via promoting the replication of human immunodeficiency virus 1 (HIV-1) [3, 4]. Therefore, the identification of the hidden miRNA-disease associations (MDAs) can contribute to the better prevention and curation of diseases [5]. However, traditional "wet experiments" in biomedicine are often time-consuming and labor-intensive, characterizing by the lack of directionality to a certain extent

*Correspondence: sunyan225@126.com

School of Computer Science, Qufu Normal University, 276826 Rizhao, China

Shang *et al. BMC Genomics*   (2022) 23:686

Page 2 of 12

[6, 7]. Therefore, it is necessary and meaningful to predict MDAs through biological methods. So far, many biological computation models for predicting MDAs have been developed [8–10], which can be divided into three main categories.

The first category of computational methods is represented by models constructed on the basis of score functions. The assumption that functionally similar miRNAs are more likely to be linked with similar diseases was implemented in the computational method for inference of MDAs devised by Jiang et al. [11]. This model firstly mapped the known MDAs, and the relationships between different miRNAs and diseases into a heterogeneous network to prove the correctness of the implemented hypothesis. Then, a scoring system for predicting MDAs was designed based on the neighbors shared by miRNAs and the measurements of their shortest paths. Therefore, the final MDA score was determined by ranking the disease-related miRNAs. In turns, Chen et al. [8] proposed a different model for predicting MDAs, called WBSMDA, which expanded the angle of miRNA and disease similarity calculation, innovatively employing the Gaussian interaction profile (GIP) kernel similarity network. It not only described the relationship between different miRNAs and diseases from a new perspective, but also could be used for independent nodes without any known associations.

The second category of computational methods for predicting MDAs is represented by models constructed on the basis of complex networks. These models mapped matrices related to miRNA and disease into the networks to predict the score of MDAs. Among those models, Chen et al. [12] proposed the RWRMDA method based on the random walk with restart to predict MDAs. The highlight of the RWRMDA was that the random walk was applied to the miRNA functional similarity network to rank all disease-related miRNAs. However, the RWRMDA ignored prior information and local topological structures of isolated miRNA and disease nodes. To eliminate the above defect, Xuan et al. [13] proposed a novel model based on the miRNA-related network constructed by integrating the available information on the MDAs and their local topological structures. The resulted nodes can be assigned as marked and unmarked whose characteristics enable the establishment of the transition network, which was proportional to the similarity between all nodes. For the isolated diseases, the bilayer network of MDAs was then constructed and random walks were extended to it. You et al. [9] constructed a multi-path heterogeneous network of MDAs (PBMDA), which integrated the known human MDAs, the miRNA-related similarity, the disease-related similarity, and the GIP kernel similarity. PBMDA set the length between different nodes on this network and further adopted a depth-first search algorithm to obtain the hidden MDAs information.

The third category of computational methods used for MDAs prediction is represented by models based on the machine learning algorithms. In this case, Chen et al. [14] proposed a matrix factorization model (IMCMDA), which solved the common problem of traditional methods, that is, the inability of isolated nodes to predict dependencies. The IMCMDA constructed features for miRNAs and diseases based on the information of the miRNA and disease similarity and selected robust features through an alternate search algorithm. Finally, the IMCMDA predicted the MDAs score through a semi-supervised model that did not rely on negative samples. In recent years, more and more deep learning-related models have been introduced into MDAs prediction. Ji et al. [15] proposed a computational framework based on deep autoencoder (AEMDA). The innovation of the AEMDA was the development of a learning-based feature extraction method after constructing miRNA and disease feature representation. Furthermore, the deep autoencoder and the reconstruction error method were introduced to predict MDAs. Tang et al. [16] developed a model (MMGCN) based on a multi-view multichannel attention graph convolutional network (GCN). The MMGCN employed the GCN to obtain features of miRNAs and diseases from multiple angles, and a multi-channel attention mechanism was used to adaptively select the important features. The final embeddings were constructed by a Convolutional Neural Networks synthesizer. The final association prediction was equated to the recommendation task, and a matrix completion was applied to predict hidden MDAs. For predicting MDAs from a comprehensive and novel perspective, Chu et al. [17] developed an original model (MDA-DCNFTG) based on the GCN, which treated MDAs prediction as a node classification task. The highlight of the MDA-GCN-FTG was that it used graph sampling to predict MDAs from the perspective of feature and topological graphs based on Graph Convolutional Networks (GCNs). In addition, MDA-GCNFTG could predict not only new MDAs but also hidden association between diseases without known related miRNAs and miRNAs without known related diseases. Dai et al. [18] proposed a model for identifying potential MDAs based on the cascade forest model (MDA-CF), which integrated multi-source information to comprehensively characterize miRNAs and diseases, and used autoencoders for dimensionality reduction to obtain the optimal feature space and ranked it. A joint forest model was used in the prediction of potential MDAs.

Shang *et al. BMC Genomics*   (2022) 23:686

Page 3 of 12

In this study, we developed an MDAs prediction method based on a weighted bi-level network, named BLNIMDA. Specifically, a bi-level network is constructed based on the known MDAs as well as integrated similarities between miRNAs and diseases. In this constructed network, MDAs are defined into three categories, including strong associations, potential associations, and no associations. Then, each miRNA-disease pair (MDP) is assigned two information properties according to the bidirectional information distribution strategy. Finally, the association score of every MDP is obtained by averaging two affinity weights, which are derived from their information properties and association type. The Area Under Curve (AUC) values of five-fold cross-validation (FFCV) and leave-one-out cross-validation (LOOCV) are 0.9145 and 0.9176, respectively, which show that the BLNIMDA outperforms the other seven selected computational methods. Furthermore, the case studies based on four common diseases and miRNAs prove that the BLNIMDA has a good predictive performance. Thus, BLNIMDA can be used as a powerful tool to predict potential MDAs.

Materials and methods

Human MDAs

The data of MDAs are obtained from the HMDD v2.0 [19], which contains 5430 experimentally verified MDAs between 495 miRNAs and 383 diseases. To make better use of these information, we construct it as a matrix $A \in R^{n_m \times n_d}$, where n_d and n_m represent the number of diseases and miRNAs, respectively. In addition, if disease g and miRNA $_h$ are confirmed to have an association, then $A(g, h)$ will be 1, otherwise 0.

MiRNA function similarity

Wang et al. [20] developed a method for calculating the functional similarity of miRNAs, which was based on the assumption that similar miRNAs are more likely to be associated with similar diseases. The similarity score information of all miRNAs was obtained from http://www.cuilab.cn/files/images/cuilab/misim.zip. In this study, a $n_m * n_m$ matrix *MFS* was constructed to indicate that miRNA similarity and the function similarity between miRNA g and miRNA $_h$ can be expressed as $MFS(g, h)$.

Disease semantic similarity frame *I*

Disease semantic similarity is calculated from the hierarchical directed acyclic graph (DAG) of each disease [20]. The MeSH database (http://www.nlm.nih.gov/) contains the DAG information of all diseases. The semantic similarity scores between different diseases can be calculated by the relationship between their DAGs [19, 20]. Fig. 1 (a)

shows the DAGs of brain neoplasms and liver neoplasms, where each node represents a specific disease MeSH descriptor. The DAG for the disease Z can be denoted as $DAG_Z(N_Z, E_Z)$, where N_Z represents the node set, which includes the MeSH descriptor for disease Z and its ancestor nodes, and E_Z denotes the layer set, which includes all edges connecting the parent node to the child node DAG_Z. We assumed that l is a MeSH descriptor node in DAG_Z, and its semantic contribution value in DAG_Z is as follows:

$$\begin{cases} DS1_Z(l) = 1 & if \ l = Z \\ DS1_Z(l) = \max \left\{ \Delta * DS1(l') \right\} & if \ l \neq Z \end{cases} \quad (1)$$

where l' is the child node of l and Δ is used to indicate the semantic contribution decay, which is set to 0.5 according to previous literature [14] The semantic contribution of disease Z is defined as formula (2).

$$DV1(Z) = \sum_{l \in N_Z} DS1(l) \quad (2)$$

It can be found that the more shared MeSH descriptor nodes in the DAG of the two diseases, the more similar the two diseases are. Therefore, the formula for calculating the semantic similarity score of disease d_g and d_h is as follows:

$$SD1(d_g, d_h) = \frac{\sum_{c \in N_{d_g} \cap N_{d_h}} \left(DS1_{d_g}(c) + DS1_{d_h}(c) \right)}{DV1(d_g) + DV1(d_h)} \quad (3)$$

According to Eq. (3), c denotes shared ancestral MeSH descriptor nodes by disease d_g and d_h. The computational process of semantic similarity based on frame *I* for brain neoplasms and liver neoplasms is shown in Fig. 1 (b).

Disease semantic similarity frame *II*

To differentiate the semantic contribution values of different diseases, which appear in the same layer of the same disease DAG, another calculation frame [21] was proposed. In this case the formula for calculating the semantic contribution value of disease l is as follows:

$$DS2_Z(l) = -\log \left(\frac{N_l}{N} \right) \quad (4)$$

where N_l represents the number of disease DAGs contained the MeSH descriptor of disease l and N denotes the number of all diseases in the MeSH database, i.e., N_l/N is the probability that the MeSH descriptor of disease l is present in all DAGs in the MeSH database. Based on frame *II*, the formula for calculating the semantic similarity score of disease d_g and d_h is as follows:
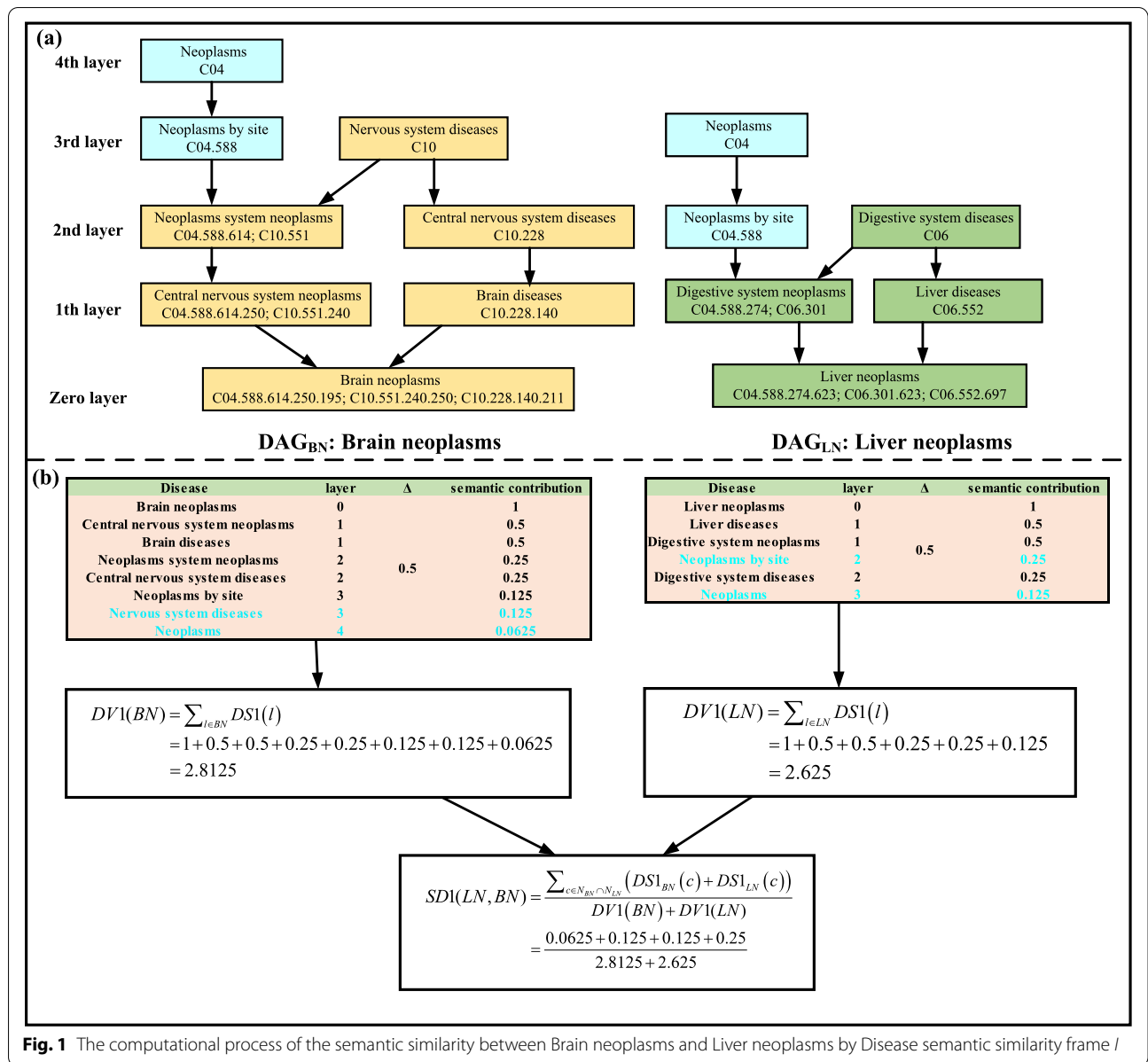
Fig. 1 The computational process of the semantic similarity between Brain neoplasms and Liver neoplasms by Disease semantic similarity frame *I*

$$SD2(d_g, d_h) = \frac{\sum_{c \in N_{d_g} \cap N_{d_h}} \left(DS2_{d_g}(c) + DS2_{d_h}(c) \right)}{DV2(d_g) + DV2(d_h)}$$

(5)

where

$$DV2(Z) = \sum_{l \in N_Z} DS2(l)$$

(6)

Finally, the similarity of different diseases at the semantic level is obtained by averaging the above two frames, which is shown as follows:

$$DS(d_g, d_h) = \frac{SD1(d_g, d_h) + SD2(d_g, d_h)}{2}$$

(7)

GIP kernel similarity

The GIP kernel [22] similarity aimed to measure the biological entities similarity based on their interaction profile information. GIP kernel similarity has been successfully introduced to the calculation of non-coding RNA and disease similarity [23]. In the adjacency matrix *A*, the *g* row denotes the correlation vector

between miRNA g and 383 diseases, and the h column indicates the correlation vector between disease h and 495 miRNAs. We used $IP(m_g)$ and $IP(d_h)$ to represent them respectively. The formula for calculating the GIP kernel similarity of miRNAs and diseases is as follows:

$$KD(d_g, d_h) = \exp\left(-\beta_d ||IP(d_g) - IP(d_h)||^2\right) \quad (8)$$

$$KM(m_g, m_h) = \exp\left(-\beta_m ||IP(m_g) - IP(m_h)||^2\right) \tag{9}$$

where

$$\beta_d = \beta_d^* / \left(\frac{\sum_{h=1}^{n} ||IP(d_h)||^2}{n_d}\right) \quad (10)$$

$$\beta_m = \beta_m^* / \left(\frac{\sum_{g=1}^{m} ||IP(m_g)||^2}{n_m}\right) \quad (11)$$

The original bandwidth β_d^* and β_m^* is set to 1.0 [24, 25].

Integrating similarity

Whether miRNA functional similarity, disease semantic similarity or GIP kernel similarity, they only provide a single aspect of similarity. Therefore, it is essential to integrate the above similarity information to obtain a more accurate and comprehensive disease or miRNA similarity. For example, if miRNA m_g and miRNA m_h have functional similarity, they will be retained, otherwise it will be equal to $KM(m_g, m_h)$. Disease similarity is integrated using the same way. Therefore, the integration method is showed as follows:

$$FMS(m_g, m_h) = \begin{cases} MS(m_g, m_h), & m_g \text{ and } m_h \text{ have functional similarity} \\ KM(m_g, m_h), & \text{otherwise} \end{cases} \quad (12)$$

$$FDS(d_g, d_h) = \begin{cases} DS(d_g, d_h), & d_g \text{ and } d_h \text{ has semantic similarity} \\ KD(d_g, d_h), & \text{otherwise} \end{cases} \quad (13)$$

An example of data processing, including the calculation of GIP kernel similarity and the process of integrating similarity, is shown in Figure S1 in the supplementary file.

BLNIMDA

On the basis of the hypothesis that functionally similar miRNAs are more likely to be linked with similar diseases, we proposed a method named BLNIMDA that combines the above-processed data, including integrated miRNA similarity, integrated disease similarity and MDAs and these data are mapped into a bi-level weighted network. The BLNIMDA predicts the MDAs based on this network and Fig. 2 provides a detailed visualization of the BLNIMDA flow. Accordingly, the BLNIMDA integrates four main computational steps, including: (i) The determination of the miRNA function similarity, the disease semantic similarity and GIP kernel similarities; (ii) The integration of the estimated similarities and mapping MDAs into a bi-level weighted network; (iii) The generation of two side information properties for each MDAs through bidirectional information construction and assignment of all MDAs into three categories, namely strong associations, potential associations, and no associations. (iv) The estimation of two affinity weights for each MDP through bidirectional information construction strategy and its association type, and then averaged as the final MDAs score. Considering the direction of miRNAs to diseases as an example, the information properties of each MDP is defined as formula (14):

$$S_1(m_g, d_h) = \frac{\sum_{k=1}^{n} HM(m_g, m_k) A(m_k, d_h)}{\sum_{k=1}^{n} FMS(m_g, m_k) A(m_k, d_h)} \quad (14)$$

where

$$HM(m_g, m_k) = \begin{cases} FMS(m_g, m_k) & FMS(m_g, m_k) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

Considering that the weak similarity nodes of m_g may affect the accuracy of prediction results, we set the parameter T to remove weak similar nodes. The MDAs are defined into three types: (i) m_g and d_h have strong association when they display unequivocal reciprocal association; (ii) m_g and d_h have potential association when they do not display direct association but the most similar node m_s to m_g has an unequivocal association with d_h. (iii) otherwise, there is not any potential association. Three types of MDAs are shown in the formula (16).

$$W_r(m_g, d_h) = \begin{cases} \exp\left(FMS(m_g, m_s)\right) & A(m_g, d_h) = 0 \text{ and } A(m_s, d_h) \neq 0 \\ \exp\left(A(m_g, d_h)\right) & A(m_g, d_h) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$(16)$$

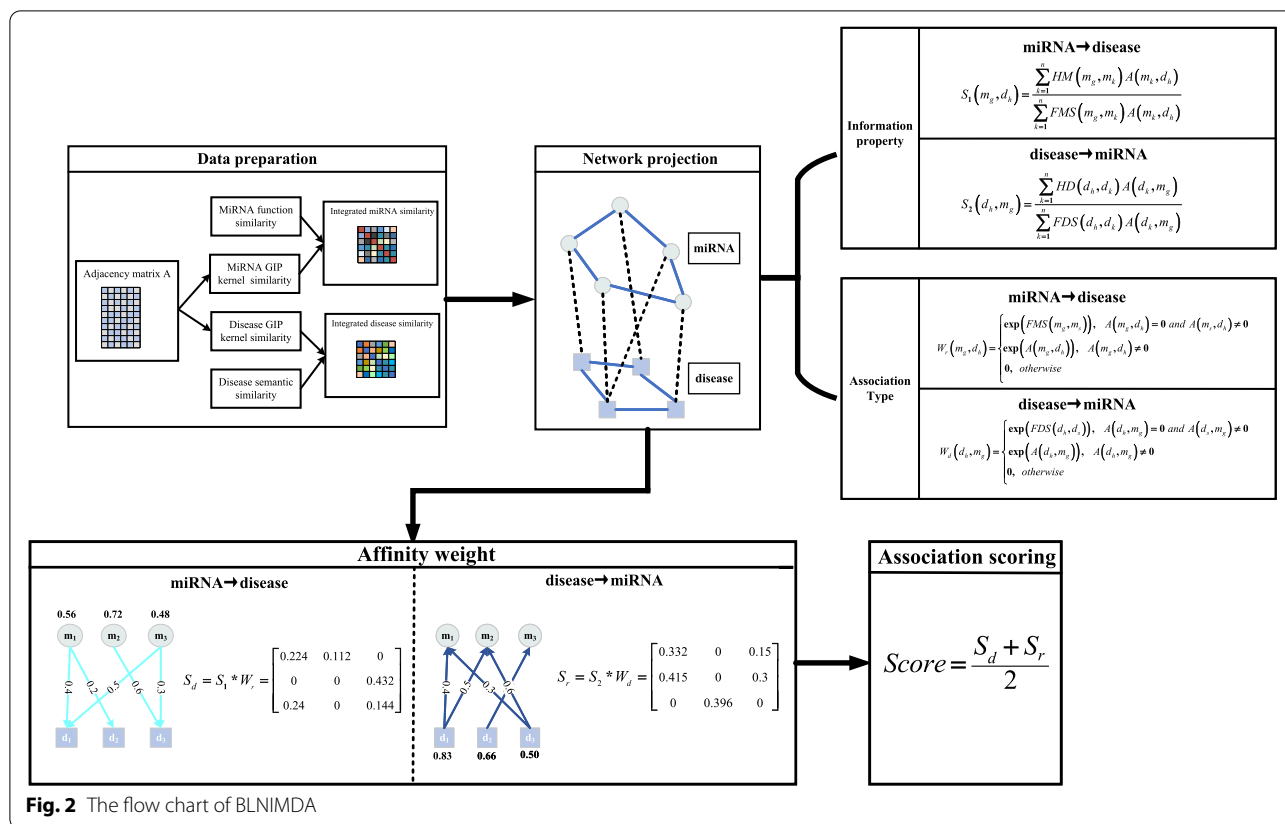Shang *et al. BMC Genomics*   (2022) 23:686

Page 6 of 12

Fig. 2 The flow chart of BLNIMDA

where m_s is the miRNA with the greatest similarity to m_g. From miRNA to disease, the affinity weight of each MDP are defined through a bidirectional information construction strategy and its association type according to formula (17):

$$S_d = S_1 * W_r \tag{17}$$

From disease to miRNA direction, the information property of each MDP is the same as the above steps, and the details are as follows:

$$S_2(d_h, m_g) = \frac{\sum_{k=1}^{n} HD(d_h, d_k) A(d_k, m_g)}{\sum_{k=1}^{n} FDS(d_h, d_k) A(d_k, m_g)} \tag{18}$$

where

$$HD(d_h, d_k) = \begin{cases} FDS(d_h, d_k) & FDS(d_h, d_k) \geq T \\ 0 & otherwise \end{cases} \tag{19}$$

The three types of miRNA-disease association are as follows:

$$W_d(d_h, m_g) = \begin{cases} \exp\left(FDS(d_h, d_s)\right) & A(d_h, m_g) = 0 \ \ and \ \ A(d_s, m_g) \neq 0 \\ \exp\left(A(d_h, m_g)\right) & A(d_h, m_g) \neq 0 \\ 0 & otherwise \end{cases} \tag{20}$$

In the same way, from disease to miRNA, the affinity weight of each MDP are determined through information property and the MDA type. The specific definition is as follows:

$$S_r = S_2 * W_d \tag{21}$$

According to formulas (14)- (21), two affinity weights for each MDP are determined, and then averaged them as the final MDAs score following the formula (22):

$$S_f = \frac{S_d + S_r}{2} \tag{22}$$

A BLNIMDA calculation example, including the generation of two side information properties, the calculation of two affinity weights for each MDP and the MDA score, is shown in Figure S2 in the supplementary file.

Results

Performance evaluation

To validate the prediction performance of the BLNIMDA, we employed cross-validation, which is considered as a reasonably comprehensive measure. In this study, FFCV and LOOCV were used to verify the performance of the BLNIMDA. For the FFCV, all MDAs were randomly

Shang *et al. BMC Genomics*   (2022) 23:686

Page 7 of 12

divided into five groups, and each group was chosen as the test sample set in turns. The remaining groups were used as the training sample sets. For LOOCV, each experimentally verified MDP was regarded in turn as the test samples where the rest of them were used as the training sample sets. The AUC and ROC were used to visualize the accuracy of the BLNIMDA in MDAs prediction. Specifically, the values closer to 1 indicated better performance of the BLNIMDA, on the contrary, the lower values of AUC revealed the worse prediction performance. By setting different thresholds, the ratio of samples above the threshold to other samples in all positive samples was taken as sensitivity (TPR), and the ratio of samples below the threshold to other samples in all negative samples was taken as 1-specificity (FPR). ROC curves were plotted by TPR and FPR under different thresholds.

Effects of parameters

In the bidirectional information construction strategy, there are some noise nodes in the miRNA or disease similarity network, which may affect the performance of the BLNIMDA. In view of this situation, we set the parameter T and adjusted the threshold of T to reduce the influence of such nodes, so as to obtain the optimal prediction results. After several experiments, the comparison of the average AUC values after 100 times FFCV when T sets different thresholds is shown in Fig. 3. It can be seen that the performance of the BLNIMDA is the best when T is set to 0.02.

Performance comparison

In order to show the superiority of the BLNIMDA, we compared seven models [8, 10, 12, 14, 24, 26, 27] for predicting the MDAs. These seven models have their characteristics. The WBSMDA and the HDMP are basic models in the research field of predicting MDAs, and the RWRMDA is a restart random walk model, the GRMDA is a graph regression method, the RLSMDA is a model based on machine learning methods, the IMCMDA is a model completed by inductive matrix, and the BNPMDA is a high-level paper proposed in recent years. All models selected for comparative analysis, including the BLNIMDA, used the same network to evaluate the performance through FFCV and LOOCV. The ROC curves of the LOOCV are shown in Fig. 4. It can be clearly found that the AUC values of WBSMDA, HDMP, RWRMDA, GRMDA, RLSMDA, IMCMDA, and BNPMDA models are 0.8030, 0.8366, 0.6850, 0.8272, 0.8426, 0.8000, 0.9028, respectively. The AUC value of the BLNIMDA is 0.9176, which is higher than the other seven models. The results of the FFCV are shown in Fig. 5. The AUC value of the BLNIMDA model is 0.9145, while the AUC values of WBSMDA, HDMP, RLSMDA, BNPMDA, RWRMDA, GRMDA and IMCMDA models are 0.8185, 0.8342, 0.8569, 0.8980, 0.6830, 0.7976 and 0.7978, respectively. All seven models are lower than the BLNIMDA. Compared with the classical graph algorithm, our BLNIMDA model introduces affinity weights evaluation and the bidirectional information distribution strategy. In particular, it calculates the association score of miRNA and disease pair through the affinity weights evaluation from the bidirectional information distribution strategy and the introduced association types.

In addition, an independent dataset was used as the test set to better verify the prediction performance of BLNIMDA. Considering that BLNIMDA uses HMDD
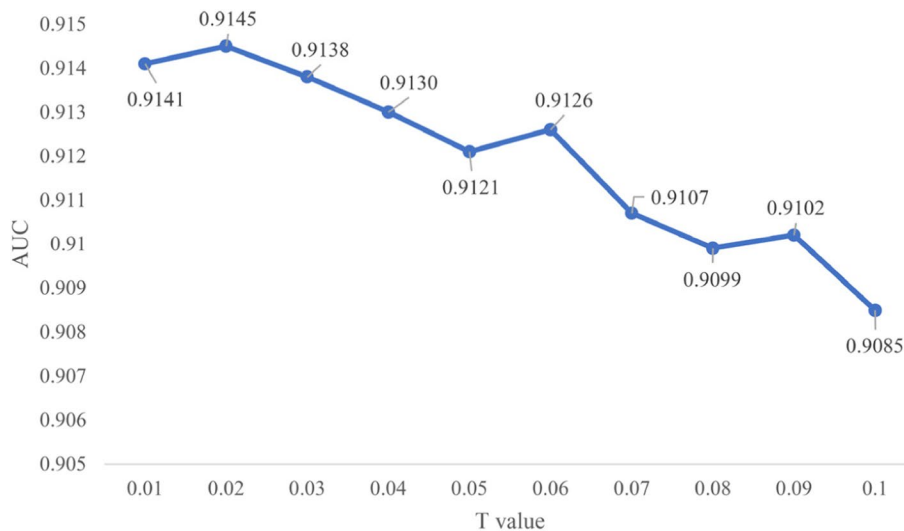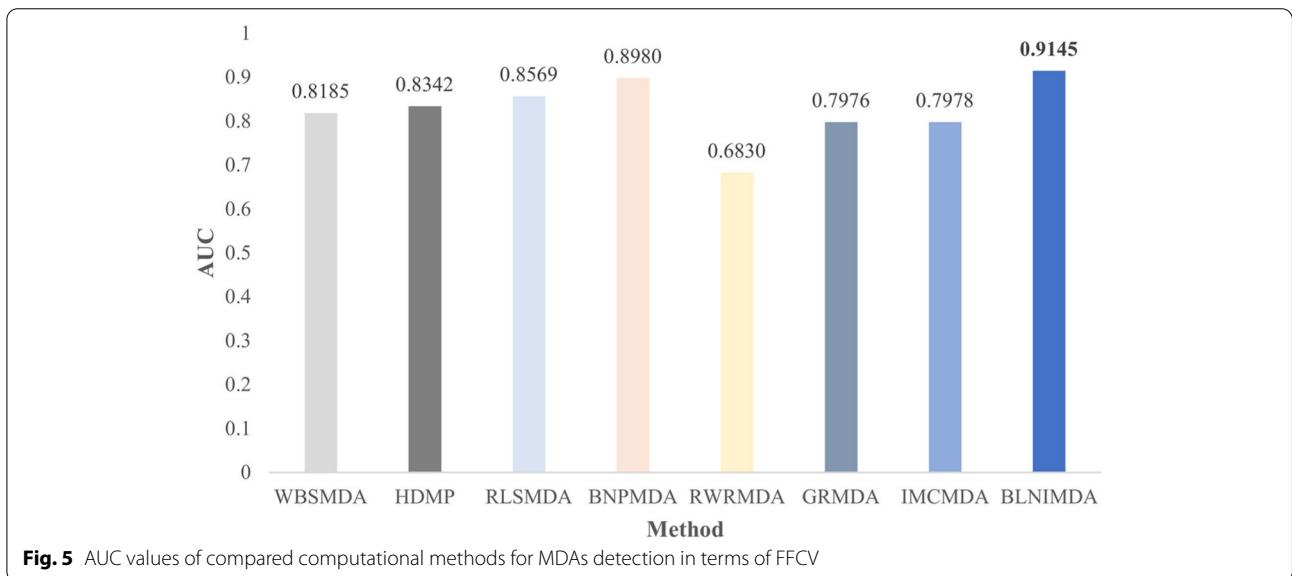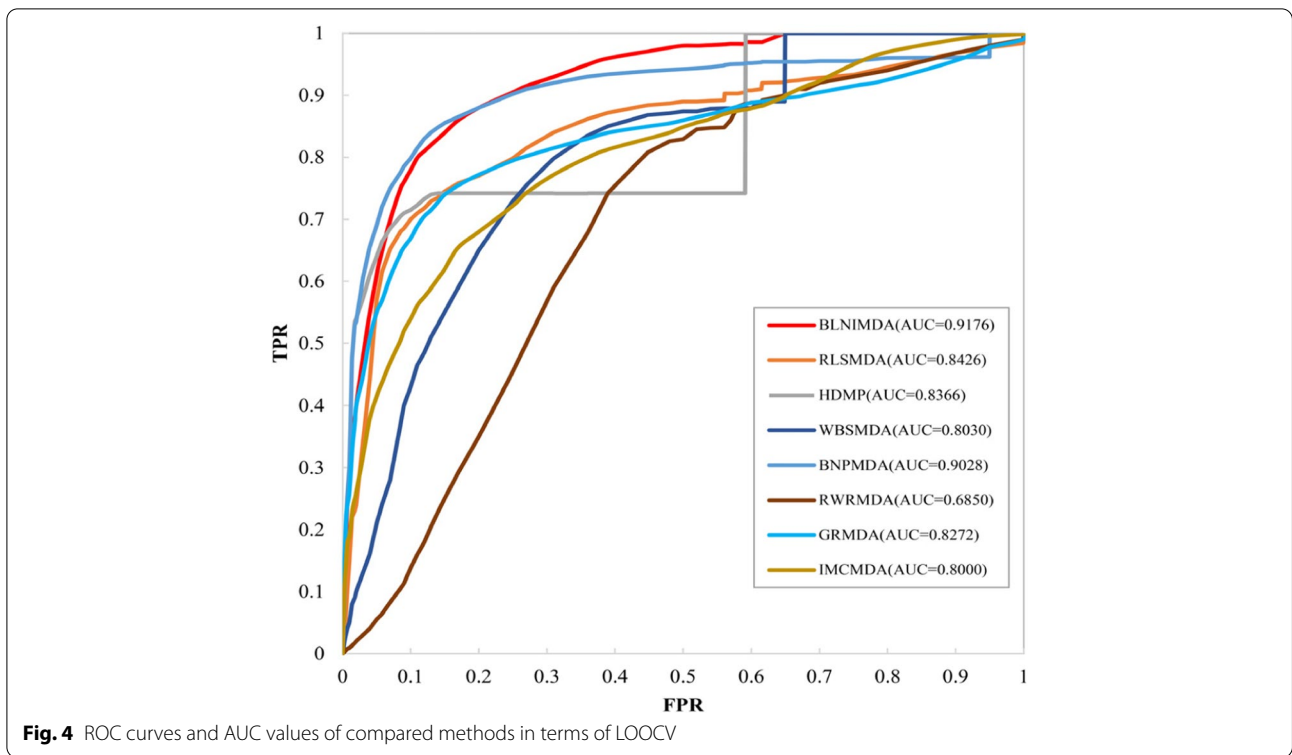

Fig. 3 AUC values of BLNIMDA depending on the different T parameter values

Fig. 4 ROC curves and AUC values of compared methods in terms of LOOCV

Fig. 5 AUC values of compared computational methods for MDAs detection in terms of FFCV

v2.0 as its training set, the MDAs in the new version of the HMDD database (HMDD v3.2) not included in the training dataset were used as a test sample. The AUCs of BLNIMDA and the other seven methods are shown in Table 1, from which it can be seen that the prediction performance of BLNIMDA in the independent test set is better than other comparison methods.

Case studies

To verify the accuracy of the BLNIMDA, we employ it to predict the association between miRNAs and four important diseases, including lung neoplasms, breast neoplasms, gastric neoplasms, and colon neoplasms.

Colon neoplasms is one of the most common malignant neoplasms that seriously threaten human health,

Shang *et al. BMC Genomics*   (2022) 23:686

Page 9 of 12

Table 1 AUC values of compared computational methods for MDAs detection in independent test set

Method	AUC
WBSMDA	0.7932
HDMP	0.8035
RLSMDA	0.8326
BNPMDA	0.8869
RWRMDA	0.6324
GRMDA	0.7574
IMCMDA	0.7658
BLNIMDA	**0.9032**

Table 2 The top five miRNAs identified by BLNIMDA to be associated with four important diseases

Disease	Rank	miRNA	Evidence	Causality
Lung Neoplasms	1	hsa-mir-21	miRCancer; HMDDv3.2	YES
	2	hsa-mir-155	miRCancer; HMDDv3.2	YES
	3	hsa-mir-146a	miRCancer; HMDDv3.2	NO
	4	hsa-mir-17	miRCancer; HMDDv3.2	YES
	5	hsa-mir-34a	miRCancer; HMDDv3.2	YES
Breast Neoplasms	1	hsa-mir-21	miRCancer; HMDDv3.2	YES
	2	hsa-mir-155	miRCancer; HMDDv3.2	YES
	3	hsa-mir-17	miRCancer; HMDDv3.2	YES
	4	hsa-mir-146a	miRCancer; HMDDv3.2	YES
	5	hsa-mir-34a	miRCancer; HMDDv3.2	YES
Gastric Neoplasms	1	hsa-mir-148a	miRCancer; HMDDv3.2	YES
	2	hsa-mir-23a	miRCancer; HMDDv3.2	YES
	3	hsa-mir-370	miRCancer; HMDDv3.2	YES
	4	hsa-mir-429	miRCancer; HMDDv3.2	YES
	5	hsa-mir-21	miRCancer; HMDDv3.2	YES
Colon Neoplasms	1	hsa-mir-145	miRCancer; HMDDv3.2	YES
	2	hsa-mir-17	HMDDv3.2	YES
	3	hsa-mir-21	miRCancer; HMDDv3.2	YES
	4	hsa-mir-126	miRCancer; HMDDv3.2	YES
	5	hsa-mir-155	miRCancer; HMDDv3.2	YES

and its mortality rate ranks third in the world [25]. According to available statistics, the number of new cases of colon neoplasms worldwide in 2008 reached 1 million, half of which led to death [28]. In recent years, the number of new cases of colon neoplasms in China has been on the rise, which can be attributed to the improvement of their diagnosis effectiveness by the advancement of endoscopy and the higher frequency of their emergence caused by changes in dietary patterns. Therefore, colon cancer has become one of the most common malignant tumors, which seriously threatens human health [29, 30]. Lungcancer is another most common malignant tumors in the world. According to incomplete statistics, about 200,000 people are diagnosed with lung cancer every year [31]. Generally speaking, patients with lung cancer have no typical clinical manifestations in the early stage, and have already been in the advanced stage at the time of diagnosis, with a poor prognosis and the 5-year survival rate of less than 15% [32]. If lung cancer can be detected early and intervened, the survival rate of patients can be significantly increased [33]. Therefore, it is of great significance to find an effective early diagnostic marker. Breast neoplasm is another one of the most frequent malignancies and occurs in breast epithelial tissue in women [34]. Approximately 1.2 million women around the world are diagnosed each year. Gastric cancer is a particularly intractable malignant tumor with the fifth highest incidence and the third highest mortality in the world highly recurrent malignant tumors, after lung cancer and colorectal [35]. Gastric cancer is difficult to detect in the early stage and has no obvious characteristics, so many patients are diagnosed in advanced stage and the mortality rate is extremely high [35]. After scoring and ranking the MDAs by the BLNIMDA, all known miRNAs related to these four diseases were removed, and the remaining information was queried and verified in the HMDDv3.2 [36], the miRCancer [37] database and related literature. According to the results shown in Table 2, it can be seen that the 20 MDPs predicted by the BLNIMDA are all confirmed from other databases.

At the same time, we randomly selected four miRNAs and screened the top five diseases with their association scores. The predicted MDAs were verified in HMDDv3.2 and miRCancer databases, and almost all of them were confirmed by available evidence data and shown in Table 3. Furthermore, causality information for the top five associations was verified by HMDD v3.2 and attached to Tables 2 and 3.

Discussion
Many miRNAs have been confirmed to have links with the occurrence of human diseases. Discovering the potential MDAs is of great significance for

Shang *et al. BMC Genomics*　　(2022) 23:686

Page 10 of 12

understanding the pathogenesis of diseases. Traditional methods for finding MDAs have some disadvantages such as low efficiency and difficult operation. Therefore, many computational models have been proposed, which are mainly divided into three categories, namely score function-based, complex network-based and machine learning-based. In this study, firstly, we proposed a computational method BLNIMDA based on a weighted bi-level network to predict the hidden MDAs. Specifically, the bi-level network is constructed based on the known MDAs as well as integrated similarities between miRNAs and diseases, in which nodes denote miRNAs and diseases. In this developed bi-level network, the MDAs are defined as strong associations, potential associations and no association according to the relationship between the miRNAs and diseases. Then, each MDP is assigned two information properties based on the bidirectional information distribution strategy. Finally, two affinity weights for each MDP are obtained from the information properties and the association type and then averaged as the final association score of every MDP. In experiments, FFCV and LOOCV are used to evaluate the performance of the BLNIMDA, and the AUC values of them are 0.9145 and 0.9176, respectively, which show that the BLNIMDA have advantages over the other seven computational methods. Furthermore, the case studies based on four common diseases and miRNAs prove that the BLNIMDA has good predictive performance. Therefore,

BLNIMDA is an effective method for predicting hidden MDAs.

The main contributions of the BLNIMDA are the ability for determining MDA type, and the introduction of two affinity weights in accordance to the bidirectional information distribution strategy and defined association types, which ensures the comprehensiveness and accuracy of the final prediction score for each MDP. However, BLNIMDA still has certain defects. First of all, the dependence on known information has not been completely eliminated. In addition, for the information between miRNAs and diseases, multiple dimensions similarity could be considered to improve prediction performance. Furthermore, BLNIMDA is essentially a traditional network method, which still needs to be improved through algorithms based on deep learning.

In the future, we will consider in-depth researches for MDAs, such as the regulatory role of miRNAs in specific diseases and the multiple type MDAs instead of taking them as binary, which have important implications for the treatment of complex human diseases.

Conclusion

In this study, a method BLNIMDA based on a weighted bi-level network was proposed for predicting the hidden associations between miRNAs and diseases. Highlights of the BLNIMDA lie in the definition of MDA

Table 3 The top five diseases identified by BLNIMDA to be associated with four major miRNAs

MiRNA	Rank	Disease	Evidence	Causality
hsa-mir-125a	1	Carcinoma	HMDDv3.2	YES
	2	Breast Neoplasms	miRCancer; HMDDv3.2	YES
	3	Lung Neoplasms	miRCancer; HMDDv3.2	YES
	4	Stomach Neoplasms	HMDDv3.2	NO
	5	Ovarian Neoplasms	miRCancer; HMDDv3.2	YES
hsa-mir-15a	1	Carcinoma, Hepatocellular	HMDDv3.2	YES
	2	Breast Neoplasms	miRCancer; HMDDv3.2	YES
	3	Melanoma	miRCancer; HMDDv3.2	YES
	4	Lung Neoplasms	HMDDv3.2	YES
	5	Colorectal Neoplasms	miRCancer; HMDDv3.2	NO
hsa-mir-150	1	Carcinoma, Hepatocellular	HMDDv3.2	YES
	2	Colorectal Neoplasms	HMDDv3.2	YES
	3	Lung Neoplasms	HMDDv3.2	YES
	4	Stomach Neoplasms	HMDDv3.2	NO
	5	Breast Neoplasms	HMDDv3.2	YES
hsa-mir-137	1	Lung Neoplasms	HMDDv3.2	YES
	2	Breast Neoplasms	HMDDv3.2	YES
	3	Carcinoma, Hepatocellular	miRCancer; HMDDv3.2	YES
	4	Stomach Neoplasms	miRCancer; HMDDv3.2	NO
	5	Melanoma	miRCancer; HMDDv3.2	YES

Shang *et al. BMC Genomics*   (2022) 23:686

Page 11 of 12

types, and the introduction of two affinity weights by the bidirectional information distribution strategy. FFCV, LOOCV and case studies based on four common diseases and miRNAs are used to evaluate the performance of the BLNIMDA. The experimental results show that BLNIMDA has achieved excellent prediction performance. However, the BLNIMDA still has room for improvement, such as integrating more comprehensive data on the miRNA and disease similarities.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-022-08908-8.

Additional file 1: Figure S1. An example of data processing, including the calculation of GIP kernel similarity and the process of integrating similarity. **Figure S2.** A BLNIMDA calculation example, including the generation of two side information properties, the calculation of two affinity weights for each MDP and the MDA score.

Availability of data and materials
The BLNIMDA is implemented in Matlab. Its source code, user manual and related experimental data are available online at https://github.com/CDMB-lab/BLNIMDA.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Ambros V. The functions of animal microRNAs. Nature. 2004;431(7006):350–5.
2. Dluzen DF, Noren Hooten N, Zhang Y, et al. Racial differences in microRNA and gene expression in hypertensive women. Rep. 2016;6:35815.
3. Chiang K, Liu HB, Rice AP. miR-132 enhances HIV-1 replication. Virology. 2013;438(1):1–4.
4. Mantri CK, Jui PD, Jyoti VM, Chandravanu CV. Cocaine Enhances HIV-1 Replication in CD4+ T Cells by Down-Regulating MiR-125b. PLoS One. 2012;7(12):e51387.
5. Zeng X, Liu L, Lü L, Zou Q. Prediction of potential disease-associated microRNAs using structural perturbation method. Bioinformatics. 2018;34(14):2425–32.
6. Ping P, Wang L, Kuang L, Ye S, Mfb I, Pei T. A Novel method for lncRNA-disease association prediction based on a lncRNA-disease association network. IEEE/ACM Trans Comput Biol Bioinform. 2019;16(2):688–93.
7. Yu J, Ping P, Wang L, Kuang L, Li X, Wu Z. A Novel probability model for lncRNA-disease association prediction based on the Naïve Bayesian Classifier. Genes. 2018;9(7):345.
8. Chen X, Clarence Yan CG, Zhang X, You ZH, Deng LX, Liu Y, Zhang YD, Dai QH. WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. Rep. 2016;6(1):21106.
9. You ZH, Huang Z-A, Zhu ZX, Yan GY, Chen X. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput Biol. 2017;13(3):e1005455.
10. Chen X, Yang JR, Guan NN, Li JQ. GRMDA: Graph Regression for MiRNA-Disease Association Prediction. Front Physiol. 2018;9:92.
11. Jiang Q, Hao Y, Wang G, Juan L, Zhang T, Teng M, Wang LY. Prioritization of disease microRNAs through a human phenome-microRNAome network. BMC Syst Biol. 2010;4:S2.
12. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA–disease associations. Mol Biosyst. 2012;8(10):2792–8.
13. Xuan P, Han K, Guo Y, Li J, Li X, Zhong Y, Zhang Z, Ding J. Prediction of potential disease-associated microRNAs based on random walk. Bioinformatics. 2015;31(11):1805.
14. Chen X, Wang L, Qu J, Guan NN, Li JQ. BNPMDA: Predicting miRNA-disease association based on inductive matrix completion. Bioinformatics. 2018;24:4256–65.
15. Ji C, Gao Z, Ma X, Wu QW, Ni JC, Zheng CH. AEMDA: inferring miRNA-disease associations based on deep autoencoder. Bioinformatics. 2021;37(1):66–72.
16. Tang XR, Luo JW, Shen C, Lai ZH. Multi-view Multichannel Attention Graph Convolutional Network for miRNA–disease association prediction. Brief Bioinform. 2021;22(6):bbab174.
17. Chu YY, Wang X, Dai Q, Wang Y, Wang Q, Peng S, Wei X, Qiu J, Russell SD, Xiong Y. MDA-GCNFTG: identifying miRNA-disease associations based on graph convolutional networks via graph sampling through the feature and topology graph. Brief Bioinform. 2021;22(6):6.
18. Dai Q, Chu Y, Li Z, Zhao Y, Mao X, Wang Y, Xiong Y, Wei D-Q. MDA-CF: Predicting MiRNA-Disease associations based on a cascade forest model by fusing multi-source information. Comput Biol Med. 2021;136:104706.
19. Li Y, Qiu CX, Tu J, Geng B, Yang JC, Jiang TZ, Cui QH. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res. 2014;42(Database issue):D1070-4.
20. Wang D, Wang J, Lu M, Song F, Cui QH. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010;26(13):1644–50.
21. Xuan P, Han K, Guo MZ, Guo YH, Li JB, Ding J, Dai QG, Li J, Teng ZX, Huang YF. Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors. PLoS One. 2013;8(8):e70204.
22. Twan VL, Sander BN, Elena M. Gaussian interaction profile kernels for predicting drug–target interaction. Bioinformatics. 2011;27(21):3036.
23. Zeng X, Wang W, Deng GS, Bing JX, Zou Q. Prediction of Potential Disease-Associated MicroRNAs by Using Neural Networks. Mol Ther-Nucl Acids. 2019;16(7):566–75.
24. Chen X, et al. Semi-supervised learning for potential human microRNA-disease associations inference. Sci Rep-UK. 2014;4:5501.
25. Lu D, Bai XC, Gui L, Li M, Zheng WS, Han XQ, Luo SQ. Arsenic trioxide-induced apoptosis of human malignant lymphoma cell lines and its mechanisms. Acad J First Med Coll PLA. 2003;23(10):997.
26. Chen X, Yan GY. Predicting miRNA-disease association based on inductive matrix completion. Bioinformatics. 2018;24:4256–65.
27. Liu XX, Zhao W, Wang W, Lin S, Yang L. Puerarin suppresses LPS-induced breast cancer cell migration, invasion and adhesion by blockage NF-κB and Erk pathway. Biomed Pharmacother. 2017;92:429–36.
28. Adlakha YK, Saini N. miR-128 exerts pro-apoptotic effect in a p53 transcription-dependent and independent manner via PUMA-Bak axis. Cell Death Dis. 2013;4(3):e542.

Shang *et al. BMC Genomics*   (2022) 23:686

Page 12 of 12

29. Ye Y, Zhuang JH, Wang GY, He SF, Ni J, Xia W. MicroRNA-139 targets fibronectin 1 to inhibit papillary thyroid carcinoma progression. Oncol Lett. 2017;14(6):7799–806.
30. Takamizawa J, Konish H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y, Mitsudomi T, Takahashi T. Reduced Expression of the let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival. Cancer Res. 2004;64(11):3753–6.
31. Cesar L, Laurence AM, Elena AO, Carlos PP, Del Oscar MH, Elizabeth JCO, Sergio RC. MetastamiRs: Non-Coding MicroRNAs Driving Cancer Invasion and Metastasis. Int J Mol Sci. 2012;13(2):1347–79.
32. Sobin LH. The World Health Organization's Histological Classification of Lung Tumors: a comparison of the first and second editions. Cancer Detect Prev. 1982;5(4):391.
33. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.
34. Zhang XL, Zhao HB, Zhang YX, Yang XD, Zhang JX, Yi M, Zhang CJ. The MicroRNA-382-5p/MXD1 Axis Relates to Breast Cancer Progression and Promotes Cell Malignant Phenotypes. J Surg Res. 2020;246:442–9.
35. Wu D, Zhang PL, Ma J, Xu JB, Yang L, Xu WD, Que HF, Chen MF, Xu HT. Serum biomarker panels for the diagnosis of gastric cancer. Cancer Med-US. 2019;8(4):1576–83.
36. Zhou H, Shi JC, Gao YX, Cui CM, Zhang S. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. Nucleic Acids Res. 2018;47(D1):D1013–7.
37. Ahmed M. miRCancerdb_helpers. figshare. Dataset 2017, 5576323.v1.

Publisher's Note