RESEARCH Open Access

A heterogeneous label propagation approach to explore the potential associations between miRNA and disease

Xing Chen^{1*}, De-Hong Zhang¹ and Zhu-Hong You^{2*}

Abstract

Background: Research on microRNAs (miRNAs) has attracted increasingly worldwide attention over recent years as growing experimental results have made clear that miRNA correlates with masses of critical biological processes and the occurrence, development, and diagnosis of human complex diseases. Nonetheless, the known miRNA-disease associations are still insufficient considering plenty of human miRNAs discovered now. Therefore, there is an urgent need for effective computational model predicting novel miRNA-disease association prediction to save time and money for follow-up biological experiments.

Methods: In this study, considering the insufficiency of the previous computational methods, we proposed the model named heterogeneous label propagation for MiRNA-disease association prediction (HLPMDA), in which a heterogeneous label was propagated on the multi-network of miRNA, disease and long non-coding RNA (IncRNA) to infer the possible miRNA-disease association. The strength of the data about IncRNA-miRNA association and IncRNA-disease association enabled HLPMDA to produce a better prediction.

Results: HLPMDA achieved AUCs of 0.9232, 0.8437 and 0.9218 ± 0.0004 based on global and local leave-one-out cross validation and 5-fold cross validation, respectively. Furthermore, three kinds of case studies were implemented and 47 (esophageal neoplasms), 49 (breast neoplasms) and 46 (lymphoma) of top 50 candidate miRNAs were proved by experiment reports.

Conclusions: All the results adequately showed that HLPMDA is a recommendable miRNA-disease association prediction method. We anticipated that HLPMDA could help the follow-up investigations by biomedical researchers.

Keywords: miRNA, Disease, miRNA-disease association, Multi-network, Label propagation

Background

MicroRNAs (miRNAs) consist of about 22 nucleotides and they are one category of endogenous short non-coding RNAs (ncRNAs) that could regulate the expression of target messenger RNAs (mRNAs) at the level of transcription and post-translation [1–4]. There are 28645 miRNAs in the 21st version of miRBase [5] including more than three thousand human miRNAs. As regulators of gene expression and protein production, on the

one hand some of miRNAs serve as negative regulators by binding to the 3'-UTRs of the target mRNAs [4]; on the other hand, the regulatory impact of some miRNAs is positive [6, 7]. Thus miRNAs have effect on cell proliferation [8], development [9], differentiation [10], apoptosis [11], metabolism [12, 13], aging [12, 13], signal transduction [14], and viral infection [10]. Moreover, evidence is mounting that miRNAs play a fundamental role in the development, progression, and prognosis of numerous human diseases [15–20]. For instance, HIV-1 replication could be enhanced by miR-132 [21] and similarly, cocaine could down-regulate miR-125b in CD4+ T cells to enhance HIV-1 replication [22]. Breast neoplasms stem cell formation could be promoted by downregulation

 $^{^2}$ Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science, Ürümqi 830011, China



^{*}Correspondence: xingchen@amss.ac.cn; zhuhongyou@ms.xjb.ac.cn

¹ School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

Chen et al. J Transl Med (2018) 16:348 Page 2 of 14

of miR-140 in basal-like early stage breast cancer [23]. In addition, compared to normal epithelium, miR-139 and miR-140 was down-regulated during lobular neoplasia progression [24]. The transcripts of certain let-7 homologs would be downregulated in human lung cancer and the low levels of let-7 would link to poor prognosis [25]. In addition, non-small-cell lung cancer relates to many other miRNAs [26–29].

Faced with a great variety of miRNAs and diseases, experimental methods for the sake of finding new associations between miRNAs and diseases, are both costly and time-consuming. In the wake of the growth of the biological datasets, the practicable computational methods are urgently necessary to greatly help identify more disease-related miRNAs and explore new perspective treatment of various important human diseases. Over the past decade, some progress has been made to uncover novel miRNA-disease associations. Most computational methods depends on the assumption that functionally similar miRNAs usually have connection with phenotypically similar diseases [30-36]. From the standpoints of network and systems biology, most computational methods belonged to the similarity measure-based approaches or machine learning-based approaches.

A functionally related miRNA network and a human phenome-microRNAome network were first constructed by Jiang et al. [37]. Then the disease phenotype similarity network, miRNA functional similarity network, and the known human disease-miRNA association network were combined together. Based on the combination, they devised a computational model of disease-miRNA prioritization, which could rank the entire human microR-NAome for investigated diseases. However, its prediction performance was ordinary because of only using miRNA neighbor information. Furthermore, Xuan et al. [38] proposed HDMP model to predict disease-related miRNA candidates on the basis of weighted k most similar neighbors. In HMDP, miRNA functional similarity was calculated through the information content of disease terms and disease phenotype similarity. Then, the miRNA family (cluster) information was considered and miRNA functional similarity was recalculated after giving higher weight to members in the same miRNA family (cluster). However, the precision was directly influenced by the number of a miRNA's neighbors. These two methods were limited by their local network similarity measure, which meant it was insufficient to simply consider miRNA neighbor information. Therefore, global network similarity measure was adopted in some studies. Chen et al. [39] proposed Random Walk with Restart for MiRNA-disease association (RWRMDA), in which random walk analysis was applied to miRNA-miRNA functional similarity network. It was a pity that this method was the unavailability for diseases with no confirmed related miRNAs despite of its passable predictive accuracy. Xuan et al. [40] further put forward a random walk method, MIDP, in which transition weights of labeled nodes were higher than unlabeled nodes. In MIDP, the side effect of the noisy data was reduced by fitting restart rate and MIDP is applicable for the disease with no related miRNAs.

Some other methods made use of the information about confirmed disease-related genes and predicted miRNA-target interactions. For instance, Shi et al. [41] developed a computational prediction method in which random walk analysis was used in the protein-protein interaction (PPI) networks. It is assumed that if a target gene of a miRNA associates with a disease, this disease is likely to be related with the miRNA. MiRNA-target interactions and disease-gene associations were integrated into a PPI network and then the functional relationship information about miRNA targets and disease genes was dug out in this PPI network. Besides, this method could serve to find miRNA-disease co-regulated modules by hierarchical clustering analysis. Mørk et al. [42] presented miRPD in which miRNA-protein-disease associations, not just miRNA-disease associations, were predicted. It was a good idea to bring in the abundant information of protein as a bridge indirectly linking the miRNA and the disease. In detail, known and predicted miRNA-protein associations were coupled with protein-disease associations from the literature to make an inference about miRNA-disease associations. In fact, the molecular bases for human diseases we had partly known accounted for less than 40% and highly accurate miRNA-target interactions can hardly be obtained. In other words, above two methods lacked solid data foundation. Chen et al. [43] proposed a model based on super-disease and miRNA for potential miRNA-disease association prediction (SDM-MDA). In view of the fact that rare miRNA-disease associations were known and many associations are 'missing', the concepts of 'super-miRNA' and 'super-disease' were introduced to improve the similarity measures of miR-NAs and diseases.

The computational methods based on machine learning could bring us some new inspiration. Xu et al. [44] constructed the miRNA-target dysregulated network (MTDN) and introduced support vector machine (SVM) classifier based on the features and changes in miRNA expression to distinguish positive miRNA-disease associations from negative associations. However, there was little confirmed information about negative samples, so improvement was needed. In view of the lack of negative samples, Chen et al. [45] developed a semi-supervised method named Regularized Least Squares for MiRNA-disease association (RLSMDA). In the framework of

Chen et al. J Transl Med (2018) 16:348 Page 3 of 14

regularized least squares, RLSMDA was a global method integrating disease semantic similarity, miRNA functional similarity and human miRNA-disease associations. RLSMDA could simultaneously prioritize all the possible miRNA-disease associations without the need of negative samples. Chen et al. [46] proposed Restricted Boltzmann machine for multiple types of miRNA-disease association prediction (RBMMMDA) by which four types of miRNA-disease associations could be identified. RBM-MMDA is the first model which could identify different types of miRNA-disease associations. There is a hypothesis that by distributional semantics, information attached to miRNAs and diseases can be revealed. Pasquier and Gardès [47] developed a model named MirAI, in which the hypothesis was investigated by expressing distributional information of miRNAs and diseases in a highdimensional vector space and then associations between miRNAs and diseases could be defined considering their vector similarity. Chen et al. [39] introduced KNN algorithm into miRNA-disease association prediction and proposed the computational model of RKNNMDA (Ranking-based KNN for MiRNA-disease association prediction).

Some previous researches paid attention to the network tool-based prediction model. For instance, Xuan et al. [40] divided network nodes into labeled nodes and unlabeled nodes and gave them different transition weights. The restart of walking could determine the walking distance, so the negative effect of noisy data would be lessened. Specially, the information from different layers of the miRNA-disease bilayer network was weighed differently. Then, Chen et al. [48] developed Within and Between Score for MiRNA-disease association prediction (WBSMDA) in which for the first time, Gaussian interaction profile kernel similarity for diseases and miRNAs were combined with miRNA functional similarity, disease semantic similarity and miRNA-disease associations. Chen et al. [49] further proposed Heterogeneous graph inference for miRNA-disease association prediction (HGIMDA) and the heterogeneous graph was constructed by the combination of miRNA functional similarity, disease semantic similarity, Gaussian interaction profile kernel similarity, and miRNA-disease associations. Similar to random walk, HGIMDA was an iterative process for the optimal solutions based on global network similarity. In aspect of AUC, HGIMDA reached 0.8781 and 0.8077 after implementing global and local LOOCV, respectively. Li et al. [50] put forward MCMDA (Matrix Completion for MiRNA-disease association prediction) in which a matrix completion algorithm was introduced and the lowly ranked miRNA-disease matrix was updated efficiently. WBSMDA, HGIMDA and MCMDA apply to the disease (miRNA) without any proved related miRNAs (diseases). MaxFlow is a combinatorial prioritization algorithm proposed by Yu et al. [51]. Besides the same type of data used in WBSMDA, MaxFlow also introduced the information about disease phenotypic similarity, miRNA family and miRNA cluster. Then a directed miRNAome-phenome network graph was constructed and every weighted edges were seen as flow capacity. The association possibility was defined as the flow quantity from the miRNA node to the investigated disease node. You et al. [52] proposed Path-Based computational model for MiRNA-disease association prediction (PBMDA). A heterogeneous graph, including three interlinked sub-graphs, was constructed by the same data as in WBSMDA and depth-first search algorithm was applied to predict possible existing miRNAdisease associations. Chen et al. [53] summed up the relatively important miRNA-disease association prediction approach.

More links should exist between miRNAs and diseases than we had learned. However, the computational methods aforementioned were limited by the utilization of inaccurate information (such as miRNA-target interactions), the selection of parameter values, the combination of different classifiers in the different networks or spaces, etc. In pursuit of the higher predictive accuracy, we proposed heterogeneous label propagation for MiRNA-disease association prediction (HLPMDA) for underlying miRNA-disease association prediction. In HLPMDA, heterogeneous data (miRNA similarity, disease similarity, miRNA-disease association, long non-coding RNA (lncRNA)-disease association and miRNA-lncRNA interaction) were integrated into a heterogeneous network [54]. Then, disease-related miRNA prioritization problem was formulated as an optimization problem. In details, within-network smoothness and cross-network consistency were considered here. HLPMDA achieved AUCs of 0.9232, 0.8437 and 0.9218 ± 0.0004 based on global/local LOOCV and 5-fold cross validation, respectively. Both in local and global LOOCV, HLPMDA was better than previous methods. In the case studies of three human diseases, 47, 49 and 46 out top 50 predicted miR-NAs for esophageal neoplasms, breast neoplasms and lymphoma were verified by some recent experimental research.

Methods

Human miRNA-disease associations

There are 5430 human miRNA-diseases associations between 383 diseases and 495 miRNAs, which were obtained from the Human microRNA Disease Database version 2.0 [55]. For convenience, the adjacency matrix $S_{1,2}$ represented known miRNAs-disease associations. If miRNA m(j) is associated with disease d(i), $S_{1,2}(i, j) = 1$;

Chen et al. J Transl Med (2018) 16:348 Page 4 of 14

otherwise, $S_{1,2}(i, j) = 0$. In addition, variable nm and nd indicated the number of involved miRNAs and diseases, respectively.

IncRNA-disease associations

Because we aim to predict latent miRNA-disease association, we looked for the lncRNAs that associate with the disease contained in $S_{1,2}$, or interacted with the miR-NAs contained in $S_{1,2}$. As a result, 1089 lncRNAs (from LncRNADisease database [56] and starBase v2.0 database [57] matched the above conditions. For the convenience of subsequent calculations, the adjacency matrix $S_{2,3} \in \mathbb{R}^{383 \times 1089}$ was constructed to represent known lncRNA-disease associations. If lncRNA l(j) is associated with disease d(i), $S_{2,3}$ (i, j)=1; otherwise, $S_{2,3}$ (i, j)=0. Variable nl means the number of involved lncRNAs. The known lncRNA-disease associations came from LncRNA database disease (http://www.cuilab.cn/lncrnadise ase) which provided many experimentally confirmed lncRNA-disease associations and we deleted duplicate associations with different evidences. Finally 251 different confirmed lncRNA-disease associations were selected out and in fact they only had something to do with 150 lncRNAs and 63 diseases so $S_{2.3}$ was a sparse matrix.

miRNA-IncRNA interactions

Similarly, the adjacency matrix $S_{1,3} \in R^{495 \times 1089}$ was constructed to represent known miRNA–lncRNA interaction. If miRNA ms(i) is interacted with lncRNA l(j), $S_{1,3}$ (i,j)=1; otherwise, $S_{1,3}$ (i,j)=0. MiRNA–lncRNA interaction dataset was downloaded from starBase v2.0 database [57] (http://starbase.sysu.edu.cn/), which provided the most comprehensive experimentally confirmed miRNA–lncRNA interactions based on large scale CLIP-Seq data. Then we deleted duplicate interactions and 9088 different confirmed lncRNA–miRNA interactions were selected out. Similar to $S_{2,3}$, $S_{1,3}$ was also a sparse matrix in which the interactions were only about 246 miRNAs rather than all the 495 miRNAs.

MiRNA functional similarity

It was assumed in the previous work [58] that functional similar miRNAs often correlate with phenotypically similar diseases. Based on this important assumption, miRNA functional similarity score was calculated and the related data could be downloaded from http://www.cuilab.cn/files/images/cuilab/misim.zip. Analogously, the miRNA functional similarity network was represented by miRNA functional similarity matrix FS, in which functionally similar between miRNA m(i) and m(j) is denoted by the entity FS(m(i), m(j)).

Disease semantic similarity model

There are two kinds of models to calculate disease semantic similarity. Directed acyclic graph (DAG) is a finite directed graph but there is no directed circle in it. DAG consists of finite vertices and edges, with each edge directed from one node (parent) to another (child), and it is impossible to start at a node n and follow a consistently-directed sequence of edges that eventually loops back to n again. DAG served as a tool to describe the relationships among involved diseases in many previous studies [45, 48, 49, 52]. According to the data from the National Library of Medicine (http://www.nlm. nih.gov/), the relationship of different diseases could be measured by the disease DAG based on the MeSH descriptor of Category C. For example, for the DAG of esophageal neoplasms (see Fig. 1), 'Neoplasms' points to 'Neoplasms by Site', so 'Neoplasms' is the parent of child 'Neoplasms by Site'. The disease D was represented by DAG(D) = (D, T(D), E(D)), in which T(D) is the node set representing disease D itself and its ancestor (its parent and above), E(D) is the corresponding direct edges from the parent to the child [58]. According to [38], the semantic value of disease D could be calculated as follows:

$$DV(D) = \sum_{d \in T(D)} D_D(d)$$
 (1)

where

$$D_D(d) = \begin{cases} 1, & \text{if } d = D \\ \max\{\Delta * D_D(d') | d' \in \text{children of } d\}, & \text{if } d \neq D \end{cases}$$
(2)

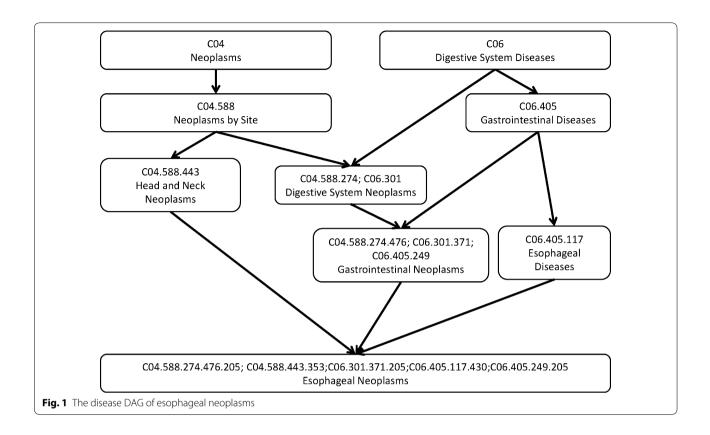
where Δ is the semantic contribution factor. For disease D, the contribution of itself to the semantic value of disease D was 1 and the longer distance between D and other disease was, the smaller semantic contribution was. If disease terms are in the same layer, they would have the same contribution to the semantic value of disease D.

There is a wildly accepted assumption that the more part of two diseases' DAGs are sharing, the more semantic similarity they have. The semantic similarity between disease d(i) and d(j) can be defined as follows:

$$DS1(d(i), d(j)) = \frac{\sum_{t \in T(d(i)) \cap T(d(j))} D_{d(i)}(t) + D_{d(j)}(t)}{DV(d(i)) + DV(d(j))}$$
(3)

Furthermore, there is another model for disease similarity calculation [38] and it was adopted in this study. It is observed that in the same layer of DAG(A), different diseases terms may appear in the different numbers of disease DAGs. For instance, there are two diseases in the same layer of DAG(A), if one disease appears in less disease DAGs than the other, it is obvious that the former is more specific than the latter. So we assigned them with different contributions, and the former's contribution

Chen et al. J Transl Med (2018) 16:348 Page 5 of 14



factor should be higher than the latter. The contribution of disease term t in DAG(A) to the semantic value of disease A is defined as follows:

$$C2A(t) = -\log\left(\frac{DAG_t}{nd}\right) \tag{4}$$

where DAG_t represents the number of DAGs including t. The semantic similarity between two diseases were defined as follows:

$$DS2(d(i), d(j)) = \frac{\sum_{t \in T(A) \cap T(B)} C2_A(t) + C2_B(t)}{C2(d(i)) + C2(d(j))}$$
(5)

So the final disease semantic similarity was defined as follows:

$$DS = \frac{DS1 + DS2}{2} \tag{6}$$

Gaussian interaction profile kernel similarity for diseases and miRNAs

In order to make the most of the topologic information from known miRNA-disease association network, Gaussian interaction profile kernel similarity for diseases are calculated on the assumption that analogic diseases are likely to associate with functionally similar miRNAs and vice versa [20, 58–60]. The ith row of the adjacency matrix $S_{1,2}$ is taken out as a new binary vector, IP(d(i)). Obviously, IP(d(i)) illustrate the associative or non-associative

situation between disease d(i) and all miRNAs involved in this study and it is called interaction profiles of disease d(i). According to [61], Gaussian kernel similarity between two diseases, d(i) and d(j), could be calculated as follows:

$$KD(d(i), d(j)) = exp(-\gamma_d || IP(d(i)) - IP(d(j))||^2)$$

where γ_d is a parameter for the kernel bandwidth control, and it was calculated through the normalization of a new bandwidth parameter Y_d' by the average number of associations with miRNAs for all the diseases.

$$\gamma_d = \frac{\gamma_d'}{\frac{1}{n^d} \sum_{i=1}^{nd} \|D(d(i))\|^2}$$
 (8)

Similarly, Gaussian interaction profile kernel similarity between two miRNAs (m(i) and m(j)) is calculated as follows:

$$KM(m(i), m(j)) = exp(-\gamma_m || IP(m(i)) - IP(m(j)) ||^2)$$
(9)

$$\gamma_m = \gamma'_m / \left(\frac{1}{nm} \sum_{i=1}^{nm} \| IP(m(i)) \|^2 \right)$$
 (10)

where IP(m(i)) and IP(m(j)) represent *i*th column and the *j*th column of the adjacency matrix $S_{1,2}$; γ_m is a

Chen et al. J Transl Med (2018) 16:348 Page 6 of 14

parameter for the kernel bandwidth control, and it was calculated through the normalization of a new bandwidth parameter Y_m' by the average number of associated diseases for all the miRNAs. According to [62] and for the simplicity of calculations, we set $\gamma_d = \gamma_m = 1$.

Integrated similarity for miRNAs and diseases

Here, according to [48], let S_1 represent the integrated miRNA similarity matrix and S_2 be the integrated disease similarity matrix.

The hetero-sub-networks (bipartite networks) are $G_{i,j} = (V_i \cup V_j, E_{i,j})$ for i,j=1,2,3, and i< j, where i,j=1,2,3 for miRNAs, diseases and lncRNAs, respectively. E_i represents the set of edges between vertices in the vertex set V_i of homo-sub-network G_i . And $E_{i,j}$ represents the set of edges between a vertex in V_i to a vertex in V_i .

On the base of heterogeneous network G, we measure the weight of homo-sub-network edge (i, j) by bipartite

$$S_1(m(i), m(j)) = \begin{cases} FS(m(i), m(j)), & \text{if } m(i) \text{ and } m(j) \text{ have functional similarity} \\ KM(m(i), m(j)), & \text{otherwise} \end{cases}$$
 (11)

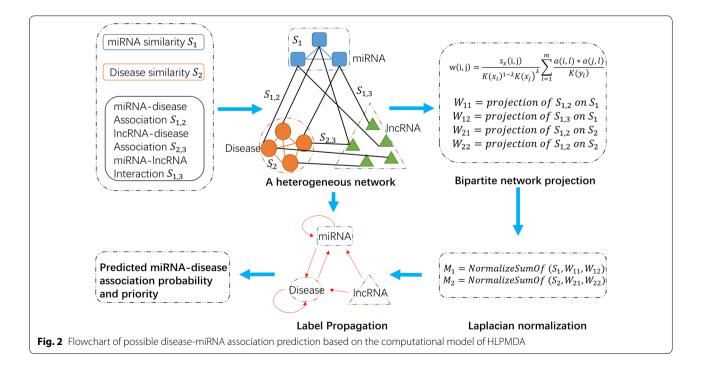
$$S_2(d(i), d(j)) = \begin{cases} DS(d(i), d(j)), & \text{if } d(i) \text{ and } d(j) \text{ have semantic similarity} \\ KD(d(i), d(j)), & \text{otherwise} \end{cases}$$
(12)

HLPMDA

HLPMDA is motivated by Heter-LP [63]. As shown in Fig. 2, the heterogeneous network constructed based on the above data included three kinds of nodes (miRNAs, diseases, and lncRNAs) and five kinds of edges (miRNA similarity, disease similarity, miRNA-disease association, miRNA-lncRNA interaction and lncRNA-disease association). Thus a heterogeneous network G=(V, E) was constructed with two homo-sub-networks and three hetero-sub-networks (see Fig. 2). The homo-sub-networks are defined as $G_i=(V_pE_i)$ where i=1, 2 for miRNAs and diseases, respectively.

network projection, a weighted one-mode projection technique from [63, 64]. Let the adjacency matrix A represent one bipartite network, in which there are two nonempty disjoint vertex sets X and Y. S_x is the similarity matrix of vertex set X and S_x (i, j) is the entry of row i and column j in S_x ; $K(x_i)$ represents the degrees of vertices x_i in G; W is the projected matrix of A onto X and the corresponding calculation process is:

$$w(i,j) = \frac{s_x(i,j)}{K(x_i)^{1-\lambda}K(x_j)^{\lambda}} \sum_{l=1}^{m} \frac{a(i,l)*a(j,l)}{K(y_l)}$$
(13)



Chen et al. J Transl Med (2018) 16:348 Page 7 of 14

where i,j belong to identical homo-sub-networks; w(i, j) is the entry of row i and column j in W; 0 < k < 1 is diffusion parameter of the projection (in this study we set k = 0.5); a(i, l) represents the weight of edge (x_i, y_l) in G. If there is no edge from i to j, w(i, j) = 0.

Next, label propagation was applied on miRNA-disease hetero-sub-network by means of the information from other homo-sub-networks and hetero-sub-networks. Table 1 shows the main pseudo-code of HLP-MDA. Firstly, let y_1 , y_2 and y_3 be the label vectors that represent miRNA, disease and lncRNA, respectively. y_1 , y_2 and y_3 were initialized to zero. Secondly, all associations ($S_{1,2}$ and $S_{2,3}$) and interactions ($S_{1,3}$) were projected onto similarity matrices (S_1 and S_2) using the weighted one-mode projection technique as described above. Four projected matrices came out (W_{11}) is the projection of $S_{1,2}$ on S_1 ; W_{12} is the projection of $S_{1,3}$ on S_1 ; W_{21} is the projection of $S_{1,2}$ on S_2 ; W_{22} is the projection of $S_{2,3}$ on S_2). Thirdly, four projected matrices $(W_{11}, W_{12} \text{ and } W_{21}, W_{22})$ were integrated with corresponding similarity matrices $(S_1 \text{ or } S_2)$ respectively, with the help of the Laplacian normalization (M_1 is the Laplacian normalization of S_1 , W_{11} and W_{12} ; M_2 is the Laplacian normalization of S_2 , W_{21} and W_{22}). Taking M_1 as an example, the Laplacian normalization is defined by

$$M(i,j) = S_1(i,j) + W_{11}(i,j) + W_{12}(i,j)$$
 (14)

$$M(i,j) = \begin{cases} 1, & i = j \\ \frac{M(i,j)}{\sqrt{d(i)d(j)}}, & i \neq j \end{cases}$$
 (15)

where d(i) is the sum of ith row of the matrix M, and if d(i) = 0, d(i) = 1.

Then in label propagation phase, there were three iterative loops. In each loop, the label of the investigated miRNA (disease or lncRNA) was set to one and others to zero. The label propagation function is applied, and output matrices, $F_{1,2}$ and $F_{2,1}$, are updated. Finally, the predictive matrix F for underlying miRNA-disease associations could be obtained and then all predictive scores could be ranked in descending order.

According to the previous study [63], the convergence of label propagation iteration (*LabelPropagation* function) in the algorithm HLPMDA could be determined (the relevant proof can be found in [63]). So in order to reduce the time complexity and space complexity of HLPMDA, the complex part, i.e. *Label-Propagation* function was replaced by the following equation:

Table 1 The illustration of the HLPMDA algorithm

Algorithm: HLPMDA

Input:

- (1) σ : total convergence threshold
- (2) σ' : homogenous convergence threshold
- (3) α : diffusion parameter of label propagation
- (4) λ : diffusion parameter of projection
- (5) y_1, y_2, y_3 : vectors of initial label values
- (6) S_1, S_2 : homo-subnetwork matrices
- (7) $S_{1,2}$, $S_{1,3}$, $S_{2,3}$: hetero-subnetwork matrices

Output:

F: hetero-subnetwork matrices of final label values

- $F_{1,2} = F_{2,1} = 0$
- 2 $W_{11} = projection of S_{1,2} on S_1$
- $3 \hspace{1cm} W_{12} = projection \ of \ S_{1,3} \ on \ S_{1}$
- $4 W_{21} = projection of S_{1,2} on S_2$
- 5 $W_{22} = projection of S_{2,3} on S_2$
- $6 M_1 = NormalizeSumOf(S_1, W_{11}, W_{12})$
- 7 $M_2 = NormalizeSumOf(S_2, W_{21}, W_{22})$
- For i=1 to y_1 .length
 - 8.1 $y_1[i] = 1, y_1[j] = 0 \text{ for all } j \neq i$
 - $y_2 = y_3 = 0$
 - 8.3 $f_1 = (I \alpha M_1)^{-1} [(1 \alpha)^2 y_1 + (1 \alpha)^3 S_{1,2} y_2 + (1 \alpha)^3 S_{1,3} y_3]$
 - 8.4 $F_{1,2}[i,] = f_1$
- For i=1 to y_2 .length
 - 9.1 $y_2[i] = 1, y_1[j] = 0 \text{ for all } j \neq i$
 - $9.2 y_1 = y_3 = 0$
 - 9.3 $f_2 = (I \alpha M_2)^{-1} [(1 \alpha)^2 y_2 + (1 \alpha)^3 S_{2,1} y_1 + 1 \alpha 3S2,3y3$
 - .4 $F_{2,1}[,i] = f_2$
- $F = (F_{1,2} + F_{2,1})/2$

Chen et al. J Transl Med (2018) 16:348 Page 8 of 14

Table 1 (continued)

```
NormalizeSumOf (S, W, W'):
        d = 0 //a vector with S.numberOfRows length
        for i = 1 to S.numberOfRows
                   for i=1 to S.numberOfColumns
        12.1
                            M[i, j] = S[i, j] + W[i, j] + W'[i, j]
                12.1.2
                            d[i] = d[i] + M[i, j]
                   if (d/i) = 0 d/i = 1
        for i = 1 to M.numberOfRows
        13.1
                   for j=1 to M.numberOfColumns
                           if (i=j) M[i, j] = 1
                            else if (M[i, j]!=0) M[i, j] = \frac{M[i, j]}{\sqrt{d[i]d[j]}}
14
        return M
```

is the comparison range. In local LOOCV, a comparison was made between test sample and the miRNAs without known association with the investigated disease. Whereas in global LOOCV, a comparison was made between test sample and all the miRNA-disease pairs without confirmed associations. In 5-fold cross validation, all the known miRNA-disease associations in HMDD v2.0 were divided into five sets with equal sizes, where four sets trained the model and the other set tested the model. For fear of the performance difference due to the samples divisions, all associations were randomly divided 100 times and the results of all 100 times were averaged to derive the final evaluation result.

If the test sample ranked higher than the given threshold, it was a successful prediction. Next, Receiver operating characteristics (ROC) curve was drawn where true positive rate (TPR, sensitivity) was plotted versus false positive rate (FPR, 1-specificity) at different thresholds. Sensitivity represents the ratio of successful predictions to the test samples. Specificity represents the percentage of negative miRNA-disease pairs which were ranked lower than the threshold. Area under the ROC curve (AUC) could be calculated to show predictive capability of MDMMDA. The closer that AUC is to 1, the better predictive capability the method is. AUC=0.5 means the random performance.

$$f_1 = (I - \alpha M_1)^{-1} \left[(1 - \alpha)^2 y_1 + (1 - \alpha)^3 S_{1,2} y_2 + (1 - \alpha)^3 S_{1,3} y_3 \right]$$
(16)

$$f_2 = (I - \alpha M_2)^{-1} \left[(1 - \alpha)^2 y_2 + (1 - \alpha)^3 S_{2,1} y_1 + (1 - \alpha)^3 S_{2,3} y_3 \right]$$
(17)

where f_1 and f_2 are label vectors that represent the predictive result for the investigated miRNA with all diseases or the investigated disease with all miRNAs; I is the identity matrix; $S_{2,1} = (S_{2,1})^T$; α is a constant parameter and we set $\alpha = 0.1$ referring to the similar study [63].

Results

Cross validation

In order to evaluate the predictive performance of HLP-MDA, global LOOCV, local LOOCV and 5-fold cross validation were executed based on the known miRNA-disease associations from HMDD v2.0 [55]. Then, HLPMDA was compared with ten state-of-the-art computational methods: PBMDA [52], MCMDA [50], Max-Flow [51], HGIMDA [49], RLSMDA [45], HDMP [38] WBSMDA [48], MirAI [47], MIDP [40] and RWRMDA [65].

In LOOCV, each proved miRNA-disease association was regarded as a test sample in turn while other known associations were used as training set of the model. The difference between local and global LOOCV

As illustrated in Fig. 3, HLPMDA achieved AUCs of 0.9232, 0.8437 and 0.9218 ± 0.0004 in the global LOOCV, local LOOCV and 5-fold CV, respectively, which shows a better predictive capability than other ten methods: PBMDA [52], MCMDA [50], MaxFlow [51], HGIMDA [49], RLSMDA [45], HDMP [38] WBSMDA [48], MirAI [47], MIDP [40] and RWRMDA [65]. (RWRMDA and MIDP are random walk-based method and this two method could be implemented only after determine the disease, so there are no global LOOCV results about them. MiRAI lacked the results of global LOOCV, either. Because during the caculation of MiRAI, the association scores for different diseases were not comparable.) Besides, MiRAI implemented on our data sets had a lower AUC (0.6299) than described in the origin literature [47], due to the data sparsity problem of collaborative filtering algorithm that MiRAI was based on.

Case studies

To be specific, three malignant human diseases, esophageal neoplasms, breast neoplasms and Lymphoma were

Chen et al. J Transl Med (2018) 16:348 Page 9 of 14

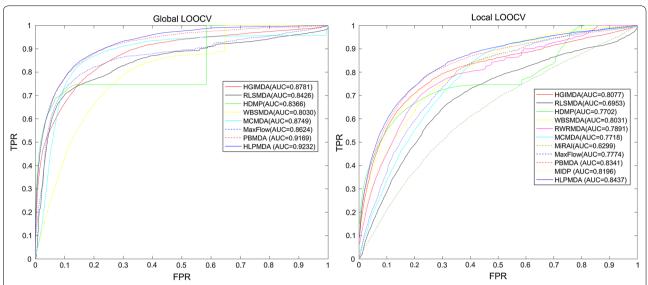


Fig. 3 Predictive capability comparisons between HLPMDA and ten classical models of disease-miRNA association prediction (PBMDA, MCMDA, MaxFlow, HGIMDA, RLSMDA, HDMP, WBSMDA, MirAl, MIDP, and RWRMDA) in terms of ROC curve and AUC based on local and global LOOCV, respectively. As a result, HLPMDA achieved AUCs of 0.9232 and 0.8437 in the global and local LOOCV, significantly outperforming all the previous classical models

selected out to execute three kind of case studies (each kind of case studies investigate one disease).

In the first kind of case studies, data came from HMDD v2.0 and then the prediction results were checked up in miR2Disease [66] and dbDEMC database [67] (another two well-known miRNA-disease association databases). This kind of case studies is about esophageal neoplasms. Esophageal neoplasm is a common malignant tumor worldwide and it affects more males than females [68]. In terms of pathological characteristics, there are two main subtype of esophageal neoplasms: esophageal squamous cell carcinoma (ESCC) and esophageal adenocarcinoma (EAC) [68]. ESCC remains the main subtype of esophageal neoplasms [68]. Survival rate of esophageal neoplasms is improving but remains poor [69]. So more esophageal neoplasms related miRNAs may help detect, diagnose and treat esophageal neoplasms earlier. Until now, some miRNAs have been found associated with esophageal neoplasms. For example, after 24- and/ or 72-h treatment of esophageal neoplasms by Chemotherapy, 13 miRNAs (miR-199a-5p, miR-302f, miR-320a, miR-342-3p, miR-425, miR-455-3p, miR-486-3p, miR-519c-5p, miR-548d-5p, miR-617, miR-758, miR-766, miR-1286) were deregulated [70]. By HLPMDA, the candidate miRNAs of esophageal neoplasms were ranked and then checked up by miR2Disease and dbDEMC. As a result, all of the top 10 and 47 out of the top 50 candidate miRNAs could be proved to be related with esophageal neoplasms (see Table 2). Besides, all candidate miRNAs were ranked by HLPMDA for all the diseases in HMDD

v2.0 (see Additional file 1). We hope that these prediction results could help the corresponding experimental research in the future.

In the second kind of case studies, data also came from HMDD v2.0 but the investigated disease-related miRNAs were removed in order to evaluate the predictive capability for those diseases without any known associated miRNAs. Then the prediction results were checked up in HMDD v2.0, miR2Disease and dbDEMC database. This kind of case studies is about breast neoplasms. Breast neoplasms (Breast cancer) is the second leading cause of women cancer death in the US and the breast cancer death rates of black women remain higher than whites nationally [71]. Some miRNAs have been proved to correlate with Breast neoplasms and the corresponding treatment. For example, by decreasing TrkB and Bmi1 expression, miR-200c sensitizes breast cancer cells to doxorubicin treatment [72]. Furthermore, in human breast cancer cells miRNA-200 family alterations relates to mesenchymal and drug-resistant phenotypes [73]. By HLPMDA, the candidate miRNAs of Breast neoplasms were ranked and then checked up by HMDD v2.0, miR2Disease and dbDEMC. As a result, all of the top 10 and 49 out of the top 50 candidate miRNAs could be proved to be related with Breast neoplasms (see Table 3).

In the third kind of case studies, data came from HMDD v1.0 and then the prediction results were checked up in HMDD v2.0, miR2Disease and dbDEMC database, just for the sake of examining the robustness of HLP-MDA on the different dataset. This kind of case studies

Chen et al. J Transl Med (2018) 16:348 Page 10 of 14

Table 2 HLPMDA was implemented to predict potential esophageal neoplasms-related miRNAs based on the known miRNA-disease association from HMDD v2.0 (left column: top 1–25; right column: top 26–50)

miRNA	Evidence	miRNA	Evidence
hsa-mir-125b	dbDEMC	hsa-mir-127	dbDEMC
hsa-mir-200b	dbDEMC	hsa-mir-429	dbDEMC
hsa-mir-17	dbDEMC	hsa-mir-181b	dbDEMC
hsa-mir-18a	dbDEMC	hsa-mir-181a	dbDEMC
hsa-mir-16	dbDEMC	hsa-mir-107	dbDEMC; miR2Disease
hsa-mir-221	dbDEMC	hsa-mir-24	dbDEMC
hsa-mir-1	dbDEMC	hsa-mir-182	dbDEMC
hsa-mir-222	dbDEMC	hsa-mir-30c	dbDEMC
hsa-let-7e	dbDEMC	hsa-mir-93	dbDEMC
hsa-let-7f	Unconfirmed	hsa-mir-18b	dbDEMC
hsa-mir-19b	dbDEMC	hsa-mir-199b	dbDEMC
hsa-mir-29a	dbDEMC	hsa-mir-133b	dbDEMC
hsa-let-7d	dbDEMC	hsa-mir-132	dbDEMC
hsa-let-7i	dbDEMC	hsa-mir-195	dbDEMC
hsa-mir-9	dbDEMC	hsa-mir-30a	dbDEMC
hsa-let-7 g	dbDEMC	hsa-mir-191	dbDEMC
hsa-mir-218	Unconfirmed	hsa-mir-15b	dbDEMC
hsa-mir-10b	dbDEMC	hsa-mir-302b	dbDEMC
hsa-mir-146b	dbDEMC	hsa-mir-224	dbDEMC
hsa-mir-142	dbDEMC	hsa-mir-194	dbDEMC; miR2Disease
hsa-mir-125a	dbDEMC	hsa-mir-20b	dbDEMC
hsa-mir-7	dbDEMC	hsa-mir-124	dbDEMC
hsa-mir-29b	dbDEMC	hsa-mir-302c	dbDEMC
hsa-mir-106a	dbDEMC	hsa-mir-122	Unconfirmed
hsa-mir-106b	dbDEMC	hsa-mir-335	dbDEMC

is about Lymphoma originating in the lymphatic hematopoietic system, which accounts for more than one-fifth of all cancer cases [71]. According to the tumor cells, there are two categories of lymphoma: Hodgkin lymphomas (HL) and the non-Hodgkin lymphomas (NHL) [74, 75]. It is very hard for HL to be detected at early stages [74, 75]. Some miRNAs were found associated with lymphoma. For instance, there are different expressions of miR-150 between lymphoma and small lymphocytic leukemia [76], and specifically, miR-150 is a tumor suppressor in malignant lymphoma [77]. Besides, EBV-positive Burkitt lymphoma differentiation can be induced by reexpression of miR-150 targeting c-Myb [78]. By HLP-MDA, the candidate miRNAs of lymphoma were ranked and then checked up by HMDD v2.0, miR2Disease and dbDEMC. As a result, 9 of the top 10 and 46 out of the top 50 candidate miRNAs could be proved to be related with lymphoma (see Table 4).

Discussion

The reliability and availability of HLPMDA lied in the following several aspects. Firstly, HMDD as well as other biological datasets provided a solid foundation for the subsequent prediction steps. Secondly, the introduction of lncRNA data and the application of bipartite network projection help profile the relationship between one miRNA and another miRNA, between one disease and another disease. There is a widely accepted view that more data may help produce a better output. Adding the corresponding lncRNA data brings more information to the problem of latent miRNA-disease association prediction. It is a fresh perspective and it was proved to be an advantageous improvement by the performance of HLP-MDA. Bipartite network projection also dug out more implicit message that made the prediction more accurate. In addition, the heterogeneous label propagation is a useful algorithm based on the local and global feature in the constructed network, with no need of negative examples. In recent years, the network approach has been relatively widely adopted in some fields of bioinformatics [79–81]. The major cause is that similarity, links, associations, interactions and relationships among the research targets (like miRNA, diseases and so on) in the network approach become easier to be represented, calculated, analyzed and tested by some math tools, together with some descriptive expressions transformed into quantitative representations. As a result, it indeed helps improve the effectiveness of the prediction. Finally, according to NanoString's Hallmarks of Cancer Panel collection (https://doi.org/10.1016/j.j.ph.) ://www.nanostring.com/), it is proved that a part of the miRNAs' targets is related to cancer hallmarks [82, 83], which were found to be associated with the corresponding genes. So our work may be helpful for the further research about cancer hallmarks, genes and miRNA.

However, HLPMDA is undeniably limited by following factors which are also the room to improve HLPMDA. First, the data about miRNA and disease is not ample enough. For instance, the known miRNA-disease associations have a large degree of sparsity (labeled miRNAdisease associations only accounts for 2.86% of 189,585 miRNA-disease pairs). It is believed that more data could promote the performance of the computational model. Therefore, with more information about miRNA, disease and some other objects (like genes, drugs, targets and so on) related to one or both of them put to use [84], predictive power of HLPMDA would be stronger. Second, it may be unfair for different miRNAs or diseases because the known information about every item is not relatively equivalent. Therefore, HLPMDA may cause advantageous bias to miRNAs or diseases which have more

Chen et al. J Transl Med (2018) 16:348 Page 11 of 14

Table 3 HLPMDA was implemented to predict potential breast neoplasms-related miRNAs based on the known miRNA-disease association from HMDD v2.0 while the associations about breast neoplasms were removed and then the prediction results were checked up in HMDD v2.0, miR2Disease and dbDEMC database (left column: top 1–25; right column: top 26–50)

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	HMDD dbDEMC miR2Disease	hsa-mir-205	HMDD; dbDEMC miR2Disease
hsa-mir-155	HMDD dbDEMC miR2Disease	hsa-mir-203	HMDD; dbDEMC miR2Disease
hsa-mir-125b	HMDD miR2Disease	hsa-mir-1	HMDD; dbDEMC
hsa-mir-145	HMDD; dbDEMC miR2Disease	hsa-mir-34c	HMDD; dbDEMC
hsa-mir-146a	HMDD dbDEMC miR2Disease	hsa-mir-19b	HMDD; dbDEMC
hsa-mir-221	HMDD dbDEMC miR2Disease	hsa-mir-375	HMDD; dbDEMC
hsa-mir-17	HMDD miR2Disease	hsa-mir-199a	HMDD; dbDEMC
hsa-mir-34a	HMDD dbDEMC	hsa-mir-100	HMDD; dbDEMC
hsa-let-7a	HMDD dbDEMC miR2Disease	hsa-mir-29a	HMDD; dbDEMC
hsa-mir-20a	HMDD miR2Disease	hsa-mir-210	HMDD; dbDEMC miR2Disease
hsa-mir-31	HMDD dbDEMC miR2Disease	hsa-mir-15a	HMDD; dbDEMC
hsa-mir-200c	HMDD dbDEMC miR2Disease	hsa-let-7c	HMDD; dbDEMC
hsa-mir-222	HMDD dbDEMC miR2Disease	hsa-mir-182	HMDD; dbDEMC miR2Disease
hsa-mir-126	HMDD dbDEMC miR2Disease	hsa-mir-34b	HMDD; dbDEMC
hsa-mir-18a	HMDD dbDEMC miR2Disease	hsa-mir-183	HMDD; dbDEMC
hsa-mir-92a	HMDD	hsa-mir-141	HMDD; dbDEMC miR2Disease
hsa-mir-200b	HMDD dbDEMC miR2Disease	hsa-mir-101	HMDD; dbDEMC miR2Disease
hsa-mir-16	HMDD dbDEMC	hsa-mir-146b	HMDD; dbDEMC miR2Disease
hsa-let-7b	HMDD dbDEMC	hsa-mir-181a	HMDD; dbDEMC miR2Disease
hsa-mir-142	Unconfirmed	hsa-let-7d	HMDD; dbDEMC miR2Disease

Table 3 (continued)

miRNA	Evidence	miRNA	Evidence
hsa-mir-9	HMDD dbDEMC miR2Disease	hsa-mir-10b	HMDD; dbDEMC miR2Disease
hsa-mir-200a	HMDD dbDEMC miR2Disease	hsa-let-7e	HMDD; dbDEMC
hsa-mir-223	HMDD dbDEMC	hsa-mir-29c	HMDD; dbDEMC miR2Disease
hsa-mir-19a	HMDD dbDEMC	hsa-mir-218	HMDD; dbDEMC
hsa-mir-143	HMDD dbDEMC miR2Disease	hsa-let-7 g	HMDD; dbDEMC

Table 4 HLPMDA was implemented to predict potential lymphoma-related miRNAs based on the known miRNA-disease association from HMDD v1.0 and then the prediction results were checked up in HMDD v2.0, miR2Disease and dbDEMC database (left column: top 1–25; right column: top 26–50)

miRNA	Evidence	miRNA	Evidence
hsa-mir-21	HMDD; dbDEMC	hsa-mir-127	dbDEMC
hsa-mir-155	HMDD; dbDEMC	hsa-mir-141	dbDEMC
hsa-mir-221	dbDEMC	hsa-mir-199a	dbDEMC
hsa-let-7a	dbDEMC	hsa-mir-29c	HMDD; dbDEMC
hsa-mir-146a	HMDD; dbDEMC	hsa-mir-24	HMDD; dbDEMC
hsa-mir-125b	Unconfirmed	hsa-mir-146b	Unconfirmed
hsa-mir-222	dbDEMC	hsa-mir-200a	HMDD; dbDEMC
hsa-mir-34a	dbDEMC	hsa-mir-143	dbDEMC
hsa-let-7b	dbDEMC	hsa-mir-181a	HMDD; dbDEMC
hsa-let-7e	dbDEMC	hsa-mir-25	dbDEMC
hsa-mir-223	dbDEMC	hsa-mir-150	HMDD; dbDEMC
hsa-let-7d	dbDEMC	hsa-mir-106a	dbDEMC
hsa-mir-29b	dbDEMC	hsa-mir-125a	HMDD; dbDEMC
hsa-mir-29a	dbDEMC	hsa-mir-373	dbDEMC
hsa-mir-145	dbDEMC	hsa-mir-101	HMDD; dbDEMC
hsa-mir-9	dbDEMC	hsa-mir-451	dbDEMC
hsa-mir-200b	HMDD; dbDEMC	hsa-mir-93	dbDEMC
hsa-let-7i	dbDEMC	hsa-mir-214	dbDEMC
hsa-let-7f	dbDEMC	hsa-mir-34c	Unconfirmed
hsa-let-7c	dbDEMC	hsa-mir-34b	dbDEMC
hsa-let-7 g	dbDEMC	hsa-mir-191	dbDEMC
hsa-mir-126	HMDD; dbDEMC	hsa-mir-339	dbDEMC
hsa-mir-181b	dbDEMC	hsa-mir-194	dbDEMC
hsa-mir-106b	dbDEMC	hsa-mir-132	dbDEMC
hsa-mir-1	dbDEMC	hsa-mir-429	Unconfirmed

Chen et al. J Transl Med (2018) 16:348 Page 12 of 14

known association (or interaction) records. Last but not the least, the parameters in HLPMDA were set according to the previous similar studies and our experience. We have not thought a lot of the parameters but there may exist better parameters which could bring about more accurate prediction results.

Data collection, database construction, data analysis, mining and testing about miRNA-disease associations has become an important field in bioinformatics. As we all know, there are strong connections in many fields of biology. The research of miRNA-disease association relates to protein—protein interaction, miRNA-target interaction, miRNA—lncRNA interaction, drug, environmental factor, etc. In the future, we believe that this field need to obtain more data and to be integrated with other research areas for the sake of producing predictive synergy with more integrated data.

Conclusion

It is valuable to seek the underlying miRNA-disease associations. In this paper, on the grounds that functionally similar miRNAs were likely to correlate with similar diseases and vice versa, heterogeneous label propagation for MiRNA-disease association prediction (HLPMDA) was proposed. AUCs of HLPMDA are 0.9232 (global LOOCV), 0.8437 (local LOOCV) and 0.9218 ± 0.0004 (5-fold CV). In three case studies, the accurate rates were all higher than 85%. Furthermore, three kinds of case studies were implemented for further evaluations. As a result, 47 (esophageal neoplasms), 49 (breast neoplasms) and 46 (lymphoma) of top 50 candidate miRNAs were proved by experiment reports. All the results sufficiently showed the reliability of HLPMDA in predicting possible disease-miRNA associations. HLPMDA will be a valuable computational tool for miRNA-disease association prediction and miRNA biomarker identification for human disease.

Additional file

Additional file 1. All candidate miRNAs were ranked by HLPMDA for all the diseases in HMDD v2.0. Prediction results could be obtained publicly for further research and experimental validation.

Abbreviations

MiRNA: microRNA; LncRNA: long non-coding RNA; LOOCV: leave-one-out cross validation; 5-fold CV: 5-fold cross validation; ROC: receiver-operating characteristics curve; AUC: the area under ROC curve.

Authors' contributions

XC conceived the project, developed the prediction method, designed and implemented the experiments, analyzed the result, and wrote the paper. DHZ designed and implemented the experiments, analyzed the result, wrote the paper, and revised the paper. ZHY revised the paper. All authors read and approved the final manuscript.

Acknowledgements

We thank anonymous reviewers for very valuable suggestions.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

XC was supported by National Natural Science Foundation of China under Grant No. 61772531.

Publisher's Note

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

Received: 23 June 2018 Accepted: 4 December 2018 Published online: 11 December 2018

References

- Ambros V. The functions of animal microRNAs. Nature. 2004;431:350–5.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116:281–97.
- 3. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. Nature. 2004;431:343–9.
- Ambros V. microRNAs: tiny regulators with great potential. Cell. 2001;107:823–6.
- Kozomara A, Griffithsjones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 2011;39:D152–7.
- Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. Science. 2007;318:1931–4.
- Jopling CL, Yi MK, Lancaster AM, Lemon SM, Sarnow P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. Science. 2005;309:1577–81.
- 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;33:1290–7.
- Karp X, Ambros V. Encountering MicroRNAs in cell fate signaling. Science. 2005;310:1288–9.
- 10. Miska EA. How microRNAs control cell division, differentiation and death. Curr Opin Genet Dev. 2005;15:563.
- 11. Xu P, Guo M, Hay BA. MicroRNAs and the regulation of cell death. Trends Genet. 2005;20:617–24.
- Alshalalfa M, Alhajj R. Using context-specific effect of miRNAs to identify functional associations between miRNAs and gene signatures. BMC Bioinformatics. 2013;14:S1.
- Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136:215–33.
- Cui Q, Yu Z, Purisima EO, Wang E. Principles of microRNA regulation of a human cellular signaling network. Mol Syst Biol. 2014;2:46.
- Alvarez-Garcia I, Miska EA. MicroRNA functions in animal development and human disease. Development. 2005;132:4653.
- Meola N, Gennarino VA, Banfi S. microRNAs and genetic diseases. PathoGenetics. 2009;2:7.
- Lynam-Lennon N, Maher SG, Reynolds JV. The roles of microRNA in cancer and apoptosis. Biol Rev. 2009;84:55–71.
- 18. Esquelakerscher A, Slack FJ. Oncomirs—microRNAs with a role in cancer. Nat Rev Cancer. 2006;6:259.
- 19. Latronico MV, Catalucci D, Condorelli G. Emerging role of microRNAs in cardiovascular biology. Circ Res. 2007;101:1225.

- 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:e3420.
- Chiang K, Liu H, Rice AP. miR-132 enhances HIV-1 replication. Virology. 2013;438:1.
- Mantri CK, Jui PD, Velamarti MJ, Dash CCV. Cocaine enhances HIV-1 replication in CD4+ T cells by down-regulating MiR-125b. PLoS ONE. 2012:7:e51387.
- 23. 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. 2013;33:2589.
- Giricz O, Reynolds PA, Ramnauth A, Liu C, Wang T, Stead L, Childs G, Rohan T, Shapiro N, Fineberg S. Hsa-miR-375 is differentially expressed during breast lobular neoplasia and promotes loss of mammary acinar polarity. J Pathol. 2012;226:108–19.
- Wiemer EAC. The role of microRNAs in cancer: no small matter. Eur J Cancer. 2007;43:1529–44.
- Yang C, Sun C, Liang X, Xie S, Huang J, Li D. Integrative analysis of micro-RNA and mRNA expression profiles in non-small-cell lung cancer. Cancer Gene Ther. 2016;23:90–7.
- Sun CC, Li SJ, Zhang F, Zhang YD, Zuo ZY, Xi YY, Wang L, Li DJ. The novel miR-9600 suppresses tumor progression and promotes paclitaxel sensitivity in non-small-cell lung cancer through altering STAT3 expression. Mol Ther Nucleic Acids. 2016;5:e387.
- Sun CC, Li SJ, Yuan ZP, Li DJ. MicroRNA-346 facilitates cell growth and metastasis, and suppresses cell apoptosis in human non-small cell lung cancer by regulation of XPC/ERK/Snail/E-cadherin pathway. Aging (Albany NY). 2016;8:2509–24.
- Sun C, Li S, Zhang F, Xi Y, Wang L, Bi Y, Li D. Long non-coding RNA NEAT1 promotes non-small cell lung cancer progression through regulation of miR-377-3p-E2F3 pathway. Oncotarget. 2016;7:51784–814.
- 30. Perez-Iratxeta C, Wjst M, Bork P, Andrade MA. G2D: a tool for mining genes associated with disease. BMC Genet. 2005;6:1–9.
- 31. Perez-Iratxeta C, Bork P, Andrade MA. Association of genes to genetically inherited diseases using data mining. Nat Genet. 2002;31:316–9.
- Aerts S, Lambrechts D, Maity S, Loo PV, Coessens B, Smet FD, Tranchevent LC, Moor BD, Marynen P, Hassan B. Gene prioritization through genomic data fusion. Nat Biotechnol. 2006;24:537.
- 33. Chen X, Wang L, Qu J, Guan N-N, Li J-Q. Predicting miRNA-disease association based on inductive matrix completion. Bioinformatics. 2018. https://doi.org/10.1093/bioinformatics/bty503.
- Chen X, Huang L, Xie D, Zhao Q. EGBMMDA: extreme gradient boosting machine for MiRNA-disease association prediction. Cell Death Dis. 2018;9:3.
- Chen X, Huang L. LRSSLMDA: Laplacian regularized sparse subspace learning for MiRNA-disease association prediction. PLoS Comput Biol. 2017;13:e1005912
- Chen X, Zhou Z, Zhao Y. ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction. RNA Biology. 2018;15(6):807– 818. https://doi.org/10.1080/15476286.2018.1460016.
- 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.
- 38. Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, Liu Y, Dai Q, Li J, Teng Z. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. PLoS ONE. 2013;8:e70204.
- Chen X, Wu QF, Yan GY. RKNNMDA: ranking-based KNN for MiRNA-disease association prediction. RNA Biol. 2017. https://doi.org/10.1080/15476 286.2017.1312226.
- 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;31:1805–15.
- 41. Shi H, Xu J, Zhang G, Xu L, Li C, Wang L, Zhao Z, Wei J, 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;7:1–12.
- 42. Mørk S, Pletscher-Frankild S, Palleja CA, Gorodkin J, Jensen LJ. Proteindriven inference of miRNA-disease associations. Bioinformatics. 2014;30:392–7.
- 43. Chen X, Jiang ZC, Xie D, Huang DS, Zhao Q, Yan GY, You ZH. A novel computational model based on super-disease and miRNA for potential miRNA-disease association prediction. Mol BioSyst. 2017;13:1202.

- Xu J, Li CX, Lv JY, Li YS, Xiao Y, Shao TT, Huo X, Li X, Zou Y, Han QL. Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer. Mol Cancer Ther. 2011;10:1857–66.
- 45. Chen X, Yan GY. Semi-supervised learning for potential human microRNA-disease associations inference. Sci Rep. 2014;4:5501.
- 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.
- 47. Pasquier C, Gardès J. Prediction of miRNA-disease associations with a vector space model. Sci Rep. 2016;6:27036.
- 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.
- Chen X, Clarence YC, Zhang X, You ZH, Huang YA, Yan GY. HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. Oncotarget. 2016;7:65257–69.
- Li JQ, Rong ZH, Chen X, Yan GY, You ZH. MCMDA: matrix completion for MiRNA-disease association prediction. Oncotarget. 2017;8:21187–99.
- Yu H, Chen X, Lu L. Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. Sci Rep. 2017;7:43792.
- 52. 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:e1005455.
- Chen X, Xie D, Zhao Q, You ZH: MicroRNAs and complex diseases: from experimental results to computational models. *Brief Bioinform* 2017.
- Chen X, Yan CC, Zhang X, You Z-H. Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief Bioinform. 2017;18:558–76.
- 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;42:1070–4.
- Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, Zhang Q, Yan G, Cui Q. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res. 2013;41:983–6.
- Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNAceRNA, miRNA-ncRNA and protein–RNA interaction networks from largescale CLIP-Seq data. Nucleic Acids Res. 2014;42:D92.
- Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010;26:1644–50.
- 59. Sanghamitra B, Ramkrishna M, Ujjwal M, Zhang MQ. Development of the human cancer microRNA network. Silence. 2010;1:6.
- Goh K, Cusick ME, Valle D, Childs B, Vidal M, Barabási A. The human disease network. Proc Natl Acad Sci USA. 2007;104:8685–90.
- Laarhoven TV, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics. 2011;27:3036.
- Chen X, Yan GY. Novel human IncRNA-disease association inference based on IncRNA expression profiles. Bioinformatics. 2013;29:2617–24.
- Lotfi Shahreza M, Ghadiri N, Mousavi SR, Varshosaz J, Green JR. Heter-LP: a heterogeneous label propagation algorithm and its application in drug repositioning. J Biomed Inform. 2017;68:167–83.
- Zhou T, Ren J, Medo M, Zhang YC. Bipartite network projection and personal recommendation. Phys Rev E Stat Nonlin Soft Matter Phys. 2007;76:046115.
- Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNAdisease associations. Mol BioSyst. 2012;8:2792.
- 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:D98–104.
- Yang Z, Ren F, Liu C, He S, Sun G, Gao Q, Yao L, Zhang Y, Miao R, Cao Y. dbDEMC: a database of differentially expressed miRNAs in human cancers. BMC Genomics. 2010;11(Suppl 4):55.
- 68. He B, Yin B, Wang B, Xia Z, Chen C, Tang J. MicroRNAs in esophageal cancer (review). Mol Med Rep. 2012;6:459.
- Berry MF. Esophageal cancer: staging system and guidelines for staging and treatment. J Thorac Dis. 2014;6(Suppl 3):S289.
- Hummel R, Wang T, Watson DI, Michael MZ, Van der Hoek M, Haier J, Hussey DJ. Chemotherapy-induced modification of microRNA expression in esophageal cancer. Oncol Rep. 2011;26:1011–7.

Chen et al. J Transl Med (2018) 16:348 Page 14 of 14

- Desantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. CA Cancer J Clin. 2017;67:439.
- Kopp F, Oak PS, Wagner E, Roidl A. miR-200c sensitizes breast cancer cells to doxorubicin treatment by decreasing TrkB and Bmi1 expression. PLoS ONE. 2012;7:e50469.
- Tryndyak VP, Beland FA, Pogribny IP. E-cadherin transcriptional downregulation by epigenetic and microRNA-200 family alterations is related to mesenchymal and drug-resistant phenotypes in human breast cancer cells. Int J Cancer. 2010;126:2575–83.
- Gibcus JH, Tan LP, Harms G, Schakel RN, De JD, Blokzijl T, Möller P, Poppema S, Kroesen BJ, Van der Berg A. Hodgkin lymphoma cell lines are characterized by a specific miRNA expression profile. Neoplasia. 2009; 11:167, IN166-176, IN169.
- Xie L, Ushmorov A, Leithäuser F, Guan H, Steidl C, Färbinger J, Pelzer C, Vogel MJ, Maier HJ, Gascoyne RD. FOXO1 is a tumor suppressor in classical Hodgkin lymphoma. Blood. 2012;119:3503.
- Iqbal J, Shen Y, Liu Y, Fu K, Jaffe ES, Liu C, Liu Z, Lachel CM, Deffenbacher K, Greiner TC. Genome-wide miRNA profiling of mantle cell lymphoma reveals a distinct subgroup with poor prognosis. Blood. 2012;119:4939–48
- Watanabe A, Tagawa H, Yamashita J, Teshima K, Nara M, Iwamoto K, Kume M, Kameoka Y, Takahashi N, Nakagawa T, et al. The role of micro-RNA-150 as a tumor suppressor in malignant lymphoma. Leukemia. 2011;25:1324–34.

- 78. Chen S, Wang Z, Dai X, Pan J, Ge J, Han X, Wu Z, Zhou X, Zhao T. Reexpression of microRNA-150 induces EBV-positive Burkitt lymphoma differentiation by modulating c-Myb in vitro. Cancer Sci. 2013;104:826–34.
- Wang E, Zaman N, McGee S, Milanese JS, Masoudi-Nejad A, O'Connor-McCourt M. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. Semin Cancer Biol. 2015;30:4–12.
- Chen X, Xie D, Wang L, Zhao Q, You Z-H, Liu H. BNPMDA: bipartite network projection for MiRNA-disease association prediction. Bioinformatics. 2018. https://doi.org/10.1093/bioinformatics/bty333.
- Chen X, Yin J, Qu J, Huang L. MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. PLoS Comput Biol. 2018;14:e1006418.
- 82. Gao S, Tibiche C, Zou J, Zaman N, Trifiro M, O'Connor-McCourt M, Wang E. Identification and construction of combinatory cancer hallmark-based gene signature sets to predict recurrence and chemotherapy benefit in stage II colorectal cancer. JAMA Oncol. 2016;2:37–45.
- 83. McGee SR, Tibiche C, Trifiro M, Wang E. Network analysis reveals a signaling regulatory loop in the PIK3CA-mutated breast cancer predicting survival outcome. Genomics Proteom Bioinform. 2017;15:121–9.
- 84. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, Zhang Y. Drug-target interaction prediction: databases, web servers and computational models. Brief Bioinform. 2016;17:696.

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

- fast, convenient online submission
- $\bullet\,$ 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

