Journal of Biomedical Informatics 80 (2018) 26–36

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Human disease MiRNA inference by combining target information based on heterogeneous manifolds

Pingjian Ding[a], Jiawei Luo[a,*], Cheng Liang[b], Qiu Xiao[a], Buwen Cao[a]

[a] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410083, China
[b] School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

ARTICLE INFO

Keywords:
miRNA-disease associations
Heterogeneousnetwork
Manifold learning
Graph regularization
Target information

ABSTRACT

The emergence of network medicine has provided great insight into the identification of disease-related molecules, which could help with the development of personalized medicine. However, the state-of-the-art methods could neither simultaneously consider target information and the known miRNA-disease associations nor effectively explore novel gene-disease associations as a by-product during the process of inferring disease-related miRNAs. Computational methods incorporating multiple sources of information offer more opportunities to infer disease-related molecules, including miRNAs and genes in heterogeneous networks at a system level. In this study, we developed a novel algorithm, named inference of Disease-related MiRNAs based on Heterogeneous Manifold (DMHM), to accurately and efficiently identify miRNA-disease associations by integrating multi-omics data. Graph-based regularization was utilized to obtain a smooth function on the data manifold, which constitutes the main principle of DMHM. The novelty of this framework lies in the relatedness between diseases and miRNAs, which are measured via heterogeneous manifolds on heterogeneous networks integrating target information. To demonstrate the effectiveness of DMHM, we conducted comprehensive experiments based on HMDD datasets and compared DMHM with six state-of-the-art methods. Experimental results indicated that DMHM significantly outperformed the other six methods under fivefold cross validation and de novo prediction tests. Case studies have further confirmed the practical usefulness of DMHM.

1. Introduction

The expression of messenger RNAs is suppressed in a sequence-specific manner by microRNAs that consist of small endogenous non-coding RNAs [1,2]. A mounting number of studies have indicated that miRNAs are important components in the cell. Specifically, studies have proved the vital roles of miRNAs in various biological processes, such as cell growth [3], cell development [3], cell cycle regulation [4], cell apoptosis [5], stress responses [6] and tumor invasion [7]. Furthermore, the strong relationships between miRNAs and diseases have been verified by numerous biological observations and studies [8,9]. The accumulating knowledge of disease-related miRNAs could be of great help to achieve pathological classifications, individualized diagnoses and disease treatments [10]. However, it still remains challenging for biologists to discover the underlying miRNA-disease associations extensively. Consequently, there is an urgent need to develop powerful computational approaches that could effectively uncover miRNA-disease associations.

With the emergence of large-scale genomic, transcriptomic and proteomic data, network-based methods have recently become popular tools for predicting disease-related miRNAs [11,12]. Archana et al. [13] simultaneously considered motif, sequential and network features to infer immune specific miRNAs from other non-immune miRNAs. However, as we known, it is currently hard or even impossible to obtain negative immune system diseases-related miRNAs. Xu et al. [14] showed aberrant behavior in prostate cancer by analyzing miRNA-target associations for sequences in the constructed miRNA-target dysregulated network (MTDN). Sun et al. [15] determined miRNA-disease associations by using a two-stage filtration procedure (miRPD). Chen et al. [16] inferred different types of miRNA-disease relationships by using a Restricted Boltzmann machine (RBMMMDA). Chen et al. [17] proposed a semi-supervised method for predicting disease-related miRNAs based on a regularized least squares method (RLSMDA). RLSMDA could prioritize all miRNA-disease associations simultaneously. Xuan et al. [18] utilized weighted k nearest neighbors to prioritize miRNA-disease associations (HDMP). Chen et al. [19] predicted disease-related miRNAs using random walking with restart on the miRNA functional similarity network (RWRMDA). Chen et al. [20]

* Corresponding author.
E-mail address: luojiawei@hnu.edu.cn (J. Luo).

https://doi.org/10.1016/j.jbi.2018.02.013
Received 25 September 2017; Received in revised form 11 February 2018; Accepted 21 February 2018
Available online 24 February 2018
1532-0464/ © 2018 Elsevier Inc. All rights reserved.

uncovered potential miRNA-disease associations based on the within and between scores for many complex diseases (WBSMDA). Chen et al. [21] developed a hybrid graph-based recommendation approach to reveal disease-related miRNAs, which simultaneously considered the network structure and node attribution (HAMDA). Luo et al. [22] adopted a strategy of collective prediction method to predict potential miRNA-disease associations (CPTL). Chen et al. [23] integrated miRNA functional similarity [24], disease semantic similarity and Gaussian interaction profile kernel similarity into a heterogeneous graph to uncover disease-related miRNAs (HGIMDA). Li et al. [25] designed a matrix computational algorithm that could efficiently update the low-rank miRNA-disease matrix to identify the disease-miRNA associations (MCMDA). In particular, network medicine is conducive to completely understand the molecular complexities of diseases [26,27]. Meanwhile, many studies have established relationships between diseases and genes using network-based methods [28–30].

Many network-based methods for uncovering miRNA-disease associations are mainly based on the following two principles [11,12]: (1) miRNAs are related to complex diseases through the target genes being regulated by them [14,31]; (2) target diseases tend to be involved with miRNAs that are related to similar diseases of the target disease [32–34]. Recently, some related papers were published based on the second principle [35–37]. For most methods based on the second principle, miRNA functional similarity calculated by MISIM (miRNA similarity) [24] based on the known miRNA-disease associations was utilized to prioritize disease-related miRNAs. Although many researchers recognize these two principles, there is still a lack of systematic approaches to automatically integrate the target information into the prediction of disease-related miRNAs by taking these two principles together. In addition, inference of disease-related miRNAs and prediction of gene-disease associations have been considered to be two separate tasks in previous studies. However, it would be more reasonable to make direct inference of disease-miRNA associations by incorporating target information.

In this paper, we propose a novel DMHM algorithm to accurately and efficiently infer disease miRNAs based on a heterogeneous manifold [38–40], which integrates miRNA-disease association inferences and disease-related gene predictions into one unified framework. Compared with previous methods based on a homogeneous miRNA network (Fig. 1(A)), homogeneous gene network (Fig. 1(B)), and a bi-layer heterogeneous network (Fig. 1(C)), the heterogeneous miRNA-gene-disease network in this method involves three different types of nodes (diseases, miRNAs and genes), as well as five types of relationships (Fig. 1(D)). Prior knowledge from existing databases [41–44] was used to construct disease-miRNA relationships, disease-gene relationships, miRNA-gene relationships and gene-gene relationships, while the disease semantic similarity was obtained to construct disease-disease relationships [45]. DMHM takes the first principle into account by utilizing miRNA-target interactions and disease-gene associations. Furthermore, the second principle is considered in the proposed algorithm by combining known miRNA-disease associations with disease similarity (Fig. 2(A)). The inference of disease-related miRNAs or genes could be formulated as a binary classification problem; that is, to decide whether a link exists on this heterogeneous network. A semi-supervised prediction algorithm based on a heterogeneous manifold that propagates information across the heterogeneous network is proposed to uncover the missing edges. The guilt-by-association [46] that has been validated to be effective in many previous studies [32,47,48] was also adopted in our method. As a unique characteristic, DMHM automatically incorporates target information into disease-associated miRNA prediction. Although the main goal of DMHM is to discover disease miRNAs, it is worth noting that new disease-gene relationships could be predicted simultaneously in heterogeneous biological networks as well.

As reported in previous papers [22,33], the performance of the second principle-based methods is always better than that of the first principle-based approach. To validate the performance of DMHM, we compared its performance with that of a homogeneous miRNA model (such as: MIDP [33]) (Fig. 1(B)) and bi-layers heterogeneous model (such as: RLSMDA (Regularized Least Squares for MiRNA-Disease Association) [32], MIDPE [33], WBSMDA (Within and Between Score for MiRNA-Disease association) [34], CPTL (Collective Prediction based on Transductive learning) [22] and HDMP (Human Disease-related MiRNA Prediction) [18]) (Fig. 1(C)) consisting of only diseases and miRNAs, which are all based on miRNA functional similarity. Furthermore, these six approaches are all related to the second principle. In the experiment, fivefold cross validation and de novo prediction were implemented to verify the effectiveness of DMHM on large-scale omics data from existing databases. Compared with the state-of-the-art methods, DMHM successfully achieved the best prediction results with the highest area under the receiver operating characteristics (AUC) and precision. By taking two diseases as case studies, we showed several top-ranked miRNA-disease associations that were validated by dbDEMC 2.0 (database of Differentially Expressed MicroRNA in Cancer) [49] and miR2Disease [50].

2. Materials

2.1. Network model for predicting miRNA-disease associations

Most computational methods that have been proposed for prioritizing or predicting disease-related miRNAs are based on network models, including homogeneous network models (Fig. 1(A) and (B)) and bi-layer heterogeneous network models (Fig. 1(C)). However, the goal of the proposed method was to rank the unlabeled nodes and find potential miRNA-disease associations in a complex heterogeneous miRNA-gene-disease network (Fig. 1(D)).

2.1.1. Homogeneous network model for disease miRNA inference

Traditionally, disease-related miRNA inferences based on a homogeneous network model (Fig. 1(A) and (B)), i.e., a model consisting of only one type of node, as well as links, mainly rely on a gene interaction network or miRNA similarity network. Homogeneous network-based methods can be divided mainly into two categories: (1) the miRNA network on which the prediction process is performed (such as MIDP [33] and RWRMDA [19]); and (2) prediction methods using a gene network based on the assumption that miRNAs are related to complex diseases by the target genes regulated by them (such as miR_DG [31] and Shi's method [51]).

2.1.2. Bi-layer heterogeneous network model for disease miRNA prediction

In contrast, a bi-layer heterogeneous network (Fig. 1(C)) for predicting disease-associated miRNAs usually consists of two types of nodes (disease and miRNA) and three types of links (disease-disease links, miRNA-miRNA links and disease-miRNA links). Given a heterogeneous network, $G_{DM} = \{V = \{D,M\}, E = \{R_{DD}, R_{MM}, R_{DM}\}\}$, $D = \{d_1, d_2, \ldots, d_{nD}\}$ and $M = \{m_1, m_2, \ldots, m_{nM}\}$ are the node sets consisting of diseases and miRNAs, respectively, while R_{DD}, R_{MM} and R_{DM} denote links between disease pairs, miRNA pairs and disease-miRNA pairs. These links are constructed by incorporating miRNA-miRNA similarities, disease-disease similarities and known disease-miRNA associations. The inference problem of disease-related miRNAs could then be formulated as uncovering the missing edges in the bi-layer heterogeneous network. Bi-layer heterogeneous networks involving miRNAs and diseases have recently become a popular tool for predicting miRNA-disease associations (such as CPTL [22], MIDPE [33], KRLSM [36], BRWH [37], RLSMDA [32], PBMDA [35] and NetCBI [52]).

2.1.3. Heterogeneous miRNA-gene-disease network model for identifying miRNA-disease associations

However, methods based on bi-layer heterogeneous networks can only consider miRNAs with at least one related disease as a candidate.

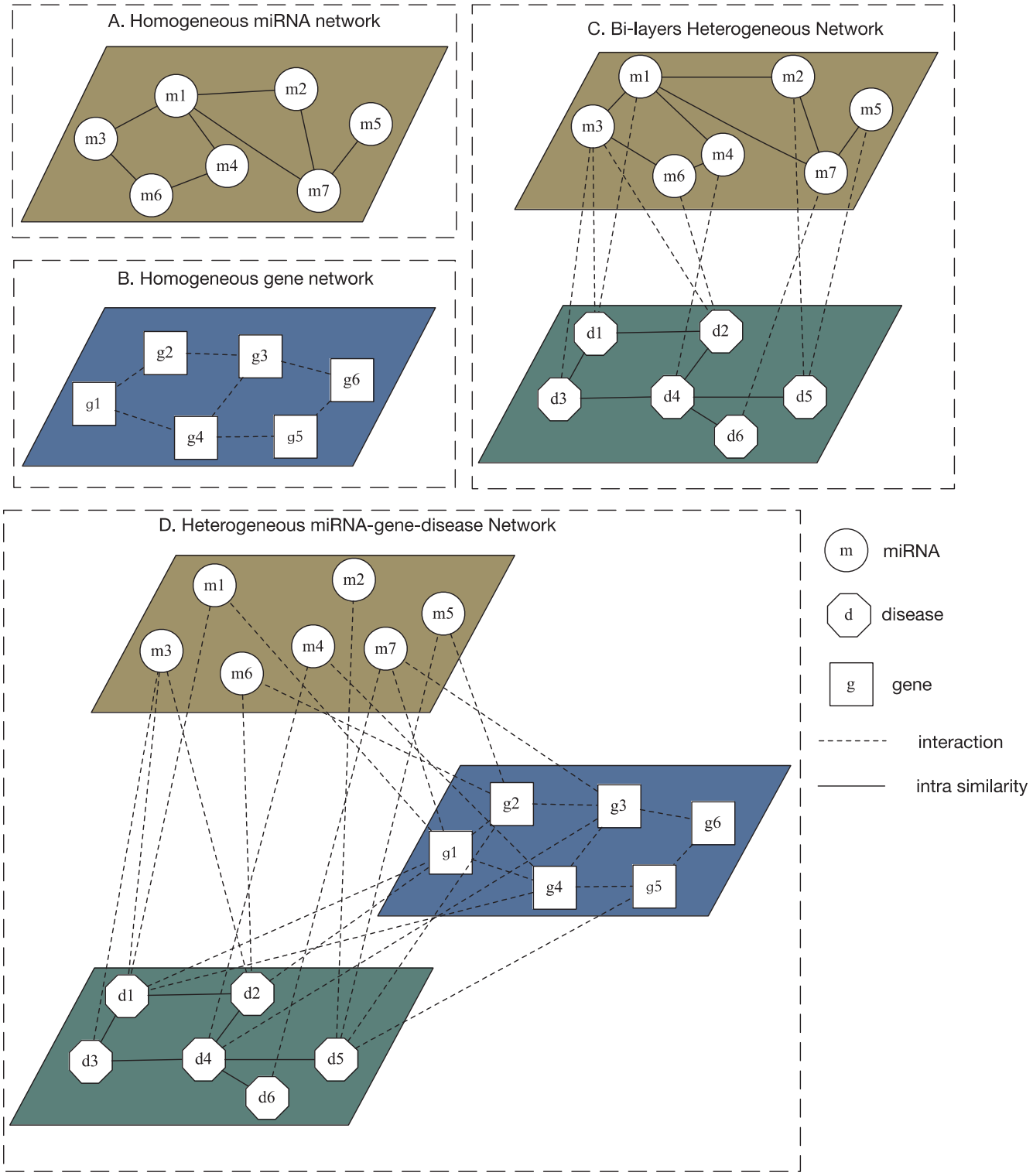


Fig. 1. Homogeneous network and heterogeneous network model; (A) homogeneous miRNA network; (B) homogeneous gene network; (C) bi-layer heterogeneous network; (D) heterogeneous miRNA-gene-disease network.

Therefore, it is required that an effective method incorporating target information considers more miRNAs as disease-related candidates. A more complex heterogeneous miRNA-gene-disease network model (Fig. 1(D)), which consists of three types of nodes (miRNA, gene and disease) and five types of links (miRNA-gene, miRNA-disease, disease-gene, gene-gene, miRNA-miRNA), was introduced. Given a heterogeneous miRNA-gene-disease network $G_{MGD} = \{V = \{M,G,D\}, E = \{R_{DD}, R_{GG}, R_{DG}, R_{DM}, R_{MG}\}\}$, $D = \{d_1,d_2,\ldots,d_{nD}\}$, $M = \{m_1, m_2, \ldots, m_{nM}\}$ and $G = \{g_1, g_2, \ldots, g_{nG}\}$ are the node sets representing diseases, miRNAs and genes, while R_{DD}, R_{GG}, R_{DG}, R_{DM} and R_{MG} denote the sets of links between disease-disease, gene-gene, disease-gene, disease-miRNA and miRNA-gene, respectively. The network schema [53] of heterogeneous miRNA-gene-disease is provided in Fig. 2(B). The goal of DMHM is to establish new links between miRNAs and diseases based on

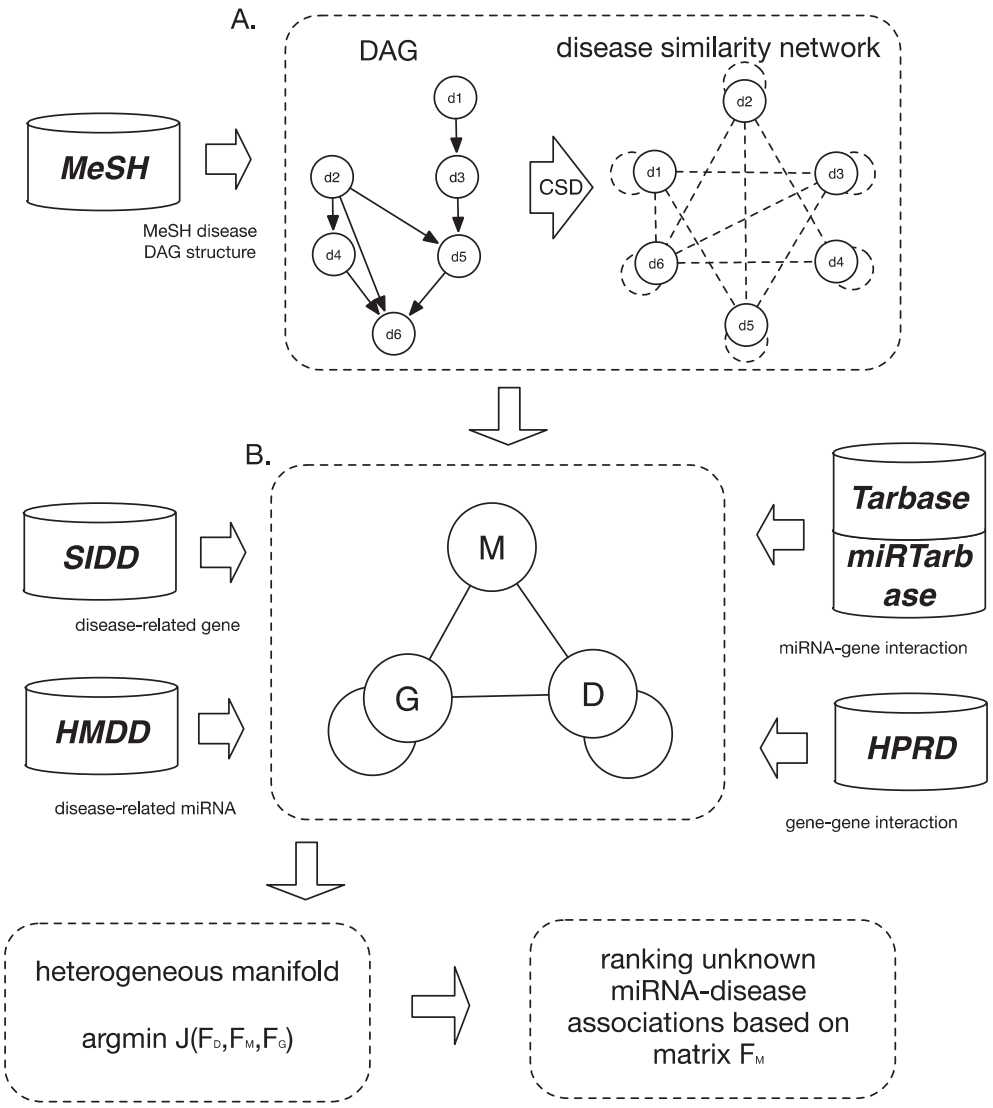


Fig. 2. The overall workflow of DMHM; (A) construction of a disease similarity network; (B) network schema of a heterogeneous miRNA-gene-disease network.

a heterogeneous miRNA-gene-disease network.

2.2. Dataset

Information on diseases, miRNAs and genes are downloaded from different data sources to construct the heterogeneous network. Several existing databases provide known disease-miRNA relationships, disease-gene relationships, miRNA-gene relationships and gene-gene relationships. Furthermore, disease semantic similarity is calculated by employing the existing methods.

2.2.1. Disease similarity measures

Cosine similarity of diseases (CSD) [45] is employed to measure disease semantic similarity based on MeSH disease terms [54]. Based on the description in CSD, the disease semantic similarity can be calculated in two steps. A hierarchical directed acyclic graph (DAG) from the MeSH database (www.ncbi.nlm.nih.gov/) was first used to obtain attribute features of a disease. Let δ denote the shortest distance between the disease d and the *i-th* node that is an ancestor node of d, and $|D|$ denote the number of diseases in the MeSH database, the *i-th* element of V_d representing disease d could be computed as

$$V_d(i) = \begin{cases} 1 & if\ i\text{-}th = d \\ \frac{1}{\delta + 1} & if\ i\text{-}th \in \{the\ ancestor\ nodes\ of\ d\} \\ 0 & otherwise \end{cases}$$

Then, cosine similarity was performed to calculate the semantic similarity of a disease pair d_i-d_j:

$$\mathrm{CSD}(d_i,d_j) = \frac{V_{di} * V_{dj}^T}{\|V_{di}\| \|V_{dj}\|}$$

where V_{di} and V_{dj} are the feature vectors of disease d_i and d_j, respectively. The semantic similarity of two diseases is set to the corresponding value in the disease similarity matrix R_{DD}. A link between two diseases with larger similarity than zero is added to construct the disease network (Fig. 2(A)).

2.2.2. Disease gene associations

We downloaded disease-related genes from the SIDD database [42] (http://mlg.hit.edu.cn/SIDD), which combines five databases of gene-disease associations: GeneRIF [55], OMIM [56], CTD [57], GAD [58] and SpliceDisease [59]. Disease-gene associations are described by the adjacency matrix R_{DG}. The corresponding value in the matrix R_{DG} is set to 1 if the relationship exists and 0 otherwise. We used R_{GD} to denote

the transpose of R_{DG}.

2.2.3. Disease miRNA associations

Initial disease-related microRNA sets are downloaded from the HMDD database [43] (http://cmbi.bjmu.edu.cn/hmdd), which provides miRNA-disease associations validated by biological experiments. Disease-miRNA associations are described by an adjacency matrix R_{DM}. The corresponding value in the matrix R_{DM} is set to 1 if the relationship exists and 0 otherwise. Matrix R_{MD} is the transpose of R_{DM}.

2.2.4. MiRNA target interactions

MiRTarbase is a database of curated miRNA-target interactions (http://mirtarbase.mbc.nctu.edu.tw/), which collects miRNA-target interactions validated by reporter assays and next generation sequencing experiments [41]. DIANA-Tarbase (http://www.microrna.gr/tarbase) [60] is a reference database aiming to catalogue published experimentally supported miRNA-gene interactions. Here, we obtained miRNA-gene interactions from miRTarbase and DIANA-Tarbase v8 to construct the adjacency matrix R_{MG}. The corresponding value in the matrix R_{MG} is set to 1 if the relationship exists and 0 otherwise. Matrix R_{GM} is the transpose of R_{MG}.

2.2.5. Gene interactions

The Human Protein Reference Database (HPRD) [44] has accumulated abundant proteomic information on human proteins, which can be freely accessed from http://www.hprd.org/. Gene-gene interactions are described by an adjacency matrix R_{GG}. The corresponding value in the matrix R_{GG} is set to 1 if the relationship exists and 0 otherwise.

2.2.6. Data preparation

Disease-miRNA relationships, gene-disease relationships and miRNA-target interactions with incorrect disease names, miRNA names or gene names are filtered out. Then, different miRNA transcripts corresponding to the same mature miRNAs are combined [61] as done in previous works [24,36,37]. To reduce the impact of genes with low degrees on the performance of predicting disease-related miRNAs, we only keep those genes with at least one miRNA-gene interaction, one disease-gene association and one gene-gene interaction. Finally, there are 330 diseases, 1726 miRNAs, 6180 genes, 5219 miRNA-disease associations, 19,476 disease-gene associations, 25,031 gene-gene interactions and 242,860 miRNA-gene interactions in the constructed heterogeneous network (supplementary material 1).

3. Methods

3.1. DMHM

The main idea of DMHM is to minimize the cost of the objective function. The rationale behind our work is the achievement of smooth function on the data manifold to predict disease-related miRNAs using graph-based regularization in the heterogeneous miRNA-gene-disease network. In DMHM, disease semantic similarity is first computed based on MeSH disease DAG structure. Then, a heterogeneous disease-miRNA-gene network is constructed by combing disease semantic similarity, disease-gene associations, disease-miRNA associations, gene-gene associations and miRNA-gene associations. Finally, DMHM inferred new disease-miRNA relationships based on heterogeneous manifolds [40]. Likewise, a novel disease-gene relationship between similar diseases and similar genes could be predicted [62]. The overall workflow of DMHM is illustrated in Fig. 2.

Let n_D, n_M and n_G represent the number of diseases, miRNAs and genes, respectively. For each disease, miRNA and gene, we aimed to compute a relatedness matrix $F_i \in R_{iD}$ of size $n_i * n_D$, $i \in \{D,M,G\}$, where each F_M^{pq} measures the confidence that the p-th miRNA relates to the q-th disease. Then, the q-th disease-related miRNAs could be inferred via the value in the q-th column of F_M. Additionally, three matrices: $Y_M = R_{MD}$, $Y_G = R_{GD}$, and $Y_D = R_{DD}$ are designed to utilize a priori knowledge for measuring the relatedness between miRNA-disease pairs, gene-disease pairs as well as disease pairs, respectively. Diagonal matrix D_{ij} of size $n_i * n_i$, $i,j \in \{D,M,G\}$ is constructed, where its (p,p)-th element is the sum of the p-th row of R_{ij} of size $n_i * n_j$, $i,j \in \{D,M,G\}$. Based on the assumption that the predictive results should be consistent with prior knowledge in the heterogeneous biological network, the cost function $J(F_D,F_M,F_G)$ is defined as follows:

$$J(F_D,F_M,F_G) = \sum_{i,j\in\{D,M,G\}} \sum_{p=1}^{n_i} \sum_{q=1}^{n_j} R_{ij}^{pq} \left\| \frac{F_i^p}{\sqrt{D_{ij}^{pp}}} - \frac{F_j^q}{\sqrt{D_{ji}^{qq}}} \right\|^2 + \sum_{i\in\{D,M,G\}} \|F_i - Y_i\|^2$$

where D_{ij}^{pp} denotes the (p,p)-element of D_{ij}, and D_{ji}^{qq} denotes the (q,q)-element of D_{ji}. The first term in the cost function, smooths the information in a heterogeneous network, which formulates that two miRNAs (diseases, genes) associating with the same disease should be similar. The impact of the popularity of nodes is reduced by employing a Laplacian matrix, which dominates the relatedness estimations. Furthermore, the difference between the prior knowledge and the inference results is minimized by the second term.

The normalized form of each adjacency matrix for describing an algorithm is expediently defined as follows:

$$S_{ij} = D_{ij}^{-(1/2)} R_{ij} D_{ji}^{-(1/2)} i,j \in \{D,M,G\}$$

After some algebraic transformations (Appendix A), we could obtain the closed form solution of the cost function as follows:

$$F_i = (5I_i - 2S_{ii})^{-1} * \left(\sum_{j\in\{D,M,G\}, j\neq i} S_{ij} F_j + Y_i \right), i \in \{D,M,G\}$$

where matrix $5I_i - 2S_{ii}, i \in \{D,M,G\}$ is easily proven to be a positive definite and invertible (Appendix B). Furthermore, because there are no links between miRNA pairs, i.e., matrix S_{MM} is a zero matrix, the matrix F_M could be described as follows:

$$F_M = 0.2*(S_{MD}F_D + S_{MG}F_G + Y_M)$$

The process of disease-related miRNAs inference based on a heterogeneous manifold, DMHM, is summarized in Algorithm 1.

Algorithm 1 (*DMHM*).

Input: miRNA-target interactions; known disease-related miRNAs; disease-gene associations; disease-disease similarities, gene-gene interactions

Preprocessing phase:

1. construct the adjacency matrix R_{ij}, i,j ∈ {D,M,G}, where R_{MM} is the zero matrix;
2. normalize R_{ij} by $S_{ij} = D_{ij}^{-(1/2)} R_{ij} D_{ji}^{-(1/2)} i,j \in \{D,M,G\}$;

Prediction phase:

3. set Y_i by $Y_i = R_{iD}$, i ∈ {D,M,G};
4. solve the minimization problem
   $F_M,F_G,F_D = argmin_{F_M,F_G,F_D} J(F_M,F_G,F_D)$ using closed form:

For miRNA:

$F_M(t+1) = 0.2*(S_{MD}F_D(t) + S_{MG}F_G(t) + Y_M)$;

For gene:

$F_G(t+1) = (5I_G - 2S_{GG})^{-1} * (S_{GM}F_M(t) + S_{GD}F_D(t) + Y_G)$;

For disease:

$F_D(t+1) = (5I_D - 2S_{DD})^{-1} * (S_{DM}F_M(t) + S_{DG}F_G(t) + Y_D)$;

when F_M converges // $F_M(t+1) - F_M(t) \leqslant threshhold$.

5. all the unlabeled miRNA-disease interactions are ranked based on matrix F_M. miRNA-disease associations with high rankings are potential disease-related miRNAs.

Table 1
Comparison of various computational approaches under the fivefold cross validation scheme.

Disease	#miRNA	DMHM	MIDPE	MIDP	WBSMDA	RLSMDA	CPTL	HDMP
Breast neoplasms	202	**0.9317**	0.8138	0.8061	0.7543	0.8017	0.8039	0.7956
Colorectal neoplasms	147	**0.9533**	0.8059	0.7994	0.7048	0.797	0.7955	0.812
Glioblastoma	96	**0.9467**	0.8001	0.7901	0.7721	0.7675	0.8082	0.8276
Heart failure	120	**0.951**	0.7979	0.7814	0.717	0.7625	0.7923	0.7864
Hepatocellular carcinoma	214	**0.9328**	0.7706	0.7491	0.7132	0.7423	0.7633	0.7549
Lung neoplasms	132	**0.9759**	0.9017	0.892	0.8001	0.8696	0.8916	0.9064
Melanoma	141	**0.9361**	0.8204	0.8115	0.7594	0.8008	0.8219	0.8232
Non-small-cell lung carcinoma	95	**0.9574**	0.8589	0.8501	0.7895	0.8339	0.8506	0.8441
Ovarian neoplasms	114	**0.9745**	0.8845	0.8931	0.8174	0.8841	0.8926	0.8913
Pancreatic neoplasms	99	**0.9754**	0.9013	0.8823	0.8066	0.8585	0.8953	0.8913
Prostatic neoplasms	118	**0.9508**	0.8376	0.8298	0.8271	0.8052	0.8301	0.8519
Renal cell carcinoma	107	**0.932**	0.8139	0.8035	0.7658	0.7844	0.8029	0.8103
Stomach neoplasms	173	**0.9326**	0.7863	0.7793	0.7431	0.773	0.7629	0.7802
Urinary bladder neoplasms	92	**0.9524**	0.8442	0.8274	0.8148	0.785	0.8359	0.8712
Average AUC		**0.9502**	0.8312	0.8211	0.7704	0.8047	0.8248	0.8319

The bold value represents the greatest value in the corresponding row.

3.2. Evaluation metrics

To systematically evaluate the prediction performance of each method, two metrics including precision and AUC (the area under the ROC curve) are used [63]. Precision indicates how excellent results are generated for miRNA-disease associations with top ranks. The AUC is utilized to assess the overall performance.

To evaluate whether DMHM outperforms other methods for inferring miRNA-disease associations with high ranks, precision is adopted and defined as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

where *TP* and *FP* represent true positive and false positive for the given disease, respectively.

Furthermore, Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) is implemented to assess the global prediction performance. The true positive rate (TPR) and the false positive rate (FPR) with varying cut-off values for plotting ROC curves could be calculated as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \ \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

where *TN* and *FN* denote the number of negative and positive samples with lower rank than a given cutoff for the given disease, respectively.

4. Experiments and results

In this section, we first compare DMHM with six state-of-the-art methods on a gold standard dataset using these evaluation metrics. Then, we confirm the ability of DMHM to uncover disease-related miRNAs via case studies. Finally, we demonstrate the performance of DMHM for predicting gene-disease associations.

4.1. Performance comparison with other methods

The performance of DMHM compared with that of six state-of-the-art computational prediction methods (i.e., RLSMDA [32], MIDP [33], MIDPE [33], WBSMDA [34], CPTL [22] and HDMP [18]) was analyzed. Specifically, RLSMDA utilized a simple weighted average operation to combine the results generated from two semi-supervised classifiers implementing in disease space and miRNA space, respectively. Both MIDP and MIDPE adopt a random walk strategy, with the only difference being the network models on which they could apply. WBSMDA inferred potential disease-related miRNAs using the Gaussian interaction profile kernel similarity network. CPTL adopted transductive learning as a local prediction model to collectively predict disease-related miRNAs. HDMP prioritized miRNA-disease associations based on weighted k as the most similar neighbors, which combined information of miRNA families and clusters. The HMDD dataset was utilized as the experimental dataset. We implemented all methods in C++ with Xcode, and employed a C++ template library eigen3 for matrix manipulation (downloaded from http://eigen.tuxfamily.org/). Furthermore, the source code of DMHM is available at https://github.com/KDDing/DMHM.

4.1.1. Five-fold cross validation

Fivefold cross validation was implemented to systematically evaluate the prediction performance of each method. In fivefold cross validation, we randomly divided the associations of each disease into five subsets of equal size. Each subset was used as a testing test in turn. As mentioned above, we compared DMHM with six state-of-the-art methods, i.e., RLSMDA [32], MIDP [33], MIDPE [33], WBSMDA [34], CPTL [22] and HMDP [18], to evaluate the prediction performances of our method. For each method, we set optimal values for the parameters as suggested by the authors. In particularly, we set $nm = 1$, $nd = 1$ and $w = 0.9$ for RLSMDA, $r_Q = 0.4$ and $r_U = 0.1$ for MIDP, $a = 0.9$ and $r = 0.8$ for MIDPE, $\lambda_{MM} = 0.2$, $\lambda_{MD} = 0.4$, $\lambda_{DD} = 0.2$, $\lambda_{DM} = 0.4$, and $iter = 25$ for CPTL, $k = 10$ for HDMP. To make fair comparisons, we implemented fivefold cross validation for 14 diseases with more than 80 confirmed associations in the same version of the HMDD database, since diseases with few miRNAs could not be sufficiently used to evaluate the prediction performance. As shown in Table 1, the AUCs averaged over 50 runs of DMHM, MIDPE, MIDP, WBSMDA, RLSMDA, CPTL and HDMP were 0.9502, 0.8312, 0.8211, 0.7704, 0.8047, 0.8248 and 0.8319, respectively. DMHM achieved the highest averaged AUC value, which was 0.119, 0.1291, 0.1798, 0.1455, 0.1254 and 0.1183 higher than the other six prediction methods, respectively.

The number of miRNA candidates of DMHM is different from that of the other methods, since MIDP, MIDPE, WBSMDA, RLSMDA, CPTL and HDMP could not address isolated miRNAs without related diseases. Moreover, an equal number of positive samples is generated from HMDD for different predictive methods. As researchers are generally more interested in the predicted top-ranked miRNAs, we further used precision to measure the predicted top-ranked associations for these 14 diseases [22]. The average precision, from top 10 to top 50, are shown in Fig. 3. We could see that DMHM also outperformed other methods in terms of average precision, especially in the range from top-10 to top-30. Taken together, it could be seen that DMHM could not only consider more miRNAs as candidates but also achieve better performance in terms of AUC and precision of the top-ranked associations.

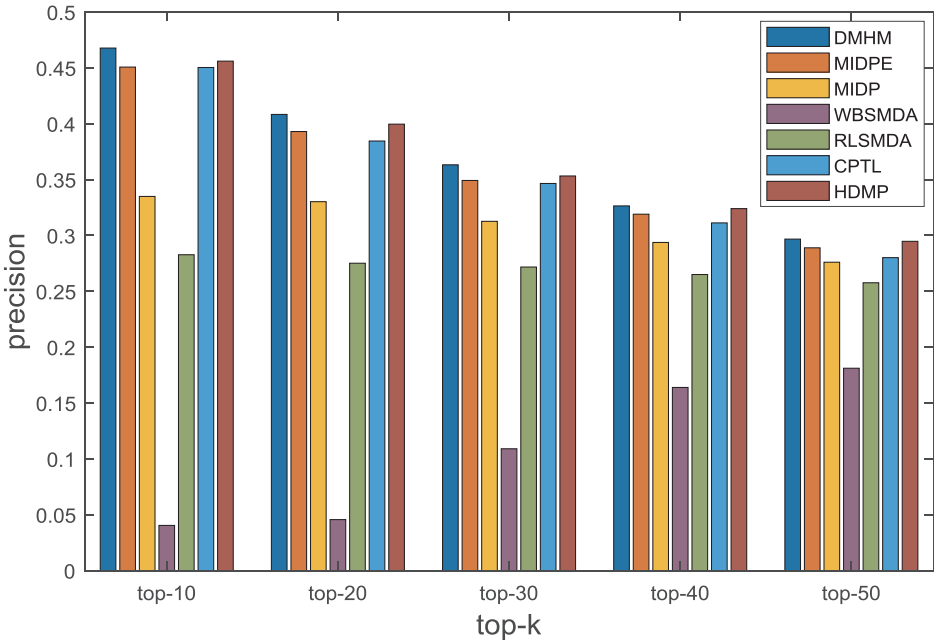


Fig. 3. The comparison between DMHM and other methods in terms of precision.

4.1.2. De novo disease-miRNA prediction test

A de novo prediction test was employed to illustrate the capability of DMHM to uncover new disease-related miRNAs. To perform the test, all the known disease *d*-related miRNA associations were removed, and these associations were considered as positive samples in the testing dataset. Since MIDP, WBSMDA, CPTL and HDMP could not address diseases without known related miRNAs, only the comparisons among DMHM, MIDPE and RLSMDA were implemented. In addition, because miRNAs are closely related to the digestive system and urinary system [64,65], two common important diseases (stomach neoplasms and prostatic neoplasms) belonging to the digestive system and urinary system diseases were taken as instances to evaluate the ability to uncover new disease-related miRNAs. The ROCs of stomach neoplasms and prostatic neoplasms are illustrated in Fig. 4 under a de novo prediction test.

Here, we removed 173 and 118 known miRNA-disease associations for stomach neoplasms and prostatic neoplasms, respectively. As shown in Fig. 4, the AUC values of stomach neoplasms corresponding to the three methods (DMHM, RLSMDA and MIDPE) were 0.9229, 0.7841 and 0.775, respectively. For prostatic neoplasms, the AUCs were 0.9468, 0.8487 and 0.8236, respectively. DMHM achieved significantly better performance than RLSMDA and MIDPE. This might be attributed to the fact that the miRNA similarities calculated from disease-related miRNA is greatly influenced by the removed miRNA-disease associations. Therefore, the bi-layer heterogeneous network was constructed with known miRNA-disease associations, disease similarity and miRNA similarity, which may not be a good way to predict new disease-associated miRNAs. Furthermore, we conducted the top-ranked miRNAs for stomach neoplasms and prostatic neoplasms under the de novo prediction test (Fig. 5). As shown in Fig. 5, DMHM still performed slightly better than other two approaches in terms of precision.

4.2. Case studies

It was reported in many articles that miRNAs have a relationship with the digestive system and urinary system [64,65]. The roles that miRNAs play in the digestive system and urinary system diseases could be explored by disease-miRNA relationships. Therefore, we conducted two case studies to validate the prediction capability of DMHM. Esophageal neoplasms were selected as typical instances of digestive system diseases. Meanwhile, we selected kidney neoplasms as the representative instance of urinary system diseases. We then used dbDEMC 2.0 [49] and miR2Disease[50] as the benchmark databases to verify the predicted associations based on the HMDD dataset. Therefore, the

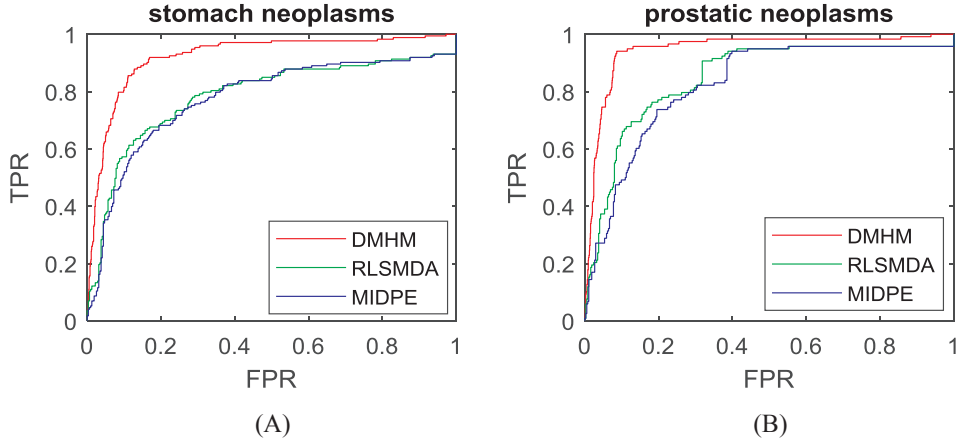


Fig. 4. The ROC curves of various methods for two diseases without related miRNAs.

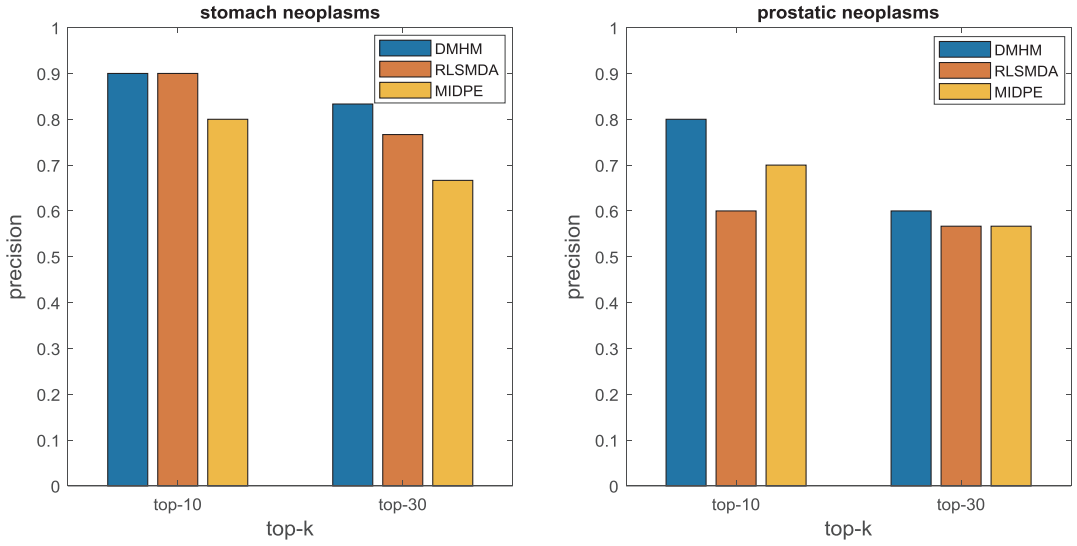


Fig. 5. The precisions for stomach neoplasms and prostatic neoplasms under de novo prediction test.

number of known miRNA-disease associations that are not included in HMDD are used to estimate the performance of DMHM.

An esophageal neoplasm with a poor prognosis is a type of digestive carcinoma. There are many symptoms with esophageal neoplasms, such as difficult swallowing, weight loss and so on. Winther et al. [66] suggested that mir-21 and mir-375 could be acted as prognostic biomarkers for esophageal cancer. We carried out DMHM to yield esophageal neoplasms-related miRNAs (Supplementary Table 1). The prediction results demonstrated that 9 of the top-10 and 45 of the top-50 predicted miRNAs were confirmed to be related to esophageal neoplasms by the two independent benchmark databases (dbDEMC 2.0 and miR2Disease). Furthermore, the detailed information of the top-five esophageal cancer-related miRNAs were described to discuss the underlying biological mechanisms. Zhu et al. [67] demonstrated that aberrant expression level of mir-16 could suppress cell apoptosis as well as regulate RECK and SOX6 to promote growth, which is important to the pathogenesis of esophageal squamous cell carcinoma. Zhang et al. [68] described that target of the anticancer activity of tanshinone IIA is metabolic regulation of mir-122/PKM2 in human esophageal cancer cells. Liu et al. [69] concluded that growth of esophageal carcinoma cell could be repressed by exogenous expression of mir-29a with down-regulating Nfia and activating the Notch signaling pathway. mir-15b could act as the prognostic marker of esophageal cancer [70]. Jing et al. [71] indicated that migration and invasion could be impeded by mir-17 with TGF-β/ITGB6 pathway for esophageal squamous cell carcinoma.

Kidney neoplasms are rapidly increasing cancer types. Many studies have proved that miRNAs have prognostic and predictive relevance in urological carcinomas. MicroRNA biomarkers are a promising tool in the early detection and diagnosis of kidney cancer [72]. We implemented DMHM to identify kidney neoplasm-associated miRNAs (Supplementary Table 2). Similarly, 7 of the top-10 and 41 of the top-50 predicted miRNAs have been confirmed to be related to kidney neoplasms with the two independent benchmark databases (dbDEMC 2.0 and miR2Disease). Furthermore, the detailed information of the top-five kidney cancer-related miRNAs were described to discuss the underlying biological mechanisms. Alvarez et al. [73] demonstrated that glucose and TGF- β1 could upregulated miR-1207-5p that is a PVT1-derived miRNA and abundantly expressed in kidney cells. PVT1, a long non-coding RNA, has been related with kidney cancer. Li et al. [74] concluded that mir-155 is targeted at BACH1 to play an important role in kidney cancer. Yu et al. [75] reported that CD44 in kidney cancer is the target of mir-34a which could repress the growth, metastasis and tube formation of kidney cancer cell. Wang et al. [76] concluded that oncogenic SPAG9 in 786-O and ACHN cells could be directly regulated by miR-200a-3p. The cell cycle and the growth of the renal cancer cell lines could significantly decrease due to the silencing of SPAG9. Zununi et al. [77] described that kidney allograft dysfunction and IFTA are associated with the expression levels of the aberrant urinary mir-21 and mir-200b, which could be acted as diagnostic biomarkers for kidney allograft monitoring. Moreover, different expression of mir-200b is in clear cell renal cell cancer compared to normal counterparts [78]. Moreover, the top-50 predicted miRNAs related to different diseases with more than 80 confirmed miRNAs are contained in the supplementary material 2.

4.3. Disease-gene associations prediction

DMHM could also explore novel disease-gene relationships as a by-product. MiRNA information could be considered as an additional source to uncover disease-related genes compared with network-based disease-related genes inference methods (such as: RWRHN [62], EPU [79] and PUDI [80]). Therefore, fivefold cross validation was implemented to demonstrate the capability of DMHM for predicting disease-related genes (Fig. 6). Thirty-one diseases with more than 200 related genes were taken as target diseases. As shown in Fig. 6, DMHM achieved an average AUC of 0.7866 for disease-related gene identification. Furthermore, all gene-disease associations from SIDD were utilized as a training set. The top-50 predictive genes that were related to 31 diseases with more than 200 confirmed genes were prioritized in supplementary material 3.

5. Discussion

Based on the principle that phenotypically similar diseases tend to associate with functionally similar miRNAs and vice versa, DMHM integrates target information to predict miRNA-disease associations using heterogeneous manifolds in a heterogeneous miRNA-gene-disease network.

DMHM has the following advantages. First, DMHM considers disease-related gene identification into miRNA-disease association predictions based on a single framework. Furthermore, for some existing methods (such as RLSMDA [32], MIDP [33], MIDPE [33], WBSMDA [34] and PBMDA [35]), only miRNAs from the HMDD database are considered as candidates of given diseases. However, these miRNAs are

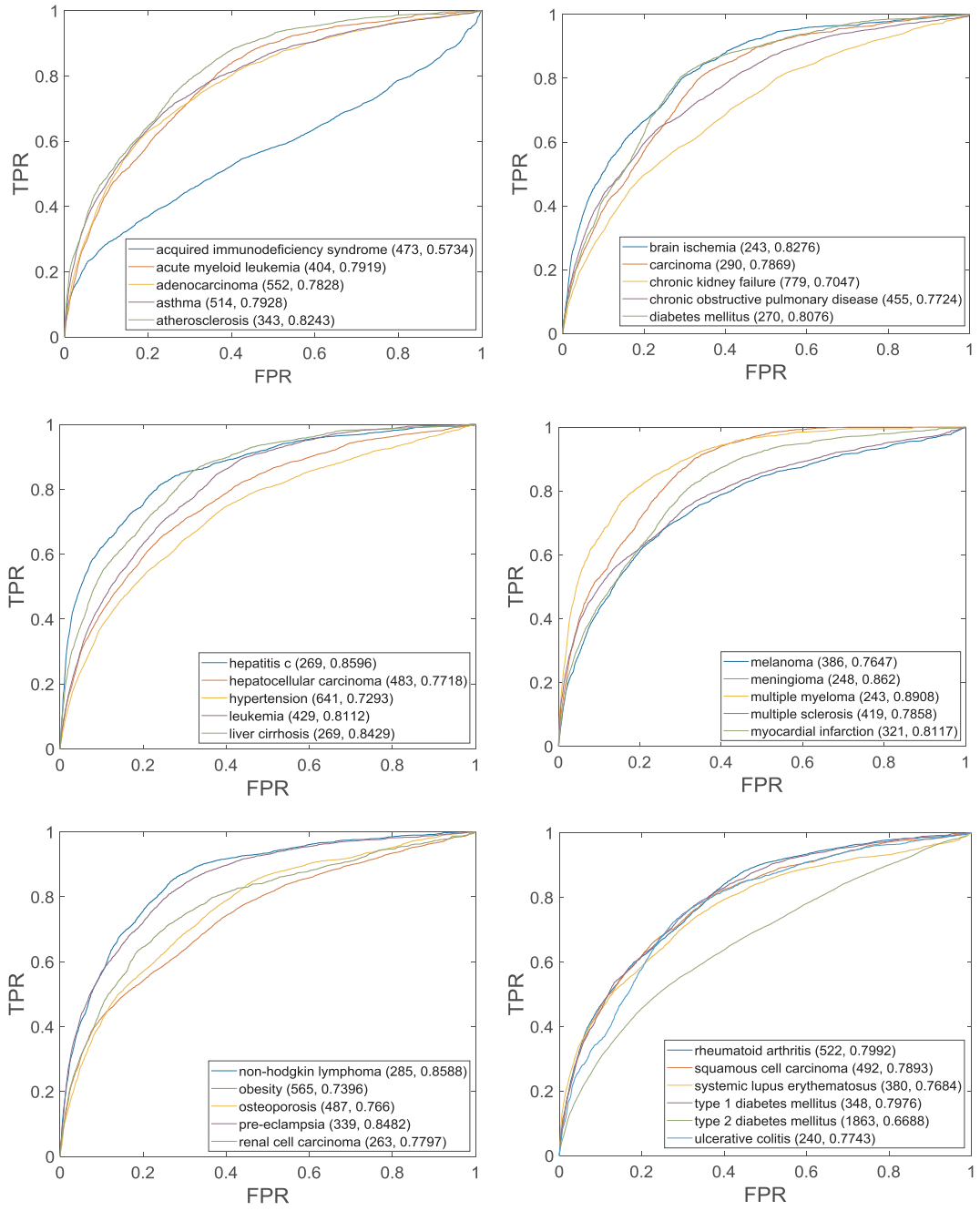


Fig. 6. The ROC curves of DMHM for predicting disease genes under the fivefold cross validation scheme.

only a small fraction of the discovered miRNAs. Second, experimental results showed that the performances of some methods are highly influenced by the values selected for the parameters (such as: RLSMDA [32]). In contrast, DMHM does not take any parameters in the prediction process and is thus more robust. Third, to dramatically improve the performance of prediction methods in the bi-layer heterogeneous network, DMHM combines different types of related information in the reliable heterogeneous miRNA-gene-disease network. Fourth, DMHM is a semi-supervised learning method for uncovering disease-related miRNAs, which could address difficulties in achieving negative samples of miRNA-disease associations.

Although DMHM could achieve favorable performance in a heterogeneous miRNA-gene-disease network, there is still room for improvement of the proposed method. A reasonable attempt would be integrating more diverse sources of information. Furthermore, weights for gene-gene interactions could be defined based on the functional similarity using GO terms, which may eliminate the noisy data in our framework and further improve the prediction accuracy. Meanwhile, node attribution may be more conducive for improving the performance of inferring miRNA-disease associations.

6. Conclusions

In this article, a heterogeneous manifold-based approach DMHM, which integrates target information in a heterogeneous miRNA-gene-disease network, has been proposed for miRNA-disease association predictions. DMHM iteratively obtained confidence between the disease and miRNA to predict potential miRNAs for each disease. Compared with six state-of-the-art methods in predicting miRNA-disease associations, DMHM achieved better performances under a fivefold cross validation scheme and de novo prediction test in terms of AUC and precision. Case studies on esophageal neoplasms and kidney neoplasms

further illustrated that DMHM could uncover new miRNA-disease associations supported by dbDEMC 2.0 and miR2Disease. Therefore, DMHM makes sense for further study to discover disease-related miRNAs, which benefits diagnosis, treatment, prevention and prognosis of human disease. Furthermore, microRNAs always cooperatively regulate gene expression. In the future, we plan to discover co-functional microRNA pairs based on a heterogeneous network.

Conflict of interest

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

Acknowledgment

This work has been supported by the National Natural Science Foundation of China (Grant No. 61572180, 61602283), Shandong Provincial Natural Science Foundation (Grant No. ZR2016FB10), the Key Foundation of the Education Department of Hunan Province (Grant No. 17A037) and Hunan Provincial Innovation Foundation for Postgraduate (Grant No. CX2017B102). The author PingjianDing is thankful for the funding from Chinese Scholarship Council (CSC).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2018.02.013.

References

[1] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2) (2004) 281–297.
[2] V. Ambros, microRNAs: tiny regulators with great potential, Cell 107 (7) (2001) 823–826.
[3] V. Ambros, MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing, Cell 113 (6) (2003) 673–676.
[4] M. Carleton, M.A. Cleary, P.S. Linsley, MicroRNAs and cell cycle regulation, Cell Cycle 6 (17) (2007) 2127–2132.
[5] F. Petrocca, et al., E2F1-regulated microRNAs impair TGFβ-dependent cell-cycle arrest and apoptosis in gastric cancer, Cancer Cell 13 (3) (2008) 272–286.
[6] A.K. Leung, P.A. Sharp, MicroRNA functions in stress responses, Molecular Cell 40 (2) (2010) 205–215.
[7] L. Ma, J. Teruya-Feldstein, R.A. Weinberg, Tumour invasion and metastasis initiated by microRNA-10b in breast cancer, Nature 449 (7163) (2007) 682–688.
[8] M. Esteller, Non-coding RNAs in human disease, Nat. Rev. Genet. 12 (12) (2011) 861.
[9] S.M. Hammond, MicroRNAs as tumor suppressors, Nat. Genetics 39 (5) (2005) 582–583.
[10] J. Weidhaas, Using microRNAs to understand cancer biology, Lancet Oncol. 11 (2) (2010) 106–107.
[11] X. Zeng, X. Zhang, Q. Zou, Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks, Briefings Bioinform. 17 (2) (2016) 193–203.
[12] Q. Zou, et al., Similarity computation strategies in the microRNA-disease network: a survey, Briefings Funct. Genom. (2015) elv024.
[13] A. Prabahar, J. Natarajan, Prediction of microRNAs involved in immune system diseases through network based features, J. Biomed. Inform. 65 (2017) 34–45.
[14] J. Xu, et al., Prioritizing candidate disease miRNAs by topological features in the miRNA target–dysregulated network: Case study of prostate cancer, Mol. Cancer Therap. 10 (10) (2011) 1857–1866.
[15] X. Sun, et al., mirPD: a pattern-based approach for identifying microRNAs from deep sequencing data, Digital Signal Process. 23 (6) (2013) 1887–1896.
[16] X. Chen, et al., RBMMMDA: predicting multiple types of disease-microRNA associations, Sci. Reports 5 (2015).
[17] X. Chen, G.-Y. Yan, Semi-supervised learning for potential human microRNA-disease associations inference, Sci. Reports 4 (2014).
[18] P. Xuan, et al., Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors, PloS One 8 (8) (2013) e70204.
[19] X. Chen, M.-X. Liu, G.-Y. Yan, RWRMDA: predicting novel human microRNA–disease associations, Mol. BioSystems 8 (10) (2012) 2792–2798.
[20] X. Chen, et al., WBSMDA: within and between score for MiRNA-disease association prediction, Sci. Reports 6 (2016).
[21] X. Chen, et al., HAMDA: hybrid approach for MiRNA-disease association prediction, J. Biomed. Inform. 76 (2017) 50–58.
[22] J. Luo, et al., Collective prediction of disease-associated miRNAs based on transduction learning, IEEE/ACM Trans. Comput. Biol. Bioinform. (2016).
[23] X. Chen, et al., HGIMDA: heterogeneous graph inference for miRNA-disease association prediction, Oncotarget 7 (40) (2016) 65257–65269.
[24] D. Wang, et al., Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases, Bioinformatics 26 (13) (2010) 1644–1650.
[25] J.-Q. Li, et al., MCMDA: Matrix completion for MiRNA-disease association prediction, Oncotarget 8 (13) (2017) 21187.
[26] A.-L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nat. Rev. Genetics 12 (1) (2011) 56–68.
[27] K.-I. Goh, I.-G. Choi, Exploring the human diseasome: the human disease network, Briefings Funct. Genom. 11 (6) (2012) 533–542.
[28] X. Wang, N. Gulbahce, H. Yu, Network-based methods for human disease gene prediction, Briefings Funct. Genom. 10 (5) (2011) 280–293.
[29] X. Zhou, et al., Human symptoms–disease network, Nat. Commun. 5 (2014).
[30] N.T. Doncheva, T. Kacprowski, M. Albrecht, Recent approaches to the prioritization of candidate disease genes, Wiley Interdisciplinary Rev.: Syst. Biol. Med. 4 (5) (2012) 429–442.
[31] X.-M. Zhao, et al., Identifying cancer-related microRNAs based on gene expression data, Bioinformatics 31 (8) (2015) 1226–1234.
[32] X. Chen, G.-Y. Yan, Semi-supervised learning for potential human microRNA-disease associations inference, Sci. Reports 4 (2014) 5501.
[33] P. Xuan, et al., Prediction of potential disease-associated microRNAs based on random walk, Bioinformatics (2015) btv039.
[34] X. Chen, et al., WBSMDA: Within and Between Score for MiRNA-Disease Association prediction, Sci. Reports 6 (2016) 21106.
[35] Z.-H. You, et al., PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction, PLoS Comput. Biol. 13 (3) (2017) e1005455.
[36] J. Luo, et al., Predicting microRNA-disease associations using Kronecker Regularized Least Squares based on heterogeneous omics data, IEEE Access 5 (2017) 2503–2513.
[37] J. Luo, Q. Xiao, A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network, J. Biomed. Inform. 66 (2017) 194–203.
[38] H. Huang, et al., Locally learning heterogeneous manifolds for phonetic classification, Comput. Speech Language 38 (2016) 28–45.
[39] J. Hou, R. Nayak, Robust clustering of multi-type relational data via a heterogeneous manifold ensemble, 2015 IEEE 31st International Conference on Data Engineering (ICDE), IEEE, 2015.
[40] W. Zhao, Z. Guan, Z. Liu, Ranking on heterogeneous manifolds for tag recommendation in social tagging services, Neurocomputing 148 (2015) 521–534.
[41] C.-H. Chou, et al., miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database, Nucl. Acids Res. 44 (D1) (2015) D239–D247.
[42] L. Cheng, et al., SIDD: a semantically integrated database towards a global view of human disease, PloS One 8 (10) (2013) e75504.
[43] Y. Li, et al., HMDD v2. 0: a database for experimentally supported human microRNA and disease associations, Nucl. Acids Res. (2013) gkt1023.
[44] T.K. Prasad, et al., Human protein reference database—2009 update, Nucl. Acids Res. 37 (Suppl 1) (2009) D767–D772.
[45] P. Ding, et al., A path-based measurement for human miRNA functional similarities using miRNA-disease associations, Sci. Reports 6 (2016) 32533.
[46] D. Altshuler, M. Daly, L. Kruglyak, Guilt by association, Nat. Genetics 26 (2) (2000) 135–137.
[47] A. Ezzat, et al., Drug-target interaction prediction with graph regularized matrix factorization, IEEE/ACM Trans. Comput. Biol. Bioinform. (2016).
[48] W. Wang, et al., Drug repositioning by integrating target information through a heterogeneous network model, Bioinformatics 30 (20) (2014) 2923–2930.
[49] Z. Yang, et al., dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers, Nucl. Acids Res. 45 (D1) (2017) D812–D818.
[50] Q. Jiang, et al., miR2Disease: a manually curated database for microRNA deregulation in human disease, Nucl. Acids Res. 37 (1) (2009) D98–104.
[51] H. Shi, et al., Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes, BMC Syst. Biol. 7 (1) (2013) 101.
[52] H. Chen, Z. Zhang, Similarity-based methods for potential human microRNA-disease association prediction, BMC Med. Genom. 6 (1) (2013) 12.
[53] C. Shi, et al., A survey of heterogeneous information network analysis, IEEE Trans. Knowledge Data Eng. 29 (1) (2017) 17–37.
[54] C.E. Lipscomb, Medical Subject Headings (MeSH), Bull. Med. Library Assoc. 88 (3) (2000) 265.
[55] J.D. Osborne, et al., GeneRIF is a More Comprehensive, Current and Computationally Tractable Source of Gene-Disease Relationships than OMIM, Oxford University, 2006.
[56] J.S. Amberger, et al., OMIM. org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders, Nucl. Acids Res. 43 (D1) (2014) D789–D798.
[57] A.P. Davis, et al., Comparative Toxicogenomics Database: a knowledgebase and discovery tool for chemical-gene-disease networks, Nucl. Acids Res. 37 (1) (2009) D786–92.
[58] K.G. Becker, et al., The genetic association database, Nat. Genetics 36 (5) (2004) 431–432.
[59] J. Wang, et al., SpliceDisease database: linking RNA splicing and disease, Nucl. Acids Res. 40 (Database issue) (2012) D1055–9.
[60] D. Karagkouni, et al., DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions, Nucl. Acids Res. (2017).
[61] Y. Liu, et al., Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources, IEEE/ACM Trans. Comput. Biol. Bioinform. (2016) p. 1-1.
[62] J. Luo, S. Liang, Prioritization of potential candidate disease genes by topological

similarity of protein-protein interaction network and phenotype data, J. Biomed. Inform. 53 (2015) 229.
[63] D.M.W. Powers, Evaluation: from precision, recall and F-Factor to ROC, informedness, markedness & correlation, J. Mach. Learn. Technol. 2 (2011) 2229–3981.
[64] P. Correa, The epidemiology of gastric cancer, Gastric Cancer 12 (1) (2002) 111–127.
[65] E. Wienholds, et al., MicroRNA expression in zebrafish embryonic development, Science 309 (5732) (2005) 310–311.
[66] M. Winther, et al., Evaluation of miR-21 and miR-375 as prognostic biomarkers in esophageal cancer, Acta Oncol. 54 (9) (2015) 1582.
[67] Y. Zhu, et al., MiR-16 induced the suppression of cell apoptosis while promote proliferation in esophageal squamous cell carcinoma, Cell. Physiol. Biochem. 33 (5) (2014) 1340–1348.
[68] H.-S. Zhang, et al., Tanshinone IIA inhibits human esophageal cancer cell growth through miR-122-mediated PKM2 down-regulation, Arch. Biochem. Biophys. 598 (2016) 50–56.
[69] C. Liu, et al., miR-29a activates Hes1 by targeting Nfia in esophageal carcinoma cell line TE-1, Oncol. Lett. 9 (1) (2015) 96–102.
[70] A. Bansal, et al., Feasibility of microRNAs as biomarkers for Barrett's esophagus progression: a pilot cross-sectional, phase 2 biomarker study, Am. J. Gastroenterol. 106 (6) (2011) 1055–1063.
[71] C. Jing, et al., MicroRNA-17/20a impedes migration and invasion via TGF-β/ITGB6 pathway in esophageal squamous cell carcinoma, Am. J. Cancer Res. 6 (7) (2016) 1549.
[72] M. Schubert, K. Junker, J. Heinzelmann, Prognostic and predictive miRNA biomarkers in bladder, kidney and prostate cancer: Where do we stand in biomarker development? J. Cancer Res. Clin. Oncol. 142 (8) (2016) 1–23.
[73] M.L. Alvarez, et al., Role of microRNA 1207–5P and its host gene, the long non-coding RNA Pvt1, as mediators of extracellular matrix accumulation in the kidney: implications for diabetic nephropathy, PLoS One 8 (10) (2013) e77468.
[74] S. Li, et al., microRNA-155 silencing inhibits proliferation and migration and induces apoptosis by upregulating BACH1 in renal cancer cells, Mol. Med. Reports 5 (4) (2012) 949–954.
[75] G. Yu, et al., miRNA-34a suppresses cell proliferation and metastasis by targeting CD44 in human renal carcinoma cells, J. Urol. 192 (4) (2014) 1229–1237.
[76] X. Wang, et al., MicroRNA-200a-3p suppresses tumor proliferation and induces apoptosis by targeting SPAG9 in renal cell carcinoma, Biochem. Biophys. Res. Commun. 470 (3) (2016) 620–626.
[77] S.Z. Vahed, et al., Dysregulation of urinary miR-21 and miR-200b associated with interstitial fibrosis and tubular atrophy (IFTA) in renal transplant recipients, Clin. Biochem. 50 (1) (2017) 32–39.
[78] F.C. Tsz-fung, et al., Differential expression profiling of microRNAs and their potential involvement in renal cell carcinoma pathogenesis, Clin. Biochem. 43 (1) (2010) 150–158.
[79] P. Yang, et al., Ensemble positive unlabeled learning for disease gene identification, PloS One 9 (5) (2014) e97079.
[80] P. Yang, et al., Positive-unlabeled learning for disease gene identification, Bioinformatics 28 (20) (2012) 2640–2647.