

RESEARCH

Open Access



FKL-Spa-LapRLS: an accurate method for identifying human microRNA-disease association

Limin Jiang^{1,2}, Yongkang Xiao³, Yijie Ding⁴, Jijun Tang^{1,2,5} and Fei Guo^{1*}

From 29th International Conference on Genome Informatics
Yunnan, China. 3-5 December 2018

Abstract

Background: In the process of post-transcription, microRNAs (miRNAs) are closely related to various complex human diseases. Traditional verification methods for miRNA-disease associations take a lot of time and expense, so it is especially important to design computational methods for detecting potential associations. Considering the restrictions of previous computational methods for predicting potential miRNAs-disease associations, we develop the model of FKL-Spa-LapRLS (Fast Kernel Learning Sparse kernel Laplacian Regularized Least Squares) to break through the limitations.

Result: First, we extract three miRNA similarity kernels and three disease similarity kernels. Then, we combine these kernels into a single kernel through the Fast Kernel Learning (FKL) model, and use sparse kernel (Spa) to eliminate noise in the integrated similarity kernel. Finally, we find the associations via Laplacian Regularized Least Squares (LapRLS). Based on three evaluation methods, global and local leave-one-out cross validation (LOOCV), and 5-fold cross validation, the AUCs of our method achieve 0.9563, 0.8398 and 0.9535, thus it can be seen that our method is reliable. Then, we use case studies of eight neoplasms to further analyze the performance of our method. We find that most of the predicted miRNA-disease associations are confirmed by previous traditional experiments, and some important miRNAs should be paid more attention, which uncover more associations of various neoplasms than other miRNAs.

Conclusions: Our proposed model can reveal miRNA-disease associations and improve the accuracy of correlation prediction for various diseases. Our method can be also easily extended with more similarity kernels.

Keywords: MiRNA-disease association, Similarity kernel, Fast kernel learning, Sparse kernel, Laplacian regularized least squares

Background

MicroRNAs (miRNAs) are some of non-coding RNAs with 20 ~ 25 nucleotides [1]. In the process of post-transcription, miRNAs are a part of messenger RNA (mRNA) sequences and affect protein synthesis [2–4]. Some previous studies have proved that miRNAs are related to various diseases including cancers. For example, the expression level of *hsa-mir-21* leads

to more than 125 diseases, such as Alzheimer Disease, Diabetes Mellitus, Lymphoma and so on. Thus, the research of miRNAs is helpful for the diagnosis and treatment of diseases [5]. The traditional experiments to detect the associations between miRNAs and diseases are time-consuming and expensive [6]. Therefore, it is especially important to find potential miRNA-disease associations by the computational methods [7]. Previous researches achieved massive miRNA-disease associations through the traditional experiments, and some databases have been constructed for miRNA-disease associations.

*Correspondence: fguo@tju.edu.cn

¹School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China

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



Human MicroRNA Disease Database (HMDD) [8] collects 572 miRNAs, 378 Disease and 10368 miRNA-disease associations. The miR2Disease [9] includes 349 miRNAs, 163 disease and 3273 miRNA-disease associations. The dbDEMC contains 2224 miRNAs, 36 cancer types and 20037 miRNA-disease associations through the high-throughput methods. Thus, these associations promote the development of the computing methods.

Up to now, it has achieved excellent performance that people find the potential disease-miRNA associations by the computational methods [10–14]. Most of these methods are based on the assumption that miRNAs with high similarity apt to be related with similar diseases and vice versa [15, 16]. Xuan et al. [17] proposed HDMP that achieves a score for one miRNA by weighting k most similar neighbors, and a larger score has higher possibility to associate with a specific disease, but HDMP can't work for a new disease without known related miRNAs. Jiang et al. [18] devised a hypergeometric distribution-based model to calculate the score of each miRNA for a specific disease, and the miRNA with larger score tend to cause this disease. Scores of above two methods are based on miRNA neighbor information, which ignores entire informations of miRNA similarity network. Many models find miRNA-disease associations based on the similarity networks [19–23]. Chen et al. developed the RWRMDA model [24], which uses the information of miRNA functional similarity network and known miRNA-disease association network, and utilizes the random walk model to find the potential miRNA-disease association. However, RWRMDA is faced with the same problem as HDMP, because of the initial nonzero vector. Therefore, Chen et al. [25] proposed WBSMDA to find the potential association by integrating the miRNA functional similarity network, disease semantic similarity and known miRNA-disease association network. For the similarity between two miRNAs/diseases, WBSMDA integrates Gaussian Interaction Profile (GIP) kernel similarity for miRNA and disease, and calculates the association probability for miRNA-disease pair using Within-Score and Between-Score of disease and miRNA. Gu et al. [26] developed NCPMDA by constructing novel similarity kernel for miRNA and disease via the matrix operation and calculating the space projection scores of miRNA and disease. The final score between miRNA and disease is calculated by combining two space projection scores. The predictive performance of NCPMDA is superior over the previous methods when working for a disease without any known related miRNAs [13].

Many previous models are based on defining a cost function and minimizing this cost function. Chen et al. [27] developed RLSMDA, a semi-supervised method, which minimizes the Regularized Least Squares cost

function and uncovers the potential miRNAs associated with various diseases. After that, Chen et al. [28] proposed LRSSLMDA, which is used to reveal the potential association between miRNA and disease. LRSSLMDA constructs comprehensive statistical features and graph theoretic features by combining the miRNA and disease similarity kernels. Then, Laplacian regularization term is used to add objective function. Experimental results demonstrate that LRSSLMDA is a valuable computational model. In addition, many previous methods are based on machine learning algorithms [29, 30], matrix completion [31–33] and graph theory [34]. For example, Shen et al. [35] proposed CMFMDA that uses WKNKN to estimate association probability for unknown associations between miRNA and disease, and uses Collaborative Matrix Factorization to uncover the potential association. You et al. [36] developed PBMDA that constructs a heterogeneous graph by integrating five networks, gets all scores of paths for a miRNA-disease pair, and calculates the miRNA-disease association possibility through the sum of all path score. PBMDA gets a remarkable performance to find the potential miRNA-disease association.

All above methods have achieved remarkable results, but there are still different limitations or restrictions. For example, most of the existing methods are based on the assumption that miRNAs with high similarity apt to be related with similar diseases. About constructing miRNA and disease similarity kernel, most researches use the functional similarity and GIP kernel similarity for miRNA, and use the semantic similarity and GIP kernel similarity for disease. To integrate two similarity kernels, lots of works only tend to accumulate or average [29, 37, 38]. Therefore, there is an urgent need to propose an effective method for integrating multiple miRNA and disease similarity kernels [39].

In this paper, we firstly extract the miRNA functional similarity, the miRNA sequence similarity and GIP kernel similarity for miRNA, and the disease semantic similarity, disease functional similarity and GIP kernel similarity for disease. Then, we use the Fast Kernel Learning method to construct one miRNA similarity kernel and one disease similarity kernel. Finally, we propose a novel Sparse Laplacian Regularized Least Squares method to uncover the miRNA-disease association. Here, three evaluation methods are used to assess performance, including global Leave-One-Out Cross Validation (global LOOCV), local Leave-One-Out Cross Validation (local LOOCV) and 5-fold cross validation (5-fold CV). In these three evaluation methods, our method obtains the remarkable performance (AUCs of 0.9563, 0.8398 and 0.9535, respectively) compared with other nine models. And also, we use case studies of eight Neoplasms for further analyzing the

performance of our method. We find that 47 of top 50 candidates are confirmed to have associations with Lymphoma in global verification, and all top 50 candidates are confirmed to have associations with Breast and Colorectal Neoplasms in local verification. Moreover, we find that some of the miRNAs need to be paid more attention to uncover more associations with various neoplasms, including hsa-mir-106b, hsa-mir-19b, hsa-mir-29c, hsa-mir-1, hsa-mir-29a and so on.

Methods

We firstly use three miRNA similarity kernels and three disease similarity kernels to uncover potential miRNA-disease associations, respectively. Then, we combine these similarity kernels into a miRNA similarity kernel and a disease similarity kernel using Fast Kernel Learning, and sparse two similarity kernels after combination. Finally, we use Laplacian Regularized Least Squares to construct a loss function and get predicted association matrix from

miRNA and disease space, respectively. Figure 1 is the flow chart of our method.

Human miRNA-disease associations dataset

In this paper, the set of miRNAs is denoted by $M = \{m_i\}_{i=1}^m$, and the set of diseases is denoted by $D = \{d_j\}_{j=1}^n$, where m and n are the numbers of miRNAs and diseases respectively. The associations between miRNAs and diseases can be downloaded from HMDD database, which include 5430 associations between 495 miRNAs and 383 diseases. The associations are represented by a binary matrix $Y \in R^{m \times n}$, where $y_{ij} \in \{0, 1\}$. If a miRNA m_i is association with a disease d_j , y_{ij} is set to 1; otherwise, y_{ij} is set to 0;

MiRNA similarity

Basing on the assumption that miRNAs with high similarity tend to be associated with the same disease, we extract three classes of miRNA similarity, including functional

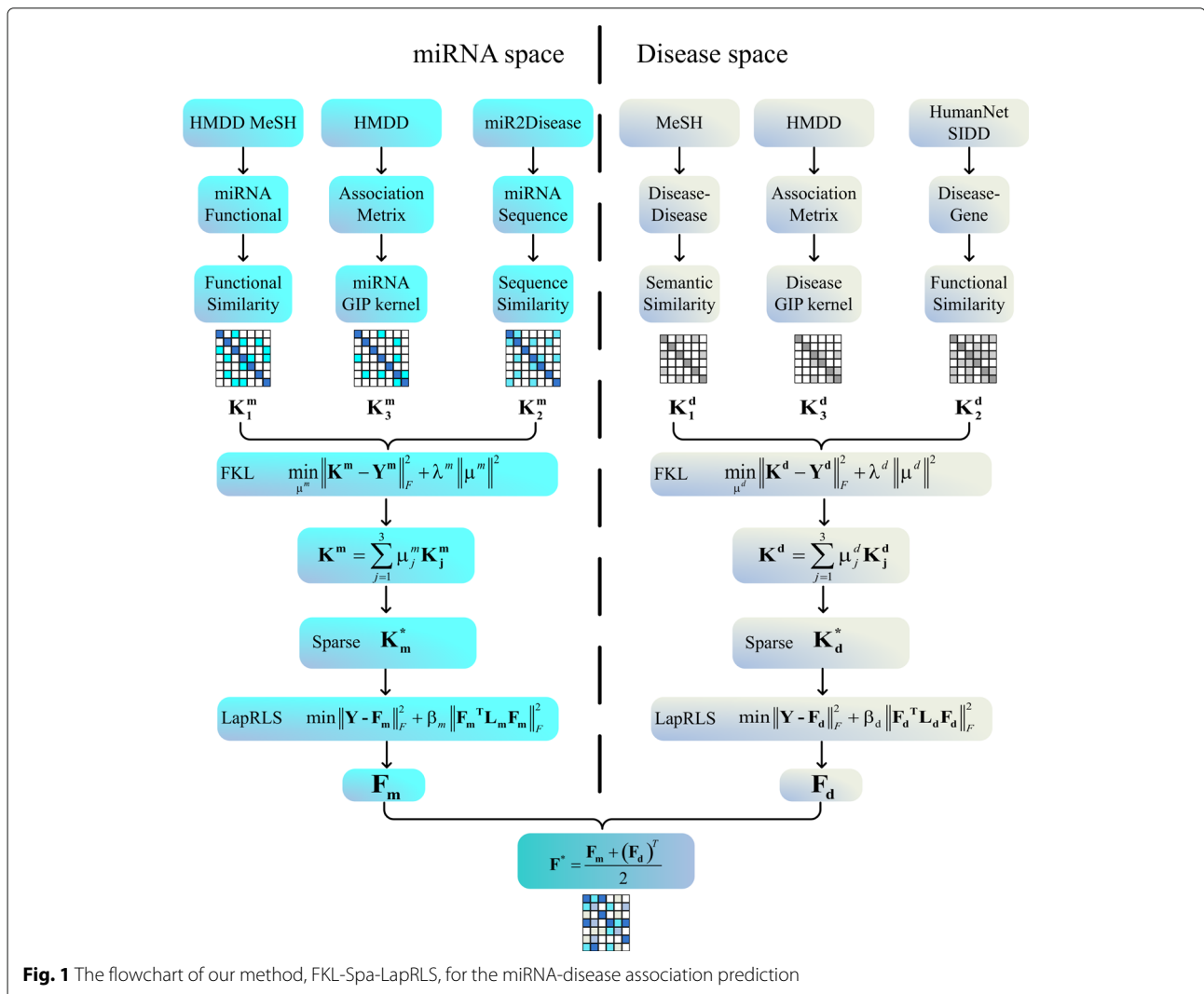


Fig. 1 The flowchart of our method, FKL-Spa-LapRLS, for the miRNA-disease association prediction

similarity, sequence similarity and Gaussian Interaction Profile (GIP) kernel similarity.

MiRNA functional similarity

In the previous works, the MISIM method [40] proposed by Cui et al calculated the score of miRNA functional similarity. We extract 495 functional similarity score through MISIM and construct kernel $K_1^m \in R^{m \times m}$ to represent the miRNA functional similarity network, in which $K_1^m(m_i, m_j)$ is the functional similarity score between miRNAs m_i and m_j .

MiRNA sequence similarity

All 495 miRNA sequences are downloaded from miRBase database [41]. We extract miRNA sequence similarity using the Needleman-Wunsch Algorithm and get kernel $K_2^m \in R^{m \times m}$ to represent the miRNA similarity of sequence network, in which $K_2^m(m_i, m_j)$ is the similarity of sequence score between miRNA m_i and m_j .

GIP kernel similarity for miRNAs

GIP the kernel similarity [29, 38, 42] between miRNAs m_i and m_j is denoted as $K_3^m \in R^{m \times m}$ and the calculation method is as Eq. (1)

$$K_3^m(m_i, m_j) = \exp(-\gamma_m \| IP(m_i) - IP(m_j) \|^2) \quad (1)$$

where $IP(m_i) \in R^{1 \times n}$ denotes the interaction profiles of miRNA m_i by observing whether miRNA m_i is associated with each disease or not, that is to say, the i -th row of the associations matrix Y ; γ_m is used for kernel bandwidth control, which is set to -1 in this paper.

Disease similarity

We extract three classes of disease similarity, including semantic similarity, functional similarity and GIP kernel similarity.

Disease semantic similarity

In the previous research [37, 40], disease $d(i)$ can be described as a node in Directed Acyclic Graph(DAG) based on the MeSH [43] database (<https://www.nlm.nih.gov/bsd/disted/meshtutorial/themeshdatabase/>), and denoted as $DAG_{d_i} = (d_i, T_{d_i}, E_{d_i})$, in which T_{d_i} is the set of all ancestor nodes of d_i including node d_i itself and E_{d_i} is the set of corresponding links. A semantic score of each disease $t \in T_{d_i}$ can be calculated by Eq. (2).

$$D_{d_i}(t) = \begin{cases} 1 & \text{if } t = d_i \\ \max \{ \Delta * D_{d_i}(t') | t' \in \text{children of } t \} & \text{if } t \neq d_i \end{cases} \quad (2)$$

where Δ is the semantic contribution factor, which is set to 0.5 in this paper.

Then, we define the semantic score of disease d_i by Eq. (3).

$$DV(d_i) = \sum_{t \in T_{d_i}} D_{d_i}(t) \quad (3)$$

Therefore, we denote the disease semantic similarity as $K_1^d \in R^{n \times n}$ and the disease semantic similarity value between d_i and d_j is calculated by Eq. (4).

$$K_1^d(d_i, d_j) = \frac{\sum_{t \in T_{d_i} \cap T_{d_j}} (D_{d_i}(t) + D_{d_j}(t))}{DV(d_i) + DV(d_j)} \quad (4)$$

Disease functional similarity

The associations between disease-gene and gene-gene are widely used to understand disease similarity [44]. From the HumanNet [45] database, we download the interactions of genes and one interaction has an log likelihood score (LLS) that measure the probability of a functional linkage between genes. The LLS scores are normalized by Eq. (5)

$$LLS^*(g_i, g_j) = \frac{LLS(g_i, g_j) - LLS_{min}}{LLS_{max} - LLS_{min}} \quad (5)$$

where $LLS(g_i, g_j)$ represents LLS between the i -th and j -th genes; $LLS^*(g_i, g_j)$ represents the LLS score after normalization; LLS_{min} and LLS_{max} indicate the minimum and maximum LLS scores in HumanNet respectively.

The functional similarity score between two genes is defined as Eq. (6)

$$FS(g_i, g_j) = \begin{cases} 1 & \text{if } i = j \\ LLS^*(g_i, g_j) & \text{if } i \neq j \text{ and } e(i, j) \in S_{HumanNET} \\ 0 & \text{if } i \neq j \text{ and } e(i, j) \notin S_{HumanNET} \end{cases} \quad (6)$$

where $S_{HumanNET}$ indicates the gene-gene associations in the HumanNet database; $e(i, j)$ indicates the association between i -th and j -th genes.

Then, the functional similarity score between a gene g and a gene set G is defined as Eq. (7).

$$F_G(g) = \max_{g_i \in G} (FS(g, g_i)) \quad (7)$$

In many cases, a disease d_i is related to many genes, which is defined as gene set G_i , the associations between disease and genes are download from SIDD [46]. The disease functional similarity score is defined as Eq. (8)

$$K_2^d(d_i, d_j) = \frac{\sum_{g_k \in G_j} F_{G_i}(g_k) + \sum_{g_s \in G_i} F_{G_j}(g_s)}{|G_j| + |G_i|} \quad (8)$$

GIP kernel similarity for diseases

Similar to calculation of GIP kernel similarity for miRNA, GIP kernel similarity for disease is denoted as $K_3^d \in R^{n \times n}$, calculated as Eq. (9).

$$K_3^d(d_i, d_j) = \exp(-\gamma_d \| IP(d_i) - IP(d_j) \|^2) \quad (9)$$

where $IP(d_i) \in R^{m \times 1}$ denotes the interaction profiles of disease d_i by observing whether disease d_i is associated with each miRNA or not, that is to say, the i -th column of the associations matrix Y ; γ_d is used for kernel bandwidth control, which is set to -1 in this paper.

Fast kernel learning

Considering that a single similarity kernel cannot cover all information between miRNAs, we integrate K_1^m, K_2^m for K_3^m to get a new integrated similarity kernel $K^m \in R^{m \times m}$ using the method of Fast Kernel Learning (FKL) [47]. We define K^m as Eq. (10).

$$K^m = \sum_{j=1}^3 \mu_j^m K_j^m \tag{10}$$

It is believed that K^m should be close to the associations matrix Y . We define the miRNAs associations similarity as Eq. (11).

$$Y^m = Y Y^T \tag{11}$$

Therefore, we would like to find $\mu^m \in R^{3 \times 1}$ using the following Eq. (12) to minimize the distance between K^m and Y^m .

$$\min_{\mu^m} \|K^m - Y^m\|_F^2 \tag{12}$$

where $\|K^m - Y^m\|_F^2 = \sum_i \sum_j (K_{ij}^m - Y_{ij}^m)^2$.

To avoid overfitting in learning procedure, a regularization term should be added to equation as Eq. (13).

$$\begin{aligned} \min_{\mu^m} & \|K^m - Y^m\|_F^2 + \lambda^m \|\mu^m\|^2 \\ \text{s.t.} & \mu_j^m \geq 0, j = 1, 2, 3 \\ & \sum_{j=1}^3 \mu_j^m = 1 \end{aligned} \tag{13}$$

where λ^m is set to 200 in this paper.

We use the matlab R2017a CVX to solve this optimization problem and obtain the integrate parameter $\mu^m \in R^{1 \times 3}$ for miRNA functional similarity, miRNA sequence similarity and GIP kernel similarity. Therefore, the integrated miRNA similarity kernel is defined as Eq. (14).

$$K^m = \sum_{j=1}^3 \mu_j^m K_j^m \tag{14}$$

Similarly, we obtain the integrate parameter $\mu^d \in R^{1 \times 3}$ for disease semantic similarity, disease functional similarity and GIP kernel similarity by FKL, and the integrated disease similarity kernel is defined as Eq. (15).

$$K^d = \sum_{j=1}^3 \mu_j^d K_j^d \tag{15}$$

Laplacian regularized least squares

Given the similarity kernels of miRNAs and diseases, we use Sparse Laplacian Regularized Least Squares (Spa-LapRLS) to get a new association matrix, and find potential miRNA-disease associations. It includes Sparse kernel model and LapRLS model.

Sparse kernel model

We use a Top- k Neighbor model to reduce noise in integrated similarity kernel. For the miRNA subspace, we construct a weight matrix $w_m \in R^{m \times m}$ for K^m , whose elements are defined as Eq. (16), by the Top- k Neighbor method.

$$w_m(i, j) = \begin{cases} 1 & \text{if } K^m(i, j) > \max(T(k, i), T(k, j)) \\ 0.5 & \text{if } K^m(i, j) \in [\min(T(k, i), T(k, j)), \max(T(k, i), T(k, j))] \\ 0 & \text{if } K^m(i, j) < \min(T(k, i), T(k, j)) \end{cases} \tag{16}$$

where k satisfies condition $0 < k < m$; $T(k, i)$ represents the k -th largest element of the i -th row in K^m and $T(k, j)$ represents the k -th largest element of the j -th column in K^m .

Therefore, we record the denoised miRNA similarity kernel as Eq. (17)

$$K_m^* = w_m \circ K^m \tag{17}$$

Similarity, we also calculate the denoised disease similarity kernel as $K_d^* \in R^{n \times n}$.

LapRLS for miRNA-disease interaction prediction

Given a pair of similarity kernels for miRNA K_m^* and disease K_d^* , we first use the Least Squares on the two subspace, and add Laplacian Regularization term to avoid overfitting. For miRNA subspace, the objective function of LapRLS [48] is defined as Eq. (18)

$$\min_{F_m} \|Y - F_m\|_F^2 + \beta_m \|F_m^T L_m F_m\|_F^2 \tag{18}$$

where $F_m = K_m^* \alpha_m \in R^{m \times n}$ is the predictive association matrix from miRNA; $L_m = D_m^{-\frac{1}{2}} (D_m - K_m^*) D_m^{-\frac{1}{2}}$, in which D_m is the diagonal matrix of K_m^* in the form of $D_m(i, i) = \sum_{j=1}^m K_m^*(i, j)$; β_m is the regularization coefficients, which is set to 2^{-5} in this paper; α_m is renewed by the function Eq. (19) in [48].

$$\alpha_m = \arg \min_{\alpha_m \in R^{m \times n}} \left\{ \|Y - K_m^* \alpha_m\|_F^2 + \beta_m \|\alpha_m^T K_m^* L_m K_m^* \alpha_m\|_F^2 \right\} \tag{19}$$

The derivation of the optimization algorithm are presented in [48].

In this way, the predicted associations matrix for all miRNA-disease pairs from the view of miRNAs are obtained as Eq. (20).

$$F_m = K_m^* (K_m^* + \beta_m L_m K_m^*)^{-1} Y \tag{20}$$

Similarly, we can get the predicted associations matrix for all miRNA-disease pairs from the view of miRNAs as Eq. (21)

$$F_d = K_d^* (K_d^* + \beta_d L_d K_d^*)^{-1} Y^T \tag{21}$$

where $F_d = K_d^* \alpha_d \in R^{n \times m}$; β_d is the regularization coefficients, which is set to 2^{-5} in this paper.

In the end, the predicted associations matrix from the view of miRNA and disease is defined as Eq. (22)

$$F^* = \frac{F_m + F_d^T}{2} \tag{22}$$

where $F^* \in R^{m \times n}$.

Results and discussion

In this section, we study the performance of our method from different aspects on prediction of unknown miRNA-disease associations. First, we establish three evaluation methods and two assessment indicators to evaluate the accuracy of our method. Second, we analyze the performance of our method with different parameters by using 10-fold CV and local LOOCV. Third, we employ 10-fold CV and local LOOCV to analyze the performance of the FKL model. Fourth, we compare the performance of LapRLS with multiple matrix factorization method. Fifth, we compare the performance of FKL-Spa-LapRLS with nine outstanding methods. Finally, for a further validation, we implement the global and local verifications on eight neoplasms for case studies.

Evaluation criteria

In this paper, we implement 10-fold CV, global LOOCV and local LOOCV to evaluate the prediction accuracy of our method. In the 10-fold CV, all miRNA-disease associations are randomly divided into ten uncrossed groups, one of which is regarded as test set and the other nine groups are used for training set in turns. In the global LOOCV, all 5430 miRNA-disease verified associations are regarded as objective research sample, and each association is left in turns served as a test sample and other known associations are regarded as training sample. In the local LOOCV, only considering miRNAs for a specific disease, for disease $d(i)$, each miRNA related to $d(i)$ is left out as test set, and other associations are regarded as training set. All the miRNA-disease associations in test set are reseted as 0 in the association matrix Y .

In our study, we use Area Under Curve (AUC) and Area Under the Precision-Recall curve (AUPR) to establish the assessment criteria for method prediction. AUC is the area under the receiver operating characteristic (ROC) created by plotting true positive rate against false positive rate at various threshold settings. An AUC value of 1 indicates perfect performance and an AUC of 0.5 indicates random performance. AUPR is the area under the

curve created by plotting precision against recall at various threshold setting. The greater the value of AUPR, the better performance of the model.

Parameter selection

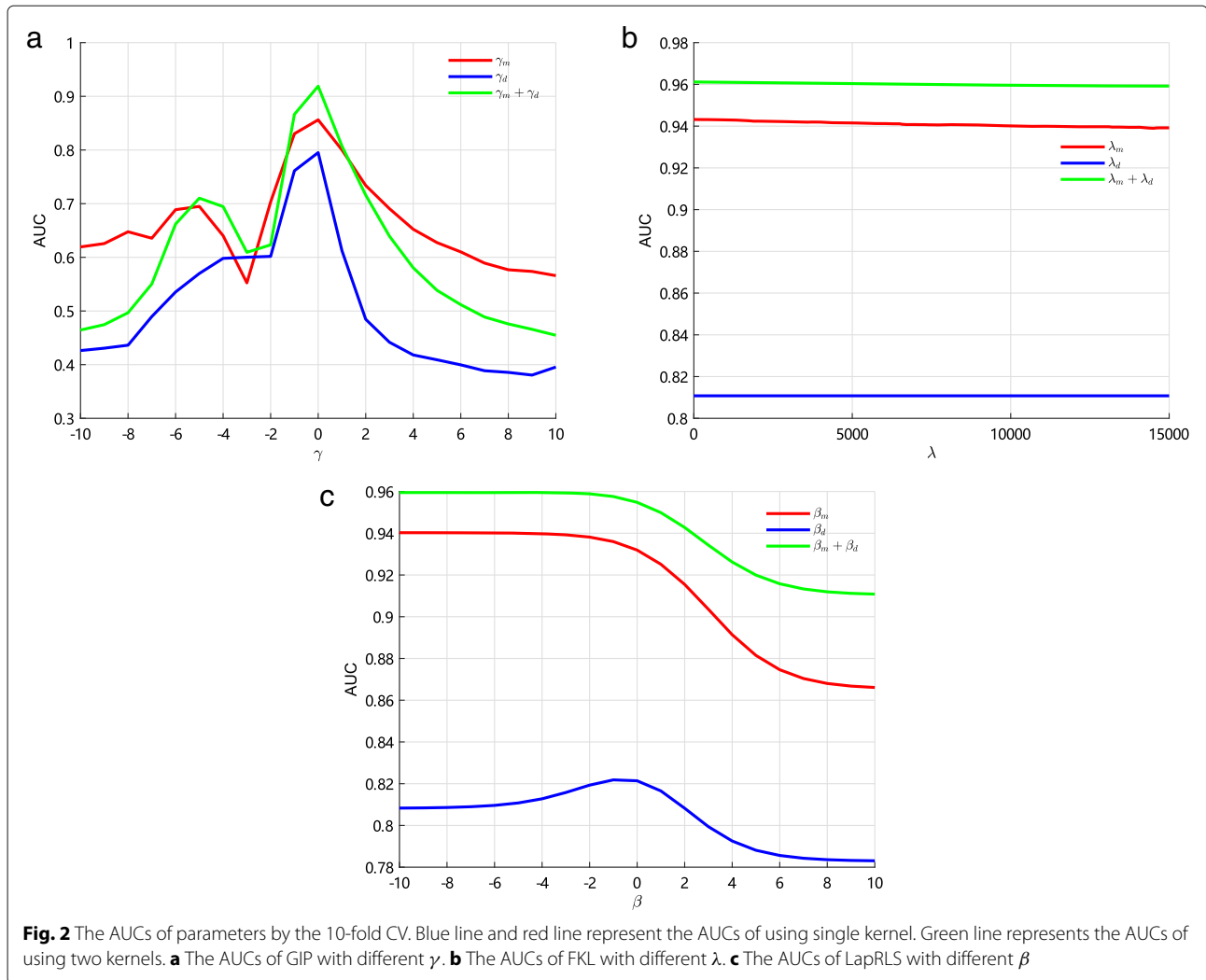
In this section, we use 10-fold CV and local LOOCV to analyze several parameters, including γ_m , γ_d , λ_m , λ_d , β_m , β_d and k value.

The γ_m and γ_d are the parameters in the process of constructing GIP kernel similarity for miRNA and diseases, respectively. We just use GIP kernel similarity to predict potential miRNA-disease associations and use 10-fold CV to evaluate performance of GIP kernel with different parameters. Then, we take γ_m and γ_d from -10 to 10 with step 1 and calculate AUCs, respectively. The results are shown in Fig. 2a. It shows that the performance of GIP similarity kernel is sensitive to γ_m and γ_d , and the optimal AUC is obtained when γ_m and γ_d equal to 0 . However, the $K_{m,3}$ and $K_{d,3}$ are matrices with ones in all elements according to Eqs. (1) and (9) when two parameters equal to 0 . Therefore, we adopt suboptimal $\gamma_m = -1$ and $\gamma_d = -1$ in this paper. Since most of elements in GIP similarity kernel are more than 1 , we need to normalize GIP similarity kernel before integrating multiple kernels.

The λ_m and λ_d are the regularization coefficients of FKL. We use different λ_m and λ_d to integrate three miRNA similarity kernels and three disease similarity kernels, respectively. Then we use integrated similarity kernel and LapRLS to uncover potential associations and use 10-fold CV to evaluate performance of FKL with different parameters. The λ_m and λ_d are gradually varying from 0 to 15000 with step 100 in order to find the best value. The results are shown in Fig. 2b. It can be found that AUC keeps small fluctuation in the range between 0 to 15000 . It demonstrates that FKL is insensitive to regularization coefficient. So, λ_m and λ_d are set to 200 in this paper.

The β_m and β_d are the regularization coefficients of LapRLS. We take β_m and β_d from 2^{-10} to 2^{10} , respectively. We adopt 10-fold CV to evaluate performance of LapRLS with different parameters. The results are shown in Fig. 2c. It can be found that AUC keeps small fluctuation in the range between 2^{-10} to 2^{-2} , and AUC has obvious change when β_m and β_d greater than 2^{-2} . We select the optimal β_m and β_d by the highest AUC value and set β_m and β_d as 2^{-5} in this paper.

Meanwhile, k value in the process of sparse kernel is an important parameter in this paper. We use 10-fold CV and local LOOCV to analyze k value. The value of k is taken from 20 to 250 with step 5 , are shown in Fig. 3. It can be clearly seen that the process of sparse kernel has positive effect on the discovery of potential miRNA-disease associations. In this study, k value is set to 20 in the 10-fold CV and global LOOCV, and is set to 40 in the local LOOCV.



FKL performance analysis

In this section, we analyze the performance of FKL. First, we compare FKL with single kernel and average kernel by the 10-fold CV and local LOOCV. Then, we compare FKL with two multiple kernels learning method by the 10-fold CV and local LOOCV.

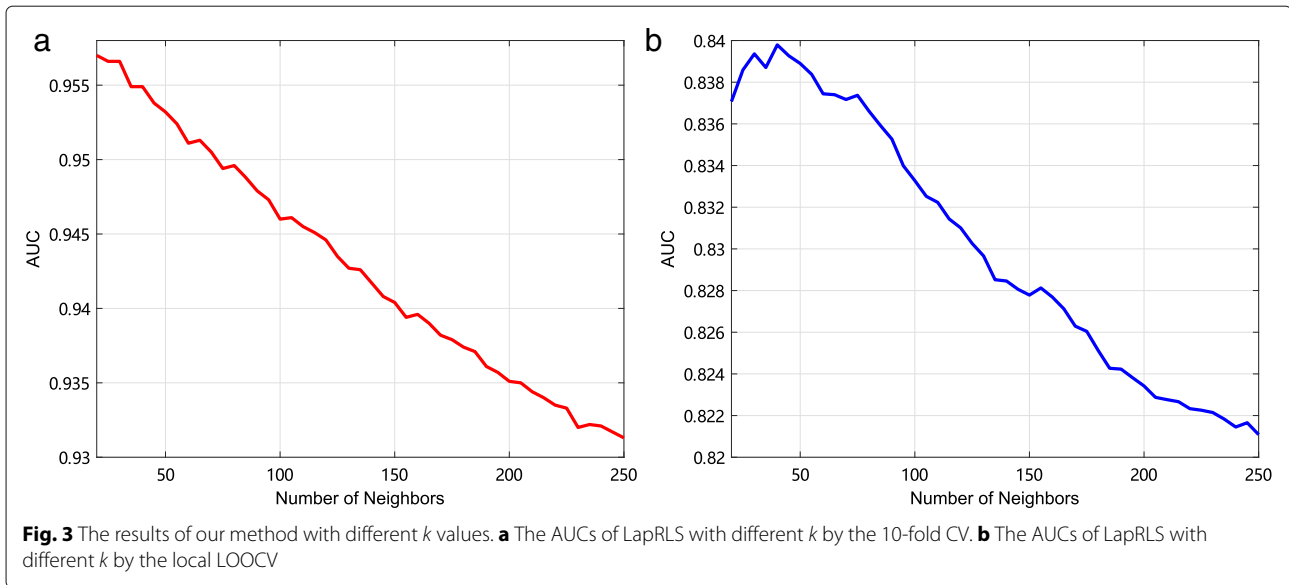
Comparison with single kernel and average kernel

We compare the prediction performance of FKL with three single similarity kernels and an average similarity kernels by using 10-fold CV and local LOOCV methods. The experiments are remarked as following.

$$\begin{cases} K_1^m \ \& \ K_1^d & K_1 \\ K_2^m \ \& \ K_2^d & K_2 \\ K_3^m \ \& \ K_3^d & K_3 \\ \text{avg}(K_1^m, K_2^m, K_3^m) \ \& \ \text{avg}(K_1^d, K_2^d, K_3^d) & \text{AVG} \\ K_m^* \ \& \ K_d^* & \text{FKL} \end{cases} \quad (23)$$

The comparison results obtained by the 10-fold CV and local LOOCV are shown in Fig. 4.

In the 10-fold CV, The AUC of FKL is the highest among five curves, and the AUC difference between the FKL model and the K_1 is slight but the difference in AUPR is obvious. Local LOOCV is a measure that can express model performance excellently when we handle a new disease not having known associations with miRNA. In Fig. 4, the AUC of average kernel is greater than FKL kernel. In the process of KFL, we need to find a optimized μ to weight kernels. Here, we get $\mu^m = (0.6610, 0.3390, 1.1562 \times 10^{-9})$ and $\mu^d = (1, 9.1453 \times 10^{-10}, 7.3854 \times 10^{-10})$, that is to say, the miRNA functional similarity kernel and the miRNA sequence similarity kernel are more important than GIP kernel similarity, and disease semantic similarity kernel is the most important in the three kernels. The model loses a part of information in the weighting process. However, a new disease not having any known association



with miRNA needs more detail information from different aspects. The average kernel method satisfies this requirement of more detail informations. That is why the AUC of FKL model is lower than average kernel, but the AUPR of FKL model is higher than average kernel method. Moreover, AUPR can evaluate the classifier performance better when dealing with unbalanced dataset. Therefore, it demonstrates that the FKL model is most significant in all kinds of models.

Comparison with other multiple kernel learning methods

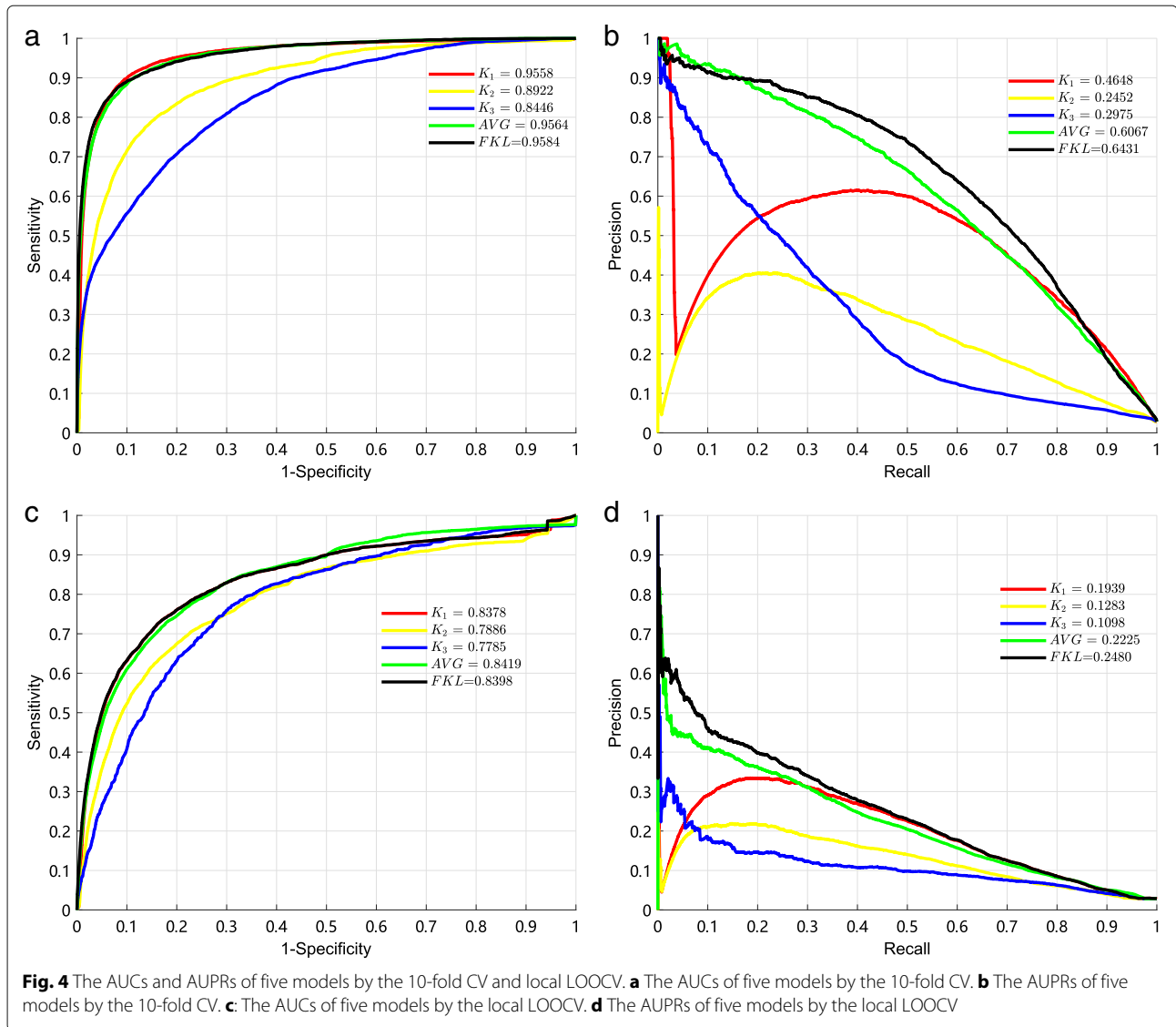
Several multiple kernel learning methods have been proposed to predict microRNA-disease associations, including Kronecker regularized least squares (KRLS) [39, 49] and kernelized Bayesian matrix factorization (KBMF) [32, 50]. We compare FKL with these two methods to integrate the similarity kernels to predict potential associations, respectively. Then, we use 10-fold CV and local LOOCV to evaluate performance of these three methods. The comparison results are shown in Fig. 5. In the 10-fold CV, it can be observed that the best AUC of 0.9584 and the best AUPR of 0.6431 are obtained by FKL. Comparing with KRLS, FKL achieves AUC improvement of 0.0162 (0.9584 over 0.9422) and AUPR improvement of 0.1201 (0.6431 over 0.5230). Comparing with KBMF, FKL achieves AUC improvement of 0.0598 (0.9584 over 0.8986) and AUPR improvement of 0.2005 (0.6431 over 0.4426). In local LOOCV, it can be observed that the best AUC of 0.8398 and the best AUPR of 0.2480 are also obtained by FKL. It shows that FKL is excellent at the aspect of uncovering associations between miRNAs and diseases.

Comparison with matrix factorization

The matrix factorization (MF) methods are widely used for different bioinformatics applications, including Protein-Protein interactions (PPI) prediction, drug-target interaction (DTI) prediction, drug response prediction, and so on. Therefore, we compare sparse LapRLS with four MF methods, including Similarity-Regularized Matrix Factorization (SRMF) [51], Collaborative Matrix Factorization (CMF) [52], Neighborhood Regularized Logistic Matrix Factorization (NRLMF) [53] and Graph Regularized Matrix Factorization (GRMF) [54]. We use the same integrated similarity kernels and these five methods to predict potential associations, and adopt 10-fold CV to evaluate performance of different methods. The results are shown in Fig. 6. In 10-fold CV, it can be observed that the best AUC of 0.9584 and the best AUPR of 0.6431 are obtained by spa-LapRLS. In local LOOCV, it can be observed that the best AUC of 0.8398 and the best AUPR of 0.2480 are also obtained by sparse LapRLS. It demonstrates that sparse LapRLS is reliable for predicting miRNA-disease associations.

Comparison with other methods

We furtherly compare the performance of FKL-Spa-LapRLS with nine computational prediction models (i.e., PBMDA [36], MCMMDA [31], MaxFlow, NCPMDA [26], WBSMDA [25], HDMP [17], RLSMDA [27], LRSSLMDA [28], HGIMDA [55]), and the comparisons are shown in Table 1. In the local LOOCV, FKL-Spa-LapRLS gets an AUC of 0.8398, which is slightly under performance of NCPMDA (0.8584) and LRSSLMDA (0.8418). However, in the global LOOCV, our method gets an AUC of 0.9563,



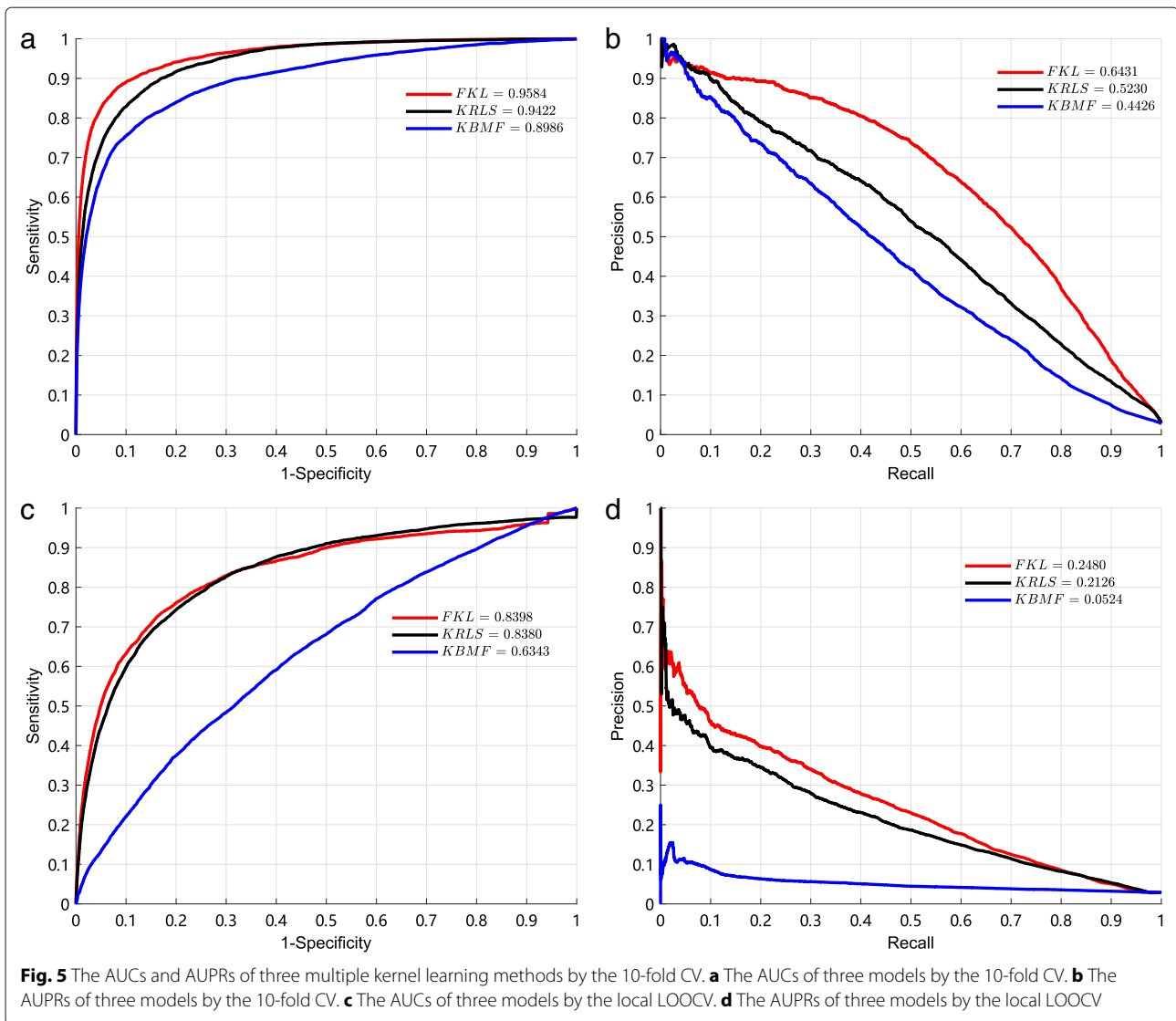
which is significant superior to the result of other methods. In the 5-fold, FKL-Spa-LapRLS obtains an AUC of 0.9535, which also has a great outperformance than other methods. Therefore, FKL-Spa-LapRLS improves the prediction performance of disease-miRNA associations from different evaluation measures.

Case studies

In this section, we study several important diseases to further validate the predictive power of our method. We utilize the known miRNA-disease associations included in HMDD to find the potential miRNA-disease associations not included in HMDD, and verify the predicted results through two independent databases (dbDEMOC [56] and miR2Disease [9]). In fact, dbDEMOC and miR2Disease are commonly utilized to be benchmark datasets for

many models, such as PBMDA and LRSSLMDA. The dbDEMOC database includes 2224 miRNAs, 36 cancer types and 20037 miRNA-disease associations by the high-throughput method, and our model predicts the top five disease, including Colon Neoplasms, Gastric Neoplasms, Pancreatic Neoplasms, Colorectal Neoplasms and Esophageal Neoplasms. Furthermore, in previous work, Kidney Neoplasms, Breast Neoplasms and Lymphoma were used to infer their underlying associated miRNAs. Therefore, we use case studies of eight diseases to analyze the performance of FKL-Spa-LapRLS in this section.

We implement two methods, global validation and local validation, to evaluate the predicted performance of our method in case studies. In global verification, 5430 known miRNA-disease associations in HMDD are used as a training set to discover the potential associations. For each



disease, we extract top 50 candidate associations that can't be covered by training set. And we get all of 400 candidate associations that are checked by dbDEMC and miR2Disease databases. In the local validation, all known associations that are related to a special disease are reset to unknown ones. We use other known associations as training set to discover the potential associations. we also extract top 50 candidate associations for this special disease. And we obtain all of 400 candidate associations that are checked by the HMDD, miR2Disease and dbDEMC databases.

The verification results of eight diseases are listed in Table 2. In Table 2, the global verification is the number of confirmed associations by dbDEMC and miR2Disease in top 50 miRNAs. And the local verification is the number of identified associations by HMDD, dbDEMC and

miR2Disease. In Table 2, we can find that 47 of top 50 candidates are associated with lymphoma confirmed by global verification, and local verification confirms that all top 50 candidates are associated with breast and Colorectal Neoplasms.

The results of case studies and some special miRNAs are shown in Figs. 7 and 8 (detail results in Additional files 1, 2, 3, 4, 5, 6, 7 and 8). The green lines are the confirmed miRNA-disease associations, the red lines are the unconfirmed miRNA-disease associations, the black nodes are the eight neoplasms, and the brown nodes are the predicted miRNAs associated with diseases. There are 400 associations in Fig. 7, and we can find that most of the miRNA-disease associations are confirmed by the global verification. In addition, there are many miRNAs that are only related to Breast Neoplasms but they have

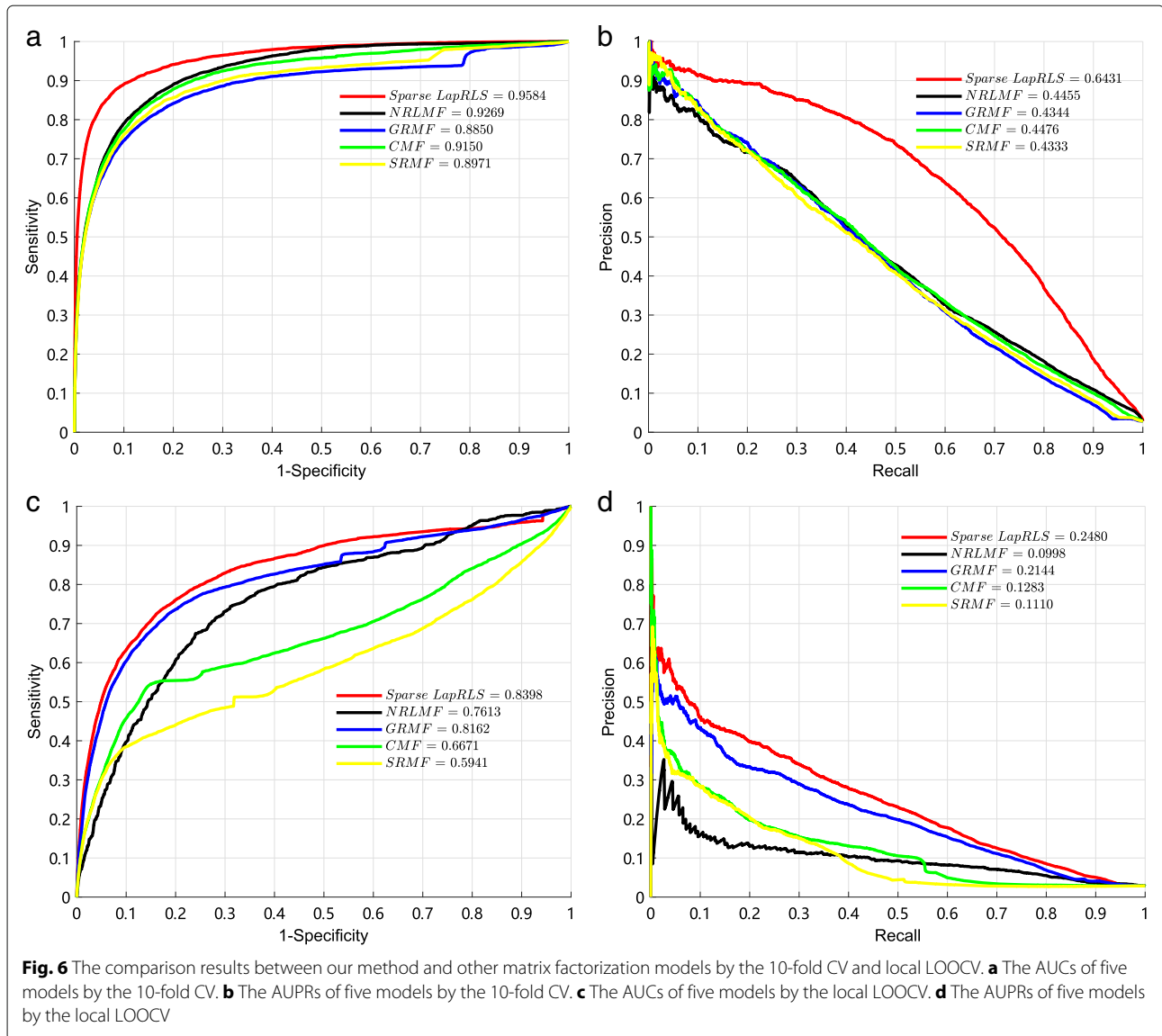


Table 1 The comparison results between our method and other nine computational models

Methods	Global LOOCV	Local LOOCV	5-fold CV
FKL-Spa-LapRLS	0.9563	0.8398	0.9535
PBMDA	0.9169	0.8341	0.9172
MCMDA	0.8749	0.7718	0.8767
MaxFlow	0.8624	0.7774	0.8579
NCPMDA	0.9073	0.8584	0.8763
WBSMDA	0.8030	0.8031	0.8185
HDMP	0.8366	0.7702	0.8342
RLSMDA	0.8426	0.6953	0.8569
LRSSLMDA	0.9178	0.8418	0.9181
HGIMDA	0.8781	0.8077	—

The boldface is the best value in the column

Table 2 The verification results about eight neoplasms types

Disease name	Global verification	Local verification
Colon Neoplasms	44	48
Gastric Neoplasms	42	40
Pancreatic Neoplasms	45	50
Colorectal Neoplasms	45	50
Esophageal Neoplasms	39	46
Kidney Neoplasms	43	43
Breast Neoplasms	39	50
Lymphoma	47	48

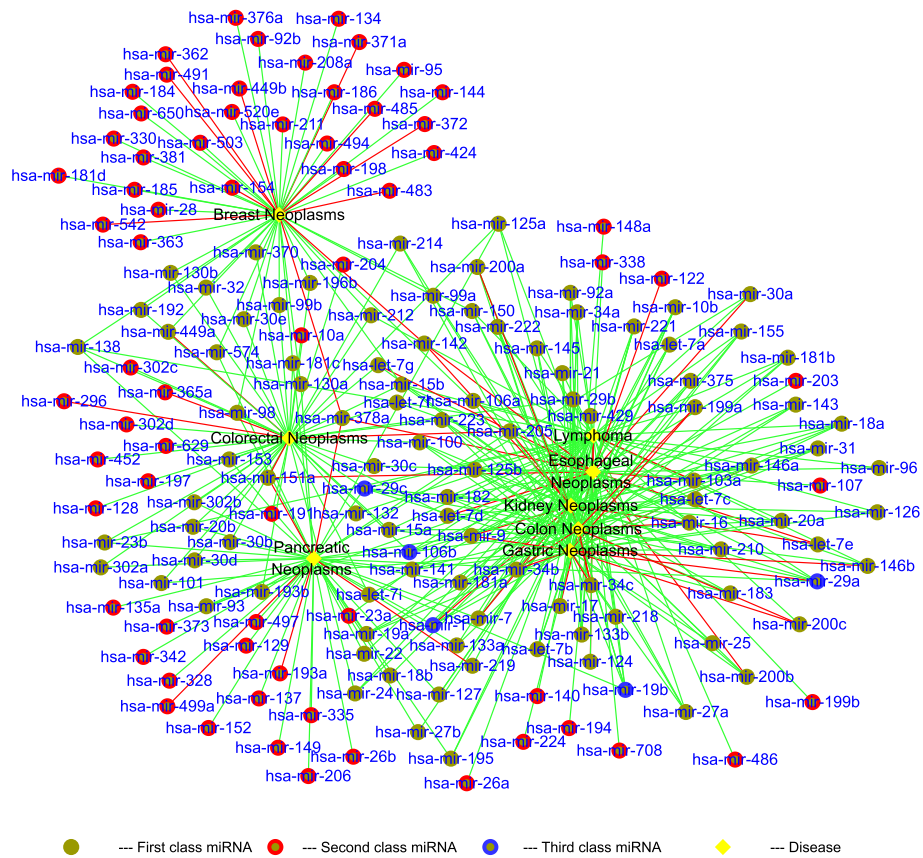


Fig. 7 The case studies by global verification. The green lines are the confirmed candidate associations. The red lines are the unconfirmed candidate associations. The black nodes are the disease. The brown nodes are the candidate miRNAs. First class miRNA represents miRNA associated with multiple diseases. Second class miRNA represents miRNA associated with one disease. Third class miRNA represents important miRNA associated with more than six diseases

nothing to do with other diseases. And there are nine associations are unconfirmed. The reason is that of total 495 miRNAs in the training set, 202 have been linked to Breast Neoplasms, so there is a large possibility that the remaining miRNAs have no association with it. Similarly, there are 11 miRNAs related to Esophageal Neoplasms but not confirmed. The reason is that there are already 74 miRNAs associated with the Esophageal Neoplasms in the training set. On the other hand, there are a few unconfirmed miRNAs associated with other six diseases. In Fig. 7, we can see that hsa-mir-106b, hsa-mir-19b and hsa-mir-29c are associated with six out of eight diseases, and these miRNAs should be paid more attention to reveal more associations. Moreover, hsa-mir-1 and hsa-mir-29a are expected to be associated with five diseases out of eight diseases, but these associations still have not been verified by valid experiment. In Fig. 8, we can find that most of miRNAs work on various diseases. For a special disease with unknown associations with miRNAs, our method can reveal the miRNAs associated with it, and only 26 associations out of 400 cannot be confirmed

by known experiments. These unconfirmed associations need to be paid more attention. Especially for hsa-let-7a, hsa-let-7b, hsa-mir-125b, hsa-mir-126, hsa-mir-145, hsa-mir-155, hsa-mir-181b, hsa-mir-20a, hsa-mir-21, hsa-mir-34a, hsa-mir-92a, these miRNAs are associated with all diseases. And we find that the related miRNAs among eight Neoplasms are highly similar. Therefore, it is very important to find more diseases related to these 11 miRNAs.

Conclusions

In this paper, we propose a FKL-Spa-LapRLS model to uncover potential miRNA-disease associations. We demonstrate that the KFL model is more importance than the average kernel method using 10-fold CV and local LOOCV, and the process of sparse kernel has a positive effect on noise elimination in similarity network. The LapRLS method contributes to accuracy of finding potential miRNA-disease associations.

FKL-Spa-LapRLS has been compared with nine prediction methods that have got excellent performance

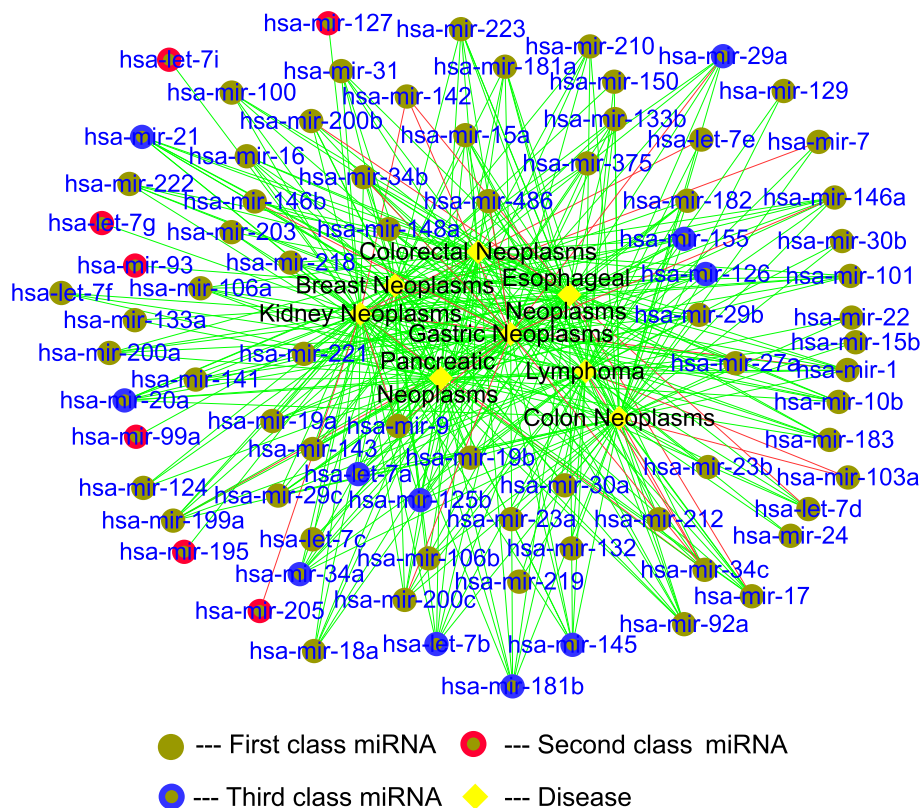


Fig. 8 The case studies by local verification. The green lines are the confirmed candidate associations. The red lines are the unconfirmed candidate associations. The black nodes are the disease. The brown nodes are the candidate miRNAs. First class miRNA represents miRNA associated with multiple diseases. Second class miRNA represents miRNA associated with one disease. Third class miRNA represents important miRNA associated with more than six diseases

for prediction of miRNA-disease associations, including PBMDA, MCMDA, MaxFlow, NCPMDA, WBSMDA, HDMP, RLSMDA, LRSSLMDA and HGIMDA. FKL-Spa-LapRLS has the significantly highest accuracy in 5-fold CV and global LOOCV, albeit weakly lower than NCPMDA and LRSSLMDA in local LOOCV. To further analyze the performance of FKL-Spa-LapRLS, we implement case studies of eight Neoplasms. We find that 47 of top 50 candidates are confirmed to be associated with Lymphoma in global verification and all the top 50 candidates are confirmed to be associated with Breast and Colorectal Neoplasms in local verification, and some miRNAs need to be paid more attention.

Of course, FKL-Spa-LapRLS also have some limitations that need to be improved in the future. For example, our method needs more similarity kernels that are constructed by many information about gene-disease, disease-disease and miRNA-miRNA, and it would lose some detail information in the process of FKL when handling a new disease without the known associations with miRNAs.

Additional files

- Additional file 1:** Table S1. The top 50 predicted miRNAs related to Colon Neoplasms. (XLSX 11 kb)
- Additional file 2:** Table S2. The top 50 predicted miRNAs related to Gastric Neoplasms. (XLSX 10 kb)
- Additional file 3:** Table S3. The top 50 predicted miRNAs related to Pancreatic Neoplasms. (XLSX 11 kb)
- Additional file 4:** Table S4. The top 50 predicted miRNAs related to Colorectal Neoplasms. (XLSX 10 kb)
- Additional file 5:** Table S5. The top 50 predicted miRNAs related to Esophageal Neoplasms. (XLSX 11 kb)
- Additional file 6:** Table S6. The top 50 predicted miRNAs related to Kidney Neoplasms. (XLSX 10 kb)
- Additional file 7:** Table S7. The top 50 predicted miRNAs related to Breast Neoplasms. (XLSX 11 kb)
- Additional file 8:** Table S8. The top 50 predicted miRNAs related to Lymphoma. (XLSX 11 kb)

Abbreviations

CMF: Collaborative matrix factorization; CV: Cross validation; FKL: Fast kernel learning; GIP: Gaussian interaction profile; GRMF: Graph regularized matrix factorization; HMDD: Human microRNA disease database; KBMF: Kernelized Bayesian matrix factorization; KRLS: Kronecker regularized least squares;

LapRLS: Laplacian regularized least squares; LLS: Log likelihood score; LOOCV: Leave-one-out cross validation; NRLMF: Neighborhood regularized logistic matrix factorization; SRMF: Similarity-regularized matrix factorization

Acknowledgements

Authors would like to thank the reviewers for their helpful comments on the original manuscript. Authors are grateful to the conference committee of The 29th International Conference on Genome Informatics (GIW 2018).

Funding

This work is supported by a grant from the National Science Foundation of China (NSFC 61772362) and the Tianjin Research Program of Application Foundation and Advanced Technology (16JCQNJC00200). Publication costs are funded by the NSFC 61772362.

Availability of data and materials

The code and all supporting data files are available from <https://github.com/guofei-tju/FKL-Spa-LapRLS>.

About this supplement

This article has been published as part of *BMC Genomics Volume 19 Supplement 10, 2018: Proceedings of the 29th International Conference on Genome Informatics (GIW 2018): genomics*. The full contents of the supplement are available online at <https://bmcbgenomics.biomedcentral.com/articles/supplements/volume-19-supplement-9>.

Authors' contributions

FG, YD and LJ conceived and designed the experiments; LJ performed the experiments and analyzed the data; YX wrote the paper. FG and JT supervised the experiments and reviewed the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflict of interest.

Publisher's Note

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

Author details

¹School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China. ²Tianjin University Institute of Computational Biology, Tianjin University, Tianjin, China. ³School of Chemical Engineering and Technology, Tianjin University, Tianjin, China. ⁴School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou, China. ⁵Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA.

Published: 31 December 2018

References

- Shi H, Zhang G, Zhou M, Cheng L, Yang H, Wang J, et al. Integration of Multiple Genomic and Phenotype Data to Infer Novel miRNA-Disease Associations. *Plos ONE*. 2016;11(2):e0148521.
- Zou Q, Li J, Hong Q, Lin Z, Wu Y, Shi H, et al. Prediction of MicroRNA-Disease Associations Based on Social Network Analysis Methods. *Biomed Res Int*. 2015;2015(10):810514.
- Yuan D, Cui X, Wang Y, Zhao Y, Li H, Hu S, et al. Enrichment Analysis Identifies Functional MicroRNA-Disease Associations in Humans. *Plos ONE*. 2015;10(8):e0136285.
- Zou Q, Li J, Song L, Zeng X, Wang G. Similarity computation strategies in the microRNA-disease network: a survey. *Brief Funct Genom*. 2016;15(1):55.
- Zeng X, Liu L, Lu L, Zou Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics*. 2018;34:2425–32.
- Zeng X, Zhang X, Zou Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief Bioinform*. 2016;17(2):193.
- Mørk S, Pletscher-Frankild S, Palleja CA, Gorodkin J, Jensen LJ. Protein-driven inference of miRNA-disease associations. *Bioinformatics*. 2014;30(3):392.
- Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*. 2014;42(Database issue):D1070.
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009;37(1):D98–104.
- Peng L, Peng M, Liao B, Huang G, Liang W, Li K. Improved low-rank matrix recovery method for predicting miRNA-disease association. *Sci Rep*. 2017;7(1):6007.
- Luo J, Ding P, Liang C, Chen X. Semi-supervised prediction of human miRNA-disease association based on graph regularization framework in heterogeneous networks. *Neurocomputing*. 2018;294:29–38.
- Zhao Q, Xie D, Liu H, Wang F, Yan GY, Chen X. SSCMDA: spy and super cluster strategy for MiRNA-disease association prediction. *Oncotarget*. 2018;9(2):1826–42.
- Liu Y, Zeng X, He Z, Quan Z. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans Comput Biol Bioinform*. 2016;PP(99):1–1.
- Shi H, Xu J, Zhang G, Xu L, Li C, Wang L, et al. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst Biol*. 2013;7(1):1–12.
- Luo J, Xiao Q. A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J Biomed Inform*. 2017;66:194–203.
- Lan W, Wang J, Li M, Liu J, Wu FX, Pan Y. Predicting microRNA-disease associations based on improved microRNA and disease similarities. *IEEE/ACM Trans Comput Biol Bioinform*. 2016;PP(99):1–1.
- Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, et al. Correction: Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors. *Plos ONE*. 2013;8(9):e70204.
- Jiang Q, Hao Y, Wang G, Juan L, Zhang T, Teng M, et al. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst Biol*. 2010;4(S1):S2.
- Pasquier C, Gardès J. Prediction of miRNA-disease associations with a vector space model. *Sci Rep*. 2016;6:27036.
- Yu Q, Zhang H, Cheng L, Xiao D. KATZMDA: Prediction of miRNA-disease associations based on KATZ model. *IEEE Access*. 2017;PP(99):1–1.
- Nalluri JJ, Kamapantula BK, Barh D, Jain N, Bhattacharya A, Almeida SSD, et al. DISMIRA: Prioritization of disease candidates in miRNA-disease associations based on maximum weighted matching inference model and motif-based analysis. *BMC Genom*. 2015;16 Suppl 5(S5):S12.
- Liao B, Ding S, Chen H, Li Z, Cai L. Identifying human microRNA-disease associations by a new diffusion-based method. *J Bioinform Comput Biol*. 2015;13(04):1550014.
- Zeng X, Liao Y, Liu Y, Zou Q. Prediction and Validation of Disease Genes Using HeteSim Scores. *IEEE/ACM Trans Comput Biol Bioinform*. 2016;99:1–1.
- Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. *Mol BioSyst*. 2012;8(10):2792.
- Chen X, Yan CC, Zhang X, You ZH, Deng L, Liu Y, et al. WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Sci Rep*. 2016;6:21106.
- Gu C, Bo L, Li X, Li K. Network Consistency Projection for Human miRNA-Disease Associations Inference. *Sci Rep*. 2016;6:36054.
- Chen X, Yan GY. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci Rep*. 2014;4:5501.
- Chen X, Huang L. LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for MiRNA-Disease Association prediction. *Plos Comput Biol*. 2017;13(12):e1005912.
- Fu L, Peng Q. A deep ensemble model to predict miRNA-disease association. *Sci Rep*. 2017;7(1):14482.
- Jiang Q, Wang G, Zhang T, Wang Y. Predicting human microRNA-disease associations based on support vector machine. *Int J Data Min Bioinform*. 2011;8(3):282–93.
- Li JQ, Rong ZH, Chen X, Yan GY, You ZH. MCMMA: Matrix completion for MiRNA-disease association prediction. *Oncotarget*. 2017;8(13):21187.

32. Lan W, Wang J, Li M, Liu J, Pan Y. Predicting microRNA-disease associations by integrating multiple biological information. In: IEEE International Conference on Bioinformatics and Biomedicine. Bioinformatics and Biomedicine; 2015. p. 183–8.
33. Zeng X, Ding N, Rodríguez-Patón A, Quan Z. Probability-based collaborative filtering model for predicting gene–disease associations. *BMC Med Genomics*. 2017;10(5):76.
34. Chen X, Guan NN, Li JQ, Yan GY. GIMDA: Graphlet interaction-based MiRNA-disease association prediction. *J Cel Mol Med*. 2018;22(3):1548–61.
35. Shen Z, Zhang YH, Han K, Nandi AK, Honig B, Huang DS. miRNA-Disease Association Prediction with Collaborative Matrix Factorization. *Complexity*. 2017;2017(9):1–9.
36. You ZH, Huang ZA, Zhu Z, Yan GY, Li ZW, Wen Z, et al. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *Plos Comput Biol*. 2017;13(3):e1005455.
37. You ZH, Wang LP, Chen X, Zhang S, Li XF, Yan GY, et al. PRMDA: personalized recommendation-based MiRNA-disease association prediction. *Oncotarget*. 2017;8(49):85568–83.
38. Peng L, Chen Y, Ma N, Chen X. NARRMDA: negative-aware and rating-based recommendation algorithm for miRNA-disease association prediction. *Mol BioSyst*. 2017;13:2650–59.
39. Chen X, Niu YW, Wang GH, Yan GY. MKRMDA: multiple kernel learning-based Kronecker regularized least squares for MiRNA–disease association prediction. *J Transl Med*. 2017;15(1):251.
40. 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(13):1644–50.
41. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(Database issue):D68.
42. Chen X, Niu YW, Wang GH, Yan GY. HAMDA: Hybrid Approach for MiRNA-Disease Association prediction. *J Biomed Inform*. 2017;76:50–58.
43. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Jama*. 1994;271(14):1103–8.
44. Luo J, Xiao Q, Liang C, Ding P. Predicting MicroRNA-Disease Associations Using Kronecker Regularized Least Squares Based on Heterogeneous Omics Data. *IEEE Access*. 2017;5(99):2503–13.
45. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res*. 2011;21(7):1109.
46. Liang C, Wang G, Li J, Zhang T, Xu P, Wang Y. SIDD: A Semantically Integrated Database towards a Global View of Human Disease. *Plos ONE*. 2013;8(10):e75504.
47. He J, Chang SF, Xie L. Fast kernel learning for spatial pyramid matching: *Computer Vision and Pattern Recognition*; 2008, pp. 1–7.
48. Xia Z, Zhou X, Sun Y, Wu LY. Semi-supervised Drug-Protein Interaction Prediction from Heterogeneous Spaces, Vol. 4; 2010. p. S6.
49. Nascimento ACA, Prudencio RBC, Costa IG. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinformatics*. 2016;17(1):46.
50. Gonen M, Kaski S. Kernelized Bayesian Matrix Factorization. *IEEE Trans Pattern Anal Mach Intell*. 2014;36(10):2047–60.
51. Wang L, Li X, Zhang L, Gao Q. Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer*. 2017;17(1):513.
52. Zheng X, Ding H, Mamitsuka H, Zhu S. Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining; 2013. p. 1025–33.
53. Liu Y, Wu M, Miao C, Zhao P, Li X. Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction. *PLoS Comput Biol*. 2016;12(2):e1004760.
54. Ezzat A, Zhao P, Wu M, Li X, Kwok CK. Drug-Target Interaction Prediction with Graph Regularized Matrix Factorization. *IEEE/ACM Trans Comput Biol Bioinform*. 2017;14(3):646–56.
55. Chen X, Yan CC, Zhang X, You ZH, Huang YA, Yan GY. HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget*. 2016;7(40):65257–69.
56. Yang Z, Ren F, Liu C, He S, Sun G, Gao Q, et al. dbDEMC: a database of differentially expressed miRNAs in human cancers. *BMC Genomics*. 2010;11(Suppl 4):1–8.

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

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

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

