GNMFLMI: Graph Regularized Nonnegative Matrix Factorization for Predicting LncRNA-MiRNA Interactions

Mei-Neng Wang[1], Zhu-Hong You[2, *], Li-Ping Li[2], Leon Wong[2], Zhan-Heng Chen[2], Cheng-Zhi Gan[1]

[1]School of Mathematics and Computer Science, Yichun University, Yichun Jiangxi 336000, China

[2]Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Sciences, Urumqi 830011, China

*** Correspondence:** zhuhongyou@ms.xjb.ac.cn

Abstract: Long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) have been involved in various biological processes. Emerging evidence suggests that the interactions between lncRNAs and miRNAs play an important role in regulating of genes and the development of many diseases. Due to the limited scale of known lncRNA-miRNA interactions, and expensive time and labor costs for identifying them by biological experiments, more accurate and efficient lncRNA-miRNA interactions computational prediction approach urgently need to be developed. In this work, we proposed a novel computational method, GNMFLMI, to predict lncRNA-miRNA interactions using graph regularized nonnegative matrix factorization. More specifically, the similarities both lncRNA and miRNA are calculated based on known interaction information and their sequence information. Then, the affinity graphs for lncRNAs and miRNAs are constructed using the *p*-nearest neighbors, respectively. Finally, a graph regularized nonnegative matrix factorization model is developed to accurately identify potential interactions between lncRNAs and miRNAs. To evaluate the performance of GNMFLMI, five-fold cross validation experiments are carried out. GNMFLMI achieves the AUC value of 0.9769 which outperforms the compared methods NMF and CNMF. In the case studies for lncRNA nonhsat159254.1 and miRNA hsa-mir-544a, 20 and 16 of the top-20 associations predicted by GNMFLMI are confirmed, respectively. Rigorous experimental results

demonstrate that GNMFLMI can effectively predict novel lncRNA-miRNA interactions, which can provide guidance for relevant biomedical research.

Keywords: lncRNA-miRNA interaction, nonnegative matrix factorization, graph regularization, lncRNA-miRNA similarity.

1 Introduction

With the development of next-generation sequencing, specific biological mechanisms can be better understood from the wide-ranging biomolecular interactions in the genome. Long non-coding RNAs (lncRNAs) and microRNAs (miRNAS) were previously thought to be non-functional sequences in the process of gene evolution [1]. In fact, they not only play an important role in cell differentiation, somatic development and other life processes, but also can participate in the occurrence of disease through interaction [1]. LncRNA is a type of non-coding RNA (ncRNA) located in the nucleus or cytoplasm of more than 200nt in length which has no obvious protein-coding function and exists in any branch of life [2]. Depending on the positional relationship of the coding genes, lncRNAs can be divided into five categories (i.e. bidirectional, antisense, sense, introverted and intergenic) [3-5]. Because of lncRNA has tissue specificity, cell specificity, spatiotemporal specificity, developmental stage specificity and disease specificity, it is widely involved in cell differentiation, metabolism and proliferation, and is closely associated with many complex diseases [6, 7]. More and more evidences have shown that lncRNAs can silence or activate genes by regulating histone modification, DNA methylation, mRNA splicing and chromatin remodeling in a variety of ways, such as epigenetics, transcriptional regulation, and post-transcriptional control and so on [8]. As a new focus of regulation for gene expression, lncRNA plays a biological role mainly through signal function, bait function, scaffold function and guiding function [9]. Even though the experiment has identified more than 58 000 human lncRNA genes, Only a few lncRNAs have been functionally characterized, such as H19, HOTAIR and Malat, Most of them are still functionally uncharacterized [10].

LncRNAs participate in the regulation of expressed proteins via specific mechanism involving multiple biological interactions such as lncRNA-mRNA, lncRNA-ncRNA and

lncRNA-protein interactions [11]. Therefore, it is necessary to construct a network of biomolecular interactions mediated by lncRNAs, which is very useful for revealing the underlying mechanisms and biological functions of lncRNAs [12]. As a bait for miRNA, lncRNAs can inhibit the binding of miRNA to target gene mRNA, and can also act as an endogenous miRNA sponge to inhibit miRNA expression [13, 14]. With the accumulation of knowledge on miRNA function, the lncRNA-miRNA interaction network can help us better understand the complex functions of lncRNA [7]. MiRNA is a class of non-coding short sequence RNAs of 18-25nt in length that are widely found in eukaryotes and are highly conservative, spatiotemporal-specific and tissue-specific [15]. A miRNA molecule can regulate the expression of up to 200 target genes, and about one-third of human genes are regulated by miRNAs [15]. Up to now, miRNAs are considered to be the most important gene regulators in cell differentiation, development, growth, and tumorigenesis, progression, metastasis, and drug resistance [16]. In the occurrence and development of human tumors, some miRNAs can act as both an oncogenes and a cancer suppressor genes [17, 18]. For example, certain miRNAs are associated with the development of ductal carcinoma in situ (DCIS) to invasive carcinoma, especially miR-210, miR221 and let-7d, which are down-regulated in situ carcinoma but up-regulated in invasive carcinoma [17].

In recent years, More and more studies have shown that both lncRNA and miRNA play critical roles in various biological processes and human complex diseases [19, 20]. It has been systematically studied that the lncRNA-miRNA interactions exert regulation role in some human complex diseases [21-23]. In many diseased cells, lncRNA is discovered to have a certain quantitative relationship with some miRNAs, this quantitative relationship is closely associated with the occurrence and development of diseases [24]. For example, in the renal cell carcinoma (RCC), after Malta silencing, miR-205 expression is up-regulated, but cell proliferation, migration and invasion are inhibited. Conversely, the expression of Malta is significantly reduced after overexpression of miR-205，experiments showed that there is a mutual regulation relationship between Malta and miR-205 [25]. The detailed understanding of interactions between lncRNAs and miRNAs in disease is very helpful for new biomarker discovery and treatment methods exploration [26].

However，identifying the interactions between lncRNAs and miRNAs is expensive and time-consuming by biological experiment.

To accelerate the process of identifying interactions between biomolecules, many computational methods have been proposed and effectively used for predicting relationships (e.g. miRNA-disease associations, protein-protein interactions and lncRNA-protein interactions), including manifold learning, manifold embedding and semi-supervised learning, etc. [27-29]. The computational methods for predicting miRNA-target interactions usually have the following common rules, including site accessibility, seed matching, free energy and protection [10, 30]. However, many miRNA-target identification methods were proposed originally for mRNA targets that may not be able to identify the interactions between lncRNAs and miRNAs, or even contradictory [31, 32]. Huang and Chan proposed the EPLMI calculation model based on the assumption that lncRNAs tend to interact collaboratively with miRNAs of similar expression profiles, and constructs bipartite graph via known lncRNA-miRNA interactions for prediction[33]. Huang *et al.* proposed the GBCF computational model based on known interaction network to obtain a top-*k* probability ordering list of individual lncRNA or miRNA for prediction [34]. Although the above two methods have better predictive effects in the known lncRNA-miRNA interaction network, they cannot be applied to new lncRNA or miRNA. The predictive effect of interactions between molecules can be improved by integrating biological information from different sources [35-37]. In fact, prediction of lncRNA-miRNA interactions can be considered as a recommender system problem [38, 39]. Accumulated studies have shown that matrix factorization is an effective method which has been successfully used in recommender system for data representation，and already widely applied in the field of bioinformatics [40-42].

In this paper, we propose a new calculation model, GNMFLMI, to predict lncRNA-miRNA interactions using graph regularized nonnegative matrix factorization (NMF) [43]. The model is based on the assumption that functionally similar lncRNAs (or miRNAs) are more possibility to interact with a same miRNA (or lncRNA) [44] . GNMFLMI fully exploits miRNA/lncRNA sequence information and known lncRNA-miRNA interaction network to calculate miRNA/lncRNA similarity. Subsequently, the graph spaces of miRNA

and lncRNA were constructed based on the local invariance hypothesis of intrinsic geometric space [45-47], which promote similar lncRNAs/miRNAs to be close enough to each other in the lncRNA/miRNA space. We evaluated the performance of our method by five-fold cross validation and compared the performance with standard NMF [43], RNMF [48]. The experiment results show that GNMFLMI is better than other compared methods, and can effectively predict the novel lncRNA-miRNA interactions.

2 Materials and methods

2.1 Benchmark Dataset

The lncRNA-miRNA interactions dataset used in this work were obtained from the lncRNASNP2 database in January, 2019. This dataset is collected and collated by Ya-Ru Miao *et al.* [49] and is accessible to academic users at http://bioinfo.life.hust.edu.cn/lncRNASNP. We downloaded the known lncRNA-miRNA interactions and the duplicated entries were removed. After the preprocessing, 8634 experimentally verified lncRNA–miRNA interactions were obtained, containing 262 miRNAs and 468 lncRNAs. In the experiment, all known lncRNA-miRNA interactions provided by lncRNASNP2 dataset were used as positive samples, and other unknown lncRNA-miRNA interactions as negative samples. The lncRNA-miRNA interactions adjacency matrix $Y \in R^{r \times n}$ was constructed based on lncRNASNP2 database, where r is the number of lncRNAs, n is the number of miRNAs. If the lncRNA $l(i)$ was verified to be interacted with miRNA $m(j)$, the element $Y(i,j)$ was assigned the value of 1, otherwise it is 0.

In this study ,we let $L = \{l_1, l_2, \cdots, l_r\}$ and $M = \{m_1, m_2, \cdots, m_n\}$ which denote the set of r lncRNAs and n miRNAs. The ith row vector of matrix Y, $Y(l_i) = (Y_{i1}, Y_{i2}, \cdots, Y_{in})$; the jth column vector of matrix Y, $Y(m_j) = (Y_{1j}, Y_{2j}, \cdots, Y_{rj})$. $Y(l_i)$ and $Y(m_j)$ represent the interaction profiles for lncRNA l_i and miRNA m_j, respectively.

2.2 Related work

2.2.1 The standard nonnegative matrix factorization (NMF)

Nonnegative matrix factorization (NMF) is an effective algorithm which has been successfully used in recommender system for data representation [41]. This algorithm

divides a large original matrix Y into two low-dimensional nonnegative matrices: a basis matrix U and a coefficient matrix V, and make their product as close as possible to the original matrix. Recent years, NMF has been successfully utilized to predict potential associations of lncRNA-protein [50], miRNA-disease [51], Microbe-disease [52], drug-target [53], CircRNA-disease [54], etc. In this work, the lncRNA-miRNA interaction adjacency matrix $Y \in R^{468\times262}$ is divided into $U \in R^{k\times468}$ and $V \in R^{k\times262}$, k is the sub-space dimensionality ($k < rn/(r+n)$). So that:

$$Y \cong U^T V \qquad s.t.\ U \geq 0, V \geq 0 \tag{1}$$

The objective function for predicting lncRNA-miRNA interactions can be formulated as following:

$$\min_{U,V} \|Y - U^T V\|_F^2 \qquad s.t.\ U \geq 0,\ V \geq 0 \tag{2}$$

Where, $\|\cdot\|_F$ denotes the Frobenius norm. $U,\ V \geq 0$ means that all elements of U and V are nonnegative. According to matrix properties $\|A\|_F^2 = Tr(A^T A)$, $Tr(A^T) = Tr(A)$, and $Tr(AB) = Tr(BA)$, we can obtain:

$$\begin{aligned} \|Y - U^T V\|_F^2 &= Tr\big((Y - U^T V)(Y - U^T V)^T\big) \\ &= Tr(YY^T) - 2Tr(YV^T U) + Tr(U^T V V^T U) \end{aligned} \tag{3}$$

Where $Tr(\cdot)$ is the trace of a matrix.

Lee and Seung [43] propose the nonnegative matrix factorization algorithm and this algorithm is based on the multiplicative update rules of U and V, the update rules are as follows:

$$u_{ki} \leftarrow u_{ki} \frac{(VY^T)_{ki}}{(VV^T U)_{ki}} \tag{4}$$

$$v_{kj} \leftarrow v_{kj} \frac{(UY)_{kj}}{(UU^T V)_{kj}} \tag{5}$$

2.2.2 Constrained nonnegative matrix factorization (CNMF)

Due to the standard NMF cannot ensure the U and V smoothness. We can use the Tikhonov (L_2) in the standard NMF to solve this problem [55]. Pauca *et al.* proposed the following constrained nonnegative matrix factorization (CNMF) formulation [48]:

$$\min_{U,V} \|Y - U^T V\|_F^2 + \beta(\|U\|_F^2 + \|V\|_F^2) \qquad s.t.\ U \geq 0,\ V \geq 0 \tag{6}$$

where β is the sparseness constraint coefficient which is used adjust the sparsity of U

and V via L_2-norm. The Eq. (6) can be written as:

$$\|Y - U^T V\|_F^2 + \beta(\|U\|_F^2 + \|V\|_F^2)$$
$$= Tr(YY^T) - 2Tr(YV^T U) + Tr(U^T VV^T U) + \beta Tr(UU^T) + \beta Tr(VV^T) \quad (7)$$

The multiplicative update rules of U and V are as follows:

$$u_{ki} \leftarrow u_{ki} \frac{(VY^T)_{ki}}{(VV^T U + \beta U)_{ki}} \quad (8)$$

$$v_{kj} \leftarrow v_{kj} \frac{(UY)_{kj}}{(UU^T V + \beta V)_{kj}} \quad (9)$$

the details of multiplicative update rules are in section 2.3.5.

2.3 Graph regularized nonnegative matrix factorization for predicting lncRNA-miRNA interactions (GNMFLMI)

2.3.1 methods overview

In this study， we propose a new calculation model, GNMFLMI, to predict lncRNA-miRNA interactions. The GNMFLMI can be summarized into three steps, and its framework is shown in Figure 1. First, the similarity matrices of lncRNA and miRNA are calculated based on lncRNA/miRNA sequence information and known lncRNA-miRNA interaction network. Second, we construct the affinity graphs for lncRNAs and miRNAs using *p*-nearest neighbors. Finally, graph regularized nonnegative matrix factorization is performed to calculate the lncRNA-miRNA interaction scores.

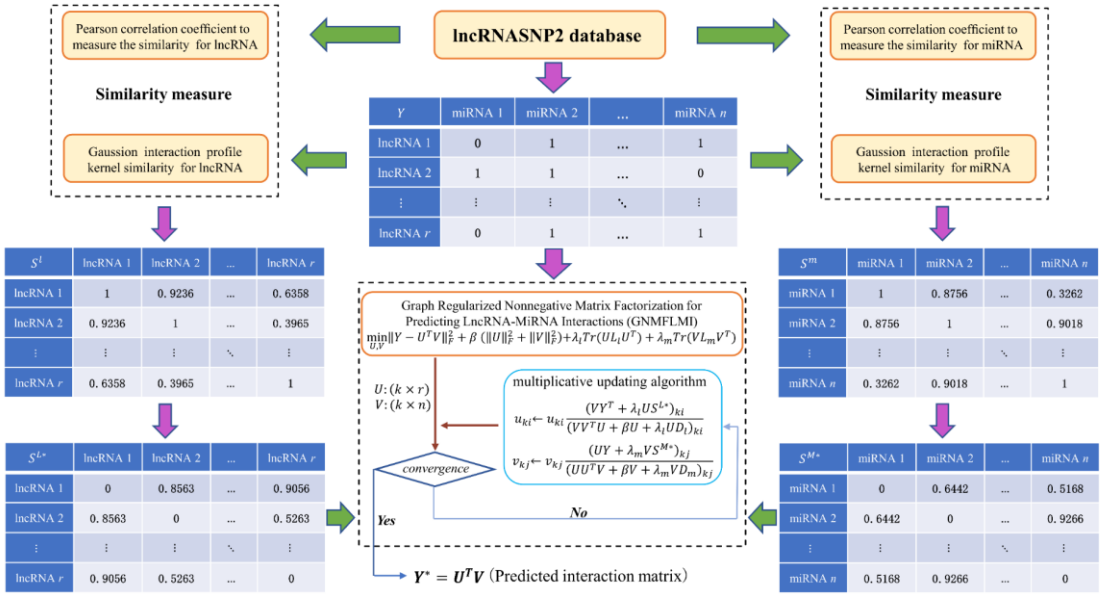

Fig. 1. Flowchart of GNMFLMI for predicting potential lncRNA-miRNA interaction.

2.3.2 Similarity measure

In our method, we need to construct the similarity matrices for lncRNAs and miRNAs separately， so the similarity both each pair of lncRNA-lncRNA and each pair of miRNA-miRNA need to be determined. In this study, two different types of lncRNA/miRNA similarity were measured via integrating diverse sources of information. The first type of lncRNA/miRNA similarity is calculated using Gaussian interaction profile (GIP) kernel based on known lncRNA-miRNA interaction network [56]. The second type is calculated using Pearson correlation coefficient (PCC) based on lncRNA/miRNA sequence information [57]. Subsequently, the overall similarity matrices for lncRNAs and miRNAs were constructed based on above two types of similarity.

Similarity based on Gaussian interaction profile (GIP) kernel. According to the assumption that functionally similar lncRNAs tend to interact with the similar miRNAs, and Gaussian interaction profile (GIP) kernel has been widely used to compute the molecule similarity [56, 58] . The Gaussian interaction profile kernel similarity of lncRNA and miRNA can be constructed according to the topologic information of known lncRNA-miRNA interaction network. Thus, we use GIP kernel to calculate the similarity $L_{GIP}(l_i, l_j)$ between lncRNA l_i and lncRNA l_j as following:

$$L_{GIP}(l_i, l_j) = \exp\left(-\gamma_l \left\| Y(l_i) - Y(l_j) \right\|^2\right) \tag{10}$$

where

$$Y = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{r,1} & y_{r,1} & \cdots & y_{r,n} \end{bmatrix} \tag{11}$$

$$\gamma_l = \frac{1}{\frac{1}{r}\sum_{i=1}^{r} \|Y(l_i)\|^2} \tag{12}$$

Y is the adjacent matrix of lncRNA-miRNA interaction based on lncRNASNP2 database. r and m are the number of lncRNAs and miRNAs, respectively. The size of L_{GIP} is $r \times r$, $Y(l_i)$ is the ith row of the adjacent matrix Y, γ_l is the kernel width parameter.

Similar to lncRNAs, the Gaussian interaction profile kernel similarity $M_{GIP}(m_i, m_j)$ of miRNA m_i and miRNA m_j can be calculated as follows:

$$M_{GIP}(m_i, m_j) = exp\left(-\gamma_m \left\| Y(m_i) - Y(m_j) \right\|^2\right) \tag{13}$$

where

$$\gamma_m = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\|Y(m_i)\|^2} \tag{14}$$

The size of M_{GIP} is $n \times n$, $Y(m_i)$ is the ith column of the adjacent matrix Y, γ_m is the kernel width parameter.

Similarity based on Pearson correlation coefficient (PCC). We download the expression profiles for lncRNAs and miRNAs from lncRNASNP2 database. For each lncRNA/miRNA, the values of expression profile can be obtained. Pearson correlation coefficient (PCC) has been widely applied to study expression profiles in bioinformatics [59, 60]. PCC of lncRNA/miRNA is calculated based on the lncRNA/miRNA expression profile values. For example, Given two lncRNAs l_i and l_j, the expression profiles are denoted as $X_l = \{x_{l1}, x_{l2}, \cdots, x_{lt}\}$ and $Z_l = \{z_{l1}, z_{l2}, \cdots, z_{lt}\}$. The similarity between lncRNA l_i and lncRNA l_j is calculated as follows:

$$L_{pcc}(l_i, l_j) = \frac{\left|\sum_{i=1}^{t}(x_{li} - \bar{X}_l)(z_{li} - \bar{Z}_l)\right|}{\sqrt{\sum_{i=1}^{t}(x_{li} - \bar{X}_l)^2}\sqrt{\sum_{i=1}^{t}(z_{li} - \bar{Z}_l)^2}} \tag{15}$$

where, t is the number of attributes of the expression profile, $\bar{X}_l$ and $\bar{Z}_l$ denote the average value of X_l and Z_l, respectively. Generally, the larger $L_{pcc}(l_i, l_j)$ represents the more similarly expression of lncRNA l_i and lncRNA l_j.

Similar to lncRNAs, the similarity between miRNA m_i and miRNA m_j can be calculated by Pearson correlation coefficient as follows.

$$M_{pcc}(m_i, m_j) = \frac{\left|\sum_{i=1}^{t}(x_{mi} - \bar{X}_m)(z_i - \bar{Z}_m)\right|}{\sqrt{\sum_{i=1}^{t}(x_{mi} - \bar{X}_m)^2}\sqrt{\sum_{i=1}^{t}(z_{mi} - \bar{Z}_m)^2}} \tag{16}$$

where, $X_m = \{x_{m1}, x_{m2}, \cdots, x_{mt}\}$ and $Z_m = \{z_{m1}, z_{m2}, \cdots, z_{mt}\}$ denote the expression profiles of miRNA m_i and miRNA m_j. $\bar{X}_m$ and $\bar{Z}_m$ denote the average value of X_m and Z_m, respectively.

Construct the overall similarity for lncRNAs and miRNAs. In this study, the Gaussian interaction profile kernel similarity and Pearson correlation coefficient similarity of lncRNA and miRNA are calculated, respectively. After that, the functional similarity between lncRNA l_i and lncRNA l_j is defined according to [61, 62], the final similarity matrix S^L of lncRNA is calculated as follows:

$$S^L(l_i, l_j) = \frac{L_{GIP}(l_i,l_j)+L_{pcc}(l_i,l_j)}{2} \quad (17)$$

where S^L is r-order square matrix, $S^L(l_i, l_j)$ represents the similarity score between lncRNA l_i and lncRNA l_j.

Based on the same method, the final similarity matrix S^M of miRNA is calculated as follows:

$$S^M(m_i, m_j) = \frac{M_{GIP}(m_i,m_j)+M_{pcc}(m_i,m_j)}{2} \quad (18)$$

where S^M is n-order square matrix, $S^M(m_i, m_j)$ represents the similarity score between miRNA m_i and miRNA m_j.

2.3.3 Sparsification of the similarity matrices

Recent studies on manifold learning theories and spectral graph have shown that the scattered nearest neighbors of data points can effectively model local geometric structure [53, 63]. In graph regularized matrix factorization, the nearest neighbor graph can promote close lncRNAs (or miRNAs) to be sufficiently close to each other in the lncRNA space (or miRNA space) [46, 64]. That is, it can preserve the local geometries of the original data. In this study, the affinity graphs (S^{L*} and S^{M*}) for lncRNAs and miRNAs are constructed using p-nearest neighbor graph, respectively. We set p=5, and the weight matrix G^l is generated based on the lncRNA similarity matrix S^L as follows:

$$G_{ij}^l = \begin{cases} 1 & i \in N_p(l_j) \,\&\, j \in N_p(l_i) \\ 0 & i \notin N_p(l_j) \,\&\, j \notin N_p(l_i) \\ 0.5 & otherwise \end{cases} \quad (19)$$

where $N_p(l_i)$ and $N_p(l_j)$ denote the sets of p-nearest neighbors to lncRNA l_i and lncRNA l_j, respectively. Subsequently, the sparse similarity matrix S^{L*} for lncRNAs is defined as:

$$\forall\, i, j, \qquad S_{ij}^{L*} = S_{ij}^L G_{ij}^l \quad (20)$$

The same procedure for miRNAs, the sparse similarity matrix S^{M*} can be obtained by the similarity matrix S^M of miRNA.

2.3.4 The model of GNMFLMI

In the Euclidean space, the standard nonnegative matrix factorization in Eq. (2) fails

to discover the intrinsic geometrical and discriminating structure of the data space [65, 66]. To avoid overfitting and enhance generalization capability, we use the Tikhonov (L_2) regularization in Eq. (2) to guarantee the U and V smoothness (i.e. Eq. (6)) [55]. At the same time, graph regularization is used to ensure that the relative positions of data points in the lncRNA feature space or miRNA feature space are unchanged [46]. The objective function of graph regularized nonnegative matrix factorization can be defined as follows:

$$\min_{U,V} \|Y - U^T V\|_F^2 + \lambda_l \sum_{i \le j}^{r} \left\|u_i - u_j\right\|^2 S_{ij}^{L*} + \lambda_m \sum_{i \le j}^{n} \left\|v_i - v_j\right\|^2 S_{ij}^{M*} + \beta \left(\|U\|_F^2 + \|V\|_F^2\right) \qquad s.t.\ U \ge 0,\ V \ge 0 \tag{21}$$

and

$$\begin{aligned} R_l = \sum_{i \le j}^{r} \left\|u_i - u_j\right\|^2 S_{ij}^{L*} &= \sum_{j=1}^{r} u_j^T u_j \sum_{i,j=1}^{r} S_{ij}^{L*} - \sum_{i,j=1}^{r} u_i^T u_j S_{ij}^{L*} \\ &= \sum_{j=1}^{r} u_j^T u_j D_{jj} - \sum_{i,j=1}^{r} u_i^T u_j S_{ij}^{L*} \\ &= Tr(UD_l U^T) - Tr(US^{L*}U^T) = Tr(UL_l U^T) \end{aligned} \tag{22}$$

Similarly,

$$R_m = \sum_{i \le j}^{n} \left\|v_i - v_j\right\|^2 S_{ij}^{M*} = Tr(VL_m V^T) \tag{23}$$

where λ_l and λ_m are the graph regularization parameters, u_i and v_j denote the ith and jth columns of U and V, respectively. R_l and R_m are the graph regularization terms of lncRNA and miRNA. We hope that if two data points l_i and l_j are close, u_i and u_j are also close to each other by minimizing R_l (or R_m). $L_l = D_l - S^{L*}$ and $L_m = D_m - S^{M*}$ represent the graph Laplacian matrices for S^{L*} and S^{M*} [67], respectively; D_l and D_m are the diagonal matrices whose diagonal elements are column (or row) sums of S^{L*} and S^{M*}, respectively. The Eq. (21) can be rewritten as:

$$\min_{U,V} \|Y - U^T V\|_F^2 + \beta \left(\|U\|_F^2 + \|V\|_F^2\right) + \lambda_l Tr(UL_l U^T) + \lambda_m Tr(VL_m V^T) \qquad s.t.\ U \ge 0,\ V \ge 0 \tag{24}$$

According to trace properties of matrix, the objective function Eq. (24) can be transformed into:

$$\|Y - U^T V\|_F^2 + \beta \left(\|U\|_F^2 + \|V\|_F^2\right) + \lambda_l Tr(UL_l U^T) + \lambda_m Tr(VL_m V^T)$$

$$= Tr\left((Y - U^TV)(Y - U^TV)^T\right) + \beta\left(Tr(UU^T) + Tr(VV^T)\right)$$
$$+\lambda_l Tr(UL_lU^T) + \lambda_m Tr(VL_mV^T)$$
$$= Tr(YY^T) - 2Tr(YV^TU) + Tr(U^TVV^TU) + \beta Tr(U^TU) + \beta Tr(V^TV)$$
$$+\lambda_l Tr(UL_lU^T) + \lambda_m Tr(VL_mV^T) \quad (25)$$

2.3.5 Model optimization

Because of the objective functions of GNMFLMI is not convex, finding the global minima is unrealistic by optimization algorithm. However, the local minima can be achieved via algorithm. In this study, the Lagrange Multiplier method was introduced to solve the optimization problem in Eq. (25). Let $\psi = \{\varphi_{ki}\}$ and $\Phi = \{\phi_{kj}\}$, the Lagrange multipliers φ_{ki} and ϕ_{kj} are used to restrict the $u_{ki} \geq 0$ and $v_{kj} \geq 0$, respectively. The Lagrange function $\mathcal{H}$ can be constructed as:

$$\mathcal{H} = Tr(YY^T) - 2Tr(YV^TU) + Tr(U^TVV^TU) + \beta Tr(U^TU) + \beta Tr(V^TV)$$
$$+\lambda_l Tr(UL_lU^T) + \lambda_m Tr(VL_mV^T) + Tr(\psi U) + Tr(\Phi V) \quad (26)$$

The partial derivatives of the function $\mathcal{H}$ to U and V are:

$$\frac{\partial \mathcal{H}}{\partial u} = -2VY^T + 2VV^TU + 2\beta U + 2\lambda_l UL_l + \psi \quad (27)$$

$$\frac{\partial \mathcal{H}}{\partial v} = -2UY + 2UU^TV + 2\beta V + 2\lambda_m VL_m + \Phi \quad (28)$$

According to the Karush–Kuhn–Tucker (*KKT*) conditions [68], $\varphi_{ki}u_{ki} = 0$ and $\phi_{kj}v_{kj} = 0$, we get the following equations for u_{ki} and v_{kj}:

$$-(VY^T)_{ki}u_{ki} + (VV^TU)_{ki}u_{ki} + (\beta U)_{ki}u_{ki} + [\lambda_l U(D_l - S^{L*})]_{ki}u_{ki} = 0 \quad (29)$$

$$-(UY)_{kj}v_{kj} + (UU^TV)_{kj}v_{kj} + (\beta V)_{kj}v_{kj} + [\lambda_m V(D_m - S^{M*})]_{kj}v_{kj} = 0 \quad (30)$$

Finally, the updating rules can be determined as follows:

$$u_{ki} \leftarrow u_{ki}\frac{(VY^T + \lambda_l US^{L*})_{ki}}{(VV^TU + \beta U + \lambda_l UD_l)_{ki}} \quad (31)$$

$$v_{kj} \leftarrow v_{kj}\frac{(UY + \lambda_m VS^{M*})_{kj}}{(UU^TV + \beta V + \lambda_m VD_m)_{kj}} \quad (32)$$

We update the nonnegative matrices U and V according to Eq. (31) and Eq. (32) until convergence or reaching the iteration upper limit. Ultimately, the new prediction lncRNA-miRNA interaction matrix Y^* can be calculated by $Y^* = U^TV$. In general, the larger value of the element in prediction matrix Y^*, it is more likely that lncRNA/miRNA

interacts with the corresponding miRNA/lncRNA. That is, for each miRNA, we can sort the lncRNA in descending order according to the value of the element, the top ranked lnRNAs in each column of Y^* are more likely to be associated with the corresponding miRNA. The same is true for each lncRNA. Table 1 generalizes the procedure of GNMFLMI for predicting lncRNA-miRNA interactions.

Table 1. Algorithm description of graph regularized nonnegative matrix factorization for predicting lncRNA-miRNA interactions.

Algorithm 1: GNMFLMI Algorithm

Input: Adjacency matrix $Y \in R^{r\times n}$, subspace dimensionality k, neighborhood size p, sparseness constraint coefficients β and regularization coefficients λ_l, λ_m.
Output: Predicted association matrix Y^*.

1. randomly initialize two nonnegative matrices $U \in R^{k\times r}$ and $V \in R^{k\times n}$;
2. construct the similarity matrices S^L and S^M
   S^L and S^M ← Based on Gaussian interaction profile kernel similarity and Pearson correlation coefficient;
3. construct the weight matrix G^l and G^m using p-nearest neighbors;
4. $L = \{l_1, l_2, \cdots, l_r\}$, $M = \{m_1, m_2, \cdots, m_n\}$;
5. for each lncRNA $l_a \in L$ do
   LNN = *K Nearest Known Neighbors*(l_a, S^L, p);
   end for
6. for each disease $m_b \in D$ do
   MNN= *K Nearest Known Neighbors*(m_b, S^M, p);
   end for
7. Sparse similarity matrices $S_{ij}^{L*} = S_{ij}^L G_{ij}^l$, $S_{ij}^{M*} = S_{ij}^M G_{ij}^m$;
8. repeat
   update U and V by the following rules:

$$u_{ki} \leftarrow u_{ki} \frac{(VY^T + \lambda_l U S^{L*})_{ki}}{(VV^T U + \beta U + \lambda_l U D_l)_{ki}}$$

$$v_{kj} \leftarrow v_{kj} \frac{(UY + \lambda_m V S^{M*})_{kj}}{(UU^T V + \beta V + \lambda_m V D_m)_{kj}}$$

9. until convergence or reaching upper limit of the iteration
10. calculate the predicted association matrix $Y^* = U^T V$;
11. return Y^*.

3 Results and discussion

3.1 Experimental settings

In this study, to estimate the performance of GNMFLMI on predicting lncRNA-miRNA interactions, the five-fold cross validation experiments were performed on the lncRNASNP2 dataset and compare our method with the following approaches: NMF and RNMF. In the five-fold cross validation， we randomly divide 8634 known lncRNA-miRNA interaction samples into five equal subsets. For each cross validation experiment, one of the subsets was used as the test set and the other four subsets were used as the training set.

The receiver operating characteristics (ROC) curve and AUC (area under the receiver operating characteristics curve) are widely used to estimate the performance [69, 70]. A larger value of AUC represents the better prediction performance of model. The confusion matrix can be obtained by setting different thresholds, the sensitivity (Sen.) and specificity (Spe.) are calculated as:

$$Sen. = \frac{TP}{TP+FN} \tag{33}$$

$$Spe. = \frac{TN}{TN+FP} \tag{34}$$

where FN and TP represent the number of false negative samples and true positive samples, respectively; FP and TN represent the number of false positive samples and true negative samples, respectively. FPR is false positive rate (FPR=1-Spe.), TPR is true positive rate (TPR=Sen.). In addition, precision (Pre.), accuracy (Acc.) and F1-Score are also used as general measurements.

$$Pre. = \frac{TP}{TP+Fp} \tag{35}$$

$$Acc. = \frac{TN+TP}{TN+TP+FN+Fp} \tag{36}$$

$$F1-Score = \frac{2\times Pre.\times Sn}{Pre.+Sn} \tag{37}$$

In this paper, the parameters are chose based on the grid search. There are five parameters in our method: neighborhood size p, subspace dimensionality k, sparseness constraint coefficient β and graph regularization coefficients λ_l , λ_m. The parameter combinations were determined from the following ranges: $\{65,140\}$ for k, $\{0.0001,0.001,0.01\}$ for β and the ranges of $\lambda_l = \lambda_m \in \{0.001,0.01,0.1\}$. Based on

the studies of Cai *et al.* [46] and Li *et al.* [65], we set $p = 5$. Finally, the parameter values are $p = 5$, k=80, $\beta = 0.01$, $\lambda_l = \lambda_m = 0.1$.

3.2 Cross validation

We compared the performance of GNMFLMI with computational approaches NMF and CNMF on the lncRNASNP2 dataset. Figure 2, Figure 3 and Figure 4 plot ROC curves and calculate the average AUC values of NMF, CNMF and GNMFLMI, respectively. Table 2 lists the AUC values of GNMFLMI, CNMF and NMF under five-fold cross validation. GNMFLMI achieved the AUC value of 0.9769, which higher than the AUC values of NMF 0.9344 and CNMF 0.9510. The experiment results show that the prediction performance of GNMFLMI outperform the NMF and CNMF.

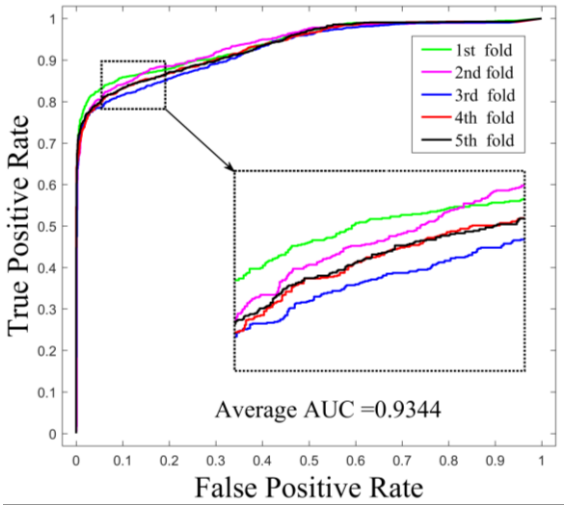

Fig. 2. The ROC curves of NMF in lncRNA-miRNA interaction prediction under 5-fold cross validation.

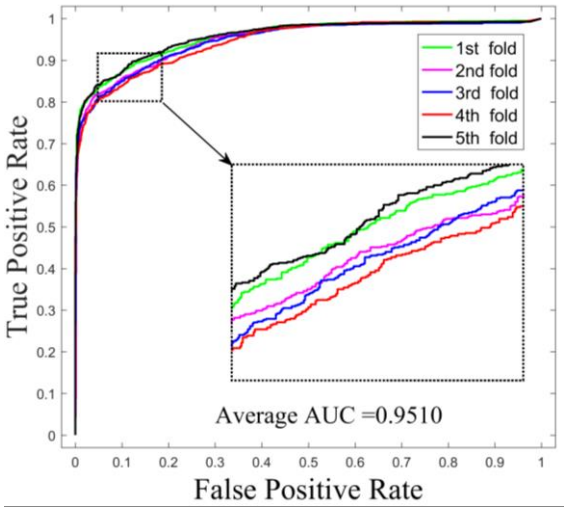

Fig. 3. The ROC curves of CNMF in lncRNA-miRNA interaction prediction under 5-fold cross validation.

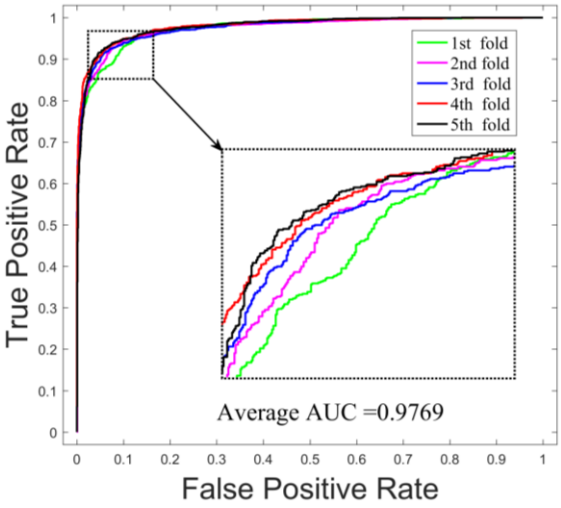

Fig. 4. The ROC curves of GNMFLMI in lncRNA-miRNA interaction prediction under 5-fold cross validation.

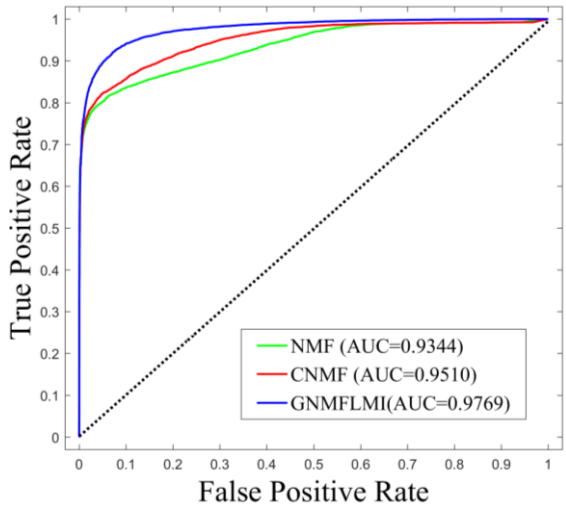

Fig. 5. The ROC curves of three methods.

In addition, the sensitivity, precision, accuracy and F1-Score for these methods were calculated at different specificity. As shown in Table 3, when specificity is 95%, the average sensitivities of GNMFLMI, NMF and CNMF are 89.40%, 80.17% and 82.08%, respectively. The sensitivity of GNMFLMI is 9.23% and 7.32% higher than NMF and CNMF. When specificity is 90%, GNMFLMI achieves the average sensitivity of 94.20%, which is still 10.63% and 8.47% higher than NMF and CNMF, respectively. In figure 5, we can also discover that the ROC curve of GNMFLMI is always above CNMF and NMF. These results further demonstrate that the performance of GNMFLMI is better than CNMF and NMF.

Table 2. The average AUC values and standard deviations obtained by various methods under five-fold cross validation (CV) on the lncRNASNP2 dataset.

Methods	The AUC values under five-fold CV					Average
	1st	2nd	3rd	4th	5th	
NMF	0.9399	0.9383	0.9267	0.9331	0.9341	0.9344±0.0052
CNMF	0.9562	0.9488	0.9476	0.9453	0.9572	0.9510±0.0054
GNMFLMI	0.9748	0.9752	0.9763	0.9799	0.9781	0.9769±0.0022

Table 3. The average sensitivity, precision, accuracy and F1-Score values of GNMFLMI and existing methods at different specificity.

Methods	Spe.(%)	Sen.(%)	Pre.(%)	Acc.(%)	F1-Score(%)
NMF	95.00	80.17	94.15	87.59	86.59
CNMF	95.00	82.08	94.28	89.46	88.83
GNMFLMI	95.00	89.40	94.72	92.21	91.98
NMF	90.00	83.57	89.29	86.78	86.33
CNMF	90.00	85.73	89.53	87.86	87.54
GNMFLMI	90.00	94.20	90.38	92.09	92.25

3.3 Case study

Case studies are carried out to further verify the capability of GNMFLMI on predicting novel interactions between lncRNAs and miRNAs. Here, we left out an arbitrary lncRNA-miRNA interaction (i.e. removing its interactions from the lncRNASNP2 dataset) to verify if its interactions would be discovered successfully. According to [71], the prediction performance is greatly affected by nearest neighbor information. If the new lncRNA l_i

(miRNA m_i) and lncRNA l_j (miRNA m_j) are close to each other, it may be easy to accurately predict the interactions between them, and vice versa. In this work, we select the lncRNA and miRNA which the similarity to the nearest neighbor is low (according to S^{L*} and S^{M*}, respectively) to validate the capability of model on predicting novel interactions. However, the standard NMF and CNMF fails to discover the novel interactions in this case.

Table 4. The top 20 novel interactions predicted by GNMFLMI for nonhsat-159254.1 on the lncRNASNP2 dataset.

Rank	miRNAs	Evidences
1	hsa-mir-590-3p	confirmed
2	hsa-mir-150-5p	confirmed
3	hsa-mir-374b-5p	confirmed
4	hsa-mir-374a-5p	confirmed
5	hsa-mir-30d-5p	confirmed
6	hsa-mir-30e-5p	confirmed
7	hsa-mir-30a-5p	confirmed
8	hsa-mir-30c-5p	confirmed
9	hsa-mir-30b-5p	confirmed
10	hsa-mir-200a-3p	confirmed
11	hsa-mir-141-3p	confirmed
12	hsa-mir-205-5p	confirmed
13	hsa-mir-216a-5p	confirmed
14	hsa-mir-4465	confirmed
15	hsa-mir-26a-5p	confirmed
16	hsa-mir-1297	confirmed
17	hsa-mir-26b-5p	confirmed
18	hsa-mir-363-3p	confirmed
19	hsa-mir-25-3p	confirmed
20	hsa-mir-92a-3p	confirmed

Table 5. The top 20 novel interactions predicted by GNMFLMI for hsa-mir-544a on the lncRNASNP2 dataset.

Rank	LncRNAs	Evidences
1	nonhsat137542.2	confirmed
2	nonhsat137541.2	confirmed
3	nonhsat137559.2	confirmed
4	nonhsat159243.1	confirmed
5	nonhsat137558.2	confirmed
6	nonhsat022125.2	confirmed
7	nonhsat159246.1	confirmed
8	nonhsat159242.1	confirmed
9	nonhsat159254.1	confirmed
10	**nonhsat096369.2**	**unconfirmed**
11	**nonhsat096376.2**	**unconfirmed**
12	**nonhsat096375.2**	**unconfirmed**
13	nonhsat022145.2	confirmed
14	nonhsat159244.1	confirmed
15	nonhsat159252.1	confirmed
16	**nonhsat198591.1**	**unconfirmed**
17	nonhsat159248.1	confirmed
18	nonhsat022132.2	confirmed
19	nonhsat007675.2	confirmed
20	nonhsat007699.2	confirmed

For lncRNA-nonhsat159254.1, we remove all miRNAs which interact with this lncRNA from the lncRNASNP2 dataset, the remaining known interactions are used to train the model of GNMFLMI. Then, all candidate miRNAs are sorted in descending order according to the predicted interaction scores. Table 4 gives the top 20 predicted interactions for nonhsat159254.1, the top 20 predicted interactions were verified by databases. The same procedure is performed for miRNA-hsa-mir-544a, Table 5 lists the top 20 predicted interactions for hsa-mir-544a, 16 out of the top 20 candidate lncRNAs were verified by databases.

It is worth noting that the nonhsat159254.1 and hsa-mir-544a prediction can be considered two difficult cases. Specifically, the similarities both nonhsat159254.1 and hsa-mir-544a to their nearest neighbors lncRNA and miRNA are as low as 0.1786 (according to S^{L*}) and 0.3658 (according to S^{M*}), respectively. Such low similarity makes it more difficult to predict the interactions between them. According to the above two cases, it is shown that GNMFLMI can effectively predict novel and challenging interactions between lncRNAs and miRNAs， which can provide valuable information for biological experiments

4 Conclusion

The interactions between lncRNAs and miRNAs constitute a complex molecular regulatory network, and studies have confirmed that their interactions are closely related to the occurrence and development of various diseases. Identifying lncRNA-miRNA interactions can help people better understand the complex disease mechanisms. In this paper, we propose a new method, GNMFLMI, for lncRNA-miRNA interaction prediction. Different from other traditional methods, GNMFLMI guides the matrix factorization via constructing graph Laplacian regularizations of lncRNAs and miRNAs, and uses Lagrange multipliers method to optimize the objective function. This method can also be applied into other similar association prediction (e.g. small molecular-miRNA and mRNA–protein associations). Five-fold cross validation and case studies were conducted to validate the performance of GNMFLMI， the experiment results show that our method outperforms the other compared methods and can identify potential lncRNA-miRNA interactions. We

believe that our approach is helpful for clinical research. In future work, more different biological information can be integrated to further improve prediction performance of model.

Acknowledgements

This work was supported in part by the NSFC Excellent Young Scholars Program, under Grant 61722212, in part by the National Natural Science Foundation of China, under Grants 61702444, 61572506, in part by the Pioneer Hundred Talents Program of Chinese Academy of Sciences, in part by the Chinese Postdoctoral Science Foundation, under Grant 2019M653804, in part by the West Light Foundation of the Chinese Academy of Sciences, under Grant 2018-XBQNXZ-B-008.

Competing interests

The authors declare that they have no competing interests.

References

1. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP: **A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?** *Cell* 2011, **146**(3):353-358.
2. Schaukowitch K, Kim T-K: **Emerging epigenetic mechanisms of long non-coding RNAs**. *Neuroscience* 2014, **264**:25-38.
3. Guttman M, Russell P, Ingolia NT, Weissman JS, Lander ES: **Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins**. *Cell* 2013, **154**(1):240-251.
4. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL: **RNA maps reveal new RNA classes and a possible function for pervasive transcription**. *Science* 2007, **316**(5830):1484-1488.
5. Mercer TR, Dinger ME, Mattick JS: **Long non-coding RNAs: insights into functions**. *Nature reviews genetics* 2009, **10**(3):155.
6. Kurihara M, Shiraishi A, Satake H, Kimura AP: **A conserved noncoding sequence can function as a spermatocyte-specific enhancer and a bidirectional promoter for a ubiquitously expressed gene and a testis-specific long noncoding RNA**. *Journal of molecular biology* 2014, **426**(17):3069-3093.
7. Chen X, You ZH, Yan GY, Gong DW: **IRWRLDA: improved random walk with restart for lncRNA-disease association prediction**. *Oncotarget* 2016, **7**(36):57919-57931.
8. Sun M, Kraus WL: **From discovery to function: the expanding roles of long noncoding RNAs in physiology and disease**. *Endocrine reviews* 2015, **36**(1):25-64.
9. Roberts TC, Morris KV, Weinberg MS: **Perspectives on the mechanism of transcriptional regulation by long non-coding RNAs**. *Epigenetics* 2014, **9**(1):13-20.
10. Quinn JJ, Chang HY: **Unique features of long non-coding RNA biogenesis and function**. *Nature Reviews Genetics* 2016, **17**(1):47.
11. Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H: **starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data**. *Nucleic acids research* 2013, **42**(D1):D92-D97.
12. You Z, Lei Y, Ji Z, Zhu Z: **A novel approach to modelling protein-protein interaction networks**. In: *Advances in Swarm Intelligence.* Springer Berlin Heidelberg; 2012: 49-57.
13. Tay Y, Rinn J, Pandolfi PP: **The multilayered complexity of ceRNA crosstalk and competition**. *Nature* 2014, **505**(7483):344.
14. Guo G, Kang Q, Zhu X, Chen Q, Wang X, Chen Y, Ouyang J, Zhang L, Tan H, Chen R: **A long noncoding RNA critically regulates Bcr-Abl-mediated cellular transformation by acting as a**

competitive endogenous RNA. *Oncogene* 2015, **34**(14):1768.

15. Sui W, Lin H, Peng W, Huang Y, Chen J, Zhang Y, Dai Y: **Molecular dysfunctions in acute rejection after renal transplantation revealed by integrated analysis of transcription factor, microRNA and long noncoding RNA**. *Genomics* 2013, **102**(4):310-322.
16. Augoff K, McCue B, Plow EF, Sossey-Alaoui K: **miR-31 and its host gene lncRNA LOC554202 are regulated by promoter hypermethylation in triple-negative breast cancer**. *Molecular cancer* 2012, **11**(1):5.
17. Di Leva G, Cheung DG, Croce CM: **miRNA clusters as therapeutic targets for hormone-resistant breast cancer**. *Expert review of endocrinology & metabolism* 2015, **10**(6):607-617.
18. Huang Y-A, You Z-H, Li L-P, Huang Z-A, Xiang L-X, Li X-F, Lv L-T: **EPMDA: an expression-profile based computational model for microRNA-disease association prediction**. *Oncotarget* 2017, **8**(50):87033.
19. Chen X, Yan CC, Zhang X, You ZH: **Long non-coding RNAs and complex diseases: from experimental results to computational models**. *Briefings in Bioinformatics* 2016, **18**(4):558.
20. Wang L, You Z-H, Chen X, Li Y-M, Dong Y-N, Li L-P, Zheng K: **MTRDA: Using logistic model tree to predict miRNA-disease associations by fusing multi-source information of sequences and similarities**. *PLOS Computational Biology* 2019, **15**(3):e1006865.
21. Yang G, Lu X, Yuan L: **LncRNA: a link between RNA and cancer**. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 2014, **1839**(11):1097-1109.
22. You Z-H, Wang L-P, Chen X, Zhang S, Li X-F, Yan G-Y, Li Z-W: **PRMDA: personalized recommendation-based MiRNA-disease association prediction**. *Oncotarget* 2017, **8**(49):85568.
23. Yi H-C, You Z-H, Zhou X, Cheng L, Li X, Jiang T-H, Chen Z-H: **ACP-DL: A Deep Learning Long Short-Term Memory Model to Predict Anticancer Peptides Using High Efficiency Feature Representation**. *Molecular Therapy - Nucleic Acids* 2019.
24. Yoon J-H, Abdelmohsen K, Gorospe M: **Functional interactions among microRNAs and long noncoding RNAs**. In: *Seminars in cell & developmental biology: 2014*. Elsevier: 9-14.
25. Hirata H, Hinoda Y, Shahryari V, Deng G, Nakajima K, Tabatabai ZL, Ishii N, Dahiya R: **Long noncoding RNA MALAT1 promotes aggressive renal cell carcinoma through Ezh2 and interacts with miR-205**. *Cancer research* 2015, **75**(7):1322-1331.
26. You Z, Wang S, Gui J, Zhang S: **A Novel Hybrid Method of Gene Selection and Its Application on Tumor Classification**. In: *Advanced Intelligent Computing Theories and Applications With Aspects of Artificial Intelligence*. Springer Berlin Heidelberg; 2008: 1055-1068.
27. You Z-H, Yin Z, Han K, Huang D-S, Zhou X: **A semi-supervised learning approach to predict synthetic genetic interactions by combining functional and topological properties of functional gene network**. *Bmc Bioinformatics* 2010, **11**(1):343.
28. Lei Y-K, You Z-H, Ji Z, Zhu L, Huang D-S: **Assessing and predicting protein interactions by combining manifold embedding with multiple information integration**. *BMC bioinformatics* 2012, **13**(Suppl 7):S3.
29. Chen Z-H, Li L-P, He Z, Zhou J-R, Li Y, Wong L: **An Improved Deep Forest Model for Predicting Self-Interacting Proteins From Protein Sequence Using Wavelet Transformation**. *Frontiers In Genetics* 2019, **10**.
30. Chen X, Zhang D-H, You Z-H: **A heterogeneous label propagation approach to explore the potential associations between miRNA and disease**. *Journal of translational medicine* 2018, **16**(1):348.
31. Li Y, Chen J, Zhang J, Wang Z, Shao T, Jiang C, Xu J, Li X: **Construction and analysis of lncRNA-lncRNA synergistic networks to reveal clinically relevant lncRNAs in cancer**. *Oncotarget* 2015, **6**(28):25003.
32. Liu B, Fang L, Liu F, Wang X, Chen J, Chou K-C: **Identification of real microRNA precursors with a pseudo structure status composition approach**. *PloS one* 2015, **10**(3):e0121501.
33. Huang Y-A, Chan KC, You Z-H: **Constructing prediction models from expression profiles for large scale lncRNA–miRNA interaction profiling**. *Bioinformatics* 2017, **34**(5):812-819.
34. Huang Z-A, Huang Y-A, You Z-H, Zhu Z, Sun Y: **Novel link prediction for large-scale miRNA-lncRNA interaction network in a bipartite graph**. *BMC medical genomics* 2018, **11**(6):113.
35. Meng F-R, You Z-H, Chen X, Zhou Y, An J-Y: **Prediction of drug–target interaction networks from the integration of protein sequences and drug chemical structures**. *Molecules* 2017, **22**(7):1119.
36. Li ZW, You ZH, Chen X, Li LP, Huang DS, Yan GY, Nie R, Huang YA: **Accurate prediction of protein-protein interactions by integrating potential evolutionary information embedded in PSSM profile and discriminative vector machine classifier**. *Oncotarget* 2017, **8**(14):23638.
37. You Z-H, Li L, Yu H, Chen S, Wang S-L: **Increasing reliability of protein interactome by**

combining heterogeneous data sources with weighted network topological metrics. In: *Advanced Intelligent Computing Theories and Applications.* Springer Berlin Heidelberg; 2010: 657-663.

38. Luo X, Zhou M, Li S, Xia Y, You Z, Zhu Q, Leung H: **An efficient second-order approach to factorize sparse matrices in recommender systems**. *IEEE Transactions on Industrial Informatics* 2015, **11**(4):946-956.
39. Luo X, Zhou M, Li S, You Z, Xia Y, Zhu Q: **A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method**. *IEEE transactions on neural networks and learning systems* 2016, **27**(3):579-592.
40. Jiang X, Hu X, Xu W: **Microbiome data representation by joint nonnegative matrix factorization with laplacian regularization**. *IEEE/ACM transactions on computational biology and bioinformatics* 2015, **14**(2):353-359.
41. Luo X, Zhou M, Xia Y, Zhu Q: **An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems**. *IEEE Transactions on Industrial Informatics* 2014, **10**(2):1273-1284.
42. Luo X, Zhou M, Leung H, Xia Y, Zhu Q, You Z-H, Li S: **An Incremental-and-Static-Combined Scheme for Matrix-Factorization-Based Collaborative Filtering**. *IEEE Trans Automation Science and Engineering* 2016, **13**(1):333-343.
43. Lee DD, Seung HS: **Learning the parts of objects by non-negative matrix factorization**. *Nature* 1999, **401**(6755):788.
44. Wu X, Jiang R, Zhang MQ, Li S: **Network‐based global inference of human disease genes**. *Molecular systems biology* 2008, **4**(1).
45. He X, Niyogi P: **Locality preserving projections**. In: *Advances in neural information processing systems: 2004*. 153-160.
46. Cai D, He X, Han J, Huang TS: **Graph regularized nonnegative matrix factorization for data representation**. *IEEE transactions on pattern analysis and machine intelligence* 2010, **33**(8):1548-1560.
47. Zhu L, You Z-H, Huang D-S, Wang B: **t-LSE: A Novel Robust Geometric Approach for Modeling Protein-Protein Interaction Networks**. *PloS one* 2013, **8**(4):e58368.
48. Pauca VP, Piper J, Plemmons RJ: **Nonnegative matrix factorization for spectral data analysis**. *Linear algebra and its applications* 2006, **416**(1):29-47.
49. Miao Y-R, Liu W, Zhang Q, Guo A-Y: **lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs**. *Nucleic acids research* 2017, **46**(D1):D276-D280.
50. Zhang T, Wang M, Xi J, Li A: **LPGNMF: Predicting Long Non-coding RNA and Protein Interaction Using Graph Regularized Nonnegative Matrix Factorization**. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2018.
51. Xiao Q, Luo J, Liang C, Cai J, Ding P: **A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations**. *Bioinformatics* 2017, **34**(2):239-248.
52. Liu Y, Wang S-L, Zhang J-F: **Prediction of Microbe–Disease Associations by Graph Regularized Non-Negative Matrix Factorization**. *Journal of Computational Biology* 2018, **25**(12):1385-1394.
53. Ezzat A, Zhao P, Wu M, Li X-L, Kwoh C-K: **Drug-target interaction prediction with graph regularized matrix factorization**. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 2017, **14**(3):646-656.
54. Wei H, Liu B: **iCircDA-MF: identification of circRNA-disease associations based on matrix factorization**. *Briefings in bioinformatics* 2019.
55. Guan N, Tao D, Luo Z, Yuan B: **Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent**. *IEEE Transactions on Image Processing* 2011, **20**(7):2030-2048.
56. Huang Y-A, You Z-H, Chen X, Huang Z-A, Zhang S, Yan G-Y: **Prediction of microbe–disease association from the integration of neighbor and graph with collaborative recommendation model**. *Journal of translational medicine* 2017, **15**(1):209.
57. Wang L, You Z-H, Chen X, Li J-Q, Yan X, Zhang W, Huang Y-A: **An ensemble approach for large-scale identification of protein-protein interactions using the alignments of multiple sequences**. *Oncotarget* 2017, **8**(3):5149.
58. van Laarhoven T, Nabuurs SB, Marchiori E: **Gaussian interaction profile kernels for predicting drug–target interaction**. *Bioinformatics* 2011, **27**(21):3036-3043.
59. Li M, Zheng R, Zhang H, Wang J, Pan Y: **Effective identification of essential proteins based on priori knowledge, network topology and gene expressions**. *Methods* 2014, **67**(3):325-333.
60. Kang H, Lee J, Yu S: **Differential Co-Expression Networks using RNA-seq and microarrays in Alzheimer's disease**. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine*

(BIBM): 2016. IEEE: 1907-1908.

61. Chen X: **KATZLDA: KATZ measure for the lncRNA-disease association prediction**. *Scientific reports* 2015, **5**:16840.
62. Chen X, Yin J, Qu J, Huang L: **MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction**. *PLoS computational biology* 2018, **14**(8):e1006418.
63. You Z, Zhang S, Li L: **Integration of genomic and proteomic data to predict synthetic genetic interactions using semi-supervised learning**. In: *Emerging Intelligent Computing Technology and Applications With Aspects of Artificial Intelligence.* Springer Berlin Heidelberg; 2009: 635-644.
64. You Z-H, Lei Y-K, Gui J, Huang D-S, Zhou X: **Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data**. *Bioinformatics* 2010, **26**(21):2744-2751.
65. Li X, Cui G, Dong Y: **Graph regularized non-negative low-rank matrix factorization for image clustering**. *IEEE transactions on cybernetics* 2016, **47**(11):3840-3853.
66. Wang J-Y, Almasri I, Gao X: **Adaptive graph regularized nonnegative matrix factorization via feature selection**. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012): 2012*. IEEE: 963-966.
67. Liu X, Zhai D, Zhao D, Zhai G, Gao W: **Progressive image denoising through hybrid graph Laplacian regularization: A unified framework**. *IEEE Transactions on image processing* 2014, **23**(4):1491-1503.
68. Facchinei F, Kanzow C, Sagratella S: **Solving quasi-variational inequalities via their KKT conditions**. *Mathematical Programming* 2014, **144**(1-2):369-412.
69. Hanley JA, McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve**. *Radiology* 1982, **143**(1):29-36.
70. Fawcett T: **An introduction to ROC analysis pattern recognition letter**. 2006.
71. Wassermann AM, Geppert H, Bajorath Jr: **Ligand prediction for orphan targets using support vector machines and various target-ligand kernels is dominated by nearest neighbor effects**. *Journal of chemical information and modeling* 2009, **49**(10):2155-2167.