

RESEARCH

Open Access



# LncRNA-miRNA interaction prediction through sequence-derived linear neighborhood propagation method with information combination

Wen Zhang<sup>1\*</sup>, Guifeng Tang<sup>2</sup>, Shuang Zhou<sup>3</sup> and Yanqing Niu<sup>4\*</sup>

From IEEE International Conference on Bioinformatics and Biomedicine 2018  
Madrid, Spain. 3-6 December 2018

## Abstract

**Background:** Researchers discover lncRNAs can act as decoys or sponges to regulate the behavior of miRNAs. Identification of lncRNA-miRNA interactions helps to understand the functions of lncRNAs, especially their roles in complicated diseases. Computational methods can save time and reduce cost in identifying lncRNA-miRNA interactions, but there have been only a few computational methods.

**Results:** In this paper, we propose a sequence-derived linear neighborhood propagation method (SLNPM) to predict lncRNA-miRNA interactions. First, we calculate the integrated lncRNA-lncRNA similarity and the integrated miRNA-miRNA similarity by combining known lncRNA-miRNA interactions, lncRNA sequences and miRNA sequences. We consider two similarity calculation strategies respectively, namely similarity-based information combination (SC) and interaction profile-based information combination (PC). Second, the integrated lncRNA similarity-based graph and the integrated miRNA similarity-based graph are respectively constructed, and the label propagation processes are implemented on two graphs to score lncRNA-miRNA pairs. Finally, the weighted averages of their outputs are adopted as final predictions. Therefore, we construct two editions of SLNPM: sequence-derived linear neighborhood propagation method based on similarity information combination (SLNPM-SC) and sequence-derived linear neighborhood propagation method based on interaction profile information combination (SLNPM-PC). The experimental results show that SLNPM-SC and SLNPM-PC predict lncRNA-miRNA interactions with higher accuracy compared with other state-of-the-art methods. The case studies demonstrate that SLNPM-SC and SLNPM-PC help to find novel lncRNA-miRNA interactions for given lncRNAs or miRNAs.

**Conclusion:** The study reveals that known interactions bring the most important information for lncRNA-miRNA interaction prediction, and sequences of lncRNAs (miRNAs) also provide useful information. In conclusion, SLNPM-SC and SLNPM-PC are promising for lncRNA-miRNA interaction prediction.

**Keywords:** lncRNA-miRNA interactions, Integrated similarity, Label propagation

\* Correspondence: [zhangwen@mail.hzau.edu.cn](mailto:zhangwen@mail.hzau.edu.cn);  
[niuyanqing@mail.scuec.edu.cn](mailto:niuyanqing@mail.scuec.edu.cn)

<sup>1</sup>College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

<sup>4</sup>School of Mathematics and Statistics, South-Central University for Nationalities, Wuhan 430074, China

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



## Background

Non-coding RNAs (ncRNAs) are a class of RNAs that are not translated into functional proteins [1]. ncRNAs can be classified into many types, e.g. long non-coding RNA, circular RNA, snRNA, etc. Long non-coding RNAs (lncRNAs) are a kind of ncRNAs whose lengths are more than 200 nucleotides [2]. Studies [3, 4] show that a great number of lncRNAs are involved in many biological processes, such as cell proliferation, chromatin remodeling, gene imprinting and immune response. More importantly, some researchers discovered that lncRNAs are associated with severe diseases such as prostate cancer and gastric cancer [5–10].

lncRNAs play functional roles by interacting with other biological molecules (DNAs, RNAs and proteins), and the studies on lncRNA-biomolecule interactions help to characterize the functions of lncRNAs. For example, lncRNA loc285194 can interact with p53 gene and act as a tumor suppressor [11]; lncRNA PVT1 interacts with FOXM1 protein and promotes gastric cancer progression [12]. For a long time, researchers have been paying attention to lncRNA-DNA interactions [13, 14] or lncRNA-protein interactions [15, 16]. Recently, some researchers discover [17] that lncRNAs can act as decoys or sponges to regulate the behavior of miRNAs. For example, the lncRNA H19 is found to modulate let-7 family of miRNAs [18]. Therefore, exploring lncRNA-miRNA interactions contributes to understanding the complicated functions of lncRNAs.

Previous studies conduct wet experiments to identify lncRNA-miRNA interactions. For example, Amanda et al. [18] carry out *in vivo* crosslinking combined with affinity purification experiments to explore the interaction between lncRNA H19 and miRNA let-7. Based on the crosslinking and real-time PCR (RT-qPCR) experiment, their results demonstrated that lncRNA H19 can physically interact with let-7 *in vivo*. Zhang et al. [19] once studied the miRNA miR-7's function in breast cancer stem cell (BCSCs) and its associated lncRNA. By implementing ChIP-PCR and Double-Luciferase Reporter assay, they find that the downregulation of miR-7 in BCSCs might be indirectly attributed to lncRNA HOTAIR. The wet methods are time-consuming and labor-intensive; thus, it is important to perform *in silico* prediction to refine the candidate list for further validation experiments.

Recently, researchers introduce machine learning techniques into the lncRNA-biomolecule interaction prediction, especially the lncRNA-protein interaction [20–25]. However, only a few lncRNA-miRNA interaction prediction methods have been proposed. Huang et al. [26] propose a method named EPLMI, which relies on the assumption that lncRNAs having similar expression profiles are prone to associate with a cluster of miRNAs that have similar expression profiles. Huang et al. [27] develop a novel group preference Bayesian

collaborative filtering model called GBCF, which picks up a top-k probability ranking list for an individual miRNA or lncRNA based on known miRNA-lncRNA interaction network. Hu et al. [28] predict lncRNA-miRNA interactions by integrating the expression similarity network and the sequence similarity network, and develop a method named INLMI. Nevertheless, these methods have several limitations, which inspire us to develop better models. Firstly, existing methods rely on several features of lncRNAs and miRNAs, such as sequences, expression profiles and target genes, but expression profiles and target genes are not available for all lncRNAs (or miRNAs). Secondly, many lncRNAs and miRNAs do not have any known interaction, but a desirable model should be capable of predicting their interactions.

In this paper, we propose a sequence-derived linear neighborhood propagation method (SLNPM) to predict lncRNA-miRNA interactions. First, we calculate integrated lncRNA-lncRNA similarity and integrated miRNA-miRNA similarity by combining known lncRNA-miRNA interactions, lncRNA sequences and miRNA sequences. As the extension of our previous work [29], we consider two integrated similarity calculation strategies, namely similarity-based information combination (SC) and interaction profile-based information combination (PC). Second, the integrated lncRNA similarity-based graph and the integrated miRNA similarity-based graph are respectively constructed, and the label propagation processes are respectively implemented on two graphs to score lncRNA-miRNA pairs. Finally, the averages of their outputs are adopted as final predictions. In this way, we construct two editions of SLNPM based on similarity information combination (SLNPM-SC) and based on interaction profile information combination (SLNPM-PC). The experimental results show that SLNPM-SC and SLNPM-PC predict lncRNA-miRNA interactions with higher accuracy compared with other state-of-the-art methods. We also analyze the prediction capability of SLNPM-SC and SLNPM-PC for lncRNAs (or miRNAs) which do not have any known interaction, and the case studies demonstrate that SLNPM-SC and SLNPM-PC help to find novel interactions which do not exist in our dataset.

This paper makes the following contributions: (1) the proposed SLNPM models make use of diverse information to achieve high-accuracy performances; (2) the proposed SLNPM models can deal with the lncRNAs (or miRNAs) that do not have any known interaction.

## Datasets and methods

### Datasets

There are several datasets about lncRNAs, miRNAs and lncRNA-miRNA interactions, such as lncRNASNP [17],

NONCODE [30], miRBase [31] and miRmine [32]. LncRNASNP [17] contains experimentally validated lncRNA-related SNPs and lncRNA-miRNA interactions, which can facilitate to study lncRNAs' functions. NONCODE [30] is an integrated knowledge database of non-coding RNAs (ncRNAs). The ncRNA sequences and related information (e.g. function, cellular role, cellular location, chromosomal information, etc.) in NONCODE have been confirmed manually by consulting relevant literature. MiRBase [31] is a comprehensive database about miRNAs, containing published miRNA sequences and annotation. The database miRmine [32] provides high-quality human miRNA-Seq and miRNA expression profiles.

To compile our datasets, we first download data from LncRNASNP, and obtain 8091 experimentally verified lncRNA-miRNA interactions. After removing duplicated associations, there remain 5118 interactions between 780 lncRNAs and 275 miRNAs. Then, we collect lncRNA's sequences from NONCODE and collect miRNAs' sequences from miRbase. Thus, sequences are available for 642 lncRNAs and 275 miRNAs. Next, we obtain expression profiles of lncRNAs in 24 human tissues from NONCODE, and obtain expression profiles of miRNAs in 16 types of human tissues and 24 types of cell types from miRmine. The expression profiles are available for 417 lncRNAs and 265 miRNAs. Therefore, we compile a dataset named SLNPM-S by removing lncRNAs and miRNAs whose sequences or expression profiles are unavailable. Similarly, we compile a dataset named SLNPM-L by removing lncRNAs and miRNAs whose sequences are unavailable. SLNPM-S serves as the main dataset for model training and performance evaluation, and SLNPM-L is used for the case study. Table 1 summarizes the details of two datasets.

**Linear neighborhood similarity measure**

In previous work [33, 34], we proposed a novel similarity measure named linear neighborhood similarity (LNS), and successfully solved several problems in bioinformatics [24, 35–37]. In this paper, we adopt the linear neighborhood similarity measure (LNS) to calculate lncRNA-lncRNA similarity and miRNA-miRNA similarity. Here we first introduce the detailed process of LNS.

Given  $n$ -dimensional feature vectors  $x_1, x_2, \dots, x_m$ , these feature vectors are considered as the data points in the feature space. We concentrate the vectors row by row to obtain the  $n \times m$  matrix  $X$ , where  $x_i$  is the  $i$  th

row of the matrix  $X$ . It is assumed that each data point can be reconstructed by the linear weighted sum of neighboring data points. Generally, nearest neighbors based on the Euclidean distance are selected for each data point  $x_i$ , and the ratio of the neighbors (selected nearest neighbors vs all neighboring data points) is called neighborhood ratio, denoted by  $K$ .  $N(x_i)$  is the set of selected nearest neighbors of  $x_i$ . By minimizing the reconstructive errors for all data points, we present the following optimization problem:

$$\min_W \frac{1}{2} \|X - (C \odot W)X\|_F^2 + \frac{\mu}{2} \sum_{i=1}^m \|(C \odot W)e\|_2^2 \tag{1}$$

$$s.t. (C \odot W)e = e, W \geq 0$$

where  $C$  is an indicator matrix.  $C(i, j) = 1$  if  $x_j \in N(x_i)$ ; else  $C(i, j) = 0$ ;  $C(i, i) = 0$ .  $\|\cdot\|_F$  is the Frobenius-norm.  $e = (1, 1, \dots, 1)^T$ , and  $\odot$  is Hadamard product.  $\mu$  is the tradeoff parameter.  $W$  is a  $m \times m$  weight matrix, where the  $i$ th row indicates the data points' reconstruction contributions to the data point  $x_i$ .

To solve the objection function (1), we introduce the Lagrange function:

$$L = \frac{1}{2} \|X - (C \odot W)X\|_F^2 + \frac{\mu}{2} \|(C \odot W)e\|_2^2 - \lambda^T ((C \odot W)e - e) - \text{tr}(\Phi^T W) \tag{2}$$

where  $\Phi$  is Lagrange multiplier. Differentiating  $L$  with respect to  $W$ , we have:

$$\nabla_W L = C \odot ((C \odot W)XX^T + \mu(C \odot W)ee^T - XX^T - \lambda e^T) - \Phi^T$$

By Complementary slackness condition, we obtain:

$$((C \odot W)XX^T + \mu(C \odot W)ee^T - XX^T - \lambda e^T)_{ij} W_{ij} C_{ij} = 0$$

So  $W_{ij}$  can be written as:

$$W_{ij} = \begin{cases} \frac{W_{ij}(XX^T + \lambda e^T)_{ij}}{((C \odot W)XX^T + \mu(C \odot W)ee^T)_{ij}} & x_j \in N(x_i) \\ 0 & x_j \notin N(x_i) \end{cases} \tag{3}$$

But there still exists  $\lambda$  in (3), and (2) has the equivalent form:

**Table 1** Summary of SLNPM-S and SLNPM-L datasets

Dataset	LncRNAs	MiRNAs	Interactions	Features
SLNPM-S	417	265	2272	Sequences, Expression Profiles
SLNPM-L	642	275	3784	Sequences

$$\begin{aligned} \min_{\omega_i} L^i &= \frac{1}{2} \left\| x_i - \sum_{i_j: x_{i_j} \in N(x_i)} \omega_{i,i_j} \right\|^2 \\ &\quad + \frac{\mu}{2} \left( \sum_{i_j: x_{i_j} \in N(x_i)} |\omega_{i,i_j}| \right)^2 \\ &= \frac{1}{2} \omega_i^T G^i \omega_i + \frac{\mu}{2} \|\omega_i\|_1^2 \end{aligned} \tag{4}$$

$$s.t. e^T \omega_i = 1, \omega_i \geq 0$$

where  $G^i$  is the Gram Matrix whose entry is  $(x_i, x_{i_j})(x_i, x_{i_k})^T$ . The Lagrange function of (4) is:

$$L^i = \frac{1}{2} \omega_i^T G^i \omega_i + \frac{\mu}{2} \|\omega_i\|_1^2 - \lambda_i (e^T \omega_i - 1) - \eta^T \omega_i \tag{5}$$

By Karush–Kuhn–Tucker (KKT) conditions, we obtain:

$$\begin{cases} \nabla_{\omega_i} L^i = G^i \omega_i + \mu e e^T \omega_i - \lambda_i e - \eta = 0 \\ \nabla_{\lambda_i} L^i = e^T \omega_i - 1 = 0 \\ \eta \geq 0, \omega_i \geq 0, \eta_j \omega_{i,i_j} = 0 \end{cases}$$

Then, it can be inferred that:

$$\omega_i^T \nabla_{\omega_i} L^i = \omega_i^T G^i \omega_i + \mu (\omega_i^T e)^2 - \lambda_i \omega_i^T e = 0$$

So:

$$\lambda_i = \left( \omega_i^T G^i \omega_i + \mu (e^T \omega_i)^2 \right) / e^T \omega_i$$

The reconstruction error  $\frac{1}{2} \omega_i^T G^i \omega_i \approx 0$ . If  $\omega_i$  is the optimal solution for (5),  $e^T \omega_i - 1 = 0$ . So  $\lambda_i \approx \mu$ . Let  $\lambda = \mu e$ . Then we obtain:

$$W_{ij} = \begin{cases} \frac{W_{ij}(XX^T + \mu e e^T)_{ij}}{((C \odot W)XX^T + \mu(C \odot W)ee^T)_{ij}} & x_j \in N(x_i) \\ 0 & x_j \notin N(x_i) \end{cases} \tag{6}$$

Weight matrix  $W$  is updated according to Eq. (6) until convergence.

### Sequence similarity and interaction profile similarity

In this section, we introduce mathematical notations for lncRNA (and miRNA) interaction profile, lncRNA (and miRNA) sequence similarity and lncRNA (and miRNA) interaction profile similarity. Given lncRNAs  $L_1, \dots, L_i, \dots, L_l$  and miRNAs  $M_1, \dots, M_j, \dots, M_m$ , their pairwise interactions are represented by a  $l \times m$  interaction matrix  $Y$ , where  $Y_{ij} = 1$  if the lncRNA  $L_i$  interacts with the miRNA  $M_j$ , otherwise  $Y_{ij} = 0$ . By using the interaction matrix  $Y$ , we define the interaction profiles for lncRNAs and miRNAs. The interaction profile of lncRNA  $L_i$  is a binary vector specifying the absence or presence of its interactions with every miRNA, and corresponds to the  $i$ th row of  $Y$ , namely  $Y(i, :)$ . The interaction profile of a

miRNA  $M_j$  is a binary vector encoding the absence or presence of its interactions with every lncRNA, and corresponds to the  $j$ th row of  $Y$ , namely  $Y(:, j)$ .

lncRNA sequences and miRNA sequences provide important information for exploring their functions, and the  $k$ -mer [38] is a popular sequence-derived feature, which describes repeated patterns of sequences. There exist four types of nucleotides i.e. A, C, G and T/U for both lncRNA sequences and miRNA sequences. For the  $k$ -mer feature, we count the frequencies of  $4^k$  types of  $k$ -length contiguous subsequences along lncRNA (miRNA) sequences. More specifically, for a lncRNA (or miRNA) sequence  $x$ , the  $k$ -mer feature of the sequence is defined as  $f_k(x) = (d_1, d_2, \dots, d_{4^k})$ , where  $d_i$  is the occurrence frequency of corresponding  $k$ -length contiguous subsequences. In this work, we set  $k = 5$ , and we present lncRNAs and miRNAs with their corresponding  $k$ -mer vectors. Then, we calculate sequence similarities for  $l$  lncRNAs, denoted as a  $l \times l$  matrix  $S_{LSF}$ , by using the linear neighborhood similarity measure (LNS). Similarly, we utilize LNS to calculate sequence similarities for  $m$  miRNAs, denoted as a  $m \times m$  matrix  $S_{MSF}$ .

Related studies [39–41] adopt biological molecules' interaction profiles in prediction models and achieve high-accuracy performance. These studies reveal the importance of interaction profiles in predicting unknown associations. Based on the interaction matrix  $Y$ , lncRNAs  $L_1, \dots, L_i, \dots, L_l$  are represented by interaction profiles  $Y(1, :), \dots, Y(i, :), \dots, Y(l, :)$ , and miRNAs  $M_1, \dots, M_j, \dots, M_m$  are represented by interaction profiles  $Y(:, 1), \dots, Y(:, j), \dots, Y(:, l)$ . Then, we can respectively calculate interaction profile similarities for  $l$  lncRNAs, denoted as a  $l \times l$  matrix  $S_{LIP}$ , using the linear neighborhood similarity measure; we calculate interaction profile similarities for  $m$  miRNAs, denoted as a  $m \times m$  matrix  $S_{MIP}$ .

### Sequence-derived linear neighborhood propagation method

Since we have the sequence feature and interaction profiles for lncRNAs (miRNAs), we integrate diverse information of lncRNAs (or miRNAs) to develop prediction models. On the one hand, information integration can lead to improved performances. On the other hand, there exist lncRNAs (miRNAs) that have no known interaction with any miRNA (lncRNA), and the interaction profiles are unavailable for these lncRNAs (miRNAs). The information integration can deal with such lncRNAs (miRNAs). Here, we propose a sequence-derived linear neighborhood propagation method (SLNPM) and consider two strategies: similarity-based information combination (SC) and interaction profile-based information combination (PC) to integrate diverse features and meanwhile address above-mentioned

problems. Thus, we present two editions of SLNPM: sequence-derived linear neighborhood propagation method based on similarity information combination (SLNPM-SC) and sequence-derived linear neighborhood propagation method based on interaction profile information combination (SLNPM-PC). The flowchart of two prediction models is shown in Fig. 1.

**Similarity-based information combination**

In this section, we propose the similarity-based information combination strategy to build the sequence-derived linear neighborhood propagation model, abbreviated as SLNPM-SC.

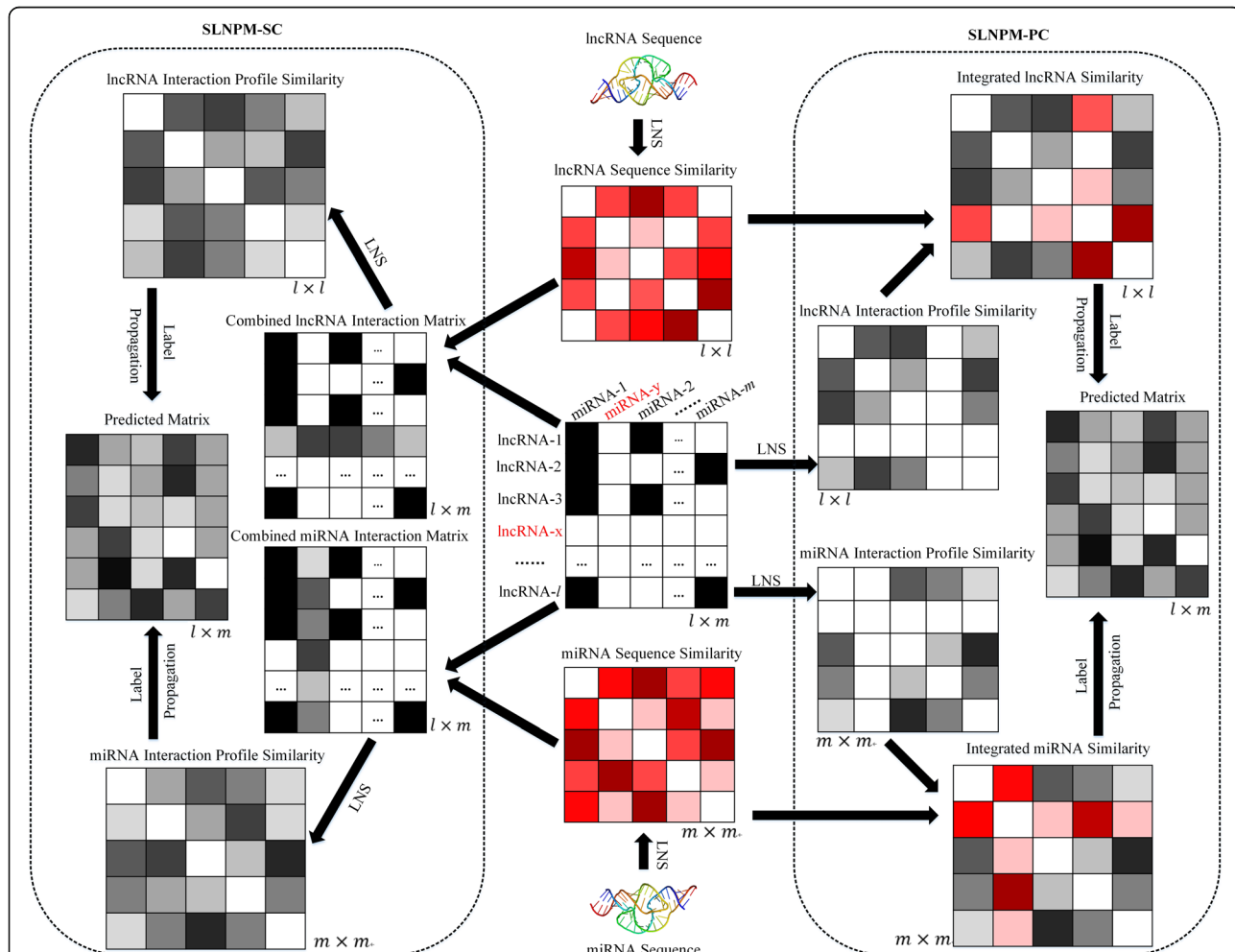
For a lncRNA  $L_i$  (miRNA  $M_j$ ), which has no interaction with any miRNA (lncRNA), its interaction profile is an all-zero vector. We cannot calculate the interaction

profile similarities for lncRNAs (miRNAs) without interactions. Therefore, entries in the  $i$  th ( $j$  th) row and  $i$  th ( $j$  th) column of the lncRNA (miRNA) interaction profile similarity matrix  $S_{LIP}$  ( $S_{MIP}$ ) are all zeros. The similarity-based information combination strategy is described below.

First, we calculate the sequence similarity  $S_{LSF}$  for all lncRNAs, and calculate the interaction profile similarity  $S_{LIP}$  for lncRNAs with interaction information. Then, we calculate the integrated similarity  $S_{LIS}$  for lncRNAs by:

$$S_{LIS}(i,:) = \begin{cases} S_{LIP}(i,:) & \text{if } L_i \text{ has interactions} \\ S_{LSF}(i,:) & \text{otherwise} \end{cases} \quad (7)$$

Similarly, we calculate the sequence similarity  $S_{MSF}$  for all miRNAs, and calculate the interaction profile similarity  $S_{MIP}$  for miRNAs with interaction information. Then,



**Fig. 1** Workflow of the sequence-derived linear neighborhood propagation method. The figure explains two models: SLNPM-SC and SLNPM-PC. SLNPM-SC integrates sequence similarity and interaction profile similarity to obtain combined similarities, and then makes predictions based on the combined similarities; SLNPM-PC utilizes the sequence similarities to complement the interaction profiles, and then calculates the interaction profile similarity to make predictions

we calculate the integrated similarity  $S_{MIS}$  for miRNAs by:

$$S_{MIS}(j, :) = \begin{cases} S_{MIP}(j, :) & \text{if } M_j \text{ has interactions} \\ S_{MSF}(j, :) & \text{otherwise} \end{cases} \quad (8)$$

Then, we construct a directed graph based on the integrated lncRNA similarity matrix  $S_{LIS}$ , and construct another directed graph based on the integrated miRNA similarity matrix  $S_{MIS}$ . Considering miRNA  $M_j$ , the  $j$ th column of interaction matrix  $Y$  is regarded as the initial labels of all nodes (lncRNAs) in the integrated lncRNA similarity-based graph. The label information is iteratively propagated in the graph until convergence, and the details about label propagation can refer to [42]. The prediction matrix  $P^l$  with size  $l \times m$  is obtained. Similarly, considering lncRNA  $L_i$ , the  $i$ th row of interaction matrix  $Y$  is regarded as the initial labels of all nodes (miRNAs) in the integrated miRNA similarity-based graph, and the  $l \times m$  prediction matrix  $P^m$ . Finally, the prediction result of SLNPM-SC model is produced by:

$$P_{SLNPM-SC} = \beta P^l + (1-\beta)P^m \quad (9)$$

where  $0 \leq \beta \leq 1$  is the weighted coefficient.

### Interaction profile-based information combination

In this section, we propose the interaction profile-based information combination strategy to build a sequence-derived linear neighborhood propagation model, abbreviated as SLNPM-PC.

The interaction profiles of lncRNAs (miRNAs) without any interaction information are unavailable, and corresponding rows (columns) in the interaction matrix are all zeros. The interaction profile-based information integration strategy is described below.

For miRNA  $L_i$ , which does not have any interaction, its interaction profile is complemented by the sequence information,

$$Y(i, :) = \frac{1}{Q_i} \sum_{i_k \in N(L_i)} S_{LSF}(i, i_k) Y(i_k, :) \quad (10)$$

where  $N(L_i)$  is the set of  $k$  most similar lncRNAs to the lncRNA  $L_i$  based on lncRNA sequence similarity  $S_{LSF}$ , and each of similar lncRNAs has at least one association with miRNAs.  $Q_i$  is the sum of similarity between the lncRNA  $L_i$  and  $k$  most similar lncRNAs,  $Q_i = \sum_{i_k \in N(L_i)} S_{LSF}(i, i_k)$ .

Similarly, for miRNA  $M_j$ , which does not have any interaction, its interaction profile is complemented by the sequence information,

$$Y(:, j) = \frac{1}{Q_j} \sum_{j_k \in N(M_j)} S_{MSF}(j, j_k) Y(:, j_k) \quad (11)$$

where  $N(M_j)$  is the set of  $k$  most similar miRNAs for the

miRNA  $M_j$  based on miRNA sequence similarity  $S_{MSF}$ , and each of similar miRNAs has at least one association with lncRNAs.  $Q_j$  is the sum of similarity between the miRNA  $M_j$  and  $k$  most similar miRNAs,  $Q_j = \sum_{j_k \in N(M_j)} S_{MSF}(j, j_k)$ .

After complementing interaction profiles by using lncRNA (miRNA) sequence similarities, we can calculate interaction similarity matrices for lncRNA and miRNA respectively. Then, we construct prediction models based on lncRNA-lncRNA similarity graph and miRNA-miRNA similarity graph by using label propagation, and the prediction models produce the prediction matrices  $P^m$  and  $P^l$ . The final prediction matrix  $P_{SLNPM-PC}$  is produced by a weighted average of two prediction matrices,

$$P_{SLNPM-PC} = \beta P^l + (1-\beta)P^m \quad (12)$$

where  $0 \leq \beta \leq 1$  is the weighted coefficient.

## Results and discussion

### Evaluation metrics

Here, we adopt 5-fold cross-validation (5-CV) to evaluate prediction models. Specifically, we randomly split known lncRNA-miRNA interactions into five subsets. In each fold, we keep one subset as the testing set, and use others as the training set. All the prediction models are built on the interactions in the training set, and then make predictions for other lncRNA-miRNA pairs. Then, the predictions and real labels (interactions or not) for these pairs are used to calculate evaluation metrics: the area under receiver-operating characteristic curve (AUC), the area under precision-recall curve (AUPR), sensitivity (SEN), specificity (SPEC), precision (PREC), accuracy (ACC) and F-measure (F).

The area under the precision-recall curve (AUPR) and the area under the ROC curve (AUC) are adopted as the evaluation metrics. AUPR and AUC evaluate the performances of prediction models regardless of any threshold. We also adopt binary classification metrics to measure the models, i.e. recall (REC), specificity (SP), precision (PR), accuracy (ACC) and F1-measure (F1). In the experiments, we run 20 runs of 5-CV for each model and adopt averages.

### Parameter settings

In this study, both SLNPM-SC and SLNPM-PC have two major components: the linear neighborhood similarity calculation and similarity-based label propagation. The linear neighborhood similarity has the parameter: neighbor number  $K$ , and the label propagation has the parameter: absorbing probability  $\alpha$ .  $\beta$  is a tradeoff parameter in the final prediction phase. Here, we consider different combinations of following values: {10%, 20%, 30%, ..., 90%} of number of data points for  $K$ , {0.1, 0.2,

0.3, ..., 0.9} for  $\alpha$  and {0, 0.05, 0.1, ..., 0.95, 1} for  $\beta$  to build SLNPM-SC model and SLNPM-PC model, and then evaluate the influence of parameters. All the experiments are conducted with 5-fold cross-validation on SLNPM-S dataset. The result shows that SLNPM-SC model achieves the best AUPR score of 0.6033 when  $K=80\%$ ,  $\alpha=0.4$  and  $\beta=0.25$  and SLNPM-PC model produces the best AUPR score of 0.5996 when  $K=90\%$ ,  $\alpha=0.4$  and  $\beta=0.25$ .

For simplicity, we use the parameter setting in the SLNPM-SC model for analysis. Firstly, we set  $\beta=0.25$  and then evaluate the influence of  $K$  and  $\alpha$  on the performances of SLNPM-SC model. The AUPR scores of SLNPM-SC models with different combinations of  $K$  value and  $\alpha$  value are visualized in Fig. 2 (a). This figure indicates that the parameter  $\alpha$  has great impact on the performance of SLNPM-SC model. More specifically, when  $\alpha$  becomes greater, the performances first increase and then decrease after a peak. Besides, better performance can also be obtained as the neighborhood ratio  $K$  keeps increasing. This result might be the consequence of more neighbors' information being considered to calculate similarity. Then, we fix  $K=0.8$  and  $\alpha=0.4$  and evaluate the influence of parameter  $\beta$  in the prediction model. Note that  $\beta$  is a tradeoff parameter between lncRNA-based prediction and miRNA-based prediction. The parameter  $\beta=1$  means that SLNPM-SC only utilizes the lncRNA-lncRNA similarity information in lncRNA-miRNA interaction prediction. Vice versa, SLNPM-SC only uses the miRNA-miRNA similarity information when  $\beta=0$ . All the results are summarized and shown in Fig. 2 (b) and denote that the prediction model produces

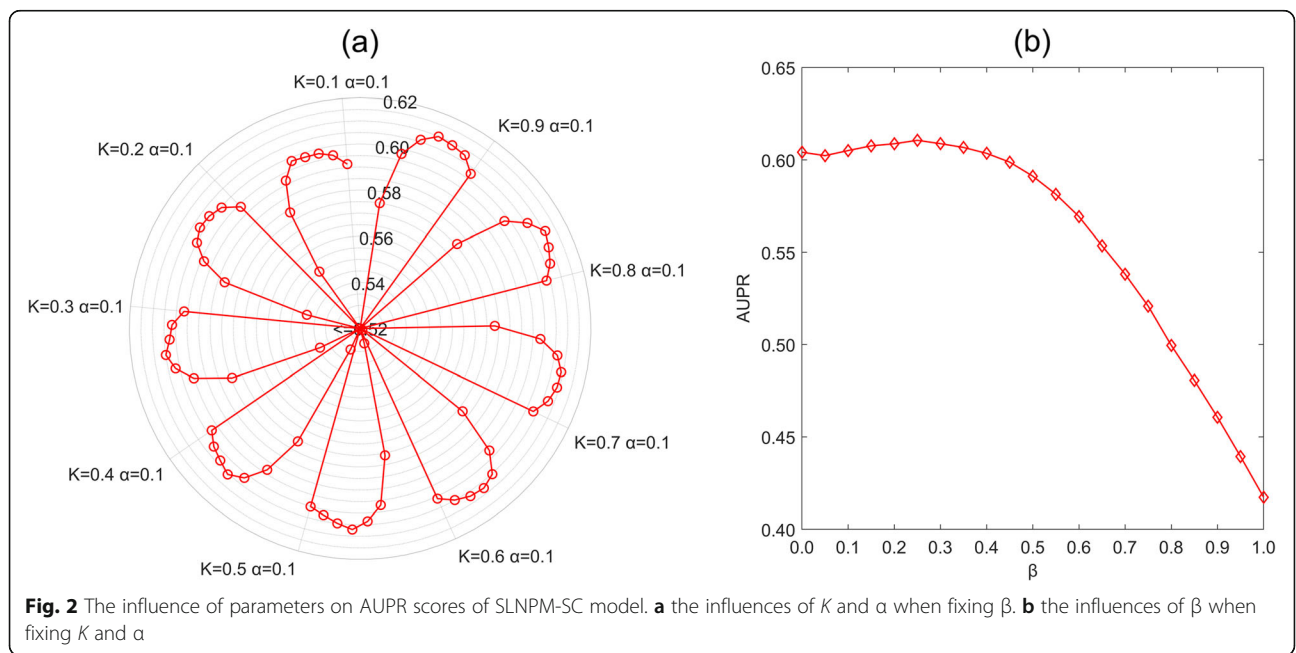
the best result when  $\beta=0.25$ . This result demonstrates the SLNPM-SC model depends more on the miRNA information-based component than the lncRNA information-based component (0.75 VS. 0.25).

Therefore, we adopt  $K=80\%$ ,  $\alpha=0.4$  and  $\beta=0.25$  for SLNPM-SC model and  $K=90\%$ ,  $\alpha=0.4$  and  $\beta=0.25$  for the SLNPM-PC model in the following sections.

**Results of SLNPM-SC and SLNPM-PC**

SLNPM-SC integrates sequence similarity and interaction profile similarity to obtain combined similarities, and then makes predictions based on the combined similarities; SLNPM-PC utilizes the sequence similarities to complement the interaction profiles and then calculates the interaction profile similarity to make predictions.

To demonstrate the superiority of the SLNPM-SC and SLNPM-PC, we build several similar models by using individual features or other similarity measures. First, we respectively build sequence-derived linear neighbor propagation (SLNPM) models based on either interaction profile similarities or sequence similarities. Since existing work [43] ever used the expression profiles of lncRNAs and miRNAs in predicting lncRNA-miRNA interactions, we calculate the expression profile similarity by using linear neighborhood similarity measure (LNS) and build the SLNPM model. We also calculate the sequence similarity by using the Smith-Waterman algorithm (SW) [44] and build the SLNPM model. The performances of the above models are evaluated on SLNPM-S dataset by using 5-CV, and results are shown in Table 2. Clearly, SLNPM-SC and SLNPM-PC produce



**Fig. 2** The influence of parameters on AUPR scores of SLNPM-SC model. **a** the influences of  $K$  and  $\alpha$  when fixing  $\beta$ . **b** the influences of  $\beta$  when fixing  $K$  and  $\alpha$

better results than other SLNPM models, indicating the effectiveness of two information combination strategies. The SLNPM model produced by LNS has better performances than the SLNPM model produced by SW, demonstrating the LNS can better measure lncRNA-lncRNA similarity and miRNA-miRNA similarity than SW. Moreover, the SLNPM models which utilize interaction profile similarities outperform other SLNPM models based on individual feature similarities, revealing the importance of interaction profiles.

Previous studies [26, 29] and our experimental results demonstrate that interaction profiles are critical for predicting lncRNA-miRNA associations. However, interaction profiles of some lncRNAs (miRNAs) are unavailable. Therefore, the models which mainly rely on interaction profiles cannot make predictions for such lncRNAs (miRNAs), and thus we solve this problem with the proposed information combination strategies which utilize the biological feature: lncRNA (miRNA) sequences. Besides, we notice that expression profiles can also describe lncRNAs (miRNAs), and relevant study [28] shows expression profiles play a crucial role in lncRNA-miRNA interactions. To compare the effectiveness of different information sources used in the combination strategy, we respectively utilize sequences and expression profiles to build SLNPM-SC and SLNPM-PC. The performances of these models are evaluated by 5-CV and detailed results are displayed in Table 3. Specifically, we calculate the lncRNA expression profile similarity and miRNA expression profile similarity by using linear neighborhood similarity measure, and build SLNPM-SC (M2) model and SLNPM-PC model (M4), our original SLNPM-SC model(M1) and SLNPM-PC model(M3) based on sequence similarity are denoted by M1 and M3 respectively. Clearly, the SLNPM models based on the sequence similarity can lead to much better performances than the SLNPM models based on expression profile similarity.

Since we implement 20 runs of 5-CV for each model, we can obtain 20 AUPR scores and 20 AUC scores of each model. Further, we test the statistical difference between SLNPM-SC models (M1 and M2) and SLNPM-PC models (M3 and M4) by using the paired t-test. For the SLNPM-SC models, the *P*-values are 7.97E-27 (M2

VS. M1) and 1.07E-10 (M2 VS. M1) respectively in terms of the AUPR scores and AUC scores. For the SLNPM-PC models, considering the AUPR scores and AUC scores, the *P*-values are 1.24E-22 (M3 VS. M4) and 1.63E-04 (M3 VS. M4), respectively. The experimental results show that two editions of sequence-derived linear neighborhood propagation method (M1 and M3) can statistically outperform the SLNPM models based on expression information (M2 and M4) in terms of AUPR and AUC (*P*-value < 0.05).

### Comparison with state-of-the-art methods

To the best of our knowledge, there are only a few machine-learning based methods for lncRNA-miRNA interaction prediction. Here, we adopt EPLMI [26] and INLMI [28] as benchmark methods. EPLMI is a two-way diffusion model which uses the known lncRNA-miRNA interaction-based bipartite graph and expression profiles to predict lncRNA-miRNA interaction. We implement EPLMI using its publicly available source code. INLMI [28] integrates the expression similarity network and the sequence similarity network to predict lncRNA-miRNA interactions, and we implement this model according to descriptions in [28]. Since predicting lncRNA-miRNA interactions can be considered as a link prediction task, we adopt several network link inference methods as baseline methods, i.e. the collaborative filtering method (CF) [45] and the resource allocation algorithm (RA) [46]. The collaborative filtering method takes known lncRNA-miRNA interactions as a bipartite graph and exploits external information, such as expression profiles to calculate the lncRNA-lncRNA similarity and miRNA-miRNA similarity. Then, CF method finds neighbors for each lncRNA and each miRNA, and then predicts unknown interactions by utilizing a weighted average of its neighbors' interacting miRNAs/lncRNAs, then combines the lncRNAs' neighbors-based prediction and the miRNAs' neighbor-based prediction with a tradeoff parameter. The resource allocation algorithm also formulates lncRNAs/miRNAs as nodes and lncRNA-miRNA interactions as links in a bipartite graph. Interaction information is iteratively propagated from miRNAs to their linked lncRNAs, and then the information is allocated from lncRNAs back to miRNAs. After finite iteration,

**Table 2** Performances of SLNPM models based on different information sources

Information Source	Similarity Computing	AUPR	AUC	REC	SP	PR	ACC	F1
Expression Profiles	LNS	0.0305	0.6981	0.0415	0.9974	0.0763	0.9935	0.0518
	SW	0.1358	0.8245	0.2515	0.9956	0.1989	0.9925	0.2191
Sequences	LNS	0.1856	0.8596	0.2883	0.9962	0.2436	0.9932	0.2621
	LNS	0.5981	0.8756	0.5993	0.9990	0.7180	0.9973	0.6500
Interaction Profiles	LNS	0.5981	0.8756	0.5993	0.9990	0.7180	0.9973	0.6500
SLNPM-SC		0.6033	0.9115	0.6043	0.9989	0.7028	0.9972	0.6469
SLNPM-PC		0.5996	0.9006	0.6092	0.9989	0.7087	0.9973	0.6522



**Table 3** Performances of SLNPM models based on different similarities combinations

Models	Model and Information Source	AUPR	AUC	REC	SP	PR	ACC	F1
M1	SLNPM-SC (combining sequence similarity)	0.6033	0.9115	0.6043	0.9989	0.7028	0.9972	0.6469
M2	SLNPM-SC (combining expression profile similarity)	0.3962	0.9000	0.5669	0.9973	0.4734	0.9955	0.5135
M3	SLNPM-PC (complementing IP with sequence similarity)	0.5996	0.9006	0.6092	0.9989	0.7087	0.9973	0.6522
M4	SLNPM-PC (complementing IP with expression profile similarity)	0.5236	0.8980	0.5787	0.9983	0.5929	0.9966	0.5843

IP interaction profile

final resources for miRNAs are probabilities that the lncRNA interacts with these miRNAs. EPLMI and RA have no parameter. INLMI has a parameter that represents the dimension of latent variable in the non-negative matrix factorization. CF has a trade-off parameter for the lncRNAs' neighbor-based prediction and the miRNAs' neighbor-based prediction. We tuned the parameters of INLMI and CF, and adopted the values that produce the best results.

All models are evaluated on SLNPM-S dataset by using 5-CV. As shown in Table 4, SLNPM-SC model achieves AUPR score of 0.6033 and AUC score of 0.9115, and SLNPM-PC model produces AUPR score of 0.5996 and AUC score of 0.9006. The performances of the proposed models are far better than EPLMI (AUPR score of 0.0706 and AUC score of 0.8494), INLMI (AUPR score of 0.0723 and AUC score of 0.8477), RA (AUPR score of 0.5078 and AUC score of 0.8637) and CF (AUPR score of 0.2363 and AUC score of 0.8610). There are several reasons why SLNPM-SC and SLNPM-PC have excellent prediction performances. On one hand, the linear neighborhood similarity measure effectively calculates the lncRNA-lncRNA similarities and miRNA-miRNA similarities. On the other hand, the integrated similarities and complemented interaction profile make use of diverse information.

In the computational predictions, the top-ranked predictions are very important and reflect the performances of models. Here, we check up on the top-ranked predictions ranging from top 100 to top 1000, and figure out how many real interactions can be predicted. As shown in Fig. 3, SLNPM-SC model and SLNPM-PC model perform better than the other three methods when checking up on top-ranked predictions. In the top 100 predictions, EPLMI, INLMI, RA, CF, SLNPM-SC and SLNPM-

**Table 4** Performances of different models on SLNPM-S dataset

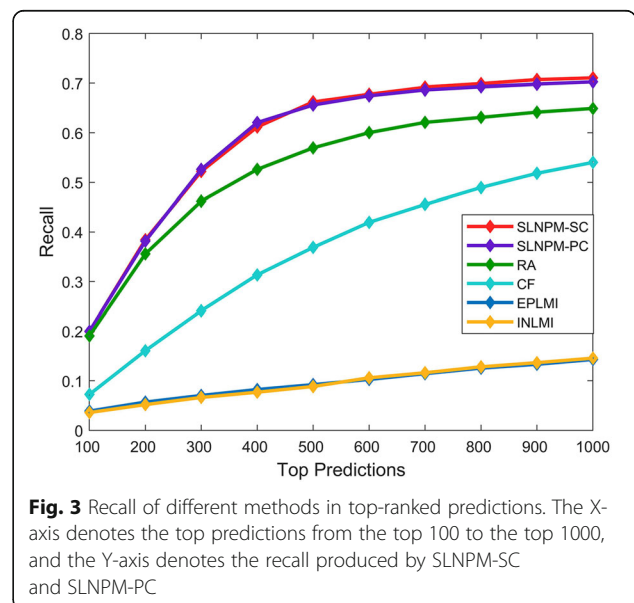
Methods	AUPR	AUC	REC	SP	PR	ACC	F1
EPLMI	0.0706	0.8494	0.1373	0.9962	0.0883	0.9939	0.1055
INLMI	0.0723	0.8477	0.1531	0.9956	0.0867	0.9935	0.1086
RA	0.5078	0.8637	0.5129	0.9987	0.6299	0.9967	0.5631
CF	0.2363	0.8610	0.4599	0.9956	0.3089	0.9934	0.3684
SLNPM-SC	0.6033	0.9115	0.6043	0.9989	0.7028	0.9972	0.6469
SLNPM-PC	0.5996	0.9006	0.6092	0.9989	0.7087	0.9973	0.6522

PC find out 18, 19, 87, 33, 91 and 91 real interactions respectively. Importantly, SLNPM-SC model and SLNPM-PC model can respectively predict 71 and 70% of interactions when only verifying top 1000 predictions.

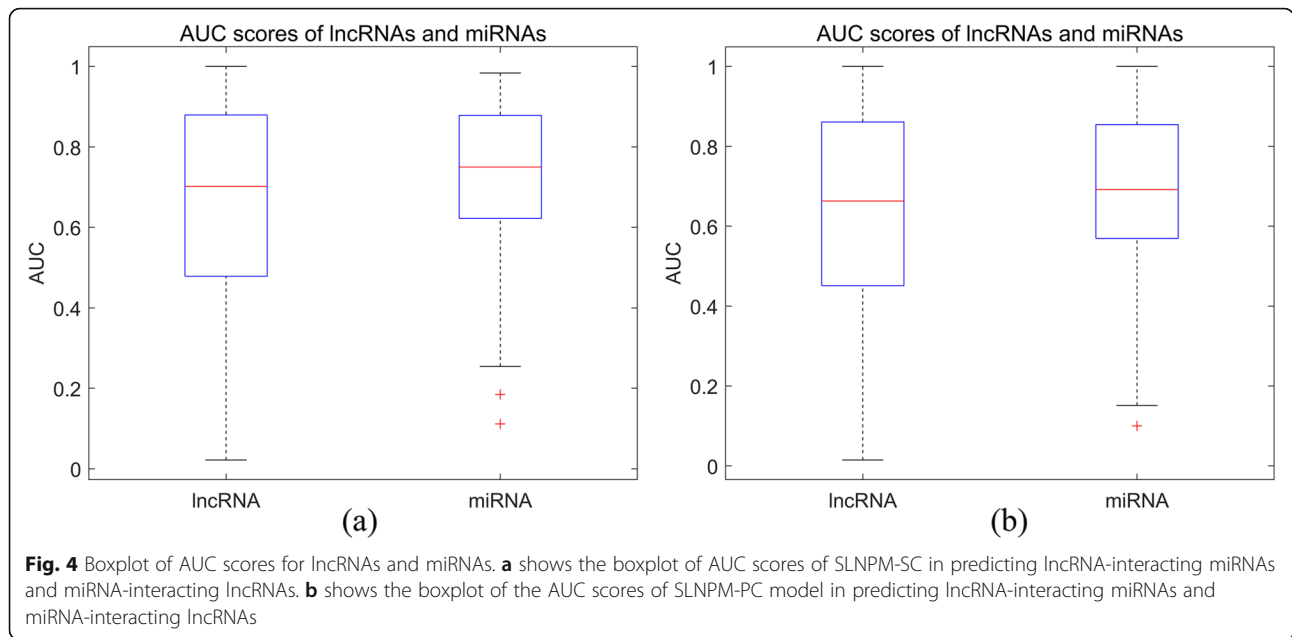
**Case studies**

In this section, we conduct the experiments on SLNPM-L dataset to demonstrate the practical capability of SLNPM-SC and SLNPM-PC for the lncRNA-miRNA interaction prediction.

First, we analyze the performances of SLNPM-SC and SLNPM-PC for predicting lncRNAs (miRNAs) interacted with a specific miRNA (lncRNA). In the experiment, we remove the interactions of a specific lncRNA or the interactions of a specific miRNA in our dataset, and build the SLNPM-SC model and SLNPM-PC model to predict the removed interactions. For every lncRNA or miRNA, we adopt the prediction scores and real labels (interaction or non-interaction) to calculate the AUC scores. We conduct the statistical analysis on the results for lncRNAs and miRNAs, and draw the boxplot. As shown in Fig. 4, the medians of lncRNAs and miRNAs are all larger than 0.65, indicating SLNPM-SC model and SLNPM-PC model can produce satisfying



**Fig. 3** Recall of different methods in top-ranked predictions. The X-axis denotes the top predictions from the top 100 to the top 1000, and the Y-axis denotes the recall produced by SLNPM-SC and SLNPM-PC



results in predicting lncRNA-interacting miRNAs and miRNA-interacting lncRNAs.

Further, we build the SLNPM-SC model and SLNPM-PC model based on SLNPM-L dataset to predict novel lncRNA-miRNA interactions, which are not included in the SLNPM-L dataset. Since the SLNPM-L dataset is compiled from lncRNASNP [17], the predictions are validated by other databases and publicly available literature. We take the lncRNA “MALAT1” and the miRNA “hsa-miR-17-5p” as examples, and respectively build prediction models (SLNPM-SC and SLNPM-PC) to predict miRNAs interacting with “MALAT1” and lncRNAs interacting with “hsa-miR-17-5p”. The lncRNA MALAT1 (metastasis-associated lung adenocarcinoma transcript 1), a bona fide long noncoding RNA, is

reported to be closely related with lung cancer and is one of the first discovered cancer-associated lncRNAs [47, 48]. The miRNA hsa-miR-17-5p, also known as miR-17, is identified as a member of solid cancer miRNA signature [49], and also acts as both an oncogene and a tumor suppressor in different cellular contexts [50, 51].

The top 10 predictions for the lncRNA “MALAT1” and the miRNA “hsa-miR-17-5p” are shown in Table 5. Both SLNPM-SC and SLNPM-PC correctly predict that hsa-miR-1 can interact with the lncRNA “MALAT1”. The study [60] reported that MALAT1 was identified as the target of miRNA hsa-miR-1, and MALAT1 could directly bind with hsa-miR-1, and level of miRNA hsa-miR-1 was negatively associated with that of MALAT1

**Table 5** Top 10 prediction of LNPM-SC and SLNPM-PC for lncRNA “MALAT1” and miRNA “hsa-miR-17-5p”

NO	SLNPM-SC				SLNPM-PC			
	MALAT1	miRNAs	Confirmed?	hsa-miR-17-5p	MALAT1	miRNAs	Confirmed?	hsa-miR-17-5p
1	hsa-miR-1	YES [52]	lnc-SNRPN-8	N.A.	hsa-miR-1	YES [52]	lnc-SNRPN-8	N.A.
2	hsa-miR-101-3p	YES [53]	KCNQ1OT1	N.A.	hsa-miR-101-3p	YES [53]	KCNQ1OT1	N.A.
3	hsa-miR-206	YES [54]	XIST	YES [55]	hsa-miR-142-3p	YES [56]	XIST	YES [55]
4	hsa-miR-210-3p	N.A.	lnc-COL9A2-1	N.A.	hsa-miR-206	YES [54]	lnc-COL9A2-1	N.A.
5	hsa-miR-216a-5p	YES [57]	lnc-ALYREF-1	N.A.	hsa-miR-210-3p	N.A.	lnc-ALYREF-1	N.A.
6	hsa-miR-329-3p	N.A.	COX10-AS1	YES [55]	hsa-miR-216a-5p	YES [57]	COX10-AS1	YES [55]
7	hsa-miR-335-5p	N.A.	lnc-NFAT5-2	N.A.	hsa-miR-335-5p	N.A.	lnc-NFAT5-2	N.A.
8	hsa-miR-3529-5p	N.A.	lnc-ACER2-1	YES [58]	hsa-miR-376b-3p	YES [52]	lnc-ACER2-1	YES [58]
9	hsa-miR-362-3p	N.A.	lnc-LUZP1-1	YES [55]	hsa-miR-455-5p	YES [59]	lnc-LUZP1-1	YES [55]
10	hsa-miR-376b-3p	YES [52]	lnc-NMRK1-1	N.A.	hsa-miR-876-5p	YES [52]	lnc-NMRK1-1	N.A.

N.A not available.

in breast cancer tissues. In general, SLNPM-SC successfully identifies 5 miRNAs interacting with the lncRNA “MALAT1” and 4 lncRNAs interacting with the miRNA “hsa-miR-17-5p”; SLNPM-SC identifies 8 miRNAs interacting with the lncRNA “MALAT1” and 4 lncRNAs interacting with the miRNA “hsa-miR-17-5p”. Therefore, both SLNPM-SC and SLNPM-PC can predict novel lncRNA-miRNA interactions with high accuracy.

## Conclusions

lncRNA-miRNA interactions are critical to many biological events, and exploring these interactions contributes to understanding lncRNA's functions. In this work, we propose a computational method named the sequence-derived linear neighborhood propagation method (SLNPM). SLNPM makes the best use of lncRNA sequences, miRNA sequences and known lncRNA-miRNA interactions to predict novel lncRNA-miRNA interactions. To deal with the miRNAs (or lncRNAs) without interaction information, we introduce two information combination strategies: similarity-based information combination and interaction profile-based information combination, and develop two editions of SLNPM: SLNPM-SC and SLNPM-PC. The proposed models are compared with benchmark methods and baseline methods. The experimental results show that the interaction profiles are very important for the high-accuracy performances of SLNPM-SC and SLNPM-PC, and the information combination strategies further improve performances. The prediction capabilities of proposed models are also tested by case studies, and predicted lncRNAs (miRNAs) for the given miRNA (lncRNAs) are confirmed by existing literature. In conclusion, SLNPM-SC and SLNPM-PC are promising for lncRNA-miRNA interaction prediction. However, SLNPM has several parameters, and it costs a large amount of time to determine optimal parameters. How to effectively tune parameters of SLNPM is our future consideration.

## Abbreviations

5-CV: 5-fold cross-validation; AUC: Area under ROC curve; AUPR: Area under the precision-recall curve; IP: Interaction profile; SLNPM-PC: Sequence-derived linear neighborhood propagation method based on interaction profile information combination; SLNPM-SC: Sequence-derived linear neighborhood propagation method based on similarity information combination

## Acknowledgements

Not applicable.

## About this supplement

This article has been published as part of *BMC Genomics Volume 20 Supplement 11, 2019: Selected articles from the IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2018: genomics*. The full contents of the supplement are available online at <https://bmcbgenomics.biomedcentral.com/articles/supplements/volume-20-supplement-11>.

## Authors' contributions

WZ designed the study, implemented the algorithm and drafted the manuscript. GT implemented the algorithm and drafted the manuscript. SZ, YN helped prepare the data and draft the manuscript. All authors read and approve the final manuscript.

## Funding

Publication costs are funded by National Key Research and Development Program (2018YFC0407904), the National Natural Science Foundation of China (61772381, 61572368) and Huazhong Agricultural University Scientific & Technological Self-innovation Foundation. The funders have no role in the design of the study and collection analysis, and interpretation of data and writing the manuscript.

## Availability of data and materials

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>College of informatics, Huazhong Agricultural University, Wuhan 430070, China. <sup>2</sup>School of Computer Science, Wuhan University, Wuhan 430072, China. <sup>3</sup>Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China. <sup>4</sup>School of Mathematics and Statistics, South-Central University for Nationalities, Wuhan 430074, China.

Published: 20 December 2019

## References

- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009;10(3):155–9.
- Hung T, Chang HY. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol.* 2010;7(5):582–5.
- Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet.* 2014;15(1):7–21.
- Turner M, Galloway A, Vigorito E. Noncoding RNA and its associated proteins as regulatory elements of the immune system. *Nat Immunol.* 2014;15(6):484–91.
- Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li RH, Hennig S, Mosquera JM, Pauwels J, Park K, Kossai M, et al. The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat Commun.* 2014;5:1–3.
- Xia T, Liao Q, Jiang X, Shao Y, Xiao B, Xi Y, Guo J. Long noncoding RNA associated-competing endogenous RNAs in gastric cancer. *Sci Rep.* 2014;4:6088.
- Quagliata L, Matter MS, Piscuoglio S, Arabi L, Ruiz C, Proccino A, Kovac M, Moretti F, Makowska Z, Boldanova T. lncRNA HOTTIP / HOXA13 expression is associated with disease progression and predicts outcome in hepatocellular carcinoma patients. *Hepatology.* 2014;59(3):911.
- Zheng HT, Shi DB, Wang YW, Li XX, Xu Y, Tripathi P, Gu WL, Cai GX, Cai SJ. High expression of lncRNA MALAT1 suggests a biomarker of poor prognosis in colorectal cancer. *Int J Clin Exp Pathol.* 2014;7(6):3174.
- Fang JS, Li YJ, Liu R, Pang XC, Li C, Yang RY, He YY, Lian WW, Liu AL, Du GH. Discovery of multitarget-directed ligands against Alzheimer's disease through systematic prediction of chemical protein interactions. *J Chem Inf Model.* 2015;55(1):149–64.
- Sun H, Wang G, Peng Y, Zeng Y, Zhu QN, Li TL, Cai JQ, Zhou HH, Zhu YS. H19 lncRNA mediates 17 $\beta$ -estradiol-induced cell proliferation in MCF-7 breast cancer cells. *Oncol Rep.* 2015;33(6):3045–52.
- Qian L, Jianguo H, Nanjiang Z, Ziqiang Z, Ali Z, Zhaohui L, Fangting W, Yin-Yuan M. lncRNA loc285194 is a p53-regulated tumor suppressor. *Nucleic Acids Res.* 2013;41(9):4976–87.
- Xu MD, Wang Y, Weng W, Wei P, Qi P, Zhang Q, Tan C, Ni SJ, Dong L, Yang Y. A positive feedback loop of lncRNA-PVT1 and FOXM1 facilitates gastric cancer growth and invasion. *Clin Cancer Res.* 2016;23(8):2071.
- Simon MD. Capture hybridization analysis of RNA targets (CHART). *Curr Protoc Mol Biol.* 2013;21(21 25):1–6.
- Berghoff EG, Clark MF, Chen S, Cajigas J, Leib DE, Kohtz JD. Evt2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development.* 2013;140(21):4407–16.

15. Hao YJ, Wu W, Li H, Yuan J, Luo JJ, Zhao Y, Chen RS. NPinter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database-Oxford*. 2016;:1–5. <https://doi.org/10.1093/database/baw057>
16. Wang TJ, Xie HW. Drug target proteins prediction with network topological indices. *Res J Biotechnol*. 2014;9(12):76–81.
17. Gong J, Liu W, Zhang J, Miao X, Guo AY. lncNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res*. 2015;43(Database issue):D181–6.
18. Kallen AN, Xiao-Bo Z, Jie X, Chong Q, Jing M, Lei Y, Lingeng L, Chaochun L, Jae-Sung Y, Haifeng Z. The imprinted H19 lncRNA antagonizes let-7 microRNAs. *Mol Cell*. 2013;52(1):101–12.
19. Hongyi Z, Kai C, Jing W, Xiaoying W, Kai C, Fangfang S, Longwei J, Yunxia Z, Jun D. MiR-7, inhibited indirectly by lincRNA HOTAIR, directly inhibits SETDB1 and reverses the EMT of breast cancer stem cells by downregulating the STAT3 pathway. *Stem Cells*. 2015;32(11):2858–68.
20. Zhang W, Qu QL, Zhang YQ, Wang W. The linear neighborhood propagation method for predicting long non-coding RNA - protein interactions. *Neurocomputing*. 2018;273:526–34.
21. Ao L, Zang Q, Sun D, Wang M. A text feature-based approach for literature mining of lncRNA-protein interactions. *Neurocomputing*. 2016;206:73–80.
22. Hu H, Zhu C, Ai H, Zhang L, Zhao J, Zhao Q, Liu H. LPI-ETSLP: lncRNA-protein interaction prediction using eigenvalue transformation-based semi-supervised link prediction. *Mol BioSyst*. 2017;13(9):1781–7.
23. Zheng X, Yang W, Kai T, Zhou J, Guan J, Luo L, Zhou S. Fusing multiple protein-protein similarity networks to effectively predict lncRNA-protein interactions. *BMC Bioinformatics*. 2017;18(Suppl 12):420.
24. Zhang W, Yue X, Tang G, Wu W, Huang F, Zhang X. SFPEL-LPI: sequence-based feature projection ensemble learning for predicting lncRNA-protein interactions. *PLoS Comput Biol*. 2018;14(12):e1006616.
25. Zhang T, Wang M, Xi J, Ao L. LPGNMF: Predicting Long Non-coding RNA and Protein Interaction Using Graph Regularized Nonnegative Matrix Factorization. *IEEE/ACM Trans Comput Biol Bioinform*. 2018;PP(99):1–1.
26. Huang YA, Chan K, You ZH. Constructing Prediction Models from Expression Profiles for Large Scale lncRNA-miRNA Interaction Profiling. *Bioinformatics*. 2017;34(5):812–9.
27. 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 Med Genet*. 2018;11(6):113.
28. Hu P, Huang Y-A, Chan KCC, You Z-H. Discovering an Integrated Network in Heterogeneous Data for Predicting lncRNA-miRNA Interactions. *Cham: Springer*; 2018. p. 539–45.
29. Zhang W, Tang G, Wang S, Chen Y, Zhou S, Li X. Sequence-derived linear neighborhood propagation method for predicting lncRNA-miRNA interactions. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2018.
30. Fang S, Zhang L, Guo J, Niu Y, Wu Y, Li H, Zhao L, Li X, Teng X, Sun X, et al. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res*. 2018;46(D1):D308–14.
31. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(D1):D68–73.
32. Panwar B, Omenn GS, Guan YF. miRmine: a database of human miRNA expression profiles. *Bioinformatics*. 2017;33(10):1554–60.
33. Zhang W, Chen Y, Tu S, Liu F, Qu Q. Drug side effect prediction through linear neighborhoods and multiple data source integration. 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, 2016, pp. 427–34.
34. Zhang W, Yue X, Chen YL, Lin WR, Li BL, Liu F, Li XH. Predicting drug-disease associations based on the known association bipartite network. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 2017. p. 503–9.
35. Zhang W, Yue X, Huang F, Liu R, Chen Y, Ruan C. Predicting drug-disease associations and their therapeutic function based on the drug-disease association bipartite network. *Methods*. 2018;145:51–9.
36. Zhang W, Jing K, Huang F, Chen Y, Li B, Li J, Gong J. SFLLN: a sparse feature learning ensemble method with linear neighborhood regularization for predicting drug-drug interactions. *Inf Sci*. 2019;497:189–201.
37. Zhang W, Li Z, Guo W, Yang W, Huang F. A fast linear neighborhood similarity-based network link inference method to predict microRNA-disease associations. *IEEE/ACM transactions on computational biology and bioinformatics*, Early Access, <https://doi.org/10.1109/TCBB.2019.2931546>.
38. Li DF, Luo LQ, Zhang W, Liu F, Luo F. A genetic algorithm-based weighted ensemble method for predicting transposon-derived piRNAs. *Bmc Bioinformatics*. 2016;17:329.
39. Zhang W, Chen YL, Li DF. Drug-Target Interaction Prediction through Label Propagation with Linear Neighborhood Information. *Molecules*. 2017;22(12):2056.
40. Zhang W, Chen YL, Liu F, Luo F, Tian G, Li XH. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *Bmc Bioinformatics*. 2017;18:18.
41. Zhang W, Yue X, Liu F, Chen YL, Tu SK, Zhang XN. A unified frame of predicting side effects of drugs by using linear neighborhood similarity. *BMC Syst Biol*. 2017;11:101.
42. Wen Z, Weitai Y, Xiaoting L, Feng H, Fei L. The bi-direction similarity integration method for predicting microbe-disease associations. *IEEE Access*. 2018;6:38052–61.
43. Huang YA, Chan KCC, You ZH. Constructing prediction models from expression profiles for large scale lncRNA-miRNA interaction profiling. *Bioinformatics*. 2018;34(5):812–9.
44. Smith TF, Waterman MS, Burks C. The statistical distribution of nucleic acid similarities. *Nucleic Acids Res*. 1985;13(2):645–56.
45. Schafer JB, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. *ACM Trans Inf Syst*. 2004;22(1):5–53.
46. Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, Zhang YC. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc Natl Acad Sci U S A*. 2010;107(10):4511–5.
47. Gutschner T, Hämmerle M, Diederichs S. MALAT1 — a paradigm for long noncoding RNA function in cancer. *J Mol Med*. 2013;91(7):791–801.
48. Tony G, Monika HM, Moritz E, Jeff H, Youngsoo K, Alexey R, Gayatri A, Marion S, Matthias G. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res*. 2013;73(3):1180–9.
49. Volinia S, Calin G, Liu C-G, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, et al. A microRNA expression signature of human solid tumors define cancer gene targets. *Proc Natl Acad Sci U S A*. 2006; 103:2257–61.
50. Cloonan N, Brown MK, Steptoe AL, Wani S, Chan WL, Forrest AR, Kolle G, Gabrielli B, Grimmond SM. The miR-17-5p microRNA is a key regulator of the G1/S phase cell cycle transition. *Genome Biol*. 2008;9(8):R127.
51. Li H, Bian C, Liao L, Li J, Zhao RC. miR-17-5p promotes human breast cancer cell migration and invasion through suppression of HBP1. *Breast Cancer Res Treat*. 2011;126(3):565–75.
52. Jin C, Yan B, Lu Q, Lin Y, Ma L. Reciprocal regulation of Hsa-miR-1 and long noncoding RNA MALAT1 promotes triple-negative breast cancer development. *Tumour Biol*. 2015;37(6):7383–94.
53. Wang H, Li W, Zhang G, Lu C, Chu H, Rui Y, Zhao G. MALAT1/miR-101-3p/MCL1 axis mediates cisplatin resistance in lung cancer. *Oncotarget*. 2018;9(7):7501–12.
54. Wang SH, Zhang WJ, Wu XC, Zhang MD, Weng MZ, Zhou D, Wang JD, Quan ZW. Long non-coding RNA Malat1 promotes gallbladder cancer development by acting as a molecular sponge to regulate miR-206. *Oncotarget*. 2016;7(25):37857–67.
55. Jun-Hao L, Shun L, Hui Z, Liang-Hu Q, Jian-Hua Y. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92.
56. Xia C, Liang S, He Z, Zhu X, Chen R, Chen J. Metformin, a first-line drug for type 2 diabetes mellitus, disrupts the MALAT1/miR-142-3p sponge to decrease invasion and migration in cervical cancer cells. *Eur J Pharmacol*. 2018;830:59–67.
57. Zhang Y, Tang X, Shi M, Wen C, Shen B. MiR-216a decreases MALAT1 expression, induces G2/M arrest and apoptosis in pancreatic cancer cells. *Biochem Biophys Res Commun*. 2017;483(2):816–22.
58. Wang P, Li J, Zhao W, Shang C, Jiang X, Wang Y, Zhou B, Bao F, Qiao H. A novel lncRNA-miRNA-mRNA triple network identifies lncRNA RP11-363E7.4 as an important regulator of miRNA and gene expression in gastric cancer. *Cell Physiol Biochem*. 2018;47(3):1025–41.
59. Li L, Yang Z, Wang Y, Zhang Y, Zhou Y, Wang W, Lin L, Su W. Long non-coding RNA MALAT1 promote triple-negative breast cancer progression by regulating miR-204 expression. *Biosci Rep*. 2016;9:969–77.
60. Liu R, Li J, Lai Y, Liao Y, Liu R, Qiu W. Hsa-miR-1 suppresses breast cancer development by down-regulating K-ras and long non-coding RNA MALAT1. *Int J Biol Macromol*. 2015;81:491–7.

## Publisher's Note

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