

RESEARCH

Open Access



# GKLOMLI: a link prediction model for inferring miRNA–lncRNA interactions by using Gaussian kernel-based method on network profile and linear optimization algorithm

Leon Wong<sup>1,6</sup>, Lei Wang<sup>1,2\*</sup>, Zhu-Hong You<sup>3\*</sup>, Chang-An Yuan<sup>1</sup>, Yu-An Huang<sup>3</sup> and Mei-Yuan Cao<sup>4,5</sup>

\*Correspondence:  
leiwang@gxas.cn;  
zhuhongyou@gmail.com

<sup>1</sup> Guangxi Key Lab of Human-machine Interaction and Intelligent Decision, Guangxi Academy of Sciences, Nanning 530007, China

<sup>2</sup> College of Information Science and Engineering, Zaozhuang University, Zaozhuang 277160, China

<sup>3</sup> School of Computer Science, Northwestern Polytechnical University, Xi'an 710139, China

<sup>4</sup> School of Electrical and Electronic Engineering, Guangdong Technology College, Zhaoqing 526100, China

<sup>5</sup> Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor, Malaysia

<sup>6</sup> Institute of Machine Learning and Systems Biology, School of Electronics and Information Engineering, Tongji University, 200092 Shanghai, China

## Abstract

**Background:** The limited knowledge of miRNA–lncRNA interactions is considered as an obstruction of revealing the regulatory mechanism. Accumulating evidence on *Human* diseases indicates that the modulation of gene expression has a great relationship with the interactions between miRNAs and lncRNAs. However, such interaction validation via crosslinking-immunoprecipitation and high-throughput sequencing (CLIP-seq) experiments that inevitably costs too much money and time but with unsatisfactory results. Therefore, more and more computational prediction tools have been developed to offer many reliable candidates for a better design of further bio-experiments.

**Methods:** In this work, we proposed a novel link prediction model based on Gaussian kernel-based method and linear optimization algorithm for inferring miRNA–lncRNA interactions (GKLOMLI). Given an observed miRNA–lncRNA interaction network, the Gaussian kernel-based method was employed to output two similarity matrixes of miRNAs and lncRNAs. Based on the integrated matrix combined with similarity matrixes and the observed interaction network, a linear optimization-based link prediction model was trained for inferring miRNA–lncRNA interactions.

**Results:** To evaluate the performance of our proposed method, *k*-fold cross-validation (CV) and leave-one-out CV were implemented, in which each CV experiment was carried out 100 times on a training set generated randomly. The high area under the curves (AUCs) at  $0.8623 \pm 0.0027$  (2-fold CV),  $0.9053 \pm 0.0017$  (5-fold CV),  $0.9151 \pm 0.0013$  (10-fold CV), and  $0.9236$  (LOO-CV), illustrated the precision and reliability of our proposed method.

**Conclusion:** GKLOMLI with high performance is anticipated to be used to reveal underlying interactions between miRNA and their target lncRNAs, and deciphers the potential mechanisms of the complex diseases.

**Keywords:** Computational biology, miRNA–lncRNA interaction, Link prediction, Competing endogenous RNA (ceRNA), Gaussian kernel



## Introduction

In pre-genomic eras, the genetic central dogma plays a vital role in deciphering the message flow of genetic material [1]. However, along with the in-depth studies, it was found that the biological mechanism is far more complex than the dogma [2, 3]. From the recent researches, more and more non-coding RNA (ncRNA) that cannot directly translate into protein has been found that it can function as regulation in most of the biological processes [4]. The ncRNA makes up more than 98% of total RNA. There are many criteria for ncRNA classification. According to the length of RNA and the significance, microRNA (miRNA) with 20–25 nt and long non-coding RNA (lncRNA) with > 200 nt are two main kinds of ncRNA. miRNA can combine with the Argonaute proteins and function as the RNA-induced silencing complex (RISC), which leads to target mRNA degradation and translation repression [5, 6]. lncRNA has similarity with protein-coding RNA in splicing structure and length [7]. The competing endogenous RNA (ceRNA) hypothesis indicates that crosstalk among RNAs is existing, which results in various RNA regulation in vivo [3, 8].

The research on miRNA–target interaction (MTI) is a hot topic for its regulatory role in proliferation and apoptosis, cell differentiation, cellular transport, transcriptional and post-transcriptional regulation, epigenetic regulation, cell cycle control, tumorigenesis, and organ or tissue development [9–11]. It is known that some specific miRNAs can act as response elements, also called MREs. Gene expression can be inhibited when RNAs need to bind with the same MREs competitively [12]. It is reported that miRNAs can regulate lncRNAs for which some specific lncRNAs have similar structures with mRNA [13]. However, more and more studies release that miRNA–lncRNA interactions (MLIs) play essential roles in human diseases such as tumors, cancer, and vasculature [14–16]. It is believed that more effective therapeutic approaches can be developed by further investigating the regulation activities among miRNAs and lncRNAs [10, 13].

To reveal the underlying mechanisms of regulatory mediated by miRNAs and lncRNAs, the interactions between them need to be validated [11]. Owing to the next-generation bioengineering techniques, it is conducive to help researchers to carry out the bio-experiments in high throughput, obtain the results, and build a normative database [17, 18]. Based on the cumulative bio-data, the researchers in other fields can easily get access and make more efforts for further investigation.

The experimental identification method can intuitively detect the functions of miRNAs acting on their targets. Many well-established online databases collect miRNA-related records with details, and update periodically, such as miR2disease, miRCancer, OncomiRDB, DIANA-TarBase, microRNA.org, miRGate, miRDB, miRNAMap and human microRNA disease database (HMDD) [19–27]. By using the database management techniques and web-based techniques, these records can be freely accessed and browsed through multiple filters. For instance, DIANA-TarBase v8.0 collects more than 665,800 miRNA interactomes with details about publication, cell lines, tissue, and experiments, as well as implements an intuitive interface with searching filters including species, cell type, tissue, regulation type, and other options. Therefore, accumulating bio-data facilitates researchers to make further progress.

Biogenetical deregulation of some specific miRNAs can lead to various human diseases such as miR-17-92 at malignant lymphoma [28], miR-206 at sensing motor neuron [29, 30], and miR-1 at cardio-genesis [31]. However, the validation progress of miRNA interactomes is still underway. The regulation functions of most miRNAs are far from clear, for which it may be affected by the evolution of species, gene mutation, dynamic *in vivo*, and other uncertain factors [32].

To date, an optimal framework for revealing mechanisms of ncRNA-target duplexes is to integrate experimental and computational approaches iteratively [33, 34]. By using the limited understanding of miRNAs and the related targets, computational methods can be developed to predict potential interactions that have high probabilities to be true positive ones in further biological experiments. Many existing computational prediction tools are combined with some wide-used principles such as evolutionary conservation status, seed sequence complementarity, target-site abundance, target-site accessibility, free energy, G-U wobble, and local AU flanking content [5]. There are many miRNA-target prediction tools based on different combination of principles that can improve the performance of prediction to a certain extent.

Specifically, LncTar was proposed to predict the RNA target of lncRNA by integrating the primer-dimer prediction method implemented by PerlPrimer, and normalized free energy that can measure the relative stability of base pairing between RNAs [35]. A pattern-based approach rna22 was proposed to identify miRNAs and their corresponding heteroduplex, which facilitates users to obtain prediction results with different binding sites that are recognized by using Teiresias algorithm [36]. A parameter-free prediction model, Probability of Interaction by Target Accessibility (PITA), was proposed to apply target-site accessibility principle to detect miRNA-target duplexes according to the energy change in procedure of RNAs' formation and unpairing [37]. miRanda was proposed to predict the miRNA:target based on evolutionary conservation, seed sequence complementarity, G-U wobble and free energy [38]. TargetScan was developed based on multiple principles such as local AU content, seed sequence complementarity, target-site accessibility, target-site abundance, evolutionary conservation, G-U wobble, compensatory pairing and free energy [39]. Moreover, some web-based prediction tools with powerful computing are developed to facilitate researchers to dispose large-scale data, such as STarMirDB, DIANA-microsT-CDS, miRGate and miRDB. Although these existing tools were well-established, many efforts still need to be made on the researches of miRNA:target.

Most of the tools mentioned above rigidly relied on the conserved seed match, which may yield a high false-negative rate because of the sophistication of the mechanism *in vivo* [40–45]. In recent years, many existing *in silico* prediction methods of miRNA:target were developed based on machine learning (ML) algorithms. ML algorithms can make it possible to do the prediction task more applicable and effective for various species and miRNA-related targets by learning the most potential features from the limited experimental data. Nevertheless, more and more flexible methods were developed to predict miRNA:target by constructing an effective interaction network with more entities and combining the side information.

Due to the vital role of MLI, some specific prediction algorithms have been proposed for their particularity. Yu et al. developed a resource allocation-based algorithm called

LCBNI that integrated the similarities based on the sequence profile [46]. LNRLMI was proposed to use the expression profile of RNAs to construct the RNA similarities, in which the co-expression information among RNAs can be utilized for model training through the integrated network [47]. LMNLMI is a matrix completion model combined with multiple profiles for prediction [48]. INLMI and EPLMI are both based on a two-way diffusion [49, 50]. The difference is that INLMI computed the final rating scores by taking the average of the results from the sequence-based and expression profile-based model, and EPLMI trained two sub-models where one is based on miRNA expression profile-related weighting network and one is based on lncRNA expression profile-related weighting network.

In this work, motivated by the previous works, we here propose a linear optimization-based method combined with Gaussian kernel-based network similarity for inferring miRNA–lncRNA interactions. In detail, A Gaussian kernel-based method was employed to construct the similarity on the given observed miRNA–lncRNA interaction network. Then, the training matrix was constructed by combining the interaction network and the similarities of miRNAs and lncRNAs. Finally, a link prediction model was trained by using the well-constructed training matrix. For performance evaluation,  $k$ -fold cross-validation (CV) experiments were implemented by setting  $k$  to 2, 5, and 10. Also, leave-one-out cross-validation (LOO-CV) was carried out. Specifically, by carrying out each  $k$ -fold CV with a randomly given interaction profile 100 times, our proposed method yielded high AUCs at  $0.8623 \pm 0.0027$  (2-fold CV),  $0.9053 \pm 0.0017$  (5-fold CV),  $0.9151 \pm 0.0013$  (10-fold CV), and 0.9236 (LOO-CV). From the results yielded by the existing methods and our proposed method, it is believed that our proposed method can yield reliable results for obtaining more potential miRNA–lncRNA interactions and uncover the underlying regulatory mechanisms.

## Results

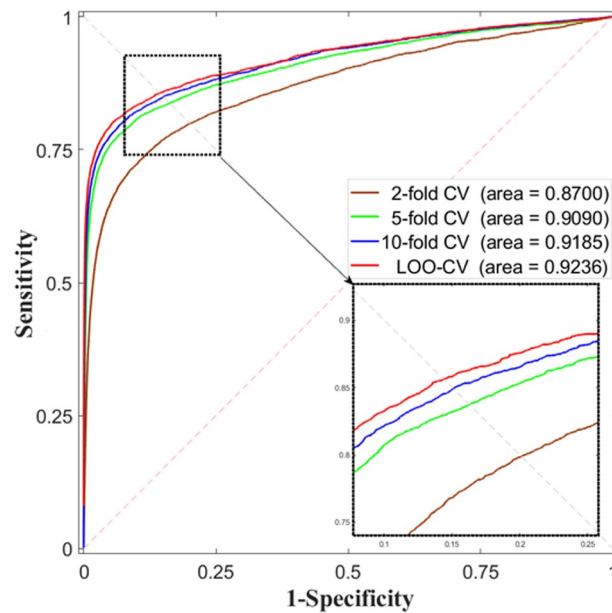
### Using $k$ -fold cross-validation for performance evaluation

For the purpose of performance evaluation, it is universal to apply the  $k$ -fold cross-validations (CV) method that can help to measure more precisely to a certain extent. In this paper, 4 kinds of CV experiments were implemented, including 2-fold CV, 5-fold CV, 10-fold CV, and leave-one-out CV (LOO-CV). A smaller value of  $k$  indicates the less training samples. Before training model, all samples are shuffled and divided into  $k$  parts equally.  $k - 1$  parts are for training and the rest for testing. When employing 5-fold CV, 5118 samples of MLIs used in our work are divided into 4095 samples for training and 1023 samples for testing. The 5-fold CV is widely used for comparison among models.

On account of that training model with different training sets yield different performance, each  $k$ -fold CV is implemented for 100 times. 2-fold, 10-fold CV and LOO-CV were implemented and the results are listed in Table 1. As a result, our proposed method that is integrated with Gaussian kernel-based network similarity yielded the average AUCs at 0.9053 and a low standard deviation at 0.0017 in 5-fold CV. Also, the AUCs are at  $0.8623 \pm 0.0027$  (2-fold CV),  $0.9151 \pm 0.0013$  (10-fold CV), and 0.9236 (LOO-CV). To better illustrate the performance, all the receiver operation characteristics (ROC) curves were plotted in Fig. 1.

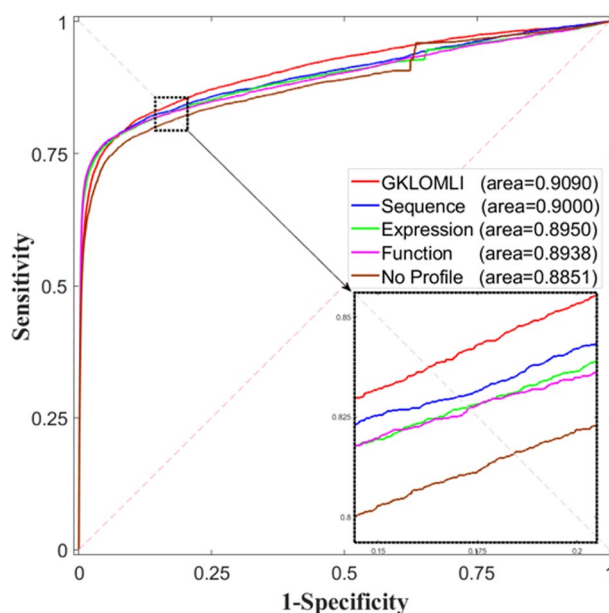
**Table 1** Performance comparison by implementing CV experiments on different profiles

| Profile      | 2-fold CV       | 5-fold CV       | 10-fold CV      | LOO-CV |
|--------------|-----------------|-----------------|-----------------|--------|
| No profile   | 0.8306 ± 0.0037 | 0.8806 ± 0.0020 | 0.8910 ± 0.0014 | 0.8997 |
| Expression   | 0.8378 ± 0.0033 | 0.8903 ± 0.0021 | 0.9017 ± 0.0014 | 0.9112 |
| Bio-function | 0.8389 ± 0.0031 | 0.8890 ± 0.0022 | 0.9008 ± 0.0018 | 0.9111 |
| Sequence     | 0.8515 ± 0.0031 | 0.8941 ± 0.0020 | 0.9040 ± 0.0016 | 0.9123 |
| GKLOMLI      | 0.8623 ± 0.0027 | 0.9053 ± 0.0017 | 0.9151 ± 0.0013 | 0.9236 |

**Fig. 1** ROC curves of GKLOMLI by implementing 2-fold, 5-fold, 10-fold CV and LOO-CV

### Experiment results in different types of RNA profile-based similarity

To demonstrate the outstanding performance of our proposed method, we here implement  $k$ -fold CV with the other three types of profiles and without profile. The results are listed in Table 1. The corresponding ROC curves were plotted in Fig. 2. From the results of 2-fold CV, the AUCs are at  $0.8306 \pm 0.0037$  (no profile),  $0.8378 \pm 0.0033$  (expression profile),  $0.8389 \pm 0.0031$  (bio-function), and  $0.8515 \pm 0.0031$  (sequence), respectively. From the results of 5-fold CV, the AUCs are at  $0.8806 \pm 0.0020$  (no profile),  $0.8903 \pm 0.0021$  (expression profile),  $0.8890 \pm 0.0022$  (bio-function), and  $0.8515 \pm 0.0031$  (sequence), respectively. Also,  $p$  values between our proposed method and the methods with the other three types of profiles and without profile are calculated at  $3.4646e-154$  (no profile),  $1.6730e-95$  (expression profile),  $1.0972e-107$  (bio-function) and  $3.7514e-93$  (sequence), respectively. From the results of 10-fold CV, the AUCs are at  $0.8910 \pm 0.0014$  (no profile),  $0.8903 \pm 0.0021$  (expression profile),  $0.9008 \pm 0.0018$  (bio-function), and  $0.9040 \pm 0.0016$  (sequence), respectively. From the results of LOO-CV, the AUCs are at 0.8997 (no profile), 0.9112 (expression profile),  $0.9008 \pm 0.0018$  (bio-function), and 0.9123 (sequence), respectively.



**Fig. 2** ROC curves by implementing 5-fold CV on different profiles and no profile

From these results, the model without any profile yields the lowest AUC overall. From the results of the models based on different bio-profiles, sequence-based model performs well. It should be noticed that the collection of bio-profiles are incomplete except the sequence information, which indicated that the lack of information tends to lead to a low AUC. From the 5-fold results, the  $p$  values are lower than 0.05, which indicates our method has better performance than those methods.

#### Comparison with different prediction methods

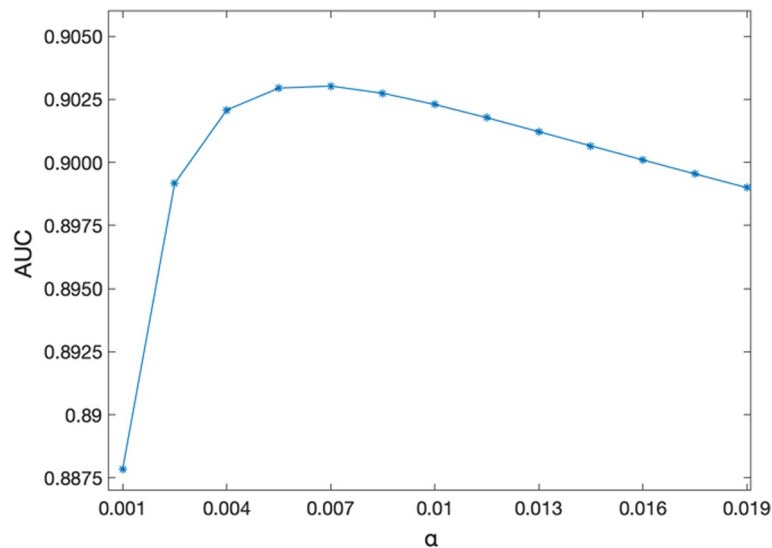
To better illustrate the good performance of our proposed method, the current widely used prediction methods were introduced for comparison, including LCBNI [46], LNRLMI [47], LMNLMI [48], INLMI [50], and EPLMI [49]. Yu et al. proposed a method named LCBNI to predict interactions between miRNAs and lncRNAs by employing a resource allocation algorithm combined with the sequence-based similarity of RNAs. LNRLMI method was proposed to apply the co-expression mechanism in a miRNA–lncRNA interaction network to construct the expression profile-based similarity and integrate that for prediction. LMNLMI method was first proposed to integrate three types of biological profiles relating to biological function, sequence information, and expression. INLMI employed a non-negative matrix factorization method on an integrated network utilizing interaction profile, sequence information-based, and expression profile-based similarities. EPLMI was first proposed to predict miRNA–lncRNA interactions by using a two-way diffusion model based on the expression profiles of miRNA and lncRNA.

From the 5-fold CV experiment results listed in Table 2, LNRLMI and EPLMI achieved the average AUCs at 0.8960 and 0.8447, respectively. And, LCBNI, LMNLMI and INLMI achieved the best AUCs at 0.8982, 0.8926 and 0.8517, respectively. Our proposed method yielded the best AUC at 0.9090 and the average AUC at 0.9053. The



**Table 2** Performance comparison of different existing methods

| Method | LCBNI [46] | LNRLMI [47]     | LMNLMI [48] | INLMI [50] | EPLMI [49]      | GKLOMLI         |
|--------|------------|-----------------|-------------|------------|-----------------|-----------------|
| AUC    | 0.8982     | 0.8960 ± 0.0015 | 0.8926      | 0.8517     | 0.8447 ± 0.0017 | 0.9053 ± 0.0017 |

**Fig. 3** AUCs of different parameter  $\alpha$ 

outperformance of our proposed method indicated that it can be a better method for inferring miRNA–lncRNA interactions.

### Parameter selection

In order to study the sensitivity of GKLOMLI to the parameter  $\alpha$ , 13 different numbers for parameter  $\alpha$  are used in the range of 0.001–0.019 (0.001, 0.0025, 0.004, ..., 0.019) with step of 0.0015, when implementing 5-fold CV experiments. Based on these parameters setting, the results are plotted (see Fig. 3), in which the highest AUC of 0.9030 is obtained with  $\alpha$  at 0.007. The distribution of AUC on a bell-shape curve indicates that it is easy to optimize the proposed model. Moreover, the distribution tends to reach the peak as  $\alpha = 0.0055$  and  $AUC = 0.9029$ . Thus, it demonstrates the robustness of the model to the setting of parameter.

### Discussion

Our proposed method proposed to transform the side information of RNAs into similarity and integrate that for further model training and prediction. The extra profiles of RNAs have been released constantly, which can be used for bioanalysis. In machine learning, fusing side information into former original data can offer more features of the entity information. To some extent, the integrated features can help train a model with better performance. The state-of-the-art methods are proposed to use different bio-information to improve the model performance. Specifically, LCBNI is based on sequence information. LNRLMI and EPLMI introduced expression profile of RNAs.

INLMI integrated sequence information and expression profile. LMNLMI utilized bio-functional profile, sequence information and expression profile. The extra bio-information is introduced for improvement, but the results point out that more is not always better. However, the biological profiles of some specific RNAs haven't been released, which leads to information loss. Such a problem largely results in low accuracy. In the pipeline of our proposed method, the integrated network completely relies on the network profile, which indicates its general application in many link prediction problems. The limited data of the interaction profile can be fully utilized by employing the Gaussian kernel-based method to transform them into a similarity matrix for information supplement. In further experiments,  $k$ -fold CV was implemented for performance evaluation. From the LOO-CV and  $k$ -fold CV experiments with parameter  $k$  at 2, 5, and 10, the results illustrated that more information offered can yield a higher AUC that means better performance. By integrating different biological profiles, the model based on the interaction profile yielded the highest AUC that indicated the Gaussian kernel-based method employed on a network profile is effective and reliable. Comparing the experiment results among the existing studies on miRNA–lncRNA interaction prediction, our proposed method performs well.

## Conclusions

The process of investigation of miRNA–lncRNA interactions still needs to make more effort for its essential role in potential regulation mechanisms in ceRNA network. In this paper, we proposed to employ the Gaussian kernel-based method on interaction profile to construct the network similarity and a linear optimization-based link prediction model. From all experiment results, it suggests that (1) ceRNA network contains the underlying information and can be extracted for network integration, and (2) information fusion is helpful for further model training in collaborative interactions between miRNAs and lncRNAs. Our proposed method can be a powerful tool to decipher the underlying regulation mechanisms in ceRNA network.

## Methods

### Dataset

In this work, there are 4 kinds of profiles used for investigating the interactions of miRNA–lncRNA from various databases, including miRNA–lncRNA interaction profile, sequence information, bio-functional profile, and expression profile. The interaction profile describes the miRNAs interact with lncRNA or not, in which value of 1 denotes interaction, and value of 0 denotes non-interaction. Sequence information is that an RNA is a fix-length string of 4 kinds of bases including A, U, C and G. The bio-functional profile is arranged by a kind of gene annotation and interactions between RNAs and their targets. The expression profile is a series of numerical values that represent the expression level of a specific RNA in a cell line or tissues in vivo.

MLIs are obtained from the lncRNASNP database (released in Feb. 2017), was download at <http://bioinfo.life.hust.edu.cn/lncRNASNP> [18]. 8091 miRNA–lncRNA target pairs are provided from the database derived from 108 CLIP-Seq datasets, which are strongly related to bio-experimental studies. After processing data de-duplication, 5118 MLIs involving in



275 miRNAs and 780 lncRNAs are used for constructing the miRNA–lncRNA interaction network.

The sequence information is available in miRbase and LNCipedia databases at <http://www.mirbase.org/> and <https://lncipedia.org/>, respectively [51, 52].

The bio-functional profiles of miRNAs are collected from miRTarBase of version 6.1 [53, 54]. 272 bio-functional profiles of miRNAs are collected. To obtain the functional profile of lncRNAs, the Lnc-GFP method was applied to learn the probable bio-function of lncRNAs from the coding-non-coding co-expression network.

The expression profiles of miRNAs and lncRNAs are downloaded from the microRNA.org database and the NONCODE database, respectively [19, 55]. 8 cell lines and 16 types of tissues are involved in the expression profiles of lncRNAs, as well as 172 dimensions of cell lines and tissues in the expression profiles of miRNAs. The numbers of expression profiles of miRNAs and lncRNAs are 230 and 450, respectively.

### Constructing RNA similarity for side information

In the pipeline of our proposed method, RNA similarities are regarded as the side information in the constructed observed miRNA–lncRNA interaction network, considering the observation that RNAs with similar clusters of targeting RNAs have more similarities in their profiles. Four kinds of similarities were constructed, including Gaussian kernel-based network similarity, sequence-based similarity, bio-function similarity, and expression profile-based similarity.

The Gaussian kernel is widely used in many fields for its efficiency in refining useful information from any input.

Given a profile  $PF \in \mathbb{R}^{n \times d}$  with  $n$  RNA samples and  $d$  dimensions of the profile, the Gaussian kernel-based similarity value between the  $i$ -th and  $j$ -th sample is calculated as follow:

$$GKS(r(i), r(j)) = \exp\left(-\gamma_r PF(r(i)) - PF(r(j))^2\right), \tag{1}$$

where  $\gamma_r$  denotes the Gaussian kernel bandwidth. Its definition is as follow:

$$\gamma_r = \left[ \left( \sum_{i=1}^n PF(r(i))^2 \right) / n \right]^{-1}. \tag{2}$$

To construct sequence-based similarity, a sequence alignment method was employed. The Needleman–Wunsch method is encapsulated in the *pairwise2* package of *Biopython* under the environment of Python. In the parameter setting, the identification score, gap-open penalty, and gap-open extending penalty were at 2,  $-0.5$ , and  $<0.1$ , respectively.

The function profile of RNA is related to so many entities. A set theory-based method was applied to construct the function similarity. The definition of the algorithm is as follow:

$$FS(r_a, r_b) = \frac{\text{card}(RA(r_a) \cap RA(r_b))}{\sqrt{\text{card}(RA(r_a)) \cdot \text{card}(RA(r_b))}} \tag{3}$$

where  $r_a$  and  $r_b$  denote two RNAs, and  $RA(\cdot)$  denotes the functional annotations of RNA.

The Pearson correlation coefficient measure is one of the high-efficiency similarity measurements. The definition of it is as follow:

$$PCCM(r_a, r_b) = \frac{\sum_{i=1}^N (EX(r_a, i) - \overline{EX(r_a)})(EX(r_b, i) - \overline{EX(r_b)})}{\sqrt{\sum_{i=1}^N (EX(r_a, i) - \overline{EX(r_a)})^2 \sum_{i=1}^N (EX(r_b, i) - \overline{EX(r_b)})^2}} \quad (4)$$

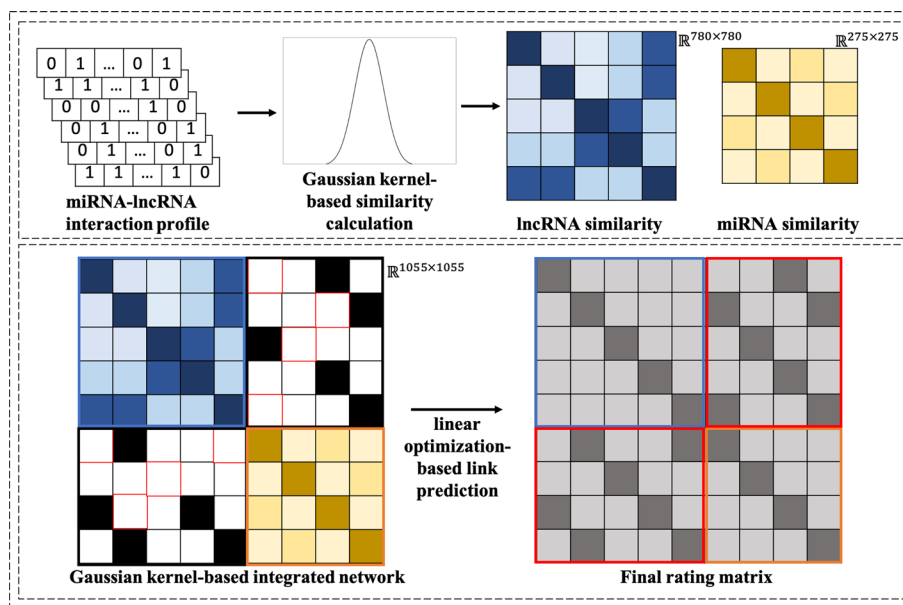
where  $EX(r_a, i)$  denotes the  $i$ -th value of  $N$  elements in the expression profile of RNA  $r_a$  and  $\overline{EX(r_a)}$  denotes the mean value of the expression profile of RNA  $r_a$ .

**A linear optimization-based method for inferring miRNA–lncRNA interactions**

In the pipeline of our proposed method, a semi-supervised learning algorithm is introduced to infer miRNA–lncRNA interactions from a constructed informative network that contains side information (see Fig. 4). In detail, in the framework of the proposed prediction model, we first integrate the given observed miRNA–lncRNA interaction network with the side information (bio-similarity) mentioned in the last section. Then, a linear optimization-based model is trained based on the integrated network. Notice that the similarities of miRNA and lncRNA are derived from the same kind of bio-profile.

Since more and more researches support the assumption that two RNAs with more similarities in profiles have a greater possibility of interacting with a common target RNA cluster, fusing the side information for training can be conducive to the investigation of miRNA–lncRNA interaction.

Given the similarity matrixes of miRNA and lncRNA, i.e.,  $Sim_{miRNA}$  and  $Sim_{lncRNA}$ , the integrated network is constructed with the adjacent matrix of miRNA–lncRNA interactions  $Adj \in \mathbb{R}^{n_{miRNA} \times n_{lncRNA}}$  as follow:



**Fig. 4** Flowchart of GKLOMLI model

$$Adj' = \begin{bmatrix} Sim_{miRNA} & Adj \\ Adj^\top & Sim_{lncRNA} \end{bmatrix} \tag{5}$$

where  $Adj'$  can be regarded as a weighted graph  $G(V, E, W)$  with vertices  $V$ , edges  $E$ , and weighting  $W$ . Noted that  $Adj'$  is a real symmetric matrix, in which each row of a vertices in  $Adj'$  denotes a pathway regarded as edges with weights to other vertices. A target rating matrix related to  $Adj'$  and a contribution matrix  $C$  can be denoted as follow:

$$RS = Adj' \cdot C \tag{6}$$

Each element  $RS(i, j)$  can be represented by unfolding as follow:

$$RS(i, j) = \sum_k Adj'(i, k) \cdot C(k, j) \tag{7}$$

Then, we set a target function to get a final target matrix  $RS$  by solving it as a linear optimization problem. The definition is as follow:

$$\min_C \alpha Adj' - Adj' \cdot C + C \tag{8}$$

where  $C$  should be with a small magnitude and  $\alpha$  is a parameter for constraint. To solve the above equation, it can be treated as a minimum optimization problem by using the Frobenius-2 norm as follow:

$$\begin{aligned} F &= \alpha Adj' - Adj' \cdot C^2 + C^2 \\ &= \alpha Tr \left[ (Adj' - Adj' C)^\top (Adj' - Adj' C) \right] + Tr(C^\top C) \\ &= \alpha Tr \left( Adj'^\top Adj' - Adj'^\top Adj' C - C^\top Adj'^\top Adj' + C^\top Adj'^\top Adj' C \right) + Tr(C^\top C) \end{aligned} \tag{9}$$

$$\frac{\partial F}{\partial C} = \alpha \left( 2Adj'^\top Adj' C - 2Adj'^\top Adj' \right) + 2C \tag{10}$$

Let the above equation to be zero,  $C$  can be obtained:

$$C^* = \alpha \left( \alpha Adj'^\top Adj' + E \right)^{-1} Adj'^\top Adj' \tag{11}$$

where  $E$  is the identity matrix related to  $Adj'$ . The final rating matrix can be represented as follow:

$$RS = Adj' C^* \tag{12}$$

**Acknowledgements**

The authors would like to thank all anonymous reviewers for their constructive advice.

**Author contributions**

Conceptualization, LW (Leon Wong) and LW (Lei Wang); methodology, LW (Leon Wong), LW (Lei Wang) and Y-AH; validation, C-AY, Z-HY and M-YC; writing-original draft preparation, LW (Leon Wong); writing-review and editing, LW (Lei Wang) and Z-HY; investigation, LW (Leon Wong); funding acquisition, LW (Lei Wang) and Z-HY. All authors have read and agreed to the published version of the manuscript.

**Funding**

This work was supported in part by STI 2030-Major Projects, under Grant 2021ZD0200403, in part by the Guangxi Post-doctoral Special Funding Project, the Natural Science Foundation of Guangxi, under Grant 2022JJD170019, the National

Natural Science Foundation of China, under Grants 62172355, the Guangxi Science and Technology Base and Talent Special Project under Grant 2021AC19394 and 2021AC19354.

#### Availability of data and materials

The data pertaining to the present study has been included in table and/or figure form in the present manuscript. And all datasets and computational code underlying this study are available in an online archive <https://github.com/leon-lg-wong/GKLOMLI>.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 30 August 2022 Accepted: 27 April 2023

Published online: 08 May 2023

#### References

- Crick F. Central dogma of molecular biology. *Nature*. 1970;227(5258):561–3. <https://doi.org/10.1038/227561a0>.
- Costello A, Badran AH. Synthetic biological circuits within an orthogonal central dogma. *Trends Biotechnol*. 2020;39:59–71.
- Cesana M, Daley GQ. Deciphering the rules of ceRNA networks. *Proc Natl Acad Sci*. 2013;110(18):7112–3. <https://doi.org/10.1073/pnas.1305322110>.
- Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, León-Ricardo BX, et al. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*. 2014;159(1):148–62. <https://doi.org/10.1016/j.cell.2014.08.028>.
- Ab Mutalib N-S, Sulaiman SA, Jamal R. Computational tools for microRNA target prediction. In: Wei LK, editor. Computational epigenetics and diseases. Cambridge: Academic Press; 2019. p. 79–105.
- Kawamata T, Seitz H, Tomari Y. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol*. 2009;16(9):953. <https://doi.org/10.1038/nsmb.1630>.
- Beermann J, Piccoli M-T, Viereck J, Thum T. Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiol Rev*. 2016;96(4):1297–325. <https://doi.org/10.1152/physrev.00041.2015>.
- 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):353–8.
- Pasquinelli AE. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet*. 2012;13(4):271–82.
- Bassett AR, Azzam G, Wheatley L, Tibbit C, Rajakumar T, McGowan S, et al. Understanding functional miRNA–target interactions in vivo by site-specific genome engineering. *Nat Commun*. 2014;5(1):1–11. <https://doi.org/10.1038/ncomms5640>.
- Tang X, Feng D, Li M, Zhou J, Li X, Zhao D, et al. Transcriptomic analysis of mRNA–lncRNA–miRNA interactions in hepatocellular carcinoma. *Sci Rep*. 2019;9(1):1–12. <https://doi.org/10.1038/s41598-019-52559-x>.
- Paraskevopoulou MD, Karagkouni D, Vlachos IS, Tastsoglou S, Hatzigeorgiou AG. microCLIP super learning framework uncovers functional transcriptome-wide miRNA interactions. *Nat Commun*. 2018;9(1):1–16. <https://doi.org/10.1038/s41467-018-06046-y>.
- Paraskevopoulou MD, Hatzigeorgiou AG. Analyzing miRNA–lncRNA interactions. In: Feng Y, Zhang L, editors. Long non-coding RNAs. Berlin: Springer; 2016. p. 271–86.
- Kataoka M, Wang D-Z. Non-coding RNAs including miRNAs and lncRNAs in cardiovascular biology and disease. *Cells*. 2014;3(3):883–98. <https://doi.org/10.3390/cells3030883>.
- Ballantyne M, McDonald R, Baker A. lncRNA/MicroRNA interactions in the vasculature. *Clin Pharmacol Ther*. 2016;99(5):494–501. <https://doi.org/10.1002/cpt.355>.
- Shen L, Liu F, Huang L, Liu G, Zhou L, Peng L. VDA-RWLRLS: An anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares. *Comput Biol Med*. 2022;140:105119.
- Ho PY, Yu AM. Bioengineering of noncoding RNAs for research agents and therapeutics. *Wiley Interdiscip Rev RNA*. 2016;7(2):186–97. <https://doi.org/10.1002/wrna.1324>.
- Gong J, Liu W, Zhang J, Miao X, Guo A-Y. lncNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res*. 2015;43(D1):D181–6.
- Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA. org resource: targets and expression. *Nucleic Acids Res*. 2008;36(suppl\_1):D149–53. <https://doi.org/10.1093/nar/gkm995>.
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009;37(suppl\_1):D98–104. <https://doi.org/10.1093/nar/gkn714>.
- Xie B, Ding Q, Han H, Wu D. miRCancer: a microRNA–cancer association database constructed by text mining on literature. *Bioinformatics*. 2013;29(5):638–44. <https://doi.org/10.1093/bioinformatics/btt014>.

22. Wang D, Gu J, Wang T, Ding Z. OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics*. 2014;30(15):2237–8. <https://doi.org/10.1093/bioinformatics/btu155>.
23. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res*. 2018;46(D1):D239–45. <https://doi.org/10.1093/nar/gkx1141>.
24. Andrés-León E, González Peña D, Gómez-López G, Pisano DG. miRGate: a curated database of human, mouse and rat miRNA–mRNA targets. *Database*. 2015. <https://doi.org/10.1093/database/bav035>.
25. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res*. 2015;43(D1):D146–52. <https://doi.org/10.1093/nar/gku1104>.
26. Hsu S-D, Chu C-H, Tsou A-P, Chen S-J, Chen H-C, Hsu PW-C, et al. miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. *Nucleic Acids Res*. 2007;36(suppl\_1):D165–9. <https://doi.org/10.1093/nar/gkm1012>.
27. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, et al. HMDD v3.0: a database for experimentally supported human microRNA–disease associations. *Nucleic Acids Res*. 2019;47(D1):D1013–7. <https://doi.org/10.1093/nar/gky1010>.
28. Jiang C, Bi C, Jiang X, Tian T, Huang X, Wang C, et al. The miR-17~92 cluster activates mTORC1 in mantle cell lymphoma by targeting multiple regulators in the STK11/AMPK/TSC/mTOR pathway. *Br J Haematol*. 2019;185(3):616–20. <https://doi.org/10.1111/bjh.15591>.
29. Shi G, Zeng P, Zhao Q, Zhao J, Xie Y, Wen D, et al. The regulation of miR-206 on BDNF: a motor function restoration mechanism research on cerebral ischemia rats by meridian massage. *Evid Based Complement Alternat Med*. 2022;2022:8172849. <https://doi.org/10.1155/2022/8172849>.
30. Valsecchi V, Anzilotti S, Serani A, Laudati G, Brancaccio P, Guida N, et al. miR-206 reduces the severity of motor neuron degeneration in the facial nuclei of the brainstem in a mouse model of SMA. *Mol Ther*. 2020;28(4):1154–66. <https://doi.org/10.1016/j.jymthe.2020.01.013>.
31. Sadat-Ebrahimi SR, Rezabakhsh A, Aslanabadi N, Asadi M, Zafari V, Shanebandi D, et al. Novel diagnostic potential of miR-1 in patients with acute heart failure. *PLoS ONE*. 2022;17(9):e0275019. <https://doi.org/10.1371/journal.pone.0275019>.
32. Peng L, Wang F, Wang Z, Tan J, Huang L, Tian X, et al. Cell–cell communication inference and analysis in the tumour microenvironments from single-cell transcriptomics: data resources and computational strategies. *Brief Bioinform*. 2022;23(4):326. <https://doi.org/10.1093/bib/bbac234>.
33. Peng L, Tan J, Tian X, Zhou L. EnANNDeep: an ensemble-based lncRNA–protein interaction prediction framework with adaptive k-nearest neighbor classifier and deep models. *Interdiscip Sci Comput Life Sci*. 2022;14(1):209–32.
34. Lihong P, Wang C, Tian X, Zhou L, Li K. Finding lncRNA–protein interactions based on deep learning with dual-net neural architecture. *IEEE/ACM Trans Comput Biol Bioinform*. 2021;19:3456–68.
35. Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, et al. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform*. 2015;16(5):806–12.
36. Loher P, Rigoutsos I. Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics*. 2012;28(24):3322–3. <https://doi.org/10.1093/bioinformatics/bts615>.
37. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet*. 2007;39(10):1278–84.
38. Turner DA. Miranda: a non-strict functional language with polymorphic types. In: *Conference on functional programming languages and computer architecture*. Springer; 1985, p. 1–16.
39. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4:e05005. <https://doi.org/10.7554/eLife.05005>.
40. Yi H-C, You Z-H, Zhou X, Cheng L, Li X, Jiang T-H, et al. ACP-DL: a deep learning long short-term memory model to predict anticancer peptides using high-efficiency feature representation. *Mol Ther Nucleic Acids*. 2019;17:1–9.
41. Wang L, You Z-H, Huang D-S, Zhou F. Combining high speed ELM learning with a deep convolutional neural network feature encoding for predicting protein–RNA interactions. *IEEE/ACM Trans Comput Biol Bioinform*. 2018;17:972–80.
42. Zheng K, You Z-H, Wang L, Zhou Y, Li L-P, Li Z-W. Dbmda: a unified embedding for sequence-based miRNA similarity measure with applications to predict and validate miRNA–disease associations. *Mol Ther Nucleic Acids*. 2020;19:602–11. <https://doi.org/10.1016/j.omtn.2019.12.010>.
43. Wang M-N, You Z-H, Wang L, Li L-P, Zheng K. LDGRNMF: lncRNA–disease associations prediction based on graph regularized non-negative matrix factorization. *Neurocomputing*. 2020;424:236–45.
44. You Z, Wang S, Gui J, Zhang S. A novel hybrid method of gene selection and its application on tumor classification. In: *International conference on intelligent computing*. Springer; 2008, p. 1055–68.
45. Chen Z-H, You Z-H, Li L-P, Wang Y-B, Wong L, Yi H-C. Prediction of self-interacting proteins from protein sequence information based on random projection model and fast Fourier transform. *Int J Mol Sci*. 2019;20(4):930. <https://doi.org/10.3390/ijms20040930>.
46. Yu Z, Zhu F, Tianl G, Wang H. LCBNI: link completion bipartite network inference for predicting new lncRNA–miRNA interactions. In: *2018 IEEE International conference of safety produce informatization (IICSPI)*. IEEE; 2018, p. 873–7.
47. Wong L, Huang YA, You ZH, Chen ZH, Cao MY. LNRLM: linear neighbour representation for predicting lncRNA–miRNA interactions. *J Cell Mol Med*. 2020;24(1):79–87. <https://doi.org/10.1111/jcmm.14583>.
48. Hu P, Huang Y-A, Chan KC, You Z-H. Learning multimodal networks from heterogeneous data for prediction of lncRNA–miRNA interactions. *IEEE/ACM Trans Comput Biol Bioinform*. 2019;17:1516–24.
49. Huang Y-A, Chan KC, You Z-H. Constructing prediction models from expression profiles for large scale lncRNA–miRNA interaction profiling. *Bioinformatics*. 2018;34(5):812–9. <https://doi.org/10.1093/bioinformatics/btx672>.
50. Hu P, Huang Y-A, Chan KC, You Z-H. Discovering an integrated network in heterogeneous data for predicting lncRNA–miRNA interactions. In: *International conference on intelligent computing*. Springer; 2018, p. 539–45.
51. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 2014;42(D1):D68–73.

52. Volders P-J, Helsens K, Wang X, Menten B, Martens L, Gevaert K, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* 2013;41(D1):D246–51. <https://doi.org/10.1093/nar/gks915>.
53. Chou C-H, Chang N-W, Shrestha S, Hsu S-D, Lin Y-L, Lee W-H, et al. miRTarBase 2016: updates to the experimentally validated miRNA–target interactions database. *Nucleic Acids Res.* 2016;44(D1):D239–47. <https://doi.org/10.1093/nar/gkv1258>.
54. Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, et al. miRTarBase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res.* 2011;39(suppl\_1):D163–9. <https://doi.org/10.1093/nar/gkq1107>.
55. Bu D, Yu K, Sun S, Xie C, Skogerbø G, Miao R, et al. NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.* 2012;40(D1):D210–5. <https://doi.org/10.1093/nar/gkr1175>.

### Publisher's Note

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

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

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

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

