BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 04/08 pp261–272
Volume 2, Number 4, December 2019
DOI: 10.26599/BDMA.2019.9020010

Prediction of miRNA-circRNA Associations Based on *k*-NN Multi-Label with Random Walk Restart on a Heterogeneous Network

Zengqiang Fang and Xiujuan Lei*

Abstract: Circular RNAs (circRNAs) play important roles in various biological processes, as essential non-coding RNAs that have effects on transcriptional and posttranscriptional gene expression regulation. Recently, many studies have shown that circRNAs can be regarded as micro RNA (miRNA) sponges, which are known to be associated with certain diseases. Therefore efficient computation methods are needed to explore miRNA-circRNA interactions, but only very few computational methods for predicting the associations between miRNAs and circRNAs exist. In this study, we adopt an improved random walk computational method, named KRWRMC, to express complicated associations between miRNAs and circRNAs. Our major contributions can be summed up in two points. First, in the conventional Random Walk Restart Heterogeneous (RWRH) algorithm, the computational method simply converts the circRNA/miRNA similarity network into the transition probability matrix; in contrast, we take the influence of the neighbor of the node in the network into account, which can suggest or stress some potential associations. Second, our proposed KRWRMC is the first computational model to calculate large numbers of miRNA-circRNA associations, which can be regarded as biomarkers to diagnose certain diseases and can thus help us to better understand complicated diseases. The reliability of KRWRMC has been verified by Leave One Out Cross Validation (LOOCV) and 10-fold cross validation, the results of which indicate that this method achieves excellent performance in predicting potential miRNA-circRNA associations.

Key words: miRNA-circRNA associations; heterogeneous network; multi-label; random walk restart

1 Introduction

A decade ago, biologists discovered that circular RNAs (circRNAs) are present in human cells and tissue[1]. Compared with linear RNAs, circRNAs have tended to draw less attention due to their unusual nature. Nonetheless, with the development of identification technology, an increasing number of studies are providing evidence that circRNAs play significant roles in the whole biological process[2–4]. A circRNA is a noncoding RNA structured as a closed RNA loop and lacking the free ending 5' and 3' ends that are present in linear RNAs. Due to limitations in earlier detection technologies[5], researchers were only able to identify those RNAs that have free 5' and 3' ends; therefore, closed loop circRNAs were neglected. With improvements in detection methods[6], researchers can now use advanced detection technologies to identify novel circRNAs. Recently, some researchers have used expression profiles or RNA sequence data to conduct tissue specificity experiments[7,8]. In addition, a growing number of studies have shown that circRNAs can work or function as micro RNA (miRNA) sponges[9–11], and circRNAs can also regulate and modulate the expression of their ancestral genes[12–15]. Since the beginning of last century, miRNAs have gradually attracted attention because of their biomarking function. In recent years,

• Zengqiang Fang and Xiujuan Lei are with the College of Computer Science, Shaanxi Normal University, Xi'an 710119, China. E-mail: fangzq@snnu.edu.cn; xjlei@snnu.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2018-11-25; revised: 2019-04-16; accepted: 2019-04-17

research has revealed that miRNAs play a variety of roles in biological processes[16–18]. Especially, many studies have been done based on eukaryotes. There is also much evidence indicating that miRNAs are closely related to various diseases[19–22]. Therefore, predicting the potential associations between circRNAs and the disease-related miRNAs is important for promoting future work in disease prediction. Beyond that, these aforementioned works can help people to obtain a greater overall comprehension of RNAs and related diseases. Although high throughput technology and biological experiments have been applied widely to identify intricate biological associations, such techniques remain expensive and time-consuming. Therefore, to predict potential associations between different biological molecules, many computational methods have been suggested in order to sharply reduce the time and cost. The abundance of research along these lines includes work on the prediction of candidate diseases related genes[23–25], miRNA-disease associations[26–29], long non-coding RNA (lncRNA)-disease associations[30,31], drug targets[32,33], disease-related environmental factors[34], and miRNA-lncRNA associations[29,35]. From previous studies we find that biological networks have been widely and successfully applied to predict complicated biological associations. In this study, we propose a novel computational method called KRWRMC to predict associations between miRNA and circRNA. Our method uses the k-Nearest Neighbor (k-NN) algorithm based on the Random Walk Restart (RWR) method to predict miRNA and circRNA interactions. The heterogeneous network is based on three basic subnetworks: a miRNA functional similarity subnetwork, a circRNA functional similarity subnetwork, and a miRNA-circRNA association subnetwork. Some related information can be disregarded. Taking the miRNA-circRNA association subnetwork into consideration, each miRNA entry may be associated with the same circRNAs, meaning that we can regard the miRNA-related circRNAs as labels. The circRNA-related miRNAs can also be regarded as labels for specific circRNAs. Based on the above idea, we can obtain new similarity scores between each miRNA pair, through which some of the miRNA similarity scores may increase and some may decrease. This can highlight some miRNA, which are specific neighbors of miRNA associated with circRNAs, as an increased miRNA similarity score will stress the association between a neighbor miRNA of the specific miRNA and the specific miRNA-related circRNA. We then follow the same procedures on the circRNA similarity subnetwork. Finally, the improved RWR method is adopted on the reconstructed heterogeneous network.

2 Method

2.1 Network

2.1.1 miRNA-circRNA interaction network

To build the miRNA-circRNA interaction network, an initial miRNA-circRNA associations dataset is downloaded from the starBase database[36] (http://starBase.sysu.edu.cn/index.php). starBase is a comprehensive bioinformatics database containing a variety of interactions between different biological molecules, such as lncRNA-miRNA interactions, miRNA-pseudogene interactions, and miRNA-sncRNA interactions. starBase includes 130 000+ miRNA-circRNA interactions, with 276 miRNA entries and 7018 circRNA entries. We select 226 miRNA entries and 905 circRNA entries from the initial dataset, as shown in Table 1. Finally, the miRNA-circRNA interaction adjacency matrix A is constructed, made up of 19 000+ interactions. In matrix A, an entry $A(i, j)$ in row i and column j is equal to 1 if miRNA i and circRNA j are associated, otherwise it is equal to 0.

2.1.2 miRNA functional similarity network

To calculate the miRNA functional similarity score, the miRNA and its related gene data are downloaded from the miRTarBase database[37,38], containing miRNAs and their target genes (http://mirtarbase.mbc.nctu.edu.tw/php/index.php). Gene Ontology (GO) data, downloaded from the Human Protein Reference Database (HPRD)[39] (http://www.hprd.org/), is then used to measure the similarity score between each pair of miRNAs. Based on our initial dataset, the miRNA-circRNA interaction network has 226 miRNA entries to extract. After we pre-process the GO data obtained from the HPRD database, we obtain GO terms and related genes entries, which are utilized to calculate the miRNA functional similarity scores.

Some similarity measuring methods have been

Table 1 Number of experimental data.

Number of miRNA-circRNA interactions	19 644
Number of miRNAs	226
Number of circRNAs	905

proposed by previous studies[40]. The common-term-based method[40] is adopted here to calculate the similarity scores, by which the more GO terms are shared by two different miRNA target genes G_{m_i} and G_{m_j}, the higher is their similarity score. The matrix MS is denoted as the miRNA functional similarity network, in which an entry MS(i, j) can be calculated by Eq. (1):

$$\mathrm{MS}(i,j)=\frac{\left|G_{m_i}\cap G_{m_j}\right|}{\left|G_{m_i}\cup G_{m_j}\right|} \tag{1}$$

where G_{m_i} and G_{m_j} are the sets of the whole GO terms of the target genes which are matched by both miRNAs i and j.

2.1.3 circRNA functional similarity network

To calculate the circRNA functional similarity score, circRNA target genes are downloaded from the circBase database[41] (http://www.circbase.org/). There are 90 000+ circRNA target genes collected in the circBase database. 905 circRNA entries are derived from the initial miRNA-circRNA interactions dataset and its GO terms. Simultaneously, the circRNA-related gene GO terms are downloaded from the HPRD database. The circRNA functional similarity network is denoted as CS, with the entry CS(i, j) calculated by the following equation:

$$\mathrm{CS}(i,j)=\frac{|G_{c_i}\cap G_{c_j}|}{|G_{c_i}\cup G_{c_j}|} \tag{2}$$

where G_{c_i} and G_{c_j} are the sets of the whole GO terms of the target genes which are matched by both circRNA i and j.

2.1.4 Heterogeneous network

Our initial heterogeneous network is constructed based on three subnetworks: the miRNA-circRNA association network, miRNAs functional similarity network, and circRNA functional similarity network.

According to the aforementioned processes, we know that $A=\{a(i,j)\}_{i=1,j=1}^{n,l}$ denotes the miRNA-circRNA association network, A^{T} denotes the transposed matrix of the miRNA-circRNA association matrix, $\mathrm{MS}=\{M(i,j)\}_{i=1,j=1}^{l,l}$ denotes the miRNA functional similarity network, and $\mathrm{CS}=\{C(i,j)\}_{i=1,j=1}^{n,n}$ denotes the circRNA functional similarity network, where a represents a miRNA-circRNA node in the miRNA-circRNA association network, M represents a node in the miRNA functional similarity network, and C represents a node in the circRNA functional similarity network. Based on the above three subnetworks, we realize the heterogeneous network H as established by Eq. (3):

$$H=\begin{bmatrix}\mathrm{MS} & A\\ A^{\mathrm{T}} & \mathrm{CS}\end{bmatrix} \tag{3}$$

2.2 KRWRMC model

In this study, we propose a computational method with a heterogeneous network based on the improved RWR algorithm. The random walk algorithm has been widely applied to heterogeneous networks for a variety of biological applications, such as for predicting the potential associations between diseases and candidate genes[42–44], miRNA[23,45,46] and lncRNA[47–49], and for seeking drug targets[50,51]. Therefore, we have developed KRWRMC to predict associations between miRNA and circRNA, using the k-NN algorithm based on the RWR method. The process of building a KRWRMC model is shown in Fig. 1. Firstly, the miRNA-circRNA association network is used to build a k-NN based link graph for miRNA and circRNA, which are depicted as K_m and K_c, respectively, and calculated in accordance with the below Eqs. (5) and (6). Secondly, we integrate the functional similarity network of miRNA/circRNA with the k-NN based link graph for miRNA/circRNA, denoting the final miRNA/circRNA similarity network as W_m and W_c, respectively. When all of the above steps are complete, the heterogeneous network H is converted into a transition heterogeneous network with a multi-label learning algorithm using the k-NN based link graphs of miRNA and circRNA. The initial heterogeneous network only has the weighted relationship between edge pairs. The greater the weight of the edges, the higher is the possibility of turning to the node during the transfer process. After we convert the weighted matrix into a transition matrix, the RWR algorithm is applied to the heterogeneous network; this step can be described as the beginning of the iteration process on the seed nodes in the graph. Each iteration makes a choice between selecting adjacent nodes based on their transition potential or returning to the starting (seed) node. In this algorithm, the parameter γ is used to indicate the restart probability and $1-\gamma$ to represent the probability of moving to adjacent nodes. The probability distribution remains stationary after the iteration. After converting all the weighted matrixes and adjacency matrices into transition matrixes, the heterogeneous network is reconstructed and is defined as follows:

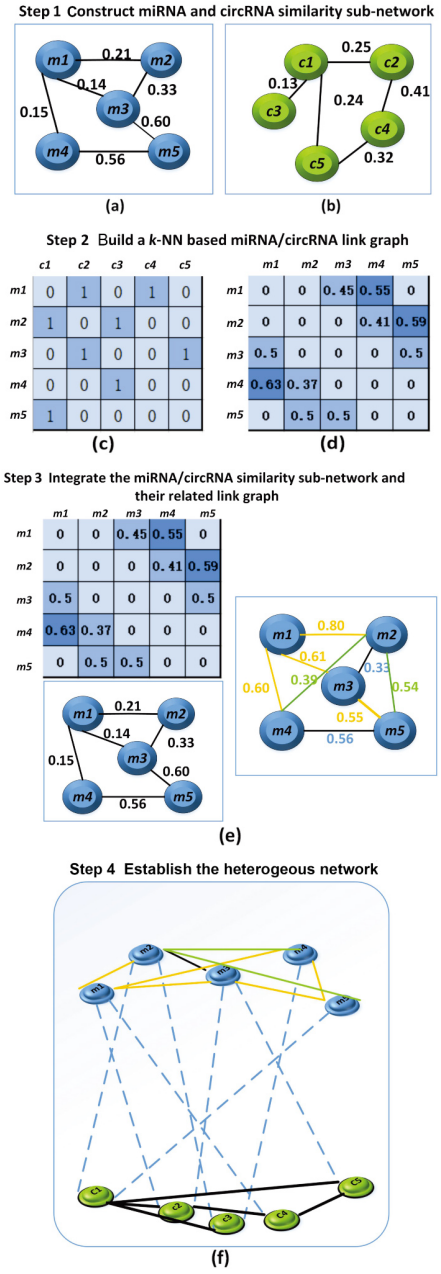

Fig. 1 Flowchart of KRWRMC framework. There are four main steps to construct the model (*m* is represented miRNA and *c* is represented circRNA.). Step 1: Based on the GO data to establish the miRNA/circRNA similarity sub-network. Step 2: Based on the miRNA-circRNA network to build the *k*-NN based link graph of miRNA/circRNA. Step 3: integrating GO similarity sub-network and its related link graph. Step 4: Based on the integrated miRNA/circRNA sub-networks and the miRNA-circRNA sub-network, the final heterogeneous network was set up. It is clear that the green lines denote the inferred novel similarity scores, the yellow ones denote the changing weight scores between the initial associations, and the black ones represent the initial association weight scores.

$$T = \begin{bmatrix} T_{mm} & T_{mc} \\ T_{cm} & T_{cc} \end{bmatrix} \tag{4}$$

If only miRNA functional similarity is taken into account, some potential associations are neglected in the miRNA function similarity network. Accordingly, we could regard circRNA entries which are in the predicted circRNA-miRNA associations as miRNA related item labels, which means that when predicting the probability of associations between a specific miRNA and circRNA entries, each circRNA entry can be regarded as a label. Here, a classification algorithm, multi-label *k*-NN[52], is adopted to calculate the potential similarity score between each miRNA-miRNA pair. Thereby, we build a *k*-NN based on the link graph[53] of miRNA. The main idea of the link graph is that we see the instances that are being predicted as the labels; for each instance $(m_i, L_i) \in A$, m_i represents the miRNA items and L_i represents the related labels found in the miRNA-circRNA associations matrix. On the basis that different items have different labels, we can use this to measure the similarity between two items.

The definition of a *k*-NN based link graph can be described as follows:

$$\begin{aligned} G^m &= \{V^m, E^m\}, \\ V^m &= \{v_i | (m_i, L_i) \in A\}, \\ E^m &= \{(v_i^m, v_j^m | v_i^m, v_j^m \in V^m, \\ &\quad v_i^m \in \mathrm{kNN}(v_j^m), v_j^m \in \mathrm{kNN}(v_i^m), i \neq j)\}. \end{aligned}$$

Based on the above definition, we can calculate the weight of the *k*-NN based link graph. According to the *k*-NN based link graph of miRNA, some potential associations are created between each miRNA pairs, which can show or highlight some significant interactions as follows:

$$K_m(i,j) = \begin{cases} 0, & i = j \text{ or } (v_i^m, v_j^m) \notin E^m; \\ \dfrac{\mathrm{dis}(v_i^m, v_j^m)}{\sum\limits_{v_p^m \in \mathrm{kNN}(v_i^m)} \mathrm{dis}(v_i^m, v_p^m)}, & (v_i^m, v_j^m) \in E^m \end{cases} \tag{5}$$

Euclidean distance is adopted to measure the distance between the vertices v_i^m, v_j^m, K_m; K_m is a normalized matrix based on the row. It is worth remarking that K_m is not a symmetrical matrix. The hidden circRNA association matrix K_c can be obtained in the same way, as follows:

$$K_c(i,j)=\begin{cases}0, & i=j \text{ or } (v_i^c,\ v_j^c)\notin E^c;\\ \dfrac{\mathrm{dis}(v_i^c,\ v_j^c)}{\sum\limits_{v_p^c\in \mathrm{kNN}(v_i^c)}\mathrm{dis}(v_i^c,\ v_p^c)}, & (v_i^c,\ v_j^c)\in E^c\end{cases} \tag{6}$$

After we obtain two k-NN based link graphs of weighted matrixes K_m and K_c, some extra edges can be extracted or additional weighting scores added to measure the probability of miRNA to miRNA transition and circRNA to circRNA transition. Finally, the networks K_m and MS are integrated to establish a fusion miRNA similarity matrix W_m, which can be described as follows:

$$W_m(i,j)=\frac{\mathrm{MS}(i,j)+K_m(i,j)}{2} \tag{7}$$

Equally, we can gain the final similarity matrix from circRNA to circRNA named W_c with same way which can be described as follows:

$$W_c(i,j)=\frac{\mathrm{CS}(i,j)+K_c(i,j)}{2} \tag{8}$$

There are four different kinds of transition matrices[47,54] in Eq. (4): T_{mm} and T_{cc} are intra-subnetwork transition matrixes and T_{mm} and T_{cm} are inter-subnetwork transition matrixes. The transition probability matrix from miRNA to miRNA T_{mm} can be defined as follows:

$$T_{mm}=\begin{cases}W_c(i,j)\Big/\sum\limits_{k=1}^{n}W_c(i,k), & \text{if } \sum\limits_{k=1}^{n}A(i,k)=0;\\ (1-\zeta)W_c(i,j)\Big/\sum\limits_{k=1}^{n}W_c(i,k), & \text{otherwise}\end{cases} \tag{9}$$

The transition probability matrix from circRNA to circRNA T_{cc} can be defined as follows:

$$T_{cc}=\begin{cases}W_c(i,j)\Big/\sum\limits_{k=1}^{m}W_c(i,k), & \text{if } \sum\limits_{k=1}^{n}A(i,k)=0;\\ (1-\theta)W_c(i,j)\Big/\sum\limits_{k=1}^{n}W_c(i,k), & \text{otherwise}\end{cases} \tag{10}$$

The transition probability from miRNA to circRNA matrix T_{mc} can be defined as follows:

$$T_{mc}=\begin{cases}0, & \text{if } \sum\limits_{k=1}^{m}A(i,k)=0;\\ \zeta A(i,j)\Big/\sum\limits_{k=1}^{m}A(i,k), & \text{otherwise}\end{cases} \tag{11}$$

The transition probability matrix from circRNA to miRNA T_{cm} can be defined as follows:

$$T_{cm}=\begin{cases}0, & \text{if } \sum\limits_{k=1}^{n}A(k,i)=0;\\ \theta A(i,j)\Big/\sum\limits_{k=1}^{n}A(k,i), & \text{otherwise}\end{cases} \tag{12}$$

where the parameters ζ and θ represent the possibility that one seed node jumps from miRNA subnetwork to circRNA subnetwork, or vice versa. We then apply the RWR algorithm on the final rebuilt heterogeneous transition network T. The algorithm can be described as starting from seed nodes and walking with an equal probability to its neighbor nodes or, with a different probability, returning to itself. The formal formula for this algorithm is as follows[23]:

$$p_{t+1}=(1-\gamma)T^{\mathrm{T}}p_t+\gamma p_0 \tag{13}$$

where T is the final transition matrix calculated from the aforementioned context and $\gamma\in(0,1)$ is the restart probability, meaning that the walker will hold a probability of γ to return to itself during each iteration. The initial vector is $p_0=\begin{bmatrix}(1-\eta)m_0\\ \eta c_0\end{bmatrix}$ and the sum of the probabilities is equal to 1. The vector m_0 can be initialized as follows. We set starting seed node i equal to 1 and other neighbor nodes are assigned as 0. Equally, the dimension of vector c_0 will be initialized at an equal probability for the seed nodes, thus $c_0=\begin{bmatrix}\frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n}\end{bmatrix}$. After several iterations, each value of the vector p_{t+1} will remain stable, which denotes that result has converged[55,56]. We adopt L1-norm to measure whether the probability has reached a stable state, which is assumed if the value of $p_{t+1}-p_t$ is less than 10^{-10}. The value of each p_{t+1} is the transition probability from seed node i to its neighbor nodes.

3 Results

3.1 Leave One Out Cross Validation (LOOCV)

In this study, we adopt LOOCV and 10-fold cross validation to evaluate the performance of our computational method. For a given miRNA i, there are many associations between circRNA and miRNA. We select one miRNA i related circRNA j each time to serve as the hidden associations, using this hidden data as a test dataset and treating the rest of the known associations as training data to predict the excluded data. For example, if there are n associations between

miRNA i and the entire set of circRNAs, each remaining known association is regarded as test data. In addition, an extra iteration is carried out where we let all of the unknown associations serve as test data. Each time, we obtain the probability of miRNA i being related to circRNA j. Finally, the results we obtain from each iteration can be combined into a probability vector, and then the probability values ranked in descending order to change the threshold. When the probability is greater than the threshold, there could be an association between miRNA i and circRNA j. According to the changing threshold value, we can draw the Receiver Operating Characteristic (ROC) curve and calculate the Area Under Curve (AUC) value, which can then be utilized to measure the accuracy of each prediction result.

Using this method, after ranking all of the association scores between each miRNA and circRNA, we obtained a final AUC value of 0.920, which is shown in Fig. 2 to be superior to other computational methods[57–59]. Some disease-related miRNA-circRNA prediction results are denoted in Fig. 3, which show that the performance of our proposed algorithm is far better than conventional network algorithms. To obtain more reliable evaluation results, a 10-fold cross validation is also adopted to illustrate the performance of our computational method, the results of which is shown in Fig. 4.

To further investigate the performance of our proposed method, we adopt the prediction rate of associations between some special diseases related miRNAs and circRNAs, such as hsa-miR-200b-3p, hsa-miR-30e-5p, and hsa-miR-320b. We compare the number of correct miRNA-circRNA associations achieved in the top 10, 20, and 50 predictions using different computational methods, obtaining the

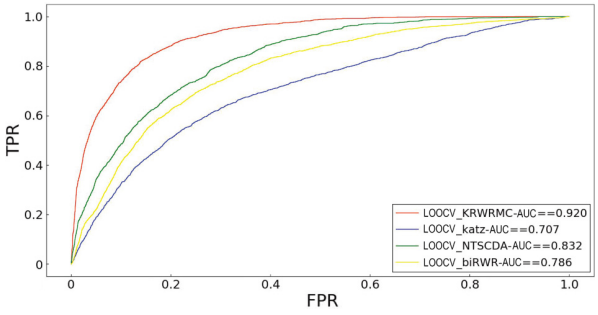

Fig. 2 AUC value and AOC curve of our method. TPR represents true positive rate and FPR represents false positive rate.

results shown in Figs. 5 – 7. The results show that our method's prediction of miRNA hsa-miR-107, which has strong connection with breast cancer[60,61], is superior to other computational methods; in the top 50 predictions, KATZ, Bi-RWR, NTS, and RWRKNN pick out 30, 33, 37, and 40 existing associations, respectively. Additionally, our method also achieves excellent results in the prediction of hsa-miR-30e-5p, which has a strong relationship with schizophrenia[62,63]; in the top 50 predictions, KATZ, Bi-RWR, NTS, and RWRKNN pick out 14, 13, 21, and 37 real associations, respectively. According to these biological experiment results, we can conclude that our proposed computational method performs reliably in predicting the potential miRNA-circRNA associations.

3.2 Parameter effects

There are several parameters that need to be set in this study, some of which have been already mentioned above. The parameter η controls the trend of one subnetwork conveying to another. For instance, if η is more than 0.5, the seed node has a higher probability to walk on the miRNA subnetwork; otherwise, the seed node may prefer to walk on the circRNA subnetwork. The parameter γ is the restarting probability, which controls the likelihood of returning to the beginning node. Poor results will be obtained if γ is less than 0.5. The final parameter K is the number of neighbors of miRNA or circRNA, which be applied to the multi-label based on k-NN algorithm. To optimize the model parameters, Fig. 8 shows the results obtained using different K values, with the values of other parameters fixed. From this we see that KRWRMC achieves optimal performance when the K value is 6. From previous work[47,54], we find that different restart probability values have very small effects on the results. Parameter η being set to 0.7 has demonstrated excellent performance in prior work. Therefore, we set the five parameters to the following values: $\zeta = 0.9$, $\theta = 0.8$, $\eta = 0.1$, $\gamma = 0.7$, and $K = 6$.

4 Conclusion

In this study, we propose a novel computational method called KRWRMC to predict associations between miRNA and circRNA. KRWRMC is based on a heterogeneous network consisting of three subnetworks: the miRNA similarity subnetwork based on GO similarity, the circRNA similarity subnetwork

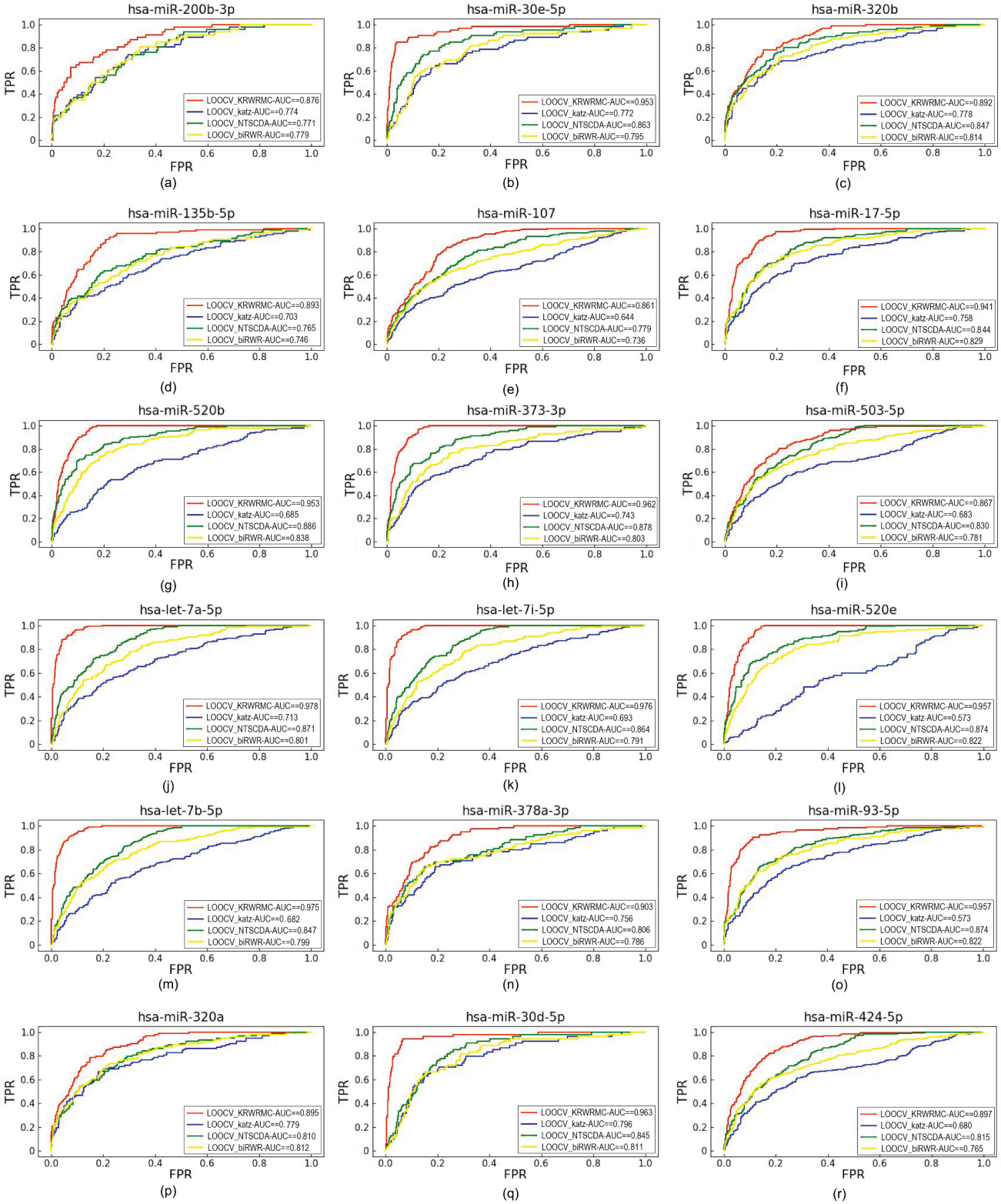

Fig. 3 KRWRMC is compared with other methods by LOOCV.

based on GO similarity, and the miRNA-circRNA association subnetwork. An essential factor behind the superior performance of KRWRMC compared to traditional computational methods is that we take miRNA-circRNA associations into consideration, which means that we not only measure the similarity between the miRNAs and circRNAs based on GO data but also take the effects on their similarity of miRNA or

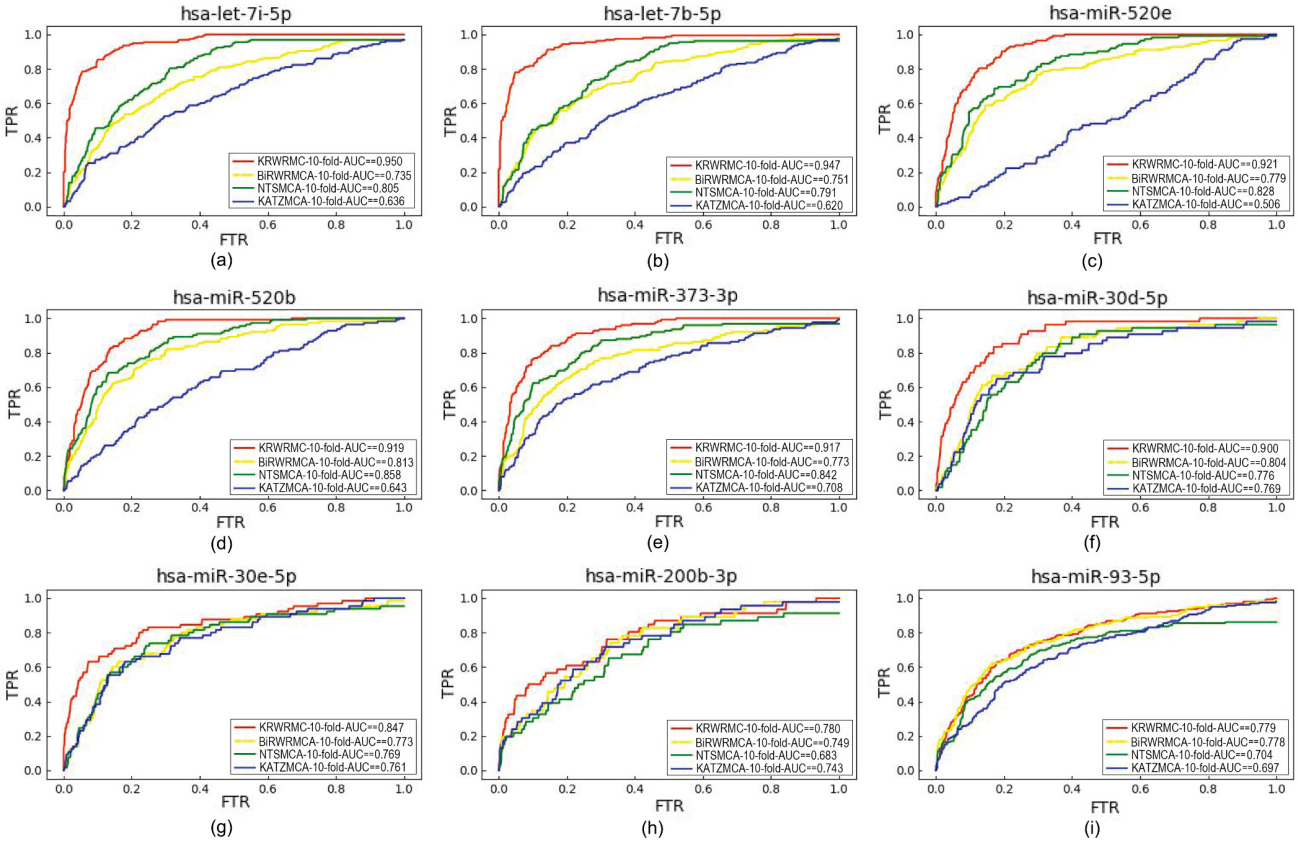

Fig. 4 KRWRMC is compared with other methods by 10-fold cross validation.

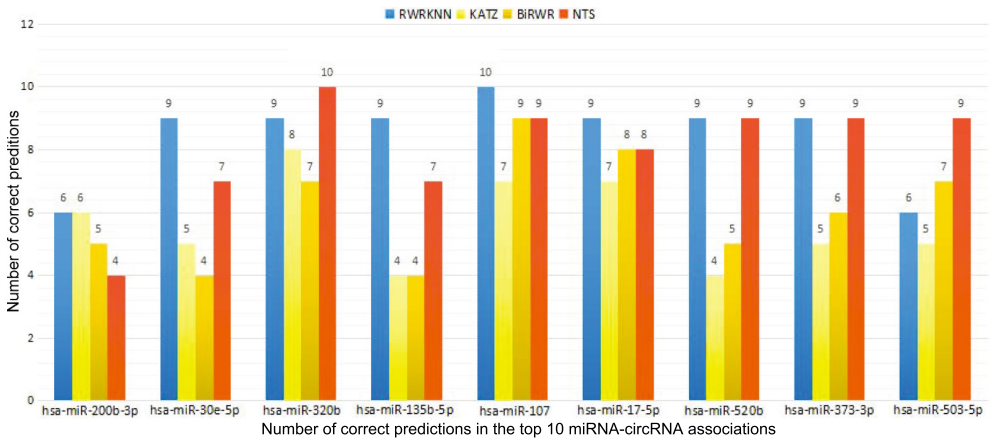

Fig. 5 Number of correct predictions in the top 10 miRNA-circRNA associations.

circRNA neighbors into account, applying a multi-label learning algorithm to evaluate how much influences are made by these neighbors. Considering the influence of biological molecules that can enhance associations upon the same class provides greater capacity to predict some potential associations. The performance of our method is evaluated by LOOCV and 10-fold cross validation, results of which show KRWRMC to be an accurate and reliable computational method for predicting miRNA-circRNA associations.

Although the results of LOOCV and 10-fold cross validation are much better than existing computational methods, there are still some limitations in our model. First and foremost, many parameters need to be confirmed in this model, and future work is needed to identify the most suitable parameter values. Second, based on the basic property of the RWR algorithm, KRWRMC would not be suitable for a situation in which there is no known association, and it cannot be used to infer associations which do not have any

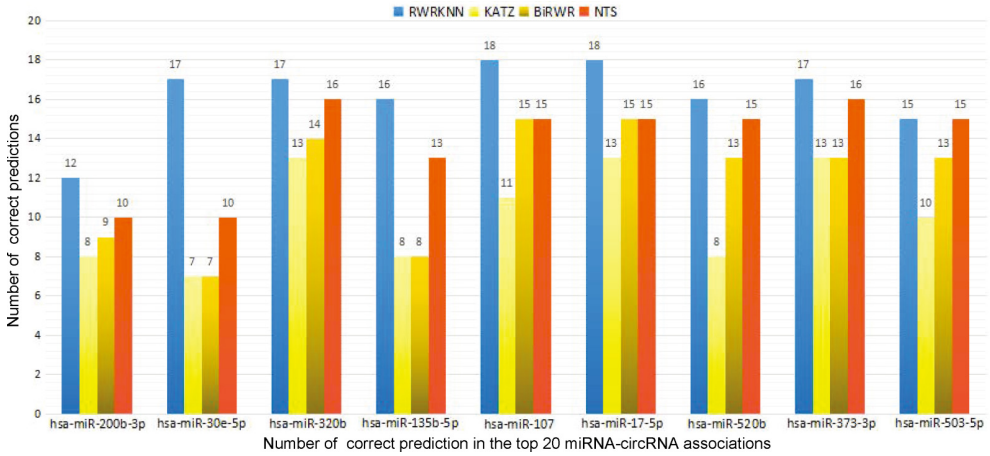

Fig. 6 Number of correct predictions in the top 20 miRNA-circRNA associations.

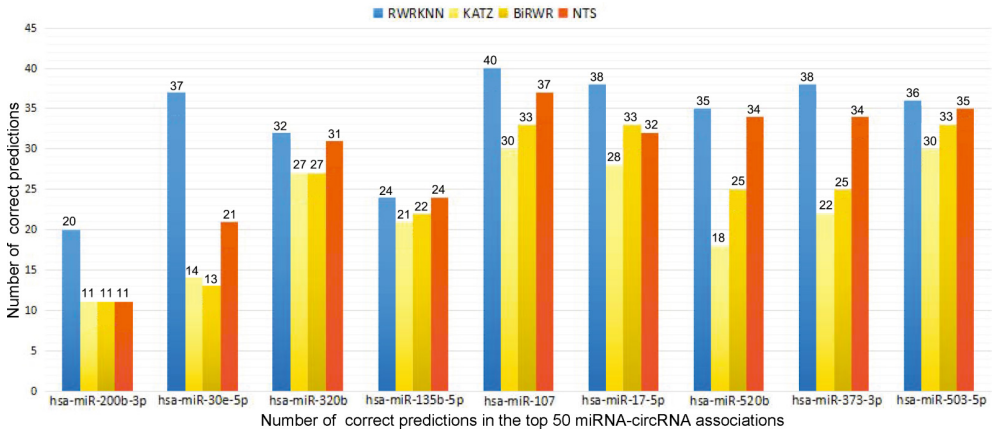

Fig. 7 Number of correct predictions in the top 50 miRNA-circRNA associations.

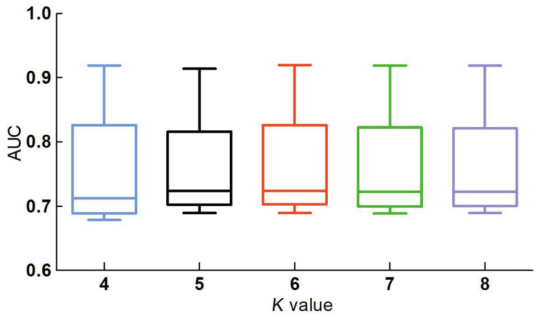

Fig. 8 According to the top 50 average AUC values based on the different *K* values, the influence of parameter *K* of the model performance is analyzed.

GO data. Furthermore, we will take more biological data into account in future work, which can help our model become more reliable and provide a better comprehension of the biological perspective.

Acknowledgment

This paper was supported by the National Natural Science Foundation of China (No. 61672334) and the Fundamental Research Funds for the Central Universities (No. GK201901010).

References

[1] K. Andreeva and N. G. F. Cooper, Circular RNAs: New players in gene regulation, *Adv. Biosci. Biotechnol.*, vol. 6, no. 6, pp. 433–441, 2015.

[2] H. L. Sanger, G. Klotz, D. Riesner, H. J. Gross, and A. K. Kleinschmidt, Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures, *Proc. Natl. Acad. Sci. USA*, vol. 73, no. 11, pp. 3852–3856, 1976.

[3] J. Salzman, C. Gawad, P. L. Wang, N. Lacayo, and P. O. Brown, Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types, *PLoS One*, vol. 7, no. 2, p. e30733, 2012.

[4] S. B. Qu, X. S. Yang, X. L. Li, J. L. Wang, Y. Gao, R. Z. Shang, W. Sun, K. F. Dou, and H. M. Li, Circular RNA: A new star of noncoding RNAs, *Cancer Lett.*, vol. 365, no. 2, pp. 141–148, 2015.

[5] W. R. Jeck and N. E. Sharpless, Detecting and

characterizing circular RNAs, *Nat. Biotechnol.*, vol. 32, no. 5, pp. 453–461, 2014.

[6] P. L. Wang, Y. Bao, M. C. Yee, S. P. Barrett, G. J. Hogan, M. N. Olsen, J. R. Dinneny, P. O. Brown, and J. Salzman, Circular RNA is expressed across the eukaryotic tree of life, *PLoS One*, vol. 9, no. 4, p. e95116, 2014.

[7] Y. C. Liu, Y. J. Chiu, J. R. Li, C. H. Sun, C. C. Liu, and H. D. Huang, Biclustering of transcriptome sequencing data reveals human tissue-specific circular RNAs, *BMC Genomics*, vol. 19, no. S1, p. 958, 2018.

[8] T. Y. Xu, J. Wu, P. Han, Z. M. Zhao, and X. F. Song, Circular RNA expression profiles and features in human tissues: A study using RNA-seq data, *BMC Genomics*, vol. 18, no. S6, p. 680, 2017.

[9] T. B. Hansen, T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard, and J. Kjems, Natural RNA circles function as efficient microRNA sponges, *Nature*, vol. 495, no. 7441, pp. 384–388, 2013.

[10] L. F. Thomas and P. Sætrom, Circular RNAs are depleted of polymorphisms at microRNA binding sites, *Bioinformatics*, vol. 30, no. 16, pp. 2243–2246, 2014.

[11] F. R. Kulcheski, A. P. Christoff, and R. Margis, Circular RNAs are miRNA sponges and can be used as a new class of biomarker, *J. Biotechnol.*, vol. 238, pp. 42–51, 2016.

[12] S. Memczak, M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, et al., Circular RNAs are a large class of animal RNAs with regulatory potency, *Nature*, vol. 495, no. 7441, pp. 333–338, 2013.

[13] Y. Zhang, X. O. Zhang, T. Chen, J. F. Xiang, Q. F. Yin, Y. H. Xing, S. S. Zhu, L. Yang, and L. L. Chen, Circular intronic long noncoding RNAs, *Mol. Cell*, vol. 51, no. 6, pp. 792–806, 2013.

[14] Z. Y. Li, C. Huang, C. Bao, L. Chen, M. Lin, X. L. Wang, G. L. Zhong, B. Yu, W. C. Hu, L. M. Dai, et al., Exon-intron circular RNAs regulate transcription in the nucleus, *Nat. Struct. Mol. Biol.*, vol. 22, no. 3, pp. 256–264, 2015.

[15] R. Ashwal-Fluss, M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evantal, S. Memczak, N. Rajewsky, and S. Kadener, circRNA biogenesis competes with pre-mRNA splicing, *Mol. Cell*, vol. 56, no. 1, pp. 55–66, 2014.

[16] A. M. Cheng, M. W. Byrom, J. Shelton, and L. P. Ford, Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis, *Nucl. Acids Res.*, vol. 33, no. 4, pp. 1290–1297, 2005.

[17] A. C. Mallory, D. V. Dugas, D. P. Bartel, and B. Bartel, MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs, *Curr. Biol.*, vol. 14, no. 12, pp. 1035–1046, 2004.

[18] Y. M. Cai, X. M. Yu, S. N. Hu, and J. Yu, A brief review on the mechanisms of miRNA regulation, *Genomics Proteomics Bioinformatics*, vol. 7, no. 4, pp. 147–154, 2009.

[19] N. Yanaihara, N. Caplen, E. Bowman, M. Seike, K. Kumamoto, M. Yi, R. M. Stephens, A. Okamoto, J. Yokota, T. Tanaka, et al., Unique microRNA molecular profiles in lung cancer diagnosis and prognosis, *Cancer Cell*, vol. 9, no. 3, pp. 189–198, 2006.

[20] D. D. Taylor and C. Gercel-Taylor, MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer, *Gynecol. Oncol.*, vol. 110, no. 1, pp. 13–21, 2008.

[21] L. Zhang, J. Huang, N. Yang, J. Greshock, M. S. Megraw, A. Giannakakis, S. Liang, T. L. Naylor, A. Barchetti, M. R. Ward, et al., microRNAs exhibit high frequency genomic alterations in human cancer, *Proc. Natl. Acad. Sci. USA*, vol. 103, no. 24, pp. 9136–9141, 2006.

[22] E. J. Lee, Y. Gusev, J. M. Jiang, G. J. Nuovo, M. R. Lerner, W. L. Frankel, D. L. Morgan, R. G. Postier, D. J. Brackett, and T. D. Schmittgen, Expression profiling identifies microRNA signature in pancreatic cancer, *Int. J. Cancer*, vol. 120, no. 5, pp. 1046–1054, 2007.

[23] S. Köhler, S. Bauer, D. Horn, and P. N. Robinson, Walking the interactome for prioritization of candidate disease genes, *Am. J. Hum. Genet.*, vol. 82, no. 4, pp. 949–958, 2008.

[24] H. Yang, P. N. Robinson, and K. Wang, Phenolyzer: Phenotype-based prioritization of candidate genes for human diseases, *Nat. Methods*, vol. 12, no. 9, pp. 841–843, 2015.

[25] D. Shriner, T. M. Baye, M. A. Padilla, S. J. Zhang, L. K. Vaughan, and A. E. Loraine, Commonality of functional annotation: A method for prioritization of candidate genes from genome-wide linkage studies, *Nucl. Acids Res.*, vol. 36, no. 4, p. e26, 2008.

[26] Q. Zou, J. J. Li, Q. Q. Hong, Z. Y. Lin, Y. Wu, H. Shi, and Y. Ju, Prediction of microRNA-disease associations based on social network analysis methods, *BioMed Res. Int.*, vol. 2015, pp. 810514, 2015.

[27] X. Chen, M. X. Liu, and G. Y. Yan, RWRMDA: Predicting novel human microRNA-disease associations, *Mol. BioSyst.*, vol. 8, no. 10, pp. 2792–2798, 2012.

[28] X. X. Zeng, X. Zhang, and Q. Zou, Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks, *Brief. Bioinform.*, vol. 17, no. 2, pp. 193–203, 2016.

[29] X. Chen, Q. F. Wu, and G. Y. Yan, RKNNMDA: Ranking-based KNN for miRNA-disease association prediction, *RNA Biol.*, vol. 14, no. 7, pp. 952–962, 2017.

[30] J. Y. Wang, R. X. Ma, W. Ma, J. Chen, J. C. Yang, Y. G. Xi, and Q. H. Cui, LncDisease: A sequence based bioinformatics tool for predicting lncRNA-disease associations, *Nucl. Acids Res.*, vol. 44, no. 9, p. e90, 2016.

[31] X. F. Yang, L. Gao, X. L. Guo, X. H. Shi, H. Wu, F. Song, and B. B. Wang, A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases, *PLoS One*, vol. 9, no. 1, p. e87797, 2014.

[32] M. A. Doyle, R. B. Gasser, B. J. Woodcroft, R. S. Hall, and S. A. Ralph, Drug target prediction and prioritization: Using orthology to predict essentiality in parasite genomes, *BMC Genomics*, vol. 11, p. 222, 2010.

[33] F. Prado-Prado, X. García-Mera, P. Abeijón, N. Alonso, O. Caamaño, M. Yáñez, T. Gárate, M. Mezo, M. González-Warleta, L. Muiño, et al., Using entropy of drug and protein graphs to predict FDA drug-target network: Theoretic-experimental study of MAO inhibitors and hemoglobin peptides from *Fasciola hepatica*, *Eur. J. Med. Chem.*, vol. 46, no. 4, pp. 1074–1094, 2011.

[34] X. Chen, M. X. Liu, Q. H. Cui, and G. Y. Yan, Prediction of disease-related interactions between microRNAs and environmental factors based on a semi-supervised classifier, *PLoS One*, vol. 7, no. 8, p. e43425, 2012.

[35] Y. A. Huang, K. C. C. Chan, and Z. H. You, Constructing prediction models from expression profiles for large scale lncRNA-miRNA interaction profiling, *Bioinformatics*, vol. 34, no. 5, pp. 812–819, 2018.

[36] J. H. Li, S. Liu, H. Zhou, L. H. Qu, and J. H. Yang, starBase v2.0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, *Nucl. Acids Res.*, vol. 42, no. D1, pp. D92–D97, 2014.

[37] C. H. Chou, N. W. Chang, S. Shrestha, S. D. Hsu, Y. L. Lin, W. H. Lee, C. D. Yang, H. C. Hong, T. Y. Wei, S. J. Tu, et al., miRTarBase 2016: Updates to the experimentally validated miRNA-target interactions database, *Nucl. Acids Res.*, vol. 44, no. D1, pp. D239–D247, 2016.

[38] S. D. Hsu, F. M. Lin, W. Y. Wu, C. Liang, W. C. Huang, W. L. Chan, W. T. Tsai, G. Z. Chen, C. J. Lee, C. M. Chiu, et al., miRTarBase: A database curates experimentally validated microRNA-target interactions, *Nucl. Acids Res.*, vol. 39, no. S1, pp. D163–D169, 2011.

[39] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, et al., Human protein reference database—2009 update, *Nucl. Acids Res.*, vol. 37, no. S1, pp. D767–D772, 2009.

[40] T. Price, F. I. Peña III, and Y. R. Cho, Survey: Enhancing protein complex prediction in PPI networks with GO similarity weighting, *Interdiscip. Sci. Comput. Life Sci.*, vol. 5, no. 3, pp. 196–210, 2013.

[41] P. Glažar, P. Papavasileiou, and N. Rajewsky, circBase: A database for circular RNAs, *RNA*, vol. 20, no. 11, pp. 1666–1670, 2014.

[42] Y. J. Li and J. Y. Li, Disease gene identification by random walk on multigraphs merging heterogeneous genomic and phenotype data, *BMC Genomics*, vol. 13, no. S7, p. S27, 2012.

[43] J. Zhu, Y. F. Qin, T. G. Liu, J. Wang, and X. Q. Zheng, Prioritization of candidate disease genes by topological similarity between disease and protein diffusion profiles, *BMC Bioinform.*, vol. 14, no. Suppl 5, p. S5, 2013.

[44] C. F. Sun, Q. Z. Li, L. Z. Cui, H. Li, and Y. L. Shi, Heterogeneous network-based chronic disease progression mining, *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 25–34, 2019.

[45] Y. S. Liu, X. X. Zeng, Z. Y. He, and Q. Zou, Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 14, no. 4, pp. 905–915, 2017.

[46] H. L. Chen and Z. P. Zhang, Prediction of associations between OMIM diseases and microRNAs by random walk on OMIM disease similarity network, *Sci. World J.*, vol. 2013, p. 204658, 2013.

[47] J. Sun, H. B. Shi, Z. Z. Wang, C. J. Zhang, L. Liu, L. T. Wang, W. W. He, D. P. Hao, S. L. Liu, and M. Zhou, Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network, *Mol. BioSyst.*, vol. 10, no. 8, pp. 2074–2081, 2014.

[48] X. Chen, Z. H. You, G. Y. Yan, and D. W. Gong, IRWRLDA: Improved random walk with restart for lncRNA-disease association prediction, *Oncotarget*, vol. 7, no. 36, pp. 57919–57931, 2016.

[49] M. Zhou, X. J. Wang, J. W. Li, D. P. Hao, Z. Z. Wang, H. B. Shi, L. Han, H. Zhou, and J. Sun, Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network, *Mol. BioSyst.*, vol. 11, no. 3, pp. 760–769, 2015.

[50] X. Chen, M. X. Liu, and G. Y. Yan, Drug-target interaction prediction by random walk on the heterogeneous network, *Mol. BioSyst.*, vol. 8, no. 7, pp. 1970–1978, 2012.

[51] L. C. Chen, H. Y. Yeh, C. Y. Yeh, C. R. Arias, and V. W. Soo, Identifying co-targets to fight drug resistance based on a random walk model, *BMC Syst. Biol.*, vol. 6, p. 5, 2012.

[52] G. D. Guo, H. Wang, D. Bell, Y. X. Bi, and K. Greer, KNN model-based approach in classification, in *Proc. OTM Confederated Int. Conf. the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, Catania, Italy, 2003, pp. 986–996.

[53] X. Xia, X. H. Yang, S. P. Li, C. Wu, and L. L. Zhou, RW.KNN: A proposed random walk KNN algorithm for multi-label classification, in *Proc. 4^{th} Workshop on Workshop for PhD Students in Information & Knowledge Management*, Glasgow, UK, 2011, pp. 87–90.

[54] Y. J. Li and J. C. Patra, Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network, *Bioinformatics*, vol. 26, no. 9, pp. 1219–1224, 2010.

[55] J. Weston, A. Elisseeff, D. Y. Zhou, C. S. Leslie, and W. S. Noble, Protein ranking: From local to global structure in the protein similarity network, *Proc. Natl. Acad. Sci. USA*, vol. 101, no. 17, pp. 6559–6563, 2004.

[56] T. Can, O. Çamoglu, and A. K. Singh, Analysis of protein-protein interaction networks using random walks, in *Proc. 5^{th} Int. Workshop on Bioinformatics*, Chicago, IL, USA, 2005, pp. 61–68.

[57] G. U. Ganegoda, J. X. Wang, F. X. Wu, and M. Li,

Prediction of disease genes using tissue-specified gene-gene network, *BMC Syst. Biol.*, vol. 8, no. Suppl 3, p. S3, 2014.

[58] D. D. Sun, A. Li, H. Q. Feng, and M. H. Wang, NTSMDA: Prediction of miRNA-disease associations by integrating network topological similarity, *Mol. BioSyst.*, vol. 12, no. 7, pp. 2224–2232, 2016.

[59] M. Q. Xie, T. Hwang, and R. Kuang, Prioritizing disease genes by bi-random walk, in *Proc. 16th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining*, Kuala Lumpur, Malaysia, 2012, pp. 292–303.

[60] M. G. Schrauder, R. Strick, R. Schulz-Wendtland, P. L. Strissel, L. Kahmann, C. R. Loehberg, M. P. Lux, S. M. Jud, A. Hartmann, A. Hein, et al., Circulating micro-RNAs as potential blood-based markers for early stage breast cancer detection, *PLoS One*, vol. 7, no. 1, p. e29770, 2012.

[61] C. Polytarchou, D. Iliopoulos, and K. Struhl, An integrated transcriptional regulatory circuit that reinforces the breast cancer stem cell state, *Proc. Natl. Acad. Sci. USA*, vol. 109, no. 36, pp. 14470–14475, 2012.

[62] X. Y. Sun, J. Lu, L. Zhang, H. T. Song, L. Zhao, H. M. Fan, A. F. Zhong, W. Niu, Z. M. Guo, Y. H. Dai, et al., Aberrant microRNA expression in peripheral plasma and mononuclear cells as specific blood-based biomarkers in schizophrenia patients, *J. Clin. Neurosci.*, vol. 22, no. 3, pp. 570–574, 2015.

[63] Y. Xu, F. Li, B. Zhang, K. R. Zhang, F. Q. Zhang, X. Z. Huang, N. Sun, Y. Ren, M. Q. Sui, and P. Z. Liu, MicroRNAs and target site screening reveals a pre-microRNA-30e variant associated with schizophrenia, *Schizophr. Res.*, vol. 119, nos. 1–3, pp. 219–227, 2010.

Xiujuan Lei received the PhD degree from Northwestern Polytechnical University, China, in 2005. She is currently a professor at College of Computer Science, Shaanxi Normal University, China. Her current research interests include bioinformatics, intelligent computing, pattern recognition, and data mining.

Zengqiang Fang is currently pursuing the MS degree in Shaanxi Normal University, China. His research interests include machine learning, bioinformatics, and data mining.