



OPEN ACCESS

EDITED BY

George Tsiamis,
University of Patras,
Greece

REVIEWED BY

Mete Yilmaz,
Bursa Technical University,
Turkey
ChangQing Yu,
Xijing University,
China

*CORRESPONDENCE

Shunhan Yao
✉ 2128402005@st.gxu.edu.cn

SPECIALTY SECTION

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 06 July 2022

ACCEPTED 30 November 2022

PUBLISHED 22 December 2022

CITATION

Yao D, Nong L, Qin M, Wu S and
Yao S (2022) Identifying circRNA-miRNA
interaction based on multi-biological
interaction fusion.
Front. Microbiol. 13:987930.
doi: 10.3389/fmicb.2022.987930

COPYRIGHT

© 2022 Yao, Nong, Qin, Wu and Yao. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Identifying circRNA-miRNA interaction based on multi-biological interaction fusion

Dunwei Yao^{1,2}, Lidan Nong³, Minzhen Qin^{1,2},
Shengbin Wu^{2,4} and Shunhan Yao^{5*}

¹Department of Gastroenterology, The People's Hospital of Baise, Baise, China, ²The Southwest Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China, ³Department of Child Healthcare, Baise Maternal and Child Hospital, Baise, China, ⁴Department of Pulmonary and Critical Care Medicine, The People's Hospital of Baise, Baise, China, ⁵Medical College of Guangxi University, Nanning, China

CircRNA is a new type of non-coding RNA with a closed loop structure. More and more biological experiments show that circRNA plays important roles in many diseases by regulating the target genes of miRNA. Therefore, correct identification of the potential interaction between circRNA and miRNA not only helps to understand the mechanism of the disease, but also contributes to the diagnosis, treatment, and prognosis of the disease. In this study, we propose a model (IIMCCMA) by using network embedding and matrix completion to predict the potential interaction of circRNA-miRNA. Firstly, the corresponding adjacency matrix is constructed based on the experimentally verified circRNA-miRNA interaction, circRNA-cancer interaction, and miRNA-cancer interaction. Then, the Gaussian kernel function and the cosine function are used to calculate the circRNA Gaussian interaction profile kernel similarity, circRNA functional similarity, miRNA Gaussian interaction profile kernel similarity, and miRNA functional similarity. In order to reduce the influence of noise and redundant information in known interactions, this model uses network embedding to extract the potential feature vectors of circRNA and miRNA, respectively. Finally, an improved inductive matrix completion algorithm based on the feature vectors of circRNA and miRNA is used to identify potential interactions between circRNAs and miRNAs. The 10-fold cross-validation experiment is utilized to prove the predictive ability of the IIMCCMA. The experimental results show that the AUC value and AUPR value of the IIMCCMA model are higher than other state-of-the-art algorithms. In addition, case studies show that the IIMCCMA model can correctly identify the potential interactions between circRNAs and miRNAs.

KEYWORDS

circRNA-miRNA interaction, multi-biological interaction fusion, inductive matrix completion, network embedding, computational method

1. Introduction

Different from traditional linear non-coding RNA, circRNA is a new type of non-coding RNA with a closed loop structure (3' and 5' in circRNA are connected together; Wilusz and Sharp, 2013; Lan et al., 2015b). The unique molecular structure of circRNA ensures that it cannot be affected by RNA exonuclease. In addition, the expression of circRNA is more stable and not easily degraded than other linear non-coding RNA. Further experiments proved that circRNA is rich in miRNA binding sites, which can act as a miRNA sponge in cells to splice, transcribe, and modify the expression of parental genes (Qu et al., 2015; Rybak-Wolf et al., 2015).

Recent experimental results show that circRNA plays an important role in many diseases. For example, quantitative real-time PCR (qRT-PCR) detection found that circRNA BCRC-3 is low expressed in bladder cancer tissue cells. Moreover, circRNA BCRC-3 can directly bind to miRNA miR-182-5p, and then act as a sponge for miRNA miR-182-5p to promote the activity of its target genes. Therefore, circRNA BCRC3 can be used as a tumor suppressor to inhibit the proliferation of bladder cancer cells (Xie et al., 2018). The expression of circRNA hsa_circ_0008068 is significantly down-regulated in prostate cancer cells. There are multiple binding sites between the circRNA and the anticancer miRNA miR-145-3p. CircRNA hsa_circ_0008068 can play an anti-cancer role in prostate cancer cells by regulating miR-145-3p and its target gene WISP1. Therefore, circRNA hsa_circ_000806 may be an important target for the diagnosis and treatment of prostate cancer (Zheng et al., 2020).

With the continuous development of high-throughput sequencing technology, more and more circRNA-miRNA-disease interactions have been confirmed. At the same time, a large number of databases have been developed to store the basic information of circRNA and interactions related to circRNA such as circBase (Glažar et al., 2014), circBank (Liu et al., 2019), circad (Rophina et al., 2020), and circR2Cancer (Lan et al., 2020c). As a benchmark database in the circRNA field, the circBase database stores basic information related to circRNA such as the position of circRNA, the genomic length, the spliced sequence length, and the gene symbol (Glažar et al., 2014). circBank is a professional database dedicated to standardizing circRNA naming (Liu et al., 2019). This database not only provides basic information about circRNA, but also names some newly discovered circRNAs uniformly. The Circad database collects 1,388 experimentally verified circRNA-disease interactions from five different species (Homo sapiens, mice, rats, chickens, and wild boars; Rophina et al., 2020). CircR2Cancer is a new database that stores circRNA-cancer interactions. This database not only stores experimentally verified circRNA-cancer interactions but also circRNA-miRNA interactions and miRNA-cancer interactions (Lan et al., 2020c). In addition to storing experimentally verified interactions, the circR2Cancer database also stores basic information about circRNA and diseases.

The emergence of circRNA-related databases provides a data basis for circRNA-related interaction prediction based on computational methods. Compared with traditional biological identification method, the interaction prediction model based on the computational method has higher accuracy and less time consumption. Guo et al. (2022) presented a computational model to predict circRNA-miRNA interactions by using Word2vec, Structural Deep Network Embedding, Convolutional Neural Network, and Deep Neural Network. Qian et al. (2022) proposed a computational model (CMASG) for circRNA-miRNA interactions prediction based on graph neural network and singular value decomposition. It utilized the graph neural network to learn feature representations of nodes and the lightGBM to predict circRNA-miRNA association. Lan et al. (2021b) developed a computational framework (NECMA) to identify interactions between circRNAs and miRNAs by using network embedding. It extracted features of circRNA and miRNA based on network embedding and predict circRNA-miRNA associations based on neighborhood regularization logic matrix decomposition and inner product. He et al. (2022) proposed a computational approach (GCNCMI) to predict the potential interactions between circRNAs and miRNAs based on graph convolutional neural network. It used the graph convolutional neural network to exact the potential interactions of adjacent nodes and then utilized the embedded representations generated by each layer to predict the final score. Qian et al. (2021) introduced a computational framework (CMIVGSD), to predict circRNA-miRNA interaction by using singular value decomposition and graph variational auto-encoders. Yu et al. (2022) proposed a computational model (SGCNCMI) to identify circRNA-miRNA interactions by combining multimodal information and graph convolutional neural network. Wang et al. (2022) presented a computing method (KGDCMI) to predict the interactions between circRNA and miRNA based on multi-source information fusion. It exacts RNA attribute information from sequence and similarity and captures the behavior information in RNA association based on graph-embedding algorithm. Then, the principal component analysis is used to obtain feature vector, and further the deep neural network is utilized to identify potential circRNA-miRNA interactions. Fang and Lei (2019) fused circRNA-miRNA interaction network, circRNA functional similarity network, and miRNA functional similarity network to construct a circRNA-miRNA heterogeneous network. Then use the K-nearest neighbor algorithm based on restart random walk to predict the potential interaction of circRNA and miRNA.

In this paper, we propose a circRNA-miRNA interaction prediction model (IIMCCMA) based on multi-biological interaction data. This model uses experimentally verified circRNA-miRNA interaction, circRNA-cancer interaction, and miRNA-cancer interaction to construct circRNA-miRNA adjacency matrix, circRNA-cancer adjacency matrix, and miRNA-cancer adjacency matrix, respectively. On the basis of the above adjacency matrix, this model uses Gaussian kernel function and cosine function to calculate circRNA GIP kernel similarity and circRNA functional similarity, as well as miRNA GIP kernel similarity and miRNA functional similarity. In order to reduce the negative impact of noise or redundant information in the known circRNA-miRNA interaction

on the prediction model, the IIMCCMA model first uses the known circRNA-miRNA interaction to construct a heterogeneous network. Then we use the network embedding algorithm to extract the potential feature vectors of circRNA and miRNA in heterogeneous networks. In order to make full use of the information contained in different data sources, this model uses a feature fusion method to integrate the similarity features and topological features of entities in the interaction network to form circRNA fusion features and miRNA fusion features, respectively. Finally, on the basis of circRNA fusion features and miRNA fusion features, an improved inductive matrix completion algorithm is used to predict the potential interaction of circRNA and miRNA. The 10-fold cross-validation experiment was used to evaluate the predictive performance of the IIMCCMA model. The experimental results show that the IIMCCMA model achieves better performance than other advanced interaction prediction models. In addition, the case study results show that the IIMCCMA model can correctly identify the potential interaction between circRNA and miRNA.

2. Materials and methods

2.1 Materials

We use two datasets as gold standard set inhere which is downloaded from circR2Cancer (Lan et al., 2020c) and KGNACDA (Lan et al., 2022a). In dataset 1, there are 756 interactions between 514 circRNAs and 461 miRNAs, 647 interactions between 514 circRNAs and 62 cancers, and 732 interactions between 461 miRNAs and 62 cancers. In dataset 2, there are 330 circRNAs, 79 diseases and 245 miRNAs, 346 circRNA-disease interactions, 146 circRNA-miRNA interactions, and 106 miRNA-disease interactions. Further, we construct an adjacency matrix to represent the above-mentioned interaction network. The adjacency matrix CM represents the circRNA-miRNA interactions. If circRNA CM_i is related to miRNA CM_j , $CM(i,j) = 1$, otherwise, $CM(i,j) = 0$. Similarly, the adjacency matrix CC represents the circRNA-cancer interactions. If circRNA CC_i is related to cancer CC_j , $CC(i,j) = 1$, otherwise, $CC(i,j) = 0$. The adjacency matrix MC represents the miRNA-cancer interactions. If miRNA is related to cancer MC_j , $MC(i,j) = 1$, otherwise, $MC(i,j) = 0$.

2.2 circRNA and miRNA similarity calculation

Based on the assumption that circRNAs with similar functions are often associated with similar miRNAs (Lan et al., 2020a, 2021a, 2022c), circRNA GIP kernel similarity and miRNA GIP kernel similarity are calculated based on the circRNA-miRNA interaction network, respectively. We define GCS to represent the Gaussian interaction profile kernel similarity network of circRNA.

The definition of GIP kernel similarity between circRNA c_i and circRNA c_j is as follows:

$$GCS(c_i, c_j) = \exp\left(-\gamma_{cs} CM(i, :) - CM(j, :)^2\right)$$

$$\gamma_{cs} = 1 / \left(\frac{1}{n_{circ}} \sum_{i=1}^{n_{circ}} CM(i, :)^2 \right)$$

where, $CM(i, :)$ represents the i -th row of the circRNA-miRNA interaction network CM . n_{circ} represents the number of rows of the interaction network CM . γ_{cs} represents the kernel bandwidth.

Similarly, we define GMS to represent the Gaussian interaction profile kernel similarity network of miRNA. The definition of GIP kernel similarity between miRNA m_i and miRNA m_j is as follows:

$$GMS(m_i, m_j) = \exp\left(-\gamma_{ms} CM(:, i) - CM(:, j)^2\right)$$

$$\gamma_{ms} = 1 / \left(\frac{1}{n_{mi}} \sum_{i=1}^{n_{mi}} CM(:, i)^2 \right)$$

where, $CM(:, i)$ represents the i -th column of the circRNA-miRNA interaction network CM . n_{mi} represents the number of columns of the interaction network CM . γ_{ms} represents the kernel bandwidth.

In addition, we also use the cosine function to calculate the circRNA functional similarity and the miRNA functional similarity on the basis of circRNA-cancer interaction network and miRNA-cancer interaction network. The cosine similarity measures the similarity between two vectors by the angle between two vectors in a two-dimensional space. If the two vectors point in the same direction, it means that the two vectors are more similar, otherwise, the similarity is lower. Therefore, according to the above cosine similarity theory, the circRNA functional similarity and miRNA functional similarity are defined as follows:

$$CCS(c_i, c_j) = \frac{\sum_{i=1}^{n_{circ}} CC(i, :) \times CC(j, :)}{\sqrt{\sum_{i=1}^{n_{circ}} CC(i, :)^2} \times \sqrt{\sum_{i=1}^{n_{circ}} CC(j, :)^2}}$$

$$CMS(m_i, m_j) = \frac{\sum_{i=1}^{n_{mi}} MC(i, :) \times MC(j, :)}{\sqrt{\sum_{i=1}^{n_{mi}} MC(i, :)^2} \times \sqrt{\sum_{i=1}^{n_{mi}} MC(j, :)^2}}$$

where CCS and CMS represent the circRNA functional similarity network and the miRNA functional similarity network,

respectively. $CCS(c_i, c_j)$ represents the functional similarity between circRNA c_i and circRNA c_j . $CC(i, :)$ represents the i -th row in the circRNA-cancer network CC . n_{circ} represents the number of rows in the network CC . $CMS(m_i, m_j)$ represents the functional similarity between miRNA m_i and miRNA m_j . $MC(i, :)$ represents the i -th row in the miRNA-cancer network MC . n_{mi} represents the number of rows in the network MC .

In order to make better use of the circRNA and the miRNA similarity characteristics, we integrate the above two similarities to obtain the circRNA similarity $circ_{sim}$ and miRNA similarity mi_{sim} , which are defined as follows:

$$circ_{sim}(c_i, c_j) = \begin{cases} GCS(c_i, c_j), & \text{if } CCS(c_i, c_j) = 0 \\ CCS(c_i, c_j), & \text{otherwise} \end{cases}$$

$$mi_{sim}(m_i, m_j) = \begin{cases} GMS(m_i, m_j), & \text{if } CMS(m_i, m_j) = 0 \\ CMS(m_i, m_j), & \text{otherwise} \end{cases}$$

where $circ_{sim}(c_i, c_j)$ represents the integrated similarity between circRNA c_i and circRNA c_j . $mi_{sim}(m_i, m_j)$ represents the integrated similarity between miRNA m_i and miRNA m_j . GCS represents circRNA GIP kernel similarity. CCS represents the circRNA functional similarity. In the same way, GMS represents miRNA GIP kernel similarity. CMS represents the miRNA functional similarity.

2.3 Potential feature extraction and fusion of circRNA and miRNA

In order to reduce the influence of noise or redundant information in the known circRNA-miRNA interaction network, we construct the heterogeneous network $H_{circ-mi}$. The heterogeneous network is composed of the circRNA-miRNA interaction adjacency matrix CM and the transposed matrix CM^T of the circRNA-miRNA adjacency matrix. It is defined as follows:

$$H_{circ-mi} = \begin{bmatrix} 0 & CM \\ CM^T & 0 \end{bmatrix}$$

After obtaining the heterogeneous network, the NetMF algorithm (Qiu et al., 2018) is used to obtain the circRNA-miRNA latent feature matrix with size equals to $(m+n) \times d$. Among them, m represents the number of circRNA in the heterogeneous network and $H_{circ-mi}$. n represents the number of miRNAs. d represents the dimensions of circRNA and miRNA low-dimensional space vectors. Experiments have verified that the model has the best prediction effect when the dimension of the low-dimensional space vector of circRNA and miRNA is set to 16.

In order to make full use of the information of different interaction, we use a fusion method to fuse the circRNA and miRNA topological features ($circ_{Net}$, mi_{Net}) obtained through the NetMF algorithm with the integrated circRNA similarity features and miRNA similarity features, respectively. The fused information can not only describe the characteristics of different data sources, but also describe the complex relationship between circRNA and miRNA more comprehensively. The fusion feature of circRNA $circ_feature$ and the fusion feature of miRNA $mi_feature$ are defined as follows:

$$circ_f = [circ_{Net}, circ_{sim}]$$

$$mi_f = [mi_{Net}, mi_{sim}]$$

where $circ_{Net}$ and mi_{Net} represent the topological characteristics of circRNA and miRNA based on the NetMF algorithm, respectively. $circ_{sim}$ and mi_{sim} represents the circRNA integrated similarity and miRNA integrated similarity, respectively.

2.4 Prediction of potential interaction between circRNA and miRNA

In this paper, we propose a circRNA-miRNA interaction prediction model (IIMCCMA) based on an improved inductive matrix completion algorithm. This model is implemented based on the known circRNA-miRNA interaction, the fusion feature of circRNA and the fusion feature of miRNA. The specific implementation process of the IIMCCMA model is shown in Figures 1A,B.

Many studies have found that the sparsity problem of biological interaction networks is very serious. Taking the circRNA-miRNA interaction network used in this paper as an example, the circRNA-miRNA interaction network CM is composed of 756 interactions between 514 circRNAs and 461 miRNAs. Obviously, the interaction network CM is very sparse (the matrix density is 0.0032). In addition, in the calculation process of the inductive matrix completion algorithm (Jain and Dhillon, 2013; Lan et al., 2015a; Si et al., 2016; Nazarov et al., 2018), due to the high sparsity of the known interaction matrix, a relatively large amount of effective information will be lost in the process of low-dimensional mapping, which will affect the prediction effect of the circRNA-miRNA potential interaction prediction model. Therefore, in order to alleviate the negative impact of the high sparsity of the interaction network on the model, we modify the mapping method of the low-rank matrix in the inductive matrix completion algorithm. Specifically, in order to better protect the structural information in the sparse matrix, we perform multiple low-dimensional

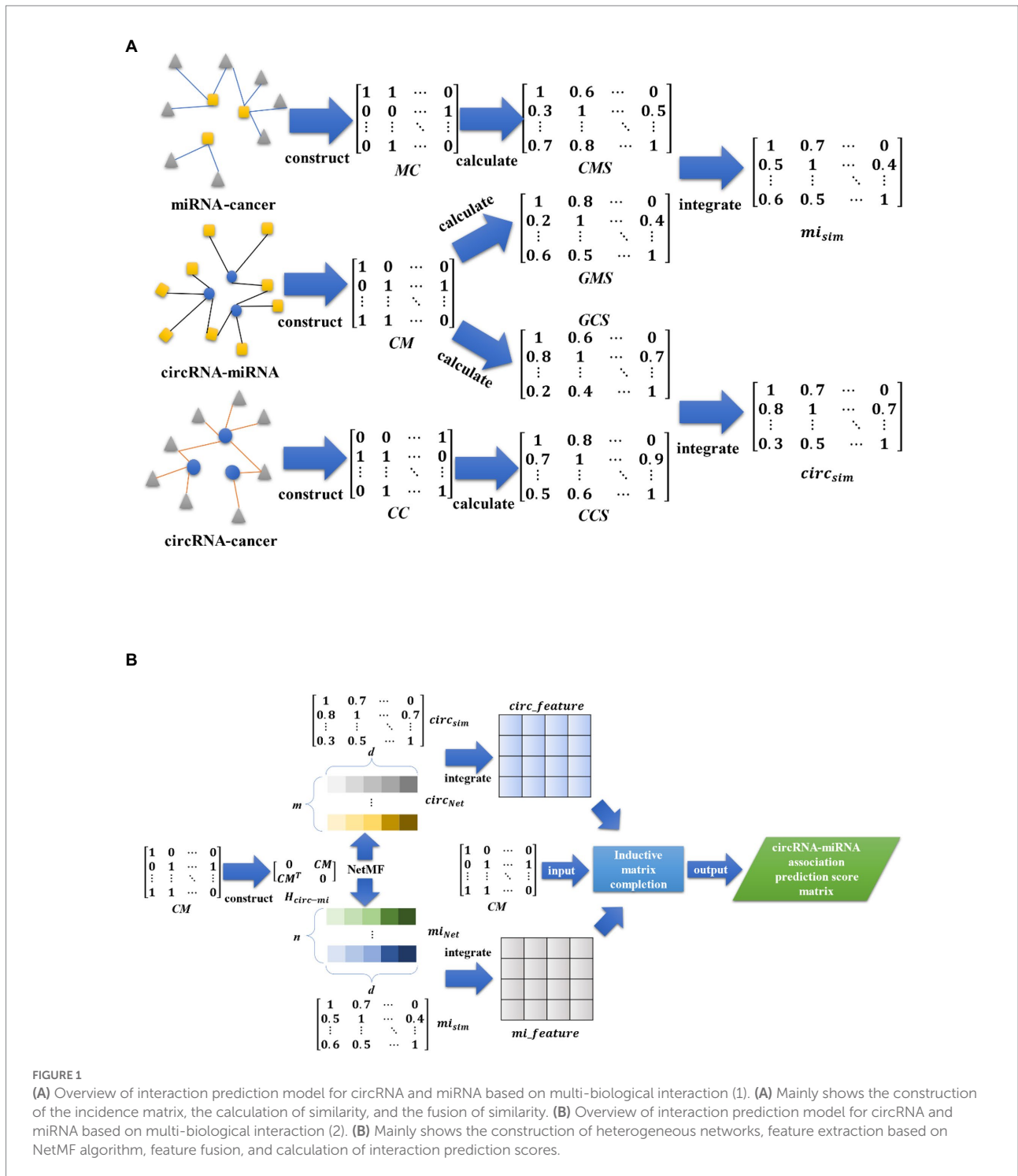


FIGURE 1

(A) Overview of interaction prediction model for circRNA and miRNA based on multi-biological interaction (1). (A) Mainly shows the construction of the incidence matrix, the calculation of similarity, and the fusion of similarity. (B) Overview of interaction prediction model for circRNA and miRNA based on multi-biological interaction (2). (B) Mainly shows the construction of heterogeneous networks, feature extraction based on NetMF algorithm, feature fusion, and calculation of interaction prediction scores.

mapping operations to obtain multiple low-rank matrices with different dimensions. Then we use low-rank matrices of different dimensions to calculate the potential interaction prediction scores of circRNA and miRNA. Finally, the prediction score matrix calculated from the low-rank matrix of different dimensions is integrated to realize the potential interaction prediction of circRNA and miRNA.

In summary, the objective function of the circRNA-miRNA potential interaction prediction model based on the fusion feature and the improved inductive matrix completion algorithm is as follows:

$$\min_{W, H} \frac{1}{2} \|CM - CM_{Pre}\|_F^2 + \frac{\theta_1}{2} \|W_i^{d_i}\|_F^2 + \frac{\theta_2}{2} \|H_i^{d_i}\|_F^2, W_i^{d_i} \geq 0, H_i^{d_i} \geq 0$$

$$CM_{Pre} = mi_f W_i^{d_1} H_i^{d_1 T} circ_f^T$$

where CM represents the known circRNA-miRNA interaction matrix. CM_{pre} represents the predicted circRNA-miRNA interaction matrix. $W_i^{d_1}$ and $H_i^{d_1}$ represent the d -dimensional low-rank matrix obtained through the i -th complete iteration of the circRNA-miRNA interaction matrix. θ_1 and θ_2 represent the regularization parameters. According to the previous research, we set $\theta_1 = \theta_2 = 1 \cdot \|F\|_F$ represents the Frobenius norm of the matrix (F-norm). $\frac{\theta_1}{2} \|W_i^{d_1}\|_F^2$ and $\frac{\theta_2}{2} \|H_i^{d_1}\|_F^2$ are used to prevent overfitting. In order to find the minimum value of the objective function, we first set up the random dense matrices of $W_i^{d_1}$ and $H_i^{d_1}$, and then update the matrices $W_i^{d_1}$ and $H_i^{d_1}$ through iterative equations. When the convergence condition is met, we will stop iteration. The iterative equation are defined as follows:

$$W_i^{d_1} \leftarrow W_i^{d_1} \frac{mi_f^T * CM * circ_f H_{ini}}{mi_f^T mi_f W_{ini} H_{ini}^T circ_f^T circ_f H_{ini} + \theta_1 W_{ini}}$$

$$H_i^{d_1} \leftarrow H_i^{d_1} \frac{circ_feature^T CM^T mi_feature W_{ini}}{circ_feature^T circ_feature H_{ini} W_{ini}^T mi_feature^T mi_feature W_{ini} + \theta_2 H_{ini}}$$

where $circ_feature$ and $mi_feature$ represent the fusion characteristics of circRNA and the fusion characteristics of miRNA, respectively. $circ_feature^T$ and $mi_feature^T$ represent the transposition matrix of the circRNA fusion feature matrix and the transposition matrix of the miRNA fusion feature matrix, respectively. CM^T represents the transposed matrix of the known circRNA-miRNA interaction matrix. W_{ini} and H_{ini} represent the initial random dense matrix of the low-rank matrix $W_i^{d_1}$ and $H_i^{d_1}$, respectively.

Finally, the calculation method of the circRNA and miRNA correlation prediction score matrix is defined as follows:

$$Pre_{circRNA-miRNA} = \frac{\sum_i^k mi_feature W_i^{d_1} H_i^{d_1 T} circ_feature}{k}$$

where $Pre_{circRNA-miRNA}$ represents the final circRNA-miRNA potential interaction prediction score matrix. Each item in the matrix represents the interaction probability score between circRNA and miRNA. The higher the score, the greater the probability that there is an exact interaction between circRNA and miRNA. k indicates the number of complete iterations of the iterative equation. The best prediction performance is obtained when $k = 2$ and $d_1 = 128, d_2 = 64$.

2.5 Performance evaluation

In order to evaluate the performance of model in predicting the potential interaction between circRNA and miRNA, the 10-fold cross-validation experiment is used to evaluate the performance. In 10-fold cross-validation, the known circRNA-miRNA interactions are randomly divided into 10 subsets. Then, in each round of cross-validation experiments, nine subsets are taken from 10 subsets as the training set for model training and the remaining subset is used as the test set. The final interaction prediction score of circRNA and miRNA is obtained. The higher the score, the higher the probability that there is a biological interaction between circRNA and miRNA. Afterward, we ranked the interaction prediction scores between circRNA and miRNA in descending order. Then, the true positive rate (TPR) and false positive rate (FPR) are calculated by modifying the threshold. The calculation of TPR and FPR are defined as follows:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

Finally, a receiver operating curve (ROC) based on the true positive rate and false positive rate is plotted, and the area under the ROC curve (AUROC value) is calculated to evaluate the predictive ability of the model. Similarly, the area of the curve (AUPR value) based on *precision* and *recall* is also used to evaluate the performance of the predictive model. The calculation of *precision* and *recall* is defined as follows:

$$Precision = \frac{TP}{TP + FP}$$

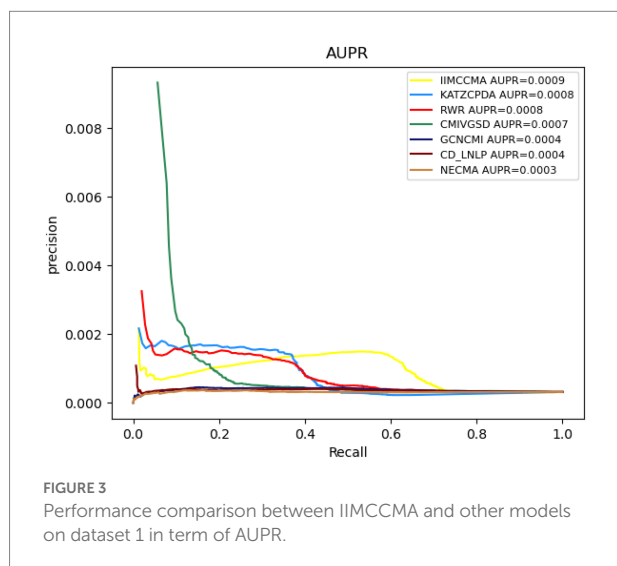
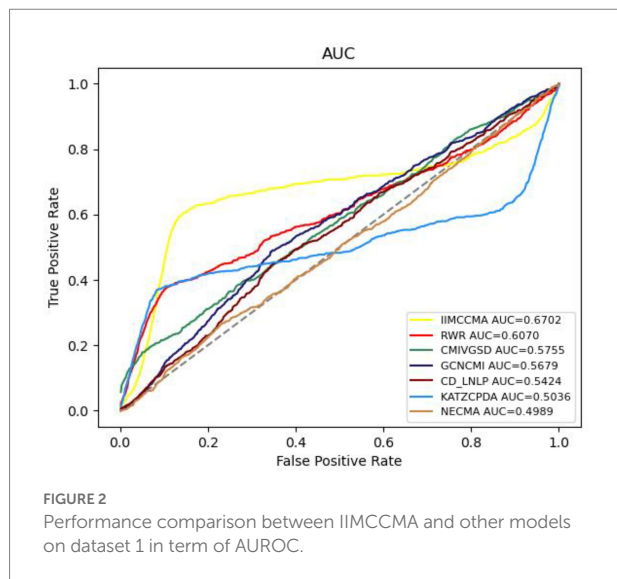
$$Recall = \frac{TP}{TP + FN}$$

where TP means that the classifier predicts the number of positive samples in the actual positive samples. FP represents the number of positive samples is predicted in the actual negative samples. TN means that the classifier predicts the number of negative samples in the actual negative samples. FN indicates the number of actual positive samples that are predicted to be negative.

3. Results and discussion

3.1 Compare with other models

In order to further demonstrate the performance of IIMCCMA, we compare it with the other six prediction methods



(NECMA; Lan et al., 2021b, GCNCMI; He et al., 2022, CMIVGSD; Qian et al., 2021, CCD-LNLP; Zhang et al., 2019, RWR; Vural et al., 2019, and KATZCPDA; Fan et al., 2018). As shown in Figure 2, under the 10-fold cross-validation experiment on dataset 1, the AUROC value of NECMA is 0.4898, the AUROC value of CMIVGSD is 0.5755, the AUROC value of GCNCMI is 0.5679, the AUROC value of CD-LNLP is 0.5424, the AUROC value of RWR is 0.6070, the AUROC value of KATZCPDA is 0.5036, and the AUROC value of IIMCCMA is 0.6702. Therefore, from the experimental results, it can be found that the IIMCCMA model has a higher AUROC value than other interaction prediction models on dataset 1.

As shown in Figure 3, under the 10-fold cross-validation experiment on dataset 1, the AUROC value of NECMA is 0.0003, the AUPR value of CMIVGSD is 0.0007, the AUPR value of GCNCMI is 0.0004, the AUPR value of CD-LNLP is 0.0004, the AUPR value of RWR is 0.0008, the AUPR value of KATZCPDA is

0.0008, and the AUPR value of the IIMCCMA model is 0.0009. It can be found from the experimental results that the IIMCCMA model achieves a higher AUPR value than the other models on dataset 1.

The Figure 4 shows the performance comparison in term of AUROC on dataset 2. It can be found that the AUROC value of NECMA is 0.5021, the AUROC value of CMIVGSD is 0.7081, the AUROC value of GCNCMI is 0.4789, the AUROC value of CD-LNLP is 0.6751, the AUROC value of RWR is 0.6729, the AUROC value of KATZCPDA is 0.6292, and the AUROC value of IIMCCMA is 0.7333. It demonstrates that IIMCCMA outperforms than other prediction models on dataset 2.

As shown in Figure 5, the AUPR value of NECMA is 0.0002, the AUPR value of CMIVGSD is 0.0011, the AUPR value of GCNCMI is 0.0002, the AUPR value of CD-LNLP is 0.0008, the AUPR value of RWR is 0.0007, the AUPR value of KATZCPDA is 0.0006, and the AUPR value of the IIMCCMA model is 0.0011. It can be found that the IIMCCMA model achieves a higher AUPR value than the other models on dataset 2. In conclusion, under the 10-fold cross-validation experiment, we can find that the IIMCCMA has achieved higher AUROC and AUPR values than the other prediction models. Thus, it can be proved that the IIMCCMA performs better in the potential circRNA-miRNA interactions identification.

3.2 Ablation experiment

In order to verify the effectiveness of the improvements of IIMCCMA, we conduct ablation experiment on dataset 1: CircRNA-miRNA potential interaction prediction model based on multi-source similarity and inductive matrix completion (IIMCCMA without improved IMC and topological features). CircRNA-miRNA potential interaction prediction model based on fusion features and inductive matrix completion (IIMCCMA without improved IMC). We adopt the 10-fold cross-validation experiment and use the AUROC value as the evaluation metrics. As shown in Figure 6, the AUROC value of the circRNA-miRNA potential interaction prediction model based on multi-source similarity (IIMCCMA without improved IMC and topological features) is 0.6728. The AUROC value of the circRNA-miRNA potential interaction prediction model based on fusion features (IIMCCMA without improved IMC) is 0.6816. The AUROC value of IIMCCMA is 0.6938. In summary, based on the original inductive matrix completion algorithm, fusion of similarity features and topological features can improve the predictive ability of the model. Adding improved inductive matrix completion on the basis of fusion features can further improve the performance of the prediction model.

3.3 Case study

In order to prove the ability of the circRNA-miRNA potential interaction model (IIMCCMA) based on the multi-source

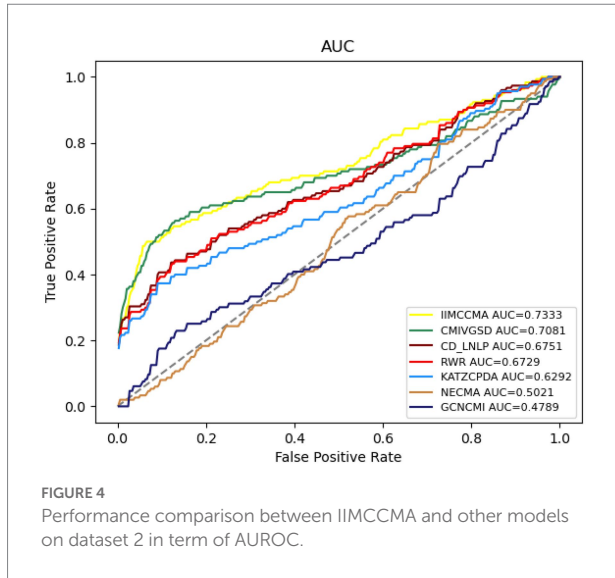


FIGURE 4 Performance comparison between IIMCCMA and other models on dataset 2 in term of AUROC.

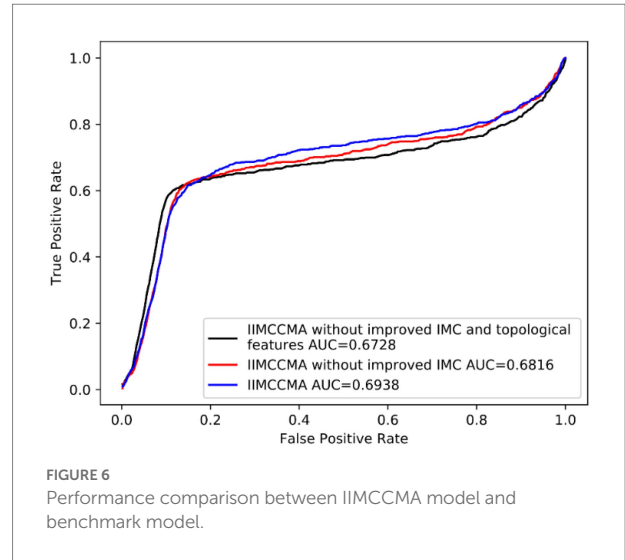


FIGURE 6 Performance comparison between IIMCCMA model and benchmark model.

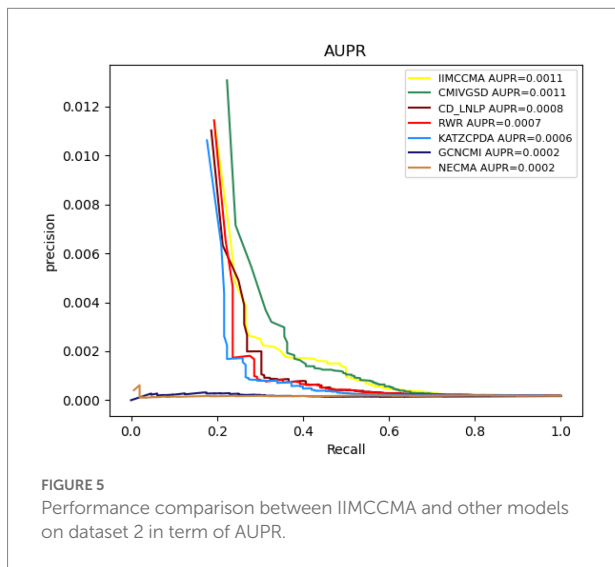


FIGURE 5 Performance comparison between IIMCCMA and other models on dataset 2 in term of AUPR.

TABLE 1 Case study based on microRNA miR-145-5p.

Rank	CircRNA	Evidence	Reference
1	hsa_circ_0058063	PMID: 30362519	Sun et al. (2019a)
2	hsa_circRNA_101981	PMID: 30136305	He et al. (2018)
3	hsa_circRNA_091420	PMID: 30136305	He et al. (2018)
4	hsa_circ_100242	PMID: 32218853	Zhu et al. (2020)
5	circPTN	PMID: 31511040	Chen et al. (2019)
6	circPVT1	PMID: 31986409	Zheng and Xu (2020)
7	hsa_circRNA_101996	PMID: 30136305	He et al. (2018)
8	circCEP128	PMID: 30939216	Sun et al. (2019b)
9	hsa_circ_0003855	PMID: 31776711	Zhang et al. (2020)
10	hsa_circ_0001955	PMID: 31822654	Yao et al. (2019)

biological interaction data to identify the potential interaction between circRNA and miRNA. This paper builds a case study based on miRNA miR-145-5p. Finally, this paper selects the top 10 circRNAs predicted by the IIMCCMA model that are related to miRNA miR-145-5p, and manually searches the existing literature to prove their relevance.

The top 10 circRNAs related to miRNA miR-145-5p predicted by the IIMCCMA model are shown in Table 1. From Table 1, 10 circRNAs related to miRNA miR-145-5p (hsa_circ_0058063, hsa_circRNA_101981, hsa_circRNA_091420, hsa_circ_100242, circPTN, circPVT1, hsa_circRNA_101996, circCEP128, hsa_circ_0003855, and hsa_circ_0001955) have been confirmed by existing literature. Specifically, the first circRNA hsa_circ_0058063 can be used as the sponge of miRNA miR-145-5p to regulate the expression of miRNA target gene CDK6 and promote the development of bladder cancer (Sun et al., 2019a). In prostate

cancer cells, the expression pattern of the second-ranked circRNA hsa_circRNA_101981 was significantly down-regulated. Further experiments showed that miRNA miR-145-5p can regulate the expression of circRNA hsa_circRNA_101981 (He et al., 2018). The expression pattern of the third-ranked circRNA hsa_circRNA_091420 in prostate cancer cells was significantly upregulated. Overexpressed miRNA miR-145-5p can inhibit the expression of circRNA hsa_circRNA_091420 (He et al., 2018). Experimental results show that the fourth-ranked circRNA hsa_circ_100242 can interact with miRNA miR145-5p in bladder cancer cells (Zhu et al., 2020). Experiments show that the fifth-ranked circRNA circPTN is overexpressed in glioma cells and tissues. Further experiments showed that circRNA circPTN can sponge miRNA miR-145-5p and thus play a carcinogenic effect in glioma cells (Chen et al., 2019). CircRNA circPVT1, ranked sixth, was significantly up-regulated in lung adenocarcinoma cells. Experiments show that in lung adenocarcinoma cells, circRNA circPVT1 can be used as a ceRNA for miRNA miR145-5p (Zheng and Xu, 2020). Experiments show that the seventh-ranked circRNA hsa_circRNA_101996 can interact with miRNA miR-145-5p in prostate cancer cells. In addition, overexpressed

miRNA miR-1455p can inhibit the expression of circRNA hsa_circRNA_101996 (He et al., 2018). The eighth-ranked circRNA circCEP128 can promote the development of bladder cancer by regulating miRNA miR-145-5p and miRNA's target gene MYD88 (Sun et al., 2019b). The expression pattern of circRNA hsa_circ_0003855, ranked ninth, was significantly increased in gastric cancer cells. Experimental results show that circRNA hsa_circ_0003855 can take on the sponge effect of miRNA miR-145-5p to promote the proliferation and migration of gastric cancer cells (Zhang et al., 2020). The tenth-ranked circRNA hsa_circ_0001955 can assume the role of miRNA miR-145-5p sponge. Additionally, the downregulated circRNA hsa_circ_0001955 can inhibit the growth of hepatocellular carcinoma tumors (Yao et al., 2019). In summary, through the case study results based on miRNA miR-145-5p, it can be found that the IIMCCMA model can correctly identify the potential biological interaction between circRNA and miRNA.

4. Conclusion

Experiments show that circRNA can play an important role in cancer as a miRNA sponge. Therefore, correct identification of the interaction between circRNA and miRNA not only helps to understand the complex disease mechanism, but also contributes to the diagnosis, treatment and prognosis of the disease. Based on circRNA-miRNA interaction, circRNA-cancer interaction and miRNA-cancer interaction, this paper proposes a circRNA-miRNA potential interaction prediction model based on multi-source biological interaction data, IIMCCMA. This model first uses the Gaussian kernel function to calculate the GIP kernel similarity of circRNA and the GIP kernel of miRNA based on the circRNA-miRNA interaction network. Then, on the basis of the circRNA-cancer interaction network and the miRNA-cancer interaction network, the cosine function is used to calculate the functional similarity of circRNA and miRNA, respectively. Afterward, the different similarities of circRNAs and the different similarities of miRNAs were integrated separately. The known circRNA-miRNA interaction network is used to construct a heterogeneous network for extracting topological features of circRNA and miRNA, and the network embedding algorithm (NetMF) is used to obtain the low-dimensional space vectors of circRNA and miRNA, respectively. Finally, based on the fusion features, an improved inductive matrix completion algorithm is used to predict the potential interaction between circRNA and miRNA. In order to test the performance of the IIMCCMA, this paper selects four circRNA-disease potential interaction prediction models for comparison. The 10-fold cross-validation results show that compared with the other four models, the IIMCCMA achieved higher AUROC and AUPR values. Therefore, it is proved that IIMCCMA has better predictive ability. Moreover, the results of a case study based on miRNA miR-1455p show that the IIMCCMA model can correctly identify the potential interaction between circRNA and miRNA.

Although, the IIMCCMA model has shown excellent performance in predicting the potential interaction between circRNA and miRNA. However, there are still some shortcomings and limitations. (1) The imbalance of positive and negative samples in interaction data. Because the efficiency of identifying circRNA and miRNA interactions through biological experiments is low, in the existing circRNA-miRNA interaction network, the experimentally verified interactions are far less than the unknown interactions. The sparse circRNA-miRNA interaction network greatly affects the performance of the prediction model (Lan et al., 2016b, 2020b; Lei et al., 2020). Therefore, in the follow-up work, we will try to pre-fill the original interaction matrix to alleviate the sparsity of the known interaction network and enhance the performance of the model. (2) Parameter setting. There are a certain number of parameters in the IIMCCMA model that need to be set manually. The quality of the parameters needs to be confirmed through experimental verification. In addition, too many parameters will reduce the learning and generalization capabilities of the model. Therefore, no parameter or self-learning parameter model will be the main work in the future (Lan et al., 2016a, 2017, 2022b; Chen et al., 2021).

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SY designed the work. DY and LN performed all the experiments. DY, LN, MQ, and SW wrote the manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Chen, J., Chen, T., Zhu, Y., Li, Y., Zhang, Y., Wang, Y., et al. (2019). circPTN sponges miR-145-5p/miR-330-5p to promote proliferation and stemness in glioma. *J. Exp. Clin. Cancer Res.* 38, 1–17. doi: 10.1186/s13046-019-1376-8
- Chen, Q., Lai, D., Lan, W., Wu, X., Chen, B., Liu, J., et al. (2021). ILDMSF: inferring associations between long non-coding RNA and disease based on multi-similarity fusion. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1106–1112. doi: 10.1109/TCBB.2019.2936476
- Fan, C., Lei, X., and Wu, F. X. (2018). Prediction of CircRNA-disease interactions using KATZ model based on heterogeneous networks. *Int. J. Biol. Sci.* 14, 1950–1959. doi: 10.7150/ijbs.28260
- Fang, Z., and Lei, X. (2019). Prediction of miRNA-circRNA interactions based on k-NN multi-label with random walk restart on a heterogeneous network. *Big Data Min. Anal.* 2, 261–272. doi: 10.26599/BDMA.2019.9020010
- Glažar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA* 20, 1666–1670. doi: 10.1261/rna.043687.113
- Guo, L. X., You, Z. H., Wang, L., Yu, C. Q., Zhao, B. W., Ren, Z. H., et al. (2022). A novel circRNA-miRNA association prediction model based on structural deep neural network embedding. *Brief. Bioinform.* 23:bbac391. doi: 10.1093/bib/bbac391
- He, J. H., Han, Z. P., Zhou, J. B., Chen, W. M., Lv, Y. B., He, M. L., et al. (2018). MiR-145 affected the circular RNA expression in prostate cancer LNCaP cells. *J. Cell. Biochem.* 119, 9168–9177. doi: 10.1002/jcb.27181
- He, J., Xiao, P., Chen, C., Zhu, Z., Zhang, J., and Deng, L. (2022). GCNCMI: a graph convolutional neural network approach for predicting circRNA-miRNA interactions. *Front. Genet.* 13:959701. doi: 10.3389/fgene.2022.959701
- Jain, P., and Dhillon, I. S. (2013). Provable inductive matrix completion. arXiv [Preprint]. doi: 10.48550/arXiv.1306.0626
- Lan, W., Dong, Y., Chen, Q., Liu, J., Wang, J., Chen, Y. P. P., et al. (2021a). IGNSCDA: predicting CircRNA-disease associations based on improved graph convolutional network and negative sampling. *IEEE/ACM Trans. Comput. Biol. Bioinform.* doi: 10.1109/TCBB.2021.3111607 (Epub ahead of print).
- Lan, W., Dong, Y., Chen, Q., Zheng, R., Liu, J., Pan, Y., et al. (2022a). KGANCD: predicting circRNA-disease associations based on knowledge graph attention network. *Brief. Bioinform.* 23:bbab494. doi: 10.1093/bib/bbab494
- Lan, W., Lai, D., Chen, Q., Wu, X., Chen, B., Liu, J., et al. (2020a). LDICDL: lncRNA-disease association identification based on collaborative deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.*
- Lan, W., Lai, D., Chen, Q., Wu, X., Chen, B., Liu, J., et al. (2020b). LDICDL: lncRNA-disease interaction identification based on collaborative deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.*
- Lan, W., Li, M., Zhao, K., Liu, J., Wu, F. X., Pan, Y., et al. (2017). LDAP: a web server for lncRNA-disease association prediction. *Bioinformatics* 33, 458–460. doi: 10.1093/bioinformatics/btw639
- Lan, W., Wang, J., Li, M., Liu, J., Li, Y., Wu, F. X., et al. (2016a). Predicting drug-target interaction using positive-unlabeled learning. *Neurocomputing* 206, 50–57. doi: 10.1016/j.neucom.2016.03.080
- Lan, W., Wang, J., Li, M., Liu, J., and Pan, Y. (2015a). “Predicting microRNA-disease associations by integrating multiple biological information.” in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 183–188.
- Lan, W., Wang, J., Li, M., Liu, J., Wu, F. X., and Pan, Y. (2016b). Predicting microRNA-disease associations based on improved microRNA and disease similarities. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15, 1774–1782. doi: 10.1109/TCBB.2016.2586190
- Lan, W., Wang, J., Li, M., Peng, W., and Wu, F. (2015b). Computational approaches for prioritizing candidate disease genes based on PPI networks. *Tsinghua Sci. Technol.* 20, 500–512. doi: 10.1109/TST.2015.7297749
- Lan, W., Wu, X., Chen, Q., Peng, W., Wang, J., and Chen, Y. P. (2022b). GANLDA: graph attention network for lncRNA-disease associations prediction. *Neurocomputing* 469, 384–393. doi: 10.1016/j.neucom.2020.09.094
- Lan, W., Zhang, H., Dong, Y., Chen, Q., Cao, J., Peng, W., et al. (2022c). DRGCNCDA: predicting circRNA-disease interactions based on knowledge graph and disentangled relational graph convolutional network. *Methods* 208, 35–41. doi: 10.1016/j.ymeth.2022.10.002
- Lan, W., Zhu, M., Chen, Q., Chen, B., Liu, J., Li, M., et al. (2020c). CircR2Cancer: A manually curated database of interactions between circRNAs and cancers. Database, 2020
- Lan, W., Zhu, M., Chen, Q., Chen, J., Ye, J., Liu, J., et al. (2021b). Prediction of circRNA-miRNA associations based on network embedding. *Complexity* 2021, 1–10. doi: 10.1155/2021/6659695
- Lei, X., Mudiyansele, T. B., Zhang, Y., Bian, C., Lan, W., Yu, N., et al. (2020). A comprehensive survey on computational methods of non-coding RNA and disease interaction prediction. *Brief. Bioinform.*
- Liu, M., Wang, Q., Shen, J., Yang, B. B., and Ding, X. (2019). Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol.* 16, 899–905. doi: 10.1080/15476286.2019.1600395
- Nazarov, I., Shirokikh, B., Burkina, M., Fedonin, G., and Panov, M. (2018). Sparse group inductive matrix completion. arXiv [Preprint]. doi: 10.48550/arXiv.1804.10653
- Qian, Y., Zheng, J., Jiang, Y., Li, S., and Deng, L. (2022). Prediction of circRNA-miRNA association using singular value decomposition and graph neural networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1, 1–9. doi: 10.1109/TCBB.2022.3222777
- Qian, Y., Zheng, J., Zhang, Z., Jiang, Y., Zhang, J., and Deng, L. (2021). “CMIVGSD: circRNA-miRNA interaction prediction based on Variational graph auto-encoder and singular value decomposition.” in *2021 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 205–210.
- Qiu, J., Dong, Y., Ma, H., Li, J., Wang, K., and Tang, J. (2018). “Network embedding as matrix factorization: unifying deepwalk, line, ptv, and node2vec.” in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 459–467.
- Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., et al. (2015). Circular RNA: a new star of noncoding RNAs. *Cancer Lett.* 365, 141–148. doi: 10.1016/j.canlet.2015.06.003
- Rophina, M., Sharma, D., Poojary, M., and Scaria, V. (2020). Circad: A comprehensive manually curated resource of circular RNA associated with diseases. Database, 2020.
- Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885. doi: 10.1016/j.molcel.2015.03.027
- Si, S., Chiang, K. Y., Hsieh, C. J., Rao, N., and Dhillon, I. S. (2016). “Goal-directed inductive matrix completion.” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1165–1174.
- Sun, M., Zhao, W., Chen, Z., Li, M., Li, S., Wu, B., et al. (2019a). Circ_0058063 regulates CDK6 to promote bladder cancer progression by sponging miR-145-5p. *J. Cell. Physiol.* 234, 4812–4824. doi: 10.1002/jcp.27280
- Sun, M., Zhao, W., Chen, Z., Li, M., Li, S., Wu, B., et al. (2019b). Circular RNA CEP128 promotes bladder cancer progression by regulating Mir-145-5p/Myd88 via MAPK signaling pathway. *Int. J. Cancer* 145, 2170–2181. doi: 10.1002/ijc.32311
- Vural, H., Kaya, M., and Alhaji, R. (2019). “A model based on random walk with restart to predict circRNA-disease interactions on heterogeneous network.” in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 929–932.
- Wang, X. F., Yu, C. Q., Li, L. P., You, Z. H., Huang, W. Z., Li, Y. C., et al. (2022). KGDCMI: a new approach for predicting circRNA-miRNA interactions from multi-source information extraction and deep learning. *Front. Genet.* 13:958096. doi: 10.3389/fgene.2022.958096
- Wilusz, J. E., and Sharp, P. A. (2013). A circuitous route to noncoding RNA. *Science* 340, 440–441. doi: 10.1126/science.1238522
- Xie, F., Li, Y., Wang, M., Huang, C., Tao, D., Zheng, F., et al. (2018). Circular RNA BCRC-3 suppresses bladder cancer proliferation through miR-182-5p/p27 axis. *Mol. Cancer* 17, 1–12. doi: 10.1186/s12943-018-0892-z
- Yao, Z., Xu, R., Yuan, L., Xu, M., Zhuang, H., Li, Y., et al. (2019). Circ_0001955 facilitates hepatocellular carcinoma (HCC) tumorigenesis by sponging miR-516a-5p to release TRAF6 and MAPK11. *Cell Death Dis.* 10, 1–12. doi: 10.1038/s41419-019-2176-y
- Yu, C. Q., Wang, X. F., Li, L. P., You, Z. H., Huang, W. Z., Li, Y. C., et al. (2022). SGCNCMI: a new model combining multi-modal information to predict circRNA-related miRNAs, diseases and genes. *Biology* 11:1350. doi: 10.3390/biology11091350
- Zhang, Z., Wang, C., Zhang, Y., Yu, S., Zhao, G., and Xu, J. (2020). CircDUSP16 promotes the tumorigenesis and invasion of gastric cancer by sponging miR-145-5p. *Gastric Cancer* 23, 437–448. doi: 10.1007/s10120-019-01018-7
- Zhang, W., Yu, C., Wang, X., and Liu, F. (2019). Predicting CircRNA-disease interactions through linear neighborhood label propagation method. *IEEE Access* 7, 83474–83483. doi: 10.1109/ACCESS.2019.2920942
- Zheng, Y., Chen, C. J., Lin, Z. Y., Li, J. X., Liu, J., Lin, F. J., et al. (2020). Circ_KATNAL1 regulates prostate cancer cell growth and invasiveness through the miR-145-3p/WISP1 pathway. *Biochem. Cell Biol.* 98, 396–404. doi: 10.1139/bcb-2019-0211
- Zheng, F., and Xu, R. (2020). CircPVT1 contributes to chemotherapy resistance of lung adenocarcinoma through miR-145-5p/ABCC1 axis. *Biomed. Pharmacother.* 124:109828. doi: 10.1016/j.biopha.2020.109828
- Zhu, Z., Chang, F., Liu, J., Wang, J., and Zhang, X. (2020). Comprehensive circular RNA profiling reveals the regulatory role of circ_100242/miR-145 pathway in bladder cancer. *Oncol. Lett.* 19, 2971–2978. doi: 10.3892/ol.2020.11380