

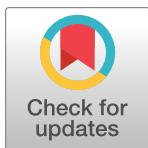
RESEARCH ARTICLE

lncRNA-disease association prediction based on the weight matrix and projection score

Bo Wang *, Chao Zhang, Xiao-xin Du, Xiao-dong Zheng, Jing-you Li

College of Computer and Control Engineering, Qiqihar University, Qiqihar, People's Republic of China

* bowangdr@qqhr.edu.cn



OPEN ACCESS

Citation: Wang B, Zhang C, Du X-x, Zheng X-d, Li J-y (2023) lncRNA-disease association prediction based on the weight matrix and projection score. PLoS ONE 18(1): e0278817. <https://doi.org/10.1371/journal.pone.0278817>

Editor: Sriparna Saha, Indian Institute of Technology Patna, INDIA

Received: January 7, 2022

Accepted: November 25, 2022

Published: January 3, 2023

Copyright: © 2023 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets generated and/or analysed during the current study are available in the HMDD repository, <http://www.cuilab.cn/>; MNDR repository, <http://www.rna-society.org/mndr/>; starBase repository, <http://starbase.sysu.edu.cn/starbase2/index.php>; Additional file 1 Known miRNA-disease associations obtained from HMDD. Additional file 2 Known miRNA-lncRNA associations obtained from starBase v2.0. Additional file 3 Known lncRNA-disease associations obtained from MNDR v2.0. These are third-party data that no one else has

Abstract

With the development of medical science, long noncoding RNA (lncRNA), originally considered as a noise gene, has been found to participate in a variety of biological activities. Several recent studies have shown the involvement of lncRNA in various human diseases, such as gastric cancer, prostate cancer, lung cancer, and so forth. However, obtaining lncRNA-disease relationship only through biological experiments not only costs manpower and material resources but also gains little. Therefore, developing effective computational models for predicting lncRNA-disease association relationship is extremely important. This study aimed to propose an lncRNA-disease association prediction model based on the weight matrix and projection score (LDAP-WMPS). The model used the relatively perfect lncRNA-miRNA relationship data and miRNA-disease relationship data to predict the lncRNA-disease relationship. The integrated lncRNA similarity matrix and the integrated disease similarity matrix were established by fusing various methods to calculate the similarity between lncRNA and disease. This study improved the existing weight algorithm, applied it to the lncRNA-miRNA-disease triple network, and thus proposed a new lncRNA-disease weight matrix calculation method. Combined with the improved projection algorithm, the lncRNA-miRNA relationship and miRNA-disease relationship were used to predict the lncRNA-disease relationship. The simulation results showed that under the Leave-One-Out-Cross-Validation framework, the area under the receiver operating characteristic curve of LDAP-WMPS could reach 0.8822, which was better than the latest result. Taking adenocarcinoma and colorectal cancer as examples, the LDAP-WMPS model was found to effectively infer the lncRNA-disease relationship. The simulation results showed good prediction performance of the LDAP-WMPS model, which was an important supplement to the research of lncRNA-disease association prediction without lncRNA-disease relationship data.

Introduction

According to the traditional central principle, RNA is divided into messenger RNA (mRNA) and noncoding RNA (ncRNA). mRNA is the medium for DNA to be transcribed into a protein, while ncRNA has always been regarded as noise and has no real effect. However, the

special access rights, and others will be able to access them in the same way as the author.

Funding: This work was supported in part by the grants of the Undergraduate Universities Fundamental Research Funding Project of Heilongjiang Province awarded to JYL (No. 135509115). No additional external funding was received for this study.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AUC, areas under ROC curve; LDAP-WMPS, lncRNA-Disease Association Prediction Based On Weight Distribution And Projection Score; PR, false positive rates; LMDN, the lncRNA-miRNA-disease tripartite network; LMDN, an updated lncRNA-miRNA-disease association tripartite network; lncRNA, long non-coding RNAs lncRNA; LOOCV, Leave-One Out Cross Validation; PR, true positive rates.

sequencing results showed that in the whole human gene pool, less than 5% of DNA and RNA were involved in protein transcription, while other genes were involved in RNA transcription that could not be encoded, that is, the number of ncRNA was far greater than that of coding RNA [1]. In 1998, two American scientists, Andrew Farr and Craig Mello, jointly published a paper on the discovery of RNA interference mechanism in the journal *Nature*. They believed that RNA interference existed in all organisms, and RNA played a regulatory role in gene expression [2], virus infection [3, 4], immune system [5], and so forth, thus bringing biological research into a new stage. Then, the research on ncRNA gradually increased, among which the research on long ncRNAs (lncRNAs) has been one of the hot topics. lncRNA is a kind of ncRNA whose nucleotide length is more than 200. In previous studies, it was considered to be the noise generated in the process of transcription [6, 7]. Nowadays, lncRNA has been found to be involved in all aspects of cell life cycle, including transcription [8], cell differentiation [9], cell transport [10], apoptosis [11], metabolic process [12], and so on. Moreover, lncRNA has also been found to be associated with various human diseases [13], including leukemia [14, 15], diabetes [16, 17], prostate cancer [18, 19], lung cancer [20, 21], colon cancer [22, 23], cardiovascular disease [24, 25], and so on. lncRNA participates in diseases through abnormal sequence and spatial structure, abnormal expression level, and abnormal interaction with binding proteins, thus affecting human health [26, 27].

Therefore, linking lncRNA with diseases can help realize the early detection of diseases, the targeted treatment of diseases, and the systematic understanding of the etiological characteristics of complex diseases. The biological experiments related to lncRNA cost a lot of money and time to carry out because of the complex lncRNA-disease relationship. Computer-aided experiment has become an effective research method. These experiments can effectively predict the complex lncRNA-disease relationship. The datasets in the open lncRNA database are used to verify the prediction results. The prediction of the lncRNA-disease relationship is of great significance in biology, medicine, and other fields. In the field of biology, computer-aided experiments can reduce the cost of experiments and improve the success rate of experiments. In the field of medicine, computer-aided experiments can help researchers identify lncRNAs related to various diseases and understand the pathogenesis of diseases at the molecular level so as to effectively prevent and treat diseases [28].

The proposed prediction model is developing rapidly. Many prediction models, including CircRNA-disease association prediction model [29], miRNA-disease association prediction model [30, 31], lncRNA-miRNA association prediction model [32, 33], and lncRNA-disease association prediction model, have greatly enriched the relationship between computer science and biology. This paper mainly proposes a new prediction method for lncRNA-disease association prediction. The following is an analysis of some previous lncRNA-disease association prediction models: The proposed model can be divided into two categories based on the experimental data. The first model relies only on the lncRNA-disease relationship information. Specifically, we can predict the lncRNA-disease relationship through the association information between lncRNAs and diseases. For example, Xie et al. proposed a new method for human lncRNA disease association prediction based on network consistent projection (NCPHLDA) [34]. The model integrates lncRNA cosine similarity network and disease cosine similarity network. At the same time, it has no requirements for parameters and has good prediction performance. However, there are some limitations. If the known lncRNA disease correlation is very small, the prediction results will be biased. Chen et al. developed the NCMCMDA [35] model, innovatively combines neighborhood constraints with matrix completion, providing a new idea to use similarity information used to aid forecasting. However, NCMCMDA also has limitations. Currently known miRNA deficiency-disease association may affect the long road of NCMCMDA performance expansion data. Secondly, how to effectively select parameter

information, miRNA similarity information and low order constraints to balance the influence of disease similarity still needs further research. Based on disease semantic similarity and lncRNA-disease relationship information, Zhang et al. developed an LDAI-ISPS model to predict the potential lncRNA-disease relationship through network consistency [36]. The model integrated Gaussian interaction profile central similarity to calculate disease similarity and lncRNA similarity, which made up the incompleteness of the similarity network construction only with semantic similarity. However, this method still had limitations in that the predicted results were biased toward the diseases with more related lncRNAs or the lncRNAs with more related diseases. The other model integrated multiple data; collected multiple biological data, such as lncRNA, miRNA, protein, disease, and so on; and integrated these data into matrix or heterogeneous network to infer the potential lncRNA-disease relationship. For example, Fu et al. proposed an lncRNA disease association prediction method (MFLDA) based on matrix decomposition [37]. In this way, the weight of the data source and the correlation matrix of the disease can be assigned to the data source with smaller weight to speculate the potential association of lncRNA disease. The biggest advantage of this model is that it can easily predict the correlation between different research objects by classifying various heterogeneous data sources. However, MFLDA prefers to study data sparse matrix. Its performance depends on low-quality and irrelevant internal relational data sources, but it does not get rid of the use of lncRNA disease association attribute information. Yu and Wang et al. developed an NBCLDA model [38], which integrated a variety of organisms to construct a new tripartite network, including miRNA-disease, miRNA-lncRNA, and lncRNA-disease relationship and interaction. Then, a quadruple network was constructed, and a naïve Bayesian classifier was applied for the prediction. The important limitation of the naïve Bayesian classifier was that the information of negative samples was required. Therefore, unlabeled lncRNA-disease pairs were always randomly selected as negative samples, which could seriously influence the prediction performance. Yu et al. proposed a new model CFNHC [39], which was an improvement of the original NBCLDA model. It combined collaborative filtering with naïve Bayes and inferred the potential lncRNA-disease relationship by calculating the relationship score between lncRNA and disease. Although the introduction of a collaborative filtering algorithm effectively improved the prediction ability of CFNHC, it still failed to resolve the limitations of the naïve Bayesian model.

Most of the prediction of lncRNA-disease correlation needed to know the correlation between lncRNA and diseases. However, the known lncRNA-disease relationship is quite rare. To solve the aforementioned problems, this study proposed an lncRNA-disease association prediction model based on the weight matrix and projection score (LDAP-WMPS). The model used the relatively perfect lncRNA-miRNA relationship data and miRNA-disease relationship data to predict the lncRNA-disease relationship. The integrated lncRNA similarity matrix and the integrated disease similarity matrix were established by fusing various methods to calculate the similarity between lncRNA and disease. On this basis, the weight algorithm was improved and applied to the lncRNA-miRNA-disease triple network. Based on the network, a new lncRNA-disease weight matrix calculation method was proposed. Combined with the improved projection algorithm, the lncRNA-miRNA relationship and the miRNA-disease relationship were used to predict the lncRNA-disease relationship. The simulation results showed that based on the Leave-One-Out-Cross-Validation (LOOCV) framework, the area under the receiver operating characteristic (ROC) curve (AUC) of LDAP-WMPS could reach 0.8822, which was better than the latest result. Taking adenocarcinoma and colorectal cancer as examples, LDAP-WMPS was found to effectively infer the lncRNA-disease relationship.

Materials and methods

Dataset and preprocessing

The known lncRNA-disease relationship dataset was downloaded from the MNDRv2.0 database (2017 edition) [40]. The known miRNA-disease relationship datasets were downloaded from the HMDD database (2018 edition) [41]. The known lncRNA-miRNA relationship dataset was downloaded from the Starbase v2.0 database (2015 edition) [42]. After data cleaning and name unification, three datasets D_{LM} , D_{MD} , and D_{LD} were retrieved. The D_{LM} database comprised 1089 different lncRNAs and 246 different miRNAs; the D_{MD} database comprised 246 different miRNAs and 373 different diseases; and the D_{LD} comprised 1089 different lncRNAs and 373 different diseases. The D_{LD} dataset was not used as the training set, but only as the test set. The D_{MD} and D_{LM} datasets were analyzed and transformed into adjacency matrices. Taking lncRNA-miRNA relationship dataset as an example, the adjacency matrix A_{LM} was constructed. The lncRNA was listed as the row, and miRNA was listed as the column. If the miRNA in row j interacted with lncRNA in column i , then $A_{LM}(i,j) = 1$; else, $A_{LM}(i,j) = 0$. Similarly, the adjacency matrix A_{MD} was constructed.

Cosine similarity for diseases

The principle of disease cosine similarity was based on the assumption that if disease i and disease j were similar to each other, which is a commonly used similarity calculation method [43]. Then the binary vectors $A_{MD}(:, i)$ and $A_{MD}(:, j)$ should also be similar to each other. The same assumption should also be true for diseases. According to the known miRNA-disease relationship data, the cosine similarity for disease between miRNA and disease was calculated as:

$$CD(i,j) = \frac{A_{MD}(:,i) * A_{MD}(:,j)}{\|A_{MD}(:,i)\| \|A_{MD}(:,j)\|} \quad (1)$$

Where $A_{MD}(:, i)$ is the i th column vector in the adjacency matrix of miRNA and disease, which represents the relationship feature of disease i .

Jaccard similarity for diseases

Similarity measurement is the core of a prediction model. Cosine similarity is widely used in related prediction researches. However, in many practical applications, the sparsity of evaluation data is too high, and the calculation of cosine similarity between diseases produces misleading results. Compared with the traditional similarity measurement method, the Jaccard method improves the disadvantage that the cosine similarity only considers the disease score and ignores other information. It is especially suitable for data with high sparsity. The Jaccard similarity for disease between miRNA and disease was calculated as follows:

$$JD(i,j) = \frac{A_{MD}(:,i) \cap A_{MD}(:,j)}{A_{MD}(:,i) \cup A_{MD}(:,j)} \quad (2)$$

Where $A_{MD}(:, i)$ is the i th column vector in the adjacency matrix of miRNA and disease, which represents the relationship feature of disease i . Similarly, $A_{MD}(:, j)$ represents the relationship feature of miRNA j ; $A_{MD}(:,i) \cap A_{MD}(:,j)$ is the number of miRNAs associated with disease i and disease j ; and $A_{MD}(:,i) \cup A_{MD}(:,j)$ is the sum of miRNAs related to disease i and disease j .

Integrated disease similarity

The two similarity calculation methods were integrated, the shortcomings of various similarity calculation methods were reduced to a certain extent, and the prediction ability of unknown relationships was greatly increased. Integrating disease semantic similarity and cosine similarity for diseases gave:

$$IDS(i, j) = \begin{cases} CD(i, j) & \text{if } CD(i, j) \neq 0; \\ JD(i, j) & \text{if } CD(i, j) = 0; \end{cases} \quad (3)$$

Cosine similarity for lncRNA

Similar to the disease cosine similarity calculation method, the cosine similarity for lncRNA between lncRNA and miRNA was calculated as follows:

$$CL(i, j) = \frac{A_{LM}(i, :) * A_{LM}(j, :)^T}{\|A_{LM}(i, :)\| \|A_{LM}(j, :)\|} \quad (4)$$

Where $A_{LM}(i, :)$ is the i th row vector in the adjacency matrix of lncRNA and miRNA, which represents the relationship feature of lncRNA i .

Jaccard similarity for lncRNA

Similar to the disease Jaccard similarity calculation method, the Jaccard similarity for lncRNA between lncRNA and miRNA was calculated as follows:

$$JL(i, j) = \frac{A_{LM}(i, :) \cap A_{LM}(j, :)^T}{A_{LM}(i, :) \cup A_{LM}(j, :)^T} \quad (5)$$

Where $A_{LM}(i, :)$ is the i th row vector in the adjacency matrix of lncRNA and miRNA, which represents the relationship feature of lncRNA i . Similarly, $A_{LM}(j, :)$ represents the relationship feature of lncRNA j ; $A_{LM}(i, :) \cap A_{LM}(j, :)^T$ is the number of miRNAs associated with lncRNA i and lncRNA j ; and $A_{LM}(i, :) \cup A_{LM}(j, :)^T$ is the sum of miRNAs related to lncRNA i and lncRNA j .

Integrated lncRNA similarity

Similar to the disease-integrated similarity calculation method, integrating miRNA similarity MS and cosine similarity CL for lncRNA gave:

$$ILS(i, j) = \begin{cases} CL(i, j) & \text{if } CL(i, j) \neq 0; \\ JL(i, j) & \text{if } CL(i, j) = 0; \end{cases} \quad (6)$$

Establishment of lncRNA-disease weight matrix

Weight assignment algorithm [44, 45] is often used in the association prediction of the lncRNA dual network. The correlation score between lncRNA and diseases could be obtained through weight distribution. This was further improved and applied to the lncRNA-miRNA-disease triple network, as shown in Fig 1. Taking L to M as an example, the first step was defined as follows:

$$f(M_j) = \sum_{i=1}^m \frac{a_{ij}f(L_i)}{k(L_i)} \quad (7)$$

Where m is the number of lncRNAs, $k(L_i)$ gives the number of miRNAs related to lncRNA i ,

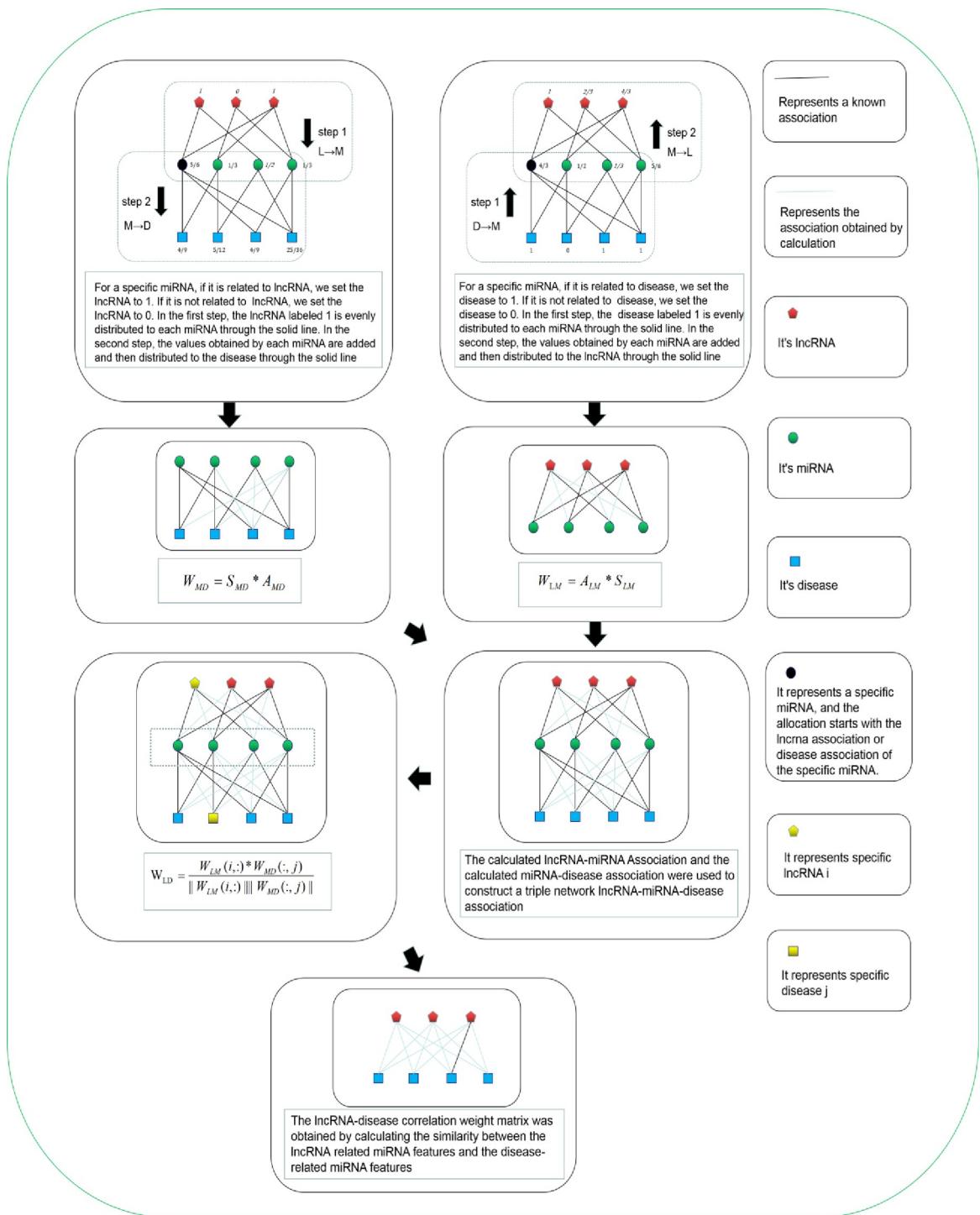


Fig 1. Flow chart of lncRNA-disease association weight matrix construction.

<https://doi.org/10.1371/journal.pone.0278817.g001>

and a_{ij} represents an entity in the lncRNA-miRNA matrix A_{LM} . $f(L_i)$ represents a binary vector formed by miRNA j corresponding to all lncRNA-miRNA relationships (if lncRNA is associated with miRNA, the value is 1; otherwise, the value is 0).

The second step was M to D, defined as:

$$f(D_e) = \sum_{j=1}^n \frac{b_{je}}{k(M_j)} \sum_{i=1}^m \frac{a_{ij} f(L_i)}{k(L_i)} \quad (8)$$

Where n is the number of miRNAs, and e is the number of diseases. $k(M_j)$ is the number of diseases related to miRNA j, and b_{je} represents an entity in the miRNA-disease matrix A_{MD} . $f(D_e)$ could be expressed as:

$$f(D_e) = \sum_{i=1}^m s^{md} * f(L_i) \quad (9)$$

Combining Eqs (8) and (9), the following formula was obtained:

$$s^{md} = \frac{1}{k(L_i)} \sum_{j=1}^n \frac{b_{je} a_{ij}}{k(M_j)} \quad (10)$$

In the aforementioned formula, $S_{MD} = \{s^{md}\}_{n \times n}$ is the score of miRNA-disease relationship. The miRNA-disease relationship weight matrix was defined as:

$$W_{MD} = S_{MD} * A_{MD} \quad (11)$$

Similarly, the weight matrix W_{LM} from D to M to L was defined as:

$$s^{lm} = \frac{1}{k(D_e)} \sum_{j=1}^n \frac{b_{je} a_{ij}}{k(M_j)} \quad (12)$$

$$W_{LM} = A_{LM} * S_{LM} \quad (13)$$

For lncRNA i , the potential relationship characteristics between miRNAs and lncRNA i were calculated; and for disease j , the potential relationship characteristics between miRNAs and disease j were also calculated. $W_{LM}(i, :)$ was used to represent the eigenvalue of miRNA associated with lncRNA i , and $W_{MD}(:, j)$ to represent the eigenvalue of miRNA associated with disease j . Then, the weight between lncRNA and disease was defined as follows:

$$w^{ij} = \frac{W_{LM}(i, :) * W_{MD}(:, j)}{\|W_{LM}(i, :)\| \|W_{MD}(:, j)\|} \quad (14)$$

Building LDAP-WMPS prediction model

The flowchart of the LDAP-WMPS model is shown in Fig 2. The LDAP-WMPS model was divided into three parts: the first step calculated the disease projection score; the second step calculated the lncRNA projection score; and the third step fused the disease projection score and the lncRNA projection score proportionally, which were then normalized to get the prediction score matrix [46].

The disease projection score was defined by the following formula:

$$DSP(i, j) = \frac{W_{LD}(i, :) \times IDS(:, j)}{\|W_{LD}(i, :)\|} \quad (15)$$

where $W_{LD}(i, :)$ is the vector formed by the i th row of the lncRNA-disease weight matrix, which represents the relationship score between lncRNA i and various diseases. Its calculation

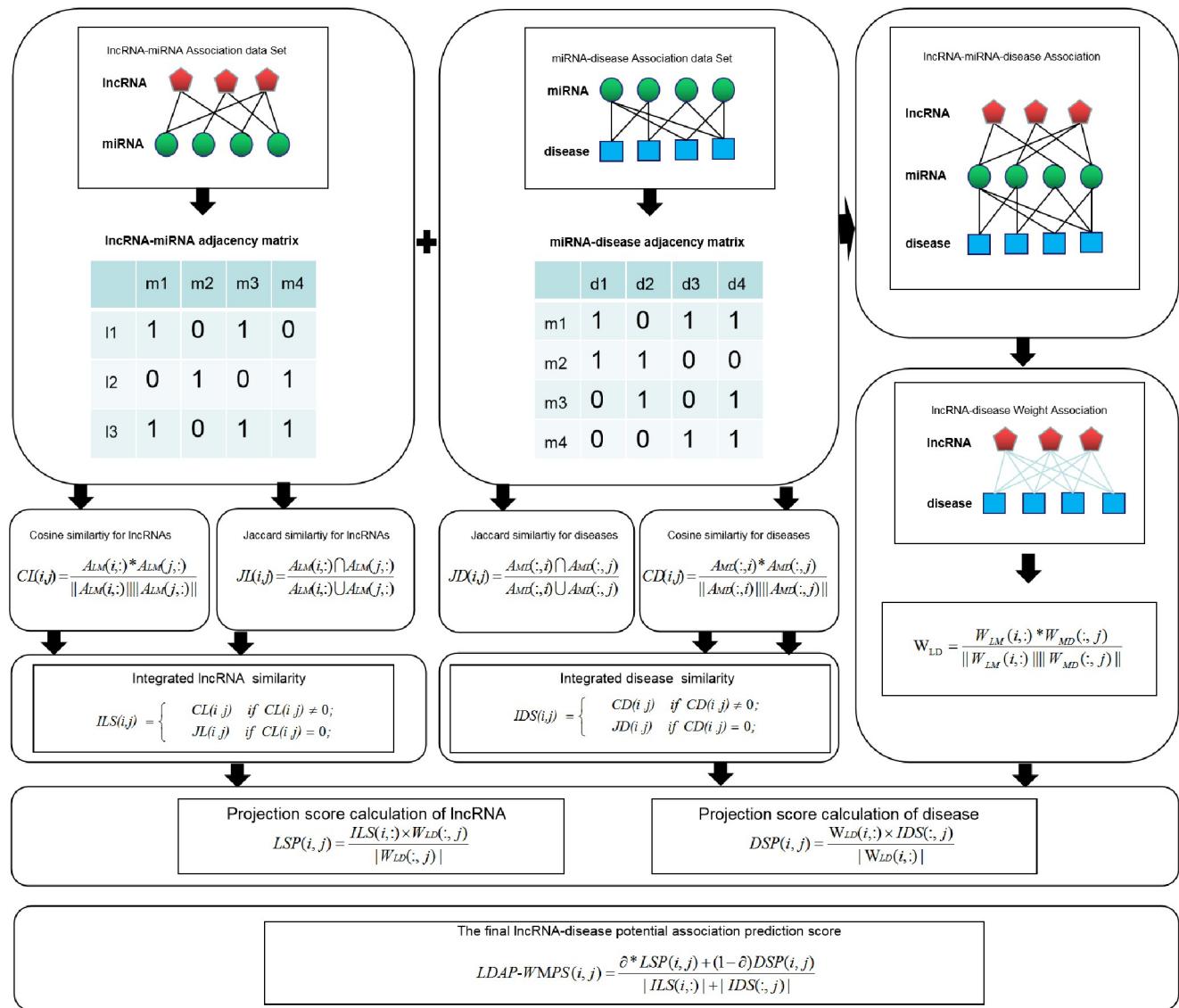


Fig 2. Flow chart of LDAP-WMPS applied to lncRNA-disease association prediction.

<https://doi.org/10.1371/journal.pone.0278817.g002>

process is shown in Fig 1. $IDS(:,j)$ is the vector formed by column j of the integrated disease similarity matrix, which represents the vector composed of the similarity between disease j and other diseases. $|W_{LD}(i,:)|$ represents the module length of disease i -related lncRNA component vector. $DSP(i,j)$ is the projection score of the disease. The multidimensional similarity relation was transformed into a concrete value by projection.

The projection score of lncRNA was defined as follows:

$$LSP(i,j) = \frac{ILS(i,:) \times W_{LD}(:,j)}{\|W_{LD}(:,j)\|} \quad (16)$$

In the aforementioned formula, $ILS(i,:)$ is the vector formed by the i th row of the functional similarity matrix of lncRNA, which represents the vector composed of the similarity between

lncRNA i and other kinds of lncRNA. $W_{LD}(:, j)$ is the vector formed by column j of lncRNA-disease relationship weight matrix, which represents the relationship score between disease j and various lncRNAs. $|W_{LD}(:, j)|$ is the module length of lncRNA i -related disease component vector. $LSP(i, j)$ is the projection score of lncRNA.

The final lncRNA-disease potential association prediction score matrix was formed by fusing lncRNA projection score with the disease projection score, defined as:

$$LDAP - WDPS(i, j) = \frac{\partial * LSP(i, j) + (1 - \partial)DSP(i, j)}{|ILS(i, :)| + |IDS(:, j)|} \quad (17)$$

where $LDAP - WDPS(i, j)$ is the final relationship score between lncRNA i and disease j . $|ILS(i, :)|$ is the module length of the lncRNA composition vector similar to lncRNA i in integrated lncRNA similarity matrix, and $|IDS(:, j)|$ is the module length of the disease composition vector similar to disease j in integrated disease similarity matrix. ∂ is the proportion of the lncRNA projection score and the disease projection score in the fusion score calculation.

Results

Performance evaluation

The performance of the LDAP-WMPS model was evaluated using the LOOCV framework, and each known disease-lncRNA relationship was left out in turn as a test sample. How well this test sample was ranked relative to the candidate samples (all the disease-lncRNA pairs without the evidence to confirm their relationships) was evaluated. When the rank of this test sample exceeded the given threshold, this model was considered to provide a successful prediction. The evaluation process (LOOCV) came from the reference [39]. For a more detailed code, the code was link in reference [39]. The results were compared with other prediction models using LOOCV, and with other prediction models for LOOCV. The true-positive rate (TPR) and false-positive rate (FPR) were calculated to obtain the ROC and the AUC for intuitive evaluation:

$$TPR = \frac{TP}{TP + FN} \quad (18)$$

$$FPR = \frac{FP}{FP + TN} \quad (19)$$

The ROC curve was drawn with TPR and FPR, and the AUC was calculated.

Comparison with other advanced models

The LDAP-WMPS model was compared with other advanced models to prove the effectiveness of the LDAP-WMPS model. Considering that the dataset used in this model was the same as those of NBCLDA [38] and CFNMC [39] models, the NBCLDA and CFNMC models were chosen as the comparison models. The ROC and AUC were obtained by applying three different models to the same dataset. After comparison, the LDAP-WMPS model was slightly better than the other methods in the ROC curve, and the AUC reached the value 0.8822. The highest AUC values of the NBCLDA and CFNMC models were 0.8521 and 0.8576, respectively. The results showed that the proposed method was slightly better than the CFNMC method. The results are shown in Table 1 and Figs 3 and 4. All models were statistically tested, and the P value was less than 0.05, indicating that all models had statistical significance.

Table 1. AUC values of LDAP-WMPS model and other models under LOOCV framework under the same dataset.

Method	AUC	AUPR	P value
LDAP-WMPS	0.8822	0.0355	<0.05
NBCLDA	0.8521	0.0238	<0.05
CFNBC	0.8576	0.0241	<0.05

<https://doi.org/10.1371/journal.pone.0278817.t001>

Analysis of parameters

In the proposed model, a parameter ∂ was introduced. The range of the parameter ∂ was [0,1]. When $\partial = 0$, only the disease projection score was used for the final score calculation; when $\partial = 1$, only the lncRNA projection score was used for the final score calculation. The results are shown in Fig 5. Obviously, when $\partial = 0.52$, AUC reached the highest value of 0.8822. The models using and not using weight matrix were evaluated, respectively, to further prove the effectiveness of the proposed lncRNA-disease weight matrix, and the results are shown in Fig 6. It was obvious that the weight matrix of the proposed model effectively improved the prediction ability.

Case studies

Tumor refers to a new organism formed by the proliferation of local tissue cells under the action of various oncogenic factors because this new organism is mostly space-occupying massive protuberance, also known as vegetation. According to the cellular characteristics of tumors and the degree of harm to the body, tumors are divided into benign tumors and malignant tumors: benign tumors can be removed by surgery and do not metastasize and relapse. Malignant tumors, often called cancer, are easy to metastasize, difficult to cure by surgery, and have a possibility of recurrence after cure [47]. Adenocarcinoma and colorectal cancer were studied to further prove the practicability of LDAP-WMPS in lncRNA-disease association prediction. The first 20 pieces of information about LDAP-WMPS predicting adenocarcinoma and colorectal cancer are shown in Tables 2 and 3, respectively.

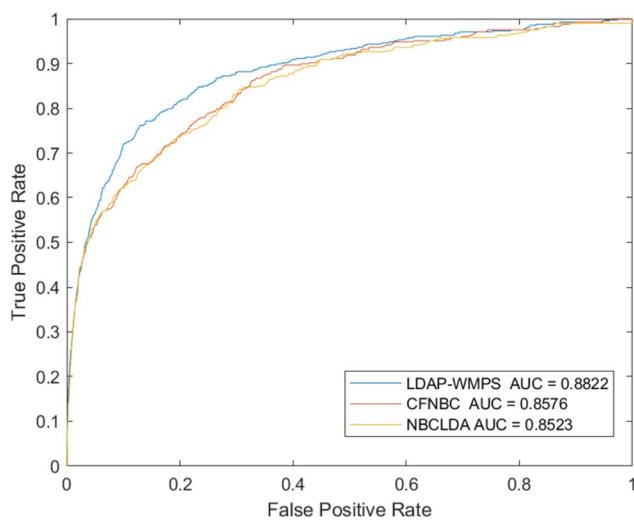


Fig 3. The performance of LDAP-WMPS and others models in terms of ROC curves and AUCs based on 407 known lncRNA-disease associations under the framework of LOOCV.

<https://doi.org/10.1371/journal.pone.0278817.g003>

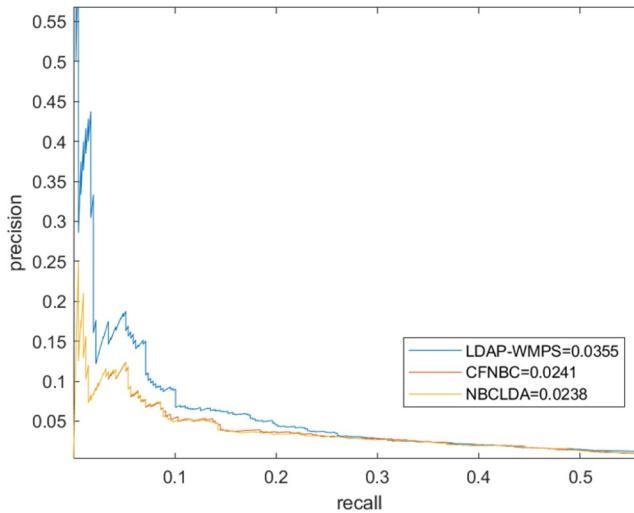


Fig 4. The performance of LDAP-WMPS and others models in terms of PR curves and AUPRs based on 407 known lncRNA-disease associations under the framework of LOOCV.

<https://doi.org/10.1371/journal.pone.0278817.g004>

Colorectal cancer is a common cancer type. Its incidence rate and mortality rate are high in the world. In 2018 alone, the number of new cases reached nearly 2 million, and the number of deaths was nearly 900,000. Some data showed that about 5.2% of men and 4.8% of women were at risk of colorectal cancer in the United States, and the mortality caused by colorectal cancer was close to 33% [48]. Many studies showed that lncRNA was closely related to colorectal cancer. In the prediction results of this study, 12 of the first 20 lncRNAs associated with colorectal cancer had been already proved by relevant medicine: lncRNA XIST expedited metastasis and modulated epithelial-mesenchymal transition in colorectal cancer [49]; lncRNA SNHG16 promoted colorectal cancer cell proliferation, migration, and epithelial-mesenchymal transition through miR-124-3p/MCP-1 [50]; and lncRNA MALAT1 promoted the colorectal cancer malignancy by increasing DCP1A expression and miR203 downregulation [51]. The lncRNA HCG18 promoted the growth and invasion of colorectal cancer cells through

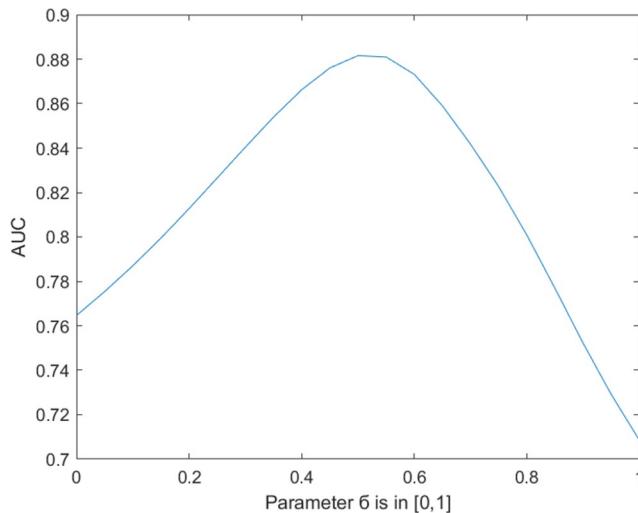


Fig 5. Influence curve of parameter a on AUC within [0,1].

<https://doi.org/10.1371/journal.pone.0278817.g005>

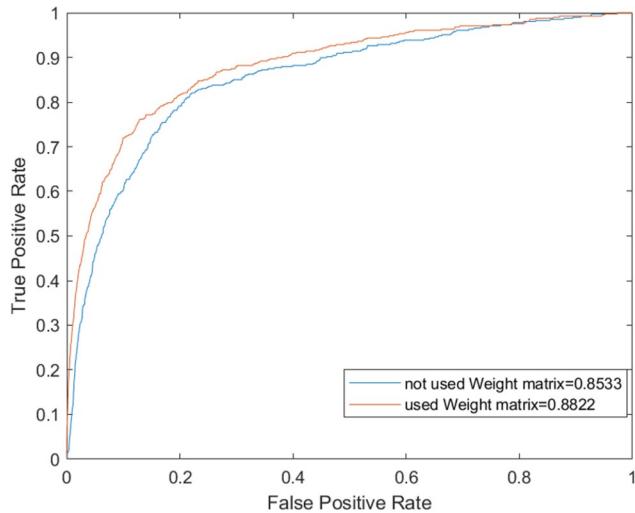


Fig 6. Comparison of ROC curve calculated with weight matrix and ROC curve calculated without weight matrix.

<https://doi.org/10.1371/journal.pone.0278817.g006>

sponging miR-1271 and upregulating MTDH [52]. lncRNA FGD5-AS1 promoted colorectal cancer cell proliferation, migration, and invasion through upregulating CDCA7 via sponging miR-302e [53]. lncRNA TUG1 mediated 5-fluorouracil resistance by acting as a competing endogenous RNA of miR-197-3p in colorectal cancer [54].

Adenocarcinoma is a kind of lung cancer. It is least related to smoking, accounting for 40% of primary adenocarcinoma. It is often located in the peripheral part of the lung, but also involves the pleura and the formation of associated scarring and pleural effusion. Extensive resection should be performed because of the invasive growth of adenocarcinoma. The rate of lymph node metastasis of adenocarcinoma is high, which can be as high as 36%-47%. It is easy to relapse and has a poor prognosis. Lin Guoji reported 68 cases of adenocarcinoma. The 5-year and 10-year cure rates were 43.9% and 29.0%, respectively [55]. In the prediction results of the proposed model, 14 of the first 20 lncRNAs associated with adenocarcinoma had been already proved by relevant medicine: lncRNA XIST promoted human lung adenocarcinoma cells to cisplatin resistance via let-7i/BAG-1 axis [56]; lncRNA MALAT1 promoted gastric adenocarcinoma through the miR-181a-5p/AKT3 axis [57]; lncRNA CTB-89H12.4 regulated

Table 2. Top 20 lncRNA of colorectal neoplasms predicted by LDAP-WMPS.

lncRNA	Evidence (PMID)	Rank
XIST	28837144	1
MALAT1	25031737;21503572	3
DCP1A	29964337	4
KCNQ1OT1	16965397;11340379	5
NEAT1	30185232	8
OIP5-AS1	29773344	9
HCG18	31854468	10
FGD5-AS1	31332696	13
TUG1	31528224	14
RP4-773N10.5	31966592	15
SNHG16	32859986	18
GAS5	31619268	20

<https://doi.org/10.1371/journal.pone.0278817.t002>

Table 3. Top 20 lncRNA of adenocarcinoma predicted by LDAP-WMPS.

lncRNA	Evidence (PMID)	Rank
XIST	28961027	1
MALAT1	31480991	3
DCP1A	25089265	4
RP6-24A23.7	28299977	5
KCNQ1OT1	30932685	7
HCG18	32559619	8
NEAT1	30036873	9
OIP5-AS1	32669972	10
CTB-89H12.4	26975529	12
FGD5-AS1	33416094	13
SNHG16	31580045	16
SENP3-EIF4A1	32602848	17
TUG1	29960845	18
LINC00662	33108738	20

<https://doi.org/10.1371/journal.pone.0278817.t003>

phosphatase and tensin homolog expression in prostate cancer [58]; lncRNA HCG18 acted on the oncogene in lung adenocarcinoma and enhanced lung adenocarcinoma progression by targeting miR-34a-5p/HMMR axis [59]; and lncRNA SNHG16 promoted cell proliferation and invasion in lung adenocarcinoma via sponging let-7a-5p [60].

Next, we took the XIST gene as an example for further analysis to verify whether it might be associated with Colorectal cancer. In our study, we divided all Colorectal cancer patient samples into high and low expression groups. This phenomenon was observed by survival analysis. That, the survival time of Colorectal cancer patients in the XIST gene high expression group was relatively short, as shown in Fig 7.

Discussion

Investigating the lncRNA-disease relationship is not only of great significance to the treatment of diseases but also helpful to explore the mystery of the human body. Using artificial

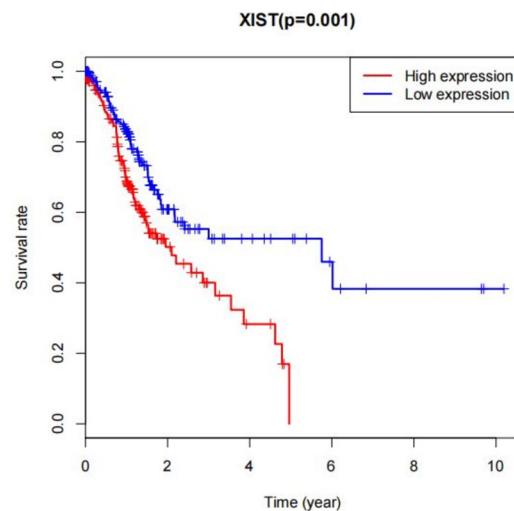


Fig 7. Survival period of XIST in high and low expression.

<https://doi.org/10.1371/journal.pone.0278817.g007>

intelligence to mine the existing medical data not only improves the use rate of data but also speeds up the process of medical intelligence. In this study, a computational model LDAP-WMPS was proposed. In this model, a weight allocation algorithm based on the lncRNA-miRNA-disease triple network and an lncRNA-disease relationship weight calculation method were proposed. The lncRNA-disease weight matrix was combined with the improved projection algorithm to calculate the relationship between each lncRNA and disease the interaction between lncRNA and disease information was obtained. Compared with the other three models, LDAP-WMPS was slightly better in AUC. Twelve of the first 20 lncRNAs were confirmed to predict the relationship between adenocarcinoma and colorectal cancer, which also proved the reliability of LDAP-WMPS. In addition, the proposed model was based on the lncRNA-miRNA relationship and miRNA-disease relationship to achieve the prediction of the lncRNA-disease relationship. The present relatively perfect lncRNA-miRNA relationship dataset and miRNA-disease relationship dataset to predict the lncRNA-disease relationship could effectively avoid the current lack of lncRNA-disease relationship data in data prediction. However, the proposed model also had some limitations. Many kinds of data were required for prediction, such as the miRNA-disease relationship dataset and lncRNA-disease relationship dataset. At the same time, the density of the aforementioned two datasets had a great impact on the final prediction results.

Conclusions

The main contributions of this study were as follows: (1) An integrated lncRNA similarity calculation method and an integrated disease similarity calculation method were proposed. The similarity was calculated by a variety of similarity calculation methods, which could effectively avoid the defects of insufficient similarity obtained by a single similarity calculation method and improve the prediction ability of the model to unknown relationships. (2) Based on the weight distribution of lncRNA-miRNA-disease triple network, a method of lncRNA-disease weight calculation was proposed. This method could effectively associate lncRNA-miRNA dataset with miRNA-disease dataset and help in indirectly predicting lncRNA-disease relationship through lncRNA-miRNA dataset and miRNA-disease dataset. (3) The existing consistency projection scoring formula was improved, and the proportion of the projection of the lncRNA part and the projection of disease part was adjusted in the final score to improve the prediction ability. (4) The lncRNA-disease relationship could be predicted by the LDAP-WMPS model without relying on the known lncRNA-disease relationship data.

Supporting information

S1 Data.

(ZIP)

Acknowledgments

The authors thank all those who have made suggestions for this article.

Author Contributions

Conceptualization: Bo Wang.

Data curation: Xiao-xin Du.

Formal analysis: Chao Zhang.

Investigation: Jing-you Li.

Methodology: Xiao-xin Du.

Validation: Jing-you Li.

Visualization: Xiao-dong Zheng.

Writing – original draft: Chao Zhang.

References

1. Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet*. 2006; 15 Spec No 1:R17–R29. <https://doi.org/10.1093/hmg/ddl046> PMID: 16651366.
2. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet*. 2020; 21(2):102–117. <https://doi.org/10.1038/s41576-019-0184-5> PMID: 31729473.
3. Yi K, Zhang Y, Wang Y, et al. Long noncoding RNA and its role in virus infection and pathogenesis. *Front Biosci (Landmark Ed)*. 2019; 24:777–789. <https://doi.org/10.2741/4750> PMID: 30844712.
4. Chen L, Zhou Y, Li H. lncRNA, miRNA and lncRNA-miRNA interaction in viral infection. *Virus Res*. 2018; 257:25–32. <https://doi.org/10.1016/j.virusres.2018.08.018> PMID: 30165080.
5. Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol*. 2017; 18(9):962–972. <https://doi.org/10.1038/ni.3771> PMID: 28829444.
6. Guttman M, Garber M, Levin JZ, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs [published correction appears in *Nat Biotechnol*. 2010 Jul;28(7):756]. *Nat Biotechnol*. 2010; 28(5):503–510. <https://doi.org/10.1038/nbt.1633> PMID: 20436462.
7. Hüttenhofer A, Schattner P, Polacek N. Non-coding RNAs: hope or hype?. *Trends Genet*. 2005; 21(5):289–297. <https://doi.org/10.1016/j.tig.2005.03.007> PMID: 15851066.
8. Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription?. *Sci Adv*. 2017; 3(9):eaao2110. <https://doi.org/10.1126/sciadv.aao2110> PMID: 28959731.
9. Ju C, Liu R, Zhang YW, et al. Mesenchymal stem cell-associated lncRNA in osteogenic differentiation. *Biomed Pharmacother*. 2019; 115:108912. <https://doi.org/10.1016/j.bioph.2019.108912> PMID: 31048188.
10. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*. 2016; 17(1):47–62. <https://doi.org/10.1038/nrg.2015.10> PMID: 26666209.
11. Zhao W, Geng D, Li S, Chen Z, Sun M. lncRNA HOTAIR influences cell growth, migration, invasion, and apoptosis via the miR-20a-5p/HMGA2 axis in breast cancer. *Cancer Med*. 2018; 7(3):842–855. <https://doi.org/10.1002/cam4.1353> PMID: 29473328.
12. Ferrè F, Colantoni A, Helmer-Citterich M. Revealing protein-lncRNA interaction. *Brief Bioinform*. 2016; 17(1):106–116. <https://doi.org/10.1093/bib/bbv031> PMID: 26041786.
13. Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. *Cancer Res*. 2017; 77(15):3965–3981. <https://doi.org/10.1158/0008-5472.CAN-16-2634> PMID: 28701486.
14. Fernando TR, Contreras JR, Zampini M, et al. The lncRNA CASC15 regulates SOX4 expression in RUNX1-rearranged acute leukemia. *Mol Cancer*. 2017; 16(1):126. <https://doi.org/10.1186/s12943-017-0692-x> PMID: 28724437.
15. Delás MJ, Sabin LR, Dolzhenko E, et al. lncRNA requirements for mouse acute myeloid leukemia and normal differentiation. *Elife*. 2017; 6:e25607. <https://doi.org/10.7554/elife.25607> PMID: 28875933.
16. Feng SD, Yang JH, Yao CH, et al. Potential regulatory mechanisms of lncRNA in diabetes and its complications. *Biochem Cell Biol*. 2017; 95(3):361–367. <https://doi.org/10.1139/bcb-2016-0110> PMID: 28177764.
17. Suwal A, Hao JL, Liu XF, et al. NONRATT021972 long-noncoding RNA: A promising lncRNA in diabetes-related diseases. *Int J Med Sci*. 2019; 16(6):902–908. <https://doi.org/10.7150/ijms.34200> PMID: 31337964.
18. Hua JT, Ahmed M, Guo H, et al. Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. *Cell*. 2018; 174(3):564–575.e18. <https://doi.org/10.1016/j.cell.2018.06.014> PMID: 30033362.
19. Wu M, Huang Y, Chen T, et al. lncRNA MEG3 inhibits the progression of prostate cancer by modulating miR-9-5p/QKI-5 axis. *J Cell Mol Med*. 2019; 23(1):29–38. <https://doi.org/10.1111/jcmm.13658> PMID: 30565858.

20. Loewen G, Jayawickramarajah J, Zhuo Y, Shan B. Functions of lncRNA HOTAIR in lung cancer. *J Hematol Oncol*. 2014; 7:90. <https://doi.org/10.1186/s13045-014-0090-4> PMID: 25491133.
21. Zhang YX, Yuan J, Gao ZM, Zhang ZG. lncRNA TUC338 promotes invasion of lung cancer by activating MAPK pathway. *Eur Rev Med Pharmacol Sci*. 2018; 22(2):443–449. https://doi.org/10.26355/eurrev_201801_14193 PMID: 29424901.
22. Huang JZ, Chen M, Chen D, et al. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol Cell*. 2017; 68(1):171–184.e6. <https://doi.org/10.1016/j.molcel.2017.09.015> PMID: 28985503.
23. Wu Q, Meng WY, Jie Y, Zhao H. lncRNA MALAT1 induces colon cancer development by regulating miR-129-5p/HMGB1 axis. *J Cell Physiol*. 2018; 233(9):6750–6757. <https://doi.org/10.1002/jcp.26383> PMID: 29226325.
24. Huang Y. The novel regulatory role of lncRNA-miRNA-mRNA axis in cardiovascular diseases. *J Cell Mol Med*. 2018; 22(12):5768–5775. <https://doi.org/10.1111/jcmm.13866> PMID: 30188595.
25. Bär C, Chatterjee S, Thum T. Long Noncoding RNAs in Cardiovascular Pathology, Diagnosis, and Therapy. *Circulation*. 2016; 134(19):1484–1499. <https://doi.org/10.1161/CIRCULATIONAHA.116.023686> PMID: 27821419.
26. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015; 21(11):1253–1261. <https://doi.org/10.1038/nm.3981> PMID: 26540387.
27. Li J, Xuan Z, Liu C. Long non-coding RNAs and complex human diseases. *Int J Mol Sci*. 2013; 14(9):18790–18808. <https://doi.org/10.3390/ijms140918790> PMID: 24036441.
28. Chen X, Sun YZ, Guan NN, et al. Computational models for lncRNA function prediction and functional similarity calculation. *Brief Funct Genomics*. 2019; 18(1):58–82. <https://doi.org/10.1093/bfgp/ely031> PMID: 30247501.
29. Wang CC, Han CD, Zhao Q, Chen X. Circular RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform*. 2021 Nov 5; 22(6):bbab286. <https://doi.org/10.1093/bib/bbab286> PMID: 34329377.
30. Liu D, Huang Y, Nie W, Zhang J, Deng L. SMALF: miRNA-disease associations prediction based on stacked autoencoder and XGBoost. *BMC Bioinformatics*. 2021 Apr 28; 22(1):219. <https://doi.org/10.1186/s12859-021-04135-2> PMID: 33910505.
31. Chen X, Li TH, Zhao Y, Wang CC, Zhu CC. Deep-belief network for predicting potential miRNA-disease associations. *Brief Bioinform*. 2021 May 20; 22(3):bbaa186. <https://doi.org/10.1093/bib/bbaa186> PMID: 34020550.
32. Zhang L, Liu T, Chen H, Zhao Q, Liu H. Predicting lncRNA-miRNA interactions based on interactome network and graphlet interaction. *Genomics*. 2021; 113(3):874–880. <https://doi.org/10.1016/j.ygeno.2021.02.002> PMID: 33588070.
33. Zhang L, Yang P, Feng H, Zhao Q, Liu H. Using Network Distance Analysis to Predict lncRNA-miRNA Interactions. *Interdiscip Sci*. 2021; 13(3):535–545. <https://doi.org/10.1007/s12539-021-00458-z> PMID: 34232474.
34. Fu G, Wang J, Domeniconi C, Yu G. Matrix factorization-based data fusion for the prediction of lncRNA-disease associations. *Bioinformatics*. 2018 May 1; 34(9):1529–1537. <https://doi.org/10.1093/bioinformatics/btx794> PMID: 29228285.
35. Chen X, Sun LG, Zhao Y. NCMCMDA: miRNA-disease association prediction through neighborhood constraint matrix completion. *Brief Bioinform*. 2021 Jan 18; 22(1):485–496. <https://doi.org/10.1093/bib/bbz159> PMID: 31927572.
36. Zhang Y, Chen M, Li A, Cheng X, Jin H, Liu Y. LDAI-ISPS: lncRNA-Disease Associations Inference Based on Integrated Space Projection Scores. *Int J Mol Sci*. 2020; 21(4):1508. <https://doi.org/10.3390/ijms21041508> PMID: 32098405.
37. Wang Y, Yu G, Wang J, Fu G, Guo M, Domeniconi C. Weighted matrix factorization on multi-relational data for lncRNA-disease association prediction. *Methods*. 2020 Feb 15; 173:32–43. <https://doi.org/10.1016/j.ymeth.2019.06.015> PMID: 31226302.
38. Yu J, Ping P, Wang L, Kuang L, Li X, Wu Z. A Novel Probability Model for lncRNA-Disease Association Prediction Based on the Naïve Bayesian Classifier. *Genes (Basel)*. 2018; 9(7):345. <https://doi.org/10.3390/genes9070345> PMID: 29986541.
39. Yu J, Xuan Z, Feng X, Zou Q, Wang L. A novel collaborative filtering model for lncRNA-disease association prediction based on the Naïve Bayesian classifier. *BMC Bioinformatics*. 2019; 20(1):396. <https://doi.org/10.1186/s12859-019-2985-0> PMID: 31315558.
40. Cui T, Zhang L, Huang Y, et al. MNDR v2.0: an updated resource of ncRNA-disease associations in mammals. *Nucleic Acids Res*. 2018; 46(D1):D371–D374. <https://doi.org/10.1093/nar/gkx1025> PMID: 29106639.

41. Li Y, Qiu C, Tu J, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* 2014; 42(Database issue):D1070–D1074. <https://doi.org/10.1093/nar/gkt1023> PMID: 24194601.
42. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* 2014; 42(Database issue):D92–D97. <https://doi.org/10.1093/nar/gkt1248> PMID: 24297251.
43. Xie G, Huang Z, Liu Z, Lin Z, Ma L. NCPHLDA: a novel method for human lncRNA-disease association prediction based on network consistency projection. *Mol Omics.* 2019 Dec 2; 15(6):442–450. <https://doi.org/10.1039/c9mo00092e> PMID: 31686064.
44. Yang X, Gao L, Guo X, et al. A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One.* 2014; 9(1):e87797. <https://doi.org/10.1371/journal.pone.0087797> PMID: 24498199.
45. Gu C, Liao B, Li X, Li K. Network Consistency Projection for Human miRNA-Disease Associations Inference. *Sci Rep.* 2016 Oct 25; 6:36054. <https://doi.org/10.1038/srep36054> PMID: 27779232; PMCID: PMC5078764.
46. Ding L, Wang M, Sun D, Li A. TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph. *Sci Rep.* 2018 Jan 18; 8(1):1065. <https://doi.org/10.1038/s41598-018-19357-3> PMID: 29348552.
47. Gerrand C, Athanasou N, Brennan B, et al. UK guidelines for the management of bone sarcomas. *Clin Sarcoma Res.* 2016; 6:7. <https://doi.org/10.1186/s13569-016-0047-1> PMID: 27148438.
48. Brambilla E, Travis WD, Colby TV, Corrin B, Shimosato Y. The new World Health Organization classification of lung tumours. *Eur Respir J.* 2001; 18(6):1059–1068. <https://doi.org/10.1183/09031936.01.00275301> PMID: 11829087.
49. Chen DL, Chen LZ, Lu YX, et al. Long noncoding RNA XIST expedites metastasis and modulates epithelial-mesenchymal transition in colorectal cancer. *Cell Death Dis.* 2017; 8(8):e3011. <https://doi.org/10.1038/cddis.2017.421> PMID: 28837144.
50. Chen ZY, Wang XY, Yang YM, et al. lncRNA SNHG16 promotes colorectal cancer cell proliferation, migration, and epithelial-mesenchymal transition through miR-124-3p/MCP-1. *Gene Ther.* 2020; 10:1038/s41434-020-0176-2. <https://doi.org/10.1038/s41434-020-0176-2> PMID: 32859986
51. Wu C, Zhu X, Tao K, et al. MALAT1 promotes the colorectal cancer malignancy by increasing DCP1A expression and miR203 downregulation. *Mol Carcinog.* 2018; 57(10):1421–1431. <https://doi.org/10.1002/mc.22868> PMID: 29964337.
52. Li S, Wu T, Zhang D, Sun X, Zhang X. The long non-coding RNA HCG18 promotes the growth and invasion of colorectal cancer cells through sponging miR-1271 and upregulating MTDH/Wnt/β-catenin. *Clin Exp Pharmacol Physiol.* 2020; 47(4):703–712. <https://doi.org/10.1111/1440-1681.13230> PMID: 31854468.
53. Li D, Jiang X, Zhang X, Cao G, Wang D, Chen Z. Long noncoding RNA FGD5-AS1 promotes colorectal cancer cell proliferation, migration, and invasion through upregulating CDCA7 via sponging miR-302e. *In Vitro Cell Dev Biol Anim.* 2019; 55(8):577–585. <https://doi.org/10.1007/s11626-019-00376-x> PMID: 31332696.
54. Wang M, Hu H, Wang Y, et al. Long non-coding RNA TUG1 mediates 5-fluorouracil resistance by acting as a ceRNA of miR-197-3p in colorectal cancer. *J Cancer.* 2019; 10(19):4603–4613. <https://doi.org/10.7150/jca.32065> PMID: 31528224.
55. Kuhn E, Morbini P, Cancellieri A, Damiani S, Cavazza A, Comin CE. Adenocarcinoma classification: patterns and prognosis. *Pathologica.* 2018; 110(1):5–11. <https://doi.org/10.259909> PMID: 30259909.
56. Sun J, Pan LM, Chen LB, Wang Y. lncRNA XIST promotes human lung adenocarcinoma cells to cis-platin resistance via let-7i/BAG-1 axis. *Cell Cycle.* 2017; 16(21):2100–2107. <https://doi.org/10.1080/15384101.2017.1361071> PMID: 28961027.
57. Lu Z, Luo T, Pang T, et al. MALAT1 promotes gastric adenocarcinoma through the MALAT1/miR-181a-5p/AKT3 axis. *Open Biol.* 2019; 9(9):190095. <https://doi.org/10.1098/rsob.190095> PMID: 31480991.
58. Du Z, Sun T, Hacisuleyman E, et al. Integrative analyses reveal a long noncoding RNA-mediated sponge regulatory network in prostate cancer. *Nat Commun.* 2016; 7:10982. <https://doi.org/10.1038/ncomms10982> PMID: 26975529.
59. Li W, Pan T, Jiang W, Zhao H. HCG18/miR-34a-5p/HMMR axis accelerates the progression of lung adenocarcinoma. *Biomed Pharmacother.* 2020; 129:110217. <https://doi.org/10.1016/j.biopha.2020.110217> PMID: 32559619.
60. Guo Y, Zhen H, Chen W, Sun Y, Li K, Feng E. Long non-coding RNA SNHG16 promotes cell proliferation and invasion in lung adenocarcinoma via sponging let-7a-5p. *Minerva Chir.* 2019; 74(6):509–511. <https://doi.org/10.23736/S0026-4733.19.08140-9> PMID: 31580045.