IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2020, accepted June 10, 2020, date of publication June 15, 2020, date of current version June 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3002588

A Novel lncRNA-Disease Association Prediction Model Using Laplacian Regularized Least Squares and Space Projection-Federated Method

MIN CHEN[1], YAN PENG[2], ANG LI[1], YINWEI DENG[1,3], AND ZEJUN LI[1]

[1]School of Computer Science and Technology, Hunan Institute of Technology, Hengyang 421002, China
[2]School of International Communication, Hunan Institute of Technology, Hengyang 421002, China
[3]Hainan Key Laboratory for Computational Science and Application, Haikou 571158, China

Corresponding authors: Yan Peng (annie_pengyan@hnit.edu.cn), Ang Li (liang@hnit.edu.cn), and Yinwei Deng (37619569@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772192 and Grant 61672223, in part by the Nature Science Foundation of Hunan Province, China, under Grant 2018JJ2085 and Grant 2019JJ40064, in part by the Scientific Research Project of Education Department of Hunan Province, China, under Grant 19A125 and Grant 19B142, in part by the Major Cultivation Projects of Hunan Institute of Technology under Grant 2017HGPY001, and in part by the Science and Technology Innovative Research Team of Hunan Institute of Technology under Grant TD18005.

ABSTRACT Recent studies indicated that numerous long noncoding RNAs (lncRNAs) are closely related to human diseases and can serve as potential biomarkers and drug targets for complex diseases. Therefore, identifying lncRNAs associated with diseases through computational methods is conducive to the exploration of disease pathogenesis. Most previous studies had shortcomings, such as low prediction accuracy, the need for negative samples, and weak generalization. Such studies established shallow prediction models and failed to fully capture the complex relationships among lncRNA–disease associations, lncRNA similarity, and disease similarity. LRLSSP, a new computational method based on Laplacian regularized least squares (LRLS) and space projection was used to predict candidate disease lncRNAs in this study. LRLSSP deeply integrates information on lncRNA similarity, disease similarity, and known lncRNA–disease associations. The estimated score of lncRNA–disease association was obtained through LRLS, and network projection was utilized to reliably predict disease-related lncRNAs. Leave-one-out cross validation(LOOCV) was implemented to evaluate the prediction performance of LRLSSP. Results showed that LRLSSP performed was better than other state-of-the-art methods in predicting lncRNA–disease associations. In addition, case studies conducted on melanoma,cervical cancer, ovarian cancer and breast cancer indicated that LRLSSP can discover potential and novel lncRNA–disease associations. Overall, the results demonstrated that LRLSSP may serve as a reliable and effective computational tool for disease-related lncRNAs prediction.

INDEX TERMS Disease similarity, lncRNA similarity, Laplacian regularized least squares, space projection, computational prediction model.

I. INTRODUCTION

LncRNAs are RNA molecules that are not translated into proteins and exceed 200 nt in length. lncRNAs have long been considered as transcriptional noise. Considerable evidence suggests that lncRNAs play fundamental and key regulatory roles in important biological processes, including chromatin remodeling, epigenetic regulation, genomic splicing, immune response, and cell cycle control [1]–[4]. The mutations and dysregulation of lncRNAs are related to the occurrence and development of various complex diseases [5]. For example, HOTAIR promotes serous ovarian cancer cell proliferation by regulating cell cycle arrest and apoptosis [6]. TDRG1 can enhance the tumorigenicity of endometrial cancer by binding to and targeting VEGF-A proteins [7]. MEG3 is involved in the epigenetic regulation of epithelial–mesenchymal transformation in lung cancer cell lines [8]. In addition, GAS5 promotes bladder cancer cell apoptosis by inhibiting the transcription of EZH2 [9]. Therefore, studying the associations between lncRNAs and

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

diseases can help researchers to discover the pathogenesis of complex human diseases at the molecular level and facilitate their diagnosis and treatment. However, revealing the associations between lncRNAs and diseases through biological experiments requires considerable manpower, materials, and financial resources; thus, efficient computational methods should be used to predict the reliable associations between lncRNAs and diseases and to serve as a powerful supplement to biological experiments [10]. In light of this reason, further computational methods for identifying disease-related lncRNAs are urgently needed.

With the increase in available biological data, computational models have been developed to predict the association between lncRNAs and diseases [11]–[13]. These prediction methods can be divided into two categories. The first category includes network-based models, which are based on the general assumption that ''functionally related lncRNAs tend to phenotypically similar diseases, and vice versa'' [14]. Ping *et al.* [14] constructed a bipartite network based on known lncRNA-disease associations, and then found the network was closely followed a power law distribution. Finally, they predicted the lncRNA–disease associations by calculating the length of the path. Several researchers successfully implemented the random walk algorithm to infer the potential lncRNA -disease relationship, miRNA-disease relationship, and gene-disease relationship on heterogeneous networks. [15]–[26]. Sun *et al.* [15] constructed a functional similarity network of lncRNAs on the basis of a semantic similarity network of diseases and a known lncRNA–disease association network,then they implemented random walk with restart on the network. This method is a global prediction method but cannot be used to predict lncRNAs without known associations. Chen *et al.* [23] proposed an improved random walk with restart algorithm, Li *et al.* [27] proposed an improved local random walk method for lncRNA–disease association prediction to eliminated the limitations of traditional restart random walk algorithm. Fan *et al.* [28] first used positive pointwise mutual information to construct a large-scale lncRNA–disease heterogeneous network,and then implemented Random Walk with restart algorithm to predict lncRNA disease associations; Hu *et al.* [29] used interaction profile and gene ontology information to build lncRNA-disease network, and then utilized bi-random walks to predict lncRNA–disease associations. HyperGeometric distribution [10], label propagation algorithm [30], KATZ [31], and space projection [32]–[34] have been used to predict lncRNA–disease association, but although their accuracy is low. Xiao *et al.* [35] designed a method based on a simple path with limited length in a heterogeneous network to predict disease-related lncRNAs, but given the extremely simple short-decay function, its prediction performance should be improved with a machine learning method.

The second category includes machine learning-based models. Zhao *et al.* [36] and Yu *et al.* [37] used naive Bayesian classifier model, whereas Lan *et al.* [38] and Chen *et al.* [39] used bagging support vector machine (SVM) to predict lncRNA–disease association. Although such machine learning methods can predict lncRNA–disease associations, they are faced with a huge challenge, that is, the uncertain influence caused by the selection difficulty and bias of negative samples. To solve the problem as to the necessity of negative samples for machine learning methods, Chen and Yan [40] applied Laplacian regularized least squares algorithm (LRLSLDA) to predict lncRNA–disease associations. Chen *et al.* [41] proposed a new lncRNA similarity calculation method named LNCSIM, whereas Huang *et al.* [42] proposed another lncRNA similarity calculation method named ILNCSIM; Chen *et al.* [43] proposed a fuzzy measure-based lncRNA functional similarity calculation method and implemented LRLSLDA on the basis of these similarity construction methods, all of which achieved good prediction results.

Many researchers attempted to use deep learning in the field of lncRNA association prediction. Xuan *et al.* proposed a variety of methods on the basis of convolutional neural network for the prediction of disease-related lncRNAs [44]–[46]. In recent years, other machine learning-based models, such as randomized matrix completion algorithm [47], inductive matrix completion algorithm [48], and matrix factorization algorithm [49], have been put forward to predict lncRNA-associated diseases. Liu *et al.* [50] proposed a weighted graph regulated collaborative matrix factorization method to identify lncRNA–disease associations; Xuan *et al.* [51] used the probabilistic matrix factorization method to predict disease-related lncRNAs; Zhu *et al.* [52] integrated alternative lead squares and matrix factorization to identify lncRNA–disease associations; Zeng *et al.* [53] integrated deep learning and matrix factorization to identify lncRNA-disease association. However, parameters are difficult to determine via such methods. Xuan *et al.* [51] used the probabilistic matrix factorization method to predict disease-related lncRNAs.

From the above analysis,although existing computational methods have made considerable contribution to revealing lncRNA–disease associations, they still have different limitations. In this study, a reliable method for lncRNA–disease association prediction was proposed on the basis of LRLS and space projection. LRLSSP mainly comprises three steps. First, Gaussian interaction profile kernel similarity,disease semantic similarity, and lncRNA functional similarity are used to reconstruct an accurate similarity network for lncRNAs and diseases. Second, LRLS is used to estimate the association between lncRNAs and diseases to alleviate the effect of sparsity of lncRNA–disease association and positive and unlabeled problems. Finally, a network space projection is used to refine the lncRNA–disease association prediction effect. The area Under receiving operating characteristic (ROC) Curve (AUC) obtained by implementing LOOCV in three datasets is 0.9377, 0.9383 and 0.9467, respectively,which outperforms those of the three state-of-the-art methods, namely, NCPLDA [32], LDAI-ISPS [33], and, IIRWR [24], and to demonstrate the effectiveness

of LRLSSP. Moreover, LRLSSP was utilized in case studies of melanoma,cervical cancer, ovarian cancer and breast cancer to further validate its effectiveness. Overall, LRLSSP is effective in predicting lncRNA–disease associations and can be further applied to reveal other biological associations.

II. MATERIALS AND DATA PREPARATION

A. KNOWN lncRNA–DISEASE ASSOCIATIONS

Chen *et al.* [54] built the LncRNADisease, an lncRNA and disease database, which is publicly accessible at http://www.cuilab.cn/lncrnadisease, to investigate the associations between lncRNAs and human disease The 2013, 2015,and 2017 versions of the database were obtained for this work. After the data were processed, 156 lncRNAs (Supplementary Material S1), 190 diseases (Supplementary Material S2), and 352 lncRNA–disease associations (Supplementary Material S3) were obtained from the 2013 version (marked as dataset 1). A total of 285 lncRNAs (Supplementary Material S4), 226 diseases (Supplementary Material S5), and 621 lncRNA–disease associations (Supplementary Material S6) were obtained from the 2015 version. (marked as dataset 2). A total of 828 lncRNAs (Supplementary Material S7), 314 diseases (Supplementary Material S8), and 1695 lncRNA–disease associations (Supplementary Material S9) were obtained from the 2017 version. (marked as dataset 3). The matrix $L = \{l_1, l_2, \cdots, l_{nl}\}$ represents the lncRNA sets, the matrix $D = \{d_1, d_2, \cdots, d_{nd}\}$ denotes that disease sets, and the Boolean matrix $LD = (ld_{ij})_{nl\times nd}$ indicates lncRNA–disease associations. The value of ld_{ij} is 1, indicating that lncRNA l_i is related to disease d_j; otherwise, its value is 0.

B. DISEASE–DISEASE SEMANTIC SIMILARITY

In accordance with the work of Wang *et al.* [55], the disease semantic similarity was calculated on the basis of a disease's tree attributes in Medical Subject Headings. In this method, each disease corresponded to a directed acyclic graph. If two diseases share more disease items than others, the similarity between the two diseases would be remarkable. The data were obtained directly from the works of Zhang *et al.* [33], Li *et al.* [32] and Wang *et al.* [24] (Supplementary Materials S10,S11,and S12) and represented by matrix $DD = (dd_{ij})_{nd\times nd}$.

C. LncRNA–lncRNA FUNCTIONAL SIMILARITY

Jie *et al.* [15] used disease semantic similarity and known disease–lncRNA associations to calculate lncRNA functional similarity. In the present study, lncRNA–lncRNA functional similarity was obtained on the basis of relevant studies.

The calculation of the functional similarity of lncRNAs can be roughly divided into two steps. (1) Given any two lncRNAs l_i and l_j whose associated diseases are denoted as two vectors $D^{(l_i)} = \{d_{1'}, d_{2'}, \cdots, d_{m'}\} = \{d_{i'}\}_m \subset D$ and $D^{(l_i)} = \{d_{1''}, d_{2''}, \cdots, d_{m''}\} = \{d_{i''}\}_n \subset D$, respectively, the functional similarity of lncRNA l_i and l_j could be defined as follows:

$$ll_{ij} = \frac{\sum\limits_{d_t\in\mathbf{D}^{(l_i)}} S(d_t, \mathbf{D}^{(l_j)}) + \sum\limits_{d_t\in\mathbf{D}^{(l_j)}} S(d_t, \mathbf{D}^{(l_i)})}{m+n} \tag{1}$$

where m and n are the known numbers of diseases associated with lncRNAs l_i and l_j, respectively. $S(d_{i'}, D^{(l_i)})$ represents the correlation score between a given disease $d_{i'}$ and a given disease set $D^{(l_i)}$ calculated as follows:

$$S(d_{i'}, \mathbf{D}^{(l_j)}) = \max_{d_t\in\mathbf{D}^{(l_j)}} (dd_{i't}) \tag{2}$$

The matrix $LL = (ll_{ij})_{nl\times nl}$ was used to represent such data (Supplementary Materials S13,S14, and S15).

D. CENTRAL SIMILARITY OF THE GAUSSIAN INTERACTION PROFILE

A small number of known lncRNA–disease associations may lead to the sparse similarity matrix of diseases (lncRNAs). The central similarities of the Gaussian interaction profile were used to solve this problem and to address this defect. On the basis of Laarhoven's work [56], the similarity score between disease d_i and d_j is defined as follows:

$$gd_{ij} = \exp(-\gamma_d \|\mathbf{LD}(:, i) - \mathbf{LD}(:, j)\|^2) \tag{3}$$

where $\mathbf{LD}(:, i)$ is the ith column of the matrix LD, representing known lncRNA information associated with diseases d_i, γ_d is the Gaussian kernel bandwidth, and the calculation formula is as follows:

$$\gamma_d = \frac{\gamma'_d}{\frac{1}{nd}\sum_{i=1}^{nd} \|\mathbf{LD}(:, i)\|^2} \tag{4}$$

where based on previous studies [43], γ'_d is set to 1. Furthermore, the matrix $\mathbf{GD} = (gd_{ij})_{nd\times nd}$ was used to represent Gaussian interaction profile central similarities between diseases.

For lncRNAs, the central similarity of the Gaussian interaction profile between lncRNA l_i and lncRNA l_j could be calculated in a similar manner:

$$gl_{ij} = \exp(-\gamma_l \|\mathbf{LD}(i, :) - \mathbf{LD}(j, :)\|^2) \tag{5}$$

where $\mathbf{LD}(i, :)$ is the ith row of matrix $\mathbf{LD}_{nl\times nd}$, representing known lncRNA information associated with lncRNA l_i, and γ_d is calculated as follows:

$$\gamma_l = \frac{\gamma'_l}{\frac{1}{nl}\sum_{i=1}^{nl} \|\mathbf{LD}(i, :)\|^2} \tag{6}$$

where γ'_l is also set to 1, and the matrix $\mathbf{GL} = (gl_{ij})_{nl\times nl}$ is used to represent the central similarities of the Gaussian interaction profile between lncRNAs.

E. INTEGRATED SIMILARITY FOR lncRNAs AND DISEASES

After LL, GD, and GL were calculated, they were integrated with DD to obtain the final disease similarity matrix (denoted by $\mathbf{DD}^* = (dd_{ij}^*)_{nd\times nd}$) and the final lncRNA similarity matrix (denoted by $\mathbf{LL}^* = (ll_{ij}^*)_{nl\times nl}$). Given different diseases d_i and d_j, if the semantic similarity between the two is 0, then the final similarity value is set as the central similarity of the Gaussian interaction profile between the two diseases; otherwise, the final similarity value is set as the semantic similarity between the two diseases. The final similarity between lncRNAs was also calculated in the same way. For clarity, the detailed formula is described as follows:

$$dd_{ij}^* = \begin{cases} dd_{ij}, & \text{if } dd_{ij} \neq 0 \\ gd_{ij}, & \text{otherwise} \end{cases} \tag{7}$$

$$11_{ij}^* = \begin{cases} 11_{ij}, & \text{if } 11_{ij} \neq 0 \\ g_{ij}^1, & \text{otherwise} \end{cases} \tag{8}$$

III. LRLSSP METHOD

In this work, LRLSSP was used to uncover the association between lncRNAs and diseases. LRLSSP comprises three steps: (1) data preparation, (2) score estimation of lncRNA–disease association, and (3) score refinement of lncRNA–disease association.

In data preparation (Equations 7 and 8), the integrated disease similarity was obtained using semantic similarity and Gaussian interaction profile central similarities, and the integrated lncRNA similarity was obtained using functional similarity and the central similarities of the Gaussian interaction profile. LRLS algorithm was used to estimate the lncRNA–disease association score to prioritize lncRNA–disease interactions based on the integrated network. In the score refinement of lncRNA–disease association, space projection was used to obtain the final prediction scores of lncRNA–disease association. Figure 1 shows the entire workflow.

A. SCORE ESTIMATION OF lncRNA–DISEASE ASSOCIATION

LRLS algorithm is a data-dependent manifold regulative framework, and it has been successfully applied to drug–protein interaction prediction [57] and miRNA–disease association prediction [58].

From the perspective of diseases, the estimated scores can be obtained by solving the optimization problem of the following equation:

$$\min_{LD_d} \left\{ \|LD - LD_d\|_F^2 + \alpha \left\| LD_d D_{DD^*} LD_d^T \right\|_F^2 \right\} \tag{9}$$

where LD_d is the estimated score matrix of lncRNA–disease association based on disease space; D_{DD^*} is a diagonal matrix; $D_{DD^*}(i,i)$ is the sum of the ith row element of DD^*; $\|\cdot\|_F$ is the tradeoff parameter, and $\|\cdot\|_F$ is Frobenius norm. After Equation (9) was solved, the estimated score matrix of lncRNA–disease association was obtained on the basis of disease space LD_d as follows [57]:

$$LD_d = DD^*(DD^* + \alpha D_{DD^*} DD^*)^{-1} LD^T \tag{10}$$

In the same way, the estimated score matrix of lncRNA–disease association can be obtained on the basis of lncRNA space LD_l as Equation (11):

$$LD_l = LL^*(LL^* + \beta D_{LL^*} LL^*)^{-1} LD \tag{11}$$

Then, the score estimation matrix of the lncRNA–disease association can be defined as follows:

$$LD_e = \frac{LD_d^T + LD_l}{2} \tag{12}$$

B. SCORE REFINEMENT OF THE lncRNA–DISEASE ASSOCIATION

On the basis of the estimated lncRNA–disease association scores obtained from LRLS in the previous section, matrix space projection was used to obtain the final lncRNA–disease association scores. Thus, the prediction performance of the proposed model can be improved.

Here, the predicted score calculation of the association between lncRNA l_i and disease d_j was considered an example.The projection of the final lncRNA similarity matrix $\mathbf{LL}^*$ in the lncRNA–disease association score estimation matrix LD_e was used as the projection score on the basis of the lncRNA space to fully capture the information of disease similarity,lncRNA similarity,and lncRNA–disease associations. The score between lncRNA l_i and disease d_j was calculated as follows:

$$\mathbf{LD}_{pl}(i,j) = \frac{\mathbf{LL}^*(i,:) \times \mathbf{LD}_e(:,j)}{\|\mathbf{LD}_e(:,j)\|} \tag{13}$$

where $\|\mathbf{LD}_e(:,j)\|$ is the 2-norm of $\mathbf{LD}_e(:,j)$, and $\mathbf{LD}_{pl}(i,j)$ is the score of space projection based on lncRNA space between lncRNA l_i and disease d_j.

Similarly, the space projection scores can be obtained from the disease similarity network $\mathbf{DD}^*$ by

$$\mathbf{LD}_{pd}(i,j) = \frac{\mathbf{DD}^*(j,:) \times \mathbf{LD}_e^T(:,i)}{\left\|\mathbf{LD}_e^T(:,i)\right\|} \tag{14}$$

where $\left\|\mathbf{LD}_e^T(:,i)\right\|$ is the 2-norm of $\mathbf{LD}_e^T(:,i)$.

The prediction matrix based on lncRNA space and disease space are $\mathbf{LD}_{pl}$ and $\mathbf{LD}_{pd}$, respectively. Finally, $\mathbf{LD}_{pl}$ and $\mathbf{LD}_{pd}$ were integrated to obtain the final prediction score matrix between lncRNAs and diseases as follows:

$$\mathbf{LD}^* = \frac{\mathbf{LD}_{pl} + \mathbf{LD}_{pd}^T}{2} \tag{15}$$

IV. EXPERIMENTS AND RESULTS

A. PERFORMANCE EVALUATION OF LRLSSP

LOOCV was implemented on the two datasets to evaluate the predictive performance of LRLSSP. During the experiment, each lncRNA–disease association was used as the test sample, and other lncRNA–disease associations were considered the training samples until all lncRNA–disease associations

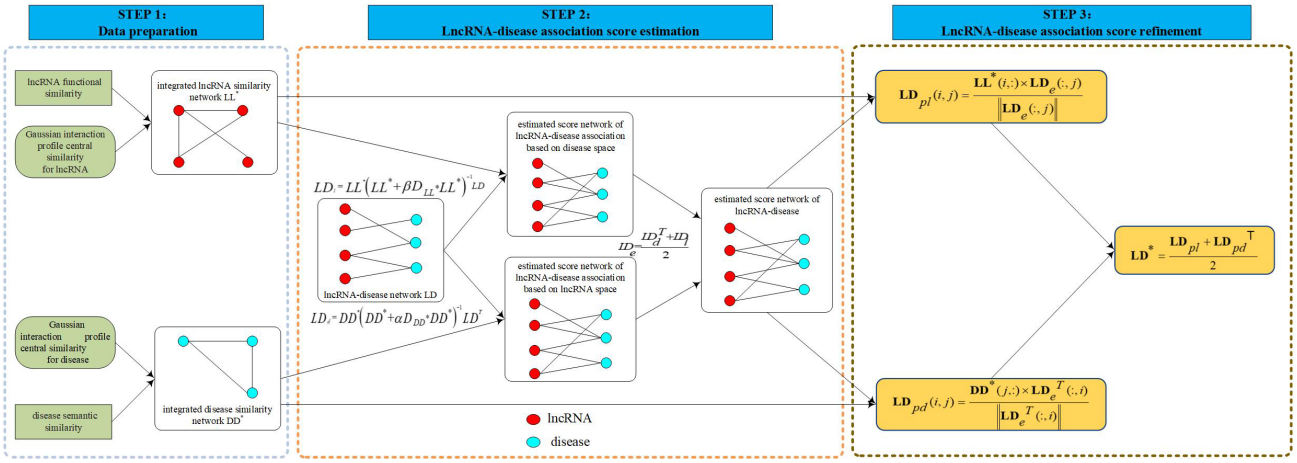

FIGURE 1. The flowchart of the whole modeling procedure.

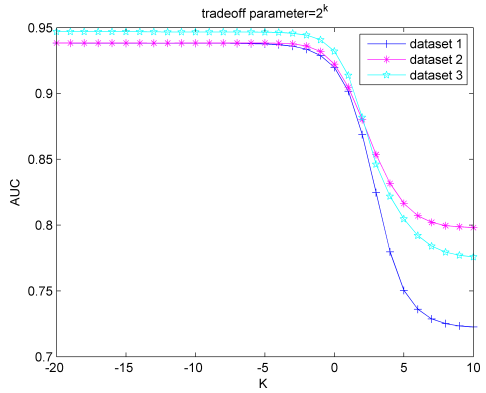

FIGURE 2. Influence of parameter variation on model predictive accuracy.

were tested once. The ROC curve and AUC were used as performance criteria for the evaluation. The higher the AUC value, the better is the model prediction performance. In addition, we also drew PR curve with recall as abscissa and precision as ordinate, and calculatedAUPR. The larger the AUPR value, the better the prediction performance.

B. PARAMETER SELECTION

The method proposed in this paper includes two parameters, namely, the tradeoff parameters α and β, which are equal. The tradeoff parameters were gradually increased from 2^{-20} to 2^{10} for LOOCV, and the AUC value was calculated to obtain the optimal parameter. Figure 2 shows the results of the LOOCV experiments implemented in the three datasets. In the figure, blue represents the AUC variation diagram on dataset 1,and magenta denotes the AUC variation diagram on dataset 2,and green represets the AUC variation diagram on dataset 3. The variation trends of AUC value on the three datasets were almost identical. When the tradeoff parameter was increased from 2^{-20} to 2^{-10}, the AUC value remained almost constant. As the parameter increased from 2^{-10} to 2^{-5}, the AUC value decreased slightly. When the parameter increased from 2^{-5} to 2^{0}, the AUC decreased gradually. As the parameter increased from 2^{0} to 2^{10}, the AUC decreased significantly. Therefore, the parameters of three datasets were set as 2^{-10}.

C. COMPARISON OF PREDICTIVE PERFORMANCE UNDER DIFFERENT SITUATIONS

LRLSSP mainly includes the score estimation and refinement of lncRNA–disease associations. This method first uses the LRLS algorithm to estimate the lncRNA–disease association score and uses matrix space projection to refine the lncRNA–disease association score. The prediction performance of LRLSSP was evaluated considering the following scenarios: (1) predictive performance using only LRLS(LRLS), (2) predictive performance using only matrix space projection (SP), (3) lncRNA–disease association score estimation via space projection and refinement via LRLS (SPLRLS), and (4) estimation of the lncRNA–disease association score via LRLS and refinement via matrix space projection (LRLSSP). Figure 3 plots the ROC curve of the four scenarios mentioned in dataset 1. The ROC curve of LRLSSP was superior to those of the other scenarios. The AUC of LRLSSP was 0.9377, whereas those of LRLS, SP, and SPLRLS were 0.7299, 0.7342, and 0.7343, respectively. LOOCV in dataset 2 further confirmed that LRLSSP had the best prediction effect with an AUC value of 0.9383, whereas the AUC values of LRLS, SP, and SPLRLS were 0.8037, 0.8034, and 0.8038, respectively. Figure 4 shows the ROC curves of the four scenarios in dataset 2. Figure 5 shows the results of LOOCV of four different scenarios on dataset 3. The AUC value of LRLSSP is 0.9467, while that of LRLS, SP, and SPLRLS are 0.8813, 0.8840, and 0.8846, respectively. The results show that the prediction performance of LRLSSP is much better than that of LRLS, SP, and SPLRLS.

D. COMPARISON OF LRLSSP WITH EXISTING STATE-OF-THE-ART METHODS

Thus far, state-of-the-art models for lncRNA–disease association include NCPLDA [32], IIRWR [24], and

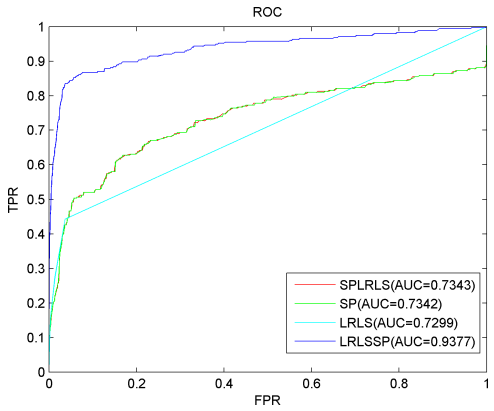

FIGURE 3. ROC curves and AUC values based on LOOCV in different situations in the dataset 1.

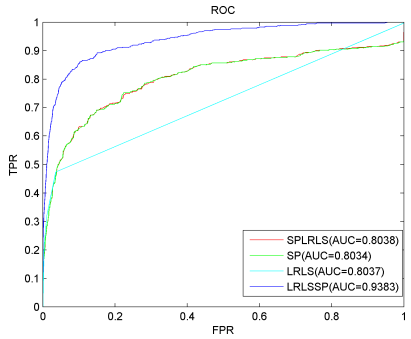

FIGURE 4. ROC curves and AUC values based on LOOCV in different situations in the dataset 2.

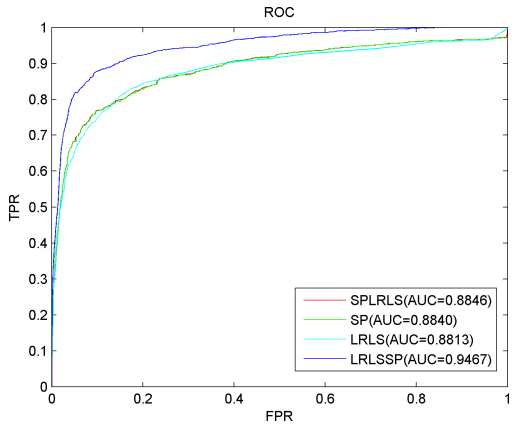

FIGURE 5. ROC curves and AUC values based on LOOCV in different situations in the dataset 3.

LDAI-ISPS [33]. The performance of LRLSSP was assessed by comparing it with three methods in the framework of LOOCV. A fair comparison was ensured by setting the optimal parameters for NCPLDA, IIRWR, and LDAI-ISPS as described in the corresponding literature. Figure 6 shows the ROC curve and AUC values of LRLSSP, NCPLDA, and LDAI-ISPS on dataset 1. The AUC values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS were 0.9377, 0.9107, 0.7883,and 0.9154, respectively. LRLSSP showed

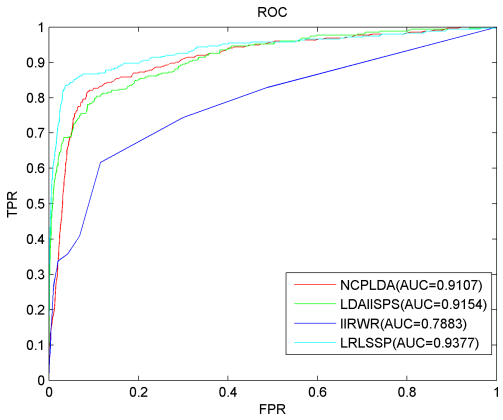

FIGURE 6. ROC curves and AUC values of LRLSSP and other three methods in the dataset 1.

the best performance among the tested methods, and its AUC values were 2.96%,16.47%, and 2.44% higher than those of the other methods. The AUC values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS in dataset 2 were 0.9383, 0.9012, 0.8230,and 0.8341, respectively. The AUC values obtained for LRLSSP were 3.95%,12.29%,and 12.49% higher than those of NCPLDA and LDAI-ISPS, respectively (Figure 7).AUC values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS in dataset 3 were 0.9467, 0.9307, 0.8745, 0.8455 (Figure 8), respectively. AUC values of LRLSSP were 1.69%, 7.63%, and 10.69% higher than those of NCPLDA, IIRWR, and LDAI-ISPS.

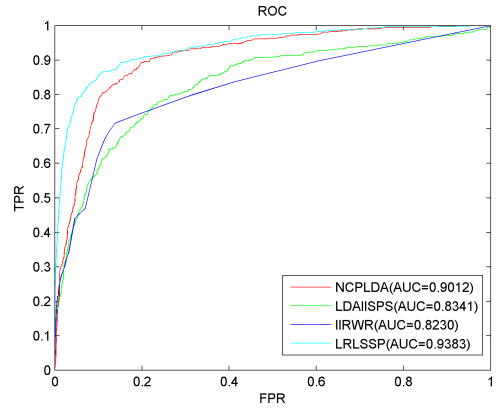

FIGURE 7. ROC curves and AUC values of LRLSSP and other three methods in the dataset 2.

AUPR was also used to evaluate the prediction performance of these methods. In dataset 1, the AUPR values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS were 0.5642, 0.1886, 0.1319, and 0.4516 (Figure 9), respectively. In dataset 2, the AUPR values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS were 0.3992, 0.1378, 0.1724, and 0.1768 (Figure 10), respectively. In dataset 3, the AUPR values of LRLSSP, NCPLDA, IIRWR, and LDAI-ISPS were 0.3175, 0.1355, 0.1163, and 0.0820 (Figure 11), respectively. The above two indexes of LRLSSP were superior to NCPLDA,

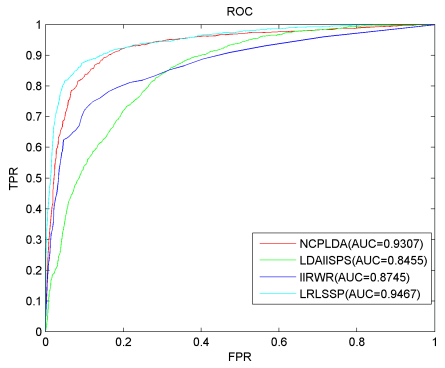

FIGURE 8. ROC curves and AUC values of LRLSSP and other three methods in the dataset 3.

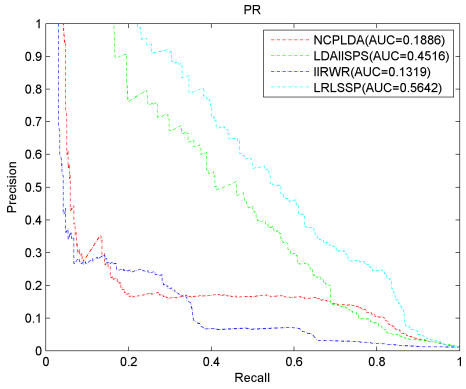

FIGURE 9. PR curves and AUPR values of LRLSSP and other three methods in the dataset 1.

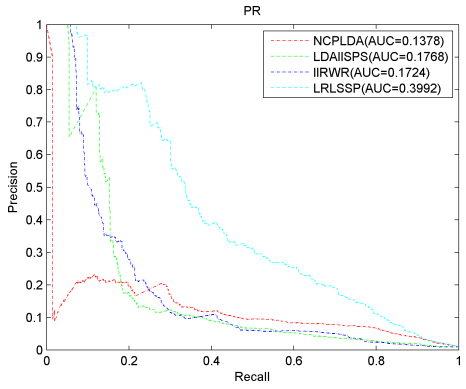

FIGURE 10. PR curves and AUPR values of LRLSSP and other three methods in the dataset 2.

IIRWR,and LDAI-ISPS in three different data sets, which fully showed that LRLSSP was a reliable lncRNA-disease association prediction tool.

E. PREDICTION OF ISOLATED DISEASES AND NEW lncRNAs

Prediction experiments for diseases without any known associated lncRNAs(isolated diseases)were performed to further assess the capability of LRLSSP to discover potential lncRNAs for isolated diseases. In the experiments, one

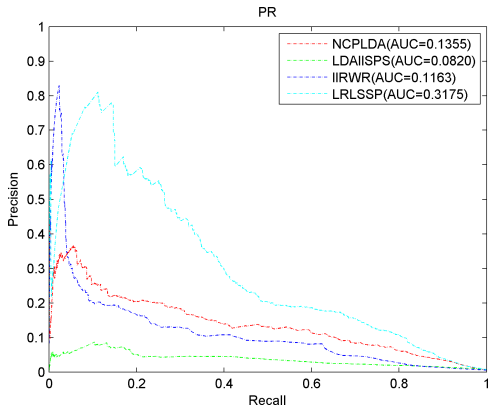

FIGURE 11. PR curves and AUPR values of LRLSSP and other three methods in the dataset 3.

disease was selected as the test sample. The rest of the associations related to other diseases were used as the training sample until all the diseases were verified as test sample once. For each disease, all related lncRNAs were removed to simulate isolated diseases. The cross-validation results in Figure 12 demonstrate that LRLSSP has a good prediction performance for isolated diseases, and the AUC values in dataset 1, dataset 2, and dataset 3 were 0.8732, 0.8833, and 0.8909,respectively.

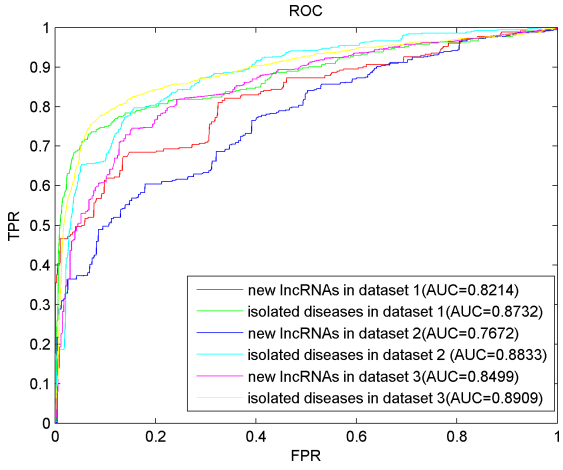

FIGURE 12. Predictions of new lncRNAs and isolated diseases.

Similarly, a lncRNA was selected as the test sample to verify the predictive capability of LRLSSP for new lncRNA-related diseases. Other existing associations were considered the training sample until all lncRNAs were verified once as test samples. As in the simulation of isolated diseases, all associated diseases for each lncRNA were deleted to simulate new lncRNAs, and all candidate diseases were prioritized using information from other lncRNA-related diseases. As shown in Figure 12, for the cross-validation of new lncRNAs, LRLSSP obtained an AUC value of 0.8214,0.7672,and 0.8499 in dataset 1, dataset 2, and dataset 3,respectively. Thus, the generalization capability

of LRLSSP in predicting lncRNAs without any known associated diseases was verified.

F. CASE STUDIES

Two specific diseases, namely, melanoma and cervical cancer, were selected for a case study to further estimate the performance of LRLSSP in predicting potential disease-related lncRNAs. During simulation, first took the known lncRNA–disease associations in dataset 1 as training samples, then took the rest unknown associations in dataset 1 as test verification set, and finally verified the prediction results in LncRNADisease v2.0 and relevant literature. Table 1 lists the top five melanoma- and cervical cancer-related lncRNAs predicted using the LRLSSP model and the corresponding validation evidence.

Melanoma, an aggressive skin cancer, is the seventh most malignant tumor in females and the fifth in males [59] [60]. Over the past few decades, the incidence and mortality of melanoma steadily increased [61]. Here, LRLSSP was implemented to predict lncRNAs associated with melanoma and to identify novel molecular associations as prognostic or therapeutic markers. Table 1 shows that four of the top five predicted lncRNAs related to melanoma were confirmed in the LncRNADisease database, and only BCYRN1 was not verified. Furthermore, Li *et al.* [61] observed that MEG3 inhibits the progression of malignant melanoma through the inactivated Wnt signaling pathway; Wang *et al.* [62] suggested that PVT1 promotes melanoma progression; Chen *et al.* [63] revealed that RNA PVT1 can be used as a diagnostic biomarker for melanoma. These pieces of evidence further prove the reliable performance of LRLSSP in the prediction of potential melanoma-related lncRNA. Although no evidence can prove that BCYRN1 is associated with melanoma, researchers may find evidence of its role in the future.

Cervical cancer is one of the gynecological tumors with the highest cancer-related deaths worldwide [64]. The potential therapeutic targets of cervical cancer revealed by the analysis of lncRNA association will be helpful for its treatment. Therefore, early detection of cervical cancer is necessary, and a better understanding of the molecular mechanisms of cervical cancer growth and metastasis will help in the discovery of novel specific biomarkers. Here, LRLSSP was implemented to discover lncRNAs that may be related to cervical cancer. As shown in Table 1, all the predicted results of the top five cervical cancer-related lncRNAs were confirmed in the LncRNADisease v2.0. These cases suggest the good performance of LRLSSP in predicting new potential lncRNA–disease associations.

In recent years, many new diseases with no known lncRNA association information have been gradually discovered. Many models are inapplicable to predict the lncRNAs associated with isolated diseases. The performance of LRLSSP was further evaluated by using it to predict lncRNAs associated with isolated diseases. Here, ovarian cancer and breast cancer were selected for case analysis.

In order to simulate isolated diseases, we removed the association between the disease and all lncRNA when implementing LRLSSP.With ovarian cancer as an example, 5 known association information of lncRNAs related to ovarian cancer were observed in dataset 1 and deleted. This procedure ensured that no known lncRNAs were associated with ovarian cancer, and only other available information was used to predict lncRNAs associated with it.

Tables 2 shows that the top five predicted ovarian cancer-related lncRNAs are confirmed by the LncRNADisease v2.0. Furthermore, manual literature search revealed that MEG3 can affect the proliferation, invasion, and migration of ovarian cancer cells under the regulation of lncRNA PTEN [65].

Similarly, all 14 known lncRNAs associated with breast cancer were detected in the present work from dataset 1. LRLSSP was implemented to predict breastcancer-related lncRNAs, and the predicted results were ranked. The top five predicted results are listed in Table 2, and the evidence of association between all lncRNAs and breast cancer was found in LncRNADisease v2.0.

The above experimental results further prove the reliable performance of LRLSSP in the prediction of isolated disease-related lncRNAs. The findings also address the problem with many lncRNA–disease association prediction models that cannot be applied to the prediction of isolated disease-related lncRNAs.

V. DISCUSSION

LncRNA variation and disorders can lead to many diseases, and LncRNAs associated with diseases caught the attention of many researchers. The identification and prediction of the relationship between lncRNA and diseases are helpful in understanding lncRNA function and pathogenesis. However, the existing biological experimental methods for identifying the association between lncRNA and diseases are expensive and time consuming. Thus, computational models can be used as effective complement to biological experimental verification. Researchers have developed various computational models to predict lncRNA-related diseases.

This work proposed a LRLSSP model based on LRLS algorithm and matrix space projection to predict potential lncRNA–disease associations. The LRLSSP shows excellent performance in predicting potentially unknown lncRNA–disease interactions and effectively predicts isolated diseases and new lncRNAs.

We first evaluated the performance of the model itself to fairly evaluate the performance of LRLSSP model. Given that LRLSSP integrates the LRLS and matrix SP algorithms, four conditions were considered in evaluating the performance of the model: (1) predictive performance using LRLS only(LRLS), (2) predictive performance using the matrix SP only(SP), (3) score estimation of lncRNA–disease association via SPLRLS, and (4) score estimation of lncRNA–disease association via LRLSSP. The AUC values of LRLSSP, LRLS, SP, and SPLRLS obtained by LOOCV

TABLE 1. The top 5 potential melanoma and cervical cancer-related lncRNAs predicted by LRLSSP based on dataset 1.

Disease	lncRNA name	Evidences	RANK
melanoma	MALAT1	LncRNADisease	1
melanoma	MEG3	LncRNADisease	2
melanoma	PVT1	LncRNADisease	3
melanoma	BCYRN1	Unconfirmed	4
melanoma	HOTAIR	LncRNADisease	5
cervical cancer	MEG3	LncRNADisease	1
cervical cancer	PVT1	LncRNADisease	2
cervical cancer	CDKN2B-AS1	LncRNADisease	3
cervical cancer	HOTAIR	LncRNADisease	4
cervical cancer	LSINCT5	LncRNADisease	5

TABLE 2. The top 5 lncRNAs associated with ovarian cancer and BREAST cancer were predicted by LRLSSP with hiding all known related lncRNAs based on dataset 1.

Disease	lncRNA name	Evidences	RANK
ovarian cancer	H19	LncRNADisease	1
ovarian cancer	DNM3OS	LncRNADisease	2
ovarian cancer	MALAT1	LncRNADisease	3
ovarian cancer	MEG3	LncRNADisease	4
ovarian cancer	PVT1	LncRNADisease	5
breast cancer	H19	LncRNADisease	1
breast cancer	BCAR4	LncRNADisease	2
breast cancer	MIR31HG	LncRNADisease	3
breast cancer	PINC	LncRNADisease	4
breast cancer	MALAT1	LncRNADisease	5

on dataset 1 were 0.9377, 0.7299, 0.7342, and 0.7343, respectively. The AUC values of LRLSSP, LRLS, SP, and SPLRLS obtained by LOOCV on dataset 2 were 0.9383, 0.8037, 0.8034, and 0.8038, respectively.The AUC values of LRLSSP, LRLS, SP, and SPLRLS obtained by LOOCV on dataset 3 were 0.9467, 0.7448, 0.8840, and 0.8846, respectively.

The performance of LRLSSP was compared with those of two advanced models (NCPLDA, IIRWR, and LDAI-ISPS). The AUC values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS obtained by LOOCV on dataset 1 were 0.9377, 0.9107, 0.7883,and 0.9154, respectively. Those from dataset 2 were 0.9383 (LRLSSP), 0.9012 (NCPLDA), 0.8230(IIRWR),and 0.8341 (LDAI-ISPS). AUC values of LRLSSP, NCPLDA, IIRWR,and LDAI-ISPS in dataset 3 were 0.9467, 0.9307, 0.8745, 0.8455, respectively.The prediction results of LRLSSP are better than those of the other methods.

Each disease (lncRNA) was simulated as an isolated disease (a new lncRNA)and cross-validated to assess the

prediction performance of LRLSSP for new lncRNAs and isolated diseases. The cross-validation results for isolated diseases in dataset 1, dataset 2, and dataset 3 were 0.8732 and 0.8833 and 0.8909, respectively, and the results for new lncRNAs in dataset 1,dataset 2,and dataset were 3 0.8214, 0.7672, and 0.8499, respectively. In addition, two specific diseases, melanoma and cervical cancer, were selected as cases to evaluate the predictive power of LRLSSP and to further verify its reliability in predicting the relationship between a potential lncRNA and a disease. The performance of LRLSSP was validated by using it to predict lncRNAs associated with isolated diseases. In the follow-up experiment, ovarian cancer and breast cancer were selected as cases. The case study further validated LRLSSP's excellent predictive performance.

The high prediction performance of LRLSSP is mainly due to the following aspects. First, a reasonable network of disease (lncRNA) similarity was constructed. The kernel similarity of the Gaussian interaction profile was used to address the deficiency of disease semantic similarity (lncRNA functional similarity), which effectively improved the accuracy of the similarity between diseases (lncRNAs). Second, LRLSSP integrates the LRLS algorithm and matrix space projection to construct a reliable prediction framework for lncRNA–disease association. The LRLS algorithm was first used to estimate the lncRNA–disease association to alleviate the problems of the sparsity of known lncRNA–disease association. Matrix space projection was used to obtain accurate association scores between lncRNA–disease associations. Matrix space projection also provided unprecedented and comprehensive information on disease space, lncRNA space, and known lncRNA–disease associations. Thus, the potential lncRNA–disease association could be predicted, and disease-related lncRNAs could be isolated. LRLSSP can effectively predict the lncRNA–disease association but with limitations. First, the construction of disease and lncRNA similarity networks urgently needs to integrate considerable omics data. Second, the present prediction method is based on known lncRNA–disease interactions, which bias the disease with known associated lncRNAs.

VI. CONCLUSIONS

Identifying lncRNA–disease associations aids in understanding the molecular mechanisms of a disease. In this work, a novel computational method, called LRLSSP, was proposed on the basis of LRLS algorithm to predict potential lncRNA–disease interactions. LRLSSP is a global prediction method that can simultaneously predict the association between diseases and lncRNAs and can be implemented in the prediction of isolated diseases and new lncRNAs. Its advantages include the lack of negative sample and having one parameter requirement. Compared with other state-of-the-art methods, LRLSSP has higher prediction accuracy on potential lncRNA–disease association. This work demonstrated the reliable performance of LRLSSP in the prediction of lncRNA and disease association. In conclusion, LRLSSP is a promising method for predicting lncRNA–disease associations and may promote related research on complex human diseases.

ACKNOWLEDGMENT

(Min Chen and Yan Peng are co-first authors.)

REFERENCES

[1] X. Song, G. Cao, L. Jing, S. Lin, X. Wang, J. Zhang, M. Wang, W. Liu, and C. Lv, "Analysing the relationship between lnc RNA and protein-coding gene and the role of lnc RNA as ce RNA in pulmonary fibrosis," *J. Cellular Mol. Med.*, vol. 18, no. 6, pp. 991–1003, 2014.

[2] K. C. Wang and H. Y. Chang, "Molecular mechanisms of long noncoding RNAs," *Mol. Cell*, vol. 43, no. 6, pp. 904–914, Sep. 2011.

[3] O. Wapinski and H. Y. Chang, "Corrigendum: Long noncoding RNAs and human disease:[trends in cell biology 21 (2011), 354–361]," *Trends Cell Biol.*, vol. 21, no. 10, p. 561, 2011.

[4] Y.-Z. Sun, D.-H. Zhang, Z. Ming, and J.-Q. Li, "DLREFD: A database providing associations of long non-coding RNAs, environmental factors and phenotypes," *Database, J. Biol. Databases Curation*, vol. 2017, Oct. 2017, Art. no. bax084, doi: 10.1093/database/bax084.

[5] S. U. Schmitz, P. Grote, and B. G. Herrmann, "Mechanisms of long noncoding RNA function in development and disease," *Cellular Mol. Life Sci.*, vol. 73, no. 13, pp. 2491–2509, Jul. 2016.

[6] J.-J. Qiu, Y. Wang, J.-X. Ding, H.-Y. Jin, G. Yang, and K.-Q. Hua, "The long non-coding RNA HOTAIR promotes the proliferation of serous ovarian cancer cells through the regulation of cell cycle arrest and apoptosis," *Experim. Cell Res.*, vol. 333, no. 2, pp. 238–248, May 2015.

[7] S. Chen, L.-L. Wang, K.-X. Sun, Y. Liu, X. Guan, Z.-H. Zong, and Y. Zhao, "LncRNA TDRG1 enhances tumorigenicity in endometrial carcinoma by binding and targeting VEGF–A protein," *Biochim. et Biophys. Acta (BBA) Mol. Basis Disease*, vol. 1864, no. 9, pp. 3013–3021, Sep. 2018.

[8] M. Terashima, S. Tange, A. Ishimura, and T. Suzuki, "MEG3 long noncoding RNA contributes to the epigenetic regulation of epithelial-mesenchymal transition in lung cancer cell lines," *J. Biol. Chem.*, vol. 292, no. 1, pp. 82–99, Jan. 2017.

[9] M. Wang, C. Guo, L. Wang, G. Luo, C. Huang, Y. Li, D. Liu, F. Zeng, G. Jiang, and X. Xiao, "Long noncoding RNA GAS5 promotes bladder cancer cells apoptosis through inhibiting EZH2 transcription," *Cell Death Disease*, vol. 9, no. 2, pp. 1–16, Feb. 2018.

[10] X. Chen, "Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA," *Sci. Rep.*, vol. 5, no. 1, Oct. 2015, Art. no. 013186.

[11] X. Chen, Y.-Z. Sun, N.-N. Guan, J. Qu, Z.-A. Huang, Z.-X. Zhu, and J.-Q. Li, "Computational models for lncRNA function prediction and functional similarity calculation," *Briefings Funct. Genomics*, vol. 18, no. 1, pp. 58–82, Feb. 2019.

[12] X. Chen, C. C. Yan, X. Zhang, and Z.-H. You, "Long non-coding RNAs and complex diseases: From experimental results to computational models," *Briefings Bioinf.*, vol. 18, no. 4, pp. 558–576, 2016.

[13] W. Lan, L. Huang, D. Lai, and Q. Chen, "Identifying interactions between long noncoding RNAs and diseases based on computational methods," in *Computational Systems Biology, Methods and Protocols*, T. Huang, Ed. New York, NY: Springer, 2018, pp. 205–221.

[14] P. Ping, L. Wang, L. Kuang, S. Ye, M. F. B. Iqbal, and T. Pei, "A novel method for LncRNA-disease association prediction based on an lncRNA-disease association network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 2, pp. 688–693, Mar. 2019.

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

[16] Y. Liu, R. Zhang, F. Qiu, K. Li, Y. Zhou, D. Shang, and Y. Xu, "Construction of a lncRNA-PCG bipartite network and identification of cancer-related lncRNAs: A case study in prostate cancer," *Mol. BioSyst.*, vol. 11, no. 2, pp. 384–393, 2015.

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

[18] G. U. Ganegoda, M. Li, W. Wang, and Q. Feng, "Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations," *IEEE Trans. Nanobiosci.*, vol. 14, no. 2, pp. 175–183, Mar. 2015.
[19] G. Yu, G. Fu, C. Lu, Y. Ren, and J. Wang, "BRWLDA: Bi-random walks for predicting lncRNA-disease associations," *Oncotarget*, vol. 8, no. 36, 2017, Art. no. 060429.
[20] Y. Hu, M. Zhou, H. Shi, H. Ju, Q. Jiang, and L. Cheng, "Measuring disease similarity and predicting disease-related ncRNAs by a novel method," *BMC Med. Genomics*, vol. 10, no. S5, p. 71, Dec. 2017.
[21] C. Xu, Y. Ping, H. Zhao, S. Ning, P. Xia, W. Wang, L. Wan, J. Li, L. Zhang, L. Yu, and Y. Xiao, "LncNetP, a systematical lncRNA prioritization approach based on ceRNA and disease phenotype association assumptions," *Oncotarget*, vol. 8, no. 70, 2017, Art. no. 114603.
[22] Q. Yao, L. Wu, J. Li, L. G. Yang, Y. Sun, Z. Li, S. He, F. Feng, H. Li, and Y. Li, "Global prioritizing disease candidate lncRNAs via a multi-level composite network," *Sci. Rep.*, vol. 7, no. 1, p. 39516, Apr. 2017.
[23] X. Chen, Z.-H. You, G.-Y. Yan, and D.-W. Gong, "IRWRLDA: Improved random walk with restart for lncRNA-disease association prediction," *Oncotarget*, vol. 7, no. 36, 2016, Art. no. 057919.
[24] L. Wang, Y. Xiao, J. Li, X. Feng, Q. Li, and J. Yang, "IIRWR: Internal inclined random walk with restart for LncRNA-disease association prediction," *IEEE Access*, vol. 7, pp. 54034–54041, 2019.
[25] L. Zhu, F. Su, Y. Xu, and Q. Zou, "Network-based method for mining novel HPV infection related genes using random walk with restart algorithm," *Biochim. et Biophys. Acta (BBA) Mol. Basis Disease*, vol. 1864, no. 6, pp. 2376–2383, Jun. 2018.
[26] Y. Liu, X. Zeng, Z. He, and Q. Zou, "Inferring MicroRNA-disease associations by random walk on a heterogeneous network with multiple data sources," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 4, pp. 905–915, Jul. 2017.
[27] J. Li, H. Zhao, Z. Xuan, J. Yu, X. Feng, B. Liao, and L. Wang, "A novel approach for potential human LncRNA-disease association prediction based on local random walk," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Aug. 14, 2019, doi: 10.1109/TCBB.2019.2934958.
[28] X.-N. Fan, S.-W. Zhang, S.-Y. Zhang, K. Zhu, and S. Lu, "Prediction of lncRNA-disease associations by integrating diverse heterogeneous information sources with RWR algorithm and positive pointwise mutual information," *BMC Bioinf.*, vol. 20, no. 1, p. 87, Dec. 2019.
[29] J. Hu, Y. Gao, J. Li, Y. Zheng, J. Wang, and X. Shang, "A novel algorithm based on bi-random walks to identify disease-related lncRNAs," *BMC Bioinf.*, vol. 20, no. S18, p. 569, Nov. 2019.
[30] Y. Liu, X. Feng, H. Zhao, Z. Xuan, and L. Wang, "A novel network-based computational model for prediction of potential LncRNA–disease association," *Int. J. Mol. Sci.*, vol. 20, no. 7, p. 1549, Mar. 2019.
[31] X. Chen, "KATZLDA: KATZ measure for the lncRNA-disease association prediction," *Sci. Rep.*, vol. 5, no. 1, Dec. 2015, Art. no. 016840.
[32] G. Li, J. Luo, C. Liang, Q. Xiao, P. Ding, and Y. Zhang, "Prediction of LncRNA-disease associations based on network consistency projection," *IEEE Access*, vol. 7, pp. 58849–58856, 2019.
[33] Y. Zhang, M. Chen, A. Li, X. Cheng, H. Jin, and Y. Liu, "LDAI-ISPS: LncRNA–disease associations inference based on integrated space projection scores," *Int. J. Mol. Sci.*, vol. 21, no. 4, p. 1508, Feb. 2020.
[34] G. Xie, Z. Huang, Z. Liu, Z. Lin, and L. Ma, "NCPHLDA: A novel method for human lncRNA-disease association prediction based on network consistency projection," *Mol. Omics*, vol. 15, no. 6, pp. 442–450, 2019.
[35] X. Xiao, W. Zhu, B. Liao, J. Xu, C. Gu, B. Ji, Y. Yao, L. Peng, and J. Yang, "BPLLDA: Predicting lncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network," *Frontiers Genet.*, vol. 9, p. 411, Oct. 2018.
[36] T. Zhao, J. Xu, L. Liu, J. Bai, C. Xu, Y. Xiao, X. Li, and L. Zhang, "Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features," *Mol. BioSyst.*, vol. 11, no. 1, pp. 126–136, 2015.
[37] J. Yu, P. Ping, L. Wang, L. Kuang, X. Li, and Z. Wu, "A novel probability model for LncRNA–disease association prediction based on the Naïve Bayesian classifier," *Genes*, vol. 9, no. 7, p. 345, Jul. 2018.
[38] W. Lan, M. Li, K. Zhao, J. Liu, F.-X. Wu, Y. Pan, and J. Wang, "LDAP: A Web server for lncRNA-disease association prediction," *Bioinformatics*, vol. 33, no. 3, pp. 458–460, Feb. 2017.
[39] Q. Chen, D. Lai, W. Lan, X. Wu, B. Chen, Y.-P.-P. Chen, and J. Wang, "ILDMSF: Inferring associations between long non-coding RNA and disease based on multi-similarity fusion," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Aug. 20, 2019, doi: 10.1109/TCBB.2019.2936476.
[40] X. Chen and G.-Y. Yan, "Novel human lncRNA–disease association inference based on lncRNA expression profiles," *Bioinformatics*, vol. 29, no. 20, pp. 2617–2624, Oct. 2013.
[41] X. Chen, C. Clarence Yan, C. Luo, W. Ji, Y. Zhang, and Q. Dai, "Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity," *Sci. Rep.*, vol. 5, no. 1, Sep. 2015, Art. no. 011338.
[42] Y.-A. Huang, X. Chen, Z.-H. You, D.-S. Huang, and K. C. Chan, "ILNCSIM: Improved lncRNA functional similarity calculation model," *Oncotarget*, vol. 7, no. 18, 2016, Art. no. 025902.
[43] X. Chen, Y.-A. Huang, X.-S. Wang, Z.-H. You, and K. C. Chan, "FML-NCSIM: Fuzzy measure-based lncRNA functional similarity calculation model," *Oncotarget*, vol. 7, no. 29, p. 45948, 2016.
[44] P. Xuan, N. Sheng, T. Zhang, Y. Liu, and Y. Guo, "CNNDLP: A method based on convolutional autoencoder and convolutional neural network with adjacent edge attention for predicting lncRNA–disease associations," *Int. J. Mol. Sci.*, vol. 20, no. 17, p. 4260, Aug. 2019.
[45] P. Xuan, L. Jia, T. Zhang, N. Sheng, X. Li, and J. Li, "LDAPred: A method based on information flow propagation and a convolutional neural network for the prediction of disease-associated lncRNAs," *Int. J. Mol. Sci.*, vol. 20, no. 18, p. 4458, Sep. 2019.
[46] P. Xuan, Y. Cao, T. Zhang, R. Kong, and Z. Zhang, "Dual convolutional neural networks with attention mechanisms based method for predicting disease-related lncRNA genes," *Frontiers Genet.*, vol. 10, p. 416, May 2019.
[47] W. Li, S. Wang, J. Xu, G. Mao, G. Tian, and J. Yang, "Inferring latent disease-lncRNA associations by faster matrix completion on a heterogeneous network," *Frontiers Genet.*, vol. 10, p. 769, Sep. 2019.
[48] C. Lu, M. Yang, F. Luo, F.-X. Wu, M. Li, Y. Pan, Y. Li, and J. Wang, "Prediction of lncRNA–disease associations based on inductive matrix completion," *Bioinformatics*, vol. 34, no. 19, pp. 3357–3364, Oct. 2018.
[49] G. Fu, J. Wang, C. Domeniconi, and G. Yu, "Matrix factorization-based data fusion for the prediction of lncRNA–disease associations," *Bioinformatics*, vol. 34, no. 9, pp. 1529–1537, May 2018.
[50] J.-X. Liu, Z. Cui, Y.-L. Gao, and X.-Z. Kong, "WGRCMF: A weighted graph regularized collaborative matrix factorization method for predicting novel LncRNA-disease associations," *IEEE J. Biomed. Health Informat.*, early access, Apr. 13, 2020, doi: 10.1109/JBHI.2020.2985703.
[51] Z. Xuan, J. Li, J. Yu, X. Feng, B. Zhao, and L. Wang, "A probabilistic matrix factorization method for identifying lncRNA-disease associations," *Genes*, vol. 10, no. 2, p. 126, Feb. 2019.
[52] W. Zhu, K. Huang, X. Xiao, B. Liao, Y. Yao, and F.-X. Wu, "ALSBMF: Predicting lncRNA-disease associations by alternating least squares based on matrix factorization," *IEEE Access*, vol. 8, pp. 26190–26198, 2020.
[53] M. Zeng, C. Lu, F. Zhang, Z. Lu, F.-X. Wu, Y. Li, and M. Li, "LncRNA–disease association prediction through combining linear and non-linear features with matrix factorization and deep learning techniques," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 577–582.
[54] G. Chen, Z. Wang, D. Wang, C. Qiu, M. Liu, X. Chen, Q. Zhang, G. Yan, and Q. Cui, "LncRNADisease: A database for long-non-coding RNA-associated diseases," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D983–D986, Nov. 2012.
[55] D. Wang, J. Wang, M. Lu, F. Song, and Q. Cui, "Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases," *Bioinformatics*, vol. 26, no. 13, pp. 1644–1650, Jul. 2010.
[56] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug-target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, Nov. 2011.
[57] Z. Xia, L.-Y. Wu, X. Zhou, and S. T. Wong, "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces," *BMC Syst. Biol.*, vol. 4, no. 2, p. S6, 2010.
[58] L. Jiang, Y. Ding, J. Tang, and F. Guo, "MDA-SKF: Similarity kernel fusion for accurately discovering miRNA-disease association," *Frontiers Genet.*, vol. 9, p. 618, Dec. 2018.
[59] S. C. Trotter, N. Sroa, R. R. Winkelmann, T. Olencki, and M. Bechtel, "A global review of melanoma follow-up guidelines," *J. Clin. Aesthetic Dermatology*, vol. 6, no. 9, p. 18, 2013.
[60] L. Wu, L. Zhu, Y. Li, Z. Zheng, X. Lin, and C. Yang, "LncRNA MEG3 promotes melanoma growth, metastasis and formation through modulating miR-21/E-cadherin axis," *Cancer Cell Int.*, vol. 20, no. 1, pp. 1–14, Dec. 2020.

[61] P. Li, Y. Gao, J. Li, Y. Zhou, J. Yuan, H. Guan, and P. Yao, "LncRNA MEG3 repressed malignant melanoma progression via inactivating wnt signaling pathway," *J. Cellular Biochem.*, vol. 119, no. 9, pp. 7498–7505, Sep. 2018.
[62] B.-J. Wang, H.-W. Ding, and G.-A. Ma, "Long noncoding RNA PVT1 promotes melanoma progression via endogenous sponging miR-26b," *Oncol. Res. Featuring Preclin. Clin. Cancer Therapeutics*, vol. 26, no. 5, pp. 675–681, Jun. 2018.
[63] X. Chen, G. Gao, S. Liu, L. Yu, D. Yan, X. Yao, W. Sun, D. Han, and H. Dong, "Long noncoding RNA PVT1 as a novel diagnostic biomarker and therapeutic target for melanoma," *BioMed Res. Int.*, vol. 2017, pp. 1–9, Feb. 2017.
[64] R. D. Burk, Z. Chen, C. Saller, K. Tarvin, A. L. Carvalho, C. Scapulatemponeto, H. C. Silveira, and J. H. Fregnani, "Integrated genomic and molecular characterization of cervical cancer," *Nature*, vol. 543, no. 7645, pp. 378–384, Mar. 2017.
[65] J. Wang, W. Xu, Y. He, Q. Xia, and S. Liu, "LncRNA MEG3 impacts proliferation, invasion, and migration of ovarian cancer cells through regulating PTEN," *Inflammation Res.*, vol. 67, nos. 11–12, pp. 927–936, Dec. 2018.

MIN CHEN received the Ph.D. degree in computer science and technology from Hunan University, Changsha, China. He is currently a Professor with the Hunan Institute of Technology, Hengyang, Hunan, China. His current research interests include bioinformatics and machine learning.

YAN PENG received the master's degree from Central South University, Changsha, China, in 2012. Her research interests include bioinformatics and hospitality management.

ANG LI received the master's degree in computer science from Central South University, Changsha, China. He is currently an Associate Professor with the Hunan Institute of Technology, Hengyang, Hunan, China. His research interests include data mining and computational biology.

YINWEI DENG received the master's degree in software engineering from the Central South University of Forestry and Technology, in 2017. Her current research interests include bioinformatics, predicting lncRNA, and disease association.

ZEJUN LI received the Ph.D. degree in computer science and technology from Hunan University, Changsha, China. He is currently a Professor with the Hunan Institute of Technology, Hengyang, Hunan, China. His current research interests include bioinformatics and data mining.