BMC Bioinformatics

RESEARCH Open Access

Prediction of lncRNA-disease association based on a Laplace normalized random walk with restart algorithm on heterogeneous networks

Liugen Wang[1], Min Shang[1], Qi Dai[2] and Ping-an He[1*]

*Correspondence: pinganhe@zstu.edu.cn
[1] School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China
Full list of author information is available at the end of the article

Abstract

Background: More and more evidence showed that long non-coding RNAs (lncRNAs) play important roles in the development and progression of human sophisticated diseases. Therefore, predicting human lncRNA-disease associations is a challenging and urgently task in bioinformatics to research of human sophisticated diseases.

Results: In the work, a global network-based computational framework called as LRWRHLDA were proposed which is a universal network-based method. Firstly, four isomorphic networks include lncRNA similarity network, disease similarity network, gene similarity network and miRNA similarity network were constructed. And then, six heterogeneous networks include known lncRNA-disease, lncRNA-gene, lncRNA-miRNA, disease-gene, disease-miRNA, and gene-miRNA associations network were applied to design a multi-layer network. Finally, the Laplace normalized random walk with restart algorithm in this global network is suggested to predict the relationship between lncRNAs and diseases.

Conclusions: The ten-fold cross validation is used to evaluate the performance of LRWRHLDA. As a result, LRWRHLDA achieves an AUC of 0.98402, which is higher than other compared methods. Furthermore, LRWRHLDA can predict isolated disease-related lnRNA (isolated lnRNA related disease). The results for colorectal cancer, lung adenocarcinoma, stomach cancer and breast cancer have been verified by other researches. The case studies indicated that our method is effective.

Keywords: lncRNA-disease associations, Similarity network, Heterogeneous network, LRWRHLDA, Ten-fold cross validation, AUC

Background

The disease is an abnormal life activity process that occurs due to the disorder of homeostasis after the body is damaged by the cause of the disease under certain conditions. Currently, many studies have confirmed that there is a complex cross-regulation relationship among diseases, genes, lncRNAs, and miRNAs [1–4].

Many researches have shown that although the proportion of encoded proteins in the human genome is less than 2%, under certain conditions, most of all nucleotides are detectably transcribed [5]. Among the various types of non-protein-coding transcripts, long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) has attracted more and more attention. Among them, lncRNAs are defined as non-coding RNA with a length greater than 200 nucleotides [6]; miRNAs are an RNA molecule with a length of about 19–25 nucleotides that exists widely in eukaryotes [7].

The lncRNAs play an important role in a variety of biological mechanisms, such as epigenetic regulation, chromatin remodeling, gene transcription, protein transport, cell transportation [8]. The function of lncRNAs can be divided into the following categories: Transcription interference; Inducing chromatin remodeling and nucleosome modification; Regulating alternative splicing mode; Generating endogenous siRNAs; Regulating protein activity; Structure or Tissue function; Change the location of protein; Precursor of small RNA [5, 9, 10], et al.

Many researchers found that the expression or functional abnormalities of lncRNAs are closely related to the occurrence of human diseases, including cancers and degenerative neurological diseases, which seriously endanger human health. For example: The lncRNA HOTAIR overexpression increases breast cancer cell proliferation [11, 12]. The lncRNA AFAP1-AS1 has abnormal expression in cholangiocarcinoma, gallbladdercancer, hepatocellular carcinoma, gastric cancer, colorectal cancer, esophageal cancer [13]. The lncRNA HOXA-AS2 may be a biomarker for the treatment of gastric cancer, et al. [14]. There is a close correlation between lncRNA PCGEM1 and osteoarthritis [15]. Therefore, lncRNAs can be used as an important biomarker for the diagnosis of diseases.

The identification of lncRNA-diseases association includes biological experimental verification methods and computational model predictions. For example, based on the biological experiments, Faghihi et al. [16] found that the expression of BACE1-AS can promote the rapid feed forward regulation of β-secretase in Alzheimer's disease. Applying the RT-PCR technology and Northern blot analysis, Hu et al. [17] confirmed and verified that H19 may become a new target for colon cancer anti-tumor therapy. The results of biological experimental are reliable, however, they are time-consuming and costly.

Recently, the computational model attracted more and more attention, in which various data resources can be integrated, to identify the lncRNA-disease association. For instance, based on a semi-supervised learning framework, the Laplacian regularized least squares for lncRNA-disease association calculation model (LRLSLDA) was suggested to predict potential disease-related lncRNA models [18]. Integrating genome, regulome and transcriptome data, the naive Bayesian classifier was proposed to identify cancer-related lncRNAs [19]. Similarly, based on disease-gene cluster association scores, a machine learning method was suggested to predict potential lncRNA-disease associations [20]. Combining the incremental principal component analysis (IPCA) and random forest (RF) algorithm, a machine learning model, called as IPCARF, was applied to predict the lncRNA-disease associations [21].

In the process of finding lncRNA-disease associations, the method of matrix factorization has also been widely used. For instance, the dual-network integrated logistic matrix factorization and Bayesian optimization model has been used for lncRNA-disease associations (DNILMF-LDA) [22]. In addition, the weighted graph regularized collaborative matrix factorization (WGRCMF), dual sparse collaborative matrix factorization (DSCMF) and the multi-label fusion collaborative matrix factorization (MLFCMF) were applied to construct model for prediction of lncRNA-disease associations [23–25].

Based on the hypothesis that lncRNAs with similar functions may be related to diseases with similar phenotypes, some researchers have proposed several calculation methods based on biological networks to predict disease-related lncRNAs.

In addition, integrating the lncRNA and the disease similarity network, and the lncRNA-disease association network. BPLLDA model based on paths of fixed lengths in a heterogeneous lncRNA-disease association network was proposed to predict lncRNA-disease associations [26]. Furthermore, some random walk models on these heterogeneous networks were suggested to predict the relationship between lncRNA and disease [27–29]. For example, Sun et al. [27] proposed the random walk with restart method on a lncRNA functional similarity network (RWRlncD). Gu et al. [28] proposed a global network-based random walk with restart algorithm on lncRNA seed nodes and disease seed nodes to predict the relationship between lncRNA and disease (GrWLDA). Based on the heterogeneous network through the lncRNA, disease, and gene similarity network, MHRWR model was proposed based on random walk with restart algorithm on the global network [29].

Following the random walk with restart model, in the paper, a new computational model based on Laplacian normalized random walk with restart algorithm in a heterogeneous network was proposed to predict the association between lncRNA and disease. Firstly, the disease semantic similarity (lncRNA function similarity, gene function similarity, miRNA function similarity) is calculated. And then, based on the association of lncRNA and disease (miRNA and gene), the Gaussian interaction profile kernel similarity of lncRNA and disease (miRNA and gene) are calculated. The lncRNA function similarity (disease semantic similarity, miRNA function similarity, gene function similarity) is integrated with the Gaussian interaction profile kernel similarity for lncRNAs (diseases, miRNAs, genes) to construct the isomorphic networks. Furthermore, the Laplace normalized random walk with restart algorithm on heterogeneous networks is developed to predict potential lncRNA-disease association. As a result, our method obtains reliable AUCs of 0.98402 in the ten-fold cross validation. The performance of our method is superior to other similar methods. Moreover, case studies on colorectal cancer, lung adenocarcinoma, stomach cancer and breast cancer also demonstrate the reliability of our model.

Methods

Experimental data sources

In the paper, the databases involved in lncRNA-disease associations mainly include LncRNADisease database [30, 31], EVLncRNAs database [32], Lnc2Cancer database [33], MNDR v3.1 database [34], et al. Similarly, the lncRNA-miRNA association comes from the integrated data of DIANA-LncBase database [35], LncAcTdb 2.0 database [36], MiRcode database [37], and StarBase database [38]. The lncRNA-gene association comes from the integrated data of LncRNADisease database [30, 31], LncAcTdb 2.0 database [36] and LncRNA2Target v2.0 database [39]. The miRNA-disease association comes from the integrated data of MNDR v3.1 database [34], HMDD database [40] and MiR2Disease database [41]. The miRNA-gene association comes from the data of MiRTarBase database [42]. The gene-disease association comes from the integrated data of DisGeNET database [43], CREEDS database [44], and DISEASES database [45].

Due to the different databases may have different names for the same biomolecule, so we need to perform data error correction and data cleaning on the data sets obtained from the database (mainly includes deleting duplicates, mistake, vacant data). In addition, the names of biomolecules of the same type from different databases are unified. In order to improve the comprehensiveness of the data and further improve the accuracy and scope of the prediction, the union of the related data of the above database was considered.

For lncRNA, the intersection of three database, lncRNA-disease, lncRNA-gene and lncRNA-miRNA association set obtained from all databases, were considered to construct the lncRNA similarity network. There are 814 lncRNA in the work (Fig. 1). Finally, 2476 miRNAs, 7986 genes, and 217 diseases were remained to research. At the same time, we also summarize some basic characteristics of the X–Y association dataset (e.g., the average degree) of the dataset in Table 1. And X and Y both stand for lncRNA, disease, gene, miRNA.

Calculate the similarity matrix

LncRNA functional similarity matrix

Similar to the method of Sun et al. [27], the functional similarity of two lncRNAs was computed as following:

Supposing lncRNA l_1 is associated with the disease group D_1 ($D_1 = \{d_{1i} | 1 \leq i \leq a\}$), and lncRNA l_2 is associated with the disease group D_2 ($D_2 = \{d_{2j} | 1 \leq j \leq b\}$), the similarity between disease d_{11} and a disease group D_2 is defined as follows:

$$S(d_{11}, D_2) = \max_{d_2 \in D_2} (Sim(d_{11}, d_2)), \tag{1}$$

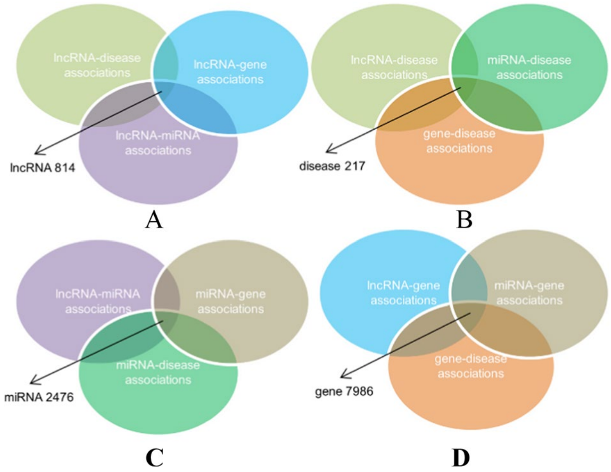
lncRNA-disease associations
lncRNA-gene associations
lncRNA-miRNA associations
lncRNA 814
A
lncRNA-disease associations
miRNA-disease associations
gene-disease associations
disease 217
B
lncRNA-miRNA associations
miRNA-gene associations
miRNA-disease associations
miRNA 2476
C
lncRNA-gene associations
miRNA-gene associations
gene-disease associations
gene 7986
D

Fig. 1 Ultimately retained the number of lncRNA, disease, miRNA, gene node. **A** lncRNA. **B** Disease. **C** miRNA. **D** Gene

Table 1 The basic characteristics of the X–Y association dataset

X	Y	Total	Total of associations	Average degree of X	Average degree of Y	Max degree of X	Max degree of Y
lncRNA		814					
	Disease	217	3434	4.2	15.8	90	418
	miRNA	2476	38,010	46.7	15.4	2284	98
	Gene	7986	14,987	18.4	1.9	2178	35
miRNA		2476					
	Disease	217	27,174	11.0	125.2	81	2453
	Gene	7986	216,934	87.6	27.2	1374	355
Gene		7986					
	Disease	217	37,277	4.7	171.8	74	6066

where $Sim(d_{11}, d_2)$ is the disease semantic similarity of diseases d_{11} and d_2. Then, the functional similarity between lncRNA l_1 and l_2 is defined as:

$$LS(l_1, l_2) = \frac{\sum\limits_{1 \le i \le a} S(d_{1i}, D_2) + \sum\limits_{1 \le j \le b} S(d_{2j}, D_1)}{a + b}. \tag{2}$$

Disease semantic similarity matrix

The Disease Ontology (DO) provides open-source ontology for the integration of biomedical data that is associated with human disease [46]. The terms in DO are diseases or ideas of disease-related that are organized in a directed acyclic graph (*DAG*). Applying the method of Wang et al. [47, 48], the semantic similarity of diseases is calculated as following:

Given disease d, its DAG graph can be expressed as $DAG(d) = (Ans(d), E(d))$, where $Ans(d)$ represents the set of the node, including node and its ancestor nodes, $E(d)$ represents the edge set of the corresponding direct link from the parent node d to the child node. That is the $E(d)$ denotes the relationship between different diseases. Based on DAG graph, the contribution of disease term d to the semantic value of disease T and the semantic value of disease T itself can be computed by the following two steps:

$$\begin{cases} D_T(d) = 1 & if\ d = T, \\ D_T(d) = \max\{\Delta * D_T(d') | d' \in chidren\ of\ d\} & if\ d \neq T, \end{cases} \tag{3}$$

$$DV(T) = \sum_{d \in Ans(d)} D_T(d), \tag{4}$$

where Δ is the semantic contribution attenuation factor and its value ranged from 0 to 1. As the direct distance between disease d and its ancestor diseases increases, the contribution of these ancestral diseases to the semantic value of disease d will gradually decrease. The semantic similarity between diseased d_1 and diseased d_2 is calculated by Eq. (5):

$$DS(d_1, d_2) = \frac{\sum\limits_{d \in (Ans(d_1) \cap Ans(d_2))} (D_{d_1}(d) + D_{d_2}(d))}{DV(d_1) + DV(d_2)}. \tag{5}$$

MiRNA functional similarity matrix

Similar to the Wang et al. [47] method, the functional similarity of two miRNAs can be defined as following:

Assuming that miRNA m_1 is associated with the disease group D_3 ($D_3 = \{d_{3k}|1 \leq k \leq c\}$) and miRNA m_2 is associated with the disease group D_4 ($D_4 = \{d_{4z}|1 \leq z \leq e\}$). The similarity of a disease d_{31} and a disease group D_4 is defined as follows:

$$S(d_{31}, D_4) = \max_{d_4 \in D_4} (Sim(d_{31}, d_4)), \tag{6}$$

and the functional similarity between miRNA m_1 and m_2 is computed by Eq. (7):

$$MS(m_1, m_2) = \frac{\sum\limits_{1 \leq k \leq c} S(d_{3k}, D_4) + \sum\limits_{1 \leq z \leq e} S(d_{4z}, D_3)}{c + e}. \tag{7}$$

Gene function similarity matrix

The Gene Ontology (*GO*) database is the world's largest informatics resource on the functions of genes [49]. For a *GO* node *A*, *DAG*=(*Ans*(*A*), *E* (*A*)) is its directed acyclic graph, where *Ans*(*A*) represents the set of all ancestors of node *A* (including node *A*); *E* (*A*) represents the set of edges connecting each node in *DAG*. For any *GO* node, assuming *t* is the ancestor of *A*, or *t*=*A*, $S_A(t)$ of *t*'s contribution to *A* is defined by Eq. (8):

$$\begin{cases} S_A(t) = 1 & if \ t = A, \\ S_A(t) = \max\{\Delta * S_A(t')|t' \in chidren \ of \ t\} & if \ t \neq A, \end{cases} \tag{8}$$

where Δ is the semantic contribution attenuation factor and its value ranged from 0 to 1. As the direct distance between gene *A* and its ancestor genes increases, the contribution of these ancestral genes to the semantic value of gene *A* will gradually decrease. The semantic contribution $S_V(A)$ of node *A* is defined as follows:

$$S_V(A) = \sum\nolimits_{t \in Ans(A)} S_A(t). \tag{9}$$

Then the semantic similarity of nodes *A* and *B* is calculated by Eq. (10):

$$S_{GO}(A, B) = \frac{\sum_{t \in (Ans(A) \cap t \in Ans(B))} (S_A(t) + S_B(t))}{S_V(A) + S_V(B)}. \tag{10}$$

The similarity of a go node *g* and a *GO* node set $G = \{\mathrm{go}_1, go_2, \ldots, go_f\}$ is defined as:

$$S(g, G) = \max_{1 \leq i \leq f} (S_{GO}(g, go_i)). \tag{11}$$

Assuming that the GO term set annotations of genes G_1 and G_2 are $GO_1 = \{\mathrm{go}_{11}, go_{12}, \ldots, go_{1m}\}$ and $GO_2 = \{\mathrm{go}_{21}, go_{22}, \ldots, go_{2n}\}$, respectively, the similarity of the two genes G_1 and G_2 is calculated by Eq. (12) [50]:

$$GS(G_1, G_2) = \frac{\sum\limits_{1 \le i \le m} S(go_{1i}, GO_2) + \sum\limits_{1 \le j \le n} S(go_{2j}, GO_1)}{m + n}. \tag{12}$$

Gaussian interaction profile kernel similarity for lncRNAs and diseases

Because there are many zeros in the matrix *LS*, *DS*, *MS* and *GS*, this will cause the sparsity of the matrix, which may lead to the inaccuracy of the prediction results. To avoid such scenario, we introduce the Gaussian interaction profile kernel similarity [51, 52].

Firstly, the $m \times n$ matrix *LD* represents the association matrix of lncRNA and disease, the elements are only 0 and 1. For example, if lncRNA l_i is related to disease d_j, *LD* (*i*, *j*) = 1, otherwise *LD* (*i*, *j*) = 0.

In the same way, we can define the lncRNA-miRNA association matrix *LM*, lncRNA-gene association matrix *LG*, disease-gene association matrix *DG*, miRNA-gene association matrix *MG*, miRNA-disease association matrix *MD*, respectively.

The Gaussian interaction profile kernel similarity of lncRNA l_i and l_j is defined as following:

$$GaL(l_i, l_j) = \exp(-r_l||IP(l_i) - IP(l_j)||^2), \tag{13}$$

$$r_l = r_l'/(\frac{1}{m}\sum_{i=1}^{m}||IP(l_i)||). \tag{14}$$

where *IP* (l_i) is a binary vector, which represents the *i*th row of the lncRNA-disease association matrix *LD*, and *m* represents the number of lncRNAs. $r_l^{'}$ is a regulation parameter of the kernel bandwidth parameter of r_l. According to the previous research, it is set to 1.

Similarly, the Gaussian interaction profile kernel similarity of disease d_i and d_j is defined as:

$$GaD(d_i, d_j) = \exp(-r_d||IP(d_i) - IP(d_j)||^2), \tag{15}$$

$$r_d = r_d'/(\frac{1}{n}\sum_{i=1}^{n}||IP(d_i)||). \tag{16}$$

where *IP* (d_i) is a binary vector, which represents the *i*th column of the lncRNA-disease association matrix *LD* and *n* is the number of diseases. $r_d' = 1$, it is a regulation parameter of the kernel bandwidth parameter of r_d.

Gaussian interaction profile kernel similarity for MiRNAs and genes

The Gaussian interaction profile kernel similarity calculation method of miRNA and gene is similar to that of lncRNA and disease, but the correlation matrix *MG* is used here. Therefore, we similarly define as follows: *IP* (m_i)is a binary vector, which represents the *i*-th row of the matrix *MG* and *h* is the number of miRNAs. $r_m' = 1$, it is a regulation parameter of the kernel bandwidth parameter of r_m. *IP* (g_i) is a binary vector, which represents the *i*th column of the matrix *MG* and *k* is the number of genes. $r_g' = 1$, it is a regulation parameter of the kernel bandwidth parameter of r_g.

Integration of similarities between lncRNAs, miRNAs, genes, and diseases

We integrate the lncRNA functional similarity (disease semantic similarity, miRNA functional similarity, gene functional similarity) with the Gaussian interaction profile kernel similarity for lncRNAs (diseases, miRNAs, genes) as follows:

$$LL = \begin{cases} GaL(l_i, l_j) & if\ l_i\ or\ l_j \in NL, \\ LS(l_i, l_j) & else. \end{cases} \tag{17}$$

$$DD = \begin{cases} GaD(d_i, d_j) & if\ d_i\ or\ d_j \in ND, \\ DS(d_i, d_j) & else. \end{cases} \tag{18}$$

$$MM = \begin{cases} GaM(m_i, m_j) & if\ m_i\ or\ m_j \in NM, \\ MS(m_i, m_j) & else. \end{cases} \tag{19}$$

$$GG = \begin{cases} GaG(g_i, g_j) & if\ g_i\ or\ g_j \in NG, \\ GS(g_i, g_j) & else. \end{cases} \tag{20}$$

where *NL* is the set of lncRNAs with no functional similarity with any other lncRNAs, *ND* is the set of diseases with no sematic similarity with any other disease, *NM* is the set of miRNAs with no functional similarity with any other miRNAs, and *NG* is the set of genes with no functional similarity with any other genes. By definition, *LL*, *DD*, *MM* and *GG* are symmetric.

The heterogeneous network

Based on the novel lncRNA similarity matrix *LL*, diseases similarity matrix *DD,* miRNA similarity matrix *MM*, and gene similarity matrix *GG*, four isomorphic networks include lncRNA similarity network, disease similarity network gene similarity network and miRNA similarity network were constructed, as shown in Fig. 2. In addition, a heterogeneous network through these four similarity networks and their interrelation ships were built based on six association matrix *LD, LM, LG, MD, MG, DG,* as shown in Fig. 3.

The random walk with restart

Based on the heterogeneous network, the random walk with restart (RWR) on the heterogeneous network to predict lncRNA-disease association was defined as follows [53]:

$$P^{t+1} = (1-\lambda)WP^t + \lambda P^0, \tag{21}$$

where P^0 is the initial probability vector, P^t is the probability vector in which the *i*th element is the probability of detecting the random walk at node *i* at step *t.* λ is the restart probability, and its value ranged from 0 to 1. *W* is the probability transition matrix and W_{ij} denotes the transition probability from node *i* to *j*, when the L_1 norm of P^{t+1} and P^t is less than 10^{-6}, it can be considered that reaches a stable state, meanwhile, the stable probability P^{∞} can be obtained.

The probability transition matrix *W* is constructed in this paper as follows:

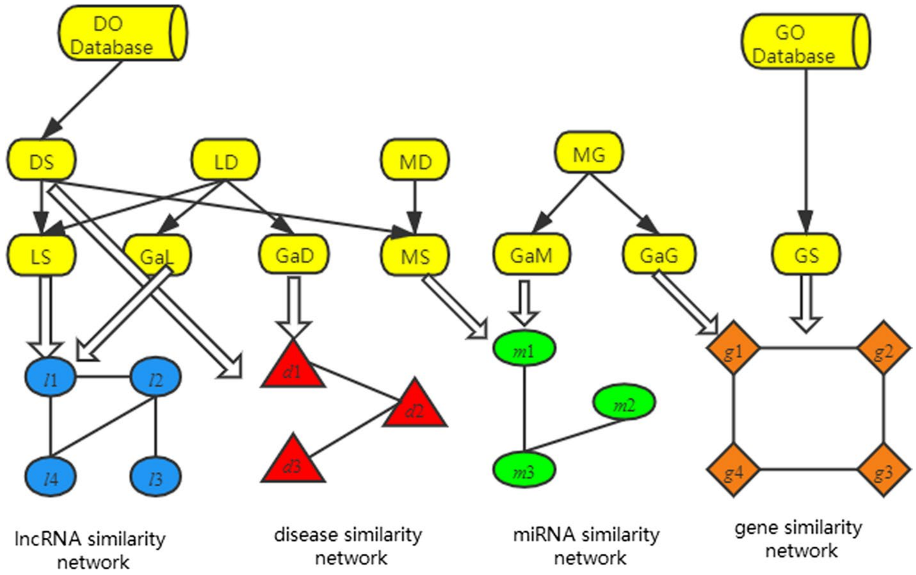

Fig. 2 Construction of similarity network of lncRNAs, diseases, miRNAs and genes

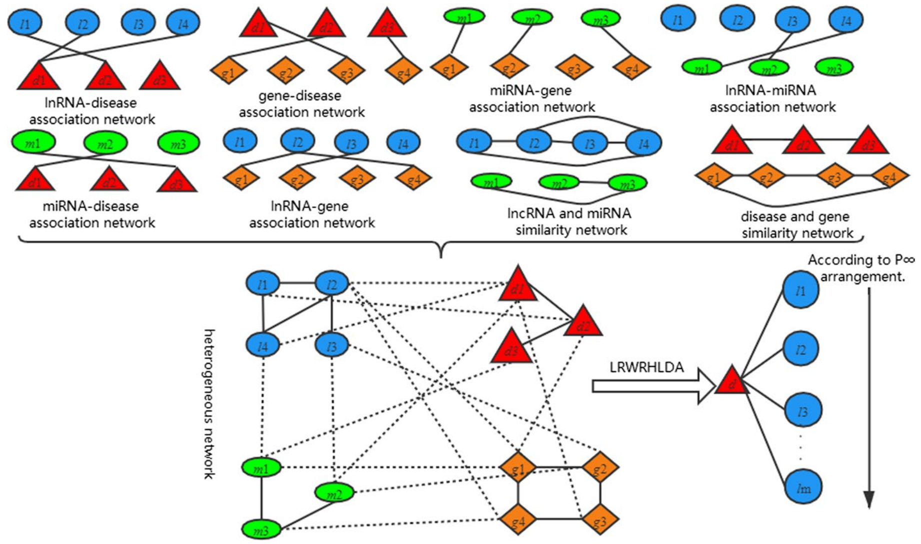
lnRNA-disease association network
gene-disease association network
miRNA-gene association network
lnRNA-miRNA association network
miRNA-disease association network
lnRNA-gene association network
lncRNA and miRNA similarity network
disease and gene similarity network
heterogeneous network
LRWRHLDA
According to P∞ arrangement.
l1
l2
l3
l4
lm
d1
d2
d3
m1
m2
m3
g1
g2
g3
g4

Fig. 3 Construction of heterogeneous network, and rank of lncRNA according to the stable probability of lncRNA by LRWRHLDA

$$W = \begin{pmatrix} W_{LL} & W_{LM} & W_{LG} & W_{LD} \\ W_{ML} & W_{MM} & W_{MG} & W_{MD} \\ W_{GL} & W_{GM} & W_{GG} & W_{GD} \\ W_{DL} & W_{DM} & W_{DG} & W_{DD} \end{pmatrix}. \tag{22}$$

Among them, the matrix W includes four intra-transition matrices and twelve inter-transition matrices. W_{LL} is the intra-transition matrix of lncRNA similarity network. W_{DD}, W_{MM} and W_{GG} are similar to W_{LL} and represent the intra-transition matrix of disease similarity network, miRNA similarity network, and gene similarity network, respectively. W_{LM} is defined as the transition matrix from lncRNA network to miRNA network. W_{LG}, W_{LD}, W_{ML}, W_{MG}, W_{MD}, W_{GL}, W_{GM}, W_{GD}, W_{DL}, W_{DM} and W_{DG} are defined similar to W_{LM}.

Laplacian normalization

Given the matrix $A = A\ (i, j)$, the diagonal matrix D is defined as follows, if $i = j$, then $D\ (i, j)$ is equal to the sum of the ith row of matrix A, otherwise $D\ (i, j) = 0$, then the Laplace normalization of matrix A is defined as [54, 55]:

$$\overrightarrow{A}\ (i,j) = \frac{A(i,j)}{\sqrt{D(i,i)D(j,j)}}. \tag{23}$$

Therefore, W_{LM} and W_{LL} can be obtained by the following two steps:

The probability of transition from l_i to m_j is as follows:

$$\overrightarrow{LM}(i,j) = \begin{cases} \dfrac{LM(i,j)}{\sqrt{\sum\limits_i LM(i,j) \sum\limits_j LM(i,j)}} & if\ \sum\limits_i LM(i,j) \sum\limits_j LM(i,j) \neq 0, \\ 0 & else. \end{cases} \tag{24}$$

$$W_{LM}(i,j) = \begin{cases} P_{LM} * \dfrac{\overrightarrow{LM}(i,j)}{\sum\limits_j \overrightarrow{LM}(i,j)} & if\ \sum\limits_j \overrightarrow{LM}(i,j) \neq 0, \\ 0 & else. \end{cases} \tag{25}$$

The probability of transition from l_i to l_j is as follows:

$$\overrightarrow{LL}(i,j) = \begin{cases} \dfrac{LL(i,j)}{\sqrt{\sum\limits_i LL(i,j) \sum\limits_j LL(i,j)}} & if\ \sum\limits_i LL(i,j) \sum\limits_j LL(i,j) \neq 0, \\ 0 & else. \end{cases} \tag{26}$$

$$W_{LL}(i,j) = \begin{cases} \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) = 0, \sum\limits_j \overrightarrow{LG}(i,j) = 0, \sum\limits_j \overrightarrow{LD}(i,j) = 0, \\ (1 - P_{LM}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) \neq 0, \sum\limits_j \overrightarrow{LG}(i,j) = 0, \sum\limits_j \overrightarrow{LD}(i,j) = 0, \\ (1 - P_{LG}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) = 0, \sum\limits_j \overrightarrow{LG}(i,j) \neq 0, \sum\limits_j \overrightarrow{LD}(i,j) = 0, \\ (1 - P_{LD}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) = 0, \sum\limits_j \overrightarrow{LG}(i,j) = 0, \sum\limits_j \overrightarrow{LD}(i,j) \neq 0, \\ (1 - P_{LM} - P_{LG}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) \neq 0, \sum\limits_j \overrightarrow{LG}(i,j) \neq 0, \sum\limits_j \overrightarrow{LD}(i,j) = 0, \\ (1 - P_{LM} - P_{LD}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) \neq 0, \sum\limits_j \overrightarrow{LG}(i,j) = 0, \sum\limits_j \overrightarrow{LD}(i,j) \neq 0, \\ (1 - P_{LG} - P_{LD}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) = 0, \sum\limits_j \overrightarrow{LG}(i,j) \neq 0, \sum\limits_j \overrightarrow{LD}(i,j) \neq 0, \\ (1 - P_{LM} - P_{LG} - P_{LD}) * \overrightarrow{LL}(i,j) \Big/ \sum\limits_j \overrightarrow{LL}(i,j) & if\ \sum\limits_j \overrightarrow{LM}(i,j) \neq 0, \sum\limits_j \overrightarrow{LG}(i,j) \neq 0, \sum\limits_j \overrightarrow{LD}(i,j) \neq 0. \end{cases} \tag{27}$$

where P_{LM} (P_{LG}, P_{LD}) is the parameter which represents the transition probability from lncRNA similarity network to miRNA (gene, disease) similarity network and its value ranged from 0 to 1. Besides, $P_{LM}{=}P_{ML}$, $P_{LG}{=}P_{GL}$, $P_{LD}{=}P_{DL}$, $P_{MG}{=}P_{GM}$, $P_{MD}{=}P_{DM}$, $P_{GD}{=}P_{DG}$. Similarly, other intra-transition matrix and inter-transition matrix can be defined.Applying the Laplacian normalization, all elements of probability transition matrix W can be obtained.The calculation formula of P^0 is as follows:

$$P^0 = \begin{pmatrix} P_L * U_{L0} \\ P_M * U_{M0} \\ P_G * U_{G0} \\ (1 - P_L - P_M - P_G) * U_{D0} \end{pmatrix}. \tag{28}$$

Among them, the parameters P_L, P_M, P_G, $1-P_L-P_M-P_G$ represent the importance of lncRNA similarity network, miRNA similarity network, gene similarity network and disease similarity network, respectively. Their values ranged from 0 to 1. U_{L0} represents the initial probability of the lncRNA similarity network, which is equal probabilities and is assigned to all seed nodes in the lncRNA similarity network. The sum of U_{L0} is 1. The initial probability U_{M0} and U_{G0} are similar to U_{L0}. U_{D0} represents the initial probability of the disease similarity network, for disease d, the initial transition probability of disease d is 1, and the transition probability of other diseases is 0.

Finally, the Laplace normalized random walk with restart algorithm is used to predict related lncRNAs scores (see Fig. 3). The method was called as LRWRHLDA (the Laplace normalized random walk with restart algorithm in heterogeneous networks to predict the lncRNA-disease association).

Results

Performance evaluation

In this paper, ten-fold cross validation is used to evaluate the performance of our model. In the ten-fold cross validation, all known lncRNA-disease interactions are randomly divided into ten folds. For each experiment, nine subsets are regarded as training samples and the remaining one subset is treated as test samples. After completing the test, predicted scores are generated. Then, we rank test samples and unknown lncRNA-disease interactions. The corresponding predicted result of test samples is considered as true positive (TP) when the predicted relevance score is greater than the threshold. Otherwise, considered as false negative (FN). Similarly, for the unknown lncRNA-disease interactions, the corresponding predicted result consider as false positive (FP) when the predicted relevance score is greater than the threshold. Otherwise, considered as true negative (TN). Then, the true positive rates (TPR), the false positive rates (FPR), recall and precision are calculated as follow:

$$TPR = recall = \frac{TP}{TP + FN}, \tag{29}$$

$$FPR = \frac{FP}{FP + TN}, \tag{30}$$

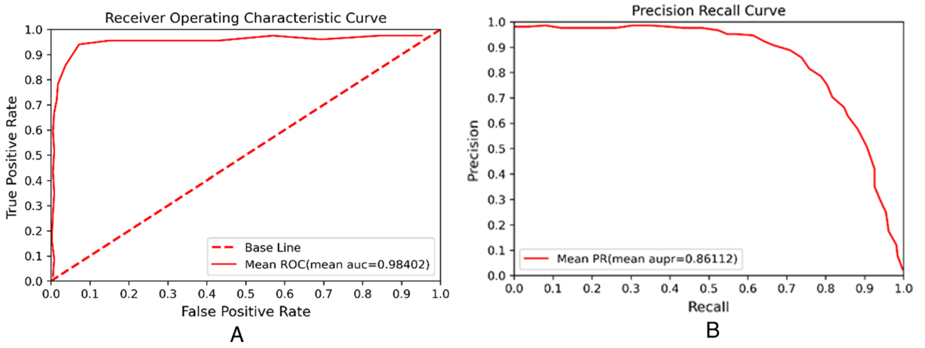

Fig. 4 The performance of LRWRHLDA by ten-fold cross validation. **A** The average ROC curve. **B** The average PR curve

Table 2 The AUC (AUPR) value for each experiment and mean AUC (AUPR) value

Test	AUC	AUPR
Test 1	0.98284	0.85944
Test 2	0.98326	0.86120
Test 3	0.98416	0.86039
Test 4	0.98531	0.85861
Test 5	0.98435	0.86283
Test 6	0.98309	0.86184
Test 7	0.98497	0.86178
Test 8	0.98337	0.86041
Test 9	0.98374	0.86227
Test 10	0.98510	0.86240
Mean	0.98402	0.86112

$$precision = \frac{TP}{TP + FP}. \tag{31}$$

Finally, the receiver operating characteristic (ROC) curve and precision-recall curve (PR) curve are drawn as shown in Fig. 4. The area under the ROC curve (AUC) and the area under the PR curve (AUPR) are used to evaluate the performance of our method. The range of AUC, AUPR are all from 0 to 1. When the parameters are set to $P_{LM}=P_{LG}=P_{LD}=P_{MG}=P_{MD}=P_{GD}=0.2$, $P_L=0.4$, $P_M=0.1$, $P_G=0.1$, $\lambda=0.7$, the results of ten experiments are shown in Table 2.

Comparison with different predicted methods using ten-fold cross validation

In order to compare with other models, the data in this paper is applied to the BPLLDA model [26], the RWRlncD model [27], GrwLDA model [28] and the MHRWR model [29].

As a result, the ROC curves under ten-fold cross validation of LRWRHLDA, RWRlncD, GrwLDA, BPLLDA and MHRWR were plotted in Fig. 5.

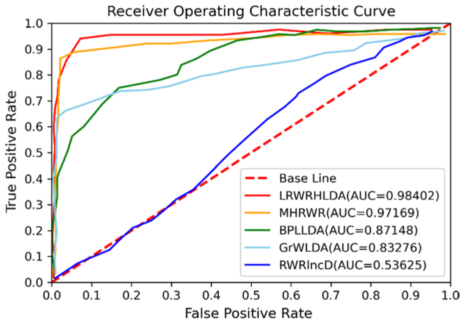

Fig. 5 The ROC curve and AUC of LRWRHLDA, RWRlncD, GrWLDA, BPLLDA and MHRWR in predicting lncRNA-disease associations by the ten-fold cross validation

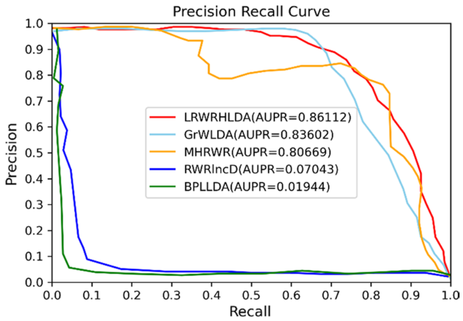

Fig. 6 The PR curve and AUPR of LRWRHLDA, RWRlncD, GrWLDA, BPLLDA and MHRWR in predicting lncRNA-disease associations by ten-fold cross validation

Table 3 The AUC and AUPR values when λ taking different values from 0.1 to 0.9, in which other parameters were fixed

λ	AUC	AUPR
0.1	0.95693	0.45169
0.2	0.96973	0.57582
0.3	0.97582	0.66525
0.4	0.97947	0.73139
0.5	0.98184	0.78217
0.6	0.98338	0.82477
0.7	0.98402	0.86112
0.8	0.98346	0.89130
0.9	0.98058	0.91417

P_{LM}, P_{LG}, P_{LD}, P_{MG}, P_{MD}, P_{GD} are all 0.2, $P_M = 0.1$, $P_G = 0.1$ and $P_L = 0.4$

As can be seen, LRWRHLDA has an AUC of 0.98402 and outperformed RWRlncD (0.53625), GrwLDA (0.83276), BPLLDA (0.87148) and MHRWR (0.97169). In summary, LRWRHLDA is better than other model in lncRNA-disease association prediction.

The area under PR curve (AUPR) is also used to evaluate the performance of LRWRHLDA model, BPLLDA model [26], the RWRlncD model [27], GrwLDA model [28] and MHRWR model [29] to avoid overestimates the performance of these methods (see Fig. 6).

It can be seen from Fig. 6 that the AUPR value of LRWRHLDA is also higher than other models.

Effects of parameters

There are ten parameters in our model, including the transition probability P_{LM}, P_{LG}, P_{LD}, P_{MG}, P_{MD}, P_{GD} between networks; the weight of the subnet P_L, P_M, P_G; and the restart probability λ. Due to too many parameters and our limited computing resources, we arbitrarily fixed nine of these parameters in the paper and only discussed the impact of restart probability λ with the ten-fold cross validation in our model. The results are shown in Table 3. As can be seen, based on the AUC index, the parameter λ has less influence on the performance of LRWRHLDA, when $\lambda = 0.7$. Based on the AUPR index, when λ is equal to 0.9, the AUPR value reaches the maximum. And observing Table 3, the results showed that the restart probability λ has powerful effects on our model.

Case study

Case studies on predicted lncRNA-disease associations

It is known that lncRNAs play critical roles in the development of many diseases. To evaluate the ability of LRWRHLDA in inferring potential lncRNA-disease associations, we use all known lncRNA-disease associations in LD as training data to assess the potential of predicted associations by our model.

The stable probability P^{∞} can be used as a measure of proximity to the seed lncRNAs. If P^{∞} (lncRNA i) > P^{∞} (lncRNA j), then lncRNA i will be in closer proximity to

Table 4 The predicted top10 potential lncRNAs for four cancers by LRWRHLDA

Rank	LncRNA	Evidence	LncRNA	Evidence
Colorectal cancer			Lung adenocarcinoma	
1	CASC19	MNDR v3.1	ZNF295-AS1	MNDR v3.1
2	ENST00000535511	PMID: 28177879	LINC01969	MNDR v3.1
3	RP4	PMID: 29531464	PRKCZ-AS1	MNDR v3.1
4	CTNNAP1	PMID: 27487124	PIK3CD-AS2	MNDR v3.1
5	LINC01021	PMID: 29262524	GMDS-AS1	PMID: 31860169
6	UCOO2KMD.1	MNDR v3.1	FAM83A-AS1	MNDR v3.1
7	UICLM	MNDR v3.1	ACTA2-AS1	MNDR v3.1
8	UCC	MNDR v3.1	LINC00635	MNDR v3.1
9	N-BLR	MNDR v3.1	LINC01207	PMID: 26693067
10	RP11-317J10.2	MNDR v3.1	LINC00941	MNDR v3.1
Stomach cancer			Breast cancer	
1	M59227	MNDR v3.1	LNC015192	MNDR v3.1
2	LOC150622	MNDR v3.1	LINC00993	MNDR v3.1
3	MUC2	MNDR v3.1	LINC00894-002	MNDR v3.1
4	AKR7L	MNDR v3.1	AC008268.1	MNDR v3.1
5	AC110615.1	MNDR v3.1	MIR2052HG	MNDR v3.1
6	AC079089.1	MNDR v3.1	ST8SIA6-AS1	MNDR v3.1
7	PWRN1	MNDR v3.1	PAX8-AS1-N	MNDR v3.1
8	SUCLG2-AS1	MNDR v3.1	PRLB	MNDR v3.1
9	AL162586.1	MNDR v3.1	PDCD4-AS1	MNDR v3.1
10	LINC01856	MNDR v3.1	ADARB2-AS1	MNDR v3.1

the seed lncRNAs than lncRNA *j* in the lncRNA similarity network. As a result, all candidate lncRNAs can be ranked according to the P^{∞}, and the top ranked lncRNAs can be expected to have a high probability of being associated with the disease of interest. The novel lncRNA-disease associations are ranked according to the stable probability of LRWRHLDA. To validate the predictions, we use literature or the following those databases: LncRNADisease [30], LncRNADisease v2.0 [31], MNDR v3.1 [34], lnCAR [56]. Specifically, we list the top 10 lncRNAs associated with four diseases, including colorectal cancer, lung adenocarcinoma, stomach cancer and breast cancer. According to P^{∞}, the top 10 results were shown in Table 4 (the detailed results see Additional file 1: Table-S1).

Colorectal cancer is the third most common cancer diagnosed in the US. While the incidence and the mortality rate of colorectal cancer has decreased due to effective cancer screening measures, there has been an increase in number of young patients diagnosed in colon cancer due to unclear reasons at this point of time [57]. Lung adenocarcinoma is one of the main types of lung cancer, which belongs to non-small cell carcinoma. The incidence of lung adenocarcinoma is mainly female and non-smokers [58]. Stomach cancer is the fifth most common cancer and the third most common cause of cancer death globally [59]. The most majority of stomach cancers are adenocarcinomas, with no obvious symptoms in the early stage. They are often similar to the symptoms of chronic gastric diseases such as gastritis and gastric ulcers, and easily ignore. Moreover, the current early diagnosis rate of stomach cancer is still low. Breast cancer is a malignant tumor that occurs in the epithelial tissue of the breast. At present, breast cancer has become a major public health problem in the current society, and its cause is not yet fully understood. In the world, breast cancer is an important cause of human suffering and premature mortality among women [60].

In Table 4, the six potential lncRNA-disease associations were confirmed in the literature except the existing lncRNA-disease associations in the database, in which included ENST00000535511-colorectal cancer, RP4-colorectal cancer, CTNNAP1-colorectal cancer, LINC01021-colorectal cancer, GMDS-AS1-lung adenocarcinoma, LINC01207-lung adenocarcinoma. These results demonstrated that the predictive performance of the proposed method.

Case studies on predicted novel diseases and novel lncRNAs

For each disease, it is deemed as a novel disease and all its related lncRNAs are removed to predict potential lncRNAs related the disease. All the candidate lncRNAs were ranked according to P^{∞} and lncRNAs with high scores were expected to be potentially related with investigated disease *d*. Depend on P^{∞}, the top 10 results were listed in Table 5 (the detailed results see Additional file 2: Table-S2).

Analogously, the stable probability P^{∞} can be also used as a measure of proximity to the seed diseases. All the candidate diseases were ranked according to P^{∞} and diseases with high scores were expected to be potentially related with investigated lncRNA. To evaluate the ability of our model to predict new lncRNAs, we analyzed two lncRNAs

Table 5 The predicted top 10 novel lncRNAs-related for four cancers by LRWRHLDA

Rank	LncRNA	Evidence	LncRNA	Evidence
Colorectal cancer			Lung adenocarcinoma	
1	CARL	Unconfirmed	FOXP4-AS1	lnCAR
2	CASC19	MNDR v3.1	NEXN-AS1	lnCAR
3	MCM3AP-AS1	PMID: 32982409	VPS9D1-AS1	lnCAR
4	AL358334.2	lnCAR	TERC	lnCAR
5	AL157400.4	lnCAR	AC018413.1	unconfirmed
6	LAMA5-AS1	lnCAR	AL157838.1	lnCAR
7	HNF1A-AS1	MNDR v3.1	TUBB2A	unconfirmed
8	RGMB-AS1	lnCAR	AC019197.1	lnCAR
9	C21ORF62-AS1	lnCAR	Z93930.2	lnCAR
10	CASC8	MNDR v3.1	SATB2-AS1	PMID: 34249715
Stomach cancer			Breast cancer	
1	SSBP3-AS1	lnCAR	LINC00652	lnCAR
2	AC103740.1	lnCAR	TAPT1-AS1	lnCAR
3	AF117829.1	Unconfirmed	AC007823.1	lnCAR
4	AC092910.3	lnCAR	LHX1-DT	PMID: 33194577
5	RAB30-AS1	lnCAR	KLF3-AS1	MNDR v3.1
6	AC093157.1	lnCAR	FGF14-AS2	MNDR v3.1
7	GATA2-AS1	lnCAR	KCNK15-AS1	MNDR v3.1
8	PCA3	lnCAR	AC107959.2	lncRNADisease v2.0 (predicted)
9	TERC	MNDR v3.1	AP003486.1	Unconfirmed
10	AC087164.1	lnCAR	LINC00993	MNDR v3.1

Table 6 The predicted top 10 novel diseases-related for H19 and HOTAIR by LRWRHLDA

H19			HOTAIR	
Rank	Disease	Evidence	Disease	Evidence
1	Carcinoma	lncRNADisease v2.0	Parkinson's disease	MNDR v3.1
2	Parkinson's disease	MNDR v3.1	Carcinoma	lncRNADisease v2.0
3	Colon cancer	MNDR v3.1	Colon cancer	MNDR v3.1
4	Stomach cancer	MNDR v3.1	Liver cancer	MNDR v3.1
5	Liver cancer	MNDR v3.1	Stomach cancer	MNDR v3.1
6	Pancreatic cancer	MNDR v3.1	Pancreatic cancer	MNDR v3.1
7	Kidney cancer	MNDR v3.1	Colorectal cancer	MNDR v3.1
8	Schizophrenia	lncRNADisease v2.0 (predicted)	Kidney cancer	MNDR v3.1
9	Colorectal cancer	MNDR v3.1	Colorectal carcinoma	MNDR v3.1
10	Glioblastoma	lncRNADisease	Melanoma	lncRNADisease

including H19 and HOTAIR. For each lncRNA, it is removed all its related diseases in predicting potential diseases. According to P^{∞}, the top 10 results were showed in Table 6 (the detailed results see Additional file 3: Table-S3).

Observing Table 5, we can find that thirty-five of the top ten lncRNAs associations with four cancers were validated by the database or literature. However, other five cancer-lncRNA associations, colorectal cancer-CARL, stomach cancer-AF117829.1, breast cancer-AP003486.1, lung adenocarcinoma-AC018413.1 and lung adenocarcinoma-TUBB2A have not been confirmed by the database or literature. It implies our method can predict more additional lncRNA-disease associations.

From Table 6, in both cases, all top ten associated diseases were validated by the database. In summary, LRWRHLDA achieves favorable performances in predicting novel disease-associated lncRNAs and novel lncRNA-associated diseases.

Discussion

At present, many studies have shown that lncRNA has an important influence on the physiological process of diseases. Because traditional biological experiments are time-consuming and costly, it is necessary to develop a computational model to predict the association between lncRNA and disease.

In this paper, a new model-LRWRHLDA based on the Laplace normalized random walk with restart algorithm in heterogeneous network was constructed to predict potential lncRNA-disease associations. The ten-fold cross validation test is applied to evaluate the prediction performance of our method. In comparison with the state-of-the-art prediction methods, our method can achieve better performance in terms of AUC values. Moreover, case studies of colorectal cancer, lung adenocarcinoma, stomach cancer and breast cancer are implemented to further demonstrate that it could be a useful method for predicting potential relationships between lncRNAs and diseases as well.

However, our method has some limitations. Firstly, since we have 10 parameters, the selection and adjustment of parameters still face some difficulties. Secondly, because of our model is based on four networks, there are too many nodes in the network. In the random walk process, the more nodes there are, the longer the random walk time will be. In the future, we will continue to improve the model.

Conclusion

In this study, we proposed an effective method, LRWRHLDA, which is based on the Laplace normalized random walk with restart algorithm in heterogeneous network to predict the potential lncRNA and disease association. First, a heterogeneous network based on lncRNA, disease, miRNA, gene similarity network and their correlation networks were constructed. Then, we calculate the probability transition matrix by Laplace normalization. Finally, the potential lncRNA-disease associations were predicted by the random walk with restart over heterogeneous networks. Furthermore, LRWRHLDA can predict isolated disease-related lnRNA (isolated lnRNA-related disease). Our method is evaluated comprehensively by ten-fold cross validation and case studies in comparison with other methods. The results show that our method has higher prediction accuracy.

Abbreviations

lncRNAs: Long non-coding RNAs; miRNA: MicroRNA; LRWRHLDA: Prediction the potential lncRNA-disease associations based on Laplace normalized random walk with restart algorithm in heterogeneous networks; ROC: Receiver operating characteristic; TPR: True positive rates; FPR: False positive rates; AUC: Areas under ROC curve; PR: Precision-recall; AUPR: The area under the precision-recall curve.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-021-04538-1.

Additional file 1. In this file we provide the results of stable probability of lncRNA when LRWRHLDA run over for four cancers based on the LD matrix.

Additional file 2. In this file we provide the results of stable probability of lncRNA when LRWRHLDA run over when delete related lncRNAs of the cancer.

Additional file 3. In this file we provide the results of stable probability of lncRNA when LRWRHLDA run over when delete related cancer of the lncRNAs.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61772027, 61772028), key research and development plan of Zhejiang Province (2021C02039).

Authors' contributions

LW, MS, QD and PH designed the study. LW and MS carried out analyses and wrote the program. LW and PH wrote the paper. All authors read and approved the final manuscript.

Funding

This research is partly sponsored by the National Natural Science Foundation of China (No. 61772027, 61772028), key research and development plan of Zhejiang Province (2021C02039). The funding bodies did not play any roles in the design of the study, in the collection, analysis, or interpretation of data, or in writing the manuscript.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. The code (executable code and source code) and data for this study are available at https://github.com/wang-124/LRWRHLDA.git.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China. [2]College of Life Science, Zhejiang Sci-Tech University, Hangzhou 310018, China.

Received: 8 October 2021 Accepted: 15 December 2021
Published online: 04 January 2022

References

1. Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nat Rev Genet. 2012;13(8):523–36.
2. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov. 2017;16(3):203–22.
3. Bhan A, Soleimani M, Mandal SS. Long noncoding RNA and cancer: a new paradigm. Cancer Res. 2017;77(15):3965–81.
4. Dai LY, Liu JX, Zhu R, Wang J, Yuan SS. Logistic weighted profile-based bi-random walk for exploring MiRNA-disease associations. J Comput Sci Technol. 2021;36(2):276–87.
5. Jarroux J, Morillon A, Pinskaya M. History, discovery, and classification of lncRNAs. Adv Exp Med Biol. 2017;1008:1–46.
6. Li J, Li Z, Zheng W, Li X, Wang Z, Cui Y, et al. LncRNA-ATB: an indispensable cancer-related long noncoding RNA. Cell Prolif. 2017;50(6):e12381.
7. Lu TX, Rothenberg ME. MicroRNA. J Allergy Clin Immunol. 2018;141(4):1202–7.
8. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol. 2013;14(11):699–712.
9. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. RNA Biol. 2013;10(6):925–33.
10. Li Z, Ho IHT, Li X, Xu D, Wu WKK, Chan MTV, et al. Long non-coding RNAs in the spinal cord injury: novel spotlight. J Cell Mol Med. 2019;23(8):4883–90.
11. Xue X, Yang YA, Zhang A, Fong KW, Kim J, Song B, et al. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. Oncogene. 2016;35(21):2746–55.
12. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071–6.

13. Ji D, Zhong X, Jiang X, Leng K, Xu Y, Li Z, et al. The role of long non-coding RNA AFAP1-AS1 in human malignant tumors. Pathol Res Pract. 2018;214(10):1524–31.
14. Wang J, Su Z, Lu S, Fu W, Liu Z, Jiang X, et al. LncRNA HOXA-AS2 and its molecular mechanisms in human cancer. Clin Chim Acta. 2018;485:229–33.
15. Zhao Y, Xu J. Synovial fluid-derived exosomal lncRNA PCGEM1 as biomarker for the different stages of osteoarthritis. Int Orthop. 2018;42(12):2865–72.
16. Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. Nat Med. 2008;14(7):723–30.
17. Hu Q, Wang YB, Zeng P, Yan GQ, Xin L, Hu XY. Expression of long non-coding RNA (lncRNA) H19 in immunodeficient mice induced with human colon cancer cells. Eur Rev Med Pharmacol Sci. 2016;20(23):4880–4.
18. Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics. 2013;29(20):2617–24.
19. Zhao T, Xu J, Liu L, Bai J, Xu C, Xiao Y, et al. Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features. Mol Biosyst. 2015;11(1):126–36.
20. Yuan Q, Guo X, Ren Y, Wen X, Gao L. Cluster correlation based method for lncRNA-disease association prediction. BMC Bioinform. 2020;21(1):180.
21. Zhu R, Wang Y, Liu JX, Dai LY. IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier. BMC Bioinform. 2021;22(1):175.
22. Li Y, Li J, Bian N. DNILMF-LDA: prediction of lncRNA-disease associations by dual-network integrated logistic matrix factorization and bayesian optimization. Genes (Basel). 2019;10(8):608.
23. Liu JX, Cui Z, Gao YL, Kong XZ. WGRCMF: a weighted graph regularized collaborative matrix factorization method for predicting novel LncRNA-disease associations. IEEE J Biomed Health Inform. 2021;25(1):257–65.
24. Liu JX, Gao MM, Cui Z, Gao YL, Li F. DSCMF: prediction of LncRNA-disease associations based on dual sparse collaborative matrix factorization. BMC Bioinform. 2021;22(Suppl 3):241.
25. Gao MM, Cui Z, Gao YL, Wang J, Liu JX. Multi-label fusion collaborative matrix factorization for predicting LncRNA-disease associations. IEEE J Biomed Health Inform. 2021;25(3):881–90.
26. Xiao X, Zhu W, Liao B, Xu J, Gu C, Ji B, et al. BPLLDA: predicting lncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network. Front Genet. 2018;9:411.
27. Sun J, Shi H, Wang Z, Zhang C, Liu L, Wang L, et al. Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. Mol Biosyst. 2014;10(8):2074–81.
28. Gu C, Liao B, Li X, Cai L, Li Z, Li K, et al. Global network random walk for predicting potential human lncRNA-disease associations. Sci Rep. 2017;7(1):12442.
29. Zhao X, Yang Y, Yin M. MHRWR: prediction of lncRNA-disease associations based on multiple heterogeneous networks. IEEE/ACM Trans Comput Biol Bioinform. 2020;PP.
30. Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, et al. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res. 2013;41(Database issue):D983–6.
31. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. Nucleic Acids Res. 2019;47(D1):D1034–7.
32. Zhou B, Ji B, Liu K, Hu G, Wang F, Chen Q, et al. EVLncRNAs 2.0: an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments. Nucleic Acids Res. 2021;49(D1):D86-91.
33. Gao Y, Shang S, Guo S, Li X, Zhou H, Liu H, et al. Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. Nucleic Acids Res. 2021;49(D1):D1251–8.
34. Ning L, Cui T, Zheng B, Wang N, Luo J, Yang B, et al. MNDR v3.0: mammal ncRNA-disease repository with increased coverage and annotation. Nucleic Acids Res. 2021;49(D1):D160–4.
35. Paraskevopoulou MD, Georgakilas G, Kostoulas N, Reczko M, Maragkakis M, Dalamagas TM, et al. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. Nucleic Acids Res. 2013;41(Database issue):D239–45.
36. Wang P, Li X, Gao Y, Guo Q, Wang Y, Fang Y, et al. LncACTdb 2.0: an updated database of experimentally supported ceRNA interactions curated from low- and high-throughput experiments. Nucleic Acids Res. 2019;47(D1):D121–7.
37. Jeggari A, Marks DS, Larsson E. MiRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics. 2012;28(15):2062–3.
38. 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):D92–7.
39. Cheng L, Wang P, Tian R, Wang S, Guo Q, Luo M, et al. LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. Nucleic Acids Res. 2019;47(D1):D140–4.
40. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, et al. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. Nucleic Acids Res. 2019;47(D1):D1013–7.
41. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. MiR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009;37(Database issue):D98-104.
42. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. MiRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res. 2020;48(D1):D148–54.
43. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1):D845–55.
44. Wang Z, Monteiro CD, Jagodnik KM, Fernandez NF, Gundersen GW, Rouillard AD, et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. Nat Commun. 2016;7:12846.
45. Pletscher-Frankild S, Pallejà A, Tsafou K, Binder JX, Jensen LJ. DISEASES: text mining and data integration of disease-gene associations. Methods. 2015;74:83–9.
46. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res. 2019;47(D1):D955–62.

47. Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. Bioinformatics. 2010;26(13):1644–50.
48. Li J, Gong B, Chen X, Liu T, Wu C, Zhang F, et al. DOSim: an R package for similarity between diseases based on disease ontology. BMC Bioinform. 2011;12:266.
49. The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330–8.
50. Wang JZ, Du Z, Payattakool R, Yu PS, Chen CF. A new method to measure the semantic similarity of GO terms. Bioinformatics. 2007;23(10):1274–81.
51. Laarhoven TV, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics. 2011;27(21):3036–43.
52. Ganegoda GU, Li M, Wang W, Feng Q. Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations. IEEE Trans Nanobiosci. 2015;14(2):175–83.
53. Li Y, Patra JC. Genome-wide inferring gene–phenotype relationship by walking on the heterogeneous network. Bioinformatics. 2010;26(9):1219–24.
54. Wen Y, Han G, Anh VV. Laplacian normalization and bi-random walks on heterogeneous networks for predicting lncRNA-disease associations. BMC Syst Biol. 2018;12(Suppl 9):122.
55. Zhao ZQ, Han GS, Yu ZG, Li J. Laplacian normalization and random walk on heterogeneous networks for disease-gene prioritization. Comput Biol Chem. 2015;57:21–8.
56. Zheng Y, Xu Q, Liu M, Hu H, Xie Y, Zuo Z, et al. LnCAR: a comprehensive resource for lncRNAs from cancer arrays. Cancer Res. 2019;79(8):2076–83.
57. Thanikachalam K, Khan G. Colorectal cancer and nutrition. Nutrients. 2019;11(1):164.
58. Song Q, Shang J, Yang Z, Zhang L, Zhang C, Chen J, et al. Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. J Transl Med. 2019;17(1):70.
59. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. Lancet. 2020;396(10251):635–48.
60. Coughlin SS. Epidemiology of breast cancer in women. Adv Exp Med Biol. 2019;1152:9–29.

Publisher's Note

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