Mathematical Biosciences 315 (2019) 108229

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Predicting lncRNA-disease associations using network topological similarity based on deep mining heterogeneous networks

Zhang Hui[a], Liang Yanchun[a,b], Peng Cheng[a], Han Siyu[a], Du Wei[a,*], Li Ying[a,*]

[a] College of Computer Science and Technology, Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
[b] Zhuhai Laboratory of Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Jilin University, Zhuhai 519041, China

ARTICLE INFO

Keywords:
Deep learning
lncRNA-disease association prediction
Similarity measure
Biological network science

ABSTRACT

A kind of noncoding RNA with length more than 200 nucleotides named long noncoding RNA (lncRNA) has gained considerable attention in recent decades. Many studies have confirmed that human genome contains many thousands of lncRNAs. LncRNAs play significant roles in many important biological processes, including complex disease diagnosis, prognosis, prevention and treatment. For some important diseases such as cancer, lncRNAs have been novel candidate biomarkers. However, the role of lncRNAs in human diseases is still in its infancy, and only a small part of lncRNA-disease associations have been experimentally verified. Predicting lncRNA-disease association is an important way to understand the mechanism and function of lncRNA involved in diseases to enrich the annotations of lncRNA. Therefore, it is urgent to prioritize lncRNAs potentially associated with diseases. Biological system is a highly complex heterogenous network involved different molecules. Therefore, the algorithms based on network methods have been extensively applied in information fields which can provide a quantifiable characterization for the networks characterizing multifarious biological systems. A heterogeneous network topology possessing abundant interactions between biomedical entities is rarely utilized in similarity-based methods for predicting lncRNA-disease associations based on the array of varying features of lncRNAs and diseases. DeepWalk, encoding the relations of nodes in a continuous vector space, is an extension of language model and unsupervised learning from sequence-based word to network. In this article, we present a novel lncRNA-disease association prediction method based on DeepWalk, which enhances the existing association discovery methods through a topology-based similarity measure. We integrate the heterogeneous data to construct a Linked Tripartite Network which is a heterogeneous network containing three types od nodes which generated from bioinformatics linked datasets and use DeepWalk method to extract topological structure features of the nodes in the linked tripartite network for calculating similarities. Our proposed method can be separated into the following steps: Firstly, we integrate heterogeneous data to construct a Linked Tripartite Network: containing the topological interactions of known lncRNA-disease, lncRNA-microRNA and microRNA-disease. Secondly, the topological structure features of the nodes are extracted based on DeepWalk. Thirdly, similarity scores of disease-disease pairs and lncRNA-lncRNA pairs are computed based on the topology of this network. Finally, new lncRNA and disease associations are discovered by rule-based inference method with lncRNA-lncRNA similarities. Our proposed method shows superior predictive performance for prediction of lncRNA-disease associations based on topological similarity from heterogenous network. The AUC value is used to show the performance of our method. The similarity measurement using network topology based on DeepWalk provide a novel perspective which is different from the similarity derived from sequence or structure information.

Availability: All the data and codes are freely availability at: https://github.com/Pengeace/lncRNA-disease-link.

* Corresponding authors.
E-mail addresses: weidu@jlu.edu.cn (W. Du), liying@jlu.edu.cn (Y. Li).

https://doi.org/10.1016/j.mbs.2019.108229
Received 30 September 2018; Received in revised form 12 May 2019; Accepted 16 July 2019
Available online 16 July 2019

1. Introduction

A kind of noncoding RNA with length more than 200 nucleotides is named long noncoding RNA (lncRNA) [1–3], which has gained considerable attention in recent decades. Many studies have confirmed that the human genome contains many thousands of lncRNAs. A large quantity of lncRNAs play significant roles in many important biological processes including chromatin modification, transcriptional and post-transcriptional regulation, genomic splicing, differentiation, immune responses and so on [4–6]. The mutations and malfunctions of lncRNAs are closely related to human diseases such as neurological disorders [7], blood diseases [8], cardiovascular diseases [9], and various cancers [10]. LncRNAs have been involved in complex disease diagnosis, prognosis, prevention, and treatment [11–13]. LncRNAs have been novel candidate biomarkers for cancers.

The experimental approaches for lncRNA-disease association prediction are expensive and time-consuming. The database lncRNADisease [14] has collected the associations between more than 100 diseases and more than 250 lncRNAs. However, the NONCODE database [15] has collected more than 90,000 human lncRNAs. The majority relationship between lncRNAs and diseases are still unknown. It is therefore urgent to put forward the computational approaches to identify a novel lncRNA and disease associations.

In recent years, some computational models have been proposed to identify potential associations between lncRNAs and diseases based on network science and machine learning algorithms. Biological system is a highly complex heterogenous network involved different molecules. Therefore, the algorithms based on network science have been extensively applied in information fields which can provide a quantifiable characterization for the networks characterizing multifarious biological systems. Multifarious computational models have been proposed to identify lncRNA-disease associations or lncRNA-protein interactions relationships by integrating heterogeneous data sources and machine learning algorithms [16–19]. Chen et al. [5] proposed a computational model of LRLSLDA to predict potential disease-related lncRNAs based on the semi-supervised learning method which is in Laplacian regularized least squares framework. Furthermore, LRLSLDA does not require a negative sample and can produce reliable results based on integrating lncRNA expression profile and known lncRNA-disease associations. Based on the assumption that functional similar lncRNAs tend to associated with similar diseases, Chen et al. [20] developed a new lncRNA-disease association model LRLSLDA-LNCSIM, integrating disease semantic similarity and lncRNA functional similarity with lncRNA expression similarity, using lncRNA Gaussian interaction profile kernel similarity and disease Gaussian interaction profile kernel similarity in LRLSLDA. Huang et al. further proposed novel lncRNA and disease association prediction model [21], in which the general hierarchical structure information of disease directed acyclic graphs are used for disease similarity calculation based on an edge-based method. Zhao et al. [22] developed a naïve Bayesian classifier-based model based on integrating multi-omic data, genomic, regulome and transcriptome data, to identify new cancer-related lncRNAs by known cancer-related lncRNAs. The limitation of supervised classifiers was negative samples were obtained by randomly selecting unlabeled lncRNA-disease pairs. Wang and Cui et al. proposed a sequence based computation model to predict lncRNA-disease association based on the crosstalk between lncRNAs and microRNAs [23]. LDAP [24] is a web server for lncRNA-disease association prediction by integrating multiple biological data resources based on lncRNA similarities and disease similarities which employed the geometric mean of matrix to fuse different data resources while the SVM is used to predict potential lncRNA-disease associations. BRWLDA [25] is a model that performs Bi-Random Walks to predict new lncRNA disease associations and utilizes multiple heterogeneous data to construct the lncRNA functional similarity network, and Disease Ontology to construct a disease network which are the bases for constructing a directed bi-relational network.

In addition, another type of computational approaches for lncRNA-disease association are based on integrating known lncRNA-disease association network, disease similarity network and lncRNA similarity network to construct heterogeneous network and implement global network similarity-based models to obtain potential associations between lncRNAs and diseases. These global network similarity-based models are based on random walk and various network propagation algorithms. Sun et al. [26] proposed a global network-based computational method named RWRlncD to infer potential human lncRNA-disease associations. RWRlncD implemented a random walk algorithm with restart (RWR) based on constructing lncRNA-disease association network, disease similarity network and lncRNA functional similarity network. Zhou et al. [16] proposed RWRHLD method to predict lncRNA-disease associations based on the assumption that lncRNAs with more common miRNA interaction partners tend to be associated with similar diseases. RWRHLD constructed a heterogeneous network by integrating lncRNA and miRNA-associated lncRNA crosstalk network, disease-disease similarity network and the known lncRNA-disease association network and implemented a random walk on it. IRWRLDA [17] is a model to predict novel lncRNA-disease associations by integrating known lncRNA-disease associations, disease semantic similarity, and various lncRNA similarity measures. Ganegoda et al. proposed a kernel based random walk with restart in heterogeneous network model (KRWRH) to predict new lncRNA-disease associations, which incorporates with disease-disease similarity network, lncRNA-lncRNA similarity network and known lncRNA-disease association network [27]. Gaussian interaction profile kernel was used to calculate the similarities of diseases and lncRNAs in KRWRH. KATZLD [28] is the model of KATZ measure which is a graph-based computational method which transforms the problem of link prediction into a problem of calculating similarities between nodes in a heterogeneous network for lncRNA-disease association prediction by integrating known lncRNA-disease associations, lncRNA expression profiles, lncRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity. It is worth mentioning that KATZLDA can work for diseases without known related lncRNAs and lncRNAs without known associated diseases.

It is worth mentioning that in recent years, there are many methods based on network analysis to predict the association between lncRNA and disease, but these methods construct heterogeneous networks with more diverse nodes, not only lncRNA and disease, but also microRNA, drug and ceRNA. LncPriCNet [29] used a multi-level composite network which integrated genes, lncRNAs, phenotypes and their associations to prioritize candidate lncRNAs associated with diseases by random walking with restart (RWR) algorithm to a multi-level network. HGLDA [20] is a novel model of Hypergeometric distribution for lncRNA disease association prediction by integrating miRNA-disease associations and lncRNA-miRNA interactions. HGLDA can predict without relying on lncRNA disease associations. Matrix Factorization based LncRNA–Disease Association prediction model (MFLDA) [30] is proposed to account for the quality and relevance of different heterogeneous data sources, MFLDA first encodes relevant data sources related to lncRNAs or diseases in individual relational data matrices, such as lncRNA-miRNA associations, lncRNA-gene interactions, lncRNA-gene function associations, lncRNA-disease associations, miRNA-gene interactions, miRNA-disease associations, Gene Ontology annotations, gene-disease associations, gene- drug associations, gene-gene interactions and drug-drug interactions. And MFLDA can preset weights for these matrices. DisLncPri [31] is an improved method for lncRNA-disease association prediction based on ceRNA theory and functional genomics data which lncRNAs can be mapped to nine functional genomics contexts through their mRNA interactions. A lncRNA-disease association network is constructed based on the available lncRNA-disease associations as a bipartite network and uses the propagation algorithm for Network analysis. In the process of constructing network, the two biological networks derived from it have become a highlight,

Table 1
Summary of computational methods for the prediction of lncRNA-disease associations.

Method	Brief description
RWRHLD [16]	Prioritisation of candidate lncRNA-disease associations by integrating heterogenous networks
IRWRLDA [17]	Improved random walk with restart for lncRNA-disease association prediction
LRLSLDA [5]	Laplacian regularised least squares for lncRNA-disease association inference.
LRLSLDA-LNCSIM [20]	A new lncRNA-disease association model integrating disease semantic similarity and lncRNA functional similarity with lncRNA expression similarity
Zhao et al. [22]	Naïve Bayesian classifier to identify cancer-related lncRNAs
[a]lncDisease [23]	A sequence-based computation model to predict lncRNA-disease association.
RWRlncD [26]	A network-based computation model performing random walk with restart on the network to predict lncRNA-disease association
KRWRH [27]	Inference of lncRNA-disease associations using Gaussian interaction profile kernel and random walk with restart.
[b]LncPriCNet [29]	Using heterogenous networks and random walk with restart to predict lncRNA-disease.
KATZLDA [28]	A network-based method with the KATZ centrality measure.
[c]LDAP [24]	A web server for lncRNA-disease association prediction
HGLDA [37]	A Hyper Geometric distribution model for lncRNA-disease association
[d]MFLDA [30]	Method by matrix factorisation based on fusion framework to identify lncRNA-disease associations.
BRWLDA [25]	Method bi-random walk with restart to network of lncRNA functional similarities and disease associations
DisLncPri [31]	A disease lncRNA prioritisation method based on ceRNA theory and functional genomics data
lncDN/DlncN [32]	LncRNA-implicated disease networks (lncDN) and disease-associated lncRNA networks (DlncN)
[e]ncPred [33]	Tripartite network-based method which integrates information on ncRNAs, targeting and their associations with diseases.

Availability:.
[a] http://www.cuilab.cn/lncRNADisease.
[b] https://cran.r-project.org/src/contrib/Archive/LncPriCNet/.
[c] http://bioinformatics.csu.edu.cn/ldap.
[d] http://mlda.swu.edu.cn/data_eng.php.
[e] http://alpha.dmi.unict.it/ncPred/.

including “lncRNA-implicated disease network”(lncDN) and “disease-associated lncRNA network”(DlncN) [32]. With the help of data which in a network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases from the previous method [32], ncPred [33] can predict novel ncRNA-disease associations whose aim is to compute association's prediction starting from a tripartite network which integrates information on ncRNAs, targeting and their associations with diseases in order to improve prediction quality and accuracy. We will provide a more comprehensive summary of the tool resources or network resources related to lncRNA-disease associations in Table 1.

A heterogeneous network topology possessing abundant interactions between biomedical entities is rarely utilized in similarity-based methods for predicting lncRNA-disease associations based on the array of varying features of lncRNAs and diseases. Deep learning has been used to network representation to reveal topology features of vertices of a large network that can be adapted in accommodating the similarity-based solutions such as DeepWalk, node2vec and LINE [34–36], which can achieve the vector-based representation of nodes. DeepWalk, encoding the relations of nodes in a continuous vector space, is an extension of language model and unsupervised learning from sequence-based word to graph.

In this article, we present a novel lncRNA-disease association prediction method based on DeepWalk, which enhances the existing association discovery methods through a topology-based similarity measure. We integrate heterogeneous data to construct a Linked Tripartite Network containing three types of nodes in this heterogeneous network and use DeepWalk method to extract topological structure features of the nodes in the tripartite network for calculating similarities. The AUC value is used to evaluate the performance of our proposed method. Furthermore, the inferred potential lncRNA-disease associations by our proposed are verified by literatures mining technology. These results show that our proposed method is superior, which can support further biological experiments and promote research productivity.

2. Materials and methods

2.1. Method overview

The overview of proposed method is shown in Fig. 1. Our proposed method can be separated into the following steps: Firstly, we integrate heterogeneous data to construct a Linked Tripartite Network: containing the interactions of lncRNA-disease, lncRNA-microRNA and microRNA-disease. Secondly, the topological structural features of the nodes are extracted based on DeepWalk. Thirdly, similarity scores of disease-disease pairs and lncRNA-lncRNA pairs are inferred based on the topology feature vectors of the nodes. Finally, new lncRNA and disease associations are discovered by rule-based inference method by means of lncRNA-lncRNA similarities.

2.2. Linked Tripartite Network construction

A heterogeneous network we constructed and called Linked Tripartite network consists of three types of nodes, lncRNAs, microRNAs and diseases. Linked Tripartite network contains three kinds of relationships, including lncRNA-disease associations, lncRNA-microRNA interactions and microRNA-disease associations. And relationships in the Linked Tripartite network are obtained as follows: LncRNA-disease associations were downloaded from lncRNADisease database [14] and Lnc2Cancer database [38]. LncRNA-disease associations dataset from lncRNADisease database is experimentally supported, which integrated 2947 lncRNA-disease associations and 475 lncRNA interaction entries, including 914 lncRNAs and 329 diseases. 1488 lncRNA-cancer associations are obtained from lnc2Cancer database, including 666 lncRNAs and 97 human cancers. After getting rid of duplicate lncRNA-disease associations, 1454 lncRNA-disease associations including 804 lncRNAs and 288 diseases were obtained. We collected miRNA-disease associations from HMDD database [39], miRCancer database [40], miR2Disease database [41] and the data provided in the method PBMDA [42]. From HMDD database, 5430 high-quality experimentally verified human miRNA-diseases associations about 572 miRNAs and 378 diseases are downloaded. From miRCancer database, 5562 high-quality experimentally verified human miRNA-cancer associations about 44,353 miRNAs and 184 cancers are obtained. From miR2Disease database, 3273 human miRNA-disease associations about 349 miRNAs and 163 diseases are downloaded. Finally, we can use the data in the article named PBMDA [42]. PBMDA constructed a heterogeneous graph consisting of three interlinked subgraphs and further adopted depth-first search algorithm to infer potential miRNA-disease associations based on path, which integrated

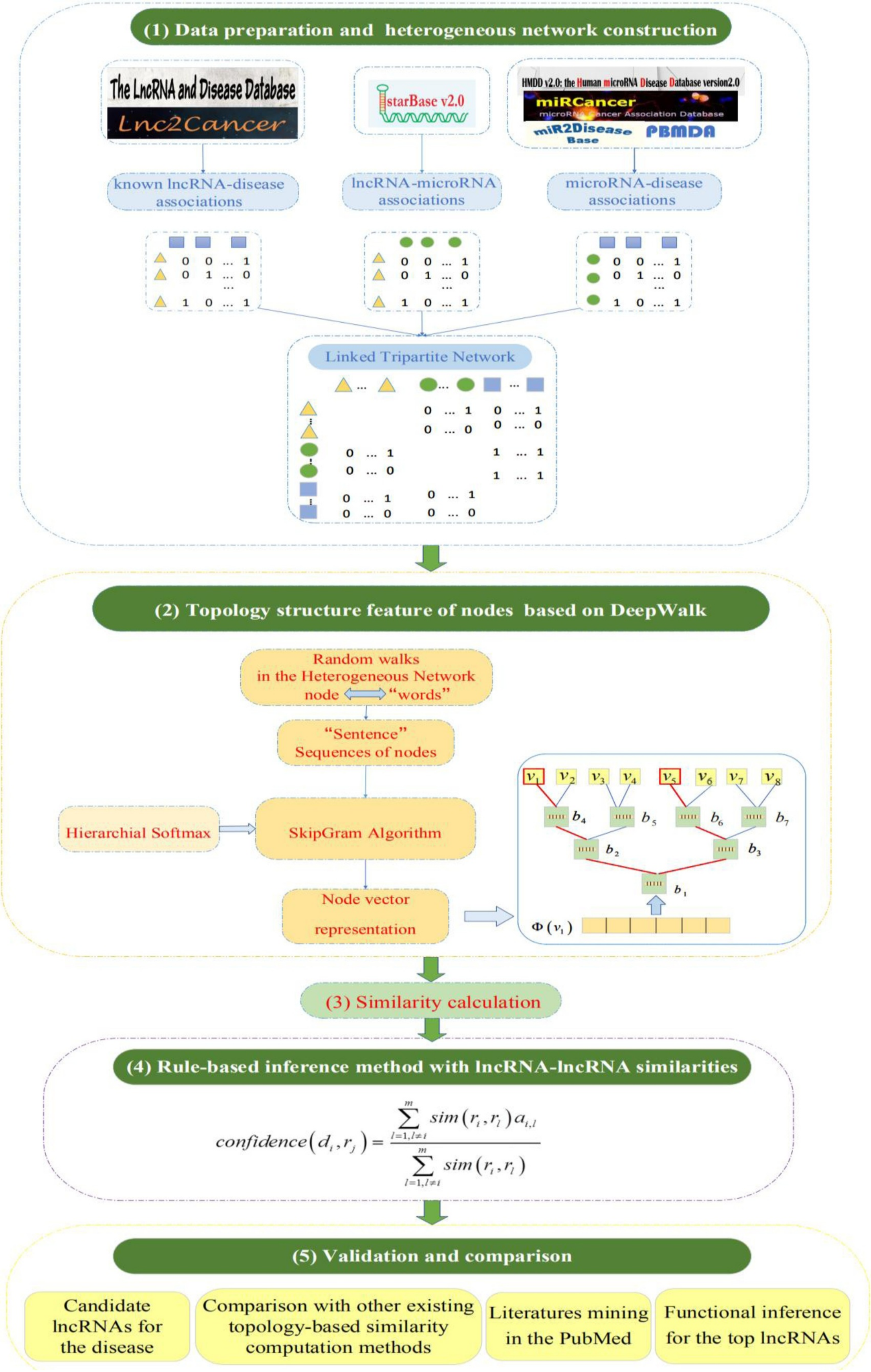
(1) Data preparation and heterogeneous network construction
The LncRNA and Disease Database
Lnc2Cancer
starBase v2.0
HMDD v2.0: the Human microRNA Disease Database version2.0
miRCancer
microRNA Cancer Association Database
miR2Disease Base
PBMDA
known lncRNA-disease associations
lncRNA-microRNA associations
microRNA-disease associations
Linked Tripartite Network
(2) Topology structure feature of nodes based on DeepWalk
Random walks in the Heterogeneous Network
node ⟺ "words"
"Sentence" Sequences of nodes
Hierarchial Softmax
SkipGram Algorithm
Node vector representation
v_1 v_2 v_3 v_4 v_5 v_6 v_7 v_8
b_4 b_5 b_6 b_7
b_2 b_3
b_1
$\Phi(v_1)$
(3) Similarity calculation
(4) Rule-based inference method with lncRNA-lncRNA similarities
$confidence(d_i, r_j) = \frac{\sum_{l=1, l\neq i}^{m} sim(r_i, r_l) a_{i,l}}{\sum_{l=1, l\neq i}^{m} sim(r_i, r_l)}$
(5) Validation and comparison
Candidate lncRNAs for the disease
Comparison with other existing topology-based similarity computation methods
Literatures mining in the PubMed
Functional inference for the top lncRNAs

Fig. 1. The overall workflow of our method.

known human miRNA-disease associations from HMDD database, miRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity for miRNAs and diseases. After integration of miRNA-disease associations and getting rid of duplicate associations, 11,835 miRNA-disease associations were obtained, including 1080 miRNAs and 590 diseases. And lncRNA-miRNA interactions were downloaded from starBase v2.0 database [43], which provided the most comprehensive experimentally confirmed lncRNA-miRNA interactions based on large scale CLIP-Seq data. After getting rid of duplicate interactions, 10,198 lncRNA-miRNA interactions were obtained, including 277 miRNAs and 1127 lncRNAs.

2.3. *Topology structure feature of the nodes based on DeepWalk*

DeepWalk, a deep learning method, is utilized to vectorize the vertices in the heterogeneous network to reveal the topology features of the nodes for calculating the similarities within a heterogeneous network which generated from biomedical linked datasets. Here, DeepWalk is used to compute the vector-based representation of the nodes including lncRNAs and diseases. DeepWalk has two main components. First, for each vertex $\mathrm{v_i}$, γ random walks with length t are conducted, with $\mathrm{v_i}$ as the starting vertex. Second, the vertex representation is updated with the SkipGram algorithm [44] for each walk. SkipGram maximizes the co-occurrence likelihood of the vertices that come into view within a window w using an independent assumption as follows:

$$\Pr(\{v_{i-w,}\cdots,v_{i+w}\}\setminus v_i|\Phi(v_i)) = \prod_{j=i-w,j\neq i}^{i+w} \Pr(v_j|\Phi(v_i))$$

where Φ denotes the latent topological representation associated with every vertex v_i. Φ is represented by a $|V| \times d$ matrix, where $|V|$ is the cardinality of vertex set V, and d is the dimension of the vertex vector. To speed up the training time, $\Pr(v_j|\Phi(v_i))$ is factorized with Hierarchical Softmax [45] by allocating the vertices to the leaves of a binary tree, and $\Pr(v_j|\Phi(v_i))$ is then computed as follows:

$$\Pr(v_j|\Phi(v_i)) = \prod_{l=1}^{\lceil log|V|\rceil} 1/\left(1 + e^{-\Phi(v_i)\cdot\psi(b_l)}\right)$$

where $\psi(b_l)$ represents the parent of tree node $b_l\cdot(b_0, b_1\cdots b_{\lceil log|V|\rceil})$ is the sequence of tree nodes to identify

v_j, where b_0 = root and b + $\lceil log|V|\rceil$ + $= v_j$.

After completing the training, the output of DeepWalk is a latent topological representation of nodes in the network. Therefore, the similarity of two vertices u and v can be computed by using cosine similarity as follows:

$$\mathrm{sim}(u, v) = \frac{\sum_{k=1}^{d} u_k v_k}{\sum_{k=1}^{d} u_k^2 \sqrt{\sum_{k=1}^{d} u_k^2}}$$

where d is the dimension, and u_i and v_i are the components of vector u and v, respectively.

2.4. *Rule-based inference method*

We adapted a rule-based inference method with lncRNA-lncRNA similarity, which was inspired by complex network theory [33] to predict disease-related lncRNA candidates with lncRNA-lncRNA similarities. The similar lncRNAs tend to associated with the same disease. The rule-based inference method predicts a lncRNA-disease association s(d_i, r_j), if a disease d_i is associated with a lncRNA that has a similar lncRNA r_j. For a pair of (d_i, r_j), a confidence score of the pair is calculated as follows:

$$\mathrm{confidence}(d_i, r_j) = \frac{\sum_{l=1,l\neq j}^{m} sim(r_i, r_l)a_{i,l}}{\sum_{l=1,l\neq j}^{m} sim(r_i, r_l)}$$

where $sim(r_i, r_l)$ is the similarity between r_i and r_l, and $a_{i,l} = 1$ if there is an existing association between d_i and r_l otherwise $a_{i,l} = 0$.

For a disease d_i or a lncRNA r_j as the input query, the confidences are normalized as follows:

$$\mathrm{nomarlized\ Confidence}(d_i, r_j) = \frac{\mathrm{confidence}(d_i, r_j) - Max(\cdot, r_j)}{Max(\cdot, r_j) - Min(\cdot, r_j)}$$

where $Max(\cdot, r_j)$ and $Min(\cdot, r_j)$ represent the maximum and minimum associated confidence score of lncRNA r_j with diseases that have no known association with r_j, respectively.

3. Result

3.1. *Evaluation of prediction performance and compared with other methods*

To perform a proper evaluation of our proposed method, we utilize AUC value to demonstrate its superior performance. The predicted potential lncRNA-disease associations and the top-ranking associations are further verified by biomedical literatures mining. These results afford convincing evidence of the good performance of our method as well as potential value in supporting further biological experiments and promoting research productivity.

First, all the microRNA-lncRNA associations, microRNA-disease associations and lncRNA-disease associations are integrated to a comprehensive Linked Tripartite network. Then, DeepWalk is employed to generate the embedding representations of each microRNA, lncRNA and disease. Next, we apply rule-based inference method to calculate the relevance score for each lncRNA disease pair. The inferred association scores are used to calculate the AUC value. In AUC calculation, we regard all the known lncRNA-disease associations as positive pairs and the rest lncRNA disease pairs as negative pairs. The AUC value is 0.9316, which shows the performance of our method. The AUC of our method for some diseases are shown in the Table 2. We also utilize five-cross-validation, we obtain an AUC value of 0.915. Before cross validation, we mark some of the data in the lncRNA-disease associations network as interaction pairs and put them in the network for training to ensure that the network is connected. Then, the remaining data other than the markers are divided into five parts to cross validation.

In order to comprehensively assess the predictive ability of our method to predict lncRNA-disease associations, we compare our method about some diseases with two methods: RWRlncD [26] and RWRHLD [16]. The prediction results of our method, RWRlncD and RWRHLD are 0.9316, 0.5432 and 0.9231, respectively. We choose some cancers with high morbidity and mortality, as well as typical central nervous system degenerative diseases on which researchers focus closely. We list the top 10 LncRNA cases associated with the diseases including breast cancer, prostate cancer, pancreas cancer, ovarian cancer, Huntington's disease and Alzheimer's disease in Table 3. For each of the

Table 2
The AUC of our method for some diseases.

Disease name	AUC
Breast cancer	0.9668
Prostate cancer	0.9871
B-cell neoplasms	0.9367
Ovarian cancer	0.9835
Burkitts lymphoma	0.8654
Alzheimer's disease	0.9954
Beckwith-Wiedemann syndrome	0.7510

Table 3
The number of occurrences of each of predicted top 10 lncRNAs in the PubMed database with lncRNA.

Alzheimer's disease			Huntington's disease		
LncRNA name	Score	PubMed-hits	LncRNA name	Score	PubMed-hits
NCRMS	0.009365	194	Epist	0.010557887	242
BACE1-AS	0.023338	44	TINCR	0.013694799	29
OVAL	0.010345	25	BDNF-AS	0.030364109	11
HESRG	0.00995	21	TCONS_l2_00010365	0.011427827	9
PANDA	0.008817	21	OVAL	0.011593268	6
BDNF-AS	0.016106	20	REST/CoREST-regulated lncRNAs	0.034662853	2
TARID	0.010035	9	HTTAS	0.031974289	2
CCND1	0.008601	9	GDNFOS	0.021079279	2
BC200	0.012978	8	TUG1	0.014362848	2
51A	0.024278802	6	TARID	0.013245289	2
Ovarian cancer			Pancreas cancer		
LncRNA name	Score	PubMed-hits	LncRNA name	Score	PubMed-hits
KRAS1P	0.021621	463	KRAS1P	0.035358	1369
OVAL	0.03915	113	ICR	0.035444	109
TARID	0.028407	47	PTHLH	0.033788	37
TINCR	0.024443	30	HOTAIR	0.045301	18
HTTAS	0.021301	27	MALAT1	0.041686	17
HOTAIR	0.023752	25	H19	0.040227	15
UCA1	0.023648	9	AATBC	0.036427	14
PTCSC	0.021945	9	MINA	0.041407	12
MEG3	0.023257	6	HOTTIP	0.062951	11
GAS5	0.02157	6	PVT1	0.040546	9
Prostate cancer			Breast cancer		
LncRNA name	Score	PubMed-hits	LncRNA name	Score	PubMed-hits
Epist	0.047157	1544	IRAIN	0.080266	2967
PCA3	0.094911	447	SKP2	0.07347	129
DRAIC	0.101867	444	NRG1	0.078224	90
TCONS_l2_00010365	0.045406	93	UCA1	0.071403	27
PRINS	0.081916	71	BCAR4	0.105443	19
PCGEM1	0.097514	29	BC200	0.088655	8
PCAT-1	0.062119	22	CCAT2	0.077033	8
SChLAP1	0.097437	21	ZFAS1	0.073645	7
PRNCR1	0.066866	17	NKILA	0.103446	6
PCGEM1	0.101266	13	SRA1	0.087546	6

diseases listed above, we looked up the number of occurrences of each of the predicted top 10 lncRNAs and the associated disease in the PubMed database, which contains two other methods that do not predict successful association pairs. The top 10 predicted lncRNAs for each of the above six diseases have been verified by the biomedical literature are also shown in Table 3.

We listed lncRNA-disease association predictions which the top 10 lncRNAs related to breast cancer by our method, RWRlncD and RWRHLD, respectively. The predicted results of each method were verified by the number of PubMed-Hits in Table 4 and PubMed Central-Hits (PMC—Hits) in Table 5. By means of comparison, our method performs better in terms of the number of PubMed-Hits.

3.2. Functional inference of the top lncRNAs

One of the most important molecular mechanisms of lncRNAs is their interactions with proteins. Therefore, predicting lncRNA-protein interactions is an important way to explore the function and enrich annotation of lncRNAs in depth. We intend to infer the functional information of lncRNAs using predicted proteins that interact with lncRNAs. Here, the computational model named LncADeep [46] is used to predict lncRNA-protein interactions. LncADeep outperforms state-of-the-art methods which predicts lncRNA-protein interactions based on deep learning model, using both sequence and structure information.

The process of functional inference is mainly divided into three steps. First, for the six diseases mentioned above, the top lncRNAs

Table 4
Predicted breast cancer related lncRNAs, which were ranked in top 10 by our method, RWRlncD or RWRHLD, and predicted results of each method correspond to the PubMed-Hits.

LncRNA	Rank	PubMed-Hits (Our method)	LncRNA	PubMed-Hits (RWRlncD)	LncRNA	PubMed-Hits (RWRHLD)
IRAIN	1	2967	LINC00900	0	ACTA2-AS1	0
SKP2	2	129	LINC00638	0	TINCR	2
NRG1	3	90	AC062029.1	0	CBR3-AS1	0
UCA1	4	27	ITGA9-AS1	0	TUSC8	1
BCAR4	5	19	RP11-66B24.4	0	BDNF-AS	1
BC200	6	8	C6orf3	0	LINC00271	0
CCAT2	7	8	LL0XNC01-116E7.2	0	LINC00900	0
ZFAS1	8	7	LINC01004	0	LINC00638	0
NKILA	9	6	MORF4L2-AS1	0	AC062029.1	0
SRA1	10	6	RP4-583P15.10	1	ITGA9-AS1	0

Table 5
Predicted results of each method correspond to the PubMed Central-Hits (PMC−Hits).

LncRNA	Rank	PMC−Hits (Our method)	LncRNA	PMC−Hits (RWRlncD)	LncRNA	PMC−Hits (RWRHLD)
IRAIN	1	21	LINC00900	1	ACTA2-AS1	3
SKP2	2	129	LINC00638	2	TINCR	161
NRG1	3	1230	AC062029.1	1	CBR3-AS1	13
UCA1	4	735	ITGA9-AS1	2	TUSC8	3
BCAR4	5	112	RP11-66B24.4	1	BDNF-AS	189
BC200	6	184	C6orf3	0	LINC00271	4
CCAT2	7	289	LL0XNC01-116E7.2	1	LINC00900	1
ZFAS1	8	7	LINC01004	1	LINC00638	2
NKILA	9	77	MORF4L2-AS1	1	AC062029.1	1
SRA1	10	172	RP4-583P15.10	151	ITGA9-AS1	2

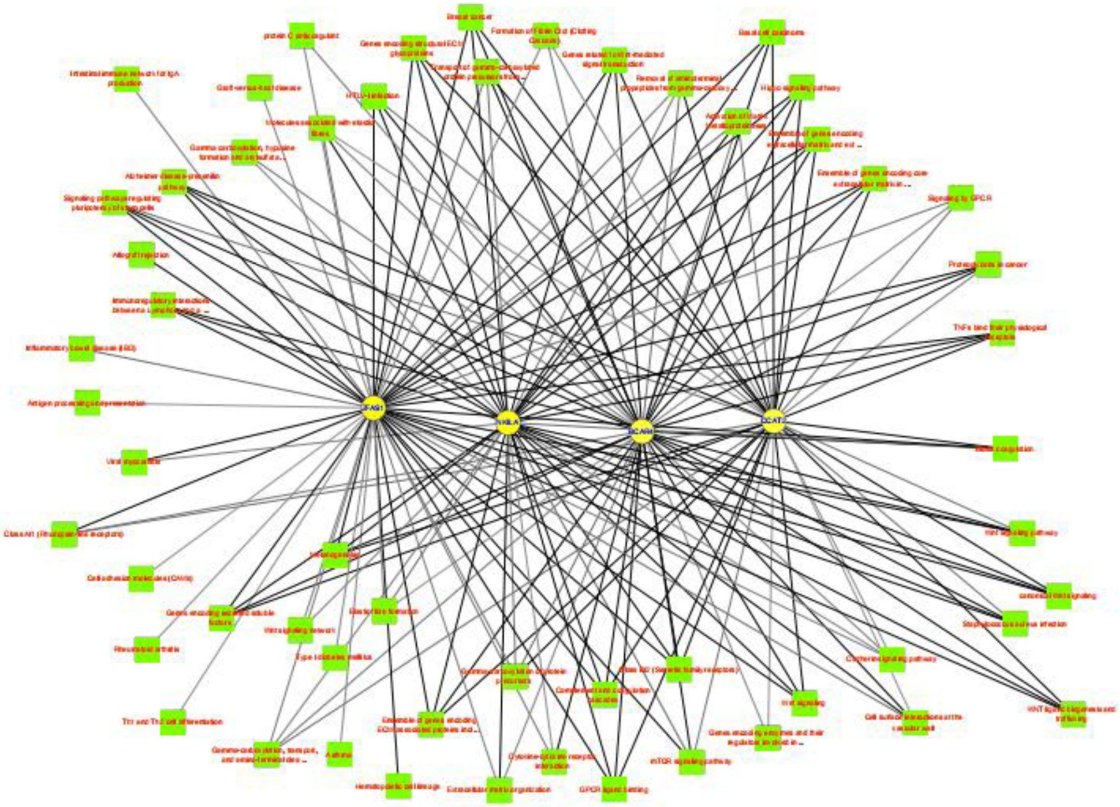
ZFAS1
NKILA
BCAR4
CCAT2
Breast cancer
protein C anticoagulant
Genes encoding structural ECM glycoproteins
Formation of Fibrin Clot (Clotting Cascade)
Basal cell carcinoma
Intestinal immune network for IgA production
Graft-versus-host disease
HTLV-I infection
Transport of gamma-carboxylated protein precursors from ...
Genes related to PIP3-mediated signal transduction
Removal of aminoterminal propeptides from gamma-carboxy ...
Hippo signaling pathway
Molecules associated with elastic fibers
Activation of Matrix Metalloproteinases
Ensemble of genes encoding extracellular matrix and ext ...
Gamma carboxylation, hypusine formation and arylsulfata ...
Ensemble of genes encoding core extracellular matrix in ...
Alzheimer disease-presenilin pathway
Signaling by GPCR
Signaling pathways regulating pluripotency of stem cells
Allograft rejection
Proteoglycans in cancer
Immunoregulatory interactions between a Lymphoid and a ...
TNFs bind their physiological receptors
Inflammatory bowel disease (IBD)
Antigen processing and presentation
Blood coagulation
Viral myocarditis
Class A/1 (Rhodopsin-like receptors)
Wnt signaling pathway
Cell adhesion molecules (CAMs)
canonical Wnt signaling
Genes encoding secreted soluble factors
Elastic fibre formation
Staphylococcus aureus infection
Rheumatoid arthritis
Wnt signaling network
Type I diabetes mellitus
Gamma-carboxylation of protein precursors
Class B/2 (Secretin family receptors)
Cadherin signaling pathway
Complement and coagulation cascades
Wnt signaling
Cell surface interactions at the vascular wall
WNT ligand biogenesis and trafficking
Th1 and Th2 cell differentiation
Ensemble of genes encoding ECM-associated proteins incl ...
Genes encoding enzymes and their regulators involved in ...
Gamma-carboxylation, transport, and amino-terminal cleav ...
Asthma
Cytokine-cytokine receptor interaction
mTOR signaling pathway
Hematopoietic cell lineage
Extracellular matrix organization
GPCR ligand binding

Fig. 2. Enrichment pathways of the top lncRNAs associated with breast cancer.

associated with the disease were identified using the proposed model. Then, for the top lncRNAs, we found all the associated protein sets predicted by the LncADeep model. Finally, these protein sets were enriched and analyzed to complete the functional inference of disease-related top lncRNAs in case study. For pathway enrichment analysis, Fisher's exact test for the significance test and Benjamini–Hochberg (BH) method for the multiple testing correction with keeping enriched pathways whose adjusted P-value is <0.05 are used. To increase visualization, the enriched pathways of the top lncRNAs associated with breast cancer and ovarian cancer respectively are shown in Figs. 2 and 3. From Figs. 2 and 3, there are many pathways related to cancer.

4. Discussion and conclusion

Identifying novel lncRNA-disease associations become more important for exploring disease pathogenesis with the large number of lncRNAs recognized. In this paper, we utilized the integrated lncRNA-disease, microRNA-disease and lncRNA-microRNA associations to construct a heterogeneous network. And we adopted a deep learning algorithm, DeepWalk, to determine the similarity of each lncRNA-lncRNA pair based on a comprehensive heterogeneous network named Linked Tripartite Network. Then, with a rule-based inference method, similarity measures were assembled to compute the association likelihood of each candidate disease-lncRNA pair. To validate the prediction accuracy of our approach, cross validation was implemented with a lncRNA-disease association dataset. In addition, the top 10 predicted lncRNAs for each of popular human diseases have been verified by the latest experimental literature. Our method performs better in terms of the number of PubMed-Hits and PubMed Central-Hits by comparison with RWRlncD and RWRHLD.

It is worth to point out that our method has some biases. The lncRNA functional similarity network is constructed depending on known lncRNA-disease associations. It would affect the performance of the method. Since the disease names in the LncRNADisease database are not standardized, we avoided the bias of finding a closely matched phenotype for disease similarity scores, so we chose a rule-based approach with lncRNA-lncRNA similarity scores for lncRNA-disease prediction. Hence, calculating lncRNA similarity by integrating more data would benefit the improvement of predictive ability. In the future, we will add new networks such as protein entity network and construct

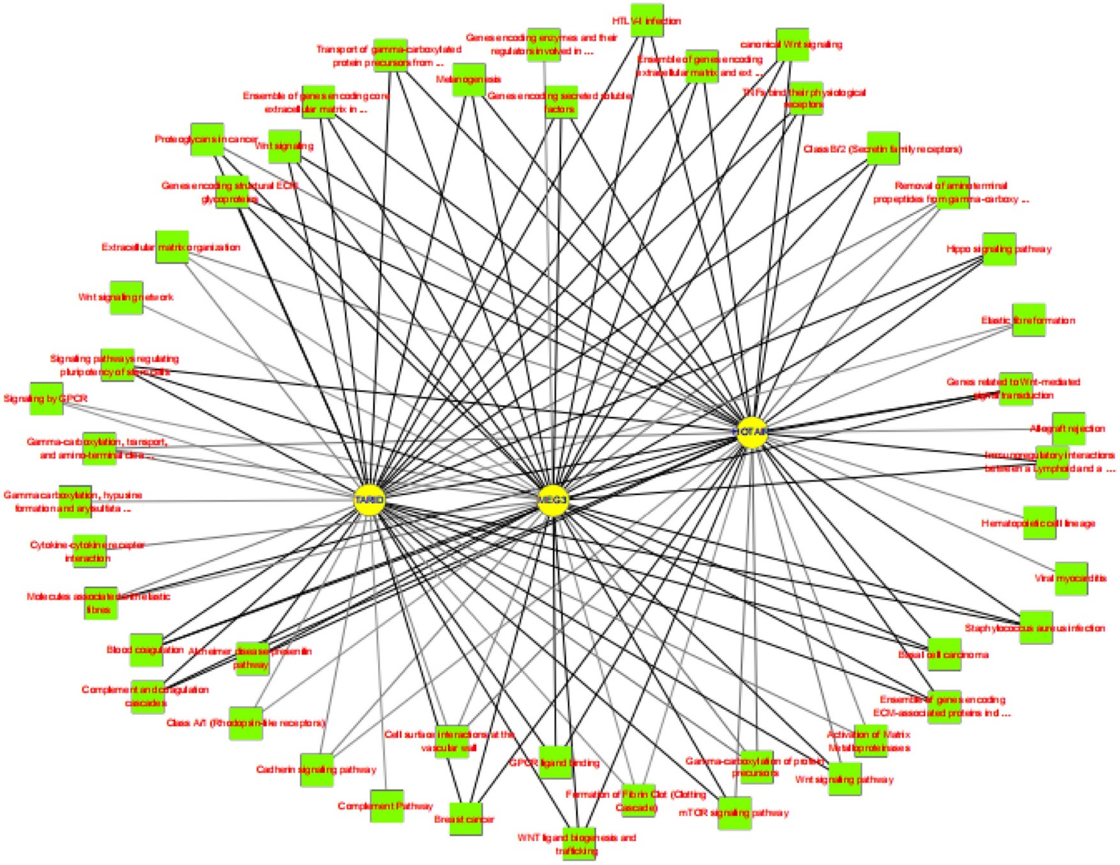
Transport of gamma-carboxylated protein precursors from ...
Genes encoding enzymes and their regulators involved in ...
HTLV-I infection
canonical Wnt signaling
Melanogenesis
Ensemble of genes encoding extracellular matrix and ext ...
Ensemble of genes encoding core extracellular matrix in ...
Genes encoding secreted soluble factors
TNFs bind their physiological receptors
Proteoglycans in cancer
Wnt signaling
Class B/2 (Secretin family receptors)
Genes encoding structural ECM glycoproteins
Removal of aminoterminal propeptides from gamma-carboxy ...
Extracellular matrix organization
Hippo signaling pathway
Wnt signaling network
Elastic fibre formation
Signaling pathways regulating pluripotency of stem cells
Genes related to Wnt-mediated signal transduction
Signaling by GPCR
HOTAIR
Allograft rejection
Gamma-carboxylation, transport, and amino-terminal clea ...
Immunoregulatory interactions between a Lymphoid and a ...
Gamma carboxylation, hypusine formation and arylsulfata ...
TARID
MEG3
Hematopoietic cell lineage
Cytokine-cytokine receptor interaction
Viral myocarditis
Molecules associated with elastic fibres
Staphylococcus aureus infection
Blood coagulation
Alzheimer disease-presenilin pathway
Basal cell carcinoma
Complement and coagulation cascades
Ensemble of genes encoding ECM-associated proteins incl ...
Class A/1 (Rhodopsin-like receptors)
Cell surface interactions at the vascular wall
Activation of Matrix Metalloproteinases
GPCR ligand binding
Gamma-carboxylation of protein precursors
Cadherin signaling pathway
Wnt signaling pathway
Complement Pathway
Formation of Fibrin Clot (Clotting Cascade)
Breast cancer
mTOR signaling pathway
WNT ligand biogenesis and trafficking

Fig. 3. Enrichment pathways of the top lncRNAs associated with ovarian cancer.

disease similarity including more data integration to predict the potential lncRNA-disease associations.

Our method can select and weigh heterogeneous data sources, thus achieving a superior performance. The method we proposed to predict lncRNA-disease associations also map heterogenous data sources onto homologous networks. The shared structure of heterogenous data sources can be explored and exploited by our method. Furthermore, it can readily integrate various heterogenous data sources to predict associations between different types of entities, such as lncRNA-protein associations, associations between genes and GO terms and so on. How to efficiently and jointly construct heterogenous network is an interesting future work and may further improve the performance of method we proposed. In the future, we will add new networks such as protein entity network and construct disease similarity including more data integration to predict the potential lncRNA-disease associations.

Currently, the functions of most lncRNAs remain unknown, and the knowledge of disease-related lncRNAs are very limited and there are still too many candidates to experimentally validate. The method prioritizes these candidate lncRNAs, allowing researchers to select high-ranking disease related lncRNAs and test their functions. The lncRNA-disease associations might provide clues as to the functional mechanisms of lncRNAs. Overall, it is a useful way for lncRNA-disease prioritization and provides better understanding of the molecular mechanisms of human disease at the lncRNA level, which may uncover new diagnostic and therapeutic opportunities. The strategy of the multi-level composite network based on integrated heterogenous data could be used in other fields of biomedicine, such as disease, drug and target discovery.

Funding

This work was supported by the National Natural Science Foundation of China (61572105, 61872418, and 71774154), Natural Science Foundation of Jilin Province (20180101331JC and 20180101050JC), Zhuhai Premier-Discipline Enhancement Scheme, Guangdong Premier Key-Discipline Enhancement Scheme, and Graduate Innovation Fund of Jilin University (2017124), the Fundamental Research Funds for the Central Universities.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] M. Guttman, P. Russell, N.T. Ingolia, J.S. Weissman, E.S. Lander, Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins, Cell 154 (2013) 240–251, https://doi.org/10.1016/j.cell.2013.06.009.

[2] M. Esteller, Non-coding RNAs in human disease, Nat. Rev. Genet. 12 (2011) 861–874, https://doi.org/10.1038/nrg3074.

[3] K.C. Wang, H.Y. Chang, Molecular mechanisms of long noncoding RNAs, Mol. Cell 43 (2011) 904–914, https://doi.org/10.1016/j.molcel.2011.08.018.

[4] O. Wapinski, H.Y. Chang, Long noncoding RNAs and human disease, Trends Cell Biol. 21 (2011) 354–361, https://doi.org/10.1016/j.tcb.2011.04.001.

[5] X. Chen, G.Y. Yan, Novel human lncRNA-disease association inference based on lncRNA expression profiles, Bioinformatics 29 (2013) 2617–2624, https://doi.org/10.1093/bioinformatics/btt426.

[6] T.R. Mercer, M.E. Dinger, J.S. Mattick, Insights into functions, Nat. Rev. Genet. 10 (2009) 155–159, https://doi.org/10.1038/nrg2521.

[7] R. Johnson, Long non-coding RNAs in Huntington's disease neurodegeneration, Neurobiol. Dis. 46 (2012) 245–254, https://doi.org/10.1016/j.nbd.2011.12.006.

[8] M. Ouimet, S. Drouin, M. Lajoie, M. Caron, P. St-Onge, R. Gioia, C. Richer, D. Sinnett, A childhood acute lymphoblastic leukemia-specific lncRNA implicated in prednisolone resistance, cell proliferation, and migration, Oncotarget 8 (2017) 7477–7488, https://doi.org/10.18632/oncotarget.13936.

[9] A. Congrains, K. Kamide, R. Oguro, O. Yasuda, K. Miyata, E. Yamamoto, T. Kawai, H. Kusunoki, H. Yamamoto, Y. Takeya, K. Yamamoto, M. Onishi, K. Sugimoto, T. Katsuya, N. Awata, K. Ikebe, Y. Gondo, Y. Oike, M. Ohishi, H. Rakugi, Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of

ANRIL and CDKN2A/B, Atherosclerosis 220 (2012) 449–455, https://doi.org/10.1016/j.atherosclerosis.2011.11.017.

[10] X. Chen, C.C. Yan, X. Zhang, Z.-H. You, Long non-coding RNAs and complex diseases: from experimental results to computational models, Brief. Bioinform. 18 (2016), https://doi.org/10.1093/bib/bbw060 bbw060.

[11] X. Chen, C.C. Yan, X. Zhang, Z.-H. You, Long non-coding RNAs and complex diseases: from experimental results to computational models, Brief. Bioinform. 18 (2016), https://doi.org/10.1093/bib/bbw060 bbw060.

[12] Y. Zhang, Y. Tao, Q. Liao, Long noncoding RNA: a crosslink in biological regulatory network, Brief. Bioinform. (2017) 1–16, https://doi.org/10.1093/bib/bbx042.

[13] Y. Gu, T. Chen, G. Li, X. Yu, Y. Lu, H. Wang, L. Teng, LncRNAs: emerging biomarkers in gastric cancer, Future Oncol. 11 (2015) 2427–2441, https://doi.org/10.2217/fon.15.175.

[14] G. Chen, Z. Wang, D. Wang, C. Qiu, M. Liu, X. Chen, Q. Zhang, G. Yan, Q. Cui, LncRNADisease: a database for long-non-coding RNA-associated diseases, Nucleic Acids Res. 41 (2013) 983–986, https://doi.org/10.1093/nar/gks1099.

[15] S. Fang, L. Zhang, J. Guo, Y. Niu, Y. Wu, H. Li, L. Zhao, X. Li, X. Teng, X. Sun, L. Sun, M.Q. Zhang, R. Chen, Y. Zhao, NONCODEV5: a comprehensive annotation database for long non-coding RNAs, Nucleic Acids Res. 46 (2018) D308–D314, https://doi.org/10.1093/nar/gkx1107.

[16] M. Zhou, X. Wang, J. Li, D. Hao, Z. Wang, H. Shi, L. Han, H. Zhou, J. Sun, Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network, Mol. BioSyst. 11 (2015) 760–769, https://doi.org/10.1039/C4MB00511B.

[17] X. Chen, Z.-H. You, G.-Y. Yan, D.-W. Gong, IRWRLDA: improved random walk with restart for lncRNA-disease association prediction, Oncotarget 7 (2016) 57919–57931, https://doi.org/10.18632/oncotarget.11141.

[18] W. Peng, W. Lan, Z. Yu, J. Wang, Y. Pan, A framework for integrating multiple biological networks to predict microRNA-disease associations, IEEE Trans. Nanobiosci. 14 (2016), https://doi.org/10.1109/TNB.2016.2633276 1–1.

[19] X. Chen, C.C. Yan, X. Zhang, Z.-H. You, Long non-coding RNAs and complex diseases: from experimental results to computational models, Brief. Bioinform. 18 (2016), https://doi.org/10.1093/bib/bbw060 bbw060.

[20] X. Chen, Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA, Sci. Rep. 5 (2015) 1–11, https://doi.org/10.1038/srep13186.

[21] Y.-A. Huang, X. Chen, Z.-H. You, D.-S. Huang, K.C.C. Chan, ILNCSIM: improved lncRNA functional similarity calculation model, Oncotarget 7 (2016) 7–14, https://doi.org/10.18632/oncotarget.8296.

[22] T. Zhao, J. Xu, L. Liu, J. Bai, C. Xu, Y. Xiao, X. Li, L. Zhang, Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features, Mol. Biosyst. 11 (2015) 126–136, https://doi.org/10.1039/C4MB00478G.

[23] Q. Wang, Junyi, Ma, Ruixia, Cui, LncDisease:a sequence based bioinformatics tool for predicting lncRNA-disease association, (2016). doi:10.1093/narlgkw093.

[24] W. Lan, M. Li, K. Zhao, J. Liu, F. Wu, Y. Pan, Subject section LDAP : a web server for lncRNA-disease asso- ciation prediction, (2016) 3–5.

[25] G. Yu, G. Fu, C. Lu, Y. Ren, J. Wang, BRWLDA: bi-random walks for predicting lncRNA-disease associations, Oncotarget (2017), https://doi.org/10.18632/oncotarget.19588.

[26] J. Sun, H. Shi, Z. Wang, C. Zhang, L. Liu, L. Wang, W. He, D. Hao, S. Liu, M. Zhou, Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network, Mol. BioSyst. 10 (2014) 2074–2081, https://doi.org/10.1039/c3mb70608g.

[27] G.U. Ganegoda, M. Li, W. Wang, Q. Feng, Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations, IEEE Trans. Nanobiosci. 14 (2015) 175–183.

[28] X. Chen, KATZLDA: KATZ measure for the lncRNA-disease association prediction, Sci. Rep. 5 (2015) 1–11, https://doi.org/10.1038/srep16840.

[29] Q. Yao, L. Wu, J. Li, L.G. Yang, Y. Sun, Z. Li, S. He, F. Feng, H. Li, Y. Li, Global prioritizing disease candidate lncRNAs via a multi-level composite network, Sci. Rep. 7 (2017) 1–14, https://doi.org/10.1038/srep39516.

[30] G. Fu, J. Wang, C. Domeniconi, G. Yu, Matrix factorization-based data fusion for the prediction of lncRNA – disease associations, Bioinformatics 34 (2017) 1–9, https://doi.org/10.1093/bioinformatics/btx794.

[31] P. Wang, Q. Guo, Y. Gao, H. Zhi, Y. Zhang, Y. Liu, Improved method for prioritization of disease associated lncRNAs based on ceRNA theory and functional genomics data, Oncotarget 8 (2017) 4642–4655.

[32] X. Yang, L. Gao, X. Guo, X. Shi, H. Wu, F. Song, B. Wang, A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases, PLoS One 9 (2014), https://doi.org/10.1371/journal.pone.0087797.

[33] S. Alaimo, R. Giugno, A. Pulvirenti, ncPred : ncRNA-disease association prediction through tripartite network-based inference, Front. Bioeng. Biotechnol. 2 (2014) 1–8, https://doi.org/10.3389/fbioe.2014.00071.

[34] B. Perozzi, R. Al-Rfou, S. Skiena, DeepWalk: online learning of social representations Bryan, Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '14, 2014, pp. 701–710, , https://doi.org/10.1145/2623330.2623732.

[35] A. Grover, J. Leskovec, Node2Vec, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '16, 2016, pp. 855–864, , https://doi.org/10.1145/2939672.2939754.

[36] J. Tang, M. Qu, LINE : large-scale information network embedding categories and subject descriptors, ACM World Wide Web. 2015, pp. 1067–1077, , https://doi.org/10.1145/2736277.2741093.

[37] X. Chen, Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA, Sci. Rep. 5 (2015) 1–12, https://doi.org/10.1038/srep13186.

[38] S. Ning, J. Zhang, P. Wang, H. Zhi, J. Wang, Y. Liu, Y. Gao, M. Guo, M. Yue, L. Wang, X. Li, Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers, Nucleic Acids Res. 44 (2016) D980–D985, https://doi.org/10.1093/nar/gkv1094.

[39] Y. Li, C. Qiu, J. Tu, B. Geng, J. Yang, T. Jiang, Q. Cui, HMDD v2, A database for experimentally supported human microRNA and disease associations, Nucleic Acids Res. 42 (2014) 1070–1074, https://doi.org/10.1093/nar/gkt1023.

[40] B. Xie, Q. Ding, H. Han, D. Wu, MiRCancer: a microRNA-cancer association database constructed by text mining on literature, Bioinformatics 29 (2013) 638–644, https://doi.org/10.1093/bioinformatics/btt014.

[41] Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, M. Li, G. Wang, Y. Liu, miR2Disease: a manually curated database for microRNA deregulation in human disease, Nucleic Acids Res. 37 (2009) 98–104, https://doi.org/10.1093/nar/gkn714.

[42] Z.H. You, Z.A. Huang, Z. Zhu, G.Y. Yan, Z.W. Li, Z. Wen, X. Chen, PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction, PLoS Comput. Biol. 13 (2017), https://doi.org/10.1371/journal.pcbi.1005455.

[43] J.-H. Li, S. Liu, H. Zhou, L.-H. Qu, J.-H. Yang, starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Res. 42 (2014) D92–D97, https://doi.org/10.1093/nar/gkt1248.

[44] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, (2013) 1–12. doi:10.1162/153244303322533223.

[45] A. Mnih, G.E. Hinton, A scalable hierarchical distributed language model, Adv. Neural Inf. Process. Syst. (2008) 1–8 doi:10.1.1.205.5467.

[46] C. Yang, L. Yang, M. Zhou, H. Xie, C. Zhang, M.D. Wang, H. Zhu, LncADeep : an ab initio lncRNA identification and functional annotation tool based on deep learning, Bioinformatics 34 (2018) 3825–3834, https://doi.org/10.1093/bioinformatics/bty428.