Computers in Biology and Medicine 153 (2023) 106527

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

# Data resources and computational methods for lncRNA-disease association prediction

Nan Sheng [a], Lan Huang [a,*], Yuting Lu [b], Hao Wang [c], Lili Yang [a,d], Ling Gao [a], Xuping Xie [a], Yuan Fu [e], Yan Wang [a,b,**]

[a] *Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, China*
[b] *School of Artificial Intelligence, Jilin University, Changchun, China*
[c] *Department of Hepatopancreatobiliary Surgery, Second Affiliated Hospital of Harbin Medical University, Harbin, China*
[d] *Department of Obstetrics, The First Hospital of Jilin University, Changchun, China*
[e] *Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, Ceredigion, United Kingdom*

ARTICLE INFO

*Keywords:*
Long non-coding RNAs
LncRNA-disease association prediction
Data resources
Computational methods

ABSTRACT

Increasing interest has been attracted in deciphering the potential disease pathogenesis through lncRNA-disease association (LDA) prediction, regarding to the diverse functional roles of lncRNAs in genome regulation. Whilst, computational models and algorithms benefit systematic biology research, even facilitate the classical biological experimental procedures. In this review, we introduce representative diseases associated with lncRNAs, such as cancers, cardiovascular diseases, and neurological diseases. Current publicly available resources related to lncRNAs and diseases have also been included. Furthermore, all of the 64 computational methods for LDA prediction have been divided into 5 groups, including machine learning-based methods, network propagation-based methods, matrix factorization- and completion-based methods, deep learning-based methods, and graph neural network-based methods. The common evaluation methods and metrics in LDA prediction have also been discussed. Finally, the challenges and future trends in LDA prediction have been discussed. Recent advances in LDA prediction approaches have been summarized in the GitHub repository at https://github.com/sheng-n/lncRNA-disease-methods.

## 1. Introduction

Molecular biology has shown that RNA can be divided into messenger RNAs, which have the ability to encode proteins, and non-coding RNAs (ncRNAs), which does not have the ability to encode proteins [1,2]. NcRNAs, especially long non-coding RNAs (lncRNAs) (a type of RNA longer than 200 nucleotides), play an important role in numerous life activities, such as transcription, translation, epigenetic regulation, splicing, differentiation, immune responses, and cell cycle control [3,4]. Recent progress suggests that the involvement of lncRNAs in variety of human diseases [2,5,6]. Therefore, exploring the complex relationship between lncRNAs and diseases will serve to understand the disease pathogenesis better, and it also benefits the development of lncRNA-based pharmacology.

LncRNAs impact almost every aspect of gene expression, from DNA transcription to mRNA splicing, RNA decay and translation. Many lncRNAs have been shown to promote cancer development and progression, for instance, lncRNA H19 expression disorder linked with different types of cancers, such as lung cancer, breast cancer, and gastric cancer [7]. LncRNA H19 can induce protein LIN28B to accelerate the proliferation of lung cancer cells through sponge miR-196b, which leads to a worse progression [8]. Plasma H19, in particular, has the potential to serve as a breast cancer biomarker [9]. LncRNA H19 and related miR-675 act as oncogenes, promoting proliferation and inhibiting apoptosis in gastric cancer [10]. Furthermore, H19 involves in several other human diseases, like male infertility [11] and Silver-Russell syndrome [12]. LncRNA ANRIL and lncRNA UFC1 have been shown to be oncogenes in non-small cell lung cancer (NSCLC) [13,14]. Gupta et al.

* Corresponding author.
** Corresponding author. Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, China.
*E-mail addresses:* huanglan@jlu.edu.cn (L. Huang), wy6868@jlu.edu.cn (Y. Wang).

https://doi.org/10.1016/j.compbiomed.2022.106527
Received 20 September 2022; Received in revised form 8 December 2022; Accepted 31 December 2022
Available online 2 January 2023

analyzed the oncogenic roles of lncRNA ANRIL and lncRNA UFC1 through the Boolean network and suggested that these lncRNAs may be involved in cell fate decisions such as senescence and apoptosis in NSCLC through the miR-34a/Myc axis [15]. In addition, various lncRNAs, such as UCA1 [16], HOTAIR [17] and MALAT1 [18], have been verified to be associated with human prostate cancer.

With the development of sequencing technologies, more lncRNAs have been annotated in human genome. Meanwhile, biological experiments, on the other hand, have discovered that some lncRNAs can be employed as potential biomarkers in clinical diagnosis, treatment, and prevention. However, traditional biological experiments, such as reverse transcription-quantitative polymerase chain reaction (RT-qPCR), RNA pull down, RNA-RIP and other techniques can never provide a panorama of the transcriptomics within living cells, which hinder the steps toward pathogenesis studies of human diseases. Thanks to the accumulation of high-throughput data, researchers have developed many public biomedical databases and bioinformatics tools for studying biomedical problems at the molecular level. The computational approach based on integrating multi-source data for LDA prediction helps to identify promising disease-related lncRNAs. More importantly, biological experiments can benefit from these studies. For example, Li et al. [19] predicted the breast cancer-associated lncRNA MNX1-AS1 and confirmed it by Li et al. [20] in 2020. This article showed that MNX1-AS1 could promote the development of triple-negative breast cancer via enhancing phosphorylation of Stat3. Xie et al. predicted a correlation between the lncRNA PTENP1 and cervical cancer [21]. Later, Fan et al. demonstrated that TENP1 prevented cervical cancer progression by competitively binding to miR-106b [22].

In Summary, computational model-based LDA prediction approaches can efficiently and systematically identify disease-related lncRNAs. As shown in Fig. 1, the workflow of the computational LDA prediction method is demonstrated. Firstly, researchers need to collect available data from public biomedical data sources related to lncRNAs and diseases, and these multi-source data provide useful information for predicting LDAs. Secondly, benefiting from the advances of computer science, various models have been developed to predict potential LDAs. Thirdly, to verify and evaluate the prediction performance of the proposed model, multiple evaluation methods and metrics are employed in the same dataset to compare it with the available state-of-the-art methods. Finally, biomedical experiments are used by biologists to validate the disease-associated candidate lncRNAs, which are predicted by the computational scientists. However, in this paper, the relevant biological experimental verification is not the focus of our investigation.

Benefiting from the collection and storage of lncRNA- and disease-related data, LDA prediction methods based on computational models have been widely presented. Chen et al. presented two reviews of computational methods for LDA prediction. Among them [23], was published in 2017 and reviewed the existing lncRNA knowledge databases and LDA prediction methods. The other review by Chen et al. focuses on describing lncRNA function prediction models and lncRNA functional similarity calculations [24]. In recent years, however, a growing number of computational models have been developed, particularly computational methods based on deep learning and graph neural network. Some old lncRNA- and disease-related databases have been discontinued for access. Therefore, there is an urgent need to present a new review to summarize and outline the computational methods and the data resources for LDA prediction. Compared with previous reviews, this paper provides comprehensive review of lncRNA- and disease-related data resources, as well as available computational methods to discover potential LDAs.

In this review, we present a comprehensive review of computational methods and data resources used for LDA prediction, as well as a systematic taxonomy of them. In conclusion, our contributions are as follows:

(1) LncRNA- and disease-related data resources for LDA prediction are collected and presented, covering LDA data, lncRNA attribute information data (expression profiles, sequences), lncRNA interaction information data (genes, miRNAs, proteins), disease attribute information data (semantics, phenotypes), disease interaction information data (genes, miRNAs). These databases are still updated and accessible as of today, and they are widely utilized in computational methods.

(2) We surveyed practically all of the known computational methods for LDA prediction. These methods are systematically classified and divided into 5 categories, including machine learning, network propagation, matrix factorization and completion, deep learning, and graph neural networks. In particular, a detailed

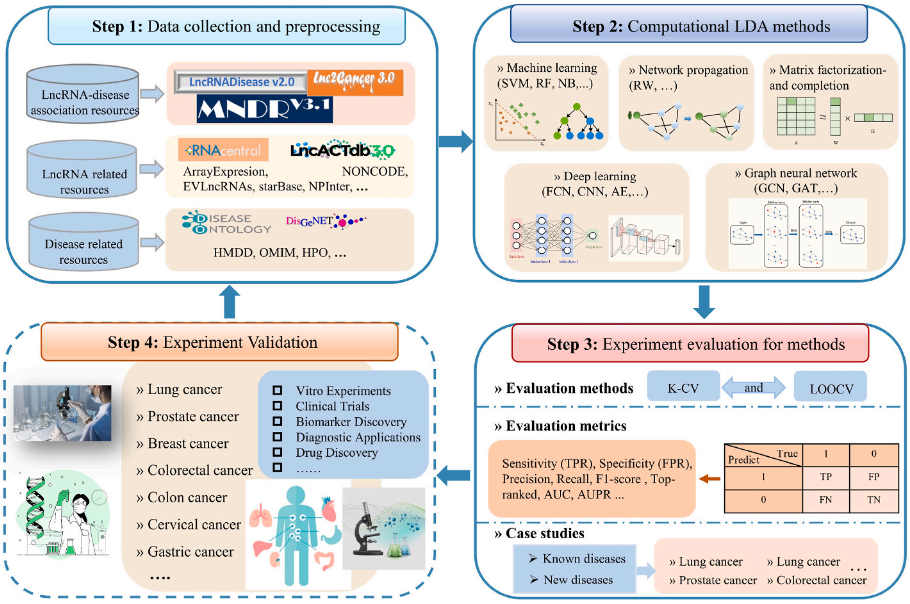

**Fig. 1.** The workflow of computational lncRNA-disease associations. Step 1: Collection and processing of available data related to lncRNAs and diseases from various public biomedical data sources. Step 2: Development of computational methods to predict LDAs. Step 3: Evaluate the prediction performance of computational methods, and obtain the candidate LDAs. Step 4: Validation of the candidate LDAs by using wet-lab experiments.

overview of deep learning- and graph neural network-based methods are presented.

(3) The evaluation methods and metrics commonly used in LDA prediction are investigated. These evaluation metrics are frequently employed and can guide researchers in future studies to effectively evaluate and validate the prediction capability of their proposed methods.

(4) Relevant databases and the latest computational methods for LDA prediction are summarized in the GitHub repository.

The review is organized as follows. Section 2, describes representative human disease-associated lncRNAs. Section 3, outlines the lncRNA- and disease-related data resources. Section 4, summarizes and provides an overview of 5 classes of classical models for LDA prediction. Section 5, presents evaluation methods and metrics for assessing LDA prediction performance. Section 6, discusses the current challenges and future trends.

## 2. LncRNAs related to diseases

The expression or function abnormality of lncRNA is closely related to disease development and progression [25]. Three common types of diseases and their related lncRNAs are described below, including cancers, cardiovascular diseases, and neurological diseases.

### *2.1. LncRNA and cancer*

According to the latest data from International Agency for Research on Cancer (IARC), breast cancer is the most common cancer in the world [26]. Many lncRNAs express abnormally in the cancer tissue. For example, the highly expressed lncRNA BANCR is significantly associated with worse prognosis [27]. LncRNA GAS5, on the contrary, a decreased expression can induce EMT which promotes lung metastasis of breast cancer, suggesting its potential of being a therapeutic target in breast cancer [28]. It has been demonstrated that lncRNA UCA1 exerts oncogenic activity in lung cancer, and promotes lung cancer cell proliferation and migration by sponging miR-193a to upregulate HMGB1 expression [29]. The lncRNA TINCR, conversely, can regulate miR-544a/FBXW7 axis to suppress proliferation and invasion in cancer, indicating its therapeutic value of lung cancer therapy [30]. There are also other lncRNAs widely involved in cancer. For example, lncRNA MALAT1 has been found overexpressed in a variety of cancers, including breast cancer, liver cancer, kidney cancer, gastric cancer, and lung cancer [31]. Recent research has revealed that the lncRNA HOTAIR can facilitate protein-protein interactions that affect multiple pathways of cancer, whilst promote tumor growth by regulating the expression and function of miRNAs [32].

### *2.2. LncRNA and cardiovascular disease*

The prevalence of cardiovascular disease continues to increase worldwide [33]. LncRNAs play a vital role in the development of pathological processes of cardiovascular pathogenesis, such as myocardial infarction (MI), heart failure (HF) and cardiac hypertrophy (CH), and have a considerable impact on the prognosis and survival of patients [34]. Studies have shown that lncRNA XIST can promote MI by targeting miR-130a-3p [35]. According to analysis of mRNA, miRNA, and lncRNA expression in cardiomyocytes, Lee et al. discovered that lncRNA expression profiles were more sensitive to different etiology of HF [36]. Kumarswamy et al. studied circulating lncRNAs in plasma of HF patients and found that mitochondrial lncRNA was down-regulated in early MI, but up-regulated in later stages [37]. Wo et al. found that lncRNA CHRF expression was up-regulated and miR-93 expression was down-regulated whilst analyzing mice and cellular models of CH [38]. CHRF promotes cardiac hypertrophy by regulating Akt3 via miR-93, therefore has been considered as a potential therapeutic target. Likely, lncRNA MALAT1 and COL1A1 have been shown to be linked to many cardiovascular diseases [39,40].

### *2.3. LncRNA and neurological disease*

Environmental exposures and genetic disorders has been linked to several neurodegenerative diseases [41]. Double et al. proposed that protein aggregation, oxidative stress, inappropriate inflammatory responses and increased apoptosis contribute to neurological diseases [42]. LncRNAs played important role in neurological diseases, such as Alzheimer’s Disease (AD), Parkinson’s disease (PD), and Huntington’s disease (HD) [43]. AD causes the degeneration of brain cells and currently affects approximately 50 million people worldwide [44]. Studies have shown that lncRNA BACE1-AS and MALAT1 are closely related to AD [45,46]. Plasma BACE1-ASS levels were found to be significantly higher in AD patients than the control people, implying that BACE1-AS could be used as a blood biomarker for AD [45]. PD is the second most prevalent neurodegenerative disease [47]. Researchers have found that lncRNA SNHG1, UCA1 and MEG3 were significantly upregulated in PD patients than the health control people, and showed that lncRNA could be the potential of biomarkers of PD [48–50]. In addition, lncRNA NEAT1 and MEG3 were associate with HD [51,52].

## 3. Data resources

With the development of high-throughput sequencing experimental technologies, an amount of genomics, transcriptomics, proteomics, and metabolomics data has been generated. Researchers started collecting and integrating these data and storing them in databases, which are publicly available for further study and analysis. To some extent, the emergence of public databases has facilitated computational scientists in downloading data and developing computational methods. Firstly, this section reviews LDA data resources, which supplies important target label dataset for the LDA prediction methods. The performance of LDA prediction can be improved by combining multi-source data. Then, we present commonly utilized lncRNA- and disease-related data resources, which provide crucial support for computational approaches. We not only provide a brief description of the latest versions of the databases, but also list corresponding web links. In particular, we manually removed some databases that appeared in the computational approach but are no longer accessible. As shown in Table 1, we divided these databases into five categories: LDA data resources, lncRNA attribute information data resources, lncRNA interaction information data resources, disease attribute information data resources, disease interaction information data resources.

### *3.1. LncRNA-disease association data resources*

#### *3.1.1. LncRNADisease [53] (http://www.rnanut.net/lncrnadisease/)*

Chen et al. developed the LncRNADisease database that collected and curated 480 experimentally validated LDAs, involving 321 lncRNAs and 166 diseases [54]. Recently, Bao et al. have upgraded the database to LncRNADisease v2.0, which contains experimentally and computationally supported data. Among them, the experimentally supported data are obtained from relevant publications by searching the PubMed using the keywords. The researchers next used manual literature curation to obtain 10,564 experimentally verified associations between 19, 166 lncRNAs and 529 diseases. In addition, LncRNADisease v2.0 additionally recorded 1004 circRNA-disease associations including 823 circRNAs. Specifically, the database also provides computationally supported associations based on computational algorithms, such as LRLSLDA [55], LDAP [56], and RWRlncD [57]. Furthermore, it provides a user-friendly web interface that allows users to browse, search or download experimental (prediction) supported LDAs data and lncRNA interaction data.

**Table 1**
lncRNA- and disease-related data resources.

| LncRNA-disease association data resources | | | |
|---|---|---|---|
| Databases | Latest version | Description | URL |
| LncRNADisease [53] | LncRNADisease v2.0 | Documents 19,166 lncRNAs, 529 diseases, and 10,564 association in Homo sapiens, Mus musculus, Rattus norvegicus and Gallus gallus | http://www.rnanut.net/lncrnadisease/ |
| Lnc2Cancer [58] | Lnc2Cancer v3.0 | Collects 2659 lncRNAs, 216 cancers and 9254 associations in human | http://bio-bigdata.hrbmu.edu.cn/lnc2cancer/ |
| MNDR [60] | MNDR v3.1 | Records 39,880 lncRNA, over 1600 diseases and 295,834 associations in 11 mammals | http://bio-bigdata.hrbmu.edu.cn/lnc2cancer/ |
| **LncRNA attribute information data resources** | | | |
| **Databases** | **Latest version** | **Description** | **URL** |
| NONCODE [61] | NONCODE v6.0 | Records the comprehensive knowledge database of ncRNA from 39 species | http://www.noncode.org/ |
| RNAcentral [62] | RNAcentral v19 | Documents ncRNA sequences for 296 species | https://rnacentral.org/ |
| EVLncRNAs [63] | EVLncRNAs v2.0 | Collects sequence, structure, function and phenotype information of experimentally validated lncRNAs | https://www.sdklab-biophysics-dzu.net/EVLncRNAs2/ |
| NPInter [64] | NPInter v4.0 | Documents the regulatory interactions between ncRNAs and other biomolecules | http://bigdata.ibp.ac.cn/npinter4/ |
| LncACTdb [65] | LncACTdb v3.0 | Records experimentally supported ceRNA interactions and comprehensive annotations | http://bio-bigdata.hrbmu.edu.cn/LncACTdb/ |
| LNCipedia [66] | LNCipedia v5.2 | Collects annotations and sequence information for 1555 human lncRNAs | https://lncipedia.org |
| lncRNAWiki [70] | – | Collects 105,255 non-redundant lncRNA transcripts information from ENCODE, NONCODE, and LNCippedia | http://lncrna.big.ac.cn |
| lncRNome [74] | – | Records information on over 17,000 human lncRNAs, and provides experimental and predicted lncRNA-protein interactions | http://genome.igib.res.in/lncRNome |
| LncExpDB [75] | – | Contains 101,293 high quality human lncRNA gene expression profiles from 1977 samples of 337 biological conditions across 9 biological contexts | https://ngdc.cncb.ac.cn/lncexpdb |
| **LncRNA interaction information data resources** | | | |
| **Databases** | **Latest version** | **Description** | **URL** |
| starBase [76] | ENCORI | Records lncRNA-miRNA, miRNA-mRNA, ncRNA-RNA interactions | https://starbase.sysu.edu.cn/ |
| LncRNA2target [77] | LncRNA2Target v3.0 | Collects experimentally supported lncRNA and target relationships | http://www.bio-annotation.cn/lncrna2target/index.jsp |
| RAID [78] | RAID v2.0 | Records 40,668 lncRNA-associated RNA-protein interactions and 34,790 lncRNA-associated RNA-RNA interactions | https://www.rna-society.org/raid2 |
| **Disease attribute information data resources** | | | |
| **Databases** | **Latest version** | **Description** | **URL** |
| Disease Ontology [79] | – | Provides a human-readable and machine-interpretable diseases corpus through a common language | https://disease-ontology.org/ |
| HPO [80] | – | A comprehensive logical standard for describing and computationally analyzing phenotypic abnormalities found in human diseases | https://hpo.jax.org/app/ |
| OMIM [81] | – | Describes genes with known sequence and phenotypes | http://www.ncbi.nlm.nih.gov/omim |
| **Disease interaction information data resources** | | | |
| **Databases** | **Latest version** | **Description** | **URL** |
| DisGeNet [82] | DisGeNet v7.0 | Collects 30,170 diseases, 21,671 genes, and 1,124,942 associations | https://www.disgenet.org/ |
| HMDD [83] | HMDD v3.2 | Records 893 diseases, 1206 miRNAs, and 35,547 associations in human | http://www.cuilab.cn/hmdd |

### *3.1.2. Lnc2Cancer [58] (http://bio-bigdata.hrbmu.edu.cn/lnc2cancer/)*

The Lnc2Cancer, proposed by Ning et al., is a manually curated database of lncRNA-cancer associations [59]. All cancer-related lncRNAs in it are experimentally supported and manually curated from the published literature. Currently, the Lnc2Cancer database has been upgraded to Lnc2Cancer v3.0. The current version records human 9254 lncRNA-cancer associations and 1049 circRNA-cancer associations that are experimentally supported and manually curated from >15,000 relevant publications, involving 2659 lncRNAs, 216 cancers, 743 circRNAs. Lnc2Cancer v3.0 provides a flexible data access pathway that allows users to browse, query, and download all experimentally supported associations. In particular, the single cell and RNA-seq web tools are also available in the current database.

### *3.1.3. MNDR [60] (http://www.rna-society.org/mndr/)*

MNDR is a database that focuses on relationship between ncRNA and disease in mammals. The database integrates experimentally supported and predicted ncRNA-disease associations that are manually curated from the published literature and other resources. MNDR has now been updated to MNDR v3.1, removing some controversial literature. MNDR v3.1 documents 5 types of ncRNAs-disease association, including 393,651 associations between 6301 miRNAs and diseases, 295,834 associations between 39,880 lncRNAs and diseases, 300,630 associations between 20,256 circRNAs and diseases, 13,624 associations between 10,894 piRNAs and diseases, and 1573 associations between 521 snoRNAs and diseases. In addition, MNDR v3.1 covered 11 species, including Homo sapiens, Mus musculus, Rattus norvegicus, Oryctolagus cuniculus, Sus scrofa, Macaca mulatta, Canis lupus familiarizes, Pan troglodytes, Calli trichinae, Ovis aries, Bos taurus. Similarly, search, query, and download functions have all been designed.

### *3.2. LncRNA attribute information data resources*

#### *3.2.1. NONCODE [61] (http://www.noncode.org/)*

NONCODE is a comprehensive repository of ncRNAs, especially lncRNAs, which integrates resources from related publications and other public databases. Currently, the database has been updated to NONCODE v6.0, and it contains information on 39 species, involving 16 animals and 23 plants. It provides basic information, including the location, strand, exon number, length, and sequence of lncRNAs, as well as advanced information, including the expression profile, exosome expression profile, conservation information, predicted function, and disease relation. The database is frequently used to gain the sequence and expression information.

#### *3.2.2. RNAcentral [62] (https://rnacentral.org/)*

RNAcentral is a database for collating ncRNA sequence information of 296 species. With 19 versions updated, the RNAcentral database has become a significant part of the RNA research process and a key source of extracted sequence data for academic and commercial groups.

#### *3.2.3. EVLncRNAs [63] (https://www.sdklab-biophysics-dzu.net/EVLncRNAs2/)*

EVLncRNAs is a comprehensive database of experimentally validated functional lncRNAs, that collects sequence, structure, function, and phenotypic information of experimentally validated lncRNAs from all species. EVLncRNAs has been updated to EVLncRNAs v2.0 with relevant data manually extracted from nearly 19,000 published literature. Specifically, EVLncRNAs v2.0 contains 4010 lncRNAs from 124 species.

#### *3.2.4. NPInter [64] (http://bigdata.ibp.ac.cn/npinter4/)*

NPInter documents the regulatory interactions between ncRNAs and other biomolecules in 35 organisms, and which are primarily collected and validated by dual manual literature mining and high-throughput sequencing data processing. NPInter has been updated to NPInter v4.0, which contains most other types of ncRNAs, such as lncRNAs, miRNAs, snoRNAs, snRNAs, and circRNAs. This database is frequently used to obtain lncRNA- and miRNA-interaction data.

#### *3.2.5. LncACTdb [65] (http://bio-bigdata.hrbmu.edu.cn/LncACTdb/)*

LncACTdb is a database for collecting, storing, and analyzing experimentally supported ceRNA interactions and comprehensive annotations. LncACTdb has been updated to LncACTdb v3.0, which collects published literature through the PubMed database, and manually curates experimentally supported ceRNA interactions. LncACTdb v3.0 has been expanded to 25 species and 537 diseases/phenotypes, containing 913 lncRNAs, 1723 mRNAs, 337 circRNAs, and 19 pseudogene interactions.

#### *3.2.6. LNCipedia [66] (https://lncipedia.org)*

LNCipedia is a public database of lncRNA sequences and annotations. The current version LNCipedia v5.2 contains annotations and sequence information for 1555 human lncRNAs from 2482 lncRNA publications. These lncRNA annotations are from Ensembl [67], RefSeq [68], and FANTOM CAT [69]. The database provides a download function which contains various exports in BED, GFF, GTF and FASTA formats.

#### *3.2.7. lncRNAWiki [70] (http://lncrna.big.ac.cn)*

The lncRNAWiki is a community-managed and collected human lncRNA knowledgebase. It integrates human lncRNAs information from multiple different sources. Currently, lncRNAWiki integrates lncRNA sequence and annotation information from three data sources, including GENCODE [71] (version 19; 23,898 human lncRNA transcripts), NONCODE [72] (version 4.0; 95,135 human lncRNA transcripts) and LNCippedia [73] (version 2.1; 32,181 human lncRNA transcripts). Finally, 105,255 non-redundant lncRNA transcripts were obtained by using blastn across these three data sources. Specifically, it not only allowed different users to edit, update and manage existing lncRNAs, but also allowed any user to add newly identified lncRNAs.

#### *3.2.8. lncRNome [74] (http://genome.igib.res.in/lncRNome)*

The lncRNome is a comprehensive lncRNAs knowledgebase. This database records information on over 17,000 lncRNAs in humans. lncRNome provides information on the types, chromosomal locations, biological function descriptions and disease associations of lncRNAs. In addition, the database provides experimental and prediction datasets of RNA-protein interactions for lncRNAs.

#### *3.2.9. LncExpDB [75] (https://ngdc.cncb.ac.cn/lncexpdb)*

LncExpDB is a comprehensive database for human lncRNA expression, covering lncRNA gene expression in various biological contexts. The database contains 101,293 high quality human lncRNA gene expression profiles from 1977 samples of 337 biological conditions across 9 biological contexts. LncExpDB provides a user-friendly web interface for easy data query, browsing, visualization and access.

### *3.3. LncRNA interaction information data resources*

#### *3.3.1. starBase [76] (https://starbase.sysu.edu.cn/)*

According to CLIP-seq, degradome-seq and RNA-RNA interactions omics, the starBase records the experimentally supported interactions of miRNA-lncRNA, miRNA-mRNA, ncRNA-RNA, RNA-RNA, RBP-ncRNA and RBP-mRNA. Currently, starBase has been renamed ENCORI, and it provides molecular interactions information between 23 species. ENCORI is the most comprehensive miRNA-lncRNA interactions network supported by CLIP-Seq experiments to date. As a result, it is a widely utilized database for acquiring lncRNA-miRNA interaction information in designing LDA prediction computational models.

#### *3.3.2. LncRNA2target [77] (http://www.bio-annotation.cn/lncrna2target/index.jsp)*

LncRNAs play a very important role in the regulation of gene expression, and the online database LncRNA2target is developed to store information on lncRNA-regulated target genes. LncRNA2target has been updated to LncRNA2Target v3.0, which records experimentally supported lncRNA-target relationships through reviewing the published literature. Specifically, the database collects all differentially expressed genes after the lncRNA knockdown or overexpression. Human and mouse lncRNA-target gene associations can be employed from this database.

#### *3.3.3. RAID [78] (https://www.rna-society.org/raid2)*

RAID provides RNA-associated crosstalk, including RNA-RNA and RNA-protein interactions, through manually curating the literature and integrating other database resources. RAID has been updated to version 2.0 (RAID v2.0), and including 40,668 lncRNA-associated RNA-protein interactions and 34,790 lncRNA-associated RNA-RNA interactions.

### *3.4. Disease attribute information data resources*

#### *3.4.1. Disease Ontology [79] (https://disease-ontology.org/)*

Disease Ontology (DO) provides a human-readable and machine-interpretable diseases corpus through a common language, which semantically integrates disease and medical vocabulary terms. The DO aims to produce a clear definition for each disease in etiology-based disease classification so that they can be used and applied consistently when annotating biomedical data. The disease semantic similarity has been widely calculated using directed acyclic graphs (DAGs) of DO, which can be constructed based on disease terms.

#### *3.4.2. HPO [80] (https://hpo.jax.org/app/)*

The Human Phenotype Ontology (HPO) aims to provide a comprehensive logical standard for describing and computationally analyzing phenotypic abnormalities in human diseases. Currently, there are over 13,000 terms and 156,000 annotations on genetic disorders in the HPO. The HPO contains multiple types of phenotypic information, and phenotypic terms can be used to construct DAGs to represent phenotypic features and relationships. A number of computational measures of disease phenotypic similarity have been proposed and are widely used to improve the performance of LDAs.

#### *3.4.3. OMIM [81] (http://www.ncbi.nlm.nih.gov/omim)*

OMIM is a knowledge of human genes and genetic diseases, developed to support human genetics research, education and clinical genetics practice. OMIM updates the data monthly, and this genetic information can help researchers understand the occurrence of disease at the molecular level.

### *3.5. Disease interaction information data resources*

#### *3.5.1. DisGeNet [82] (https://www.disgenet.org/)*

To explore the issue of pairwise genomic variation, DisGeNet database integrates disease-associated gene and variant data from the scientific literature and other databases. DisGeNet has been upgraded to DisGeNet v7.0, which contains 1,135,045 gene-disease associations between 21,671 genes and 30,170 diseases, and 369,554 variant-disease associations covering 194,515 variants and 14,155 diseases. DisGeNet is the primary source database for disease-gene associations when designing the LDAs prediction approach.

#### *3.5.2. HMDD [83] (http://www.cuilab.cn/hmdd)*

MiRNAs are an important type of regulatory RNAs that play a vital role in a variety of biological processes, and miRNAs dysfunction is related to many diseases. HMDD is a widely used database that records experimentally supported human miRNA-disease associations. HMDD has now been launched to HMDD v3.2, which manually curated 35,547 experimentally supported miRNA-disease associations, containing 1206 miRNAs and 893 diseases from 19,280 publications.

## 4. Computational methods for lncRNA-disease association prediction

As traditional biological experiments are time-consuming and expensive, computational methods that integrate different prior information can identify potential LDAs faster and lower cost. To achieve this goal, many LDA prediction methods have been developed in recent years to assist in screening disease-associated lncRNA markers. As shown in Fig. 2, we introduced almost all existing computational approaches for predicting LDA published from 2013 to 2022. These models are classified into 5 categories: machine learning-based methods (see Table 2), network propagation-based methods (see Table 3), matrix factorization- and completion-based methods (see Table 4), deep learning-based methods (see Table 5), graph neural network-based methods (see Table 6).

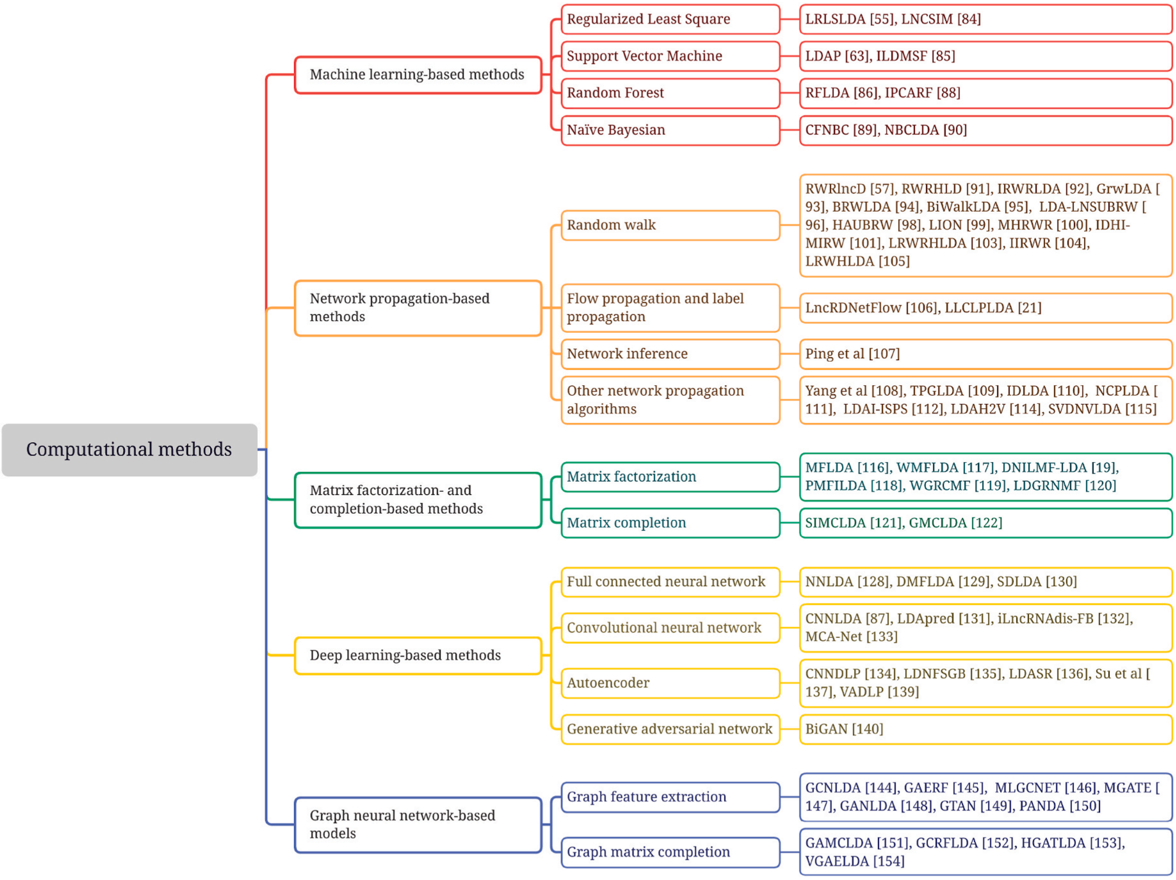
Computational methods
Machine learning-based methods
Regularized Least Square
LRLSLDA [55], LNCSIM [84]
Support Vector Machine
LDAP [63], ILDMSF [85]
Random Forest
RFLDA [86], IPCARF [88]
Naïve Bayesian
CFNBC [89], NBCLDA [90]
Network propagation-based methods
Random walk
RWRlncD [57], RWRHLD [91], IRWRLDA [92], GrwLDA [93], BRWLDA [94], BiWalkLDA [95], LDA-LNSUBRW [96], HAUBRW [98], LION [99], MHRWR [100], IDHI-MIRW [101], LRWRHLDA [103], IIRWR [104], LRWHLDA [105]
Flow propagation and label propagation
LncRDNetFlow [106], LLCLPLDA [21]
Network inference
Ping et al [107]
Other network propagation algorithms
Yang et al [108], TPGLDA [109], IDLDA [110], NCPLDA [111], LDAI-ISPS [112], LDAH2V [114], SVDNVLDA [115]
Matrix factorization- and completion-based methods
Matrix factorization
MFLDA [116], WMFLDA [117], DNILMF-LDA [19], PMFILDA [118], WGRCMF [119], LDGRNMF [120]
Matrix completion
SIMCLDA [121], GMCLDA [122]
Deep learning-based methods
Full connected neural network
NNLDA [128], DMFLDA [129], SDLDA [130]
Convolutional neural network
CNNLDA [87], LDApred [131], iLncRNAdis-FB [132], MCA-Net [133]
Autoencoder
CNNDLP [134], LDNFSGB [135], LDASR [136], Su et al [137], VADLP [139]
Generative adversarial network
BiGAN [140]
Graph neural network-based models
Graph feature extraction
GCNLDA [144], GAERF [145], MLGCNET [146], MGATE [147], GANLDA [148], GTAN [149], PANDA [150]
Graph matrix completion
GAMCLDA [151], GCRFLDA [152], HGATLDA [153], VGAELDA [154]

**Fig. 2.** Taxonomy and representation of computational methods for LDA prediction, and grouped them based on their underlying computational models consisting of machine learning, network propagation, matrix factorization and completion, deep learning, and graph neural network.

**Table 2**
Machine learning-based computational methods for predicting disease-related lncRNAs.

| Method | Description | Experiment evaluation | Source code |
|---|---|---|---|
| LRLSLDA [55] | A first computational method based on Regularized Least Squares to predict LDAs | LOOCV (AUC) | Unavailable |
| LNCSIM [84] | Developing two novel lncRNA functional similarity Laplacian calculation models to improve LRLSLDA | LOOCV (AUC) | Unavailable |
| LDAP [63] | A computational method that integrates multiple biological data and uses bagging SVM classifier | LOOCV (AUC) | Unavailable |
| ILDMSF [85] | A framework based on multi-similarity fusion and the support vector machine | LOOCV (AUC top-r) | Unavailable |
| RFLDA [86] | A prediction method that implements random forest and feature selection | 5-cv (AUC, AUPR) | https://github.com/ydkvictory/RFLDA |
| IPCARF [88] | A method that combines incremental principal component analysis and random forest model to predict | 10-cv (AUC) | https://github.com/zhurong1942/IPCARF_zr1 |
| CFNBC [89] | An approach based on Naïve Bayesian classifier and collaborative filtering model with an updated tripartite network | LOOCV (AUC) | https://github.com/jingwenyu18/CFNBC |
| NBCLDA [90] | A method that constructs global quadruple network and naïve Bayesian classifier | LOOCV (AUC, AUPR) | Unavailable |

**Table 3**
Network propagation-based computational methods for predicting disease-related lncRNAs.

| Method | Description | Experiment evaluation | Source code |
|---|---|---|---|
| RWRlncD [57] | A global network method based on random walk with restart of a lncRNA functional similarity network | LOOCV (AUC) | Unavailable |
| RWRHLD [91] | A rank-based method that performs random walk with restart on heterogeneous network | LOOCV (AUC) | Unavailable |
| IRWRLDA [92] | A model of improving random walk with restart that used lncRNA expression similarity and disease semantic similarity to set the initial probability vector | LOOCV (AUC) | Unavailable |
| GrwLDA [93] | A global network based on random walk model | LOOCV (AUC, AUPR) | Unavailable |
| BRWLDA [94] | A model that incorporates multiple heterogeneous data and performs bi-random walks | LOOCV (AUC) | Unavailable |
| BiWalkLDA [95] | A novel method that uses bi-random walks and solving cold-start | LOOCV (AUC, top-r) | https://github.com/screamer/BiwalkLDA |
| LDA-LNSUBRW [96] | An approach based on linear neighborhood similarity and unbalanced bi-random walk | LOOCV, 5-cv (AUC) | https://github.com/JIAWEI1234/LDA-LNSUBRW |
| HAUBRW [98] | A method based on hybrid algorithm and unbalanced bi-random walk | LOOCV, 5-cv (AUC, AUPR, Acc, Pre, Sen, F1, Mcc) | Unavailable |
| LION [99] | An approach using network diffusion within the multi-level network | Randomly shuffling node labels (AUC) | Unavailable |
| MHRWR [100] | A prediction approach using random walk with restart on multi-layer network | LOOCV (AUC, top-r) | https://github.com/yangyq505/MHRWR |
| IDHI-MIRW [101] | A model that integrates diverse heterogeneous information sources with random walk with restart algorithm | LOOCV (AUC) | https://github.com/NWPU-903PR/IDHI-MIRW |
| LRWRHLDA [103] | A global network computational framework using Laplace normalized random walk with restart algorithm on heterogeneous network | 10-cv (AUC, AUPR) | https://github.com/wang-124/LRWRHLDA |
| IIRWR [104] | A prediction model based on internal inclined random walk with restart | LOOCV, 5-cv, 10-cv (AUC) | Unavailable |
| LRWHLDA [105] | A framework based on improving local random walks | LOOCV, 2-cv, 5-cv (AUC) | Unavailable |
| LncRDNetFlow [106] | A global network-based method using flow propagation | LOOCV, 5-cv (AUC) | Unavailable |

(continued on next page)

### 4.1. Machine learning-based methods

Recently, machine learning techniques have rapidly progressed, providing powerful computational methods for LDA prediction. These methods aim to define the LDA prediction problem as a binary classification task, with lncRNA and disease as instances, and the information of lncRNA and disease as features. Known and experimentally validated LDAs are treated as positive samples, whereas unknown and unexperimentally validated LDAs are considered as negative samples. Then, classical machine learning classification models, such as regularized least squares (RLS), support vector machine (SVM), random forest (RF), and Naive Bayesian (NB) are applied to the LDA prediction methods, as shown in Table 2.

#### 4.1.1. Regularized least square

Chen et al. proposed a semi-supervised learning method LRLSLDA, based on Laplacian regularized least square (LRLS) [55]. To our knowledge, LRLSLDA is the first computational method that has been developed for predicting LDAs. Benefited that the LncRNADisease database, which was established in 2012, collected experimentally validated LDA data from publications. This method calculated lncRNA expression profile similarity, as well as lncRNA and disease Gaussian Interaction Profile Kernel (GIPK) similarities by using lncRNA expression profiles and LDAs. Based on the assumption that similar diseases tend to be associated with similar lncRNAs, LRLS was utilized to predict LDAs. Although LRLSLDA cannot achieve significant prediction performance, more importantly, the method provided valuable experience for future research of lncRNAs and diseases. Based on the assumption that functionally similar lncRNA tend to be associated with similar diseases, and vice versa. Chen et al. proposed two new methods LNCSIM for calculating lncRNA functional similarity [84]. By quantitatively measuring the similarity of two lncRNA-related diseases, the functional similarity between lncRNA and lncRNA can be calculated. The disease

**Table 3** (*continued*)

| Method | Description | Experiment evaluation | Source code |
|---|---|---|---|
| | algorithm to integrate multiple networks | | |
| LLCLPLDA [21] | A method using locality-constrained linear coding and label propagation | LOOCV, 5-cv (AUC) | Unavailable |
| Ping et al. [107] | A novel model that calculates the first- and second-order similarities of lncRNAs and diseases based lncRNA-disease bipartite network | LOOCV (AUC) | Unavailable |
| Yang et al. [108] | A propagation algorithm on coding-non-coding genes-disease bipartite network | LOOCV (AUC) | Unavailable |
| TPGLDA [109] | A novel approach via resource allocation algorithm on lncRNA-disease-gene tripartite graph | LOOCV (AUC, Sn, Acc, Pre, Mcc, top-r) | https://github.com/USTC-HIlab/TPGLDA |
| IDLDA [110] | A diffusion model to calculate the information conveyed in the lncRNA-disease bipartite network | LOOCV (AUC) | Unavailable |
| NCPLDA [111] | A method based on network consistency projection | LOOCV (AUC) | https://github.com/ghli16/NCPLDA |
| LDAI-ISPS [112] | A network-based space projection scores method | LOOCV (AUC) | Unavailable |
| LDAH2V [114] | A computational framework using meta-path across multiple networks | 10-cv (AUC) | Unavailable |
| SVDNVLDA [115] | A novel model based on Singular Value Decomposition and node2vec methods | 10-cv (AUC, AUPR, Acc, Sen, Spe, Pre, Mcc) | https://github.com/iALKing/SVDNVLDA |

semantic similarity can be measured based on the disease DAGs. Then, two lncRNA functional similarities were introduced into the LRLSLDA model.

### 4.1.2. *Support vector machine*

Lan et al. developed an efficient calculation method, named LDAP [56]. LDAP measured 2 types of lncRNA similarities and 5 types of disease similarities, including lncRNA sequence similarity, lncRNA expression profile similarity, disease phenotype similarity, disease GIPK similarity, disease annotation similarity, disease topology similarity and disease-distance similarity (according protein-protein interaction and disease-gene association), which were integrated using Karcher mean. Considering the imbalance between positive and negative samples, bagging SVM was employed as classifier to predict LDAs. Similar to LDAP, Chen et al. proposed a computational framework ILDMSF that combined lncRNA and disease similarity for predicting LDAs [85]. To incorporate similarity, ILDMSF adopted the similarity network fusion algorithm and the K-nearest neighbor (KNN) algorithm. Bagging SVM was utilized to predict disease-related candidate lncRNAs.

### 4.1.3. *Random forest*

Yao et al. proposed a method (called RFLDA) based on RF and feature selection to integrate information among lncRNA, miRNA and disease [86]. The lncRNA-disease pairwise feature matrix was constructed using a method similar to CNNLDA [87]. RF was performed to predict the association scores of lncRNAs and diseases. Zhu et al. developed IPCARF, a new computational method that combined incremental principal component analysis (IPCA) and RF for predicting LDAs [88]. The disease similarity and lncRNA similarity were exploited to build the lncRNA-disease pairwise vectors. The method adopted IPCA to project the input vector into a low-dimensional space and utilized RF to predict the association probability of node pairs.

**Table 4**
Matrix factorization- and completion-based computational methods for predicting disease-related lncRNAs.

| Method | Description | Experiment evaluation | Source code |
|---|---|---|---|
| MFLDA [116] | A method using matrix tri-factorization on multiple heterogeneous data source | 5-cv (AUC) | http://mlda.swu.edu.cn/codes.php?Name=MFLDA |
| WMFLDA [117] | A model that integrates multi-relational data with a weighted matrix factorization | 5- cv (AUC, AUPR) | http://mlda.swu.edu.cn/codes.php?name=WMFLDA |
| DNILMF-LDA [19] | A framework based on dual-network integrated logistic matrix factorization | 10-cv (AUC, AUPR) | Unavailable |
| PMFILDA [118] | A computational method based on KNN algorithm and probabilistic matrix factorization | LOOCV (AUC) | Unavailable |
| WGRCMF [119] | A method based on graph regularization collaborative matrix factorization | 10-cv (AUC) | Unavailable |
| LDGRNMF [120] | A computational approach using graph regularized nonnegative matrix factorization | 5-cv (AUC) | Unavailable |
| SIMCLDA [121] | A method based on inductive matrix completion | LOOCV (AUC, top-r) | https://github.com//bioinfomaticsCSU/SIMCLDA |
| GMCLDA [122] | A method based on geometric matrix completion | LOOCV (AUC, top-r) | https://github.com/bioinfomaticsCSU/GMCLDA |

### 4.1.4. *Naive Bayesian*

Yu et al. proposed CFNBC, a new collaborative filtering model based on Naive Bayesian Classifier [89]. The model integrated LDAs, lncRNA-miRNA interactions, and miRNA-disease associations to construct a lncRNA-miRNA-disease (LMD) tripartite network. Considering that there were very few known associations and interactions among lncRNA, miRNA, disease in the tripartite network, lncRNA and disease common neighbors would be extremely useful. Collaborative filtering algorithm was utilized to update the LMD network, and improve the number of common neighbors of lncRNA and disease nodes. NB classifier was applied to predict LDAs. It was worth noting that integrating more lncRNA- and disease-related data can improve the prediction performance. Yu et al. proposed a new approach, NBCLDA, which introduced the gene-lncRNA interaction network, gene-disease association network, and gene-miRNA interaction network in the LMD tripartite network to reconstruct a lncRNA-miRNA-gene-disease global quadruple network [90]. Similarly, the LDA scores can be obtained using NB classifier.

## 4.2. *Network propagation-based methods*

Besides training a classifier to predict LDAs, the network-based approach is also widely utilized. It is dedicated to reveal potential novel relationships at the network level by constructing networks to represent the correlations between lncRNAs and diseases. Network-based methods, unlike machine learning-based methods, often do not require negative samples. In recent years, many network-based methods have been proposed, especially the random walk, which has become a

**Table 5**
Deep learning-based computational methods for predicting disease-related lncRNAs.

| Method | Description | Experiment evaluation | Source code |
|---|---|---|---|
| NNLDA [128] | A framework using deep neural networks and matrix factorization | 10-cv (top-r) | https://github.com/gao793583308/NNLDA |
| DMFLDA [129] | A deep learning framework based on deep matrix factorization model | LOOCV, 5-cv (AUC, top-r) | https://github.com/CSUBioGroup/DMFLDA |
| SDLDA [130] | A hybrid computational framework using singular value decomposition and deep learning | LOOCV (AUC, AUPR, top-r) | https://github.com/CSUBioGroup/SDLDA |
| CNNLDA [87] | A method based on dual convolutional neural networks with attention mechanisms | 5-cv (AUC, AUPR, top-r) | Unavailable |
| LDApred [131] | A method based on information flow propagation and convolutional neural networks | 5-cv (AUC, AUPR, top-r) | Unavailable |
| iLncRNAdis-FB [132] | A computational predictor based on convolution neural network to integrate different data sources using the feature block | 5-cv (AUC, AUPR) | Unavailable |
| MCA-Net [133] | A deep learning method based on multi-feature coding and attention convolutional neural network | 10-cv (AUC, Acc, Sen, Spe, Pre, Mcc) | Unavailable |
| CNNDLP [134] | A deep learning method using convolutional neural network with the attention mechanism and convolutional autoencoder | 5-cv (AUC, AUPR, top-r) | Unavailable |
| LDNFSGB [135] | An effective method using feature similarity and gradient boosting algorithm | LOOCV, 10-cv (AUC, Acc, Sen, Spe, Pre, Mcc) | https://github.com/MLMIP/LDNFSGB |
| LDASR [136] | A computational framework based on autoencoder and rotation forest | LOOCV, 5-cv (AUC, Acc, Sen, Spe, Pre, Mcc) | Unavailable |
| Su et al. [137] | A network representation learning method using stacked autoencoder, Node2vec and XGBoost | 5-cv (AUC, AUPR, Acc, Sen, Spe, Pre, MCC) | Unavailable |
| VADLP [139] | A prediction model based on convolutional and variance autoencoders | 5-cv (AUC, AUPR, top-r) | Unavailable |
| BiGAN [140] | A computational model using bidirectional generative adversarial network | 5-cv, 10-cv (AUC, AUPR, Acc, F1, Mcc) | https://github.com/TomasYang001/BiGAN-lncRNA-disease-associations-prediction.git |

**Table 6**
Graph neural network-based computational methods for predicting disease-related lncRNAs.

| Method | Description | Experiment evaluation | Source code |
|---|---|---|---|
| GCNLDA [144] | A novel method based on graph convolution and convolutional neural network | 5-cv (AUC, AUPR, top-r) | Unavailable |
| GAERF [145] | A computational method based on graph autoencoder and random forest | 5-cv (AUC, AUPR) | Unavailable |
| MLGCNET [146] | A framework using multi-layer aggregation graph convolutional network and extra trees | 5-cv (AUC, AUPR) | https://github.com/QingwWu/MLGCNET |
| MGATE [147] | A method using multi-channel graph attention autoencoder | 5-cv (AUC, AUPR, Acc, F1, Mcc, top-r) | https://github.com/sheng-n/MGATE |
| GANLDA [148] | An end-to-end computational model based on graph attention network | 10-cv (AUC, AUPR) | Unavailable |
| GTAN [149] | A novel method based on graph neural network with attribute-level attention mechanisms and multi-layer convolutional neural networks | 5-cv, 10-cv, 20-cv (AUC, AUPR, top-r) | Unavailable |
| PANDA [150] | A graph-based method using variational graph autoencoder | 90% edges for training and 10% for testing (AUC, AUPR, Acc, F1) | Unavailable |
| GAMCLDA [151] | A computational framework based on graph autoencoder matrix completion | 10-cv (AUC, AUPR, Pre, F1, top-r) | Unavailable |
| GCRFLDA [152] | A method using graph convolutional matrix completion with conditional random field and attention mechanism | 5-cv (AUC, AUPR) | https://github.com/jademyC1221/GCRFLDA |
| HGATLDA [153] | A heterogeneous graph attention network framework based on meta-paths | 5-cv (AUC, AUPR, ACC, Pre, Rec, F1) | Unavailable |
| VGAELDA [154] | An end-to-end model based on variational inference and graph autoencoders | 5-cv (AUC, AUPR, Sen, Acc, Pre, F1, Mcc, top-r) | https://github.com/zhanglabNKU/VGAELDA |

popular tool for predicting LDAs, as shown in Table 3.

*4.2.1. Random walk*

Based on the assumption that functionally related genes tend to be associated with phenotypically similar diseases, Sun et al. developed a novel method for calculating lncRNA functional similarity. The method proposed the global network-based computational model, RWRlncD, which integrated the LDA network, disease similarity network, and lncRNA functional similarity network [57]. Potential LDAs were inferred by using the random walk with restart (RWR) on the lncRNA functional similarity network. Based on the biological premise that lncRNAs with more common miRNA interactions partners were often related to similar diseases, Zhou et al. proposed RWR based computational model RWRHLD to predict LDAs, which introduced lncRNA-miRNA interaction information [91]. The model integrated three networks to construct the LDA heterogeneous network, including the miRNA-related lncRNA-lncRNA crosstalk network, disease-disease similarity network, and LDA network. LDA scores were obtained by RWR in the lncRNA-disease heterogeneous network. However, due to the few available lncRNA-miRNA interactions, the lncRNA-lncRNA crosstalk network is incomplete, which may lead to prediction bias.

The traditional RWR is used known disease-related lncRNAs to set the initial probabilities. Whereas, for any given disease, if a disease has no known associated lncRNAs (isolated disease), RWR cannot perform the corresponding work. In addition, the construction of lncRNA-lncRNA network is also a problem that needs to be solved. Chen et al.

proposed IRWRLDA, based on an improved RWR to predict LDAs [92]. IRWRLDA integrated three networks, containing the lncRNA-lncRNA similarity network (integrating lncRNA expression profile similarity and lncRNA GIPK similarity), the disease similarity network, and the LDA network. Additionally, the novelty of IRWRLDA was that it combined lncRNA expression similarity and disease semantic similarity to set the initial probability matrix of RWR, and obtained disease-associated lncRNAs by RWR on the lncRNA-lncRNA network. Thus, IRWRLDA could be applied to diseases with no known associated lncRNAs. Gu et al. developed a global network-based RWR named GrwLDA to predict LDAs, which can also be applied to diseases without known related lncRNAs and lncRNAs without known related diseases [93]. A lncRNA-disease heterogeneous network was constructed by employing lncRNA similarity, disease similarity, and LDA. GrwLDA performed RWR from lncRNA nodes and disease nodes, respectively, then integrated the two results as the final LDA scores.

To comprehensively consider the structural differences between lncRNA networks and disease networks, several Bi-Random Walks (BRW)-based algorithms are proposed. Yu et al. developed the BRW-based method to predict LDAs, named BRWLDA, which calculated lncRNA-lncRNA similarity by integrating lncRNA-miRNA interactions, miRNA-disease associations, and lncRNA-gene functional associations [94]. The lncRNA sub-network (lncRNA-lncRNA similarity), disease sub-network (disease semantic similarity), and LAD network were utilized to construct a directed bi-relational network. Then, BRWLDA adopted the BRW to identify the associations between lncRNAs and diseases. Hu et al. proposed a computational LDA prediction model (called BiWalkLDA) in the framework of BRW to solve the cold-start problem by calculating the interaction profile of lncRNAs using the mean value of the neighbors of lncRNA [95]. Similarly, the BRW was performed in the lncRNA-disease heterogeneous network. The known LDAs are few, resulting in a sparse association matrix. Xie et al. improved the preprocessing operation in LDA prediction, and proposed a computational model (named LDA-LNSUBRW) based on linear neighborhood similarity and unbalanced bi-random walk [96]. The model recalculated the interaction score between lncRNAs and diseases by using the weighted K nearest known neighbors (WKNKN) algorithm [97] for known LDAs. Considering that there were many isolated nodes in the lncRNA functional similarity network and disease semantic similarity, linear neighbor similarities between lncRNAs and diseases were calculated, and lncRNA-disease binary network was constructed. Unbalance BRW in the binary network was performed to predict the disease-associated candidate lncRNAs. Meanwhile, Xie et al. introduced Hybrid algorithm and unbalance BRW to predict potential LDAs (named HAUBRW) [98].

The network-based approaches allow flexible integration of multiple sources data with lncRNAs- and diseases-related for improving the LDA prediction performance. Multiple biomedical entities, such as proteins, genes, miRNAs, etc., can be contained in the multi-layer heterogeneous networks that are built. Sumathipala et al. proposed a method LION, which built a lncRNA-protein-disease tripartite network based on lncRNA-protein interactions, protein-protein interactions, and disease-protein associations, and performed RW network diffusion algorithm to obtain the proximity between lncRNAs and diseases [99]. The heterogeneous network contains complex relationships (inter- and intra-relationships) between the different types of nodes, while the intra-relationships between nodes of the same type contained in the homogeneous network should not be neglected. Zhao et al. combined the intra-network (the lncRNA, disease, and gene similarities), as well as the inter-network (the lncRNA-disease, lncRNA-gene, and disease-gene associations) to construct a lncRNA-gene-disease triple-layer network with six sub-networks [100]. An RWR-based prediction model, MHRWR, was developed to predict potential LDAs. Fan et al. developed IDHI-MIRW, which integrates more information to calculate multiple types similarities of lncRNA and disease, including lncRNA expression profile similarity, miRNA-associated lncRNA kernel similarity, protein-associated lncRNA kernel similarity, disease DoSim, miRNA-associated disease kernel similarity, and gene-associated disease kernel similarity [101]. Topological information of each similarity network was captured by performing RWR and PPMI. The topological similarity of each node pair was measured using PPMI [102], and the final lncRNA similarity network and disease similarity network were achieved by using average fusion of 3 types of lncRNA similarities and disease similarities. Finally, the lncRNA-disease heterogeneous network was constructed by combining the lncRNA similarity network, disease similarity network and known LDA network, and performing RWR on them to predict LDAs. Wang et al. proposed LRWRHLDA, which integrates six heterogeneous and four homogeneous networks (similarity networks) consisting of lncRNAs, genes, miRNAs, and diseases to build a lncRNA-gene-miRNA-disease global network [103]. RWR was then used to predict the LDAs.

There were several improved RW algorithms, IIRWR [104], and LRWHLDA [105]. IIRWR introduces the concept of disease clique, which promotes the process of the walker to the next node. LRWHLDA designed a local RW-based method for predicting potential LDAs, that fully utilized local topological information in heterogeneous networks.

#### 4.2.2. *Flow propagation and label propagation*

Zhang et al. proposed LncRDNetFlow, a flow propagation algorithm that integrated multi-sources information to predict potential LDAs [106]. Firstly, the model combined three intra-networks (lncRNA-lncRNA similarity work, disease-disease similarity network, and protein-protein interaction network) and three inter-networks (LDA network, disease-protein association network, and lncRNA-protein interaction network) to construct a lncRNA-protein-disease global network. The network propagation algorithm was employed to measure the correlation values between disease and each lncRNA. Xie et al. proposed a network-based label propagation algorithm, LLCLPLDA, to predict LDAs [21]. Firstly, the lncRNA-lncRNA similarity matrix and disease-disease similarity matrix were extracted from known LDAs by using locality-constrained linear coding (LLC). Then, the final LDA scores were calculated by combining lncRNA expression profile similarity as well as disease semantic similarity and applying label propagation algorithm.

#### 4.2.3. *Network inference*

Unlike other methods, Ping et al. achieved better performance by only using known LDA bipartite networks to predict disease-associated lncRNAs [107]. Firstly, based on the assumption that two lncRNA (disease) nodes are similar if they have common neighboring disease (lncRNA) nodes. The model calculated the first-order similarities of lncRNAs and diseases. Secondly, based on the assumption that if two lncRNA (disease) nodes do not have common neighboring nodes but are related to similar disease (lncRNA) nodes, then they are also considered that there is some degree of similarity between them. The authors measured the second-order similarity of lncRNAs and diseases. Finally, the first- and second-order similarities of lncRNAs and diseases were integrated, and the LDA score matrix was predicted by network inference.

#### 4.2.4. *Other network propagation algorithms*

Yang et al. integrated LDA and protein-coding gene-disease association, and constructed a coding-non-coding genes-disease bipartite network to reflect the association between disease and disease-causing genes [108]. A propagation algorithm was implemented on the bipartite network to infer LDAs. Similarly, Ding et al. combined LDAs and disease-gene associations to construct a lncRNA-disease-gene tripartite graph that can better delineate the heterogeneity of coding-non-coding genes-disease associations [109]. A resource allocation algorithm, called TPGLDA, was utilized in the tripartite graph to accurately identify potential LDAs. Based on the assumption that phenotypically similar diseases tend to be related to functionally similar lncRNAs, and vice versa.

Wang et al. utilized an improved diffusion model IDLDA to calculate the information conveyed in the lncRNA-disease bipartite network, which was quantified to represent the LDAs [110]. Li et al. presented NCPLDA, which fused lncRNA functional similarity, disease semantic similarity, lncRNA and disease GIPK similarities to generate lncRNA similarity and disease similarity [111]. The LDA matrix was reconstructed using the WKNKN algorithm, which took into account the sparsity of the original LDA matrix. Network consistency projections in lncRNA space and disease space were performed separately to obtain projection scores, and the results from these two spaces were combined to predict the LDA scores. Similar to NCPLDA, Zhang et al. proposed a network-based space projection scores method (called LDAI-ISPS) to predict LDAs [112]. This model utilized lncRNA similarity (lncRNA functional similarity and lncRNA GIPK similarity) and disease similarity (disease semantic similarity and disease GIPK similarity), and known LDA networks. The final LDA scores were integrated by the two space projection scores constructed by the lncRNA network and disease network, respectively. Deng et al. developed a method LDAH2V, which constructed LMD heterogenous information network to learn the meta-paths and feature vectors of lncRNA-disease pairs by using HIN2Vec [113], and employed the gradient boosting tree to predict association scores [114]. Li et al. proposed a new model SVDNVLDA that captures linear and nonlinear features of lncRNA and disease using SVD and node2vec methods. The linear and nonlinear features of each entity were connected and XGBoost classifier was adopted to identify LDAs [115].

### 4.3. Matrix factorization- and completion-based methods

The LDA prediction can be thought of recommender system problem. Recent studies have shown that matrix factorization and matrix completion techniques are effective methods that have been widely used for data representation in recommendation tasks. The main purpose of matrix factorization is to discover two low-dimensional matrices that can approximate the original input matrix, which can recover and fill in the missing associations. In comparison to other methods, computational methods based on matrix factorization and completion do not require negative samples to train the model, and can be more flexible to fuse lncRNA- and disease-related multi-source data. As shown in Table 4, we summarize the calculation methods based on matrix factorization and matrix completion.

#### 4.3.1. Matrix factorization

Integrating multi-source data of lncRNA and disease can improve prediction performance by using more prior knowledge from different perspectives. The matrix factorization-based on MFLDA model was developed by Fu et al., which integrates multi-information for predicting disease-related lncRNAs [116]. The model took into account relationships among six different types of objects, including lncRNA, miRNA, genes, gene ontology, and disease ontology. Firstly, multi-relational data matrices into low-rank matrices using matrix tri-factorization. Secondly, the weights were preset for these low-rank matrices to selectively integrate these multi-source data. Finally, the low-rank matrices and weight matrices were iteratively optimized, and the LDA matrix was reconstructed based on the optimized low-rank matrices and weights. However, considering that MFLDA ignored the different correlations of intra-relationship matrices for the same type of objects, Wang et al. proposed a new weight matrix factorization-based method, WMFLDA, to improve the prediction performance [117].

Li et al. developed a new LDA prediction method, the dual network integrated logistic matrix factorization, DNILMF-LDA [19]. The lncRNA expression similarity matrix and the disease semantic similarity matrix were transformed into the lncRNA kernel matrix and the disease kernel matrix by the conversion algorithm. Then, nonlinear fusion technology was utilized to combine the lncRNA and disease GIPK matrix into a lncRNA and disease kernel matrices, respectively. Finally, DNILMF-LDA adopted logistic matrix factorization and similarity information to predict the interaction probability of lncRNAs with diseases. Xuan et al. integrated the lncRNA-miRNA interaction network, miRNA-disease association network, and LDA network to construct a weighted lncRNA-disease network [118]. Because there were few known LDAs, disease similarities, and lncRNA similarities, KNN algorithms was performed to reconstruct the weighted lncRNA-disease network. The computational model PMFILDA, based on probabilistic matrix factorization, was proposed to infer LDAs.

Liu et al. introduced graph regularization into the collaborative matrix factorization, and developed the WGRCMF, a new computational model for predicting LDAs [119]. The model utilized an adjustment parameter to integrate lncRNA expression similarity, lncRNA GIPK similarity, disease semantic similarity, and disease GIPK similarity. The input association matrix was decomposed into two low-rank latent feature matrices using weighted graph regularized collaborative matrix factorization. Wang et al. considered the p-nearest neighbor graph to recalculate the affinity graph of lncRNA and disease utilizing lncRNA and disease similarities [120]. Secondly, the lncRNA-disease adjacency matrix was reconstructed using WKNKN. A model based on graph regularized non-negative matrix factorization, LDGRNMF, was utilized to predict potential LDAs.

#### 4.3.2. Matrix completion

The goal of the matrix-completion methods is to discover new disease-related candidate lncRNAs by filling in the unknown elements of the LDA matrix. Lu et al. presented SIMCLDA, an approach based on inductive matrix completion, to predict disease-associated lncRNAs [121]. The approach calculated lncRNA GIPK similarity and disease Jaccard similarity by using known LDAs. The primary feature vectors of lncRNA and disease were extracted using the principal component analysis (PCA) algorithm. Based on the assumption that similar lncRNAs interacted with similar diseases, SIMCLDA recomputed the lncRNA-disease interaction profiles based on the neighbors of lncRNAs. The association matrix was reconstructed using the primary feature vectors and lncRNA-disease interaction profiles. Meanwhile, Lu et al. introduced lncRNA sequence similarity and developed a computational model based on geometric matrix completion, GMCLDA, for predicting LDAs [122].

### 4.4. Deep learning-based methods

In the last decade, the development of deep learning has successfully facilitated research in pattern recognition and data mining. Many fields, such as speech recognition, image processing, natural language processing, and computational chemistry, which once relied heavily on manual feature engineering to extract features, have recently been transformed by various deep learning models. Meanwhile, deep learning has been widely used in bioinformatics, including drug repositioning [123,124], miRNA-disease association prediction [125], lncRNA-protein interaction prediction [126], lncRNA-miRNA interaction prediction [127], and other applications using effective models to explore various biological problems. In this section, we discuss deep learning-based LDA prediction models, including the fully connected neural network (FCN), convolutional neural network (CNN), autoencoder (AE), and generative adversarial network (GAN), as shown in Table 5.

#### 4.4.1. Full connected neural network

Considering that FCN can learn complex nonlinear features of input data, Hu et al. proposed a deep learning model, NNLDA, based on MF and FCN to predict LDAs [128]. NNLDA was divided into two parts: MF and deep. MF extracted node pair representation from lncRNA and disease feature vectors. FCN was applied in the deep part to capture the complex relationship between lncRNA and disease. Finally, the feature representations of the two parts were concatenated together, and the final association scores were predicted using FCN layer. Unlike NNLDA, Zeng et al. proposes DMFLDA, used the FCN to generate the dense vector representation of the nodes by feeding the lncRNA vector and disease vector, which were derived from the rows and columns of the LDA matrix [129]. The LDA score was obtained utilizing FCN layer with sigmoid activation, which fused the outputs of two different networks using the element-wise multiplication. In addition, Zeng et al. presented SDLDA, a framework that combined singular value decomposition (SVD) and FCN to learn linear and nonlinear features of lncRNA and disease nodes in order to predict disease-associated lncRNAs [130]. The LDA matrix was projected into a low-dimensional space using SVD preserved linear features of lncRNAs and diseases. The lncRNA feature vector and disease feature vector were input to FCN to extract nonlinear low-dimensional features of lncRNA and disease nodes. Finally, combining the linear feature vector and the nonlinear feature vector of the lncRNA-disease pairwise, and appling FCN layer to predict the associated probability of pairwise.

### 4.4.2. *Convolutional neural network*

Deep learning can effectively capture complex feature representations of data. For predicting LDAs, Xuan et al. developed a dual CNN with attention mechanism based on CNNLDA, that integrates multiple data sources between lncRNA, miRNA and disease [87]. CNNLDA constructed feature matrix of lncRNA and disease pairwise, based on the biological assumption that a lncRNA and a disease were more likely to be associated with each other when they have similar relationships, associations and interactions with more common lncRNAs (diseases, miRNAs). The CNN was performed in the left part of the model to learn global representation of the nodes from the feature matrix. Considering different connection relationships between nodes and different features of nodes had different contributions to the prediction. On the right part of the model, CNNLDA proposes relationship-level attention and feature-level attention to learn the attention representations of node pairwise. Finally, to integrate these two representations and predict LDAs, an additional FCN layer was employed. LDApred was proposed by Xuan et al. [131]. LDApred and CNNLDA employ the same strategy to construct the feature matrix of node pairwise on the left part of the model, while the former used FCN layer to predict the association score $score_1$ of node pairwise. Information flow propagation algorithm was applied to the right part of the model, to enhance the similarity and association information of lncRNA and disease. The association score $score_2$ was then predicted using CNN and FCN. To acquire the final association scores, integrated $score_1$ and $score_2$ by weighting sum.

Wei et al. built three-dimensional feature blocks using the lncRNA similarity matrix, disease similarity matrix, and LDA, and CNN were utilized to predict the LDA scores (iLncRNAdis-FB) [132]. Firstly, iLncRNAdis-FB integrated LDAs, lncRNA-miRNA interactions, miRNA-disease associations, lncRNA expression profile similarity, lncRNA sequence similarity, and disease semantic similarity to construct four feature matrix blocks of LDA pairwise. Secondly, CNN was utilized to filter noise in the feature blocks and extract high-level features of the node pairs. Finally, LDA scores were predicted using two additional FCN layers. Different from the fusion similarity approach by iLncRNAdis-FB, Zhang et al. developed a deep learning model MCA-Net, which used multi-feature coding and CNN with attention to predict LDAs [133]. MCA-Net measured six similarities of lncRNA and disease that efficiently represent the lncRNA and disease feature vectors. By assigning different weights and integrating various types of information of lncRNA and disease, the similar feature matrix of lncRNA and disease was encoded. Finally, constructed node pair feature matrices, attention CNN and FNC layers were performed to improve feature extraction and predict LDA scores.

### 4.4.3. *Autoencoder*

AE is an end-to-end unsupervised learning method that extracts low-dimensional representation of input data. There are very few known LDAs, resulting in LDA matrices are typically very sparse. To efficiently learn the low-dimensional and dense representation of the nodes. Xuan et al. proposed a deep learning model, CNNDLP, based on the combination of CNN and convolutional autoencoder (CAE) to predict LDAs [134]. CNNDLP mainly consists of two frameworks, the left and the right. The left framework employed CNN with an attention mechanism to extract low-dimensional feature vectors of node pairs and FCN layer to predict the association score $score_1$. To reconstruct the node pair feature matrix, the first-order similarity and second-order similarity of lncRNA and disease were computed. In the right framework, CNNDLP employed CAE to extract the low-dimensional feature representation of the node pairs, and the FCN layer to predict the association score $score_2$. The final LDA score was measured by integrating the two scores $score = \lambda score_1 + (1 - \lambda) score_2$. Unlike CNNDLP, Zhang et al. captured low-dimensional global and local information of nodes from lncRNA-disease heterogeneous networks using FCN autoencoder (named LDNFSGB) [135]. Gradient boosting tree was employed to predict the LDA scores. Guo et al. developed LDASR, which used lncRNA and disease GIPK similarity and disease semantic similarity to construct lncRNA-disease pair vectors [136]. FCN autoencoders were applied to reduce feature dimensionality and extract optimal features. The association probabilities of lncRNAs and diseases were calculated using a rotating forest-based classifier. Su et al. presented a new model constructed a complex and comprehensive molecular correlations network by integrating 9 types of interactions between lncRNAs, proteins, miRNAs, diseases, and drugs [137]. The network embedding aims to convert the node representation from the original high-dimensional space to a low-dimensional space. In the molecular association network, Node2vec [138] was utilized to capture the network structural information of the nodes, and to represent the behavioral features of the nodes. The attribute features of lncRNAs and disease were extracted from the lncRNA sequences using k-mer, and the disease semantic similarity applying autoencoder, respectively. XGBoost was trained as a classifier to predict LDAs.

Variational autoencoder (VAE) is a type of generative model that learns potential attributes and constructs new elements from probability distributions in the latent variable space. Sheng et al. developed VADLP, a deep learning model based on CAE and VAE to predict disease-related lncRNAs [139]. A triple-layer heterogeneous graph with weighted inter-layer edges and intra-layer edges was constructed to exploit the similarities and correlations between lncRNAs, miRNAs, and diseases. VADLP defined three representations, including node attribute representation, node topology representation and feature distribution representation. CAE was performed to extract pairwise attribute representations from the embedding matrix of pairwise, which was built from heterogeneous graphs using embedding strategy. The pairwise topological representation was obtained by the random walk algorithm, the CAE was utilized to learn the hidden topological structure relationships of the lncRNA and disease. Feature distribution representation was modeled by VAE to capture the potential relationships between lncRNAs and diseases. Finally, VADLP leveraged the representation-level attention fusion module to adaptively integrate the three representations, and predicted the LDA scores through FCN layer.

### 4.4.4. *Generative adversarial network*

GAN is also a generative model that optimizes the network structure using game-based training approach, which contains three components, encoder, generator (decoder), and discriminator. Yang et al. proposed a computational model based on bidirectional generative adversarial networks, BiGAN, to predict LDAs, which using lncRNA GIPK similarity, lncRNA sequence similarity, disease semantic similarity, and disease GIPK similarity [140]. LncRNA-disease pairwise vectors were built by integrating lncRNA and disease similarities. Encoder was modeled to reduce the dimensionality of the pairwise vector, and mapped to the latent feature space in $E(x)$. To generate new feature data $G(z)$ based on the learned features, noise $z$ was randomly sampled from the latent space and inputted to the generator. $E(x)$ and $G(z)$ were fed into the discriminator to determine whether the input data was real or false. If the data originates from the encoder $E(x)$, its label was set to 1 by the discriminator. The data, conversely, comes from the generator $G(z)$, whose label was set to 0. Finally, BiGAN used the result of the discriminator as the association score of lncRNA-disease pairwise.

### *4.5. Graph neural network-based methods*

Although deep learning is widely utilized to capture the hidden patterns between lncRNAs and diseases. However, we cannot ignore that the lncRNA-disease interaction network is an undirected graph network, which is graph structured data without rules. In recent years, graph neural networks, such as graph convolutional network (GCN) [141], graph autoencoders (GAE) [142], graph attention network (GAT) [143], which have shown superior performance on many tasks, such as social networks, natural language processing, and computer vision. In particular, GCN, proposed by Kipf et al. has been increasingly applied in bioinformatics [141]. It is a graph model for graph data that can extract nodes feature information and aggregate node neighborhood information effectively. Here, we summarize the computational methods based on graph neural networks (GNN) into two categories: graph feature extraction and graph matrix completion, as shown in Table 6.

#### *4.5.1. Graph feature extraction*

Considering the sparsity of associations between lncRNAs and diseases, the end-to-end model based on GNN was employed to efficiently extract feature representations of the nodes. Meanwhile, machine learning-based classifiers are applied to predict the association scores of lncRNA-disease pairs. To improve LDA prediction, Xuan et al. proposed GCNLDA, CNN and graph convolutional autoencoder (GCA) based approach [144]. The model was divided into two parts: the left and the right parts. In the left part, the authors construct a triple-layer heterogeneous network by integrating association and similarity information between lncRNA, miRNA and disease. Then, GCNLDA used GCA to extract low-dimensional topological representations of lncRNA and disease nodes. Finally, FCN was utilized to predict the association score $score_1$ of the node pairs. In the right part, the pair of lncRNA and disease embedding matrix was built, based on the relevant biological premises, which were used as the input to the CNN to learn the low-dimensional features of the node pair. Similarly, FCN was also adopted to predict the association probability $score_2$. This method integrated the left and right scores as the final association probability of the pairwise of lncRNA and disease through a weighted sum. Considering the overly complex structure of the GCNLDA model, many parameters need to be adjusted. Wu et al. developed a method GAEFR, for predicting LDAs that was based only on graph autoencoder (GCA) [145]. Similarly, GAERF constructed the heterogeneous network by combining the correlation and similarity between lncRNA, miRNA and disease. Then, GCA was utilized to capture latent low-dimensional feature representations of lncRNA and disease nodes from LMD network. The feature vectors were fed into a RF classifier, which was used to predict disease-related lncRNAs. In addition, Wu et al. additionally considered the vanishing gradient problem when encoding with multi-layer GCA (named MLGCNET) [146]. Therefore, layer aggregate was applied to merge each layer feature map to obtain lncRNA and disease feature representations. Extra Trees was classifier that predicts the LDA scores. Sheng et al. considered homogeneous and heterogeneous information in the LMD graph, and used a multi-channel graph attention autoencoder (called MGATE) to extract comprehensive and subtle information of lncRNA and disease nodes from complex-, inter- and intra-graphs [147]. Finally, RF was performed to predict the association scores between lncRNAs and diseases.

In comparison to GCN, GAT can assign different weights to the neighbors of nodes for information aggregation, and also does not require corresponding matrix operations. Lan et al. proposed an end-to-end model, GANLDA, based on GAT to predict disease-related lncRNAs [148]. The model integrated LDAs, lncRNA-gene associations, lncRNA-go associations, and lncRNA-miRNA interactions as the original features of lncRNA, and united disease-gene associations and disease-miRNA associations as the original features of disease. Because the original features of lncRNA and disease were very sparse and noisy, the PCA was utilized to reduce the dimension of the original features. Then, the neighboring feature information of the nodes was then captured by GANLDA using GAT. Finally, the pairwise association scores were predicted using multi-layer perceptron (MLP). Xuan et al. not only employed GAT to extract topological representations of lncRNA and disease nodes respectively, but also used CNN with attribute-level attention mechanisms to learn attribute representations of node pairs (called GTAN) [149]. FCN was used to combine node feature representations, and predicted the LDA scores.

The variational graph autoencoders (VGAE) was a probabilistic method for describing latent space embedding that may efficiently learn a smooth latent space representation of input data, instead of generating single value to represent each latent space. Silva et al. combined LDA, lncRNA transcript information, lncRNA sequence information, and disease symptoms information, to propose a method PANDA based on VGAE to extract the features and edge representation of nodes [150]. Finally, FCN was applied to predict potential candidate LDAs.

#### *4.5.2. Graph matrix completion*

The LDA prediction can be regarded as the link prediction problem. As a result, from the perspective of link prediction, matrix completion of LDA prediction can be considered. Matrix completion-based link prediction was applied to many fields and has a lot of practical applications. With the progress of GNN on graph structure data, Wu et al. developed GAMCLDA, a computational model based on graph autoencoder matrix completion, to predict disease-associated lncRNAs [151]. The method used GCN to encode local features of lncRNA-disease heterogeneous graph, and to learn latent factor vectors of lncRNAs and diseases. The decoder reconstructs the LDA matrix by inner-product the factor vectors of lncRNAs and the factor vectors of diseases. Considering that conditional random field (CRF) can better preserve the similarity information of nodes in GCN for node classification tasks. Fan et al. introduced GCRFLDA, a graph matrix completion prediction method that utilized the encoder composed of CRF and attention mechanism to learn node embeddings [152]. As decoder, the inner product of lncRNA embedding and disease embedding generated the LDA score matrix.

Considering that meta-paths in graph data can capture the specific semantic information of the graph, and different meta-paths represent different semantic information, Zhao et al. proposed a graph attention completion framework based on meta-path, HGATLDA, to improve LDA prediction using meta-path information of nodes [153]. HGATLDA used the transformation matrix to project lncRNA expression profile similarity and lncRNA GIPK similarity into lncRNA feature space, as well as projects disease semantic similarity and disease GIPK similarity into disease feature space, and combined LDA network constructs lncRNA-disease heterogeneous graph. HGATLDA took into account five types of meta-paths information, including disease-lncRNA-disease, lncRNA-lncRNA, disease-disease, lncRNA-disease-lncRNA, lncRNA-disease-lncRNA, and lncRNA-disease. In each meta-path, a multi-head GAT network was employed to extract path information of lncRNAs and disease. Meta-path-based subgraph attention was used to fuse node information and reconstruct the LDA matrix using neural inductive matrix completion.

In addition, Shi et al. developed VGAELDA, a graph model based on VGAE and GAE, to predict LDAs [154]. The model mainly consisted of two representation learning frameworks, feature inference network framework and label propagation network framework. In the feature inference framework, the VGAE was employed to learn representations from lncRNA and disease features. In the label propagation network framework, the GAE utilized known LDA networks to propagate labels and learn network representations of nodes. Alternate training with the variational expectation maximization (EM) algorithm was performed to enhance VGAELDA ability to extract efficient low-dimensional features from high-dimensional features.

## 5. Experimental evaluation

### 5.1. Evaluation methods

K-fold cross-validation (k-CV) and LOOCV are extensively adopted to analyze the prediction performance to systematically evaluate the effectiveness of the computational approaches. Among these, k-CV mainly includes 5-CV and 10-CV. These approaches typically treat the experimentally verified LDAs as positive samples, whereas the unverified associations as negative samples. For k-CV experiments, all positive samples are randomly divided into k equal size sets. At each fold, selected k-1 sets of positive samples and randomly selected negative samples whose size is equal to the number of samples with k-1 positive samples for training. For testing, the remaining one set of positive samples and negative samples are used. For LOOCV experiments, one positive sample is selected each time for testing, while the rest of samples are utilized for training.

### 5.2. Evaluation metrics

Unknown interactions between lncRNAs and diseases are generally referred to as candidate associations. The tasks for predicting LDAs can be regarded as binary classification tasks. As a result, many machine learning evaluation metrics, which are typically utilized for classification tasks, are employed to assess the proposed computational methods prediction performance. The receiver operating characteristic curve (ROC) and the precision-recall (PR) are often employed to assess the proposed computational methods performance. The receiver operating characteristic (ROC) curve is plotted based on true positive rate (TPR), also named Sensitivity (Sen), as the vertical axis and false positive rate (FPR), also called Specificity (Spe), as the horizontal axis. Based on the precision as the vertical axis and recall as the horizontal axis, the precision-recall (PR) curve is obtained. The area under the ROC curve (AUC) and the area under the PR curve (AUPR) can better reflect the performance of the different methods, and the higher values indicate better prediction performance of the approaches. TPR (FPR) is defined as the proportion of correctly (incorrectly) identified positive (negative) samples among all the positive (negative) samples. The precision (Pre) represents the percentage of the correctly identified positive samples among all the samples, which are predicted as positive and recall (Rec) is the same as TPR. There are a number of other metrics that are used, including Accuracy (Acc), Matthews Correlation Coefficient (Mcc), F1-score (F1), top-ranked (top-r). To further validate the capacity of the above computational approaches to predict LDAs, one or several diseases are usually investigated in case studies and top-ranked LDAs are selected for further validation. For case studies, two approaches are typically performed to get disease-related potential candidate lncRNAs. (1) All known LDAs in the dataset are used for training, while the remaining unknown LDAs are predicted by the trained model. (2) All the original lncRNA related information of the investigated disease is removed from the training dataset. The purpose of this strategy is employed to prove the ability of the proposed model to predict lncRNAs associated with new diseases. In both of these approaches, for each disease, the lncRNA candidates are ranked in descending order based on their LDA scores. Finally, the researchers examined how many disease-associated top-ranked lncRNAs are verified using public databases and published literature.

## 6. Conclusions and perspectives

Exploring the complex relationship between lncRNAs and diseases can provide a novel insight for the diagnosis and treatment of diseases. Compared with traditional biological experiments, computational methods for identifying candidate LDAs are more valuable and promising, which can reduce time, save cost, and diminish the risk of failure. Based on large-scale biological data and multiple models, computational methods for LDA identification are implemented. In this review, we present a comprehensive insight of computational models and data resources aimed on identifying potential associations between lncRNAs and human diseases. Firstly, we list representative disease associated with lncRNAs, contains human cancers, cardiovascular diseases, and neurological diseases. Secondly, we present available lncRNA- and disease-related data resources. Thirdly, we summarize the computational models for identifying LDAs so far and divide them into 5 categories: machine learning, network propagation, matrix factorization and completion, deep learning, and graph neural network based methods. Fourthly, we outline commonly used evaluation methods and metrics in computational methods. Finally, we discuss the future challenges and perspectives of lncRNAs and diseases research.

It is worth noting that each method has advantages and limitations, and users need to choose the appropriate method according to their requirements. Most methods based on machine learning can effectively predict new lncRNAs and new diseases. However, they are often based on shallow models and can achieve poor performance. Nowadays, network propagation-based methods are an effective tool for inferring potential LDAs, which can integrate known LDA networks, lncRNA-relate networks, and disease-related networks. Potential LDAs are predicted using propagation algorithms such as RWR. These methods have the disadvantage that they are the over-reliance on known LDAs to construct reliable networks. Matrix factorization- and completion-based methods allow for flexible integration of data, but make it difficult to learn the deeper features of nodes. The key advantage of deep learning-based methods is that they can learn the deep features and non-linear representation of nodes to the extent that they can achieve better prediction performance. However, these models often have many training parameters and are prone to overfitting problems. In addition, deep learning requires large amounts of data to train the model and has weak explanations. Graph neural network-based methods can aggregate node topological structure information and attribute information, but it is difficult to achieve good prediction results for isolated nodes (i.e., new diseases and new lncRNAs). Overall, deep learning- and graph neural network-based methods can achieve significantly better performance.

As more multi-source data on lncRNAs and diseases available in the future, the prediction accuracy of these methods may be further improved. From the data perspective, we review the public data resources involved in LDA prediction. However, these biological data, which are lncRNA- and disease-related derived from different sources, tend to be incomplete, noisy, unreliable, and inconsistent. As a result, how to introduce amount of available heterogeneous data and integrate them into a unified work is also a hot topic for future research.

Benefiting from the above collection of data, many computational prediction methods are developed for discovering potential LDAs. We grouped these prediction methods into 5 categories: machine learning, network propagation, matrix factorization and completion, deep learning, and graph neural networks. Recently, with the development of deep learning and graph neural network, prediction methods based on these two types of models have shown better prediction performance [87,132,145,147]. However, as mentioned above, each computational method, has its applicability and limitations. Therefore, it is also a challenge to combine different computational models by analyzing their characteristics or to develop simple computational models to improve their prediction performance.

Furthermore, LDA prediction methods tend to require both positive samples and negative samples for training prediction models. However, true negative samples are not available. Most of current studies try to address this problem by randomly selecting negative samples from unlabeled data, which may affect model performance because they are often not true negative samples. Therefore, in the future work designing a more reasonable negative sample selection strategy is important.

Additionally, there is an essential problem that many computational models have been developed by computational scientists, but they all perform case studies to analyze the prediction results by seeking evidence from existing public databases and published literature. However, this leads to a lack of cooperation between computational scientists and biologists in verifying prediction results, thus falling short of the goal of reducing costs and improving efficiency through computational methods. In conclusion, researchers need to collaborate more closely to reveal the complex relationship between lncRNAs and diseases, and to provide new perspectives on disease diagnosis, treatment, and prognosis.

## Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62072212), the Development Project of Jilin Province of China (Nos. 20220508125RC, 20210508060RQ), and the Chinese Postdoctoral Science Foundation (No. 2021M691211).

## References

[1] J. Beermann, M.-T. Piccoli, J. Viereck, et al., Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches, Physiol. Rev. 96 (4) (2016) 1297–1325, https://doi.org/10.1152/physrev.00041.2015.

[2] X. Zhang, R. Hong, W. Chen, et al., The role of long noncoding RNA in major human disease, Bioorg. Chem. 92 (2019), 103214, https://doi.org/10.1016/j.bioorg.2019.103214.

[3] Kevin C. Wang, Y. Chang Howard, Molecular mechanisms of long noncoding RNAs, Mol. Cell 43 (6) (2011) 904–914, https://doi.org/10.1016/j.molcel.2011.08.018.

[4] M. Kazimierczyk, M.K. Kasprowicz, M.E. Kasprzyk, et al., Human long noncoding RNA interactome: detection, characterization and function, Int. J. Mol. Sci. 21 (3) (2020), https://doi.org/10.3390/ijms21031027.

[5] O. Wapinski, H.Y. Chang, Long noncoding RNAs and human disease, Trends Cell Biol. 21 (6) (2011) 354–361, https://doi.org/10.1016/j.tcb.2011.04.001.

[6] M. Huarte, The emerging role of lncRNAs in cancer, Nat. Med. 21 (11) (2015) 1253–1261, https://doi.org/10.1038/nm.3981.

[7] S. Ghafouri-Fard, M. Esmaeili, M. Taheri, H19 lncRNA: roles in tumorigenesis, Biomed. Pharmacother. 123 (2020), 109774, https://doi.org/10.1016/j.biopha.2019.109774.

[8] J. Ren, J. Fu, T. Ma, et al., LncRNA H19-elevated LIN28B promotes lung cancer progression through sequestering miR-196b, Cell Cycle 17 (11) (2018) 1372–1380, https://doi.org/10.1080/15384101.2018.1482137.

[9] K. Zhang, Z. Luo, Y. Zhang, et al., Circulating lncRNA H19 in plasma as a novel biomarker for breast cancer, Cancer Biomarkers 17 (2016) 187–194, https://doi.org/10.3233/CBM-160630.

[10] J. Yan, Y. Zhang, Q. She, et al., Long noncoding RNA H19/miR-675 Axis promotes gastric cancer via FADD/Caspase 8/Caspase 3 signaling pathway, Cell. Physiol. Biochem. 42 (6) (2017) 2364–2376, https://doi.org/10.1159/000480028.

[11] F. Nasri, B. Gharesi-Fard, B. Namavar Jahromi, et al., Sperm DNA methylation of H19 imprinted gene and male infertility, Andrologia 49 (10) (2017), e12766, https://doi.org/10.1111/and.12766.

[12] D. Bartholdi, M. Krajewska-Walasek, K. Õunap, et al., Epigenetic mutations of the imprinted IGF2-H19 domain in Silver-Russell syndrome (SRS): results from a large cohort of patients with SRS and SRS-like phenotypes, J. Med. Genet. 46 (3) (2009) 192, https://doi.org/10.1136/jmg.2008.061820.

[13] F-q Nie, M. Sun, J-s Yang, et al., Long noncoding RNA ANRIL promotes non-small cell lung cancer cell proliferation and inhibits apoptosis by silencing KLF2 and P21 expression, Mol. Cancer Therapeut. 14 (1) (2015) 268–277, https://doi.org/10.1158/1535-7163.MCT-14-0492.

[14] X. Zang, J. Gu, J. Zhang, et al., Exosome-transmitted lncRNA UFC1 promotes non-small-cell lung cancer progression by EZH2-mediated epigenetic silencing of PTEN expression, Cell Death Dis. 11 (4) (2020) 215, https://doi.org/10.1038/s41419-020-2409-0.

[15] S. Gupta, R.F. Hashimoto, Dynamical analysis of a Boolean network model of the oncogene role of lncRNA ANRIL and lncRNA UFC1 in non-small cell lung cancer, Biomolecules 12 (3) (2022) 420, https://doi.org/10.3390/biom12030420.

[16] Y. Yu, F. Gao, Q. He, et al., lncRNA UCA1 functions as a ceRNA to promote prostate cancer progression via sponging miR143, Mol. Ther. Nucleic Acids 19 (2020) 751–758, https://doi.org/10.1016/j.omtn.2019.11.021.

[17] M. Taheri, M. Habibi, R. Noroozi, et al., HOTAIR genetic variants are associated with prostate cancer and benign prostate hyperplasia in an Iranian population, Gene 613 (2017) 20–24, https://doi.org/10.1016/j.gene.2017.02.031.

[18] D. Xue, H. Lu, H.-Y. Xu, et al., Long noncoding RNA MALAT1 enhances the docetaxel resistance of prostate cancer cells via miR-145-5p-mediated regulation of AKAP12, J. Cell Mol. Med. 22 (6) (2018) 3223–3237, https://doi.org/10.1111/jcmm.13604.

[19] Y. Li, J. Li, N. Bian, Dnilmf-Lda, Prediction of lncRNA-disease associations by dual-network integrated logistic matrix factorization and Bayesian optimization, Genes 10 (8) (2019) 608, https://doi.org/10.3390/genes10080608.

[20] J. Li, Q. Li, D. Li, et al., Long non-coding RNA MNX1-AS1 promotes progression of triple negative breast cancer by enhancing phosphorylation of Stat3, Front. Oncol. 10 (2020) 1108, https://doi.org/10.3389/fonc.2020.01108.

[21] G. Xie, S. Huang, Y. Luo, et al., LLCLPLDA: a novel model for predicting lncRNA-disease associations, Mol. Genet. Genom. 294 (6) (2019) 1477–1486, https://doi.org/10.1007/s00438-019-01590-8.

[22] Y. Fan, W. Sheng, Y. Meng, et al., LncRNA PTENP1 inhibits cervical cancer progression by suppressing miR-106b, Artif. Cell Nanomed. Biotechnol. 48 (1) (2020) 393–407, https://doi.org/10.1080/21691401.2019.1709852.

[23] X. Chen, C.C. Yan, X. Zhang, et al., Long non-coding RNAs and complex diseases: from experimental results to computational models, Briefings Bioinf. 18 (4) (2017) 558–576, https://doi.org/10.1093/bib/bbw060.

[24] X. Chen, Y.-Z. Sun, N.-N. Guan, et al., Computational models for lncRNA function prediction and functional similarity calculation, Brief. Func. Genom. 18 (1) (2019) 58–82, https://doi.org/10.1093/bfgp/ely031.

[25] L. Statello, C.-J. Guo, L.-L. Chen, et al., Gene regulation by long non-coding RNAs and its biological functions, Nat. Rev. Mol. Cell Biol. 22 (2) (2021) 96–118, https://doi.org/10.1038/s41580-020-00315-9.

[26] M.J. Cardoso, K. Mokbel, Locoregional therapy in de novo metastatic breast cancer. The unanswered question, Breast 58 (2021) 170–172, https://doi.org/10.1016/j.breast.2021.05.002.

[27] K.X.L.Z.H. Lou, P. Wang, et al., Long non-coding RNA BANCR indicates poor prognosis for breast cancer and promotes cell proliferation and invasion, Eur. Rev. Med. Pharmacol. Sci. 22 (5) (2018) 1358–1365, https://doi.org/10.26355/eurrev_201803_14479.

[28] Y.-X. Ding, K.-C. Duan, S.-L. Chen, Low expression of lncRNA-GAS5 promotes epithelial-mesenchymal transition of breast cancer cells in vitro, Nan fang yi ke da xue xue bao, J. South. Med. Univ. 37 (11) (2017) 1427–1435, https://doi.org/10.3969/j.issn.1673-4254.2017.11.01.

[29] H. Wu, C. Zhou, Long non-coding RNA UCA1 promotes lung cancer cell proliferation and migration via microRNA-193a/HMGB1 axis, Biochem. Biophys. Res. Commun. 496 (2) (2018) 738–745, https://doi.org/10.1016/j.bbrc.2018.01.097.

[30] X. Liu, J. Ma, F. Xu, et al., TINCR suppresses proliferation and invasion through regulating miR-544a/FBXW7 axis in lung cancer, Biomed. Pharmacother. 99 (2018) 9–17, https://doi.org/10.1016/j.biopha.2018.01.049.

[31] K. Su, N. Wang, Q. Shao, et al., The role of a ceRNA regulatory network based on lncRNA MALAT1 site in cancer progression, Biomed. Pharmacother. 137 (2021), 111389, https://doi.org/10.1016/j.biopha.2021.111389.

[32] T. Rajagopal, S. Talluri, R.L. Akshaya, et al., HOTAIR LncRNA: a novel oncogenic propellant in human cancer, Clin. Chim. Acta 503 (2020) 1–18, https://doi.org/10.1016/j.cca.2019.12.028.

[33] T. Vos, A.A. Abajobir, K.H. Abate, et al., Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet 390 (10100) (2017) 1211–1259, https://doi.org/10.1016/S0140-6736(17)32154-2.

[34] Y. Fang, Y. Xu, R. Wang, et al., Recent advances on the roles of LncRNAs in cardiovascular disease, J. Cell Mol. Med. 24 (21) (2020) 12246–12257, https://doi.org/10.1111/jcmm.15880.

[35] T. Zhou, G. Qin, L. Yang, et al., LncRNA XIST regulates myocardial infarction by targeting miR-130a-3p, J. Cell. Physiol. 234 (6) (2019) 8659–8667, https://doi.org/10.1002/jcp.26327.

[36] J.-H. Lee, C. Gao, G. Peng, et al., Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts, Circ. Res. 109 (12) (2011) 1332–1341, https://doi.org/10.1161/CIRCRESAHA.111.249433.

[37] R. Kumarswamy, C. Bauters, I. Volkmann, et al., Circulating long noncoding RNA, LIPCAR, predicts survival in patients with heart failure, Circ. Res. 114 (10) (2014) 1569–1575, https://doi.org/10.1161/CIRCRESAHA.114.303915.

[38] Y. Wo, J. Guo, P. Li, et al., Long non-coding RNA CHRF facilitates cardiac hypertrophy through regulating Akt3 via miR-93, Cardiovasc. Pathol. 35 (2018) 29–36, https://doi.org/10.1016/j.carpath.2018.04.003.

[39] Y. Yan, D. Song, X. Song, et al., The role of lncRNA MALAT1 in cardiovascular disease, IUBMB Life 72 (3) (2020) 334–342, https://doi.org/10.1002/iub.2210.

[40] X. Hua, Y.-Y. Wang, P. Jia, et al., Multi-level transcriptome sequencing identifies COL1A1 as a candidate marker in human heart failure progression, BMC Med. 18 (1) (2020) 2, https://doi.org/10.1186/s12916-019-1469-4.

[41] M. Zhang, P. He, Z. Bian, Long noncoding RNAs in neurodegenerative diseases: pathogenesis and potential implications as clinical biomarkers, Front. Mol. Neurosci. (2021) 161, https://doi.org/10.3389/fnmol.2021.685143.

[42] K.L. Double, S. Reyes, E.L. Werry, et al., Selective cell death in neurodegeneration: why are some neurons spared in vulnerable regions? Prog.

Neurobiol. 92 (3) (2010) 316–329, https://doi.org/10.1016/j.pneurobio.2010.06.001.

[43] Y. Chen, J. Zhou, LncRNAs: macromolecules with big roles in neurobiology and neurological diseases, Metab. Brain Dis. 32 (2) (2017) 281–291, https://doi.org/10.1007/s11011-017-9965-8.

[44] Z. Breijyeh, R. Karaman, Comprehensive review on Alzheimer's disease: causes and treatment, Molecules 25 (24) (2020), https://doi.org/10.3390/molecules25245789.

[45] S.N. Fotuhi, M. Khalaj-Kondori, M.A. Hoseinpour Feizi, et al., Long non-coding RNA BACE1-AS may serve as an Alzheimer's disease blood-based biomarker, J. Mol. Neurosci. 69 (3) (2019) 351–359, https://doi.org/10.1007/s12031-019-01364-2.

[46] L. Li, Y. Xu, M. Zhao, et al., Neuro-protective roles of long non-coding RNA MALAT1 in Alzheimer's disease with the involvement of the microRNA-30b/CNR1 network and the following PI3K/AKT activation, Exp. Mol. Pathol. 117 (2020), 104545, https://doi.org/10.1016/j.yexmp.2020.104545.

[47] E. Taghizadeh, S.M. Gheibihayat, F. Taheri, et al., LncRNAs as putative biomarkers and therapeutic targets for Parkinson's disease, Neurol. Sci. 42 (10) (2021) 4007–4015, https://doi.org/10.1007/s10072-021-05408-7.

[48] T.F.J. Kraus, M. Haider, J. Spanner, et al., Altered long noncoding RNA expression precedes the course of Parkinson's disease—a preliminary report, Mol. Neurobiol. 54 (4) (2017) 2869–2877, https://doi.org/10.1007/s12035-016-9854-x.

[49] L. Cai, L. Tu, T. Li, et al., Downregulation of lncRNA UCA1 ameliorates the damage of dopaminergic neurons, reduces oxidative stress and inflammation in Parkinson's disease through the inhibition of the PI3K/Akt signaling pathway, Int. Immunopharm. 75 (2019), 105734, https://doi.org/10.1016/j.intimp.2019.105734.

[50] Y. Quan, J. Wang, S. Wang, et al., Association of the plasma long non-coding RNA MEG3 with Parkinson's disease, Front. Neurol. 11 (2020), 532891, https://doi.org/10.3389/fneur.2020.532891.

[51] J.-S. Sunwoo, S.-T. Lee, W. Im, et al., Altered expression of the long noncoding RNA NEAT1 in Huntington's disease, Mol. Neurobiol. 54 (2) (2017) 1577–1586, https://doi.org/10.1007/s12035-016-9928-9.

[52] K. Chanda, S. Das, J. Chakraborty, et al., Altered levels of long NcRNAs Meg3 and Neat1 in cell and animal models of huntington's disease, RNA Biol. 15 (10) (2018) 1348–1363, https://doi.org/10.1080/15476286.2018.1534524.

[53] Z. Bao, Z. Yang, Z. Huang, et al., LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases, Nucleic Acids Res. 47 (D1) (2019) D1034–D1037, https://doi.org/10.1093/nar/gky905.

[54] G. Chen, Z. Wang, D. Wang, et al., LncRNADisease: a database for long-non-coding RNA-associated diseases, Nucleic Acids Res. 41 (D1) (2013) D983–D986, https://doi.org/10.1093/nar/gks1099.

[55] X. Chen, G.-Y. Yan, Novel human lncRNA-disease association inference based on lncRNA expression profiles, Bioinformatics 29 (20) (2013) 2617–2624, https://doi.org/10.1093/bioinformatics/btt426.

[56] W. Lan, M. Li, K. Zhao, et al., LDAP: a web server for lncRNA-disease association prediction, Bioinformatics 33 (3) (2016) 458–460, https://doi.org/10.1093/bioinformatics/btw639.

[57] J. Sun, H. Shi, Z. Wang, et al., Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network, Mol. Biosyst. 10 (8) (2014) 2074–2081, https://doi.org/10.1039/C3MB70608G.

[58] Y. Gao, S. Shang, S. Guo, 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. 49 (D1) (2021) D1251–D1258, https://doi.org/10.1093/nar/gkaa1006.

[59] S. Ning, J. Zhang, P. Wang, et al., Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers, Nucleic Acids Res. 44 (D1) (2016) D980–D985, https://doi.org/10.1093/nar/gkv1094.

[60] L. Ning, T. Cui, B. Zheng, et al., MNDR v3.0: mammal ncRNA-disease repository with increased coverage and annotation, Nucleic Acids Res. 49 (D1) (2021) D160–D164, https://doi.org/10.1093/nar/gkaa707.

[61] Y. Zhao, H. Li, S. Fang, et al., NONCODE 2016: an informative and valuable data source of long non-coding RNAs, Nucleic Acids Res. 44 (D1) (2016) D203–D208, https://doi.org/10.1093/nar/gkv1252.

[62] R.N. Consortium, RNAcentral 2021: secondary structure integration, improved sequence search and new member databases, Nucleic Acids Res. 49 (D1) (2021) D212–D220, https://doi.org/10.1093/nar/gkaa921.

[63] B. Zhou, B. Ji, K. Liu, et al., EVLncRNAs 2.0: an updated database of manually curated functional long non-coding RNAs validated by low-throughput experiments, Nucleic Acids Res. 49 (D1) (2021) D86–D91, https://doi.org/10.1093/nar/gkaa1076.

[64] X. Teng, X. Chen, H. Xue, et al., NPInter v4.0: an integrated database of ncRNA interactions, Nucleic Acids Res. 48 (D1) (2020) D160–D165, https://doi.org/10.1093/nar/gkz969.

[65] P. Wang, Q. Guo, Y. Qi, et al., LncACTdb 3.0: an updated database of experimentally supported ceRNA interactions and personalized networks contributing to precision medicine, Nucleic Acids Res. 50 (D1) (2022) D183–D189, https://doi.org/10.1093/nar/gkab1092.

[66] P.-J. Volders, J. Anckaert, K. Verheggen, et al., LNCipedia 5: towards a reference set of human long non-coding RNAs, Nucleic Acids Res. 47 (D1) (2018) D135–D139, https://doi.org/10.1093/nar/gky1031.

[67] D.R. Zerbino, P. Achuthan, W. Akanni, et al., Ensembl 2018, Nucleic Acids Res. 46 (D1) (2017) D754–D761, https://doi.org/10.1093/nar/gkx1098.

[68] N.A. Oleary, M.W. Wright, J.R. Brister, et al., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation, Nucleic Acids Res. 44 (D1) (2015) D733–D745, https://doi.org/10.1093/nar/gkv1189.

[69] C.-C. Hon, J.A. Ramilowski, J. Harshbarger, et al., An atlas of human long non-coding RNAs with accurate 5′ ends, Nature 543 (7644) (2017) 199–204, https://doi.org/10.1038/nature21374.

[70] L. Liu, Z. Li, C. Liu, et al., LncRNAWiki 2.0: a knowledgebase of human long non-coding RNAs with enhanced curation model and database system, Nucleic Acids Res. 50 (D1) (2021) D190–D195, https://doi.org/10.1093/nar/gkab998.

[71] J. Harrow, A. Frankish, J.M. Gonzalez, et al., GENCODE: the reference human genome annotation for the ENCODE Project, Genome Res. 22 (9) (2012) 1760–1774, https://doi.org/10.1080/15384101.2018.1482137.

[72] C. Xie, J. Yuan, H. Li, et al., NONCODEv4: exploring the world of long non-coding RNA genes, Nucleic Acids Res. 42 (D1) (2013) D98–D103, https://doi.org/10.1093/nar/gkt1222.

[73] P.-J. Volders, K. Helsens, X. Wang, et al., LNCipedia: a database for annotated human lncRNA transcript sequences and structures, Nucleic Acids Res. 41 (D1) (2012) D246–D251, https://doi.org/10.1093/nar/gks915.

[74] D. Bhartiya, K. Pal, S. Ghosh, et al., lncRNome: a comprehensive knowledgebase of human long noncoding RNAs, Database (2013), https://doi.org/10.1093/database/bat034.

[75] Z. Li, L. Liu, S. Jiang, et al., LncExpDB: an expression database of human long non-coding RNAs, Nucleic Acids Res. 49 (D1) (2020) D962–D968, https://doi.org/10.1093/nar/gkaa850.

[76] J.-H. Li, S. Liu, H. Zhou, et al., starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data, Nucleic Acids Res. 42 (D1) (2014) D92–D97, https://doi.org/10.1093/nar/gkt1248.

[77] L. Cheng, P. Wang, R. Tian, et al., LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse, Nucleic Acids Res. 47 (D1) (2019) D140–D144, https://doi.org/10.1093/nar/gky1051.

[78] Y. Yi, Y. Zhao, C. Li, et al., RAID v2.0: an updated resource of RNA-associated interactions across organisms, Nucleic Acids Res. 45 (D1) (2016) D115–D118, https://doi.org/10.1093/nar/gkw1052.

[79] L.M. Schriml, E. Mitraka, J. Munro, et al., Human Disease Ontology 2018 update: classification, content and workflow expansion, Nucleic Acids Res. 47 (D1) (2019) D955–D962, https://doi.org/10.1093/nar/gky1032.

[80] S. Köhler, L. Carmody, N. Vasilevsky, et al., Expansion of the human phenotype ontology (HPO) knowledge base and resources, Nucleic Acids Res. 47 (D1) (2019) D1018–D1027, https://doi.org/10.1093/nar/gky1105.

[81] J.S. Amberger, C.A. Bocchini, F. Schiettecatte, et al., OMIM.org: online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders, Nucleic Acids Res. 43 (D1) (2015) D789–D798, https://doi.org/10.1093/nar/gku1205.

[82] J. Piñero, J.M. Ramírez-Anguita, J. Saüch-Pitarch, et al., The DisGeNET knowledge platform for disease genomics: 2019 update, Nucleic Acids Res. 48 (D1) (2020) D845–D855, https://doi.org/10.1093/nar/gkz1021.

[83] Z. Huang, J. Shi, Y. Gao, et al., HMDD v3.0: a database for experimentally supported human microRNA-disease associations, Nucleic Acids Res. 47 (D1) (2019) D1013–D1017, https://doi.org/10.1093/nar/gky1010.

[84] X. Chen, C. Clarence Yan, C. Luo, et al., Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity, Sci. Rep. 5 (1) (2015), 11338, https://doi.org/10.1038/srep11338.

[85] Q. Chen, D. Lai, W. Lan, et al., ILDMSF: inferring associations between long non-coding RNA and disease based on multi-similarity fusion, IEEE ACM Trans. Comput. Biol. Bioinf 18 (3) (2021) 1106–1112, https://doi.org/10.1109/TCBB.2019.2936476.

[86] D. Yao, X. Zhan, X. Zhan, et al., A random forest based computational model for predicting novel lncRNA-disease associations, BMC Bioinf. 21 (1) (2020) 126, https://doi.org/10.1186/s12859-020-3458-1.

[87] P. Xuan, Y. Cao, T. Zhang, et al., Dual convolutional neural networks with attention mechanisms based method for predicting disease-related lncRNA genes, Front. Genet. 10 (2019), https://doi.org/10.3389/fgene.2019.00416.

[88] R. Zhu, Y. Wang, J.-X. Liu, et al., IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier, BMC Bioinf. 22 (1) (2021) 175, https://doi.org/10.1186/s12859-021-04104-9.

[89] J. Yu, Z. Xuan, X. Feng, et al., A novel collaborative filtering model for LncRNA-disease association prediction based on the Naïve Bayesian classifier, BMC Bioinf. 20 (1) (2019) 396, https://doi.org/10.1186/s12859-019-2985-0.

[90] J. Yu, P. Ping, L. Wang, et al., A novel probability model for LncRNA-disease association prediction based on the Naïve Bayesian classifier, Genes 9 (7) (2018) 345, https://doi.org/10.3390/genes9070345.

[91] M. Zhou, X. Wang, J. Li, et al., Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network, Mol. Biosyst. 11 (3) (2015) 760–769, https://doi.org/10.1039/C4MB00511B.

[92] X. Chen, Z.-H. You, G.-Y. Yan, et al., IRWRLDA: improved random walk with restart for lncRNA-disease association prediction, Oncotarget 7 (36) (2016) 57919–57931, https://doi.org/10.18632/oncotarget.11141.

[93] C. Gu, B. Liao, X. Li, et al., Global network random walk for predicting potential human lncRNA-disease associations, Sci. Rep. 7 (1) (2017), 12442, https://doi.org/10.1038/s41598-017-12763-z.

[94] G. Yu, G. Fu, C. Lu, et al., BRWLDA: bi-random walks for predicting lncRNA-disease associations, Oncotarget 8 (36) (2017) 60429–60446, https://doi.org/10.18632/oncotarget.19588.

[95] J. Hu, Y. Gao, J. Li, et al., A novel algorithm based on bi-random walks to identify disease-related lncRNAs, BMC Bioinf. 20 (18) (2019) 569, https://doi.org/10.1186/s12859-019-3128-3.

[96] G. Xie, J. Jiang, Y. Sun, Lda-Lnsubrw, lncRNA-disease association prediction based on linear neighborhood similarity and unbalanced bi-random walk, IEEE ACM Trans. Comput. Biol. Bioinf (2020), https://doi.org/10.1109/TCBB.2020.3020595, 1-1.

[97] A. Ezzat, P. Zhao, M. Wu, et al., Drug-target interaction prediction with graph regularized matrix factorization, IEEE ACM Trans. Comput. Biol. Bioinf 14 (3) (2017) 646–656, https://doi.org/10.1109/TCBB.2016.2530062.

[98] G. Xie, C. Wu, G. Gu, et al., HAUBRW: Hybrid algorithm and unbalanced bi-random walk for predicting lncRNA-disease associations, Genomics 112 (6) (2020) 4777–4787, https://doi.org/10.1016/j.ygeno.2020.08.024.

[99] M. Sumathipala, E. Maiorino, S.T. Weiss, et al., Network diffusion approach to predict LncRNA disease associations using multi-type biological networks: LION, Front. Physiol. 10 (2019) 888, https://doi.org/10.3389/fphys.2019.00888.

[100] X. Zhao, Y. Yang, M. Yin, MHRWR: prediction of lncRNA-disease associations based on multiple heterogeneous networks, IEEE ACM Trans. Comput. Biol. Bioinf 18 (6) (2021) 2577–2585, https://doi.org/10.1109/TCBB.2020.2974732.

[101] X.-N. Fan, S.-W. Zhang, S.-Y. Zhang, et al., Prediction of lncRNA-disease associations by integrating diverse heterogeneous information sources with RWR algorithm and positive pointwise mutual information, BMC Bioinf. 20 (1) (2019) 87, https://doi.org/10.1186/s12859-019-2675-y.

[102] V. Gligorijević, M. Barot, R. Bonneau, deepNF: deep network fusion for protein function prediction, Bioinformatics 34 (22) (2018) 3873–3881, https://doi.org/10.1093/bioinformatics/bty440.

[103] L. Wang, M. Shang, Q. Dai, et al., Prediction of lncRNA-disease association based on a Laplace normalized random walk with restart algorithm on heterogeneous networks, BMC Bioinf. 23 (1) (2022) 5, https://doi.org/10.1186/s12859-021-04538-1.

[104] L. Wang, Y. Xiao, J. Li, et al., IIRWR: internal inclined random walk with restart for LncRNA-disease association prediction, IEEE Access 7 (2019) 54034–54041, https://doi.org/10.1109/ACCESS.2019.2912945.

[105] J. Li, H. Zhao, Z. Xuan, et al., A novel approach for potential human LncRNA-disease association prediction based on local random walk, IEEE ACM Trans. Comput. Biol. Bioinf 18 (3) (2021) 1049–1059, https://doi.org/10.1109/TCBB.2019.2934958.

[106] J. Zhang, Z. Zhang, Z. Chen, et al., Integrating multiple heterogeneous networks for novel LncRNA-disease association inference, IEEE ACM Trans. Comput. Biol. Bioinf 16 (2) (2019) 396–406, https://doi.org/10.1109/TCBB.2017.2701379.

[107] P. Ping, L. Wang, L. Kuang, et al., A novel method for LncRNA-disease association prediction based on an lncRNA-disease association network, IEEE ACM Trans. Comput. Biol. Bioinf 16 (2) (2019) 688–693, https://doi.org/10.1109/TCBB.2018.2827373.

[108] X. Yang, L. Gao, X. Guo, et al., A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases, PLoS One 9 (1) (2014), e87797, https://doi.org/10.1371/journal.pone.0087797.

[109] L. Ding, M. Wang, D. Sun, et al., TPGLDA: novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph, Sci. Rep. 8 (1) (2018) 1065, https://doi.org/10.1038/s41598-018-19357-3.

[110] Q. Wang, G. Yan, Idlda, An improved diffusion model for predicting LncRNA–disease associations, Front. Genet. 10 (2019), https://doi.org/10.3389/fgene.2019.01259.

[111] G. Li, J. Luo, C. Liang, et al., Prediction of LncRNA-disease associations based on network consistency projection, IEEE Access 7 (2019) 58849–58856, https://doi.org/10.1109/ACCESS.2019.2914533.

[112] Y. Zhang, M. Chen, A. Li, et al., LDAI-ISPS: LncRNA-disease associations inference based on integrated space projection scores, Int. J. Mol. Sci. 21 (4) (2020) 1508, https://doi.org/10.3390/ijms21041508.

[113] Fu T-y, Lee W-C, Lei Z. HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning, Proceedings of the 2017 ACM on Conference on Information and Knowledge Management 2017:1797-1806. http://doi.org/10.1145/3132847.3132953.

[114] L. Deng, W. Li, J. Zhang, LDAH2V: exploring meta-paths across multiple networks for lncRNA-disease association prediction, IEEE ACM Trans. Comput. Biol. Bioinf 18 (4) (2021) 1572–1581, https://doi.org/10.1109/TCBB.2019.2946257.

[115] J. Li, J. Li, M. Kong, et al., SVDNVLDA: predicting lncRNA-disease associations by Singular Value Decomposition and node2vec, BMC Bioinf. 22 (1) (2021) 538, https://doi.org/10.1186/s12859-021-04457-1.

[116] G. Fu, J. Wang, C. Domeniconi, et al., Matrix factorization-based data fusion for the prediction of lncRNA-disease associations, Bioinformatics 34 (9) (2017) 1529–1537, https://doi.org/10.1093/bioinformatics/btx794.

[117] Y. Wang, G. Yu, J. Wang, et al., Weighted matrix factorization on multi-relational data for LncRNA-disease association prediction, Methods 173 (2020) 32–43, https://doi.org/10.1016/j.ymeth.2019.06.015.

[118] Z. Xuan, J. Li, J. Yu, et al., A probabilistic matrix factorization method for identifying lncRNA-disease associations, Genes 10 (2) (2019) 126, https://doi.org/10.3390/genes10020126.

[119] J.X. Liu, Z. Cui, Y.L. Gao, et al., WGRCMF: a weighted graph regularized collaborative matrix factorization method for predicting novel LncRNA-disease associations, IEEE J. Biomed. Health Inform. 25 (1) (2021) 257–265, https://doi.org/10.1109/JBHI.2020.2985703.

[120] M.-N. Wang, Z.-H. You, L. Wang, et al., LDGRNMF: LncRNA-disease associations prediction based on graph regularized non-negative matrix factorization, Neurocomputing 424 (2021) 236–245, https://doi.org/10.1016/j.neucom.2020.02.062.

[121] C. Lu, M. Yang, F. Luo, et al., Prediction of lncRNA-disease associations based on inductive matrix completion, Bioinformatics 34 (19) (2018) 3357–3364, https://doi.org/10.1093/bioinformatics/bty327.

[122] C. Lu, M. Yang, M. Li, et al., Predicting human lncRNA-disease associations based on geometric matrix completion, IEEE J. Biomed. Health Inform. 24 (8) (2020) 2420–2429, https://doi.org/10.1109/JBHI.2019.2958389.

[123] L. Gao, H. Cui, T. Zhang, et al., Prediction of drug-disease associations by integrating common topologies of heterogeneous networks and specific topologies of subnets, Briefings Bioinf. 23 (1) (2022) bbab467, https://doi.org/10.1093/bib/bbab467.

[124] P. Xuan, L. Gao, N. Sheng, et al., Graph convolutional autoencoder and fully-connected autoencoder with attention mechanism based method for predicting drug-disease associations, IEEE J. Biomed. Health Inform. 25 (5) (2021) 1793–1804, https://doi.org/10.1109/JBHI.2020.3039502.

[125] J. Peng, W. Hui, Q. Li, et al., A learning-based framework for miRNA-disease association identification using neural networks, Bioinformatics 35 (21) (2019) 4364–4371, https://doi.org/10.1093/bioinformatics/btz254.

[126] L. Huang, S. Jiao, S. Yang, et al., LGFC-CNN: prediction of lncRNA-protein interactions by using multiple types of features through deep learning, Genes 12 (11) (2021), https://doi.org/10.3390/genes12111689.

[127] S. Yang, Y. Wang, Y. Lin, et al., LncMirNet: predicting LncRNA-miRNA interaction based on deep learning of ribonucleic acid sequences, Molecules 25 (19) (2020), https://doi.org/10.3390/molecules25194372.

[128] J. Hu, Y. Gao, J. Li, et al., Deep learning enables accurate prediction of interplay between lncRNA and disease, Front. Genet. 10 (2019), https://doi.org/10.3389/fgene.2019.00937.

[129] M. Zeng, C. Lu, Z. Fei, et al., DMFLDA: a deep learning framework for predicting lncRNA-disease associations, IEEE ACM Trans. Comput. Biol. Bioinf 18 (6) (2021) 2353–2363, https://doi.org/10.1109/TCBB.2020.2983958.

[130] M. Zeng, C. Lu, F. Zhang, et al., SDLDA: lncRNA-disease association prediction based on singular value decomposition and deep learning, Methods 179 (2020) 73–80, https://doi.org/10.1016/j.ymeth.2020.05.002.

[131] P. Xuan, L. Jia, T. Zhang, et al., LDAPred: a method based on information flow propagation and a convolutional neural network for the prediction of disease-associated lncRNAs, Int. J. Mol. Sci. 20 (18) (2019) 4458, https://doi.org/10.3390/ijms20184458.

[132] H. Wei, Q. Liao, B. Liu, iLncRNAdis-FB: identify lncRNA-disease associations by fusing biological feature blocks through deep neural network, IEEE ACM Trans. Comput. Biol. Bioinf 18 (5) (2021) 1946–1957, https://doi.org/10.1109/TCBB.2020.2964221.

[133] Y. Zhang, F. Ye, X. Gao, M.C.A. Net, Multi-feature coding and attention convolutional neural network for predicting lncRNA-disease association, IEEE ACM Trans. Comput. Biol. Bioinf (2021), https://doi.org/10.1109/TCBB.2021.3098126, 1-1.

[134] P. Xuan, N. Sheng, T. Zhang, et al., CNNDLP: a method based on convolutional autoencoder and convolutional neural network with adjacent edge attention for predicting lncRNA-disease associations, Int. J. Mol. Sci. 20 (17) (2019) 4260, https://doi.org/10.3390/ijms20174260.

[135] Y. Zhang, F. Ye, D. Xiong, et al., LDNFSGB: prediction of long non-coding rna and disease association using network feature similarity and gradient boosting, BMC Bioinf. 21 (1) (2020) 377, https://doi.org/10.1186/s12859-020-03721-0.

[136] Z.-H. Guo, Z.-H. You, Y.-B. Wang, et al., A learning-based method for LncRNA-disease association identification combing similarity information and rotation forest, iScience 19 (2019) 786–795, https://doi.org/10.1016/j.isci.2019.08.030.

[137] X. Su, Z. You, H. Yi, Prediction of LncRNA-disease associations based on network representation learning, IEEE Int. Conf. Bioinform. Biomed. (2020) 1805–1812, https://doi.org/10.1109/BIBM49941.2020.9313139, 2020.

[138] A. Grover, J. Leskovec, node2vec: scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016, pp. 855–864, https://doi.org/10.1145/2939672.2939754.

[139] N. Sheng, H. Cui, T. Zhang, et al., Attentional multi-level representation encoding based on convolutional and variance autoencoders for lncRNA-disease association prediction, Briefings Bioinf. 22 (3) (2020) bbaa067, https://doi.org/10.1093/bib/bbaa067.

[140] Q. Yang, X. Li, BiGAN, LncRNA-disease association prediction based on bidirectional generative adversarial network, BMC Bioinf. 22 (1) (2021) 357, https://doi.org/10.1186/s12859-021-04273-7.

[141] T.N. Kipf, MJapa Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the 5th International Conference on Learning Representations, 2016. https://openreview.net/forum?id=SJU4ayYgl.

[142] T.N. Kipf, M. Welling, Variational Graph Auto-Encoders, 2016, https://doi.org/10.48550/arXiv.1611.07308 arXiv:1611.07308.

[143] P. Veličković, G. Cucurull, A. Casanova, et al., Graph attention networks, in: Proceedings of the 6th International Conference on Learning Representations, 2017. https://openreview.net/forum?id=rJXMpikCZ.

[144] P. Xuan, S. Pan, T. Zhang, et al., Graph convolutional network and convolutional neural network based method for predicting lncRNA-disease associations, Cells 8 (9) (2019) 1012, https://doi.org/10.3390/cells8091012.

[145] Q.-W. Wu, J.-F. Xia, J.-C. Ni, et al., GAERF: predicting lncRNA-disease associations by graph auto-encoder and random forest, Briefings Bioinf. 22 (5) (2021) bbaa391, https://doi.org/10.1093/bib/bbaa391.

[146] Q.W. Wu, R.F. Cao, J. Xia, et al., Extra trees method for predicting LncRNA-disease association based on multi-layer graph embedding aggregation, IEEE ACM Trans. Comput. Biol. Bioinf (2021), https://doi.org/10.1109/TCBB.2021.3113122, 1-1.

[147] N. Sheng, L. Huang, Y. Wang, et al., Multi-channel graph attention autoencoders for disease-related lncRNAs prediction, Briefings Bioinf. 23 (2) (2022) bbab604, https://doi.org/10.1093/bib/bbab604.

[148] W. Lan, X. Wu, Q. Chen, et al., GANLDA: graph attention network for lncRNA-disease associations prediction, Neurocomputing 469 (2022) 384–393, https://doi.org/10.1016/j.neucom.2020.09.094.

[149] P. Xuan, L. Zhan, H. Cui, et al., Graph triple-attention network for disease-related LncRNA prediction, IEEE J. Biomed. Health Inform. (2021), https://doi.org/10.1109/JBHI.2021.3130110, 1-1.

[150] A.B.O.V. Silva, E.J. Spinosa, Graph Convolutional Auto-Encoders for predicting novel lncRNA-Disease associations, IEEE ACM Trans. Comput. Biol. Bioinf (2021), https://doi.org/10.1109/TCBB.2021.3070910, 1-1.

[151] X. Wu, W. Lan, Q. Chen, et al., Inferring LncRNA-disease associations based on graph autoencoder matrix completion, Comput. Biol. Chem. 87 (2020), 107282, https://doi.org/10.1016/j.compbiolchem.2020.107282.

[152] Y. Fan, M. Chen, X. Pan, GCRFLDA: scoring lncRNA-disease associations using graph convolution matrix completion with conditional random field, Briefings Bioinf. 23 (1) (2021) bbab361, https://doi.org/10.1093/bib/bbab361.

[153] X. Zhao, X. Zhao, M. Yin, Heterogeneous graph attention network based on meta-paths for lncRNA-disease association prediction, Briefings Bioinf. 23 (1) (2021) bbab407, https://doi.org/10.1093/bib/bbab407.

[154] Z. Shi, H. Zhang, C. Jin, et al., A representation learning model based on variational inference and graph autoencoder for predicting lncRNA-disease associations, BMC Bioinf. 22 (1) (2021) 136, https://doi.org/10.1186/s12859-021-04073-z.