Computational Methods and Applications for Identifying Disease-Associated IncRNAs as Potential Biomarkers and Therapeutic Targets

Congcong Yan,1 Zicheng Zhang,1 Siqi Bao,1 Ping Hou,1 Meng Zhou,1 Chongyong Xu,2 and Jie Sun1

1School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325027, P.R. China; 2Department of Radiology, The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou 325027, P.R. China

Long non-coding RNAs (IncRNAs) have been recognized as critical components of a broad genomic regulatory network and play pivotal roles in physiological and pathological processes. Identification of disease-associated IncRNAs is becoming increasingly crucial for fundamentally improving our understanding of molecular mechanisms of disease and developing novel biomarkers and therapeutic targets. Considering lower efficiency and higher time and labor cost of biological experiments, computer-aided inference of disease-associated RNAs has become a promising avenue for facilitating the study of IncRNA functions and provides complementary value for experimental studies. In this study, we first summarize data and knowledge resources publicly available for the study of IncRNA-disease associations. Then, we present an updated systematic overview of dozens of computational methods and models for inferring IncRNA-disease associations proposed in recent years. Finally, we explore the perspectives and challenges for further studies. Our study provides a guide for biologists and medical scientists to look for dedicated resources and more competent tools for accelerating the unraveling of disease-associated IncRNAs.

INTRODUCTION

Advances in genomic and transcriptional analyses have markedly expanded our knowledge of the genomic dark matter and revealed that only about 2% of the human genome encodes protein-coding genes, and the vast majority are transcribed into non-coding RNAs (ncRNAs).1–2 Long ncRNAs (IncRNAs), constituting the biggest class of ncRNAs, were arbitrarily defined as ncRNAs with more than 200 nt in length.3–4 There is increasing evidence that IncRNAs are hidden critical components of a broad genomic regulatory network involved in gene transcription, epigenetic regulation, and post-transcriptional regulation, and they thus play pivotal roles in a wide variety of biological processes.5–6 A large number of IncRNAs with oncogenic or tumor-suppressor function have been found, highlighting the emerging role of IncRNAs in complex diseases.7–9

Identification of disease genes is a significant and challenging task in biomedical research. Systematic identification of disease-associated IncRNAs not only contributes to our understanding of the underlying molecular mechanisms of complex diseases, but it also has been shown to have the intrinsic advantage over protein-coding genes in disease diagnosis, prognosis, and treatment.10–17 Despite increasing efforts that have been taken to explore the function of IncRNAs and their implications in diseases, the vast majority of IncRNAs are not functionally well characterized, and their associations with diseases remain unknown. Low-throughput biological experiments in vivo or in vitro have been extensively used to dissect disease-related IncRNAs. Although the exact association between disease and IncRNAs, as well as the pathogenic mechanism of IncRNAs, could be elucidated through in vivo or in vitro experiments, these low-throughput biological experiments tend to be time-consuming, expensive, and inefficient when faced with tens of thousands of IncRNAs with unknown function. With the application of high-throughput technologies (e.g., microarray and next-generation sequencing) to disease transcriptomes, a large number of dysregulated IncRNAs were identified to be associated with disease. However, results of high-throughput technologies contained much noise, and most of the dysregulated IncRNAs tend to be unrelated rather than causal IncRNAs because the aberrant expression in disease is not sufficient evidence to confirm the causal association between IncRNAs and diseases. With large-scale available heterogeneous data resources of IncRNAs and diseases, great efforts have been devoted to system-level inference of IncRNA-disease association through computational or bioinformatics approaches, which constitute a complement to wet-lab experiments.18,19

In this study, we present an overview of the computer-aided inference of the IncRNA-disease association. First, data sources accessible to the IncRNA-disease association study are introduced in detail. Second, novel computational methods and software...
tools, as well as their application in lncRNA-disease association prediction, are summarized and reviewed. Finally, we explore the future perspectives and challenges in this field.

RESULTS
In this section, we reviewed dozens of novel computational methods in inferring the lncRNA-disease association proposed in recent years. Based on the core idea of the algorithm, these computational methods could be divided into four categories: (1) matrix completion-based methods, (2) recommendation algorithm-based methods, (3) resource allocation-based methods, and (4) integration-based methods.

Matrix Completion-Based Methods
The universal characteristic of the matrix completion-based methods is to complete the data set with missing values in the form of a matrix. As shown in Figure 1, three matrices, including the lncRNA-disease association matrix, lncRNA-lncRNA matrix, and disease-disease matrix, were obtained. Then feature extraction was accomplished based on the above three matrices to obtain lncRNA feature vectors and disease feature vectors. Finally, matrix completion methods were performed on the lncRNA-disease association matrix to acquire the lncRNA-disease association.

Li et al.19 developed a computational model of faster randomized matrix completion for latent disease lncRNA association (named FRMCLDA) that used the faster singular value threshold (fSVT) algorithm to predict lncRNA-disease associations based on the idea of matrix completion. FRMCLDA uses the disease similarity matrix, lncRNA similarity matrix, lncRNA-disease association matrix, and transpose matrix of the association matrix to construct the adjacency matrix, which improves the prediction performance by fitting the adjacency matrix. LDAPM and SIMCLDA also use a matrix completion approach, but the difference is that LDAPM denotes the approximated matrix as the multiplication of the two matrices.20,21 TSSR exploits learned representation matrices as feature matrices to reconstruct the original matrix.22

Resource-Allocation-Based Methods
Resource allocation is the allocation of available resources to each node. To predict the lncRNA-disease association, resource allocation is based on the initial value of the multi-data source matrices as a possible value for the relationship between nodes. The process is demonstrated in Figure 2. Resource allocation-based methods are built on data from multiple sources, such as lncRNA-disease association,
miRNA-disease association, miRNA-IncRNA association, and so on. The heterogeneous multilayer network is constructed, and the edges are weighted according to the corresponding values of the matrix. The IncRNA-disease scoring matrix was produced by post-processing resource allocation on the heterogeneous network (Table 2).

Resource allocation has been implemented in more than a dozen methods for predicting the IncRNA-disease association. Fan et al. proposed a computational model of IDHI-MIRW by integrating diverse heterogeneous information sources with positive pointwise mutual information and random walk with restart algorithm. Xiao et al. developed BPLLDA to predict IncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network. Zhang et al. proposed a rule-based inference method on the linked tripartite network, which was constructed by integrating heterogeneous data with deep learning algorithms. Some other information had also been introduced to allocate resources for improving prediction performance.

The first category of recommendation algorithm-based methods is the content-based recommendation. Content-based recommendation refers to the recommendation of similar nodes of previous related nodes for this node. For example, NCPLDA, proposed by Li et al., measured the IncRNA-disease association score based on network consistency projection. The second category of recommendation algorithm-based methods is based on collaborative filtering. Collaborative filtering refers to adding other nodes similar to the nodes and using the information of these nodes to make inferences. CFNBC, developed by Yu et al., and NBCLDA, proposed by Yu et al., both used collaborative filtering on multi-data matrices to uncover a new relationship between IncRNA and disease and then took advantage of a naive Bayesian classifier to determine whether there is an association between IncRNA and disease in the set of the IncRNA-associated node and disease-associated node. A similarity correlation fusion method introduced neighbor information and then was used to predict the association by making the original matrix fit as well as possible in ILDMSF and SKF-LDA. BLM-NPAI, developed by Cui et al., introduced the nearest profile to get the final prediction results after constructing the local model of IncRNA and disease. Another method, proposed by Ping et al., measured the one-step neighbor of a node based on simrank measure when there was no common neighbor. DCSLDA, proposed by Zhao et al., calculated the shortest path between IncRNA and disease and the distance correlation coefficient to construct the final matrix. The third category of recommendation algorithm-based methods is based on matrix factorization, including MFLDA, WMFLDA, and PMFILDA. The three methods above are based on matrix factorization, and the difference is that the latter two add weight and probability separately. Li et al. developed a computational method, DNILMF-LDA, that is

### Table 1. Overview of Categories and Corresponding Method/Tool for Acquiring IncRNA-IncRNA Association

| Categories               | Method/Tool            | Data Types               | Data Resources                      | References                 |
|--------------------------|------------------------|--------------------------|-------------------------------------|----------------------------|
| Sequence similarity      | EMBOSS Needle tool     | IncRNA sequence          | LncRNADisease, UCSC, LNCipedia      | Needleman and Wunsch       |
| Functional similarity    | LNCSIM                 | IncRNA-disease association, MeSH descriptors | LncRNADisease, Lnc2Cancer, MNDR, MeSH | Chen et al.                |
| Functional similarity    | ILNCSIM                | IncRNA-disease association, MeSH descriptors | MNDR, Lnc2Cancer, MeSH             | Huang et al.               |
| Functional similarity    | NA                     | IncRNA-gene association, protein-protein interaction | LncRNA2Target, StarBase, HPRD     | Paik et al.                |
| Functional similarity    | NA                     | IncRNA-miRNA association  | StarBase                            | Zhao et al.                |
| Expression similarity    | Spearman/Pearson       | IncRNA expression profiles | Array Express, UCSC Genome Bioinformatics | Chen and Yan              |
| Cosine similarity        | cosine similarity      | IncRNA-disease association | MNDR, Lnc2Cancer, LncRNADisease     | Cheng et al.               |

NA, not applicable.
anchored in the neighborhood regularized logistic matrix factorization and optimizes the above parameters to predict interaction probabilities. NNLDA, determined by Hu et al., solved some of the disadvantages of traditional matrix factorization by changing the training method and the loss function and adding a fully connected layer. DSCMF combines matrix factorization and collaborative filtering to predict associations efficiently by introducing neighbor information.

**Multi-model Integration-Based Methods**

Multi-model integration methods have also been proposed to overcome the shortcomings of the single model and improve prediction performance (Table 4). A combination of matrix completion ideas and recommendation system ideas was applied in three models to predict potential lncRNA-disease associations, including LDASR, ECLDA, and weighted bagging LightGBM model. Three methods (CNNLDA, CNNDLP, and GCNLDA) were developed by Xuan et al. to construct the final module through the integration of the convolutional module and attention module. LDAPred, proposed by Xuan et al., introduced the convolutional neural network based on the integration of resource allocation and matrix completion.

**DISCUSSION**

During the past decade, it has been well documented that lncRNAs play a critical role in nearly all biological processes and have become an emerging paradigm of human disease research. Identification of disease-associated lncRNAs is becoming increasingly important for fundamentally improving our understanding of molecular mechanisms and developing novel therapeutic targets, and thus has attracted more and more attention in the scientific community and is becoming one of the hotspots in medical research. Although current experimental studies in vitro and in vivo could directly link identified lncRNAs with disease phenotypes, they are affected by the limitation of lower efficiency and higher time and labor cost. Taking into account the limitations of experimental studies, high-throughput technologies were then implemented, leading to exponential growth in the number of dysregulated lncRNAs in diseases. However, the aberrant expression of lncRNAs is not sufficient evidence for ascribing to them a functional role in disease. Therefore, efficient and accurate identification and functional elucidation of disease-associated lncRNAs are in their infancy and remain a major challenge. With the rapidly increasing quantity and quality of bioinformatics databases and resources in lncRNAs and diseases, computer-aided inference of disease-associated RNAs has become a promising avenue for facilitating the unraveling of the functional role of lncRNAs in diseases and provides complementary value for experimental studies. A large number of computational models,
algorithms, and tools have been developed and proposed, compensating for this dearth.

In this work, we first summarize data and knowledge sources available for the lncRNA-disease association study, which contains databases of lncRNAs, diseases, and known lncRNA-disease associations. We then present a detailed overview of previously proposed computational methods for inferring lncRNA-disease associations. Based on the core idea implemented, these computational methods can be divided into four categories: (1) matrix completion-based methods, (2) recommendation algorithm-based methods, (3) resource allocation-based methods, and (4) multi-model integration-based methods. Despite that the performance of each computational method is very great according to the reports in their own studies, one emerging critical issue is that most of these methods used different data sources as their training dataset and carried out cross-validation on their dataset, lacking benchmark performance evaluation. These computational methods have distinct limitations and weaknesses, which are noted at follows. First, matrix completion-based methods considered the feature vectors of lncRNA and disease to improve the accuracy of the prediction. However, these algorithms hold the disadvantage of poor robustness. The ranks of diverse datasets are likely to vary widely. Second, because experimentally verified lncRNA-disease associations are still too incomplete, resource allocation-based methods need to consider the prediction of separate nodes or integrate additional biological information. However, although this can improve prediction accuracy, some interactions from other databases may contain some noise to interfere with prediction results. In addition, recommendation algorithm-based methods are stated separately. Content-based recommendation models only need node prior knowledge, but new levels of disease-lncRNA associations cannot be recognized. Although collaborative filtering-based recommendation methods complement this shortcoming, the spare lncRNA-disease association matrix is harmful to the recommendation, and the complexity and time cost of the algorithm will sharply increase when the amount of data is too large. Additionally, matrix factorization-based recommendation methods reduce space complexity by mapping a matrix to a product of low-dimensional matrices. These methods also make it easy to add additional data from different sources and use the intrinsic structure. Matrix factorization-based methods also have the same weaknesses as collaborative filtering-based recommendation methods. The above computational approaches can thus complement each other. Therefore, multi-model integration-based methods were proposed to achieve better performance when investigating the association between lncRNAs and diseases. Finally, only several computational approaches have been developed as online web tools, and most are still theoretical studies that hampered their use for biologists and medical scientists.

With the rapidly increasing knowledge for the functional mechanism of lncRNAs, several challenges that would be helpful to improve the accuracy and practicality of the predictors could be highlighted. It is well known that the majority of lncRNAs exhibited precise subcellular localization, thus performing regulatory roles in a spatiotemporal manner. Therefore, some interactions between lncRNAs and other biological molecules (DNA, RNA, and proteins) used in previous predictors are derived from prediction and do not exist in the real biological world. Therefore, co-localization information of lncRNAs and other biological molecules should be considered. Additionally, it has also been observed that lncRNAs were expressed in highly cell type-specific, tissue-specific, and disease-specific manners. Therefore, more molecular information in the appropriate biological contexts should be introduced into predictors that are more suitable for the specific disease. Finally, the prediction results of these computational approaches are only descriptive associations, and the specific association type (e.g., casual or non-causal association of lncRNA with the disease) is still a challenging task and needs to be answered. The implementation of efficient and reliable computational predictions, together with systematic biological experiments, will greatly accelerate the study of lncRNA functions and mechanisms in physiological and pathological conditions.

**MATERIALS AND METHODS**

**Databases and Knowledge Bases**

NONCODE ([http://www.noncode.org/](http://www.noncode.org/)) collects and integrates data from PubMed and other resources via text mining. Users can use
CNCI to predict their protein-encoding potential and display the results of functional annotation and enrichment through the ncFANs online website. The current version covers expression, function, sequence, structure, disease relevance of lncRNA, and other factors. Compared with other lncRNA databases, NONCODE stores more information about lncRNA transcripts and unique annotations.

LncRBase (http://bicresources.jcbose.ac.in/zhumur/lncrbase) is an annotation database resource for analyzing lncRNA functions based on feature sequences. It has recorded transcript information about 133,361 human lncRNA entries and 83,201 mouse lncRNA entries. Information about the lncRNA subtypes and small ncRNA-lncRNA associations is included. The database also provides microarray probes mapped to specific lncRNAs and expression in tissues.

LncBook (http://bigd.big.ac.cn/lncbook) collects information on 268,848 experimentally verified and predicted lncRNAs (including 1,867 functional lncRNAs) and includes information on related functions, diseases, expressions, methylation, mutations, and miRNA interactions (via software prediction). The team developed a database called LncRNAWiki, which is an integrated database. LncRNAWiki has set up a model of collaborative annotation. Then LncBook has been established to organize large-scale annotations systematically as a complement to LncRNAWiki.

MONOCLdb (https://www.monocldb.org/) contains 20,728 lncRNAs from the sequencing of virus-infected lungs of eight respiratory-infected mice, of which 5,329 were differentially expressed. These differentially expressed lncRNAs are annotated by different methods (enrichment methods, as well as module-based and rank-based methods). The correlation score of lncRNA expression profiles and six phenotypic data were determined as pathogenic associations.

lncRNome (http://genome.igib.res.in/lncRNome) is a comprehensive database of human lncRNAs, which collects information on annotation, sequence, structure, interacting proteins, genomic variations, conservation, and epigenetic modifications for more than 18,000 lncRNAs. Annotation is manually curated from literature and databases, including associated diseases, related literature, and the mapping of disease-associated variation in lncRNA gene loci.

LncRNASNP (http://bioinfo.life.hust.edu.cn/lncRNASNP/) mainly sorts information on the single nucleotide polymorphism (SNP) loci located on the lncRNA gene in humans and mice. The cancer mutations in lncRNA transcripts and lncRNA expression in cancer, the predicted interactions of miRNAs and associated diseases, and the impact of variations on lncRNA structures were integrated into LncRNASNP. This database also collects experimentally verified and predicted disease-lncRNA associations in humans.
LncRNADisease (http://www.rnanut.net/lncrnadisease/) is an open-access database that has been updated to version 2.0.27 As one of the more commonly used databases, LncRNA Disease 2.0 provides 10,564 experimental lncRNA-disease associations and 195,395 computational lncRNA-disease associations in four species. Additionally, a confidence score was obtained for each pair of lncRNA-disease associations based on known experimental information. LncRNADisease also collects lncRNA regulatory networks.

Lnc2Cancer (http://www.bio-bigdata.net/lnc2cancer) collected 4,989 comprehensive experimentally supported associations between 1,614 lncRNAs and 165 human cancers.28,29 These records were built through text mining on the PubMed database. The database consists of three classifications of relationships between lncRNAs and cancers: circulating, drug-resistant, and prognostic-related lncRNAs. Additionally, it collects transcription factor (TF), microRNA (miRNA), variant, and methylation molecular information on the regulation of lncRNAs.

MNDR (http://www.rna-society.org/mndr/) is built through manual curation of scientific literature.30 The current release (MNDR v2.0) has recorded 261,042 entries including six species and 1,416 diseases from 26,600 studies. Detailed and comprehensive annotations for lncRNA-disease associations are presented at the bottom of each record, including the data from articles and evidence support.

EVlncRNAs (http://biophy.dzu.edu.cn/EVLncRNAs) is a database that intends to include all lncRNA-disease associations that are validated by low-throughput experiments.31 The database includes 1,543 lncRNAs from 77 species, 886 of which are associated with 338 diseases, along with experimental information. For other lncRNAs that are not associated with diseases, their functional information is collected.

NSDNA (http://www.bio-bigdata.net/nsdna/) is an online knowledge base of ncRNA-nervous system disease (NSD) associations.32 It contains 24,713 entries of associations covering 142 NSDs and 8,593 ncRNAs from more than 1,300 articles, of which 4,608 lncRNAs-NSDs are included. Users can brows by ncRNAs, diseases, species, or tissue name. Also, if searching for data, users can select low-throughput or high-throughput experimental data or both.

Nc2Eye (http://nc2eye.bio-data.cn/) is the first high-quality manually curated ncRNAomics knowledge base associated with eye disease and includes 1,147 lncRNA-associated entries.33

### Computational Methods for Acquiring lncRNA-lncRNA Associations

Most of the computational methods for inferring lncRNA-disease association are based on lncRNA-lncRNA association data. Therefore, acquiring a high-quality lncRNA-lncRNA association is critical for improving performance in predicting the lncRNA-disease association. We have summarized the currently available computational methods for acquiring the lncRNA-lncRNA association. In general, these methods could be basically divided into four categories (Table 1): sequence similarity-based methods, functional similarity-based methods, cosine similarity-based methods, and expression-based methods.

| Method Name | Computational Principle | Data Types | Available Tool (Package or Code) | References |
|-------------|-------------------------|------------|---------------------------------|------------|
| SIMCLDA     | inductive matrix completion, singular value decomposition | lncRNA-disease association, lncRNA-lncRNA association, disease-disease association | code (https://github.com/bioinformaticsCSU/SIMCLDA) | Lu et al.47 |
| LDAPM       | inductive matrix completion, singular value decomposition | lncRNA-disease association, lncRNA-lncRNA association, disease-disease association | NA | Fraidouni and Zaruba48 |
| FRMCLDA     | faster randomized matrix completion, faster singular value threshold | lncRNA-disease association, lncRNA-lncRNA association, disease-disease association | code (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6749816/bin/Table_7.docx) | Li et al.36 |
| TSSR        | sparse self-representation | lncRNA-disease association, lncRNA-lncRNA association, disease-disease association | code (https://github.com/Oyl-CityU/TSSR) | Ou-Yang et al.39 |

NA, not applicable.
Functional Similarity-Based Methods

Based on the assumption that if lncRNA-related molecules have a similar function, lncRNA functions are identical, several functional similarity-based computational methods were developed. Chen et al. 35 proposed a method, named LNCSIM, to measure the semantic similarity of lncRNAs’ associated two groups of diseases. They integrated two semantic similarity models to achieve better performance. Two models both collected diseases’ MeSH descriptors and constructed a directed acyclic graph (DAG). Then, the contribution of disease term \( t \) to disease \( A \) was calculated, which is also the difference between these two models. In the first model, it was calculated as Equation 2. Since the contribution of other diseases to the semantic value of the disease decreases with the increase of the distance between this disease and disease \( A \), the decay factor is added. For the second model, diseases that appear in DAG(A) and are less common in other diseases, DAGs have a more significant contribution that can be calculated in Equation 3:

\[
C_2(A) = - \log \frac{\text{the number of DAGs including } t}{\text{the number of diseases}}
\]

Therefore, the semantic value of disease \( A \) is defined as the sum of contributions from ancestral diseases and disease \( A \) itself:

\[
C_1(A) = \sum_{t \in \text{DAG}(A)} C_1(t)
\]

Thus, the semantic similarity between two diseases \( A \) and \( B \) is calculated based on the common nodes of DAG(A) and DAG(B):

\[
SS(A, B) = \frac{\sum_{t \in \text{DAG}(A) \cap \text{DAG}(B)} (C_2(A) + C_2(B))}{C(A) + C(B)}
\]

Finally, lncRNA functional similarity was obtained by calculating the average of the similarities between the two groups of diseases.

Huang et al. 36 proposed an improved model called ILNCSIM, which introduced information content (IC) and focused on the hierarchical structure of disease DAGs. First, the information

\[
C_2(t) = - \log \frac{\text{the number of DAGs including } t}{\text{the number of diseases}}
\]

Therefore, the semantic value of disease \( A \) is defined as the sum of contributions from ancestral diseases and disease \( A \) itself:

\[
C_1(A) = \sum_{t \in \text{DAG}(A)} C_1(t)
\]

Thus, the semantic similarity between two diseases \( A \) and \( B \) is calculated based on the common nodes of DAG(A) and DAG(B):

\[
SS(A, B) = \frac{\sum_{t \in \text{DAG}(A) \cap \text{DAG}(B)} (C_2(A) + C_2(B))}{C(A) + C(B)}
\]

Finally, lncRNA functional similarity was obtained by calculating the average of the similarities between the two groups of diseases.
Content value was calculated. Information content of disease term \( a \) is defined as the negative log-likelihood of each term:

\[
IC(i) = - \log p(i)
\]  

(6)

Second, IC-based distances were used to calculate the most informative common ancestors (MICAs) and the most informative leaf (MIL):

\[
DIST_{IC}(i, j) = |IC(i) - IC(j)|
\]  

(7)

Third, components \( \alpha \), to measure the specificity of MICA, \( \beta \), to measure the generality of two disease terms, and \( \gamma \), to estimate the total IC-based distances between two terms and their MICA, are computed:

\[
\alpha = DIST_{IC} (\text{ROOT, MICA}),
\]

(8)

\[
\beta = \frac{DIST_{IC} (\text{node}_i, \text{MIL}_i) + DIST_{IC} (\text{node}_j, \text{MIL}_j)}{2}
\]

(9)

\[
\gamma = DIST (\text{MICA, node}_i) + DIST (\text{MICA, node}_j)
\]

(10)

Fourth, based on the above three equations, to compute the semantic similarity of two diseases:

\[
SS_{\text{disease}} (\text{term}_i, \text{term}_j) = \frac{\alpha}{\alpha + \beta} + \frac{1}{1 + \gamma}
\]

(11)

Finally, IncRNA functional similarity was measured by calculating the average of the similarities between the two groups of diseases.

ICod measures disease similarity by scoring disease-related gene similarity, and researchers applied this idea to IncRNAs. The contribution value of each miRNA for IncRNA divided by the number of miRNAs associated with the process is the sum of contributions of commonly associated lncRNA-lncRNA association, which introduced lncRNA-miRNA associations and was defined as for Equation 14 and 15. The process is the sum of contributions of commonly associated miRNA divided by the number of miRNAs associated with two lncRNAs. The contribution value of each miRNA for IncRNA is computed by Equation 14. \( D(i) \) and \( D(j) \) are the number of lncRNA-related edges and IncRNA-related edges, respectively:

\[
C_C (m_i) = - \log_2 \left( \frac{\text{the number of } m_i \text{ - relate lncRNAs}}{\text{the number of lncRNAs - miRNA associations}} \right)
\]  

(14)

\[
LS(i, j) = \frac{\exp \left( \sum_{m_i \in D(i) \cap D(j)} C_C (m_i) \right)}{|D(i)| + |D(j)| - |D(i) \cap D(j)|}
\]  

(15)

Expression Similarity-Based Methods

IncRNAs are expressed in highly cell type-specific, tissue-specific, and disease-specific manners. Consequently, the expression similarity between two IncRNAs is an important point. The co-expression relationship between IncRNAs measured by a Pearson or Spearman correlation coefficient was commonly used to infer the IncRNA-IncRNA association.

Cosine Similarity-Based Methods

The concept of cosine is the origin of mathematics. Cheng et al. proposed a computational method called IntNetLncSim, which linked cosine similarity with IncRNA similarity. The similarity between two lncRNAs, \( \text{lnc}_1 \) and \( \text{lnc}_2 \), was calculated as follows:

\[
LS(\text{lnc}_1, \text{lnc}_2) = \frac{\sum_{i=1}^{d} w_{1,i} \cdot w_{2,i}}{\sqrt{\sum_{i=1}^{d} w_{1,i}^2} \cdot \sqrt{\sum_{i=1}^{d} w_{2,i}^2}}
\]

(18)

where \( w_{1,i} \) represents the vector values of IncRNA on the \( i \)th dimension.

Computational Methods for Acquiring Disease-Disease Associations

Disease-disease association data are critical data used in most of the computational methods for inferring IncRNA-disease association. In general, these methods for obtaining disease-disease association can be divided into four categories (Table 2): semantic similarity-based methods, functional similarity-based methods, Gaussian interaction profile kernel similarity-based methods, and cosine similarity-based methods.

Semantic Similarity-Based Methods

Semantic similarity between diseases is one of the commonly used methods that use mesh descriptors or disease ontology terms and determine the similarity of two disease terms based on the information content of common ancestral terms. A package named DOSE developed by Yu and Wang can calculate disease semantic similarity. Other methods also obtained the same results by mathematical formulas. The detailed algorithm is described in LNCSIM of IncRNA functional similarity.
Functional Similarity-Based Methods

Functional similarity-based methods were achieved by using the Jaccard coefficient to measure the similarity and difference of disease-related gene ontology.\(^{44}\) Disease-gene interaction and gene-gene ontology interaction were used and calculated as Equation 19. GO\(_i\) represents the gene ontology terms related to disease \(i\):

\[
DS(i, j) = \frac{|G_O_i \cap G_O_j|}{|G_O_i \cup G_O_j|}
\]

(19)

Zhao et al.\(^ {40}\) proposed a model to infer the disease-disease association in their computational method, which introduced disease-miRNA associations and was defined as Equation 20 and 21. The process is the sum of contributions of commonly associated miRNA divided by the average number of associations with the number of miRNAs associated with two diseases. The contribution value of each miRNA for a disease is computed by Equation 20. \(D(i)\) and \(D(j)\) are the number of disease-related edges and disease-related edges, respectively:

\[
C_D(m_i) = -\log\left(\frac{\text{the number of } m_i \text{ - relate diseases}}{\text{the number of disease - miRNA associations}}\right)
\]

(20)

\[
DS(i, j) = \frac{\exp\left(\sum_{m_i \in D(i) \cap D(j)} C_D(m_i)\right)}{|D(i)| + |D(j)| - |D(i) \cap D(j)|}
\]

(21)

Gaussian Interaction Profile Kernel Similarity-Based Methods

Based on the assumption that genes with similar functions tend to be associated with a similar disease, Chen and Yan\(^ {41}\) applied the Gaussian interaction profile kernel (also called the radial basis function kernel) similarity to measure disease-disease association as in Equation 22. IP\((i)\) indicates the row of disease, in the disease-IncRNA association matrix:

\[
KD(i, j) = \exp\left(-\gamma_d ||IP(i) - IP(j)||^2\right)
\]

(22)

The parameter \(\gamma_d\) controls the kernel bandwidth, which is defined as follows:

\[
\gamma_d = \gamma_d' \left(\frac{1}{\text{nd}} \sum_{i=1}^{\text{nd}} ||IP(i)||^2\right)
\]

(23)

where nd denotes the number of the contained diseases. \(\gamma_d'\) is a novel bandwidth parameter by the average number of associations with IncRNAs per disease.

Cosine Similarity-Based Methods

In a previous study, Hamaneh and Yu\(^ {45}\) linked cosine similarity with disease similarity. We used the IncRNA-disease association matrix to replace the original matrix for display. The similarity between the two diseases, \(dis_1\) and \(dis_2\), was calculated as follows:

\[
\text{LS}(\text{dis}_1, \text{dis}_2) = \frac{\sum_{i=1}^{\text{nd}} w_{1,i} \cdot w_{2,i}}{\sqrt{\sum_{i=1}^{\text{nd}} w_{1,i}^2} \cdot \sqrt{\sum_{i=1}^{\text{nd}} w_{2,i}^2}}
\]

(24)

where \(w_{i,1}\) represents the vector values of \(\text{dis}_1\) on the \(i\)th dimension.

Computational Methods for Inferring IncRNA-Disease Associations

In this section, we review dozens of novel computational methods in inferring the IncRNA-disease associations proposed in recent years. Based on the core idea of the algorithm, these computational methods can be divided into four categories: matrix completion-based methods, recommendation algorithm-based methods, resource allocation-based methods, and multi-model integration-based methods.

Matrix Completion-Based Methods

The universal characteristic of the matrix completion-based methods is to complete the dataset with missing values in the form of a matrix. As shown in Figure 1, three matrices, including the IncRNA-disease association matrix, IncRNA-IncRNA matrix, and disease-disease matrix, were obtained. Then, feature extraction is accomplished based on the above three matrices to obtain IncRNA feature vectors and disease feature vectors. Finally, matrix completion methods were conducted on the IncRNA-disease association matrix to acquire the IncRNA-disease scoring matrix based on IncRNA feature vectors and disease feature vectors (Table 3).

Li et al.\(^ {46}\) developed a computational model of faster randomized matrix completion for latent disease-IncRNA association (named FRMCLDA) that used the fSVT algorithm to predict IncRNA-disease associations based on the idea of matrix completion. FRMCLDA uses the disease similarity matrix, IncRNA similarity matrix, IncRNA-disease association matrix, and transpose matrix of the association matrix to construct the adjacency matrix, which improves the prediction performance by fitting the adjacency matrix. LDAPM and SIMCLDA also use a matrix completion approach, but the difference is that LDAPM denotes the approximated matrix as the multiplication of the two matrices.\(^ {47,48}\) Also, TSSR exploits learned representation matrices as feature matrices to reconstruct the original matrix.\(^ {49}\)

Resource Allocation-Based Methods

Resource allocation is used to allocate available resources to each node. To predict the IncRNA-disease association, resource allocation is based on the initial value of the multi-data source matrices as a possible value for the relationship between nodes. The process is demonstrated in Figure 2. Resource allocation-based methods are built on data from multiple sources, such as IncRNA-disease association, miRNA-disease association, miRNA-IncRNA association, and so forth. The heterogeneous multilayer network is constructed, and the edges are weighted according to the corresponding values of the
matrix. The lncRNA-disease scoring matrix was produced by post-processing resource allocation on the heterogeneous network (Table 4).

Resource allocation has been implemented in more than a dozen methods for predicting the lncRNA-disease association. Fan et al. proposed a computational model of IDHH-MIRW by integrating diverse heterogeneous information sources with positive pointwise mutual information and random walk with a restart algorithm. Xiao et al. developed BPLLDA to predict lncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network. Zhang et al. proposed a rule-based inference method on the linked tripartite network, which was constructed by integrating heterogeneous data with deep-learning algorithms. Some other information had also been introduced to allocate resources for improving prediction performance. For example, LION and DisIncRF introduced protein information and genome-wide tissue expression profiles, which are aided by protein-coding genes. NBLDA and LLCLPLDA both constructed four matrices and used the label propagation algorithm to resource allocation. By constructing a disease weight matrix based on the similarity between the lncRNA disease set and the specified disease, IIRWR introduced the concept of disease clique and added the weight of disease linkages to the traveling network. Lap-BiRWRHLDA and BiWalkLDA both used Laplacian normalization and a bi-random walk algorithm on similarity networks. The other two methods, TPGLDA and ncPred, allocated resources from the disease to lncRNAs and other nodes, respectively, but the difference is that the resources are returned to the initial nodes.

Recommendation Algorithm-Based Methods

The common characteristic of the recommendation algorithm-based methods is to recommend a node that may be related to another node. It mainly includes content-based recommendation, collaborative filtering, and matrix factorization. The process is depicted in Figure 3. After applying a recommendation system algorithm to multi-data matrices, recommendation matrices at multiple levels (e.g., such as lncRNA, miRNA) are obtained. Finally, the possibility of the potential relationship between lncRNA and disease is measured through the combination of the recommendation matrices (Table 5).

The first category of recommendation algorithm-based methods is the content-based recommendation. Content-based recommendation refers to recommending similar nodes of previous related nodes for this node. For example, NCPLDA, proposed by Li et al., measured the lncRNA-disease association score based on network consistency projection. The second category of recommendation algorithm-based methods is based on collaborative filtering. Collaborative filtering refers to adding other nodes similar to the nodes and using the information of these nodes to make inferences. CFNBC developed by Yu et al. and NBCLDA proposed by Yu et al. both used collaborative filtering on multi-data matrices to uncover a new relationship between lncRNA and disease and then took advantage of a naive Bayesian classifier to determine whether there is an association between lncRNA and disease in the set of lncRNA-associated node and disease-associated node. A similarity correlation fusion method introduced neighbor information and then was used to predict the association by making the original matrix fit as good as possible in ILDMSF and SKF-LDA. Another method proposed by Ping et al. measured the one-step neighbor of a node based on SimRank measure when there is no common neighbor. DCSLDA, proposed by Zhao et al., calculated the shortest path between lncRNA and disease and distance correlation coefficient to construct the final matrix. The third category of recommendation algorithm-based methods is based on matrix factorization, including MFLDA, WMFLDA, and PMFILDA. These three methods are based on matrix factorization, and the difference is that the latter two add weight and probability separately. Li et al. developed a computational method, DNILMF-LDA, which is anchored in the neighborhood-regularized logistic matrix factorization and optimizes the above parameters to predict interaction probabilities. NNLDA, determined by Hu et al., solved some of the disadvantages of traditional matrix factorization by changing the training method and the loss function and adding a fully connected layer. Additionally, DSCMF combines matrix factorization and collaborative filtering to predict associations efficiently by introducing neighbor information.

Multi-Model Integration-Based Methods

Multi-model integration methods have also been proposed to overcome the shortcomings of the single model and improve prediction performance (Table 6). A combination of matrix completion ideas and recommendation system ideas was applied in three models to predict potential lncRNA-disease associations, including LDASR, ECLDA, and the weighted bagging LightGBM model. Three methods (CNNLDA, CNNDLP, and GCNLDLA) were developed by Xuan et al. to construct the final module through the integration of the convolutional module and attention module. LDAPred, proposed by Xuan et al., introduced the convolutional neural network based on the integration of resource allocation and matrix completion.

DISCUSSION

During the past decade, it has been well documented that lncRNAs play a critical role in nearly all biological processes and have become an emerging paradigm of human disease research. Identification of disease-associated lncRNAs is becoming increasingly important for fundamentally improving our understanding of molecular mechanisms and developing novel therapeutic targets, and thus has attracted more and more attention in the scientific community and is becoming one of the hotspots in medical research. Although current experimental studies in vitro and in vivo could directly link identified lncRNAs with disease phenotypes, they are affected by the limitation of lower efficiency and higher time and labor cost. Taking into account the limitations of experimental studies, high-throughput technologies were then implemented, leading to exponential growth in the
number of dysregulated lncRNAs in diseases. However, the aberrant expression of lncRNAs is not sufficient evidence for ascribing to them a functional role in disease. Therefore, efficient and accurate identification and functional elucidation of disease-associated lncRNAs are in their infancy and remain a major challenge. With the rapidly increasing quantity and quality of bioinformatics databases and resources in lncRNAs and diseases, computer-aided inference of disease-associated RNAs has become a promising avenue for facilitating the unraveling of the functional role of lncRNAs in diseases and provides complementary value for experimental studies. A large number of computational models, algorithms, and tools have been developed and proposed, compensating for this dearth.

In this work, we first summarize data and knowledge sources available for the lncRNA-disease association study, which contains databases of lncRNAs, diseases, and known lncRNA-disease associations. We then present a detailed overview of previously proposed computational methods for inferring lncRNA-disease associations. Based on the core idea implemented, these computational methods can be divided into four categories: (1) matrix completion-based methods,
Second, because experimentally verified IncRNA-disease associations are still too incomplete, resource allocation-based methods need to consider the prediction of separate nodes or integrate additional biological information. However, although this can improve prediction accuracy, some interactions from other databases may contain some noise to interfere with prediction results. In addition, recommendation algorithm-based methods are stated separately. Content-based recommendation models only need node prior knowledge, but new levels of disease-IncRNA associations cannot be recognized. Although collaborative filtering-based recommendation methods complement this shortcoming, the sparse IncRNA-disease association matrix is harmful to the recommendation, and the complexity and time cost of the algorithm will sharply increase when the amount of data is too large. Additionally, matrix factorization-based recommendation methods reduce space complexity by mapping a matrix to a product of low-dimensional matrices. These methods also make it easy to add additional data from different sources and use the intrinsic structure. Matrix factorization-based methods also have the same weaknesses as collaborative filtering-based recommendation methods. The above computational approaches can thus complement each other. Therefore, multi-model integration-based methods were proposed to achieve better performance when investigating the association between IncRNAs and diseases. Finally, only several computational approaches have been developed as online web tools, and most are still theoretical studies that hampered their use for biologists and medical scientists.

With the rapidly increasing knowledge for the functional mechanism of IncRNAs, several challenges that would be helpful to improve the accuracy and practicality of the predictors could be highlighted. It is well known that the majority of IncRNAs exhibited precise subcellular localization, thus performing regulatory roles in a spatiotemporal manner. Therefore, some interactions between IncRNAs and other biological molecules (DNA, RNA, and proteins) used in previous predictors are derived from prediction and do not exist in the real biological world. Therefore, co-localization information of IncRNAs and other biological molecules should be considered. Additionally, it has also been observed that IncRNAs were expressed in highly cell type-specific, tissue-specific, and disease-specific manners. Therefore, more molecular information in the appropriate biological contexts should be introduced into predictors that are more suitable for the specific disease. Finally, the prediction results of these computational approaches are only descriptive associations, and the specific association type (e.g., casual or non-causal association of IncRNA with the disease) is still a challenging task and needs to be answered. The implementation of efficient and reliable computational predictions,
together with systematic biological experiments, will greatly accelerate the study of lncRNA functions and mechanisms in physiological and pathological conditions.

AUTHORS CONTRIBUTIONS
J.S., C.X., and M.Z. designed the study. C.Y., Z.Z., S.B., and P.H. collected and reviewed literature. J.S., M.Z., and C.Y. drafted the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST
The authors declare no competing interests.

ACKNOWLEDGMENTS
This study was supported by the National Natural Science Foundation of China (grant 61973240) and the Scientific Research Foundation for Talents of Wenzhou Medical University (QTJ18029, and QTJ18024). The funders had no roles in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

REFERENCES
1. Johnson, J.M., Edwards, S., Shoemaker, D., and Schadt, E.E. (2005). Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet. 21, 93–102.
2. Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput. Biol. 4, e1000176.
3. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802.
4. Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 155–159.
5. Fang, Y., and Fullwood, M.J. (2016). Roles, functions, and mechanisms of long non-coding RNAs in cancer. Genomics Proteomics Bioinformatics 14, 42–54.
6. Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484–1488.
7. Huarte, M. (2015). The emerging role of lncRNAs in cancer. Nat. Med. 21, 1253–1261.
8. Zeng, W., Wang, F., Ma, Y., Liang, X.C., and Chen, P. (2019). Dysfunctional mechanism of liver cancer mediated by transcription factor and non-coding RNA. Curr. Bioinform. 14, 100–107.
9. Tang, W., Wen, S., Yang, Z., Teschendorff, A.E., and Zou, Q. (2018). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics 34, 398–406.
10. Zhao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., and Zhou, M. (2019). Computational identification of mutator-derived IncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. Brief. Bioinform. bbz118.
11. Chenthamaratam, S.W., Gruhl, F., Mattick, J.S., and Dinger, M.E. (2013). Long noncoding RNAs and the genetics of cancer. Br. J. Cancer 108, 2419–2425.
12. Zhou, M., Hu, L., Zhang, Z., Wu, N., Sun, J., and Su, J. (2018). Recurrence-associated long non-coding RNA signature for determining the risk of recurrence in patients with colon cancer. Mol. Ther. Nucleic Acids 12, 518–529.
13. Zhou, M., Zhang, Z., Zhao, H., Bao, S., Cheng, L., and Sun, J. (2018). An immune-related six-lncRNA signature to improve prognosis prediction of glioblastoma multiforme. Mol. Neurobiol. 55, 3684–3697.
14. Zhou, M., Zhao, H., Xu, W., Bao, S., Cheng, L., and Sun, J. (2017). Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. Mol. Cancer 16, 16.
15. Zou, Q., and Ma, Q. (2020). The application of machine learning to disease diagnosis and treatment. Math. Biosci. 320, 108365.
16. Liao, Z.J., Li, D.P., Wang, X.R., Li, L.S., and Zou, Q. (2018). Cancer diagnosis through isomiR expression with machine learning method. Curr. Bioinform. 13, 57–63.
17. Sun, J., Zhang, Z., Bao, S., Yan, C., Hou, P., Wu, N., Su, J., Xu, L., and Zhou, M. (2020). Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. J. Immunother. Cancer 8, e001100.
18. Zhou, M., Wang, X., Li, J., Hao, D., Wang, Z., Shi, H., Han, L., Zhou, H., and Sun, J. (2015). Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. Mol. Biosyst. 11, 760–769.
19. Sun, J., Shi, H., Wang, Z., Zhang, C., Liu, L., Wang, L., He, W., Hao, D., Liu, S., and Zhou, M. (2014). Inferring novel lncRNA-disease associations based on a random walk model of an lncRNA functional similarity network. Mol. Biosyst. 10, 2074–2081.
20. Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, Y., Xing, T., Sun, X., et al. (2018). NONCODEv5: a comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res. 46 (D1), D308–D314.
21. Chakraborty, S., Deh, A., Maji, R.K., Saha, S., and Ghosh, Z. (2014). lncRNABase: an enriched resource for lncRNA information. PLoS ONE 9, e108010.
22. Ma, L., Gao, J., Liu, L., Du, Q., Li, Z., Zou, D., Bajic, V.B., and Zhang, Z. (2019). LncBook: a curated knowledgebase of human long non-coding RNAs. Nucleic Acids Res. 47 (D1), D128–D134.
23. Ma, L., Li, A., Zou, D., Xu, X., Xiao, L., Yu, J., Bajic, V.B., and Zhang, Z. (2015). LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. Nucleic Acids Res. 43, D187–D192.
24. Bosse, L., Titchekh, N., Lalinski, L.E., Ferris, M.T., Eisfeld, A.J., Green, R.R., Thomas, M.J., Tsoncik-Go, J., Schroth, G.P., Kawakoa, Y., et al. (2014). Annotation of long non-coding RNAs expressed in collaborative cross founder mice in response to respiratory virus infection reveals a new class of interferon-stimulated transcripts. RNA Biol. 11, 875–890.
25. Bhati, D., Pal, K., Ghosh, S., Kapoor, S., Jalali, S., Panwar, B., Jain, S., Sat, S., Sengupta, S., Sachidanandan, C., et al. (2013). LncRNAname: a comprehensive knowledgebase of human long noncoding RNAs. Database (Oxford) 2013, bat034.
26. Xiao, Y.-R., Liu, W., Zhang, Q., and Guo, A.-Y. (2018). LncRNASSP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. Nucleic Acids Res. 46 (D1), D276–D280.
27. Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q., and Dong, D. (2019). LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. Nucleic Acids Res. 47 (D1), D1034–D1037.
28. Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., Li, X., Fang, Y., Shen, W., Xu, Y., et al. (2019). Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. Nucleic Acids Res. 47 (D1), D1028–D1033.
29. Ning, S., Zhang, J., Wang, P., Zhi, H., Wang, J., Liu, Y., Gao, Y., Yue, M., Yue, W., Wang, L., and Li, X. (2016). Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. Nucleic Acids Res. 44 (D1), D980–D985.
30. Cui, T., Zhang, L., Huang, Y., Yi, T., Tan, P., Zhao, Y., Hu, Y., Xu, L., Li, E., and Wang, D. (2018). MNDR v2.0: an updated resource of ncRNA–disease associations in mammals. Nucleic Acids Res. 46 (D1), D371–D374.
31. Zhou, B., Zhao, H., Hu, J., Gao, C., Dou, X., Song, F., et al. (2018). EVLncRNAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments. Nucleic Acids Res. 46, D100–D105.
32. Wu, X., and Zhang, Z., Cheng, L., Xiong, L., et al. (2017). LncRNA and DNA methylation markers. Nucleic Acids Res. 35, 10808–10815.
33. Zhang, Y., Xue, Z., Guo, F., Xu, X., and Chen, H. (2020). NCCyte: a curated lncRNAomics knowledgebase for bridging basic and clinical research in eye diseases. Front. Cell. Dev. Biol. 8, 75.
34. Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.
35. Chen, X., Yan, C.C., Luo, C., Ji, W., Zhang, Y., and Dai, Q. (2015). Constructing IncRNA functional similarity network based on IncRNA disease associations and disease semantic similarity. Sci. Rep. 5, 11338.
36. Huang, Y.-A., Chen, X., You, Z.-H., Huang, D.-S., and Chan, K.C. (2016). ILNCSIM: improved IncRNA functional similarity calculation model. Oncotarget 7, 25902–25914.
37. Paik, H., Heo, H.-S., Ban, H.J., and Cho, S.B. (2014). Unraveling human protein interaction networks underlying co-occurrences of diseases and pathological conditions. J. Transl. Med. 12, 99.
38. Chen, Q., Dai, D., Lan, W., Wu, X., Chen, B., Chen, Y.P., and Wang, J. (2019). ILDSMF: inferring associations between long non-coding RNA and disease based on multi-similarity fusion. IEEE/ACM Trans. Comput. Biol. Bioinform. Published online August 20, 2019. https://doi.org/10.1109/TCBB.2019.2936476.
39. Zou, Q., Li, J., Song, L., Zeng, X., and Wang, G. (2016). Similarity computation strategies in the microRNA-disease network: a survey. Brief. Funct. Genomics 15, 55–64.
40. Zhao, H., Kuang, L., Wang, L., and Xuan, Z. (2018). A novel approach for predicting disease-IncRNA associations based on the distance correlation set and information of the miRNAs. Comput. Math. Methods Med. 2018, 6747453.
41. Chen, X., and Yan, G.-Y. (2013). Novel human IncRNA-disease association inference based on IncRNA expression profiles. Bioinformatics 29, 2617–2624.
42. Cheng, L., Shi, H., Wang, Z., Hu, Y., Yang, H., Zhou, C., Sun, J., and Zhou, M. (2016). IntNetLncSim: an integrative network analysis method to infer human IncRNA functional similarity. Oncotarget 7, 47864–47874.
43. Yu, G., Wang, L.G., Yan, G.R., and He, Q.Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics 31, 608–609.
44. Mathur, S., and Dinakarpandian, D. (2012). Finding disease similarity based on implicit semantic similarity. J. Biomed. Inform. 45, 363–371.
45. Hamaneh, M.B., and Yu, Y.-K. (2014). Relating diseases by integrating gene associations and information flow through protein interaction network. PLoS ONE 9, e105636.
46. Li, W., Wang, S., Xu, J., Mao, G., Tian, G., and Yang, J. (2019). Inferring latent disease-IncRNA associations by faster matrix completion on a heterogeneous network. Front. Genet. 10, 769.
47. Lu, C., Yang, M., Luo, F., Wu, F.-X., Li, M., Pan, Y., Li, Y., and Wang, J. (2018). Prediction of IncRNA-disease associations based on inductive matrix completion. Bioinformatics 34, 3357–3364.
48. Fraidouni, N., and Zaruba, G. (2019). A Matrix Completion Approach for Predicting IncRNA-disease association. In Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP) (CSREA Press), pp. 61–66, https://csce.ucsc.edu/cur/books/2019/LFS/CSREA2019/BIO458.pdf.
49. Ou-Yang, L., Huang, J., Zhang, X.-F., Li, Y.-R., Sun, Y., He, S., and Zhu, J. (2019). IncRNA-disease association prediction using two-side sparse self-representation. Front. Genet. 10, 476.
50. Fan, X.-N., Zhang, S.-W., Zhang, S.-Y., Zhu, K., and Lu, S. (2019). Prediction of IncRNA-disease associations by integrating diverse heterogeneous information sources with RWR algorithm and positive pointwise mutual information. BMC Bioinformatics 20, 87.
51. Xiao, X., Zhu, W., Liao, B., Xu, J., Gu, C., Ji, B., Yao, Y., Peng, L., and Yang, J. (2018). BPLLDA: predicting IncRNA-disease associations based on simple paths with limited lengths in a heterogeneous network. Front. Genet. 9, 411.
52. Zhang, H., Liang, Y., Peng, C., Han, S., Du, W., and Li, Y. (2019). Predicting IncRNA-disease associations using network topological similarity based on deep mining heterogeneous networks. Math. Biosci. 315, 108229.
53. Sumathipala, M., Mairoino, E., Weiss, S.T., and Sharma, A. (2019). Network Diffusion Approach to Predict lncRNA disease associations using multi-type biological networks: LION. Front. Physiol. 10, 888.
In lncRNA-disease associations. In Intelligent Computing Methodologies. ICIC 2019. Lecture Notes in Computer Science, volume 11645, D.S. Huang, Z.K. Huang, and A. Hussain, eds (Springer), pp. 318–326.

76. Guo, Z.-H., You, Z.-H., Wang, Y.-B., Yi, H.-C., and Chen, Z.-H. (2019). A Learning-Based Method for LncRNA-Disease Association Identification Combining Similarity Information and Rotation Forest. iScience 19, 786–795.

77. Chen, X., and Liu, X. (2018). A weighted bagging LightGBM model for potential lncRNA-disease association identification. In Bio-Inspired Computing: Theories and Applications. BIC-TA 2018. Communications in Computer and Information Science, volume 951, J. Qiao, X. Zhao, L. Pan, X. Zuo, X. Zhang, Q. Zhang, and S. Huang, eds (Springer), pp. 307–314.

78. Guo, Z.-H., You, Z.-H., Li, L.-P., Wang, Y.-B., and Chen, Z.-H. (2019). Combining high speed ELM with a CNN feature encoding to predict lncRNA-disease associations. In Intelligent Computing Theories and Application. ICIC 2019. Lecture Notes in Computer Science, volume 11644, D.S. Huang, K.H. Jo, and Z.K. Huang, eds (Springer), pp. 406–417.

79. Xuan, P., Cao, Y., Zhang, T., Kong, R., and Zhang, Z. (2019). Dual convolutional neural networks with attention mechanisms based method for predicting disease-related lncRNA genes. Front. Genet. 10, 416.

80. Xuan, P., Sheng, N., Zhang, T., Liu, Y., and Guo, Y. (2019). CNNNDLP: a method based on convolutional autoencoder and convolutional neural network with adjacent edge attention for predicting lncRNA-disease associations. Int. J. Mol. Sci. 20, 4260.

81. Xuan, P., Pan, S., Zhang, T., Liu, Y., and Sun, H. (2019). Graph convolutional network and convolutional neural network based method for predicting lncRNA-disease associations. Cells 8, 1012.

82. Xuan, P., Jia, L., Zhang, T., Sheng, N., Li, X., and Li, J. (2019). 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, 4458.

83. Wang, Y., Yu, G., Wang, J., Fu, G., Guo, M., and Domeniconi, C. (2020). Weighted matrix factorization on multi-relational data for lncRNA-disease association prediction. Methods 173, 32–43.

84. Cui, Z., Liu, J.-X., Gao, Y.-L., Zhu, R., and Yuan, S.-S. (2020). lncRNA-disease associations prediction using bipartite local model with nearest profile-based association inferring. IEEE J. Biomed. Health Inform. 24, 1519–1527.