OXFORD

Long non-coding RNAs and complex diseases: from experimental results to computational models

Xing Chen*, Chenggang Clarence Yan*, Xu Zhang* and Zhu-Hong You

Corresponding authors. Xing Chen, School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China.
E-mail: xingchen@amss.ac.cn; Zhu-Hong You, School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116,
China. E-mail: zhuhongyou@gmail.com
*The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Abstract

LncRNAs have attracted lots of attentions from researchers worldwide in recent decades. With the rapid advances in both experimental technology and computational prediction algorithm, thousands of lncRNA have been identified in eukaryotic organisms ranging from nematodes to humans in the past few years. More and more research evidences have indicated that lncRNAs are involved in almost the whole life cycle of cells through different mechanisms and play important roles in many critical biological processes. Therefore, it is not surprising that the mutations and dysregulations of lncRNAs would contribute to the development of various human complex diseases. In this review, we first made a brief introduction about the functions of lncRNAs, five important lncRNA-related diseases, five critical disease-related lncRNAs and some important publicly available lncRNA-related databases about sequence, expression, function, etc. Nowadays, only a limited number of lncRNAs have been experimentally reported to be related to human diseases. Therefore, analyzing available lncRNA–disease associations and predicting potential human lncRNA–disease associations have become important tasks of bioinformatics, which would benefit human complex diseases mechanism understanding at lncRNA level, disease biomarker detection and disease diagnosis, treatment, prognosis and prevention. Furthermore, we introduced some state-of-the-art computational models, which could be effectively used to identify disease-related lncRNAs on a large scale and select the most promising disease-related lncRNAs for experimental validation. We also analyzed the limitations of these models and discussed the future directions of developing computational models for lncRNA research.

Key words: long non-coding RNA; complex disease; lncRNA–disease association prediction; computational model; machine learning; biological network

LncRNA

According to the well-known central dogma of molecular biology, genetic information is stored in protein-coding genes [1–6]. Therefore, non-coding RNAs (ncRNAs) have been considered to be transcriptional noise for a long time until more and more evidences showed up and challenged this traditional view [7]. Protein-coding genes only account for approximately 1.5% of the whole genome, which means more than 98% of the human

genome does not encode protein sequences [8]. Furthermore, the proportion of non-protein-coding sequence correspondingly increases along with the complexity of organisms [9]. Recently, increasing evidences have revealed that ncRNAs play a critical role in multiple fundamental and important biological processes [10]. Based on transcript lengths, ncRNAs can be further divided into small ncRNAs and long ncRNAs (lncRNAs). LncRNAs are a major class of important heterogeneous ncRNAs with the lengths more than 200 nucleotides [11–13].

Recently, lncRNAs have attracted much attention from researcher because increasing evidences indicated that lncRNAs play critical roles in multiple biological processes based on diverse underlying mechanisms, such as epigenetic regulation, chromatin remodeling, gene transcription, protein transport, trafficking, cell differentiation, organ or tissue development, cellular transport, metabolic processes and chromosome dynamics [14–20]. Accumulating evidences have further demonstrated that mutations and dysregulations of these lncRNAs are associated with the development and progression of various complex human diseases [21], such as prostate cancer [22, 23], colon cancer [24], lung cancer [25], Alzheimer's diseases (AD) [26], cardiovascular diseases [27], leukemia [28], diabetes [29], AIDS [30] and neurodegeneration diseases [31]. For instance, lncRNA HOTAIR, PCA3 and UCA1 have been treated as potential biomarkers of hepatocellular carcinoma recurrence [32], prostate cancer aggressiveness [33] and bladder cancer detection, respectively [34, 35]. However, the general features of most lncRNAs, such as structure, transcriptional regulation, functions and molecular mechanisms in multiple biological processes or various diseases, still largely remain elusive [15, 16].

LncRNA discovery and classification

With the emergence of sequencing technologies and computational algorithms for lncRNA discovery, more and more lncRNAs are being identified and characterized at a rapid pace in eukaryotic organisms ranging from nematodes to humans [36–39]. For example, the discoveries of two well-known lncRNAs, H19 and X-inactive-specific transcript (Xist), could be traced back to the early 1990s based on the traditional gene mapping [40–44]. Guttman et al. [45] developed a new genome-wide approach and identified 1600 novel large intervening non-coding RNAs (lincRNAs) across four mouse cell types using chromatin marks for promoter regions and gene bodies and gene expression data. Furthermore, they developed a functional genomics approach to assign putative functions to each lincRNA and demonstrate various critical functional roles of lincRNAs [45]. Cabili et al. [46] presented an integrative approach to build the human lincRNA catalog including more than 8000 lincRNAs across 24 different human cell types and tissues based on chromatin marks and RNA-sequencing data and characterize them by more than 30 properties, such as their sequence, structure and orthology features. A large number of lncRNAs have been recorded in biological databases such as lncRNAdb [39, 47], NONCODE [48–52], PLncDB [53] and LNCipedia [37, 38]. For example, there are 487 164 lncRNA transcripts and 324 646 lncRNA genes from 16 species (such as human, mouse, cow and rat) in NONCODE [48–52].

Increasing evidences reveal that human transcriptome is much more complex than what we thought. Based on their different features, lncRNAs could be further divided into the different subgroups as follows [16]: LincRNA [45, 54], long intronic ncRNA [55, 56], transcribed pseudogene [57, 58], transcribed ultraconserved region [28], natural antisense transcript (NAT)

[59–61], promoter-associated long RNA [62], promoter upstream transcript [63], repetitive element-associated ncRNA [64–66] and enhancer-like ncRNA [7, 67]. LncRNAs could also be classified in the following three ways according to their positions relative to protein-coding genes [68, 69]: sense or antisense (classified according to whether the lncRNAs are on the same strand of the nearest protein-coding genes or not) [70], divergent or convergent (classified according to in which way the lncRNAs are transcribed compared with the nearest protein-coding genes) [12] and intronic or intergenic (classified according to the lncRNAs' relative locations to protein-coding genes: inside the introns of a protein-coding gene or in the interval regions between two protein-coding genes) [12, 71, 72].

LncRNA function

In the past, the functionality of lncRNAs caused much controversy (even regarded as transcriptional noises) because of their relatively less cross-species conservation, lower expression levels and higher tissue specificity than protein-coding genes [69, 73, 74]. Furthermore, experimental results indicated that lncRNAs tend to have longer, but fewer, exons [11, 46, 75]. With the rapid development of biological technology and computational models, a growing number of evidences suggest that lncRNAs are involved in almost the whole life cycle of cells through different mechanisms [15, 76]. LncRNAs are confirmed to play diverse and important roles in many fundamental and critical biological processes, including transcriptional and post-transcriptional regulation, epigenetic regulation, organ or tissue development, cell differentiation and apoptosis, cell cycle control, cellular transport, metabolic processes, chromosome dynamics, etc. [3, 18, 77–82].

More and more examples indicated that lncRNAs take the role of signal, decoy, scaffold and guide capacities at almost every stage of gene expression [83]. In addition, considering the fact that lncRNA is large and has a complex secondary and tertiary structure, recent studies also revealed that lncRNAs could bind to DNA, RNA or protein and modulate their functions [7]. Specially, Fendrr, an important and essential lncRNAs for heart and body wall development in the mouse, could interact directly with DNA [84]. It has been observed that the functional properties of lncRNAs are mainly related to their secondary structures [79]. Furthermore, the chromatin modification could be caused by the transcription-independent and transcription-dependent mechanisms of lncRNAs [85–87]. LncRNAs could also be involved in epigenetic silencing by recruiting chromatin remodeling complexes [86]. It is further observed that some lncRNAs usually interact with more than one chromatin-modifying complex [86]. For example, molecular investigations revealed that lncRNAs such as Kcnq1ot1, Airn, Xist and HOTAIR are associated with chromatin remodeling complexes such as Polycomb repressive complexes 1 and 2 (PRC1 and PRC2) [86, 88–96]. In addition, the mutations and dysregulations of lncRNAs are confirmed to be associated with diverse human diseases [83].

Although there are a large number of annotated lncRNAs, only a few lncRNAs have been extensively studied for the identification of their possible functions and the possible molecular mechanism underlying [40, 45]. Therefore, it is a big challenge for both experimental researches and computational biology to accurately identify the functions of lncRNAs [45].

LncRNA–disease associations

Considering the various functions of lncRNAs, it is no surprise to find that the mutations and dysregulations of lncRNAs are closely related to the development and progression of many kinds of human diseases [2, 16, 69, 82, 97–99], such as breast cancer [21, 100], prostate cancer [22, 23], hepatocellular cancer (HCC) [101], colon cancer [24], bladder cancer [34], thyroid cancer [102], lung cancer [25, 103], ovarian cancer [104], AD [26], diabetes [29, 105] and AIDS [30]. Based on the comprehensive lncRNA–disease associations in the lncRNADisease database (http://www.cuilab.cn/lncrnadisease), there have been more than 200 diseases associated with various lncRNAs and more than 300 lncRNAs playing critical roles in various human complex diseases [106].

LncRNAs could function as potential biomarkers for disease diagnosis, treatment and prognosis, and potential drug targets for drug discovery and clinical treatment [97]. For example, lncRNA HOTAIR is treated to be potential biomarker of HCC recurrence and breast cancer detection based on its overexpression from hundreds to even nearly two-thousand-fold in the quantitative Polymerase Chain Reaction (PCR) [21, 107]. Furthermore, lncRNA PCA3 has been confirmed to be related to the formation of prostate cancer aggressiveness by showing 60 times expression levels in prostate tumors compared with normal tissues [33]. lncRNA BC200 is expressed in many kinds of cancers, such as breast, cervix, esophagus, lung, ovary, parotid and tongue cancer, but not in corresponding normal tissues [108]. Another example is lncRNA UCA1, which could contribute to the diagnosis of bladder cancer [35]. In summary, many lncRNAs have been connected to more than one disease, and one disease can be associated with various lncRNAs. Some representative human complex diseases and lncRNAs were introduced as follows.

Breast cancer

Breast cancer is one of the most frequently diagnosed cancer which comprises 22% of all cancers in women worldwide [109, 110]. Histopathological features of breast cancer, such as tumor size, grade and lymph node status, could assist the diagnosis of breast cancer [111]. Experiments indicate that multiple molecular alterations could cause the formation of breast cancer. Especially, many lncRNAs were known to be associated with the formation and development of breast cancer. Some lncRNAs' overexpression could enhance the carcinogenicity of breast cancer cells [112]. For example, lncRNA H19 has great effects in primary breast carcinomas [113, 114]. Down-regulation of H19 significantly reduced the anchorage-independent growth of breast cancer as well as lung cancer [115]. Besides, lncRNA BC200 was found to be expressed in the breast cancer and could be used to predict the tumor development which would benefit the diagnosis and treatment of breast cancer [108, 116]. Furthermore, CDKN2B-AS1 mainly expressed co-clustered with p14/ARF in human breast tumors [117]; GAS5 was also linked with breast cancer because its transcript levels were significantly reduced compared to unaffected normal breast epithelia [118, 119]; amplification of PVT1 could contribute to the pathophysiology of breast cancer [104]. What's more, XIST, KCNQ1OT1 and NEAT1 were also experimentally confirmed to be closely related to breast cancer [120–122].

Lung cancer

Lung cancer is the leading cause of cancer-related deaths worldwide, with the mortality even higher than the combination of colon, breast and prostate cancers [105, 123–125]. Furthermore, the data collected in the recent 5 years further suggested that the survival rate of lung cancer patients (∼%15) is much lower than other cancers [126]. According to the disease patterns and treatment strategies, lung cancer could be roughly divided into non-small cell lung cancer (NSCLC) (80.4%) and small cell lung cancer (SCLC) (16.8%) [124]. Biological experiments demonstrated that lncRNA BCYRN1 was expressed in the tissues of the lung, breast, cervix, esophagus, ovary, parotid and tongue cancer, but it was not expressed in corresponding normal tissues [108]. LncRNA H19 was also confirmed to be associated with lung cancer. Experiments showed that lung cancer cell clonogenicity and anchorage-independent growth would be significantly decreased when H19 was downregulated [113]. Besides, the expression of the tumor suppressor lncRNA GAS5 was also found significantly downregulated in lung cancer tissues [127].

Hepatocelluar carcinoma (HCC)

As the third leading cause of cancer deaths worldwide with the surveillance rates below 20%, HCC is a big threat to human healthy in many countries [128–130]. As far as we know, many factors are closely related to the formation of HCC, such as the infection with hepatitis B virus (HBV) or hepatitis C virus (HCV), aflatoxin B1 intake, alcohol consumption, non-alcoholic fatty liver disease and some hereditary diseases [128, 131]. Especially, the incidence of HBV and HCV is high in Asia and Africa, which largely leads to the development of HCC [132, 133]. Recently, more and more evidences demonstrated that lncRNAs have been involved in HCC. LncRNA Dreh can modify the expression and reorganization of vimentin through binding to vimentin to inhibit HCC metastasis [134, 135]. In addition, lncRNA HOTAIR, LALR and HULC can impact proliferation of hepatoma cells through targeting various key regulators of different pathways in HCC [128, 135]. Particularly, HULC's depletion gave rise to a significant abnormality of several genes related to HCC [136]. Furthermore, LncRNA ATB was suggested to be associated with poor prognosis of HCC since it could promote HCC cell invasion and the invasion-metastasis cascade in HCC [137]. Braconi et al. [135] also found that the expression of MEG3 was markedly reduced in four human HCC cell lines compared with normal liver cells. Another downregulated lncRNA LET played a critical role in hypoxia-induced metastasis in HCC [135, 138].

Alzheimer's disease

According to recent studies, the number of people with dementia worldwide is increasing at a rapid pace [139]. AD is a chronic progressive neurodegenerative disorder, which is caused by the loss of synapses and neurons in specific brain regions such as the CA1 region of hippocampus [140, 141]. Accumulating researches indicated that lncRNAs such as BACE1-AS and BC200 were closely related to AD. For example, the expression of BACE1-AS could drive rapid feed-forward regulation of beta-secretase in AD [26]. Furthermore, compared with age-matched normal brains, significant upregulation of BC200 RNA was found in brain areas that are involved in AD [142, 143]. Furthermore, BC200 expression levels tend to increase with the progression of AD [142, 143].

Heart failure (HF)

HF is a complex clinical syndrome with high concurrent rate and mortality rate [144–146]. Recent studies have found several lncRNAs associated with HF (such as Fendrr [84], Trpm3 and Scarb2 [144]) and revealed the critical functions of these lncRNAs in heart development and HF. These lncRNAs would have important therapeutic potential for HF [147]. For example, tissue-specific lncRNA Fendrr is an essential regulator of heart development [84]. Furthermore, lncRNA Nkx2-5 is a genetic modifier of myotonic muscular dystrophy RNA toxicity, which has important functionality in heart dysfunction [19]. The mitochondrial lncRNA LIPCAR was downregulated early after myocardial infarction but upregulated in later stages. Therefore, LIPCAR could be used to predict survival for the patients with HF and identify the state of patients' cardiac remodeling independent to other risk markers associated with cardiovascular deaths [148].

MEG3

Recent studies showed that some lncRNAs, such as MEG3, HOTAIR, lincRNA-p21 and MALAT-1, work as "tumor-suppressor ncRNAs" or "oncogenic ncRNAs" and play a major role in the development of various cancers (breast cancer, lung cancer, HCC, colon cancer, chronic myeloid leukemia, prostate cancer, etc.) [25]. For example, a pituitary-derived MEG3 isoform could inhibit cancer cell proliferation to some extent [25, 149]. The locus of MEG3 has been predicted to be associated with the pathogenesis and progression of several kinds of tumors, such as meningiomas, nasopharyngeal carcinoma, colorectal carcinoma and leukemia [150]. It was observed that the DLK1-MEG3 locus was silenced and there was no allele loss at the MEG3 gene locus in human non-functioning pituitary tumors [150–152]. Furthermore, the imprinted DLK1-MEG3 gene region on chromosome 14q32.2 would also have influence on susceptibility to type 1 diabetes [153].

H19

H19 has been used as sensitivity diagnostic marker of many important human diseases [154, 155]. For example, upregulated H19 can regulate ID2 expression to promote bladder cancer cell proliferation [156]. Downregulated H19 can stimulated melanogenesis in melisma and may cause melanoma [157]. Furthermore, epigenetic dysregulation of H19 was associated with diseases such as pituitary adenoma and Prader–Willi syndrome [158]. Studies showed that about 37% of patients with Wilms' tumor may be caused by H19 epimutation [159]. H19 could be used to distinguish whether disease is geneogenous for patients with Beckwith–Wiedemann syndrome [160]. In addition, H19 is also frequently overexpressed in myometrium and stroma during pathological endometrial proliferative events and thus may function as tumor suppressor of kidney cancer [154, 161].

HOTAIR

The expression level of HOTAIR would significantly increase in various cancers such as breast cancer [21], lung cancer [162] and HCC [32, 163]. The expression of HOTAIR in primary breast tumors has been treated as an effective prognosis marker of patient survival [164] considering that it showed positive association with breast cancer invasiveness and metastasis [21]. HOTAIR was also confirmed to be upregulated in lung cancer cells based on a three-dimensional organotypic culture model [165]. As a potentially useful biomarker and drug target in malignant gastrointestinal stromal tumor (GIST), frequent upregulation of HOTAIR was detected in GIST [166]. Furthermore, HOTAIR was also regarded as a negative prognostic factor in both primary tumors and blood of colorectal cancer patients [167]. HOTAIR can also be used as an independent prognostic factor of tumor recurrence for HCC patients after liver transplantation [32, 106]. Another example demonstrated that HOTAIR could function as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer [167, 168].

MALAT1

MALAT1 was found to be overexpressed in many solid tumors such as lung cancer, cervical cancer, colorectal cancer and HCC [98]. Specially, it was regarded as a decisive regulator of the metastasis phenotype of lung cancer cells [169] because of its regulation of alternative splicing [170]. Furthermore, MALAT1 expression is three-fold higher in metastasizing tumors like NSCLC than in non-metastasizing tumors. As the oncogene of bladder cancer and kidney cancer, MALAT1 also plays a critical role in cell migration and tumor metastasis [169, 171]. MALAT1 was also treated as a putative marker for prostate cancer [172].

PVT1

PVT1 has close associations with various complex diseases. For example, it has been demonstrated that PVT1 may contribute to the development and progression of diabetic nephropathy [29]. Furthermore, the overexpression of PVT1 caused by genomic abnormalities contributed to ovarian pathogenesis [104]. What's more, the identification of chromosome 15 locus for plasmacytoma variant (6; 15) translocations suggested that PVT1 is associated with some murine T lymphomas [173]. In addition, PVT1 works as the site of reciprocal translocations to immunoglobulin loci in tumors like Burkitt's lymphoma and plasmacytomas [174].

Databases

A plenty of lncRNA-related databases have been constructed recently, including databases annotating lncRNA's sequences or structures such as LNCipedia [37, 38], providing comprehensive information of lncRNAs such as NONCODE [48–52] and lncRNAwiki [175], displaying the experimentally confirmed lncRNA–disease associations such as lncRNADisease [106] and Lnc2Cancer [176], and collecting lncRNA-related interactions such as LncRNA2Target [177] and DIANA-LncBase [178].

Databases collecting comprehensive information of lncRNAs

LNCipedia
(http://www.lncipedia.org/) [37, 38]

The latest version of this database is LNCipedia 3.1, which contains 111 685 annotated human lncRNA transcripts obtained from different sources. It also provides some additional information such as protein-coding potential, secondary structure information and microRNA (miRNA) binding sites. The database is publicly available, which allows users to download the information they need or query new information of lncRNAs, such as sequences and structures.

NONCODE database

(http://www.bioinfo.org/noncode/) [48–52]

NONCODE database is an integrated knowledge database including almost all traditional ncRNA classes (except tRNAs and rRNAs). In particular, the expression profiles and predicted functions of these lncRNAs are also included in it. It also provides a service of lncRNA identification. Users can convert the RefSeq or Ensembl ID to NONCODE ID on NONCODE. In the latest version of NONCODE 2016, the number of lncRNAs has increased sharply from 21 083 to 527336 compared with NONCODE v4.0. Specially, there are 167 150 and 130 558 lncRNAs about human and mouse, respectively. NONCODE 2016 further introduces the information of conservation annotation and lncRNA–disease associations.

LncRBase

(http://bicresources.jcbose.ac.in/zhumur/lncrbase) [179]

LncRBase collects the information of 216 562 lncRNA transcript entries in human and mouse. The basic lncRNA transcript features and additional details on genomic location, overlapping small non-coding RNAs, associated Repeat Elements, associated imprinted genes and lncRNA promoter information are all included in it. It allows users to search for the datasets through selecting one property of lncRNA.

lncRNAWiki

(http://lncrna.big.ac.cn) [175]

lncRNAWiki is a community-curated resource of lncRNA knowledge. The lncRNA sequences and annotation information in it are collected from three databases: GENCODE (version 19; 23 898 human lncRNA transcripts) [11, 12], NONCODE (version 4.0; 95 135 human lncRNA transcripts) [48–52] and LNCipedia (version 2.1; 32 181 human lncRNA transcripts) [37, 38]. Finally, 105 255 non-redundant lncRNA transcripts are obtained from these resources. The classifications of lncRNAs based on genomic location are provided in this database. LncRNAWiki allows users to edit or download the information, or add the newly identified lncRNAs to it [175].

lncRNome

(http://genome.igib.res.in/lncRNome) [180]

lncRNome is an evidence-based resource for over 17 000 lncRNAs in human. Each lncRNA has several properties: the types, chromosomal locations, description on the biological functions and disease associations of lncRNAs. Users can enter the lncRNA's name and obtain the corresponding information about it. In addition to the information mentioned above, the methylation and histone modification, single nucleotide polymorphisms, miRNA binding sites and integrated validated lncRNA–protein interactions are all available.

lncRNAdb

(http://www.lncrnadb.org) [39, 47]

lncRNAdb aims to summarize the knowledge of eukaryotic lncRNAs in an easily accessible and searchable format. Each entry contains information of nucleotide sequences, genomic context, gene expression data derived from the Illumina Body Atlas, structural information, subcellular localization, conservation and function with referenced literature of each entry. It allows users to search for the information about lncRNAs and submit new entries.

GreeNC

(http://greenc.sciencedesigners.com) [181]

To facilitate the study of lncRNAs for the plant research, the GreeNC database was developed to provide information about sequence, genomic coordinates, coding potential and folding energy for all the identified lncRNAs in 37 plant species and six algae. Among more than 190 000 transcripts, more than 120 000 transcripts are annotated as lncRNAs with high confidence, with 30% of them from the *Triticum aestivum* (17.8%) and *Zea mays* (8.2%).

Databases about SNP and lncRNAs

SNP@lincTFBS

(http://bioinfo.hrbmu.edu.cn/SNP_lincTFBS) [182]

SNP@lincTFBS was designed to promote the study and understanding of lincRNA-associated variants and provide improved convenience to identify the function of the abundance of discrepant lincRNA expression in human diseases. It contains 5835 lincRNAs, 6665 single nucleotide polymorphisms (SNPs) mapped within 6614 potential transcription factor binding sites (TFBSs) of 2423 human lincRNAs, 33 181 TFBSs of 3839 human lincRNAs from ucsc dataset and 323 256 TF peaks of 4831 human lincRNAs from ChIPSeq dataset. Users can search SNP or TFBSs of human lincRNAs. This important database has great significance in identification of disease-associated lincRNA candidates.

LncRNASNP

(http://bioinfo.life.hust.edu.cn/lncRNASNP) [183]

LncRNASNP is a resource including SNPs in human/mouse lncRNAs, SNP effects on lncRNA structure and lncRNA–miRNA binding. There are 495 729 SNPs in 32 108 human lncRNA transcripts of 17 436 lncRNA genes for browse or search. In addition, users can obtain the targeted lncRNAs of a miRNA through selecting the miRNA's name in the blank.

Databases collecting lncRNA-related interactions

DIANA-LncBase

(http://www.microrna.gr/LncBase) [178]

DINAN-LncBase is used to illustrate the assumed miRNA–lncRNA functional interactions. It consists of two distinct modules: the Experimental Module and the Prediction Module. There are more than 5000 experimentally supported interactions between 2958 lncRNAs and 120 miRNAs included in the Experimental module. Furthermore, there are more than 10 million computationally predicted interactions between 56 097 lncRNAs and 3078 miRNAs and their corresponding detailed information in the Prediction module, which is calculated based on the latest version of a state-of-the-art algorithm, DIANA-microT-CDS.

LncRNA2Target

(http://www.lncrna2target.org) [177]

LncRNA2Target is a resource of differentially expressed genes (target genes of an lncRNA) after lncRNA knockdown or overexpression. The target genes regulated by an lncRNA and the regulatory lncRNAs of a specific target gene are all available for users to search and browse. In this database, there are 26 410 human lncRNA-target associations between 82 lncRNAs and 11 605 target genes and 67 152 mouse lncRNA-target associations between 134 lncRNAs and 14 762 target genes. It also allows users to download the manually curated lncRNA-target association data in the database or submit new data to the database.

Databases collecting lncRNA–disease associations

LncRNADisease

(http://www.cuilab.cn/lncrnadisease) [106].

Chen et al. developed the LncRNADisease database that integrated more than 1000 lncRNA–disease entries and 475 lncRNA interaction entries, including 321 lncRNAs and 221 diseases from ∼500 publications. LncRNADisease curates lncRNA interactions in various levels, including the interactions with protein, RNA, miRNA and DNA. It also provides the predicted associations between human diseases and 1564 human lncRNAs. It is also a platform that integrated tool(s) which could effectively predict novel lncRNA–disease associations. Furthermore, it allows users to browse, search or download the experimentally supported lncRNA–disease association data or lncRNA interaction data and submit new entries. Finally, users can predict potential disease-lncRNA associations based on the computational models developed in literature [184] (described in detail in the following sections) and then download the predicted association results. The prediction would be implemented by identifying lncRNAs within the regions of 50 kb from any of the disease-related genes based on the genomic context of the lncRNAs and known disease-gene associations.

Lnc2Cancer

(http://www.bio-bigdata.net/lnc2cancer) [176]

Lnc2Cancer is a manually curated database that aims to provide a high-quality and integrated resource for exploring the mechanisms and functions of cancer related lncRNAs. It contains 1239 entries of associations between 579 human lncRNAs and 93 human cancers, which are collected from more than 1,500 published papers. The lncRNA and cancer name, the lncRNA expression pattern, experimental techniques, a brief functional description, the original reference and additional annotation information are all provided by Lnc2Cancer.

MNDR

(http://www.rna-society.org/mndr) [185]

MNDR is a repository focused on diverse ncRNA–disease relationships in mammals that aims to provide a platform to globally view the ncRNA-mediated disease network. Totally, 807 lncRNA-associated, 229 miRNA-associated, 13 piRNA-associated and 100 snoRNA-associated entries are integrated from three mammals (866, 251 and 32 from Homo sapiens, Mus musculus and Rattus norvegicus, respectively).

Computational models

As more and more research evidences have indicated that the mutations and dysregulations of lncRNAs are closely connected to diverse human diseases, more attentions have been paid on the clarity of the functions of lncRNAs and their associations with human diseases [186–188]. Especially, computational models could be effective ways for the identification of potential lncRNA functions and lncRNA–disease associations. Here, we proposed the framework of constructing powerful computational models to predict potential lncRNA–disease associations, which includes three kinds of feasible and important research schemas.

LncRNA–disease associations could be predicted based on powerful computational models in the following three ways. First, we could construct machine learning-based models to predict potential lncRNA–disease associations based on training samples (known disease-related lncRNAs) and unlabeled samples (disease–lncRNA pairs without any known association evidences). Then, we could integrate known lncRNA–disease association network, disease similarity network and lncRNA similarity network to construct heterogeneous network and implement global network similarity-based models (such as random walk and various propagation algorithms) to uncover potential associations between lncRNAs and diseases. Most of these methods cannot be applied to new diseases (diseases without any known associated lncRNAs) and/or new lncRNAs (lncRNAs without any known associated diseases or known miRNA interaction partners). Finally, considering the fact that a plenty of disease–gene associations and disease–miRNA associations have been obtained [189–195], we could obtain potential lncRNA–disease associations based on known disease-related genes/miRNAs by constructing the relationships between gene/ miRNAs and lncRNAs based on their expression levels and regulation relationship.

Machine learning-based models

Laplacian Regularized Least Squares for LncRNA–Disease Association (LRLSLDA)

Chen et al. [35] developed the powerful computational model of LRLSLDA to predict potential disease-related lncRNAs based on the semi-supervised learning framework (see Figure 1). To our knowledge, LRLSLDA is the first lncRNA–disease association prediction model, which is developed based on the basic assumption that similar diseases tend to have associations with functionally similar lncRNAs. LRLSLDA integrates the known disease–lncRNA associations and lncRNA expression profiles to jointly capture the potential associations between disease and lncRNA. LRLSLDA obtains an AUC of 0.7760 in the Leave-One-Out Cross Validation (LOOCV), significantly improving the performance of previous methods which are used to solve the similar computational biology problems. More importantly, LRLSLDA does not need the information of negative samples, which are really difficult to obtain in practical problems. Of course, there are also some limitations in the LRLSLDA. For example, many parameters appear in the model and how to select the parameters is still not well solved. Furthermore, two different scores from lncRNA and disease spaces would be obtained for the same lncRNA–disease pair.

LRLSLDA–LNCRNA functional SIMilarity calculation model (LNCSIM)

Based on the assumption that functional similar lncRNAs are always associated with similar diseases, Chen et al. [8] developed two novel LNCSIMs by calculating semantic similarity between their associated disease groups (see Figure 2). The difference between these two models (LNCSIM1 and LNCSIM2) lies in the calculation of disease semantic similarity based on disease directed acyclic graph (DAG), which could effectively represent the relationships among different diseases. When disease semantic similarity and lncRNA functional similarity (calculated by LNCSIM) are integrated with lncRNA expression similarity, lncRNA Gaussian interaction profile kernel similarity and disease Gaussian interaction profile kernel similarity used in the previous study of LRLSLDA, new lncRNA–disease association model, LRLSLDA–LNCSIM, is obtained, which could further improve the performance of LRLSLDA for lncRNA–disease association prediction. As a result, we obtained the reliable AUCs of 0.8130 and 0.8198 in LOOCV based on two versions of lncRNA similarity scores. Limitations also existed in this method. Considering the fact that the method is based on the known lncRNA–disease associations, prediction results may produce
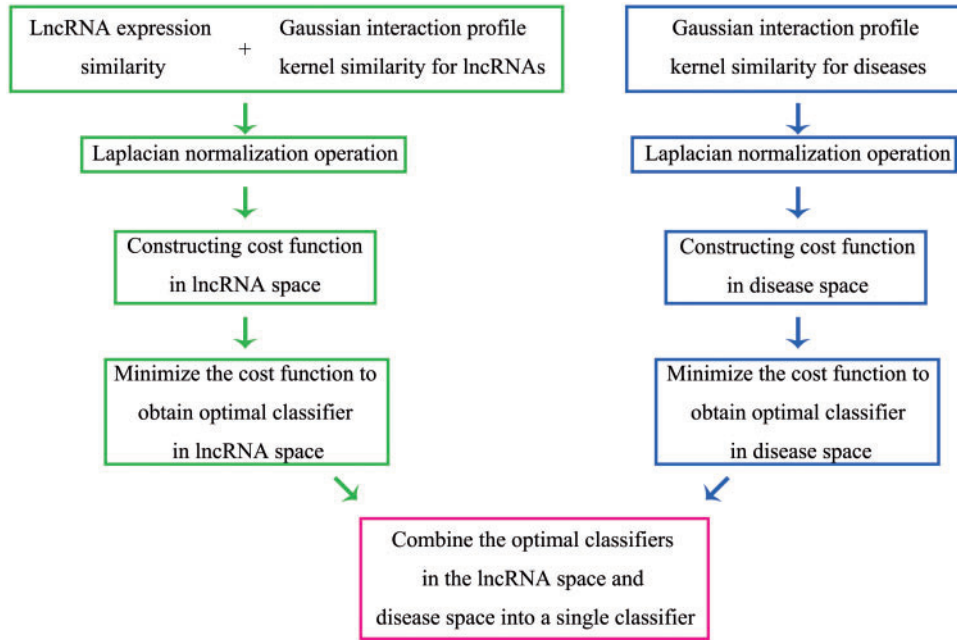
Figure 1. The flowchart of LRLSLDA which have described the basic steps to predict lncRNA–disease associations based on LRLSLDA.
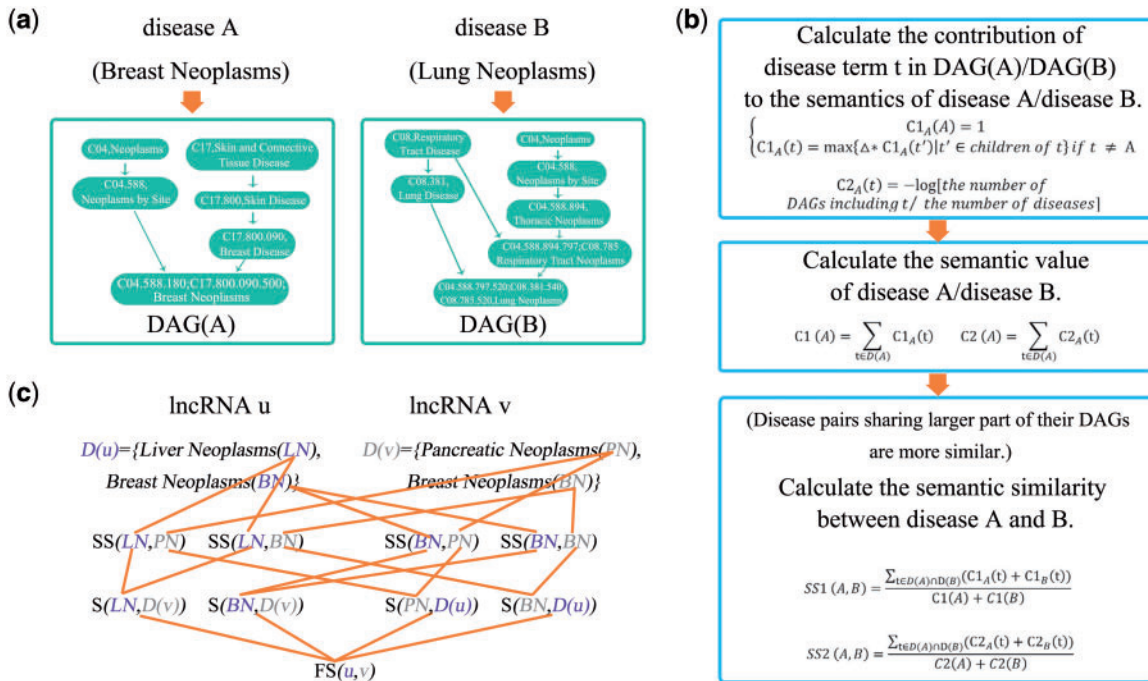
Figure 2. The flowchart of LNCSIM which have described the basic ideas of calculating functional similarity between two lncRNAs: (**A**) constructed the DAGs for disease A and B which are associated with lncRNA *u* and *v*; (**B**) calculated semantic similarity between disease A and B; (**C**) calculate the similarity score between two disease groups associated with lncRNA *u* and *v*. and then obtained functional similarity between them.

the bias to lncRNAs with more known associated diseases. What's more, the selection of semantic contribution decay factor has not been well solved.

LRLSLDA–Improved LNCRNA functional SIMilarity calculation model (ILNCSIM)

Huang et al. further developed the ILNCSIM based on the assumption that lncRNAs with similar biological functions tend to be involved in similar diseases [196]. ILNCSIM was combined with the previously proposed model LRLSLDA to quantify lncRNA–disease association probabilities by using computed lncRNA functional similarity and disease semantic similarity. The main difference between ILNCSIM and previous methods is that ILNCSIM retains the general hierarchical structure information of disease DAGs for disease similarity calculation based on an edge-based method. As a result, LRLSLDA–ILNCSIM obtained

AUCs of 0.9316 and 0.9074 based on MNDR and Lnc2cancer databases in the LOOCV and AUCs of 0.9221 and 0.9033 for MNDR and Lnc2cancer database in 5-fold cross validation, respectively. Limitations also existed in ILNCSIM. For example, the similarity scores used in the model can be further optimized by adding constant terms in calculation. The calculation result was also influenced by the lack of unrecorded but real lncRNA–disease associations. Finally, ILNCSIM still failed to integrate other types of lncRNA-related or disease-related data from biological databases. It is undoubted that the prediction performance will be further improved by integrating those additional data.

Naïve Bayesian classifier

Using known cancer-related lncRNAs, Zhao et al. [76] developed a naïve Bayesian classifier-based model based on the integration of multi-omic data, genomic, regulome and transcriptome data, to identify new cancer-related lncRNAs. The model was evaluated based on 10-fold cross validation on re-annotated publicly available exon array data of multiple cancer types and knockdown data of orthologous lncRNAs on mice. As a result, the proposed model showed a good performance and successfully identified 707 potential cancer-related lncRNAs. The important limitation of supervised classifiers, such as support vector machine (SVM) and naïve Bayesian classifier used here, is that they need the information of negative samples, which are unavailable in the current study. Therefore, they always randomly select unlabeled lncRNA–disease pairs as negative samples, which would seriously influence the prediction performance.

Biological network-based models

RWRlncD

Based on the assumption that functionally related lncRNAs tend to be associated with phenotypically similar diseases, Sun et al. [197] proposed a global network-based computational method named RWRlncD based on an lncRNA–lncRNA functional similarity network. By constructing lncRNA–disease association network, disease similarity network and lncRNA functional similarity network, RWRlncD was proposed to infer potential human lncRNA–disease associations by implementing random walk with restart (RWR) on the lncRNA functional similarity network. RWRlncD obtained an AUC of 0.822 in LOOCV based on known experimentally verified lncRNA–disease associations. However, this method cannot be applied to the diseases without any known associated lncRNAs. The prediction performance of RWRlncD would be further improved when more lncRNA–disease associations and more accurate lncRNA functional similarity measures are available in the future.

RWR on lncRNA–PCG bipartite network

Liu et al. constructed a protein-coding gene (PCG)–lncRNA bipartite network based on lncRNAs and PCGs expression profiles in prostate cancer and protein interaction datasets and further predict cancer-related lncRNAs based on RWR [198]. However, this method was seriously affected by the incomplete protein interaction datasets.

RWRHLD

Based on the assumption that lncRNAs with more common miRNA interaction partners tend to be associated with similar diseases, Zhou et al. [199] proposed the computational model of RWRHLD to identify potential lncRNA–disease associations (see Figure 3). RWRHLD integrated three networks (miRNA-associated lncRNA–lncRNA crosstalk network by calculating shared miRNA interaction partners for each lncRNA pair, disease–disease similarity network and known lncRNA–disease association network) into a heterogeneous network and implemented a random walk on it. RWRHLD obtained a reliable AUC value of 0.871 in LOOCV based on known experimentally verified lncRNA–disease associations. However, RWRHLD is only applied to lncRNAs with known lncRNA–miRNA interactions. Furthermore, the incomplete coverage of lncRNA crosstalk network and lncRNA–disease association network will probably produce some biased predictions.

Kernel-based Random Walk with Restart in Heterogeneous (KRWRH)

The computational model of KRWRH network was proposed to predict new disease–lincRNA associations using three networks: disease–disease similarity network, lincRNA–lincRNA similarity network and known lincRNA–disease association network [200]. These networks will be integrated to construct a heterogeneous network. Then, RWR would be implemented on this heterogeneous network. The experimental results in LOOCV showed that KRWRH was able to predict known and unknown disease–lincRNA associations with a reliable performance.

KATZLDA

Chen et al. [201] developed another model called KATZLDA by integrating known lncRNA–disease associations, lncRNA expression profiles, lncRNA functional similarity, disease semantic similarity and Gaussian interaction profile kernel similarity to uncover potential lncRNA–disease associations (see Figure 4). KATZLDA first transforms link prediction into similarity calculation between nodes and further transforms similarity calculation into counting the number of walks connecting lncRNA node and disease node in the heterogeneous network and calculating the lengths of their walks to jointly decide the potential association probability. As a result, KATZLDA obtained reliable AUCs of 7175, 0.7886 and 0.7719 in the local LOOCV, global LOOCV and 5-fold cross validation, respectively. It is important that KATZLDA could be effectively applied to new diseases and lncRNAs without any known associations. The prediction performance of KATZLDA can be further improved by integrating more information such as disease phenotypic similarity, known disease–genes/miRNAs associations and various lncRNA-related interactions. However, KATZLDA may cause the bias to diseases with more known related lncRNAs and lncRNAs with more known associated diseases or/and more known miRNA interaction partners.

Propagation algorithm on coding-non-coding gene–disease bipartite network

Yang et al. [202] constructed coding-non-coding gene–disease bipartite network based on known disease genes and lncRNA–disease associations and further implemented a propagation algorithm on this bipartite network to infer the underlying lncRNA–disease associations. As a result, the method obtained an AUC of 0.7881 in LOOCV. However, the lack of interactions between non-coding genes and protein-coding genes and lncRNA functional annotations affected the performance of this method.

Models not based on known lncRNA–disease associations

In the above two subsections, all the computational models need the known lncRNA–disease associations to implement prediction. However, even nowadays, known experimentally
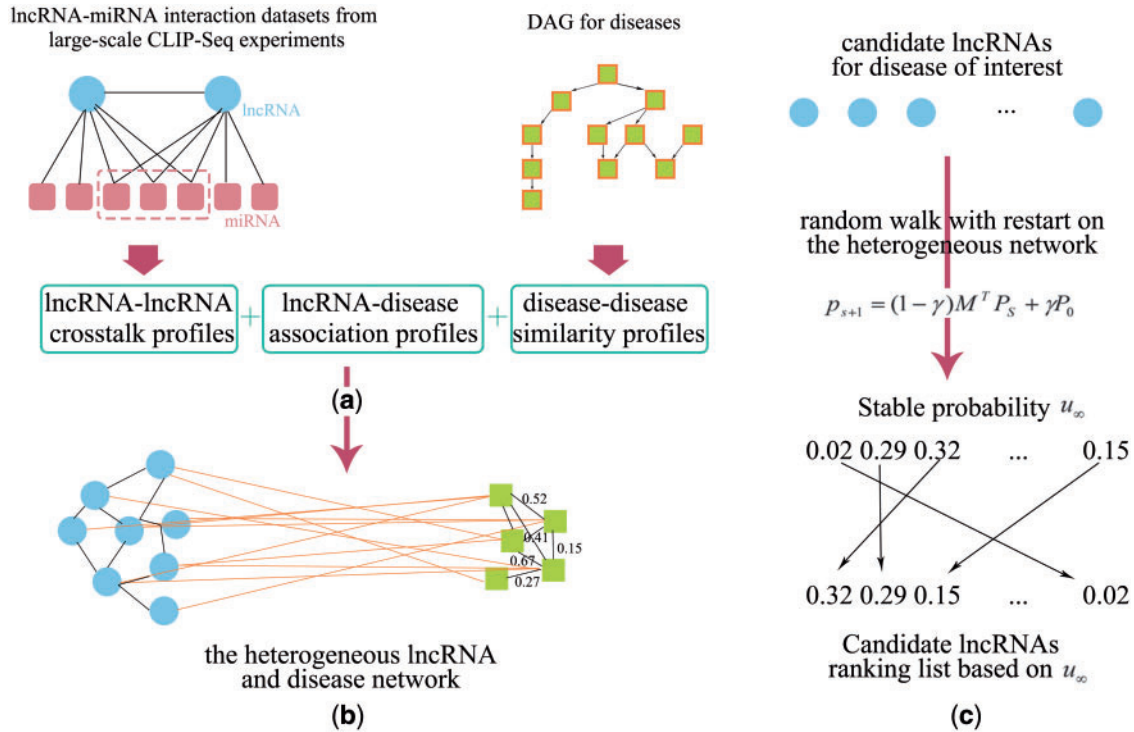
Figure 3: The flowchart shows the three steps of RWRHLD: (**A**) constructing the lncRNA-miRNA interaction network based on the "ceRNA hypothesis" and the disease–disease similarity network based on disease DAG structure; (**B**) constructing the heterogeneous lncRNA–disease network by integrating lncRNA crosstalk network, disease similarity network, and experimentally confirmed lncRNA–disease association network; (**C**) implementing random walk on the heterogeneous network and obtaining a stable probability to rank candidate lncRNAs. A colour version of this figure is available at BIB online: https://academic.oup.com/bib.
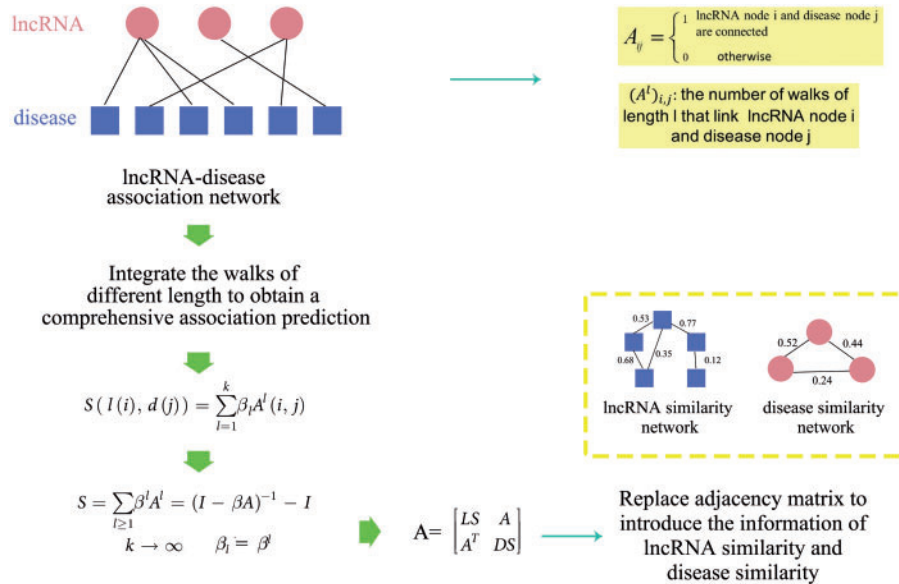
Figure 4: The flowchart of KATZLDA which demonstrates the basic ideas of adopting Katz measure for predicting lncRNA–disease associations. A colour version of this figure is available at BIB online: https://academic.oup.com/bib.

confirmed lncRNA–disease associations are still very limited. Therefore, researchers start to predict lncRNA–disease association based on the known disease-related genes/miRNAs and the relationships between lncRNAs and genes/miRNAs.

Computational framework based on disease genes
Liu et al. [203] developed the first computational method without the need to rely on known lncRNA–disease associations to predict potential human lncRNA–disease associations by integrating known human disease genes and expression profiles of human lncRNAs and gene (see Figure 5). In this method, the lncRNAs were divided into two parts: tissue-specific and non-tissue-specific lncRNAs. They first calculated the tissue specificity scores based on the expression levels of all lncRNAs in different tissues. Then, for tissue-specific lncRNAs, this computational framework infers that there could be potential
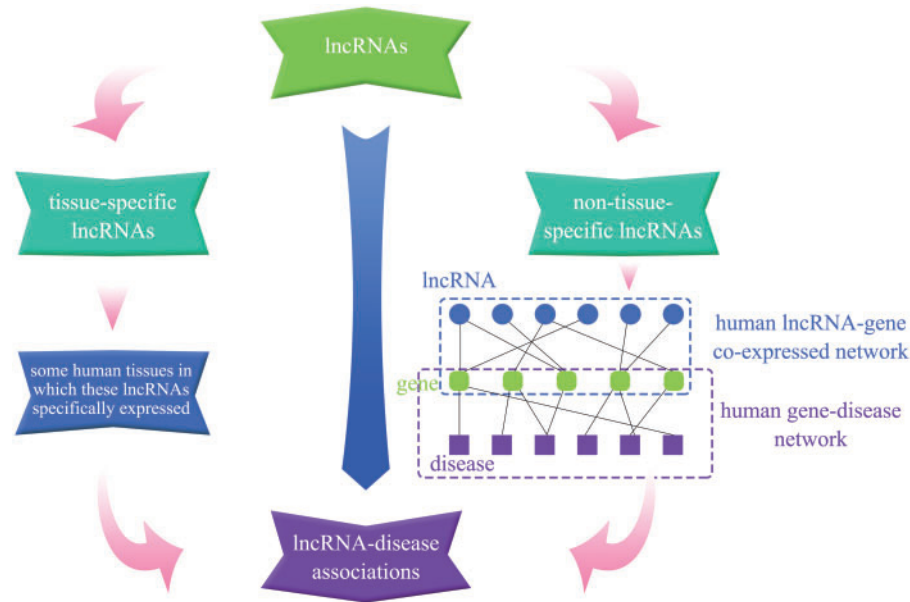
Figure 5: This method consists the following four steps: calculating tissue specificity score and dividing all the lncRNAs into tissue-specific and non-tissue-specific lncRNAs; predicting potential lncRNA–disease associations for tissue-specific lncRNAs; constructing gene–lncRNA co-expression relationships for all the non-tissue-specific lncRNAs by computing Spearman's correlation coefficients between their expression profiles; performing disease enrichment and predicting potential lncRNA–disease associations for non-tissue-specific lncRNAs. A colour version of this figure is available at BIB online: https://academic.oup.com/bib.

associations between this lncRNAs with diseases related with these human tissues. Furthermore, it could obtain related diseases for non-tissue-specific lncRNAs based on disease–gene associations and gene–lncRNA co-expression relationship. The model obtained an AUC of 0.7645 in LOOCV and the prediction accuracy of 0.89 for non-tissue-specific lncRNAs. However, this method cannot predict the associated lncRNAs for diseases with no related gene records.

Genomic location-based method

Li et al. [184] proposed a computational method based on genome location to globally screen the human lncRNAs potentially involved in vascular disease. Ten lncRNAs predicted to be associated with vascular smooth muscle cells were selected for further experimental validation to test the accuracy of the method. As a result, eight of the 10 lncRNAs (80%) were confirmed. The experimental result demonstrated the reliable prediction performance of this method and its potential value for the identification of novel lncRNAs for the diagnosis and therapy of vascular disease. However, the application scope of this method is extremely limited because not all the lncRNAs have neighbor genes and even if this lncRNA has neighbor genes, it may be not functionally related with its neighbor genes.

HyperGeometric distribution for LncRNA–Disease Association (HGLDA)

Chen [204] developed a novel computational model of HGLDA inference by integrating miRNA–disease associations and lncRNA–miRNA interactions (see Figure 6). In addition, Chen also constructed a model of LncRNA Functional Similarity Calculation based on the information of MiRNA (LFSCM) to calculate lncRNA functional similarity combining disease semantic similarity, miRNA–disease associations and lncRNA–miRNA interactions. As a result, HGLDA obtained an AUC of 0.7621 in LOOCV although it did not rely on any known disease–lncRNA associations. HDLDA has a reliable performance of predicting potential disease–lncRNA associations and could be useful in

detecting biomarkers for human disease diagnosis, treatment, prognosis and prevention. However, HGLDA cannot be applied to those lncRNAs without any known miRNA interaction partners. Furthermore, considering the calculation of LFSCM, it tends to cause bias to lncRNAs with more miRNA interaction partners or/and lncRNAs with miRNA interaction partners which has been associated with more diseases.

Case studies and experimental validations

A plenty of computational models mentioned above have been successfully applied to potential disease–lncRNA association prediction. For the prediction, known lncRNA–disease associations in the databases, such as lncRNADisease [106], *MNDR* [185] and *Lnc2Cancer* [176], are used as training samples. Some of prediction results have been further confirmed by biological experiments (see Table 1). Case studies about six kinds of important human cancers are summarized as follows.

Colon cancer

Researchers have implemented the computational models of LRLSLDA–ILNCSIM [196], KATZLDA [201], HGLDA [204] and LRLSLDA–LNCSIM [8] to predict potential colon cancer-related lncRNAs. As a result, we experimentally confirmed six lncRNAs out of top 20 potential predictions based on LRLSLDA–ILNCSIM. Furthermore, four and seven out of top 10 predicted lncRNAs based on LRLSLDA–LNCSIM and KATZLDA were confirmed based on various biological experiments. For example, PVT1 (3rd in the prediction results of KATZLDA) was confirmed to be functionally correlated with the proliferation and invasion of colon cancer cells based on real-time PCR and considered to be a potential independent colon cancer biomarker for disease detection and patient survival [205]. For the computational model of HGLDA, predicted lncRNAs with false discovery rate (FDR) less than 0.05 were selected as potential colon cancer-related lncRNAs and five of them were experimentally confirmed. For example, considering the frequent occurrence of loss of
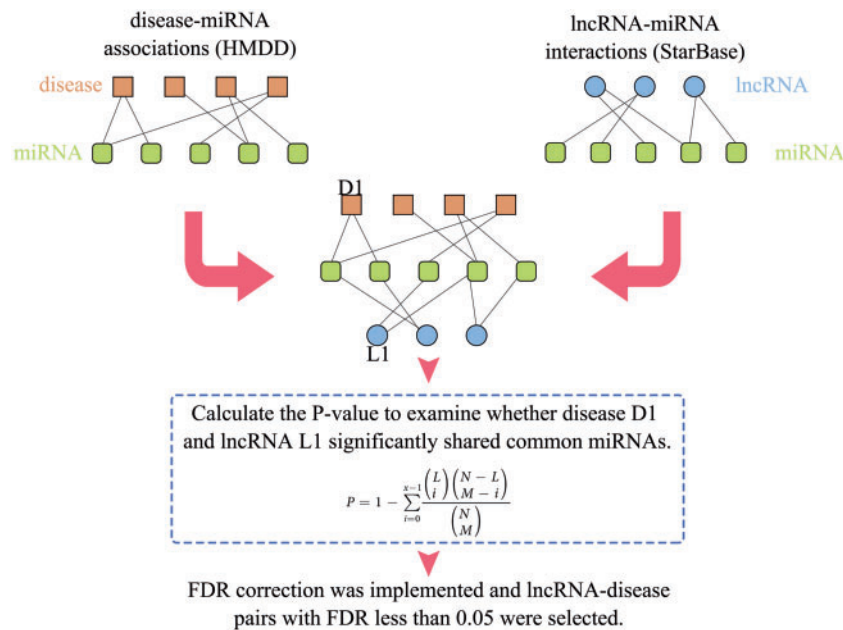
Figure 6: The flowchart of HGLDA which showed the basic idea of predicting potential lncRNA–disease associations by integrating disease–miRNA associations and lncRNA-miRNA interactions. The *P* value was obtained for each lncRNA–disease pair to examine whether they have significantly common associated miRNAs. Then FDR correction was implemented to all these *P* values. At last, the lncRNA–disease pairs whose FDR was less than 0.05 were selected for experimental validation. A colour version of this figure is available at BIB online: https://academic.oup.com/bib.

imprinting of KCNQ1OT1 in colon cancer, it has been considered as an effective biomarker for disease diagnosis [206].

Lung cancer

The computational models of LRLSLDA–LNCSIM [8], LRLSLDA–ILNCSIM [196] and HGLDA [204] have been used for lung cancer–lncRNA association prediction. As a result, three out of top 10 (LRLSLDA–LNCSIM) and seven out of top 20 (LRLSLDA–ILNCSIM) predictions were experimentally confirmed. For example, UCA1 in the 3rd of the prediction result have been confirmed to provide the high diagnostic ability for NSCLC [207]. Furthermore, seven out of all the potential lung cancer-related lncRNAs with FDR less than 0.05 were experimentally confirmed. For example, MALAT1 is an important lung cancer metastasis biomarker, which could promote lung cancer cell motility by regulating motility related gene expression [208]. TUG could affect NSCLC cell proliferation by epigenetically regulating the expression of HOXB788 [209].

Breast cancer

HGLDA [204] was applied to breast cancer for associated lncRNA prediction and seven potential lncRNAs with significant FDR less than 0.05 have been confirmed based on biological experiments. For example, NEAT1 could play critical role in nicotine-induced breast cancer development. Further experiments indicated that breast cancer patients with high NEAT1 expression tend to have low survival rate [122, 210].

Prostate cancer

As an important human complex disease, many researchers paid much attention to predicting prostate cancer–lncRNA associations based on the computational models such as LRLSLDA–ILNCSIM [196]. Six associations were successfully predicted by computational models, such as the associations between H19, CBR3-AS1, MEG3, UCA1, KCNQ1OT1, LINCRNA-P21 and prostate cancer.

Gastric cancer

KATZLDA [201] has been successfully applied to identify potential associations between human lncRNAs and gastric cancer. Six out of top 10 predicted lncRNAs (H19, CDKN2B-AS1, MEG3, PVT1, MALAT1 and HOTAIR) have been confirmed by the experimental evidences. For examples, H19 was ranked 1st in the prediction list. Its associations with gastric cancer have been confirmed by both microarray and Qrt-PCR. In the experiments, H19 was the most upregulated lncRNA among all the 135 differentially expressed lncRNAs in gastric cancer tissues [211]. Another prediction result, MALAT1 in the 5th of the prediction results, has been confirmed to induce gastric cancer cell proliferation and have frequently upregulated expression in gastric cancer cell lines [212].

Renal cancer

Potential renal cancer–lncRNA associations have been predicted based on the computational model of KATZLDA [201]. H19, MEG3, PVT1, UCA1 and MALAT1 in the top 10 prediction results have been confirmed by biological experiments.

Discussion and conclusion

More and more lncRNAs are being identified and characterized at a rapid pace with the advances in transcriptome arrays and deep sequencing [36]. Furthermore, lncRNAs are confirmed to play critical roles in multiple biological processes [3, 18, 77–80]. Therefore, there is no surprise that lncRNAs have been closely involved in the origin and development of various human complex diseases based on a growing body of evidences [16]. The roles of lncRNAs in multiple biological processes or various diseases seem to be much more complex than what we have known from GWAS studies as well as the studies of disease processes [36]. However, so far, very little annotated lncRNAs have obvious functional annotations for the lack of evolutionary conservation of lncRNAs, the lack of common biogenesis or mechanism of action for lncRNAs, and the absence of unified

Table 1. Predicted lncRNA–disease associations based on various computational models were successfully experimentally confirmed

Model	Disease	lncRNA	Rank
LRLSLDA–ILNCSIM	Colon cancer	UCA1	3
KATZLDA			6
LRLSLDA–ILNCSIM	Colon cancer	HOTAIR	13
KATZLDA			4
HGLDA			FDR < = 0.05
LRLSLDA–ILNCSIM	Colon cancer	H19	1
HGLDA			FDR < = 0.05
LRLSLDA–LNCSIM1			2
LRLSLDA–LNCSIM2			2
LRLSLDA–ILNCSIM	Colon cancer	XIST	14
HGLDA			FDR < = 0.05
KATZLDA	Colon cancer	KCNQ1OT1	7
HGLDA			FDR < = 0.05
KATZLDA	Colon cancer	MALAT1	2
HGLDA			FDR < = 0.05
KATZLDA	Colon cancer	PVT1	3
LRLSLDA–LNCSIM1			5
LRLSLDA–LNCSIM2			5
KATZLDA	Colon cancer	CRNDE	9
LRLSLDA–LNCSIM1			1
LRLSLDA–LNCSIM2			1
LRLSLDA–LNCSIM1	Colon cancer	CASC2	8
LRLSLDA–LNCSIM2			9
LRLSLDA–ILNCSIM	Colon cancer	MEG3	16
LRLSLDA–ILNCSIM	Colon cancer	HULC	19
KATZLDA	Colon cancer	CDKN2B-AS1	1
LRLSLDA–ILNCSIM	Lung cancer	HOTAIR	4
HGLDA			FDR < = 0.05
LRLSLDA–LNCSIM1			2
LRLSLDA–LNCSIM2			2
LRLSLDA–ILNCSIM	Lung cancer	GAS5	10
HGLDA			FDR < = 0.05
LRLSLDA–LNCSIM1			9
LRLSLDA–LNCSIM2			9
LRLSLDA–ILNCSIM	Lung cancer	UCA1	3
LRLSLDA–LNCSIM1			7
LRLSLDA–LNCSIM2			8
LRLSLDA–ILNCSIM	Lung cancer	BC200	1
LRLSLDA–ILNCSIM	Lung cancer	XIST	8
LRLSLDA–ILNCSIM	Lung cancer	MEG3	17
LRLSLDA–ILNCSIM	Lung cancer	LSINCT5	20
HGLDA	Lung cancer	EPB41L4A-AS1	FDR < = 0.05
HGLDA	Lung cancer	MALAT1	FDR < = 0.05
HGLDA	Lung cancer	TUG1	FDR < = 0.05
HGLDA	Lung cancer	H19	FDR < = 0.05
HGLDA	Lung cancer	NEAT1	FDR < = 0.05
LRLSLDA–ILNCSIM	Prostate cancer	H19	1
LRLSLDA–ILNCSIM	Prostate cancer	CBR3-AS1	2
LRLSLDA–ILNCSIM	Prostate cancer	UCA1	3
LRLSLDA–ILNCSIM	Prostate cancer	KCNQ1OT1	13
LRLSLDA–ILNCSIM	Prostate cancer	LINCRNA-P21	14
LRLSLDA–ILNCSIM	Prostate cancer	MEG3	15
KATZLDA	Gastric cancer	MALAT1	5
KATZLDA	Gastric cancer	H19	1
KATZLDA	Gastric cancer	CDKN2B-AS1	2
KATZLDA	Gastric cancer	MEG3	3
KATZLDA	Gastric cancer	PVT1	4
KATZLDA	Gastric cancer	HOTAIR	7
KATZLDA	Renal cancer	UCA1	8
KATZLDA	Renal cancer	H19	1
KATZLDA	Renal cancer	MEG3	3
KATZLDA	Renal cancer	PVT1	4
KATZLDA	Renal cancer	MALAT1	6
HGLDA	Breast cancer	MALAT1	FDR < = 0.05
LRLSLDA–LNCSIM1			15
HGLDA	Breast cancer	H19	FDR < = 0.05
HGLDA	Breast cancer	CDKN2B-AS1	FDR < = 0.05
HGLDA	Breast cancer	NEAT1	FDR < = 0.05

(continued)

Table 1. Continued

Model	Disease	lncRNA	Rank
LRLSLDA			4
HGLDA	Breast cancer	XIST	FDR < = 0.05
HGLDA	Breast cancer	KCNQ1OT1	FDR < = 0.05
HGLDA	Breast cancer	HOTAIRM1	FDR < = 0.05
LRLSLDA–LNCSIM1	Brain ischemia	B2 SINE RNA	12
LRLSLDA–LNCSIM2			13
LRLSLDA–LNCSIM1	Lung adenocarcinoma	MEG3	7
LRLSLDA–LNCSIM2			5
LRLSLDA–LNCSIM1	Lung adenocarcinoma	BCYRN1	8
LRLSLDA–LNCSIM2			6
LRLSLDA–LNCSIM1	Colorectal neoplasia	HOTAIR	6
LRLSLDA–LNCSIM2			8
LRLSLDA–LNCSIM1	Colorectal neoplasia	KCNQ1OT1	10
LRLSLDA–LNCSIM2			14
LRLSLDA–LNCSIM1	Colorectal neoplasia	MALAT1	9
LRLSLDA–LNCSIM2			11
LRLSLDA–LNCSIM1	Heroin abuse	MEG3	2
LRLSLDA–LNCSIM2			4
LRLSLDA–LNCSIM1	Heroin addiction	MIAT	11
LRLSLDA–LNCSIM2			7
LRLSLDA–LNCSIM1	Lung adenocarcinoma	H19	3
LRLSLDA–LNCSIM2			2
LRLSLDA–LNCSIM1	cervix cancer	H19	14
LRLSLDA	Alzheimer disease	HAR1A	17
LRLSLDA	Alzheimer disease	HAR1B	18
LRLSLDA	bladder cancer	TUG1	18
LRLSLDA	melanoma	BANCR	10

resources to annotate lncRNAs [213]. Especially, compared with a large amount of lncRNA-related biological data about sequence and expression produced by a plenty of experimental studies, only a few lncRNAs have been extensively studied to annotate their possible functions and identify their potential associations with various human complex diseases.

The prediction of lncRNA–disease associations is of great significance in biological, medical and other fields [214]. Recently, scientists focused on building computational models to predict new lncRNA–disease associations, which will help understand the biogenesis, regulation and function of lncRNAs and human disease molecular mechanism at lncRNA level, identify the associations between lncRNAs and diseases and design biomarker and drug for human disease diagnosis, treatment, prognosis and prevention [7, 8, 35, 201, 203, 204]. Based on computational models, the association probability between lncRNAs and diseases could be quantified and lncRNA–disease pairs with higher scores could be selected for further biological experimental validation. In this way, we could effectively decrease the time and the cost of biological experiments. Therefore, computational models could provide a powerful guidance and support to the research of identifying novel lncRNA–disease associations. Computational approaches could also be used to predict potential functions of lncRNAs, identify novel lncRNA genes and construct potential regulatory networks between lncRNAs and other molecules at various levels [7].

In this paper, we summarized the functions of lncRNAs, five important lncRNA-related diseases, five critical disease-related lncRNAs and some important publicly available lncRNA-related databases about sequence, expression, function, etc. Then, we introduced some state-of-the-art computational models for disease-related lncRNAs identification on a large scale, which could be used to select most promising disease-related lncRNAs for biological experiment validation. Computational models consist of machine learning-based models, biological network-

based models and models without the need to rely on known lncRNA–disease associations. Most of these models integrate different types of biological datasets to implement prediction. For the computational model construction, similarity calculation has important impact on accuracy of lncRNA–disease association prediction models. Therefore, how to develop effective computational models to construct lncRNA functional similarity and reasonably integrate the similarity scores from different biological information is also a hot topic worthy of further research. We also analyze the limitations of these models and discuss the future directions of computational lncRNA research.

Machine learning-based models have their advantages and disadvantages. The key advantage of most of these approaches is that almost all the models can effectively predict novel lncRNA–disease associations for lncRNAs with at least one known associated diseases and diseases with at least one known associated lncRNAs. Some models, such as LRLSLDA–LNCSIM and LRLSLD–-ILNCSIM, could be applied to predict lncRNA-associated diseases by integrating lncRNA similarity network, disease similarity network and experimentally confirmed lncRNA–disease associations. With more lncRNA–diseases associations available in the future, prediction accuracy could be further improved. Furthermore, the semi-supervised models such as LRLSLDA could integrate positive lncRNA–disease associations and unlabeled lncRNA–disease pairs to implement effective prediction, which solve the problems of obtaining negative lncRNA–disease associations. However, supervised learning-based models, such as SVM and naïve Bayesian classifier, seriously rely on negative samples which are difficult to obtain. The problems of parameter section, classifier combination and prediction bias also exist in the current machine learning-based computational models.

Nowadays, network has become an effective tool in predicting potential lncRNA–disease associations. Successful network-based models would have critical impact on timely diagnosis, personalized treatment, prognosis and personalized prevention of diseases at the level of lncRNAs. Biological network-based computational models tend to integrate known lncRNA–disease association network, disease semantic/phenotypic similarity network, and lncRNA functional similarity network obtained from known lncRNA–disease associations or lncRNA–miRNA interactions. RWR or various propagation algorithms are used to implement potential predictions on constructed heterogeneous network. The important disadvantage of most of these methods is that they may not obtain prediction results for new diseases and/or new lncRNAs. Furthermore, the incomplete coverage of lncRNA–miRNA interaction network, protein interaction network and lncRNA–disease networks will probably produce some biased prediction for lncRNAs with more known associated diseases or miRNA interaction partners which has been associated with more diseases. Nowadays, a wide range of lncRNA-related databases and web servers have been built, providing a variety of resources of lncRNAs. Therefore, making full use of different types of heterogeneous data sources will help to greatly improve the predict performance of computational predictive models. Therefore, the future direction of the network-based methods could be summarized as follows. On one hand, more heterogeneous networks should be integrated, such as lncRNA–disease network, disease similarity network and lncRNA functional similarity network and lncRNA-related various interaction networks. On the other hand, new network-based computational models should be implemented on this heterogeneous network rather than the single network. In this way, for the lncRNAs without known associated diseases, we still can obtain potential associated diseases of this lncRNA based on the known heterogeneous network as long as there is at least one reachable path in the network.

As for the models which do not rely on known lncRNA–disease associations, they use other biological datasets to predict potential lncRNA–disease associations, such as gene–disease associations or miRNA–disease associations. Therefore, the incomplete human disease-associated gene/miRNA dataset will greatly affect the prediction performance of these computational models. Furthermore, the computational model based on gene genomic context could be limited by the fact that not all the lncRNAs have functionally related neighbor genes.

For most of the computational models mentioned above, the prediction performance was evaluated based on cross validation. However, recently, Park et al. [215] demonstrated that the performance evaluation based on cross validation is different for in-sample and out-of-sample associations. We have developed a computational model named LRLSLDA to predict potential lncRNA–disease associations and further evaluated the performance of LRLSLDA based on the new validation framework proposed by Park et al. [35]. As a result, LRLSLDA obtained an excellent predictive performance in different test classes. Therefore, for the lncRNA–disease association prediction, it is very important and necessary to report cross validation performance for all the four independent test classes.

Key Points

- We made a brief introduction of the functions of lncRNAs, five important lncRNA-related diseases, five critical disease-related lncRNAs and some important publicly available lncRNA-related databases about sequence, expression, function, etc.
- Developing effective computational models to predict potential lncRNA–disease associations from heterogeneous biological data could benefit not only better understanding of human complex diseases mechanism at lncRNA level but also biomarker detection for complex human diseases diagnosis, treatment, prognosis and prevention
- LncRNA–disease associations could be predicted based on powerful computational models in the three ways, including machine learning-based models, network-based models and models without the need to rely on known lncRNA–disease associations.
- Various computational models for potential lncRNA–disease association prediction have their advantages and disadvantages.
- Making full use of different types of heterogeneous data sources could benefit more effective identification of new lncRNA–disease interactions.

Funding

References

1. Yanofsky C. Establishing the triplet nature of the genetic code. *Cell* 2007;**128**:815–8.
2. Taft RJ, Pang KC, Mercer TR, *et al.* Non-coding RNAs: regulators of disease. *J Pathol* 2010;**220**:126–39.

3. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 2009;**23**:1494–504.

4. Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008;**322**:1845–8.

5. Bertone P, Stolc V, Royce TE, *et al*. Global identification of human transcribed sequences with genome tiling arrays. *Science* 2004;**306**:2242–6.

6. Crick F, Barnett L, Brenner S, *et al*. General nature of the genetic code for proteins. *Nature* 1961;**192**:1227–32.

7. Mohanty V, Gökmen-Polar Y, Badve S, *et al*. Role of lncRNAs in health and disease—size and shape matter. *Brief Funct Genomics* 2015;**14**:115–29.

8. Chen X, Yan CC, Luo C, *et al*. Constructing lncRNA functional similarity network based on lncRNA–disease associations and disease semantic similarity. *Sci Rep* 2015;**5**:11338.

9. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 2007;**29**:288–99.

10. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genetics* 2011;**12**:861–74.

11. Harrow J, Frankish A, Gonzalez JM, *et al*. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;**22**:1760–74.

12. Derrien T, Johnson R, Bussotti G, *et al*. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012;**22**:1775–89.

13. Guttman M, Russell P, Ingolia NT, *et al*. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 2013;**154**:240–51.

14. Zhao W, Luo J, Jiao S. Comprehensive characterization of cancer subtype associated long non-coding RNAs and their clinical implications. *Sci Rep* 2014;**4**:6591.

15. Lu Q, Ren S, Lu M, *et al*. Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics* 2013;**14**:651.

16. Li J, Xuan Z, Liu C. Long non-coding RNAs and complex human diseases. *Int J Mol Sci* 2013;**14**:18790–808.

17. Bussemakers MJ, van Bokhoven A, Verhaegh GW, *et al*. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 1999;**59**:5975–9.

18. Managadze D, Rogozin IB, Chernikova D, *et al*. Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. *Genome Biol Evol* 2011;**3**:1390–404.

19. Schonrock N, Harvey RP, Mattick JS. Long noncoding RNAs in cardiac development and pathophysiology. *Circ Res* 2012;**111**:1349–62.

20. Bhartiya D, Kapoor S, Jalali S, *et al*. Conceptual approaches for lncRNA drug discovery and future strategies. *Expert Opin Drug Discov* 2012;**7**:503–13.

21. Gupta RA, Shah N, Wang KC, *et al*. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010;**464**:1071–6.

22. Chung S, Nakagawa H, Uemura M, *et al*. Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer Sci* 2011;**102**:245–52.

23. Cui Z, Ren S, Lu J, *et al*. The prostate cancer-up-regulated long noncoding RNA PlncRNA-1 modulates apoptosis and proliferation through reciprocal regulation of androgen receptor. *Urol Oncol* 2013;**31**:1117–23.

24. Pibouin L, Villaudy J, Ferbus D, *et al*. Cloning of the mRNA of overexpression in colon carcinoma-1: a sequence overexpressed in a subset of colon carcinomas. *Cancer Genet Cytogenet* 2002;**133**:55–60.

25. Zhang X, Zhou Y, Mehta KR, *et al*. A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells. *J Clin Endocrinol Metab* 2003;**88**:5119–26.

26. Faghihi MA, Modarresi F, Khalil AM, *et al*. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nat Med* 2008;**14**:723–30.

27. Congrains A, Kamide K, Oguro R, *et al*. Genetic variants at the 9p21 locus contribute to atherosclerosis through modulation of ANRIL and CDKN2A/B. *Atherosclerosis* 2012;**220**:449–55.

28. Calin GA, Liu C-G, Ferracin M, *et al*. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 2007;**12**:215–29.

29. Alvarez ML, Di Stefano JK. Functional characterization of the plasmacytoma variant translocation 1 gene (PVT1) in diabetic nephropathy. *PLoS One* 2011;**6**:e18671.

30. Zhang Q, Chen C-Y, Yedavalli VS, *et al*. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio* 2013;**4**:e00596–12.

31. Johnson R. Long non-coding RNAs in Huntington's disease neurodegeneration. *Neurobiol Dis* 2012;**46**:245–54.

32. Yang Z, Zhou L, Wu L-M, *et al*. Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation. *Ann Surg Oncol* 2011;**18**:1243–50.

33. van Poppel H, Haese A, Graefen M, *et al*. The relationship between Prostate CAncer gene 3 (PCA3) and prostate cancer significance. *BJU Int* 2012;**109**:360–6.

34. Zhang Z, Hao H, Zhang C, *et al*. Evaluation of novel gene UCA1 as a tumor biomarker for the detection of bladder cancer. *Zhonghua Yi Xue Za Zhi* 2012;**92**:384–7.

35. Chen X, Yan G-Y. Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* 2013;**29**:2617–24.

36. Yang G, Lu X, Yuan L. LncRNA: a link between RNA and cancer. *Biochim Biophys Acta (BBA)—Gene Regul Mech* 2014;**1839**:1097–109.

37. Volders P-J, Verheggen K, Menschaert G, *et al*. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* 2015;**43**:4363–4.

38. Volders P-J, Helsens K, Wang X, *et al*. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* 2013;**41**:D246–51.

39. Amaral PP, Clark MB, Gascoigne DK, *et al*. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res* 2011;**39**:D146–51.

40. Qiu M-T, Hu J-W, Yin R, *et al*. Long noncoding RNA: an emerging paradigm of cancer research. *Tumor Biol* 2013;**34**:613–20.

41. Brannan CI, Dees EC, Ingram RS, *et al*. The product of the H19 gene may function as an RNA. *Mol Cell Biol* 1990;**10**:28–36.

42. Brockdorff N, Ashworth A, Kay GF, *et al*. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 1992;**71**:515–26.

43. Brown CJ, Hendrich BD, Rupert JL, *et al*. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 1992;**71**:527–42.

44. Feil R, Walter J, Allen ND, *et al*. Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes. *Development*. 1994;**120**:2933–43.

45. Guttman M, Amit I, Garber M, *et al*. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009;**458**:223–7.

46. Cabili MN, Trapnell C, Goff L, *et al*. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 2011;**25**:1915–27.

47. Quek XC, Thomson DW, Maag JL, *et al*. lncRNAdb v2. 0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res* 2015;**43**:D168–73.

48. Liu C, Bai B, Skogerbø G, *et al*. NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res* 2005;**33**:D112–5.

49. He S, Liu C, Skogerbø G, *et al*. NONCODE v2. 0: decoding the non-coding. *Nucleic Acids Res* 2008;**36**:D170–2.

50. Bu D, Yu K, Sun S, *et al*. NONCODE v3. 0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res* 2012;**40**:D210–5.

51. Xie C, Yuan J, Li H, *et al*. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* 2014;**42**:D98–103.

52. Zhao Y, Li H, Fang S, *et al*. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res* 2016;**44**:D203–8.

53. Jin J, Liu J, Wang H, *et al*. PLncDB: plant long non-coding RNA database. *Bioinformatics* 2013;**29**:1068–71.

54. Guttman M, Garber M, Levin JZ, *et al*. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol* 2010;**28**:503–10.

55. Tahira AC, Kubrusly MS, Faria MF, *et al*. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Mol Cancer* 2011;**10**:1476–59.

56. Louro R, Smirnova AS, Verjovski-Almeida S. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 2009;**93**:291–8.

57. Poliseno L, Salmena L, Zhang J, *et al*. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;**465**:1033–8.

58. Salmena L, Poliseno L, Tay Y, *et al*. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 2011;**146**:353–8.

59. Yin Y, Zhao Y, Wang J, *et al*. antiCODE: a natural sense-antisense transcripts database. *BMC Bioinformatics* 2007;**8**:319.

60. Katayama S, Tomaru Y, Kasukawa T, *et al*. Antisense transcription in the mammalian transcriptome. *Science* 2005;**309**:1564–6.

61. Lehner B, Williams G, Campbell RD, *et al*. Antisense transcripts in the human genome. *Trends Genet* 2002;**18**:63–5.

62. Kapranov P, Cheng J, Dike S, *et al*. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 2007;**316**:1484–8.

63. Preker P, Nielsen J, Kammler S, *et al*. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 2008;**322**:1851–4.

64. Liu W-M, Chu W-M, Choudary PV, *et al*. Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* 1995;**23**:1758–65.

65. Espinoza CA, Goodrich JA, Kugel JF. Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA* 2007;**13**:583–96.

66. Mariner PD, Walters RD, Espinoza CA, *et al*. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* 2008;**29**:499–509.

67. Kim T-K, Hemberg M, Gray JM, *et al*. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010;**465**:182–7.

68. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012;**81**:145–66.

69. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* 2009;**136**:629–41.

70. Nam J-W, Bartel DP. Long noncoding RNAs in C. elegans. *Genome Res* 2012;**22**:2529–40.

71. Tsai M-C, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Res* 2011;**71**:3–7.

72. Ma H, Hao Y, Dong X, *et al*. Molecular mechanisms and function prediction of long noncoding RNA. *Sci World J* 2012;**2012**:541786.

73. Mercer TR, Dinger ME, Sunkin SM, *et al*. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* 2008;**105**:716–21.

74. Pauli A, Valen E, Lin MF, *et al*. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res* 2012;**22**:577–91.

75. Hrdlickova B, de Almeida RC, Borek Z, *et al*. Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim Biophys Acta* 2014;**1842**:1910–22.

76. Zhao T, Xu J, Liu L, *et al*. Identification of cancer-related lncRNAs through integrating genome, regulome and transcriptome features. *Mol Biosyst* 2015;**11**:126–36.

77. Lander ES, Linton LM, Birren B, *et al*. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860–921.

78. Qureshi IA, Mattick JS, Mehler MF. Long non-coding RNAs in nervous system function and disease. *Brain Res* 2010;**1338**:20–35.

79. Liao Q, Liu C, Yuan X, *et al*. Large-scale prediction of long non-coding RNA functions in a coding–non-coding gene co-expression network. *Nucleic Acids Res* 2011;**39**:3864–78.

80. Moran VA, Perera RJ, Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res* 2012;**40**:6391–400.

81. Zhu J, Fu H, Wu Y, *et al*. Function of lncRNAs and approaches to lncRNA–protein interactions. *Sci China Life Sci* 2013;**56**:876–85.

82. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009;**10**:155–9.

83. Li X, Wu Z, Fu X, *et al*. lncRNAs: insights into their function and mechanics in underlying disorders. *Mutat Res/Rev Mutat Res* 2014;**762**:1–21.

84. Grote P, Wittler L, Hendrix D, *et al*. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* 2013;**24**:206–14.

85. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 2013;**14**:699–712.

86. Saxena A, Carninci P. Long non-coding RNA modifies chromatin. *Bioessays* 2011;**33**:830–9.

87. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev* 2009;**23**:1831–42.

88. Pandey RR, Mondal T, Mohammad F, *et al*. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 2008;**32**:232–46.

89. Rinn JL, Kertesz M, Wang JK, *et al*. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 2007;**129**:1311–23.

90. Khalil AM, Guttman M, Huarte M, *et al*. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* 2009;**106**:11667–72.

91. Nagano T, Mitchell JA, Sanz LA, *et al*. The air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 2008;**322**:1717–20.

92. Terranova R, Yokobayashi S, Stadler MB, *et al*. Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 2008;**15**:668–79.

93. Tsai M-C, Manor O, Wan Y, *et al*. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 2010;**329**:689–93.

94. Umlauf D, Goto Y, Cao R, *et al*. Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* 2004;**36**:1296–300.

95. Zhao J, Sun BK, Erwin JA, *et al*. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 2008;**322**:750–6.

96. Kaneko S, Li G, Son J, *et al*. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev* 2010;**24**:2615–20.

97. Spizzo R, Almeida MI, Colombatti A, *et al*. Long non-coding RNAs and cancer: a new frontier of translational research&quest. *Oncogene* 2012;**31**:4577–87.

98. Cheetham S, Gruhl F, Mattick J, *et al*. Long noncoding RNAs and the genetics of cancer. *Br J Cancer* 2013;**108**:2419–25.

99. Gutschner T, Diederichs S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol* 2012;**9**:703–19.

100. Guffanti A, Iacono M, Pelucchi P, *et al*. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. *BMC Genomics* 2009;**10**:163.

101. Wang J, Liu X, Wu H, *et al*. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res* 2010;**38**:5366–83.

102. Jendrzejewski J, He H, Radomska HS, *et al*. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci* 2012;**109**:8646–51.

103. Ji P, Diederichs S, Wang W, *et al*. MALAT-1, a novel noncoding RNA, and thymosin β4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 2003;**22**:8031–41.

104. Guan Y, Kuo W-L, Stilwell JL, *et al*. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin Cancer Res* 2007;**13**:5745–55.

105. Pasmant E, Sabbagh A, Vidaud M, *et al*. ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J* 2011;**25**:444–8.

106. Chen G, Wang Z, Wang D, *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* 2013;**41**:D983–6.

107. Yang F, Zhang L, Huo XS, *et al*. Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology* 2011;**54**:1679–89.

108. Chen W, Böcker W, Brosius J, *et al*. Expression of neural BC200 RNA in human tumours. *J Pathol* 1997;**183**:345–51.

109. Donahue HJ, Genetos DC. Genomic approaches in breast cancer research. *Brief Funct Genomics* 2013;**12**:391–6.

110. Karagoz K, Sinha R, Arga KY. Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. *Omics: A J Integr Biol* 2015;**19**:115–30.

111. Meng J, Li P, Zhang Q, *et al*. A four-long non-coding RNA signature in predicting breast cancer survival. *J Exp Clin Cancer Res* 2014;**33**:84.

112. Wang KC, Chang HY. Molecular mechanisms of long non-coding RNAs. *Mol Cell* 2011;**43**:904–14.

113. Barsyte-Lovejoy D, Lau SK, Boutros PC, *et al*. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer Res* 2006;**66**:5330–7.

114. Lottin S, Adriaenssens E, Dupressoir T, *et al*. Overexpression of an ectopic H19 gene enhances the tumorigenic properties of breast cancer cells. *Carcinogenesis* 2002;**23**:1885–95.

115. Tessier CR, Doyle GA, Clark BA, *et al*. Mammary tumor induction in transgenic mice expressing an RNA-binding protein. *Cancer Res* 2004;**64**:209–14.

116. Iacoangeli A, Lin Y, Morley EJ, *et al*. BC200 RNA in invasive and preinvasive breast cancer. *Carcinogenesis* 2004;**25**:2125–33.

117. Pasmant E, Laurendeau I, Héron D, *et al*. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 2007;**67**:3963–9.

118. Mourtada-Maarabouni M, Pickard M, Hedge V, *et al*. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* 2009;**28**:195–208.

119. Huarte M, Guttman M, Feldser D, *et al*. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 2010;**142**:409–19.

120. Vincent-Salomon A, Ganem-Elbaz C, Manié E, *et al*. X inactive–specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. *Cancer Res* 2007;**67**:5134–40.

121. Rodriguez BA, Weng Y-I, Liu T-M, *et al*. Estrogen-mediated epigenetic repression of the imprinted gene cyclin dependent kinase inhibitor 1C in breast cancer cells. *Carcinogenesis* 2011;**32**:812–21.

122. Bavarva JH, Tae H, Settlage RE, *et al*. Characterizing the genetic basis for nicotine induced cancer development: a transcriptome sequencing study. *PLoS One* 2013;**8**:e67252.

123. Wood SL, Pernemalm M, Crosbie PA, *et al*. Molecular histology of lung cancer: from targets to treatments. *Cancer Treat Rev* 2015;**41**:361–75.

124. White NM, Cabanski CR, Silva-Fisher JM, *et al*. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome Biol* 2014;**15**:429.

125. Liu J, Lee W, Jiang Z, *et al*. Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. *Genome Res* 2012;**22**:2315–27.

126. Liu B, Fang L, Chen J, *et al*. miRNA-dis: microRNA precursor identification based on distance structure status pairs. *Mol Biosyst* 2015;**11**:1194–204.

127. Shi X, Sun M, Liu H, *et al*. A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Mol Carcinog* 2013;**54**:E1–12.

128. He Y, Meng X-M, Huang C, *et al*. Long noncoding RNAs: novel insights into hepatocelluar carcinoma. *Cancer Lett* 2014;**344**: 20–7.

129. Venook AP, Papandreou C, Furuse J, *et al*. The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. *Oncologist* 2010;**15**:5–13.

130. Edenvik P, Davidsdottir L, Oksanen A, *et al*. Application of hepatocellular carcinoma surveillance in a European setting. What can we learn from clinical practice? *Liver Int* 2015;**35**:1862–71.

131. Song P, Feng X, Zhang K, *et al*. Screening for and surveillance of high-risk patients with HBV-related chronic liver disease: promoting the early detection of hepatocellular carcinoma in China. *Biosci Trends* 2013;**7**:1–6.

132. Wei Q, Guo P, Mu K, *et al*. Estrogen suppresses hepatocellular carcinoma cells through ERβ-mediated upregulation of the NLRP3 inflammasome. *Lab Invest* 2015;**95**:804–16.

133. Fu SC, Huang YW, Wang TC, *et al*. Increased risk of hepatocellular carcinoma in chronic hepatitis B patients with new onset diabetes: a nationwide cohort study. *Aliment Pharmacol Ther* 2015;**41**:1200–9.

134. Zhang L, Yang F, Yuan J-H, *et al*. Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. *Carcinogenesis* 2012;**34**:577–86.

135. Huang J-L, Zheng L, Hu Y-W. Characteristics of long noncoding RNA and its relation to hepatocellular carcinoma. *Carcinogenesis* 2013;**35**:507–14.

136. Panzitt K, Tschernatsch MM, Guelly C, *et al*. Characterization of HULC, a novel gene with striking up-regulation in hepatocellular carcinoma, as noncoding RNA. *Gastroenterology* 2007;**132**:330–42.

137. Yuan J-H, Yang F, Wang F, *et al*. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell* 2014;**25**:666–81.

138. Yang F, Huo X-s, Yuan S-x, *et al*. Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol Cell* 2013;**49**:1083–96.

139. Schonrock N, Götz J. Decoding the non-coding RNAs in Alzheimer's disease. *Cell Mol Life Sci* 2012;**69**:3543–59.

140. Palop JJ, Chin J, Mucke L. A network dysfunction perspective on neurodegenerative diseases. *Nature* 2006;**443**:768–73.

141. Tan L, Yu J-T, Hu N, *et al*. Non-coding RNAs in Alzheimer's disease. *Mol Neurobiol* 2013;**47**:382–93.

142. Ng S-Y, Lin L, Soh BS, *et al*. Long noncoding RNAs in development and disease of the central nervous system. *Trends Genet* 2013;**29**:461–8.

143. Mus E, Hof PR, Tiedge H. Dendritic BC200 RNA in aging and in Alzheimer's disease. *Proc Natl Acad Sci USA* 2007;**104**: 10679–84.

144. Li D, Chen G, Yang J, *et al*. Transcriptome analysis reveals distinct patterns of long noncoding RNAs in heart and plasma of mice with heart failure. *PLoS One* 2013;**8**:e77938.

145. Kaya Z, Leib C, Katus HA. Autoantibodies in heart failure and cardiac dysfunction. *Circ Res* 2012;**110**:145–58.

146. Barsheshet A, Brenyo A, Goldenberg I, *et al*. Sex-related differences in patients' responses to heart failure therapy. *Nat Rev Cardiol* 2012;**9**:234–42.

147. Papait R, Kunderfranco P, Stirparo GG, *et al*. Long noncoding RNA: a new player of heart failure? *J Cardiovasc Transl Res* 2013;**6**:876–83.

148. Kumarswamy R, Bauters C, Volkmann I, *et al*. Circulating long noncoding RNA, LIPCAR, predicts survival in patients with heart failure. *Circ Res* 2014;**114**:1569–75.

149. Benetatos L, Vartholomatos G, Hatzimichael E. MEG3 imprinted gene contribution in tumorigenesis. *Int J Cancer* 2011;**129**:773–9.

150. Zhao J, Dahle D, Zhou Y, *et al*. Hypermethylation of the promoter region is associated with the loss of MEG3 gene expression in human pituitary tumors. *J Clin Endocrinol Metab* 2005;**90**:2179–86.

151. Gejman R, Batista DL, Zhong Y, *et al*. Selective loss of MEG3 expression and intergenic differentially methylated region hypermethylation in the MEG3/DLK1 locus in human clinically nonfunctioning pituitary adenomas. *J Clin Endocrinol Metab* 2008;**93**:4119–25.

152. Cheunsuchon P, Zhou Y, Zhang X, *et al*. Silencing of the imprinted DLK1-MEG3 locus in human clinically nonfunctioning pituitary adenomas. *Am J Pathol* 2011;**179**:2120–30.

153. Wallace C, Smyth DJ, Maisuria-Armer M, *et al*. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat Genet* 2010;**42**:68–71.

154. Martens-Uzunova ES, Böttcher R, Croce CM, *et al*. Long noncoding RNA in prostate, bladder, and kidney cancer. *Eur Urol* 2014;**65**:1140–51.

155. Jiang YJ, Bikle DD. LncRNA: a new player in 1α, 25 (OH) 2 vitamin D3/VDR protection against skin cancer formation. *Exp Dermatol* 2014;**23**:147–50.

156. Luo M, Li Z, Wang W, *et al*. Long non-coding RNA H19 increases bladder cancer metastasis by associating with EZH2 and inhibiting E-cadherin expression. *Cancer Lett* 2013;**333**:213–21.

157. Kim NH, Lee CH, Lee AY. H19 RNA downregulation stimulated melanogenesis in melasma. *Pigment Cell Melanoma Res* 2010;**23**:84–92.

158. Shi X, Sun M, Liu H, *et al*. Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett* 2013;**339**: 159–66.

159. Scott RH, Murray A, Baskcomb L, *et al*. Stratification of Wilms tumor by genetic and epigenetic analysis. *Oncotarget* 2012;**3**:327–35.

160. De Baun MR, Niemitz EL, Mc Neil DE, *et al*. Epigenetic alterations of H19 and LIT1 distinguish patients with Beckwith–Wiedemann syndrome with cancer and birth defects. *Am J Hum Genet* 2002;**70**:604–11.

161. Lottin S, Adriaenssens E, Berteaux N, *et al*. The human H19 gene is frequently overexpressed in myometrium and stroma during pathological endometrial proliferative events. *Eur J Cancer* 2005;**41**:168–77.

162. Nakagawa T, Endo H, Yokoyama M, *et al*. Large noncoding RNA HOTAIR enhances aggressive biological behavior and is associated with short disease-free survival in human non-small cell lung cancer. *Biochem Biophys Res Commun* 2013;**436**:319–24.

163. Geng Y, Xie S, Li Q, *et al*. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. *J Int Med Res* 2011;**39**:2119–28.

164. Zhang A, Xu M, Mo Y-Y. Role of the lncRNA-p53 regulatory network in cancer. *J Mol Cell Biol* 2014;**6**:181–91.

165. Zhuang Y, Wang X, Nguyen HT, *et al*. Induction of long intergenic non-coding RNA HOTAIR in lung cancer cells by type I collagen. *J Hematol Oncol* 2013;**6**:35.

166. Niinuma T, Suzuki H, Nojima M, *et al*. Upregulation of miR-196a and HOTAIR drive malignant character in gastrointestinal stromal tumors. *Cancer Res* 2012;**72**:1126–36.

167. Svoboda M, Slyskova J, Schneiderova M, *et al*. HOTAIR long non-coding RNA is a negative prognostic factor not only in

primary tumors, but also in the blood of colorectal cancer patients. *Carcinogenesis* 2014;**35**:1510–5.

168. Liu X, Sun M, Nie F, *et al*. Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol Cancer* 2014;**13**:92.

169. Gutschner T, Hämmerle M, Eißmann M, *et al*. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 2013;**73**:1180–9.

170. Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod Pathol* 2013;**26**:155–65.

171. Ying L, Chen Q, Wang Y, *et al*. Upregulated MALAT-1 contributes to bladder cancer cell migration by inducing epithelial-to-mesenchymal transition. *Mol Biosyst* 2012;**8**:2289–94.

172. Guay C, Jacovetti C, Nesca V, *et al*. Emerging roles of noncoding RNAs in pancreatic *β*-cell function and dysfunction. *Diabetes Obes Metab* 2012;**14**:12–21.

173. Graham M, Adams JM, Cory S, Murine T. lymphomas with retroviral inserts in the chromosomal 15 locus for plasmacytoma variant translocations. *EMBO J* 1985;**4**:675–81.

174. Beck-Engeser GB, Lum AM, Huppi K, *et al*. Pvt1-encoded microRNAs in oncogenesis. *Retrovirology* 2008;**5**:4.

175. Ma L, Li A, Zou D, *et al*. LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res* 2014;**43**:D187–92.

176. Ning S, Zhang J, Wang P, *et al*. Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res* 2016;**44**:D980–5.

177. Jiang Q, Wang J, Wu X, *et al*. LncRNA2Target: a database for differentially expressed genes after lncRNA knockdown or overexpression. *Nucleic Acids Res* 2015;**43**:D193–6.

178. Paraskevopoulou MD, Georgakilas G, Kostoulas N, *et al*. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res* 2013;**41**:D239–45.

179. Chakraborty S, Deb A, Maji RK, *et al*. LncRBase: an enriched resource for lncRNA information. *PloS One* 2014;**9**:e108010.

180. Bhartiya D, Pal K, Ghosh S, *et al*. lncRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database* 2013;**2013**:bat034.

181. Gallart AP, Pulido AH, de Lagrán IAM, *et al*. GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Res* 2016;**44**:D1161–6.

182. Ning S, Zhao Z, Ye J, *et al*. SNP@ lincTFBS: an integrated database of polymorphisms in human LincRNA transcription factor binding sites. *PloS One* 2014;**9**:e103851.

183. Gong J, Liu W, Zhang J, *et al*. lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res* 2015;**43**:D181–6.

184. Li J, Gao C, Wang Y, *et al*. A bioinformatics method for predicting long noncoding RNAs associated with vascular disease. *Sci China Life Sci* 2014;**57**:852–7.

185. Wang Y, Chen L, Chen B, *et al*. Mammalian ncRNA-disease repository: a global view of ncRNA-mediated disease network. *Cell Death Dis* 2013;**4**:e765.

186. Li J, Ma W, Zeng P, *et al*. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinformatics* 2015;**16**:806–12.

187. Ferrè F, Colantoni A, Helmer-Citterich M. Revealing protein–lncRNA interaction. *Brief Bioinformatics* 2016;**17**:106–16.

188. Yotsukura S, Hancock T, Natsume-Kitatani Y, *et al*. Computational recognition for long non-coding RNA (lncRNA): software and databases. *Brief Bioinformatics* 2016;bbv114.

189. Chen X, Liu MX, Cui QH, *et al*. Prediction of disease-related interactions between MicroRNAs and environmental factors based on a semi-supervised classifier. *PLoS One* 2012;**7**:e43425.

190. Chen X, Liu MX, Yan G. RWRMDA: predicting novel human microRNA-disease associations. *Mol Biosyst* 2012;**8**:2792–8.

191. Chen X, Yan CC, Zhang X, *et al*. RBMMMDA: predicting multiple types of disease-microRNA associations. *Sci Rep* 2015;**5**:13877.

192. Chen X, Yan G-Y. Semi-supervised learning for potential human microRNA–disease associations inference. *Sci Rep* 2014;**4**:5501.

193. Chen X, Yan GY, Liao XP. A novel candidate disease genes prioritization method based on module partition and rank fusion. *OMICS: J Integr Biol* 2010;**14**:337–56.

194. Chen X. miREFRWR: a novel disease-related microRNA–environmental factor interactions prediction method. *Mol Biosyst* 2016;**12**:624–33.

195. Chen X, Yan CC, Zhang X, *et al*. WBSMDA: within and between score for MiRNA–disease association prediction. *Sci Rep* 2016;**6**:21106.

196. Huang Y, Chen X, You Z, *et al*. ILNCSIM: improved lncRNA functional similarity calculation model. *Oncotarget* 2016;**7**:25902–14.

197. Sun J, Shi H, Wang Z, *et al*. Inferring novel lncRNA–disease associations based on a random walk model of a lncRNA functional similarity network. *Mol Biosyst* 2014;**10**:2074–81.

198. Liu Y, Zhang R, Qiu F, *et al*. Construction of a lncRNA–PCG bipartite network and identification of cancer-related lncRNAs: a case study in prostate cancer. *Mol Biosyst* 2015;**11**:384–93.

199. Zhou M, Wang X, Li J, *et al*. Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. *Mol Biosyst* 2015;**11**:760–9.

200. Ganegoda GU, Li M, Wang W, *et al*. Heterogeneous network model to infer human disease-long intergenic non-coding RNA associations. *IEEE Trans Nanobiosci* 2015;**14**:175–83.

201. Chen X. KATZLDA: KATZ measure for the lncRNA–disease association prediction. *Sci Rep* 2015;**5**:16840.

202. Yang X, Gao L, Guo X, *et al*. A network based method for analysis of lncRNA-disease associations and prediction of lncRNAs implicated in diseases. *PLoS One* 2014;**9**:e87797.

203. Liu M-X, Chen X, Chen G, *et al*. A computational framework to infer human disease-associated long noncoding RNAs. *PLoS One* 2014;**9**:e84408.

204. Chen X. Predicting lncRNA–disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci Rep* 2015;**5**:13186.

205. Takahashi Y, Sawada G, Kurashige J, *et al*. Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *Br J Cancer* 2014;**110**:164–71.

206. Tanaka K, Shiota G, Meguro M, *et al*. Loss of imprinting of long QT intronic transcript 1 in colorectal cancer. *Oncology* 2000;**60**:268–73.

207. Wang H-M, Lu J-H, Chen W-Y, *et al*. Upregulated lncRNA-UCA1 contributes to progression of lung cancer and is closely related to clinical diagnosis as a predictive biomarker in plasma. *Int J Clin Exp Med* 2015;**8**:11824–30.

208. Tano K, Mizuno R, Okada T, *et al*. MALAT-1 enhances cell motility of lung adenocarcinoma cells by influencing the expression of motility-related genes. *FEBS Lett* 2010;**584**:4575–80.

209. Zhang E, Yin D, Sun M, *et al*. P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression. *Cell Death Dis* 2014;**5**:e1243.

210. Choudhry H, Albukhari M, Morotti A, *et al. Tumor hypoxia induces nuclear paraspeckle formation through HIF-2α dependent transcriptional activation of NEAT1 leading to cancer cell survival, Oncogene* 2014;34:4546.

211. Guo X, Xia J, Deng K. Long non-coding RNAs: emerging players in gastric cancer, tumor. *Biology* 2014;**35**:10591–600.

212. Wang J, Su L, Chen X, *et al*. MALAT1 promotes cell proliferation in gastric cancer by recruiting SF2/ASF. *Biomed Pharmacother* 2014;**68**:557–64.

213. Jalali S, Kapoor S, Sivadas A, *et al*. Computational approaches towards understanding human long noncoding RNA biology. *Bioinformatics* 2015;**31**:2241–51.

214. Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol* 2011;**21**:354–61.

215. Park Y, Marcotte EM. Flaws in evaluation schemes for pair-input computational predictions. *Nat Methods* 2012;**9**:1134–6.