

## RESEARCH ARTICLE

# Functional characterization of disease/comorbidity-associated lncRNA

Jing Tang<sup>1,2,†</sup>, Yongheng Wang<sup>2,†</sup>, Jianbo Fu<sup>1,†</sup>, Xianglu Wu<sup>3</sup>, Zhijie Han<sup>1,2</sup>, Chuan Wang<sup>2</sup>, Maiyuan Guo<sup>2</sup>, Yingxiang Wang<sup>2,3</sup>, Yubin Ding<sup>3,\*</sup>, Bo Yang<sup>1,\*</sup>, Feng Zhu<sup>1,\*</sup>

<sup>1</sup> College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

<sup>2</sup> College of Basic Medicine, Chongqing Medical University, Chongqing 400016, China

<sup>3</sup> Joint International Research Laboratory of Reproductive and Development, Department of Reproductive Biology, School of Public Health, Chongqing Medical University, Chongqing 400016, China

\* Correspondence: dingyb@cqmu.edu.cn, yang924@zju.edu.cn, zhufeng@zju.edu.cn

Received July 8, 2020; Revised October 24, 2020; Accepted November 17, 2020

**Background:** Functional characterization of the long noncoding RNAs (lncRNAs) in disease attracts great attention, which results in a limited number of experimentally characterized lncRNAs. The major problems underlying the lack of experimental verifications are considered to come from the significant false-positive assignments and extensive genetic-heterogeneity of disease. These problems are even worse when it comes to the functional characterization in comorbidity (simultaneous/sequential presence of multiple diseases in a patient, and showing much wider prevalence, poorer treatment-response and longer illness-course than a single disease).

**Methods:** Herein, FCCLnc was developed to characterize lncRNA function by (1) integrating diverse SNPs that were associated with 193 diseases standardized by International Classification of Diseases (ICD-11), (2) condition-specific expression of lncRNAs, (3) weighted correlation network of lncRNAs and protein-coding neighboring genes.

**Results:** FCCLnc can characterize lncRNA function in both disease and comorbidity by not only controlling false discovery but also tolerating their disease heterogeneity. Moreover, FCCLnc can provide interactive visualization and full download of lncRNA-centered co-expression network.

**Conclusion:** In summary, FCCLnc is unique in characterizing lncRNA function in diverse diseases and comorbidities and is highly expected to emerge to be an indispensable complement to other available tools. FCCLnc is accessible at <https://idrblab.org/fcclnc/>.

**Keywords:** comorbidity; long noncoding RNA; functional characterization; disease-associated SNPs; guilt-by-association

**Author summary:** Functional characterization of the long noncoding RNAs (lncRNAs) in disease attracts great attention, but has significant false-positive assignments and extensive genetic-heterogeneity of disease especially in comorbidity. Herein, FCCLnc was developed to characterize lncRNA function in diverse diseases and comorbidities, which can be expected to be an indispensable complement to other available tools.

## INTRODUCTION

Functional characterization of the long noncoding RNA (lncRNA) has attracted considerable attentions [1–4] due to their pivotal role in suppressing DNA synthesis [5],

transcriptionally/post-transcriptionally regulating RNAs [6], and modulating the process of protein translation [7]. A typical protocol for inferring the lncRNA function in a system is the guilt-by-association based on differential expression analysis [8]. The weighted correlation network analysis (WGCNA) based on the guilt-by-association

<sup>†</sup> These authors contributed equally to this work.

principle is a well-known type of correlation network [9,10], and often used in combination with differential expression analysis (WGCNA-DEA) for lncRNA function prediction [11–14]. With the recent discovery of lncRNAs as the key regulator of disease pathogenesis [15–18] and drug resistance [19,20], the WGCNA-DEA is anticipated to reveal lncRNA-disease associations [21] and accelerate the experimental discovery of lncRNA function for a studied disease [22]. Till now, this method has been successfully applied to describe the disease phenotype via analyzing co-expression network [13], and identify some lncRNAs associated with disease metastasis [23].

However, only 6,000 (~7%) of over 90,000 lncRNAs in human genome have been characterized as “disease-associated” by experiments [21,24], and the major problems leading to this lack of experimentally verified associations include: (1) the significant number of false positive function assignments [25] and (2) the extensive genetic heterogeneity of many diseases [26,27]. On the one hand, the disease-associated lncRNAs tend to have great inter-individual expression variabilities [8,28], which is largely ignored by differential expression testing [29]. In other words, WGCNA-DEA method can only capture the differential expression of lncRNAs at the population level [29], which brings about a large number of false positive predictions [25]. On the other hand, the disease-associated single-nucleotide polymorphisms (SNPs) can not only alter lncRNAs’ expression level but also modify their secondary structure [30,31]. Since structural variations are not taken into consideration by differential expression analysis, the application of WGCNA-DEA alone may not be capable of fully describing the genetic heterogeneity of complex diseases [26,27,32].

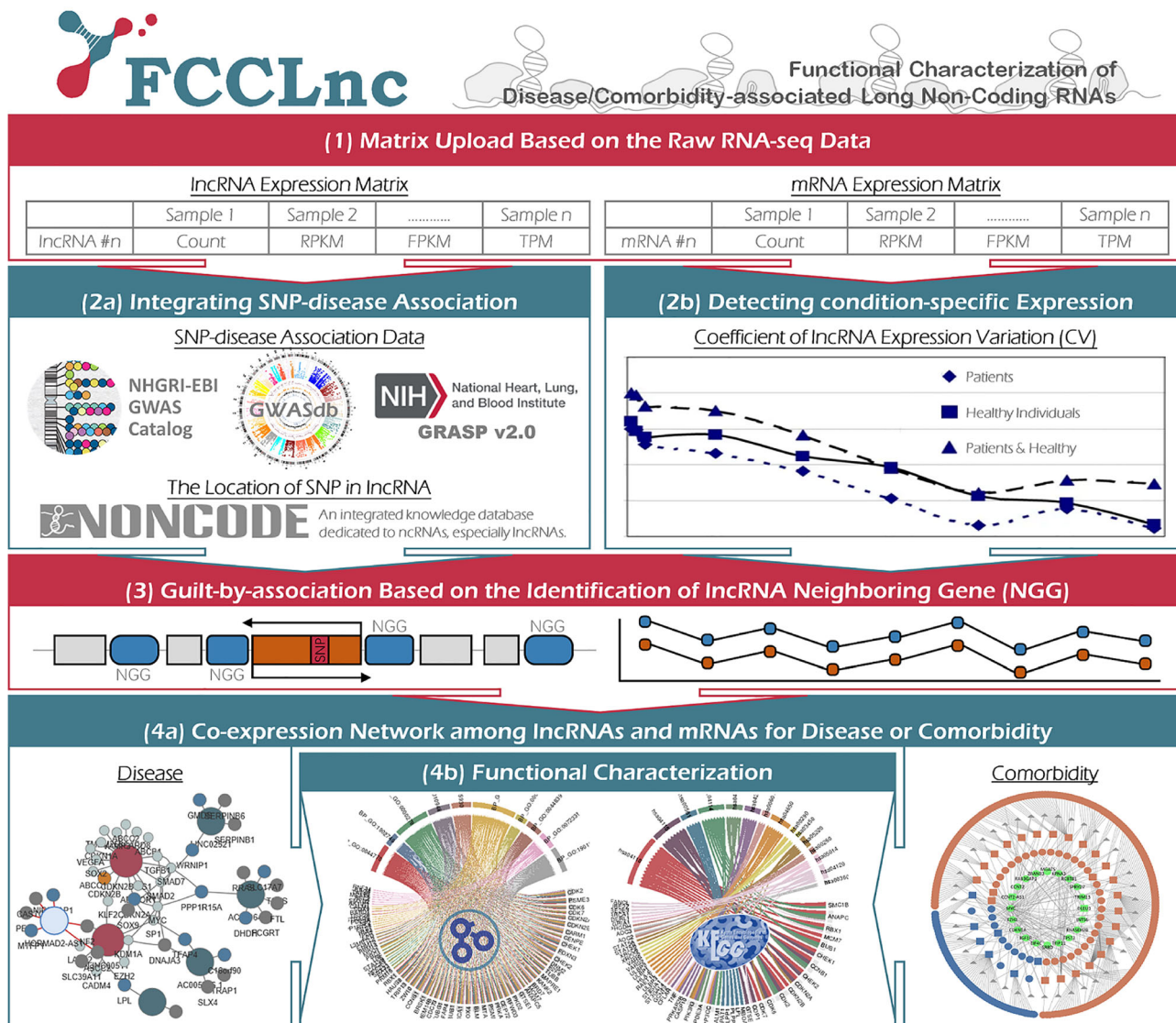
To cope with these major problems, the algorithm for detecting condition-specific expression is proposed to be in place of differential expression testing for describing inter-individual expression variability [8]. Particularly, this variability is assessed using the standard measure ‘coefficient of variation (CV)’ [33]. A low value of CV denotes a lncRNA in normal cell (health control), while a high value represents disease-related lncRNA [8,33]. Moreover, due to the superiority of SNPs in representing disease heterogeneity [34,35], disease-associated SNPs have been integrated with lncRNA expression level for predicting the function of lncRNAs in heterogeneous diseases such as malignancy [36] and diabetes [34]. Thus, it is essential and of great interest to have a tool that is capable of characterizing the lncRNA functions by not only controlling the false discovery rate but also tolerating the disease heterogeneity [8,37–39].

More importantly, due to its much wider prevalence, longer illness-course, and poorer treatment-response than

single disorder, the comorbidity (a simultaneous or sequential presence of multiple indications in one patient) has emerged to be one of the most vital tasks in clinical medicine [40–42], for instance, psychiatric comorbidity [43,44]. The regulation by lncRNAs is critical to the pathogenesis of various comorbidities [45], and is expected to be used to improve the efficacy of current therapies [46]. Compared with the single disease, comorbidity is even more heterogeneous, and shows greater inter-individual variabilities in lncRNA expression [47,48]. Thus, it is also necessary to make a tool effective in characterizing the function of lncRNAs in any comorbidity of interest [45,46].

A variety of powerful online tools have been designed to facilitate the functional characterization of lncRNA [8]. Some of the tools are dedicated as databases that offer the experimentally verified and/or computationally predicted lncRNA data [49–63]. The others are popular web-servers including: AnnoLnc2 [64], Co-LncRNA [65], FARNALnc [25], Lnc-GFP [66], LncRNA2Function [67], LncRNAs2Pathways [68], ncFANs [69], and so on. The majority of the web-servers are based on the construction of co-expression network and/or the differential expression analysis, but neither integrate disease-associated SNP data nor detect condition-specific expression [65–69]. AnnoLnc2 is unique in systematically annotating the newly identified human lncRNA and provides disease-associated SNPs for the annotated lncRNAs [64], but it is specifically for the lncRNAs associated with 39 cancer types [64]. Moreover, none of the available online tools can effectively characterize the function of lncRNAs in comorbidity. It is thus urgently needed to construct online tools that can characterize lncRNA functions (via simultaneously integrating disease-associated SNP data and detecting the condition-specific expression of lncRNAs) for a significantly increased number of diseases and comorbidities. However, no such tool is yet available.

In this study, a novel web-server FCCLnc was therefore constructed. As illustrated in Fig. 1, the FCCLnc is unique in (1) integrating diverse SNP data that are identified to be associated with 193 diseases (standardized by the latest version of WHO International Classification of Diseases, ICD-11 [70–72]), (2) enabling the functional characterization of lncRNAs in all 193 diseases and various comorbidities (the combinations among multiple diseases from those 193 diseases), (3) reducing the false characterization through detecting the condition-specific expression of lncRNA, and (4) offering interactive visualization of the lncRNA-centered co-expression networks. In summary, FCCLnc is distinguished for its capacity of characterizing lncRNA functions for a significantly increased number of diseases and comorbidities, and is therefore expected to emerge to be an indispensable complement to other available tools.



**Figure 1. The workflow of FCCLnc.** There were four key steps: (1) matrix upload based on raw RNA-seq data (upload the expression matrices of both lncRNAs and mRNAs), (2) identification of disease-associated lncRNAs (through integrating SNP-disease association data, and detecting condition-specific expression), (3) constructing the co-expression network among lncRNAs and mRNAs (by a WGCNA based on the neighboring genes of the studied lncRNAs), and (4) lncRNA functional characterization based on the co-expression network among lncRNAs and mRNAs (GO and KEGG enrichments were conducted to facilitate the functional annotation).

FCCLnc can be freely accessible (without login requirement) at: <https://idrblab.org/fcclnc/>.

## RESULTS AND DISCUSSION

### Web service provided by and operating procedure adopted in FCCLnc

To make the usage of FCCLnc convenient, the operating procedure implemented in this tool was provide via four sequential steps (illustrated in Fig. 1). STEP (1): matrices upload based on raw RNA-seq data (upload the expres-

sion matrices of lncRNAs & mRNAs); STEP (2): identification of disease-associated lncRNAs (through integrating SNP-disease association data, and detecting condition-specific expression); STEP (3): constructing a co-expression network among lncRNAs and mRNAs (by the guilt-by-association based on the neighboring genes of the studied lncRNAs); STEP (4): functional characterization of lncRNAs based on the newly constructed co-expression network (GO & KEGG enrichments were conducted to facilitate the functional annotation). The general workflow of FCCLnc integrating all four steps was illustrated in Fig. 1. Detailed user manual and website

demo were systematically provided in the ‘Manual’ panel of FCCLnc.

The systematic analysis on the functions provided by FCCLnc could reveal its uniqueness. First, a direct upload of lncRNA & mRNA expression matrices was allowed in STEP (1), which made it possible to analyze clinical/experimental sequencing data. Such function is very crucial for the researchers working in clinical or precision medicine [40–42]. Second, FCCLnc stood out among available tools by not only integrating SNP-disease association data but also detecting condition-specific expressions of lncRNAs in its STEP (2). These novel features would make it capable of characterizing lncRNA function by simultaneously controlling FDR and tolerating disease heterogeneity [8,37]. Third, the characterization of lncRNA function in comorbidity was realized, for the first time, in FCCLnc by overlapping disease-associated genes among comorbid diseases. FCCLnc enabled the functional characterizations for a very wide range of diseases (193 diseases in total standardized by ICD-11), which included 19 infectious diseases, 36 cancers, 12 immune system disorders, 10 metabolic diseases, 40 neurodevelopmental disorders, 17 digestive system diseases, 15 circulatory system disorders, 10 respiratory system diseases, 4 genitourinary system disorders, 10 musculoskeletal system diseases, 4 developmental anomalies, and 16 other disorders of skin, eyes, ear or mastoid process. The combinations among multiple diseases from those 193 would result in the widest coverage of comorbidities so far. Finally, lncRNA-

centered co-expression network was generated by FCCLnc for the interactive visualization and download. User can navigate to any step of interest through the “BACK” or “NEXT” button of each step. The resulting network for disease/comorbidity could be downloaded in the format of ‘.html’ (for visualization and text mining) and ‘.cys’ (supporting further network analysis in Cytoscape [73]). Experimentally verified lncRNAs were also collected and integrated into FCCLnc for enhancing the visualization of the co-expression network.

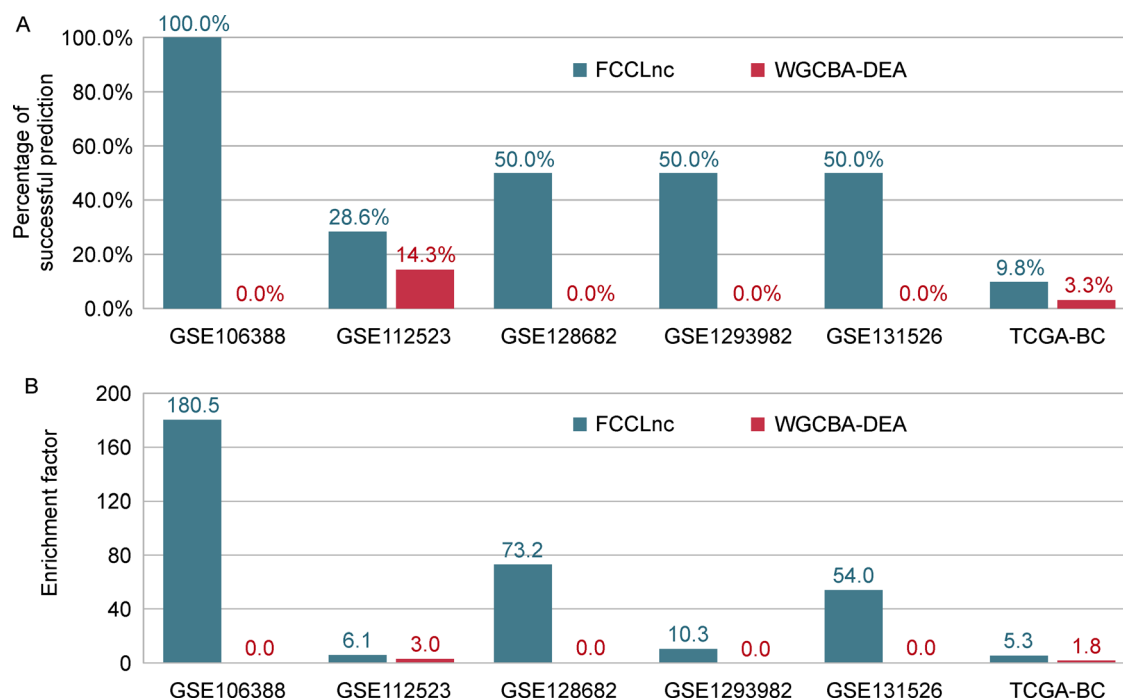
### Characterizing the function of lncRNAs in a particular disease by FCCLnc

In order to evaluate the capacity of FCCLnc in characterizing the function of lncRNAs in disease, six datasets were collected and shown in Table 1, which included: GSE106388 [74], GSE112523 [75], GSE128682 [76], GSE129398 [77], GSE131526 [78], and TCGA-BC [79]. These datasets were transcriptomic data of diverse diseases (including asthma, schizophrenia, ulcerative colitis, obesity, type-I diabetes and breast cancer, respectively). As shown in Fig. 2, the performances of FCCLnc and one of the most frequently utilized methods (WGCNA-DEA) [8] were compared based on two key indexes. First, the percentage of successful prediction (PSP) was utilized to measure method’s success rate (%) in characterizing experimentally verified lncRNAs [21]. As shown in Fig. 2A, the PSPs of FCCLnc varied (from 9.8% for TCGA-BC to

**Table 1** Eight benchmark datasets collected for case study

Dataset ID	Sample details of each dataset	Classification of disease by ICD-11	Expression unit	Number of lncRNAs & mRNAs
GSE106388 [74]	15 mild asthma patients & 4 healthy controls	CA23	Reads count	12,994 lncRNAs & 19,364 mRNAs
GSE112523 [75]	29 SCZ patients & 28 healthy controls	6A20	Reads count	12,179 lncRNAs & 18,437 mRNAs
GSE128682 [76]	14 ulcerative colitis patients & 16 healthy controls	DD71	Reads count	1,756 lncRNAs & 17,355 mRNAs
GSE129398 [77]	12 obesity people & 10 controls of low BMI	5B81	Reads count	822 lncRNAs & 14,300 mRNAs
GSE131526 [78]	12 type-1 diabetes patients & 3 healthy controls	5A10	Reads count	6,798 lncRNAs & 19,614 mRNAs
TCGA-BC [79]	115 breast cancer patients & 113 healthy controls	2C60	FPKM	14,097 lncRNAs & 19,631 mRNAs
GSE78936 [80]	28 SCZ patients & 30 bipolar disorder patients	6A20 & 6A60	FPKM	5,494 lncRNAs & 18,025 mRNAs
GSE133099 [76]	6 type-2 diabetes patients & 6 obesity people	5A11 & 5B81	Reads count	8,929 lncRNAs & 17,518 mRNAs

Note: The first six datasets were the case-control study for single disease (standardized by the latest version of ‘International Classification of Diseases’ provided by the World Health Organization, ICD-11) and the last two datasets were the patients of two different diseases (used to analyze case studies of the corresponding comorbidity). BMI: body mass index; FPKM: fragments per kilobase of exon per million; ICD-11: international classification of diseases 11th revision; SCZ: schizophrenia.



**Figure 2. Performance comparison between FCCLnc and WGCNA-DEA based on two indexes.** (A) The percentage of successful prediction (PSP), and (B) enrichment factor (EF). Particularly, the PSPs (%) in characterizing the experimentally verified 'true' lncRNAs were used as the first index for evaluating the performances. Meanwhile, the EFs were adopted to indicate the concentration of the experimentally verified lncRNAs among the prediction results of a method compared to their concentration throughout the entire lncRNAs in the uploaded expression matrix. EF value is no less than zero. Only when EF is larger than 1, there is an enrichment. The larger the EF, the lower the FDR. Comparisons were based on the six benchmarks shown in Table 1.

100% for GSE106388) and the PSPs of WGCNA-DEA also differed greatly (from 0% for four datasets to 14.3% for GSE112523). It was clear to see that, for all datasets, the PSPs of FCCLnc could significantly surpass those of WGCNA-DEA, which showed the good performances of FCCLnc on characterizing the experimentally verified lncRNAs. Second, the enrichment factor (EF) was further used to assess method's ability to control the false characterization. As shown in Fig. 2B, the EFs of FCCLnc differed greatly (from 5.3 for TCGA-BC to 180.5 for GSE106388) and the EFs of WGCNA-DEA also varied (from 0.0 for four datasets to 3.0 for GSE112523). It was also obvious that, for all datasets, the EFs of FCCLnc were consistently better than those of WGCNA-DEA, which indicated the superior capacity of FCCLnc in controlling the false characterization of lncRNA function.

Based on a breast cancer dataset TCGA-BC collected from The Cancer Genome Atlas [79], the performances of FCCLnc and WGCNA-DEA were further compared based on the enrichment analysis of KEGG pathways and GO terms. As illustrated in Fig. 3A, besides the chord diagram on the left side, two additional pie charts were

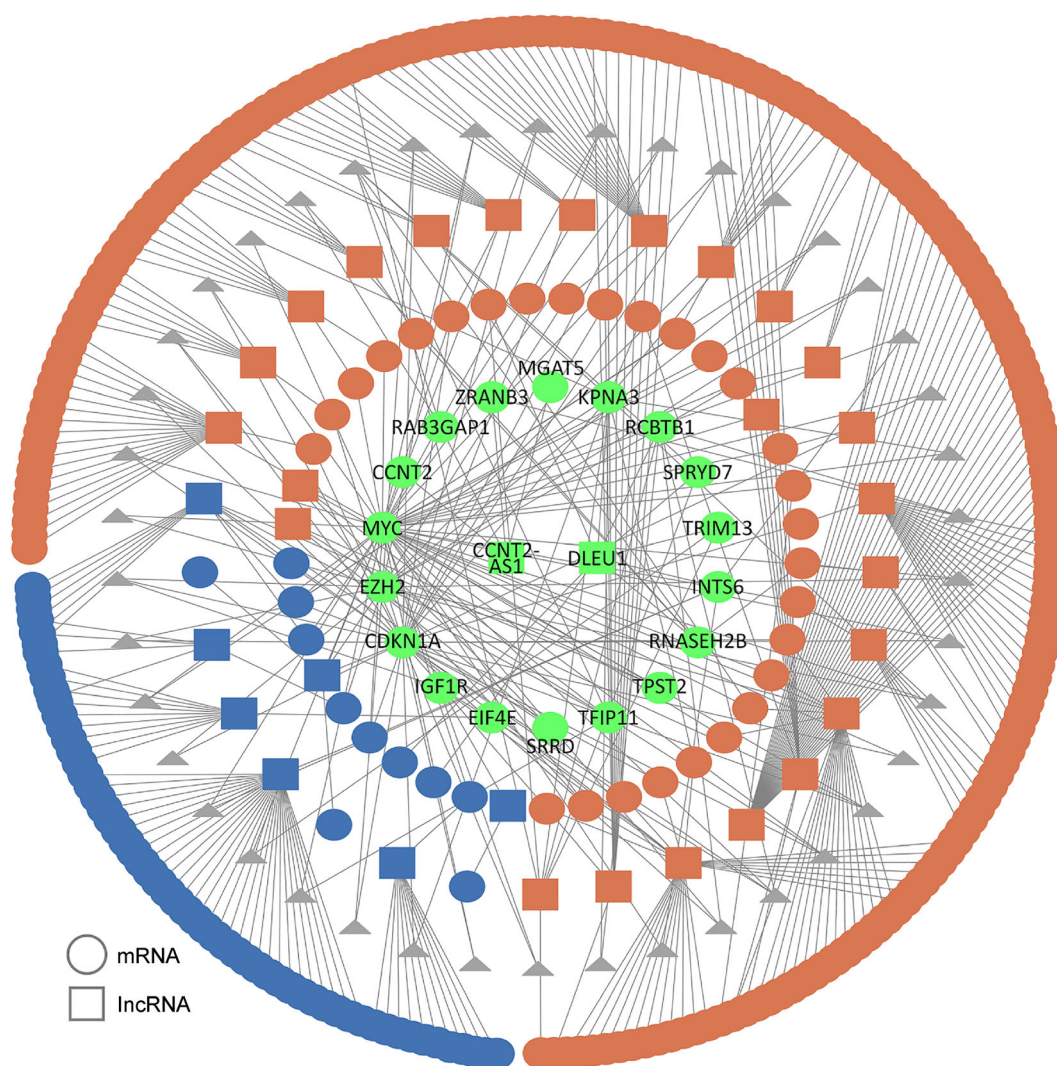
drawn to illustrate the percentages of pathways (enriched based on FCCLnc (upper) and WGCNA-DEA (lower)) that were further confirmed by literature search to be closely related to breast cancer. As shown, 95% of the pathways enriched ( $p$ -value < 0.05) based on FCCLnc were also confirmed by previously reported literatures (the detail descriptions were provided in Supplementary Table S1), and 60% of the enriched pathways ( $p$ -value < 0.05) were identified by FCCLnc only. When it comes to WGCNA-DEA, the percentage confirmed by previous reports decreased to 79%, and only 55% of the enriched pathways were solely identified by WGCNA-DEA. Like pathways, the results of GO terms enrichment were illustrated in Fig. 3B. 83% of the GO terms enriched (adjusted  $p$ -value < 0.05) based on FCCLnc were confirmed by previously reported literatures (the detail descriptions were provided in Supplementary Table S2) and 78% of those enriched GO terms (adjusted  $p$ -value < 0.05) were found by FCCLnc only. For WGCNA-DEA, its percentage confirmed by previous reports was only 66%. All in all, these results further validated the good performance of FCCLnc in lncRNA functional annotation.



### Characterizing the function of lncRNAs in certain comorbidity by FCCLnc

LncRNA regulation is essential to the pathogenesis of various comorbidities [45]. Therefore, the ability of FCCLnc to characterize the function of lncRNAs in certain comorbidity was evaluated. Particularly, two datasets were collected (as shown in Table 1), which included GSE133099 [76] (containing patients of type-2 diabetes and obesity), and GSE78936 [80] (containing patients of schizophrenia and bipolar disorder). Comorbidity-associated lncRNAs were therefore identified by finding the genetic factors (“common disease genes”) shared by the comorbid diseases [81,82]. Figure 4 showed

the lncRNAs and mRNAs co-expression networks in obesity patients comorbid with type-2 diabetes (identified using GSE133099 [76]). As shown, two lncRNAs (DLEU1 and CCNT2-AS1) and their eighteen co-expressed mRNAs were found to be common disease genes. Recent study has revealed the essential roles of DLEU1 in thyroid hormone synthesis [83]. Since thyroid hormones are strongly associated with both obesity [84–86] and diabetes [87,88], DLEU1 may be considered to be a critical regulator of this comorbidity. Meanwhile, although there were few studies on CCNT2-AS1, its co-expressed mRNA CCNT2 (Fig. 4) had been reported to possess genetic variations that were significantly associated with obesity and diabetes [89].



**Figure 4.** The co-expression network illustrating the regulation among lncRNAs and mRNAs in obesity patients comorbid with type-2 diabetes. The network was constructed by FCCLnc based on dataset GSE133099 [76]. The lncRNAs/mRNAs solely involved in obesity or type-2 diabetes were shown in blue or orange, respectively, while the common disease genes between obesity or type-2 diabetes were highlighted in green. Two common disease lncRNAs (green square) shared by both diseases were identified which included DLEU1 and CCNT2-AS1.

Similarly, Supplementary Fig. S1 showed the co-expression network between lncRNAs and mRNAs in the schizophrenia patients comorbid with bipolar disorder (constructed using GSE78936 [80]). As shown, three lncRNAs (LINC00243, MIR3681HG and TSBP1-AS1) and seven co-expressed mRNAs were discovered to be common disease genes. Although there were few reports on those identified lncRNAs, some of their co-expressed mRNAs (Supplementary Fig. S1) have been reported to be strongly associated with both comorbid diseases. For example, the mRNA *ATAT1*, co-expressed with LINC00243, was found to be extensively associated with the pathogenesis of schizophrenia [90] and bipolar disorder [91]; the *ROCK2*, co-expressed with MIR3681HG, was reported to play an important role in both schizophrenia [92] and bipolar disorder [93]. All in all, it is feasible to use FCCLnc to characterize the function of lncRNAs in certain comorbidity.

## CONCLUSIONS

FCCLnc is distinguished for its capacity to characterize the lncRNA functions for a significantly increased number of diseases and comorbidities, and it is expected to emerge as an indispensable complement to other available tools. With the rapid accumulation of next generation sequencing data, FCCLnc and other powerful tools could collectively promote various life science researches, including pathological study, precision medicine, drug/target discovery, disease association, and biomarker identification.

## MATERIALS AND METHODS

### Discovering the potentially disease-associated lncRNAs by SNP-disease associations

The majority of disease-associated SNPs located in lncRNAs modify the secondary structures or influence expression levels, thereby affecting their regulatory function, hence contributing to the development of disease [30]. In this study, SNP-disease association data were therefore collected to facilitate the identification of potentially disease-associated lncRNAs. First, the SNP-disease associations were collected from three well-known sources: GWASdb [94], NHGRI-EBI GWAS Catalog [95] and GRASP2 [96]. This led to 17,842 SNPs located in the lncRNAs associated with ( $p$ -value < 0.001) 156 diseases (standardized by the latest version of ‘International Classification of Diseases’ provided by the World Health Organization, ICD-11). Moreover, a literature review of over 350 recent publications yielded 4,616 additional SNPs located in the lncRNAs associated with 72 standardized diseases. All in all, 24,339

associations between 193 standardized diseases and 22,458 SNPs were collected for subsequent analysis. Second, the chromosome information of lncRNAs was downloaded from the NONCODEV5 (human reference genome hg38) [53] to match the disease-associated SNPs to the lncRNA region. Finally, 10,936 lncRNAs (with at least one disease-associated SNP) were identified to be “potentially disease-associated”.

### Detecting the inter-individual variability of lncRNA by condition-specific expression

The lncRNAs transcribed in a certain disease indication tend to show high expression variability, and the algorithm for detecting condition-specific expression is therefore proposed to be in place of differential expression testing for describing such inter-individual expression variability [8]. The variability could be assessed using the standard measure ‘coefficient of variation (CV)’ [33,97,98]. A low value of CV denotes a lncRNA in normal cell (healthy individual), while a high value represents the disease-related lncRNA [8,33]. Herein, the CV is first defined as the ratio between the standard deviation of the lncRNA expression levels measured across the patients and its mean [33]. Using those “potentially disease-associated” lncRNAs identified in previous section, their CV values were then calculated and ranked. Finally, top-N ranked lncRNAs (N = 100, 200, 300, 400 or 500 depending on user’s preference) were selected and identified as “disease-associated” ones for the subsequent functional characterization.

### Constructing the co-expression network based on lncRNAs’ neighboring genes

Co-expressed genes were known as more likely co-regulated and functionally associated, which made the identification of the co-expressed neighboring protein-coding gene helpful in assigning lncRNA function [8,69,99,100]. To construct the co-expression network between disease-associated lncRNAs and their neighboring genes, the WGCNA [9] based on the guilt-by-association principle was used to compute co-expression network of lncRNAs with their neighboring genes. Herein, the comprehensive data of 96,308 lncRNAs and 19,975 protein coding genes were first collected from NONCODEV5 (human reference genome hg38) [53] and GENCODEV31 (human reference genome hg38) [51], respectively. Then, the neighboring genes within 5 kb [101], 10 kb [102], 20 kb [103], 50 kb [104], 70 kb [105], 100 kb [106], 200 kb [107], 300 kb [108], 400 kb [109] and 500 kb [110] up/downstream of the studied lncRNAs were calculated based on the collected location information, which resulted in a collection of neighboring genes

of the studied disease-associated lncRNAs. WGCNA [9] was used to compute a co-expression network based on the studied lncRNAs and their neighboring genes. The resulting co-expression network was illustrated in FCCLnc, and it could be downloaded in the formats of '.html' (for visualization analysis) and '.cys' (supporting the network analysis in Cytoscape [73]). In the meantime, ~6,000 experimentally verified lncRNAs were collected from LncRNADisease [21] and LncRNA2Target [50]. Such data were integrated into FCCLnc for enhancing the visualization of the co-expression network by highlighting the lncRNAs with experimental verification information.

### Annotating the lncRNA function based on gene ontology and KEGG pathway

The Gene Ontology (GO) [111] and Kyoto Encyclopedia of Genes and Genome (KEGG) pathway [112,113] were widely adopted to characterize the function of disease-associated lncRNAs. Herein, the GO annotations (containing biological processes, molecular functions, and cellular components) as well as KEGG pathways of proteins-coding gene were downloaded from the gene set enrichment analysis (GSEA) database [114]. Then, GO and KEGG enrichment analysis were performed using the mRNAs identified by FCCLnc. In particular, the statistical significance of GO terms and KEGG pathways enrichments were evaluated by hypergeometric test with  $p$ -value less than 0.05 [115]. Finally, a chord diagram illustrating the enrichment results was displayed directly in FCCLnc online server.

### Characterizing the lncRNA function in comorbidity using common disease genes

The mechanism of lncRNAs in the comorbid diseases were frequently explained by their shared genetic factors (namely "common disease genes") [81,82]. In other words, direct overlap of the disease-associated genes among comorbid disease was found to be one of the critical factors for explaining the corresponding comorbidity [81,82]. Thus, the upload of RNA expression matrices that containing the data of multiple diseases was allowed in FCCLnc. First, the RNA expression data of each disease were analyzed using the sequential steps discussed in the first three sections of Materials and Methods, which resulted in multiple co-expression networks (each contained a great many of lncRNAs & mRNAs). Then, a direct overlap of disease-associated RNAs among multiple diseases was conducted, which identified a set of common disease genes (RNAs). Third, multiple co-expression networks were linked together based on this set of common disease genes, and the

resulting network is the co-expression network of the corresponding comorbidity. Finally, the common disease lncRNAs were therefore characterized as comorbidity-associated, and their function was further annotated by their co-expressed mRNAs. The enrichment results of multiple co-expression networks could also be overlapped to facilitate the functional annotation.

### Evaluating the ability of FCCLnc to characterize the function of lncRNA in disease

Two indexes were adopted here to evaluate the ability of FCCLnc to characterize the function of disease-associated lncRNAs, both of which were based on the experimentally validated disease-associated lncRNAs. These indexes included: (1) PSP, and (2) EF. Particularly, the experimentally verified disease-associated lncRNAs were collected directly from LncRNADisease [21], which provided a number of "true" lncRNAs for each of those 193 diseases, and the PSPs (%) of FCCLnc in characterizing these true lncRNAs were used as the first index for evaluating the performances. Moreover, the EF was adopted here to denote the concentration of the experimentally verified lncRNAs among the prediction results of FCCLnc compared to their concentration throughout the entire lncRNAs in expression matrix, which was known to be effective in assessing false discovery by fully considering the real-world true lncRNAs [115]. EF could be represented by:

$$EF = \frac{N_{\text{trueFCCLnc}}/N_{\text{FCCLnc}}}{N_{\text{true}}/N_{\text{all}}},$$

where  $N_{\text{FCCLnc}}$  denoted the number of lncRNAs characterized by FCCLnc as 'disease-associated';  $N_{\text{trueFCCLnc}}$  represented the number of 'true' lncRNAs successfully characterized by the FCCLnc as 'disease-associated';  $N_{\text{all}}$  was the total number of lncRNAs in the expression matrix uploaded by users; and  $N_{\text{true}}$  indicated the number of 'true' lncRNAs in the uploaded expression matrix based on the LncRNADisease data [21]. EF value is no less than zero. Only when an EF is larger than 1, there is an enrichment. The larger the EF, the lower the FDR.

The weighted correlation network analysis based on differential expression lncRNA (WGCNA-DEA) frequently utilized method for inferring the lncRNA functions [12,13] was assessed here to compare with FCCLnc. Based on RNA-seq data of control-case studies, WGCNA-DEA was implemented based on two procedures. First, differential expression analysis was conducted for finding the differential lncRNA by R package DESeq2 [116]. Second, the weighted correlation network between differential lncRNA and protein-coding neighboring genes was built.

## Required data formats of FCCLnc input files and server implementation details

There were two input files for FCCLnc analysis that provided the expression matrices of lncRNAs and mRNAs, respectively. The required formats of both input files were almost the same, which should provide two RNA-by-sample matrices in csv format. Particularly, the sample ID and class of samples are sequentially provided in the first two rows of input file. The sample ID is uniquely assigned according to users' preferences. Title of the second row must be kept to "label" without change during the analysis, and class of samples indicates the sample groups. For characterizing the lncRNA function in single disease, there should be only two types of group names in each of those two input files (using the exact word 'control' to indicate healthy individuals, and denoting all disease samples by disease name. For example, users could use 'diabetes' or 'type-II diabetes' to label the samples of type-II diabetes patients). For characterization in comorbidity, all samples should be labeled by their disease name. For instance, when studying the anxiety comorbidity in schizophrenia, user could use 'Anxiety' and 'SCZ' to denote the class of samples in the uploaded expression matrices. The sample ID and class of samples must be identical in both matrices. The unique RNA IDs (NONCODE ID and Ensemble Gene ID for lncRNA and mRNA, respectively) should be provided. The expression matrix could be reads counts, transcripts per kilobase million (TPM), reads per kilobase million (RPKM) or fragments per kilobase million (FPKM). Example files can be directly downloaded from the "Analysis" panel of FCCLnc.

FCCLnc website is deployed on a server running Cent OS Linux v7.0 operating system, Apache Tomcat servlet container and Apache HTTP web server v2.2.15. Its interface was developed by R v3.6.2 and R package Shiny v0.13.1 running on Shiny-server v1.4.1.759 [117,118]. A variety of packages in R package were utilized in the background processes [119], which included SpeCond, htmlTable, visNetwork, networkD3, htmltools, WGCNA, shinythemes, shiny, shinyjs, shinyBS, and shinydashboard. The popular browsers such as Google Chrome, Mozilla Firefox, Safari, and Internet Explorer (11 or later) were compatible with the official website of FCCLnc (without login requirement).

## The sample datasets collected for validating the case studies in this work

To test the utility of FCCLnc, the expression matrices of both lncRNA and mRNA were collected from Gene Expression Omnibus (GEO) [120] and The Cancer Genome Atlas (TCGA) [79]. There were eight benchmark

datasets in total collected for analyzing case study. As shown in Table 1, the first six datasets were used for the case-control studies [121] for six single disease, and the last two datasets contained the patients of two different diseases (utilized for the case study of the corresponding comorbidity). Sample details in each benchmark dataset were shown in Table 1, which included the detail of expression unit and the number of lncRNAs and mRNAs.

## SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at <https://doi.org/10.15302/J-QB-021-0247>.

## AUTHOR CONTRIBUTIONS

F.Z. conceived the idea, supervised the work and wrote the manuscript. J.T., Y.W. and J.F. performed the research. J.T., Y.W. and J.F. constructed the web-server and wrote the scripts. J.T., Y.W., J.F., X.W., Z.H., C.W., M.G., Y. W., Y.D. and B.Y. prepared and analyzed the data. All authors reviewed and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

This work was funded by the National Natural Science Foundation of China (81872798 & U1909208), Natural Science Foundation of Zhejiang Province (LR21H300001), National Key R&D Program of China (2018YFC0910500), Leading Talent of the "Ten Thousand Plan" – National High-Level Talents Special Support Plan of China; Fundamental Research Fund for Central Universities (2018QNA7023), "Double Top-Class" University Project (181201\*194232101), and Key R&D Program of Zhejiang Province (2020C03010). This work was supported by Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare; Alibaba Cloud; Information Technology Center of Zhejiang University.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Jing Tang, Yongheng Wang, Jianbo Fu, Xianglu Wu, Zhijie Han, Chuan Wang, Maiyuan Guo, Yingxiong Wang, Yubin Ding, Bo Yang and Feng Zhu declare that they have no conflict of interests or financial conflicts to disclose.

All procedures performed in studies were in accordance with the ethical standards of the institution or practice at which the studies were conducted, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or

exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Kopp, F. and Mendell, J. T. (2018) Functional classification and experimental dissection of long noncoding RNAs. *Cell*, 172, 393–407
- Zhou, J., Zhang, S., Wang, H. and Sun, H. (2017) LncFunNet: an integrated computational framework for identification of functional long noncoding RNAs in mouse skeletal muscle cells. *Nucleic Acids Res.*, 45, e108
- Antonov, I. V., Mazurov, E., Borodovsky, M. and Medvedeva, Y. A. (2019) Prediction of lncRNAs and their interactions with nucleic acids: benchmarking bioinformatics tools. *Brief. Bioinform.*, 20, 551–564
- Yin, J., Sun, W., Li, F., Hong, J., Li, X., Zhou, Y., Lu, Y., Liu, M., Zhang, X., Chen, N., *et al.* (2020) VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res.*, 48, D1042–D1050
- Jiang, S., Cheng, S. J., Ren, L. C., Wang, Q., Kang, Y. J., Ding, Y., Hou, M., Yang, X. X., Lin, Y., Liang, N., *et al.* (2019) An expanded landscape of human long noncoding RNA. *Nucleic Acids Res.*, 47, 7842–7856
- Stojic, L., Niemczyk, M., Orjalo, A., Ito, Y., Ruijter, A. E., Uribe-Lewis, S., Joseph, N., Weston, S., Menon, S., Odom, D. T., *et al.* (2016) Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. *Nat. Commun.*, 7, 10406
- Kondrashov, A. V., Kieffmann, M., Ebnet, K., Khanam, T., Muddashetty, R. S. and Brosius, J. (2005) Inhibitory effect of naked neural BC1 RNA or BC200 RNA on eukaryotic *in vitro* translation systems is reversed by poly(A)-binding protein (PABP). *J. Mol. Biol.*, 353, 88–103
- Signal, B., Gloss, B. S. and Dinger, M. E. (2016) Computational approaches for functional prediction and characterisation of long noncoding RNAs. *Trends Genet.*, 32, 620–637
- Fu, T. T., Tu, G., Ping, M., Zheng, G. X., Yang, F. Y., Yang, J. Y., Zhang, Y., Yao, X. J., Xue, W. W. and Zhu, F. (2020) Subtype-selective mechanisms of negative allosteric modulators binding to group I metabotropic glutamate receptors. *Acta Pharmacol. Sin.*, doi: 10.1038/s41401-020-00541-z
- Yang, Q., Li, B., Chen, S., Tang, J., Li, Y., Li, Y., Zhang, S., Shi, C., Zhang, Y., Mou, M., *et al.* (2021) MMEASE: Online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis. *J. Proteomics*, 232, 104023
- Qu, S., Shi, Q., Xu, J., Yi, W. and Fan, H. (2020) Weighted gene coexpression network analysis reveals the dynamic transcriptome regulation and prognostic biomarkers of hepatocellular carcinoma. *Evol. Bioinform. Online*, 16, 1176934320920562
- Zhou, Y., Lutz, P. E., Wang, Y. C., Ragoussis, J. and Turecki, G. (2018) Global long non-coding RNA expression in the rostral anterior cingulate cortex of depressed suicides. *Transl. Psychiatry*, 8, 224
- Li, Y. H., Li, X. X., Hong, J. J., Wang, Y. X., Fu, J. B., Yang, H., Yu, C. Y., Li, F. C., Hu, J., Xue, W. W., *et al.* (2020) Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs. *Brief. Bioinform.*, 21, 649–662
- Tang, J., Mou, M., Wang, Y., Luo, Y. and Zhu, F. (2020) MetaFS: Performance assessment of biomarker discovery in metaproteomics. *Brief. Bioinform.*, bbaa105
- Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., Zhang, Q., Yan, G. and Cui, Q. (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, 41, D983–D986
- Chen, Y. G., Satpathy, A. T. and Chang, H. Y. (2017) Gene regulation in the immune system by long noncoding RNAs. *Nat. Immunol.*, 18, 962–972
- Tang, J., Wang, Y., Fu, J., Zhou, Y., Luo, Y., Zhang, Y., Li, B., Yang, Q., Xue, W., Lou, Y., *et al.* (2020) A critical assessment of the feature selection methods used for biomarker discovery in current metaproteomics studies. *Brief. Bioinform.*, 21, 1378–1390
- Hong, J., Luo, Y., Mou, M., Fu, J., Zhang, Y., Xue, W., Xie, T., Tao, L., Lou, Y. and Zhu, F. (2020) Convolutional neural network-based annotation of bacterial type IV secretion system effectors with enhanced accuracy and reduced false discovery. *Brief. Bioinform.*, 21, 1825–1836
- Lu, C., Wei, Y., Wang, X., Zhang, Z., Yin, J., Li, W., Chen, L., Lyu, X., Shi, Z., Yan, W., *et al.* (2020) DNA-methylation-mediated activating of lncRNA SNHG12 promotes temozolomide resistance in glioblastoma. *Mol. Cancer*, 19, 28
- Yin, J., Li, F., Zhou, Y., Mou, M., Lu, Y., Chen, K., Xue, J., Luo, Y., Fu, J., He, X., *et al.* (2021) INTEDE: interactome of drug-metabolizing enzymes. *Nucleic Acids Res.*, 49, D1233–D1243
- Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q. and Dong, D. (2019) LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, 47, D1034–D1037
- Parikhshak, N. N., Swarup, V., Belgard, T. G., Irimia, M., Ramaswami, G., Gandal, M. J., Hartl, C., Leppa, V., Ubieta, L. T., Huang, J., *et al.* (2016) Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, 540, 423–427
- Yang, Y., Chen, L., Gu, J., Zhang, H., Yuan, J., Lian, Q., Lv, G., Wang, S., Wu, Y., Yang, Y. T., *et al.* (2017) Recurrently deregulated lncRNAs in hepatocellular carcinoma. *Nat. Commun.*, 8, 14421
- Volders, P. J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P. and Vandesompele, J. (2019) LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.*, 47, D135–D139
- Alam, T., Uludag, M., Essack, M., Salhi, A., Ashoor, H., Hanks, J. B., Kapfer, C., Mineta, K., Gojbori, T. and Bajic, V. B. (2017) FARNAs: knowledgebase of inferred functions of non-coding RNA transcripts. *Nucleic Acids Res.*, 45, 2838–2848
- Liley, J., Todd, J. A. and Wallace, C. (2017) A method for

- identifying genetic heterogeneity within phenotypically defined disease subgroups. *Nat. Genet.*, 49, 310–316
27. Yan, X., Liang, A., Gomez, J., Cohn, L., Zhao, H. and Chupp, G. L. (2017) A novel pathway-based distance score enhances assessment of disease heterogeneity in gene expression. *BMC Bioinformatics*, 18, 309
  28. Kornienko, A. E., Dotter, C. P., Guenzl, P. M., Gisslinger, H., Gisslinger, B., Cleary, C., Kralovics, R., Pauler, F. M. and Barlow, D. P. (2016) Long non-coding RNAs display higher natural expression variation than protein-coding genes in healthy humans. *Genome Biol.*, 17, 14
  29. Peng, F., Wang, R., Zhang, Y., Zhao, Z., Zhou, W., Chang, Z., Liang, H., Zhao, W., Qi, L., Guo, Z., *et al.* (2017) Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma. *Mol. Cancer*, 16, 98
  30. Castellanos-Rubio, A. and Ghosh, S. (2019) Disease-associated SNPs in inflammation-related lncRNAs. *Front. Immunol.*, 10, 420
  31. Han, Z., Xue, W., Tao, L., Lou, Y., Qiu, Y. and Zhu, F. (2020) Genome-wide identification and analysis of the eQTL lncRNAs in multiple sclerosis based on RNA-seq data. *Brief. Bioinform.*, 21, 1023–1037
  32. Li, P., Guo, M., Wang, C., Liu, X. and Zou, Q. (2015) An overview of SNP interactions in genome-wide association studies. *Brief. Funct. Genomics*, 14, 143–155
  33. Ecker, S., Pancaldi, V., Rico, D. and Valencia, A. (2015) Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. *Genome Med.*, 7, 8
  34. Li, L., Cheng, W. Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E. P. and Dudley, J. T. (2015) Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.*, 7, 311ra174
  35. Nguyen, Q. and Carninci, P. (2016) Expression specificity of disease-associated lncRNAs: toward personalized medicine. *Curr. Top. Microbiol. Immunol.*, 394, 237–258
  36. Shah, M. Y., Ferracin, M., Pileczki, V., Chen, B., Redis, R., Fabris, L., Zhang, X., Ivan, C., Shimizu, M., Rodriguez-Aguayo, C., *et al.* (2018) Cancer-associated rs6983267 SNP and its accompanying long noncoding RNA *CCAT2* induce myeloid malignancies via unique SNP-specific RNA mutations. *Genome Res.*, 28, 432–447
  37. Liu, S. J. and Lim, D. A. (2018) Modulating the expression of long non-coding RNAs for functional studies. *EMBO Rep.*, 19, e46955
  38. Wang, Y., Li, F., Zhang, Y., Zhou, Y., Tan, Y., Chen, Y. and Zhu, F. (2020) Databases for the targeted COVID-19 therapeutics. *Br. J. Pharmacol.*, 177, 4999–5001
  39. Tang, J., Wang, Y., Luo, Y., Fu, J., Zhang, Y., Li, Y., Xiao, Z., Lou, Y., Qiu, Y. and Zhu, F. (2020) Computational advances of tumor marker selection and sample classification in cancer proteomics. *Comput. Struct. Biotechnol. J.*, 18, 2012–2025
  40. Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S. and Guthrie, B. (2012) Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*, 380, 37–43
  41. Zhang, Y., Ying, J. B., Hong, J. J., Li, F. C., Fu, T. T., Yang, F. Y., Zheng, G. X., Yao, X. J., Lou, Y., Qiu, Y., *et al.* (2019) How does chirality determine the selective inhibition of histone deacetylase 6? A lesson from Trichostatin A enantiomers based on molecular dynamics. *ACS Chem. Neurosci.*, 10, 2467–2480
  42. Cui, X., Yang, Q., Li, B., Tang, J., Zhang, X., Li, S., Li, F., Hu, J., Lou, Y., Qiu, Y., *et al.* (2019) Assessing the effectiveness of direct data merging strategy in long-term and large-scale pharmacometabonomics. *Front. Pharmacol.*, 10, 127
  43. Xue, W., Yang, F., Wang, P., Zheng, G., Chen, Y., Yao, X. and Zhu, F. (2018) What contributes to serotonin-norepinephrine reuptake inhibitors' dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem. Neurosci.*, 9, 1128–1140
  44. McIntyre, R. S., Rosenbluth, M., Ramasubbu, R., Bond, D. J., Taylor, V. H., Beaulieu, S. and Schaffer, A., and the Canadian Network for Mood and Anxiety Treatments (CANMAT) Task Force. (2012) Managing medical and psychiatric comorbidity in individuals with major depressive disorder and bipolar disorder. *Ann Clin Psychiatry*, 24, 163–169
  45. Kato, M. and Natarajan, R. (2014) Diabetic nephropathy-emerging epigenetic mechanisms. *Nat. Rev. Nephrol.*, 10, 517–530
  46. Reddy, M. A., Zhang, E. and Natarajan, R. (2015) Epigenetic mechanisms in diabetic complications and metabolic memory. *Diabetologia*, 58, 443–455
  47. Geronazzo-Alman, L., Guffanti, G., Eisenberg, R., Fan, B., Musa, G. J., Wicks, J., Bresnahan, M., Duarte, C. S. and Hoven, C. (2018) Comorbidity classes and associated impairment, demographics and 9/11-exposures in 8,236 children and adolescents. *J. Psychiatr. Res.*, 96, 171–177
  48. Schuckit, M. A. (2006) Comorbidity between substance use disorders and psychiatric conditions. *Addiction*, 101, 76–88
  49. Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., Li, X., Fang, Y., Shen, W., Xu, Y., *et al.* (2019) Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.*, 47, D1028–D1033
  50. Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., Zhou, W., Liu, G., Jiang, H. and Jiang, Q. (2019) LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.*, 47, D140–D144
  51. Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, 47, D766–D773
  52. Ma, L., Cao, J., Liu, L., Du, Q., Li, Z., Zou, D., Bajic, V. B. and Zhang, Z. (2019) LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.*, 47, D128–D134
  53. Yang, Q. X., Wang, Y. X., Li, F. C., Zhang, S., Luo, Y. C., Li, Y., Tang, J., Li, B., Chen, Y. Z., Xue, W. W., *et al.* (2019)

- Identification of the gene signature reflecting schizophrenia's etiology by constructing artificial intelligence-based method of enhanced reproducibility. *CNS Neurosci. Ther.*, 25, 1054–1063
54. Fu, J., Tang, J., Wang, Y., Cui, X., Yang, Q., Hong, J., Li, X., Li, S., Chen, Y., Xue, W., *et al.* (2018) Discovery of the consistently well-performed analysis chain for SWATH-MS based pharmacoproteomic quantification. *Front. Pharmacol.*, 9, 681
  55. Miao, Y. R., Liu, W., Zhang, Q. and Guo, A. Y. (2018) lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Res.*, 46, D276–D280
  56. Paraskevopoulou, M. D., Vlachos, I. S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T., *et al.* (2016) DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.*, 44, D231–D238
  57. Zheng, L. L., Li, J. H., Wu, J., Sun, W. J., Liu, S., Wang, Z. L., Zhou, H., Yang, J. H. and Qu, L. H. (2016) deepBase v2.0: identification, expression, evolution and function of small RNAs, lncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res.*, 44, D196–D202
  58. Li, X. X., Yin, J., Tang, J., Li, Y., Yang, Q., Xiao, Z., Zhang, R., Wang, Y., Hong, J., Tao, L., *et al.* (2018) Determining the balance between drug efficacy and safety by the network and biological system profile of its therapeutic target. *Front. Pharmacol.*, 9, 1245
  59. Quek, X. C., Thomson, D. W., Maag, J. L., Bartonicek, N., Signal, B., Clark, M. B., Gloss, B. S. and Dinger, M. E. (2015) lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.*, 43, D168–D173
  60. Fu, T., Zheng, G., Tu, G., Yang, F., Chen, Y., Yao, X., Li, X., Xue, W. and Zhu, F. (2018) Exploring the binding mechanism of metabotropic glutamate receptor 5 negative allosteric modulators in clinical trials by molecular dynamics simulations. *ACS Chem. Neurosci.*, 9, 1492–1502
  61. Liu, K., Yan, Z., Li, Y. and Sun, Z. (2013) Linc2GO: a human lincRNA function annotation resource based on ceRNA hypothesis. *Bioinformatics*, 29, 2221–2222
  62. Xue, W., Wang, P., Tu, G., Yang, F., Zheng, G., Li, X., Li, X., Chen, Y., Yao, X. and Zhu, F. (2018) Computational identification of the binding mechanism of a triple reuptake inhibitor amitriptyline for the treatment of major depressive disorder. *Phys. Chem. Chem. Phys.*, 20, 6606–6616
  63. Liao, Z. J., Li, D. P., Wang, X. R., Li, L. S. and Zou, Q. (2018) Cancer diagnosis through isomiR expression with machine learning method. *Curr. Bioinform.*, 13, 57–63
  64. Ke, L., Yang, D. C., Wang, Y., Ding, Y. and Gao, G. (2020) AnnoLnc2: the one-stop portal to systematically annotate novel lncRNAs for human and mouse. *Nucleic Acids Res.*, 48, W230–W238
  65. Zhao, Z., Bai, J., Wu, A., Wang, Y., Zhang, J., Wang, Z., Li, Y., Xu, J. and Li, X. (2015) Co-LncRNA: investigating the lncRNA combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data. *Database (Oxford)*, 2015, bav082
  66. Guo, X., Gao, L., Liao, Q., Xiao, H., Ma, X., Yang, X., Luo, H., Zhao, G., Bu, D., Jiao, F., *et al.* (2013) Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res.*, 41, e35
  67. Wang, P., Zhang, X., Fu, T., Li, S., Li, B., Xue, W., Yao, X., Chen, Y. and Zhu, F. (2017) Differentiating physicochemical properties between addictive and nonaddictive ADHD drugs revealed by molecular dynamics simulation studies. *ACS Chem. Neurosci.*, 8, 1416–1428
  68. Li, Y. H., Xu, J. Y., Tao, L., Li, X. F., Li, S., Zeng, X., Chen, S. Y., Zhang, P., Qin, C., Zhang, C., *et al.* (2016) SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PLoS One*, 11, e0155290
  69. Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S., Cao, Q., Xue, W., Chen, N. and Zhu, F. (2016) Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Sci. Rep.*, 6, 38881
  70. The Lancet. (2019) ICD-11. *Lancet*, 393, 2275
  71. Wang, Y., Zhang, S., Li, F., Zhou, Y., Zhang, Y., Wang, Z., Zhang, R., Zhu, J., Ren, Y., Tan, Y., *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, 48, D1031–D1041
  72. Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G., *et al.* (2018) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.*, 46, D1121–D1127
  73. Otasek, D., Morris, J. H., Bouças, J., Pico, A. R. and Demchak, B. (2019) Cytoscape Automation: empowering workflow-based network analysis. *Genome Biol.*, 20, 185
  74. Ravi, A., Koster, J., Dijkhuis, A., Bal, S. M., Sabogal Piñeros, Y. S., Bonta, P. I., Majoor, C. J., Sterk, P. J. and Lutter, R. (2019) Interferon-induced epithelial response to rhinovirus 16 in asthma relates to inflammation and FEV<sub>1</sub>. *J. Allergy Clin. Immunol.*, 143, 442–447.e10
  75. Pai, S., Li, P., Killinger, B., Marshall, L., Jia, P., Liao, J., Petronis, A., Szabó, P. E. and Labrie, V. (2019) Differential methylation of enhancer at IGF2 is associated with abnormal dopamine synthesis in major psychosis. *Nat. Commun.*, 10, 2046
  76. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, 41, D991–D995
  77. Herring, B. P., Chen, M., Mihaylov, P., Hoggatt, A. M., Gupta, A., Nakeeb, A., Choi, J. N. and Wo, J. M. (2019) Transcriptome profiling reveals significant changes in the gastric muscularis externa with obesity that partially overlap those that occur with idiopathic gastroparesis. *BMC Med. Genomics*, 12, 89
  78. Speake, C., Skinner, S. O., Berel, D., Whalen, E., Dufort, M. J., Young, W. C., Odegard, J. M., Pesenacker, A. M., Gorus, F. K., James, E. A., *et al.* (2019) A composite immune signature

- parallels disease progression across T1D subjects. *JCI Insight*, 4, e126917
79. Hutter, C. and Zenklusen, J. C. (2018) The Cancer Genome Atlas: creating lasting value beyond its data. *Cell*, 173, 283–285
  80. Hu, J., Xu, J., Pang, L., Zhao, H., Li, F., Deng, Y., Liu, L., Lan, Y., Zhang, X., Zhao, T., *et al.* (2016) Systematically characterizing dysfunctional long intergenic non-coding RNAs in multiple brain regions of major psychosis. *Oncotarget*, 7, 71087–71098
  81. Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M. and Barabási, A. L. (2007) The human disease network. *Proc. Natl. Acad. Sci. USA*, 104, 8685–8690
  82. Ko, Y., Cho, M., Lee, J. S. and Kim, J. (2016) Identification of disease comorbidity through hidden molecular mechanisms. *Sci. Rep.*, 6, 39433
  83. Leng, L., Zhang, C., Ren, L. and Li, Q. (2019) Construction of a long noncoding RNA-mediated competitive endogenous RNA network reveals global patterns and regulatory markers in gestational diabetes. *Int J Mol Med*, 43, 927–935
  84. Brent, G. A. (2012) Mechanisms of thyroid hormone action. *J. Clin. Invest.*, 122, 3035–3043
  85. Pearce, E. N. (2012) Thyroid hormone and obesity. *Curr. Opin. Endocrinol. Diabetes Obes.*, 19, 408–413
  86. Biondi, B. (2010) Thyroid and obesity: an intriguing relationship. *J. Clin. Endocrinol. Metab.*, 95, 3614–3617
  87. Sinha, R. A., Singh, B. K. and Yen, P. M. (2014) Thyroid hormone regulation of hepatic lipid and carbohydrate metabolism. *Trends Endocrinol. Metab.*, 25, 538–545
  88. Biondi, B., Kahaly, G. J. and Robertson, R. P. (2019) Thyroid dysfunction and diabetes mellitus: two closely associated disorders. *Endocr. Rev.*, 40, 789–824
  89. Broholm, C., Olsson, A. H., Perfilyev, A., Hansen, N. S., Schrölkamp, M., Strasko, K. S., Scheele, C., Ribøl-Madsen, R., Mortensen, B., Jørgensen, S. W., *et al.* (2016) Epigenetic programming of adipose-derived stem cells in low birthweight individuals. *Diabetologia*, 59, 2664–2673
  90. Forstner, A. J., Basmanav, F. B., Mattheisen, M., Böhmer, A. C., Hollegaard, M. V., Janson, E., Strengman, E., Priebe, L., Degenhardt, F., Hoffmann, P., *et al.* (2014) Investigation of the involvement of MIR185 and its target genes in the development of schizophrenia. *J. Psychiatry Neurosci.*, 39, 386–396
  91. Venkatasubramanian, G. (2015) Understanding schizophrenia as a disorder of consciousness: biological correlates and translational implications from quantum theory perspectives. *Clin. Psychopharmacol. Neurosci.*, 13, 36–47
  92. Swanger, S. A., Mattheyses, A. L., Gentry, E. G. and Herskowitz, J. H. (2016) ROCK1 and ROCK2 inhibition alters dendritic spine morphology in hippocampal neurons. *Cell. Logist.*, 5, e1133266
  93. Ross, K. A. (2011) Evidence for somatic gene conversion and deletion in bipolar disorder, Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis, type-1 diabetes, and type-2 diabetes. *BMC Med.*, 9, 12
  94. Li, M. J., Liu, Z., Wang, P., Wong, M. P., Nelson, M. R., Kocher, J. P., Yeager, M., Sham, P. C., Chanock, S. J., Xia, Z., *et al.* (2016) GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, 44, D869–D876
  95. Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, 47, D1005–D1012
  96. Zhong, C., Yang, Y. and Yooseph, S. (2019) GRASP2: fast and memory-efficient gene-centric assembly and homolog search for metagenomic sequencing data. *BMC Bioinformatics*, 20, 276
  97. Tang, J., Fu, J., Wang, Y., Luo, Y., Yang, Q., Li, B., Tu, G., Hong, J., Cui, X., Chen, Y., *et al.* (2019) Simultaneous improvement in the precision, accuracy, and robustness of label-free proteome quantification by optimizing data manipulation chains. *Mol. Cell. Proteomics*, 18, 1683–1699
  98. Yang, Q., Hong, J., Li, Y., Xue, W., Li, S., Yang, H. and Zhu, F. (2020) A novel bioinformatics approach to identify the consistently well-performing normalization strategy for current metabolomic studies. *Brief. Bioinform.*, 21, 2142–2152
  99. Ros, G., Pegoraro, S., De Angelis, P., Sgarra, R., Zucchelli, S., Gustincich, S. and Manfioletti, G. (2020) HMGA2 antisense long non-coding RNAs as new players in the regulation of HMGA2 expression and pancreatic cancer promotion. *Front. Oncol.*, 9, 1526
  100. Han, Z. J., Xue, W. W., Tao, L. and Zhu, F. (2018) Identification of novel immune-relevant drug target genes for Alzheimer's disease by combining ontology inference with network analysis. *CNS Neurosci. Ther.*, 24, 1253–1263
  101. Wang, M., Yuan, D., Tu, L., Gao, W., He, Y., Hu, H., Wang, P., Liu, N., Lindsey, K. and Zhang, X. (2015) Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol.*, 207, 1181–1197
  102. Cabili, M. N., Dunagin, M. C., McClanahan, P. D., Biaesch, A., Padovan-Merhar, O., Regev, A., Rinn, J. L. and Raj, A. (2015) Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.*, 16, 20
  103. Werner, M. S., Sullivan, M. A., Shah, R. N., Nadadur, R. D., Grzybowski, A. T., Galat, V., Moskowitz, I. P. and Ruthenburg, A. J. (2017) Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. *Nat. Struct. Mol. Biol.*, 24, 596–603
  104. Teimuri, S., Hosseini, A., Rezaenasab, A., Ghaedi, K., Ghoheud, E., Etemadifar, M., Nasr-Esfahani, M. H. and Megraw, T. L. (2018) Integrative analysis of lncRNAs in Th17 cell lineage to discover new potential biomarkers and therapeutic targets in autoimmune diseases. *Mol. Ther. Nucleic Acids*, 12, 393–404
  105. Li, S., Yu, X., Lei, N., Cheng, Z., Zhao, P., He, Y., Wang, W. and Peng, M. (2017) Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. *Sci. Rep.*, 7, 45981
  106. Wang, X., Yang, C., Guo, F., Zhang, Y., Ju, Z., Jiang, Q., Zhao, X., Liu, Y., Zhao, H., Wang, J., *et al.* (2019) Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein

- bulls with high and low sperm motility. *Sci. Rep.*, 9, 2092
107. Schultz, B. M., Gallicio, G. A., Cesaroni, M., Lupey, L. N. and Engel, N. (2015) Enhancers compete with a long non-coding RNA for regulation of the *Kcnq1* domain. *Nucleic Acids Res.*, 43, 745–759
  108. Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., *et al.* (2010) Long noncoding RNAs with enhancer-like function in human cells. *Cell*, 143, 46–58
  109. Pyfrom, S. C., Luo, H. and Payton, J. E. (2019) PLAIDOH: a novel method for functional prediction of long non-coding RNAs identifies cancer-specific LncRNA activities. *BMC Genomics*, 20, 137
  110. Khyzha, N., Khor, M., DiStefano, P. V., Wang, L., Matic, L., Hedin, U., Wilson, M. D., Maegdefessel, L. and Fish, J. E. (2019) Regulation of *CCL2* expression in human vascular endothelial cells by a neighboring divergently transcribed long noncoding RNA. *Proc. Natl. Acad. Sci. USA*, 116, 16410–16419
  111. The Gene Ontology Consortium. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, 47, D330–D338
  112. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, 45, D353–D361
  113. Yang, Q., Li, B., Tang, J., Cui, X., Wang, Y., Li, X., Hu, J., Chen, Y., Xue, W., Lou, Y., *et al.* (2020) Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief. Bioinform.*, 21, 1058–1068
  114. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, 102, 15545–15550
  115. Hong, J., Luo, Y., Zhang, Y., Ying, J., Xue, W., Xie, T., Tao, L. and Zhu, F. (2020) Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning. *Brief. Bioinform.*, 21, 1437–1447
  116. Love, M. I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15, 550
  117. Yang, Q., Wang, Y., Zhang, Y., Li, F., Xia, W., Zhou, Y., Qiu, Y., Li, H. and Zhu, F. (2020) NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res.*, 48, W436–W448
  118. Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., Chen, Y., Xue, W., Li, X. and Zhu, F. (2017) NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.*, 45, W162–W170
  119. Tang, J., Fu, J., Wang, Y., Li, B., Li, Y., Yang, Q., Cui, X., Hong, J., Li, X., Chen, Y., *et al.* (2020) ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief. Bioinform.*, 21, 621–636
  120. Clough, E. and Barrett, T. (2016) The gene expression omnibus database. *Methods Mol. Biol.*, 1418, 93–110
  121. Li, F., Zhou, Y., Zhang, X., Tang, J., Yang, Q., Zhang, Y., Luo, Y., Hu, J., Xue, W., Qiu, Y., *et al.* (2020) SSizer: determining the sample sufficiency for comparative biological study. *J. Mol. Biol.*, 432, 3411–3421