ncFANs v2.0: an integrative platform for functional annotation of non-coding RNAs

Yuwei Zhang [1,2,†], **Dechao Bu**[3,†], **Peipei Huo**[4], **Zhihao Wang**[4], **Hao Rong**[1,2], **Yanguo Li**[5], **Jingjia Liu**[6], **Meng Ye**[1], **Yang Wu**[3], **Zheng Jiang**[7], **Qi Liao**[1,2,*] and **Yi Zhao** [3,*]

[1]The Affiliated Hospital of Medical School of Ningbo University, Ningbo, Zhejiang, 315000, China, [2]Department of Preventative Medicine, Zhejiang Provincial Key Laboratory of Pathological and Physiological Technology, School of Medicine, Ningbo University, Ningbo, Zhejiang, 315000, China, [3]Pervasive Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China, [4]Luoyang Branch of Institute of Computing Technology, Chinese Academy of Sciences, Henan, 471000, China, [5]I nstitute of Drug Discovery Technology, Ningbo University, Ningbo, Zhejiang, 315000, China, [6]Ningbo Institute of Life and Health Industry, University of Chinese Academy of Sciences, Ningbo, Zhejiang, 315000, China and [7]Department of Colorectal Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100021, China

ABSTRACT

Increasing evidence proves the essential regulatory roles of non-coding RNAs (ncRNAs) in biological processes. However, characterizing the specific functions of ncRNAs remains a challenging task, owing to the intensive consumption of the experimental approaches. Here, we present an online platform ncFANs v2.0 that is a significantly enhanced version of our previous ncFANs to provide multiple computational methods for ncRNA functional annotation. Specifically, ncFANs v2.0 was updated to embed three functional modules, including ncFANs-NET, ncFANs-eLnc and ncFANs-CHIP. ncFANs-NET is a new module designed for data-free functional annotation based on four kinds of pre-built networks, including the co-expression network, co-methylation network, long non-coding RNA (lncRNA)-centric regulatory network and random forest-based network. ncFANs-eLnc enables the one-stop identification of enhancer-derived lncRNAs from the *de novo* assembled transcriptome based on the user-defined or our pre-annotated enhancers. Moreover, ncFANs-CHIP inherits the original functions for microarray data-based functional annotation and supports more chip types. We believe that our ncFANs v2.0 carries sufficient convenience and practicability for biological researchers and facilitates unraveling the regulatory mechanisms of ncRNAs. The ncFANs v2.0 server is freely available at http://bioinfo.org/ncfans or http://ncfans.gene.ac.

GRAPHICAL ABSTRACT
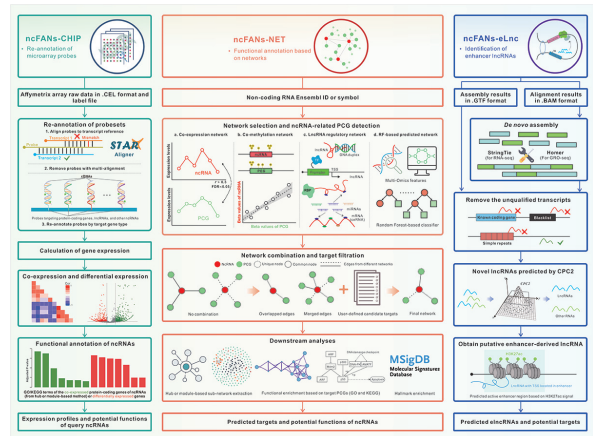
INTRODUCTION

The eukaryotic genome comprises a large portion of non-coding genes with limited protein-coding potentials. Over the past decade, non-coding RNAs (ncRNAs) such as long non-coding RNA (lncRNA), microRNA (miRNA), transfer RNA (tRNA) and small nucleolar RNA (snoRNA) have been widely verified to interact with protein-coding genes (PCGs) and participate in various biological processes through diversified molecular mechanisms (1–4). Nevertheless, compared to the relatively well-annotated functions of

PCGs, much less is known about the regulatory mechanisms of ncRNA. Moreover, for the rapid growth of the newly identified ncRNAs and the disproportionate supply of experimental validation, characterizing the roles of ncRNAs remains a long way ahead and requires assistance from computational approaches.

Based on high-throughput data and some intrinsic features of ncRNAs such as structure and sequence information, some bioinformatics methods have been designed for the functional annotation of ncRNAs. For example, nc-FANs (5) is the earliest online server for lncRNA functional annotation using microarray-based co-expression patterns between lncRNAs and PCGs. Besides, lncFunTK (6) integrates multi-Omics data and discovers the interactive genes of lncRNAs to evaluate the importance and predict the specific functions of lncRNAs. AnnoLnc2 (7) provides a scheme to identify new lncRNAs, annotate their functions based on lncRNA–PCG co-expression relationships and predict the subcellular localization. However, almost all tools only served lncRNAs and cannot be used for other kinds of ncRNAs. Furthermore, some of them have entered a low-support stage and even been inaccessible. Moreover, most of the currently available tools assigned the functions of the related PCGs to the query ncRNAs, but they rarely considered the condition-specific nature of the ncRNA-PCG relationships, especially the tissue and disease specificity. All these limitations urge the development of a more powerful functional annotator.

Therefore, we present <u>nc</u>RNA <u>F</u>unctional <u>AN</u>notation <u>S</u>erver (ncFANs) v2.0, the updated version of our previous ncFANs (5), which supports more kinds of ncRNAs, embeds more diversified annotation methods and provides a user-friendly interface. Compared to the old version, nc-FANs v2.0 comprises three functional modules (Figure 1), including the new ncFANs-NET for data-free functional annotation based on four kinds of ncRNA-PCG networks, ncFANs-eLnc for enhancer-derived lncRNA (el-ncRNA) identification and the enhanced ncFANs-CHIP for microarray data-based functional annotation. Our nc-FANs v2.0, thus, carries sufficient practicability for biological researchers. This web server is freely available at http://bioinfo.org/ncfans or http://ncfans.gene.ac/ and opens to all users, and there is no login requirement.
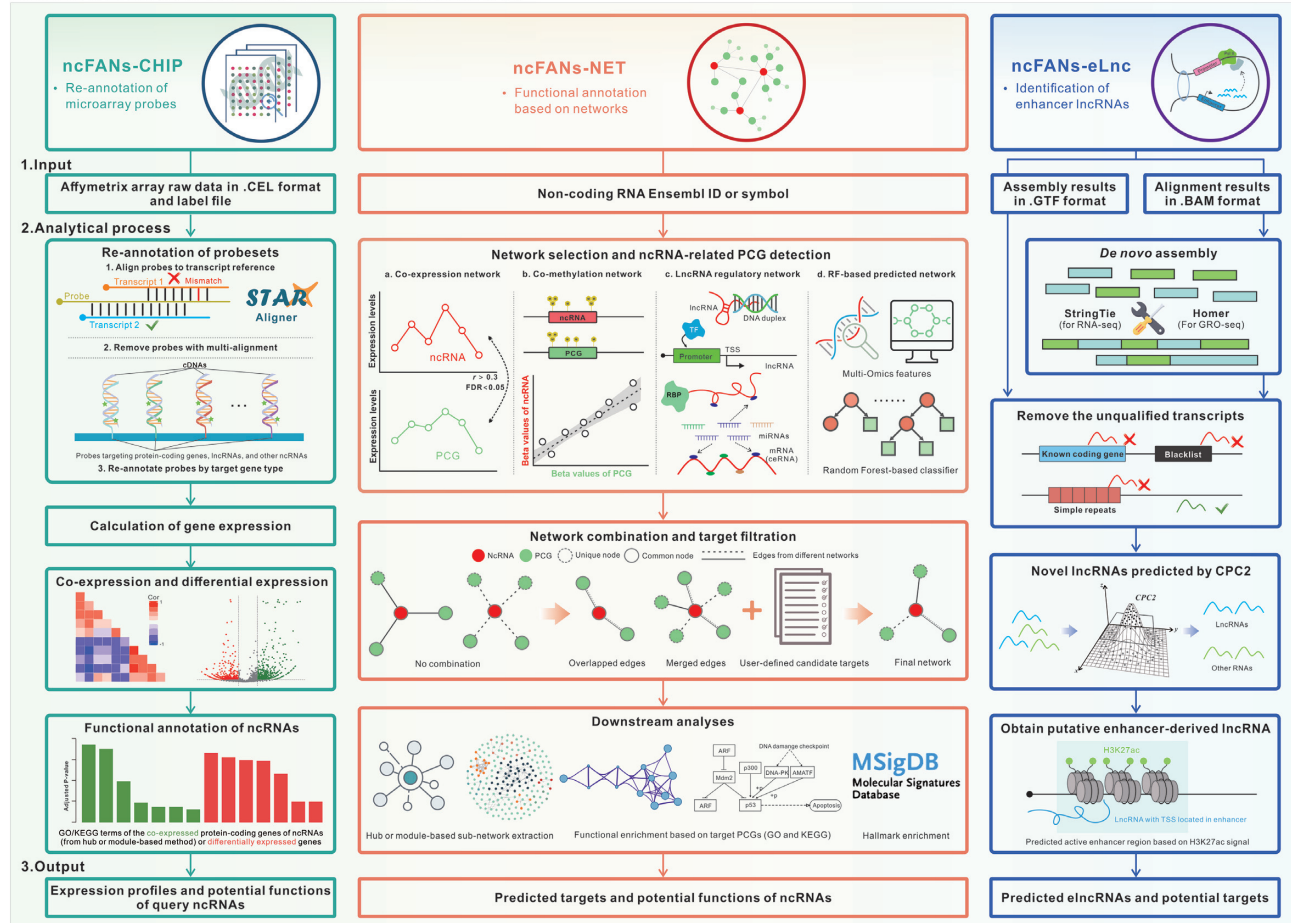
RESULTS

ncFANs-NET for pre-built network-based functional annotation

ncFANs-NET is a new module designed for the data-free functional annotation of human ncRNAs based on four kinds of pre-built ncRNA-PCG networks (Figure 2), including the normal tissue- and cancer-specific co-expression network, co-methylation network, general lncRNA-centric regulatory network, and random forest-based lncRNA-PCG interactive network. Users only need to input the query ncRNAs and set the parameters of each chosen network. Notably, the co-expression and co-methylation networks support five kinds of ncRNAs (see the table at http://ncfans.gene.ac/statistics/), while the other two networks are dedicated to lncRNAs. As our previous studies have pro-

posed these networks (5,8–12), the construction workflows will be briefly described as below:

1. Co-expression network. The strategy adopted to construct the co-expression network is similar to that we previously proposed (9). First, RNA-seq data of 54 normal tissues and 33 cancer types were downloaded from GTEx (13) and the TCGA portal (https://www.cancer.gov/tcga). The sample sizes and gene numbers are recorded at http://ncfans.gene.ac/statistics/. Next, the R package WGCNA (14) was used to calculate the Spearman correlation coefficients (Rho), Fisher's asymptotic *P*-value, and topological overlap measure (TOM) (15). A pair of genes with adjusted *P*-value (FDR correction) <0.05, Rho and TOM greater than user-defined cutoffs will be co-expressed. Apart from the condition-specific networks, users can also choose the overall normal or cancer network, which is defined as the pairs of genes co-expressed in a certain proportion of tissues or cancer types.

2. Co-methylation network. Recent studies proposed that gene pair's co-methylation patterns always indicated their similar molecular functions and thus can be used for ncRNA functional annotation (8,16–18). Our strategy for co-methylation network construction is shown in the following. For normal tissues, a total of 40 human methylation profiles based on Bisulfite-seq were collected from the MethBank database (19) and classified by methylation types and genomic regions. For each gene, the methylation level in the specific condition of methylation type and the genomic region was calculated. For cancers, Illumina 450K array data were downloaded from the TCGA database (https://www.cancer.gov/tcga) of which the probes targeting the promoter regions (±2 kb around transcription start site) were reserved for further analyses. Next, the transcripts with at least three consistent probes in the promoter regions (the average Spearman correlation coefficients among the consistent probes should be >0.2) were considered to be valid and reliable records. Moreover, the average beta values of the probes targeting the same promoter were assigned as the gene methylation levels. If a gene has multiple detectable isoforms and alternative promoters, only the maximum methylation level was assigned to the gene. Finally, the correlations between the gene methylation levels were calculated using the R package WGCNA (14) as described in the above section.

3. LncRNA-centric regulatory network (LncRnet). LncRnet is composed of the lncRNA-PCG relationships detected by multi-Omics data and software, which is also proposed by our team (10,11,20). To construct the lncRNA-centric regulatory network, we investigated the lncRNA-PCG relationships in aspects of transcription factor (TF) binding activities, competing endogenous RNA (ceRNA) mechanism, binding activities of RNA-binding proteins (RBPs) and lncRNA–DNA triplexes. The specific workflows of obtaining these relationships are similar to that used in the published study (10). Briefly, for lncRNA–TF interaction, a total of 11 356 ChIP-seq datasets involved in 1354 TFs were obtained from the Cistrome database (21). The ncRNAs with TF

Figure 1. The workflows of three functional modules of ncFANs v2.0. ncFANs v2.0 is composed of ncFANs-CHIP for microarray data-based functional annotation, ncFANs-NET for data-free functional annotation based on pre-built networks and ncFANs-eLnc for identification of enhancer-derived lncR-NAs.

peaks located in the promoter regions were considered to be regulated. For the ceRNA mechanism, 2656 mature human miRNAs sequences were downloaded from the miRBase database (22). Next, the lncRNA–miRNA and mRNA–miRNA interactions predicted by miRanda (23) with default parameters were reserved to investigate the competing possibility and strength between lncR-NAs and mRNAs, which were assessed by the hypergeometric test and MuTaME scores as previously described (10). For RBP-binding activities, both experimental and computational methods are supported. In experimental method, the lncRNA–RBP interactions identified by CLIP-seq data were collected from POSTAR2 database (24), in which 216 RBPs and 12 963 lncRNAs are involved. In computational method, the sequences of 1803 RBPs annotated in GO:0003723 were obtained from the Uniprot database (25), of which the binding possibility to lncRNAs was predicted by lncPro (26) with default parameter. For lncRNA–DNA triplex, we provided the information predicted by Triplexator (27), which merely serves as a reference for users and is not incorporated into the network construction due to the limited reliability. Finally, all obtained lncRNA–PCG relationships

were integrated with the protein–protein interactions (PPIs) from the STRING database (28) to construct the overall interaction network.

4. Random forest-based network (RF-based network). In this part, the lncRNA–PCG relationships were obtained based on our recently published method (12). According to the user-defined cutoff, ncFANs-NET will extract the qualified lncRNA–PCG interactions and corresponding PPIs from the STRING database (28) to construct the overall RF-based network.

After obtaining the relationships between ncRNAs and PCGs, ncFANs-NET will annotate the functions of ncR-NAs using hub- and module-based methods described by Liao *et al.* (9). In the hub-based methods, ncFANs-NET selected the directly connected PCGs of the hub ncRNAs in the network and assigned the enriched functions of these PCGs to ncRNAs. The functional enrichment analyses, including gene ontology (GO), KEGG pathway and MSigDB hallmark enrichment (29) were performed using our self-made scripts (see https://github.com/zhangyw0713/FunctionEnrichment). In the module-based methods, SPICi (30) in the unweighted mode and with de-
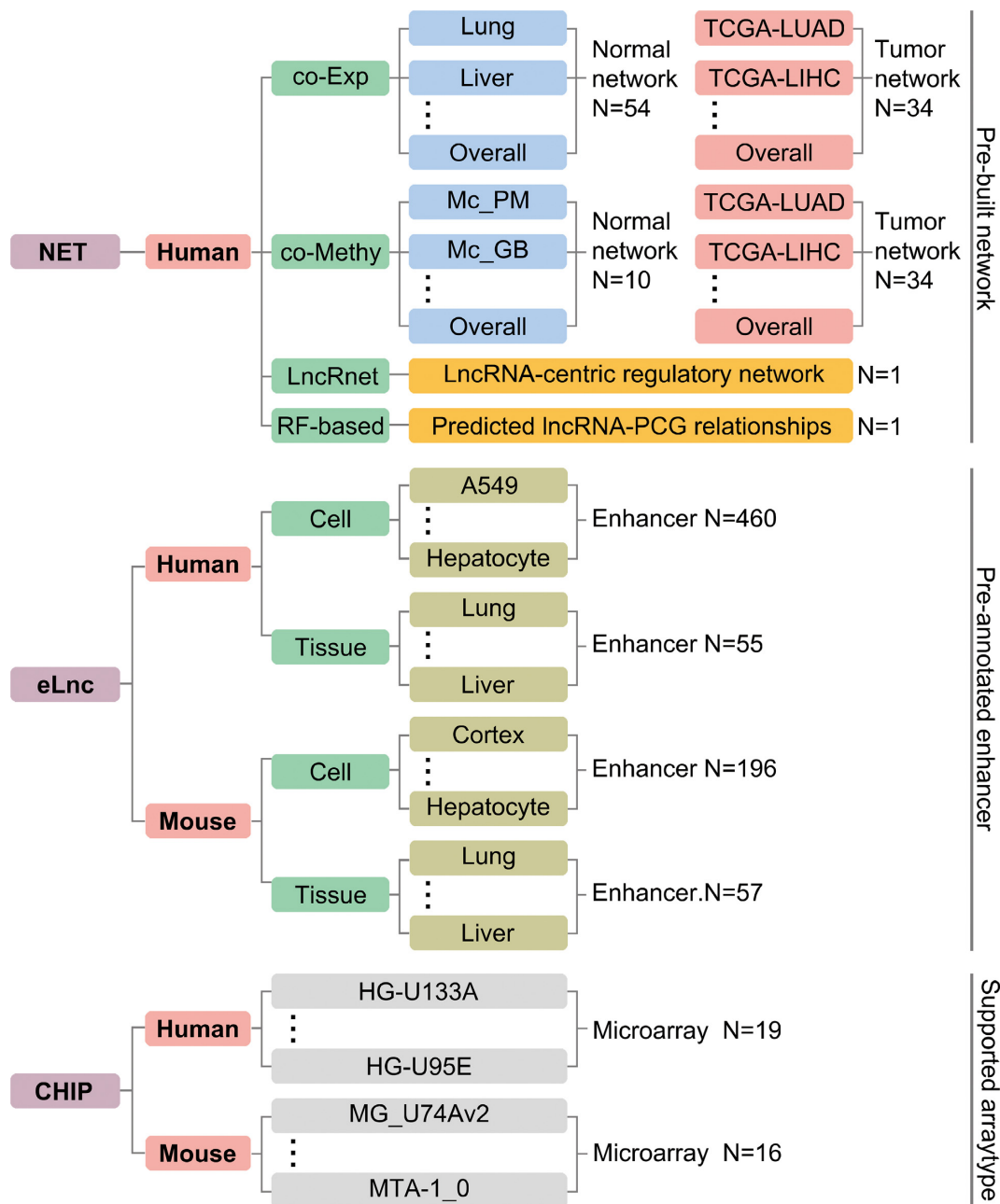
Figure 2. Data architecture of ncFANs v2.0, including the pre-built networks of ncFANs-NET, the pre-annotated enhancers of ncFANs-eLnc and the supported array type of ncFANs-CHIP. The full lists of these data can be fetched at http://ncfans.gene.ac/statistics/.

fault parameter replaced the previous Markov cluster algorithm (MCL) to extract the close-connected modules from the overall network, owing to the overwhelming performance and speed. According to the assumption that the genes with similar functions tend to be concentrically distributed, the functions of ncRNAs can be represented by the enriched functions of PCGs in the same module.

To explicitly demonstrate the usage of ncFANs-NET, we performed a case study based on three differentially expressed ncRNAs (lncRNA *PVT1*, lncRNA *LINC00265* and snoRNA *AL356356.1*) in human col-

orectal cancer (CRC). First, 3 query ncRNAs and 4234 differentially expressed PCGs in CRC which were identified by our previous study (11) and served as candidate targets were input (Figure 3A). Next, four kinds of networks were chosen (Figure 3A) of which the co-expression and co-methylation networks can be cancer-specific and thus set as TCGA-COAD. The other parameters are shown in Figure 3B. After receiving the launch instruction from the 'Run' button, ncFANs-NET started the network-based functional annotation and returned three parts of results. In the first part (Figure 4A), ncFANs-NET provided the basic information of the query ncRNAs especially the
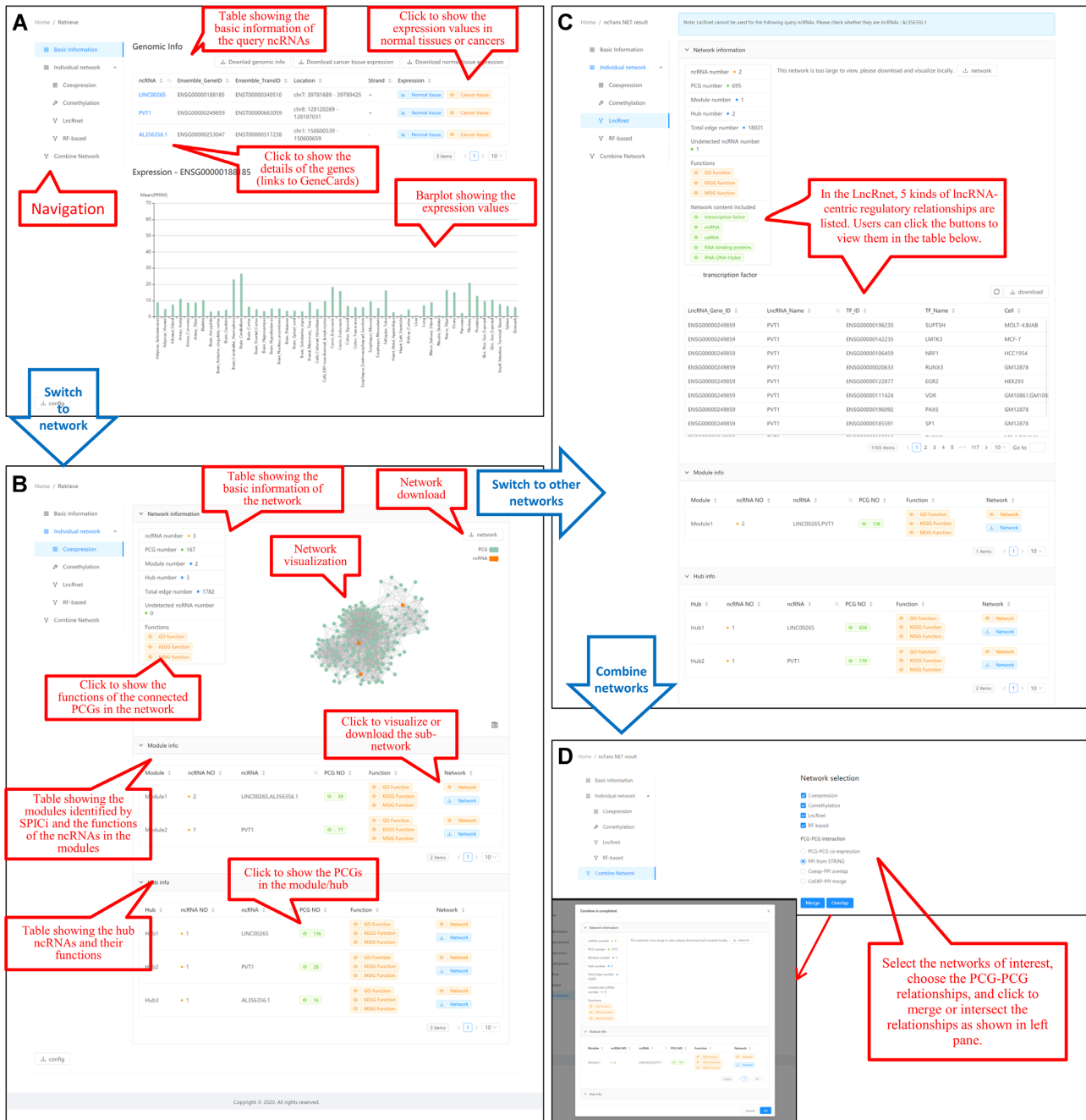
Figure 3. Input interface of ncFANs-NET. Users need to (**A**) input the query ncRNAs and optional target PCGs set and choose the background networks. The parameters of the networks are shown in (**B**).

expression levels indicating their biological importance in the tissues or cancers. Next, the functional annotation results based on the selected networks were displayed in the second part (Figure 4B). For example, in the result based on the co-expression network, query ncRNAs-related network was extracted and visualized in a force-directed layout as shown in Figure 4B. Moreover, 2 modules and 3 ncRNA hubs were respectively identified. As we can see, lncRNA *LINC00265* and snoRNA *AL356356.1* were simultaneously clustered in module 1, indicating their similar regulatory functions in molecular biosynthesis in CRC. Notably, the only significantly enriched hallmark 'mitotic splendid' implied their potential role in mitosis and cell division.

Moreover, the results based on other networks can be viewed via the left navigation bar (Figure 4C), which provides more clues for the complex functions of ncRNAs. Also, the ncRNA–PCG relationships predicted by different kinds of networks can be optionally merged or intersected. The PCG–PCG relationships can also be re-defined to obtain a more comprehensive or reliable result (Figure 4D).

ncFANs-eLnc for the identification of elncRNAs

ncFANs-eLnc provides identification of elncRNAs, which possibly exert enhancer-like function. In this module, users need to upload either the *de novo* assembled transcriptome in GTF format (highly recommended) or the alignment results of RNA-seq or GRO-seq data in BAM format. For alignment results, owing to storage and network issues, we set a threshold on the BAM file size (<200 mb), which should be suitable for the elncRNA identification on a certain chromosome (e.g. chr1) or some segments of interests. If users want to analyze the whole transcriptome data, they are encouraged to perform *de novo* assembly locally by using our strategy (see the commands in FAQ page at http://ncfans.gene.ac/help/) and then upload the GTF file instead. Next, users either provide the enhancer regions or select the pre-annotated cell- and tissue-specific enhancers (Figure 2). Our strategy adopted to annotate the enhancers is modified from that previously proposed (31). First, H3K27ac modification ChIP-seq data were collected from the Cistrome database (21). Peaks entirely located

Figure 4. Output interface of ncFANs-NET. (**A**) Users are first provided with the basic information of the query ncRNAs. (**B** and **C**) And they can switch to view the annotated functions of the ncRNAs based on the selected networks. (**D**) Finally, the ncRNA-PCG relationships obtained from different networks can be intersected or merged to re-annotate the functions.

within blacklisted regions or promoters which were defined as the ±2 kb regions around transcription start sites were filtered out. Considering that the ChIP-seq peaks could be narrower than the elncRNA transcription region, we selected the ±3 kb regions around the center of these tentative enhancers as potential elncRNA-transcribing enhancers. After receiving the defined transcriptome and enhancers, ncFANs-eLnc first removes the transcripts overlapped with the known coding genes, blacklisted regions and simple repeats and then employ CPC2 (32) with default parame-

ters to identify the novel ncRNAs. Novel lncRNAs with transcription start site located in enhancers are defined as elncRNAs.

In the output interface of ncFANs-eLnc, the details of the results, including the chromatin coordinates of the elncRNAs, the enhancer locations and the putative targets, are provided (Figure 5). The genes adjacent to the elncRNAs (±1 Mb) are defined as the putative targets and the known lncRNAs overlapped with the novel elncRNAs are also extracted, which can be viewed and

Figure 5. Output interface of ncFANs-eLnc. A table showing the information of the identified elncRNAs and enhancers were provided.

downloaded by clicking on the blue buttons in the penultimate and last columns. Moreover, if the enhancer regions uploaded by users are overlapped with the promoters of the annotated lncRNAs, ncFANs-eLnc will also return the information of these lncRNAs which can be accessed by clicking on the button shown in Figure 5. Besides, the genome browser based on JBrowse (33) is provided for users to directly view the location of elncRNAs and enhancers.

ncFANs-CHIP for microarray data-based functional annotation

ncFANs-CHIP retains the original function of our previous ncFANs (5), which predicts the functions of ncRNAs based on the re-annotated microarray data. The workflow is shown in Figure 1. Specifically, we first collected the description files in CDF format of 19 human and 16 mouse Affymetrix microarrays (Figure 2, see the full list at http://ncfans.gene.ac/statistics/). The probes of these arrays were re-annotated to PCGs and 6 kinds of ncRNAs (see the table at http://ncfans.gene.ac/statistics/) using the modified pipeline (5,9). The modification is that we used the new transcriptome reference sequences from GENCODE mouse vm24 and human v33 (34), and the mapper STAR (35) to replace the previous BLASTn (36). Of note, the tRNA reference sequences were collected from GtRNAdb 2.0 (37). Next, when users upload their raw microarray data in CEL format and define the array type, the gene expression values will be calculated based on the corresponding re-annotated probes. The differential expression analysis and the local coding-noncoding co-expression network construction with multiple flexible parameters such as fold change, *P*-value, correlation coefficient and TOM are optional for users (see the input interface at http://ncfans.gene.ac/ncFansChip/). If the co-expression network construction is chosen, the functional annotation will automatically launch and the functions of ncRNAs will be predicted us-

ing the hub- and module-based methods as described in ncFANs-NET.

After a task is submitted, the results will be returned in several minutes, or users can retrieve the results on the homepage using the allocated task ID. Here, we provided a case study based on the Affymetrix HG-U133A_2 array data GSE16919 in human embryonic stem cells (hESC) to exemplify the output. The parameters are default. As shown in Figure 6A, ncFANs-CHIP returned the gene composition of the expression profile calculated from the re-annotated probes. Besides, the results of the differential expression analysis and co-expression network results were also returned as we chose the options. For the differential expression analysis (Figure 6B), the detailed information of 275 up-regulated and 196 down-regulated genes in the nicotinamide-treated hESCs, such as fold change and adjusted *P*-values, was displayed along with the buttons for downloading the full lists. Furthermore, the bar plots showing the enriched functions of the up- or down-regulated genes in the nicotinamide-induced hESC differentiation alteration can be viewed and saved via clicking on the yellow buttons. For the co-expression network construction in hESCs (Figure 6C), ncFANs-CHIP provided the basic information of the network, such as the included ncRNA and PCG numbers, the hubs and the modules with at least one ncRNA, and the enriched functions of the PCGs. In this study, as the network was too large to view (nodes >1000 or edges >10000), users were encouraged to download the network in Cytoscape (38) file format to visualize it locally. Moreover, the information of the modules and hubs extracted from the overall co-expression network and used for ncRNA functional annotation follow (Figure 6C). In these two parts, the ncRNAs, the directly connected PCGs in the hub/module, and their potential functions are listed. For example, in this study, we detected the lncRNA *DLEU2L*, *MIR22HG*, *AL450384.2* were clustered together in module 1, and the connected PCGs were enriched in the regulation of histone methylation, which in-
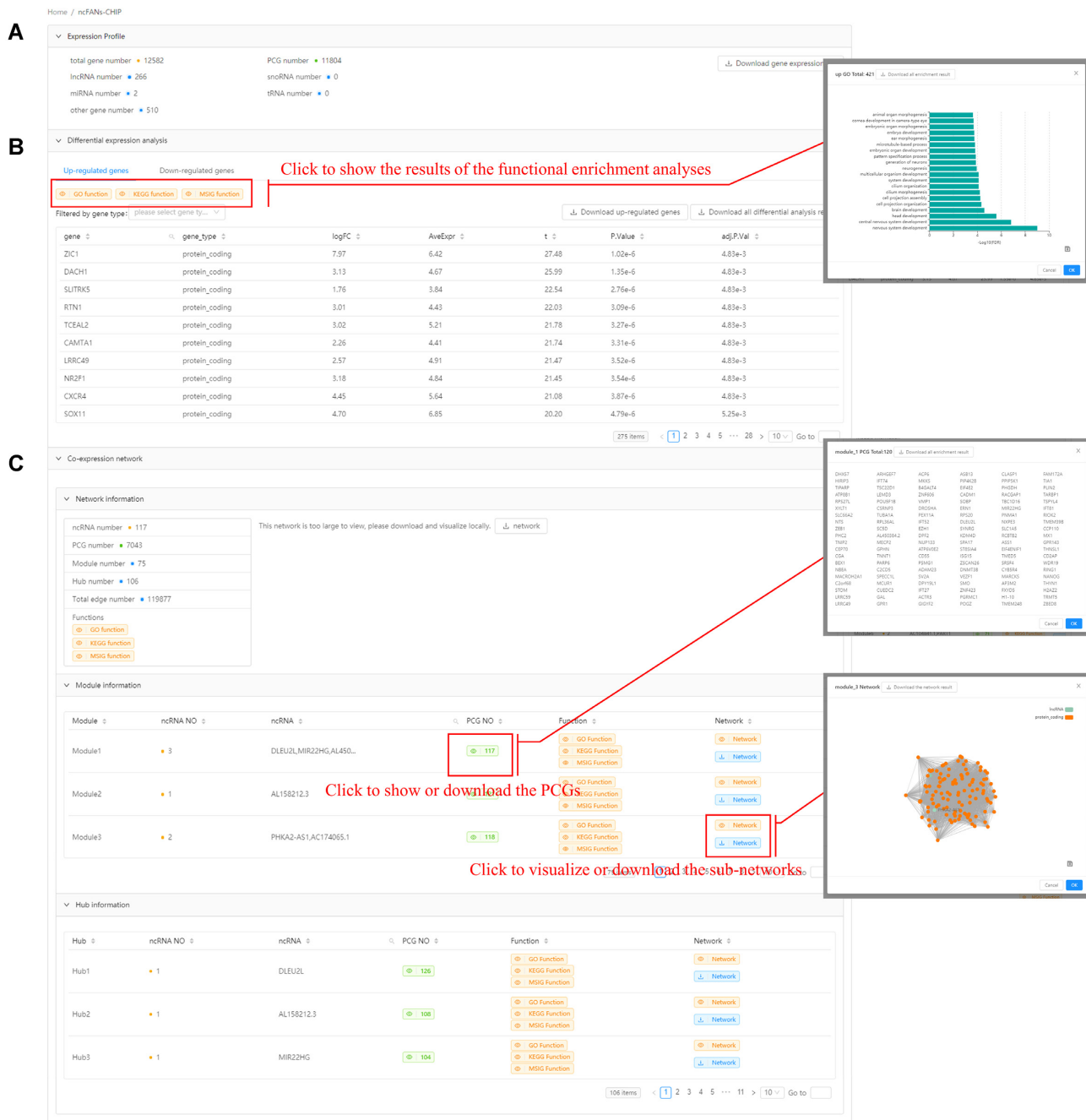
Figure 6. Output interface of ncFANs-CHIP. ncFANs-CHIP returns (**A**) the number of ncRNAs and PCGs of re-annotated microarray data, (**B**) the information of the differentially expressed genes and (**C**) the results of functional annotation based on the co-expression network.

dicated the possibly similar functions of these three lncR-NAs in hESC differentiation through influencing histone methylation.

DISCUSSION

As the successor of ncFANs, ncFANs v2.0 inherits the original features of microarray-based functional annotation and is invested with prominent enhancement in the diversity of annotation methods and applicability for more kinds of ncRNAs, along with a user-friendly and interactive inter-

face. The main improvements include but are not limited to the below items:

1. ncFANs v2.0 embeds a new module, ncFANs-NET, to provide data-free functional annotation, in which only ncRNAs and the background networks should be defined.
2. ncFANs-NET provides multiple kinds of annotation methods based on the pre-built networks, including the condition-specific co-expression and co-methylation networks and the versatile LncRnet and RF-based net-

works. Significantly, the condition-specific networks enable more accurate identification of the connected PCGs and functional annotation.

3. ncFANs-eLnc is the first one-stop platform for elncRNA identification.

4. Compared to the previous version, ncFANs-CHIP supports more types of microarray data and more kinds of ncRNAs including snoRNAs, rRNAs, miRNAs and so on (see the table at http://ncfans.gene.ac/statistics/).

Despite of the enhanced functions of our new ncFANs v2.0, there are still some shortcomings. In the future, we will put continuous effort into improving it in the following aspects. First, more species will be supported in all three modules to broaden the application fields. Moreover, the module for local co-expression network construction and functional annotation based on user-defined sequencing data will be developed. Next, in ncFANs-NET, we will keep updating the disease-specific networks, which should not be confined to cancer but include more diseases with sufficient data, such as Alzheimer's disease and HIV infection. In ncFANs-eLnc, the sets of the pre-annotated enhancer regions will be expanded to support more kinds of tissues and cell lines. It is expected that our ncFANs will benefit from our efforts and users' feedback, thus becoming increasingly powerful and popular.

DATA AVAILABILITY

The ncFANs v2.0 server is publicly available at http://bioinfo.org/ncfans or http://ncfans.gene.ac.

FUNDING

REFERENCES

1. Fabbri,M., Girnita,L., Varani,G. and Calin,G.A. (2019) Decrypting noncoding RNA interactions, structures, and functional networks. *Genome Res.*, **29**, 1377–1388.
2. Singh,I., Contreras,A., Cordero,J., Rubio,K., Dobersch,S., Gunther,S., Jeratsch,S., Mehta,A., Kruger,M., Graumann,J. *et al.* (2018) MiCEE is a ncRNA-protein complex that mediates epigenetic silencing and nucleolar organization. *Nat. Genet.*, **50**, 990–1001.
3. Slack,F.J. and Chinnaiyan,A.M. (2019) The role of non-coding RNAs in oncology. *Cell*, **179**, 1033–1055.
4. Teng,X., Chen,X., Xue,H., Tang,Y., Zhang,P., Kang,Q., Hao,Y., Chen,R., Zhao,Y. and He,S. (2020) NPInter v4.0: an integrated database of ncRNA interactions. *Nucleic Acids Res.*, **48**, D160–D165.
5. Liao,Q., Xiao,H., Bu,D., Xie,C., Miao,R., Luo,H., Zhao,G., Yu,K., Zhao,H., Skogerbo,G. *et al.* (2011) ncFANs: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res.*, **39**, W118–W124.
6. Zhou,J., Huang,Y., Ding,Y., Yuan,J., Wang,H. and Sun,H. (2018) lncFunTK: a toolkit for functional annotation of long noncoding RNAs. *Bioinformatics*, **34**, 3415–3416.
7. 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.
8. Liao,Q., He,W., Liu,J., Cen,Y., Luo,L., Yu,C., Li,Y., Chen,S. and Duan,S. (2015) Identification and functional annotation of lncRNA genes with hypermethylation in colorectal cancer. *Gene*, **572**, 259–265.
9. Liao,Q., Liu,C., Yuan,X., Kang,S., Miao,R., Xiao,H., Zhao,G., Luo,H., Bu,D., Zhao,H. *et al.* (2011) Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.*, **39**, 3864–3878.
10. Zhang,Y., Tao,Y., Ji,H., Li,W., Guo,X., Ng,D.M., Haleem,M., Xi,Y., Dong,C., Zhao,J. *et al.* (2019) Genome-wide identification of the essential protein-coding genes and long non-coding RNAs for human pan-cancer. *Bioinformatics*, **35**, 4344–4349.
11. Zhang,Y., Tao,Y., Li,Y., Zhao,J., Zhang,L., Zhang,X., Dong,C., Xie,Y., Dai,X., Zhang,X. *et al.* (2018) The regulatory network analysis of long noncoding RNAs in human colorectal cancer. *Funct. Integr. Genomics*, **18**, 261–275.
12. Zhang,Y., Yi,T., Ji,H., Zhao,G., Xi,Y., Dong,C., Zhang,L., Zhang,X., Zhao,J. and Liao,Q. (2020) Designing a general method for predicting the regulatory relationships between long noncoding RNAs and protein-coding genes based on multi-omics characteristics. *Bioinformatics*, **36**, 2025–2032.
13. Consortium,G.T. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
14. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.*, **9**, 559.
15. Ravasz,E., Somera,A.L., Mongru,D.A., Oltvai,Z.N. and Barabasi,A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–1555.
16. Akulenko,R. and Helms,V. (2013) DNA co-methylation analysis suggests novel functional associations between gene pairs in breast cancer samples. *Hum. Mol. Genet.*, **22**, 3016–3022.
17. Ma,X., Yu,L., Wang,P. and Yang,X. (2017) Discovering DNA methylation patterns for long non-coding RNAs associated with cancer subtypes. *Comput. Biol. Chem.*, **69**, 164–170.
18. Wei,Y., Dong,S., Zhu,Y., Zhao,Y., Wu,C., Zhu,Y., Li,K. and Xu,Y. (2019) DNA co-methylation analysis of lincRNAs across nine cancer types reveals novel potential epigenetic biomarkers in cancer. *Epigenomics*, **11**, 1177–1190.
19. Li,R., Liang,F., Li,M., Zou,D., Sun,S., Zhao,Y., Zhao,W., Bao,Y., Xiao,J. and Zhang,Z. (2018) MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Res.*, **46**, D288–D295.
20. Zhang,Y., Tao,Y. and Liao,Q. (2018) Long noncoding RNA: a crosslink in biological regulatory network. *Brief. Bioinform.*, **19**, 930–945.
21. Zheng,R., Wan,C., Mei,S., Qin,Q., Wu,Q., Sun,H., Chen,C.H., Brown,M., Zhang,X., Meyer,C.A. *et al.* (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.*, **47**, D729–D735.
22. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
23. Betel,D., Koppal,A., Agius,P., Sander,C. and Leslie,C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.
24. Zhu,Y., Xu,G., Yang,Y.T., Xu,Z., Chen,X., Shi,B., Xie,D., Lu,Z.J. and Wang,P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
25. UniProt,C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
26. Mann,C.M., Muppirala,U.K. and Dobbs,D. (2017) Computational prediction of RNA-protein interactions. *Methods Mol. Biol.*, **1543**, 169–185.
27. Buske,F.A., Bauer,D.C., Mattick,J.S. and Bailey,T.L. (2012) Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.*, **22**, 1372–1381.
28. Szklarczyk,D., Gable,A.L., Lyon,D., Junge,A., Wyder,S., Huerta-Cepas,J., Simonovic,M., Doncheva,N.T., Morris,J.H., Bork,P. *et al.* (2019) STRING v11: protein-protein association networks with

increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, D607–D613.

29. Liberzon,A., Birger,C., Thorvaldsdottir,H., Ghandi,M., Mesirov,J.P. and Tamayo,P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.

30. Jiang,P. and Singh,M. (2010) SPICi: a fast clustering algorithm for large biological networks. *Bioinformatics*, **26**, 1105–1111.

31. Zhang,Z., Lee,J.H., Ruan,H., Ye,Y., Krakowiak,J., Hu,Q., Xiang,Y., Gong,J., Zhou,B., Wang,L. *et al.* (2019) Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat. Commun.*, **10**, 4562.

32. Kang,Y.J., Yang,D.C., Kong,L., Hou,M., Meng,Y.Q., Wei,L. and Gao,G. (2017) CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.*, **45**, W12–W16.

33. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.

34. 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.

35. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

36. Boratyn,G.M., Thierry-Mieg,J., Thierry-Mieg,D., Busby,B. and Madden,T.L. (2019) Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinform.*, **20**, 405.

37. Chan,P.P. and Lowe,T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.*, **44**, D184–D189.

38. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.