

eRNA-IDO: A One-stop Platform for Identification, Interactome Discovery, and Functional Annotation of Enhancer RNAs

Yuwei Zhang ^{1,2,#}, Lihai Gong ^{3,#}, Ruofan Ding ³, Wenyan Chen ³, Hao Rong ⁴,
Yanguo Li ⁵, Fawziya Shameem ¹, Korakkandan Arshad Ali ¹, Lei Li ^{3,*}, Qi Liao ^{1,*}

¹School of Public Health, Health Science Center, Ningbo University, Ningbo 315211, China

²Biomedical Big Data Center, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China

³Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518107, China

⁴School of Clinical Medicine, Health Science Center, Ningbo University, Ningbo 315211, China

⁵Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China

*Corresponding authors: liaoqi@nbu.edu.cn (Liao Q), lei.li@szbl.ac.cn (Li L).

#Equal contribution.

Handling Editor: Leng Han

Abstract

Growing evidence supports the transcription of enhancer RNAs (eRNAs) and their important roles in gene regulation. However, their interactions with other biomolecules and their corresponding functionality remain poorly understood. In an attempt to facilitate mechanistic research, this study presents eRNA-IDO, the first integrative computational platform for the identification, interactome discovery, and functional annotation of human eRNAs. eRNA-IDO comprises two modules: eRNA-ID and eRNA-Anno. Functionally, eRNA-ID can identify eRNAs from *de novo* assembled transcriptomes. eRNA-ID includes eight kinds of enhancer makers, enabling users to customize enhancer regions flexibly and conveniently. In addition, eRNA-Anno provides cell/tissue-specific functional annotation for both new and known eRNAs by analyzing the eRNA interactome from prebuilt or user-defined networks between eRNAs and protein-coding genes. The prebuilt networks include the Genotype-Tissue Expression (GTEx)-based co-expression networks in normal tissues, The Cancer Genome Atlas (TCGA)-based co-expression networks in cancer tissues, and omics-based eRNA-centric regulatory networks. eRNA-IDO can facilitate research on the biogenesis and functions of eRNAs. The eRNA-IDO server is freely available at http://bioinfo.szbl.ac.cn/eRNA_IDO/.

Key words: Enhancer RNA; Identification; Interactome; Functional annotation; Web server.

Introduction

Over the past decade, a growing number of studies have reported the pervasive transcription of non-coding RNAs (ncRNAs) from active enhancer regions, termed enhancer RNAs (eRNAs). Due to the dynamic nature of enhancer activity across different tissues and lineages, eRNA transcription exhibits high specificity in biological contexts [1]. Once regarded as “transcription noise” or “byproduct” [2], eRNAs have now been shown to play crucial roles in various biological processes and diseases, such as cardiovascular development [3] and cancer [4]. Mechanistically, eRNAs can promote enhancer–promoter loops (E–P loops) and are involved in epigenetic regulation by interacting with other biomolecules, including components of cohesion or mediator [5,6], and histone acetyltransferases CBP/p300 [4,7]. Furthermore, eRNAs interact with transcription elongation factors to facilitate the pause-release of RNA polymerase II (RNAPII), thus controlling transcription elongation.

With the growing interest in eRNA functionality, several databases have been developed to characterize the transcription and potential targets of eRNAs, such as eRNABase [8], Human enhancer RNA Atlas (HeRA) [9], the Cancer eRNA Atlas (TCeA) [10], Animal-eRNAdb [11], and eRNA in cancer (eRic) [12]. Nonetheless, these databases only provide

information on annotated eRNA loci and enhancer regions but do not allow the evaluation of novel eRNAs. Additionally, several platforms exist for functional annotation of ncRNAs, but they are not well-suited for eRNAs. For example, ncRNA functional annotation server (ncFANs) v2.0 [13] requires known ncRNA identifiers as input, but most eRNAs lack a reference ID or symbol. AnnoLnc2 [14] allows the prediction of the functions of novel long non-coding RNAs (lncRNAs) based on co-expression networks but does not consider cell/tissue specificity and does not provide eRNA-specific characteristics such as histone modification, chromatin architecture, and interactive molecules. At present, a comprehensive platform for eRNA functional annotation is still lacking.

This study introduces eRNA-IDO, the first one-stop platform for human eRNA identification, interactome discovery, and functional annotation (Figure 1). eRNA-IDO comprises two available modules, namely eRNA-ID and eRNA-Anno. eRNA-ID enables users to define enhancers and identify enhancer-derived ncRNAs from an uploaded *de novo* assembled transcriptome. eRNA-Anno predicts eRNA functions by discovering eRNA-connected protein-coding genes (PCGs) in normal/cancer co-expression and eRNA-centric regulatory networks. Furthermore, eRNA-IDO offers the capacity to utilize prebuilt data as well as user-defined data, providing a practical and convenient tool for biological researchers. This

Received: 12 April 2023; Revised: 20 June 2024; Accepted: 5 August 2024.

© The Author(s) 2024. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.



eRNA-IDO

Identification & Functional Annotation



• eRNA-ID

- Identification of eRNAs



• eRNA-Anno

- Interactome discovery and functional annotation of eRNAs

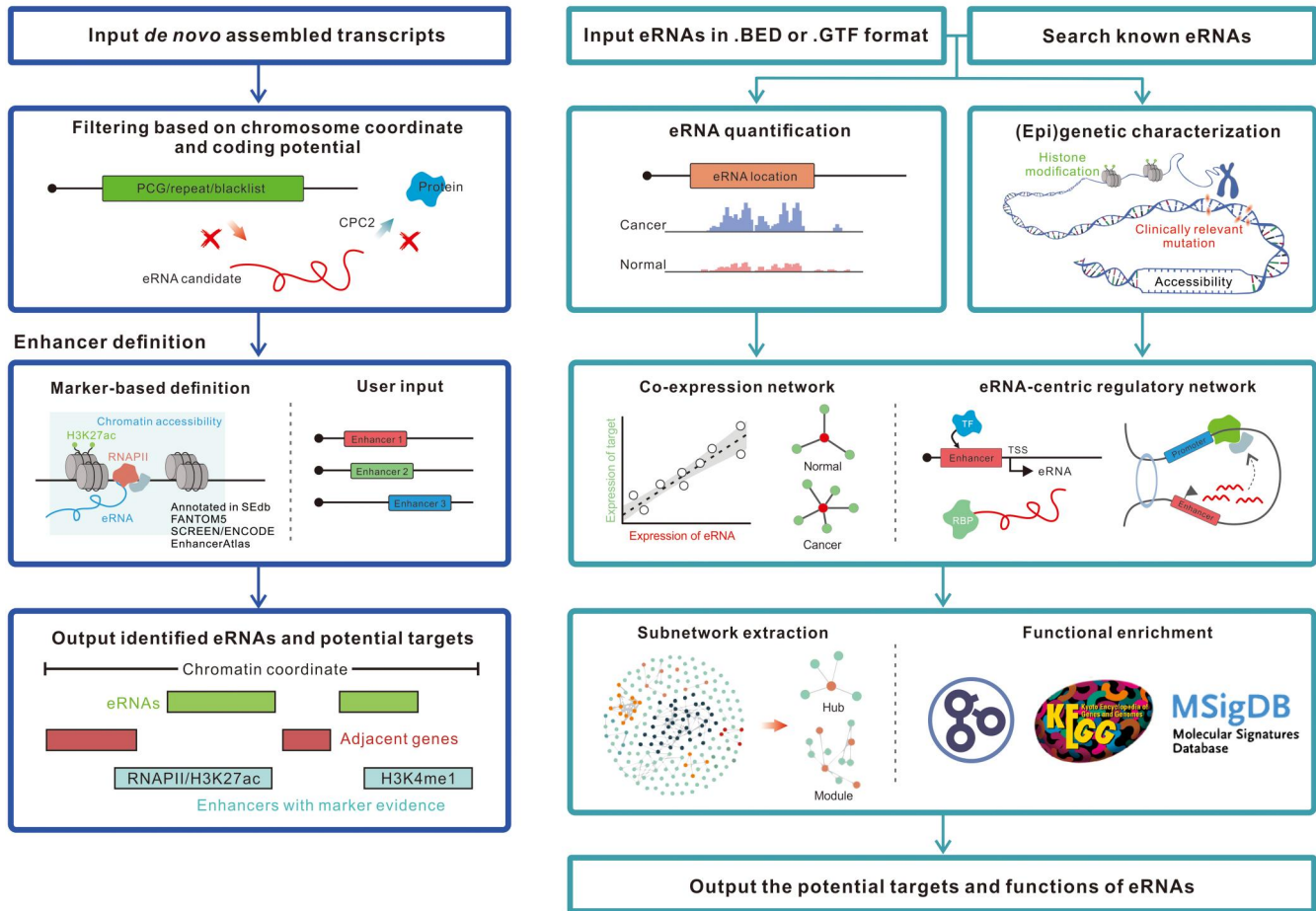


Figure 1 The workflow of eRNA-IDO

eRNA-IDO comprises two functional modules: eRNA-ID for eRNA identification and eRNA-Anno for interactome discovery and functional annotation. eRNA, enhancer RNA; PCG, protein-coding gene; SEdb, super-enhancer database; FANTOM5, functional annotation of the mammalian genome 5; ENCODE, ENcyclopedia of DNA Elements; SCREEN, Search Candidate *cis*-Regulatory Elements by ENCODE; RNAPII, RNA polymerase II; KEGG, Kyoto Encyclopedia of Genes and Genomes; MSigDB, Molecular Signatures Database; RBP, RNA-binding protein; TSS, transcription start site; TF, transcription factor; BED, browser extensible data; GTF, gene transfer format; CPC2, Coding Potential Calculator 2.

web server is freely available at http://bioinfo.szbl.ac.cn/eRNA_IDO/ and is open to all users without a login requirement.

Method

Workflow and data architecture of eRNA-ID

The schematic workflow of eRNA-ID is illustrated in Figure 1 (left panel). The processing of *de novo* assembled transcripts is initiated from user-provided RNA sequencing (RNA-seq) or global run-on sequencing (GRO-seq) data. The transcripts overlapping with annotated PCGs, simple repeats,

and blacklisted regions are removed based on the GENCODE v33 reference [15]. Thereafter, the coding potential of the remaining transcripts is evaluated by Coding Potential Calculator 2 (CPC2) [16] with default parameters, and ncRNAs transcribed from enhancer regions are identified as eRNAs. Enhancer regions can be either uploaded by users in Browser Extensible Data (BED) format or defined using our marker buffet. The marker buffet comprises eight kinds of enhancer markers, including H3K27ac (Table S1), H3K4me1 (Table S2), chromatin accessibility (Table S3), RANPII binding (Table S4), super-enhancers from super-enhancer database (SEdb) 2.0 [17], EnhancerAtlas 2.0 [18]

enhancers, functional annotation of the mammalian genome 5 (FANTOM5) [19] enhancers, and Search Candidate *cis*-Regulatory Elements by ENCYClopedia of DNA Elements (ENCODE) (SCREEN) [20] enhancers. The markers are optionally overlapped or merged (using BEDTools multiinter/merge) to obtain high-confidence or comprehensive enhancer profiles. The ± 3 kb regions around the center of the selected markers are defined as potential enhancer regions. These markers are cell-/tissue-specific, except those from FANTOM5 and SCREEN databases. The data type, source, and number of biosamples of these enhancer markers are listed in Table 1. Finally, eRNA-ID outputs the chromatin locations, adjacent genes (± 1 Mb), and enhancers of predicted eRNAs.

Workflow and data architecture of eRNA-Anno

The schematic workflow of eRNA-Anno is depicted in Figure 1 (right panel). The chromatin coordinates of novel eRNAs in BED/gene transfer format (GTF) format or the identifiers of known eRNAs annotated in HeRA [9] and eRic [12] databases are input in eRNA-Anno. For known eRNAs, the ENSR identifiers, chromatin coordinates, and adjacent genes (within ± 1 Mb) are accepted. Below is a detailed description of each procedure.

eRNA quantification

The expression levels of known eRNAs are obtained from HeRA and eRic. When chromatin coordinates of novel eRNAs are input, RNA-seq data from TCGA (<https://portal.gdc.cancer.gov/>) and GTEx portal [21] are used to quantify eRNA expression. Subsequently, eRNA expression levels are estimated based on the read coverage from BigWig files to shorten the processing time using the following formula:

$$FPKM = \frac{\sum(Cov) \times 10^9}{R \times L \times T}$$

where $\sum(Cov)$ represents the total read coverage of a given eRNA region, R is read length, L is eRNA length, and T indicates the total mapped reads of the library.

Profiling genetic/epigenetic landscape

eRNA-Anno portrays a genetic/epigenetic landscape for eRNAs, including chromatin accessibility, clinically relevant mutation, and histone modification (H3K27ac and H3K4me1). Histone modification and chromatin accessibility are characterized based on chromatin immunoprecipitation sequencing (ChIP-seq) data and assay for transposase-

accessible chromatin using sequencing (ATAC-seq)/DNase I hypersensitive site sequencing (DNase-seq) data, respectively, from the Cistrome Data Browser [22] (Table S1–S3). Finally, clinically relevant mutations within the query eRNA regions are collected from ClinVar [23] and the Catalogue Of Somatic Mutations In Cancer (COSMIC) [24] database.

eRNA-PCG network construction

Thereafter, a co-expression network between eRNAs and PCGs and an eRNA-centric regulatory network are constructed. The connected genes in the co-expression network are defined as the potential interactome of eRNAs. Both user-uploaded expression matrix and publicly available data are supported for the co-expression network. Publicly available data refer to RNA-seq data of 52 normal tissues from the GTEx portal [21] and 31 cancer types from the TCGA portal (Table S5). In addition, the toolkit GCEN [25] is used to calculate Spearman correlation coefficients and adjusted P values. The significant eRNA-PCG pairs are selected to construct the co-expression network according to user-defined thresholds.

For the eRNA-centric regulatory network, the relationships between eRNAs and transcription factors (TFs), RNA-binding proteins (RBPs), and E–P loops are analyzed. The eRNA–TF interactions are identified based on 11,356 ChIP-seq datasets from the Cistrome Data Browser [22], which involve 1354 TFs and 642 cells/tissues (Table S4). Furthermore, the eRNA–RBP interactions are obtained based on 518 cross-linking immunoprecipitation sequencing (CLIP-seq) datasets from the post-transcriptional regulation coordinated by RBP (POSTAR3) database [26], which involve 221 RBPs and 34 cells/tissues (Table S6). TFs and RBPs with peaks located within eRNA regions are defined as potential regulators of eRNAs. E–P loops identified by 198 HiChIP experiments across 108 cell types (Table S7) are collected from HiChIPdb [27]. The loops harboring anchors overlapping with query eRNAs are defined as eRNA-mediated loops.

Subnetwork extraction

Subsequently, eRNA-Anno extracts hubs/modules from the overall network to obtain tightly connected PCGs of query eRNAs. During this process, SPICi [28] in the unweighted mode (with default parameters) is utilized for module extraction.

Functional enrichment analyses

Functional enrichment analyses, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, and Molecular Signatures Database (MSigDB)

Table 1 Data type, source, and number of biosamples of enhancer markers

Marker	Data type	Data source	No. of biosamples	Ref.
Chromatin accessibility	ATAC-seq/DNase-seq	Cistrome	371	[22]
H3K27ac	ChIP-seq	Cistrome	555	[22]
H3K4me1	ChIP-seq	Cistrome	364	[22]
RNAPII binding	ChIP-seq	Cistrome	166	[22]
FANTOM5 enhancer	–	FANTOM5	–	[19]
SCREEN enhancer	–	SCREEN	–	[20]
EnhancerAtlas enhancer	–	EnhancerAtlas 2.0	197	[18]
Super-enhancer	–	SEdb 2.0	1705	[17]

Note: ATAC-seq, assay for transposase-accessible chromatin using sequencing; DNase-seq, DNase I hypersensitive site sequencing; ChIP-seq, chromatin immunoprecipitation sequencing; RNAPII, RNA polymerase II; FANTOM5, functional annotation of the mammalian genome 5; ENCODE, ENCYClopedia of DNA Elements; SCREEN, Search Candidate *cis*-Regulatory Elements by ENCODE; SEdb, super-enhancer database.

hallmark enrichment [29], are performed based on hypergeometric tests using our in-house scripts (<https://github.com/zhangyw0713/FunctionEnrichment>).

Results

Web interface of eRNA-ID

eRNA-ID is designed for eRNA identification based on *de novo* assembled transcriptomes. In the input interface (http://bioinfo.szbl.ac.cn/eRNA_IDO/eRNA-ID), users are required to upload a transcriptome profile in GTF format, which can be generated from RNA-seq and GRO-seq data, and define enhancer regions using either provided marker buffet or by uploading a customized BED file. eRNA-ID adopts an analytical workflow similar to ncFANs-eLnc [13] to identify eRNAs (see Method). As shown in Table S8, the major advantage of eRNA-ID over ncFANs is its prebuilt buffet of eight kinds of enhancer markers (H3K27ac, H3K4me1, chromatin accessibility, RNAPII binding, SEdb 2.0 super-enhancers [17], and three types of enhancer annotations from EnhancerAtlas 2.0 [18], FANTOM5 [19], and SCREEN [20] databases), enabling users to customize enhancer regions of interest. For example, users may require high-confidence enhancer regions simultaneously labeled by multiple markers or may want to obtain as many enhancers as possible by merging multiple markers. The processing procedure of eRNA-ID is fast. For instance, a GRO-seq-derived transcriptome with 3483 transcripts (SRA008244) took 45 s, and a RNA-seq-derived *de novo* transcriptome with 222,848 transcripts (GSM2824220) took 88 s (default parameters).

In the output interface of eRNA-ID (Figure 2), the chromatin coordinates, enhancers, and putative targets (adjacent genes within ± 1 Mb of eRNAs) of identified eRNAs are displayed in a table. Users can also view the information in a genome browser based on JBrowse [30]. Moreover, functional annotation can be conducted for these novel eRNAs by clicking on the “Deliver eRNA to eRNA-Anno” button.

Web interface of eRNA-Anno

eRNA-Anno is designed for the network-based interactome discovery and functional annotation of eRNAs. In this module, users input the chromatin coordinates of novel eRNAs (Figure 3A) or the identifiers/locations of known eRNAs annotated in HeRA [9] and eRic [12] databases (Figure 3B), followed by network selection and parameter setting. eRNA-Anno first quantifies the eRNA expression levels based on RNA-seq data from TCGA and GTEx portal. As hundreds of RNA-seq samples require a long processing time, the read coverages from BigWig files were used to speed up the quantification (see Method). To examine the reliability of this method, the expression levels of known eRNAs acquired via this method were correlated with those based on the canonical featureCounts [31] method obtained from HeRA and eRic databases. The results reveal that our method is highly correlated with the canonical method (Figure S1A and B) and is approximately 400 times faster (Figure S1C).

eRNA-Anno annotates the functions of eRNAs by discovering their interactomes based on eRNA-PCG networks, including normal co-expression networks based on GTEx expression profiles [21], cancer co-expression networks based on TCGA expression profiles (<https://portal.gdc.cancer.gov/>), and eRNA-centric regulatory networks. Co-expression relationships are widely used to annotate the

functions of eRNAs [32–34]. Additionally, eRNAs have been reported to exert regulatory functions by interacting with other biomolecules, such as TFs [35–37], RBPs [4,38,39], and target genes activated by E–P loops [40,41]. Therefore, the regulatory network can be used for eRNA functional annotation, similar to approaches used for other ncRNAs [13,42–45]. The network construction procedure is detailed in the Method section. Parameters include tissue/cancer type of expression profile, co-expression coefficient, significance threshold, biosamples of interaction relationships, and epigenetic landscape (Figure 3C and D).

Upon receiving launch instructions, eRNA-Anno initiates the analytical procedure (see Method) to identify the potential targets of query eRNAs from selected networks and annotate their functions based on hub-based and module-based strategies. The whole procedure typically takes tens of minutes, depending on the number of input eRNAs (Figure S2). Hence, users are recommended to set an email notification or record the task ID for result retrieval when submitting a task with a large set of eRNAs.

In the output interface, eRNA-Anno provides basic information about eRNAs (*i.e.*, location and expression, epigenetic landscape, and disease relevance) and putative targets and functions based on various networks. In the “Location and expression” section, chromatin coordinates, expression levels in normal and cancer samples, adjacent genes (≤ 1 Mb), and overlapping super-enhancers are listed in the table (Figure 4A). Furthermore, eRNA-Anno profiles active enhancer markers (H3K27ac and H3K4me1) and chromatin accessibility of eRNA regions to evaluate the activity of enhancers where eRNAs are transcribed (Figure 4B). Considering that mutations in eRNA regions are often related to eRNA expression and subsequent disease development [46], clinically relevant mutations within query eRNA regions are displayed in the “Disease relevance” section (Figure 4C) and can be visualized in the genome browser (Figure 4D). Moreover, the interactome and predicted functions of eRNAs based on the selected networks are displayed in the second part (Figure 5). For example, in a cancer co-expression network (Figure 5A), the eRNA-PCG network is visualized in a force-directed layout, and the functions of connected PCGs are provided (Figure 5B). Since genes with similar functions tend to be concentrically distributed, eRNA-Anno extracts hubs and modules composed of tightly connected genes from the overall network (Figure 5C). The functions of query eRNAs can be inferred by the functions of the PCGs within the same module or hub (Figure 5D).

In addition, the eRNA-centric regulatory network (Figure 5E) provides a visualization of the relationships of eRNAs with TFs, RBPs, and E–P loops in multiple modes, including network topology, table, and genome browser. Similarly, the functions of eRNAs can be inferred by the related biomolecules in the overall network, modules, or hubs. The results of individual networks can be combined into a summary (Figure 6).

A case study demonstrating the usage of eRNA-Anno

Since the input interface has many user-dependent options and the output interface displays interactive information, a case study is described to introduce the usage and interpret the results obtained from eRNA-Anno. CCAT1 and LINC02257, which have been characterized as colon cancer-associated

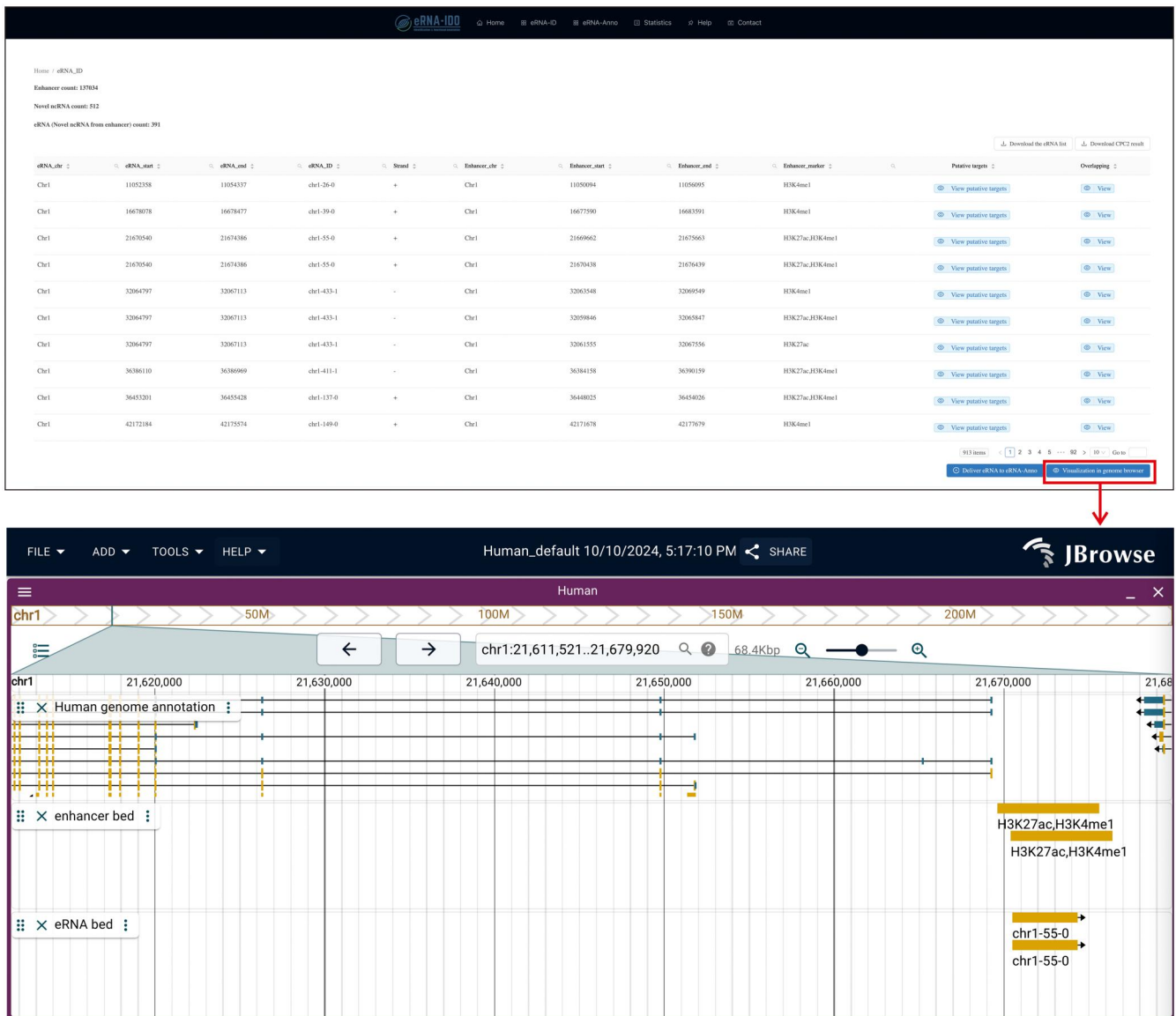


Figure 2 The output interface of eRNA-ID

The predicted eRNA locations, enhancer regions, markers for active enhancers, putative targets (adjacent genes), and overlapping lncRNAs are displayed in a table and can be visualized in the genome browser. Additional details are shown in the demo: http://bioinfo.szbl.ac.cn/eRNA_IDO/retrieve/?taskId=5a9LFXS8oGCm. lncRNA, long non-coding RNA.

eRNAs [47,48], were analyzed in this study and input in GTF format. Subsequently, “TCGA-COAD” and “GTEx-Colon Transverse” were chosen, co-expression and regulatory networks were selected, the parameters were set, and eRNA-IDO was launched, as depicted in Figure 3.

In the output interface, eRNA-Anno revealed that both CCAT1 and LINC02257 exhibited higher expression levels in colorectal cancer (Figure 4A) and showed enriched active enhancer markers (Figure 4B), which is consistent with previous studies [47,48]. Additionally, the genomic regions of CCAT1 and LINC02257 harbor carcinoma-associated mutations (Figure 4C), indicating their clinical significance. Subsequently, the co-expression network in colon adenocarcinoma was further investigated to evaluate the interactome and functions of CCAT1 and LINC02257. The topology of the co-expression network revealed limited connections between CCAT1 and LINC02257 (Figure 5A), indicating their independent regulatory roles. Furthermore, functional

enrichment analysis of the co-expressed PCGs revealed that CCAT1 and LINC02257 were potentially enriched in translation and cell cycle pathways (Figure 5B). The CCAT1-containing module precisely pinpointed the role of CCAT1 in regulating cell cycle (Figure 5C and D), which conforms to previous findings [30,49]. Moreover, the eRNA-centric regulatory network detected the interactive TFs, RBPs, and genes involved in E-P loops. These interactive molecules were enriched in cell cycle and cancer pathways, suggesting the similar role of CCAT1 and LINC02257 (Figure 5E). Additionally, a genome browser based on JBrowse [30] visualized eRNA locations and the mutational, epigenetic, and interactive landscapes (Figure 5E). Finally, the nodes and edges from two separate networks were overlapped to determine high-confidence interactions of CCAT1 in a cell cycle-related module (Figure 6), including some previously reported targets such as CDK4 [50] and SOX4 [51]. This case study demonstrates the application of eRNA-Anno, displaying its ability

A

Novel eRNA Known eRNA **Move to input page for known eRNAs**

▼ User input

1. Upload a query eRNA list (<50Mb is recommended) [load example](#)

GTF BED

Click or drag file to this area to upload

2. Input candidate target genes to constrain the network (Optional) [load example](#)

Click or drag file to this area to upload

▼ Expression quantification

Please define expression profiles for eRNA quantification and co-expression network construction

Based on the preconstructed expression profiles

Cancer

Normal

or users upload the expression profiles (both below are required) [load example](#)

1. eRNA expression profile 2. Expression of potential targets

Click or drag file to this area to upload

Click or drag file to this area to upload

▼ Network Selection

> Normal and cancer co-expression network YES

> Regulatory relationships YES

▼ Genetic/Epigenetic landscape

> Epigenetic feature YES

> Mutation and Disease relevance YES

B

Novel eRNA Known eRNA

▼ User input

Input query eRNA list in the below box [load example](#) or upload the list via the button

upload

ENSR_ID Chromatin_coordinate Adjacent_gene

Input eRNAs in following format:

```
ENSR0000032472
ENSR00000261094
ENSR00000261096
ENSR0000032520
```

Or for chromatin coordinates:
chr5:15911700-15923700
chr9:63104200-63110200

C

▼ Network Selection

▼ Normal and cancer co-expression network YES

NOTE: co-expression network construction is based on the expression profiles you selected / uploaded in above "Expression quantification" section

▼ Advanced options (optional network parameters)

Minimal correlation coefficient between eRNA and protein-coding gene

Minimal correlation coefficient between two protein-coding genes

Significance threshold(PDR)

Select the negative/positive correlation

D

▼ Regulatory relationships YES

Transcription factor
Overall / Overall

RNA binding protein
Overall / Overall

Enhancer-promoter loop
Overall / Overall

Choose biosample

Figure 3 The input interface of eRNA-Anno

A. The input contents include a potential eRNA list, optional target candidates, parameters for eRNA quantification, network selection, and genetic/epigenetic landscape. **B.** The input interface for known eRNAs annotated in HeRA [9] and eRic [12]. **C.** Parameters for the construction of the co-expression network. **D.** Parameters for the construction of the eRNA-centric regulatory network. HeRA, Human enhancer RNA Atlas; eRic, eRNA in cancer.

to comprehensively and reliably predict eRNA interactome and functions.

Discussion

As a web server dedicated to eRNA analysis, eRNA-IDO provides a convenient method for eRNA identification, interactome discovery, and functional annotation. The major advantages of eRNA-IDO include but are not limited to the following. First, eRNA-ID includes eight kinds of enhancer markers, offering a more convenient and customized approach for enhancer definition compared to ncFANs-eLnc

[13], which only includes the H3K27ac marker. Second, eRNA-Anno is applicable to both novel and known eRNAs. Considering the poor characterization of eRNAs, the applicability to novel eRNAs grants eRNA-Anno higher flexibility and biological practicability compared to other tools requiring known identifiers (such as ncFANs v2.0 [13]) and other databases [9–12]. The detailed comparison between eRNA-IDO and ncFANs v2.0 is displayed in Table S8. Third, biological context-specific expression and interaction profiles are prebuilt in eRNA-Anno. Compared to tools without biological specificity such as AnnoLnc2 [14], eRNA-Anno is expected to provide more precise findings for *in vivo*



Figure 4 The output interface of eRNA-Anno showing the basic information of query eRNAs CCAT1 and LINC02257. **A.** The locations and expression levels of CCAT1 and LINC02257. **B.** The epigenetic landscape. **C.** Clinically relevant mutations within the genomic regions of CCAT1 and LINC02257. **D.** The genome browser visualization accessible via the "Visualization in genome browser" button. Further details are shown in the demo: http://bioinfo.szbl.ac.cn/eRNA_IDO/retrieve/?taskid=97XPLicEaj4eUY/.



Figure 5 The output interface of eRNA-Anno showing the interactomes and functions of CCAT1 and LINC02257

A. The co-expression network of CCAT1 and LINC02257 in human colorectal cancer. **B.** The enriched KEGG pathways of CCAT1- and LINC02257-connected PCGs. **C.** Visualization of the CCAT1-containing module. **D.** The enriched KEGG pathways of PCGs within the CCAT1-containing module. **E.** The CCAT1-centric and LINC02257-centric regulatory network.

investigations. Moreover, the prebuilt profiles facilitate the use of the service. Finally, eRNA-IDO is the first one-stop platform for eRNA identification, interactome discovery, and functional annotation.

Nevertheless, the limitations of the study should be acknowledged and overcome. First, eRNA-IDO is currently designed for human data, and additional species will be

supported in the future. Second, some characteristics such as the m⁶A modification [52] and RNA structure [53,54] are essential for eRNA functionality but are not evaluated by eRNA-IDO. Third, the current iteration of eRNA-IDO only considers normal tissue and cancer. In the future, a larger number of disease-specific and cell-specific expression and interaction profiles will be incorporated. Hopefully, eRNA-

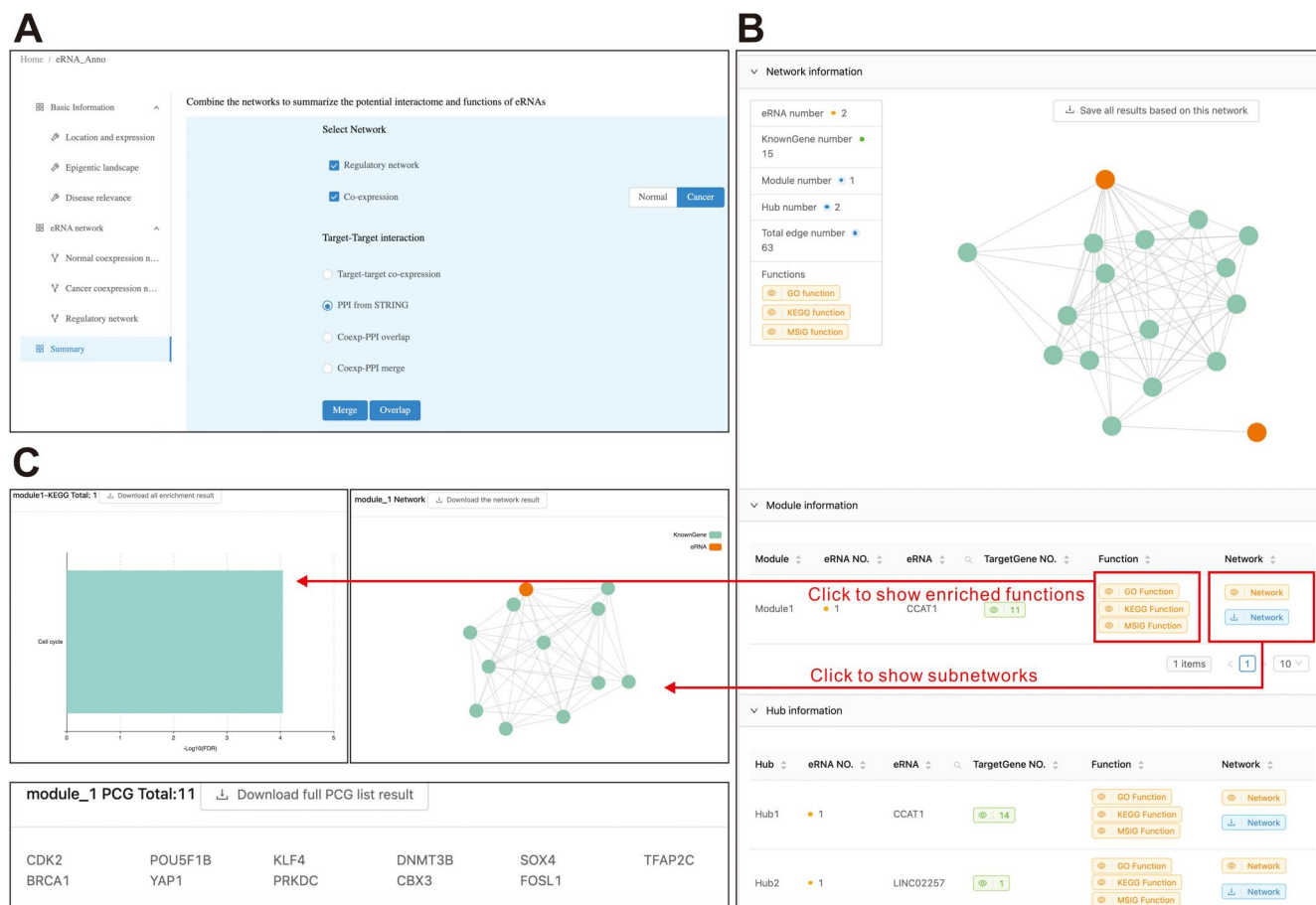


Figure 6 Summary of the interactome and functions of query eRNAs based on the combination of co-expression network and regulatory network

A. Parameter settings for network combination. **B.** A high-confidence network comprising the overlapping nodes and edges generated for CCAT1 and LINC02257. **C.** The CCAT1-containing module indicating its interactive genes and functions in cell cycle regulation.

IDO will benefit from user feedback and develop into a more powerful tool upon continuous updates.

Data availability

The eRNA-IDO web server is available at http://bioinfo.szbl.ac.cn/eRNA_IDO/.

CRediT author statement

Yuwei Zhang: Data curation, Methodology, Investigation, Writing – original draft. **Lihai Gong:** Methodology, Software, Visualization. **Ruofan Ding:** Data curation, Visualization. **Wenyan Chen:** Data curation, Investigation. **Hao Rong:** Data curation, Investigation. **Yanguo Li:** Data curation. **Fawziya Shameem:** Writing – review & editing. **Korakkandan Arshad Ali:** Writing – review & editing. **Lei Li:** Resources, Conceptualization, Supervision, Writing – review & editing. **Qi Liao:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. All authors have read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (<https://doi.org/10.1093/gpbjnl/qzae059>).

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32300430, 32100533, and 31970630), the Open Grant Funds from Shenzhen Bay Laboratory (Grant No. SZBL2021080601001 to Lei Li), the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY21C060002), and the Ningbo Major Research and Development Plan Project (Grant Nos. 2023Z226 and 2023Z171), China.

ORCID

0000-0002-8395-3020 (Yuwei Zhang)
 0009-0007-0880-8177 (Lihai Gong)
 0000-0003-1667-9866 (Ruofan Ding)
 0000-0002-4821-9306 (Wenyan Chen)
 0000-0002-1650-0670 (Hao Rong)
 0000-0001-9064-0033 (Yanguo Li)
 0009-0009-5011-7665 (Fawziya Shameem)
 0009-0002-9962-0543 (Korakkandan Arshad Ali)
 0000-0003-3924-2544 (Lei Li)
 0000-0001-6796-104X (Qi Liao)

References

- [1] Lam MT, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 2014; 39:170–82.
- [2] Han Z, Li W. Enhancer RNA: what we know and what we can achieve. *Cell Prolif* 2022;55:e13202.
- [3] Ounzain S, Pedrazzini T. Super-enhancer lncs to cardiovascular development and disease. *Biochim Biophys Acta* 2016; 1863:1953–60.
- [4] Jiao W, Chen Y, Song H, Li D, Mei H, Yang F, et al. HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. *Oncogene* 2018;37:2728–45.
- [5] Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 2013; 494:497–501.
- [6] Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 2013;498:516–20.
- [7] Bose DA, Donahue G, Reinberg D, Shiekharrar R, Bonasio R, Berger SL. RNA binding to CBP stimulates histone acetylation and transcription. *Cell* 2017;168:135–49.e22.
- [8] Song C, Zhang G, Mu X, Feng C, Zhang Q, Song S, et al. eRNAbase: a comprehensive database for decoding the regulatory eRNAs in human and mouse. *Nucleic Acids Res* 2024;52:D81–91.
- [9] Zhang Z, Hong W, Ruan H, Jing Y, Li S, Liu Y, et al. HeRA: an atlas of enhancer RNAs across human tissues. *Nucleic Acids Res* 2021;49:D932–8.
- [10] Chen H, Liang H. A high-resolution map of human enhancer RNA loci characterizes super-enhancer activities in cancer. *Cancer Cell* 2020;38:701–15.e5.
- [11] Jin W, Jiang G, Yang Y, Yang J, Yang W, Wang D, et al. Animal-eRNAdb: a comprehensive animal enhancer RNA database. *Nucleic Acids Res* 2022;50:D46–53.
- [12] Zhang Z, Lee JH, Ruan H, Ye Y, Krakowiak J, Hu Q, et al. Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat Commun* 2019; 10:4562.
- [13] Zhang Y, Bu D, Huo P, Wang Z, Rong H, Li Y, et al. ncFANs v2.0: an integrative platform for functional annotation of non-coding RNAs. *Nucleic Acids Res* 2021;49:W459–68.
- [14] Ke L, Yang DC, Wang Y, Ding Y, Gao G. AnnoLnc2: the one-stop portal to systematically annotate novel lncRNAs for human and mouse. *Nucleic Acids Res* 2020;48:W230–8.
- [15] Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. *Nucleic Acids Res* 2021; 49:D916–23.
- [16] Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res* 2017;45:W12–6.
- [17] Wang Y, Song C, Zhao J, Zhang Y, Zhao X, Feng C, et al. SEdb 2.0: a comprehensive super-enhancer database of human and mouse. *Nucleic Acids Res* 2023;51:D280–90.
- [18] Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res* 2020;48:D58–64.
- [19] Abugessaisa I, Ramiłowski JA, Lizio M, Severin J, Hasegawa A, Harshbarger J, et al. FANTOM enters 20th year: expansion of transcriptomic atlases and functional annotation of non-coding RNAs. *Nucleic Acids Res* 2021;49:D892–8.
- [20] The ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoshani N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020; 583:699–710.
- [21] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;369:1318–30.
- [22] Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res* 2019;47:D729–35.
- [23] Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:D1062–7.
- [24] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 2019;47:D941–7.
- [25] Chen W, Li J, Huang S, Li X, Zhang X, Hu X, et al. GCEN: an easy-to-use toolkit for gene co-expression network analysis and lncRNAs annotation. *Curr Issues Mol Biol* 2022;44:1479–87.
- [26] Zhao W, Zhang S, Zhu Y, Xi X, Bao P, Ma Z, et al. POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res* 2022; 50:D287–94.
- [27] Wang S, Gao S, Zeng Y, Zhu L, Mo Y, Wong CC, et al. N⁶-methyladenosine reader YTHDF1 promotes ARHGEF2 translation and RhoA signaling in colorectal cancer. *Gastroenterology* 2022; 162:1183–96.
- [28] Jiang P, Singh M. SPICi: a fast clustering algorithm for large biological networks. *Bioinformatics* 2010;26:1105–11.
- [29] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417–25.
- [30] Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* 2023;24:74.
- [31] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–30.
- [32] Cai H, Liang J, Jiang Y, Tan R, Hou C, Hou J. Integrative analysis of N⁶-methyladenosine-related enhancer RNAs identifies distinct prognosis and tumor immune micro-environment patterns in head and neck squamous cell carcinoma. *Cancers (Basel)* 2022; 14:4657.
- [33] Chen X, Yuan J, Xue G, Campanario S, Wang D, Wang W, et al. Translational control by DHX36 binding to 5'UTR G-quadruplex is essential for muscle stem-cell regenerative functions. *Nat Commun* 2021;12:5043.
- [34] Yao P, Lin P, Gokoolparsadh A, Assareh A, Thang MW, Voineagu I. Coexpression networks identify brain region-specific enhancer RNAs in the human brain. *Nat Neurosci* 2015; 18:1168–74.
- [35] Allen MA, Andrysk Z, Dengler VL, Mellert HS, Guarnieri A, Freeman JA, et al. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. *Elife* 2014;3:e02200.
- [36] Azofeifa JG, Allen MA, Hendrix JR, Read T, Rubin JD, Dowell RD. Enhancer RNA profiling predicts transcription factor activity. *Genome Res* 2018;28:334–44.
- [37] Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D, et al. Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. *Genome Res* 2018;28:159–70.
- [38] Bai X, Li F, Zhang Z. A hypothetical model of *trans*-acting R-loops-mediated promoter-enhancer interactions by Alu elements. *J Genet Genomics* 2021;48:1007–19.
- [39] Huang Z, Yu H, Du G, Han L, Huang X, Wu D, et al. Enhancer RNA lnc-CES1-1 inhibits decidual cell migration by interacting with RNA-binding protein FUS and activating PPAR γ in URPL. *Mol Ther Nucleic Acids* 2021;24:104–12.
- [40] Arnold PR, Wells AD, Li XC. Diversity and emerging roles of enhancer RNA in regulation of gene expression and cell fate. *Front Cell Dev Biol* 2020;7:377.

- [41] Harrison LJ, Bose D. Enhancer RNAs step forward: new insights into enhancer function. *Development* 2022;149:dev200398.
- [42] Zhang Y, Tao Y, Li Y, Zhao J, Zhang L, Zhang X, et al. The regulatory network analysis of long noncoding RNAs in human colorectal cancer. *Funct Integr Genomics* 2018;18:261–75.
- [43] Zhang Y, Tao Y, Liao Q. Long noncoding RNA: a crosslink in biological regulatory network. *Brief Bioinform* 2018;19:930–45.
- [44] Chen L, Zhang W, Li DY, Wang X, Tao Y, Zhang Y, et al. Regulatory network analysis of LINC00472, a long noncoding RNA downregulated by DNA hypermethylation in colorectal cancer. *Clin Genet* 2018;93:1189–98.
- [45] Luo C, Tao Y, Zhang Y, Zhu Y, Minyao DN, Haleem M, et al. Regulatory network analysis of high expressed long non-coding RNA LINC00941 in gastric cancer. *Gene* 2018;662:103–9.
- [46] Hu X, Wu L, Yao Y, Ma J, Li X, Shen H, et al. The integrated landscape of eRNA in gastric cancer reveals distinct immune subtypes with prognostic and therapeutic relevance. *iScience* 2022;25:105075.
- [47] McClelland ML, Mesh K, Lorenzana E, Chopra VS, Segal E, Watanabe C, et al. CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest* 2016;126:639–52.
- [48] Xiao J, Liu Y, Yi J, Liu X. LINC02257, an enhancer RNA of prognostic value in colon adenocarcinoma, correlates with multi-omics immunotherapy-related analysis in 33 cancers. *Front Mol Biosci* 2021;8:646786.
- [49] Liu Z, Chen Q, Hann SS. The functions and oncogenic roles of CCAT1 in human cancer. *Biomed Pharmacother* 2019;115:108943.
- [50] Li JL, Li R, Gao Y, Guo WC, Shi PX, Li M. LncRNA CCAT1 promotes the progression of preeclampsia by regulating CDK4. *Eur Rev Med Pharmacol Sci* 2018;22:1216–23.
- [51] Hu B, Zhang H, Wang Z, Zhang F, Wei H, Li L. LncRNA CCAT1/miR-130a-3p axis increases cisplatin resistance in non-small-cell lung cancer cell line by targeting SOX4. *Cancer Biol Ther* 2017;18:974–83.
- [52] Lee JH, Wang R, Xiong F, Krakowiak J, Liao Z, Nguyen PT, et al. Enhancer RNA m⁶A methylation facilitates transcriptional condensate formation and gene activation. *Mol Cell* 2021;81:3368–85.e9.
- [53] Cheng JH, Pan DZ, Tsai ZT, Tsai HK. Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. *Sci Rep* 2015;5:12648.
- [54] Ren C, Liu F, Ouyang Z, An G, Zhao C, Shuai J, et al. Functional annotation of structural ncRNAs within enhancer RNAs in the human genome: implications for human disease. *Sci Rep* 2017;7:15518.