DATABASE
The Journal of Biological Databases and Curation

CenhANCER: a comprehensive cancer enhancer database for primary tissues and cell lines

Zhi-Hui Luo[1,*,‡], Meng-Wei Shi[2,3,4,5,‡], Yuan Zhang[2,3,4,5,‡], Dan-Yang Wang[2,3], Yi-Bo Tong[2,3], Xue-Ling Pan[2,3,4,5] and ShanShan Cheng [ORCID][1,*]

[1]Department of Epidemiology and Biostatistics, Ministry of Education Key Laboratory of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, 13 Hangkong Road, Wuhan, Hubei 430030, P.R. China
[2]Hubei Hongshan Laboratory, College of Biomedicine and Health, Huazhong Agricultural University, No. 1, Shizishan Street, Wuhan, Hubei 430070, China
[3]Hubei Key Laboratory of Agricultural Bioinformatics, College of Life Science and Technology, Huazhong Agricultural University, No. 1, Shizishan Street, Wuhan, Hubei 430070, China
[4]Shenzhen Institute of Nutrition and Health, Huazhong Agricultural University, 97 Buxin Road, Shenzhen 518000, China
[5]Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, 97 Buxin Road, Shenzhen 518000, China

*Corresponding author: Tel: +86 18163358670; Email: luozhihui1123@163.com
Correspondence may also be addressed to ShanShan Cheng. Email: sscheng@hust.edu.cn
‡These authors contributed equally to this work.

Abstract

Enhancers, which are key tumorigenic factors with wide applications for subtyping, diagnosis and treatment of cancer, are attracting increasing attention in the cancer research. However, systematic analysis of cancer enhancers poses a challenge due to the lack of integrative data resources, especially those from tumor primary tissues. To provide a comprehensive enhancer profile across cancer types, we developed a cancer enhancer database CenhANCER by curating public resources including all the public H3K27ac ChIP-Seq data from 805 primary tissue samples and 671 cell line samples across 41 cancer types. In total, 57 029 408 typical enhancers, 978 411 super-enhancers and 226 726 enriched transcription factors were identified. We annotated the super-enhancers with chromatin accessibility regions, cancer expression quantitative trait loci (eQTLs), genotype-tissue expression eQTLs and genome-wide association study risk single nucleotide polymorphisms (SNPs) for further functional analysis. The identified enhancers were highly consistent with accessible chromatin regions in the corresponding cancer types, and all the 10 super-enhancer regions identified from one colorectal cancer study were recapitulated in our CenhANCER, both of which testified the high quality of our data. CenhANCER with high-quality cancer enhancer candidates and transcription factors that are potential therapeutic targets across multiple cancer types provides a credible resource for single cancer analysis and for comparative studies of various cancer types.

Database URL: http://cenhancer.chenzxlab.cn/

Introduction

Enhancers, which are short DNA regions commonly marked by the histone modification H3K27 acetylation (H3K27ac), can positively regulate spatiotemporal gene expression (1) and have been widely explored in a variety of cancer types (2). For example, enhancers display distinct profiles in four histotypes of ovarian cancer, which indicates their application value in cancer subtyping (3). The activity of H3K27ac is correlated with clinical characteristics of acute myeloid leukemia, suggesting its prognostic application in cancer (4). A therapeutic strategy based on enhancer-blocking bromodomain inhibitors is devised for colorectal cancer, implying its therapeutic potential (5). Furthermore, recent studies have revealed the *cis*-regulatory effect of enhancers in disease development with the aid of risk single nucleotide polymorphisms (SNPs) identified from genome-wide association studies (GWASs) (6, 7). All these findings show the key role of enhancers in cancer research.

Although cancer onset and progression are widely acknowledged to be associated with enhancer activities, the functions and mechanisms of regulation of most enhancers remain elusive due to the limited integrative data sources and the challenging nature of functional analysis. Although H3K27ac biosamples have been rapidly accumulated in recent years, there is no large-scale pan-cancer project with this type of data, while the existing large epigenomic projects such as ENCODE and Roadmap mainly involve normal tissues and cell lines (8).

To provide comprehensive cancer enhancer resources, we established a user-friendly database CenhANCER (http://cenhancer.chenzxlab.cn/) by manually curating all public

H3K27ac ChIP-Seq data of cancer primary tissues and cell lines from Gene Expression Omnibus (GEO) database, and these data were reliable for enhancer identification (9). Typical enhancers, super-enhancers and transcription factors (TFs) were identified by standard pipelines. To annotate enhancers and their association with cancer, we integrated TCGA ATAC-seq (10), cancer expression quantitative trait locus (eQTL) (11), genotype-tissue expression (GTEx) eQTL (12) and GWAS risk SNP data (13) with the enhancers. CenhANCER is thus a database that consolidates high-quality cancer enhancer candidates and TFs that are potential therapeutic targets, a resource that will be highly valuable for cancer research.

Materials and methods

The database was established through data integration, enhancer identification, enhancer annotation and TF enrichment (Figure 1).

Data sources

We collected cancer-related ChIP-Seq samples from GEO database by a text-mining pipeline we developed. Notably, all the information from National Center for Biotechnology Information was obtained by using the 'Bio.Entrez' python package. First, we retrieved 1203 study projects (GSE accessions) from GEO with '(H3K27ac[All Fields] AND ("neoplasms"[MeSH Terms] OR cancer[All Fields])) AND "Homo sapiens"[Organism] AND "gse"[Filter]' as keyword. Then, we extracted 51 748 samples from these projects. Second, the metadata containing sample attribute descriptive information in the form of free text were extracted from the xml files obtained in the Sequence Read Archive (SRA). Third, to ensure that the samples were H3K27ac ChIP-seq data, these samples whose attributes contained no 'ChIP-Seq' or 'H3K27ac' were filtered. The primary tissue samples with their attributes containing 'cell line' were filtered out. Then, we manually curated the samples without any chemical treatment from 2317 samples of 251 projects and further integrated their corresponding input and adjacent samples. We limited cell line samples to those with attributes containing 'cell line' or 'cell_line' and ensured that the selected samples were not treated with drugs. Finally, we obtained 805 primary cancer tissues and 671 cancer cell lines across 41 cancer types. Each cancer type is manually assigned an abbreviation (Supplementary Table S1).

Detection of typical enhancers and super-enhancers

The raw fastq data were extracted from SRA files with fastq-dump (14), and their quality control was performed using fastqc (15). The low-quality reads were excluded. Then, adapters were removed by fastp with default parameters (16). Clean reads were mapped to the
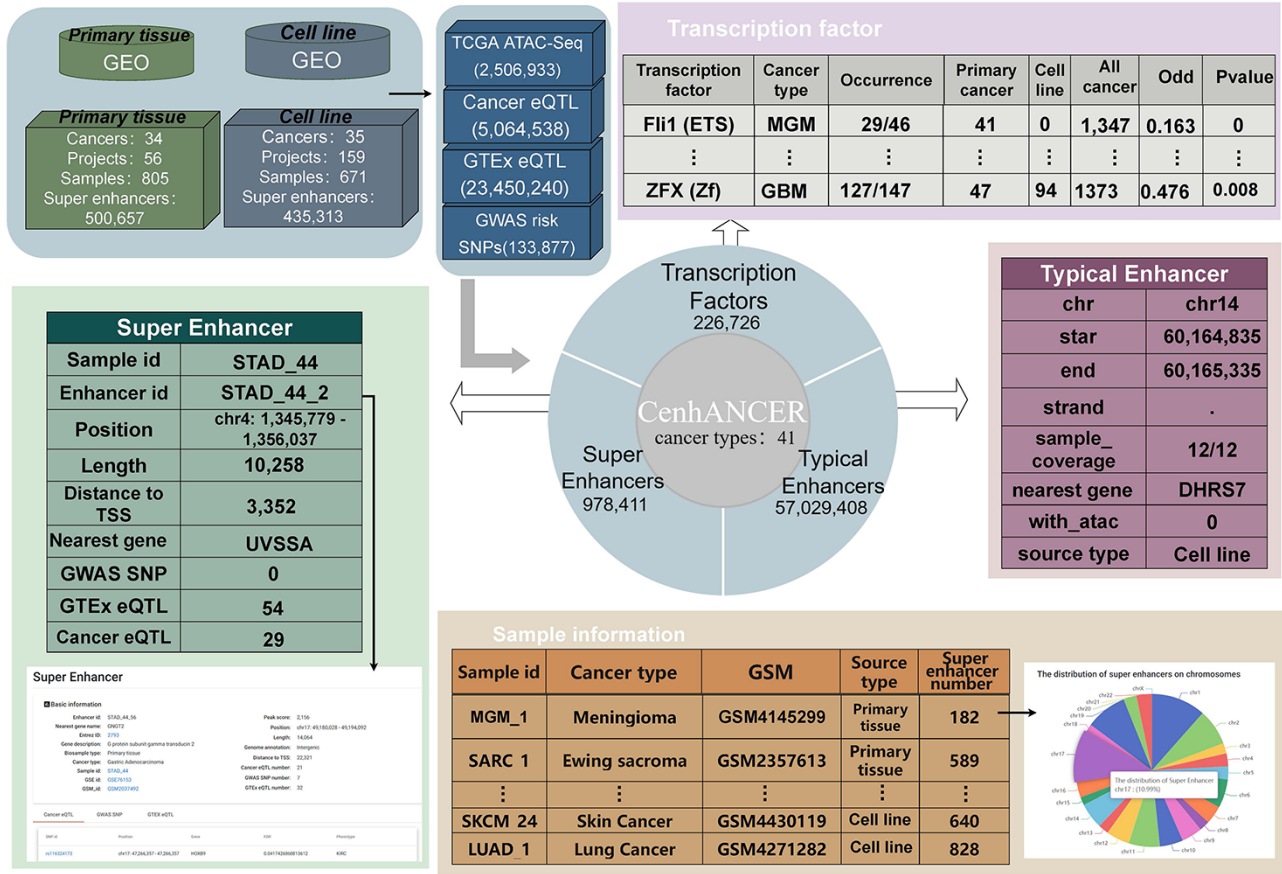

Figure 1. CenhANCER database structure and content. The main data include ChIP-Seq data of H3K27ac samples covering cancer primary tissues and cell line samples. ATAC-seq ACRs, cancer eQTLs, GTEx eQTLs and GWAS risk SNPs are supplementary annotation for enhancers. CenhANCER consist of four major parts, namely, sample information, super-enhancers, typical enhancers and TFs.

human reference genome (GRCh38, http://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_38/GRCh38.primary_assembly.genome.fa.gz). We processed single-end data by bowtie (version 1.2.1) [17] with the parameter '–sensitive' and paired-end data by bowtie2 (version 2.3.3) [18] with parameter '–chunkmbs 320 -m 1 –best'. Duplicate reads were removed by Picard.

Typical enhancers were identified from narrow peaks called by macs2 with parameters '–nomodel –keep-dup all -p 1E-5 –extsize 147'. We removed the peaks mapped to mitochondria and decoyed contigs. The peaks in the ENCODE hg38 blacklist were also excluded. Narrow peaks were handled according to Kundaje's pipeline (https://github.com/ENCODE-DCC/chip-seq-pipeline2). All the narrow peaks were normalized to a fixed width of 500 bp by extending 250 bp from the peak center to both upstream and downstream. All the narrow peaks in one cancer type were merged to obtain a cancer-specific peak set. Specifically, in the merging process, two overlapped peaks from two independent samples were merged into one, the peak with a higher score was retained and all the peaks were subjected to pairwise merging to obtain one peak set. Then, the peaks of another sample were merged into the previously merged peak set. After obtaining the overall set of one cancer, we intersected the narrow peaks of all the samples with the overall set. Only the peaks overlapped with those from at least two samples were identified as a reproducible cancer-associated enhancer.

Super-enhancers were identified from each sample by Rank Ordering of Super-Enhancers with the parameters '-t 2500 -s 12500' based on the peaks and the mapped bam file [19].

Overlaps of the typical enhancer and accessible chromatin region

The accessible chromatin regions (ACRs, ATAC-seq peak) of 23 cancer types were downloaded from a recent TCGA publication [10]. Of these 23 cancers, 17 cancers were overlapped with those from our CenhANCER. The ACRs were intersected with the cancer-specific enhancers by 'bedtools intersect' (https://bedtools.readthedocs.io/en/latest/) with default parameters.

Risk SNP–associated super-enhancers

We downloaded the known GWAS SNPs from the European Bioinformatics Institute GWAS catalog [13]. We normalized the phenotype of each disease with our previously developed tool pyMeSHSim [20]. To obtain the cancer terms, we filtered the terms that did not belong to 'C04 (neoplasms)'. Finally, a total of 366 GWAS projects of cancer were obtained. Then, we retrieved the lead SNPs from these projects and added SNPs with linkage disequilibrium (LD) $r^2 > 0.8$ to the GWAS lead SNPs. This LD information was obtained from the R package 'haploR' with the embeded data from the haploreg website [21]. In addition, SNP mutations for cancer types were downloaded from TCGA by TCGAbiolinks [22]. All the SNPs were mapped to the super-enhancers by 'bedtools intersect'.

GTEx eQTL–associated super-enhancers

The eQTLs obtained from the GTEx project were linked to the cancer super-enhancers and their target genes. The *.signif_variant_gene_pairs.txt.gz files were downloaded from GTEx database (https://gtexportal.org/home/datasets). We intersected the SNPs of eQTL with each super-enhancer with bedtools. Only significant eQTLs (P-value < 0.05) were retained for subsequent analysis.

TF enrichment with typical enhancers

For each sample, binding motifs of TFs were enriched with the corresponding typical enhancers, while transcriptional start sites (TSS) ± 2 kb were removed to exclude an effect from the nucleosome-free regions. The TF binding motifs were identified using the HOMER script (findMotifsGenome.pl) with the parameter of '-size 600' [23]. The typical enhancer peaks excluding regions overlapping with TSS ± 2 kb were used as target sequences. An approximately same number of sequences as the number of target sequences will be randomly picked from the genome as background sequences. Then, a hypergeometric test was conducted for occurrences of each TF motif in the target and background sequences [1],

$$P_{hyper} = \sum_{i=t}^{min(T,\,t+b)} \frac{\binom{t+b}{i}\binom{T+B-t-b}{T-i}}{\binom{T+B}{T}} \quad (1)$$

where T and B are the number of target and background sequences; t and b are the subset of target and background sequences containing at least one occurrence of the motif; i is the number of observed sequences containing the motif. p_{hyper} is the P-value of the hypergeometric test.

Then, HOMER applied the Benjamini–Hochberg false discovery rate method to adjust the P-values of all TFs (embedded in HOMER) in each sample. Significantly enriched TF binding motifs were selected based on a P-adj < 0.05. Then, the frequency of the TFs enriched from the samples of each cancer type was summarized. We estimated the odds ratio of each TF in each cancer type through Fisher's exact test.

Results
Data summary

Several enhancer-related databases had been reported in recent research studies, and the difference between CenhANCER and these databases is summarized in Table 1. CancerEnd is a cancer-associated database based on expressed enhancer candidates rather than histone modifications [24]. OncoBase provided the relationship between cancer somatic mutations and regulatory elements, which were mainly recognized in normal samples [25]. OncoCis uncovered somatic mutations located in *cis*-regulatory elements within limited cancer cell line data [26]. ENdb and DiseaseEnhancer contained manually collected and experimentally verified disease-associated high-quality enhancers, which makes their total records far less than databases based on high-throughput data [27, 28]. SEdb and SEanalysis were two comprehensive web platforms for super-enhancers mainly from normal ChIP-Seq samples [29, 30]. To the best of our knowledge, there is still no comprehensive enhancer analysis for all cancer types, due to limited data and the lack of an integrated data source. Here, we provided the first and the largest cancer enhancer resource of ChIP-Seq data from tumor primary tissues and cell lines.

CenhANCER covered 805 primary tissue samples across 34 cancer types and 671 cell line samples across 35 cancer types. Specifically, it included 2 506 933 ATAC-Seq

Table 1. Comparison between CenhANCER and relevant databases

| Database	Data source	Expression enhancer	Typical enhancer	Super-enhancer	ChIP-Seq
CenhANCER	Tumor tissue/cell lines	No	Yes	Yes	Yes
CancerEnD	Tumor tissue	Yes	No	No	No
OncoBase	Normal tissue/cell lines	No	Yes	Yes	Yes
OncoCis	Cell lines	No	Yes	Yes	Yes
ENdb	Normal tissue/cell lines	No	Yes	Yes	Yes
DiseaseEnhancer	Literatures	No	Yes	Yes	No
SEdb	Normal tissue/cell lines	No	No	Yes	Yes
SEanalysis	Normal tissue/cell lines	No	No	Yes	Yes
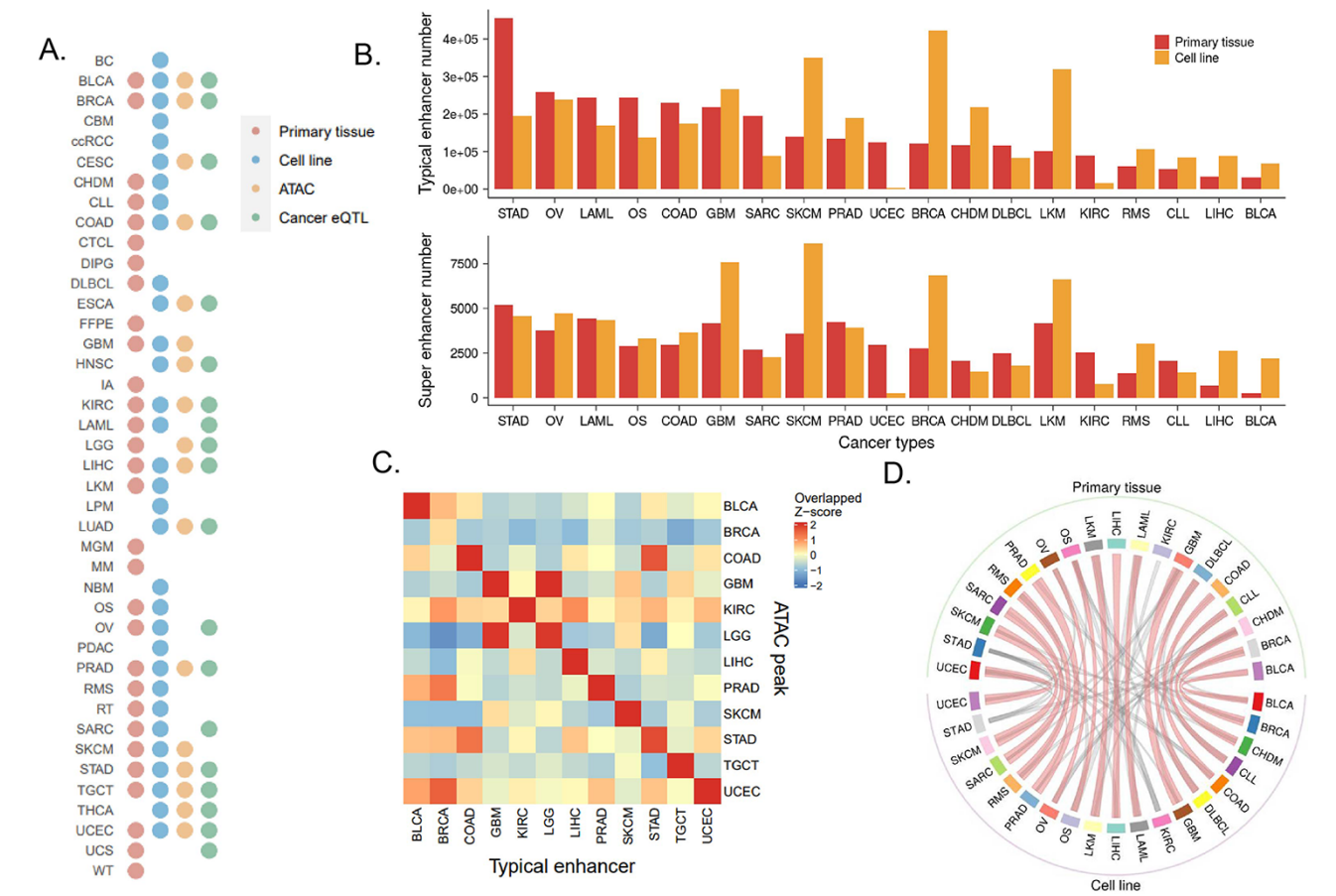
Figure 2. Overview of cancer typical enhancers and super-enhancers. (A) Distribution of samples of primary tissue H3K27ac, cell line H3K27ac, cancer eQTL and primary tissue ATAC-Seq in cancer types. (B) The upper panel exhibits the distribution of reproducible typical enhancer number in cancer types, and the lower panel displays the distribution of merged super-enhancer number in cancer types. (C) The heatmap of the overlap between the primary tissue typical enhancer and the primary tissue ATAC-Seq ACRs. (D) Circos plot of the typical enhancer overlap between primary tissue and cell lines. Each tissue is linked to the top three overlapped cell lines, and the thickened linkage line indicates the same cancer type shared by tissue and cell line.

ACRs across 17 cancers, 5 064 538 cancer eQTLs across 19 cancers, 23 450 240 GTEx eQTLs and 133 877 GWAS risk SNPs (Figure 1, Figure 2A). In total, we identified 57 029 408 typical enhancers, 978 411 super-enhancers (Figure 2B) and 226 726 enriched TFs.

To assess the quality of our identified enhancers, we examined the consistency between the typical enhancers and the ATAC-Seq ACRs, since the ACRs were highly correlated with enhancer distribution (31,32). Specifically, the typical enhancers for each cancer type were intersected with the corresponding ATAC-Seq ACRs. For both primary tissues and cell lines, the overlap between the typical enhancers and the corresponding ATAC-Seq ACRs in the same cancer was higher than

that in different cancers (Figure 2C, Figure S1). Interestingly, the typical enhancers covering more samples had a higher overlap with the ACRs for both primary tissues and cell lines (Figure S2A and B). Furthermore, an overlap between the primary tissue enhancers and the cell line enhancers was higher in the same cancer than in different cancers (Figure 2D). All these results indicated a high quality of the typical enhancers in CenhANCER.

Database access

We designed a user-friendly web platform for data exploration. The users can visit six functional pages including

'Browser', 'SuperEnhancer', 'TypicalEnhancer', 'TranscriptionFactor', 'Download' and 'Help' through the navigation bar in the home page (Figure S3A–F). On the 'Browser' page, users can access any sample by selecting one cancer type on the left menu. After clicking one sample, more detailed information will be provided (Figure S3A and B). The summarized super-enhancers can be searched through their loci or nearest genes on the 'SuperEnhancer' page (Figure S3C). After clicking one super-enhancer, the users can get its corresponding annotation information (Figure S3D). The reproducible cancer typical enhancer set, categorized by cancer type and data type, is provided on the 'TypicalEnhancer' page (Figure S3E). The 'TranscriptionFactor' page provides information including TF's name, occurrence number in a certain cancer and the odds ratio. Only TFs with $P < 0.01$ are taken into account (Figure S3F). All data in CenhANCER can be downloaded on the 'Download' page. A detailed tutorial usage is available on the 'Help' page.

Usage example
To illustrate the usage of CenhANCER in cancer enhancer exploration, we provide an example of super-enhancer analysis. A study of colorectal cancer (32) whose data have not been released is not covered by CenhANCER. In this study, *IER3*, *LIF*, *SLC7A5*, *CYP2S1*, *PHF19*, *RNF43*, *CEBPB*, *TBC1D16*, *TNFRSF6B* and *VEGFA* have been identified as variant super-enhancers loci–associated genes. We retrieve these genes in the 'SE' page via the 'Search by gene' with 'Source' option set as 'None', 'Cancer' option as 'None' or 'COAD' (colorectal cancer) and 'Gene' option as the gene symbol. Accordingly, we obtained the related super-enhancer loci in colorectal cancer (CRC) and all cancer samples. Ten out of 10 genes (100%) reported in the study are successfully detected in our database. The occurrence frequency of related super-enhancers in CRC is higher than that in other cancers, and this result is consistent for both primary tissue and cell line samples (Figure 3A and B). These findings suggest the high consistency of our data with the unpublished data. The *RNF43*-, *CYP2S1*- and *LIF*-related super-enhancers are shown as colorectal cancer–specific super-enhancers since the odds ratio of these three gene-related super-enhancers are >20 ($P < 0.05$) in primary tissues and >15 in cell lines ($P < 0.05$) (Figure 3C). To validate these three genes, we downloaded the cancer gene expression data by TCGAbiolinks (22) and explored the expression of these these genes in 11 cancers, which had corresponding H3K27ac ChIP-Seq data in CenhANCER. Interestingly, COAD was the cancer type where *RNF43*, *CYP2S1* and *LIF* showed the highest expression (Figure 3D), suggesting that these three genes' cancer-specific functions might be attributed to their adjacent super-enhancers. Furthermore, the super-enhancers of *CYP2S1* were also stomach adenocarcinoma
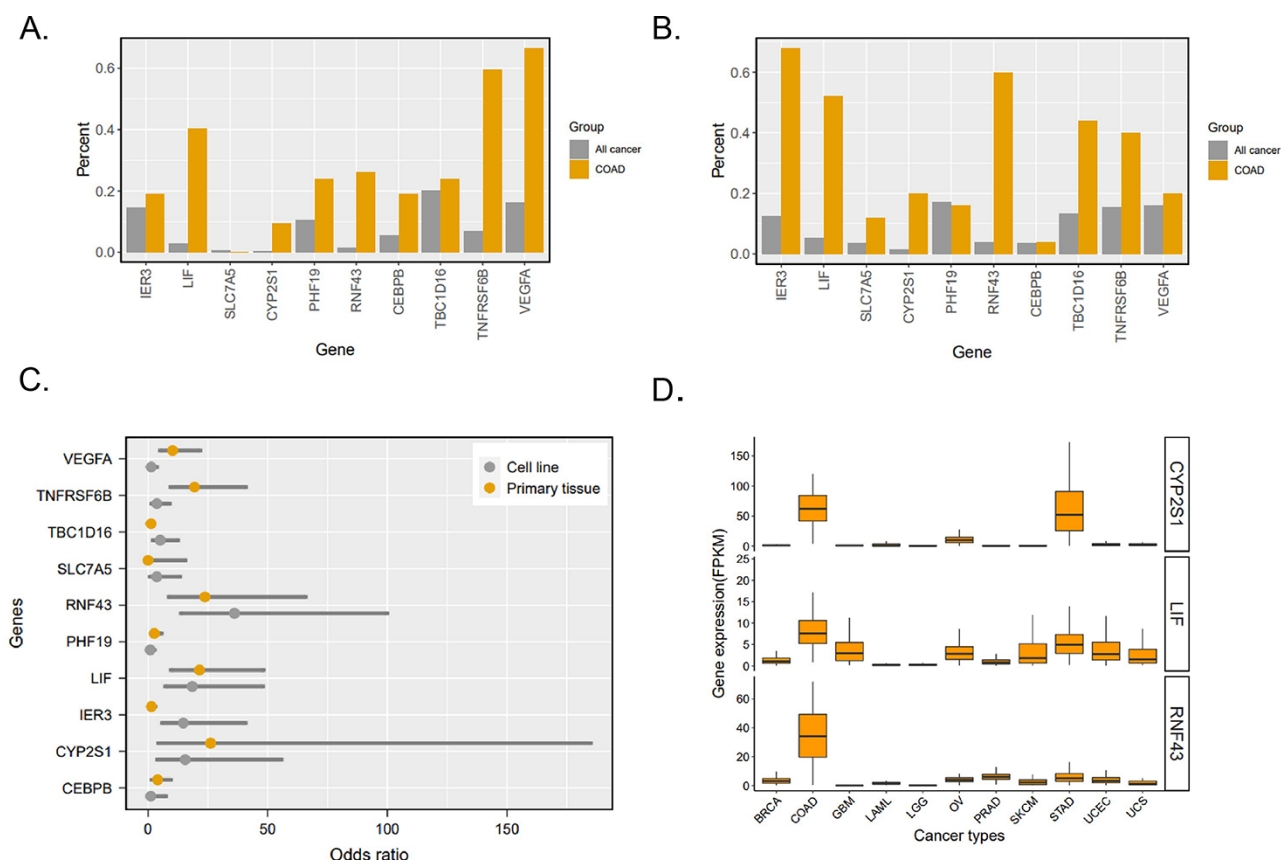

Figure 3. Analysis of 10 gene-related super-enhancers in colorectal cancer. (A) Percentage of 10 gene-related super-enhancers detected in primary tissue samples. 'All cancer' indicates cancer types excluding COAD. (B) Percentage of 10 gene-related super-enhancers detected in cell line samples. 'All cancer' indicates cancer types excluding COAD. (C) Ten genes' odds ratio of super-enhancers detected in colorectal cancer to those detected in all other cancers. (D) Three genes' expression (fragments per kilobase of exon per million mapped fragments, FPKM) in different cancer types according to TCGA RNA-Seq.

(STAD)–specific (Figures S4 and S5). As expected, the *CYP2S1* exhibited a higher gene expression in STAD than in other cancers.

To further verify the target genes that the super-enhancers regulate, we retrieved the eQTLs and SNPs located in super-enhancer regions related to the above-mentioned 10 genes. When a gene's eQTL or SNP is overlapped with an enhancer region of the same gene, it is highly possible for the enhancer to regulate this gene (10, 33). Finally, we obtained four *TNFRSF6B*-related GWAS risk SNPs in the gene's adjacent super-enhancer regions from the 'colorectal cancer' phenotype (Supplementary Table S2). We obtained 34 GTEx eQTL SNPs in the super-enhancer regions of *LIF*, *PHF19*, *TNFRSF6B* and *CYP2S1* (Supplementary Table S3) from 'Colon Sigmoid' or 'Colon Transverse' tissues. These results indicated that there might be a regulatory connection between these genes and their adjacent super-enhancers. It is worth noting that two (*LIF* and *CYP2S1*) out of the three colorectal cancer–specific genes aforementioned were identified from eQTLs and SNPs analyses. These results further confirmed that the cancer-specific functions might be attributed to their adjacent super-enhancers. All the aforementioned analyses suggest the in-depth applicability of CenhANCER in cancer research.

Discussion

CenhANCER is a comprehensive open resource to provide cancer enhancer profiles including typical enhancers, super-enhancers and potential TFs interacting with the enhancers. Convenient retrieval interface has been developed for enhancers, samples and cancer types. All processed data have been integrated for download. CenhANCER is the first ChIP-Seq data-focused cancer enhancer database, covering a large number of primary tissues and cell lines. It also integrated TCGA ATAC-Seq ACRs, GTEx eQTLs, cancer eQTLs and GWAS SNPs as supplementary function annotation. CenhANCER utilizes a semi-auto data collection method, making it easy to update its dataset. Since enhancers exhibit strong regulation function in cancer biology, more and more data will be generated. The cancer samples in our database are updated in real time, and it will provide valuable epigenetic information for both experimental and computational researchers.

Supplementary material

Supplementary material is available at *Database* online.

Code availability

All codes used for analysis and plotting are available at https://github.com/luozhhub/chip-seq-pyflow.

Funding

Conflict of interest

None declared.

References

1. Peng,Y. and Zhang,Y. (2018) Enhancer and super-enhancer: positive regulators in gene transcription. *Anim. Models Exp. Med.*, **1**, 169–179.
2. Sur,I. and Taipale,J. (2016) The role of enhancers in cancer. *Nat. Rev. Cancer*, **16**, 483–493.
3. Corona,R.I., Seo,J.-H., Lin,X. *et al.* (2020) Non-coding somatic mutations converge on the PAX8 pathway in ovarian cancer. *Nat. Commun.*, **11**, 2020.
4. Thoms,J.A., Truong,P., Subramanian,S. *et al.* (2021) Disruption of a GATA2-TAL1-ERG regulatory circuit promotes erythroid transition in healthy and leukemic stem cells. *Blood*, **138**, 1441–1455.
5. Orouji,E., Raman,A.T., Singh,A.K. *et al.* (2022) Chromatin state dynamics confers specific therapeutic strategies in enhancer subtypes of colorectal cancer. *Gut*, **71**, 938–949.
6. Boix,C.A., James,B.T., Park,Y.P. *et al.* (2021) Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*, **590**, 300–307.
7. Nasser,J., Bergman,D.T., Fulco,C.P. *et al.* (2021) Genome-wide enhancer maps link risk variants to disease genes. *Nature*, **593**, 238–243.
8. Moore,J.E., Purcaro,M.J., Pratt,H.E. *et al.* (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, **583**, 699–710.
9. Barrett,T., Suzek,T. O., Troup,D. B. *et al* (2005) NCBI GEO: mining millions of expression profiles–database and tools. *Nucleic Acids Res.*, **33**, D562–D566.
10. Corces,M.R., Granja,J.M., Shams,S. *et al.* (2018) The chromatin accessibility landscape of primary human cancers. *Science*, **362**, eaav1898.
11. Gong,J., Mei,S., Liu,C. *et al.* (2018) PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.*, **46**, D971–D976.
12. Aguet,F., Anand,S. and Ardlie,K.G., GTEx Consortium. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
13. MacArthur,J., Bowler,E., Cerezo,M. *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.
14. Leinonen,R., Sugawara,H., Shumway,M. *et al.* (2010) The sequence read archive. *Nucleic Acids Res.*, **39**, D19–D21.
15. Andrews,S. (2010) *FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
16. Chen,S., Zhou,Y., Chen,Y. *et al.* (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.
17. Langmead,B., Trapnell,C., Pop,M. *et al.* (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, 1–10.
18. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
19. Whyte,W.A., Orlando,D.A., Hnisz,D. *et al.* (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
20. Luo,Z.-H., Shi,M.-W., Yang,Z. *et al.* (2020) pyMeSHSim: an integrative python package for biomedical named entity recognition, normalization, and comparison of MeSH terms. *BMC Bioinform.*, **21**, 1–14.
21. Zhbannikov,I.Y., Arbeev,K., Ukraintseva,S. *et al.* (2017) haploR: an R package for querying web-based annotation tools. *F1000Research*, **6**, 97.

22. Colaprico,A., Silva,T.C., Olsen,C. *et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.*, **44**, e71–e71.

23. Heinz,S., Benner,C., Spann,N. *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

24. Kumar,R., Lathwal,A., Kumar,V. *et al.* (2020) CancerEnD: a database of cancer associated enhancers. *Genomics*, **112**, 3696–3702.

25. Li,X., Shi,L., Wang,Y. *et al.* (2019) OncoBase: a platform for decoding regulatory somatic mutations in human cancers. *Nucleic Acids Res.*, **47**, D1044–D1055.

26. Perera,D., Chacon,D., Thoms,J.A. *et al.* (2014) OncoCis: annotation of cis-regulatory mutations in cancer. *Genome Biol.*, **15**, 1–14.

27. Zhang,G., Shi,J., Zhu,S. *et al.* (2018) DiseaseEnhancer: a resource of human disease-associated enhancer catalog. *Nucleic Acids Res.*, **46**, D78–D84.

28. Bai,X., Shi,S., Ai,B. *et al.* (2020) ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res.*, **48**, D51–D57.

29. Jiang,Y., Qian,F., Bai,X. *et al.* (2019) SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res.*, **47**, D235–D243.

30. Qian,F.-C., Li,X.-C., Guo,J.-C. *et al.* (2019) SEanalysis: a web tool for super-enhancer associated regulatory analysis. *Nucleic Acids Res.*, **47**, W248–W255.

31. Klemm,S.L., Shipony,Z. and Greenleaf,W.J. (2019) Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.*, **20**, 207–220.

32. Li,Q. L., Lin,X., Yu,Y. L. *et al.* (2021) Genome-wide profiling in colorectal cancer identifies PHF19 and TBC1D16 as oncogenic super enhancers. *Nat. Commun.* **12**

33. Moore,J.E., Pratt,H.E., Purcaro,M.J. *et al.* (2020) A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. *Genome Biol.*, **21**, 1–16.