Published online 16 October 2022 *Nucleic Acids Research, 2023, Vol. 51, Database issue* ***D57–D69***
https://doi.org/10.1093/nar/gkac893

ChromLoops: a comprehensive database for specific protein-mediated chromatin loops in diverse organisms

Qiangwei Zhou[1,2], Sheng Cheng[1,2], Shanshan Zheng[1,2], Zhenji Wang[1,2], Pengpeng Guan[1,2], Zhixian Zhu[1,2], Xingyu Huang[1,2], Cong Zhou[1,2] and Guoliang Li[1,2,*]

[1]National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China and [2]Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, Key Laboratory of Smart Farming for Agricultural Animals, 3D Genomics Research Center, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

Received August 14, 2022; Revised September 14, 2022; Editorial Decision September 26, 2022; Accepted October 03, 2022

ABSTRACT

Chromatin loops (or chromatin interactions) are important elements of chromatin structures. Disruption of chromatin loops is associated with many diseases, such as cancer and polydactyly. A few methods, including ChIA-PET, HiChIP and PLAC-Seq, have been proposed to detect high-resolution, specific protein-mediated chromatin loops. With rapid progress in 3D genomic research, ChIA-PET, HiChIP and PLAC-Seq datasets continue to accumulate, and effective collection and processing for these datasets are urgently needed. Here, we developed a comprehensive, multispecies and specific protein-mediated chromatin loop database (ChromLoops, https://3dgenomics.hzau.edu.cn/chromloops), which integrated 1030 ChIA-PET, HiChIP and PLAC-Seq datasets from 13 species, and documented 1 491 416 813 high-quality chromatin loops. We annotated genes and regions overlapping with chromatin loop anchors with rich functional annotations, such as regulatory elements (enhancers, super-enhancers and silencers), variations (common SNPs, somatic SNPs and eQTLs), and transcription factor binding sites. Moreover, we identified genes with high-frequency chromatin interactions in the collected species. In particular, we identified genes with high-frequency interactions in cancer samples. We hope that ChromLoops will provide a new platform for studying chromatin interaction regulation in relation to biological processes and disease.

INTRODUCTION

Chromatin loops are a basic feature of eukaryotic genomes and can link regulatory elements, such as enhancers or transcription factor-binding sites (TFBS), physically close to their target genes. Aberrant chromatin loops may affect development and disease. For example, point mutations in the enhancer regulatory region ZRS (chromosome 7q36) result in preaxial polydactyly by regulating the ectopic expression of the *SHH* gene in mice (1,2). A 640-kb noncoding region at 8q24 contains distal cis-acting enhancers that control *Myc* expression in the developing face (3). Deletion of this region leads to modest facial morphological changes in mice and, sporadically, to cleft lip/cleft palate (CL/P) (3). In addition to enhancer elements, silencer elements, which repress gene expression through long-range chromatin interactions, also play important roles in human and mouse cell fate and development (4–6). Deletion of certain silencers in mice leads to transcriptional derepression of their interacting genes and pleiotropy for developmental phenotypes, including embryonic lethality (5). Furthermore, some recent reports showed that intragenic CTCF-mediated chromatin loops could regulate alternative splicing across individuals (7–9).

A few methods, including ChIA-PET, HiChIP and PLAC-Seq, have been proposed to capture genome-wide chromatin interactions mediated by specific proteins such as RNA polymerase II (RNAPII), transcription factors (TFs), and histone proteins, resulting in highly specific detection of chromatin loops (10–13). Practically, specific protein-mediated chromatin loops have higher resolution and are better suited for further study of chromatin regulatory mechanisms. With the rapid development of Three-Dimensional (3D) genomics research, ChIA-PET, HiChIP and PLAC-Seq datasets continue to accumulate. However,

*To whom correspondence should be addressed. Tel: +86 27 87285078; Fax: +86 27 87286876; Email: guoliang.li@mail.hzau.edu.cn

data generation and analysis of chromatin interactions are not an easy task for every independent laboratory. Effective collection and processing of these datasets are essential and urgently needed for uncovering the gene regulation mechanism of chromatin interactions.

More importantly, the regulation of genes through long-range chromatin loops could strongly influence cell development and disease. Genome-wide association studies (GWASs) have revealed many single nucleotide polymorphisms (SNPs) associated with complex diseases, and most of these SNPs are in noncoding chromosomal regions (14–16). However, it is unclear which genes are affected by disease-associated SNPs and their regulatory mechanisms. Currently, we know that the target genes that form physical interactions with GWAS-identified SNP regions can be identified by information on long-range chromatin interactions. For example, a common Parkinson's disease-associated SNP in a noncoding distal enhancer element upregulates the expression of α-synuclein (*SNCA*), a key gene implicated in the pathogenesis of Parkinson's disease (17). A recent study shows that multiple enhancers spanning an ultralong distance (>megabases) can interact with one another to modulate gene expression and disease risks through 3D chromosomal interactions (18). Since long-distance enhancer mutations are unlikely to occur simultaneously, this long-distance enhancer co-regulation provides robustness for gene expression. Nonetheless, it has also been shown that not all disruptions in chromatin interactions lead to changes in gene expression (19–21). And more systematic and complex regulatory mechanisms between long-range chromatin interactions and gene expression need to be further studied. Undoubtedly, in-depth studies of gene expression regulation through chromatin interactions rely heavily on comprehensive and reliable functional annotations. It is necessary to build a chromatin loop database to integrate and analyze loop results and annotate the genes and regions involved in the interactions. This will help accelerate the research on and discovery of chromatin loop regulatory mechanisms.

Here, we developed a comprehensive, multispecies and specific protein-mediated chromatin loop database (ChromLoops, https://3dgenomics.hzau.edu.cn/chromloops). The current version of ChromLoops encompasses 1 491 416 813 high-quality chromatin loops from 1030 datasets across 366 samples from 13 species, such as *Homo sapiens* and *Mus musculus*. These datasets include all ChIA-PET, HiChIP and PLAC-Seq data available from the NCBI Gene Expression Omnibus (GEO) (22). We annotated genes and regions overlapping with chromatin loop anchors on the basis of various types of information, including regulatory elements (enhancers, super-enhancers and silencers), variations (SNPs and quantitative trait loci (QTLs)), TFs, alternative splicing information, circRNAs, transcriptome-wide association study (TWAS) information, chromosome open accessible information, gene expression information and DNA methylation information. In addition, ChromLoops provides the search modules Genes, Regions, SNPs, Loop Circos, Browsers and Samples, showing the results of long-range chromatin interactions in individual or all samples. Moreover, physical interactions between distal regulatory elements play a key role in regulating gene expression. To characterize these interactions in cancer and normal samples, we manually curated the human datasets and provided genes with high-frequency chromatin interactions in cancer and normal samples and genes with high-frequency cancer-specific interactions. In a similar way, we also provided genes with high-frequency chromatin interactions in collected species.

To the best of our knowledge, ChromLoops is the first well-resourced and comprehensive database for specific protein-mediated chromatin loops. We hope that this comprehensive, multispecies chromatin loop database will stimulate research on various aspects of chromatin interactions.

MATERIALS AND METHODS

Database implementation

MySQL (version 5.7.26) was used to organize the database. Django and Vue were used for web interface development (Figure 1). ECharts (http://echarts.baidu.com) and D3 (https://d3js.org) were used as graphical visualization frameworks. The Network.js plugin (https://github.com/yuleicul/vue-network-d3) was used for chromatin interaction network visualization. WashU Epigenome Browser (release 52.5.0) (23) was used to construct the 'ChromLoops Browser' module, which shows chromatin interactions in arc mode and heatmap mode. The database has a convenient web interface to facilitate searching, browsing and downloading the results of chromatin loops (Figure 2A).

Chromatin loop data collection

ChIA-PET, HiChIP and PLAC-Seq are currently the most commonly used techniques to detect genome-wide chromatin interactions based on specific protein enrichment. To construct a comprehensive chromatin loop database, we searched the NCBI GEO database, downloaded all datasets (368 ChIA-PET, 597 HiChIP and 65 PLAC-Seq datasets) available as of November 2021, and filtered the low-quality data (loops < 500). These datasets originated from 13 species, including *Homo sapiens, Mus musculus, Arabidopsis thaliana* and *Oryza sativa* (Figure 2B and C). Finally, 1030 available datasets were manually curated and selected from 1072 datasets (Figure 2D). The datasets contained in ChromLoops are mainly enriched by proteins such as H3K27ac, H3K4me3, CTCF and RNAPII (Figure 2E).

Processing of ChIA-PET/HiChIP/PLAC-Seq data

For ChIA-PET data analysis, we used ChIA-PET Tool (V3) (24), a ChIA-PET data analysis tool developed in our laboratory. In addition, it is important to unify the analyses of all datasets for database construction. We updated ChIA-PET Tool (V3) to analyze HiChIP and PLAC-Seq data, and the code is available from GitHub (https://github.com/GuoliangLi-HZAU/ChIA-PET_Tool_V3). Therefore, the ChIA-PET Tool (V3) was used to analyze all chromatin interaction data in this study.

Filtering data with few interaction loops

To ensure high-quality analysis results, we filtered the data based on the number of chromatin loops analyzed and pre-

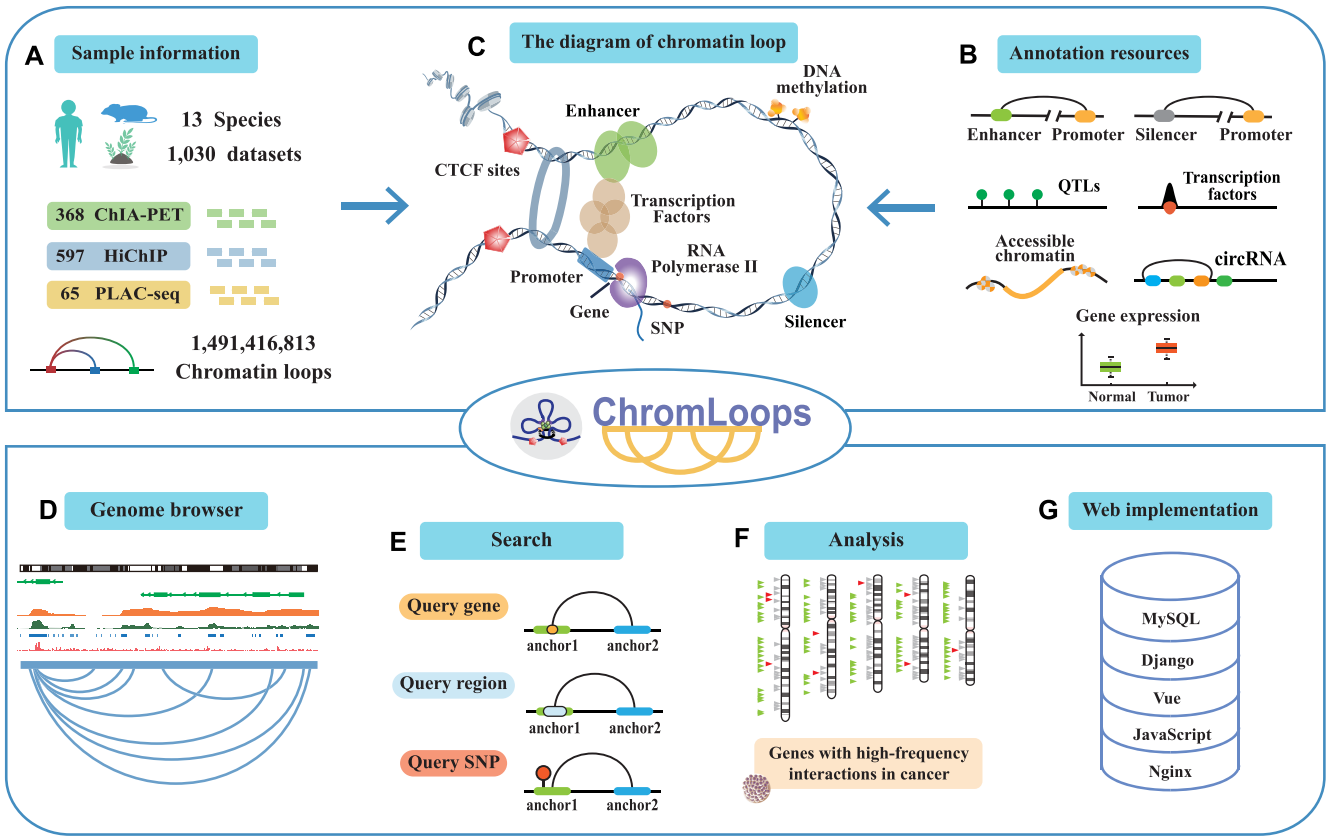

Figure 1. The procedure for ChromLoops database construction. (**A**) Sample information included in ChromLoops. ChromLoops contains 1030 datasets collected from the NCBI GEO for 13 species. (**B**) Annotation resources included in ChromLoops. We annotated loop anchor genes and regions with genomic variations, enhancers, silencers and transcription factors (TFs). (**C**) Chromatin loop diagram. (**D**) Genome browser in ChromLoops. (**E**) Search modules of ChromLoops. (**F**) Analysis functions included in ChromLoops. (**G**) Web implementation. The ChromLoops database was constructed with MySQL, Django and Vue tools.

served the high-quality data with ≥500 chromatin loops. Additionally, we quantified the interaction numbers, interaction distance and interaction frequency of each high-quality dataset.

Annotation resource collection and processing

To promote deeper research of chromatin loop regulatory mechanisms, we provided genetic and epigenetic annotations for each chromatin loop anchor, namely, regulatory elements (enhancers, super-enhancers and silencers), variations (SNPs from different sources, such as common SNPs, somatic SNPs, risk SNPs, drug SNPs, linkage disequilibrium (LD) SNPs, dsQTLs, eQTLs and meQTLs), TFs (TFBSs and motif changes), alternative splicing events, circRNAs, TWAS associations, chromosome open access information (ATAC and DHSs), gene expression information and DNA methylation information. BEDTools (25) was used to annotate corresponding information to the chromatin loop anchors. Interactive tables and bar plots were used to show details of the annotations.

(1) Enhancers: We downloaded and screened 5 796 148 enhancers and 1 046 168 super-enhancers from ENdb (26), EnhancerAtlas (27), FANTOM5 (28), GeneHancer (29), RAEdb (30), scEnhancer (31), SEA (32), dbSUPER (33) and VISTA (34). (2) Silencers: A total of 3 557 008 silencers were collected from SilencerDB (35). (3) Variations: ChromLoops also contains a wealth of SNP information, including common SNPs, disease SNPs and meQTLs. See Table 1 for details. (4) TFs: In total, 152 677 632 TFBSs were collected from ENCODE (36), ReMap2022 (37), Cistrome (38), ChIP-Atlas (39) and GTRD (40). In addition, 107 023 225 motif changes were collected from VARAdb (41) and OncoBase (42). (5) CircRNAs: We downloaded and merged 3 557 008 circRNAs from CircNet2 (43), CSCD2 (44), TSCD (45) and CircAtlas (46). (6) Alternative splicing: We collected 2 233 946 gene alternative splicing events from OncoSplicing (47). Moreover, we collected 7 941 822 cancer gene splicing events from CancerSplicingQTL (48). (7) TWAS associations: TWAS associations were downloaded from WebTWAS (49), including 1 088 735 gene–disease associations. (8) Chromosome open access information: ATAC annotations were collected from Cistrome (38) and TCGA (50), yielding 4 228 583 results. DHS annotations were collected from ENCODE (36), yielding 62 112 668 DHSs. (9) DNA methylation: DNA methylation levels of gene body and promoter were collected from the ASMdb database (51).

We first removed the repeated information of downloaded annotation resources and further converted the genomic locations of the resources from the hg19 human

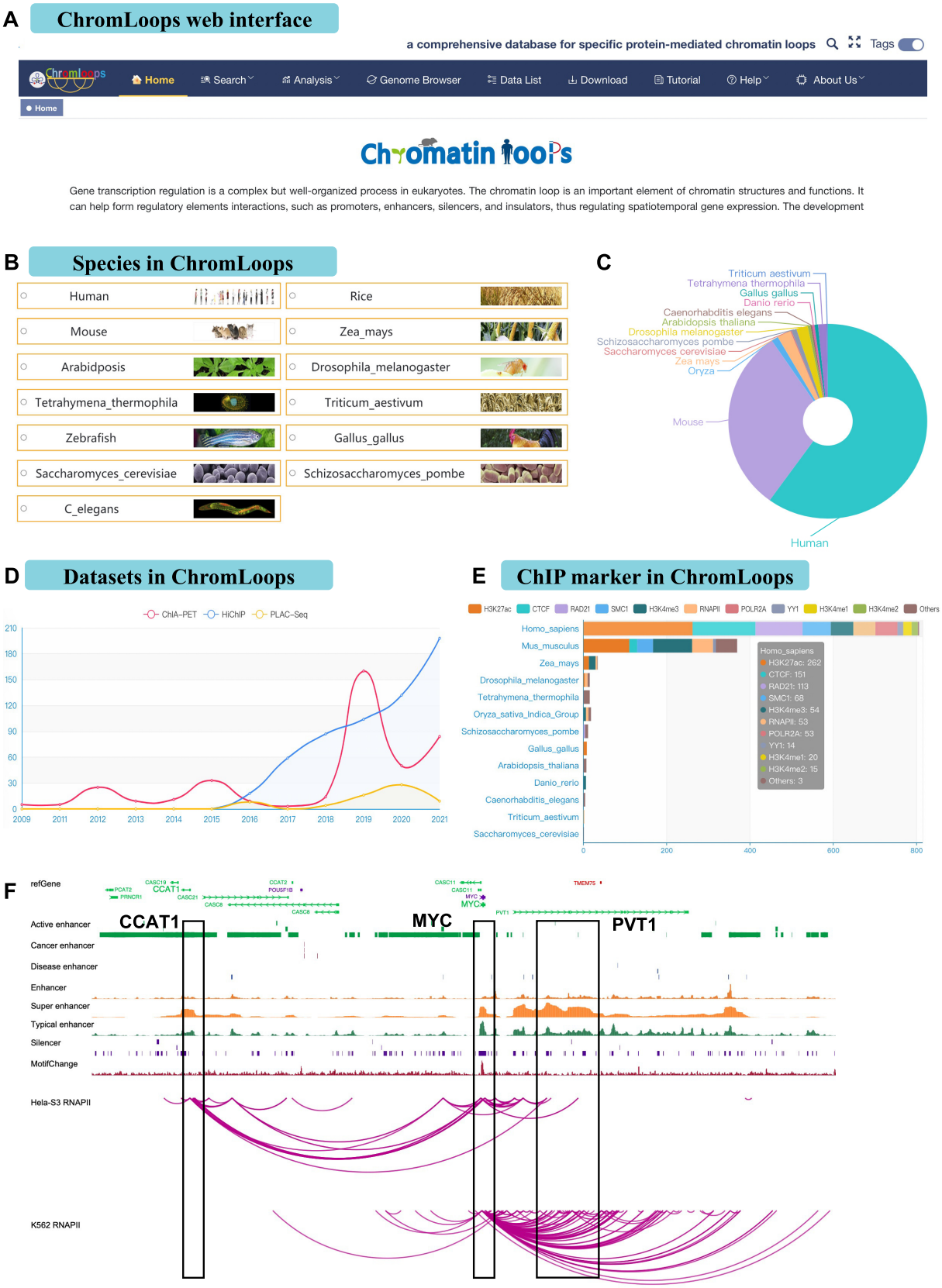

Figure 2. Overview of the ChromLoops database. (**A**) The homepage of the web interface. (**B**) Main species included in ChromLoops. (**C**) The proportion of ChIA-PET, HiChIP and PLAC-Seq data from various species in ChromLoops. (**D**) Statistics of ChIA-PET, HiChIP and PLAC-Seq data collected in ChromLoops per year. (**E**) Statistics of ChIP marker in ChromLoops. (**F**) An example of a genome browser screenshot around the *MYC* gene region in human HeLa and K562 cells (chr8:126542875–128709625, 2.2 Mb). Black boxes highlight *CCAT1, MYC* and *PVT1*.

Table 1. SNP resources

Variation type	Source	Number of SNPs
common SNPs	dbSNP(52)	12 675 975
somatic SNPs	Cosmic(53), TCGA (50), ICGC (54)	35 223 134
risk SNPs	ClinVar(55), GRASP v2.0 (56), GWAS Catalog (57), GWASdb v2.0 (58), DisGeNET (59), VARAdb (41)	2 986 804
drug SNPs	PharmGKB(60)	3449
eQTLs	GTEx(61), HaploReg (62), OncoBase (42), PancanQTL (63)	20 387 710
dsQTLs	OncoBase (42)	115 097
meQTLs	mQTLdb(64) and Pancan-meQTL (65)	52 555 932

assembly. For all subsequent analyses, the genomic intervals were converted to hg38 assembly coordinates using the UCSC Browser's standalone 'liftOver' tool and whole-genome pairwise alignment of the two assemblies (hg19Tohg38.over.chain). Then, we annotated the loop genes and anchor regions with BEDTools (25).

Genes with high-frequency chromatin loops in cancer

High-quality chromatin loops were retained through loop quality filtering, and only the loops with a PET count ≥ 3 were used for the following analysis. Subsequently, we manually screened and sorted all cancer-related samples and corresponding normal sample data. We screened out genes with high-frequency interactions in cancer and normal samples by calculating the gene long-range interaction frequency. Furthermore, we calculated the interaction frequency of each gene in all samples and obtained the genes with high-frequency interactions in the species.

RESULTS

Web interface

A user-friendly web interface (Figure 2A) is provided to allow users to query the database through multiple modules: (i) 'Loop gene search', a module that shows all interactions of the queried gene in individual or all samples; (ii) 'Loop region search', a module that lists all interactions of this chromosome region in individual or all samples; (iii) 'Loop SNP search', a module that displays all interactions overlapping with this SNP in individual or all samples; (iv) 'Circos search', a module that shows all interactions of one or more chromosome regions; (v) 'Browser search', a module that supplies interaction loops with arc mode in selected samples; (vi) 'Sample search', a retrieval module for online illustration of the data information and loop results in a specific sample; (vii) 'Genes with high-frequency chromatin interactions analysis', modules with detailed tables and heatmaps of high-frequency interaction genes in cancer samples and collected species; (viii) 'ChromLoops browser', a browser for browsing and searching chromatin loops and other functional elements; (ix) 'Download', a module that links to the download websites of the resources and (x) 'Tutorial', 'Help' and 'About Us', modules with detailed tutorials and documentation.

Genome browser and data visualization

The genome browser, which was developed using WashU Epigenome Browser, can easily display the chromatin interaction information of regions or genes in arc and heatmap modes. Users are able to select different samples and view the chromatin interaction results of a specified gene or region. In addition, the WashU browser provides 'gene plot', 'screenshot' and other functions. In particular, 'screenshot' gives users the ability to save and download the results in PDF/SVG format. Moreover, users can view the enhancer, silencer, SNP, and motif change information that we have prepared in the browser. To better introduce and demonstrate the application of the browser, we give an example: the *MYC* gene. *MYC* is one of the most widely investigated cancer-causing genes and is implicated in the formation, maintenance and progression of several different cancer types. Here, we demonstrate chromatin interactions around the *MYC* gene in human cervical cancer HeLa cells and leukemia K562 cells with RNAPII ChIP markers. Figure 2F shows the chromatin interactions of the *MYC* gene and information on enhancers, silencers and motif changes in the corresponding region. Studies have shown that *CCAT1* promotes cervical cancer cell proliferation and invasion (66). The *PVT1* region with enhancer-like activity interacts with the *MYC* gene (67). Consistent with the results, the *MYC* gene had chromatin interactions with *CCAT1* in HeLa cells and with *PVT1* in K562 cells (Figure 2F). Moreover, users can upload local data for viewing, and the browser allows the uploading of files in formats supported by the WashU browser, such as BEDPE and BED.

Functions of ChromLoops

Loop gene search. The 'Loop gene search' page provides a search function for the chromatin interactions of a gene in individual or all samples. The search results include (i) the basic information of the searched gene; (ii) information on the chromatin interactions mediated by the gene in individual or all samples; (iii) the gene annotation information, including that on variations, enhancers, silencers, TFs, circRNA, TWAS associations, chromosome open access and alternative splicing sites and (iv) the gene expression information in cancer and normal samples.

For instance, previous studies have shown that *ERBB2* is amplified or overexpressed in 20–30% of invasive breast carcinomas, and its amplification is an important predictive biomarker for identifying patients with breast cancer (68–70). To better illustrate the 'Loop gene search' function, we take the *ERBB2* gene in MCF-7 breast cancer cells with the RNAPII ChIP marker as an example. Figure 3A displays the search page information. The result details can be seen in Figure 3B. Figure 3C shows the basic information of the *ERBB2* gene, and Figure 3D shows

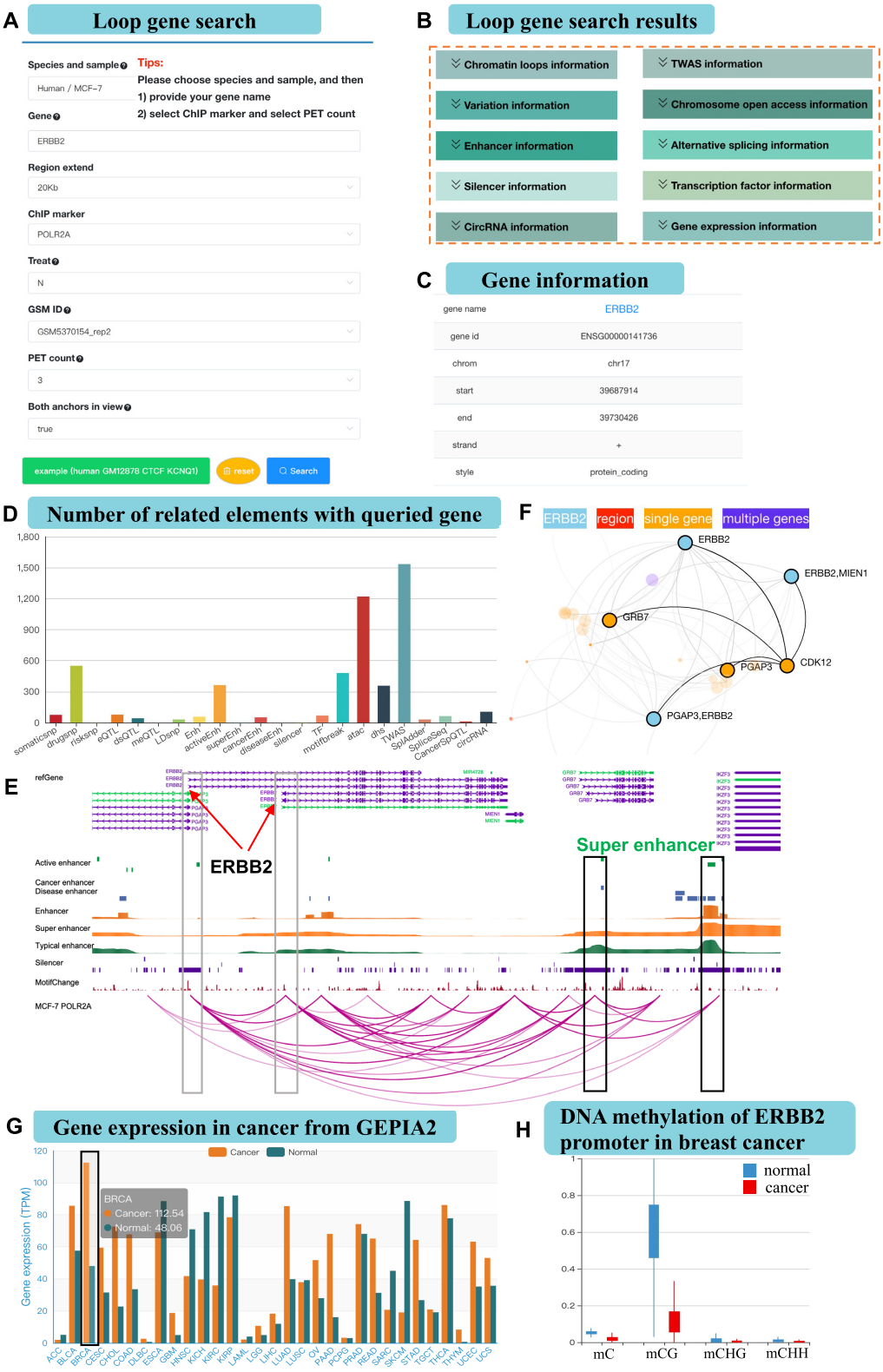

Figure 3. Loop gene analysis. (**A**) Loop gene search page. (**B**) The details of the loop gene search results. The results contain genomic variations, enhancers, silencers, transcription factors (TFs) and so on. (**C**) Basic information of the *ERBB2* gene. (**D**) Annotations covering the *ERBB2* gene. (**E**) Genome browser screenshot around the *ERBB2* gene. Red arrows indicate the *ERBB2* gene promotors. Black boxes highlight the super-enhancer regions. (**F**) The chromatin interaction network of the *ERBB2* gene in MCF-7 breast cancer cells. (**G**) The expression level of *ERBB2* from the GEPIA2 database. Black box highlights the gene expression level of *ERBB2* in breast cancer tissue (BRCA). (**H**) The DNA methylation level of *ERBB2* promoter from the ASMdb database.

the annotation results for *ERBB2,* such as variations, enhancers, silencers, and TFs. In the ChromLoops database, we found 558 drug-related SNPs, 58 enhancers, 2 super-enhancers and 52 cancer-related enhancers covering the *ERBB2* gene. We can also see that the *ERBB2* gene in MCF-7 remotely regulates genes such as *PAGP3* and *CDK12* and may be regulated by distal elements such as enhancers (56 distal-enhancers). These regulatory elements may play an important role in studying *ERBB2,* especially its long-range regulation. This is the first database integrated with such abundant regulatory elements and chromatin loops for the regulation of *ERBB2*. Similar results could be obtained for other genes of interest.

Figure 3E presents the interaction results for the *ERBB2* gene in the chr17:39677852–39761202 region in MCF-7 cells. Strong chromatin interactions between super-enhancers and *ERBB2* in MCF-7 cells were demonstrated in the results. The interaction network of *ERBB2* is reflected in Figure 3F. In addition, to obtain the gene expression and DNA methylation levels in cancer/normal cells and tissues, ChromLoops provides an association analysis with GEPIA2 (47), CCLE (71), ENCODE (cell line and primary cell) (36), GTEx (72) and ASMdb (51). The expression of the *ERBB2* gene in breast cancer (BRCA) tissue was significantly higher than that in normal tissue (Figure 3G). The DNA methylation level of the *ERBB2* promoter was significantly decreased in breast cancer (Figure 3H). It is consistent with the knowledge that genes with high methylation levels in the promoter have low expression levels.

Loop region search. Users can search for chromatin interactions of a region in individual or all samples on the 'Loop region search' page. The search results include genes and chromatin interaction information for the searched region. In addition, similar to the 'gene search' results, the 'region search' results also provide detailed statistical information and abundant annotation information (including variations, enhancers, silencers, TFs, circRNAs, TWAS associations, chromosome open access information, and alternative splicing information).

Loop SNP search. One of the main challenges in SNP studies is identifying the target genes of a SNP of interest. With ChromLoops, we can search the target genes of the SNP of interest for chromatin interactions. A query on the 'Loop SNP search' page provides the chromatin interactions of the queried SNP involved in individual or all samples and the annotation of the interacting targets. For example, SNP rs3851179 is an Alzheimer's disease-associated SNP located within a super-enhancer (73). The partial search results for SNP rs3851179 in a hippocampus sample are shown in Figure 4A and B. The results showed that multiple types of enhancers and motifs covered SNP rs3851179. The 'Loop SNP search' module can facilitate related research on the regulatory mechanism of the SNP of interest. In addition, the results page displays the chromatin interactions mediated by the super-enhancer containing rs3851179. Statistical results show that 18 distal genes have long-range chromatin interactions with SNP rs3851179. Among the results, it has been reported that the interaction between rs3851179 and the *PICALM* gene is associated with a risk of developing Alzheimer's disease (74) (Supplementary Figure S1).

It should be noted that we also provide the functions 'Loop gene search', 'Loop region search' and 'Loop SNP search' in all samples. On the corresponding search pages, users can select 'All samples' to search for a gene, region or SNP in all samples. For instance, Figure 4C shows the search results for SNP rs606231147 in all samples, including A673 cells, H9 cells, H1 cells and other cells or tissues. The results of a keyword (*SHH* gene) search showed rs606231147-related loops in all samples (Figure 4C). Researchers have demonstrated that SNP rs606231147 in the enhancer regulatory region ZRS (located in intron 5 of *LMBR1)* results in preaxial polydactyly by regulating ectopic expression of the gene *SHH (*75,76). Figure 4D shows an example of the interactions between rs606231147 and the *SHH* gene in H1 RAD21 data in ChromLoops. These search results can help reveal the target genes and regulatory mechanisms of disease-associated SNPs.

Loop Circos search. On the 'Loop Circos search' page, users are able to search for chromatin interactions involving one or more regions in a specific sample. This function allows users to easily view the long-range chromatin regulations between multiple regions. For example, in the GM12878 cell line, multiple interchromosomal interactions of the enhancer-promoter (chr1, chr4, chr7 and chr15) were found, and these interaction genes including B-cell-specific highly expressed genes (77). The chromatin interactions between chr1:46.51–56.75 Mb, chr4: 33–44.4 Mb, chr 7:70.4–81.6 Mb, and chr15: 58.6–70 Mb can be seen in Supplementary Figure S2.

Sample search. A query on the 'Sample search' page displays the basic sample information and statistical information of the results. For example, K562 cells are the first established human immortalized myeloid leukemia cell line, which is highly valuable for leukemia research. We take the search of the K562 cell line as an example. Details for the searched example of human K562 cells with RNAPII include (i) basic information, such as NCBI GEO ID, sample description, and ChIP marker information (Supplementary Figure S3A); (ii) statistical information for the interaction results (Supplementary Figure S3B); (iii) a loop span distribution density plot (Supplementary Figure S3C); (iv) a Circos diagram of genome-wide chromatin interactions of K562 cells (Supplementary Figure S4A); (v) a detailed list of chromatin loops, including the interaction anchor region and covered genes (Supplementary Figure S4B) and (vi) the genome browser interaction information for K562 cells (Supplementary Figure S4C).

Genes with high-frequency chromatin interactions in cancer. After analyzing the data of all samples, we further carried out the following work: (i) to explore the characteristics of chromatin interactions in cancer, we screened the datasets related to cancer and then identified the cancer-specific genes with high-frequency chromatin interactions, and the results are shown in Figure 5. Figure 5A and B demonstrates the 'high-frequency interaction gene in cancer' search page and the heatmap of the distribution of

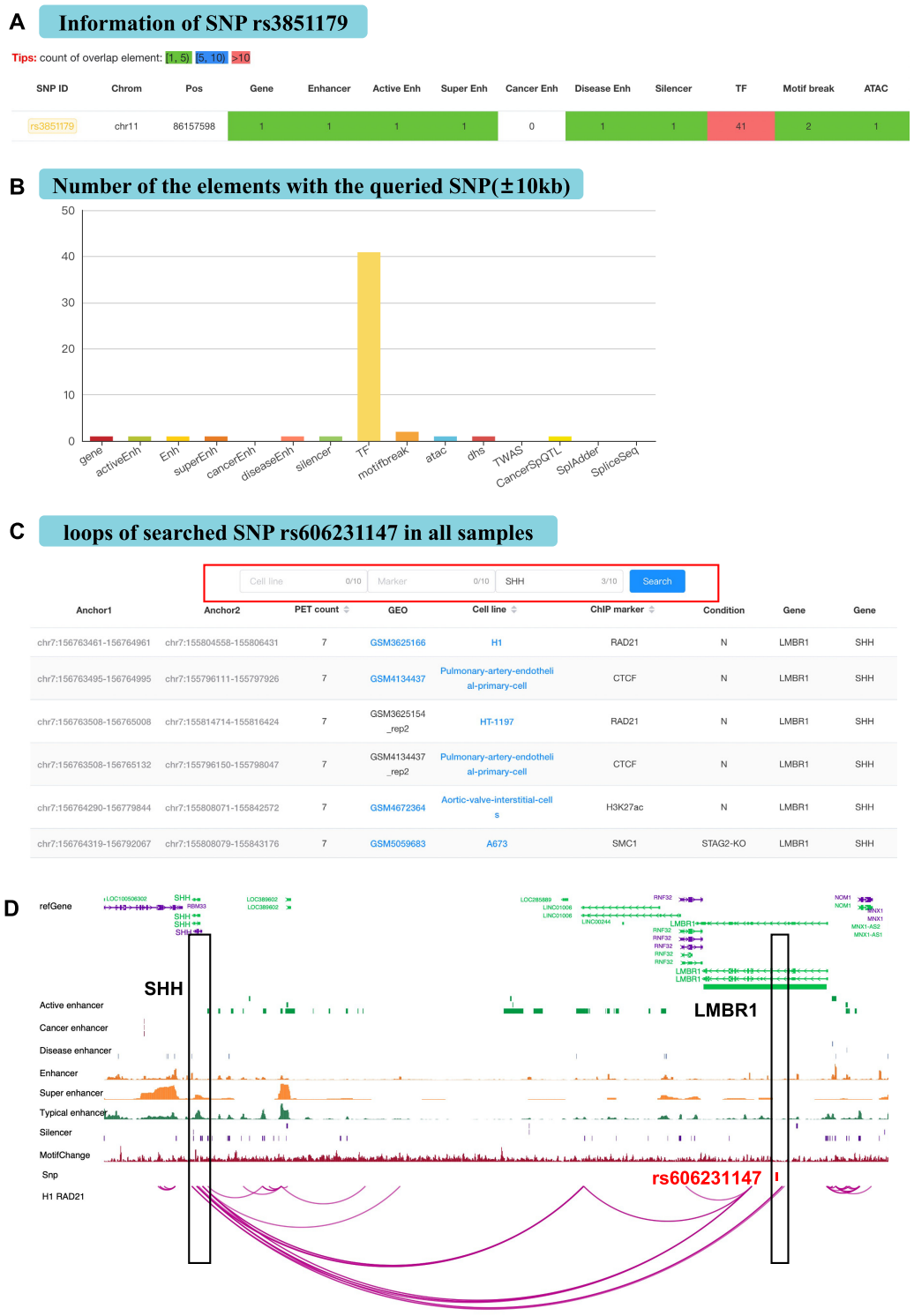

Figure 4. Loop SNP analysis. (**A**) The basic information for SNP rs3851179. (**B**) The distribution of genes, enhancers, silencers, etc., at SNP rs3851179. (**C**) Chromatin interaction loop information for SNP rs606231147 in all samples. The red box highlights the keyword search function, including the cell line, ChIP marker or gene. (**D**) Genome browser screenshot between SNP rs606231147 and the *SHH* gene in H1 cells with RAD21 data. The right black box highlights SNP rs606231147, and the left black box highlights the *SHH* gene.

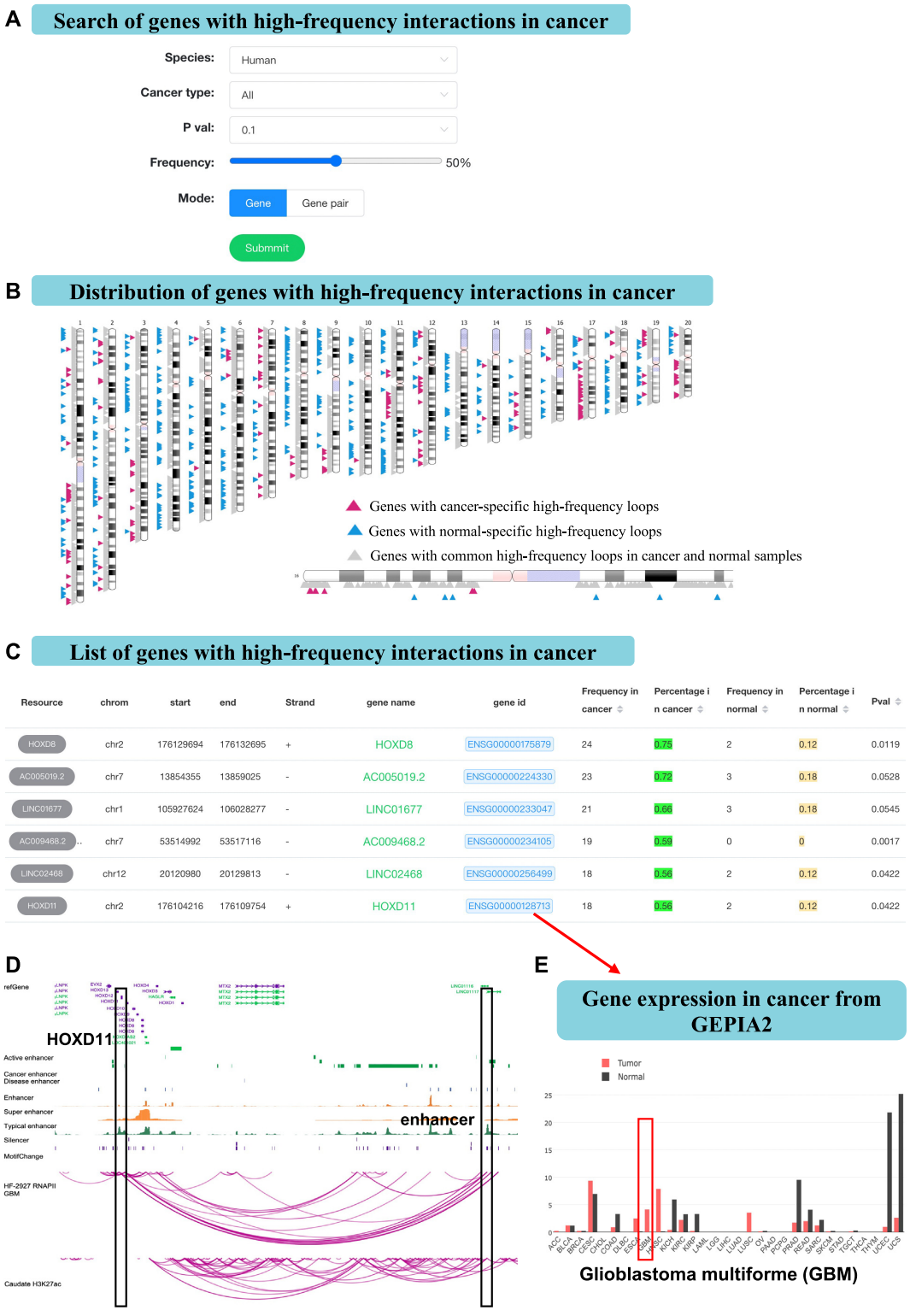

Figure 5. Genes with high-frequency chromatin interactions in cancer. (**A**) The search page for genes with high-frequency interactions. (**B**) The distribution of genes with high-frequency chromatin interactions on chromosomes in cancer and normal samples. (**C**) A list of genes with high-frequency interactions in brain cancer and normal samples. (**D**) Example of the *HOXD11* gene with cancer-specific high-frequency chromatin interactions. (**E**) The expression of gene *HOXD11* with the high-frequency interactions in cancer.

genes on chromosomes. Figure 5C shows the detailed list of high-frequency interaction genes in cancer and normal samples. Previous studies demonstrated that abnormal expression of *HOXD11* promotes the malignant behavior of glioma cells (78,79). Our results indicate that *HOXD11* is a specific high-frequency interaction gene in glioma (Figure 5C), and there are obvious chromatin interactions between an enhancer enrichment region and the gene *HOXD11* in glioma (Figure 5D). This 3D specific interaction may be one of the reasons for the high expression of *HOXD11* in glioma. Figure 5E shows the expression of the *HOXD11* gene in normal and cancer samples. (ii) We also analyzed and obtained the high-frequency interaction genes in different species. These results for high-frequency interaction genes in cancer and in species will provide a new perspective for research on long-range chromatin interaction regulation, species specificity and interspecies comparison.

Download and others. The 'Download' module provides the statistics of enhancers, silencers, variations and other annotation resources, as well as the download function for chromatin loops. With the 'Contact Us' module, users can contact us and submit data links not contained in our database. We will perform data quality control and analysis and then update the results in the database as soon as possible.

DISCUSSION AND FUTURE DIRECTIONS

The first database for specific protein-mediated chromatin loops

At present, there are several browsers and databases for 3D genomes mainly based on Hi-C data. Browsers such as the Hi-C data browser (80), WashU epigenome browser (23) and 3D genome browser (81) have provided heatmap visualization functions for 3D genomic data, especially Hi-C data. The Delta browser (82) can visualize Hi-C data as well as the 3D physical architecture of genomes. Databases such as the 3DGD (83) and 3DIV (84) databases mainly provide topologically associated domain (TAD) analysis and visualization results based on Hi-C data. 3DSNP (85) enables linking noncoding SNPs to their 3D interacting genes based on human Hi-C data. These browsers and databases have provided a good visualization and analysis platform for 3D genomics research in recent years. Although the 3DSNP database presents a few chromatin loops (∼75 362) based on human Hi-C data, most of these tools focus on TAD-level analysis and visualization. Specific protein-mediated chromatin loops have higher resolution and are better suited for functional annotation to further study chromatin regulatory mechanisms.

Here, we analyzed and obtained high-resolution results for specific protein-mediated chromatin loops based on all publicly available ChIA-PET, HiChIP, and PLAC-Seq data. ChromLoops, a database for specific protein-mediated chromatin loops in diverse organisms, was constructed; it also integrates search, analysis, visualization and download functions. Crucially, we further annotated the loop genes and regions with abundant functional annotation information, such as regulatory elements, variations and TFs. More importantly, we identified high-frequency chromatin interactions in individual and all cancer samples. It will be helpful to explore the characteristics of specific long-range chromatin interactions in relation to cancer and provide a new theoretical basis for the regulation of key cancer genes. ChromLoops is a comprehensive and powerful database for promoting research on 3D genomic functions, long-range chromatin interaction regulation and gene transcription regulation in association with cellular function and disease.

Future directions

In future versions, ChromLoops will continue to be updated as follows: (i) We will further collect and analyze new protein-mediated chromatin interaction data (including ChIA-PET, HiChIP, PLAC-Seq and HiCuT data (86)) from different sources and species. HiCuT is a new technique for detecting protein-mediated chromatin interactions that has just been published this year (86). At present, there are only 12 datasets. (ii) We will provide additional functional annotation resources and online functions in the database based on user feedback. We will keep ChromLoops up to date to ensure its value as a user-friendly database. We expect that ChromLoops will contribute to research on chromatin interactions in relation to cellular function.

DATA AVAILABILITY

ChromLoops is a database with online and open access, available at https://3dgenomics.hzau.edu.cn. Any constructive comments and suggestions are welcome to send to Prof. Guoliang Li at email address guoliang.li@mail.hzau.edu.cn.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Mr. Hao Liu from the National Key Laboratory of Crop Genetic Improvement for essential help in running the high-throughput computing clusters. We thank the group members for providing feedback on the database. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

FUNDING

National Key Research and Development Program of China [2021YFC2701201]; National Natural Science Foundation of China [31970590]. Funding for open access charge: National Key Research and Development Program of China [2021YFC2701201]; National Natural Science Foundation of China [31970590].

Conflict of interest statement. None declared.

REFERENCES

1. Lettice,L.A., Heaney,S.J.H., Purdie,L.A., Li,L., de Beer,P., Oostra,B.A., Goode,D., Elgar,G., Hill,R.E. and de Graaff,E. (2003) A long-range shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.*, **12**, 1725–1735.
2. Visel,A., Rubin,E.M. and Pennacchio,L.A. (2009) Genomic views of distant-acting enhancers. *Nature*, **461**, 199–205.
3. Uslu,V.V., Petretich,M., Ruf,S., Langenfeld,K., Fonseca,N.A., Marioni,J.C. and Spitz,F. (2014) Long-range enhancers regulating myc expression are required for normal facial morphogenesis. *Nat. Genet.*, **46**, 753–758.
4. Pang,B. and Snyder,M.P. (2020) Systematic identification of silencers in human cells. *Nat. Genet.*, **52**, 254–263.
5. Ngan,C.Y., Wong,C.H., Tjong,H., Wang,W., Goldfeder,R.L., Choi,C., He,H., Gong,L., Lin,J., Urban,B. *et al.* (2020) Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat. Genet.*, **52**, 264–272.
6. Doni Jayavelu,N., Jajodia,A., Mishra,A. and Hawkins,R.D. (2020) Candidate silencer elements for the human and mouse genomes. *Nat. Commun.*, **11**, 1061.
7. Ruiz-Velasco,M., Kumar,M., Lai,M.C., Bhat,P., Solis-Pinson,A.B., Reyes,A., Kleinsorg,S., Noh,K.-M., Gibson,T.J. and Zaugg,J.B. (2017) CTCF-Mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. *Cell Syst.*, **5**, 628–637.
8. Grubert,F., Srivas,R., Spacek,D.V., Kasowski,M., Ruiz-Velasco,M., Sinnott-Armstrong,N., Greenside,P., Narasimha,A., Liu,Q., Geller,B. *et al.* (2020) Landscape of cohesin-mediated chromatin loops in the human genome. *Nature*, **583**, 737–743.
9. Alharbi,A.B., Schmitz,U., Bailey,C.G. and Rasko,J.E.J. (2021) CTCF as a regulator of alternative splicing: new tricks for an old player. *Nucleic Acids Res.*, **49**, 7825–7838.
10. Li,G., Ruan,X., Auerbach,R.K., Sandhu,K.S., Zheng,M., Wang,P., Poh,H.M., Goh,Y., Lim,J., Zhang,J. *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, **148**, 84–98.
11. Tang,Z., Luo,O.J., Li,X., Zheng,M., Zhu,J.J., Szalaj,P., Trzaskoma,P., Magalska,A., Wlodarczyk,J., Ruszczycki,B. *et al.* (2015) CTCF-Mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
12. Giambartolomei,C., Seo,J.-H., Schwarz,T., Freund,M.K., Johnson,R.D., Spisak,S., Baca,S.C., Gusev,A., Mancuso,N., Pasaniuc,B. *et al.* (2021) H3K27ac HiChIP in prostate cell lines identifies risk genes for prostate cancer susceptibility. *Am. J. Hum. Genet.*, **108**, 2284–2300.
13. Okuyama,K., Strid,T., Kuruvilla,J., Somasundaram,R., Cristobal,S., Smith,E., Prasad,M., Fioretos,T., Lilljebjorn,H., Soneji,S. *et al.* (2019) PAX5 is part of a functional transcription factor network targeted in lymphoid leukemia. *PLoS Genet.*, **15**, e1008280.
14. Javierre,B.M., Burren,O.S., Wilder,S.P., Kreuzhuber,R., Hill,S.M., Sewitz,S., Cairns,J., Wingett,S.W., Várnai,C., Thiecke,M.J. *et al.* (2016) Lineage-Specific genome architecture links enhancers and Non-coding disease variants to target gene promoters. *Cell*, **167**, 1369–1384.
15. Mumbach,M.R., Satpathy,A.T., Boyle,E.A., Dai,C., Gowen,B.G., Cho,S.W., Nguyen,M.L., Rubin,A.J., Granja,J.M., Kazane,K.R. *et al.* (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.*, **49**, 1602–1612.
16. Li,Z., Chen,J., Yu,H., He,L., Xu,Y., Zhang,D., Yi,Q., Li,C., Li,X., Shen,J. *et al.* (2017) Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.*, **49**, 1576–1583.
17. Soldner,F., Stelzer,Y., Shivalila,C.S., Abraham,B.J., Latourelle,J.C., Barrasa,M.I., Goldmann,J., Myers,R.H., Young,R.A. and Jaenisch,R. (2016) Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression. *Nature*, **533**, 95–99.
18. Lin,X., Liu,Y., Liu,S., Zhu,X., Wu,L., Zhu,Y., Zhao,D., Xu,X., Chemparathy,A., Wang,H. *et al.* Nested epistasis enhancer networks for robust genome regulation. *Science*, **377**, 1077–1085.
19. Sun,F., Chronis,C., Kronenberg,M., Chen,X.F., Su,T., Lay,F.D., Plath,K., Kurdistani,S.K. and Carey,M.F. (2019) Promoter-Enhancer communication occurs primarily within insulated neighborhoods. *Mol. Cell*, **73**, 250–263.
20. Nora,E.P., Goloborodko,A., Valton,A.L., Gibcus,J.H., Uebersohn,A., Abdennur,N., Dekker,J., Mirny,L.A. and Bruneau,B.G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, **169**, 930–944.
21. Belokopytova,P.S., Nuriddinov,M.A., Mozheiko,E.A., Fishman,D. and Fishman,V. (2020) Quantitative prediction of enhancer-promoter interactions. *Genome Res.*, **30**, 72–84.
22. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
23. Li,D., Purushotham,D., Harrison,J.K., Hsu,S., Zhuo,X., Fan,C., Liu,S., Xu,V., Chen,S., Xu,J. *et al.* (2022) WashU epigenome browser update 2022. *Nucleic Acids Res.*, **47**, W158–W165.
24. Li,G., Sun,T., Chang,H., Cai,L., Hong,P. and Zhou,Q. (2019) Chromatin interaction analysis with updated ChIA-PET tool (V3). *Genes (Basel)*, **10**, 554.
25. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
26. Bai,X., Shi,S., Ai,B., Jiang,Y., Liu,Y., Han,X., Xu,M., Pan,Q., Wang,F., Wang,Q. *et al.* (2020) ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res.*, **48**, D51–D57.
27. Gao,T. and Qian,J. (2020) EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.*, **48**, D58–D64.
28. Kawaji,H., Kasukawa,T., Forrest,A., Carninci,P. and Hayashizaki,Y. (2017) The FANTOM5 collection, a data series underpinning mammalian transcriptome atlases in diverse cell types. *Sci. Data*, **4**, 170113.
29. Fishilevich,S., Nudel,R., Rappaport,N., Hadar,R., Plaschkes,I., Iny Stein,T., Rosen,N., Kohn,A., Twik,M., Safran,M. *et al.* (2017) GeneHancer: genome-wide integration of enhancers and target genes in genecards. *Database (Oxford)*, **2017**, bax028.
30. Cai,Z., Cui,Y., Tan,Z., Zhang,G., Tan,Z., Zhang,X. and Peng,Y. (2019) RAEdb: a database of enhancers identified by high-throughput reporter assays. *Database (Oxford)*, **2019**, bay140.
31. Gao,T., Zheng,Z., Pan,Y., Zhu,C., Wei,F., Yuan,J., Sun,R., Fang,S., Wang,N., Zhou,Y. *et al.* (2022) scEnhancer: a single-cell enhancer resource with annotation across hundreds of tissue/cell types in three species. *Nucleic Acids Res.*, **50**, D371–D379.
32. Chen,C., Zhou,D., Gu,Y., Wang,C., Zhang,M., Lin,X., Xing,J., Wang,H. and Zhang,Y. (2020) SEA version 3.0: a comprehensive extension and update of the super-enhancer archive. *Nucleic Acids Res.*, **48**, D198–D203.
33. Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.
34. Visel,A., Minovitsky,S., Dubchak,I. and Pennacchio,L.A. (2007) VISTA enhancer Browser–a database of tissue-specific human enhancers. *Nucleic Acids Res.*, **35**, D88–D92.
35. Zeng,W., Chen,S., Cui,X., Chen,X., Gao,Z. and Jiang,R. (2021) SilencerDB: a comprehensive database of silencers. *Nucleic Acids Res.*, **49**, D221–D228.
36. Luo,Y., Hitz,B.C., Gabdank,I., Hilton,J.A., Kagda,M.S., Lam,B., Myers,Z., Sud,P., Jou,J., Lin,K. *et al.* (2020) New developments on the encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Res.*, **48**, D882–D889.
37. Hammal,F., de Langen,P., Bergon,A., Lopez,F. and Ballester,B. (2022) ReMap 2022: a database of human, mouse, drosophila and arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res.*, **50**, D316–D325.
38. Wang,Q., Huang,J., Sun,H., Liu,J., Wang,J., Wang,Q., Qin,Q., Mei,S., Zhao,C., Yang,X. *et al.* (2014) CR cistrome: a chip-Seq database for chromatin regulators and histone modification linkages in human and mouse. *Nucleic Acids Res.*, **42**, D450–D458.
39. Zou,Z., Ohta,T., Miura,F. and Oki,S. (2022) ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating chip-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res.*, **50**, W175–W182.

40. Kolmykov,S., Yevshin,I., Kulyashov,M., Sharipov,R., Kondrakhin,Y., Makeev,V.J., Kulakovskiy,I.V., Kel,A. and Kolpakov,F. (2021) GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.*, **49**, D104–D111.
41. Pan,Q., Liu,Y.J., Bai,X.F., Han,X.L., Jiang,Y., Ai,B., Shi,S.S., Wang,F., Xu,M.C., Wang,Y.Z. *et al.* (2021) VARAdb: a comprehensive variation annotation database for human. *Nucleic Acids Res.*, **49**, D1431–D1444.
42. Li,X., Shi,L., Wang,Y., Zhong,J., Zhao,X., Teng,H., Shi,X., Yang,H., Ruan,S., Li,M. *et al.* (2019) OncoBase: a platform for decoding regulatory somatic mutations in human cancers. *Nucleic Acids Res.*, **47**, D1044–D1055.
43. Chen,Y., Yao,L., Tang,Y., Jhong,J.H., Wan,J., Chang,J., Cui,S., Luo,Y., Cai,X., Li,W. *et al.* (2022) CircNet 2.0: an updated database for exploring circular RNA regulatory networks in cancers. *Nucleic Acids Res.*, **50**, D93–D101.
44. Feng,J., Chen,W., Dong,X., Wang,J., Mei,X., Deng,J., Yang,S., Zhuo,C., Huang,X., Shao,L. *et al.* (2022) CSCD2: an integrated interactional database of cancer-specific circular RNAs. *Nucleic Acids Res.*, **50**, D1179–D1183.
45. Xia,S., Feng,J., Lei,L., Hu,J., Xia,L., Wang,J., Xiang,Y., Liu,L., Zhong,S., Han,L. *et al.* (2017) Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform*, **18**, 984–992.
46. Wu,W., Ji,P. and Zhao,F. (2020) CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. *Genome Biol.*, **21**, 101.
47. Zhang,Y., Yao,X., Zhou,H., Wu,X., Tian,J., Zeng,J., Yan,L., Duan,C., Liu,H., Li,H. *et al.* (2022) OncoSplicing: an updated database for clinically relevant alternative splicing in 33 human cancers. *Nucleic Acids Res.*, **50**, D1340–D1347.
48. Tian,J., Wang,Z., Mei,S., Yang,N., Yang,Y., Ke,J., Zhu,Y., Gong,Y., Zou,D., Peng,X. *et al.* (2019) CancerSplicingQTL: a database for genome-wide identification of splicing QTLs in human cancer. *Nucleic Acids Res.*, **47**, D909–D916.
49. Cao,C., Wang,J., Kwok,D., Cui,F., Zhang,Z., Zhao,D., Li,M.J. and Zou,Q. (2022) webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. *Nucleic Acids Res.*, **50**, D1123–D1130.
50. Cancer Genome Atlas Research, N., Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The cancer genome atlas pan-cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
51. Zhou,Q., Guan,P., Zhu,Z., Cheng,S., Zhou,C., Wang,H., Xu,Q., Sung,W.K. and Li,G. (2022) ASMdb: a comprehensive database for allele-specific DNA methylation in diverse organisms. *Nucleic Acids Res.*, **50**, D60–D71.
52. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic. Acids. Res.*, **29**, 308–311.
53. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E. *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–D947.
54. International Cancer Genome, C., Hudson,T.J., Anderson,W., Artez,A., Barker,A.D., Bell,C., Bernabe,R.R., Bhan,M.K., Calvo,F., Eerola,I. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
55. Landrum,M.J., Chitipiralla,S., Brown,G.R., Chen,C., Gu,B., Hart,J., Hoffman,D., Jang,W., Kaur,K., Liu,C. *et al.* (2020) ClinVar: improvements to accessing data. *Nucleic Acids Res.*, **48**, D835–D844.
56. Eicher,J.D., Landowski,C., Stackhouse,B., Sloan,A., Chen,W., Jensen,N., Lien,J.P., Leslie,R. and Johnson,A.D. (2015) GRASP v2.0: an update on the genome-wide repository of associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.
57. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2019) The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
58. Li,M.J., Liu,Z., Wang,P., Wong,M.P., Nelson,M.R., Kocher,J.P., Yeager,M., Sham,P.C., Chanock,S.J., Xia,Z. *et al.* (2016) GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **44**, D869–D876.
59. Pinero,J., Bravo,A., Queralt-Rosinach,N., Gutierrez-Sacristan,A., Deu-Pons,J., Centeno,E., Garcia-Garcia,J., Sanz,F. and Furlong,L.I. (2017) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833–D839.
60. Barbarino,J.M., Whirl-Carrillo,M., Altman,R.B. and Klein,T.E. (2018) PharmGKB: a worldwide resource for pharmacogenomic information. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **10**, e1417.
61. Lonsdale,J., Thomas,J., Salvatore,M., Phillips,R., Lo,E., Shad,S., Hasz,R., Walters,G., Garcia,F., Young,N. *et al.* (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
62. Ward,L.D. and Kellis,M. (2016) HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.*, **44**, D877–D881.
63. Gong,J., Mei,S., Liu,C., Xiang,Y., Ye,Y., Zhang,Z., Feng,J., Liu,R., Diao,L., Guo,A.Y. *et al.* (2018) PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.*, **46**, D971–D976.
64. Gaunt,T.R., Shihab,H.A., Hemani,G., Min,J.L., Woodward,G., Lyttleton,O., Zheng,J., Duggirala,A., McArdle,W.L., Ho,K. *et al.* (2016) Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.*, **17**, 61.
65. Gong,J., Wan,H., Mei,S., Ruan,H., Zhang,Z., Liu,C., Guo,A.Y., Diao,L., Miao,X. and Han,L. (2019) Pancan-meQTL: a database to systematically evaluate the effects of genetic variants on methylation in human cancer. *Nucleic Acids Res.*, **47**, D1066–D1072.
66. Shen,H., Wang,L., Xiong,J., Ren,C., Gao,C., Ding,W., Zhu,D., Ma,D. and Wang,H. (2019) Long non-coding RNA CCAT1 promotes cervical cancer cell proliferation and invasion by regulating the miR-181a-5p/MMP14 axis. *Cell Cycle*, **18**, 1110–1121.
67. Shigeyasu,K., Toden,S., Ozawa,T., Matsuyama,T., Nagasaka,T., Ishikawa,T., Sahoo,D., Ghosh,P., Uetake,H., Fujiwara,T. *et al.* (2020) The PVT1 lncRNA is a novel epigenetic enhancer of MYC, and a promising risk-stratification biomarker in colorectal cancer. *Mol. Cancer*, **19**, 155.
68. Yi,Z., Rong,G., Guan,Y., Li,J., Chang,L., Li,H., Liu,B., Wang,W., Guan,X., Ouyang,Q. *et al.* (2020) Molecular landscape and efficacy of HER2-targeted therapy in patients with HER2-mutated metastatic breast cancer. *NPJ Breast Cancer*, **6**, 59.
69. Owens,M.A., Horten,B.C. and Da Silva,M.M. (2004) HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clin. Breast Cancer*, **5**, 63–69.
70. Yaziji,H., Goldstein,L.C., Barry,T.S., Werling,R., Hwang,H., Ellis,G.K., Gralow,J.R., Livingston,R.B. and Gown,A.M. (2004) HER-2 testing in breast cancer using parallel tissue-based methods. *JAMA*, **291**, 1972–1977.
71. Barretina,J., Caponigro,G., Stransky,N., Venkatesan,K., Margolin,A.A., Kim,S., Wilson,C.J., Lehar,J., Kryukov,G.V., Sonkin,D. *et al.* (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
72. Consortium, G.T. (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
73. Schmitt,A.D., Hu,M., Jung,I., Xu,Z., Qiu,Y., Tan,C.L., Li,Y., Lin,S., Lin,Y., Barr,C.L. *et al.* (2016) A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.*, **17**, 2042–2059.
74. Baig,S., Joseph,S.A., Tayler,H., Abraham,R., Owen,M.J., Williams,J., Kehoe,P.G. and Love,S. (2010) Distribution and expression of picalm in alzheimer disease. *J. Neuropathol. Exp. Neurol.*, **69**, 1071–1077.
75. Norbnop,P., Srichomthong,C., Suphapeetiporn,K. and Shotelersuk,V. (2014) ZRS 406A>G mutation in patients with tibial hypoplasia, polydactyly and triphalangeal first fingers. *J. Hum. Genet.*, **59**, 467–470.
76. Wieczorek,D., Pawlik,B., Li,Y., Akarsu,N.A., Caliebe,A., May,K.J., Schweiger,B., Vargas,F.R., Balci,S., Gillessen-Kaesbach,G. *et al.* (2010) A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes werner mesomelic syndrome (WMS) while complete ZRS duplications underlie haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum. Mutat.*, **31**, 81–89.

77. Deshpande,A.S., Ulahannan,N., Pendleton,M., Dai,X., Ly,L., Behr,J.M., Schwenk,S., Liao,W., Augello,M.A., Tyer,C. *et al.* (2022) Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol*, **40**, 1488–1499.
78. Giannakis,M., Hodis,E., Jasmine Mu,X., Yamauchi,M., Rosenbluh,J., Cibulskis,K., Saksena,G., Lawrence,M.S., Qian,Z.R., Nishihara,R. *et al.* (2014) RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat. Genet.*, **46**, 1264–1266.
79. Fang,L., Ford-Roshon,D., Russo,M., O'Brien,C., Xiong,X., Gurjao,C., Grandclaudon,M., Raghavan,S., Corsello,S.M., Carr,S.A. *et al.* (2022) RNF43 G659fs is an oncogenic colorectal cancer mutation and sensitizes tumor cells to PI3K/mTOR inhibition. *Nat. Commun.*, **13**, 3181.
80. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
81. Wang,Y., Song,F., Zhang,B., Zhang,L., Xu,J., Kuang,D., Li,D., Choudhary,M.N.K., Li,Y., Hu,M. *et al.* (2018) The 3D genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.*, **19**, 151.
82. Tang,B., Li,F., Li,J., Zhao,W. and Zhang,Z. (2018) Delta: a new web-based 3D genome visualization and analysis platform. *Bioinformatics*, **34**, 1409–1410.
83. Li,C., Dong,X., Fan,H., Wang,C., Ding,G. and Li,Y. (2014) The 3DGD: a database of genome 3D structure. *Bioinformatics*, **30**, 1640–1642.
84. Kim,K., Jang,I., Kim,M., Choi,J., Kim,M.S., Lee,B. and Jung,I. (2021) 3DIV update for 2021: a comprehensive resource of 3D genome and 3D cancer genome. *Nucleic Acids Res.*, **49**, D38–D46.
85. Quan,C., Ping,J., Lu,H., Zhou,G. and Lu,Y. (2022) 3DSNP 2.0: update and expansion of the noncoding genomic variant annotation database. *Nucleic Acids Res.*, **50**, D950–D955.
86. Sati,S., Jones,P., Kim,H.S., Zhou,L.A., Rapp-Reyes,E. and Leung,T.H. (2022) HiCuT: an efficient and low input method to identify protein-directed chromatin interactions. *PLoS Genet.*, **18**, e1010121.