Published online 29 October 2018 *Nucleic Acids Research, 2019, Vol. 47, Database issue* ***D235–D243***
doi: 10.1093/nar/gky1025

SEdb: a comprehensive human super-enhancer database

Yong Jiang†, Fengcui Qian†, Xuefeng Bai†, Yuejuan Liu, Qiuyu Wang, Bo Ai, Xiaole Han, Shanshan Shi, Jian Zhang, Xuecang Li, Zhidong Tang, Qi Pan, Yuezhu Wang, Fan Wang and Chunquan Li*

School of Medical Informatics, Daqing Campus, Harbin Medical University, Daqing 163319, China

Received August 10, 2018; Revised October 12, 2018; Editorial Decision October 15, 2018; Accepted October 17, 2018

ABSTRACT

Super-enhancers are important for controlling and defining the expression of cell-specific genes. With research on human disease and biological processes, human H3K27ac ChIP-seq datasets are accumulating rapidly, creating the urgent need to collect and process these data comprehensively and efficiently. More importantly, many studies showed that super-enhancer-associated single nucleotide polymorphisms (SNPs) and transcription factors (TFs) strongly influence human disease and biological processes. Here, we developed a comprehensive human super-enhancer database (SEdb, http://www.licpathway.net/sedb) that aimed to provide a large number of available resources on human super-enhancers. The database was annotated with potential functions of super-enhancers in the gene regulation. The current version of SEdb documented a total of 331 601 super-enhancers from 542 samples. Especially, unlike existing super-enhancer databases, we manually curated and classified 410 available H3K27ac samples from >2000 ChIP-seq samples from NCBI GEO/SRA. Furthermore, SEdb provides detailed genetic and epigenetic annotation information on super-enhancers. Information includes common SNPs, motif changes, expression quantitative trait locus (eQTL), risk SNPs, transcription factor binding sites (TFBSs), CRISPR/Cas9 target sites and Dnase I hypersensitivity sites (DHSs) for in-depth analyses of super-enhancers. SEdb will help elucidate super-enhancer-related functions and find potential biological effects.

INTRODUCTION

Super-enhancers are a large cluster of transcriptionally active enhancers enriched in enhancer-associated chromatin characteristics (1). Compared to typical enhancers, super-enhancers are larger, exhibit higher transcription factor density (2,3), and are frequently associated with key lineage-specific genes that control cell state and differentiation in somatic cells (4). In cancer cells, super-enhancers drive the expression of critical oncogenes such as CACNA1H (5), LMO1 (6), RARA (7) and TAL1 (8), suggesting that cancer cells generate super-enhancers at oncogenes that are involved in tumor pathogenesis (9). Mack et al. discovered 15 important super-enhancers in ependymoma. In the absence of any of 15 super-enhancers, the survival rate of ependymoma cancer cells was reduced by at least 50% (5). In neuroblastomas, super-enhancer-associated TFs networks may mediate lineage differentiation of normal development, leading to epigenetic regulation of neuroblastoma and internal heterogeneity of tumors (10). A large number of disease-associated sequence variations are preferentially enriched in super-enhancers of disease-related cell types (11). For example, disease-associated SNPs for autoimmune diseases such as rheumatoid arthritis are often located in super-enhancer regions (12). The causal SNP rs539846, which is localized to a super-enhancer in intron 3 of B cell lymphoma 2-modifying factor, influences chronic lymphocytic leukemia susceptibility through altering a conserved RELA-binding motif (13). Oldridge *et al.* found that carcinogenic dependence in tumor cells is due to the difference in polymorphisms between super-enhancer elements in the first intron of LMO1, which binds and directly regulates LMO1 expression (6). Together, these studies demonstrate the importance of super-enhancers in addressing key issues associated with cancer biology and cell differentiation. The

*To whom correspondence should be addressed. Tel: +86 15004591078; Fax: +86 459 8153035; Email: lcqbio@163.com
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

studies highlight the important and widespread utility of super-enhancers in biological and medical research.

Previous studies showed that the histone H3K27ac mark is an efficient and robust means of super-enhancer demarcation (1,7,14). Although several super-enhancer databases have been developed such as dbSUPER (15) and SEA (16). These databases are effective data sources for super-enhancer investigation. Existing databases provide only basic information about super-enhancers, such as their genome location, cell or tissue types and associated genes (17). However, with the rapid development of human epigenetics studies, human H3K27ac ChIP-seq datasets are accumulating. The effective collection and processing of these data are urgently needed. More importantly, a number of studies show that super-enhancer-associated SNPs and TFs strongly influence human disease and biology processes (6,11,13). Follow-up studies of super-enhancers largely depend on subsequent reliable regulatory annotation (1). Therefore, building a human super-enhancer database is necessary to integrate, analyze, and reveal the regulatory mechanism of super-enhancers to accelerate research and discovery of their functions.

To this end, we developed a comprehensive human super-enhancer database (SEdb, http://www.licpathway.net/sedb). SEdb focuses on providing a large number of available resources on human super-enhancers. It annotates their potential cell specific functions in gene regulation. The current version of SEdb documented a total of 331 601 super-enhancers from 542 samples, including samples from NCBI GEO/SRA (18,19), ENCODE (20), Roadmap (20,21) and GGR (Genomics of Gene Regulation Project) (20). Furthermore, SEdb provides detailed genetic and epigenetic information about super-enhancers including common SNPs, motif changes, eQTLs, risk SNPs, TFBSs, CRISPR/Cas9 target sites, DHSs and enhancers. The database supports the display of SNP effects on regulatory motifs for performing in-depth analyses of super-enhancers. SEdb is a comprehensive human super-enhancer database that integrates multiple functions of storage, browsing, annotation, and analysis. It could become a powerful work platform for mining deep functions and finding relevant regular patterns about super-enhancers.

DATA SOURCE AND PROCESSING

Identification of super-enhancers

In SEdb, we collected the 542 publicly available human H3K27ac samples for more than 240 tissues and cell types. To ensure the quality of super-enhancer identification, each of the H3K27ac samples collected by SEdb needs to contain H3K27ac ChIP-seq and the corresponding input control sequencing data. First, we integrated H3K27ac ChIP-seq data from NCBI GEO/SRA (18,19), ENCODE (20), Roadmap (20,21) and GGR (20) (Figure 1). Notably, we downloaded the data for ENCODE, Roadmap and GGR from the ENCODE/Roadmap website (www.encodeproject.org). In the process of screening NCBI GEO/SRA data, we did not consider samples that appeared in ENCODE, Roadmap or GGR, to prevent duplication. Furthermore, all data from ENCODE, Roadmap, GGR and GEO/SRA were further de-duplicated by manual screening according to the unique GEO/SRA series number. Second, to identify super-enhancers, Bowtie (v0.12.9) (1,22) was used for sequence alignment and to map ChIP-seq reads to hg19 reference genomes downloaded from UCSC Genome Bioinformatics (23). Third, for the sequence alignment file (in .sam format) generated by Bowtie, MACS14 (v1.4.2) (24) was used to identify enhancer enrichment regions. Fourth, the ROSE (9) algorithm was used to identify super-enhancers region as 'python ROSE_main.py -g hg19 -i *******.gff –c *******_input.sort.bam -r *******_cas.sort.bam -o ******* -s 12500'. In the recognition process, H3K27ac peaks within ±1 kb of transcription start sites were subtracted and the enhancer sutured at a distance of 12 500 bp before ranking stitched enhancers according to H3K27ac ChIP-seq occupancy rates. Finally, a threshold was determined according to the geometric inflection point to distinguish between enhancers and super-enhancers (1,9). These steps identified 331 601 super-enhancers and 1 992 738 super-enhancer elements in the samples.

Annotation of super-enhancers

To mine the deeper functions of super-enhancers, we provided genetic and epigenetic annotations for each super-enhancer including common SNPs, motif changes, eQTLs, risk SNPs, TFBSs, CRISPR/Cas9 target sites, DHSs and enhancers. We used BEDTools (v2.25.0) (25) to annotate corresponding information to super-enhancers and displayed details of the annotation using interactive tables.

Common SNPs/Linkage disequilibrium SNPs/Risk SNPs. We obtained 38,063,729 common SNPs from dbSNP release 150 (26). For common SNPs, linkage disequilibrium (LD) was calculated using phased genotype information accompanying the 1000 Genomes Project phase 3 (27). We used VCFTools (v0.1.13) (28) to filter out SNPs with a minimum allele frequency (MAF) less than 0.05. We used plink (v1.9) (29) to calculate SNPs for MAF >0.05 in the LD ($r^2 = 0.8$) of five super-populations (African, Ad Mixed American, East Asian, European and South Asian). For risk SNPs, genome-wide association studies (GWAS) results were obtained from a table curated by the GWAS Catalog (30) and GWASdb v2.0 (31) collection, which provided functional annotations for SNPs and insertion/deletions variants in the human disease/traits.

Motif changes. To annotate the effects of mutations on motifs, position weight matrices were collected from TRANSFAC (32) and JASPAR (33). The R package atSNP (34) was used to calculate binding affinities of mutations to motifs. For SNPs with MAF >0.05 of 1000 Genomes Project phase 3 (27) located in super-enhancer regions, a 30-bp region upstream and downstream of SNPs was calculated. After calculation, SEdb included 254 545 586 motif changes.

EQTLs. Human eQTL datasets were downloaded and merged from GTEx v5.0 (35,36), HaploReg (37) and PancanQTL (38). Data from GTEx v5.0 and HaploReg mainly included pairs of eQTL-gene relationships in different tissues. PancanQTL data included pairs of eQTL-gene relationships for different cancers in TCGA (https://tcga-data.

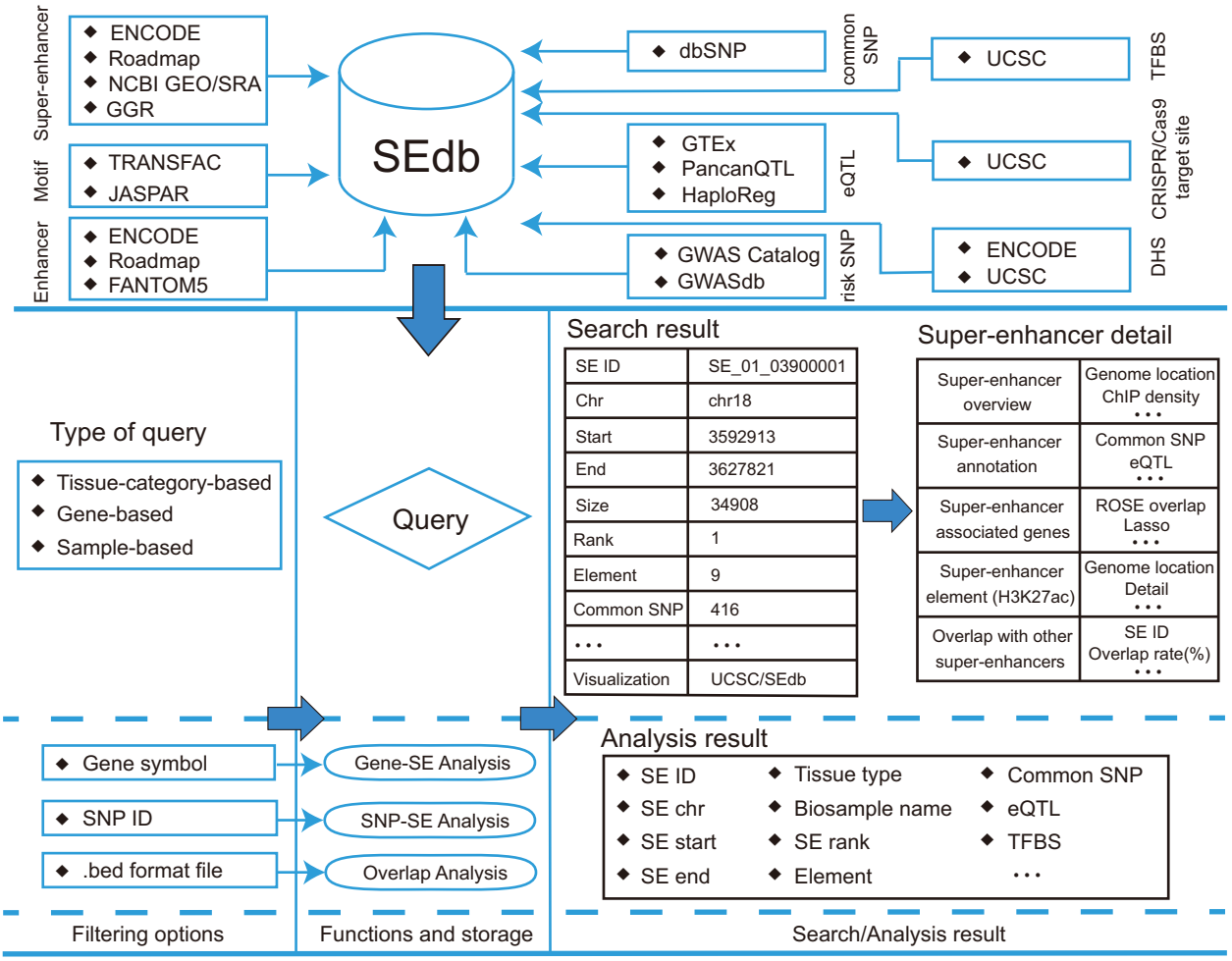

Figure 1. Database content and construction. SEdb-calculated super-enhancers based on H3K27ac ChIP-seq data. Genetic and epigenetic annotations were collected or calculated including common SNPs, eQTLs, risk SNPs, TFBSs, CRISPR/Cas9 target sites, DHSs, enhancers, motif changes and LD SNPs. Users query super-enhancers using three types: tissue-category-based query, gene-based query and sample-based advanced query. SEdb includes analytical tools and personalized genome browser to discover potential biological effects of super-enhancers. DHS: Dnase I hypersensitivity site, TFBS: transcription factor binding site, eQTL: expression quantitative trait locus.

nci.nih.gov/tcga). We mapped SNPs of eQTL and annotated them to super-enhancers regions, and the genes regulated by the SNP were provided for potential target genes for associated super-enhancers.

DHSs. DHS annotation was downloaded from UCSC (23) and ENCODE (20) for 69 860 705 DHSs of 293 samples. We match corresponding DNA hypersensitivity data to super-enhancers of sample/cell type in the database if DNA hypersensitivity data are available for that sample/cell type.

Enhancers. To annotate more enhancers within super-enhancer regions, we downloaded predicted enhancers from three major genome/epigenome annotation projects. We obtained 14 867 092 enhancers from ENCODE (20) and Roadmap (21) predicted by ChromHMM (39) method. We obtained 65,423 enhancers from FANTOM5 (40,41) predicted by cap analysis of gene expression (CAGE) (42).

TFBSs. TFBSs were downloaded from UCSC (23), including 5 797 266 TFBSs.

CRISPR/Cas9 target sites. CRISPR/Cas9 target sites can be used to induce precise cleavage of endogenous genomic loci in biological cells (43). The CRISPR/Cas9 target site annotation displays DNA sequence that targets the CRISPR RNA sequence and transcribed region within 200 bp of the genomic regions (23,44). CRISPR/Cas9 target sites were downloaded from UCSC (23), predicted by the tool CRISPOR (43), which helps to design, evaluate, and clone guide sequences for the CRISPR/Cas9 system.

Super-enhancers associated genes

Read density of H3K27ac ChIP-seq data around the TSS can be used to estimate which genes in a cell type are expressed (9). We therefore calculated the closest active genes of super-enhancers using H3K27ac ChIP-seq data, and proximities to genes using the Young *et al.* algorithm (45). Transcript ranking was obtained by grading the H3K27ac

reading density in the ±1 kb region around the TSS for each transcript in each sample. The transcript was then assigned to the gene and the repeat gene was deleted, and the first two-thirds of the ranked gene was recognized as the closest active gene. Finally, the closest active genes of super-enhancers were obtained according to their proximities (45–47). The closest active genes of super-enhancers were used as the default gene-based query interface to query super-enhancers according to super-enhancer-associated genes.

In addition, we used five other strategies to obtain super-enhancer-associated genes. Three of five strategies were the ROSE (9) method for predicting associated genes including overlap, proximal and closest. The other two were enhancer target–gene algorithms Lasso (48) and PreSTIGE (49), we directly downloaded the relevant results. When enhancers of samples were located in super-enhancer regions, the target gene of the enhancer was considered to be associated genes of the super-enhancer. The super-enhancer-associated genes obtained from all six strategies are provided in SEdb and can be used as the gene-based query interface in SEdb.

DATABASE USE AND ACCESS

A search interface for retrieving super-enhancers

The top navigation bar of SEdb is designed to help users access the database features (Figure 2A). SEdb provides a variety of query methods, including tissue-category-based, gene-based and sample-based advanced queries. Based on the tissue query, users can query the super-enhancer for all samples of a particular type of tissue. In the gene-based query, users can query a gene of interest and SEdb will return all super-enhancers that match the super-enhancer–gene relationship for all samples. In the sample-based advanced query, users determine the scope of the super-enhancer query by determining the sample and genome location for the results of interest. Brief information on the search results is displayed in a table on the results page (Figure 2B). The interactive table describes the super-enhancer's ID coded by SEdb (SE ID), genome location, size, rank, number of elements that are constituents of the super-enhancer, visualization and statistics of annotation in the region. For each sample, super-enhancers and typical enhancers can be downloaded from the results page. The results page also displays search parameters, sample description information, and usage parameters for the software. Users click 'SE ID' for details about the super-enhancer. In addition to general information about the super-enhancer (Figure 2C), SEdb lists more detailed annotation information including common SNPs, eQTLs, risk SNPs, TFBSs, CRISPR/Cas9 target sites, DHSs and enhancers (Figure 2D). For example, for common SNPs, SEdb provides the number of common SNPs in the super-enhancer region and details about common SNPs within the super-enhancer, computing the effect of SNPs on regulatory motifs (Figure 2D). Genes potentially associated with the super-enhancer are provided through five different identification strategies (Figure 2E). SEdb provides details and analysis of each element of the super-enhancer (Figure 2F). Detailed information for the elements is viewed by clicking 'Detail'. A detailed page of the element includes annotation and summary information. SEdb provides links to other super-enhancers identified in samples that overlap with this super-enhancer, viewed in detail by clicking 'SE ID'. This is equivalent to regional enrichment analysis (Figure 2G).

User-friendly browsing of samples

The 'Data-Browse' page is an interactive and alphanumerically sortable table that allows users to quickly search for samples and customize filters using 'Data sources', 'Biosample type', 'Tissue type' and 'Biosample name' (Figure 2J). Users use the 'Show entries' drop-down menu to change the number of records per page. To view super-enhancers for a given sample, users click on 'Sample ID'.

Online analysis tools

Using the 'Gene-SE analysis' tool, users submit a gene and analyze super-enhancers associated with it via relationships between the super-enhancer and associated genes identified under different strategies (9,45,48,49) (Six strategies: the closest active gene, ROSE overlap, ROSE proximal, ROSE closest, Lasso and PreSTIGE) are obtained from determined and indeterminate samples (Figure 2H). SEdb also links to external resources including NCBI Gene (50,51), GeneCards (52), UniProt (53) and Wikipedia (https://www.wikipedia.org). With the 'SNP-SE analysis' tool, users submit a common SNP and find super-enhancers in which it appears, the super-enhancer's annotation information and LD SNPs of five super-populations (Figure 2I). With the 'Overlap analysis' tool, users submit a 'bed' file and identify super-enhancers with overlapping relationships with the submitted regions by setting the percentage of overlap (Figure 2N). SEdb supports analysis tools of external links for search results and super-enhancer elements such as GREAT (54) and Galaxy (55).

Personalized genome browser and data visualization

To help users view proximity information of super-enhancers in genomes, we developed a personalized genome browser using JBrowse (56) with useful tracks (Figure 2K). Users see the proximity of super-enhancers to nearby genes, genome segments, SNPs, common SNPs, risk SNPs, DHSs, enhancers, TFBS conserved, TFBS by ChIP-seq and conservative score. SEdb exhibits super-enhancer-associated pie charts of chromosome distribution and histograms of annotation statistics (Figure 2C). Relationships between super-enhancers and associated genes are displayed using a D3 network visualization plugin (Figure 2E). SEdb also links to visualize data in the UCSC genome browser (23) by adding custom tracks.

CGI interface

For data sharing, a CGI program was built for SEdb. Users such as website developers provide a genome location and use the SEdb website to determine the super-enhancers that overlap with the location. Data obtained from the feedback is displayed directly on the platform.

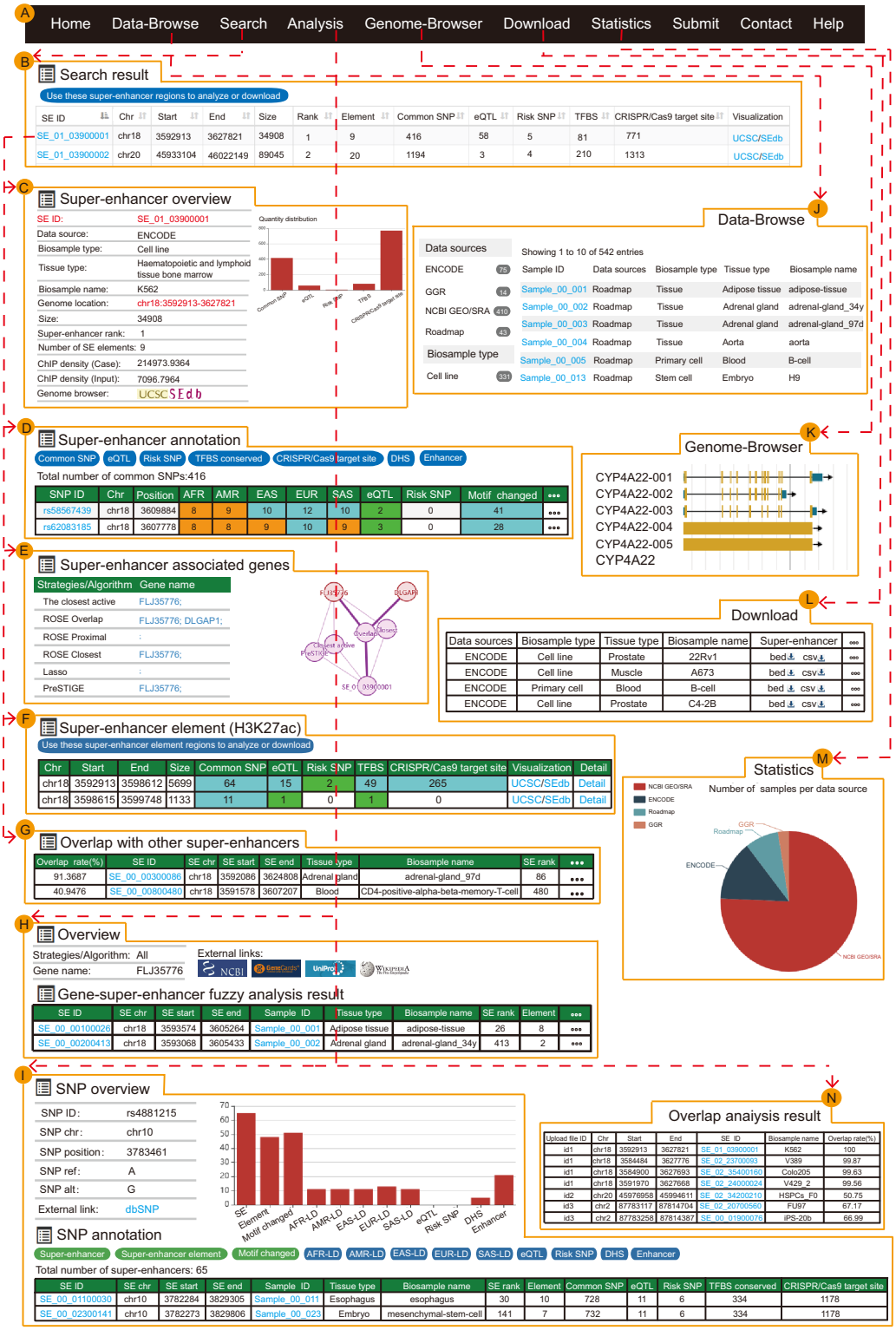

Figure 2. The main functions and usage of SEdb. (**A**) Top navigation bar help users use functions of this database. (**B**) Table of search results including super-enhancer ID coded by SEdb (SE ID), Chr, Start, End, Size, Rank, Element, Common SNP, eQTL, Risk SNP, TFBS, CRISPR/Cas9 target site and Visualization (genome browser). (**C**) Overview of super-enhancer. (**D**) Detailed interactive table of annotation information. (**E**) Genes potentially associated with super-enhancers are provided through six identification strategies. Network diagram of their relationships. (**F**) Interactive table of super-enhancer elements, related annotation information and analysis tools. (**G**) Other super-enhancers identified in samples overlapping the super-enhancer. (**H**) Analysis of super-enhancers related to a query gene via relationships between super-enhancers and associated genes under different strategies. (**I**) Analysis of a common SNP, super-enhancers it appears in, and annotation of the common SNP. (**J**) Browse the sample details. (**K**) Visualization of genome browser for genetic and epigenetic information. (**L**) Data download. (**M**) Quantitative statistics of data sources in SEdb. (**N**) This is an overlap analysis tool that uses the super-enhancers provided in SEdb to annotate the user-submitted regions.

Data download and statistics

SEdb provides downloads of super-enhancers and super-enhancer elements for each sample, including bed format and csv format (Figure 2L). Considering that typical enhancers may also be of importance to users, typical enhancers are also provided on the download page of the database. In addition, the database supports the packaged downloads of all enhancers, including super-enhancer and typical enhancer sets. The export of data is also supported for search results of interest to the users. In the 'Statistics' page, we can visually see the statistics of SEdb, including digital display and graphical display (Figure 2M). In addition, samples information for DHSs and enhancers are also provided.

Data submission

Database updates are critical to their sustainability. Users can share their H3K27ac data to SEdb. To ensure data quality control, we recommend that users submit the GEO/SRA series number for the raw data on the 'Submit' page. For data from other sources, the submitter needs to provide the corresponding accessible URL for storing the raw fastq files and the sample information for the submitted data (see 'Submit' page of SEdb for the detailed process). We will update the data dynamically according to number of samples, to ensure the timely release of the data.

SYSTEM DESIGN AND IMPLEMENTATION

The current version of SEdb was developed using MySQL 5.7.17 (http://www.mysql.com) and runs on a Linux-based Apache Web server (http://www.apache.org). We used PHP 7.0 (http://www.php.net) for server-side scripting. We designed and built the interactive interface using Bootstrap v3.3.7 (https://v3.bootcss.com) and JQuery v2.1.1 (http://jquery.com). We used ECharts (http://echarts.baidu.com) (57) and D3 (https://d3js.org) as a graphical visualization framework, and JBrowse (http://jbrowse.org) is the genome browser framework. We recommend using a modern web browser that supports the HTML5 standard such as Firefox, Google Chrome, Safari, Opera or IE 9.0+ for the best display.

The SEdb database is freely available to the research community using the web link (http://www.licpathway.net/sedb). Users are not required to register or login to access features in the database.

DISCUSSION

The emerging importance of super-enhancers in human diseases and biological processes, coupled with their exquisite tissue-specificity, raises the need for comprehensive super-enhancer catalogs of human. The existing databases, including dbSUPER and SEA, are based on data mainly from the ENCODE/Roadmap project and integrate other research results. These databases contain significantly fewer human ChIP-seq samples than NCBI GEO/SRA. Therefore, we created SEdb, a comprehensive human super-enhancer database with a large number of human samples. SEdb integrated 542 human H3K27ac samples from NCBI GEO/SRA, ENCODE, Roadmap and GGR, and calculated 331 601 super-enhancers based on these data. By manually curating and classifying 410 available H3K27ac samples from >2000 ChIP-seq samples from the NCBI GEO/SRA. To ensure the quality of super-enhancer identification, each of the H3K27ac samples collected by SEdb needs to contain H3K27ac ChIP-seq and the corresponding input control sequencing data. Furthermore, a sample, as well as cell type, will be contained in the database if super-enhancers were successfully identified in the sample by H3K27ac ChIP-seq and corresponding input control sequencing. The number of samples in SEdb was more than 5-fold the samples in dbSUPER. SEA was updated to new version 2.0. Compared to the previous version, ChIP-seq samples from more species, including humans, were supported. However, the number of human H3K27ac samples in SEdb was 353 more than the SEA v2.0.

SEdb provides a user-friendly interface to search, browse, analyze and visualize information about super-enhancers. SEdb has rich annotations and element information and useful analysis tools and visualizations. Table 1 compares SEdb with other super-enhancer databases for information and functions, showing the SEdb advantages. SEdb provides: (i) comprehensive genetic and epigenetic annotation of super-enhancers including common SNPs, motif changes, eQTLs, risk SNPs, TFBSs, CRISPR/Cas9 target sites, DHSs and enhancers, and user-friendly displays with interactive tables; (ii) online analysis tools such as 'Gene-SE analysis', 'SNP-SE analysis', 'Overlap analysis' and analysis tools from external links such as GREAT and Galaxy for search results; (iii) a customized genome browser for user-friendly visualizing of genomic context information of super-enhancers and links for visualizing data in the UCSC genome browser by adding custom tracks; (iv) user-friendly browsing of samples; (v) a CGI interface that can be easily used and quickly generate super-enhancers that overlap with the user-submitted genome location; (vi) detailed internal information on super-enhancer elements, including related annotations and related analysis tools; (vii) overlapping contacts with other super-enhancers in different samples.

SEdb is a super-enhancer database with the largest number of human super-enhancers and samples and the most comprehensive annotation information about super-enhancers. Because sequence variants in super-enhancer regions increase the risk of common human diseases, detailed genetic information on super-enhancers such as risk SNPs, eQTLs and motif changes are provided in SEdb. The current version of SEdb mainly focuses on super-enhancers identified by H3K27ac data through the ROSE algorithm, though typical enhancers can also be identified by ROSE. We therefore also provided general information on typical enhancers in SEdb, considering that these may be of interest to users of this database. However, given that the number of typical enhancers in a sample is much greater than that of super-enhancers, no further detailed annotation of typical enhancers is provided in SEdb. In future versions, we will provide more annotation information of super-enhancers, especially for typical enhancers, and practical analysis tools. In order to keep the data update, we add more annotation information and practical analysis tools. We believe that

Table 1. Comparison of SEdb with other databases that are based on human super-enhancer-related data and functions (20 June 2018)

Function type	Data type/Specific function	SEdb	dbSUPER	SEA v2.0
Interaction table /annotation	Number of human samples	542	102	189
	Number of human super-enhancers	331 601	69 205	164 398
	Strategies of super-enhancer associated genes[a]	6	3	3
	Common SNP	✓		
	Motif changed	✓		
	eQTL	✓		
	Risk SNP	✓		
	TFBS	✓		✓
	CRISPR/Cas9 target site	✓		✓
	DHS	✓		
	Enhancer[b]	✓		
	LD SNP	✓		
Genome browser	Super-enhancers	✓		✓
	Super-enhancer elements	✓		
	Genome segments	✓		
	SNP	✓		
	Common SNP	✓		✓
	Risk SNP	✓		✓
	TFBS conserved	✓		
	TFBS by ChIP-seq	✓		✓
	CRISPR/Cas9 target site	✓		✓
	DHS	✓		
	Enhancer	✓		
	Conservative score	✓		✓
Analysis functions	Gene-SE analysis	✓		
	SNP-SE analysis	✓		
	Overlap analysis	✓	✓	✓
	Region analysis[c]	✓	✓	✓
Data browse	Simple browse[d]	✓	✓	✓
	Browse based on samples classification[e]	✓		
	Alphanumerically sortable table	✓		
CGI tool	Genome location overlap[f]	✓		
Other functions	Super-enhancer element annotation	✓		
	Overlap with other super-enhancers in different samples	✓		

[a]Super-enhancer-associated genes obtained by different strategies or algorithms.
[b]ChromHMM method or CAGE to predict enhancers.
[c]External link to GREAT and Galaxy.
[d]Simple browser function of super-enhancer samples.
[e]Classification of samples including Data sources, Biosample type, Tissue type and Biosample name.
[f]Quickly generate super-enhancers that overlap with the user-submitted genome location.

SEdb can promote researches and discovery of more potential biological effects of super-enhancers.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [81572341, 61601150] (in part); Natural Science Foundation of Heilongjiang Province [JJ2016ZR1232/F2016031]; Yu Weihan Outstanding Youth Training Fund of Harbin Medical University. Funding for open access charge: National Natural Science Foundation of China [81572341, 61601150].
Conflict of interest statement. None declared.

REFERENCES

1. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saintandré,V., Sigova,A.A., Hoke,H.A. and Young,R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
2. Chapuy,B., Mckeown,M.R., Lin,C.Y., Monti,S., Roemer,M.G., Qi,J., Rahl,P.B., Sun,H.H., Yeda,K.T. and Doench,J.G. (2013) Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell*, **24**, 777–790.
3. Christensen,C.L., Kwiatkowski,N., Abraham,B.J., Carretero,J., Alshahrour,F., Zhang,T., Chipumuro,E., Herterspie,G.S., Akbay,E.A. and Altabef,A. (2014) Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. *Cancer Cell*, **26**, 909–922.
4. Amaral,P.P. and Bannister,A.J. (2014) Re-place your BETs: the dynamics of super enhancers. *Mol. Cell*, **56**, 187.
5. Mack,S.C., Pajtler,K.W., Chavez,L., Okonechnikov,K., Bertrand,K.C., Wang,X., Erkek,S., Federation,A., Song,A. and Lee,C. (2017) Therapeutic targeting of ependymoma as informed by oncogenic enhancer profiling. *Nature*, **553**.
6. Oldridge,D.A., Wood,A.C., Weichertleahey,N., Crimmins,I., Sussman,R., Winter,C., Mcdaniel,L.D., Diamond,M., Hart,L.S. and Zhu,S. (2015) Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature*, **528**, 418.
7. Mckeown,M.R., Corces,M.R., Eaton,M.L., Fiore,C., Lee,E., Lopez,J.T., Chen,M.W., Smith,D., Chan,S.M. and Koenig,J.L. (2017) Super-Enhancer analysis defines novel epigenomic subtypes of Non-APL AML including an RARα dependency targetable by SY-1425, a potent and selective RARα agonist. *Cancer Discov.*, **7**, 1136–1153.

8. Mansour,M.R., Abraham,B.J., Anders,L., Berezovskaya,A., Gutierrez,A., Durbin,A.D., Etchin,J., Lawton,L., Sallan,S.E. and Silverman,L.B. (2014) Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*, **346**, 1373–1377.
9. Lin,C.Y. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
10. Van,G.T., Koster,J., Valentijn,L.J., Zwijnenburg,D.A., Akogul,N., Hasselt,N.E., Broekmans,M., Haneveld,F., Nowakowska,N.E. and Bras,J. (2017) Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat. Genet.*, **49**, 1261–1266.
11. Parker,S.C., Stitzel,M.L., Taylor,D.L., Orozco,J.M., Erdos,M.R., Akiyama,J.A., van Bueren,K.L., Chines,P.S., Narisu,N. and Black,B.L. (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 17921–17926.
12. Vahedi,G., Kanno,Y., Furumoto,Y., Jiang,K., Parker,S.C., Erdos,M.R., Davis,S.R., Roychoudhuri,R., Restifo,N.P. and Gadina,M. (2015) Stretch-Enhancers delineate Disease-Associated regulatory nodes in T cells. *Nature*, **520**, 558–562.
13. Radhika,K., Sava,G.P., Speedy,H.E., Sílvia,B., Martín-Subero,J.I., Studd,J.B., Gabriele,M., Law,P.J., Puente,X.S. and David,M.G. (2016) Genetic predisposition to chronic lymphocytic leukemia is mediated by aBMFSuper-Enhancer polymorphism. *Cell Rep.*, **16**, 2061–2067.
14. Whyte,W., Orlando,D., Hnisz,D., Abraham,B., Lin,C., Kagey,M., Rahl,P., Lee,T.I. and Young,R. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
15. Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.
16. Wei,Y., Zhang,S., Shang,S., Zhang,B., Li,S., Wang,X., Wang,F., Su,J., Wu,Q. and Liu,H. (2016) SEA: a super-enhancer archive. *Nucleic Acids Res.*, **44**, D172–D179.
17. Wang,Z., Zhang,Q., Zhang,W., Lin,J.R., Cai,Y., Mitra,J. and Zhang,Z.D. (2017) HEDD: Human Enhancer Disease Database. *Nucleic Acids Res.*, **46**, D113–D120.
18. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H. and Sherman,P.M. (2012) NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.*, **39**, 1005–1010.
19. Kodama,Y., Shumway,M. and Leinonen,R. (2012) The sequence read archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, 54–56.
20. Karmakar,S. (2012) An integrated encyclopedia of DNA elements in human genome. *Nature*, **489**, 57–74.
21. Bernstein,B.E., Stamatoyannopoulos,J.A., Costello,J.F., Ren,B., Milosavljevic,A., Meissner,A., Kellis,M., Marra,M.A., Beaudet,A.L. and Ecker,J.R. (2010) The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.*, **28**, 1045–1048.
22. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
23. Rosenbloom,K.R., Armstrong,J., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M. *et al.* (2015) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.*, **43**, D670.
24. Yong,Z., Tao,L., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M. and Wei,L. (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, 1–9.
25. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841.
26. 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.
27. 1000 Genomes Project Consortium, Abecasis,G.R., Auton,A., Brooks,L.D., DePristo,M.A., Durbin,R.M., Handsaker,R.E., Kang,H.M., Marth,G.T. and McVean,G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
28. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., Depristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T. and Sherry,S.T. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
29. Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A., Bender,D., Maller,J., Sklar,P., de Bakker,P.I., Daly,M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
30. Welter,D., Macarthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T. and Hindorff,L. (2014) The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, 1001–1006.
31. 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.
32. Matys,V., Kelmargoulis,O.V., Fricke,E., Liebich,I., Land,S., Barredirrie,A., Reuter,I., Chekmenev,D., Krull,M. and Hornischer,K. (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
33. Khan,A., Fornes,O., Stigliani,A., Gheorghe,M., Castro-Mondragon,J.A., Van,d.L.R., Bessy,A., Chã¨Neby,J., Kulkarni,S.R. and Tan,G. (2017) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, **77**, e43.
34. Zuo,C., Shin,S. and Keleş,S. (2015) atSNP: transcription factor binding affinity testing for regulatory SNP detection. *Bioinformatics*, **31**, 3353–3355.
35. Lonsdale,J., Thomas,J., Salvatore,M., Phillips,R., Lo,E., Shad,S., Hasz,R., Walters,G., Garcia,F. and Young,N. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **13**, 307–308.
36. Consortium,G. (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648.
37. Ward,L.D. and Kellis,M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, 930–934.
38. Gong,J., Mei,S., Liu,C., Xiang,Y., Ye,Y., Zhang,Z., Feng,J., Liu,R., Diao,L. and Guo,A.Y. (2017) PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.*, **46**, D971.
39. Ernst,J. and Kellis,M. (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods*, **9**, 215–216.
40. Andersson,R., Gebhard,C., Miguelescalada,I., Hoof,I., Bornholdt,J., Boyd,M., Chen,Y., Zhao,X., Schmidl,C. and Suzuki,T. (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.
41. Kawaji,H. (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462–470.
42. Kodzius,R., Kojima,M., Nishiyori,H., Nakamura,M., Fukuda,S., Tagami,M., Sasaki,D., Imamura,K., Kai,C. and Harbers,M. (2006) CAGE: cap analysis of gene expression. *Nat. Methods*, **687**, 211–222.
43. Haeussler,M., Kai,S., Eckert,H., Eschstruth,A., Mianné,J., Renaud,J.B., Schneidermaunoury,S., Shkumatava,A., Teboul,L. and Kent,J. (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.*, **17**, 148.
44. Bae,S., Kweon,J., Kim,H.S. and Kim,J.S. (2014) Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods*, **11**, 705–706.
45. Saintandré,V., Federation,A.J., Lin,C.Y., Abraham,B.J., Reddy,J., Lee,T.I., Bradner,J.E. and Young,R.A. (2016) Models of human core transcriptional regulatory circuitries. *Genome Res.*, **26**, 385–396.
46. Odom,D.T., Dowell,R.D., Jacobsen,E.S., Lena,N., Alexander,R.P., Danford,T.W., Gifford,D.K., Ernest,F., Bell,G.I. and Young,R.A. (2006) Core transcriptional regulatory circuitry in human hepatocytes. *Mol.Syst.Biol.*, **2**, doi:10.1038/msb4100059.
47. Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zucker,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L. and Jenner,R.G. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
48. Cao,Q., Anyansi,C., Hu,X., Xu,L., Xiong,L., Tang,W., Mok,M.T.S., Cheng,C., Fan,X. and Gerstein,M. (2017) Reconstruction of

enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nat. Genet.*, **49**, 1428.
49. Corradin,O., Saiakhova,A., Akhtarzaidi,B., Myeroff,L., Willis,J., Cowpersal,R.L., Lupien,M., Markowitz,S. and Scacheri,P.C. (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.*, **24**, 1.
50. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaudnissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J. and Mcgarvey,K.M. (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, 756–763.
51. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54.
52. Safran,M., Dalah,I., Alexander,J., Rosen,N., Iny,S.T., Shmoish,M., Nativ,N., Bahir,I., Doniger,T. and Krug,H. (2010) GeneCards Version 3: the human gene integrator. *Database (Oxford)*. **2010**, baq020.
53. Consortium, U.P. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, 204–212.
54. Mclean,C.Y., Bristor,D., Hiller,M., Clarke,S.L., Schaar,B.T., Lowe,C.B., Wenger,A.M. and Bejerano,G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.*, **28**, 495–501.
55. Blankenberg,D., Coraor,N., Von,G.K., Taylor,J. and Nekrutenko,A. (2011) Integrating diverse databases into an unified analysis framework: a Galaxy approach. *Database (Oxford)*, **2011**, bar011.
56. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoztorres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E. and Stein,L. (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
57. Li,D., Mei,H., Shen,Y., Su,S., Zhang,W., Wang,J., Zu,M. and Chen,W. (2018) ECharts: a declarative framework for rapid construction of web-based visualization. *Vis. Informatics*, **2**, 136–146.