

ENdb: a manually curated database of experimentally supported enhancers for human and mouse

Xuefeng Bai[†], Shanshan Shi[†], Bo Ai[†], Yong Jiang[†], Yuejuan Liu[†], Xiaole Han, Mingcong Xu, Qi Pan, Fan Wang, Qiuyu Wang, Jian Zhang, Xuecang Li, Chenchen Feng, Yanyu Li, Yuezhu Wang, Yiwei Song, Ke Feng and Chunquan Li*

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

Received August 15, 2019; Revised October 10, 2019; Editorial Decision October 11, 2019; Accepted October 16, 2019

ABSTRACT

Enhancers are a class of *cis*-regulatory elements that can increase gene transcription by forming loops in intergenic regions, introns and exons. Enhancers, as well as their associated target genes, and transcription factors (TFs) that bind to them, are highly associated with human disease and biological processes. Although some enhancer databases have been published, most only focus on enhancers identified by high-throughput experimental techniques. Therefore, it is highly desirable to construct a comprehensive resource of manually curated enhancers and their related information based on low-throughput experimental evidences. Here, we established a comprehensive manually-curated enhancer database for human and mouse, which provides a resource for experimentally supported enhancers, and to annotate the detailed information of enhancers. The current release of ENdb documents 737 experimentally validated enhancers and their related information, including 384 target genes, 263 TFs, 110 diseases and 153 functions in human and mouse. Moreover, the enhancer-related information was supported by experimental evidences, such as RNAi, in vitro knockdown, western blotting, qRT-PCR, luciferase reporter assay, chromatin conformation capture (3C) and chromosome conformation capture-on-chip (4C) assays. ENdb provides a user-friendly interface to query, browse and visualize the detailed information of enhancers. The database is available at <http://www.licpathway.net/ENdb>.

INTRODUCTION

Gene transcription is regulated by genomic elements, such as enhancers, promoters, silencers and insulators in a tissue-

specific and spatiotemporal-specific manner (1,2). Among such elements, enhancers are a class of *cis*-regulatory elements that can facilitate the transcriptional activity of genes by forming loops in intergenic regions, introns and exons. A large number of enhancers, as well as their corresponding target genes and transcription factors (TFs), were found to be highly associated with human disease and biological processes (3). Active enhancers are always marked by chromatin marks (e.g. H3K4me1 and H3K27ac) and can be identified through high-throughput and low-throughput experimental methods (4–8). The most commonly used low-throughput methods include RNAi (9), in vitro knockdown (10), electrophoretic mobility shift assay (EMSA) (11), western blotting (12), qRT-PCR (13), luciferase reporter assays (14), chromatin conformation capture (3C) (15) and chromosome conformation capture-on-chip (4C) (16) assays. These assays have been used to directly and quantitatively assess enhancer activity, as well as that of their related regulatory elements. Low-throughput techniques are undoubtedly more accurate for the identification of enhancers compared with high-throughput techniques. Conversely, high-throughput experimental techniques provide approaches to investigate huge amounts of computationally-predicted enhancers and genome-wide features. To date, the human genome has been extensively cataloged with over one million candidate enhancers (17). Additionally, the number of publications on enhancers increases each year, which will advance in research will provide for collecting more reliable evidences of enhancer.

Some enhancer databases, as well as super-enhancer databases, have been published, including EnhancerAtlas (18), HACER (19), HEDD (20), EnhancerDB (21), DiseaseEnhancer (22) and EnDisease (23). These databases have provided valuable resources for enhancer investigation (24–33); however, most of them only focus on enhancers identified by high-throughput experimental techniques. Recognizing that potential false positives exist in high-throughput experiments, several low-throughput experimentally-supported enhancer databases were built

*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 five authors should be regarded as Joint First Authors.

(22,23,34). For example, the VISTA Enhancer Browser (34) is a central resource for experimentally-validated human and mouse fragments with gene enhancer activity assessed in transgenic mice. The VISTA Enhancer Browser group performed numerous transgenic mouse assays to determine the activity of putative enhancers. DiseaseEnhancer and ENdisease differ from VISTA Enhancer Browser in terms of their strategy of collecting enhancers. Those database manually curated low-throughput experimentally-supported enhancers by reviewing hundreds of publications. However, they are specific in their respective purposes (i.e. focus on diseases). Therefore, it is desirable to construct a comprehensive resource of manually curated enhancers and related information (e.g. target genes, TFs binding to enhancers, functions and diseases) derived from low-throughput experimental evidence.

Here, we established a comprehensive, manually curated enhancer database for human and mouse (ENdb, <http://www.licpathway.net/ENdb>), which aimed to provide a resource for experimentally-supported enhancers, and to annotate the details of such enhancers. Enhancers were manually examined from a review of 1,590 published studies. In general, if the genomic location of enhancers could be retrieved and enhancer activity could be experimentally validated, the enhancers were stored in ENdb. The current release of ENdb documents 737 experimentally validated enhancers and related information, including 384 target genes, 263 TFs, 110 diseases and 153 functions in human and mouse. ENdb provides a user-friendly interface to query, browse and visualize detailed information about enhancers. We hope that this elaborate database, which is supported by strong experimental evidence, can serve as an important catalyst for future research.

DATA COLLECTION AND DATABASE CONTENT

To ensure high quality data collection, we referred to the steps involved in manual collection from other databases, such as DiseaseEnhancer (22), Lnc2Cancer (35) and EVLncRNAs (36). The abstracts of all publications with the keyword matching ‘enhancer,’ were initially retrieved from the PubMed database (37). We manually reviewed a total of 23 462 abstracts to obtain the pertinent enhancer-related publications for human and mouse. A total of 4641 publications associated with enhancers were retained for human and mouse. The following workflow was mainly applied for extracting enhancer information from the publications. First, we filtered 4601 publications based on two criteria: (i) the genomic location of enhancers could be obtained from the publications (e.g. from the figures, or from in-text descriptions); and (ii) the enhancers were supported by at least one form of experimental evidence. After filtering publications based on those criteria, 1590 publications were retained. Second, we reviewed the full text of the 1590 publications in detail and extracted the general information of enhancers and related information (e.g. target genes, TFs binding to enhancers, functions and diseases). Notably, general information of enhancers was extracted from the publications, including the PubMed ID of the publication, the enhancer name, chromosome and positional information, experimental techniques (e.g. luciferase

reporter assay, EMSA, western blotting and 3C), experimental samples (cell line or tissue), the experimental class of enhancers (‘low-throughput’, ‘low- + high-throughput’), species (human or mouse) and reference genome (hg19 or mm10). Third, the enhancer-related information was also collected, including the TFs binding to enhancers, target genes, functions and diseases, as well as the type of experimental assay performed, such as luciferase reporter assay, qRT-PCR, RNAi, in vitro knockdown, western blotting, 3C and 4C. For example, the experimental evidence for TFs (e.g. ChIP-qPCR and EMSA) was used to confirm the binding of transcription factors to enhancers. The relationships (38) between enhancers and target genes were divided into those supported by ‘strong evidence’ and ‘less strong evidence’. If evidence was derived from 3C, 4C, 5C or CRISPR/Cas9, a label of ‘strong evidence’ was assigned. Otherwise, ‘less strong evidence’ was assigned. For example, RNAi used to knockdown transcripts of TFs that bind the enhancers to test expression of target genes. Thus, we considered RNAi as ‘less strong evidence’ for relationships between enhancers and target genes. Finally, we used a previously standardized classification scheme described (35) to map disease names to Disease Ontology (39) and MeSH (40) databases. The names of target genes and TFs associated with enhancers were replaced with official or recommended names. All enhancers and related information (e.g., target genes, TFs binding to enhancers, functions and diseases) were manually extracted from the selected publications by at least two researchers. Moreover, ENdb provided links to external resources, including PubMed ID in NCBI, HGNC, dbSNP and diseases in the Disease Ontology and MeSH databases.

During data collection, it was a challenge to extract information on the genomic locations of enhancers. However, we found that many publications provided important clues to determine such location information. Thus, by reviewing the figures, in-text descriptions provided in the publications, we could obtain accurate location information. For example, some enhancers and related genomic locations were only provided in the figures of publications. In such cases, we determined the location of enhancers by carefully reviewing the figures in the full text. For some cases where only single point within an enhancer, such as ‘enhancer locating at upstream xxx kb from the transcription start sites’, is provided in original publications, the ± 1 kb region centered on this point was used as the genomic location of the enhancer (17). All genomic locations for human and mouse were unified in the corresponding hg19/mm10 genomic coordinates.

DATABASE STATISTICS

The current version of ENdb contains 737 enhancers, 384 target genes, 263 TFs, 110 diseases, 153 functions and 34 SNPs for human and mouse. There are 425 enhancers for human and 312 for mouse (Figure 1A). We counted the number of publications each year from 2008 to 2018, that involved in enhancers, enhancers with information of target genes or those with TF binding data (Figure 1B). We found that the number of enhancer-related studies steadily increased over the past decade. Enhancers were mainly lo-

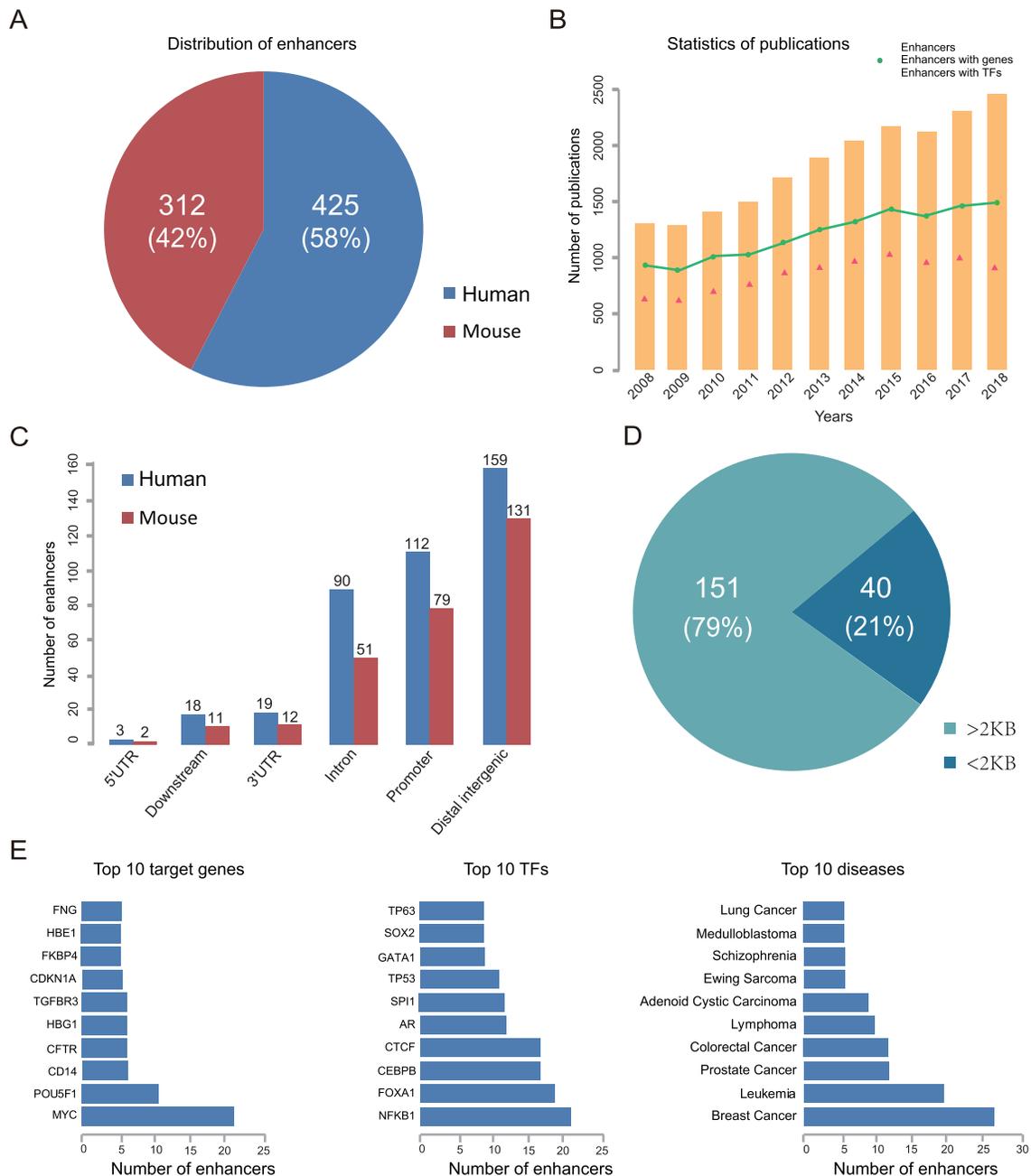


Figure 1. Statistics of enhancers in ENdb. (A) The distribution of enhancers for human and mouse. (B) The annual number of publications on enhancers and their related information each year. (C) The distribution of enhancers in genomic elements. (D) The distribution of enhancers in the promoter and distant regions. (E) The number of enhancers associated with the top 10 target genes (left), TFs (middle) and diseases (right).

cated in distal intergenic (39.7%), promoter (27.9%), intron (22.5%), downstream (4.4%), 3'UTR (4.8%), and merely containing 0.7% 5'UTR regions for human (Figure 1C). Mouse enhancers were mainly located in distal intergenic (45.8%), promoter (27.6%), intron (17.8%), 3'UTR (4.2%), downstream (3.9%) and 5'UTR (0.7%) regions (Figure 1C). Consistent with statistics from the DiseaseEnhancer database, most of enhancers were concentrated at the distal intergenic and near promoters. A possible explanation for enhancers being concentrated near promoters is that they are likely to be super-enhancers with large sizes or that they

function as promoters in higher chromosomal organization. For enhancers located at promoter regions, we calculated the distance between the enhancer and the transcription start site of their target gene. The results revealed that the distances of only 40 (21%) enhancers was <2KB (Figure 1D). Thus, those data demonstrated that most enhancers located at promoter regions regulate distant genes, and do not enhance the closest gene.

The top 10 target genes ranked by the number of associated enhancers represented by 18% of the curated target genes associated with enhancers (Figure 1E). The proto-

oncogene MYC was one of the top target genes ranked based on the number of associated enhancers, and was affected by 22 enhancers. For example, a NOTCH-bound MYC enhancer played essential roles in the T-Cell Acute Lymphoblastic Leukemia (41), while in colon cancer, MYC expression was concurrently regulated by enhancers and complexes (42). On average, 14 enhancers were associated with the top 10 TFs that were ranked based on the number of associated enhancers in the ENdb database (Figure 1E). A large number of TFs were also able to bind various enhancers and exert their functions (e.g. SPI1, CTCF, FOXA1, AR and TP53). The top 10 diseases ranked by the number of associated enhancers represented by 53.5% of the curated disease-associated enhancers. A total of 27 enhancers were involved in breast cancer, 20 in leukemia, and 12 in prostate cancer (Figure 1E).

USER INTERFACE

ENdb provides a user-friendly interface to browse, search, download and visualize detailed information about enhancers (Figure 2A). The ‘Data-Browse’ page is organized as an interactive and alphanumerically sortable table that allows users to quickly browse through ‘Enhancer ID’, ‘Enhancer symbol’, ‘Genomic location’, ‘Biosample name’, ‘Enhancer experiment’ and ‘Disease’ (Figure 2B). The sub-menu of the ‘Browse’ page provides fuzzy search functions according to different conditions (Figure 2B). Simultaneously, users can browse enhancers by clicking Enhancer ID, and a list of matched entries is returned. The ‘Detail’ page contains the general information of enhancers, such as PubMed ID of publication, the enhancer symbol, chromosome and positional information, experimental techniques (e.g., luciferase reporter assay, EMSA, western blotting, 3C), the level of evidence for the relationships between enhancers and target genes (e.g., ‘strong evidence’ or ‘less strong evidence’), experimental samples (cell line and/or tissue), species (human or mouse) and reference gene (hg19 or mm10) (Figure 2B). The enhancer-associated network graphically displays the regulatory information of the enhancer, in which the relative positions of the enhancer and target genes are provided (Figure 2C). In the ‘Search’ page, ENdb allows users to search by a variety of keywords, facilitating searches by returning the records with the closest match. Users can search enhancers via four paths, including ‘Search by enhancer’, ‘Search by target gene’, ‘Search by TF’, ‘Search by SNP’ (Figure 2D). We also provide a function to search by gene name. Using this function, Users can input gene names, including gene official symbols or aliases to search the information associated with enhancers.

In order to help users intuitively view the proximity information of enhancers in the genome, ENdb provides two genome browsers (GIVE and JBrowse; Figure 2E) (43,44). Hundreds of tracks of histone modifications and chromatin interaction data can be browsed using the genome browser. GIVE enables automatic generation of interactive visualization webpages for user-selected chromosome conformation data. In addition, in the browse and search pages, users can link to UCSC (45), GIVE and Genome-Browse via the genomic position of enhancers. This allows users to visualize enhancer-related information, such as the relative posi-

tions of enhancers and target genes. ENdb allows users to download all the obtained data in the ‘Download’ page. In addition, ENdb also offers a submission page that enables researchers to submit novel experimentally-supported enhancers and their related information.

SYSTEM DESIGN AND IMPLEMENTATION

The general workflow and features of ENdb was developed using MySQL (version 5.7.17) (<https://www.mysql.com/>). The ENdb was designed and the interactive interface was built using Bootstrap (version 3.3.7) and JQuery (version 2.1.1). The ECharts (<http://echarts.baidu.com>) and D3 (<https://d3js.org>) were used as a graphical visualization framework, and JBrowse (<http://jbrowse.org>) was used as the browser framework. We recommend using the database with a modern web browser that supports HTML5, such as Firefox, Google Chrome, Safari, Opera or IE 9.0+. ENdb is freely available to the research community using the web link provided (<http://www.licpathway.net/ENdb>). Users are not required to register or login to access features in the database.

DISCUSSION

Enhancers are critical elements for regulating the transcription, and influence the expression of genes, as well as developmental and disease-related processes. In recent years, several projects have been developed, including ENCODE (17), FANTOM5 (25), Roadmap (26) and Blueprint/IHEC (46). Many candidate enhancers have been identified and enhancer databases have been established to provide a resource for collecting the detailed information for such enhancers. However, most of those databases focus on enhancers identified by high-throughput experimental techniques, while low-throughput techniques (e.g. RNAi, EMSA, western blotting, qRT-PCR, luciferase reporter assay, 3C and 4C) are more accurate in identifying enhancers and in exploring their upstream/downstream regulatory mechanisms. Therefore, we developed ENdb, a comprehensive database of all enhancers that were validated by low-throughput experiments in human and mouse.

We compared ENdb to the other three experimentally validated enhancer-related databases (VISTA Enhancer Browser, DiseaseEnhancer and ENdisease), and the results showed that 629 (85.3%) enhancers were uniquely collected in our database. The VISTA Enhancer Browser is an experimentally validated enhancer database, in which gene enhancer activity is accessed in transgenic mice (34). The VISTA Enhancer Browser group performed many transgenic mouse assays to determine the activity of putative enhancers based on candidate regions from seven publications. However, DiseaseEnhancer and ENdisease focus on manually curated enhancers associated with diseases from a large number of publications (22,23). Compared to DiseaseEnhancer and ENdisease, ENdb adopted a strict strategy to collect 737 (human, mouse) enhancers validated by low-throughput experiments from >1000 publications. Notably, the genomic location of enhancers must be retrievable from publications to be included in ENdb. Enhancer activity must also be directly validated by low-throughput



Figure 2. Database content and construction. (A) ENdb provides a user-friendly interface to browse, search, download, submit and visualize detailed information about enhancers. (B) The ‘Browse’ page allows users to quickly browse the information of enhancer and provides fuzzy search functions according to different conditions. (C) The ‘Detail’ page provides the general information of enhancer and the enhancer-associated information. The enhancer-associated information is graphically displayed in the page. (D) The ‘Search’ page allows users to search by a variety of keywords to return the records via four paths. Users also search the information of enhancers via the official symbols or aliases of genes. (E) Users view the proximity information of enhancers in the genome in the ‘Genome-Browse’ page.

experiments for inclusion in the ENdb. Many enhancers in DiseaseEnhancer were indirectly collected through mapping single nucleotide variants (SNVs) to enhancers identified by high-throughput experiments (22). SNV functions were validated by biological experiments. Additionally, ENdisease contains enhancers validated by both low and high-throughput experiments. However, the number of enhancers validated by low-throughput experiments is 434, which is less than that contained in ENdb. More impor-

tantly, compared to other databases, ENdb focuses on collecting experimentally-validated enhancers and associated information, including target genes, TFs, diseases and functions. Thus, ENdb has considerable potential to complement other enhancer-related databases.

Enhancers and super-enhancers, can drive high expression of genes involved in cell identity and disease. For example, super-enhancers frequently regulate lineage specific genes that control cell state and differentiation in so-

matic cells (47). In cancer cells, enhancers drive the expression of critical oncogenes, such as MYC (48), RUNX1 (49), MYOD (50) and BCL6 (51). Interestingly, negative regulators of enhancer-associated genes including the TFs CEBPA, IRF8, IRF1 and ETV6, can suppress cancer cell growths (52). Tsukiji *et al.* found that the expression of Shh was achieved through a novel regulatory element comprised of TFs and multiple enhancers in mouse embryonic endodermal tissues (53). Importantly, emerging evidence suggests that disease-associated SNPs are preferentially enriched in enhancers and can make essential contributions to disease progression (54). The causal SNP, rs58692659, in enhancer regions disrupts DNA-binding and enhancer activity (55). Together, these studies demonstrate the importance of enhancers and associated annotation information in human disease and biological processes. Our effort to establish ENdb was prompted by the great need of researchers to access a comprehensive dataset of experimentally validated enhancers and associated information, including genomic location, target genes, TFs, functions and diseases. Researchers in need of such a comprehensive database include cellular/molecular biologists, geneticists and data scientists. In the future, we will continue to manually curate newly validated enhancer associations and update the database. We will also consider continuously adding experimentally-supported enhancers from different species. We believe that ENdb can serve as a reliable and valuable resource of enhancers information to understand human diseases and biological processes.

FUNDING

National Natural Science Foundation of China [81572341, 61601150]; The Fundamental Research Funds for the Provincial Universities [JFXN201912]; Natural Science Foundation of Heilongjiang Province [JJ2016ZR1232/F2016031]; YuWeihan 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

- Korkmaz,G., Lopes,R., Ugalde,A.P., Nevedomskaya,E., Han,R., Myacheva,K., Zwart,W., Elkon,R. and Agami,R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 192–198.
- Weintraub,A.S., Li,C.H., Zamudio,A.V., Sigova,A.A., Hannett,N.M., Day,D.S., Abraham,B.J., Cohen,M.A., Nabet,B., Buckley,D.L. *et al.* (2017) YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, **171**, 1573–1588.
- Liu,W., Ma,Q., Wong,K., Li,W., Ohgi,K., Zhang,J., Aggarwal,A. and Rosenfeld,M.G. (2013) Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. *Cell*, **155**, 1581–1595.
- Rada-Iglesias,A., Bajpai,R., Swigut,T., Brugmann,S.A., Flynn,R.A. and Wysocka,J. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**, 279–283.
- Ernst,J., Melnikov,A., Zhang,X., Wang,L., Rogov,P. and Mikkelsen,T.S. (2016) Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat. Biotechnol.*, **34**, 1180–1190.
- Visel,A., Blow,M.J., Li,Z., Zhang,T., Akiyama,J.A., Holt,A., Plajzer-Frick,I., Shoukry,M., Wright,C., Chen,F. *et al.* (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, **457**, 854–858.
- Solberg,N. and Krauss,S. (2013) Luciferase assay to study the activity of a cloned promoter DNA fragment. *Methods Mol. Biol.*, **977**, 65–78.
- Caplen,N.J. and Mousset,S. (2003) Short interfering RNA (siRNA)-mediated RNA interference (RNAi) in human cells. *Ann. N. Y. Acad. Sci.*, **1002**, 56–62.
- Zhang,X., Choi,P.S., Francis,J.M., Imielinski,M., Watanabe,H., Cherniack,A.D. and Meyerson,M. (2016) Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.*, **48**, 176–182.
- Mochizuki,Y., Chiba,T., Kataoka,K., Yamashita,S., Sato,T., Kato,T., Takahashi,K., Miyamoto,T., Kitazawa,M., Hata,T. *et al.* (2018) Combinatorial CRISPR/Cas9 approach to elucidate a Far-Upstream enhancer complex for Tissue-Specific Sox9 expression. *Dev. Cell*, **46**, 794–806.
- Yakhnin,A.V., Yakhnin,H. and Babitzke,P. (2012) Gel mobility shift assays to detect protein-RNA interactions. *Methods Mol. Biol.*, **905**, 201–211.
- Ohmura,S., Mizuno,S., Oishi,H., Ku,C.J., Hermann,M., Hosoya,T., Takahashi,S. and Engel,J.D. (2016) Lineage-affiliated transcription factors bind the Gata3 Tce1 enhancer to mediate lineage-specific programs. *J. Clin. Invest.*, **126**, 865–878.
- Joo,J.Y., Schaukowitz,K., Farbiak,L., Kilaru,G. and Kim,T.K. (2016) Stimulus-specific combinatorial functionality of neuronal c-fos enhancers. *Nat. Neurosci.*, **19**, 75–83.
- Angrisano,T., Sacchetti,S., Natale,F., Cerrato,A., Pero,R., Keller,S., Peluso,S., Perillo,B., Avvedimento,V.E., Fusco,A. *et al.* (2011) Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. *Nucleic Acids Res.*, **39**, 1993–2006.
- Harmston,N. and Lenhard,B. (2013) Chromatin and epigenetic features of long-range gene regulation. *Nucleic Acids Res.*, **41**, 7185–7199.
- Sexton,T., Kurukuti,S., Mitchell,J.A., Umlauf,D., Nagano,T. and Fraser,P. (2012) Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat. Protoc.*, **7**, 1335–1350.
- ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Gao,T., He,B., Liu,S., Zhu,H., Tan,K. and Qian,J. (2016) EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics*, **32**, 3543–3551.
- Wang,J., Dai,X., Berry,L.D., Cogan,J.D., Liu,Q. and Shyr,Y. (2019) HACER: an atlas of human active enhancers to interpret regulatory variants. *Nucleic Acids Res.*, **47**, D106–D112.
- Wang,Z., Zhang,Q., Zhang,W., Lin,J.R., Cai,Y., Mitra,J. and Zhang,Z.D. (2018) HEDD: human enhancer disease database. *Nucleic Acids Res.*, **46**, D113–D120.
- Kang,R., Zhang,Y., Huang,Q., Meng,J., Ding,R., Chang,Y., Xiong,L. and Guo,Z. (2019) EnhancerDB: a resource of transcriptional regulation in the context of enhancers. *Database*, **2019**, bay141.
- Zhang,G., Shi,J., Zhu,S., Lan,Y., Xu,L., Yuan,H., Liao,G., Liu,X., Zhang,Y., Xiao,Y. *et al.* (2018) DiseaseEnhancer: a resource of human disease-associated enhancer catalog. *Nucleic Acids Res.*, **46**, D78–D84.
- Zeng,W., Min,X. and Jiang,R. (2019) EnDisease: a manually curated database for enhancer-disease associations. *Database*, **2019**, baz020.
- Ashoor,H., Kleftogiannis,D., Radovanovic,A. and Bajic,V.B. (2015) DENdb: database of integrated human enhancers. *Database*, **2015**, bav085.
- Abugessaisa,I., Noguchi,S., Carninci,P. and Kasukawa,T. (2017) The FANTOM5 computation ecosystem: genomic information hub for promoters and active enhancers. *Methods Mol. Biol.*, **1611**, 199–217.
- Kundaje,A., Meuleman,W., Ernst,J., Bilenky,M., Yen,A., Heravi-Moussavi,A., Kheradpour,P., Zhang,Z., Wang,J., Ziller,M.J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
- Cunningham,F., Achuthan,P., Akanni,W., Allen,J., Amode,M.R., Armean,I.M., Bennett,R., Bhai,J., Billis,K., Boddu,S. *et al.* (2019) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
- Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.

29. Wei, Y., Zhang, S., Shang, S., Zhang, B., Li, S., Wang, X., Wang, F., Su, J., Wu, Q., Liu, H. *et al.* (2016) SEA: a super-enhancer archive. *Nucleic Acids Res.*, **44**, D172–D179.
30. Jiang, Y., Qian, F., Bai, X., Liu, Y., Wang, Q., Ai, B., Han, X., Shi, S., Zhang, J., Li, X. *et al.* (2019) SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res.*, **47**, D235–D243.
31. 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*, **2017**, 1–17.
32. Keen, J.C. and Moore, H.M. (2015) The genotype-tissue expression (GTEx) project: linking clinical data with molecular analysis to advance personalized medicine. *J Pers Med*, **5**, 22–29.
33. 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*, **2019**, bay140.
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**, D88–92.
35. Gao, Y., Wang, P., Wang, Y., Ma, X., Zhi, H., Zhou, D., Li, X., Fang, Y., Shen, W., Xu, Y. *et al.* (2019) Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res.*, **47**, D1028–D1033.
36. Zhou, B., Zhao, H., Yu, J., Guo, C., Dou, X., Song, F., Hu, G., Cao, Z., Qu, Y., Yang, Y. *et al.* (2018) EVLncRNAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.*, **46**, D100–D105.
37. Sayers, E.W., Agarwala, R., Bolton, E.E., Brister, J.R., Canese, K., Clark, K., Connor, R., Fiorini, N., Funk, K., Hefferon, T. *et al.* (2019) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **47**, D23–D28.
38. Chou, C.H., Shrestha, S., Yang, C.D., Chang, N.W., Lin, Y.L., Liao, K.W., Huang, W.C., Sun, T.H., Tu, S.J., Lee, W.H. *et al.* (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **46**, D296–D302.
39. Kibbe, W.A., Arze, C., Felix, V., Mittra, E., Bolton, E., Fu, G., Mungall, C.J., Binder, J.X., Malone, J., Vasant, D. *et al.* (2015) Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.*, **43**, D1071–D1078.
40. Bhattacharya, S., Ha-Thuc, V. and Srinivasan, P. (2011) MeSH: a window into full text for document summarization. *Bioinformatics*, **27**, i120–128.
41. Yashiro-Ohtani, Y., Wang, H., Zang, C., Arnett, K.L., Bailis, W., Ho, Y., Knoechel, B., Lanauze, C., Louis, L., Forsyth, K.S. *et al.* (2014) Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E4946–E4953.
42. Yochum, G.S. (2011) Multiple Wnt/ss-catenin responsive enhancers align with the MYC promoter through long-range chromatin loops. *PLoS One*, **6**, e18966.
43. Cao, X., Yan, Z., Wu, Q., Zheng, A. and Zhong, S. (2018) GIVE: portable genome browsers for personal websites. *Genome Biol.*, **19**, 92.
44. Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elisk, C.G., Lewis, S.E., Stein, L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
45. Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J. *et al.* (2010) The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.*, **38**, D613–D619.
46. Fernandez, J.M., de la Torre, V., Richardson, D., Royo, R., Puiggras, M., Moncunill, V., Frangkogianni, S., Clarke, L., Flicek, P., Rico, D. *et al.* (2016) The BLUEPRINT data analysis portal. *Cell Syst.*, **3**, 491–495.
47. Loots, G.G., Kneissel, M., Keller, H., Baptist, M., Chang, J., Collette, N.M., Ovcharenko, D., Plajzer-Frick, I. and Rubin, E.M. (2005) Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Res.*, **15**, 928–935.
48. Qiu, M., Bao, W., Wang, J., Yang, T., He, X., Liao, Y. and Wan, X. (2014) FOXA1 promotes tumor cell proliferation through AR involving the Notch pathway in endometrial cancer. *BMC Cancer*, **14**, 78.
49. Markova, E.N., Kantidze, O.L. and Razin, S.V. (2011) Transcriptional regulation and spatial organisation of the human AML1/RUNX1 gene. *J. Cell Biochem.*, **112**, 1997–2005.
50. Mousavi, K., Zare, H., Dell’orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L. and Sartorelli, V. (2013) eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell*, **51**, 606–617.
51. Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G. *et al.* (2013) Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell*, **24**, 777–790.
52. Pelish, H.E., Liao, B.B., Nitulescu, I.I., Tangpeerachaikul, A., Poss, Z.C., Da Silva, D.H., Caruso, B.T., Arefolov, A., Fadeyi, O., Christie, A.L. *et al.* (2015) Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature*, **526**, 273–276.
53. Tsukiji, N., Amano, T. and Shiroishi, T. (2014) A novel regulatory element for Shh expression in the lung and gut of mouse embryos. *Mech. Dev.*, **131**, 127–136.
54. Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S.C., Erdos, M.R., Davis, S.R., Roychoudhuri, R., Restifo, N.P., Gadina, M. *et al.* (2015) Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*, **520**, 558–562.
55. Pasquali, L., Gaulton, K.J., Rodriguez-Segui, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J.J., Moran, I., Gomez-Marín, C., van de Bunt, M. *et al.* (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.*, **46**, 136–143.