*D172–D179* *Nucleic Acids Research, 2016, Vol. 44, Database issue* *Published online 17 November 2015*
*doi: 10.1093/nar/gkv1243*

# SEA: a super-enhancer archive

**Yanjun Wei[1,†], Shumei Zhang[1,†], Shipeng Shang[1,†], Bin Zhang[1], Song Li[1], Xinyu Wang[1], Fang Wang[1], Jianzhong Su[1], Qiong Wu[2], Hongbo Liu[1,*] and Yan Zhang[1,*]**

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China and [2]School of Life Science and Technology, State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150001, China

Received August 05, 2015; Revised October 28, 2015; Accepted October 30, 2015

## ABSTRACT

**Super-enhancers are large clusters of transcriptional enhancers regarded as having essential roles in driving the expression of genes that control cell identity during development and tumorigenesis. The construction of a genome-wide super-enhancer database is urgently needed to better understand super-enhancer-directed gene expression regulation for a given biology process. Here, we present a specifically designed web-accessible database, Super-Enhancer Archive (SEA, http://sea.edbc.org). SEA focuses on integrating super-enhancers in multiple species and annotating their potential roles in the regulation of cell identity gene expression. The current release of SEA incorporates 83 996 super-enhancers computationally or experimentally identified in 134 cell types/tissues/diseases, including human (75 439, three of which were experimentally identified), mouse (5879, five of which were experimentally identified), *Drosophila melanogaster* (1774) and *Caenorhabditis elegans* (904). To facilitate data extraction, SEA supports multiple search options, including species, genome location, gene name, cell type/tissue and super-enhancer name. The response provides detailed (epi)genetic information, incorporating cell type specificity, nearby genes, transcriptional factor binding sites, CRISPR/Cas9 target sites, evolutionary conservation, SNPs, H3K27ac, DNA methylation, gene expression and TF ChIP-seq data. Moreover, analytical tools and a genome browser were developed for users to explore super-enhancers and their roles in defining cell identity and disease processes in depth.**

## INTRODUCTION

Super-enhancers are genome regions that are large clusters of transcriptional enhancers (1). The term 'super-enhancer' was used for the first time by Chen *et al.* in 2004 (2). Super-enhancers differ from typical enhancers in size, transcription factor density and content, and the ability to activate transcription (3). As key gene regulators, super-enhancers are assembled in a cell type-specific manner, which is closely related to the binding of numerous transcription factors and the enrichment of H3K27ac (4,5).

Super-enhancers are regarded as the regulators of cell identity, stem cell pluripotency and even tumor pathogenesis (6–8). They have been reported to drive the cell type-specific expression of genes that define cell identity during development and disease (9,10), as shown by the super-enhancers of hair follicle stem cells that drove reporter gene expression *in vivo* (11). In another study, the CRISPR/Cas9 genome editing strategy revealed that a super-enhancer was responsible for over 90% of *Sox2* expression in mouse embryonic stem cells (12). Super-enhancers play important roles in normal development, although their aberrant assembly can enhance the abnormal expression of key genes and contribute to multiple diseases including Alzheimer's disease, type 1 diabetes mellitus and cancers (13). For example, the quintessential oncogene *MYC* is regulated by super-enhancers in most cancers (1,14–16), while the super-enhancers associated with genes encoding the transcription factors RUNX1, FOSL2 and BHLHE40 are critical for the mesenchymal transformation of brain tumors (17).

Several studies have identified novel cancer therapeutics directed at components of super-enhancers in diverse tumor types (18). For example, tumor oncogenes including *MYC* can be selectively inhibited by disrupting super-enhancers in cancers (18,19), while *CDK7* inhibition suppresses super-enhancer-linked oncogenic transcription (20–22). The deletion of super-enhancer constituents using a CRISPR/Cas9-based approach in colorectal cancer cells previously reduced the expression of the associated gene, suggesting that

*To whom correspondence should be addressed. Tel: +86 86669617; Fax: +86 86669617; Email: tyozhang@ems.hrbmu.edu.cn
Correspondence may also be addressed to Hongbo Liu. Tel: +86 86669617; Fax: +86 86669617; Email: hongbo919@gmail.com
†These authors contributed equally to this work as the first authors.

cancer therapeutics targeting super-enhancer components may be particularly effective in tumor cells (9).

At present, there are several methods for discovering super-enhancers, including computational method based on factors such as Med1 (19), BRD4 (10) and H3K27ac (1), and experimental technology such as ChIA-PET (8), CRISPR/Cas9 (9), RT-qPCR (23) and so on. Because of the larger size of super-enhancers, it was difficult and time consuming to experimentally identify novel super-enhancers from a range of cell types/tissues/diseases (12). It is known that H3K27ac can identify a large fraction of super-enhancers formed by the master transcription factors and it has been profiled in a broad range of samples from multiple species (4). Thus, H3K27ac can be used as the surrogate mark of super-enhancers and imported into ROSE software developed by Loven *et al.* (19) to identify the super-enhancers in a broad range of samples from multiple species. Briefly, it stitches constituent enhancers together if they are within a certain distance and ranks the enhancers by their input-subtracted signal of H3K27ac. Then super-enhancers are separated from typical enhancers by identifying a threshold point of H3K27ac signal versus enhancer rank (4). For example, H3K27ac was used to identify the super-enhancers in five human DLBCL lines and a normal tonsil sample (3). Zhou *et al.* also used H3K27ac signals to characteristic Epstein–Barr virus super-enhancers (5). These results indicate that H3K27ac is a robust marker of super-enhancers in various cell types, normal tissues and diseases. Recently, the development of several epigenome projects, such as ENCODE (24), modENCODE (25) and Human Epigenome Roadmap (26), have provided a genome-wide landscape of H3K27ac in many cell types/tissues/diseases. This raised the possibility of integrating these data sets to systematically identify a comprehensive map of super-enhancers in multiple species.

Therefore, the comprehensive collection and more and more prediction of candidate super-enhancers would be useful for the systemic exploration of their potential roles in human disease or biology processes (14,27). To this end, we developed a specifically designed database, Super-Enhancer Archive (SEA, http://sea.edbc.org), which focuses on integrating experimentally and computationally identified super-enhancers and annotating their potential roles in the regulation of cell identity gene expression in a cell type-specific manner. Recently, another database for super-enhancers dbSUPER at http://bioinfo.au.tsinghua.edu.cn/dbsuper/ was proposed by Khan *et al.* (28). In contrast to dbSUPER, SEA (i) currently stores integrated super-enhancers of four species, including humans, mice, *Drosophila melanogaster* and *Caenorhabditis elegans*; (ii) displays the cell type/tissue/disease specificity of super-enhancers by comparative analysis, which is useful for the analysis of the super-enhancer regulatory function; (iii) identifies related genes of super-enhancers based on the genome location shown in flexible online networks; (iv) identifies CRISPR/Cas9 target sites of super-enhancers, which is useful for further functional analysis of super-enhancers for that the Cas9 nucleases can be directed by short RNAs to induce precise cleavage at endogenous genomic loci in organism cells; (v) supplements the classification of publications related to super-enhancers according to research content; (vi) offers an online analytical tool for genomic region enrichment analysis and another for H3K27ac cell type specificity analysis; and (vii) develops a genome browser and supports customized and user-friendly views on a genomic scale. In brief, as a comprehensive super-enhancer resource, SEA provides a convenient platform to store, annotate and readily analyse super-enhancers.

## OVERVIEW AND DATA PROCESSING

In this study, we took advantage of the available public H3K27ac data sets and the published super-enhancer lists to construct a comprehensive list of super-enhancers in mutiple cell types/tissues/diseases of four species (Figure 1). First, we examined the super-enhancers identified by Hnisz *et al.* in 86 human cell types/tissues using the ROSE program and based on the H3K27ac ChIP-Seq data set (1). Second, for other cell types/tissues in humans and other species, we obtained the H3K27ac raw data set profiled by ChIP-Seq from the SRA database (29). Bowtie2 (version 2.2.5) (30) was used to map ChIP-Seq reads to reference genomes downloaded from UCSC Genome Bioinformatics (31), including those of human (hg19), mouse (mm9), *D. melanogaster* (dm6) and *C. elegans* (ce10). For each sample, ROSE was used to obtain potential super-enhancers as 'python ROSE_main.py -g hg19 -i /data/Human.gff -r /data/SRR*******_sort.bam -o /data/SRR******* -s 12500'. And it was run with a stitching distance of 12 500 bp, i.e. enhancers were allowed to be stitched together within 12 500 bp. Besides, we manually collected eight super-enhancers that had been identified and confirmed by experimental methods in published papers.

In this way, we incorporated 83 996 super-enhancers that had been experimentally discovered or computationally identified in 134 cell types, tissues, or diseases from humans, mice, *D. melanogaster* and *C. elegans* (Table 1). In the case of humans, there were 75 439 super-enhancers from 99 cell types/tissues including 30 disease cell lines/tissues. Among these human super-enhancers, 58 283 were predicted by Hnisz *et al.* in 86 human cell types/tissues based on the H3K27ac ChIP-Seq data set (1), 1109 were predicted by Loven *et al.* in three human cell types/tissues based on the Med1 ChIP-Seq data set (19), 629 were predicted by Kwiatkowski *et al.* in the human Jurkat cell based on the H3K27ac ChIP-Seq data set (21), 15 415 were predicted by SEA in 15 human cell types/tissues based on the H3K27ac ChIP-Seq data set and 3 had been confirmed experimentally (3,14). In the case of mice, there were 5879 super-enhancers from 21 cell types/tissues, including 5 that had been experimentally confirmed (8,9,12). For *D. melanogaster*, there were 1774 computationally identified super-enhancers from 11 tissues, and 904 computationally identified super-enhancers from three tissues of *C. elegans*. The detailed number of samples for each cell type/tissue/disease per species that used to identify super-enhancers by different researchers were listed in the 'Supplement Table'.

On the convenience of using of super-enhancers, its' IDs/names in SEA are coded using a specific convention: <reference genome>_<cell-

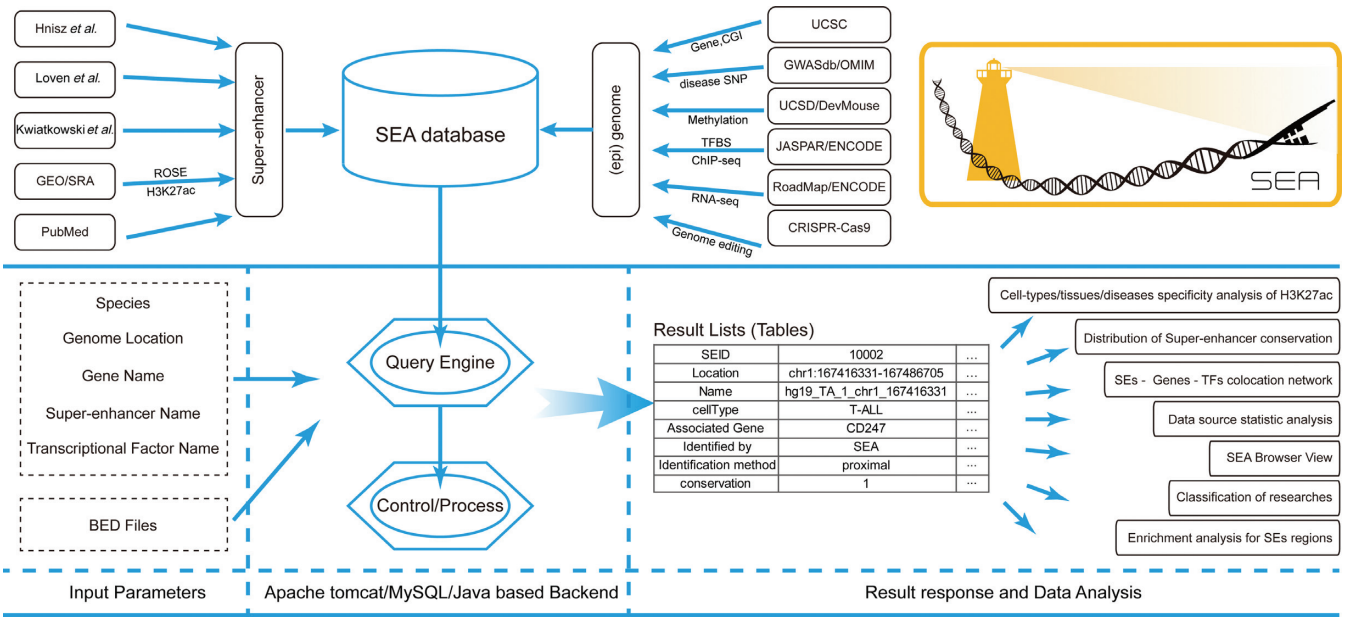


**Figure 1.** Database content and construction. SEA took advantage of the available public H3K27ac data sets to identify super-enhancers in different cell types/tissues/diseases of four species. And it also stores genetic and epigenetic information related to super-enhancers. Users can input multiple options to the query engine to acquire the super-enhancers. It enable users to do genomic region enrichment analysis and cell type specificity analysis. All search and analysis results can be downloaded as flat format for further analysis.

**Table 1.** SEA data content and statistics as of September 24, 2015

| | Total num. | Data source | Num. | Types num. | Human | Mouse | D. melanogaster | C. elegans |
|---|---|---|---|---|---|---|---|---|
| Super-Enhancer | 83 996 | Hnisz *et al.* (1) | 58 283 | 86 | 58 283 | 0 | 0 | 0 |
| | | Loven *et al.* (19) | 1109 | 3 | 1109 | 0 | 0 | 0 |
| | | Kwiatkowski *et al.* (21) | 629 | 1 | 629 | 0 | 0 | 0 |
| | | SEA (c) | 23 967 | 50 | 15 415 | 5874 | 1774 | 904 |
| | | SEA (e) | 8 | 4 | 3 | 5 | 0 | 0 |
| H3K27ac | 196 | SRA (29) | 122 | 98 | 122 | 0 | 0 | 0 |
| | | GEO (37) | 48 | 21 | 0 | 48 | 0 | 0 |
| | | modENCODE (25) | 26 | 14 | 0 | 0 | 15 | 11 |
| DNA Methylation | 26 | GEO (37) | 26 | 26 | 21 | 5 | 0 | 0 |
| Expression | 35 | Roadmap (26) | 20 | 20 | 20 | 0 | 0 | 0 |
| | | ENCODE (38) | 15 | 15 | 0 | 15 | 0 | 0 |
| Reference Genome | 4 | UCSC (31) | 4 | - | hg19 | mm9 | dm6 | ce10 |
| CpG Islands | 66 916 | UCSC (31) | 66 916 | - | 28 691 | 16 026 | 22 199 | 0 |
| SNP | 14 730 558 | GWASdb (39) | 314 237 | - | 314 237 | 0 | 0 | 0 |
| | | OMIM (40) | 148 | | 148 | 0 | 0 | 0 |
| | | UCSC (31) | 14 416 173 | | 14 416 173 | 0 | 0 | 0 |
| TFBS | 1 104 318 | JASPAR (41) | 1 104 318 | - | 803 489 | 267 980 | 13 357 | 19 492 |
| TF ChIP-seq data | 98 | ENCODE (38) | 98 | 11 | 98 | 0 | 0 | 0 |
| CRISPR/Cas9 | 1 211 142 | CRISPR Genome Engineering Resources (42) | 1 211 142 | - | 518 266 | 137 529 | 482 150 | 73 197 |

(c)Computationally identified super-enhancers by ROSE.
(e)Experimentally identified super-enhancers from literature search.

type/tissue/disease>_<number of repeat samples>_<chromosome>_<start site>. Underscores were used as the delimiters to clearly separate the items. With this convention, each super-enhancer has been coded like 'hg19_H1_chr3_181423059', which tells that this is a super-enhancer identified in human H1 cell. Additionally, the experimentally identified super-enhancers in SEA were marked by adding the suffix '*' behind the super-enhancers' name, taking 'hg19_GCBC_chr3_188892648*' for example.

In addition, SEA also stores genetic and epigenetic information related to super-enhancers, including sequence conservation, information about associated genes, CRISPR/Cas9 target sites, H3K27ac ChIP-seq data, DNA methylation, RNA-seq data, disease and common single-nucleotide polymorphisms, transcriptional factor binding sites and TF ChIP-seq data (Table 1).

The genes associated with experimentally discovered super-enhancer were defined by the context of published literature, while the genes associated with predicted super-

enhancers were assigned by its' closest gene referred to the description in Kwiatkowski *et al.*'s paper. In order to simplify the back-end database, the H3K27ac ChIP-seq data, DNA methylation, RNA-seq data and TF ChIP-seq data were pre-processed following these procedures: (i) extend to upstream and downstream in the center of the super-enhancer regions with 2-fold of their original length; (ii) merge the extended regions to larger regions; and (iii) calculate the methylation level or counts of ChIP-seq/RNA-seq reads falling within per 100 bp in the merged regions, which is used as the estimation of the signal of methylation/H3K27ac/TF/expression. Then, these data sets were designed as browser view tracks for visualization.

For each TF, the enrichment score were calculated following this formula below:

$$\text{enrichment score} = \frac{N(TFBS\ of\ TF\ in\ SE) * L(refgenome)}{L(SE) * N(TFBS\ of\ TF\ in\ refgenome)}$$

where N(TFBS of TF in SE) represents the number of TFBS referred to specific TF in the super-enhancer, L(refgenome) represents the length of the reference genome, L(SE) represents the length of super-enhancer and N(TFBS of TS in refgenome) represents the number of TFBS referred to the tested TF in the whole reference genome. Briefly, the enrichment score is used to depict the intensity of transcription factor binding sites along the super-enhancer regions. And it is suggested that the high enrichment score may be a possible active super-enhancer.

## DATABASE USE AND ACCESS

### Using the search tool to retrieve super-enhancers

The search tool supports accurate searching (Figure 2). Users can specify detailed information about super-enhancers such as species, genome location, gene name, cell type/tissue, super-enhancer name and transcription factor. The search tool results include genomic location, super-enhancer ID coded by SEA, associated gene, cell type specificity, sequence conservation and CRISPR/Cas9 target sites (Figure 2A). Users can obtain detailed information about a particular super-enhancer following the links appended in this table, while information about the means used to define the super-enhancer, its' reporter, nearby genes and transcription factors are given on the detailed results page (Figure 2B). Moreover, the comparison between cell types/tissues is shown in a data matrix. Links to the SEA-browser enable visualization of the information in a genomic context (Figure 2C). The search result also provides useful links to other gene annotation databases including NCBI Gene (32), GeneCards (33) and UniProt (34). In addition, the statistic page provides statistics charts (Figure 2D) and the download page (Figure 2E) provides the download links for all data sources used in SEA.

### Online analysis tools for super-enhancers

Using the advanced search for super-enhancers or the interface to build the bed file on the 'Data Analysis' page, users can obtain a '.bed' file (Figure 2F), which can be downloaded for future study or directly sent to the Genomic Regions Enrichment of Annotations Tool (GREAT) (35), for the functional significance analysis of these super-enhancers. Users can also submit genomic region(s) to SEA through the second tab control on the 'Data Analysis' page, which may be stored in a '.bed' file. Then, the submitted genomic regions were converted to super-enhancers' regions through back-end overlapping analysis. Next, the average histone modification level of H3K27ac in the overlapped super-enhancer regions were calculated for each selected cell type/tissue, with which the heat map (Figure 2G) was generated by the JavaScript tool 'FusionCharts'. The row and the column in the heat map represent a particular super-enhancer and a special cell type/tissue, respectively. In order to demonstrate the difference between super enhancer and other genome regions such as typical enhancer, the enrichment score calculating tool for TFs was embedded into the third tab control on the Data Analysis page. The background of the enrichment analysis is the whole genome or the genome regions given in a custom designed BED file by user. Then, SEA will calculate the enrichment score for each TF falling into the region of the specified super-enhancers.

### SEA-browser

The SEA-browser was developed to visualize super-enhancer information in a (epi)genomic context (Figure 2C). Users can observe the proximity of super-enhancers to nearby genes, transcription factor binding sites with their ChIP-seq peaks, and cell type-specific super-enhancers, CRISPR/Cas9 target sites, H3K27ac modification pattern, DNA methylation level, expression of RNA-seq data, as well as disease single-nucleotide polymorphisms. Users can also add or reduce data tracks by clicking the selector at the bottom of the page. In addition, a 'Custom Track' tool has been developed to allow users to upload their own sequencing data as well as visualization.

### A case application of SEA in the systematic analysis of super-enhancers close to human gene *SOX2*

On the Search page, we selected 'Human' as the species and input '*SOX2*' as the gene name (Figure 3A). After clicking the 'Search' button at the bottom, the search result page listed eight super-enhancers related to human *SOX2* in different cell types/tissues (Figure 3B), which enriched in nine molecular functions analyzed by GREAT (Figure 3C). Because *SOX2* is a key gene in stem cells (36), we selected the super-enhancer (SEID: 52416) in human stem cell H1. Clicking the blue-colored SEID (52416) revealed details of this super-enhancer on an information page (Figure 3D). This showed the location of two genes (*SOX2* and *SOX2-OT*) near the super-enhancer. The network and tables suggested the presence of five transcription factors with eight binding sites (including the well-known super-enhancer markers CEBP and MYOD) within this super-enhancer, of which some were essential for the transcription regulation of *SOX2*.

To examine the relationship between the super-enhancer and *SOX2*, we navigated to the SEA-Browser via the links provided on the search result page or the detailed informa-

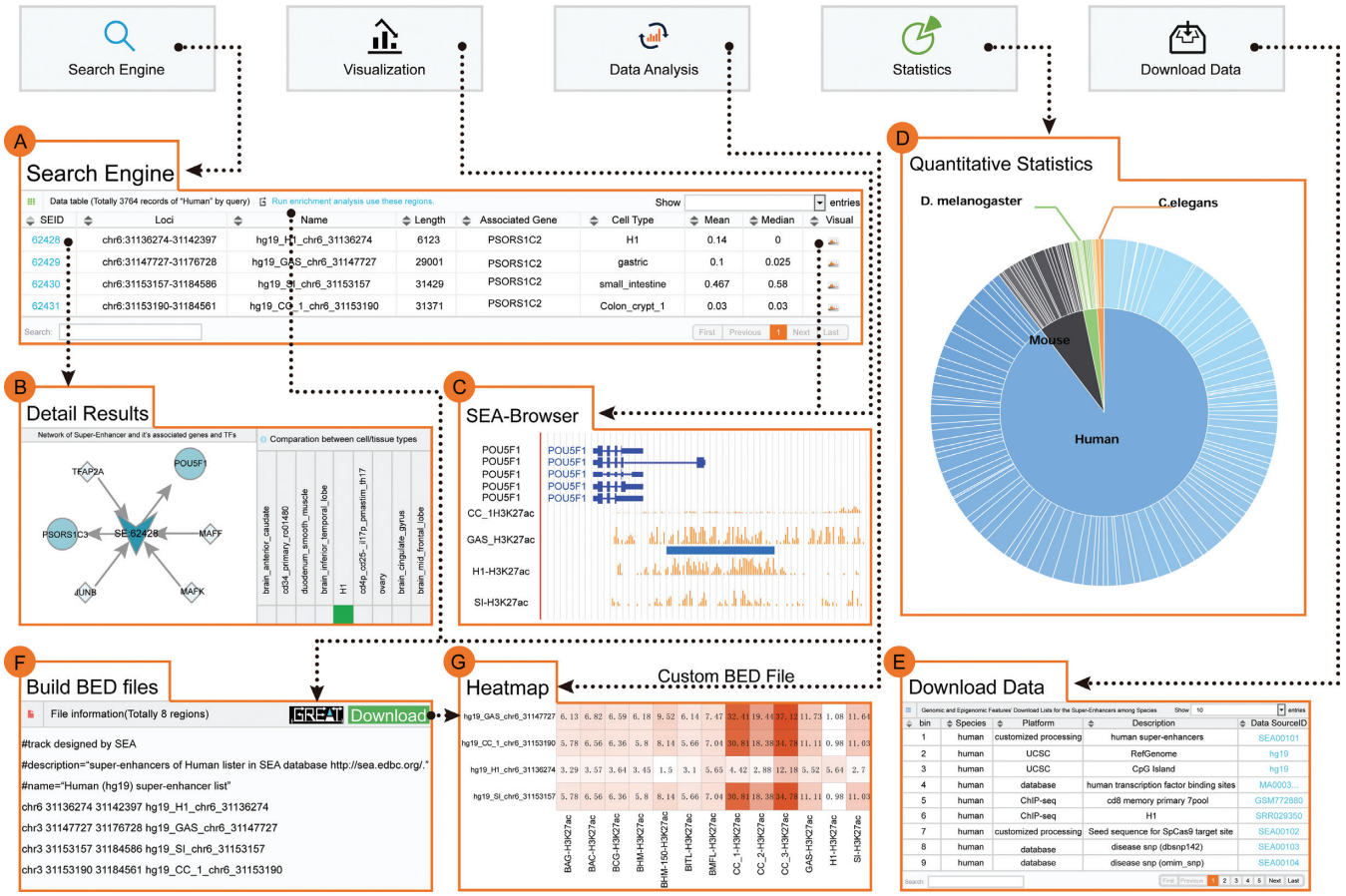


**Figure 2.** A comprehensive view of SEA query results. (**A**) The search tool results table, including SEID, super-enhancer's name, genomic location, associated gene, region length, cell types/tissues and conservative score. (**B**) Details of a selected super-enhancer from the results table showing the means used to define the super-enhancer, its' reporter, the nearby genes and transcription factors, together with the comparison among cell/tissue types. (**C**) Visualization of genetic and epigenetic information in the SEA-browser, such as H3K27ac modification pattern, DNA methylation level, expression of RNA-seq and so on. (**D**) Quantitative statistics of super-enhancers and the sources in SEA. (**E**) Data sources listed in the download table. (**F**) Genomic region enrichment analysis tool for super-enhancers. (**G**) Heat map result for the cell type specificity analysis of H3K27ac modification pattern. Each of the row and the column in the heat map represented a particular super-enhancer and a special cell type/tissue, respectively.

tion page. The genetic and epigenetic contexts of the super-enhancer were visualized within the SEA-Browser (Figure 3E), and showed that *SOX2* is localized in the super-enhancer. The track of super-enhancer shows the joint genomic regions of super-enhancers. In addition, single-base conservation score in the track named as 'Conservation' provided an immediately clear view for the conservative property of the super-enhancer region. The super-enhancer was also shown to include three CpG islands, which are key regulatory regions for DNA methylation. The SEA-Browser enabled us to check the cell type/tissue/disease specificity of the super-enhancer and its H3K27ac modification. The super-enhancer was shown to be stem cell specific and to be specifically enriched by H3K27ac, indicating its importance in the regulation of *SOX2* in stem cells. Using the link to GeneCards, we discovered more information about human *SOX2*. In short, SEA provided a search engine and tools for the gene-centric systematic analysis of super-enhancers and their roles in defining cell identity and disease.

## SYSTEM DESIGN AND IMPLEMENTATION

SEA was constructed based on three major software components: an Apache Tomcat web server, a MySQL relational database and Java-based computational services. The backstage processing programs were written in Java, which are available on request. The web services were developed using Apache Struts2, a Java web application framework, and iBATIS, a persistence framework that automates the mapping between MySQL databases and objects in Java, both of which help guarantee the high performance and stability of the web services. Browser-based interfaces were built in HTML/JSP and AJAX. The Apache Batik SVG toolkit was used to render, generate and manipulate the SVG dynamically. IE 9.0+ and Firefox with 1440*900 (or higher) are preferred to achieve the best display effect. SEA is available at http://sea.edbc.org.

## FUTURE DEVELOPMENT

To build a super-enhancer database focusing on multi-species, continued efforts will be made to update the SEA data and improve the SEA-browser and database func-

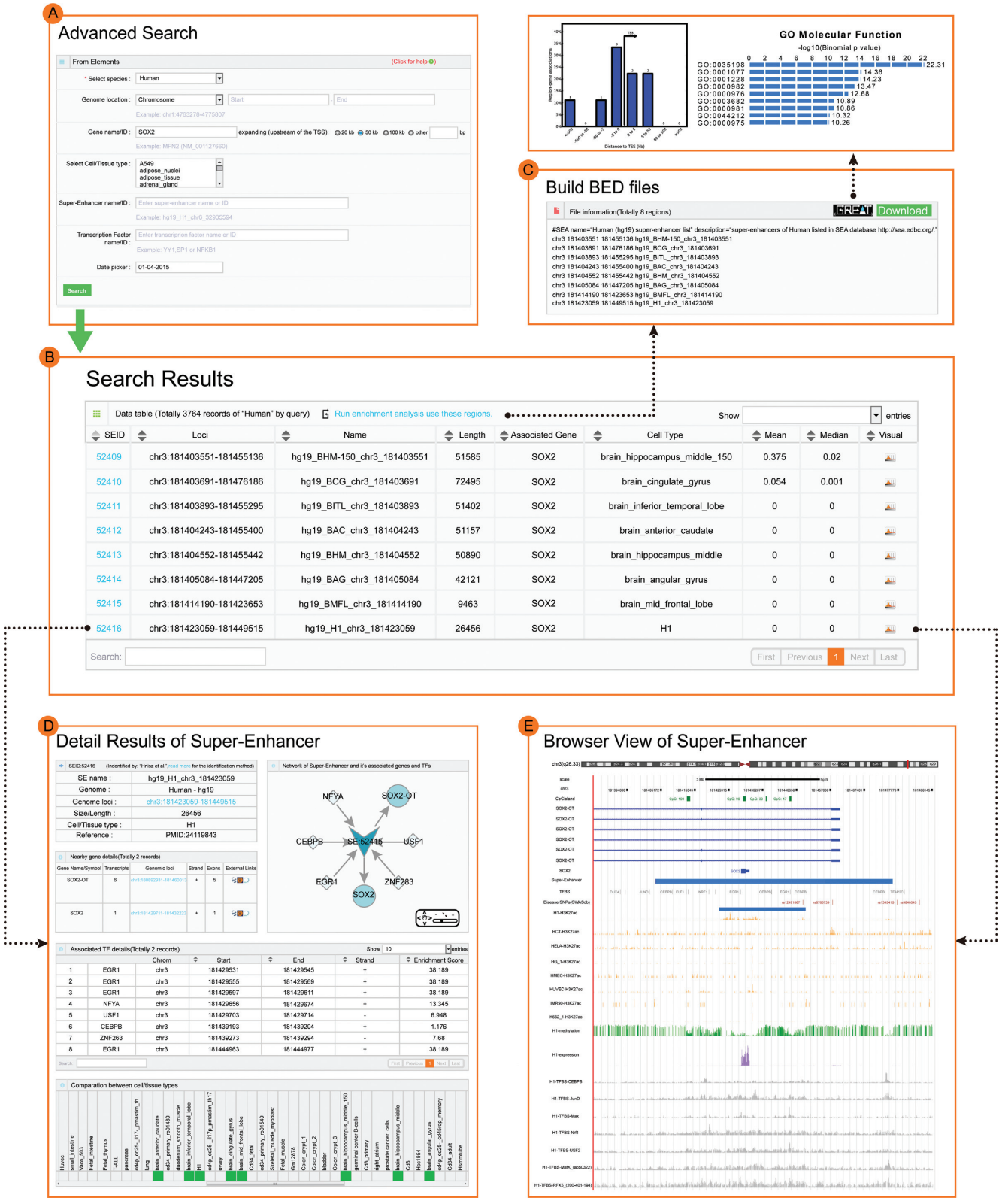


**Figure 3.** The usage of SEA by the example gene *SOX2*. (**A**) Input 'Human' as the species and '*SOX2*' as the gene name in the search form. (**B**) The search results related to human *SOX2* in different cell types/tissues. (**C**) Enrichment results of the eight super-enhancers from the search results analyzed by GREAT. (**D**) Details of the super-enhancer (SEID: 52416) in human stem cell H1 on an information page. (**E**) Visualization of human gene *SOX2* and super-enhancers nearby.

tionality. In next years, genome-wide super-enhancer information will become accumulating faster for many cell-types/tissues/diseases in more species, we will continuously collect the latest data sets to keep SEA up-to-date. Currently, nearly all super-enhancers in SEA were predicted by computational method based on H3K27ac. The super-enhancers identified or confirmed by experiment would be recorded into SEA. And more genetic and epigenetic information related to super-enhancer would be added into this database for further study. We expect that the continuous efforts to develop and improve SEA will contribute to our understanding of super-enhancers and their roles in defining cell identity and disease.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank the NIH Roadmap Epigenomics, ENCODE, modENCODE and other projects for generating and sharing the data used in this paper. We thank Richard A. Young and his colleagues for sharing ROSE program to this work.

## FUNDING

National Natural Science Foundation of China [31371334, 61403112, 81573021 and 61402139]; Natural Science Foundation of Heilongjiang Province [ZD2015003]; Innovation and Technology special Fund for researchers of Harbin Science and Technology Bureau [2015RAXXJ052]; Innovation Research Fund for Graduate Students of Harbin Medical University [YJSCX2014–23HYD]. Funding for open access charge: National Natural Science Foundation of China [31371334, 61403112, 81573021 and 61402139]; Natural Science Foundation of Heilongjiang Province [ZD2015003]; Innovation and Technology special Fund for researchers of Harbin Science and Technology Bureau [2015RAXXJ052]; Innovation Research Fund for Graduate Students of Harbin Medical University [YJSCX2014-23HYD].
*Conflict of interest statement.* None declared.

## REFERENCES

1. Hnisz,D., Abraham,B.J., Lee,T.I., Lau,A., Saint-Andre,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. Chen,Y., Yao,B., Zhu,Z., Yi,Y., Lin,X., Zhang,Z. and Shen,G. (2004) A constitutive super-enhancer: homologous region 3 of Bombyx mori nucleopolyhedrovirus. *Biochem. Biophys. Res. Commun.*, **318**, 1039–1044.
3. 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.
4. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
5. Ding,J., Huang,X., Shao,N., Zhou,H., Lee,D.F., Faiola,F., Fidalgo,M., Guallar,D., Saunders,A., Shliaha,P.V. *et al.* (2015) Tex10 Coordinates Epigenetic Control of Super-Enhancer Activity in Pluripotency and Reprogramming. *Cell Stem Cell*, **16**, 653–668.
6. Chen,K., Chen,Z., Wu,D., Zhang,L., Lin,X., Su,J., Rodriguez,B., Xi,Y., Xia,Z., Chen,X. *et al.* (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet.*, **47**, 1149–1157.
7. Qian,J., Wang,Q., Dose,M., Pruett,N., Kieffer-Kwon,K.R., Resch,W., Liang,G., Tang,Z., Mathe,E., Benner,C. *et al.* (2014) B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell*, **159**, 1524–1537.
8. Dowen,J.M., Fan,Z.P., Hnisz,D., Ren,G., Abraham,B.J., Zhang,L.N., Weintraub,A.S., Schuijers,J., Lee,T.I., Zhao,K. *et al.* (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*, **159**, 374–387.
9. Hnisz,D., Schuijers,J., Lin,C.Y., Weintraub,A.S., Abraham,B.J., Lee,T.I., Bradner,J.E. and Young,R.A. (2015) Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol. Cell*, **58**, 362–370.
10. Di Micco,R., Fontanals-Cirera,B., Low,V., Ntziachristos,P., Yuen,S.K., Lovell,C.D., Dolgalev,I., Yonekubo,Y., Zhang,G., Rusinova,E. *et al.* (2014) Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.*, **9**, 234–247.
11. Adam,R.C., Yang,H., Rockowitz,S., Larsen,S.B., Nikolova,M., Oristian,D.S., Polak,L., Kadaja,M., Asare,A., Zheng,D. *et al.* (2015) Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature*, **521**, 366–370.
12. Brown,J.D., Lin,C.Y., Duan,Q., Griffin,G., Federation,A.J., Paranal,R.M., Bair,S., Newton,G., Lichtman,A.H., Kung,A.L. *et al.* (2014) NF-kappaB directs dynamic super enhancer formation in inflammation and atherogenesis. *Mol. Cell*, **56**, 219–231.
13. Mansour,M.R., Abraham,B.J., Anders,L., Berezovskaya,A., Gutierrez,A., Durbin,A.D., Etchin,J., Lawton,L., Sallan,S.E., Silverman,L.B. *et al.* (2014) Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*, **346**, 1373–1377.
14. Herranz,D., Ambesi-Impiombato,A., Palomero,T., Schnell,S.A., Belver,L., Wendorff,A.A., Xu,L., Castillo-Martin,M., Llobet-Navas,D., Cordon-Cardo,C. *et al.* (2014) A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.*, **20**, 1130–1137.
15. Affer,M., Chesi,M., Chen,W.D., Keats,J.J., Demchenko,Y.N., Tamizhmani,K., Garbitt,V.M., Riggs,D.L., Brents,L.A., Roschke,A.V. *et al.* (2014) Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia*, **28**, 1725–1735.
16. Tolani,B., Gopalakrishnan,R., Punj,V., Matta,H. and Chaudhary,P.M. (2014) Targeting Myc in KSHV-associated primary effusion lymphoma with BET bromodomain inhibitors. *Oncogene*, **33**, 2928–2937.
17. Carro,M.S., Lim,W.K., Alvarez,M.J., Bollo,R.J., Zhao,X., Snyder,E.Y., Sulman,E.P., Anne,S.L., Doetsch,F., Colman,H. *et al.* (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature*, **463**, 318–325.
18. Mottok,A. and Gascoyne,R.D. (2015) Bromodomain inhibition in diffuse large B-cell lymphoma–giving MYC a brake. *Clin. Cancer Res.*, **21**, 4–6.
19. Loven,J., Hoke,H.A., Lin,C.Y., Lau,A., Orlando,D.A., Vakoc,C.R., Bradner,J.E., Lee,T.I. and Young,R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
20. Chipumuro,E., Marco,E., Christensen,C.L., Kwiatkowski,N., Zhang,T., Hatheway,C.M., Abraham,B.J., Sharma,B., Yeung,C., Altabef,A. *et al.* (2014) CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell*, **159**, 1126–1139.
21. Kwiatkowski,N., Zhang,T., Rahl,P.B., Abraham,B.J., Reddy,J., Ficarro,S.B., Dastur,A., Amzallag,A., Ramaswamy,S., Tesar,B. *et al.* (2014) Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature*, **511**, 616–620.
22. Christensen,C.L., Kwiatkowski,N., Abraham,B.J., Carretero,J., Al-Shahrour,F., Zhang,T., Chipumuro,E., Herter-Sprie,G.S., Akbay,E.A., Altabef,A. *et al.* (2014) Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor. *Cancer Cell*, **26**, 909–922.

23. Li,Y., Rivera,C.M., Ishii,H., Jin,F., Selvaraj,S., Lee,A.Y., Dixon,J.R. and Ren,B. (2014) CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PloS One*, **9**, e114485.
24. ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
25. Celniker,S.E., Dillon,L.A., Gerstein,M.B., Gunsalus,K.C., Henikoff,S., Karpen,G.H., Kellis,M., Lai,E.C., Lieb,J.D., MacAlpine,D.M. *et al.* (2009) Unlocking the secrets of the genome. *Nature*, **459**, 927–930.
26. Chadwick,L.H. (2012) The NIH Roadmap Epigenomics Program data resource. *Epigenomics*, **4**, 317–324.
27. Yin,J.W. and Wang,G. (2014) The Mediator complex: a master coordinator of transcription and cell lineage development. *Development*, **141**, 977–987.
28. Khan,A. and Zhang,X. (2015) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, doi:10.1093/nar/gkv1002 .
29. Kodama,Y., Shumway,M., Leinonen,R. and International Nucleotide Sequence Database Collaboration. (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–D56.
30. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
31. Karolchik,D., Barber,G.P., Casper,J., Clawson,H., Cline,M.S., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M. *et al.* (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–D770.
32. Brown,G.R., Hem,V., Katz,K.S., Ovetsky,M., Wallin,C., Ermolaeva,O., Tolstoy,I., Tatusova,T., Pruitt,K.D., Maglott,D.R. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
33. Safran,M., Dalah,I., Alexander,J., Rosen,N., Iny Stein,T., Shmoish,M., Nativ,N., Bahir,I., Doniger,T., Krug,H. *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database*, baq020.
34. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
35. 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.
36. Qin,Y., Zhou,C., Wang,N., Yang,H. and Gao,W.Q. (2015) Conversion of Adipose Tissue-Derived Mesenchymal Stem Cells to Neural Stem Cell-Like Cells by a Single Transcription Factor, Sox2. *Cell. Reprogramming*, **17**, 221–226.
37. 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.
38. ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
39. Li,M.J., Wang,P., Liu,X., Lim,E.L., Wang,Z., Yeager,M., Wong,M.P., Sham,P.C., Chanock,S.J. and Wang,J. (2012) GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **40**, D1047–D1054.
40. Amberger,J.S., Bocchini,C.A., Schiettecatte,F., Scott,A.F. and Hamosh,A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
41. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.Y., Chou,A., Ienasescu,H. *et al.* (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **42**, D142–D147.
42. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.