Published online 11 November 2021 *Nucleic Acids Research, 2022, Vol. 50, Database issue* ***D371–D379***
https://doi.org/10.1093/nar/gkab1032

scEnhancer: a single-cell enhancer resource with annotation across hundreds of tissue/cell types in three species

Tianshun Gao [1,2,*], Zilong Zheng [1], Yihang Pan [1,2], Chengming Zhu [2], Fuxin Wei [3], Jinqiu Yuan [1,2], Rui Sun [1,2], Shuo Fang [1,4], Nan Wang [2], Yang Zhou [1] and Jiang Qian [5,6]

[1]Big Data Center, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen 518107, P.R. China, [2]Scientific Research Center, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen 518107, P.R. China, [3]Department of Orthopaedics, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen 518107, P.R. China, [4]Department of Oncology, The Seventh Affiliated Hospital of Sun Yat-sen University, Shenzhen 518107, P.R. China, [5]The Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD 21231, USA and [6]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA

Received August 15, 2021; Revised October 04, 2021; Editorial Decision October 13, 2021; Accepted October 19, 2021

ABSTRACT

Previous studies on enhancers and their target genes were largely based on bulk samples that represent 'average' regulatory activities from a large population of millions of cells, masking the heterogeneity and important effects from the sub-populations. In recent years, single-cell sequencing technology has enabled the profiling of open chromatin accessibility at the single-cell level (scATAC-seq), which can be used to annotate the enhancers and promoters in specific cell types. A comprehensive resource is highly desirable for exploring how the enhancers regulate the target genes at the single-cell level. Hence, we designed a single-cell database scEnhancer (http://enhanceratlas.net/scenhancer/), covering 14 527 776 enhancers and 63 658 600 enhancer-gene interactions from 1 196 906 single cells across 775 tissue/cell types in three species. An unsupervised learning method was employed to sort and combine tens or hundreds of single cells in each tissue/cell type to obtain the consensus enhancers. In addition, we utilized a cis-regulatory network algorithm to identify the enhancer-gene connections. Finally, we provided a user-friendly platform with seven useful modules to search, visualize, and browse the enhancers/genes. This database will facilitate the research community towards a functional analysis of enhancers at the single-cell level.

INTRODUCTION

The DNA cis-regulatory elements, such as distal enhancers and promoters, determine the transcriptional regulation of tissue/cell type-specific gene expression in development, embryogenesis, immunity, homeostasis and diseases (1–5). To date, the annotation for enhancers or super-enhancers increased rapidly through many large-scale resources, including EnhancerAtlas, SEA, Endb, CancerEnD, SEdb, HACER, RAEdb, HEDD, DiseaseEnhancer, GeneHancer, DENdb, dbSUPER and VISTA (6–19). These resources annotate enhancers from bulk datasets and measured only average enhancer activities in large populations of cells, masking the heterogeneity and key effects among and within the sub-populations (e.g. sub-cell types) containing small numbers of cells (3,20–22) (Supplementary Figure S1). Since enhancers are tissue/cell type-specific, enhancer identification based on the single-cell level can better reveal the cellular specificity to determine the differences of gene expression on phenotypes across tissue/cell types (23,24). Therefore, single-cell resources are ideal for exploring how the enhancers regulate the target genes at ultra-high resolution in an accurate cell type-specific manner.

Assay of Transposase-Accessible Chromatin using sequencing (ATAC-seq) is a powerful tool for epigenomic profiling of cell type-specific chromatin accessibility (25). It was reported that at least 50% and around 25% of the bulk ATAC-seq peaks fell into the enhancer and promoter regions, respectively (26). Thus, the peaks called from ATAC-seq mainly represent cis-regulatory elements, including enhancers and promoters, and can be used to annotate tissue/cell type-specific enhancers or promoters (22,27,28). At the single-cell level, scATAC-seq studies have been

*To whom correspondence should be addressed. Tel: +86 18126408738; Fax: +86 0755 81206211; Email: gts.hust@gmail.com

widely applied to identify cell type-specific enhancers or enhancer–promoter interactions (1,3,4,29–34). Especially, accessible sites in the human genome identified by scATAC-seq displayed a high overlap with 75% of experimentally validated active enhancers in VISTA (3,9). Using the Graphical LASSO, a single-cell cis-regulatory network algorithm, Cicero, was developed (33) and enabled identification of genome-wide enhancer–promoter connections on a large scale (1,3,4,29–32,34). For example, Domcke *et al* utilized Cicero to identify 6.3 million unique pairs of cis-regulatory elements in 54 human cell types (3). These indicate that scATAC-seq may be an ideal single-cell sequencing technique for annotating enhancers and enhancer–promoter interactions.

Here, we constructed a single-cell enhancer database, scEnhancer, based on an improved unsupervised learning approach previously developed in our bulk enhancer database, EnhancerAtlas 2.0 (6). This method was used to integrate many genomic datasets to derive a consensus annotation of enhancers. It displayed several characteristics: (i) it was based on a well-designed score voting strategy for ranking and combining a large set of unlabelled data (7,35); (ii) we replaced the Pearson correlation with the Jaccard index, which was appropriate for the binary nature of scATAC-seq data, for computing similarity among all single-cell datasets (36); (iii) in contrast to the only 12 independent high-throughput datasets used in EnhancerAtlas 2.0, the new method could process tens or hundreds of single-cell datasets with different qualities as measured by the average number of fragments per cell; (iv) the Cicero results were used as a filtering condition for identification of the final single-cell enhancers (33). We also leveraged Cicero to generate enhancer–promoter connections of high quality (1,3,4,29–32,34). In some aspects, as a comprehensive single-cell enhancer resource, scEnhancer possesses tremendous advantages: (i) it has profiled 1 196 906 single cells and annotated a total of 14 527 776 enhancers in 775 tissue/cell types across three species; (ii) a suitable combination of an improved unsupervised learning method and a cis-network algorithm was applied to identify the enhancers and enhancer-gene interactions and (iii) a user-friendly platform with seven functional modules and the browser options were designed for searching, visualizing, drawing, and browsing enhancer or enhancer–promoter profiles. These will facilitate the analysis of enhancers at the single-cell level for the research community.

MATERIALS AND METHODS

Single-cell data collection and integration

To identify single-cell enhancers, we collected raw or processed (e.g. by cellranger) single-cell datasets with peak and tissue/cell-type annotations from several scATAC-seq resources, including the NCBI GEO datasets (37), Signac analysis with 10X Genomics data (38), DESCARTES (3), LungMap (39), Mouse sci-ATAC-seq Atlas (4), Fly ATAC Atlas (1) and MPAL-Single-Cell-2019 (29). All the samples in human, mouse, and fly were mapped to genome builds GRCh37/hg19, GRCm37/mm9 and BDGP5/dm3, respectively, by liftOver (40).

Integration of tissue/cell-type specific binary matrix

To obtain the tissue/cell-type specific binary matrix, we first converted the processed or raw scATAC-seq data into a large standard matrix with labelled cell types. In most scATAC-seq projects, the single-cell data are usually presented in many file formats, such as MatrixMarket (4), h5 (38), txt (41), RangedSummarizedExperiment (42), Seurat RDS (3), or even the raw fastq files (43). Here, we used the functions in R language and cellranger-atac 1.2.0 (30) to transform these different formats of single-cell data into large standard matrices for the subsequent extraction of small cell-type specific matrices (Supplementary Table S1). Data in MatrixMarket or h5 formats were transformed into the standard matrix via 'Matrix::readMM' and 'Read10X_h5' in Signac (38) while the txt datasets could be read as data.frame and then converted into the matrices. For the RangedSummarizedExperiment dataset, we first parsed its R structure to obtain the matrix, peaks, cell barcodes and cell type annotations from its 'assay', 'colnames', 'colData', 'rowRanges' slots (42). Analogously, we parsed the dataset in Seurat RDS and extracted the matrix with peak, barcode, and cell type information from the slots 'GetAssayData', 'assays RNA/peaks' and 'meta.data', respectively (3). The raw fastq single-cell datasets could be transformed into a large standard matrix by cellranger-atac (30) that processes peak and cell callings into MatrixMarket or h5 formats (43). Finally, we removed the irregular datasets with fewer than 200 peaks. Signac and cellranger-atac tools can be downloaded and installed from: https://satijalab.org/signac/ and https://support.10xgenomics.com/single-cell-atac/software/downloads/1.2.

After integrating the large standard matrix containing barcodes of all mixed cell types, we extracted cell-type specific single-cell datasets from the matrix based on the cell type annotation information in the metadata. To make the single-cell datasets comparable, we binarized them to normalize each dataset. In each dataset, the peak signal was set to '1' ('open') for at least one read and '0' ('closed') otherwise in the absence of reads (36). Thus, all the single-cell datasets for one tissue/cell type were merged and consolidated into a binary cell-type matrix. In this cell-type matrix, single cells (i.e. columns) with less than 200 peaks (i.e. rows) were removed, as well as the peaks without any signal in all single cells or in uncommon chromosomes (e.g. chr1_random). We also removed the cell types with <50 single cells. Genomic coordinates of peaks with other genome builds in human and mouse were converted by liftOver to GRCh37/hg19 and GRCm37/mm9, respectively (40).

Generation of consensus single-cell enhancers

We designed an improved unsupervised method to identify the bulk consensus enhancers from 12 types of independent datasets in EnhancerAtlas 2.0 (6). Here, we modified the method to determine the weights of hundreds of single-cells and combine them to generate consensus single-cell enhancers. On average, there were hundreds of single cells per cell type in our integrated data (Table 1). For these datasets we hypothesized that one dataset was high-quality if it was highly correlated with the other datasets and low-quality otherwise (7). Since single-cell datasets are binary,

to weigh more on the 'open' peaks across the whole genome rather than the 'close' peaks, we applied the Jaccard coefficient that had been used for scATAC-seq clustering analysis (36,44), to measure the correlation between any two single cells (e.g. $c_i$and c_j) based on the overlapping degree in open chromatin regions:

$$J_{C_iC_j} = \frac{|C_i \cap C_j|}{|C_i \cup C_j|} = \frac{|C_i \cap C_j|}{|C_i| + |C_j| - |C_i \cap C_j|}$$

where $|C_i|$, $|C_j|$ and $|C_i \cap C_j|$ represent the number of 'open' peaks in C_i, C_j and their overlap, respectively. For a cell type with n single-cell datasets, a Jaccard similarity matrix for all combined datasets was integrated as:

$$\begin{bmatrix} J_{C_1C_1} & \cdots & J_{C_1C_i} & \cdots & J_{C_1C_n} \\ \vdots & & \vdots & & \vdots \\ J_{C_iC_1} & \cdots & J_{C_iC_i} & \cdots & J_{C_iC_n} \\ \vdots & & \vdots & & \vdots \\ J_{C_nC_1} & \cdots & J_{C_nC_i} & \cdots & J_{C_nC_n} \end{bmatrix}$$

Using the Jaccard matrix, we measured the weight of any single-cell dataset C_i as:

$$w_{C_i} = \frac{\sum_{j=1}^{n} J_{C_iC_j}}{\sum_{j=1,k=1}^{n} J_{C_jC_k}} \quad (j,k \in [1,n],\ j \neq i,\ j \neq k)$$

By combining all single cells into one cell type, the signal score of any consensus single-cell peak i could be defined as:

$$S_{consensus}(i) = \sum_{j=1}^{n} w_{C_j} S_{C_j}(i)$$

where the $S_{C_j}(i)$ represents the signal score of the peak i in the single-cell dataset C_j.

Furthermore, we removed the single-cell peaks overlapping with promoter or exon regions. We also used the experimentally validated silencers in SilencerDB (45) as a key filter to remove the single-cell peaks that overlapped with silencers. Finally, one consensus single-cell peak should satisfy two the requirements: (i) The signal of single-cell peak should be larger than 95% of the random signals calculated by shuffling the peaks in each single cell; (ii) Cicero connection score ($\geq$0.1) is required to display the interaction of single-cell peak with at least one gene promoter.

To evaluate the accuracy of single-cell enhancers identified by this approach, we extracted the experimentally validated active enhancers from the VISTA database (9) as the gold standard. We compared them with bulk enhancers from the EnhancerAtlas 2.0 (6) in four human tissue/cell types. For single-cell or bulk enhancers in one cell type, the ones that overlapped with VISTA enhancers were classified as the positives while the others remained as the negatives. The sensitivity and specificity of the single-cell or bulk enhancers were computed on a base pair basis. We used the area under the receiver operating characteristic (AROC) to evaluate the performance for single-cell or bulk enhancers. The results showed that single-cell enhancers in brain, heart, eye and cranial nerve had much more overlaps with VISTA enhancers than the ones in bulk enhancers, as well as an average higher performance measured by AROC than bulk enhancers (Supplementary Figure S2). This indicated that single-cell enhancers were more accurately annotated than bulk enhancers.

Identification of enhancer–promoter interactions in scATAC-seq data

To identify the enhancer–promoter interactions in cell types, we employed the single-cell *cis*-regulatory network tool, Cicero, which had been widely used in many scATAC-seq projects to identify all the distal elements (e.g. enhancers)–promoter connections on a genome-wide basis (3,4,29–34). Because the scATAC-seq binary data for each tissue/cell type are extremely sparse, it is difficult to make accurate estimates of the co-accessibility score of chromatin accessibility loci with no normalization of the matrix (33,36). To successfully use Cicero, we transformed the binary matrix into a term frequency-inverse document frequency (TF-IDF) matrix using the latent semantic indexing (LSI) method for aggregating similar single cells to obtain denser counts in each peak (1,4). For each cell type, the binary count matrix M was converted into TF-IDF matrix as following:

$$M_{TF} = t(t(M)/colSums(M))$$

$$IDF = log(1 + ncol(M)/rowSums(M))$$

$$M_{TF-IDF} = IDF \times M_{TF}$$

where $colSums(M)$ and $rowSums(M)$ represent the sum of each column or row of the matrix M,respectively, while $col(M)$ is the number of columns in M.

We then performed the Singular Value Decomposition (SVD) on the transformed TF-IDF matrix to reduce the dimensionality (1,4). Since the first dimension was highly correlated with the read depth, only the 2nd to 50th dimensions were passed to UMAP for 2D visualisation. Finally, Cicero used the UMAP coordinates and normalization matrix as the standard input to calculate the co-accessibility scores among peaks within a limited distance in DNA by the Graphical LASSO algorithm. Applying Cicero to each cell type across the three species with a cut-off of value >0.1, we annotated 63 658 600 enhancer–promoter connections involving 4 942 303 promoters, and 14 527 776 enhancers across 775 tissue/cell types in human, mouse and fly.

Implementation of scEnhancer

We developed a powerful web server, scEnhancer, for single-cell analysis of enhancers and enhancer–promoter interactions. scEnhancer adopted the Linux CentOS7 with a new web configuration of nginx (1.20.1)-php (5.4.16)-mySQL (5.7.34) in a new web configuration to build the website. In addition, we employed perl (5.16.3) for the fast processing of text files with large data. Moreover, the HTML5 Canvas API and Javascript with a drawing module were utilized together to establish a genome browser for displaying enhancer/gene distributions for single-cell or consensus datasets of tissue/cell types. We also set up a two-handle

Table 1. The numbers of tissue/cell types, single-cell enhancers, single-cell datasets, average peaks per cell, single-cell promoters and enhancer–gene interactions for three species

	Homo sapiens	*Mus musculus*	*Drosophila melanogaster*	Total
Tissue/cell types	543	185	47	775
Enhancers	1 13 73 862	26 94 616	4 59 298	1 45 27 776
Single cells	10 47 052	1 30 481	19 373	11 96 906
Average peaks per cell	6212	5371	3427	5003
Promoters	39 05 212	9 42 549	94 542	49 42 303
Enhancer-gene interactions	5 08 61 554	1 14 58 766	13 38 280	6 36 58 600

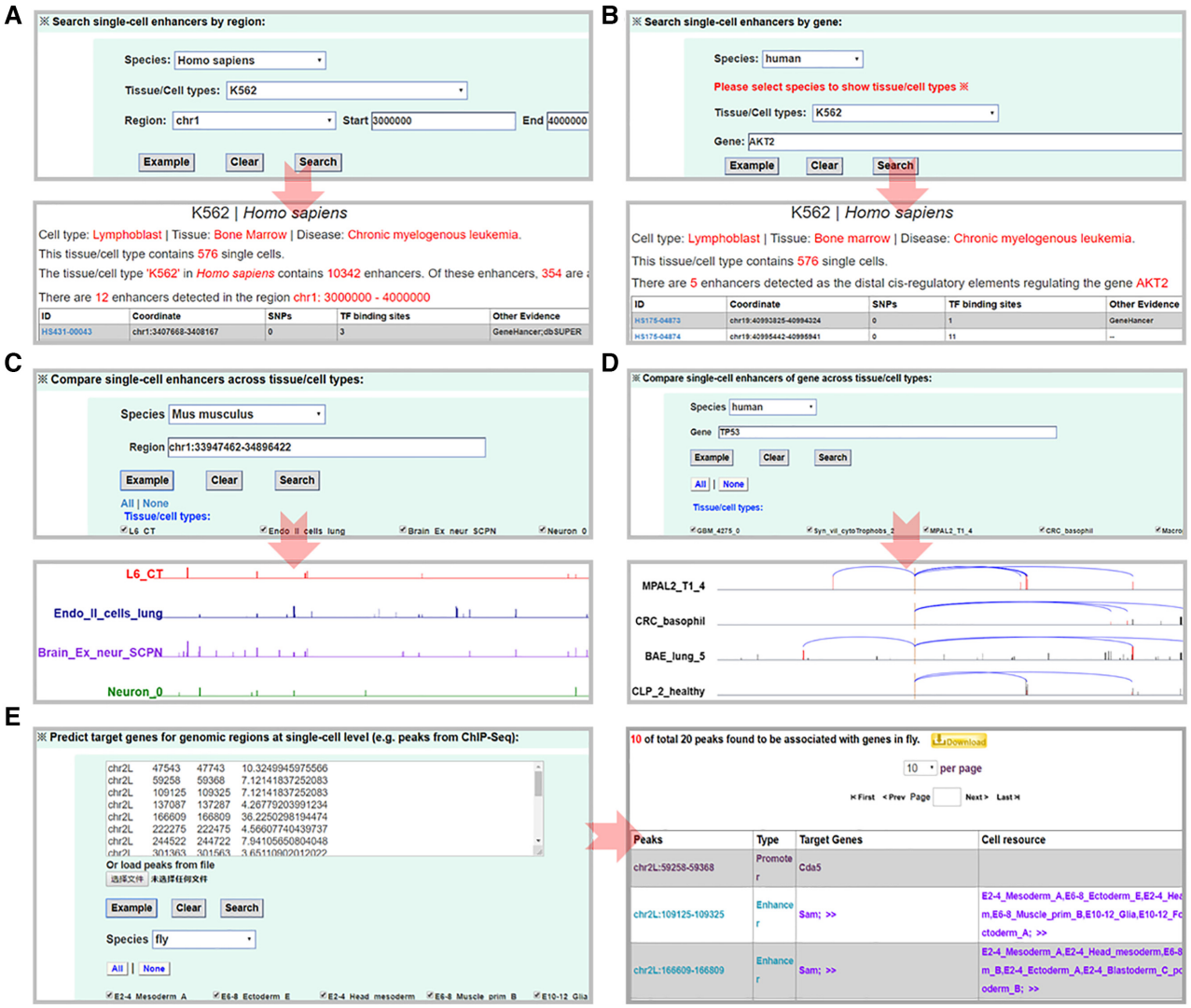
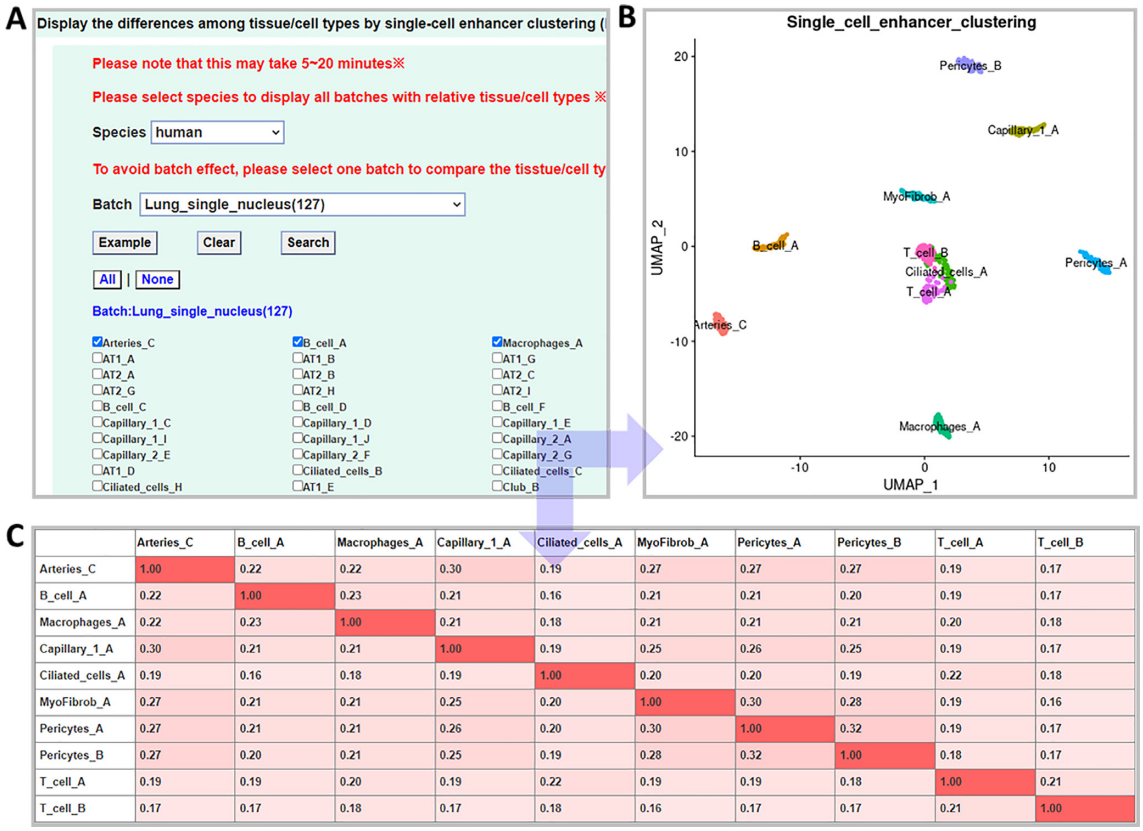


Figure 1. Simple search options. (**A**) Searching for single-cell enhancers by the input of a genomic region. (**B**) Querying for the enhancers around the input target gene. (**C**) Searching and comparing the distribution of single-cell enhancers across the selected tissue/cell types. (**D**) Fixing the target gene and comparing the enhancers regulating the gene across the selected tissue/cell types. (**E**) Checking whether the input genomic regions were single-cell enhancers, promoters or not in known tissue/cell types.

slider in Canvas to scale the genomic regions. Especially, using several packages including Signac (1.2.1), Seurat (4.0.0), and ggplot2 (3.3.5) in R language (4.0.3), we successfully designed a powerful module with a function of online plotting capabilities to graphically display the differences among any selected group of tissue/cell types by single-cell clustering analysis. Several useful analytic tools on the homepage were available for users to compare single-cell enhancers at different levels. The current version of scEnhancer can run in Windows, Mac and Linux systems and supports common web browsers such as Google, Safari, Microsoft Edge and Firefox.



	Arteries_C	B_cell_A	Macrophages_A	Capillary_1_A	Ciliated_cells_A	MyoFibrob_A	Pericytes_A	Pericytes_B	T_cell_A	T_cell_B
Arteries_C	1.00	0.22	0.22	0.30	0.19	0.27	0.27	0.27	0.19	0.17
B_cell_A	0.22	1.00	0.23	0.21	0.16	0.21	0.21	0.20	0.19	0.17
Macrophages_A	0.22	0.23	1.00	0.21	0.18	0.21	0.21	0.21	0.20	0.18
Capillary_1_A	0.30	0.21	0.21	1.00	0.19	0.25	0.26	0.25	0.19	0.17
Ciliated_cells_A	0.19	0.16	0.18	0.19	1.00	0.20	0.20	0.19	0.22	0.18
MyoFibrob_A	0.27	0.21	0.21	0.25	0.20	1.00	0.30	0.28	0.19	0.16
Pericytes_A	0.27	0.21	0.21	0.26	0.20	0.30	1.00	0.32	0.19	0.17
Pericytes_B	0.27	0.20	0.21	0.25	0.19	0.28	0.32	1.00	0.18	0.17
T_cell_A	0.19	0.19	0.20	0.19	0.22	0.19	0.19	0.18	1.00	0.21
T_cell_B	0.17	0.17	0.18	0.17	0.18	0.16	0.17	0.17	0.21	1.00

Figure 2. Advanced search with scATAC-seq enhancer matrix. (**A**) Browsing all tissue/cell types by clicking on the name or image of that species. (**B**) Displaying the differences among selected tissue/cell types by scATAC-seq clustering analysis. (**C**) Showing the similarities among selected tissue/cell types by the Jaccard index.

RESULTS

Database statistics

We employed a modified unsupervised learning approach and the Cicero algorithm to build the single-cell enhancer resource scEnhancer (Supplementary Figure S3). To date, scEnhancer has catalogued 775 tissue/cell types, including 14 527 776 consensus single-cell enhancers and 63 658 600 enhancer-gene interactions from 1 196 906 single cells in three species (Table 1). We overlaid SNPs from GWAS (46) or TF binding motifs from JASPAR (47) with enhancer regions and found that the single-cell enhancers were much enriched for SNPs and TF binding sites. We also summarized the number of single cells, enhancers, and enhancer-gene connections in all tissue/cell types for each species (Supplementary Tables S2–S4). The integrated cell types covered many cancers and nearly all tissues in human and mouse (Supplementary Tables S2 and S3). Most cell types display high quality with an average of >3000 peaks per cell (Supplementary Tables S2–S4). The final consensus single-cell enhancer was determined by the possible functional evidence from enhancer–promoter interactions as defined by Cicero (33). As more and more cell type-specific marker genes were identified (48), we will use these marker genes to confirm the cluster's cell type in the scATAC-seq clustering analysis to confirm the cell types of clusters and then predict single-cell enhancers even when single-cell datasets are not labeled with cell type information.

Simple search

We designed five user-friendly analytical modules in scEnhancer for a simple search of single-cell enhancers (Figure 1): (i) Search for single-cell enhancers by region (Figure 1A). (ii) Search for single-cell enhancers by a gene (Figure 1B). (iii) Compare single-cell enhancers from different tissue/cell types (Figure 1C). (iv) Compare enhancers of a gene in different tissue/cell types (Figure 1D). (v) Predict target genes in genomic regions at the single-cell level (e.g. peaks from ChIP-Seq) (Figure 1E). Users can search for the enhancers by region in any tissue/cell of any species. In each module, an 'Example' button can facilitate users to give input in one-step for a simple search. We allowed the gene name or ID input from several common gene/protein resources, including Ensembl, EMBL,UCSC, PDB, FlyBase, RefSeq and UniProt (49–55). These modules serve as easy-to-use web interfaces for users to search, visualize and download single-cell enhancers and enhancer–promoter connections in any genomic region or any tissue/cell type(s).

Advanced search

We developed two powerful modules with several R packages as 'advanced search' to graphically display the differences/similarities among cell types or reveal the cell type specificity of enhancers at the single-cell level: (i) Display the differences among tissue/cell types by single-cell enhancer clustering (Figure 2). (ii) Identify tissue/cell

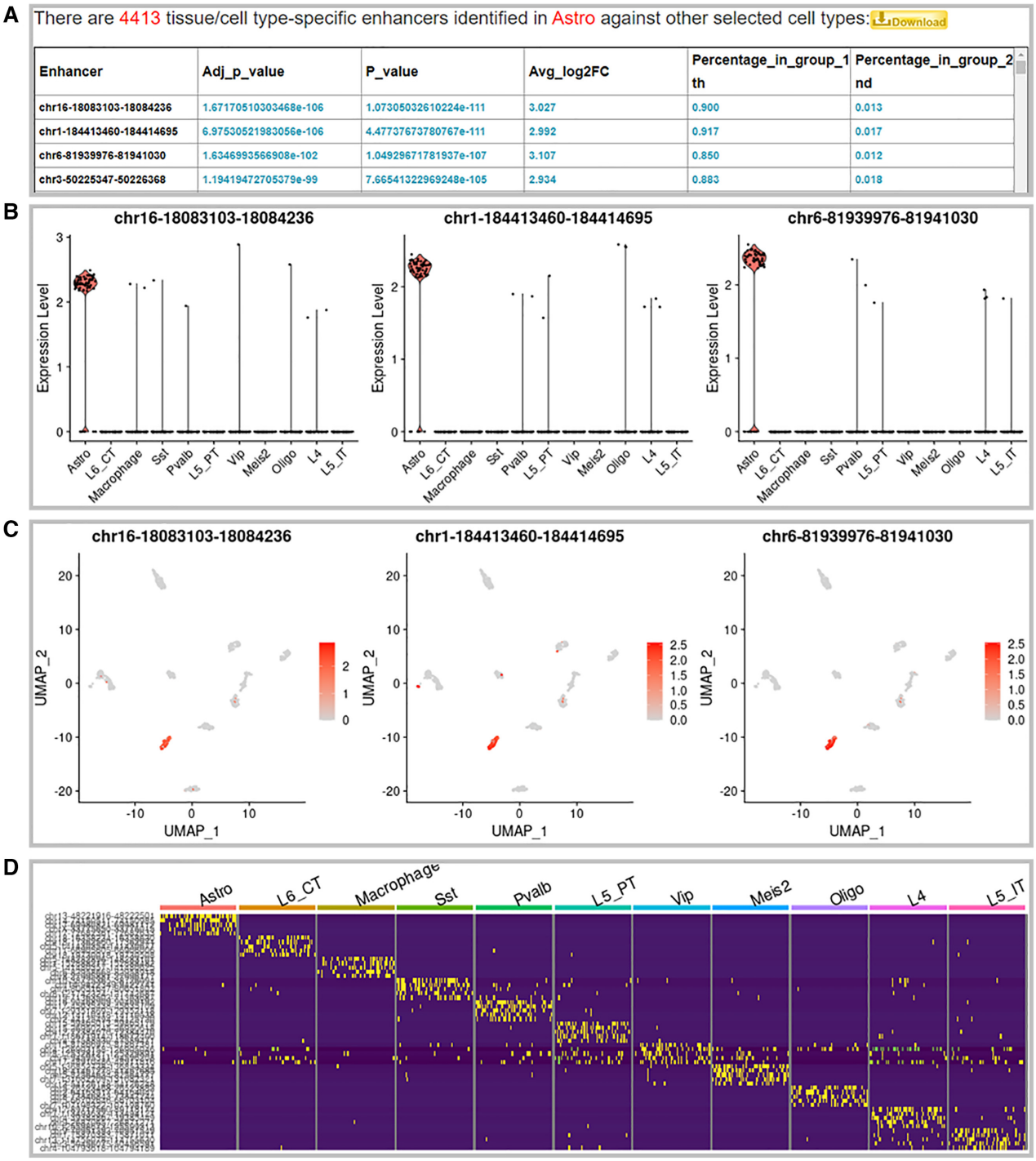


Figure 3. Identification of cell type-specific single-cell enhancers. (**A**) A list of cell type-specific enhancers for one cell type against the reference cell types. (**B**) Cell type specificities of the top three enhancers using VlnPlot. (**C**) The feature enrichments of the top three enhancers across all cell types using FeaturePlot. (**D**) Heatmap based on the top five single-cell enhancers that distinguish each cell type from the others.

type-specific enhancers at the single-cell level (Figure 3). To avoid the batch effects to the greatest extent, we assigned each tissue/cell type to a batch and compared different tissue/cell types within the same batch. Based on the equipment or technology platforms, we classified all the cell types into 10, 8 and 1 batches in human, mouse and fly, respectively.

In the first module, the users can select a group of cell types of interest in the same batch to observe their differences or similarities among them (Figure 2A). Merging the scATAC-seq matrices of the selected cell types, the differences among selected tissue/cell types can be displayed by DimPlot of Signac (38) (Figure 2B). In addition, this module can also calculate and present the

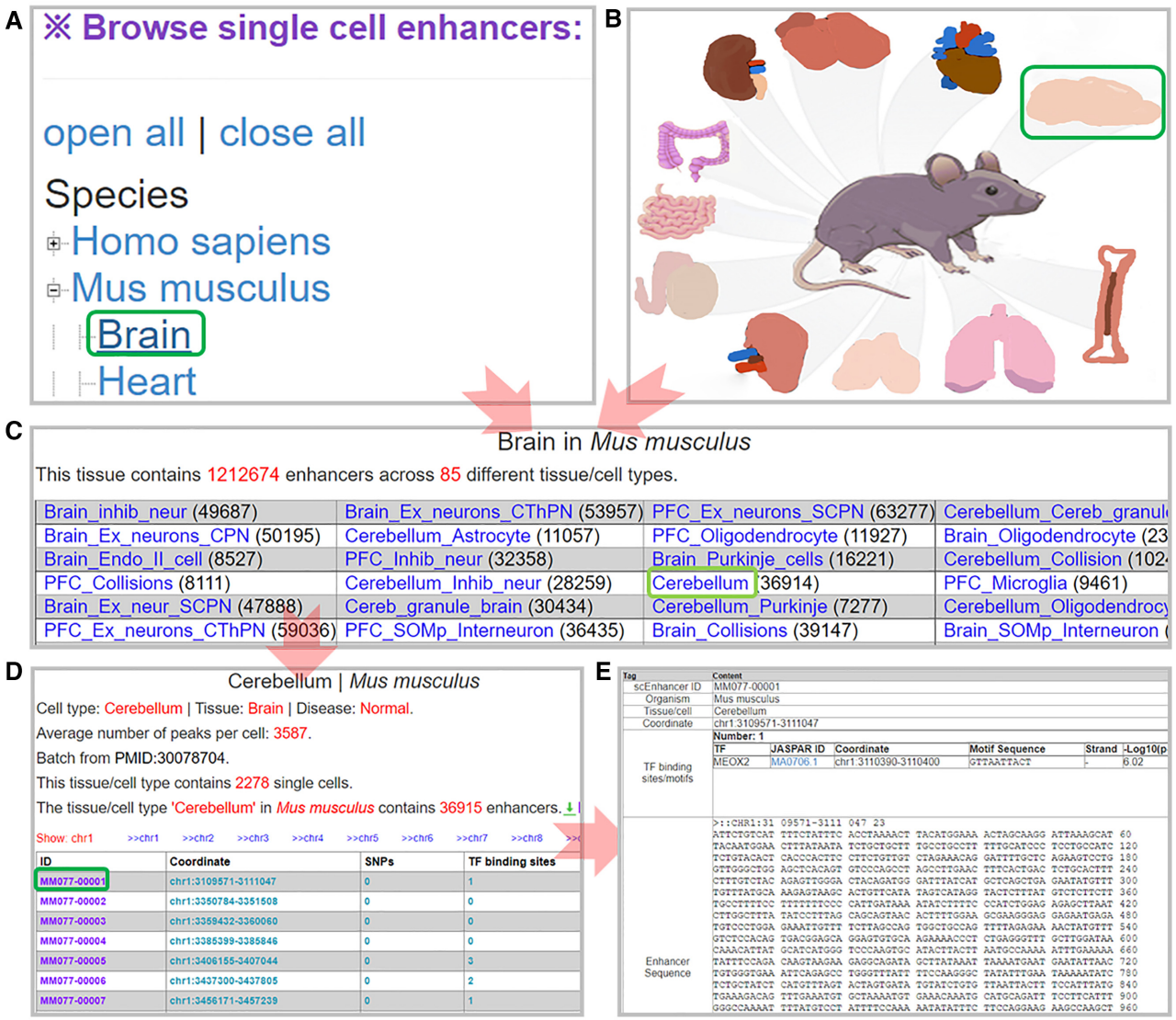


Figure 4. Single-cell enhancer browser. (**A**) List tissue/cell types by clicking on the name of the species or by clicking on the image of the species (**B**). (**C**) Selecting a tissue/cell type to browse all enhancers. The number in parentheses indicates the number of enhancers in that tissue/cell type. (**D**) A table of all the enhancers in the selected tissue/cell type. (**E**) A summary table describing the available features of the selected enhancer.

similarities among selected tissue/cell types by Jaccard index (Figure 2C).

To reveal the cell type specificity of single-cell enhancers, the users can select a batch of interest and a primary cell type in which specific enhancers were found and select the reference cell types to compare with (Figure 3). By clicking on 'Search', a list of cell type-specific enhancers will be obtained (Figure 3A). Moreover, the cell-type specificity of the identified single-cell enhancers can be displayed by particular analyses, such as VlnPlot, FeaturePlot, and Heatmap (Figure 3B–D).

Browser of single-cell enhancers

A browser page was provided in scEnhancer for accessing the single-cell enhancers. By clicking on the species name or image, the users can browse any tissue/cell type, any chromosome, and any single-cell enhancer, generating a summary table including the genomic coordinate of the enhancer, the contained GWAS SNPs (46), TF binding sites (47), relative super-enhancers (8), diseases (17) and DNA sequences (Figure 4).

CONCLUSIONS

scEnhancer is the first database to annotate enhancers or enhancer–promoter interactions at the single-cell level. It contains 50 861 554, 11 458 766 and 1 338 280 enhancer–promoter connections involving 3 905 212, 942 549 and 94 542 promoters, 11 373 862, 2 694 616 and 459 298 enhancers across 543, 185 and 47 tissue/cell types in human, mouse and fly, respectively. We believe this is the most comprehensive enhancer database that includes the largest number of enhancer-related datasets at the single-cell level.

DATA AVAILABILITY

A webserver with multiple analytic tools and deep browser capabilities is available at http://www.enhanceratlas.net/scenhancer and no login is required for all users to access the website. Tutorials for performing the scEnhancer analytic tools are freely provided at http://www.enhanceratlas.net/scenhancer/help.php. All the data including single-cell enhancers, promoters, enhancer–promoter interactions, SNPs/Motifs in enhancers, and scATAC matrix in tissue/cell types could be downloaded in http://www.enhanceratlas.net/scenhancer/download.php

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Bin Xia and Qiangsheng He for useful suggestions.

FUNDING

National Natural Science Foundation of China [32100434, 81972135]; Shenzhen's introduction of talents and research start-up [392020]. Funding for open access charge: National Natural Science Foundation of China [32100434, 81972135]; Shenzhen's introduction of talents and research start-up [392020].
Conflict of Interest Statement. None declared.

REFERENCES

1. Cusanovich,D.A., Reddington,J.P., Garfield,D.A., Daza,R.M., Aghamirzaie,D., Marco-Ferreres,R., Pliner,H.A., Christiansen,L., Qiu,X., Steemers,F.J. *et al.* (2018) The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*, **555**, 538–542.
2. Rickels,R. and Shilatifard,A. (2018) Enhancer logic and mechanics in development and disease. *Trends Cell Biol.*, **28**, 608–630.
3. Domcke,S., Hill,A.J., Daza,R.M., Cao,J., O'Day,D.R., Pliner,H.A., Aldinger,K.A., Pokholok,D., Zhang,F., Milbank,J.H. *et al.* (2020) A human cell atlas of fetal chromatin accessibility. *Science*, **370**, eaba7612.
4. Cusanovich,D.A., Hill,A.J., Aghamirzaie,D., Daza,R.M., Pliner,H.A., Berletch,J.B., Filippova,G.N., Huang,X., Christiansen,L., DeWitt,W.S. *et al.* (2018) A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell*, **174**, 1309–1324.
5. Wimmers,F., Donato,M., Kuo,A., Ashuach,T., Gupta,S., Li,C., Dvorak,M., Foecke,M.H., Chang,S.E., Hagan,T. *et al.* (2021) The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. *Cell*, **184**, 3915–3935.
6. Gao,T. and Qian,J. (2020) EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.*, **48**, D58–D64.
7. 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.
8. Chen,C., Zhou,D., Gu,Y., Wang,C., Zhang,M., Lin,X., Xing,J., Wang,H. and Zhang,Y. (2020) SEA version 3.0: a comprehensive extension and update of the super-enhancer archive. *Nucleic. Acids. Res.*, **48**, D198–D203.
9. Visel,A., Minovitsky,S., Dubchak,I. and Pennacchio,L.A. (2007) VISTA Enhancer Browser–a database of tissue-specific human enhancers. *Nucleic Acids Res.*, **35**, D88–D92.
10. Ashoor,H., Kleftogiannis,D., Radovanovic,A. and Bajic,V.B. (2015) DENdb: database of integrated human enhancers. *Database (Oxford)*, **2015**, bav085.
11. Cai,Z., Cui,Y., Tan,Z., Zhang,G., Tan,Z., Zhang,X. and Peng,Y. (2019) RAEdb: a database of enhancers identified by high-throughput reporter assays. *Database (Oxford)*, **2019**, bay140.
12. Fishilevich,S., Nudel,R., Rappaport,N., Hadar,R., Plaschkes,I., Iny Stein,T., Rosen,N., Kohn,A., Twik,M., Safran,M. *et al.* (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*, **2017**, bax028.
13. 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.
14. Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res.*, **44**, D164–D171.
15. Kumar,R., Lathwal,A., Kumar,V., Patiyal,S., Raghav,P.K. and Raghava,G.P.S. (2020) CancerEnD: a database of cancer associated enhancers. *Genomics*, **112**, 3696–3702.
16. 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.
17. 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.
18. 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.
19. Bai,X., Shi,S., Ai,B., Jiang,Y., Liu,Y., Han,X., Xu,M., Pan,Q., Wang,F., Wang,Q. *et al.* (2020) ENdb: a manually curated database of experimentally supported enhancers for human and mouse. *Nucleic Acids Res.*, **48**, D51–D57.
20. Buenrostro,J.D., Wu,B., Litzenburger,U.M., Ruff,D., Gonzales,M.L., Snyder,M.P., Chang,H.Y. and Greenleaf,W.J. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, **523**, 486–490.
21. Cusanovich,D.A., Daza,R., Adey,A., Pliner,H.A., Christiansen,L., Gunderson,K.L., Steemers,F.J., Trapnell,C. and Shendure,J. (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science*, **348**, 910–914.
22. Corces,M.R., Granja,J.M., Shams,S., Louie,B.H., Seoane,J.A., Zhou,W., Silva,T.C., Groeneveld,C., Wong,C.K., Cho,S.W. *et al.* (2018) The chromatin accessibility landscape of primary human cancers. *Science*, **362**, eaav1898.
23. Przytycki,P.F. and Pollard,K.S. (2021) CellWalker integrates single-cell and bulk data to resolve regulatory elements across cell types in complex tissues. *Genome Biol.*, **22**, 61.
24. Heinz,S., Romanoski,C.E., Benner,C. and Glass,C.K. (2015) The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.*, **16**, 144–154.
25. Buenrostro,J.D., Giresi,P.G., Zaba,L.C., Chang,H.Y. and Greenleaf,W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
26. Yan,F., Powell,D.R., Curtis,D.J. and Wong,N.C. (2020) From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.*, **21**, 22.
27. Thurman,R.E., Rynes,E., Humbert,R., Vierstra,J., Maurano,M.T., Haugen,E., Sheffield,N.C., Stergachis,A.B., Wang,H., Vernot,B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.
28. Daugherty,A.C., Yeo,R.W., Buenrostro,J.D., Greenleaf,W.J., Kundaje,A. and Brunet,A. (2017) Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. *Genome Res.*, **27**, 2096–2107.
29. Granja,J.M., Klemm,S., McGinnis,L.M., Kathiria,A.S., Mezger,A., Corces,M.R., Parks,B., Gars,E., Liedtke,M., Zheng,G.X.Y. *et al.* (2019) Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.*, **37**, 1458–1465.
30. Satpathy,A.T., Granja,J.M., Yost,K.E., Qi,Y., Meschi,F., McDermott,G.P., Olsen,B.N., Mumbach,M.R., Pierce,S.E., Corces,M.R. *et al.* (2019) Massively parallel single-cell chromatin

landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.*, **37**, 925–936.
31. Cao,J., Cusanovich,D.A., Ramani,V., Aghamirzaie,D., Pliner,H.A., Hill,A.J., Daza,R.M., McFaline-Figueroa,J.L., Packer,J.S., Christiansen,L. *et al.* (2018) Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, **361**, 1380–1385.
32. Yoo,S., Kim,J., Lyu,P., Hoang,T.V., Ma,A., Trinh,V., Dai,W., Jiang,L., Leavey,P., Duncan,L. *et al.* (2021) Control of neurogenic competence in mammalian hypothalamic tanycytes. *Sci. Adv.*, **7**, eabg3777.
33. Pliner,H.A., Packer,J.S., McFaline-Figueroa,J.L., Cusanovich,D.A., Daza,R.M., Aghamirzaie,D., Srivatsan,S., Qiu,X., Jackson,D., Minkina,A. *et al.* (2018) Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell*, **71**, 858–8718.
34. Sinnamon,J.R., Torkenczy,K.A., Linhoff,M.W., Vitak,S.A., Mulqueen,R.M., Pliner,H.A., Trapnell,C., Steemers,F.J., Mandel,G. and Adey,A.C. (2019) The accessible chromatin landscape of the murine hippocampus at single-cell resolution. *Genome Res.*, **29**, 857–869.
35. Parisi,F., Strino,F., Nadler,B. and Kluger,Y. (2014) Ranking and combining multiple predictors without labeled data. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1253–1258.
36. Fang,R., Preissl,S., Li,Y., Hou,X., Lucero,J., Wang,X., Motamedi,A., Shiau,A.K., Zhou,X., Xie,F. *et al.* (2021) Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.*, **12**, 1337.
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. Stuart,T., Srivastava,A., Lareau,C. and Satija,R. (2020) Multimodal single-cell chromatin analysis with Signac. bioRxiv doi: https://doi.org/10.1101/2020.11.09.373613, 10 November 2020, preprint: not peer reviewed.
39. Wang,A., Chiou,J., Poirion,O.B., Buchanan,J., Valdez,M.J., Verheyden,J.M., Hou,X., Kudtarkar,P., Narendra,S., Newsome,J.M. *et al.* (2020) Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of SARS-CoV2 host genes. *Elife*, **9**, e62522.
40. Hinrichs,A.S., Karolchik,D., Baertsch,R., Barber,G.P., Bejerano,G., Clawson,H., Diekhans,M., Furey,T.S., Harte,R.A., Hsu,F. *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–D598.
41. Xiong,L., Xu,K., Tian,K., Shao,Y., Tang,L., Gao,G., Zhang,M., Jiang,T. and Zhang,Q.C. (2019) SCALE method for single-cell ATAC-seq analysis via latent feature extraction. *Nat. Commun.*, **10**, 4576.
42. Granja,J.M., Corces,M.R., Pierce,S.E., Bagdatli,S.T., Choudhry,H., Chang,H.Y. and Greenleaf,W.J. (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.*, **53**, 403–411.
43. Ranzoni,A.M., Tangherloni,A., Berest,I., Riva,S.G., Myers,B., Strzelecka,P.M., Xu,J., Panada,E., Mohorianu,I., Zaugg,J.B. *et al.* (2021) Integrative single-cell RNA-Seq and ATAC-Seq analysis of human developmental hematopoiesis. *Cell Stem Cell*, **28**, 472–487.
44. Baker,S.M., Rogerson,C., Hayes,A., Sharrocks,A.D. and Rattray,M. (2019) Classifying cells with Scasat, a single-cell ATAC-seq analysis tool. *Nucleic Acids Res.*, **47**, e10.
45. Zeng,W., Chen,S., Cui,X., Chen,X., Gao,Z. and Jiang,R. (2021) SilencerDB: a comprehensive database of silencers. *Nucleic Acids Res.*, **49**, D221–D228.
46. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
47. Fornes,O., Castro-Mondragon,J.A., Khan,A., van der Lee,R., Zhang,X., Richmond,P.A., Modi,B.P., Correard,S., Gheorghe,M., Baranasic,D. *et al.* (2020) JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.*, **48**, D87–D92.
48. Zhang,X., Lan,Y., Xu,J., Quan,F., Zhao,E., Deng,C., Luo,T., Xu,L., Liao,G., Yan,M. *et al.* (2019) CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.*, **47**, D721–D728.
49. The UniProt, C. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.
50. Haft,D.H., DiCuccio,M., Badretdin,A., Brover,V., Chetvernin,V., O'Neill,K., Li,W., Chitsaz,F., Derbyshire,M.K., Gonzales,N.R. *et al.* (2018) RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.*, **46**, D851–D860.
51. Marygold,S.J., Crosby,M.A., Goodman,J.L. and FlyBase,C. (2016) Using FlyBase, a Database of Drosophila Genes and Genomes. *Methods Mol. Biol.*, **1478**, 1–31.
52. Velankar,S., Alhroub,Y., Best,C., Caboche,S., Conroy,M.J., Dana,J.M., Fernandez Montecelo,M.A., van Ginkel,G., Golovin,A., Gore,S.P. *et al.* (2012) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, **40**, D445–D452.
53. Casper,J., Zweig,A.S., Villarreal,C., Tyner,C., Speir,M.L., Rosenbloom,K.R., Raney,B.J., Lee,C.M., Lee,B.T., Karolchik,D. *et al.* (2018) The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.*, **46**, D762–D769.
54. McWilliam,H., Li,W., Uludag,M., Squizzato,S., Park,Y.M., Buso,N., Cowley,A.P. and Lopez,R. (2013) Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.*, **41**, W597–W600.
55. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.