

# SingleQ: a comprehensive database of single-cell expression quantitative trait loci (sc-eQTLs) cross human tissues

Zhiwei Zhou<sup>1,2,†</sup>, Jingyi Du<sup>1,2,†</sup>, Jianhua Wang<sup>3,†</sup>, Liangyi Liu<sup>1,†</sup>, M. Gracie Gordon<sup>4,5,6,7</sup>, Chun Jimmie Ye<sup>5,6,8,9,10,11,12</sup>, Joseph E. Powell<sup>13,14</sup>, Mulin Jun Li<sup>①</sup><sup>3,†,\*</sup> and Shuquan Rao<sup>1,2,†,\*</sup>

<sup>1</sup>State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, 288 Nanjing Road, Tianjin 300020, China

<sup>2</sup>Tianjin Institutes of Health Science, 28 Tuanbo Avenue, Tianjin 301600, China

<sup>3</sup>Department of Pharmacology, Tianjin Key Laboratory of Inflammation Biology, School of Basic Medical Sciences, National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, 22 Qixiangtai Road, Tianjin 300070, China

<sup>4</sup>Biological and Medical Informatics Graduate Program, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>5</sup>Division of Rheumatology, Department of Medicine, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>6</sup>Institute for Human Genetics, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>7</sup>Department of Bioengineering and Therapeutic Sciences, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>8</sup>Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>9</sup>Department of Epidemiology and Biostatistics, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>10</sup>Parker Institute for Cancer Immunotherapy, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>11</sup>Chan Zuckerberg Biohub, 499 Illinois Street, San Francisco, CA 94158, USA

<sup>12</sup>Bakar Computational Health Sciences Institute, University of California, 500 Parnassus Avenue, San Francisco, CA 94143, USA

<sup>13</sup>Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, 384 Victoria Street, Sydney, NSW 2010, Australia

<sup>14</sup>UNSW Cellular Genomics Futures Institute, University of New South Wales, UNSW Sydney, Sydney, NSW 2052, Australia

\*Corresponding author: Tel/Fax: +86 13880239302; Email: [raoshuquan@ihcamsac.cn](mailto:raoshuquan@ihcamsac.cn)

Correspondence may also be addressed to Mulin Jun Li. Tel/Fax: +86 15889735158; E-mail: [mulin0424.li@gmail.com](mailto:mulin0424.li@gmail.com)

†Co-first authors.

\*Senior authors.

Citation details: Zhou, Z., Du, J., Wang, J. *et al.* SingleQ: a comprehensive database of single-cell expression quantitative trait loci (sc-eQTLs) cross human tissues. *Database* (2024) Vol. 2024: article ID baae010; DOI: <https://doi.org/10.1093/database/baae010>

## Abstract

Mapping of expression quantitative trait loci (eQTLs) and other molecular QTLs can help characterize the modes of action of disease-associated genetic variants. However, current eQTL databases present data from bulk RNA-seq approaches, which cannot shed light on the cell type- and environment-specific regulation of disease-associated genetic variants. Here, we introduce our Single-cell eQTL Interactive Database which collects single-cell eQTL (sc-eQTL) datasets and provides online visualization of sc-eQTLs across different cell types in a user-friendly manner. Although sc-eQTL mapping is still in its early stage, our database curates the most comprehensive summary statistics of sc-eQTLs published to date. sc-eQTL studies have revolutionized our understanding of gene regulation in specific cellular contexts, and we anticipate that our database will further accelerate the research of functional genomics.

Database URL: <http://www.sqraolab.com/scqtl>

## Introduction

Functional interpretation of disease-associated genetic variants remains a significant challenge in the post-genome-wide association studies (GWAS) era (1). Mapping of expression quantitative trait loci (eQTLs) and other molecular QTLs can help characterize the modes of action of disease-associated genetic variants and identify the putative target genes they regulate. Efforts, such as Genotype-Tissue Expression (GTEx)

(2) and eQTL-Gen (3), have identified eQTLs across a variety of tissues but have used bulk RNA-seq approaches, which cannot shed light on the cell type- and environment-specific regulation of disease-associated genetic variants.

Recent advancements in single-cell technologies have enabled eQTL analysis at single-cell resolution. Compared with bulk RNA sequencing which averages gene expression across cell types and cell states, single-cell assays capture the

Received 4 September 2023; Revised 9 January 2024; Accepted 11 February 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

transcriptional states of individual cells (4). Single-cell eQTL (sc-eQTL) mapping can identify context-dependent eQTLs that vary with cell states, including some that colocalize with disease variants identified in genome-wide association studies, thus holds great potential for prioritizing therapeutic targets and pathways driving disease pathogenesis (5–19). Although significant progress has been made in the field of sc-eQTL mapping, a comprehensive database summarizing sc-eQTLs across human tissues is still lacking.

In this context, we collected all sc-eQTL datasets published to date and built a Single-cell eQTL Interactive Database (SingleQ) which provides online visualization of sc-eQTLs across different cell types in a user-friendly manner. Briefly, our database offers the following key features.

(i) Our database curates the most comprehensive summary statistics of sc-eQTLs from 273 different cell types and annotates 77 467 cell type-specific eGenes.

(ii) Cell type-specific sc-eQTLs can be queried with four searching options by either genetic variant, gene symbol, genomic location or chromosome region, allowing it to be friendly for any user.

(iii) Summary statistics of sc-eQTLs can be browsed by both cell type and genes centered on genetic variant or genomic location. More importantly, our database used popular tools, such as LocusZoom.js and Tabix, to visualize sc-eQTLs and relevant information in a single page, allowing users to identify cell type-specific sc-eQTLs easily and to prioritize target genes.

(iv) All sc-eQTL summary statistics can be downloaded for further customized analysis.

## Materials and methods

### Data collection

We collected all sc-eQTL studies from PubMed and Google Scholar with the following searching strategy: (single-cell expression quantitative trait loci) OR (single-cell eQTL) OR (sc-eQTL). Additional relevant studies were collected by screening the reference lists of studies in hand. Each study was manually assessed for suitability of inclusion, and sc-eQTL summary statistics were downloaded, processed, harmonized and visualized in our SingleQ database (<http://www.sqraolab.com/scqtl>). Additionally, we manually curated cell type annotations to provide detailed information of each cell type.

### Genetic variant information unification

Since the description of genetic variants from different sc-eQTL datasets might be heterogeneous, we synchronized Single Nucleotide Polymorphism Database (dbSNP) IDs with the ones from the most recently released dbSNP build 156 (20). For genetic variants that provided chromosome positions only, we first used LiftOver (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) to convert them to GRCh37 (Genome Reference Consortium Human Build 37) (21) positions and then filled in the reference (or major) and alternative (or minor) alleles of genetic variants. For sc-eQTLs, the effective allele is the alternative allele (otherwise indicated elsewhere).

### Standardization of sc-eQTL summary statistics

Since diverse strategies were used for eQTL mapping in different studies, the format of eQTL summary statistics varied

across studies. We therefore manually harmonized the format of sc-eQTL summary statistics, and the following items were included in our online database, including chromosome number, base position, rsID, ENSEMBL gene ID, effective allele, non-effect allele, minimum allele frequency,  $\beta$  value, standard error and  $P$ -value. We used our custom scripts to fill out any information missing in certain studies.

### Database design

SingleQ was built on a Python-based web framework. The sc-eQTL summary statistics and relevant information are stored in PostgreSQL or retrieved using Tabix (22). Several dynamic web pages are implemented using HyperText Markup Language, Cascading Style Sheets, jQuery and related JavaScript modules. Graphical visualization and tabular presentation of retrieved data are accomplished using JavaScript modules like LocusZoom.js (23) and DataTable.js (<https://datatables.net/>).

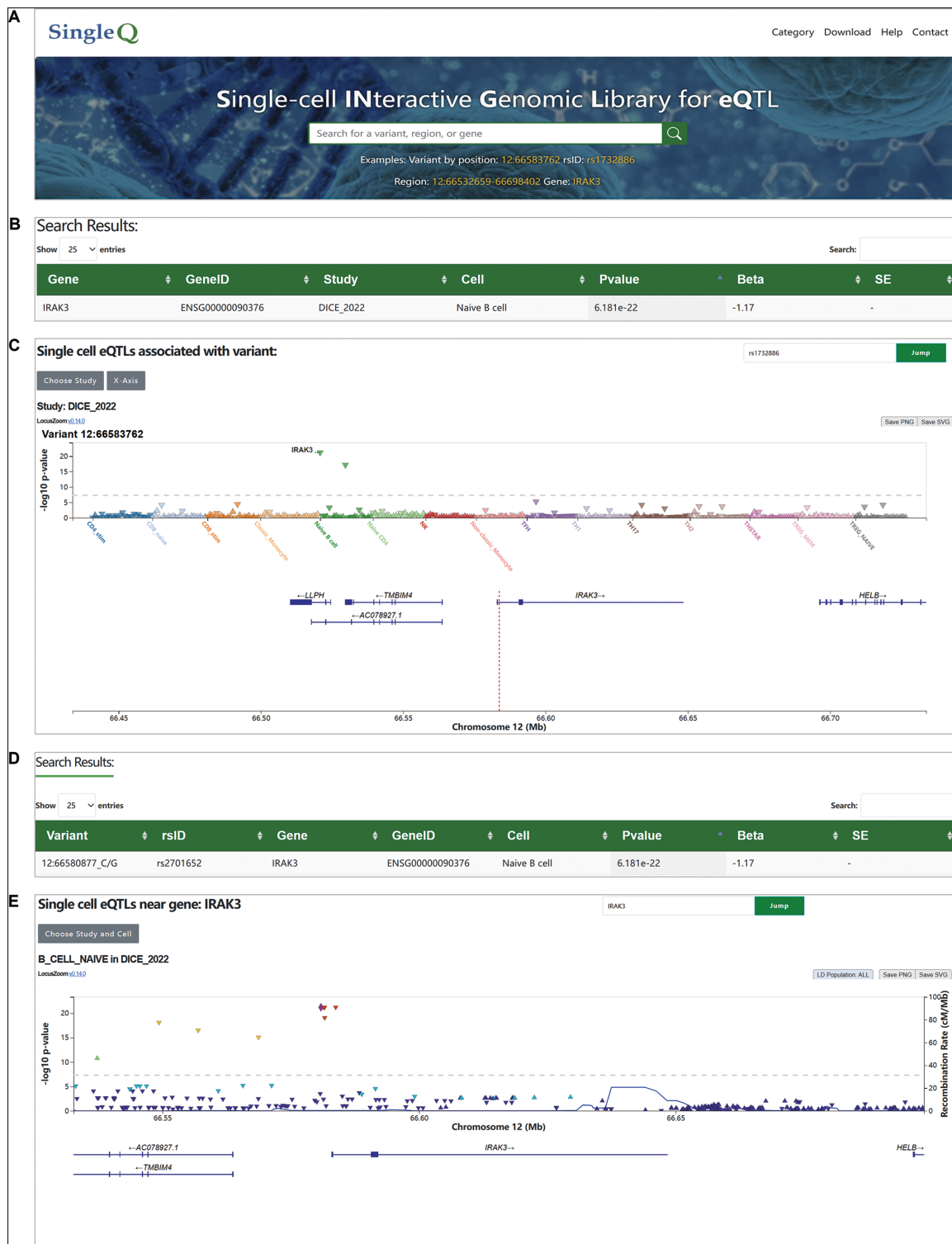
## Results

### Overview of SingleQ database

As of July 2023, we retrieved 15 independent sc-eQTL studies from which sc-eQTL summary statistics are available. For each study, sc-eQTL summary statistics were downloaded and harmonized based on the most recent dbSNP build 156. Briefly, SingleQ sc-eQTL database curated up to 77 467 eQTL summary statistics from 273 unique cell types covering different developmental stages of diverse tissues or cell states (Supplementary Table S1). To ensure uniform nomenclature, SingleQ mapped them to fine-grained terms (Supplementary Table S2).

We provide a user-friendly web interface for users to search, browse and download data. SingleQ allows users to retrieve sc-eQTL information from four perspectives: genetic variant by position, rsID, gene symbol and genomic region that spans no more than 200 kb (Figure 1A). When querying an individual variant, SingleQ displays all eQTLs between the genetic variant of interest and genes located within 2 Mb centered on the variant across all cell types and states (Figure 1B). In addition to summary statistics, SingleQ provides LocusZoom.js visualization of eQTLs across all available cell types and cell states from the chosen study (Figure 1C). Each triangle plot represents a unique eQTL with one specific gene nearby, where the Y-axis indicates the  $-\log_{10}(P\text{-val})$  of eQTLs and the X-axis shows cell types or cell states distinguished by different colors. Using the 'X-Axis' button on the top left, users can browse the eQTLs either by cell types/states or gene symbols. Detailed information, such as study ID, cell type or state, genetic variant, gene symbol or ID,  $P$ -val and beta, can be obtained by hovering the mouse over the triangle plot. Using the button 'Choose Study' on the top left, users can browse across different studies.

When querying a gene symbol or chromosome region, SingleQ returns all eQTLs between the gene of interest and genetic variants located within 2 Mb upstream and downstream across all cell types and states (Figure 1D). The eQTL plots are visualized by LocusZoom.js (Figure 1E), with each triangle plot representing a unique eQTL with the gene of interest, where the Y-axis and X-axis display the  $-\log_{10}(P\text{-val})$  of eQTLs and genomic region within 100 kb centered on the gene of interest, respectively.



**Figure 1. Web interface of SingleQ database.** (A) Browser navigation bar and search box of SingleQ with an example. (B) Example of results obtained through variant search. (C) Example of LocusZoom plot in the results page of variant search. (D) Example of results obtained through region search. (E) Example of LocusZoom plot in the results page of region search.

Collectively, through single-cell eQTL data filtering and visualization, SingleQ aids in uncovering potential cell type-specific regulatory effects.

### Example search

We used a previously reported case to illustrate how SingleQ helps users to interpret the cell type- or state-specific

regulatory effect of genetic variants. The example involves the genetic variant rs1732887 associated with acute lung injury. The region containing rs1732887 (−1464 A/G) is expected to be a highly conserved putative binding site of the *FOXP3* transcription factor, where the alternative allele G of rs1732887 might disrupt the binding site (24). Clinically, upregulation of the *IRAK3* gene nearby rs1732887 was observed in



website. This database provides researchers to explore sc-eQTLs through queries based on position, rsID, gene symbol and genomic region allowing for interactive visualization of cell type-specific eQTLs from diverse perspectives. Although the field of sc-eQTLs is still in its infancy, we anticipate that our sc-eQTL database will deliver on its promise to facilitate the elucidation of the molecular mechanisms underlying genetic associations with complex diseases. Since peripheral blood samples are more easily obtained than other tissue samples, more than half of the sc-eQTL annotations in the current version of SingleQ database are from peripheral blood mononuclear cells. As single-cell eQTL research continues to evolve rapidly, the SingleQ database will be continuously updated. Subsequent versions will further enhance database functionalities, aiming to provide more comprehensive and valuable information. In the future, we will continue to update SingleQ by adding more cell type- or state-eQTLs and enriching the functional modules to make SingleQ a powerful tool for investigating genetic regulation.

## Supplementary Material

Supplementary material is available at *Database* online. SingleQ is freely available online at <http://www.sqraolab.com/scqtl>.

## Author Contributions

Conceptualization, S.R. and M.J.L.; methodology, Z.Z., J.D. and J.W.; Dataset collection and website construction, Z.Z., J.D. and L.L.; writing—original draft, Z.Z. and J.D.; writing—review & editing, S.R., M.J.L.; supervision, S.R.

## Conflict of interest

None declared.

## Acknowledgements

We sincerely thank Prof. Gosia Trynka and Prof. Anna Lorenc (Wellcome Sanger Institute, Wellcome Genome Campus) for providing sc-eQTL datasets from T cells. This work was supported by the CAMS Innovation Fund for Medical Sciences (2021-I2M-1-041 to S.R.); the National Key R&D Program of China (2021YFA1102300 to S.R.); the Tianjin Municipal Science and Technology Commission Grant (21JCQNJC01220 to S.R.); the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2021-RC310-015 to S.R.); the Science, Technology & Innovation Project of Xiongan New Area (2022XAGG0142 to S.R.).

## References

- Broekema,R.V., Bakker,O.B. and Jonkers,I.H. (2020) A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol.*, **10**, 190221.
- Consortium,G. (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
- Vosa,U., Claringbould,A., Westra,H.J. *et al.* (2021) Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.*, **53**, 1300–1310.
- Kang,J.B., Raveane,A., Nathan,A. *et al.* (2023) Methods and insights from single-cell expression quantitative trait loci. *Annu. Rev. Genomics Hum. Genet.*, **24**, 277–303.
- Bryois,J., Calini,D., Macnair,W. *et al.* (2022) Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.*, **25**, 1104–1112.
- Cuomo,A.S.E., Seaton,D.D., McCarthy,D.J. *et al.* (2020) Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.*, **11**, 810.
- Elorbany,R., Popp,J.M., Rhodes,K. *et al.* (2022) Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLoS Genet.*, **18**, e1009666.
- Jerber,J., Seaton,D.D., Cuomo,A.S.E. *et al.* (2021) Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.*, **53**, 304–312.
- Nathan,A., Asgari,S., Ishigaki,K. *et al.* (2022) Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature*, **606**, 120–128.
- Oelen,R., de Vries,D.H., Brugge,H. *et al.* (2022) Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. *Nat. Commun.*, **13**, 3267.
- Ota,M., Nagafuchi,Y., Hatano,H. *et al.* (2021) Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*, **184**, 3006–3021.e3017.
- Perez,R.K., Gordon,M.G., Subramaniam,M. *et al.* (2022) Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science*, **376**, eabf1970.
- Schmiedel,B.J., Gonzalez-Colin,C., Fajardo,V. *et al.* (2022) Single-cell eQTL analysis of activated T cell subsets reveals activation and cell type-dependent effects of disease-risk variants. *Sci. Immunol.*, **7**, eabm2508.
- Schmiedel,B.J., Singh,D., Madrigal,A. *et al.* (2018) Impact of genetic polymorphisms on human immune cell gene expression. *Cell*, **175**, 1701–1715.e1716.
- Soskic,B., Cano-Gamez,E., Smyth,D.J. *et al.* (2022) Immune disease risk variants regulate gene expression dynamics during CD4(+) T cell activation. *Nat. Genet.*, **54**, 817–826.
- van der Wijst,M.G.P., Brugge,H., de Vries,D.H. *et al.* (2018) Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.*, **50**, 493–497.
- Yazar,S., Alquicira-Hernandez,J., Wing,K. *et al.* (2022) Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science*, **376**, eabf3041.
- Natri,H.M., Azodi,C.B.D., Peter,L. *et al.* (2023) Cell type-specific and disease-associated eQTL in the human lung. *bioRxiv*, [10.1101/2023.03.17.533161](https://doi.org/10.1101/2023.03.17.533161).
- Resztak,J.A., Wei,J., Zilioli,S. *et al.* (2023) Genetic control of the dynamic transcriptional response to immune stimuli and glucocorticoids at single-cell resolution. *Genome Res.*, **33**, 839–856.
- Sherry,S.T., Ward,M.H., Kholodov,M. *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
- Church,D.M., Schneider,V.A., Graves,T. *et al.* (2011) Modernizing reference genome assemblies. *PLoS Biol.*, **9**, e1001091.
- Li,H. (2011) Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics*, **27**, 718–719.
- Boughton,A.P., Welch,R.P., Flickinger,M. *et al.* (2021) LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics*, **37**, 3017–3018.
- Pino-Yanes,M., Ma,S.F., Sun,X. *et al.* (2011) Interleukin-1 receptor-associated kinase 3 gene associates with susceptibility to acute lung injury. *Am. J. Respir. Cell Mol. Biol.*, **45**, 740–745.

25. Karczewski,K.J., Francioli,L.C., Tiao,G. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
26. Sollis,E., Mosaku,A., Abid,A. *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.*, **51**, D977–D985.
27. Kang,R., Zhang,Y., Huang,Q. *et al.* (2019) EnhancerDB: a resource of transcriptional regulation in the context of enhancers. *Database*, **2019**, bay141.
28. Zhong,T., Wang,W., Liu,H. *et al.* (2023) eccDNA Atlas: a comprehensive resource of eccDNA catalog. *Briefings Bioinf.*, **24**, bbad037.

