

# xQTLatlas: a comprehensive resource for human cellular-resolution multi-omics genetic regulatory landscape

Yuran Jia<sup>1</sup>, Hongchao Dong<sup>1</sup>, Linhao Li<sup>2</sup>, Fang Wang<sup>1</sup>, Liran Juan<sup>3</sup>, Yadong Wang<sup>2,4,\*</sup>, Hongzhe Guo<sup>1,4,\*</sup> and Tianyi Zhao<sup>1,2,4,\*</sup>

<sup>1</sup>Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China

<sup>2</sup>School of Medicine and Health, Harbin Institute of Technology, Harbin 150001, China

<sup>3</sup>School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, China

<sup>4</sup>Zhengzhou Research Institute, Harbin Institute of Technology, Harbin 450000, China

\*To whom correspondence should be addressed. Tel: +86 13613619293; Email: zty2009@hit.edu.cn

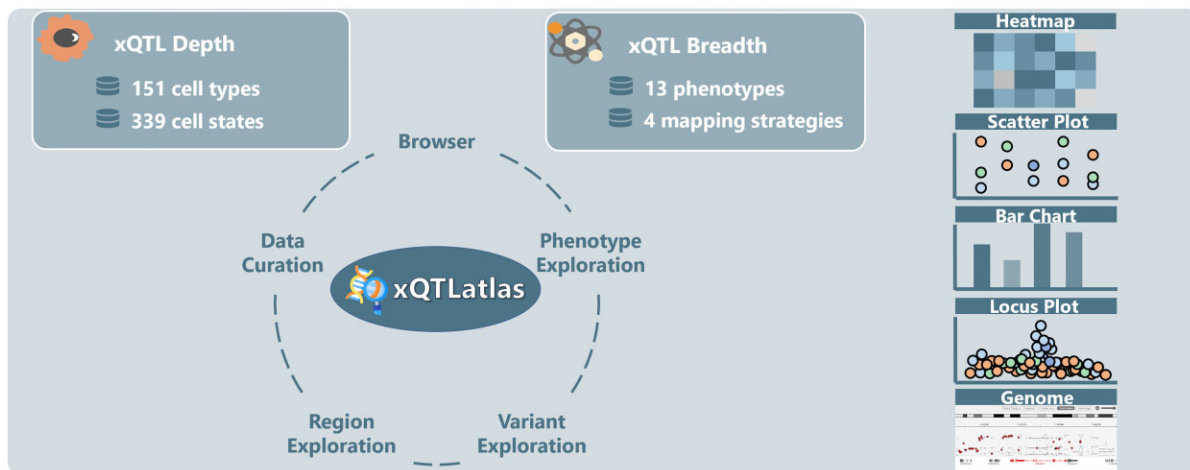
Correspondence may also be addressed to Yadong Wang. Email: ydwang@hit.edu.cn

Correspondence may also be addressed to Hongzhe Guo. Email: hzguo@hit.edu.cn

## Abstract

Understanding how genetic variants influence molecular phenotypes in different cellular contexts is crucial for elucidating the molecular and cellular mechanisms behind complex traits, which in turn has spurred significant advances in research into molecular quantitative trait locus (xQTL) at the cellular level. With the rapid proliferation of data, there is a critical need for a comprehensive and accessible platform to integrate this information. To meet this need, we developed xQTLatlas (<http://www.hitxqtl.org.cn/>), a database that provides a multi-omics genetic regulatory landscape at cellular resolution. xQTLatlas compiles xQTL summary statistics from 151 cell types and 339 cell states across 55 human tissues. It organizes these data into 20 xQTL types, based on four distinct discovery strategies, and spans 13 molecular phenotypes. Each entry in xQTLatlas is meticulously annotated with comprehensive metadata, including the origin of the tissue, cell type, cell state and the QTL discovery strategies utilized. Additionally, xQTLatlas features multiscale data exploration tools and a suite of interactive visualizations, facilitating in-depth analysis of cell-level xQTL. xQTLatlas provides a valuable resource for deepening our understanding of the impact of functional variants on molecular phenotypes in different cellular environments, thereby facilitating extensive research efforts.

## Graphical abstract



## Introduction

Genome-wide association studies (GWAS) have substantially enriched our understanding of genetic variants linked to human traits and diseases. Nevertheless, a predominant number of these variants are found in non-coding regions, complicating the elucidation of their roles and the molecular pathways

they influence (1,2). The advancement of high-throughput technologies has significantly facilitated quantitative trait locus (xQTL) analysis, establishing connections between functional variants and molecular phenotypes at diverse biological levels (3,4), including gene expression (eQTL) (5), chromatin accessibility (caQTL) (6), histone modification (hQTL)

Received: July 22, 2024. Revised: August 26, 2024. Editorial Decision: September 5, 2024. Accepted: September 13, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.

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.

(7) and DNA methylation (mQTL) (8). These resources enhance our comprehensive understanding of the genetic architecture, revealing causal relationships and functional impacts of genetic variants across multidimensional biological levels.

Large-scale xQTL studies such as GTEx (9), BLUEPRINT (10) and PsychENCODE (11) have advanced our understanding of the downstream causal effects of genetic variants. However, analyses based on averaged signals from tissue samples may obscure the specific cellular contexts in which functional variants influence their associated molecular pathways. Recent advancements in cell-resolution analysis have substantially enhanced our understanding of how genetic variations influence molecular phenotypes within specific cellular contexts (12–16). This progress reveals a complex interplay between genetic factors and cellular environments, crucial for unraveling the nuances of genetic regulation (17–21). For instance, Natri *et al.* utilized single-cell RNA sequencing (scRNA-seq) on lung tissues from pulmonary fibrosis patients and healthy controls, mapping eQTLs across 38 cell types, which revealed both shared and distinct cell-type-specific regulatory effects (22). Nathan *et al.* explored the impact of genetic variations in memory T cells, demonstrating how cellular dynamics, including cytotoxicity and regulatory capacity, significantly influence these variations (19). A pivotal direction for future genetic research is the precise characterization of the cellular contexts in which disease-associated variants modulate molecular phenotypes. This effort is essential for uncovering the molecular and cellular mechanisms contributing to disease susceptibility, deepening our understanding of the involved pathways, and facilitating the development of targeted therapeutic strategies (23–29). scQTLbase (30) and SingleQ (31) are pioneering databases specifically dedicated to housing eQTL research based on scRNA-seq. scQTLbase includes cellular-dependent eQTL (cd-eQTL) resources from 17 distinct studies, while SingleQ maintains a collection of 15 scRNA-seq resources. In addition to single-cell sequencing technology, there are other strategies for discovering cd-eQTL. These strategies include using purified cell type omics data to map cellular-dependent xQTL (cd-xQTL) for understanding genetic regulation in homogeneous cell populations (32,33), artificial intelligence models to detect cell type-specific regulatory effects (34), and cell deconvolution techniques to examine the interaction between genotype and cell abundance (35). Each of these strategies offers unique insights. However, current data resources do not comprehensively collect and organize cd-eQTL from these diverse approaches. Furthermore, the lack of cellular resolution genetic regulatory structure data at other molecular levels limits the depth and breadth of existing databases. These limitations highlight the urgent need for a comprehensive resource that not only systematically organizes cell resolution multi-omics genetic structures but also supports in-depth exploration and analysis.

To address these issues and meet the growing demand for a comprehensive and integrated cellular-level genetic regulatory landscape database, we propose xQTL Atlas (xQTLAtlas). xQTLAtlas integrates 796 datasets from 151 cell types and 339 cell states across 55 human tissues, covering 20 xQTL types. All cd-xQTL summary statistical datasets in xQTLAtlas have been processed through standardized pipelines. These data enable researchers to systematically study the regulatory mechanisms of risk variants across multidimensional biological levels in various cellular contexts. xQTLAtlas offers multi-level data exploration capabilities and unique interactive visu-

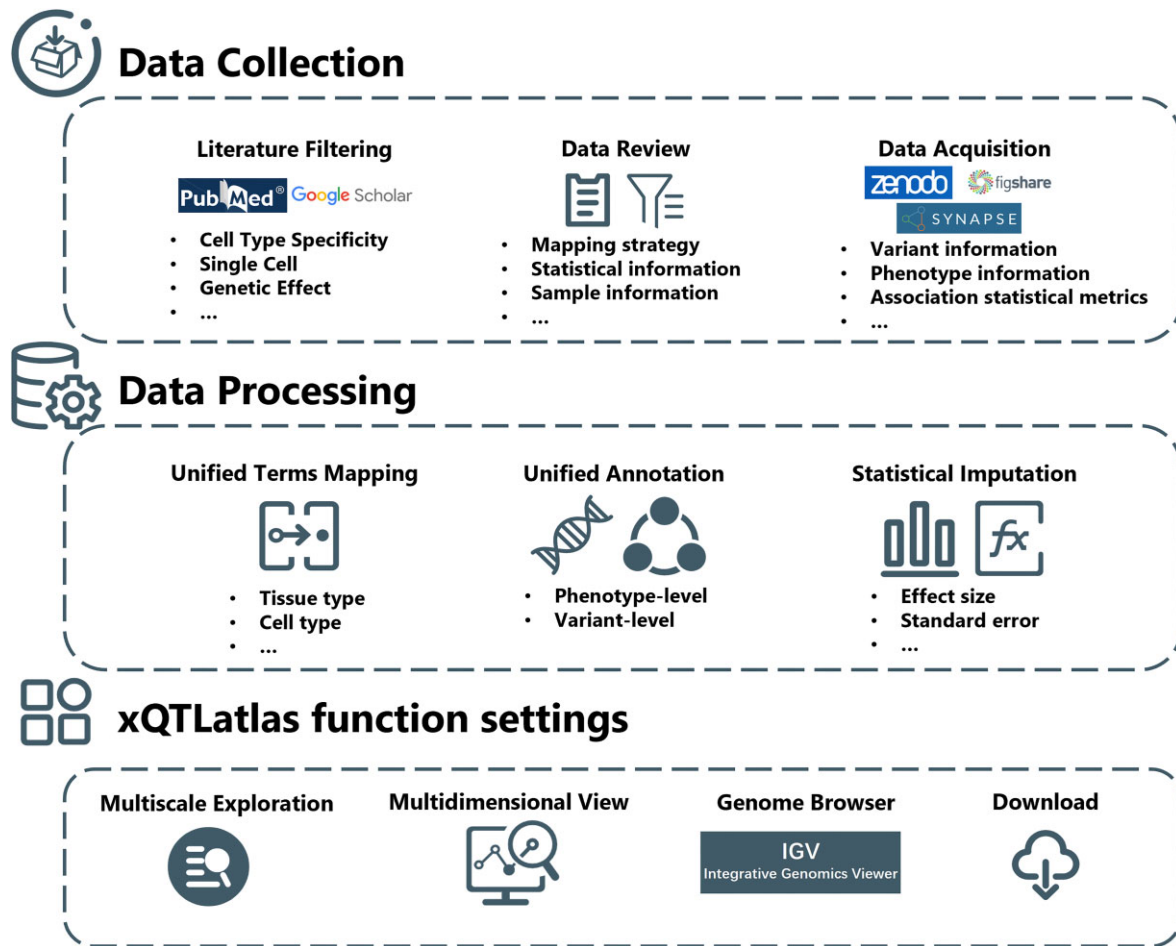
alization tools, enabling cross-cell type, phenotype and genetic variants association analysis views. Additionally, xQTLAtlas has significantly improved the user interface based on user-centered design principles, enhancing visibility and usability, and creating an environment conducive to intuitive data visualization and simplified queries. To our knowledge, xQTLAtlas is the first comprehensive database to systematically organize multi-omics genetic structures of functional variants at the cellular level. We anticipate that xQTLAtlas will become an invaluable resource for researchers, driving a deeper understanding of complex genetic regulatory mechanisms and promoting the development of precision medicine.

## Materials and methods

### Summary statistics collection and curation

In constructing our comprehensive atlas, xQTLAtlas, we meticulously implemented rigorous filtering criteria to ensure comprehensive and accurate content (Figure 1). We began by conducting searches on NCBI PubMed and Google Scholar using predefined keywords and terms related to cell type specificity and xQTL analysis to identify relevant studies. We integrated datasets from various cd-xQTL mapping strategies including cell sorting, single-cell sequencing, interaction analysis and computational discoveries to maintain data integrity. We included studies featuring cells from diverse biological backgrounds—normal, treated, diseased or other stimulated conditions—and rigorously selected data from published research articles, excluding those lacking essential details such as critical information on variants, traits or statistical measures. Due to the limited availability and incomplete statistics of cell-level trans-xQTL data, trans-xQTLs are not included in the current version of xQTLAtlas. All data are sourced from repositories such as Zenodo (<https://zenodo.org/>), FigShare (<https://figshare.com/>), GEO (36), Synapse (<https://www.synapse.org/>) and additional links provided in the original publications. xQTLAtlas includes research from both genome-wide and regional cd-xQTL mapping, providing comprehensive coverage and ensuring that data can be traced back to the demographic information of the populations studied. We map these populations to categories defined in the 1000 Genomes Project (37)—AFR (African), AMR (Admixed American), EAS (East Asian), EUR (European) and SAS (South Asian), annotating datasets with multiple populations as ‘MIX.’ Given the variability of tissues, cell types/states and phenotypes in xQTL studies, we methodically categorized and annotated datasets from multiple sources. This methodical approach ensures that each cd-xQTL dataset can be precisely traced back to its original source, such as specific tissues, cell types or mapping strategies.

For all genetic variants, we systematically remapped the chromosomal locations from the original publications to the Genome Reference Consortium Human Build 38 (GRCh38). When the coordinates were documented in GRCh37, we employed LiftOver (38) to transition these to GRCh38 coordinates. Simultaneously, we updated the original dbSNP ID to the dbSNP156 release (39). Alleles for each cd-xQTL were preserved as documented from their original publications. The variety of cd-xQTL mapping strategies used across different studies led to significant differences in the format of summary statistics. We extracted key statistical metrics from each qualifying publication, including affected molecular



**Figure 1.** Overview of xQTLatlas. xQTLatlas pipeline begins with quality control using predefined publication filters, followed by a detailed examination and extraction of original variant-phenotype association data. Following this, statistical data are standardized and imputed, culminating in the development of functional modules for xQTL analysis and visualization.

phenotypes and their annotations, genetic variants and associated statistical measures. Drawing on established practices from existing databases (4,30), we standardized the format of cd-xQTL summary statistics within xQTLatlas. This standardization ensures that each entry consistently includes detailed information about phenotypes and variant locations, alleles, associated *P*-values, standard errors, effect sizes and false discovery rates. Where key statistical measures were absent in the original datasets, we supplemented these by imputing values based on available additional information. For example, in cases where standard error were not provided, we calculated them using available *P*-values and effect size.

To augment xQTLatlas, we have integrated systematic variant annotations from VannoPortal (40), providing detailed information on allele frequency, linkage disequilibrium, evolutionary conservation, functional evidence and variant pathogenicity predictions across various tissues and cell types. Furthermore, insights into potential causal relationships between risk variants and complex traits from CAUSALdb (41) have been incorporated to enhance the database's comprehensiveness.

#### Ontology standardization and categorization

xQTLatlas encompasses a wide array of molecular phenotypes, necessitating the standardization of terminology for

both phenotypes and cell types to enhance the consistency and usability of the data. For each entry, we provide detailed annotations, including the phenotype type, genomic regions and specific names. Each phenotype is meticulously defined with specific description and annotation standards. For example, methylation phenotypes are annotated using probe IDs and genomic regions, whereas transcription and gene phenotypes are standardized based on genomic positions and names, employing the latest GENCODE (42) Release 46 annotations from the GRCh38 version. Phenotypes lacking standardized annotations are described by chromosomal positions, such as 'chr6:32472840–32473244' for caQTL, and specific histone modifications like 'H3K4me1(chr2:2131351–2132034)' and 'H3K27ac(chr6:32270482–32272947)'.

In terms of organizing cell type classification in xQTLatlas, we conducted a detailed manual review of each cell type as reported in the original studies. For each entry in xQTLatlas, we aligned and standardized annotations according to the tissue source, cell type, cell state and cd-xQTL mapping strategies provided in the original research. The annotations of cell states include specific features described in the publications, such as particular gene states or conditions like 'ATF1+' for active transcription factor presence or 'interferon stimulation 24 h' to indicate cellular response to treatment. This systematic review and annotation process ensures that each cell type in xQTLatlas is accurately represented and reflects the detailed

cellular context provided by the original studies. Cell types from studies that lack specific tissue source information and are derived from *in vitro* cultures are classified under ‘Other’. Moreover, we standardized terms from the original publications by mapping them to the naming conventions of the Human Cell Landscape (43) and GTEx (44), converting abbreviations to their full forms and aligning them with widely accepted reference names.

## Database design

xQTLatlas was constructed using a robust framework that incorporates MySQL (<https://www.mysql.com>) for database management and Java for backend development. Primary and composite indexing strategies were employed to optimize performance. A primary key index based on unique identifiers was used to expedite data retrieval and maintain data integrity. Additionally, a composite index, including chromosome and position columns, was implemented to accelerate specific queries. To further enhance performance, additional columns such as cell type, variant, phenotype and position were indexed, and a range-based partitioning strategy was adopted, segmenting the dataset into discrete, non-overlapping sections of 10 million bp each. The user interface of xQTLatlas was developed with HTML, Vue.js (<https://vuejs.org/>) and Plotly.js (<https://github.com/plotly/plotly.js/>), incorporating JavaScript libraries for enhanced interaction. Furthermore, IGV.js (45) was integrated for interactive genomic visualization. For optimal performance, we recommend accessing xQTLatlas using web browsers like Google Chrome or Microsoft Edge, ensuring efficient and effective use of its functionalities.

## Results

### Overview of xQTLatlas

The current version of xQTLatlas integrates summary statistics from 61 independent cd-xQTL studies, meticulously designed to ensure completeness and accuracy. Each dataset is manually annotated to include fundamental details such as tissue source, cell type and state. The entries in xQTLatlas are categorized into four main types based on the xQTL discovery methods detailed in the original publications: purified cell type xQTL, single cell xQTL, cell type/state interaction xQTL and *in silico* inferred cell type xQTL. These are further divided into 796 datasets affecting 13 molecular phenotypes (Supplementary Figure S1). Single-cell sequencing, which allow for a finer examination of cell heterogeneity and diversity within tissues, contribute the most extensive datasets, representing 46.8% of the total in xQTLatlas (Supplementary Figure S2). Additionally, studies employing purified cell type and single-cell sequencing for xQTL mapping are the most common sources of data, accounting for 56.5% and 25% of xQTLatlas, respectively (Supplementary Figure S3). xQTLatlas currently includes nearly 1.5 billion significant cd-xQTL entries ( $P$ -value  $< 0.05$ ), covering 151 cell types and 339 cell states from 55 human tissues, providing an integrated genetic regulatory landscape with unmatched depth and breadth. xQTLatlas ensures accessibility through a user-friendly portal, enabling researchers without extensive computational backgrounds to engage effortlessly with the data. It features interactive modules for visualizing cd-xQTL summary statistics, including heatmaps, locus plots

and scatter plots, along with a genome browser for visualizing genome-wide genetic associations. Moreover, all datasets and user-retrieved entries are downloadable in standardized formats for further analysis.

### Comprehensive data exploration module

xQTLatlas provides researchers with the capability to comprehensively explore cd-xQTL data from multiple dimensions. Using the data exploration interface (Figure 2A), users can efficiently delve into all functional variants and their associations with phenotypes, organized systematically by xQTL type (Figure 2B). xQTLatlas supports multi-level exploration across genetic variants, molecular phenotypes and regions of interest, enabling precise targeting.

**Exploring variant.** When searching for specific variants using their genomic position or dbSNP ID in variant search mode, users are initially presented with an overview organized by xQTL type and then by tissue type, as shown in Figure 2B. Users can then select a xQTL type of interest for further exploration. Upon choosing a specific subset of data, an interactive heatmap is displayed with the horizontal axis representing cell types and the vertical axis representing phenotypes. Each data cell in the heatmap represents a specific phenotype-variant association statistic, with the color intensity indicating the number of phenotype-variant associations (Figure 2C). Clicking on a data cell in the heatmap reveals the association strength and effect size for functional variants across different phenotypes within that cell type, where the height of the bars in the histogram is determined by the median  $-\log_{10}$   $P$ -value of the phenotype-variant associations, and the color represents the median effect size (Figure 2D). Additionally, the extent of specific phenotype-variant associations across cell types is also displayed, with data points colored according to the original publication (Figure 2E). Beneath the cross-cell type view at the bottom of the page, the summary statistics table is dynamically updated to reflect the user-selected data subset, simplifying the process of accessing and downloading relevant data (Figure 2F). Additionally, this dynamic table includes annotations of cellular context and phenotypes, as well as association  $P$ -values, effect size, standard error and false discovery rate for each xQTL. It also includes external links to the original publications and detailed information about the genetic variants. We have also integrated GTEx, VannoPortal and CAUSALDB, providing users with systematic and context-specific variant annotations and causal effects. This integrated approach not only enhances user interaction by providing tailored information but also facilitates the direct extraction of specific xQTL data from the interface, significantly boosting the utility of xQTLatlas for targeted genetic research.

**Exploring phenotype.** In the phenotype search mode, users start with an overview similar to the variant mode. However, in the interactive heatmap, the vertical axis represents different functional variants rather than phenotypes, with each cell in the heatmap denoting specific phenotype-variant association statistics. The color intensity of each cell is determined by the median  $-\log_{10}$   $P$ -value (Figure 3A). Clicking on a heatmap cell displays a locus plot for the selected cell type, illustrating the distribution of functional variants associated with the phenotype. The most significant cd-xQTL are specially marked (Figure 3B), and the color of the data points is determined by the publication source. Below this, a visualization shows the cross-cell type associations of variants





**Figure 2.** Demonstration of variant exploration in xQTLAtlas. **(A)** The data exploration page of xQTLAtlas supports three different modes. **(B)** The data overview interface after variant query is organized by xQTL type in the first level menu and by tissue in the second level menu. **(C)** Interactive heatmaps for data statistics. **(D)** Further statistical graphs based on xQTL statistics provide an overview of the association between different functional variants and phenotypes. **(E)** Comparison of the strength of the association between variant and phenotype in different cell types. **(F)** Summary statistics table.

(Figure 3C), and all modules enable comprehensive interaction with the data points. Like in the variant mode, the summary statistics table below provides cd-xQTL information.

**Exploring regions of interest.** Similar to the phenotype search mode, xQTLAtlas enables customized exploration of the regulatory landscape within specific genomic regions. To ensure optimal performance, we recommend limiting the search range to within 200 kb. The interface features a search box in the overview section that allows users to query genetic associations for specific phenotypes within defined intervals. The data are organized in interactive heatmaps, locus plots and scatter plots, following the same user-friendly format as in the phenotype search mode (Supplementary Figure S4). This provides a consistent and efficient user experience.

### Genome-wide association strength browser module

xQTLAtlas includes an interactive genomic browser that supports tailored exploration of diverse cd-xQTL datasets. Within the ‘Browser’ page, datasets are categorized by tissue, cell type, cell state, cd-xQTL mapping strategy and source of publication. Users can customize their analysis by adding multiple tracks to compare multi-omics cd-xQTL datasets. xQTL datasets of interest can be visualized by selecting them through the ‘Show Track List’ button (Figure 3D), and users have the capability to save the comparative genomic structure of a specific region as an SVG file. In addition, xQTLAtlas supports precise localization of genomic regions, phenotypes or variants in tracks, and highlights specific cd-xQTL data to better target user needs. Clicking on data points reveals detailed in-

formation about each cd-xQTL, including the dbSNP ID and the associated *P*-value. For instance, xQTLAtlas illustrates a visualization of the genetic variants impacts on gene expression, histone modification and methylation within the same genomic region in naive CD4 + T cells (Figure 3D). In this context, functional variants exert a more significant regulatory effect on molecular phenotypes at the levels of methylation and histone modifications than on gene expression. This indicates that epigenetic modifications play a predominant role in shaping the functional landscape of this region, potentially directing differential gene activity that could affect cellular behavior and responses. This underscores the complex interplay between different molecular mechanisms and emphasizes the utility of xQTLAtlas in dissecting the multilayered genetic influences on cellular functions.

### Discussion and future directions

We have developed xQTLAtlas, a comprehensive database designed to explore genetic regulatory landscapes at the cellular resolution. xQTLAtlas supports multiscale data exploration and offers multidimensional visualization and analysis capabilities. It incorporates data from an array of cell types, cell states and cd-xQTL mapping strategies, enabling in-depth analysis to uncover the intricate regulatory structures of functional variants. xQTLAtlas is distinguished by several innovative features: (i) The current version has meticulously curated hundreds of cd-xQTL datasets, organized in a hierarchical tree structure to facilitate efficient data access based on unique properties. (ii) Multi-level data exploration capabilities and distinctive interactive visualization modules



**Figure 3.** Demonstration of phenotype exploration and genome browser in xQTLatlas. **(A)** Interactive heatmaps for data statistics. **(B)** View of genetic regulatory structures for specific phenotypes. **(C)** View of phenotype-variant associations across cell types. **(D)** Genome browser of xQTLatlas visualizes the genetic architecture of different molecular phenotypes within the same genomic region in naive CD4 + T cell.

allow users to explore genetic regulatory landscapes across various scales. (iii) A genetic structural landscape browser for visualizes the association structures of multi-omics quantitative trait locus across the genome-wide. For further guidance, refer to the Tutorial page on xQTLatlas.

Compared to other resources that encompass partial cellular-resolution xQTL data, xQTLatlas exhibits significant advantages. Specifically, QTLbase (4) focuses primar-

ily on tissue-level data, consequently providing limited cellular resolution xQTL. Furthermore, databases like scQTLbase, ImmuneRegulation (46) and SingleQ lack diversity in their xQTL mapping strategies, cell types, molecular phenotypes and visualization techniques. xQTLatlas integrates an extensive range of genetic regulatory resources based on various mapping strategies and molecular phenotypes, enhancing both the depth and breadth of the data. Additionally, by stan-

standardizing cell type terminology and systematically annotating functional variants along with their potential causal relationships, xQTLatlas significantly advances its practical utility. xQTLatlas features advanced visualization and exploration tools that enable interactive, multi-scale exploration, comparison and analysis of xQTL data. Its user-friendly interface significantly enhances researchers' capacity to perform comprehensive studies of the genetic regulatory landscape, surpassing the functionalities of existing databases. As our understanding of the complexity and heterogeneity of biological systems deepens, tools like xQTLatlas become increasingly essential. It significantly advances the exploration of genetic regulatory mechanisms at the cellular level, which is crucial for elucidating how genetic variants operate across various cellular contexts.

We plan to regularly update xQTLatlas, adding new data and features every 6 months to ensure our users have access to the latest and most relevant data in cd-xQTL research. With the evolution of new paradigms in population genetics, such as *in vitro* quantitative trait locus studies (47,48), we are positioned to explore the effects of genetic variants in rare cell types and unravel the molecular basis of common genetic diseases. Additionally, we aim to incorporate a broader range of cell phenotypes, including response, migration and stress states, to serve a wider research community. Future plans also include expanding our analysis pipelines and visualization methods to enhance the functionality of xQTLatlas.

## Data availability

xQTLatlas is freely accessible to users with no registration needed at <http://www.hitxqtl.org.cn/>. All the datasets curated can be downloaded from xQTLatlas website for the purpose of research.

## Supplementary data

Supplementary Data are available at NAR Online.

## Acknowledgements

We would like to thank a number of users for reporting bugs and providing constructive feedback. We would also like to thank the researchers for publishing and sharing their data. We thank the developers of IGV for their efforts in implementing xQTL track.

Y.J.: Data Curation, Software Development, Original Draft Writing, Review and Editing, Visualization. H.D.: Software Development. L.L.: Data Curation. F.W.: Original Draft Writing, Review and Editing. L.J.: Software Development. Yadong Wang: Conceptualization, Project Supervision, Review and Editing. H.G.: Conceptualization, Project Supervision, Review and Editing. T.Z.: Conceptualization, Project Supervision, Review and Editing.

## Funding

National Natural Science Foundation of China [62202125].

## Conflict of interest statement

The authors declare no competing interests.

## References

1. Qi,T, Wu,Y, Fang,H, Zhang,F, Liu,S, Zeng,J and Yang,J. (2022) Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.*, **54**, 1355–1363.
2. Visscher,P.M., Wray,N.R., Zhang,Q, Sklar,P, McCarthy,M.I., Brown,M.A. and Yang,J. (2017) 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.
3. Ng,B, White,C.C., Klein,H.-U., Sieberts,S.K., McCabe,C., Patrick,E., Xu,J., Yu,L., Gaiteri,C., Bennett,D.A., *et al.* (2017) An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.*, **20**, 1418–1426.
4. Zheng,Z., Huang,D., Wang,J., Zhao,K., Zhou,Y., Guo,Z., Zhai,S., Xu,H., Cui,H., Yao,H., *et al.* (2020) QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res.*, **48**, D983–D991.
5. Vösa,U., Claringbould,A., Westra,H.-J., Bonder,M.J., Deelen,P., Zeng,B., Kirsten,H., Saha,A., Kreuzhuber,R., Yazar,S., *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.
6. Bryois,J., Garrett,M.E., Song,L., Safi,A., Giusti-Rodriguez,P., Johnson,G.D., Shieh,A.W., Buil,A., Fullard,J.F., Roussos,P., *et al.* (2018) Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.*, **9**, 3121.
7. Kundu,K., Tardaguila,M., Mann,A.L., Watt,S., Pongstingl,H., Vasquez,L., Von Schiller,D., Morrell,N.W., Stegle,O., Pastinen,T., *et al.* (2022) Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for 12 immune-mediated diseases. *Nat. Genet.*, **54**, 251–262.
8. Oliva,M., Demanelis,K., Lu,Y., Chernoff,M., Jasmine,F., Ahsan,H., Kibriya,M.G., Chen,L.S. and Pierce,B.L. (2023) DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. *Nat. Genet.*, **55**, 112–122.
9. The,G.C., Aguet,F., Anand,S., Ardlie,K.G., Gabriel,S., Getz,G.A., Graubert,A., Hadley,K., Handsaker,R.E., Huang,K.H., *et al.* (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
10. Chen,L., Ge,B., Casale,F.P., Vasquez,L., Kwan,T., Garrido-Martin,D., Watt,S., Yan,Y., Kundu,K., Ecker,S., *et al.* (2016) Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell*, **167**, 1398–1414.
11. Wang,D., Liu,S., Warrell,J., Won,H., Shi,X., Navarro,F.C.P., Clarke,D., Gu,M., Emani,P., Yang,Y.T., *et al.* (2018) Comprehensive functional genomic resource and integrative model for the human brain. *Science*, **362**, eaat8464.
12. Cuomo,A.S.E., Seaton,D.D., McCarthy,D.J., Martinez,I., Bonder,M.J., Garcia-Bernardo,J., Amatya,S., Madrigal,P., Isaacson,A., Buettner,F., *et al.* (2020) Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.*, **11**, 810.
13. van der Wijst,M.G.P., de Vries,D.H., Groot,H.E., Trynka,G., Hon,C.C., Bonder,M.J., Stegle,O., Nawijn,M.C., Idaghdour,Y., van der Harst,P., *et al.* (2020) The single-cell eQTLGen consortium. *eLife*, **9**, e52155.
14. Yazar,S., Alquicira-Hernandez,J., Wing,K., Senabouth,A., Gordon,M.G., Andersen,S., Lu,Q., Rowson,A., Taylor,T.R.P., Clarke,L., *et al.* (2022) Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science*, **376**, eabf3041.
15. Cuomo,A.S.E., Nathan,A., Raychaudhuri,S., MacArthur,D.G. and Powell,J.E. (2023) Single-cell genomics meets human genetics. *Nat. Rev. Genet.*, **24**, 535–549.



16. Yu, X., Xu, X., Zhang, J. and Li, X. (2023) Batch alignment of single-cell transcriptomics data using deep metric learning. *Nat. Commun.*, **14**, 960.
17. Teng, M., Du, D., Chen, D. and Irizarry, R.A. (2021) Characterizing batch effects and binding site-specific variability in ChIP-seq data. *NAR Genom Bioinform*, **3**, lqab098.
18. Liu, X., Zhao, B., Shaw, T.I., Fridley, B.L., Duckett, D.R., Tan, A.C. and Teng, M. (2022) Summarizing internal dynamics boosts differential analysis and functional interpretation of super enhancers. *Nucleic Acids Res.*, **50**, 3115–3127.
19. Nathan, A., Asgari, S., Ishigaki, K., Valencia, C., Amariuta, T., Luo, Y., Beynor, J.I., Baglaenko, Y., Suliman, S., Price, A.L., *et al.* (2022) Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature*, **606**, 120–128.
20. Aquino, Y., Bisiaux, A., Li, Z., O'Neill, M., Mendoza-Revilla, J., Merklings, S.H., Kerner, G., Hasan, M., Libri, V., Bondet, V., *et al.* (2023) Dissecting human population variation in single-cell responses to SARS-CoV-2. *Nature*, **621**, 120–128.
21. Song, L., Sun, X., Qi, T. and Yang, J. (2023) Mixed model-based deconvolution of cell-state abundances (MeDuSA) along a one-dimensional trajectory. *Nat. Comput. Sci.*, **3**, 630–643.
22. Natri, H.M., Del Azodi, C.B., Peter, L., Taylor, C.J., Chugh, S., Kendle, R., Chung, M.-i., Flaherty, D.K., Matlock, B.K., Calvi, C.L., *et al.* (2024) Cell-type-specific and disease-associated expression quantitative trait loci in the human lung. *Nat. Genet.*, **56**, 595–604.
23. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C. and Plagnol, V. (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.*, **10**, e1004383.
24. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, **48**, 481–487.
25. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. and Meyre, D. (2019) Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.*, **20**, 467–484.
26. Liu, H., Zhang, W., Zou, B., Wang, J., Deng, Y. and Deng, L. (2020) DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res.*, **48**, D871–D881.
27. Li, Y., Qiao, G., Gao, X. and Wang, G. (2022) Supervised graph co-contrastive learning for drug–target interaction prediction. *Bioinformatics*, **38**, 2847–2854.
28. Li, Y., Qiao, G., Wang, K. and Wang, G. (2022) Drug–target interaction prediction via multi-channel graph neural networks. *Brief Bioinform*, **23**, bbab346.
29. Cao, C., Shao, M., Zuo, C., Kwok, D., Liu, L., Ge, Y., Zhang, Z., Cui, F., Chen, M., Fan, R., *et al.* (2024) RAVAR: a curated repository for rare variant–trait associations. *Nucleic Acids Res.*, **52**, D990–D997.
30. Ding, R., Wang, Q., Gong, L., Zhang, T., Zou, X., Xiong, K., Liao, Q., Plass, M. and Li, L. (2024) scQTLbase: an integrated human single-cell eQTL database. *Nucleic Acids Res.*, **52**, D1010–D1017.
31. Zhou, Z., Du, J., Wang, J., Liu, L., Gordon, M.G., Ye, C.J., Powell, J.E., Li, M.J. and Rao, S. (2024) SingleQ: a comprehensive database of single-cell expression quantitative trait loci (sc-eQTLs) across human tissues. *Database*, **2024**, baee010.
32. Momozawa, Y., Dmitrieva, J., Théâtre, E., Deffontaine, V., Rahmouni, S., Charlotiaux, B., Crins, F., Docampo, E., Elansary, M., Gori, A.-S., *et al.* (2018) IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.*, **9**, 2427.
33. Bossini-Castillo, L., Glinos, D.A., Kunowska, N., Golda, G., Lamikanra, A.A., Spitzer, M., Soskic, B., Cano-Gamez, E., Smyth, D.J., Cattermole, C., *et al.* (2022) Immune disease variants modulate gene expression in regulatory CD4+ T cells. *Cell Genom.*, **2**, 100117.
34. Xiao, Y., Wang, J., Li, J., Zhang, P., Li, J., Zhou, Y., Zhou, Q., Chen, M., Sheng, X., Liu, Z., *et al.* (2023) An analytical framework for decoding cell type-specific genetic variation of gene regulation. *Nat. Commun.*, **14**, 3884.
35. Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., *et al.* (2020) Cell type-specific genetic regulation of gene expression across human tissues. *Science*, **369**, eaaz8528.
36. Clough, E., Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, J.F., Thomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., *et al.* (2024) NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res.*, **52**, D138–D144.
37. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
38. Lee, C.M., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., Nassar, L.R., Powell, C.C., *et al.* (2020) UCSC Genome Browser enters 20th year. *Nucleic Acids Res.*, **48**, D756–D761.
39. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, D308–D311.
40. Huang, D., Zhou, Y., Yi, X., Fan, X., Wang, J., Yao, H., Sham, P.C., Hao, J., Chen, K. and Li, M.J. (2022) VannoPortal: multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. *Nucleic Acids Res.*, **50**, D1408–D1416.
41. Wang, J., Huang, D., Zhou, Y., Yao, H., Liu, H., Zhai, S., Wu, C., Zheng, Z., Zhao, K., Wang, Z., *et al.* (2020) CAUSALdb: a database for disease/trait causal variants identified using summary statistics of genome-wide association studies. *Nucleic Acids Res.*, **48**, D807–D816.
42. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, J., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
43. Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., *et al.* (2020) Construction of a human cell landscape at single-cell level. *Nature*, **581**, 303–309.
44. Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A.V., *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
45. Robinson, J.T., Thorvaldsdottir, H., Turner, D. and Mesirov, J.P. (2023) igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*, **39**, btac830.
46. Kalayci, S., Selvan, M.E., Ramos, J., Cotsapas, C., Harris, E., Kim, E.-Y., Montgomery, R.R., Poland, G., Pulendran, B., Tsang, J.S., *et al.* (2019) ImmuneRegulation: a web-based tool for identifying human immune regulatory elements. *Nucleic Acids Res.*, **47**, W142–W150.
47. Panopoulos, A.D., D'Antonio, M., Benaglio, P., Williams, R., Hashem, S.I., Schultdt, B.M., DeBoever, C., Arias, A.D., Garcia, M., Nelson, B.C., *et al.* (2017) iPSCORE: a resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Rep.*, **8**, 1086–1100.
48. Aygün, N., Liang, D., Crouse, W.L., Keele, G.R., Love, M.I. and Stein, J.L. (2023) Inferring cell-type-specific causal gene regulatory networks during human neurogenesis. *Genome Biol.*, **24**, 130.