SCVdb: a comprehensive platform for integrating single-cell chromatin accessibility regions with causal variants

Zheng-Min Yu[1,2,†], Feng-Cui Qian[1,2,†], Qiao-Li Fang[2,†], Xiang-Yang Meng[2,†], Yan-Yu Li[1,2], Chen-Chen Feng[2], Li-Dong Li[1], Bing-Long Li[3], Hui Jiang[2], Qiu-Yu Wang[1,2], Xuan Fan[4], Jin-Cheng Guo[4,*] and Chun-Quan Li[1,2,3,5,6,*]

[1]The First Affiliated Hospital & Hunan Provincial Key Laboratory of Multi-omics and Artificial Intelligence of Cardiovascular Diseases, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

[2]School of Computer, University of South China, Hengyang, Hunan, 421001, China

[3]Insititute of Biochemistry and Molecular Biology, Hengyang Medical College, University of South China, Hengyang, Hunan 421001, China

[4]School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing 100029, China

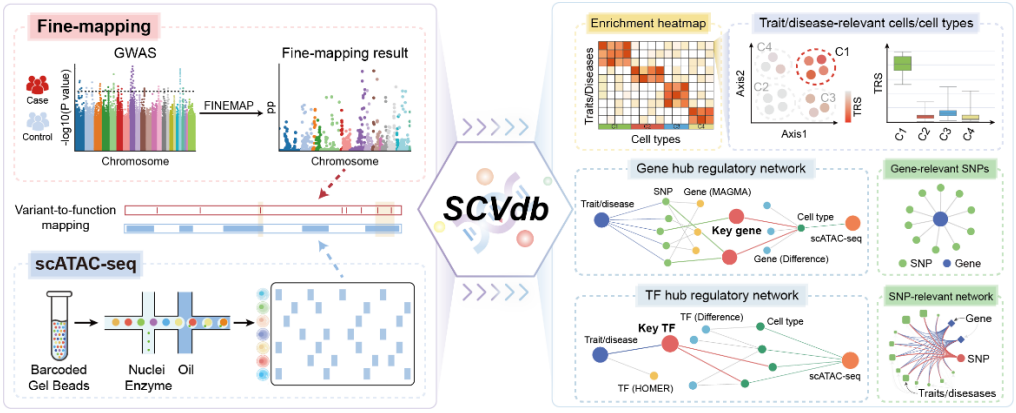
[5]Hunan Provincial Maternal and Child Health Care Hospital, National Health Commission Key Laboratory of Birth Defect Research and Prevention, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

[6]Key Laboratory of Rare Pediatric Diseases, Ministry of Education, University of South China, Hengyang, Hunan, 421001, China

Abstract

Integrating causal variant effects with single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq) enables a more effective elucidation of the roles and impacts of genetic variations at the single-cell level. With the accumulation of genome-wide association studies (GWAS) and single-cell genomic data, there is an urgent need for comprehensive analysis and efficient exploration of these data to uncover the underlying biological processes. To address this, we developed SCVdb (https://bio.liclab.net/scvdb/), a user-friendly database, which aims to provide trait-relevant cell populations at single-cell resolution. The current version of SCVdb integrated ~180 scATAC-seq datasets and ~15,800 fine-mapping results, generating 15.95 billion trait-cell pairs, offering valuable resources for exploring the functional localization of single-cell variations. To enhance the understanding of how phenotypic associations are mapped to single-cell data, SCVdb provides a wealth of detailed information, including trait-relevance scores (TRSs) for individual cells, cell-type-specific differential gene activities and transcription factors (TFs) motif enrichments, trait-associated gene interactions, trait-associated TF interactions, and regulatory networks linking cell types to traits. For these comprehensive analysis results, SCVdb offers users convenient interface to search, browse, analyze, and visualize relationships between traits and cell populations at single-cell resolution.

Graphical abstract



Introduction

Genome-wide association studies (GWAS) provide a broad foundation for the study of genetic variant mechanisms underlying human complex phenotypes or diseases, such as the manifestation of genetic factors in various traits or diseases (1-6), risk prediction for various complex diseases (3,7-15). Given that the functional effects of genetic variants vary across tissues and cell types, identifying the genetic associations between genetic variants and cell types is of great significance for in-depth understanding of disease mechanisms and the development of targeted therapeutic strategies (4,16-20). Single-cell sequencing technologies offer unprecedented opportunities to identify cellular heterogeneity within complex diseases. To date, several databases integrating single-cell data with GWAS data have emerged,

such as sc2GWAS and SC2disease (21,22). However, existing resources and methods primarily focus on single-cell RNA sequencing (scRNA-seq) data. Notably, it has been confirmed that most of the identified disease-relevant genetic variants are located outside the exome (1,23). Approximately 80-90% of the relevant variant sites are found in non-coding regions (24-26). Therefore, in relevant traits or diseases, the activity of regulatory elements often overlaps with causal genetic variant sites to identify the interested cell types, i.e., enrichment in functional cell types (24,27,28). In practice, a major limitation is the computational burden imposed by a large number of significant variants due to the linkage disequilibrium relationship between effect variants (1,6,27,29). Therefore, the method that assumes the presence of a causal variant effect among multiple effect variant sites is most commonly used, which facilitates large-scale data analysis (5,27,29). In the context of the above observations, genetic fine-mapping can effectively address the problem and help to effectively elucidate the role and impact of genetic variation at the association-to-cell level (25,30).

Currently, the advancement of single-cell genomics has given rise to single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq) technology (31), facilitating the exploration of the functional roles of genetic variants in non-coding regions and at the single-cell level (32). The integration of causal variant effects with scATAC-seq enables the precise identification of genetic variants that play pivotal roles in specific cell types (25). However, compared to scRNA-seq, scATAC-seq data commonly exhibits high sparsity and technical noise (33), significantly increasing the difficulty of performing cell-type-specific genetic inference based on such data. To address this challenge, recently developed innovative methods, g-chromVAR (27) and SCAVENGE (25), effectively integrate causal variant effects with scATAC-seq maps using co-localization techniques, successfully identifying cell types significantly implicated in specific traits. The SCAVENGE method, building upon g-chromVAR, incorporates network propagation to mitigate the sparsity of single-cell data, thereby enhancing analytical performance. These pioneering methods systematically infer the specific cellular contexts in which phenotype-associated variants exert their effects, laying a solid foundation for elucidating the causal mechanisms of phenotypic variation. However, challenges such as lengthy computation times and dependence on high-performance computing in large-scale acquisition and exploration of genetic variation at the single-cell level have hindered convenient utilization for researchers. Simultaneously, the rapid accumulation of large-scale scATAC-seq data and causal variant data has created an urgent need for centralized integration, allowing researchers to effectively utilize these resources to explore the functional roles and impacts of genetic association loci at single-cell resolution.

Here, we developed SCVdb, a platform for documenting variant-to-function mapping relationships at single-cell resolution (Figure 1). It aims to accommodate an extremely vast number of trait-relevant cell pairs at the single-cell level by integrating large-scale scATAC-seq data with complex traits. SCVdb provides comprehensive annotations, gene pathway enrichment analysis, and context-specific regulatory networks for these associated pairs. The current version of SCVdb accommodates over 1,342,000 single cells (derived from ~180 manually collected scATAC-seq samples) and over 15,800 traits (filtered based on fine-mapping results). Applying g-chromVAR and SCAVENGE, the database has established over 15.9 billion trait-cell relationships. The fine-mapping results encompass a comprehensive disease spectrum, including all 22 major categories and over 250 subcategories from the International Classification of Diseases (ICD-10) (34). Moreover, the dataset spans comprehensive phenotypic information, covering indicator, drug and chemical compound, etc. To ensure result reliability, we processed scATAC-seq profiles and genetic fine-mapping results independently via standardized pipelines, using consistent parameters tailored to each data type. To empower researchers to access the landscape underlying potential enrichment associations between phenotypes or diseases and cell types, we provide detailed and abundant regulation information, including TRSs for cells, cell-type-specific differential gene activity and transcription factor (TF), trait-relevant genes and TFs, and systematically constructed trait-cell type association networks, etc. In conclusion, SCVdb is a user-centric database that facilitates the exploration and interpretation of single-cell genetic data through structured displays and dynamic visualizations. It offers intuitive workflows, customizable parameters and comprehensive data accessibility.

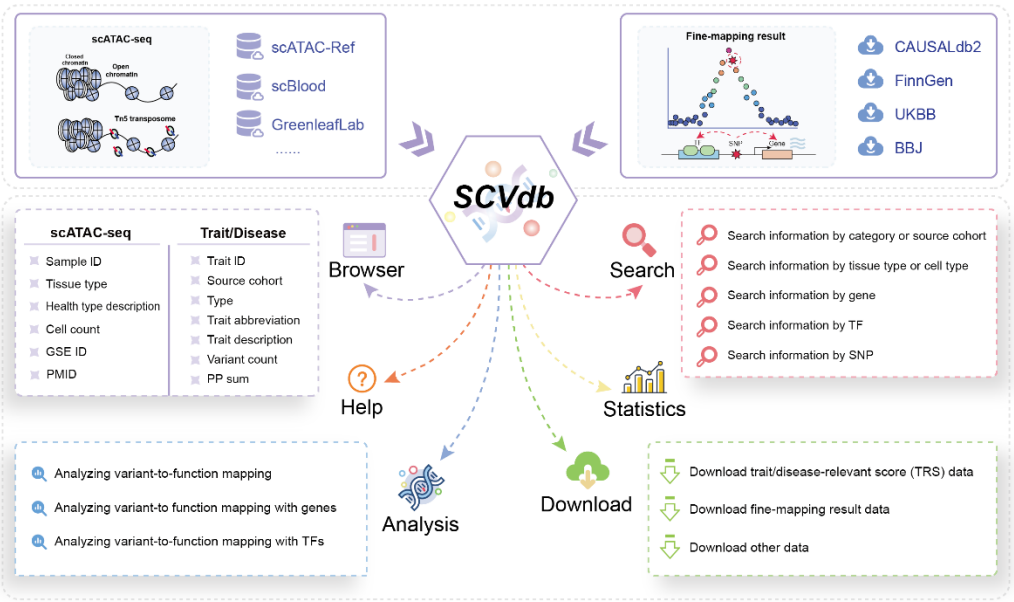


Figure 1. | The architecture of SCVdb. The SCVdb platform has amassed a comprehensive collection of scATAC-seq annotated with cell type labels, as well as extensive fine-mapping results, all preprocessed through a unified pipeline. SCVdb supports multiple functions, including browsing, searching, analyzing, statistics and downloading.

Material and methods

Collection and processing of fine-mapping data.

We collected fine-mapping results of 16,445 traits from four major resources: CAUSALdb2 (35), FinnGen (36,37), UKBB (37,38) and BBJ (37,39) (Figure 1). In order to ensure the reliability of the fine-mapping data in subsequent integrated analyses, we performed uniform preprocessing of the data from different cohorts. For each trait, retain the variants with a causal variant probability value (PP) > 0.001 calculated by FINEMAP (25,40). As a result, 15,805 traits were filtered from the initial 16,445 traits and used for subsequent analysis (Supplementary Table 1). To harmonize genomic coordinates between variants and scATAC-seq data, we performed LiftOver (41) to convert variant positions to match the reference genome version used in single-cell analysis. Next, we manually categorized them into a broad array of classifications, including diseases, indicators, drugs, chemical compound, health care, treatments, therapy and symptom, etc. For the disease category, we specifically annotated them according to ICD-10, encompassing all 22 major disease categories to provide an intuitive and convenient reference for diseases (Supplementary Table 2).

Collection and processing of single-cell chromatin accessibility datasets.

We obtained ~180 scATAC-seq datasets with cell type labels from multiple sources: scATAC-Ref (42), scBlood (43), GreenleafLab laboratory (https://github.com/GreenleafLab/MPAL-Single-Cell-2019) (25) and the PlaqView single-cell data portal (https://www.plaqview.com/) (44) (Figure 1, Supplementary Table 3). These data involve approximately 20 tissue types including blood, brain, skin, heart, kidney, and liver, encompassing around 200 cell types such as lymphocytes, cardiomyocytes, fibroblasts, and tumor cells, and covering more than 1,342,000 cells in total. To ensure data standardization, we used the SnapATAC2 (45) software for unified preprocessing and quality control, including feature selection and doublets identification. Next, we use SnapATAC2 to derive cellular gene activity profiles and identify differential TFs across cell types. To identify differentially active genes among cell types, we employ SCANPY (46), performing the differential activity analysis using the Wilcoxon test. We conducted gene set enrichment analysis using GSEApy on the top 500 genes ranked by differential gene scores for each cell type in each sample.

Estimation of trait relevance for each single cell.

Identifying causal variants relevant cell types underlying the observed associations will provide important reference resources for understanding the complex phenotypic manifestations and disease pathogenesis. Therefore, building upon the collection and preprocess of fine-mapping data across 15,805 traits and

scATAC-seq datasets from ~180 samples, we next linked key genetic variants to cellular functions, which will help to discover the cell-specific mechanisms behind traits. Specifically, we systematically link genetic variants to their functional cellular contexts at single-cell resolution using g-chromVAR and SCAVENGE methods (25,27), with peak-by-cell matrix generated from scATAC-seq data and fine-mapping results of traits as inputs. For each trait, we compute bias-corrected Z-scores of per cell using g-chromVAR by integrating variant posterior probabilities and chromatin accessibility at linked peaks (27). SCAVENGE was applied with default parameters to calculate TRSs for each cell. Briefly, cells ranked high based on g-chromVAR Z-scores were selected as seed cells. A mutual k-nearest neighbor (M-*k*NN) graph was constructed based on latent semantic indexing (LSI) features of the scATAC-seq data, and network propagation was performed to propagate seed cell information, generating TRSs that reflect each cell's relevance to the trait (25). As a result, with ~180 scATAC-seq samples and 15,805 traits, a total of ~2,844,900 trait-sample associations were generated, where each trait is linked to every cell within each sample. Each association is annotated with both g-chromVAR Z-scores and SCAVENGE TRSs to quantify trait relevance at single-cell resolution. A small fraction of these results shows entirely zero TRS values across all cells in specific samples for certain traits, indicating no detectable enrichment of the corresponding trait in those samples. In contrast, a total of over 1,923,500 trait-sample associations exhibited non-zero TRS values in at least one cell per sample, demonstrating detectable trait enrichment. Among these, there are over 13.06 million associations linking traits to distinct cell types within specific samples and more than 15.95 billion trait/disease-cell associations at the single-cell level (Supplementary Table 4).

Trait-relevant genes and TFs

To systematically study causal variant effects, we provide comprehensive trait-related analysis resources: (1) Gene-based analysis: We performed gene-based association analysis using MAGMA (47) to help investigate the potential effects of causal variants on associated genes. In brief, we first use Entrez eUtils to match variant sites lacking rsID annotations with dbSNP to ensure the integrity of MAGMA enrichment (48). Then, variant loci were mapped to genes using MAGMA's annotation function based on GENCODE v41 (49), ensuring consistency with the single-cell data processing pipeline. Gene-level association analysis was performed using MAGMA, incorporating LD structure from the 1000 Genomes Project to account for linkage disequilibrium among variants. (2) Functional enrichment analysis: To characterize trait-associated functions, we performed gene set enrichment analysis using GSEApy (50) across Gene Ontology (BP/CC/MF) (51), KEGG pathways (52), and GWAS Catalog (53) annotations. (3) Motif analysis: We performed motif enrichment analysis using HOMER (54) with the parameter $-size\ 1000$ to assess variant impact on TF binding affinity.

Construction of trait-cell type associations

To help understand how traits affect biological functions and disease susceptibility through specific cell types, we constructed trait-cell type association network linking by gene and TF, referred to as "Gene hub network" and "TF hub network" respectively. The both networks integrate data from two dimensions genetic variation and cell type specificity. (1) Gene hub network: The genetic variation dimension refers to the variant-relevant genes obtained through MAGMA analysis, which represent the variant-gene associations network. Since a trait or disease have recorded numerous susceptibility causal variant loci, this constitutes the trait-variant association network. The cell type specificity dimension is derived from cell type-specific differentially active genes in single-cell samples, forming the gene-cell type association network. These two dimensions of data are used to construct a global regulatory landscape of the trait-variant-gene-cell type-single cell sample network. The key genes, obtained by overlapping the genes contained in two dimensions, represent candidate marker genes that play enriched functional roles and effects in specific cell types for a trait or disease. (2) TF hub network: The genetic variation dimension refers to the trait- or disease-relevant TFs obtained through HOMER analysis, forming the trait-TF association network. From the cell type-specific dimension, we obtained the TF-cell type association network. From these two-dimensional data, we construct a trait-TF-cell type-single cell sample network regulatory landscape. The key TFs were obtained by overlapping the TFs included in the two dimensions.

Gene annotation.

To further explore the manifestations of complex phenotypes and the underlying mechanisms of diseases, SCVdb provides more (epi) genetic annotation data for variant-relevant genes (Supplementary Table 5). We curated genomic regulatory elements including 26.75 million typical enhancers (TEs) and 1.33 million

super-enhancers (SEs). Specifically, TEs were compiled from SEA v3 (55) and SEdb v2 (56), while SEs were aggregated from SEA v3, dbSUPER (57), and SEdb v2. We also incorporated over 37.3 million common SNPs from dbSNP b151 (48) and over 96,900 risk SNPs from GWAS ATLAS (6). These annotations were map to variant-relevant genes using BEDTools (58) software. Additionally, we further annotated genes with 101.5 million expression quantitative trait loci (eQTLs) sourced from GTEx v10 (59), directly linking genetic variants to genes.

Database implementation

The latest version of the SCVdb database operates on CentOS 7.7.1908. We deploy and release the front-end and back-end projects to the remote server separately using Dockerfile in the integrated development environments WebStorm 2025.1.1.1 and IntelliJ IDEA 2025.1.1.1. The remote server has pre-installed Docker 19.03.5 (https://www.docker.com/), and the back-end API along with the front-end pages are reverse-proxied through Nginx 1.22.0 (https://nginx.org/). The project employs a technical architecture that separates the front-end from the back-end. On the back-end side, business logic processing is built upon the Spring Boot 3.0.5 framework (https://spring.io/projects/spring-boot), which is based on Java 17.0.1 (https://dev.java/). For database management, MyBatis-plus 3.5.7 (https://github.com/baomidou/mybatis-plus) serves as the ORM framework, connecting to a MySQL (https://www.mysql.com/) structured database set up through the Docker container version mysql:8.0.32. To enhance system performance, we have introduced the Redis (https://redis.io/) 6.2.11-alpine Docker container version as a caching mechanism. The front-end is developed using the Vue 3.2.4 (https://vuejs.org/) framework within a Node.js v16.13.0 (https://nodejs.org/en) environment. For front-end page development, Axios 0.21.4 (https://www.axiosdev.com.au/) is used for data interaction with the back-end API, while Bootstrap v5.1.3 (https://getbootstrap.com/) and Element-UI (element-plus 2.2.0) (https://element-plus.org/en-US/) provide page layout and style design tools. Font Awesome 6.1.1 (https://fontawesome.com/) provides icon style support. Echarts 5.3.1 (https://echarts.apache.org/en/index.html), Plotly 2.23.0 (https://plotly.com/) and CanvasXpress 38.4.1 (https://canvasxpress.org/) are used for graph visualization. To ensure the smoothest browsing experience, it is recommended that users access the website with modern web browsers that support HTML5 standard, such as **Firefox**, Google Chrome and Edge.

Database content and usage

A user-friendly interface for browsing datasets

The "Data Browser" page of SCVdb offers users two intuitive ways to explore detailed data: browsing by scATAC-seq datasets and by traits (Figure 2A). The scATAC-seq dataset-based browsing is further divided into sample-based and cell type-based browsing modes. Users can filter samples by "tissue type" and "health status", while cell type-based browsing can filter by "tissue type" and "cell type". Trait-based browsing supports filtering by "Type", "ICD-10 category", "ICD-10 subcategory" and "Source cohort". Upon clicking on any interesting "Sample ID" or "Trait ID", detailed information is presented in a new webpage featuring the following eight panels: (1) Overview module: This module provides comprehensive descriptions and visualizations for selected entries. For single-cell datasets, it details key information such as the tissue origin of samples, the number of cell types, cell count, etc., and visually presents the proportional distribution of cell types via pie charts (Figure 2B). For trait data, it offers information such as trait type, ICD-10 disease categories, the number of associated genetic variant sites, etc., along with pie charts showing the chromatin distribution of variant sites (Figure 2C). (2) Trait relevance score Module: This module provides information and visualizations on the associations between target traits and single-cell data calculating by g-chromVAR and SCAVENGE. It includes pie charts of enrichment proportions for single-cell samples or traits (Figure 2D), UMAP plots annotated by cell type and TRSs, and box plots showing the distribution of TRSs across different cell types (Figure 2E). By presenting data from multiple perspectives, this module serves as an important basis for identifying key regulatory cell types associated with genetic traits. (3) Cell type-specific differential genes module: This module provides a heatmap of cell type-specific differential genes and a bubble plot of function enrichment analysis result, offering biological functional insights to support the exploration of trait-relevant cell types (Figure 2F). (4) Trait-relevant genes module: For selected traits, this module presents MAGMA gene-level analysis results and performs function enrichment analysis on related genes. Clicking on a gene allows further access to detailed annotation information, including enhancers, super-enhancers, risk SNPs, and more. (5) Gene hub network module: Trait- or disease-cell type association networks primarily document and reveal key genes underlying phenotypic manifestations or pathogenic mechanisms, as well as the potential regulatory

relationships among these genes. (Figure 2G). (6) Cell type-specific differential TFs module: This module provides tables and heatmaps of differential TFs across different cell types. (7) Trait-relevant TFs module: This module provides a list of TFs enriched in trait-relevant variants predicted by HOMER. (8) TF hub network module: Trait- or disease-cell type association networks primarily document and reveal key TFs underlying phenotypic manifestations or pathogenic mechanisms, as well as the potential regulatory relationships among these gene TFs. (Figure 2H).

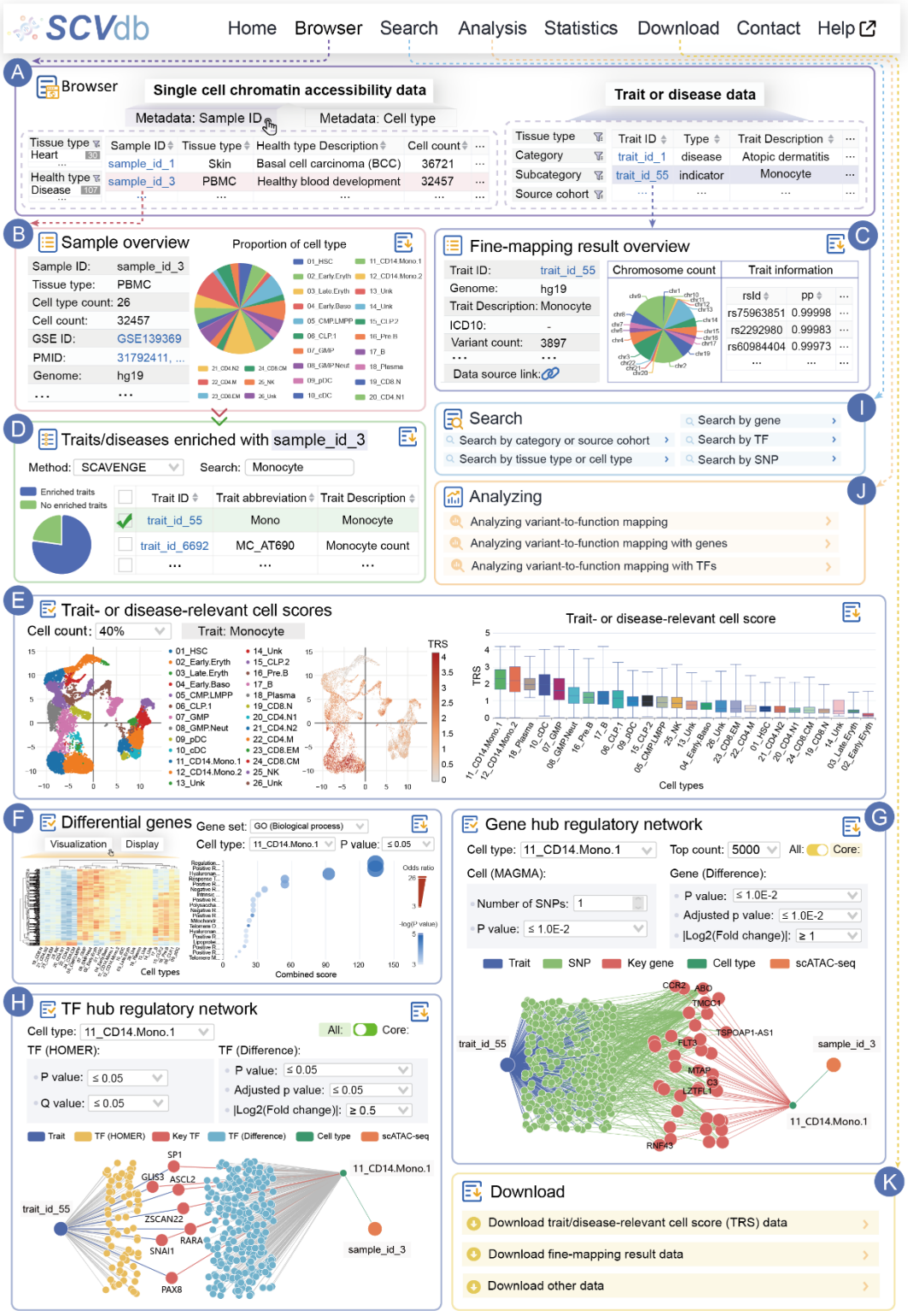


Figure 2. | Main functions and usage of SCVdb. (**A**) An interface for efficient browsing of both scATAC-seq data and fine-mapping results. (**B**) Detailed information for each scATAC-seq sample. (**C**) Detailed information for each trait or disease. (**D**) Enriching trait or disease information for each single-cell scATAC-seq sample. (**E**) Visualization of trait or disease-enriched cells. (**F**) Heatmap and table of differentially active genes across cell types for each scATAC-seq sample, along with functional enrichment bubble plots for highly scored genes. (**G**) Gene hub regulatory network for each trait/disease-single cell sample pair. (**H**) TF hub regulatory network for each trait/disease-single cell sample pair. (**I**) Five search modes are provided. (**J**) Three online analysis tools are provided. (**K**) SCVdb provides a comprehensive collection of downloadable data, including scATAC-seq profiles, trait/disease data, TRS result data, MAGMA gene-level association data, HOMER motif analysis data, differential gene activity and TF activity data, genome annotation data and additional resources.

An interface for convenient retrieving in the "Search" page

The SCVdb search interface provides five query modes for exploring trait-cell associates, including "Search by category or source cohort", "Search by SNP", "Search by tissue or cell type", "Search by gene", and "Search by TF" (Figure 2I). With the "Search by Category or Source Cohort" query, users can flexibly select search strategies by switching between the "Source cohort" and "Category" tabs. After entering a specific "Source cohort" or "Category" entry, the system will return the corresponding list of traits. By clicking a "Trait ID", users can further view detailed information about cells associated with that trait. The "Search by SNP" mode allows users to enter an rsID number to retrieve trait information linked to that particular genetic variant. The "Search by Tissue or Cell Type" mode enables rapid retrieval of relevant scATAC-seq samples. Clicking on the "Sample ID" in the search results enables users to view detailed sample information. In the "Search by Gene" mode, users can query a target gene to obtain its basic information, associated traits and scATAC-seq samples, and regulatory annotations. Similarly, in the "Search by TF" mode, users can retrieve the basic information of a specific TF, related traits and scATAC-seq samples.

Online analysis tools

SCVdb provides three analysis tools to provide users with multi-level genetic regulation research, including "Variant-to-function mapping", "Variant-to-function mapping with genes" and "Variant-to-function mapping with TFs" (Figure 2J). In "Variant-to-function mapping", users can input multiple traits and select specific sample to obtain regulatory potential scores for these traits across various cell types. The "Variant-to-function mapping with genes" and "Variant-to-function mapping with TFs" enables users to input custom gene or TF sets, and flexibly set differential analysis parameters (Log2(Fold change) and p-value) and trait associated parameters (SNP number and p-value). With a single-click analysis, the tool provides scATAC-seq samples and traits related to the target genes or TFs, facilitating the interpretation of genetic regulatory mechanisms of genes or TFs in specific cell types.

Data download

The "Download" page provides a centralized access portal for all downloadable data on SCVdb, designed to help users conveniently obtain the research resources (Figure 2K). Users can quickly locate target datasets through categorized navigation, including: (i) TRS for each single cell in each sample generated through g-chromVAR and SCAVENGE methods; (ii) gene and TF associated analysis results; (iii) gene regulation annotation results, and (iv) scATAC-seq associated analysis results. Additionally, we also provide downloads for all table data and all visualization graphics in detail page.

Case study

Basal cell carcinoma (BCC) is the most common cancer in humans worldwide, and its incidence is on the rise (60-63). Previous fine-mapping studies have identified several genetic variants with potential causal effects within susceptibility regions for BCC. To further explore the potential pathogenic mechanisms of causal variant effects in the single-cell context of BCC, we utilized the "Search information by tissue type or cell type" function of SCVdb to analysis and identify disease-relevant cell populations. We entered "Skin" into the search box and clicked "Search", retrieving the scATAC-seq data table. We selected the single-cell sample "sample_id_1" related to BCC (Figure 3A). Then, in the "Overview" module, detailed information for "sample_id_1" is displayed, which contains 20 cell types and over 3,600 cells (Figure 3B). The "Trait-relevant cell score" module provides a trait table related with this BCC sample, as well as UMAP plots and boxplots showing the TRS scores for each specific trait in each cell. Next, we search for “BCC” in the trait table and select the relevant basal cell carcinoma disease (e.g., trait_id_894) to further investigate its associated cell populations (Figure 3C).

UMAP visualization showed that for trait_id_894, the TRSs in tumor cells (including the four subtypes Tumor 1-4) are significantly higher than those of other cell types. The box plot further confirms the identification of tumor cells closely relevant to the disease within the tumor microenvironment (Figure 3D). This demonstrated that SCVdb has capability to accurately map traits to biologically relevant cell types (60). Notably, when selecting other BCC diseases (e.g., trait_id_895), these analysis results also show good reproducibility (Supplementary Figure 1). To explore functional implications of this trait on tumor cells, we examined differentially active genes in tumor cells. The heatmap revealed overexpression of validated BCC drivers including: *RCC2*, *GLI1*, *CASC15*, *OCA2*, *IRF4,* etc (62,64). The *GLI* family genes encode terminal TFs of the Hedgehog (Hh) signaling pathway, and existing research confirms that hyperactivation of this pathway is a key driver of BCC development (65). The gene *RCC2*, *GLI1*, *CASC15*, *OCA2*, *IRF4* were identified as a trait_id_894-relevant gene based on MAGMA gene-level analysis (47). GWAS (Catalog) term enrichment analysis results of trait_id_894-relevant genes (Figure 3F) demonstrated that trait_id_894 was significantly enriched for multiple skin tumor-relevant functional terms, including "Basal Cell Carcinoma", "Cutaneous Melanoma" and "Cutaneous Squamous Cell Carcinoma". Leveraging the "Gene hub network" module function, SCVdb further constructed a gene hub regulatory network from trait_id_894 to tumor cell subtype with high TRSs (Figure 3G). In the candidate marker genes identified by this network, we found multiple reported BCC susceptibility genes, such as *RCC2*, *PADI6*, *CASC15*, *RGS22*, *LINC00111* and *ZBTB10* (61,62,64,66). In addition, it also includes genes that have been reported to influence pigmentation in Europeans, such as *ASIP*, *SLC45A2* and *OCA2* (61,67-69). To further validate our findings, we used the identified candidate marker gene *RCC2* as an anchor to elucidate its potential regulatory mechanism in BCC (64). In the "Trait-relevant genes" module, click on the *RCC2* gene to enter the gene details page. On this page, in the "Regulation region" panel, we observe that in the epigenomic annotation of the *RCC2* gene, risk SNPs, eQTLs, TEs, and SEs are all confirmed to be associated with skin-relevant diseases (Figure 3I). Meanwhile, in the "Gene-relevant traits or diseases" panel, locate the "trait_id_894" and click the corresponding "View" button to access detailed enrichment information of RCC2 in this trait. Within the "gene-relevant SNP regulatory network" section at the bottom, we identify a causal variant, *rs57142672*, which is located within 500 kb of the *RCC2* transcription start site (TSS). In the pathogenic mechanism of BCC, the regulatory relationship involving *RCC2-rs57142672* has been previously reported in the literature (64). SCVdb provides an opportunity for further validation of causal variants. By clicking on *rs57142672* to enter the SNP details page, we find that *rs57142672* primarily exhibits causal effects in BCC, malignant skin tumors, and cancer. These findings fully demonstrate that SCVdb exhibits high reliability in the identification of cell types associated with traits or diseases, providing researchers with a trustworthy data resource.

Furthermore, within the TF-hub association network from trait_id_894 to tumor cell subtype with high TRSs, we observed a significant correlation between Hypoxia-Inducible Factor 1 Alpha (*HIF1A*) and tumor cell subpopulations (Figure 3H). It is noteworthy that during cancer development, HIFs are described as oncogenes or tumor suppressors under different environmental conditions (70). Previous articles have reported that the *HIF1A* gene plays a potential downregulating role in BCC (71). This result provides guidance for TFs that are highly probable to perform potential regulatory functions in cell types with significant enrichment of traits or diseases.

In summary, SCVdb encompasses multi-level, multi-faceted regulatory annotations for variant-to-function mapping relationships, primarily through cellular trait or disease enrichment analysis, differential gene and differential TF analysis across cell types, MAGMA gene-level analysis, HOMER motif analysis, gene pathway enrichment analysis, gene and TF hub regulatory networks and SNP-relevant regulatory networks. This assists researchers in exploring and revealing the phenotypic manifestations and pathogenic mechanisms of causal variant effects within the cellular environment.

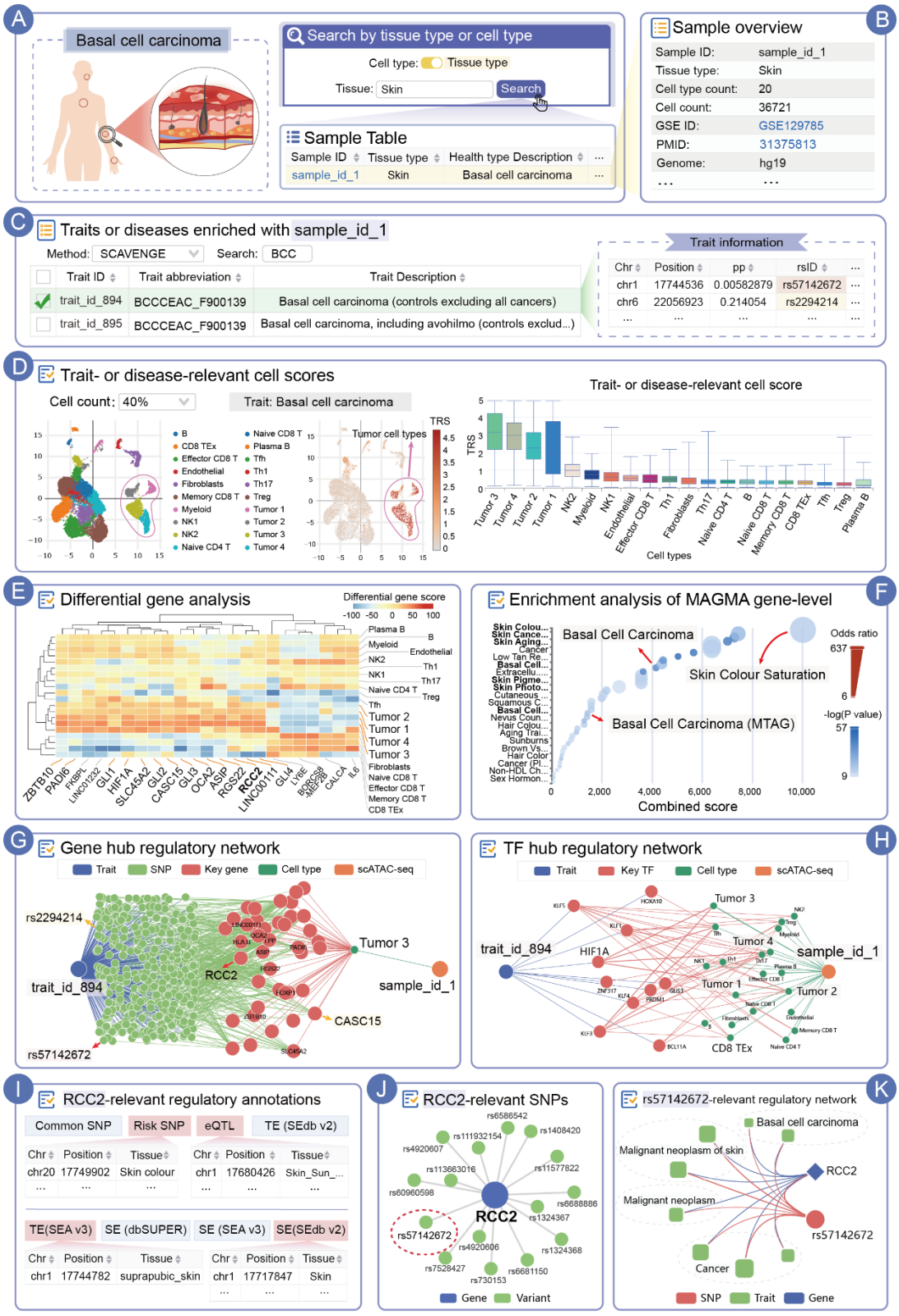


Figure 3. | Case study of basal cell carcinoma (BCC) sample. (**A**) Search for BCC using "Search information by tissue type or cell type". (**B**) Detailed information for BCC sample. (**C**) Enriching trait or disease information for BCC sample. (**D**) Visualization of trait or disease-enriched cells. (**E**) Differential gene heatmap of BCC sample. (**F**) Bubble plot of MAGMA gene-level enrichment analysis for "trait_id_894" disease. (**G**) Gene hub regulatory pathway network from "trait_id_894" to "Tumor 3". (**H**) TF hub regulatory network from "trait_id_894" to "Tumor 3". (**I**) Gene-relevant regulatory annotations. (**J**) Gene-relevant SNP network. (**K**) SNP-relevant regulatory network.

Discussion

Deciphering cell-specific regulatory mechanisms of causal variants on traits or diseases, and exploring the regulatory links between these traits or diseases and fate-relevant cell types, is crucial for genetic medicine progress. To meet this challenge, we developed the SCVdb database platform, leveraging g-chromVAR and SCAVENGE algorithms. SCVdb effectively integrates scATAC-seq and fine-mapping data, providing causal variant to function mapping at single-cell resolution, which elucidating the role and mechanisms of genetic variation at the cellular level. SCVdb not only offers comprehensive trait- or disease-cell and trait- or disease-cell type association relationships but also identifies potential key regulatory genes and TFs involved in these associations. Additionally, it provides epigenetic annotations and gene functional enrichment analyses. These features deliver detailed insights into the functional exploration of complex phenotypes and diseases within cellular contexts. In summary, SCVdb is a user-friendly database that facilitates the searching, browsing, analyzing and visualizing of information related to single-cell genetics. As a comprehensive platform that integrates human single-cell chromatin accessibility data with causal variant effects, SCVdb provides researchers with access to a wealth of scATAC-seq datasets, fine-mapping results and TRS data, along with a variety of analytical tools. SCVdb is expected to significantly advance research in single-cell genetics and clinical medicine.

To further enhance the value of the SCVdb database, we are currently constrained by the limited availability of scATAC-seq data with clear cell-type labels, as well as the accuracy of methods like SCAVENGE. Moving forward, we will continue to incorporate newly emerging scATAC-seq data while also refining our algorithmic pipeline to improve accuracy. We are confident that through continuous improvements, SCVdb will better serve the scientific community by providing researchers with richer and more valuable information resources.

Data availability

The research community can access information freely in the SCVdb without registration or logging in. The URL for SCVdb is https://bio.liclab.net/scvdb/.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Natural Science Foundation of China [62301246, 62272212]; The Science and Technology Innovation Talent Program of Hunan Province of China [2024RC1062]; The Innovation Platform and Talent Program [2023TP1047]; Natural Science Foundation of Hunan Province [2025JJ50401, 2025JJ50105]; Clinical Research 4310 Program of the University of South China [20224310NHYCG05]; Scientific research Project of the Education Department of Hunan Province [24B0417]; The Fundamental Research Funds for the Central Universities (2025-JYB-JBGS-026); Beijing Nova Program [20240484661]; Noncommunicable Chronic Diseases-National Science and Technology Major Project [2024ZD0530800]

Conflict of interest statement

None declared.

References

1. 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. *The American Journal of Human Genetics*, **101**, 5-22.
2. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G. and Meyre, D. (2019) Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, **20**, 467-484.
3. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T. *et al.* (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*, **50**, 1219-1224.

4. 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.
5. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203-209.
6. Watanabe, K., Stringer, S., Frei, O., Umicevic Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M. and Posthuma, D. (2019) A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*, **51**, 1339-1348.
7. Torkamani, A., Wineinger, N.E. and Topol, E.J. (2018) The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*, **19**, 581-590.
8. Chatterjee, N., Shi, J. and Garcia-Closas, M. (2016) Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*, **17**, 392-406.
9. Mars, N., Koskela, J.T., Ripatti, P., Kiiskinen, T.T.J., Havulinna, A.S., Lindbohm, J.V., Ahola-Olli, A., Kurki, M., Karjalainen, J., Palta, P. *et al.* (2020) Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med*, **26**, 549-557.
10. Inouye, M., Abraham, G., Nelson, C.P., Wood, A.M., Sweeting, M.J., Dudbridge, F., Lai, F.Y., Kaptoge, S., Brozynska, M., Wang, T. *et al.* (2018) Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J Am Coll Cardiol*, **72**, 1883-1893.
11. Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J.P., Chen, T.H., Wang, Q., Bolla, M.K. *et al.* (2019) Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*, **104**, 21-34.
12. Elliott, J., Bodinier, B., Bond, T.A., Chadeau-Hyam, M., Evangelou, E., Moons, K.G.M., Dehghan, A., Muller, D.C., Elliott, P. and Tzoulaki, I. (2020) Predictive Accuracy of a Polygenic Risk Score-Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. *JAMA*, **323**, 636-645.
13. Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H. *et al.* (2021) The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat Genet*, **53**, 420-425.
14. Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R. and Domingue, B. (2019) Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*, **10**, 3328.
15. Wand, H., Lambert, S.A., Tamburro, C., Iacocca, M.A., O'Sullivan, J.W., Sillari, C., Kullo, I.J., Rowley, R., Dron, J.S., Brockman, D. *et al.* (2021) Improving reporting standards for polygenic scores in risk prediction studies. *Nature*, **591**, 211-219.
16. Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J. *et al.* (2015) The support of human genetic evidence for approved drug indications. *Nat Genet*, **47**, 856-860.
17. King, E.A., Davis, J.W. and Degner, J.F. (2019) Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet*, **15**, e1008489.

18. Plenge, R.M., Scolnick, E.M. and Altshuler, D. (2013) Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*, **12**, 581-594.
19. Suhre, K., Arnold, M., Bhagwat, A.M., Cotton, R.J., Engelke, R., Raffler, J., Sarwath, H., Thareja, G., Wahl, A., DeLisle, R.K. *et al.* (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*, **8**, 14357.
20. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C., Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M. *et al.* (2021) Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res*, **49**, D1302-D1310.
21. Yin, M., Feng, C., Yu, Z., Zhang, Y., Li, Y., Wang, X., Song, C., Guo, M. and Li, C. (2025) sc2GWAS: a comprehensive platform linking single cell and GWAS traits of human. *Nucleic Acids Research*, **53**, D1151-D1161.
22. Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., Hao, J. and Peng, J. (2021) SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res*, **49**, D1413-D1419.
23. Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J. and Richards, J.B. (2017) Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics*, **19**, 110-124.
24. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190-1195.
25. Yu, F., Cato, L.D., Weng, C., Liggett, L.A., Jeon, S., Xu, K., Chiang, C.W.K., Wiemels, J.L., Weissman, J.S., de Smith, A.J. *et al.* (2022) Variant to function mapping at single-cell resolution through network propagation. *Nature Biotechnology*, **40**, 1644-1653.
26. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K. *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*, **47**, 1228-1235.
27. Ulirsch, J.C., Lareau, C.A., Bao, E.L., Ludwig, L.S., Guo, M.H., Benner, C., Satpathy, A.T., Kartha, V.K., Salem, R.M., Hirschhorn, J.N. *et al.* (2019) Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nature Genetics*, **51**, 683-693.
28. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F. *et al.* (2021) Genome-wide enhancer maps link risk variants to disease genes. *Nature*, **593**, 238-243.
29. Zhang, S., Zhang, H., Zhou, Y., Qiao, M., Zhao, S., Kozlova, A., Shi, J., Sanders, A.R., Wang, G., Luo, K. *et al.* (2020) Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science*, **369**, 561-565.
30. Schaid, D.J., Chen, W. and Larson, N.B. (2018) From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, **19**, 491-504.
31. 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.
32. Corces, M.R., Shcherbina, A., Kundu, S., Gloudemans, M.J., Fresard, L., Granja, J.M., Louie, B.H., Eulalio, T., Shams, S., Bagdatli, S.T. *et al.* (2020) Single-cell epigenomic analyses

implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat Genet*, **52**, 1158-1168.

33. Hao, G., Fan, Y., Yu, Z., Su, Y., Zhu, H., Wang, F., Chen, X., Yang, Y., Wang, G., Wong, K.C. *et al.* (2025) Topological identification and interpretation for single-cell epigenetic regulation elucidation in multi-tasks using scAGDE. *Nat Commun*, **16**, 1691.
34. Organization, W.H. (2019) International statistical classification of diseases and related health problems (ICD-10) (10th ed., 2019 version). *https://icd.who.int/browse10/2019/en*.
35. Wang, J., Ouyang, L., You, T., Yang, N., Xu, X., Zhang, W., Yang, H., Yi, X., Huang, D., Zhou, W. *et al.* (2025) CAUSALdb2: an updated database for causal variants of complex traits. *Nucleic Acids Res*, **53**, D1295-D1301.
36. Kurki, M.I., Karjalainen, J., Palta, P., Sipila, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A. *et al.* (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*, **613**, 508-518.
37. Kanai, M.e.a. (2021) Insights from complex trait fine-mapping across diverse populations. *medRxiv*.
38. Wang, Q.S., Kelley, D.R., Ulirsch, J., Kanai, M., Sadhuka, S., Cui, R., Albors, C., Cheng, N., Okada, Y., Biobank Japan, P. *et al.* (2021) Leveraging supervised learning for functionally informed fine-mapping of cis-eQTLs identifies an additional 20,913 putative causal eQTLs. *Nat Commun*, **12**, 3394.
39. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M. *et al.* (2021) A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*, **53**, 1415-1424.
40. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S. and Pirinen, M. (2016) FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics*, **32**, 1493-1501.
41. Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N. *et al.* (2019) The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*, **47**, D853-D858.
42. Qian, F.-C., Zhou, L.-W., Zhu, Y.-B., Li, Y.-Y., Yu, Z.-M., Feng, C.-C., Fang, Q.-L., Zhao, Y., Cai, F.-H., Wang, Q.-Y. *et al.* (2024) scATAC-Ref: a reference of scATAC-seq with known cell labels in multiple species. *Nucleic Acids Research*, **52**, D285-D292.
43. Zhao, Y., Yu, Z.-M., Cui, T., Li, L.-D., Li, Y.-Y., Qian, F.-C., Zhou, L.-W., Li, Y., Fang, Q.-L., Huang, X.-M. *et al.* (2024) scBlood: A comprehensive single-cell accessible chromatin database of blood cells. *Computational and Structural Biotechnology Journal*, **23**, 2746-2753.
44. Turner, A.W., Hu, S.S., Mosquera, J.V., Ma, W.F., Hodonsky, C.J., Wong, D., Auguste, G., Song, Y., Sol-Church, K., Farber, E. *et al.* (2022) Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat Genet*, **54**, 804-816.
45. Zhang, K., Zemke, N.R., Armand, E.J. and Ren, B. (2023) SnapATAC2: a fast, scalable and versatile tool for analysis of single-cell omics data. *bioRxiv*.
46. Wolf, F.A., Angerer, P. and Theis, F.J. (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*, **19**, 15.
47. de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized

gene-set analysis of GWAS data. *PLoS Comput Biol*, **11**, e1004219.

48. 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**, 308-311.
49. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S. *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*, **22**, 1760-1774.
50. Fang, Z., Liu, X. and Peltz, G. (2023) GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*, **39**.
51. Gene Ontology, C. (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res*, **49**, D325-D334.
52. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. and Ishiguro-Watanabe, M. (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*, **51**, D587-D592.
53. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Gunes, O., Hall, P., Hayhurst, J. *et al.* (2023) The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*, **51**, D977-D985.
54. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**, 576-589.
55. 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.
56. Wang, Y., Song, C., Zhao, J., Zhang, Y., Zhao, X., Feng, C., Zhang, G., Zhu, J., Wang, F., Qian, F. *et al.* (2023) SEdb 2.0: a comprehensive super-enhancer database of human and mouse. *Nucleic Acids Res*, **51**, D280-D290.
57. Khan, A. and Zhang, X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Res*, **44**, D164-171.
58. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841-842.
59. Coorens, T.H.H., Guillaumet-Adkins, A., Kovner, R., Linn, R.L., Roberts, V.H.J., Sule, A., Van Hoose, P.M. and d, G.C. (2025) The human and non-human primate developmental GTEx projects. *Nature*, **637**, 557-564.
60. 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. *Nature Biotechnology*, **37**, 925-936.
61. Stacey, S.N., Helgason, H., Gudjonsson, S.A., Thorleifsson, G., Zink, F., Sigurdsson, A., Kehr, B., Gudmundsson, J., Sulem, P., Sigurgeirsson, B. *et al.* (2015) New basal cell carcinoma susceptibility loci. *Nat Commun*, **6**, 6825.
62. Stacey, S.N., Gudbjartsson, D.F., Sulem, P., Bergthorsson, J.T., Kumar, R., Thorleifsson, G., Sigurdsson, A., Jakobsdottir, M., Sigurgeirsson, B., Benediktsdottir, K.R. *et al.* (2008) Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat Genet*, **40**, 1313-1318.

63. Ressler, J.M., Plaschka, M., Silmbrod, R., Bachmayr, V., Shaw, L.E., Silly, T., Zila, N., Stepan, A., Kusienicka, A., Tschandl, P. *et al.* (2025) Efficacy and tolerability of neoadjuvant therapy with Talimogene laherparepvec in cutaneous basal cell carcinoma: a phase II trial (NeoBCC trial). *Nat Cancer*, **6**, 51-66.
64. Chahal, H.S., Wu, W., Ransohoff, K.J., Yang, L., Hedlin, H., Desai, M., Lin, Y., Dai, H.J., Qureshi, A.A., Li, W.Q. *et al.* (2016) Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat Commun*, **7**, 12510.
65. Yao, C.D., Haensel, D., Gaddam, S., Patel, T., Atwood, S.X., Sarin, K.Y., Whitson, R.J., McKellar, S., Shankar, G., Aasi, S. *et al.* (2020) AP-1 and TGFss cooperativity drives non-canonical Hedgehog signaling in resistant basal cell carcinoma. *Nat Commun*, **11**, 5079.
66. Stacey, S.N., Sulem, P., Gudbjartsson, D.F., Jonasdottir, A., Thorleifsson, G., Gudjonsson, S.A., Masson, G., Gudmundsson, J., Sigurgeirsson, B., Benediktsdottir, K.R. *et al.* (2014) Germline sequence variants in TGM3 and RGS22 confer risk of basal cell carcinoma. *Hum Mol Genet*, **23**, 3045-3053.
67. Gudbjartsson, D.F., Sulem, P., Stacey, S.N., Goldstein, A.M., Rafnar, T., Sigurgeirsson, B., Benediktsdottir, K.R., Thorisdottir, K., Ragnarsson, R., Sveinsdottir, S.G. *et al.* (2008) ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat Genet*, **40**, 886-891.
68. Han, J., Qureshi, A.A., Nan, H., Zhang, J., Song, Y., Guo, Q. and Hunter, D.J. (2011) A germline variant in the interferon regulatory factor 4 gene as a novel skin cancer risk locus. *Cancer Res*, **71**, 1533-1539.
69. Stacey, S.N., Sulem, P., Masson, G., Gudjonsson, S.A., Thorleifsson, G., Jakobsdottir, M., Sigurdsson, A., Gudbjartsson, D.F., Sigurgeirsson, B., Benediktsdottir, K.R. *et al.* (2009) New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet*, **41**, 909-914.
70. Andrysik, Z., Bender, H., Galbraith, M.D. and Espinosa, J.M. (2021) Multi-omics analysis reveals contextual tumor suppressive and oncogenic gene modules within the acute hypoxic response. *Nat Commun*, **12**, 1375.
71. Wan, C. and Li, Y. (2020) Integrative analysis of mRNA-miRNA-TFs reveals the key regulatory connections involved in basal cell carcinoma. *Arch Dermatol Res*, **312**, 133-143.