

scGRN: a comprehensive single-cell gene regulatory network platform of human and mouse

Xuemei Huang^{1,2,3,5,†}, Chao Song^{1,2,5,6,†}, Guorui Zhang^{1,2,5,7,†}, Ye Li^{1,2,5,7,†}, Yu Zhao^{1,2,3,5}, Qinyi Zhang^{2,7}, Yuexin Zhang^{1,2,5,8}, Shifan Fan^{1,2,3,5}, Jun Zhao¹, Liyuan Xie^{1,2,3,5} and Chunquan Li^{1,2,3,4,5,*}

¹The First Affiliated Hospital & MOE Key Lab of Rare Pediatric Diseases, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

²Hunan Provincial Key Laboratory of Multi-omics and Artificial Intelligence of Cardiovascular Diseases & College of Basic Medical Sciences, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

³School of Computer, University of South China, Hengyang, Hunan, 421001, China

⁴Hunan 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

⁵The First Affiliated Hospital, Cardiovascular Lab of Big Data and Imaging Artificial Intelligence, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

⁶The First Affiliated Hospital, Department of Cardiology, Hengyang Medical School, University of South China, Hengyang, China

⁷Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

⁸The First Affiliated Hospital, Institute of Cardiovascular Disease, Hengyang Medical School, University of South China, Hengyang, Hunan, 421001, China

*To whom correspondence should be addressed. Tel: +86 13272311691; Fax: +86 0734 8279018; Email: lcqbio@163.com

†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

Abstract

Gene regulatory networks (GRNs) are interpretable graph models encompassing the regulatory interactions between transcription factors (TFs) and their downstream target genes. Making sense of the topology and dynamics of GRNs is fundamental to interpreting the mechanisms of disease etiology and translating corresponding findings into novel therapies. Recent advances in single-cell multi-omics techniques have prompted the computational inference of GRNs from single-cell transcriptomic and epigenomic data at an unprecedented resolution. Here, we present scGRN (<https://bio.liclab.net/scGRN/>), a comprehensive single-cell multi-omics gene regulatory network platform of human and mouse. The current version of scGRN catalogs 237 051 cell type-specific GRNs (62 999 692 TF–target gene pairs), covering 160 tissues/cell lines and 1324 single-cell samples. scGRN is the first resource documenting large-scale cell type-specific GRN information of diverse human and mouse conditions inferred from single-cell multi-omics data. We have implemented multiple online tools for effective GRN analysis, including differential TF–target network analysis, TF enrichment analysis, and pathway downstream analysis. We also provided details about TF binding to promoters, super-enhancers and typical enhancers of target genes in GRNs. Taken together, scGRN is an integrative and useful platform for searching, browsing, analyzing, visualizing and downloading GRNs of interest, enabling insight into the differences in regulatory mechanisms across diverse conditions.

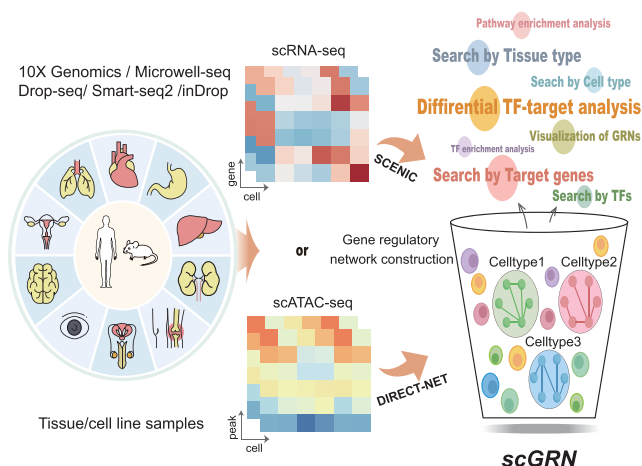
Received: August 15, 2023. Revised: September 19, 2023. Editorial Decision: October 2, 2023. Accepted: October 12, 2023

© The Author(s) 2023. 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-NonCommercial License

(<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Graphical abstract



Introduction

The intricate patterns of gene expression are governed and shaped, to a large extent, by the regulation of transcription factors (TFs), which are sequence specific DNA-binding proteins located in or around *cis*-regulatory elements (CREs) (1). Changes in the abundance or activity of TFs can lead to an increase or decrease in the transcription of their downstream targets (2,3). Gene regulation by TFs is not a linear process, but rather takes place in the context of complex networks that encompass the regulatory interactions between multiple TFs and their downstream target genes (4), termed gene regulatory networks (GRNs), enabling the control of timing, condition and amount where genes are expressed. Uncovering the topology and dynamics of GRNs is fundamental to understanding the establishment and reprogramming of cellular identity and how cell fate is decided (5,6). These networks can also be used to model the changes in gene expression under different conditions, enabling insight into the mechanisms of differential gene expression at a system level (4).

Reconstructing GRNs is a long-standing objective in biology, and a variety of approaches have been created to mine TF–target regulatory relationships and map them onto graphic diagrams (7–12). Early researches on reconstructing GRNs leveraged experimentally validated regulation events compiled in databases (13,14). The advent of next-generation sequencing technology has facilitated the massive *de novo* GRN inference from bulk TF-binding or transcriptomics data (10,15). However, the limitation caused by mixed measures across cell types in a tissue sample has not been overcome, due to the bulk profiling nature of these high-throughput data. The last decade has seen a revolution in single-cell technology, and GRN reconstruction methods using single-cell omics data have been gradually applied to infer cell type-specific TF–gene interactions (16,17). Single-cell RNA sequencing (scRNA-seq) is presently the most commonly used method for inferring single-cell GRNs (scGRNs). For example, SCENIC integrates scRNA-seq co-expression networks with motif binding information to infer GRNs (12). The advent of single-cell transposase-accessible chromatin sequencing (scATAC-seq) has enabled the profiling of open chromatin at a single-cell and genome-wide scale (18), and GRN reconstruction using large-scale epigenomic information generated by scATAC

data turns into a complementary option for refining the inference of TF–gene regulatory relationships. Recently, DIRECT-NET was proposed by Zhang *et al.* (19), where GRNs can be reconstructed using scATAC data alone. Constructing GRNs is a major focus in systems biology because it can provide significant information for drug design and medical research. For example, Kragsteen *et al.* performed gene regulatory analysis to identify the target genes of Norn TFs and found that Tcf21 was a key transcriptional regulator of Norn cells (20). Jain *et al.* discovered several novel targets for the treatment of systemic sclerosis-associated interstitial lung disease (SSc-ILD) by studying the GRNs of SSc-ILD and healthy control lung samples (21).

Multiple resources devoted to providing TF–gene relationships have been created over the past few years, with the purpose of characterizing genome-wide regulatory events (15,22–24). Yet with few exceptions, the TF–target gene regulatory relationships predicted were derived from bulk omics data, and were thus unable to characterize the genome-wide regulatory events specific to a cell type. TRRUST (22,25) was the first database of literature-curated TF–target interactions constructed by using sentence-based text mining, making it a useful benchmark for the computational inference and reconstruction of GRNs. The Gene Regulatory Network Database (‘GRAND’) (23) was developed to provide a sample specific GRN resource, where the differences in phenotypic variations between patients such as sex, age and ethnicity were considered. The database, hTFtarget (15), provided TF–target gene regulatory relationships of human by integrating ChIP-seq data and TF binding site prediction. While the above databases have provided a comprehensive resource for retrieving TF–target gene relationships, an indepth analysis of GRNs at single cell resolution is unachievable because the data is derived from bulk omics data. This limitation has been partly addressed by GRNdb (24), which reconstructs GRNs using scRNA-seq data of diverse human and mouse conditions collected from public databases. However, the continuous accumulation of single-cell transcriptome data has prompted us to create a more comprehensive resource documenting scGRNs (see Supplementary Table S1). Furthermore, using single-cell transcriptome data alone may overlook the effects of epigenomic information, such as open chromatin regions, on

inferring the regulatory relationships between TFs and their target genes.

We have therefore developed scGRN (<https://bio.liclab.net/scGRN/>), a comprehensive single-cell multi-omics gene regulatory network platform across diverse human and mouse conditions, which aimed to record massive GRN information computationally inferred from scRNA-seq and scATAC-seq data, and to provide detailed epigenetics annotations of target genes. At present, scGRN catalogs a total of 62 999 692 cell type-specific TF–target gene pairs, which covers 160 tissues/cell lines and 1324 single-cell samples. These samples have been manually curated from NCBI GEO/SRA (26,27), ENCODE (28), Arrayexpress (29) databases and included multiple single-cell sequencing platforms. To provide a better user experience, scGRN provided multiple types of visualization methods for GRN related information and detailed functional annotations of promoters, super enhancers and typical enhancers in TF binding regions. We have also developed three online analysis tools, which include differential network analysis, TF enrichment analysis and pathway downstream analysis. We anticipate that scGRN will become an integrative and useful platform for exploring potential functions and regulatory mechanisms in diseases and biological processes.

Materials and methods

Data collection and curation

We collected scRNA-seq samples from multiple public databases, including NCBI GEO, ENCODE and ArrayExpress. The scRNA-seq datasets in the GEO database were automatically crawled and downloaded with a text-mining based pipeline where the keywords used for querying data included ‘scRNA-seq’, ‘single cell RNA sequencing’, ‘single-cell RNA-seq’, ‘*Homo sapiens*’ and ‘*Mus musculus*’. After manually reviewing each record collected, we filtered out those samples without raw gene expression profiles. We also downloaded scRNA-seq samples of human and mouse from ENCODE and ArrayExpress. For scATAC-seq data, we collected raw fastq files from NCBI GEO/SRA, which were acquired in a similar manner as from the GEO database. In addition, we manually collated and proofread the meta information for each scRNA-seq and scATAC-seq, including species, data type, tissue type, sequencing platform, cell numbers and related publications.

Data preprocessing

To overcome the influence of technical differences and noise, we adopted a unified quality control pipeline for the preprocessing of scRNA-seq and scATAC-seq data, respectively. First, we performed quality control on the raw gene expression profiles of scRNA-seq samples using the Seurat package (version 4.3.0) (30). Then we normalized the expression profiles using the normalization method ‘LogNormalize’ of global-scaling with a default scale factor and adopted PCA linear dimensional reduction (30). For scATAC-seq data, we first processed raw fastq files by cellranger-atac and then conducted stringent quality control using Signac (version 1.9.0) (31), to remove low-quality cells. In addition, we also normalized the cell-peak matrix based on the term frequency-inverse document frequency (TF-IDF) method. The low-dimensional representation of each cell was extracted by performing singular

value decomposition on a transformed single-cell chromatin accessibility peak count matrix via TF-IDF (31).

Cell clustering and annotation of scRNA and scATAC data

After data preprocessing, cell clustering of scRNA-seq data was performed using the Seurat standard pipeline based on the normalized gene expression profiles. The cell type of each cluster was subsequently identified using the automated annotation method, SingleR (version 2.2.0) (32). For scATAC-seq data, we used the Signac R package for cell clustering of each sample. We also used the ‘ATACCalculateGenescore’ function in the MAESTRO R package (version 1.5.1) (33) to convert peak matrices to gene activity score matrices and performed cell type annotation using SingleR. Clustering of cell types from both the scRNA-seq and scATAC-seq was visualized using the uniform manifold approximation and projection (‘UMAP’) dimensional reduction method.

Gene regulatory network reconstruction

From single-cell transcriptomics data

We adopted the pySCENIC pipeline (version 0.11.2) (34) to infer GRNs using the gene expression matrix of scRNA-seq datasets and the known TF-motif annotations (Supplementary Figure S1A). First, we used GENIE3 (10) and GRN-BOOST2 (35) to identify potential TF targets based on TF–gene co-expression modules, respectively. However, GRNs predicted using co-expression alone contained many false positives and indirect targets. To overcome this limitation, each co-expression module was then pruned using RcisTarget (34) to identify direct targets (regulons) with motif support. These motifs were from two databases (10 kb around the TSS and 500 bp upstream from TSS), and we retained those annotated to the corresponding TF having a Normalized Enrichment Score (‘NES’) >3.0. Finally, the activities of these regulons were quantified in each cell with AUCell (34), and we calculated the cell-type specific score of each regulon using the regulon activity score (‘RAS’) matrix and cell type barcode information. Cell type-specific regulons were defined as those with RSS >0.1 and ranked in the top 10 of each cell type.

From single-cell chromatin accessibility data

We used the DIRECT-NET (version 1.0.0) (19) to reconstruct GRNs from single-cell chromatin accessibility data (i.e. scATAC data) (Supplementary Figure S1B). First, we transformed the cell-peak matrix of each scATAC sample with MAESTRO (33) to a gene activity matrix, from which differentially expressed genes (DEGs) of each cell-type cluster with Bonferroni corrected $P \leq 0.01$ and average \log_2 -fold-change (\log_2FC) values of ≥ 0.25 were detected. For each DEG, 500bp upstream from transcription start site (TSS) was treated as a promoter, and peaks within 250 kb upstream and downstream the TSS were defined as candidate functional regions (19). We then constructed a new feature matrix, where the rows represented similar cells and the columns represented promoters and candidate regulatory regions, by aggregating signals across similar cells, which were identified from a k-nearestneighbor (‘KNN’) graph (default $k = 50$). Third, we regressed the accessibility of promoter expression levels by the accessibility of distally candidate functional regions, and those regions with importance scores higher than the maximum of the median of importance scores and value of 0.001 were

regarded as high confidence functional CREs. Finally, cell type-specific GRNs were reconstructed by identifying TFs bound to the promoters and HC CREs using motif enrichment analysis, by using the `motifmatchr` function in the ChromVAR package (version 1.20.2) (36) to identify the enriched TFs of differential accessible HC CREs. Moreover, we also calculated Spearman's correlation scores between TFs and target genes to provide references.

Annotation of target genes in GRNs

We first obtained 1 717 744 super-enhancer (SE) regions and 79 709 120 typical-enhancer (TE) regions from SEDb 2.0 (37), which were identified from 1739 human and 931 mouse H3K27ac ChIP-seq samples. Then, we mapped the target genes of these SE and TE regions using four linking strategies, including the closest active gene, overlapping gene, proximal gene and closest gene (38). We defined the basal domain 2 kb upstream and downstream of the TSSs as the promoter regions of target genes.

To identify TFs bound to target genes' promoters, SEs and TEs, we collected 51 616 973 and 32 985 444 non-redundant binding regions of 817 human and 648 mouse TFs, respectively, across various cell lines and tissue types from ReMap 2022 (39). TF binding peaks that overlapped with the promoter, SE or TE regions of target genes in all GRNs were identified using BEDTools (v2.25.0) (40). We further collected >3000 DNA binding motifs of ~700 TFs from TRANSFAC and MEME, and used FIMO to identify Motif occurrences within gene promoter, SE or TE regions with a threshold of $P < 1e^{-6}$.

Platform use and access

Overview

The main framework and functions of scGRN are shown in Figure 1. The scGRN currently contains 1324 scRNA-seq and scATAC-seq samples from multiple sequencing platforms, such as 10X Genomics, Drop-seq, Microwell-seq, in-Drop, Smart-seq, HyDrop, sciATAC-seq and others. These samples included 6 808 724 cells from 160 tissues/cell lines, and samples of both disease and healthy conditions were included. We used a unified pipeline and software parameters for GRN inference from scRNA-seq and scATAC-seq data, respectively. The scGRN provided visualizations such as clustering, cell type annotation, heatmap of transcription factor activity, and inferred cell type-specific GRNs for each sample. Moreover, scGRN provided detailed functional annotations of promoters, super enhancers and typical enhancers in TF binding regions. In addition, the scGRN provided three online analysis tools for users, including TF enrichment analysis, differential network analysis and pathway downstream analysis. Overall, scGRN was a user-friendly platform to query, analyze and visualize information associated with scGRNs.

A search interface for retrieving scGRNs datasets

scGRN enables users to search, browse, analyze, visualize and download GRNs of interest (Figure 2A). We provided four different query methods on the 'Search' page for searching scGRN-related information, including 'Search by TFs', 'Search by Target Genes', 'Search by Tissue Types' and 'Search

by Cell Types' (Figure 2C). The first query mode, 'Search by TFs', was designed for biologists who were interested in some particular TFs. The users first selected a species (*Homo sapiens* or *Mus musculus*) and data type (scRNA-seq or scATAC-seq) they intend to study and then input the TFs of interest. Finally, after inputting this information, clicking the 'Search' button directs the users to all scGRN datasets associated with the inputted TFs. The query mode 'Search by Target Genes', similar to 'Search by Target Genes', was designed specifically for users who were interested in specific target genes mapped in scGRNs. The only difference was that the users should enter the target gene names of interest, rather than TF names. When users are interested in a particular tissue type or cell type, 'Search by Tissue Types' and 'Search by Cell Types' should be the optimal query mode. In the 'Search by Tissue Types' query mode, the users first select a species and data type they want to study and then the tissue type of interest. In the 'Search by Cell Types' search mode, the users query scGRNs in cell types of interest by selecting a species, data type and entering the interested cell types. Finally, clicking the 'Search' button directs the users to the corresponding datasets in tissue types or cell types that were searched.

A summary table of associated samples is presented in the result page, and brief information of each sample, such as 'Sample ID', 'Species' and 'Platform', is shown in this table. The user is directed to a details page after clicking a particular 'Sample ID' where three panels are presented (Figure 2D). These include 'Sample Overview', 'Visualization' and 'TF-target network'. We provided a basic description for the selected sample in the left part of the 'Sample Overview' panel, and detailed information, such as cell distribution, statistics on regulatory networks and quality control results, are shown with graph representations in the right part. Regarding scRNA-seq samples in the visualization panel, we provided multiple visualization methods for SCENIC (GENIE3) and SCENIC (GRNBOOST2), including 'Cell cluster & TF activity', 'Regulon module', 'Regulon specificity' and 'Regulon activity', respectively. We provide 'Cell cluster & TF activity', 'TF module' and 'TF activity heatmap' in the visualization panel for scATAC-seq data. The basic information of each TF in the visualization panel was also provided, which was collected from the AnimalTFDB 4.0 (41) database. The GRN information is presented in the TF-target network panel where the search involving GRNs is shown in the form of networks. We support network selection based on inference method, cluster-cell type and TF names, and the TF-target gene brief information is organized as a table, which is placed at the bottom. After clicking 'Target Gene' in the TF-target network information, a detailed description of each gene is displayed on a new page, including gene overview, annotation of TF binding regions of target genes of interest, and gene expression atlas from different sources, such as GTEx (42), CCLE (43), TCGA (<https://cancergenome.nih.gov/>) and ENCODE.

A user-friendly interface for browsing scGRNs datasets

The 'Browse' page was organized as an interactive and alphanumerically sortable table that allows users to quickly browse samples and customize filters including 'Species', 'Platform', 'Data Type', 'Tissue Type' and 'Biosample Type'

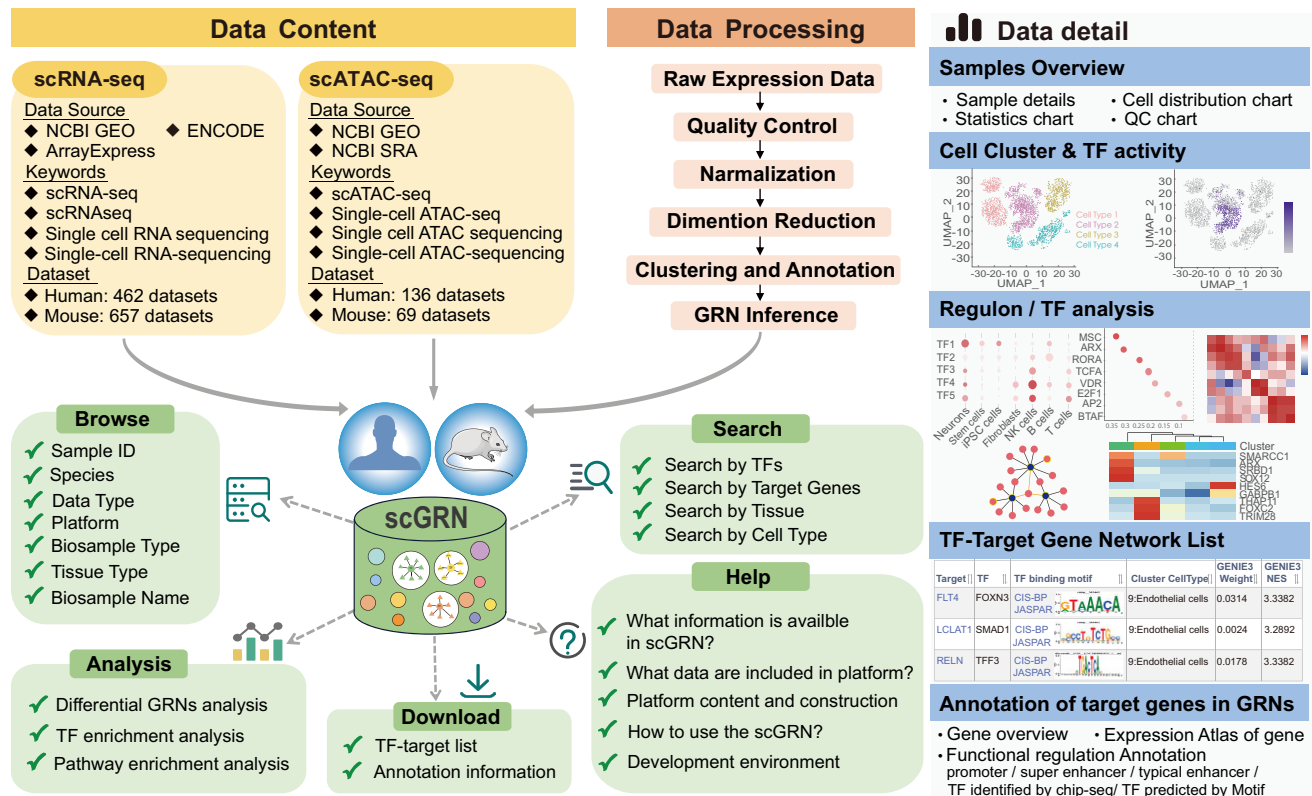


Figure 1. Platform content and construction. scGRN collected abundant single-cell multi-omics data of human and mouse and used a unified pipeline to infer GRNs. scGRN supported multiple functions for browsing, searching, analyzing, visualizing and downloading GRN-related information.

(Figure 2B). Users can use the ‘Show entries’ drop-down menu to obtain a different number of records per page. To further view the details of a given sample, users only need to click on the ‘Sample ID’.

Effective online tools for gene regulatory network analysis

We have implemented three online tools for effective gene regulatory network analysis, including ‘Differential TF–target network analysis’, ‘TF enrichment analysis’ and ‘Pathway downstream analysis’ (Figure 2F). ‘Differential TF–target network analysis’ was implemented to compare the network difference of any two samples. First, the users determine the species, data type and the network inference method. Then, the users should select two samples to be analyzed. Finally, clicking ‘Analyze’ will return the differential network analysis results, including ‘Sample Overview’, ‘Global differential network’ and ‘Differential network between cell types’. Using ‘TF enrichment analysis’, the users should first provide a list of genes or upload a file containing the gene names, and then set the parameters used by the hypergeometric test. At the same time, the samples for TF enrichment should also be selected. Running the analysis will return the enrichment results of all genes submitted by users, which are organized as a summary table containing the detailed TF enrichment information of these genes. For improving intuition, we also provided a bubble chart and bar plot of the enrichment results. The function of ‘Pathway downstream analysis’ is to identify the enriched pathways of TFs or target genes in a selected sample. Clicking ‘Analyze’ will return the enriched pathways in the selected pathway database.

Data download and help interface

Graphs and tables in scGRN facilitate downloading, such as the visualization results and TF information (Figure 2G). The scGRN provides multiple download methods in the download page, such as batch download of all samples or some samples. We supported the download of marker genes, cell type annotation, UMAP coordinates, raw and cell type specific GRNs of all the samples in scGRN. We also provided the download of functional annotations and the expression atlas of target genes. Additionally, the ‘Help’ page provides a detailed tutorial for users.

Case study

Case study of BRCA1. The human breast cancer type 1 (BRCA1) is a tumor suppressor gene responsible for repairing DNA damage and maintaining chromosomal stability. Loss or mutations in BRCA1 can cause an increased risk for breast cancer (44), suggesting BRCA1 plays a central role in tumorigenesis (45). Transcription factor, E2F1, has been proposed by Siervi *et al.* to be involved in both the activation and repression of BRCA1 activity (46). To validate the regulatory relationship between E2F1 and BRCA1, we queried scGRN with ‘Search by Target genes’ (Figure 3A). First, the ‘Species’ and ‘Data Type’ options were defaulted to ‘human’ and ‘scRNA-seq’, respectively. Then, we input ‘BRCA1’ gene in the ‘Genes’ box and clicked the ‘Search’ button. The search results page then listed all samples that involved the BRCA1 gene. We retained samples with ‘Tissue Type’ breast, and a breast cancer sample with ID ‘sample_1_h_433’ was identified (Figure 3B). We ranked TF-BRCA1 regulatory relationships by GENIE3 weight or NES score and found that E2F1 was the top ranked

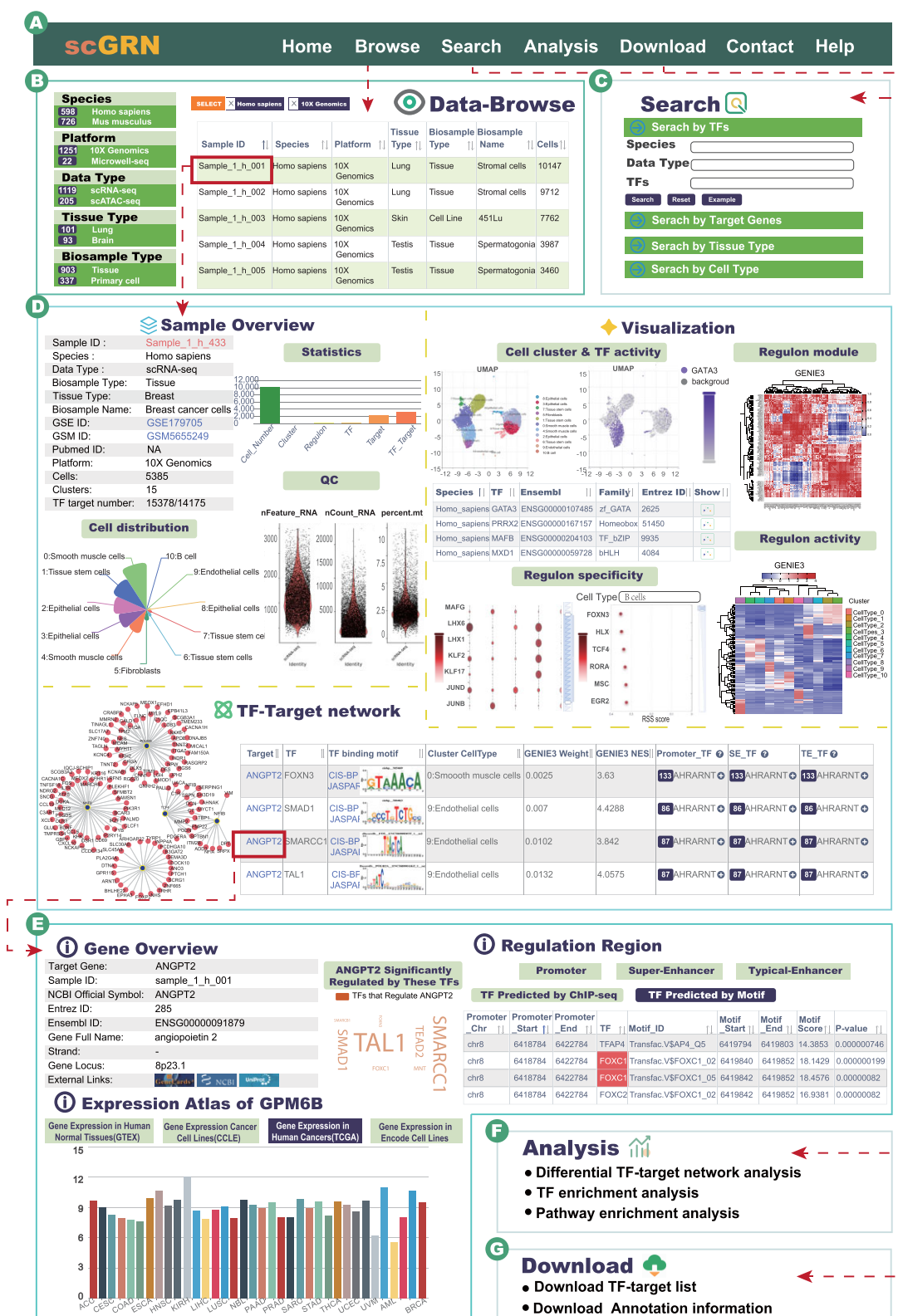


Figure 2. Main functions and usage of scGRN. (A) Navigation bar of scGRN. (B) Browse page of scGRN. (C) Four query modes are provided including 'Search by TFs', 'Search by Target Genes', 'Search by Tissue Type' and 'Search by Cell Type'. (D) Detailed information of each sample including 'Sample Overview', 'Visualization' and 'TF-target Network'. (E) Detailed information of each target gene. (F) Three online analysis tools are provided including 'Differential TF-target network analysis', 'TF enrichment analysis' and 'Pathway downstream analysis'. (G) scGRN allows users to download TF-target list and annotation information.

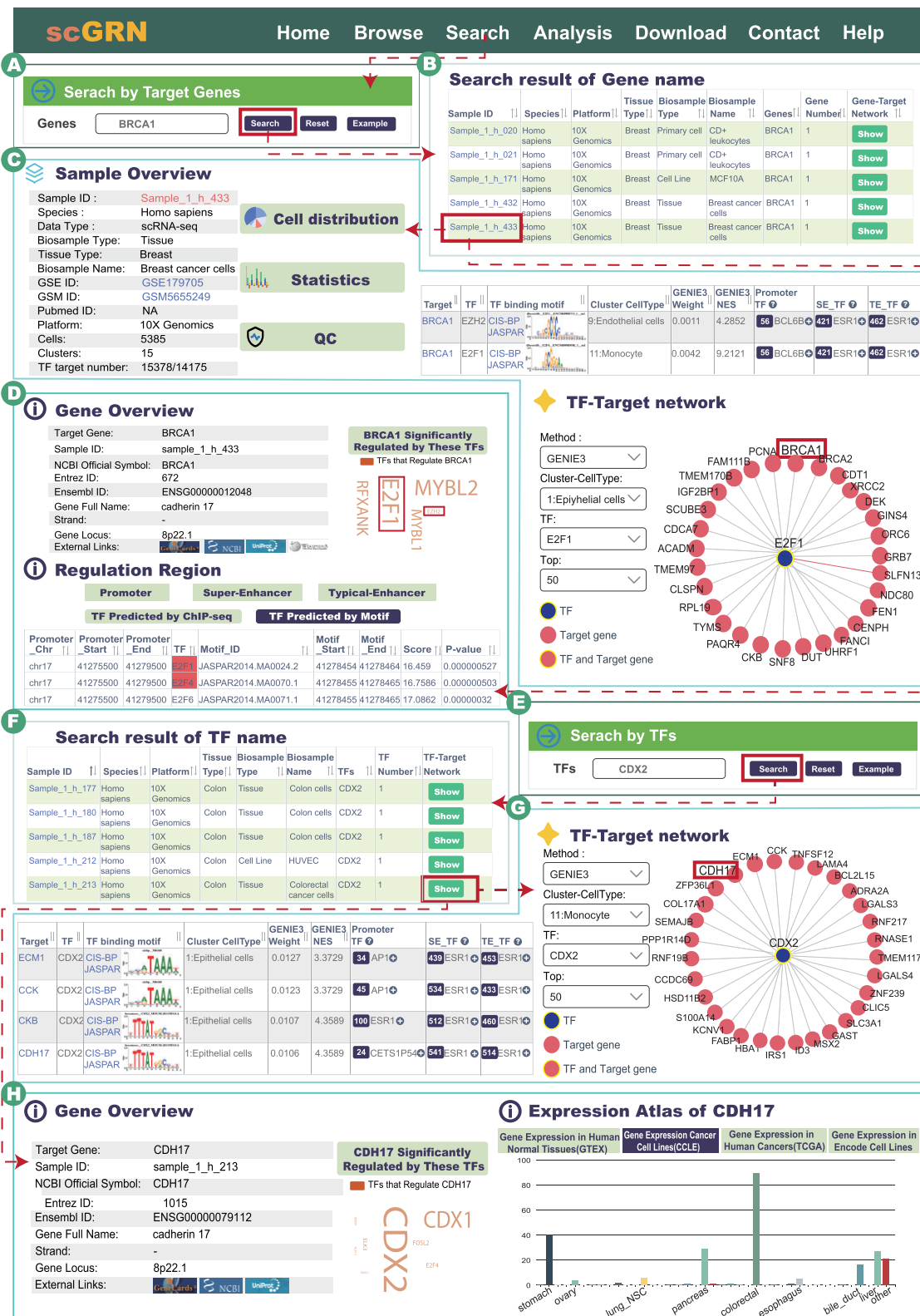


Figure 3. Case studies of scGRN. (A) *BRCA1* is used as the input of 'Search by Target Genes'. (B) Search results of *BRCA1*. (C) Regulatory relationships between *E2F1-BRCA1* and *EZH2-BRCA1* are validated in 'Sample_1_h_433'. (D) Detailed information of *BRCA1* including 'Gene Overview' and 'Regulation Region'. (E) *CDX2* is used as the input of 'Search by TFs'. (F) Search result of *CDX2*. (G) Regulatory relationship between *CDH17* and *CDX2* is validated in 'Sample_1_h_213'. (H) Detailed information of *CDH17*.

TF in this breast cancer sample (Figure 3C). Transcription factor, EZH2, is a critical regulator of cellular memory and mutations or over-expression of EZH2 have been linked to multiple cancers, including breast, prostate, melanoma and bladder cancers. Gonzalez *et al.* reported that EZH2 regulated the nuclear/cytoplasmic shuttling of BRCA1 in benign and breast cancer cells (47). We have also successfully verified this regulatory relationship in ‘sample_1_h_433’ (Figure 3D). The detailed information of the *BRCA1* gene is shown in Figure 3D, including ‘Gene Overview’ and ‘Regulation region’.

Case study of CDX2. We used colorectal cancer (CRC) as another case to illustrate the use of the scGRN. The transcription factor, CDX2, plays an essential role in intestinal development and differentiation and has been linked to the progression of CRC (48,49). Reduced expression of CDX2 is commonly associated with more advanced tumor stage, vessel invasion and metastasis (50). To understand the mechanisms of CDX2 underlying intestinal cell fate specification, Hinoi *et al.* (51) used high-density oligonucleotide arrays to identify downstream target genes associated with the overexpression of CDX2 in a colon cancer cell line. They found that the gene expression of cadherin 17 (*CDH17*) was strongly associated with CDX2. To validate this regulatory relationship, we queried scGRN by ‘Search by TFs’ with CDX2 as input (Figure 3E). Then we retained those samples with ‘Biosample Name’ as ‘Colorectal cancer cells’, and a sample with ID ‘sample_1_h_213’ was identified (Figure 3F). We ranked CDX2-target gene relationships in this sample and found that *CDH17* was the top-ranked target gene for CDX2 in this sample (Figure 3G). The detailed information of the *CDH17* gene is shown in Figure 3H including ‘Gene Overview’ and ‘Expression Atlas of *CDH17*’, and we found that *CDH17* was overexpressed in colorectal cancer cell lines.

Case study of online analysis. To illustrate the usage of scGRN analysis, we used the ‘TF enrichment analysis’ to identify the regulatory TFs of gene sets of interest. First, we obtained 1791 DEGs ($\log_2\text{FCI} > 1$, $P\text{-adj} < 0.05$) of cardiac hypertrophy disease from the Cis-Cardio (52) database as input (Supplementary Table S2). We then selected the species as mouse (default values were used for other parameters, such as $P\text{-value} < 0.01$) and the sample with ID ‘sample_2_m_021’ related to myocardial infarction. Finally, we clicked the ‘Analyze’ button for enrichment analysis (Supplementary Figure S2A). In the ‘TF enrichment analysis’ result page, we found that Gata4 was significantly enriched and ranked in the top10 of all TFs with an enrichment $P\text{-value} = 4.29\text{e-}41$ (Supplementary Figure S2B). During the development of the heart, Gata4 regulates the proliferation, differentiation and fate determination of cardiomyocytes (53). Previous studies have reported that Gata4 was not only a crucial regulator in cardiovascular disease, and also a marker gene of cardiomyocytes (54). Consistent with previous studies, we verified that the activity of transcription factor Gata4 was highly expressed in cardiomyocytes in ‘sample_2_m_021’ (Supplementary Figure S2C). In addition, an important transcription factor, Nkx2-5, was also enriched, which has been used as a treatment target for cardiovascular disease (55).

Discussion

Uncovering the topology and dynamics of GRNs has profound implications for the clinical therapeutics of many diseases (56), such as cancers, cardiovascular diseases and neu-

rological disorders, which are associated with dysregulations of gene regulations. Such knowledge could be used, for example, to inhibit or activate key nodes in gene regulatory networks, thus achieving interventions and treatments for many diseases. The recent advances in single-cell techniques have prompted the rapid accumulation of single-cell multi-omics data, enabling the reconstruction of large-scale cell type- and state-specific GRNs at an unprecedented resolution. There are few databases that archive single-cell regulatory network data except GRNdb, HTCA (57) and AgeAnno (58,59), and we noted that the scGRNs were far from comprehensive in terms of quantity and coverage for sample conditions. Moreover, there is currently no existing platform documenting scGRNs inferred from single-cell multi-omics data (i.e. scRNA-seq and scATAC-seq). More importantly, previous studies have shown that the TFs regulate the transcriptional activity of their target genes by binding to promoters, SE and TE regions (60,61). For example, TFs bind to promoters to precisely control gene activity and expression, thereby determining cell fate and function (62). Super enhancers or enhancers can recruit multiple TFs to form transcription complexes to participate in gene regulation (61,63). Thus, providing a large number of epigenetic annotations of target genes has great biological significance for the study of GRNs, which can help researchers understand the topology and functions of GRNs.

The above limitations have prompted us to establish scGRN, a user-friendly platform to query, analyze and visualize information associated with scGRNs. scGRN is the first and currently the largest resource documenting GRN information computationally inferred from single-cell multi-omics data. The advantages of scGRN are reflected in multiple aspects, including: (i) interactive TF–target gene network visualization and detailed TF–target gene pair tables; (ii) detailed epigenetic annotations about TFs binding to promoters, SEs, and TEs of downstream target genes; (iii) heatmap and scatter diagrams of TF activity and basic TF information tables; (iv) a user-friendly browsing interface; (v) four useful search methods to access corresponding GRNs; and (vi) useful and full-featured online analysis tools such as ‘Differential TF–target network analysis’, ‘TF enrichment analysis’ and ‘Pathway downstream analysis’. scGRN therefore has the potential to benefit cell/molecular biologists, geneticists and data scientists, in identifying differences in regulatory mechanisms across diverse conditions.

Multiple methods have been created to reconstruct GRNs from a variety of data types, such as bulk transcriptomics data, TF binding data, scRNA-seq data and scATAC-seq data. WGCNA (64) reconstructs undirected gene co-expression networks from bulk transcriptomics data, whereas the resulting network contains a large number of false positive associations and lacks interpretability. SCODE (65) and SINGE (66) are GRN inference methods for reconstructing GRNs from ordered scRNA-seq data. However, compared to GENIE3 and GRNBOOST2, these two methods perform poorly in terms of accuracy (11). Methods such as GENIE3 (10) and its faster implementation GRNBOOST2 (35) address these limitations, enabling the reconstruction of directed GRNs from bulk transcriptomics data or scRNA-seq data. However, inferring gene regulatory networks from transcriptome data alone still has a large number of false positives due to the neglect of other mechanisms involved in gene regulation, such as TF binding. Based on GENIE3 and GRNBOOST2, SCENIC (12) generates cell type-specific directed GRNs by

exploiting TF–gene co-expression patterns and TF binding motif data. DIRECT-NET (19) is the only existing method supporting cell type-specific GRNs inference from scATAC-seq data, where GRNs are reconstructed by regressing the accessibility of each gene promoter using the accessibility values of candidate *cis*-regulatory elements (CREs), and thus the influences of distal regulatory elements on inferring TF–gene relationships are taken into account. Recent advances in single-cell techniques have enabled the simultaneous profiling of transcriptomics and chromatin accessibility data, thus greatly improving the GRN reconstruction performance compared using a single type of data.

A limitation of our study was that scGRN currently only documents scGRNs predicted using scRNA-seq or scATAC-seq data, and not including those inferred using both scRNA-seq and scATAC-seq data due to their limited available samples at this stage. However, continuous advances in single-cell multi-omics techniques have led to the development of novel computational methods for more reliable scGRNs inferences using multimodal profiling samples, such as ANANSE (59), CellOracle (67), DeepMAPS (68) and SCENIC+ (69). We will constantly monitor this emerging area and update scGRN in future versions. We believe that scGRN will become an integrative and useful platform for identifying potential functions and regulatory mechanisms of scGRNs in diseases and biological processes.

Data availability

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

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Natural Science Foundation of China [62171166]; Research Foundation of the First Affiliated Hospital of University of South China for Advanced Talents [20210002-1005 USCAT-2021-01]; China Postdoctoral Science Foundation [2019M661311]; Postdoctoral Science Foundation of Heilongjiang Province of China [LBH-Z19075]; Natural Science Foundation of Hunan Province [2023JJ40594, 2023JJ30536]; Clinical Research 4310 Program of the University of South China [20224310NHYCG05]; Funding for open access charge: National Natural Science Foundation of China [62171166].

Conflict of interest statement

None declared.

References

- Chen, K. and Rajewsky, N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.
- Hobert, O. (2008) Gene regulation by transcription factors and microRNAs. *Science*, **319**, 1785–1786.
- Feng, C., Song, C., Liu, Y., Qian, F., Gao, Y., Ning, Z., Wang, Q., Jiang, Y., Li, Y., Li, M., *et al.* (2020) KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res.*, **48**, D93–D100.
- Macneil, L.T. and Walhout, A.J. (2011) Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.*, **21**, 645–657.
- Olson, E.N. (2006) Gene regulatory networks in the evolution and development of the heart. *Science*, **313**, 1922–1927.
- Davidson, E.H. and Erwin, D.H. (2006) Gene regulatory networks and the evolution of animal body plans. *Science*, **311**, 796–800.
- Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.*, **2**, 343–372.
- Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Caestani, C., Yuh, C.-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C.S., *et al.* (2002) A genomic regulatory network for development. *Science*, **295**, 1669–1678.
- Snyder, M. and Gallagher, J.E.G. (2009) Systems biology from a yeast omics perspective. *FEBS Lett.*, **583**, 3895–3899.
- Huynh-Thu, V.A., Irrthum, A., Wehenkel, L. and Geurts, P. (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One*, **5**, e12776.
- Pratapa, A., Jaliha, A.P., Law, J.N., Bharadwaj, A. and Murali, T.M. (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods*, **17**, 147–154.
- Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., *et al.* (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*, **14**, 1083–1086.
- Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V., Jagodnik, K.M., Kropiwnicki, E., Wang, Z. and Ma'ayan, A. (2019) ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.*, **47**, W212–W224.
- Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D. and Saez-Rodriguez, J. (2019) Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.*, **29**, 1363–1375.
- Zhang, Q., Liu, W., Zhang, H.M., Xie, G.Y., Miao, Y.R., Xia, M. and Guo, A.Y. (2020) hTFtarget: a comprehensive database for regulations of Human transcription factors and their targets. *Genomics Proteomics Bioinformatics*, **18**, 120–128.
- Shu, H., Zhou, J., Lian, Q., Li, H., Zhao, D., Zeng, J. and Ma, J. (2021) Modeling gene regulatory networks using neural network architectures. *Nat. Comput. Sci.*, **1**, 491–501.
- Dong, X., Tang, K., Xu, Y., Wei, H., Han, T. and Wang, C. (2022) Single-cell gene regulation network inference by large-scale data integration. *Nucleic Acids Res.*, **50**, e126.
- 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.
- Zhang, L., Zhang, J. and Nie, Q. (2022) DIRECT-NET: an efficient method to discover *cis*-regulatory elements and construct regulatory networks from single-cell multiomics data. *Sci. Adv.*, **8**, eabl7393.
- Kragesteen, B.K., Giladi, A., David, E., Halevi, S., Geirsdóttir, L., Lempke, O.M., Li, B., Bapst, A.M., Xie, K., Katzenelenbogen, Y., *et al.* (2023) The transcriptional and regulatory identity of erythropoietin producing cells. *Nat. Med.*, **29**, 1191–1200.
- Papazoglou, A., Huang, M., Bulik, M., Lafyatis, A., Tabib, T., Morse, C., Sembrat, J., Rojas, M., Valenzi, E. and Lafyatis, R. (2022) Epigenetic regulation of profibrotic macrophages in systemic sclerosis-associated interstitial lung disease. *Arthritis Rheumatol.*, **74**, 2003–2014.
- Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., *et al.* (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.*, **46**, D380–D386.

23. Ben Guebila, M., Lopes-Ramos, C.M., Weighill, D., Sonawane, A.R., Burkholz, R., Shamsaei, B., Platig, J., Glass, K., Kuijjer, M.L. and Quackenbush, J. (2022) GRAND: a database of gene regulatory network models across human conditions. *Nucleic Acids Res.*, **50**, D610–D621.
24. Fang, L., Li, Y., Ma, L., Xu, Q., Tan, F. and Chen, G. (2021) GRNdb: decoding the gene regulatory networks in diverse human and mouse conditions. *Nucleic Acids Res.*, **49**, D97–D103.
25. Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., *et al.* (2015) TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.*, **5**, 11432.
26. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, J.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
27. Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J.R. and O'Sullivan, C. (2022) The sequence read Archive: a decade more of explosive growth. *Nucleic Acids Res.*, **50**, D387–D390.
28. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., *et al.* (2020) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.*, **48**, D882–D889.
29. Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., Snow, C., Fonseca, N.A., Petryszak, R., Papatheodorou, I., *et al.* (2019) ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res.*, **47**, D711–D715.
30. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.
31. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. and Satija, R. (2021) Single-cell chromatin state analysis with Signac. *Nat. Methods*, **18**, 1333–1341.
32. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., *et al.* (2019) Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.*, **20**, 163–172.
33. Wang, C., Sun, D., Huang, X., Wan, C., Li, Z., Han, Y., Qin, Q., Fan, J., Qiu, X., Xie, Y., *et al.* (2020) Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol.*, **21**, 198.
34. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., *et al.* (2020) A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.*, **15**, 2247–2276.
35. Moerman, T., Aibar Santos, S., Bravo Gonzalez-Blas, C., Simm, J., Moreau, Y., Aerts, J. and Aerts, S. (2019) GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, **35**, 2159–2161.
36. Schep, A.N., Wu, B., Buenrostro, J.D. and Greenleaf, W.J. (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods*, **14**, 975–978.
37. 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.
38. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I. and Young, R.A. (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, **153**, 320–334.
39. Hammal, F., de Langen, P., Bergon, A., Lopez, F. and Ballester, B. (2022) ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res.*, **50**, D316–D325.
40. Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
41. Shen, W.K., Chen, S.Y., Gan, Z.Q., Zhang, Y.Z., Yue, T., Chen, M.M., Xue, Y., Hu, H. and Guo, A.Y. (2023) AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. *Nucleic Acids Res.*, **51**, D39–D45.
42. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
43. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
44. Easton, D.F., Ford, D. and Bishop, D.T. (1995) Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.*, **56**, 265–271.
45. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., *et al.* (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, **266**, 66–71.
46. De Siervi, A., De Luca, P., Byun, J.S., Di, L.J., Fufa, T., Haggerty, C.M., Vazquez, E., Moiola, C., Longo, D.L. and Gardner, K. (2010) Transcriptional autoregulation by BRCA1. *Cancer Res.*, **70**, 532–542.
47. Gonzalez, M.E., DuPrie, M.L., Krueger, H., Merajver, S.D., Ventura, A.C., Toy, K.A. and Kleer, C.G. (2011) Histone methyltransferase EZH2 induces Akt-dependent genomic instability and BRCA1 inhibition in breast cancer. *Cancer Res.*, **71**, 2360–2370.
48. Silberg, D.G., Swain, G.P., Suh, E.R. and Traber, P.G. (2000) Cdx1 and Cdx2 expression during intestinal development. *Gastroenterology*, **119**, 961–971.
49. Yu, J., Liu, D., Sun, X., Yang, K., Yao, J., Cheng, C., Wang, C. and Zheng, J. (2019) CDX2 inhibits the proliferation and tumor formation of colon cancer cells by suppressing wnt/ β -catenin signaling via transactivation of GSK-3 β and Axin2 expression. *Cell Death. Dis.*, **10**, 26.
50. Graule, J., Uth, K., Fischer, E., Centeno, I., Galván, J.A., Eichmann, M., Rau, T.T., Langer, R., Dawson, H., Nitsche, U., *et al.* (2018) CDX2 in colorectal cancer is an independent prognostic factor and regulated by promoter methylation and histone deacetylation in tumors of the serrated pathway. *Clinical Epigenetics*, **10**, 120.
51. Hinoi, T., Lucas, P.C., Kuick, R., Hanash, S., Cho, K.R. and Fearon, E.R. (2002) CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. *Gastroenterology*, **123**, 1565–1577.
52. Song, C., Zhang, Y., Huang, H., Wang, Y., Zhao, X., Zhang, G., Yin, M., Feng, C., Wang, Q., Qian, F., *et al.* (2023) Cis-cardio: a comprehensive analysis platform for cardiovascular-relevant cis-regulation in human and mouse. *Mol. Ther. Nucleic Acids*, **33**, 655–667.
53. Oka, T., Maillet, M., Watt, A.J., Schwartz, R.J., Aronow, B.J., Duncan, S.A. and Molkentin, J.D. (2006) Cardiac-specific deletion of Gata4 reveals its requirement for hypertrophy, compensation, and myocyte viability. *Circ. Res.*, **98**, 837–845.
54. Heineke, J., Auger-Messier, M., Xu, J., Oka, T., Sargent, M.A., York, A., Klevitsky, R., Vaikunth, S., Duncan, S.A., Aronow, B.J., *et al.* (2007) Cardiomyocyte GATA4 functions as a stress-responsive regulator of angiogenesis in the murine heart. *J. Clin. Invest.*, **117**, 3198–3210.
55. Anderson, D.J., Kaplan, D.I., Bell, K.M., Koutsis, K., Haynes, J.M., Mills, R.J., Phelan, D.G., Qian, E.L., Leitoguinho, A.R., Arasaratnam, D., *et al.* (2018) NKX2-5 regulates human cardiomyogenesis via a HEY2 dependent transcriptional network. *Nat. Commun.*, **9**, 1373.
56. Cha, J. and Lee, I. (2020) Single-cell network biology for resolving cellular heterogeneity in human diseases. *Exp. Mol. Med.*, **52**, 1798–1808.
57. Pan, L., Shan, S., Tremmel, R., Li, W., Liao, Z., Shi, H., Chen, Q., Zhang, X. and Li, X. (2023) HTCA: a database with an in-depth

- characterization of the single-cell human transcriptome. *Nucleic Acids Res.*, **51**, D1019–D1028.
58. Huang,K., Gong,H., Guan,J., Zhang,L., Hu,C., Zhao,W., Huang,L., Zhang,W., Kim,P. and Zhou,X. (2023) AgeAnno: a knowledgebase of single-cell annotation of aging in human. *Nucleic Acids Res.*, **51**, D805–D815.
 59. Xu,Q., Georgiou,G., Frölich,S., van der Sande,M., Veenstra,G.J.C., Zhou,H. and van Heeringen,S.J. (2021) ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. *Nucleic Acids Res.*, **49**, 7966–7985.
 60. Zhang,P., Zhang,H. and Wu,H. (2022) iPro-WAEL: a comprehensive and robust framework for identifying promoters in multiple species. *Nucleic Acids Res.*, **50**, 10278–10289.
 61. Spitz,F. and Furlong,E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.
 62. Castellanos,M., Mothi,N. and Muñoz,V. (2020) Eukaryotic transcription factors can track and control their target genes using DNA antennas. *Nat. Commun.*, **11**, 540.
 63. Jia,Q., Chen,S., Tan,Y., Li,Y. and Tang,F. (2020) Oncogenic super-enhancer formation in tumorigenesis and its molecular mechanisms. *Exp. Mol. Med.*, **52**, 713–723.
 64. Langfelder,P. and Horvath,S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.*, **9**, 559.
 65. Matsumoto,H., Kiryu,H., Furusawa,C., Ko,M.S.H., Ko,S.B.H., Gouda,N., Hayashi,T. and Nikaido,I. (2017) SCODE: an efficient regulatory network inference algorithm from single-cell RNA-seq during differentiation. *Bioinformatics*, **33**, 2314–2321.
 66. Deshpande,A., Chu,L.F., Stewart,R. and Gitter,A. (2022) Network inference with Granger causality ensembles on single-cell transcriptomics. *Cell Rep.*, **38**, 110333.
 67. Kamimoto,K., Stringa,B., Hoffmann,C.M., Jindal,K., Solnica-Krezel,L. and Morris,S.A. (2023) Dissecting cell identity via network inference and in silico gene perturbation. *Nature*, **614**, 742–751.
 68. Ma,A., Wang,X., Li,J., Wang,C., Xiao,T., Liu,Y., Cheng,H., Wang,J., Li,Y., Chang,Y., *et al.* (2023) Single-cell biological network inference using a heterogeneous graph transformer. *Nat. Commun.*, **14**, 964.
 69. Bravo González-Blas,C., De Winter,S., Hulselmans,G., Hecker,N., Matetovici,I., Christiaens,V., Poovathingal,S., Wouters,J., Aibar,S. and Aerts,S. (2023) SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods.*, **20**, 1355–1367.