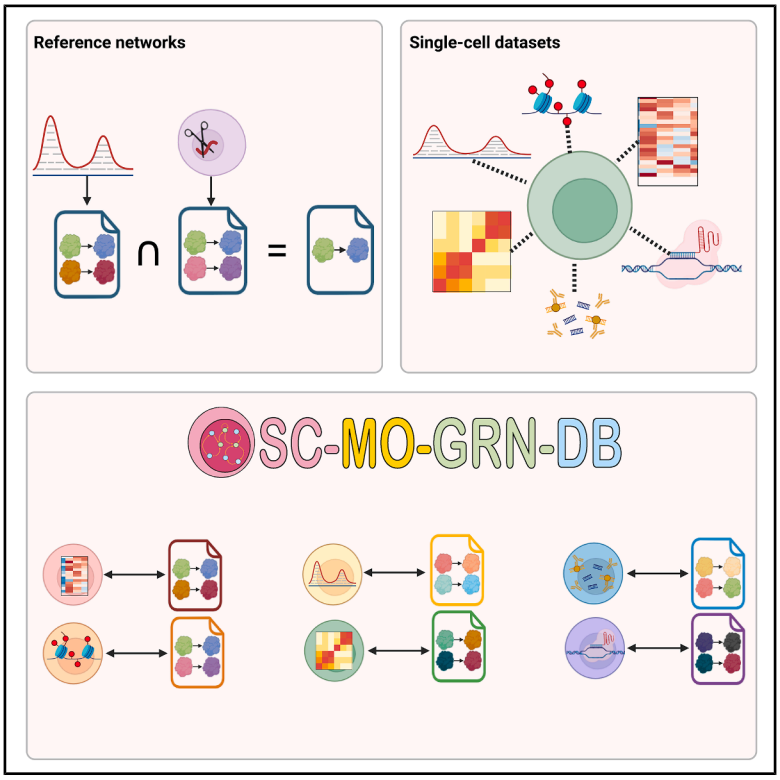
iScience

Article

SC-MO-GRN-DB: A comprehensive repository for single-cell multiomic gene regulatory networks

Graphical abstract



Highlights

- A single-cell multiomics gene regulatory network database for analysis and unbiased benchmark
- Evidence-based reference networks built with TF localization and perturbation data
- Standardized single-cell multiomic datasets paired with reference networks
- Intuitive interface supporting easy access and exploration of data

Authors

Hannah Valensi, Karamveer Karamveer, Eric Moeller, Sami Eren Ozdogan, Rachael M. Edwards, Yasin Uzun

Correspondence

yuzun@pennstatehealth.psu.edu

In brief

Medicine; Biological sciences; Genetics

Valensi et al., 2026, iScience *29*, 115323
April 17, 2026
https://doi.org/10.1016/j.isci.2026.115323

CellPress

iScience

CellPress
OPEN ACCESS

Article

SC-MO-GRN-DB: A comprehensive repository for single-cell multiomic gene regulatory networks

Hannah Valensi,[1] Karamveer Karamveer,[1] Eric Moeller,[1] Sami Eren Ozdogan,[2] Rachael M. Edwards,[3] and Yasin Uzun[1,4,5,*]
[1]Department of Pediatrics, Pennsylvania State University College of Medicine, Hershey, PA, USA
[2]Department of Computer Science and Engineering, Pennsylvania State University, State College, PA, USA
[3]Department of Neuroscience and Experimental Therapeutics, Pennsylvania State University College of Medicine, Hershey, PA, USA
[4]Department of Molecular and Precision Medicine, Pennsylvania State University College of Medicine, Hershey, PA, USA
[5]Lead contact
*Correspondence: yuzun@pennstatehealth.psu.edu
https://doi.org/10.1016/j.isci.2026.115323

SUMMARY

Gene regulatory networks (GRNs) are essential models for understanding gene expression, cell differentiation, and cellular function. Existing GRN resources primarily focus on transcriptomic data and often overlook the epigenetic mechanisms of gene regulation, limiting evaluation and improvement of GRN inference methods. To address this challenge, we developed SC-MO-GRN-DB, a publicly accessible database of experimentally validated GRNs alongside tissue-matched single-cell multiomic datasets across human and mouse tissues. This repository includes ground-truth GRNs comprising over 22 million regulatory edges, curated from high-confidence experimental datasets. In addition, it hosts multiomic single-cell datasets totaling more than two million cells across six molecular modalities: single-cell RNA sequencing (scRNA-seq), chromatin accessibility (scATAC-seq), chromatin immunoprecipitation (scChIP-seq), DNA methylation (scDNA-Met), chromatin conformation (scHi-C), and gene perturbation screens (scCRISPR-seq). By offering standardized input datasets paired with experimentally supported ground-truth networks, SC-MO-GRN-DB provides a platform for the development, benchmarking, and validation of GRNs with single-cell multiomic data.

INTRODUCTION

The regulation of gene expression is a fundamental process that determines how cells acquire their identity, respond to environmental cues, and maintain their functions. Central to this regulation are transcription factors (TFs), which recognize and bind to specific DNA sequences to activate or repress their target genes (TGs).[1] Rather than acting in isolation, TFs operate within interconnected gene regulatory networks (GRNs) that capture the dependencies among regulators and their downstream targets.[2] These networks provide a framework for understanding how genetic programs are established,[3] how cellular states transition during development or disease,[2,4] and how dysregulation of TF activity contributes to pathological processes.[1]

Over the past two decades, a diverse range of strategies for inferring the structure and function of GRNs have been tested.[5–10] Prior to the high-throughput sequencing (HTS) era, attempts to reconstruct GRNs relied on experimentally validated TF-TG interactions compiled from the literature, ChIP-chip, or small-scale molecular assays with limited size and scope.[11–14] The advent of HTS technologies enabled large-scale inference of GRNs using data types such as RNA-seq, ChIP-seq, and related assays, providing genome-wide views of transcriptional regulation.[15] Single-cell RNA-seq and multiomic assays further facilitated the development of algorithms and resources that capture regulatory programs with elevated power and resolution.[7–10]

A variety of GRN-focused databases have since been developed to aggregate TF-gene interactions, providing benchmarks and reference points for the field.[14,16–19] While these resources have been instrumental in advancing systems-level studies, most rely on bulk data or focus narrowly on transcriptomic-only data. Although bulk profiling captures genome-wide activity patterns, it reflects population averages across millions of cells. This masks cell type-specific regulation and obscures heterogeneity in the cell population.[20,21] Consequently, bulk-based networks are limited in representing the full spectrum of regulatory mechanisms that shape gene expression, including epigenetic modifications, chromatin accessibility, higher-order chromatin organization, and direct perturbation effects.

The introduction of single-cell technologies has provided unprecedented resolution for characterizing TF activity and network structure at the level of individual cells.[22] Both single-cell transcriptomics and epigenomics have been instrumental in enabling the construction of GRNs that better capture dynamic regulatory programs and the molecular diversity of tissues.[23–25] These advances have driven the development of computational approaches that leverage single-cell omics to infer more precise and context-dependent regulatory networks.[26] Although recent efforts have incorporated single-cell transcriptomic data into GRN resources, a collection of multiple single-cell modalities with experimentally supported ground-truth networks is lacking. This gap constrains both our

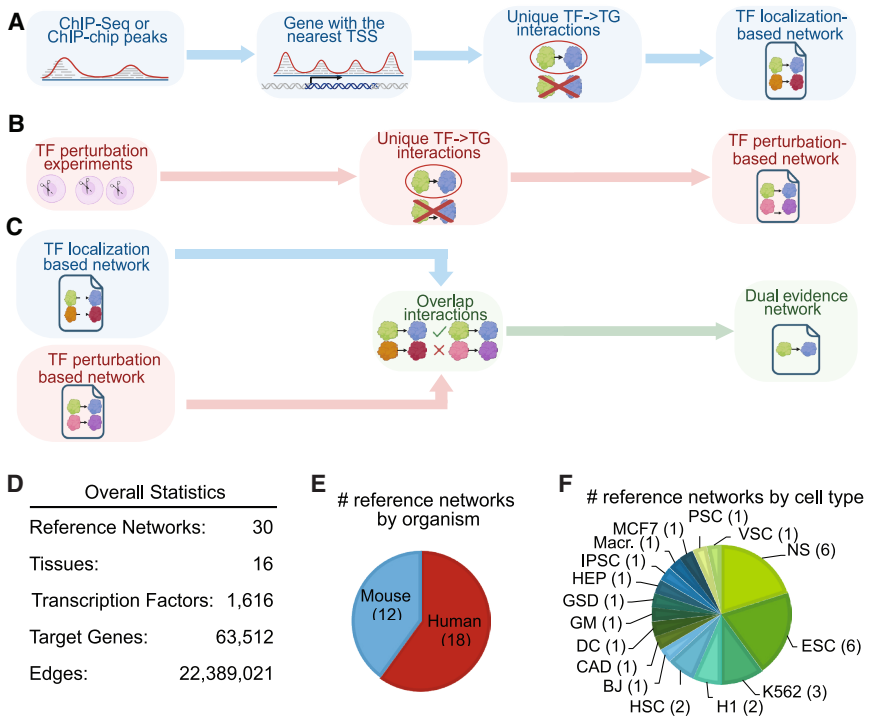


Figure 1. Reference network curation process and summary statistics

(A) TF localization network construction flow chart.
(B) TF perturbation network construction flow chart.
(C) Dual evidence network construction flow chart.
(D) Network overall statistics table.
(E) Number of networks per organism pie chart.
(F) Number of networks per cell type pie chart.

mechanistic understanding of transcriptional regulation and the evaluation of new computational methods for GRN inference.

Evaluating and comparing GRN inference methods remains a major challenge. Each computational study typically constructs its own benchmarking datasets from disparate sources, requiring extensive preprocessing to match biological context or experimental modality. As each individual tool employs a different set of criteria for selecting or filtering reference networks, results are often incomparable across studies. This fragmentation slows down methodological progress, as each group must rebuild their evaluation framework from scratch rather than relying on standardized benchmarking. Although previous efforts[8,27] have provided useful reference datasets, they require substantial manual curation and are limited in their coverage of single-cell and multiomic modalities. The field has therefore lacked a unified, comprehensive, and ready-to-use resource that pairs high-confidence ground-truth networks with the corresponding single-cell data necessary for rigorous benchmarking.

To overcome these challenges, we developed SC-MO-GRN-DB (https://scmogrndb.psu.edu), a comprehensive database that unifies experimentally validated reference GRNs with harmonized single-cell multiomic datasets. This resource catalogs more than 22 million high-confidence regulatory edges derived from experimental evidence and contains over 2 million cells across six molecular single-cell modalities: scRNA-seq, scATAC-seq, scChIP-seq, scDNA methylation, scHi-C, and scCRISPR-based perturbation screens. All datasets have been systematically processed and standardized to facilitate cross-study and cross-modality comparisons. In addition to ground truth GRNs, the database provides a curated collection of single-cell multiomic data that supports reproducible analyses and benchmarking of computational methods. By bridging transcriptomic and epigenomic layers of regulation, SC-MO-GRN-DB establishes a unique foundation for studying GRNs in diverse cellular contexts and for accelerating discoveries in systems biology, disease mechanisms, and therapeutic target identification.

RESULTS

Current GRN research is built upon two fundamental components: ground truth (reference) networks and single-cell datasets. Ground truth networks curated through various types of genetic data serve as a reference for evaluating inferred GRNs. Depending on the type of evidence, these networks can either be tissue- or organism-specific. Single-cell sequencing datasets are used as input data for methods that computationally infer GRNs. We curated both types of data to provide a foundation for GRN research, especially in the context of single-cell genomics.

Reference networks

To establish reliable benchmarks for GRN inference, we curated high-confidence, experimentally validated TF-TG interaction networks. For curating tissue and cell type-specific networks, we employed complementary strategies. One approach that we utilized is to employ TF localization data from ChIP-Seq and ChIP-chip experiments to identify physical binding events (Figure 1A). Widely used for evaluating inferred networks,[28–30] this approach assumes that TFs bind to the DNA regions that are in proximity to their TGs to initiate transcription.

Although ChIP-Seq and ChIP-chip data provide TF localization information, it does not guarantee a regulatory interaction, leading to potential false positives. As an alternative, TF perturbation assays coupled with gene expression profiling provide functional evidence of gene regulation. These assays reveal how altering TF activity via genetic manipulation influences downstream gene expression, and therefore, provide causal evidence of regulation. More specifically, perturbation-based data from genetic knockdowns, knockouts, and overexpression assays capture the functional consequences of changes in TF activity on downstream TG expression. For the genetic perturbation experiment of a specific TF, the genes showing significant differential expression in the readout as a response to the perturbation event are considered as its targets. We also curated reference

networks using TF perturbation data obtained from available resources (Figure 1B). Although generating such data at scale remains technically challenging, CRISPR-based technologies streamline the process, expanding the feasibility of systematically capturing TF-gene relationships across many conditions.

To reduce potential false positives, we also built dual-evidence networks by intersecting TF localization- and perturbation-based networks for the same cell type (Figure 1C). These networks require both types of evidence (localization and perturbation) to hold at the same time for each individual TF-TG interaction, increasing the confidence. However, this approach decreases network size, as for a TF to be included for a specific cell type, experimental data need to be available for both assays. It should be noted that this reduced coverage, while enhancing confidence, results in sparser TF regulons, which may influence benchmarking and downstream analyses (Figure S1). This design reflects the broader tradeoff between confidence and comprehensiveness. Dual-evidence networks prioritize high-confidence interactions, whereas single-evidence networks offer more extensive coverage, and the choice between them may depend on the goals of a given analysis.

In addition to experimentally derived networks, SC-MO-GRN-DB includes smaller sets of literature-curated, text-mined, and protein-protein interaction (PPI)-based networks to broaden biological and contextual coverage. Literature-curated networks are assembled manually by expert curators who extract TF-TG relationships from published studies. These networks often capture high-quality interactions with direct experimental evidence but can be limited by human bias and incomplete coverage, as not all regulatory relationships are reported or curated consistently. PPI-based networks leverage protein-protein interaction data[31,32] to infer potential co-regulatory relationships among TFs and other transcriptional regulators. Although PPI-based networks often lack tissue specificity, they expand SC-MO-GRN-DB's coverage, allowing both detailed and exploratory analyses across diverse biological conditions.

Automated text mining has also been used as an approach to derive gene networks from the published literature.[33] These methods extract TF-to-TG associations by processing vast amounts of scientific literature and can be useful for hypothesis generation and for supplementing areas with limited experimental data. Although text mining-based networks can include indirect associations, they can be useful for some cell types where tissue-specific experimental data are not available and are used for studying genetic interactions and evaluating inference models.[8] Therefore, they are included in SC-MO-GRN-DB, in addition to the cell type-specific reference networks supported with experimental data.[31]

Curated with the described approaches, SC-MO-GRN-DB hosts 30 curated reference networks spanning 16 tissues across human and mouse in total. These networks contain 1,616 TFs, 63,512 TGs, and 22,389,021 regulatory edges (Figure 1D). Networks are distributed across 18 human and 12 mouse tissues (1E) encompassing varied biological contexts including embryonic stem cells, hematopoietic stem cells (HSCs), cortical area development (CAD), dendritic cells (DCs), gonadal sex determination (GSD), induced pluripotent stem cells (IPSCs), blood macrophages (Macr.), pluripotent stem cells (PSCs), ventral spinal cord (VSC), and cell lines BJ, GM12878, HepG2, K562, H1, and MCF7 (Figure 1F).

The sizes of curated reference networks vary markedly, reflecting differences in experimental scope, evidence depth, and biological context (Figures 2 and S2). The literature-based networks include the lowest numbers of TFs as regulators, limiting their usability for comprehensive studies whereas general, non-specific networks have greatest TFs, on the opposite end of the spectrum (Figure 2A). Tissue-specific networks that are based on experimental evidence span the space between these two groups, where the number of TFs tends to be higher for heavily studied cell types and cell lines such as ESC, HSC, and K562.

In terms of overall regulon size (overall TG set size; Table S1), literature-based networks are also limited (Figure 2B). Experimentally derived networks are larger in size when compared to others (Figure 2B). As these networks are based on genome-wide experiments, they cover a large set of targets and have larger sizes than PPI and text mining-based networks. The total number of edges (TF-to-TG interactions) increase linearly with target set sizes (Figure 2C). A similar trend is seen for the regulons of individual TFs. Literature-based and text mining-based networks are limited in size compared to experimentally derived reference networks with large regulons (Figure 2D).

SC-MO-GRN-DB enables researchers to explore reference networks directly through an intuitive web interface. The browsing functionality allows researchers to navigate by organism, tissue, and evidence type, presenting summaries of the size of the networks via TF, TG, and edge counts (Figure 3A). This hierarchical view facilitates comparison across datasets and quick assessment of network scope and quality. Complementing this, the search function enables targeted exploration of specific TFs, TGs, or TF-TG pairs across all networks (Figure 3B). Search results display the supporting evidence types dynamically, allowing the investigators to trace how each regulatory relationship is supported by localization, perturbation, dual evidence, or literature-based sources. These interactive features streamline discovery and cross-validation of transcriptional regulatory information within and across biological systems.

Single-cell datasets

As the complementary component to reference networks, SC-MO-GRN-DB hosts curated single-cell datasets. Each single-cell dataset is paired with a corresponding tissue-type-specific reference network, enabling direct benchmarking of inferred GRNs against experimentally supported ground-truth interactions for the same biological context. The quality and composition of the single-cell data have a strong effect on the accuracy of inferred GRNs. Large-scale single-cell datasets provide elevated power to detect rare cell states and subtle regulatory relationships; however, excessive cell numbers without appropriate curation can introduce technical noise and increase computational burden. Datasets that are smaller in size can fail to detect rare cell states and miss subtle regulatory relationships, but can still capture core regulatory programs and high-frequency interactions. The size of the datasets in SC-MO-GRN-DB is curated to ensure both sufficient cellular coverage and biological relevance.

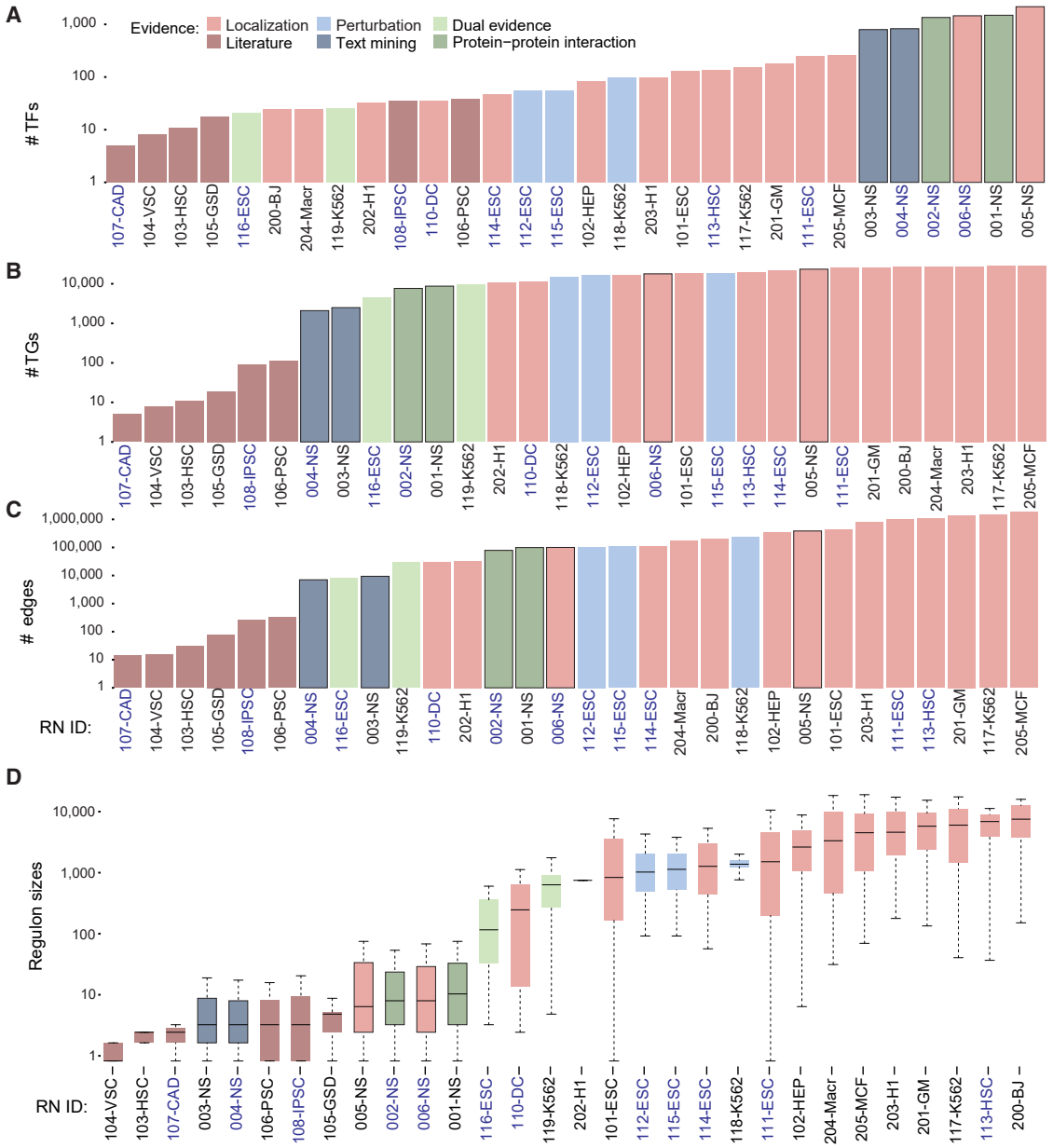


Figure 2. Reference network sizes

Bars and boxes are colored with respect to the evidence type. The *y* axes are displayed in logarithmic scale for visibility purposes. Solid border lines indicate non-specific networks. Networks derived from TF binding localization data and protein-protein interaction evidence tend to exhibit higher TF, TG, and edge counts, reflecting broader regulatory coverage, whereas perturbation-based and literature-curated networks are generally sparser and more selective. Labels for mouse datasets are colored with blue and human datasets are black. Middle lines show median values for each group. Box edges represent 25% and 75% quartiles. Whiskers extend 1.5 times the inter-quartile range from the individual box edges within the limits of data points.

(A) Total number of TFs for each reference network.

(B) Overall TG set sizes for each reference network (each TG counted only once, regardless of the regulator).

(C) Total number of TF-to-TG edges for each reference network.

(D) Distribution of individual regulon sizes (TG counts) for individual TFs for each reference network.

In this context, we curated single-cell multiomic datasets spanning six modalities: scRNA-seq (single-cell RNA sequencing), scATAC-seq (single-cell assay for transposase-accessible chromatin using sequencing), scChIP-seq (single-cell chromatin immunoprecipitation sequencing), scDNA methylation (single-cell DNA methylation profiling), scHi-C (single-cell high-throughput

A

Evidence Type: All | Organism: All | Cell Type: All

Select	Reference Network ID	Evidence Type	Organism	Cell Type	Number of TFs	Number of TGs	Number of Edges	PMID	Download
☐	RN001	PPI	Human	Nonspecific	1489	8712	99149	31907445	Download
☐	RN002	PPI	Mouse	Nonspecific	1350	7636	78567	31907445	Download
☐	RN003	TextMining	Human	Nonspecific	795	2492	9384	29087512	Download

B

Transcription Factor (TF): RUNX1 | Target Gene (TG): e.g., BRCA1 | Search

Select	RN ID	Evidence Type	Organism	Tissue	TF Name	TGs per TF	PMID	Download
☐	RN001	PPI	Human	Nonspecific	RUNX1	161	31907445	Download
☐	RN002	PPI	Mouse	Nonspecific	RUNX1	131	31907445	Download
☐	RN003	TextMining	Human	Nonspecific	RUNX1	40	29087512	Download
☐	RN004	TextMining	Mouse	Nonspecific	RUNX1	17	29087512	Download

Figure 3. Reference network investigation functionalities

(A) Reference network browsing functionality.

(B) Reference network search functionality.

chromosome conformation capture), and scCRISPR-seq (single-cell CRISPR-based perturbation sequencing) (Figure 4A). Particular emphasis was placed on joint assays that simultaneously measure gene expression and chromatin accessibility, especially scRNA-seq with scATAC-seq, providing complementary layers of information critical for accurate GRN reconstruction. For each dataset, only wild-type or control cells were retained to ensure that the data reflected baseline cellular states. When datasets contained multiple annotated cell types, cells were separated into individual subsets so that each dataset corresponded to a single, untreated, unmodified cell type. This curation strategy ensures that single-cell data are optimally matched to the corresponding reference networks, which are likewise cell type-specific and derived from unperturbed contexts.

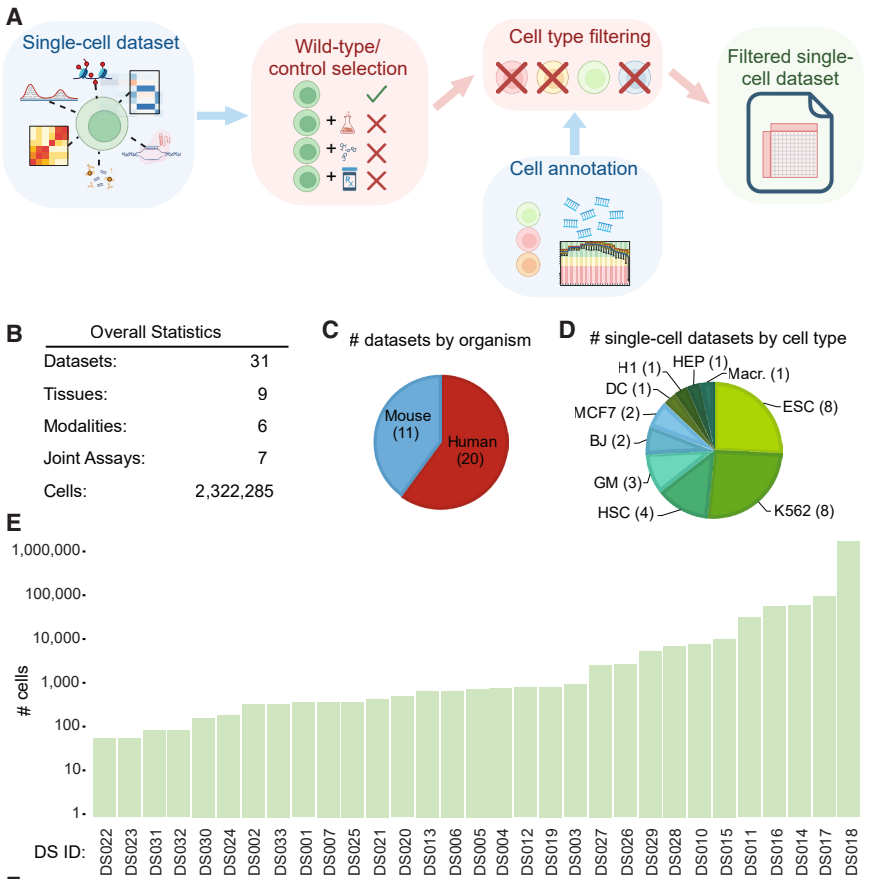
Our repository contains 31 harmonized single-cell datasets covering nine tissues for human and mouse, totaling 2,322,285 cells, and spans six modalities with seven joint assays (Figure 4B). Datasets cover both human and mouse tissues (Figure 4C) and cell type representation includes ESC (n = 8), K562 (n = 8), HSC (n = 4), GM12878 (n = 3), BJ (n = 2), MCF7 (n = 2), and single datasets from DC, H1, HepG2, and macrophages (Figure 4D).

The number of cells per dataset in SC-MO-GRN-DB varies widely, reflecting differences in experimental design, sequencing depth, and biological system complexity (Figure 4E). Dataset sizes range from as few as 63 cells in smaller targeted studies to nearly two million cells in larger-scale studies. This extensive range ensures that the repository includes both high-resolution, small-scale datasets ideal for mechanistic validation and large-scale datasets suited for benchmarking scalability and statistical robustness. Variation in dataset size also enables researchers to assess how GRN inference methods perform across datasets with differing sparsity, depth, and biological diversity.

Similar to the case of reference networks, SC-MO-GRN-DB provides multiple ways to explore these curated datasets. The individual datasets can be browsed by molecular modality, organism, and cell type, and or multiple datasets can be selected as a batch through the interface (Figure 4F). Each dataset is summarized with its total number of cells, cell type, and source study. This functionality enables researchers to quickly identify datasets relevant to their study design, whether they seek large-scale multiomic assays or more narrowly defined, cell type-specific datasets.

DISCUSSION

Benchmarking GRN inference methods is challenging because computational methods often train or validate models using



Select	Dataset ID	Data Modality	Organism	Cell Type	Number of Cells	PMID or DOI	Download
☐	DS001	scRNA	Mouse	ESC	422	31907445	Download
☐	DS002	scRNA	Mouse	DC	384	31907445	Download
☐	DS003	scRNA	Mouse	HSC	1072	31907445	Download

Figure 4. Single-cell dataset curation process and summary statistics

(A) Single-cell datasets with different modalities (scRNA-Seq, scATAC-Seq, scChIP-Seq, scHiC, and scBS-Seq) matching with cell types of the reference networks are curated from the literature. When there are multiple cell types in the data, the relevant cell types are filtered using cell type annotations associated with the dataset. Cleaned, filtered single-cell datasets are deposited in the database.

(B) Overall statistics for the single-cell datasets in the repository.

(C) Number of single-cell datasets by organism.

(D) Number of single-cell datasets for each individual cell type. ESC, embryonic stem cells; Hep, HepG2 human hepatocellular carcinoma cell line; HSC, hematopoietic stem cells; DC, dendritic cells; K562, human erythroleukemia cell line; GM, GM12878 human B-lymphoblastoid cell line; BJ, human fibroblast cell line; H1, human embryonic stem cell line; Macr, macrophage; MCF7, human breast cancer cell line.

(E) Number of cells per single-cell dataset.

(F) Browsing interface for single-cell datasets using data modality (such as scRNA-Seq and scATAC-Seq), organism (human/mouse) and cell type (such as embryonic stem cells, HSCs, and macrophage).

ground-truth networks derived from heterogeneous evidence types.[26] While this variability complicates direct comparisons, SC-MO-GRN-DB mitigates this issue by providing paired reference networks and curated input datasets for each cell type and modality, which facilitates fair and standardized benchmarking. Furthermore, most current GRN inference tools rely primarily on ChIP-seq or other localization-based ground-truth networks.[34] By including TF knock-down/knockout perturbation networks in addition to localization and perturbation networks, SC-MO-GRN-DB presents alternative approaches for ground truth selection.

Ground-truth networks differ not only in scope but also in the type of evidence supporting each TF-TG interaction. Traditional GRN resources rely primarily on TF localization assays such as ChIP-seq, which identify physical DNA-binding events but cannot confirm their regulatory effect. SC-MO-GRN-DB expands beyond this paradigm by including perturbation-based networks. These networks are derived from knockdown, knockout, and CRISPR interference studies, which reveal the functional consequences of TF perturbation. We also provide dual-evidence networks, which intersect localization and perturbation data to provide high-confidence interactions. We include text-mined, literature-curated, and PPI-based networks to contribute breadth and contextual depth. Literature-curated networks deliver precision through expert annotation but remain limited by scope. Text-mined networks offer a large spanning, robust network size, but require careful interpretation to avoid inaccurate associations. PPI-based networks capture regulatory cofactors and signaling interactions that shape TF activity but may not directly indicate transcriptional regulation. Collectively, these complementary evidence types make SC-MO-GRN-DB a flexible benchmarking resource adaptable to both high-confidence and exploratory GRN studies.

Together with these curated reference networks, the accompanying single-cell datasets provide the essential foundation for benchmarking GRN inference methods across modalities and cell types. The variety of tissues across both organisms allows systematic evaluation of model generalizability, while the inclusion of multiomic and joint assays enables integrative approaches that more faithfully capture the multilayered nature of transcriptional regulation. The scale of over two million cells supports robust statistical inference and motivates the use of advanced computational methods that can fully leverage the richness of the data. By providing complementary, cell type-specific datasets aligned with curated reference networks, SC-MO-

GRN-DB supports the construction and benchmarking of GRN inference tools.

SC-MO-GRN-DB is therefore designed to support flexible mapping between single-cell datasets and reference networks derived from independent sources. For a given tissue or cell type, the database may include multiple single-cell datasets and multiple reference networks, reflecting differences in experimental conditions, molecular modalities, and evidence types. These resources are not intended to be paired in a one-to-one manner. Instead, compatible datasets and reference networks can be combined to support GRN construction and evaluation. Our recommendation is to leverage all available datasets rather than relying on a single pairing, as this enables more robust and reproducible analyses. For benchmarking and method evaluation, an all-to-all strategy, testing each single-cell dataset against each reference network, allows for the assessment of consistency and performance across data sources. For biological discovery, results may be combined across datasets and reference networks to construct consensus GRNs that capture shared regulatory signals. This design provides multiple, independent lines of evidence to support GRN construction, validation, and comparative analysis.

In parallel with the rapid expansion of single-cell datasets, a wide range of machine learning and deep learning approaches have been developed to model regulatory and disease-associated processes in genomic data.[35–37] These methods span diverse applications, including regulatory interaction prediction, GRN inference, cell clustering and annotation, and disease association modeling, often leveraging attention mechanisms, graph-based representations, or deep autoencoder frameworks.[38–40] While such approaches have demonstrated strong performance within specific tasks and datasets, they are typically evaluated using custom data processing pipelines and heterogeneous reference networks, limiting reproducibility and cross-method comparison. By providing standardized, experimentally supported reference networks paired with tissue-matched single-cell multiomic datasets, SC-MO-GRN-DB offers a unifying resource to support the systematic evaluation and future development of these computational approaches.

Beyond evidence type, the scale of both networks and datasets strongly influences their interpretive and benchmarking value. By encompassing over 1,000 TFs and tens of thousands of TGs, SC-MO-GRN-DB captures broad regulatory landscapes suitable for large-scale method evaluation. The database includes networks ranging from small, focused systems with only a few regulators and targets to extensive, high-density networks containing nearly 1,000 TFs, nearly 30,000 TGs, and more than a million edges. This variability reflects the underlying experimental scope and biological diversity of the source data. Network properties such as the number of TFs, TGs, and edges determine how much regulatory complexity is represented, while dataset properties, particularly cell count and gene coverage, define the resolution at which those networks can be inferred. Larger networks and datasets offer statistical robustness and broader coverage but may introduce noise and redundancy. Smaller or more focused systems yield clearer signals but risk underrepresenting biological diversity. SC-MO-GRN-DB balances these trade-offs by providing both diverse and cell type-specific networks. By providing a range of network sizes, SC-MO-GRN-DB enables users to tailor analyses for either comprehensive exploration or mechanistic precision.

The curated pairing of networks and single-cell datasets is central to SC-MO-GRN-DB's design. Each dataset has been matched to an experimentally validated reference network representing the same tissue, cell type, and biological condition. This alignment ensures that benchmarking remains interpretable, as both the reference network and the corresponding dataset describe the same regulatory context. Such systematic curation reduces ambiguity in evaluation and allows for fair comparison between methods, even across distinct molecular modalities.

Unlike repositories that rely heavily on large-scale text-mining or automated inference,[8,14,19] SC-MO-GRN-DB emphasizes curated, experimentally supported content (Table S2). Each network and dataset has been[34] standardized and annotated to facilitate reproducibility and meaningful comparison. This design principle ensures that the database delivers resources with a strong focus on accuracy and biological relevance, directly supporting the construction and evaluation of GRN inference tools.

A defining strength of SC-MO-GRN-DB is its coverage of molecular modalities encompassed within the single-cell datasets. While many current GRN inference tools have focused primarily on transcriptomics or chromatin accessibility as input for GRN inference,[41] SC-MO-GRN-DB incorporates scRNA-seq, scATAC-seq, scChIP-seq, scDNA methylation, scHi-C, and scCRISPR-seq datasets. This diversity enables researchers to investigate regulatory mechanisms that extend beyond gene expression alone, such as TF binding, epigenetic modifications, three-dimensional chromatin organization, and causal perturbation effects.[42] Joint assays, such as scRNA-seq with scATAC-seq, provide complementary perspectives on gene regulation by simultaneously measuring transcriptional output and chromatin accessibility. Similarly, scHi-C captures long-range regulatory interactions, while scCRISPR-seq offers direct perturbation-based evidence of causality. Together, these multiomic layers create new opportunities for tool development that more faithfully capture the complexity of transcriptional regulation. By providing such a comprehensive resource, SC-MO-GRN-DB positions itself as a foundation for next-generation GRN inference methods that integrate across molecular modalities.

By pairing high-confidence reference networks with curated multiomic datasets, SC-MO-GRN-DB serves not only as a repository of curated data but also as a practical platform for advancing both the benchmarking and construction of GRNs in single-cell multiomic contexts. This resource reduces the burden on researchers to locate and standardize datasets, which will help to accelerate the development of more accurate multimodal GRN inference tools. This comprehensive framework supports the exploration of transcriptional regulation across multiple molecular layers, from gene expression and chromatin accessibility to epigenetic modifications, three-dimensional genome architecture, and perturbation-driven causality. As a result, SC-MO-GRN-DB enables the systematic construction and evaluation of context-specific GRNs, empowering researchers to generate insights into cellular regulatory programs, disease mechanisms, and potential

therapeutic targets in a reproducible and biologically meaningful manner.

SC-MO-GRN-DB is curated from currently available public data sources, including reference networks and single-cell datasets, with a specific emphasis on multiomics. The single-cell genomics field is growing rapidly, and new datasets are being shared with the research community continuously. As a living resource, SC-MO-GRN-DB is designed for continued expansion as new tissues, modalities, and evidence types become available. In this sense, it serves as a seed repository, a scalable foundation for GRN inference, capable of growing into a central hub for the development, comparison, and validation of computational models of transcriptional regulation, similar to the UCI repository for machine learning.[43]

Limitations of the study

The present release includes data from 16 tissue types and is biased toward well-characterized cell lines and abundant cell populations, while rare cell types and disease-associated states remain underrepresented. This limitation reflects the availability of high-confidence reference networks and matched single-cell multiomic datasets rather than constraints of the database framework. In particular, the construction of ground truth reference networks requires experimentally validated regulatory interactions, and for many TFs, such evidence is only available through dedicated perturbation or ChIP-based experiments. As a result, ground truth coverage remains incomplete and uneven across TFs and cell types. To address this limitation, we plan to continuously expand the database by incorporating new ground truth networks and single-cell datasets. A key focus of future development will be the inclusion of more truly paired multi-omic datasets.

While integration methods exist for unpaired modalities, such approaches can introduce uncertainty and batch effects, potentially reducing the reliability of downstream GRN inference. Therefore, prioritizing true multiomics (joint assay) datasets will provide better data alignment, reduce preprocessing burden for users, and enhance the robustness of multi-modal analyses. Overall, the database is designed as a living resource, with ongoing updates aimed at improving data coverage, methodological reliability, and support for increasingly complex GRN inference workflows.

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to and will be fulfilled by the lead contact, Yasin Uzun (yuzun@pennstatehealth.psu.edu).

Materials availability

This study did not generate new reagents.

Data and code availability

- SC-MO-GRN-DB is publicly accessible at https://scmogrndb.psu.edu.
- The scripts that are used to build SC-MO-GRN-DB are deposited into and publicly accessible at: https://github.com/UzunLab/SC-MO-GRN-DB.

ACKNOWLEDGMENTS

The research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM150616.

AUTHOR CONTRIBUTIONS

Conceptualization was performed by Y.U.; data curation was carried out by H.V., Y.U., and R.E.; formal analysis, project administration, visualization, and the majority of the investigation were conducted by H.V.; methodology was developed by H.V. with input from Y.U.; software development was led by H.V. with technical support from S.E.O.; validation and quality control were performed by H.V., K.K., and E.M.; resources were provided by Y.U.; supervision was provided by Y.U.; writing of the original draft was performed by H.V.; and writing-review and editing were contributed by H.V., K.K., E.M., and Y.U.

DECLARATION OF INTERESTS

Y.U. is the owner of Systems Biology Consulting & Analytics LLC (Systems Bio). However, Systems Bio did not provide any support for this manuscript and has no financial or other interests related to its content.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Construction of reference networks
  - Curation of single-cell datasets
  - Interactive browsing
  - Interactive search
  - Analytical functions
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2026.115323.

Received: November 13, 2025
Revised: January 9, 2026
Accepted: March 6, 2026
Published: March 11, 2026

REFERENCES

1. Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.
2. Macneil, L.T., and Walhout, A.J.M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res. *21*, 645–657.
3. Feigin, C., Li, S., Moreno, J., and Mallarino, R. (2023). The GRN concept as a guide for evolutionary developmental biology. J. Exp. Zool. B Mol. Dev. Evol. *340*, 92–104.
4. Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796–800.
5. Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. Annu. Rev. Genomics Hum. Genet. *2*, 343–372.
6. Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., et al. (2002). A genomic regulatory network for development. Science *295*, 1669–1678.

7. 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.
8. Pratapa, A., Jalihal, 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.
9. 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.
10. Yuan, Q., and Duren, Z. (2025). Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. Nat. Biotechnol. *43*, 247–257.
11. 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.
12. 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.
13. Liu, Z.-P., Wu, C., Miao, H., and Wu, H. (2015). RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. Database *2015*, bav095.
14. 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.
15. Angelini, C., and Costa, V. (2014). Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. Front. Cell Dev. Biol. *2*, 51.
16. 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. Genom. Proteom. Bioinform. *18*, 120–128.
17. Ben Guebila, M., Lopes-Ramos, C.M., Weighill, D., Sonawane, A., Burkholz, R., Shamsaei, B., Platig, J., Glass, K., Kuijjer, M., and Quackenbush, J. (2022). GRAND: a database of gene regulatory network models across human conditions. Nucleic Acids Res. *50*, D610–D621.
18. 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.
19. Huang, X., Song, C., Zhang, G., Li, Y., Zhao, Y., Zhang, Q., Zhang, Y., Fan, S., Zhao, J., Xie, L., and Li, C. (2024). scGRN: a comprehensive single-cell gene regulatory network platform of human and mouse. Nucleic Acids Res. *52*, D293–D303.
20. Fiers, M.W.E.J., Minnoye, L., Aibar, S., Bravo González-Blas, C., Kalender Atak, Z., and Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. Brief. Funct. Genomics *17*, 246–254.
21. Cha, J., and Lee, I. (2025). Single-cell network biology enabling cell-type-resolved disease genetics. Genomics Inform. *23*, 10.
22. Tabula Sapiens Consortium, Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896.
23. Chen, S., Lake, B.B., and Zhang, K. (2019). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat. Biotechnol. *37*, 1452–1457.
24. Liu, L., Liu, C., Quintero, A., Wu, L., Yuan, Y., Wang, M., Cheng, M., Leng, L., Xu, L., Dong, G., et al. (2019). Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. Nat. Commun. *10*, 470.
25. Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103–1116.e20.
26. Karamveer & Uzun, Y., and Uzun, Y. (2024). Approaches for Benchmarking Single-Cell Gene Regulatory Network Methods. Bioinform. Biol. Insights *18*, 11779322241287120.
27. Xu, H., Baroukh, C., Dannenfelser, R., Chen, E.Y., Tan, C.M., Kou, Y., Kim, Y.E., Lemischka, I.R., and Ma'ayan, A. (2013). ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. Database *2013*, bat045.
28. Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., DREAM5 Consortium; Kellis, M., Collins, J.J., and Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. Nat. Methods *9*, 796–804.
29. Sikora-Wohlfeld, W., Ackermann, M., Christodoulou, E.G., Singaravelu, K., and Beyer, A. (2013). Assessing computational methods for transcription factor target gene identification based on ChIP-seq data. PLoS Comput. Biol. *9*, e1003342.
30. Aibar, S., González-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.
31. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646.
32. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. *30*, 187–200.
33. Altman, R.B., Bergman, C.M., Blake, J., Blaschke, C., Cohen, A., Gannon, F., Grivell, L., Hahn, U., Hersh, W., Hirschman, L., et al. (2008). Text mining for biology–the way forward: opinions from leading scientists. Genome Biol. *9*, S7.
34. Loers, J.U., and Vermeirssen, V. (2024). A single-cell multimodal view on gene regulatory network inference from transcriptomics and chromatin accessibility data. Brief. Bioinform. *25*, bbae382.
35. Yuan, L., Zhao, L., Lai, J., Jiang, Y., Zhang, Q., Shen, Z., Zheng, C.H., and Huang, D.S. (2024). iCRBP-LKHA: Large convolutional kernel and hybrid channel-spatial attention for identifying circRNA-RBP interaction sites. PLoS Comput. Biol. *20*, e1012399.
36. Yuan, L., Zhao, J., Shen, Z., Zhang, Q., Geng, Y., Zheng, C.H., and Huang, D.S. (2023). iCircDA-NEAE: Accelerated attribute network embedding and dynamic convolutional autoencoder for circRNA-disease associations prediction. PLoS Comput. Biol. *19*, e1011344.
37. Yuan, L., Sun, S., Zhang, Q., Li, H.T., Shen, Z., Hu, C., Zhao, X., Ye, L., Zheng, C.H., and Huang, D.S. (2024). Identification of ferroptosis-related lncRNAs for predicting prognosis and immunotherapy response in non-small cell lung cancer. Future Gener. Comput. Syst. *159*, 204–220.
38. Yuan, L., Zhao, L., Jiang, Y., Shen, Z., Zhang, Q., Zhang, M., Zheng, C.H., and Huang, D.S. (2024). scMGATGRN: a multiview graph attention network-based method for inferring gene regulatory networks from single-cell transcriptomic data. Brief. Bioinform. *25*, bbae526.
39. Yuan, L., Xu, Z., Meng, B., and Ye, L. (2025). scAMZI: attention-based deep autoencoder with zero-inflated layer for clustering scRNA-seq data. BMC Genom. *26*, 350.
40. Yuan, L., Sun, S., Jiang, Y., Zhang, Q., Ye, L., Zheng, C.H., and Huang, D.S. (2024). scRGCL: a cell type annotation method for single-cell RNA-seq data using residual graph convolutional neural network with contrastive learning. Brief. Bioinform. *26*, bbae662.
41. Badia-I-Mompel, P., Wessels, L., Müller-Dott, S., Trimbour, R., Ramirez Flores, R.O., Argelaguet, R., and Saez-Rodriguez, J. (2023). Gene

regulatory network inference in the era of single-cell multi-omics. Nat. Rev. Genet. *24*, 739–754.

42. Stock, M., Losert, C., Zambon, M., Popp, N., Lubatti, G., Hörmanseder, E., Heinig, M., and Scialdone, A. (2025). Leveraging prior knowledge to infer gene regulatory networks from single-cell RNA-sequencing data. Mol. Syst. Biol. *21*, 214–230.
43. Dua, D., and Graff, C. (2017). UCI Machine Learning. http://archive.ics.uci.edu/ml.
44. Frankish, A., Carbonell-Sala, S., Diekhans, M., Jungreis, I., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Arnan, C., Barnes, I., et al. (2023). GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Res. *51*, D942–D949.
45. Zou, Z., Ohta, T., and Oki, S. (2024). ChIP-Atlas 3.0: a data-mining suite to explore chromosome architecture together with large-scale regulome data. Nucleic Acids Res. *52*, W45–W53.
46. Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.H., Brown, M., Zhang, X., Meyer, C.A., and Liu, X.S. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. *47*, D729–D735.
47. Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L., et al. (2017). Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res. *45*, D658–D662.
48. Raghav, P., Kumar, R., Lathwal, A., and Sharma, N. (2024). Computational Biology for Stem Cell Research (Elsevier).
49. Feng, C., Song, C., Song, S., Zhang, G., Yin, M., Zhang, Y., Qian, F., Wang, Q., Guo, M., and Li, C. (2024). KnockTF 2.0: a comprehensive gene expression profile database with knockdown/knockout of transcription (co-)factors in multiple species. Nucleic Acids Res. *52*, D183–D193.
50. 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.
51. Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., et al. (2007). ArrayExpress–a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. *35*, D747–D750.
52. Peidli, S., Green, T.D., Shen, C., Gross, T., Min, J., Garda, S., Yuan, B., Schumacher, L.J., Taylor-King, J.P., Marks, D.S., et al. (2024). scPerturb: harmonized single-cell perturbation data. Nat. Methods *21*, 531–540.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Constructed reference networks	This paper	https://scmogrndb.psu.edu/
Filtered single cell datasets	This paper	https://scmogrndb.psu.edu/
Software and algorithms		
Code for constructing reference networks from raw ChIP-seq data	This paper	https://github.com/UzunLab/SC-MO-GRN-DB
R version 4.2.3	R Core Team	https://www.R-project.org/
GENCODE v38 human (hg38) mouse (mm39)	Frankish, A. et al.[44]	https://www.gencodegenes.org/
Other		
Website	This paper	https://scmogrndb.psu.edu/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study did not generate new experimental data or involve human participants, animals, cell lines, or primary cell cultures. All analyses were performed using publicly available datasets obtained from previously published studies and community repositories. Therefore, no experimental model or study participant information is applicable.

Sex, age, genotype, and sample allocation were not controlled by the present study and were reported as provided by the original sources. Limitations related to incomplete metadata in original datasets are acknowledged.

METHOD DETAILS

Construction of reference networks

To establish a foundation of high-confidence GRNs, we systematically collected multiple complementary evidence types. Each line of evidence provides unique strengths for capturing TF-TG interactions, and their intersection reduces bias and increases network reliability. TF localization data were derived from ChIP-seq and ChIP-chip experiments, which map genome-wide TF binding sites. Datasets were collected from BEELINE,[8] ChIP-Atlas,[45] Cistrome,[46,47] and ESCAPE,[27,48] as well as from individual publications (Table S3). The binding peaks were assigned to the gene with the nearest transcription start site (TSS) based on the genomic distance, using the "`nearestTSS`" function in edgeR package (v4.0.16) with gene annotations from GENCODE[44] v38 for human (hg38) and mouse (mm39) genomes.

TF perturbation networks were constructed from loss-of-function and gain-of-function experiments, including TF knockdown, knockout, and overexpression assays curated from KnockTF2.0,[49] ESCAPE,[27,48] and additional primary publications (Table S3). The ground-truth networks were standardized into two-column lists of TF-to-TG interactions for consistency across networks. In addition, protein-protein interaction data from STRING[31] (obtained through BEELINE[8]) were incorporated to capture regulatory complexes, recognizing that such interactions represent indirect regulation rather than direct TF-DNA binding. Text-mined networks were obtained from TRRUST,[14] which uses sentence-level text mining, along with individual studies that provided experimentally validated TF-TG relationships. Literature-curated networks were obtained from BEELINE,[8] which selected published Boolean models from four different cell types in both mouse and human. To further increase confidence, we generated dual-evidence networks by intersecting localization- and perturbation-based networks within the same cellular context.

Once networks were standardized, they were filtered using only known TFs[50] for both human and mouse to retain only TF-TG interactions. All reference networks were annotated with metadata, including article identifier, organism, tissue or cell type, evidence type, and summary statistics (number of TFs, TGs, and edges). The full repository of networks and their associated metadata is accessible through the SC-MO-GRN-DB web interface.

Curation of single-cell datasets

To provide high-quality input data for benchmarking and inference of GRNs, we curated single-cell multiomic datasets spanning a wide range of tissues, modalities, and organisms (Table S4). Datasets were obtained from BEELINE,[8] publicly available repositories such as Gene Expression Omnibus (GEO) and ArrayExpress,[51] community resources including scPerturb,[52] and individual publications (Table S5). We retained only datasets that included wild-type or untreated control cells, as perturbed or genetically modified

cells may confound baseline regulatory relationships. When datasets included multiple annotated cell types, cells were stratified into homogenous groups according to provided metadata, ensuring that each dataset corresponded to a single, unmodified cellular state. For each dataset, we excluded cells with poor quality metrics to ensure consistency across studies. All curated datasets were annotated with information about organism, tissue or cell type, molecular modality, number of cells, and primary literature reference (Table S5). The complete collection, including article identifiers and summary statistics, is made available through the SC-MO-GRN-DB web interface.

Interactive browsing

The SC-MO-GRN-DB web interface was designed to facilitate intuitive exploration of reference networks and curated single-cell datasets. Reference networks are indexed by evidence type, organism, and cell type, enabling users to rapidly identify networks of interest based on experimental design or biological context. Each entry is annotated with standardized metadata, including the number of TFs, TG, regulatory edges, and source publication, thereby allowing direct comparison across networks. Similarly, single-cell datasets are indexed by modality, organism, and cell type, with annotations for the number of cells and originating study. Individual entries can be browsed to inspect specific datasets or perform batch selection and download for larger analyses. The browsing interface was optimized for both usability and reproducibility, ensuring that researchers can efficiently identify and obtain the resources most relevant to their work.

Interactive search

To support targeted exploration of regulatory relationships, we implemented a search module that allows queries by TF, TG, or TF-TG pair. Searches by TF return all networks containing the specified TF, along with the number of unique targets it regulates in each context. Searches by TG return all networks in which the gene appears as a target, while searches by TF-TG pair return only networks in which the specified interaction is present. All results are displayed with network-level metadata, including evidence type, organism, cell type, number of edges, and associated article identifier, ensuring that users can assess the context and reliability of each interaction. Search results can be exported individually or as batches, allowing seamless integration into downstream analyses. This functionality supports both hypothesis-driven investigations, for example, querying a candidate TF for its regulatory footprint across tissues, and broader network-level benchmarking tasks. By combining precise search capabilities with standardized metadata, the search module enhances the flexibility and practical utility of SC-MO-GRN-DB for diverse research applications.

Analytical functions

To support quantitative benchmarking and downstream evaluation of GRNs, SC-MO-GRN-DB provides an analysis module that implements standardized comparison, enrichment, and network-level control analyses. For GRN benchmarking, user-uploaded networks are compared against selected SC-MO-GRN-DB reference networks using edge-level overlap. True positives (TP) are defined as inferred TF-TG interactions present in the reference network, false positives (FP) as inferred interactions absent from the reference, and false negatives (FN) as reference interactions not recovered by the inferred network. From these quantities, performance metrics including precision recall, F1 score, Jaccard index, and area under the precision-recall curve (AUC) are computed. Summary statistics, including the total number of user-uploaded edges, reference edges, and TP, FP, and FN counts, are reported for each comparison.

To enable functional interpretation of regulatory relationships, enrichment analysis is performed using a hypergeometric test. Given a TF-specific target gene set, a pathway or gene set of interest, and a background gene set representing all expressed genes, the test evaluates whether TF targets are significantly overrepresented in the pathway relative to the background. Enrichment statistics and associated p-values are reported, allowing hypothesis-driven assessment of TF-pathway associations.

To characterize and compare reference network structure, control analyses based on graph-theoretic measures are implemented. For a given reference network, centrality metrics, including degree centrality, betweenness centrality, closeness centrality, and PageRank, are computed for each TF, providing complementary measures of regulatory influence within the network. In addition, pairwise comparison of reference networks is supported by quantifying overlap between TF target sets. For each TF shared between two networks, the number of common targets and targets unique to each network are calculated, enabling direct comparison of regulatory programs across networks.

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantitative and statistical analyses in this study were performed using the comparison, enrichment, and control functions described in the method details section. These analyses were used to evaluate overlap between networks, assess enrichment of transcription factor target genes in gene sets of interest, and summarize properties of reference networks.

Exact values of *n* (defined as the number of transcription factors, target genes, regulatory edges, or networks, depending on the analysis) are reported in the corresponding tables and figures. Because this study did not involve experimental group comparisons or biological replicates, no sample size estimation or randomization procedures were applicable.