scRegulate: Single-Cell Regulatory-Embedded Variational Inference of Transcription Factor Activity from Gene Expression

Mehrdad Zandigohar[1], Jalees Rehman[1,2], Yang Dai[1,*]

[1]Department of Biomedical Engineering, University of Illinois Chicago, Illinois, United States

[2]Department of Biochemistry and Molecular Genetics, University of Illinois Chicago, Illinois, United States

*To whom correspondence should be addressed.

Abstract

Motivation: Accurately inferring transcription factor (TF) activity from single-cell RNA sequencing (scRNA-seq) data remains a fundamental challenge in computational biology. While existing methods rely on statistical models, motif enrichment, or prior-based inference, they often suffer from deterministic assumptions about regulatory relationships, reliance on static regulatory databases, or lack of interpretability. Moreover, few approaches can effectively integrate prior biological knowledge with data-driven inference to capture novel, dynamic, and context-specific regulatory interactions.

Results: To address these limitations, we develop scRegulate, a generative deep learning framework that leverages variational inference to infer TF activities while incorporating gene regulatory network (GRN) priors. By integrating structured biological constraints with a probabilistic latent space model, scRegulate offers a scalable and biologically interpretable solution for prediction of regulatory interactions from scRNA-seq data. We comprehensively benchmark scRegulate using multiple public experimental and synthetic datasets generated from GRouNdGAN to demonstrate its ability to infer TF activities and GRNs that are consistent with the underlying ground-truth regulatory interactions. scRegulate outperforms existing TF inference methods, achieving AUROC values of 0.71-0.86 and AUPRC values of 0.80-0.95 on three synthetic datasets. Additionally, scRegulate accurately recapitulates experimentally validated TF knockdown effects on a Perturb-seq dataset, achieving a mean $\log_2$ fold change of -0.66 to -16.98 ($p \leq 8.06\times10^{-13}$) for key TFs such as ELK1, EGR1, and CREB1. Applied to the PBMC scRNA-seq data, scRegulate reconstructs cell-type-specific GRNs and identifies differentially active TFs that align with known immune regulatory pathways. Furthermore, we show that scRegulate's TF embeddings capture meaningful transcriptional heterogeneity, enabling accurate clustering of cell types. Collectively, our results establish scRegulate as a powerful, interpretable, and scalable framework for inferring TF activities and regulatory networks from single-cell transcriptomics.

Availability: All datasets and results are available on GitHub at github.com/YDaiLab/scRegulate.

Contact: yangdai@uic.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Transcription factors (TFs) drive cellular identity and function by regulating gene expression programs (Lambert *et al.*, 2018; Voss and Hager, 2014). When TFs bind to specific DNA motifs and recruit regulatory machinery, they modulate the activation or repression of target genes in a context-dependent manner. Their precise activity is essential not only for normal cellular processes but also for survival, development, and response to environmental cues. Their dysregulation can therefore contribute to a variety of diseases and cancers, making the accurate inference of TF activity essential to understanding cell fate decisions and disease mechanisms.

Single-cell RNA sequencing (scRNA-seq) has revolutionized the study of gene regulation by enabling high-resolution profiling of transcriptomes in individual cells. Numerous computational methods exist for inferring TF activity from single-cell transcriptomes (Hecker *et al.*, 2023). decoupleR (Badia-i-Mompel *et al.*, 2022) rapidly estimates TF activity by coupling a statistical method with a prior knowledge network, such as the univariate linear model or an ensemble of different models. While it offers superior speed advantages due to its simplicity, its capacity is limited in capturing complex regulatory relationships. SCENIC integrates known TF-target gene databases and motif enrichment analysis to infer a TF regulon (i.e., target genes of the TF) based on co-expression that may not be necessarily expected between the TF and a target gene (Aibar *et al.*, 20(Aibar *et al.*, 2017). In addition, its dependance on static motif databases may introduce an unknown level of false positive rate in inferred TF regulons in poorly annotated systems. Bayesian factor models like BITFAM (Gao *et al.*, 2021) bypass co-expression and motif-based inference by incorporating prior knowledge of TF binding data retrieved from public TF-ChIP databases. However, the reliance on linear assumptions

may limit their ability to capture the complex, non-linear transcriptional dynamics present in heterogeneous cell populations.

Multiomic approaches, such as SCENIC+ (Bravo González-Blas *et al.*, 2023) and CellOracle (Kamimoto *et al.*, 2023), and BIOTIC (Cao *et al.*, 2025) attempt to enhance inference by integrating chromatin accessibility alongside transcriptomic data. These methods can provide more accurate TF activity prediction. However, they require high-quality enhancer annotations or complementary ATAC-seq chromatin data, which are not always available in single-cell studies. Additionally, their reliance on integrating diverse modalities introduces additional complexity and data harmonization challenges.

GRNs are inherently context-dependent and cell-type-specific, necessitating models that can adapt dynamically to transcriptional profiles of different cell types or cell states. However, the inherent sparsity and noise in scRNA-seq data complicate inference efforts. In light of these challenges, recent developments in generative models, particularly variational autoencoders (VAEs), have shown promise in extracting meaningful representations from high-dimensional scRNA-seq data (Lopez *et al.*, 2020; Gayoso *et al.*, 2022). While models like SCALE (Xiong *et al.*, 2019) and other adaptations of scVI (Svensson *et al.*, 2020) have made the decoder structure more interpretable, they do not inherently incorporate biological knowledge into the inference process.

To address these challenges, we develop scRegulate, a novel framework that explicitly embeds gene regulatory network (GRN) relationships in a VAE architecture. Specifically, scRegulate integrates curated TF-target interactions from existing regulatory datasets into a non-linear generative model. Unlike previous approaches that infer static or bulk-level GRNs, scRegulate dynamically tailors regulatory networks to reflect the unique transcriptional landscapes of different cell populations. This capability provides unprecedented resolution in studying cell-type-specific regulatory mechanisms and their roles in cellular identity and function, allowing for dynamic refinement of TF-target interactions. Through comprehensive benchmarking, we demonstrate that scRegulate outperforms existing methods in accuracy, robustness, and biological relevance. Thus, scRegulate is a powerful and flexible tool for unraveling the complexities of transcriptional regulation at single-cell resolution. The implementation and benchmarking of scRegulate are available on GitHub (see "Availability" section).

2 Materials and Methods

2.1. Overview

scRegulate is a TF activity inference framework that integrates prior knowledge of TF and target genes into a VAE architecture (Figure 1). The model follows a structured three-phase approach: (1) in the prior initialization phase, the weights connecting the TF activity layer to the output layer are initialized using the initial GRN, thereby enforcing known TF-gene regulatory interactions; (2) in the dynamic inference phase, regulatory weights are optimized, with prior constraints gradually relaxed to allow for the discovery of new TF-target interactions; and (3) in the cell-type-specific fine-tuning phase, GRNs are optimized per cell-type to refine TF activity patterns for more accurate representation of cell-type-specific regulation. The output of inferred TF-activities and weighted regulatory interactions can be used for downstream analysis, including identifying differentially active TFs, cell-type-specific TF-gene interactions, and functional enrichment analyses for the target genes. The three core modules in the scRegulate pipeline, i.e., data preprocessing, TF activity inference via VAE, and downstream analysis, are implemented using Scanpy and AnnData, adhering to best practices in scalable single-cell omics analysis.

2.2. Data Preprocessing

The input data consists of a scRNA-seq expression matrix and an optional GRN prior. The TFLink (Müller-Dott *et al.*, 2023), a high-quality collection of signed TF-gene interactions for 1186 TFs, is leveraged as default to initialize our regulatory network. The scRNA-seq data undergoes standard preprocessing using Scanpy, including filtering out low-quality cells based on mitochondrial gene content, selecting highly variable genes, and performing log-normalization. Further filtering is applied such that only genes expressed in at least 10 cells and cells expressing at least 3 genes are retained. Additionally, genes must have at least one associated TF in the prior GRN to be included. TFs regulating fewer than 10 target genes are removed. These preprocessing steps are intended to reduce the risk of inferring spurious regulatory links from transcription factors with limited target information.

2.3. Model Architecture and Implementation

The VAE architecture consists of fully connected layers with ReLU activation functions, forming both encoder and decoder networks. The encoder (1-3 layers) maps input gene expression profiles to a lower-dimensional latent space, while the decoder (1-2 layers) first infers TF activity and then reconstructs gene expression using a trainable GRN. Latent variables are inferred via the reparameterization trick, and the model is trained using variational inference. Optimization details, including loss functions and hyperparameter selection, are described in Section 2.3.4. The VAE model was implemented using PyTorch (Paszke *et al.*, 2019).

2.3.1. Variational Inference of TF Activity

scRegulate employs a VAE to infer latent TF activity from gene expression data. The input gene expression vector $\mathbf{x}$ is mapped to a latent representation $\boldsymbol{z}$ using an encoder network parameterized by mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}$:

$$\boldsymbol{\mu}, \boldsymbol{\sigma} = q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$$

here, $q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x})$ denotes the encoder network, which outputs the mean μ and standard deviation $\boldsymbol{\sigma}$ of the latent variable $\mathbf{z}$ given the input $\mathbf{x}$.

The latent variable $\mathbf{z}$ is obtained by shifting and scaling a standard normal random variable $\boldsymbol{\epsilon}$ drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I})$, also known as the reparameterization trick:

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$$

The latent representation $\mathbf{z}$ is then decoded to estimate TF activities embeddings $\hat{\mathbf{e}}_{\mathrm{TF}}$:

$$\hat{\mathbf{e}}_{\mathrm{TF}} = p_{\boldsymbol{\theta}}(\mathbf{e}_{\mathrm{TF}}|\mathbf{z})$$

here, $\boldsymbol{e}_{\mathrm{TF}}$ represents the unobserved true TF activity embedding that the model aims to estimate. These inferred TF activities are then mapped to gene expression through the GRN weight matrix W_{GRN}:

$$\hat{\mathbf{x}} = \mathbf{W}_{\mathrm{GRN}} \cdot \hat{\mathbf{e}}_{\mathrm{TF}}$$

where $\hat{\mathbf{x}}$ is the reconstructed gene expression vector, $\hat{\mathbf{e}}_{\mathrm{TF}}$ is the inferred TF activity embedding vector, and $\mathbf{W}_{\mathrm{GRN}}$ is the gene regulatory network weight matrix defining TF-gene interactions. This explicit relationship defined by the last layer in the neural network is one of the key features of scRegulate, allowing direct examination of TF effects on target gene expression and enhancing model interpretability.

2.3.2. Integration of Prior GRNs into TF Activity Estimation

To initialize $\hat{\mathbf{e}}_{\mathrm{TF}}$, scRegulate employs a Univariate Linear Model (ULM) inspired by (Badia-i-Mompel *et al.*, 2022), where TF activities are estimated as a weighted sum of all target gene expression values in the prior GRN (Supplementary Note 1). To balance the contribution between the prior-driven estimate $\mathbf{e}_{\mathrm{TF}}^{\mathrm{ULM}}$ and the data-driven estimate $p_\theta(\mathbf{e}_{\mathrm{TF}}|\mathbf{z})$ from decoder, the contribution of $\mathbf{e}_{\mathrm{TF}}^{\mathrm{ULM}}$is gradually reduced over training by a dynamic scheduling parameter α:

$$\hat{\mathbf{e}}_{\mathrm{TF}} = (1-\alpha)\mathbf{e}_{\mathrm{TF}}^{\mathrm{ULM}} + \alpha p_\theta(\mathbf{e}_{\mathrm{TF}}|\mathbf{z})$$

where $\alpha \in [0, \alpha_{max}]$, and α follows an increasing schedule to transition from prior-driven $\mathbf{e}_{\mathrm{TF}}^{\mathrm{ULM}}$ toward data-driven inference $p_\theta(\mathbf{e}_{\mathrm{TF}}|\mathbf{z})$. Further,

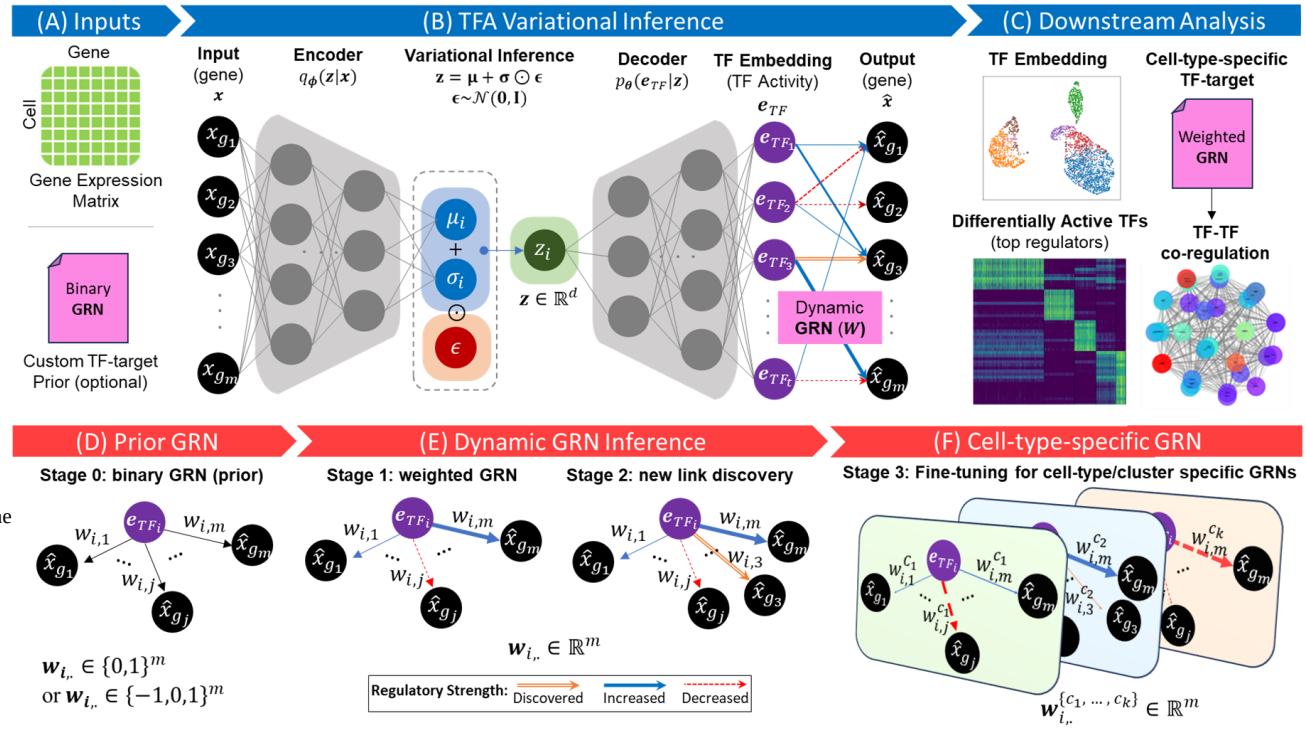

Figure 1. Overview of scRegulate. A schematic representation of scRegulate's approach to inferring TF activity and regulatory networks from single-cell gene expression data. **(A)** Inputs include a gene expression matrix and an optional TF-target gene prior GRN. **(B)** TF activity is inferred through a variational autoencoder framework, where the encoder maps input data into a latent space (z), followed by a TF embedding layer to predict gene expression and capture TF activities, informed by the dynamic GRN. **(C)** Downstream analysis leverages TF embeddings for clustering, identification of differentially active TFs. Construction of cell-type-specific weighted GRNs, provides insights into TF-target regulation and TF-TF co-regulation. **(D)** Prior GRN initializes the regulatory network using a binary or ternary TF-target prior. **(E)** Dynamic GRN inference refines the network by adjusting regulatory strengths (edge weights of TF-target gene links), discovering new connections, and evolving to a weighted GRN. **(F)** Cell-type-specific GRN fine-tunes the GRN for individual cell types or clusters, resulting in tailored regulatory networks. This comprehensive workflow integrates transcription factor activity inference with dynamic GRN updates, enabling cell-type-specific regulatory insights.

scheduling of α follows a controlled incremental approach, transitioning from its initial value α_{start} to its maximum $\alpha_{\max}$ in discrete steps defined by the hyperparameter $\Delta\alpha$. The update rule for α is formulated as:

$$\alpha^{(t+1)} = \min\left(\alpha_{\max}, \alpha^{(t)} + \Delta\alpha\right)$$

here, each training iteration is denoted as t.

2.3.3. GRN Prior Enforcement and Dynamic Updating

To reduce noise in the inferred regulatory interactions, weak regulatory interaction in the $\mathbf{w}_{\text{GRN}}$ are penalized by enforcing sparsity as L1 regularization:

$$\mathcal{L}_{\text{GRN}} = \gamma \sum_i \left\|\mathbf{w}_{\text{GRN},i}\right\|_1$$

where γ is the regularization parameter linearly scheduled over training epochs, and the sum is taken over all elements $\mathbf{w}_{\text{GRN},i}$, the vector of regulatory weights for gene i. The GRN prior is enforced using a mask factor $m(t)$, which gradually transitions from 1 to 0 over training epochs:

$$m(t) = \frac{1}{1 + \exp\left(-\frac{t - T/2}{T/20}\right)}$$

where T is the total number of epochs. This ensures that initial regulatory constraints are respected while allowing for new regulatory interactions to emerge. The mask factor is applied to regulate updates to $\mathbf{W}_{\text{GRN}}$:

$$\mathbf{W}_{\text{GRN}}^{t+1} = m(t)\mathbf{W}_{\text{prior}} + (1 - m(t))\mathbf{W}_{\text{GRN}}^{t}$$

This equation updates the GRN weights by blending the previous weight matrix with the newly learned weights, modulated by the mask $m(t)$, ensuring a smooth transition from prior-based regulation to learned interactions.

2.3.4. Loss Functions, Optimization, and Training Details

The VAE is trained to minimize the total Loss which is the sum of the Evidence Lower Bound (ELBO) and the GRN regularization loss:

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{ELBO}} + \mathcal{L}_{\text{GRN}}$$

To ensure model generalizability and prevent overfitting to noise, we apply an 85%-15% split between the training and validation subsets during model training. The optimizer used for training is Adam, with an adaptive learning rate scheduler that dynamically adjusts based on validation loss. Early stopping is triggered when no significant improvements are observed in validation loss over a set number of epochs, preventing unnecessary training cycles. No explicit test set is used from the same dataset, as benchmarking is performed using independent datasets for GRN inference and TF activity evaluation. More details regarding the training are available in the Supplementary Note S2.

2.3.5. Benchmark Datasets

Multiple datasets were used to evaluate the clustering of TF embeddings, GRN, and TF activity inference. To investigate how well the TF embedding captures the cell heterogeneity in the data, we used a combination of publicly available scRNA-seq datasets, including three Tabula Muris datasets (i.e., heart, lung, and brain) (The Tabula Muris Consortium *et al.*, 2018), PBMC, and human middle temporal gyrus (MTG) scRNA-seq (Allen Institute for Brain Science, 2022) datasets where ground-truth cell-type labels are available.

For GRN benchmarking, we utilized GRouNdGAN (Zinati *et al.*, 2024), a simulated single-cell dataset with imposed regulatory interactions, making it a controlled benchmark for testing network reconstruction

accuracy. To ensure robust performance evaluation and stability across different subsets, we employed proportional stratified sketching to generate five independent subsets. This approach ensures that each subset retains the full diversity of cell types while maintaining proportional representation across the dataset, reducing bias introduced by random sampling. The procedure involved clustering the dataset using Leiden algorithm, followed by sketch-based subsampling within each cluster using geosketch. Each pseudo-cluster was assigned a subset size proportional to its overall representation in the dataset. This method ensures that smaller clusters were not underrepresented while maintaining a well-balanced distribution across all five subsets. GRN accuracy was measured using AUROC and AUPRC (Supplementary Note 3), computed over the intersection of target genes between scRegulate-inferred networks and the imposed ground-truth networks in GRouNdGAN. The intersection accounts for prior inconsistencies and provides a fair basis for evaluation. Additionally, splitting the dataset into multiple subsets provides a bounded range for GRN performance metrics, ensuring that results are not biased by a single instance of dataset partitioning.

For TF activity benchmarking, we employed a Perturb-seq dataset, in which specific TFs were inhibited via CRISPR interference (CRISPRi) (Dixit *et al.*, 2016). The experimental TF knockdown effects were used as a reference for evaluating inferred TF activities.

Finally, to demonstrate real-world applicability, we applied scRegulate to a PBMC 3k dataset, a well-characterized immune cell dataset. This allowed for an in-depth exploration of GRN structures and TF activity inference at single-cell resolution, with extensive analysis of TF-TF interactions, differentially active TFs, and functional gene enrichment analysis.

Dataset details, including the number of cells and genes, are provided in Supplementary Table S1.

2.4. Benchmarking Strategy and Evaluation Metrics

2.4.1. Clustering Performance Evaluation

Leiden clustering is applied to the TF embeddings at multiple resolutions (0.2-3.0), and cluster assignments are compared to ground truth labels using standard clustering metrics. Each cluster is assigned a label based on maximum voting. Specifically, for each cluster, we tally the ground-truth cell-type labels of all the cells within that cluster. The label that appears most frequently is then assigned as the representative label for the cluster. This process enables the computation of clustering evaluation metrics (ARI, NMI, macro F1) against the known cell-type annotations. Formal definitions and computational formulas for these evaluation metrics are provided in Supplementary Note 2. To assess the robustness of scRegulate's embedded TF clustering, we simulated dropout effects by applying a masking process where each gene expression value X_{ij} was retained with probability $1-p$ and set to zero otherwise, using:

$U_{ij} \sim \text{Uniform}(0,1), D_{ij} = 1(U_{ij} > p), \boldsymbol{X}' = \boldsymbol{X} \odot \boldsymbol{D}, p \in \{0.1, 0.3\}$

This mimics the sparsity commonly observed in single-cell RNA-seq data.

2.4.2. GRN Inference Evaluation

GRN inference performance is evaluated using AUROC and AUPRC, computed via standard numerical integration. The inferred weighted GRNs are binarized at varying thresholds and compared against a gold-standard binary GRN to assess performance stability. To evaluate GRN consistency across cell types, we compute Pearson correlation between TF-target interaction matrices without additional normalization or thresholding. Additionally, TF-TF co-regulation similarity is assessed by computing Pearson correlation across rows of $\boldsymbol{W}_{GRN}$, both within and across cell types, focusing on a biologically relevant subset of TFs. All equations and detailed formulations are provided in Supplementary Note 3.

2.4.3. Benchmarking TF Activity using Real Perturb-seq Data

The performance of scRegulate in inferring TF activity was evaluated using a Perturb-seq dataset, where experimental perturbations (i.e., knockdowns) of key TFs were performed. In this dataset, the changes in gene expression resulting from these perturbations serve as independent ground-truth indicators of TF activity. Specifically, we compared the predicted TF activity shifts against the experimentally observed changes using log_2 fold change (LFC). In addition, Wilcoxon rank-sum tests were performed to assess the capability of scRegulate in distinguishing significantly perturbed TFs from non-perturbed ones.

2.5. Application of scRegulate to Real PBMC Data

To illustrate the utility of scRegulate on real-world data, we applied the framework to a well-characterized PBMC 3k dataset. This dataset, comprising peripheral blood mononuclear cells from healthy donors, was preprocessed following standard scRNA-seq pipelines inspired by scanpy. scRegulate was then used to infer cell-type-specific TF activities and reconstruct GRNs at single-cell resolution. The results were then investigated for comparative and functional biological interpretations.

2.5.1. Comparative Analysis of the Inferred TF Activity

We performed a differential analysis of TF activities across cell types. Specifically, we applied the Wilcoxon rank-sum test to compare TF activity profiles between groups, retaining only those TFs with an adjusted p-value below 0.05. From this subset, we ranked the differentially active TFs based on the highest LFC to prioritize the most robustly modulated regulators.

2.5.2. Functional Validation of the GRN

Building upon the analysis of TF activities, we next focused on validating the cell-type-specific GRN inferred by scRegulate within the PBMC dataset. We evaluated the similarity of GRNs across cell types by computing Pearson correlation coefficients between their TF-target interaction matrices, and then hierarchically cluster the resulting correlations to identify groups of cell types with similar regulatory profiles. This analysis enabled us to quantify the conservation of regulatory interactions and to identify clusters of TFs that co-regulate similar sets of target genes. In addition, we conducted TF-TF co-regulation analyses. We first selected the top differentially active TFs from each cell type along with their corresponding GRN weight vectors. We then computed the cosine similarity between each pair of these TFs, both within the same cell type and across different cell types, to quantify the degree of co-regulation. Assume we have two GRN vectors $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$, representing the regulatory weight profiles of a TF from two different cell types (or conditions). The cosine similarity between these two vectors is given by:

$$\text{Cosine Similarity}(\boldsymbol{w_1}, \boldsymbol{w_2}) = \frac{\boldsymbol{w_1} \cdot \boldsymbol{w_2}}{\|\boldsymbol{w_1}\| \|\boldsymbol{w_2}\|}$$

Where $\boldsymbol{w_1} \cdot \boldsymbol{w_2}$ is the dot product of the two vectors and $\|\boldsymbol{w}\|$ is the Euclidian norm of a vector $\boldsymbol{w}$. This analysis allowed us to identify TFs that share similar regulatory profiles and to reveal potential shared regulatory mechanisms across cell types.

3 Results

3.1. scRegulate TF Embeddings Capture Cellular Heterogeneity

scRegulate's TF embeddings provide a structured representation of cellular heterogeneity. Clustering analyses revealed that scRegulate's TF embeddings achieved highly accurate cell-type separations, surpassing other methods in key clustering metrics (Figures 2A, and S1). For PBMC, scRegulate achieved an ARI of 0.814 ± 0.042, NMI of 0.821 ± 0.021, F1-score of 0.868 ± 0.118, and AUC of 0.930 ± 0.058, significantly outperforming SCENIC, decoupleR and BIOTIC (Wilcoxon signed-rank test, $p < 0.01$). Similar trends were observed across the other datasets, with

scRegulate achieving an

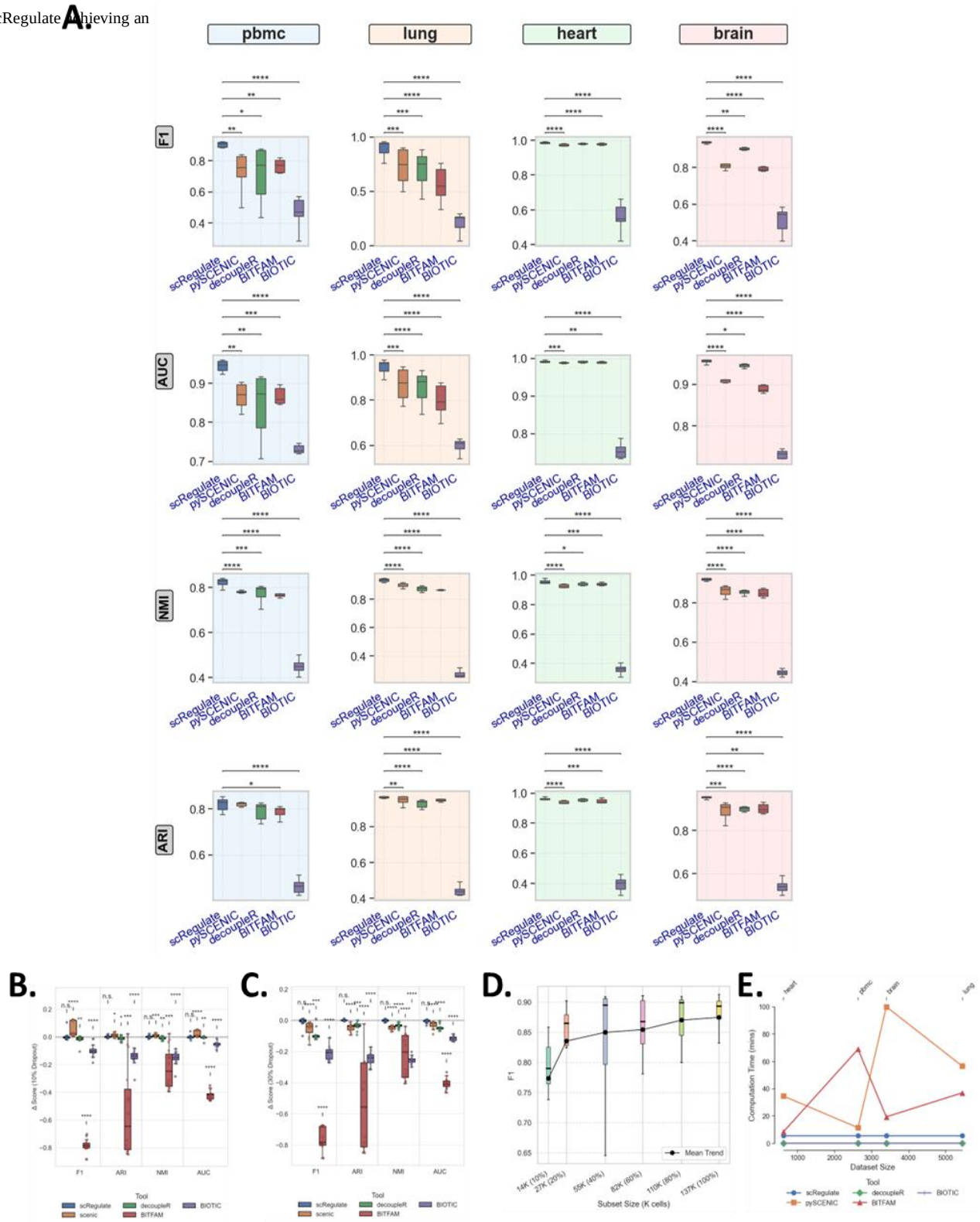

Figure 2. Comparison of TF-based embedding methods on four scRNA-seq datasets. (A) Clustering performance (ARI, NMI, F1, and AUC) of four methods (scRegulate, SCENIC, BITFAM, decoupleR and BIOTIC) on TF activities using the PBMC, mouse brain, mouse lung, and mouse heart datasets. Each boxplot displays performance using different Leiden clustering resolutions; asterisks indicate statistical significance based on p-values ($p < 0.05$: *; $p < 0.01$: **; $p < 0.001$: ***). **(B)** Effect of 10% simulated dropout and **(C)** 30% simulated dropout on using the brain Tabula Muris dataset, highlighting each method's robustness to partial TF information. **(D)** F1-score across dataset sizes for the MTG dataset, showing performance scalability. **(E)** Computation times for each method across the four datasets, illustrating differences in runtime as dataset size and complexity vary. In all plots, error bars or box boundaries represent variability across different Leiden resolutions.

heart, and 0.953 ± 0.015 in brain, outperforming the other methods.

To assess the robustness of scRegulate's embedded TF clustering, we simulated dropout effects by randomly removing 10% and 30% of gene expression values, mimicking the sparsity commonly observed in single-cell data (Figures 2B,C and S2). scRegulate maintained high clustering performance across different Leiden resolutions, whereas the alternative methods exhibited significant performance degradation (Wilcoxon signed-rank test, $p < 0.01$). scRegulate achieved consistently high F1-scores as dataset size increased on the MTG dataset, demonstrating its scalability for large-scale single-cell applications (Figures 2D, S3). Finally, we evaluated computation times across all methods and observed that scRegulate efficiently scales with dataset size, outperforming SCENIC and BITFAM in runtime while maintaining high accuracy (Figure 2E). Taking together, these results highlight scRegulate's ability to generate biologically meaningful embeddings that more accurately reflect transcriptional variation across cell types, reinforcing its utility for robust single-cell clustering.

3.2. scRegulate Outperforms Existing Methods in GRN Inference on Gold-Standard Synthetic Data

Across all benchmarked synthetic datasets (Dahlin, PBMC, and Tumor), scRegulate consistently outperformed SCENIC and BITFAM based on AUROC and AUPRC metrics (Figures 3A, and S4). Notably, in the PBMC dataset, scRegulate achieved the highest AUPRC (0.948) and a strong AUROC (0.707), significantly surpassing pySCENIC (AUPRC = 0.746, AUROC = 0.568) and BITFAM (AUPRC = 0.398, AUROC = 0.537). A similar trend was observed in the Tumor dataset, where scRegulate's AUPRC reached 0.858, exceeding both pySCENIC (0.744) and BITFAM (0.490). The Dahlin dataset followed this pattern, with scRegulate achieving an AUPRC of 0.796, while pySCENIC and BITFAM scored lower at 0.588 and 0.313, respectively. These results demonstrated that scRegulate consistently inferred gene regulatory networks with higher precision compared to the other approaches.

3.3. scRegulate Outperforms Alternatives in Predicting TF Activity from Experimental Knockdowns

To assess scRegulate's ability to correctly infer TF activity, we evaluated its performance on the Perturb-seq dataset. Since CRISPRi leads to transcriptional repression, we expect negative log-fold changes (LFCs) in the inferred TF activities when comparing the knockdown cells vs. control cells. We focused on double-knockdown perturbations to minimize off-target effects and provide a more reliable benchmarking framework for TF activity inference. Across all tested perturbations, scRegulate consistently inferred significantly reduced TF activity in the knockdown cells compared to controls (Wilcoxon signed-rank test, $p < 0.01$). Notably, for ELK1, scRegulate detected a mean TF activity decrease of -0.66, with a LFC of -0.68 ($p = 8.06\times10^{-13}$). Similar trends were observed for EGR1 (LFC = -0.76, $p = 4.36\times10^{-7}$) and CREB1 (LFC = -16.98, $p = 3.00\times10^{-6}$), confirming that scRegulate correctly captured expected repression patterns (Figures 3B, and S5). Compared to the alternative methods, scRegulate demonstrated higher precision in detecting TF knockdowns. For ELK1, both SCENIC and decoupleR failed to detect significant repression ($p > 0.7$), while BITFAM incorrectly inferred increased activity (LFC = 0.18, $p = 0.0068$). In contrast, scRegulate's results were more consistent with ground truth expectations, achieving an AUROC of 0.87 and AUPRC of 0.84, outperforming decoupleR (AUROC ≈ 0.72, AUPRC ≈ 0.60, $p = 0.008$). These results confirm that scRegulate provides an accurate and biologically consistent framework for inferring TF activity from the CRISPRi-based Perturb-seq data, correctly capturing repression effects where other methods struggle.

3.4. scRegulate Uncovers Cell-Type-Specific Regulatory Programs in Real PBMC Data

Applying scRegulate to the PBMC scRNA-seq data revealed distinct TF activity patterns across 8 immune cell types (Figure 4A). In contrast, the latent representation of gene expression based on principal component analysis (PCA) resulted in overlapping clusters, demonstrating that scRegulate's regulatory-centric embedding provides a more biologically informative representation of single-cell states (Figure S6). Differential analysis of TF activity identified significantly active TFs specific to the immune cell types (Figure 4B). PAX5 is a well-established master regulator of B cell lineage commitment and maintenance, orchestrating B cell identity through repression of alternative lineage genes (Medvedovic *et al.*, 2011). Additionally, BCL6 and IRF4 were recovered, both essential for germinal center formation and plasma cell differentiation, respectively (Ochiai *et al.*, 2013). In CD4 T cells, canonical regulators GATA3 and BATF were identified, consistent with their established roles in Th2 differentiation and effector T cell programming (Schraml *et al.*, 2009). CD8 T cells exhibited high activity of RUNX3, EOMES, STAT4, STAT5A, and TBX21, which collectively orchestrate cytotoxic T cell lineage specification, expansion, and effector function(Kallies *et al.*, 2009). For NK cells, ID2, EOMES, and TBX21critical for cytotoxic development and lineage commitment were among the top-scoring TFs (Gordon *et al.*, 2012). In the myeloid lineage, SPI1 (PU.1) was consistently identified across

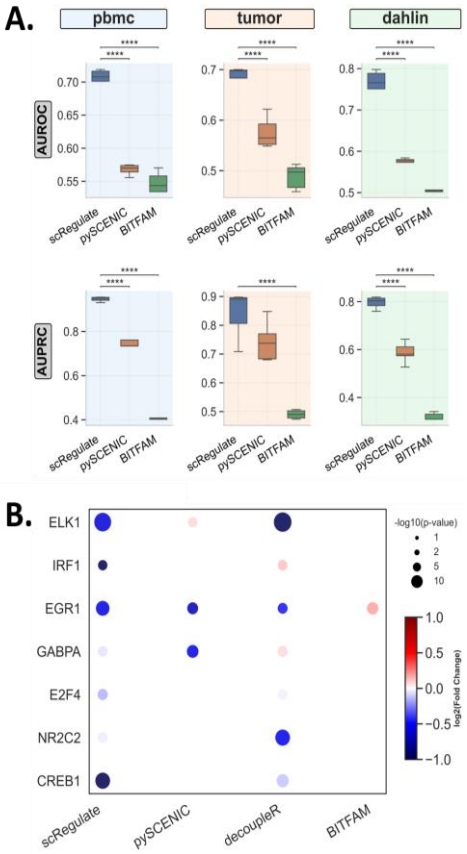

Figure 3. Comparison of scRegulate with Alternative Tools on GRN inference. (A) GRNs inferred from the three synthetic gold standard data of PBMC, tumor, and Dahlin. **(B)** TF activity inferred using real perturb-seq dataset with double knockdown effect ($p < 0.0001$: ****).

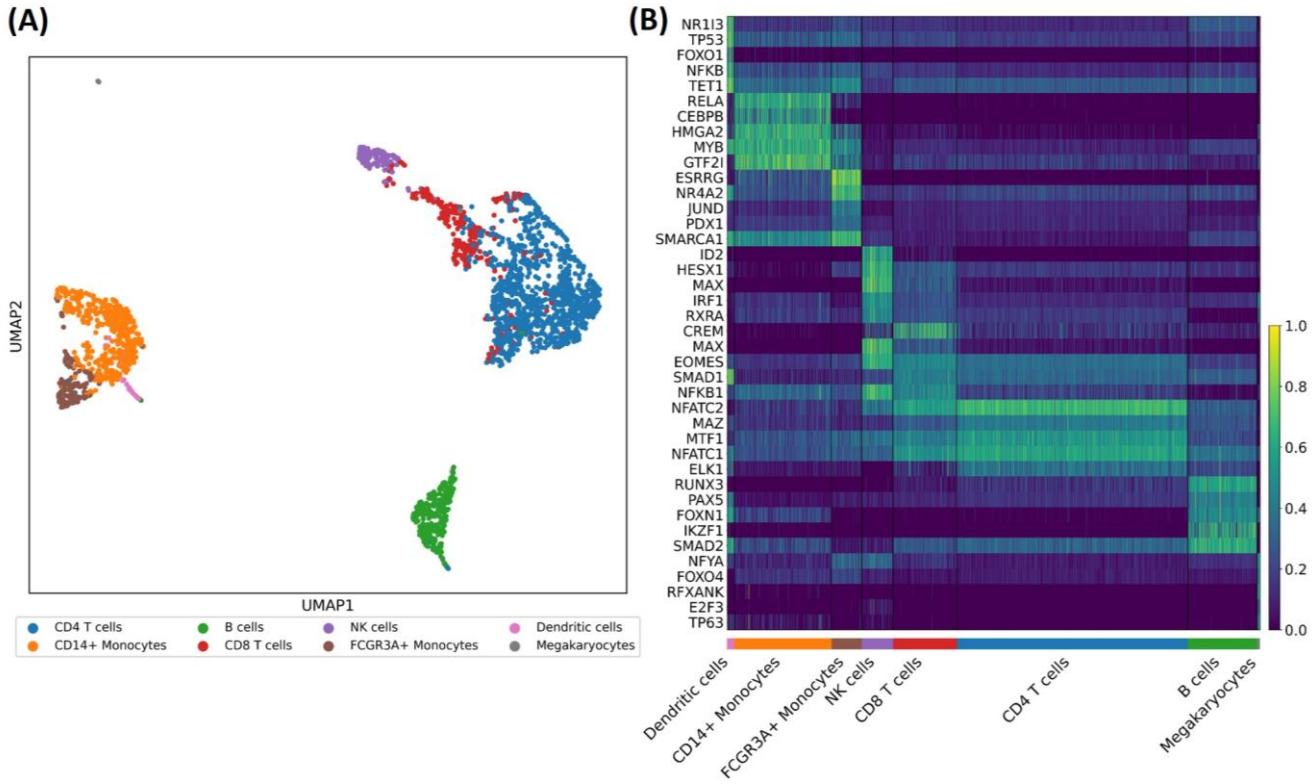

Figure 4. Application of scRegulate in PBMC dataset. (A) UMAP visualization of the TF embeddings demonstrates the cell type heterogeneity, revealing eight distinct cell types. **(B)** Heatmap showcasing the top 5 differentially active TFs per cell type, including TFs well-known for regulating their respective cell types. **(C)** UMAP highlighting the activity of cell type-specific TFs within their corresponding cell types.

monocytes and dendritic cells, reflecting its central role in myeloid fate specification(Heinz *et al.*, 2010). CEBPB and CEBPE further supported monocyte identity through their roles in granulocyte-monocyte differentiation (Kim *et al.*, 2024). The recovery of these hallmark TFs provides independent support for the accuracy and interpretability of our inferred TF activity profiles. (Figure S7).

Hierarchical clustering of inferred GRNs for the eight cell types further supported the biological relevance of scRegulate's predictions (Figure S8). CD4 and CD8 T cell subtypes clustered together, reflecting their shared lineage, while monocyte subsets grouped with dendritic cells, mirroring their myeloid origins. Notably, megakaryocytes formed a distinct cluster, highlighting their unique regulatory program separate from immune cells. To explore the structure of these inferred GRN networks, we examined the top TF–target interactions across cell types (Figure S8). Key co-regulatory targets emerged, for example, CEBPA, and JUND co-regulate inflammatory genes in monocytes and PAX5 activating B-cell lineage genes. Functional enrichment analysis confirmed that the inferred top targets for top differentially active TFs were highly relevant to their associated cell identity (Figure 5). Finally, a TF–TF co-regulatory network revealed that cell-type-specific TFs form tightly connected regulatory modules, with T-cell-associated TFs clustering together, myeloid TFs forming distinct regulatory units, and megakaryocyte TFs positioned independently (Figure S8).

Overall, scRegulate provides a biologically consistent and interpretable map of TF activity and regulatory interactions in PBMCs. By identifying key regulators, mapping TF–target interactions, and revealing higher-order regulatory relationships, scRegulate provides deeper insights into cell-type-specific transcriptional programs in single-cell data.

4 Discussion and conclusions

We developed scRegulate, an interpretable VAE to infer GRNs using scRNA-seq data. By embedding prior knowledge of TF-gene interactions, scRegulate can infer cell-type-specific GRNs, with the ability to predict novel gene targets, and TF activity estimation for each cell. scRegulate leverages ULM-based initialization, ensuring TF activity inference starts from a biologically meaningful estimate rather than full randomness. This improves convergence, stability, and reproducibility, reducing sensitivity to initialization biases common in deep generative models. To facilitate the generation of additional insights on the regulatory programs, scRegulate also perform downstream analyses, such as clustering of cells using TF activities, identification of differentially active TFs across different cell types, and constructing TF-TF coregulatory network using WGCNA (Zhang and Horvath, 2005).

Our benchmarking demonstrated that scRegulate outperformed the existing methods in clustering accuracy and GRN inference across PBMC, lung, heart, and brain datasets compared to decoupleR, SCENIC and BITFAM. For the CRISPRi Perturb-seq data, scRegulate accurately predicted negative TF activity changes in knockdown cells, outperforming other methods. These findings highlight its strength in recovering GRNs across diverse contexts. In the PBMC case study, we demonstrated its ability to recover multiple known lineage-defining TFs, such as BATF in T cells, CEBPA in monocytes, and PAX5 in B cells, while hierarchical

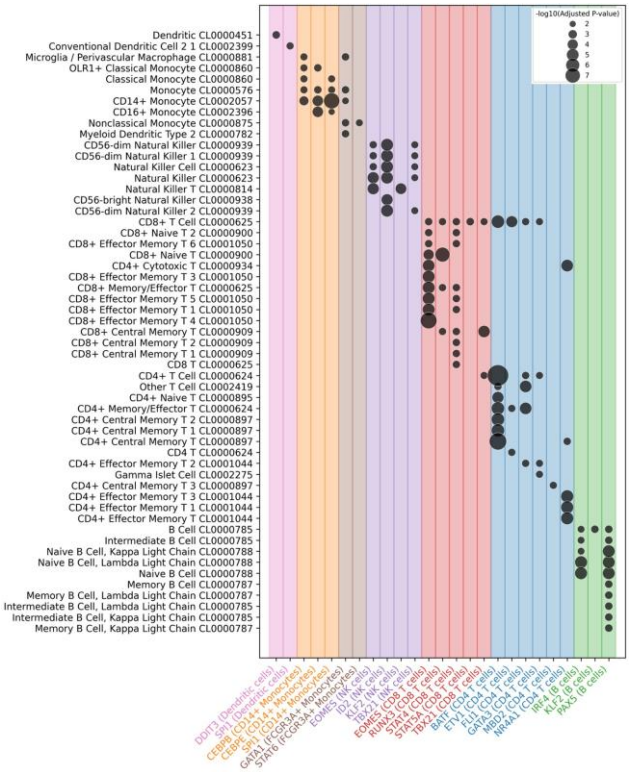

Figure 5. Insights into Cell-Type-Specific Regulatory Networks and Functional Enrichment. Enriched GO terms (Azimuth Human Cell Atlas) based on the top 1% target genes of differentially active TFs from cell type-specific GRN, aligning with cell-type identities.

clustering of inferred GRNs reflected expected lineage relationships. Compared to gene-expression-based clustering, TF embeddings provided a more structured, biologically interpretable representation of cellular heterogeneity.

Unlike existing methods, scRegulate jointly models TF activity embeddings and GRNs in a cell-type-specific manner. It provides two key outputs: (1) a TF activity embedding to identify differentially active TFs and (2) distinct GRNs for each cell type. In contrast, other methods infer TF activity and a global GRN, but do not model cell-type-specific networks. In addition, scRegulate leverages all genes that pass basic filtering criteria, overcoming the limitation of approaches that rely on highly variable genes. This expansion allows to retain broader regulatory signals for GRN inference. Finally, scRegulate's ability to construct unique regulatory networks per cell type make it unique in offering a high-resolution view of transcriptional regulation at single-cell granularity.

A key limitation of scRegulate is its dependence on prior GRNs, which may introduce biases if the prior is incomplete. Additional hyperparameter tuning may be required in certain datasets where regulatory landscapes differ significantly from available priors. As a potential remedy, the prior GRN can be enhanced by incorporating putative TF-gene interactions inferred from motif analysis of chromatin accessibility regions if scATAC-seq data are available under the same biological context, as detailed in our previous work (Zandigohar *et al.*, 2024). Computational cost is another consideration, as training a variational model is more resource-intensive than simpler approaches. However, scRegulate's prior-guided learning improves interpretability and result stability, making it more robust than purely data-driven methods. Utilization of GPU acceleration (CUDA) can significantly overcome computational bottlenecks for faster and scalable training on large single-cell datasets. Future improvements could directly integrate scATAC-seq to refine TF-target inference. Expanding scRegulate to model regulatory dynamics across conditions, disease states or perturbation responses, would further enhance its applicability. While every method has trade-offs, scRegulate fills an important gap between prior-driven inference and deep-learning-based discovery, making it a valuable tool for dissecting single-cell gene regulation.

Acknowledgements

We sincerely thank Brandon Lukas for his valuable feedback on this manuscript. This work was supported by the Division of Diabetes, Endocrinology, and Metabolic Diseases of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) through the Diabetic Complications Consortium under Grants DK076169 and DK115255. The authors declare no conflicts of interest.

Data Availability and Supplementary Data

Supplementary data are available at Bioinformatics online. All datasets used in this study are publicly available. Tabula Muris (lung, heart, brain) Smart-seq2 data were downloaded from Figshare with pre-annotated cell types. Benchmarking datasets and gold-standard GRNs were obtained from the GRouNdGAN repository, including PBMC, Tumor, and Dahlin datasets. All code, filtering steps, and the scRegulate package are available at https://github.com/YDaiLab/scRegulate, with full reproducibility ensured using a global seed of 42.

References

Aibar,S. *et al.* (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat Methods*, **14**, 1083–1086.

Allen Institute for Brain Science (2022) Allen Cell Types Database -- Human MTG 10x.

Badia-i-Mompel,P. *et al.* (2022) decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, **2**, vbac016.

Bravo González-Blas,C. *et al.* (2023) SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods*, **20**, 1355–1367.

Cao,L. *et al.* (2025) BIOTIC: a Bayesian framework to integrate single-cell multiomics for transcription factor activity inference and improve identity characterization of cells. *Briefings in Bioinformatics*, **26**, bbaf013.

Dixit,A. *et al.* (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*, **167**, 1853-1866.e17.

Gao,S. *et al.* (2021) A Bayesian inference transcription factor activity model for the analysis of single-cell transcriptomes. *Genome Res.*, **31**, 1296–1311.

Gayoso,A. *et al.* (2022) A Python library for probabilistic analysis of single-cell omics data. *Nat Biotechnol*, **40**, 163–166.

Gordon,S.M. *et al.* (2012) The Transcription Factors T-bet and Eomes Control Key Checkpoints of Natural Killer Cell Maturation. *Immunity*, **36**, 55–67.

Hecker,D. *et al.* (2023) Computational tools for inferring transcription factor activity. *Proteomics*, **23**, 2200462.

Heinz,S. *et al.* (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, **38**, 576–589.

Kallies,A. *et al.* (2009) Blimp-1 Transcription Factor Is Required for the Differentiation of Effector CD8+ T Cells and Memory Responses. *Immunity*, **31**, 283–295.

Kamimoto,K. *et al.* (2023) Dissecting cell identity via network inference and in silico gene perturbation. *Nature*, **614**, 742–751.

Kim,S. *et al.* (2024) Transcription factor C/EBPα is required for the development of Ly6C^{hi} monocytes but not Ly6C^{lo} monocytes. *Proc. Natl. Acad. Sci. U.S.A.*, **121**, e2315659121.

Lambert,S.A. *et al.* (2018) The Human Transcription Factors. *Cell*, **172**, 650–665.

Lopez,R. *et al.* (2020) Enhancing scientific discoveries in molecular biology with deep generative models. *Molecular Systems Biology*, **16**, e9198.

Medvedovic,J. *et al.* (2011) Pax5. In, *Advances in Immunology*. Elsevier, pp. 179–206.

Müller-Dott,S. *et al.* (2023) Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. *Nucleic Acids Research*, **51**, 10934–10949.

Ochiai,K. *et al.* (2013) Transcriptional Regulation of Germinal Center B and Plasma Cell Fates by Dynamical Control of IRF4. *Immunity*, **38**, 918–929.

Paszke,A. *et al.* (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library.

Schraml,B.U. *et al.* (2009) The AP-1 transcription factor Batf controls TH17 differentiation. *Nature*, **460**, 405–409.

Svensson,V. *et al.* (2020) Interpretable factor models of single-cell RNA-seq via variational autoencoders. *Bioinformatics*, **36**, 3418–3421.

The Tabula Muris Consortium *et al.* (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, **562**, 367–372.

Voss,T.C. and Hager,G.L. (2014) Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet*, **15**, 69–81.

Xiong,L. *et al.* (2019) SCALE method for single-cell ATAC-seq analysis via latent feature extraction. *Nat Commun*, **10**, 4576.

Zandigohar,M. *et al.* (2024) Transcription Factor Activity Regulating Macrophage Heterogeneity during Skin Wound Healing. *The Journal of Immunology*, **213**, 506–518.

Zhang,B. and Horvath,S. (2005) A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, **4**.

Zinati,Y. *et al.* (2024) GRouNdGAN: GRN-guided simulation of single-cell RNA-seq data using causal generative adversarial networks. *Nat Commun*, **15**, 4055.