EnsembleRegNet: Interpretable Deep Learning for Transcriptional Network Inference from Single-Cell RNA-Seq

Duaa Mohammad Alawad[1], Ataur Katebi[2,3], Md Tamjidul Hoque[1,*]

dmalawad@uno.edu, a.katebi@neu.edu, thoque@uno.edu

[1]Department of Computer Science, University of New Orleans, New Orleans, LA, USA.

[2]Department of Bioengineering, Northeastern University, Boston, MA, USA.

[3]Center for Theoretical Biological Physics, Northeastern University, Boston, MA, USA.

Abstract

Gene regulatory networks (GRNs) govern gene expression and cellular identity, but accurately inferring their structure from high-dimensional single-cell RNA sequencing (scRNA-seq) data remains a major challenge. Here, we present EnsembleRegNet, a deep learning framework that infers transcription factor (TF)–target gene relationships by integrating an ensemble encoder-decoder and multilayer perceptron (MLP) architecture. EnsembleRegNet utilizes Hodges-Lehmann estimator (HLE)-based binarization, case-deletion analysis, motif enrichment using RcisTarget, and regulon activity scoring with AUCell to enhance both robustness and biological interpretability. Extensive evaluations across simulated and real scRNA-seq datasets demonstrate that EnsembleRegNet outperforms existing GRN inference methods, including SCENIC and SIGNET, in both clustering performance and regulatory accuracy. By uncovering cell-type-specific regulatory modules and enhancing interpretability, EnsembleRegNet offers a scalable and biologically grounded framework for exploring transcriptional regulation. Its demonstrated performance establishes a new benchmark for GRN inference and highlights its promise for applications in disease modeling, biomarker discovery, and cellular reprogramming.

* Corresponding author: thoque@uno.edu

Keywords: Ensemble, encoder-decoder, deep learning, gene regulatory network inference, cell clustering

1. Introduction

Gene expression regulation is essential for cellular identity, development, and the onset or progression of various diseases [1][2]. At its core, this regulation is orchestrated by gene regulatory networks (GRNs), which map the interactions between transcription factors (TFs) and their target genes across diverse biological contexts. GRNs encode the logic of cellular behavior by linking TFs to their regulatory targets [3, 4][5][6]. These networks are dynamic and often cell-type-specific, making them crucial for understanding both normal biological processes and pathological states. However, deciphering GRNs remains challenging due to the complexity and context-dependence of these regulatory relationships.

Transcription factors (TFs) serve as master regulators that bind specific DNA sequences to control gene expression programs [7][8][9]. While their regulatory influence is well known, inferring direct TF-target relationships from high-dimensional gene expression data, particularly in heterogeneous single-cell contexts, remains a non-trivial task. These networks delineate the intricate patterns of regulatory interactions that orchestrate cellular operations, response to external stimuli, and transitions between distinct states, highlighting the dynamic and interlinked nature of cellular regulation [10].

Single-cell RNA sequencing (scRNA-seq) has revolutionized our ability to dissect cellular heterogeneity, allowing for the capture of gene expression at single-cell resolution across thousands of cells [11][12][13]. This wealth of data offers unprecedented opportunities for reconstructing GRNs with fine-grained biological fidelity. However, technical challenges such as

sparsity, dropout noise, and non-linear dependencies limit the effectiveness of many inference approaches.

Despite methodological progress, inferring GRNs from scRNA-seq remains difficult. Challenges include (1) the complexity of regulatory interactions involving both direct and indirect effects, (2) identifying regulators in rare or transitional cell types, and (3) high false positive rates stemming from purely statistical or linear modeling approaches [14][15][16][17].

In computational biology, a wide array of methods has been developed to infer gene regulatory networks (GRNs) from single-cell RNA sequencing (scRNA-seq) data. These approaches span six major model categories, each targeting different facets of regulatory complexity. Linear regression methods, such as SINCERITIES developed by Nan Papili Gao et al., employ regularized regression on time-stamped scRNA-seq data to identify directed gene relationships based on temporal expression dynamics [18]. Differential equation models like SCODE, proposed by Hirotaka Matsumoto and colleagues, use ordinary differential equations to reconstruct dynamic regulatory networks from differentiation trajectories [19]. Information-theoretic models, exemplified by PIDC from Thalia E. Chan's team, apply multivariate information measures and partial information decomposition to uncover dependencies among gene triplets [20].

Bayesian network approaches have incorporated autoregressive moving-average models within variational Bayesian frameworks to analyze pseudo-temporally ordered data, demonstrating flexibility in modeling regulatory uncertainty and temporal progression [21]. Boolean networks offer a discrete logic-based alternative for capturing binary gene activation states, reflecting diverse regulatory patterns across datasets [22]. Finally, causal inference frameworks like Scribe, developed by Qiu et al., employ restricted directed information to detect causal regulatory

interactions, emphasizing the limitations of pseudo-time approximations and suggesting RNA velocity as a potential solution for capturing more accurate temporal coupling in single-cell analyses [23].

Deep learning has been widely applied to scRNA-seq data analysis, particularly for tasks like dropout imputation [24] and cell clustering [25]. Building on this foundation, several recent studies have extended deep learning toward gene regulatory modeling and related biological tasks. For example, **scRGCL** [26] Yuan *et al.,* 2025 employs residual graph convolutional networks with contrastive learning to improve cell type annotation across heterogeneous scRNA-seq datasets, demonstrating that GCN-based models can effectively capture graph-structured transcriptional relationships. **scAMZI** [27] integrates an attention-based autoencoder with a zero-inflated layer to jointly perform feature extraction and clustering, improving robustness to dropout noise and uncovering more accurate cellular subpopulations. Moving closer to GRN inference, **scMGATGRN** [28] introduces a multiview graph attention mechanism that combines gene co-expression, pseudo-time, and similarity graphs, thereby enhancing detection of transcriptional regulatory edges. These advances notwithstanding, the application of deep learning specifically to transcriptional regulatory network inference remains limited. Challenges include modeling non-linear TF–target dependencies, ensuring scalability to large single-cell datasets, and providing biologically interpretable outputs. To address these gaps, we introduce EnsembleRegNet, a deep learning framework that infers GRNs by integrating an encoder–decoder architecture with a multilayer perceptron (MLP). The model operates on the premise that a transcription factor (TF) strongly associated with a target gene's expression likely regulates it.

EnsembleRegNet comprises six integrated components. First, high-quality data preprocessing ensures that scRNA-seq inputs are properly filtered and normalized for downstream

analysis. The second step employs an ensemble of encoder-decoder and MLP models to predict TF-target interactions with improved accuracy and robustness. Third, motif enrichment is validated using RcisTarget, which scores the likelihood of TF binding to predicted targets based on DNA motif data [29]. Fourth, AUCell quantifies TF activity at the single-cell level by calculating area under the curve (AUC) scores for each regulon [30]. In the fifth stage, cell clustering is performed based on regulon activity, enabling the identification of cell-type-specific regulatory programs. Lastly, network visualization reveals the underlying structure of GRNs and highlights key transcriptional regulators, supporting biological interpretation and hypothesis generation.

2. Method

In this section, we outline the framework of our study, which is structured around three core components: firstly, the dataset that serves as the benchmark for assessing the efficacy of our proposed methodology; secondly, a comprehensive description of the EnsembleRegNet framework, which is a useful tool for identifying important regulatory modules driving various biological processes; and thirdly, the cluster validation indices that are implemented to authenticate the clustering outcomes derived from our method.

2.1 Benchmark Dataset

In this study, the performance of EnsembleRegNet in clustering tasks was assessed using three single-cell RNA sequencing (scRNA-seq) datasets obtained from the Gene Expression Omnibus (GEO) database. These datasets encompassed a diverse set of biological samples: cerebral cortex cells from mice (GSE60361), consisting of 3005 single cells derived from a group of 67 mice [31]; human skin cutaneous melanoma (SKCM, GSE115978) cells obtained from 31 melanoma tumors

[32], totaling 7186 single cells; and cells from head and neck squamous cell carcinoma (HNSCC, GSE103322), specifically from oral cavity tumors of 18 patients, summing up to 5902 single cells [33]. For a comparative analysis of cell clustering, we employed not only SCENIC [2] but also SIGNET [34], which is recognized as a benchmark in the field.

2.2 EnsembleRegNet Framework

The EnsembleRegNet approach is structured around six key components, as depicted in Figure 1: Data preprocessing and binarization, predicting the interaction between transcription factors and target genes, RcisTarget utilization, AUCell analysis, cell clustering, and network visualization.

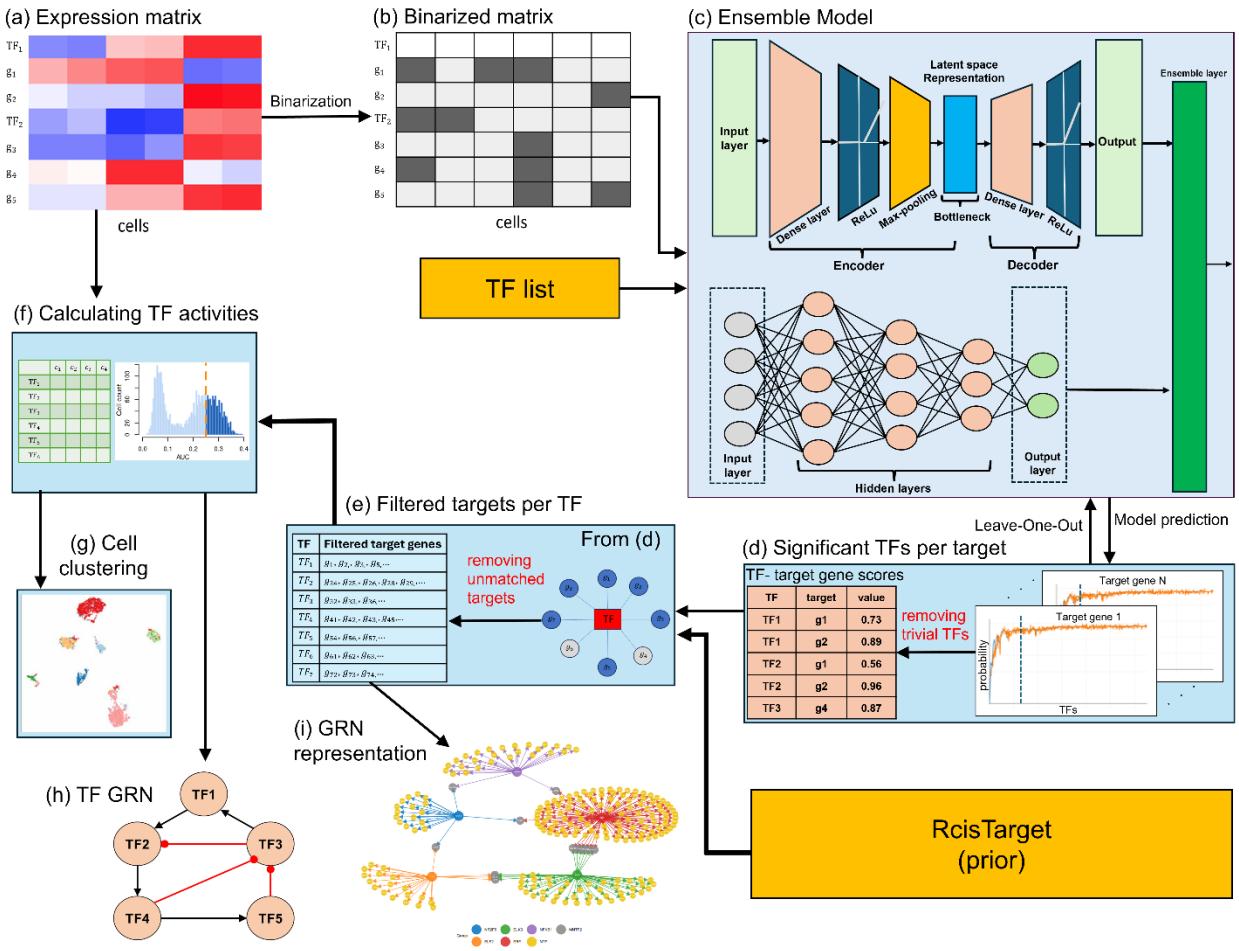


Figure 1. Overview of EnsembleRegNet framework: An integrative approach for deducing gene regulatory networks from scRNA-seq data. The process begins with the preprocessing of scRNA-seq data and applying the binarization

method to form a binary matrix (top left panels *a*, *b*), which, along with a transcription factor (TF) list (top middle), feeds into the ensemble model comprising an encoder, decoder, and multilayer perceptron (MLP) (top right panel *c*). The framework includes a mechanism for refining TF-gene interactions by removing less significant TFs, shown through a threshold-based filtering process (middle right panel *d*). RcisTarget is then employed to integrate existing biological knowledge into the network (bottom right), which helps in generating a curated gene list after the exclusion of non-unmatched targets (middle panel *e*). The outcomes from RcisTarget facilitate the visualization of various gene regulatory networks, elucidating the interactions among the chosen TFs (bottom middle *i*). Subsequently, the AUCell tool computes TF activities based on RcisTarget's results (middle left panel *f*), while these TF activities can be used to construct the TF-TF network (bottom left panel *h*). Finally, the framework leverages AUC values derived from AUCell to perform cell clustering (middle left panel *g*), which aids in categorizing cells based on regulatory network activities.

2.2.1 Data preprocessing and binarization:

In the preprocessing of our single-cell RNA sequencing (scRNA-seq) data, a rigorous quality control pipeline was implemented to ensure the integrity and reliability of the subsequent analyses. Initially, the raw gene expression matrix was filtered using the *Scanpy* library (*sc.pp.filter_genes*), where genes expressed in fewer than six cells were removed [35]. This step is crucial for eliminating genes with sparse expression across the dataset, which are typically considered as noise and can skew downstream analyses during ensemble model training. Following gene filtering, we conducted a cell quality assessment focused on mitochondrial gene content. Mitochondrial genes are pivotal quality indicators in scRNA-seq data; an elevated proportion of mitochondrial gene expression is often associated with apoptotic or lysed cells. Depending on the species under study (mouse or human), genes were classified as mitochondrial ('mt-' for mouse, 'MT-' for human), and cells exhibiting abnormal mitochondrial gene expression (>5% of total expression) or an excessively high gene count (>9000 genes) were excluded. This step is fundamental for removing potential outliers, such as doublets or dying cells, which can significantly affect the accuracy of the analysis.

Post quality control, the dataset underwent a normalization step using the *sc.pp.normalize_total* function from the *scanpy* package. This normalization process scales the gene expression levels in each cell to sum up to a constant total (10,000 in our case). Normalization is a critical step in scRNA-seq data processing to account for differences in sequencing depth across cells, thus enabling a more accurate comparison of gene expression profiles.

Finally, to stabilize the variance and improve the interpretability of the data, a log transformation (*sc.pp.log1p* function) was applied. This transformation mitigates the impact of highly expressed genes and compresses the wide range of expression values into a scale that enhances the performance of downstream statistical analyses. Log transformation is particularly effective in scRNA-seq data due to the inherent sparsity and skewed distribution of gene expression levels.

In the second phase of our data processing methodology, we employed the Hodges-Lehmann Estimator (HLE) for binarization of the gene expression count matrix [34]. The rationale behind using HLE lies in its robustness as a non-parametric estimator of the median, particularly beneficial in dealing with skewed distributions commonly encountered in scRNA-seq data.

Our data consists of a count matrix Z representing gene expression levels across cells. Let's denote R as the total number of cells and C as the total number of filtered genes. These genes are further categorized into F transcription factor (TF) genes and N non-transcription factor (NTF) genes. Thus, the structure of the count matrix Z is as follows in **Equation 1:**

$$Z = [Z_1 \quad Z_2] = \begin{bmatrix} z_{11} & \cdots & z_{1F} & z_{1,F+1} & \cdots & z_{1,F+N} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ z_{R1} & \cdots & z_{RF} & z_{R,F} & \cdots & z_{R,F+N} \end{bmatrix}_{R \times C} \quad (1)$$

To apply HLE for binarization, we first calculate the estimator for each pair of data points within the matrix. Specifically, for any two data points a_i and a_j, the HLE is determined as the median of the averages of all possible pairs as defined in **Equation 2:**

$$HLE = median_{i,j \in \{1,\ldots,n\}} \frac{a_i + a_j}{2} \quad (2)$$

For each column l of the count matrix Z, the threshold for binarization, denoted as D_l, is set to the HLE calculated from that column as defined in **Equation 3**:

$$D_l = median_{i,j \in \{1,\ldots,R\}} \frac{Z_{ic} + Z_{jc}}{2} \quad (3)$$

Subsequently, the count matrix Z is transformed into a binary matrix B by setting the elements in each column to l if they are equal to or exceed the threshold D_l , and to 0 otherwise. This binary transformation is mathematically represented as in **Equation 4**:

$$B = (b_{rc})_{R \times C}, b_{rc} = \mathbb{I}\{z_{rc} \geq D_l\} \quad (4)$$

This binarization step is pivotal in simplifying the gene expression data into a binary format, which is instrumental for certain downstream analyses such as gene regulatory network inference. By reducing the complexity of the data, we aim to enhance the interpretive clarity and computational efficiency of subsequent analytical processes.

The use of the Hodges-Lehmann estimator for binarization provides a non-parametric and robust threshold that is inherently less sensitive to extreme outliers and skewed distributions common in scRNA-seq data. This makes it well-suited for a wide range of datasets without requiring adjustment. For exceptionally large or noisy datasets, users may consider applying an additional variance-stabilizing transformation prior to binarization.

2.2.2 Predicting the interactions between transcription factors and genes.

In the second step, to find the relation between TF and gene, we introduce an innovative ensemble model that synergizes the capabilities of an Encoder-Decoder Classifier with a Multilayer Perceptron (MLP) to significantly enhance classification accuracy. This ensemble model is designed to exploit the complementary strengths of its constituent parts, thereby providing a more nuanced and effective approach to pattern recognition and classification tasks.

Encoder-Decoder Classifier

The Encoder-Decoder Classifier is at the heart of our model's ability to process and interpret complex input patterns. The encoder component transforms an input vector x of dimensionality input_dim into a compressed, lower-dimensional encoding h with dimensionality encoding_dim [36]. This compression is achieved through a linear transformation followed by a ReLU activation function, as stated in **Equation 5:**

$$h=\text{ReLU}\ (W_e x + b_e) \quad (5)$$

Here $W_e x,\ b_e$ represent the encoder's weights and biases, respectively. This step is critical for capturing the deep, latent representations of the input data, facilitating the model's feature extraction and dimensionality reduction capabilities. The decoder, as mentioned in **Equation 6**, then attempts to reconstruct the original input from this encoding, effectively mapping h back to the input space through another series of linear transformations and ReLU activations:

$$x'=\text{ReLU}\ (W_d h + b_d) \quad (6)$$

with $W_d h, b_d$ represent the decoder's weights and biases, respectively. This process not only aids in validating the encoder's ability to retain essential information through the encoding but also ensures that the reconstructed inputs maintain the core characteristics necessary for accurate classification.

Subsequently, the classifier takes the decoded output x' and processes it through a feedforward network to classify it into two categories. The use of ReLU activations followed by a linear transformation without a sigmoid activation layer is intentional, as we employ the *'nn.CrossEntropyLoss'* for training, which requires raw logits as defined in **Equation 7** and **Equation 8.**

$$Z = \text{ReLU}\ (W_{c1}x' + b_{c1}) \quad (7)$$

$$y = (W_{c2}z + b_{c2}) \quad (8)$$

where W_{c1} , b_{c1} , W_{c2}, and b_{c2} are the weights and biases of the classifier layers.

This component's design is pivotal for understanding complex data structures, enabling the model to perform accurate classifications based on deeply learned and reconstructed input features.

Multilayer Perceptron (MLP)

Complementing the Encoder-Decoder Classifier, the MLP serves as a straightforward yet powerful pattern recognition mechanism [37]. It directly exploits the raw input features without the intermediate encoding-decoding process, efficiently navigating the input space to classify data points. The MLP's architecture consists of three fully connected layers with ReLU activations, and is dynamically tailored to the input size, ensuring optimal complexity and preventing overfitting or underfitting, as mentioned in **Equations 9, 10, and 11**:

$$x_1 = \text{ReLU}\ (W_1 x + b_1) \quad (9)$$

$$x_2 = \text{ReLU}\ (W_2 x_1 + b_2) \quad (10)$$

$$y_{MLP} = \text{ReLU}\ (W_3 x_2 + b_3) \quad (11)$$

Ensemble Model and L1 Regularization

The essence of our ensemble model lies in its strategic combination of the outputs from both the Encoder-Decoder Classifier and the MLP. y_{mlp} represents the output from the Multilayer Perceptron, which processes one aspect of the input data, focusing on capturing nonlinear relationships and patterns. On the other hand, $y_{encoder-decoder}$ is the output from the Encoder-Decoder Classifier, which is designed to focus on differentiating and reconstructing the input features, thereby providing a robust set of features for classification or prediction.

By averaging these outputs with a preferential weighting towards the MLP (setting w=4 in Equation 12), we create a balanced, integrated response:

$$y_{ensemble} = \frac{\text{w}(y_{MLP}) + y_{encoder-decoder}}{5} \quad (12)$$

This weighting factor w offers a tunable parameter for users to adjust the influence of each model component. The value of 4 was found to be optimal for the datasets in this study, favoring the MLP's direct pattern recognition. Users may adjust this value (e.g., increasing it for noisier data or decreasing it to grant more influence to the encoder-decoder's reconstructed features) to tailor the ensemble for specific dataset characteristics.

This ensemble approach not only enriches the model's representation of the input data by blending abstract features with direct patterns but also improves decision-making robustness, effectively outperforming individual models in accuracy and generalizability.

To further enhance the model's performance and prevent overfitting, we employ L1 regularization across the model's weights as defined in **Equation 13** [38]. This technique encourages sparsity in the weight matrix, which can lead to more interpretable models and potentially better generalization on unseen data:

$$L_1 = \lambda_{L1} \sum_i |w_i| \tag{13}$$

where λ_{L1} is the regularization strength and w_i are the model weights. Integrating L_1 regularization in the training process ensures that our model remains efficient and effective. Beyond improving generalization, the induced sparsity directly supports interpretability by isolating the most relevant transcription factors with non-zero weights, thereby highlighting candidate key regulators in a biologically meaningful way.

The L1 regularization strength λ was chosen to promote sparsity and prevent overfitting. This value is a key tunable parameter; increasing λ will result in a sparser, more interpretable model at the potential cost of predictive power, while decreasing it may improve accuracy on large datasets with many true regulatory interactions.

Upon establishing a predictive model that associates all genes with a specific target gene, we delved into identifying transcription factors (TFs) within the gene pool that have a marked impact on the target gene. This investigation utilized the case-deletion method to convert the initial dataset into a format where the influence of individual TFs could be evaluated in isolation. This technique involves systematically setting the expression level of each TF to zero, thereby generating a series of modified datasets. These datasets are then used to forecast the expression

status of the target gene. This systematic case-deletion analysis provides a transparent view of model decision-making, allowing us to quantify the relative importance of each TF for predicting target gene expression.

From a mathematical perspective, we define a prediction function $f_n : D^L \rightarrow D$ that maps the space of all genes to the space of the n_{th} target gene, where $D = \{0, 1\}$, and n belongs to the set $\{1, \dots, N\}$. For every m in the set $\{1, \dots, M\}$, the m_{th} case-deleted matrix is constructed as following in **Equation 14**:

$$B_m = \begin{bmatrix} b_{11} & \cdots & b_{1F} & 0 & b_{1,m+1} & \dots b_M \dots & b_{1,L} \\ \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{R1} & \dots & b_{RF} & 0 & b_{K,m+1} & \dots b_{K,M\dots} & b_{KL} \end{bmatrix}_{K \times L} \quad (14)$$

The outcome, $f_n\ (B_m)$ represents the predicted expression of the n_{th} target gene upon exclusion of the m_{th} TF gene's effect. We assess the prediction's accuracy across the dataset using the **Equation 15**:

$$Acc_{mn} = 1 - \frac{1}{K}\ ||f_n(B_m) - \ b_{M+n}\ ||_{0,} \quad (15)$$

where b_{M+n} denotes the $(M+n)_{th}$ column of matrix B, representing the actual label of the n_{th} target gene and $||.\ ||_{0,}$ calculates the count of non-zero entries in the vector. The accuracy matrix, ACC is defined as $ACC = (ACC_{mn})_{M \times N}$. We propose that a TF's setting to zero resulting in significant accuracy fluctuations for a target gene suggests a strong influence of that TF on the gene. To identify such influential TFs with statistical significance, we apply outlier analysis through the generalized extreme Studentized deviate test (gESD test) which can detect up to r outliers by performing r Grubbs tests [39].

The hypothesis test is structured as follows:

$$H_{0:}\{\ There\ are\ no\ outliers\ in\ the\ dataset.\}$$

$$H_1:\{\textit{There are up to } r \textit{ outliers in the dataset.}\}$$

For each iteration, i_{th} calculation, the R statistic is calculated as mentioned in **Equation 16:**

$$R_i = \frac{max_i\,|x_i - \bar{x}|}{s} \tag{16}$$

where $\bar{x}, s$ are the mean and standard deviation, respectively, excluding the $i-1$ identified outliers. We then have the r test statistics $R_1, \dots, R_r$. Corresponding to the r test statistics, the critical values are determined by:

$$\lambda_i = \frac{(n-i)_{tp}(n-i-1)}{\sqrt{(n-i+1)(n-i-1+t_p^2(n-i-1))}}\,, i = 1, \dots, r \tag{17}$$

with $t_p(v)$ being the 100 p^{th} percentile of the t-distribution with v degrees of freedom and $p = 1 - \frac{a}{2(n-i-1)}$. The maximum i for which $R_i > \lambda_i$ indicates the number of outliers. In practical settings, r may vary from 30 to 80, reflecting the extent of potential regulatory interactions between feature TFs and the target gene, and α is typically set at 0.05 to highlight TF-NTF pairs for further analysis using RcisTarget trans-screening and AUCell scoring methodologies.

This section introduced an ensemble model that combines an Encoder-Decoder Classifier and a Multi-Layer Perceptron (MLP) to improve classification accuracy in identifying transcription factor (TF) and gene interactions. The Encoder-Decoder Classifier captures complex patterns within the data, while the MLP focuses on recognizing key features directly, allowing for a robust understanding of TF-gene relationships. By integrating the outputs of both models with a preference for MLP-driven outputs, the ensemble approach enhances the predictive capacity and accuracy of regulatory interactions. This model is a core component of EnsembleRegNet, enabling it to provide deeper insights into the regulatory dynamics within scRNA-seq data and supporting accurate downstream analyses in gene regulatory network construction.

2.2.3 RcisTarget utilization

In third step in our framework, we utilized RcisTarget, reference [29], to authenticate and refine the transcription factor (TF)-target gene regulons. RcisTarget operates by taking three key inputs: a motif dataset, motif annotation, and a specific gene set. With these inputs, it computes the enrichment score for a TF motif within the promoter regions of a set target gene group. This score is instrumental in assessing the influence a TF might have over its target genes. For our study, we focused on those modules that exhibited notably high enrichment scores, identifying these as regulons for subsequent, more detailed examination.

RcisTarget is a bioinformatics tool designed to identify transcription factor binding motifs (TFBS) that are over-represented in a list of genes or genomic regions. The process begins with the selection of DNA motifs that are significantly over-represented near the transcription start site (TSS) of the genes in the gene set. This selection is facilitated by a database containing genome-wide cross-species rankings for each motif.

After identifying these motifs, RcisTarget then annotates them to transcription factors (TFs). Motifs that exhibit a high Normalized Enrichment Score (NES) are particularly focused on. The final step in the RcisTarget process is the prediction of candidate target genes. These are the genes within the gene set that are ranked above a certain threshold in the context of the identified motifs.

The primary tunable parameter in this step is the Normalized Enrichment Score (NES) threshold for retaining regulons. We applied a stringent threshold to ensure high-confidence regulons. For studies aiming for a more comprehensive, exploratory network, this threshold can be lowered. Conversely, for a very high-confidence, focused network, the threshold can be increased. Thus, RcisTarget not only validates predicted TF–target links but also increases the

interpretability of EnsembleRegNet by anchoring predictions in known regulatory motifs and filtering out spurious associations.

2.2.4 AUCell Analysis

In the fourth step of our EnsembleRegNet, we employed AUCell, a computational tool designed to evaluate the activity levels of gene regulatory networks within individual cell samples. AUCell operates by taking gene sets of interest as input and subsequently transforming the expression matrix of genes in each cell into an Area Under the Curve (AUC) matrix [40]. This AUC matrix effectively quantifies the regulatory activities of each Transcription Factor (TF) within each cell. The resulting AUC matrix serves as a crucial resource for downstream analyses, enabling us to classify cells based on their regulatory profiles and conduct in-depth investigations. By leveraging the information contained within the AUC matrix, we gain insights into the intricate regulatory mechanisms operating within individual cells and across different cellular populations.

Moreover, the generated regulon list, derived from the AUCell analysis, provides valuable insights for the reconstruction and exploration of Gene Regulatory Networks (GRNs). Regulons, defined as sets of genes under the control of a common regulatory element, offer a comprehensive view of the transcriptional landscape within cellular systems. By elucidating the connections between TFs and their target genes, the regulon list facilitates the elucidation of regulatory circuits and the identification of key drivers governing cellular behavior. In summary, AUCell serves as a powerful tool in our analysis toolkit, allowing us to extract nuanced information regarding the regulatory dynamics within single cells. The resulting regulon list not only aids in the classification of cells but also lays the foundation for comprehensive investigations into the underlying regulatory architecture of biological systems.

2.2.5 Cell clustering

In the fifth step of our experimental framework, cell clustering assumes a pivotal role, facilitating the categorization of cells into distinct groups based on similarities in their gene expression patterns. These patterns, intricately influenced by transcription factor (TF) activities, offer valuable insights into the underlying regulatory dynamics within cellular systems [41].

Following the utilization of AUCell, the activity of cell regulons is condensed into a structured matrix, where columns represent individual cells and rows denote specific regulons. In the binary regulon activity matrix, each cell's regulatory activity is discretely indicated by a “1” if the corresponding regulon is active, and “0” otherwise. Conversely, the AUC activity matrix encompasses continuous AUC values for each cell-regulon pair, allowing for a nuanced representation of regulatory activity levels.

Through the application of clustering analyses to these regulon activity matrices, distinct groups of regulons with recurrent activity patterns across discrete subsets of cells are identified, thereby delineating regulatory networks within the cellular landscape. Notably, the binary activity matrix accentuates higher-order similarities among cells, effectively mitigating batch effects and technical biases. Conversely, the AUC matrix affords a more refined perspective, enabling the discernment of subtle variations in regulatory activity.

For visualization purposes, *t*-distributed Stochastic Neighbor Embedding (t-SNE) techniques are predominantly employed. Rigorous testing of t-SNE results across various perplexity values and distance metrics ensures the robustness of our analyses. It is noteworthy that t-SNE analyses are conducted on complete regulon activity matrices, affording comprehensive insights into the intricate regulatory dynamics governing cellular behavior [42].

2.2.6 Network visualization

The visualization of gene regulatory networks (GRNs) poses a significant challenge, given the complexity and inherent noise stemming from numerous interactions within the cellular environment [43]. In the final step of our proposed framework, we endeavor to visualize the gene regulatory network in two distinct forms. The first form comprises a comprehensive gene regulatory network, elucidating the intricate web of connections between various transcription factors (TFs) and their corresponding target genes. This expansive network encapsulates a wealth of detail, mapping the multifaceted interactions among numerous TFs and their respective targets. However, its extensive nature often yields a dense and noisy network, which can prove overwhelming and challenging to interpret owing to the sheer volume of data it encompasses.

To resolve this issue and isolate the essential elements of this intricate system, a specialized network called *the transcription factor gene regulatory network* is designed. This refined network is crucial for its ability to succinctly capture the essence of the gene regulatory network (GRN), emphasizing the critical regulatory functions of transcription factors (TFs). It streamlines the GRN by focusing solely on the direct interactions between TFs, excluding indirect interactions, interactions between TFs and their target genes, and other less significant connections. Through this process, the TF gene regulatory network efficiently removes unnecessary information, spotlighting the key pathways and regulatory mechanisms fundamental to the dynamics of gene expression.

This distilled network accentuates the most influential TFs, positioning them as key regulators in the genetic circuitry, and reveals how they orchestrate gene expression patterns that drive cellular processes. By concentrating on these salient connections, the TF gene regulatory

network serves as an essential tool for researchers, offering a more tractable and less cluttered visualization of cellular regulation. It acts as a refined lens, through which the fundamental architecture and regulatory hierarchies within the cell are brought to light, providing insights into the orchestration of biological functions and the maintenance of cellular homeostasis [44].

2.2.7 Computational Environment and Reproducibility

All experiments were performed on a standardized workstation equipped with an Intel Core i9-14900HX processor (24 cores, 32 logical threads, 2.20 GHz), 32 GB RAM, and an NVIDIA GPU with 8 GB dedicated memory, running Microsoft Windows 11 Home.

The software environment was based on Python 3.8.18 with the following key packages and versions: PyTorch 1.13.1+cu117, torchvision 0.14.1+cu117, torchaudio 0.13.1, Scanpy 1.9.6, scikit-learn 1.3.2, SciPy 1.10.1, NumPy 1.23.4, pandas 2.0.3, pySCENIC 0.11.2, matplotlib 3.5.1, seaborn 0.12.2, openpyxl 3.0.9, cytoolz 0.11.0, along with python-igraph, louvain (conda-forge), and multicore-tSNE.

To ensure reproducibility, we fixed the random seed to 42 and limited CPU threads to 24. Reported runtimes (**Table 4**) reflect single-run executions under identical conditions, thereby allowing fair comparisons between EnsembleRegNet, GENIE3, and PIDC. This setup balances practical scalability with transparency, enabling replication of results across independent computing environments.

2.3 Cluster Validation

In the context of cluster analysis, validation methodologies are essential for verifying the efficacy of clustering outcomes. These methodologies are generally bifurcated into two distinct categories: internal and external validation approaches [45]. Internal validation relies exclusively on the intrinsic properties of the dataset, devoid of any external reference points or predetermined labels [46]. Conversely, external validation techniques utilize a predefined set of labels or benchmarks as a standard for evaluating the clustering's precision [47]. In the spectrum of validation techniques, this study has elected to employ a cadre of indices that are both well-established within the literature and encompass a spectrum of both internal and external validation metrics, facilitating a holistic assessment of clustering validity. Regarding internal validation, the indices are chosen based on their ability to encapsulate two fundamental dimensions of clustering quality: compactness within clusters and separation between them. An optimal clustering solution is one where intra-cluster compactness is minimized while inter-cluster separation is maximized. To quantify these attributes, the following metrics have been selected for evaluation:

1- Root-mean-square standard deviation (RMSSTD): is a metric used to assess the compactness of clusters in a dataset. A lower RMSSTD value suggests that data points within each cluster are closer to their cluster's centroid, indicating a potentially higher quality of clustering [48]. The calculation of RMSSTD involves summing the squared distances between data points and their respective cluster centroids and then taking the square root of this sum after normalization with certain factors. Let's dissect this further:

c_i represents the centroid of the cluster i.

n_i denotes the number of data points in the cluster i.

P is the total number of data points in the dataset.

The RMSSTD is calculated using the **Equation 18**:

$$\text{RMSSTD} = \sqrt{\frac{\sum_i \sum_{x \in c_i} \|x - c_i\|^2}{P \sum_i (n_i - 1)}} \tag{18}$$

RMSSTD quantifies the average distance of points from their cluster centroids, adjusted for cluster size and total data points. Smaller RMSSTD values imply that the clusters are more tightly grouped, which is usually desirable in cluster analysis.

2-The Modified Hubert Γ statistic is a measure used to assess the quality of clustering by examining the dissimilarity between pairs of objects in two different clusters [46]. Here is a breakdown of the components of the Modified Hubert Γ statistic:

$$\Gamma = \frac{2}{n(n-1)} \sum_{x \in D} \sum_{y \in D} d(x, y) \,.\, d_{x \in C_i \,, y \in C_j} (c_i, c_j) \tag{19}$$

Where:

- d is a distance measure between two data points.
- D is the dataset.
- C_i and C_j are clusters.
- c_i and c_i are the centroids of the clusters C_i and C_j respectively.
- n is the number of data points.

The term $d_{x \in C_i \,, y \in C_j} (c_i, c_j)$ represents the distance between centroids of the clusters to which x and y belong, respectively. This equation calculates a weighted sum of pairwise distances between data points, with the weights being the distances between the centroids of the clusters to which the data points belong. The resulting statistic, Γ, indicates the clustering quality. A high Γ value suggests that the clusters are compact and well-separated, indicating good clustering.

Conversely, a low Γ value would suggest that the clusters are scattered and overlapping, indicating poorer clustering quality.

3-The Silhouette index, or Silhouette score, is a metric for assessing the efficacy of a clustering algorithm. It is calculated for each object in the dataset and takes on values between -1 and 1. A high Silhouette score indicates that the object is well placed within its own cluster and distant from objects in neighboring clusters [50], as mentioned in **Equation 20**.

$$S = \frac{1}{NC} \sum_{i} \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x) - a\,(x)}{\max[\, b(x), a\,(x)]} \tag{20}$$

where:

- (a) is the average distance from the i^{th} data point to the other data points in the same cluster.
- $b(x)$ is the smallest average distance from the i^{th} data point to data points in a different cluster, minimized over clusters.
- C_i is the i^{th} cluster.
- ni is the number of data points in cluster C_i .
- NC is the number of clusters.

A high average Silhouette score across all objects suggests that the clustering structure makes sense, while a low or negative average score could indicate that the clustering has either too many or too few clusters. Therefore, the Silhouette score provides an objective measure to determine the appropriateness of the number of clusters used.

4-The R-squared index, often denoted as $\mathcal{R}^2$**,** is a statistical measure that quantifies the proportion of the variance in a dataset that is accounted for by the clustering algorithm's model [51][3]. Here's a detailed breakdown of the elements within the R-squared formula as defined in **Equation 21**:

$$R^2 = \frac{\sum_{x \in D} ||x - c||^2 - \sum_i \sum_{i \in C_i} ||x - c_i||^2}{\sum_{x \in D} ||x - c||^2} \quad (21)$$

where:

- D is the dataset.
- C_i is the i^{th} cluster.
- c is the centroid of the entire dataset D.
- c_i is the centroid of the i^{th} cluster.

The $\mathcal{R}^2$ index essentially compares the total variance of the dataset to the amount of variance explained by the clustering. A high $\mathcal{R}^2$ value indicates that the clustering has captured a large proportion of the dataset's variance, suggesting a good fit between the data points and the clustering model.

In contrast to the RMSSTD, which assesses the tightness or compactness of individual clusters by looking at how close the points in each cluster are to their centroid, the $\mathcal{R}^2$ index offers a more global perspective. It assesses how much of the data's overall variance is explained by the differences between clusters, thus providing insight into the overall clustering structure relative to the dataset's inherent variability.

For external validation, the chosen methodology capitalizes on a benchmark or a set of predefined labels to objectively measure the clustering algorithm's performance relative to a known ground truth. This approach addresses the correspondence between the induced clustering and an external criterion. The metrics adopted for external validation in this study include the following 1 to 9:

5-The Adjusted Rand Index (ARI) serves as a refined metric for assessing the congruence between two data clusters: one representing the ground truth and the other derived from a clustering algorithm's output [52]. As a normalized index, the ARI mitigates the influence of chance agreement, thus offering a more rigorous evaluation of clustering results compared to the unadjusted Rand Index. Values of ARI range from -1 to 1, with a score of 1 signifying an exact match between the clustering and the ground truth, a score near 0 indicating no better agreement than would be expected by random chance, and a negative score reflecting a discordance worse than random. The calculation of the ARI is formalized as **Equation 22:**

$$\text{ARI}=\frac{RI\ -E[RI]}{1-\ E(RI)} \tag{22}$$

Here, E[RI] is the expected value under random conditions of the Rand Index, a measure of similarity between two cluster assignments, and the denominator normalizes the index to account for the maximum possible Rand Index value, which is 1. The adjustment for chance in the computation of ARI ensures a more accurate depiction of an algorithm's clustering performance, particularly when compared to random clustering, thus providing critical insights into the algorithm's effectiveness in revealing the intrinsic groupings within the data.

6-The Normalized Van Dongen (NVD) metric is a quantitative measure designed to evaluate the degree of dissimilarity between two distinct clusters, one representing the ground truth and the other derived from a clustering algorithm's output [4]. The essence of this metric is to capture the extent of overlap between clusters from two separate clusters labeled X and Y. It achieves this by identifying the maximal intersections—namely, the largest subsets of data points shared between clusters across the two partitions. The NVD then executes a normalization procedure, where the

sum of these maximal intersections is divided by the total number of data points within the dataset, denoted by N. This normalization serves to scale the metric, facilitating comparisons regardless of the dataset's size. Mathematically, the NVD is represented as mentioned in **Equation 23**:

$$\text{NVD (X, Y)} = \frac{2N - \sum_{i=1}^{k} max_i\, n_{ij} + \sum_{j=1}^{l} max_j\, n_{ij}}{2N} \quad (23)$$

In this formula:

- $k\ and\ l$ denotes the number of clusters within clusters X and Y, respectively.
- N signifies the aggregate number of data points in the dataset.
- n_{ij} corresponds to the number of data points that are in the i-th cluster of X and the j-th cluster of Y.

The NVD metric's value ranges between 0 and 1, where a lower score indicates a higher similarity between the two clusters and a higher score indicates greater dissimilarity. The dual summation reflects the consideration of each cluster's contribution to the overall partitioning, thus offering a comprehensive assessment of the partitioning's quality. This measure is particularly useful when comparing the output of different clustering algorithms or tracking the consistency of a single algorithm's performance across varying parameters or datasets.

7-The F-score metric is a harmonizing of concepts of precision and recall within its calculation. Specifically, within the context of clustering, the F-score is computed by considering each cluster relative to every corresponding cluster in the ground truth [53]. The process involves selecting the highest F-score from these comparisons to stand as the representative score for each cluster. The overall F-score for the clustering outcome is then determined through an aggregation of these individual scores, weighted by the size of each cluster in the solution set. This is formally expressed as in **Equation 24**:

$$F = \frac{1}{N}\sum_{i=1}^{k} n_i \max \frac{2\, n_{ij}}{n_i + m_j} \quad (24)$$

where:

- n_i is the number of items in cluster c_i .
- m_j is the number of items in cluster g_j .
- n_{ij} is the number of common items between c_i and g_j

The overall F-score for the clustering is thus an aggregate measure, obtained by weighting the individual maximum F-scores of each cluster by the number of items it contains, normalized by the total number of items N across all clusters. This comprehensive score provides insight into the clustering algorithm's accuracy, reflecting its ability to not only include most items of a true cluster but also to exclude non-members effectively.

8-The Pair Sets Index (PSI) serves as a valuable metric for comparing the similarity between two clusters within a dataset, one representing the ground truth and the other derived from a clustering algorithm's output, offering crucial insights in cluster analysis [52]. PSI focuses on examining pairs of data points and assesses whether they are grouped together or separated into distinct clusters. The computation of PSI involves the summation of fractions representing the proportion of items in each cluster of the first partitioning that also belong to a cluster in the second partitioning. This computation is expressed as in **Equations 25, 26, 27, and 28:**

$$\begin{Bmatrix} \frac{S - E(s)}{K - E(s)} & , \quad S \geq E(S)\,,\ K > 1 \\ 0 & , \quad S < E(S), \end{Bmatrix} \quad (25)$$

$$S = \sum_{i=1}^{K} \frac{n_{ij}}{\max(n_i, m_j)} \quad (26)$$

$$e_{ij} = \sum_{i=1}^{K} \frac{n_i \times m_j}{N} \quad (27)$$

$$\mathrm{E(S)} = \sum_{i=1}^{K} \frac{e_{ij}}{\max(n_i, m_j)} \quad (28)$$

Where:

- n_{ij} denotes the number of items common to cluster i and cluster j, n_i is the count of items in cluster i, and m_j is the count of items in cluster j.
- S represents the sum of the overlaps n_{ij} normalized by the maximum count of items in each cluster (n_i and m_j respectively).
- e_{ij} denotes the expected overlap between clusters i and j, calculated based on the counts of items in each cluster and the total number of items in the dataset.
- $E(S)$ represents the expected value of S, calculated by normalizing e_{ij}.

The normalization process ensures that PSI accounts for the expected overlap between two independent clusters, adjusting for the size and distribution of clusters within the dataset. By normalizing PSI with $E(S)$, the resulting metric offers a meaningful comparison of clustering similarity, reflecting the resemblance of the clustering structures rather than being solely influenced by cluster sizes. This normalization enhances the interpretability and reliability of PSI as a measure of clustering similarity.

9-Normalized Mutual Information (NMI) is an insightful measure for comparing the similarity between two clusters of the same data set, one of which represents the ground truth and the other is derived from a clustering algorithm's output, embodying a normalization of the mutual information (MI) score. It essentially quantifies the amount of shared information between two

clustering assignments, P and G, where P might represent the predicted clustering and G the ground truth.

$$NMI = \frac{MI\ (P\ , G)}{\frac{1}{2}\big(H(P) + H(G)\big)} \tag{29}$$

In this context, $MI\ (P, G)$ represents the mutual information between the two clustering schemes, which measures the mutual dependence between them. $H(P)$ and $H(G)$ are the entropies of P and G respectively, signifying the amount of uncertainty or randomness in the individual clustering assignments [53].

By normalizing the mutual information by the average of the entropies of the two clusters, NMI adjusts for the size of the clusters and allows for a comparison that is not biased by the number of clusters or the number of data points within them. The NMI value ranges from 0 to 1, where 0 indicates no mutual information (completely independent clusters), and 1 denotes perfect correlation between the two clustering assignments. Thus, NMI is particularly useful for validating clustering algorithms against a known ground truth by offering a standardized measure of the quality of the clustering.

3. Results

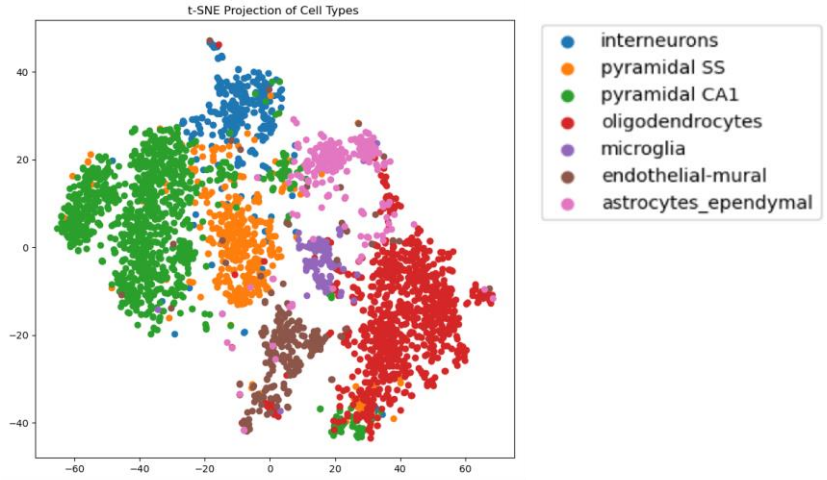
In this section, we present the results of a comparative analysis of gene regulatory network (GRN) inference methods, specifically evaluating the performance of EnsembleRegNet in comparison to SIGNET and SCENIC across multiple datasets. We assess the effectiveness of these methods using a range of external and internal clustering validation metrics to provide a comprehensive

evaluation of their clustering accuracy and consistency. By conducting detailed comparisons, we highlight the strengths and weaknesses of each approach, with a particular focus on EnsembleRegNet's ability to capture the underlying structure of single-cell RNA sequencing (scRNA-seq) data. Additionally, we visualize the GRNs and transcription factor interactions, offering deeper insights into the regulatory dynamics within the datasets. This analysis contributes to a better understanding of GRN inference methods and demonstrates the robustness of EnsembleRegNet as a powerful tool for biological data analysis.

3.1 Analytical Comparisons with Existing Methods.

To assess the effectiveness of EnsembleRegNet, we replicated the SIGNET and SCENIC methods across three datasets and employed both external and internal cluster validation techniques to compare the clustering outcomes derived from each method.

(a) EnsembleRegNet



(b) SIGNET

(c) SCENIC

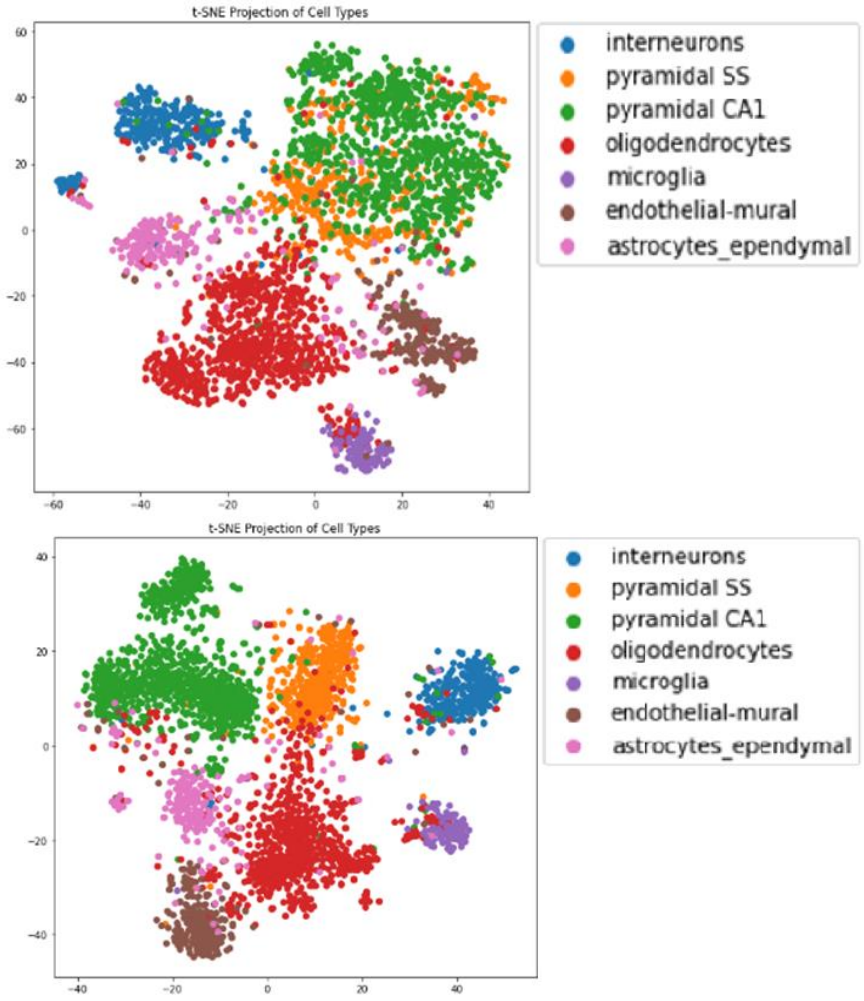


Figure 2: Comparative t-SNE plots displaying cell types from mouse cerebral cortex scRNA-seq dataset (GSE60361). These plots illustrate the effectiveness of different clustering algorithms: (a) EnsembleRegNet, (b) SIGNET, and (c) SCENIC. The figure highlights the superior intra-cluster compactness and inter-cluster separation achieved by EnsembleRegNet, indicating a more definitive categorization of cell types when contrasted with the other methodologies.

Table 1: Comparison of External and Internal Clustering Validation Metrics for EnsembleRegNet, SIGNET, and SCENIC on the GSE60361 Dataset.

Cluster Validation	**Measure**	**EnsembleRegNet**	**SIGNET**	**SCENIC**
External clustering validation	The Adjusted Rand Index (ARI)	**0.544**	0.469	0.509
	Normalized Mutual Information (NMI)	**0.592**	0.554	0.500
	F-score	**0.139**	0.151	0.026
	Normalized Van Dongen (NVD) score	**0.236**	0.300	0.240
	Pair Sets Index (PSI) score	**0.725**	0.680	0.699
Internal clustering validation	RMSSTD	0.227	0.258	**0.186**
	Modified Hubert Γ statistic	5362.284	5715.164	**3186.7275**
	Silhouette index	**0.416**	0.438	0.451
	R-squared index	**0.899**	0.873	0.882

The best score values are **bold-faced.**

Table 1 shows a comparative evaluation of external and internal clustering validation metrics for EnsembleRegNet, SIGNET, and SCENIC on the GSE60361 dataset, providing insights into each algorithm's accuracy in data grouping and the intrinsic quality of the resulting clusters

For external clustering validation, EnsembleRegNet consistently surpasses both SIGNET and SCENIC, as detailed in Table 1. It achieves the highest Adjusted Rand Index (ARI) of 0.544, reflecting strong alignment with the true data labels, followed by SCENIC at 0.509 and SIGNET at 0.469. EnsembleRegNet also leads in Normalized Mutual Information (NMI) with a score of 0.592, indicating its superior ability to capture the data's intrinsic structure compared to the other algorithms. Additionally, EnsembleRegNet attains the highest Pair Sets Index (PSI) score at 0.725, underscoring its robustness in clustering accuracy relative to SCENIC and SIGNET.

The Normalized Van Dongen (NVD) score further emphasizes EnsembleRegNet's clustering precision, with a lower score of 0.236, signaling closer alignment with the actual data classifications than both SIGNET's 0.300 and SCENIC's 0.240. While EnsembleRegNet's F-score is slightly lower than SIGNET's (0.139 vs. 0.151), its consistently high performance across ARI, NMI, PSI, and NVD scores demonstrates its effectiveness in producing accurate clusters that closely reflect the true structure within the dataset.

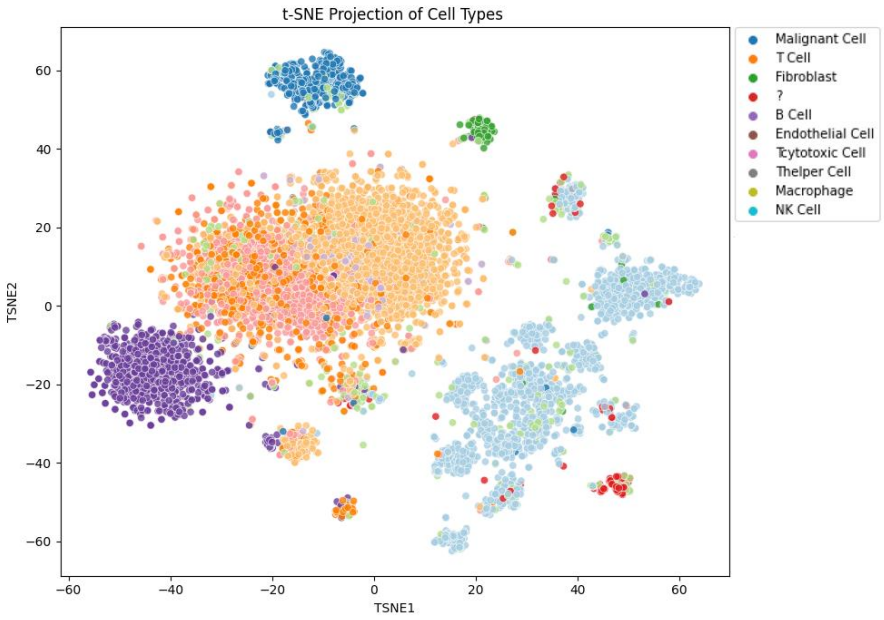
For internal clustering validation, EnsembleRegNet continues to exhibit strong performance in forming cohesive clusters. It achieves an RMSSTD score of 0.227, indicating tight clusters that are more cohesive than those by SIGNET, though not as compact as those by SCENIC. EnsembleRegNet also leads in the Modified Hubert Γ statistic, with a value of 5362.284, highlighting its ability to generate well-defined clusters.

Although SCENIC has a slightly higher Silhouette index of 0.451 compared to EnsembleRegNet's 0.416, which suggests a marginally better cluster separation, EnsembleRegNet reaches the highest R-squared index of 0.899. This value indicates that EnsembleRegNet's clustering captures the dataset's variance more effectively than SIGNET's 0.873 or SCENIC's 0.882.

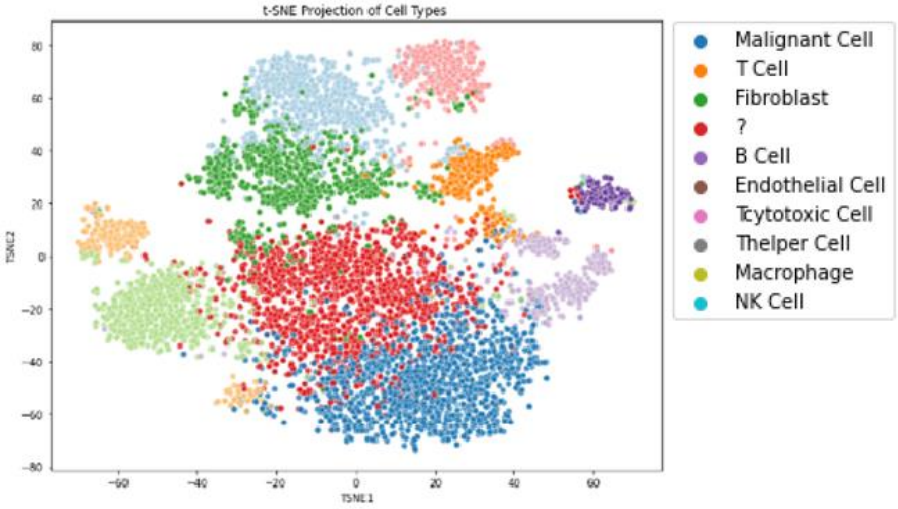
In summary, EnsembleRegNet outperforms both SIGNET and SCENIC, showing superior accuracy in clustering and internal structure. Its consistently strong results across external and internal validation metrics confirm its capability to create well-defined, accurate clusters, positioning it as the most effective algorithm for the GSE60361 dataset.

(a)

EnsembleRegNet



(b) SIGNET



(c) SCENIC

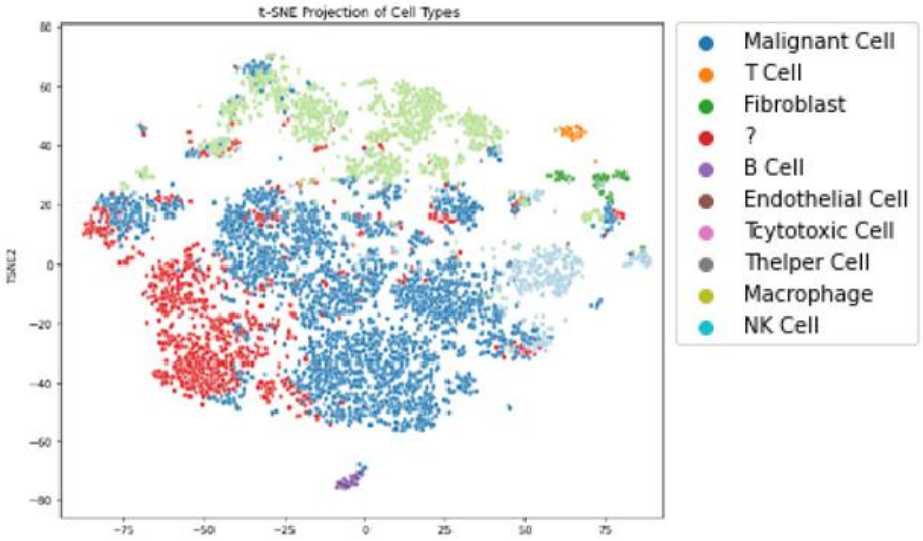


Figure 3: Comparative t-SNE plots displaying cell types from the human skin cutaneous melanoma single-cell RNA sequencing (scRNA-seq) data (GSE115978). These plots demonstrate the clustering algorithms' effectiveness when applied to TF activities, calculated using (a) EnsembleRegNet, (b) SIGNET, and (c) SCENIC. The visualizations emphasize the distinction in cluster density and separation between clusters, particularly noting EnsembleRegNet's enhanced performance in creating tight, well-delineated clusters that suggest a clearer classification of cell types compared to the other methods.

Table *2*: Comparison of External and Internal Clustering Validation Metrics for EnsembleRegNet, SIGNET, and SCENIC on the GSE115978 Dataset.

	Measure	**EnsembleRegNet**	**SIGNET**	**SCENIC**
External clustering validation	The Adjusted Rand Index (ARI)	**0.378**	0.0121	0.007
	Normalized Mutual Information (NMI)	**0.524**	0.024	0.019
	F-score	**0.075**	0.095	0.0189
	Normalized Van Dongen (NVD) score	**0.364**	0.754	0.893
	Pair Sets Index (PSI) score	**0.684**	0.511	0.507

Internal clustering validation	RMSSTD	**0.126**	0.175	0.186
	Modified Hubert Γ statistic	5594.133	9383.478	**9553.364**
	Silhouette index	**0.454**	0.412	0.417
	R-squared index	**0.924**	0.9123	0.903

The best score values are **bold-faced.**

Table 2 provides an evaluative comparison of external and internal clustering validation metrics for EnsembleRegNet, SIGNET, and SCENIC on the GSE115978 dataset, revealing insights into the algorithms' data grouping accuracy and the intrinsic qualities of these clusters.

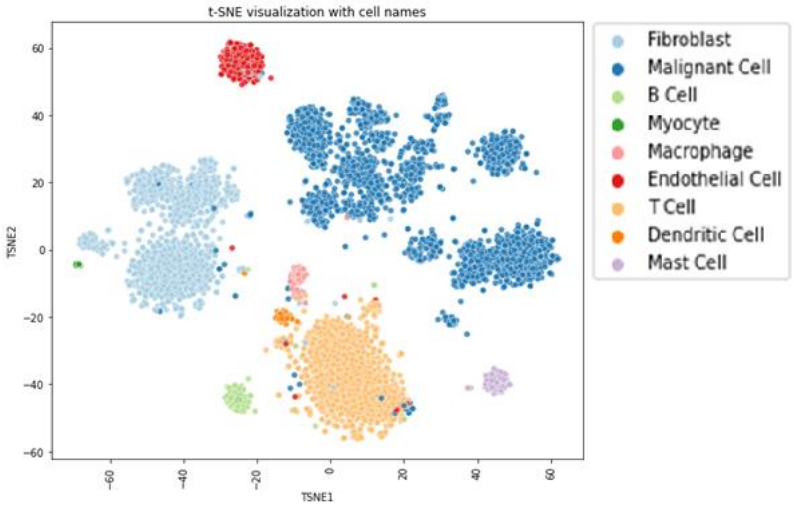
In external clustering validation, EnsembleRegNet consistently outperforms SIGNET and SCENIC. With an Adjusted Rand Index (ARI) of 0.378 and Normalized Mutual Information (NMI) of 0.524, EnsembleRegNet demonstrates a stronger alignment with true cluster labels, outperforming SIGNET and SCENIC, which show much lower values in these metrics. EnsembleRegNet also leads in the F-score, underscoring its relative accuracy in clustering. Although none of the algorithms achieve ideal results in the Normalized Van Dongen (NVD) score, EnsembleRegNet's score of 0.364 is notably lower than the higher misalignment seen in SCENIC (0.893), indicating EnsembleRegNet's improved conformity to the actual data classification. EnsembleRegNet also achieves the highest Pair Sets Index (PSI) score at 0.684, suggesting more accurate cluster formations compared to SIGNET and SCENIC.

Internal clustering validation further emphasizes EnsembleRegNet's superior clustering structure. It shows the lowest RMSSTD value (0.126), indicating the tightest and most cohesive clusters. EnsembleRegNet also has a slight edge in the Silhouette index, demonstrating better separation between clusters. Its high R-squared index of 0.924 highlights its effective alignment with the dataset's natural variance, marking EnsembleRegNet as the strongest performer in capturing the dataset's intrinsic structure. Though SIGNET and SCENIC show good internal

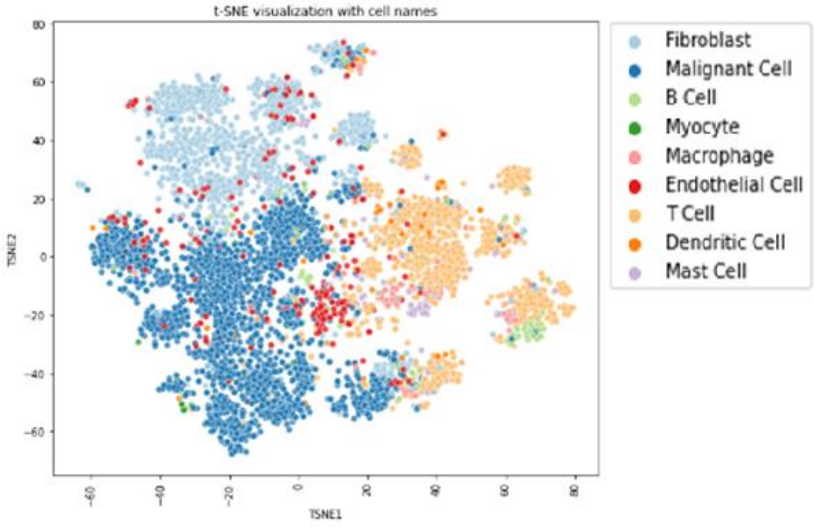
structure through high Modified Hubert Γ statistics, they do not match EnsembleRegNet's overall cohesiveness and clarity in cluster separation.

In summary, EnsembleRegNet outperforms both SIGNET and SCENIC, demonstrating superior clustering accuracy and cohesion. Its strong results in both external and internal validation metrics underscore its effectiveness in creating well-defined, meaningful clusters that best represent the dataset's structure.

(a) EnsembleRegNet



(b) SIGNET



(c) SCENIC

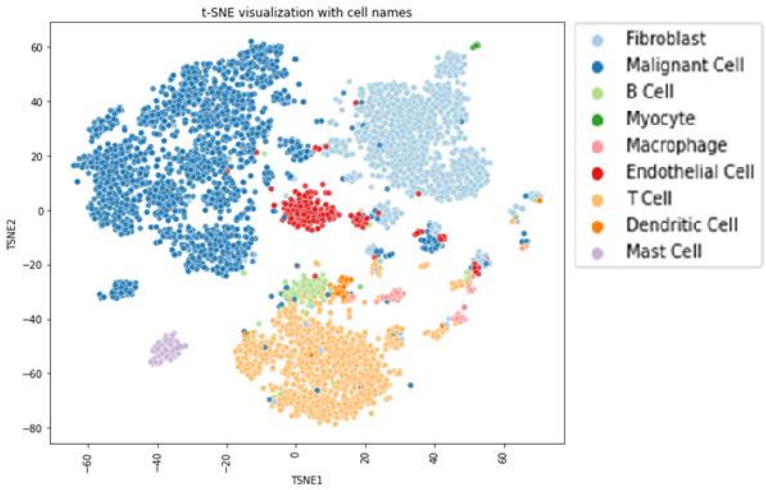


Figure 4: Comparative t-SNE visualizations illustrating cell type clustering within the human head and neck squamous cell carcinoma scRNA-seq data (GSE103322). These plots demonstrate the clustering performances of distinct computational strategies: (a) EnsembleRegNet, (b) SIGNET, and (c) SCENIC. The graphical representation emphasizes the enhanced cluster density and the distinction between clusters facilitated by EnsembleRegNet, signifying its heightened ability to categorize cell types with greater precision than the alternative approaches.

Table 3: Comparison of External and Internal Clustering Validation Metrics for EnsembleRegNet, SIGNET (MLP), and SCENIC (RF) on the GSE103322 Dataset.

Cluster validation	Measure	EnsembleRegNet	SIGNET	SCENIC
External clustering validation	The Adjusted Rand Index (ARI)	**0.466**	0.011	0.261
	Normalized Mutual Information (NMI)	**0.685**	0.023	0.581
	F-score	0.060	0.076	**0.084**
	Normalized Van Dongen (NVD) score	**0.246**	0.755	0.342
	Pair Sets Index (PSI) score	**0.629**	0.511	0.422
Internal clustering validation	RMSSTD	**0.153**	0.174	0.194
	Modified Hubert Γ statistic	6878.118	**9315.471**	6248.413
	Silhouette index	**0.484**	0.311	0.335
	R-squared index	**0.925**	0.812	0.825

The best score values are **bold-faced.**

Table 3 presents a comparative analysis of external and internal clustering validation metrics for EnsembleRegNet, SIGNET, and SCENIC on the GSE103322 dataset, highlighting EnsembleRegNet's superior performance across most metrics.

For external clustering validation, EnsembleRegNet surpasses both SIGNET and SCENIC, demonstrating a markedly stronger alignment with true cluster labels. With an Adjusted Rand

Index (ARI) of 0.466—nearly double SCENIC's 0.261 and an extraordinary 42 times greater than SIGNET's 0.011—it sets a new benchmark for accurately matching true data labels. EnsembleRegNet also excels in Normalized Mutual Information (NMI), achieving a score of 0.685—well above SCENIC's 0.581 and far surpassing SIGNET's 0.023. EnsembleRegNet again leads in the Normalized Van Dongen (NVD) score with a superior 0.246, reflecting significantly lower misalignment with true classifications compared to SCENIC's 0.342 and SIGNET's high misalignment at 0.755. Furthermore, EnsembleRegNet achieves the highest Pair Sets Index (PSI) score of 0.629, underscoring its unmatched clustering accuracy over both SCENIC and SIGNET.

For internal clustering validation, EnsembleRegNet again leads in quality metrics, achieving the lowest RMSSTD at 0.153, indicating the most cohesive clusters among the three methods. Its Silhouette index of 0.484 suggests well-separated clusters, surpassing SIGNET's 0.311 and SCENIC's 0.335. EnsembleRegNet also attains the highest R-squared index at 0.925, showing its strong alignment with the dataset's natural variance. While SIGNET achieves a high Modified Hubert Γ statistic of 9315.471, indicating solid internal structure, EnsembleRegNet's overall cluster cohesiveness and separation provide a more robust clustering outcome.

In summary, EnsembleRegNet outperforms both SIGNET and SCENIC on the GSE103322 dataset, demonstrating superior clustering accuracy and internal consistency. Its consistently high scores across both external and internal validation metrics confirm its effectiveness in producing well-defined, meaningful clusters that closely align with the dataset's inherent structure.

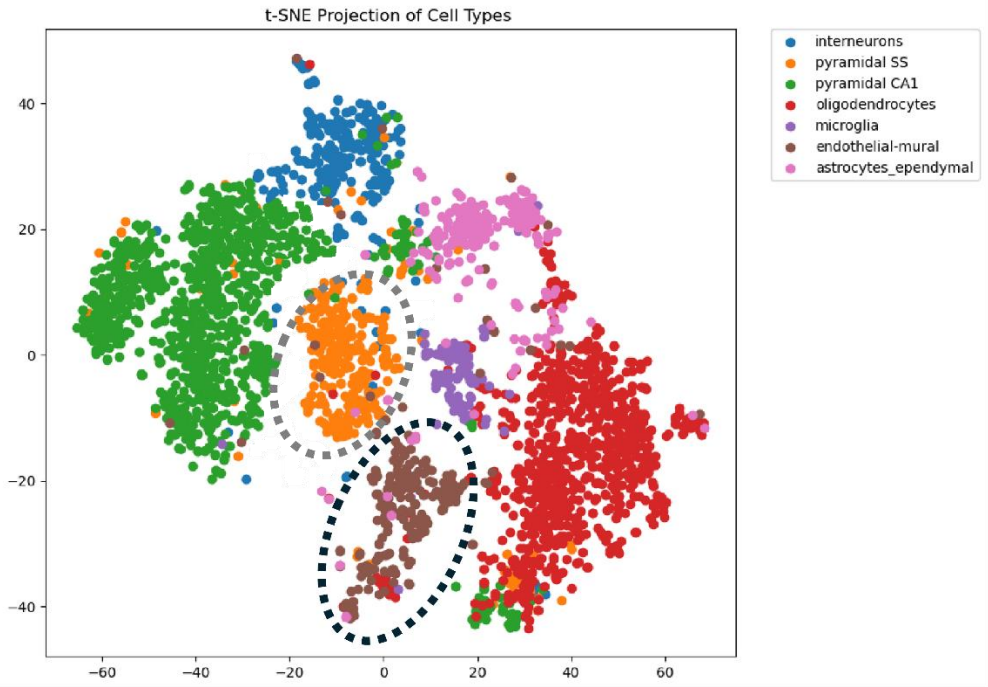


Figure (5): t-SNE visualization of a mouse cell dataset highlighting the distinct separation between the endothelial-mural cluster (in brown) and the pyramidal SS cluster (in orange), with the delineation emphasized by Z-score normalization.

Further analysis depicted in **Figure 5** shows a scatter plot that underscores EnsembleRegNet's capability in separating data point clusters, each symbolizing different cell types in the GSE60361 dataset. The plot emphasizes the segregation between the endothelial-mural cluster (in brown) and the pyramidal SS cluster (in orange), visually evaluating clustering effectiveness. The separation between these clusters was quantified using z-score analysis for each cluster.

The z-score quantifies the dispersion of data points by measuring the distance of each point from the cluster's mean, expressed in units of standard deviation. For the endothelial-mural and pyramidal SS clusters, we computed z-scores for every point to determine the mean and standard deviation of these scores within each cluster. Such statistical computations yield insights into the cohesiveness and extent of spread among the data points.

Our analysis revealed that the interneurons cluster exhibits a low standard deviation in z-scores, suggesting a tight clustering with data points situated in close proximity to one another. This tight grouping denotes a high degree of similarity among the endothelial-mural, manifesting as a concentrated cluster on the scatter plot. In contrast, a higher standard deviation of z-scores within the pyramidal SS cluster indicates a greater level of dispersion among its data points. This suggests that the pyramidal SS cluster is more spread out, potentially alluding to the presence of subclusters or a wider range of variance within this grouping.

The scatter plot itself, with its distinct color-coding and clear boundary delineations, provides a visual corroboration of the statistical analysis. It highlights the compact nature of the endothelial-mural cluster in contrast to the more diffuse spread of the pyramidal SS cluster. Dashed lines encircling particular clusters serve to draw attention to these regions, underscoring areas where the integrity and distinctness of the clusters are of particular interest. This visual and statistical synergy in **Figure 5** confirms EnsembleRegNet's capability in achieving discernible separation between varied cell type clusters.

3.2 Comparative Benchmarking with GENIE3 and PIDC Using Experimental Gold Standards

To further strengthen validation, we benchmarked EnsembleRegNet against two additional established GRN inference methods, GENIE3 [54] and PIDC [55], on the widely used 10x Genomics PBMC 10k dataset. For evaluation, we constructed an expanded gold standard by integrating TRRUST v2 (curated TF–target interactions) [56] with DoRothEA regulons [57], thereby capturing experimentally supported regulatory relationships with broad coverage.

Performance was assessed using Macro-AUROC and Macro-AUPRC. These metrics extend the standard AUROC/AUPR by computing the score for each transcription factor (TF) individually and then averaging across all TFs. This macro-averaging strategy ensures that each TF contributes equally to the final performance score, regardless of how many validated targets are associated with it. This distinction is critical because curated gold standards such as TRRUST and DoRothEA exhibit high variability in regulon size: a few regulators (e.g., *STAT3, NFKB1*) have hundreds of known targets, while many others have only a handful. If global AUROC/AUPR were used, large regulons would dominate the metric, potentially obscuring whether the method performs well on smaller but biologically important TFs. By contrast, Macro-AUROC/Macro-AUPRC provides a stricter and fairer evaluation of generalization across the entire TF set, preventing bias toward hub regulators.

In addition to predictive accuracy, we assessed computational efficiency to address scalability concerns. All benchmarking experiments were conducted under the standardized environment described in Section 2.2.7. As summarized in **Table 4**, EnsembleRegNet achieved a Macro-AUROC of 0.687 and Macro-AUPRC of 0.0116, outperforming GENIE3 (AUROC 0.588, AUPRC 0.0109) and showing competitive performance relative to PIDC (AUROC 0.691, AUPRC 0.0116). Importantly, EnsembleRegNet completed the inference in 447 minutes, compared to 545 minutes for GENIE3 and 661 minutes for PIDC. These results demonstrate that EnsembleRegNet not only excels in clustering performance and recovery of validated regulatory edges but also offers improved scalability and runtime efficiency when applied to large-scale scRNA-seq datasets.

Table 4. Benchmarking EnsembleRegNet against GENIE3 and PIDC on the PBMC 10k dataset using TRRUST and DoRothEA as gold standard**.** Performance is reported as Macro-AUROC, Macro-AUPRC, and runtime (in minutes). All experiments were performed on an Intel Core i9 (2.20 GHz, 32 GB RAM, 8 GB GPU). EnsembleRegNet achieves superior accuracy compared to GENIE3, is competitive with PIDC, and demonstrates lower computational cost than both alternatives.

Method	Macro_AUROC	Macro_AUPRC	Runtime (min)

EnsembleRegNet	0.687	0.0116	447
GENIE3	0.588	0.0109	545
PIDC	0.691	0.0126	661

3.3 Visualization of gene regulatory network

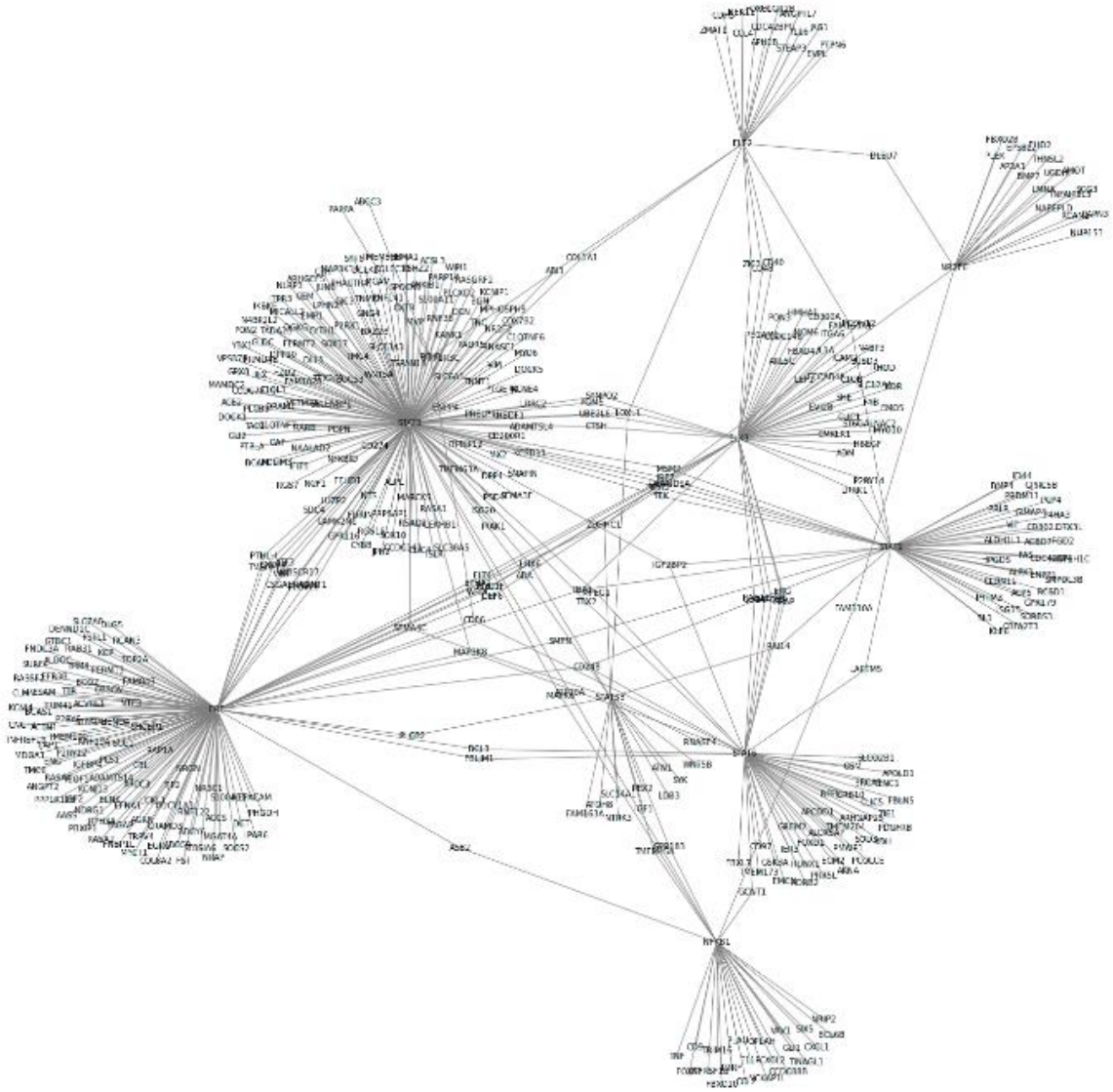
Inferring gene regulatory networks is pivotal for pushing the boundaries of our understanding of biology, enhancing strategies for disease treatment and prevention, and catalyzing the development of innovative technologies within the fields of biotechnology and medicine. To this end, our framework is designed to visualize two specific types of gene regulatory networks, as illustrated in **Figures 6(a)** and **6(b)**, providing a comprehensive tool for researchers and practitioners alike.

In Figure 6(a), the framework showcases a detailed gene regulatory network associated with GSE60361. This visualization captures the intricate relationships between transcription factors (TFs) and their target genes and maps the interactions among the TFs themselves. The depiction of these networks is exhaustive, offering an extensive view of the genetic interactions that underpin regulatory processes. However, the richness of information in this type of network introduces complexity and potential noise, which can obscure critical patterns and relationships essential for understanding gene regulation.

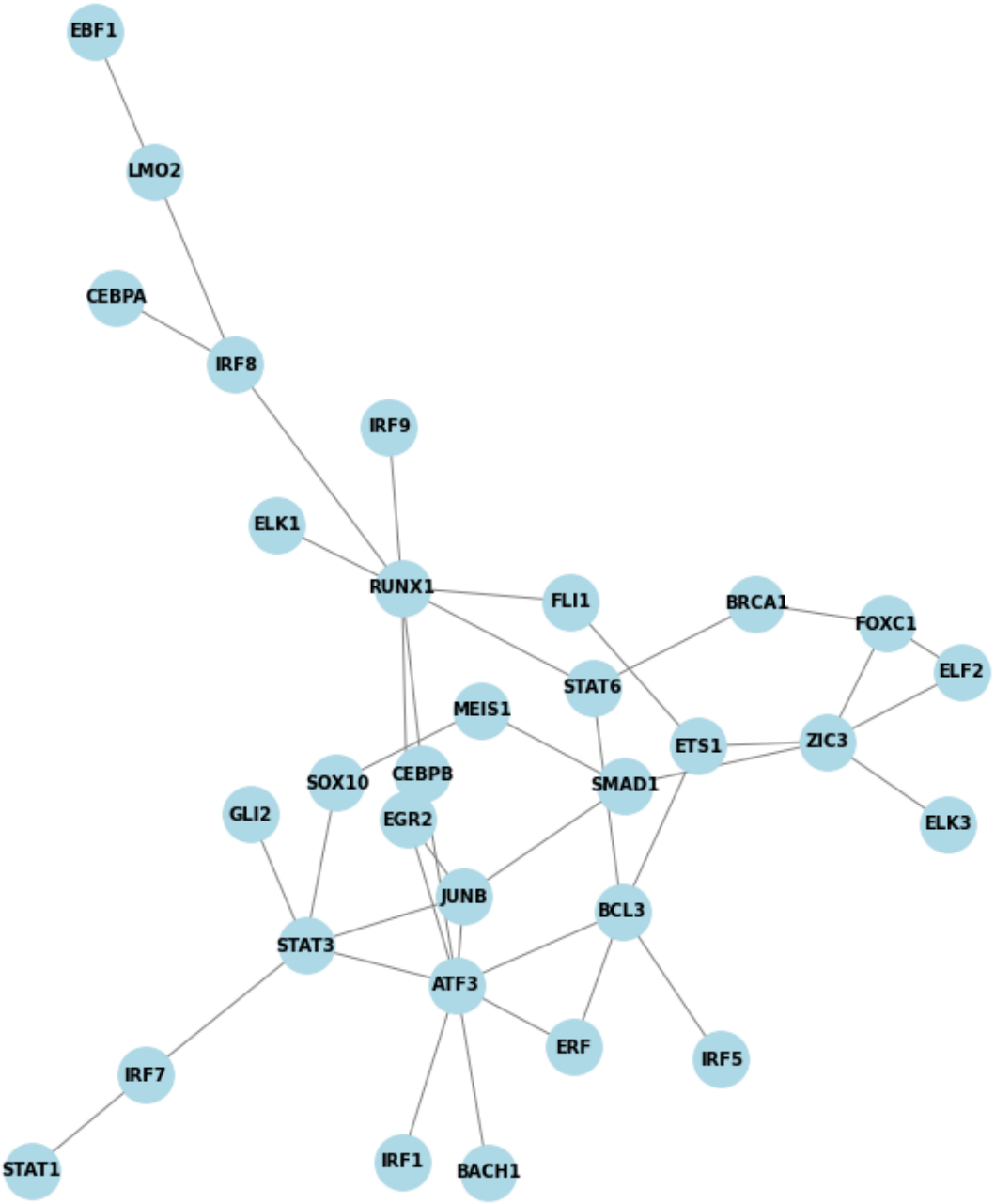
To mitigate this challenge and complement the detailed view provided by the comprehensive gene regulatory network, our framework further constructs a transcription factor gene regulatory network. This network serves as a distilled abstraction of the more detailed network, specifically focusing on the interactions between transcription factors. By simplifying the network to its core components, it provides a more accessible and general overview of how

transcription factors relate to each other, stripping away the extraneous details that might detract from a clear understanding of the regulatory dynamics.

To enhance interpretability, we also distinguished known versus novel interactions in these visualizations. Several TF–TF edges in **Figure 6(b)** are corroborated by curated databases and experimental studies, such as *CEBPB – EGR2* and *IRF7 – STAT1* (documented in TRRUST) and *EBF1 – LMO2* [58]. Likewise, *STAT3 – JUNB* has experimental support in stem cell regulation [59]. At the same time, the networks highlight regulatory interactions not currently captured in TRRUST or the literature, pointing to potentially novel links that warrant future exploration.



(a)



(b)

Figure 6: Visualization of gene regulatory networks. This figure provides a dual visualization of gene regulatory networks from the GSE60361 dataset: (a) presents a complex, comprehensive gene regulatory network displaying numerous interactions, while (b) focuses on the interactions between transcription factors to simplify the network by presenting its core components, providing a more accessible and general overview of how transcription factors interact with each other.

EnsembleRegNet demonstrates a consistent advantage over alternative methods like SIGNET and SCENIC, excelling in several key performance metrics. The model not only achieves high accuracy in grouping cells based on gene expression profiles but also provides enhanced sensitivity in identifying regulatory interactions. EnsembleRegNet's balanced use of external and internal validation metrics underscores its ability to form coherent clusters that align closely with the biological structures in the data. By combining clustering precision and interpretability, EnsembleRegNet proves itself to be a powerful tool for exploring complex gene regulatory networks across diverse biological contexts, potentially advancing insights into cellular behaviors and regulatory mechanisms.

4. Conclusions

Drawing upon the remarkable advancements made with deep learning in the realms of cell clustering and dropout imputation within single-cell RNA sequencing (scRNA-seq) data analysis, this study ventured into the application of deep learning techniques for the inference of Gene Regulatory Networks (GRNs) using scRNA-seq data. Our approach, termed EnsembleRegNet, leverages an innovative ensemble of encoder-decoder architectures with Multilayer Perceptron Bagging by using the Hodges-Lehmann Estimator (HLE) for binarization to transform the expression matrix from scRNA-seq data into a binary matrix indicative of gene activity. This matrix then serves as input to the ensemble model, which discerns the regulatory impact of transcription factors (TFs) on target genes by evaluating the consequences of TF omission from input features on gene expression prediction. A TF's removal that results in significant prediction accuracy variation for a gene's expression earmarks that gene as a regulatory target, forming what we refer to as a regulon. These regulons undergo further refinement through motif enrichment analysis to ensure biological relevance and specificity.

Our model's effectiveness was benchmarked against other methodologies using three publicly available scRNA-seq datasets for cell clustering. This provided robust evidence of EnsembleRegNet's superior sensitivity in identifying functional regulons and clustering rare cell types. Moreover, it affords analytical utilities for GRN construction and visualization, enhancing the interpretability and usability of generated insights for downstream analyses. The strength of the EnsembleRegNet framework is its robustness, achieved through the use of non-parametric statistics and regularized models. While the default parameters performed consistently well across the diverse datasets presented here, we have provided guidance within the Methods section on key parameters that users may adjust to tailor the inference to specific biological contexts or data characteristics, such as extreme sparsity or very large cell numbers.

The superior performance and sensitivity of EnsembleRegNet stem from the integration of MLP models, bootstrap-driven bias reduction, and motif enrichment–based refinement via RcisTarget. These design choices enable the framework to capture complex regulatory dynamics more effectively than existing methods, as reflected in its recovery of a greater number of ChIP-seq–validated targets per regulon. In comparative evaluations across GSE60361, GSE115978, and GSE103322, EnsembleRegNet outperformed in adjusted rand index (ARI), normalized mutual information (NMI), Normalized Van Dongen (NVD), and Pair Sets Index (PSI) scores, while also achieving stronger internal validation through lower RMSSTD and higher Modified Hubert Γ statistics.

Taken together, EnsembleRegNet demonstrates a significant advancement in the precision of GRN inference from scRNA-seq data and establishes a new benchmark for detecting regulatory interactions and clustering cell types. The method is particularly well-suited for scenarios where cell-type–specific regulatory programs, rare subpopulations, or disease-associated transcriptional

rewiring are of interest, such as in cancer biology, immunology, and developmental systems. Its integration of binarization, ensemble deep learning, and motif enrichment provides distinct advantages over alternative tools in contexts requiring both predictive accuracy and biological interpretability. However, like other deep learning approaches, EnsembleRegNet requires sufficient sample size to avoid overfitting, can be computationally demanding for very large datasets, and may need additional preprocessing or careful threshold adjustments in sparse or noisy data. Furthermore, while EnsembleRegNet enhances interpretability through RcisTarget, AUCell, and case-deletion analysis of TF importance, it cannot fully distinguish direct from indirect regulatory effects without complementary experimental validation.

In conclusion, EnsembleRegNet offers a robust, interpretable, and versatile framework for GRN inference from scRNA-seq data, clearly highlighting the contexts in which it excels while acknowledging the conditions where caution is warranted.

Author Contributions

Data collection and processing: DA; conceived and designed the experiments: DA, AK, and MTH; performed the experiments: DA; analyzed the data: DA, AK, and MTH; contributed reagents/materials/analysis tools: MTH; wrote the paper: DA, AK, and MTH. All authors have read and agreed to publish the final version of the manuscript.

Data sharing plans

All model code & data, and a working software version of the tool are freely available and can be accessed at https://github.com/DuaaAlawad/EnsembleRegNet.git

Funding

The research reported in the paper was partially supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P2O GM103424-21.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] R. W. Carthew, "Gene Regulation and Cellular Metabolism: An Essential Partnership," *Trends Genet.*, vol. 37, no. 4, pp. 389–400, Apr. 2021, doi: 10.1016/j.tig.2020.09.018.

[2] T. A. Wynn, A. Chawla, and J. W. Pollard, "Macrophage biology in development, homeostasis and disease," *Nature*, vol. 496, no. 7446, pp. 445–455, Apr. 2013, doi: 10.1038/nature12034.

[3] L. T. MacNeil and A. J. M. Walhout, "Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression," *Genome Res.*, vol. 21, no. 5, pp. 645–657, May 2011, doi: 10.1101/gr.097378.109.

[4] S. Sinha *et al.*, "Behavior-related gene regulatory networks: A new level of organization in the brain," *Proc. Natl. Acad. Sci.*, vol. 117, no. 38, pp. 23270–23279, Sep. 2020, doi: 10.1073/pnas.1921625117.

[5] D. S. Johnston *et al.*, "The Mouse Epididymal Transcriptome: Transcriptional Profiling of Segmental Gene Expression in the Epididymis1," *Biol. Reprod.*, vol. 73, no. 3, pp. 404–413, Sep. 2005, doi: 10.1095/biolreprod.105.039719.

[6] B. M. Turner, "Epigenetic responses to environmental change and their evolutionary implications," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 364, no. 1534, pp. 3403–3418, Nov. 2009, doi: 10.1098/rstb.2009.0125.

[7] A.-L. Todeschini, A. Georges, and R. A. Veitia, "Transcription factors: specific DNA binding and specific gene regulation," *Trends Genet.*, vol. 30, no. 6, pp. 211–219, Jun. 2014, doi: 10.1016/j.tig.2014.04.002.

[8] S. Hahn and E. T. Young, "Transcriptional Regulation in *Saccharomyces cerevisiae* : Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators," *Genetics*, vol. 189, no. 3, pp. 705–736, Nov. 2011, doi: 10.1534/genetics.111.127019.

[9] H. Bolouri, *Computational Modeling of Gene Regulatory Networks — A Primer*. PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., 2008. doi: 10.1142/p567.

[10] S. Sherekar and G. A. Viswanathan, "Boolean dynamic modeling of cancer signaling networks: Prognosis, progression, and therapeutics," *Comput. Syst. Oncol.*, vol. 1, no. 2, p. e1017, Jun. 2021, doi: 10.1002/cso2.1017.

[11] A. Adil, V. Kumar, A. T. Jan, and M. Asger, “Single-Cell Transcriptomics: Current Methods and Challenges in Data Acquisition and Analysis,” *Front. Neurosci.*, vol. 15, p. 591122, Apr. 2021, doi: 10.3389/fnins.2021.591122.

[12] D. Jovic, X. Liang, H. Zeng, L. Lin, F. Xu, and Y. Luo, “Single-cell RNA sequencing technologies and applications: A brief overview,” *Clin. Transl. Med.*, vol. 12, no. 3, p. e694, Mar. 2022, doi: 10.1002/ctm2.694.

[13] X. Tang, Y. Huang, J. Lei, H. Luo, and X. Zhu, “The single-cell sequencing: new developments and medical applications,” *Cell Biosci.*, vol. 9, no. 1, p. 53, Dec. 2019, doi: 10.1186/s13578-019-0314-y.

[14] J. Wang, A. Ma, Q. Ma, D. Xu, and T. Joshi, “Inductive inference of gene regulatory network using supervised and semi-supervised graph neural networks,” *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 3335–3343, 2020, doi: 10.1016/j.csbj.2020.10.022.

[15] M. Ishikawa *et al.*, “RENGE infers gene regulatory networks using time-series single-cell RNA-seq data with CRISPR perturbations,” *Commun. Biol.*, vol. 6, no. 1, p. 1290, Dec. 2023, doi: 10.1038/s42003-023-05594-4.

[16] T. Celià-Terrassa and M. K. Jolly, “Cancer Stem Cells and Epithelial-to-Mesenchymal Transition in Cancer Metastasis,” *Cold Spring Harb. Perspect. Med.*, vol. 10, no. 7, p. a036905, Jul. 2020, doi: 10.1101/cshperspect.a036905.

[17] X. Chen, D. Xie, Q. Zhao, and Z.-H. You, “MicroRNAs and complex diseases: from experimental results to computational models,” *Brief. Bioinform.*, vol. 20, no. 2, pp. 515–539, Mar. 2019, doi: 10.1093/bib/bbx130.

[18] N. Papili Gao, S. M. M. Ud-Dean, O. Gandrillon, and R. Gunawan, “SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles,” *Bioinformatics*, vol. 34, no. 2, pp. 258–266, Jan. 2018, doi: 10.1093/bioinformatics/btx575.

[19] H. Matsumoto *et al.*, “SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation,” *Bioinformatics*, vol. 33, no. 15, pp. 2314–2321, Aug. 2017, doi: 10.1093/bioinformatics/btx194.

[20] T. E. Chan, M. P. H. Stumpf, and A. C. Babtie, “Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures,” *Cell Syst.*, vol. 5, no. 3, pp. 251-267.e3, Sep. 2017, doi: 10.1016/j.cels.2017.08.014.

[21] M. Sanchez-Castillo, D. Blanco, I. M. Tienda-Luna, M. C. Carrion, and Y. Huang, “A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data,” *Bioinformatics*, vol. 34, no. 6, pp. 964–970, Mar. 2018, doi: 10.1093/bioinformatics/btx605.

[22] F. K. Hamey, S. Nestorowa, S. J. Kinston, D. G. Kent, N. K. Wilson, and B. Göttgens, “Reconstructing blood stem cell regulatory network models from single-cell molecular profiles,” *Proc. Natl. Acad. Sci.*, vol. 114, no. 23, pp. 5822–5829, Jun. 2017, doi: 10.1073/pnas.1610609114.

[23] X. Qiu *et al.*, “Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe,” *Cell Syst.*, vol. 10, no. 3, pp. 265-274.e11, Mar. 2020, doi: 10.1016/j.cels.2020.02.003.

[24] Y. Xu, Z. Zhang, L. You, J. Liu, Z. Fan, and X. Zhou, "scIGANs: single-cell RNA-seq imputation using generative adversarial networks," *Nucleic Acids Res.*, vol. 48, no. 15, pp. e85–e85, Sep. 2020, doi: 10.1093/nar/gkaa506.

[25] Q. Liu, S. Chen, R. Jiang, and W. H. Wong, "Simultaneous deep generative modelling and clustering of single-cell genomic data," *Nat. Mach. Intell.*, vol. 3, no. 6, pp. 536–544, May 2021, doi: 10.1038/s42256-021-00333-y.

[26] L. Yuan, S. Sun, Y. Jiang, Q. Zhang, L. Ye, C.-H. Zheng, and D.-S. Huang, "scRGCL: a cell type annotation method for single-cell RNA-seq data using residual graph convolutional neural network with contrastive learning," *Briefings in Bioinformatics*, vol. 26, no. 1, p. bbae662, Jan. 2025, doi: 10.1093/bib/bbae662.

[27] L. Yuan, Z. Xu, B. Meng, et al., "scAMZI: attention-based deep autoencoder with zero-inflated layer for clustering scRNA-seq data," *BMC Genomics*, vol. 26, p. 350, 2025, doi: 10.1186/s12864-025-11511-2.

[28] L. Yuan, L. Zhao, Y. Jiang, Z. Shen, Q. Zhang, M. Zhang, C.-H. Zheng, and D.-S. Huang, "scMGATGRN: a multiview graph attention network–based method for inferring gene regulatory networks from single-cell transcriptomic data," *Briefings in Bioinformatics*, vol. 25, no. 6, p. bbae526, Nov. 2024, doi: 10.1093/bib/bbae526.

[29] S. Aibar *et al.*, "SCENIC: single-cell regulatory network inference and clustering," *Nat. Methods*, vol. 14, no. 11, pp. 1083–1086, Nov. 2017, doi: 10.1038/nmeth.4463.

[30] B. Van de Sande *et al.*, "A scalable SCENIC workflow for single-cell gene regulatory network analysis," *Nat. Protoc.*, vol. 15, no. 7, pp. 2247–2276, Jul. 2020, doi: 10.1038/s41596-020-0336-2.

[31] A. Zeisel *et al.*, "Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq," *Science*, vol. 347, no. 6226, pp. 1138–1142, Mar. 2015, doi: 10.1126/science.aaa1934.

[32] L. Jerby-Arnon *et al.*, "A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade," *Cell*, vol. 175, no. 4, pp. 984-997.e24, Nov. 2018, doi: 10.1016/j.cell.2018.09.006.

[33] S. V. Puram *et al.*, "Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer," *Cell*, vol. 171, no. 7, pp. 1611-1624.e24, Dec. 2017, doi: 10.1016/j.cell.2017.10.044.

[34] Q. Luo, Y. Yu, and X. Lan, "SIGNET: single-cell RNA-seq-based gene regulatory network prediction using multiple-layer perceptron bagging," *Brief. Bioinform.*, vol. 23, no. 1, p. bbab547, Jan. 2022, doi: 10.1093/bib/bbab547.

[35] F. A. Wolf, P. Angerer, and F. J. Theis, "SCANPY: large-scale single-cell gene expression data analysis," *Genome Biol.*, vol. 19, no. 1, p. 15, Dec. 2018, doi: 10.1186/s13059-017-1382-0.

[36] T. A. Geddes *et al.*, "Autoencoder-based cluster ensembles for single-cell RNA-seq data analysis," *BMC Bioinformatics*, vol. 20, no. S19, p. 660, Dec. 2019, doi: 10.1186/s12859-019-3179-5.

[37] J. Orbach, "Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms.," *Arch. Gen. Psychiatry*, vol. 7, no. 3, p. 218, Sep. 1962, doi: 10.1001/archpsyc.1962.01720030064010.

[38] K. Fukushima, “Visual Feature Extraction by a Multilayered Network of Analog Threshold Elements,” *IEEE Trans. Syst. Sci. Cybern.*, vol. 5, no. 4, pp. 322–333, 1969, doi: 10.1109/TSSC.1969.300225.

[39] B. Rosner, “Percentage Points for a Generalized ESD Many-Outlier Procedure,” *Technometrics*, vol. 25, no. 2, pp. 165–172, May 1983, doi: 10.1080/00401706.1983.10487848.

[40] S. A. Stein Aerts, “AUCell: Analysis of ‘gene set’ activity in single-cell RNA-seq data.” VIB-KU Leuven Center for Brain \& Disease Research, Laboratory of Computational Biology. [Online]. Available: http://scenic.aertslab.org

[41] S. Zhang, X. Li, J. Lin, Q. Lin, and K.-C. Wong, “Review of single-cell RNA-seq data clustering for cell-type identification and characterization,” *RNA*, vol. 29, no. 5, pp. 517–530, May 2023, doi: 10.1261/rna.078965.121.

[42] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, “Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets,” *Nat. Commun.*, vol. 10, no. 1, p. 5415, Nov. 2019, doi: 10.1038/s41467-019-13055-y.

[43] R. Li *et al.*, “Inflammatory Gene Regulatory Networks in Amnion Cells Following Cytokine Stimulation: Translational Systems Approach to Modeling Human Parturition,” *PLoS ONE*, vol. 6, no. 6, p. e20560, Jun. 2011, doi: 10.1371/journal.pone.0020560.

[44] D. Mercatelli, L. Scalambra, L. Triboli, F. Ray, and F. M. Giorgi, “Gene regulatory network inference resources: A practical overview,” *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1863, no. 6, p. 194430, Jun. 2020, doi: 10.1016/j.bbagrm.2019.194430.

[45] J. Handl, J. Knowles, and D. B. Kell, “Computational cluster validation in post-genomic data analysis,” *Bioinformatics*, vol. 21, no. 15, pp. 3201–3212, Aug. 2005, doi: 10.1093/bioinformatics/bti517.

[46] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu, “Understanding of Internal Clustering Validation Measures,” in *2010 IEEE International Conference on Data Mining*, Sydney, Australia: IEEE, Dec. 2010, pp. 911–916. doi: 10.1109/ICDM.2010.35.

[47] H. Xiong and Z. Li, “Clustering Validation Measures,” in *Data Clustering*, 1st ed., C. C. Aggarwal and C. K. Reddy, Eds., Chapman and Hall/CRC, 2018, pp. 571–606. doi: 10.1201/9781315373515-23.

[48] P. Rujasiri and B. Chomtee, “Comparison of Clustering Techniques for Cluster Analysis”.

[49] Heng Zhao, Jimin Liang, and Haihong Hu, “Clustering Validity Based on the Improved Hubert \Gamma Statistic and the Separation of Clusters,” in *First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC’06)*, Beijing, China: IEEE, 2006, pp. 539–543. doi: 10.1109/ICICIC.2006.250.

[50] K. R. Shahapure and C. Nicholas, “Cluster Quality Analysis Using Silhouette Score,” in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, sydney, Australia: IEEE, Oct. 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.

[51] S Anitha and M. Metilda, “AN EXTENSIVE INVESTIGATION OF OUTLIER DETECTION BY CLUSTER VALIDATION INDICES,” 2019, doi: 10.13140/RG.2.2.26801.63848.

[52] M. Rezaei and P. Franti, “Set Matching Measures for External Cluster Validity,” *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 8, pp. 2173–2186, Aug. 2016, doi: 10.1109/TKDE.2016.2551240.

[53] N. X. Vinh, J. Epps, and J. Bailey, “Information theoretic measures for clusterings comparison: is a correction for chance necessary?,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, Montreal Quebec Canada: ACM, Jun. 2009, pp. 1073–1080. doi: 10.1145/1553374.1553511.

[54] V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, and P. Geurts, “Inferring regulatory networks from expression data using tree-based methods,” *PLoS One*, vol. 5, no. 9, p. e12776, Sep. 2010, doi: 10.1371/journal.pone.0012776.

[55] D. Chan, M. T. Hsu, M. Hoon, D. R. Meyer, J. J. Drum, M. Kellis, and B. J. Wold, “Gene regulatory inference from single-cell expression data using partial information decomposition and context,” *Nature Communications*, vol. 8, no. 1, p. 1502, Nov. 2017, doi: 10.1038/s41467-017-01543-z.

[56] H. Han, J. W. Cho, S. Lee, A. Yun, H. Kim, D. Bae, S. Yang, *et al.*, “TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions,” *Nucleic Acids Research*, vol. 46, no. D1, pp. D380–D386, Jan. 2018, doi: 10.1093/nar/gkx1013.

[57] L. Garcia-Alonso, C. H. Holland, M. M. Ibrahim, D. Turei, and J. Saez-Rodriguez, “Benchmark and integration of resources for the estimation of human transcription factor activities,” *Genome Research*, vol. 29, no. 8, pp. 1363–1375, Aug. 2019, doi: 10.1101/gr.240663.118.

[58] K. Hirano, H. Hosokawa, M. Koizumi, Y. Endo, T. Yahata, K. Ando, and K. Hozumi, “LMO2 is essential to maintain the ability of progenitors to differentiate into T-cell lineage in mice,” *eLife*, vol. 10, p. e68227, Aug. 2021, doi: 10.7554/eLife.68227.

[59] P.-Y. Bourillot, I. Aksoy, V. Schreiber, F. Wianney, H. Schulz, O. Hummel, N. Hubner, and P. Savatier, “Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm differentiation in cooperation with Nanog,” *Stem Cells*, vol. 27, no. 4, pp. 757–768, Apr. 2009, doi: 10.1002/stem.110.