Research
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Constructing Cell-Specific Causal Networks of Individual Cells for Depicting Dynamical Biological Processes

Xinzhe Huang[1], Luonan Chen[1,2*], and Xiaoping Liu[1*]

[1]Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China. [2]School of Mathematical Sciences and School of AI, Shanghai Jiao Tong University, Shanghai 200240, China.

*Address correspondence to: xpliu@ucas.ac.cn (X.L.); lnchen@sjtu.edu.cn (L.C.)

Citation: Huang X, Chen L, Liu X. Constructing Cell-Specific Causal Networks of Individual Cells for Depicting Dynamical Biological Processes. *Research* 2025;8:Article 0743. https://doi.org/10.34133/research.0743

Submitted 7 April 2025
Revised 19 May 2025
Accepted 26 May 2025
Published 27 June 2025

Copyright © 2025 Xinzhe Huang et al. Exclusive licensee Science and Technology Review Publishing House. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Causal inference is crucial in biological research, as it enables the understanding of complex relationships and dynamic processes that drive cellular behavior, development, and disease. Within this context, gene regulatory network (GRN) inference serves as a key approach for understanding the molecular mechanisms underlying cellular function. Despite substantial advancements, challenges persist in GRN inference, particularly in dynamic rewiring, inferring causality, and context specificity. To tackle these issues, we present single cell-specific causal network (SiCNet), a novel causal network construction method that utilizes single-cell gene expression profiles and a causal inference strategy to construct molecular regulatory networks at a single-cell level. Additionally, SiCNet utilizes cell-specific network information to construct network outdegree matrix (ODM), enhancing the performance of cell clustering. It also enables the construction of context-specific GRNs to identify key regulators of fate transitions for diverse processes such as cellular reprogramming and development. Furthermore, SiCNet can delineate the intricate dynamic regulatory processes involved in development, providing deep insights into the mechanisms governing cellular transitions and the gene regulation across developmental stages.

Introduction

The single-cell RNA sequencing (scRNA-seq) technology has revolutionized molecular biology by providing high-resolution, large-scale transcriptome data at the single-cell level [1,2]. With this technology, researchers can identify rare or novel cell types in tissues, study cell differentiation, and uncover the molecular mechanisms behind diseases with computational methods [3–5]. Many biological processes, such as transcriptional regulation and epigenetic modifications, involve molecular interactions among transcriptional products and regulatory elements. Deciphering these complex regulation relationships at a single-cell level is crucial for understanding cellular heterogeneity and uncovering the underlying mechanisms of biological functions and disease progression. Causal relationships, which are central to molecular biology, are key to understanding the underlying mechanisms of biological processes. Unlike correlations, causality can distinguish which gene is the cause and which gene is the effect [6]. Generally, the cause gene is defined as the regulator and the effect gene is defined as the target gene. This is particularly important in understanding disease mechanisms, identifying drug targets, and reconstructing gene regulatory networks (GRNs). By focusing on causality, researchers can develop more precise treatments, gain deeper insights into biological systems, and advance fields such as synthetic biology and personalized medicine [7,8].

The inference of GRNs has become one of the key approaches for elucidating molecular causality in biological systems. Most well-known methods inferring GRNs for single-cell data, such as GENIE3 [9], GRNBoost2 [10], and PIDC [11], are limited to construct GRNs at the cell type level, but not at the individual cell level. These methods fail to account for the heterogeneity among individual cells in a population. Recently, methods like SSN [12] and PSSN [13] have been proposed to construct relevance networks for individual cells. However, these methods primarily focus on correlations between genes, potentially missing causal relationships and more complex regulatory interactions.

In this paper, we proposed a novel method to construct causal networks for individual cells, referred to as the single cell-specific causal network (SiCNet), to address a challenge of inferring causal relationships from single-cell expression profile. By incorporating a statistical concept rooted in causal inference called CVP [14], we applied this causal inference framework to construct cell-specific causal networks. By analyzing cell-specific causal networks, we can identify key regulators and potential regulatory relationships in the context of cancer, shedding light on cancer progression. Furthermore, by integrating these cell-specific causal networks, we can represent the states of individual samples and detect critical states by calculating dynamic network biomarker (DNB) scores. Additionally, we aim to uncover regulatory processes involved

in cell reprogramming and dynamic regulatory mechanisms during cellular differentiation, providing insights into factors associated with reprogramming efficiency and hematopoietic specification.

Results

Overview of SiCNet

Based on the previous study [14], we used this causal network inference framework to develop a new approach, SiCNet, for predicting causal relationship at the single-cell level. The SiCNet is a data-driven and model-free algorithm for inferring cell-specific causal relationships between genes based on single-cell transcriptome (Fig. 1A).

To construct a causal network for a single cell, an additional single-cell transcriptomic dataset was required as a reference dataset. This reference dataset was used to establish an initial causal network based on the CVP method [14], and the initial causal network is as the reference causal network (Fig. 1B and Methods). Specifically, we employed 5-fold cross-validation to partition the reference dataset into 5 equal subsets. For each iteration, 4 subsets were used for model training, while the remaining subset served for model prediction. The causal inference was performed pairwise between genes. We trained 2 distinct regression models for each gene pair: one incorporating the presence of the cause gene, and another excluding it. The predictive performance of each model was evaluated by calculating the average root mean square error (RMSE), yielding ε_1^r for models including the cause gene and ε_2^r for those excluding it. The causal strength index (CSI; ω_r) between the 2 genes was then quantified as the logarithm of the ratio between these RMSE values $[ln\left(\varepsilon_2^r/\varepsilon_1^r\right)]$. A positive CSI ($\omega_r > 0$) indicates a statistically significant causal effect between the gene pair, whereas a nonpositive value suggests the absence of such causal relationship (Fig. 1B and Methods). The subscript/superscript "r" in all symbols denotes the "reference". The reference causal network captured general causal relationships between genes in the reference cells, serving as a benchmark of the initial GRN.

After obtaining the reference causal network, the gene expression data of a single cell were used to infer the causal relationships between genes (Fig. 1C and Methods). Specifically, for a target gene regulated by several other genes (regulators) in the reference network, we can predict the expression of this gene from the genes who regulate the target gene. Then, we can predict the target gene's expression again after removing a regulating gene (regulator). If the prediction loss increases after removing a regulator, it indicates that the removed regulator has a causal influence on this target gene, thereby revealing its regulatory role in the causal network. Similarly, the causal relationship is determined by the sign of the CSI, which is represented by ω^c (Fig. 1C and Methods). The subscript/superscript "c" in all symbols denotes the "specific cell". The cell-specific causal network was obtained by collecting regulation relationship of every gene, and it can reflect the individualized causal relationships adapted to the gene expression profile in particular cell. To enhance the reliability of the causal network prediction, a protein–protein interaction (PPI) network from the STRING database was used as the background network (Fig. 1 and Methods).

By identifying the set of target genes for each regulator, the regulatory activity of each gene can be quantified by counting the number of target genes in the cell-specific network. Based on this, a network outdegree matrix (ODM) was generated for further analysis (Fig. 1D and Methods). The ODM retains the same dimensions as the original gene expression matrix (GEM), with the same number of rows and columns, but it represents the higher-order regulatory information of each gene in the network. This matrix can be seamlessly analyzed using any standard scRNA-seq algorithm for tasks such as cell clustering, dimensionality reduction, and pseudo-trajectory analysis, simply by substituting the original GEM with our ODM.

The performance of SiCNet in clustering on scRNA-seq datasets

To verify whether the SiCNet method can effectively characterize cellular heterogeneity, we constructed cell-specific networks for 5 scRNA-seq datasets (see Data Availability and Table S1) with explicit cellular annotations. For each dataset, we transformed the cell-specific networks constructed by SiCNet into an ODM for dimensionality reduction, visualization, and clustering analysis (see Methods).

For the purpose of visualizing the datasets, we employed *t*-distributed stochastic neighbor embedding (*t*-SNE) [15] to reduce the dimensionality of both the GEM and the ODM (Fig. 2A and B). Notably, the ODM visualizations exhibit clearer demarcation and clustering between different cell types, particularly in the Enge, Darmanis, and Fink datasets (Fig. 2B), where cell populations are more distinctly separated compared with the GEM visualizations (Fig. 2A). For example, in the Enge dataset, the representation of ODM clearly shows more compact and well-defined clusters, especially for cell types such as alpha (orange) and ductal (green), which are more tightly grouped compared to the GEM visualization (Fig. 2B). Similarly, in the Darmanis dataset, cell types such as oligodendrocyte precursor cells (OPCs) (cyan) and oligodendrocytes (orange) in ODM are more tightly grouped and clearly separated, while in the corresponding GEM visualization, these cell types reveal a more dispersed distribution with less defined separations (Fig. 2B). Overall, these results suggested that the ODM enhances the resolution and clarity of cell type distinctions, offering superior performance in terms of visualizing complex and high-dimensional data compared with the traditional GEM (Fig. 2B).

To quantitatively compare the clustering results of the ODM with the GEM, we applied 5 single-cell clustering algorithms, including SC3 [16], CIDR [17], Scanpy [18], SIMLR [19], and Supercell [20], across the 5 datasets. These clustering algorithms are widely used in single-cell analysis to uncover distinct cell populations. To evaluate the clustering performance, we calculated the adjusted rand index (ARI), adjusted mutual information (AMI), normalized mutual information (NMI), and Fowlkes–Mallows index (FMI) (Fig. 2C and D and Fig. S1). The results demonstrated that the representation of ODM outperformed the representation of GEM across most datasets, as evidenced by higher scores in the majority of clustering performance metrics. For example, in the Enge dataset, the ODM achieved higher ARI and AMI across all clustering algorithms, compared to the GEM (Fig. 2C and D). Specifically, ODM yielded an ARI of 0.58 and an AMI of 0.67 for the SC3 algorithm, compared to GEM's ARI of 0.33 and AMI of 0.53. Similarly, for the SIMLR algorithm, ODM achieved an ARI of 0.35 and AMI of 0.51, while GEM showed an ARI of 0.27 and AMI of 0.47 (Fig. 2C and D). Similar trend was observed in the Darmanis and Goolam datasets, particularly with algorithms like Supercell and Scanpy (Fig. 2C and

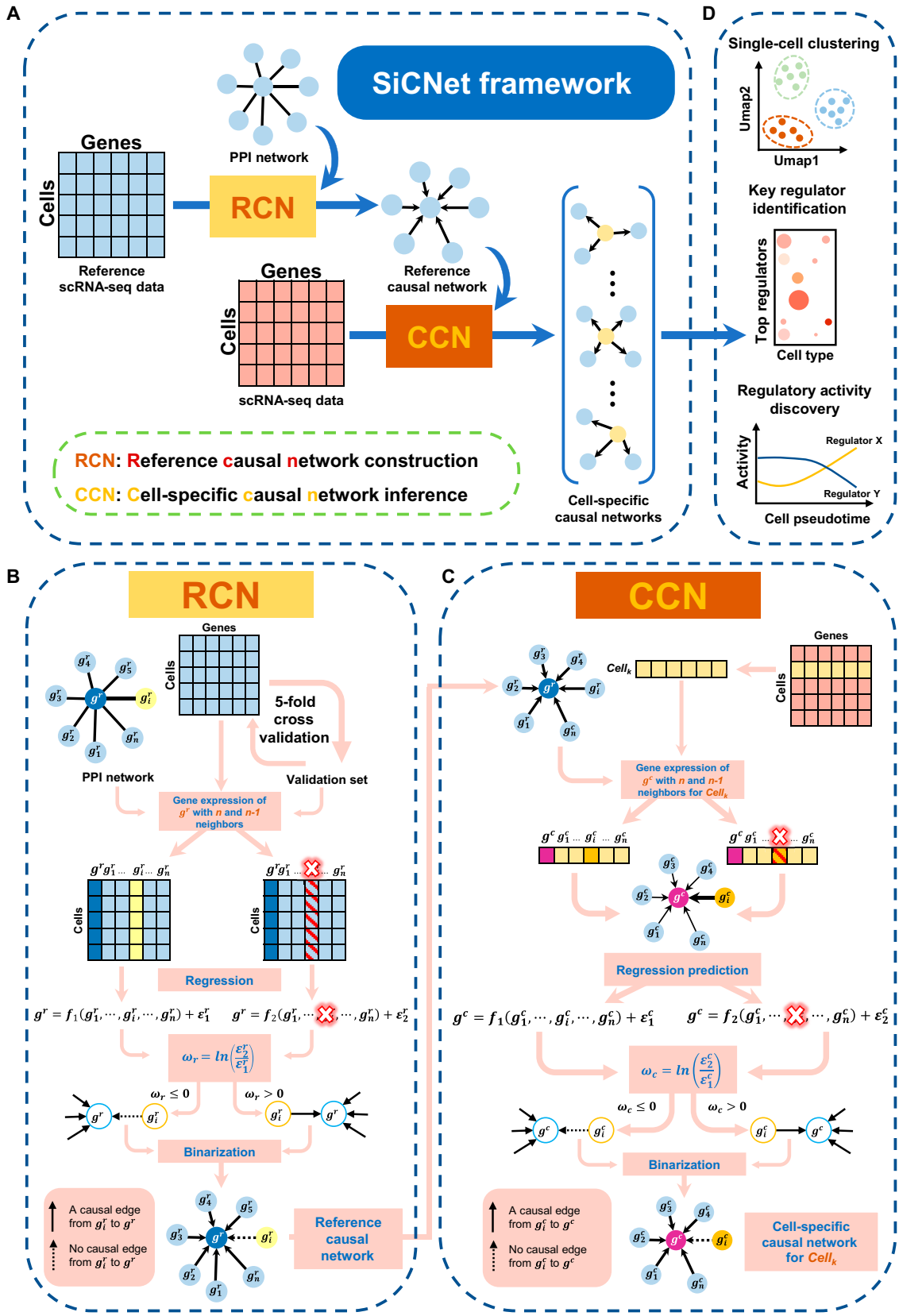


Fig. 1. An overview of the SiCNet framework. (A) The SiCNet method takes reference scRNA-seq data and objective scRNA-seq data, along with protein–protein interaction (PPI) information, as inputs. It generates, as output, a set of causal networks for each cell. (B) Framework for inferring an initial reference causal network. (C) Framework for predicting single cell-specific causal networks. (D) The cell-specific networks generated as output are used for downstream analysis, including clustering, identifying key regulators, and discovering patterns of active regulators.

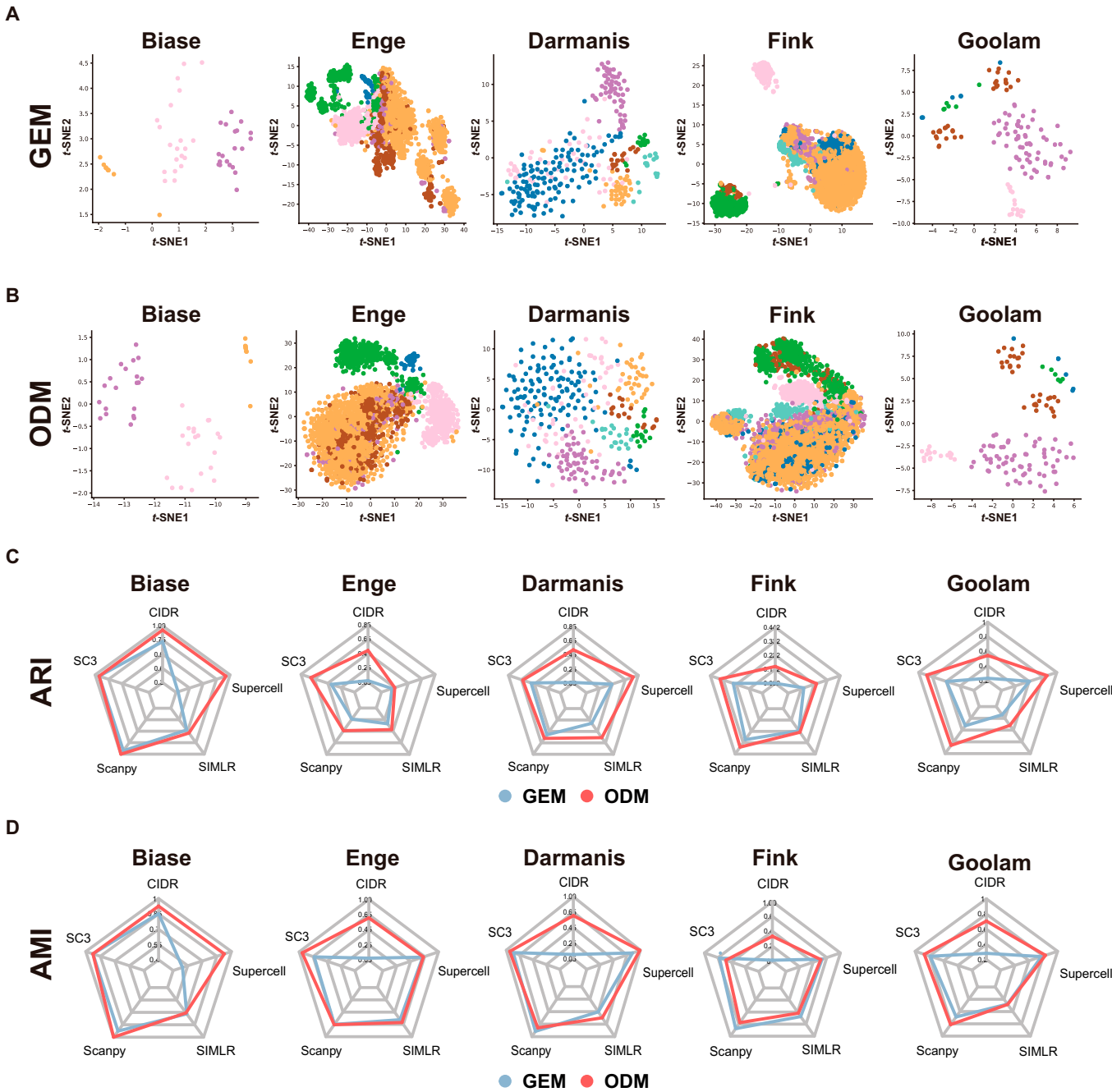


Fig. 2. Five benchmark scRNA-Seq datasets demonstrated superior clustering performance with ODM compared to GEM over existing methods. (A) *t*-SNE visualization of GEM for the datasets. Different colors represent different cell types. (B) *t*-SNE visualization of ODM for the datasets. Different colors represent different cell types. (C) The ARI metrics are used to measure performance of clustering. Clustering methods SC3, CIDR, Scanpy, SIMLR, and Supercell are used to cluster datasets. (D) The AMI metrics are used to measure performance of clustering. Clustering methods SC3, CIDR, Scanpy, SIMLR, and Supercell are used to cluster datasets. GEM, gene expression matrix; ODM, network outdegree matrix; ARI, adjusted rand index; AMI, adjusted mutual information; *t*-SNE, *t*-distributed stochastic neighbor embedding.

D). In the Darmanis dataset, ODM achieved an ARI of 0.63 and an AMI of 0.69 with Supercell, whereas GEM had an ARI of 0.37 and an AMI of 0.60. For the Scanpy algorithm, ODM reached an ARI of 0.47 and AMI of 0.59, outperforming GEM with an ARI of 0.42 and AMI of 0.63 (Fig. 2C and D). In the Goolam dataset, ODM demonstrated superior performance with an ARI of 0.86 and AMI of 0.83 for Supercell, compared to GEM's ARI of 0.61 and AMI of 0.77. For Scanpy, ODM showed an ARI of 0.86 and AMI of 0.81, while GEM had an ARI of 0.53 and AMI of 0.69 (Fig. 2C and D). These results highlight the robustness and effectiveness of the ODM in uncovering distinct cell populations across diverse datasets.

Revealing colorectal cancer progression patterns through SiCNet

To gain insight into the molecular mechanisms underlying disease at a single-cell network level, we obtained a human colorectal cancer scRNA-seq dataset [21], which includes cells from the healthy adjacent tissues as well as the border and core

regions of the tumor tissues (Fig. 3A). It is worth noting that the colorectal cancer primarily originates from colon epithelial cells that lose genomic stability. This instability promotes the accumulation of mutations in tumor suppressor genes and oncogenes [22]. These mutations drive the malignant transformation of colon epithelial cells, leading to the formation of tumor core regions that expands outward through cycles of clonal expansion [23,24]. Therefore, in this study, we focused specifically on epithelial cells across 3 regions of tissue. The number of cells in the healthy tissue was 1,144. In the border region, the number of cells was 2,812, while in the core region, it was 2,212.

We applied the SiCNet method to construct cell-specific networks for each cell region and transformed these networks into 3 ODMs (see Methods): the ODM of the healthy adjacent region, the ODM of the border region, and the ODM of the core region. Notably, each of these ODMs includes the same number of genes as in the GEM and the corresponding number of cells specific to each region. Subsequently, we concatenated the 3 ODMs along the cell dimension, generating a single integrated ODM (6,168 cells) with all original genes retained. This new ODM consolidated information from the 3 regions. Finally, we performed dimensionality reduction and clustering on this new ODM matrix to uncover cancer progression patterns across the 3 regions. Compared with GEM, the ODM can better separate the cell populations of the border and core regions (Fig. 3B), further demonstrating that the ODM derived from the SiCNet method can effectively capture the characteristics and specificities of each cell. It is possible that the healthy adjacent tissues may gradually undergo malignant transformation at the tumor border region, with the lesions becoming more pronounced in the core region of the tumor. We performed a 2-sided Student's *t* test on the ODM data between healthy adjacent and core regions to identify the key driver regulators of colorectal cancer progression (see Methods). For each regulator, the *P* value was adjusted for multiple testing using false discovery rate (FDR) correction (see Methods).

We finally identified 26 regulators with enhanced regulatory activity from the healthy adjacent region to the core region, aligning with previously reported oncogenic transformation patterns essential for the growth and development of colorectal tumors (Fig. 3C and D, Figs. S2 and S3, and Table S2). For example, the increased activity of *ABL1* ($P = 2.20 \times 10^{-219}$) is proved to be associated with decreased metastasis-free survival and/or lower overall survival in colorectal cancer [25,26]. The increased activity of *NCOA1* ($P = 6.07 \times 10^{-157}$) contributes to colorectal cancer progression by enhancing Hedgehog signaling [27]. The increased activity of *ATF3* ($P = 4.89 \times 10^{-278}$) drives colorectal cancer cell proliferation [28,29]. The increased activity of *JUN* ($P < 1.00 \times 10^{-278}$) triggers the activation of the c-Jun N-terminal kinase (JNK) and mitogen-activated protein kinase (MAPK) pathways in colorectal cancer [30,31]. The increased activity of *EGR2* ($P < 1.00 \times 10^{-278}$) facilitates the persistence of colorectal cancer stem cells [32]. In terms of gene expression, some genes, particularly *NCOA1* and *EGR2*, showed no significant differential expression between the healthy adjacent and core regions (Figs. S4 to S6 and Table S3).

To comprehensively evaluate the potential biological functions of these regulators, we conducted an enrichment analysis using the pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and the hallmark gene sets from the Molecular Signatures Database (MSigDB). The KEGG enrichment analysis revealed that the regulators were enriched into the peroxisome proliferator-activated receptor (PPAR) signaling pathway, insulin resistance pathway, and the citrate cycle pathway by hypergeometric test with *P* values 1.49×10^{-4}, 1.12×10^{-2}, and 1.21×10^{-2} (Fig. 3E). All of these pathways are associated with the metabolic pathways and critical for the development of colorectal cancer [33]. These signaling pathways suggest that metabolic dysregulation is one of the major contributing factors in the development and progression of colorectal cancer. It is consistent with the existing knowledge [34–37]. The Hallmark analysis highlighted the enrichment of tumor necrosis factor-α (TNF-α) signaling and hypoxia signaling by hypergeometric test with *P* values 5.24×10^{-6} and 1.25×10^{-2} (Fig. 3F). This signaling indicated that the inflammatory processes were involved in colorectal cancer, and that the molecular interactions in the tumor microenvironment also played a critical role [38,39]. This is in line with previous reports suggesting that the evasion of tumor immune surveillance and low levels of oxygen support cancer cell survival, both of which are the features of the core region of colorectal cancer [40,41].

Regulatory analysis and identifying critical states of colorectal cancer

We then counted the number of active and inactive regulatory connections—where active connections represent regulatory connections present and inactive connections indicate their absence—in each cell-specific network across the healthy, border, and core region, and identified 133 significant up-regulation links by chi-square test that may be associated with the cancer progression from the healthy adjacent region to the core region (Fig. 3G and Table S4). Among these significant links, several regulated genes (target genes) were found to play a crucial role in colorectal cancer (Fig. 3G). For instance, *RAC3* is regulated by *ABL1* (Fig. 3G), and this enhanced regulation has been shown to contribute to chemoresistance in colon cancer cells and support the maintenance of cancer cell stemness [42,43]. Additionally, *NR3C2* is regulated by *NCOA1* (Fig. 3G), and this intensified regulation activates the Wnt/β-catenin signaling pathway, enhancing the invasion of colon cancer cells [44]. *FOSB* is increasingly regulated by *ATF3* from the healthy adjacent region to the core region (Fig. 3G), and it plays a critical role in regulating inflammatory molecule expression, which influences the therapeutic effects of drugs on colorectal cancer [45]. Moreover, *CDK5R1* has been identified as a prognostic marker associated with poor survival in colorectal cancer [46]. We also performed the KEGG enrichment analysis to confirm the association between these regulated genes and the colorectal cancer, highlighting their roles in cancer-associated signaling pathways (Fig. 3H). Consistent with previous reports, the MAPK signaling pathway, phosphatidylinositol 3-kinase (PI3K)/AKT signaling pathway, and Wnt/β-catenin signaling pathway were found to play critical roles in the progression and development of colorectal cancer (Fig. 3H) [47]. In addition, metabolic reprogramming-related pathways, such as the ERBB and FOXO signaling pathways, were also significantly enriched (Fig. 3H), indicating that the dysregulation of these processes promote cancer growth [48–50]. These findings further support the existing evidence that metabolic dysregulation is a hallmark of colorectal cancer, and it can offer potential avenues for therapeutic interventions by targeting cancer cell metabolism.

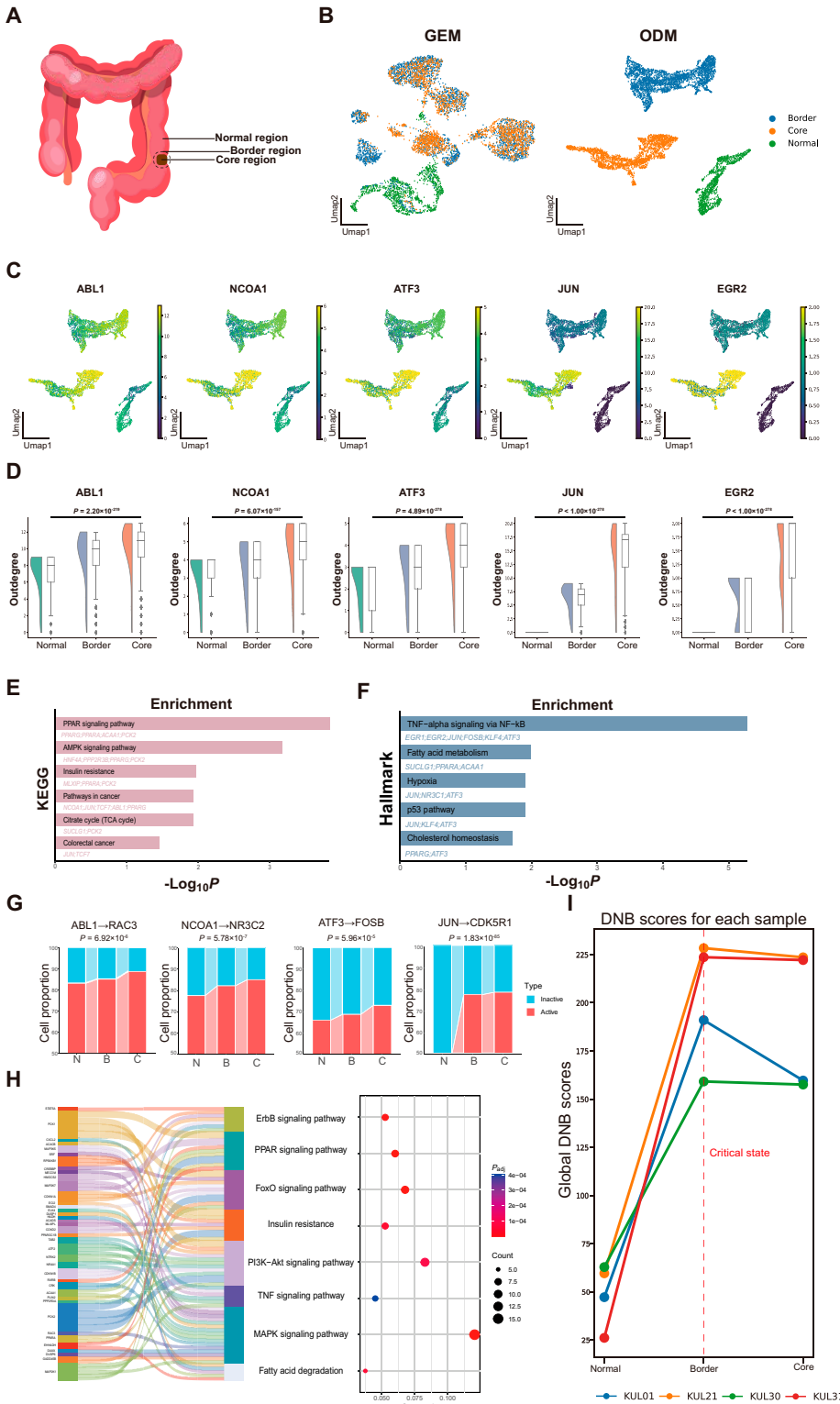


Fig. 3. SiCNet reveals colorectal cancer progression regulatory patterns. (A) Diagram illustrating the colorectal, with the healthy adjacent tissues as well as the border and core regions of the tumor tissues. (B) Leiden clustering and UMAP visualization of GEM and ODM for the datasets. (C) UMAP visualization showing key regulators with high regulatory activity across the gradient from the healthy adjacent region to the core region. (D) Corresponding statistical analysis of the key regulators' regulatory activity from the healthy adjacent region to the core region. *P* values were calculated using a 2-sided Student's *t* test. (E) The identified regulators are significantly enriched in KEGG database to illustrate the progression of colorectal cancer. *P* values were calculated using hypergeometric test. (F) The identified regulators are significantly enriched in MSigDB database to illustrate the progression of colorectal cancer. *P* values were calculated using hypergeometric test. (G) Significantly up-regulated links found across 3 regions to illustrate the downstream regulation of colorectal cancer. *P* values were calculated using chi-square test. (H) Regulated genes from up-regulated links are significantly enriched in KEGG database to illustrate the downstream regulation of colorectal cancer. (I) Identification of critical states for cancer progression for each sample. KEGG, Kyoto Encyclopedia of Genes and Genomes; MSigDB, Molecular Signatures Database; UMAP, Uniform Manifold Approximation and Projection; GEM, gene expression matrix; ODM, network outdegree matrix.

DNB is a model to detect the early warning signal before disease onset [51] based on biological network. In this dataset, the border region represents the transitional zone where cells shift from a normal (healthy region) to a malignant state (core region), and it can be considered as a critical state from normal to disease. Therefore, we aggregated the cell-specific networks of each region to characterize each sample and used the *l*-DNB method [52] to detect critical states (Note S2). Undoubtedly, in most samples, the border region was indeed detected as a critical state (Fig. 3I and Fig. S7). This implied the potential of SiCNet method in applying the DNB approach to detect critical states.

Overall, our SiCNet method provides a cell-specific causal network perspective that effectively characterizes the progression of colorectal cancer in epithelial cells, strongly supporting our current understanding on the development of cancer.

Uncovering key regulators during reprogramming

Cellular reprogramming is the process of converting differentiated cells back to a pluripotent state, which has significant implications in regenerative medicine and patient-specific disease modeling [53–56]. However, the regulatory mechanisms driving this process remain poorly understood.

To gain deeper insights into this dynamic process, we gathered single-cell expression dataset capturing the reprogramming of mouse embryonic fibroblasts (MEFs) into pluripotent stem cells (mESCs). The dataset included 6 time points: the initial MEFs, fully reprogrammed mESCs, and intermediate stages spanning days 3 to 12 (days 3, 6, 9, and 12) (Fig. 4A). We applied the SiCNet method to construct cell-specific networks for the datasets with the 6 time points. These networks were then integrated into an ODM (see Methods). Using the Leiden clustering algorithm [57], we identified 6 distinct clusters (C0 to C5) (Fig. 4B). Each cluster represents a specific cell type in the reprogramming process. Obviously, based on the composition of cells from different time points within each cluster (Fig. 4C and D), C4 is MEF-specific, while C5 is mESC-specific, and the remaining clusters represented a mix of cells spanning different time points. C4 represents the starting cell type of reprogramming, C5 represents the end cell type, and the remaining clusters represent intermediate cell types of the process. This pattern, where cells from different time points are mixed within the clusters, reflects the inherent heterogeneity of the reprogramming system, consistent with its dynamics and stochasticity [58,59].

For each cluster, we calculated the proportion of active regulatory links (edges) in cell-specific networks that appeared in the cells of that cluster. Specifically, the proportion of an edge is defined as the number of cells containing the edge divided by the total number of cells in the cluster. Subsequently, we applied a cutoff to remove edges that were present in less than 70% of the cells in the cluster. This filtering step ensured that the retained regulatory interactions were consistently present in the majority of cells within the cluster. As a result, these retained regulatory interactions in each cluster formed a cell type-specific network. Among these networks, we identified the top transcriptional regulators associated with each of the cell clusters. For example, the MEF-specific cluster (C4) is enriched in *Hoxb4*, *Hoxb7*, *Dlx1*, and *Sox2*, all of which are related to development and differentiation processes [60–63]. In contrast, the mESC-specific cluster (C5) is characterized by the enrichment of *Stat5a*, *Smad3*, *Trp53*, and *Sin3a*, which play essential roles in maintaining pluripotency and regulating reprogramming [64–67]. Based on the earlier designation of the MEF-specific cluster (C4) as the starting point and the mESC-specific cluster (C5) as the endpoint of the reprogramming process, we employed the Slingshot approach [68] to infer the cellular trajectories across the 6 clusters (Fig. 4E).

To investigate the regulatory mechanisms during the reprogramming process, we focused on the key components of each cell type-specific network and compared the differences in regulatory networks across cell types (clusters). As the first key transition point (Fig. 4F), C2 contains the regulators such as *Ccna2*, *Cdca8*, and *Bub1b*, which are known to be involved in cell cycle progression (Fig. 4F). In contrast, these regulators are less prevalent or absent in C3. Consequently, C2 displays a higher cell cycle-related regulatory activity compared to C3. This distinct pattern of regulatory activity in C2 aligns with previous findings highlighting the critical role of cell cycle regulation during reprogramming [69,70].

During the transition from C4 to C2 and then to C0, in addition to the aforementioned cell cycle-related genes, other genes such as *Stat5a*, *Stat1*, *Src*, and *Jun*, which are closely associated with cell proliferation and differentiation, also show active regulatory involvement. This indicates that regulatory processes, such as the cell cycle and cell proliferation, become progressively more active during this transition (Fig. 4F). Moreover, the gene like *Trp53*, which is associated with apoptosis and DNA repair, also begins to exhibit regulatory activity during this transition (Fig. 4F). The processes of apoptosis and DNA repair are essential for cellular reprogramming, as they ensure the survival and proper differentiation of reprogrammed cells [71–73]. Next, the second transition point is identified at C0, from which 2 distinct trajectories emerge as C0 to C1 and C0 to C5 (Fig. 4F). Compared to C1, C5 has more regulatory links associated with maintaining genome stability, which involve regulators such as *Gnas*, *Ctnnb1*, *Trp53*, and *Hdac1* (Fig. 4F). These regulators contribute to a tight regulation network that ensures cellular homeostasis and stemness during reprogramming [74,75]. We summarized the entire dynamic process using GO enrichment analysis and calculated the Jaccard similarity between these cell types (clusters) (Fig. 4G and Methods). As expected, the similarity gradually decreased as the reprogramming process progressed.

Overall, the results derived from our SiCNet method demonstrate its effectiveness in uncovering the core regulatory mechanisms during the reprogramming process. Building on these findings, the newly identified regulators present potential targets for perturbation, enabling us to assess their impact on reprogramming efficiency.

Identifying the developmental tendency and the dynamics of regulation in human hematopoiesis

To further demonstrate the utility of SiCNet in differential fields, we obtained scRNA-seq data of bone marrow mononuclear cells (BMMC) from healthy donors (Fig. 5A and B) [76,77]. We constructed the cell-specific network for each individual cell based on SiCNet, and then integrated and filtered the networks for each cell type to obtain the cell type-specific networks. Subsequently, we identified the top regulators exhibiting potential regulatory activity for each cell type (Fig. 5C). For example, *EGR1*, a gene known for its role in hematopoiesis, plays a critical role in the development and maturation of erythroid cells [78]. In addition, *PIK3CA*, a key regulator of the

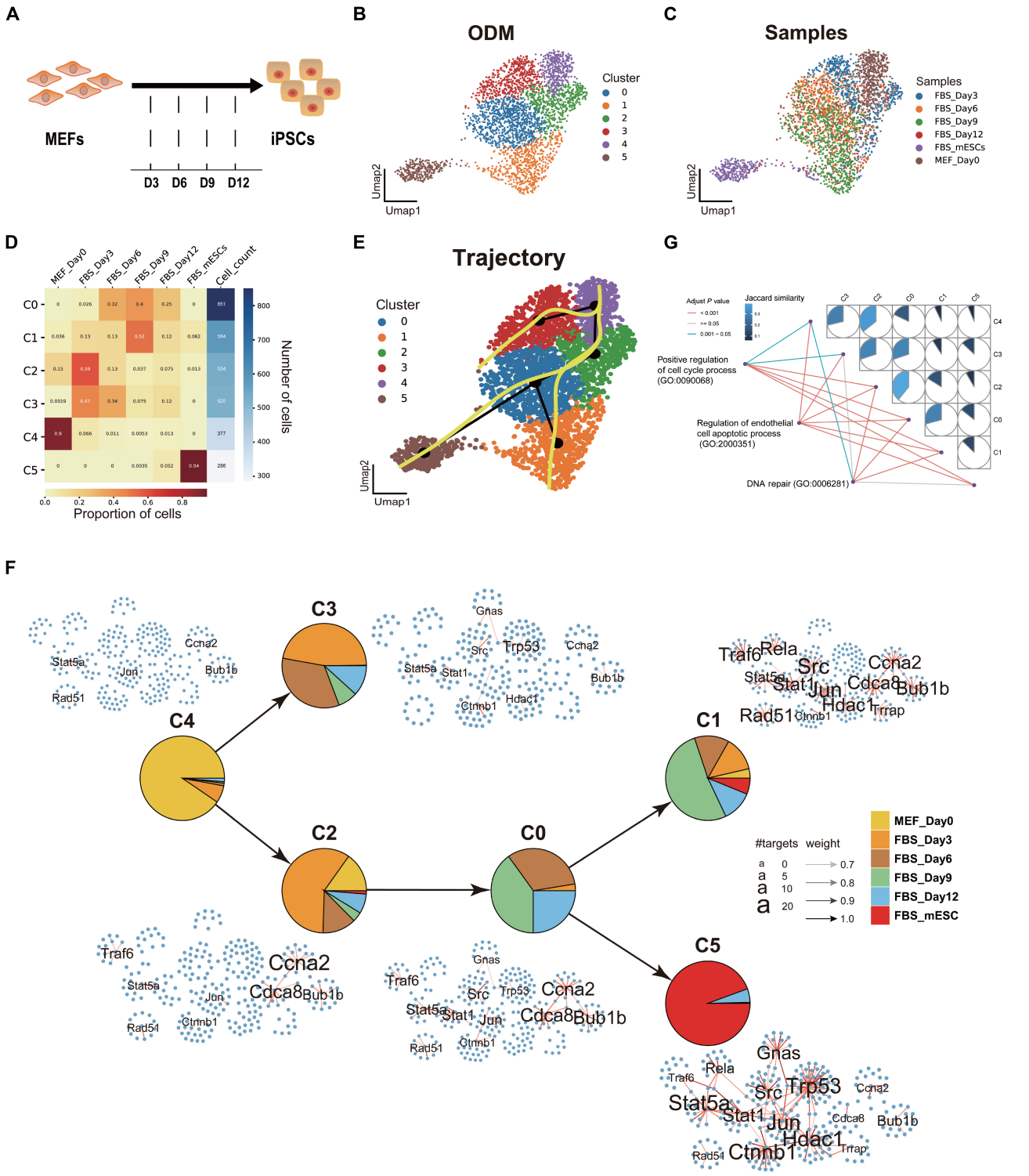


Fig. 4. Constructing the cell type-specific networks to uncover the key regulators during reprogramming. (A) Diagram illustrating the mouse cellular reprogramming process and the sampling time point for the datasets. (B) UMAP of cell clusters on the ODM data of the data from mouse cellular reprogramming. (C) UMAP of cell clusters on the sample labels of the data from mouse cellular reprogramming. (D) Distribution of the sample labels in each cell cluster. (E) Inferred cell lineage trajectory for the 6 cell clusters with ODM data. (F) Cell lineage structure constructed through the inferred trajectory. The layout of each network is the same; edges present in a particular cell cluster are shown in red. Labeled nodes correspond to regulators with different targets. (G) Gene Ontology (GO) enrichment analysis on the regulators identified from each refined cell cluster-specific network. Jaccard similarity was calculated among these cell clusters with respective cell cluster-specific networks.

PI3K/AKT signaling pathway, is essential for B cell receptor (BCR) signaling, which is crucial for B cell development and function [79]. Furthermore, *TRAF6* contributes to the activation of nuclear factor-κB (NF-κB) and AP-1 transcription factors during $CD4^+$ T cell development [80,81]. These findings indicate that the SiCNet method can accurately identify the

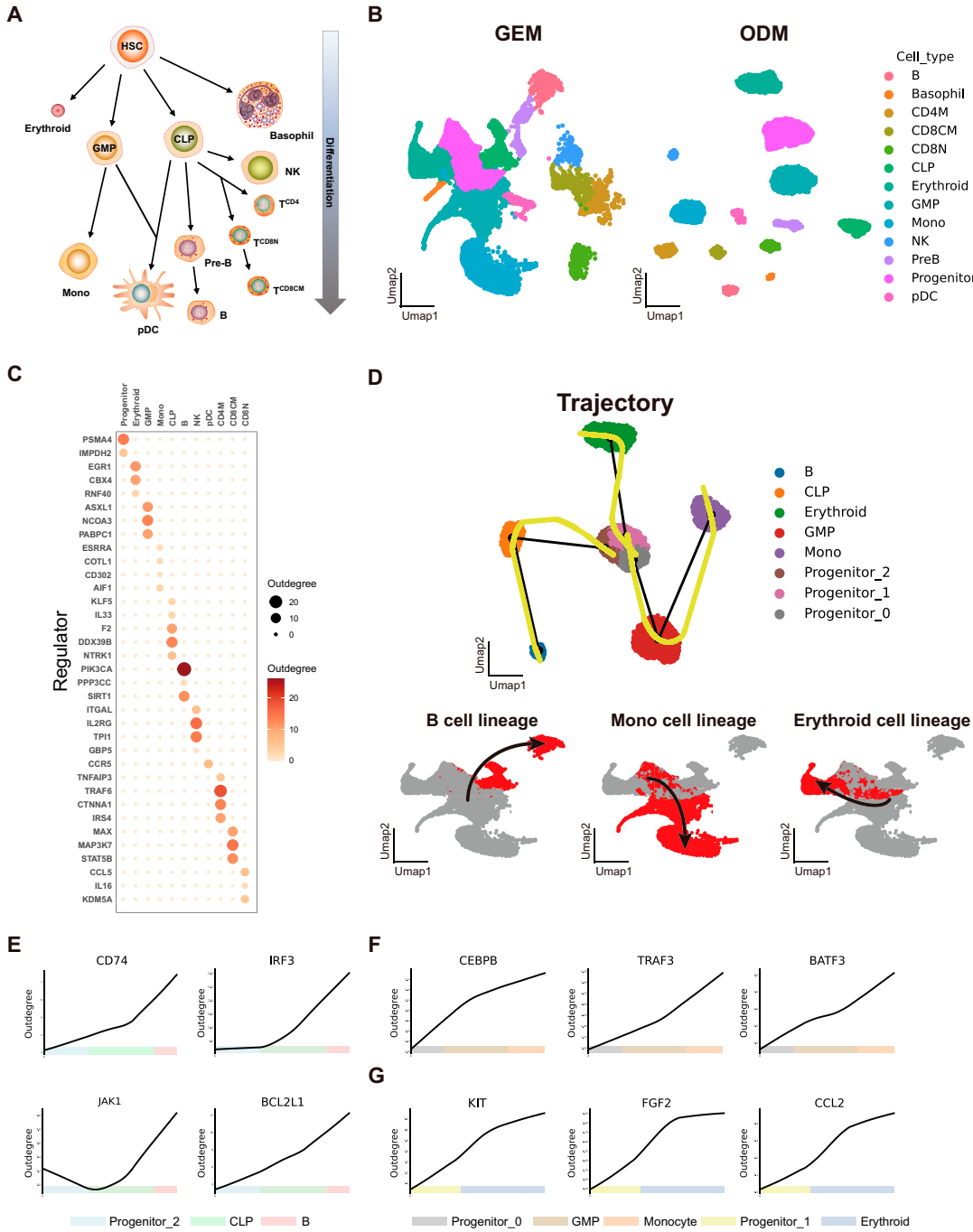


Fig. 5. Identifying the developmental tendency of progenitor cells and discovering the dynamics of regulation in human hematopoiesis. (A) Hierarchical cartoon representation of the human hematopoietic cell types identified in scRNA-seq profiles of bone marrow mononuclear cells (BMMC) from healthy donors. pDC, plasmacytoid dendritic cell; Pre-B, pre-B cell; CLP, common lymphoid progenitor; GMP, granulocyte–monocyte progenitor; NK, natural killer cell; TCD4, $CD4^+$ T cell; TCD8N, $CD8^+$ naive T cell; TCD8CM, $CD8^+$ central memory T cell; TCD4M, memory $CD4^+$ T cells. (B) UMAP of the human hematopoietic cell types identified in GEM and ODM data. (C) Top regulator identified from each refined cell type-specific network based on target count. (D) Inferred potential development trajectories are divided into 3 lineages: B cell lineage, monocyte lineage, and erythroid cell lineage. (E) Examples of regulators with activated patterns of regulatory activity found in B cell lineage. The blue–green–red strip represents the development trajectories of progenitor–CLP–B lineage. (F) Examples of regulators with activated patterns of regulatory activity found in monocyte lineage. The gray–brown–orange strip represents the development trajectories of progenitor–GMP–Mono lineage. (G) Examples of regulators with activated patterns of regulatory activity found in erythroid cell lineage. The yellow–violet strip represents the development trajectories of progenitor–erythroid lineage.

specific regulatory genes in the differentiation and maturation of various immune cell types in human hematopoiesis.

Rather than focusing on the static regulation, we place more emphasis on the dynamic changes during the developmental process. By applying the SiCNet method, we can precisely model the dynamic regulations involved in hematopoietic development. This approach allows us to capture transient states and regulatory shifts, which provides a clearer and more comprehensive understanding of the underlying mechanisms. For hematopoietic progenitor cells, we consider that their

developmental regulation may perform directional tendencies. We classified the progenitor cells into distinct subtypes and used Slingshot [68] to infer the developmental trajectories of these subtypes toward 3 lineages: B cells, monocytes, and erythroid cells (Fig. 5D). Using these cells along the lineage as a representation of developmental order, we can investigate the dynamic regulations of the development-associated regulators and uncover potential regulatory processes. For each regulator, its regulatory activity along the developmental lineage can be represented as a curve showing its changes from progenitor cells to a specific cell type. This curve is fitted based on the number of target genes regulated by the regulator in each cell along the lineage.

In the lineage of B cell development (Fig. 5E, Fig. S8, and Table S5), we found a series of increased activity of regulators. For example, *CD74*, a cell surface receptor for the cytokine macrophage migration inhibitory factor (MIF), plays a critical role in B cell development [82]. Our findings suggest that the number of target genes regulated by *CD74* progressively increases from progenitor cells to mature B cells. This observation is consistent with previous studies, which demonstrate that alterations in *CD74* expression significantly promote B cell differentiation and survival [83]. Given these characteristics, we believe that *CD74* may play a significant role in directing hematopoietic stem cells (HSCs) to differentiate into the B cell lineage by establishing new regulatory pathways. Similarly, *IRF3* is known to play central roles in cell differentiation and development, as well as in signaling downstream of various pathways, such as the Toll-like receptor (TLR) and BCR signaling pathways [84]. Importantly, BCR signaling is crucial for regulating B cell activation and survival, contributing to a balanced immune response [85]. The increase in *IRF3* regulation may integrate various signals from the BCR to promote the differentiation of naïve B cells into mature forms, such as plasma cells and memory B cells, which is essential for effective immune function [86]. In addition, *JAK1* is essential for maintaining the survival of mature B cells in peripheral tissues, ensuring effective immune responses and the generation of memory B cells. In the early stages of B cell development within the bone marrow, the increase of *JAK1* activation promotes the proliferation and differentiation of pro-B cells into pre-B cells by transducing signals from both the BCR and cytokines [87]. *BCL2L1*, also known as *BCL-XL*, plays a critical role in B cell development by protecting cells from programmed cell death. It helps promote the survival of B cells at various stages, particularly during the transitional and mature phases [88,89]. The increased activity of *BCL2L1* also promotes B cell maturation in a highly regulated manner [90].

In the lineage of monocyte cell development (Fig. 5F, Fig. S9, and Table S6), we identified several potential regulators that may be involved in promoting the survival and proliferation of monocytes during myeloid lineage differentiation. Specifically, *CEBPB* modulates the NF-κB signaling pathway to promote inflammatory responses, further facilitating the generation and functional maturation of monocytes [91–93]. Its increased regulation is required for the terminal differentiation and maturation of granulocytes and monocytes [94]. Moreover, the upregulation of *TRAF3* is essential for monocyte activation, as it plays a pivotal role in cytokine signaling and cell survival, thereby facilitating the differentiation and immune function of monocytes [95,96]. In addition, we found that *BATF3* has a high activity during the development process, and it may play a pivotal role in driving the differentiation and functional maturation of monocytes. This suggests that *BATF3* may regulate critical genes involved in immune responses, contributing to the efficient transition of progenitor cells into fully developed monocytes. Furthermore, *BATF3* may influence the ability of these monocytes to mediate immune functions, such as antigen presentation and cytokine production.

In the lineage of erythroid cell development (Fig. 5G, Fig. S10, and Table S7), we found that regulators like *KIT*, *FGF2*, and *CCL2* exhibit up-regulated activity, suggesting their essential roles in promoting erythropoiesis. *KIT* is a receptor tyrosine kinase that plays a critical role in erythroid progenitor cell survival and differentiation. Binding its ligand, stem cell factor (SCF), *KIT* activates downstream signaling pathways such as the PI3K/AKT and MAPK/extracellular signal–regulated kinase (ERK) pathways. These pathways promote cell proliferation and prevent apoptosis during early erythroid development [97,98]. The increase of *KIT* activation is also proved to modulate terminal erythroid differentiation and survival [99]. *FGF2* is the prototype member of the fibroblast growth factor (FGF) family and interacts with its receptors (FGFRs) to mediate the receptor phosphorylation and the activation of signaling pathways, such as RAS/MAPK and PI3K/AKT pathways, which are involved in cell proliferation and differentiation [100]. In addition, it can enhance the formation and maintenance of blood cells [101,102]. Surprisingly, our findings suggest that *CCL2* may play a potential regulatory role in erythroid development. A previous study has shown that *CCL2* interacts with *ACKR1* to maintain an important physiological role in the erythroid lineage [103].

In summary, the SiCNet method effectively identifies known regulatory mechanisms while also uncovering potential regulatory insights within hematopoietic development. This capability enhances our challenge to explore both established and novel regulatory pathways, thereby advancing our knowledge of hematopoiesis and its associated mechanisms.

Discussion

In this paper, we present SiCNet, a method for causal network inference, to construct cell-specific network and analyze dynamic regulations from single-cell transcriptome data. The SiCNet method integrates causal inference principles and prior knowledge-based networks to predict gene-level causal relationships in individual cells. Considering the variability in gene expression data and the robustness of gene regulation, we transform cell-specific networks into ODM for downstream dimension reduction and clustering analysis, which allows us to identify cell populations with greater accuracy than GEM. Importantly, we can use the SiCNet method to uncover regulatory insights at the network level in cancer progression, cellular reprogramming, and hematopoiesis. These regulations cannot be captured by current single-cell analysis, which focused solely on gene expression changes. In contrast, our SiCNet method measures the biological effects of each gene within the context of the cell-specific network. Furthermore, we capture dynamic regulatory changes at the single-cell level by focusing on individual cell differences rather than population-level trends. Notably, in addition to modeling at the single-cell level, our SiCNet method can also be applied to single samples to construct sample-specific networks. Moreover, SiCNet can be extended to spatial transcriptomics data, offering a powerful tool for capturing spatially dependent regulatory information that is crucial for understanding tissue architecture and function (Note S8).

GRN inference has long been a fundamental challenge in biology, with many unresolved issues, such as handling high-dimensional data with insufficient sample sizes, constructing networks from multiple samples, and addressing the complexities of time-series data, including temporal dependencies and feedback mechanisms. The SiCNet method incorporates causal inference while constructing networks for individual cells, offering a solution to some of these challenges. However, our approach does not account for nonlinearity, non-steady-state conditions, or mechanistic models, nor does it integrate other potentially informative data modalities, such as unspliced mRNAs, DNA methylation, chromatin accessibility, or cell lineage tracking. Despite these limitations, SiCNet assumes steady states and still uncovers rich biological information in non-steady-state developmental systems.

Single-cell technologies have revolutionized GRN inference and analysis, enabling comparisons at the level of individual cells across multiple modalities. Methods like SiCNet, which focus on single-cell-specific GRN inference, can reveal cellular heterogeneity and quantify gene regulation with higher throughput and resolution. This approach offers a broader perspective than individual gene expression analysis and paves the way for deeper biological discoveries and investigations.

Methods

Details of SiCNet algorithm

SiCNet is a data-driven and model-free algorithm for inferring gene-level causal relationships at the single-cell level (Fig. 1A). This method operates in 2 steps: (a) constructing a reference causal network from a designated reference dataset, and (b) inferring cell-specific causal networks for individual cells based on the reference network. The reference network serves as the benchmark to capture general causal relationships in gene regulation, while the cell-specific causal network utilizes the gene expression profile of an individual cell in combination with the reference network to infer causal relationships between genes.

Reference causal network construction

The first step is to infer the initial causal network, also referred to as the reference causal network, from a designated reference dataset (Fig. 1B). This dataset can be any single-cell gene expression data, but to make the results more meaningful, it is recommended to use data relevant to the research objectives. For example, in cancer research, it could correspond to normal adjacent tissue data, while in developmental studies, it might refer to progenitor cell-related data. The cell types within this set may consist of either a single-cell type or a mixture of multiple cell types. To enhance the interpretability of the inferred networks, we integrated the PPI network as prior knowledge into the inference process. We selected a gene g^r and its first-order neighbor genes $g_i^r \in \{g_1^r, g_2^r, \cdots, g_n^r\}$, as a unit (or subnetwork of g^r) (Fig. 1B). The r in g^r refers to "reference", indicating the gene expression profile of g^r comes from the reference dataset. We assume that all these neighbors of gene $g^r$are the cause of g^r, implying that g^r is regulated by these neighbor genes. Then, we inferred the causal relationship between g^r and each g_i^r, respectively. In detail, we utilized a linear regression model to represent the relationship between g^r and the presence or absence of g_i^r, as shown in Eqs. 1 and 2.

$$g^r = f_1\left(g_i^r, Z^r\right) + \varepsilon_1^r = \sum_{j=1}^{n} \alpha_j g_j^r + \varepsilon_1^r \tag{1}$$

$$g^r = f_2\left(Z^r\right) + \varepsilon_2^r = \sum_{j=1, j\neq i}^{n} \beta_j g_j^r + \varepsilon_2^r \tag{2}$$

Here, the functions $f_1\left(g_i^r, Z^r\right)$ and $f_2(Z^r)$ model the relationships between g^r and its neighbors, with the key difference being whether g_i^r is present or not. Accordingly, Z^r represents the set of all first-order neighbor genes of g^r, excluding g_i^r. The parameters α_j and β_j denote the regression coefficients corresponding to Eqs. 1 and 2, respectively. The residuals ε_1^r and ε_2^r represent the RMSE of the linear model corresponding to Eqs. 1 and 2, respectively. Notably, to ensure more reliable results, we applied 5-fold cross-validation on the samples (cells). Particularly, the dataset was randomly split into 5 approximately equal-sized subsets. In each iteration, one subset was used as the validation set, while the remaining 4 subsets were used as the training set. The linear regression model was trained on the training set and tested on the validation set to compute the loss values. This process was repeated for all 5 subsets, and the losses for each model were averaged across the 5 iterations to obtain the final loss values, which were referred to as ε_1^r and ε_2^r.

The causal relationship between g_i^r and g^r can be quantified by a CSI $\omega_r\left(g_i^r, g^r\right)$, which is described in Eq. 3.

$$\omega_r\left(g_i^r, g^r\right) = \ln\left(\frac{\varepsilon_2^r}{\varepsilon_1^r}\right) \tag{3}$$

The positive ω_r indicates the existence of a causal relationship from the gene g_i^r to the gene g^r; otherwise, the absence of this causal relationship is inferred. Thus, by determining the sign of the CSI for each g_i^r to g^r, we construct a reference causal network that serves as the benchmark for predicting cell-specific causal networks in the next step.

Cell-specific causal network inference

The next step is to infer cell-specific causal networks for individual cells based on the reference causal network (Fig. 1C). Similarly, we selected a gene g^c and its first-order cause (regulatory) genes $g_i^c \in \{g_1^c, g_2^c, \ldots, g_m^c\}$ from the reference causal network as a unit (or subnetwork of g^c). The c in g^c refers to "specific cell", indicating that the gene expression profile of g^c comes from the particular cell. We used a set of gene expression of g_i^c to estimate the expression of g^c in each individual cell, as shown in Eq. 4. Then, we estimated the expression of g^c when the corresponding cause (regulatory) gene g_i^c was absent, as shown in Eq. 5.

$$g^c = f_1\left(g_i^c, Z^c\right) + \varepsilon_1^c = \sum_{j=1}^{n} \alpha_j' g_j^c + \varepsilon_1^c \tag{4}$$

$$g^c = f_2\left(Z^c\right) + \varepsilon_2^c = \sum_{j=1, j\neq i}^{n} \beta_j' g_j^c + \varepsilon_2^c \tag{5}$$

Here, $f_1\left(g_i^c, Z^c\right)$ and $f_2(Z^c)$ represent the estimated functions derived from the reference causal network. Accordingly, Z^c represents the set of all first-order cause (regulatory) genes of g^c, excluding g_i^c. The parameters α_j' and β_j' denote the regression

coefficients corresponding to Eqs. 4 and 5, respectively. The residuals ε_1^c and ε_2^c denote the RMSE between the actual expression of g^c and its corresponding predicted values $f_1\left(g_i^c, Z^c\right)$ and $f_2(Z^c)$, respectively.

Similarly, the causal relationship between g_i^c and g^c can be determined by the sign of $\omega_c\left(g_i^c, g^c\right)$ within the single cell, as shown in Eq. 6. If the value of ω_c exceeds zero, it indicates the existence of a causal influence of g_i^c on g^c; otherwise, no such causal relationship exists.

$$\omega_c\left(g_i^c, g^c\right) = \ln\left(\frac{\varepsilon_2^c}{\varepsilon_1^c}\right) \tag{6}$$

By performing causal predictions for each gene pair (g_i^c to g^c) in every individual cell, we construct cell-specific causal networks.

Background network

To obtain a background network as the basis for initial causal network construction, we downloaded the PPI network from the STRING database version 12.0 [104] (https://string-db.org/) and filtered out the edges with associated scores below 700.

Network binarization

To simplify the calculation and subsequent analyses, we binarized both the reference causal network and all cell-specific networks. Thus, for the edge of gene g_i to gene g_j ($i \neq j$) in the network, we have:

$$edge_{g_i \to g_j} = \begin{cases} 0, & \omega \leq 0 \\ 1, & \omega > 0 \end{cases} \tag{7}$$

Network ODM from SiCNet method

Cell-specific networks constructed by SiCNet can be applied to various biological studies at the network level. However, the number of features described in a network in most scRNA-seq analyses is quite large. For instance, there are at least 100,000 gene pairs or direct edges supported by the STRING PPI network. In this paper, we transformed the cell-specific network from SiCNet method into a network ODM to capture network features while simultaneously reducing dimensionality, although SiCNets can also be directly used or reduced through other methods for clustering analysis. Thus, for gene g_i in the network of cell k, we have:

$$ODM_{g_i}^k = \sum_{i=1, i \neq j}^{m} edge_{g_i \to g_j} \tag{8}$$

Then, we can get a matrix ODM with $m \times n$ elements, where m represents the number of genes and n represents the number of cells. g_i denotes the regulator in the network, and g_j represents the regulated gene (target gene). In the ODM, the count of g_j is considered as the outdegree of g_i. We use the NetworkX Python package to identify the outdegree of the network in each cell.

The ODM retains the same dimensions as the original GEM, with the same number of rows and columns, but it represents the higher-order information of each gene within the network rather than the gene expression levels. This matrix can be seamlessly analyzed using any standard scRNA-seq algorithm for cell clustering, dimensionality reduction, and pseudo-trajectory analysis by simply substituting the original GEM with our ODM. Thus, our SiCNet method introduces a novel approach for analyzing scRNA-seq data at the network level.

Benchmark datasets

In this work, we collected 5 single-cell datasets to demonstrate the advantages of our method in cell clustering. The cell type annotations were derived from the original publications, where they were determined by the authors using their biological expertise. Detailed descriptions and sources of all datasets are provided below.

The Biase dataset contains 49 cells with 3 types: 2-cell mouse embryos, 4-cell mouse embryos, and zygotes [105].

Enge dataset contains pancreatic cells from 8 human volunteers, comprising a total of 2,282 cells across 6 distinct cell types [106].

Darmanis dataset is a human brain dataset characterized by its complexity, comprising 7 different cell types and 331 cells [107].

Fink dataset is a total of 3,045 single-cell transcriptomes of stromal cells from human adult ureters, encompassing 7 distinct cell types [108].

Goolam dataset consists of sequencing results from mouse embryo cells including 5 cell types and 124 cells [109].

All these datasets can be obtained at the Data Availability section.

Student's *t* test

The "scipy.stats.ttest_ind" function in Python was employed to implement Student's t test for the different outdegrees or gene expression between 2 groups, such as the healthy adjacent region and the core region in Fig. 3. The P values were adjusted for multiple testing using the FDR method.

Chi-square test

The "scipy.stats.chi2_contingency" function in Python was employed to implement chi-square test on the rate of activity change for each edge from the healthy adjacent region to the core region in Fig. 3, with P values adjusted for multiple testing using FDR.

Evaluation metrics

ARI [110], AMI [111], NMI [112], and FMI [113] are used as metrics to evaluate the clustering performance.

The ARI measures the similarity between 2 cluster sets, adjusting for the chance grouping of elements. It considers all pairs of samples and counts pairs that are assigned in the same or different clusters in the predicted and true clusters. The ARI is defined as:

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \frac{\left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}}{0.5\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \frac{\left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right]}{\binom{n}{2}}} \tag{9}$$

where n is the total number of samples, n_{ij} is the number of samples in both cluster i and cluster j, and a_i and b_j are the sum of elements in cluster i of the true clustering and cluster j of the predicted clustering, respectively.

AMI adjusts the mutual information score for chance, providing a normalized measure of similarity between 2 cluster sets. It considers how much information is shared between the 2 cluster sets, accounting for the expected mutual information between 2 random cluster sets. The AMI is defined as:

$$AMI = \frac{MI(U,V) - \mathbb{E}[MI(U,V)]}{max(H(U),H(V)) - \mathbb{E}[MI(U,V)]} \quad (10)$$

where $MI(U,V)$ is the mutual information between the true labels U and predicted labels V, $H(U)$ and $H(V)$ are the entropies of U and V, and $\mathbb{E}[MI(U,V)]$ is the expected mutual information.

NMI measures the similarity between 2 cluster sets, normalized to scale between 0 (no mutual information) and 1 (perfect correlation). It is defined as:

$$NMI = \frac{MI(U,V)}{\sqrt{H(U) \cdot H(V)}} \quad (11)$$

where $MI(U,V)$ is the mutual information between the true labels U and predicted labels V, and $H(U)$ and $H(V)$ are the entropies of U and V.

The FMI is a geometric mean of precision and recall used for evaluating the quality of clustering by comparing the pairwise similarity between the true and predicted clusters. It is defined as:

$$FMI = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}} \quad (12)$$

where TP (true positive) is the number of pairs correctly clustered together, FP (false positive) is the number of pairs incorrectly clustered together, and FN (false negative) is the number of pairs incorrectly not clustered together.

Acknowledgments

Funding: This study was supported by the National Natural Science Foundation of China (nos. T2341007, T2350003, 12131020, 42450084, 42450135, 12326614, and 12426310), the Zhejiang Provincial Natural Science Foundation of China (no. LZ22C060001), the research funds of Hangzhou Institute for advanced study, UCAS (nos. 2024HIAS-Y016 and 2024HIAS-P004), the National Key Research and Development Program of China (2022YFA1004800, 2025YFF1207900), TianYuan Special Funds of the National Natural Science Foundation of China (no. 12026607), The Zhejiang Province Vanguard Goose-Leading Initiative (no. 2025C01114), and JST Moonshot R&D (no. JPMJMS2021).

Author contributions: X.L. conceived the research. X.H. performed the real data analysis. X.H. wrote the manuscript, and X.L. and L.C. revised the manuscript. All authors read and approved the final manuscript.

Competing interests: The authors declare that they have no competing interests.

Data Availability

The benchmark datasets in this work were collected from the NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) database with accession numbers GSE57249, GSE81547, GSE67835 and GSE184111 and the ArrayExpress database, hosted by EMBL-EBI (https://www.ebi.ac.uk/), with accession number E-MTAB-3321. The single-cell RNA-Seq data for colorectal cancer were obtained from GEO with accession number GSE144735. The mouse cellular reprogramming scRNA-seq data were accessed through GEO with accession number GSE108222. The human blood scRNA-seq data were available through GEO with accession number GSE139369. The source code of the SiCNet algorithm is available at https://github.com/Huang-XZ-Sandy/SiCNet.

Supplementary Materials

Supplementary Notes S1 to S10
Figs. S1 to S24
Tables S1 to S8

References

1. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, Zappia L, Lücken MD, Strobl DC, Henao J, Curion F, et al. Best practices for single-cell analysis across modalities. *Nat Rev Genet*. 2023;24(8):550–572.
2. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med*. 2022;12(3):Article e694.
3. Hou W, Ji Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *Nat Methods*. 2024;21(8):1462–1465.
4. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol*. 2018;14(8):479–492.
5. Saadatpour A, Lai S, Guo G, Yuan G-C. Single-cell analysis in cancer genomics. *Trends Genet*. 2015;31(10):576–586.
6. Jiao L, Wang Y, Liu X, Li L, Liu F, Ma W, Guo Y, Chen P, Yang S, Hou B. Causal inference meets deep learning: A comprehensive survey. *Research*. 2024;7:Article 0467.
7. Wen Y, Huang J, Guo S, Elyahu Y, Monsonego A, Zhang H, Ding Y, Zhu H. Applying causal discovery to single-cell analyses using CausalCell. *eLife*. 2023;12:Article e81464.
8. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C, Kannan S. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Syst*. 2020;10(3):265–274.e11.
9. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE*. 2010;5(9):Article e12776.
10. Moerman T, Aibar Santos S, Bravo González-Blas C, Simm J, Moreau Y, Aerts J, Aerts S. GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics*. 2019;35(12):2159–2161.
11. Chan TE, Stumpf MP, Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst*. 2017;5(3):251–267.e3.
12. Liu X, Wang Y, Ji H, Aihara K, Chen L. Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res*. 2016;44(22):Article e164.

13. Huang Y, Chang X, Zhang Y, Chen L, Liu X. Disease characterization using a partial correlation-based sample-specific network. *Brief Bioinform*. 2021;22(3):Article bbaa062.
14. Zhang Y, Li Q, Wang J, Chang X, Chen L, Liu X. Causal network inference based on cross-validation predictability. *Commun Phys*. 2025;8(1):173.
15. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9(11):2579–2605.
16. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, et al. SC3: Consensus clustering of single-cell RNA-seq data. *Nat Methods*. 2017;14(5):483–486.
17. Lin P, Troup M, Ho JW. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol*. 2017;18:Article 59.
18. Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol*. 2018;19(1):Article 15.
19. Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat Methods*. 2017;14(4):414–416.
20. Bilous M, Tran L, Cianciaruso C, Gabriel A, Michel H, Carmona SJ, Pittet MJ, Gfeller D. Metacells untangle large and complex single-cell transcriptome networks. *BMC Bioinformatics*. 2022;23(1):Article 336.
21. Lee H-O, Hong Y, Etlioglu HE, Cho YB, Pomella V, Van den Bosch B, Vanhecke J, Verbandt S, Hong H, Min J-W, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet*. 2020;52(6):594–603.
22. Dekker E, Tanis PJ, Vleugels JL, Kasi PM, Wallace MB. Colorectal cancer. *Lancet*. 2019;394(10207):1467–1480.
23. Malki A, ElRuz RA, Gupta I, Allouch A, Vranic S, Al Moustafa A-E. Molecular mechanisms of colon cancer progression and metastasis: Recent insights and advancements. *Int J Mol Sci*. 2020;22(1):Article 130.
24. Nunes L, Li F, Wu M, Luo T, Hammarström K, Torell E, Ljuslinder I, Mezheyeuski A, Edqvist P-H, Löfgren-Burström A, et al. Prognostic genome and transcriptome signatures in colorectal cancers. *Nature*. 2024;633(8028):137–146.
25. Sonoshita M, Itatani Y, Kakizaki F, Sakimura K, Terashima T, Katsuyama Y, Sakai Y, Taketo MM. Promotion of colorectal cancer invasion and metastasis through activation of NOTCH–DAB1–ABL–RHOGEF protein TRIO. *Cancer Discov*. 2015;5(2):198–211.
26. Luttman JH, Colemon A, Mayro B, Pendergast AM. Role of the ABL tyrosine kinases in the epithelial–mesenchymal transition and the metastatic cascade. *Cell Commun Signal*. 2021;19(1):Article 59.
27. Guo P, Chen Q, Peng K, Xie J, Liu J, Ren W, Tong Z, Li M, Xu J, Zhang Y, et al. Nuclear receptor coactivator SRC-1 promotes colorectal cancer progression through enhancing GLI2-mediated hedgehog signaling. *Oncogene*. 2022;41(20):2846–2859.
28. Wu Z-Y, Wei Z-M, Sun S-J, Yuan J, Jiao S-C. Activating transcription factor 3 promotes colon cancer metastasis. *Tumour Biol*. 2014;35(8):8329–8334.
29. Lu X, Zhong L, Lindell E, Veanes M, Guo J, Zhao M, Salehi M, Swartling FJ, Chen X, Sjöblom T, et al. Identification of ATF3 as a novel protective signature of quiescent colorectal tumor cells. *Cell Death Dis*. 2023;14(10):Article 676.
30. Wang H, Wen C, Chen S, Li W, Qin Q, He L, Wang F, Chen J, Ye W, Li W, et al. ROS/JNK/C-Jun pathway is involved in chaetocin induced colorectal cancer cells apoptosis and macrophage phagocytosis enhancement. *Front Pharmacol*. 2021;12:Article 729367.
31. Fang JY, Richardson BC. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol*. 2005;6(5):322–327.
32. Regan JL, Schumacher D, Staudte S, Steffen A, Lesche R, Toedling J, Jourdan T, Haybaeck J, Golob-Schwarzl N, Mumberg D, et al. Identification of a neural development gene expression signature in colon cancer stem cells reveals a role for *EGR2* in tumorigenesis. *iScience*. 2022;25(7):Article 104498.
33. La Vecchia S, Sebastián C. Metabolic pathways regulating colorectal cancer initiation and progression. *Semin Cell Dev Biol*. 2020;98:63–70.
34. Jayakrishnan T, Mariam A, Farha N, Rotroff DM, Aucejo F, Barot SV, Conces M, Nair KG, Krishnamurthi SS, Schmit SL, et al. Plasma metabolomic differences in early-onset compared to average-onset colorectal cancer. *Sci Rep*. 2024;14(1):Article 4294.
35. Cheng HS, Yip YS, Lim EKY, Wahli W, Tan NS. PPARs and tumor microenvironment: The emerging roles of the metabolic master regulators in tumor stromal–epithelial crosstalk and carcinogenesis. *Cancers*. 2021;13(9):Article 2153.
36. You M, Yuan S, Shi J, Hou Y. PPARδ signaling regulates colorectal cancer. *Curr Pharm Des*. 2015;21(21):2956–2959.
37. Yuan C, Bao Y, Sato K, Nimptsch K, Song M, Brand-Miller JC, Morales-Oyarvide V, Zoltick ES, Keum N, Wolpin BM, et al. Influence of dietary insulin scores on survival in colorectal cancer patients. *Br J Cancer*. 2017;117(7):1079–1087.
38. Mercogliano MF, Bruni S, Mauro F, Elizalde PV, Schillaci R. Harnessing tumor necrosis factor alpha to achieve effective cancer immunotherapy. *Cancers*. 2021;13(3):Article 564.
39. Al Obeed OA, Alkhayal KA, Al Sheikh A, Zubaidi AM, Vaali-Mohammed M-A, Boushey R, Mckerrow JH, Abdulla M-H. Increased expression of tumor necrosis factor-α is associated with advanced colorectal cancer stages. *World J Gastroenterol*. 2014;20(48):18390–18396.
40. Tang Y, Weng X, Liu C, Li X, Chen C. Hypoxia enhances activity and malignant behaviors of colorectal cancer cells through the STAT3/microRNA-19a/PTEN/PI3K/AKT axis. *Anal Cell Pathol*. 2021;2021(1):Article 4132488.
41. Fletcher T, Thompson AJ, Ashrafian H, Darzi A. The measurement and modification of hypoxia in colorectal cancer: Overlooked but not forgotten. *Gastroenterol Rep*. 2022;10:Article goac042.
42. Rubio MF, Lira MC, Rosa FD, Sambresqui AD, Salazar Güemes MC, Costas MA. RAC3 influences the chemoresistance of colon cancer cells through autophagy and apoptosis inhibition. *Cancer Cell Int*. 2017;17:Article 111.
43. Panelo LC, Machado MS, Rubio MF, Jaworski F, Alvarado CV, Paz LA, Urtreger AJ, Vazquez E, Costas MA. High RAC3 expression levels are required for induction and maintaining of cancer cell stemness. *Oncotarget*. 2017;9(5):5848–5860.
44. Nie K, He Z-J, Kong L-J. NR3C2 affects the proliferation and invasiveness of colon cancer cells through the Wnt/β-catenin signaling pathway. *J Cancer Res Clin Oncol*. 2024;150(9):Article 411.
45. Cervantes-Madrid DL, Nagi S, Asting Gustafsson A. FosB transcription factor regulates COX-2 expression in colorectal cancer cells without affecting PGE2 expression. *Oncol Lett*. 2017;13(3):1411–1416.

46. Dastjerdi S, Haghparast A, Amroabadi JM, Dolatabadi NF, Mirzaei S, Zamani A, Hashemi M, Mahdevar M, Ghaedi K. Elevated CDK5R1 expression associated with poor prognosis, proliferation, and drug resistance in colorectal and breast malignancies: CDK5R1 as an oncogene in cancers. *Chem Biol Interact*. 2022;368:Article 110190.
47. Koveitypour Z, Panahi F, Vakilian M, Peymani M, Seyed Forootan F, Nasr Esfahani MH, Ghaedi K. Signaling pathways involved in colorectal cancer progression. *Cell Biosci*. 2019;9(1):Article 97.
48. Wang N, Cao Y, Si C, Shao P, Su G, Wang K, Bao J, Yang L. Emerging role of ERBB2 in targeted therapy for metastatic colorectal cancer: Signaling pathways to therapeutic strategies. *Cancers*. 2022;14(20):Article 5160.
49. Song Y, Chen M, Wei Y, Ma X, Shi H. Signaling pathways in colorectal cancer implications for the target therapies. *Mol Biomed*. 2024;5(1):Article 21.
50. Farhan M, Wang H, Gaur U, Little PJ, Xu J, Zheng W. FOXO signaling pathways as therapeutic targets in cancer. *Int J Biol Sci*. 2017;13(7):815–827.
51. Aihara K, Liu R, Koizumi K, Liu X, Chen L. Dynamical network biomarkers: Theory and applications. *Gene*. 2022;808:Article 145997.
52. Liu X, Chang X, Leng S, Tang H, Aihara K, Chen L. Detection for disease tipping points by landscape dynamic network biomarkers. *Natl Sci Rev*. 2019;6(4):775–785.
53. Grath A, Dai G. Direct cell reprogramming for tissue engineering and regenerative medicine. *J Biol Eng*. 2019;13(1):Article 14.
54. Zhang S, Pyne S, Pietrzak S, Halberg S, McCalla SG, Siahpirani AF, Sridharan R, Roy S. Inference of cell type-specific gene regulatory networks on cell lineages from single cell omic datasets. *Nat Commun*. 2023;14(1):Article 3064.
55. Silva-Pedrosa R, Salgado AJ, Ferreira PE. Revolutionizing disease modeling: The emergence of organoids in cellular systems. *Cells*. 2023;12(6):Article 930.
56. Cerneckis J, Cai H, Shi Y. Induced pluripotent stem cells (iPSCs): Molecular mechanisms of induction and applications. *Signal Transduct Target Ther*. 2024;9(1): Article 112.
57. Traag VA, Waltman L, Van Eck NJ. From Louvain to Leiden: Guaranteeing well-connected communities. *Sci Rep*. 2019;9(1):Article 5233.
58. de Lima Camillo LP, Quinlan RB. A ride through the epigenetic landscape: Aging reversal by reprogramming. *GeroScience*. 2021;43(2):463–485.
59. Zeng N, Tang W, Wu Y, Fan H, Xie S, Cao N. Harnessing stem cell and lineage reprogramming technology to treat cardiac fibrosis. *Cell Regen*. 2023;12(1):Article 39.
60. Izawa K, Yamazaki S, Becker HJ, Bhadury J, Kakegawa T, Sakaguchi M, Tojo A. Activated HoxB4-induced hematopoietic stem cells from murine pluripotent stem cells via long-term programming. *Exp Hematol*. 2020;89:68–79.e7.
61. Monterisi S, Lo Riso P, Russo K, Bertalot G, Vecchi M, Testa G, Di Fiore PP, Bianchi F. HOXB7 overexpression in lung cancer is a hallmark of acquired stem-like phenotype. *Oncogene*. 2018;37(26):3575–3588.
62. Wonders C, Anderson S. Beyond migration: Dlx1 regulates interneuron differentiation. *Nat Neurosci*. 2005;8(8):979–981.
63. Zhang S, Cui W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J Stem Cells*. 2014;6(3):305–311.
64. Tang Y, Tian X. JAK-STAT3 and somatic cell reprogramming. *JAKSTAT*. 2013;2(4):Article e24935.
65. Bhowmick NA, Chytil A, Plieth D, Gorska AE, Dumont N, Shappell S, Washington MK, Neilson EG, Moses HL. TGF-ß signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science*. 2004;303(5659):848–851.
66. Lin T, Lin Y. p53 switches off pluripotency on differentiation. *Stem Cell Res Ther*. 2017;8(1):Article 44.
67. Feng J, Zhu F, Ye D, Zhang Q, Guo X, Du C, Kang J. Sin3a drives mesenchymal-to-epithelial transition through cooperating with Tet1 in somatic cell reprogramming. *Stem Cell Res Ther*. 2022;13(1):Article 29.
68. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. 2018;19(1):Article 477.
69. Ebrahimi B. Reprogramming barriers and enhancers: Strategies to enhance the efficiency and kinetics of induced pluripotency. *Cell Regen*. 2015;4(1):Article 10.
70. Kuno A, Nishimura K, Takahashi S. Time-course transcriptome analysis of human cellular reprogramming from multiple cell types reveals the drastic change occurs between the mid phase and the late phase. *BMC Genomics*. 2018;19(1):Article 9.
71. Zheng P. Current understanding of genomic stability maintenance in pluripotent stem cells: Genomic stability maintenance in pluripotent stem cells. *Acta Biochim Biophys Sin*. 2022;54(6):858–863.
72. Chen ACH, Peng Q, Fong SW, Lee KC, Yeung WSB, Lee YL. DNA damage response and cell cycle regulation in pluripotent stem cells. *Genes*. 2021;12(10):1548.
73. Zhang J, Zhao L, Fu Y, Liu F, Wang Z, Li Y, Zhao G, Sun W, Wu B, Song Y, et al. Reprogramming efficiency and pluripotency of mule iPSCs over its parents. *Biol Reprod*. 2023;108(6):887–901.
74. Fu X, Wu S, Li B, Xu Y, Liu J. Functions of p53 in pluripotent stem cells. *Protein Cell*. 2020;11(1):71–78.
75. Cascio CL, McNamara JB, Melendez EL, Lewis EM, Dufault ME, Sanai N, Plaisier CL, Mehta S. Nonredundant, isoform-specific roles of HDAC1 in glioma stem cells. *JCI Insight*. 2021;6(17):Article e149232.
76. Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, Parks B, Gars E, Liedtke M, Zheng GX, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol*. 2019;37(12):1458–1465.
77. Wang L, Trasanidis N, Wu T, Dong G, Hu M, Bauer DE, Pinello L. Dictys: Dynamic gene regulatory network dissects developmental continuum with single-cell multiomics. *Nat Methods*. 2023;20(9):1368–1378.
78. Min IM, Pietramaggiori G, Kim FS, Passegué E, Stevenson KE, Wagers AJ. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell*. 2008;2(4):380–391.
79. Abdelrasoul H, Werner M, Setz CS, Okkenhaug K, Jumaa H. PI3K induces B-cell development and regulates B cell identity. *Sci Rep*. 2018;8(1):1327.
80. So T, Nagashima H, Ishii N. TNF receptor-associated factor (TRAF) signaling network in CD4+ T-lymphocytes. *Tohoku J Exp Med*. 2015;236(2):139–154.
81. Muto G, Kotani H, Kondo T, Morita R, Tsuruta S, Kobayashi T, Luche H, Fehling HJ, Walsh M, Choi Y, et al. TRAF6 is essential for maintenance of regulatory T

cells that suppress Th2 type autoimmunity. *PLOS ONE*. 2013;8(9):Article e74639.
82. David K, Friedlander G, Pellegrino B, Radomir L, Lewinsky H, Leng L, Bucala R, Becker-Herman S, Shachar I. CD74 as a regulator of transcription in normal B cells. *Cell Rep*. 2022;41(5):Article 111572.
83. Gil-Yarom N, Radomir L, Sever L, Kramer MP, Lewinsky H, Bornstein C, Blecher-Gonen R, Barnett-Itzhaki Z, Mirkin V, Friedlander G, et al. CD74 is a novel transcription regulator. *Proc Natl Acad Sci USA*. 2017;114(3):562–567.
84. Thompson CD, Matta B, Barnes BJ. Therapeutic targeting of IRFs: Pathway-dependence or structure-based? *Front Immunol*. 2018;9:2622.
85. Yam-Puc JC, Zhang L, Zhang Y, Toellner K-M. Role of B-cell receptors for B-cell development and antigen-induced differentiation. *F1000Res*. 2018;7:Article 429.
86. Herzog S, Reth M, Jumaa H. Regulation of B-cell proliferation and differentiation by pre-B-cell receptor signalling. *Nat Rev Immunol*. 2009;9(3):195–205.
87. Garrido-Trigo A, Salas A. Molecular structure and function of Janus kinases: Implications for the development of inhibitors. *J Crohn's Colitis*. 2020;14(Supplement_2):S713–S724.
88. Adams JM, Cory S. The Bcl-2 protein family: Arbiters of cell survival. *Science*. 1998;281(5381):1322–1326.
89. Vikström IB, Slomp A, Carrington EM, Moesbergen LM, Chang C, Kelly GL, Glaser SP, Jansen J, Leusen JH, Strasser A, et al. MCL-1 is required throughout B-cell development and its loss sensitizes specific B-cell subsets to inhibition of BCL-2 or BCL-XL. *Cell Death Dis*. 2016;7(8):e2345.
90. Lopez-Hoyos M, Carrio R, Merino J, Merino R. Regulation of B cell apoptosis by Bcl-2 and Bcl-XL and its role in the development of autoimmune diseases. *Int J Mol Med*. 1998;1(2):475–558.
91. Huber R, Pietsch D, Panterodt T, Brand K. Regulation of C/EBPβ and resulting functions in cells of the monocytic lineage. *Cell Signal*. 2012;24(6):1287–1296.
92. Zhang Z, Bossila EA, Li L, Hu S, Zhao Y. Central gene transcriptional regulatory networks shaping monocyte development in bone marrow. *Front Immunol*. 2022;13:1011279.
93. Kim S, Chen J, Ou F, Liu T-T, Jo S, Gillanders WE, Murphy TL, Murphy KM. Transcription factor C/EBPα is required for the development of Ly6Chi monocytes but not Ly6Clo monocytes. *Proc Natl Acad Sci USA*. 2024;121(15):Article e2315659121.
94. Hirai H, Zhang P, Dayaram T, Hetherington CJ, Mizuno S-i, Imanishi J, Akashi K, Tenen DG. C/EBPβ is required for 'emergency' granulopoiesis. *Nat Immunol*. 2006;7(7):732–739.
95. Zhu S, Lalani AI, Jin J, Sant'Angelo D, Covey LR, Liu K, Young HA, Ostrand-Rosenberg S, Xie P. The adaptor protein TRAF3 is an immune checkpoint that inhibits myeloid-derived suppressor cell expansion. *Front Immunol*. 2023;14:1167924.
96. Lalani AI, Moore CR, Luo C, Kreider BZ, Liu Y, Morse HC, Xie P. Myeloid cell TRAF3 regulates immune responses and inhibits inflammation and tumor development in mice. *J Immunol*. 2015;194(1):334–348.
97. Dzierzak E, Philipsen S. Erythropoiesis: Development and differentiation. *Cold Spring Harb Perspect Med*. 2013;3(4):Article a011601.
98. Schippel N, Sharma S. Dynamics of human hematopoietic stem and progenitor cell differentiation to the erythroid lineage. *Exp Hematol*. 2023;123:1–17.
99. Haas N, Riedt T, Labbaf Z, Baßler K, Gergis D, Fröhlich H, Gütgemann I, Janzen V, Schorle H. Kit transduced signals counteract erythroid maturation by MAPK-dependent modulation of erythropoietin signaling and apoptosis induction in mouse fetal liver. *Cell Death Differ*. 2015;22(5):790–800.
100. Akl MR, Nagpal P, Ayoub NM, Tai B, Prabhu SA, Capac CM, Gliksman M, Goy A, Suh KS. Molecular and clinical significance of fibroblast growth factor 2 (FGF2/bFGF) in malignancies of solid and hematological cancers for personalized therapies. *Oncotarget*. 2016;7(28):44735.
101. Bruveris FF, Ng ES, Stanley EG, Elefanty AG. VEGF, FGF2, and BMP4 regulate transitions of mesoderm to endothelium and blood cells in a human model of yolk sac hematopoiesis. *Exp Hematol*. 2021;103:30–39.e2.
102. Meng X, Baylink DJ, Sheng M, Wang H, Gridley DS, Lau K-HW, Zhang X-B. Erythroid promoter confines FGF2 expression to the marrow after hematopoietic stem cell gene therapy and leads to enhanced endosteal bone formation. *PLOS ONE*. 2012;7(5):Article e37569.
103. Duchene J, Novitzky-Basso I, Thiriot A, Casanova-Acebes M, Bianchini M, Etheridge SL, Hub E, Nitz K, Artinger K, Eller K, et al. Atypical chemokine receptor 1 on nucleated erythroid cells regulates hematopoiesis. *Nat Immunol*. 2017;18(7):753–761.
104. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, et al. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638–D646.
105. Biase FH, Cao X, Zhong S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell RNA sequencing. *Genome Res*. 2014;24(11):1787–1796.
106. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, Kim SK, Quake SR. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell*. 2017;171(2):321–330.e14.
107. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Gephart MGH, Barres BA, Quake SR. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA*. 2015;112(23):7285–7290.
108. Fink EE, Sona S, Tran U, Desprez P-E, Bradley M, Qiu H, Eltemamy M, Wee A, Wolkov M, Nicolas M, et al. Single-cell and spatial mapping identify cell types and signaling networks in the human ureter. *Dev Cell*. 2022;57(15):1899–1916.e6.
109. Goolam M, Scialdone A, Graham SJ, Macaulay IC, Jedrusik A, Hupalowska A, Voet T, Marioni JC, Zernicka-Goetz M. Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. *Cell*. 2016;165(1):61–74.
110. Hubert L, Arabie P. Comparing partitions. *J Classif*. 1985;2:193–218.
111. Vinh NX, Epps J, Bailey J. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *J Mach Learn Res*. 2010;11:2837–2854.
112. Strehl A, Ghosh J. Cluster ensembles—A knowledge reuse framework for combining multiple partitions. *J Mach Learn Res*. 2002;3:583–617.
113. Fowlkes EB, Mallows CL. A method for comparing two hierarchical clusterings. *J Am Stat Assoc*. 1983;78(383):553–569.