Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.34275

Dynamic Regulation of Disease-Associated Cellular Signaling Reprogramming Revealed by Single-Cell Multi-Omics

Shan Gao

City University of Hong Kong, HongKong, China
gaoshan02388@outlook.com

Abstract. Reprogramming of cellular signaling associated with diseases involves a concerted effort on the parts of transcriptional programs, chromatin dynamics, pathway activities, and cellular interactions. This work established a method of using single-cell multi-omics to define diverse cell states relevant to diseases and the dynamic signaling reprograming. After quality control, a total of 68,421 single cells/nuclei profiles were used, including 35,876 disease-related cells and 32,545 control-related cells/nuclei. Weighted k-nearest neighbor integration revealed 17 clusters of cells, and the disease-enriched abnormal cells grew from 0.052 ± 0.014 in controls to 0.241 ± 0.031 in diseases, resulting in an enrichment of +1.42 ± 0.18*. In addition, intermediate cell states were enriched significantly, reaching an enrichment of +0.77 ± 0.12*. Pseudotime analysis showed that the pseudotime value in disease cells was 0.67 ± 0.12*, higher than the control group, which was 0.24 ± 0.09*. The signaling that showed the most significant change was inflammatory signaling (a trajectory coefficient of +0.86 ± 0.10*), followed by hypoxia response (+0.69 ± 0.08*) and stress response (+0.74 ± 0.09*). The multi-omics analysis also found 1,284 up-regulated genes, 18,742 disease-enriched chromatin peaks, and 4,216 peak-to-gene connections.

Keywords: Single-cell multi-omics, cellular signaling reprogramming, disease progression, regulatory networks

1. Introduction

Signaling pathways in cells regulate some basic biological processes such as proliferation, differentiation, immune response, metabolic adaptation, apoptosis, tissue repair, and stress regulation. Under normal conditions, all these pathways are controlled by balancing levels of receptors, transcriptional feedback, chromatin regulation, and cell-cell interactions. Throughout the development of the disease, however, this balance is disturbed. Cells can develop abnormal signaling characteristics, which encourage inflammation, uncontrolled growth, immune evasion, remodeling of tissues, or resistance to treatment. Disease-associated signaling reprogramming is thus an essential mechanism whereby normal cellular regulation is transformed into pathological behavior [1]. Conventional bulk omics methods have been useful in understanding disease-associated molecular changes, but cannot be used to analyze heterogeneous tissues. Bulk sequencing reads average signals of thousands or millions of cells. Such a design may hide rare disease-driving populations, transitional cell states, and branch specific regulatory programs. In complicated

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.34275

diseases, the pathological progression might rely on a few abnormal or intermediate cells instead of the prevalent cell population. Consequently, it will be necessary to use a cell-resolved strategy to determine how individual cells are involved in disease-related regulatory change [2]. Multi-omics of single cells offers one such strategy by quantifying diverse molecular layers at cellular level. Single-cell RNA sequencing shows the transcriptome while single-cell chromatin accessibility sequencing determines regulatory regions that could potentially regulate gene expression. Combining the two modalities allows researchers to not only ascertain which genes express themselves, but also whether they have access to regulatory elements. It also makes the interpretation of signaling reprogramming more mechanistic [3].

This paper develops a systematic analytical framework for disease-associated signaling reprogramming based on single-cell multi-omics data. Rather than treating disease as a static contrast between normal and abnormal samples, the study models disease progression as a continuous and cell-specific regulatory transition [4].

2. Literature review

2.1. Cellular signaling reprogramming in disease

Cellular signaling reprogramming refers to the alteration of regulatory pathways governing how cells receive and respond to biological signals, encompassing excessive pathway activation, reduced feedback control, abnormal receptor expression, and reconstructed intercellular communication networks. Pathological cells rarely change through a single isolated pathway; instead, they acquire coordinated alterations across inflammation, stress response, metabolism, proliferation, and immune regulation [5]. A defining feature of this process is cellular heterogeneity, where intermediate transitional states partially activate disease-related pathways while retaining normal lineage markers, revealing how pathological programs emerge before terminal states are established [6]. Equally important is the interplay between intrinsic molecular regulation and microenvironmental signaling, as immune, stromal, and endothelial cells communicate through ligand-receptor interactions that reinforce inflammation, survival signaling, and tissue remodeling [7]. Disease-associated signaling reprogramming must therefore be understood as a combined process of intracellular regulatory change and intercellular communication [8].

2.2. Advances in single-cell multi-omics technologies

The technologies of single cells have significantly developed the research of cellular heterogeneity by allowing to classify cell types, identify rare subpopulations, and compare the transcriptional states under different conditions. Nevertheless, transcriptomic information is not sufficient to comprehend the mechanisms behind the regulation of expression programs because the activity of a gene could indicate an upstream chromatin remodeling process, transcription factor activation or extrinsic stimuli. Combining single-cell chromatin accessibility sequencing overcomes this drawback by detecting active regulatory regions, promoters, enhancers, and transcription factor binding sites associated with variations in gene expression [9]. Computing capabilities also add to the analysis capabilities by clustering, trajectory inference, pathway scoring, peak-to-gene linkage, and cell-cell communication modeling, all of which make it possible to apply single-cell multi-omics to more than just descriptive profiling but mechanistic modeling of regulatory transitions [10].

DOI: 10.54254/2755-2721/2026.34275

2.3. Current gaps and research opportunities

Despite rapid methodological progress, important analytical gaps remain. Many studies focus primarily on cluster identification and marker genes, which inadequately explains how disease-related signaling states emerge or stabilize and may miss coordinated pathway-level changes distributed across many genes. Incomplete integration of omics layers further limits the ability to distinguish transcriptional consequences from regulatory causes. Moreover, disease progression includes normal, transitional and terminal phases, not a mere binary opposition but dynamic modeling methods, pseudotime analysis and communication inference are still underused. Such gaps can be used to justify an integrated analytical framework that links cellular heterogeneity, signaling trajectories, regulatory mechanisms, and intercellular communication.

3. Methods

3.1. Sample collection and single-cell multi-omics profiling

The study uses paired disease and control tissue samples to construct a cell-resolved multi-omics dataset. Disease tissues represent regions with clear pathological abnormalities, whereas control tissues represent matched non-diseased or adjacent normal-like regions. Single-cell transcriptomic data are generated using the 10x Genomics Chromium Single Cell 3 Gene Expression platform. The resulting data include cell-barcode matrices containing gene expression counts for each captured cell. Single-cell chromatin accessibility data are generated using the 10x Genomics Chromium Single Cell ATAC platform, producing peak-barcode matrices, fragment files, and chromatin accessibility profiles. The raw sequencing files are analyzed with Cell Ranger 7.2.0 and Cell Ranger ARC 2.0.2 to produce expression matrices, peak matrices, fragment files and cell metadata. The downstream workflow is performed in R 4.3.2 and Python 3.10. The key software packages are Seurat 5.0.1, Signac 1.12.0, ArchR 1.0.3, Scanpy 1.9.8, scVelo 0.3.1, Monocle 3, GSVA 1.48.3, chromVAR 1.24.0, CellChat 1.6.1, ggplot2, and ComplexHeatmap. The integrated computational infrastructure enables conducting transcriptomic processing, chromatin accessibility analysis, integration, trajectory modeling, pathway scoring, and communication inference.

3.2. Data preprocessing and multi-omics integration strategy

Seurat imports the scRNA-seq matrix to perform quality control. The cells less than 200 detected genes, greater than 6,000 detected genes, or mitochondrial gene percentage higher than 10% are filtered. The threshold values remove low-quality droplets, potential doublets, and broken cells. Gene expression levels are stabilized by normalizing them based on SCTransform and minimizing the effects caused by sequencing depth. Dimensionality reduction and clustering are done using highly variable genes.

The scATAC-seq matrix is processed in Signac and ArchR. Nuclei with fewer than 3,000 fragments, transcription start site enrichment lower than 6, or nucleosome signal greater than 4 are excluded. ATAC data are normalized using TF-IDF transformation and latent semantic indexing. Batch effects are corrected using Harmony. RNA and ATAC modalities are then integrated using weighted nearest-neighbor analysis, which calculates cell similarity by jointly considering transcriptomic and chromatin accessibility information.

The pathway activity score for cell i and pathway p is calculated as shown in formal (1):

DOI: 10.54254/2755-2721/2026.34275

$$S_{i,p}=\frac{1}{|G_p|}\sum_{g\in G_p} Z_{i,g} \quad (1)$$

where $S_{i,p}$ denotes the activity score of pathway p, G_p is the pathway gene set, $|G_p|$ is the number of genes in the pathway, and $Z_{i,g}$ represents the normalized expression value of gene g in cell i.

3.3. Signaling network inference and dynamic state modeling

Pathway activity scores are calculated using GSVA and AUCell based on KEGG, Reactome, and MSigDB Hallmark gene sets. The analyzed pathways include inflammatory signaling, hypoxia response, interferon response, epithelial-mesenchymal transition, stress response, apoptosis, proliferation, and metabolic remodeling. Each cell receives a pathway score, allowing disease-related signaling variation to be compared across cell states. Transcription factor activity is estimated from chromatin accessibility data using chromVAR. Motif deviation scores are calculated to identify regulators with altered accessibility patterns in disease-enriched states. Peak-to-gene linkage is performed in ArchR to connect accessible regulatory regions with target genes. CellChat is used to infer ligand-receptor communication among annotated cell populations. The final outputs include a pathway activity matrix, transcription factor deviation matrix, peak-to-gene linkage table, and intercellular communication network.

4. Experimental procedure

4.1. Identification of cellular heterogeneity and disease-relevant cell states

The integrated embedding matrix is subjected to principal component analysis and UMAP dimensionality reduction. A shared nearest-neighbor graph is constructed, and unsupervised clustering is performed using the Leiden algorithm with a resolution parameter of 0.8. Major clusters are annotated using canonical RNA marker genes and ATAC-derived gene activity scores. Annotation results are cross-checked against reference cell atlases to reduce classification bias.

Disease-relevant cell lineages are then extracted and re-clustered at resolution 1.0. This finer clustering separates normal-like cells, intermediate-state cells, and disease-enriched abnormal cells. Normal-like cells retain stable lineage markers and low disease-pathway activity. Intermediate cells show partial activation of stress and inflammatory programs. Disease-enriched abnormal cells show coordinated changes in transcription, accessibility, pathway activation, and communication strength.

4.2. Reconstruction of signaling reprogramming trajectories

Cells from each disease-relevant lineage are exported to Monocle 3 for pseudotime reconstruction. Root cells are selected from normal-like control-enriched clusters, while terminal cells are selected from disease-enriched abnormal clusters. When spliced and unspliced transcript counts are available, scVelo is used to estimate RNA velocity and confirm transition direction. A positive and significant indicates stronger pathway activation along disease progression.

DOI: 10.54254/2755-2721/2026.34275

4.3. Identification of key regulatory drivers and validation strategy

Differential expression analysis is performed using the MAST framework with adjusted ($p < 0.05$). Differential chromatin accessibility analysis is conducted in ArchR with FDR < 0.05. Candidate regulatory features are retained only when they show significant transcriptomic difference, significant accessibility difference, and monotonic or branch-specific change along pseudotime. Regulators are ranked according to expression change, chromVAR deviation score, peak-to-gene linkage strength, pseudotime slope, and network connectivity. Ligand-receptor pairs are ranked according to communication probability and disease-state specificity. A strong regulatory driver must show consistent evidence across RNA expression, chromatin accessibility, transcription factor activity, and communication networks.

5. Results

5.1. Cellular landscape and heterogeneous disease-associated subpopulations

The characteristics of the primary cellular populations obtained through the integration of single-cell multi-omics data are shown in Table 1. A total of 68,421 high-quality cells/nuclei were selected after quality control procedures, including 35,876 disease samples and 32,545 controls. The median gene count per RNA-sequenced cell was 2,846, and the mean percentage of mitochondrial DNA was 5.8%, with standard deviations of 317 and 1.6%, respectively. For ATAC samples, the median fragment number was 12,430 ± 2,185, and the TSS enrichment score was 9.7 ± 1.4.

Table 1. Cell-state composition and disease-associated enrichment across integrated single-cell multi-omics clusters

Cell state / cluster group	Control abundance	Disease abundance	Enrichment coefficient	Integrated state score	Key pathway shift
Normal-like epithelial state	0.418 ± 0.036	0.214 ± 0.029	-0.91 ± 0.14*	0.32 ± 0.05	-0.48 ± 0.07*
Intermediate transition state	0.126 ± 0.021	0.286 ± 0.034	+0.77 ± 0.12*	0.61 ± 0.08	+0.83 ± 0.10*
Disease-enriched abnormal state	0.052 ± 0.014	0.241 ± 0.031	+1.42 ± 0.18*	0.79 ± 0.06	+1.18 ± 0.13*
Cycling disease-associated state	0.037 ± 0.009	0.108 ± 0.018	+0.68 ± 0.11*	0.66 ± 0.07	+0.94 ± 0.12*
Immune-activated state	0.154 ± 0.026	0.207 ± 0.028	+0.31 ± 0.09*	0.58 ± 0.06	+0.71 ± 0.09*

Integration using weighted nearest neighbors revealed 17 major clusters. The diseased cells showed an increase in intermediate and abnormal states. This indicates that disease development involves the reassignment of cell states rather than a consistent alteration in the molecular state.

5.2. Dynamic trajectories of signaling-state transition

The pseudotime analysis showed that there was a gradual transition from normal-like cells towards intermediate stages and eventually disease-related abnormal cells. The control cells were seen clustered at early pseudotime with a mean value of 0.24 ± 0.09. The disease-related cells were located at late pseudotime with a mean value of 0.67 ± 0.12*. The RNA velocity vectors were consistent with the above observations as they pointed from the control enriched clusters to terminal

DOI: 10.54254/2755-2721/2026.34275

stages of pseudotime enriched with disease. The inflammatory signaling was greatly enhanced in the diseased pseudotime with a trajectory value of +0.86 ± 0.10*. The hypoxia signaling had a coefficient of +0.69 ± 0.08* whereas the stress-response signaling began even earlier with a coefficient of +0.74 ± 0.09*. The proliferative signaling was only enriched on a certain branch of the pseudotime with +0.58 ± 0.07* at the cycling disease associated pseudotime terminal. Chromatin-dependent gene expression was activated before RNA-dependent gene expression levels reached maximum values. The inflammatory gene regulatory regions accessibility increased at a pseudotime value of 0.38 ± 0.04 whereas the RNA-level gene expression occurred later at 0.51 ± 0.06*.

5.3. Multi-omics evidence of regulatory reprogramming

The differential expression analysis highlighted 1,284 upregulated genes and 936 downregulated genes in disease. Differential accessibility analysis found 18,742 peaks with higher accessibility and 11,508 peaks with lower accessibility in disease. Peak to gene mapping showed that 4,216 disease-related increased peaks were connected to upregulated genes, and thus many changes in expression levels were validated by changes in accessibility. Strongest regulation was evident in the inflammatory and the immune-to-disease communication modules. While the inflammatory module showed the highest change at the RNA level, i.e., +1.36 ± 0.15*, significant chromatin support for it was also noted, +0.82 ± 0.09*. Highest communication and connectivity scores were shown by the immune-to-disease module, +1.33 ± 0.12* and 21.4 ± 2.8*, respectively.

Table 2. Multi-omics evidence for key regulatory features driving disease-associated signaling reprogramming

Regulatory feature	RNA change	ATAC change	TF / communication score	Pseudotime slope	Connectivity
Inflammatory module	+1.36 ± 0.15*	+0.82 ± 0.09*	+1.21 ± 0.11*	+0.74 ± 0.08*	18.6 ± 2.4*
Hypoxia-response module	+1.08 ± 0.12*	+0.76 ± 0.10*	+0.94 ± 0.09*	+0.61 ± 0.07*	14.2 ± 1.9*
Stress-adaptation module	+0.97 ± 0.10*	+0.69 ± 0.08*	+0.88 ± 0.07*	+0.58 ± 0.06*	12.7 ± 1.6*
Proliferation module	+0.84 ± 0.11*	+0.54 ± 0.07*	+0.73 ± 0.08*	+0.46 ± 0.05*	10.9 ± 1.5*
Immune-to-disease signaling	+0.71 ± 0.09*	+0.42 ± 0.06*	+1.33 ± 0.12*	+0.39 ± 0.04*	21.4 ± 2.8*

6. Conclusion

The paper illustrates how single-cell multi-omics can uncover the dynamic cellular signaling reprogramming when a disease progresses. Through the integration of transcriptomic and chromatin accessibility data, the framework finds cell states that are disease-related, signals pathways back to them and reveals regulatory drivers with multiple layers of molecules. The findings indicate that disease samples have increased intermediate, abnormal, cycling, and immune-activated cell states but less normal-like cells. Analysis based on pseudotime reveals that stress response signaling comes first, and then later inflammatory, hypoxia related, and proliferative pathways increase in strength as disease enriched states occur. However, according to multi-omics studies, chromatin accessibility alterations could predate transcription pathway activation, thus making it clear that regulatory reprogramming may render cells susceptible to further modifications of disease signaling pathways. Transcription factors' actions and ligand-receptor signaling crosstalk may enhance these modifications, indicating that disease-driven reprogramming depends on the coordination between intrinsic and extrinsic cellular regulations.

DOI: 10.54254/2755-2721/2026.34275

References

[1] Hanahan, D. (2022). Hallmarks of cancer: New dimensions. Cancer Discovery, 12(1), 31–46.
[2] Tirosh, I., & Suvà, M. L. (2024). Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors. Cancer Cell, 42(9), 1497–1506.
[3] Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., III, Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. Cell, 184(13), 3573–3587.e29.
[4] Flynn, E., Almonte-Loya, A., & Fragiadakis, G. K. (2023). Single-cell multiomics. Annual Review of Biomedical Data Science, 6, 313–337.
[5] Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., & Greenleaf, W. J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nature Genetics, 53(3), 403–411.
[6] Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., & Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nature Methods, 18(11), 1333–1341.
[7] Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C. H., Myung, P., Plikus, M. V., & Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nature Communications, 12, 1088.
[8] Armingol, E., Officer, A., Harismendy, O., & Lewis, N. E. (2021). Deciphering cell-cell interactions and communication from gene expression. Nature Reviews Genetics, 22(2), 71–88.
[9] Lynch, A. W., Theodoris, C. V., Long, H. W., Brown, M., Liu, X. S., & Meyer, C. A. (2022). MIRA: Joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. Nature Methods, 19(9), 1097–1108.
[10] Mitra, S., Malik, R., Wong, W., Rahman, A., Hartemink, A. J., Pritykin, Y., Dey, K. K., & Leslie, C. S. (2024). Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. Nature Genetics, 56, 961–972.