Proceedings of ICBioMed 2026 Symposium: Biomedical Engineering and Smart Medicine
DOI: 10.54254/2753-8818/2026.HZ38683

Integrative Single-Cell and Spatial Multi-Omics Analysis of Cancer Stem Cell Niches and Plasticity: Methodological Framework and Research Advances

Yixin Ye

College of Life Science,Hubei University, Wuhan, China
easonye_2023@qq.com

Abstract. Cancer stem cells (CSCs) are closely associated with intratumoral heterogeneity, therapy resistance, relapse, and metastasis. Increasing evidence suggests that CSCs should not be regarded solely as a fixed population defined by a small set of markers, but rather as dynamic cellular states shaped by the tumor microenvironment. Single-cell RNA sequencing (scRNA-seq) resolves transcriptional heterogeneity and cellular composition, whereas spatial transcriptomics (ST) preserves in situ tissue organization. Their integration therefore provides a powerful strategy for dissecting CSC niches and plasticity. This review summarizes the major computational routes for integrative scRNA-seq and ST analysis, including malignant-cell and CSC-state identification, cell-cell communication inference, pseudotime and RNA-velocity analysis, spatial mapping, and cellular-neighborhood modeling. Representative tools and their assumptions, strengths, and limitations are compared. We further propose a standardized workflow from single-cell characterization to spatial validation and dynamic modeling, and discuss representative applications in hepatocellular carcinoma, small-cell lung cancer, and endometrial carcinoma. Key challenges include unpaired samples, spatial-resolution limitations, prior-database bias, the non-causal nature of trajectory inference, and the difficulty of localizing rare CSC states. Future directions include single-cell/subcellular spatial technologies, multimodal integration, graph-based deep learning, and causal modeling.

Keywords: cancer stem cells, single-cell RNA sequencing, spatial transcriptomics, multi-omics integration, cell-cell communication, spatial niche, plasticity

1. Introduction

Cancer stem cells (CSCs) are generally described as a population of cells that possess self-renewal and differentiation capacities and are capable of sustaining tumor growth. This concept has provided an important theoretical framework for explaining the hierarchical organization within tumors, drug resistance, and recurrence. Nevertheless, CSCs exhibit substantial heterogeneity across different tumor types and treatment stages, and their phenotypes cannot always be stably captured by a small number of canonical markers [1]. In hepatocellular carcinoma, classic CSC markers such as CD13, CD24, EPCAM, CD44, and CD133 are shared with normal stem/progenitor cells; relying solely on

© 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: 10.54254/2753-8818/2026.HZ38683

surface markers may not only cause biological misinterpretation but also increase the risk of off-target therapeutic effects [2]. Consequently, current CSC research is gradually shifting from a "fixed subpopulation" perspective toward a more integrated view of "dynamic cell states" and "niche dependence."

The CSC niche is a local microenvironment composed of malignant cells, immune cells, fibroblasts, endothelial cells, the extracellular matrix, and factors such as hypoxia and nutrient status. This niche can maintain stemness and promote state transitions through ligand–receptor signaling, metabolic exchange, matrix remodeling, and immunosuppression [2, 3]. Traditional bulk transcriptome sequencing reflects averaged signals from mixed cell populations and cannot readily address questions such as "which cells are in a highly stem-like state," "which cells are adjacent to these cells," "what signals may maintain their state," or "whether these states exhibit continuous transitions." scRNA-seq and ST provide cellular-state resolution and spatial in situ information, respectively; their integration allows CSC research to proceed along four dimensions: cell identity, spatial location, cell–cell communication, and dynamic evolution.

Reviews have addressed single-cell annotation, data integration, communication inference, and spatial multi-omics [4-6]. Yet CSC niche analysis still lacks coherent tool-chaining rationale, relies on single metrics for CSC definition, lacks single-cell/spatial cross-validation, and overinterprets dynamic inference. This review focuses on "method selection and levels of evidence," outlining a computational framework for CSC niches and plasticity and proposing a reproducible integrated workflow.

2. Methodological foundations and core questions in CSC niche research

2.1. From "marker-based definition" to "multi-evidence CSC state identification"

CSC identification is the starting point for subsequent communication, spatial, and trajectory analyses, and it is also the step most prone to introducing bias. Relying solely on single markers such as CD44, EPCAM, or CD133 is insufficient to capture the stemness heterogeneity across tumor types, patients, and treatment states. A more robust strategy generally includes three layers of evidence: first, identifying malignant cells through copy-number variation inference, known tumor cell markers, and patient-derived information; second, within the malignant cells, assessing stemness-related gene modules or differentiation potential; and third, combining pseudotime, spatial location, clinical associations, or functional experiments to increase confidence in CSC state assignment.

CytoTRACE infers differentiation potential from transcriptional diversity and expressed gene counts, with higher scores indicating a less differentiated state [7]. In hepatocellular carcinoma, Fan et al. defined CSCs and non-CSCs using the top and bottom 25% of CytoTRACE scores among malignant cells, respectively, and constructed a 107-gene CSC signature via WGCNA. Notably, this "top 25%" threshold was study-specific, not a universal criterion; CytoTRACE reflects transcriptome-level differentiation potential and cannot replace functional validation such as clonogenicity or tumor-initiating capacity. For new cancer types or cohorts, continuous stemness scores with multi-method consistency checks are preferable to fixed quantiles.

2.2. Complementarity of single-cell and spatial technologies

scRNA-seq offers relatively high transcriptome coverage, making it suitable for cell clustering, malignant-cell identification, gene-module analysis, and state-transition modeling, but tissue

DOI: 10.54254/2753-8818/2026.HZ38683

dissociation inevitably destroys spatial architecture. ST preserves spatial coordinates and can be used to identify local enrichment, neighborhood composition, and spatial interactions; however, different technologies vary markedly in gene coverage and spatial resolution. Conventional spot-based ST often contains multiple cells per spot and requires deconvolution using single-cell references. Newer imaging-based or high-density spatial technologies can approach single-cell or even subcellular resolution but usually involve trade-offs in terms of targeted gene panels, cost, and computational scale.

Thus, CSC niche research should not be simplistically viewed as "scRNA-seq discovery followed by ST plotting." Instead, a bidirectional validation relationship should be established: single-cell data are used to define candidate CSC states and their molecular features, while spatial data are used to test whether these states form stable local niches; conversely, specific neighborhoods, boundaries, or co-localization patterns found in space can also guide the re-screening of key cell subsets and signaling axes in single-cell data.

3. Four core computational routes

Four core computational routes together form a multi-layer analytical pipeline for CSC niche research. First, CSC subpopulation identification and annotation should distinguish malignant from non-malignant cells and represent CSC states as both discrete labels and continuous stemness scores. Second, cell–cell communication inference, using tools such as CellChat [8], CellPhoneDB [9], NicheNet [10], SpaTalk [11], and COMMOT [12], should prioritize biologically meaningful signals rather than enumerate significant ligand–receptor pairs. Third, state-transition analyses such as pseudotime (Monocle, Slingshot [13]) and RNA velocity (scVelo [14]) should treat stemness as a plastic, dynamic state rather than a static label. Fourth, spatial mapping, cellular-neighborhood, and niche analysis (e.g., cell2location [15], Squidpy [16], and CosMx-based studies [17]) should elevate the unit of analysis from single cells to recurrent multicellular structures. Together, these routes support a multi-layer chain of evidence—"co-expression → proximity → functional association"—for screening candidate CSC niche mechanisms.

Table 1. Functional comparison of commonly used computational tools in CSC niche research

Analytical step	Representative tools	Core rationale and advantages	Major limitations and recommendations
CSC state identification	CytoTRACE; marker/module scoring; inferCNV	Identify malignant cells first, then use differentiation potential and stemness modules to define candidate CSC states; suitable for constructing continuous stemness spectra.	Scores are affected by sequencing depth, proliferation status, and tumor type; a single threshold should not replace functional CSC definitions.
Cell-cell communication	CellChat; CellPhoneDB; NicheNet	Infer interactions from ligand-receptor co-expression; NicheNet further links to downstream target genes.	Rely on prior databases and mRNA expression; results mainly generate mechanistic hypotheses and should be combined with spatial and experimental validation.
Dynamic plasticity	Slingshot; Monocle; scVelo	Pseudotime delineates state continua; RNA velocity provides short-term directional information.	Static data cannot prove true lineages; results depend on model assumptions, starting points, and splicing kinetics.

DOI: 10.54254/2753-8818/2026.HZ38683

Table 1. (continued)

Spatial mapping	cell2location; anchoring; RCTD, etc.	Map cell types/states to spots or tissue coordinates using scRNA-seq references.	Cross-sample differences and missing references reduce mapping accuracy; rare CSCs especially require uncertainty assessment.
Spatial communication	SpaTalk；COMMOT	Incorporate spatial distance, graph structures, or optimal transport constraints into ligand-receptor inference.	Spatial proximity does not equal true molecular contact; resolution differences across platforms alter neighborhood definitions.
Spatial neighborhood	Squidpy; custom adjacency graphs/neighborhood clustering	Identify recurrent local cellular combinations, enrichment relationships, and spatial structures.	Neighborhood scale and k-value choices affect results; reproducibility should be verified across multiple samples.

4. A standardized integrative analysis workflow for CSC niches

Based on the above methods, CSC niche research can be organized into six mutually validating steps, rather than simply linking multiple software packages in a serial manner.

First, Single-cell preprocessing and malignant-cell resolution. After QC, batch correction, annotation, and patient stratification, malignant cells are identified via CNV inference and tumor markers. Patient/sample information should be retained to avoid erasing genuine inter-patient differences through over-integration.

Second, multi-evidence definition of candidate CSC states. Within the malignant cells, compute CytoTRACE, stemness-gene-set scores, and related transcriptional programs, and evaluate their correlations with cell-cycle, hypoxia, and stress programs. If discretization into CSC versus non-CSC is needed, data-driven thresholds or sensitivity analyses should be preferred over directly applying fixed cutoffs from other studies.

Third, screening of candidate niche communications. Use CellChat or CellPhoneDB to identify high-priority signals between CSCs and macrophages, fibroblasts, endothelial cells, and T cells; then use NicheNet or similar approaches to test whether these ligands can explain downstream transcriptional programs in CSCs. Candidate pathways should be evaluated simultaneously for expression intensity, cell-type specificity, cross-sample consistency, and known biology.

Fourth, spatial localization and cellular-neighborhood validation. Map CSC features or single-cell annotations to ST data, identify CSC-enriched regions, and assess neighboring cellular composition, hypoxia/vascular/immune status, and spatial boundaries. For key ligand–receptor pairs, further apply distance constraints using SpaTalk or COMMOT. Whenever possible, test whether the same niche recurs across multiple tissue sections or independent patients.

Fifth, dynamic plasticity modeling. Use Slingshot, Monocle, or scVelo to examine whether candidate CSC states exhibit continuous transitions with other malignant states, and link state directionality to microenvironmental exposure, pre-/post-treatment differences, or spatial location. The emphasis at this step is to propose "possible state-transition models" rather than declaring a cell lineage based solely on a single pseudotime trajectory.

Sixth, cross-modal and experimental validation. Prioritize candidate mechanisms that simultaneously satisfy "significant single-cell differential expression, spatial co-localization, communication network support, and cross-patient reproducibility." Subsequently, perform immunofluorescence/in situ hybridization, flow sorting, organoid or co-culture systems, and genetic perturbations to convert computational associations into testable mechanistic hypotheses.

DOI: 10.54254/2753-8818/2026.HZ38683

5. Representative application examples and transferable lessons

Representative applications illustrate both the value and the boundaries of method transfer. In hepatocellular carcinoma, Fan et al. linked CSC definition to spatial co-localization with SPP1+ macrophages in hypoxic regions, prioritizing SPP1–CD44 and SPP1–ITGA/ITGB axes and establishing a complete evidence chain from CSC identification to communication mechanisms. In small-cell lung cancer, Zhang et al. used CosMx spatial transcriptomics to define metastasis-enriched malignant states, immune-exclusion features, and recurrent neighborhood structures, providing an analytical paradigm—single-cell states plus neighborhood structures plus clinical outcomes—that is directly transferable to CSC research. In endometrial carcinoma and other diseases, studies combining scRNA-seq and ST have refined ligand–receptor inference with spatial constraints, while the prostate cancer study demonstrated a combinatorial strategy of single-cell macrophage markers, WGCNA, and machine-learning prognostic models [18]. Similarly, a 2026 idiopathic pulmonary fibrosis study integrated cross-omics evidence with multi-layer validation and Mendelian randomization [19]. However, when CSCs were not directly defined, such findings should be presented as transferable methods rather than direct evidence for CSC mechanisms—a distinction that improves argumentative rigor in an EI-conference review.

6. Strengths, limitations, and quality-control priorities

6.1. Strengths of the integrative framework

Compared with single-omics approaches, the integration of scRNA-seq and ST places cell states, tissue organization, and cell–cell interactions within a unified analytical framework. scRNA-seq enhances the resolution of rare states; ST provides tissue boundaries, co-localization, and neighborhood structures; communication analysis generates mechanistic hypotheses; and pseudotime and RNA velocity complement state-dynamic information. For CSCs—which are both rare and plastic—multi-layer evidence stacking is more robust than any single marker or algorithm.

6.2. Major limitations and common pitfalls

First, scRNA-seq and ST often derive from different patients or from adjacent rather than identical tissue regions; cross-sample mapping may confound genuine biological differences with technical artifacts, or vice versa. Second, conventional spot-based ST involves mixed cells per spot, making spatial localization of rare CSCs heavily dependent on reference data and deconvolution models. Third, cell-communication tools rely on prior databases, and mRNA co-expression does not guarantee that protein production, secretion, binding, and downstream responses actually occur. Fourth, pseudotime and RNA velocity primarily depict statistical state relationships and are not equivalent to lineage tracing. Fifth, CSC scores can be confounded by cell cycle, stress, and sequencing depth, requiring cautious interpretation, especially in highly proliferative tumors.

Therefore, quality control should permeate the entire workflow: report cell/spot counts, filtering criteria, and batch-correction strategies; perform key statistical tests at the patient level rather than the cell level; conduct sensitivity analyses for CSC thresholds; replicate spatial neighborhoods in independent samples; perform at least protein-level or perturbation validation for key communication axes; and clearly label trajectory directions as "inferred" rather than "proven." These details often enhance reproducibility and reviewer persuasiveness more than adding additional algorithms.

DOI: 10.54254/2753-8818/2026.HZ38683

7. Future directions

Future CSC niche research will shift from single-cell transcriptomics plus low-resolution spatial localization toward single-cell/subcellular spatial multimodal analysis. Imaging-based technologies such as Xenium, CosMx, and MERFISH, together with high-density spatial transcriptomics, will enable more precise localization of rare CSCs and their adjacent cells, reducing deconvolution-induced uncertainty. Integrating scATAC-seq, spatial proteomics, metabolomics, and epigenetic information will further help distinguish transcriptional states from truly functional stable cell states. Computationally, graph neural networks are suited for representing cell–cell adjacency, while Transformer-based architectures can support cross-modal feature integration; however, deep-learning models should be evaluated not only by prediction accuracy but also by interpretability, cross-platform generalization, and independent validation. More critically, CSC plasticity research must move from cross-sectional associations to longitudinal and perturbation designs, combining paired pre-/post-treatment sampling, lineage tracing, organoid co-culture, and spatial perturbation experiments with computational models to test whether a given niche signal truly drives CSC state transitions.

The ultimate goal for clinical translation should likewise shift from single CSC markers to quantifiable niche-based metrics, such as CSC–macrophage co-localization scores, proportions of specific spatial neighborhoods, or spatially validated communication-strength composites. Such metrics hold promise for patient stratification, resistance-risk assessment, and prioritization of combination-therapy targets, but their clinical implementation will require cross-center standardization and prospective validation.

8. Conclusion

CSCs are better understood as dynamic functional states embedded in the tumor microenvironment rather than as isolated cell populations defined solely by fixed markers. The integration of scRNA-seq and ST enables researchers to sequentially address key questions such as "which malignant cells possess stem-like features," "where they are located," "which supporting cells and candidate signals surround them," and "how these states transition." Focusing on four core method categories—CSC identification, cell–cell communication, dynamic trajectories, and spatial niches—this review proposes a six-step integrative framework that extends from single-cell dissection to spatial and experimental validation. Future efforts should not simply pile up analytical tools; rather, they should build chains of evidence that are consistent across modalities, reproducible across patients, and supported by causal validation, thereby providing a more reliable methodological foundation for targeting CSC niches and overcoming tumor resistance.

References

[1] Eun, K., Ham, S. W., & Kim, H. (2017). Cancer stem cell heterogeneity: Origin and new perspectives on CSC targeting. *BMB Reports, 50*, 117–125.

[2] Fan, G., Xie, T., Li, L., et al. (2024). Single-cell and spatial analyses revealed the co-location of cancer stem cells and SPP1+ macrophage in hypoxic region that determines the poor prognosis in hepatocellular carcinoma. *npj Precision Oncology, 8*, 75.

[3] Yu, X., et al. (2023). Integrating single-cell RNA-seq and spatial transcriptomics reveals MDK-NCL dependent immunosuppressive environment in endometrial carcinoma. *Frontiers in Immunology, 14*, 1145300.

[4] Cheng, C., Chen, W., Jin, H., & Chen, X. (2023). A review of single-cell RNA-seq annotation, integration, and cell-cell communication. *Cells, 12*, 1970.

DOI: 10.54254/2753-8818/2026.HZ38683

[5] Su, J., et al. (2024). Cell-cell communication: New insights and clinical implications. *Signal Transduction and Targeted Therapy, 9*, 196.
[6] Liu, X., et al. (2024). Spatial multi-omics: Deciphering technological landscape of integration of multi-omics and its applications. *Journal of Hematology & Oncology, 17*, 72.
[7] Gulati, G. S., Sikandar, S. S., Wesche, D. J., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science, 367*, 405–411.
[8] Jin, S., Guerrero-Juarez, C. F., Zhang, L., et al. (2021). Inference and analysis of cell-cell communication using CellChat. *Nature Communications, 12*, 1088.
[9] Efremova, M., Vento-Tormo, M., Teichmann, S. A., & Vento-Tormo, R. (2020). CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nature Protocols, 15*, 1484–1506.
[10] Browaeys, R., Saelens, W., & Saeys, Y. (2020). NicheNet: Modeling intercellular communication by linking ligands to target genes. *Nature Methods, 17*, 159–162.
[11] Shao, X., Li, C., Yang, H., et al. (2022). Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nature Communications, 13*, 4429.
[12] Cang, Z., Zhao, Y., Almet, A. A., et al. (2023). Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nature Methods, 20*, 218–228.
[13] Street, K., Risso, D., Fletcher, R. B., et al. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics, 19*, 477.
[14] Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology, 38*, 1408–1414.
[15] Kleshchevnikov, V., Shmatko, A., Dann, E., et al. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology, 40*, 661–671.
[16] Palla, G., Spitzer, H., Klein, M., et al. (2022). Squidpy: A scalable framework for spatial omics analysis. *Nature Methods, 19*, 171–178.
[17] Zhang, Z., Wu, D., Chen, R., et al. (2026). Single-cell spatial transcriptomics reveals tumor microenvironment heterogeneity in primary and lymph node-metastatic small cell lung cancer. *Cell Reports Medicine, 7*(4), 102713.
[18] Ou, Y., et al. (2024). Comprehensive scRNA-seq analysis to identify new markers of M2 macrophages for predicting the prognosis of prostate cancer. *Annals of Medicine, 56*, 2398195.
[19] Hu, Z., Wang, J., Yu, J., et al. (2026). Multi-omics integration and machine learning reveal gut-immune signatures in idiopathic pulmonary fibrosis: Insights from bulk RNA-seq, single-cell profiles, spatial transcriptomics, and experimental validation. *Frontiers in Immunology, 17*, 1730289.