Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.36973

A Systematic Review of Clustering Methods for Spatial Multi-omics: A Unified Analysis of Model Architecture and Parameter Sensitivity

Linna Fan

College of Life Sciences, North China University of Science and Technology, Tangshan, China
Fannnnna@126.com

Abstract. While preserving the spatial coordinates of tissues, spatial multi-omics technologies jointly measure gene expression, protein abundance and chromatin accessibility, offering complementary molecular perspectives for deciphering tissue microenvironment heterogeneity. Nevertheless, the interwoven problems of high-dimensional sparsity, heterogeneous modality distribution and spatial autocorrelation require spatial domain identification to not only align cross-modal information, but also balance modality-specific signals and tissue spatial continuity. Following the logical thread of "data characteristics-core challenges-integration strategies-empirical boundaries", this paper first defines the task and data properties of spatial multi-omics clustering. Key issues are then summarized as cross-modal heterogeneity, information complementarity and spatial consistency. Representative approaches including graph-structure modelling, attention and contrastive learning, probabilistic factor decomposition, ensemble consensus and biological prior knowledge are compared comprehensively from the architectural, statistical and semantic levels. On this basis, this study takes Leiden clustering as a reproducible baseline, carries out 72 groups of parameter evaluations on *MouseBrain* and *HumanTonsil* datasets, and systematically compares the effects of PCA dimension, neighbour number k and resolution on clustering granularity, consistency with reference labels and spatial continuity. The results show that higher resolution is accompanied by an increase in cluster number and a decline in weighted Moran's I; the optimal configurations for ARI, NMI and spatial continuity do not overlap in the *MouseBrain* dataset. These findings reveal clear parameter-selection sensitivity of basic clustering workflows and provide empirical references for subsequent comparisons of complex integration models under unified data, pre-processing and evaluation standards.

Keywords: spatial multi-omics, spatial domain identification, multimodal integration, Leiden clustering, parameter sensitivity

1. Introduction

Single-cell sequencing has advanced gene expression analysis to cellular resolution, uncovering cell heterogeneity masked by population-averaged measurements. Single-cell multi-omics techniques

© 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: 10.54254/2755-2721/2026.36973

such as CITE-seq and scTrio-seq [1] further connect different molecular layers within the same cell. Nevertheless, tissue dissociation destroys the in-situ coordinates of cells and their neighbourhood relationships. While spatial transcriptomic technologies including 10x Visium and Stereo-seq [2] anchor expression profiles back to tissue architecture, they are mostly limited to a single molecular modality [3, 4]. DBiT-seq [5], Xenium, CosMx, Spatial-ATAC-RNA, SPOTS, as well as Stereo-seq-based spatial proteome-transcriptome co-detection platforms, jointly measure gene expression, protein abundance or epigenetic modifications while preserving spatial locations. These techniques provide a new technical foundation for dissecting tissue structure and cell-cell interactions through multi-layer molecular evidence [6, 7].

Multi-omics emphasizes the joint measurement and integration of multiple omics layers, whereas cross-omics pays more attention to directional relationships between different layers and their synergistic utilization. Spatial domain identification serves as a core downstream task for such datasets. Compared with clustering for unimodal spatial transcriptomics, spatial multi-omics clustering needs to simultaneously handle modality-specific scale differences, sparse noise, information complementarity and spatial dependence. Consequently, existing methods adopt diverse modelling pipelines [8]. Within the broader field of single-cell multimodal integration, weighted nearest neighbour (WNN) and MOFA+ offer transferable integration insights from adaptive neighbourhood construction and latent factor modelling respectively [9, 10]. Systematic benchmark studies further demonstrate that integration performance is jointly determined by pre-processing workflows, dataset scale and evaluation metrics [11]. Based on this problem framework, this paper first describes the data foundation and clustering objectives of spatial multi-omics research. Key challenges are then summarized from three perspectives: heterogeneity, complementarity and spatial consistency, and representative methods are categorized into architectural, statistical and semantic levels. Finally, this study illustrates the parameter sensitivity of basic clustering and trade-offs among evaluation metrics via Leiden parameter assessment, providing reproducible references for understanding design trade-offs in complex integration models.

2. Data foundation and clustering task of spatial multi-omics

Spatial multi-omics refers to a technical system that acquires two or more molecular modalities from the same tissue section while retaining spatial tissue coordinates. Common modality combinations include RNA+ADT, RNA+ATAC and trimodal datasets. The transcriptome records ribonucleic acid (RNA) expression; protein modalities are usually read out via antibody-derived tags (ADT); chromatin accessibility is characterized by assay for transposase-accessible chromatin using sequencing (ATAC). Taking SPOTS as an example, this platform captures messenger RNA (mRNA) through poly(A) probes on Visium slides and simultaneously detects proteins using DNA-barcoded antibodies, thus obtaining both transcriptomic and proteomic information at each spatial spot [6]. Spatial multi-omics therefore directly maps multi-layer molecular measurements onto tissue structures, laying a data foundation for subsequent cross-modal integration and spatial domain identification.

2.1. Multi-omics molecular modalities and their spatial measurements

From the perspective of molecular information flow, spatial multi-omics covers genomic, epigenomic, transcriptomic, proteomic and metabolomic layers (Table 1). These layers sequentially describe the genetic blueprint, regulatory status, transcriptional activity, functional execution and metabolic read-outs, and characterize cell states from distinct angles. The value of multi-omics

DOI: 10.54254/2755-2721/2026.36973

integration lies in assembling multi-layer observations from identical tissues. Cross-omics analysis further connects upstream and downstream molecular states, yielding more comprehensive and interpretable molecular evidence for spatial tissue architecture [3, 4].

Table 1. Overview of major molecular modalities in spatial multi-omics

Modality layer	Main contents and research objects	Representative (spatial) detection techniques	Roles and characteristics in spatial multi-omics integration
Genome	DNA sequences and genetic variations; relatively stable genetic blueprint	Co-detection at spatial level remains immature	Defines potential biological possibilities, but does not reflect real-time functional states
Epigenome	Regulatory states such as chromatin accessibility and histone modification	Spatial-ATAC-seq, Spatial-CUT&Tag	Up-stream switch for gene expression, reveals cell-fate regulation; signals are sparse
Transcriptome	Expression levels of mRNA and non-coding RNA	10x Visium, Stereo-seq, Xenium	Reflects ongoing transcriptional activity; broad coverage yet sparse and noisy
Proteome	Antibody-labelled protein abundance (ADT)	SPOTS, CODEX, CITE-seq-based antibody platforms	Direct executor of biological functions, strongly linked to phenotypes; low dimensionality and condensed information
Metabolome	Terminal small-molecule metabolic products such as sugars and lipids	Spatial metabolomics (e.g. MALDI-MSI)	Read-out most closely related to phenotypes; in-situ co-detection with transcriptome or proteome is still under development

2.2. Rationale for focusing on transcriptome, proteome and chromatin accessibility

Current spatial multi-omics workflows generally take the transcriptome as the backbone, combined with proteome (RNA+ADT) or chromatin accessibility (RNA+ATAC) measurements. Such combinations balance information complementarity and experimental feasibility. The transcriptome broadly reflects cellular expression states, the proteome provides read-outs closer to functional execution, and chromatin accessibility lies upstream of transcriptional regulation and supplements regulatory information associated with cell fate. By contrast, spatial genomic co-detection is not yet mature, and in-situ joint profiling between metabolome and transcriptome/proteome is still in the developmental stage. With platforms such as SPOTS and Spatial-ATAC-RNA gradually maturing, paired measurements of RNA-protein or RNA-ATAC have become the dominant application scenario for spatial multi-omics clustering [6].

2.3. Task definition and analytical goals of spatial multi-omics clustering

Spatial multi-omics clustering is an unsupervised analytical task for multimodal spatial datasets. Its objective is to combine molecular features and spatial topology to partition cells or spatial spots into subgroups with similar molecular patterns and relatively coherent spatial organization [8]. These subgroups are usually interpreted as spatial domains or spatial functional domains for characterizing structural heterogeneity within the tissue microenvironment. For instance, spatial domain identification in tumour tissues can reveal the spatial organization of tumour cells, immune cells and stroma. Nevertheless, computationally derived spatial domains are not equivalent to biologically

DOI: 10.54254/2755-2721/2026.36973

annotated regions, and their biological meanings still require validation with marker gene expression, histological images or independent annotations.

2.4. Heterogeneity, sparsity and spatial dependence of spatial multi-omics data

Spatial multi-omics datasets possess three key properties that directly affect integration and clustering performance. First, high-dimensional sparsity leads to abundant missing molecular signals. Technical noise may mask true biological differences and impair similarity estimation. Second, different modalities exhibit systematic discrepancies in numeric scales, distribution patterns and noise structures. For example, RNA counts typically follow count-based distributions, while protein abundance signals are closer to continuous values. Direct feature concatenation may cause high-variance modalities to dominate the joint representation. Third, neighbouring spots often share similar molecular states. Spatial autocorrelation can serve as a valid biological prior, yet over-strong spatial constraints may blur genuine tissue boundaries. Furthermore, low-dimensional modalities such as protein data carry highly condensed information, and critical features are easily lost during unified dimensionality-reduction workflows [8]. These data-related properties collectively give rise to the three core challenges described in subsequent sections.

3. Three core challenges in clustering-oriented integration of spatial multi-omics

Derived from the above-mentioned data properties, spatial multi-omics integration can be summarized as three interrelated problems: how to correct cross-modal heterogeneity in distribution, scale and noise; how to preserve both shared cross-omics information and modality-specific signals; and how to strike a balance between spatial continuity and genuine biological boundaries. These three targets correspond to the comparability of joint representations, information completeness and spatial reliability, respectively, and jointly determine the stability of spatial domain identification.

3.1. Cross-modal heterogeneity: discrepancies in distribution, scale and noise

Cross-modal heterogeneity originates primarily from distinct measurement mechanisms across different omics layers. Count-based transcriptomic signals and continuous protein or chromatin signals differ substantially in value range, noise structure and sparsity. Uncorrected feature concatenation tends to let high-variance modalities dominate similarity measurement and increases the difficulty for models to learn non-linear cross-modal correspondences [8]. Meanwhile, unified dimensionality reduction may cause disproportionate information loss for low-dimensional modalities, since their limited feature space contains few redundant signals that can be safely discarded. Data sparsity further undermines the stability of statistical estimation.

3.2. Cross-modal complementarity: decoupling of shared and modality-specific information

Different omics modalities share partial biological signals while retaining unique discriminative information. The key to effective fusion lies not in simple feature addition, but in distinguishing cross-omics consistent components from modality-specific components [12]. If only shared signals are retained, critical clues from single modalities may be overlooked. If each modality is modelled completely independently, alignment of common cellular states will become difficult. Therefore, joint representation should establish comparability within a shared latent space while reserving sufficient capacity for modality-specific signals.

DOI: 10.54254/2755-2721/2026.36973

3.3. Spatial consistency: coordination between tissue continuity and biological boundaries

Spatial consistency requires clustering outputs to form relatively continuous regions with clear boundaries, rather than discrete fragments driven by noise. Spatial autocorrelation provides a direct biological prior for this purpose, yet constraint strength must match actual tissue architecture. Excessively strong smoothing may merge real biological boundaries, whereas weak constraints fail to suppress local noise [13]. Studies on unimodal spatial transcriptomics also demonstrate that SpaGCN and STAGATE can strengthen local spatial consistency via graph convolution and graph attention [14, 15]; BayesSpace and GraphST introduce Bayesian spatial priors and self-supervised graph-based learning, respectively [16, 17]. Accordingly, spatial domain identification should not pursue continuity blindly, but preserve reliable boundary structures while retaining molecular differences.

4. Multi-layer integration strategies for addressing core challenges

To tackle the three categories of challenges described above, existing methods can be interpreted from three interconnected layers: architectural layer, statistical layer and semantic layer. The architectural layer defines which spatial or feature-based relationships are explicitly encoded by the model. The statistical layer determines how multimodal evidence is aggregated and constraints are imposed upon these relationships. The semantic layer introduces higher-order biological priors through pathway knowledge or pre-trained models. Organizing representative approaches by modelling objects rather than publication sequence facilitates direct comparison of how different architectures handle heterogeneity, complementarity and spatial consistency.

4.1. Architectural layer: modelling of spatial neighbourhood and graph relationships

According to explicit adjacency relationships, schemes at the architectural layer can be classified into single-graph, dual-graph and multi-graph strategies. Single-graph methods mainly encode spatial proximity. For instance, NicheCompass constructs a K-nearest-neighbour (KNN) graph based on spatial coordinates and aggregates omics features from each node and its neighbourhood [18]. Dual-graph methods can model spatial proximity and feature similarity simultaneously. SpatialGlue builds a shared spatial graph and modality-specific feature graphs for each modality [8]; another alternative sets independent modality channels on shared spatial topology, as adopted by COSMOS, which encodes each modality separately before cross-modal alignment [19]. Multi-graph schemes further incorporate semantic relationships. For example, besides the spatial graph and omics feature graph, spaLLM introduces an additional semantic graph derived from large language models [20]. The expansion from single-graph to multi-graph frameworks enables models to exploit spatial proximity, modality specificity and cross-modal complementary information concurrently.

4.2. Statistical layer: joint embedding and spatial regularization

Built on predefined architectures, the statistical layer is responsible for aggregating heterogeneous evidence and generating joint embeddings. Four major pathways are included: attention-based mechanisms, contrastive learning, probabilistic factor decomposition and ensemble consensus learning. Attention-based schemes adjust contributions from spatial neighbourhoods, feature channels and modalities via data-driven weights. SpatialGlue harmonizes spatial graphs and feature graphs within each modality first, and then generates unified representations through cross-modal

DOI: 10.54254/2755-2721/2026.36973

attention [8]; spaLLM treats the semantic graph as an extra view in multi-view attention computation [20]. Contrastive learning uses positive-negative sample constraints to align cross-modal structures. COSMOS takes embeddings of original inputs as positive samples and embeddings from permuted features as negative samples to preserve spatial patterns [19], while soFusion strengthens shared signals through cross-modal contrastive learning [21]. Probabilistic factor decomposition provides an explicit generative framework. SMOPCA unites multiple modalities with shared latent factors and encodes spatial priors using spatial kernel covariance [13]. Ensemble consensus methods exploit complementarity among base clusterers. SMODEL obtains consensus representations through anchor-based decomposition, element-weighted ensemble learning and dual regularization from spatial and high-order graphs [22]. For unpaired datasets, SWITCH establishes cross-modal translation via cyclic mapping and supports joint representations with generated pseudo-paired signals [23]. Different statistical mechanisms can be combined, yet their performance gains depend on dataset scale, modality composition and prior assumptions.

4.3. Semantic layer: high-order biological knowledge-driven constraints

The semantic layer introduces high-order biological knowledge into joint representations, mainly derived from gene semantics produced by pre-trained models and well-characterized biological pathways. spaLLM encodes molecular features into semantic tokens using the pre-trained foundation multi-omics model scGPT, supplements functional associations beyond direct molecular similarity, and adjusts the weight of semantic views according to local transcriptional patterns [20, 24]. NicheCompass constrains latent dimensions with signalling or regulatory pathways, so that each latent dimension corresponds to a specific biological programme and improves the interpretability of detected spatial domains [18]. Compared with the architectural and statistical layers, the semantic layer supplies a shared functional reference for different modalities. Nevertheless, its effectiveness relies heavily on pre-training corpora, pathway coverage and compatibility with target tissues.

Table 2. Methodological comparison of representative spatial multi-omics integration models

Model	Structural Layer Principle	Core Statistical Methods	Primary Challenges Addressed	Representative Limitations
SpatialGlue [8]	Spatial graph + modality-specific feature graphs	Dual-layer attention fusion within and across modalities	Heterogeneity, complementarity, and spatial consistency	The native framework is primarily designed for two modalities; additional modalities require further implementation and extension
COSMOS [9]	Bimodal channels on a shared spatial graph	GCN encoding + WNN weighting + DGI contrastive learning + spatial regularization	Complementarity and spatial consistency	The current framework is designed for two paired modalities; simple permutation-based negative samples have limited adaptability to highly separated structures

DOI: 10.54254/2755-2721/2026.36973

Table 2. (continued)

SMOPCA [13]	Spatial kernel covariance prior	Probabilistic factorization + shared latent factors	Heterogeneity and spatial consistency	The current implementation has a time complexity of $O(n^3)$ and a space complexity of $O(n^2)$, and models only normalized data
SMODEL [22]	Basic clustering graph + spatial higher-order graph	Anchor concept decomposition + element-weighted ensemble + dual-graph regularization	Heterogeneity, complementarity, and spatial consistency	Depends on multiple basic clustering results; iterative optimization increases computational overhead
SpaMICS [12]	Spatial graph + feature graph	Dual-branch GCN + shared/specific subspace disentanglement + dual constraints	Heterogeneity, complementarity, and spatial consistency	Requires learning global self-expression relationships, resulting in relatively high computational and memory overhead for ultra-large-scale data
soFusion [21]	Bimodal encoding on a shared spatial graph	GCN + within-omics/between-omics learning + modality-specific decoding	Heterogeneity, complementarity, and spatial consistency	Currently validated mainly on bimodal data; extending to ultra-large-scale data requires substantial computational resources
spaLLM [20]	Spatial graph + omics feature graph + semantic graph	Multi-view attention + scGPT semantic representations	Heterogeneity, complementarity, and interpretability	Depends on the quality of the pretrained model and gene semantics; does not natively support image modalities at present
NicheCompass [18]	Spatial neighborhood graph	Pathway-prior-constrained conditional variational graph autoencoder	Spatial consistency and interpretability	Depends on the quality of prior databases and gene coverage; relatively weaker performance on spot-level data

5. Parameter-sensitivity evaluation of the leiden baseline

Previous sections have summarized major integration pipelines for spatial multi-omics clustering from the architectural, statistical and semantic layers. Nevertheless, the performance of sophisticated integration models can still be affected by downstream clustering parameters and evaluation criteria. To generate reproducible empirical references for this issue, this chapter systematically evaluates parameter variations of Leiden clustering [25] on the *MouseBrain* and *HumanTonsil* datasets, and investigates the relationships among clustering granularity, consistency with reference labels and spatial continuity. This analysis focuses on Leiden baselines built upon simple concatenated joint features. Its objective is to reveal parameter sensitivity, rather than to produce a unified performance ranking for integration models such as SpatialGlue and SMOPCA.

DOI: 10.54254/2755-2721/2026.36973

5.1. Datasets, pre-processing and baseline settings

Two datasets with distinct modality compositions were used in the experiments. The *MouseBrain* dataset contains 9 215 spatial spots with jointly-measured RNA expression and chromatin accessibility, and nine reference spatial domains annotated by LayerName. The two modalities were standardized via StandardScaler before feature concatenation. The *HumanTonsil* dataset includes 2 492 spatial spots with joint measurements of gene expression and protein abundance, while no independent reference labels are available. Raw concatenated count values were processed sequentially with total-sum normalization (target sum = 10 000), log1p transformation and clipping scaling (max_value = 10). Principal components were then calculated on the joint-feature matrix, a nearest-neighbour graph was constructed, and Leiden clustering was performed. A fixed random seed (seed = 0) was adopted across all parameter combinations. Leiden clustering was implemented with igraph and run for two iterations to ensure reproducible comparisons among different configurations.

5.2. Parameter grid and evaluation metrics

The parameter grid contained three PCA dimensions (20, 30, 50), three nearest-neighbour values of k (10, 20, 30) and four Leiden resolution values (0.3, 0.6, 1.0, 1.5), yielding a total of $3 \times 3 \times 4 = 36$ evaluated combinations per dataset. The actual number of clusters was recorded for each result; no resolution-search procedure was used to force outputs to match a preset cluster count. For spatial-continuity assessment, a symmetric adjacency matrix was built using five spatial neighbours. Moran's I for binary labels was computed cluster-wise [26], and aggregated into weighted Moran's I weighted by cluster size. Adjusted Rand Index (ARI) [27] and Normalized Mutual Information (NMI) [28] were further calculated against LayerName annotations for *MouseBrain*. ARI and NMI were not reported for *HumanTonsil* owing to the absence of independent reference labels. Weighted Moran's I was used only for spatial-autocorrelation comparison, without permutation-based significance testing, and no statistical inference was drawn.

5.3. Impacts of parameter variation on clustering consistency and spatial continuity

Table 3. Summary of parameter sensitivity across different Leiden resolution values

Dataset	resolution	Mean cluster number	Cluster-number range	Weighted Moran's I
MouseBrain	0.3	6.78	5–8	0.6775
MouseBrain	0.6	9.89	8–12	0.5572
MouseBrain	1	13.33	10–18	0.4648
MouseBrain	1.5	19.00	15–26	0.3571
HumanTonsil	0.3	4.44	4–6	0.5754
HumanTonsil	0.6	7.00	5–9	0.5243
HumanTonsil	1	11.00	9–14	0.3998
HumanTonsil	1.5	15.11	12–20	0.3618

For the *MouseBrain* dataset, as resolution rose from 0.3 to 1.5, the mean cluster number increased from 6.78 to 19.00 (an increment of 180.3 %), and the cluster-number range expanded from 5-8 to 15-26. Over the same interval, weighted Moran's I decreased from 0.6775 to 0.3571 (a

DOI: 10.54254/2755-2721/2026.36973

reduction of 47.3 %) (Table 3). Consistent spatial trends are illustrated in Figure 1: lower-resolution outputs yield larger, relatively continuous spatial regions. Increasing resolution subdivides original regions into multiple local clusters and refines spatial boundaries.

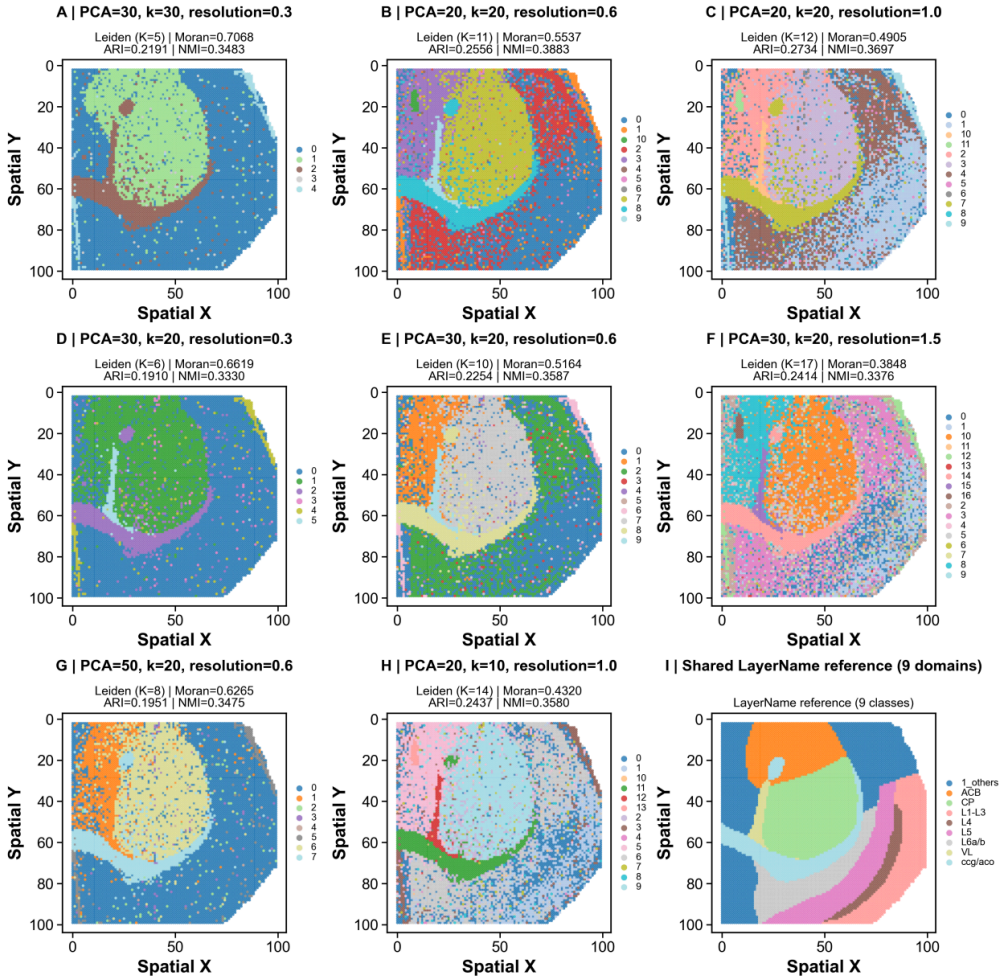

Figure 1. Parameter sensitivity of Leiden clustering on the *MouseBrain* dataset. Based on RNA-ATAC joint representations from 9 215 spatial spots, eight representative Leiden configurations (A-H) and the shared LayerName reference partition (I) are displayed. Panels A-C correspond respectively to the optimal configurations for weighted Moran's I, NMI and ARI. Panels D-F compare spatial-partition changes induced by increasing resolution from 0.3 to 1.5 under fixed PCA = 30 and k = 20. Panels G-H illustrate the separate effects of PCA-dimension and neighbour-number variation. Panel I shows the nine-class LayerName reference partition. Each sub-panel labels the actual cluster count K, ARI, NMI and weighted Moran's I

Different evaluation metrics reached their optimal values under distinct configurations in *MouseBrain*. The highest ARI (0.2734) was obtained at PCA = 20, k = 20, resolution = 1.0 with 12 clusters. The highest NMI (0.3883) occurred at PCA = 20, k = 20, resolution = 0.6 with 11 clusters. The maximum weighted Moran's I (0.7068) was achieved at PCA = 30, k = 30, resolution = 0.3 with 5 clusters. The weighted Moran's I at the ARI-optimal configuration was 0.4905, which was 0.2163 (30.6 %) lower than the spatial-continuity optimum. These results demonstrate that the optimal parameter points for reference-label consistency and spatial continuity do not coincide.

DOI: 10.54254/2755-2721/2026.36973

HumanTonsil exhibited a granularity-continuity trend consistent with *MouseBrain*. When resolution increased from 0.3 to 1.5, the mean cluster number rose from 4.44 to 15.11 (+240.0 %), while weighted Moran's I declined from 0.5754 to 0.3618 (−37.1 %). At resolution = 0.6, the mean cluster number was 7.00, yet different parameter combinations produced between 5 and 9 clusters. Two configurations yielding exactly seven clusters returned weighted Moran's I values of 0.5191 and 0.5246, respectively. Figure 2 shows that rising resolution splits formerly continuous tissue regions into more local clusters, indicating that neither identical resolution nor identical cluster count uniquely determines the final spatial partition.

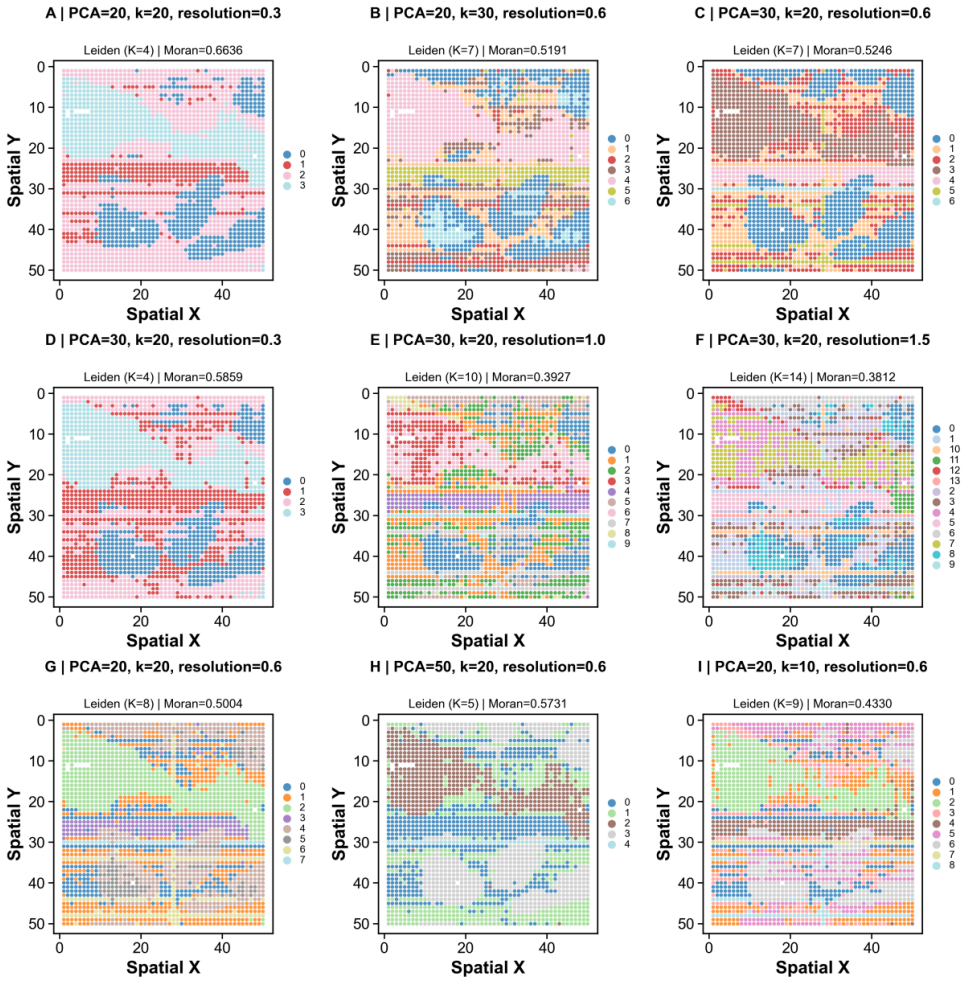

Figure 2. Parameter sensitivity of Leiden clustering on the *HumanTonsil* dataset. Based on gene-protein joint representations from 2 492 spatial spots, nine representative Leiden configurations (A-I) are presented. Panel A shows the configuration with maximum weighted Moran's I. Panels B-C produce seven clusters under different PCA-dimension and neighbour-number settings. Panels D-F demonstrate granularity shifts caused by resolution increments with fixed PCA dimension and k. Panels G-I further illustrate separate effects of PCA dimension, neighbour number and resolution. Each sub-panel labels the actual cluster count K and weighted Moran's I

DOI: 10.54254/2755-2721/2026.36973

5.4. Comprehensive comparison of parameter effects and boundaries of findings

Across both datasets, resolution exerted the most consistent granularity-regulating effect. With resolution increased from 0.3 to 1.5, mean cluster counts rose by 180.3 % (*MouseBrain*) and 240.0 % (*HumanTonsil*), while weighted Moran's I dropped by 47.3 % and 37.1 %, respectively. This shared trend confirms that higher resolution reliably generates finer spatial segmentations, although the magnitude of its influence on cluster number and spatial continuity varies substantially across datasets.

PCA dimension and neighbour-number k exerted relatively mild yet non-negligible impacts. On *MouseBrain*, raising PCA dimension from 20 to 50 reduced the mean ARI from 0.2400 to 0.2035 and the mean NMI from 0.3644 to 0.3195. Increasing k from 10 to 30 decreased the mean cluster number from 14.00 to 10.83, whereas weighted Moran's I increased from 0.4826 to 0.5404. For *HumanTonsil*, average weighted Moran's I values across the three PCA-dimension levels fell within the narrow range 0.4613–0.4711. When k was raised from 10 to 30, the mean cluster number decreased from 11.17 to 8.08. Taken together, resolution primarily controls clustering granularity, neighbour number modulates the degree of local merging, and PCA-dimension effects are more data-dependent.

Metric inconsistency further demonstrates that parameter selection constitutes a multi-objective optimization problem. In *MouseBrain*, ARI, NMI and weighted Moran's I attained their peak values under separate configurations, and the ARI-optimal setting yielded a spatial-continuity metric 30.6 % lower than the spatial-continuity optimum. In *HumanTonsil*, resolution = 0.6 produced a mean of 7.00 clusters but outputs ranged from 5 to 9 clusters under varying PCA and k values; two seven-cluster configurations returned distinct weighted Moran's I scores. Therefore, an identical cluster count does not guarantee an identical spatial structure, and no single evaluation metric can simultaneously reflect both label-reference consistency and spatial continuity.

The above conclusions are bounded to the current parameter grid, pre-processing workflow and Leiden baselines built on simple concatenated joint features. Observed parameter effects are derived from descriptive comparisons without significance testing, so causal interpretations cannot be drawn. ARI and NMI results for *MouseBrain* are conditional on LayerName annotations. Since no independent ground-truth labels are available for *HumanTonsil*, its weighted Moran's I values only reflect spatial continuity rather than the biological accuracy of clustering outputs. Computationally identified spatial domains still require biological validation using marker-gene expression, histological localization or external annotations.

6. Conclusion and outlook

Focusing on cross-modal heterogeneity, information complementarity and spatial consistency, this paper constructs an analytical framework for the integration of spatial multi-omics clustering. Representative approaches are accordingly categorized into three layers: architectural, statistical and semantic. The architectural layer defines spatial and feature-based relationships; the statistical layer completes multimodal evidence fusion; the semantic layer further incorporates pathway-related or pre-trained biological knowledge. Together, these three layers determine the comparability, information completeness and spatial reliability of joint representations. The 72 groups of Leiden-clustering results obtained from two datasets further reveal a stable yet dataset-dependent trade-off among clustering granularity, reference-label consistency and spatial continuity. This empirical experiment provides a reproducible parameter baseline, and demonstrates that comparisons among

DOI: 10.54254/2755-2721/2026.36973

complex integration models require unified standards for datasets, pre-processing workflows, parameter reporting and evaluation metrics.

Spatial multi-omics clustering still faces three major limitations concerning scalability, modality coverage and evaluation criteria. The computational complexity of some models increases quadratically or cubically with the number of spatial spots, which prevents their direct application to ultra-large-scale datasets. Most existing methods have been validated primarily on bimodal data, while their adaptability to trimodal or higher-order multimodal datasets, cross-sectional data and three-dimensional tissue data remains insufficient. Non-sequential information such as histological images has not been systematically exploited. Meanwhile, performance gains brought by large-language-model-derived or pathway-based priors are constrained by the quality of pre-training datasets and annotation coverage. Future research should prioritize the establishment of standardized benchmarks, develop scalable, interpretable integration models capable of preserving modality-specific signals, and link computationally detected spatial structures to verifiable biological meanings through independent annotations, histological evidence and cross-dataset validation.

References

[1] Stoeckius, M., Hafemeister, C., Stephenson, W., et al. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods, 14*, 865–868. https://doi.org/10.1038/nmeth.4380

[2] Ståhl, P. L., Salmén, F., Vickovic, S., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science, 353*(6294), 78–82. https://doi.org/10.1126/science.aaf2403

[3] Zahedi, R., Ghamsari, R., Argha, A., et al. (2024). Deep learning in spatially resolved transcriptomics: A comprehensive technical view. *Briefings in Bioinformatics, 25*(2), bbae082. https://doi.org/10.1093/bib/bbae082

[4] Khan, M., Arslanturk, S., & Draghici, S. (2025). A comprehensive review of spatial transcriptomics data alignment and integration. *Nucleic Acids Research, 53*(12), gkaf536. https://doi.org/10.1093/nar/gkaf536

[5] Liu, Y., Yang, M., Deng, Y., et al. (2020). High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell, 183*(6), 1665–1681.e18. https://doi.org/10.1016/j.cell.2020.10.026

[6] Ben-Chetrit, N., Niu, X., Swett, A. D., et al. (2023). Integration of whole transcriptome spatial profiling with protein markers. *Nature Biotechnology, 41*, 788–793. https://doi.org/10.1038/s41587-022-01536-3

[7] Liao, S., Heng, Y., Liu, W., et al. (2023). Integrated spatial transcriptomic and proteomic analysis of fresh frozen tissue based on Stereo-seq. *bioRxiv*. https://doi.org/10.1101/2023.04.28.538364

[8] Long, Y., Ang, K. S., Sethi, R., et al. (2024). Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nature Methods, 21*, 1658–1667. https://doi.org/10.1038/s41592-024-02316-4

[9] Hao, Y., Hao, S., Andersen-Nissen, E., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell, 184*(13), 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

[10] Argelaguet, R., Arnol, D., Bredikhin, D., et al. (2020). MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology, 21*, 111. https://doi.org/10.1186/s13059-020-02015-1

[11] Luecken, M. D., Büttner, M., Chaichoompu, K., et al. (2022). Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods, 19*, 41–50. https://doi.org/10.1038/s41592-021-01336-8

[12] Cai, G., Chen, F., Wen, K., et al. (2026). Identifying spatial domains from spatial multi-omics data using consistent and specific deep subspace learning. *Information Fusion, 125*, 103428. https://doi.org/10.1016/j.inffus.2025.103428

[13] Chen, M., Cheng, R., He, J., et al. (2025). SMOPCA: Spatially aware dimension reduction integrating multi-omics improves the efficiency of spatial domain detection. *Genome Biology, 26*, 135. https://doi.org/10.1186/s13059-025-03576-9

[14] Hu, J., Li, X., Coleman, K., et al. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods, 18*, 1342–1351. https://doi.org/10.1038/s41592-021-01255-8

[15] Dong, K., & Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications, 13*, 1739. https://doi.org/10.1038/s41467-022-29439-6

[16] Zhao, E., Stone, M. R., Ren, X., et al. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology, 39*, 1375–1384. https://doi.org/10.1038/s41587-021-00935-2

DOI: 10.54254/2755-2721/2026.36973

[17] Long, Y., Ang, K. S., Li, M., et al. (2023). Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications, 14*, 1155. https://doi.org/10.1038/s41467-023-36796-3
[18] Birk, S., Bonafonte-Pardàs, I., Miraki Feriz, A., et al. (2025). Quantitative characterization of cell niches in spatially resolved omics data. *Nature Genetics, 57*, 897–909. https://doi.org/10.1038/s41588-025-02120-6
[19] Zhou, Y., Xiao, X., Dong, L., et al. (2025). Cooperative integration of spatially resolved multi-omics data with COSMOS. *Nature Communications, 16*, 27. https://doi.org/10.1038/s41467-024-55204-y
[20] Li, L., Dong, L., Zhang, H., et al. (2025). spaLLM: Enhancing spatial domain analysis in multi-omics data through large language model integration. *Briefings in Bioinformatics, 26*(4), bbaf304. https://doi.org/10.1093/bib/bbaf304
[21] Yu, N., Li, W., Sun, X., et al. (2025). soFusion: Facilitating tissue structure identification via spatial multi-omics data fusion. *Briefings in Bioinformatics, 26*(5), bbaf513. https://doi.org/10.1093/bib/bbaf513
[22] Li, Y., Cai, G., Chen, F., et al. (2025). Unveiling spatial domains from spatial multi-omics data using dual-graph regularized ensemble learning. *Communications Biology, 8*, 945. https://doi.org/10.1038/s42003-025-08372-6
[23] Li, Z., Qu, S., Liang, H., et al. (2025). Integrative deep learning of spatial multi-omics with SWITCH. *Nature Computational Science*. https://doi.org/10.1038/s43588-025-00891-w
[24] Cui, H., Wang, C., Maan, H., et al. (2024). scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods, 21*, 1470–1480. https://doi.org/10.1038/s41592-024-02201-0
[25] Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports, 9*, 5233. https://doi.org/10.1038/s41598-019-41695-z
[26] Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1-2), 17-23. https://doi.org/10.1093/biomet/37.1-2.17
[27] Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*, 193-218. https://doi.org/10.1007/BF01908075
[28] Strehl, A., & Ghosh, J. (2002). Cluster ensembles—A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research, 3*, 583-617. https://www.jmlr.org/papers/v3/strehl02a.html