



STCC enhances spatial domain detection through consensus clustering of spatial transcriptomics data

Congcong Hu, Nana Wei, Jiyuan Yang, et al.

Genome Res. 2025 35: 1415-1428 originally published online May 12, 2025

Access the most recent version at doi:[10.1101/gr.280031.124](https://doi.org/10.1101/gr.280031.124)

References This article cites 54 articles, 5 of which can be accessed free at:
<http://genome.cshlp.org/content/35/6/1415.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

© 2025 Hu et al.; Published by Cold Spring Harbor Laboratory Press

Method

STCC enhances spatial domain detection through consensus clustering of spatial transcriptomics data

Congcong Hu,¹ Nana Wei,^{2,3} Jiyuan Yang,¹ Hua-Jun Wu,^{2,4,5} and Xiaoqi Zheng¹

¹Center for Single-Cell Omics, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China; ²Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Lymphoma, Peking University Cancer Hospital and Institute, Beijing 100142, China; ³The Guangxi Key Laboratory of Intelligent Precision Medicine, Guangxi Zhuang Autonomous Region, Nanning 530007, China; ⁴Center for Precision Medicine Multi-Omics Research, Institute of Advanced Clinical Medicine, Peking University, Beijing 100191, China; ⁵Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

The rapid advance of spatially resolved transcriptomics technologies has yielded substantial spatial transcriptomics data. Deriving biological insights from these data poses nontrivial computational and analysis challenges, of which the most fundamental step is spatial domain detection (or spatial clustering). Although a number of tools for spatial domain detection have been proposed in recent years, their performance varies across data sets and experimental platforms. It is thus an important task to take full advantage of different tools to get a more accurate and stable result through consensus strategy. In this work, we developed STCC, a novel consensus clustering framework for spatial transcriptomics data that aggregates outcomes from state-of-the-art tools using a variety of consensus strategies, including Onehot-based, average-based, hypergraph-based, and wNMF-based methods. Comprehensive assessments on simulated and real data from distinct experimental platforms show that consensus clustering significantly improves clustering accuracy over individual methods under varied input parameters. For normal tissue samples exhibiting clear layered structure, consensus clustering by integrating multiple baseline methods leads to improved results. Conversely, when analyzing tumor samples that display scattered cell type distribution patterns, integration of a single baseline method yields satisfactory performance. For consensus strategies, average-based and hypergraph-based approaches demonstrate optimal precision and stability. Overall, STCC provides a scalable and practical solution for spatial domain detection in spatial transcriptomics data, laying a solid foundation for future research and applications in spatial transcriptomics.

[Supplemental material is available for this article.]

Spatially resolved transcriptomics (SRT) techniques, which quantify gene expression profiles while preserving cellular spatial information, have emerged as a key technology for interrogating cell heterogeneity and tissue development (Bäckdahl et al. 2021). Leveraging SRT, researchers have delineated spatiotemporal transcriptomic landscapes across various developmental stages of human and mouse tissues such as brain and embryo (Sato et al. 2008; Moffitt et al. 2018; Carlberg et al. 2019; Chen et al. 2022; Zhang et al. 2023), offering useful insights into mammalian tissue development. Meanwhile, SRT has also been instrumental in elucidating mechanisms underlying complex disorders such as cancer (Berglund et al. 2018; Vickovic et al. 2019; Ji et al. 2020), rheumatoid arthritis (Vickovic et al. 2022), diabetes (Theocharidis et al. 2022), and Parkinson's (Jia et al. 2023) and Alzheimer's diseases (Chen et al. 2020) by probing disease-associated alterations in spatial gene expression patterns. Such investigations can potentially inform early biomarkers and therapeutic targets for complex diseases (Brennecke et al. 2013).

Depending on the platform adopted, SRT technologies quantify gene expression across scales of multicellular, single-cell, and subcellular resolutions (Dries et al. 2021), with representative technologies of 10x Genomics Visium, STARmap (Wang et al. 2018), and Stereo-seq (Chen et al. 2022), respectively. These multi-

resolution SRT data sets empower diverse downstream analyses including spatial domain detection (Liu et al. 2022; Yang et al. 2022; Chidester et al. 2023; Xu et al. 2024a), deconvolution (Andersson et al. 2020; Dong and Yuan 2021; Elosua-Bayes et al. 2021; Song and Su 2021), inference of spatially variable genes (SVGs) (Lopez et al. 2019; Welch et al. 2019; Abdelaal et al. 2020; Biancalani et al. 2021), cell-cell communication (Peng et al. 2022; Shao et al. 2022; Cang et al. 2023; Li et al. 2023; Raredon et al. 2023), and trajectory inference (Zhang et al. 2021; Shen et al. 2025). Among them, spatial domain detection is a crucial and pivotal step for many downstream tasks. For example, the cell type annotation of spatial domains is a prerequisite for understanding their interactions or communications within the tissue, as well as identification of spatial biomarkers across tissue. Recently, a number of algorithms including statistical (e.g., BASS [Li and Zhou 2022], BayesSpace [Zhao et al. 2021], SpatialPCA [Shang and Zhou 2022]) and deep learning (e.g., SEDR [Xu et al. 2024a], stLearn [Pham et al. 2023], SpaGCN [Hu et al. 2021], and STAGATE [Dong and Zhang 2022]) algorithms have been developed for spatial domain detection. However, their performances remain inconsistent across data sets and parameter settings (Cui et al. 2021; Hu et al. 2024; Yuan et al. 2024), making the selection of tools,

© 2025 Hu et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding authors: hjwu@pku.edu.cn, xqzheng@shsmu.edu.cn
Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280031.124>.

parameters, and integration of multiple clustering outcomes a challenging task.

One possible strategy to overcome the above challenge is consensus clustering, which aims to integrate multiple methods for enhanced accuracy and robustness (Hore et al. 2009). Consensus clustering has been widely adopted in bulk and scRNA-seq data (e.g., SC3 [Kiselev et al. 2017] and SAFE-clustering [Yang et al. 2019]). Basically, these methods apply diverse dimensionality reduction or resampling techniques on the gene expression matrix, followed by the same or different clustering algorithms to get a series of clustering outputs. Then, individual clustering results are integrated, through consensus matrices or hypergraph segmentation algorithms, to determine consensus clustering labels. Such consensus approaches have demonstrated significant potential to improve the clustering accuracy and stability of bulk and scRNA-seq transcriptomics data (Gan et al. 2018; Cui et al. 2021). However, to the best of our knowledge, consensus clustering frameworks crafted for spatial transcriptomics (ST) data remain scarce. Systematic assessments of clustering accuracy and robustness across various consensus frameworks are lacking.

To address this gap, we here encapsulated and implemented four consensus clustering frameworks, namely, Onehot-based, average-based, hypergraph-based, and weighted nonnegative matrix factorization (wNMF)-based methods, using clustering outcomes of seven typical spatial clustering algorithms as input. To systematically evaluate the performances of these four consensus strategies in terms of ST clustering accuracy and stability, we conducted comprehensive simulations and real data applications across different tissue origins and sequencing platforms. The findings of this comprehensive analysis yield a rich source of references and insights, aimed to guide and enhance future consensus clustering endeavors in ST.

Results

STCC overview

In this work, we present **s**patial **t**ranscriptome **c**onsensus **c**lustering (STCC), a consensus framework tailored for clustering of ST data (Fig. 1A). STCC amalgamates multiple clustering results from baseline algorithms through construction of a hypergraph matrix or a consensus matrix. It currently implements four consensus strategies, namely, Onehot-based, average-based, hypergraph-based, and wNMF-based to get final consensus output (for detail, see Methods). Among them, Onehot- and average-based strategies derive consensus labels by simply applying k -means clustering to the corresponding hypergraph matrix or consensus matrix; hence, they are referred to as naive consensus strategies. Two other approaches, namely, hypergraph-based and wNMF-based strategies, obtain consensus labels by employing more advanced algorithms, including hypergraph partitioning, nonnegative matrix factorization, and quadratic programming; they are termed advanced strategies. The efficacy of these consensus strategies is validated using diverse clustering evaluation metrics (for detail, see Methods).

To comprehensively assess the performance of these diverse consensus strategies, we have collected seven SRT data sets with manual annotations by pathologists as ground truth. These data sets are obtained from various sequencing platforms, species, and tissue origins (Fig. 1B); thus, they hold great diversity for evaluation of a consensus strategy. Subsequently, we applied seven spatial domain detection algorithms as baseline clustering meth-

ods to each data set (Supplemental Fig. S1A) and generated 10 distinct outcomes per algorithm by varying parameters or random seeds. Finally, we systematically evaluated the impact of different clustering inputs on the performance of consensus frameworks and specifically compared two scenarios: incorporating multiple clustering results from the same baseline clustering algorithm as input (termed as “single method”), and integrating results from all seven baseline algorithms as input (termed as “all methods”).

STCC enhances clustering performance on simulated data sets

We first evaluated the performance of consensus strategies using simulated data (Fig. 2). To achieve it, we generated a series of simulation data sets on the basis of mouse embryo data by Stereo-seq and mouse brain data by 10x Visium to investigate the influence of different factors, including numbers of spots, highly/spatially variable genes, levels of noise, and cell type numbers (for detail, see Methods). Taking SpaGCN as the baseline algorithm, we observed that when the proportions of highly variable genes (HVGs) and SVGs increase, clustering accuracy of baseline algorithm increases (from 0.5 to 0.62 for HVGs and 0.37 to 0.63 for SVGs) and reaches saturation at around 1800 HVGs and 2400 SVGs. Similar patterns are observed for all consensus strategies, highlighting the important role of HVGs and SVGs on clustering performance irrespective of single or consensus strategies usage (Fig. 2A,B). Increasing noise ratio corresponds to decreasing ARI, but consensus strategies are more tolerant across noise ratios (Fig. 2C; Supplemental Fig. S1C). For example, ARIs for average-based and Onehot-based consensus strategies decrease from 0.69 to 0.18 upon adding noise but still exceed baseline performance (Fig. 2C). We observed fluctuating but overall stable ARI as the number of cell types increased, when consensus strategies show consistently higher performance than the baseline algorithm (Fig. 2D; Supplemental Fig. S1D).

We next evaluated the impact of input cluster numbers on the performance of different consensus strategies. By fixing the true cluster number at 15 in simulation data, we varied the input cluster number k from five to 20 (Fig. 2E). The results indicate that consensus clustering methods exhibited greater robustness compared with the baseline clustering method (SpaGCN) when the input cluster number deviated from the true number. Specifically, for consensus strategies, clustering accuracy remained above 0.5 for k -values from nine to 20, peaking at $k=18$, which closely aligns with the true cluster number (Fig. 2E). Additionally, we calculated the standard deviation of the ARI for the baseline algorithms and four consensus strategies across different k -values. Compared to the baseline algorithms, the consensus strategies exhibited significantly lower standard deviations than the baseline, indicating more consistent clustering performance (Fig. 2F). Overall, consensus strategies enhanced clustering accuracy and noise resistance to a certain extent compared with the baseline clustering algorithm.

We then compared the runtime and memory usage of consensus strategies as well as baseline method. For runtime, the wNMF-based consensus strategy takes five to 10 times longer runtimes than naive strategies as the number of spots increased (Fig. 2G), which is anticipated owing to its iterative computation procedure. Regarding memory usage, the patterns are slightly different. Average-based, Onehot-based, and wNMF-based strategies generally require more memory than baseline methods (Fig. 2H), primarily owing to the necessity of constructing a $N \times N$ connectivity matrix. In contrast, the hypergraph-based consensus strategy shows markedly lower memory consumption.

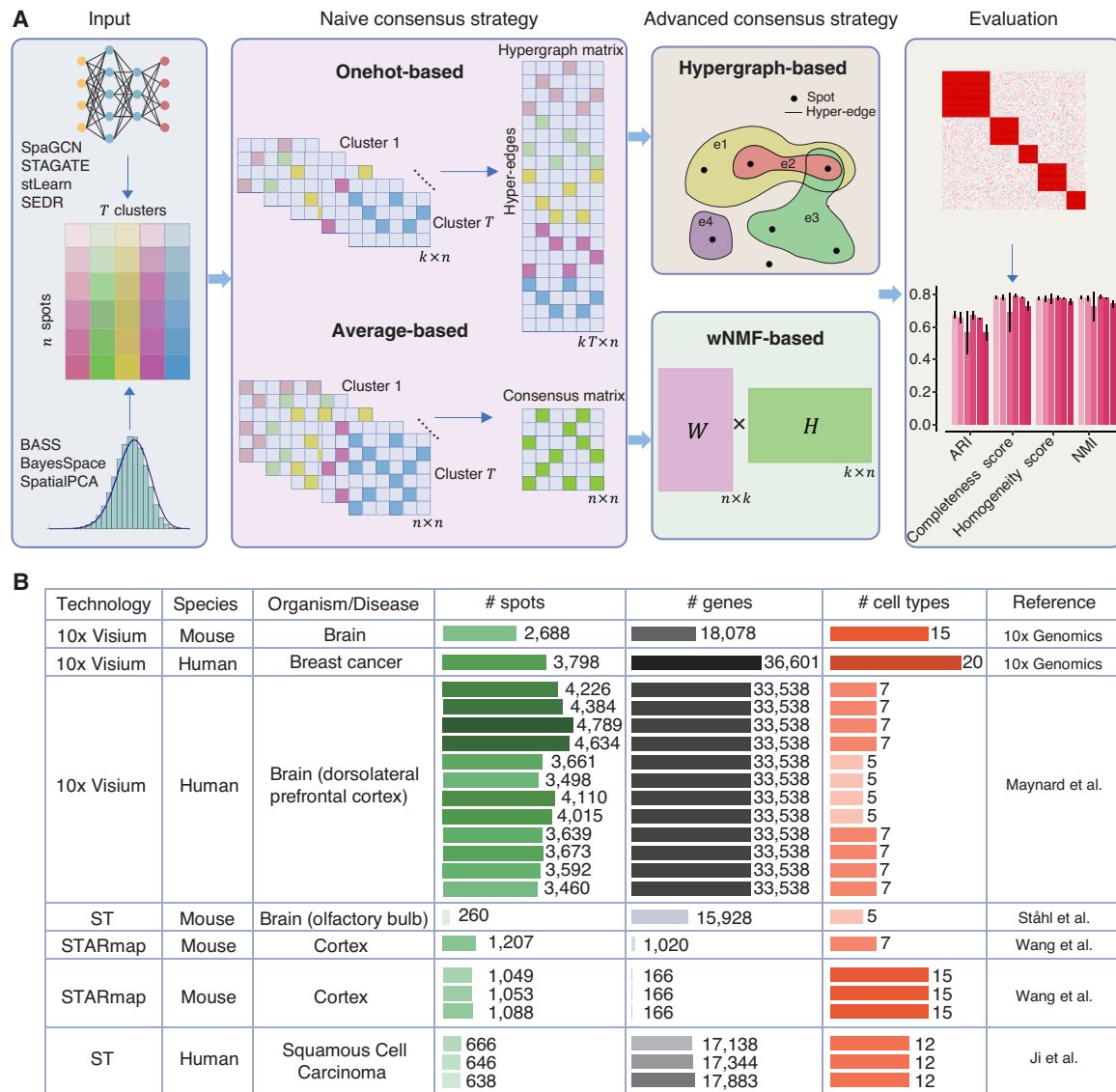


Figure 1. STCC architecture and evaluation data sets. (A) STCC architecture. STCC takes clustering results of seven baseline algorithms as input. It first constructs a hypergraph matrix and a consensus matrix and then executes two naive strategies (average-based, Onehot-based) and two advanced strategies (hypergraph-based, wNMF-based) to get consensus results. The ultimate clustering results are evaluated using seven benchmark metrics, namely, ARI, NMI, completeness score, homogeneity score, Calinski–Harabasz score, Davies–Bouldin score, and stability. (B) Details of seven benchmark data sets, including sequencing technologies, species, organism/disease types, numbers of spots, genes, and cell types, etc.

Consensus clustering achieves enhanced performance over individual baseline algorithms

We next assessed the impact of four consensus strategies on real SRT data using clustering results of single baseline algorithms as input. To this aim, we applied consensus strategies to seven baseline algorithms on mouse brain data and evaluated their performance via seven metrics, namely, ARI, NMI, completeness score, homogeneity score, Calinski–Harabasz score, Davies–Bouldin score, and stability (Fig. 3A; Supplemental Fig. S3A). We found that, in most cases, clustering outcomes show moderate enhancements after applying different types of consensus strategies. More importantly, compared to the baseline algorithms, consensus strategies show reduced variance in all evaluation metrics, especially notable in the “SEDR only” and “SpatialPCA only” scenarios (Fig. 3A). A

direct comparison of ARI scores (Fig. 3B) shows consensus strategies consistently outperform baseline algorithms, as evidenced by data points predominantly lying above the diagonal line. In Figure 3C, we visualized the outcomes of consensus clustering using BASS as baseline algorithm. Specifically, all four consensus strategies successfully reconstructed the Cortex_2 region, whereas the baseline algorithm incorrectly identifies it as striatum. Of note, the wNMF-based strategy even exclusively reconstructed the Cortex_5 structure, aligning well with the ground truth (Fig. 3C).

We next took sample 151675 from the human dorsolateral prefrontal cortex (DLPFC) data set as an illustrative case study. Our findings illustrated that, in comparison to baseline algorithms, all consensus strategies exhibit superior performance over individual baseline algorithms. Notably, evaluation metrics associated with the four consensus strategies demonstrate reduced

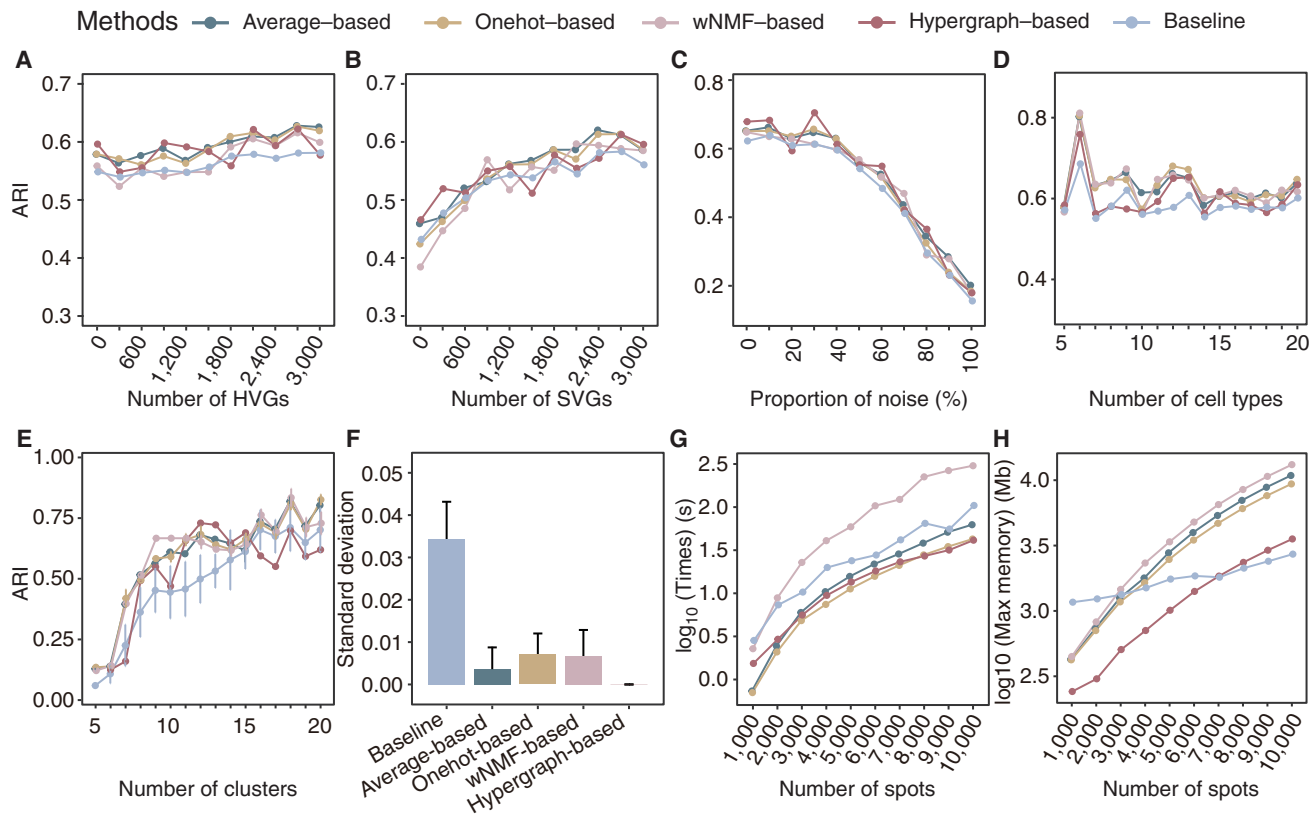


Figure 2. STCC performance on simulated data sets. (A,B) Line plots displaying the ARI for spatial domain detection on simulated data sets across different consensus strategies and baseline methods, with the y-axis showing how ARI varies as the number of added HVGs (A) and SVGs (B) increases. (ARI) Adjusted Rand index, (HVGs) highly variable genes, (SVGs) spatially variable genes. (C,D) Line plots showing the ARI by different consensus strategies and baseline method for spatial domain detection across simulated data sets change (y-axis) with the addition proportions of noise ratio (C) and number of cell types (D). (E) Line plot showing the ARI variation with the number of clusters. (F) The bar plots displaying the standard deviation of ARI from E for both the baseline algorithms and the four consensus strategies. (G,H) The runtimes (G) and maximum memory usage (H) of consensus strategies with varying spot numbers.

variances, especially in the case of “SEDR only” and “stLearn only” (Fig. 3D; Supplemental Fig. S4C). The scatter plot of ARI scores further confirms the superior performance of consensus strategies, with most data points positioned above the diagonal line (Fig. 3E). Visualization of clustering results reveals that BayesSpace fails to distinguish Layer_1 and Layer_2, whereas the Layer_1 identification by the four consensus strategies aligns closely with the ground truth (Fig. 3F). We further extended our analysis to all seven baseline algorithms, demonstrating the superior performance of consensus strategies across diverse data sets (Supplemental Figs. S2–S11). Visual representations showcasing clustering and consensus outcomes in additional data sets are also provided (Supplemental Figs. S12–S16). This comprehensive exploration enhances our insights into the robustness and applicability of consensus strategies, offering a nuanced perspective on their performance across various biological data sets.

Performance of baseline algorithms influences consensus accuracy

A prevalent hypothesis in the domain of consensus clustering is that the quality of baseline algorithm exerts a great influence on the performance of consensus strategies. To test this hypothesis in the task of spatial clustering, we ranked all baseline algorithms (from low to high) according to their average ARIs on four real

data sets and examined their final ARIs by different consensus strategies (Fig. 4A). As the ARI by baseline algorithm increases, all four consensus strategies exhibit a consistent improvement in clustering accuracy. This analysis confirms that the performance of the consensus strategies is significantly influenced by the choice of baseline algorithms. Furthermore, we observed that the clustering ARI by integrating different individual baseline algorithms on the mouse olfactory bulb data set demonstrates the biggest difference of 0.57 (from 0.14 to 0.71). The differences in the mouse brain data set, human breast cancer data set, and mouse cortex data set are 0.18 (from 0.53 to 0.71), 0.25 (from 0.35 to 0.6), and 0.28 (from 0.28 to 0.56), respectively. This inspires us that, in order to improve performance and stability when integrating different baseline algorithms, it may be helpful to explore different hyperparameters and produce diverse clustering results as input for consensus strategies.

Next, we examined the impact of the number of baseline algorithms on the performance of consensus strategies on real SRT data sets. To this end, we randomly selected a subset of baseline algorithms (from one to six for mouse brain, sample 151672 of DLPFC, and human breast cancer data; from one to five for mouse olfactory bulb data) for consensus clustering and compared them with the full model that integrating all baseline algorithms. It can be seen that for normal tissues with a layered structure, namely, mouse brain, DLPFC, and mouse olfactory bulb data, the

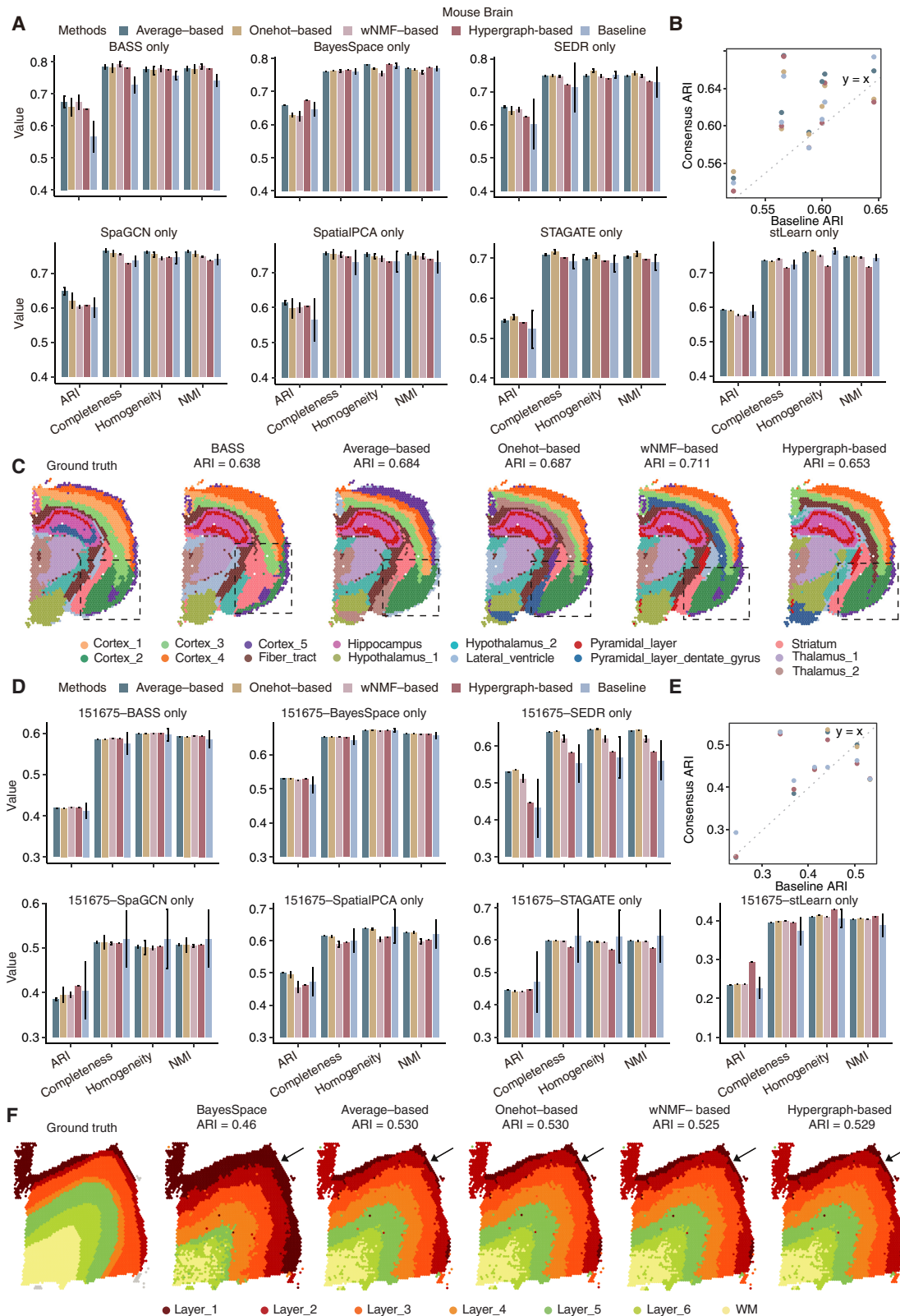


Figure 3. Performance of consensus strategies on single clustering algorithms. (A,D) Consensus clustering accuracies based on single baseline algorithms for the mouse brain data (A) and the DLPFC sample 151675 (D). Each baseline algorithm and consensus strategy are repeated 10 times to calculate the error bars. (B,E) Scatter plot comparing the ARI between baseline methods (x-axis) and consensus strategies (y-axis) for mouse brain data (B) and the DLPFC sample 151675 (E), with the diagonal line indicating equal performance. (C,F) Clustering results of mouse brain and DLPFC sample 151675 by different consensus strategies by using BASS (C) and BayesSpace (F) as baseline algorithms, respectively.

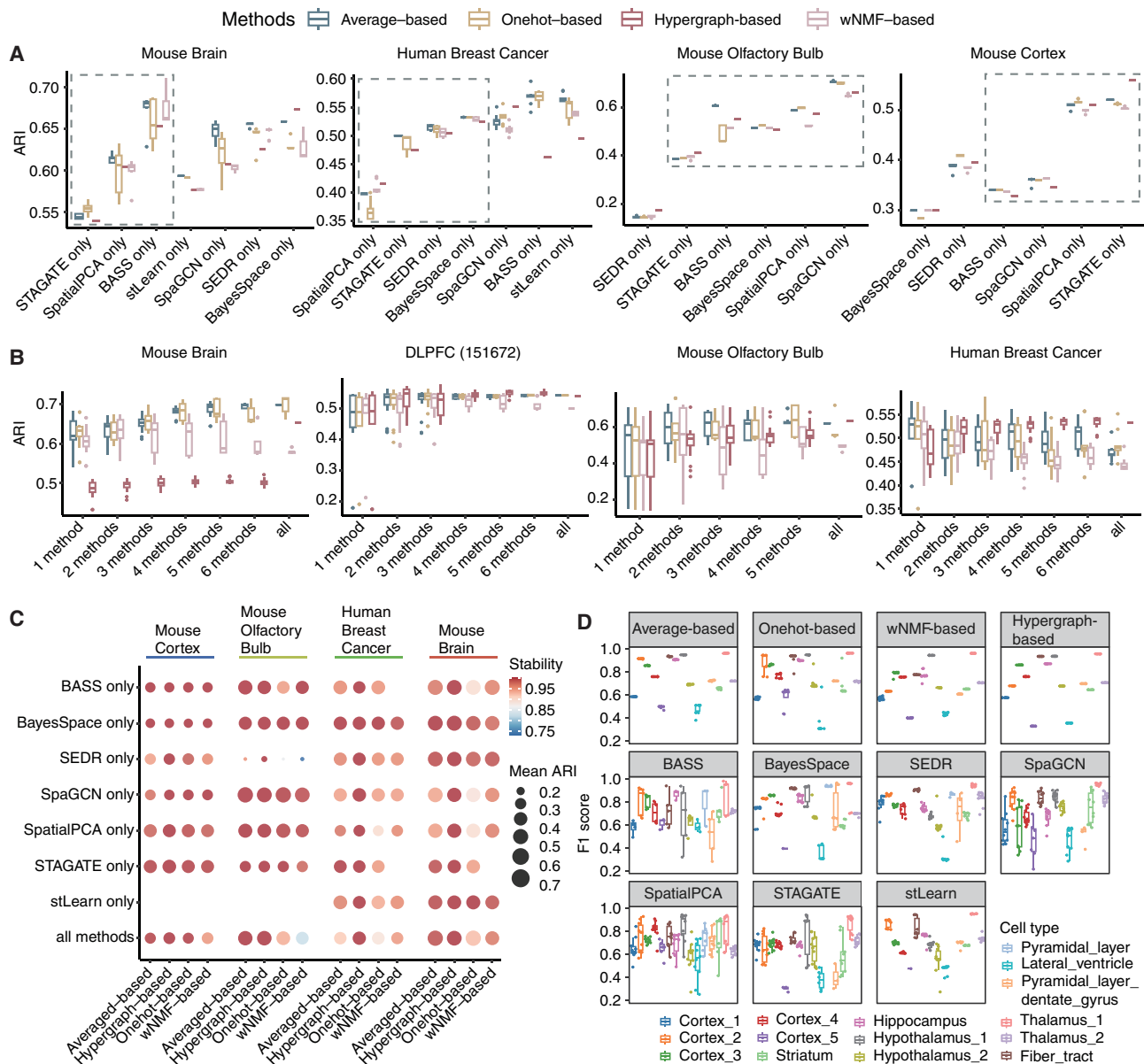


Figure 4. Comprehensive assessment of the accuracy and stability of four consensus strategies. (A) Consensus clustering results for four consensus strategies on seven baseline algorithms sorted from low to high according to average ARI. (B) Accuracies of four consensus strategies under different numbers of baseline algorithm. We randomly selected one to six baseline algorithms for mouse brain data, DLPFC sample 151672, and human breast data and selected one to five baseline algorithms for mouse olfactory bulb data and performed consensus clustering for each scenario 20 times. (C) Averaged ARI of four consensus strategies under “single method” and “all methods” situations. The size of the dots represents the mean ARI of 10 repeat experiments; color of the spots indicates stability. (D) Comparison of F1 scores across cell types for consensus strategies (top four panels) and baseline methods (bottom seven panels).

consensus performance increases with the number of baseline algorithms. The results of other samples of DLPFC data can be found in [Supplemental Figure S17](#). However, the same phenomenon is not observed for data with scattered cell type distribution patterns such as human breast cancer data; namely, integration of one single baseline algorithm achieves near optimal clustering result (Fig. 4B).

We then thoroughly evaluated overall accuracy and stability of the four consensus strategies on real data with manual annotation as the ground truth, in which accuracy was measured by average ARI across 10 consensus results, and stability measures the

consistency between 10 consensus results. Circle size in the figure signifies average ARI, with a larger circle indicating a higher value. Results consistently reveal that both average-based and hypergraph-based strategies exhibit the highest averaged ARI across all scenarios, followed by Onehot-based and wNMF-based strategies (Fig. 4C). Regarding stability, average-based and hypergraph-based consensus strategies exhibit the best resistance to randomness, whereas Onehot-based and wNMF-based strategies exhibit weaker stability (for the DLPFC data set, see [Supplemental Fig. S18A,B](#)).

To further evaluate the performance of different methods on a more detailed level, we examined the clustering accuracy at the

cell type level (Mölbart and Haghverdi 2023), as measured by F1 score (Fig. 4D). Our consensus strategies demonstrated consistently high F1 scores across diverse cell populations, including both abundant and rare cell types. In contrast, baseline methods showed considerable variability in their performance. Although some baseline methods such as BayesSpace performed reasonably well for certain cell types, they often struggled with rare cell populations and regions with thin spatial layers. This cell type-level analysis reveals that our consensus approaches effectively overcome the limitations of individual methods, providing more reliable and consistent cell type identification across different cellular populations.

Based on the above analyses, we concluded that the four consensus strategies showed substantial differences in different data sets and scenarios. Average-based and hypergraph-based consensus strategies achieve high performance in integrating baseline algorithms and demonstrate superior accuracy and stability. The optimal number of baseline algorithms to be integrated should be determined by the cell type distribution pattern of the tissue sample. This nuanced evaluation guides appropriate consensus strategies given baseline algorithm and data characteristics across data sets.

Evaluation of consensus clustering strategies on the mouse cortex data set

In this section, we applied STCC to the single-cell resolution mouse cortex data set generated by the STARmap technique (Wang et al. 2018). We focused on three consecutive tissue sections, BZ5, BZ9, and BZ14, that were assayed in the medial prefrontal cortex (mPFC) region of mice. These sections comprise 1049, 1053, and 1088 cells, respectively, each profiling the expression of the same set of 166 genes. The cells were annotated into 15 distinct cell types, including astrocytes (Astro), endothelial cells (Endo), and oligodendrocytes (Oligo), etc. We first compared the performance of four consensus strategies against individual baseline algorithms using the true number of clusters as input. Consistent with previous discoveries, the performance of consensus strategies is highly dependent on the selection of baseline algorithms. When using SEDR and SpaGCN as baselines, all four consensus strategies demonstrated significant advantages, outperforming the baselines across multiple evaluation metrics (Fig. 5A). In contrast, integrating SpatialPCA as a baseline algorithm improves clustering robustness but does not change the overall accuracy. For BASS, the average-based and Onehot-based consensus strategies surpass the baseline in terms of ARI, NMI, completeness score, and homogeneity score, whereas the hypergraph-based and wNMF-based strategies do not. No significant change in clustering accuracy is observed with BayesSpace and STAGATE as baseline algorithms (Fig. 5A). These results indicate the interplay between baseline methods and consensus strategies, highlighting the necessity to carefully select consensus strategies for STCC applications to real data sets.

In practical applications, the true number of cell types is often unknown. To address this challenge, the STCC method is equipped to automatically determine the optimal number of clusters using a silhouette-based strategy, making it particularly suited for exploratory data sets in which the true number of clusters or spatial domains remains uncertain. Specifically, it calculates the silhouette score for a specified range of cluster numbers and selects the number with the highest score as the final cluster number. As indicated in Supplemental Figure S19A, the predicted cluster num-

ber ($k=14$) with the maximal silhouette score closely approximates the true number ($k=15$). These findings highlight the resilience of specific consensus strategies in maintaining clustering accuracy despite variations in the number of clusters, reinforcing their applicability in complex ST data sets.

Finally, we visualized the clustering results and performed trajectory inference. By taking the BZ5 slice as an example, all four consensus strategies successfully identified the hierarchical structures corresponding to the eL2/3 and eL6-1 cell types in the ground-truth annotations. In contrast, the baseline algorithms, especially BayesSpace, SEDR, SpatialPCA, and SpaGCN, exhibit more chaotic clustering in these regions, failing to display clear hierarchical structures (Fig. 5B). The UMAP embeddings demonstrate that all four consensus strategies effectively preserve the underlying cluster structures, with clear separation between different cell types that closely matches the ground-truth annotations (Fig. 5C). This dimensional reduction visualization further validates the robustness of our consensus approaches in capturing biologically meaningful cellular organizations. Clustering results for the other slices are shown in Supplemental Figures S20 and S21.

Consensus clustering and trajectory inference for the squamous cell carcinoma data

We further applied STCC to highly heterogeneous human squamous cell carcinoma (SCC) samples generated by the ST technique, which comprise 12 tissue sections from four patients. Following the method of Ji et al. (2020), we focused on three tissue sections from patient 2 as a representative example (Ji et al. 2020). Unlike previous data sets, this data set lacks detailed spatial domain annotations. Therefore, we used the histopathologist-annotated tumor and nontumor regions (P2_ST_rep2) as a silver standard for model evaluation (Fig. 6A).

By incorporating the clustering results of six baseline methods as input, our consensus strategies effectively pinpoint the nontumor region at the lower part of each section (Fig. 6A). All four consensus strategies reveal clear structural patterns in this region (highlighted by black boxes), consistent with the coarse annotations. In contrast, the clustering results from the six baseline algorithms are highly scattered, lacking clear structure in the same region (Fig. 6B). Based on the clustering results from the consensus strategies, we next explored the differentiation relationship between tumor and nontumor regions. By inferring trajectories and pseudotime in this region, we identified a path tracing from nontumor regions toward tumor regions and found that the surrounded nontumor cells had a relatively earlier developmental time compared with the tumor cells (Fig. 6C). These findings provide valuable insights into the interconnectedness of the tumor region and neighboring regions, laying the foundation for a deeper understanding of tumorigenesis.

In summary, our consensus strategies provide valuable insights and robust support for downstream analyses compared with baseline methods, even for exploratory data sets in which the true number of cell types or spatial domains is unknown.

Discussion

The rapid accumulation of SRT data poses a great challenge for various downstream data analyses, especially for the most fundamental spatial domain detection step. Although a number of statistical or deep learning-based algorithms have been proposed for this task, there is no single algorithm that can achieve optimal results

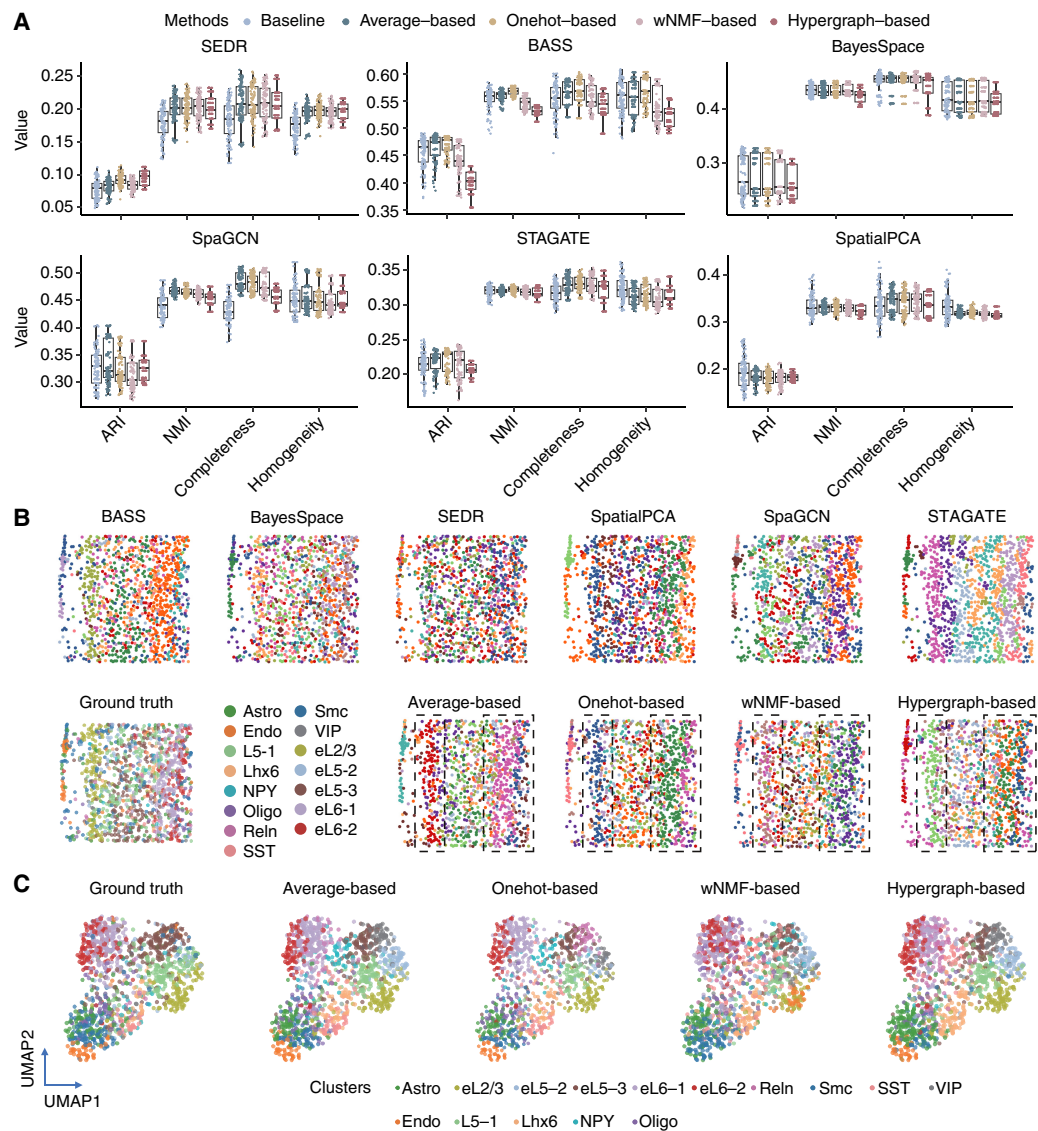


Figure 5. Performance of consensus clustering strategies on the mouse cortex data. (A) Consensus clustering accuracy of the BZ5 sample of the mouse cortex data set based on a single baseline algorithm. Each baseline algorithm and consensus strategy were repeated 10 times. (B) Visualization of clustering results of the baseline algorithm (*top*) and four consensus strategies (*bottom right*) on the BZ5 sample of mouse cortex data. (C) UMAP visualization comparing ground-truth cell type labels (*leftmost*) with clustering results from four consensus strategies.

across all data sets. Consequently, there is an urgent need for a consensus framework that can harness the strengths of different algorithms to enhance clustering accuracy and stability. Although existing consensus frameworks such as SC3 and SAFE-clustering are designed specifically for scRNA-seq data, their performance on SRT data has not been tested. It is thus an important necessity to develop and evaluate different consensus strategies for SRT data, with a particular focus on their compatibility within different pipelines, platform dependence, accuracy, and efficiency when applied to SRT data.

To fill this gap, we proposed a scalable and flexible ST consensus framework called STCC. By integrating clustering results from both single and multiple baseline algorithms, STCC generally enhances clustering performance across diverse data sets and scenarios, particularly in cases in which baseline methods encounter challenges with complex spatial structures. The comprehensive

analysis of the four consensus strategies highlights the preference for average-based consensus strategy, attributed to their simplicity, high accuracy, and stability. Notably, hypergraph-based exhibits exceptional stability and distinct advantages in the integration of vast data sets, wNMF-based underscores the enhancement in performance through consensus weighting. When it comes to the selection of consensus algorithms, we recommend (1) thoroughly exploring parameters of single baseline algorithm that could impact clustering results to attain the optimal consensus outcome, (2) adopting the more robust average-based and high effective hypergraph-based consensus strategies, and (3) selecting the appropriate number and type of baseline clustering algorithms based on the tissue origin of the samples.

Although demonstrating valuable results, our consensus framework still suffers from several limitations that should be addressed in future work. First, the current consensus framework

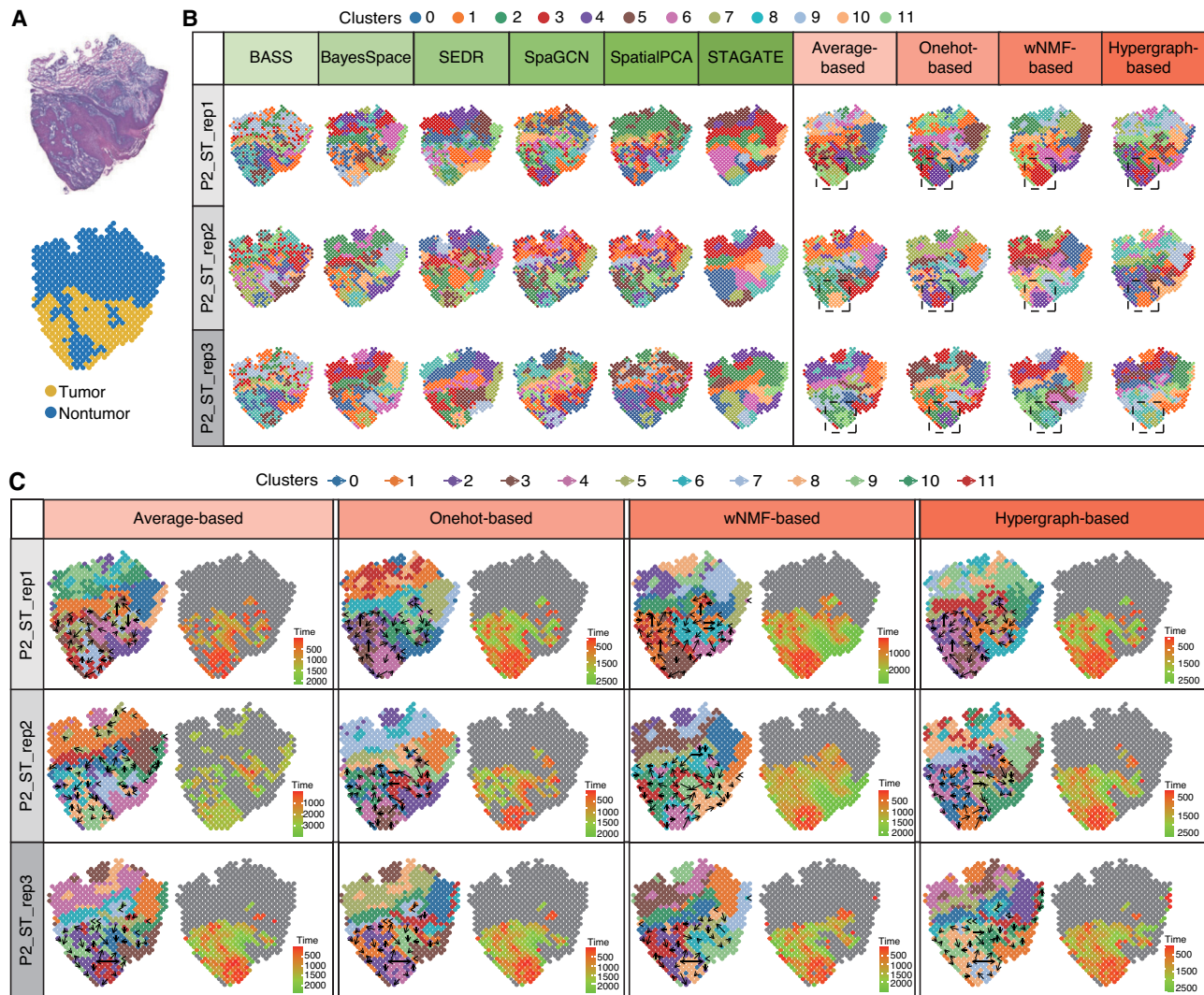


Figure 6. Exploratory analysis of squamous cell carcinoma data. (A) Hematoxylin and eosin (H&E)-stained images (*top*) of squamous cell carcinoma samples, accompanied by corresponding rough annotations (*bottom*). (B) Spatial domain detection of six baseline algorithms and four consensus strategies in the analysis of squamous cell carcinoma patient 2. (C) Trajectory inference and pseudotime analysis of four consensus strategies in squamous cell carcinoma patient 2.

takes only clustering outputs from baseline methods as input. It would be beneficial to explore the integration of various methods from the raw or processed data, such as the joint embedding from deep learning-based methods. This could potentially enhance the performance and robustness of the consensus clustering. Second, the current version of our consensus framework only incorporates seven tools, which is relatively limited considering the vast number of clustering tools currently available. To ensure a more robust consensus outcome, a broader range of clustering tools should also be incorporated into our consensus framework.

Methods

Baseline clustering algorithms for SRT data

We selected seven representative algorithms for spatial domain detection as the baseline algorithms for our consensus framework. Among them three are statistical based, namely, BASS (Li and Zhou 2022), BayesSpace (Zhao et al. 2021), and SpatialPCA (Shang and

Zhou 2022), and four of them are deep learning based, namely, SEDR (Xu et al. 2024a), stLearn (Pham et al. 2023), SpaGCN (Hu et al. 2021), and STAGATE (Dong and Zhang 2022). These tools are widely used and highly recognized for the task of spatial domain detection. In detail,

1. BayesSpace is a Bayesian statistical method designed to enhance the resolution of SRT data and perform clustering by incorporating spatial neighborhood information as priors.
2. SpatialPCA is a spatially-aware dimensionality reduction method specifically designed for SRT data. Its core concept involves using the probabilistic PCA model to infer a low-dimensional representation of gene expression data while accounting for the underlying spatial correlation structure; thus, the inferred low-dimensional components are used for downstream analyses including spatial domain detection.
3. BASS is a Bayesian hierarchical model for simultaneous cell type and spatial domain detection across multiple scales and samples.
4. stLearn is a Python software package that utilizes histological image-derived morphological distances and spatial

neighborhoods to smooth expression data to enable downstream analyses, such as spatial domain detection and trajectory inference.

5. SEDR is an autoencoder-based deep learning method that jointly captures low-dimensional embeddings of gene expression and spatial information in SRT data, which can be utilized for downstream analysis tasks such as clustering, trajectory inference, and batch effect correction.
6. SpaGCN is a graph convolutional network method to integrate multimodal data, including expression, spatial location, and histology to detect spatial domains.
7. STAGATE is a graph attention autoencoder to learn low-dimensional embeddings of SRT data by integrating gene expression and spatial information. The low-dimensional embedding can be utilized for spatial domain detection and denoising.

Simulated data for benchmarking consensus clustering algorithms

We conducted extensive simulations to evaluate the performance of the consensus strategies. The first simulation data (SimuData 1) is built from the mouse embryo data generated by Stereo-seq, with the purpose of investigating the impact of the number of spots on the runtime and maximum memory usage of consensus strategies. To achieve it, we downloaded the E12.5_E1S3 mouse embryo data from the MOSTA database, which comprises 49,908 spots quantitated by 17,337 genes, with a median gene count of 1237 per spot. We generated 15 data sets by randomly sampling a number of spots from this data set (with spot number ranging from 1000 to 15,000).

The second simulation data (SimuData 2) is obtained from mouse brain coronal data through the 10x Visium platform, which comprises 2688 spots measured on 18,078 genes (for details, see the next section). To investigate how the proportion of HVGs influences the performance of the consensus strategy, we identified 3000 HVGs using the “pp.highly_variable_genes” function in the SCANPY package (Wolf et al. 2018), leaving the remaining 15,078 genes as non-HVGs. We then created 11 data sets by combining varying numbers of HVGs with all non-HVGs, with the number of HVGs ranging from zero to 3000 in increments of 300. Analogy to SimuData 2, we constructed SimuData 3 by only replacing HVGs as SVGs inferred by “gr.spatial_autocorr” function from the Squidpy package (Palla et al. 2022).

Two additional simulation data sets (SimuData 4 and 5) are also based on the mouse brain coronal data. Among them, SimuData 4 is generated by adding random noises (with Gaussian distribution of mean zero and increasing standard deviations) on gene expression data, and SimuData 5 is constructed by selecting five to 20 cell types from all 20 annotated cell types to examine the impact of noise level and cell type number on the performance of the consensus strategy, respectively.

The final simulated data set (SimuData 6) was also constructed based on the mouse brain coronal data. It was designed to explore how the number of clusters specified in baseline algorithms affects the performance of various consensus strategies. We used cluster numbers ranging from five to 20 for the baseline algorithm SpaGCN, generating corresponding clustering results, which were then used as input for the consensus strategies.

For all above simulation data sets, we selected SpaGCN as the baseline algorithm and used its clustering outcome as input for different consensus strategies.

Real SRT data from different platforms for benchmarking consensus clustering algorithms

We downloaded seven publicly available SRT data sets with pathological annotation as gold standard to evaluate different consen-

sus strategies. The first is the mouse brain coronal data mentioned in previous data simulation section. It spans various brain regions, encompassing 15 hierarchical structures including cortex, hippocampus, hypothalamus, and pyramidal layer. The second is human breast cancer data profiled by 10x Visium platform, which comprises 3798 spots and 36,601 genes, with a median gene count of 7943 per spot. This data set was manually annotated to 20 spatial regions, including DCIS/LCIS, IDC, tumor edge, and healthy regions. The third is human DLPFC (Maynard et al. 2021) measured on the 10x Visium platform. The data set comprises 12 tissue sections with 33,538 gene expression values from three adult donors. We acquired manually annotated labels for seven laminar clusters, including six cortical layers from L1 to L6 and the white matter (WM), as the gold standard for this data set, from the original publication. The fourth is the mouse olfactory bulb data (Rep11_MOB_ST) generated by ST technology (Xu et al. 2024b), which comprises 260 spots and 15,928 genes. Following the annotation, it is categorized into five hierarchical structures, namely, granular cell layer, mitral cell layer, outer plexiform layer, glomerular layer, and olfactory nerve layer. The fifth data set is the mouse cortex data obtained from online resources provided in the original study (Wang et al. 2018), comprising 1207 spots and 1020 genes. The data set is annotated into seven layers: corpus callosum (CC), hippocampus (HPC), layer 1 (L1), layer 2/3 (L2/3), layer 4 (L4), layer 5 (L5), and layer 6 (L6). Note that the fourth and fifth data sets did not include the necessary H&E images or other histological images required to run the stLearn algorithm. As a result, these two data sets do not contain the baseline results for the stLearn method. The sixth data set consists of three consecutive sections (BZ5, BZ9, BZ14) from the mouse cortex data, which measure the same set of 166 genes across 1049, 1053, and 1088 spots, respectively. Cells in this data set were annotated as 15 distinct cell types, including but not limited to Astro, Endo, and Oligo. The final data set was generated from human SCC using the ST technology. For this study, we downloaded three consecutive sections from patient 2. The data sets comprise 666, 646, and 638 spots, which measure the expression of 17,138, 17,344, and 17,883 genes, respectively. The detailed information of the above seven benchmark data sets is summarized in Figure 1B.

Overview of consensus strategies

Our proposed consensus strategies aim to integrate clustering results from diverse baseline algorithms, which can be represented as a hypergraph or a consensus matrix. The hypergraph matrix is generated by onehot encoding of each clustering result and concatenating them row-wise, in which a column corresponds to a clustering outcome of an algorithm for each spot/single cell and a row corresponds to a cluster. The consensus matrix is obtained by aggregating the connectivity matrices from individual clustering outputs, with each entry of the consensus matrix indicating the frequency of two spots assigning to the same cluster. Based on the obtained hypergraph and consensus matrices, the above four STCC consensus strategies can be elaborated as follows.

1. Onehot-based strategy simply applies *k*-means clustering to the hypergraph matrix to obtain consensus labels.
2. Average-based strategy applies *k*-means algorithm to the consensus matrix to derive consensus labels.
3. Hypergraph-based strategy employs hypergraph partitioning algorithms on the hypergraph matrix to derive consensus labels. We currently implemented three types of hypergraph partitioning algorithms, namely, hypergraph partitioning algorithm (HGPA), metaclustering algorithm (MCLA), and cluster-based similarity partitioning algorithm (CSPA). In detail,

HGPA uses the weight of cutting hyperedges as the objective function to assess the partitioning quality of the hypergraph. MCLA first constructs a graph based on the hypergraph, with each node representing a spot and the weight of each edge measuring the number of times two spots being assigned to the same cluster. It then applies graph segmentation algorithm to partition it into categories to generate clustering labels. CSPA first computes similarities between hypergraphs (measured by Jaccard index) and then applies spectral clustering to the similarity matrix to obtain the final clustering result.

4. The wNMF-based consensus strategy assigns different weights to the connectivity matrices obtained from each baseline algorithm, that is,

$$\tilde{M} = \sum_{t=1}^T w_t M(P^t), \text{ s.t. } \sum_{t=1}^T w_t = 1, w_t \geq 0.$$

We define the binary connectivity matrix $M(P^t)$ for the t th clustering result P^t , in which each element indicates whether two spots were assigned to the same cluster. In detail,

$$M_{ij}(P^t) = \begin{cases} 1, & \text{if } P_i^t = P_j^t, \\ 0, & \text{otherwise.} \end{cases} \quad i, j = 1, \dots, n,$$

where P_i^t and P_j^t are clustering labels of the spots i and j , respectively. The total number of spots is denoted by n .

Based on the definition of the connectivity matrix, we define the distance between two clustering as

$$d(P^1, P^2) = \sum_{i,j=1}^n [M_{ij}(P^1) - M_{ij}(P^2)]^2.$$

The goal of consensus clustering is to find the optimal clustering P^* that is the closest to all clustering results obtained from individual baseline methods, which is defined under the following optimization objective:

$$\min_{P^*} J = \frac{1}{T} \sum_{t=1}^T d(P^t, P^*) = \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^n [M_{ij}(P^t) - M_{ij}(P^*)]^2.$$

Denote the consensus matrix as $\tilde{M} = \frac{1}{T} \sum_{t=1}^T M(P^t)$, and let $U_{ij} = M_{ij}(P^*)$. The objective function can then be rewritten as

$$\begin{aligned} J &= \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^n [M_{ij}(P^t) - \tilde{M}_{ij} + \tilde{M}_{ij} - U_{ij}]^2 \\ &= \Delta M^2 + \sum_{i,j=1}^n (\tilde{M}_{ij} - U_{ij})^2 \\ &= \|\tilde{M} - U\|^2, \end{aligned}$$

where $\Delta M^2 = \frac{1}{T} \sum_{t=1}^T \sum_{i,j=1}^n (M_{ij}(P^t) - \tilde{M}_{ij})^2$ is a constant, and \tilde{M} and U are matrices of \tilde{M}_{ij} and U_{ij} , respectively.

The adjacency matrix U can be represented and constrained using a cluster indicator matrix $H = \{0, 1\}^{k \times n}$, where H_{ij} indicates whether sample j is assigned to cluster i , and k represents the number of clusters. This formulation allows us to express the consensus clustering solution as $U = H^T H$, which simplifies the optimization process and ensures the clustering

constraints are satisfied. The consensus clustering problem thus becomes

$$\min_H \|\tilde{M} - H^T H\|^2. \quad (4.1)$$

Given the definition of H , $HH^T = D$, where D is a $k \times k$ diagonal matrix. However, D is unknown until the problem is solved, so we eliminate D by defining $\tilde{H} = H^T (HH^T)^{-1/2}$. This leads to the following relations:

$$H^T H = \tilde{H} D \tilde{H}^T, \quad \tilde{H}^T \tilde{H} = H^T (HH^T)^{-1} H^T = I.$$

The final optimization problem is therefore transformed into a more tractable form:

$$\min_{\tilde{H}^T \tilde{H} = I, \tilde{H}, D \geq 0} \|\tilde{M} - \tilde{H} D \tilde{H}^T\|^2, \text{ s.t. } D \text{ diagonal.}$$

The optimal values of \tilde{H} and D can be calculated by the following multiplicative updating process in each iteration until convergence:

$$\begin{aligned} \tilde{H}_{jk}^{(t+1)} &= \tilde{H}_{jk}^{(t)} \sqrt{\frac{(\tilde{M} \tilde{H} D)_{jk}}{(\tilde{H} \tilde{H}^T \tilde{M} \tilde{H} D)_{jk}}}, \\ D_{kl}^{(t+1)} &= D_{kl}^{(t)} \sqrt{\frac{(\tilde{H}^T \tilde{M} \tilde{H})_{kl}}{(\tilde{H}^T \tilde{H} D \tilde{H}^T \tilde{H})_{kl}}}. \end{aligned}$$

After solving for \tilde{H} and D , the weighted consensus clustering problem transforms into the following optimization:

$$\begin{aligned} \min_w \|\tilde{M} - \tilde{H} \tilde{H}^T\|^2 &= \min_w \text{Tr}(\tilde{M}^2 - 2\tilde{H}^T \tilde{M} \tilde{H} + \tilde{H}^T \tilde{H} \tilde{H}^T \tilde{H}) \\ &= \min_w J = w^T A w - 2b^T w + \text{const.} \\ \text{s.t. } \sum_{i=1}^T w_i &= 1, w_i \geq 0, \end{aligned}$$

where $A_{ij} = \text{Tr}(M(P^i)M(P^j)) = \sum_{kl} M(P^i)_{kl} M(P^j)_{kl}$, $b_i = \tilde{H}^T M(P^i) \tilde{H}$.

Next, a quadratic programming algorithm is employed to iteratively solve for the optimal weights w . Finally, k -means clustering is applied to the weighted consensus matrix to derive the final consensus labels.

Cluster performance evaluation metrics

We employed seven metrics to evaluate the accuracy of the consensus clustering results.

- Adjusted Rand index (ARI) is a similarity metric between two data partitions, often used for measuring clustering accuracy. It has a scale of -1 to 1 , where a higher ARI indicates more precise clustering. ARI is calculated as

$$ARI = \frac{RI - \text{Expected } RI}{\text{Max } RI - \text{Expected } RI},$$

where RI is the original Rand index, $\text{Expected } RI$ is the expected RI under random assignment, and $\text{Max } RI$ is the maximum RI value corresponding to completely correct clustering.

2. Normalized mutual information (NMI) is a measure of mutual dependence between two variables in information theory, defined as

$$NMI(X, Y) = 2 \frac{I(X; Y)}{H(X) + H(Y)},$$

where $I(X; Y)$ is the mutual information between X and Y , and $H(\cdot)$ is the Shannon entropy function.

3. Homogeneity is a measure for evaluating whether clusters contain data spots from a single true category, detailed as

$$h(U, V) = 1 - \frac{H(U|V)}{H(U)},$$

where $H(U|V)$ is conditional entropy of the true category U given clustering V , and $H(U)$ is Shannon entropy of U .

4. Completeness is a measure for measuring whether data spots from a true category are assigned to the same cluster. The equation is

$$c(U, V) = 1 - \frac{H(V|U)}{H(V)}.$$

5. Calinski–Harabasz is an index measuring cluster dispersion ratios when true labels are unavailable. In detail,

$$CH = \frac{B}{W} \times \frac{n - k}{k - 1},$$

where B is the between-cluster variance, W is the within-cluster variance, n is the number of data spots, and k is the number of clusters.

6. Davies–Bouldin is an index evaluating average within-cluster distances against between-cluster distances without true labels. Given clusters C_i and C_j , the Davies–Bouldin index is calculated as follows:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{\Delta(C_i) + \Delta(C_j)}{\delta(C_i, C_j)} \right),$$

where $\Delta(C_i)$ is the intra-cluster distance, $\delta(C_i, C_j)$ is the inter-cluster distance, and k is the number of clusters. A lower Davies–Bouldin index indicates better clustering results.

7. Stability is an indicator for measuring the consistency of clustering results, defined as

$$stability = \frac{1}{n_T} \sum_{i, j=1, i \neq j}^{n_T} ARI(C_i, C_j),$$

where n_T is the number of all possible cluster pairs of T clustering results, and C_i and C_j represent the results of the i th and j th consensus clustering, respectively.

Pseudotime and trajectory analysis

Slingshot was used to perform trajectory inference and pseudotime analysis for the human SCC samples. It takes clustering labels as input and requires specification of the starting cluster to infer tissue-wide pseudotime trajectories. In our implementation, we specified the tumor region as the starting cluster to analyze the progression patterns from tumor to surrounding areas. The detailed workflow consists of the following steps:

1. Preparation of input data, including the clustering labels (clusterLabels) and dimensional reduction coordinates from our consensus clustering results;

2. Trajectory inference by Slingshot, which constructed a minimum spanning tree connecting clusters and fitted simultaneous principal curves to identify continuous trajectories;
3. Pseudotime calculation, which assigned pseudotime values to cells based on their position along the inferred trajectories, with cells in the tumor cluster designated as the starting point; and
4. Visualization using the plot_trajectory function to generate trajectory maps that integrate pseudotime information with spatial coordinates.

Data sets

All data analyzed in this article are available from publicly available data sets. The mouse brain and human breast cancer data sets are collected from 10x Genomics (<https://support.10xgenomics.com/spatial-gene-expression/datasets>). The DLPFC data set (Maynard et al. 2021) is accessible from the spatialLIBD package (<https://github.com/LieberInstitute/HumanPilot>). The ST data (Ståhl et al. 2016) for mouse olfactory bulb tissue is accessible from <https://db.cngb.org/stomics/datasets/STDS0000017>. The STARmap data (Wang et al. 2018) for mouse visual cortex is accessible on https://www.dropbox.com/sh/f7ebheru1lbz91s/AADm6D54GSEF_XB1feRy6OSASa/visual_1020/20180505_BY3_kgenes?dl=0&subfolder_nav_tracking=1. The annotation information for STARmap data and the processed SCANPY object is provided at <https://drive.google.com/drive/folders/111nxheWlc2RXXSdiv24dex3YRaEh780my?usp=sharing>. The processed data and manual annotations for STARmap data (Wang et al. 2018) across three consecutive sections of the mouse cortex are available at GitHub (<https://github.com/zhengli09/BASS-Analysis/tree/master/data>). The ST data for human SCC (Ji et al. 2020) are available under the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) accession number GSE144240.

Software availability

The STCC framework is implemented in Python and has been made available on PyPI (<https://pypi.org/project/STCC/>). The source code for STCC is available at GitHub (<https://github.com/hucongong97/STCC>) and as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported by the National Key R&D Program of China (2024YFC2309600 to X.Z., 2021YFC1712805 to H.-J.W.), National Natural Science Foundation of China (62372286 to X.Z., 32270683 and 32470662 to H.-J.W.), Science and Technology Innovation Plan of Shanghai (23JC1403200 to X.Z.), Beijing Natural Science Foundation (5242006 to H.-J.W.), and Fundamental Research Funds for the Central Universities (BMU2021YJ064 to H.-J.W. and PKU2022LCXQ027–Clinical Medicine Plus X–Young Scholars Project, Peking University). We acknowledge the bioinformatics core in the Center for Single-Cell Omics (CSCOmics), Shanghai Jiao Tong University School of Medicine, and the high-performance computing platform in Peking University for providing bioinformatics and high-performance computing services.

Author contributions: C.H. completed the data collection, conceptualization, and writing of the paper. X.Z. and H.-J.W.

contributed to the conceptualization and writing of the paper. N.W. and J.Y. provided support during the paper revision process. All authors approved the final manuscript.

References

- Abdelaal T, Mourragui S, Mahfouz A, Reinders MJ. 2020. SpaGE: spatial gene enhancement using scRNA-seq. *Nucleic Acids Res* **48**: e107. doi:10.1093/nar/gkaa740
- Andersson A, Bergenstråhle J, Asp M, Bergenstråhle L, Jurek A, Fernández Navarro J, Lundeberg J. 2020. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Communications Biology* **3**: 565. doi:10.1038/s42003-020-01247-y
- Bäckdahl J, Franzén L, Massier L, Li Q, Jalkanen J, Gao H, Andersson A, Bhalla N, Thorell A, Rydén M, et al. 2021. Spatial mapping reveals human adipocyte subpopulations with distinct sensitivities to insulin. *Cell Metab* **33**: 1869–1882.e6. doi:10.1016/j.cmet.2021.07.018
- Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, Tarish F, Tanogljid A, Vickovic S, Larsson L, et al. 2018. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* **9**: 2419. doi:10.1038/s41467-018-04724-5
- Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, Tokcan N, Vanderburg CR, Segerstolpe Å, Zhang M, et al. 2021. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* **18**: 1352–1362. doi:10.1038/s41592-021-01264-7
- Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, et al. 2013. Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* **10**: 1093–1095. doi:10.1038/nmeth.2645
- Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, Atwood SX, Nie Q. 2023. Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods* **20**: 218–228. doi:10.1038/s41592-022-01728-4
- Carlberg K, Korotkova M, Larsson L, Catrina AI, Ståhl PL, Malmström V. 2019. Exploring inflammatory signatures in arthritic joint biopsies with spatial transcriptomics. *Sci Rep* **9**: 18975. doi:10.1038/s41598-019-55441-y
- Chen W-T, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, Corthout N, Qian X, Laláková J, Kühnemund M, Voytyuk I, et al. 2020. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**: 976–991.e19. doi:10.1016/j.cell.2020.06.038
- Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, Qiu X, Yang J, Xu J, Hao S, et al. 2022. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**: 1777–1792.e21. doi:10.1016/j.cell.2022.04.003
- Chidester B, Zhou T, Alam S, Ma J. 2023. SPICEMIX enables integrative single-cell spatial modeling of cell identity. *Nat Genet* **55**: 78–88. doi:10.1038/s41588-022-01256-z
- Cui Y, Zhang S, Liang Y, Wang X, Ferraro TN, Chen Y. 2021. Consensus clustering of single-cell RNA-seq data by enhancing network affinity. *Brief Bioinformatics* **22**: bbab236. doi:10.1093/bib/bbab236
- Dong R, Yuan G-C. 2021. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol* **22**: 145. doi:10.1186/s13059-021-02362-7
- Dong K, Zhang S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun* **13**: 1739. doi:10.1038/s41467-022-29439-6
- Dries R, Chen J, Del Rossi N, Khan MM, Sisti A, Yuan G-C. 2021. Advances in spatial transcriptomic data analysis. *Genome Res* **31**: 1706–1718. doi:10.1101/gr.275224.121
- Elosua-Bayes M, Nieto P, Mereu E, Gut I, Heyn H. 2021. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res* **49**: e50. doi:10.1093/nar/gkab043
- Gan Y, Li N, Zou G, Xin Y, Guan J. 2018. Identification of cancer subtypes from single-cell RNA-seq data using a consensus clustering method. *BMC Med Genomics* **11**: 65–72. doi:10.1186/s12920-018-0433-z
- Hore P, Hall LO, Goldfob DB. 2009. A scalable framework for clustering ensembles. *Pattern Recognit* **42**: 676–688. doi:10.1016/j.patcog.2008.09.027
- Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, Lee EB, Shinohara RT, Li M. 2021. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* **18**: 1342–1351. doi:10.1038/s41592-021-01255-8
- Hu Y, Xie M, Li Y, Rao M, Shen W, Luo C, Qin H, Baek J, Zhou XM. 2024. Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. *Genome Biol* **25**: 212. doi:10.1186/s13059-024-03361-0
- Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, Guo MG, George BM, Mollbrink A, Bergenstråhle J, et al. 2020. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**: 497–514.e22. doi:10.1016/j.cell.2020.08.043
- Jia E, Sheng Y, Shi H, Wang Y, Zhou Y, Liu Z, Qi T, Pan M, Bai Y, Zhao X, et al. 2023. Spatial transcriptome profiling of mouse hippocampal single cell microzone in Parkinson's disease. *Int J Mol Sci* **24**: 1810. doi:10.3390/ijms24031810
- Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, et al. 2017. SC3: consensus clustering of single-cell RNA-seq data. *Nat Methods* **14**: 483–486. doi:10.1038/nmeth.4236
- Li Z, Zhou X. 2022. BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies. *Genome Biol* **23**: 168. doi:10.1186/s13059-022-02734-7
- Li H, Ma T, Hao M, Guo W, Gu J, Zhang X, Wei L. 2023. Decoding functional cell–cell communication events by multi-view graph learning on spatial transcriptomics. *Brief Bioinformatics* **24**: bbad359. doi:10.1093/bib/bbad359
- Liu W, Liao X, Yang Y, Lin H, Yeong J, Zhou X, Shi X, Liu J. 2022. Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data. *Nucleic Acids Res* **50**: e72–e72. doi:10.1093/nar/gkac219
- Lopez R, Nazaret A, Langevin M, Samaran J, Regier J, Jordan MI, Yosef N. 2019. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. arXiv:1905.02269 [cs.LG]. doi:10.48550/arXiv.1905.02269
- Maynard KR, Collado-Torres L, Weber LM, Uyttingco C, Barry BK, Williams SR, Cattalini JL, Tran MN, Besich Z, Tippani M, et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**: 425–436. doi:10.1038/s41593-020-00787-0
- Moffitt JR, Bambach-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, Rubinstein ND, Hao J, Regev A, Dulac C, et al. 2018. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**: eaau5324. doi:10.1126/science.aau5324
- Mölbert C, Haghverdi L. 2023. Adjustments to the reference dataset design improve cell type label transfer. *Front Bioinform* **3**: 1150099. doi:10.3389/fbinf.2023.1150099
- Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, Rybakov S, Ibarra IL, Holmberg O, Virshup I, et al. 2022. Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* **19**: 171–178. doi:10.1038/s41592-021-01358-2
- Peng L, Wang F, Wang Z, Tan J, Huang L, Tian X, Liu G, Zhou L. 2022. Cell–cell communication inference and analysis in the tumour microenvironments from single-cell transcriptomics: data resources and computational strategies. *Brief Bioinformatics* **23**: bbac234. doi:10.1093/bib/bbac234
- Pham D, Tan X, Balderson B, Xu J, Grice LF, Yoon S, Willis EF, Tran M, Lam PY, Raghobar A, et al. 2023. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nat Commun* **14**: 7739. doi:10.1038/s41467-023-43120-6
- Raredon MSB, Yang J, Kothapalli N, Lewis W, Kaminski N, Niklason LE, Kluger Y. 2023. Comprehensive visualization of cell–cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* **39**: btac775. doi:10.1093/bioinformatics/btac775
- Sato A, Sekine Y, Saruta C, Nishibe H, Morita N, Sato Y, Sadakata T, Shinoda Y, Kojima T, Furuichi T. 2008. Cerebellar development transcriptome database (CDT-DB): profiling of spatio-temporal gene expression during the postnatal development of mouse cerebellum. *Neural Netw* **21**: 1056–1069. doi:10.1016/j.neunet.2008.05.004
- Shang L, Zhou X. 2022. Spatially aware dimension reduction for spatial transcriptomics. *Nat Commun* **13**: 7203. doi:10.1038/s41467-022-34879-1
- Shao X, Li C, Yang H, Lu X, Liao J, Qian J, Wang K, Cheng J, Yang P, Chen H, et al. 2022. Knowledge-graph-based cell–cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nat Commun* **13**: 4429. doi:10.1038/s41467-022-32111-8
- Shen X, Zuo L, Ye Z, Yuan Z, Huang K, Li Z, Yu Q, Zou X, Wei X, Xu P, et al. 2025. Inferring cell trajectories of spatial transcriptomics via optimal transport analysis. *Cell Syst* **16**: 101194. doi:10.1016/j.cels.2025.101194
- Song Q, Su J. 2021. DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief Bioinformatics* **22**: bbaa414. doi:10.1093/bib/bbaa414
- Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**: 78–82. doi:10.1126/science.aaf2403
- Theocharidis G, Thomas BE, Sarkar D, Mumme HL, Pilcher WJ, Dwivedi B, Sandoval-Schaefer T, Sirbulescu RF, Kafanas A, Mezghani I, et al. 2022.

- Single cell transcriptomic landscape of diabetic foot ulcers. *Nat Commun* **13**: 181. doi:10.1038/s41467-021-27801-8
- Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, Åijö T, Bonneau R, Bergensträhle L, Navarro JF, et al. 2019. High-definition spatial transcriptomics for in situ tissue profiling. *Nat Methods* **16**: 987–990. doi:10.1038/s41592-019-0548-y
- Vickovic S, Schapiro D, Carlberg K, Lötstedt B, Larsson L, Hildebrandt F, Korotkova M, Hensvold AH, Catrina AI, Sorger PK, et al. 2022. Three-dimensional spatial transcriptomics uncovers cell type localizations in the human rheumatoid arthritis synovium. *Communications Biology* **5**: 129. doi:10.1038/s42003-022-03050-3
- Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, et al. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**: eaat5691. doi:10.1126/science.aat5691
- Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. 2019. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**: 1873–1887.e17. doi:10.1016/j.cell.2019.05.006
- Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**: 15. doi:10.1186/s13059-017-1381-1
- Xu H, Fu H, Long Y, Ang KS, Sethi R, Chong K, Li M, Uddamvathanak R, Lee HK, Ling J, et al. 2024a. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med* **16**: 12. doi:10.1186/s13073-024-01283-x
- Xu Z, Wang W, Yang T, Li L, Ma X, Chen J, Wang J, Huang Y, Gould J, Lu H, et al. 2024b. STOmicsDB: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic Acids Res* **52**: D1053–D1061. doi:10.1093/nar/gkad933
- Yang Y, Huh R, Culpepper HW, Lin Y, Love MI, Li Y. 2019. SAFE-clustering: single-cell aggregated (from ensemble) clustering for single-cell RNA-seq data. *Bioinformatics* **35**: 1269–1277. doi:10.1093/bioinformatics/bty793
- Yang Y, Shi X, Liu W, Zhou Q, Chan Lau M, Chun Tatt Lim J, Sun L, Ng CCY, Yeong J, Liu J. 2022. SC-MEB: spatial clustering with hidden Markov random field using empirical Bayes. *Brief Bioinformatics* **23**: bbab466. doi:10.1093/bib/bbab466
- Yuan Z, Zhao F, Lin S, Zhao Y, Yao J, Cui Y, Zhang X-Y, Zhao Y. 2024. Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nat Methods* **21**: 712–722. doi:10.1038/s41592-024-02215-8
- Zhang L, Mao S, Yao M, Chao N, Yang Y, Ni Y, Song T, Liu Z, Yang Y, Li W. 2021. Spatial transcriptome sequencing revealed spatial trajectory in the non-small cell lung carcinoma. bioRxiv doi:10.1101/2021.04.26.441394
- Zhang Y, Miller JA, Park J, Lelieveldt BP, Long B, Abdelaal T, Aebermann BD, Biancalani T, Comiter C, Dzyubachyk O, et al. 2023. Reference-based cell type matching of in situ image-based spatial transcriptomics data on primary visual cortex of mouse brain. *Sci Rep* **13**: 9567. doi:10.1038/s41598-023-36638-8
- Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, Williams SR, Uyttingco CR, Taylor SE, Nghiem P, et al. 2021. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* **39**: 1375–1384. doi:10.1038/s41587-021-00935-2

Received September 16, 2024; accepted in revised form March 31, 2025.