

Inference of cell-type composition and single-cell spatial maps from spatial transcriptomics data with SWOT

Corresponding Author: Professor Lin Gao

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Wang et al. developed SWOT, a computational method for inferring cell type proportions in low-resolution spatial transcriptomics data and identifying spatial locations for each cell. The method demonstrates comparable performance to state-of-the-art techniques and offers multiple functionalities to support future research. However, I have a few questions regarding the methodology that I hope the authors can address before I evaluate the quality of the work.

Major:

1. The distance-dependent expression similarity. In Supplementary Figures S1-2, the authors assessed expression similarity based on physical distance. They interpret this as resulting from cell-cell communications. However, it could also stem from molecular diffusion (sometimes referred to as “bleeding”) in spot-level assays or from segmentation errors in single-molecule-based assays. How can the authors differentiate these effects to demonstrate that the similarity between cells goes beyond technical artifacts?
2. Simulation setting. The authors picked two datasets for simulation. One is a cortex seqFISH+ dataset, and the other is a stereo-seq brain dataset. The stereo-seq technology does not directly capture each single-cell (although the capturing area is small). Although heuristic methods can be used to infer cell types, I would be skeptical about the accuracy of cell annotation. Can the authors provide evidence that the ground truth is reliable? (For example, the authors can show purity of cell type markers of identified cells or other metrics). It is important to determine whether the annotations can be viewed as ground truth.
3. Some hyperparameters. The method involves several hyperparameters, yet I noticed limited discussion regarding them. These include λ , α , and ϵ in equation (5). Their values will impact the performance of the method, so the authors should clarify how they are determined and evaluate the robustness of the results concerning different hyperparameter values.
4. How are a and b determined in equation (5)? It does not seem reasonable because this assumes that we know the number of cells in each spot and how many times a particular cell should be used to map to spatial transcriptomics data. How are they utilized in the simulation data? Are there potential issues with information leakage?
5. Optimization method. Equation (5) only informs us about the objective function. Which optimization method did the authors use to estimate the optimizer? This is not addressed in the methods section. It also seems that calculating the third term (structured term) requires summation over $n^2 * m^2$ terms, which could be time-consuming.
6. The equation (6) seems odd. There shouldn't be a summation over spot index j in the numerator.
7. The estimation of cell number per spot lacks detail and justification. The relationship between the number of cells and the non-zero proportion must be established in real or simulated data, and the exact formula needs to be specified.

Reviewer #2

(Remarks to the Author)

The article introduces SWOT, a spatially weighted optimal transport method that deconvolutes cell-type composition in spot level spatial transcriptomics data and maps individual cells to spatial locations. A key novelty of SWOT is its approach integrating gene expression into the spot-by-spot distance matrix. While the topic is of general interest, the construction of spatially weighted distance matrix is ad hoc in nature but is novel and allows SWOT to incorporate both spatial closeness and gene expression similarity. The authors validate SWOT on simulated data and three real datasets, demonstrating its ability to reconstruct layered tissue architectures, infer single-cell coordinates, and facilitate functional annotation of tissue

neighborhoods. I have the following main comments:

1. The method involves critical hyperparameters (e.g., bandwidth b , k -nearest neighbors k , regularization terms λ , α , ε) whose values may significantly influence the results but lack empirical justification. The manuscript does not clearly justify how these values are chosen or how sensitive the results are to these choices. Moreover, the choice of a factor of 10 to weight gene expression distances (ij) in Eq. 3 and Eq. 4 also appears arbitrary. The authors should consider demonstrating more systematic strategies (e.g., cross-validation, grid search) for tuning these hyperparameters or at least provide guidelines for default settings tested on multiple datasets.
2. A key novelty of SWOT is the modification of the spot-to-spot distance to incorporate spatial coordinates and gene expression similarities. The approach is ad hoc and may oversimplify the tissue structures. It needs sensitivity analysis to verify the robustness of this method.
3. The manuscript simulated data based on single-cell resolution ST, but the Methods section does not clearly describe how the spots were generated, or how realistic gene expression characteristics and dropout events were simulated, or whether variability across cells was effectively reproduced. More details on simulation pipelines are needed.
4. The authors claim SWOT estimates the cell numbers per spot, but the estimation procedure is missing in the manuscript. Based on the GitHub repository (https://github.com/GaoLabXDU/SWOT/blob/main/SWOT/utis/cell_mapping.py at line 245), it seems to be technology specific and limited to 10X Visium, Spatial Transcriptomics, and Slide-seq. It compares each spot's zero expression proportion to a dataset-wide average, when a spot's zero ratio exceeds the average, it is assigned with a small number of cells; otherwise it will be assigned with a larger number of cells. It essentially assumes dropouts inversely correlate with the number of cells captured in a spot. However, according to Fig. S3, the inverse trend doesn't always hold, this assumption should be validated on more datasets with varying dropout rates.

Minor comments:

1. $\mathbb{R}$ symbols in the Methods section are not readable.
2. The uniform sampling of cell coordinates within spot boundaries (SWOT/SWOT/utis/cell_mapping.py at line 309) may unintentionally bias cells toward spot centers due to radial density effects.
3. The analysis of tissue cellular neighborhoods (TCNs) in Fig. 6G is not well interpreted. The authors could consider integrating cell-cell interaction methods to correlate enriched cell types with known signaling pathways or ligand-receptor pairs to enhance biological interpretability.
4. The reliance on Dijkstra's algorithm for kNN graph construction (S1.5) and the iterative optimal transport solver may have challenges in scalability for large-scale datasets. An analysis of SWOT's runtime against competing methods is helpful to evaluate its efficiency and practical utility.

Reviewer #4

(Remarks to the Author)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I thank the authors for their efforts to address my previous concerns. I only have two minor comments on the article.

1. It seems to me that Fig. 2f-g shows a systematic difference in cell type proportion (it tends to overestimate the oligodendrocyte proportion), but the article does not comment on the difference. I am curious about the reason for this bias.
2. Several figures compare the ratio between the estimated value and the ground truth. I think it makes sense to put the ratio on the log scale so that it is fairer to compare how far the ratio is from 1 (both higher and lower directions). It also helps with overcrowding. For example, it is hard to compare the performance in Figure S14a, although the authors have provided the average values in the caption.

Reviewer #2

(Remarks to the Author)

The authors have fully addressed my comments, the manuscript is much improved.

Reviewer #4

(Remarks to the Author)

The authors thoroughly addressed my previous concerns regarding their manuscript on SWOT. The revisions have substantially strengthened the manuscript, addressing all my previous concerns. The authors have improved the mathematical notation, addressed the potential bias in cell coordinate sampling, enhanced the interpretation of tissue cellular neighborhoods with cell-cell interaction analyses, and provided comprehensive runtime analyses to demonstrate the method's scalability.

The authors have significantly enhanced the paper by:

1. Providing a comprehensive analysis and justification of hyperparameter selection, including bandwidth settings, K nearest neighbors values, clustering methods, and regularization terms. New sensitivity analyses across multiple datasets demonstrate the robustness of its methods and provide clear guidance to users.
2. Through comparative analysis with alternative approaches that use only spatial coordinates or gene expression, the authors elucidate the rationale behind their spatial weighting strategy and demonstrate its necessity.
3. Expanding the description of their simulation methodology, clearly explaining how spot-resolution data were generated from single-cell resolution spatial transcriptomics data while preserving realistic gene expression characteristics, dropout events, and intercellular variability.
4. The authors thoroughly explain their cell number estimation procedure and validate it on datasets with different dropout rates. Although I still have some doubts about the method itself, the authors use sufficient validation and cell segmentation on H&E images as an independent validation method to prove the reliability of their method.
5. The authors demonstrate that SWOT remains computationally practical. With GPU-accelerated distance-matrix routines available, and runtime/memory benchmarks on large simulations show performance on par with or better than competing methods.

The manuscript now presents SWOT as a well-justified and robust method for inferring both cell-type composition and single-cell spatial maps from spot-based spatial transcriptomics data. I have no remaining concerns or revision comments.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I thank the authors for addressing my previous comments.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Reviewers' Comments

Dear Editors and Reviewers:

Thanks so much for your time and efforts on our manuscript. Also, we thank the reviewers for valuable comments and suggestions. These comments are valuable for revising and improving our manuscript. We carefully revised our manuscript according to the reviewers' suggestions and comments. The revised parts are highlighted in blue in the revised manuscript. The followings are our point-by-point responses. Also please let me know if there are remaining issues of concern. Thank you for your consideration.

Responses to Reviewer #1's comments

Comment 1.

The distance-dependent expression similarity. In Supplementary Figures S1-2, the authors assessed expression similarity based on physical distance. They interpret this as resulting from cell-cell communications. However, it could also stem from molecular diffusion (sometimes referred to as “bleeding”) in spot-level assays or from segmentation errors in single-molecule-based assays. How can the authors differentiate these effects to demonstrate that the similarity between cells goes beyond technical artifacts?

Response: Thanks for your insightful comments and questions, we answered your concerns from two aspects as follows.

(1) Upon further review of the existing literatures, we realized that interpreting distance-dependent expression similarity as cell-cell communications, namely, the notion that “*cell-cell communication leads to similar gene expression patterns in spatially adjacent cells*”, is inappropriate. We have revised the description in the Introduction section.

Existing cell-type deconvolution methods can be categorized based on whether they incorporate spatial location information. Most methods, such as SPOTlight¹⁵, RCTD¹⁶, STRIDE¹⁷, Stereoscope¹⁸, and Uniport¹⁹, do not leverage spatial information. However, previous studies^{20–22} have reported that similar cell types tend to co-localize spatially, which is a phenomenon of spatial autocorrelation, and can also be observed in real ST data (Figures S1 and S2). Therefore, neighboring locations (spots) are more likely to share similar cell-type compositions, highlighting the importance of incorporating spatial location information when estimating cell-type composition. Methods that do account for spatial information, such as CARD²² and SONAR²³, characterize spatial dependency/weight between spots based on spatial distances, which diminishes as distance increases.

(2) In addition, we have added experiments and validations regarding the statements “*neighboring locations (spots) are more likely to share similar cell-type compositions*” and “*the*

similarity between cells goes beyond technical artifacts”.

According to the existing studies¹⁻³, we conducted the following validations. Performed analyses were based on simulated and real datasets used in this study: Simulated_MBaging, MOB (Mouse Olfactory Bulb), Mouse cerebellum, and PDAC (Pancreatic Ductal Adenocarcinoma) datasets. The Simulated_MBaging dataset is newly added⁴, we designed seven grids (250, 200, 150, 130, 75, 58, 38) and generated seven corresponding datasets with different numbers of spots (290, 426, 749, 989, 2850, 4669, 9554) to support scalability evaluation. For consistency, we refer to them as: Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000, respectively. MERFISH is the raw single-cell ST data for constructing spot-resolution simulated data, which is generated by MERFISH technology. Among all simulations, we selected Simulated_MBaging_300, Simulated_MBaging_3000, and Simulated_MBaging_10000 for validation. Cell-type marker genes used in the following evaluations were obtained from previously reported literatures.

(2.1) We have conducted experiments to validate the description of “*neighboring locations are more likely to share similar cell-type compositions*” in the Discussion section of the revised manuscript.

Unlike most cell-type deconvolution methods that ignore spatial information, SWOT effectively utilizes spatial information. This approach is motivated by a phenomenon named spatial autocorrelation²⁰, where cell-type composition between spatially proximate spots tends to be more similar than those between distant spots. As validations of spatial autocorrelation in our used ST data, we found that a majority of marker genes exhibited statistically significant spatial autocorrelation, with 68.97% and 75.86% identified by Moran's I and Geary's C tests, respectively (adjusted p-value<0.05, Table S3). Semi-variance analysis of marker genes revealed a general increase with distance (Figure S27), and we observed a strong correlation in expression between each spot and its nearest neighbors (Figure S28). These findings reflected a spatial aggregation of gene expression caused by spatial autocorrelation.

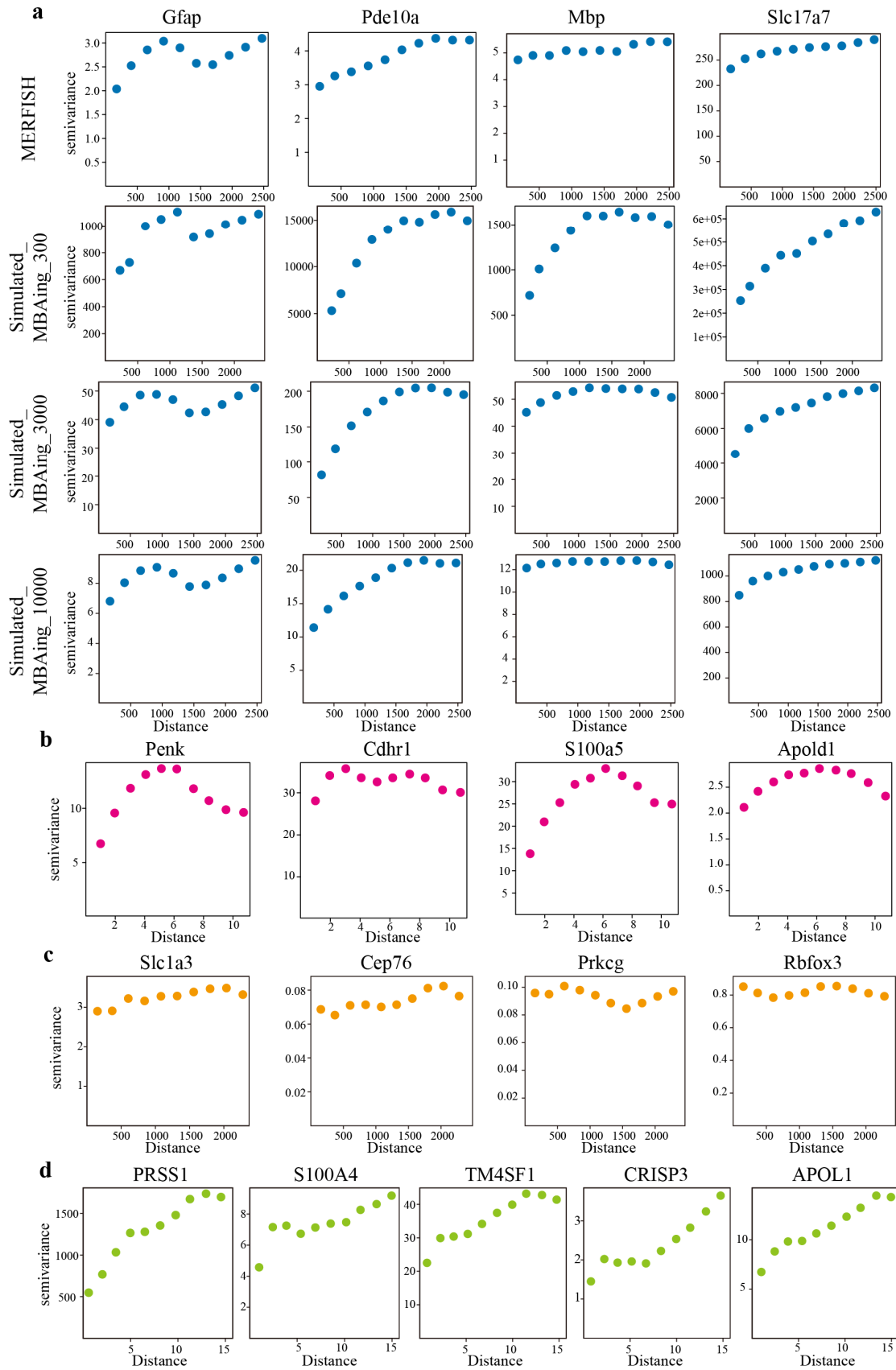
Table S3. Spatial autocorrelation analysis (Moran's I and Geary's C)

We performed spatial autocorrelation analysis using Moran's I and Geary's C metrics, which measure the dependence or correlation of a marker gene between different locations, on Simulated_MBaging (MERFISH, Simulated_MBaging_300, Simulated_MBaging_3000, and Simulated_MBaging_10000), MOB, mouse cerebellum, and PDAC datasets. In the Simulated_MBaging dataset, MERFISH is the raw single-cell ST data for constructing spot-resolution simulated data, which is generated by MERFISH technology. Statistical significance was assessed using one-sided Moran's I and one-sided Geary's C test, with p-values adjusted by Bonferroni method.

Dataset		Gene	Moran's I			Geary's C		
			Moran's I	p-value	Adjusted p-value	Geary's C	p-value	Adjusted p-value
Simulated_MBaging	MERFISH	Gfap	0.0856	1.0201e-01	8.1254e-01	0.8746	9.2450e-54	1.8490e-53
		Pde10a	0.3969	9.1573e-01	1.0000e+00	0.6155	5.4673e-01	8.4712e-01
		Mbp	0.0189	7.8675e-168	1.5735e-167	0.9791	5.2595e-21	1.0519e-20
		Slc17a7	0.1851	1.2577e-02	5.4271e-01	0.8701	8.8456e-01	1.0000e+00
	300	Gfap	0.1965	1.2026e-05	2.4053e-05	0.6813	3.1192e-05	6.2384e-05
		Pde10a	0.6807	6.7064e-45	1.3412e-44	0.3521	5.8558e-23	1.1711e-22

		Mbp	0.4365	6.3498e-19	1.2699e-18	0.5248	3.0328e-19	6.0657e-19
		Slc17a7	0.5904	9.0485e-33	1.8097e-32	0.4305	3.6250e-28	7.2501e-28
	3000	Gfap	0.1176	1.5869e-139	3.1739e-139	0.8417	4.8627e-21	9.7254e-21
		Pde10a	0.5967	5.1547e-01	9.1378e-01	0.4201	7.2478e-02	3.1247e-01
		Mbp	0.1322	1.8146e-156	3.6292e-156	0.8696	1.1867e-108	2.3734e-108
		Slc17a7	0.4544	1.0463e-02	6.7485e-02	0.5489	1.5687e-01	2.0254e-01
	10000	Gfap	0.0746	9.1547e-02	1.0287e-01	0.8701	5.4766e-41	1.0953e-40
		Pde10a	0.4736	2.3239e-01	4.5912e-01	0.5388	5.1278e-01	9.0145e-01
		Mbp	0.0432	4.0947e-185	8.1894e-185	0.9568	1.1408e-63	2.2816e-63
		Slc17a7	0.2545	1.5974e-01	2.2146e-01	0.7482	2.9512e-01	7.1545e-01
MOB		Penk	0.4325	1.8556e-23	3.7113e-23	0.5676	1.8740e-21	3.7481e-21
		Cdhr1	0.1377	6.4120e-04	1.2824e-03	0.8612	1.3789e-03	2.7579e-03
		S100a5	0.5169	1.0565e-32	2.1131e-32	0.4829	1.8564e-29	3.7128e-29
		Apold1	0.1360	7.3937e-04	1.4787e-03	0.8540	7.9663e-04	1.5932e-03
Mouse cerebellum		Slc1a3	0.0879	1.157e-01	3.3322e-01	0.9149	2.5301e-55	5.0603e-55
		Cep76	0.0167	2.8046e-24	5.6093e-24	0.9915	9.5521e-02	1.9104e-01
		Prkeg	0.0266	2.2773e-58	4.5547e-58	0.9754	1.1171e-04	2.2342e-04
		Rbfox3	0.0523	5.2719e-218	1.0543e-217	0.9678	4.2215e-20	8.4431e-20
PDAC		APOL1	0.4251	8.3302e-61	1.6661e-60	0.5583	1.0543e-47	2.1087e-47
		TM4SF1	0.4286	8.0657e-62	1.6132e-61	0.5733	2.2839e-44	4.5678e-44
		PRSS1	0.5818	7.2281e-120	1.4456e-119	0.4071	1.5362e-58	3.0723e-58
		S100A4	0.4247	5.0099e-64	1.0019e-63	0.5724	1.0011e-33	2.0022e-33
		CRISP3	0.3402	1.0759e-42	2.1519e-42	0.5877	3.3702e-29	6.7405e-29

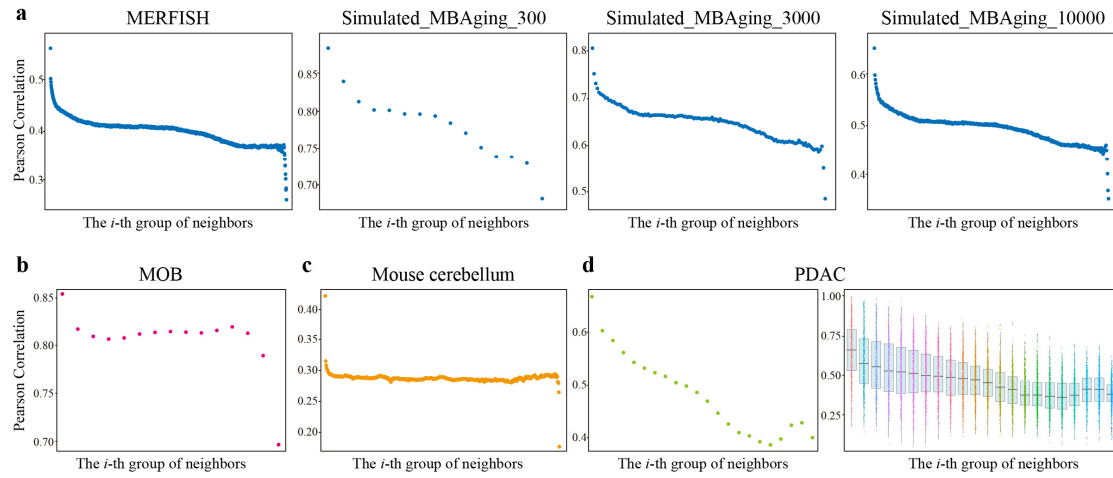
Figure S27. Semi-variograms analysis for marker genes



For (a) Simulated_MBAging (MERFISH, Simulated_MBAging_300, Simulated_MBAging_3000, and Simulated_MBAging_10000), (b) MOB, (c) mouse cerebellum, and (d) PDAC datasets, we analyzed semi-variograms to assess the variability and spatial dependence of marker genes within neighborhoods. In Simulated_MBAging dataset, MERFISH is the raw single-cell ST data, generated by MERFISH technology, for constructing spot-resolution simulated data. Marker

genes were obtained from the previously reported literatures for each dataset. The semi-variance generally increased with spatial distance, reflecting clear neighborhood spatial correlation patterns.

Figure S28. Correlation between gene expression of each spot and its neighbors

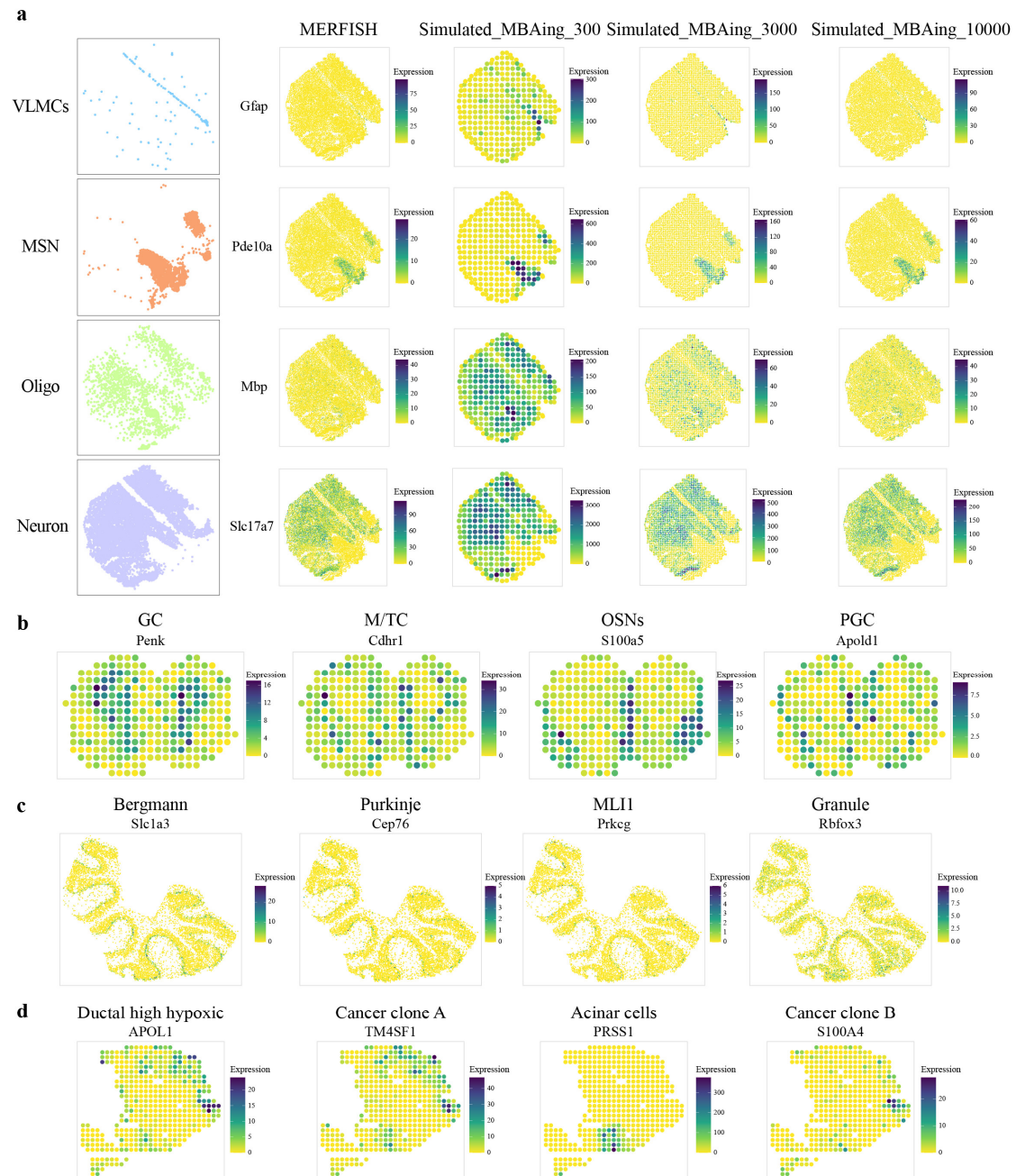


We computed the Pearson correlation coefficient between the gene expression of each spot and the average gene expression of its i -th group of neighbors in (a) Simulated_MBaging (MERFISH, Simulated_MBaging_300, Simulated_MBaging_3000, and Simulated_MBaging_10000), (b) MOB, (c) mouse cerebellum, and (d) PDAC datasets. The 1-st group of neighbors represents the nearest 20 neighbors of each spot, the 2-nd group represents the nearest neighbors from 21 to 40, and so on, and the last group represents the farthest spatial positions examined for each spot. The correlation in expression between spots decreases as distance increases. In (d) for PDAC dataset, we additionally visualized the distribution of correlation across spots in each group, each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

(2.2) To verify the similarity between spots goes beyond technical artifacts, we have added descriptions in the Discussion section of the revised manuscript.

We further distinguished the similarity between adjacent spots from potential technical artifacts, such as molecular diffusion or segmentation errors (Figure S29). Most expression patterns exhibited well-defined boundaries and cell-type-related spatial distributions, supporting the conclusion that the observed spatial patterns are biologically driven rather than artifactually induced. Collectively, these results underscored that neighboring locations tend to have similar cell-type compositions, and emphasized the necessity of incorporating spatial location information, as utilized in SWOT, when estimating cell-type compositions from spot-resolution data.

Figure S29. Spatial distribution of expression patterns for marker genes



Spatial distribution of expression density for marker genes of specific cell types in (a) Simulated_MBAging (MERFISH, Simulated_MBAging_300, Simulated_MBAging_3000, and Simulated_MBAging_10000), (b) MOB, (c) mouse cerebellum, and (d) PDAC datasets. Left panels in (a) displayed the spatial distribution of the four cell types in Simulated_MBAging dataset for reference. Color in each plot denotes the expression level of the selected marker gene. These markers were obtained from previously reported literatures. Most gene expressions exhibited clear boundaries and cell-type-related spatial distribution patterns. Only a few genes displayed diffuse or ambiguous boundaries, indicating minimal impact from technical artifacts such as molecular diffusion or segmentation errors.

Collectively, these experiments supported that neighboring locations are more likely to share similar cell-type compositions, and this similarity goes beyond technical artifacts. These phenomena emphasized the importance of spatial location information in ST data analysis.

Comment 2.

Simulation setting. The authors picked two datasets for simulation. One is a cortex seqFISH+ dataset, and the other is a stereo-seq brain dataset. The stereo-seq technology does not directly capture each single-cell (although the capturing area is small). Although heuristic methods can be used to infer cell types, I would be skeptical about the accuracy of cell annotation. Can the authors provide evidence that the ground truth is reliable? (For example, the authors can show purity of cell type markers of identified cells or other metrics). It is important to determine whether the annotations can be viewed as ground truth.

Response: We appreciate your insightful comments, and the reliability of the ground-truth is fundamental to this work. We have provided a detailed explanation about the simulated datasets, and verifications for the reliability of cell type annotation.

(1) To clarify the datasets picked for simulation, we have explained the construction process and the source of datasets and annotations as follows, and further provided detailed descriptions in the revised manuscript.

The idea of constructing simulated data in our manuscript is inspired by previous studies^{5–11}. These studies utilized existing single-cell ST data as a baseline to simulate spot-resolution ST data by establishing a spatial grid division over the spatial location of individual cells.

The datasets and cell type annotations are derived from the original publications. For Simulated_Cortex dataset, both the single-cell resolution SeqFISH+ data used for simulation and the scRNA-seq data used for reference were came from previously published studies^{6,12,13}. Single-cell ST data used the enriched genes to identify clusters, and annotated cell types based on existing marker genes¹². The cell type annotation of scRNA-seq data was obtained from another study⁶. For Simulated_MB dataset, we started from <http://116.6.21.110:8090/share/3ca60300-e8bd-4d67-8b68-14267e428523>, and selected the “/Stereopy_Demo_Data/Demo_SingleR/Reference/Mouse_brain_ref.anndata075.h5ad” as reference. We used “/Stereopy_Demo_Data/Demo_MouseBrain/SS200000135TL_D1.tissue.gef” for simulation and followed the official tutorial (<https://stereopy.readthedocs.io/en/latest/Tutorials/SingleR.html>), the cell type information of the Stereo-seq data was obtained from [14–16]. Notably, the Stereo-seq technology does not directly capture each single-cell, but Chen, A. et al.¹⁷ referred that the Stereo-seq technology can combine single-cell resolution, high sensitivity, and a large field of view to generate spatially resolved transcriptomic atlas. Nevertheless, we acknowledge that using non-single cell ST data for simulation may be inappropriate. Therefore, we present additional validations to assess the reliable of ground-truth annotations.

In the revised main text, a detailed description has been added in the “Dataset description” subsection of the Methods section.

Dataset description. Following the existing simulated data generation strategy^{36,59–61}, we constructed simulated datasets by employing single-cell resolution ST data. According to the spatial coordinates of single-cell ST data, we conducted a spatial grid partitioning to generate spot-resolution ST data. This process enables the reservation of critical biological characteristics observed in real

single-cell ST datasets, such as the realistic spatial structures, gene expression characteristics, sequencing depth, transcript detection rates, dropout events, or intercellular variability. The topology structure and spatial distribution of the original data determine the number of simulated spots, and the size of grid division determines the average number of cells per spot. For Simulated_Cortex dataset, a 200×200 grid generated 85 spots, and each containing an average of six cells, which is in line with the requirements of commonly used sequencing techniques. For Simulated_MBaging datasets, we designed seven grids (250, 200, 150, 130, 75, 58, 38) and generated seven corresponding datasets with different number of spots (290, 426, 749, 989, 2850, 4669, 9554) to support scalability evaluation. For each simulated spot, the spatial coordinates and gene expression were determined by the average coordinates and the summed expressions of all cells within that spot, respectively. The cell-type compositions, cell numbers per spot, and spatial coordinates per cell were established as the ground-truth.

In the revised main text, we have also updated Figure 2(a) to illustrate the generation of simulated data in the Results section.

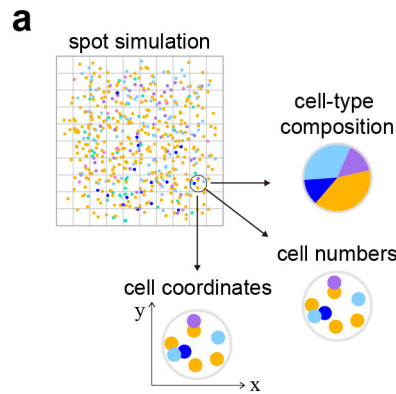


Figure 2. Performance evaluation for cell-type composition inference on simulated datasets. **a** Spot-resolution simulated ST data generation. The cell-type composition, cell numbers, cell coordinates are established as the ground-truth. Circles denote spots, points denote cells, and colored by cell types.

In the revised Supplementary text, the description in the “2.1 Simulated Datasets” subsection of the Supplementary Methods section has been revised accordingly.

2.1 Simulated datasets

For Simulated_Cortex dataset, we selected the cortex region from the sequenced mouse brain tissue by seqFISH+ technology. The single-cell resolution seqFISH+ data consisted of 523 cells expressing across 10,000 genes, and cells were annotated into six cell types based on canonical marker genes: iNeuron (inhibitory neurons), Microglia (microglia cells), endoMural (endothelial-mural cells), eNeuron (excitatory neurons), Astrocytes, and Olig (Oligodendrocytes)¹⁴. We generated a simulated spot-resolution ST dataset by applying a 200 spatial grid, which yielded 85 spots with an average of six cells per spot.

For the Simulated_MBaging datasets, we selected a snRNA-seq data from 4-week-old stage mouse brain aging tissue¹⁵ as a reference, and a single-cell resolution MERFISH data to generate simulated spot-resolution ST datasets. We filtered the MERFISH data based on the following conditions: 'donor_id=MsBrainAgingSpatialDonor_1', 'development_stage=4-week-old stage', and

'slice=0', resulting in 21,002 cells expressed on 347 genes, and labeled for 12 cell types: Astro (Astrocytes), InInter (inhibitory interneuron), Micro (microglia), Oligo (oligodendrocytes), Neuron, Endo (endothelial cells), MSN (Medium spiny neurons), OPCs (oligodendrocyte precursor cells), Peri (pericytes), Macro (macrophages), VLMCs (vascular leptomeningeal cells), and T cells. To evaluate the performance under varying spatial resolutions, we applied seven grid sizes (250, 200, 150, 130, 75, 58, 38) to generate seven Simulated_MBaging datasets with different numbers of spots (290, 426, 749, 989, 2850, 4669, 9554). For consistency, we refer to them as: Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000, respectively. The cell types in the snRNA-seq and MERFISH data were annotated based on the previously known cell-type marker genes that define major neuronal, glial, and immune cell types and genes previously reported to be highly upregulated in various cell types during aging¹⁵.

(2) Regarding the validations of the reliability in cell type annotation, we first provided descriptions from the original literatures, and further conducted a unified experiment to assess whether the annotations can be viewed as ground-truth.

(2.1) We elaborated on the efforts made in the original studies.

For single-cell ST data in Simulated_Cortex dataset, the original article¹² conducted the Pearson correlation analysis to demonstrate the unsupervised clustering of seqFISH+ correlates well with scRNA-seq data. It further illustrated the supervised mapping of seqFISH+-analyzed cortex cell clusters aligns well with those identified from scRNA-seq clusters.

For single-cell ST data in Simulated_MBaging dataset, the original article⁴ performed an integrated clustering analysis across MERFISH and snRNA-seq data, and showed good correspondence between the clustering of cell types measured by the MERFISH with the snRNA-seq. It also demonstrated that the cell type annotations in the MERFISH dataset are largely consistent with those from snRNA-seq data, with high correspondence in gene expression between the clusters from each dataset, further supporting the reliability of these annotations.

(2.2) We further assessed the purity of clusters.

The purity of clusters for each dataset can be validated using ROGUE metric¹⁸. It enables accurate and unbiased assessment of cluster purity, and provides a measure to evaluate the quality of the published and newly generated cell clusters. Accordingly, we first used differential entropy to characterize the gene expression distribution in single-cell data, and subsequently established the S-E model to capture the relationship between differential entropy and mean gene expression (Figures R1(a) and R1(b)). Each point represents a gene, and “significant genes” refers to those with notable entropy reduction (denoted as ds). For a gene, ds across all cells equals the predicted entropy minus the observed entropy, reflecting the degree of disorder or randomness in its expression. A higher ds indicates greater entropy variation for a gene within a cell type, implying that its expression differs substantially from most genes in that cluster. A cell population with large summarization of significant ds is supposed to yield a small purity score (ROGUE value) and suggest higher heterogeneity.

To directly assess cluster purity, we computed the ROGUE value for each cell type and annotated

the average ROGUE score across all cell types in each S-E plot. We plotted boxplots of ROGUE values for each cell type (Figures R1(c) and R1(d)). As shown, the ROGUE values in Simulated_Cortex and Simulated_MBaging datasets approached 1 for most cell types, indicating high purity. In contrast, the Simulated_MB dataset exhibited relatively lower ROGUE values, suggesting greater heterogeneity in its annotations.

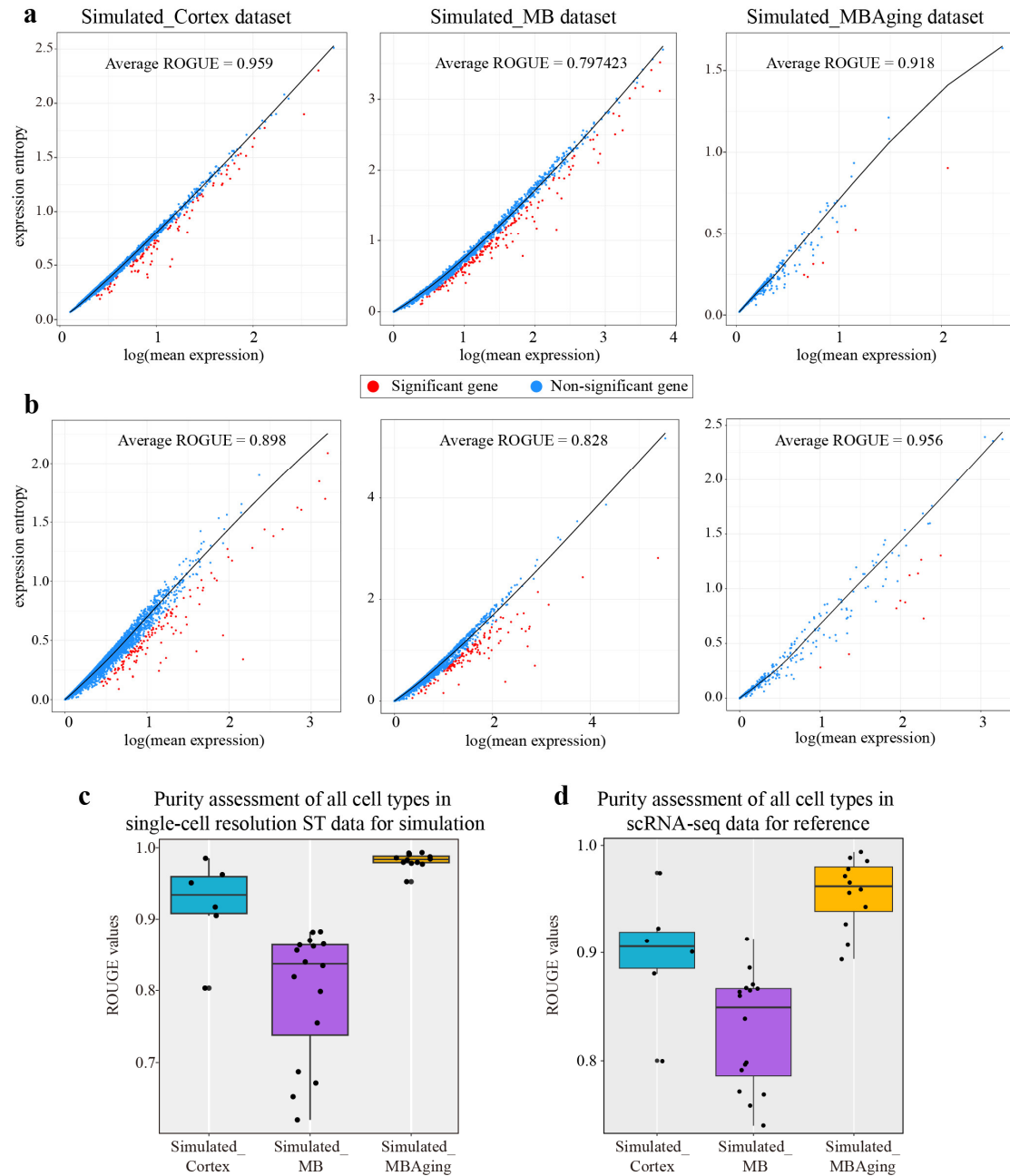


Figure R1. ROGUE measurement across all simulated datasets. **(a)** S-E plot of single-cell resolution ST data used for constructing spot-resolution Simulated_Cortex, Simulated_MB, and Simulated_MBaging datasets, respectively. **(b)** S-E plot of the scRNA-seq data used for reference in Simulated_Cortex, Simulated_MB, and Simulated_MBaging datasets, respectively. In (a) and (b), each point represents a gene. **(c)** Boxplot of computed ROGUE values for all cell types in single-cell resolution ST data. **(d)** Boxplot of computed ROGUE values for all cell types in scRNA-seq data. In (c) and (d), each point denotes a cell type, and each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

(2.3) We also calculated the area under the ROC curve (AUC) of marker genes to evaluate the

expression specificity of these markers.

The AUC values and ROC curves provide a quantitative measure of a marker gene's ability to distinguish the target cell type from others, thereby serving as an indicator of annotation accuracy based on known markers. We plotted ROC curves for a subset of marker genes across simulations and labeled with their corresponding AUC values (Figures R2-R4). As shown, the Simulated_Cortex and Simulated_MBaging datasets achieved AUC values close to 1 for most genes and cell types, indicating high specificity and reliable annotations. In contrast, the Simulated_MB dataset exhibited lower AUC values, which is consistent with the lower cluster purity revealed by the ROGUE metric (Figure R1).

The ROC curve of marker genes for each cell type in Simulated_Cortex dataset

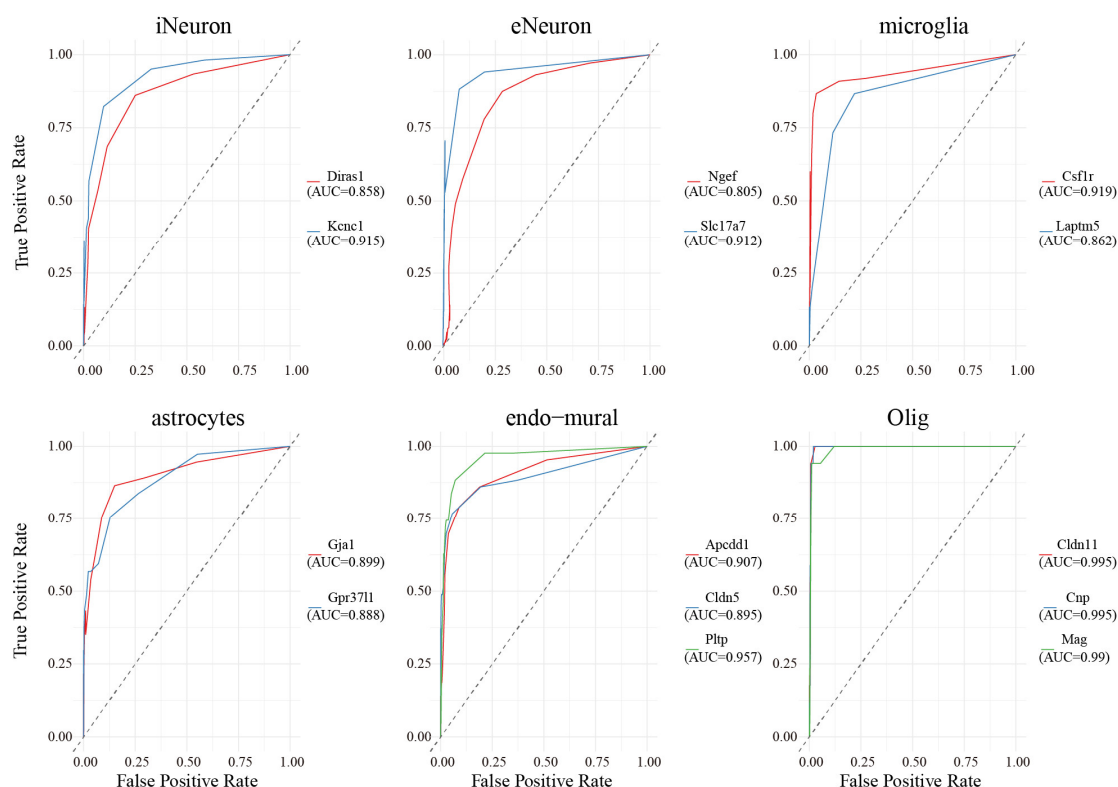


Figure R2. ROC curves of a subset of marker genes for each cell type in the Simulated_Cortex dataset. Each subplot denotes a cell type and the ROC curve of its genes is distinguished by different colors.

The ROC curve of marker genes for each cell type in Simulated_MB dataset

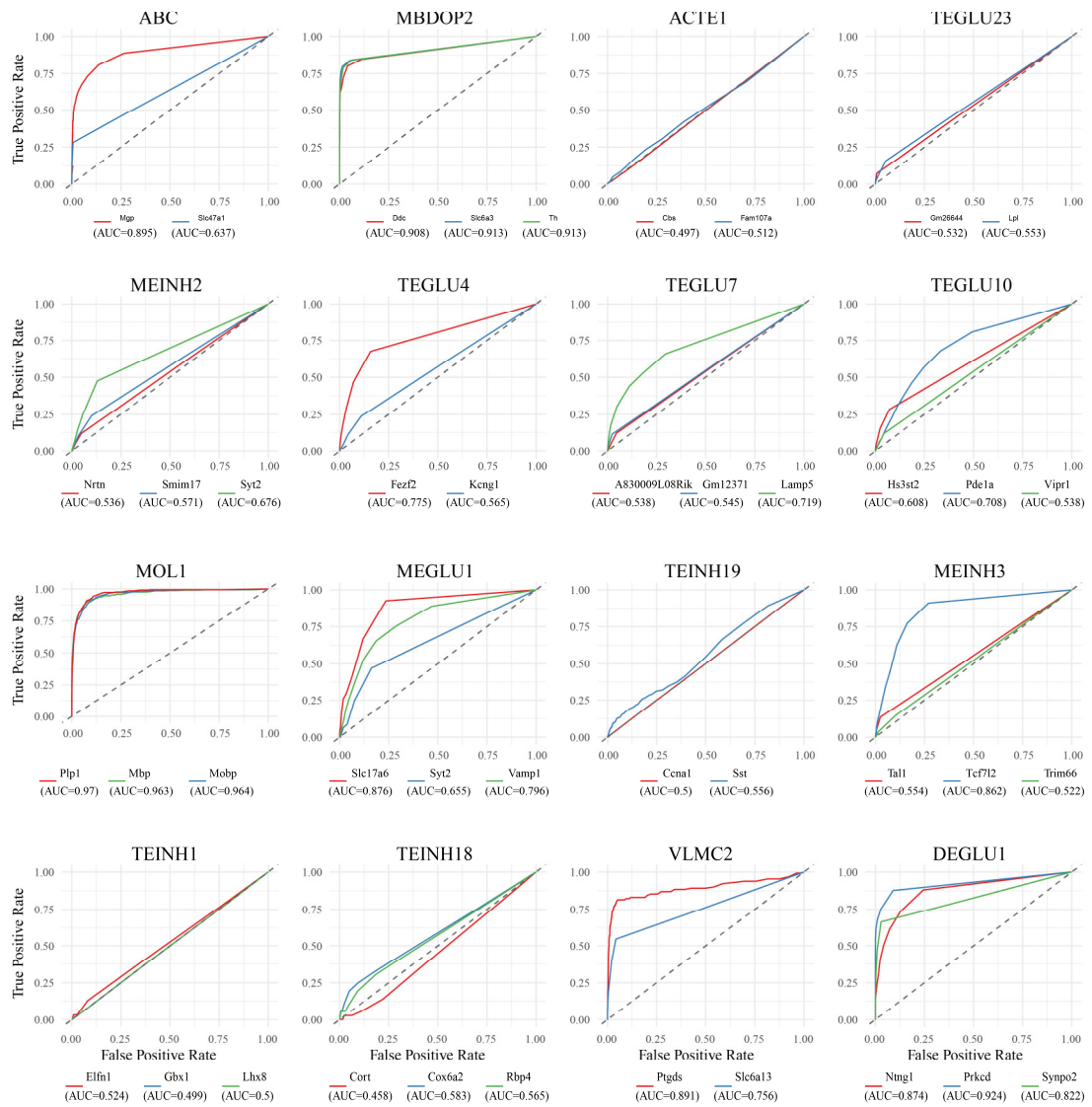


Figure R3. ROC curves of a subset of marker genes for each cell type in the Simulated_MB dataset. Each subplot denotes a cell type and the ROC curve of its genes is distinguished by different colors.

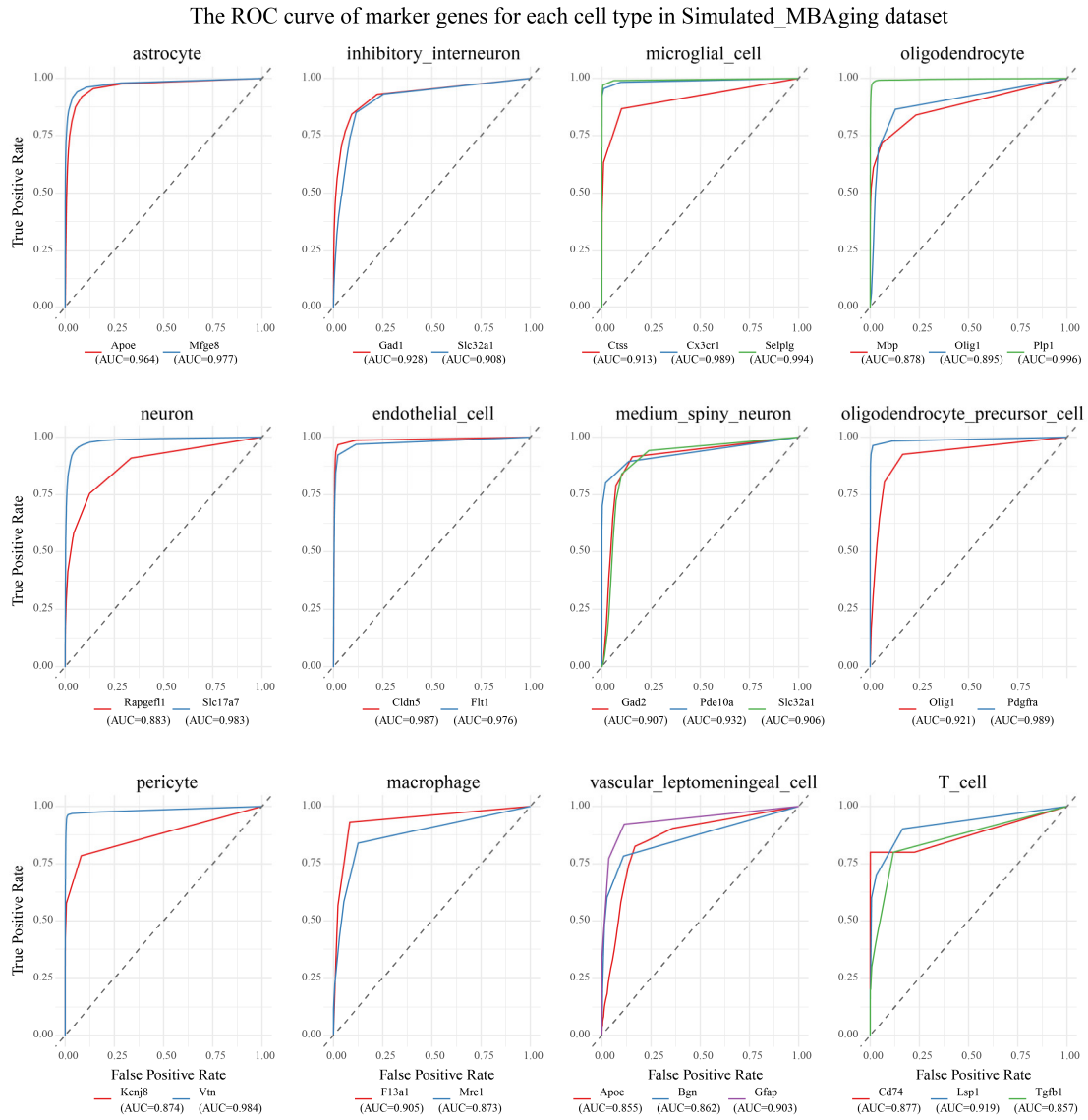


Figure R4. ROC curves of a subset of marker genes for each cell type in the Simulated_MBaging dataset. Each subplot denotes a cell type and the ROC curve of its genes is distinguished by different colors.

In summary, the above validations demonstrated that both the Simulated_Cortex and Simulated_MBaging datasets achieved reliable cell type annotation that can be considered as ground-truth. In contrast, the Simulated_MB dataset performed relatively poorly. Accordingly, in the revised manuscript, we retained the Simulated_Cortex dataset, included the newly added Simulated_MBaging dataset, and removed the Simulated_MB dataset from our analysis.

(3) We have removed results for the Simulated_MB dataset and added results for Simulated_MBaging datasets in the revised manuscript.

In subsection “Performance evaluation of SWOT on inferring cell-type composition” of the Results section, we have updated the following descriptions and figures.

The single-cell resolution ST data used to generate Simulated_MBaging datasets came from MERFISH technology in a mouse brain aging system³⁹. By setting different grid sizes, we created seven datasets with varying numbers of spots: 290, 426, 749, 989, 2850, 4669, 9554. For

consistency, we refer to them as: Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000, respectively.

In the preliminary analysis of overall rankings across all metrics and all simulations, SWOT and SONAR achieved the highest average performance (Figure 2b).

These metrics further emphasized the superior performance of SWOT compared to most other methods (Figures 2c and S4). To provide an overall assessment, we defined a combined metric, RJP, as the average of the three adjusted metrics (Figures 2d and S5). A two-sided Wilcoxon rank-sum test between SWOT and other methods on Simulated_MBaging datasets revealed statistically significant differences in performance. On the Simulated_Cortex dataset, SWOT, CARD and SONAR demonstrated commendable average performance, while in another dataset, SWOT and Stereoscope got notable results. Furthermore, we normalized these three metrics in RJP and computed their weighted average, termed composite score, to evaluate each method across all Simulated_MBaging datasets (Figures 2e and S6). With the increase of the number of spots, SWOT and CARD gained robustness, Stereoscope and Uniport achieved satisfactory results on large-scale samples, while the performance of other methods decreases.

Finally, to reveal the spatial distribution of cell types, we combined the estimated cell-type composition with spatial coordinates to obtain the spatial probability distribution for each cell type in Simulated_Cortex and Simulated_MBaging_300 datasets (Figures 2f, 2g, S7 and S8). Although Stereoscope performed well in above metrics, its inferred spatial distributions showed a near-uniform distribution of each cell type across spots, which was inconsistent with the true underlying tissue structure. In contrast, SWOT exhibited high concordance with the ground-truth spatial patterns, and successfully inferred the spatially aggregated cell types such as MSN in Simulated_MBaging_1000 and the layered structured cell type exemplified by VLMCs in Simulated_MBaging_3000 dataset (Figures 2h and S9), which corresponded with the known tissue organization. Comparatively, the proportions of these cell types inferred by other deconvolution methods were less pronounced and accurate. These results confirmed that SWOT not only outperformed most existing cell-type deconvolution methods but also provided robust and spatially coherent estimation on cellular composition.

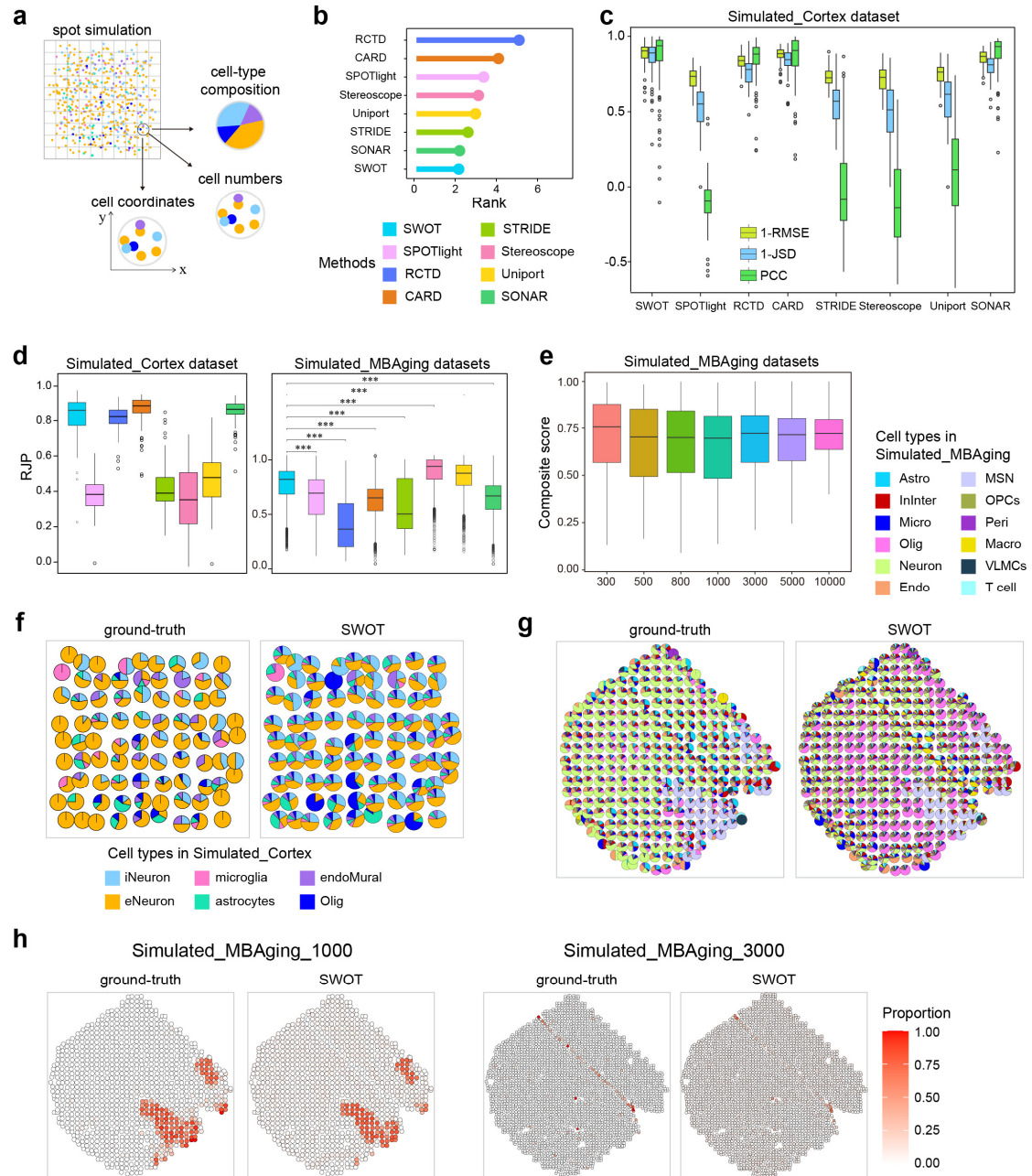
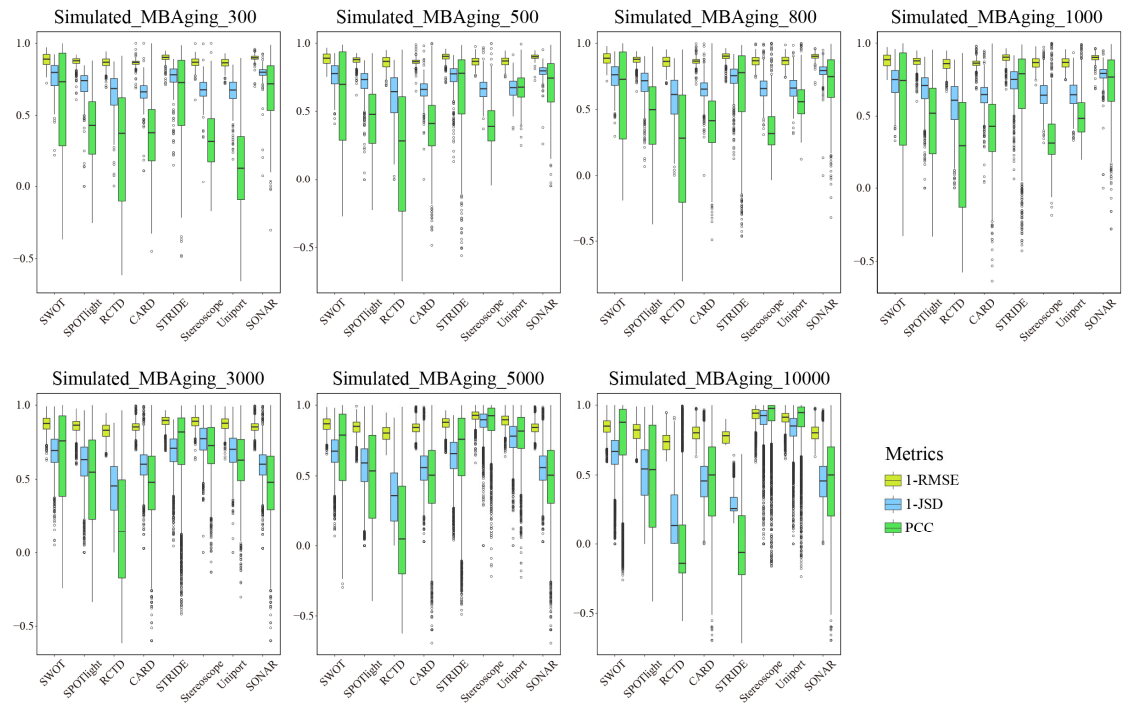


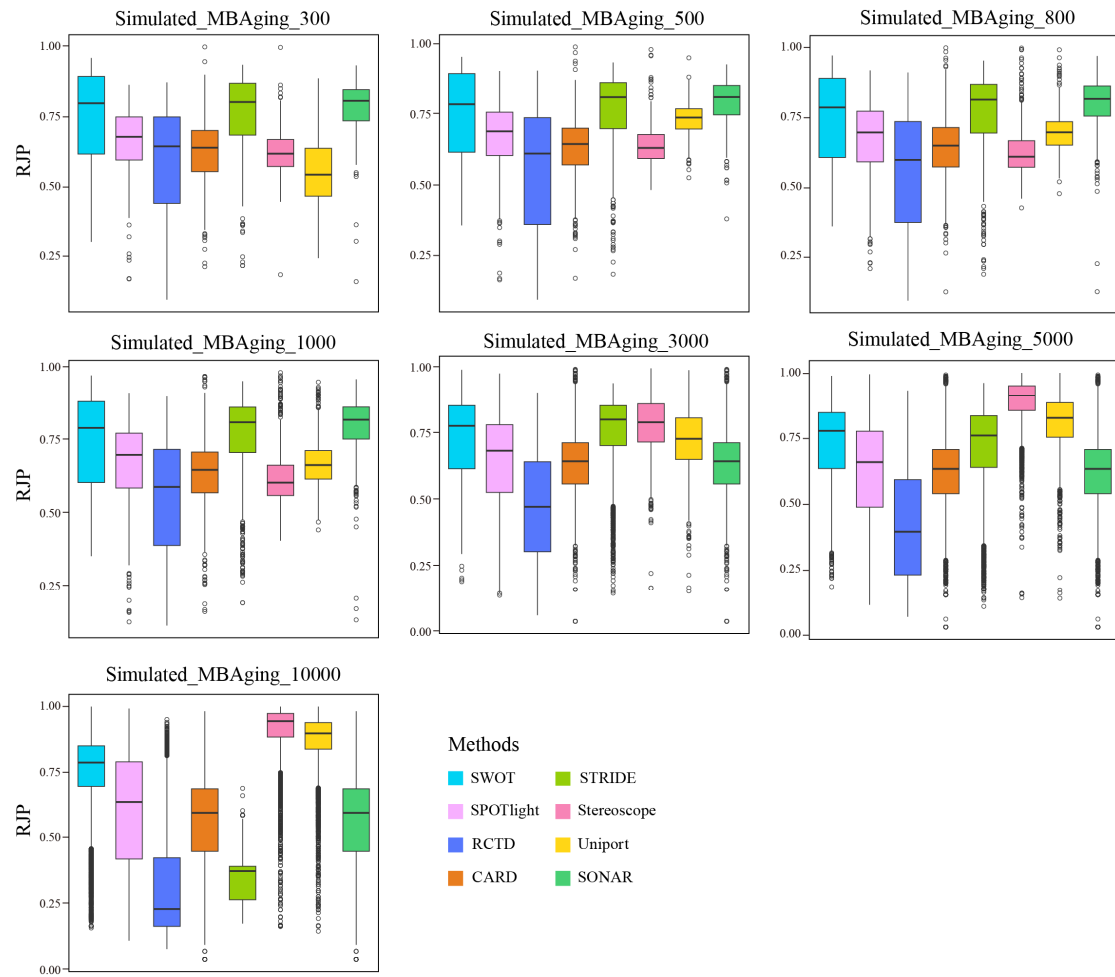
Figure 2. Performance evaluation for cell-type composition inference on simulated datasets. **a** Spot-resolution simulated ST data generation. The cell-type composition, cell numbers, cell coordinates are established as the ground-truth. Circles denote spots, points denote cells, and colored by cell types. **b** Rank of all methods on RMSE, JSD, and PCC across all simulated datasets. **c** Evaluation of the adjusted metrics for the Simulated_Cortex dataset. **d** Boxplots of RJP for Simulated_Cortex (left) and all Simulated_MBaging (right) datasets. Statistically significant differences were marked by a two-sided Wilcoxon rank-sum test in Simulated_MBaging datasets. P-values were adjusted using Benjamini-Hochberg method, and asterisks denote statistical significance (***) $p < 0.001$. **e** Boxplots of composite score across all spots in each Simulated_MBaging dataset. The horizontal axis represents the simplified name of the Simulated_MBaging dataset. In c-e, each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, whiskers extending 1.5 times the interquartile range, and points outside are outliers. **f** and **g** Spatial scatter pie of cell-type composition obtained by ground-truth and SWOT for Simulated_Cortex and Simulated_MBaging_300 datasets, respectively. Each pie denotes a spot, colored by cell types, and divided by proportions. **h** Spatial distribution of cell types MSN and VLMCs displayed in ground-truth and SWOT for the Simulated_MBaging_1000 and Simulated_MBaging_3000, respectively. Color indicates cell-type proportions of each spot.

Figure S4. Evaluation of deconvolution on Simulated_MBaging datasets



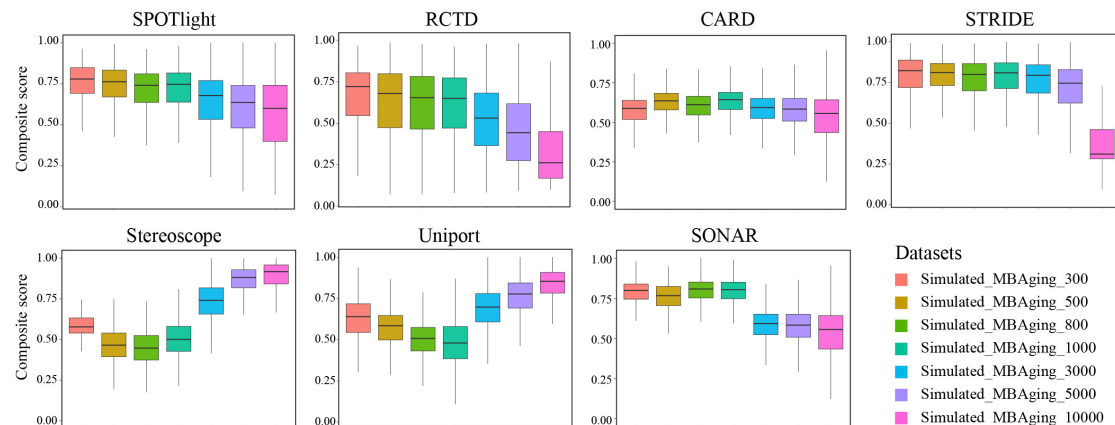
Boxplots of the $1 - RMSE$, $1 - JSD$, and PCC values across all Simulated_MB Aging datasets were generated by all deconvolution methods. Each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, whiskers extending 1.5 times the interquartile range, and points outside are outliers.

Figure S5. Performance of RJP on Simulated_MB Aging datasets



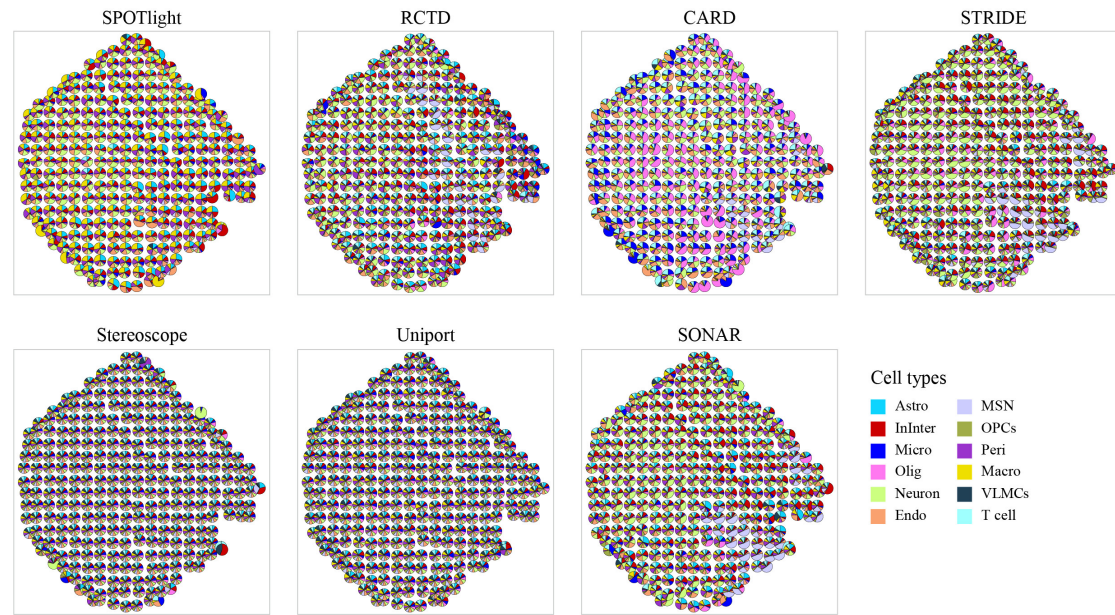
Boxplots for the mean performance of RMSE, JSD, and PCC, i.e., RJP, for all Simulated_MBaging datasets across all deconvolution methods. Each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, whiskers extending 1.5 times the interquartile range, and points outside are outliers.

Figure S6. Evaluation of composite score on Simulated_MBaging datasets



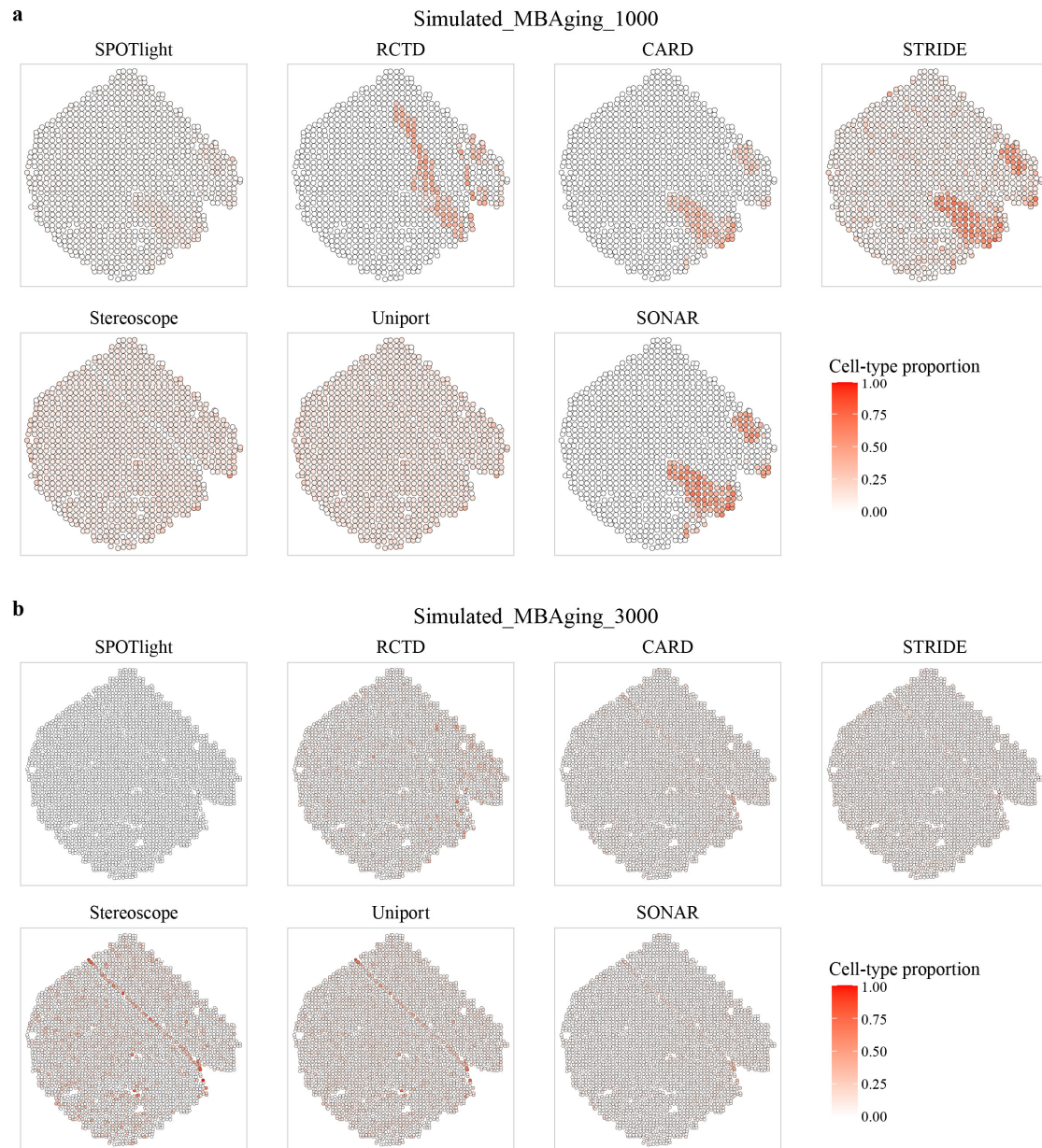
Boxplots of the composite score of seven cell-type deconvolution methods (SPOTlight, RCTD, CARD, STRIDE, Stereoscope, Uniport, and SONAR) in all Simulated_MBaging datasets. The composite score was defined as a weighted average of the normalized RMSE, JDS, and PCC. Since RMSE and JSD are negatively correlated with accuracy (lower is better), and PCC is positively correlated (higher is better), we adjusted the directionality of RMSE and JSD so that a higher composite score consistently reflects better performance. Each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

Figure S8. Spatial scatter pie of cell types on Simulated_MBaging_300 dataset



Spatial scatter pie of cell-type composition in Simulated_MBaging_300 dataset estimated by seven cell-type deconvolution methods, including: SPOTlight, RCTD, CARD, STRIDE, Stereoscope, Uniport, and SONAR. Each pie in each panel denotes a spot, colored by cell types, and divided by cell-type proportions.

Figure S9. Spatial distribution of cell types on Simulated_MBaging dataset



Spatial distribution of cell type MSN in Simulated_MBaging_1000 dataset (**a**) and cell type VLMCs in Simulated_MBaging_3000 dataset (**b**) estimated by seven cell-type deconvolution methods, including: SPOTlight, RCTD, CARD, STRIDE, Stereoscope, Uniport, and SONAR. The color in each plot indicates the cell-type proportions and the spatial distribution of the corresponding cell type in that spot.

In subsection “Performance evaluation of SWOT on inferring single-cell spatial maps” of the Results section, we have updated the following descriptions and figures.

First, in each simulated dataset, we examined the cell number estimation strategy using three quantitative metrics: (1) the ratio between the estimated and the ground-truth cell numbers per spot, (2) the Pearson correlation coefficient between the estimated and the ground-truth values across all spots, and (3) the Euclidean distance between the estimated and the ground-truth cell numbers per spot. In the Simulated_Cortex dataset, SWOT obtained average values of 1.581 for the ratio, 0.27 for the correlation, and 2.776 for the distance, indicating the accuracy of its cell number estimation (Figure 4a). Across all Simulated_MBaging datasets, SWOT also achieved competitive average results in these metrics (Figure 4b). The considerable performance of SWOT across Simulated_MBaging with different dropout rates

further supported its robustness in estimating cell numbers (Figure S15 and Table S2). Notably, CellTrek tended to overestimate the cell numbers, while CytoSPACE tended to underestimate it. Second, for the estimated spatial coordinates, we calculated a Euclidean distance between the estimated and ground-truth coordinates for SWOT and CellTrek. In Simulated_Cortex dataset, we computed the average coordinates distance for individual cells within the same cell type in each spot (Figure 4c), where smaller distances indicate more accurate estimation. We also plotted the average distance for all cell types in each Simulated_MBaging dataset (Figure 4d). The combination of accurate cell number estimation and consistently small distances across simulations demonstrated the reliability of SWOT.

Then, we observed the spatial distributions of individual cells in single-cell spatial maps inferred by SWOT, CellTrek, and CytoSPACE. The spatial distribution inferred by SWOT were consistent with the ground-truth spatial structures in Simulated_Cortex and Simualted_MBaging_1000 datasets (Figures 4e and 4f). Within the Simulated_Cortex dataset, SWOT and CytoSPACE both captured spatial patterns more accurately than CellTrek, which showed poor performance in cell type Olig. Finally, we analyzed the expression patterns of the top 100 cell-type-specific genes for cell type VLMCs, which has a significant layered structure, in Simulated_MBaging_1000 dataset (Figure 4g). The alignment of these gene expression patterns with the true spatial organization further highlighted that the single-cell spatial maps inferred by SWOT can reveal spatial patterns for cell-type-specific genes. Although CellTrek showed a more prominent and enriched gene expression signal, the pattern did not align with the actual spatial structure.

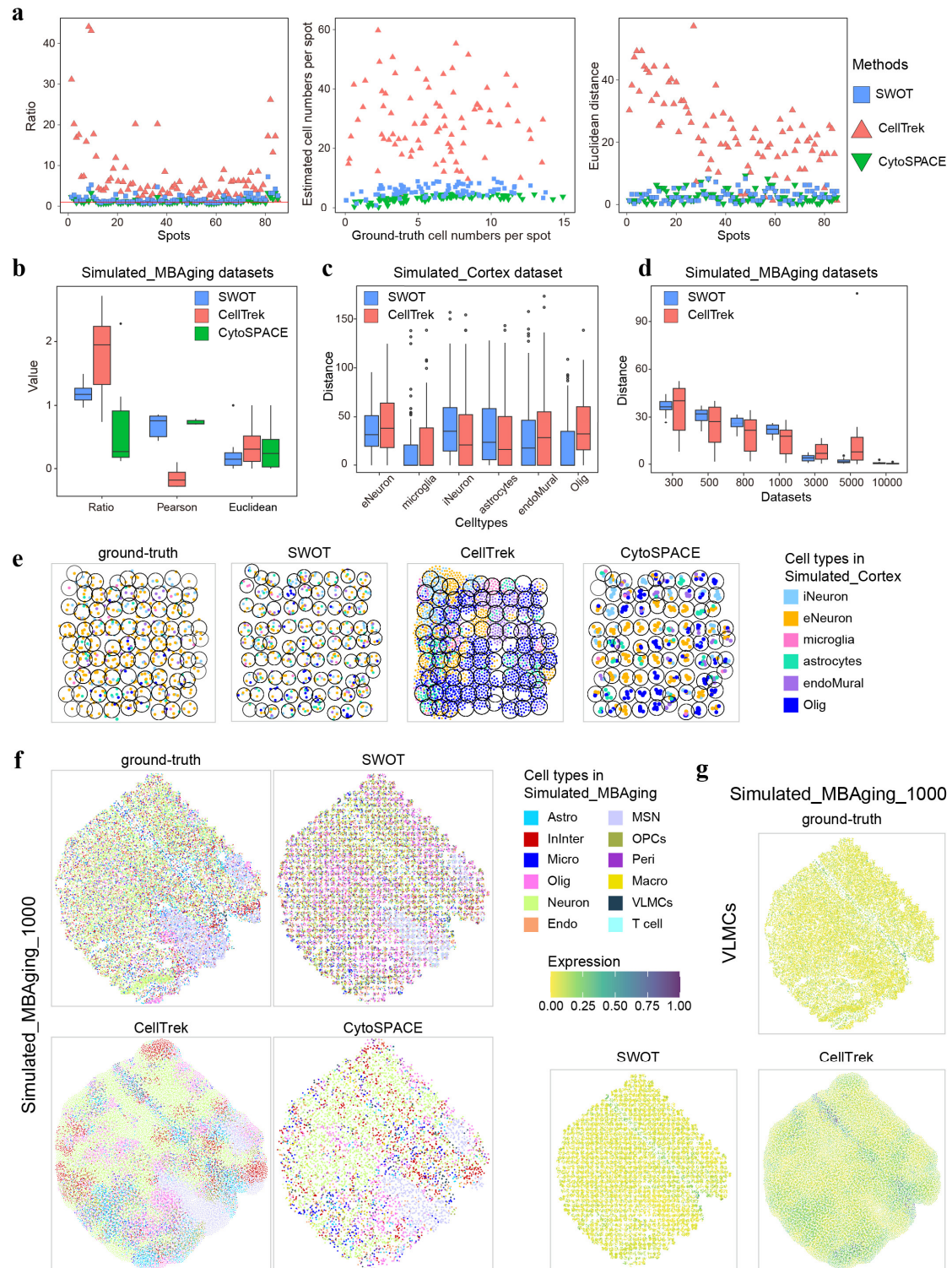


Figure 4. Performance evaluation for single-cell spatial maps inference on simulated datasets. **a** Performance of cell number estimation strategy in Simulated_Cortex dataset. Left: Ratio between estimated and ground-truth cell numbers per spot. Middle: Pearson correlation coefficient between estimated and ground-truth cell numbers across all spots. Right: Euclidean distance between estimated and ground-truth cell numbers per spot. Colors and shapes denote methods, and the straight line in the left panel indicates a ratio of 1, representing the best estimation. **b** Boxplots of the average values for the ratio, Pearson correlation coefficient, and Euclidean distance between estimated and ground-truth cell numbers across all Simulated_MBAGing datasets. **c** Distance between ground-truth and estimated cell coordinates of each cell type on Simulated_Cortex dataset. **d** Distance between ground-truth and estimated cell coordinates of all cell types on all Simulated_MBAGing datasets. The horizontal axis represents the simplified name of the dataset. In **b-d**, each box plot is the interquartile range of the

computed values, spanning from the first quartile to the third quartile, with the median as the central line, whiskers extending 1.5 times the interquartile range, and points outside are outliers. **e** Spatial distributions of cells in the inferred single-cell spatial maps generated from ground-truth, SWOT, CellTrek, and CytoSPACE on Simulated_Cortex dataset. **f** Spatial distributions of cells in the inferred single-cell spatial maps generated from ground-truth, SWOT, CellTrek, and CytoSPACE on Simulated_MBAGing_1000 dataset. In **e** and **f**, circles denote spots, points denote cells, and colored by cell types. **g** Expression patterns of cell-type-specific genes on VLMCs in single-cell spatial maps from ground-truth, SWOT, and CellTrek on Simulated_MBAGing_1000 dataset. Color represents the summed expression levels of top 50 cell-type-specific genes.

Comment 3.

Some hyperparameters. The method involves several hyperparameters, yet I noticed limited discussion regarding them. These include λ , α , and ϵ in equation (5). Their values will impact the performance of the method, so the authors should clarify how they are determined and evaluate the robustness of the results concerning different hyperparameter values.

Response: We appreciate your comment. We have added experiments and discussions to clarify how hyperparameters are determined, and evaluated the robustness of the results concerning different hyperparameter values.

In the revised main text, we have added a “Hyperparameters selection” subsection in the Methods section to discuss all hyperparameters.

Hyperparameters selection. To evaluate the robustness and sensitivity of SWOT, we conducted experiments across a range of hyperparameter settings. Involved hyperparameters include: the value of k in k NN for constructing distance matrix, the pre-clustering method and its resolutions for aggregating gene expression information, the bandwidth for defining spatial weights, the strength for controlling spatial weights, and the penalty strength for computing transport plan. Additional details are available in Supplementary Methods.

SWOT initiates by constructing a k -nearest neighbor (k NN) graph to compute pairwise distances. It involves three hyperparameters ($n_neighbors_cell$, $n_neighbors_spot$, and $n_neighbors_pos$) in three distances (expression distance between cells in scRNA-seq data, expression distance between spots in ST data, and coordinates distance between spots in ST data). Experimental evaluations showed that while increasing the k -values leads to higher time consumption, the results remain robust (Figure S34). We selected $n_neighbors_cell=5$, $n_neighbors_spot=5$, and $n_neighbors_pos=5$ as the default settings.

During constructing spatial weights, SWOT requires a pre-clustering step to integrate gene expression information from ST data. No significant differences across various clustering algorithms and resolution settings indicated SWOT's robustness (Figure S35). We chose the Leiden algorithm as default and set its resolution parameter such that the number of clusters matches the number of cell types in scRNA-seq data. Another hyperparameter is the bandwidth, which determines the radius over which each central spot aggregates information from its neighboring spots. Based on performance under different bandwidth settings (Figure S36), we balanced accuracy and computational efficiency and selected 0.1 as the default value. The hyperparameters ρ_{indiff} and ρ_{outsam} are used to control spatially weighted strength. Results across different values suggested that setting both parameters to 10 consistently yields

optimal performance across the majority of datasets (Figure S37).

In optimal transport module, solving the unbalanced and structured optimal transport problem in Eq. (5) involves three hyperparameters: λ , α , and ε . We conducted a comprehensive grid search to evaluate how different combinations of these hyperparameters affect SWOT's performance (Figures S38 and S39). Based on these evaluations, we recommend setting $\alpha=0.1$, $\lambda=100.0$, and $\varepsilon=0.1$ as the default configuration.

In the revised supplementary text, we have also supplemented a subsection “4. Setting of hyperparameter selection” in the Supplementary Methods section.

4. Setting of hyperparameter selection

The hyperparameters involved in SWOT include: the value of k in k NN for constructing distance matrix, the pre-clustering method and its resolutions for aggregating gene expression information, the bandwidth for defining spatial weights, the strength for controlling spatial weight, and the penalty strength for computing transport plan. We conducted experiments across a wide range of hyperparameter settings to investigate how each hyperparameter affects SWOT's performance.

For constructing distance matrices, SWOT begins by building a k -nearest neighbor (k NN) graph and then computing the shortest path distances between cells/spots on this graph. This process involves three hyperparameters (`n_neighbors_cell`, `n_neighbors_spot`, and `n_neighbors_pos`) in three types of distances (gene expression distance between cells in scRNA-seq data, gene expression distance between spots in ST data, and spatial coordinates distance between spots in ST data). To ensure that SWOT incorporates sufficient and informative neighbor information, we set 5 as minimum for each k , and used the floor of the logarithm of the number of cells/spots as maximum value in each dataset. By setting four distinct k -values for each of the three distances, resulting in 64 combinations per dataset, we evaluated on the accuracy and runtime across simulated and real datasets (Figure S34). Results reflected that SWOT maintained stable performance across a wide range of k -values, and increasing `n_neighbors_cell` significantly raises computational cost, as a larger k leads to a denser k NN graph. We selected `n_neighbors_cell=5`, `n_neighbors_spot=5`, and `n_neighbors_pos=5` as default. This configuration balances computational efficiency and performance, while effectively captures the influence of first-order neighbors in both gene expression and coordinates.

To assess the impact of clustering method and resolution on SWOT's performance, we compared two widely used clustering algorithms, Louvain and Leiden, with a range of resolution parameters (Figure S35). SWOT assumes that scRNA-seq and ST data share the same set of cell types; therefore, we used the number of annotated cell types in scRNA-seq data as a reference. We validated SWOT under six conditions: where the number of clusters in ST data was (i) equal to, (ii) greater than, or (iii) less than the number of known cell types, for both clustering methods. Results demonstrated high similarity in results across all six conditions for simulated and real datasets. Statistical comparisons using the Kruskal-Wallis test revealed no significant differences, indicating that SWOT is robust to both the choice of pre-clustering algorithm and resolution parameter.

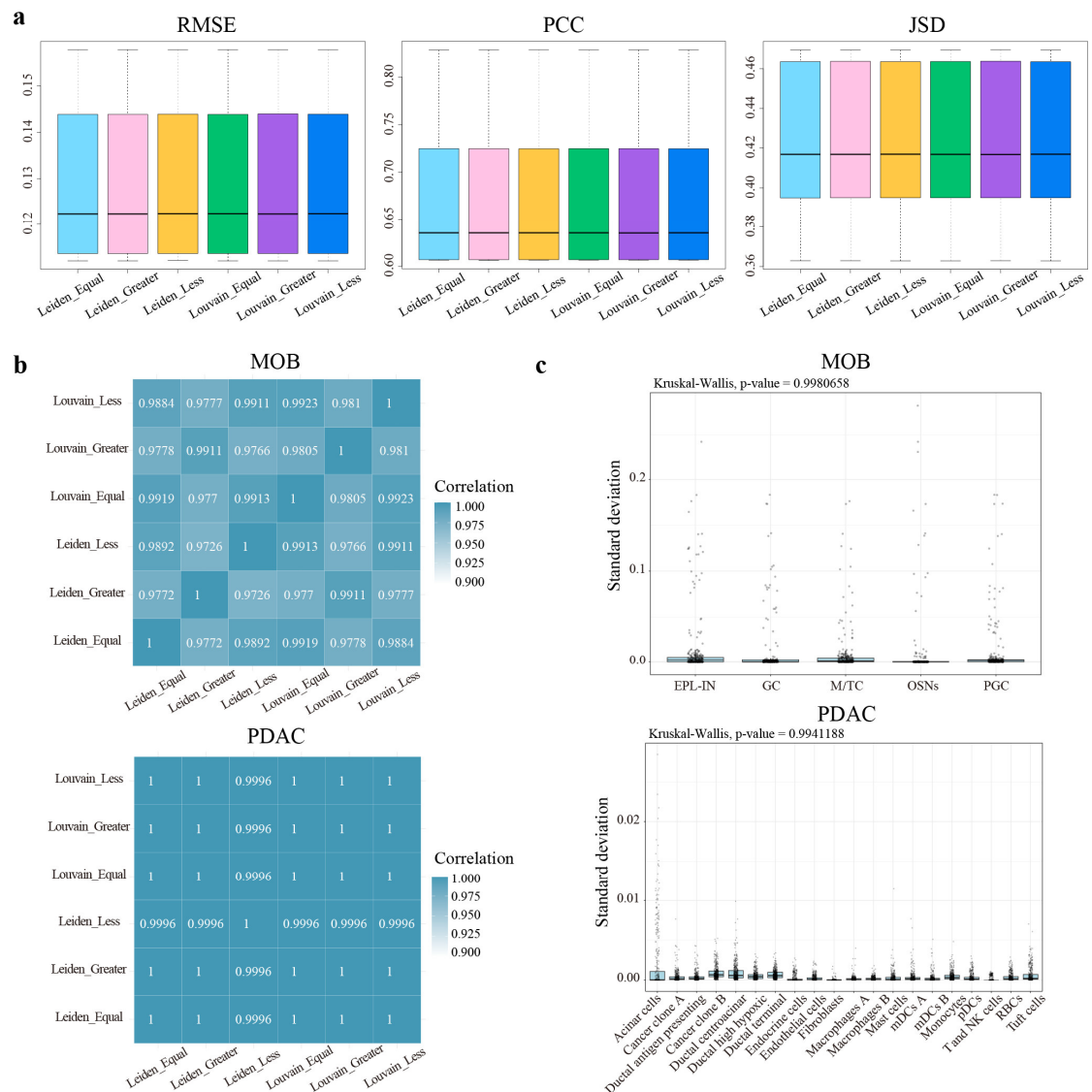
We also examined another hyperparameter involved in spatially weighted strategy, bandwidth, it determines the radius over which each central spot aggregates information from its neighboring spots.

A larger bandwidth includes more distant neighbors; however, these may be less similar to the central spot, potentially introducing noise¹⁸. Therefore, we assessed how different bandwidth values affect the performance (Figure S36). We selected six bandwidth values (0, 0.1, 0.5, 1.0, 2.5, and 5.0) for validation, where $b=0$ indicates a model without spatial weighting. Across both simulated and real datasets, the accuracy, running time, correlation and Kruskal-Wallis test on SWOT demonstrated that, except for the case $b=0$, changes among bandwidth groups were minimal.

Two hyperparameters (ρ_{indiff} and ρ_{outsam}) control the spatially weighted strength. The first ensures that within the neighborhood, the weight between spots from different clusters is lower than that between spots from the same cluster. The second ensures that outside the neighborhood, the distance between spots from the same cluster is smaller than that between spots from different clusters. For all simulated datasets, we performed a grid search strategy to examine the impact of different strength combinations (10, 100, and 1000, generated nine in total) on inferred cell-type compositions (Figure S37). Experiments on the different strengths suggested that the choice of 10 yields optimal performance across the majority of datasets.

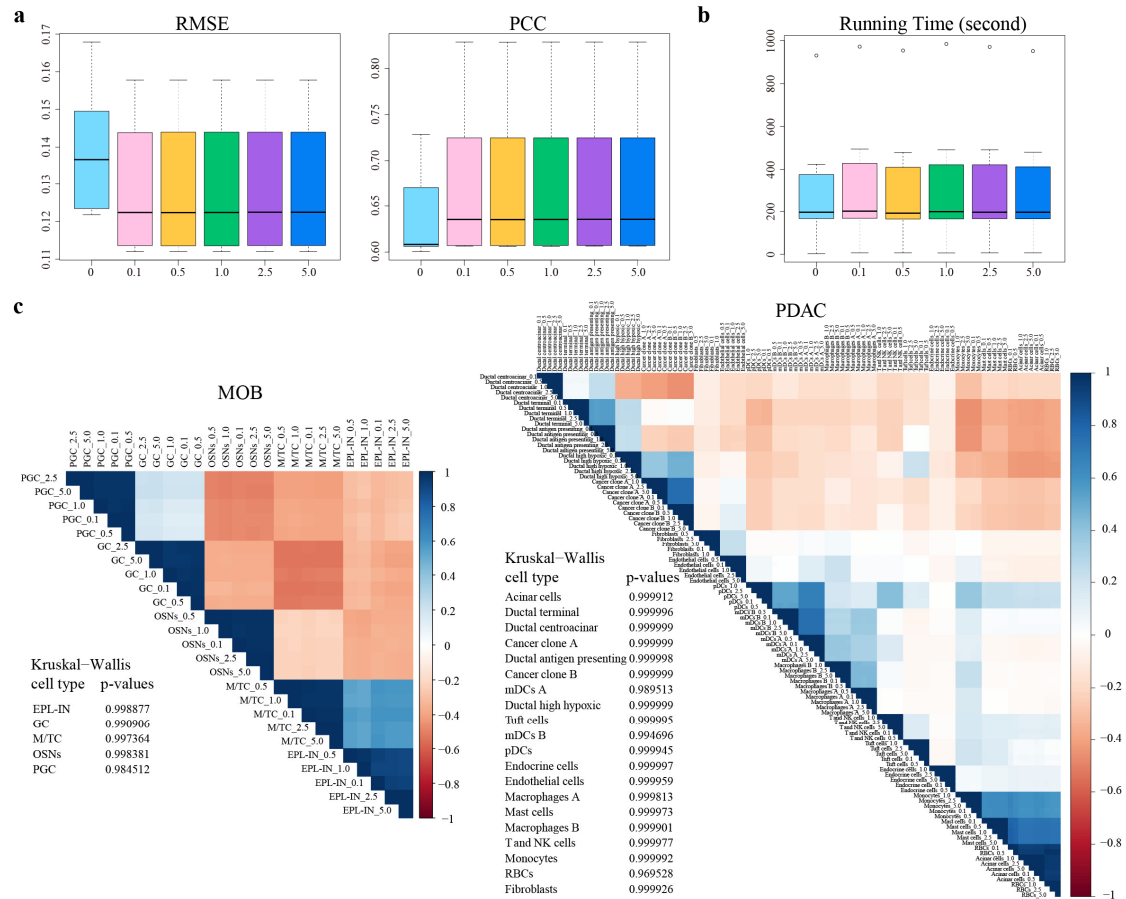
In solving the optimal transport problem, three hyperparameters are involved: λ , α , and ε , which control the penalty strengths for the unbalanced, the structural, and the entropy regularization terms, respectively. We conducted a grid search strategy over 64 combinations (4 values for each parameter) to evaluate their impact on SWOT for both simulated and real datasets (Figures S38 and S39). Based on these results, we recommend the following default values: $\alpha=0.1$, $\lambda=100.0$, and $\varepsilon=0.1$.

Figure S34. Effect of k -values in k NN for constructing distance matrix



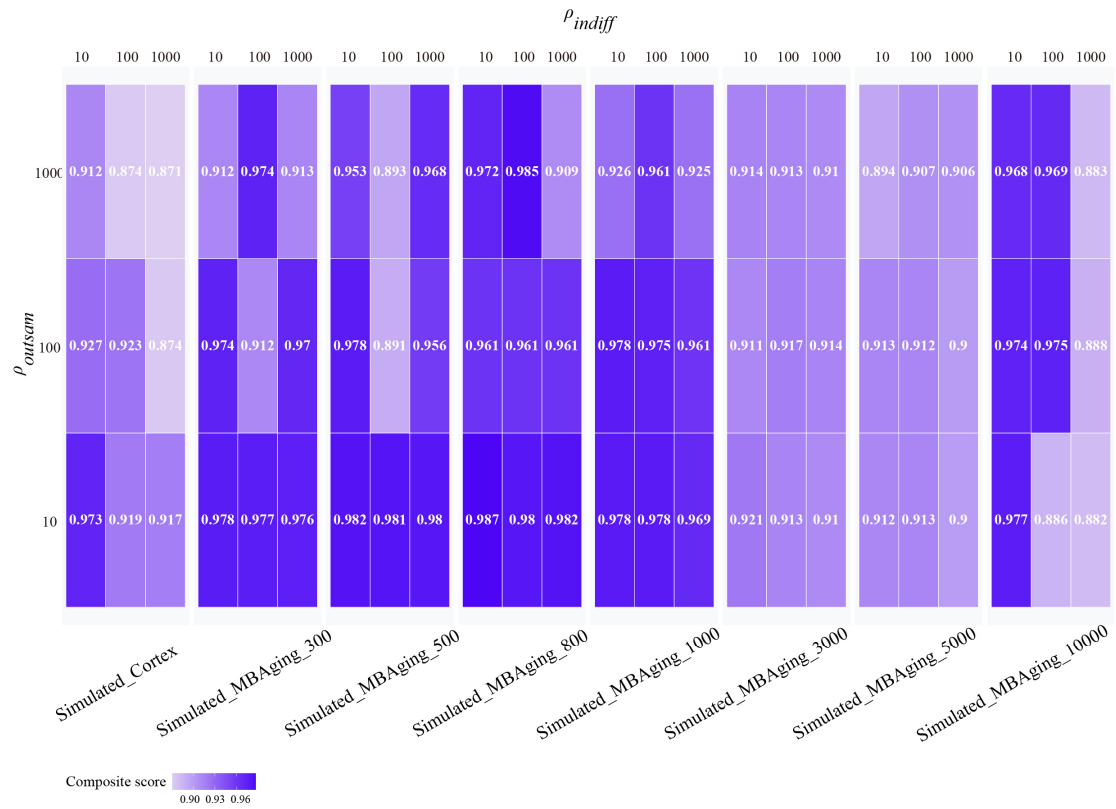
(a) Boxplots of the average RMSE, average PCC, and average JSD of the inferred cell-type compositions across all simulated datasets under each condition, with different colors representing different conditions. (b) Heatmap of the pairwise Pearson similarity of inferred cell-type composition for MOB and PDAC datasets under different settings, with more intense colors indicating higher similarity. Label in each plot is formatted as Method_Resolution (e.g., Leiden_Equal refers to using the Leiden algorithm with a resolution setting that results in the cluster numbers equal to the cell type numbers in scRNA-seq data). (c) Cell type-wise standard deviation across all spots under all conditions for MOB and PDAC datasets, with highly stable cell types represented by compact boxes. Average p-values of all cell types were marked by Kruskal-Wallis test. In (a) and (c), each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

Figure S36. Robust of bandwidth for constructing spatial weights



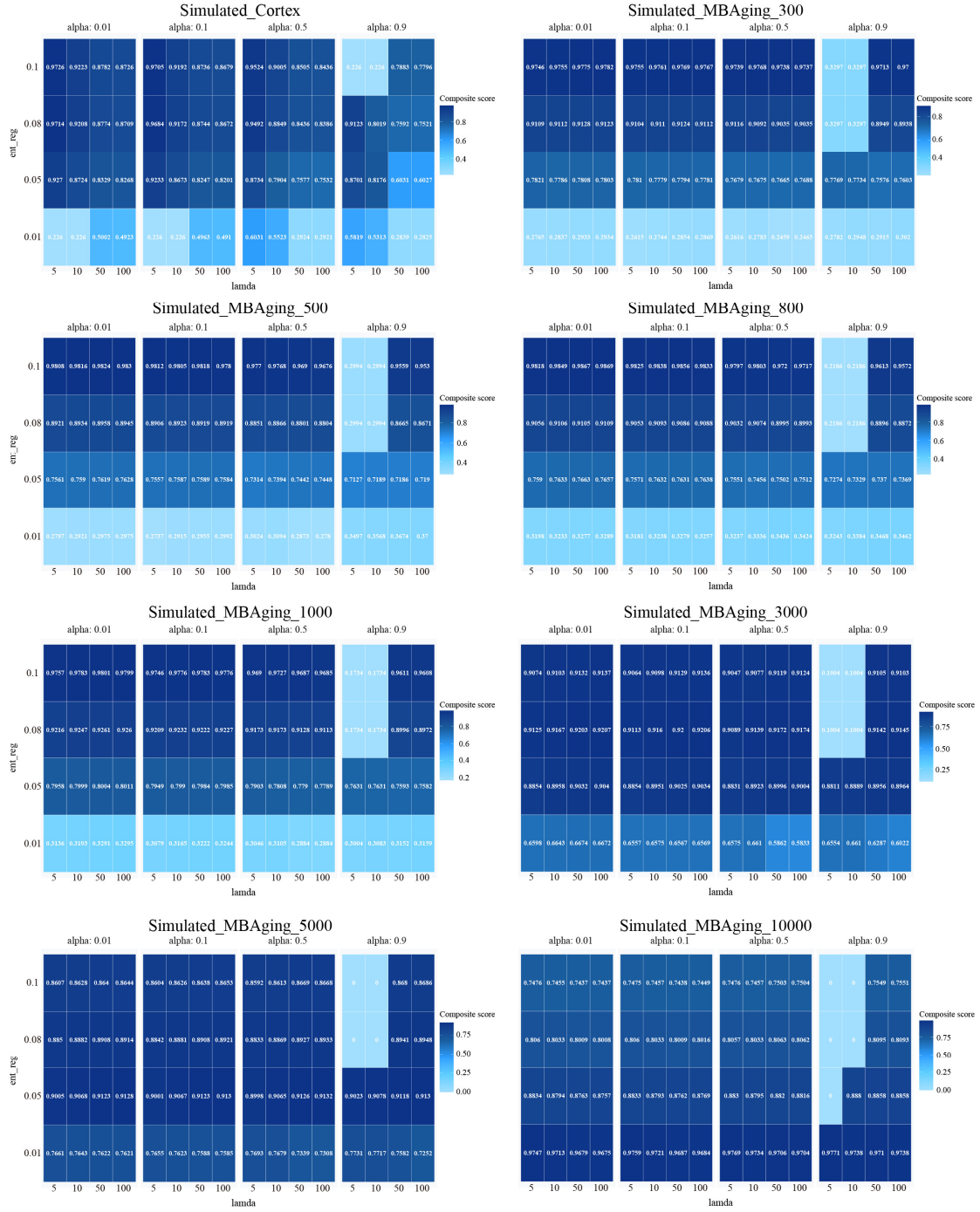
The accuracy (a) and running time (b) of SWOT in inferring cell-type composition on all simulations under various bandwidth settings: 0, 0.1, 0.5, 1.0, 2.5, and 5.0, where $b = 0$ indicates a model without spatial weighting. Boxplots displayed the average RMSE, average PCC, and running time across all simulations under each bandwidth, with different colors representing different values. (c) Heatmaps of Pearson correlation coefficient for inferred proportion of each cell type across bandwidths. Labels in the correlation plots are formatted as Celltype_Bandwidth (e.g., PGC_2.5 refers to the proportion of cell type PGC inferred under bandwidth 2.5). P-values for each cell type were marked by Kruskal-Wallis test. In (a) and (b), each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, whiskers extending 1.5 times the interquartile range, and points outside are outliers.

Figure S37. Effect of strength for controlling spatial weight



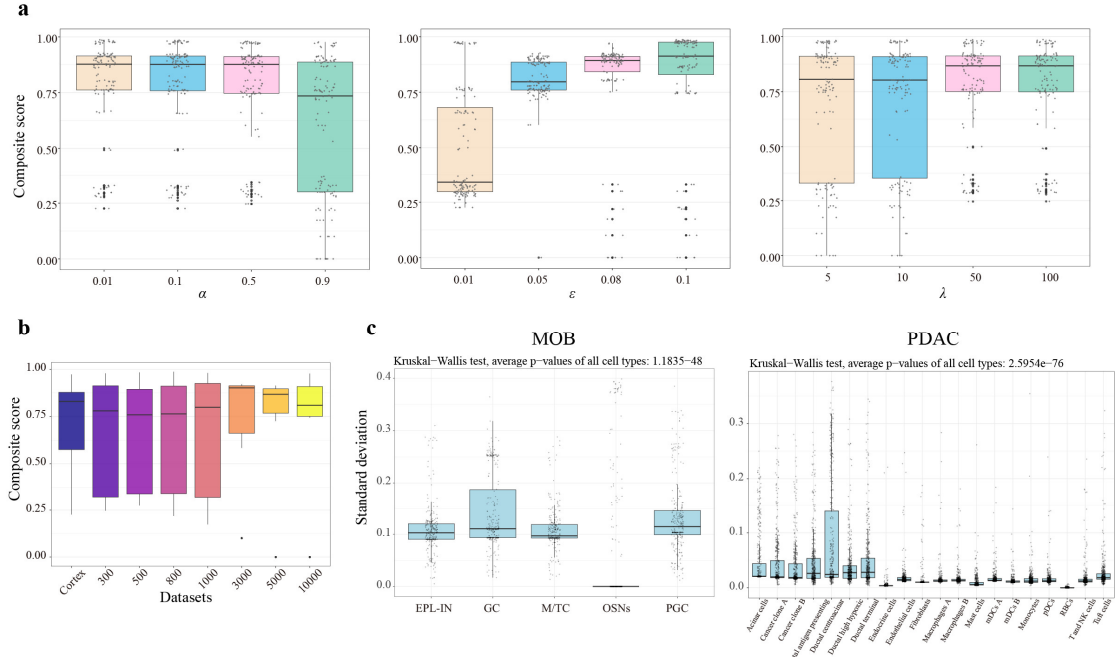
Heatmap of the composite scores across different settings of hyperparameters ρ_{indiff} and ρ_{outsam} on all simulated datasets. The composite score is calculated as a weighted average of the normalized RMSE, PCC, Spear, and JSD metrics. Each value in the heatmap represents the composite score for a particular combination, with more intense colors indicating higher scores.

Figure S38. Heatmap of hyperparameters for solving optimal transport problem



Heatmap of composite scores across all 64 combinations of hyperparameters λ , α , and ϵ , for all simulated datasets. The composite score is a weighted average of the normalized RMSE, PCC, Spear, and JSD metrics. Each value in the heatmap represents the composite score for a particular combination, with more intense colors indicating higher scores.

Figure S39. Effect of hyperparameters for solving optimal transport problem



(a) Boxplots of the composite score distributions across hyperparameters λ , α , and ε . (b) Boxplots of the composite score distributions across all simulated datasets. (c) Cell type-wise standard deviation across all spots under the 64 combinations of the three hyperparameters in MOB and PDAC datasets, and highly stable cell types are characterized by compact boxplots. Average p-values of all cell types in each dataset were marked by Kruskal-Wallis test. Each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

Comment 4.

How are a and b determined in equation (5)? It does not seem reasonable because this assumes that we know the number of cells in each spot and how many times a particular cell should be used to map to spatial transcriptomics data. How are they utilized in the simulation data? Are there potential issues with information leakage?

Response: Thank you for your comments and questions, we answered from two aspects as follows.

(1) We have appended a detailed description of how parameters a and b in Eq. (1) and Eq. (5) are determined.

In the revised main text, the added depiction was presented in subsection “Unbalanced and structured optimal transport” of the Methods section.

$\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$ are the discrete distributions of mass on the set of cells and spots. Typically, the two values are initialized by uniform distribution with each element be the same, that is $\mathbf{a} = \frac{1}{m} \mathbf{1}_m$,

$\mathbf{b} = \frac{1}{n} \mathbf{1}_n$, and determined by the number of cells in scRNA-seq and spots in ST data, respectively.

In the supplementary text, we have added definitions in “1.1 Optimal transport” subsection.

We first give the definition of the two discrete measures with weight $\mathbf{a}$ and $\mathbf{b}$ as^{2,3}:

$\mathbf{a} = \sum_{i=1}^m a_i \delta_{s_i}$ and $\mathbf{b} = \sum_{j=1}^n b_j \delta_{t_j}$. δ_{s_i} and δ_{t_j} are the Dirac measures of mass located at cell s and spot t , a_i and b_j are probability masses associated to the Dirac measures, and $\mathbf{a}$, $\mathbf{b}$ belong to the probability simplex: $\sum_m^{def} = \{\mathbf{a} \in \mathbb{R}_+^m : \sum_{i=1}^m \mathbf{a}_i = 1\}$ and $\sum_n^{def} = \{\mathbf{b} \in \mathbb{R}_+^n : \sum_{j=1}^n \mathbf{b}_j = 1\}$. Typically, the two values are initialized by uniform distribution with each element be the same, that is $\mathbf{a} = \frac{1}{m} \mathbf{1}_m$ and $\mathbf{b} = \frac{1}{n} \mathbf{1}_n$.

(2) We also provided an explanation about the two parameters.

It can be seen from the definition of these two parameters that, they are primarily determined by the number of cells in scRNA-seq data and the number of spots in ST data, and are not related to the number of cells in each spot or how many times a particular cell is mapped to each spot.

Typically, in the implementation of optimal transport module, without any prior information, these two values are initialized by a uniform distribution, and each element is the same in the source or target domains. This can be seen from the code defining $\mathbf{a}$ and $\mathbf{b}$

(https://github.com/GaoLabXDU/SWOT/blob/main/SWOT/utis/sw_usot.py at line 64). In the optimal transport module, SWOT inputs a gene expression profile with cell types from scRNA-seq data and a gene expression profile with coordinates from ST data. In the cell mapping module, SWOT inputs a cell-to-spot mapping matrix calculated from the previous module, and the output are cell-type composition, the number of cells per spot, and cell coordinates. Importantly, neither module requires prior assumptions about the number of cells per spot or the mapping times of a particular cell.

In simulated datasets, the known number of cells in each spot is only serves as a ground-truth to validate the performance of the cell mapping module, without involving the computation of any stage of SWOT. Hence, there is no potential issue with information leakage in simulations.

Comment 5.

*Optimization method. Equation (5) only informs us about the objective function. Which optimization method did the authors use to estimate the optimizer? This is not addressed in the methods section. It also seems that calculating the third term (structured term) requires summation over $n^2 * m^2$ terms, which could be time-consuming.*

Response: We thank the reviewer for the comments. Our response is organized into two parts.

(1) We have incorporated the optimization method of optimal transport problem.

In the revised main text, we have added a related description in subsection “Spatially weighted strategy” of the Methods section.

In Eq. (5) the optimal transport term and the structural term together define a Fused Gromov-

Wasserstein (FGW) distance, which results in the quadratic optimization problem. We adopt Algorithms 1 and 2 from reference 29 to solve this optimization problem. The unbalanced and entropy terms lead to an unbalanced optimal transport problem with entropic regularization^{56,57}, which solved by using a Sinkhorn algorithm to approximate the optimal transport plan. Additional details are available in Supplementary Methods.

In the revised supplementary text, we have supplemented a new subsection “1.5 Solution of optimal transport problem” in Supplementary Methods section.

1.5 Solution of optimal transport problem

In the spatially weighted unbalanced and structured optimal transport problem, the optimal transport and the structural terms together form a Fused Gromov-Wasserstein (FGW) distance¹¹. The unbalanced and the entropy correspond to an unbalanced optimal transport with entropic regularization¹². In the following, we will describe the numerical optimization problem required to compute the FGW distance and the UOT between discrete distributions.

The FGW distance is designed to align pairs of support points that are both similar in feature space and structurally close. Following the optimization framework proposed in Ref. [11], we focus on the case of two datasets, where the computation of FGW becomes a quadratic optimization problem. In this case, the optimal transport term and the structural term can be reformulated as:

$$\min_{P \in \Pi(a,b)} \langle (1-\alpha)P + \alpha L(D_{se}, SWD_{tec}) \otimes P, P \rangle \quad (8)$$

where, $\Pi(a,b) = \{P \in \mathbb{R}_+^{m \times n} \text{ s.t. } P\mathbf{1}_n = \mathbf{a}, P^T\mathbf{1}_m = \mathbf{b}\}$ and $L(D_{se}, SWD_{tec}) \otimes P$ is a tensor-matrix

multiplication and defined as $L(D_{se}, SWD_{tec}) \otimes P = (\sum_{k,l} L(D_{se}(i,k), SWD_{tec}(j,l)) P_{k,l})_{i,j}$. We first show

how to compute it efficiently for a general class of loss functions. In Ref. [8], it suggests that if the loss can be written as $L(u,v) = f_1(u) + f_2(v) - h_1(u)h_2(v)$, for functions (f_1, f_2, h_1, h_2) , then,

$L(D_{se}, SWD_{tec}) \otimes P$ can be decomposed as $L(D_{se}, SWD_{tec}) \otimes P = c_{D_{se}, SWD_{tec}} - h_1(D_{se}) \cdot P \cdot h_2(SWD_{tec})^T$,

and $c_{D_{se}, SWD_{tec}} = h_1(D_{se})\mathbf{a}\mathbf{1}_n^T + \mathbf{1}_m\mathbf{b}^T h_2(SWD_{tec})^T$ is independent of P .

This property demonstrates that for this class of losses, one can efficiently compute $L(D_{se}, SWD_{tec}) \otimes P$ with the computational complexity $O(m^2n + n^2m)$, which using only matrix multiplications, instead of the $O(m^2n^2)$ complexity of the naïve implementation in the raw optimal transport computational.

Typically, the square loss $L(u,v) = L_2(u,v) = \frac{1}{2}|u-v|^2$ is used and constructed as $f_1(u) = u^2$, $f_2(v) = v^2$, $h_1(u) = u$, and $h_2(v) = 2v$. Then, in line with Ref. [11], the gradient can be expressed with the following partial derivative:

$$G = (1-\alpha)P + 2\alpha L(D_{se}, SWD_{tec}) \otimes P \quad (9)$$

We adopt Algorithms 1 and 2 in Ref. [11] to solve the optimization problem. This approach only requires computing the gradient of Eq. (9) at each iteration and solving a classical OT problem. The line-search step involves constrained minimization of a second-order polynomial function. Although

the overall objective is non-convex, the conditional gradient method is known to converge to a local stationary point.

The second and fourth terms in Eq. (7) correspond to an UOT problem with entropic regularization, solved using the Sinkhorn algorithm^{12,13}. Similar to the classical OT problem, exact computation of UOT is computationally expensive and not scalable with respect to the number of support points n . To overcome this, the classical OT problem introduces an entropic regularization term to obtain an approximate solution. Inspired by this, entropic regularization is also incorporated into the UOT problem, as defined in Eq. (3).

According to Definition 1 in Ref. [11], to obtain an approximate transport plan for the entropically regularized UOT problem, the dual formulation is:

$$\min(u, v) := \varepsilon \sum_{i=1}^m \exp\left(\frac{u_i - D_{ij}}{\varepsilon}\right) \sum_{j=1}^n \exp\left(\frac{v_j}{\varepsilon}\right) + \lambda \langle e^{-u/\lambda}, \mathbf{a} \rangle + \lambda \langle e^{-v/\lambda}, \mathbf{b} \rangle \quad (10)$$

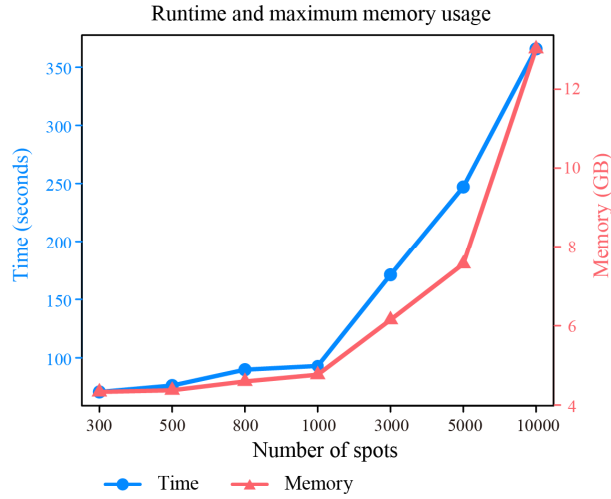
Then, using Sinkhorn algorithm to alternatively minimizes the dual function in Eq. (10) with respect to u and v , can obtain the approximate optimal transport plan for the UOT problem with entropy regularization.

(2) According to the analysis in section Supplementary Methods 1.5, the calculation complexity of the third term (structured term) in Eq. (5) can be simplified to $O(m^2n + n^2m)$. Additionally, to validate the computational efficiency of the optimal transport module and the overall SWOT, we conducted experiments on Simulated_MBAGing datasets of varying scales.

Analysis of computational efficiency has been added to the Discussion section in the revised main text.

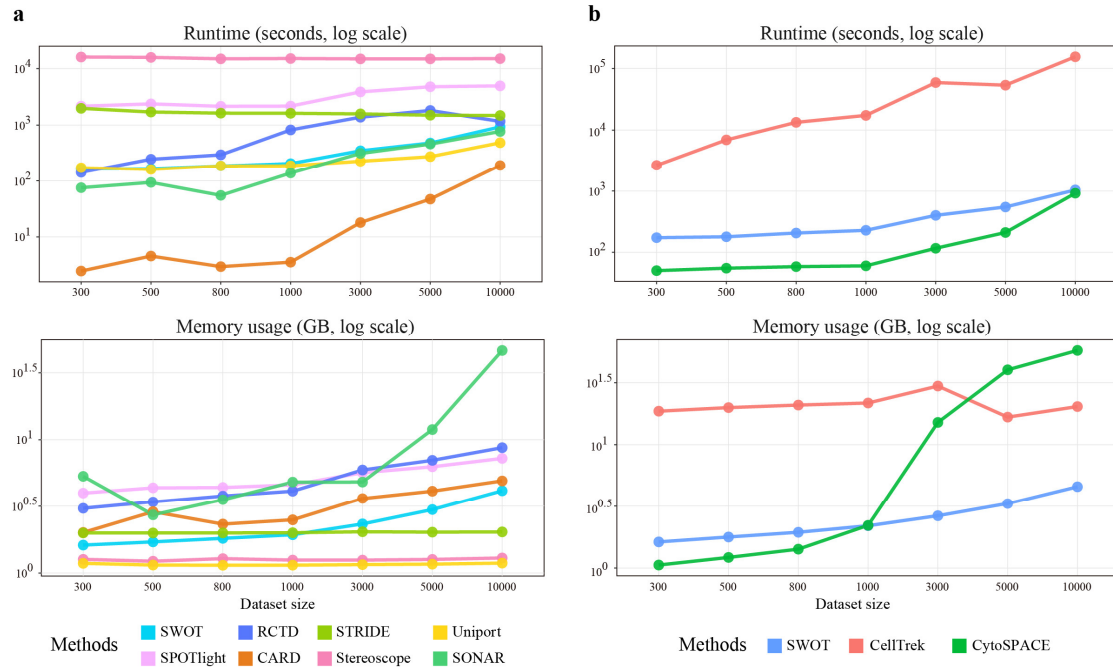
It should be noted that the primary computational cost of SWOT arises from the learning of cell-to-spot mapping in the optimal transport module, which scales with the cell numbers in scRNA-seq data and the spot numbers in ST data. Experimental results on Simulated_MBAGing datasets, with varying numbers of spots, indicated that the runtime and memory consumption of optimal transport module increase with the number of spots (Figure S32). Compared to existing methods, SWOT maintained competitive computational efficiency while delivering practical performance (Figure S33).

Figure S32. Runtime and memory usage of optimal transport module in SWOT



We used all Simulated_MBaging datasets to test the runtime (seconds, left y-axis) and maximum memory usage (GB, right y-axis) of the optimal transport module in SWOT. The Simulated_MBaging dataset consists of 15040 cells in reference (snRNA-seq) data, and the simulated ST datasets have 290, 426, 749, 989, 2850, 4669, 9554 spots, respectively. The horizontal axis represents the simplified name of all Simulated_MBaging datasets. All experiments were conducted on a workstation equipped with a 13th Gen Intel(R) Core(TM) i7-13700K CPU (3.4GHz, 16 cores) and 64GB of RAM running Windows.

Figure S33. Computational efficiency comparison of all methods



To comprehensively evaluate the computational efficiency and practical performance of SWOT, we conducted experiments on all Simulated_MBaging datasets with varying sizes, and plotted the runtime and memory consumption of all methods in two tasks: (a) cell-type composition inference and (b) single-cell spatial maps inference. Due to the wide range of values for time and memory usage among methods, we applied a log₁₀ transformation to both runtime and memory usage. Each subplot presents the trends after the logarithmic conversion, with different colors representing different methods. The Simulated_MBaging dataset has 15040 cells in reference (snRNA-seq) data, and the simulated ST datasets have 290, 426, 749, 989, 2850, 4669, 9554 spots, respectively. The horizontal axis represents the simplified name of all Simulated_MBaging datasets. All experiments were conducted on a workstation equipped with a 13th Gen Intel(R) Core(TM) i7-13700K CPU (3.4GHz, 16 cores) and 64GB of RAM running Windows.

Comment 6.

The equation (6) seems odd. There shouldn't be a summation over spot index j in the numerator.

Response: Thank you for pointing this out. We have addressed the issue as follows.

(1) In the revised main text, we have corrected an error in Eq. (6) in subsection “Estimation of cell-type proportions” of the Methods section.

Then, we denote the proportion of cell type r in spot j as M_{jr} :

$$M_{jr} = \frac{\sum_{i \in S_r^{cell}} P_{ij}^*}{|S_r^{cell}|} \quad (6)$$

(2) We also gave an explanation about the definition.

Eq. (6) is defined to estimate the cell-type proportion per spot, based on the cell type information in scRNA-seq data and transport plan derived from the optimal transport module. This strategy was originally proposed in [19] for deconvolution. It used an optimal transport plan for deconvolution by summing the probability matrix of every cluster according to given labels in scRNA data for each spot. This study demonstrated the accuracy of this strategy on real datasets¹⁹. In SWOT, we similarly obtained a transport plan, with the same objective of estimating cell-type proportion. Therefore, we followed the same way, by averaging the probability of each cell type in each spot and formulated Eq. (6) to estimate cell-type composition.

Comment 7.

The estimation of cell number per spot lacks detail and justification. The relationship between the number of cells and the non-zero proportion must be established in real or simulated data, and the exact formula needs to be specified.

Response: We appreciate your insightful comments and suggestions. We answered this question from the following three aspects.

(1) We gave an explanation about the estimation of cell numbers per spot.

Inspired by observations from previous studies^{20,21} indicating an inverse relationship between the proportion of zeros and the average total count per location in ST data, we hypothesized that spots with a higher proportion of zero expression values tend to contain fewer cells, and vice versa. Based on this principle, we established a log-linear model to represent the inverse relationship between the number of cells and the proportion of zeros. Although a general inverse trend across diverse scenarios was observed (Figure S3), the relationship does not hold consistently across all spots due to heterogeneity in spatial expression and capture efficiency. Moreover, different ST technologies differ in resolution and expected cell numbers. To account for this variability, we further refined the estimated number by adjusting them according to the specific technology used or the expected cell number ranges.

(2) We have added details and justification of cell number estimation in the revised manuscript.

For subsection “Overview of SWOT” of the Results section in the revised main text:

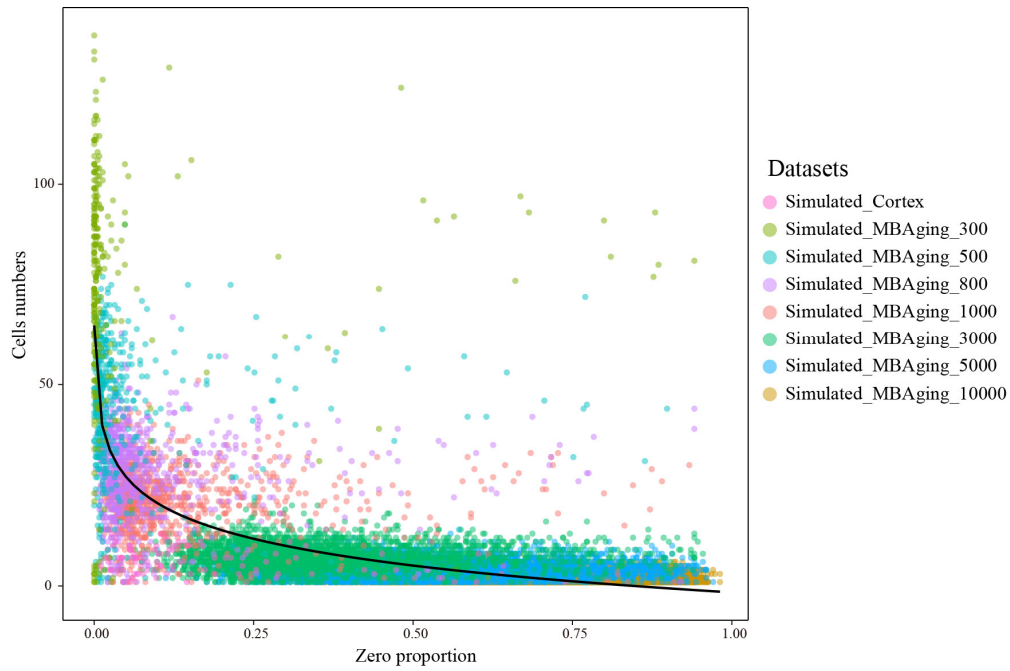
Given that the number of cells captured per spot varies across different technologies³³, and that the proportion of zero values in expression within a spot tends to be inversely proportional to the cell numbers within that spot^{33,34} (Figure S3), SWOT employs a log-linear model to estimate the cell numbers per spot.

For subsection “Estimation of cell numbers per spot” of the Methods section in the revised main text:

Estimation of cell numbers per spot. Inspired by the inverse relationship between the proportion of zeros and the average total count per location in ST data^{33,34}, we hypothesized that spots with a higher proportion of zero expression values tend to contain fewer cells, and vice versa (Figure S3). Then, SWOT estimates cell numbers per spot based on the zero proportions in gene expression and the sequencing technology that generated ST data or the predefined ranges of the expected cell numbers per spot. For each spot i , the proportion of zero values in its expression is z_i , the estimated cell numbers in spot i is c_i , and it is defined as $c_i = -10 \cdot \ln(z_i + 0.001) - 1$.

The established relationship is a log-linear model, and it describes the general inversely relationship between the zero proportions and the cell numbers per spot. Although a general inverse trend was observed, the approximation relationship does not hold consistently across all spots due to heterogeneity in spatial expression and capture efficiency. Different sequencing technologies differ in spatial resolution and expected cell numbers ranges per spot, such as 10x Genomics Visium typically captures 1–10 cells per 55 μm spot. The spatial resolution of Spatial Transcriptomics technology is 100 μm , and capturing 10–40 cells⁹. Slide-seq, characterized by a spot diameter of 10 μm , typically includes 1–3 cells^{10,33,58}. To account for this variability, we implement a post-estimation adjustment to refine the estimation according to the used technology or the expected ranges. Specifically, if a spot's estimated cell number falls outside the expected range, the estimate is adjusted by randomly assigning a value within the platform-specific range. We provide predefined expected ranges for the aforementioned three commonly platforms but do not limit to them. A 'user-defined' mode is also available, allowing users to specify custom values for the minimum and maximum number of cells per spot based on platform specifications or known knowledges. This strategy enables a flexible and adaptive estimation, ensures that each spot is assigned a biologically plausible number of cells, reduces overreliance on the log-linear assumption, and maintains consistency with the characteristics of the underlying technology.

Figure S3. Relationship between cell number and zero proportion per spot



We utilized all simulated datasets (Simulated_Cortex, Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000) to investigate the relationship between the proportion of zero values in gene expression within a spot and the number of cells present in that spot. As the true cell numbers per spot were known in all simulations, we extracted the cell numbers in each spot (y-axis) and calculated the corresponding proportion of zero values in gene expression within that spot (x-axis). Scatter plots were generated to visualize this relationship across all spots, and the black line is the fitting curve of the log-linear model.

(3) We further examined the established relationship using simulated and real datasets.

(3.1) In the revised manuscript, we have validated the relationship using simulated datasets and updated the descriptions in subsection “Performance evaluation of SWOT on inferring single-cell spatial maps” of the Results section.

First, in each simulated dataset, we examined the cell number estimation strategy using three quantitative metrics: (1) the ratio between the estimated and the ground-truth cell numbers per spot, (2) the Pearson correlation coefficient between the estimated and the ground-truth values across all spots, and (3) the Euclidean distance between the estimated and the ground-truth cell numbers per spot. In the Simulated_Cortex dataset, SWOT obtained average values of 1.581 for the ratio, 0.27 for the correlation, and 2.776 for the distance, indicating the accuracy of its cell number estimation (Figure 4a). Across all Simulated_MBaging datasets, SWOT also achieved competitive average results in these metrics (Figure 4b). The considerable performance of SWOT across Simulated_MBaging with different dropout rates further supported its robustness in estimating cell numbers (Figure S15 and Table S2). Notably, CellTrek tended to overestimate the cell numbers, while CytoSPACE tended to underestimate it. Second, for the estimated spatial coordinates, we calculated a Euclidean distance between the estimated and ground-truth coordinates for SWOT and CellTrek.

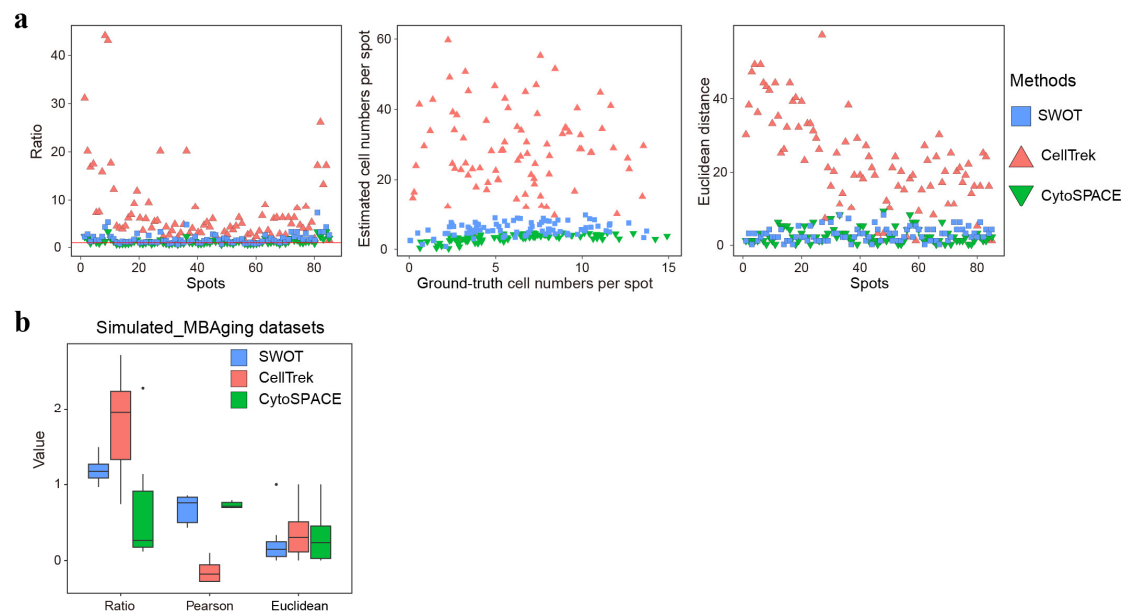
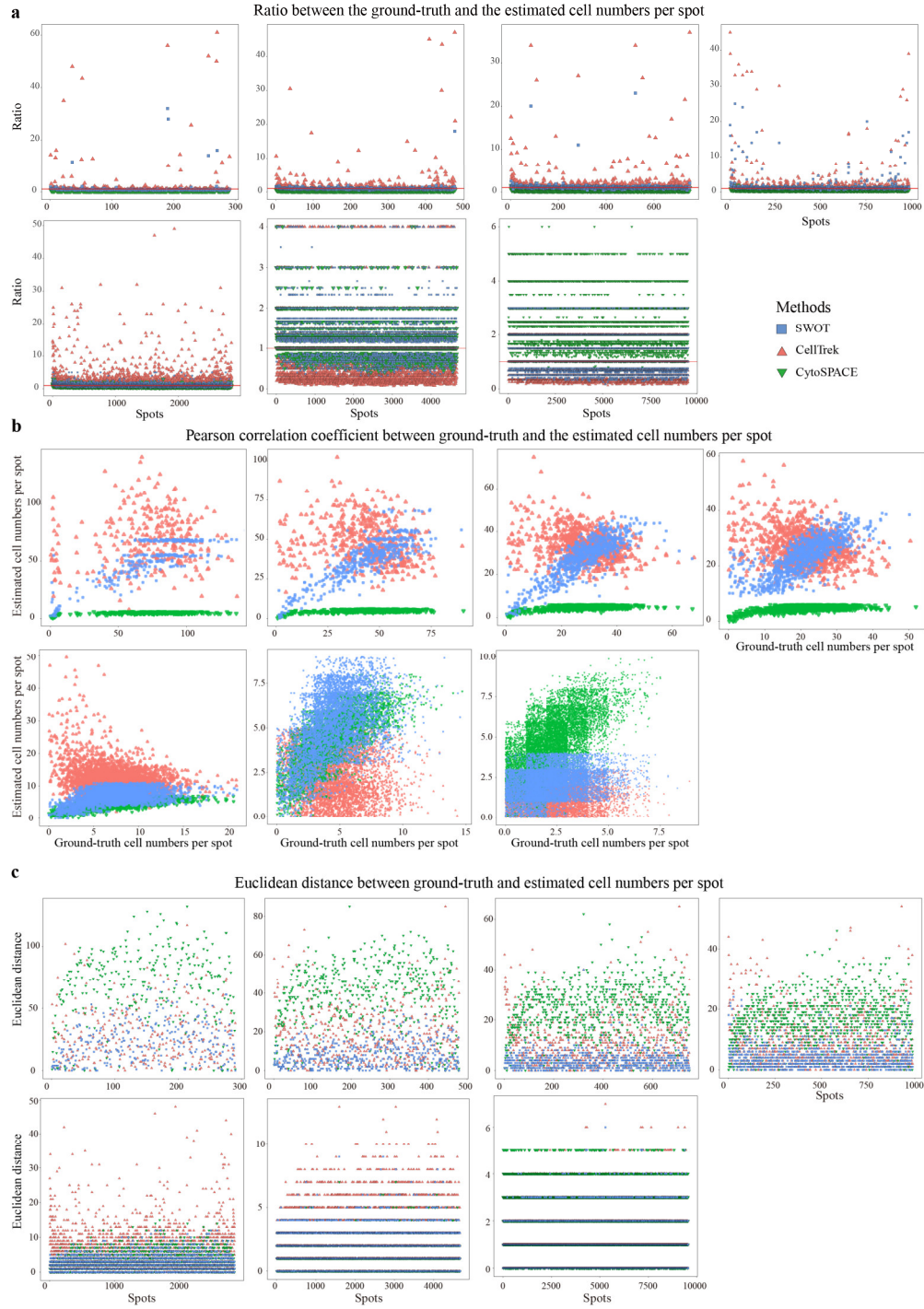


Figure 4. Performance evaluation for single-cell spatial maps inference on simulated datasets. **a** Performance of cell number estimation strategy in Simulated_Cortex dataset. Left: Ratio between estimated and ground-truth cell numbers per spot. Middle: Pearson correlation coefficient between estimated and ground-truth cell numbers across all spots. Right: Euclidean distance between estimated and ground-truth cell numbers per spot. Colors and shapes denote methods, and the straight line in the left panel indicates a ratio of 1, representing the best estimation. **b** Boxplots of the average values for the ratio, Pearson correlation coefficient, and Euclidean distance between estimated and ground-truth cell numbers across all Simulated_MBaging datasets.

Figure S15. Estimated cell numbers per spot on simulated datasets with different dropout rate



Seven Simulated_MBaging datasets (Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000) were generated under different spatial resolutions and exhibited distinct dropout rates of 0.0543, 0.0586, 0.0927, 0.1303, 0.3656, 0.5025, and 0.6781, respectively. (a) The ratio between estimated and ground-truth cell numbers for each spot. (b) The Pearson correlation coefficient between estimated and ground-truth values across all spots. (c) The Euclidean distance between estimated and ground-truth for each spot. In each subplot, all Simulated_MBaging datasets are arranged from left to right, top to bottom. Each dot in the plots represents an individual spot, with colors and shapes denoting methods. In (a), the straight line in each plot indicates a ratio value of 1, representing the best estimate of cell numbers. The average values of the three metrics (ratio, Pearson correlation coefficient, and Euclidean distance) across all spots for each dataset were summarized in Table S2.

Table S2. Evaluation of the estimated cell numbers per spot

For all Simulated_MBaging datasets, we compared the estimated cell numbers per spot by three methods (SWOT, CellTrek, and CytoSPACE) against the ground-truth using three metrics. We list the average ratio between estimated and ground-truth cell numbers per spot, the average Pearson correlation coefficient between estimated and ground-truth across all spots, and the average Euclidean distance between estimated and ground-truth values for each spot. The closer the ratio is to 1, the greater the correlation, and the smaller the distance, indicating a better estimation.

Metric	Method	Simulated_MBaging dataset						
		300	500	800	1000	3000	5000	10000
Ratio	SWOT	1.23	0.96	1.094	1.491	1.075	1.305	1.172
	CellTrek	2.714	1.948	1.854	2.071	2.403	0.742	0.801
	CytoSPACE	0.118	0.143	0.206	0.264	0.689	1.136	2.279
Pearson correlation coefficient	SWOT	0.812	0.856	0.852	0.758	0.432	0.553	0.443
	CellTrek	0.098	-0.18	-0.28	-0.282	-0.272	-0.118	0.004
	CytoSPACE	0.688	0.695	0.697	0.714	0.768	0.792	0.759
Euclidean distance	SWOT	23.183	8.219	3.979	4.256	2.206	1.526	0.708
	CellTrek	28.897	19	11.401	9.496	5.649	2.662	1.059
	CytoSPACE	67.645	39.098	23.346	16.621	2.939	0.964	2.304

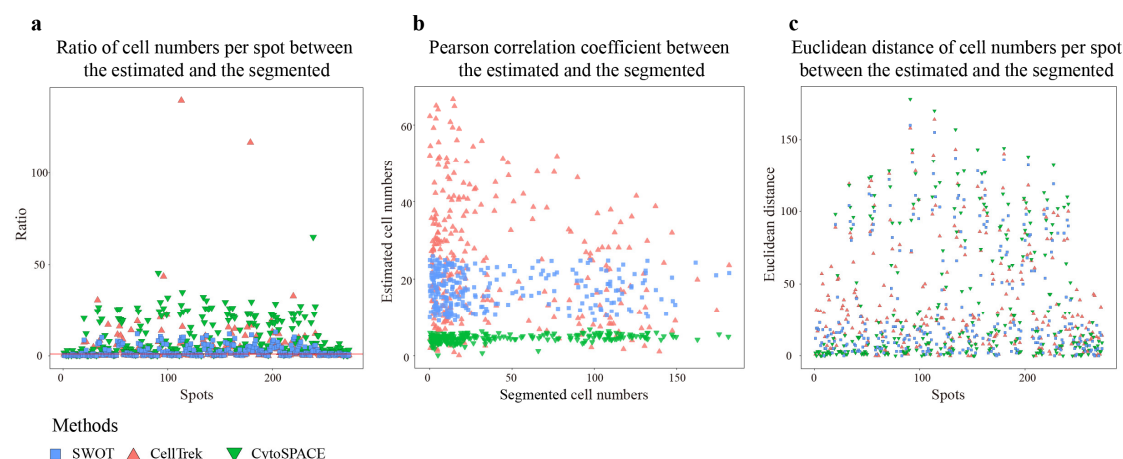
(3.2) We have also evaluated the established relationship using two real datasets.

Since the truth cell numbers, are not available in real ST data, we followed the approach used in prior previous (e.g., CytoSPACE²² and Tangram²³), and performed cell-segmentation on the H&E-stained histological images. Although segmentation results cannot be regarded as a gold standard, they provided a histomorphology perspective in measuring the spatial positions of individual cells and the number of cells. We compared the cell numbers derived from the segmentation with estimated by SWOT, CellTrek, and CytoSPACE. The experimental results demonstrated that SWOT produced a reasonable estimation of the number of cells per spot.

In the revised manuscript, we have incorporated the depictions on estimation for MOB dataset in subsection “SWOT depicts cell-type spatial distributions” of the Results section.

Since the truth cell numbers are not available in real ST data, to evaluate the reliability of the cell number estimation strategy employed by SWOT, we compared the estimated cell numbers per spot by these three methods with the segmented cell numbers obtained through cell-segmentation⁴¹ from H&E image (Figure S14). Based on the segmented numbers, SWOT showed superior performance, achieving a spot-wise ratio close to 1, the highest correlation, and the lowest Euclidean distance.

Figure S14. Estimated cell numbers per spot on MOB dataset

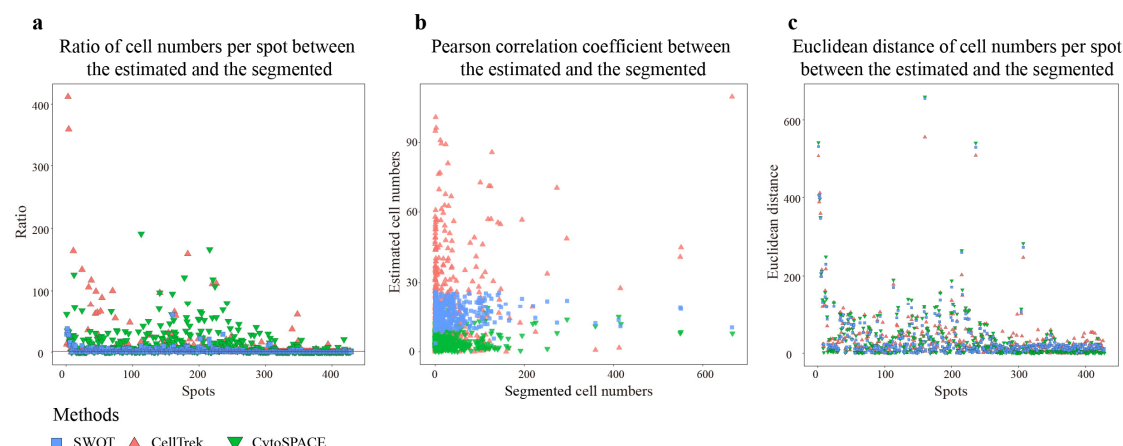


Comparison between the estimated cell numbers and the segmented cell numbers. **(a)** The ratio between the estimated and the segmented cell numbers for each spot. **(b)** The Pearson correlation coefficient between the estimated and the segmented cell numbers across all spots. **(c)** The Euclidean distance between the estimated and the segmented cell numbers for each spot. The estimated cell numbers were obtained by three methods: SWOT, CellTrek, and CytoSPACE. The segmented cell numbers were obtained through cell-segmentation of H&E image. Each dot in the plots represents a single spot, with colors and shapes denoting different methods. In (a), the straight line in each plot indicates a ratio value of 1, representing the best estimate of cell numbers. The average ratio of SWOT, CellTrek, and CytoSPACE were 2.568, 4.122, and 9.028, respectively. The average Pearson correlation coefficient of SWOT, CellTrek, and CytoSPACE were 0.4, -0.203, and 0.346, respectively. The average Euclidean distance of SWOT, CellTrek, and CytoSPACE were 34.425, 38.777, and 37.67, respectively.

In revised manuscript, we have added evaluations for PDAC dataset in subsection “SWOT facilitates annotation of tissue cellular neighborhoods functions” of the Results section.

The comparison of the estimated cell numbers with the segmented cell numbers by cell-segmentation illustrated the reliable of SWOT (Figure S25).

Figure S25. Estimated cell numbers per spot on PDAC dataset



Comparison between the estimated and the segmented cell numbers. **(a)** The ratio between the estimated and the segmented cell numbers for each spot. **(b)** The Pearson correlation coefficient between the estimated and the segmented cell numbers across all spots. **(c)** The Euclidean distance between the estimated and the segmented cell numbers for each spot. The estimated cell numbers were obtained by three methods: SWOT, CellTrek, and CytoSPACE. The segmented cell numbers were obtained through cell-segmentation of H&E image. Each dot in the plots represents a single spot, with colors and shapes denoting different methods. In (a), the straight line in each plot indicates a ratio value of 1, representing the best estimate of cell numbers. The average ratio of SWOT, CellTrek, and CytoSPACE were 2.426, 9.15, and 12.641, respectively. The average Pearson correlation coefficient of SWOT, CellTrek, and CytoSPACE were 0.13, 0.161, and 0.2, respectively. The average Euclidean distance of SWOT, CellTrek, and CytoSPACE were 35.759, 38.883, and 36.453, respectively.

In the revised main text, we have also supplemented the description of cell-segmentation in section “Cell-segmentation for H&E images” of Methods.

Cell-segmentation for H&E images. Since the truth cell numbers per spot are not available in real ST data, we performed cell-segmentation to evaluate the reliability of the cell number estimation in SWOT. Following established approaches^{26,62}, we performed segmentation on H&E staining image and measured the spatial positions of individual cells and the number of cells in spots from a histomorphology perspective. Specifically, we adopted the classical cell/nuclei segmentation method for H&E staining image implemented in Squidpy package⁴¹, and followed the provided tutorial (https://squidpy.readthedocs.io/en/stable/notebooks/examples/image/compute_segment_hne.html). This method applies a watershed segmentation algorithm based on morphological and geographical features of the tissue. The resulting segmentation masks enabled us to approximate cell counts within each spot and compare them against SWOT's predictions to evaluate the reliability of the estimation.

Responses to Reviewer #2's comments

Comment 1.

The method involves critical hyperparameters (e.g., bandwidth b , k -nearest neighbors k , regularization terms λ , α , ε) whose values may significantly influence the results but lack empirical justification. The manuscript does not clearly justify how these values are chosen or how sensitive the results are to these choices. Moreover, the choice of a factor of 10 to weight gene expression distances (ij) in Eq. 3 and Eq. 4 also appears arbitrary. The authors should consider demonstrate more systematic strategies (e.g., cross-validation, grid search) for tuning these hyperparameters or at least provide guidelines for default settings tested on multiple datasets.

Response: Thanks for your valuable suggestions. We have conducted additional experiments to clarify how the involved critical hyperparameters are chosen and how the sensitive of the results are affected by these choices.

The following performed analyses were based on simulated and real datasets used in this study: Simulated_Cortex, Simulated_MBaging, MOB (Mouse Olfactory Bulb), and PDAC (Pancreatic Ductal Adenocarcinoma) datasets. The Simulated_MBaging dataset is newly added⁴, we designed seven grids (250, 200, 150, 130, 75, 58, 38) and generated seven corresponding datasets with different numbers of spots (290, 426, 749, 989, 2850, 4669, 9554) to support scalability evaluation. For consistency, we refer to them as: Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000, respectively.

(1) In the revised main text, we have added a “Hyperparameters selection” subsection in the Methods section to discuss all hyperparameters.

Hyperparameters selection. To evaluate the robustness and sensitivity of SWOT, we conducted

experiments across a range of hyperparameter settings. Involved hyperparameters include: the value of k in k NN for constructing distance matrix, the pre-clustering method and its resolutions for aggregating gene expression information, the bandwidth for defining spatial weights, the strength for controlling spatial weights, and the penalty strength for computing transport plan. Additional details are available in Supplementary Methods.

SWOT initiates by constructing a k -nearest neighbor (k NN) graph to compute pairwise distances. It involves three hyperparameters (`n_neighbors_cell`, `n_neighbors_spot`, and `n_neighbors_pos`) in three distances (expression distance between cells in scRNA-seq data, expression distance between spots in ST data, and coordinates distance between spots in ST data). Experimental evaluations showed that while increasing the k -values leads to higher time consumption, the results remain robust (Figure S34). We selected `n_neighbors_cell`=5, `n_neighbors_spot`=5, and `n_neighbors_pos`=5 as the default settings.

During constructing spatial weights, SWOT requires a pre-clustering step to integrate gene expression information from ST data. No significant differences across various clustering algorithms and resolution settings indicated SWOT's robustness (Figure S35). We chose the Leiden algorithm as default and set its resolution parameter such that the number of clusters matches the number of cell types in scRNA-seq data. Another hyperparameter is the bandwidth, which determines the radius over which each central spot aggregates information from its neighboring spots. Based on performance under different bandwidth settings (Figure S36), we balanced accuracy and computational efficiency and selected 0.1 as the default value.

In optimal transport module, solving the unbalanced and structured optimal transport problem in Eq. (5) involves three hyperparameters: λ , α , and ε . We conducted a comprehensive grid search to evaluate how different combinations of these hyperparameters affect SWOT's performance (Figures S38 and S39). Based on these evaluations, we recommend setting α =0.1, λ =100.0, and ε =0.1 as the default configuration.

In the supplementary text, we have supplemented a subsection “4. Setting of hyperparameter selection” in Supplementary Methods section.

4. Setting of hyperparameter selection

The hyperparameters involved in SWOT include: the value of k in k NN for constructing distance matrix, the pre-clustering method and its resolutions for aggregating gene expression information, the bandwidth for defining spatial weights, the strength for controlling spatial weight, and the penalty strength for computing transport plan. We conducted experiments across a wide range of hyperparameter settings to investigate how each hyperparameter affects SWOT's performance.

For constructing distance matrices, SWOT begins by building a k -nearest neighbor (k NN) graph and then computing the shortest path distances between cells/spots on this graph. This process involves three hyperparameters (`n_neighbors_cell`, `n_neighbors_spot`, and `n_neighbors_pos`) in three types of distances (gene expression distance between cells in scRNA-seq data, gene expression distance between spots in ST data, and spatial coordinates distance between spots in ST data). To ensure that SWOT incorporates sufficient and informative neighbor information, we set 5 as minimum for each k , and used the floor of the logarithm of the number of cells/spots as maximum value in each dataset. By

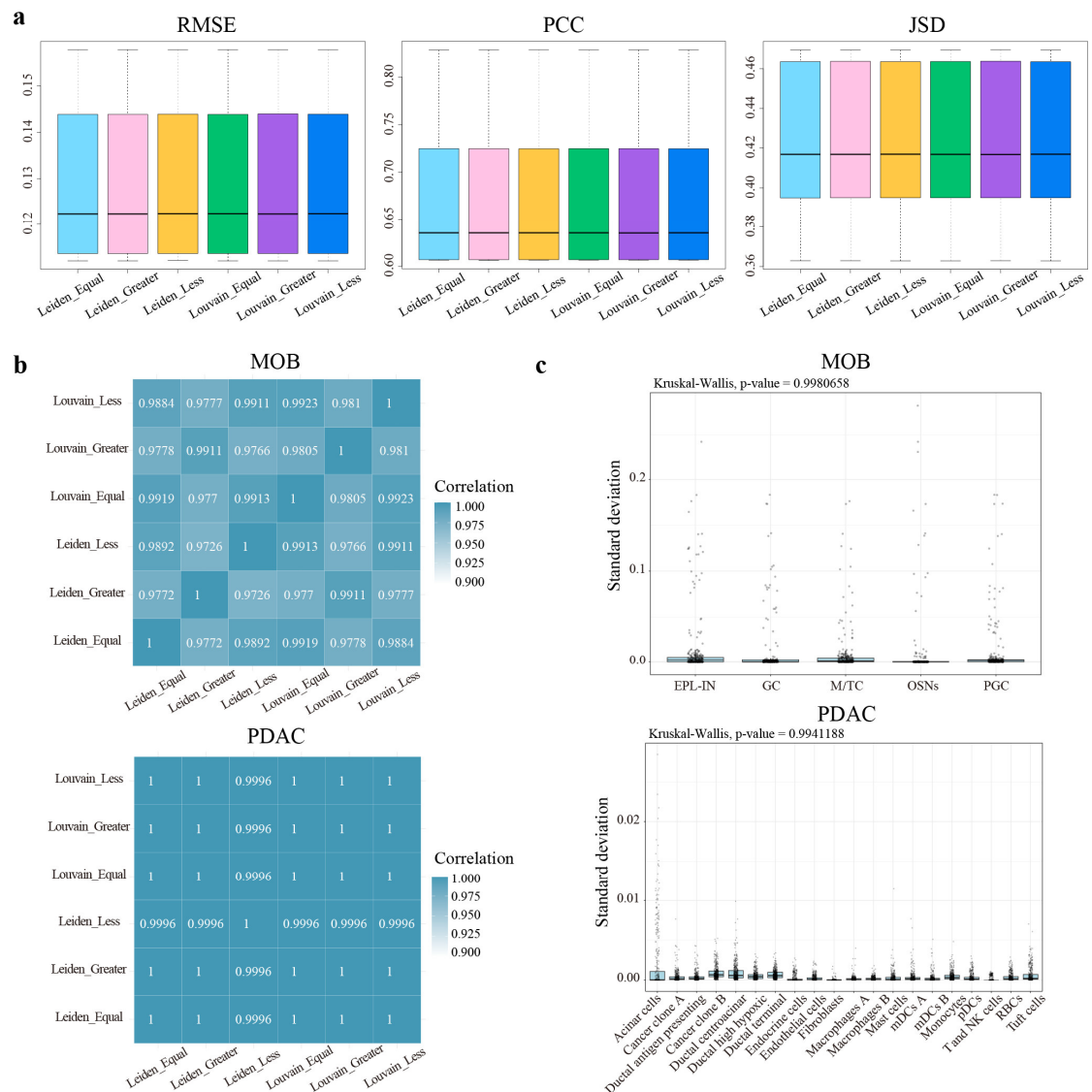
setting four distinct k -values for each of the three distances, resulting in 64 combinations per dataset, we evaluated on the accuracy and runtime across simulated and real datasets (Figure S34). Results reflected that SWOT maintained stable performance across a wide range of k -values, and increasing `n_neighbors_cell` significantly raises computational cost, as a larger k leads to a denser k NN graph. We selected `n_neighbors_cell`=5, `n_neighbors_spot`=5, and `n_neighbors_pos`=5 as default. This configuration balances computational efficiency and performance, while effectively captures the influence of first-order neighbors in both gene expression and coordinates.

To assess the impact of clustering method and resolution on SWOT's performance, we compared two widely used clustering algorithms, Louvain and Leiden, with a range of resolution parameters (Figure S35). SWOT assumes that scRNA-seq and ST data share the same set of cell types; therefore, we used the number of annotated cell types in scRNA-seq data as a reference. We validated SWOT under six conditions: where the number of clusters in ST data was (i) equal to, (ii) greater than, or (iii) less than the number of known cell types, for both clustering methods. Results demonstrated high similarity in results across all six conditions for simulated and real datasets. Statistical comparisons using the Kruskal-Wallis test revealed no significant differences, indicating that SWOT is robust to both the choice of pre-clustering algorithm and resolution parameter.

We also examined another hyperparameter involved in spatially weighted strategy, bandwidth, it determines the radius over which each central spot aggregates information from its neighboring spots. A larger bandwidth includes more distant neighbors; however, these may be less similar to the central spot, potentially introducing noise¹⁸. Therefore, we assessed how different bandwidth values affect the performance (Figure S36). We selected six bandwidth values (0, 0.1, 0.5, 1.0, 2.5, and 5.0) for validation, where $b=0$ indicates a model without spatial weighting. Across both simulated and real datasets, the accuracy, running time, correlation and Kruskal-Wallis test on SWOT demonstrated that, except for the case $b=0$, changes among bandwidth groups were minimal.

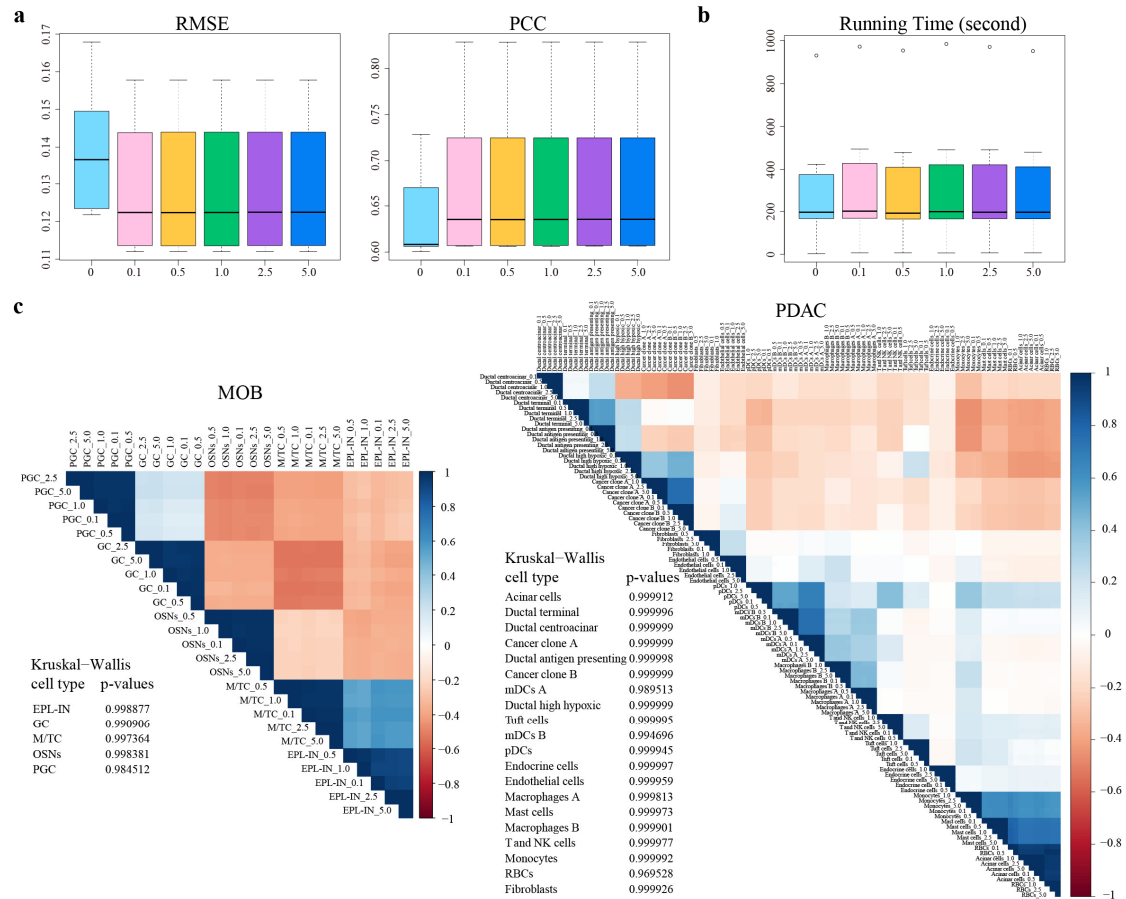
In solving the optimal transport problem, three hyperparameters are involved: λ , α , and ε , which control the penalty strengths for the unbalanced, the structural, and the entropy regularization terms, respectively. We conducted a grid search strategy over 64 combinations (4 values for each parameter) to evaluate their impact on SWOT for both simulated and real datasets (Figures S38 and S39). Based on these results, we recommend the following default values: $\alpha=0.1$, $\lambda=100.0$, and $\varepsilon=0.1$.

Figure S34. Effect of k -values in k NN for constructing distance matrix



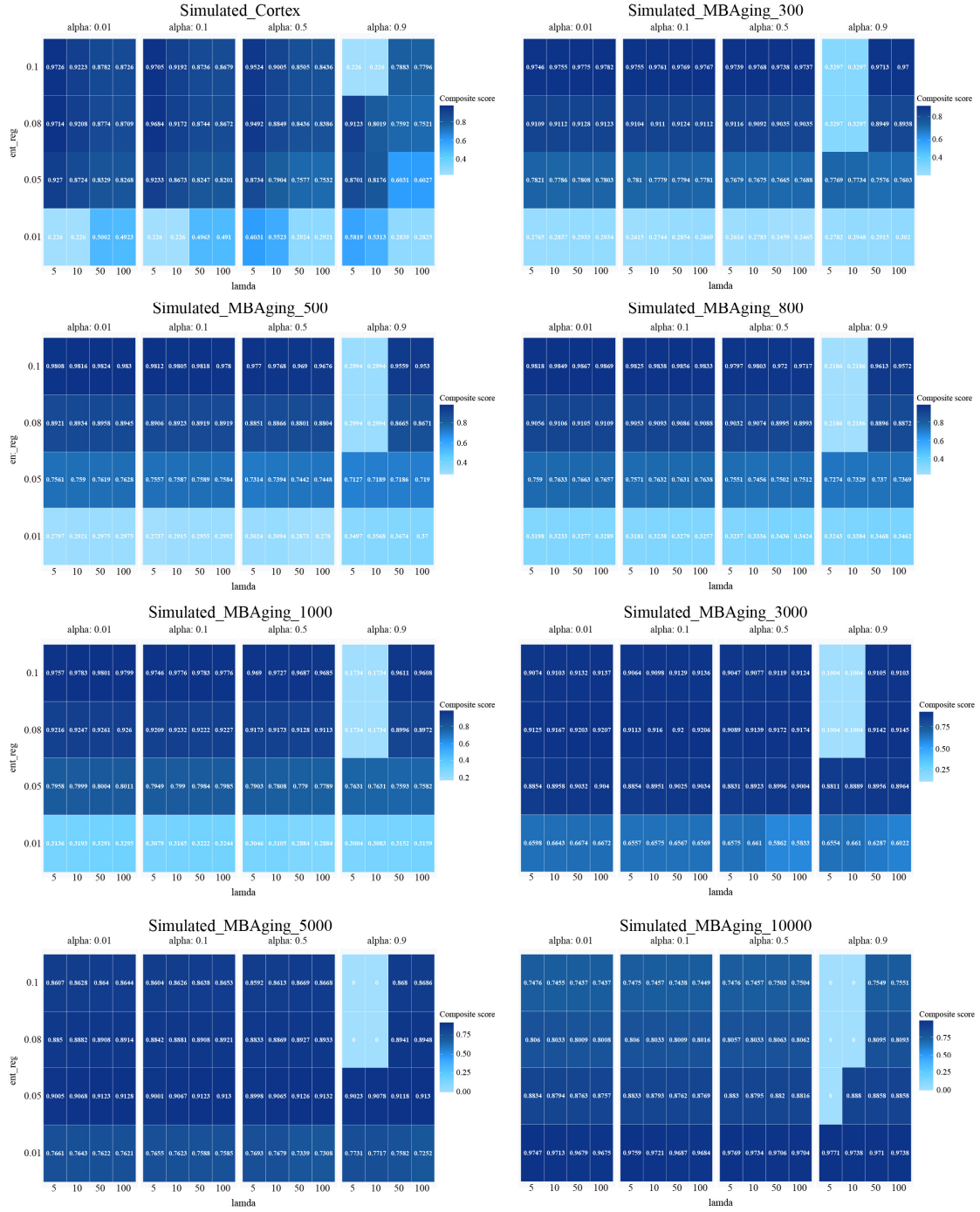
(a) Boxplots of the average RMSE, average PCC, and average JSD of the inferred cell-type compositions across all simulated datasets under each condition, with different colors representing different conditions. (b) Heatmap of the pairwise Pearson similarity of inferred cell-type composition for MOB and PDAC datasets under different settings, with more intense colors indicating higher similarity. Label in each plot is formatted as Method_Resolution (e.g., Leiden_Equal refers to using the Leiden algorithm with a resolution setting that results in the cluster numbers equal to the cell type numbers in scRNA-seq data). (c) Cell type-wise standard deviation across all spots under all conditions for MOB and PDAC datasets, with highly stable cell types represented by compact boxes. Average p-values of all cell types were marked by Kruskal-Wallis test. In (a) and (c), each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

Figure S36. Effect of bandwidth for constructing spatial weight



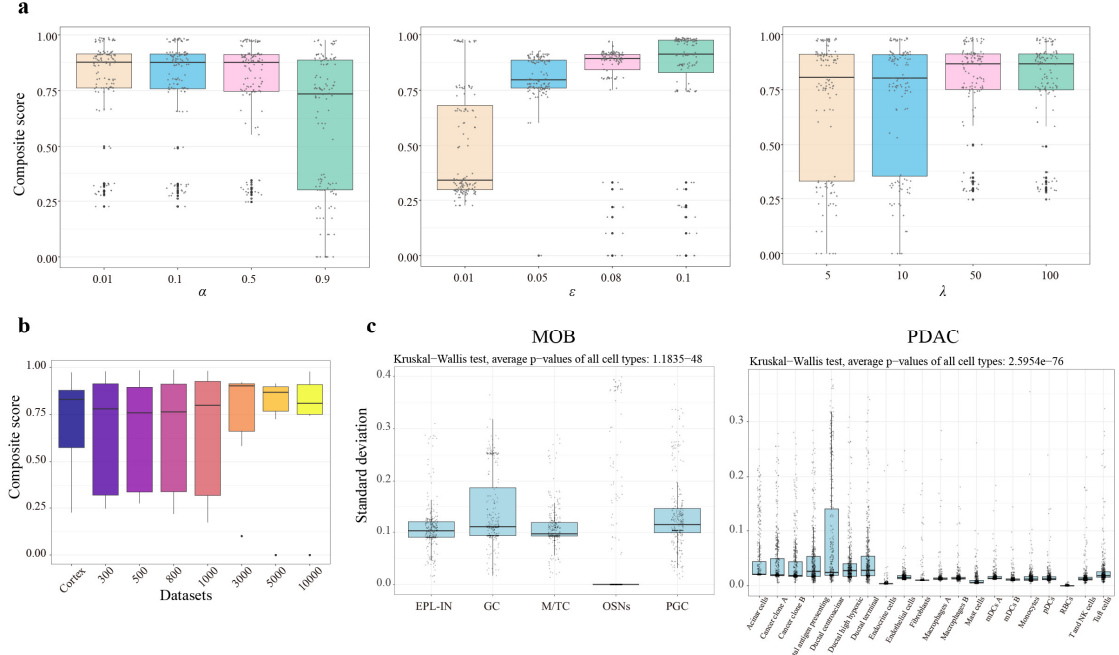
(a) Boxplots of the average RMSE, average PCC, and average JSD of the inferred cell-type compositions across all simulated datasets under each condition, with different colors representing different conditions. (b) Heatmap of the pairwise Pearson similarity of inferred cell-type composition for MOB and PDAC datasets under different settings, with more intense colors indicating higher similarity. Label in each plot is formatted as Method_Resolution (e.g., Leiden_Equal refers to using the Leiden algorithm with a resolution setting that results in the cluster numbers equal to the cell type numbers in scRNA-seq data). (c) Cell type-wise standard deviation across all spots under all conditions for MOB and PDAC datasets, with highly stable cell types represented by compact boxes. Average p-values of all cell types were marked by Kruskal-Wallis test. In (a) and (c), each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

Figure S38. Heatmap of hyperparameters for solving optimal transport problem



Heatmap of composite scores across all 64 combinations of hyperparameters λ , α , and ϵ , for all simulated datasets. The composite score is a weighted average of the normalized RMSE, PCC, Spear, and JSD metrics. Each value in the heatmap represents the composite score for a particular combination, with more intense colors indicating higher scores.

Figure S39. Effect of hyperparameters for solving optimal transport problem



(a) Boxplots of the composite score distributions across hyperparameters λ , α , and ϵ . (b) Boxplots of the composite score distributions across all simulated datasets. (c) Cell type-wise standard deviation across all spots under the 64 combinations of the three hyperparameters in MOB and PDAC datasets, and highly stable cell types are characterized by compact boxplots. Average p-values of all cell types in each dataset were marked by Kruskal-Wallis test. Each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

(2) In Eq. (3) and Eq. (4), the value “10” used to weight gene expression distance reflects two hyperparameters that control the spatial weight strength. Initially, we directly wrote “10” based on the implementation settings in our code. This indeed an oversight during manuscript writing. In the revised manuscript, we have corrected this by explicitly denoting these as hyperparameters, ρ_{indiff} and ρ_{outsam} .

We have updated the depiction in subsections “Spatially weighted strategy” and “Hyperparameters selection” of the Methods section in the revised main text.

We introduced two hyperparameters, ρ_{indiff} and ρ_{outsam} , to control the spatially weighted strength for spots inside the neighborhood but of a different cluster, and for spots outside the neighborhood but of the same cluster, respectively.

Then, a new spatial weight is defined as:

$$sw_{ij}(ps_{ij}, sp_{ij}, b, \rho) = \begin{cases} [1 - (\frac{ps_{ij}}{b})^2]^2 & ps_{ij} \leq b \text{ and } i \text{ and } j \text{ in the same cluster} \\ [1 - (\frac{ps_{ij}}{b})^2]^{2 \times (\rho_{indiff} \times sp_{ij})} & ps_{ij} \leq b \text{ and } i \text{ and } j \text{ in different cluster} \\ [1 - (\frac{sp_{ij}}{sp_{ij} + b})^2]^2 & ps_{ij} > b \text{ and } i \text{ and } j \text{ in the same cluster} \\ 0 & ps_{ij} > b \text{ and } i \text{ and } j \text{ in different cluster} \end{cases} \quad (3)$$

The redefined distance matrix between any two spots i and j is $SWD_{tec}(i, j)$ and defined as:

$$SWD_{tec}(i, j) = \begin{cases} ps_{ij} \times (1 - inter_weight) & ps_{ij} \leq b \text{ and } i \text{ and } j \text{ in the same cluster} \\ 1 / (1 + sp_{ij}^{-(inter_weight^{ps_{ij}})}) & ps_{ij} \leq b \text{ and } i \text{ and } j \text{ in different cluster} \\ ps_{ij}^{(\rho_{outsam} \times outer_weight)} & ps_{ij} > b \text{ and } i \text{ and } j \text{ in the same cluster} \\ random.uniform(0.5, 0.7) & ps_{ij} > b \text{ and } i \text{ and } j \text{ in different cluster} \end{cases} \quad (4)$$

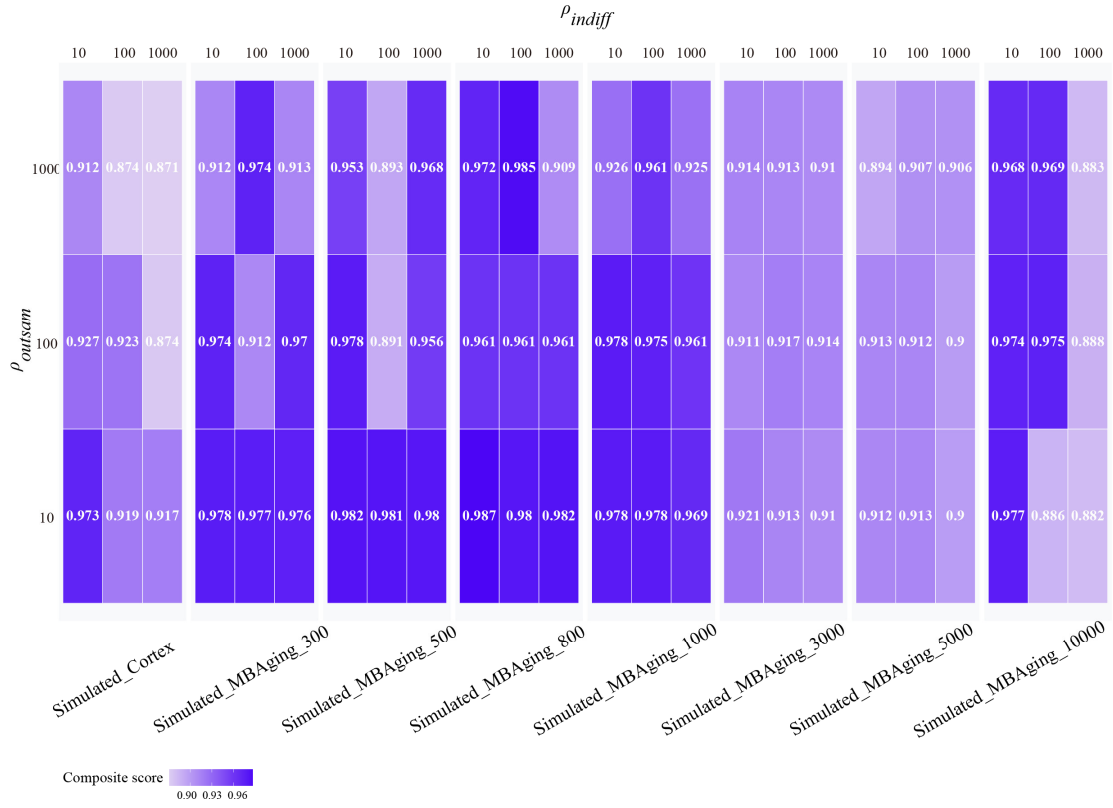
The strength in controlling spatial weight can be served as two purposes: reducing the influence of spatially adjacent spots with dissimilar gene expression on the focus spot, and enhancing the influence of spatially distant spots with similar gene expression on the focus spot. This design allows for a more nuanced integration of spatial and transcriptomic information.

The hyperparameters ρ_{indiff} and ρ_{outsam} are used to control spatially weighted strength. Results across different values suggested that setting both parameters to 10 consistently yields optimal performance across the majority of datasets (Figure S37).

In the revised supplementary text, we have also supplemented a description in subsection “4. Setting of hyperparameter selection”.

Two hyperparameters (ρ_{indiff} and ρ_{outsam}) control the spatially weighted strength. The first ensures that within the neighborhood, the weight between spots from different clusters is lower than that between spots from the same cluster. The second ensures that outside the neighborhood, the distance between spots from the same cluster is smaller than that between spots from different clusters. For all simulated datasets, we performed a grid search strategy to examine the impact of different strength combinations (10, 100, and 1000, generated nine in total) on inferred cell-type compositions (Figure S37). Experiments on the different strengths suggested that the choice of 10 yields optimal performance across the majority of datasets.

Figure S37. Effect of strength for controlling spatial weight



Heatmap of the composite scores across different settings of hyperparameters ρ_{indiff} and ρ_{outsam} on all simulated datasets. The composite score is calculated as a weighted average of the normalized RMSE, PCC, Spear, and JSD metrics. Each value in the heatmap represents the composite score for a particular combination, with more intense colors indicating higher scores.

Comment 2.

A key novelty of SWOT is the modification of the spot-to-spot distance to incorporate spatial coordinates and gene expression similarities. The approach is ad hoc and may oversimplify the tissue structures. It needs sensitivity analysis to verify the robustness of this method.

Response: Thank you for your insightful comments and questions. We answered this question from the following three aspects.

(1) First, we explained why SWOT needs to incorporate spatial coordinates and gene expression when computing spot-to-spot distance.

In ST data analysis, spatial information is essential for capturing the local microenvironmental context. Due to the presence of spatial autocorrelation, spatially adjacent spots tend to exhibit similar gene expression profiles, cell-type compositions, and tissue structures^{2,24–26}. Existing cell-type deconvolution methods have demonstrated that spatial location is effective for characterizing cell-type composition and spatial distributions within homogeneous tissue regions. However, these methods have certain limitations, for example, CARD² assigns fixed weights to neighboring spots, while SONAR²⁴ depends heavily on clustering when incorporating neighborhood information. Overall, these methods encode spatial dependence between spots through spatial distance matrices, assuming that

spatial influence decays with distance. Nevertheless, when there is significant heterogeneity among neighboring spots, such as in tissues with strong layered structures, like coronal slices of the mouse brain, aggregating mismatched neighborhood information can lead to biased estimation of cell-type composition. This may occur, for instance, when dissimilar spot types are present within a neighborhood, or when spatially distant but similar spots are excluded from the neighborhood. Therefore, we considered incorporating gene expression and spatial coordinates information when computing spot-to-spot distance matrix.

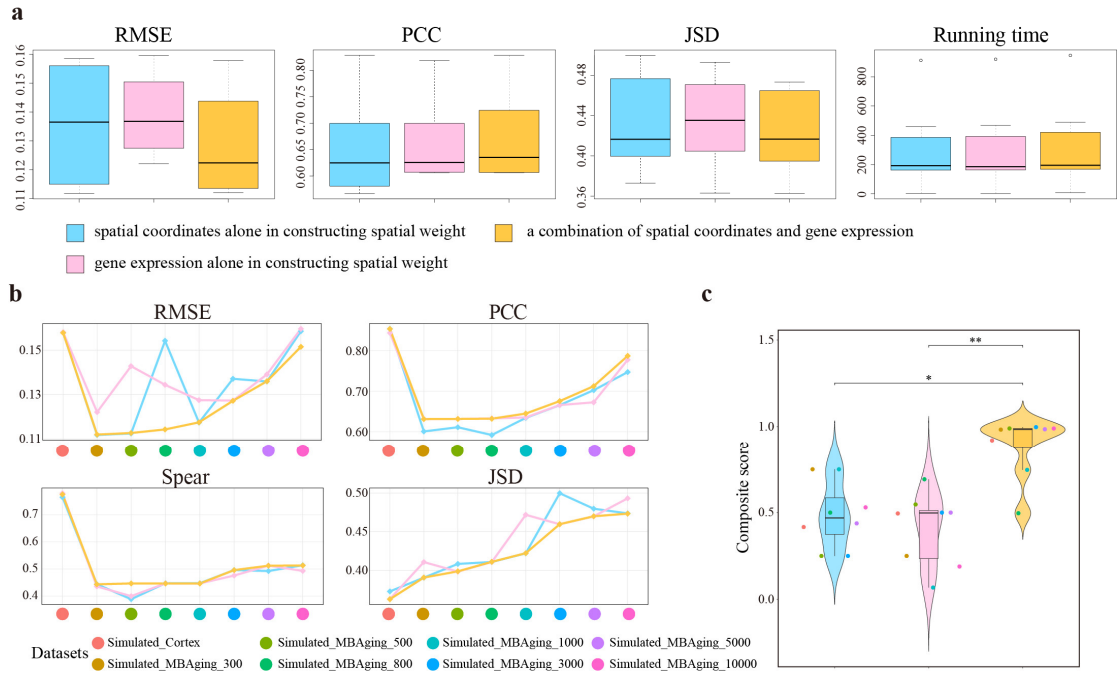
Specifically, in the optimal transport module, spot-to-spot distance focuses on how well the intra-dataset relationship is preserved by mapping²⁷. It defined by a spatially weighted strategy, and for the construction of this term, relying solely on spatial coordinates may ignore the influence of cell-type heterogeneity, while relying solely on gene expression may neglect spatial autocorrelation. Accordingly, SWOT jointly incorporated the gene expression information, derived from pre-clustering of spots, with the spatial location information, derived from the spatial neighborhood of coordinates, to uphold the intrinsic relational structure among spots in ST data as spatial weights.

(2) To demonstrate the necessary of SWOT in incorporating the two information, we constructed spatial weights (i) using spatial coordinates alone, (ii) using gene expression alone, and (iii) using both spatial coordinates and gene expression, respectively, to systematically compared the inferred cell-type composition results.

In the revised manuscript, we have provided a description in the Discussion section.

For spatially weighted strategy, relying solely on spatial coordinates may ignore the influence of cell-type heterogeneity, while relying solely on gene expression may neglect spatial autocorrelation. By incorporating gene expression and spatial neighborhood, this strategy considers the heterogeneity of spatial weights in different neighbors while preserving the spatial relationships among spots, assigning different weights to neighboring spots with varying similarities. We also illustrated the necessity of this dual incorporation by comparing it with two alternative settings that construct spatial weights using either spatial coordinates alone or gene expression alone (Figure S30).

Figure S30. Necessity of dual incorporation in spatially weighted strategy

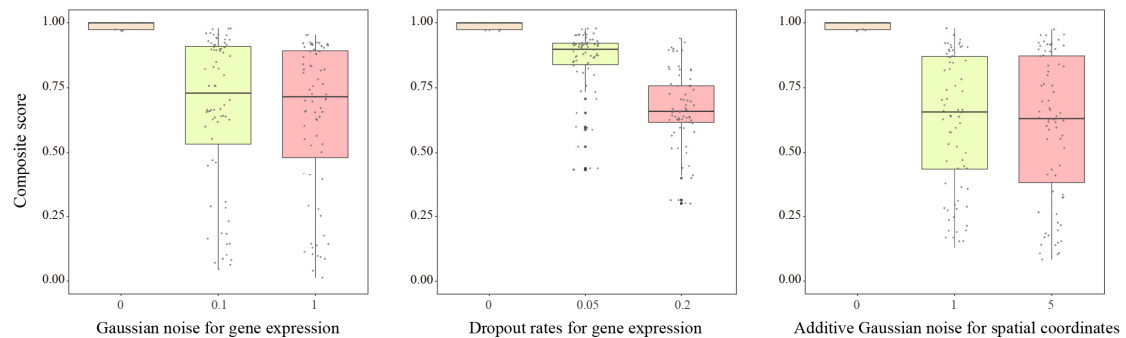


We evaluated the performance of SWOT under three settings that construct spatial weight: (i) using spatial coordinates alone, (ii) using gene expression alone, and (iii) combining both spatial coordinates and gene expression. Across all simulated datasets, we compared the accuracy and runtime for inferring cell-type composition. Boxplots in (a) summarized the RMSE, PCC, JSD, and Running time across all datasets. Line charts in (b) illustrated the average values of RMSE, PCC, Spear (Spearman correlation Coefficient), and JSD for each dataset. Among these metrics, lower RMSE and JSD values indicate better performance, whereas higher PCC and Spear values are preferred. We also exhibited the distribution of composite scores across datasets under each condition, and each point denotes a dataset (c). The composite score was defined as a weighted average of the four normalized metrics above. Since RMSE and JSD are negatively correlated with accuracy, and PCC and Spear are positively correlated, we adjusted the directionality of RMSE and JSD so that a higher composite score consistently reflects better performance. Statistical significance was assessed using pairwise two-sided Wilcoxon tests, with p-values adjusted via the Benjamini-Hochberg method, and asterisks denote statistical significance (* $p < 0.05$, ** $p < 0.01$). In (a) and (c), each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

(3) To conduct the sensitivity analysis, we introduced noise to verify the robustness of SWOT, and added a description in the Discussion section.

Additionally, we showed SWOT's robustness under multimodal noise perturbations (Figure S31).

Figure S31. Effect of noise in gene expression and spatial coordinates for SWOT



For the gene expression profiles, we incorporated Gaussian noise to simulate technical variability such as transcript capture efficiency, as well as dropout noise to mimic the dropout phenomenon commonly observed in real transcriptomics data. Specifically, we set the scale of Gaussian noise to 0.1

and 1.0, and the dropout rates to 0.05 and 0.2. For spatial coordinates, we applied additive Gaussian noise aiming to simulate potential tissue deformation and spatial misalignment during cell localization. Across all simulated datasets, we systematically assessed the impact of these noise perturbations on the performance of SWOT in inferring both cell-type compositions. The distributions of the composite score in each noise were visualized. The composite score was defined as a weighted average of normalized RMSE, PCC, Spear, and JSD. In each plot, “0” denotes the results obtained from the original, unperturbed dataset, serving as a baseline for comparison. Each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, and whiskers extending 1.5 times the interquartile range.

Comment 3.

The manuscript simulated data based on single-cell resolution ST, but the Methods section does not clearly describe how the spots were generated, or how realistic gene expression characteristics and dropout events were simulated, or whether variability across cells was effectively reproduced. More details on simulation pipelines are needed.

Response: We appreciate your insightful comments. We have clearly described the generation of the spot-resolution simulated data in the revised manuscript.

The simulation strategy in SWOT is inspired by several previous studies^{5–11}. They utilized existing single-cell ST data as a foundation, which inherently contain realistic spatial location information and gene expression characteristics, and simulated spot-resolution ST data by establishing a spatial grid division over the spatial location of individual cells. This process enables the generation of synthetic spot-resolution data while preserving critical biological characteristics observed in real single-cell ST datasets, such as sequencing depth, transcript detection rates, dropout events, and intercellular variability. Consequently, the simulated data is close to real ST profiles and suitable for evaluation.

In the revised main text, the updated description of simulated data generation was shown in subsection “Dataset description” of the Methods section, and we also updated Figure 2(a) to illustrate the generation in the Results section.

Dataset description. Following the existing simulated data generation strategy^{36,59–61}, we constructed simulated datasets by employing single-cell resolution ST data. According to the spatial coordinates of single-cell ST data, we conducted a spatial grid partitioning to generate spot-resolution ST data. This process enables the reservation of critical biological characteristics observed in real single-cell ST datasets, such as the realistic spatial structures, gene expression characteristics, sequencing depth, transcript detection rates, dropout events, or intercellular variability. The topology structure and spatial distribution of the original data determine the number of simulated spots, and the size of grid division determines the average number of cells per spot. For Simulated_Cortex dataset, a 200×200 grid generated 85 spots, and each containing an average of six cells, which is in line with the requirements of commonly used sequencing techniques. For Simulated_MBAGing datasets, we designed seven grids (250, 200, 150, 130, 75, 58, 38) and generated seven corresponding datasets with different number of spots (290, 426, 749, 989, 2850, 4669, 9554) to support scalability evaluation. For each simulated spot, the spatial coordinates and gene expression were determined by the average coordinates and the summed expressions of all cells within that spot, respectively. The cell-type

compositions, cell numbers per spot, and spatial coordinates per cell were established as the ground-truth.

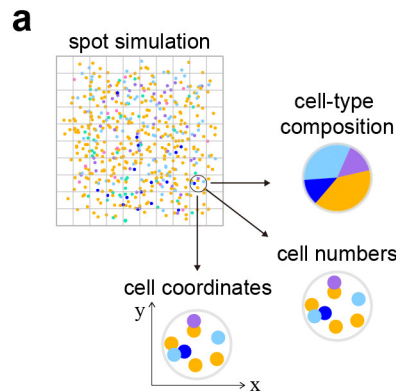


Figure 2. Performance evaluation for cell-type composition inference on simulated datasets. **a** Spot-resolution simulated ST data generation. The cell-type composition, cell numbers, cell coordinates are established as the ground-truth. Circles denote spots, points denote cells, and colored by cell types.

In the revised Supplementary text, we have modified the depiction in subsection “2.1 Simulated datasets” of the Supplementary Methods section.

2.1 Simulated datasets

For Simulated_Cortex dataset, we selected the cortex region from the sequenced mouse brain tissue by seqFISH+ technology. The single-cell resolution seqFISH+ data consisted of 523 cells expressing across 10,000 genes, and cells were annotated into six cell types based on canonical marker genes: iNeuron (inhibitory neurons), Microglia (microglia cells), endoMural (endothelial-mural cells), eNeuron (excitatory neurons), Astrocytes, and Olig (Oligodendrocytes)¹⁴. We generated a simulated spot-resolution ST dataset by applying a 200 spatial grid, which yielded 85 spots with an average of six cells per spot.

For the Simulated_MBAGing datasets, we selected a snRNA-seq data from 4-week-old stage mouse brain aging tissue¹⁵ as a reference, and a single-cell resolution MERFISH data to generate simulated spot-resolution ST datasets. We filtered the MERFISH data based on the following conditions: 'donor_id=MsBrainAgingSpatialDonor_1', 'development_stage=4-week-old stage', and 'slice=0', resulting in 21,002 cells expressed on 347 genes, and labeled for 12 cell types: Astro (Astrocytes), InInter (inhibitory interneuron), Micro (microglia), Oligo (oligodendrocytes), Neuron, Endo (endothelial cells), MSN (Medium spiny neurons), OPCs (oligodendrocyte precursor cells), Peri (pericytes), Macro (macrophages), VLMCs (vascular leptomeningeal cells), and T cells. To evaluate the performance under varying spatial resolutions, we applied seven grid sizes (250, 200, 150, 130, 75, 58, 38) to generate seven Simulated_MBAGing datasets with different numbers of spots (290, 426, 749, 989, 2850, 4669, 9554). For consistency, we refer to them as: Simulated_MBAGing_300, Simulated_MBAGing_500, Simulated_MBAGing_800, Simulated_MBAGing_1000, Simulated_MBAGing_3000, Simulated_MBAGing_5000, and Simulated_MBAGing_10000, respectively. The cell types in the snRNA-seq and MERFISH data were annotated based on the previously known cell-type marker genes that define major neuronal, glial, and immune cell types and genes previously reported to be highly upregulated in various cell types during aging¹⁵.

Comment 4.

The authors claim SWOT estimates the cell numbers per spot, but the estimation procedure is missing in the manuscript. Based on the GitHub repository

(https://github.com/GaoLabXDU/SWOT/blob/main/SWOT/utis/cell_mapping.py at line 245), it seems to be technology specific and limited to 10X Visium, Spatial Transcriptomics, and Slide-seq. It compares each spot's zero expression proportion to a dataset-wide average, when a spot's zero ratio exceeds the average, it is assigned with a small number of cells; otherwise it will be assigned with a larger number of cells. It essentially assumes dropouts inversely correlate with the number of cells captured in a spot. However, according to Fig. S3, the inverse trend doesn't always hold, this assumption should be validated on more datasets with varying dropout rates.

Response: Thanks for your valuable feedback and suggestions. We answered the issue of estimating cell numbers per spot from the following two aspects.

(1) We gave an explanation and have added an estimation procedure in the revised manuscript.

Inspired by observations from previous studies^{20,21} indicating an inverse relationship between the proportion of zeros and the average total count per location in ST data, we hypothesized that spots with a higher proportion of zero expression values tend to contain fewer cells, and vice versa. Based on this principle, we established a log-linear model to represent the inverse relationship between the cell numbers and the zero proportions. This relationship was validated in simulated datasets and generalized well across diverse scenarios (Figure S3).

Although a general inverse trend was observed, the relationship does not hold consistently across all spots due to heterogeneity in spatial expression and capture efficiency. Moreover, different ST technologies differ in resolutions and expected cell numbers per spot. To account for this variability, we further refined the estimated number by adjusting them according to the specific technology used or the expected cell number ranges.

It is important to emphasize that SWOT is not specific to any particular ST technologies and is compatible with a wide range of technologies beyond 10x Visium, Spatial Transcriptomics, and Slide-seq. Given that setting identical minimum and maximum cell number ranges across different platforms is inappropriate, SWOT incorporates the typical cell number ranges specific to each sequencing technology. For these three commonly used technologies (10x Visium, Spatial Transcriptomics, and Slide-seq), we provided predefined ranges. In addition, SWOT included a 'user-defined' option, allowing users to specify custom values for the minimum (cells_num_min) and maximum (cells_num_max) number of cells per spot based on known information. This flexibility enables more accurate and adaptable cell number estimation across varying ST technologies and datasets.

In the revised manuscript, the estimation procedure has been added in the Results and Methods sections.

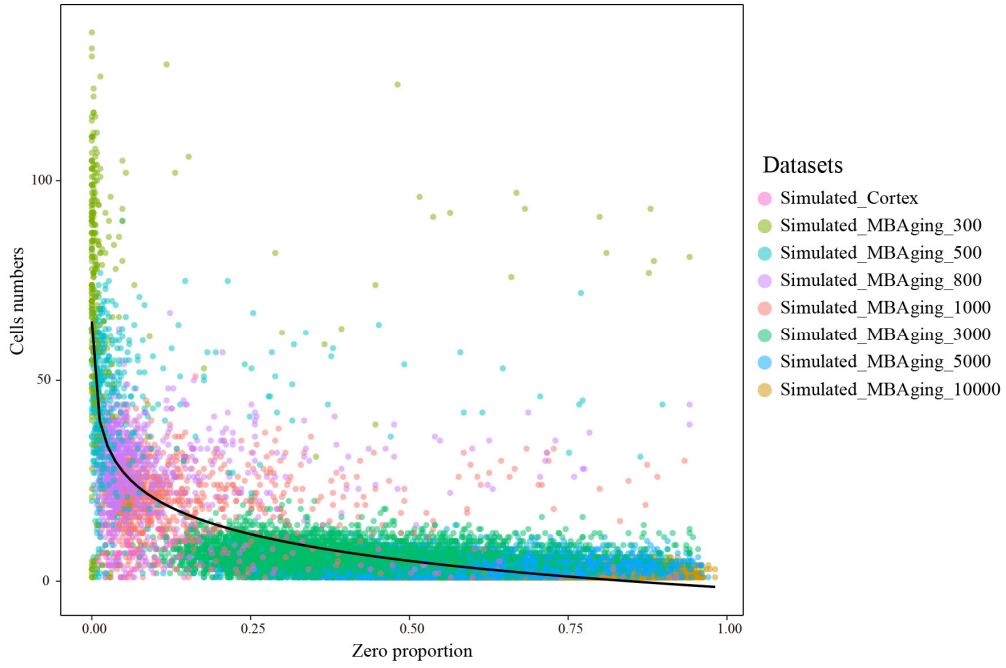
[Given that the number of cells captured per spot varies across different technologies³³, and that the proportion of zero values in expression within a spot tends to be inversely proportional to the cell](#)

numbers within that spot^{33,34} (Figure S3), SWOT employs a log-linear model to estimate the cell numbers per spot.

Estimation of cell numbers per spot. Inspired by the inverse relationship between the proportion of zeros and the average total count per location in ST data^{33,34}, we hypothesized that spots with a higher proportion of zero expression values tend to contain fewer cells, and vice versa (Figure S3). Then, SWOT estimates cell numbers per spot based on the zero proportions in gene expression and the sequencing technology that generated ST data or the predefined ranges of the expected cell numbers per spot. For each spot i , the proportion of zero values in its expression is z_i , the estimated cell numbers in spot i is c_i , and it is defined as $c_i = -10 \cdot \ln(z_i + 0.001) - 1$.

The established relationship is a log-linear model, and it describes the general inversely relationship between the zero proportions and the cell numbers per spot. Although a general inverse trend was observed, the approximation relationship does not hold consistently across all spots due to heterogeneity in spatial expression and capture efficiency. Different sequencing technologies differ in spatial resolution and expected cell numbers ranges per spot, such as 10x Genomics Visium typically captures 1–10 cells per 55 μm spot. The spatial resolution of Spatial Transcriptomics technology is 100 μm , and capturing 10–40 cells⁹. Slide-seq, characterized by a spot diameter of 10 μm , typically includes 1–3 cells^{10,33,58}. To account for this variability, we implement a post-estimation adjustment to refine the estimation according to the used technology or the expected ranges. Specifically, if a spot's estimated cell number falls outside the expected range, the estimate is adjusted by randomly assigning a value within the platform-specific range. We provide predefined expected ranges for the aforementioned three commonly platforms but do not limit to them. A 'user-defined' mode is also available, allowing users to specify custom values for the minimum and maximum number of cells per spot based on platform specifications or known knowledges. This strategy enables a flexible and adaptive estimation, ensures that each spot is assigned a biologically plausible number of cells, reduces overreliance on the log-linear assumption, and maintains consistency with the characteristics of the underlying technology.

Figure S3. Relationship between cell number and zero proportion per spot



We utilized all simulated datasets (Simulated_Cortex, Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000) to investigate the relationship between the proportion of zero values in gene expression within a spot and the number of cells present in that spot. As the true cell numbers per spot were known in all simulations, we extracted the cell numbers in each spot (y-axis) and calculated the corresponding proportion of zero values in gene expression within that spot (x-axis). Scatter plots were generated to visualize this relationship across all spots, and the black line is the fitting curve of the log-linear model.

(2) Then, we have validated the established relationship using simulated and real datasets in the revised manuscript.

(2.1) The validations and descriptions on simulated datasets, with varying dropout rates, were shown in subsection “Performance evaluation of SWOT on inferring single-cell spatial maps” of the Results section.

First, in each simulated dataset, we examined the cell number estimation strategy using three quantitative metrics: (1) the ratio between the estimated and the ground-truth cell numbers per spot, (2) the Pearson correlation coefficient between the estimated and the ground-truth values across all spots, and (3) the Euclidean distance between the estimated and the ground-truth cell numbers per spot. In the Simulated_Cortex dataset, SWOT obtained average values of 1.581 for the ratio, 0.27 for the correlation, and 2.776 for the distance, indicating the accuracy of its cell number estimation (Figure 4a). Across all Simulated_MBaging datasets, SWOT also achieved competitive average results in these metrics (Figure 4b). The considerable performance of SWOT across Simulated_MBaging with different dropout rates further supported its robustness in estimating cell numbers (Figure S15 and Table S2). Notably, CellTrek tended to overestimate the cell numbers, while CytoSPACE tended to underestimate it. Second, for the estimated spatial coordinates, we calculated a Euclidean distance between the estimated and ground-truth coordinates for SWOT and CellTrek.

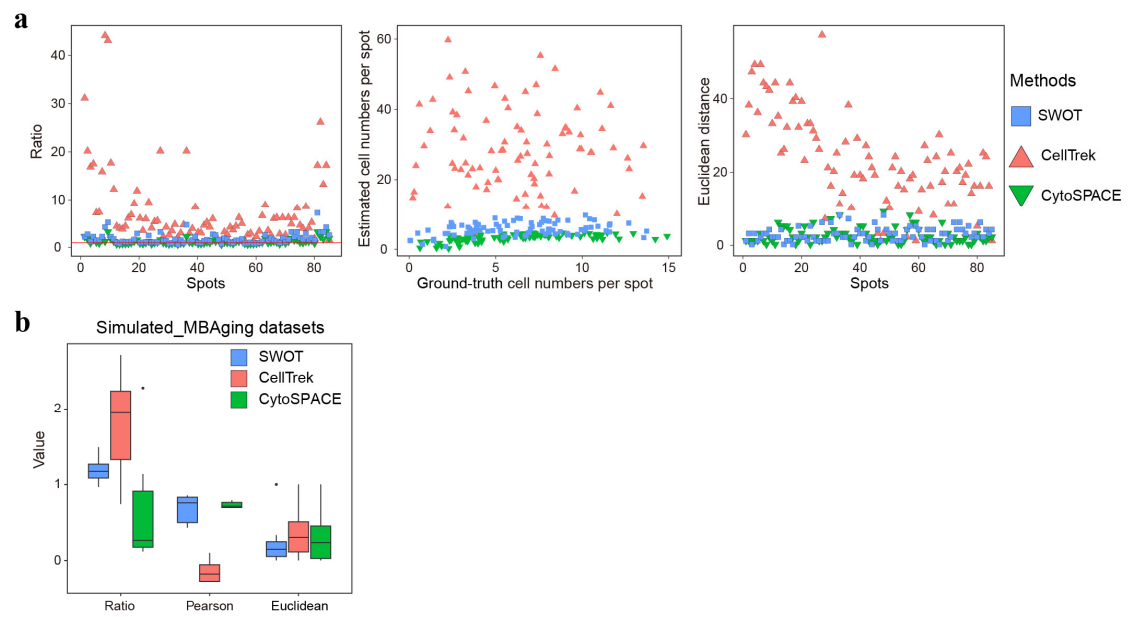
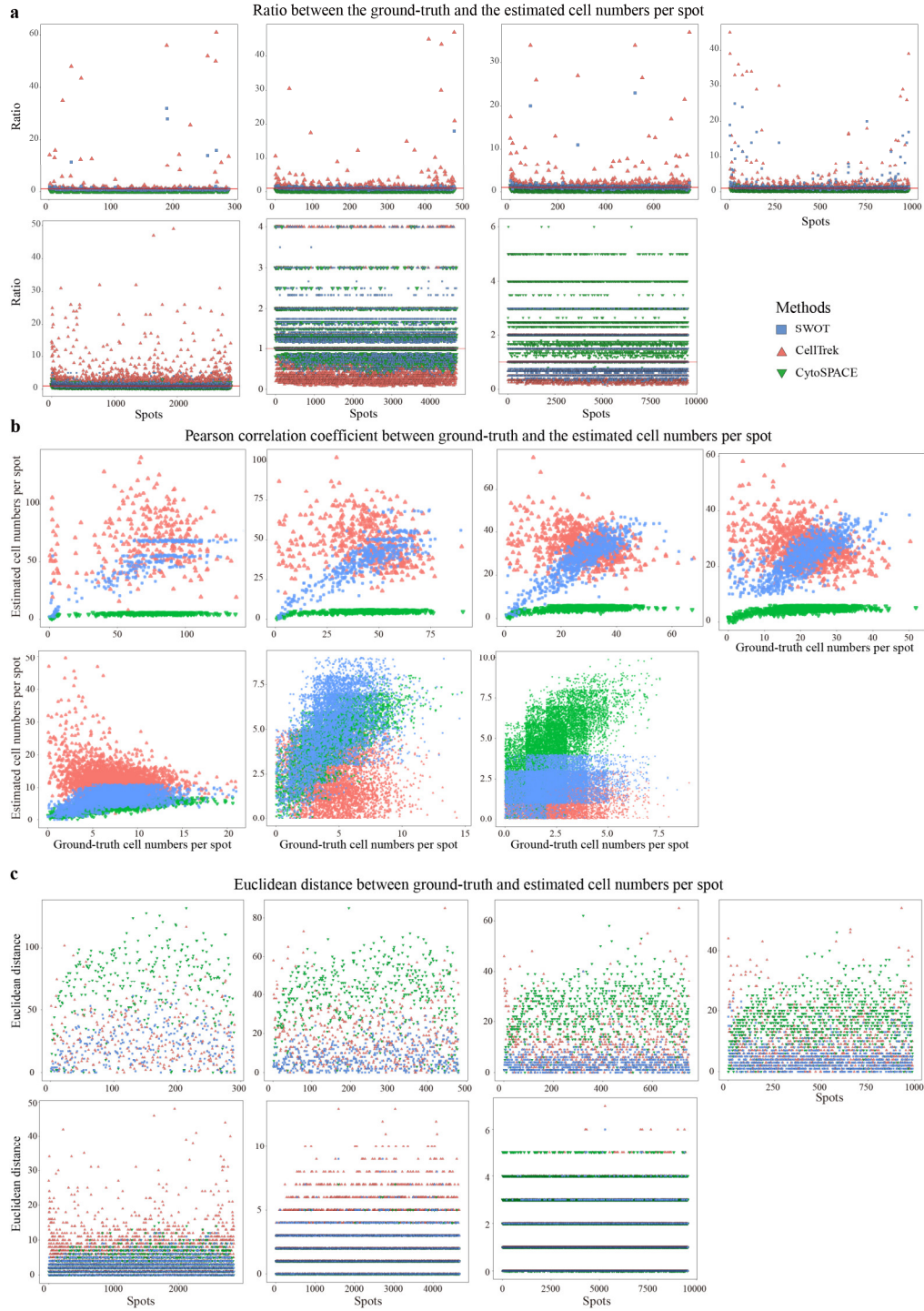


Figure 4. Performance evaluation for single-cell spatial maps inference on simulated datasets. **a** Performance of cell number estimation strategy in Simulated_Cortex dataset. Left: Ratio between estimated and ground-truth cell numbers per spot. Middle: Pearson correlation coefficient between estimated and ground-truth cell numbers across all spots. Right: Euclidean distance between estimated and ground-truth cell numbers per spot. Colors and shapes denote methods, and the straight line in the left panel indicates a ratio of 1, representing the best estimation. **b** Boxplots of the average values for the ratio, Pearson correlation coefficient, and Euclidean distance between estimated and ground-truth cell numbers across all Simulated_MBaging datasets.

Figure S15. Estimated cell numbers per spot on simulated datasets with different dropout rate



Seven Simulated_MBaging datasets (Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000) were generated under different spatial resolutions and exhibited distinct dropout rates of 0.0543, 0.0586, 0.0927, 0.1303, 0.3656, 0.5025, and 0.6781, respectively. (a) The ratio between estimated and ground-truth cell numbers for each spot. (b) The Pearson correlation coefficient between estimated and ground-truth values across all spots. (c) The Euclidean distance between estimated and ground-truth for each spot. In each subplot, all Simulated_MBaging datasets are arranged from left to right, top to bottom. Each dot in the plots represents an individual spot, with colors and shapes denoting methods. In (a), the straight line in each plot indicates a ratio value of 1, representing the best estimate of cell numbers. The average values of the three metrics (ratio, Pearson correlation coefficient, and Euclidean distance) across all spots for each dataset were summarized in Table S2.

Table S2. Evaluation of the estimated cell numbers per spot

For all Simulated_MBaging datasets, we compared the estimated cell numbers per spot by three methods (SWOT, CellTrek, and CytoSPACE) against the ground-truth using three metrics. We list the average ratio between estimated and ground-truth cell numbers per spot, the average Pearson correlation coefficient between estimated and ground-truth across all spots, and the average Euclidean distance between estimated and ground-truth values for each spot. The closer the ratio is to 1, the greater the correlation, and the smaller the distance, indicating a better estimation.

Metric	Method	Simulated_MBaging dataset						
		300	500	800	1000	3000	5000	10000
Ratio	SWOT	1.23	0.96	1.094	1.491	1.075	1.305	1.172
	CellTrek	2.714	1.948	1.854	2.071	2.403	0.742	0.801
	CytoSPACE	0.118	0.143	0.206	0.264	0.689	1.136	2.279
Pearson correlation coefficient	SWOT	0.812	0.856	0.852	0.758	0.432	0.553	0.443
	CellTrek	0.098	-0.18	-0.28	-0.282	-0.272	-0.118	0.004
	CytoSPACE	0.688	0.695	0.697	0.714	0.768	0.792	0.759
Euclidean distance	SWOT	23.183	8.219	3.979	4.256	2.206	1.526	0.708
	CellTrek	28.897	19	11.401	9.496	5.649	2.662	1.059
	CytoSPACE	67.645	39.098	23.346	16.621	2.939	0.964	2.304

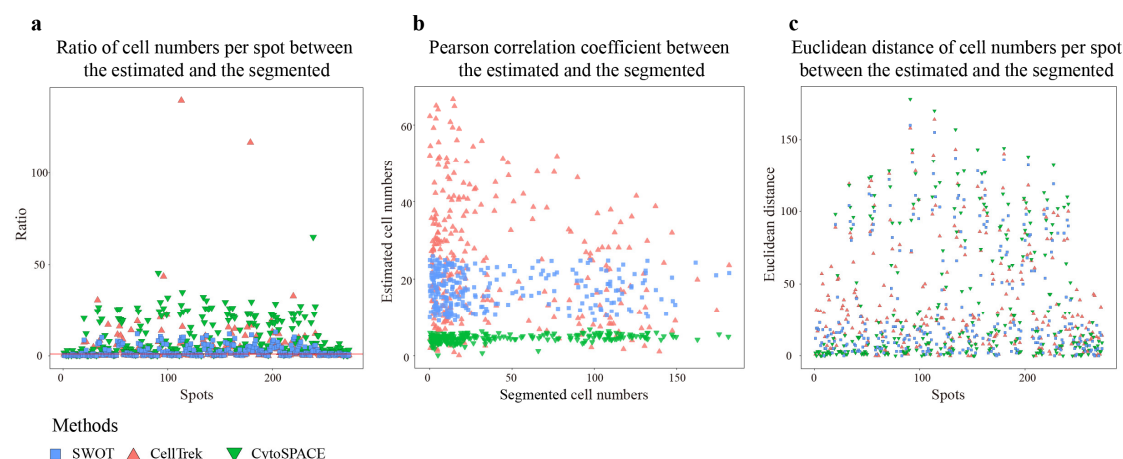
(2.2) We also evaluated the reliability cell number estimation method on real datasets (MOB and PDAC) in the revised manuscript.

Since the truth cell numbers, are not available in real ST data, we followed the approach used in prior studies such as CytoSPACE²² and Tangram²³, and performed cell-segmentation on the H&E-stained histological images. Although segmentation results cannot be regarded as a gold standard, they provided a histomorphology perspective in measuring the spatial positions of individual cells and the number of cells. We compared the cell numbers derived from the segmentation with estimated by SWOT, CellTrek, and CytoSPACE.

In subsection “SWOT depicts cell-type spatial distributions” of the Results section:

Since the truth cell numbers are not available in real ST data, to evaluate the reliability of the cell number estimation strategy employed by SWOT, we compared the estimated cell numbers per spot by these three methods with the segmented cell numbers obtained through cell-segmentation⁴¹ from H&E image (Figure S14). Based on the segmented numbers, SWOT showed superior performance, achieving a spot-wise ratio close to 1, the highest correlation, and the lowest Euclidean distance.

Figure S14. Estimated cell numbers per spot on MOB dataset

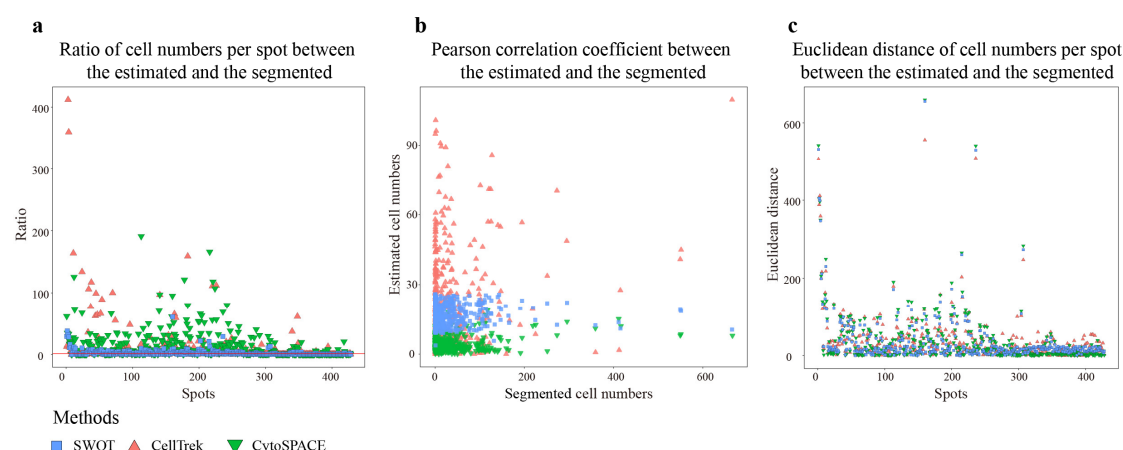


Comparison between the estimated cell numbers and the segmented cell numbers. **(a)** The ratio between the estimated and the segmented cell numbers for each spot. **(b)** The Pearson correlation coefficient between the estimated and the segmented cell numbers across all spots. **(c)** The Euclidean distance between the estimated and the segmented cell numbers for each spot. The estimated cell numbers were obtained by three methods: SWOT, CellTrek, and CytoSPACE. The segmented cell numbers were obtained through cell-segmentation of H&E image. Each dot in the plots represents a single spot, with colors and shapes denoting different methods. In (a), the straight line in each plot indicates a ratio value of 1, representing the best estimate of cell numbers. The average ratio of SWOT, CellTrek, and CytoSPACE were 2.568, 4.122, and 9.028, respectively. The average Pearson correlation coefficient of SWOT, CellTrek, and CytoSPACE were 0.4, -0.203, and 0.346, respectively. The average Euclidean distance of SWOT, CellTrek, and CytoSPACE were 34.425, 38.777, and 37.67, respectively.

In subsection “SWOT facilitates annotation of tissue cellular neighborhoods functions” of the Results section:

The comparison of the estimated cell numbers with the segmented cell numbers by cell-segmentation illustrated the reliable of SWOT (Figure S25).

Figure S25. Estimated cell numbers per spot on PDAC dataset



Comparison between the estimated and the segmented cell numbers. **(a)** The ratio between the estimated and the segmented cell numbers for each spot. **(b)** The Pearson correlation coefficient between the estimated and the segmented cell numbers across all spots. **(c)** The Euclidean distance between the estimated and the segmented cell numbers for each spot. The estimated cell numbers were obtained by three methods: SWOT, CellTrek, and CytoSPACE. The segmented cell numbers were obtained through cell-segmentation of H&E image. Each dot in the plots represents a single spot, with colors and shapes denoting different methods. In (a), the straight line in each plot indicates a ratio value of 1, representing the best estimate of cell numbers. The average ratio of SWOT, CellTrek, and CytoSPACE were 2.426, 9.15, and 12.641, respectively. The average Pearson correlation coefficient of SWOT, CellTrek, and CytoSPACE were 0.13, 0.161, and 0.2, respectively. The average Euclidean distance of SWOT, CellTrek, and CytoSPACE were 35.759, 38.883, and 36.453, respectively.

In the revised main text, we have also supplemented a subsection “Cell-segmentation for H&E images” in the Methods section.

Cell-segmentation for H&E images. Since the truth cell numbers per spot are not available in real ST data, we performed cell-segmentation to evaluate the reliability of the cell number estimation in SWOT. Following established approaches^{26,62}, we performed segmentation on H&E staining image and measured the spatial positions of individual cells and the number of cells in spots from a histomorphology perspective. Specifically, we adopted the classical cell/nuclei segmentation method for H&E staining image implemented in Squidpy package⁴¹, and followed the provided tutorial (https://squidpy.readthedocs.io/en/stable/notebooks/examples/image/compute_segment_hne.html). This method applies a watershed segmentation algorithm based on morphological and geographical features of the tissue. The resulting segmentation masks enabled us to approximate cell counts within each spot and compare them against SWOT's predictions to evaluate the reliability of the estimation.

Responses to Minor comments

Comment 1.

R symbols in the Methods section are not readable.

Response: We appreciate your feedback and have updated the mathematical symbols in subsection “Unbalanced and structured optimal transport” of the Methods section.

Three distance matrices are: $D_{se} \in \mathbb{R}^{m \times m}$ measuring gene expression distance among cells in scRNA-seq data; $D \in \mathbb{R}^{m \times n}$ measuring gene expression distance between cells and spots; and $D_{tc} \in \mathbb{R}^{n \times n}$ measuring spatial coordinates distance among spots in ST data.

where $\langle \cdot, \cdot \rangle$ is the Frobenius inner product and $P \in \mathbb{R}^{m \times n}$ is the transport plan.

$\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^n$ are the discrete distributions of mass on the set of cells and spots.

Comment 2.

The uniform sampling of cell coordinates within spot boundaries (SWOT/SWOT/utis/cell_mapping.py at line 309) may unintentionally bias cells toward spot centers due to radial density effects.

Response: We appreciate your thoughtful review. We have recognized the issue of neglecting the radial density effects when uniformly sampling of cell coordinates. This problem has addressed in our code, and we have regenerated new cell coordinates accordingly.

In line with the spatial coordinates generation method performed in previous literature²⁸, we used the uniformly sampling strategy from the polar coordinates. However, due to the nonlinear nature of geometric space, this approach may lead to radial density effects, where sampling points were overly

concentrated in the central region and sparsely distributed in the periphery^{29,30}. Although the precise location of a cell within a spot has limited impact on the assessment of cell-type composition and spatial distributions in tissue, such subtle differences may potentially influence downstream analyses, such as the identification of spatial domains and the inference of cell-cell communications.

To mitigate this bias, we have applied a square root transformation^{31,32} to the radius in our code (https://github.com/GaoLabXDU/SWOT/blob/main/SWOT/utis/cell_mapping.py at line 195). This transformation ensures a more uniform distribution of points within the circular area. We then regenerated cell coordinates and updated the corresponding results in the revised manuscript. The updated results of SWOT in both simulated and real datasets indicated that the subtle changes in spatial coordinates have a minor impact on the inferred single-cell spatial maps.

In subsection “Estimation of spatial coordinates per cell” of the Methods, we have updated the description.

To estimate the spatial coordinates of inferred single cells, we use the original coordinates of each spot as the center and sample from a uniform distribution in polar coordinates³⁵. We further apply a square root transformation to the radial component to mitigate the radial density effects, where cells tend to cluster near the center, ensuring a more even spatial distribution of the estimated coordinates.

In the revised manuscript, we have also updated related results in Figures 3-6 of the Results section.

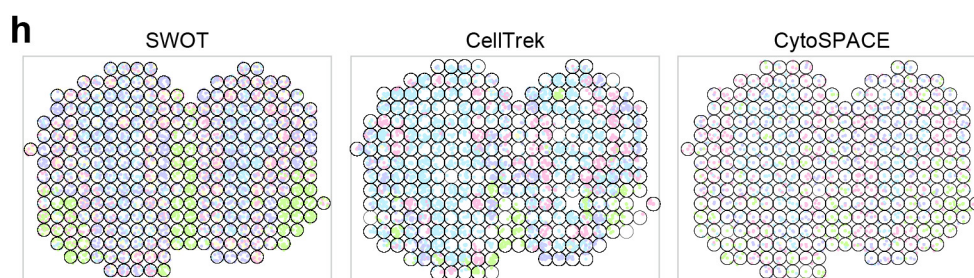


Figure 3. SWOT depicts cell-type spatial distributions on mouse olfactory bulb dataset. **h** Spatial distributions of cells in the inferred single-cell spatial maps by SWOT, CellTrek, and CytoSPACE. Circles denote spots, points denote cells, and colored by cell types.

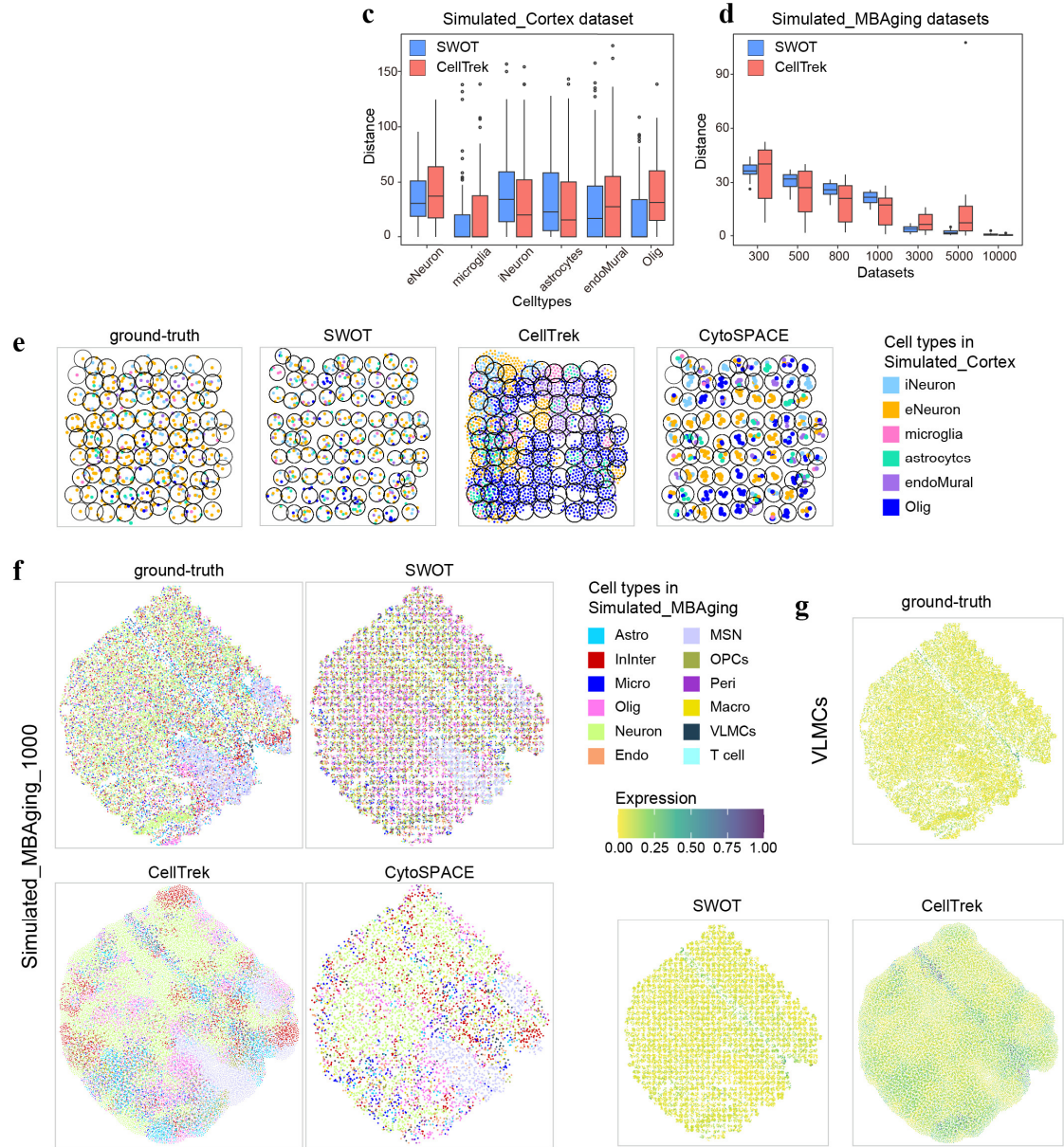


Figure 4. Performance evaluation for single-cell spatial maps inference on simulated datasets. **c** Distance between ground-truth and estimated cell coordinates of each cell type on Simulated_Cortex dataset. **d** Distance between ground-truth and estimated cell coordinates of all cell types on all Simulated_MBaging datasets. The horizontal axis represents the simplified name of the dataset. In **b-d**, each box plot is the interquartile range of the computed values, spanning from the first quartile to the third quartile, with the median as the central line, whiskers extending 1.5 times the interquartile range, and points outside are outliers. **e** Spatial distributions of cells in the inferred single-cell spatial maps generated from ground-truth, SWOT, CellTrek, and CytoSPACE on Simulated_Cortex dataset. **f** Spatial distributions of cells in the inferred single-cell spatial maps generated from ground-truth, SWOT, CellTrek, and CytoSPACE on Simulated_MBaging_1000 dataset. In **e** and **f**, circles denote spots, points denote cells, and colored by cell types. **g** Expression patterns of cell-type-specific genes on VLMCs in single-cell spatial maps from ground-truth, SWOT, and CellTrek on Simulated_MBaging_1000 dataset. Color represents the summed expression levels of top 50 cell-type-specific genes.

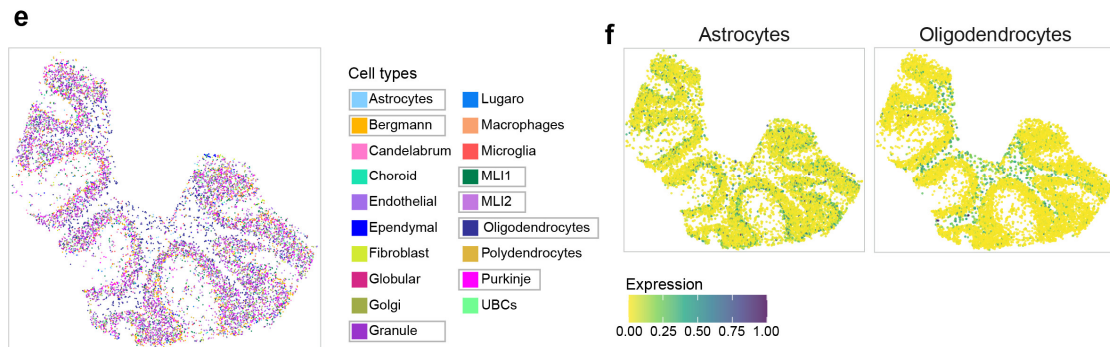
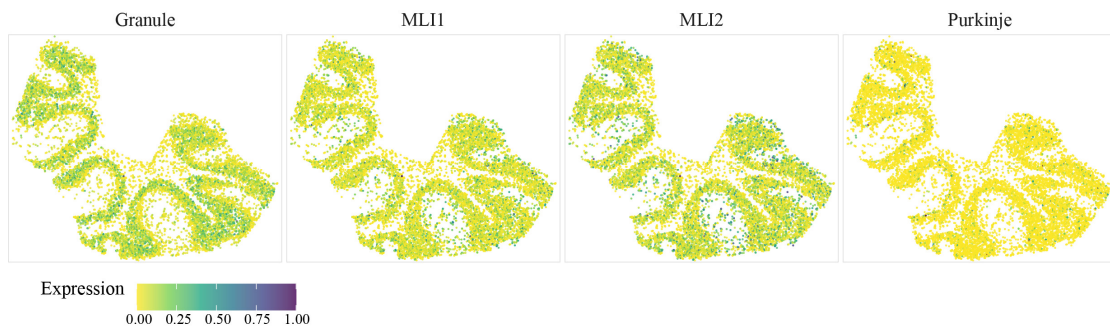


Figure 5. SWOT infers a single-cell spatial map on mouse cerebellum dataset. **e** Spatial distributions for all cell types of single cells in the inferred single-cell spatial map by SWOT. Points denote cells and colored by cell types. **f** Expression patterns of cell-type-specific genes for cell types Astrocytes and Oligodendrocytes of the single-cell spatial map inferred by SWOT. Color in **d** and **f** represents the summed expression levels of top 100 cell-type-specific genes.

Figure S20. Expression patterns of single-cell spatial map on mouse cerebellum dataset



Expression patterns of cell-type-specific genes on four cell types (Granule, MLI1, MLI, and Purkinje) of single-cell spatial maps inferred by SWOT. Color in each plot denotes the summed expression levels of the top 100 cell-type-specific genes.

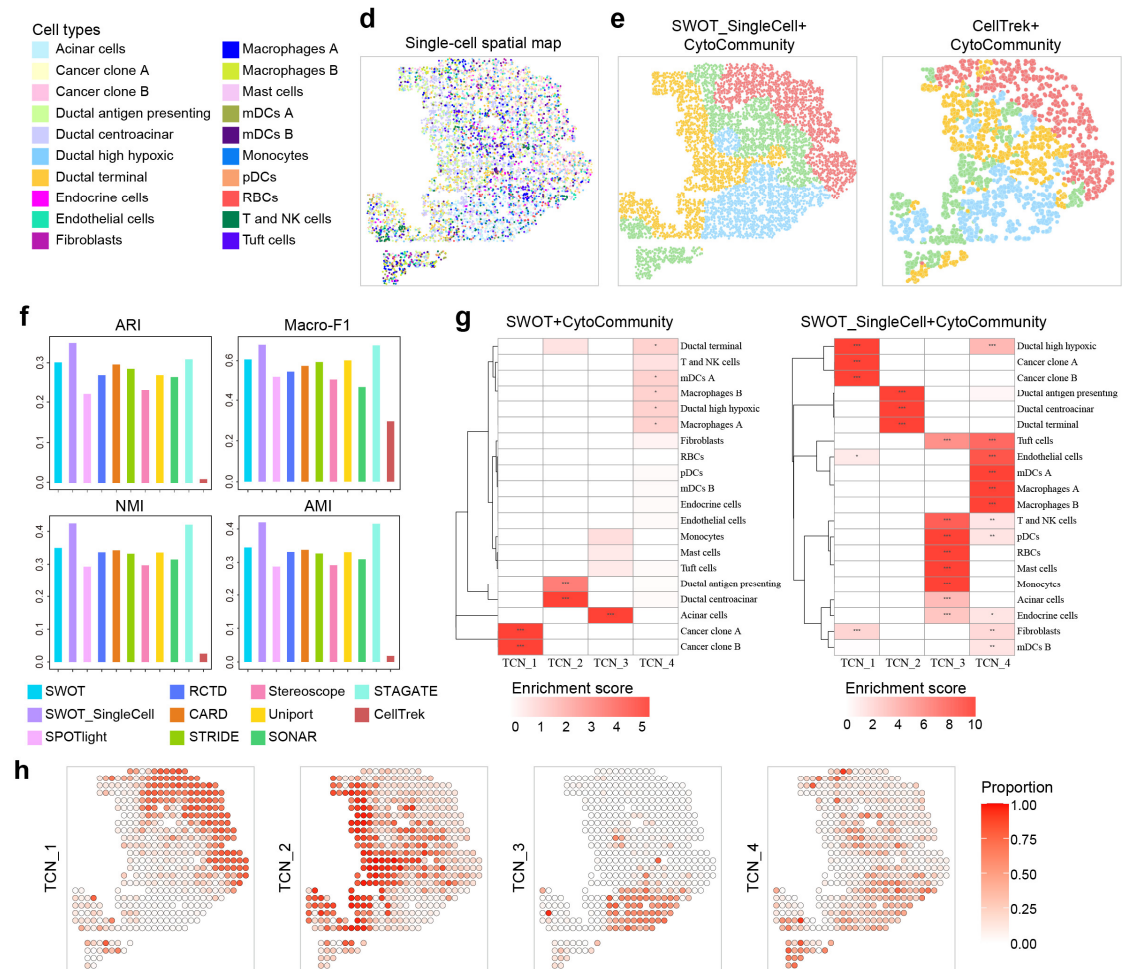


Figure 6. SWOT facilitates tissue cellular neighborhoods (TCNs) functions annotation on pancreatic ductal adenocarcinoma dataset. **d** Spatial distributions of cells in the inferred single-cell spatial map by SWOT. Points denote cells and colored by cell types. **e** TCNs identified by SWOT_SingleCell+CytoCommunity and CellTrek+CytoCommunity methods, respectively. **f** Barplots of ARI, Macro-F1, NMI, and AMI scores of identified TCNs by all methods and annotated regions. **g** Heatmaps of enrichment scores of each cell type in identified TCNs by SWOT+CytoCommunity and SWOT_SingleCell+CytoCommunity, respectively. Color represents enrichment scores. **h** Spatial distributions of enriched cell types in four TCNs. Color indicates the proportions of enriched cell types in that TCN.

Comment 3.

The analysis of tissue cellular neighborhoods (TCNs) in Fig. 6G is not well interpreted. The authors could consider integrating cell-cell interaction methods to correlate enriched cell types with known signaling pathways or ligand-receptor pairs to enhance biological interpretability.

Response: Thanks for your valuable feedback and suggestions, and we answered this problem from the following two aspects.

(1) We have provided the detailed analysis of TCNs in subsection “SWOT facilitates annotation of tissue cellular neighborhoods functions” of the Results section.

The colocalization of functionally distinct cell types within the same spatial region may reflect cellular interactions and shared contributions to regional functions. Accordingly, we functionally

annotated each TCN based on the colocalized cell types. For example, in the TCN_1 region, the colocalization of cancer cells and inflammatory fibroblasts suggested a reshape peritumoral stromal environment that could facilitate tumor cell invasion and metastasis⁴⁸. TCN_2, enriched in diverse pancreatic ductal subpopulations, may participate in cancer-associated antigens presentation, pancreatic juice secretion, and ductal maintenance within the tumor microenvironment⁴⁵. In TCN_3, representing pancreatic tissue, and TCN_4, representing stroma tissue, different infiltrating immune cell types was observed, providing important insights for the development of subsequent therapeutic strategies⁴⁹. Collectively, these findings demonstrated that the single-cell spatial map inferred by SWOT enables high-resolution identification of enriched cell types within TCNs, providing functional insights that may not be achievable through spot-resolution ST data alone.

(2) We have also performed interaction analysis and added descriptions in subsection “SWOT facilitates annotation of tissue cellular neighborhoods functions” of the Results section in the revised manuscript.

Finally, to investigate the interactions among different cell types within TCNs, we combined CellChat⁵⁰, a widely used method for inferring cell-cell communication networks from scRNA-seq or ST data. Focusing on the TCN_1 (tumor) region identified from the single-cell spatial map constructed by SWOT, we performed interaction analyze between the enriched cell types and known pathways, highlighting significant ligand-receptor interactions associated with tumor and immune (Figure 6i). The inferred spatial map captured a broader diversity of cell types enriched within tumor region, enabling a more diversity and significant communication probability. Among the identified signaling pathways, we highlighted the PLAU- and CCL-related interactions (Figure 6j). The PLAU can activate plasminogen system, promote extracellular matrix degradation, and create physical conditions conducive for tumor invasion and metastasis⁴⁸. We observed heterogeneous communication patterns of PLAU-PLAUR between two tumor subclones (Cancer clone A and Cancer clone B), reflecting their differential invasive and metastatic capacities⁵¹. Additionally, communication via CCL-related pathways between immune cells and stromal cells may regulate immune and inflammatory responses, impacting tumor growth and metastasis⁴⁹. Notably, we identified widespread signaling through the C3-C3AR1, with particularly high expression in macrophages and myeloid dendritic cells ($p < 0.01$). This implicated a key role for the complement system in PDAC-associated inflammation, potentially promoting tumor progression by recruiting myeloid cells or regulating pro-tumorigenic inflammation⁴⁹. These inferred interactions from single-cell spatial map not only enable functional refinement of TCN annotations but also provide critical insights into tumor microenvironment architecture and mechanisms of immune evasion.

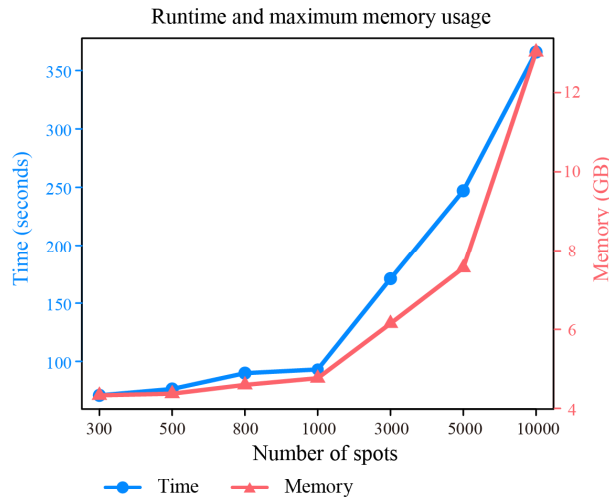
In GitHub repository

(https://github.com/GaoLabXDU/SWOT/blob/main/SWOT/utis/dis_matrix.py), we have added a Torch based code for calculating distances in GPU environment

(2) In the revised manuscript, we have appended the discussion and experiments on computational efficiency in the Discussion section.

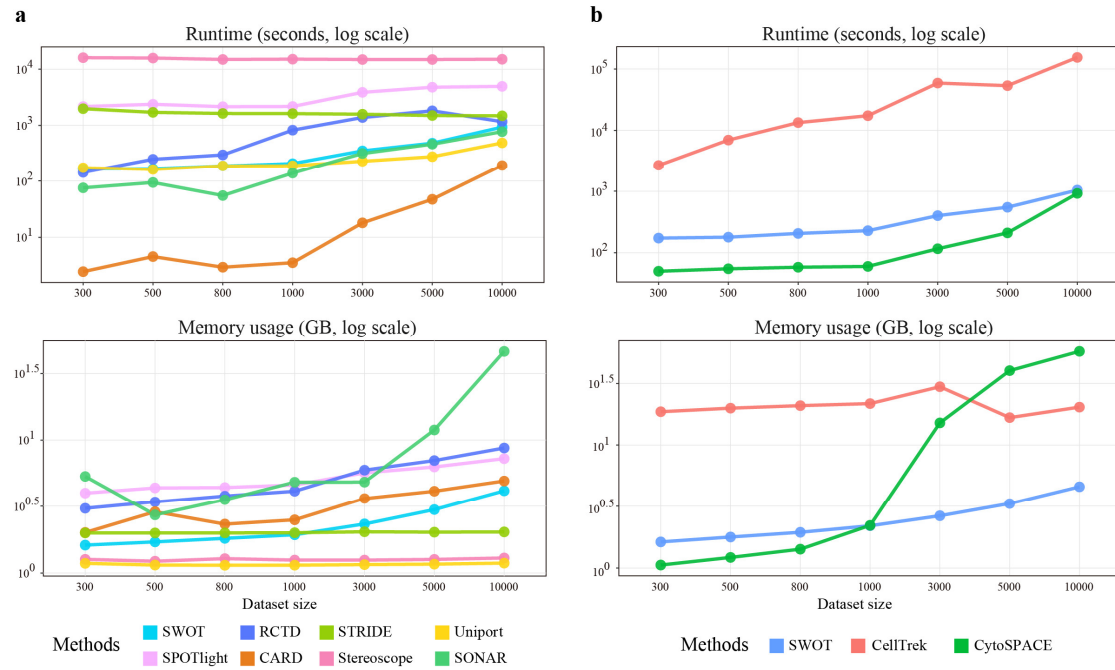
It should be noted that the primary computational cost of SWOT arises from the learning of cell-to-spot mapping in the optimal transport module, which scales with the cell numbers in scRNA-seq data and the spot numbers in ST data. Experimental results on Simulated_MBaging datasets, with varying numbers of spots, indicated that the runtime and memory consumption of optimal transport module increase with the number of spots (Figure S32). Compared to existing methods, SWOT maintained competitive computational efficiency while delivering practical performance (Figure S33).

Figure S32. Runtime and memory usage of optimal transport module in SWOT



We used all Simulated_MBaging datasets to test the runtime (seconds, left y-axis) and maximum memory usage (GB, right y-axis) of the optimal transport module in SWOT. The Simulated_MBaging dataset consists of 15040 cells in reference (snRNA-seq) data, and the simulated ST datasets have 290, 426, 749, 989, 2850, 4669, 9554 spots, respectively. The horizontal axis represents the simplified name of all Simulated_MBaging datasets. All experiments were conducted on a workstation equipped with a 13th Gen Intel(R) Core(TM) i7-13700K CPU (3.4GHz, 16 cores) and 64GB of RAM running Windows.

Figure S33. Computational efficiency comparison of all methods



To comprehensively evaluate the computational efficiency and practical performance of SWOT, we conducted experiments on all Simulated_MBaging datasets with varying sizes, and plotted the runtime and memory consumption of all methods in two tasks: **(a)** cell-type composition inference and **(b)** single-cell spatial maps inference. Due to the wide range of values for time and memory usage among methods, we applied a log10 transformation to both runtime and memory usage. Each subplot presents the trends after the logarithmic conversion, with different colors representing different methods. The Simulated_MBaging dataset has 15040 cells in reference (snRNA-seq) data, and the simulated ST datasets have 290, 426, 749, 989, 2850, 4669, 9554 spots, respectively. The horizontal axis represents the simplified name of all Simulated_MBaging datasets. All experiments were conducted on a workstation equipped with a 13th Gen Intel(R) Core(TM) i7-13700K CPU (3.4GHz, 16 cores) and 64GB of RAM running Windows.

References

1. Zormpas, E., Queen, R., Comber, A. & Cockell, S. J. Mapping the transcriptome: Realizing the full potential of spatial data analysis. *Cell* **186**, 5677–5689 (2023).
2. Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol.* **40**, 1349–1359 (2022).
3. You, Y. *et al.* Systematic comparison of sequencing-based spatial transcriptomic methods. *Nat Methods* **21**, 1743–1754 (2024).
4. Allen, W. E., Blosser, T. R., Sullivan, Z. A., Dulac, C. & Zhuang, X. Molecular and spatial signatures of mouse brain aging at single-cell resolution. *Cell* **186**, 194–208.e18 (2023)
5. Li, B. *et al.* Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods* **19**, 662–670 (2022).
6. Dong, R. & Yuan, G.-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.* **22**, 145 (2021).
7. Miller, B. F., Huang, F., Atta, L., Sahoo, A. & Fan, J. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nat Commun.* **13**, 2339 (2022).
8. Hao, M. *et al.* STEM enables mapping of single-cell and spatial transcriptomics data with transfer learning. *Commun Biol.* **7**, 56 (2024).
9. Wan, X. *et al.* Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat Commun.* **14**, 7848 (2023).
10. Liu, Y. *et al.* SDePER: a hybrid machine learning and regression method for cell-type deconvolution of spatial barcoding-based transcriptomic data. *Genome Biol* **25**, 271 (2024).

11. Tu, J.-J., Li, H.-S., Yan, H. & Zhang, X.-F. EnDecon: cell type deconvolution of spatially resolved transcriptomics data via ensemble learning. *Bioinformatics* **39**, btac825 (2023).
12. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
13. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, (2015).
14. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163–172 (2019).
15. Ba, B. *et al.* Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res.* **29**, 697–709 (2019).
16. Li, M. *et al.* CellBin: a highly accurate single-cell gene expression processing pipeline for high-resolution spatial transcriptomics. Preprint at <https://doi.org/10.1101/2023.02.28.530414> (2024).
17. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
18. Liu, B. *et al.* An entropy-based metric for assessing the purity of single cell populations. *Nat Commun.* **11**, 3155 (2020).
19. Cao, K., Gong, Q., Hong, Y. & Wan, L. A unified computational framework for single-cell data integration with optimal transport. *Nat Commun.* **13**, 7419 (2022).
20. Wang, Y. *et al.* Sprod for de-noising spatially resolved transcriptomics data based on position and image information. *Nat Methods* **19**, 950–958 (2022).
21. Zhao, P., Zhu, J., Ma, Y. & Zhou, X. Modeling zero inflation is not necessary for spatial transcriptomics. *Genome Biol.* **23**, 118 (2022).
22. Vahid, M. R. *et al.* High-resolution alignment of single-cell and spatial transcriptomes with CytoSPACE. *Nat Biotechnol.* **41**, 1543–1548 (2023).
23. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* **18**, 1352–1362 (2021).
24. Liu, Z., Wu, D., Zhai, W. & Ma, L. SONAR enables cell type deconvolution with spatially weighted Poisson-Gamma model for spatial transcriptomics. *Nat Commun.* **14**, 4727 (2023).
25. Lu, Y., Chen, Q. & An, L. SPADE: spatial deconvolution for domain specific cell-type estimation. *Commun Biol.* **7**, 469 (2024).
26. Moriel, N. *et al.* NovoSpaRc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat Protoc.* **16**, 4177–4200 (2021).
27. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun.* **11**, 2084 (2020).
28. Liao, J. *et al.* De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. *Nat Commun.* **13**, 6498 (2022).
29. Calafiore, G., Dabbene, F. & Tempo, R. Radial and uniform distributions in vector and matrix spaces for probabilistic robustness. Miller, D.E., Qiu, L. (eds) *Topics in Control and its Applications*. Springer, London 17–31 (1999).
30. Park, J., Shin, T., Yoon, S. H., Goo, J. M. & Park, J.-Y. A radial sampling strategy for uniform k-space coverage with retrospective respiratory gating in 3D ultrashort-echo-time lung imaging. *NMR Biomed.* **29**, 576–587 (2016).
31. Rouch, T. Unit disk uniform sampling. *Towards Data Science* <https://towardsdatascience.com/unit-disk-uniform-sampling-91880f3740fa/> (2024).
32. How sampling on a spherical (polar) coordinate can be biased. *IsaacTheBlog* <https://isaac-the-man.dev/posts/how-sampling-on-a-spherical-coordinate-can-be-biased/> (2024).
33. Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. *Nat Protoc.* **20**, 180–219 (2025).

Response to Reviewers' Remarks

Dear Editors and Reviewers:

Thanks so much for your time and efforts on our manuscript. Also, we thank the reviewers for valuable comments and suggestions. These comments are valuable for revising and improving our manuscript. We further carefully revised our manuscript according to the reviewers' suggestions. The revised parts are highlighted in blue in the revised manuscript. The followings are our point-by-point responses. Also please let me know if there are remaining issues of concern. Thank you for your consideration.

Responses to Reviewer #1's remarks

I thank the authors for their efforts to address my previous concerns. I only have two minor comments on the article.

1. It seems to me that Fig. 2f-g shows a systematic difference in cell type proportion (it tends to overestimate the oligodendrocyte proportion), but the article does not comment on the difference. I am curious about the reason for this bias.

Response: Thanks for your insightful comments. We answered from two aspects as follows.

(1) We provided explanations for this bias.

As shown in Figures 2f-g, S7, and S8, all deconvolution methods exhibited bias of varying degrees relative to the ground truth, both across datasets and cell types. For example, in the Simulated_Cortex dataset, methods such as SPOTlight, STRIDE, Stereoscope, and Uniport tended to underestimate the proportion of cell type Microglia. In the Simulated_MBaging_300 dataset, SWOT and CARD overestimated the Olig proportion, while RCTD and SONAR underestimated Neuron proportion. We consider that these biases may arise from the algorithm, which designed based on different mathematical model.

In our manuscript, based on the mathematical foundations of cell-type deconvolution methods¹⁻³, we selected seven representative approaches for evaluation. SPOTlight and CARD are regression-based methods that typically leverage reference to obtain cell-type-specific gene expressions, representing each spot as a weighted linear combination of different cell types. However, their primary limitation stems from the incompatibility between the assumptions of regression models and the biological reality of transcriptomics data, most notably the multicollinearity among cell types.

Probability-based models, like RCTD, STRIDE, Stereoscope, and SONAR, use the probability distribution of gene expression for each cell type to fit the mixed expression of spots. However, these models rely on data distribution assumptions that may not adequately capture the complex characteristics of real biological data. Such issues are commonly encountered in spatial deconvolution, and ongoing researches are actively addressing them^{4,5}.

SWOT and Uniport are built on optimal transport models that directly learn a mapping between cells and spots to represent their probabilistic matching relationship. Beyond estimating cell-type composition, this learned mapping can be repurposed for tasks such as correcting low quality spatial measurements, mapping cell locations⁶, reconstructing cell differentiation trajectories, predicting single-cell responses, and reconstituting tissue architectures⁷. Although the optimal transport model can bring advantages, when aligning scRNA-seq and ST data, it implicitly assumes that both modalities represent similar underlying distributions⁷. For this assumption, SWOT employed an unbalanced term, which may encounter a conflict with structured term, specifically between the global mass adjustment and the preservation of local geometry⁸, and there is no theoretically guided optimal balance point for these factors. We argue that this assumption may ultimately affect the calculated transport plan and lead to bias in estimating compositions.

We list Figures 2f-g as follows for reference:

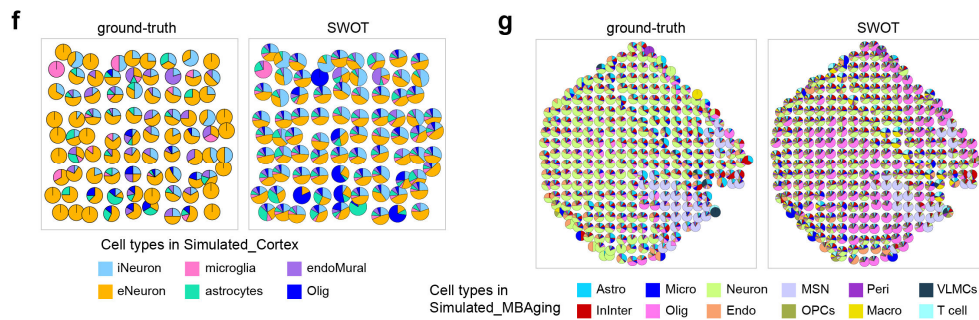


Figure 2. Performance evaluation for cell-type composition inference on simulated datasets. **f** and **g** Spatial scatter pie of cell-type composition obtained by ground-truth and SWOT for Simulated_Cortex and Simulated_MBaging_300 datasets, respectively. Each pie denotes a spot, colored by cell types, and divided by proportions.

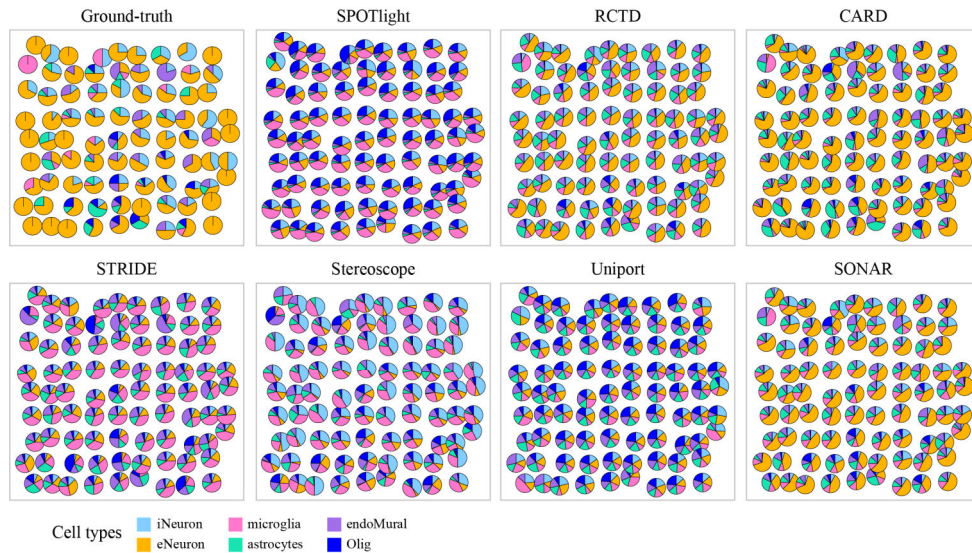
(2) We have supplemented comments in the revised manuscript.

In the revised main text, description has been added in subsection “Performance evaluation of SWOT on inferring cell-type composition” of the Results section.

Although SWOT overestimated the proportion of cell type Olig in both simulated datasets, we found that other deconvolution methods also exhibited bias of varying degrees relative to the ground-truth, across datasets and cell types.

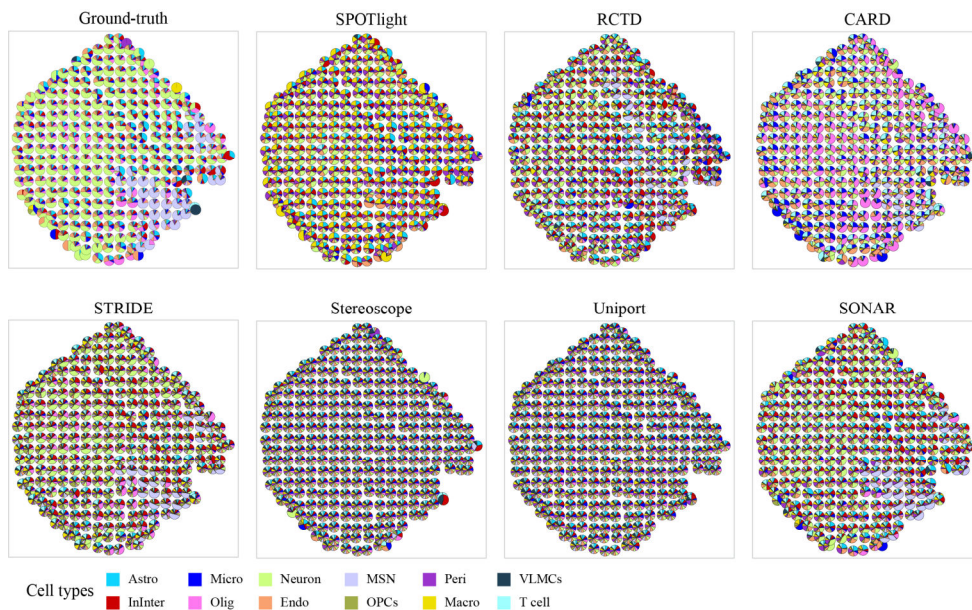
In the revised Supplementary text, we have added a “Ground-truth” subplot for comparison in Figure S7 and Figure S8.

Figure S7. Spatial scatter pie of cell types on Simulated_Cortex dataset



Spatial scatter pie of cell-type composition estimated by seven cell-type deconvolution methods, including: SPOTlight, RCTD, CARD, STRIDE, Stereoscope, Uniport, and SONAR. The first subplot is the ground-truth of cell-type composition. Each pie in each panel denotes a spot, colored by cell types, and divided by cell-type proportions.

Figure S8. Spatial scatter pie of cell types on Simulated_MBaging_300 dataset



Spatial scatter pie of cell-type composition in Simulated_MBaging_300 dataset estimated by seven cell-type deconvolution methods, including: SPOTlight, RCTD, CARD, STRIDE, Stereoscope, Uniport, and SONAR. The first subplot is the ground-truth of cell-type composition. Each pie in each panel denotes a spot, colored by cell types, and divided by cell-type proportions.

2. Several figures compare the ratio between the estimated value and the ground truth. I think it makes sense to put the ratio on the log scale so that it is fairer to compare how far the ratio is from 1 (both higher and lower directions). It also helps with overcrowding. For example, it is hard to compare the performance in Figure S14a, although the authors have provided the

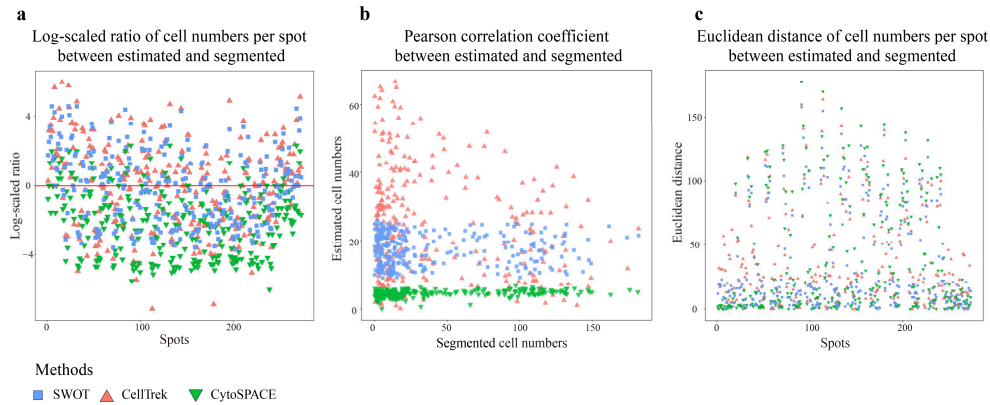
average values in the caption.

Response: We appreciate your comments and suggestions. By performing a log transformation, we have updated the presentation of the ratio between the estimated value and the ground truth in the revised manuscript.

In subsection “SWOT depicts cell-type spatial distributions” of the Results section:

Based on the segmented numbers, SWOT showed superior performance, achieving a spot-wise log-scaled ratio close to 0, the highest correlation, and the lowest Euclidean distance (Figure S14).

Figure S14. Estimated cell numbers per spot on MOB dataset



Comparison between the estimated cell numbers and the segmented cell numbers. (a) The log-scaled ratio between the estimated and the segmented cell numbers for each spot. (b) The Pearson correlation coefficient between the estimated and the segmented cell numbers across all spots. (c) The Euclidean distance between the estimated and the segmented cell numbers for each spot. The estimated cell numbers were obtained by three methods: SWOT, CellTrek, and CytoSPACE. The segmented cell numbers were obtained through cell-segmentation of H&E image. Each dot in the plots represents a single spot, with colors and shapes denoting different methods. In (a), the straight line indicates a log-scaled ratio value of 0, representing the best estimate of cell numbers. The average log-scaled ratio of SWOT, CellTrek, and CytoSPACE were -0.156, 0.015, and -2.086, respectively. The average Pearson correlation coefficient of SWOT, CellTrek, and CytoSPACE were 0.4, -0.203, and 0.346, respectively. The average Euclidean distance of SWOT, CellTrek, and CytoSPACE were 34.425, 38.777, and 37.67, respectively.

In subsection “Performance evaluation of SWOT on inferring single-cell spatial maps” of the Results section:

First, in each simulated dataset, we examined the cell number estimation strategy using three quantitative metrics: (1) the log-scaled ratio between the estimated and the ground-truth cell numbers per spot.

The considerable results of SWOT across Simulated_MBAGing datasets with different dropout rates further supported its robustness in estimating cell numbers (Figure S15 and Table S2). We also illustrated the overall performance across all Simulated_MBAGing datasets, with log-scaled ratios and Euclidean distances close to 0, and correlations close to 1 indicating consistent estimates of cell numbers with ground-truth (Figure 4b). According to the obtained log-scaled ratio values and its proportion of all spots in each dataset, CytoSPACE showed more negative values, indicating an underestimation of cell numbers, while CellTrek had more positive values, illustrating a trend of

overestimation, and SWOT produced results that were closer to 0 (with the summary 1.599 proportions in the range of -0.1~0.1).

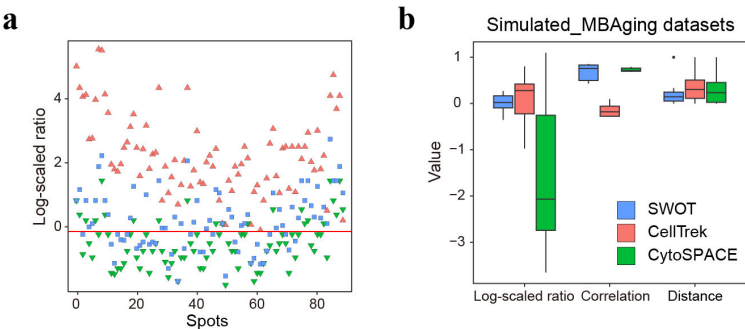
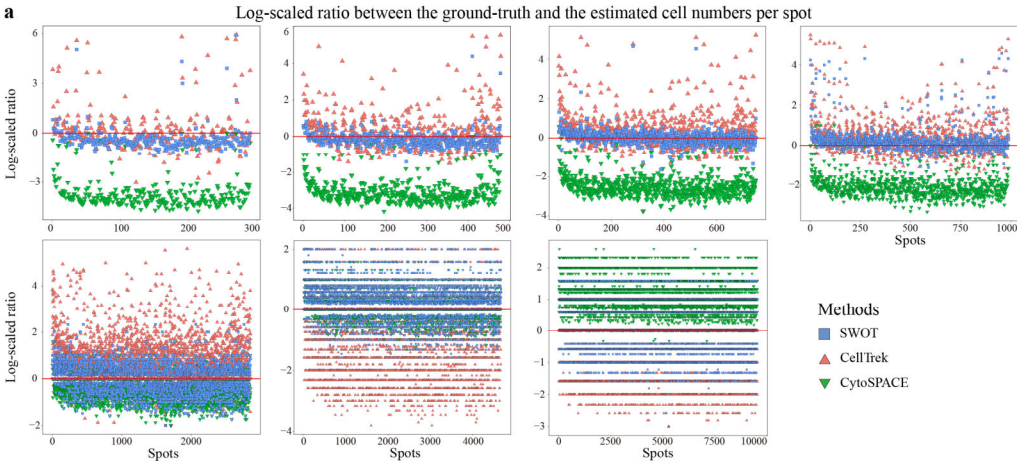


Figure 4. Performance evaluation for single-cell spatial maps inference on simulated datasets. **a** Performance of cell number estimation strategy in Simulated_Cortex dataset. Left: Log-scaled ratio between estimated and ground-truth cell numbers per spot. Middle: Pearson correlation coefficient between estimated and ground-truth cell numbers across all spots. Right: Euclidean distance between estimated and ground-truth cell numbers per spot. Colors and shapes denote methods, and the straight line in the left panel indicates a ratio of 0, representing the best estimation. **b** Boxplots of the average values for the log-scaled ratio, Pearson correlation coefficient, and Euclidean distance between estimated and ground-truth cell numbers across all Simulated_MBaging datasets.

Figure S15. Estimated cell numbers per spot on simulated datasets with different dropout rate



Seven Simulated_MBaging datasets (Simulated_MBaging_300, Simulated_MBaging_500, Simulated_MBaging_800, Simulated_MBaging_1000, Simulated_MBaging_3000, Simulated_MBaging_5000, and Simulated_MBaging_10000) were generated under different spatial resolutions and exhibited distinct dropout rates of 0.0543, 0.0586, 0.0927, 0.1303, 0.3656, 0.5025, and 0.6781, respectively. **(a)** The log-scaled ratio between estimated and ground-truth cell numbers for each spot. In each subplot, all Simulated_MBaging datasets are arranged from left to right, top to bottom. Each dot in the plots represents an individual spot, with colors and shapes denoting methods. In (a), the straight line in each subplot indicates a log-scaled ratio value of 0, representing the best estimate of cell numbers. The average values of the three metrics (log-scaled ratio, Pearson correlation coefficient, and Euclidean distance) across all spots for each dataset were summarized in Table S2.

Table S2. Evaluation of the estimated cell numbers on Simulated_MBaging dataset

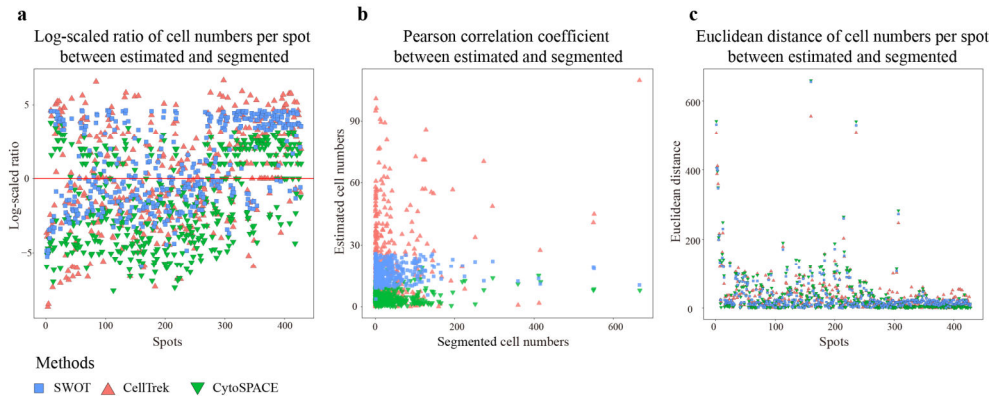
For all Simulated_MBaging datasets, we compared the estimated cell numbers per spot by three methods (SWOT, CellTrek, and CytoSPACE) against the ground-truth using three metrics. We listed the average log-scaled ratio between estimated and ground-truth cell numbers per spot, the average Pearson correlation coefficient between estimated and ground-truth across all spots, and the average Euclidean distance between estimated and ground-truth values for each spot. The closer the log-scaled ratio is to 0, the greater the correlation, and the smaller the distance, indicating a better estimation. We also calculated the proportion of all spots in each dataset and for each method whose log-scaled ratio values fell into the ranges of less than -0.1, between -0.1 and 0.1, and greater than 0.1.

Metric	Method	Simulated_MBaging dataset						
		300	500	800	1000	3000	5000	10000
Log-scaled ratio	SWOT	-0.353	-0.168	0.022	0.269	-0.02	0.254	0.079
	CellTrek	0.143	0.279	0.383	0.459	0.801	-0.974	-0.587
	CytoSPACE	-3.653	-3.043	-2.445	-2.068	-0.611	0.104	1.096
ratio < -0.1	SWOT	0.828	0.647	0.342	0.169	0.422	0.228	0.268
	CellTrek	0.528	0.411	0.290	0.272	0.151	0.731	0.554
	CytoSPACE	0.983	0.993	0.988	0.981	0.845	0.294	0.002
-0.1 ≤ ratio ≤ 0.1	SWOT	0.072	0.198	0.288	0.264	0.153	0.202	0.422
	CellTrek	0.093	0.096	0.137	0.111	0.056	0.113	0.334
	CytoSPACE	0.017	0.006	0.009	0.015	0.123	0.318	0.046
0.1 < ratio	SWOT	0.1	0.154	0.369	0.566	0.425	0.569	0.31
	CellTrek	0.379	0.492	0.573	0.616	0.793	0.155	0.111
	CytoSPACE	0	0.001	0.003	0.004	0.031	0.387	0.952
Pearson correlation coefficient	SWOT	0.812	0.856	0.852	0.758	0.432	0.553	0.443
	CellTrek	0.098	-0.18	-0.28	-0.282	-0.272	-0.118	0.004
	CytoSPACE	0.688	0.695	0.697	0.714	0.768	0.792	0.759
Euclidean distance	SWOT	23.183	8.219	3.979	4.256	2.206	1.526	0.708
	CellTrek	28.897	19	11.401	9.496	5.649	2.662	1.059
	CytoSPACE	67.645	39.098	23.346	16.621	2.939	0.964	2.304

In subsection “SWOT facilitates annotation of tissue cellular neighborhoods functions” of the Results section:

The comparison of the estimated cell numbers with the segmented cell numbers illustrated the reliable of SWOT (Figure S25).

Figure S25. Estimated cell numbers per spot on PDAC dataset



Comparison between the estimated and the segmented cell numbers. (a) The log-scaled ratio between the estimated and the segmented cell numbers for each spot. (b) The Pearson correlation coefficient between the estimated and the segmented cell numbers across all spots. (c) The Euclidean distance between the estimated and the segmented cell numbers for each spot. The estimated cell numbers were obtained by three methods: SWOT, CellTrek, and CytoSPACE. The segmented cell numbers were obtained through cell-segmentation of H&E image. Each dot in the plots represents a single spot, with colors and shapes denoting different methods. In (a), the straight line indicates a log-scaled ratio value of 0, representing the best estimate of cell numbers. The average log-scaled ratio of SWOT, CellTrek, and CytoSPACE were 0.895, 0.272, and -1.343, respectively. The average Pearson

correlation coefficient of SWOT, CellTrek, and CytoSPACE were 0.13, 0.161, and 0.2, respectively. The average Euclidean distance of SWOT, CellTrek, and CytoSPACE were 35.759, 38.883, and 36.453, respectively.

Responses to Reviewer #2's remarks

The authors have fully addressed my comments; the manuscript is much improved.

Response: Thank you very much for your recognition of our works and for the constructive suggestions you provided, which have helped us improve our work.

Responses to Reviewer #4's remarks

The authors thoroughly addressed my previous concerns regarding their manuscript on SWOT. The revisions have substantially strengthened the manuscript, addressing all my previous concerns. The authors have improved the mathematical notation, addressed the potential bias in cell coordinate sampling, enhanced the interpretation of tissue cellular neighborhoods with cell-cell interaction analyses, and provided comprehensive runtime analyses to demonstrate the method's scalability.

The authors have significantly enhanced the paper by:

- 1. Providing a comprehensive analysis and justification of hyperparameter selection, including bandwidth settings, K nearest neighbors values, clustering methods, and regularization terms. New sensitivity analyses across multiple datasets demonstrate the robustness of its methods and provide clear guidance to users.*
- 2. Through comparative analysis with alternative approaches that use only spatial coordinates or gene expression, the authors elucidate the rationale behind their spatial weighting strategy and demonstrate its necessity.*
- 3. Expanding the description of their simulation methodology, clearly explaining how spot-resolution data were generated from single-cell resolution spatial transcriptomics data while preserving realistic gene expression characteristics, dropout events, and intercellular variability.*
- 4. The authors thoroughly explain their cell number estimation procedure and validate it on datasets with different dropout rates. Although I still have some doubts about the method itself, the authors use sufficient validation and cell segmentation on H&E images as an independent validation method to prove the reliability of their method.*
- 5. The authors demonstrate that SWOT remains computationally practical. With GPU-accelerated distance-matrix routines available, and runtime/memory benchmarks on large simulations show performance on par with or better than competing methods.*

The manuscript now presents SWOT as a well-justified and robust method for inferring both cell-type composition and single-cell spatial maps from spot-based spatial transcriptomics data. I have no remaining concerns or revision comments.

Response: Thank you very much for your recognition of our works and for the

constructive suggestions you provided, which have helped us improve our work.

References

1. Gaspard-Boulinç, L. C., Gortana, L., Walter, T., Barillot, E. & Cavalli, F. M. G. Cell-type deconvolution methods for spatial transcriptomics. *Nat Rev Genet* (2025).
2. Zhang, Y. *et al.* Deconvolution algorithms for inference of the cell-type composition of the spatial transcriptome. *Comput Struct Biotechnol J* **21**, 176–184 (2023).
3. Chen, J. *et al.* A comprehensive comparison on cell-type composition inference for spatial transcriptomics data. *Brief Bioinform* **23**, bbac245 (2022).
4. Garmire, L. X. *et al.* Challenges and perspectives in computational deconvolution of genomics data. *Nat Methods* **21**, 391–400 (2024).
5. Maden, S. K. *et al.* Challenges and opportunities to computationally deconvolve heterogeneous tissue with varying cell sizes using single-cell RNA-sequencing datasets. *Genome Biol* **24**, 288 (2023).
6. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods* **18**, 1352–1362 (2021).
7. Bunne, C., Schiebinger, G., Krause, A., Regev, A. & Cuturi, M. Optimal transport for single-cell and spatial omics. *Nat Rev Methods Primers* **4**, 58 (2024).
8. Gallouët, T., Ghezzi, R. & Vialard, F.-X. Regularity theory and geometry of unbalanced optimal transport. Preprint at <https://doi.org/10.48550/arXiv.2112.11056> (2024).