

Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution

Bin Li^{1,7}, Wen Zhang^{1,2,7}, Chuang Guo^{1,7}, Hao Xu^{1,2}, Longfei Li³, Minghao Fang³, Yinlei Hu⁴, Xinye Zhang³, Xinfeng Yao¹, Meifang Tang¹, Ke Liu¹, Xuetong Zhao⁵, Jun Lin^{1,2}, Linzhao Cheng³, Falai Chen⁴, Tian Xue³ and Kun Qu^{1,2,6}✉

Spatial transcriptomics approaches have substantially advanced our capacity to detect the spatial distribution of RNA transcripts in tissues, yet it remains challenging to characterize whole-transcriptome-level data for single cells in space. Addressing this need, researchers have developed integration methods to combine spatial transcriptomic data with single-cell RNA-seq data to predict the spatial distribution of undetected transcripts and/or perform cell type deconvolution of spots in histological sections. However, to date, no independent studies have comparatively analyzed these integration methods to benchmark their performance. Here we present benchmarking of 16 integration methods using 45 paired datasets (comprising both spatial transcriptomics and scRNA-seq data) and 32 simulated datasets. We found that Tangram, gimVI, and SpaGE outperformed other integration methods for predicting the spatial distribution of RNA transcripts, whereas Cell2location, SpatialDWLS, and RCTD are the top-performing methods for the cell type deconvolution of spots. We provide a benchmark pipeline to help researchers select optimal integration methods to process their datasets.

Spatial transcriptomics approaches allow us to detect RNA transcripts in space, and these approaches have been used to investigate the spatial distribution of gene expression in various tissues and organs, including the brain¹, heart², pancreas³, and skin⁴. On the one hand, the spatial transcriptomics approaches based on *in situ* hybridization and fluorescence microscopy (image-based)—including seqFISH⁵, osmFISH⁶, and MERFISH⁷—detect the spatial distribution of transcripts with high resolution and accuracy, but they are limited in the total number of RNA transcripts that they can detect. On the other hand, spatial transcriptomics approaches based on next-generation sequencing (seq-based), such as ST⁸, 10X Visium⁹, and Slide-seq^{10,11}, can capture expressed RNAs at the whole-transcriptome scale from spots in space, but each spot (radius 10–100 μm) may contain multiple cells, which limits the spatial resolution of these approaches. The limitations of these spatial transcriptomics approaches hinder their capacity to capture whole-transcriptome-scale data at single-cell resolution in space.

To break through the limitations of spatial transcriptomics approaches, bioinformaticians have proposed and developed various integration methods to combine spatial transcriptomics and single-cell RNA-seq (scRNA-seq) data. For example, gimVI¹² employs a deep generative model to infer the likely spatial distribution of undetected transcripts; SpaGE¹³ uses the domain adaptation algorithm PRECISE¹⁴ and *k*-nearest-neighbor regression to predict the spatial distribution of undetected transcripts; Tangram¹⁵ uses non-convex optimization and a deep learning framework to learn

a spatial alignment for scRNA-seq data; Seurat¹⁶ applies canonical correlation analysis¹⁷ to embed spatial and scRNA-seq data into a common latent space, and projects cells from scRNA-seq data to the spots of the spatial transcriptomics data; LIGER¹⁸ uses both integrative non-negative matrix factorization¹⁹ and shared factor neighborhood graphs to predict gene expression levels in space; novoSpaRc²⁰ and SpaOTsc²¹ each use optimal transport methods²² to construct spatial metrics of cells on the basis of scRNA-seq data; stPlus²³ combines the auto-encoder and weighted *k*-nearest-neighbor methods to predict spatial gene expression. These integration methods enable researchers to predict the spatial distribution of undetected transcripts.

In addition, Seurat, Tangram, novoSpaRc, and SpaOTsc have the capacity to assign cells from scRNA-seq data to spatial locations in histological sections; this is useful for improving the resolution of the spatial transcriptomics data generated using spatial transcriptomics approaches, like ST or 10X Visium. Moreover, Cell2location²⁴ uses the gene expression signature of the cell subpopulations in scRNA-seq data to estimate the abundance of each cell type at each spot; RCTD²⁵ applies cell type profiles learned from scRNA-seq data and supervised learning to decompose cell type mixtures; SpatialDWLS²⁶ adopts the weighted-least-squares approach to infer cell type composition; Stereoscope²⁷ leverages the model-based probabilistic method and scRNA-seq data to deconvolve the cell mixtures in spatial data; SPOTlight²⁸ applies the seeded non-negative matrix factorization for the deconvolution

¹Department of Oncology, The First Affiliated Hospital of USTC, School of Basic Medical Sciences, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China. ²Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, China. ³Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China. ⁴School of Mathematical Sciences, University of Science and Technology of China, Hefei, China. ⁵CAS Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. ⁶CAS Center for Excellence in Molecular Cell Sciences, the CAS Key Laboratory of Innate Immunity and Chronic Disease, University of Science and Technology of China, Hefei, China. ⁷These authors contributed equally: Bin Li, Wen Zhang, Chuang Guo.

✉e-mail: qukun@ustc.edu.cn

of spots; DSTG²⁹ deconvolutes spatial transcriptomics data using graph-based convolutional networks; STRIDE³⁰ uses the topic profiles trained from scRNA-seq data to decompose cell types from spatial mixtures; DestVI³¹ adopts the variational inference and latent variable models to delineate cell type proportions. These integration methods allow researchers to predict the cell type composition of spots in histological sections.

The emergence of these integration methods has undoubtedly deepened our understanding of spatial transcriptomics data and related biological and pathological processes. However, to the best of our knowledge, no independent study has comprehensively compared the performance of these integration methods for the prediction of the spatial distribution of transcripts or for the cell type deconvolution of spots in histological sections. Here, we used multiple metrics to systematically benchmark the performance of 16 integration methods that can predict the spatial distribution of undetected transcripts, or the cell type composition of spots in histological sections (Fig. 1a), on the basis of processing of 45 paired datasets containing both spatial transcriptomics data and scRNA-seq data and 32 simulated datasets (Fig. 1b). We assessed the accuracy of each integration method in predicting the spatial distribution of transcripts, including for sparse spatial transcriptomics data that were down-sampled from the original datasets. We also evaluated the accuracy of the integration methods for the cell type deconvolution of spots in histological sections on the basis of the simulation of datasets wherein each spot could contain multiple cells of various types. Finally, we evaluated the computational resources consumed by each integration method. Our findings can help researchers choose appropriate integration methods for their datasets, and they raise interesting questions about how various processing and dataset-specific attributes influence the integration performance of these tools for spatial transcriptomics research.

Results

Benchmarking framework and datasets examined. To evaluate the performance of the 16 integration methods, we collected 45 paired spatial transcriptomics and scRNA-seq datasets from published studies^{4–7,10,15,31–61} (Fig. 1 and Supplementary Table 1). The spatial transcriptomic datasets were produced by 13 spatial transcriptomics approaches, including FISH, osmFISH, seqFISH, MERFISH, STARmap, ISS, EXseq, BaristaSeq, ST, 10X Visium, Slide-seq, Seq-scope, and HDST, and the scRNA-seq datasets were obtained by Drop-seq⁶², Smart-seq⁶³, and the 10X Chromium platform⁶⁴. We designed a pipeline to evaluate the performance of the integration methods for combining spatial and single-cell transcriptomics datasets (Fig. 1a). During preprocessing of the scRNA-seq datasets, we removed cells with fewer than 200 RNAs. For the spatial transcriptomic datasets, we generated a ‘ground truth’ using 2 criteria: for samples with <1,000 detected RNAs, we used all of the RNAs; for samples with >1,000 detected RNAs, a set of 1,000 highly variable RNAs (assessed on the basis of the coefficient of variation of each RNA; Methods) was used.

In addition, we adopted the algorithms proposed by RCTD and Stereoscope and generated 32 simulated 10X Visium datasets from 16 paired scRNA-seq datasets (Supplementary Tables 2 and 3). A simulated spot contains 5–15 cells randomly sampled from the scRNA-seq datasets (Methods), and the gene expression values of each spot represent the sum of all the cells in that spot.

After collecting the datasets, we first assessed the performance of eight integration methods, including Tangram, gimVI, SpaGE, Seurat, SpaOTsc, novoSpaRc, LIGER, and stPlus, in predicting the spatial distribution of RNA transcripts that remain undetected in spatial transcriptomics datasets. We used the 45 collected paired datasets to evaluate the accuracy of these integration methods for predicting the RNA spatial distribution. Then we down-sampled

the spatial transcriptomics data to test the performance of the integration methods for datasets with sparse expression matrices.

Beyond the prediction of the spatial distribution of RNA transcripts, Tangram, Seurat, SpaOTsc, and novoSpaRc can assign cells from scRNA-seq data to spatial locations in histological sections. Also, Cell2location, SpatialDWLS, RCTD, Stereoscope, DestVI, STRIDE, SPOTlight, and DSTG can be used to predict the cell type composition of spots in histological sections by combining spatial transcriptomics data and scRNA-seq data. All 12 of these integration methods are capable of deconvoluting cell types of the spots in the spatial transcriptomics datasets that were generated using the 10X Visium or ST platforms. To compare the performance of these integration methods in cell type deconvolution, we used datasets 4 and 10 as the basis to simulate ‘grids’ representing low-spatial-resolution datasets, and we simulated 32 datasets from the scRNA-seq data as the ground truth (Methods). Briefly, in the simulated low-resolution datasets, each gridded ‘spot’ contains 1–18 cells, similar to the spatial transcriptome datasets generated by the 10X Visium or ST approaches. Finally, we assessed the computational resources consumed by each integration method.

Methods predicting spatial distribution of RNA transcripts. We used tenfold crossvalidation (Methods) on the 45 paired datasets to evaluate the accuracy of each integration method in predicting the spatial distribution of RNA transcripts. We quantified the prediction performance of each integration method by calculating the Pearson correlation coefficient (PCC) between the expression vector of a gene in the ground truth of the spatial transcriptomics dataset and the expression vector for the same gene in the result predicted by each integration method (Methods). We first examined the prediction results of the spatial distribution for known marker genes. For example, Lein et al. reported that *Igfsf21* and *Rprm* are highly expressed in the L5/L6 layers of the cortex³⁴. Compared with the ground truth for dataset 4 (seqFISH+; Smart-seq; mouse cortex), Tangram performed the best in predicting the spatial distribution of *Igfsf21* (PCC = 0.79), and gimVI, SpaGE, and Seurat followed closely behind (PCC = 0.77, 0.71, and 0.70) (Fig. 2a). For the spatial distribution of *Rprm*, the results generated by SpaGE and Seurat had the highest PCC values (PCC = 0.79), followed by SpaOTsc, gimVI, Tangram, and LIGER (PCC = 0.78, 0.71, 0.66, 0.65) (Fig. 2b).

We also examined the predicted results of the spatial distribution of *COL17A1* in dataset 42 (ST; 10X Chromium; human squamous carcinoma). *COL17A1* is a known marker gene for basal cells of squamous carcinoma⁴. Tangram, gimVI, novoSpaRc, and SpaGE successfully predicted that *COL17A1* was highly expressed in the basal cells of squamous carcinoma; notably, the PCC values of these four integration methods were, respectively, 0.86 (Tangram), 0.84 (gimVI), 0.76 (novoSpaRc), and 0.70 (SpaGE), higher than the best result of the other integration methods (Seurat, 0.48; SpaOTsc, 0.40; LIGER, 0.31; stPlus, 0.27) (Extended Data Fig. 1a).

To further quantify the prediction accuracy of each integration method, we adopted three metrics besides PCC: (1) structural similarity index (SSIM), which combines mean value, variance, and covariance to measure the similarity between the predicted result and the ground truth; (2) root mean square error (RMSE), the absolute error between the predicted distribution and the ground truth; and (3) Jensen–Shannon divergence (JS), which uses relative information entropy to gauge the difference between two distributions. For one gene, a higher PCC/SSIM or lower RMSE/JS value indicates better prediction accuracy. We also defined an accuracy score (AS) by aggregating the four metrics (Methods) to simplify the evaluation of the accuracy of each integration method (a higher AS value indicates better performance).

Taking dataset 4 (seqFISH+; Smart-seq; mouse cortex) as an example of the image-based spatial transcriptomics approaches, Tangram, gimVI, and SpaGE clearly outperformed the other

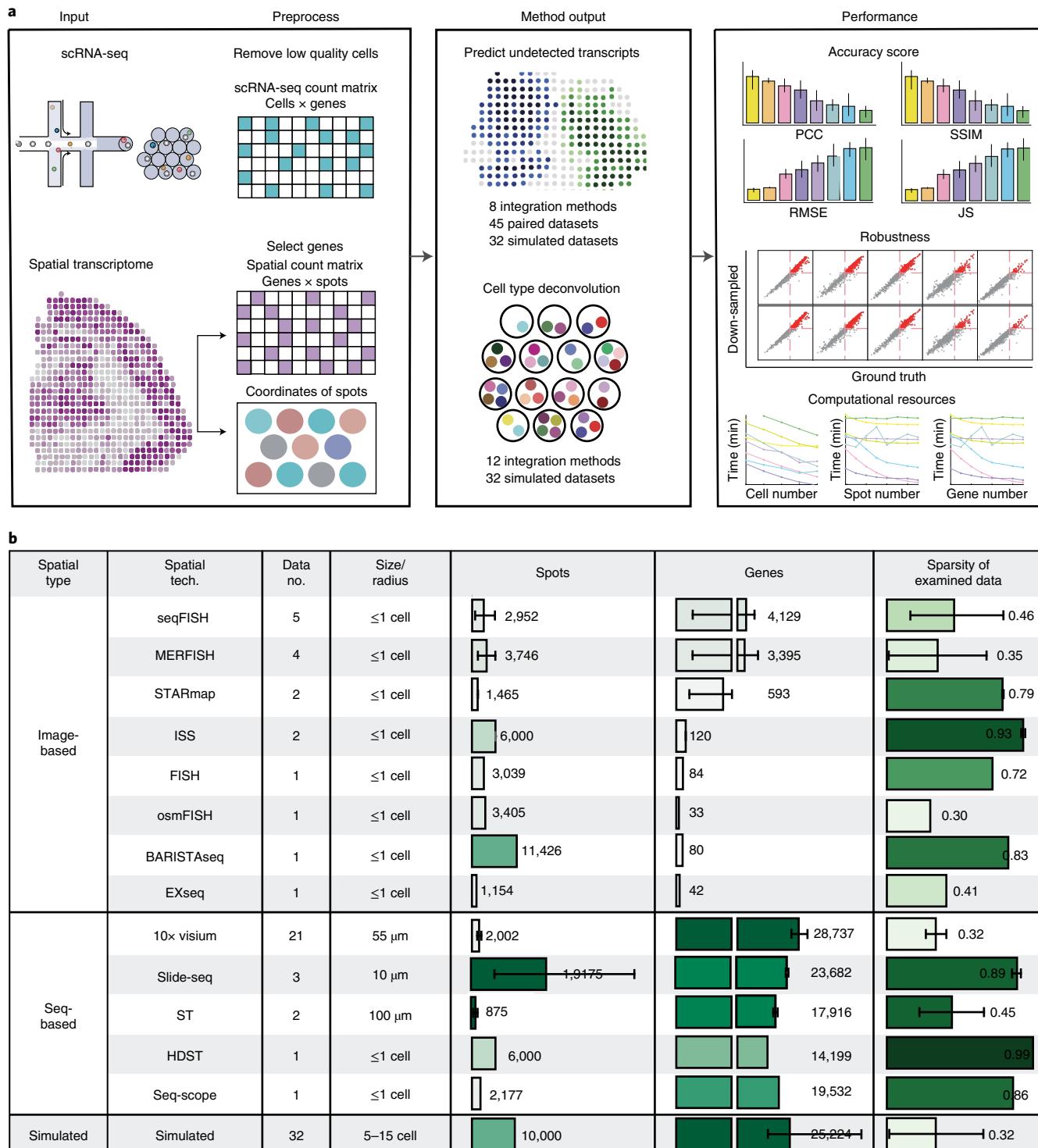


Fig. 1 | Benchmarking workflow and summary characteristics of the examined paired datasets. **a**, Schematic overview of the benchmarking workflow used to compare the performance of the integration methods for paired spatial transcriptomics and scRNA-seq datasets. We used the 16 integration methods to combine the spatial and single-cell transcriptomics data, and then compared their performance for (1) predicting the spatial distribution of RNA transcripts and (2) deconvoluting cell types of a histological spot. We also assessed the computational resources consumed by the integration methods. **b**, Information for the 45 paired datasets and 32 simulated datasets used in this study: each dataset contains both spatial transcriptomic data and scRNA-seq data for the same tissue. Detailed information for the data source is presented in the Methods and in Supplementary Table 1.

integration methods. Specifically, we found that the average PCC/SSIM of Tangram, gimVI, and SpaGE were 0.54/0.45, 0.52/0.43, and 0.49/0.39, higher than the PCC/SSIM values for the other 5 methods,

and the average RMSE/JS of these three methods were 0.94/0.18, 0.97/0.19, and 0.99/0.21, lower than the average RMSE/JS for the others (Fig. 2c). Moreover, the average AS for the Tangram, gimVI,

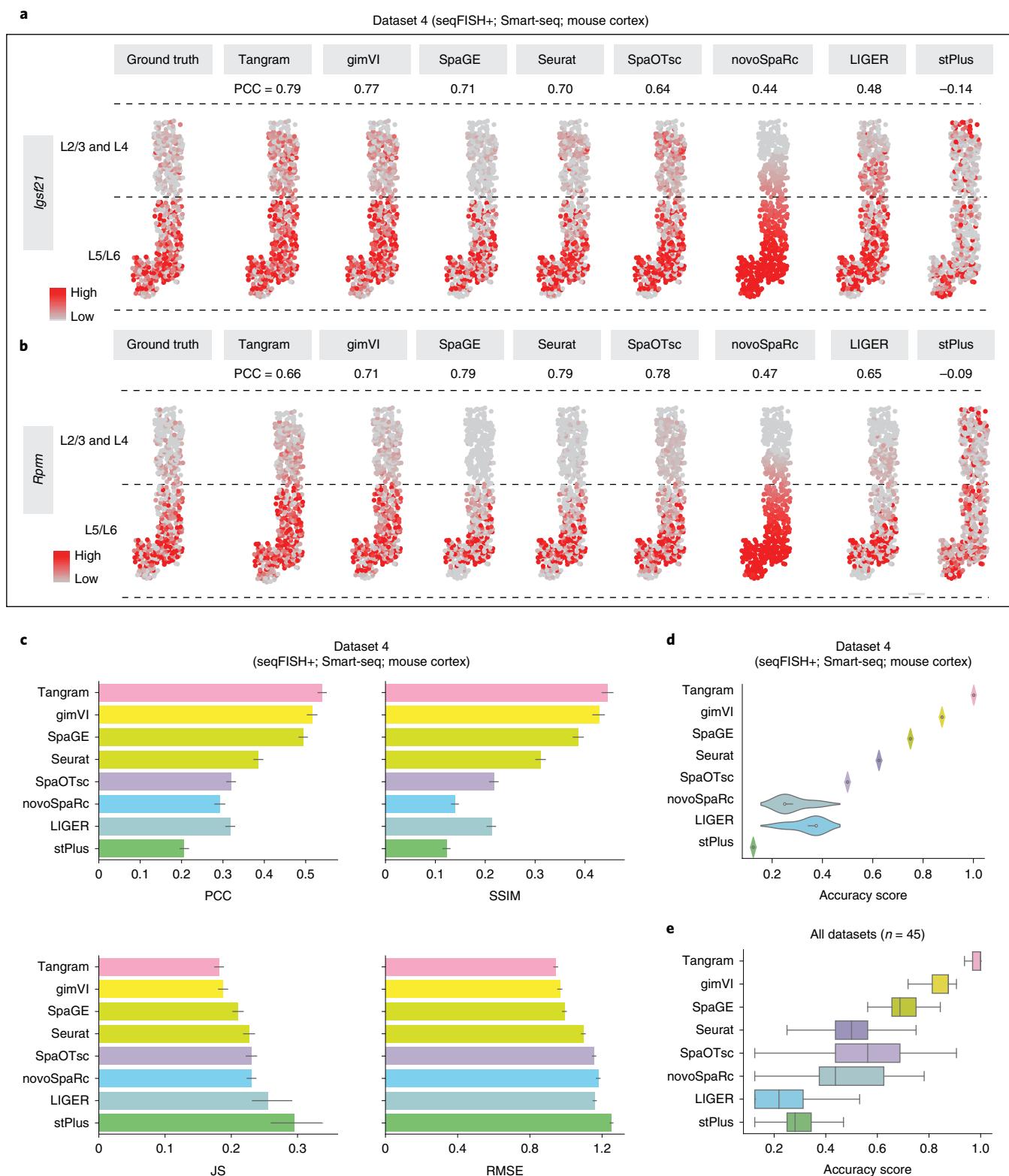


Fig. 2 | Comparing the accuracy of eight integration methods capable of predicting the spatial distribution of RNA transcripts. **a,b**, The spatial distribution of *lgsf21* (**a**) and *Rprm* (**b**) in dataset 4 (seqFISH+; Smart-seq; mouse cortex), including the ground truth and the predicted result from each of the integration methods. PCC, Pearson correlation coefficient between the expression vector of a transcript in the ground truth and that of the predicted result. **c**, The bar plots of PCC, SSIM, RMSE, and JS of each integration method in predicting the spatial distribution of transcripts in dataset 4. Data are presented as mean values \pm 95% confidence intervals; $n = 1,000$ predicted genes. **d**, The violin plot of AS (which is aggregated from the PCC, SSIM, RMSE, and JS values; see Methods) of the 8 integration methods for transcripts in dataset 4. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range; $n = 4$ benchmark metrics. **e**, Boxplots of AS of the 8 integration methods for all 45 paired datasets. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range; $n = 45$ independent datasets.

and SpaGE predictions were 1.0, 0.875, and 0.75, higher than that of the other methods (Fig. 2d). We also calculated the PCC, SSIM, RMSE, JS, and AS values of the prediction results for all transcripts in dataset 42 (as an example of the seq-based spatial transcriptomics approaches), and found that Tangram and gimVI outperformed the other integration methods on the basis of these metrics (Extended Data Fig. 1b,c).

To systematically assess the accuracy of the eight integration methods' predictions of the spatial distribution of undetected transcripts, we determined the PCC, SSIM, RMSE, JS, and AS values of their prediction results for all 45 paired datasets (Fig. 2e, Extended Data Fig. 2). The average ASs for the Tangram, gimVI, and SpaGE predictions were 0.96, 0.84, and 0.69, respectively, all of which exceed the AS values for Seurat (0.50), SpaOTsc (0.55), LIGER (0.25), novoSpaRc (0.47), and stPlus (0.31). Note that Tangram was still the best-performing integration method when we separately assessed the image-based datasets, the seq-based datasets, and the 32 simulated datasets, followed by gimVI and SpaGE (Extended Data Fig. 3a–c). Because 10X Visium, seqFISH, MERFISH, and Slide-seq have released more than 3 datasets, we further compared the ASs of the eight integration methods when processing data obtained using these four spatial transcriptomics technologies (Extended Data Fig. 3d–g). We found that Tangram, gimVI, and SpaGE outperformed other integration methods for data generated from 10X Visium, seqFISH, and MERFISH platforms, and Tangram and gimVI are top-ranked methods in processing Slide-seq datasets.

Several integration methods (for example, Seurat, LIGER, SpaGE, and stPlus) normalized the spatial transcriptomics data by default prior to integration. Here, we tested four schemes of input expression matrices: (1) raw expression matrix of spatial data and raw expression matrix of scRNA-seq data (R-R); (2) normalized expression matrix of spatial data and raw expression matrix of scRNA-seq data (N-R); (3) raw expression matrix of spatial data and normalized expression matrix of scRNA-seq data (R-N); and (4) normalized expression matrix of spatial data and normalized expression matrix of scRNA-seq data (N-N).

Interestingly, for 28 paired seq-based datasets, the transcript spatial distributions generated by Tangram, gimVI, SpaGE, Seurat, SpaOTsc, and LIGER have significantly higher PCC values when using an R-R and R-N input scheme than when using an N-R or N-N input scheme, and this trend was observed for 16 of the 28 paired datasets (P values < 0.01 , paired t -test) (Extended Data Figs. 4 and 5a); for SpaGE, Seurat, SpaOTsc, and novoSpaRc, the PCC values of the results with the R-R input scheme were higher than those with the other input schemes in 19 of the 28 paired datasets (P values < 0.01), and stPlus generated results with higher PCC values when using the R-R input scheme than the N-N input scheme in 18 of the 28 paired datasets (P value < 0.05). For 15 paired image-based datasets (Extended Data Figs. 4 and 5b), the transcript spatial distributions generated by Tangram, gimVI, SpaGE, and Seurat have higher PCC values when using the R-R or R-N input scheme than when using the N-R or N-N input scheme (11 out of 15 datasets, P value < 0.05); SpaGE, Seurat, and LIGER have higher PCC values when using the R-R input scheme than when using the other input schemes (11 out of 15 datasets, P value < 0.05); SpaOTsc has a higher PCC value when using the R-R input schemes than when using the NN input scheme (12 out of 15 datasets, P value < 0.05). Nevertheless, it should be emphasized that regardless of what input scheme was used, Tangram invariably outperformed the other integration methods (Extended Data Fig. 5c–f).

Impact of matrix sparsity. Notably, for datasets 12, 13, 40, and 44, all eight integration methods had low accuracy in predicting the spatial distribution of transcripts (that is, average PCC/SSIM < 0.3 , Extended Data Fig. 2). We investigated this apparently poor performance of the integration methods for these datasets by calculating

correlation coefficients between the four metrics (PCC, SSIM, RMSE, and JS) and considered several features of the spatial transcriptomics datasets, including the sparsity of the expression matrix (the sparsity of the spatial transcriptomics and scRNA-seq data is defined as the percentage of zero elements in the expression matrix), the number of detected genes, the number of detected spots, and the number of genes per spot. Ultimately, we found that the JS values of the 8 methods all linearly increased along with the rising of the sparsity of expression matrices (P values $< 1 \times 10^{-6}$, coefficient of determination (R^2) ≥ 0.50) (Extended Data Fig. 6).

To further characterize the impact of matrix sparsity, we next evaluated the performance of each integration method when inputting a very sparse spatial expression matrix (down-sampled from high quality datasets where sparsity was lower than 0.7). Specifically, we examined spatial transcriptomics datasets that captured $>1,000$ genes from >100 spots as high quality. To simulate expression matrices with 'high sparsity', we adopted Splatter⁶⁵ and Scuttle⁶⁶ to down-sample the non-zero elements from the original expression matrices to varying extents (Methods). We then used the original and down-sampled expression matrices as the inputs for the eight integration methods.

First, we evaluated the impact of the expression matrix sparsity in predicting the spatial distribution of known marker genes (Methods). Drew et al. reported that *Cplx1* is highly expressed in layer L5 of the cortex⁶⁷. Examining *Cplx1* in both the original and down-sampled data (down-sampling rate = 0.8) of dataset 4, we observed that the spatial distributions of *Cplx1* predicted by Tangram, gimVI, and SpaGE each had PCC values >0.7 for both the original and down-sampled data (Extended Data Fig. 7a).

We then assessed the performance of each integration method by counting the proportion of transcripts in a dataset exceeding a PCC threshold of 0.5 for both the original and down-sampled data, which we deemed the 'robustness score' (RS). For dataset 4, Tangram had the highest RS value (0.60), followed by gimVI (0.55) and then SpaGE (0.51) (Fig. 3a). Moreover, we noted that (1) the RS values decreased as the down-sampling rate increased and (2) the RS values for Tangram, gimVI, and SpaGE were consistently higher than those for Seurat, SpaOTsc, LIGER, novoSpaRc, and stPlus (Fig. 3b). A combined analysis which included the down-sampled data from 19 datasets again highlighted the strong performance of Tangram, gimVI, and SpaGE: even when the down-sampling rate reached 0.8, the average RS values of these three methods remained >0.50 (Fig. 3c, down-sampled by Splatter; Extended Data Fig. 7b,c, down-sampled by Scuttle). In summary, Tangram, gimVI, and SpaGE outperformed other integration methods in predicting the spatial distribution of transcripts for highly sparse datasets.

Performance of methods in cell type deconvolution. A common issue encountered when using spatial transcriptomics approaches like 10X Visium and ST is that each spot from a histological section may contain multiple cells, so it can be impossible to correctly assign the cell type composition of each spot. As noted above, Seurat, SpaOTsc, Tangram, and novoSpaRc are capable of assigning each cell from a scRNA-seq analysis to a spot from a spatial transcriptomics analysis, implying that they can be used to deconvolute the cell types of each spot. Moreover, Cell2location, SpatialDWLS, RCTD, Stereoscope, DestVI, STRIDE, SPOTlight, and DSTG were also designed for this purpose.

To compare the performance of the 12 integration methods in predicting the cell type composition of spots, we simulated this 'multi-cell spot problem' experienced with ST and 10X Visium datasets by 'gridding' a dataset that did not have this problem (dataset 10, acquired using STARmap; Smart-seq; mouse visual cortex). The cell type composition of each spot in dataset 10 has been reported and can be used as the ground truth when simulating a dataset with potentially ambiguous cell type assignments in each spot (Fig. 4a and

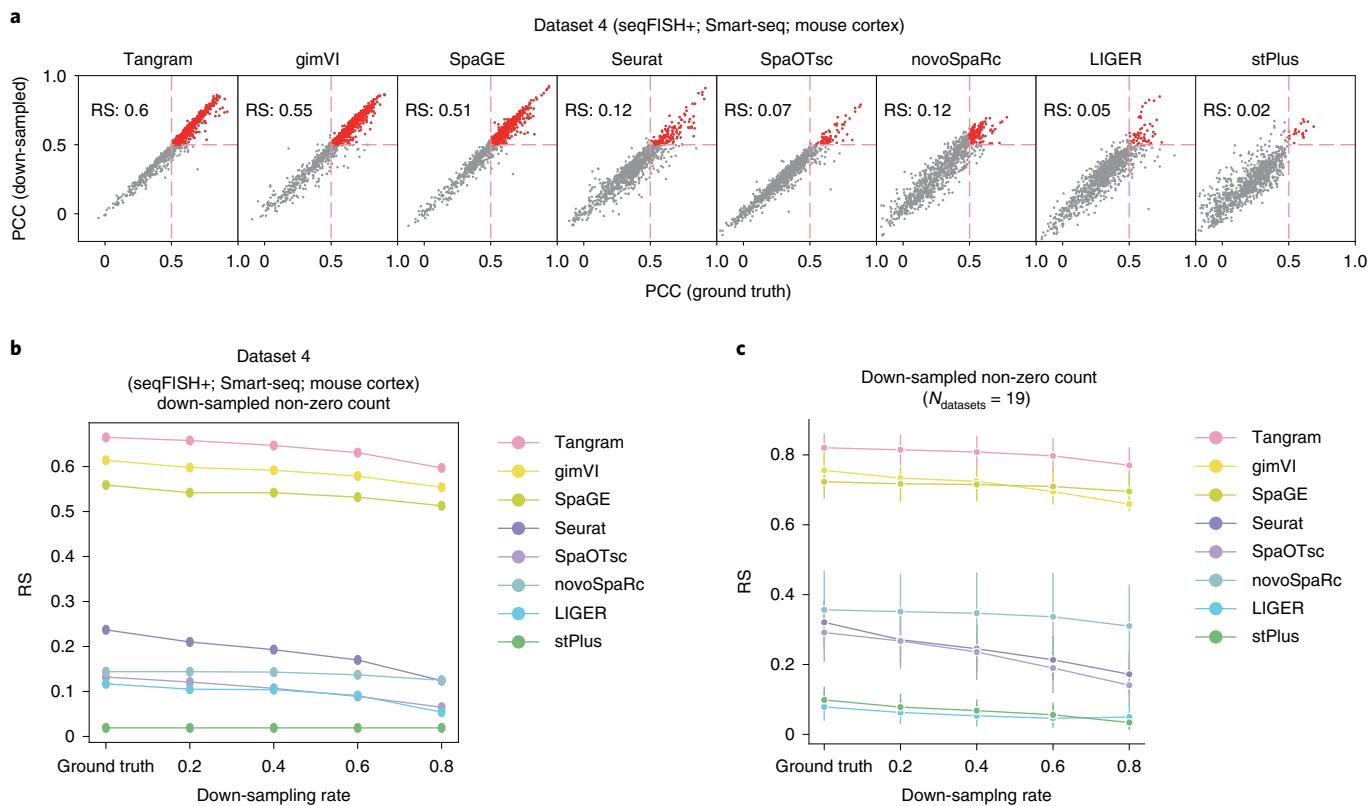


Fig. 3 | Comparing the accuracy of the eight integration methods for sparse spatial expression matrices down-sampled from the original datasets using Splatter. **a**, PCC of the spatial distribution of transcripts predicted from the original data and down-sampled data from dataset 4. The PCC values of the red-colored transcripts are greater than 0.5 for both the original and the down-sampled data. The proportion of the red-colored transcripts in all transcripts was defined as the RS. **b**, RS values of the 8 integration methods when processing sparse expression matrices down-sampled from dataset 4 at different down-sampling rates. **c**, RS values of the eight integration methods when processing the sparse expression matrices of the down-sampled datasets. The original datasets (used to generate the down-sampled datasets) capture >1,000 genes from >100 spots, and the sparsity of the expression matrices is <0.7. Data are presented as mean values \pm 95% confidence intervals; $n=19$ independent datasets.

Methods). The original dataset 10 captured 1,549 cells, corresponding to 15 cell types. After gridding, the simulated data had 189 spots, with each spot containing 1–18 cells. We plotted the locations of L4 excitatory neurons and found that RCTD and Stereoscope performed better in terms of the PCC values (0.87), followed by Tangram (0.85), Cell2location (0.83), STRIDE (0.80), SPOTlight (0.79), Seurat (0.76), SpaOTsc (0.74), and DSTG (0.71) (Fig. 4b). We then employed PCC, SSIM, RMSE, JS, and AS metrics to quantify the accuracy of the 12 integration methods in predicting the cell type composition of spots in gridded dataset 10 (Fig. 4c and Extended Data Fig. 8a). RCTD had the highest AS score (0.94), followed by Stereoscope (0.92).

We also performed the same analysis on dataset 4 (seqFISH+; Smart-seq; mouse cortex), which contains 524 cells of 14 cell types. After ‘gridding’, the simulated dataset had 72 spots (Extended Data Fig. 8b). Using the ground truth of the locations for the L5/6 excitatory neurons, we found that SpatialDWLS, RCTD, Tangram, Cell2location, and Stereoscope had PCC values of 0.88, 0.86, 0.85, 0.83, and 0.81 for the assignations of the L5/6 excitatory neurons, higher than other integration methods (Extended Data Fig. 8c). Moreover, in the prediction results for all cell types of dataset 4, SpatialDWLS, Tangram, and RCTD had the top 1, 2, and 3 ranking AS values (1.0, 0.92, and 0.83 respectively), followed by Cell2location (0.67) and Stereoscope (0.65) (Fig. 4d).

We further quantified the performance of these integration methods in cell type deconvolution of spots in the 32 simulated datasets that were synthesized from scRNA-seq datasets (Supplementary Table 2 and 3 and Methods). As the cell type information of each cell

in these scRNA-seq datasets has been reported by the data source papers, the cell type composition of a simulated spot can be inferred from the cells it contains. Note that novoSpaRc and SpaOTsc require spatial location information for each spot and thereby were excluded because spatial location information was not available in the simulated datasets. We used the 32 simulated datasets as the ground truth to assess the performance of the remaining 10 integration methods (including Seurat, Tangram, Cell2location, SpatialDWLS, RCTD, Stereoscope, DestVI, STRIDE, SPOTlight, and DSTG) in deconvoluting cell types in spots. We found that the average PCC and SSIM values of Cell2location, SpatialDWLS, and STRIDE are 0.83/0.75, 0.78/0.71, and 0.83/0.69, higher than those of the other integration methods, and the average RMSE and JS values of the Cell2location, SpatialDWLS, RCTD, and STRIDE are 0.08/0.33, 0.10/0.32, 0.096/0.37, and 0.11/0.37, lower than those of the other integration methods (Extended Data Fig. 8e). We also used the aggregation of the four metrics (that is, the AS score) to rank the performance of these integration methods in predicting cell type composition of spots, and we found that Cell2location, SpatialDWLS, RCTD, and STRIDE outperformed the other integration methods (Fig. 4e).

Computational resources. We used all the 45 paired datasets to compare the computational resources consumed by the 8 integration methods that can predict the spatial distribution of undetected transcripts (Supplementary Table 4). We used an identical CPU platform (2.2 GHz, 45 MB L3 cache, 144 CPU cores) to test each method. We are aware that gimVI and Tangram can support GPU

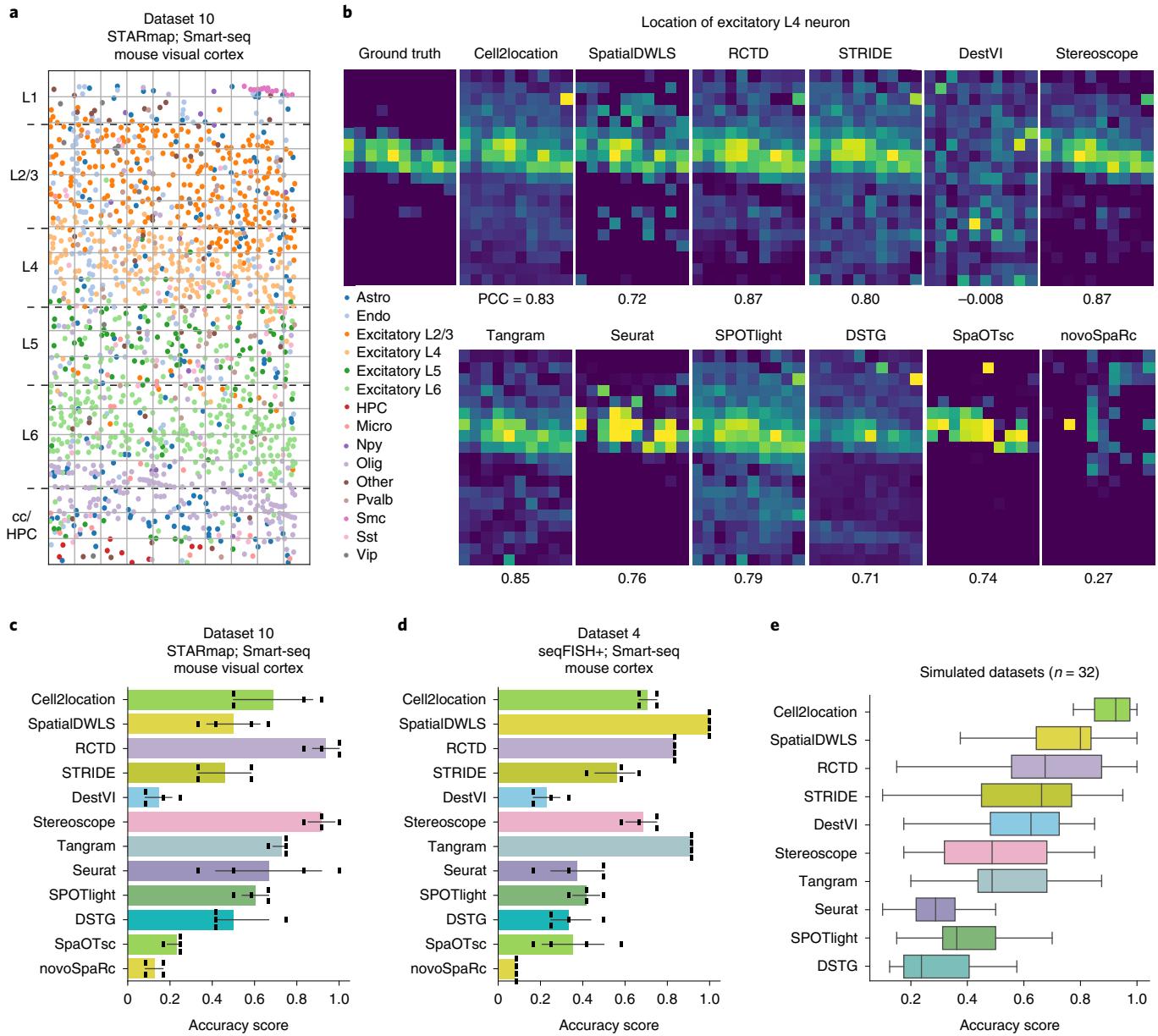


Fig. 4 | Comparing the performance of the 12 integration methods capable of deconvoluting cell types of each histological spot. **a**, A STARmap slide of dataset 10 (STARmap; Smart-seq; mouse visual cortex), with cells annotated by cell types. Each grid represents a simulated spot containing multiple cells. L1, 2/3, 4, 5, and 6: layer 1, 2/3, 4, 5, and 6; Excitatory L2/3, 4, 5, and 6: excitatory neurons in layer 2/3, 4, 5, and 6; Npy, Pvalb, Sst, and Vip: GABAergic interneuron subtypes marked by Npy, Pvalb, Sst, and Vip; Astro: astrocytes; Endo: endothelia cells; Micro: microglia; Oligo: oligodendrocytes; Smc: smooth muscle cells; Other: other unclassified cells. **b**, The proportion of L4 excitatory neurons in the spots simulated from dataset 10, including the ground truth and the predicted results of 12 integration methods. **c,d**, Bar plots of AS (aggregated from PCC, SSIM, RMSE, and JS; see Methods) of the cell type composition of the histological spots simulated from dataset 10 (**c**) and dataset 4 (**d**), predicted by 12 integration methods. Data are presented as mean values \pm 95% confidence intervals; $n=4$ benchmark metrics. **e**, Boxplots of AS of the 10 integration methods for all the 32 simulated datasets. SpaOTsc and novoSpaRc are excluded, as they require spatial location information for each spot, which is not available in the simulated datasets. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range; $n=32$ independent datasets.

processing; however, these two integration methods reported memory errors on our GPU platform (NVIDIA Tesla K80 with 12 GB memory) when processing the largest dataset 40 (19,522 spots in the spatial transcriptomics data and 26,252 cells in the scRNA-seq data). Notably, it took Seurat and LIGER less than 10 minutes of CPU time to process each dataset, and Tangram and LIGER consumed less than 32 GB of memory.

We then assessed the impacts of various data attributes (including the number of cells in scRNA-seq data, the number of spots in spatial

data, and the number of genes used for training) on the computational resources consumed by those eight integration methods. By down-sampling the number of cells and the number of spots in dataset 40 and the number of training genes in dataset 6, we found that Seurat was invariably the most computationally efficient method among the 8 integration methods for the prediction of the spatial distribution of undetected transcripts (Extended Data Fig. 9a–c).

To compare the computational resources consumed by the 10 integration methods that can deconvolute the cell types of spots,

we used a large simulated dataset (Methods) that contains 10,000 cells, 20,000 spots, and 56 cell types. For this dataset, Cell2locations reported memory errors on our GPU platform. Seurat and Tangram took less than 30 minutes of CPU time, and Stereoscope, Tangram, and DestVI consumed less than 8 GB of memory (Extended Data Fig. 9d). We then evaluated the impacts of the number of cells in scRNA-seq data, the number of spots in spatial data, and the number of cell types on computing time consumed by those 10 integration methods, and found that Tangram and Seurat are the top two most-efficient methods for processing cell type deconvolution of spots (Extended Data Fig. 9e–g).

Discussion

In this study, we benchmarked the performance of 16 integration methods capable of combining spatial transcriptomics data and single-cell transcriptomics data. We found that Tangram, gimVI, and SpaGE outperformed other integration methods for predicting the spatial distribution of transcripts, whereas Cell2location, SpatialDWLS, and RCTD were superior to other integration methods for cell type deconvolution of spots in histological sections. Our study helps researchers to choose appropriate tools and to optimize data-analysis workflows to accurately and efficiently integrate spatial transcriptomics data with scRNA-seq data. We have also provided a benchmark pipeline (<https://github.com/QuKunLab/SpatialBenchmarking>) and an instructive table (Supplementary Table 5) summarizing the properties and performance of all the benchmarked methods to guide researchers select suitable tools that match their data combinations.

Methods constructed on the basis of probabilistic models combined with negative binomial or Poisson distributions, such as gimVI, Cell2location, and RCTD, generally perform better at predicting the spatial distribution of transcripts or deconvolving cell types of spots. A deep learning algorithm was also applied in several integration methods, among which Tangram is one of the best-performing methods in predicting spatial distribution of the undetected transcripts. Technically, Tangram employs non-convex optimization in the model and selects only the optimal subset of scRNA-seq observations in the loss function. A combination of these measures may help improve the predictive power of the tools.

One observation from our comparative analysis is that the sparsity of the spatial transcriptomics expression matrix seriously affects the performance of the eight integration methods that predict the spatial distribution of RNA transcripts. There are multiple tactics that can be used to combat this sparsity issue for spatial transcriptomics expression matrices: researchers can increase the depth of sequencing, screen spots and genes with strict cut-off values to reduce the sparsity of the filtered expression matrix, or consider applying imputation algorithms (for example, SAVER⁶⁸, MAGIC⁶⁹, and WEDGE⁷⁰) to impute the zero elements in the expression matrix.

Another potential application of spatial transcriptomics is to predict ligand–receptor interactions between two cell types that are spatially close to each other. Many analytical tools have been developed for this task, such as SpaOTsc²¹, Giotto⁷¹, CellChat⁷², NicheNet⁷³, ICELLNET⁷⁴, and SingleCellSignalR⁷⁵. However, the vast discrepancies in the results from different methods make informative comparison difficult. For instance, only a small proportion (<5%) of the predicted ligand–receptor interactions were shared in the results for >3 methods (<https://github.com/QuKunLab/SpatialBenchmarking/tree/main/FigureData>). Benchmarking analysis may thereby rely on more experimental validation of cell–cell interactions in the future.

The wide diversity and fast-moving state of scRNA-seq technologies (Drop-seq, Smart-seq, and 10X Chromium) and spatial sequencing technologies (FISH, osmFISH, seqFISH, MERFISH, STARmap, ISS, EXseq, BaristaSeq, ST, 10X Visium, Slide-seq, Seq-scope, and HDST) complicates the task of this benchmarking analysis. We adopted three distinct perspectives to overcome the

challenges caused by the diversity of sequencing technologies: (1) we divided the spatial transcriptome datasets into two categories (seq-based technologies and image-based technologies); (2) we added an independent comparison of the integration methods for the set of four spatial transcriptomics technologies (MERFISH, seqFISH, Slide-seq, and 10X Visium) that have to date released more than three datasets; and (3) we used several intrinsic parameters (for example, the number of captured genes, the number of captured spots, and the sparsity of the expression matrix) to characterize the datasets generated by different spatial transcriptomics technologies. There are still aspects (that is, different numbers of genes and different spatial organization) that may affect the performance of the integration methods and the user's expectations. Nevertheless, on the basis of the current collection of datasets, we found that the performance rankings of these integration methods are barely affected by the spatial transcriptome technologies that generated these data.

Advances in spatial transcriptomics technologies, such as new versions of 10X Visium and BGI Stereo-seq⁷⁶, may enable detection of transcriptomic information for spots with diameters much smaller than cell size. However, in the near future, each spot still might not correspond exactly to a single cell when using these technologies. Moreover, considering that there is a large amount of publicly available spatial transcriptomics datasets that are highly valuable to different research communities, and the intense research efforts using spatial transcriptomics technologies that are now underway, we contend that there will be strong interest in integrating spatial transcriptomics and scRNA-seq data to determine the spatial distribution of cells or undetected transcripts.

In summary, this study presents a much-needed independent comparison of available integration methods of spatial transcriptomics and scRNA-seq data to determine the cell type deconvolution or the spatial distribution of undetected genes. Our results are also a useful resource for biologists who want to analyze their spatial transcriptomics data or methods developers who want to improve state-of-the-art technology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-022-01480-9>.

Received: 10 June 2021; Accepted: 30 March 2022;

Published online: 16 May 2022

References

- Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat. Neurosci.* **24**, 425–436 (2021).
- Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660 e1619 (2019).
- Moncada, R. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
- Ji, A. L. et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* **182**, 497–514 e422 (2020).
- Eng, C. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
- Codeluppi, S. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
- Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
- Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Visium spatial gene expression (10x Genomics, 2020).
- Stickels, R. R. et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
- Rodrigues, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

12. Lopez, R. et al. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *ICML Workshop on Computational Biology* (2019).
13. Abdelaal, T., Mourragui, S., Mahfouz, A. & Reinders, M. J. T. SpaGE: spatial gene enhancement using scRNA-seq. *Nucleic Acids Res.* **48**, e107 (2020).
14. Mourragui, S., Loog, M., van de Wiel, M. A., Reinders, M. J. T. & Wessels, L. F. A. PRECISE: a domain adaptation approach to transfer predictors of drug response from pre-clinical models to tumors. *Bioinformatics* **35**, i510–i519 (2019).
15. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
16. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e1821 (2019).
17. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
18. Welch, J. D. et al. Single-Cell Multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887 e1817 (2019).
19. Yang, Z. & Michailidis, G. A non-negative matrix factorization method for detecting modules in heterogeneous omics multi-modal data. *Bioinformatics* **32**, 1–8 (2016).
20. Nitzan, M., Karaikos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).
21. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
22. Villani, C. *Optimal Transport: Old and New* Vol. 338 (Springer, 2009).
23. Chen, S. Q., Zhang, B. H., Chen, X. Y., Zhang, X. G. & Jiang, R. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics* **37**, I299–I307 (2021).
24. Kleshcheynikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **1**–11, <https://doi.org/10.1038/s41587-021-01139-4> (2022).
25. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2021).
26. Dong, R. & Yuan, G. C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.* **22**, 145 (2021).
27. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 565 (2020).
28. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
29. Song, Q. Q. & Su, J. DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief. Bioinform.* **22**, bbaa414 (2021).
30. Sun, D., Liu, Z., Li, T., Wu, Q. & Wang, C. STRIDE: accurately decomposing and integrating spatial transcriptomics using single-cell RNA sequencing. *Nucleic Acids Res.* gkac150 (2022).
31. Lopez, R. et al. Multi-resolution deconvolution of spatial transcriptomics data reveals continuous patterns of inflammation. *Nat. Biotechnol.* in press (2022).
32. Karaikos, N. et al. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).
33. Berkeley Drosophila Transcription Network Project. <http://bdtnp.lbl.gov:8080/Fly-Net/>.
34. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
35. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
36. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. USA* **116**, 19490–19499 (2019).
37. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
38. Joglekar, A. et al. A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun.* **12**, 463 (2021).
39. Navarro, J. F. et al. Spatial transcriptomics reveals genes associated with dysregulated mitochondrial functions and stress signaling in alzheimer disease. *iScience* **23**, 101556 (2020).
40. Lohhoff, T. et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* **40**, 74–85 (2022).
41. Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
42. Han, X. et al. Mapping the mouse cell atlas by microwell-Seq. *Cell* **172**, 1091–1107 e1017 (2018).
43. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
44. Brann, D. H. et al. Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Sci. Adv.* **6**, eabc5801 (2020).
45. Cho, C. S. et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559–3572 e3522 (2021).
46. Tabula Muris, C. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
47. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
48. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 e1016 (2018).
49. McCray, T. et al. Erratum: Vitamin D sufficiency enhances differentiation of patient-derived prostate epithelial organoids. *iScience* **24**, 102640 (2021).
50. Janosevic, D. et al. The orchestrated cellular and molecular responses of the kidney to endotoxin define a precise sepsis timeline. *eLife* **10**, e62270 (2021).
51. Melo Ferreira, R. et al. Integration of spatial and single-cell transcriptomics localizes epithelial cell-immune cross-talk in kidney injury. *JCI Insight* **6**, e147703 (2021).
52. Sanchez-Ferraz, O. et al. A coordinated progression of progenitor cell states initiates urinary tract development. *Nat. Commun.* **12**, 2627 (2021).
53. Wu, S. Z. et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **53**, 1334–1347 (2021).
54. Alon, S. et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, 481 (2021).
55. Chen, X., Sun, Y. C., Church, G. M., Lee, J. H. & Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* **46**, e22 (2018).
56. Boooshaghi, A. S. et al. Isoform cell-type specificity in the mouse primary motor cortex. *Nature* **598**, 195–199 (2021).
57. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
58. Tepe, B. et al. Single-cell RNA-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell Rep.* **25**, 2689–2703 e2683 (2018).
59. Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat. Commun.* **12**, 6278 (2021).
60. McKellar, D. W. et al. Large-scale integration of single-cell transcriptomic data captures transitional progenitor states in mouse skeletal muscle regeneration. *Commun. Biol.* **4**, 1280 (2021).
61. Ratz, M. et al. Clonal relations in the mouse brain revealed by single-cell and spatial transcriptomics. *Nat. Neurosci.* **25**, 285–294 (2022).
62. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
63. Ramskold, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
64. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
65. Zappia, L., Phipson, B. & Oshlack, A. Splatter: simulation of single-cell RNA sequencing data. *Genome Biol.* **18**, 174 (2017).
66. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).
67. Drew, C. J. G., Kyd, R. J. & Morton, A. J. Complexin 1 knockout mice exhibit marked deficits in social behaviours but appear to be cognitively normal. *Hum. Mol. Genet.* **16**, 2288–2305 (2007).
68. Huang, M. et al. SAVER: gene expression recovery for single-cell RNA sequencing. *Nat. Methods* **15**, 539–542 (2018).
69. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729 (2018).
70. Hu, Y. et al. WEDGE: imputation of gene expression values from single-cell RNA-seq datasets using biased matrix decomposition. *Brief Bioinform.* **22**, bbab085 (2021).
71. Dries, R. et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78 (2021).
72. Jin, S. Q. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
73. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
74. Noel, F. et al. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089 (2021).
75. Cabello-Aguilar, S. et al. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
76. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball patterned arrays. Preprint at bioRxiv <https://doi.org/10.1101/2021.01.17.427004> (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Preprocessing of datasets. We preprocessed each dataset with the following steps. (1) Removal of low-quality cells. For scRNA-seq data, we used Seurat with parameters ‘min.features = 200’ to remove cells for which fewer than 200 RNAs were captured. (2) Normalization of the expression matrix. For spatial transcriptomics datasets, we tested both the non-normalized and normalized expression matrices for input of each integration method. To normalize the expression matrices, we used the following equation:

$$D_{ij} = \log \left(\bar{N} \times \frac{C_{ij}}{\sum_{j=1}^M C_{ij}} + 1 \right) \quad (1)$$

where C_{ij} represents the raw read count for gene i in spot j , D_{ij} represents the normalized read count for gene i in spot j , and \bar{N} is the median number of detected transcripts per cell. For scRNA-seq datasets, we normalized their expression matrix using the function ‘NormalizeData’ and default parameters in Seurat 3.2. (3) Selection of highly variable genes. For spatial transcriptomic datasets with more than 1,000 detected genes, we calculated the coefficient of variation of each gene using the following equation:

$$CV_i = \frac{\sigma_i}{\bar{u}_i} \quad (2)$$

where CV_i is the coefficient of variation of gene i ; σ_i is the s.d. of the spatial distribution of gene i in all spots; and \bar{u}_i is the average expression of gene i in all spots. We identified 1,000 genes with the highest CV_i values as highly variable genes and used their overlap with detected genes in the corresponding scRNA-seq data to construct the ground truth for each dataset. For spatial transcriptomic datasets with fewer than 1,000 detected genes, we used genes detected in both spatial transcriptomics and scRNA-seq data to build the ground truth of each dataset.

Parameter settings for integration methods. We evaluated the performance of eight integration methods, which can predict the spatial distribution of undetected transcripts, using tenfold crossvalidation. For a set of genes in the spatial data, we divided the genes into 10 portions, and iteratively used nine portions of the genes for integration (that is, the reference gene set used for training); the remaining one portion of genes was used for prediction. The parameters of each integration method were set as described below for each program.

gimVI. We followed the instructions on the gimVI website: https://docs.scvi-tools.org/en/0.8.0/user_guide/notebooks/gimvi Tutorial.ipynb. The spatial distribution of genes was obtained using the model.get_imputed_values function with parameter normalized = False.

SpaGE. We followed the guidelines on the GitHub repository of SpaGE: https://github.com/tabdelala/SpaGE/blob/master/SpaGE_Tutorial.ipynb. If the number of genes used for integration was greater than 50 (that is, $N_{gene} > 50$), we set the parameter n_pv = $N_{gene} / 2$.

Tangram. We followed the instructions on the Tangram GitHub repository: <https://github.com/broadinstitute/Tangram>. we set the parameters as modes = ‘clusters’, density = ‘rna_count_based’.

Seurat. We followed the instructions on the Seurat 3.2 website: <https://satijalab.org/seurat/archive/v3.2/integration.html>. We set the parameter reduction = ‘cca’, k.filter = NA. If $N_{gene} > 30$, we set dims = 30, otherwise we set dims = N_{gene} . The predicted spatial distribution of genes was obtained using the Seurat function ‘TransferData’.

SpaOTsc. We followed the instructions on the SpaOTsc GitHub repository: <https://github.com/zcang/SpaOTsc>. The spatial distribution of genes was obtained using the function ‘issc.transport_plan’ with parameters alpha = 0, rho = 1.0, epsilon = 0.1, scaling = False.

novoSpaRc. We followed the guidelines on the GitHub repository of novoSpaRc: https://github.com/rajewsky-lab/novospaRc/blob/master/reconstruct_drosophila embryo tutorial.ipynb. We set the parameters as alpha_linear_r = 0.5, loss_fun = ‘square_loss’, epsilon = 5×10^{-3} . We trained novoSpaRc using the expression and spatial information of the training gene set, just as we did for the other methods, to ensure the fairness of the benchmarking study.

LIGER. We followed the instructions on the LIGER GitHub repository: <https://github.com/welch-lab/liger>. The predicted spatial distribution of genes was obtained using the function ‘imputeKNN’ with parameters norm = FALSE, scale = FALSE. If $N_{gene} > 30$, we set knn_k = 30, otherwise we set knn_k = N_{gene} .

stPlus. We followed the instructions on the LIGER GitHub repository: <http://github.com/xy-chen16/stPlus>. we set tmin = 5, neighbor = 50.

We then evaluated the performance of 12 integration methods that can deconvolute the cell types of histological spots. The parameters of each integration method were set as described below for each program.

Cell2location. We followed the guidelines on the Cell2location website: https://cell2location.readthedocs.io/en/latest/notebooks/cell2location_tutorial.html. The single-cell regression model was trained with parameters max_epochs = 250, lr = 0.002. The cell2location model was obtained with parameters max_epochs = 30,000.

RCTD. We followed the guidelines on the RCTD GitHub repository: <https://raw.githubusercontent.com/dmcable/spacexr/master/vignettes/spatial-transcriptomics.html>. We set doublet_mode = ‘full’.

DestVI. We followed the guidelines on the DestVI website: https://docs.scvi-tools.org/en/stable/tutorials/notebooks/DestVI_tutorial.html. The single-cell model was trained with parameters max_epochs = 250, lr = 0.001, number of training genes = 7,000. The spatial model was trained with parameters max_epochs = 2,500.

Tangram. We followed the instructions on the Tangram GitHub repository: <https://github.com/broadinstitute/Tangram>. We set the parameters as modes = ‘clusters’, density = ‘rna_count_based’. To deconvolute the cell types in space, we invoke ‘project_cell_annotation’ to transfer the annotation to space.

Seurat. We followed the instructions on the Seurat 3.2 website: <https://satijalab.org/seurat/archive/v3.2/integration.html>. We set the parameter dim = 1:30, normalization.method = ‘SCT’.

SpatialDWLS. We followed the guidelines on the SpatialDWLS website: https://rubd.github.io/Giotto_site/articles/tut7_giotto_enrichment.html. We set the parameter as n_cell = 20.

SPOTlight. We followed the guidelines on the SPOTlight GitHub repository: <https://marcelosua.github.io/SPOTlight/>. We set the parameter as transf = ‘uv’, method = ‘nsNMF’.

Stereoscope. We followed the guidelines on the website: https://docs.scvi-tools.org/en/stable/user_guide/models/stereoscope.html. The single-cell model was trained with parameters max_epochs = 100. The spatial model was trained with parameters max_epochs = 10,000.

STRIDE. We followed the guidelines on the STRIDE website: https://stridespatial.readthedocs.io/en/latest/tutorials/Mouse_embryo.html. We set the parameter as ‘normalize’.

DSTG. We followed the instructions on the DSTG GitHub repository: <https://github.com/Su-informatics-lab/DSTG>.

SpaOTsc. We followed the instructions on the SpaOTsc GitHub repository: <https://github.com/zcang/SpaOTsc>. The spatial distribution of genes was obtained using the function ‘issc.transport_plan’ with parameters alpha=0, rho=1.0, epsilon=0.1, scaling=False.

novoSpaRc. We followed the guidelines on the GitHub repository of novoSpaRc: https://github.com/rajewsky-lab/novospaRc/blob/master/reconstruct_drosophila embryo tutorial.ipynb. We set the parameters as alpha_linear_r = 0.5, loss_fun = ‘square_loss’, epsilon = 5×10^{-3} . We trained novoSpaRc using the expression and spatial information of the training gene set, just as we did for the other methods, to ensure the fairness of the benchmarking study.

Benchmark metrics. We constructed a common pipeline to evaluate the performance of the integration methods for the 45 paired datasets. In the pipeline, we used the following five metrics to assess each integration method.

1. **PCC.** The PCC value was calculated using the following equation:

$$PCC = \frac{E[(\tilde{x}_i - \bar{u}_i)(x_i - \bar{u}_i)]}{\sigma_i \tilde{\sigma}_i} \quad (3)$$

where x_i and \tilde{x}_i are the spatial expression vectors of gene i in the ground truth and the predicted result, respectively; u_i and \tilde{u}_i are the average expression value of gene i in the ground truth and the predicted result, respectively; and σ_i and $\tilde{\sigma}_i$ are the s.d. of the spatial expression of gene i in the ground truth and the predicted result, respectively. For one gene, a higher PCC value indicates better prediction accuracy.

2. **SSIM**⁷. We first scaled the expression matrix as follows, so that the expression value of each gene was between 0 and 1:

$$x'_{ij} = \frac{x_{ij}}{\max(\{x_{11}, \dots, x_{1M}\})} \quad (4)$$

where x_{ij} denotes the expression of gene i in spot j , and M is the total number of spots. Then we used the scaled gene expression and the following equation to calculate the SSIM value of each gene:

$$SSIM = \frac{(2\tilde{u}_i u_i + C_1^2)(2\text{cov}(x'_i, \tilde{x}'_i) + C_2^2)}{(\tilde{u}_i^2 + u_i^2 + C_1^2)(\tilde{\sigma}_i^2 + \sigma_i^2 + C_2^2)} \quad (5)$$

where the definitions of u_i , \tilde{u}_i , σ_i , and $\tilde{\sigma}_i$ are similar to those for calculating the PCC value (but for scaled gene expression); C_1 and C_2 are 0.01 and 0.03, respectively; and $cov(\mathbf{x}_i, \tilde{\mathbf{x}}_i)$ is the covariance between the expression vector of gene i in the ground truth (that is, \mathbf{x}_i') and that of the predicted result (that is, $\tilde{\mathbf{x}}_i'$). For one gene, a higher SSIM value indicates better prediction accuracy.

3. RMSE. We first calculated the z-score of the spatial expression of each gene for all spots, then used the following equation to calculate RMSE:

$$RMSE = \sqrt{\frac{1}{M} \sum_{j=1}^M (\tilde{z}_{ij} - z_{ij})^2} \quad (6)$$

where z_{ij} and \tilde{z}_{ij} are the z-score of the spatial expression of gene i in spot j in the ground truth and the predicted result, respectively. For one gene, a lower RMSE value indicates better prediction accuracy.

4. JS. JS uses relative information entropy (that is, Kullback–Leibler divergence) to determine the difference between two distributions. We first calculated the spatial distribution probability of each gene as follows:

$$P_{ij} = \frac{x_{ij}}{\sum_{j=1}^M x_{ij}} \quad (7)$$

where x_{ij} denotes the expression of gene i in spot j , M is the total number of spots, and P_{ij} is the distribution probability of gene i in spot j . We then calculated the JS value of each gene using the following equations⁷⁸:

$$JS = \frac{1}{2}KL\left(\tilde{\mathbf{P}}_i \middle| \frac{\tilde{\mathbf{P}}_i + \mathbf{P}_i}{2}\right) + \frac{1}{2}KL\left(\mathbf{P}_i \middle| \frac{\tilde{\mathbf{P}}_i + \mathbf{P}_i}{2}\right) \quad (8)$$

$$KL(\mathbf{a}_i || \mathbf{b}_i) = \sum_{j=0}^M (a_{ij} \times \log \frac{a_{ij}}{b_{ij}}) \quad (9)$$

where \mathbf{P}_i and $\tilde{\mathbf{P}}_i$ are the spatial distribution probability vectors of gene i in the ground truth and the predicted result, respectively, $KL(\mathbf{a}_i || \mathbf{b}_i)$ is the Kullback–Leibler divergence between two probability distribution \mathbf{a}_i and \mathbf{b}_i , and a_{ij} and b_{ij} are the predicted probability and real probability of gene i in spot j , respectively. For one gene, a lower JS value indicates better prediction accuracy.

5. AS. We defined AS by aggregating PCC, SSIM, RMSE, and JS to evaluate the relative accuracy of the integration methods for each dataset. For one dataset, we calculated the average PCC, SSIM, RMSE, and JS of all genes predicted by each integration method. Then we sorted the PCC and SSIM values of the integration methods in ascending order to get $RANK_{PCC}$ and $RANK_{SSIM}$; the method with the highest PCC/SSIM value will have $RANK_{PCC/SSIM} = N$, and the method with the lowest PCC/SSIM value will have $RANK_{PCC/SSIM} = 1$. We also sorted the RMSE and JS values of the integration methods in descending order to get $RANK_{RMSE}$ and $RANK_{JS}$; the method with the highest RMSE/JS value will have $RANK_{RMSE/JS} = 1$, and the method with the lowest RMSE/JS value will have $RANK_{RMSE/JS} = N$. Finally, we calculated the average value of $RANK_{PCC}$, $RANK_{SSIM}$, $RANK_{RMSE}$, and $RANK_{JS}$ to obtain the AS value of each integration method, as follows:

$$AS = \frac{1}{4} (RANK_{PCC} + RANK_{SSIM} + RANK_{RMSE} + RANK_{JS}) \quad (10)$$

For a dataset, the method with the highest AS value had the best performance among the integration methods.

Simulating ‘multi-cell-spot problem’ datasets. In order to obtain multi-cell-per-spot datasets with known cell compositions at each spot, we gridded single-cell resolution spatial transcriptomics datasets 4 and 10 to simulate datasets with potentially ambiguous cell type assignments per spot. For dataset 4 (seqFISH+; Smart-seq; mouse cortex), we defined a square with 500×500 pixels (~51.5 μm) as one spot-like region to grid the seqFISH+ slide, referring to the coarse-graining procedure introduced by SpatialDWLS. We summed the expression values of all cells in a grid to simulate a spot that may contain multiple cells, and took the center of the grid as the location of the spot. The simulated data of dataset 4 had 72 spots, and we calculated the percentage of cell types in each spot as the ground truth. For dataset 10 (STARmap; Smart-seq; mouse visual cortex), we used Seurat to cluster cells, and annotated the cell type of each cluster using marker genes^{79,80}. We used marker genes *Slc17a7* and *Gad1* to annotate excitatory neurons and inhibitory neurons, respectively. The L2/3, L4, L5 and L6 excitatory neurons (eL2/3, eL4, eL5, eL6) were annotated, respectively, by marker genes *Nov*, *Rorb*, *Sulf2*, and *Pcp4*. Moreover, the VIP, SST, and PV inhibitor neurons were annotated, respectively, by marker genes *Vip*, *Sst*, and *Pvalb*. The microglia, astrocytes, oligodendrocytes, smooth-muscle, and endothelial cells were annotated by marker genes *Pdgfra*, *Aqp4*, *Enpp2*, *Mgp*, and *Bsg*, respectively. We then used a 750-pixel window to grid the STARmap slide. We summed the expression values of all cells in a grid to simulate a spot that may contain multiple cells, and took the center of the grid as the location of the spot. The simulated data of dataset 10 have 189 spots in total, and we calculated the percentage of cell types in each spot as the ground truth.

We also used PCC, SSIM, RMSE, and JS to assess the accuracy of Seurat, SpaOTsc, Tangram, and novoSpaRc in assigning cells to spatial locations in histological sections. We first counted the proportions of various types of cells in each spot. Then we introduced the cell type proportion of each spot into Eq. 3–Eq. 9 to calculate PCC, SSIM, RMSE, and JS values, which quantified the similarity (PCC/SSIM) or difference (RMSE/JS) between the predicted results and the ground truth. Finally, we used the two-sided Mann–Whitney U test to calculate the statistical significance of the difference in the prediction accuracy between different methods.

Simulating spatial datasets using scRNA-seq datasets. For the 32 simulated datasets, we devised the generation procedure of spatial transcriptomics data simulation by referring to the algorithm introduced by RCTD and Stereoscope. For each simulated spot, we first sampled cell numbers (N_c) in a uniform distribution in the range 5–15, and sampled the number of cell types (N_t) in a uniform distribution in the range 2–6. Then we assumed that these cell types have equal distribution possibility $P=1/N_t$ in the spot, and randomly assigned cells from each cell types of the scRNA-seq data to the spot. To obtain the gene expression values at each spatial location, we summed the gene expression values of all cells in one spot. Referring to the method for constructing simulated datasets used in RCTD, we used Scuttle (<http://bioconductor.org/packages/release/bioc/html/scuttle.html>) to down-sample the number of counts per spot to 10% of the original value. We can obtain the percentage of a cell type at each spot by counting the number of cells corresponding to the cell type. For the large simulated dataset used for assessing the efficiency of each integration method, we used the same algorithm as above, but set the number of spots to 20000, the number of cells to 10000, and the number of cell types to 56.

Simulating datasets with high sparsity. We used 19 of the 45 paired datasets (that is, datasets 4, 7, 18, 19, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 36, 37, 38, 42) to test the impact of the sparsity of expression matrices for each integration method. To simulate a dataset with high sparsity, we applied the Scuttle package (<http://www.bioconductor.org/packages/release/bioc/html/scuttle.html>) for the down-sampling of the spatial expression matrices of the datasets, and we also used the Splatter package to down-sample the datasets (down-sample rate = 0.2, 0.4, 0.6, and 0.8). We down-sampled each dataset 10 times at different rates to avoid errors caused by random selection. To quantify the impact of expression matrix sparsity for each integration method, we counted the percentage of genes whose PCC values of the spatial distribution predicted from the original data and the down-sampled data were both greater than 0.5, which was defined as the robustness score.

Computer platform. We ran CPU tests of the 16 integration methods on a computer cluster with four Intel Xeon E78860v4 CPUs (2.2 GHz, 45 MB L3 cache, 144 CPU cores in total) and 1 TB memory (DDR4 2,400 MHz). The GPU tests for gimVI and Tangram were performed on a computer with Intel Xeon E5-2680v4 CPU (2.4 GHz, 35 MB L3 cache, 14 CPU cores in total), 128 GB memory, and NVIDIA Tesla K80 GPU (12 GB of memory, a total of 2496 CUDA cores).

To assess the impact of various data attributes (including the number of cells in scRNA-seq data, the number of spots in spatial data, and the number of genes used for training) on the computing resources consumed by those 8 integration methods capable of predicting the spatial distribution of undetected transcripts, we down-sampled the number of cells and the number of spots in dataset 40 and down-sampled the number of shared genes in dataset 6. For the 10 integration methods that can perform cell type composition prediction of spots, we simulated a large dataset (10,000 spots, 20,000 cells) to evaluate the computer resources consumed by each method. Then, we down-sampled the number of cells, the number of spots, and the number of cell types in simulated datasets, and then evaluated the impacts of these data attributes on computing resources consumed.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

A summary of the individual accession numbers is given in Supplementary Table 1. The raw data are available from following study:
Dataset 1 (mouse gastrulation): seqFISH, <https://content.cruk.cam.ac.uk/jmlab/SpatialMouseAtlas2020/>; 10X Chromium, ‘Sample 21’ in MouseGastrulationData within the R/Bioconductor data package MouseGastrulationData.
Dataset 2 (mouse embryonic stem cell): seqFISH, <https://zenodo.org/record/3735329#.YY69HZMza3j>; Microwell-Seq, ‘EmbryonicStemCells’ in ‘MCA_BatchRemoved_Merge_dge.h5ad’ file in https://figshare.com/articles/dataset/MCA_DGE_Data/5435866.
Dataset 3 (mouse hippocampus): seqFISH, <https://ars.els-cdn.com/content/images/e1-s2.0-S0896627316307024-mm6.xlsx>; 10X Chromium, ‘HIPP_sc_Rep1_10X sample’ in GSE158450 in the GEO database.
Dataset 4 (mouse cortex): seqFISH+, <https://github.com/CaiGroup/seqFISH-PLUS>, and the spatial coordinate of each spot was generated using ‘stitchFieldCoordinates’ function in Giotto; Smart-seq, mouse primary visual

cortex (ViSp) in the dataset in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 5 (mouse olfactory bulb): seqFISH+, <https://github.com/CaiGroup/seqFISH-PLUS>; Drop-seq, [GSE148360](https://doi.org/10.1101/148360) in the GEO database.

Dataset 6 (mouse hypothalamic preoptic region): MERFISH, the eighteenth female parent mouse (animal ID = 18) in <https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248>; 10X Chromium, [GSE113576](https://doi.org/10.1101/113576) in the GEO database.

Dataset 7 (human osteosarcoma): MERFISH, the ‘B1_cell’ used in https://www.pnas.org/doi/suppl/10.1073/pnas.1912459116/suppl_file/pnas.1912459116.sd12.csv; 10X Chromium, BC22 in [GSE152048](https://doi.org/10.1101/152048) in the GEO database.

Dataset 8 (mouse primary motor cortex): MERFISH, ‘mouse1_slice162’ in <https://caltech.box.com/shared/static/dzqt6ryytmjbygai356s1z0phntsbaol.gz>; 10X Chromium, https://data.nemoarchive.org/biccn/lab/zeng/transcriptome/scell/10x_v3/mouse/processed/analysis/10X_cells_v3_AIBS.

Dataset 9 (mouse ViSP): MERFISH, <https://github.com/spacex-spacejam/data/>; Smart-seq, mouse primary visual cortex (ViSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 10 (mouse visual cortex): STARmap, ‘20180505_BY3_1kgenes’ in <https://www.starmapresources.com/data>; Smart-seq, mouse primary visual cortex (ViSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 11 (mouse prefrontal cortex): STARmap, ‘20180419_BZ9_control’ in <https://www.starmapresources.com/data>; 10X Chromium, ‘PFC_sc_Rep2_10X’ in [GSE158450](https://doi.org/10.1101/158450) in the GEO database.

Dataset 12 (human middle temporal gyrus): ISS, <https://github.com/spacex-spacejam/data>; Smart-seq, <https://portal.brain-map.org/atlas-and-data/rnaseq/human-mtg-smart-seq>.

Dataset 13 (mouse ViSP): ISS, <https://github.com/spacex-spacejam/data>; Smart-seq, mouse primary visual cortex (ViSp) in the dataset in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 14 (*Drosophila* embryo): FISH, <https://github.com/rajewsky-lab/distmap>; Drop-seq, [GSE95025](https://doi.org/10.1101/GSE95025) in the Gene Expression Omnibus (GEO) database.

Dataset 15 (mouse somatosensory cortex): osmFISH, cortical regions in <http://linnarssonlab.org/osmFISH/>; Smart-seq, mouse somatosensory cortex (SSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-smart-seq>.

Dataset 16 (mouse ViSP): BaristaSeq, <https://github.com/spacex-spacejam/data>; Smart-seq, mouse primary visual cortex (ViSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 17 (mouse ViSP): ExSeq, <https://github.com/spacex-spacejam/data>; Smart-seq, mouse primary visual cortex (ViSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 18 (mouse hindlimb muscle): 10X Visium, Vis5A in [GSE161318](https://doi.org/10.1101/GSE161318) in the GEO database; 10X Chromium, D2_Ev3 in [GSE159500](https://doi.org/10.1101/GSE159500) in the GEO database.

Dataset 19 (mouse hindlimb muscle): 10X Visium, Vis9A in [GSE161318](https://doi.org/10.1101/GSE161318) in the GEO database; 10X Chromium, D7_Ev3 in [GSE159500](https://doi.org/10.1101/GSE159500) in the GEO database.

Dataset 20 (human breast cancer): 10X Visium, ‘CID3586’ in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in [GSE176078](https://doi.org/10.1101/GSE176078) in the GEO database.

Dataset 21 (human breast cancer): 10X Visium, ‘1160920F’ in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in [GSE176078](https://doi.org/10.1101/GSE176078) in the GEO database.

Dataset 22 (human breast cancer): 10X Visium, ‘CID4290’ in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in [GSE176078](https://doi.org/10.1101/GSE176078) in the GEO database.

Dataset 23 (human breast cancer): 10X Visium, ‘CID4465’ in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in [GSE176078](https://doi.org/10.1101/GSE176078) in the GEO database.

Dataset 24 (human breast cancer): 10X Visium, ‘CID44971’ https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in [GSE176078](https://doi.org/10.1101/GSE176078) in the GEO database.

Dataset 25 (human breast cancer): 10X Visium, ‘CID4535’ in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in [GSE176078](https://doi.org/10.1101/GSE176078) in the GEO database.

Dataset 26 (zebrafish melanoma): 10X Visium, ‘Visium-A’ in [GSE159709](https://doi.org/10.1101/GSE159709) in the GEO database; 10X Chromium, ‘SingleCell-E’ in [GSE159709](https://doi.org/10.1101/GSE159709) in the GEO database.

Dataset 27 (mouse embryo): 10X Visium, ‘Visium-A1’ in [GSE160137](https://doi.org/10.1101/GSE160137) in the GEO database; 10X Chromium, ‘Pax2-GFP_SC-2’ in [GSE143806](https://doi.org/10.1101/GSE143806) in the GEO database.

Dataset 28 (human prostate): 10X Visium, ‘D25’ in [GSE159697](https://doi.org/10.1101/GSE159697) in the GEO database; 10X Chromium, ‘V8’ in [GSE142489](https://doi.org/10.1101/GSE142489) in the GEO database.

Dataset 29 (mouse kidney): 10X Visium, Sham Model in [GSE171406](https://doi.org/10.1101/GSE171406) in the GEO database; 10X Chromium, wild-type sham mouse in [GSE171639](https://doi.org/10.1101/GSE171639) in the GEO database.

Dataset 30 (mouse kidney): 10X Visium, ischemia reperfusion injury model in [GSE171406](https://doi.org/10.1101/GSE171406) in the GEO database; 10X Chromium, wild-type ischemic acute kidney injury mouse in [GSE171639](https://doi.org/10.1101/GSE171639) in the GEO database.

Dataset 31 (mouse brain): 10X Visium, ‘section1’ in [GSE153424](https://doi.org/10.1101/GSE153424) in the GEO database; 10X Chromium, ‘brain1_cx’ in [GSE153424](https://doi.org/10.1101/GSE153424) in the GEO database.

Dataset 32 (mouse prefrontal cortex): 10X Visium, ‘Visium_10X’ in [GSE158450](https://doi.org/10.1101/GSE158450) in the GEO database; 10X Chromium, ‘PFC_sc_Rep1_10X’ in [GSE158450](https://doi.org/10.1101/GSE158450) in the GEO database.

Dataset 33 (mouse hippocampus): 10X Visium, ‘Visium_10X’ in [GSE158450](https://doi.org/10.1101/GSE158450) in the GEO database; 10X Chromium, ‘HIPP_sc_Rep1_10X’ in [GSE158450](https://doi.org/10.1101/GSE158450) in the GEO database.

Dataset 34 (mouse kidney): 10X Visium, [GSE154107](https://doi.org/10.1101/GSE154107) in the GEO database; 10X Chromium, sample ‘(LPS36hr) scRNA-seq’ in [GSE151658](https://doi.org/10.1101/GSE151658) in the GEO database.

Dataset 35 (human prostate): 10X Visium, ‘ETOH’ in [GSE159697](https://doi.org/10.1101/GSE159697) in the GEO database; 10X Chromium, ‘V8’ in [GSE142489](https://doi.org/10.1101/GSE142489) in the GEO database.

Dataset 36 (mouse lymph node): 10X Visium, ‘PBS’ samples of Tissue 1 in <https://github.com/romain-lopez/DestVI-reproducibility>; 10X Chromium, ‘PBS’ samples in <https://github.com/romain-lopez/DestVI-reproducibility>.

Dataset 37 (mouse MCA205 tumor): 10X Visium, Tumor A1 of Tissue 1 in <https://github.com/romain-lopez/DestVI-reproducibility>; 10X Chromium, <https://doi.org/10.1101/GSE152050>, <https://github.com/romain-lopez/DestVI-reproducibility>.

Dataset 38 (mouse primary motor cortex): 10X Visium, https://storage.googleapis.com/tommaso-brain-data/tangram_demo/Allen_Visium_Allen1_cell_count.h5ad; 10X Chromium, ‘batch 9’ in ‘mop_sn_tutorial.h5ad’ file from <https://console.cloud.google.com/storage/browser/tommaso-brain-data>.

Dataset 39 (mouse primary motor cortex): Slide-seq, https://storage.googleapis.com/tommaso-brain-data/tangram_demo/slideseq_MOOp_1217.h5ad; 10X Chromium, ‘batch 9’ in ‘mop_sn_tutorial.h5ad’ file from <https://console.cloud.google.com/storage/browser/tommaso-brain-data>.

Dataset 40 (mouse cerebellum): Slide-seqV2, SCP948 in https://singlecell.broadinstitute.org/single_cell/; 10X Chromium, sample M003 of study SCP795 in https://singlecell.broadinstitute.org/single_cell/.

Dataset 41 (mouse hippocampus): Slide-seqV2, ‘Puck_200115_08’ in https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatial-transcriptomics-at-near-cellular-resolution-with-slide-seq_v2#study-download; Drop-seq, we randomly sampled 10,000 cells from ‘GSE116470_F_GRCm38.81.P60Hippocampus.raw.dge.txt.gz’ file in [GSE116470](https://doi.org/10.1101/GSE116470) in the GEO database.

Dataset 42 (human squamous carcinoma): ST, [GSM428432](https://doi.org/10.1101/GSM428432) in the GEO database; 10X Chromium, ‘GSE144236_cSCC_counts.txt.gz’ in [GSE144236](https://doi.org/10.1101/GSE144236) in the GEO database.

Dataset 43 (mouse hippocampus): ST, wild-type replicate 1 in <https://data.mendeley.com/datasets/6s959w2zyr1>; 10X Chromium, [GSE116470](https://doi.org/10.1101/GSE116470) in the GEO database.

Dataset 44 (mouse olfactory bulb): HDST, replicate1 in [GSE130682](https://doi.org/10.1101/GSE130682) in the GEO database; 10X Chromium, WT1 samples used from [GSE121891](https://doi.org/10.1101/GSE121891) in the GEO database.

Dataset 45 (mouse liver): Seq-scope, <https://deepblue.lib.umich.edu/data/downloads/gx41mj14n>; Smart-seq2, liver sample in [GSE109774](https://doi.org/10.1101/GSE109774) in the GEO database.

We also provide an open source website for users to download all the above datasets: https://drive.google.com/drive/folders/1pHmE9cg_tMcouV1LFJFtbyBJNp7oQo9J?usp=sharing.

Source data for figures and Extended Data Figures are provided with this paper. Source data are provided with this paper.

Code availability

We uploaded the code and scripts used for the comparative analysis and figure plotting to GitHub: <https://github.com/QuKunLab/SpatialBenchmarking>. The package can also be used to analyze user’s own datasets.

References

77. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
78. Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).
79. Basic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
80. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).

Acknowledgements

This work was supported by the National Key R&D Program of China (2020YFA0112200 to K.Q.), the National Natural Science Foundation of China grants (T2125012, 91940306, 31970858, and 31771428 to K.Q.; 32170668 to B.L.; 81871479 to J.L.), CAS Project for Young Scientists in Basic Research YSBR-005 (to K.Q.), Anhui Province Science and Technology Key Program (202003a07020021 to K.Q.) and the Fundamental Research Funds for the Central Universities (YD2070002019, WK9110000141, and WK2070000158 to K.Q.; WK910000001 to J.L.). We thank the USTC supercomputing center and the School of Life Science Bioinformatics Center for providing computing resources for this project.

Author contributions

K.Q. and B.L. conceived the project. B.L., W.Z. and C.G. designed the framework and performed data analysis with help from H.X., L.L., M.F., Y.H., X.Y., X.Z., F.C. and T.X., X.Z., M.T., K.L., J.L. and L.C. contributed to revision of the manuscript. B.L., K.Q., C.G., and W.Z. wrote the manuscript with inputs from all authors. K.Q. supervised the entire project. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

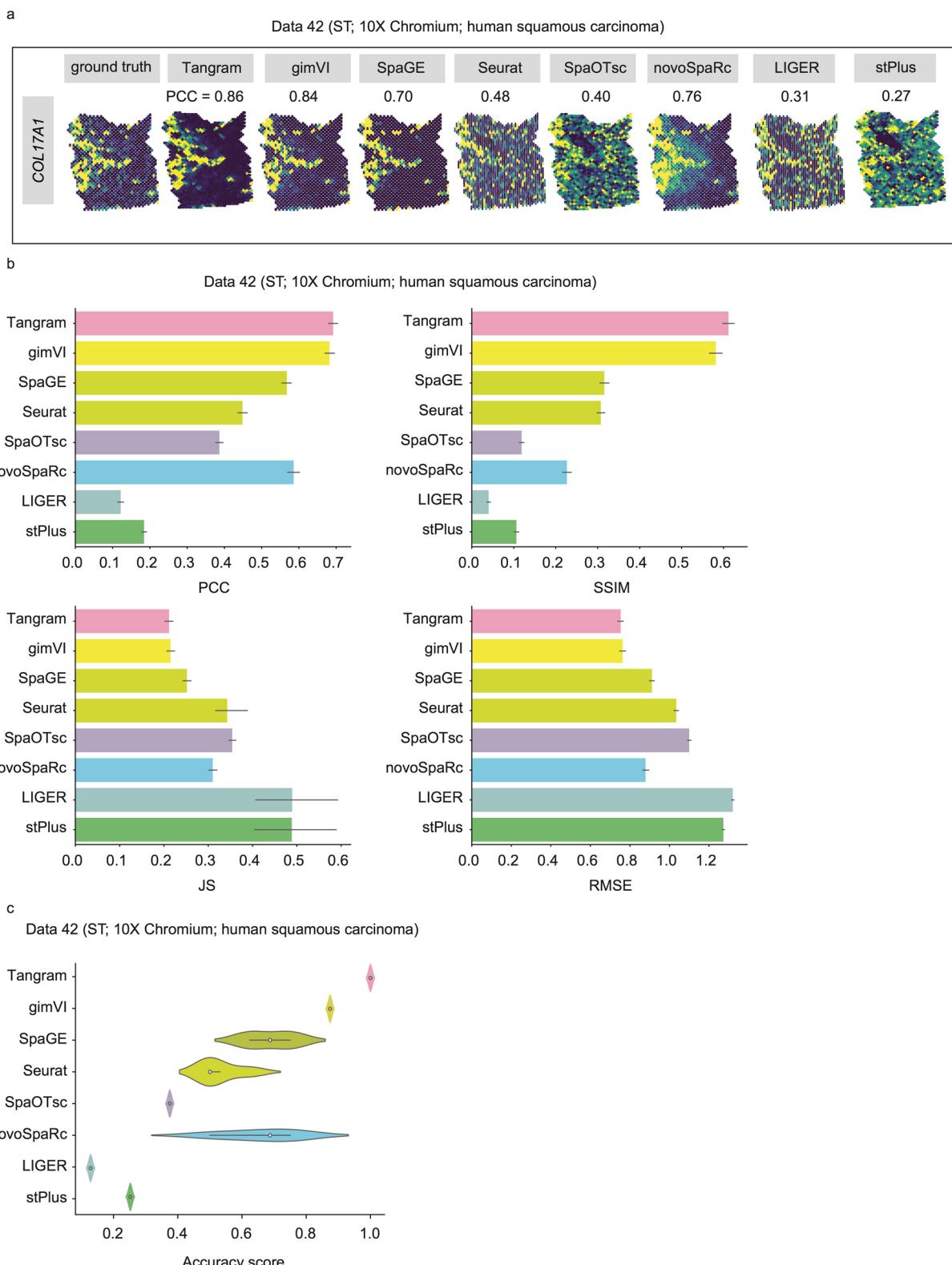
Extended data is available for this paper at <https://doi.org/10.1038/s41592-022-01480-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41592-022-01480-9>.

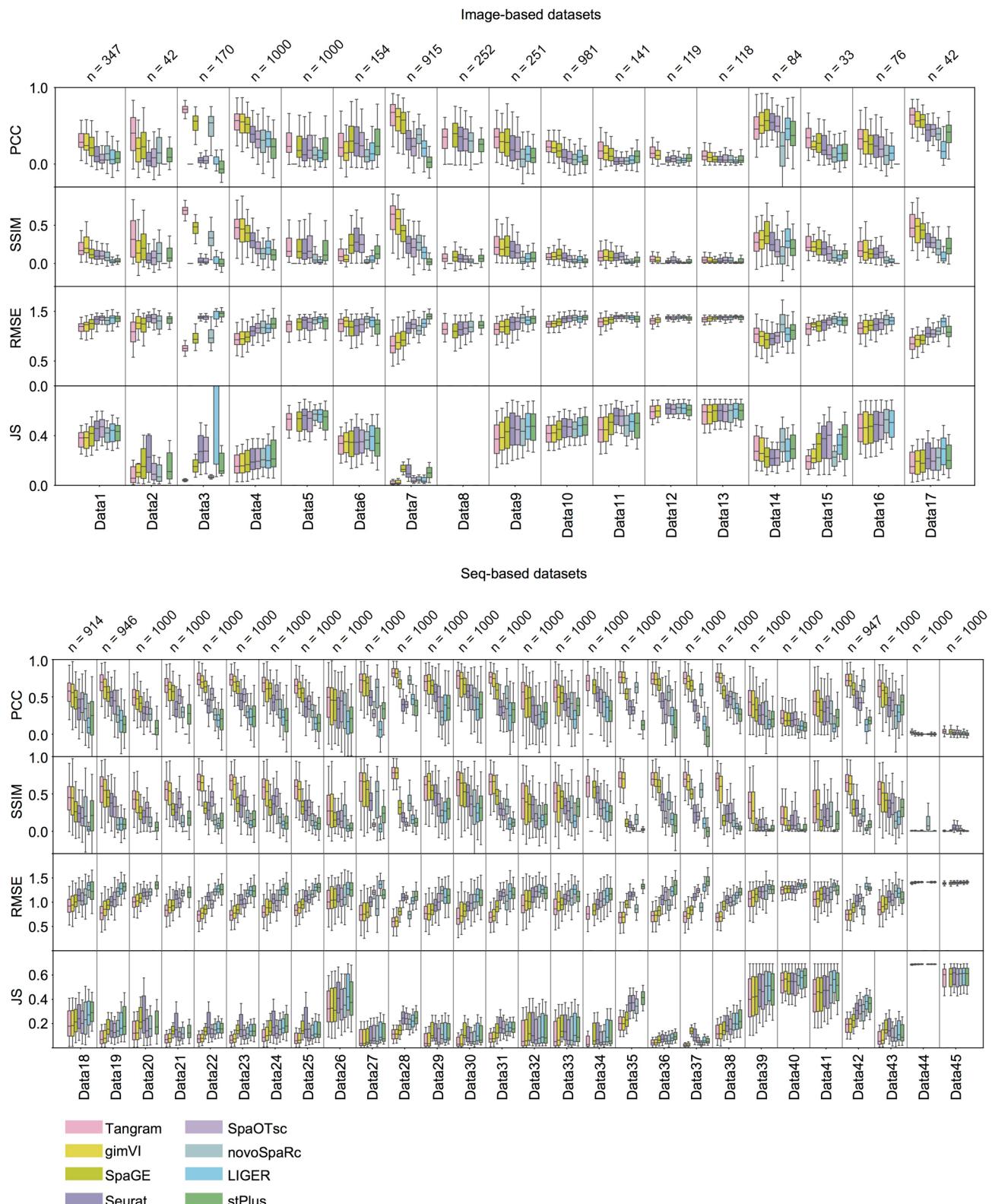
Correspondence and requests for materials should be addressed to Kun Qu.

Peer review information *Nature Methods* thanks Ahmed Mahfouz and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling editor: Lin Tang, in collaboration with the *Nature Methods* team. Peer reviewer reports are available.

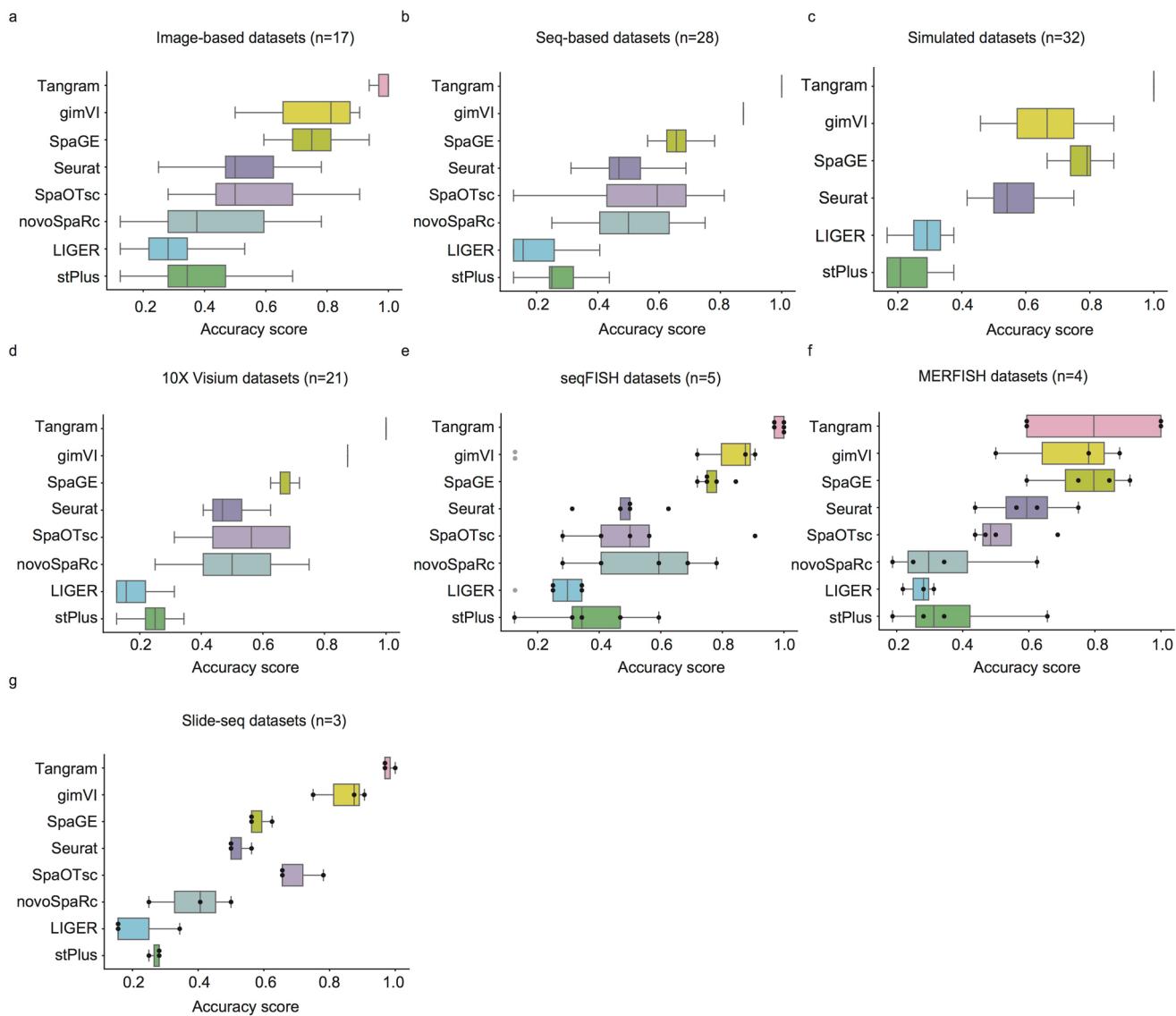
Reprints and permissions information is available at www.nature.com/reprints.



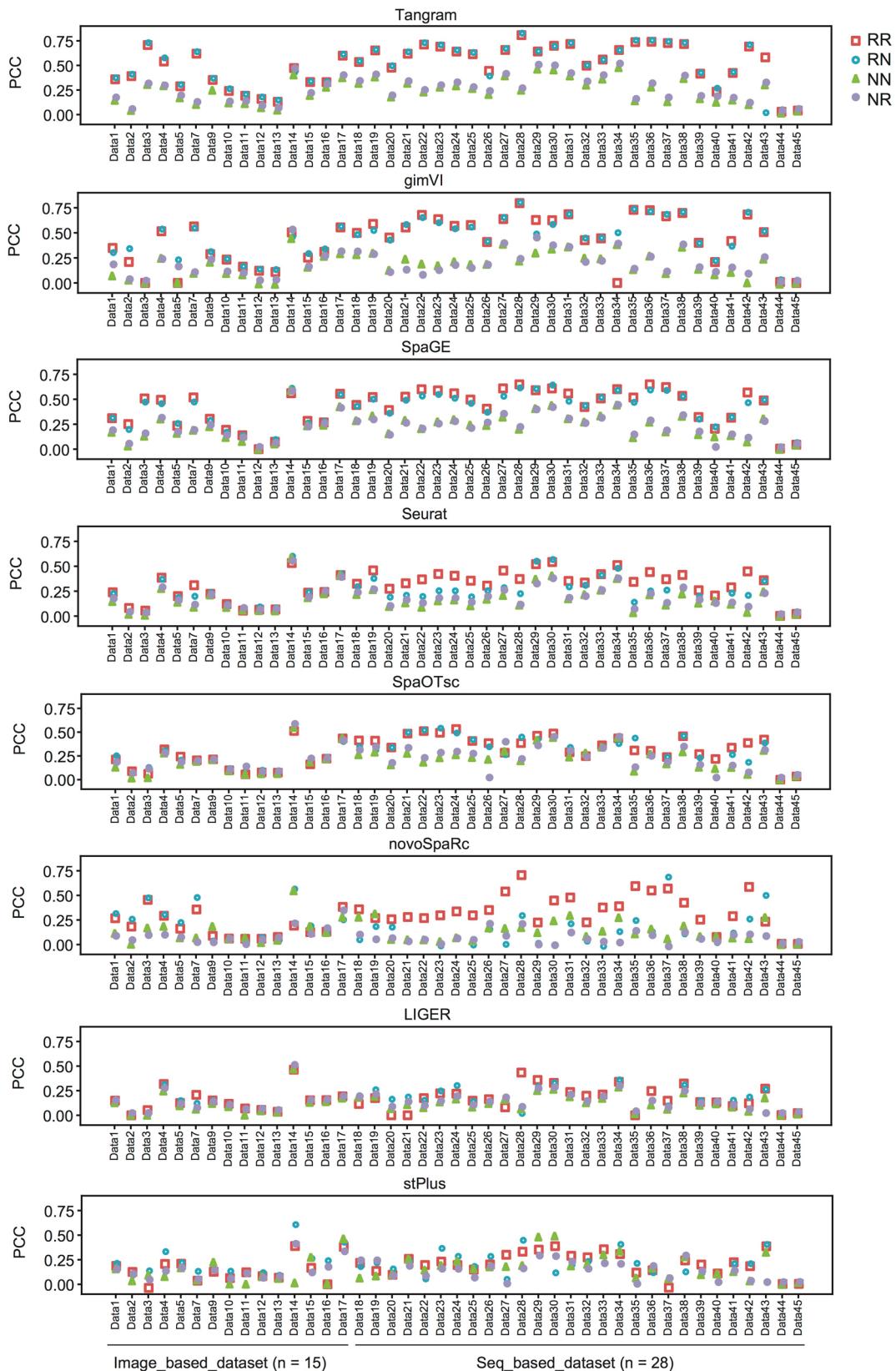
Extended Data Fig. 1 | Comparing the accuracy of eight integration methods in predicting the spatial distribution of RNA transcripts. **a**, The spatial distribution of COL17A1 in dataset 42 (ST; 10X Chromium; human squamous carcinoma), including the ground truth and prediction results from the integration methods. PCC: Pearson Correlation Coefficient between the expression vector of a transcript in the ground truth and that of the predicted result. **b**, Bar plots of PCC, SSIM, RMSE, and JS of each integration method in predicting the spatial distribution of transcripts of dataset 42. SSIM: Structural Similarity Index; RMSE: Root Mean Square Error; JS: Jensen-Shannon divergence. Data are presented as mean values \pm 95% confidence intervals; $n=948$ predicted genes. **c**, The violin plot of AS (accuracy score, aggregated from PCC, SSIM, RMSE, and JS; see Methods) of the eight integration methods for transcripts in dataset 42. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range; $n=4$ benchmark metrics.



Extended Data Fig. 2 | The boxplots of PCC, SSIM, RMSE, JS values of each integration method in predicting the spatial distribution of RNA transcripts of 45 paired spatial transcriptomics and scRNA-seq datasets. The boxplots of PCC, SSIM, RMSE, JS values of each integration method in predicting the spatial distribution of RNA transcripts of 45 paired spatial transcriptomics and scRNA-seq datasets. Center line, median; box limits, upper and lower quartiles; whiskers, 0.5x interquartile range, the number of genes for each dataset is shown at the top of each panel.

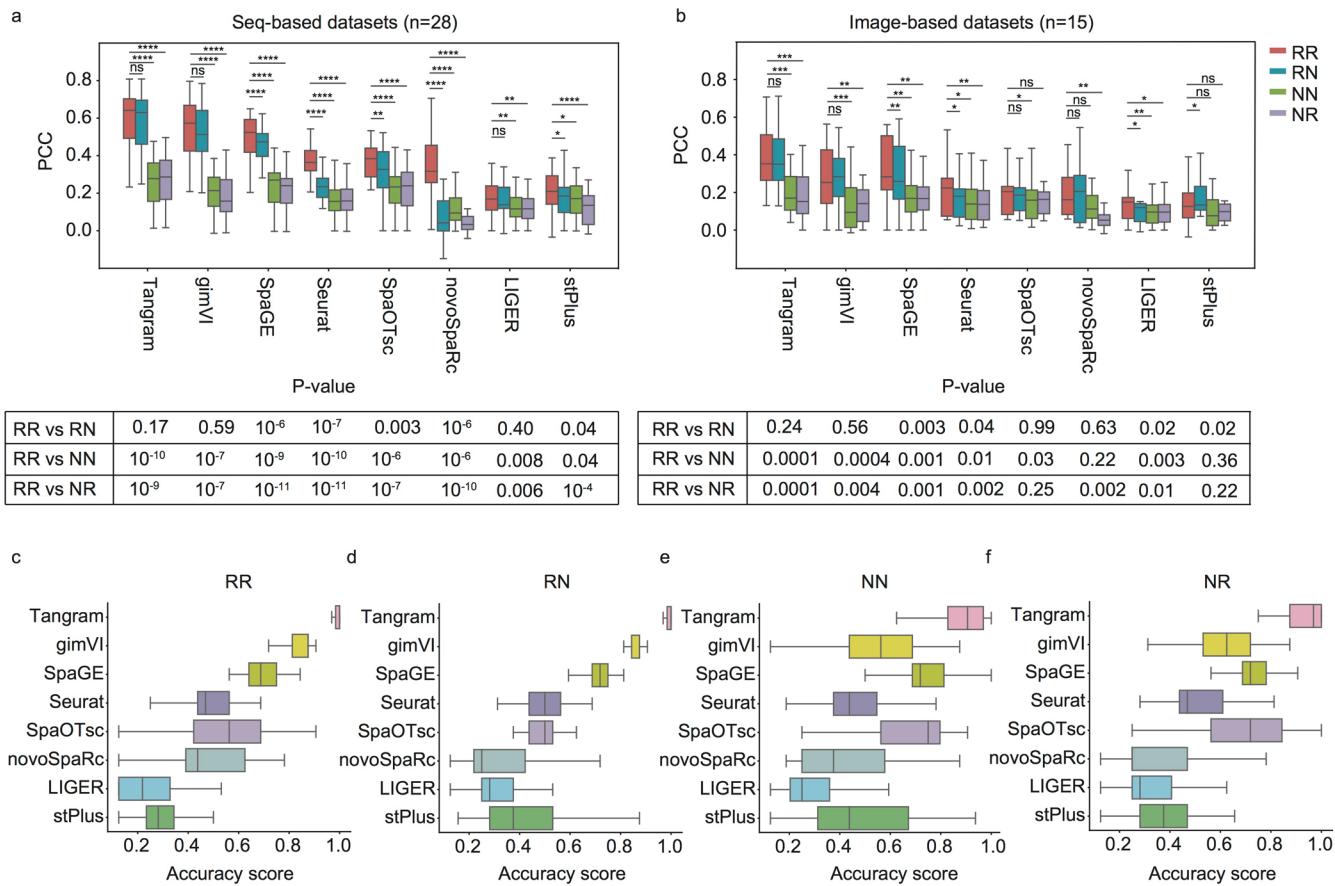


Extended Data Fig. 3 | PCC, SSIM, RMSE, JS and AS of spatial distribution of RNA transcripts predicted by each integration method for the 45 paired spatial transcriptomics and scRNA-seq datasets. a-g. Boxplots of AS (accuracy score, aggregated from PCC, SSIM, RMSE, and JS; see Methods) of the integration methods for transcripts in the 17 image-based datasets (a), 28 seq-based datasets (b), 32 simulated datasets (c), 21 10X visium datasets (d), 5 seqFISH datasets (e), 4 MERFISH datasets (f), 3 Slide-seq datasets (g). Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. Grey dots indicate the prediction result is not available, as the tool made an error when predictions.



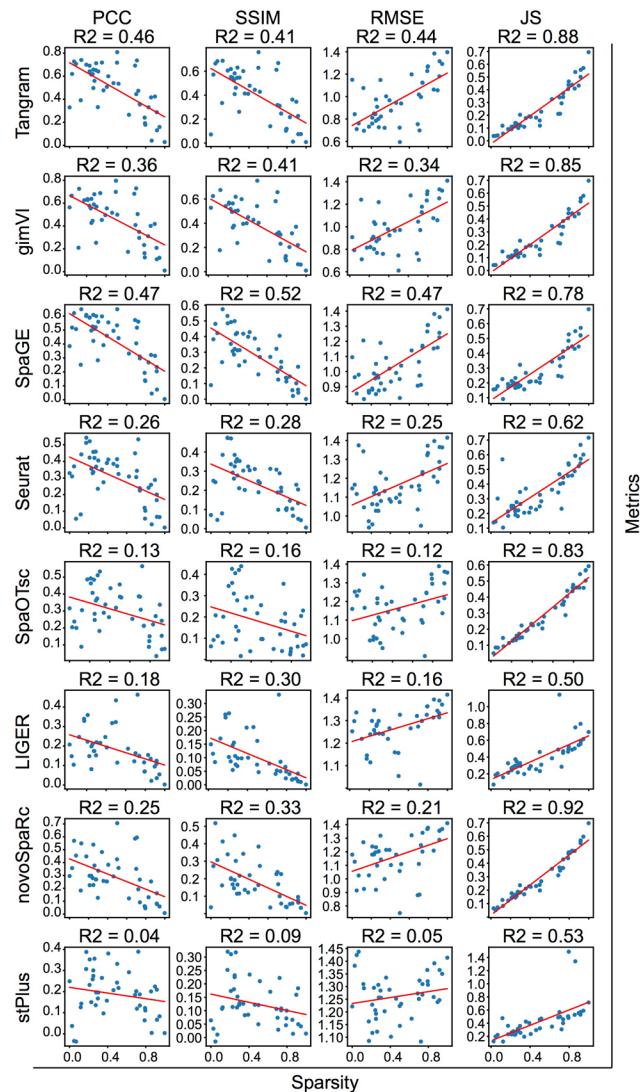
Extended Data Fig. 4 | The PCC values of each integration method when processing the raw expression matrices and the normalized expression matrices.

The PCC values of each integration method when processing the raw expression matrices and the normalized expression matrices. R-R: raw expression matrix of spatial data and raw expression matrix of scRNA-seq data; N-R: normalized expression matrix of spatial data and raw expression matrix of scRNA-seq data; R-N: raw expression matrix of spatial data and normalized expression matrix of scRNA-seq data; N-N: normalized expression matrix of spatial data and normalized expression matrix of scRNA-seq data; n=43 independent datasets. Dataset6 and Dataset8 are excluded, as the normalized expression matrix of spatial data has been normalized.



Extended Data Fig. 5 | Impact of normalization on the accuracy of eight integration methods that can predict the spatial distribution of RNA transcripts.

a, b, Boxplots of the PCC values of the eight integration methods for 28 seq-based datasets (a) or 15 image-based datasets (b) when using the four schemes of input expression matrices (that is R-R, R-N, N-R, and N-N, see their definition in the legend of Extended Data Fig. 4). For the genes predicted by each method, we removed outliers using 10%-90% confidence interval. Statistical significance was analyzed with two-sided paired t-test, *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range. **c-f**, Boxplots of the AS values of the eight integration methods for all the 45 paired datasets when using the four schemes of input expression matrices. For the genes predicted by each method, we removed outliers using 10%-90% confidence interval. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; n = 43 independent datasets.

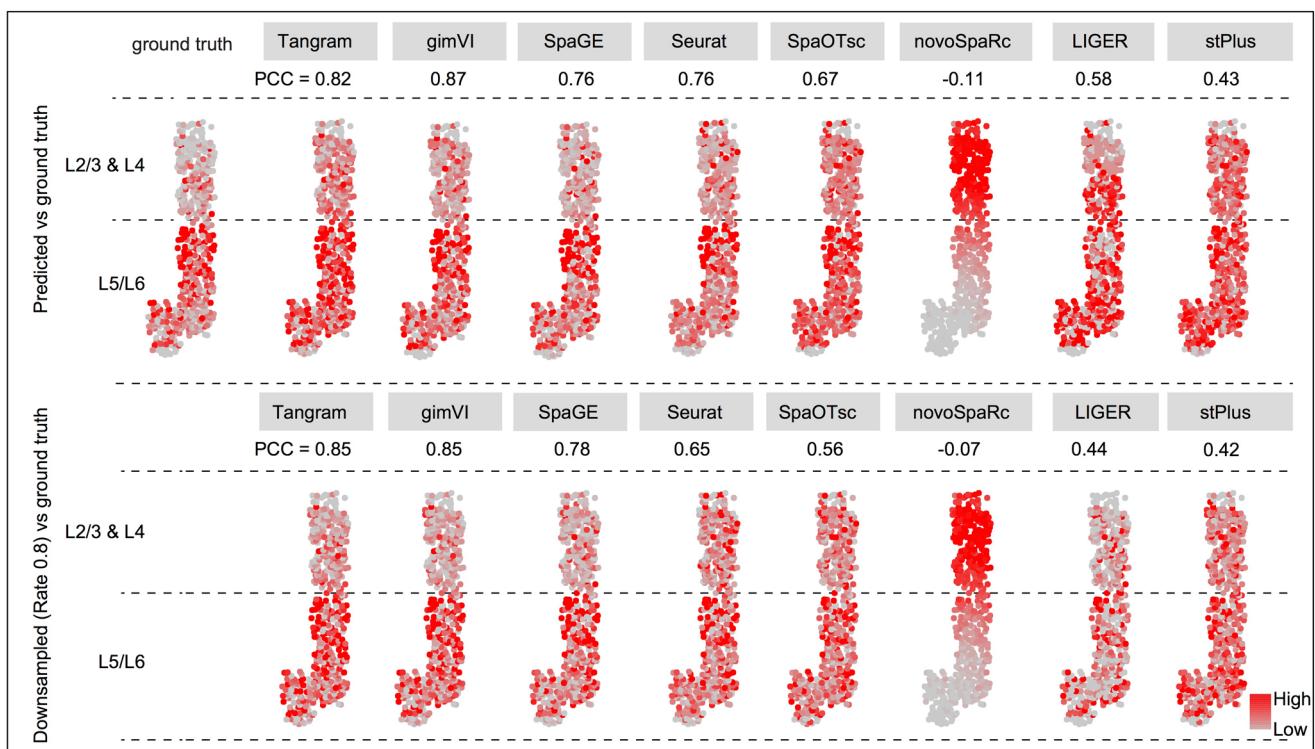


Metrics

Extended Data Fig. 6 | Correlation between the four metrics (PCC, SSIM, RMSE, and JS) and the sparsity of each examined spatial expression matrix.
 Correlation between the four metrics (PCC, SSIM, RMSE, and JS) and the sparsity of each examined spatial expression matrix. For all the eight integration methods that can predict the spatial distribution of transcripts, the JS values are linearly positively correlated with the sparsity of expression matrices of the spatial transcriptomics data ($R^2 \geq 0.50$).

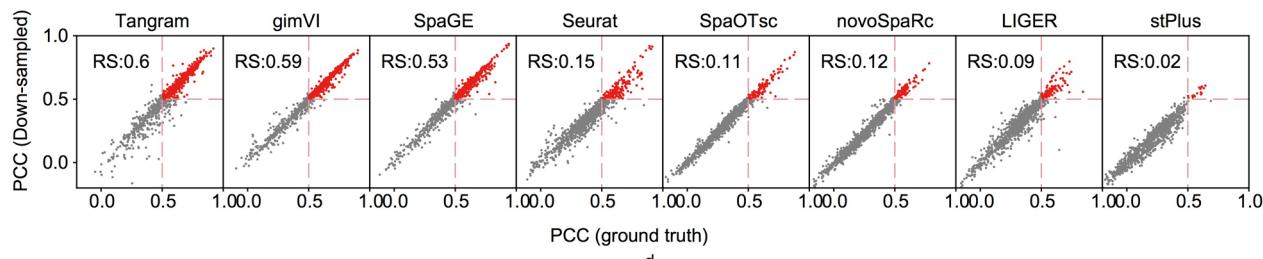
a

Dataset 4 (seqFISH+; Smart-seq; mouse cortex)
Cpxl1 expressed in L5/6 layer



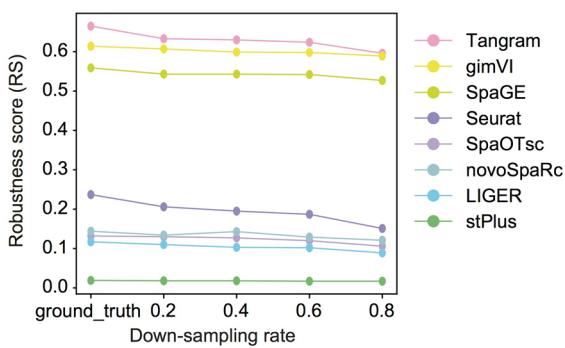
b

Dataset 4 (seqFISH+; Smart-seq; mouse cortex)

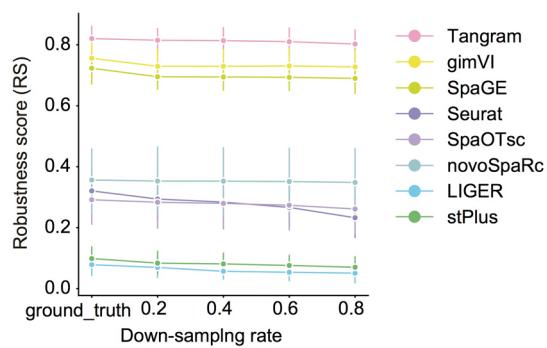


c

Dataset 4
 (seqFISH+; Smart-seq; mouse cortex)
 Down-sampled non-zero count

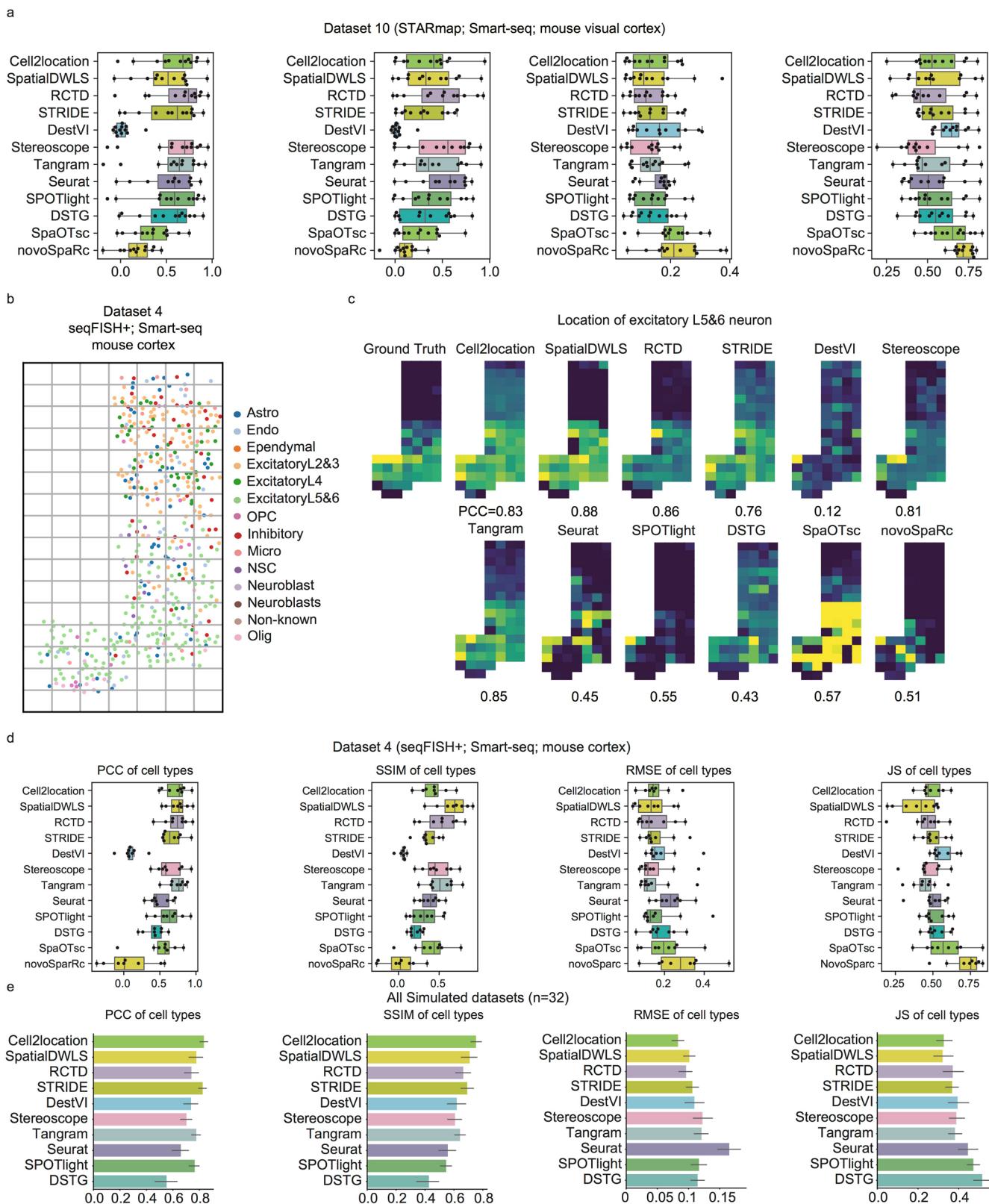


Down-sampled non-zero count
 (N Datasets = 19)



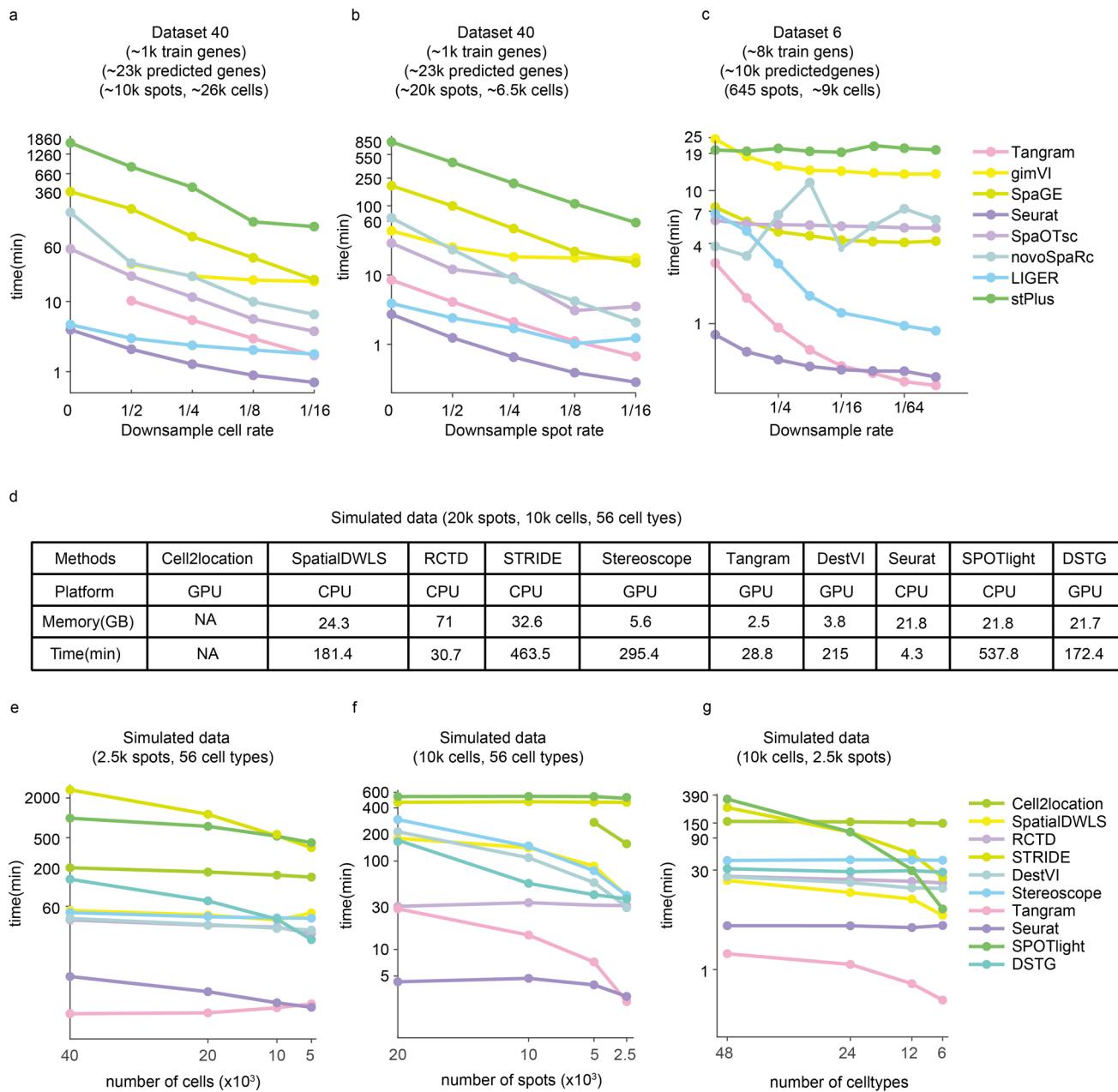
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Comparing the accuracy of the eight integration methods for sparse expression matrices down-sampled from the original datasets using Scuttle. **a**, Spatial distribution of *Cplx1* expression in dataset 4 (seqFISH+; Smart-seq; mouse cortex), predicted from the original data and down-sampled data (down-sampling rate = 0.8). **b**, PCC of the spatial distribution of transcripts predicted from the original data and down-sampled data from dataset 4. The PCC values of the red-colored transcripts are greater than 0.5 for both the original and the down-sampled data. The proportion of the red-colored transcripts in all transcripts was defined as the ‘robustness score’ (RS). **c**, RS values of the eight integration methods when processing sparse expression matrices down-sampled from dataset 4 at different down-sampling rates. **d**, RS values of the eight integration methods when processing the sparse expression matrices of the down-sampled datasets. The original datasets (used to generate the down-sampled datasets) capture >1000 genes from >100 spots, and the sparsity of the expression matrices is <0.7. Data are presented as mean values \pm 95% confidence intervals; n=19 independent datasets.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Comparing the performance of the twelve integration methods in cell type deconvolution. **a**, PCC, SSIM, RMSE, and JS values for the cell type composition of the spots simulated from dataset 10, generated by twelve integration methods. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; n=12 predicted cell types. **b**, A seqFISH+ slide of dataset 4 (seqFISH+; Smart-seq; mouse cortex) with cells annotated by cell type. Each grid represents a simulated spot containing 1–18 cells. **c**, The proportion of L5&6 excitatory neurons in the spots simulated from dataset 4, including the ground truth and the predicted results of twelve integration methods. **d**, PCC, SSIM, RMSE, and JS values for the cell type composition of the spots simulated from dataset 4, generated by twelve integration methods. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; n=8 predicted cell types. **e**, PCC, SSIM, RMSE, and JS values for the cell type composition of the spots in all the simulated datasets (n=32), generated by ten integration methods. SpaOTsc and novoSpaRc are excluded, as they require spatial location information for each spot, which is not available in the simulated datasets. Data are presented as mean values \pm 95% confidence intervals; n=32 independent datasets.



Extended Data Fig. 9 | Computer resources consumed by each integration method. **a-c.** The impact of the number of cells in scRNA-seq data (a), the number of spots in spatial data (b), and the number of genes used for training (c), on computational resources consumed by the integration methods that can predict the spatial distribution of undetected transcripts. **d.** The computer time and memory spent by the integration methods that can deconvolute cell types of histological spots, when processing a simulated dataset which contains 20000 spots in its spatial transcriptomics data and 10000 cells in its scRNA-seq data. **e-g.** The impacts of the number of cells in scRNA-seq data (e), the number of spots in spatial data (f), and the number of the cell types (g) on computational resources consumed by the integration methods that can deconvolute cell types of histological spots.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

| | |
|-----------------|--|
| Data collection | No software was used for data collection. |
| Data analysis | <p>We compared the performance of 8 integration methods for predicting the spatial distribution of undetected transcripts: gimVI (Version 0.8.0b0), SpaGE(no version), Tangram(Version 1.0.0), Seurat (Version 3.6.3), SpaOTsc (Version 0.2), LIGER (Version 0.5.0), novoSpaRc (Version 0.4.3), stPlus(Version 0.0.6).</p> <p>We compared the performance of 12 integration methods for predicting the cell type composition of spots: Cell2location(Version 0.6a0), DestVI((Version 0.14.4), SPOTlight((Version 0.1.7), SpatialDWLS(Version 1.0.4), Seurat(Version 4.0.5), Tangram((Version 1.0.0), RCTD(Version 1.2.0), Stereoscope(Version 0.14.4), STRIDE(Version 0.0.1b0), DSTG (Version 0.0.1), novoSpaRc (Version 0.4.3), SpaOTsc (Version 0.2).</p> <p>We applied the Scuttle(Version 1.0.4) and Splatte(Version 1.19.3) package for the down-sampling of the spatial expression matrices of the datasets.</p> <p>The code used in this paper is available at https://github.com/QuKunLab/SpatialBenchmarking. The file provided at https://github.com/QuKunLab/SpatialBenchmarking/Benchmarkingenvironment.yml lists the software dependencies with version numbers.</p> |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Dataset 1 : seqFISH, <https://content.cruk.cam.ac.uk/jmlab/SpatialMouseAtlas2020/>; 10X Chromium, ‘Sample 21’ in MouseGastrulationData within the R/Bioconductor data package MouseGastrulationData.

Dataset 2 : seqFISH, <https://zenodo.org/record/3735329#.YY69HZMza3J>; Microwell-Seq, ‘EmbryonicStemCells’ in ‘MCA_BatchRemoved_Merge_dge.h5ad’ file in https://figshare.com/articles/dataset/MCA_DGE_Data/5435866.

Dataset 3 : seqFISH, <https://ars.els-cdn.com/content/image/1-s2.0-S0896627316307024-mmc6.xlsx>; 10X Chromium, “HIPP_sc_Rep1_10X sample” in GSE158450 in the GEO database.

Dataset 4 : seqFISH+, <https://github.com/CaiGroup/seqFISH-PLUS>, and the spatial coordinate of each spot was generated using “stitchFieldCoordinates” function in Giotto; Smart-seq, mouse primary visual cortex (V1Sp) in the dataset in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 5 : seqFISH+, <https://github.com/CaiGroup/seqFISH-PLUS>; Drop-seq, GSE148360 in the GEO database.

Dataset 6 : MERFISH, the 18th female parent mouse (Animal ID = 18) in <https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248>; 10X Chromium, GSE113576 in the GEO database.

Dataset 7 : MERFISH, the ‘B1_cell’ used in https://www.pnas.org/doi/suppl/10.1073/pnas.1912459116/suppl_file/pnas.1912459116.sd12.csv; 10X Chromium, BC22 in GSE152048 in the GEO database.

Dataset 8 : MERFISH, ‘mouse1_slice162’ in <https://caltech.box.com/shared/static/dzqt6ryytmjbgyai356s1z0phntsba0.gz>; 10X Chromium, https://data.nemoarchive.org/biccn/lab/zeng/transcriptome/scell/10x_v3/mouse/processed/analysis/10X_cells_v3_AIBS.

Dataset 9 : MERFISH, <https://github.com/spacex-spacejam/data/>; Smart-seq, mouse primary visual cortex (V1Sp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 10 : STARmap, “20180505_BY3_1kgenes” in <https://www.staremapresources.com/data>; Smart-seq, mouse primary visual cortex (V1Sp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 11 : STARmap, “20180419_BZ9_control” in <https://www.staremapresources.com/data>; 10X Chromium, “PFC_sc_Rep2_10X” in GSE158450 in the GEO database.

Dataset 12 : ISS, <https://github.com/spacex-spacejam/data>; Smart-seq, <https://portal.brain-map.org/atlas-and-data/rnaseq/human-mtg-smart-seq>.

Dataset 13 : ISS, <https://github.com/spacex-spacejam/data>; Smart-seq, mouse primary visual cortex (V1Sp) in the dataset in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 14 : FISH, <https://github.com/rajewsky-lab/distmap>; Drop-seq, GSE95025 in the Gene Expression Omnibus (GEO) database.

Dataset 15 : osmFISH, cortical regions in <http://linnarssonlab.org/osmFISH/>; Smart-seq, mouse somatosensory cortex (SSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-smart-seq>.

Dataset 16 : BARISTASeq, <https://github.com/spacex-spacejam/data>; Smart-seq, mouse primary visual cortex (V1Sp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 17 : ExSeq, <https://github.com/spacex-spacejam/data>; Smart-seq, mouse primary visual cortex (V1Sp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 18 : 10X Visium, Vis5A in GSE161318 in the GEO database; 10X Chromium, D2_Ev3 in GSE159500 in the GEO database.

Dataset 19 : 10X Visium, Vis9A in GSE161318 in the GEO database; 10X Chromium, D7_Ev3 in GSE159500 in the GEO database.

Dataset 20 : 10X Visium, “CID3586” in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in GSE176078 in the GEO database.

Dataset 21 : 10X Visium, “1160920F” in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in GSE176078 in the GEO database.

Dataset 22 : 10X Visium, “CID4290” in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in GSE176078 in the GEO database.

Dataset 23 : 10X Visium, “CID4465” in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in GSE176078 in the GEO database.

Dataset 24 : 10X Visium, “CID44971” https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in GSE176078 in the GEO database.

Dataset 25 : 10X Visium, “CID4535” in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, ‘CID3586’ in GSE176078 in the GEO database.

Dataset 26 : 10X Visium, “Visium-A” in GSE159709 in the GEO database; 10X Chromium, “SingleCell-E” in GSE159709 in the GEO database.

Dataset 27 : 10X Visium, “Visium-A1” in GSE160137 in the GEO database; 10X Chromium, “Pax2-GFP_SC-2” in GSE143806 in the GEO database.

Dataset 28 : 10X Visium, “D25” in GSE159697 in the GEO database; 10X Chromium, “V8” in GSE142489 in the GEO database.

Dataset 29 : 10X Visium, Sham Model in GSE171406 in the GEO database; 10X Chromium, wild-type sham mouse in GSE171639 in the GEO database.

Dataset 30 : 10X Visium, Ischemia Reperfusion Injury Model in GSE171406 in the GEO database; 10X Chromium, wild-type ischemic acute kidney injury mouse in GSE171639 in the GEO database.

Dataset 31 : 10X Visium, “section1” in GSE153424 in the GEO database; 10X Chromium, “brain1_cx” in GSE153424 in the GEO database.

Dataset 32 : 10X Visium, “Visium_10X” in GSE158450 in the GEO database; 10X Chromium, “PFC_sc_Rep1_10X” in GSE158450 in the GEO database.

Dataset 33 : 10X Visium, “Visium_10X” in GSE158450 in the GEO database; 10X Chromium, “HIPP_sc_Rep1_10X” in GSE158450 in the GEO database.

Dataset 34 : 10X Visium, GSE154107 in the GEO database; 10X Chromium, sample “(LPS36hr) scRNAseq” in GSE151658 in the GEO database.

Dataset 35 : 10X Visium, “ETOH” in GSE159697 in the GEO database; 10X Chromium, “V8” in GSE142489 in the GEO database.

Dataset 36 : 10X Visium, ‘PBS’ samples of Tissue 1 in <https://github.com/romain-lopez/DestVI-reproducibility>; 10X Chromium, ‘PBS’ samples in <https://github.com/romain-lopez/DestVI-reproducibility>.

Dataset 37 : 10X Visium, Tumor A1 of Tissue 1 in <https://github.com/romain-lopez/DestVI-reproducibility>; 10X Chromium, <https://github.com/romain-lopez/DestVI-reproducibility>.

Dataset 38 : 10X Visium, <https://console.cloud.google.com/storage/browser/tommaso-brain-data>; 10X Chromium, ‘batch 9’ in ‘tangram_demo_mop_sn_tutorial.h5ad’ file from <https://console.cloud.google.com/storage/browser/tommaso-brain-data>.

Dataset 39 : Slide-seq, <https://console.cloud.google.com/storage/browser/tommaso-brain-data>; 10X Chromium, ‘batch 9’ in ‘tangram_demo_mop_sn_tutorial.h5ad’ file from <https://console.cloud.google.com/storage/browser/tommaso-brain-data>.

Dataset 40 : Slide-seqV2, SCP948 in https://singlecell.broadinstitute.org/single_cell/; 10X Chromium, sample M003 of study SCP795 in https://singlecell.broadinstitute.org/single_cell/.

Dataset 41 : Slide-seqV2, “Puck_200115_08” in https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatial-transcriptomics-at-near-cellular-resolution-with-slide-seqv2#study-download; Drop-seq, we randomly sampled 10,000 cells from ‘GSE116470_F_GRCm38.81.P60Hippocampus.raw.dge.txt.gz’ file in GSE116470 in the GEO database.

Dataset 42 : ST, GSM4284322 in the GEO database; 10X Chromium, “GSE144236_cSCC” in GSE144236 in the GEO database.

Dataset 43 : ST, wild-type replicate 1 in <https://data.mendeley.com/datasets/6s959w2zyr/1>; 10X Chromium, GSE116470 in the GEO database.

Dataset 44 : HDST, replicate1 in GSE130682 in the GEO database; 10X Chromium, WT1 samples used from GSE121891 in the GEO database.

Dataset 45 : Seq-scope, <https://deepblue.lib.umich.edu/data/downloads/gx41mj14n>; Smart-seq2, liver sample in GSE109774 in the GEO database.

The raw and processed data of these datasets are also available at : https://drive.google.com/drive/folders/1pHmE9cg_tMcouV1LFJFtbyBJNp7oQo9J?usp=sharing. A summary of these datasets is given in Supplementary Table 1.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

We used 45 paired spatial transcriptomics and scRNA-seq datasets from published studies. The spatial transcriptomic datasets were produced by 13 different spatial transcriptomics approaches. A summary of the individual accession numbers is given in Supplementary Table 1. The details of these datasets are listed as follows:
 Dataset 1 has 8425 spots in spatial transcriptomics data and 4651 cells in scRNA-seq data;
 Dataset 2 has 175 spots in spatial transcriptomics data and 9991 cells in scRNA-seq data;
 Dataset 3 has 3585 spots in spatial transcriptomics data and 8596 cells in scRNA-seq data;
 Dataset 4 has 524 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 5 has 2050 spots in spatial transcriptomics data and 31217 cells in scRNA-seq data;
 Dataset 6 has 4975 spots in spatial transcriptomics data and 31299 cells in scRNA-seq data;
 Dataset 7 has 645 spots in spatial transcriptomics data and 9234 cells in scRNA-seq data;
 Dataset 8 has 6963 spots in spatial transcriptomics data and 7240 cells in scRNA-seq data;
 Dataset 9 has 2399 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 10 has 1549 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 11 has 1380 spots in spatial transcriptomics data and 7737 cells in scRNA-seq data;
 Dataset 12 has 6000 spots in spatial transcriptomics data and 15928 cells in scRNA-seq data;
 Dataset 13 has 6000 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 14 has 3039 spots in spatial transcriptomics data and 1297 cells in scRNA-seq data;
 Dataset 15 has 3405 spots in spatial transcriptomics data and 5613 cells in scRNA-seq data;
 Dataset 16 has 11426 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 17 has 1154 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 18 has 982 spots in spatial transcriptomics data and 4748 cells in scRNA-seq data;
 Dataset 19 has 995 spots in spatial transcriptomics data and 4816 cells in scRNA-seq data;
 Dataset 20 has 4784 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 21 has 4895 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 22 has 2432 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 23 has 1211 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 24 has 1162 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 25 has 1127 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 26 has 2425 spots in spatial transcriptomics data and 1911 cells in scRNA-seq data;
 Dataset 27 has 198 spots in spatial transcriptomics data and 3415 cells in scRNA-seq data;
 Dataset 28 has 277 spots in spatial transcriptomics data and 4740 cells in scRNA-seq data;
 Dataset 29 has 1835 spots in spatial transcriptomics data and 10872 cells in scRNA-seq data;
 Dataset 30 has 2064 spots in spatial transcriptomics data and 13600 cells in scRNA-seq data;
 Dataset 31 has 3805 spots in spatial transcriptomics data and 8798 cells in scRNA-seq data;
 Dataset 32 has 3024 spots in spatial transcriptomics data and 3512 cells in scRNA-seq data;
 Dataset 33 has 3024 spots in spatial transcriptomics data and 8653 cells in scRNA-seq data;
 Dataset 34 has 1888 spots in spatial transcriptomics data and 8346 cells in scRNA-seq data;
 Dataset 35 has 744 spots in spatial transcriptomics data and 4561 cells in scRNA-seq data;
 Dataset 36 has 369 spots in spatial transcriptomics data and 7268 cells in scRNA-seq data;
 Dataset 37 has 2125 spots in spatial transcriptomics data and 7185 cells in scRNA-seq data;
 Dataset 38 has 2669 spots in spatial transcriptomics data and 3499 cells in scRNA-seq data;
 Dataset 39 has 9852 spots in spatial transcriptomics data and 3499 cells in scRNA-seq data;
 Dataset 40 has 41674 spots in spatial transcriptomics data and 26252 cells in scRNA-seq data;
 Dataset 41 has 6000 spots in spatial transcriptomics data and 10000 cells in scRNA-seq data;
 Dataset 42 has 1145 spots in spatial transcriptomics data and 48164 cells in scRNA-seq data;
 Dataset 43 has 604 spots in spatial transcriptomics data and 15095 cells in scRNA-seq data;
 Dataset 44 has 6000 spots in spatial transcriptomics data and 10259 cells in scRNA-seq data;
 Dataset 45 has 2177 spots in spatial transcriptomics data and 981 cells in scRNA-seq data;

Data exclusions

No data was excluded from the study. In the data preprocessing step of scRNA-seq data, we filtered cells that captured >200 RNA transcripts.

Replication

To make sure that the experimental findings are reproducible, we (1) verified the performance of the 8 integration methods for predicting the spatial distribution of undetected transcripts on 45 datasets and 32 simulated datasets, (2) verified the performance of the 8 integration

methods under a varied set of down-sampling settings to assess the robustness of each method. (3) verified the performance of the 10 integration methods for predicting the cell type composition of spots on 32 simulated datasets

Randomization

This is not relevant to our study because we reanalyzed publicly available data or generated our own synthetic data.

Blinding

In the comparison, all the methods were blinded to the ground truth of the spatial transcriptomics data. The outputs from the methods were then compared to the ground truth available in the respective datasets.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| | |
|-------------------------------------|-------------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | Antibodies |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | Animals and other organisms |
| <input checked="" type="checkbox"/> | Human research participants |
| <input checked="" type="checkbox"/> | Clinical data |
| <input checked="" type="checkbox"/> | Dual use research of concern |

Methods

| | |
|-------------------------------------|------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | ChIP-seq |
| <input checked="" type="checkbox"/> | Flow cytometry |
| <input checked="" type="checkbox"/> | MRI-based neuroimaging |