




A comprehensive comparison on cell-type composition inference for spatial transcriptomics data

Jiawen Chen [†], Weifang Liu [†], Tianyou Luo[†], Zhentao Yu, Minzhi Jiang, Jia Wen, Gaorav P. Gupta, Paola Giusti, Hongtu Zhu, Yuchen Yang and Yun Li 

Corresponding author. Yun Li, Department of Genetics, 120 Mason Farm Road, Campus Box 7264, University North Carolina, Chapel Hill, NC 27599, USA. Tel: (919) 843-2832; Fax: (919) 843-4682; E-mail: yunli@med.unc.edu

[†]Authors Jiawen Chen, Weifang Liu and Tianyou Luo contributed equally to this work.

Abstract

Spatial transcriptomics (ST) technologies allow researchers to examine transcriptional profiles along with maintained positional information. Such spatially resolved transcriptional characterization of intact tissue samples provides an integrated view of gene expression in its natural spatial and functional context. However, high-throughput sequencing-based ST technologies cannot yet reach single cell resolution. Thus, similar to bulk RNA-seq data, gene expression data at ST spot-level reflect transcriptional profiles of multiple cells and entail the inference of cell-type composition within each ST spot for valid and powerful subsequent analyses. Realizing the critical importance of cell-type decomposition, multiple groups have developed ST deconvolution methods. The aim of this work is to review state-of-the-art methods for ST deconvolution, comparing their strengths and weaknesses. In particular, we construct ST spots from single-cell level ST data to assess the performance of 10 methods, with either ideal reference or non-ideal reference. Furthermore, we examine the performance of these methods on spot- and bead-level ST data by comparing estimated cell-type proportions to carefully matched single-cell ST data. In comparing the performance on various tissues and technological platforms, we concluded that RCTD and stereoscope achieve more robust and accurate inferences.

Keywords: spatial transcriptomics, single-cell, cell-type deconvolution, deep learning, probabilistic modeling

Introduction

Interrogation of patterns of messenger RNAs (mRNAs) with their spatial context maintained in intact tissue sections enables simultaneous profiling of tissue anatomy and function. Recently burgeoning spatial transcriptomics (ST) technologies empower such interrogation and thus open new ways of biomedical research, holding the promise to reveal novel biological insights that can have direct clinical relevance in terms of diagnosis and treatments [1, 2]. These rapidly advancing ST technologies provide us with quantification of mRNA expression for a large number of genes, while maintaining their spatial context in the original tissue sample [3–6]. ST technologies can be classified largely into two categories, namely imaging- and sequencing-based methods [7]. Imaging-based techniques, including single-molecule fluorescence *in situ* hybridization (smFISH) [8], multiplexed error robust fluorescence *in situ* hybridization (MERFISH) [3] and non-barcoded and unamplified cyclic-

ouroboros smFISH method (osmFISH) [9], provide both quantitative measurements of RNA expression levels and information about RNA spatial localization by directly imaging individual RNA molecules in single cells. Sequencing-based techniques, such as spatial barcoding employed by the commercially available 10× Genomics Visium platform and Spatial Transcriptomics platform, are powered by placing histological sections on barcoded reverse transcription primers with unique positional barcodes followed by sequencing and computational reconstruction to capture gene expression in tissue samples [6]. We note that in our manuscript, spatial transcriptomics or ST refers to the general spatial transcriptomics field and Spatial Transcriptomics refers to the particular technology platform originally proposed by Ståhl *et al.* [6].

The emergence and rapid advancements of ST technologies offer an unprecedented way to explore transcriptional profiles in the spatially resolved context,

Jiawen Chen is a PhD student in the Department of Biostatistics at the University of North Carolina at Chapel Hill.

Weifang Liu is a PhD student in the Department of Biostatistics at the University of North Carolina at Chapel Hill.

Tianyou Luo is a PhD student in the Department of Biostatistics at the University of North Carolina at Chapel Hill.

Zhentao Yu is a PhD student in the Department of Biostatistics at the University of North Carolina at Chapel Hill.

Min-Zhi Jiang is a PhD candidate in the Department of Applied Physical Sciences at the University of North Carolina at Chapel Hill.

Jia Wen is a postdoctoral researcher in the Department of Genetics at the University of North Carolina at Chapel Hill.

Gaorav P. Gupta is an assistant professor in the Department of Radiation Oncology at the University of North Carolina at Chapel Hill.

Paola Giusti is an assistant professor in the Department of Psychiatry at the University of Florida.

Hongtu Zhu is a professor in the Department of Biostatistics at the University of North Carolina at Chapel Hill.

Yuchen Yang is an associate professor in the School of Ecology at Sun Yat-sen University.

Yun Li is a professor in the Departments of Genetics, Biostatistics and Computer Science at the University of North Carolina at Chapel Hill.

Received: February 21, 2022. Revised: May 20, 2022. Accepted: May 25, 2022

© The Author(s) 2022. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

enabling studies of myriads of aspects that were not possible without ST technologies, including identification of genes whose expressions exhibit spatial patterns, revelation of cellular subpopulations in their native spatial context, exploration of biologically relevant spatial domains, and interrogation of cell-cell communications. These spatially resolved gene expression patterns inform the corresponding tissue and organ system, capturing deeply intertwined organ structure and function. In contrast, single-cell RNA sequencing (scRNA-seq) technologies often require isolation of single cells by either fluorescence-activated cell sorting (FACS) or manual picking, to study organ structure and function [10, 11].

However, existing ST techniques are limited by the trade-off between spatial resolution and the number of genes measured. Imaging-based ST techniques can attain single cell or even subcellular resolution, but albeit theoretically possible to quantify up to 10 000 genes, most can only measure hundreds of genes with high quality and fidelity, entailing *a priori* marker gene selection and rendering them less suitable for exploratory analyses [3, 12]. Sequencing-based techniques can measure whole-transcriptome-wide gene expression, whereas these technologies can only obtain spot-level data (where each spot is of diameter 2–10 μm or 50–100 μm) that approach, but do not yet achieve, single-cell resolution [6, 13]. Therefore, downstream analyses are susceptible to confounding caused by differential cell-type compositions across spots. For example, for the identification of spatially variable genes, gene expression variation across spots could be driven by different mixtures of cell types and/or varying numbers of cells as opposed to being truly driven by spatial location. We therefore need to estimate the cell-type composition of each spot, for powerful and valid downstream analysis. Multiple ST deconvolution methods have been recently developed for this purpose of inferring spot-level cell-type mixtures. These methods each have their unique features, making it challenging for investigators to choose methods that best suit their data. We therefore need an impartial and comprehensive assessment of various state-of-the-art ST deconvolution methods. In the literature [14–23], the performance of each method has been assessed predominantly by the developers, using simulated datasets with varied assumptions presented in different studies. These comparisons are often incomplete and prone to biases in interpretation. To the best of our knowledge, no third-party comprehensive evaluation of the ST deconvolution methods has been performed using diverse real datasets. In this work, we aim to fill in this gap.

In this review, we summarize and compare computational strategies proposed for cell-type deconvolution of ST data. The review is organized as follows. We first describe 10 state-of-the-art ST deconvolution methods, highlighting several key aspects including the statistical method employed, type(s) of ST data tailored to, and method-specific unique features. We then present performance of these methods using six real ST datasets as

benchmarks. Finally, we provide practical guidelines and emphasize the advantages and drawbacks of these methods in real data applications. We note here that we only consider methods that focus squarely on cell mixture deconvolution which output cell-type proportion as a result. Methods like MIA [24] and Seurat [25] that provide other spatially deconvolved matrices like cell enrichment score for a certain area or anchor score are not included in the evaluation.

Computational methods developed for cell-type deconvolution of ST data

In recent years, a diverse collection of ST deconvolution methods has been proposed. Existing deconvolution methods for ST data can be largely classified into three groups: probabilistic methods, methods based on non-negative matrix factorization (NMF) and non-negative least squares (NNLS), and other methods (Figure 1, Table 1). The first group, probabilistic methods, includes Adroit [14], cell2location [15], DestVI [16], RCTD [17], STdeconvolve [18] and stereoscope [19], where the data distribution is explicitly or parametrically specified and inference is carried out using likelihood-based approaches. The second group, NMF and NNLS based methods, includes spatialDWLS [20] and SPOTlight [21]. Other methods including DSTG [22] and Tangram [23], estimate the cell-type proportion using some specifically designed method architecture or loss function, which we loosely classify as other methods (Figure 1, Table 1). Here we review 10 state-of-the-art methods: Adroit, cell2location, DestVI, RCTD, STdeconvolve, stereoscope, spatialDWLS, SPOTlight, DSTG and Tangram [14–23]. We briefly summarize each method below.

Accurate and Robust Method to Infer Transcriptome Composition (AdRoit) [14] is designed for bulk RNA-seq data, but it can also be used for ST data. AdRoit first selects informative genes, models their expression distributions by assuming gene counts following negative binomial distributions, and then estimates their corresponding locations and dispersion parameters. Subsequently, cross-sample variability, collinearity of expression profiles and cell-type specificity are estimated from spot-level ST data. Then, gene-wise scaling factors are estimated by jointly modeling reference scRNA-seq data and ST data under inference. These gene-wise scaling factors in AdRoit enable the method to correct for potential platform biases between the scRNA-seq reference and target ST data. Finally, these quantities are included in a weighted regularized model for inferring cell-type proportions.

Cell2location [15] adopts a Bayesian hierarchical framework. It first uses external scRNA-seq data as reference to estimate cell type-specific signatures. The observed spatial expression count matrix is then modeled with a negative binomial distribution with the mean parameter depending on reference cell-type signatures, and the overdispersion parameter modeled using an exponential-Gamma compound prior which

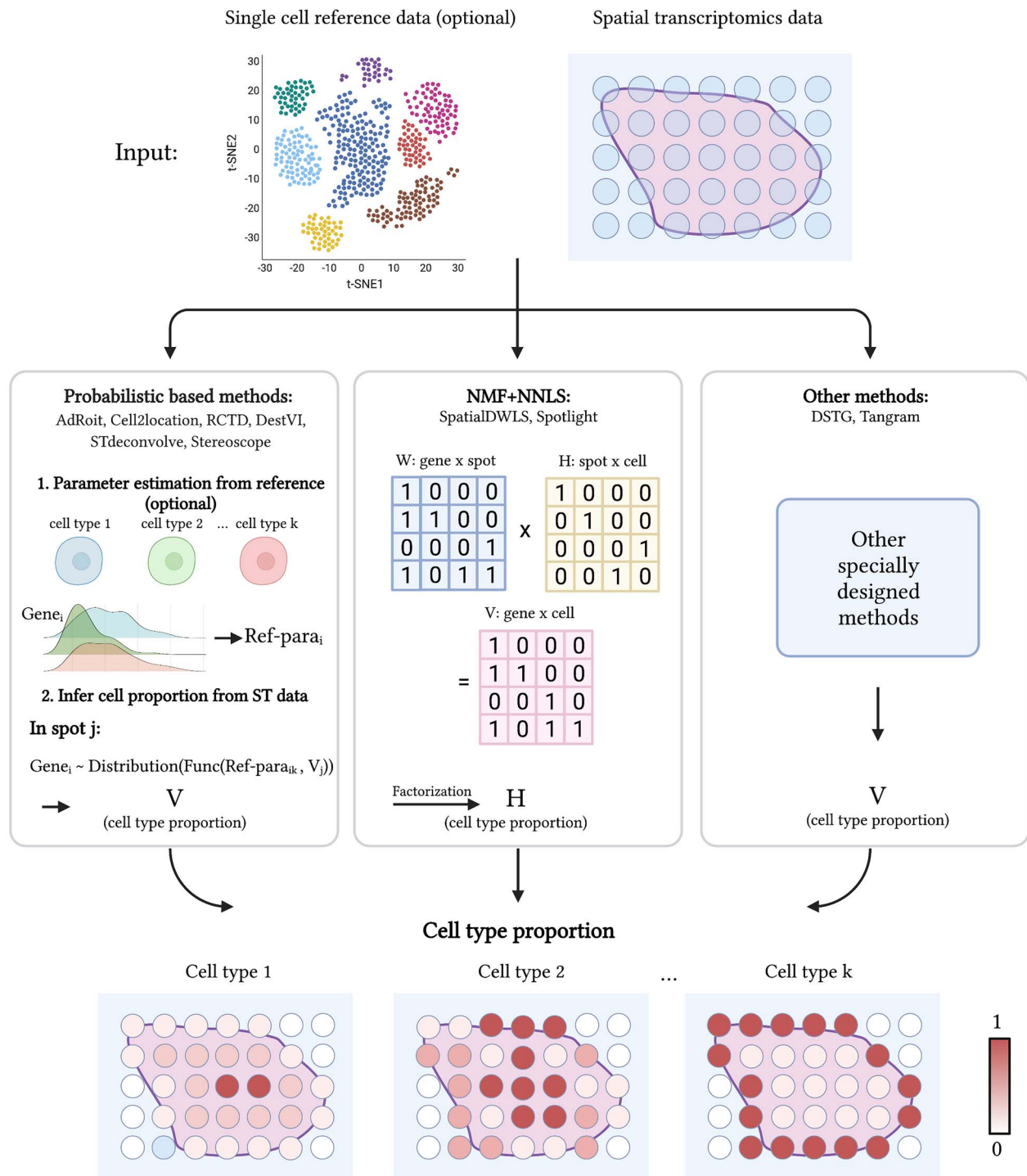


Figure 1. Summary of ST deconvolution methods. ST deconvolution methods take (target) ST data and an optional scRNA-seq reference data as input (top panel). Current ST deconvolution methods can be classified into three main categories: probabilistic-based, NMF and NNLS based, and other methods (center panel). The output of ST deconvolution methods is cell-type proportion for each spot which scales from 0 to 1 (bottom panel). Ref-para: estimated parameters of the distribution employed from the reference data. Func: the function of parameters and cell-type proportion designed by each method.

aims to make most genes have low overdispersion. Gene-specific technological sensitivity and gene- and location-specific additive shifts are included as part of the mean parameter, each individually modeled using a separate hierarchical Gamma prior. Cell2location further models the regression weights of cell-type signatures using a hierarchical Gamma prior and decomposes the regression weights into contributions from multiple

latent groups which can be interpreted as spots with shared cell-type abundance profiles, aiming to borrow strength across locations with similar cell compositions. Finally, cell2location employs variational Bayesian inference to approximate the posterior distribution and produces parameter estimates accordingly.

Deconvolution of spatial transcriptomics profiles using variational inference (DestVI) [16] is a probabilis-

Table 1. ST deconvolution methods overview

Method	Designed for ST data?	Feature selection	Inference method	Language	URLs	Reference	Published Time (bioRxiv first version)
stereoscope	Yes	Top 5000 highest expressed genes (optional)	Probabilistic, negative binomial distribution	Python	https://github.com/almaan/stereoscope	[19]	10.09.2020 (12.13.2019)
RCTD	Yes	DE genes	Probabilistic, Poisson distribution, maximum likelihood	R	https://github.com/dmcable/spacexr	[17]	02.18.2021 (05.08.2020)
SPOTlight	Yes	Highly variable genes	Non-negative matrix factorization (NMF) along with non-negative least squares (NNLS)	R	https://github.com/MarcElosua/SPOTlight_deconvolution_analysis	[21]	02.05.2021 (06.04.2020)
Tangram	Yes	Union of cell type marker genes	Optimization of self-constructed loss function	Python	https://github.com/broadinstitute/Tangram	[23]	10.28.2021 (08.30.2020)
DSTG	Yes	2000 most variable genes	Semi-supervised graph convolutional network, adaptive moment estimation algorithm	Python	https://github.com/Su-informatics-lab/DSTG	[22]	01.22.2021 (10.21.2020)
cell2location	Yes	No selection	Probabilistic, negative binomial distribution, variational Bayesian inference	Python	https://cell2location.readthedocs.io/en/latest/	[15]	01.13.2022 (11.17.2020)
AdRoit	No	Genes enriched in one or more cell types or highly variable genes	Probabilistic, non-negative least squares regression	R	https://github.com/TaoYang-dev/AdRoit	[14]	10.22.2021 (01.04.2021)
spatialDWLS	Yes	Cell type marker genes	Dampened weighted least squares (DWLS)	R	https://giottosuite.com/	[20]	05.10.2021 (02.03.2021)
DestVI	Yes	Highly variable genes	Probabilistic, latent variable models, auto-encoding variational bayes	Python	https://docs.scvi-tools.org/en/stable/user_guide/models/destvi.html	[16]	04.21.2022 (05.11.2021)
STdeconvolve	Yes	Highly variable genes	Generative probabilistic model: latent Dirichlet allocation (LDA), variational expectation-maximization algorithm	R	https://jef.works/STdeconvolve/	[18]	04.29.2022 (06.16.2021)

tic method for multi-resolution analysis of ST data. DestVI explicitly models variation within cell types via continuous latent variables instead of limiting the analysis to a discrete view of cell types. Such continuous within-cell-type variations as well as the corresponding cell type-specific profiles are learned through a conditional deep generative model, specifically, using variational inference with decoder neural networks. In this scheme, two different latent variable models (LVMs) are constructed for reference scRNA-seq (scLVM) and target ST data (stLVM), respectively. DestVI similarly assumes that the number of observed transcripts follows a negative binomial distribution. The decoder neural network trained by scLVM is employed by stLVM, and cell-type proportion is obtained using maximum-a-

posteriori (MAP) inference scheme where the number of observed transcripts in each spot is assumed to follow a weighted sum of the inferred single-cell negative binomial distributions.

Robust cell-type decomposition (RCTD) [17] is initially designed for Slide-seq data, but it could also be used on other ST data. It assumes that the observed spot-level gene counts follow a Poisson-log-normal mixture. The mean of the log-normal distribution for the library-size-normalized Poisson rate parameter is modeled with cell type-specific mean expression profiles, while accounting for platform effects by including a gene-specific random effect term. RCTD first uses external scRNA-seq reference data to estimate the mean gene expression profile of each cell type. Gene filtering is

then performed by selecting differentially expressed (DE) genes across cell types, and the variance of gene-specific platform effects is estimated. The inferred platform effects are plugged into the probabilistic model to obtain the maximum likelihood estimates (MLE) of cell-type proportions.

STdeconvolve [18] is a reference-free and unsupervised cell-type deconvolution method for ST data. The key difference between STdeconvolve and other methods is that STdeconvolve can perform cell-type deconvolution without using external scRNA-seq references. The method is built upon latent Dirichlet allocation (LDA), which has been applied in deconvolution for bulk RNA-seq data [26], to identify putative transcriptional profiles for each cell type and their proportions in each ST spot. STdeconvolve adopts the standard LDA framework [27] in the context of ST data where each spot is defined as a mixture of a predetermined number of cell types modeled by a multinomial distribution while cell-type distribution is drawn from a uniform Dirichlet distribution. STdeconvolve assumes the existence of highly co-expressed genes for each cell type and selects significantly over-dispersed genes to inform latent clusters. It also selects informative genes and provides data-driven measures to select the number of distinct clusters if not pre-specified. Although not necessary, annotation of the inferred clusters could be performed with an external scRNA-seq reference using transcriptional correlation analysis or gene set enrichment analysis.

Stereoscope [19] performs deconvolution by spatially mapping cell types using annotated scRNA-seq reference and target ST data. Stereoscope also relies on the commonly adopted assumption that gene counts from both spatial and single-cell data follow a negative binomial distribution. The method incorporates a gene-specific coefficient that is shared across all ST spots in order to correct for potential platform biases between scRNA-seq reference and target ST data. Additionally, stereoscope includes a noise term as a 'dummy' cell type to account for data asymmetry when cell types in the reference do not match perfectly with those in the target ST data. Finally, stereoscope employs MLE to estimate cell type-specific parameters from scRNA-seq reference data and uses MAP to infer cell-type mixture in the ST data.

SpatialDWLS [20] is an enrichment-based, weighted least squares method that uses dampened weighted least squares (DWLS) [28] to deconvolve ST data, where weights minimizing the overall relative error rate are selected. First, Parametric Analysis of gene set Enrichment analysis (PAGE) [29, 30] identifies likely cell types present in each ST spot by calculating the fold change of cell type-specific marker genes for each spot. Then, DWLS is applied to infer the proportions of cell types in each spot based on enrichment results. Rare cell types are removed after initial proportion estimation, followed by a second round of deconvolution, which produces the final estimates.

SPOTlight [21] is a deconvolution algorithm that employs the non-negative matrix factorization (NMF)

regression algorithm as well as the non-negative least squares (NNLS). In SPOTlight, NMF is carried out to identify cell type-specific topic profiles in scRNA-seq references and NNLS is carried out to identify spot topic profiles, which generates the deconvolution result. Besides cell-type deconvolution estimates, SPOTlight also quantifies the quality of the predicted composition by calculating the total sum of squares and the residual sum of squares.

DSTG [22] is a similarity-based semi-supervised graph convolutional network (GCN) model that can recover cell-type proportions in each ST spot. By leveraging scRNA-seq data, DSTG first constructs synthetic ST data called 'pseudo-ST' by randomly pooling two to eight cells each time from scRNA-seq data to form pseudo-ST spots. Then, to capture the similarity between spots (incorporating both pseudo and real ST data), DSTG learns a link graph by finding mutual nearest neighbors in the shared space identified by canonical correlation analysis (CCA). Based on the link graph, a semi-supervised GCN is then trained with both pseudo and real ST data, which can be used to predict cell-type proportions in the real ST data.

Tangram [23] utilizes a machine-learning-based framework and adopts specifically designed loss functions to learn a mapping that aligns scRNA-seq reference data to spatial spots, and can thereby carry out cell-type deconvolution. Tangram randomly arranges single cells to spatial locations and computes an objective function measuring the spatial correlation between real ST data and single-cell aggregated 'pseudo-spatial' data, both at gene level and at spot level. Then Tangram tries to maximize this objective function by rearranging the single cells in space to match real ST data and potentially selecting the optimal subset of single cell observations. The final output is a matrix denoting the probability of finding each cell in each ST spot. Tangram can also optionally utilize the histological or fluorescence image to carry out cell segmentation and use the estimated number of cells per spot as additional regularization in the model.

In summary, many innovative ST deconvolution tools have been developed and tailored specifically for ST data. These methods, largely in the developers' hands, have demonstrated their potential in both simulated and real datasets. However, there does not exist an impartial and comprehensive comparison of these methods. Here, we use multiple real ST datasets, encompassing both single-cell level and spot-level ST data, each with pathologist annotations (Table 2), to systematically and objectively evaluate the performance of these methods.

Benchmarking ST deconvolution methods performance

We employed three tissues to evaluate the performance of the aforementioned 10 methods [14–23] (Table 2). We used a combination of single-cell resolution and spot-level ST datasets to compare the methods in various

Table 2. Data source and reference

Data	Type	Tissue	Reference	Link
seqFISH+ 10x	Single-cell resolution ST scRNA-seq	Mouse olfactory bulb Mouse olfactory bulb	[5] [32]	https://github.com/CaiGroup/SeqFISH-PLUS https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121891
ISS 10x	Single-cell resolution ST scRNA-seq	Human heart Human heart	[31] [31]	https://github.com/Moldia/heart European Genome-phenome Archive accession number: EGAS00001003996
Spatial Transcriptomics SMART-seq	Spot-level ST scRNA-seq	Human heart Mouse brain	[31] [11]	https://www.spatialresearch.org https://portal.brain-map.org/atlas-and-data/maseq/mouse-whole-cortex-and-hippocampus-smart-seq (Here we used the data released in October 2019)
10x	Spot-level ST	Mouse brain	[15]	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-11114/
Slide-seqV2	Bead-level ST	Mouse brain	[13]	https://singlecell.broadinstitute.org/single_cell/study/SCP815/sensitive-spatial-genome-wide-expression-profiling-at-cellular-resolution#study-summary
osmFISH	Single-cell resolution ST	Mouse brain	[9]	http://linnarssonlab.org/osmFISH/

real-world scenarios. For single-cell resolution [5, 9, 31] ST datasets, we pooled the cells together according to their spatial coordinates to construct pseudo-ST spots that mimic real ST spots (Methods) [18]. In this way, we have the truth of the cell-type mixture in each pseudo spot. We assessed deconvolution performance with either internal reference (i.e. using the single-cell resolution ST dataset itself as the scRNA-seq reference) or external reference (other scRNA-seq datasets from the same tissue). For performance quantification, we used three metrics: root mean square error (RMSE), distance correlation across cell types, as well as the difference from truth for each cell type. Smaller RMSE, higher distance correlation and smaller difference from truth all indicate better performance. For spot-level ST datasets, we examined the inferred cell-type proportion according to some carefully-matched reference single-cell resolution ST dataset. We note that evaluating spot-level inference results is challenging because we do not have true spot-level cell-type compositions. Lacking gold standard truth results in uncertainty in performance quantification because the working truth we used, albeit carefully matched to the best of our knowledge, may still differ from the target spot-level ST data under inference. We reason that for tissues with well-established layered structure, such as brain cortical regions, we can at least acceptably evaluate inferred compositions of major cell types, by treating a carefully matched single-cell level ST dataset as the working truth.

Evaluation on mouse olfactory bulb (MOB)

We first evaluated the deconvolution methods on ST data from mouse olfactory bulb (MOB) [3]. We utilized single-cell level data obtained from the seqFISH+ platform [5]. This seqFISH+ dataset provides measurements of 10 000 genes, which is among the largest number of genes available in single-cell level ST data. In this dataset, 7 fields of views (FOV) of the olfactory bulb are available, containing

a total number of 2050 single cells. We cropped each FOV into 25 spots (Figure 2A) and retained only spots with non-zero cells for further analysis. For deconvolution methods that require a scRNA-seq reference, we first performed deconvolution using the internal reference (i.e. the seqFISH+ data itself as the reference). Such a perfectly matching reference eliminates the potential performance impairment caused by any systematic differences between the reference and the target ST data. Although idealistic and not realistic, this internal reference evaluation serves as a baseline assessment that provides an upper limit of methods' performance. We then proceeded with more realistic evaluations where the deconvolution methods were tested against an external reference (Methods), which allowed us to evaluate the methods' ability to handle potential batch effects between reference and ST data under inference.

Since the selection of genes is critical to deconvolution performance, we considered several gene subsets. This seqFISH+ dataset, containing 10 000 genes, allowed us to evaluate the impact of choices of genes on deconvolution performance. We considered the following three types of gene subsets. First, a 'default' gene subset was subsetted using the built-in gene selection strategy of each deconvolution method. Note that the 'default' gene subsets, using method-specific default strategies, are therefore specific to each method (Methods). For methods without a specific recommendation or built-in gene selection strategy, we used 2000 highly variable genes (HVGs) as the default. Second, we evaluated different choices of HVGs. Specifically, we considered the top 100, 500 and 1000 HVG, where HVGs were defined using the single-cell reference (Methods), restricted to the genes also available in the target ST data. Finally, we constructed gene subsets containing top cell-type marker genes. Specifically, we pooled top marker genes for each cell type (Methods) to make the number of unique genes summing approximately to 100, 500 and 1000 genes in total (Figure 2B and C).

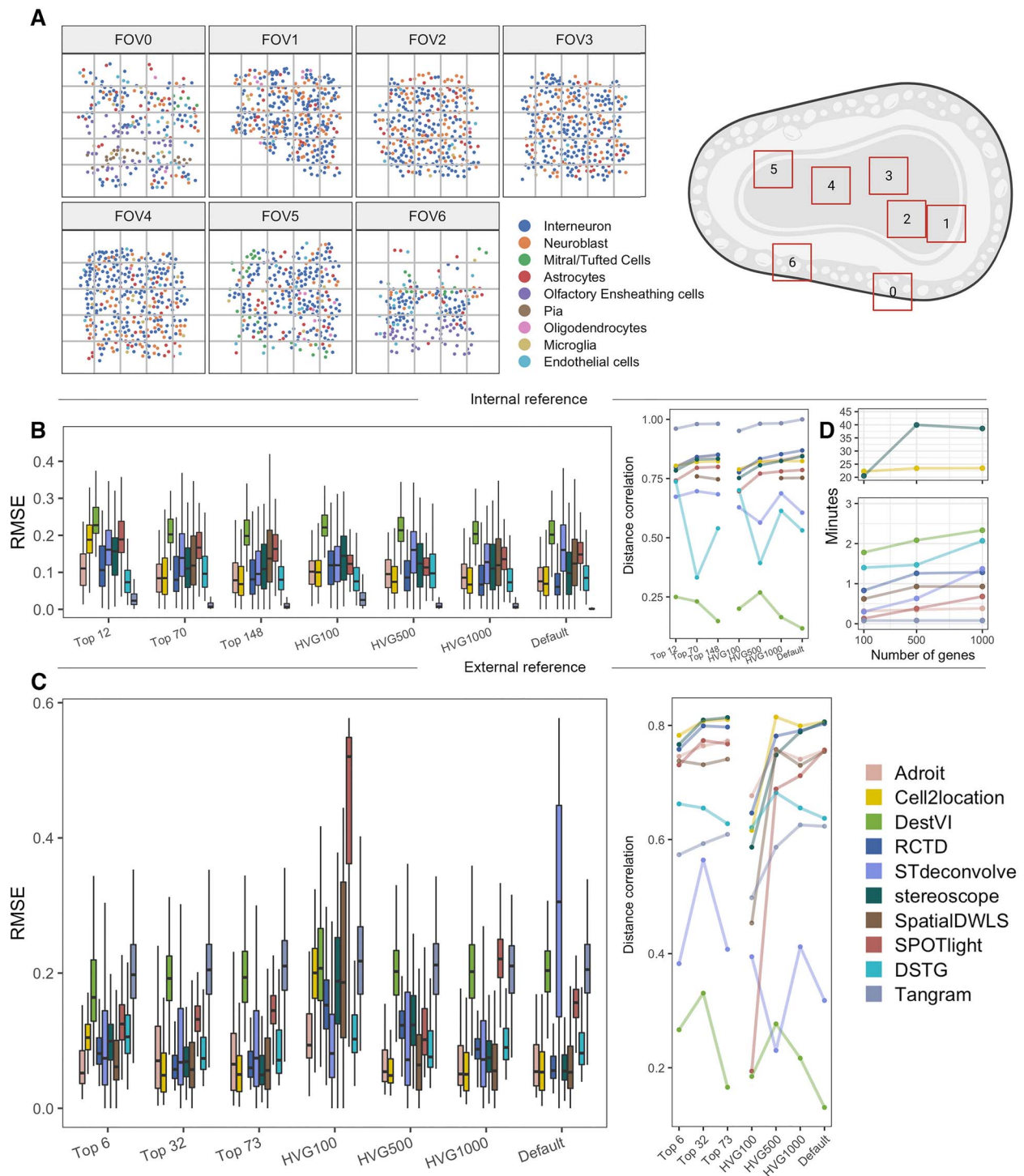


Figure 2. Evaluation on mouse olfactory bulb (MOB) data. **(A)** Overview of the cell atlas of seven fields of view in the MOB dataset. **(B)** RMSE and distance correlation of cell-type proportion estimates using internal reference from 10 methods using sets of marker genes (Top 12, Top 70, Top 148), HVGs and method-specific default gene subset (Default). **(C)** RMSE and distance correlation of cell-type proportion estimates using external reference from 10 methods using sets of marker genes (Top 6, Top 32, Top 73), HVGs and method-specific default gene subset (Default). **(D)** Computing time (in minutes) when using 100, 500, 1000 HVGs including both single-cell inference (if present) and ST spot deconvolution. Methods missing in each panel indicate that they did not produce results using the corresponding reference and ST data.

Using the internal reference, Adroit, cell2location, RCTD, DSTG and Tangram show low RMSE in their inferred results (Figure 2B). Both HVG and top marker gene subsets exhibit consistent reductions in RMSE as the number of genes increases from 100 to 1000. DSTG shows consistent results for different numbers

of nearest neighbors tested ($k = 20, 50$ and 100) when using its default gene subset (Supplementary Figure 1B). The patterns observed above with the RMSE metric remain qualitatively similar when assessed with the distance correlation metric (Figure 2B). In particular, Tangram, Adroit, cell2location, RCTD and stereoscope

inferred results display high correlation, regardless of the gene subset tested. Closer examination of the results reveals that the major deviations from truth stem from the estimated proportions of interneuron and neuroblast (Supplementary Figure 1A). These two cell types share a very close embedding space in tSNE (Supplementary Figure 2). These similarities may explain why some methods struggle to estimate their proportions accurately. SPOTlight attains relatively high correlation; however, the cell-type proportion estimates are much noisier and tend to deviate further from the true absolute proportions than the other top performers highlighted above, which also explains why SPOTlight results in better correlation but worse RMSE than spatialDWLS (Supplementary Figure 3).

Proceeding to more realistic evaluations, we performed deconvolution using an external scRNA-seq reference, also from mouse olfactory bulbs (Methods) [32]. We observed some drastic changes in the relative performances of the methods (Figure 2C). Since cell types in the external scRNA-seq reference are different from those in the ST data, we combined neuronal cell types and kept only the overlapping cell types when evaluating the performance. Therefore, the RMSE values from the internal reference are not directly comparable to those from the external reference. Among the best performers when using the internal reference, four remain among the top: Adroit, cell2location, RCTD and stereoscope, when using the default gene subset. DSTG achieves its best performance when $k = 20$ (Supplementary Figure 1D). Tangram, the No. 1 best performer when using the internal reference, however, shows substantially impaired performance when using this external reference. It appears that Tangram produces clearly biased estimates for Olfactory Ensheathing cells (OEC) and neurons, noticeably deviating from the truth (Supplementary Figures 1C and 4). DestVI exhibits similar issues with either internal or external reference, where all spots have very similar cell composition estimates (Supplementary Figures 3 and 4). Taken together patterns observed from internal and external reference, our results suggest that RCTD, cell2location and stereoscope are among the most robust to batch effects between reference scRNA-seq and target ST data.

Regarding choice of gene subsets, both RMSE and distance correlation are more drastically influenced by the number of genes when using external reference than when using the internal reference. In addition, when using external reference, most methods perform better with top cell-type marker genes than with HVG gene subsets, and most achieve the best performance with the default gene subset. Accordingly, we used the default sets of genes for analyses presented in the rest of the manuscript unless otherwise specified. For STdeconvolve, the appropriate number of clusters (representing cell types) can be determined based on prior knowledge of the ST dataset or determined by

a data-driven metric by fitting models with different numbers of clusters (detailed in Methods). To evaluate STdeconvolve, we allowed STdeconvolve to choose its optimal cluster numbers and we only kept deconvolved clusters that were successfully mapped onto real cell types with a transcriptional Pearson correlation >0.5 (Methods). Note that multiple deconvolved clusters could be mapped to the same ground truth cell type. For the gene subset analysis, we still let STdeconvolve use marker genes and HVGs from the internal reference because it is more appropriate to extract genes from the ST data itself, which is the internal reference (rather than the external reference) for the reference-free STdeconvolve method. For mapping STdeconvolve inferred clusters to cell types, however, we used external scRNA-seq references. Therefore for STdeconvolve, we had the same deconvolution results (obtained without any reference), for internal and external reference. Internal and external references made differences only in the cluster-to-cell-type mapping step. When using the default gene subset with the external reference, many of the inferred clusters could not be mapped to actual cell types (particularly for neurons) and therefore exhibited unsatisfactory performance (Supplementary Figure 1C).

We further benchmarked runtime using the sets of 100, 500 and 1000 HVGs (Figure 2D), with the internal reference. Runtime increases linearly with the number of genes in the dataset for most methods. Tangram, Adroit and SPOTlight are among the fastest methods with runtime less than a minute with 1000 genes and 164 spots under inference. The runtimes of cell2location, stereoscope, DSTG and DestVI are heavily dependent on the number of training epochs (Methods). It is difficult to choose the optimal number of training epochs to prevent underfitting. A conservative solution is to use a large number of training epochs, which will consequently increase the runtime.

Evaluation on developing human heart

To further evaluate performance impairment between internal and external references and to assess the impact of major cell types missing in the reference, we carried out analyses using data from the developing human heart [31]. The dataset we utilized is unique in that it contains both single-cell level and spot-level ST data, as well as scRNA-seq data, all derived from biological samples that are similar. The cell types in the single-cell ST data and scRNA-seq data are identical. This allows us to compare RMSE and distance correlation values between internal and external reference. Furthermore, since the spot-level ST data is based on similar biological samples, we can treat the observed cell-type proportions in the single-cell level ST data as working truth and use these to evaluate the estimated cell-type proportions.

The single-cell level ST data is at subcellular resolution, generated by the *in situ* sequencing (ISS) technology. The ISS data contains gene expression of only 69 genes including spatial marker genes and genes important for

cardiac development identified in the authors' previous ST analysis, as well as marker genes for each major cluster in scRNA-seq data. Due to the limited number of genes, no further gene filtering was applied beyond the default data preprocessing procedure (Methods). The cells in the heart ISS dataset were pooled into pseudo-spots each of size 454×424 square pixels according to their spatial coordinates (Figure 3A and B).

When using the internal reference (i.e. ISS single cells) to deconvolve pseudo-spots constructed from ISS data, Adroit, RCTD, stereoscope, DSTG and Tangram show superior performance, similar to our observations in the MOB data, but here with a much smaller number of genes (Figure 3C, Supplementary Figure 7). The atrial cardiomyocytes and ventricular cardiomyocytes are successfully mapped to the atria and ventricular body, respectively. Furthermore, smooth muscle cells are also correctly mapped to the outflow tract, and epicardial cells to the thin outer layer of the heart (Figure 3A and B, Supplementary Figure 7), all of which agree with annotations from the Cell Atlas and previous studies [31, 33, 34]. Cardiomyocytes and ventricular cardiomyocytes are similarly located within the tSNE embedding space, as well as exhibiting colocalization in the ventricular interval and the ventricle, rendering it challenging to distinguish between the two cell types (Supplementary Figures 2, 6A and 7). While STdeconvolve is able to reflect expected spatial patterns, including the differences between ventricular intervals and ventricles, mapping the inferred clusters to their actual cell-type labels remains a challenge for this reference-free method. The other methods tend to generate noisy estimates of cell-type proportions with this small number of genes (Supplementary Figure 7).

When the external reference is employed, only RCTD and stereoscope are capable of capturing the expected spatial distribution of cell types (Supplementary Figure 8). In contrast to results using the internal-reference, all the methods except cell2location suffer performance losses (Figure 3C). Cell2location displays visibly less variation in cell-type proportion estimates across spots when using the internal reference (Supplementary Figure 7). Interestingly, the results are significantly improved when using the external reference (Supplementary Figure 8). However, inferred patterns for subepicardial cells deviate from expectations, which may result from the method's inability to distinguish from epicardial cells that are similar to subepicardial cells (Supplementary Figure 8). Adroit fails to separate atrial cardiomyocytes and cardiomyocytes, which results in the large deviation from truth in both cell types (Supplementary Figure 6B, cell types (7) and (12)). Since Tangram has the option to use estimated cell numbers/cell densities as input, we additionally conducted an ablation study to explore how the performance changes for Tangram with or without the cell density information. We performed analysis in the following five different ways: (1) providing the true number of cells per spot, (2) not providing the true cell

density and using a uniform cell density instead, (3) not providing the true cell density and using a cell density proportional to the number of RNA molecules instead, (4) ignoring cell densities altogether and running the default mapping mode without cell density regularization (mode = 'cells') and (5) ignoring cell densities and running the cluster mode without cell density regularization (mode = 'clusters'). As seen in Supplementary Figure 6E, there was virtually no difference between results from the five approaches, with the only difference being that the 'cells' mode (mode (4)) was slightly worse than the others, potentially because the distribution of cell types was significantly different between external reference and the original single-cell data.

We further examined the robustness of the methods when major cell types in the ST data are missing in the scRNA-seq reference (Supplementary Figure 6C and D). We removed (12) cardiomyocytes from both internal and external reference and examined how the estimated proportions changed. We can see that with the internal reference, most methods transfer the proportion originally attributed to cardiomyocytes to similar cell types. For example, Adroit, RCTD and stereoscope results show a major increase in the estimated proportion of ventricular cardiomyocytes. For SPOTlight, the estimated proportion of cells related to (8) increases (Supplementary Figure 6C). We observe that methods generating relatively weak-differentiating patterns (e.g. cell2location, DestVI, where the estimated proportion for the major cell in each spot is not significantly higher than that for other cell types) tend to be more heavily influenced by the presence of missingness. For these methods, in the presence of missingness, the proportion of the missing cell type is separated into multiple cell types. With the external reference, stereoscope results still display a major increase in the proportion of ventricular cardiomyocytes (Supplementary Figure 6D). Other methods show varying degrees of differences from the original results when the reference still contains cardiomyocytes (Supplementary Figure 6C and D). These differences, however, are largely irrelevant because cardiomyocyte patterns, not captured in the original results, cannot and are not rescued by using a reference missing cardiomyocytes (Supplementary Figure 8). We further examine the noise term in the stereoscope estimates which is claimed to represent the cell types that are present in the ST data but not in the reference. After removing (12), instead of detecting noise in the spots with cardiomyocytes, stereoscope estimates a large proportion of noise in the atrial cardiomyocytes enriched area (Supplementary Figure 9). This may be due to the similarity between cardiomyocytes and ventricular cardiomyocytes and potentially sub-cell types in the atrial cardiomyocytes. Stereoscope assigns the proportion belonging to cardiomyocytes to ventricular cardiomyocytes. And the lack of cardiomyocytes makes stereoscopes over-interpret the part of atrial cardiomyocytes similar to cardiomyocytes to another

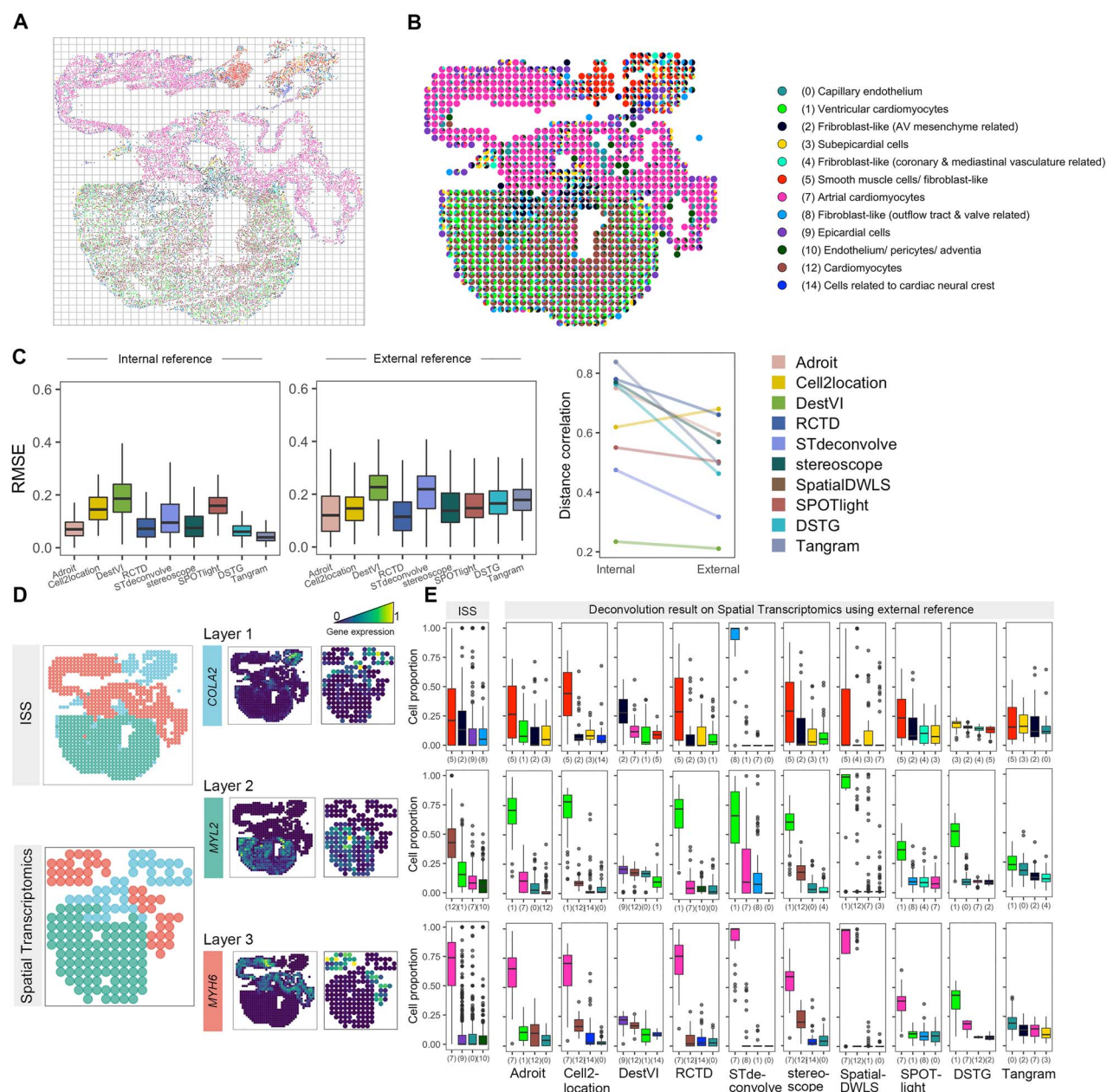


Figure 3. Evaluation on developing human heart data. (A) Overview of the cell atlas in the developing human heart PCW6.5_1 dataset. (B) Cell-type proportion in each pooled spot. Each spot is a pie chart. (C) RMSE and distance correlation of heart cell-type proportion estimates using internal and external reference from nine methods using method-specific default gene subset, on pseudo-spots constructed from single-cell resolution ISS data. (D) Biological layers inferred by BayesSpace in ISS and Spatial Transcriptomics data along with gene expression of marker genes in each layer. (E) Top four spot-wise mean cell-type proportion of each layer in ISS and inferred cell-type proportion in the Spatial Transcriptomics data using external reference. Methods missing in each panel indicate that they did not produce results using the corresponding reference and ST data.

new cell type which becomes the noise term. Since we do not acknowledge how many cell types are missing in real life and such noise estimation could be affected by multiple cell types that share the similar profile, we recommend carefully interpreting and using noise estimation.

So far, we have evaluated the methods on pseudo-spots constructed from spatially resolved single cell data so that we know the true cell-type proportions. In reality, we do not know the truth when having spot-level ST data. As a touchstone of the methods in realistic spot-level ST data, we performed deconvolution on the

spot-level developing heart ST data using the external scRNA-seq reference. This spot-level ST dataset is generated by Spatial Transcriptomics v1.0 where each spot is 100 μm in diameter [6]. Unlike the pseudo-spots we constructed, these real spots do not have true cell-type proportions for us to use in evaluations. Lacking gold standard truth for spot-level data is a challenge encountered by all evaluations presented in the ST deconvolution literature. Many publications rely on visual inspection of expected anatomical patterns or performance in downstream analysis such as clustering of spots into their expected spatial layers [19, 23]. Here,

we adopt an alternative strategy where we evaluate the estimated cell-type proportions across the tissue sample using those from a matched single-cell level ST data as the working truth. Despite differences across samples, tissue samples dissected from similar regions are expected to reflect similar biological structure. For the developing heart data, we divided the tissue into biological regions shared between spot-level and single-cell level ST data. Specifically, we observe three layers in the ISS data where the major cell types are smooth muscle cells (for layer 1), ventricular cardiomyocytes and cardiomyocytes (for layer 2) and atrial cardiomyocytes (for layer 3) (Figure 3D). We infer the layer label of each spot by performing clustering analysis on the pseudo-spots constructed from ISS data as well as the real ST spots separately using BayesSpace [35]. The inferred layers of the ISS and Spatial Transcriptomics datasets largely agree with each other, where the marker genes of each layer show similar gene expression patterns (Figure 3D). We then compared the estimated proportions of major cell types in the Spatial Transcriptomics dataset with the observed proportions of these cell types in the ISS dataset, which we treated as the working truth. We observe that most methods make reasonable inference (Figure 3E, Supplementary Figure 10). In layer 1, most methods have (5) Smooth muscle/fibroblast-like cells and (2) Fibroblast-like (AV mesenchyme related) as the top two cell types. In layer 2, stereoscope, spatialDWLS and cell2location have the same top two cell types as ISS, namely ventricular cardiomyocytes and cardiomyocytes. For almost all the methods, the proportion of ventricular cardiomyocytes is overestimated, which is consistent with the overestimation in the deconvolution performed on the ISS data (Supplementary Figures 7 and 8). In layer 3, all methods fail to capture the pattern of (9) epicardial cells. Overall, except for DestVI and Tangram which fail to capture the major cell types in some layers, all the methods perform relatively well. Among them, stereoscope, cell2location and RCTD exhibit higher agreement with ISS cell composition (Figure 3E).

Evaluation with primary somatosensory cortex

To evaluate how the methods performed on various ST platforms, we analyzed the primary somatosensory cortex area (SSp), a well-studied and well-structured tissue area [3, 9, 11, 36]. Similar to the developing heart, we again used a combination of single-cell and spot-level ST data, where the single-cell ST data comes from the osmFISH platform [9], while spot-level ST data come from 10x Visium Spatial platform [6] and Slide-seqV2 platform [13]. Compared to the Spatial Transcriptomics v1.0, Visium and Slide-seqV2 have finer resolution: 55 μm spot diameter (with a center-to-center distance of 100 μm between two consecutive spots) and $\sim 10\text{-}\mu\text{m}$ mean particle bead diameter, respectively [6].

We similarly started with deconvolution on pseudo-spots constructed from single-cell level ST data, again with internal and external reference (Figure 4A–C), where

the internal reference consists of single cells from the osmFISH single-cell level ST data and the external reference consists of 5392 single cells from scRNA-seq data in the same SSp region generated independently by Yao *et al.* [11]. The osmFISH data had only 33 genes, which we chose not to further filter. To harmonize cell-type labels between osmFISH data and the external reference, we used STANN [37] to provide cell-type labels for osmFISH cells, matching the cell-type labels in the external scRNA-seq reference. The annotated cell-type map displays patterns consistent with the adjacent MOp region chartered by MERFISH data [36] (Figure 4A). When using the internal reference, Adroit, RCTD, stereoscope, DSTG and Tangram again prove best performers, achieving low RMSE (Figure 4B). Most methods are capable of identifying layer patterns with certain methods struggling with one or two cell types. For example, DestVI assigns some proportion of L2/3 IT CTX-1 to L5 NP CTX. Similar to observations in the developing heart, cell2location results in a smoother, blended pattern, again possibly due to the small (here, only 33) number of genes available (Supplementary Figure 13). All the methods are dramatically influenced after changing to the external reference (Figure 4B and C, Supplementary Figure 14). Tangram and DSTG produce the lowest/best RMSE, while STdeconvolve and DestVI have the highest RMSE. The major performance loss when switching from internal to external reference stems from the under-estimation of L2/3 IT CTX-1, L4/5 IT CTX, L6b CTX cells and the over-estimation of Vip cells (Supplementary Figures 12 and 14). STdeconvolve is able to infer twelve clusters, but these clusters can only map to two reference cell types, which leads to suboptimal distance correlation (Supplementary Figure 16). Additionally, we attempted, as references, data from two adjacent tissues: MOp and Visp [11] in order to examine how the methods perform when using references from tissues sections that are not exactly matched. The majority of the methods maintain similar performance in comparison with the external SSp reference (Figure 4C, Supplementary Figure 12), suggesting that three regions are reasonably similar for MOp or Visp to serve as a sensible external reference for deconvolving ST data in the SSp region.

We next performed deconvolution on the Visium and Slide-seqV2 data using the external SSp reference. We compare the inferred major cell types at each spot with the osmFISH data (Figure 4D and E, Supplementary Figure 15). Adroit, cell2location, DestVI, RCTD, stereoscope and SPOTlight all show patterns of major cell types consistent with those revealed from the osmFISH cell atlas when performing deconvolution on the Visium data (Figure 4D and E). Among them, Adroit, stereoscope and SPOTlight display a smoother pattern (Supplementary Figure 15). In contrast to the performance on the pseudo-spots constructed from the osmFISH data, Tangram fails to capture the expected patterns. One possible reason is still the non-negligible difference between external scRNA-seq reference and

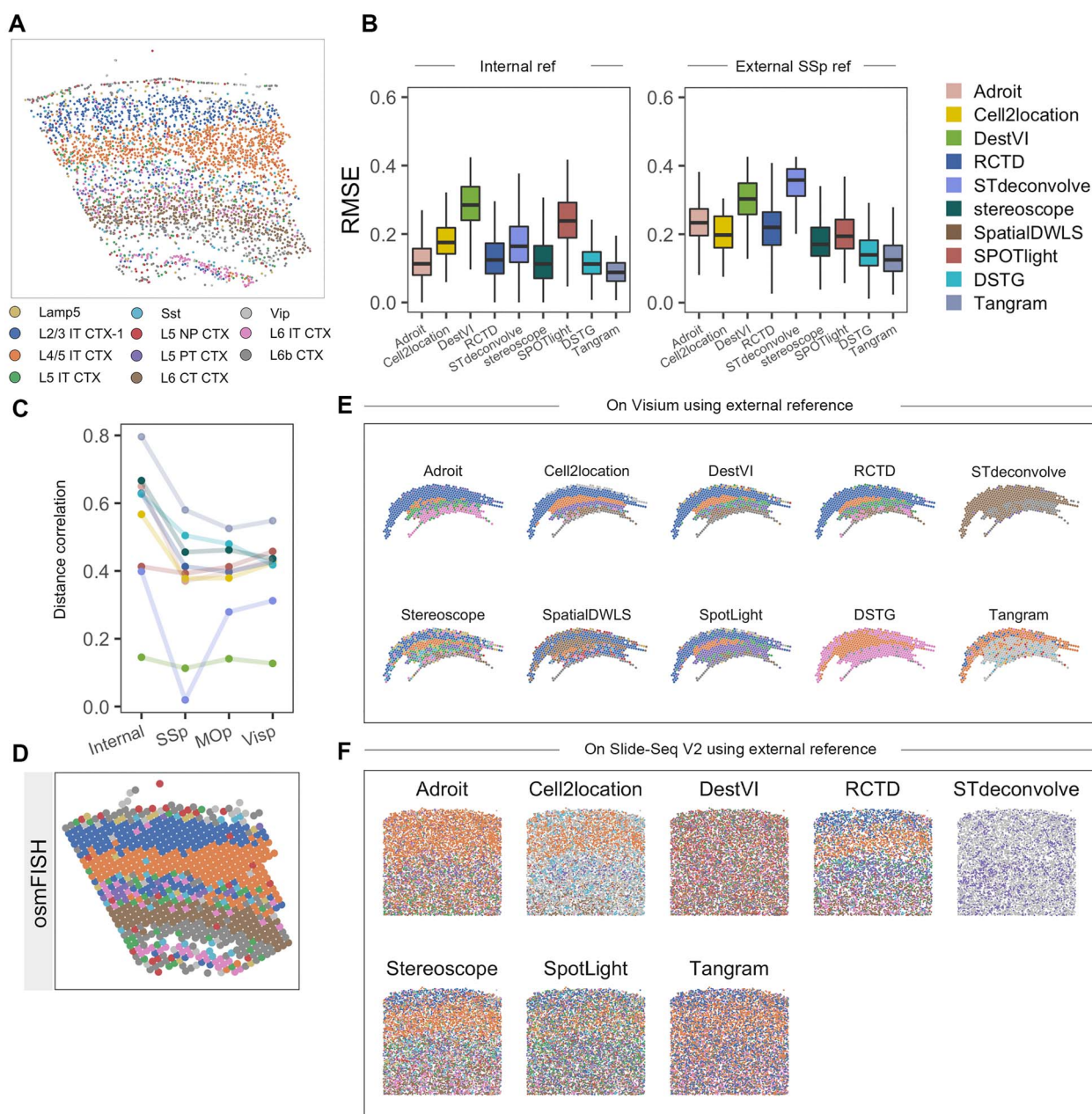


Figure 4. Evaluation on mouse SSp data. **(A)** Overview of the cell atlas in the single-cell osmFISH mouse SSp dataset. **(B)** RMSE of cell-type proportion estimates using internal and external reference from 9 methods using all genes, on pseudo-spots constructed from single-cell resolution osmFISH data. **(C)** Distance correlation of cell-type proportion estimates using internal SSp, external SSp, external Visp, external MOP references. **(D)** Major cell types of each pseudo-spot in the osmFISH data. **(E)** Major estimated cell type of each spot inferred by 10 methods on the Visium mouse SSp dataset with the default gene subset using the external SSp reference. **(F)** Major estimated cell types of each bead/spot inferred by 8 methods on the Slide-seqV2 mouse SSp dataset with the default gene subset using the external SSp reference. Methods missing in each panel indicate that they did not produce results using the corresponding reference and ST data.

ST data. When applied to the Slide-seqV2 data, RCTD and stereoscope agree the most with the osmFISH reference cell atlas. Adroit fails to accurately capture the proportion of L2/3 IT CTX-1 cells and other methods provide a rather noisy cell atlas (Figure 4F). We note that the performance of DSTG in the SSp Visium tissue differs substantially from their own evaluation. This may be due to stochasticity incurred by randomly selecting cells from scRNA-seq reference data. Moreover, hyperparameters like the number of nearest neighbors

used in constructing the graph also influence the final results.

As a conclusion, most methods, especially Tangram and DSTG, achieve excellent performance when using perfectly matched internal references. With internal reference, Adroit, cell2location, RCTD and stereoscope still provide satisfactory cell-type proportion estimation despite the limited number of genes available. RCTD and stereoscope outperform other methods when external reference is used regardless of the gene number and

platform. Cell2location performs reasonably well on Spatial Transcriptomics and Visium data when there are a sufficient number of genes.

Discussion

As ST technologies continue to evolve rapidly, we anticipate more advanced technology that captures single-cell resolution data as well as measures expression levels of as many genes as possible. Commercially available sequencing-based ST technologies, however, cannot yet achieve single-cell resolution when measuring whole transcriptome profiles. Data generated from these technologies, thus entail the inference of cell-type composition. Deconvolution of ST data has already been demonstrated to aid downstream analysis such as detecting spatial expression patterns in spot-level data [38]. It is essential to accurately estimate cell-type composition and make appropriate adjustments accordingly to ensure validity and enhance power in downstream analysis, including identification of spatial domains and spatially variable genes [39], study of cell-cell communication and evaluation of the impact on molecular function and ultimate phenotype [1, 2, 40]. An incorrect inference of cell-type composition could lead to misunderstandings of tissue structure and function, including spurious findings and impaired power in various other downstream analyses. In this review, we have benchmarked the performance of 10 ST deconvolution methods using six real datasets.

Our study evaluates the performance in three tissues: MOB, developing human heart and mouse SSP region. We used a combination of single-cell resolution ST data and spot-level ST data. The advantage of using pseudo-spots constructed from single-cell resolution ST data, by pooling cells into pseudo-spots according to their spatial coordinates, is that the true cell-type proportions are established based on the contributing cells at each constructed spot. The potential issue is that pseudo-spots may not reflect characteristics of real ST spots. We assess whether the pseudo-spots constructed from single-cell resolution ST data exhibit similar characteristics as the real spot-level ST data by checking the gene expression distribution across data sources (Supplementary Figures 17–19). Most of the genes display very similar patterns across data sources, suggesting that our evaluations on the pseudo-spots could reflect performance on real spots. We quantify the deconvolution performance using RMSE and distance correlation. We further examine the difference between ground truth and estimated cell-type proportion of each cell type.

For spot-level ST data, we do not have ground truth, rendering evaluations more challenging. As aforementioned, developers of these deconvolution methods, encountering similar difficulties, used visual inspection of expected structural patterns or assessment of performance in downstream analysis (e.g. clustering

spots into their expected anatomical layers). Here, we compare the inferred cell-type proportions to carefully matched single-cell resolution datasets.

We evaluate all methods using both internal and external references. Most methods perform reasonably well when using the internal reference, especially for Tangram and DSTG. Tangram infers cell composition by mapping cells in reference to their spatial origin, and DSTG constructs synthetic ST data by randomly pooling cells from the reference. When the internal reference is employed, the data perfectly fit the assumption of these two methods, which explains their best performance. In comparison, DestVI appears to show inferior performance when applied to pseudo-spots constructed from single-cell resolution ST data. This may be due to the lack of continuous variation within cell types or the limited number of spots/genes to train the complex latent variable model. When we evaluate the methods using an external reference, most methods show a non-negligible performance loss. Among them, RCTD and stereoscope demonstrate robustness and remain the top performers. Cell2location shows comparable performance when the gene number is sufficient (e.g. >100). When the number of genes is small, cell2location suffers dramatically, potentially because the multilayer Bayesian modeling in cell2location becomes too complicated to obtain a good fit for the shared parameters between genes. RCTD corrects potential batch effects through an additional platform effect normalization step by quantifying the random effect of each gene. Stereoscope and cell2location both incorporate a gene-specific parameter when modeling gene expression distribution. These modeling features empower the three methods to accommodate potential systematic differences between scRNA-seq reference and target ST data, resulting in their robust performance when switching from internal to external reference.

Additionally, we perform experiments to evaluate the impact of choices of genes used for inference. We have benchmarked robustness and time complexity, providing insights regarding the best choice of gene subset for each method. Most methods attain their best performance with the default gene subset and perform similarly with a comparable number of either marker genes or HVGs. SPOTlight and spatialDWLS exhibit significantly improved performance when employing marker gene subsets (in contrast to HVGs), which is due to the nature of the NMF + NNLS method that requires a marker gene profile for each cell type.

STdeconvolve, as the only reference-free method in our evaluation, shows rather unstable performance across tissue, gene subsets and references due to a few reasons. The main advantage of STdeconvolve is that no reference is needed. Its LDA framework offers a flexible and intuitive way to model spots as a finite mixture of an underlying set of cell types. LDA and thus STdeconvolve may fail to deconvolve cell types that have very similar transcriptional profiles or those that cannot

be well differentiated by differences in gene expression, volume or morphology. STdeconvolve is expected to work better with a large number of spots, ideally in thousands, while some of the ST data we evaluated contain only a few hundred spots. Another practical consideration is that we still need to annotate the clusters inferred by STdeconvolve to their corresponding cell types, which is usually achieved by using scRNA-seq data. It can be challenging to compare STdeconvolve with other reference-based methods when inferred clusters cannot be mapped to real cell types. Therefore, we effectively still need scRNA-seq data as reference, but post-inference for reference-free STdeconvolve rather than pre-inference for all other reference-based methods.

The three tissues we analyzed exhibit distinct tissue structures. There are seven FOVs of MOB that are predominantly constituted by neurons and interneurons. In comparison, a more clustered structure can be seen during the development of human heart tissue, where spots are composed of more cell types than MOB or mouse SSp, and therefore have a higher entropy. Mouse SSp is a multi-layered tissue, and cell types change as depth within the cortex increases. For each layer, there are a small number of major cell types in each spot, explaining its small entropy (Supplementary Figure 2D). The majority of the methods perform better on MOB and mouse SSp tissues, where the cellular composition is simpler than that of the developing human heart. In the developing human heart, by contrast, several similar cell types (especially cardiomyocytes, ventricular cardiomyocytes and atrial cardiomyocytes) span across the whole tissue and colocalize with other cell types. For this tissue with more complex cell-type composition, stereoscope and RCTD are observed to be more effective at distinguishing the pattern (Supplementary Figure 8).

In summary, RCTD and stereoscope exhibit consistently high performance across tissues. STdeconvolve, as the only reference-free method, has the capability for identifying tissue structure and cell mixture, but cell-type mapping must be addressed carefully. We have thoroughly evaluated various scenarios, encompassing different tissues, varying technologies and data resolution, different numbers of single cells and spots, as well as varying number and type of genes employed for analysis. Based on our results, we recommend that investigators first identify some of our evaluated scenario(s) that best match their own data and select best performing methods under these scenario(s). The choice of reference, preferably from carefully matched tissue and biological samples, is also essential for deconvolving ST data. Mismatched scRNA-seq references or references with inaccurately annotated cells could severely impair deconvolution performance. In addition, while out of the scope of this work, denoising and dimension reduction of noisy and high dimensional ST data can allow more effective information extraction [41]. We also anticipate that cell-type deconvolution further benefits from development and advancement of methods that effectively

denoise and reduce the dimension of ST data. In the meantime, we believe that our comprehensive evaluation results, along with careful review of the theoretical and modeling properties of the methods, provide useful guidelines for the deconvolution of ST data.

Methods

Evaluation metrics

For single-cell resolution ST datasets, we pooled the cells according to their spatial coordinates to construct pseudo-ST spots to mimic real ST spots. In this way, we have the truth of the cell-type mixture in each pseudo spot. Then RSME, distance correlation and differences were computed between the estimated and ground truth cell-type proportion.

We denote the ground truth cell-type proportion of cell type r at spot k as y_{rk} and the estimated proportion as \hat{y}_{rk} . RSME, difference (which is cell type-specific) and distance correlation are calculated as follows:

$RMSE_k = \sqrt{\frac{\sum_r (y_{rk} - \hat{y}_{rk})^2}{R}}$ where R is the total number of cell types.

$$\text{Difference} = \hat{y}_{rk} - y_{rk}$$

The empirical distance covariance $V^2(Y_r, \hat{Y}_r)$ of cell type r is defined by

$$V^2(Y_r, \hat{Y}_r) = (1/K^2) \sum_{k,l} A_{rkl} B_{rkl}$$

where

$$a_{rkl} = y_{rk} - y_{rl}, b_{rkl} = \hat{y}_{rk} - \hat{y}_{rl} \quad k, l = 1, \dots, K,$$

$$A_{rkl} = a_{rkl} - \overline{a_{rk}} - \overline{a_{rl}} + \overline{a_{r..}}$$

$$B_{rkl} = b_{rkl} - \overline{b_{rk}} - \overline{b_{rl}} + \overline{b_{r..}}$$

and the subscript denotes that the mean is computed for the index that it replaces. K is the total number of spots.

Similarly, $V^2(Y_r)$ is the non-negative number defined by

$$V^2(Y_r) = V^2(Y_r, Y_r) = (1/K^2) \sum_{k,l} A_{rkl}^2$$

The empirical distance correlation $dcor(Y_r, \hat{Y}_r)$ is defined as

$$dcor(Y_r, \hat{Y}_r) = V^2(Y_r, \hat{Y}_r) / \sqrt{V^2(Y_r) V^2(\hat{Y}_r)}$$

Then the distance correlation across cell types is calculated by $dcor(Y, \hat{Y}) = \frac{\sum_r dcor(Y_r, \hat{Y}_r)}{R}$ where again R is the total number of cell types [42].

Entropy to quantify tissue complexity

To characterize the complexities of the various types of tissues in our analysis, we computed entropy for each pseudo-ST spot similar to [43]. Specifically, suppose for an artificial spot k , there exists R different cell types and the true mixing proportion of cell type r is y_{rk} , then the spot-level entropy is calculated as:

$$H(k) = - \sum_{r=1}^R y_{rk} \log(y_{rk})$$

A higher entropy value indicates a more complex mixing structure for the corresponding spot, while a lower entropy value indicates that the spot is primarily composed of one or few cell types.

Data preprocessing

We utilized six data sets to test the ST deconvolution methods in this study. The data preprocessing steps are summarized as follows.

SeqFISH+ generates the single-cell level ST data from MOB. For each field of view (FOV), all cells that are located in the same 400×400 square pixel area are pooled into one spot. The single-cell level data is also used as the internal reference. Genes expressed in at least 3 cells and cells that have at least 200 features and at most 2500 features are kept in analysis. Seven fields of views are combined as one input file for all eight methods. For MOB external scRNA-seq data, we only kept genes that are present in at least 2% of cells and cells that have at least 200 features and at most 2500 features.

The human heart ISS data is also a single-cell resolution ST data. There are two slides of ISS data and we analyzed the PCW6.5_1 slide (picked randomly from the 2). Genes expressed in at least three cells are kept in analysis. Cells that are co-located in the same 454×424 square pixel area were pooled as one ST spot. For human heart Spatial Transcriptomics ST data, we chose the FH6_1000L2_CN74_D1 slide. We only kept genes that are present in at least 2% of spots and spots that have at least 200 features and at most 3000 features. For the heart external scRNA-seq data, we removed immune cells and erythrocytes to match the ISS data. Genes that are present in at least 2% of cells and cells that have at least 200 features and at most 5000 features were kept in the analysis. We note in the name of (2) and (4) cell types, 'fribroblast-like' is likely a typo and probably should be 'fibroblast-like'. We did not change it in the legend to be consistent with the original paper [31].

For the mouse brain's primary somatosensory region (SSp), we have the osmFISH data, which is a single-cell level ST data. Similarly, cells that are located in the same 800×800 square pixel area are pooled as one spot. The single-cell level data are also used as the internal reference. We used two spot-level ST data: Visium and Slide-seqV2 [13, 15]. Mouse brain single-cell data are obtained from Yao et al. [11] and only cells from the SSp, Visp and

MOp regions are kept in analysis, with cells from each region separately used as reference. For ST data, we only keep genes that are present in at least 2% of spots. For scRNA-seq data, we remove cells that are not confidently assigned a class label by the original paper. Only genes that exist in both scRNA-seq and ST data are used in deconvolution analysis.

Selection and processing of single-cell level ST data to match spot-level ST data

To reasonably interpret deconvolution results in spot-level ST data, we carefully match them to single-cell resolution ST data. Specifically, single-cell resolution ST data from similar biological samples or with similar tissue structure are selected. For the developing human heart tissue, we select the single-cell level PCW6.5_1 ISS data to match the spot-level PCW6.5 FH6_1000L2_CN74_D1 Spatial Transcriptomics slide [31]. For the SSp tissue, we crop the single-cell level osmFISH [9] data by keeping row (or y) pixel $> 22\,880$ to match the shape of spot-level Visium ST data (specifically, the ST8059048 slide). The SSp spots, as well as some ambiguous adjacent spots, are selected by our pathologist to keep in the analysis. For spot-level Slide-seqV2 data, we use the Puck_200306_03 slide and crop the data with $2300 \leq x \text{ pixel} \leq 4000$ and $500 \leq y \text{ pixel} \leq 2300$ according to Figure S4A in the Slide-seqV2 paper [13]. For MOB, we matched the Rep8 Spatial Transcriptomics [6] slide to the seqFISH+ data. We only used the MOB Spatial Transcriptomics for calculating the gene expression distribution (Supplementary Figure 17).

Gene subsetting and gene subset employed in the ST deconvolution methods

All gene subsettings are accomplished using the R package Seurat [25]. HVGs are selected using feature variance calculated by the FindVariableFeatures function with default settings. Marker genes are selected using the FindAllMarkers function with a log-fold-change threshold of 0.75. Both positive and negative markers are included in the marker gene subset. Top marker genes are selected according to P-values.

Default gene subset used in each deconvolution method

Adroit, cell2location, DestVI, stereoscope and Tangram [14–16, 19, 23] do not have a built-in gene filtering strategy. Top 2000 HVGs are used in the analysis. SPOTlight [21] uses a mixture of marker genes for each cell type and the top 500 HVGs. Note here we follow the SPOTlight pipeline and keep only positive markers when selecting marker genes. We use the default parameters of RCTD and run RCTD in full mode. By default, RCTD [17] has a built-in marker gene selection step where only genes with normalized gene expression ≥ 0.0002 are included, and it selects cell-type marker genes based on a log-fold-change threshold of 0.75. Only selected cell-type marker genes are fed into RCTD. For STdeconvolve inference, following the software pipeline, we remove genes detected

in <2% of spots or genes expressed in all spots. STdeconvolve then selects genes by choosing significantly over-dispersed genes across spots to detect transcriptionally distinct cell types. By default, only the top 1000 or fewer most over-dispersed genes are retained for STdeconvolve inference. DSTG selects the top 2000 most variable genes by default, across different cell types in the reference scRNA-seq data according to adjusted ANOVA P-values with Bonferroni correction. SpatialDWLS selects cell-type marker genes from the scRNA-seq data by Giotto's [30] built-in differential expression analysis tools.

Note that after selecting the gene subset, there may appear a situation that some of the spots do not express some of the selected gene(s). Such spots are removed from further analysis.

Other technical details

DSTG and spatialDWLS failed to run with the mouse brain SSF Slide-seq V2 data, which is the sparsest dataset among all ST data we analyzed, containing 97% zero counts. SpatialDWLS did not run for ST data with a small number of genes (e.g. ISS and osmFISH). For the gene subset analysis on MOB seqFISH+ data with internal reference, when the number of genes was moderate or small, the makeSignMatrixDWLSfromMatrix function from Giotto suite 2.0.0.997 could not build a gene signature matrix for every cell type (e.g. interneurons); therefore we did not include those results for the gene subset analysis with internal reference for spatialDWLS.

STdeconvolve employs an LDA framework [27]. Several distinctive aspects of ST data make the LDA framework an appropriate choice, including but not limited to the relatively small number of cells and cell types present in each spot, the relatively large number of spots compared to the number of cell types, and the heterogeneity of cell-type composition across spots. The LDA framework in STdeconvolve follows the standard LDA framework, where spots, cell types and genes in ST data correspond to documents, topics and words in standard LDA respectively.

STdeconvolve additionally removes spots with library size smaller than 100. We follow this guidance for the developing human heart, and mouse brain data. But for the seqFISH+ data, we choose to be more lenient and keep all spots with library size >0 since the spot number is limited (only 164 spots total). Moreover, STdeconvolve filters out genes expressed in <2% of the spots. Finally, we set the number of clusters the same as the number of cell types in the corresponding reference scRNA-seq dataset. All other parameters are set as default. To select the optimal number of cell types when not pre-specified, STdeconvolve calculates a perplexity for each model based on the posterior likelihood of the observed data conditional on deconvolved cell-type assignments. STdeconvolve also reports the number of rare cell types (average mean spot proportion < 5% across spots) to help set an upper bound on the total number of deconvolved

cell types. The optimal number of cell types is chosen with lowest perplexity and minimized number of rare cell types.

To annotate STdeconvolve inferred clusters, we followed the transcriptional correlations method described in the software documentation '[Annotating deconvolved cell-types](#)' section, where we computed Pearson correlation of transcriptional profiles between each inferred cluster and reference biological cell types from matched single-cell RNA-seq data. Inferred clusters were annotated to the reference cell type with the highest Pearson correlation that was >0.5. We used the same scRNA-seq dataset that was used as reference in other methods for fair comparison. We mapped only one reference cell type to each inferred cluster. However, multiple inferred clusters can be mapped to the same reference cell type. When this happened, we added up the estimated proportions from multiple inferred clusters as the final predicted proportion for that reference cell type for evaluation. Note here, there might be some reference cell types that had a correlation >0.5 with inferred clusters but ended up not having any inferred cluster mapped to it. This happens when the candidate inferred cluster(s) had a higher correlation with some other reference cell type. For example, for the developing human heart deconvolution with internal reference ([Supplementary Figure 11](#)), reference cell type (4) has 0.89 correlation with inferred cluster 1 and 0.92 correlation with inferred cluster 2. However, inferred cluster 1 has 0.95 correlation with reference cell type (3) and inferred cluster 2 has 0.93 correlation with reference cell type (8). With the above conditions, reference cell type (4) is excluded from evaluation.

The cell-type labels of inferred clusters can be further evaluated by performing a rank-based gene set enrichment analysis of upregulated genes in each cell type. We can choose to annotate only clusters where the enrichment P-value is significant at a certain threshold, and/or highest positive edge and enrichment score greater than certain thresholds. Such a strategy, however, may produce inconsistent results with the transcriptional correlations analysis described above. For example, in our analysis of the human heart Spatial Transcriptomics data with external reference, we observed that most clusters can be mapped to a known cell type in the scRNA-seq reference. For example, inferred cluster 2 was mapped to ventricular cardiomyocytes based on gene set enrichment analysis, but it also had the highest transcriptional correlation with atrial cardiomyocytes. While clusters 3 and 4 had the highest transcriptional correlation with fibroblast-like (outflow tract & valve related), they were mapped to epicardium-derived cells by gene set enrichment analysis ([Supplementary Figure 11](#)). In contrast, for the mouse brain osmFISH data with external reference, only cluster 11 could be mapped to a known cell type L4/5 IT CTX; however, the transcriptional correlation was only -0.01 ([Supplementary Figure 16](#)). Therefore, we chose not to rely on gene set enrichment analysis for cell-type annotation in our analysis.

For DSTG, we set $k = 100$ as the number of nearest neighbors in our analysis for the human heart ISS, mouse brain SSp 10x, and mouse brain SSp osmFISH datasets, since these data have a relatively large number of spots. For the human heart ST, MOB ST and MOB seqFISH+ default analyses, we tested $k = 20, 50$ and 100 to evaluate the impact of the number of nearest neighbors on final results. For the MOB seqFISH+ gene subset analysis, we set $k = 20$ following $k = 20$ performed best in the default analysis and also the relatively small number of spots. All other parameters were set as default unless otherwise specified.

For cell2location, we used 6000 epochs in the single-cell inference and 30 000 epochs in the ST deconvolution. For DestVI, the parameter settings are according to the 2021.10.1 tutorial. We used 500 epochs in the single-cell inference and 4000 epochs in the ST deconvolution. For stereoscope, 30 000 epochs were used in both single-cell inference and ST deconvolution. All computations for these three methods are performed using the NVIDIA GeForce RTX 3070 GPU. For DSTG, the maximum number of epochs was set to 200. For SPOTlight, 300 cells per cell type were employed in the analysis.

Tangram employs an initial step of cell segmentation to calculate the number of cells for each spot and uses it as an input to calculate the fraction of cells per spot (a spatial density prior). For pseudo-ST spots constructed from single cells, we use the true number of cells per spot as the input. For real spot-level data, we use the watershed algorithm in Squidpy [44] python package to carry out cell segmentation when histological images are available. When histological images are not available (such as for Slide-seq data) or cell segmentation results are not of reasonable quality, we use a uniform spatial density prior as input for Tangram. All computations are performed using the Tesla V100-SXM2 GPU.

T-SNE coordinates were calculated using Seurat_3.2.3 in R v3.6.0 [25]. We analyzed each scRNA-seq dataset following the standard pipeline with default Seurat parameter setting. Principal component analysis (PCA) was performed on genes using RunPCA and the top 10 PCs were used as input for running RunTSNE in Seurat.

For the mouse olfactory bulb data, we only evaluate the cell-type proportion for neurons, astrocytes, oligodendrocytes, microglia, endothelial cells and olfactory ensheathing cells when using external reference. There are many different neuron cells both in the seqFISH+ data (which were used to construct pseudo-spots) and in the external scRNA-seq reference. We pooled various neuron cell types into one cell type: Neuron. Specifically, for scRNA-seq data, we pooled seven cell types (n18-EPL-IN, Neuron_AstrocyteLike, Neuron_GC, Neuron_Inmature, Neuron_M/TC, Neuron_PGC, Neuron_Transition) as Neuron; for seqFISH+, we pooled three cell types (Interneuron, Neuroblast, Mitral/Tufted Cells) as Neuron.

To compare the performance of internal versus external reference in mouse SSp data, cells in the

osmFISH dataset were mapped, with STANN [37], to new cell types using the processed external SSp scRNA-seq as reference. We followed the pipeline described in <https://github.com/sameelab/STANN>.

Key Points

- Cell mixture inference is a critical step in the analysis of spatial transcriptomics (ST) data to mitigate potential confounding caused by differential cell-type proportions across spots in downstream analysis.
- Existing ST deconvolution methods can be classified into three groups: probabilistic-based, non-negative matrix factorization and non-negative least squares based, and other methods.
- We compare 10 ST deconvolution methods using three single cell resolution and three non-single-cell spot resolution ST datasets. We provide practical guidelines for method choice under different scenarios as well as the optimal subsets of genes to use for inference.

Supplementary data

Supplementary data are available online at [https://academic.oup.com/bib](https://academic.oup.com/bib/article/23/4/bbac245/6618233).

Acknowledgements

We thank Li lab members for providing advice on data preprocessing, method selection and feedback on the manuscript. Figure 1 is created via bioRender.

Funding

The research is partly supported by National Institutes of Health (NIH) grants (U01HG011720, U01DA052713 and P50HD103573).

Data Availability

All datasets used in this study are publicly available. The detailed reference and the websites to download them are in Table 2. The pathologist's annotation of mouse brain SSp Visium is available upon request. The code used to calculate evaluation metrics in the analysis and the processed datasets are available at <https://github.com/JiawenChenn/St-review>.

References

1. Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods* 2022;**19**:534–546.
2. Rao A, Barkley D, França GS, et al. Exploring tissue architecture using spatial transcriptomics. *Nature* 2021;**596**(7871):211–20.
3. Xia C, Fan J, Emanuel G, et al. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc Natl Acad Sci* 2019;**116**(39):19490–9.

4. Wang X, Wang X, Allen WE, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;**361**(6400):eaat5691.
5. Eng C-HL, Lawson M, Zhu Q, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019;**568**(7751):235–9.
6. Ståhl Patrik L, Ståhl PL, Salmén F, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**(6294):78–82.
7. Crosetto N, Bienko M, van Oudenaarden A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* 2015;**16**(1):57–66.
8. Femino Andrea M, Fay FS, Fogarty K, et al. Visualization of single RNA transcripts in situ. *Science* 1998;**280**(5363):585–90.
9. Codeluppi S, Borm LE, Zeisel A, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods* 2018;**15**(11):932–5.
10. Tasic B, Yao Z, Graybuck LT, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 2018;**563**(7729):72–8.
11. Yao Z, van Velthoven CTJ, Nguyen TN, et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* 2021;**184**(12):3222–3241.e26.
12. Lee JH, Daugharthy ER, Scheiman J, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat Protoc* 2015;**10**(3):442–58.
13. Stickels RR, Murray E, Kumar P, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 2021;**39**(3):313–9.
14. Yang T, Alessandri-Haber N, Fury W, et al. AdRoit is an accurate and robust method to infer complex transcriptome composition. *Communications Biology* 2021;**4**(1):1218.
15. Kleshchevnikov V, Shmatko A, Dann E, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* 2022;**40**:661–71.
16. Lopez R, Li B, Keren-Shaul H, et al. DestVI identifies continuums of cell types in spatial transcriptomics data. *Nat Biotechnol* 2022.
17. Cable DM, Murray E, Zou LS, et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol* 2022;**40**:517–526.
18. Miller BF, Huang F, et al. Reference-free cell-type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nat Commun* 2022;**13**:2339.
19. Andersson A, Bergenstråhle J, Asp M, et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun Biol* 2020;**3**(1):565.
20. Dong R, Yuan G-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol* 2021;**22**(1):145.
21. Elosua-Bayes M, Nieto P, Mereu E, et al. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res* 2021;**49**(9):e50–0.
22. Song Q, Su J. DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Brief Bioinf* 2021;**22**(5):bbaa414.
23. Biancalani T, Scalia G, Buffoni L, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods* 2021;**18**(11):1352–62.
24. Moncada, R., Barkley, D., Wagner, F. et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol*, 2020;**38**:333–342. <https://doi.org/10.1038/s41587-019-0392-8>.
25. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**(13):3573–87.e29.
26. Dey KK, Hsiao CJ, Stephens M. Visualizing the structure of RNA-seq expression data using grade of membership models. *PLoS Genet* 2017;**13**(3):e1006599.
27. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003;**3**:993–1022.
28. Tsoucas D, Dong R, Chen H, et al. Accurate estimation of cell-type composition from gene expression data. *Nat Commun* 2019;**10**(1):2975.
29. Kim S-Y, Volsky DJ. PAGE: parametric analysis of gene set enrichment. *BMC Bioinf* 2005;**6**(1):144.
30. Dries R, Zhu Q, Dong R, et al. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 2021;**22**(1):78.
31. Asp M, Giacomello S, Larsson L, et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* 2019;**179**(7):1647–60.e19.
32. Tepe B, Hill MC, Pekarek BT, et al. Single-cell RNA-Seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell Rep* 2018;**25**(10):2689–703.e3.
33. Sawada H, Rateri DL, Moorleggen JJ, et al. Smooth muscle cells derived from second heart field and cardiac neural crest reside in spatially distinct domains in the media of the ascending aorta—brief report. *Arterioscler Thromb Vasc Biol* 2017;**37**(9):1722–6.
34. Eralp I, Lie-Venema H, Bax NAM, et al. Epicardium-derived cells are important for correct development of the Purkinje fibers in the avian heart. *Anat Rec A Discov Mol Cell Evol Biol* 2006;**288A**(12):1272–80.
35. Zhao E, Stone MR, Ren X, et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* 2021;**39**(11):1375–84.
36. Zhang M, Eichhorn SW, Zingg B, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* 2021;**598**(7879):137–43.
37. Grisanti Canozo FJ, Zuo Z, Martin JF, et al. Cell-type modeling in spatial transcriptomics data elucidates spatially variable colocalization and communication between cell-types in mouse brain. *Cell Syst* 2022;**13**(1):58–70.e5.
38. Zhu J, Sun S, Zhou X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biol* 2021;**22**(1):184.
39. Hu J, Li X, Coleman K, et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 2021;**18**(11):1342–51.
40. Longo SK, Guo MG, Ji AL, et al. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* 2021;**22**(10):627–44.
41. Ni Z, Prasad A, Chen S, et al. SpotClean adjusts for spot swapping in spatial transcriptomics data. *Nat Commun* 2022;**13**:2971.
42. Gábor JS, Maria LR. Brownian distance covariance. *Ann Appl Stat* 2009;**3**(4):1236–65.
43. Andersson A, Larsson L, Stenbeck L, et al. Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun* 2021;**12**(1):6012.
44. Palla G, Spitzer H, Klein M, et al. Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* 2022;**19**(2):171–8.