

## Gene expression

# COME: contrastive mapping learning for spatial reconstruction of single-cell RNA sequencing data

Xindian Wei<sup>1</sup> , Tianyi Chen<sup>1</sup> , Xibiao Wang<sup>2</sup> , Wenjun Shen<sup>3</sup>, Cheng Liu<sup>2,\*</sup> , Si Wu<sup>4</sup>,  
Hau-San Wong<sup>1,\*</sup>

<sup>1</sup>Department of Computer Science, City University of Hong Kong, Kowloon 999077, Hong Kong

<sup>2</sup>Department of Computer Science, Shantou University, Shantou 515063, China

<sup>3</sup>Department of Bioinformatics, Shantou University Medical College, Shantou 515041, China

<sup>4</sup>Department of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China

\*Corresponding authors. Cheng Liu, Department of Computer Science, Shantou University, 243 Daxue Road, Jinping District, Shantou, 515063, China. E-mail: chengliu10@gmail.com; Hau-San Wong, Department of Computer Science, City University of Hong Kong, Tat Chee Avenue, Kowloon 999077, Hong Kong. E-mail: cshswong@cityu.edu.hk.

Associate Editor: Inanc Birol

## Abstract

**Motivation:** Single-cell RNA sequencing (scRNA-seq) enables high-throughput transcriptomic profiling at single-cell resolution. The inherent spatial location is crucial for understanding how single cells orchestrate multicellular functions and drive diseases. However, spatial information is often lost during tissue dissociation. Spatial transcriptomic (ST) technologies can provide precise spatial gene expression atlas, while their practicality is constrained by the number of genes they can assay or the associated costs at a larger scale and the fine-grained cell-type annotation. By transferring knowledge between scRNA-seq and ST data through cell correspondence learning, it is possible to recover the spatial properties inherent in scRNA-seq datasets.

**Results:** In this study, we introduce COME, a Contrastive Mapping LEarning approach that learns mapping between ST and scRNA-seq data to recover the spatial information of scRNA-seq data. Extensive experiments demonstrate that the proposed COME method effectively captures precise cell-spot relationships and outperforms previous methods in recovering spatial location for scRNA-seq data. More importantly, our method is capable of precisely identifying biologically meaningful information within the data, such as the spatial structure of missing genes, spatial hierarchical patterns, and the cell-type compositions for each spot. These results indicate that the proposed COME method can help to understand the heterogeneity and activities among cells within tissue environments.

**Availability and implementation:** The COME is freely available in GitHub (<https://github.com/cindyway/COME>)

## 1 Introduction

The advent of high-throughput single-cell RNA sequencing (scRNA-seq) technologies has enabled gene expression profiling in thousands of individual cells. This has revolutionized our ability to study cellular processes with single-cell resolution. However, a notable limitation of scRNA-seq protocols is that they necessitate the dissociation of tissues, resulting in the loss of crucial information concerning the native spatial context and cellular positioning. In contrast, this spatial information is essential for comprehending biological mechanisms such as the localization of cell types, intercellular communication, and cellular trajectories.

Spatially resolved transcriptomics can bridge the gap between cell molecular identities and organ-level tissue organization. Spatial transcriptomic (ST) technologies record both the spatial position and gene expression levels of spots on a tissue. However, imaging-based ST technologies (Williams *et al.* 2022), such as including multiplexed error-robust fluorescence *in situ* hybridization (MERFISH) (Chen *et al.* 2015), single-molecule FISH (smFISH) (Codeluppi *et al.* 2018),

spatially resolved transcript amplicon readout mapping (STARmap) (Wang *et al.* 2018), and sequential fluorescence *in situ* hybridization (seqFISH+) (Eng *et al.* 2019), can be equipment-intensive, time-consuming and have limited per-spot gene detection capacity if they offer higher spatial resolutions. In contrast to ST, scRNA-seq can capture a significantly higher number of genes for each cell, and it is relatively easier to obtain a larger number of scRNA-seq datasets (Longo *et al.* 2021). While sequence-based ST, such as 10x Visium, enables the detection of numerous gene expressions at each spot, it lacks single-cell resolution and cell-type annotations characteristic of scRNA-seq data. Therefore, establishing cell correspondences between scRNA-seq and ST datasets enables the possibility of reconstructing the spatial information for scRNA-seq cells. This enables the transfer of knowledge across the two modalities, allowing for the retrieval of the inherent spatial characteristics of single-cell data.

Early methods for cell correspondence learning between scRNA-seq and ST data are mainly based on correlating gene expression patterns across scRNA-seq and ST modalities.

Received: 3 July 2024; Revised: 12 February 2025; Editorial Decision: 18 February 2025; Accepted: 20 February 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Specifically, DistMap (Karaïskos *et al.* 2017) calculates Matthews correlation coefficients (MCCs) between cells and ranks the MCCs to assign spatial positions. Peng *et al.* (2016) used Spearman rank correlation to match cells based on co-expression similarity. Achim *et al.* (2015) developed a scoring system to rank the likelihood of cell correspondences based on expression pattern similarity and distance between putative spatial locations. These basic correlation or similarity-based approaches are prone to noise and often produce ambiguous mappings due to the high dimensionality of single-cell expression data. Additionally, methods like Seurat V3 (Stuart *et al.* 2019), Harmony (Korsunsky *et al.* 2019), LIGER (Welch *et al.* 2019), and SpaGE (Abdelaal *et al.* 2020) utilize dimension reduction techniques to embed scRNA-seq and ST data into a shared low-dimensional space for cell correspondence matching. These approaches face challenges in information loss and interpretability when projecting to a low-dimensional representation.

Recently, SpaOTsc (Cang and Nie 2020) and novoSpaRc (Moriel *et al.* 2021) construct optimal transport (OT) cost matrix using MCCs and K-Nearest Neighbors (KNNs), respectively, to measure cell-cell similarity. These approaches of outputting the mapping matrix enable various spatial associations with scRNA-seq data, facilitating tasks such as gene imputation in imaging-based ST and spot deconvolution in sequence-based ST. Moreover, deep learning models have been applied to learn nonlinear representations for discovering cell mapping between scRNA-seq and ST data. For example, Tangram (Biancalani *et al.* 2021) introduces a pairwise cosine similarity scoring method as an objective function to learn a mapping matrix. Uniport (Cao *et al.* 2022) integrates the OT distance matrix into a latent space of the Variational Autoencoder (Kingma and Welling 2014) to align two datasets but does require significant memory space for the optimization process. Meanwhile, GraphST (Long *et al.* 2023) presents a KNNs graph-based neural network approach incorporating two datasets with a projection matrix. In addition to cell correspondence learning, several methods such as gene imputation and spot deconvolution are designed for one-way information transfer from scRNA-seq data to ST. For example, in the context of imaging-based ST data, gimVI (Lopez *et al.* 2019) is a probabilistic generative model that imputes missing gene values by referencing scRNA-seq data. On the other hand, spot deconvolution methods, such as SPOTlight (Elosua-Bayes *et al.* 2021), spatialDWLS (Dong and Yuan 2021), Cell2location (Kleshcheynikov *et al.* 2022), SpatialDecon (Danaher *et al.* 2022), STRIDE (Sun *et al.* 2022), CARD (Ma and Zhou 2022), SD<sup>2</sup> (Li *et al.* 2022b), RCTD (Cable *et al.* 2022), stVAE (Li *et al.* 2023), and DOT (Rahimi *et al.* 2024), focus on leveraging sequencing-based ST to transfer cell-type annotations from scRNA-seq data to ST.

In this work, we develop a novel mapping approach that captures relationships between ST data and scRNA-seq data for the spatial reconstruction of scRNA-seq profiles. While existing methods such as SpaOTsc, novoSpaRc, Tangram, and GraphST utilize spatial mapping for the reconstruction of scRNA-seq data, they have distinct focuses and limitations. For instance, GraphST and novoSpaRc incorporate positional information and emphasize local associations by optimizing neighbor selection. SpaOTsc also leverages spatial data but assumes cell-to-cell relationships using metrics like MCCs. However, most current approaches focus primarily

on modeling local relationships, either through neighbor selection or predefined cell-to-cell interactions, and fail to automatically learn both local and global relationships through cross-representation learning. This leads to inconsistent spatial mappings. Furthermore, these methods overlook the potential of incorporating scRNA-seq cell-type information, which is crucial for more accurate learning. Cells with similar functions typically display correlated expression patterns, as noted in Harris *et al.* (2021). Therefore, an scRNA-seq cell can often be linked to multiple spatial locations sharing similar expression profiles. Integrating cell-type spatial dependencies is critical for capturing biologically meaningful interactions between cell types across tissues.

To address these limitations, our approach incorporates a contrastive learning framework that includes cell-type contrastive learning for scRNA-seq feature representation and inter-contrastive learning for both scRNA-seq and ST data. Specifically, we propose an adaptable and trainable mapping network that seamlessly integrates a structural similarity regularization loss for network optimization, through which meaningful cellular latent features can be effectively encoded to learn precise cell correspondences mapping. Furthermore, we develop a contrastive learning approach to exploit more meaningful latent feature representations across two modalities. More specifically, the contrastive learning framework comprises cell-type contrastive learning for feature representation learning of scRNA-seq data, and inter-contrastive learning for feature representation learning of both scRNA-seq and ST datasets. By leveraging the available cell-type information of scRNA-seq data and the learned mapping matrix, we aim to achieve spatial awareness and distinguish between similar cell types within the latent feature representations of both modalities. Extensive experiments on three different biological systems with different spatial reconstruction tasks have demonstrated that our method surpasses state-of-the-art methods across various metrics. The learned mapping results regarding biological insights can further verify the accurate localization of biologically meaningful cell types, extend gene spatial patterns of scRNA-seq data, and deconvolve cell-type composition. This approach allows for a more extensive knowledge of cellular interactions and functions, providing valuable insights into the spatial structure of cells in multicellular systems.

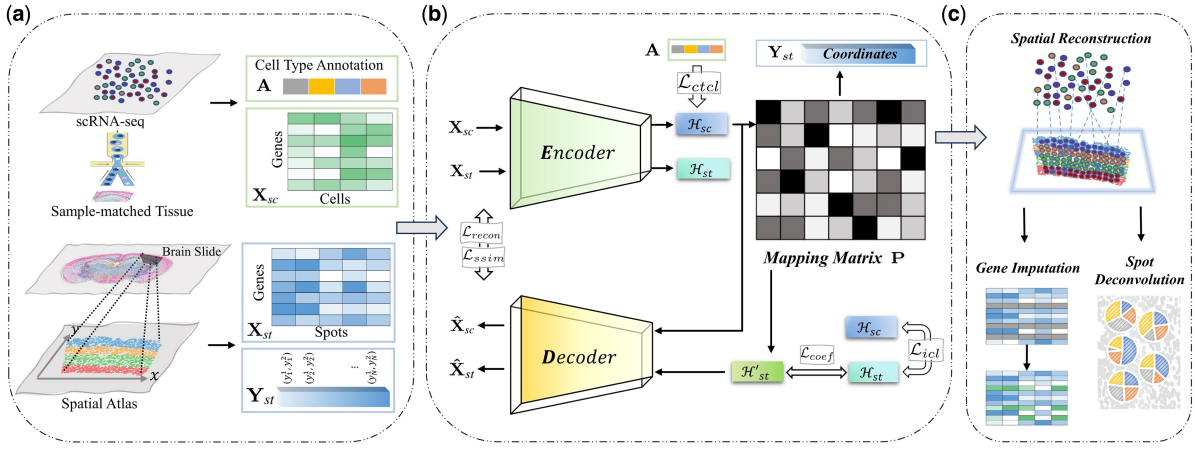
## 2 Materials and methods

In this section, we provide a detailed description of our proposed method, which seeks to spatially reconstruct scRNA-seq data by assigning a single cell to the corresponding coordinates of spatial reference. The pipeline of our approach for spatially mapping scRNA-seq data is illustrated in Fig. 1.

### 2.1 Problem formulation

The input data consists of a ST dataset  $\mathbf{X}_{st} \in \mathbb{R}^{N \times P}$  and a scRNA-seq dataset  $\mathbf{X}_{sc} \in \mathbb{R}^{M \times Q}$ , where  $N$ ,  $P$  are the number of cells and genes for the ST data and  $M$ ,  $Q$  for the scRNA-seq data.

Typically,  $Q \gg P$ , as scRNA-seq can capture a lot of gene expressions without spatial information. In practice, we can align the two modalities based on these shared genes by selecting a subset of  $K$  overlapping genes in the ST and scRNA-seq data. Therefore, by preprocessing these  $K$  overlapping genes, the data can be represented as ST data  $\mathbf{X}_{st} \in$



**Figure 1.** The pipeline of the proposed method. (a) The processing of data. (b) The network architecture of COME. The network takes ST data  $\mathbf{X}_{st}$  and scRNA-seq data  $\mathbf{X}_{sc}$  as input. To encode meaningful cellular features, the input data are encoded and decoded to latent representations ( $\mathcal{H}, \hat{\mathbf{X}}$ ). Concurrently,  $\mathcal{H}_{sc}$  is input to a mapping matrix  $\mathbf{P}$  for learning cell-spot gene expression relationships with the SSIM regularization, producing an improved target matrix. Optimization is applied to  $\mathbf{P}$  via inter-contrastive learning with cell type  $\mathbf{A}$ . By matching  $\mathbf{P}$  to  $\mathbf{Y}_{st}$ , spatial locations are predicted for  $\mathbf{X}_{sc}$ . (c) The outputs  $\mathbf{P}$  of COME can be utilized for gene imputation and spot deconvolution.

$\mathbb{R}^{N \times K}$  and scRNA-seq data  $\mathbf{X}_{sc} \in \mathbb{R}^{M \times K}$ . In this setting,  $\mathbf{Y}_{st} \in \mathbb{R}^{N \times 2}$  is the spatial coordinates matrix of ST. The key to predicting the spatial information  $\mathbf{Y}_{sc}$  for the scRNA-seq data is to learn the cell-to-spot mapping  $\mathbf{P} \in \mathbb{R}^{M \times N}$  between the ST and scRNA-seq data, and then predict  $\mathbf{Y}_{sc}$  with the learned mapping  $\mathbf{P}$ . The mapping probability matrix  $\mathbf{P}$  in our method serves dual purposes: predicting spatial information for scRNA-seq cells and performing cell spot deconvolution in ST data. For spatial prediction of scRNA-seq data, our method assigns spatial information to scRNA-seq cells by leveraging the closest ST data. Specifically, using the probability matrix  $\mathbf{P}$ , we estimate the spatial locations based on ST data, which then acts as the predicted spatial coordinates for corresponding scRNA-seq cells. In addition, for spot deconvolution in ST data analysis, where the goal is to infer cell types or mixtures of cell types at each spot, the probability matrix  $\mathbf{P}$  offers a flexible framework.  $\mathbf{P}$  allows for the prediction of mixed cell types at each spatial location, enabling more accurate modeling of complex cellular compositions in ST data.

## 2.2 Cell correspondence learning

In the proposed model, a shared autoencoder is initially employed to derive latent representations for both the ST data  $\mathbf{X}_{st}$  and the scRNA-seq data  $\mathbf{X}_{sc}$ . To capture cell-spot relationships between modalities, the reconstructed spatial data  $\hat{\mathbf{X}}_{st}$  is obtained by decoding the latent code  $\mathcal{H}_{sc}$  from the scRNA-seq data encoder  $E(\mathbf{X}_{sc})$ . A coefficient layer is also introduced to learn cell-spot matching by mapping from scRNA-seq to the spatial domain. This coefficient matrix  $\mathbf{C} \in \mathbb{R}^{M \times N}$  takes  $\mathcal{H}_{sc}$  as input and predicts corresponding spatial data via  $\mathcal{H}'_{st} = \mathbf{C}^\top \mathcal{H}_{sc}$ . The adaptive matrix  $\mathbf{C}$  captures the raw association strength without relying on predetermined relationships between cells and spots. To better capture coefficient patterns during training, the process can be formulated as follows:

$$\hat{\mathbf{X}}_{st} = D(\mathbf{C}^\top E(\mathbf{X}_{sc})), \quad (1)$$

$$\hat{\mathbf{X}}_{sc} = D(E(\mathbf{X}_{sc})). \quad (2)$$

The corresponding reconstruction loss function can be defined as follows:

$$\mathcal{L}_{recon} = \|\mathbf{X}_{st} - \hat{\mathbf{X}}_{st}\|_2^2 + \|\mathbf{X}_{sc} - \hat{\mathbf{X}}_{sc}\|_2^2, \quad (3)$$

where  $\|\cdot\|_2^2$  calculates mean squared error. We formulate the regularized coefficient loss term as:

$$\mathcal{L}_{coef} = \|\mathcal{H}_{st} - \mathcal{H}'_{st}\|_2^2 + \lambda \|\mathbf{C}\|_2^2, \quad (4)$$

where  $\lambda$  represents the weighting factors used to balance the relative importance of the regularization terms. Norm regularization  $\|\mathbf{C}\|_2^2$  is applied to the mapping matrix to enforce matrix normalization. The mapping matrix  $\mathbf{P}$  is computed by applying softmax (Goodfellow 2016) to the coefficient matrix  $\mathbf{C}$  over columns:

$$\mathbf{P}_{ij} = \frac{\exp(\mathbf{C}_{ij})}{\sum_k \exp(\mathbf{C}_{ik})} \quad (5)$$

This normalization enhances the interpretability of the mapping matrix  $\mathbf{P}$ , where each row of  $\mathbf{P}$  represents the probability of each scRNA-seq cell being located at the corresponding spatial spot.

The coefficient matrix  $\mathbf{C}$  is of full size ( $M \times N$ ), not batch-wise, allowing it to learn cell spatial patterns at the entire tissue level. Consequently, the mapping matrix  $\mathbf{P}$  captures the associations within the latent space and between ST data and scRNA-seq data.

To highlight the significance of preserving structure with the ground truth matrix, we introduce a structural similarity regularization term  $\mathcal{L}_{ssim}$  to measure the consistency of structures between matrices. This loss function is particularly sensitive to spatial pixel variations, making it ideal for assessing gene expression patterns in a global context. Given that the reconstructed spatial gene expression  $\hat{\mathbf{X}}_{st}$  is derived from the latent features  $\mathcal{H}_{sc}$  (i.e.  $\hat{\mathbf{X}}_{st} = D(\mathbf{P}^\top \mathcal{H}_{sc})$ ),  $\mathcal{L}_{ssim}$  extends

optimization beyond the latent space. To achieve this, we construct another reconstruction matrix  $\tilde{\mathbf{X}}_{st} = \mathbf{P}^\top \mathbf{X}_{sc}$  directly from the original gene expression data, and then apply  $\mathcal{L}_{ssim}$  to minimize the structural dissimilarity between the predicted and actual spatial gene expression in the global feature space. The specific form of this structural similarity measure can be expressed as follows:

$$\text{Sim}(\mathbf{X}_{st}, \tilde{\mathbf{X}}_{st}) = \frac{(2\mu_{\tilde{\mathbf{X}}_{st}}\mu_{\mathbf{X}_{st}} + \mathbf{Z}_1)(2\text{Cov}(\mathbf{X}_{st}, \tilde{\mathbf{X}}_{st}) + \mathbf{Z}_2)}{(\mu_{\mathbf{X}_{st}}^2 + \mu_{\tilde{\mathbf{X}}_{st}}^2 + \mathbf{Z}_1)(\sigma_{\mathbf{X}_{st}}^2 + \sigma_{\tilde{\mathbf{X}}_{st}}^2 + \mathbf{Z}_2)}, \quad (6)$$

where  $\text{Cov}(\cdot, \cdot)$  calculates the covariance between two entities.  $\mu$  and  $\sigma$  denote the mean and standard deviation, respectively. The terms  $\mathbf{Z}_1$  and  $\mathbf{Z}_2$  denote constant normalization terms, both derived by multiplying a constant with the maximum value of the input. Specifically, this can be expressed as:  $\mathbf{Z}_1 = a_1 \cdot \max(\mathbf{X}_{st}, \tilde{\mathbf{X}}_{st})$  and  $\mathbf{Z}_2 = a_2 \cdot \max(\mathbf{X}_{st}, \tilde{\mathbf{X}}_{st})$ . (In this context, we empirically set the constant values  $a_1 = 0.01$  and  $a_2 = 0.03$ .) We propose the following structural similarity loss as follows:

$$\mathcal{L}_{ssim} = 1 - \text{Sim}(\mathbf{X}_{st}, \tilde{\mathbf{X}}_{st}) \quad (7)$$

where the structural similarity  $\text{Sim}(\cdot, \cdot)$  ranges from 0 to 1, where higher values denote greater similarity.

### 2.3 Contrastive learning module

Moreover, we aim to develop a contrastive learning approach to uncover more meaningful latent feature representations across two modalities. By leveraging the available cell-type information and the mapping matrix, we seek to achieve spatial awareness and distinguish between similar cell types within the latent feature representations of both modalities.

Specifically, for the scRNA-seq data, cell samples within the same cell type should exhibit similar feature representations. Therefore, we develop a cell-type contrastive learning approach for the scRNA-seq data by leveraging the available cell-type information. In this intra cell-type contrastive learning, cell pairs of the same type are considered positive, whereas those of different types are considered negative. The cell-type labels are defined as  $y$ . The intra-contrastive learning can be formulated as follows:

$$\mathcal{L}_{ctcl} = -\frac{1}{M} \sum_{i=1}^M \log \frac{\sum_{y_i=y_j} f(\mathcal{H}_{sc}^i, \mathcal{H}_{sc}^j)}{\sum_{y_i \neq y_j} f(\mathcal{H}_{sc}^i, \mathcal{H}_{sc}^j)} \quad (8)$$

where  $f(a, b) = e^{\theta(a, b)/\tau}$  and inner product  $\theta(a, b) = a^T b / \|a\| \|b\|$  is used to calculate the pair's similarity, and  $\tau$  is a temperature parameter. This formulation allows us to push cell  $i$  and cell  $j$  closer together when they belong to the same cell type. Moreover, latent feature representations of scRNA-seq and ST are expected to be similar. To achieve this, we develop an inter-contrastive learning approach across the ST and scRNA-seq datasets by leveraging the learned mapping matrix  $\mathbf{P}$ . Specifically, for the scRNA-seq feature representation  $\mathcal{H}_{sc}^i$ , the associated positive sample in the ST data is  $\mathcal{H}_{st}^i$ , which has the maximum probability connection of two samples in the learned mapping matrix  $\mathbf{P}$ . Thus, the inter-contrastive loss function can be defined as:

$$\mathcal{L}_{icl} = -\frac{1}{M} \sum_{i=1}^M \log \frac{f(\mathcal{H}_{sc}^i, \mathcal{H}_{st}^i)}{\sum_{k \neq j} f(\mathcal{H}_{sc}^i, \mathcal{H}_{st}^k)} \quad (9)$$

Overall, the unified contrastive loss function can be expressed as:

$$\mathcal{L}_{cst} = \mathcal{L}_{icl} + \mathcal{L}_{ctcl} \quad (10)$$

We utilize the discrete version of the loss to optimize the network. By incorporating the aforementioned regularization methods, the optimization formulation is expressed as follows:

$$\min_{\theta_E, \theta_D, \mathbf{C}} \mathcal{L}_{recon} + \alpha_1 \mathcal{L}_{coef} + \alpha_2 \mathcal{L}_{ssim} + \alpha_3 \mathcal{L}_{cst} \quad (11)$$

where  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are weighting factors for the loss terms. The term  $\mathcal{L}_{recon}$  ensures that the autoencoder extracts informative features from the input gene expression. Due to the global structure constraint term  $\mathcal{L}_{ssim}$ , we can ensure that the variation between genes across the entire tissue is accounted for. In addition, the term  $\mathcal{L}_{coef}$  is useful for modeling the cell spatial characteristic across cell modalities, and the term  $\mathcal{L}_{cst}$  facilitates accurate correspondences between cells. All losses are optimized for encoder  $E$ , decoder  $D$ , and the mapping matrix  $\mathbf{P}$ .

## 3 Results

### 3.1 Datasets

To assess the accuracy and generalizability of our proposed mapping framework, we select spatial atlas and scRNA-seq datasets from three distinct biological systems: drosophila embryo (DE), mouse primary visual cortex (VISp), and human cancerous pancreas. As for the DE, the scRNA-seq data are from the GEO database with accession number GSE95025 (Karaikos et al. 2017), and the spatial reference utilized in the experiments is acquired from the Berkeley Drosophila Transcription Network Project (BDNTP) (Larkin et al. 2021). In the case of the mouse VISp, the corresponding Smart-Seq dataset (Tasic et al. 2018) was used. The corresponding spatial atlas is obtained from various ST technologies, MERFISH, smFISH, STARmap, seqFISH+, and 10x Visium. seqFISH+ data are primarily used for generating synthetic data (Li et al. 2022a) to assist in the quantitative evaluation of spot deconvolution. Data download instructions are included in [Supplementary.1](#) Data availability, available as [supplementary data](#) at [Bioinformatics](#) online. For the human biological system analysis, pancreatic ductal adenocarcinoma (PDAC) tissues, with a diameter extending to 100  $\mu\text{m}$ , were selected for this study. Cell-type deconvolution were

**Table 1.** The detailed information of benchmarking datasets.<sup>a</sup>

| BS                | Dataset       | Protocols        | Cells/Spots | Genes  |
|-------------------|---------------|------------------|-------------|--------|
| Drosophila Embryo | Spatial Atlas | FISH             | 3039        | 84     |
|                   | scRNA-seq     | Drop-seq         | 1297        | 8924   |
| Mouse VISp        | Spatial Atlas | STARmap          | 1549        | 1020   |
|                   |               | MERFISH          | 2399        | 254    |
|                   |               | smFISH           | 2340        | 22     |
|                   |               | seqFISH+         | 524         | 10,001 |
|                   |               | 10X Visium       | 2695        | 32,285 |
|                   |               | scRNA-seq        | Smart-seq   | 15,413 |
| PDAC              | Spatial Atlas | microarray-based | 428         | 19,736 |
|                   | scRNA-seq     | inDrop           | 1926        | 19,736 |

<sup>a</sup> "BS" denotes the biological system.

performed on 428 spatially distinct spots correlated with 1926 individual cells, encompassing the expression of 19,736 genes (Moncada *et al.* 2020). For a detailed dataset overview, please refer to Table 1.

We adhere to standardized pre-processing procedures. Specifically, the imaging-based ST data was normalized to ensure uniform total cell counts. The scRNA-seq and sequencing-based ST data were normalized based on the total counts across all genes within each cell. They were then scaled by  $10^4$  and subjected to log transformation with the addition of a pseudo-count.

### 3.2 Evaluation

After the mapping is completed, annotations can be transferred either from ST data to scRNA-seq data or vice versa. To assess the accuracy of the matching algorithm, we calculate the similarity and discrepancy between the predicted spatial ST data and the actual ST data. The evaluation metrics include the Pearson Correlation Coefficient (PCC) for measuring correlation, Structural Similarity Index (SSIM) for evaluating structural similarity, and Root Mean Square Error (RMSE) and Jensen–Shannon Divergence (JSD) for quantifying the difference between predicted and actual data. Superior matching performance is indicated by higher PCC and SSIM values, coupled with lower RMSE and JSD scores. We used a k-fold cross-validation strategy to evaluate gene expression reconstruction. For spot deconvolution and cell-type localization, we derive the distribution matrix per location. More details can be found in Supplementary.2 Validation, available as [supplementary data](#) at *Bioinformatics* online.

### 3.3 Competing methods

We compare our proposed method against several state-of-the-art cell mapping methods: Tangram (Biancalani *et al.* 2021), novoSpaRc (Moriel *et al.* 2021), SpaOTsc (Cang and Nie 2020), and GraphST (Long *et al.* 2023). Tangram, novoSpaRc, and SpaOTsc were chosen due to their widespread adoption, while GraphST represents a recently proposed deep-learning approach using graph neural networks.

### 3.4 Spatial gene reconstruction

Due to the high-throughput nature of scRNA-seq, we can detect more gene expressions at the single-cell level. Mapping scRNA-seq data onto atlases helps us to understand the spatial patterns of the extended genes, which are not on the atlas. To evaluate the effectiveness of the mapping results in predicting spatial gene expression, we first applied COME to the spatial reconstruction of DE data. Initially, we illustrated the distribution of the PCC between each gene’s actual and predicted spatial expression profiles. For each comparative analysis, one-sided paired Wilcoxon signed-rank tests are performed to ascertain whether COME yields significantly higher PCC values than baseline methods. The results in Fig. 2a demonstrate that COME significantly surpasses the performance of the leading alternative, SpaOTsc, and markedly enhances the median PCC relative to other evaluated methods (with  $P$ -value  $< .05$ ). In addition, COME consistently and significantly outperformed the PCC, SSIM, and RMSE compared to other methods on DE data. From Supplementary Fig. S1, available as [supplementary data](#) at *Bioinformatics* online, it can be observed that the coefficients learned by COME align with biological expectations. Furthermore, from Supplementary Fig. S2, available as

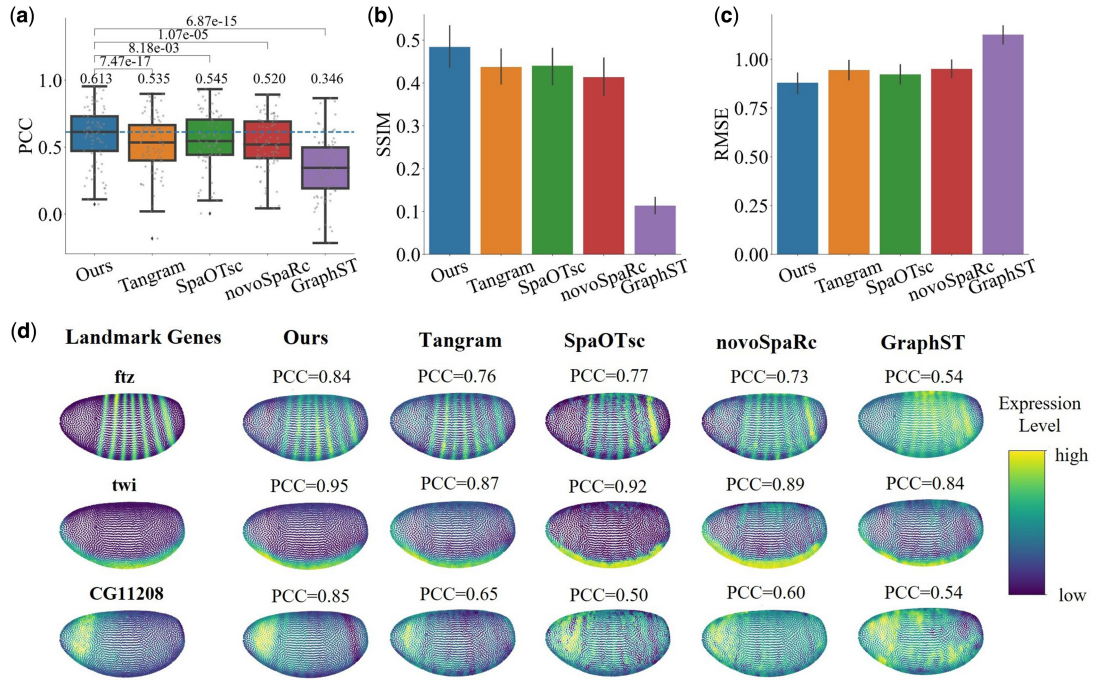
[supplementary data](#) at *Bioinformatics* online, it can be seen that COME helps to remove batch effects between ST and scRNA-seq.

We further select three typical spatial landmark genes (Karaikos *et al.* 2017, Miller *et al.* 2021) from DE and visually compare the predicted spatial patterns with their respective ground truth and other methods (as shown in Fig. 2d). A higher PCC signifies a more successful reconstruction. The results demonstrate the successful reconstruction of some well-characterized genes to show that our reconstruction method is feasible. More specifically, in drosophila embryos, the expression of the gene *twi* is typically observed in the ventral cells during the blastoderm stage (Ashraf *et al.* 1999). We observe that gene *ftz* reconstruction exhibits distinct spatial features, with only six stripes distributed within the tissue. Gene *CG11208* reconstruction is only shown on the drosophila’s dorsal, and the PCC is significantly improved when compared with the results of other methods. These three representative genes have clearer outlines than other methods and are more consistent with the ground truth.

### 3.5 Evaluation on cellular resolution spatial transcriptomic data

Here, we apply our method to infer genes and localize cell types on a large scale for both cellular and non-cellular resolution ST data of mouse VISp generated from Smart-seq. As shown in Fig. 3a–c, the gene prediction results indicate that COME surpasses all competing methods in terms of PCC across all datasets. Notably, on the STARmap dataset, the performance of COME is particularly remarkable, with a median PCC value of 0.233. This represents a significant improvement of 12% compared to the second-best Tangram method. It is worth noting that the available cell gene counts for MERFISH, and STARmap datasets are denoted as 254 and 996 each. The network trained on these datasets, focusing on gene expression in cells with higher complexity, exhibits a clear advantage. This suggests that our proposed method excels at capturing intricate correlations in gene expression within complex cells, likely due to more accurate modeling of intercellular information.

Since most atlases have coarser cell-type resolution than the scRNA-seq data, our accurate mapping can potentially lead to a better understanding of how cell types are spatially distributed within the tissue. The cerebral cortex is mainly divided into three major categories: glutamatergic neurons, GABAergic neurons, and non-neuronal cells. Among them, glutamatergic neurons are cells with distinct spatial features, meaning that the distribution of glutamatergic neuron cells, such as “L2/3 IT”, “L4”, “L5 IT”, “L5 PT”, “L6 IT”, “L6 CT”, and “L6 b” in the tissue should exhibit clear stratification. Our mapping results (Fig. 3d–f, and Supplementary Fig. S3, available as [supplementary data](#) at *Bioinformatics* online) show that scRNA-seq data mapped to different atlases show spatially layered cells in glutamatergic neurons. In the mouse VISp, different genes exhibit higher expression in different layers (Alon *et al.* 2021, Cheng *et al.* 2022, Mao and Staiger 2024). For the image-based ST data, the visualization results in Supplementary Fig. S4, available as [supplementary data](#) at *Bioinformatics* online, of marker gene and gene reconstruction are explained for their consistency. According to Supplementary Fig. S4, available as [supplementary data](#) at *Bioinformatics* online, we can see that the spatial pattern of genes predicted by smFISH based on the mapping method is consistent with the marker gene atlas of



**Figure 2.** The performance of spatial gene reconstruction of single-cell DE data. (a) PCC of different methods. The Wilcoxon signed-rank test and the median of PCC are reported. (b) SSIM of different methods. (c) RMSE of different methods. The bar chart represents the mean values of the data, with error bars indicating the  $\pm 95\%$  confidence intervals. (d) Predicted expression of test genes. The reconstruction results show the genetic patterning of the reconstructed scRNA-seq of different methods and their corresponding atlas (ground truth). The highlighted portions represent genes with high expression and vice versa.

MERFISH. For the Visium data, our method successfully reconstructs clear stratification from layer 2 to layer 6, as shown in Fig. 3f, consistent with previous studies (Wei et al. 2022, Wan et al. 2023). Our predicted results are also closer to the ground truth than other methods (Fig. 3e and f), suggesting that the predicted marker gene expression more strongly supports cell-type proportion inferred by COME. It should be highlighted that the spatial distribution of the subtypes within layer 6 is difficult to distinguish in a stratified manner. For instance, the characteristic gene of “L6 b”, *Tmem163*, unlike genes localized exclusively to particular areas, is distributed more diffusely across various layers yet predominantly accumulates at the top edges of layer 6. This distribution of “L6 b” deduced from COME aligns closely with analyses (Mao and Staiger 2024). Similarly, since IT neurons require multiple marker genes to ascertain the cell type (Fischer and Gillis 2021), the more accurate spatial distribution of scRNA-seq genes predicted using COME subsequently aids in inferring the distribution of IT neurons, which primarily within the more medial regions of the layer (Zhang et al. 2021), the comparison between the predicted results for “L5 IT” and “L5 PT” and those from Tangram can also be observed in Supplementary Fig. S5, available as supplementary data at [Bioinformatics](#) online.

Overall, our mapping method indicates that COME can recover spatial patterns for genes absent from the ST data and can further help understand the spatial distribution of cell types in tissues.

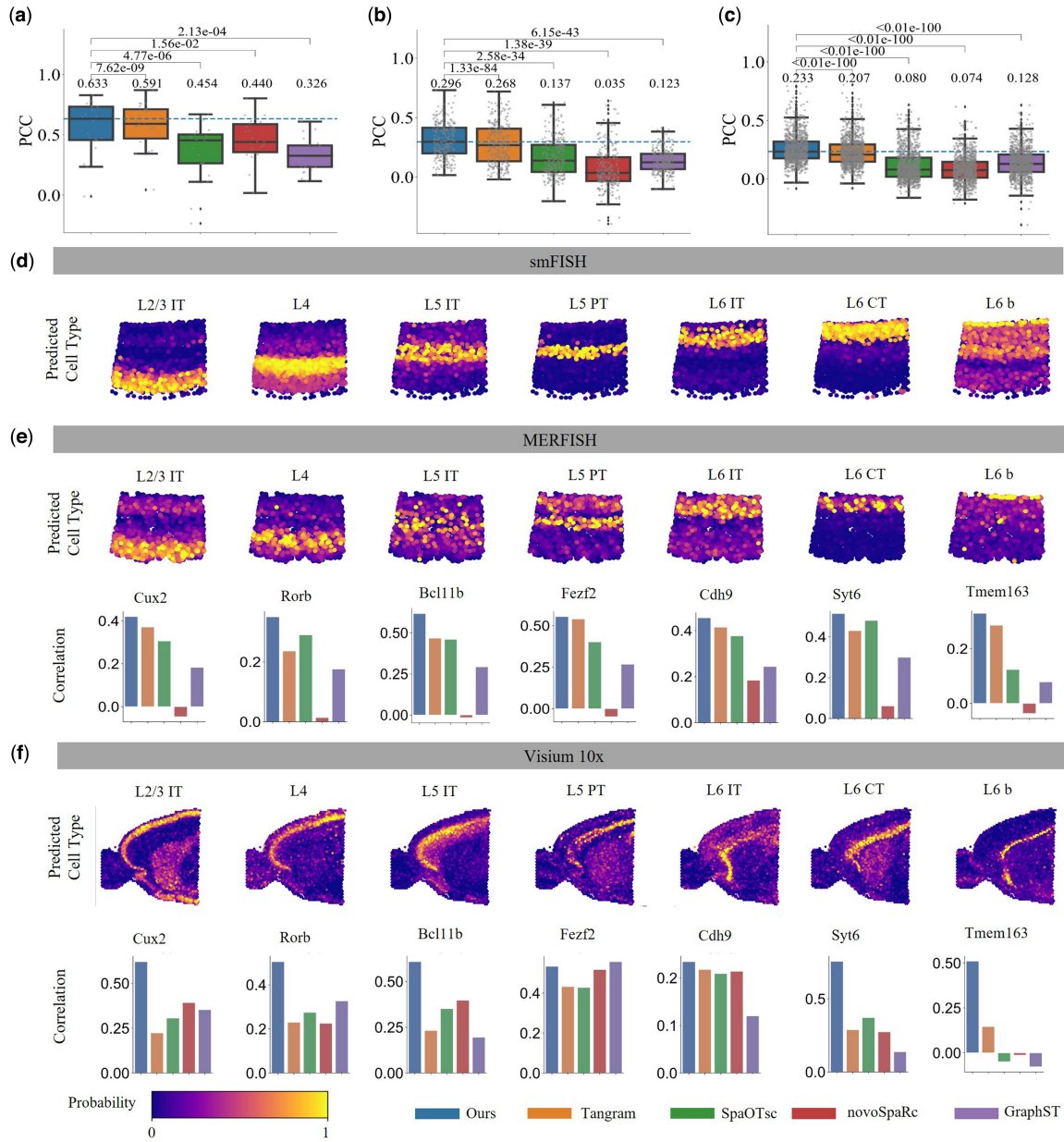
### 3.6 Spatial deconvolution

To validate the performance of COME, we use synthetic seqFISH+ ST dataset for quantitative analysis of the mapping results, and also conducted qualitative analysis using a real-world dataset (PDAC). From Supplementary Fig. S8, available as supplementary data at [Bioinformatics](#) online,

compared to the second-best mapping method, SpaOTsc, our method achieves better results in SSIM and JSD, while maintaining relatively small variances in PCC and RMSE, indicating robust performance. Notably, the incorporation of prior knowledge further enhances the effectiveness of our method (COME+), allowing for competitive performance in downstream tasks, illustrating its versatility and adaptability in various applications.

We utilized the mapping estimated by COME to perform spot deconvolution across 15 cell types in actual human PDAC ST data. As shown in Fig. 4a, the result demonstrates the distinction between cancerous and non-cancerous areas throughout the region: tumor cell types are predominantly in cancerous areas (Orange). In contrast, normal pancreatic cell types are primarily found in non-tumor areas (Blue), consistent with previous studies’ results (Moncada et al. 2020). In comparison, other methods, such as Tangram and GraphST, do not delineate a separation between cancerous and non-cancerous areas (see Supplementary Fig. S6, available as supplementary data at [Bioinformatics](#) online). Additionally, COME can distinguish between ductal epithelium (light blue) and normal pancreatic tissue (dark blue), a differentiation that other mapping methods like SpaOTsc and novoSpaRc struggle to make (see Supplementary Fig. S6, available as supplementary data at [Bioinformatics](#) online).

Moreover, COME can accurately predict the four main cell-type positions of essential tumor microenvironment (TME) constituents, including ductal and acinar cells and cancer clones A and B, as shown in Fig. 4b. Cancer cell clones A and B are localized in two subregions of the cancerous area, with clone A predominantly occupying the top subregion and clone B predominantly distributed in the lower subregion. These findings align with the actual spatial profiles of the marker genes *TM4SF1* and *S000A4*. Their marker genes



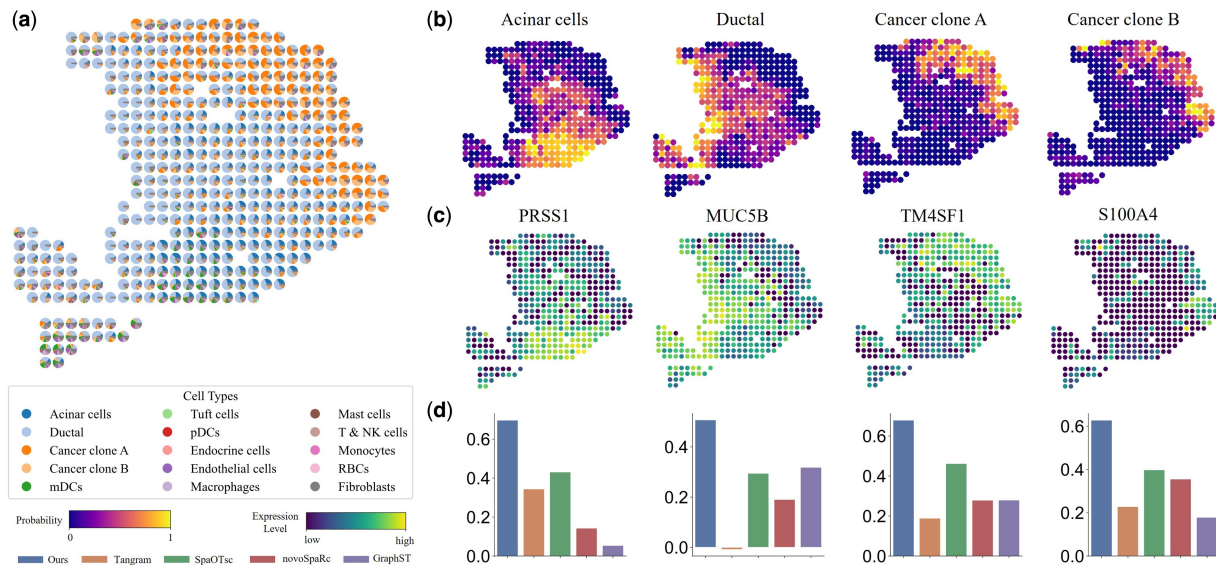
**Figure 3.** Performance of COME on the mouse VISp dataset from Smart-seq. (a) PCC of predicting gene spatial patterns on smFISH compared with different methods. (b) PCC of predicting gene spatial patterns on MERFISH compared with different methods. (c) PCC of predicting gene spatial patterns on STARmap compared with different methods. (d) Probabilistic spatial localization of scRNA-seq data on smFISH reference. (e) Probabilistic spatial localization of scRNA-seq data on MERFISH reference. In the first row, the inferred probabilities of the seven glutamatergic neuron types by COME are displayed for each spatial spot. The next row presents a comparative analysis of the PCC between the predicted expression levels and marker genes across different methods. (f) Probabilistic spatial localization of scRNA-seq data on 10x Visium reference. In the first row, the probabilities of the seven glutamatergic neuron types inferred by COME are displayed on each spot. The next row presents a comparative analysis of the PCC between the predicted expression levels and marker genes across different methods.

indicate accuracy in Fig. 4c and d demonstrates that COME outperforms other methods in the consistency of predicting marker genes by the PCC. Similarly, the cell subtype predictions (as shown in Supplementary Fig. S7, available as supplementary data at [Bioinformatics](#) online) for the TME in PDAC also conform to the previous study (Moncada *et al.* 2020).

## 4 Discussion

The essence of recovering the inherent spatial properties within scRNA-seq datasets lies in learning the cell-to-spot

relationships between scRNA-seq and different spatial transcriptomics data. In this study, we introduce a mapping method for the spatial reconstruction of scRNA-seq data to facilitate knowledge transfer. Our approach integrates mapping with structural similarity regularization to leverage cell-spot dependencies. Additionally, we incorporate a contrastive learning strategy to enhance spatial awareness and distinguish between similar cell types within the latent feature representations of both modalities. The experimental results demonstrate that the proposed method accurately infers spatial information for scRNA-seq through the learned cell-to-spot mapping. More importantly, we provide biological



**Figure 4.** Spatial deconvolution result of PDAC. (a) Visualization of the spot deconvolution results. A spatial scatter pie chart is presented, which depicts the cell-type compositions as predicted by the COME. Each individual scatter point on the plot represents a distinct spatial spot within the ST data. (b) Predicted distribution of typical clusters in PDAC. (c) Expression profiles of marker genes corresponding to the cell clusters depicted in (b). (d) The comparison of the PCC between the inferred expression levels and marker genes with other methods.

insights to verify the effectiveness and practical significance of our proposed approach.

## Acknowledgements

We thank BioRender (biorender.com) for providing technical support on drawing figures.

## Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest: None declared.

## Funding

This work was supported in part by the Research Grants Council of the Hong Kong Special Administration Region (Project No. CityU 11206622), in part by Guangdong Basic and Applied Basic Research Foundation (Project No. 2023A1515030154), in part by the Fundamental Research Funds for the Central Universities (Project No. 2024ZYGXZR077), in part by the Guangdong Basic and Applied Basic Research Foundation (Project No. 2024A1515011437), and in part by TCL Science and Technology Innovation Fund (Project No. 20231752).

## Data availability

The data underlying this article are available in Zenodo Repository, at [10.5281/zenodo.14695533](https://doi.org/10.5281/zenodo.14695533).

## References

Abdelaal T, Mourragui S, Mahfouz A *et al.* SpaGE: spatial gene enhancement using scRNA-seq. *Nucleic Acids Res* 2020;48:e107.  
Achim K, Pettit J-B, Saraiva LR *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat Biotechnol* 2015;33:503–9.

Alon S, Goodwin DR, Sinha A *et al.*; IMAXT Consortium. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* 2021;371:eaax2656.  
Ashraf SI, Hu X, Roote J *et al.* The mesoderm determinant snail collaborates with related zinc-finger proteins to control drosophila neurogenesis. *Embo J* 1999;18:6426–38.  
Biancalani T, Scalia G, Buffoni L *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nat Methods* 2021;18:1352–62.  
Cable DM, Murray E, Zou LS *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat Biotechnol* 2022;40:517–26.  
Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat Commun* 2020;11:2084.  
Cao K, Gong Q, Hong Y *et al.* A unified computational framework for single-cell data integration with optimal transport. *Nat Commun* 2022;13:7419.  
Chen KH, Boettiger AN, Moffitt JR *et al.* Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;348:aaa6090.  
Cheng S, Butrus S, Tan L *et al.* Vision-dependent specification of cell types and function in the developing cortex. *Cell* 2022;185:311–27.e24.  
Codeluppi S, Borm LE, Zeisel A *et al.* Spatial organization of the somatosensory cortex revealed by osmfish. *Nat Methods* 2018;15:932–5.  
Danaher P, Kim Y, Nelson B *et al.* Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. *Nat Commun* 2022;13:385.  
Dong R, Yuan G-C. Spatialdws: accurate deconvolution of spatial transcriptomic data. *Genome Biol* 2021;22:145.  
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:e50.  
Eng C-HL, Lawson M, Zhu Q *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019;568:235–9.  
Fischer S, Gillis J. How many markers are needed to robustly determine a cell's type? *iScience* 2021;24:103292.  
Goodfellow I. *Deep Learning*. Cambridge: MIT Press; 2016.  
Harris BD, Crow M, Fischer S *et al.* Single-cell co-expression analysis reveals that transcriptional modules are shared across cell types in the brain. *Cell Syst* 2021;12:748–56.e3.

- Karaiskos N, Wahle P, Alles J *et al.* The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 2017;358:194–9.
- Kingma DP, Welling M. Auto-encoding variational bayes. In: *Proceedings of International Conference on Learning Representations (ICLR)*, 2014. <http://arxiv.org/abs/1312.6114>
- Kleshchevnikov V, Shmatko A, Dann E *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat Biotechnol* 2022;40:661–71.
- Korsunsky I, Millard N, Fan J *et al.* Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019;16:1289–96.
- Larkin A, Marygold SJ, Antonazzo G *et al.*; FlyBase Consortium. Flybase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Research* 2021;49:D899–D907.
- Li B, Zhang W, Guo C *et al.* Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods* 2022a;19:662–70.
- Li C, Chan T-F, Yang C *et al.* Stvae deconvolves cell-type composition in large-scale cellular resolution spatial transcriptomics. *Bioinformatics* 2023;39:btad642.
- Li H, Li H, Zhou J *et al.* Sd2: spatially resolved transcriptomics deconvolution through integration of dropout and spatial information. *Bioinformatics* 2022b;38:4878–84.
- Long Y, Ang KS, Li M *et al.* Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nat Commun* 2023;14:1155.
- Longo SK, Guo MG, Ji AL *et al.* Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat Rev Genet* 2021;22:627–44.
- Lopez R, Nazaret A, Langevin M *et al.* A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. arXiv preprint, arXiv:1905.02269, 2019.
- Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat Biotechnol* 2022;40:1349–59.
- Mao X, Staiger JF. Multimodal cortical neuronal cell type classification. *Pflügers Arch* 2024;476:721–33.
- Miller BF, Bambah-Mukku D, Dulac C *et al.* Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Res* 2021;31:1843–55.
- 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;38:333–42.
- Moriel N, Senel E, Friedman N *et al.* Novosparc: flexible spatial reconstruction of single-cell gene expression with optimal transport. *Nat Protoc* 2021;16:4177–200.
- Peng G, Suo S, Chen J *et al.* Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev Cell* 2016;36:681–97.
- Rahimi A, Vale-Silva LA, Fälth Savitski M *et al.* Dot: A flexible multi-objective optimization framework for transferring features across single-cell and spatial omics. *Nat Commun* 2024;15:4994.
- Stuart T, Butler A, Hoffman P *et al.* Comprehensive integration of single-cell data. *Cell* 2019;177:1888–902.e21.
- Sun D, Liu Z, Li T *et al.* Stride: accurately decomposing and integrating spatial transcriptomics using single-cell RNA sequencing. *Nucleic Acids Res* 2022;50:e42.
- Tasic B, Yao Z, Graybiel LT *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 2018;563:72–8.
- Wan X, Xiao J, Tam SST *et al.* Integrating spatial and single-cell transcriptomics data using deep generative models with spatialscope. *Nat Commun* 2023;14:7848.
- Wang X, Allen WE, Wright MA *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;361:eaat5691.
- Wei R, He S, Bai S *et al.* Spatial charting of single-cell transcriptomes in tissues. *Nat Biotechnol* 2022;40:1190–9.
- Welch JD, Kozareva V, Ferreira A *et al.* Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* 2019;177:1873–87.e17.
- Williams CG, Lee HJ, Asatsuma T *et al.* An introduction to spatial transcriptomics for biomedical research. *Genome Med* 2022;14:68.
- Zhang M, Eichhorn SW, Zingg B *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by merfish. *Nature* 2021;598:137–43.