Supplement: Alignment and Integration of Spatial Transcriptomics Data

Ron Zeira[1], Max Land[1], Alexander Strzalkowski[1], and Benjamin J. Raphael[1]

[1]Department of Computer Science, Princeton University, Princeton NJ 08544, USA

S1 Supplementary methods

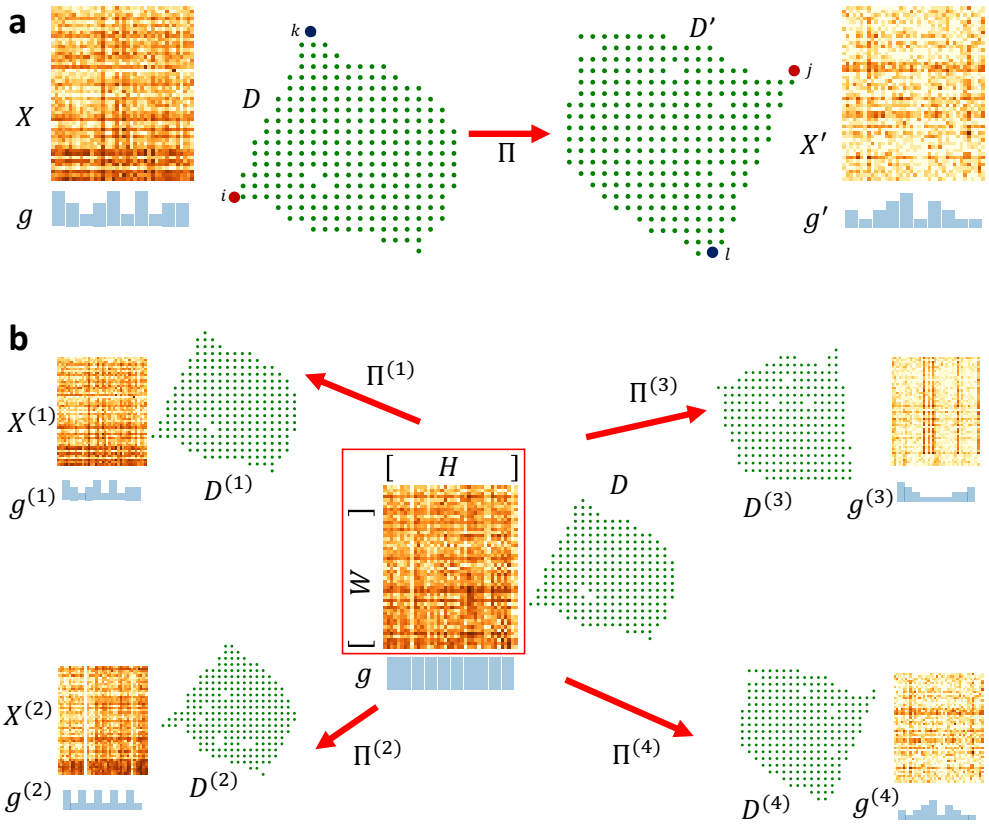


Figure S1: (a) The PAIRWISE SLICE ALIGNMENT PROBLEM aims to find a mapping $\Pi = [\pi_{ij}]$ between spots in one slice and spots in another slice while preserving the gene expression and the spatial distances of mapped spots. (b) The CENTER SLICE INTEGRATION PROBLEM aims to infer a "center" slice consisting of a low rank expression matrix $X = WH$ and a collection $\Pi^{(1)}, \ldots, \Pi^{(t)}$ of mappings from the spots of the center slice to the spots of each input slice.

S1.1 Finding optimal rotation for spatial coordinates

In this section, we seek to find a rotation and translation that of the spatial coordinates of one slice that minimizes the distances to the spatial coordinates of the other slice given a mapping. The problem of finding rotation and translation that minimizes the distances between matched set of points is a well know problem in several research fields [15, 8]. In 2d the problem is often called called Procrustes analysis, a more general linear algebra problem is called the Orthogonal Procrustes problem, and the vector weighted version is called Wahba's problem [15]. In chem-

istry/biology the solution to the 3d problem is called the Kabsch algorithm [8]. The 2d solution is based on finding the rotation angle while the general case (which also works in 2d) looks for a rotation matrix, thus it also supports reflection.

Our problem is a variation of this problem since we have a probabilistic alignment between the spots given by the mapping Π.

Problem S1. *Given ST slices with spatial coordinates $Z \in \mathbb{R}^{2\times n}$ and $W \in \mathbb{R}^{2\times n'}$ and a mapping $\Pi \in \Gamma(g, g')$, find a vector $t \in \mathbb{R}^2$ and a rotation matrix $R \in \mathbb{R}^{2\times 2}$:*

$$Q(t, R) = \sum_{i,j} \pi_{ij} \left\| z_{\cdot i} - R w_{\cdot j} - t \right\|^2. \tag{S1}$$

We first show that we can assume that no translation is needed ($t = 0$) by centering the spatial coordinates Z and W. Assuming R is fixed, we can find the optimal translation by taking the derivative of Q w.r.t. t and comparing to zero:

$$\begin{aligned}
\frac{\partial Q}{\partial t} &= -2 \sum_{i,j} \pi_{ij} (z_{\cdot i} - R w_{\cdot j} - t) \\
&= -2 \sum_i z_{\cdot i} \sum_j \pi_{ij} + 2 \sum_j w_{\cdot j} \sum_i \pi_{ij} + 2t \sum_{i,j} \pi_{ij} \\
&= -2 \sum_i z_{\cdot i} g_i + 2 \sum_j w_{\cdot j} g'_j + 2t = 0
\end{aligned}$$

We have $\hat{t} = Zg - Wg'$. By replacing the spatial coordinates $z_{\cdot i}$ with $z_{\cdot i} - Zg$ and the spatial coordinates $w_{\cdot j}$ with $w_{\cdot j} - Wg'$ we get $Q = \sum_{i,j} \pi_{ij} \left\| z_{\cdot i} - R w_{\cdot j} \right\|^2$. Therefore, centering both spatial coordinates removes the need to find a translation and we are only left with finding the optimal rotation.

We rewrite the objective Q in matrix notation:

$$\begin{aligned}
Q &= \sum_{i,j} \pi_{ij} \left\| z_{\cdot i} - R w_{\cdot j} \right\|^2 \\
&= \sum_{i,j} \pi_{ij} (z_{\cdot i}^T z_{\cdot i} + w_{\cdot j}^T R^T R w_{\cdot j} - z_{\cdot i}^T R w_{\cdot j} - w_{\cdot j}^T R^T z_{\cdot i}) \\
&= -2 \sum_{i,j} \pi_{ij} (z_{\cdot i}^T R w_{\cdot j}) + \alpha \\
&= -2 \operatorname{Tr}(Z^T R W \Pi^T) + \alpha \\
&= -2 \operatorname{Tr}(R W \Pi^T Z^T) + \alpha
\end{aligned}$$

where α us a constant independent of R.

We find the optimal rotation R that minimizes Q using SVD similar to the solution to Wahbs's problem [11]. Let $U \Sigma V^T$ be the SVD decomposition of $W \Pi^T Z^T$. Form the cyclic property of the

trace operator, we have that

$$\begin{aligned} Q &= -2\,\mathrm{Tr}(RU\Sigma V^T) + \alpha \\ &= -2\,\mathrm{Tr}(\Sigma V^T RU) + \alpha \end{aligned}$$

Notice that Σ is a positive diagonal matrix and since V, R and U are orthonormal, $V^T RU$ is also an orthonormal matrix. Therefore, the objective Q is minimized when the trace of $V^T RU$ is maximal which is attained when $V^T RU = I$. We have $R = VU^T$. We note that R may also do reflection in addition to rotation when $\det(R) = -1$.

An alternative derivation for the 2d case is done similar to Procrustes analysis. We write the rotation matrix as a function of the rotation angle θ:

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

Taking the derivative of Q with respect to θ and comparing to zero gives:

$$\begin{aligned} \frac{\partial Q}{\partial \theta} &= -2\,\mathrm{Tr}\Big(\frac{\partial R(\theta)}{\partial \theta} W\Pi^T Z^T\Big) \\ &= -2\,\mathrm{Tr}\Big(\begin{pmatrix} -\sin(\theta) & -\cos(\theta) \\ \cos(\theta) & -\sin(\theta) \end{pmatrix} W\Pi^T Z^T\Big) = 0 \end{aligned}$$

Dividing by $\cos(\theta)$ and extracting θ we have:

$$\hat{\theta} = \arctan\Big(\frac{\mathrm{Tr}\Big(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} W\Pi^T Z^T\Big)}{\mathrm{Tr}(W\Pi^T Z^T)}\Big)$$

Note that with this derivation, we allow a rotation of the coordinates but not reflection.

S1.2 Block Coordinate Descent algorithm for CENTER SLICE INTEGRATION PROBLEM

To solve the CENTER SLICE INTEGRATION PROBLEM, we propose a Block Coordinate Descent algorithm (Algorithm 1):

Algorithm 1: Block Coordinate Descent algorithm for CENTER SLICE INTEGRATION PROBLEM

Initialize W, H and set $r \leftarrow \infty$
repeat
  $r' \leftarrow r$
  for $q \leftarrow 1$ **to** t **do**
    Solve PAIRWISE SLICE ALIGNMENT PROBLEM:
    $\Pi^{(q)} \leftarrow \underset{\Pi \in \Gamma(g, g^{(q)})}{\text{argmin}} \; F(\Pi \,;\, WH, D, X^{(q)}, D^{(q)}, c, \alpha)$
  end
  Solve CENTER MAPPING NMF PROBLEM: $W, H \leftarrow \underset{W \geq 0, H \geq 0}{\text{argmin}} \sum_q \lambda_q \sum_{i,j} c(WH_{\cdot i}, x^{(q)}_{\cdot j}) \pi^{(q)}_{ij}$
  $r \leftarrow R(W, H, \Pi^{(1)}, \dots \Pi^{(t)})$
until $r' - r < \epsilon$;

The problem of finding the optimal W and H given the current mappings $\Pi^{(1)}, \dots, \Pi^{(t)}$ reduces to a new problem we call the CENTER MAPPING NMF PROBLEM:

CENTER MAPPING NMF PROBLEM. *Given t expression matrices $X^{(1)} \in \mathbb{R}_+^{p \times n_1}, \dots, X^{(t)} \in \mathbb{R}_+^{p \times n_t}$, t mapping matrices $\Pi^{(1)} \in \Gamma(g, g^{(1)}), \dots \Pi^{(t)} \in \Gamma(g, g^{(t)})$, an expression cost function c, a distribution $\lambda \in \mathbb{R}_+^t$ and parameters $0 \leq \alpha \leq 1$, $m \in \mathbb{N}$ find two low rank matrices $W \in \mathbb{R}_+^{p \times m}$ and $H \in \mathbb{R}_+^{m \times n}$ such that $X = WH$ minimizing the following objective:*

$$S(W, H) = \sum_q \lambda_q \sum_{i,j} c(WH_{\cdot i}, x^{(q)}_{\cdot j}) \pi^{(q)}_{ij} \tag{S2}$$

Solving the CENTER MAPPING NMF PROBLEM is motivated by finding a low rank expression matrix $X = WH$ that maximizes the likelihood of the following generative model when $g = \frac{1}{n}\mathbf{1}_n$ and $\lambda = \frac{1}{t}\mathbf{1}_t$:

- The random variables of the number of transcripts $x^{(q)}_{lj}$ of a gene l in spot j in slice q are independent given $WH, \Pi^{(1)}, \dots, \Pi^{(t)}$.
- The number of transcripts $x^{(q)}_{lj}$ of a gene l in spot j in slice q given that it was generated from spot i in the consensus slice has a distribution $x^{(q)}_{lj} | WH, \pi^{(q)}_{ij} \sim \text{Poisson}(n \pi^{(q)}_{ij} [WH]_{li})$.
- Therefore, the total number $x^{(q)}_{lj}$ of transcripts of a gene l in spot j in slice q is $x^{(q)}_{lj} | WH, \Pi^{(q)} \sim \text{Poisson}(n \sum_i \pi^{(q)}_{ij} [WH]_{li})$.

The negative log likelihood of the model is

$$
\begin{aligned}
-\log(\Pr(X^{(1)},\dots,X^{(t)}\ ;\ W,H,\Pi^{(1)},\dots,\Pi^{(t)})) &= \sum_q \sum_j \sum_l \Big[\sum_i n[WH]_{li}\pi_{ij}^{(q)} - x_{lj}^{(q)} \log(n \sum_i \pi_{ij}^{(q)}[WH]_{li}) \Big] + \\
&\le \sum_q \sum_j \sum_l \Big[\sum_i n[WH]_{li}\pi_{ij}^{(q)} - \sum_i n\pi_{ij}^{(q)} x_{lj}^{(q)} \log([WH]_{li}) \Big] + \beta \\
&= nS(W,H) + \beta'
\end{aligned}
$$

where β and β' are constants independent of W and H. The second transition follows from Jensen's inequality. Therefore, minimizing $S(W,H)$ with respect to W, H maximizes the likelihood of this probabilistic model.

As stated in the following theorem, the CENTER MAPPING NMF PROBLEM is equivalent to the problem of finding a weighted NMF [4] of the matrix $\bar{X} = \sum_q \lambda_q X^{(q)} {\Pi^{(q)}}^T \operatorname{diag}(\frac{1}{g})$, where $\operatorname{diag}(\frac{1}{g})$ is an $n \times n$ matrix with $\frac{1}{g_i}$ on the diagonal and zero for all other entries.

Theorem 1. *Let $\bar{X} = \sum_q \lambda_q X^{(q)} {\Pi^{(q)}}^T \operatorname{diag}(\frac{1}{g})$. We have,*

$$S(W,H) = \sum_i g_i c(WH_{\cdot i}, \bar{x}_{\cdot i}) + \tau$$

where $c(u,v) = \|u-v\|^2$ or $c(u,v) = gKL(v||u) = \sum_l v_l \log \frac{v_l}{u_l} - v_l + u_l$ or $c(u,v) = KL(v||u) = \sum_l v_l \log \frac{v_l}{u_l}$, and τ is a constant that does not depend on W, H.

Proof. Denote by $X = WH$. We first prove the theorem for the Euclidean distance $c(u,v) = \|u-v\|^2$. We write the objective function explicitly and simplify it using $\sum_j \Pi_{ij}^{(q)} = g_i$ and $\sum_q \lambda_q = 1$.

$$
\begin{aligned}
S(W,H) &= \sum_q \lambda_q \sum_i \sum_j \left\| x_{\cdot i} - x_{\cdot j}^{(q)} \right\|^2 \pi_{ij}^{(q)} \\
&= \sum_q \lambda_q \sum_i \sum_j {x_{\cdot i}}^T x_{\cdot i} \pi_{ij}^{(q)} - 2{x_{\cdot i}}^T x_{\cdot j}^{(q)} \pi_{ij}^{(q)} + \beta \\
&= \sum_i {x_{\cdot i}}^T x_{\cdot i} \sum_j \pi_{ij}^{(q)} \sum_q \lambda_q - 2 \sum_i \sum_j {x_{\cdot i}}^T \sum_q \lambda_q x_{\cdot j}^{(q)} \pi_{ij}^{(q)} + \beta \\
&= \sum_i g_i {x_{\cdot i}}^T x_{\cdot i} - 2 \operatorname{Tr}(X^T \sum_q \lambda_q X^{(q)} {\Pi^{(q)}}^T) + \beta \\
&= \operatorname{Tr}(X^T X \operatorname{diag}(g)) - 2 \operatorname{Tr}(X^T \sum_q \lambda_q X^{(q)} {\Pi^{(q)}}^T \operatorname{diag}(\frac{1}{g}) \operatorname{diag}(g)) + \beta \\
&= \operatorname{Tr}\big((X^T X - 2X^T \bar{X}) \operatorname{diag}(g)\big) + \beta \\
&= \sum_i g_i \|x_{\cdot i} - \bar{x}_{\cdot i}\|^2 + \beta'
\end{aligned}
$$

where β and β' are constants that do not depend on W, H.

Next, we prove the theorem for the generalized KL divergence $c(u, v) = gKL(v||u) = \sum_l v_l \log \frac{v_l}{u_l} - v_l + u_l$. Again, we write the objective function explicitly and simplify it:

$$\begin{aligned}
S(W,H) &= \sum_q \lambda_q \sum_i \sum_j \Big[\sum_l x_{li} - x_{lj}^{(q)} - x_{lj}^{(q)} \log(x_{li}) + x_{lj}^{(q)} \log(x_{lj}^{(q)}) \Big] \pi_{ij}^{(q)} \\
&= \sum_q \lambda_q \sum_i \sum_j \Big[\sum_l x_{li} - x_{lj}^{(q)} \log(x_{li}) \Big] \pi_{ij}^{(q)} + \gamma \\
&= \sum_i \sum_l \Big[x_{li} \sum_j \pi_{ij}^{(q)} \sum_q \lambda_q - \log(x_{li}) \sum_j \sum_q \lambda_q x_{lj}^{(q)} \pi_{ij}^{(q)} \Big] + \gamma \\
&= \sum_i g_i \sum_l \Big[x_{li} - \bar{x}_{li} \log(x_{li}) \Big] + \gamma \\
&= \sum_i g_i \mathrm{gKL}(\bar{x}_{\cdot i} || x_{\cdot i}) + \gamma'
\end{aligned}$$

where γ and γ' are constants that do not depend on W, H. This derivation also shows the Theorem holds for the KL divergence $c(u, v) = KL(v||u) = \sum_l v_l \log \frac{v_l}{u_l}$. □

As a result of Theorem 1, we can solve the CENTER MAPPING NMF PROBLEM using an algorithm for weighted NMF such as the iterative update scheme of [4]. When the distribution g of the spots in the center slice is a uniform distribution, i.e. $g = \frac{1}{n}\mathbf{1}_n$, the CENTER MAPPING NMF PROBLEM reduces to a traditional NMF [9] of the matrix $\bar{X} = n \sum_q \lambda_q X^{(q)} \Pi^{(q)^T}$.

We initialize W, H in Algorithm 1 by running NMF on one of the input slices. We stop the iterations of Algorithm 1 when the improvement in the objective function is less than $\epsilon = 10^{-3}$. We observed in both our simulated and real data experiments that the algorithm converges to a stationary point with no change in the objective function within a few iterations. Algorithm 1 outputs the low dimensional representations matrices W, H, the low rank expression matrix $X = WH$, the mappings $\Pi^{(1)}, \ldots, \Pi^{(t)}$ and the full rank expression matrix $\bar{X} = n \sum_q \lambda_q X^{(q)} \Pi^{(q)^T}$.

S1.3 Initial alignment based on spatial coordinates only

We noticed that in some cases, the conditional gradient procedure used to solve the optimal transport (OT) problems gets stuck in a local minima close to an initial uniform alignment solution $\pi_{ij} = \frac{1}{nn'}$ that assigns all pairs of spots the same value (Supplementary Figure S18). To address this issue, we provide an option in the PASTE software for the user to input an alignment Π that will be used as an initial solution to the OT solver. We found that a heuristic for initializing the OT solver based on an alignment found using the only the spatial coordinates improves the performance of PASTE for some datasets. In this heuristic, we solve an optimization problem that seeks to find a rotation of the spatial coordinates in one slice that minimizes the Wasserstein OT between

the spots of the two slices where the distance is measured according to the spatial distances between the spots. We run PASTE with and without this initialization and take mapping that gives the minimum objective function value.

Similar to Supplementary Section S1.1, we assume without loss of generality that we are given **centered** spatial coordinates $Z \in \mathbb{R}^{2\times n}$ and $W \in \mathbb{R}^{2\times n'}$ such that $Zg = (0,0)^T$ and $Wg' = (0,0)^T$.

Problem S2. *Given ST slices with centered spatial coordinates $Z \in \mathbb{R}^{2\times n}$ and $W \in \mathbb{R}^{2\times n'}$, find a mapping $\Pi \in \Gamma(g, g')$ and a rotation matrix $R \in \mathbb{R}^{2\times 2}$:*

$$\sum_{i,j} \pi_{ij} \left\| z_{\cdot i} - R w_{\cdot j} \right\|^2. \tag{S3}$$

Alvarez-Melis *et al*. proposed an alternating optimization procedure to solve a more general version of Problem S2 [2]. Namely, given a fixed rotation R, the problem boils to a standard Wasserstein OT problem, while given a fixed transportation matrix Π, the problem becomes a generalized Procrustes problem (Supplementary Section S1.1). Although Alvarez-Melis *et al*. show this Problem S2 is equivalent to Gromov OT, we found that using this formulation is easier to optimize empirically.

S1.4 Generating simulated spatial transcriptomics data

We generated simulated ST data by resampling from real ST using the following procedure.

1. Let $\mu = \frac{1}{n}\sum_{li} x_{li}$ be the empirical mean of the total read count and be $v = \frac{1}{n}\sum_i (\mu - \sum_l x_{li})^2$ the empirical variance of the total read count. We select k_i total read counts according to $k_i \sim$ NegativeBinomial(r, p), where $r = \#$ of successes and $p =$ probability of success (the same parameterization as Numpy's function). Next, we let $r = \frac{\mu^2}{v-\mu}$ and $p = \frac{\mu}{v}$, such that the $E(k_i) = \mu$, var$(k_i) = v$ respectively.

2. Generate an expression profile $x'_{\cdot i}$ for spot i according to $x'_{\cdot i} \sim$ Multinomial$(k_i, \frac{x_{\cdot i} + \delta \mathbf{1}_p}{\sum_l x_{li} + \delta p})$, where δ is a small pseudocount.

3. Generate rotated coordinates $v_{\cdot i} = z_{\cdot i}\Theta$, where Θ is a rotation matrix with an angle θ. Then, coordinates $z'_{\cdot i} = y_{\cdot \hat{j}}$ are mapped to the closest spot on the array grid $\hat{j} = \text{argmin}_j \left\| v_{\cdot i} - y_{\cdot j} \right\|$. If the grid spot $y_{\cdot \hat{j}}$ was already mapped to a previous tissue spot, spot i is discarded.

We note that due to step 3, the simulation procedure does not simply rotate the coordinates using a linear transformation. Since we restrict spot locations in our simulation to a specified grid, some of the spots of the real tissue used for simulation move their location to the closest grid spot and some get dropped. This procedure produces spot pairs that may be closer or farther on the simulated ST slice than they had been in the real data. In addition to the spatial variation introduced by step 3, the pseudocount parameter δ serves as a noise factor for the gene expression data. Intuitively, with

higher values of δ, the simulated counts become more uniform across the genes and thus are less informative. We denote by $i \sim j$ a spot i in the original slice and a spot j in the simulated slice that are mapped to one another.

S1.5 Spatial Transcriptomic Data Processing

For each ST slice, we remove spots (genes respectively) that do not contain enough transcript reads across all genes (spots respectively). We choose a relatively conservative gene filtering strategy, removing only genes that have low total counts across all spots since we did not want to bias PASTE by selecting an informative set of genes such as highly variable genes. We also tried to be conservative in removing spots in order to retain spatial structure. For example, the maximum number of spots removed from a slice in the DLPFC dataset was 127 spots, corresponding to 3% of the spots, however the median number of removed spots across all slices was only 26 (out of 4000 spots). We report in Table S1 the minimum number of reads to retain spots/genes and the remaining number of spots/genes in each slices after preprocessing.

Dataset	Min reads per spot	Min reads per gene	Remaining spots	Remaining genes
Breast cancer [13]	100	15	251–264	7453–7998
SCC [7]	100	15	588–709	7369–12241
DFLPC [12]	100	100	3431–4786	9256–12381

Table S1: The minimum number of reads required to retain spots/genes in each dataset and the number of spots/genes remaining in each slice after preprocessing. Spots and genes that had less than these total number of reads in a slice were removed. The remaining spots and genes in each dataset are also reported.

S2 Supplementary results and analyses

S2.1 Supplementary results for simulated data

S2.1.1 Comparison of PASTE to Scanorama on ST alignment simulation

We compared PASTE to an scRNA-seq integration method Scanorama [6] on simulated spatial transcriptomics data. Scanorama integrates gene expression information by resolving noise and batch effects between two or more datasets. Scanorama is not designed to output a matching between cells from RNAseq data, though it does relies on inferring nearest neighbors between cells in the given data sets. To directly compare Scanorama with PASTE in PAIRWISE SLICE ALIGNMENT, we calculated an alignment between spots of the different slices by finding a mapping that minimizes the Wasserstein optimal transport distance, where the transportation cost between the spots is taken as the Euclidean distance between the spots in the integrated gene expression datasets from Scanorama.

To compare to CENTER SLICE INTEGRATION, we ran Scanorama on the three simulated gene expression matrices of the ST slices to obtain $\widetilde{X}^{(q)}$, a batch corrected expression matrix for each slice q. Next, we compared the average difference between the true center expression matrix X and the corrected expression matrices $\widetilde{X}^{(q)}$ from Scanorama to the difference between the true expression matrix X and the integrated slice computed by PASTE; in both cases the difference between the matrices was computed using the KL divergence.

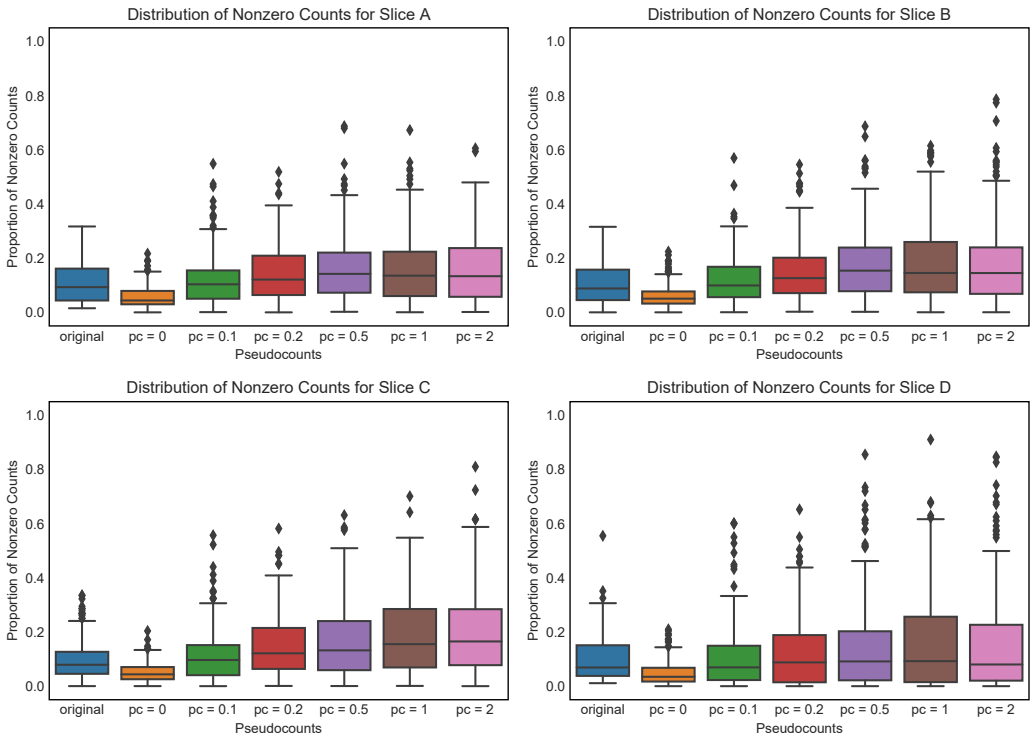


Figure S2: Distribution of the proportion of nonzero transcript counts per spot for the original four slices of breast cancer dataset [13] compared to simulated slices with varying pseudocount. Each simulated slice has $n = 254, 251, 264, 262$ spots for Slice A, B, C, D respectively. The boxes correspond to the 25%, 50% and 75% quantiles and the whiskers correspond to the 1.5 interquantile range.

a Simulation Pairwise

	Slice A	Slice B	Slice C	Slice D
Mixed	1.862	1.872	1.856	1.877
Gene Exp Only	1.862	1.872	1.856	1.877
Spatial Only	1.862	1.872	1.856	1.877

b Simulation Center

	Slice A	Slice B	Slice C	Slice D
Center, Slice S1	1.862	1.860	1.867	1.877
Center, Slice S2	1.862	1.872	1.856	1.877
Center, Slice S3	1	1	1	1

c SCC Pairwise

	Patient 2	Patient 5	Patient 9	Patient 10
Slices A, B	1.968	1.88	1.934	2.019
Slices B, C	1.986	1	2.102	1.742

d SCC Center

	Patient 2	Patient 5	Patient 9	Patient 10
Center, Slice A	1	1	1	1
Center, Slice B	1.968	1.88	1.934	2.019
Center, Slice C	1.956	1.88	2.031	1.758

Figure S3: Sparsity of the mappings Π calculated in pairwise and center alignment by PASTE. We report the average number of nonzero values per row. a) We report the sparsity of Π between a given breast cancer Slice A - D and its respective simulated slice via pairwise alignment. b) For each breast cancer Slice A-D, we simulate 3 slices (S1 - S3) and infer a center slice. Next, we report the sparsity of Π between the center slice and each slice S1 - S3. c) We report the sparsity of Π between each consecutive slices for each patient in the SCC dataset. d) We report the sparsity of Π between the inferred center slice and each original slice for each patient in the SCC dataset.

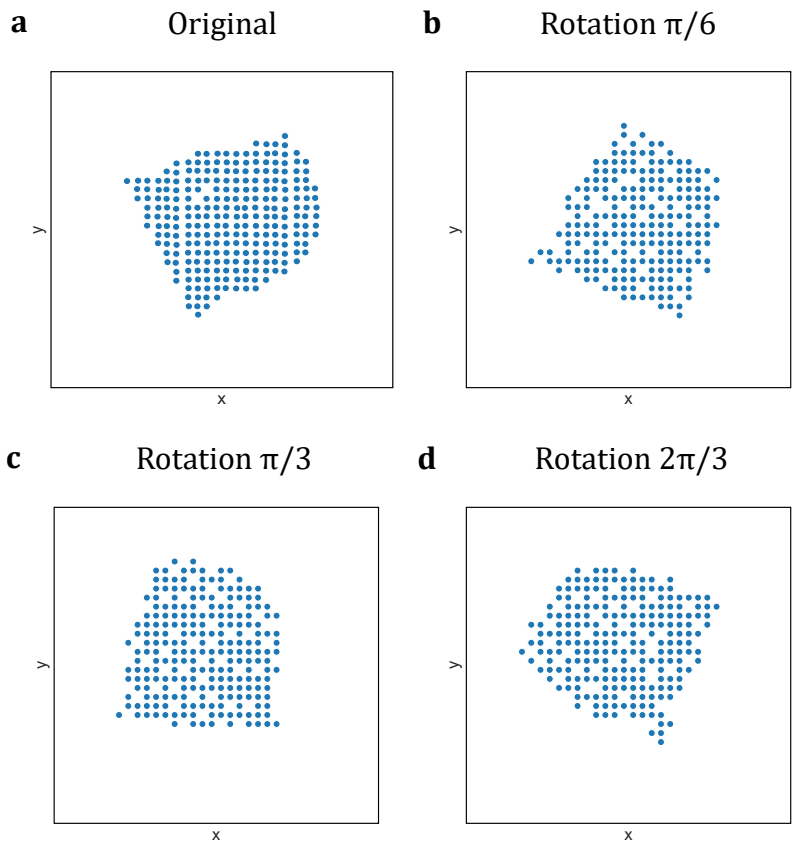


Figure S4: Spatial organization of spots used in center slice alignment simulation of slice B from the breast cancer dataset [13]. (a) Original spatial organization of spots in slice B of breast cancer dataset. (b) - (d) Simulated spatial structures obtained by rotating (a) by $\frac{\pi}{6}, \frac{\pi}{3}, \frac{2\pi}{3}$ respectively.

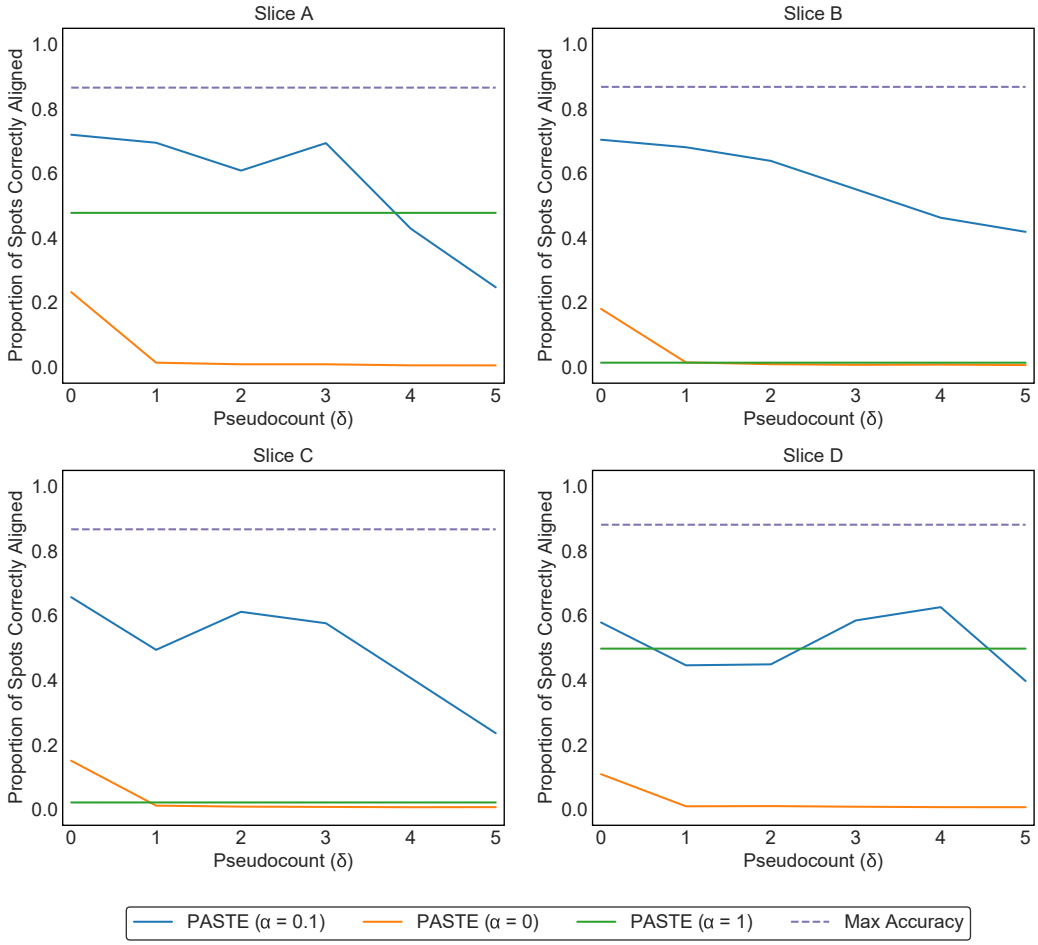


Figure S5: PASTE results on center slice integration of simulated ST slices based on four slices of breast cancer dataset [13].

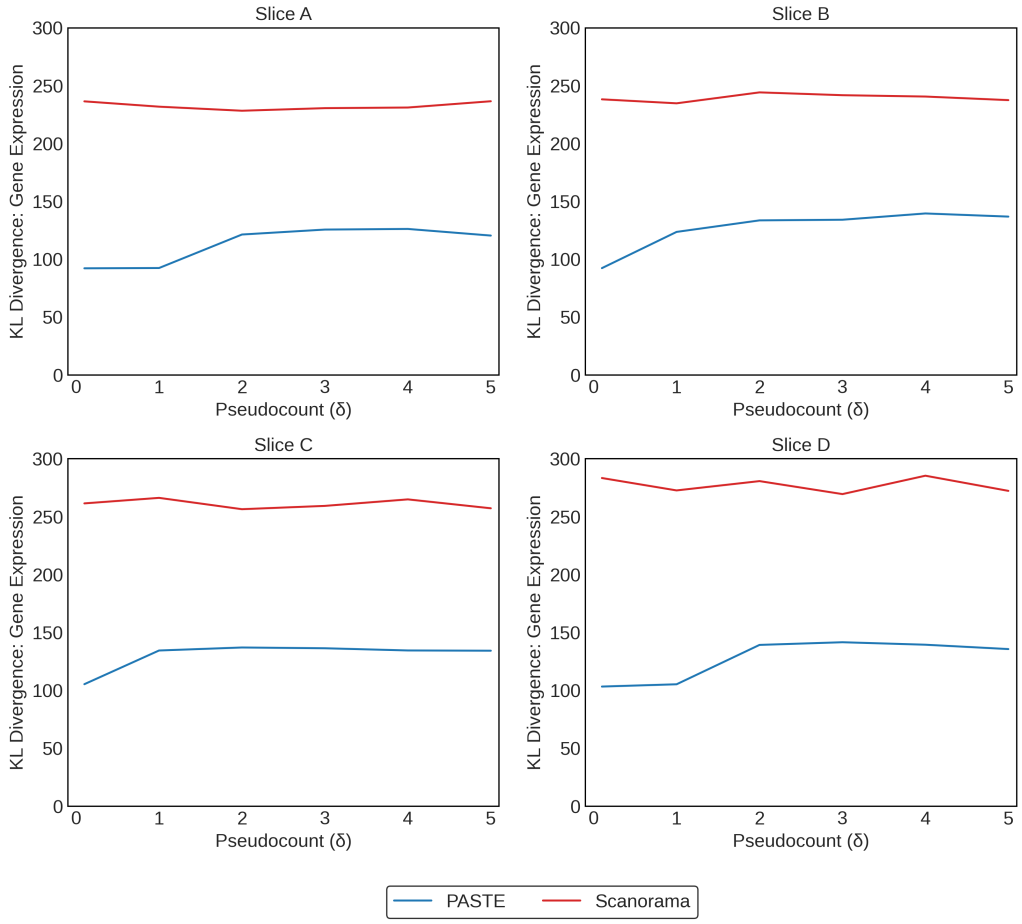


Figure S6: PASTE results on center slice integration of simulated ST slices compared to Scanorama and PASTE based on four slices of breast cancer dataset [13]. Each value is averaged over 10 simulations.

S2.2 Supplementary results for SCC data

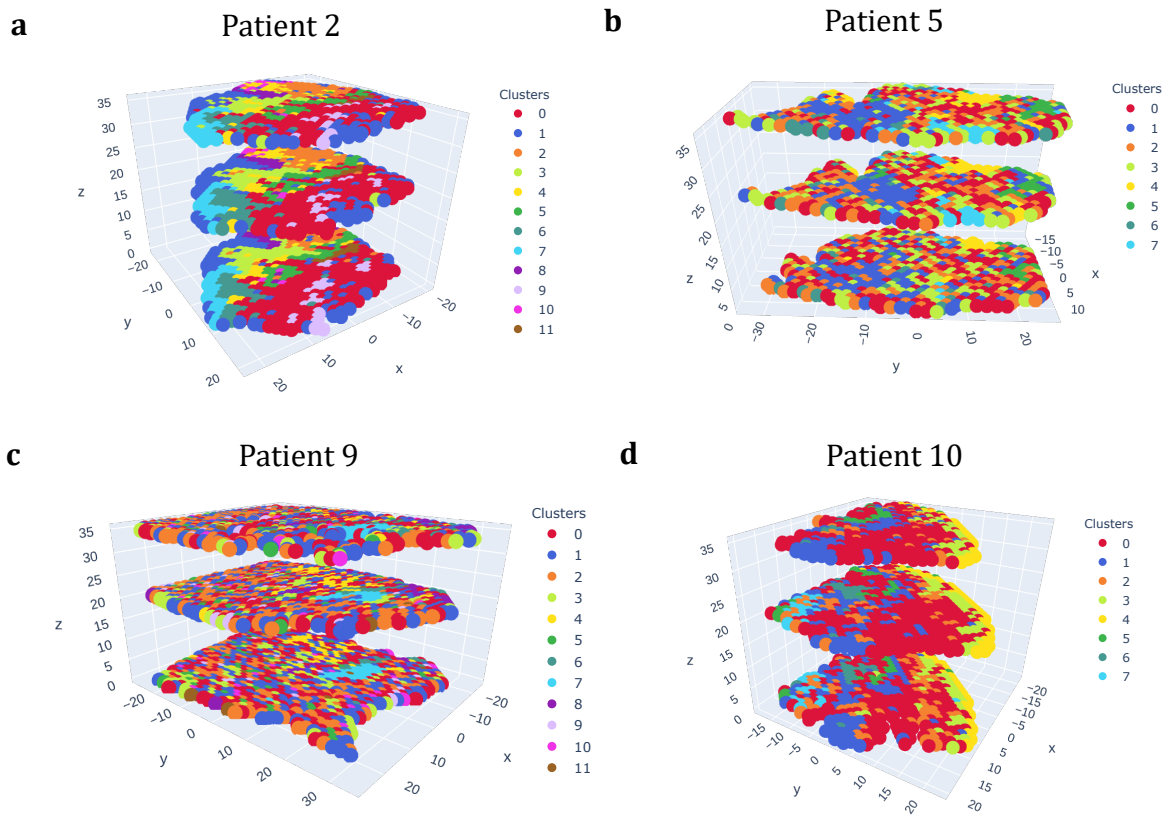


Figure S7: Stacked 3D slices alignment results on SCC data. slices are color coded according to published cluster labels from [7]. The spatial coordinates of slices were aligned using mappings calculated by PASTE with Procrustes analysis (Section S1.1). The the x, y coordinates are in 0.1mm scale while the scale of z coordinate was changed for illustrative purposes.

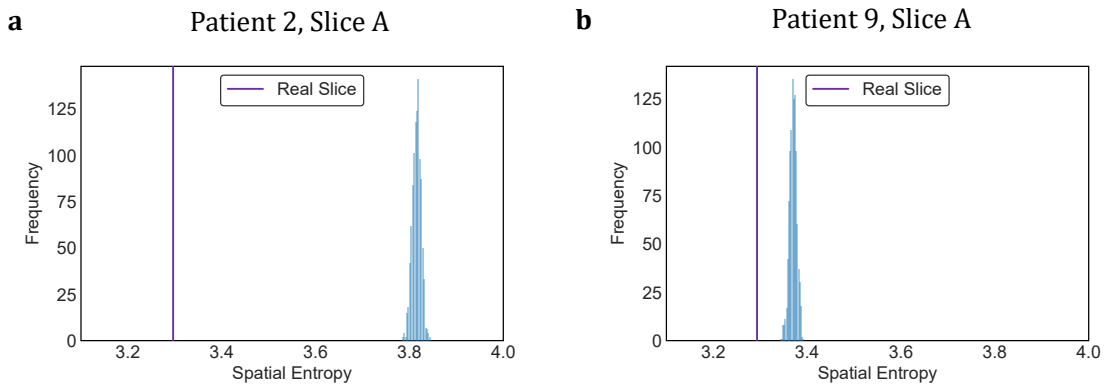


Figure S8: Histogram of spatial entropies. Given the cluster assignments for a slice, we calculated the spatial entropy for 1000 random permutations of cluster labels on the spots. This distribution was used to calculate a spatial entropy z-score for the real slice. (a) Histogram of spatial entropies for patient 2, slice A. (b) Histogram of spatial for patient 9, slice A.

S2.2.1 Effect of Downsampled SCC data on Spatial Coherence

We evaluated the effect of the read depth on the spatial coherence of clustering of the SCC ST data and the alignment accuracy of PASTE. We performed downsampling of the raw counts for ST slices of SCC Patient 2 with various fractions of the total read count and clustered the lower coverage expression. Specifically, we resampled the gene expression profile $x'_{\cdot i}$ of spot i generating a fraction $r \in [0, 1]$ out of the total read count k_i by sampling from a multinomial distribution. That is, $x'_{\cdot i} \sim \text{Multinomial}(k_i r, \frac{x_{\cdot i}+\mathbf{1}\delta}{\sum_l x_{li}+p\delta})$, where k_i is the total read count of spot i, $\delta = 0.1$ is a small pseudocount, and $r \in [0, 1]$ is the fraction of total reads sampled. We cluster all downsampled spot counts across all slices using the same analysis pipeline as in [7]. Given the cluster labels of the spots, we calculate the spatial coherence score for each slice and align adjacent downsampled slices using PASTE to calculate the proportion of aligned spots that are mapped to the same cluster across slices.

We observe that lower sequencing depth results in lower spatial coherence scores of clusters (Figure S9a). This is consistent with our observation that patients 5, 9, and 11, which have less than half the read depth of patient 2, have lower spatial coherence scores using the published clusters for these patients (Figure 3f). Moreover, this result supports our hypothesis that the lower spatial coherence scores of the other three patients are in part due to their lower read depth. Additionally, we see that higher spatial coherence score results in higher cluster mapping accuracy in pairwise alignment of adjacent slices by PASTE (Figure S9b). This is also consistent with our analysis of the spatial coherence scores of the published clusters that showed that higher spatial coherence scores in patient 2 are associated with higher mapping accuracy.

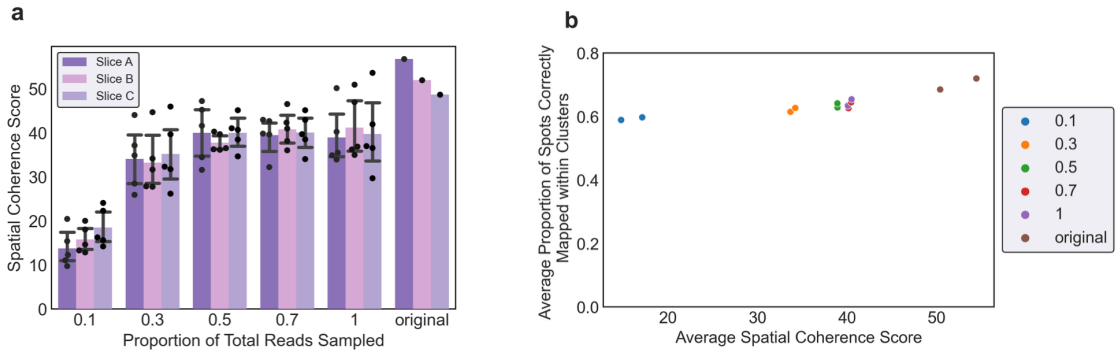


Figure S9: Results of downsampling slices from patient 2 from the SCC data with varying fractions of total reads. (a) The spatial coherence score as a function of the proportion of reads sampled. Error bars show the 95% confidence interval of the mean over 5 samples. "Original" are the clusters from [7] (b) The average spatial coherence and average proportion of spots mapped to the same cluster from PASTE 's alignment. The average is computed across the five samples from each slice.

S2.2.2 Squamous cell carcinomia (SCC) Visium Data

We evaluated PASTE on an additional SCC patient from [7] that has two slices of ST data obtained using the newer, higher-resolution Visium platform [1]. The dataset has a higher number of spots per slice (722 and 674) and a higher number of transcripts per spot (16847 median) compared to the earlier ST platform used in patients 2, 5, 9, and 10. Consistent with the ST analysis of the first four patients, we find that the center slice inferred by PASTE gives more spatially coherent clusters than the clusters obtained on individual slices (Supplementary Figure S10).

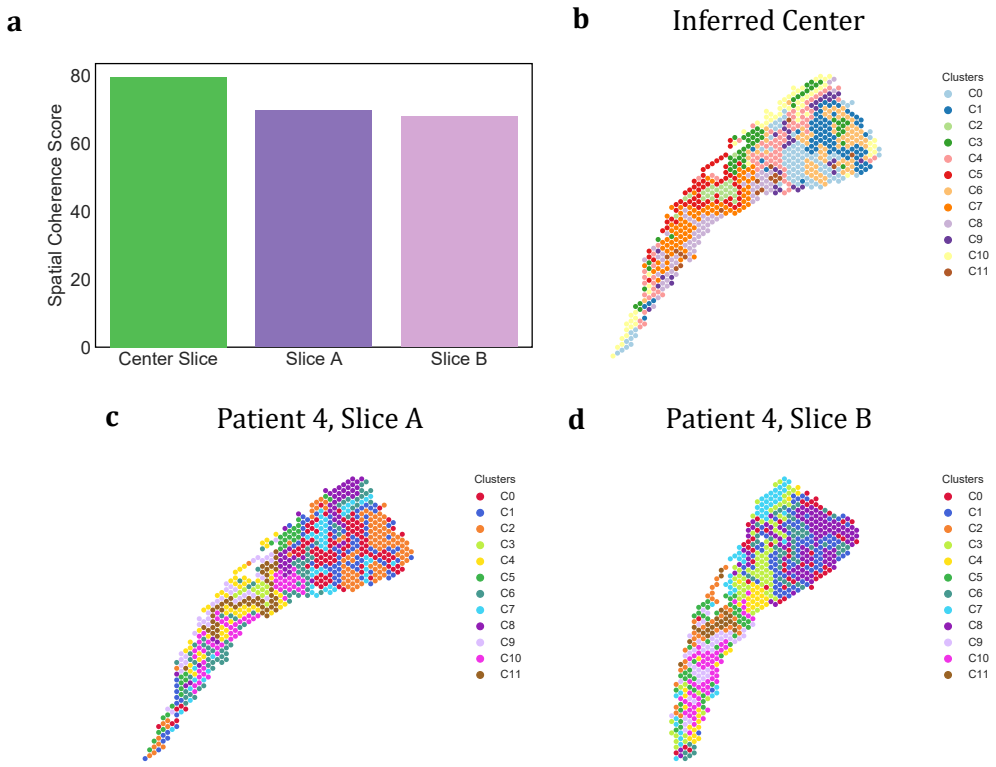


Figure S10: Center slice integration results on SCC Visium data [7]. (a) Spatial coherence comparison between center slice and real slices for patient 4. (b) Distribution of cluster-labeled spots in the inferred center slice of patient computed by PASTE. (c) Distribution of cluster-labeled spots in slice A of patient 4. (d) Distribution of cluster-labeled spots in slice B of patient 4.

S2.3 PASTE alignment of ST data from spinal cord

We applied PASTE to analyze an ST dataset from the human spinal cord [10]. These tissue slices have a fairly symmetric spatial organization, and thus this dataset is a good test of PASTE's ability to handle spatial symmetry. We performed three pairwise alignments with PASTE: (1) Alignment of the first slice (Slice A) to the second slice (Slice B); (2) Alignment of Slice A to a horizontal flip of Slice B; (3) alignment of Slice A to a vertical flip to Slice B (Supplementary Figure S11). To distinguish between the symmetrical alignments, we artificially relabled spot annotations of the two sides of symmetrical regions (e.g. vent horn and dors horn) and used these annotations to calculate the alignment accuracy between slices (Supplementary Figure S11). We did not observe changes in the accuracy between the original slice pairs and the flipped ones, showing that PASTE is robust to samples with natural, but imperfect, spatial symmetry (Supplementary Figure S12). In addition, we observe that running PASTE without spatial information ($\alpha = 0$) gives substantially lower accuracy (Supplementary Figure S12). This shows that spatial information is very useful

to correctly align symmetrical tissues, and that restricting to transcriptional information alone may result in incorrect alignments as spots may be equally likely to be mapped to either side of an axis of symmetry.

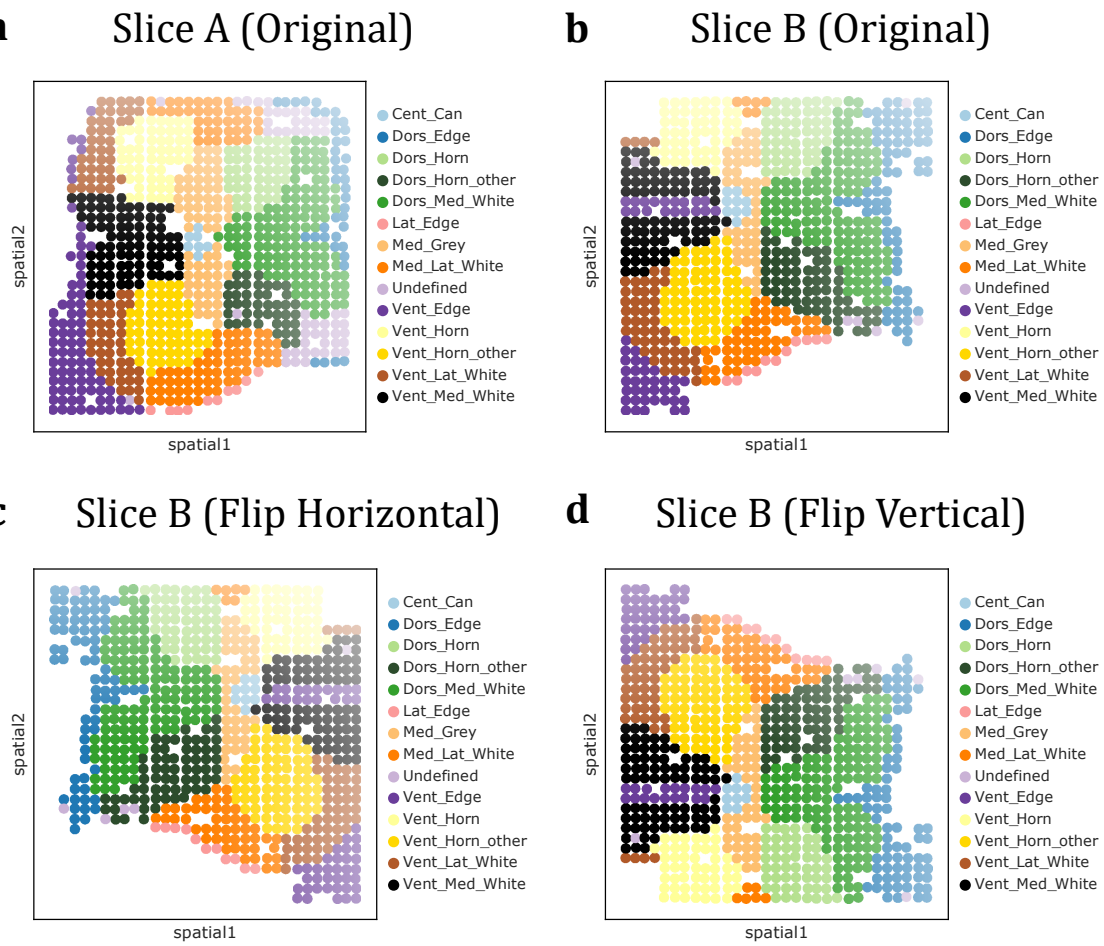


Figure S11: ST datasets from spinal cord with spots labeled according to annotation in [10]. (a-b) Slices A and B. (c) Slice B flipped horizontally (d) Slice B flipped vertically.

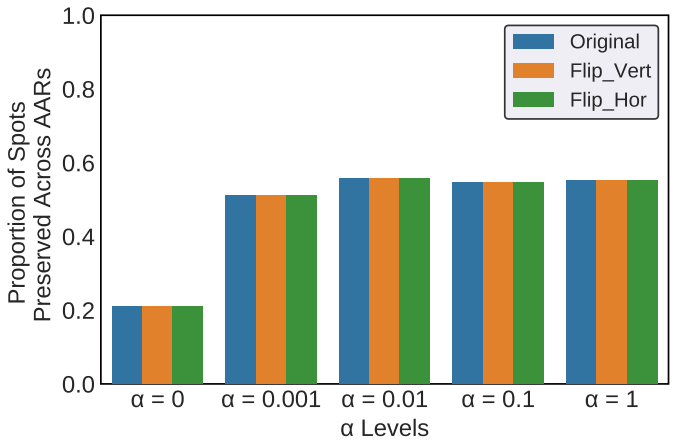


Figure S12: PASTE performance on the spinal cord ST dataset for different values of the parameter α. Each bar corresponds to the accuracy of aligning slice A to either the original slice B or slice B flipped vertically or horizontally.

S2.4 PASTE analysis of HER2 breast cancer ST data

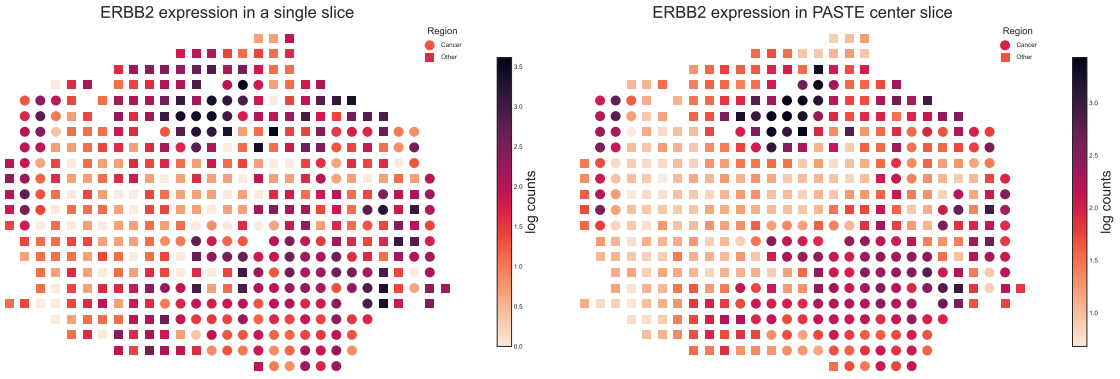
We ran PASTE on an ST dataset from Her2 breast [3], consisting of 6 patients each with 6 ST slices and 2 patients each with 3 ST slices. In the publication [3], the spots of one ST slices from each patient were manually annotated by a pathologist based on the H&E image.

We used PASTE to derive an integrated center slice per patient and clustered the spots in the integrated slice. We compared the ARI of the clusters derived from PASTE 's integrated slice and the clusters derived by Anderson *et al.* [3] using the pathology-annotated labels as a "gold standard" for each patient. We used the same number of clusters used in [3] and calculated the ARIs only over spots that were not annotated as undetermined.

We see that the clusters derived by the PASTE integration are largely comparable to the cluster assignment from Anderson *et al.* [3], but having higher ARI on 5/8 patients. Both clusterings show varying levels of agreement with the pathological annotation (Supplementary Table S2). Examining the cluster assignment from PASTE integration for Patient G, shows that PASTE is able to preserve clusters consisting of a small number of neighboring spots of cancer in situ (Extended Data Figure 4). Furthermore, the expression of ERBB2 (encoding the HER2-receptor) shows clearer agreement with pathological annotations of cancer regions in the PASTE center slice than in raw ST data of a single slice (Supplementary Figure S13).

Method/Patient	A	B	C	D	E	F	G	H
Anderson *et al.* [3]	0.158	0.268	**0.061**	**0.188**	0.084	0 .083	**0.428**	0.369
PASTE center	**0.186**	**0.32**	0.052	0.184	**0.149**	**0.147**	0.309	**0.379**

Table S2: ARIs of clusters from Andersson et al. [3] and derived from the PASTE integrated slice compared to pathological annotation in one slice from 8 breast cancer patients. Bold indicates the higher ARI for each patient .



(a) Raw ST data

(b) PASTE center slice

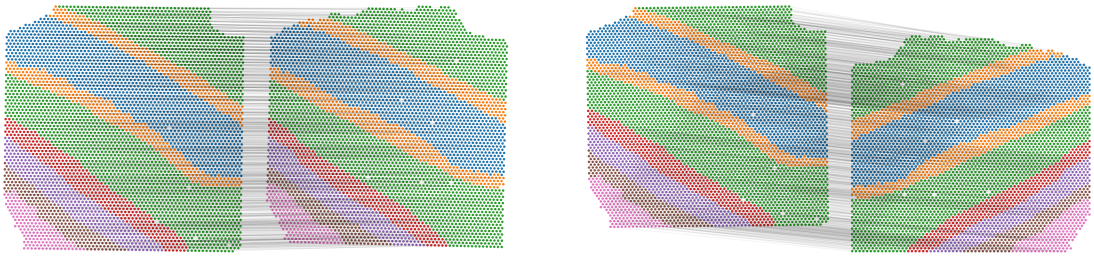
Figure S13: (a) Expression of ERBB2 (HER2-receptor) in a single ST slice of breast cancer patient G and (b) in the PASTE integrated slice. Circles indicate spots annotated as cancer regions while squares indicate regions not annotated as cancerous.

S2.5 Supplementary results and analyses for DLPFC data

Sample/Slice	A	B	C	D
I	151507	151508	151509	151510
II	151669	151670	151671	151672
III	151673	151674	151675	151676

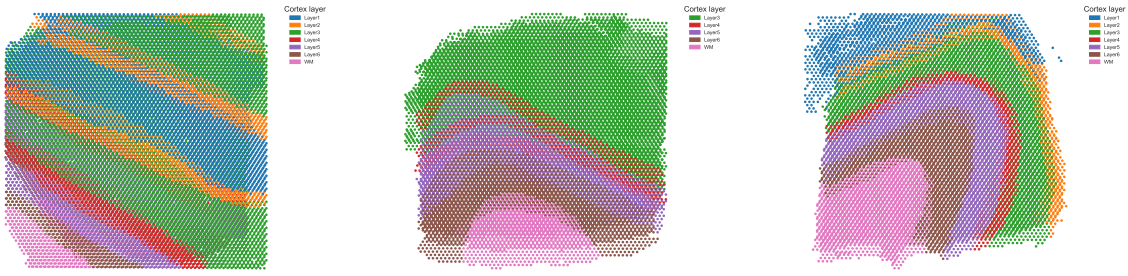
Table S3: DLPFC sample and slice names from [12].

Pairwise alignment of consecutive slices



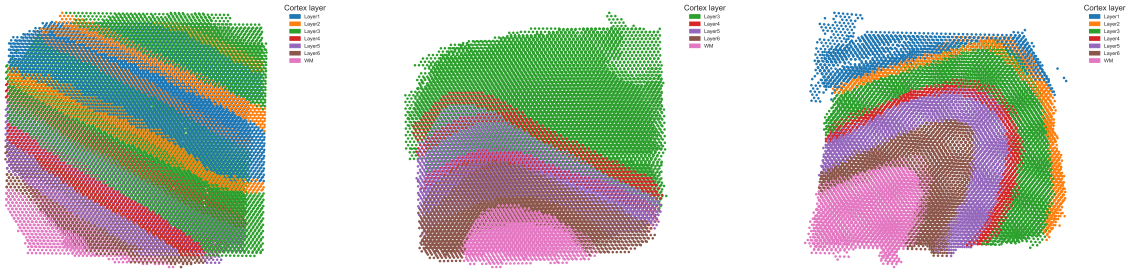
(a) PASTE (b) STUtility

Figure S14: Results of pairwise alignment and inferred coordinates of slices C and D from DLPFC Sample I using (a) PASTE and (b) STUtility. The slices are placed side by side and the spots are positioned according to the new coordinates inferred by the alignment of each method. Gray lines connect the 1000 spot pairs with highest alignment values from each method. PASTE produces an alignment that is consistent with spatial organization of slices, while STUtility produces an alignment where one slice is flipped relative to the other slice.



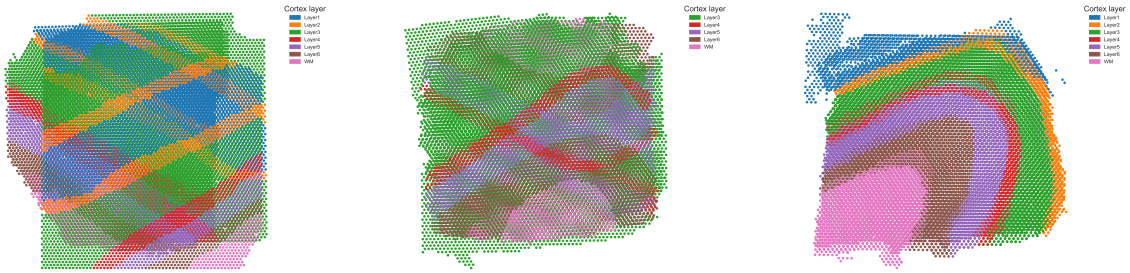
(a) Sample I (b) Sample II (c) Sample III

Figure S15: Stacking four ST slices of the DLPFC samples without alignment.



(a) Sample I (b) Sample II (c) Sample III

Figure S16: Stacking four ST slices of the DLPFC samples after alignment with PASTE



(a) Sample I (b) Sample II (c) Sample III

Figure S17: Stacking four ST slices of the DLPFC samples after alignment with STUtility

S2.5.1 The effect of the alpha parameter on PASTE performance

The α parameter in PASTE adjusts the relative contribution of expression information and spatial information in the pairwise alignment. We evaluated the effect of the α parameter on PASTE's performance on both simulated data and on the DLPFC data. On simulated data, we found that PASTE's performance varies only slightly when not extremely close to either $\alpha = 0$ (expression only) or $\alpha = 1$ (spatial only) (Extended Data Figure 2). On the DLPFC data, we found that PASTE's performance was also stable over intermediate values of $0 < \alpha < 1$, although performance is worse when using a uniform alignment to initialize the optimization (Supplementary Figure S18). In particular, the OT optimization gets stuck in the local optima close to a uniform alignment for some values of α. However, when initializing the OT procedure with an informed non-uniform alignment based on spatial similarity (Supplementary Section S1.3), PASTE performance maintains high accuracy across most values of α that are not extremely close to 0 (expression only).

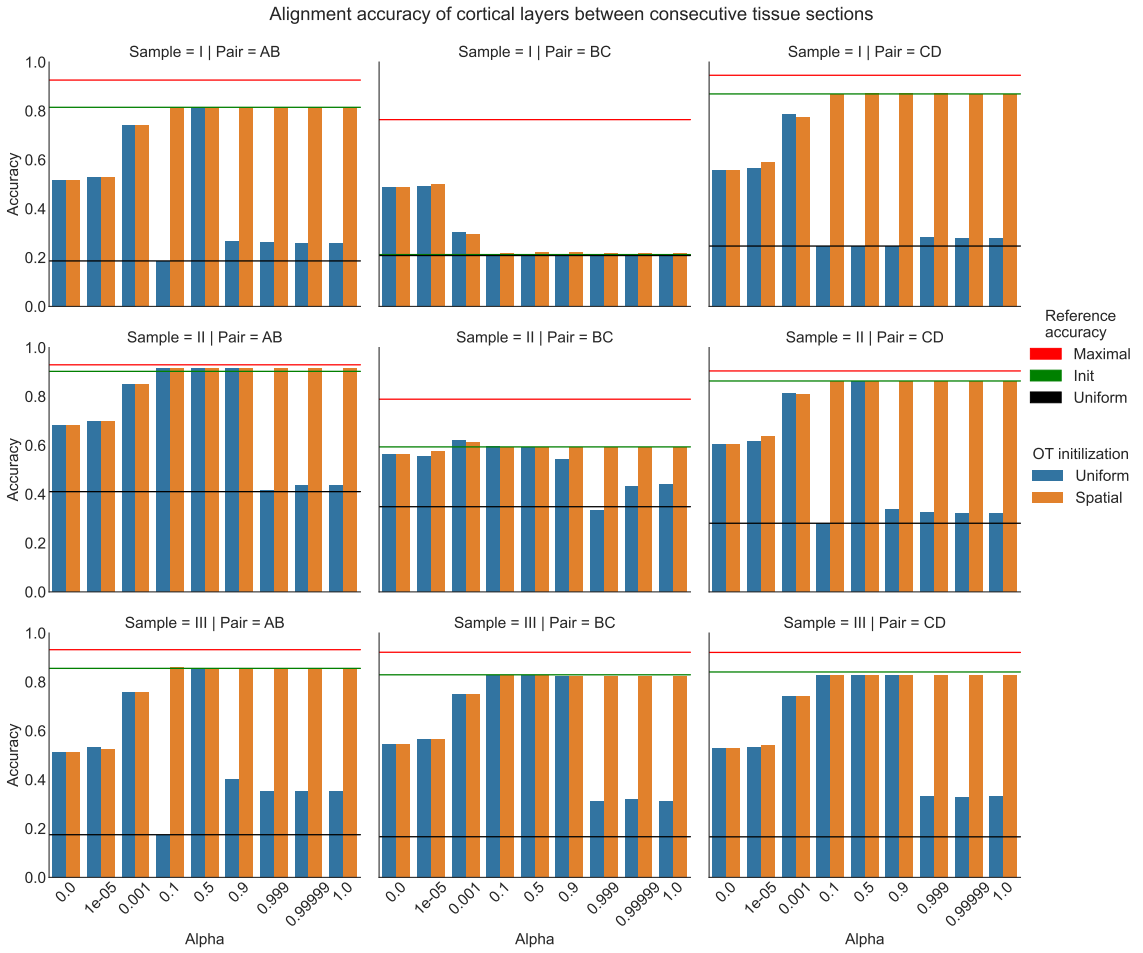


Figure S18: Alignment accuracy for consecutive tissue slices of the DLPFC datasets. For every sample and every pair of consecutive slices we ran PASTE and measured the accuracy of mapping spots to the same neocortical layers. PASTE was run with varying values of α and the OT was initialized with either a uniform mapping or a mapping based on spatial distances (Supplementary Section S1.3). Horizontal lines present reference accuracies: maximal possible accuracy between the two slices based on the number of spots from each layer (red), the accuracy of the initial alignment based on spatial data only (green), the accuracy of an alignment which gives all spot pairs the same uniform mapping (black).

Next, we examined using the value of the objective function (Equation 1) as a proxy for selecting α, reasoning that a lower objective value would correspond to better alignment. A similar approach was used by SCOT [5], a recent method based for multi-modal integration of single-cell sequencing datasets. Here, for a fixed value of α, we run PASTE using the standard procedure where we select the mapping Π that minimizes the objective function over all initializations. We

observed that the objective function value is correlated with the accuracy of the alignment (Supplementary Figure S19). However, selecting the value of α with the minimum objective value does not always yield the best accuracy (Supplementary Figure S20). We see that the α value that minimized the objective function gives close to the best accuracy in 4 out of 9 pairs of slices (Supplementary Figure S20a). Restricting to values of α in the range [0.1,0.9], the α value that minimized the objective function gives the best accuracy for PASTE in 7 out of 9 pairs of slices (Supplementary Figure S20b). This shows that using the value of the objective function can help select a value for the parameter α and that combining both transcriptional and spatial similarity ($0 < \alpha < 1$) is preferable to either of these alone.

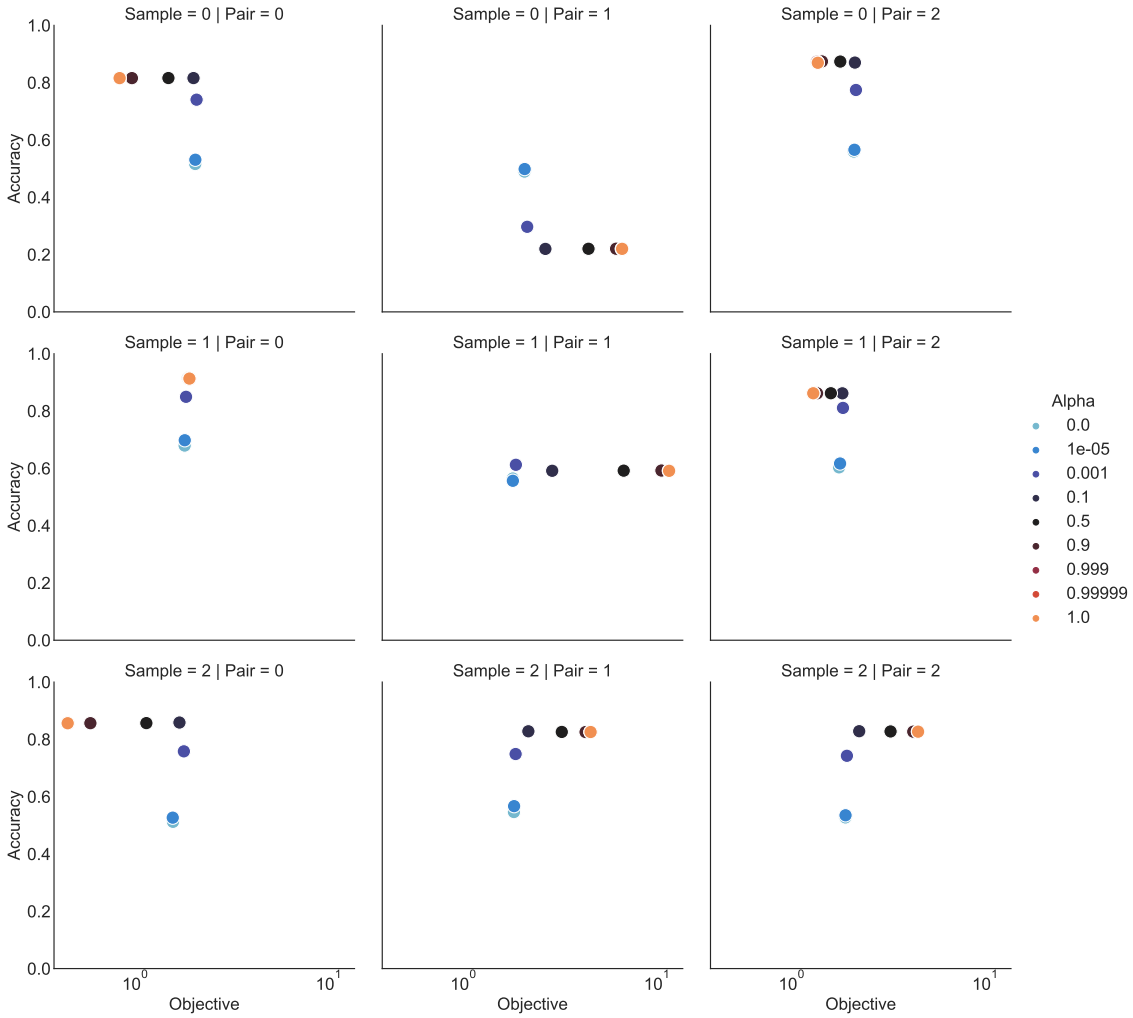
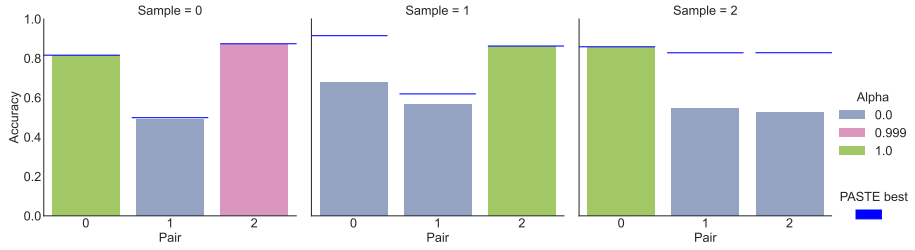
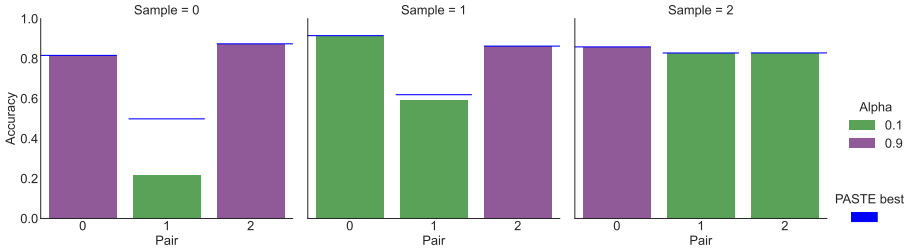


Figure S19: Alignment accuracy for consecutive tissue slices of the DLPFC datasets as a function of the value of the objective function for different values of α.



(a) All alphas



(b) Alpha in [0.1,0.9]

Figure S20: Alignment accuracy for consecutive tissue slices of the DLPFC datasets for the value of the parameter α that gave the minimum value of the objective function, as shown in Supplementary Figure S19. (a) Minimum over all values of α. (b) Minimum over α in [0.1,0.9]. Bars are colored according to value of α. The blue line indicates the best accuracy of PASTE obtained across all tested values of α.

S2.5.2 The effect of expression normalization and gene sets on PASTE performance

In this section, we examined how different gene normalization schemes, different gene expression similarity measures and different gene sets affect the performance of PASTE. We evaluated PASTE on the DLPFC data using using library-size normalized and log transformed transcript counts and calculating the expression cost function using Euclidean distance. We performed this calculation using all genes or selecting only the top 2000 highly variable genes (HVG). We compared against the default mode of PASTE that uses the KL divergence on all the genes. We observe that the alignment accuracy across the adjacent slice pairs is similar for the different cost functions with a slight advantage when using the KL divergence which achieves the highest accuracy in 7 out of 9 slice pairs (Extended Data Figure 7).

S2.5.3 PASTE with non uniform spot priors

To demonstrate that PASTE can use other distributions g on spots that encode prior information on the importance of the spots, we analyzed the DLPFC dataset using a distribution g based on the estimated number of cells per spot in the data. We estimated the number of cells per spot using the number of nuclei reported by VistoSeg (spatialLIBD) [14], a cell segmentation procedure that

identifies individual nuclei in H&E stained images. We then normalized cell counts for each spot to create a weight distribution g over spots in each slice, and used these distributions to align adjacent slices with PASTE.

We did not observe dramatic changes in the overall alignment accuracy when using the cell counts as spots weights vs. using uniform weights over spots (Supplementary Figure S21). Moreover, the small changes in accuracy were attributed to the spots being weighted differently. For example, while layer 3 spots contributed 0.4-0.62 of the weight in slices of sample II when using a uniform distribution over spots, the layer 3 spots contributed only 0.33-0.45 of the weight when using the cell counts (Supplementary Figure S22). This is because in sample II, the white matter, layers 6 and layer 5 have 3-5.5 nuclei per spots in comparison to $\approx$ 2 nuclei per spot in layers 4 and 5. As the inner layers and white matter spots tend to have more cells per spot, the weighted version of PASTE gave more emphasis to alignment of spots from these regions. This phenomena has some effect when projecting the spatial coordinates using PASTE alignments as the alignment will tend to better match the spots that contain more cells (Supplementary Figure S23).

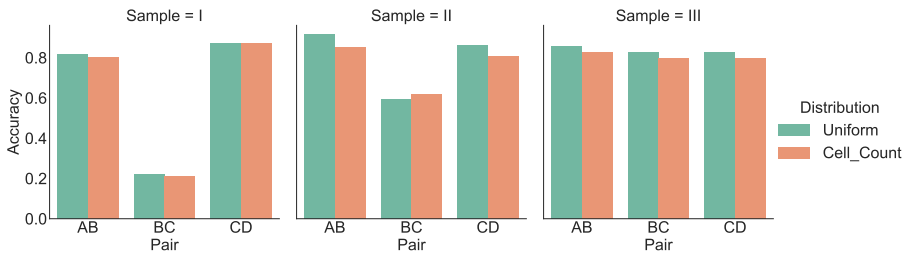


Figure S21: Accuracy of PASTE alignments with a uniform distribution over spots and with a distribution derived from estimated cell counts per spot.

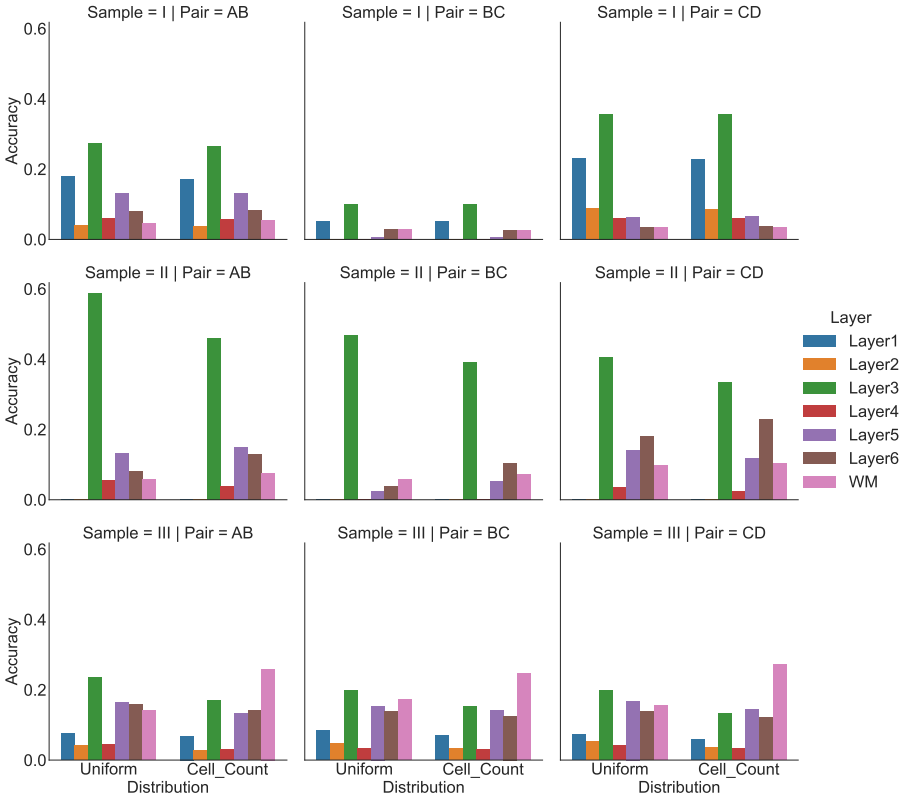


Figure S22: The proportion of alignment accuracy attributed to each layer when using PASTE with uniform or cell count distributions. The sum of accuracy over all layers corresponds to the overall accuracy presented in Figure S21.

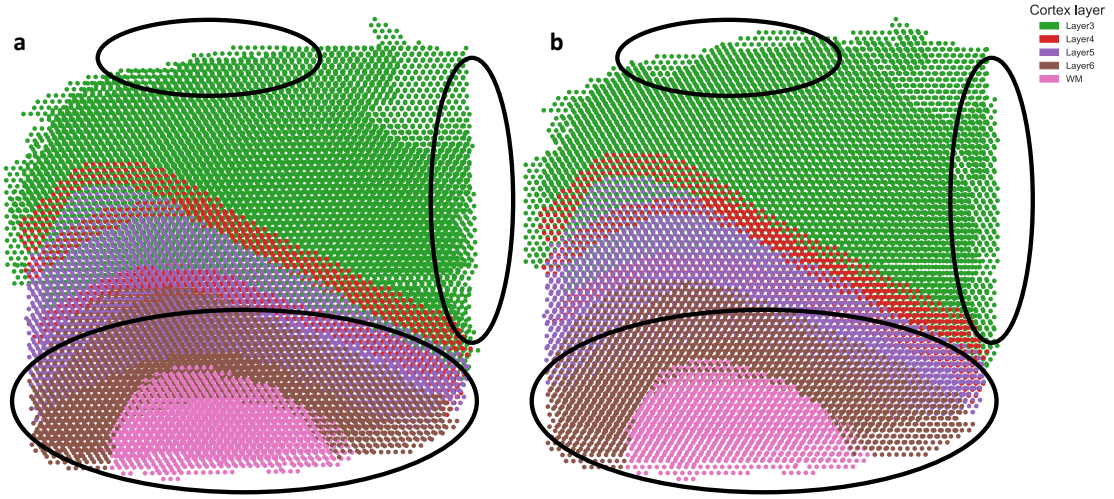
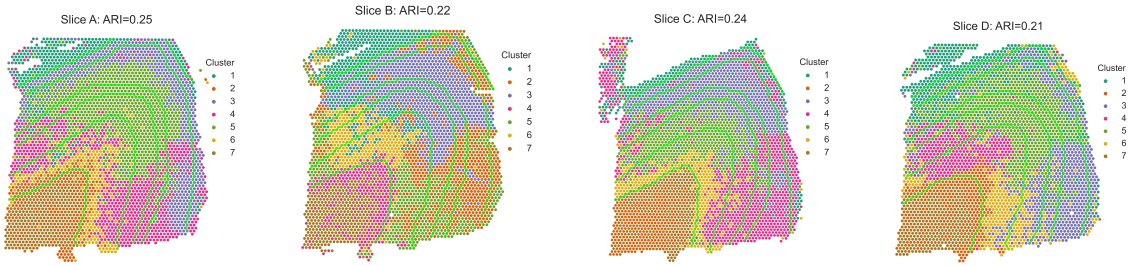


Figure S23: Projection of the slices of DLPFC sample II using PASTE alignments with (a) uniform or (b) cell count distributions. Circles indicate visual differences in alignments. (a) Using the uniform distribution leads to slightly better alignments in layer 3. (b) Using the cell count distribution leads to slightly better alignments of layer 6 and white matter which have higher cell counts.

S2.5.4 Integration of multiple slices of Sample III

	ARI				Num. marker genes				Median marker gene rank			
Method	A	B	C	D	A	B	C	D	A	B	C	D
No integration	0.25	0.22	0.24	0.21	58	57	49	44	703.5	879	753	772
Maynard *et al.*	<0.4				63				1147			
PASTE	**0.53**				80				**427**			
Scanorama	0.17	0.18	0.17	0.16	**84**				3380.5			
Seurat	0.31	0.23	0.26	0.24	79				1852			

Table S4: Comparison of PASTE, Scanorama, and Seurat in clustering spots and identifying marker genes in four slices (labeled A, B, C, and D) of spatial transcriptomics data from dorsolateral prefrontal cortex (sample III) from [12]. The second column gives the Adjusted Rand Index (ARI) of clusters derived from integrated datasets from each method to the ground truth manual annotation into layers. The third column gives the number of known layer-specific marker genes that are significantly differentially expressed (FDR < 0.01) in the integrated data from each method. The fourth column gives the median rank of known marker genes in the differential expression analysis for each layer subset model. Top performing values in each column are in bold text. For reference, we also provide results for analyzing each slice separately with no integration and the results reported in [12].



(a) Slice A (b) Slice B (c) Slice C (d) Slice D

Figure S24: Clustering results of the RNA expression for each slice of sample III separately.

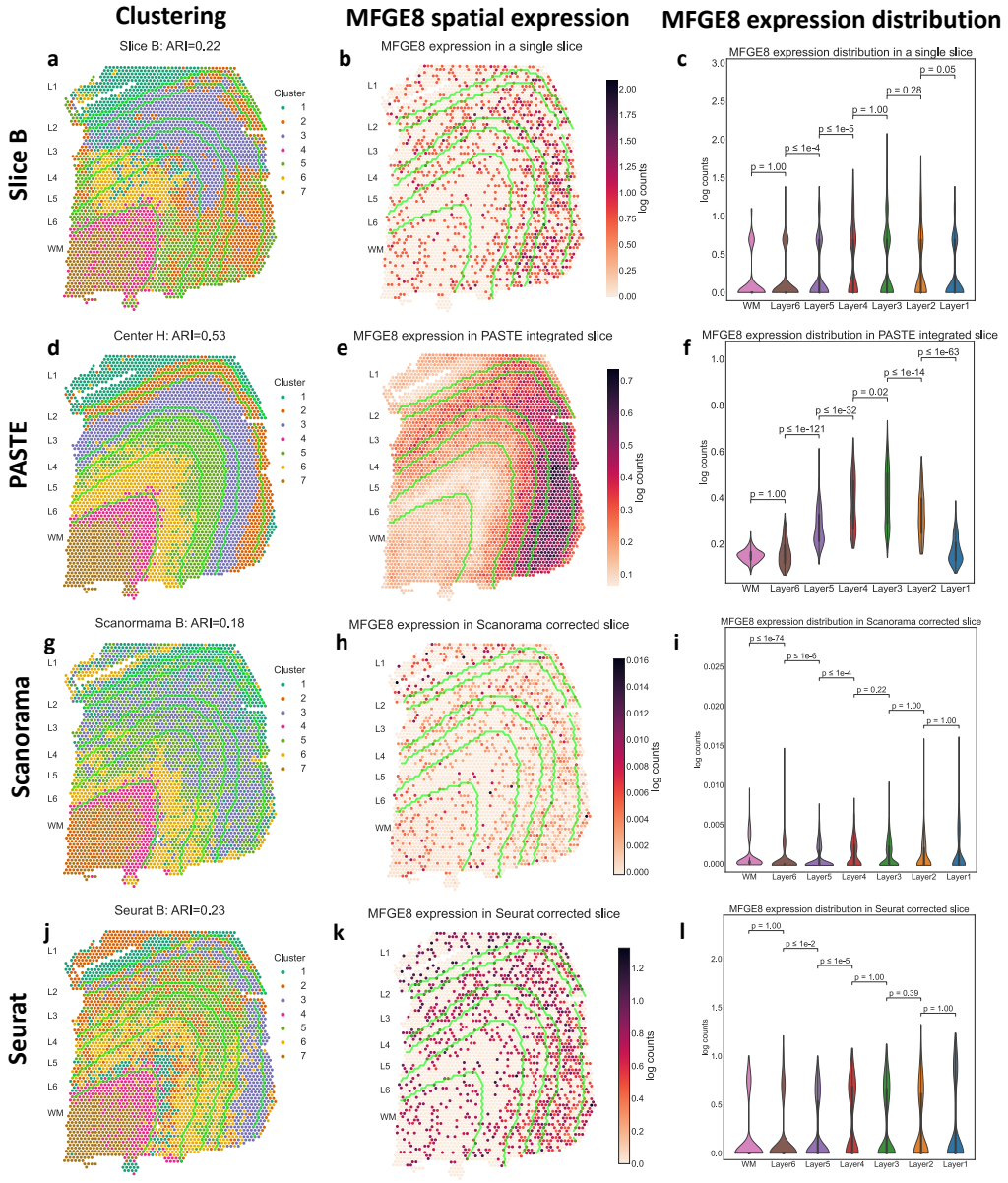
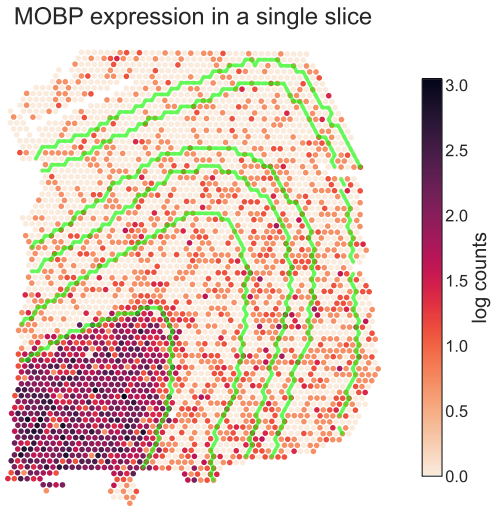
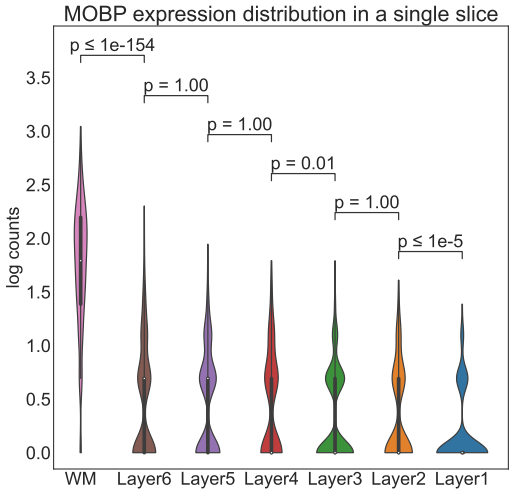


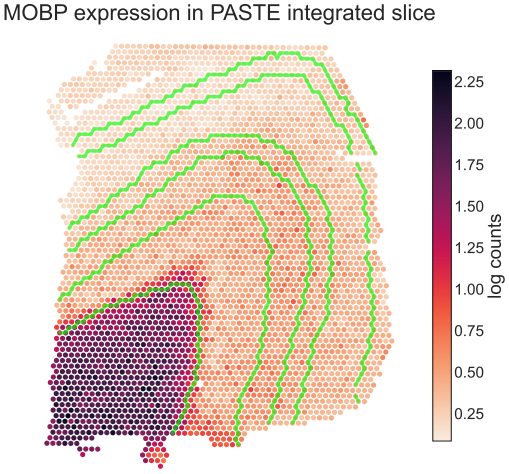
Figure S25: Comparison of gene expression clustering and *MFGE8* layer 3 marker gene expression in a single spatial transcriptomics slice (slice B, top) and integrated slices from PASTE, Scanorama and Seurat (bottom). **(a,d,g,e)** Clustering results projected onto slice B. **(b,e,h,k)** Expression of *MFGE8* in slice B and output from each integration method. **(c,f,i,l)** Distribution of expression of MFGE8 in each layer of slice B from original expression data and output from each integration method. WM and Layers 6 to 1 have 625, 614, 621, 247, 924, 224 and 380 spots respectively. Inner boxplots show the 25%, 50% and 75% quantiles of the distributions. p-values (rounded to the closest power of 10) for the difference in distribution (two-sided Mann-Whitney U test) between adjacent layers are indicated. (Panels (a-f) are reproduced from manuscript Figure 6.)



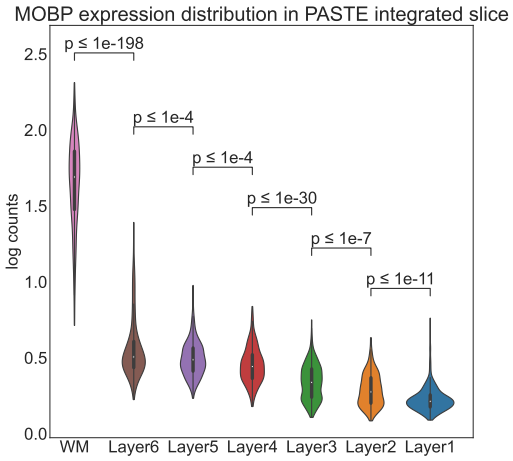
(a) MOBP spatial expression in slice B



(b) MOBP expression distribution in slice B

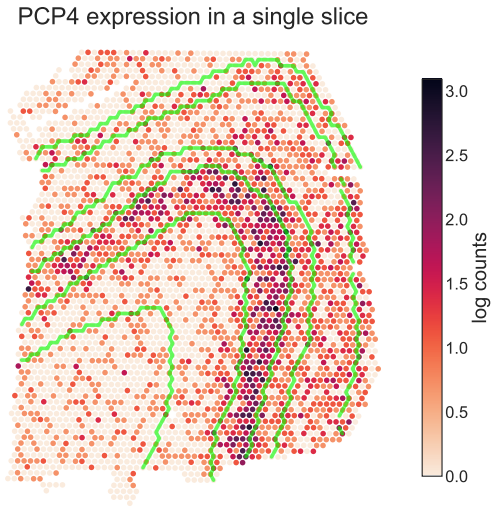


(c) MOBP spatial expression in PASTE integrated slice

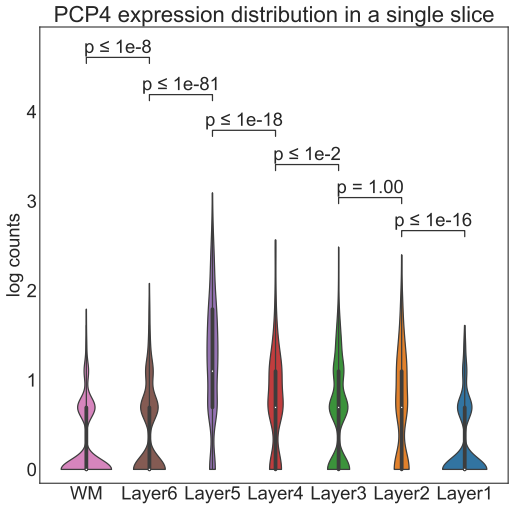


(d) MOBP expression distribution in PASTE integrated slice

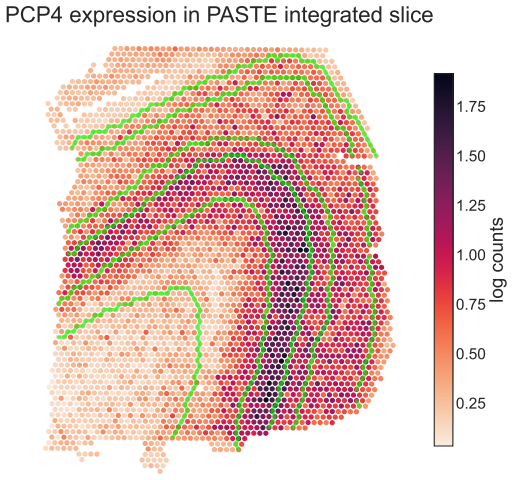
Figure S26: MOBP expression in a single slice and PASTE integrated slice. The boundaries between the layers are marked in green in a and c. WM and Layers 6 to 1 have 625, 614, 621, 247, 924, 224 and 380 spots respectively. Inner boxplots show the 25%, 50% and 75% quantiles of the distributions. p-values (rounded to the closest power of 10) for the difference in distribution (two-sided Mann-Whitney U test) between adjacent layers are indicated. MOBP is a known white matter marker gene.



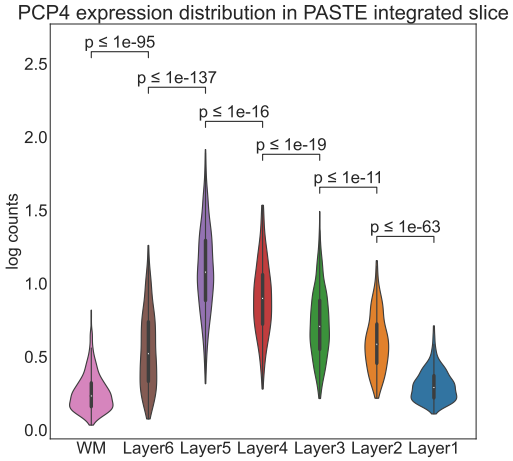
(a) PCP4 spatial expression in slice B



(b) PCP4 expression distribution in slice B



(c) PCP4 spatial expression in PASTE integrated slice



(d) PCP4 expression distribution in PASTE integrated slice

Figure S27: PCP4 expression in a single slice and PASTE integrated slice. The boundaries between the layers are marked in green in a and c. WM and Layers 6 to 1 have 625, 614, 621, 247, 924, 224 and 380 spots respectively. Inner boxplots show the 25%, 50% and 75% quantiles of the distributions. p-values (rounded to the closest power of 10) for the difference in distribution (two-sided Mann-Whitney U test) between adjacent layers are indicated. PCP4 is a known layer 5 marker gene.

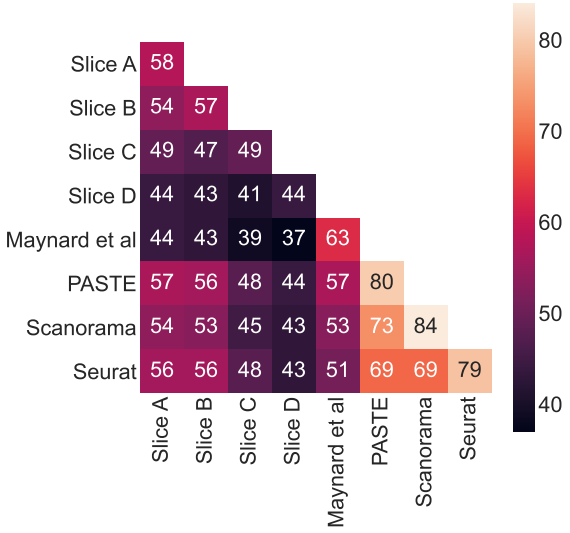


Figure S28: Number of overlapping significantly differentially expressed marker genes identified by each pair of methods.

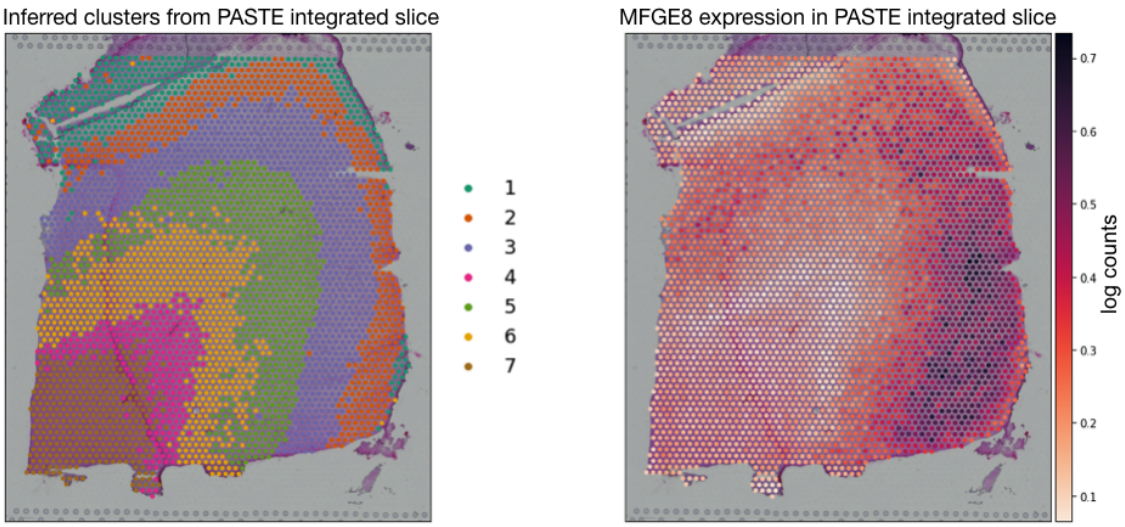


Figure S29: Combined visualization of PASTE output and H&E stained images of DLPFC data slice B, sample III. (Left) Spots are colored according to clusters obtained by clustering the low dimensional representation of the integrated center slice by PASTE. (Right) Spots are colored according to MFGE8 expression from PASTE center slice integration.

S2.5.5 The effect of dimensionality reduction on the number of significantly expressed genes

Count matrix	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	WM
A	3086	447	2290	204	980	320	3083
B	3007	412	2391	213	1059	262	3202
C	2221	416	1316	169	1021	308	2491
D	874	214	898	118	749	283	2196
PASTE low rank	9330	8902	9666	8678	9322	9110	10202
PASTE full rank	7767	2388	5674	1760	3919	2797	7099

Table S5: Number of significantly differentially expressed genes (DEGs) for each neocortical layer in different count matrices. Significance was a assessed with Wilcoxon rank sum test and thresholded at FDR adjusted p-value of <0.01.

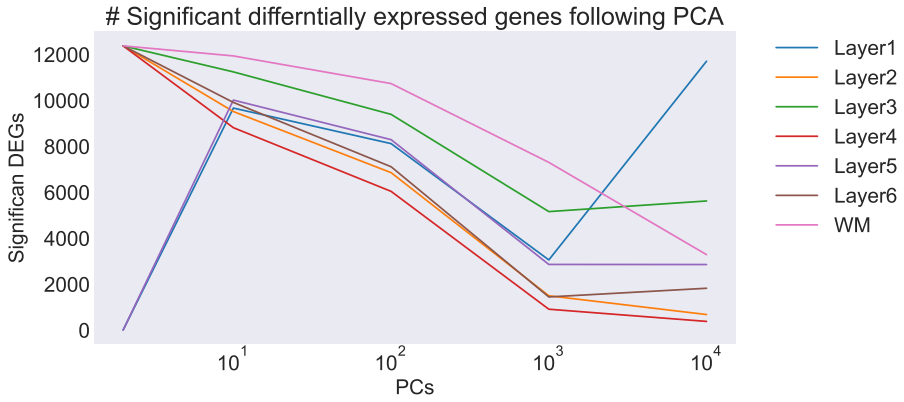


Figure S30: The number of significant DEGs as a function of the number of principle components used to lower the rank of the count matrix of slice B.

References

[1] 10x Genomics. Visium spatial gene expression: Map the whole transcriptome within the tissue context, 2019. Accessed: October 2020.

[2] David Alvarez-Melis, Stefanie Jegelka, and Tommi S. Jaakkola. Towards optimal transport with global invariances. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1870–1879. PMLR, 16–18 Apr 2019.

[3] Alma Andersson, Ludvig Larsson, Linnea Stenbeck, Fredrik Salmén, Anna Ehinger, Sunny Z. Wu, Ghamdan Al-Eryani, Daniel Roden, Alex Swarbrick, Åke Borg, Jonas Frisén, Camilla Engblom, and Joakim Lundeberg. Spatial deconvolution of her2-positive breast cancer delineates tumor-associated cell type interactions. *Nature Communications*, 12(1):6012, 2021.

[4] Vincent D Blondel, Ngoc-Diep Ho, Paul Dooren, et al. Weighted nonnegative matrix factorization and face feature extraction. In *In Image and Vision Computing*. Citeseer, 2008.

[5] Pinar Demetci, Rebecca Santorella, Björn Sandstede, William Stafford Noble, and Ritambhara Singh. Gromov-wasserstein optimal transport to align single-cell multi-omics data. *bioRxiv*, 2020.

[6] Brian Hie, Bryan Bryson, and Bonnie Berger. Efficient integration of heterogeneous single-cell transcriptomes using scanorama. *Nature Biotechnology*, 37(6):685–691, 2019.

[7] Andrew Ji, Adam Rubin, Kim Thrane, Sizun Jiang, David Reynolds, Robin Meyers, Margaret Guo, Benson George, Annelie Mollbrink, Joseph Bergenstråhle, Ludvig Larsson, Yunhao Bai, Bokai Zhu, Aparna Bhaduri, Jordan Meyers, Xavier Rovira-Clavé, S Hollmig, Sumaira Aasi, Garry Nolan, and Paul Khavari. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 182:1661–1662, 09 2020.

[8] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, Sep 1976.

[9] Daniel D. Lee and Hyunjune Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13 - Proceedings of the 2000 Conference, NIPS 2000*, Advances in Neural Information Processing Systems. Neural information processing systems foundation, January 2001. 14th Annual Neural Information Processing Systems Conference, NIPS 2000 ; Conference date: 27-11-2000 Through 02-12-2000.

[10] Silas Maniatis, Tarmo Äijö, Sanja Vickovic, Catherine Braine, Kristy Kang, Annelie Mollbrink, Delphine Fagegaltier, Žaneta Andrusivová, Sami Saarenpää, Gonzalo Saiz-Castro, Miguel Cuevas, Aaron Watters, Joakim Lundeberg, Richard Bonneau, and Hemali Phatnani. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science*, 364(6435):89–93, 2019.

[11] F. Markley and D. Mortari. Quaternion attitude estimation using vector observations. *Journal of The Astronautical Sciences*, 48:359–380, 2000.

[12] Kristen R. Maynard, Leonardo Collado-Torres, Lukas M. Weber, Cedric Uytingco, Brianna K. Barry, Stephen R. Williams, Joseph L. Catallini, Matthew N. Tran, Zachary Besich, Madhavi Tippani, Jennifer Chew, Yifeng Yin, Joel E. Kleinman, Thomas M. Hyde, Nikhil Rao, Stephanie C. Hicks, Keri Martinowich, and Andrew E. Jaffe. Transcriptome-scale

spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, 24(3):425–436, 2021.

[13] Patrik L Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O Westholm, Mikael Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

[14] Madhavi Tippani, Heena R. Divecha, Joseph L. Catallini, Lukas M. Weber, Abby Spangler, Andrew E. Jaffe, Stephanie C. Hicks, Keri Martinowich, Leonardo Collado-Torres, Stephanie C. Page, and Kristen R. Maynard. Vistoseg: a matlab pipeline to process, analyze and visualize high-resolution histology images for visium spatial transcriptomics data. *bioRxiv*, 2021.

[15] Grace Wahba. A least squares estimate of satellite attitude. *SIAM Review*, 7(3):409–409, 1965.