# DOT: Fast Cell Type Deconvolution by Optimal Transport

**Anonymous authors**
Paper under double-blind review

## Abstract

Single-cell RNA sequencing (scRNA-seq) and spatially-resolved imaging/sequencing technologies are the current cutting edge of transcriptomics data generation in biomedical research. On one hand, scRNA-seq data brings rich high-throughput information spanning the entire transcriptome, sacrificing the structural context of the cells. On the other hand, high-resolution measurements of the spatial context of cells comes with a trade-off in throughput and coverage. Combining data from these two modalities facilitates better understanding of the development and organization of complex tissues, as well as the emerging processes and function of distinct constituent cell types within the tissue. Recent approaches focus only on the expression of genes available in both modalities. They don't incorporate other relevant and available features, especially the spatial context. We propose DOT, a novel optimization framework for assigning cell types to tissue locations, ensuring a high-quality mapping by taking into account relevant but previously neglected features of the data. Our model (i) incorporates ideas from Optimal Transport theory to exploit structural similarities in the data modalities, leveraging not only joint features but also distinct features, i.e. the spatial context, (ii) introduces scale-invariant distance functions to account for differences in the sensitivity of different measurement technologies, (iii) ensures representation of rare cell types using Nash-fairness objectives, and (iv) provides control over the abundance of cell types in the localization. We present a fast implementation based on the Frank-Wolfe algorithm and we demonstrate the effectiveness of DOT on correctly assigning cell types to spatial data coming from (i) the primary motor cortex of the mouse brain, (ii) the primary somatosensory cortex of the mouse brain, and (iii) the developing human heart.

## 1 Introduction

In biological systems, the organization of cells within the tissue, their contextual cellular programs and their response to perturbations are central to better understanding intercellular communication, emergence of function, disease progression and to eventual identification of targets for therapeutic intervention (Trapnell, 2015; Arendt et al., 2016). Cell types are distinct subpopulations of cells which are often identified by known markers and/or by data-driven techniques, most commonly clustering based on transcriptomic profiles (Kiselev et al., 2019). Single-cell RNA sequencing can profile the entire transcriptome (mRNA expression of the full range of genes) of large portions of individual (single) cells. This has made scRNA-seq an essential tool for revealing distinct cell types in complex tissues and has profoundly impacted our understanding of developmental processes and the underlying mechanisms that control cellular functions (Haghverdi et al., 2016; Papalexi & Satija, 2018; Potter, 2018; Rajewsky et al., 2020). However, scRNA-seq requires dissociation of the tissue (Lee et al., 2020), losing the information about the spatial context and relationship between cells.

Recent advancements in spatially resolved transcriptomics methods present unique opportunities for analyzing the relationships between cell types in their spatial context (Marx, 2021). Spatial transcriptomics methods measure gene expression coupled with two- or three-dimensional locations, hereafter referred to as spots, and vary in two axes: spatial resolution and gene throughput. On one hand, technologies such as Multiplexed Error-Robust Fluorescence In-Situ Hybridization (MERFISH) and In-Situ Sequencing (ISS), achieve subcellular resolution (Chen et al., 2015), but are limited to measuring up to a couple of hundred pre-selected genes. On the other hand, spatially resolved RNA sequencing, such as Visium (Ståhl et al., 2016) and Slide-seq (Rodriques et al., 2019), enable high-throughput gene profiling by capturing mRNAs in-situ at the cost of spots with the size of tens of cells. Thus, there is a trade-off between the resolution and the richness of the data.

A strategy to overcome these limitations is to combine scRNA-seq data with high resolution spatial data to map dissociated cells to spatial locations or more generally to combine it with low-resolution spatial data to estimate the composition of cell types and expression in each spot. We refer to this task as deconvolution. Alternatively, we can attempt to enrich high-resolution data by predicting the expression of unmeasured genes. As the latter requires extrapolation to various degrees, machine learning and optimization methods are better suited to the deconvolution task.

Since the initial efforts to bridge this gap (Tanevski et al., 2020) there has been an increased interest in improvement and new method development (see Section 2). However, so far the methods rely on the genes that are captured both by scRNA-seq and spatial data, either neglect the *spatial* relationships between spots in the spatial data, are not using the remaining distinctly captured genes, or come with high computation cost for large instances. Neglecting the spatial context is equivalent to assuming random placement of spots in the space, which is in contrast to the established structure-function relationship of tissues. On the other hand, considering only a subset of genes limits the applicability of these methods to cases where the two data sets share several informative genes, which might not be the case when different technologies are used for profiling, or when few genes are measured in the spatial data (e.g., in MERFISH).

We address these deficiencies by incorporating ideas from the Optimal Transport (OT) theory and adapting a Gromov-Wasserstein (GW) distance (Mémoli, 2011; Peyré et al., 2016) between scRNA-seq and spatial data. We present DOT (Fast Cell Type Deconvolution by Optimal Transport), a fast and scalable optimization framework to integrate scRNA-seq and spatial data for cell type localization by solving a multi-criteria probabilistic matching problem. We summarize the main contributions of our work as follows:

(i) We propose a novel formulation for mapping cell types from scRNA-seq to spots in spatial data by casting this problem to a multi-objective probabilistic matching problem. Our model is applicable to both high- and low-resolution spatial data, in the form of inferring membership probabilities for the former and relative abundance of cell types in the latter.

(ii) We adapt a generalization of OT with a Gromov-Wasserstein objective to leverage spatial information and to go beyond the use of genes common to the two modalities.

(iii) We introduce a scale-invariant metric based on cosine-similarity to account for differences in measurement and the scale of gene expressions in scRNA-seq and spatial data.

(iv) We present a very fast implementation for our model based on the Frank-Wolfe algorithm, ensuring scalability and efficient solvability.

## 2 Related Work

**Cell type deconvolution.** Several deconvolution methods have been proposed in recent years. While most of these models are designed specifically for low-resolution spatial data, some are also applicable to high-resolution spatial data. Elosua-Bayes et al. (2021) proposed SPOTlight, which estimates relative abundance of cell types in spots using non-negative matrix factorization regression and non-negative least squares. Robust cell type decomposition (RCTD) (Cable et al., 2021) fits a statistical model by maximum-likelihood estimation, assuming a Poisson distribution for the expression of each gene at each spot. More recently, Kleshchevnikov et al. (2022) introduced cell2location, which assumes a two-step Bayesian model for inferring cell type composition of spots.

As cell type deconvolution, particularly in the high-resolution spatial data, is inherently a multiclass classification task, classification methods, such as Random Forests (Breiman, 2001), can be used for tackling this problem. However, because of the domain-specific properties of this problem, including differences in gene coverage, resolution, measurement sensitivity, and modality-specific characteristics, tailored learning mechanisms are needed. Tangram (Biancalani et al., 2021) proposes a deep learning model to find the best placement of single cells in spots using a designed loss function and can thus carry cell type information as a byproduct. Seurat V3 workflow (Stuart et al., 2019) is a widely-used toolkit for analyzing scRNA-seq data, which offers an "anchoring" technique based on mutual nearest neighbour classifier for aligning two modalities in the PCA space.

**Optimal Transport.** Optimal Transport (OT) (Villani, 2021) is a way to match with minimal cost data points/histograms between two domains embedded in possibly different spaces using different variants of the Wasserstein distance (Santambrogio, 2015; Peyré et al., 2019). Over the past decade, OT has been applied to various machine learning problems in a wide variety of contexts, including but not limited to generative modeling (An et al., 2019; Bunne et al., 2019), Wasserstein autoencoders (Tolstikhin et al., 2018), feature aggregation (Mialon et al., 2020), generalization error prediction (Chuang et al., 2021), dataset denoising (Mémoli et al., 2019), graph matching/classification (Titouan et al., 2019), and domain adaptation (Courty et al., 2016; Fatras et al., 2021).

Recently, OT has been employed in biology applications. Tong et al. (2020) model cellular dynamics as an unbalanced dynamic transport, with the goal of transporting entities from one cross sectional measurement to the next. Schiebinger et al. (2019) use OT for studying developmental time courses to infer ancestor-descendant fates and understanding the molecular programs that guide differentiation during development by incorporating temporal information and modeling cell growth over time. Similarly, Forrow & Schiebinger (2021) employ graphical models and OT to reconstruct developmental trajectories from time courses with snapshots of cell states and lineages.

## 3 Model

**Preliminaries.** Let $\mathbb{C}$ be a set of predefined cell types (CT), derived from partitioning the reference scRNA-seq data into $|\mathbb{C}|$ user-defined clusters, and $X^{\mathrm{C}}_{c,g}$ denote the mean expression of gene $g \in \mathbb{G}^{\mathrm{C}}$ in cell type $c$. Moreover, let $\mathbb{I}$ denote the set of spots in the spatial transcriptomics (ST) data. Note that the term "spot" can refer to one or a group of cells in certain spatial contexts. Each spot $i \in \mathbb{I}$ consists of spatial coordinates $\boldsymbol{x}_i \in \mathbb{R}^2$ or $\mathbb{R}^3$ and gene expressions $X^{\mathrm{S}}_{i,g}$ for $g \in \mathbb{G}^{\mathrm{S}}$, where $\mathbb{G}^{\mathrm{S}}$ is the set of genes that are measured in the spatial data. Let $n_i$ be the given size of spot $i$. When such information is not available, we set $n_i = 1$ to compute the proportion or probability of cell types in each spot rather than computing the number of cells of each type. Further, if prior information about the expected abundance of cell types in ST is available (e.g., estimated from a matched single-cell level sample), we denote the expected abundance of cell type $c$ by $r_c$. Note that $\boldsymbol{r}$ is scaled such that $\sum_{i \in \mathbb{I}} n_i = \sum_{c \in \mathbb{C}} r_c$. For convenience, we also define $\mathbb{G} = \mathbb{G}^{\mathrm{C}} \cap \mathbb{G}^{\mathrm{S}}$ as the set of genes that are common between CT and ST. In the following, unless otherwise mentioned, vectors of gene expressions are assumed to be in the space of common genes.

To assess dissimilarity between expression vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we also introduce the distance function

$$d_{\cos}(\boldsymbol{a}, \boldsymbol{b}) \coloneqq \sqrt{1 - \cos(\boldsymbol{a}, \boldsymbol{b})}, \tag{1}$$

where $\cos(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|}\langle \boldsymbol{a}, \boldsymbol{b} \rangle$. Note that $d_{\cos}$ is convex for positive vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, and is scale-invariant, in the sense that it is indifferent to the magnitudes of the vectors. This is by design, since we want to assess dissimilarity between expression vectors regardless of the measurement sensitivities of different technologies. We also note the following important property of $d_{\cos}$ (proofs given in Appendix A).

**Proposition 1.** *Unlike cosine dissimilarity (i.e.,* $1 - \cos(\cdot, \cdot)$*),* $d_{\cos}$ *is a* ***metric*** *distance function.*

**High-level model.** Our model relies on determining a "many-to-many" mapping $\boldsymbol{Y}$ of cell types in CT to spots in ST, with $Y_{c,i}$ denoting the proportion (or probability when $n_i = 1$) of spot $i \in \mathbb{I}$ that is of cell type $c \in \mathbb{C}$. A high-quality mapping should naturally match the expression of common genes across CT and ST. We ensure this by considering the following *genomic* criteria:

(i) Expression of genes in ST spots should match expression of genes mapped to spots via $\boldsymbol{Y}$.

(ii) Centroid of cell types in CT should match the centroids in ST as determined via $\boldsymbol{Y}$.

(iii) Distribution of genes in ST should be similar to distribution of genes mapped to ST via $\boldsymbol{Y}$.

Additionally, we may incorporate prior knowledge in the form of spatial location of spots as well as expected abundance of cell types using the following *auxiliary* criteria:

(iv) Spots that are both adjacent in space and have similar expression profiles should attain similar cell type profiles.

(v) If prior information about abundance of cell types is available, abundance of cell types in ST should match with the given abundances.

The genomic objectives naturally take precedence over the auxiliary objectives, especially when a large number of genes are common between CT and ST, but the auxiliary objectives are useful when the common genes are limited. Note that objective (v) is meant to provide additional control over abundance of cell types in the spatial data, but can be ignored if prior information about the abundance of cell types is not available. We elaborate on these objectives in the following.

**Formulation.** Objective (i) ensures that the vector of gene expressions in spot $i \in \mathbb{I}$ (i.e., $\boldsymbol{X}^{\mathrm{S}}_{i,:}$) is most similar to the vector of gene expressions mapped to spot $i$ through $\boldsymbol{Y}$ (i.e., $\sum_{c \in \mathbb{C}} Y_{c,i} \boldsymbol{X}^{\mathrm{C}}_{c,:}$). To achieve this objective, we minimize dissimilarity between these vectors by using

$$d_i(\boldsymbol{Y}) \coloneqq d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{i,:}, \sum\nolimits_{c \in \mathbb{C}} Y_{c,i} \boldsymbol{X}^{\mathrm{C}}_{c,:}\right). \tag{2}$$

Objective (ii) is in nature similar to objective (i). Here, we would like to minimize dissimilarity between centroid of cell type $c \in \mathbb{C}$ in CT (i.e., $\boldsymbol{X}^{\mathrm{C}}_{c,:}$) and centroid of cell type $c$ in ST as determined via $\boldsymbol{Y}$ (i.e., $\frac{1}{\rho_c} \sum_{i \in \mathbb{I}} Y_{c,i} \boldsymbol{X}^{\mathrm{S}}_{i,:}$). Given the scale-invariance property of $d_{\cos}$, we can drop $1/\rho_c$ and measure the dissimilarity between these centroids using the following distance function

$$d_c(\boldsymbol{Y}) \coloneqq d_{\cos}\left(\boldsymbol{X}^{\mathrm{C}}_{c,:}, \rho_c^{-1} \sum\nolimits_{i \in \mathbb{I}} Y_{c,i} \boldsymbol{X}^{\mathrm{S}}_{i,:}\right) = d_{\cos}\left(\boldsymbol{X}^{\mathrm{C}}_{c,:}, \sum\nolimits_{i \in \mathbb{I}} Y_{c,i} \boldsymbol{X}^{\mathrm{S}}_{i,:}\right). \tag{3}$$

Our goal in objective (iii) is to match distribution of expression of gene $g \in \mathbb{G}$ in ST (i.e., $\boldsymbol{X}^{\mathrm{S}}_{:,g}$) with the one mapped to ST through $\boldsymbol{Y}$ (i.e., $\sum_{c \in \mathbb{C}} \boldsymbol{Y}_{c,:} X^{\mathrm{C}}_{c,g}$). Hence, we minimize dissimilarity between these vectors by using

$$d_g(\boldsymbol{Y}) \coloneqq d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{:,g}, \sum\nolimits_{c \in \mathbb{C}} \boldsymbol{Y}_{c,:} X^{\mathrm{C}}_{c,g}\right). \tag{4}$$

To achieve objective (iv), we borrow ideas from Optimal Transport theory and the Gromov-Wasserstein metric. Let $\boldsymbol{M}^{\mathrm{C}}$ and $\boldsymbol{M}^{\mathrm{S}}$ be metrics in CT and ST, respectively, in that $M^{\mathrm{C}}_{c,k}$ defines distance between cell types $c$ and $k$, while $M^{\mathrm{S}}_{i,j}$ defines distance between spots $i$ and $j$. Note that these distances are defined for each dataset independently; hence, we can use the entire features in each set: the entire genome in CT, including the genes not measured in ST, and the uncommon/common genes as well as the spatial coordinates in ST (see Section 4 for how these matrices are computed). The *2-Gromov-Wasserstein* distance (Mémoli, 2011) between CT and ST for given mapping $\boldsymbol{Y}$, denoted $d_{\mathrm{GW}}(\boldsymbol{Y})$, is defined in equation 5. Minimizing $d_{\mathrm{GW}}(\boldsymbol{Y})$ ensures that similar pair of spots in ST (with respect to their locations and expressions) are not assigned to dissimilar pair of cell types in CT, and vice versa.

$$d_{\mathrm{GW}}(\boldsymbol{Y}) \coloneqq \sqrt{\sum\nolimits_{i \in \mathbb{I}} \sum\nolimits_{j \in \mathbb{I}} \sum\nolimits_{c \in \mathbb{C}} \sum\nolimits_{k \in \mathbb{C}} \left(M^{\mathrm{C}}_{c,k} - M^{\mathrm{S}}_{i,j}\right)^2 Y_{c,i} Y_{k,j}} \tag{5}$$

Let $\rho_c \coloneqq \sum_{i \in \mathbb{I}} Y_{c,i}$ denote the abundance of cell type $c$ in ST as determined by mapping $\boldsymbol{Y}$. As noted by Solomon et al. (2016), we may simplify equation 5 as stated in Proposition 2 below.

**Proposition 2.** *Define parameter* $\bar{m}_i \coloneqq \sum_{j \in \mathbb{I}} (M^S_{i,j})^2 n_j$ *and auxiliary variables* $\bar{m}_c \coloneqq \sum_{k \in \mathbb{C}} (M^C_{c,k})^2 \rho_k$ *and* $\boldsymbol{Z} \coloneqq \boldsymbol{M}^C \boldsymbol{Y} \boldsymbol{M}^S$. *GW distance function in equation 5 is equivalent to*

$$d_{GW}(\boldsymbol{Y}) = \sqrt{\sum\nolimits_{c \in \mathbb{C}} \sum\nolimits_{i \in \mathbb{I}} Y_{c,i} (\bar{m}_c + \bar{m}_i - 2 Z_{c,i})}, \tag{6}$$

Objective (v) provides optional control over abundance of cell types mapped to ST, when prior information about expected abundance of cell types is available. We employ Jensen-Shannon divergence between $\boldsymbol{\rho}$ and $\boldsymbol{r}$ to measure their dissimilarity

$$d_{\mathrm{A}}(\boldsymbol{Y}) \coloneqq \frac{1}{2} D_{\mathrm{KL}}\left(\boldsymbol{\rho} \middle\| \frac{\boldsymbol{\rho} + \boldsymbol{r}}{2}\right) + \frac{1}{2} D_{\mathrm{KL}}\left(\boldsymbol{r} \middle\| \frac{\boldsymbol{\rho} + \boldsymbol{r}}{2}\right), \tag{7}$$

where $D_{\mathrm{KL}}(\boldsymbol{p} \| \boldsymbol{q}) = \sum_j p_j \log(p_j / q_j)$ denotes the Kullback–Leibler divergence (Manning & Schutze, 1999). In addition, to avoid overfitting, we may require that all cell types are at least minimally represented in the mapping. To achieve this goal, we define

$$d_{\mathrm{R}}(\boldsymbol{Y}) \coloneqq -\sum\nolimits_{c \in \mathbb{C}} \log(\rho_c) = D_{\mathrm{KL}}(\bar{\boldsymbol{r}} \| \boldsymbol{\rho}), \tag{8}$$

where $\bar{r}_c = 1$ for all $c \in \mathbb{C}$. Equation 8 is a Nash fairness (Caragiannis et al., 2019) objective whose logarithmic form ensures presence of all cell types (i.e., $\rho_c > 0$).

**Algorithm 1:** Frank-Wolfe algorithm for DOT

```
Initialization: Setup distance matrices $M^{\mathrm{C}}$ and $M^{\mathrm{S}}$.
Set $t = 0$ and find an initial map $Y^{(0)}$ (see Appendix B.1).
while not converged do
    Compute gradient $\Delta^{(t)} = \nabla_{Y} f(Y^{(t)})$ (see Appendix B.2)
    for each spot $i \in \mathbb{I}$ do
        Find current best cell type $\hat{c} = \arg\min_{c \in \mathbb{C}} \{\Delta^{(t)}_{c,i}\}$
        Compute atom solution $\hat{Y}^{(t)}_{\hat{c},i} = n_i$ and $\hat{Y}^{(t)}_{c,i} = 0$ for $c \neq \hat{c}$
    Update $Y^{(t+1)} = Y^{(t)} + \frac{2}{2+t}(\hat{Y}^{(t)} - Y^{(t)})$
    $t \leftarrow t + 1$
```

We treat these criteria as objectives in a multi-objective optimization problem and to achieve them simultaneously (i.e., produce a Pareto-optimal solution), we optimize $\boldsymbol{Y}$ against a linear combination of these objectives as formulated below, hereafter referred to DOT model:

$$\min \quad \sum_{i \in \mathbb{I}} n_i d_i(\boldsymbol{Y}) + \lambda_{\mathrm{C}} \sum_{c \in \mathbb{C}} \rho_c d_c(\boldsymbol{Y}) + \lambda_{\mathrm{G}} \sum_{g \in \mathbb{G}} d_g(\boldsymbol{Y}) + \lambda_{\mathrm{GW}} d_{\mathrm{GW}}(\boldsymbol{Y}) + \lambda_{\mathrm{A}} d_{\mathrm{A}}(\boldsymbol{Y}) + \lambda_{\mathrm{R}} d_{\mathrm{R}}(\boldsymbol{Y}) \tag{9}$$

$$\text{w.r.t.} \quad \boldsymbol{Y} \in \mathbb{R}_+^{|\mathbb{C}| \times |\mathbb{I}|}, \boldsymbol{\rho} \in \mathbb{R}^{|\mathbb{C}|} \tag{10}$$

$$\text{s.t.} \quad \sum\nolimits_{c \in \mathbb{C}} Y_{c,i} = n_i \qquad \forall i \in \mathbb{I}, \tag{11}$$

$$\sum\nolimits_{i \in \mathbb{I}} Y_{c,i} = \rho_c \qquad \forall c \in \mathbb{C}, \tag{12}$$

where $\lambda_{\mathrm{C}}$, $\lambda_{\mathrm{G}}$, $\lambda_{\mathrm{GW}}$, $\lambda_{\mathrm{A}}$ and $\lambda_{\mathrm{R}}$ are the user-defined penalty weights, and coefficients $n_i$ and $\rho_c$ in equation 9 balance the scales of deviations in spots and cell types, respectively.

**Remark 1.** *Unlike the conventional OT formulations, DOT does not require the cell type abundances in ST (i.e., $\boldsymbol{\rho}$) to be strictly equal to their expected abundances (i.e., $\boldsymbol{r}$), and rather penalizes their deviation in the objective function.*

## 4 Algorithm

We propose a solution to the DOT model based on the Frank-Wolfe (FW) algorithm (Frank & Wolfe, 1956; Jaggi, 2013), which is a first-order method for solving non-linear optimization problems of the form $\min_{\boldsymbol{x} \in \mathbb{X}} f(\boldsymbol{x})$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a (potentially non-convex) continuously differentiable function over the convex and compact set $\mathbb{X}$. FW operates by replacing the non-linear objective function $f$ with its linear approximation $\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x}^{(0)}) + \nabla_{\boldsymbol{x}} f(\boldsymbol{x}^{(0)})^\top (\boldsymbol{x} - \boldsymbol{x}^{(0)})$ at a trial point $\boldsymbol{x}^{(0)} \in \mathbb{X}$, and solve a simpler problem $\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{X}} \tilde{f}(\boldsymbol{x})$ to produce an "atom" solution $\hat{\boldsymbol{x}}$. The algorithm then iterates by taking a convex combination of $\boldsymbol{x}^{(0)}$ and $\hat{\boldsymbol{x}}$ to produce the next trial point $\boldsymbol{x}^{(1)}$, which remains feasible thanks to convexity of $\mathbb{X}$. The FW algorithm is described in Algorithm 1, in which $f(\boldsymbol{Y})$ is the objective function in equation 9.

**Distance matrices.** Distance matrices $\boldsymbol{M}^{\mathrm{C}}$ and $\boldsymbol{M}^{\mathrm{S}}$ incorporate the features that are not shared across CT and ST. To compute $M^{\mathrm{C}}_{c,k}$, we calculate the dissimilarity between the centroids of cell types $c$ and $k$ considering all genes in CT (i.e., $\boldsymbol{X}^{\mathrm{C}}_{c,:} = (X^{\mathrm{C}}_{c,g})_{g \in \mathbb{G}^{\mathrm{C}}}$ for each $c \in \mathbb{C}$)

$$\boldsymbol{M}^{\mathrm{C}}_{c,k} = d_{\cos}(\boldsymbol{X}^{\mathrm{C}}_{c,:}, \boldsymbol{X}^{\mathrm{C}}_{k,:}).$$

The matrix $\boldsymbol{M}^{\mathrm{S}}$ captures the dissimilarity of ST spots in terms of their locations and expressions. Let $\boldsymbol{D}^1_{i,j}$ and $\boldsymbol{D}^2_{i,j}$ represent distance of spots $(i, j)$ with respect to their locations and expressions, respectively, as defined below:

$$D^1_{i,j} = \mathbf{1}_{\text{condition}} \left( \|\boldsymbol{x}_i - \boldsymbol{x}_j\| > \bar{d} \right), \qquad D^2_{i,j} = d_{\cos} \left( \boldsymbol{X}^{\mathrm{S}}_{i,:}, \boldsymbol{X}^{\mathrm{S}}_{j,:} \right),$$

where $\bar{d}$ is a given distance threshold, and $D^2_{i,j}$ is computed with respect to all genes in ST (i.e., $\mathbb{G}^{\mathrm{S}}$). Finally, we take $\boldsymbol{M}^{\mathrm{S}}$ to be the average of $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$:

$$\boldsymbol{M}^{\mathrm{S}} = (\boldsymbol{D}^1 + \boldsymbol{D}^2)/2 \tag{13}$$

**Remark 2.** *$\boldsymbol{M}^C$ is a metric in the domain of CT since $d_{\cos}$ is a metric. $\boldsymbol{M}^S$ is a metric in the domain of ST, since both $\boldsymbol{D}^1$ and $\boldsymbol{D}^2$ are metrics.*

To see why this definition of $\boldsymbol{M}^{\mathrm{S}}$ makes sense, we first note that cell types, by definition, are distinct subpopulations in the scRNA-seq data. Therefore, it is reasonable to assume that their centroids are dissimilar (i.e., $M_{c,k} \approx 1$ for $c \neq k$). This yields the following result.

**Proposition 3.** *Let $\alpha = \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}(1-M^S_{i,j})^2 n_i n_j$. Assuming that cell types are relatively distinct, so that $M^C_{c,k} \approx 1$, for $c, k \in \mathbb{C}$, $c \neq k$, then*

$$d_{GW}(\boldsymbol{Y}) \approx \sqrt{\alpha + \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\left(2M^S_{i,j} - 1\right)\langle \boldsymbol{Y}_{:,i}, \boldsymbol{Y}_{:,j}\rangle}$$

**Remark 3.** *Observe that $\langle \boldsymbol{Y}_{:,i}, \boldsymbol{Y}_{:,j}\rangle$ measures similarity between cell type profiles of spots $i$ and $j$. Therefore, $d_{GW}$ (i) rewards $\langle \boldsymbol{Y}_{:,i}, \boldsymbol{Y}_{:,j}\rangle$ when $2M^S_{i,j} - 1 \approx +1$ (i.e., encourages adjacent spots to attain similar cell types if their expressions are similar) and (ii) penalizes $\langle \boldsymbol{Y}_{:,i}, \boldsymbol{Y}_{:,j}\rangle$ when $2M^S_{i,j} - 1 \approx -1$ is close to +1 (i.e., prevents distant spots from attaining similar cell types if their expressions are different). Moreover, (iii) $d_{GW}$ is indifferent to pair $(i, j)$ when $2M^S_{i,j} - 1 \approx 0$ (i.e., if $i$ and $j$ are distant or different in expressions, but not both).*

**Producing an atom solution.** While the DOT model is not separable, its linear approximation can be decomposed to $|\mathbb{I}|$ independent subproblems, one for each spot $i \in \mathbb{I}$. This is because, unlike conventional OT formulations, we do not require the distribution of cell types (i.e., $\boldsymbol{\rho}$) to be equal to their expected distribution (i.e., $\boldsymbol{r}$), but have penalized their deviations in the objective function using $d_A$ equation 7. The subproblem $i$ then becomes

$$\min\ \left\{\langle \boldsymbol{Y}_{:,i}, \boldsymbol{\Delta}^{(t)}_{:,i}\rangle : \boldsymbol{Y}_{:,i} \in \mathbb{R}^{|\mathbb{C}|}_{+}, \sum_{c\in\mathbb{C}} Y_{c,i} = n_i\right\}$$

which, in turn, is a simple sorting problem. This property of Algorithm 1 enables it to efficiently tackle problems with large number of spots in the spatial data.

**Convergence.** Under suitable conditions, FW converges to an optimal solution in linear rate when optimizing a convex function over a polytope domain (Jaggi & Lacoste-Julien, 2015). Given the non-convex objective function in equation 9, Algorithm 1 instead obtains a first-order stationary point at a rate of $O(1/\sqrt{t})$ (Bertsekas, 2016; Wai et al., 2017). We numerically assess the convergence of Algorithm 1 at iteration $t$ using the so-called "FW-gap" (Jaggi, 2013)

$$\delta^{(t)} := \sum_{i\in\mathbb{I}}\sum_{c\in\mathbb{C}}(Y^{(t)}_{c,i} - \hat{Y}^{(t)}_{c,i})\Delta^{(t)}_{c,i}.$$

We also implemented acceleration techniques such as averaging gradients (Zhang et al., 2021b) and away steps (Jaggi & Lacoste-Julien, 2015; Garber & Meshi, 2016), but did not observe materially different convergence rates than the vanilla FW Algorithm 1.

# 5 Practical Enhancements

In this section, we introduce practical enhancements to formulation equation 9–equation 12 to incorporate the domain-specific application of this formulation.

## 5.1 Cell Type Heterogeneity

While cell types are distinct subpopulations of cells, significant variations may naturally exist within each cell type. This means, a single vector $\boldsymbol{X}^{\mathrm{C}}_{c,:}$ may not properly represent the distribution of cells within this cell type. Consequently, mapping cell types solely based on the centroid of cell types can be error-prone. To capture the intrinsic heterogeneity of cell types, we cluster each cell type into predefined $\kappa$ smaller groups using an unsupervised learning method, and produce a total of $\kappa|\mathbb{C}|$ centroids to replace the original $|\mathbb{C}|$ centroids. With this definition of centroids, we treat all terms except $d_{\mathrm{A}}$ and $d_{\mathrm{R}}$ as before. For $d_{\mathrm{A}}$ and $d_{\mathrm{R}}$, since prior information about cell types (and not sub-clusters) are available, we keep $\boldsymbol{\rho}$ to represent the abundance of original cell types by setting $\rho_c = \sum_{k\in\mathbb{K}_c}\sum_{i\in\mathbb{I}} Y_{k,i}$, where $\mathbb{K}_c$ denotes the set of sub-clusters of cell type $c$. Finally, once $\boldsymbol{Y}$ is obtained, $\sum_{k\in\mathbb{K}_c} Y_{k,i}$ determines probability that spot $i$ is of cell type $c$.

### 5.2 Sparse Mapping

As previously discussed, spatial data are either high-resolution (single-cell level) or low-resolution (multicell level). In the case of high-resolution spatial data, given that each spot corresponds to an individual cell (i.e., $n_i = 1$), it is desirable to produce sparse allocations, in the sense that we prefer $Y_{c,i}$ close to 0 or 1. In general, assuming that $Y_{c,i} \in \{0, n_i\}$, then equation 11 implies that $Y_{c,i} = n_i$ for exactly one cell type $c$ and is zero for all other cell types. Consequently, for binary $\boldsymbol{Y}$ we obtain

$$d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{i,:}, \sum\nolimits_{c\in\mathbb{C}} Y_{c,i}\boldsymbol{X}^{\mathrm{C}}_{c,:}\right) = \frac{1}{n_i}\sum\nolimits_{c\in\mathbb{C}} Y_{c,i} d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{i,:}, \boldsymbol{X}^{\mathrm{C}}_{c,:}\right),$$

which is linear in $\boldsymbol{Y}$. As linear objectives promote sparse (or corner point) solutions, we may control the level of sparsity of the mapping by introducing a parameter $\theta \in [0, 1]$ and redefining $d_i(\boldsymbol{Y})$ as

$$d_i(\boldsymbol{Y}) =(1-\theta)d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{i,:}, \sum\nolimits_{c\in\mathbb{C}} Y_{c,i}\boldsymbol{X}^{\mathrm{C}}_{c,:}\right) + \theta\frac{1}{n_i}\sum\nolimits_{c\in\mathbb{C}} Y_{c,i} d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{i,:}\boldsymbol{X}^{\mathrm{C}}_{c,:}\right). \tag{14}$$

Note that a higher value for $\theta$ yields a sparser solution. Indeed, with $\theta = 1$ and zero weights assigned to other objectives, the optimal mapping will be completely binary.

## 6 Results

We compared the performance of our method, abbreviated `DOT`, against five state of the art models in the literature: `SPOTlight` (Elosua-Bayes et al., 2021), `RCTD` (Cable et al., 2021), `cell2location` (Kleshchevnikov et al., 2022), `Tangram` (Biancalani et al., 2021), and `Seurat` Stuart et al. (2019). In Appendix C, we describe the choice of parameters for these models as well as the metrics we used for evaluating these models. We performed experiments on data coming from (i) the primary motor cortex of the mouse brain, (ii) the primary somatosensory cortex of the mouse brain, and (iii) the developing human heart, specifics of which are presented in Appendix D.

### 6.1 Results on High-Resolution Spatial Data

For our first set of experiments, we use the high-resolution MERFISH spatial data of the primary motor cortex region (MOp) of the mouse brain (Zhang et al., 2021a) as detailed in Appendix D.1. Our goal with this experiment is to evaluate the performance of different models in determining the probability distribution of cell types at each spot. Since the identity of the cell type represented by the spot is known in the MERFISH data, we can use this information as ground-truth when evaluating the performance of the different models. In addition to deconvolution methods, as baseline algorithms, we compared the methods with `SingleR` (Aran et al., 2019), which is suited for single-cell resolution data. Given the multiclass classification nature of this task, we also used `RF` (Breiman, 2001) as a multiclass classifier baseline.

The MERFISH MOp dataset contains the spatial information of 280,186 cells across 75 samples. With each sample, we created a reference scRNA-seq data using all the 280,186 cells, except the cells contained in the sample, and the 254 genes to estimate the centroids of the 99 reference cell types. We further created 12 spatial datasets for each sample (i.e., a total of 1125 spatial datasets) as follows. To simulate the effect of number of shared features between the spatial and scRNA-seq data, we assumed that only a subset of the 254 genes are available in the spatial data by selecting the first $|\mathbb{G}|$ genes, where $|\mathbb{G}| \in \{50, 75, 100, 125, 150\}$ (i.e., 20%, 30%, 40%, 50%, 60% of genes). Moreover, to simulate the effect of differences in measurement sensitivities of different technologies, we introduced random noise in the spatial data by multiplying the expression of gene $g$ in spot $i$ by $1 + \beta_{i,g}$, where $\beta_{i,g} \sim U(-\varphi, \varphi)$ with $\varphi \in \{0, 0.25, 0.5\}$.

We compare the predictive performance of `DOT` to `Seurat`, `RCTD`, `Tangram`, `SingleR` and `RF` in Figure 1. We removed `SPOTlight` and `C2L` from these plots due to their clear under-performance in the high resolution spatial data. We observe that `DOT` not only dominates the three alternatives in assigning correct cell types to the spots (Figure 1a), but also produces well-calibrated probabilities (Figure 1b) and better captures the relationships between cell types in space (Figure 1c), owing to its capacity to incorporate the spatial information in $d_{\mathrm{GW}}$ through the distance matrices. We also observe that even with very few genes in common between the spatial data and the reference scRNA-seq data (e.g., $|\mathbb{G}| \leq 75$), `DOT` is able to reliably determine the cell type of spots in the space with high accuracy. In contrast, `RCTD` fails to produce results due to lack of shared information, and `Seurat` and `Tangram` produce results with low accuracy. Additionally, we observe that `DOT` is more immune to fluctuations in expressions in the spatial data, implying the effectiveness of our $d_{\cos}$ distance function in accounting for differences in measurement scales of different technologies.

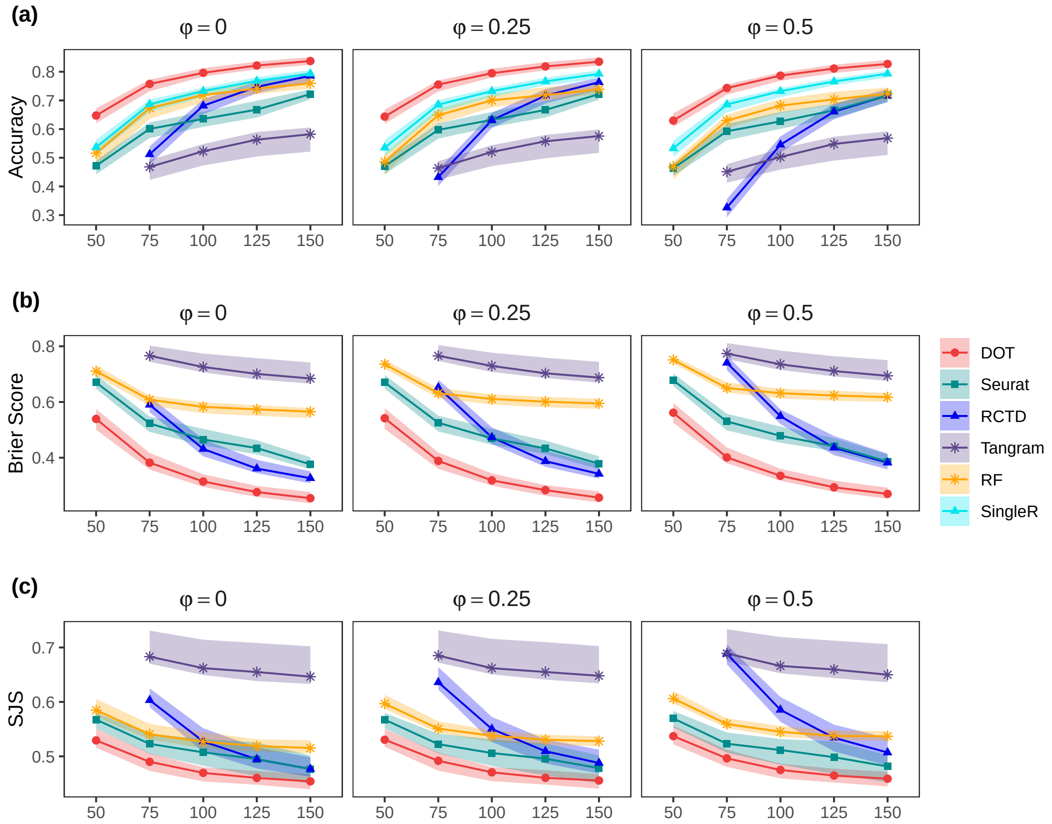


Figure 1: Predictive performance of the algorithms in the high-resolution spatial data across different number of genes in the spatial data ($x$-axis) and different noise factors ($\varphi$). Points represent the median of 75 values, and the shaded areas correspond to their interquartile interval.

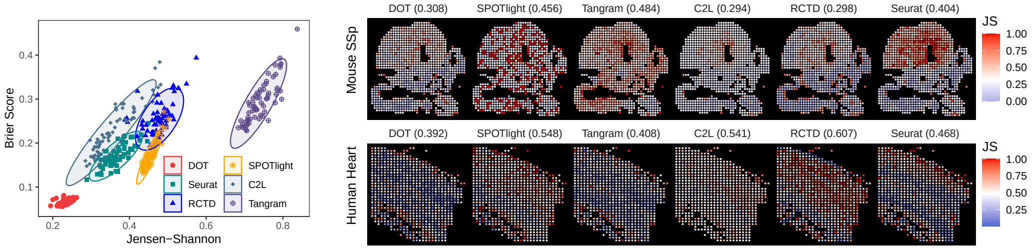
In terms of algorithmic performance (Table 1), `DOT` takes on average 441 seconds to solve each instance, which is an order of magnitude faster than `RCTD`, `Tangram`, and `RF`, and is comparable to `Seurat` and `SingleR`.

### 6.2 Results on Low-Resolution Spatial Data

We now evaluate the performance of models on low-resolution spatial data. For these experiments, since there is no ground truth for real multicell spatial data such as Visium and Slide-seq, we resort to producing ground truth multicell spatial data by pulling the adjacent cells in the high resolution spatial data of primary motor cortex of the mouse brain (MOp), primary somatosensory cortex of the mouse brain (SSp), and the developing human heart. Figure 3 in Appendix E illustrates a sample low-resolution spatial data obtained from a MERFISH MOp tissue. Note that, unlike the high-resolution spatial data, the ground truth $P_{c,i}$ now corresponds to relative abundance of cell type $c$ in spot $i$. We can therefore assess the performance of each model by comparing the probability distributions $\boldsymbol{P}_{:,i}$ and the estimated probabilities (i.e., $\boldsymbol{Y}_{:,i}$) using Brier Score or Jensen-Shannon metrics.

#### 6.2.1 Experiments on the Mouse MOp

To produce ground truth for MOp, using the common subclass annotations between MERFISH MOp and scRNA-seq MOp (Yao et al., 2021b) (see Appendix D.1), for each of the 75 MERFISH MOp samples, we randomly assigned each cell in the MERFISH MOp data to a cell in the scRNA-seq MOp data of the same subclass. Next, we lowered the resolution of spatial data by splitting each sample into regular grids of length 100µm to mimic the size and inter-distance of spots in multicell spatial transcriptomics, such as Visium. Finally, we aggregated the expression profiles of cells within each tile as the expression profile of the respective spots.



(a) Mouse MOp

(b) Mouse SSp (top) and developing human heart (bottom)

Figure 2: Predictive performance of the algorithms in the low-resolution spatial data: (a) Performance of models across 75 samples of MOp, with each point in the scatter plots denoting the average performance across all spots in the sample. (b) Distribution of performance of models on each individual spot in the SSp and human heart samples.

Figure 2a compares the performance of `DOT` against `RCTD`, `SPOTlight`, `C2L`, `Tangram` and `Seurat` in determining the cell type composition of the multicell spots. We observe that `DOT` outperforms other models with respect to both metrics. As presented in Table 1, `DOT` took on average 457 seconds to solve an instance, which proved to be more than twice faster than `Seurat`, and orders of magnitude faster than `RCTD`, `SPOTlight`, `C2L` and `Tangram`, further highlighting the superiority of `DOT` in terms of both accuracy and computational efficiency.

| Experiment | Resolution | Instances | **DOT** | **Seurat** | **RCTD** | **Tangram** | **SPOTlight** | **C2L** | **SingleR** | **RF** |
|---|---|---|---|---|---|---|---|---|---|---|
| MOp | High | 1125 | 441 | 380 | 4748 | 10141 | 7884 | 3310 | 303 | 7427 |
| MOp | Low | 75 | 457 | 1086 | 4705 | 8250 | 52825 | 6119 | — | — |
| SSp | Low | 1 | 4 | 21 | 117 | 248 | 705 | 364 | — | — |
| Heart | Low | 1 | 8 | 11 | 185 | 88 | 316 | 398 | — | — |

Table 1: Average computation times (in seconds) of different models across different experiments.

### 6.2.2 Experiments on the Mouse SSp and the Developing Human Heart

We also carried out experiments on data from the SSp region of mouse brain as well as the developing human heart to evaluate the performance of models on tissues of different structures. We employed single-cell level spatial data coming from osmFISH technology (Codeluppi et al., 2018) to produce multicell data for SSp (Appendix D.2). For the developing human heart, we used subcellular spatial data generated by the ISS platform (Asp et al., 2019) (Appendix D.3). We tested the performance of the six deconvolution methods on these two samples, results of which are illustrated in Figure 2b. Each subplot shows the distribution of prediction error based on the Jensen-Shannon divergence at each spot in the spatial data, with the average value over all spots given on top of each plot. `DOT` outperforms other models in the human heart sample and is among the best-performing models in the mouse SSp sample. Moreover, performance of `DOT` is not sensitive to different regions/cell type of the tissue (compare to `Tangram` and `Seurat` in SSp and `RCTD` in human heart). These results further highlight the competitive performance of `DOT` and its robustness in identifying the cell type composition of spots across different tissues.

## 7 Conclusion

Single-cell RNA-seq and spatially-resolved imaging/sequencing technologies, the cutting edge technologies in transcriptomic data generation, each provide a partial picture in understanding the organization of complex tissues. To obtain a full picture, computational methods aim at combining data from these two modalities. We present DOT, a fast and scalable optimization framework based on Optimal Transport theory, for assigning cell types to tissue locations by leveraging the spatial information as well as both joint and distinct genes across scRNA-seq and spatial data. Using experiments on data from mouse brain and human heart, we show that DOT predicts the cell type composition of spots in spatial data with high accuracy, outperforming the state of the art methods both in terms of predictive performance and computation time.

## 8 Reproducibility Statement

All datasets used in this study are publicly available, details of which are given in Appendix D. We implemented the methods according to the guidelines provided in the respective studies. Performance metrics, implementation details as well as computational considerations for the methods used in this study are provided in Appendix C. Implementation details of the FW algorithm are given in Section 4 and Appendix B. An anonymized downloadable source code is provided as supplementary material.

## References

Dongsheng An, Yang Guo, Na Lei, Zhongxuan Luo, Shing-Tung Yau, and Xianfeng Gu. Ae-ot: a new generative model based on extended semi-discrete optimal transport. In *International Conference on Learning Representations (ICLR)*, 2019.

Dvir Aran, Agnieszka P Looney, Leqian Liu, Esther Wu, Valerie Fong, Austin Hsu, Suzanna Chak, Ram P Naikawadi, Paul J Wolters, Adam R Abate, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology*, 20(2):163–172, 2019.

Detlev Arendt et al. The origin and evolution of cell types. *Nat. Rev. Genet.*, 17(12):744–757, 2016.

Michaela Asp et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell*, 179(7):1647–1660, 2019.

Dimitri P. Bertsekas. *Nonlinear programming*. Athena Scientific, 2016. ISBN 978-1-886529-05-2.

Tommaso Biancalani, Gabriele Scalia, Lorenzo Buffoni, Raghav Avasthi, Ziqing Lu, Aman Sanger, Neriman Tokcan, Charles R Vanderburg, Åsa Segerstolpe, Meng Zhang, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nature methods*, 18(11):1352–1362, 2021.

Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, 2001.

Charlotte Bunne, David Alvarez-Melis, Andreas Krause, and Stefanie Jegelka. Learning generative models across incomparable spaces. In *International Conference on Machine Learning*, pp. 851–861. PMLR, 2019.

Dylan M Cable et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.*, pp. 1–10, 2021.

Ioannis Caragiannis et al. The unreasonable fairness of maximum Nash welfare. *ACM Transactions on Economics and Computation (TEAC)*, 7(3):1–32, 2019.

Kok Hao Chen et al. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 348(6233), 2015.

Ching-Yao Chuang, Youssef Mroueh, Kristjan Greenewald, Antonio Torralba, and Stefanie Jegelka. Measuring generalization with optimal transport. *Advances in Neural Information Processing Systems*, 34:8294–8306, 2021.

Simone Codeluppi et al. Spatial organization of the somatosensory cortex revealed by osmfish. *Nature methods*, 15(11):932–935, 2018.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016.

Elosua-Bayes et al. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.*, 49(9):e50–e50, 2021.

Kilian Fatras et al. Unbalanced minibatch optimal transport; applications to domain adaptation. In *International Conference on Machine Learning*, pp. 3186–3197. PMLR, 2021.

Aden Forrow and Geoffrey Schiebinger. LineageOT is a unified framework for lineage tracing and trajectory inference. *Nat. Commun.*, 12(1):1–10, 2021.

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Res. Logis. Quart.*, 3(1-2):95–110, 1956.

Dan Garber and Ofer Meshi. Linear-memory and decomposition-invariant linearly convergent conditional gradient algorithm for structured polytopes. *Advances in neural information processing systems*, 29, 2016.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Laleh Haghverdi et al. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods*, 13(10):845–848, 2016.

Martin Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pp. 427–435. PMLR, 2013.

Martin Jaggi and Simon Lacoste-Julien. On the global linear convergence of Frank-Wolfe optimization variants. *Advances in Neural Information Processing Systems*, 28, 2015.

Vladimir Yu Kiselev, Tallulah S Andrews, and Martin Hemberg. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.*, 20(5):273–282, 2019.

Vitalii Kleshchevnikov, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology*, 40(5):661–671, 2022.

Jeongwoo Lee, Do Young Hyeon, and Daehee Hwang. Single-cell multiomics: technologies and data analysis methods. *Exp. Mol. Med.*, 52(9):1428–1442, 2020.

Christopher Manning and Hinrich Schutze. *Foundations of statistical natural language processing*. MIT press, 1999.

Vivien Marx. Method of the year: spatially resolved transcriptomics. *Nat. Methods*, 18(1):9–14, 2021.

Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.

Facundo Mémoli, Zane Smith, and Zhengchao Wan. The wasserstein transform. In *International Conference on Machine Learning*, pp. 4496–4504. PMLR, 2019.

Grégoire Mialon, Dexiong Chen, Alexandre d'Aspremont, and Julien Mairal. A trainable optimal transport embedding for feature aggregation. In *International Conference on Learning Representations (ICLR)*, 2020.

Efthymia Papalexi and Rahul Satija. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.*, 18(1):35–45, 2018.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-Wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pp. 2664–2672. PMLR, 2016.

Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Found. Trends Mach. Learn.*, 11(5-6):355–607, 2019.

S Steven Potter. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.*, 14(8):479–492, 2018.

Nikolaus Rajewsky et al. LifeTime and improving European healthcare through cell-based interceptive medicine. *Nature*, 587(7834):377–386, 2020.

Samuel G Rodriques et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

Geoffrey Schiebinger et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Justin Solomon et al. Entropic metric alignment for correspondence problems. *ACM Transactions on Graphics (TOG)*, 35(4):1–13, 2016.

Patrik L Ståhl et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

Tim Stuart et al. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902, 2019.

Jovan Tanevski et al. Gene selection for optimal prediction of cell position in tissues from single-cell transcriptomics data. *Life Science Alliance*, 3(11):e202000867, 2020.

Vayer Titouan et al. Optimal Transport for structured data with application on graphs. In *International Conference on Machine Learning*, pp. 6275–6284. PMLR, 2019.

Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. In *International Conference on Learning Representations (ICLR)*, 2018.

Alexander Tong et al. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In *International Conference on Machine Learning*, pp. 9526–9536. PMLR, 2020.

Cole Trapnell. Defining cell types and states with single-cell genomics. *Genome Res.*, 25(10): 1491–1498, 2015.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Hoi-To Wai et al. Decentralized Frank–Wolfe algorithm for convex and nonconvex problems. *IEEE Trans. Automat. Contr.*, 62(11):5522–5537, 2017.

Marvin N. Wright and Andreas Ziegler. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1):1–17, 2017.

Zizhen Yao et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*, 184(12):3222–3241, 2021a.

Zizhen Yao et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature*, 598(7879):103–110, 2021b.

Meng Zhang et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature*, 598(7879):137–143, 2021a.

Yilang Zhang, Bingcong Li, and Georgios B Giannakis. Accelerating frank-wolfe with weighted average gradients. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5529–5533. IEEE, 2021b.

# A Proofs

**Proof of Proposition 2.**

*Proof.* Let $g(\boldsymbol{Y}) = \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\sum_{c\in\mathbb{C}}\sum_{k\in\mathbb{C}}\left(M^{\mathsf{C}}_{c,k} - M^{\mathsf{S}}_{i,j}\right)^2 Y_{c,i}Y_{k,j}$, then

$$g(\boldsymbol{Y}) = \sum_{c\in\mathbb{C}}\sum_{i\in\mathbb{I}} Y_{c,i} \sum_{k\in\mathbb{C}}\sum_{j\in\mathbb{I}} Y_{k,j}\left((M^{\mathsf{C}}_{c,k})^2 + (M^{\mathsf{S}}_{i,j})^2 - 2M^{\mathsf{C}}_{c,k}M^{\mathsf{S}}_{i,j}\right).$$

Expanding the inner summations:

$$\sum_{k\in\mathbb{C}}\sum_{j\in\mathbb{I}} Y_{k,j}(M^{\mathsf{C}}_{c,k})^2 = \sum_{k\in\mathbb{C}}(M^{\mathsf{C}}_{c,k})^2\sum_{j\in\mathbb{I}} Y_{k,j} = \sum_{k\in\mathbb{C}}(M^{\mathsf{C}}_{c,k})^2\rho_k = \bar{m}_c \text{ (using} \sum_{j\in\mathbb{I}} Y_{k,j} = \rho_k)$$

$$\sum_{k\in\mathbb{C}}\sum_{j\in\mathbb{I}} Y_{k,j}(M^{\mathsf{S}}_{i,j})^2 = \sum_{j\in\mathbb{I}}(M^{\mathsf{S}}_{i,j})^2\sum_{k\in\mathbb{C}} Y_{k,j} = \sum_{j\in\mathbb{I}}(M^{\mathsf{S}}_{i,j})^2 n_j = \bar{m}_i \text{ (using} \sum_{k\in\mathbb{C}} Y_{k,j} = n_i)$$

$$\sum_{k\in\mathbb{C}}\sum_{j\in\mathbb{I}} Y_{k,j}M^{\mathsf{C}}_{c,k}M^{\mathsf{S}}_{i,j} = \sum_{k\in\mathbb{C}} M^{\mathsf{C}}_{c,k}\left(\boldsymbol{Y}\boldsymbol{M}^{\mathsf{S}}\right)_{k,i} = \left(\boldsymbol{M}^{\mathsf{C}}\boldsymbol{Y}\boldsymbol{M}^{\mathsf{S}}\right)_{c,i} = Z_{c,i}$$

We can therefore rewrite $g(\boldsymbol{Y})$ as:

$$g(\boldsymbol{Y}) = \sum_{c\in\mathbb{C}}\sum_{i\in\mathbb{I}} Y_{c,i}(\bar{m}_c + \bar{m}_i - 2Z_{c,i}).$$

□

**Proof of Proposition 1.**

*Proof.* Note that

$$\left\|\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} - \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\right\|^2 = \frac{\|\boldsymbol{a}\|^2}{\|\boldsymbol{a}\|^2} + \frac{\|\boldsymbol{b}\|^2}{\|\boldsymbol{b}\|^2} - 2\frac{\langle\boldsymbol{a},\boldsymbol{b}\rangle}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} = 2 - 2\cos(\boldsymbol{a},\boldsymbol{b})$$

$$\Rightarrow d_{\cos}(\boldsymbol{a},\boldsymbol{b}) = \sqrt{1-\cos(\boldsymbol{a},\boldsymbol{b})} = \sqrt{2}\|\boldsymbol{a}/\|\boldsymbol{a}\| - \boldsymbol{b}/\|\boldsymbol{b}\|\|.$$

This shows that $d_{\cos}$ is a metric since $\|\cdot\|$ is a metric. We can easily show that cosine dissimilarity (i.e., $1-\cos(\cdot,\cdot)$) is not a metric. For instance, consider $\boldsymbol{a} = (1,0,0)$, $\boldsymbol{b} = (0,1,0)$ and $\boldsymbol{c} = (x, x, \sqrt{1-2x^2})$ for arbitrary $x \in (\frac{1}{2}, \frac{1}{\sqrt{2}})$, and let $f$ denote the cosine dissimilarity. Then $f(\boldsymbol{a},\boldsymbol{b}) = 1 - \cos(\boldsymbol{a},\boldsymbol{b}) = 1$, and $f(\boldsymbol{a},\boldsymbol{c}) = f(\boldsymbol{c},\boldsymbol{b}) = 1-x$, which violates the triangular inequality since $f(\boldsymbol{a},\boldsymbol{c}) + f(\boldsymbol{c},\boldsymbol{b}) = 2-2x < 1 = f(\boldsymbol{a},\boldsymbol{b})$. It is not difficult to see that $d_{\cos}(\boldsymbol{a},\boldsymbol{c}) + d_{\cos}(\boldsymbol{c},\boldsymbol{b}) > 1$. □

**Proof of Proposition 3.**

*Proof.* Let $g(\boldsymbol{Y}) = \sum\limits_{i\in\mathbb{I}}\sum\limits_{j\in\mathbb{I}}\sum\limits_{c\in\mathbb{C}}\sum\limits_{k\in\mathbb{C}}\left(M^{\mathsf{C}}_{c,k} - M^{\mathsf{S}}_{i,j}\right)^2 Y_{c,i}Y_{k,j}$. Provided that $M^{\mathsf{C}}_{c,k} = 1$ for $c \neq k$ and $M^{\mathsf{C}}_{c,c} = 0$, we obtain

$$\begin{aligned}
g(\boldsymbol{Y}) =& \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\sum_{c\in\mathbb{C}}\left(\left(M^{\mathsf{S}}_{i,j}\right)^2 Y_{c,i}Y_{c,j} + \sum_{k\in\mathbb{C},k\neq c}\left(1-M^{\mathsf{S}}_{i,j}\right)^2 Y_{c,i}Y_{k,j}\right)\\
=& \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\sum_{c\in\mathbb{C}}\left(\left(\left(M^{\mathsf{S}}_{i,j}\right)^2 - \left(1-M^{\mathsf{S}}_{i,j}\right)^2\right) Y_{c,i}Y_{c,j} + \sum_{k\in\mathbb{C}}\left(1-M^{\mathsf{S}}_{i,j}\right)^2 Y_{c,i}Y_{k,j}\right)\\
=& \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\sum_{c\in\mathbb{C}}\left(2M^{\mathsf{S}}_{i,j}-1\right)Y_{c,i}Y_{c,j} + \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\sum_{c\in\mathbb{C}}\sum_{k\in\mathbb{C}}\left(1-M^{\mathsf{S}}_{i,j}\right)^2 Y_{c,i}Y_{k,j}\\
=& \sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{I}}\left(2M^{\mathsf{S}}_{i,j}-1\right)\langle\boldsymbol{Y}_{:,i},\boldsymbol{Y}_{:,j}\rangle + \alpha,
\end{aligned}$$

where $\alpha = \sum\limits_{i\in\mathbb{I}}\sum\limits_{j\in\mathbb{I}}\left(1-M^{\mathsf{S}}_{i,j}\right)^2\sum\limits_{c\in\mathbb{C}}\sum\limits_{k\in\mathbb{C}} Y_{c,i}Y_{k,j} = \sum\limits_{i\in\mathbb{I}}\sum\limits_{j\in\mathbb{I}}\left(1-M^{\mathsf{S}}_{i,j}\right)^2 n_i n_j$ since $\sum\limits_{c\in\mathbb{C}} Y_{c,i} = n_i$ and $\sum\limits_{k\in\mathbb{C}} Y_{k,j} = n_j$. □

# B Implementation Details of the FW Algorithm

## B.1 Initial Solution

A good quality initial solution plays a critical role in fast convergence of FW. Given the multi-objective nature of our model, we produce an initial solution as convex combination of two solutions. In the first solution, for each spot $i$ we first find cell type $\hat{c} = \arg\min_{c\in\mathbb{C}}\{d_{\cos}\left(\boldsymbol{X}^{\mathrm{S}}_{i,:}, \boldsymbol{X}^{\mathrm{C}}_{c,:}\right)\}$ and set $Y_{c,i} = n_i$ if $c = \hat{c}$ and $Y_{c,i} = 0$ otherwise. Note that this solution is optimal for the sparse case when $d_i$ is the only objective. In the second solution, we simply set $Y_{c,i} = n_i\rho_c/|\mathbb{I}|$ for each $i$ and $c$. Note that this solution is optimal for $d_{\mathrm{A}}$. We then set the initial solution as the convex combination of these two solutions, with 0.9 weight assigned to the first solution.

## B.2 Derivatives

To find the derivatives of $d_i(\boldsymbol{Y})$ and $d_c(\boldsymbol{Y})$, defined in equation 2 and equation 3, we introduce auxiliary quantities $\bar{\boldsymbol{X}}^{\mathrm{S}} := \boldsymbol{Y}^\top\boldsymbol{X}^{\mathrm{C}}$ and $\bar{\boldsymbol{X}}^{\mathrm{C}} := \boldsymbol{Y}\boldsymbol{X}^{\mathrm{S}}$, to denote the expressions mapped through $\boldsymbol{Y}$ to spots and cell types, respectively. Derivatives for $d_i(\boldsymbol{Y})$ and $d_c(\boldsymbol{Y})$ can then be calculated as:

$$\frac{\partial d_i}{\partial Y_{c,i}} = \frac{1}{\|\boldsymbol{X}^{\mathrm{S}}_{i,:}\|}\langle \boldsymbol{X}^{\mathrm{C}}_{c,:}, \boldsymbol{T}^{\mathrm{S}}_{i,:}\rangle, \qquad \frac{\partial d_c}{\partial Y_{c,i}} = \frac{1}{\|\boldsymbol{X}^{\mathrm{C}}_{c,:}\|}\langle \boldsymbol{X}^{\mathrm{S}}_{i,:}, \boldsymbol{T}^{\mathrm{C}}_{c,:}\rangle,$$

where

$$\begin{aligned} T^{\mathrm{S}}_{i,g} =& \frac{-1}{2d_i(\boldsymbol{Y})}\left(\frac{X^{\mathrm{S}}_{i,g}}{\|\bar{\boldsymbol{X}}^{\mathrm{S}}_{i,:}\|} - \frac{\bar{X}^{\mathrm{S}}_{i,g}}{\|\bar{\boldsymbol{X}}^{\mathrm{S}}_{i,:}\|^3}\langle \boldsymbol{X}^{\mathrm{S}}_{i,:}, \bar{\boldsymbol{X}}^{\mathrm{S}}_{i,:}\rangle\right), \\ T^{\mathrm{C}}_{c,g} =& \frac{-1}{2d_c(\boldsymbol{Y})}\left(\frac{X^{\mathrm{C}}_{c,g}}{\|\bar{\boldsymbol{X}}^{\mathrm{C}}_{c,:}\|} - \frac{\bar{X}^{\mathrm{C}}_{c,g}}{\|\bar{\boldsymbol{X}}^{\mathrm{C}}_{c,:}\|^3}\langle \boldsymbol{X}^{\mathrm{C}}_{c,:}, \bar{\boldsymbol{X}}^{\mathrm{C}}_{c,:}\rangle\right). \end{aligned}$$

Derivative of $\rho_c d_c(\boldsymbol{Y})$ then can be computed using the product rule. Similarly, we may derive the derivative for $d_g(\boldsymbol{Y})$ via

$$\frac{\partial d_g}{\partial Y_{c,i}} = \frac{-1}{2d_g(\boldsymbol{Y})}\frac{X^{\mathrm{C}}_{c,g}}{\|\boldsymbol{X}^{\mathrm{S}}_{:,g}\|}\left(\frac{X^{\mathrm{S}}_{i,g}}{\|\bar{\boldsymbol{X}}^{\mathrm{S}}_{:,g}\|} - \frac{Y_{c,i}}{\|\bar{\boldsymbol{X}}^{\mathrm{S}}_{:,g}\|^3}\langle \boldsymbol{X}^{\mathrm{S}}_{:,g}, \bar{\boldsymbol{X}}^{\mathrm{S}}_{:,g}\rangle\right)$$

Taking into account the simplification from Proposition 2, noting that $\bar{m}_c$ and $Z_{c,i}$ are functions of $\boldsymbol{Y}$ while $\bar{m}_i$ is constant, we can show that

$$\frac{\partial d_{\mathrm{GW}}}{\partial Y_{c,i}} = \frac{1}{2d_{\mathrm{GW}}(\boldsymbol{Y})}(2\bar{m}_c + \bar{m}_i - 4Z_{c,i}).$$

Finally, the derivatives for $d_{\mathrm{A}}$ and $d_{\mathrm{R}}$, defined in equation 7 and equation 8 respectively, can be calculated as:

$$\frac{\partial d_{\mathrm{A}}}{\partial Y_{c,i}} = \frac{1}{2}\log\left(\frac{2\rho_c}{\rho_c + r_c}\right), \qquad \frac{\partial d_{\mathrm{R}}}{\partial Y_{c,i}} = -\frac{1}{\rho_c}.$$

# C Experimental Setup

## C.1 Parameter Setting

For `DOT`, we set penalty weights $\lambda_{\mathrm{C}} = 1$ and $\lambda_{\mathrm{G}} = |\boldsymbol{n}|/|\mathbb{G}|$ to balance the scales of different objectives, where $|\boldsymbol{n}| := \sum_{i\in\mathbb{I}} n_i$. This is because both $\sum_{i\in\mathbb{I}} n_i d_i(\boldsymbol{Y})$ and $\sum_{c\in\mathbb{C}} r_c d_c(\boldsymbol{Y})$ are in the range of 0 and $|\boldsymbol{n}|$, while $0 \leq \sum_{g\in\mathbb{G}} d_g(\boldsymbol{Y}) \leq |\mathbb{G}|$. For the GW objective, it is not difficult to verify that $0 \leq d_{\mathrm{GW}}(\boldsymbol{Y}) \leq |\boldsymbol{n}|$. However, although spatial information contributes to the accuracy of cell type mapping, meaning that $\lambda_{\mathrm{GW}} > 0$ is desirable, a large value for $\lambda_{\mathrm{GW}}$ may dominate the genomic objectives $d_i(\boldsymbol{Y})$, $d_c(\boldsymbol{Y})$ and $d_g(\boldsymbol{Y})$, thus reduce accuracy. A middle-ground is to set a small positive value for $\lambda_{\mathrm{GW}}$. In our computations, we found that $\lambda_{\mathrm{GW}} = 0.1$ works best in most cases. Whenever prior information about expected abundance of cell types is available, we set $\lambda_{\mathrm{A}} = 1$ and $\lambda_{\mathrm{R}} = 1$. We computed $\rho_c$, the expected abundance of cell type $c$, based on the observed fraction of cell type $c$ in the reference scRNA-seq data multiplied by $|\boldsymbol{n}|$. We set the sparsity parameter $\theta = 1$ for high resolution spatial data, and set $\theta = 0$ for low resolution spatial data. To capture heterogeneity of cell types, we clustered each cell type into $= 10$ clusters. The distance threshold $\bar{d}$ is computed as follows. For each spot we computed its Euclidean distance to 8 closest spots in space[1], yielding $8|\mathbb{I}|$ values. We then took $\bar{d}$ as the 99th percentile of these values.

For RCTD, SPOTlight, Tangram, and C2L we used the default parameters suggested by the authors with the following exceptions. For RCTD we set the parameter UMI_min to 50 to prevent the model from removing too many cells from the data. Given the large number of cell type in the mouse MOp datasets, for SPOTlight we reduced the number of cells per cell type to 100 to enhance the computation time. Similarly, as Tangram was not able to produce results in a reasonable time for the MOp instances, we randomly selected 500 cells per cell type to reduce the computation time. For C2L, we used 20000 epochs to balance computation performance and accuracy. For Seurat and SingleR, we followed the package documentations, with functions used with default parameters. For RF we used the implementation provided in the R package ranger (Wright & Ziegler, 2017) with all parameters set at their default values.

### C.2 Performance Metrics

We compared the predictive performance of DOT against the other methods using three metrics. *Accuracy* in the context of high-resolution spatial data (i.e., when each spot corresponds to an individual cell) is the proportion of correctly classified spots (i.e., sum of the main diagonal in the confusion matrix) over all spots. To assess the accuracy of membership probabilities produced by each model, we compared the models using *Brier Score*, also known as mean squared error:

$$\text{Brier Score} = |\mathbb{I}|^{-1} \sum\nolimits_{i \in \mathbb{I}} \sum\nolimits_{c \in \mathbb{C}} (Y_{c,i} - P_{c,i})^2,$$

where $P_{c,i} = 1$ if spot $i$ is of cell type $c$ and $P_{c,i} = 0$ otherwise, and $Y_{c,i}$ is the predicted probability that spot $i$ is of cell type $c$. As Brier Score is a strictly proper scoring rule for measuring the accuracy of probabilistic predictions (Gneiting & Raftery, 2007), a model with lower Brier Score produces better-calibrated probabilities.

Besides the cell type that each spot is annotated with, we can produce a cell type probability distribution for each spot by considering the cell type of its neighboring spots, using a Gaussian smoothing kernel of the form

$$\tilde{P}_{c,i} = (\sum\nolimits_{j \in \mathbb{I}} K_{i,j})^{-1} \sum\nolimits_{j \in \mathbb{I}} K_{i,j} P_{c,j}, \qquad K_{i,j} = \exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 / 2\sigma^2\right),$$

where $\sigma$ is the kernel width parameter which we set to $0.5\bar{d}$. Note that as spot $j$ becomes closer to spot $i$, its label contributes more to the probability distribution at spot $i$. Using these probabilities, we also introduce the *Spatial Jensen-Shannon* (SJS) divergence to compare the probability distributions assigned to spots (i.e., $\boldsymbol{Y}$) with the smoothed probabilities (i.e., $\tilde{\boldsymbol{P}}$)

$$\text{SJS} = |\mathbb{I}|^{-1} \sum\nolimits_{i \in \mathbb{I}} \text{JS}(\boldsymbol{Y}_{:,i}, \tilde{\boldsymbol{P}}_{:,i}),$$

where $\text{JS}(\boldsymbol{Y}_{:,i}, \tilde{\boldsymbol{P}}_{:,i})$ is the Jensen-Shannon divergence between probability distributions $\boldsymbol{Y}_{:,i}$ and $\tilde{\boldsymbol{P}}_{:,i}$ with base 2 logarithm (Manning & Schutze, 1999), also defined in equation 7.

## D Datasets

### D.1 Mouse Primary Motor Cortex (MOp)

**MERFISH.** For high-resolution spatial transcriptomics, we used the spatially resolved cell atlas of the MOp recently generated using multiplexed error-robust fluorescence in situ hybridization (MERFISH) technology and made publicly available by Zhang et al. (2021a). The processed dataset contains normalized RNA counts of 254 genes and coordinates of the boundaries of a total of 280,186 segmented cells across 75 samples in the MOp of two adult mice, with the number of cells within each sample ranging from 1000 to 7500 cells. We computed the $(x, y)$ coordinates of the center of each cell by taking the average of the coordinates of its boundary. The study also identifies 99 trasncriptionally distinct cell types by community detection applied on a cell similarity graph. The clustering resulted in 39 excitatory neuronal cell types (clusters), 42 inhibitory neuronal cell types, 14 non-neuronal cell types, and four other cell types.

[1] We used 8 closest neighbors to mimic the number of adjacent tiles in a 2D regular grid.

**scRNA-seq.** The corresponding scRNA-seq data comes from a cell atlas of the MOp (Yao et al., 2021b), which contains the mRNA expression of the full range of genes for more than 500,000 individual cells across several omics layers. We used the scRNA-seq dataset `scRNA_10x_v2_A` generated at the Allen Institute, which contains 145,748 cells and 100 cell types. After removing the unannotated cells and low quality cell types (as categorized in the study), we retrieved 124,330 cells and 90 distinct cell types. For computational efficiency, we also selected the top 5,000 variable genes according to their means and variances (Stuart et al., 2019).

### D.2 Mouse Primary Somatosensory Cortex (SSp)

Similar to MOp, another well-studied tissue area is the primary somatosensory cortex area (SSp). Here, we used high-resolution spatial data coming from the osmFISH platform (Codeluppi et al., 2018), which contains measurements of 33 genes across 4,837 cells, as well as annotations based on 11 major cell types. For reference scRNA-seq data with matched cell types, we used the annotations independently generated by (Yao et al., 2021a) using 5,392 single cells in the same SSp region.

### D.3 Developing Human Heart

For the developing human heart, we used subcellular spatial data generated by the ISS platform (Asp et al., 2019), which contains tissue sections from human embryonic cardiac samples collected at different times. We selected the PCW6.5 slide which contains measurements of 69 genes across 17,454 cells as well as annotations of 12 major cell types. The same study also provides scRNA-seq data for similar slide, which contains matched cell types for 3,253 cells.

## E Supplementary Results

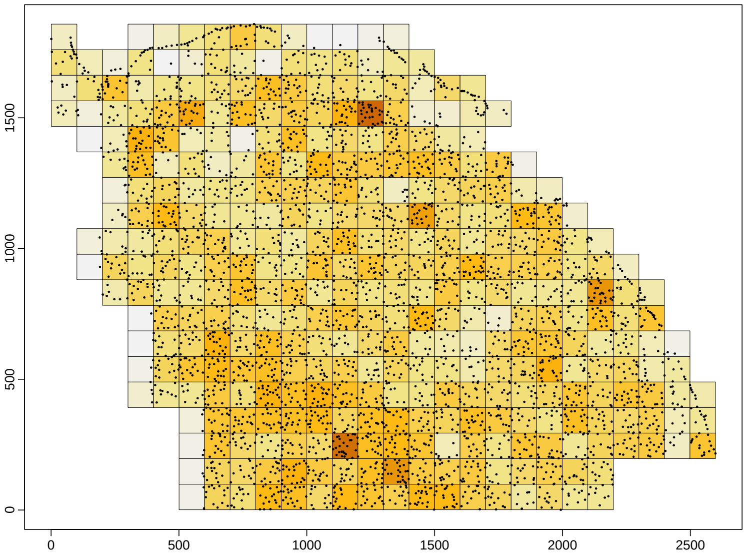


Figure 3: Synthetic multicell spatial data from MERFISH. Dots show individual cells and tiles represent multicell spots (darker tile means denser spot).