SSA: a novel method for Single-cell and Spatial transcriptomics Alignment

Presented by
Bang Tran, California State University, Sacramento

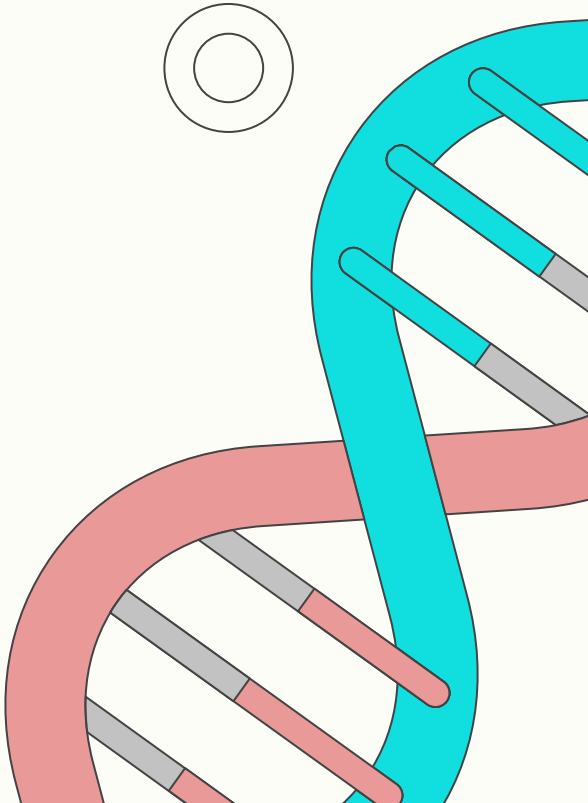
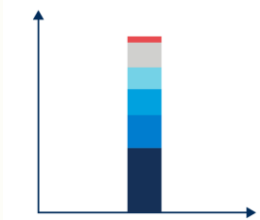


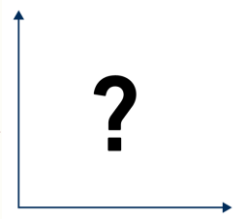
Table of contents

Single – Cell RNA Sequencing

Bulk-Cell analysis

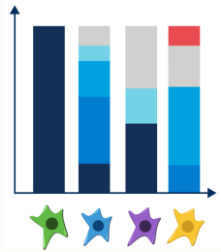


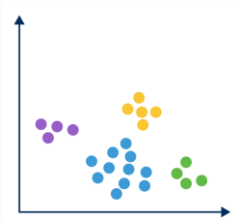
Average expression from all cells



No cellular heterogeneity revealed

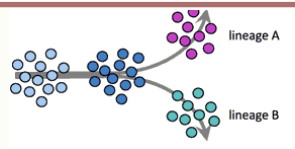
Single-Cell Analysis



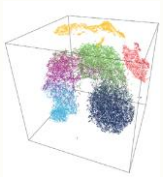


Reveal heterogeneity and subpopulation expression variability

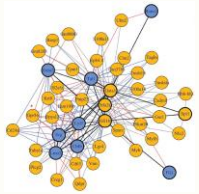
Applications



Cell lineage analysis



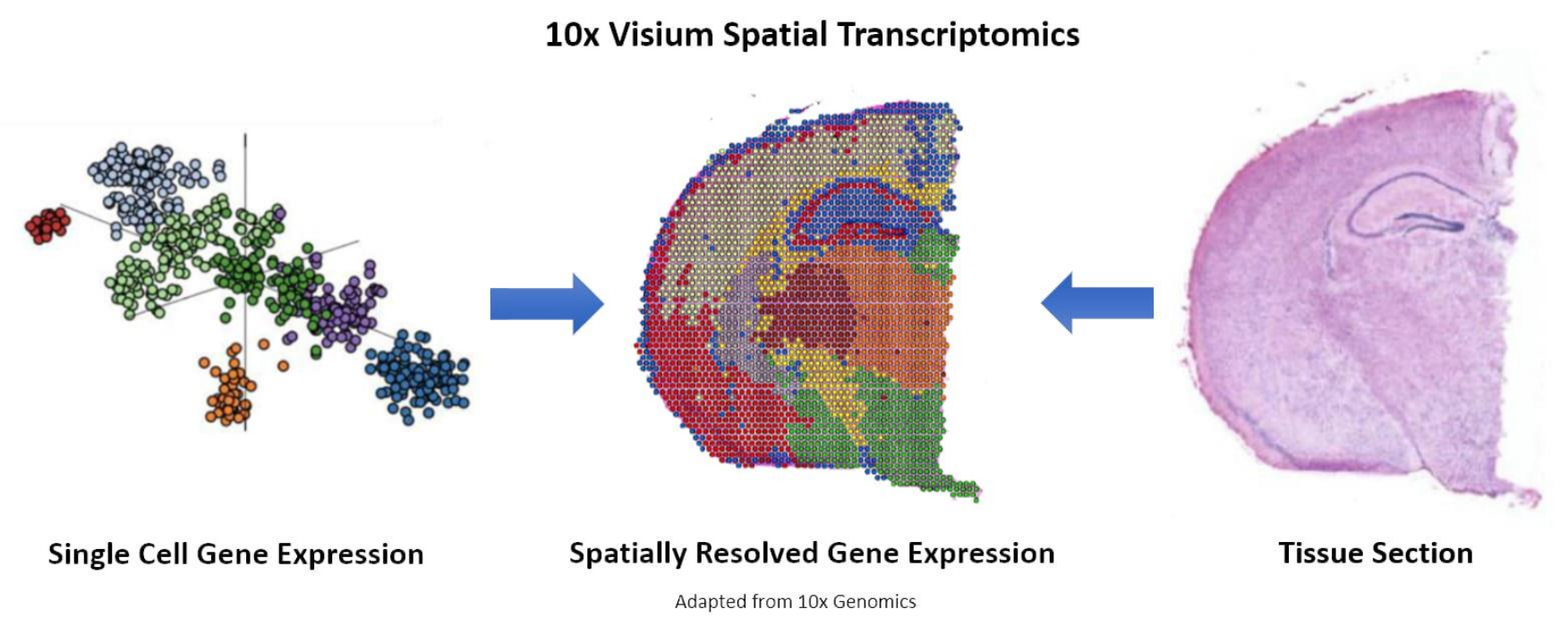
Cell type analysis



Network inference

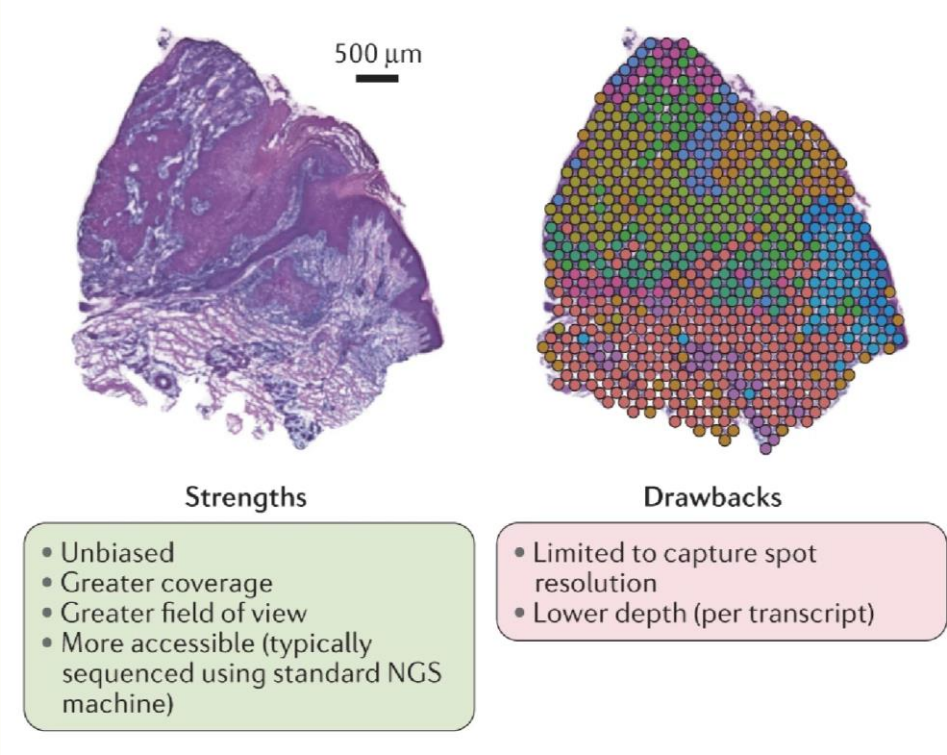
Source: 10xgenomics.com

Spatial Transcriptomics (ST)



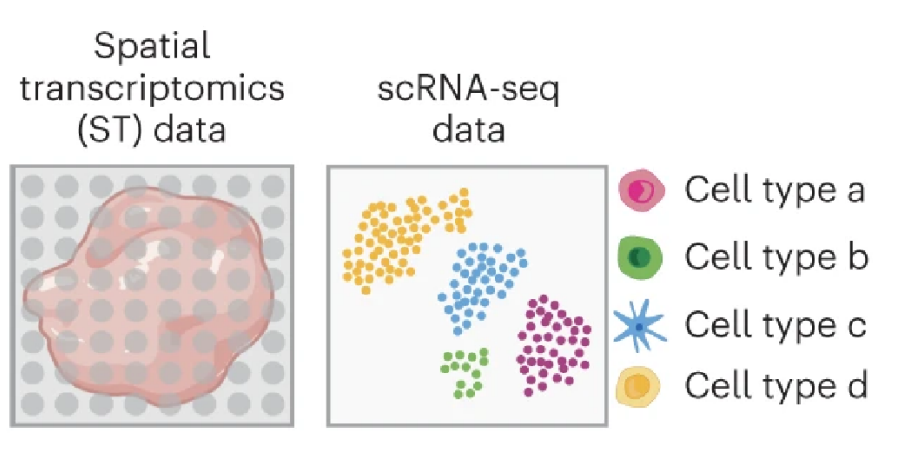
Challenges

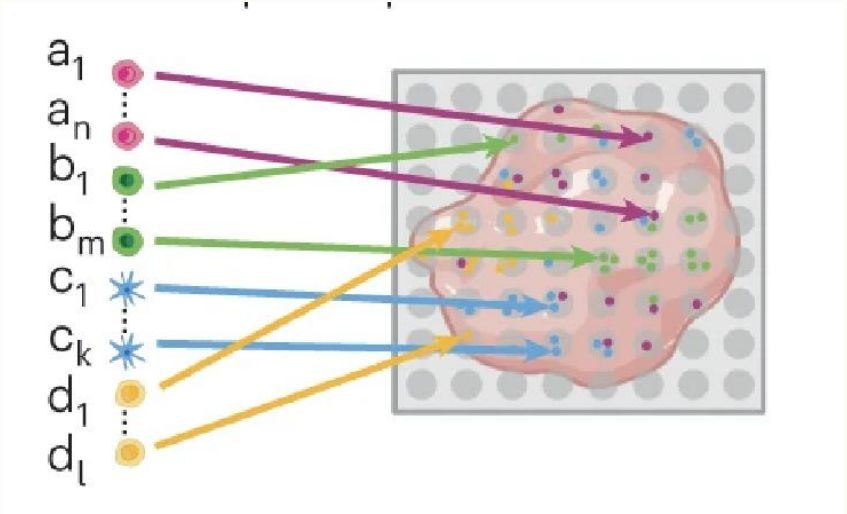
- Assays can only measure small regions with mixtures of cells
- smFISH-based techniques such as seq-FISH+ and ExM-MERFISH can only cover a small area of tissue, comprised of around 10,000 cells with a few dozens to a few hundreds of genes.
- 10x Visium platforms are restricted to a set of gene expression evaluations, in which spatial expression profile is the average expression of many cells.
- Require more specialized equipment.



https://doi.org/10.1038/s41576-021-00370-8

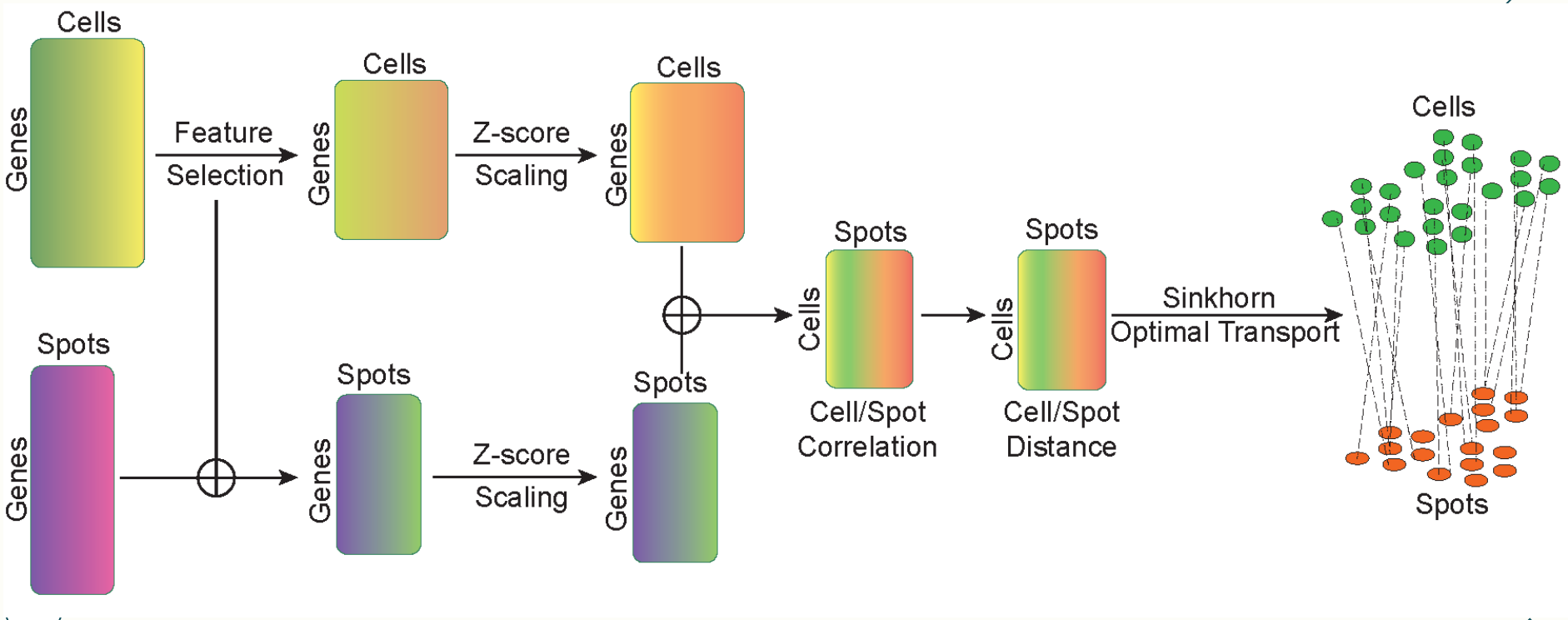
Single-cell and Spatial transcriptomics Alignment





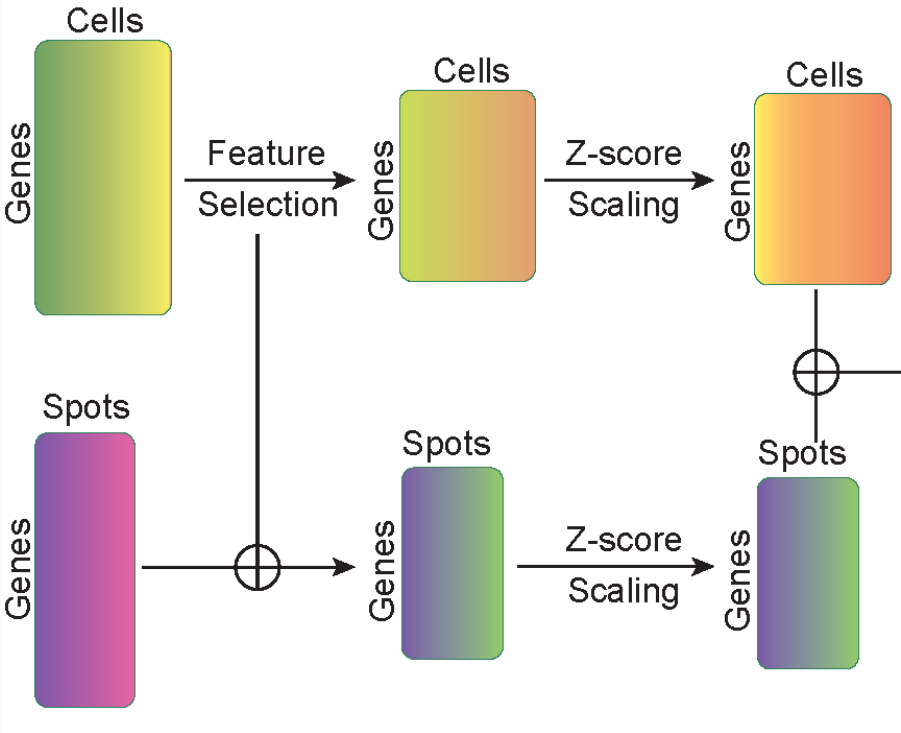
DOI: https://doi.org/10.1038/s41587-023-01697-9

SSA: a novel method for Single-cell and Spatial transcriptomics Alignment



Feature Selection and Data Transformation

- Compute the variance for each gene in the scRNA-seq data and select 5,000 genes with the highest variance.
- Subset both of the expression matrices using this gene set for consistency and comparative analysis.
- Use Z-score transformation to scale and center the data.



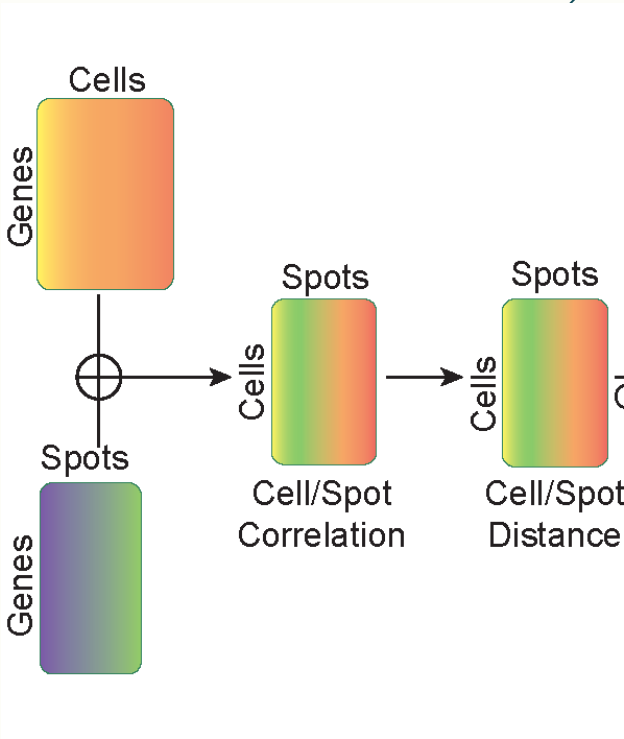
Cell to Spot Alignment using Sinkhorn Algorithm

- Calculate the pair-wise Pearson's correlation between cells and spots.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)*\sigma(Y)}$$

- Calculate the pair-wise distance between cells and spots

$$D(X,Y) = 1 - |\rho(X,Y)|$$



Cell to Spot Alignment using Sinkhorn Algorithm

1. Initialization:

- A distance matrix $D(X, Y)$ (cost matrix) that represents distances between single cells and spatial spots.
- Two probability vectors p and q that represent the expression distribution of single cells and spatial spots respectively. Here, p are row-wise sum of gene expression values for each cell in scRNA-seq data and q are row-wise sum of gene expression values for each spot in ST data.
- A regularization term $\lambda = 0.05$.
- A kernel matrix $K = e^{\frac{-\lambda * D(X,Y)}{max(D(X,Y)}}$
- Two vectors $a \in R^m$ and $b \in R^n$ with all entries equal to 1

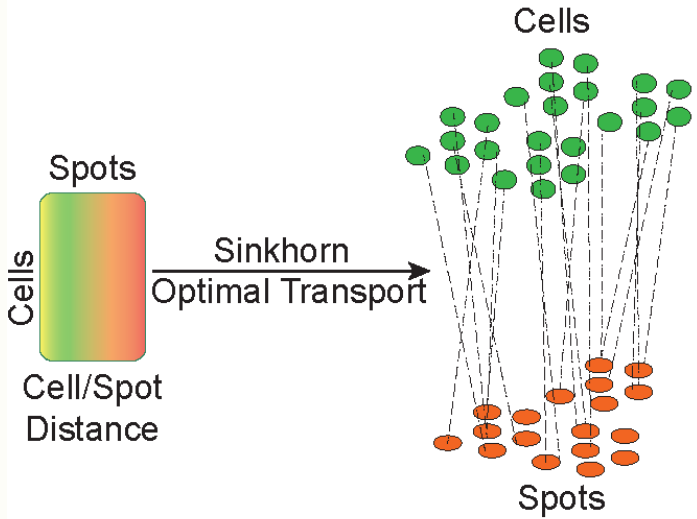
2. Update:

Repeat until convergence:

- Update: $b = \frac{q}{K^T a}$
- Update: $a = \frac{p}{Kb}$

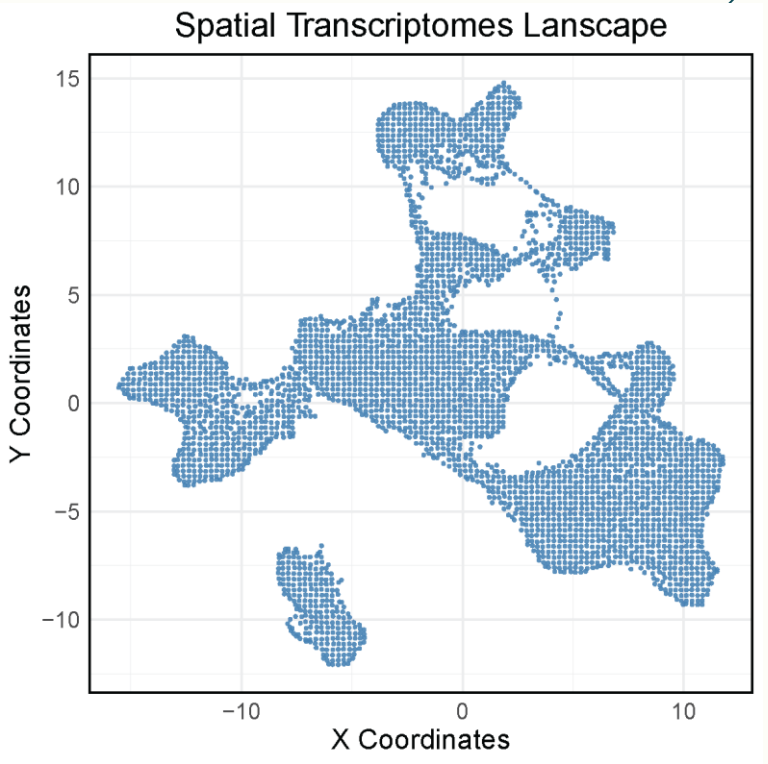
3. Compute transport plan:

$T = diag(a)Kdiag(b)$



Data Preparation

- Download high-resolution spatial transcriptomics of 100,064 cells of human breast cancer data from Gene Expression Omnibus (GEO) under accession number GSE176078.
- Partition the spatial domain into a grid structure where each grid cell, or “spot”. Centroids of cellular locations are averaging the 2-D coordinates of all cells.
- The gene expression profile for each spot is generated by averaging the gene expression levels of all cells residing within the corresponding grid cell.
- 3,615 spots and sub-sample scRNA-seq dataset 10 times



Results

Table 1: Comparisons using average Euclidean distance.

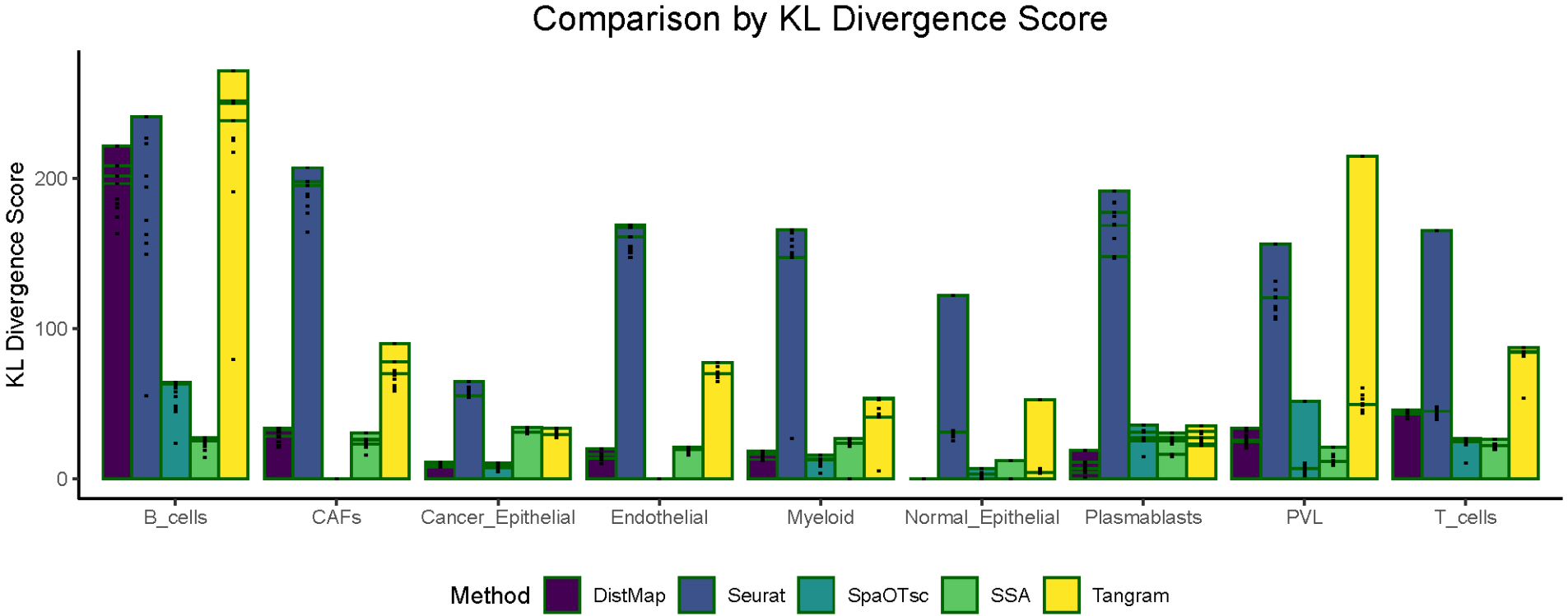
Datasets	SSA	SpaoTsc	Tangram	Seurat	DistMap
Dataset-1	**1.786**	3.007	5.364	11.627	4.624
Dataset-2	**1.802**	2.977	5.365	12.057	4.718
Dataset-3	**1.778**	2.952	5.351	11.998	4.534
Dataset-4	**1.818**	2.955	5.291	11.821	5.345
Dataset-5	**1.757**	2.970	5.357	12.072	4.519
Dataset-6	**1.738**	2.944	5.376	11.806	4.804
Dataset-7	**1.784**	2.955	5.388	12.274	4.989
Dataset-8	**1.799**	2.991	5.296	11.751	4.484
Dataset-9	**1.806**	3.076	5.351	11.961	1398.808
Dataset-10	**1.789**	2.939	5.407	12.200	4.502

Results

Table 2: Comparisons using average Manhattan distance.

Datasets	SSA	SpaoTsc	Tangram	Seurat	DistMap
Dataset-1	**2.259**	3.819	6.878	14.769	5.839
Dataset-2	**2.277**	3.781	6.874	15.309	5.967
Dataset-3	**2.245**	3.755	6.862	15.179	5.727
Dataset-4	**2.298**	3.758	6.771	14.98	6.701
Dataset-5	**2.227**	3.778	6.858	15.333	5.705
Dataset-6	**2.193**	3.747	6.895	15.096	6.040
Dataset-7	**2.252**	3.750	6.925	15.603	6.279
Dataset-8	**2.272**	3.801	6.797	14.912	5.663
Dataset-9	**2.280**	3.918	6.858	15.229	1697.566
Dataset-10	**2.263**	3.736	6.930	15.11	5.677

Results



SSA results are substantially lower than other methods across all cell types ($p = 2.838 \times 10^{-7}$ using the Wilcoxon test).

Conclusion

- We have introduced SSA for the alignment of individual cells to physical spatial locations in a tissue based on scRNA-seq and ST data.
- SSA accurately maps single-cells to spatial locations for the whole cell population with very minimal differences in distance.
- SSA efficiently reconstructs a cellular spatial map within a specific cell type.
- SSA substantially outperforms existing state-of-the-art approaches in different benchmark scenarios.
- Integrate other techniques developed in our research lab for single-cell analysis to group cells with similar patterns together before performing cells-to-spot alignment
- Extend this work to improve the data for data integration and pathway analysis, cancer research, and space biology and drug development.

Thanks!

https://tinnguyen-lab.com/home

https://webpages.csus.edu/s.tran/

NIH NATIONAL CANCER INSTITUTE
NIH NIGMS
NSF
NASA