

# A Tale of Two Spaces

Discovery of spatial biology in phenotype space via  $\Phi$ -Space ST

Jiadong Mao<sup>1</sup>, Jarny Choi<sup>2</sup> & Kim-Anh Lê Cao<sup>1</sup>

<sup>1</sup> Melbourne Integrative Genomics (MIG) & School of Mathematics and Statistics, University of Melbourne

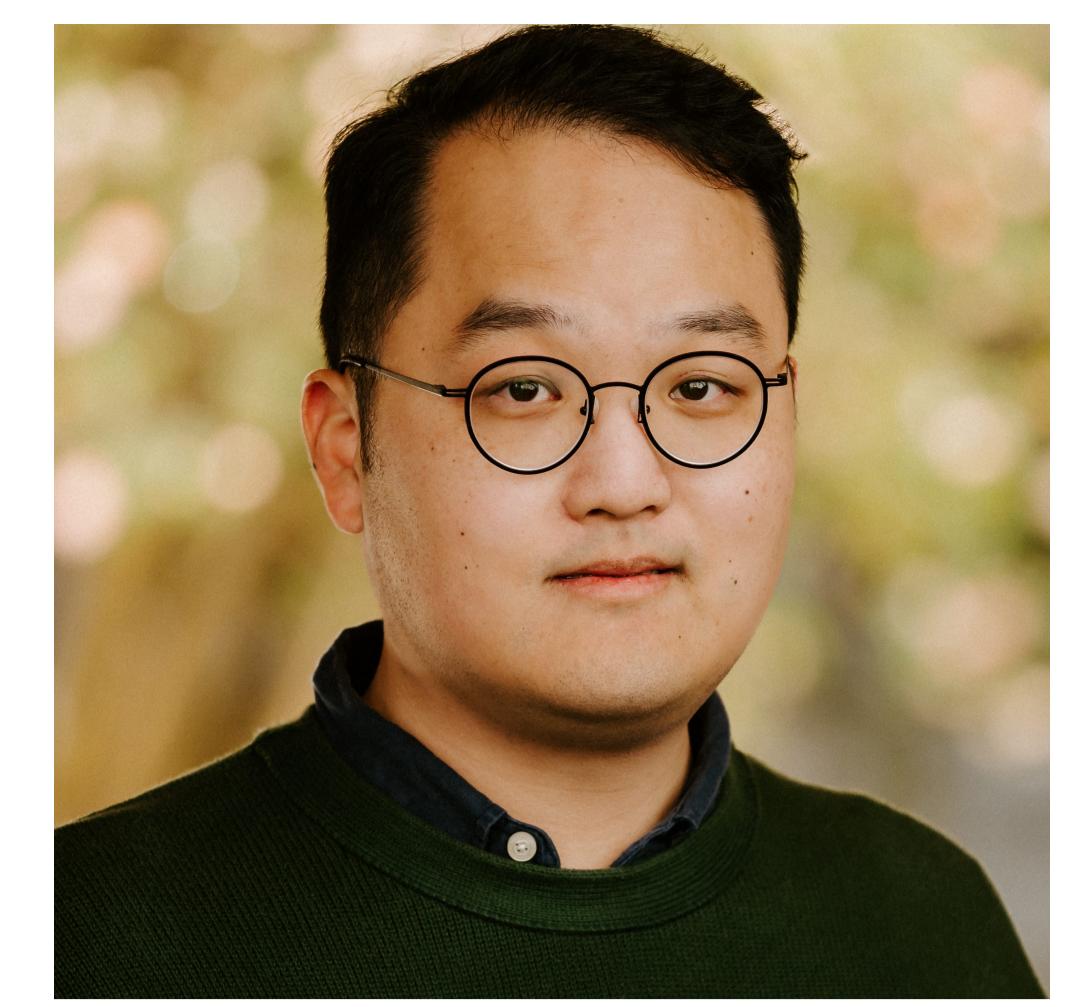
<sup>2</sup> Bioinformatics and Cellular Genomics, St Vincent's Institute



jiadong.mao@unimelb.edu.au

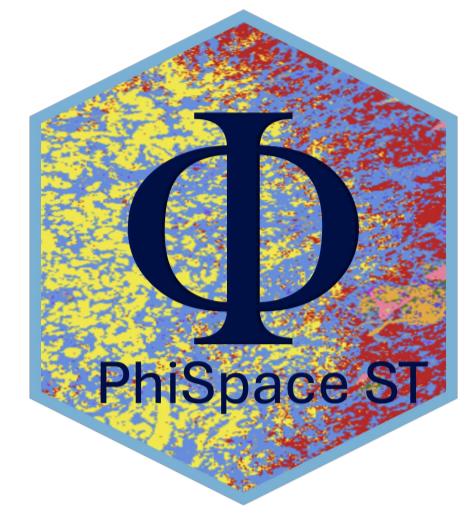


<https://github.com/jiadongm>



## Highlights

$\Phi$ -Space ST is a computationally efficient and **platform-agnostic** method for identifying cell states in spatial transcriptomics (ST) studies.



➤  $\Phi$ -Space ST annotates **continuous cell states** in ST data produced by a wide range of platforms based on scRNA-seq references.



➤  $\Phi$ -Space ST achieves **cell type deconvolution** based on **multiple reference datasets** without the need to first integrate them.

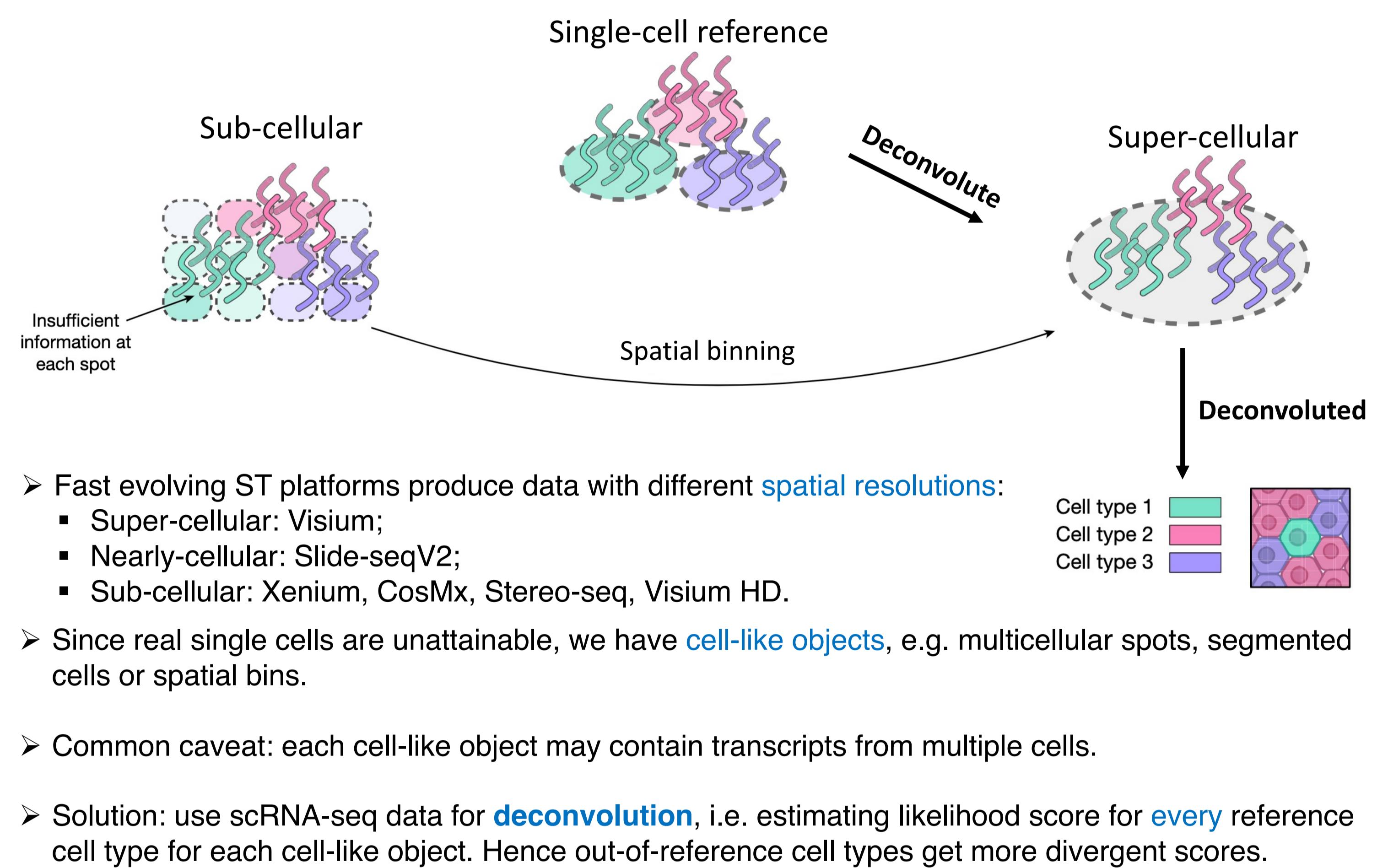
➤  $\Phi$ -Space ST provides highly interpretable annotation of cell states, including **out-of-reference ones**.

➤  $\Phi$ -Space ST is based on partial least squares (PLS) regression. It is nonparametric, computationally efficient and hence flexible to extend to emerging **spatial multiomics**.

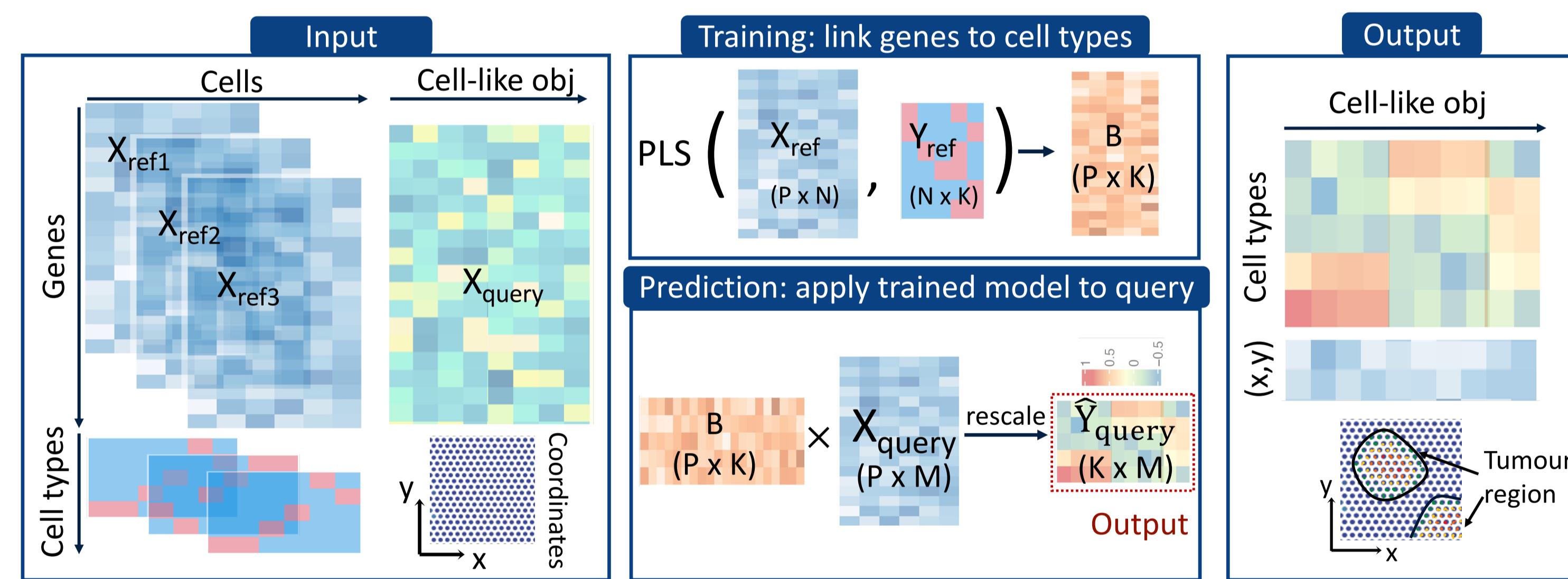
➤ The  $\Phi$ -Space R package is available on GitHub, and manuscript on bioRxiv:

Mao, J., Choi, J., & Le Cao, K.-A. (2025).  $\Phi$ -Space ST: a platform-agnostic method to identify cell states in spatial transcriptomics studies. *bioRxiv*.

## Background



## Methods

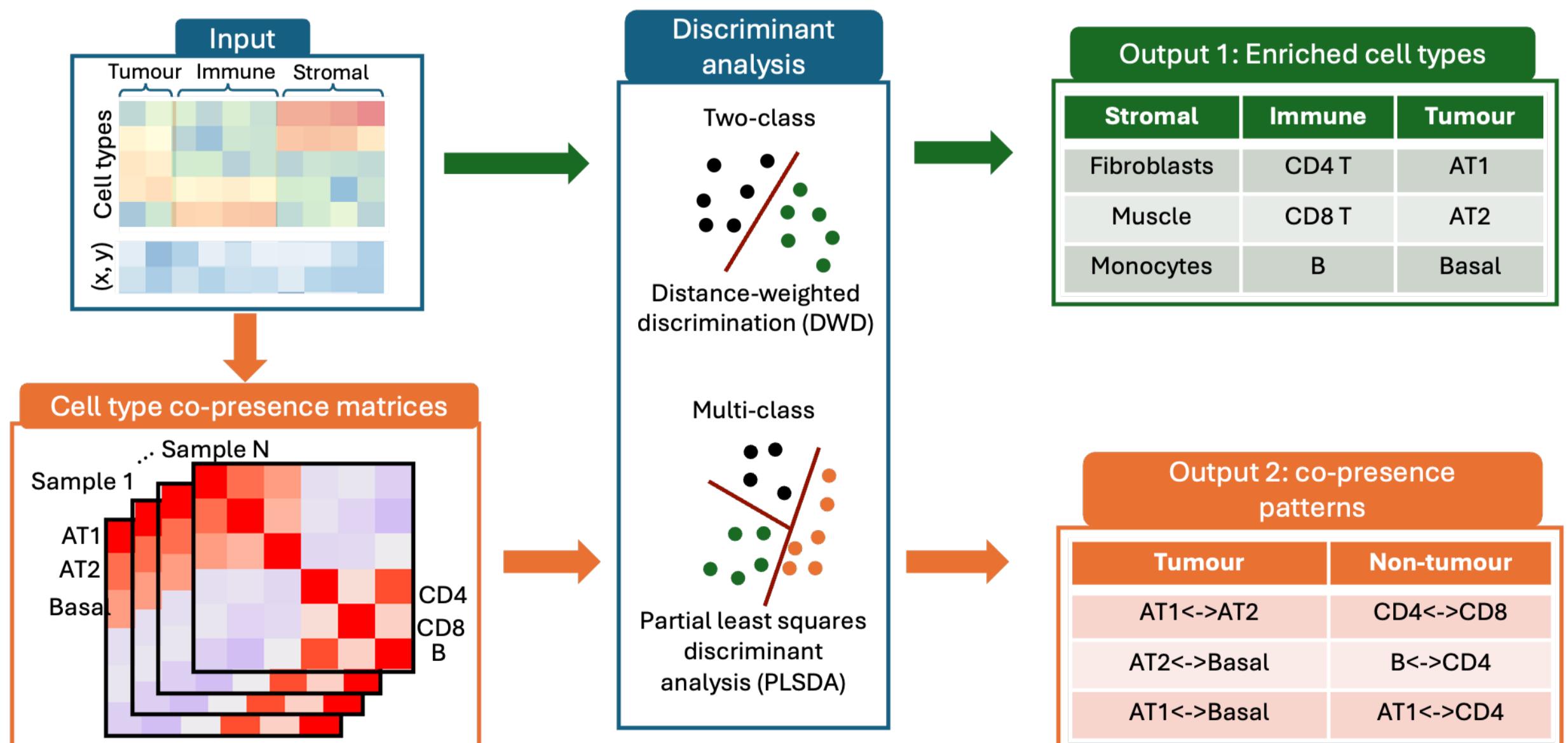


➤  $\Phi$ -Space ST is based on PLS regression. Using regression, or **soft classification**, enables continuous prediction of cell states.

➤ We train a PLS model using **each** scRNA-seq reference. Then apply each PLS model to query data.

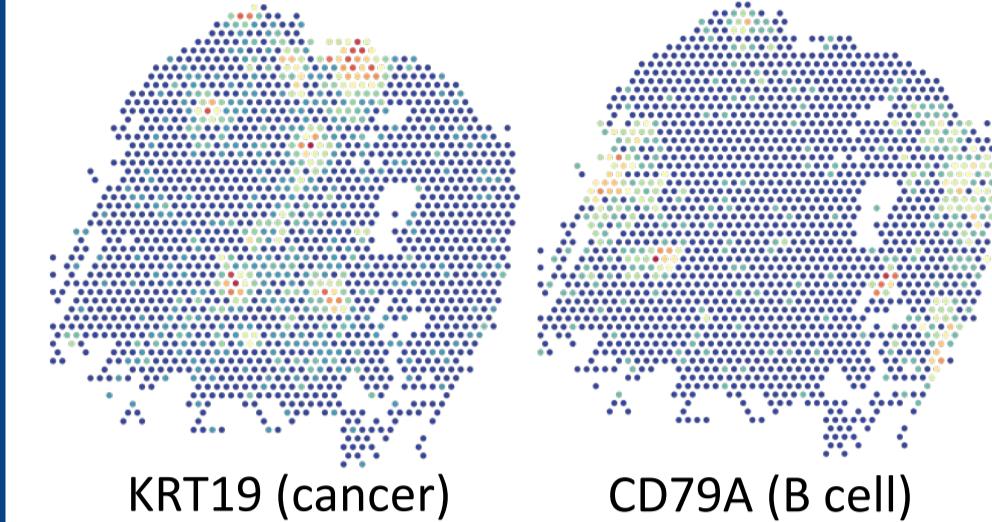
➤ The concatenation of annotations derived from all references we call **phenotype space embedding**.

➤ Phenotype space embeddings are then used as input for further **multi-sample analyses**:



## Case study: NSCLC Visium

Example: Cancerous sample P11\_T3



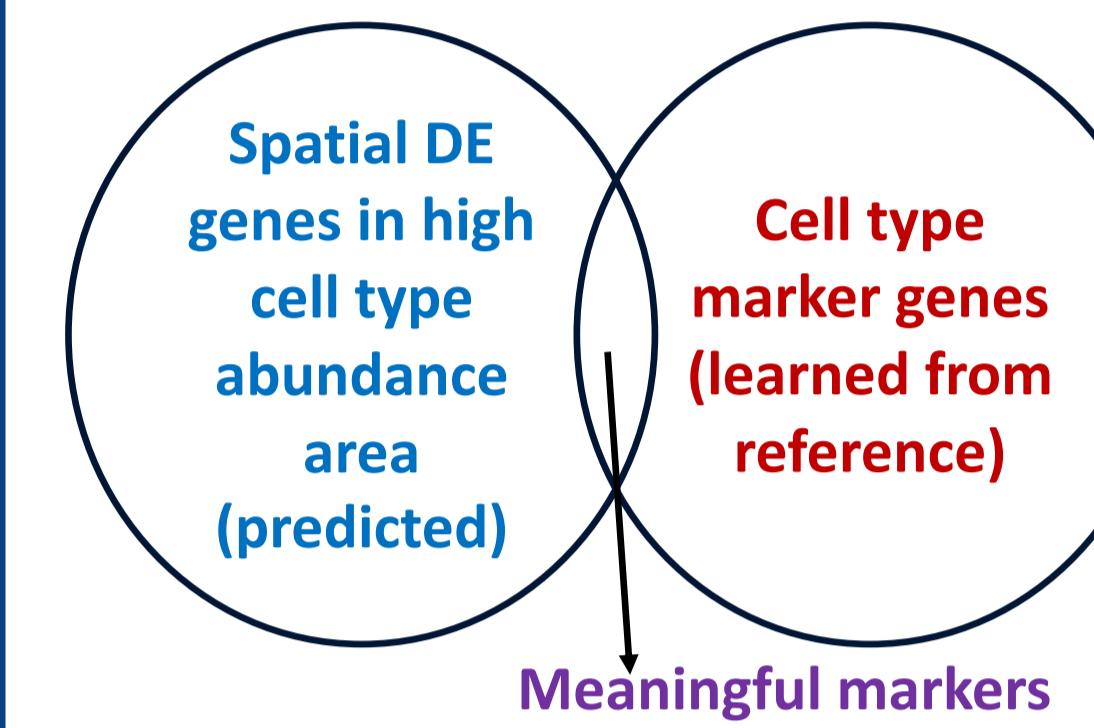
Benchmark study using **non-small cell lung cancer (NSCLC)** Visium samples [2].

- Platform: 10x Visium
- Cell-like objects: **multicellular spots** with diameter 55 $\mu$ m
- Sample info: 18 healthy and cancerous lung samples
- Reference: integrated **Human Lung Cell Atlas** (61 healthy cell types) [3]

### Challenges

- Lack of ground truth for benchmarking;
- Interpretable deconvolution when using healthy reference for cancerous query

We compare  $\Phi$ -Space ST to deconvolution methods RCTD [4], cell2location [5] and TACCO [6]. We define **number of meaningful markers** to measure their performances.

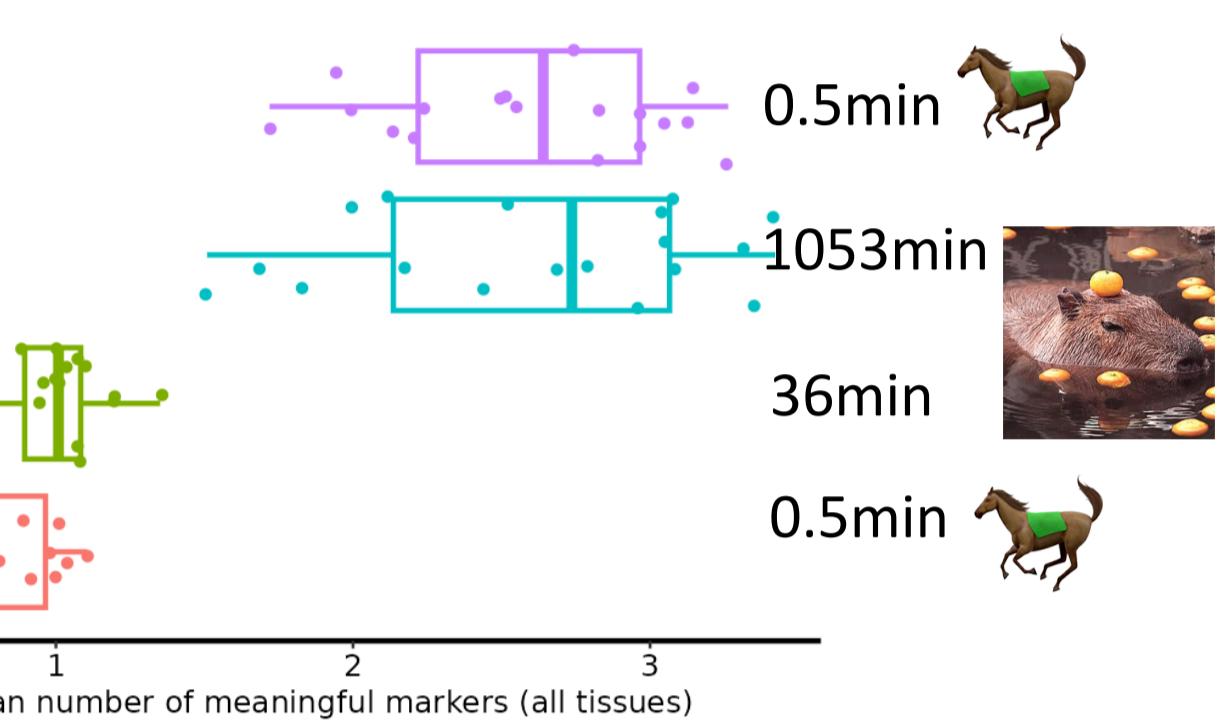


0.5min

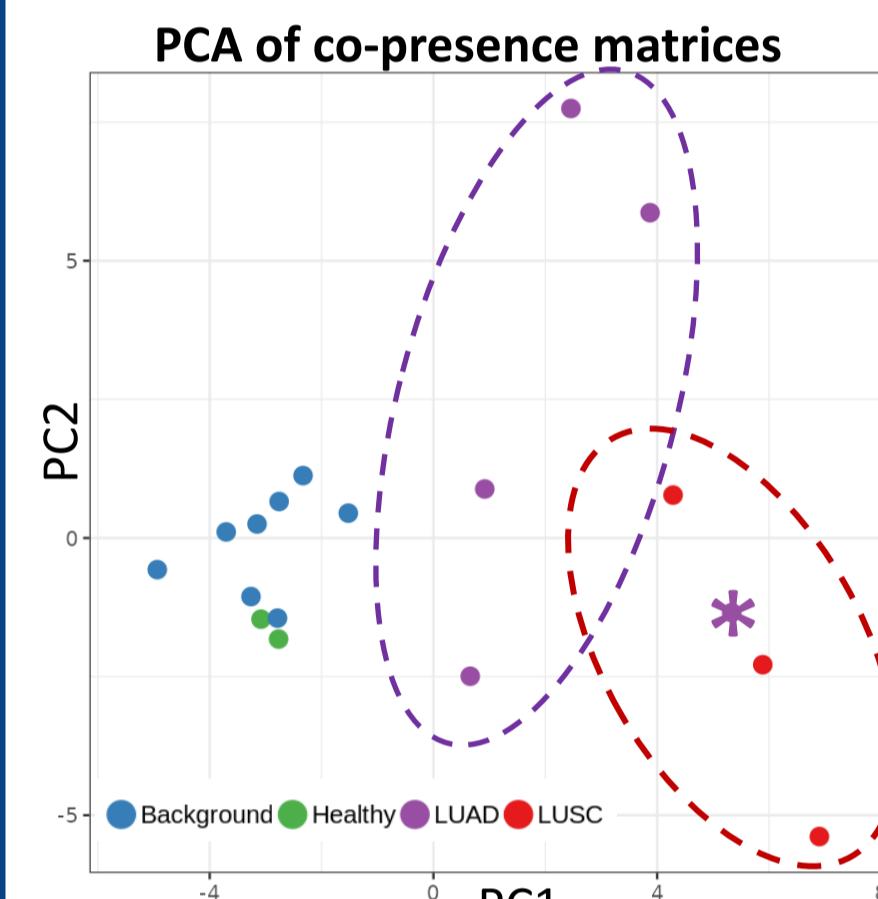
1053min

36min

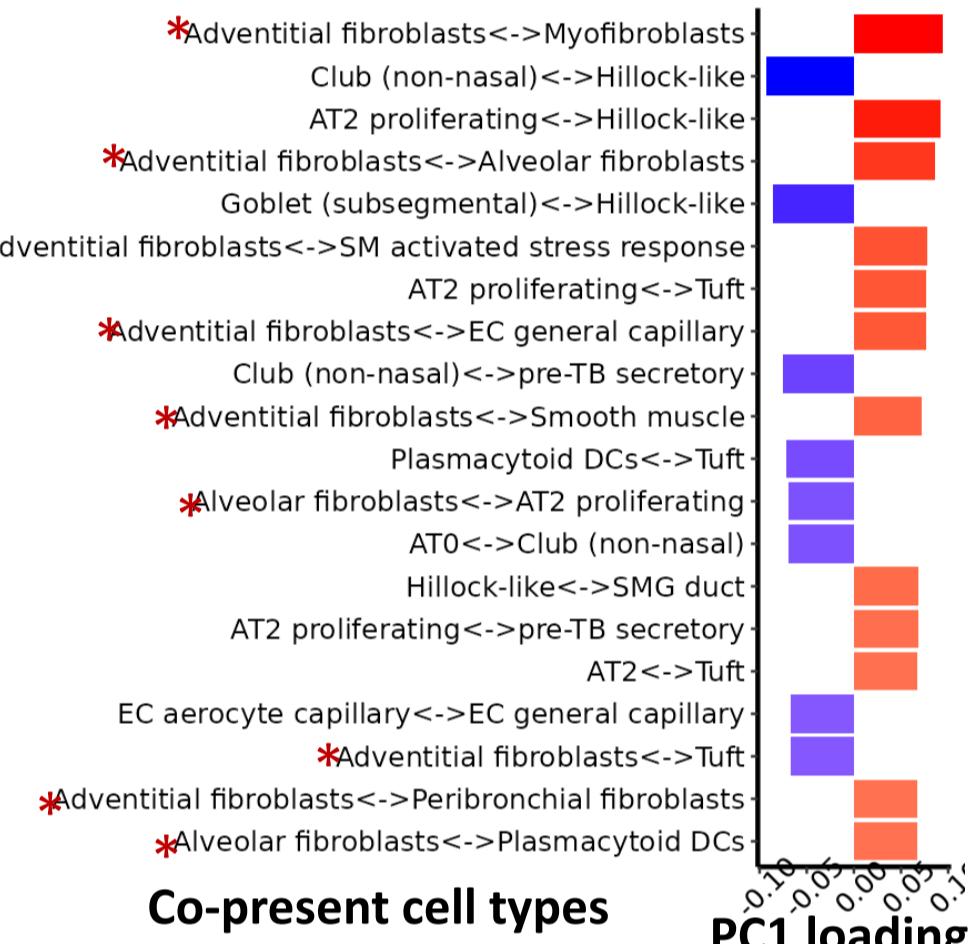
0.5min



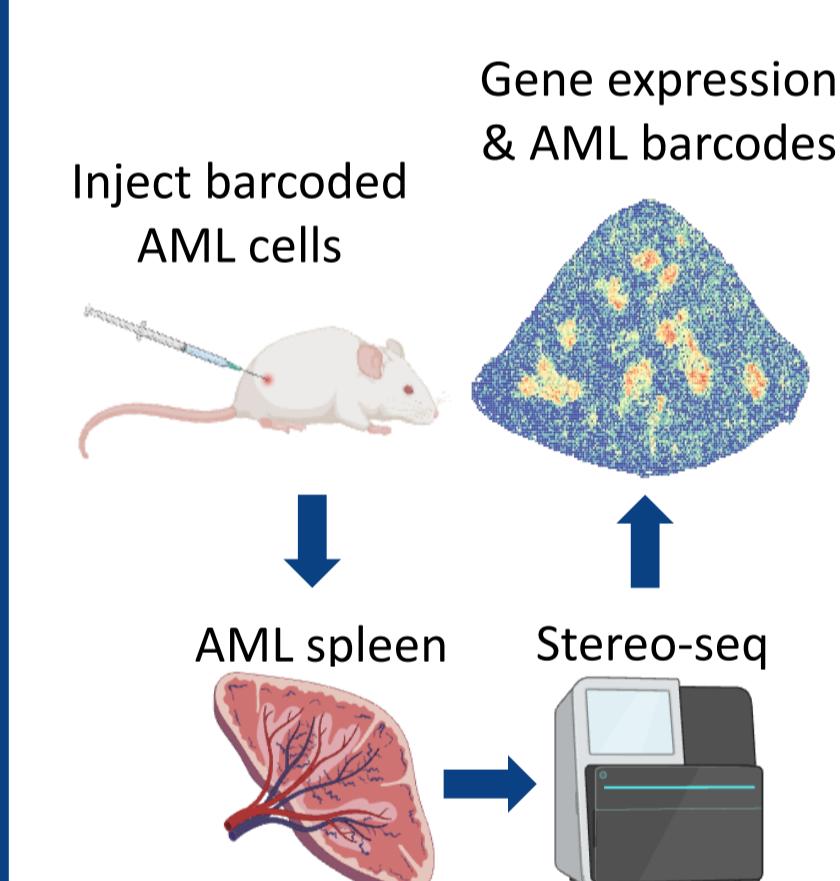
$\Phi$ -Space ST deconvolution is **meaningful** and **fast**. We then use  $\Phi$ -Space ST phenotype space embeddings for multi-sample cell type co-presence analysis.



- Different disease conditions were separated along PC1, with **lung squamous cell carcinoma (LUSC)** (worse prognosis) having larger PC1 values than **lung adenocarcinoma (LUAD)**. (\* had worst prognosis among LUAD patients.)
- Cell interaction with **fibroblasts** played a key role in separating cancer types (see PC1 loadings).



## Case study: AML Stereo-seq



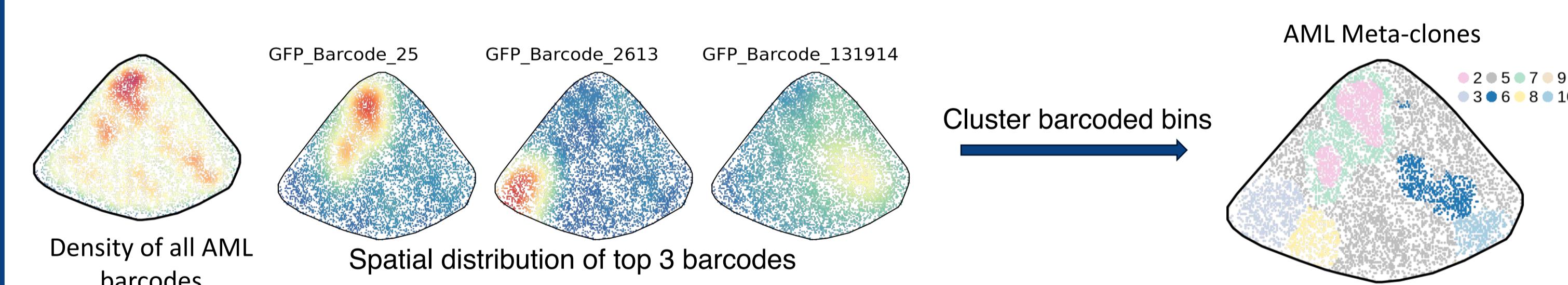
A novel workflow for **clonally-resolved** ST data [7].

- Platform: BGI Stereo-seq
- Cell like objects: **spatial bins** with side lengths 25 $\mu$ m
- Sample info:
  - 1 mouse spleen sample with **acute myeloid leukemia (AML)** cells
  - SPLINTR lineage tracing [8], same AML clone has same **barcode**
- References (all from mouse): Spleen [9]; BM: bone marrow [10]; CITE: CITE-seq spleen [11]; Neutro: neutrophils [12]

### Challenges

- Lack of cell segmentation
- Lack of reference atlas
- Liquid tumour lacking structures

We first identify 8 AML **meta-clones**, each consists of bins with similar composition of individual barcodes.



### Meta-clone-specific enriched cell types (derived from 4 references)

Meta-clone 2	Meta-clone 3	Meta-clone 6	Meta-clone 7	Meta-clone 8	Meta-clone 10
1 Granulo (BM)	Imm NK (BM)	RBC (CITE)	HPC (BM)	HPC (BM)	Naive B (BM)
2 Neutro (Spleen)	Trans B (Spleen)	RBC (Spleen)	IMM 1 (Neuro)	T (BM)	ProEryThBla (BM)
3 HPC (BM)	T (BM)	ErythBla (BM)	Neutro (Spleen)	Pre-B cycl (Spleen)	Pre-B cycl (Spleen)
4 T1 (Neuro)	MAT 3 (Neuro)	ProEryThBla (BM)	Pre-B cycl (Spleen)	Imm NK (BM)	ICOS+ Tregs (CITE)
5 Imm NK (BM)	CD8 T (Spleen)	IMM 2 (Neuro)	Pre-B (BM)	Pre-B (BM)	CD4 T (CITE)

Meta-clone-specific enriched cell types were found by comparing  $\Phi$ -Space ST phenotype space embeddings of each meta-clone and background (bins without barcode).

Meta-clones had clear spatial structures. Spatially close clones had similar cell states.

## Discussion

By developing  $\Phi$ -Space ST, we provide novel ways for analysing ST data.

- Cell segmentation is not always needed for ST studies, especially when the goal is to characterise **spatial patterns** of cell states.
- No matter how cell-like objects are defined in ST studies, due to the lack of cell isolation, the **mixture of cell types** is a persistent problem. Hence cell type deconvolution is needed.
- $\Phi$ -Space ST achieves computationally efficient and interpretable cell type deconvolution and hence serves as a **unified approach** for annotating cell states in ST studies.
- Existing deconvolution methods rely on complex statistical assumptions about data distribution, restricting them to one omics type. In contrast,  $\Phi$ -Space ST is fully nonparametric, making it straightforward to extend to other omics types and **spatial multiomics**.

## References

1. Benjamin, K., Bhandari, A., ... Bull, K. R. (2024). Multiscale topology classifies cells in subcellular spatial transcriptomics. *Nature*.
2. De Ziani, M., ... Cvejic, A. (2024). Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat. Commun.*
3. Sikkema et al. (2023). An integrated cell atlas of the lung in health and disease. *Nat. Med.*
4. Cable, D. M., ... Izarzury, R. A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotech.*
5. Kleshcheynikov, V., ... Bayraktar, O. A. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotech.*
6. Mages, S., ... Nitzan, M. (2023). TACCO unifies annotation transfer and decomposition of cell identities for single-cell and spatial omics. *Nat. Biotech.*
7. Holze, H., ... Dawson, M. A., & Vassiliadis, D. (2024). Analysis of synthetic cellular barcodes in the genome and transcriptome with BARtab and bartools. *Cell Rep. Methods*.
8. Fennell, K. A., Vassiliadis, D., ..., Dawson, M. A. (2022). Non-genetic determinants of malignant clonal fitness at single-cell resolution. *Nature*.
9. Zhang, Y., ... Fang, X. (2023). Microbiota-mediated shaping of mouse spleen structure and immune function characterized by scRNA-seq and Stereo-seq. *J. Genet. Genomics*.
10. Harris, B. D., Lee, J., & Gillis, J. (2021). A Meta-Analytic Single-Cell Atlas of Mouse Bone Marrow Hematopoietic Development. *bioRxiv*.
11. Gayoso, A., ... Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods*.
12. Ng, M. S. F., ..., Ng, L. G. (2024). Deterministic reprogramming of neutrophils within tumors. *Science*.

Acknowledgement:  
NHMRC Investigator Grant  
GNT2025648.

We thank Ms Henrietta Holze, Dr Dene Vassiliadis and Prof Mark Dawson for processed mouse spleen data and helpful discussions.

