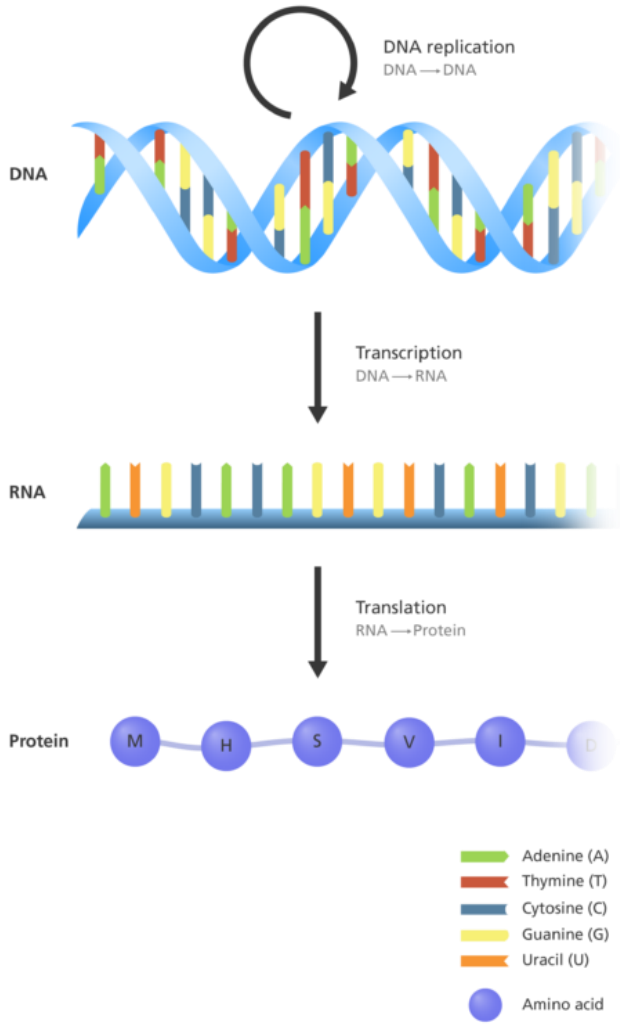
What’s so spatial about spatial transcriptomics

Jiadong Mao

Melbourne Integrative Genomics

Name of the game

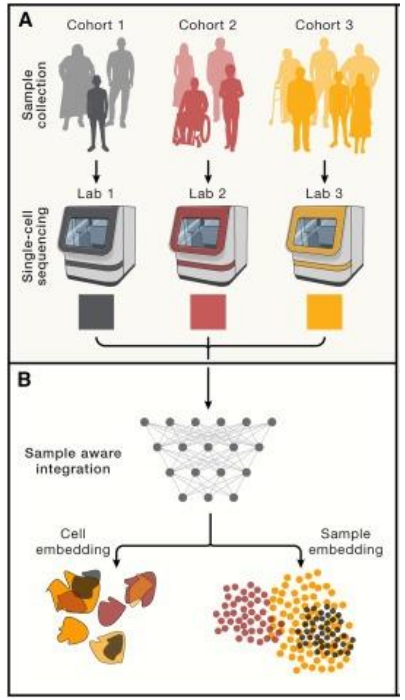
- Central dogma of molecular biology
- ‘Omics’ data & high-throughput sequencing
- Types of omics
  - Genomics
  - **Transcriptomics**
  - Proteomics
  - Metabolomics, epigenomics, …
- What’s so special about RNAs (transcripts/gene expression)
  - Cell activities and identities: T cell, B cell, …



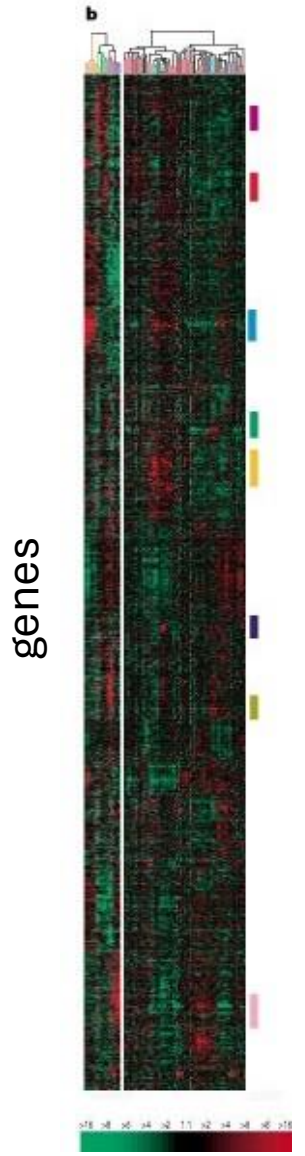
Laura Olivares Boldú / Wellcome Connecting Science

Omics and statistics

- Bulk RNA sequencing, eg microarray
- HDLSS: high dimension, low sample size
  - ~50 samples, >10,000 genes
- Variable selection, multiple testing
- Common goal: marker gene identification
  - Diagnosis, treatment, prognosis, eg cancer subtyping



Lotfollahi et al. (2024). *Cell*.



Perou et al. (2000). *Nature*.

J. R. Statist. Soc. B (2005)
67, *Part* 3, *pp.* 427–444

Geometric representation of high dimension, low sample size data

Peter Hall,
Australian National University, Canberra, Australia
J. S. Marron
University of North Carolina, Chapel Hill, USA
and Amnon Neeman
Australian National University, Canberra, Australia

JOURNAL ARTICLE

The Group Lasso for Logistic Regression FREE

Lukas Meier ✉, Sara Van De Geer, Peter Bühlmann

Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 70, Issue 1, February 2008, Pages 53–71, https://doi.org/10.1111/j.1467-9868.2007.00627.x
Published: 04 January 2008 **Article history** ▾

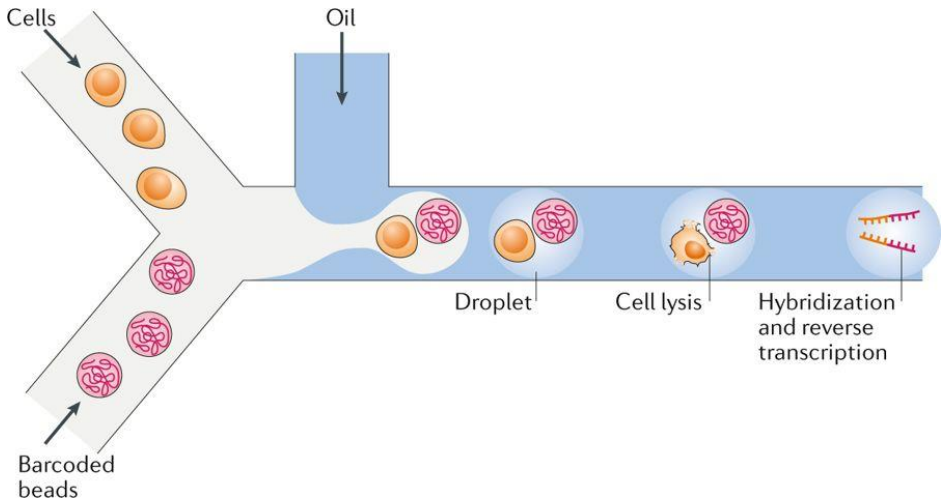
The Annals of Applied Statistics
2007, Vol. 1, No. 1, 107–129
DOI: 10.1214/07-AOAS101

ON TESTING THE SIGNIFICANCE OF SETS OF GENES

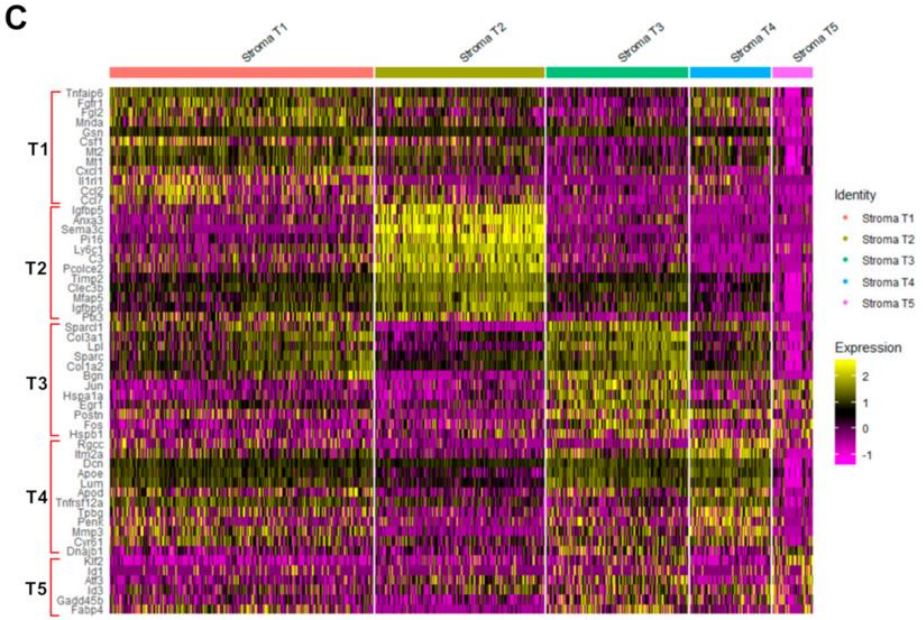
BY BRADLEY EFRON[1] AND ROBERT TIBSHIRANI[2]
Stanford University

New challenges: big omics data

- Single-cell RNA sequencing (scRNA-seq)
  - Dissolve tissue into single cells & seq
- ‘HDHSS’: High dimension, high sample size
  - >10,000 cells per sample/donor
  - >20,000 genes
- Finding marker genes at cell (type) level



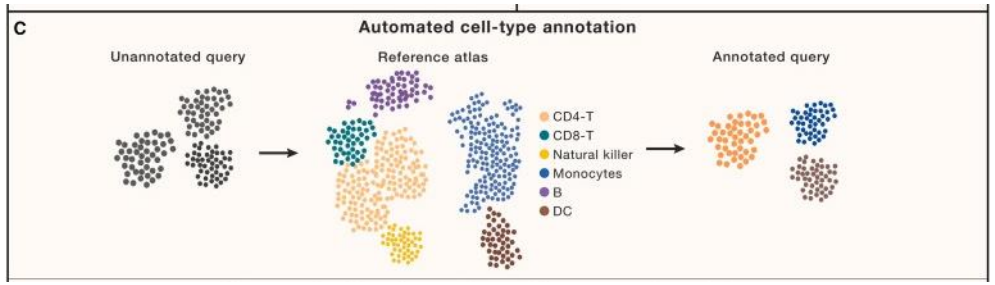
https://www.rna-seqblog.com/wp-content/uploads/2018/08/droplet.png



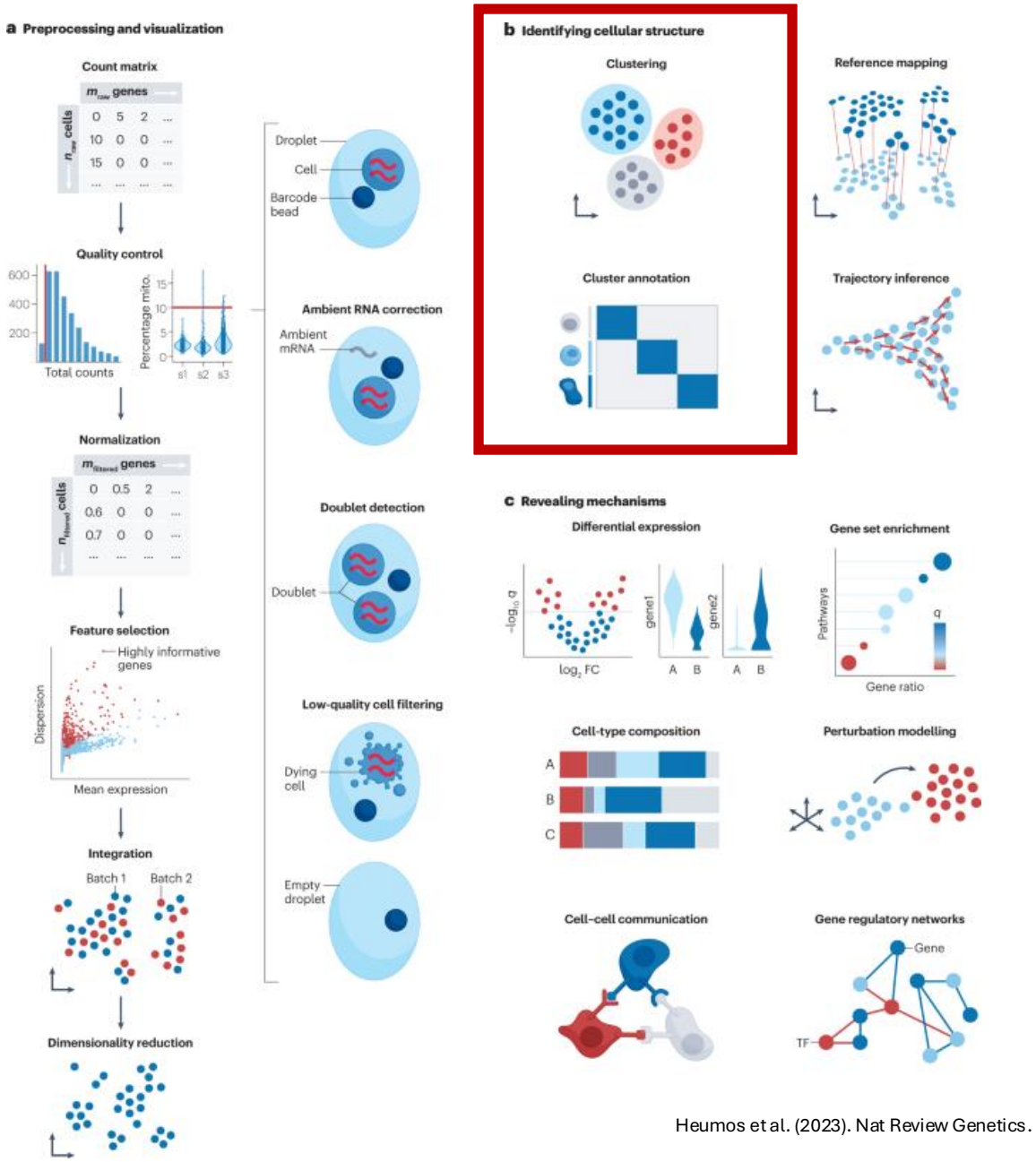
Lo et al. (2020). *Cancers*.

Cell type annotation

- Cells form (relatively) homogeneous groups
- Group cells into cell types
  - Unsupervised: cluster + manual annotation (multiple testing)
  - Supervised: train cell type classifier on reference data



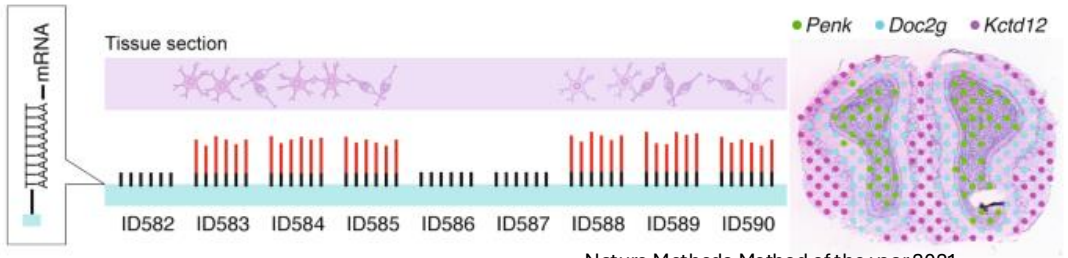
Lotfollahi et al. (2024). *Cell*.



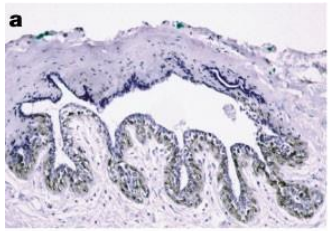
Heumos et al. (2023). Nat Review Genetics.

Adding spatial information

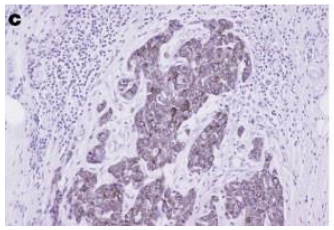
- Spatial transcriptomics (ST)
- Biology is spatial
  - 20-century tools: microscope, antibody staining, eg basal keratin
  - But limited throughput (dimensionality)
- Why ST
  - High-throughput: HD measurements in situ
  - scRNA-seq: which cells are doing what
  - ST: which cells are doing what, and **where they are doing it**
  - Example: tumour microenvironment, TLS
- Cell type annotation is still the key



Nature Methods Method of the year 2021

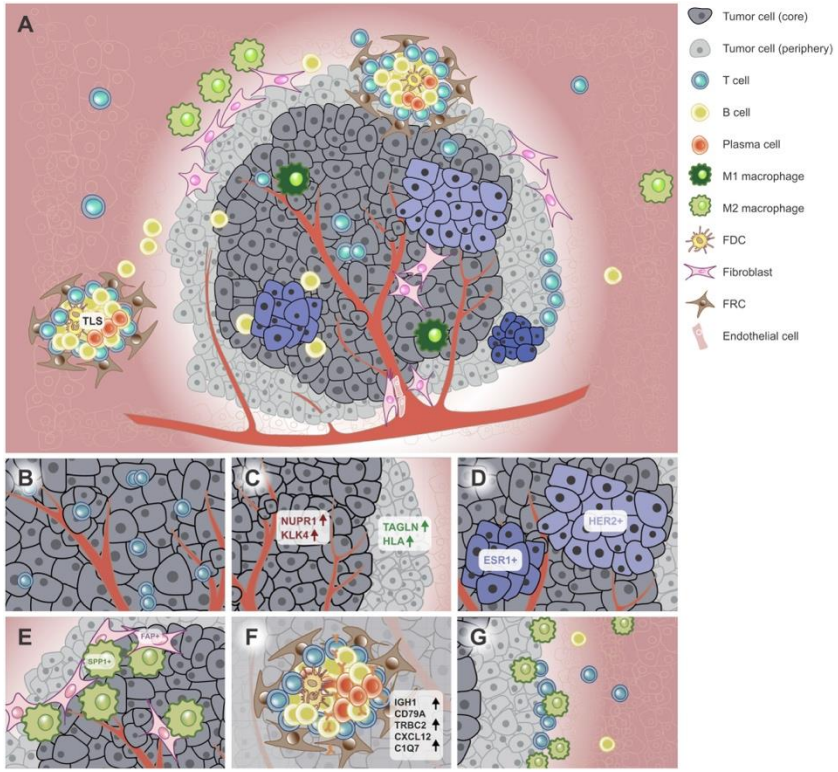


normal



tumour

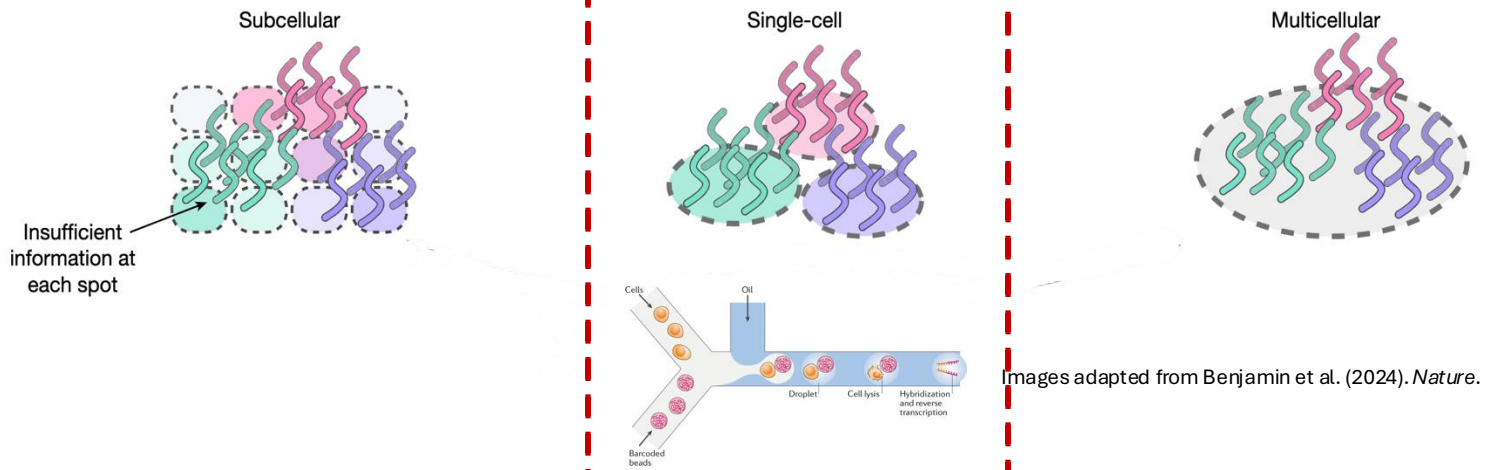
Perou et al. (2000). *Nature*.



Wait … but what is a cell in ST?

- Fast evolving tech, numerous providers
- Different spatial resolutions
  - Sub-cellular
  - Super-cellular
  - Lack of real single cells
- So, how does cell type annotation happen?

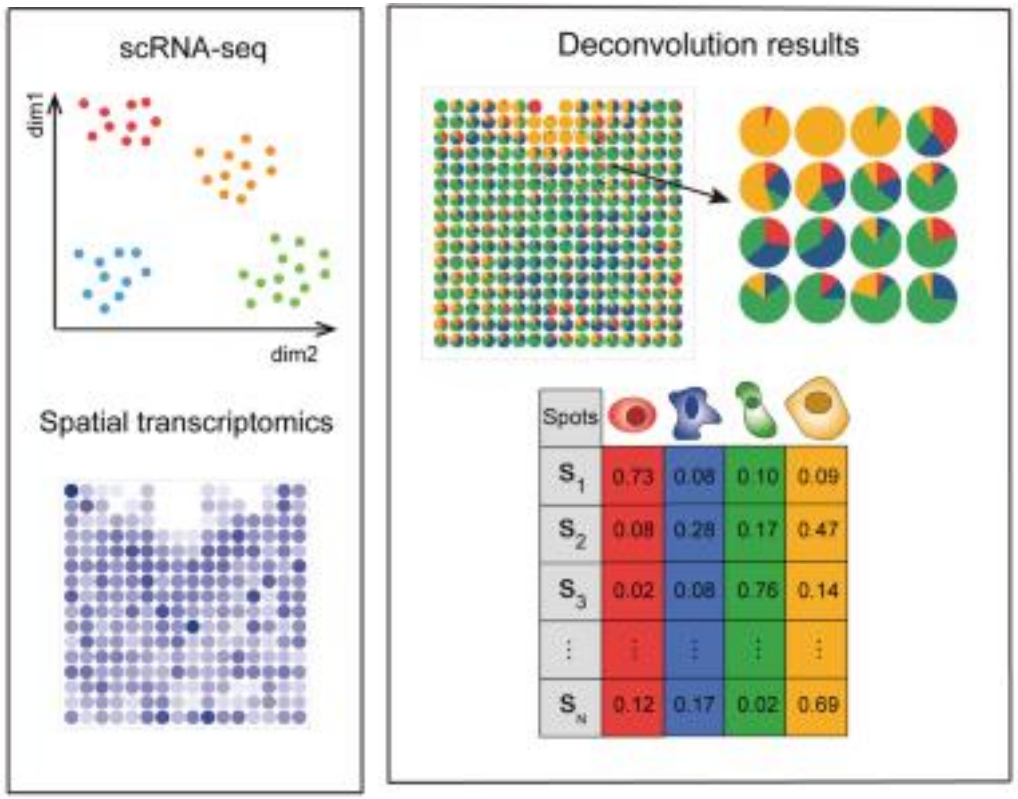
https://www.inveniagroup.com/blog/spatial-omics-technology/



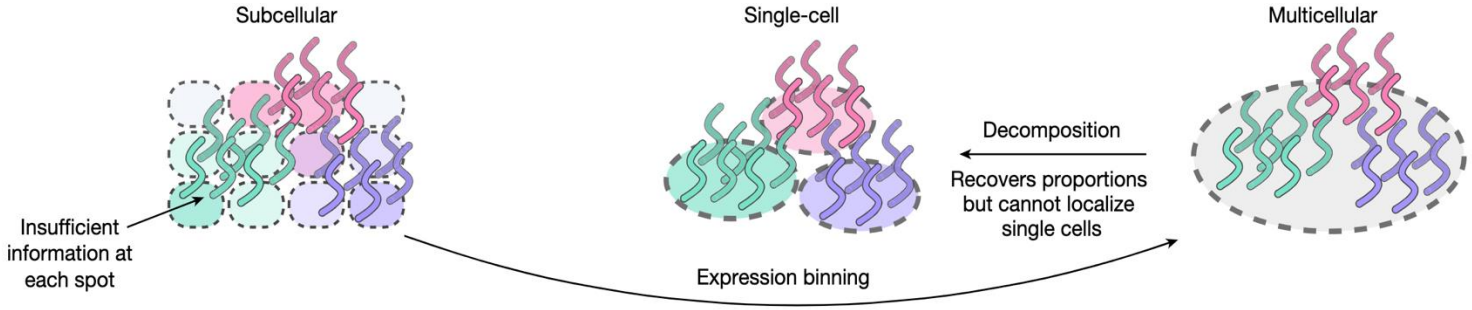
Images adapted from Benjamin et al. (2024). *Nature*.

Cell type deconvolution

- For super-cellular ST:
  - Each spot contains multiple cells
  - Use scRNA-seq as reference
  - Decompose gene expression in each spot as combination of reference cell types
- For sub-cellular ST:
  - Bin the RNAs & deconvolute



Li et al. (2023). Nat Comms



Images adapted from Benjamin et al. (2024). *Nature*.

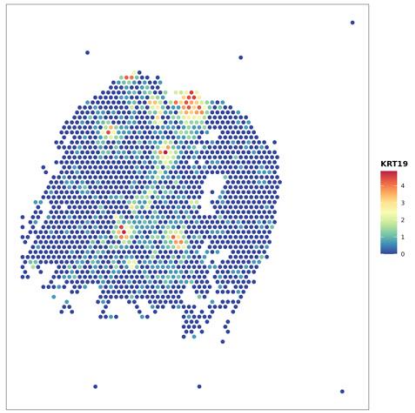
Example: Lung Cancer

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

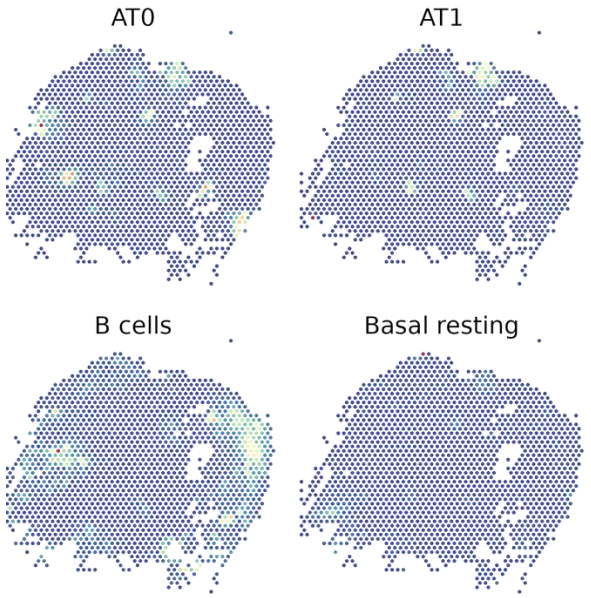
Φ-Space: Continuous phenotyping of single-cell multi-omics data

Jiadong Mao[1], Yidi Deng[1], Kim-Anh Lê Cao[1,*]

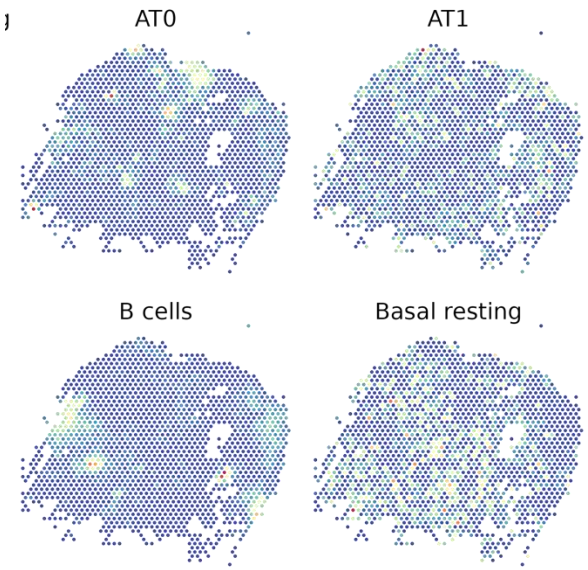
[1]Melbourne Integrative Genomics, School of Mathematics and Statistics, The University of Melbourne, Australia



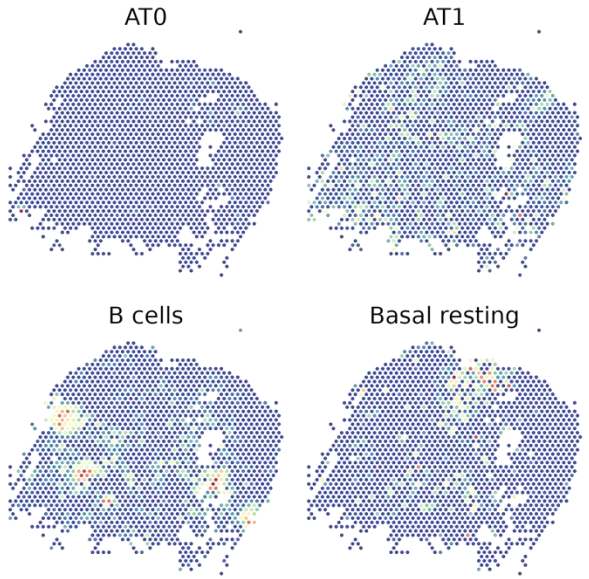
- 10x Visium lung cancer (LUAD) sample
- Each spot diameter 55μm, containing 1–10 cells
- Used lung scRNA-seq as reference (61 cell types)
- Compared 4 methods: c2l, RCTD, TACCO, PhiSpace
- Which one should I trust?



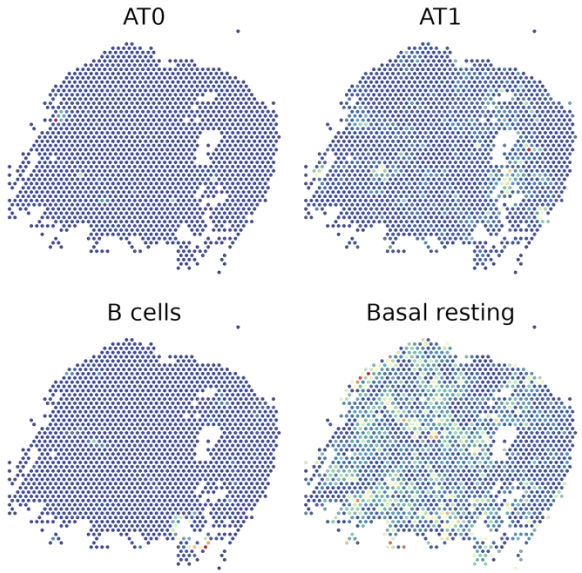
cell2location



PhiSpace



RCTD

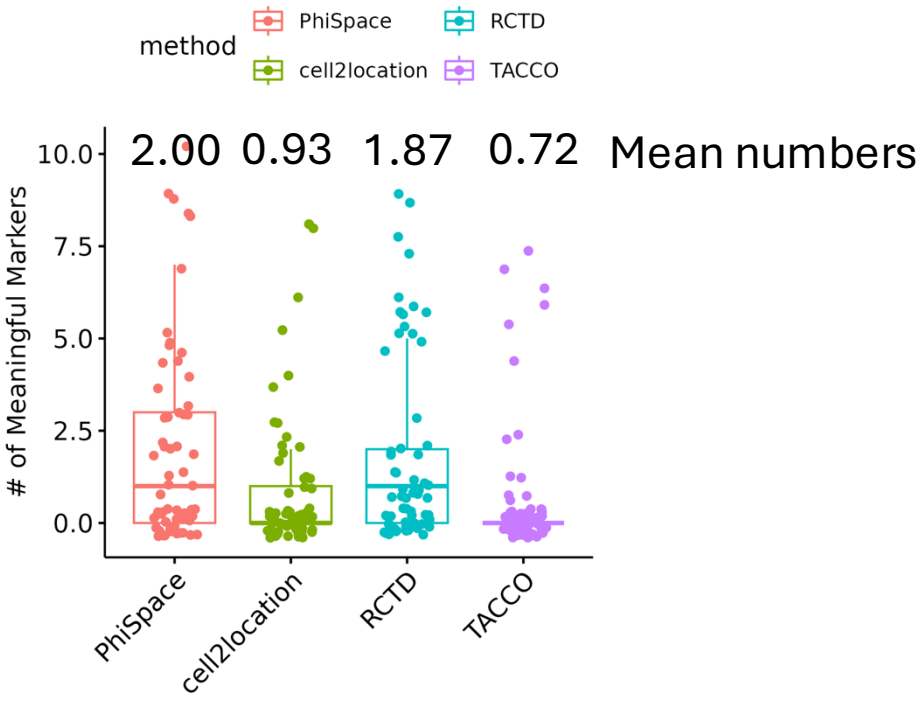


TACCO

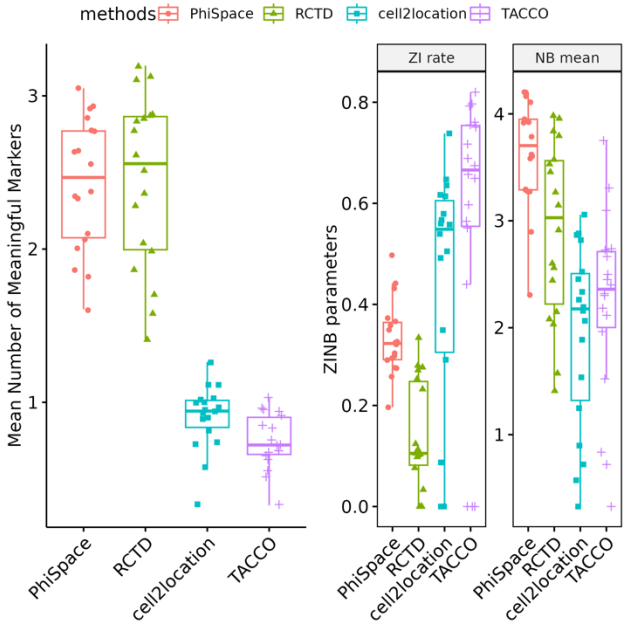
Marker analysis

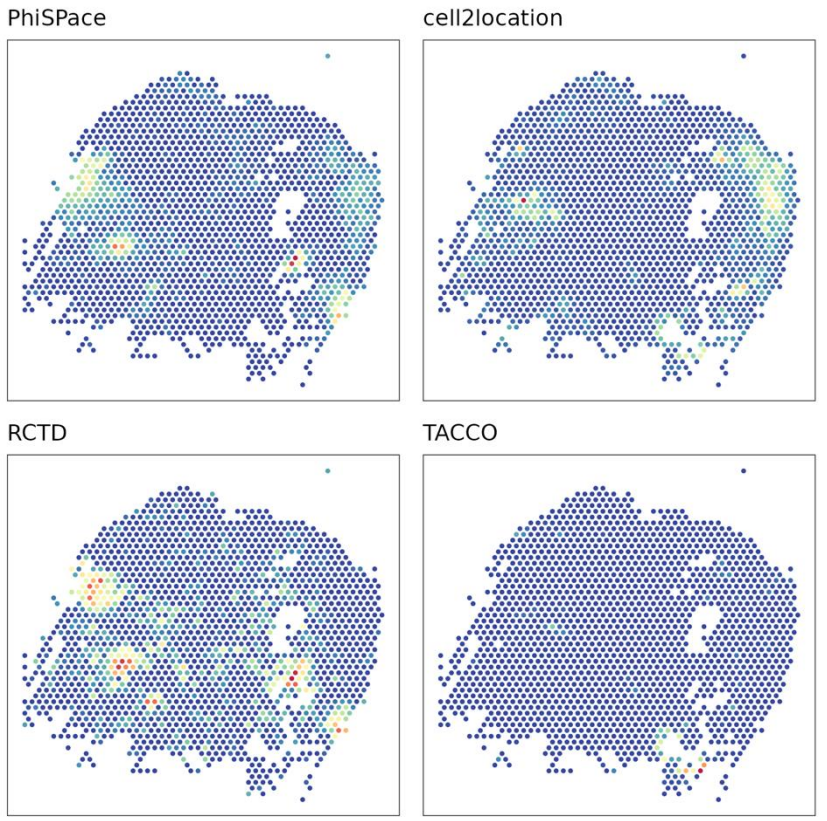
- No ground truth
- A proxy of truth:
  - If you say a spot has higher B cell proportions, that spot should contain B cell marker genes

1 tissue



18 tissues





Predicted B cell presence

Reflections

- Fast-evolving biotech, reliable stats method needed
  - 3D spatial, spatio-temporal, …
- Need fast paced stats method development
- Collaborative culture
  - Wet: Biologists, bioinformaticians;
  - Bridge: computational biologists;
  - Dry: statisticians, mathematicians, computer scientists