Beyond the Basics:
What You Can Do After Annotating Cells

Pre-conference Workshop on Single-cell and Spatial Omics Technology

AP-SciBio2025

Waradon Sungnak, PhD

Department of Microbiology, Faculty of Science

Integrative Computational BioScience (ICBS) Center

Mahidol University

19/06/2025

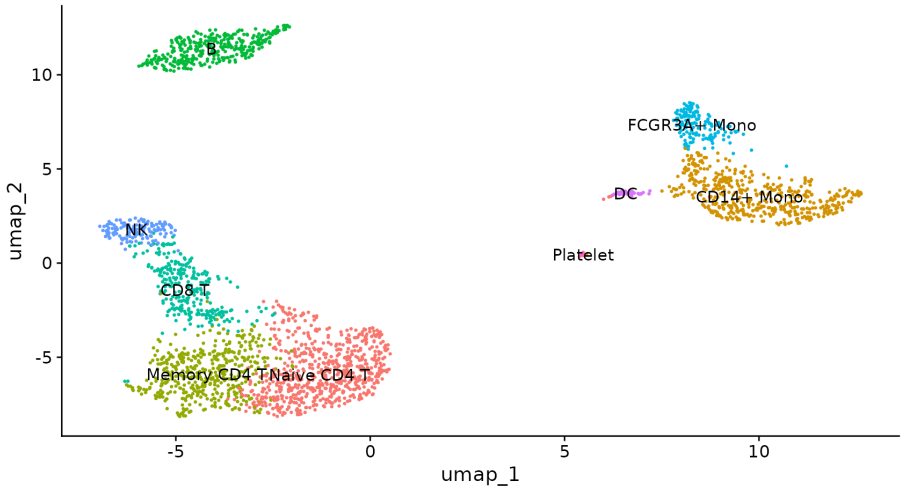
Outline

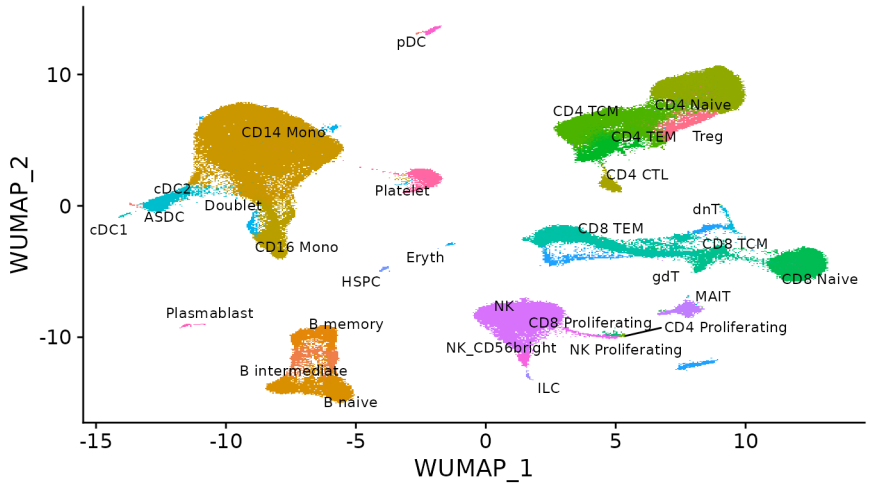
- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

More exploration on cell annotation





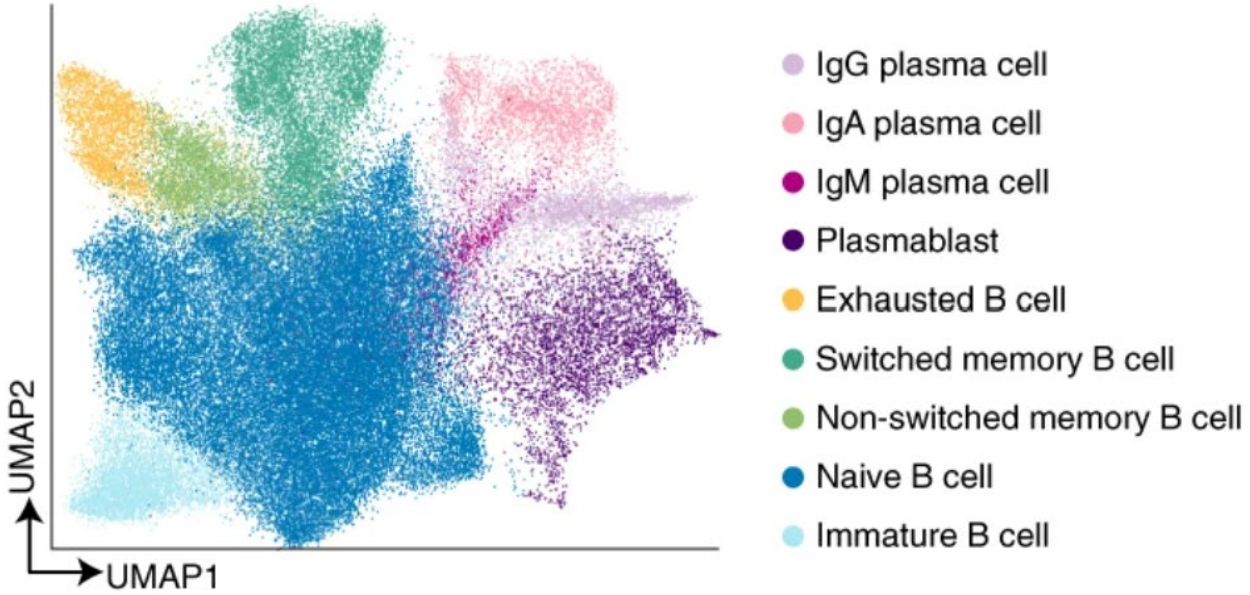
- Basic QC, clustering, and annotation
- Key techniques they can now explore more:
  - **Adjusting clustering resolutions**
  - Subsetting for more focused analysis
  - Reference-based annotation tools
    - Tools like **SingleR**, **Azimuth**, **scPred**, and **CellTypist**

Source: https://satijalab.org/seurat/

More exploration on cell annotation

- Basic QC, clustering, and annotation
- Key techniques they can now explore more:
  - Adjusting clustering resolutions
  - **Subsetting for more focused analysis**
  - Reference-based annotation tools
    - Tools like **SingleR**, **Azimuth**, **scPred**, and **CellTypist**

PBMCs -> Subset on B cells and plasmablasts/plasma cells

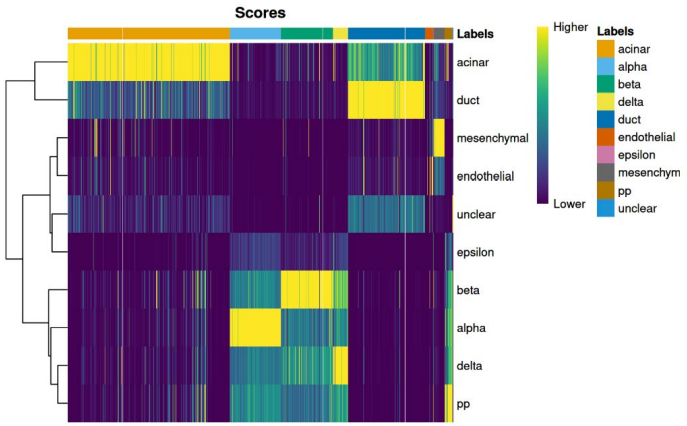


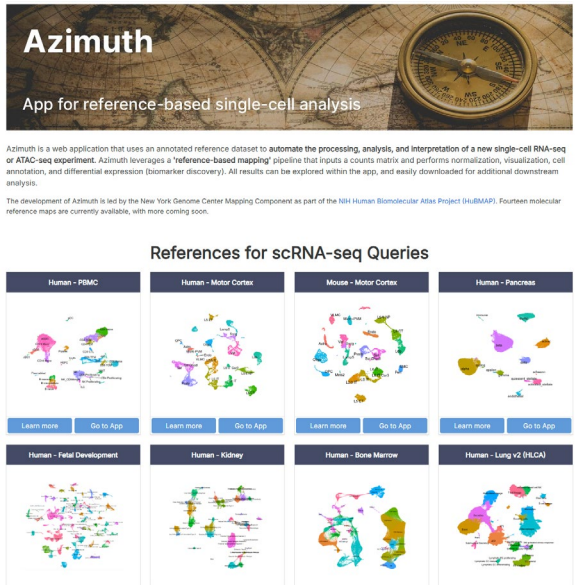
Stephenson et al. *Nat Med*. 2021

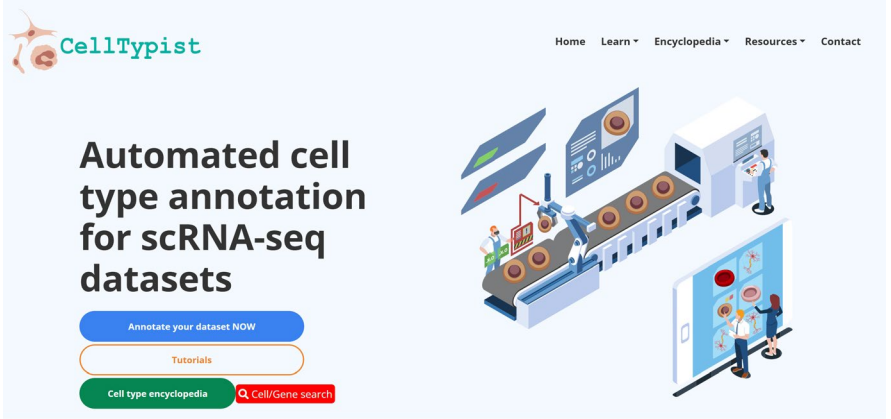
More exploring on cell annotation

- Basic QC, clustering, and annotation
- Key techniques they can now explore more:
  - Adjusting clustering resolutions
  - Subsetting for more focused analysis
  - Reference-based annotation tools
    - Tools like **SingleR**, **Azimuth**, **scPred**, and **CellTypist**

```
plotScoreHeatmap(pred.grun)
```



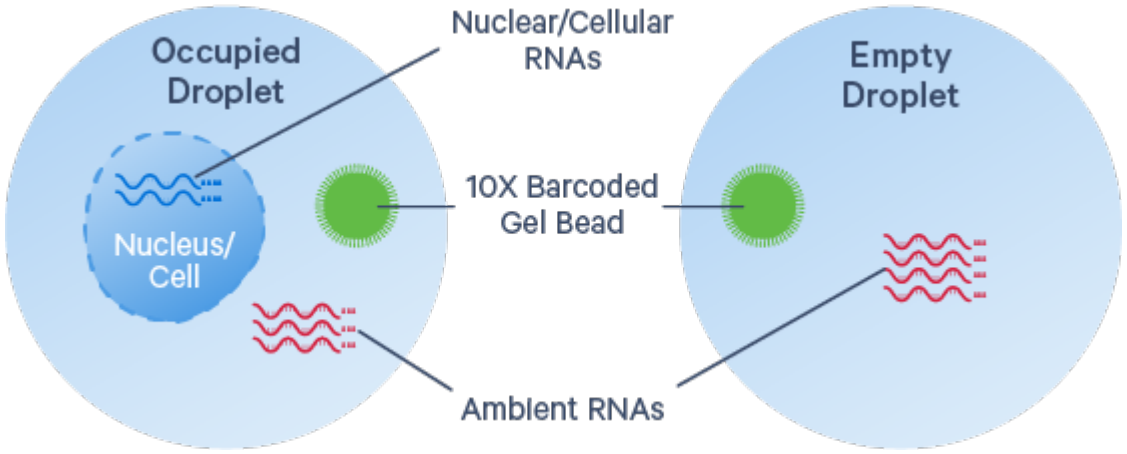




Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

Ambient RNA removal

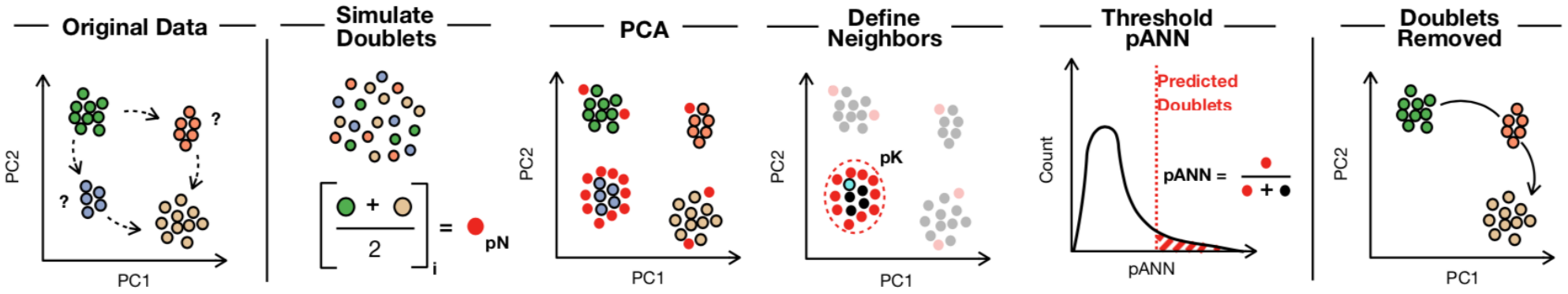


Source: 10X Genomics

- Free-floating RNA from lysed or dying cells captured in droplets
- Usual suspects:
  - Hemoglobin genes: HBA1, HBA2, HBB → from lyzed red blood cells.
  - Immunoglobulin genes: IGH*, IGK*, IGL* → from lyzed plasmablasts/plasma cells
- Apply correction before normalization
- Example tools: **SoupX, DecontX, CellBender**

Doublet detection and removal

- Doublets: two cells are captured in the same droplet
- Remove doublets before clustering and annotation
- Example tools for doublet detection and removal:
  - **DoubletFinder** (R)
  - **scDblFinder** (R)
  - **Scrublet** (Python)



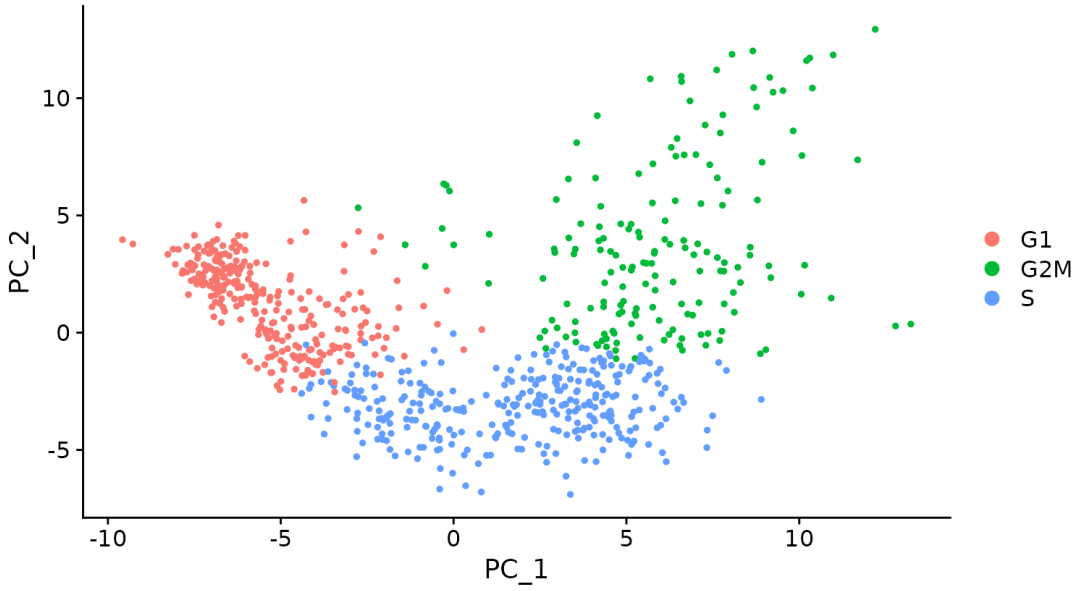
Source: https://github.com/chris-mcginnis-ucsf/DoubletFinder

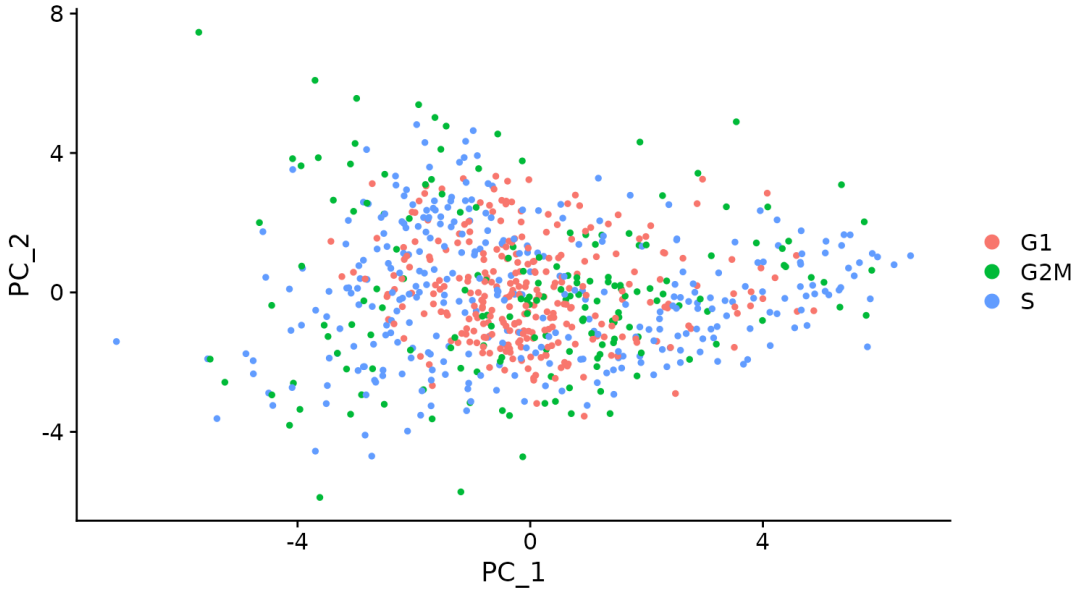
Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

Cell cycle regression

Regress out cell cycle scores (S/G2M phases) **when it's not biologically relevant**

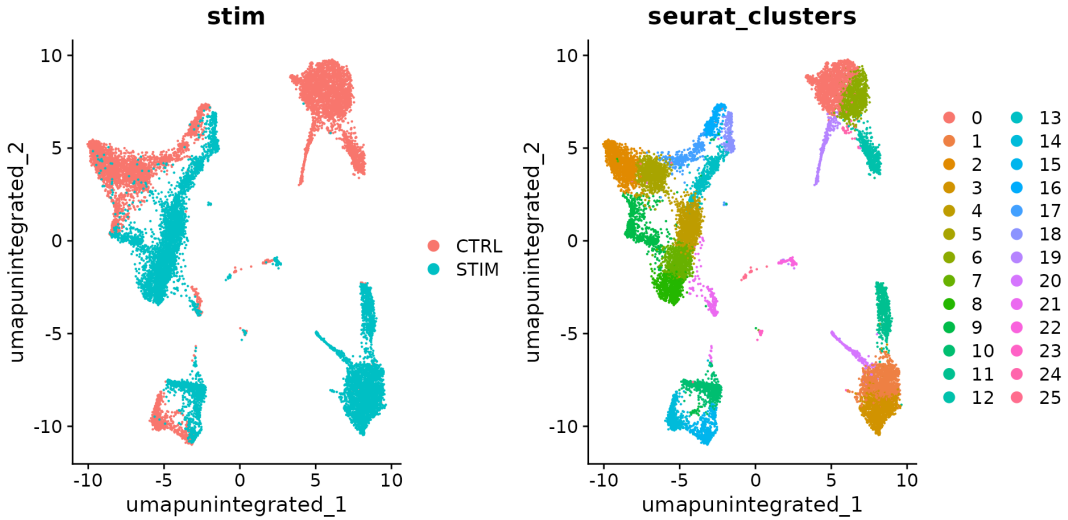




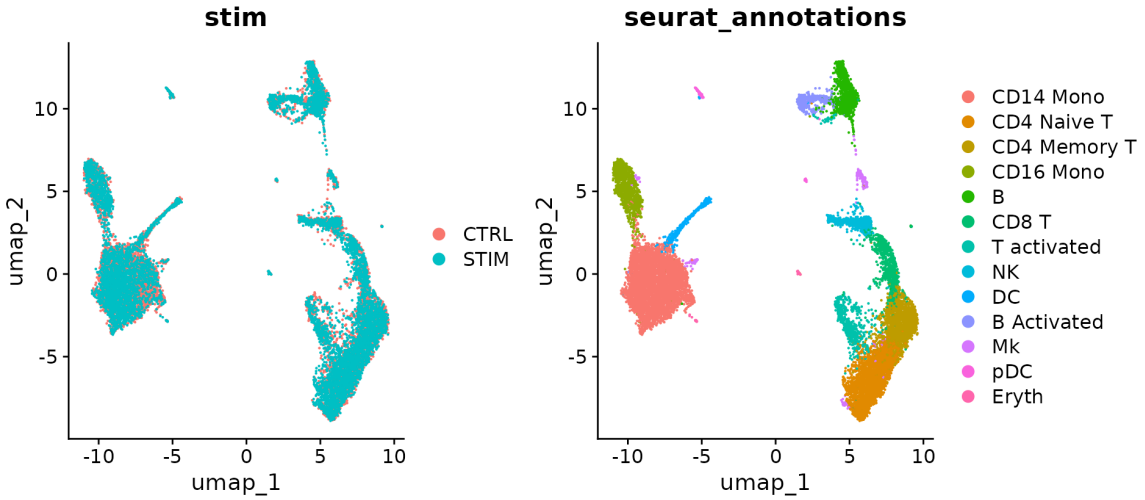
Source: https://satijalab.org/seurat/articles/cell_cycle_vignette.html

Batch correction/Data integration

Before correction

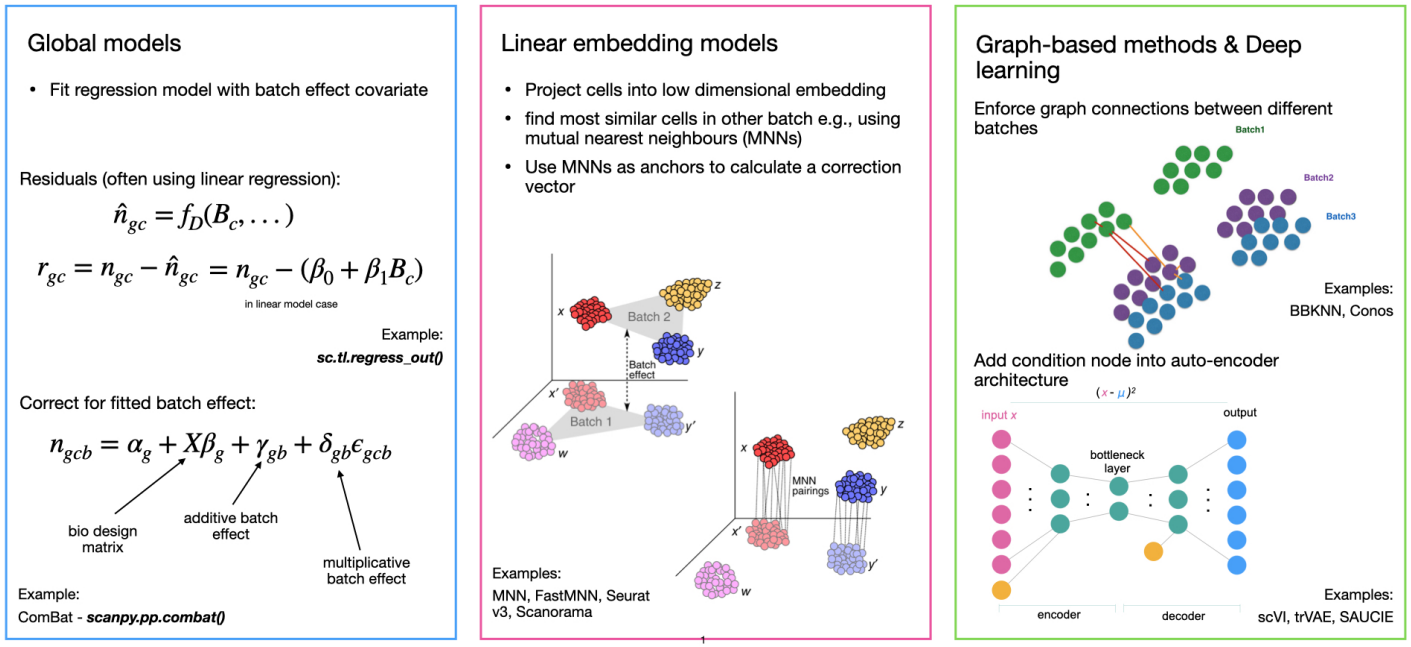


After correction



Source: https://satijalab.org/seurat/

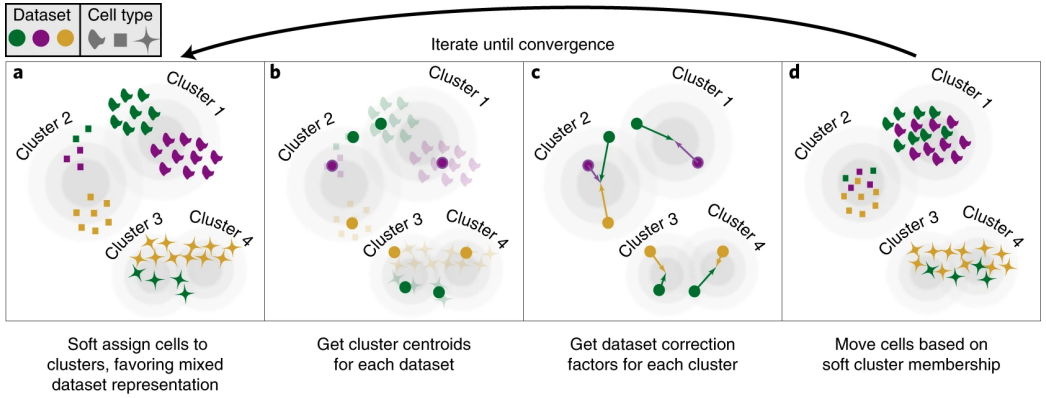
Batch correction/Data integration



Examples

- Seurat (CCA, RPCA)
- Harmony
- FastMNN
- Scanorama
- BBKNN
- scVI, scANVI, scGEN

Source: https://https://www.sc-best-practices.org/

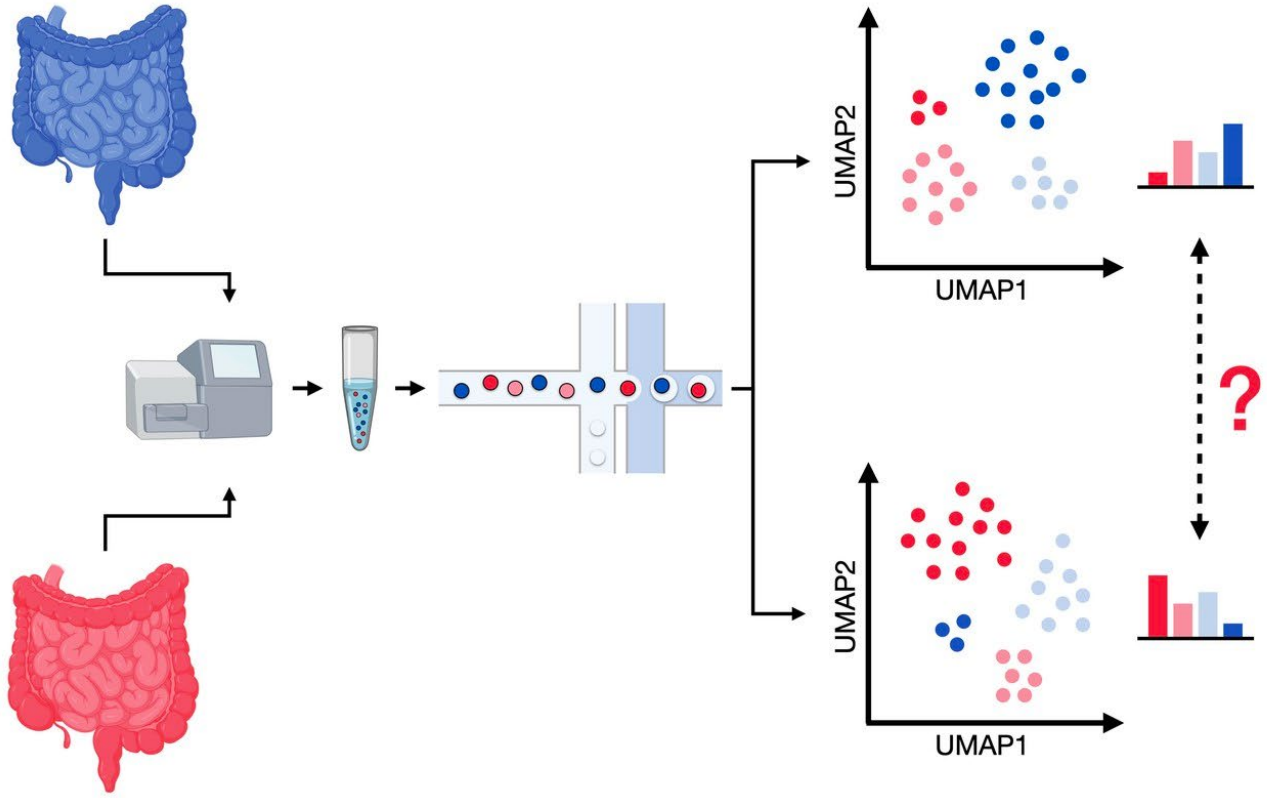


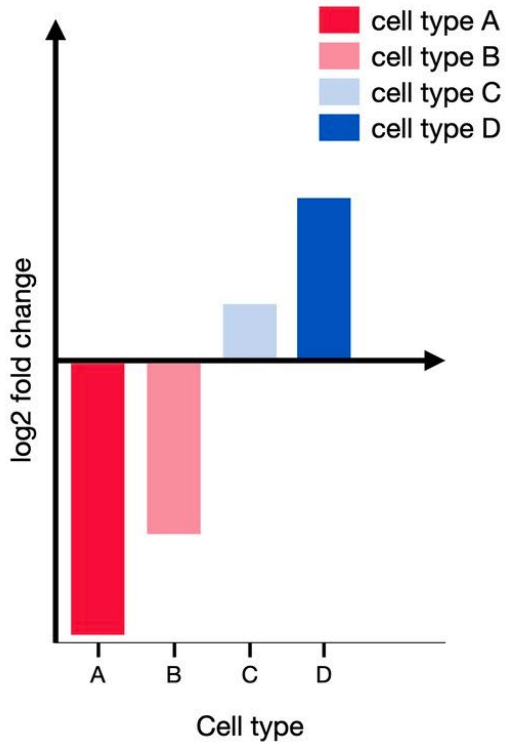
Korsunsky. *Nat Methods*. 2019

Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

Differential abundance/compositional analysis: cluster-based

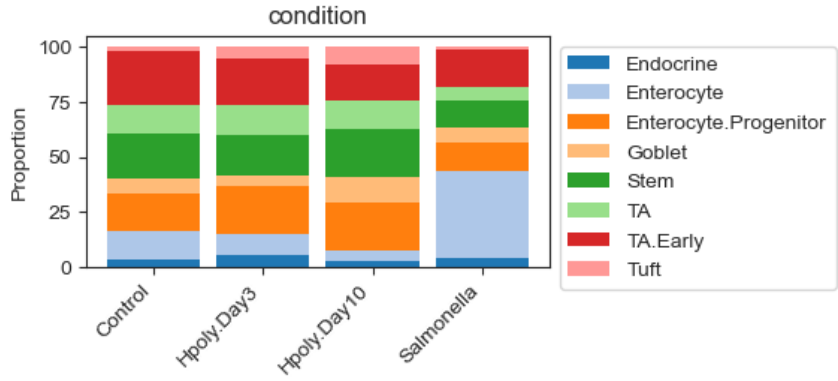


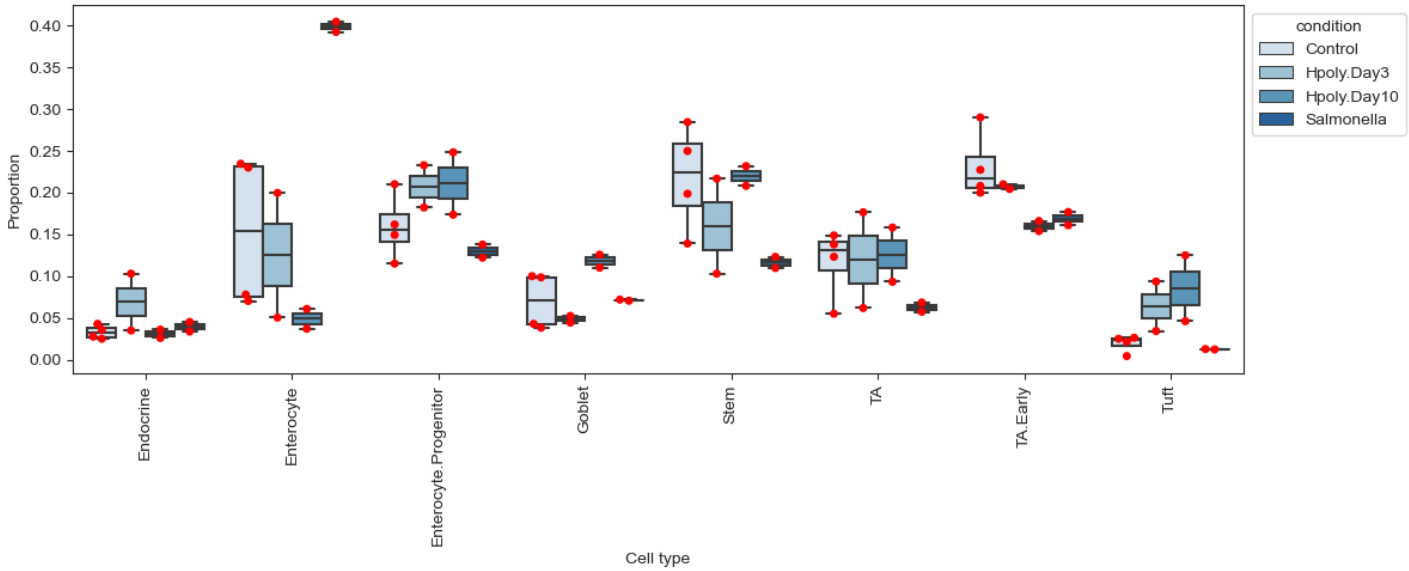


Source: https://https://www.sc-best-practices.org/

Differential abundance/compositional analysis: cluster-based

- Work on predefined clusters or cell types
- Per-sample cell proportions – can be derived by basic dplyr/pandas
- **Considerations**: adjustment for covariates, statistical tests
- Can use either frequentist (glm-based e.g. Yoshida et al. *Nature*. 2021 (glm), propeller or Bayesian approaches (e.g. **scCODA**)
- **Limitations:**
  - Relies on clustering quality
  - Can't detect **within-cluster shifts**

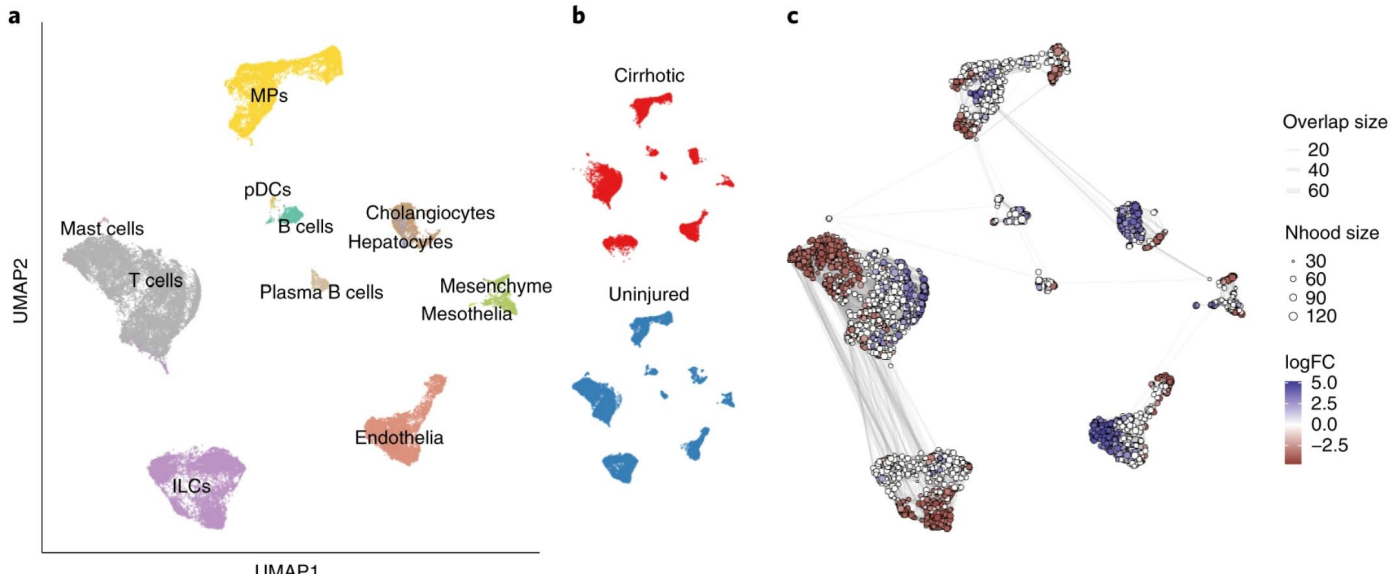


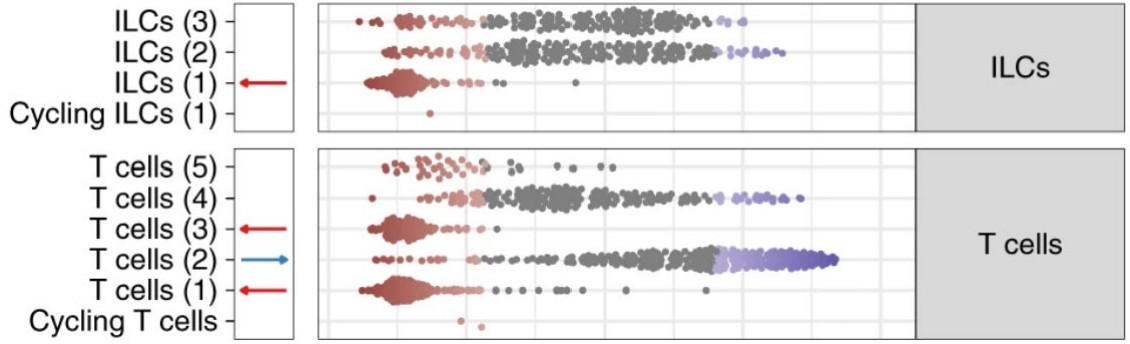


Source: https://https://www.sc-best-practices.org/

Differential abundance/compositional analysis: non-cluster-based

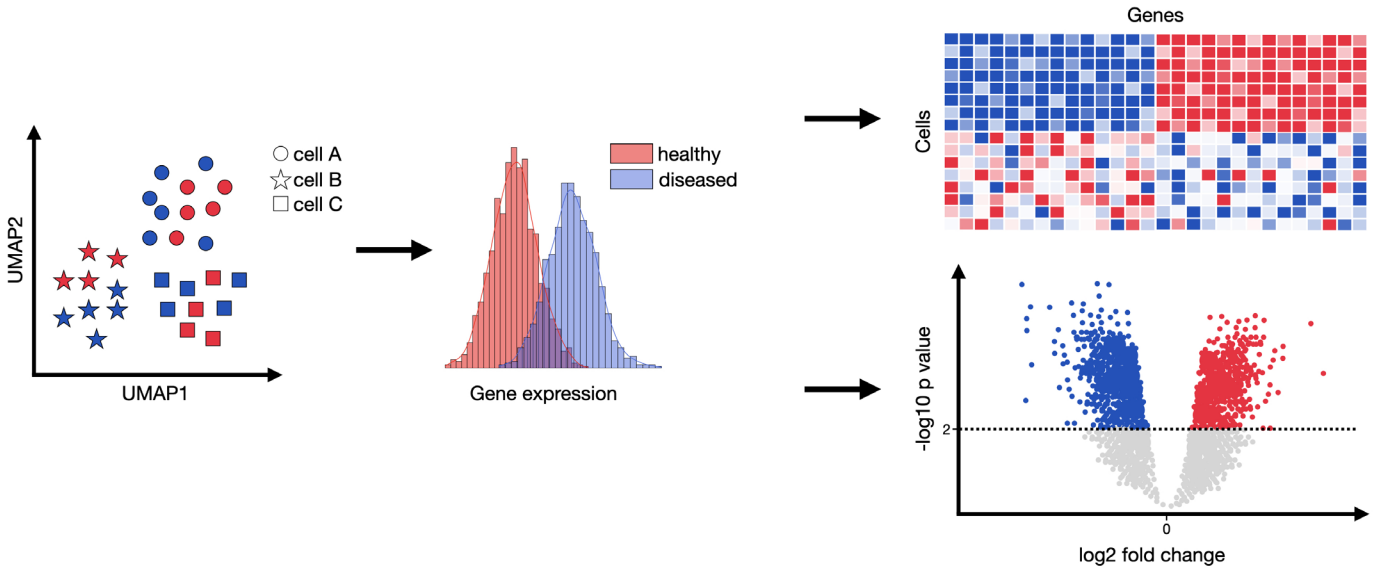
- Detect regions in expression space/latent space where **cell density differs between conditions, without relying on clusters**
- **Examples:**
  - **MiloR, MELD, Cydar, DA-seq**





Dann et al. *Nature Biotechnology*. 2022

Differential gene expression



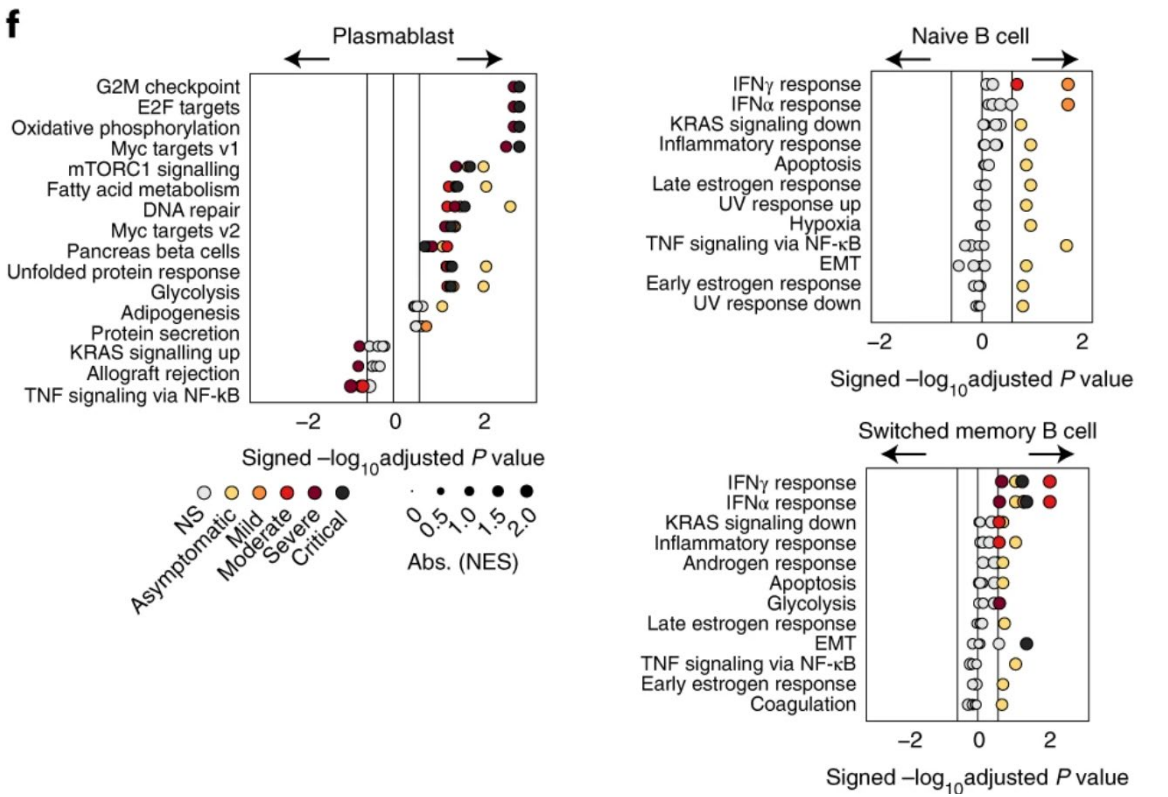
Source: https://https://www.sc-best-practices.org/

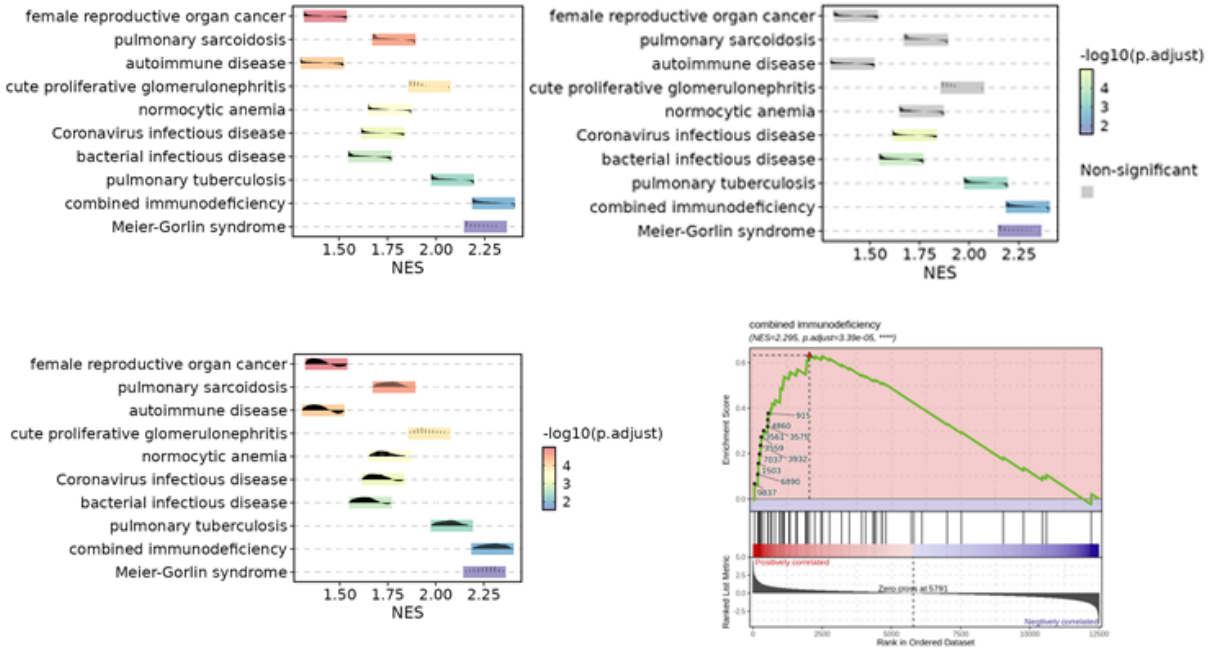
- **Consideration**
  - Single-cell level vs pseudobulk level
- **Single-cell level methods**
  - Analyze each cell as an observation -> more granular but noisy; inflated significances
  - Seurat/Scanpy (with options to choose specific methods e.g. **Wilcoxon**, MAST, t-test etc.)
- **Pseudobulk methods**
  - Aggregate cells per sample -> more robust, lower FDRs compatible with bulk RNA-seq tools e.g. **edgeR**, **DESeq2**, limma

Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

Pathway & gene set enrichment analysis



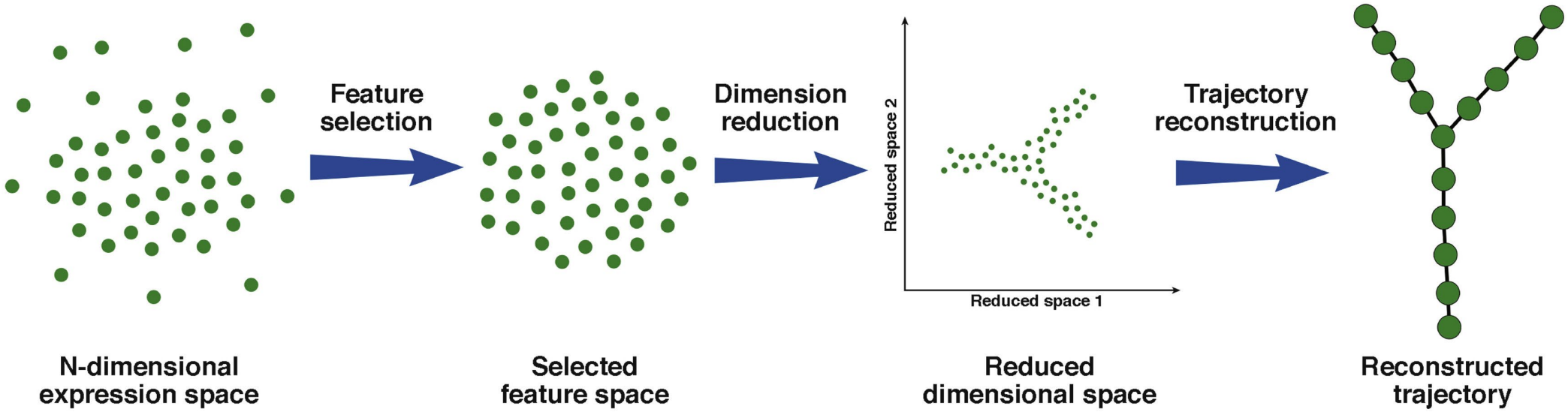


Stephenson et al. *Nature Medicine*. 2022

Source: https://pwwang.github.io/

- Examples: **fgsea, Enrichr, AUCell, GSVA, ClusterProfiler**
- Inference of gene regulatory networks (regulatory interactions between transcription factors (TFs) and target genes): **SCENIC**

Pseudotime/Trajectory analysis



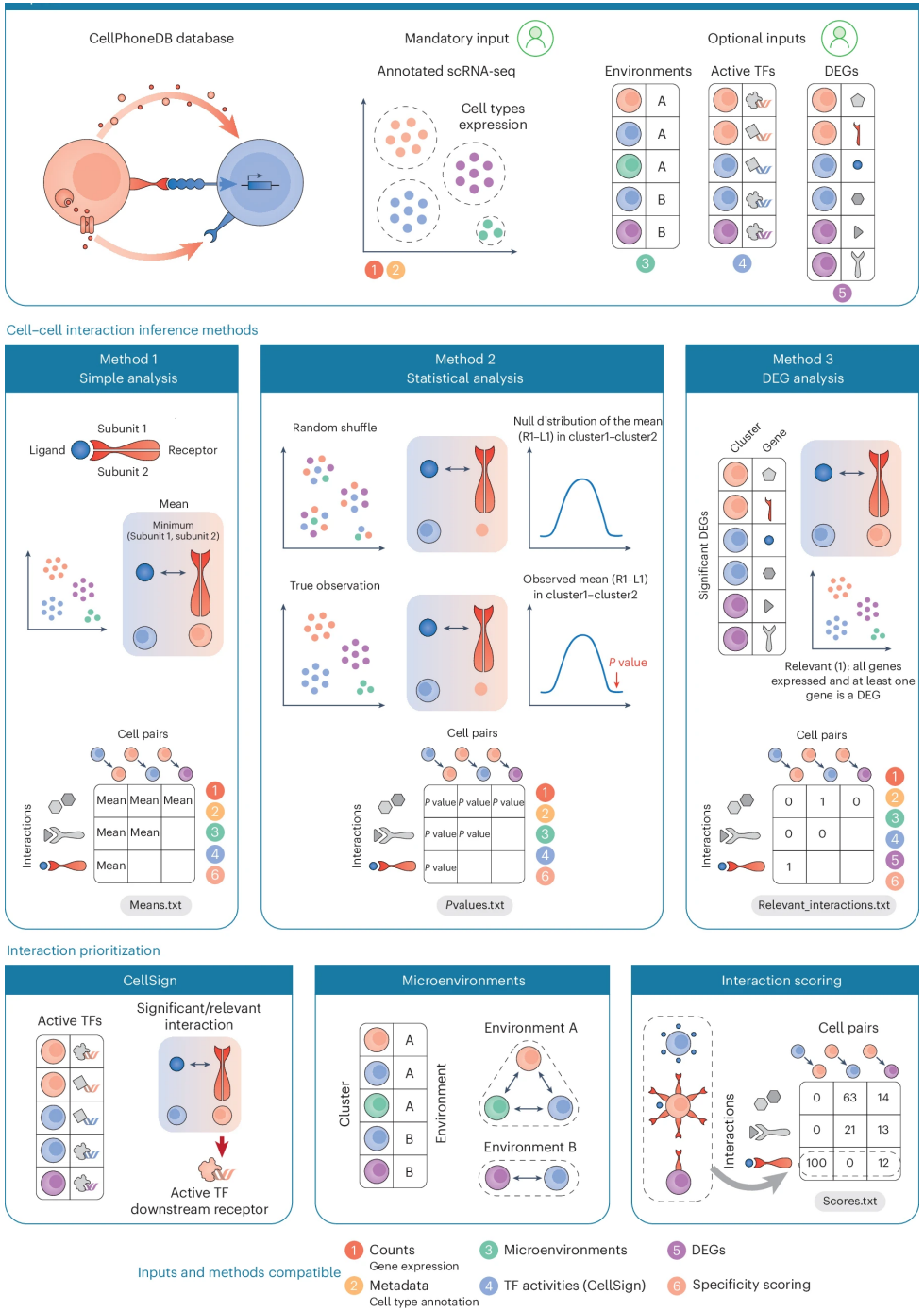
Example tools: **PAGA**, **monocle3**, Slingshot, Palantir

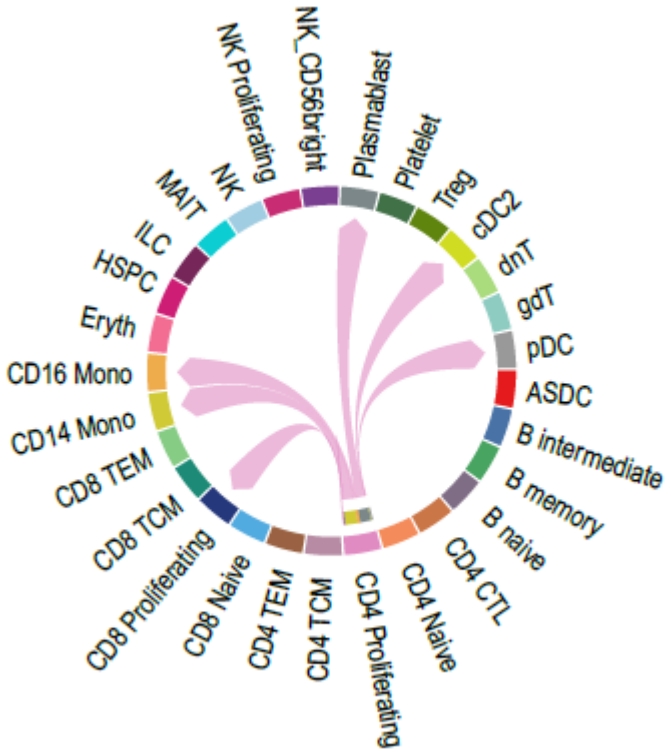
* RNA velocity-based: **scVelo** (add directionality)

Herring et al. 2018

Predicting cell-cell interactions

Tools infer signaling based on ligand–receptor expression across cell types. Often rely on curated databases.





Modeling IL-10/IL-10R communication during a viral infection (unpublished)

Example tools:

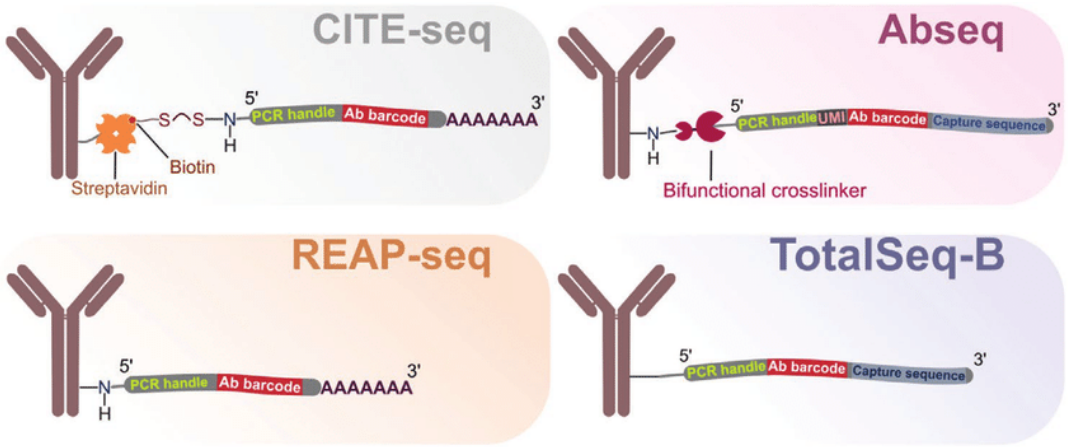
- **CellPhoneDB**
- **CellChat**
- **NicheNet**

Source: https://satijalab.org/seurat/

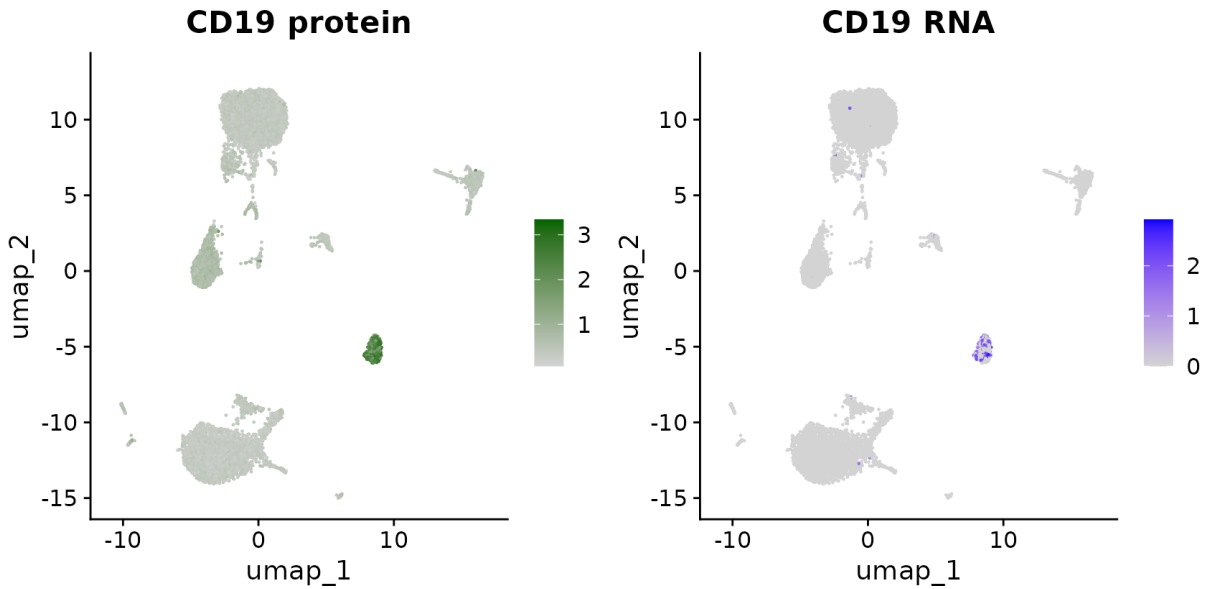
Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications

Multi-omics: protein expression

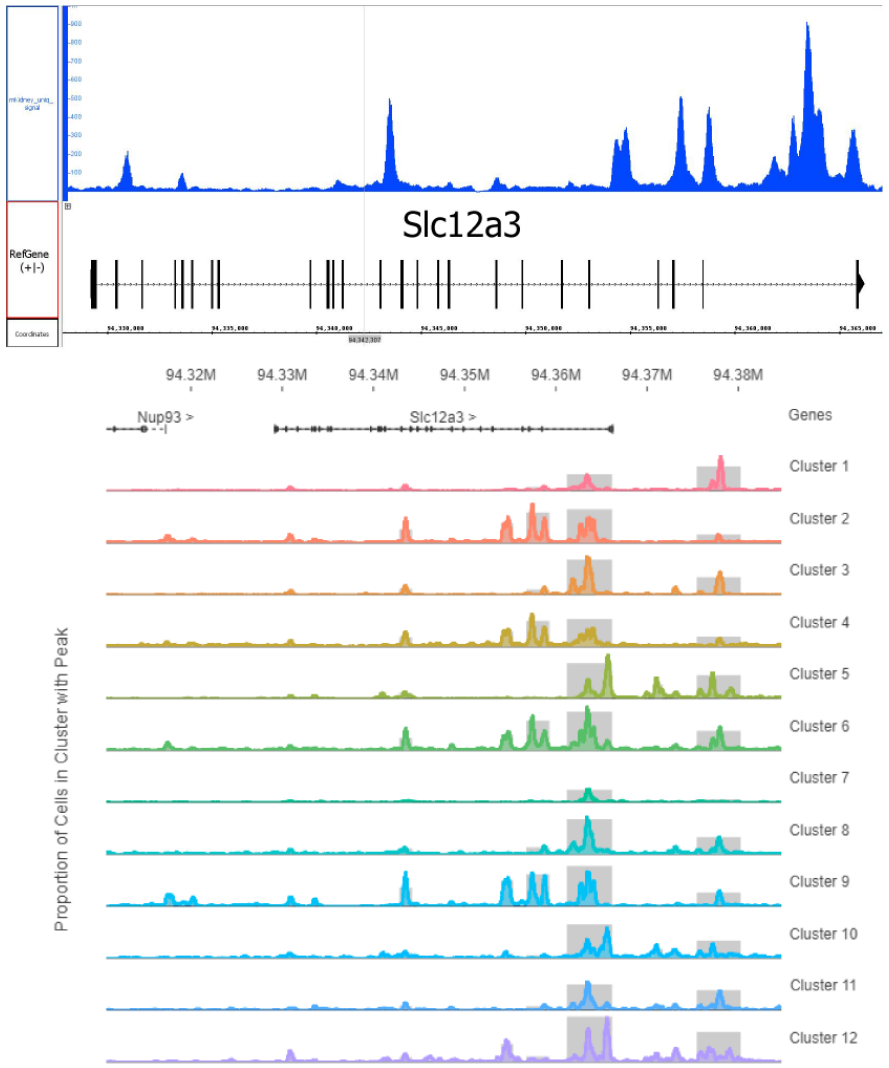


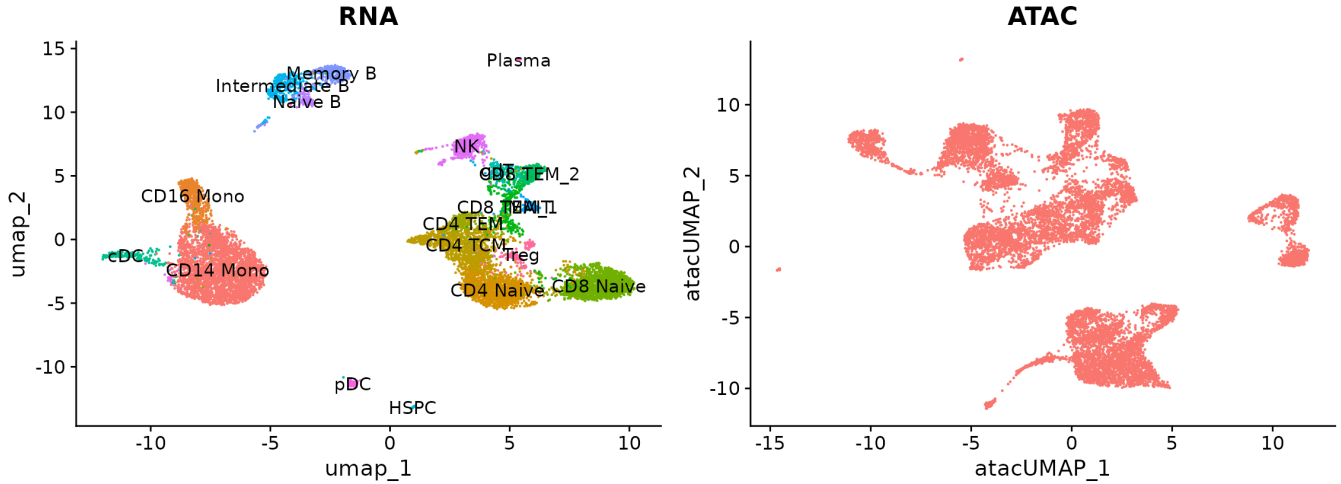
Matula et al. *Advanced Biosystems*. 2019



Source: https://satijalab.org/seurat/

Multi-omics: chromatin accessibility

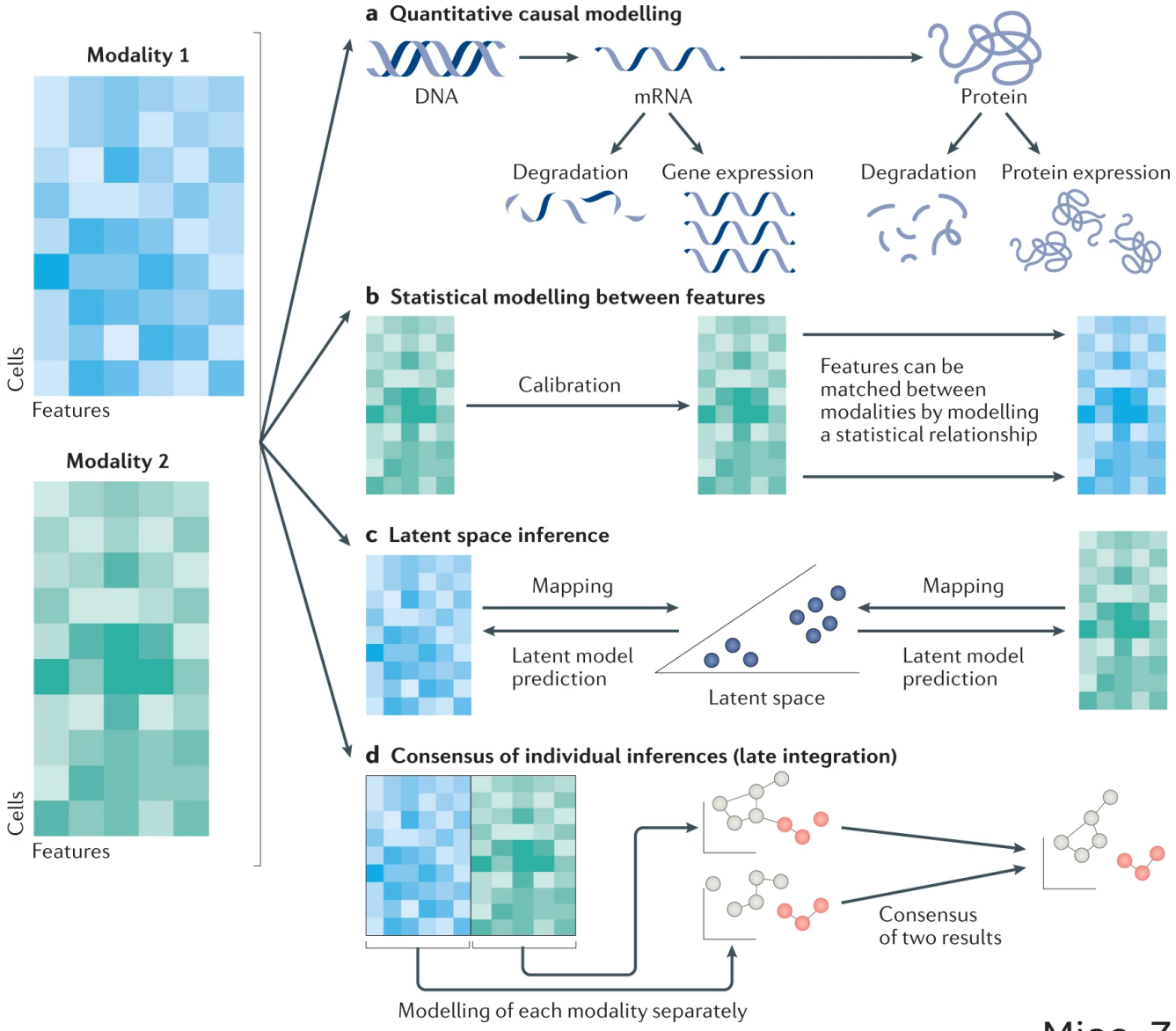




Source: https://satijalab.org/seurat/

https://www.activemotif.com/catalog/1299/single-cell-atac-seq-services

Frameworks for the integration of single-cell multi-omics data



Miao, Z. et al. *Nat Rev Nephrol*. 2021

Example of a latent space inference

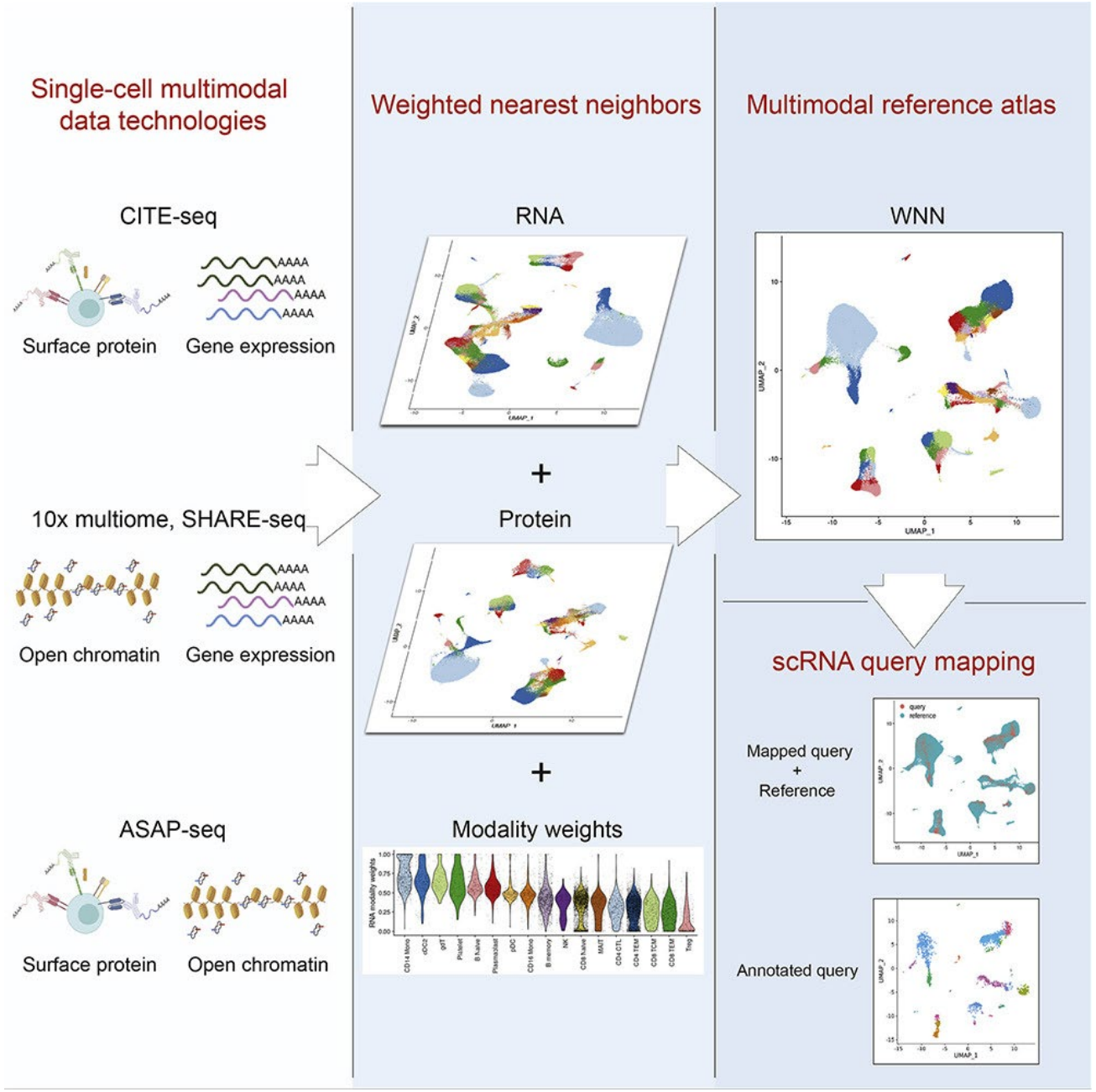
Cell

Volume 184, Issue 13, 24 June 2021, Pages 3573-3587.e29

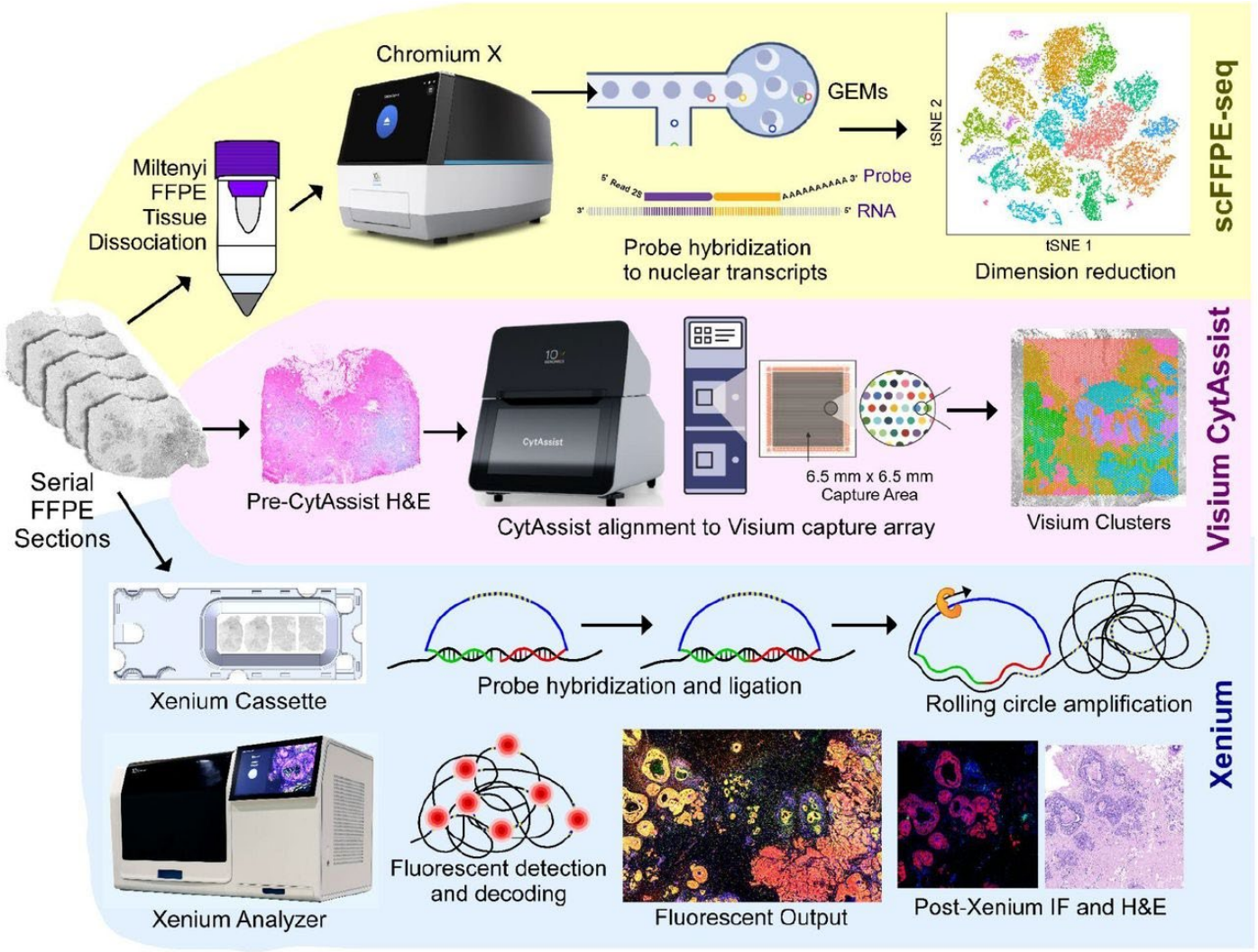
Resource

Integrated analysis of multimodal single-cell data

Yuhan Hao [1 2 10], Stephanie Hao [3 10], Erica Andersen-Nissen [4 5], William M. Mauck III [1], Shiwei Zheng [1 2], Andrew Butler [1 2], Maddie J. Lee [6], Aaron J. Wilk [6], Charlotte Darby [1], Michael Zager [7], Paul Hoffman [1], Marlon Stoeckius [3], Efthymia Papalexi [1 2], Eleni P. Mimitou [3], Jaison Jain [1], Avi Srivastava [1], Tim Stuart [1], Lamar M. Fleming [4], Bertrand Yeung [8], Angela J. Rogers [6]...Rahul Satija [1 2 11]

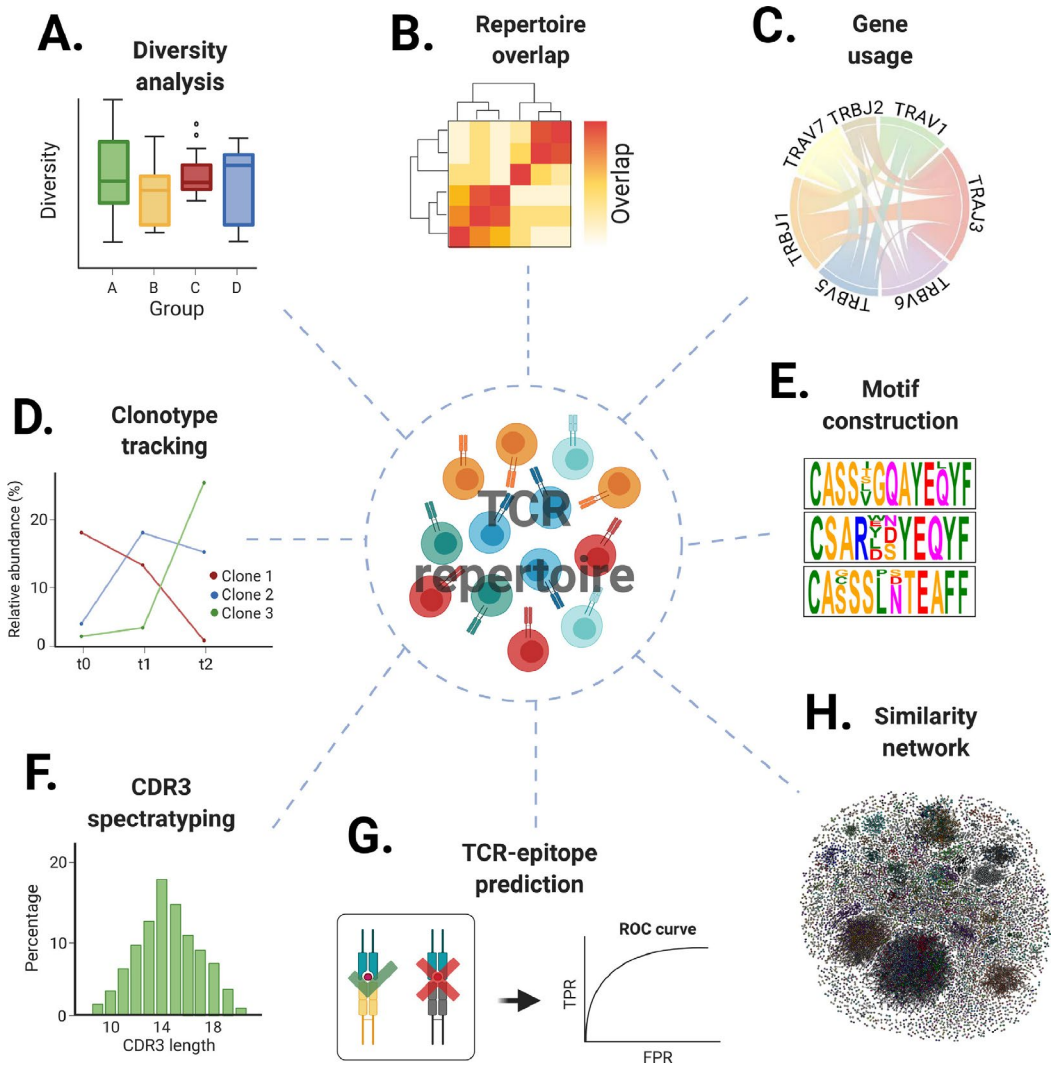


Integration with spatial omics



Janesick, Z. et al. *Nat Commun*. 2023

Immune receptor profiling (TCR/BCR Sequences)



Analysis tools examples:

- V(D)J gene /clonotype calling
  - **Cellranger V(D)J**
  - **Dandelion** (re-annotation of BCR)
- Downstream analysis
  - **scRepertoire** (R), **scirpy** (Python), **Dandelion** (Python)
  - These tools often **wrap or extend specialized utilities**

Valkiers et al. ImmunoInformatics, 2022

Science Immunology

Current Issue | First release papers | Archive | About

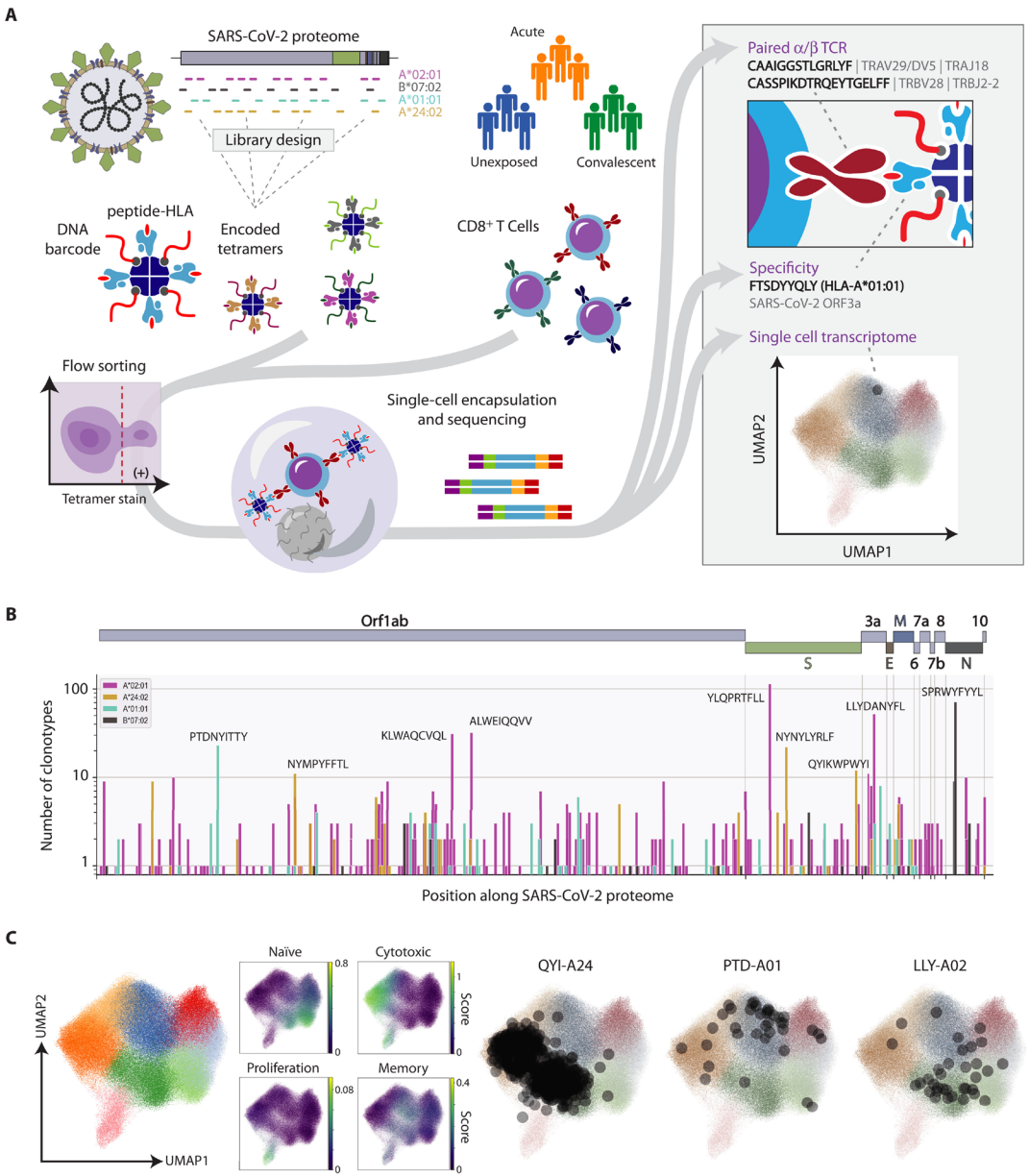
HOME > SCIENCE IMMUNOLOGY > VOL. 7, NO. 67 > ALLELIC VARIATION IN CLASS I HLA DETERMINES CD8 T CELL REPERTOIRE SHAPE AND CROSS-REACTIVE...

RESEARCH ARTICLE | CORONAVIRUS

Allelic variation in class I HLA determines CD8$^+$ T cell repertoire shape and cross-reactive memory responses to SARS-CoV-2

JOSHUA M. FRANCIS, DEL LEISTRITZ-EDWARDS, AUGUSTINE DUNN, CHRISTINA TARR, JESSE LEHMAN, CONOR DEMPSEY, ANDREW HAMEL, VIOLETA RAYON, GANG LIU, [...], AND DANIEL C. PREGIBON +25 authors Authors Info & Affiliations

- Single-cell transcriptomes
- Single-cell TCR sequences
- SAR-CoV-2 epitopes



Outline

- More exploration on cell annotation
- Additional data cleaning
- Adjusting for confounders
- Quantitative comparison across groups
- “Functional” inference
- Extensions to multi-omics
- Real-world applications


AP SCI-BIO |June 20-21, 2025 | Thailand Bangkok
AP 2025 SCI-BIO
Asia Pacific Spatial and Single-Cell Biology Innovation Conference
Suzuki Yutaka
Faisal Mahmood
Kris Chatamra
Plenary Speakers
Pimpayao Sodsai
Arutha Kulasinghe
Ruby Yun-Ju Huang
Joe Poh Sheng Yeong
Ramanuj DasGupta
Chanitra Thuwajit
Sizun Jiang
Punn Augsornworawat
Quan Nguyen
Ankur Sharma
Methichit Wattanapanich
Xi Chen
Ponpan M. Choopong
Shyam Prabhakar
Varodom Charoensawan
Woong Yang Park
Dong Sung Lee
Worachart Lert-itthiporn
Theerawit Wilaiprasitporn
Saranyoo Ponnikorn
Waradon Sungnak
Jong-Eun Park
Tim Stuart
Sampornnat Sampattavanich
Natini Jinawath

Some Resources:

- Single-cell best practices (Python-based): https://www.sc-best-practices.org/
- Benchmarking trajectory analysis methods: Saelens et al. *Nat Biotechnol*. 2019
- Benchmarking doublet detection/removal methods: Xi and Li. *Cell Syst*. 2021
- Benchmarking DE methods: Squair et al. *Nat Commun*. 2021
- Benchmarking integration methods: Luecken et al. *Nat Methods*. 2022
- Benchmarking cell-cell communication methods: Dimitrov et al. *Nat Commun*. 2022
- Benchmarking DA methods: Yi et al. *Genome Biol*. 2024