Downstream Analysis of Single Cell RNA-Seq Data

Jenny Wu

Director of Bioinformatics

Genomics Research and Technology Hub

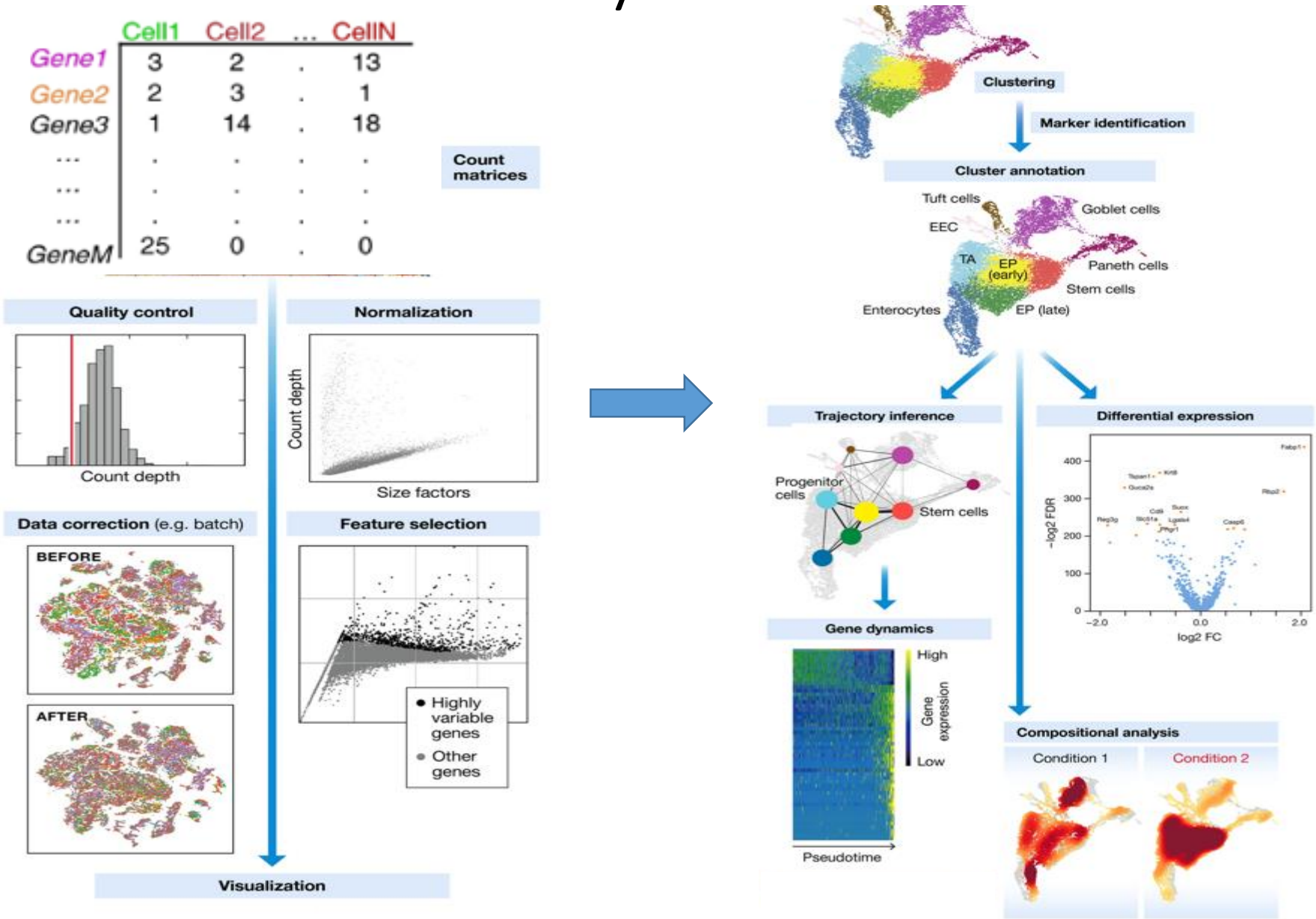
Chao Family Comprehensive Cancer Center

UC Irvine

Outline

- Downstream analysis pipeline
- Seurat Workflow
  - Data import
  - QC and filtering
  - Feature Selection
  - Dimension reduction and clustering
  - Data Visualization and integration
- Data Structure
- Advanced topics: RNA velocity, inter cellular communication, spatial transcriptomcis

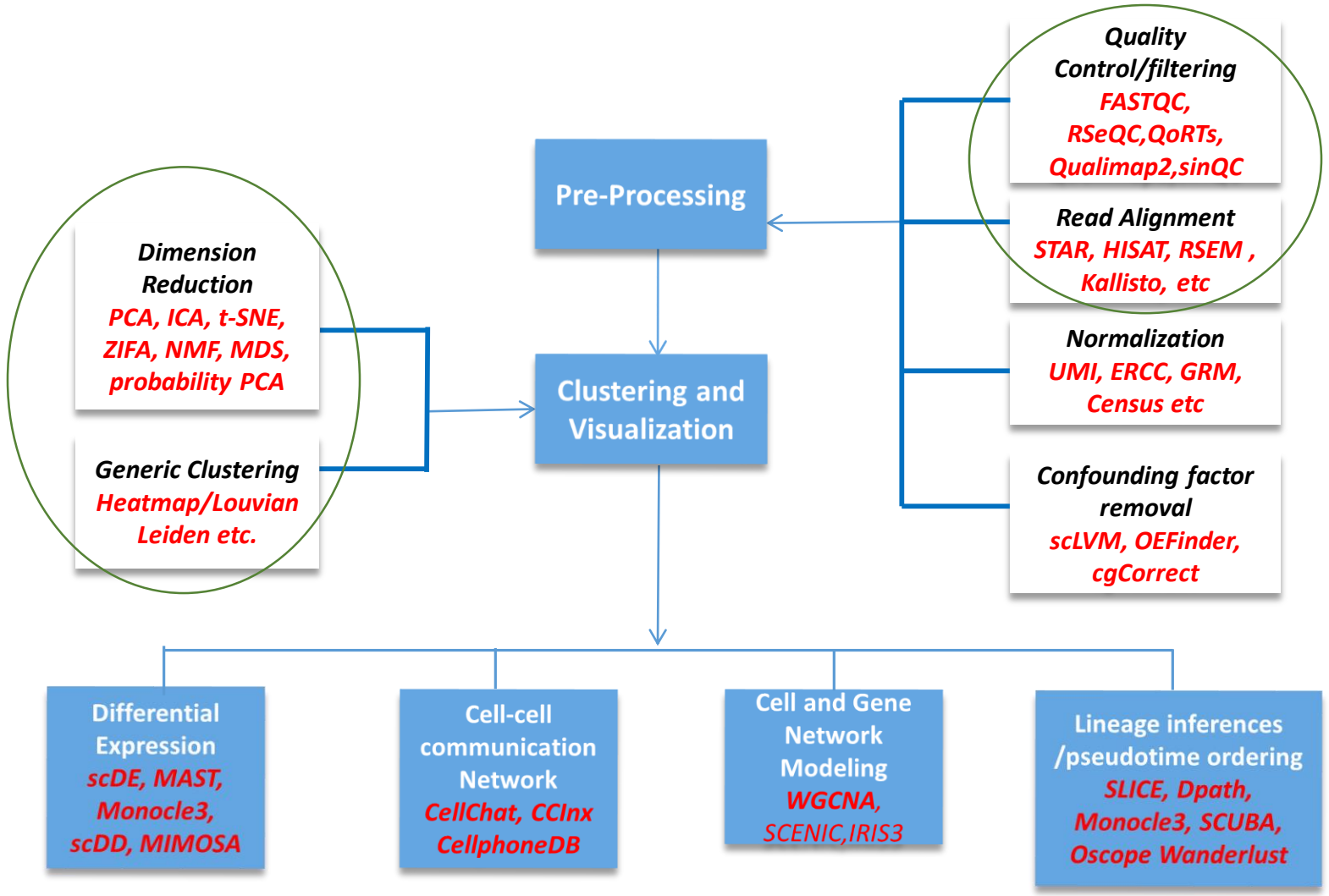
Downstream Analysis



Luecken *et al*, 2019

End to End tools for Single Cell Analysis

scRNA-seq Data Analysis Tools



CellRanger, bustools, Parsebio pipeline

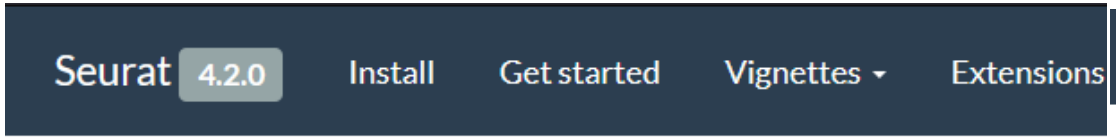
R packages: Seurat4, Monocle3, Pagoda, scater, scran
Integrated analysis: Scell, fast Project

Seurat Workflow

1. **Choose cells for downstream processing**
   - **Single cell QC**
2. **Choose genes for downstream processing**
   - **Identify genes with high dispersion**
3. **Dimensional reduction**
   - **Combat sparsity by pooling information across genes**
4. **Clustering**
   - **Identify groups of cells with similar transcriptomes**
5. **Visualization and interpretation**
   - **tSNE, marker discovery, interpretation**

Data Import

- Load count matrix into Seurat:

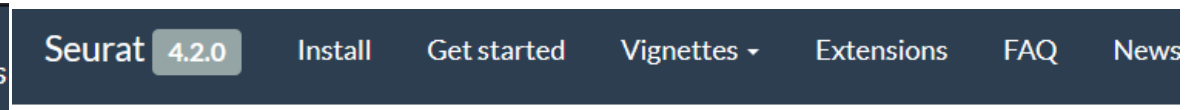


Load in data from 10X

Source: `R/preprocessing.R`

Enables easy loading of sparse data matrices provided by 10X genomics.

```
Read10X(
  data.dir,
  gene.column = 2,
  cell.column = 1,
  unique.features = TRUE,
  strip.suffix = FALSE
)
```



Read output from Parse Biosciences

Source: `R/convenience.R`

Read output from Parse Biosciences

```
ReadParseBio(data.dir, ...)
```

Arguments

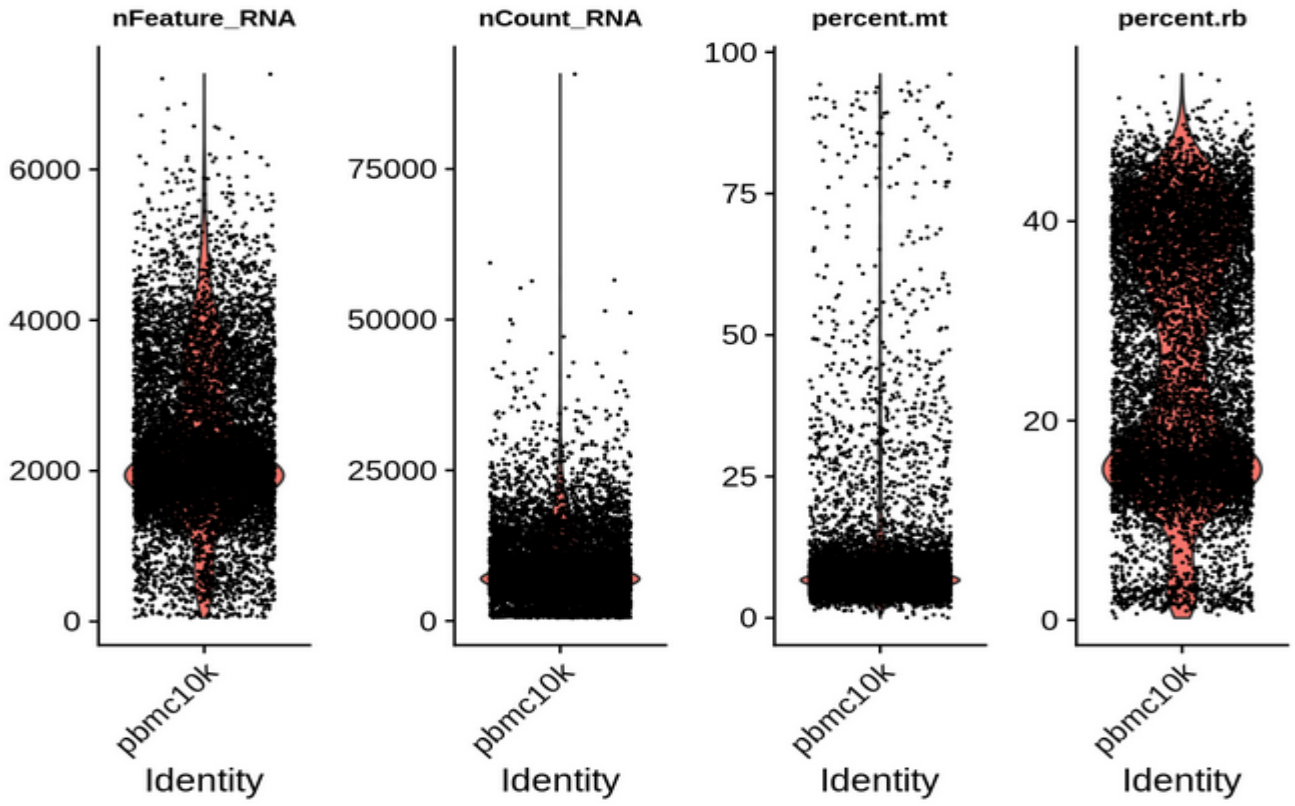
data.dir

Directory containing the data files

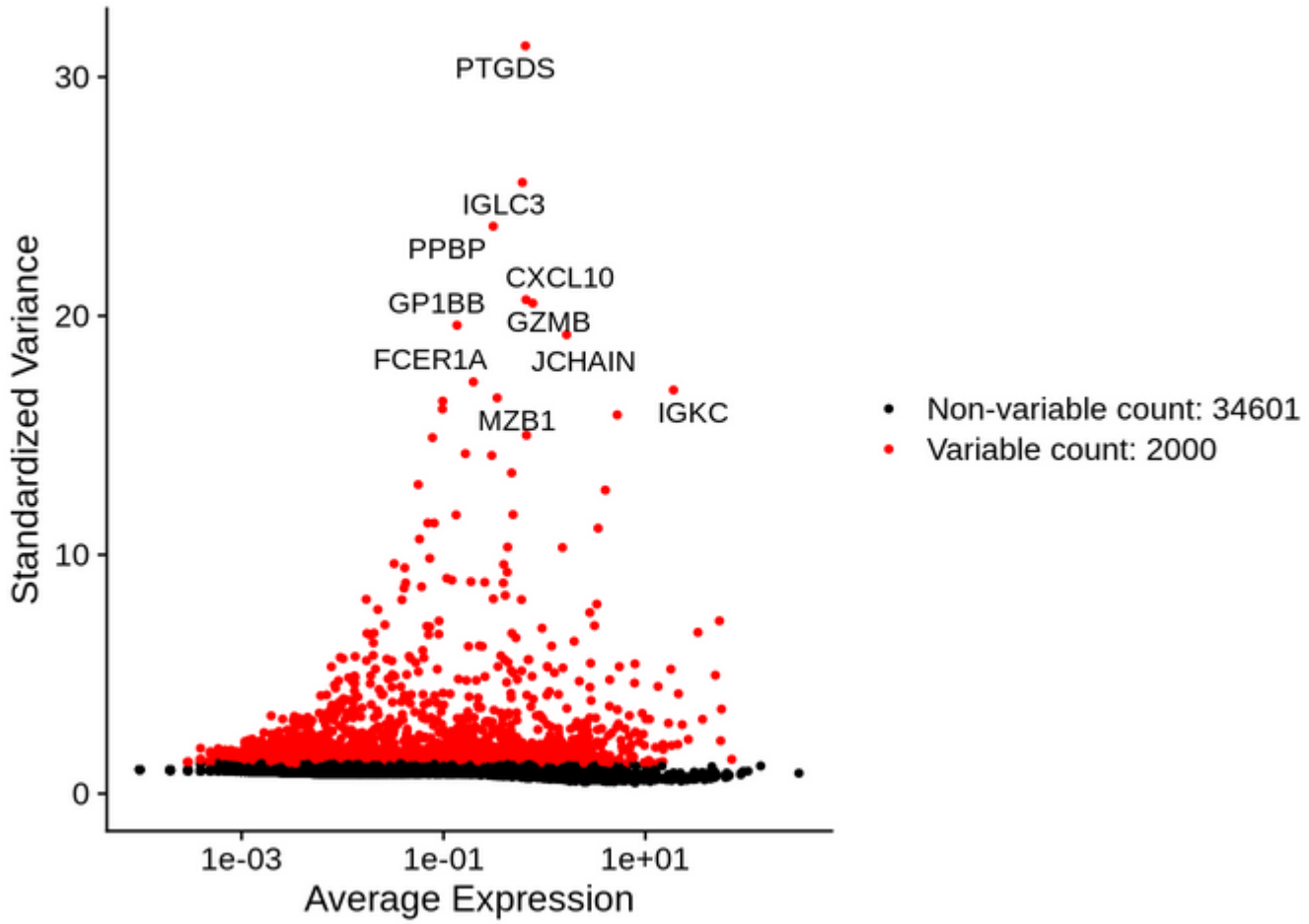
...

Extra parameters passed to `ReadMtx`

QC and Filtering

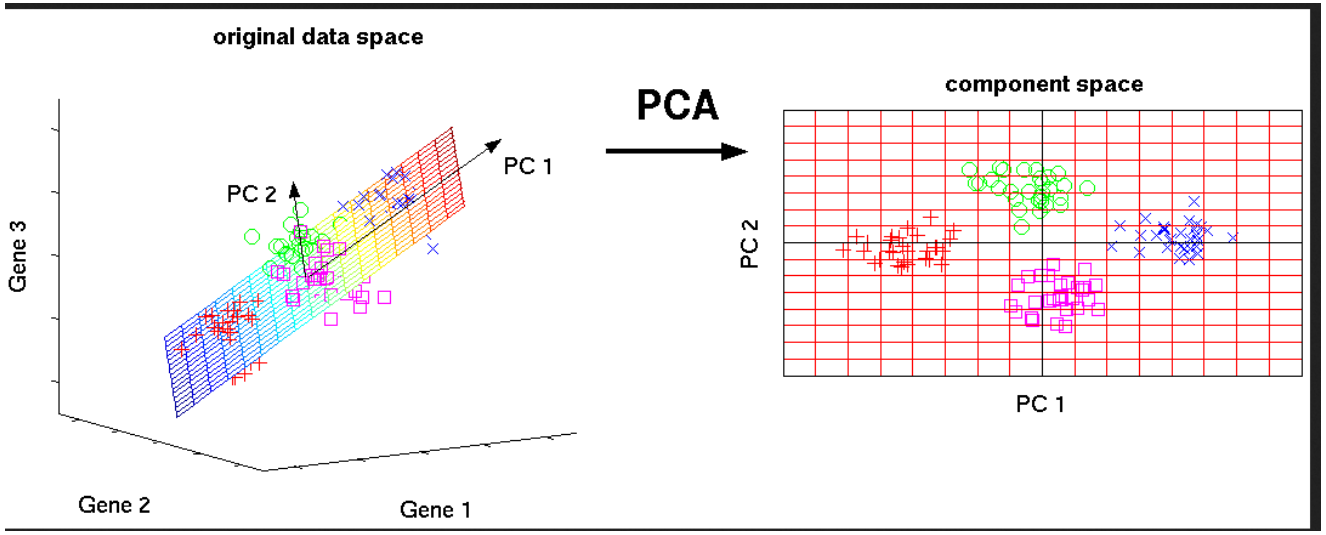


Feature Selection: HVGs



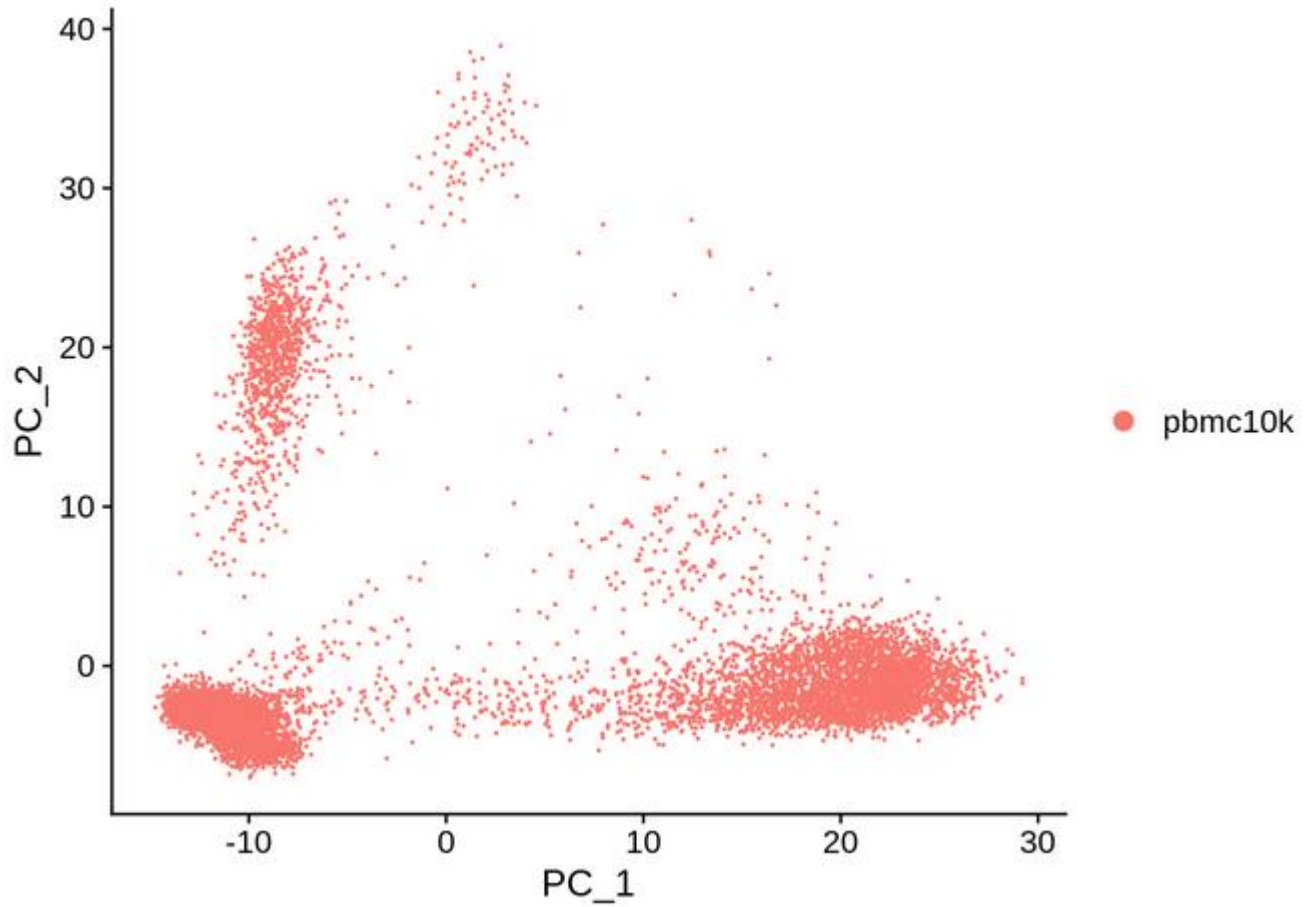
Linear Dimension Reduction

- Principle Component Analysis (PCA) is a standard technique for visualizing high dimensional data and for data pre-processing.

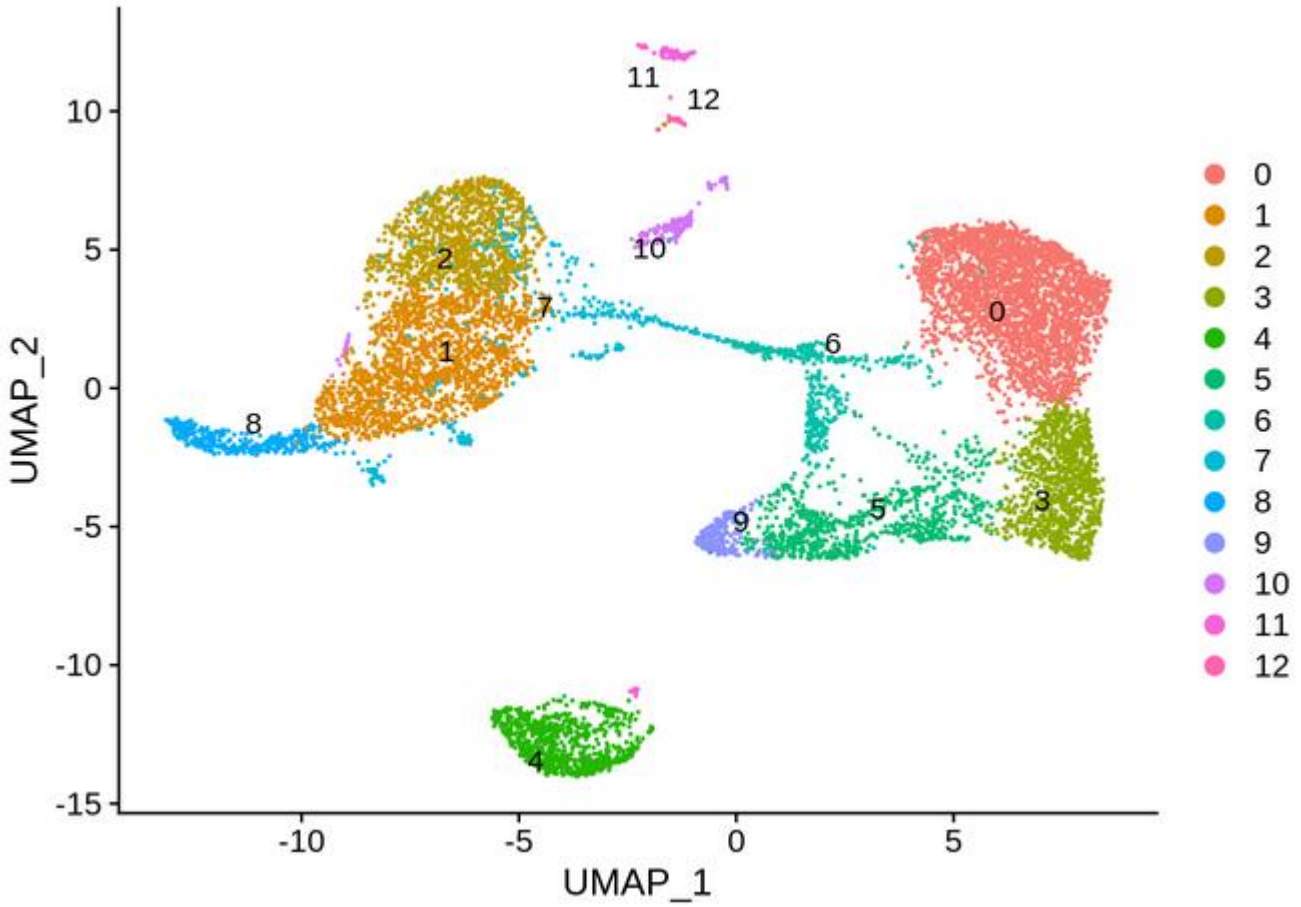


Matthias Scholz, 2015

Dimension Reduction : PCA

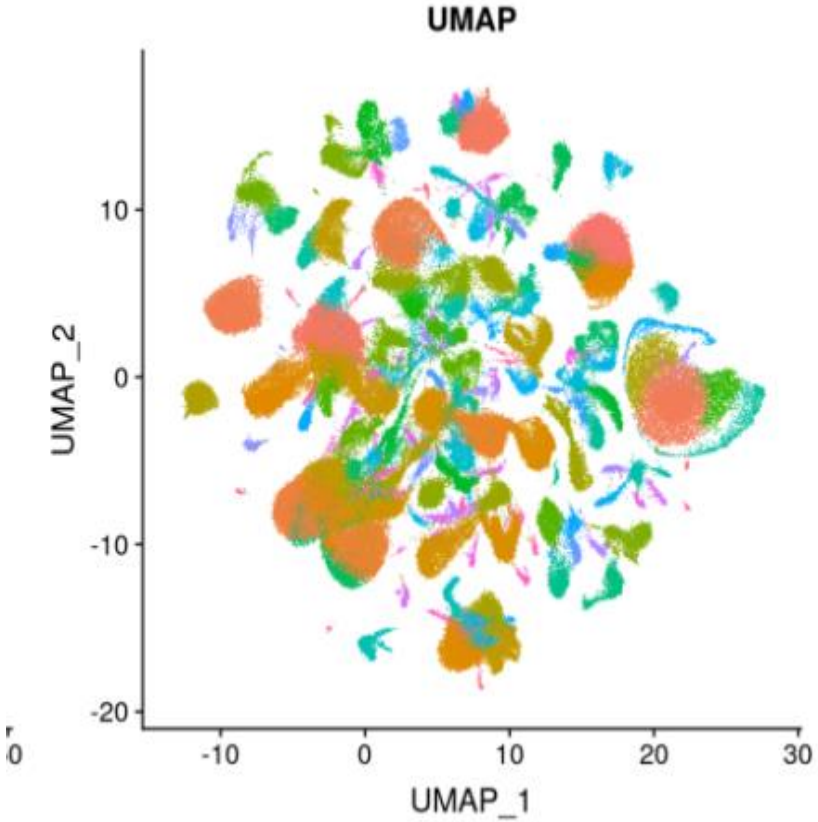


Dimension Reduction : UMAP

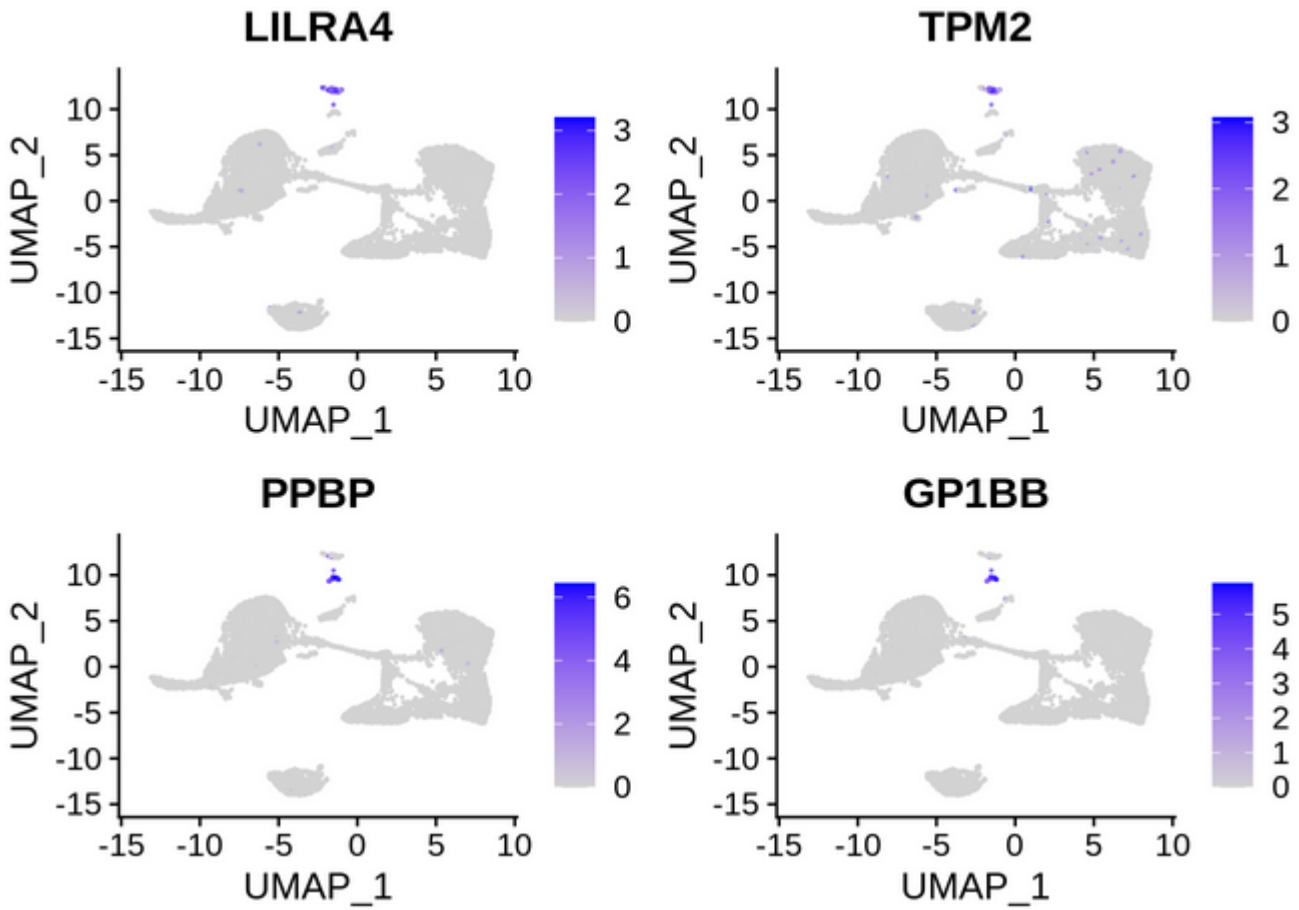


Clustering

- Distance/similarity metrics
- Algorithm choices
  - K means
  - Hierarchical
  - Density based
  - Graph partition
  - …

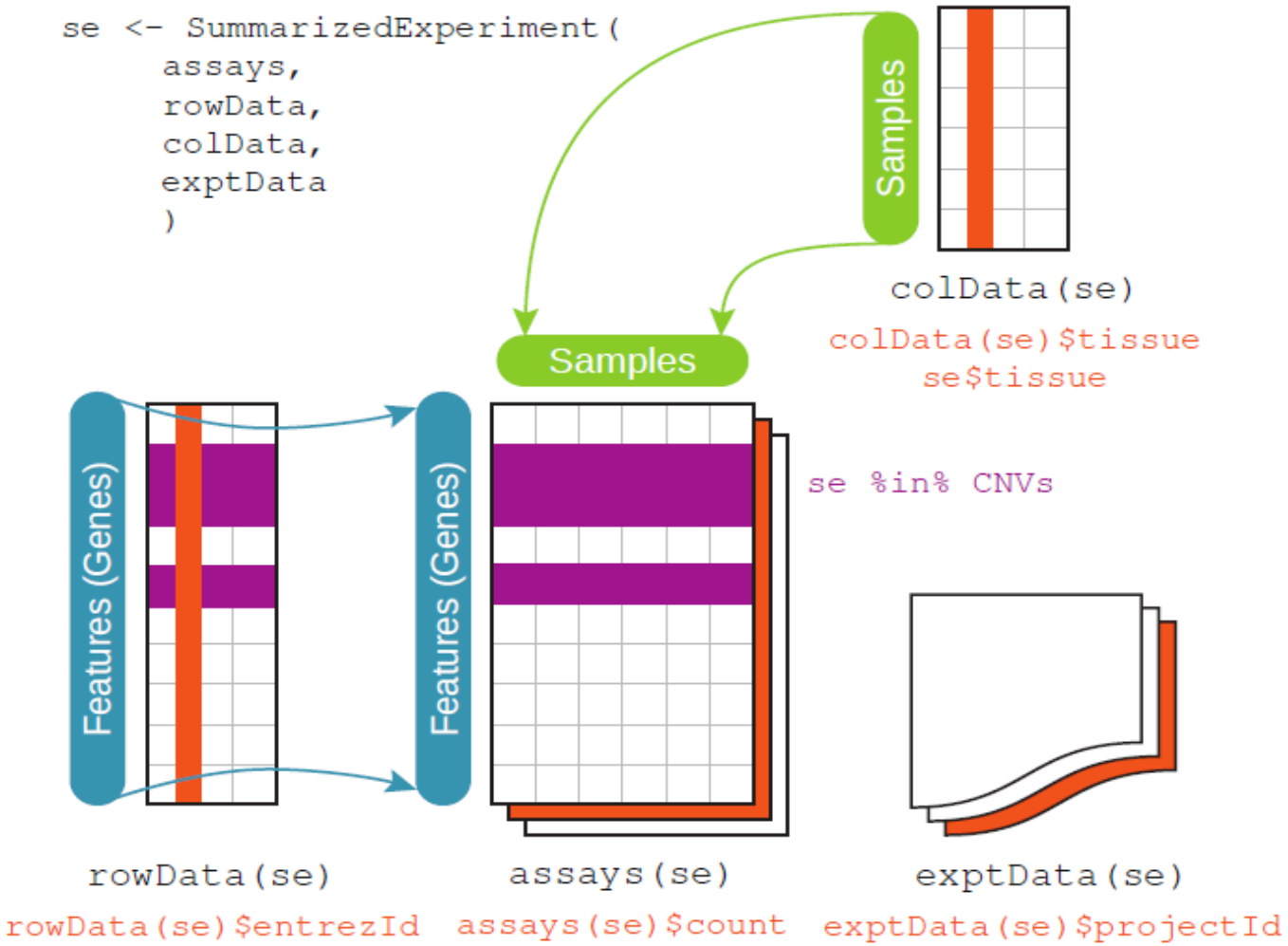


Visualization

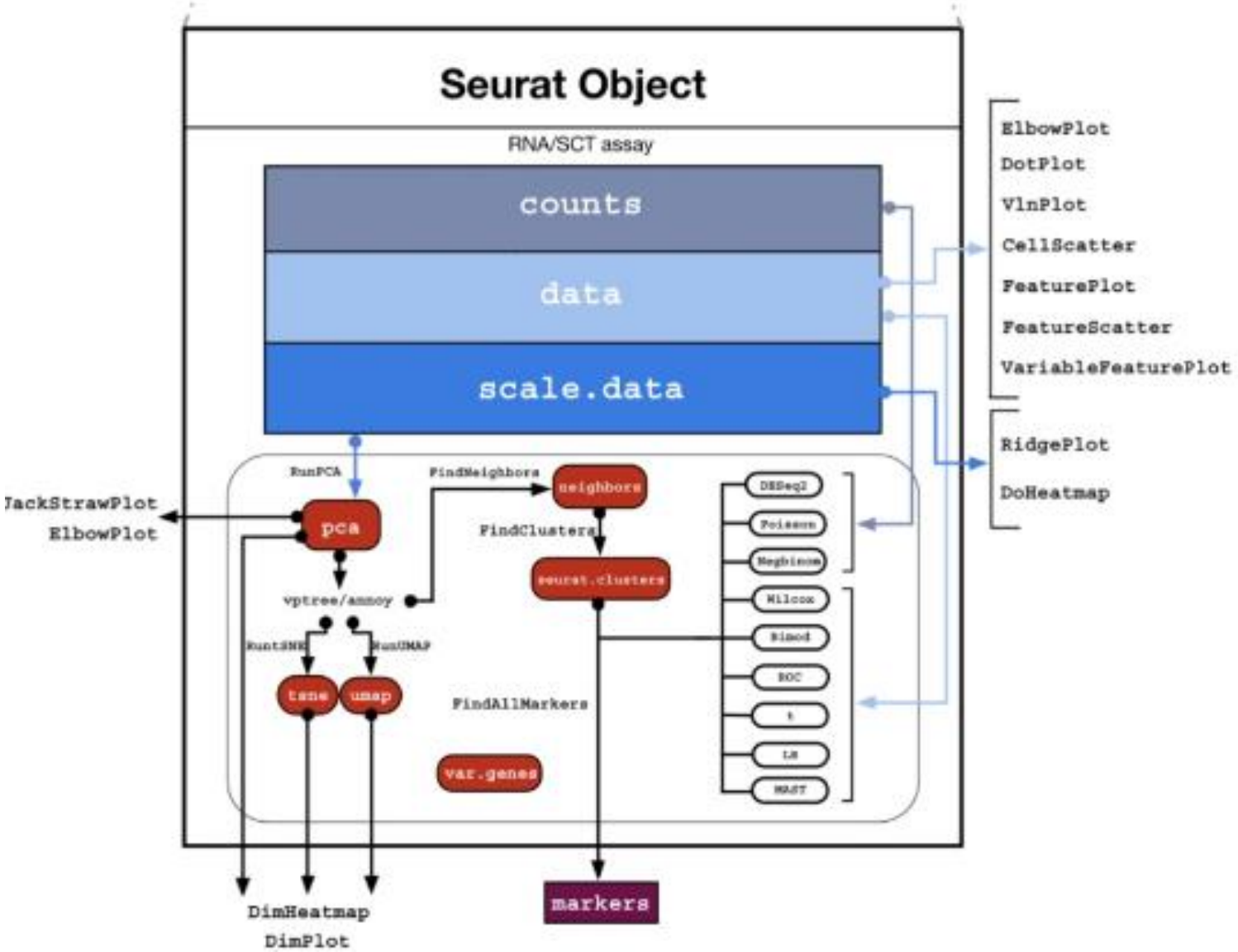


Gene Expression in R/Bioconductor:

SummarizedExperiment/ExpressionSet

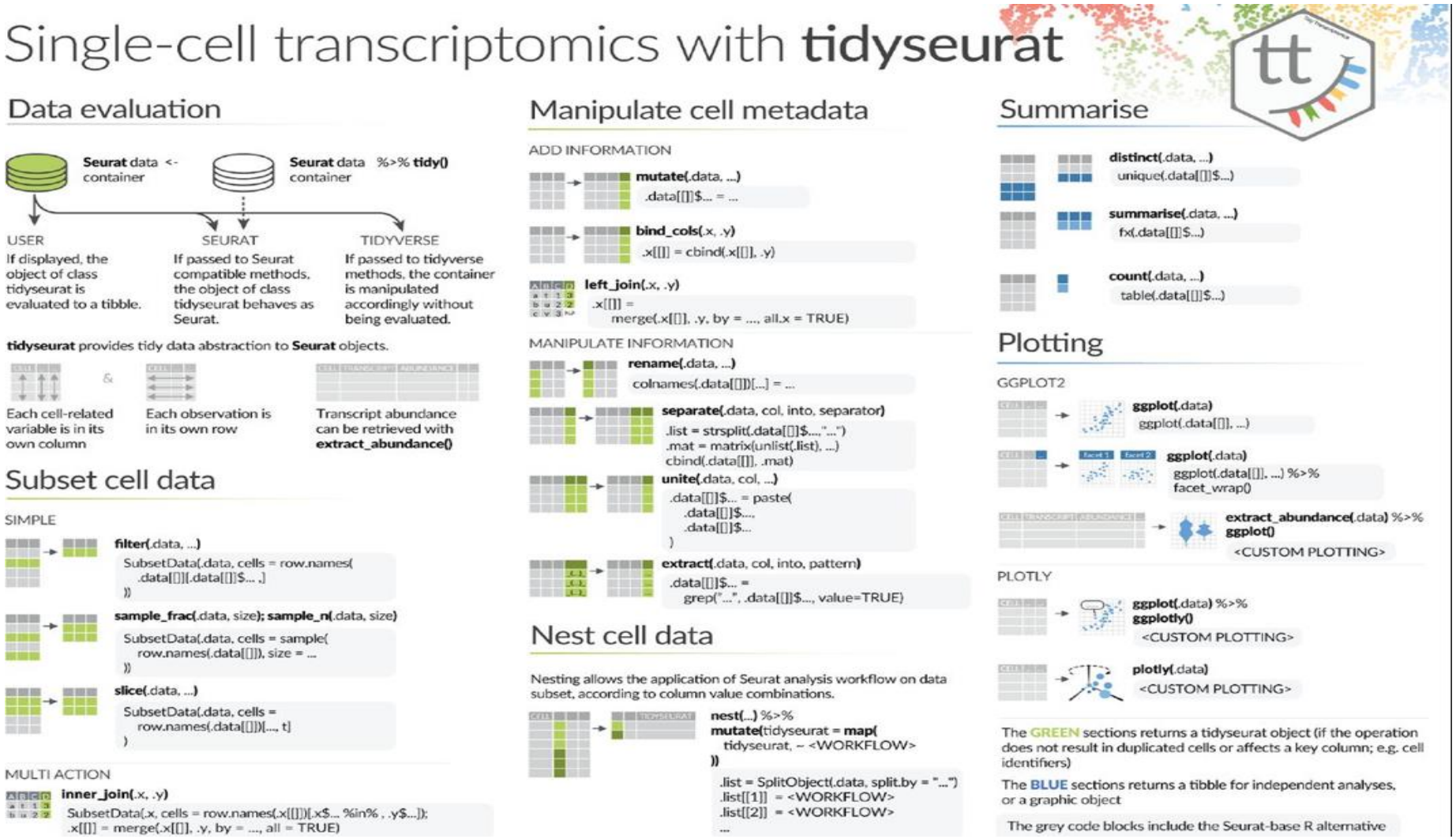


Seurat Data Structure



Pachter *et al*, 2022

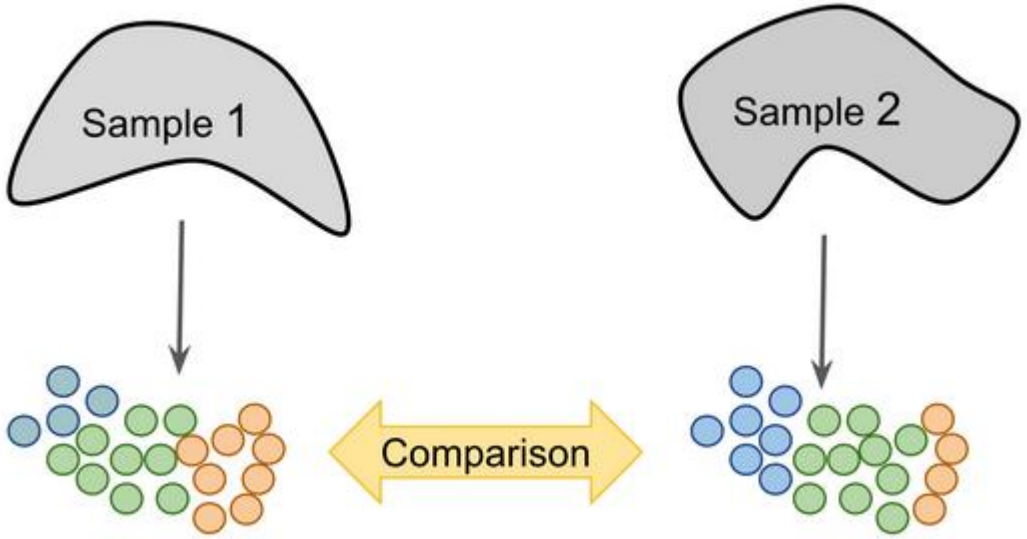
Advanced Seurat: tidyseurat



https://www.biorxiv.org/content/10.1101/2021.03.26.437294v1.full

Single Cell Data Integration

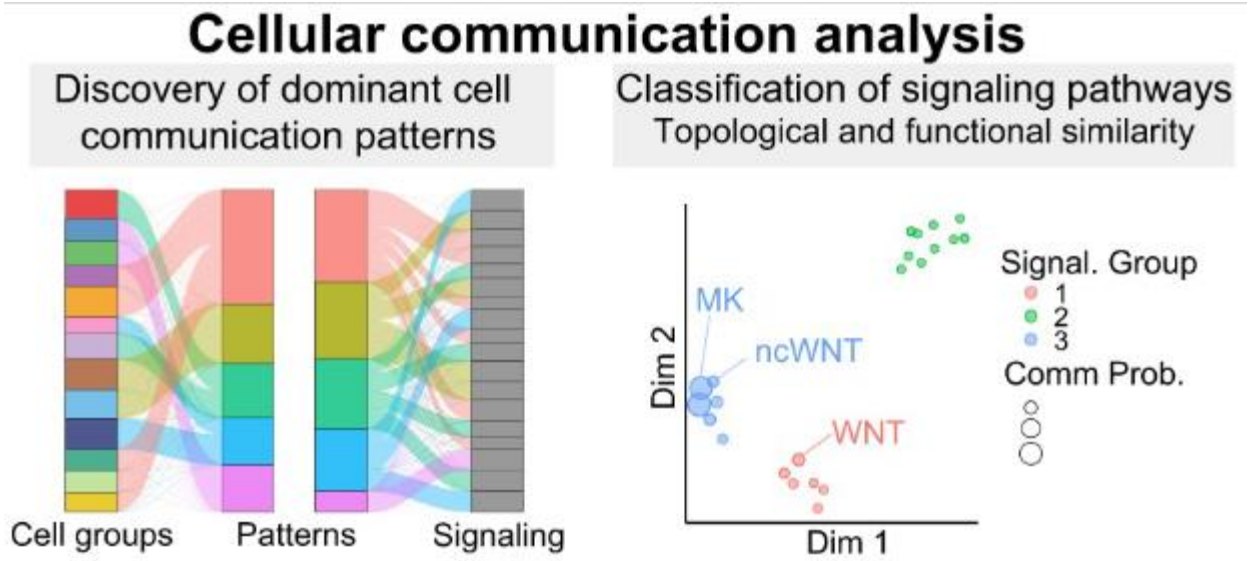
- Allows for analyzing single cell samples from different technologies and conditions at the same time
  - Harmony
  - LIGER
  - Seurat



Cell-Cell Communication Network

- Use network-based analysis and ligand-receptor biology to predict signaling between groups of cells

CellChat
Integrated Web-based CellChat Explorer

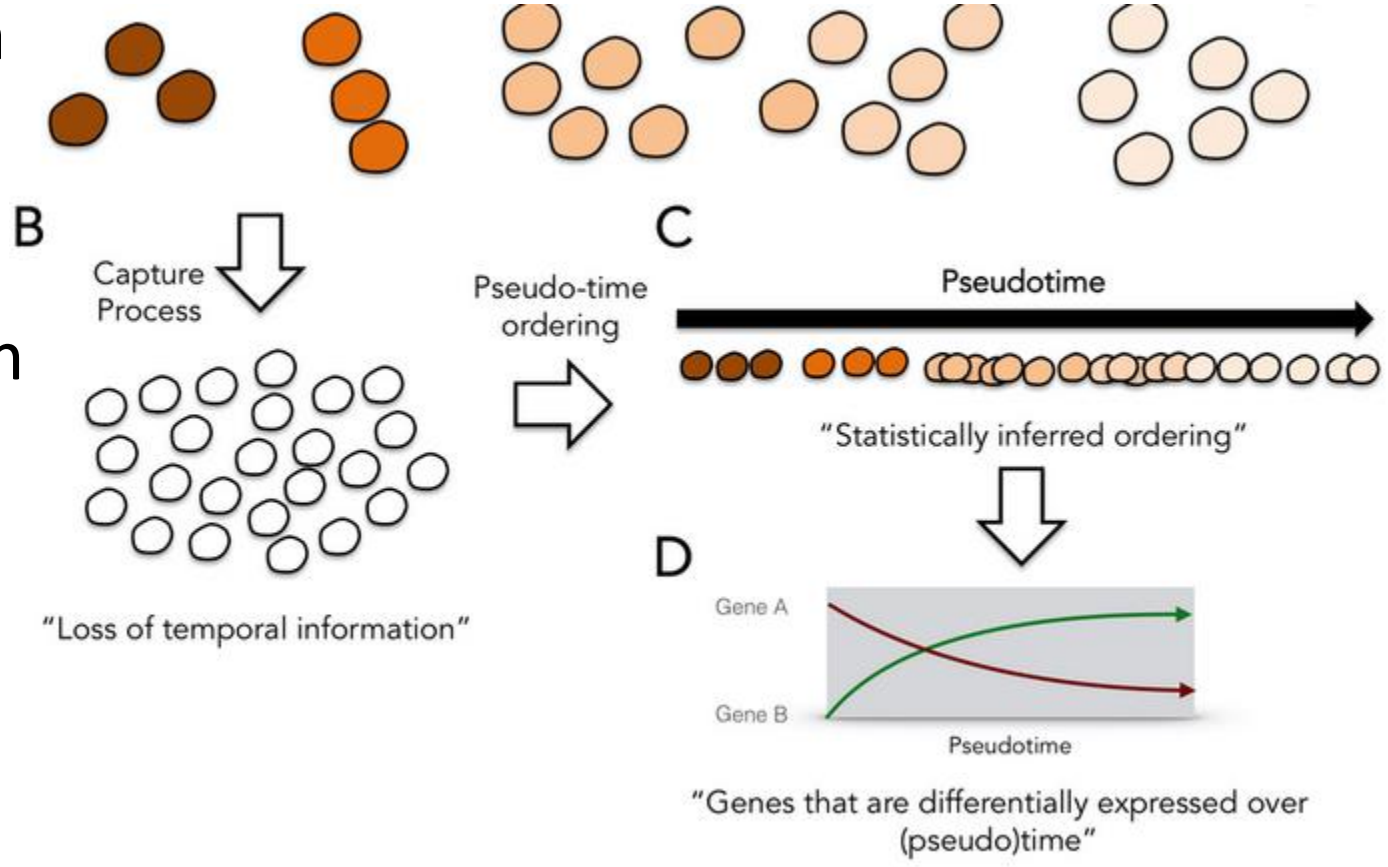


- ➢CellChat
- ➢SingleCellSignalR
- ➢CCInx

Trajectory Inferences

Determine the pattern of a dynamic process experienced by cells and then arrange cells based on their progression through the process.

- Monocle3
- URD
- PAGA



RNA Velocity

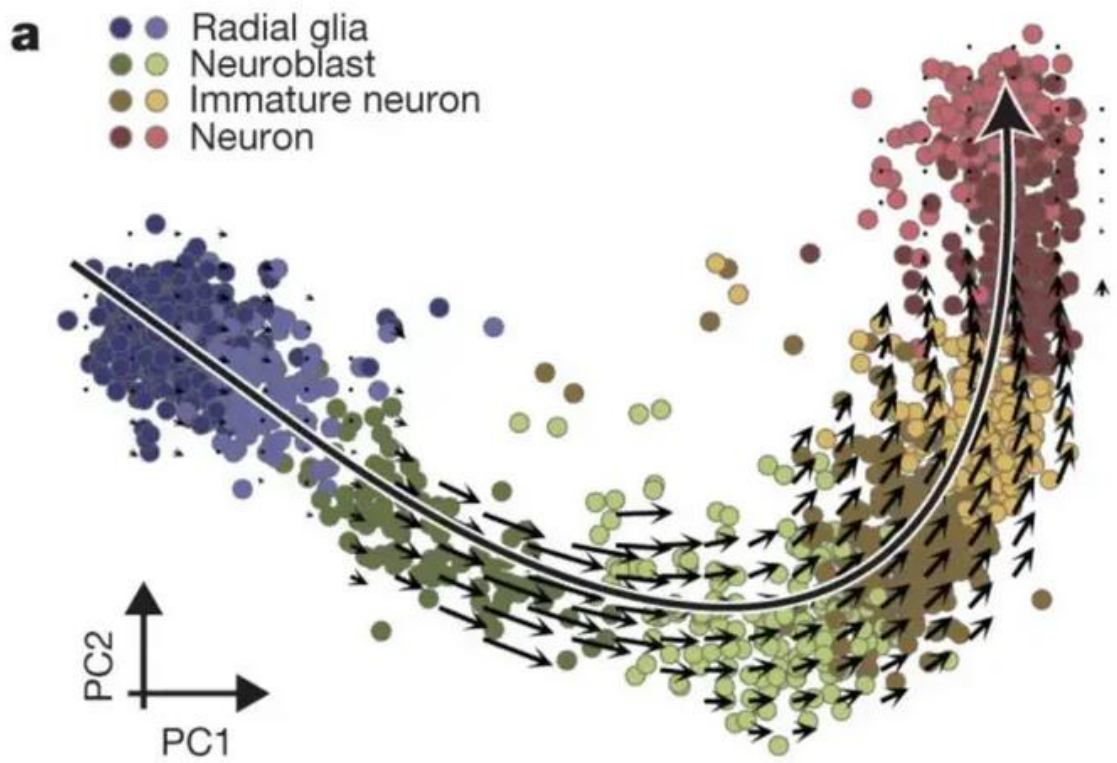
- The cell's internal compass: recovers directed information by distinguishing newly transcribed pre-mRNAs (unspliced) from mature mRNAs (spliced) reads

velocyto

Estimating RNA velocity in single cell RNA sequencing datasets

pypi v0.2.5 | downloads 154k | docs passing

scVelo - RNA velocity generalized

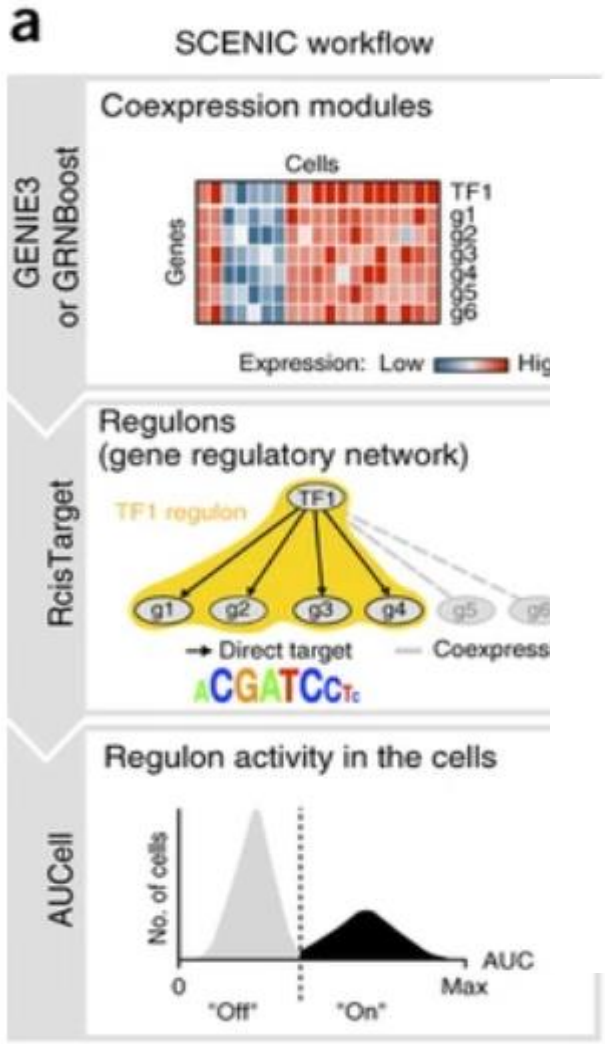


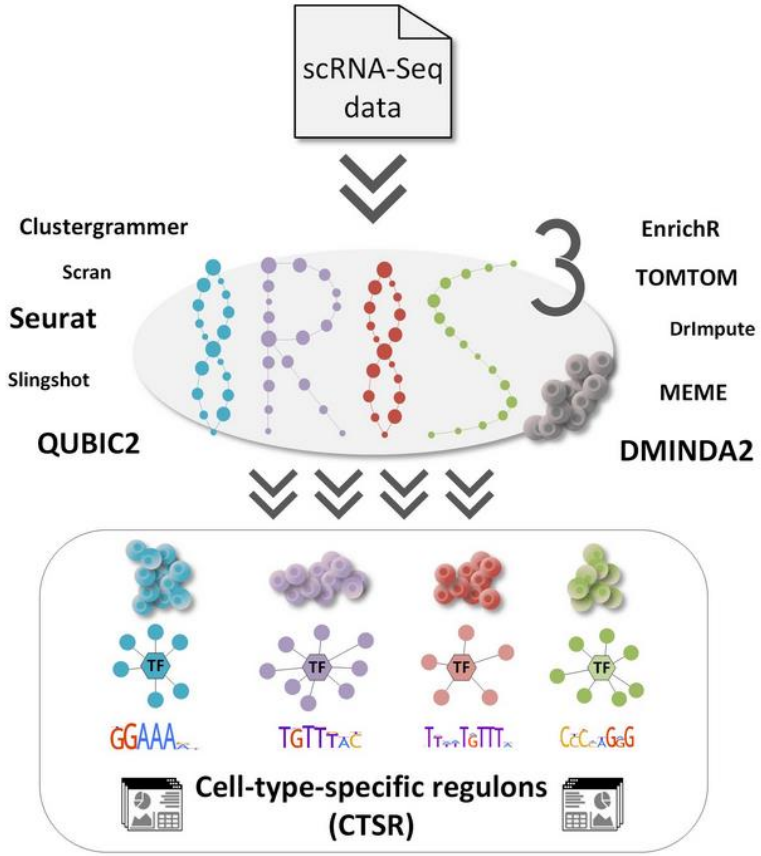
La Manno *et al*, 2018

Gene Regulatory Network Analysis

Combining cis-regulatory element analysis with single cell data to understand mechanisms driving cellular heterogeneity

- SCENIC: single-cell regulatory network inference and clustering
- IRIS3: https://bmbl.bmi.osumc.edu/iris3/

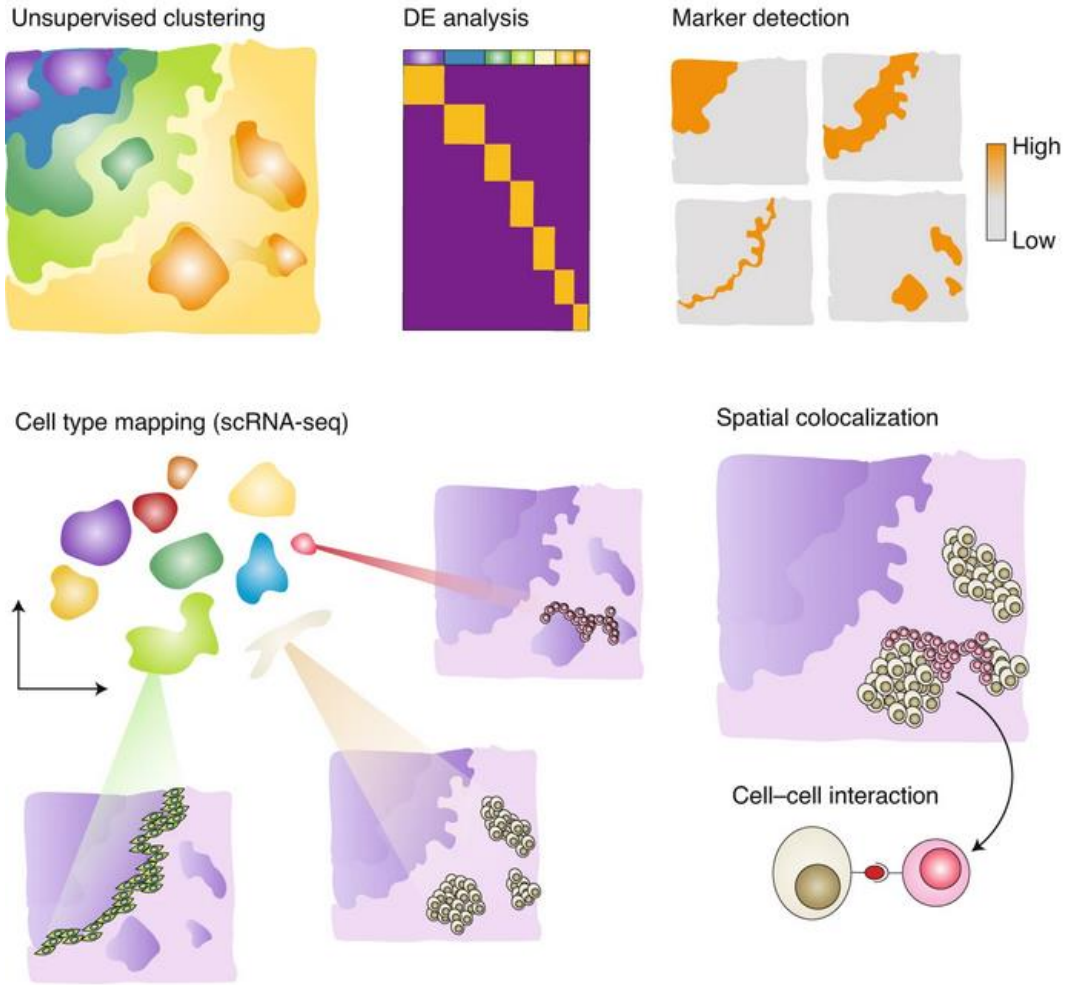




Spatial Omics Data Analysis

Challenges:

- Wide range of protocols and data processing pipelines
- Increased data volume, run time and memory usage demands more hardware and staff time
- Lack of standardized metrics or benchmarks



Larsson *et al*, 2021