DENOthe MEETING 2019

Analisi del profilo trascrizionale da "whole organ" a "single cell"

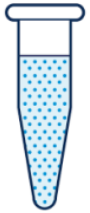
Letizia De Chiara and Roberto Semeraro

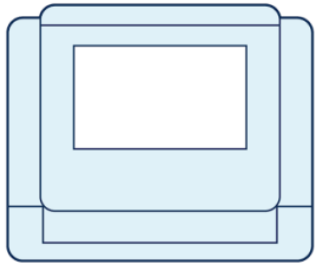
UR11. Prof. Romagnani Lab

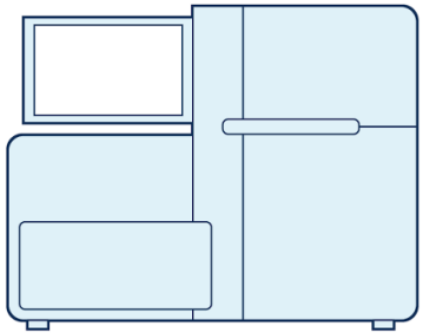
Single cell RNA-seq workflow

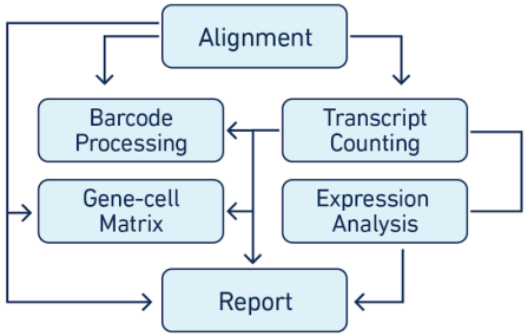
Letizia | Roberto

Input	Library Construction	Sequencing	Data Analysis	Data Visualization
Cell Suspension	Barcoding & Library Construction	Sequence Transcriptome	Pipelines	Report & Visualization



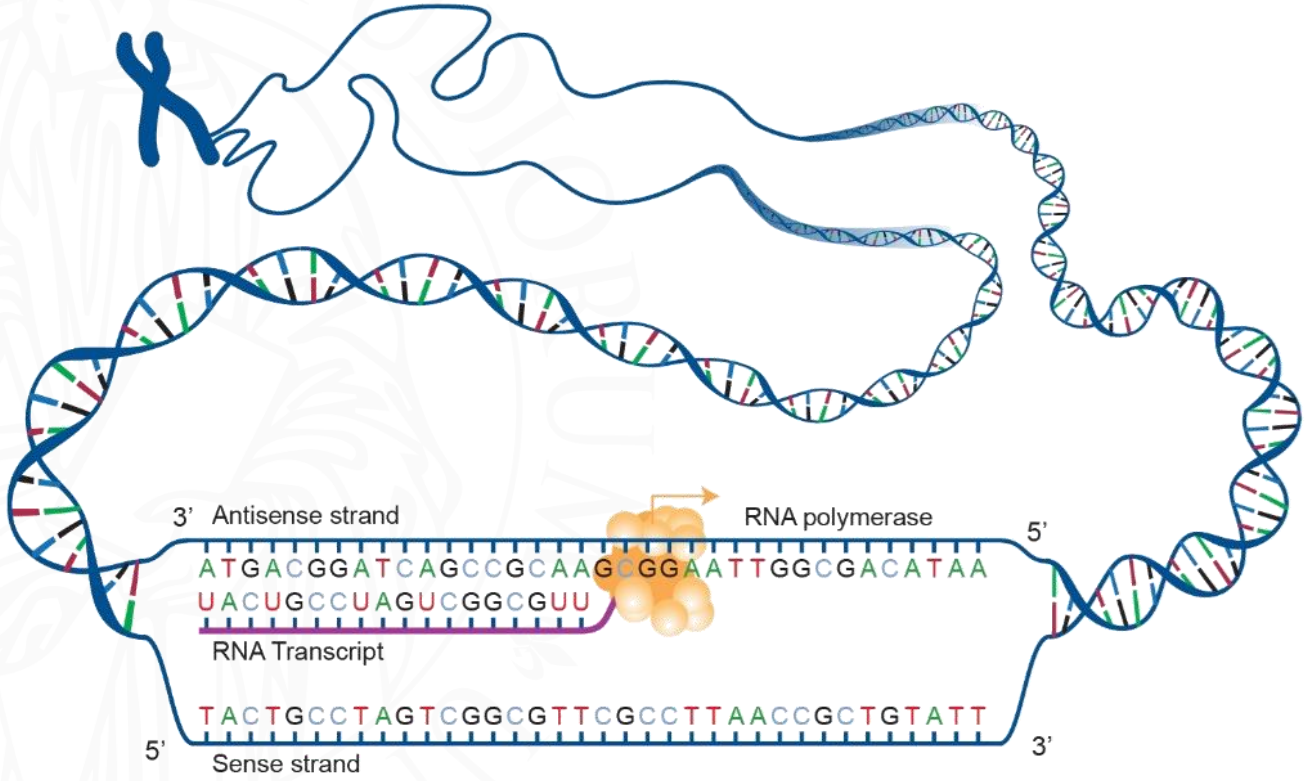




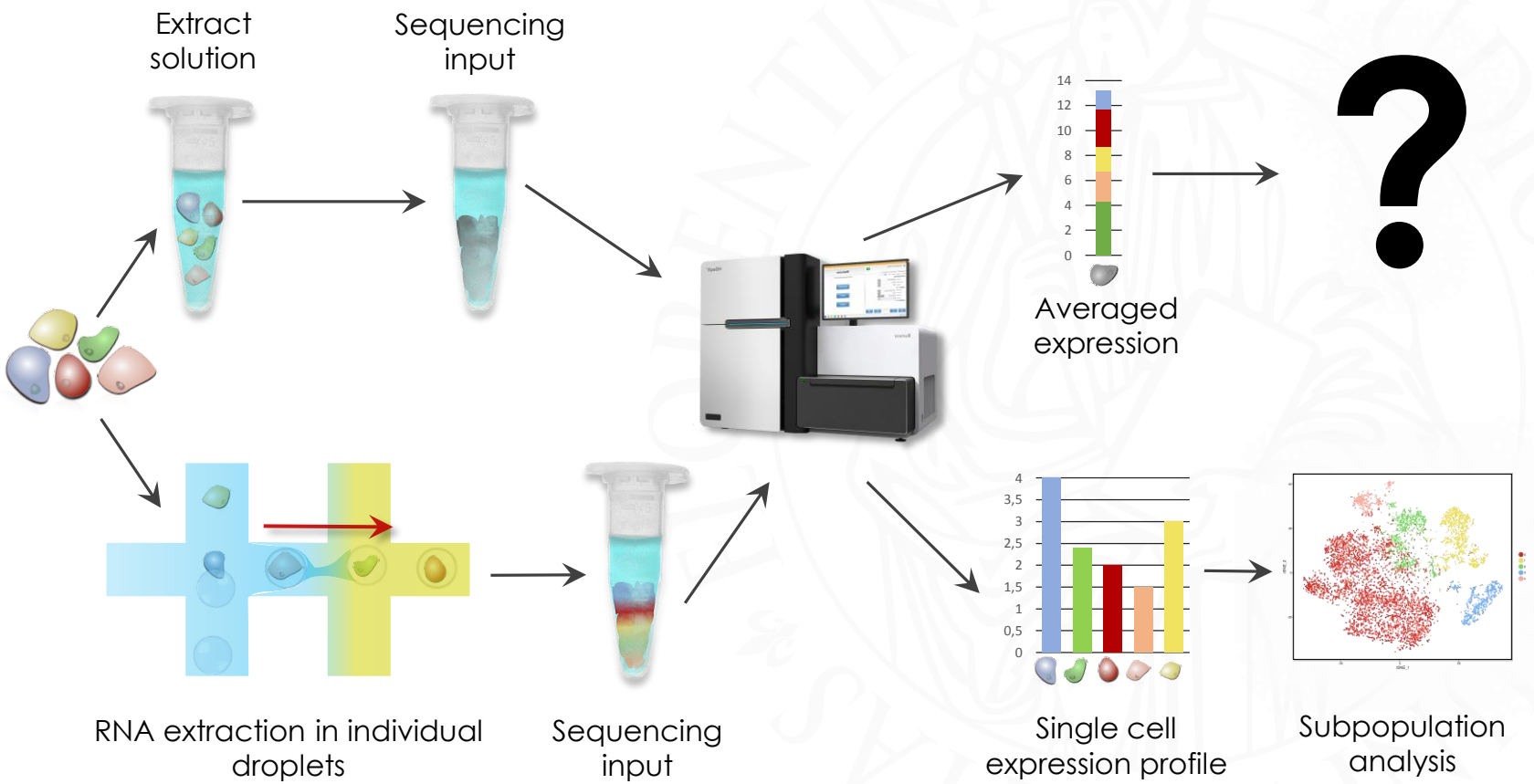


Why is it important to study cell transcriptome?

- Determine cell signature
- Connect genotype to phenotype
- Learn how individual cells respond to signals
- Study cell fate determination



Bulk RNA vs Single cell: What are the advantage?

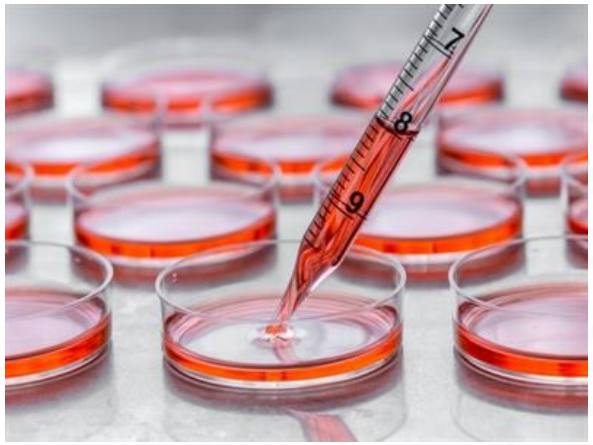


And the disvantages?

- Cell dissociation bias
- Only the more abundant cellular transcripts (≈1500) are detected
- Data are more complicated to be analyzed

Sample preparation: the main challenge

✓Cell culture

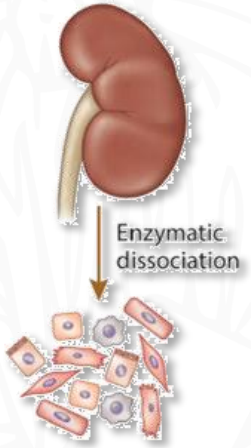


Viability:
at least 95% for cell culture
80% for tissue

↓

FACS sorting
Dead cell removal kit

✓Organ tissue

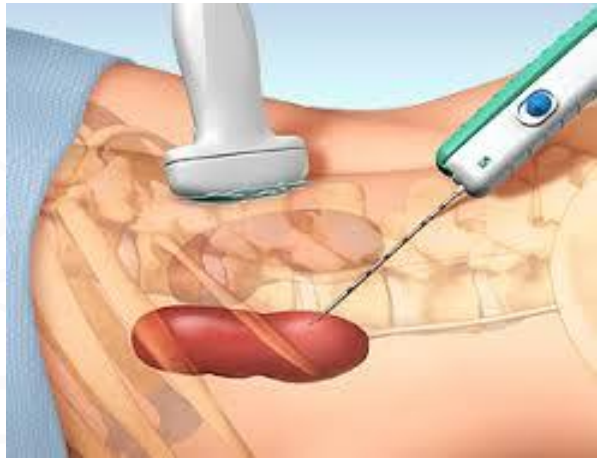


Cell dimension:
Up to 40um

↓

Nuclei isolation

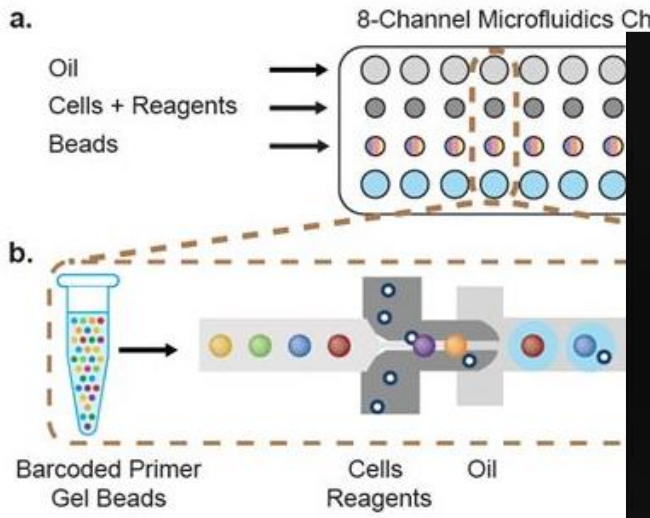
✗Biopsies

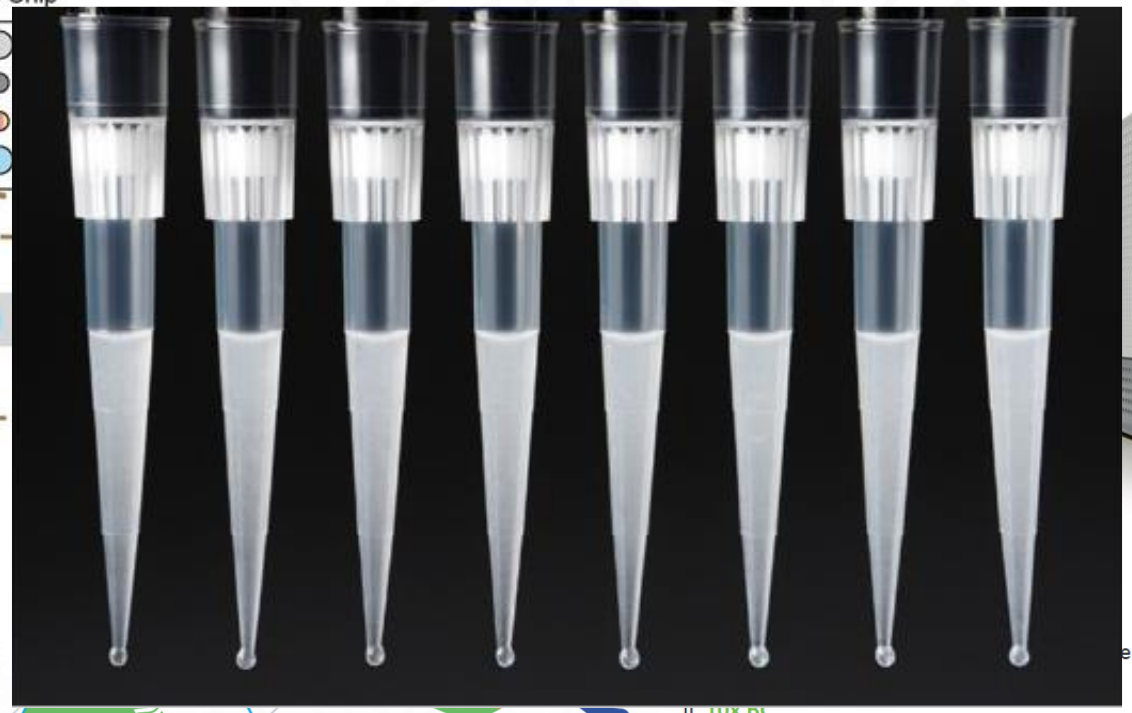


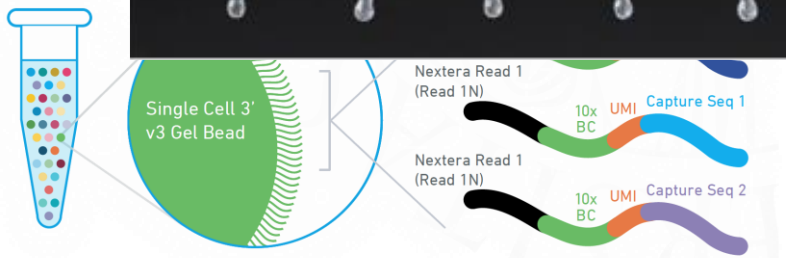
User controlled trade-off between cell numbers and doublet rate

Number of Cells	Expected Multiplet Rate (%)*
500	~0.4
1000	~0.8
5000	~3.9
10000	~7.6

Loading the sample on the 10XGenomics







ead 1 sequence

ii. 10x BC
16 nt 10x Barcode
~3.6 M defined barcode sequences;

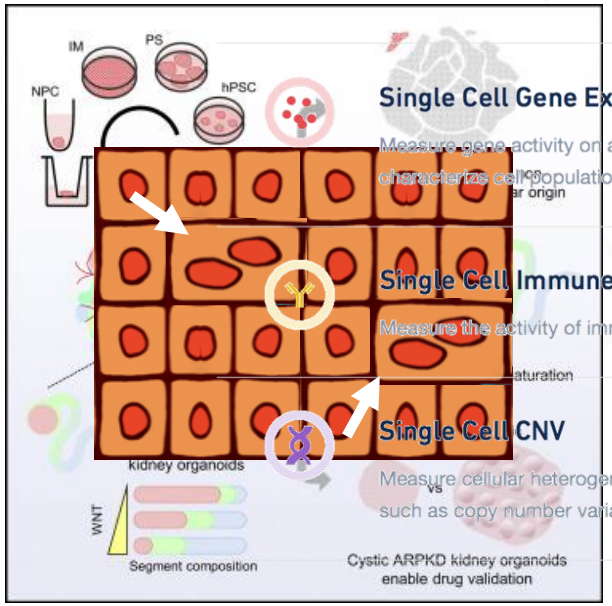
iii. UMI
12 nt Unique Molecular Identifier;

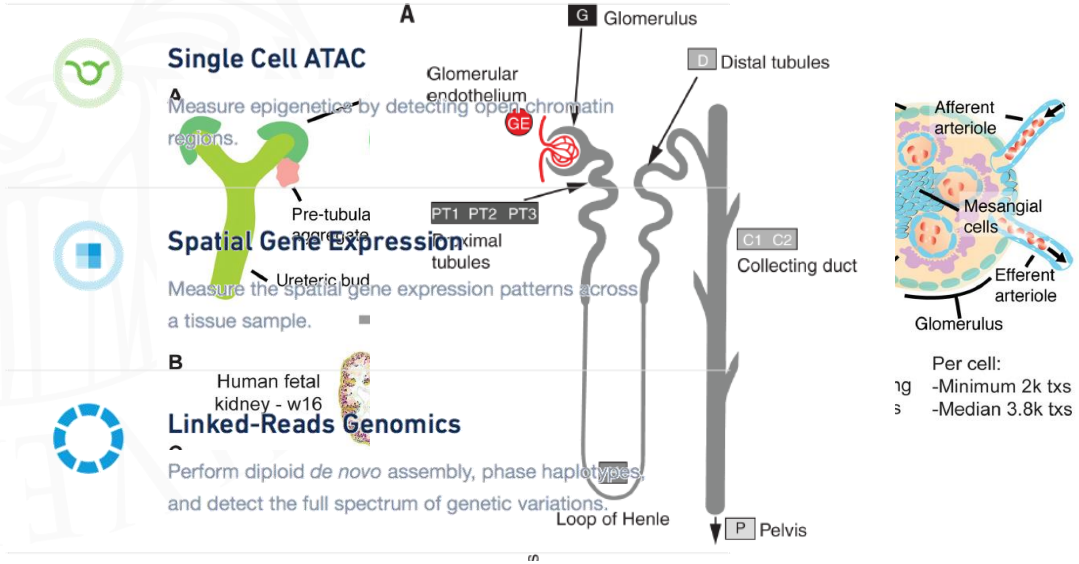
iv. Poly(dT)VN
30 nt Poly(dT) sequence

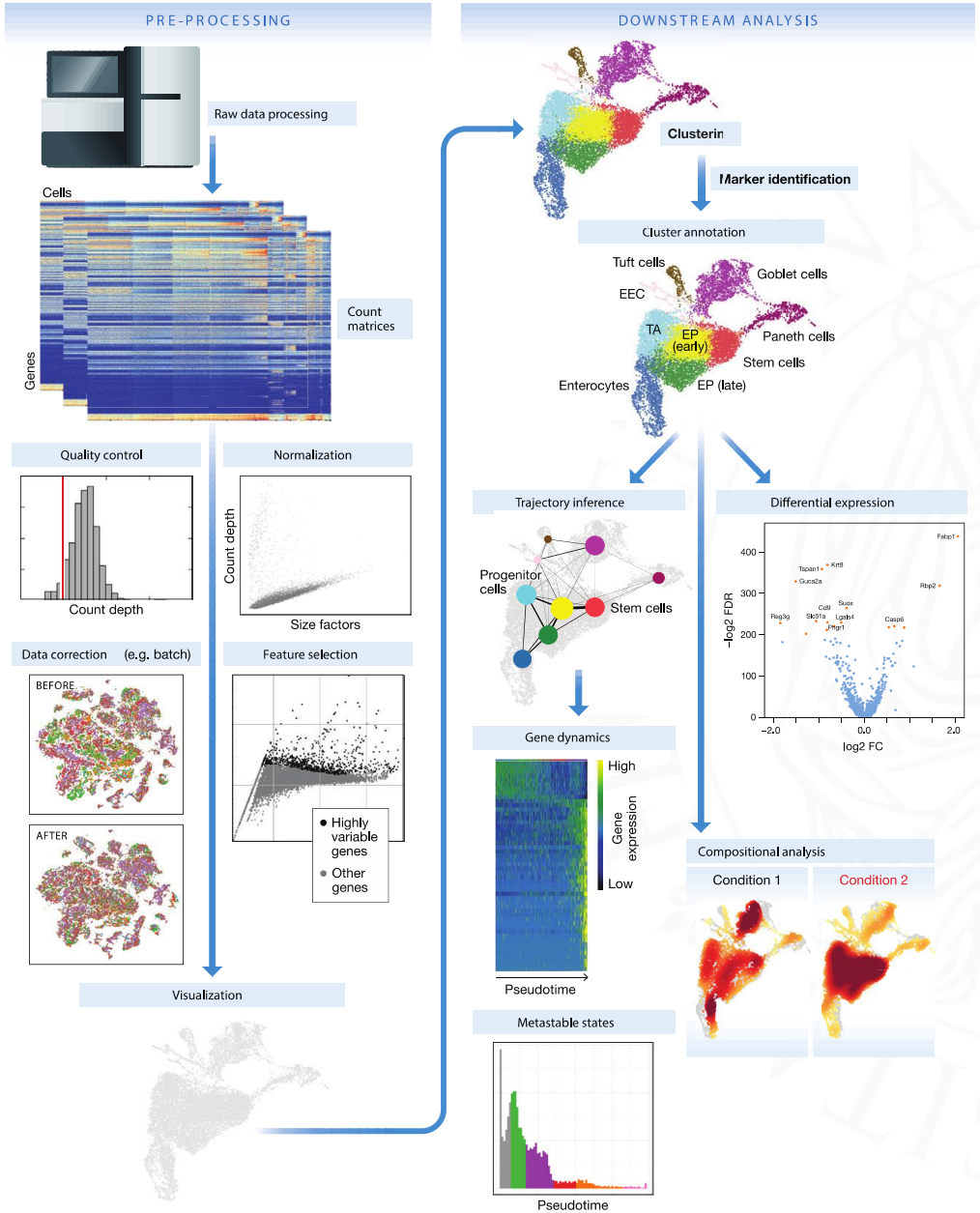
What are the applications?

- Identify stem cell population within an organ
- Analyze the expression pattern of developing organoids
- Characterize the specific signature of a cell subpopulation

- Study human organ development
- Establish cellular identity of renal tumors
- Analysis of genome







Preprocessing

- Read quality check
- Trimming
- Alignment (count matrix)
- Deduplication and quantification

Data cleaning and filtering

- Cell and Gene filtering (genes, counts, MT content)
- Normalization
- Batch correction (biological, technical)

Feature selection and visualization

- Feature selection
- log(x+1) transformation
- Dimensionality reduction (PCA, tSNE, UMAP)

Clustering

- Louvain, K-means
- Cluster annotation

Gene expression imputation

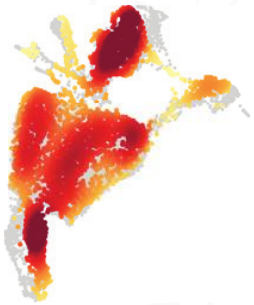
- Differential expression
- Trajectory analysis

CELL LEVEL

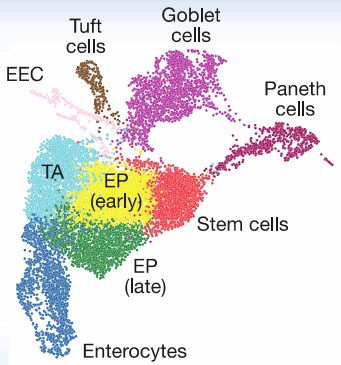
CLUSTER ANALYSIS

Clustering

Compositional analysis

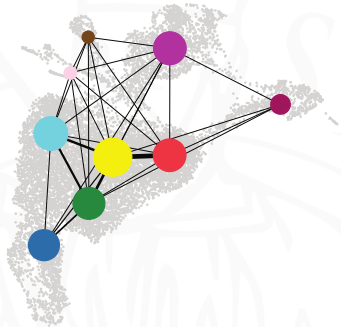


Cluster annotation

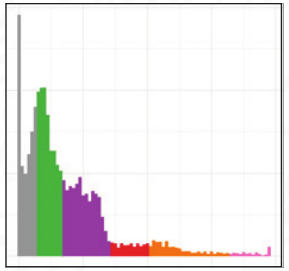


TRAJECTORY ANALYSIS

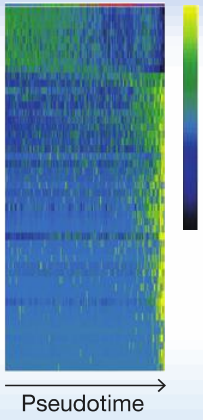
Trajectory inference



Metastable states

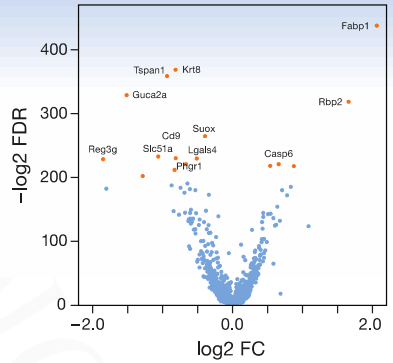


Gene expression dynamics

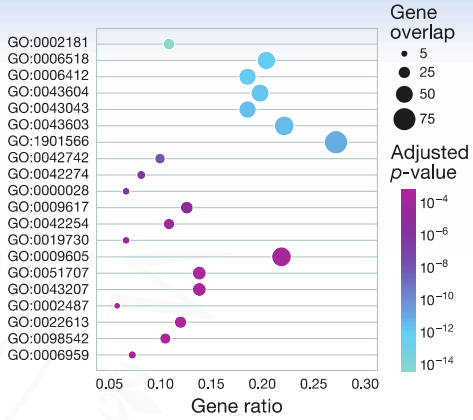


GENE LEVEL

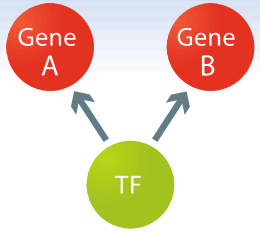
Differential expression analysis



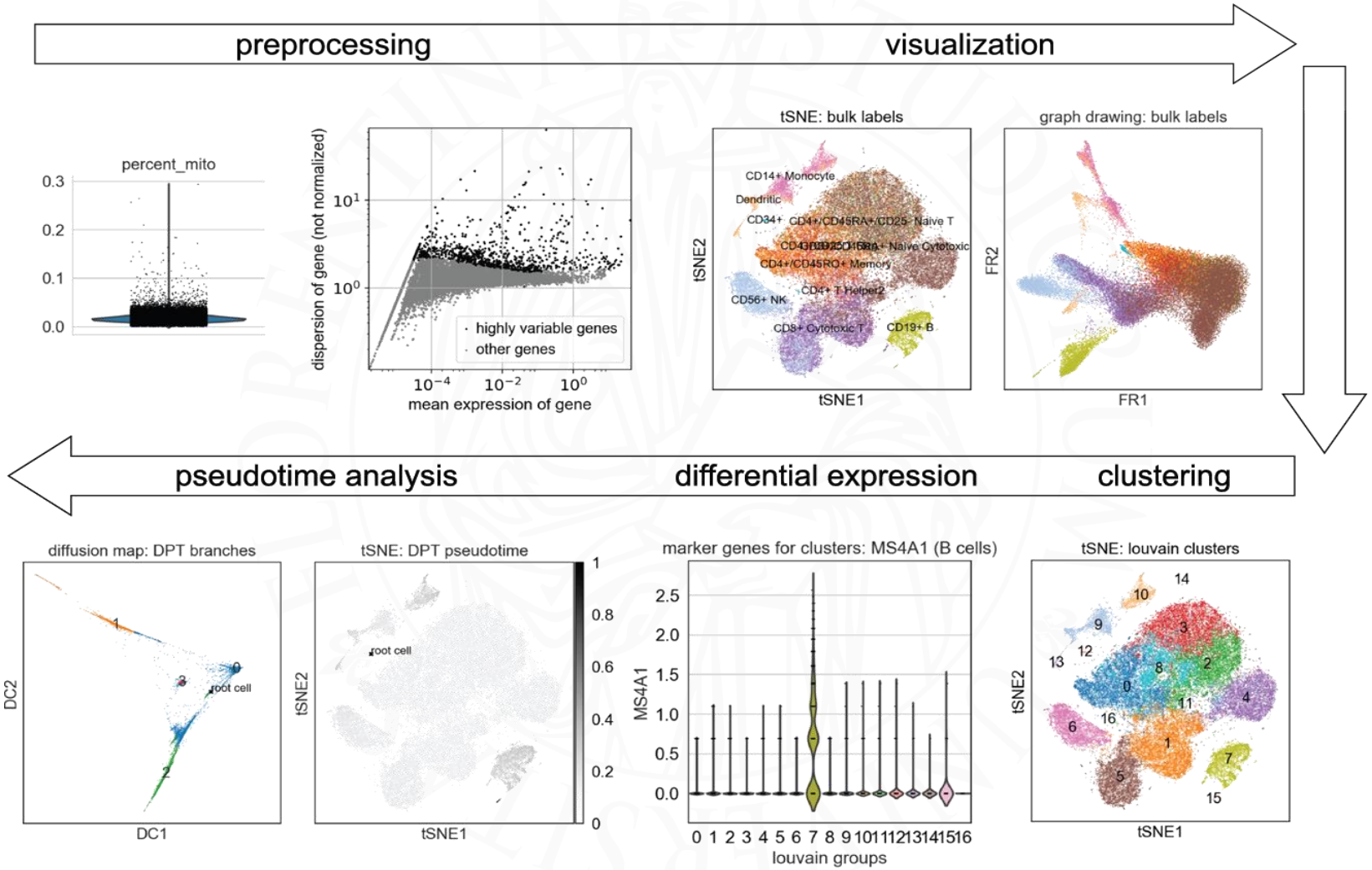
Gene set analysis



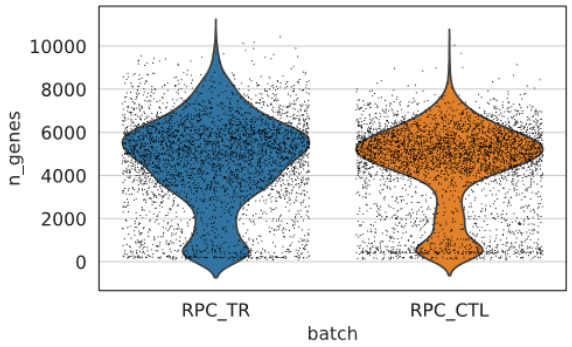
Gene regulatory networks

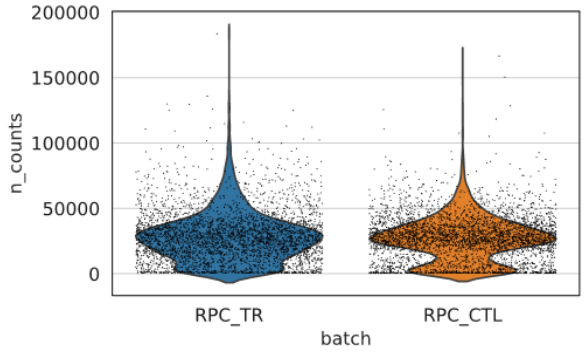


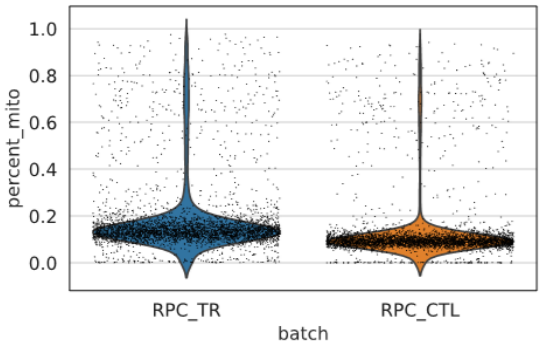
Scanpy


preprocessing
visualization
pseudotime analysis
differential expression
clustering
percent_mito
dispersion of gene (not normalized)
highly variable genes
other genes
mean expression of gene
tSNE: bulk labels
CD14+ Monocyte
Dendritic
CD34+
CD4+/CD45RA+/CD25- Naive T
CD4+/CD45RO+ Memory
CD56+ NK
CD4+ T Helper2
CD8+ Cytotoxic T
CD19+ B
tSNE1
tSNE2
graph drawing: bulk labels
FR1
FR2
diffusion map: DPT branches
root cell
DC1
DC2
tSNE: DPT pseudotime
marker genes for clusters: MS4A1 (B cells)
MS4A1
louvain groups
tSNE: louvain clusters

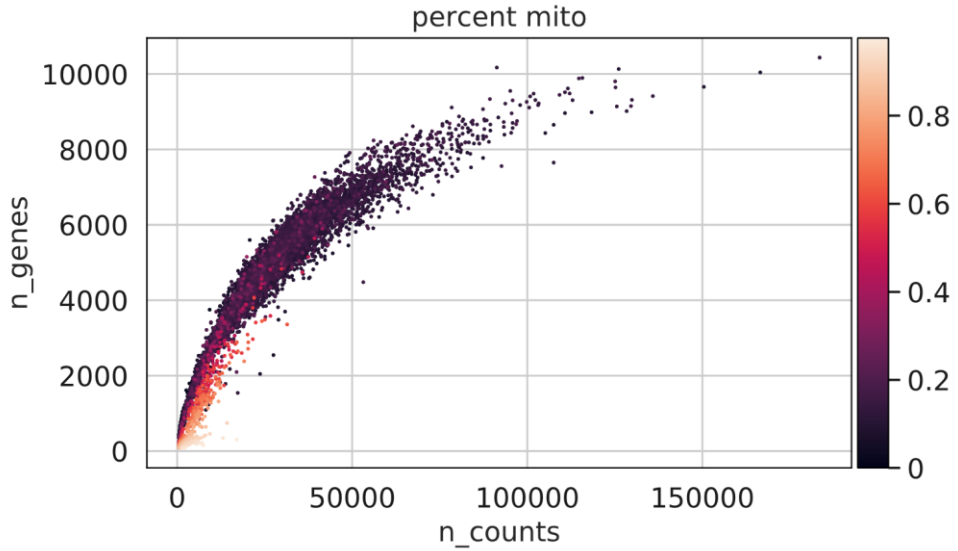
Quality control







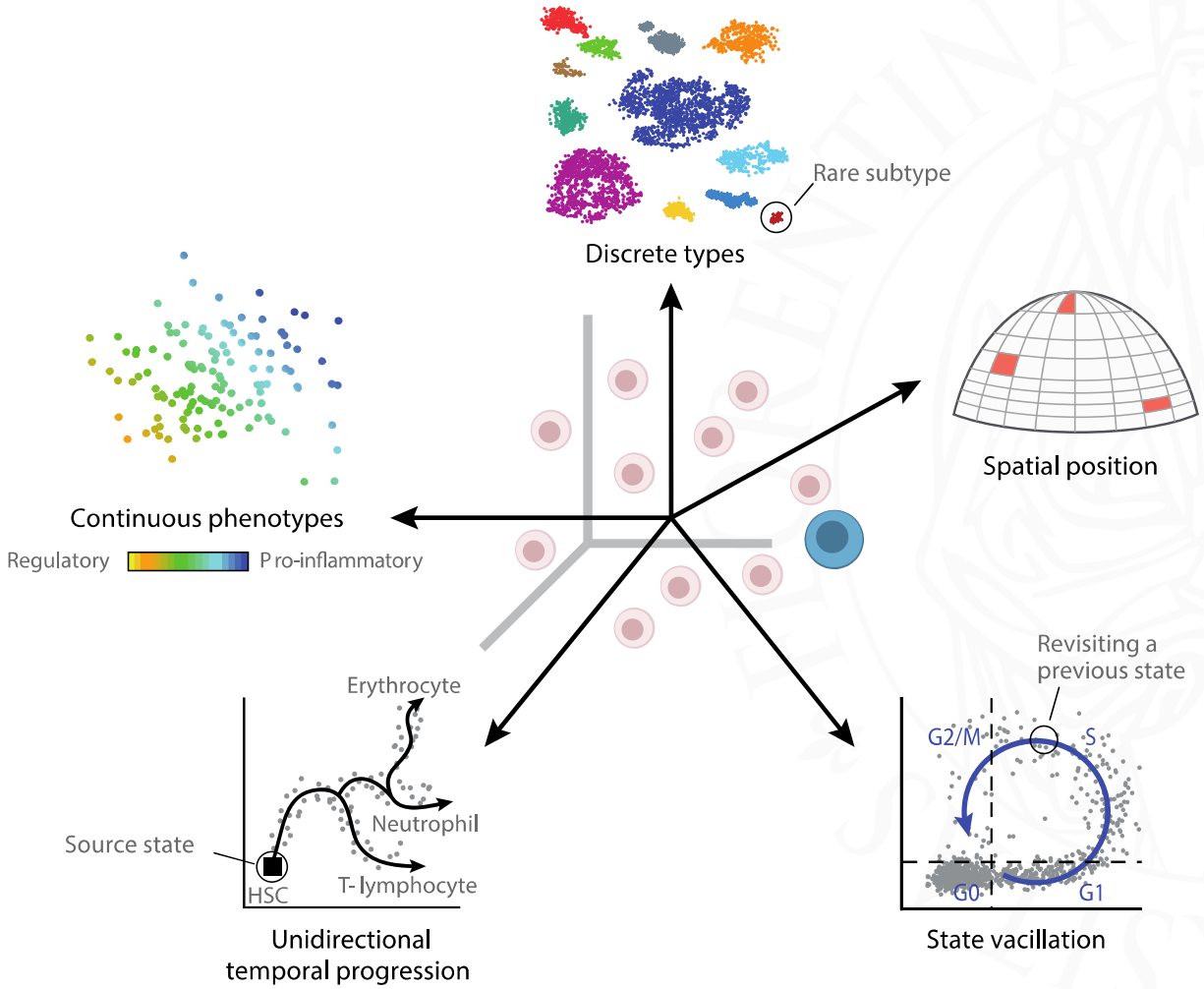
- Number of detected Genes
- Number of counts
- Mitochondrial DNA content



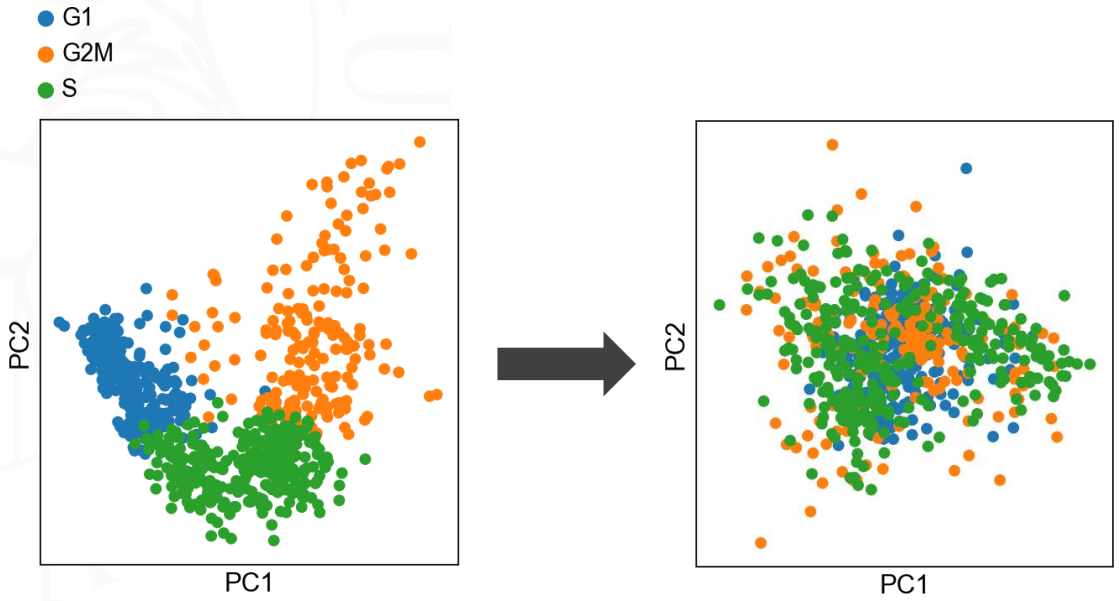
Useful to filter out broken cells and badly detected genes

```
sc.pp.filter_cells(data, min_genes=200)
sc.pp.filter_genes(data, min_cells=3)
sc.pp.filter_cells(data, min_counts=7000)
```

Normalization and Batch correction

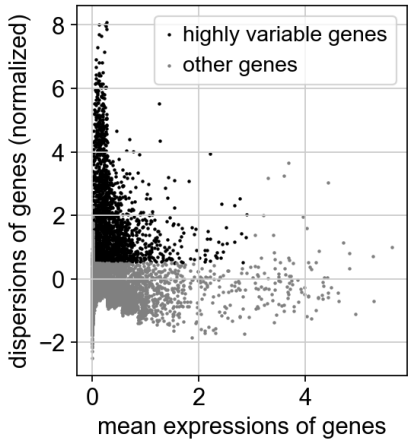


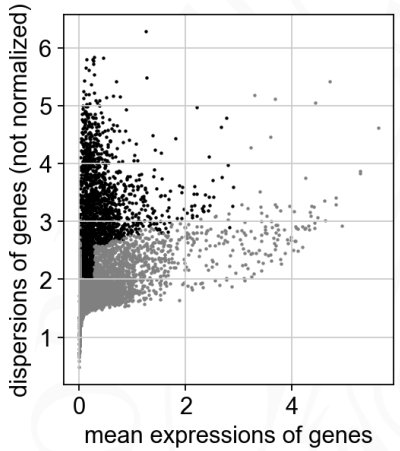
- Normalization attempts to remove the effects of count sampling
- Regressing out biological effect to uncover the underlying biological signal
- Batch correction try to mitigate biases related to different protocols, laboratories, times, etc.



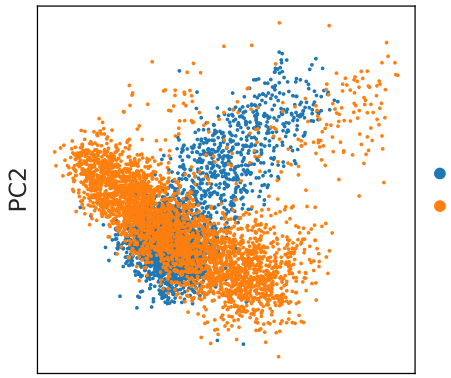
Feature selection and visualization

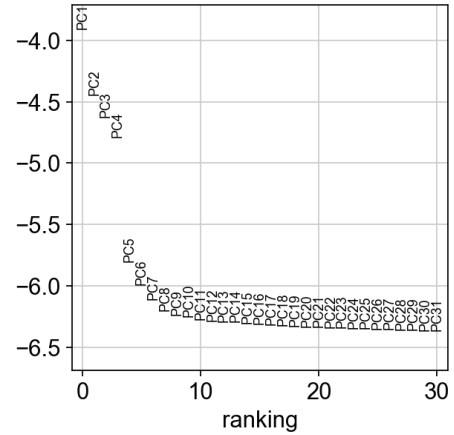
1. The dataset is filtered to keep only genes that are "informative"



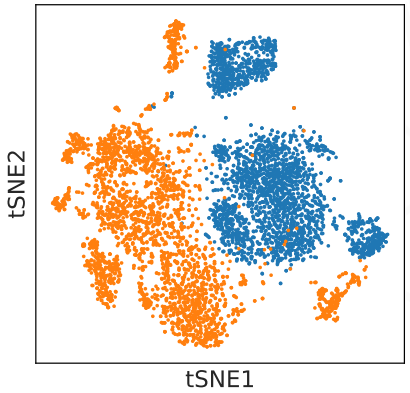


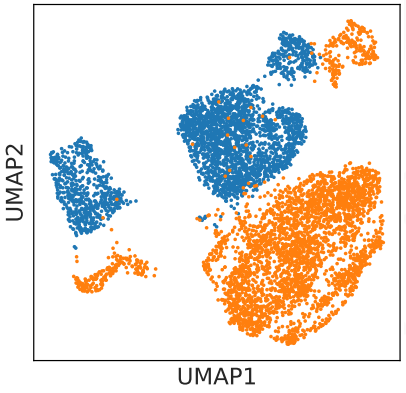
2. The dimensions of single-cell expression matrices can be further reduced (PCA)

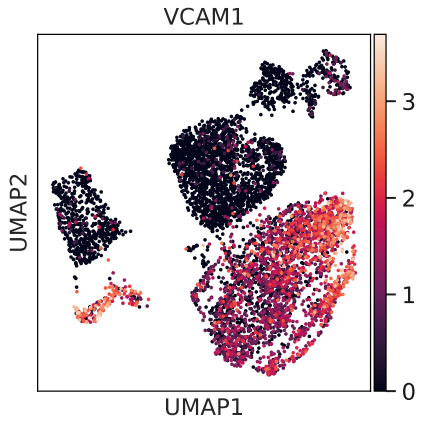




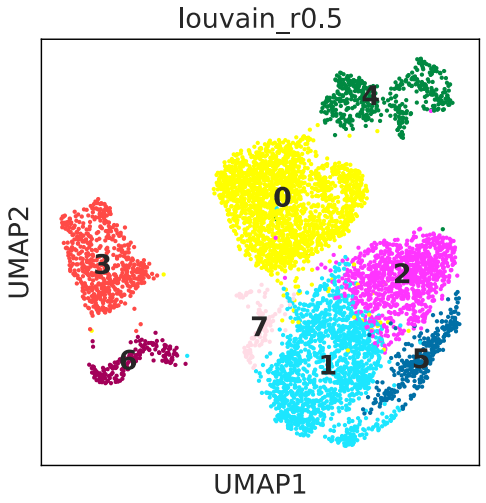
Different visualizations reveal different patterns

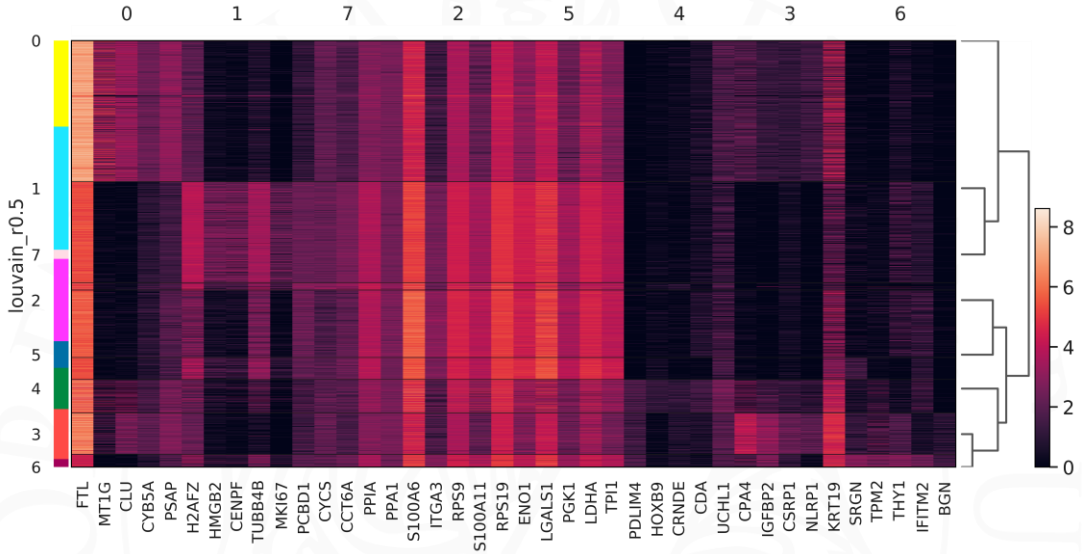


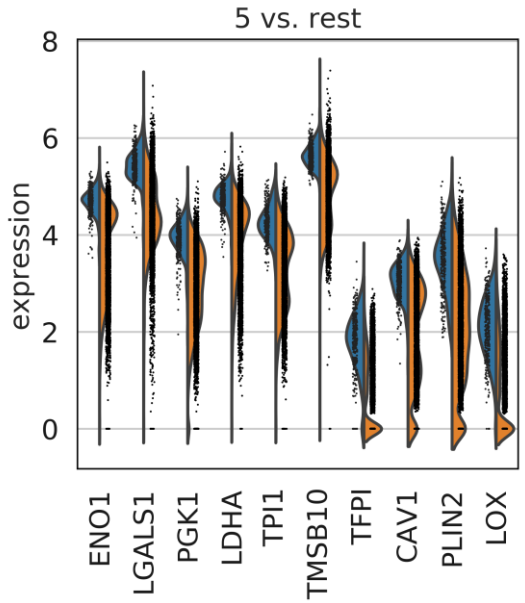


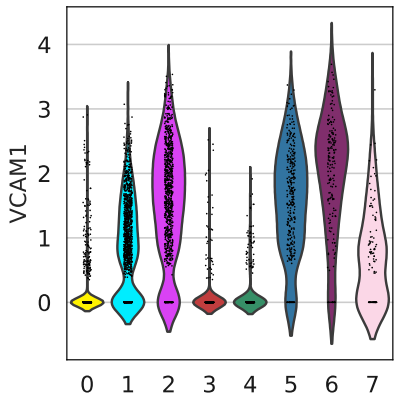


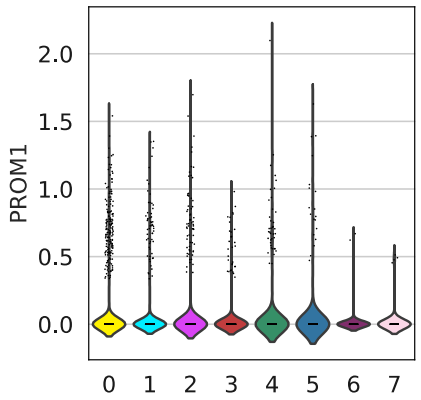
Clustering and differential analysis

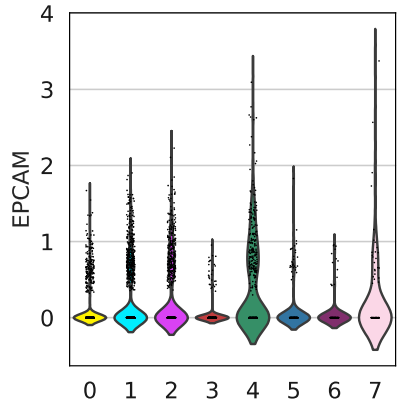






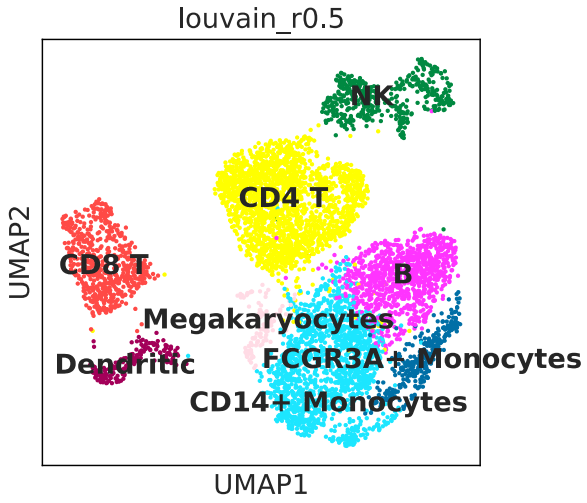


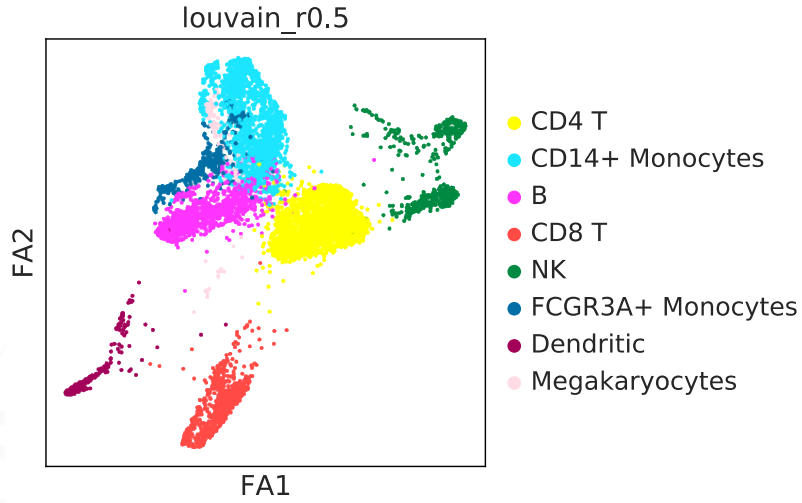


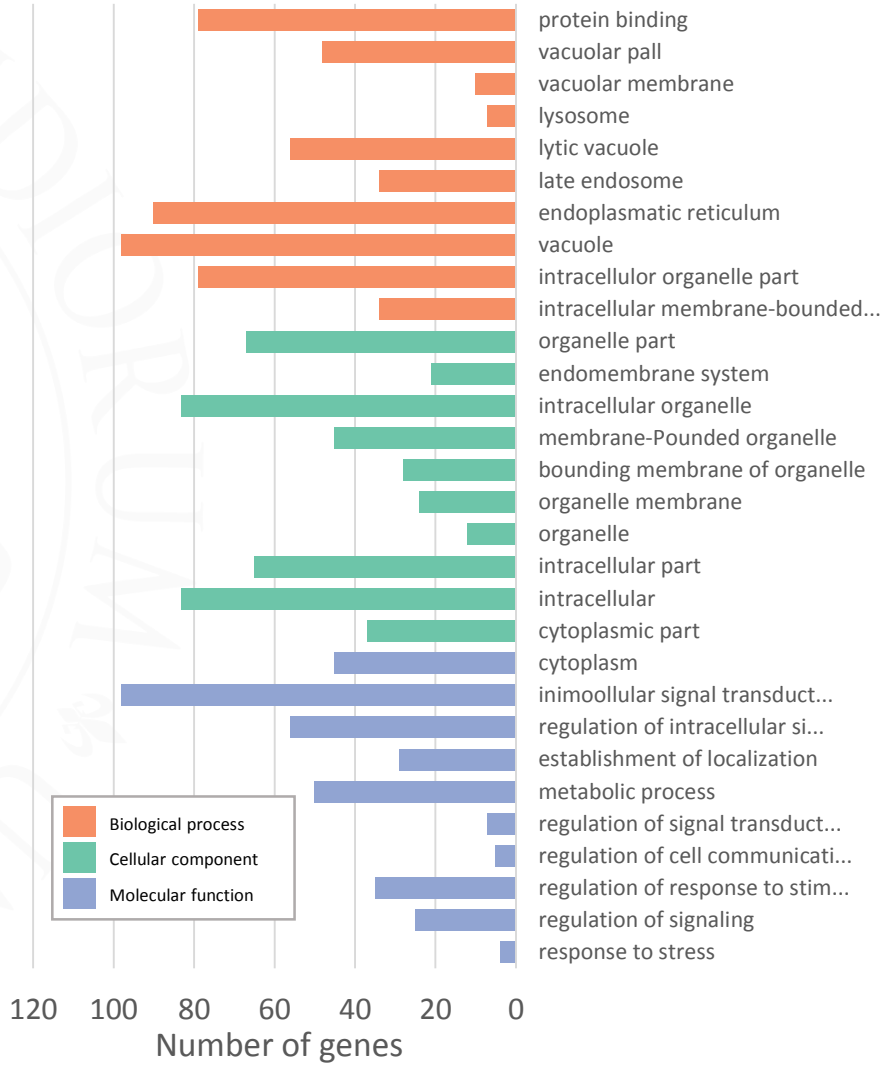


Cluster	Markers	Cell Type
0	IL7R	CD4 T cells
1	CD14, LYZ	CD14+ Monocytes
2	MS4A1	B cells
3	CD8A	CD8 T cells
4	GNLY, NKG7	NK cells
5	FCGR3A, MS4A7	FCGR3A+ Monocytes
6	FCER1A, CST3	Dendritic Cells
7	PPBP	Megakaryocytes

Cell and gene level analysis

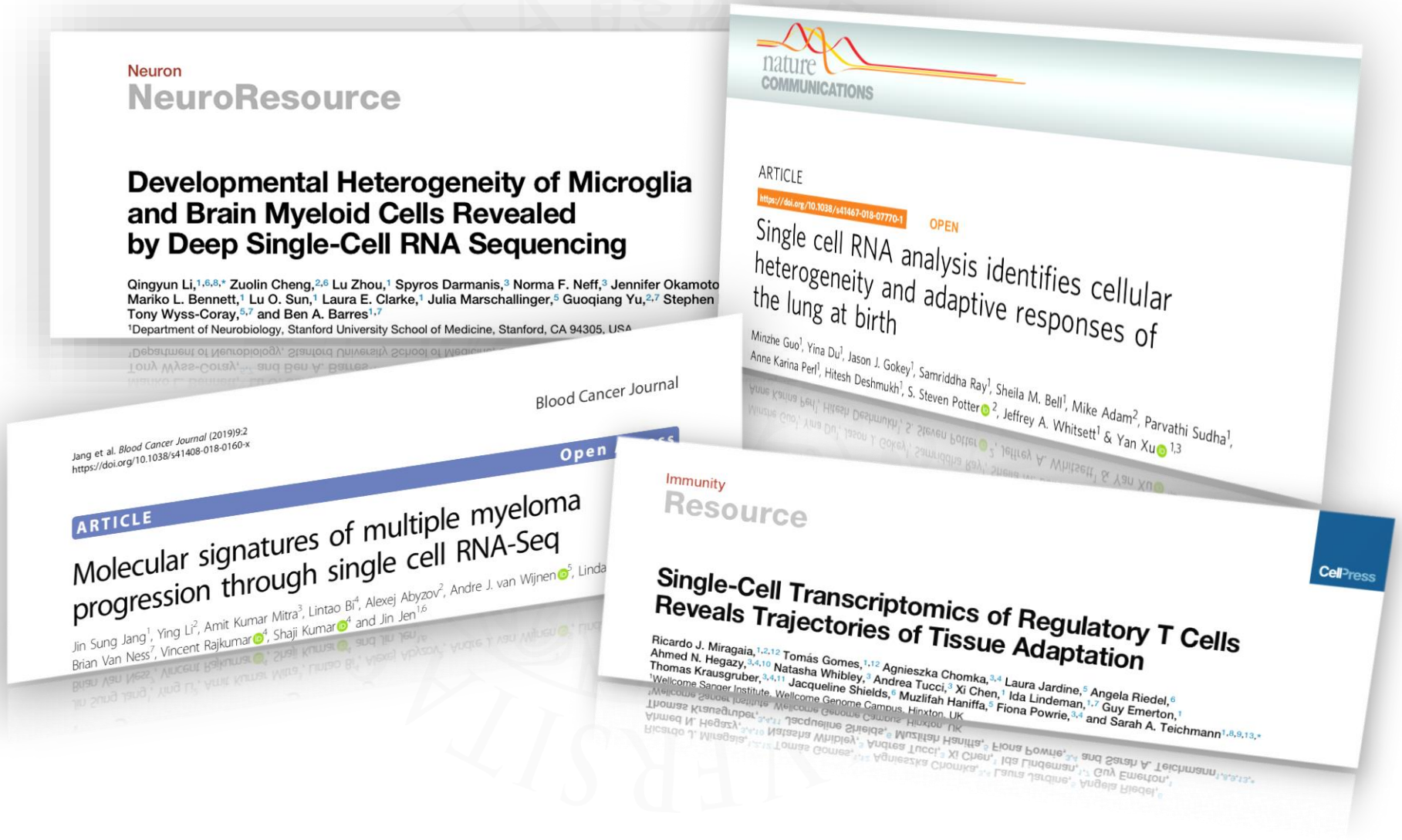






- Cell-level analysis typically focuses on the description of two structures: clusters and trajectories
- In trajectory analysis the data are regarded as a snapshot of a dynamic process
- Differential expression testing, gene set analyses and gene regulatory network inference directly investigate molecular signals in the data

Applications




Neuron
NeuroResource
Developmental Heterogeneity of Microglia and Brain Myeloid Cells Revealed by Deep Single-Cell RNA Sequencing
Jang et al. Blood Cancer Journal (2019)
https://doi.org/10.1038/s41408-018-0160-x
Open Access
ARTICLE
Molecular signatures of multiple myeloma progression through single cell RNA-Seq
nature COMMUNICATIONS
CellPress
ARTICLE
OPEN
Single cell RNA analysis identifies cellular heterogeneity and adapti
Immunity
Resource
Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation
CellPress

Thank you!