

# Introduction to scRNA-seq

03.31.2020



# Outline

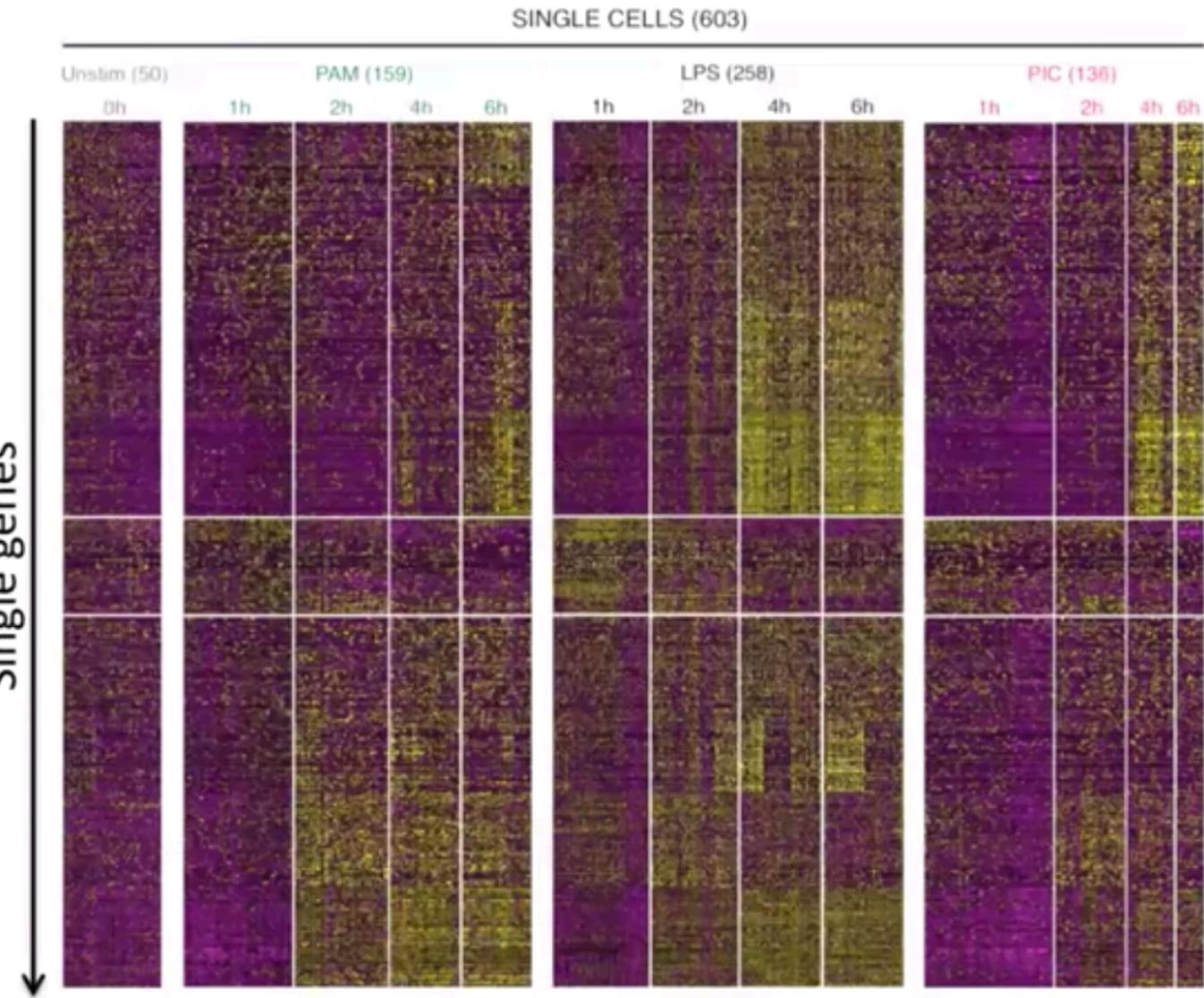
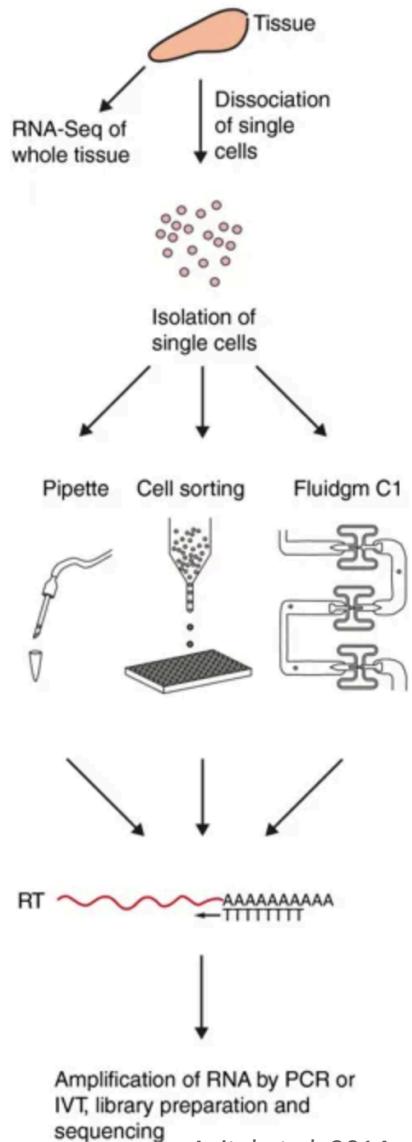
## **Part I: Single cell technologies**

- Introduction & Historical Context
- Why scRNA seq?
- Current Technologies/Methods

## **Part II: Biological insights**

- Analysis
  - Preprocessing
  - Clustering & Differential Expression
  - Trajectory Inference

# Scaling Up: scRNA-seq

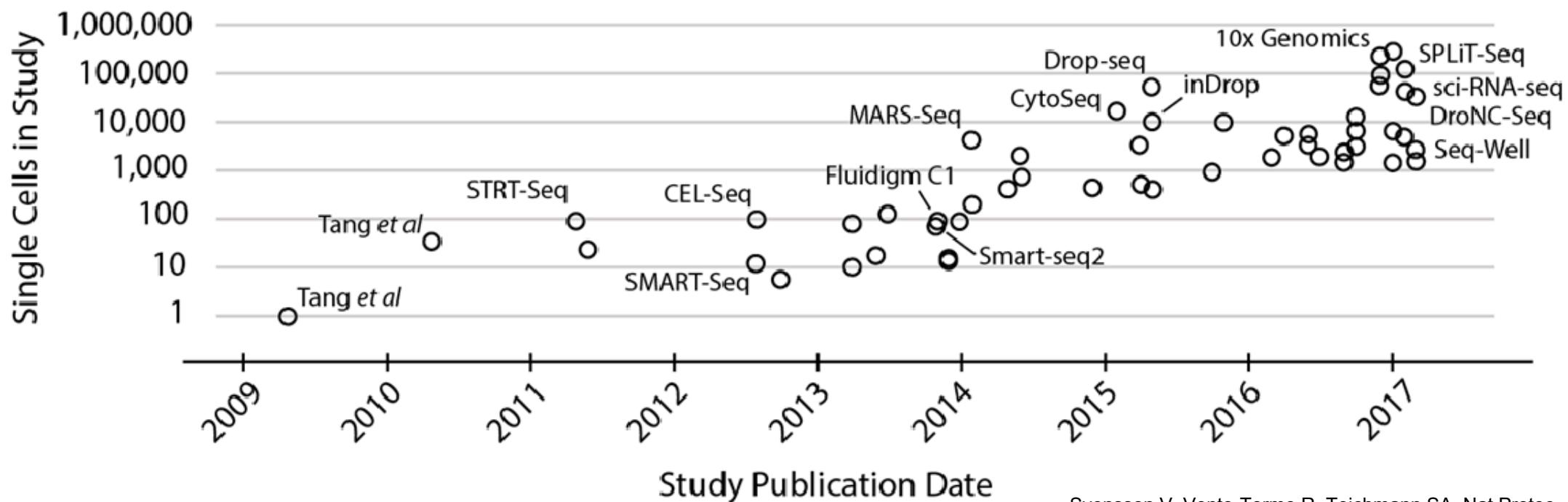


# Exponential scaling of single-cell RNA-seq in the past decade

A

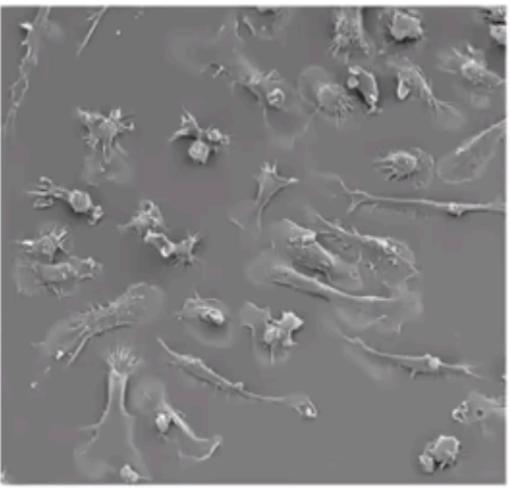


B

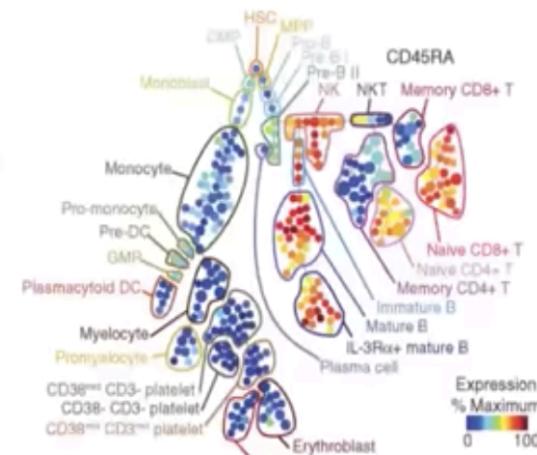


# Why Single Cells

## Cellular heterogeneity

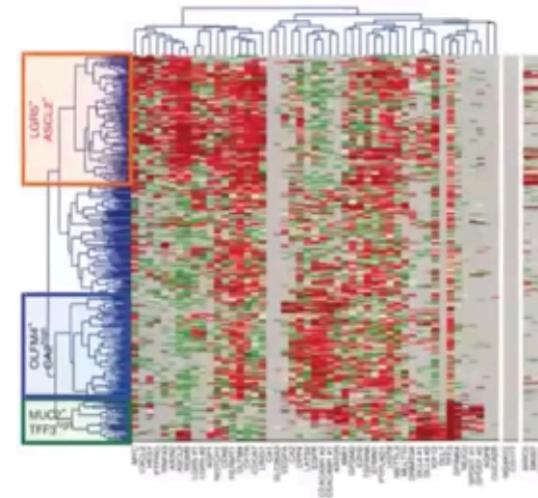


## Differentiation trajectories



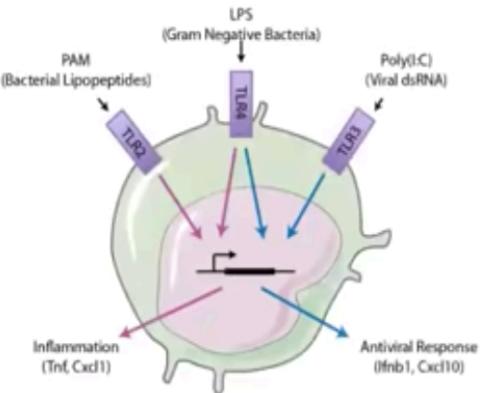
Bendall et al. (2011), Science

## Within-cell-type differences

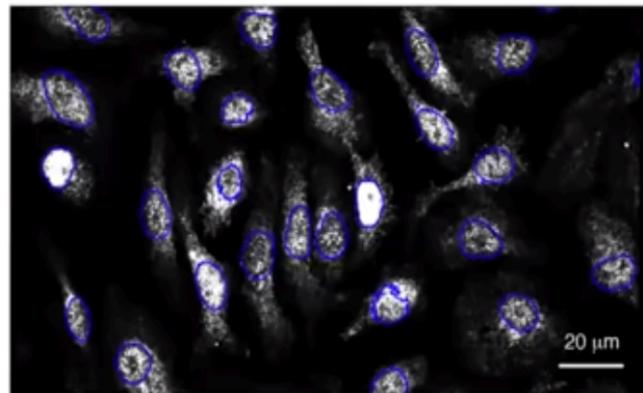


Dalerba et al. (2011), Nature Biotech

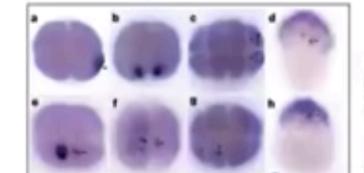
## TLR Signaling



## IRF3 Protein Levels - 4h LPS

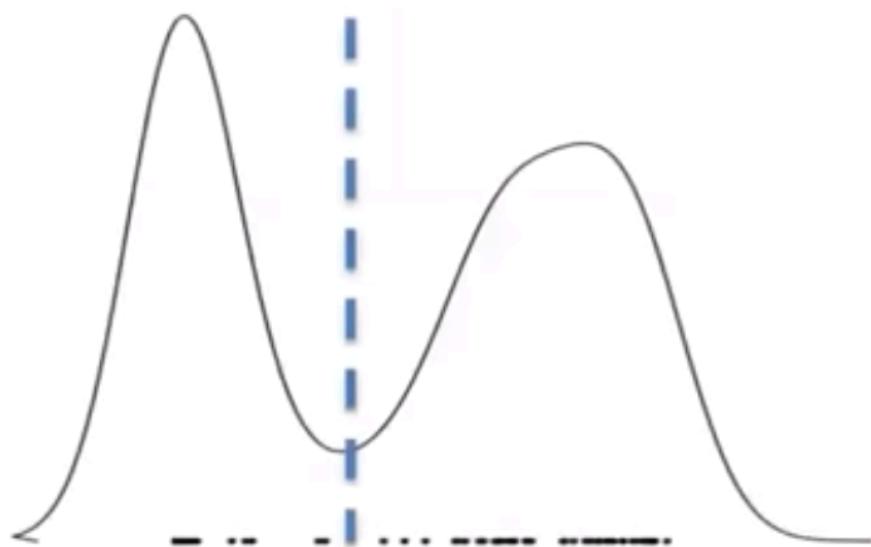


Circulating Tumor Cells

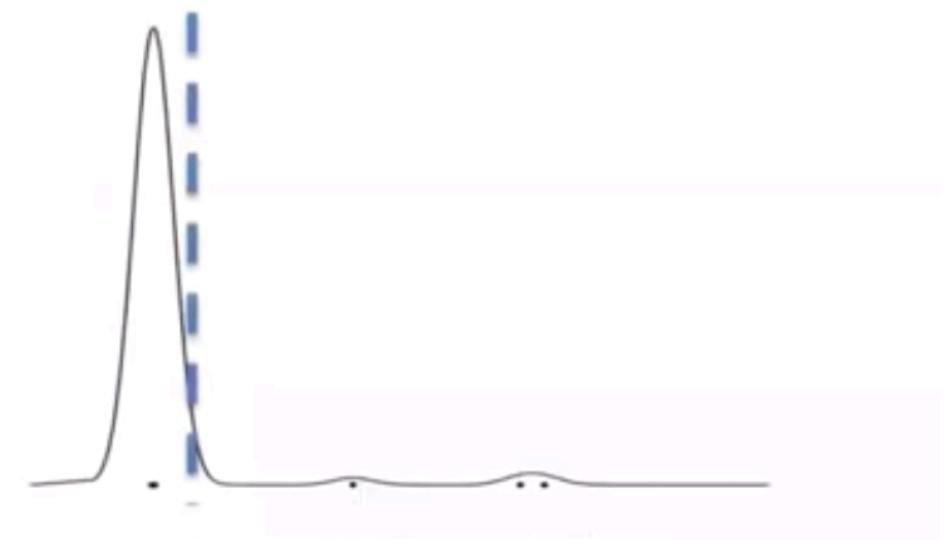


Zebrafish early embryo

# Bulk analysis does not capture the full story

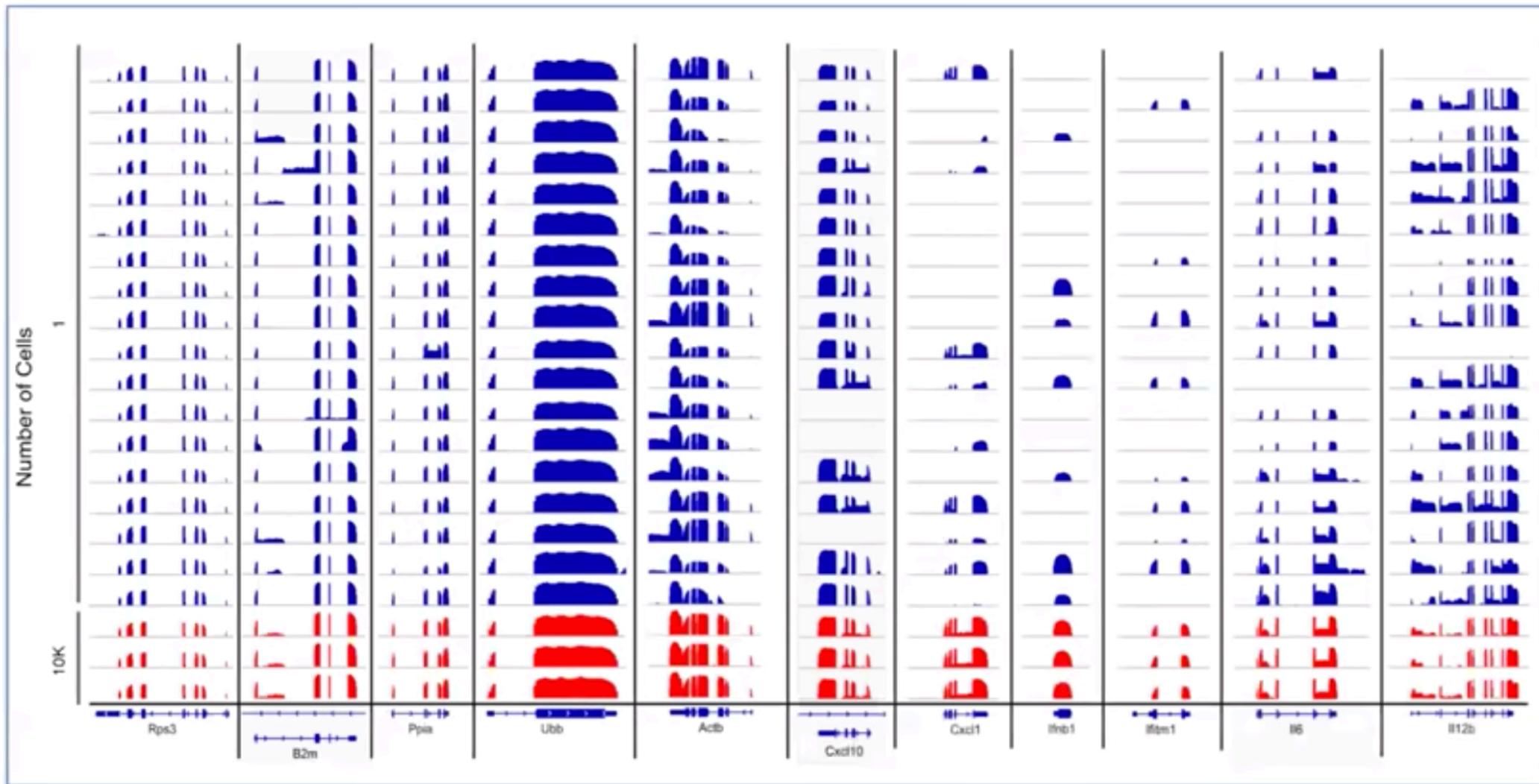


The average may not  
represent the population

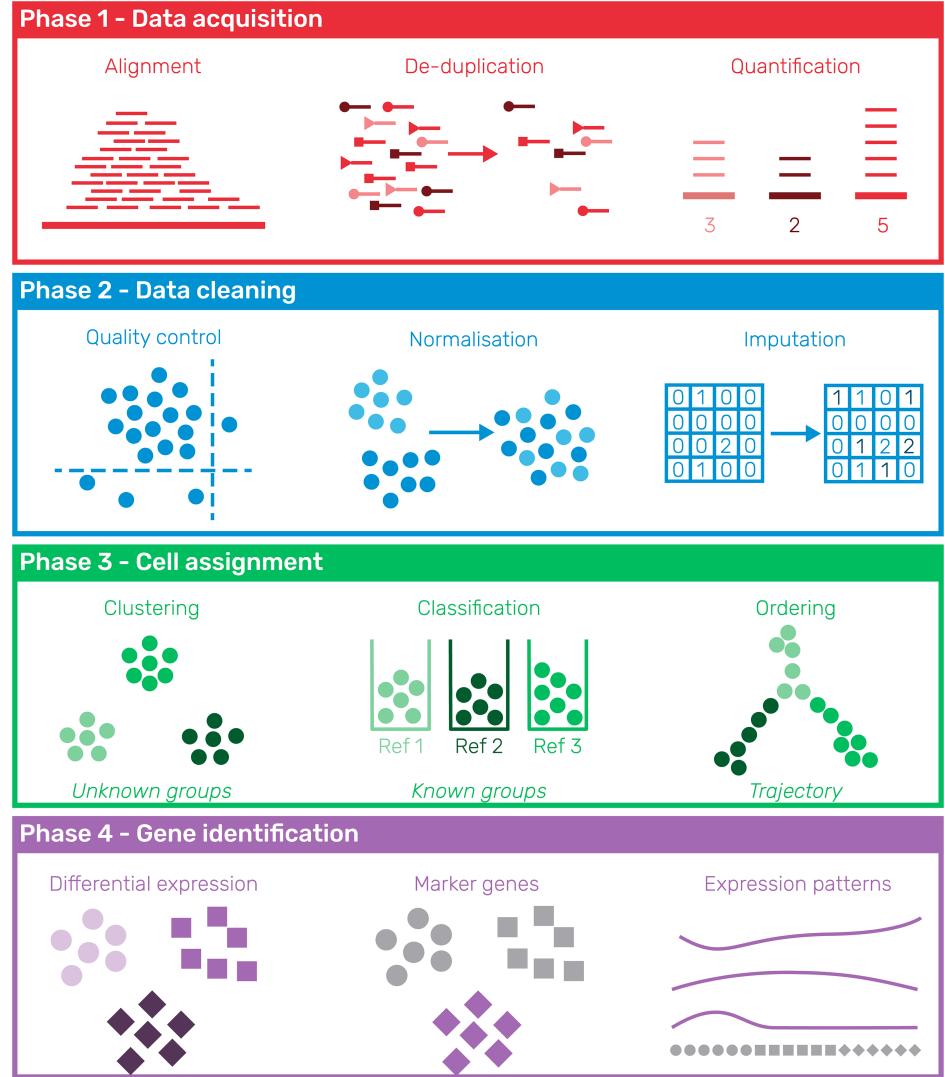


Rare signals might be lost

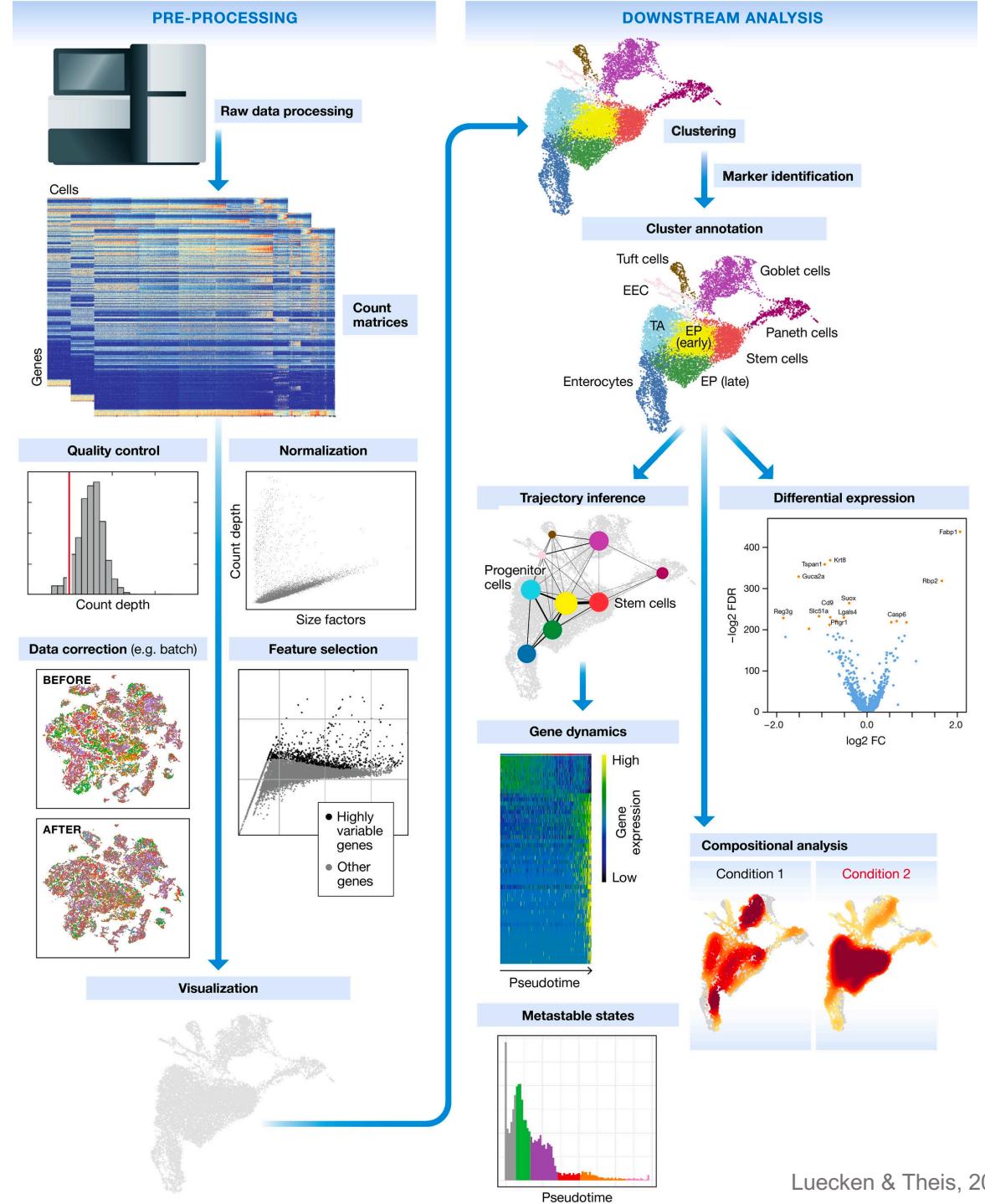
# scRNA-seq data looks like RNA-seq



# Overview

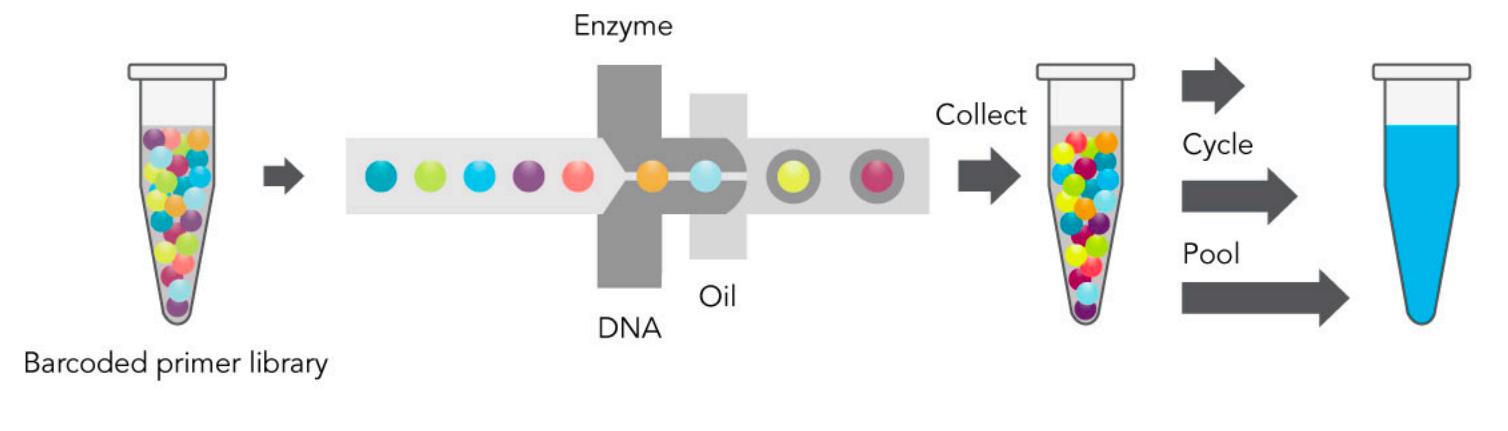
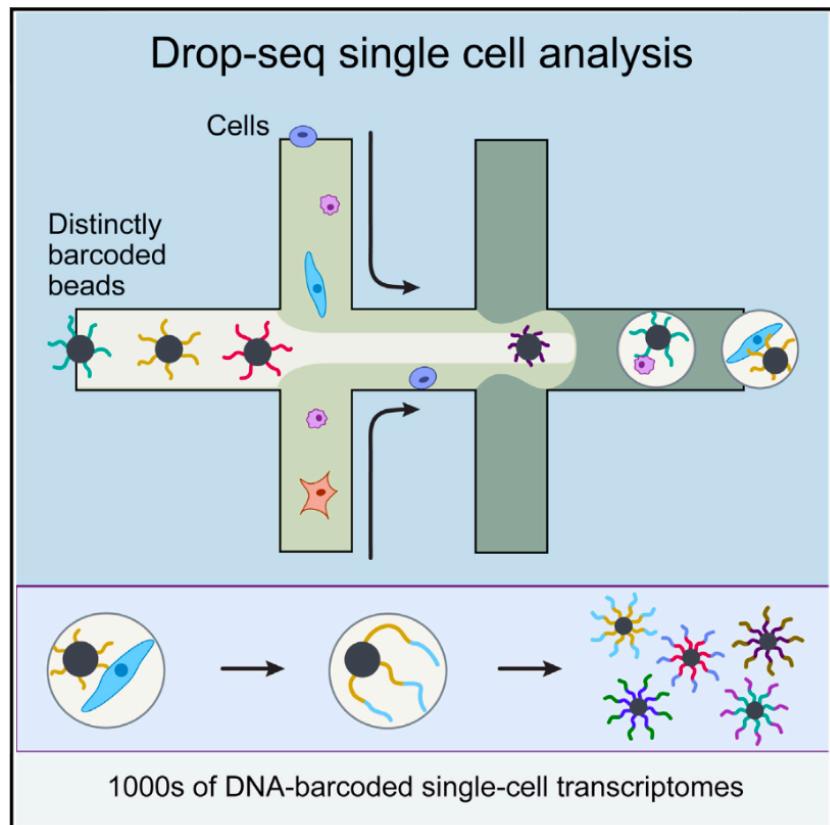


Zappia et al, 2018

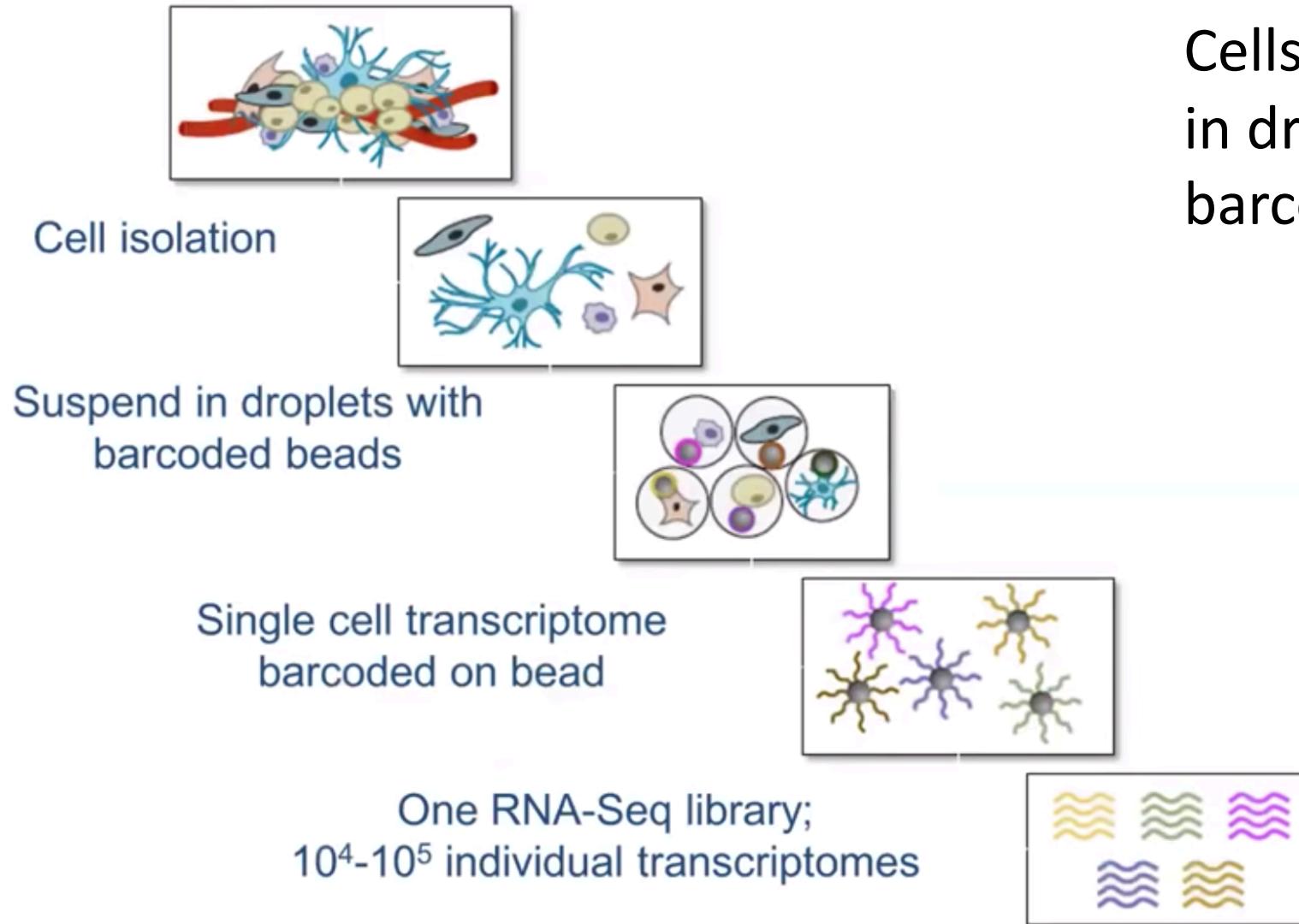


Luecken & Theis, 2019

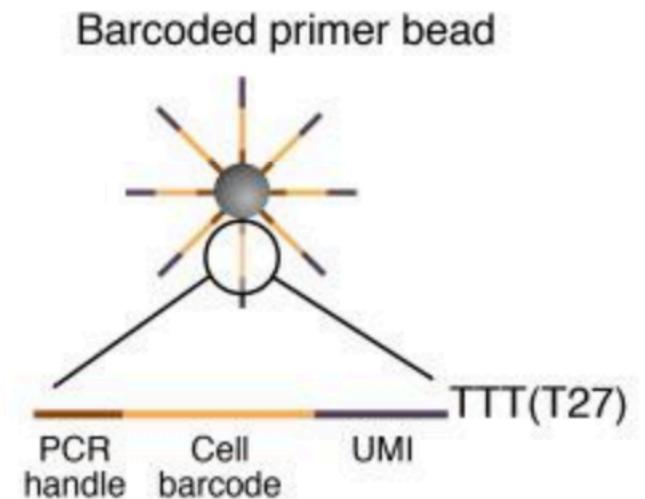
# Current Technologies



# Drop-Seq

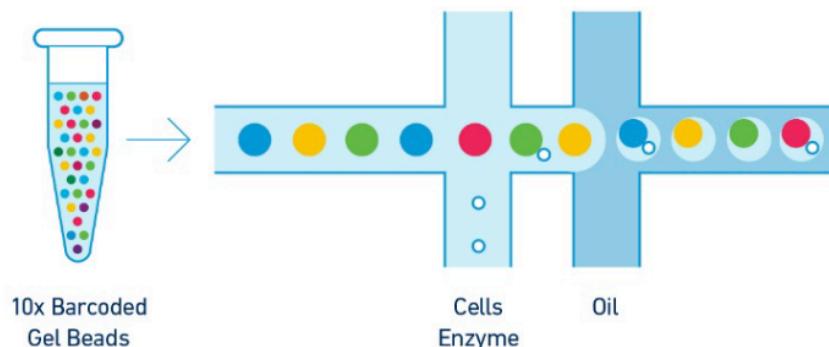
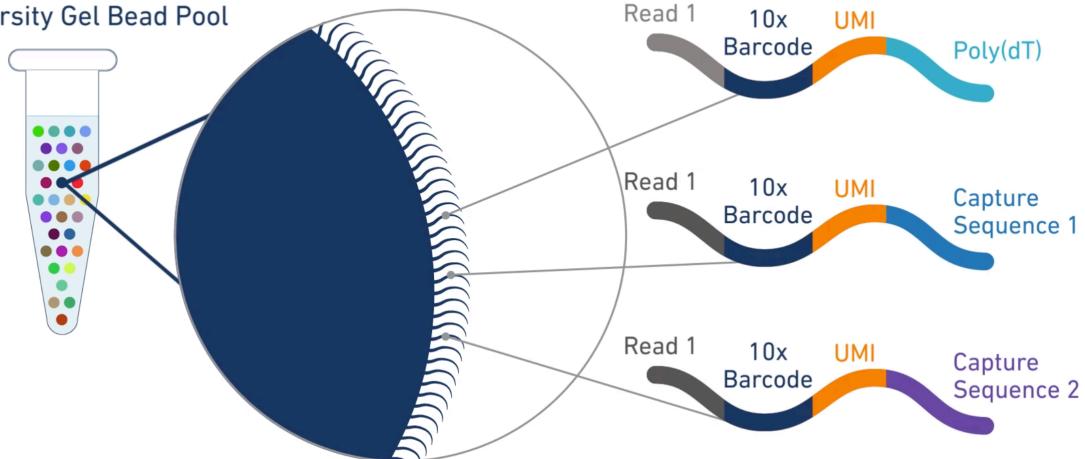


Cells isolated  
in drops with  
barcodes



# 10x Genomics

High-Diversity Gel Bead Pool



Collect



RT



Pool  
Remove Oil



Single Cell  
GEMs

10x Barcoded  
cDNA

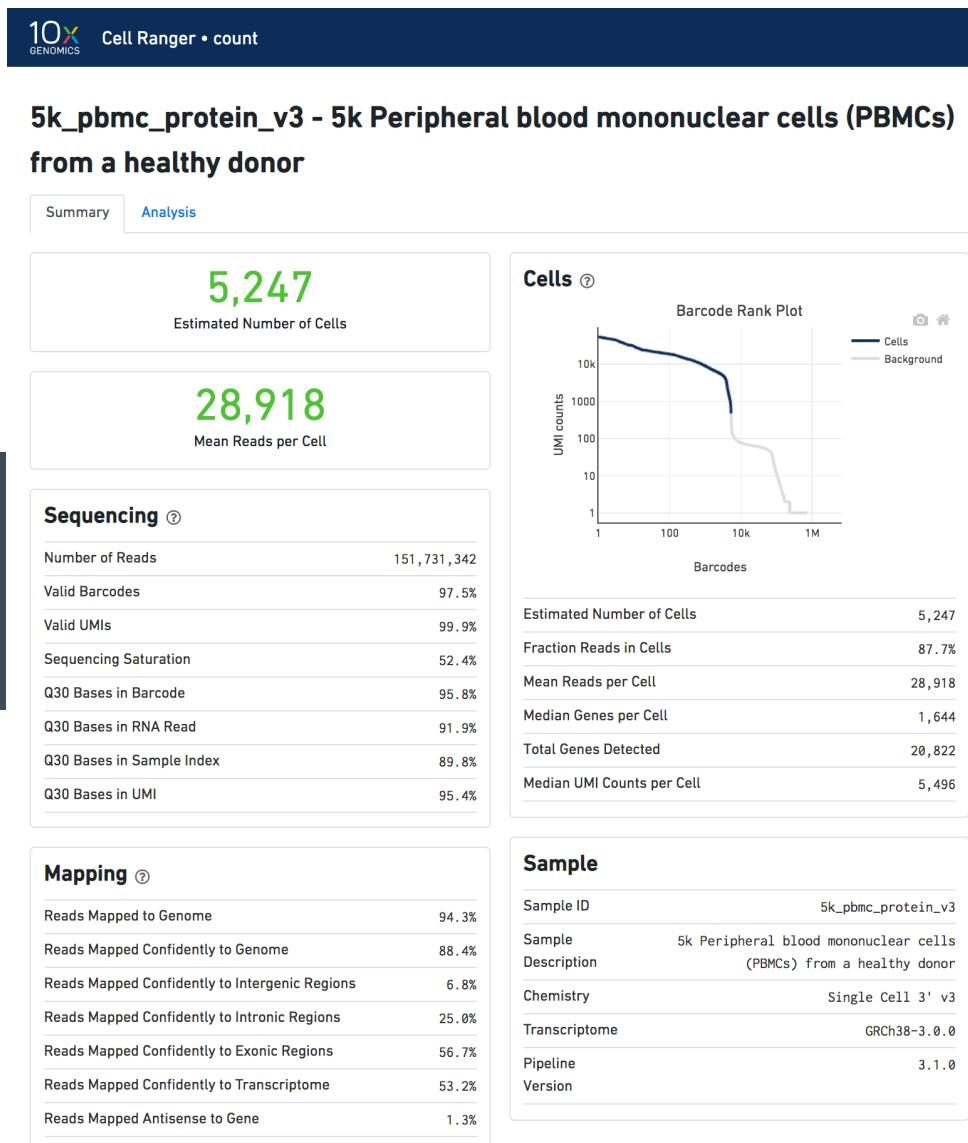
10x Barcoded  
cDNA



# 10x Genomics: Cell Ranger

- Cell Ranger count function is used to perform cell alignment to reference genome and denoising/normalisation

```
$ cd /home/jdoe/runs  
$ cellranger count --id=sample345 \  
    --transcriptome=/opt/refdata-cellranger-GRCh38-3.0.0 \  
    --fastqs=/home/jdoe/runs/HAWT7ADXX/outs/fastq_path \  
    --sample=mysample \  
    --expect-cells=1000
```



# 10x Genomics: Cell Ranger

- CellRanger ‘aggr’ is useful for combining replicates of the same sample group, but from different gem wells
- The combined replicates will maintain distinct barcodes, but will be normalised and denoised as a whole population

```
$ cd /opt/runs  
$ cellranger count --id=LV123 ...  
... wait for pipeline to finish ...  
$ cellranger count --id=LB456 ...  
... wait for pipeline to finish ...  
$ cellranger count --id=LP789 ...  
... wait for pipeline to finish ...
```

	A	B
1	library_id	molecule_h5
2	LV123	/opt/runs/LV123/outs/molecule_info.h5
3	LB456	/opt/runs/LB456/outs/molecule_info.h5
4	LP789	/opt/runs/LP789/outs/molecule_info.h5

```
library_id,molecule_h5  
LV123,/opt/runs/LV123/outs/molecule_info.h5  
LB456,/opt/runs/LB456/outs/molecule_info.h5  
LP789,/opt/runs/LP789/outs/molecule_info.h5
```

```
$ cd /home/jdoe/runs  
$ cellranger aggr --id=AGG123 \  
--csv=AGG123_libraries.csv \  
--normalize=mapped
```

# Quality Control

Based on 3 QC covariates:

- Count Depth = The number of counts per barcode
- Feature count = The number of genes per barcode
- Percent mitochondrial reads per cell (often a characteristic of damaged or dying cells)

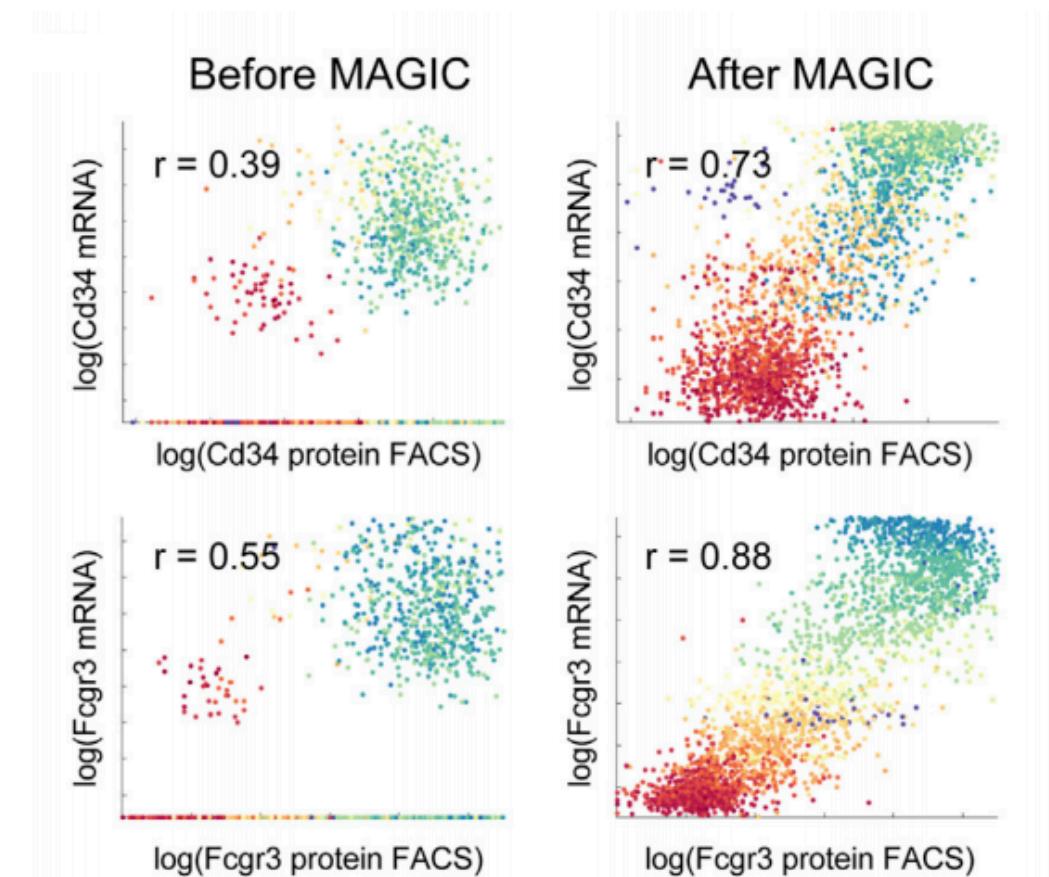
Think of a cell with a low count depth, a low number of detected genes, and a high percentage of mitochondrial reads. This is a bad cell.

Alternatively, a high count depth and a feature count could indicate a doublet cell which would not be an ideal candidate for representation.

- Three recent doublet detection tools offer more elegant and potentially better solutions (DoubletDecon: preprint: DePasquale *et al*, 2018; Scrublet: Wolock *et al*, 2019; Doublet Finder: McGinnis *et al*, 2018).

# Missing Value Imputation with MAGIC (Markov Affinity-based Graph Imputation of Cells)

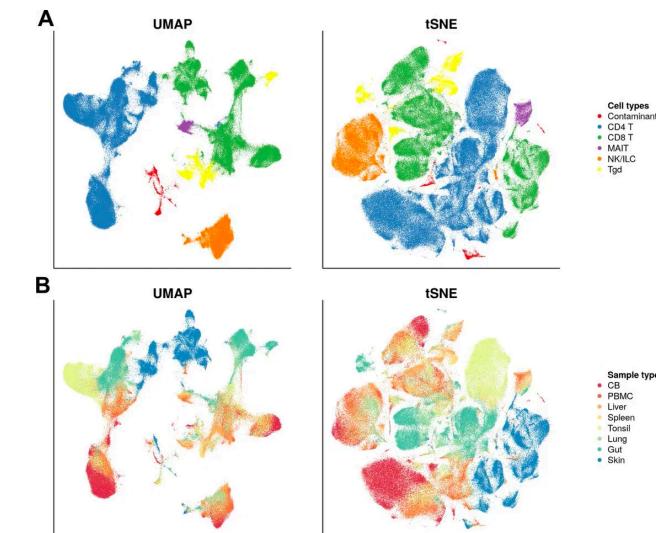
- MAGIC restores noisy and sparse single-cell data using diffusion geometry
- Imputed data are amenable to downstream analysis and can be used to predict interactions



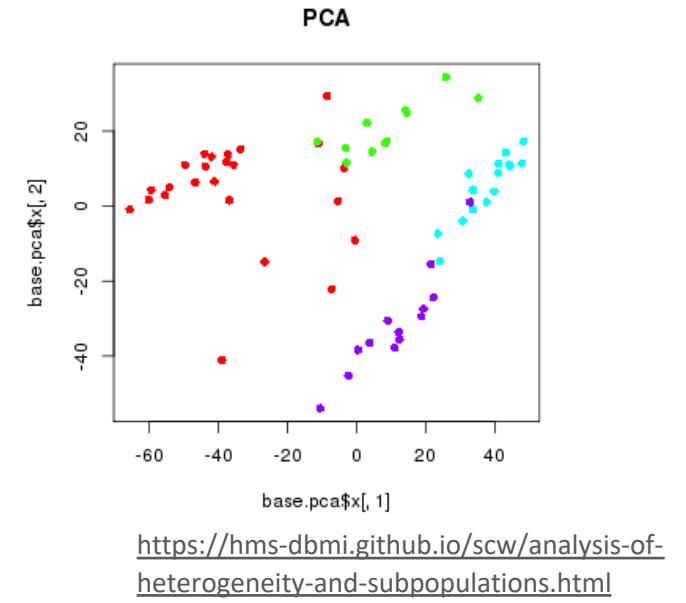
Van Dijk et al, 2018

# Clustering & Dimensionality Reduction

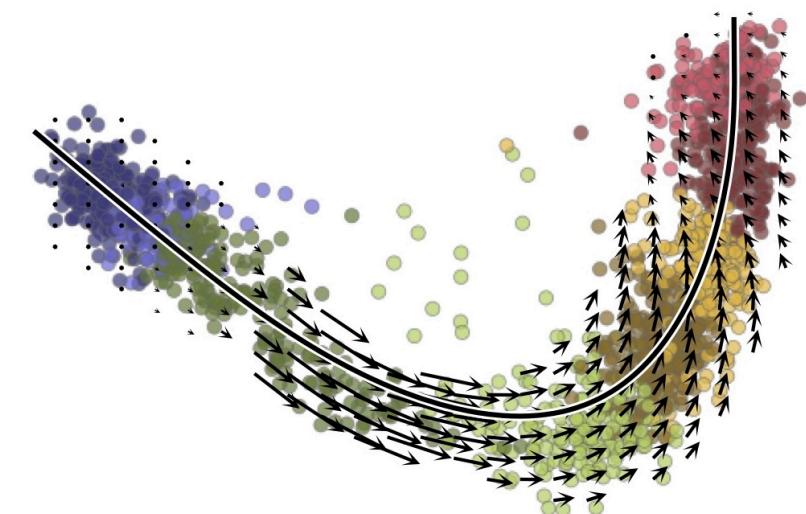
- Linear Methods (PCA)
  - Identifying batch & cell cycle effects
- Non-linear Methods (t-SNE, UMAP, RNA velocity)
  - Exploratory data analysis



<https://www.datasciencediscovery.com/index.php/tag/umap-vs-t-sne/>



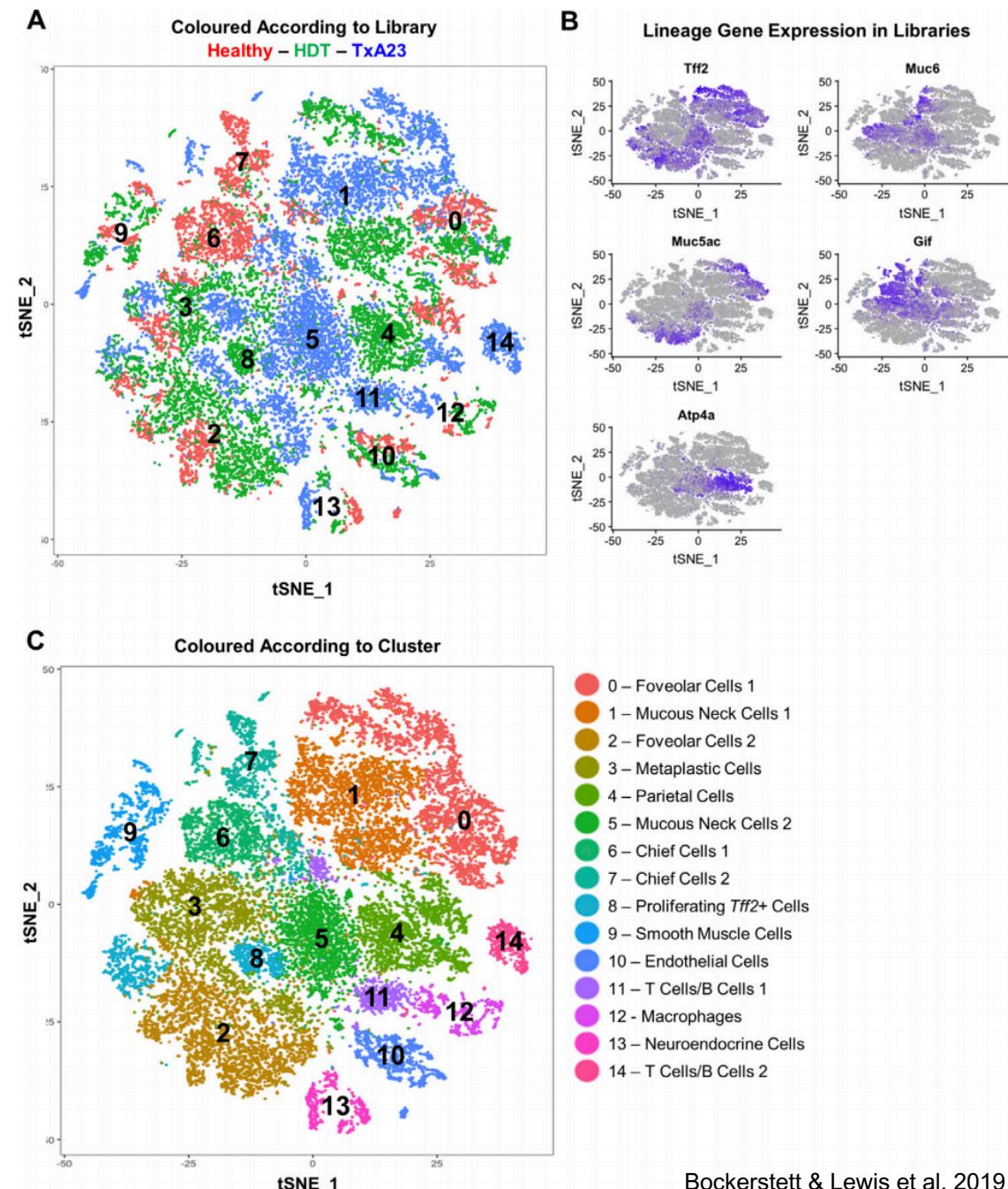
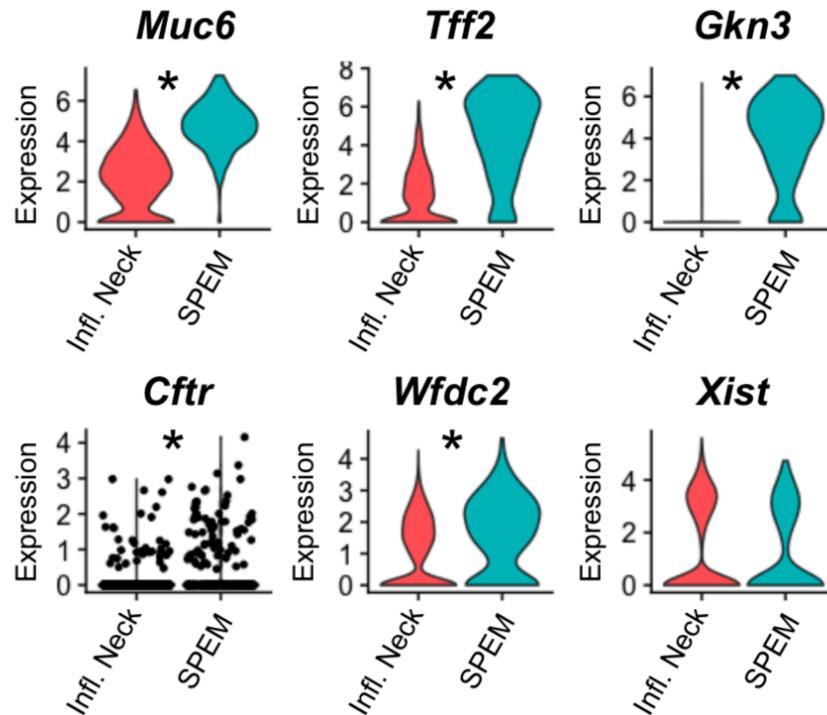
<https://hms-dbm.github.io/scw/analysis-of-heterogeneity-and-subpopulations.html>



[https://www.kallistobus.tools/velocity\\_tutorial.html](https://www.kallistobus.tools/velocity_tutorial.html)

# Dimensionality Reduction

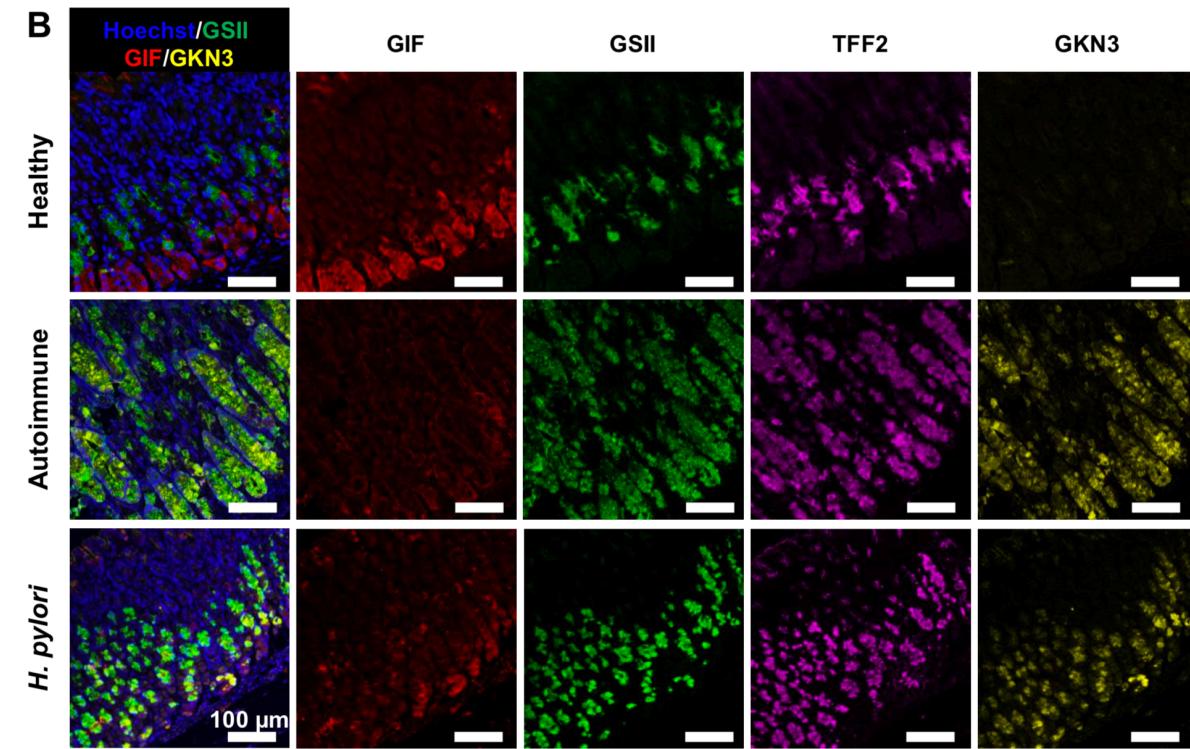
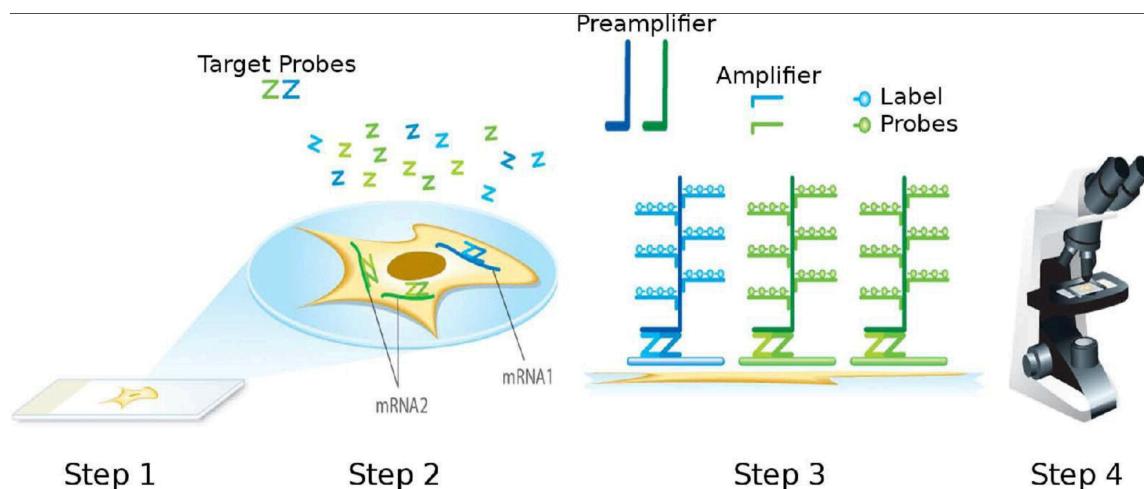
- Assign cluster identity with known biomarkers



# Novel markers can be validated

## RNA-scope

- Better signal to noise ratio achieved by multiple probes targeting different parts of the same RNA

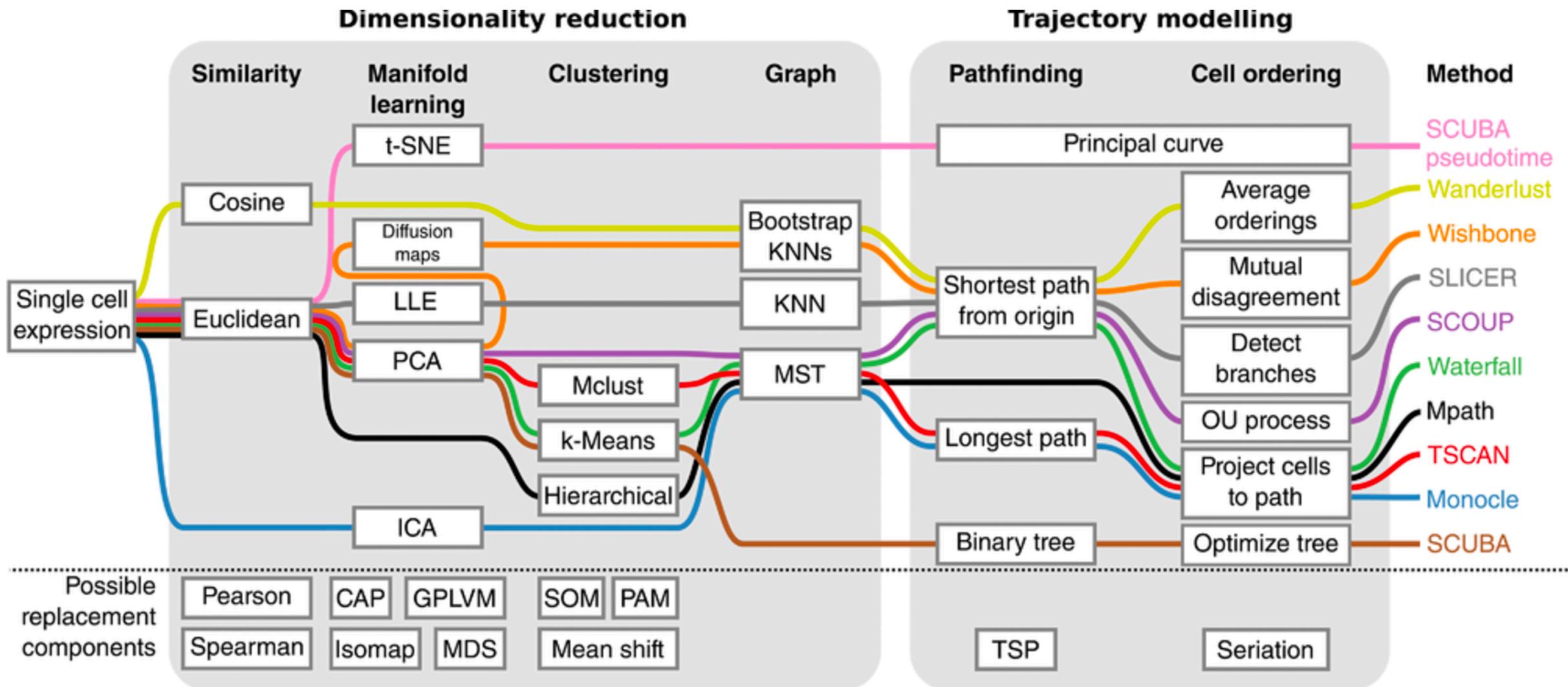


Bockerstett & Lewis et al, in review

# Pseudotime: trajectory inference

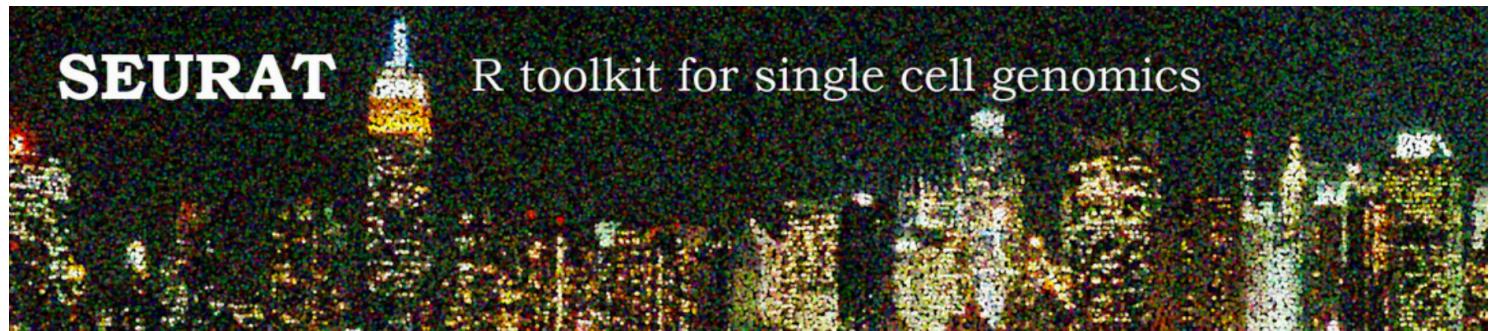
Method	SCUBA pseudotime	Wanderlust	Wishbone	SLICER	SCOUP	Waterfall	Mpath	TSCAN	Monocle	SCUBA
Visual abstract										
Structure	Linear	Linear	Single bifurcation	Branching	Branching	Linear	Branching	Linear	Branching	Branching
Robustness strategy	Principal curves	Ensemble, starting cell	Ensemble, starting cell	Starting cell	Starting population	Clustering of cells	Clustering of cells using external labelling	Clustering of cells	Differential expression	Simple model
Extra input requirements	None	Starting cell	Starting cell	Starting cell	Starting population	None	Time points	None	Time points	Time points
Unbiased	+	±	±	±	±	+	-	+	-	-
Scalability w.r.t. cells	-	-	±	±	-	±	+	+	-	±
Scalability w.r.t. genes	+	+	+	+	-	+	±	±	±	+
Code and documentation	-	±	+	±	+	±	+	+	+	±
Parameter ease-of-use	+	+	+	+	-	±	-	+	+	+

# Pseudotime: trajectory inference



# Popular analysis tools

- Seurat (Satija Lab)
  - Language: R
  - Widely used for clustering, DE testing, and multi-omics
- Monocle (Trapnell Lab)
  - Language: R
  - Native implementation of pseudotime (DDRTree)
- Scanpy (Theis Lab)
  - Language: Python
  - Native implementation of pseudotime (PAGA)



# Seurat

- Workflow for single cell data integration, visualisation and differential expression analysis
- Processed data stored in S3 objects, which are easily exported to other analysis software

