

Tools For Computational Biology – Lectures 18 & 19

Introduction to Single-cell genomics

Dominik Otto
Basic Sciences Division
Fred Hutch

dotto@fredhutch.org

Today's agenda

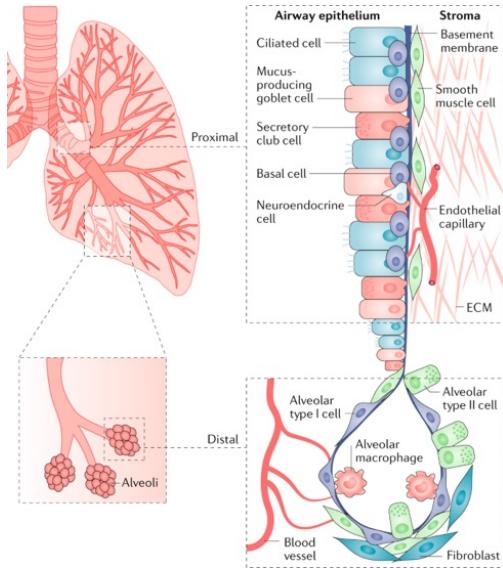
- Single-cell genomics: Why?
- Single-cell genomics: How?
- Single-cell RNA-seq preprocessing and analysis

Why single-cell genomics

Heterogeneity in tissue homeostasis

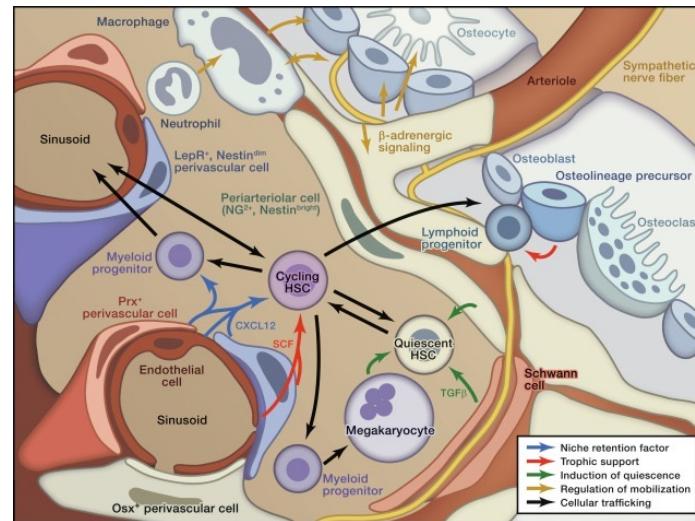
Somatosensory Cortex

Lung

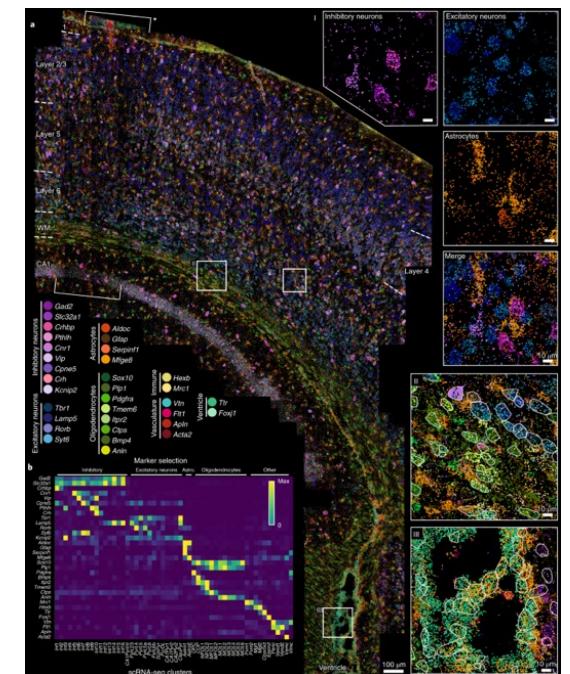


Altorki et. al., 2019

Bone Marrow

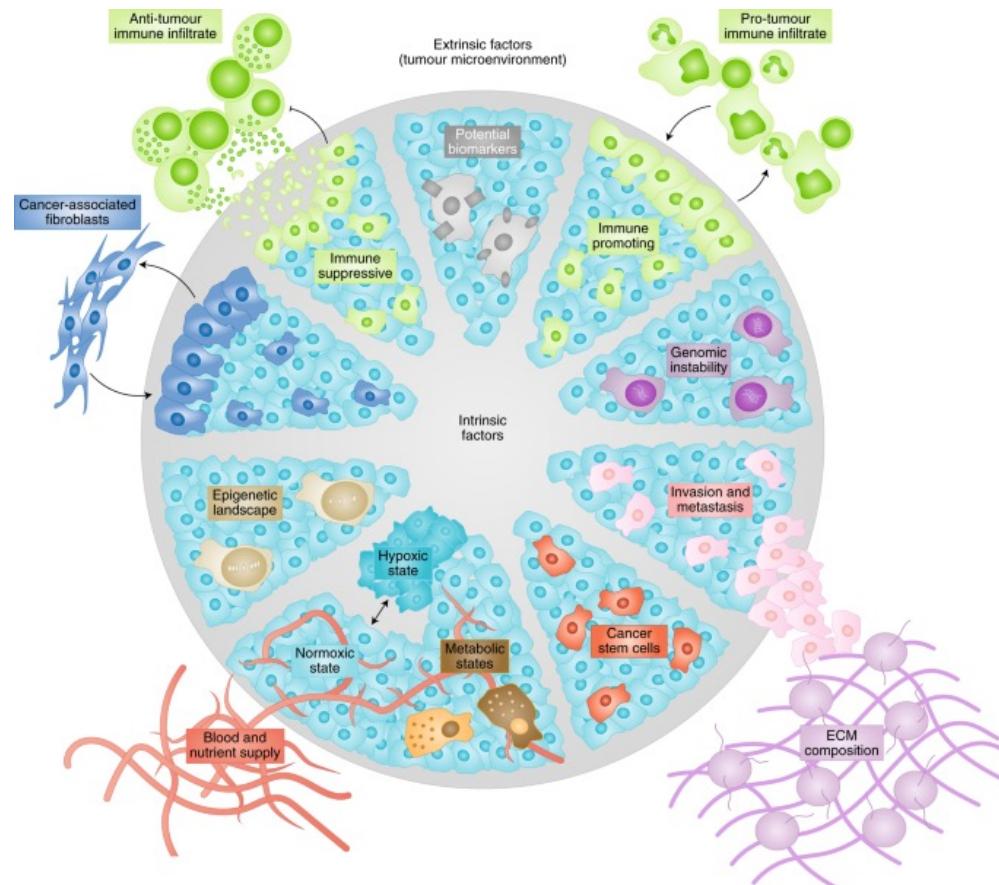


Hoffman et. al., 2020



Codeluppi et. al., 2018

Tumor heterogeneity



Lawson et. al., 2018

Single-cell technologies profile heterogeneity
instead of average



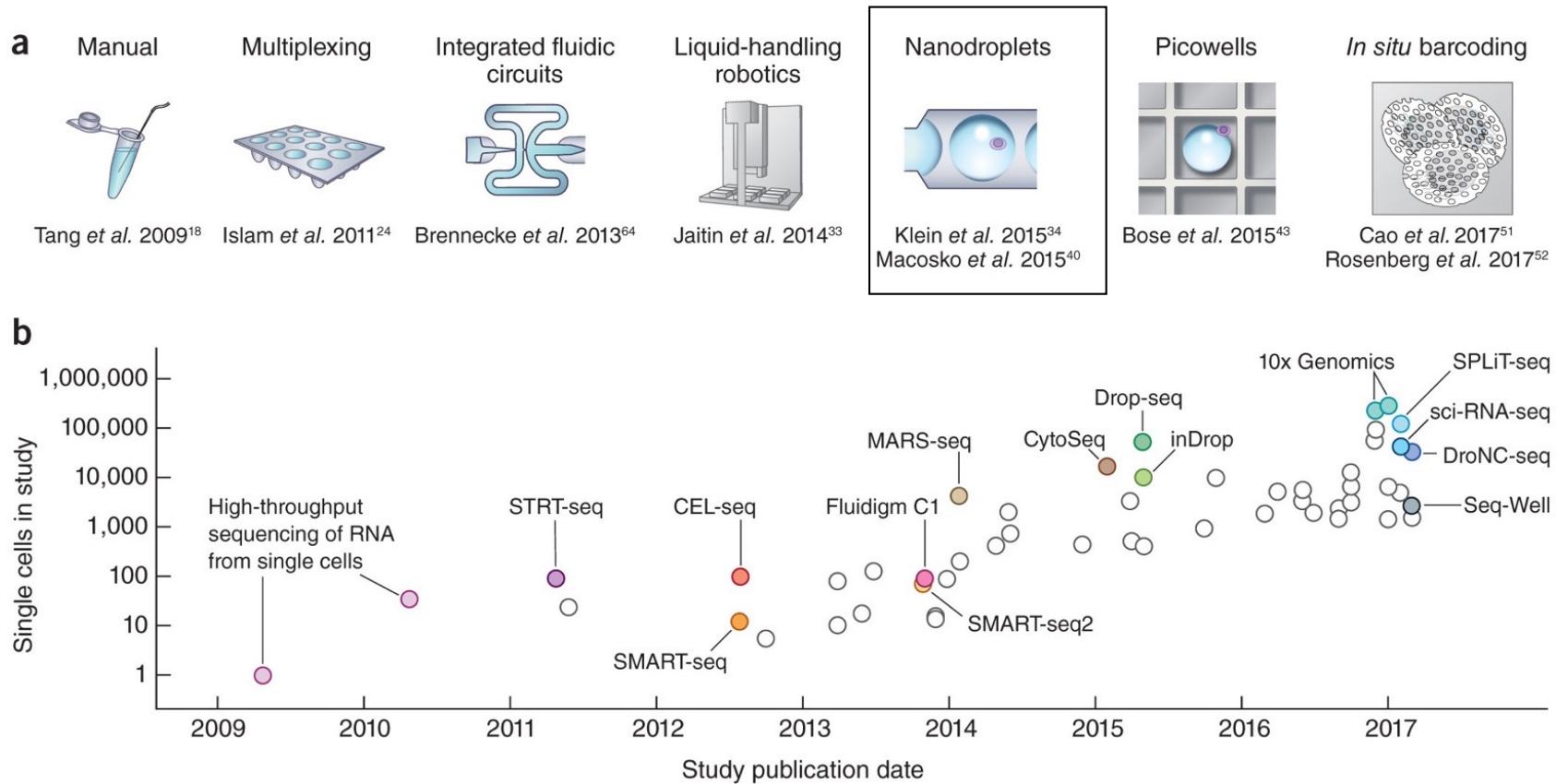
Bulk Genomics



Single-cell Genomics

Single-cell genomics: How?

Evolution of single-cell RNA-seq



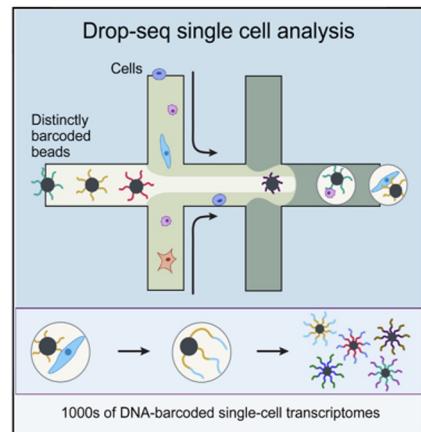
Microfluidic single-cell technologies

Moved throughput from hundreds to thousands of cells

Cell

Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets

Graphical Abstract



Resource

Authors

Evan Z. Macosko, Anindita Basu, ..., Aviv Regev, Steven A. McCarroll

Correspondence

emacosko@genetics.med.harvard.edu (E.Z.M.), mccarroll@genetics.med.harvard.edu (S.A.M.)

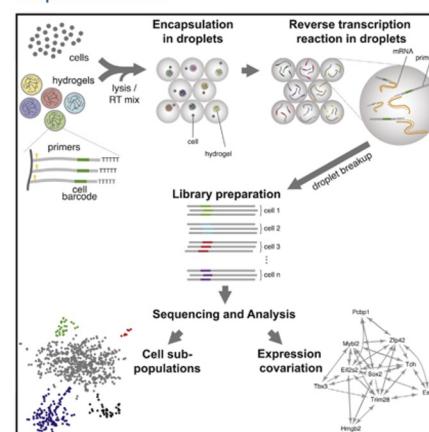
In Brief

Capturing single cells along with sets of uniquely barcoded primer beads together in tiny droplets enables large-scale, highly parallel single-cell transcriptomics. Applying this analysis to cells in mouse retinal tissue revealed transcriptionally distinct cell populations along with molecular markers of each type.

Cell

Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells

Graphical Abstract



Resource

Authors

Allon M. Klein, Linas Mazutis, ..., David A. Weitz, Marc W. Kirschner

Correspondence

weitz@seas.harvard.edu (D.A.W.), marc@hms.harvard.edu (M.W.K.)

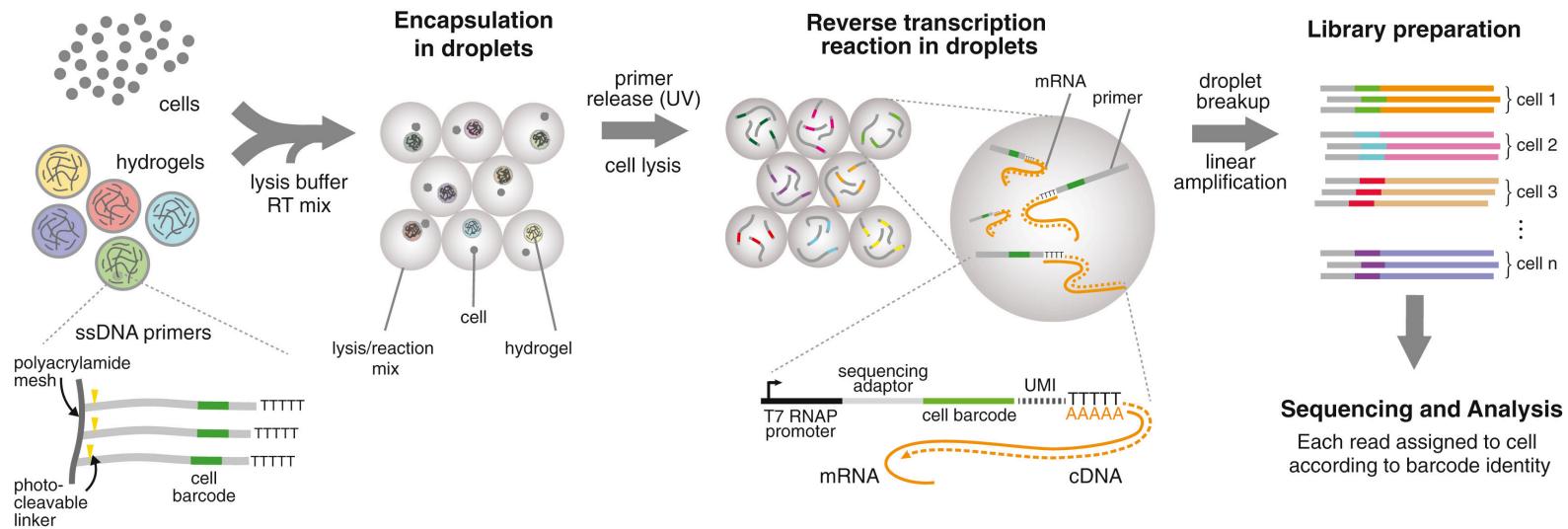
In Brief

Capturing single cells along with a set of uniquely barcoded primers in tiny droplets enables single-cell transcriptomics of a large number of cells in a heterogeneous population. Applying this analysis to mouse embryonic stem cells reveals their population structure, gene expression relationships, and the heterogeneous onset of differentiation.

Klein et al *Cell* 2015
Macosko et al *Cell* 2015

Microfluidic single-cell technologies

Moved throughput from hundreds to thousands of cells

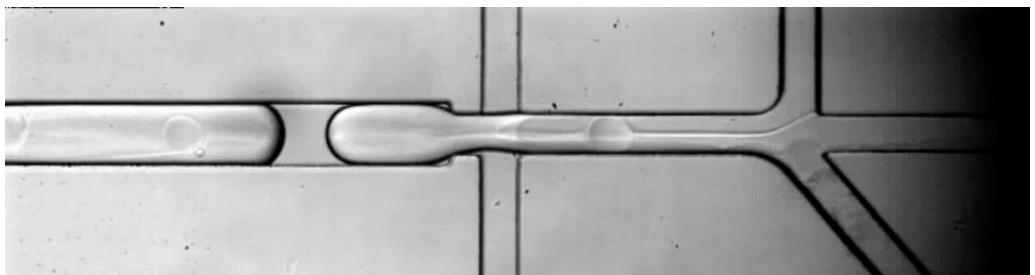
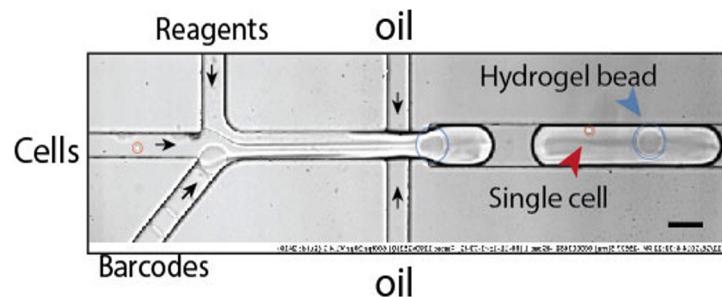


- Droplet-based processing using microfluidics
- Nanoliter scale aqueous drops in oil
- 3' End
- Bead based for cell barcoding
- Uses UMI (Unique Molecular Identifier).

Klein et al *Cell* 2015
Macosko et al *Cell* 2015

Microfluidic single-cell technologies

Moved throughput from hundreds to thousands of cells



Klein et al *Cell* 2015
Macosko et al *Cell* 2015

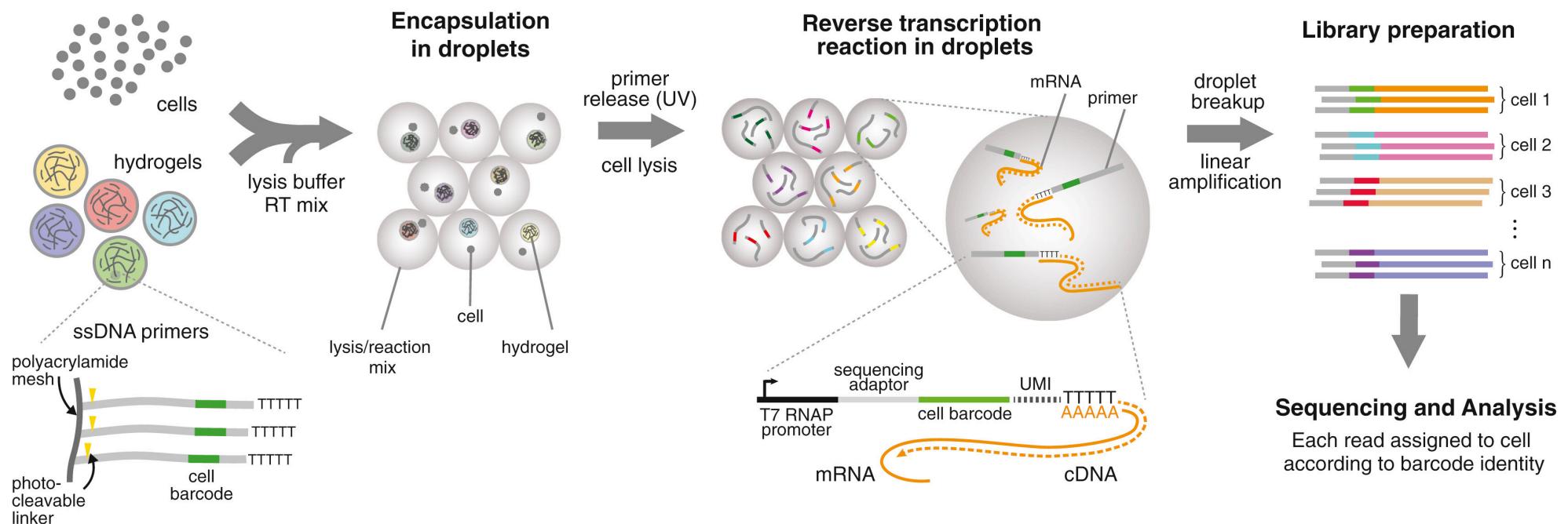
CellXGene interactive tool

```
uvx cellxgene launch 10x_pbmc.h5ad
```

```
http://localhost:5005
```

Single-cell RNA-seq: Preprocessing and Analysis

Microfluidic single-cell technologies



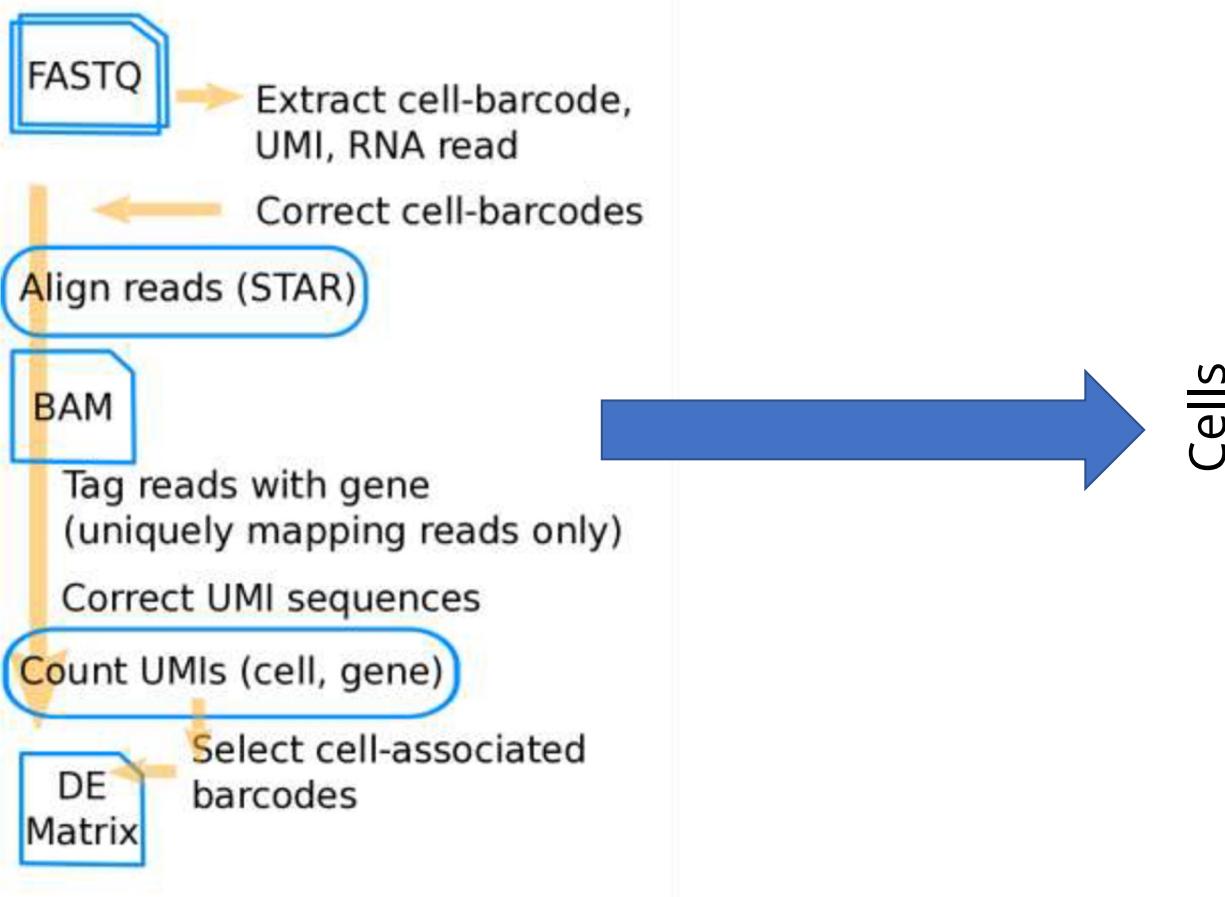
Klein et al *Cell* 2015

Cell Barcodes and UMIs

- Cell Barcode: Unique cell identifier – Whitelist
- UMI: Unique molecular identifier – Random 8mer

	Cellular barcode	UMI	
Cell 1	TTGCCGTGGTGT	GGCGGGGA	CGGTGTTA] <i>DDX51</i>
	TTGCCGTGGTGT	TATGGAGG	CCAGCACC] <i>NOP2</i>
	TTGCCGTGGTGT	TCTCAAGT	AAAATGGC] <i>ACTB</i>
Cell 2	CGTTAGATGGCA	GGGCCGGG	CTCATAGT] <i>LBR</i>
	CGTTAGATGGCA	ACGTTATA	ACCGGTAC] <i>ODF2</i>
	CGTTAGATGGCA	TCGAGATT	AGCCCTTT] <i>HIF1A</i>
Cell 3	AAATTATGACCGA	AGTTTGTA	GGGAATTAA] <i>ACTB</i> ← 2 reads, 1 molecule
	AAATTATGACCGA	AGTTTGTA	AGATGGGG]
	AAATTATGACCGA	TGTGCTTG	GACTGCAC] <i>RPS15</i>
Cell 4	GTAAACGTACC	CTAGCTGT	GATTTCT] <i>GTPBP4</i>
	GTAAACGTACC	GCAGAACT	GTGGCGT] <i>GAPDH</i>
	GTAAACGTACC	AAGGCTTG	CAAAGTTC] <i>ARL1</i> ← 2 reads, 2 molecules
	GTAAACGTACC	TTCCGGTC	TCCAGTCG]
.....			
(Thousands of cells)			

10X Preprocessing Pipeline: CellRanger



Genes

Cells

0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	0	0	62			
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	0	0	0	0			
0	38	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0		
99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0		
0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	0	0	40	0	0	0	0	0	5	0	84	0	0	0	0	0	0	0	0	0	0	0		
10	36	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	48	0	95	0	3	0	73	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0		
0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53	0	0	52	0	0	0	0	0		
0	0	3	0	0	77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54	0		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	52	0	0	0	0	94	0	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	90	0	0	0	85	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75		
0	0	22	0	0	0	0	0	0	0	0	65	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	0	0	0	0	0	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	0	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54	0		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0		
0	0	0	0	0	0	0	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26		
0	71	0	46	0	0	0	0	0	52	0	0	27	0	0	76	0	0	0	0	0	0	0	0	87		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	72	69	0	0	0	0	0	0	77		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	67	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	94	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	62	0	0	0	5	0	0	0	0	0	0	39	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	30	74	0	0	22	0	0	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	73	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	90	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	0	0	
0	90	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	12	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	8	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	11	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	75	0	0	63	0	0	16	0	0	19	0	36	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	96	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	57	0	0	92	0	0	0	0	46	0	0	0	0	99	0	0	0	0	0	0	0

Genes

Scale & Resolution X *Noise & Sparsity*

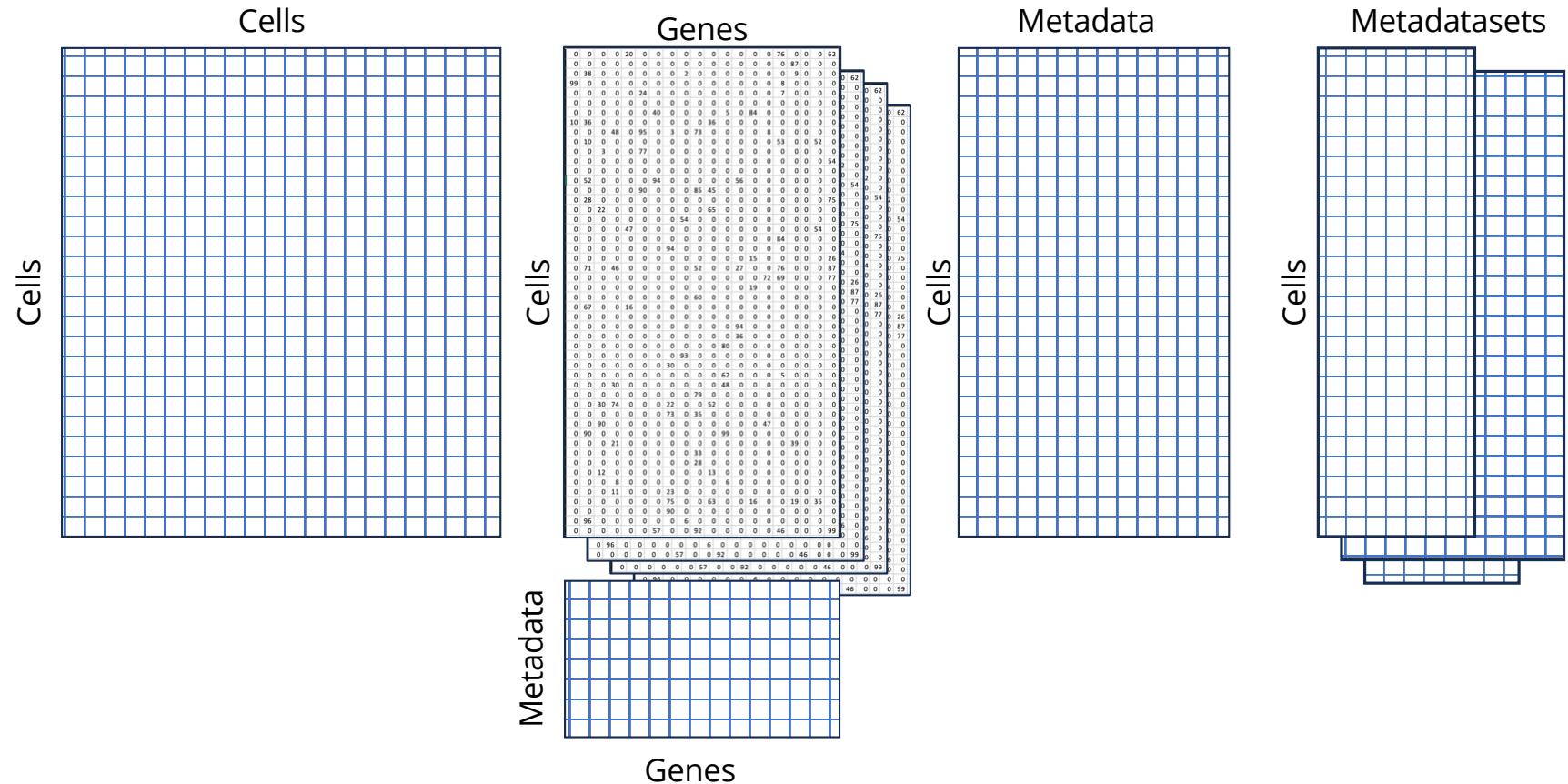


Dendritic
CD14 Monocytes
FCGR3A Monocytes
Megakaryocytes

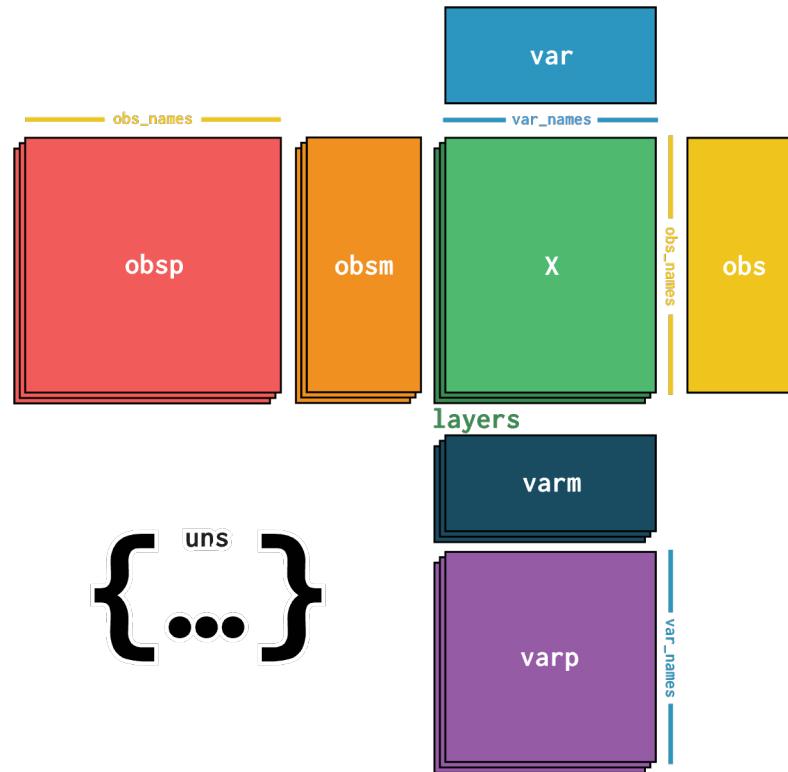
CD4 T
CD8 T
NK

B

Single-cell Data



Single-cell Data



<https://anndata.readthedocs.io>

VSCode

→ Lecture18-scRNA-seq-analysis.ipynb

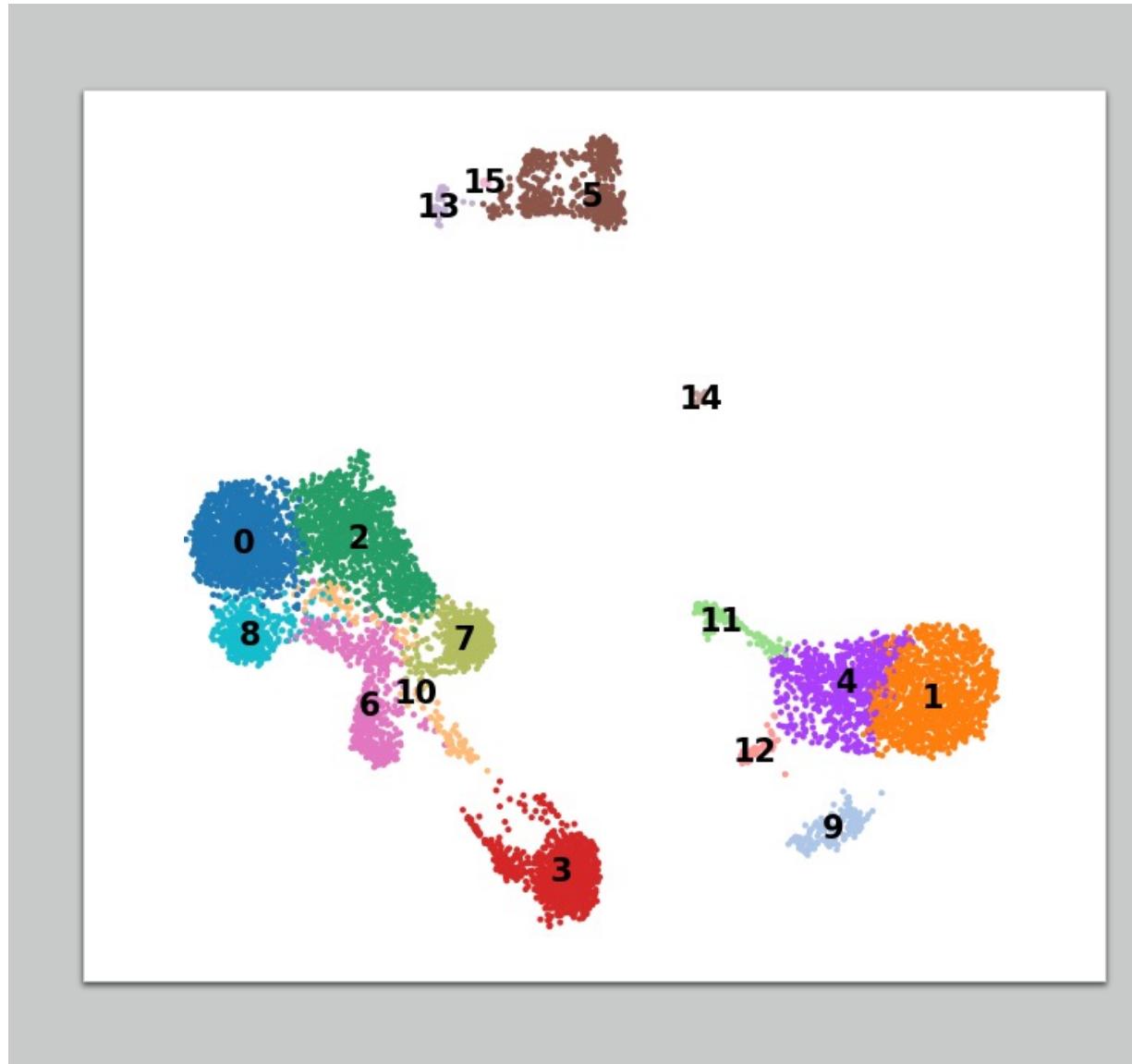
A black-box view

```
import scanpy as sc

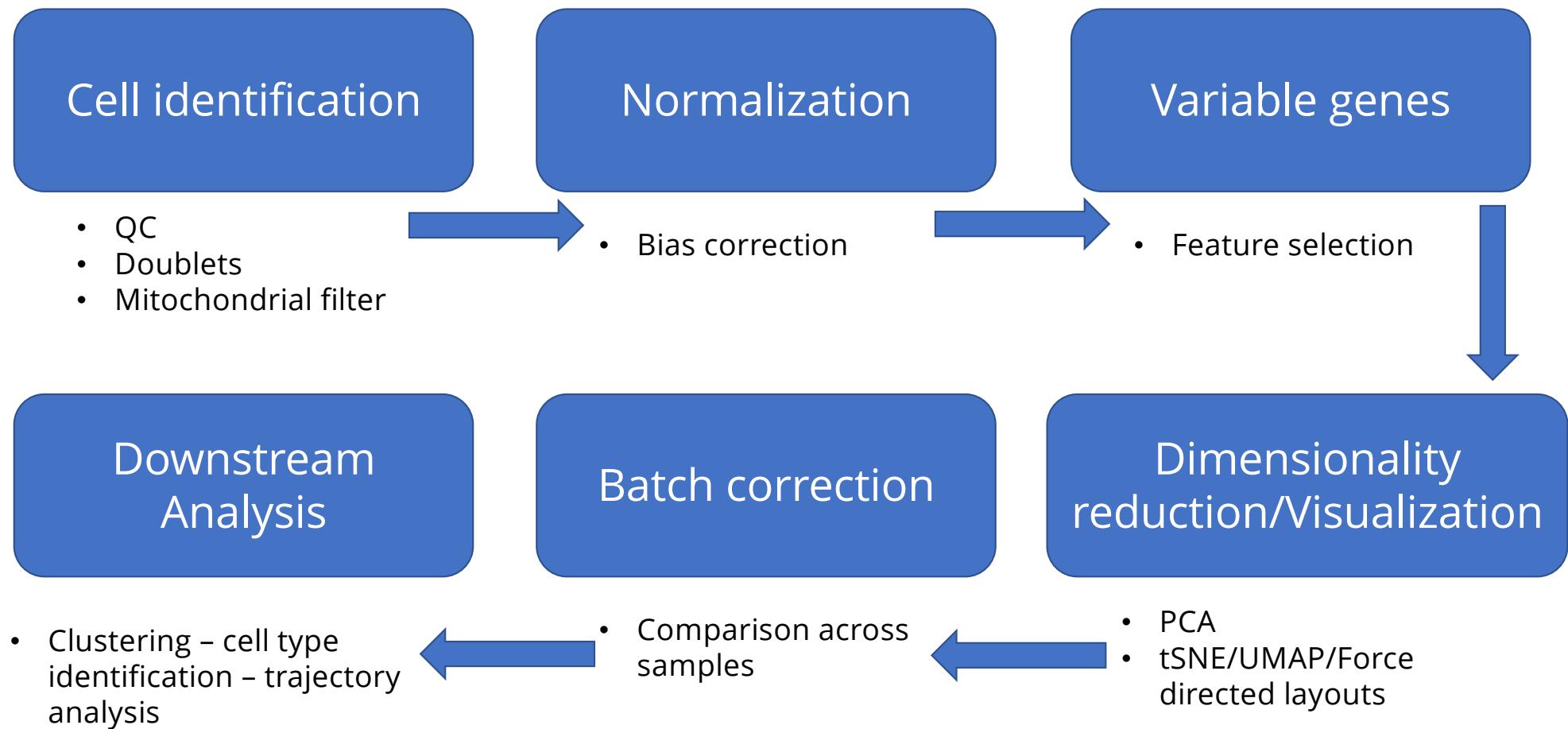
ad = sc.read( <counts file> )
sc.pp.normalize_total(ad)
sc.pp.log1p(ad)

sc.pp.highly_variable_genes(ad)
sc.pp.pca(ad)

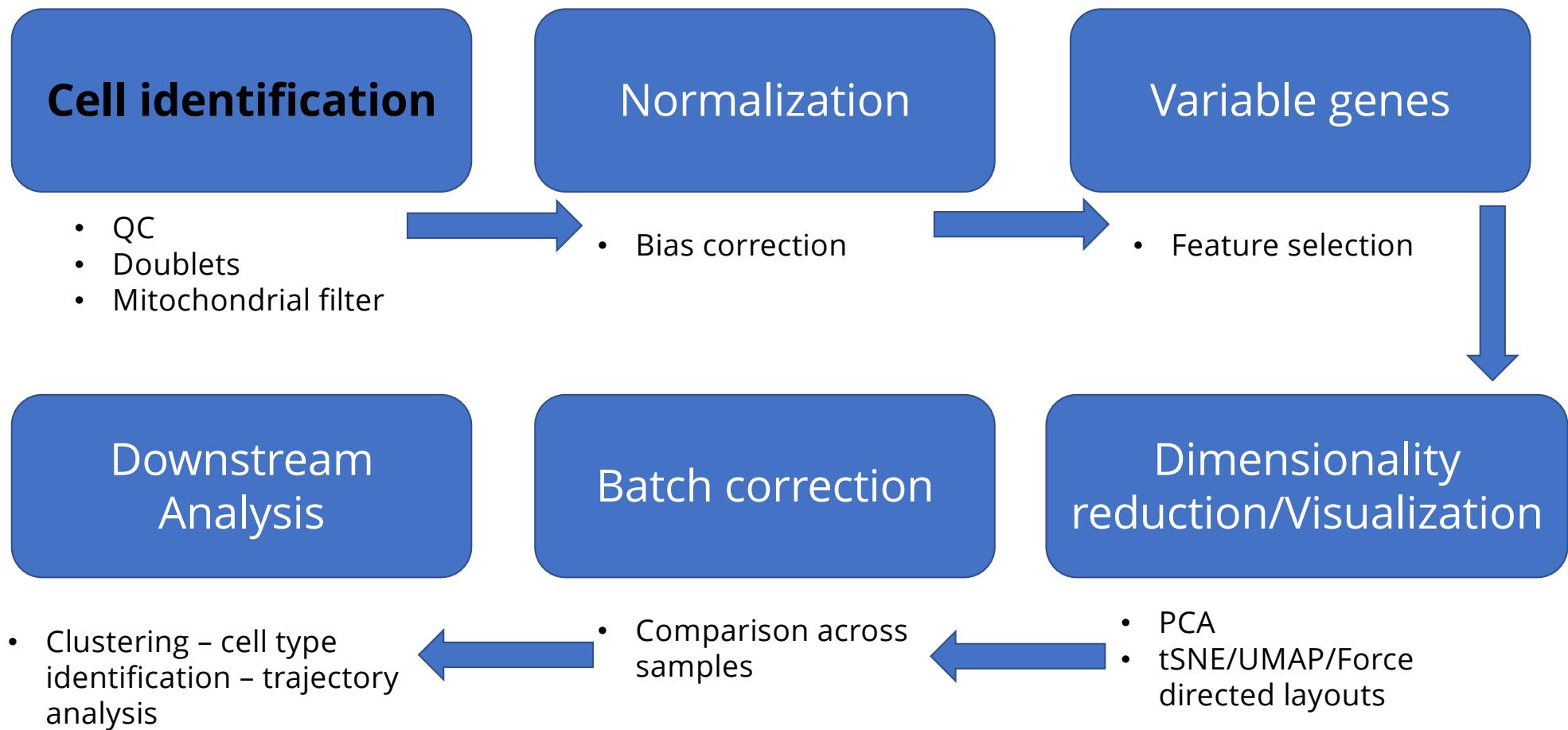
sc.pp.neighbors(ad)
sc.tl.leiden(ad)
sc.tl.umap(ad)
```



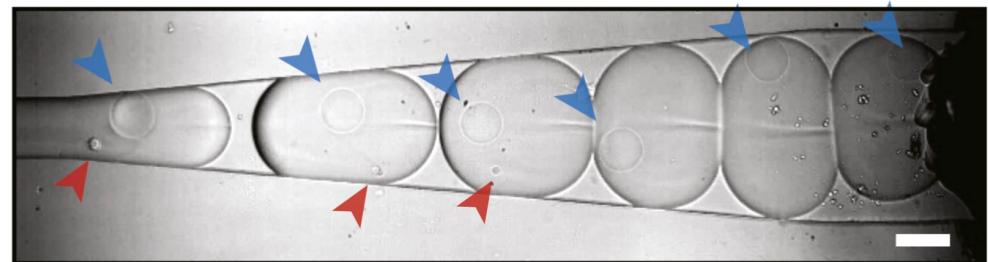
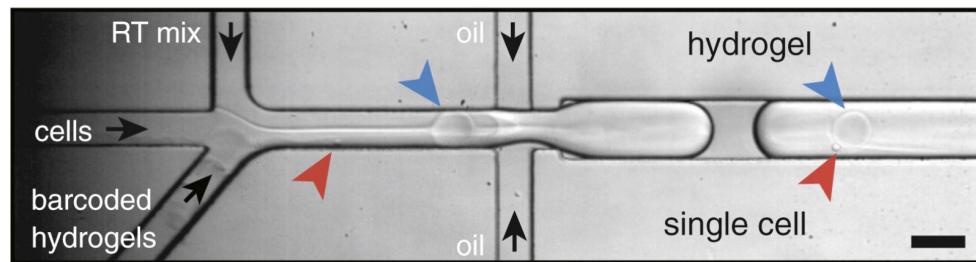
scRNA-seq analysis steps



scRNA-seq analysis steps



scRNA-seq: Empty droplets & Ambient RNA



- Most droplets do not have cells!
 - Ambient RNA

Single-cell RNA-seq

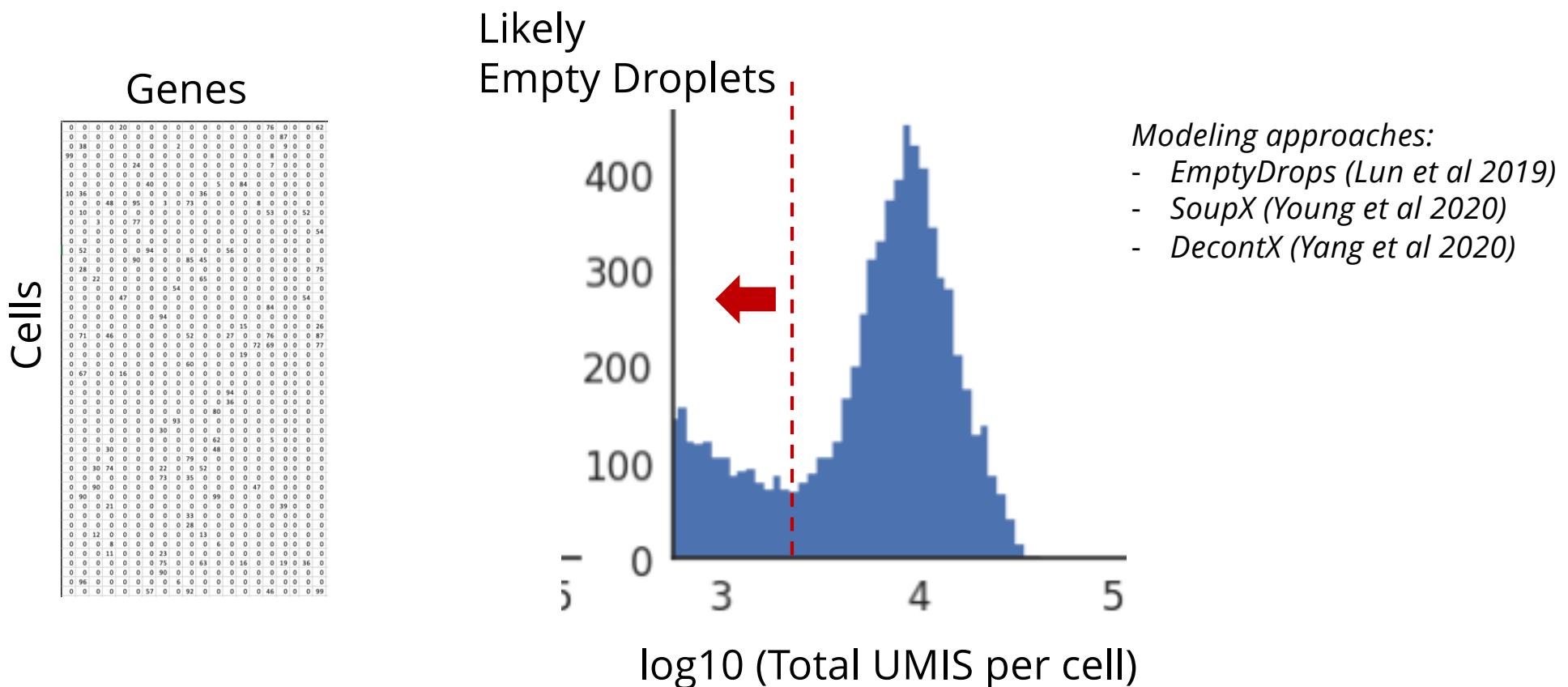
All rows are not real cells

Genes

Cells

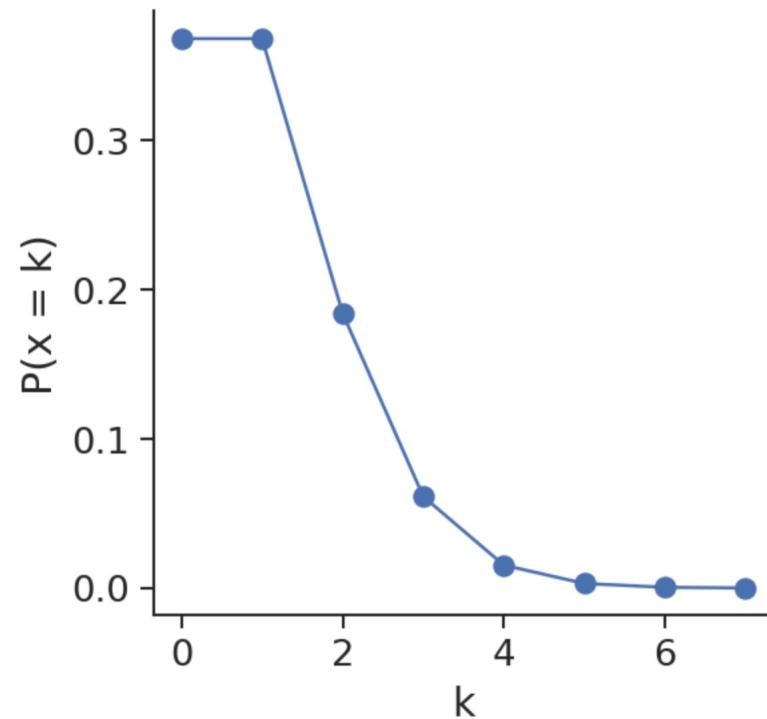
0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	76	0	0	62
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	0	0	0
0	38	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	9	0	0	0
99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0
0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	40	0	0	0	0	5	0	84	0	0	0	0	0	0	0	0
10	36	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0
0	0	0	48	0	95	0	3	0	73	0	0	0	0	8	0	0	0	0	0	0
0	10	0	0	0	0	0	0	0	0	0	0	0	0	53	0	0	52	0	0	0
0	0	3	0	0	77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	52	0	0	0	94	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	90	0	0	0	85	45	0	0	0	0	0	0	0	0	0	0
0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75
0	0	22	0	0	0	0	0	0	65	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	54	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	54	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	0	0	0	0	0
0	0	0	0	0	0	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26
0	71	0	46	0	0	0	0	0	52	0	0	27	0	0	76	0	0	0	87	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	72	69	0	0	0	77	0
0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0
0	67	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	94	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	62	0	0	0	5	0	0	0	0	0	0
0	0	0	30	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0	0	0	0	0
0	0	30	74	0	0	22	0	0	52	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	73	0	35	0	0	0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	0
0	90	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0
0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0
0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0
0	0	12	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0
0	0	0	8	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
0	0	0	11	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	75	0	0	63	0	0	16	0	0	19	0	36	0	0
0	96	0	0	0	0	0	0	90	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	57	0	92	0	0	0	0	46	0	0	0	99	0	0	0

scRNA-seq: Removal of empty droplets

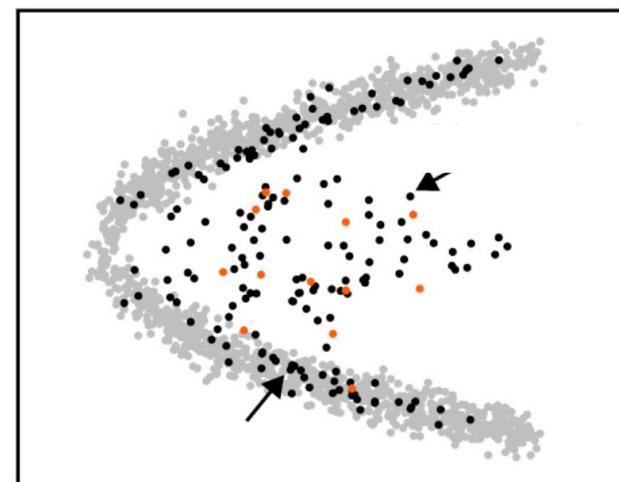
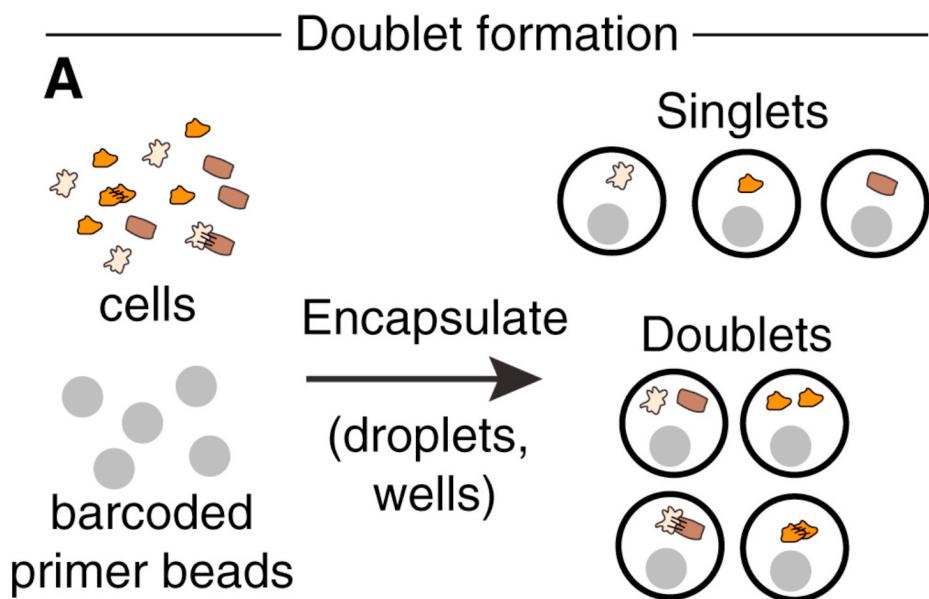


Cell containing droplets

- Cell encapsulation follows a Poisson distribution – are there reasons beyond ambient RNA, that can lead to misleading biology



Doublets

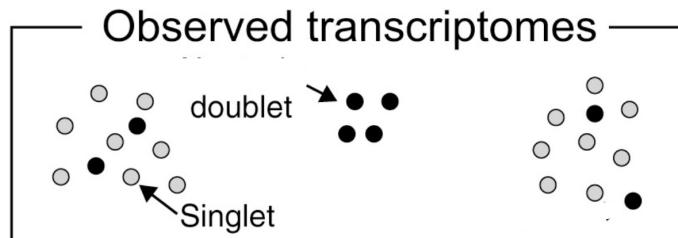


Wolock et al 2019

How to detect doublets?

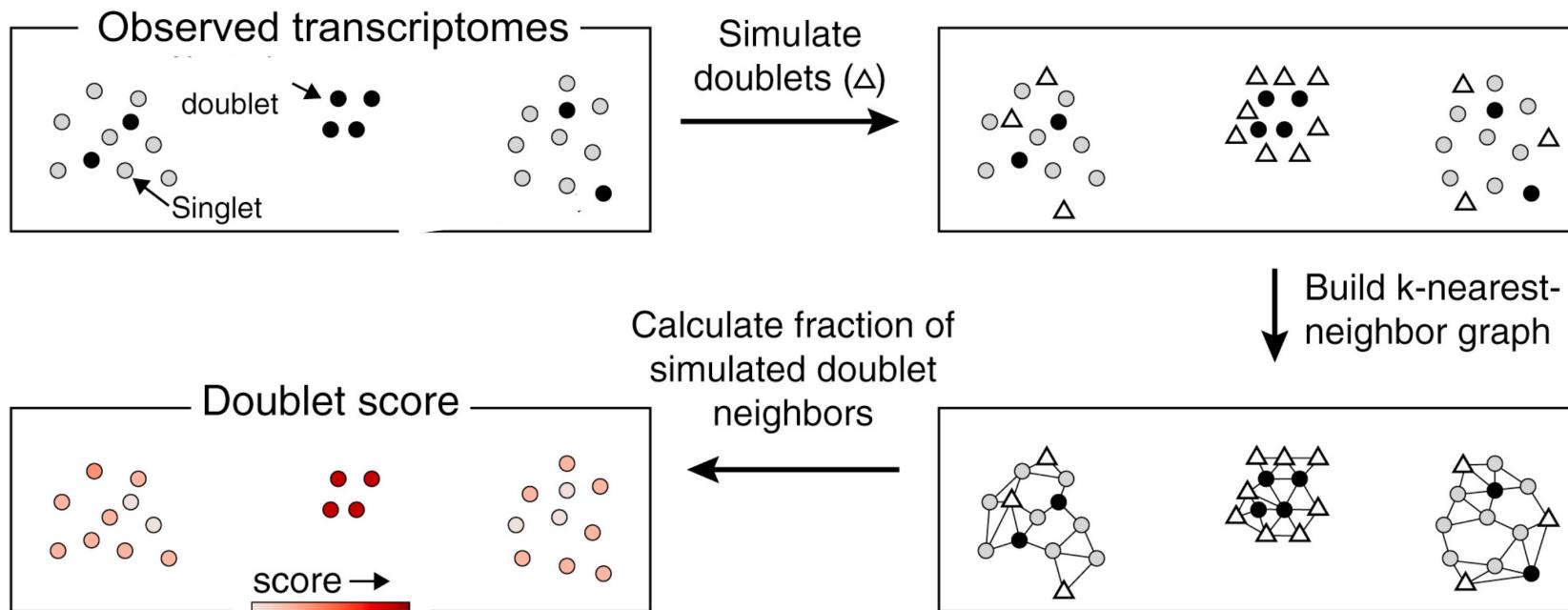
- Assume:

- Multiplets / Doublets are rare
- Constituent singlets are present



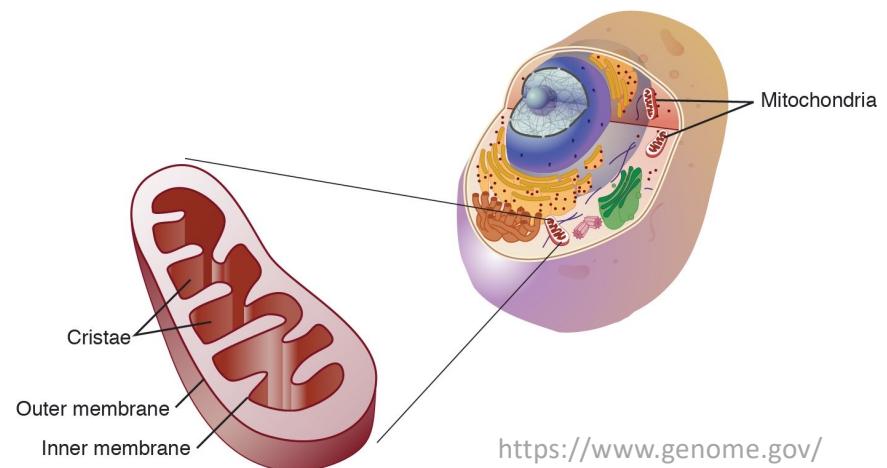
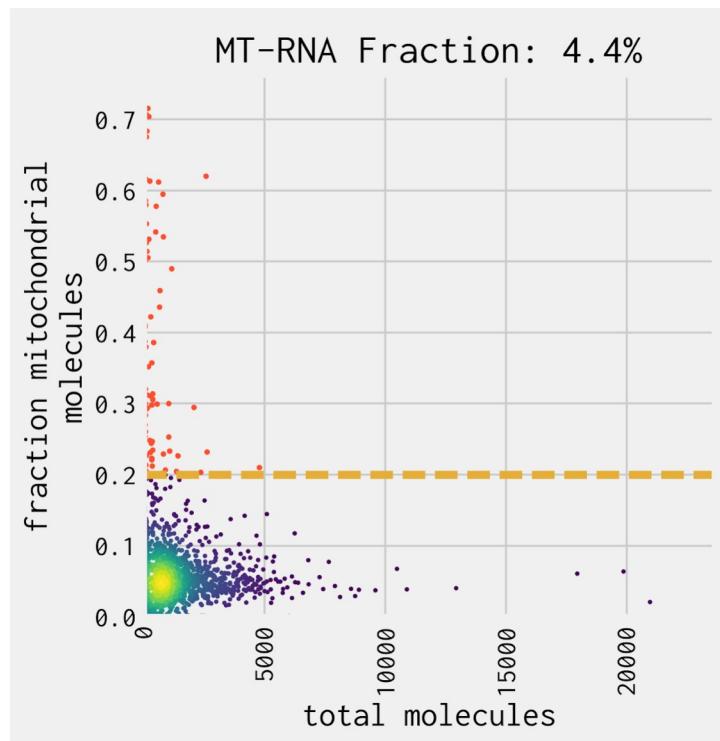
How to detect doublets?

- Assume:
 - Multiplets / Doublets are rare
 - Constituent singlets are present



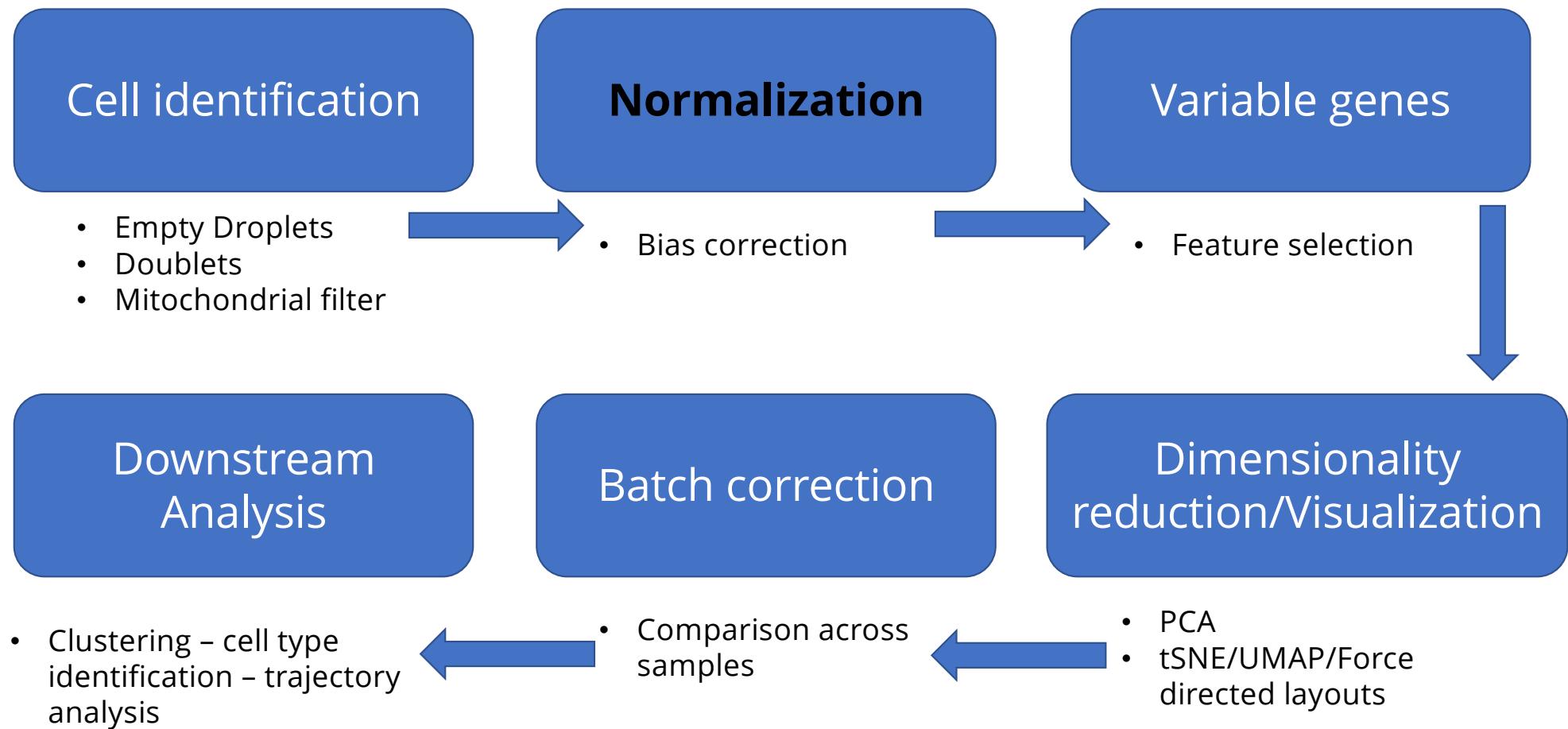
- Can this detect all possible multiplets?

Mitochondrial filter



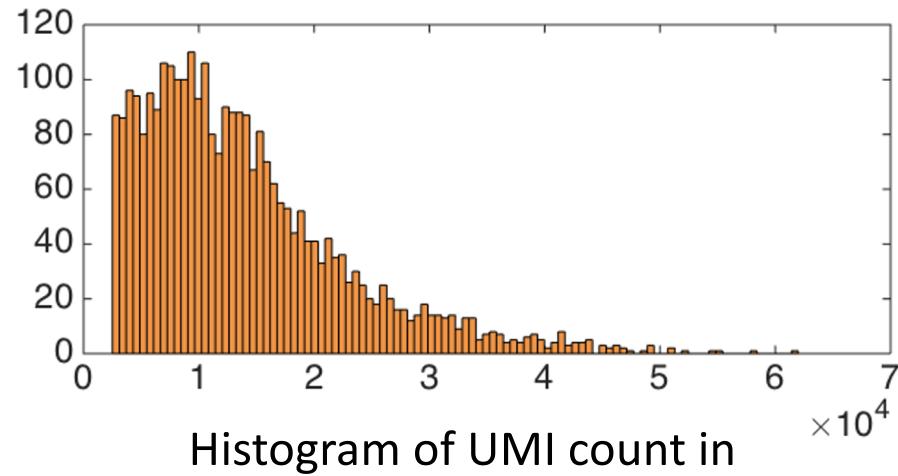
- Calculate fraction of molecules from MT genes
- Exclude cells with > 20% (optional)

scRNA-seq analysis steps



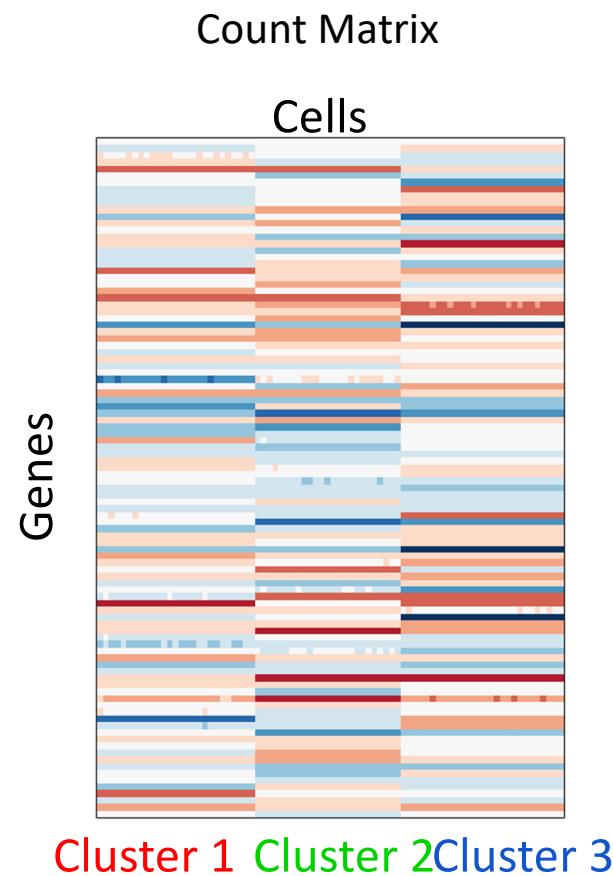
Why Normalize?

- Larger skew in distribution of total molecules (UMIs) per cell, i.e. library size
- Expression values not comparable across cells
 - *Measuring distance between cells*



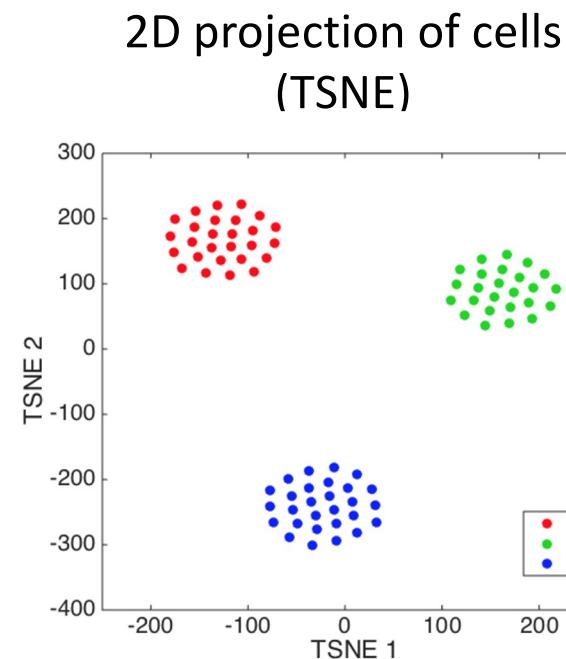
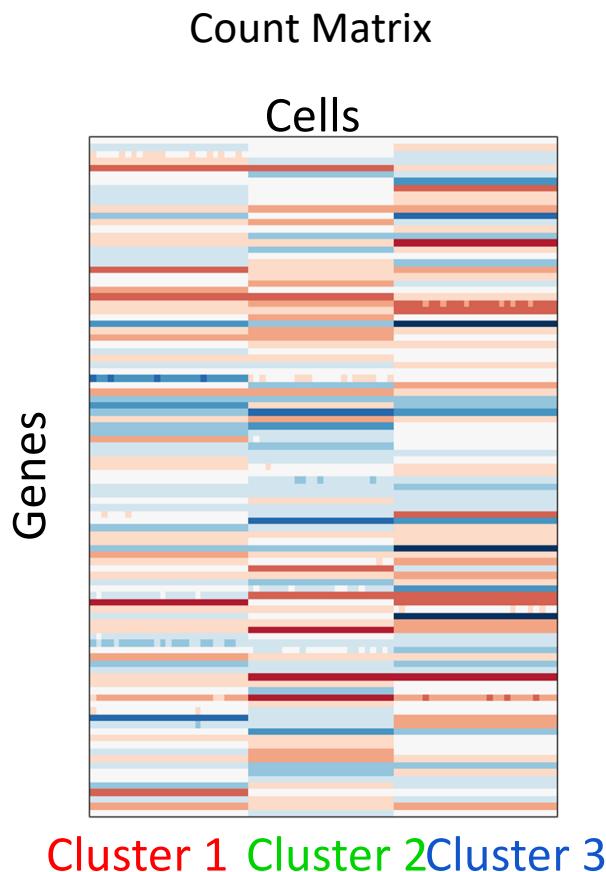
Histogram of UMI count in
example SC dataset
From Zeisel, Science 2014

Why is this problematic?

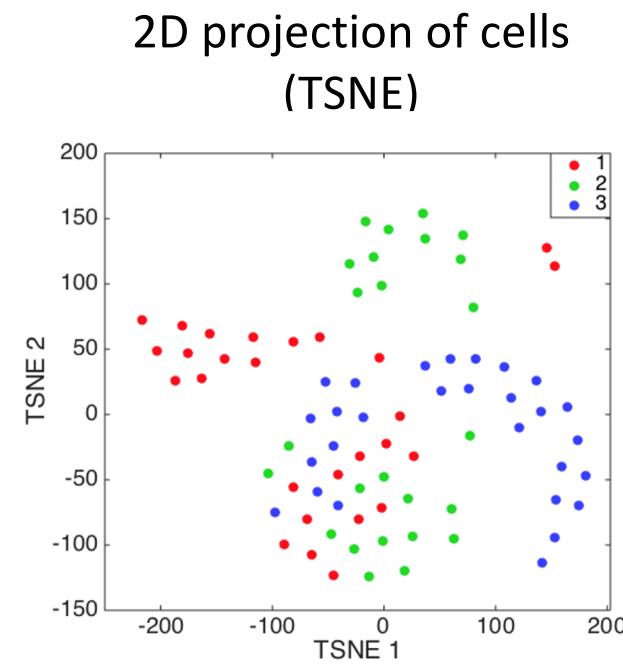
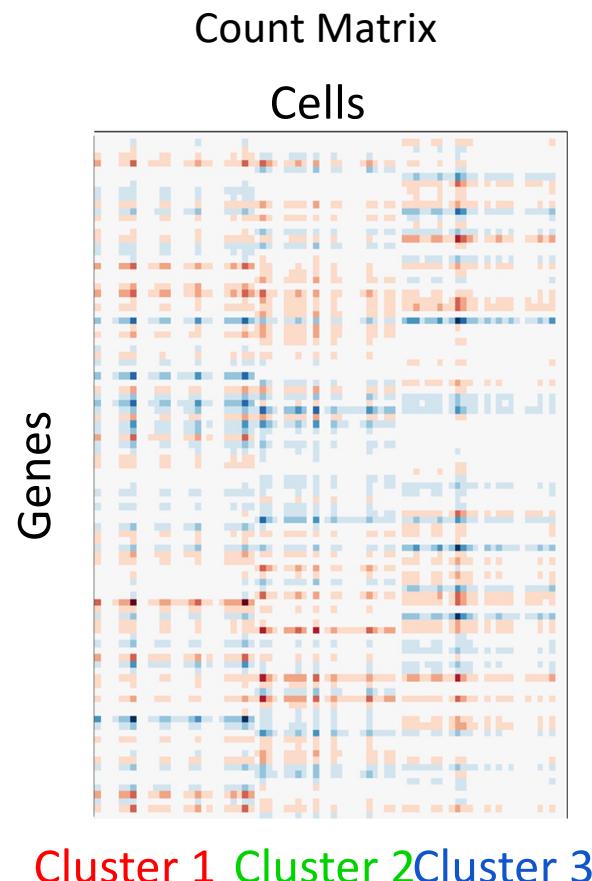


Slides courtesy of Elham Azizi

Why is this problematic?

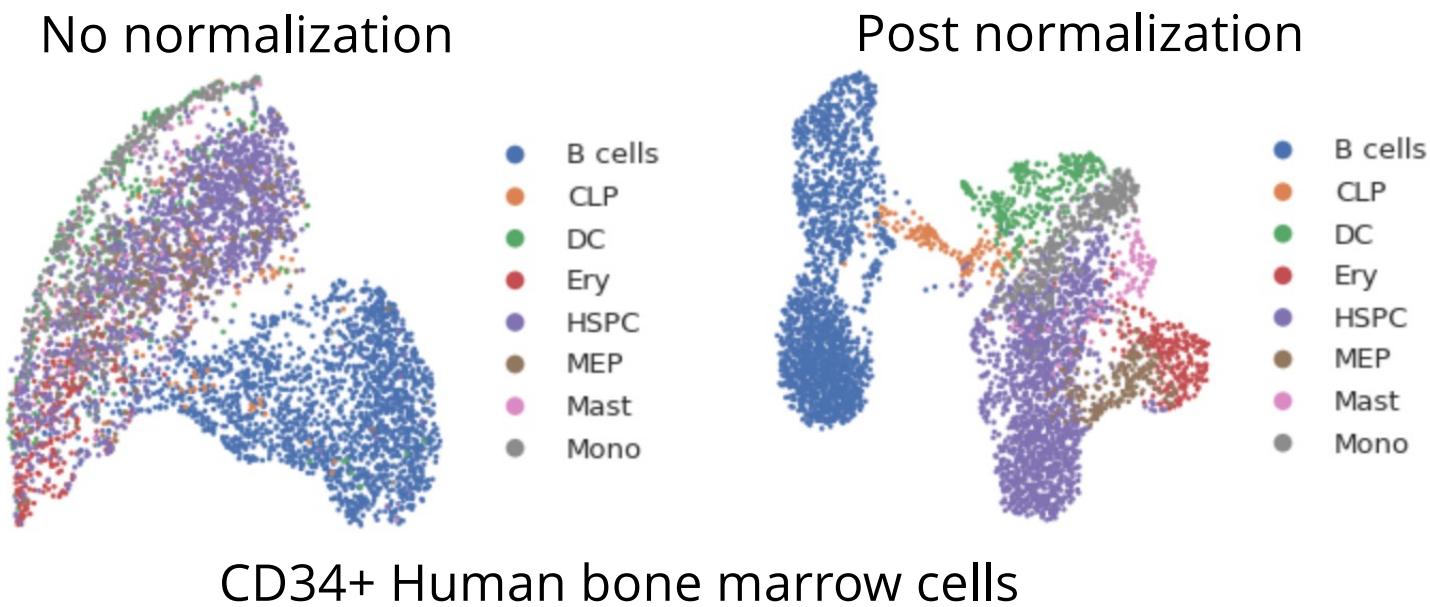


Why is this problematic?

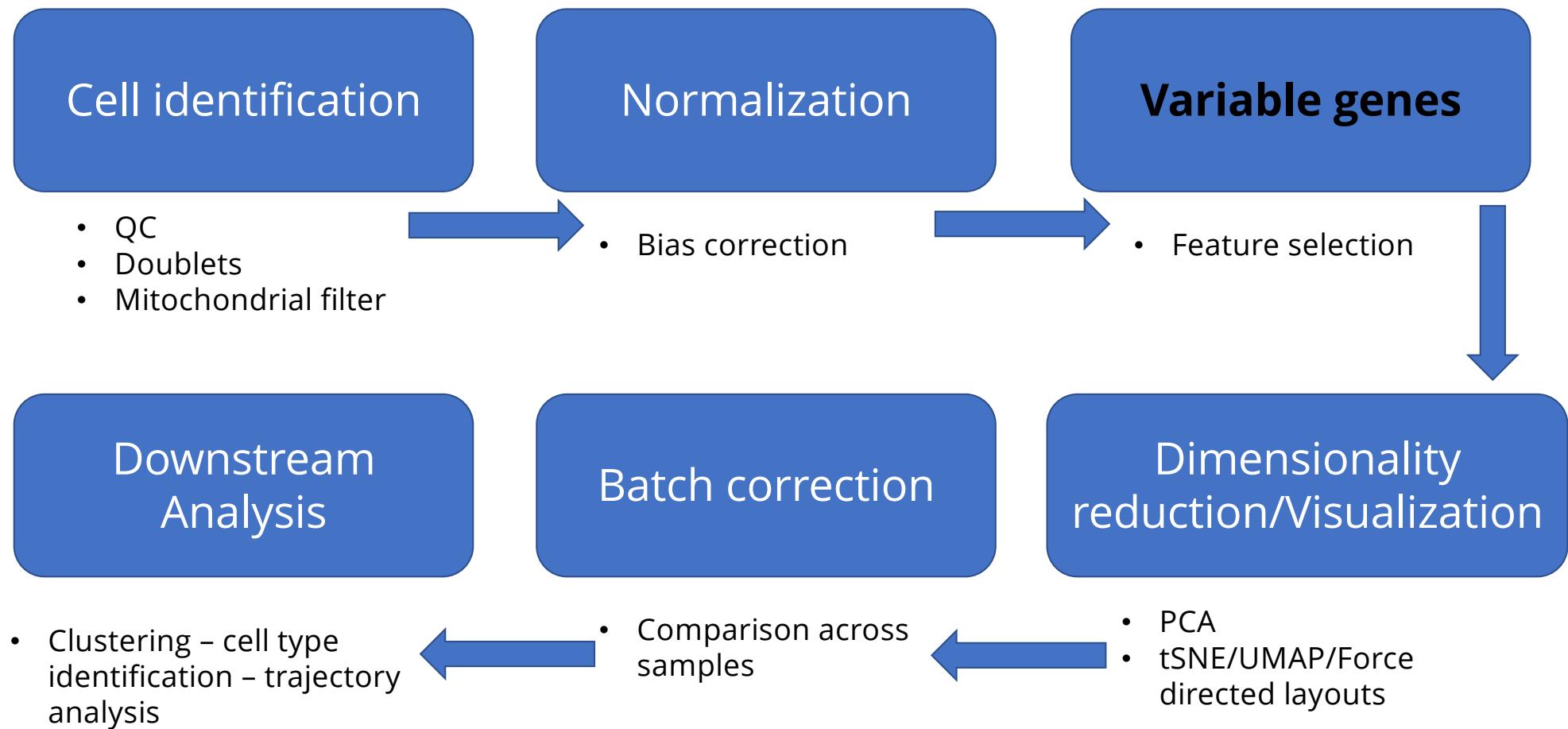


Normalization

- Global normalization:
 - Divide counts by total molecules in each cell
 - Multiply by median [To avoid numerical issues]
- Log transform of the data

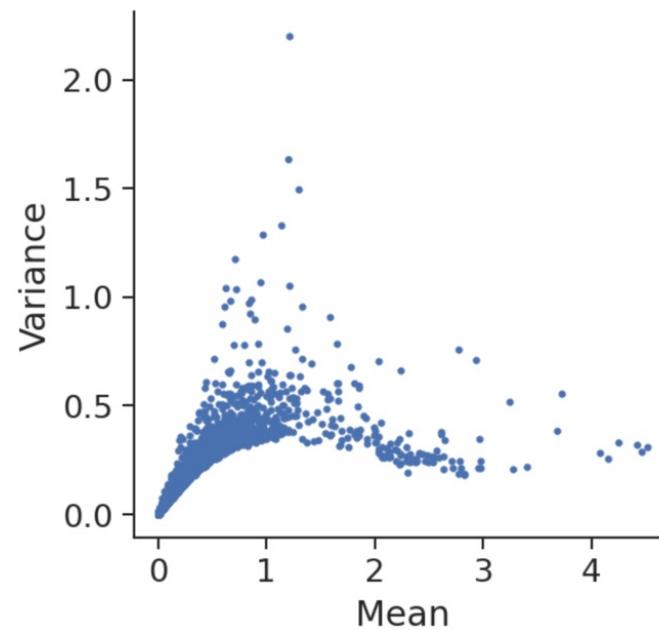


scRNA-seq analysis steps



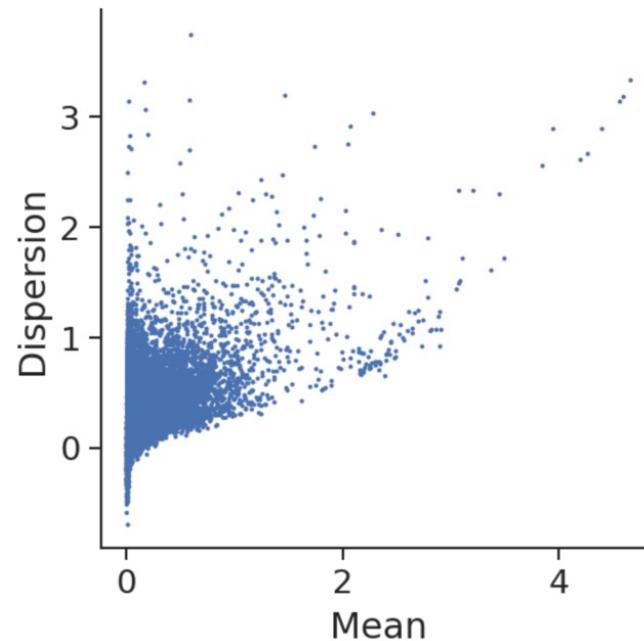
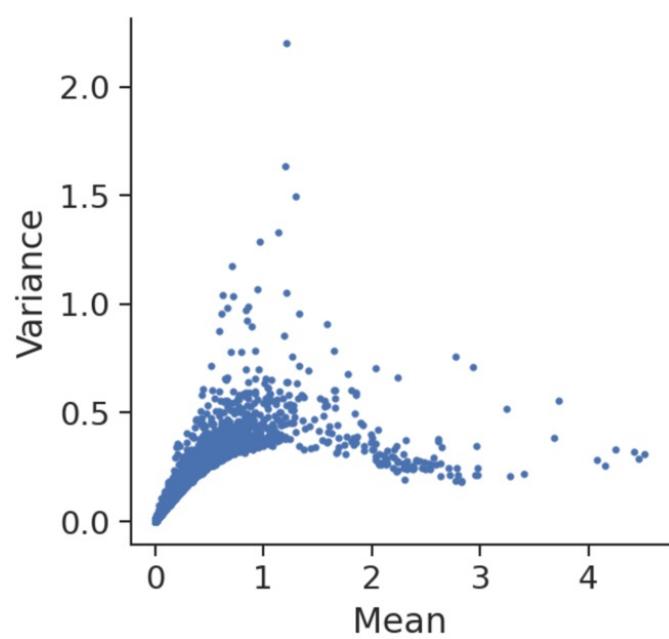
Feature selection

- Cell states are defined by expression of subsets of genes
- Goal of feature selection: Select genes that inform the biology rather than genes that represent random noise
- Possible Solution: How variable is the gene across cells?



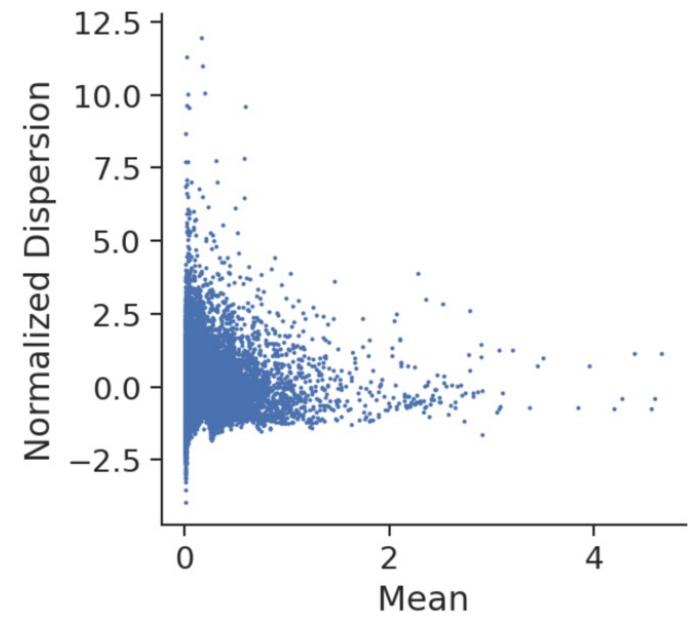
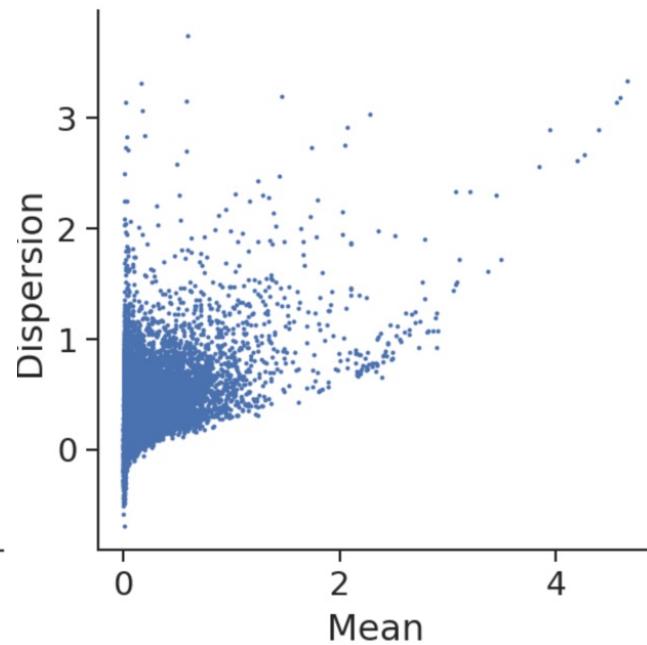
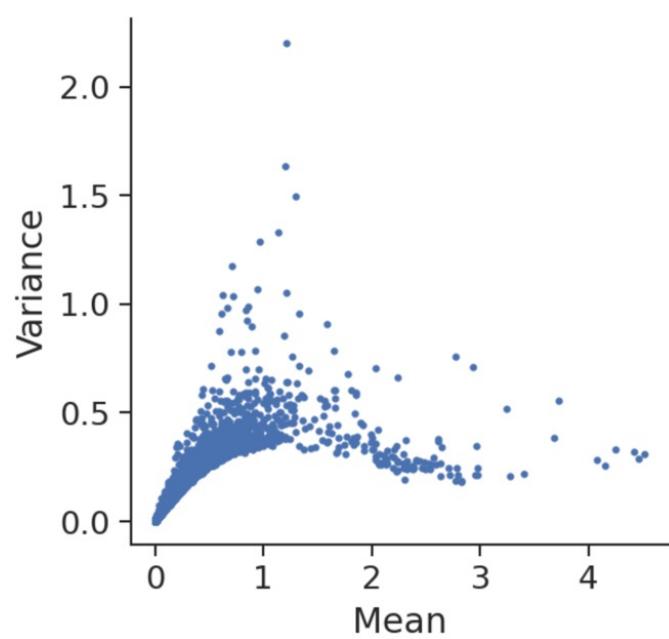
Feature selection

$$Dispersion = \frac{Variance}{Mean}$$



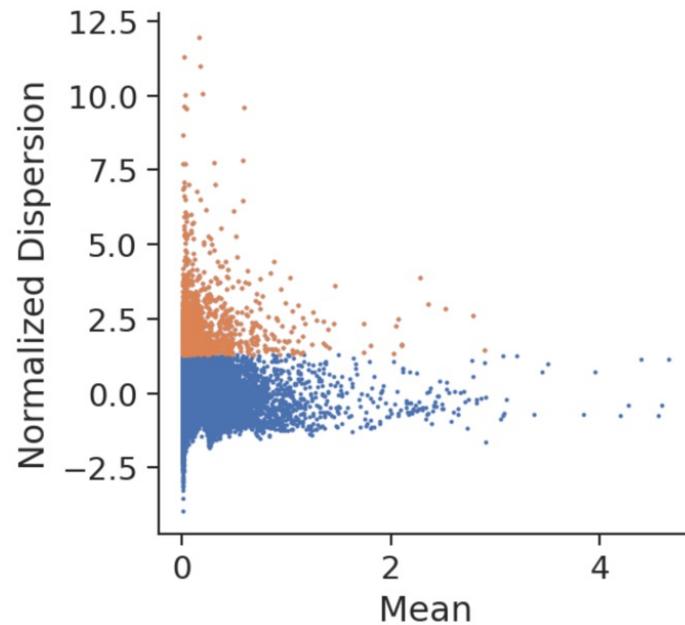
Feature selection: Share information across genes

$$\begin{aligned} \text{Normalized Dispersion} \\ = \frac{(\text{Dispersion} - \text{Mean}(\text{Dispersion in expression mean bin}))}{\text{Std}(\text{Dispersion in expression mean bin})} \end{aligned}$$



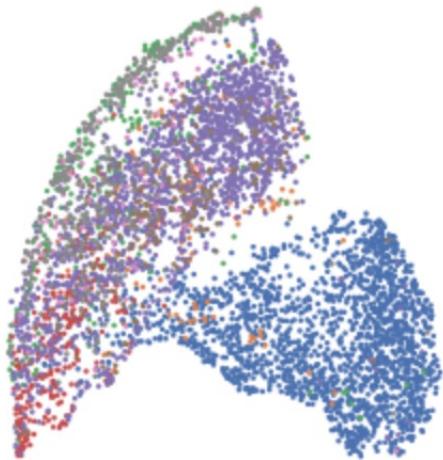
Feature selection: Share information across genes

$$\begin{aligned} \text{Normalized Dispersion} \\ = \frac{(\text{Dispersion} - \text{Mean}(\text{Dispersion in expression mean bin}))}{\text{Std}(\text{Dispersion in expression mean bin})} \end{aligned}$$

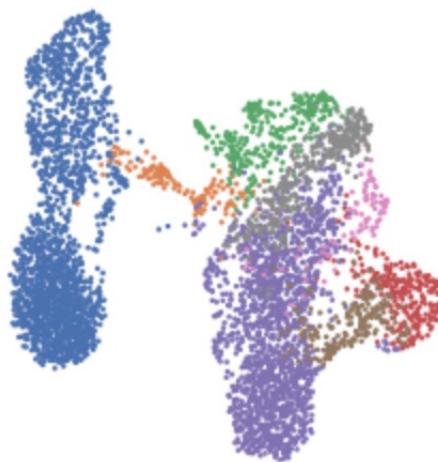


Feature selection: highly variable genes

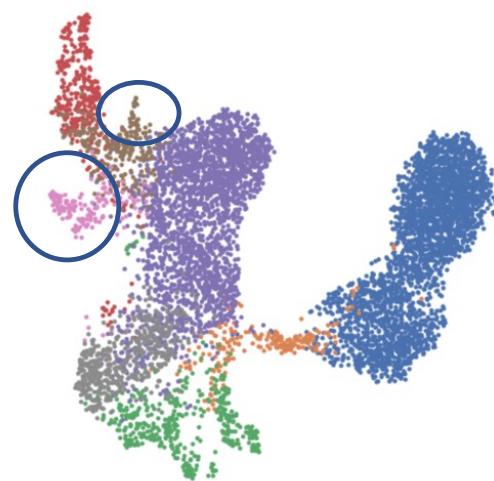
No normalization



Post normalization

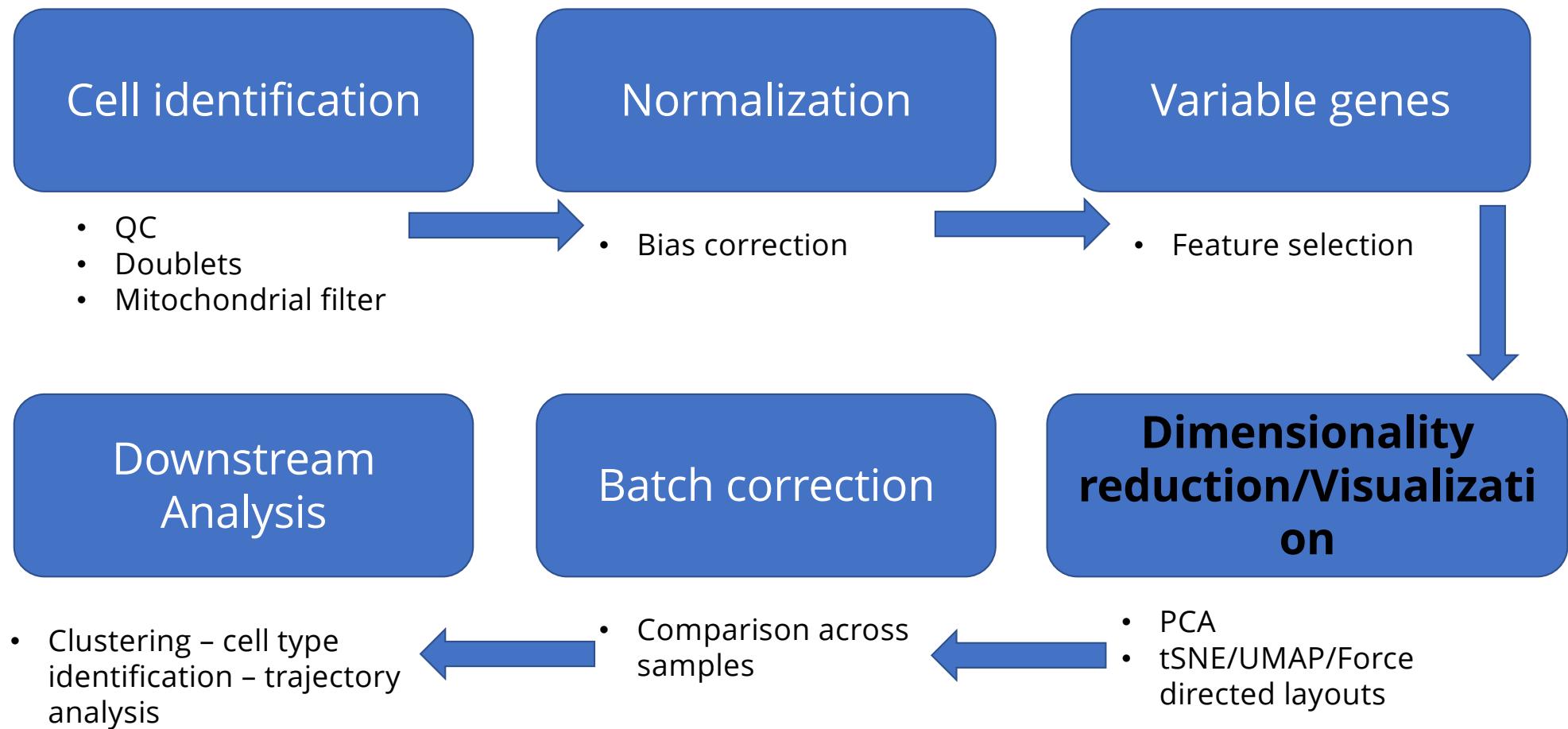


Feature selection



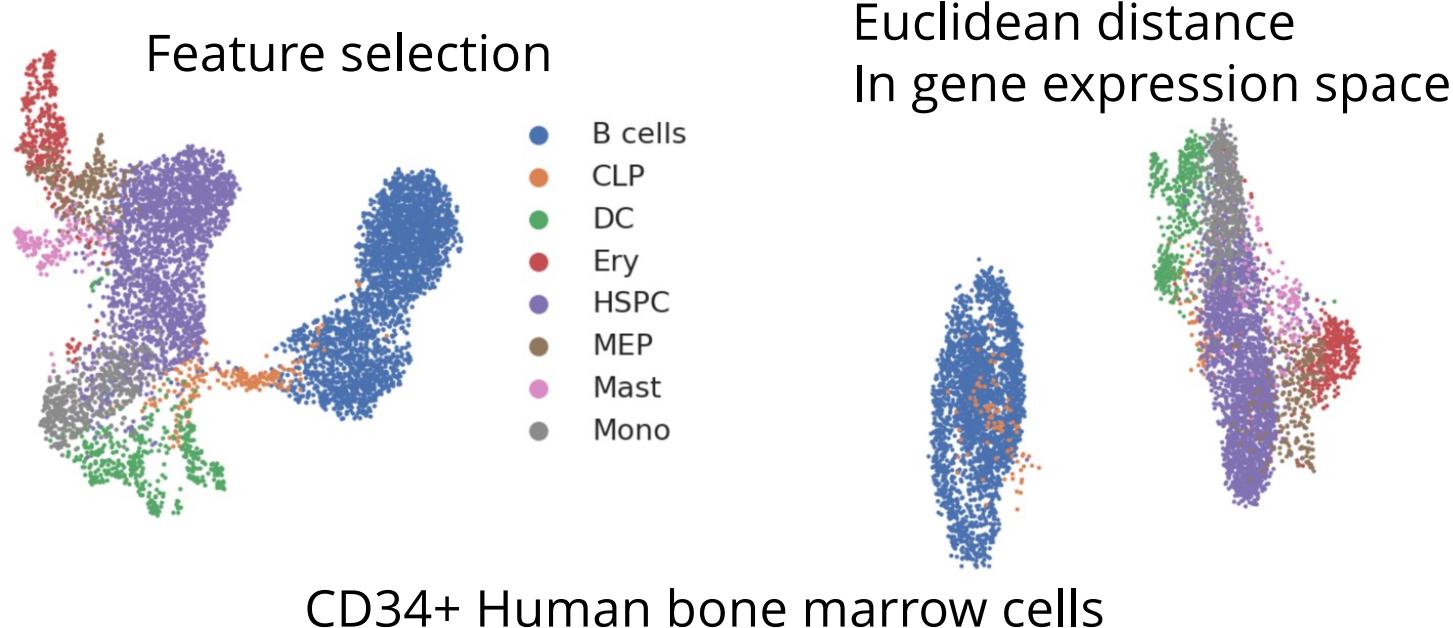
CD34+ Human bone marrow cells

scRNA-seq analysis steps



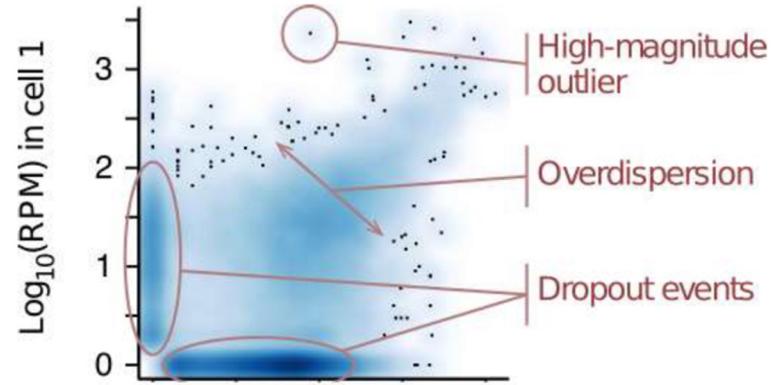
Measuring distance between cells

- Possible Solution: Euclidean distance between normalized, selected genes



Single-cell RNA Noise: Dropouts

- ~5-10% of transcripts in a cell are captured
- Further loss during reverse transcription
- Genes with higher expression have fewer zeros
- Non-zero values are also under-estimates of true counts



BRIEF COMMUNICATIONS

Bayesian approach to
single-cell differential
expression analysis



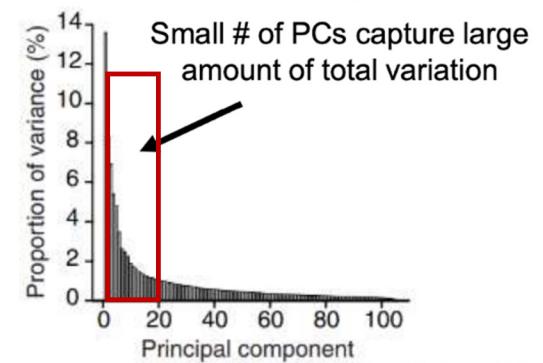
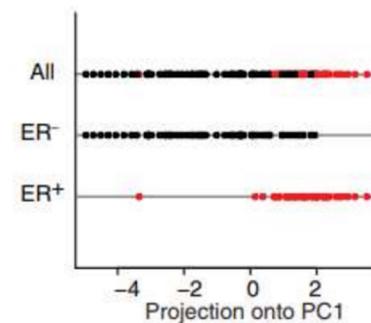
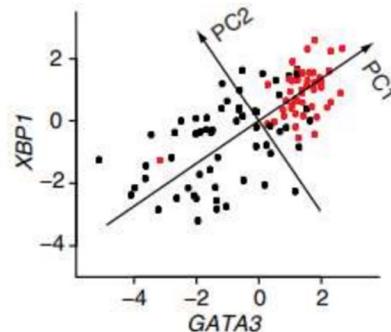
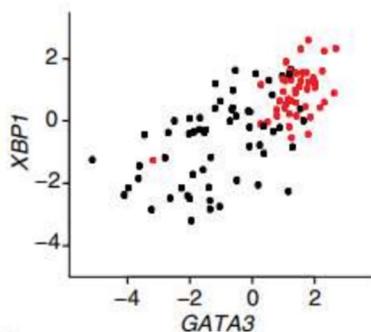
Peter V Kharchenko¹⁻³, Lev Silberstein³⁻⁵ &
David T Scadden³⁻⁵

© 2014 Nature America, Inc.

Alternative solution: PCA

- Cell states are defined by co-regulated gene modules
- PCA as a proxy to identify these genes modules

PCA effectively defines new axes through the data that capture the highest amount of variation possible



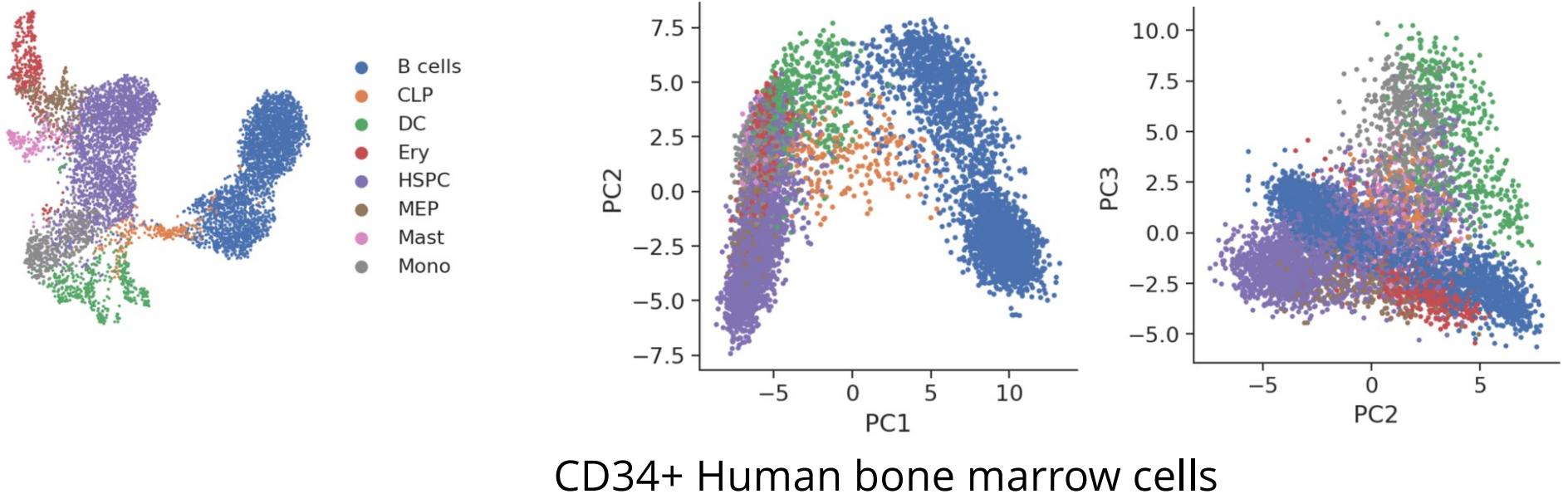
Ringér, *Nature Biotech*, 2008

- Selection of subset of PCs: Dimensionality reduction

Dimensionality Reduction

- Reduce the number of dimensions of data while preserving information
- Overcome noise
- Computational efficiency
- Visualization

PCA for visualization

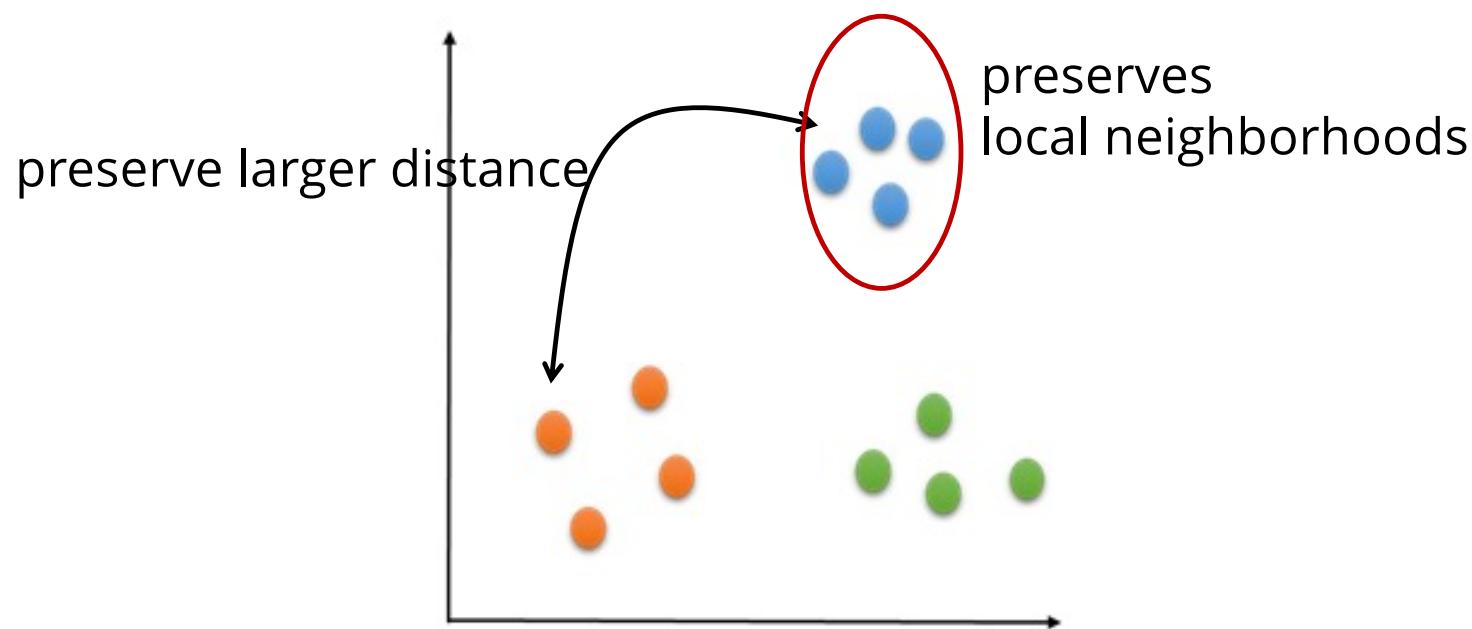


CD34+ Human bone marrow cells

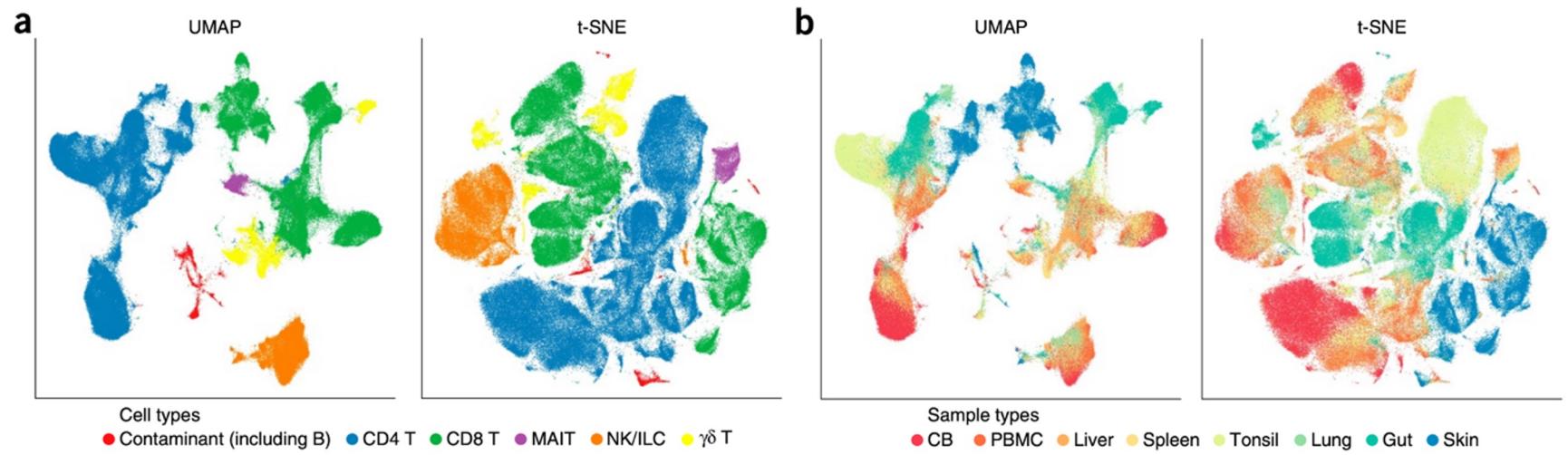
- Two axes can only capture so much information
 - Not explicitly modeled to capture as much of higher order information in 2D

Dimensionality Reduction for Visualization (tSNE, UMAP, and other force directed layouts)

- Compute a low dimensional representation that best preserves the local neighborhoods of cells

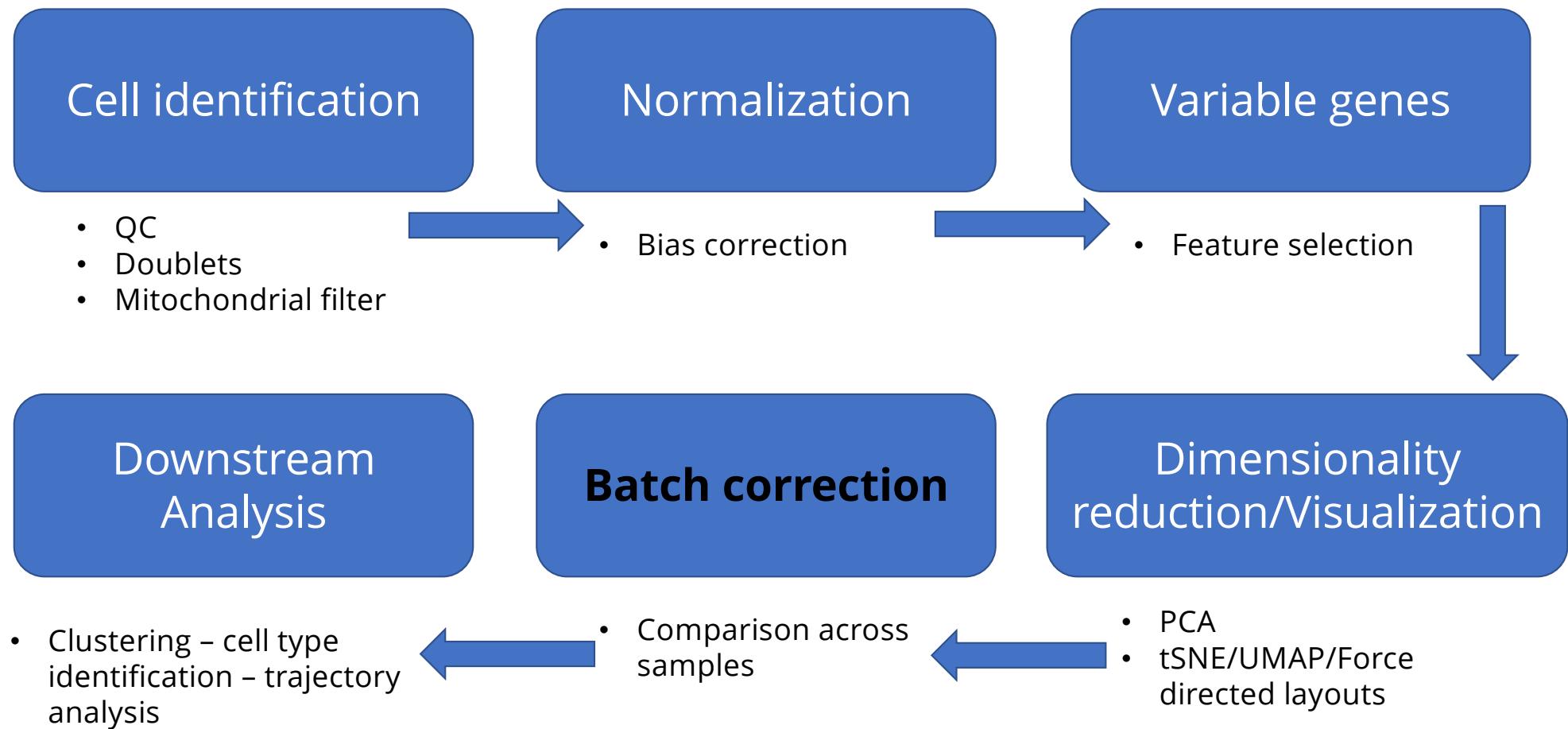


Visualization



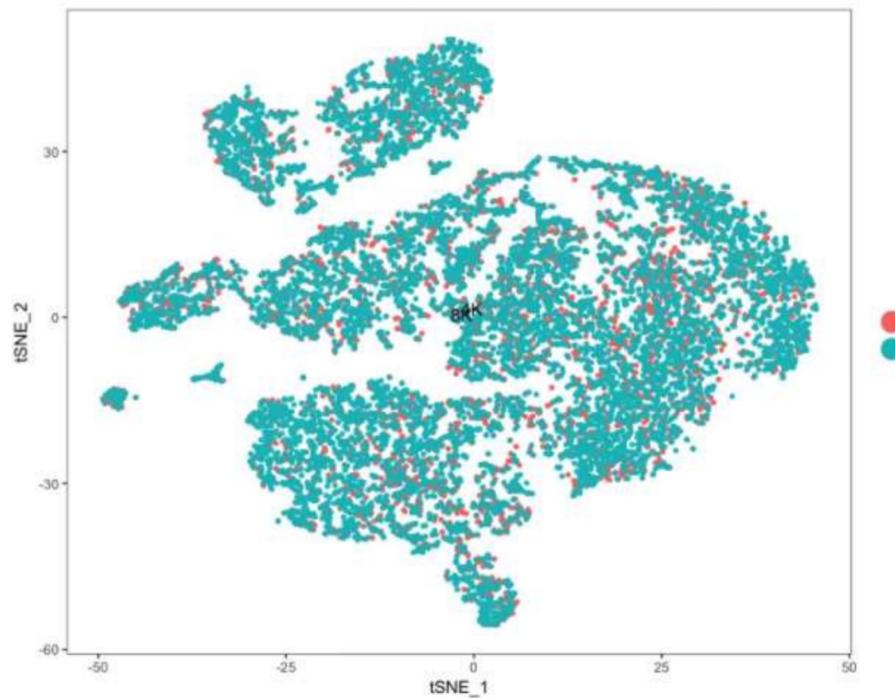
Becht et. al., 2019

scRNA-seq analysis steps

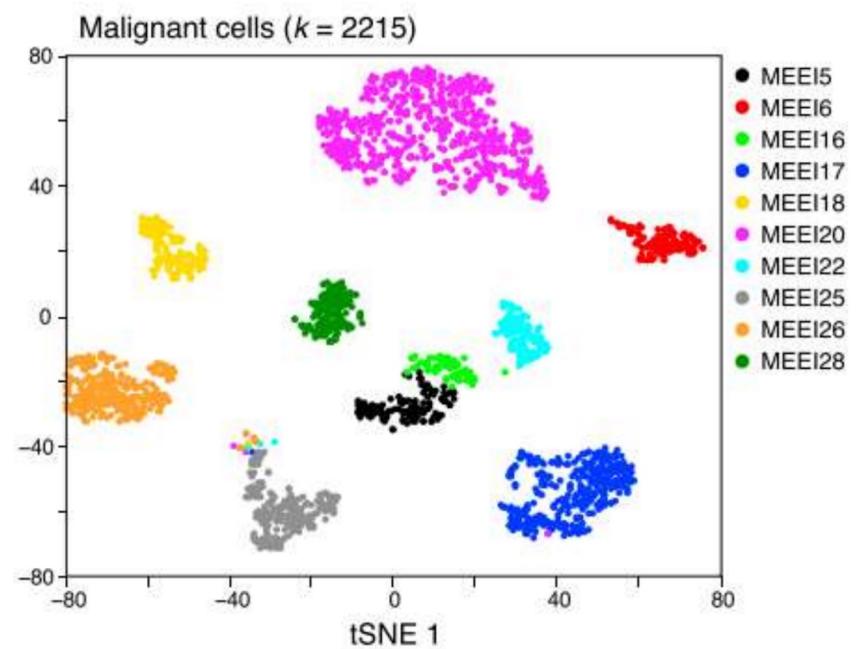


Real heterogeneity vs. technical noise

**Technical replicate of PBMCs
has near-perfect overlap**

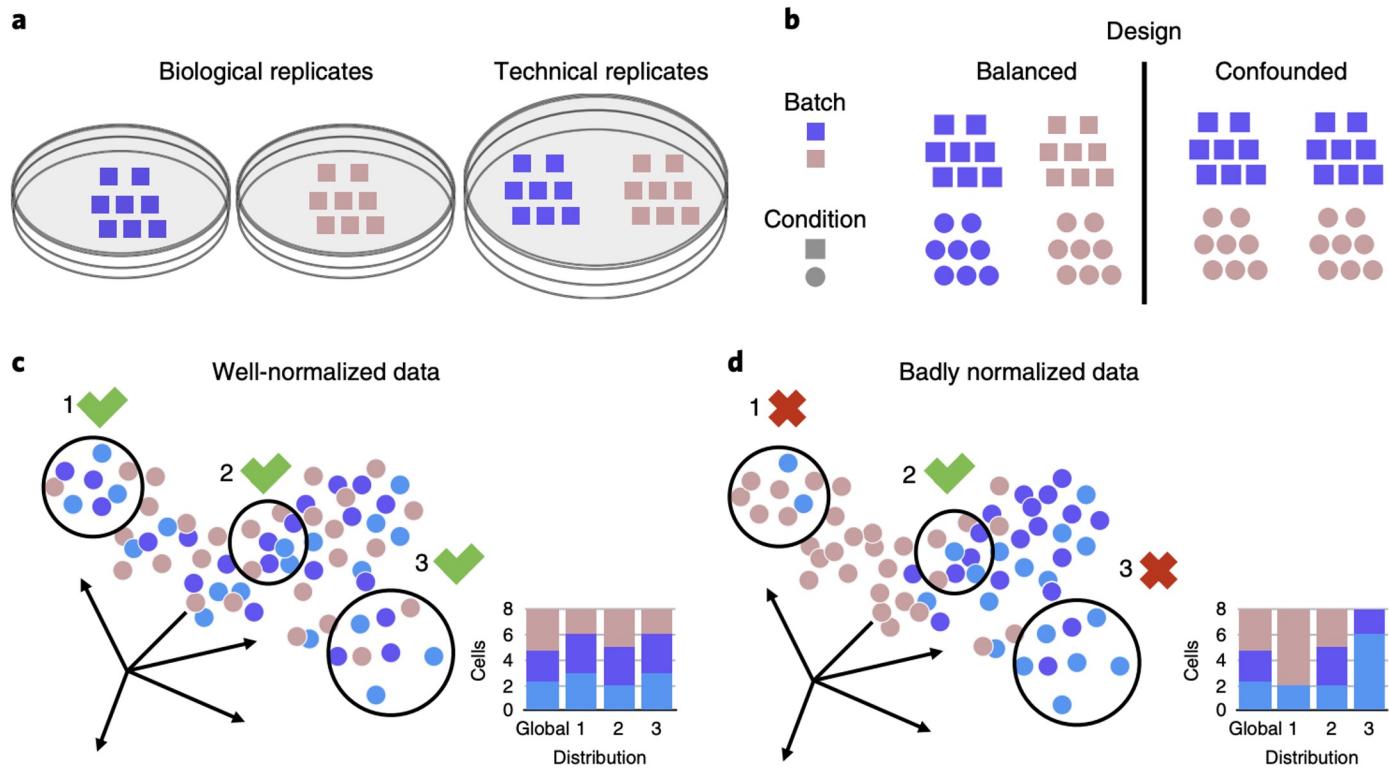


**Cancer cells dramatically
different between patients**



Assessing normalization and batch effects

Evaluate mixing of samples



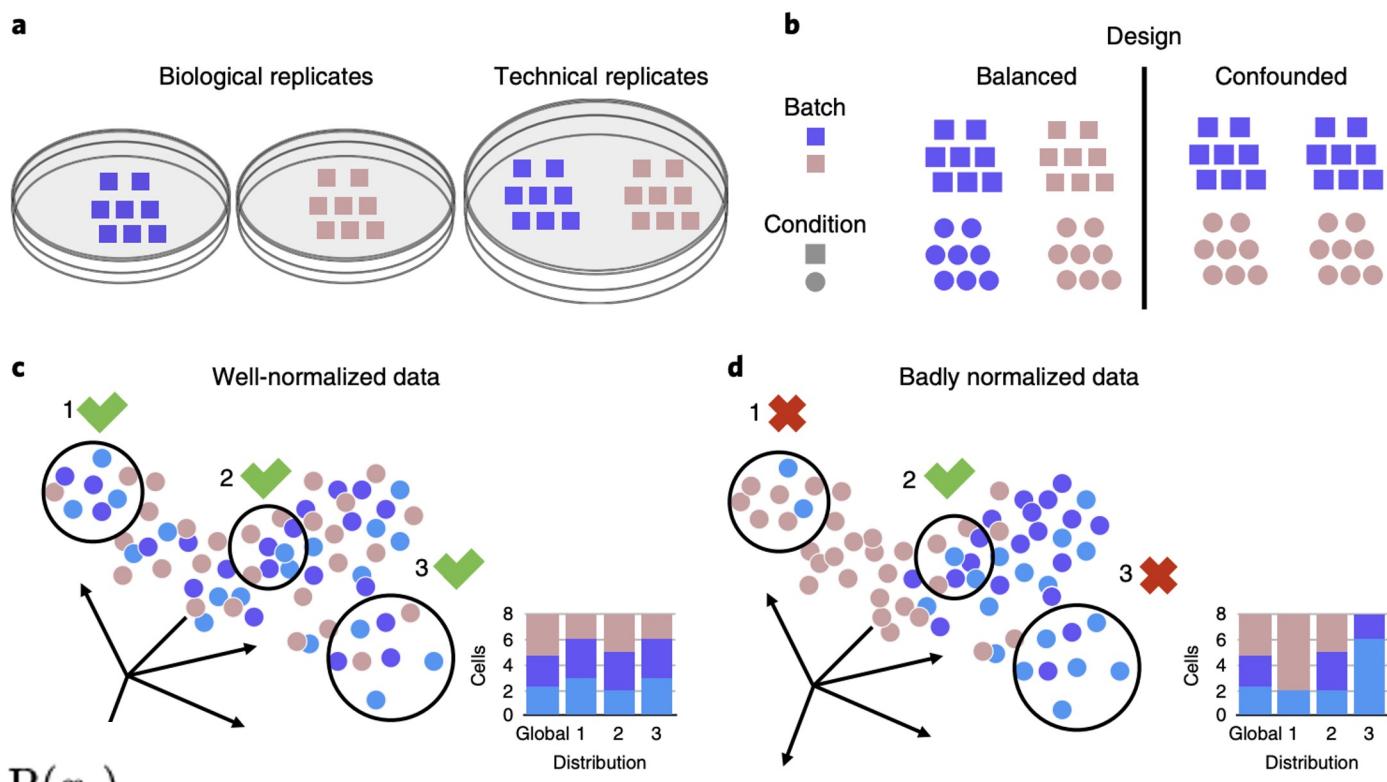
Buttner et. al, 2018

Assessing normalization and batch effects

Select random neighborhoods of fixed size

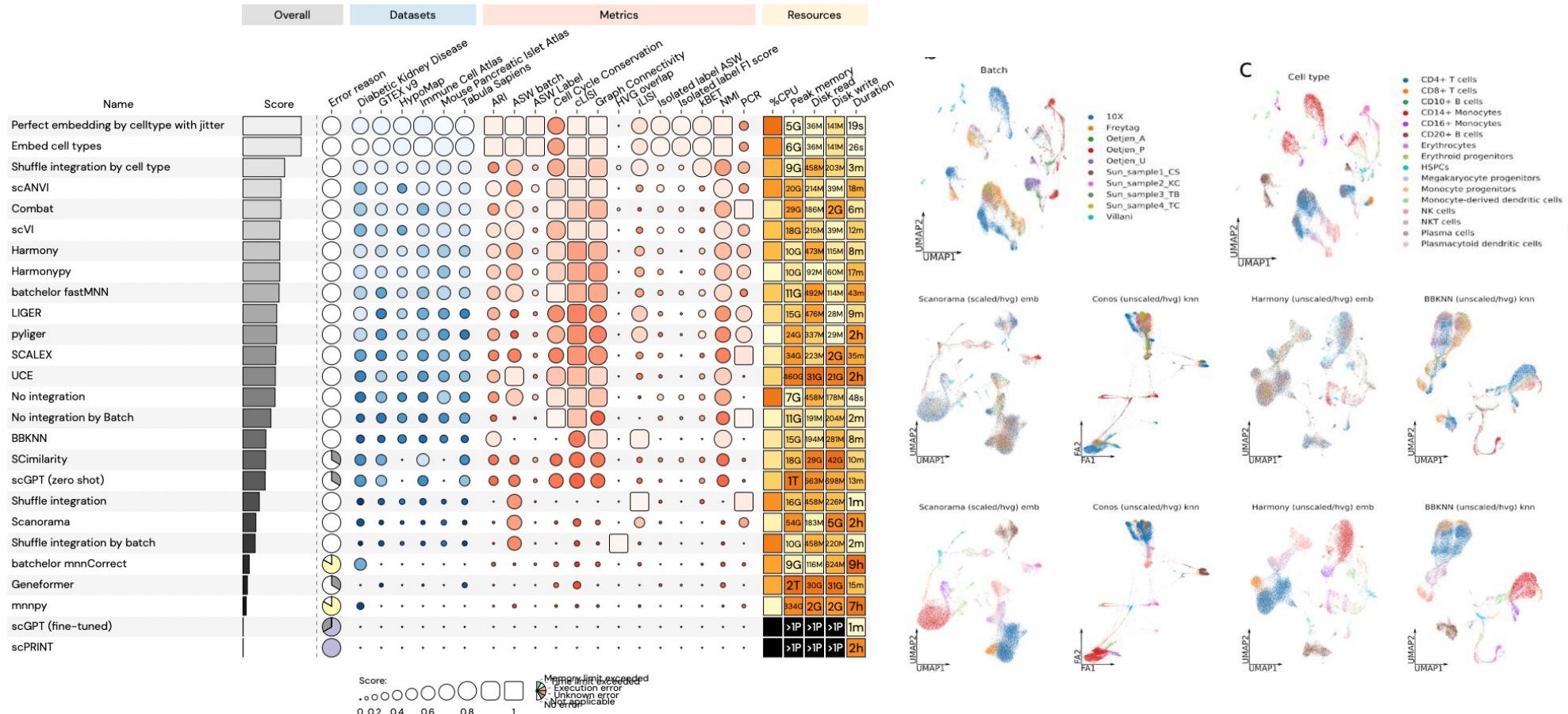
Compute Shannon Entropy of distribution across samples

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i)$$



Buttner et. al, 2018

Batch effect correction approaches

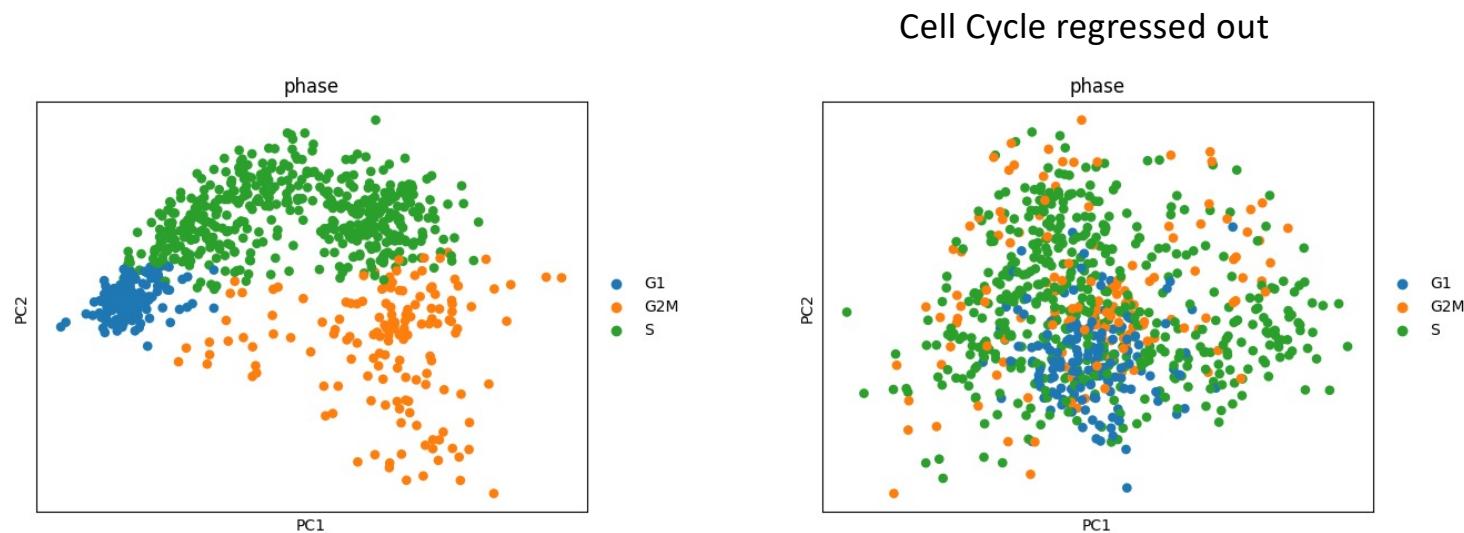


<https://openproblems.bio/> (2025)

Luecken et. al, 2020

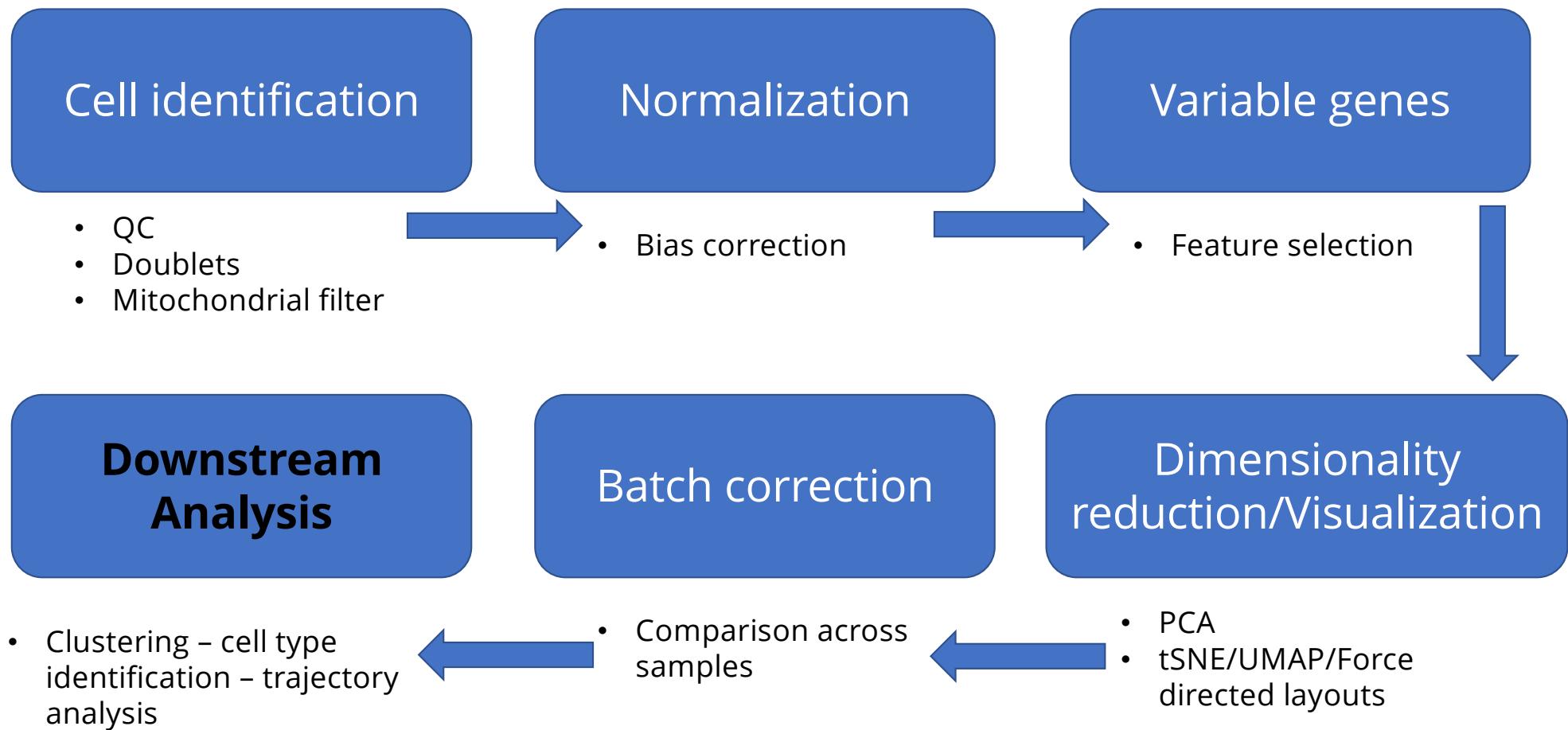
Other corrections

- Cell cycle correction
- Gene expression imputation



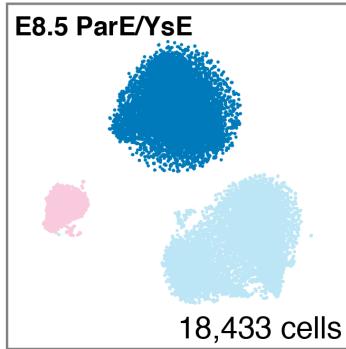
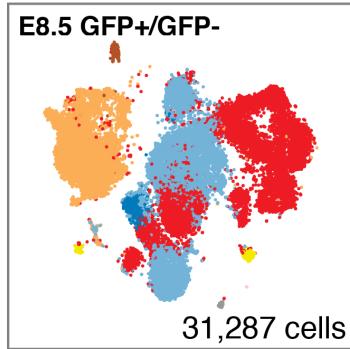
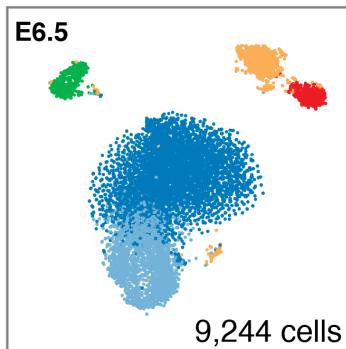
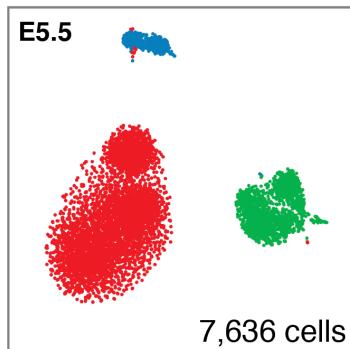
<https://scanpy.readthedocs.io/>

scRNA-seq analysis steps

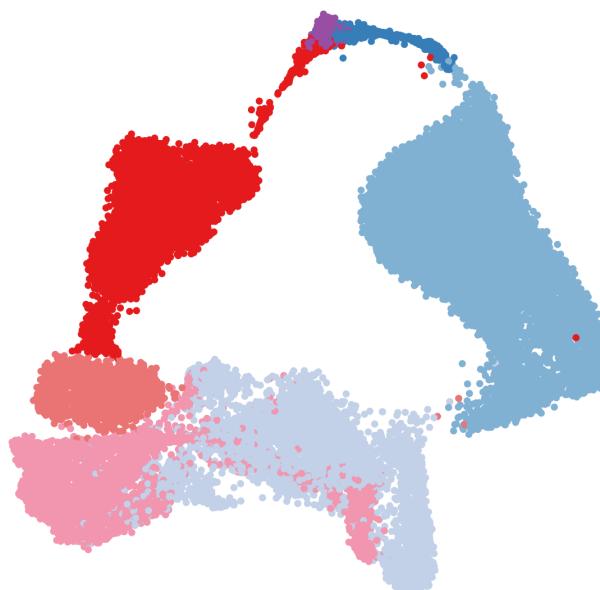


Single-cell data

Discrete clusters



Continuous trajectories



ICM
PrE
VE
Gut tube: VE
EPI
DE
Gut tube: DE

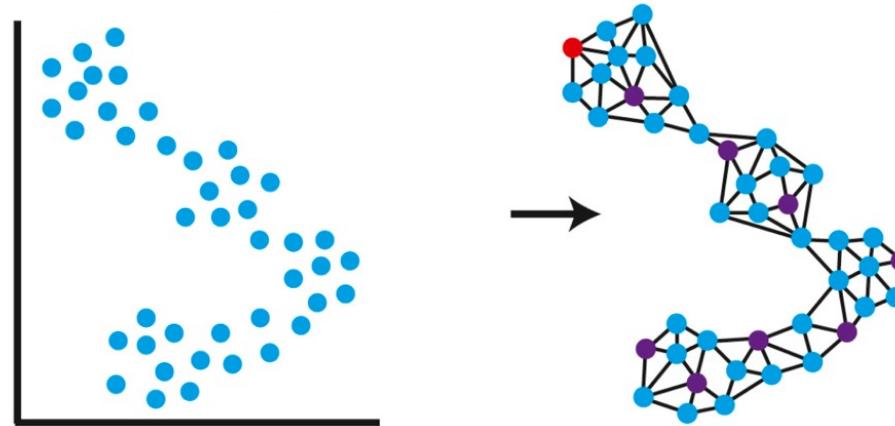
Nearest neighbor graphs

- For each point, find k nearest points using Euclidean distance

$Graph G = (V, E)$

V : Set of vertices

E : Set of edges



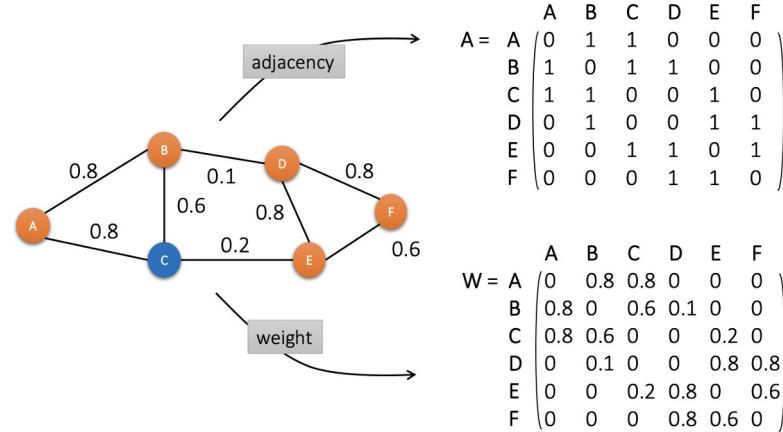
Graph adjacency matrices

- Graph can also be represented as an adjacency matrix o

Graph $G = (V, E)$

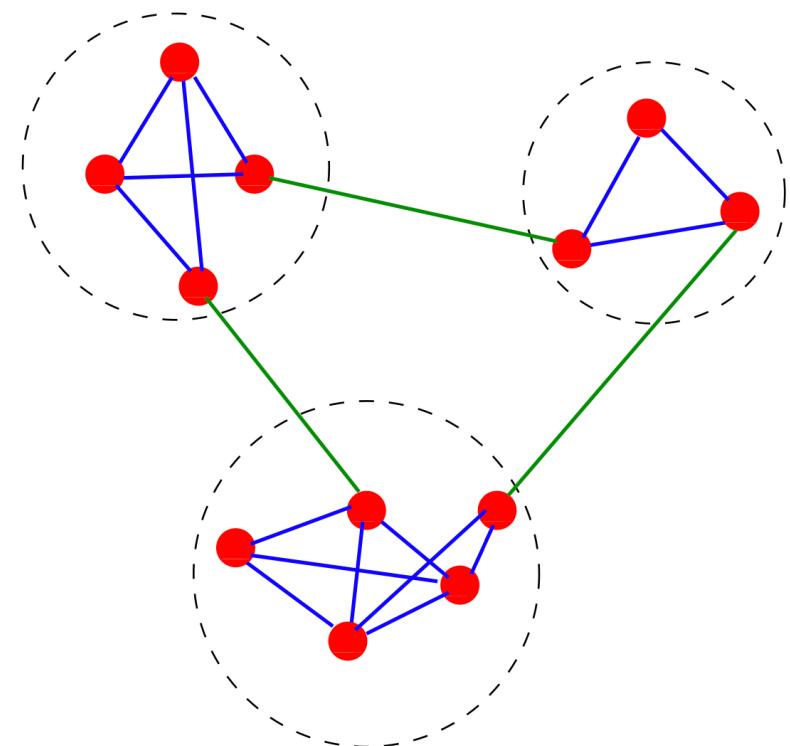
Adjacency matrix $A \in \mathbb{R}^{|V| \times |V|}$

$A_{ij} = E(i, j)$



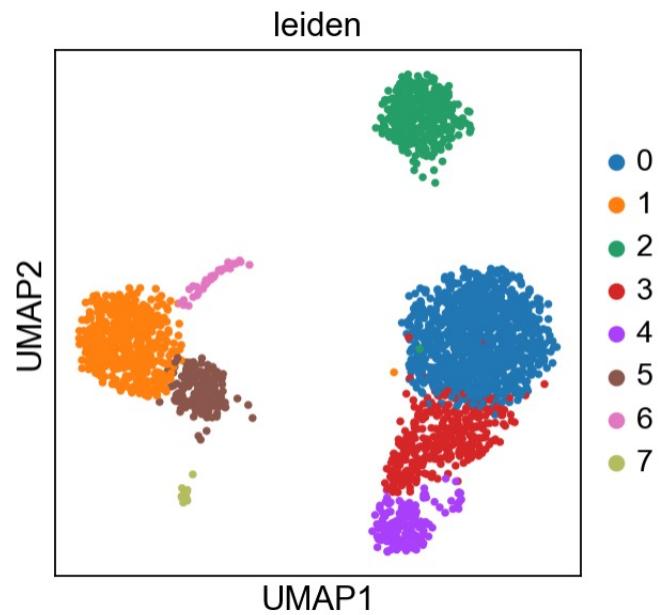
Graph based clustering

- Community detection
- Identify clusters of nodes or “communities” with high density of edges within and low density of edges across communities



Fortunato 2009

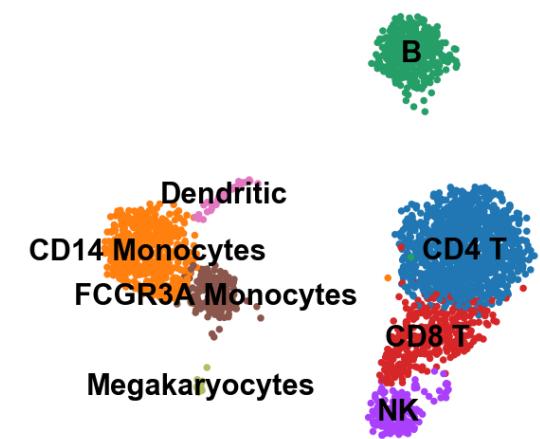
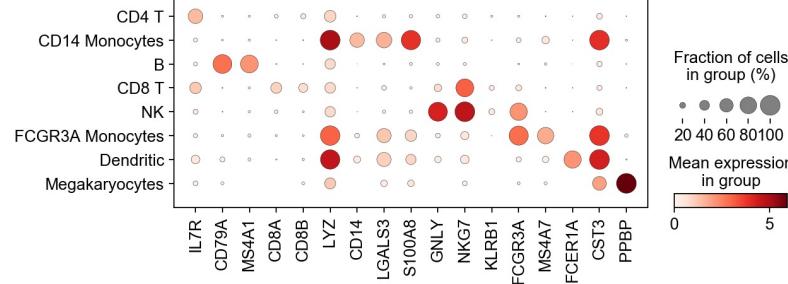
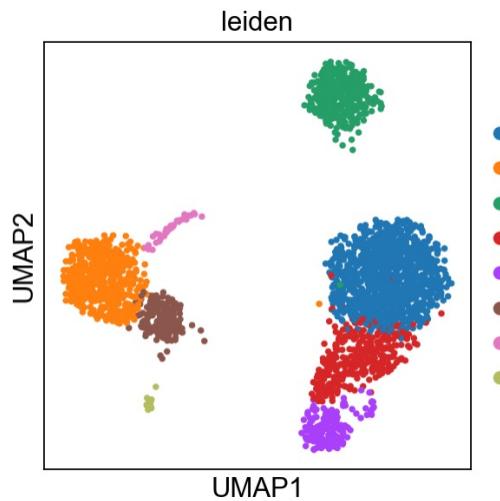
Leiden clustering in single-cell data



Butler et. al. 2018

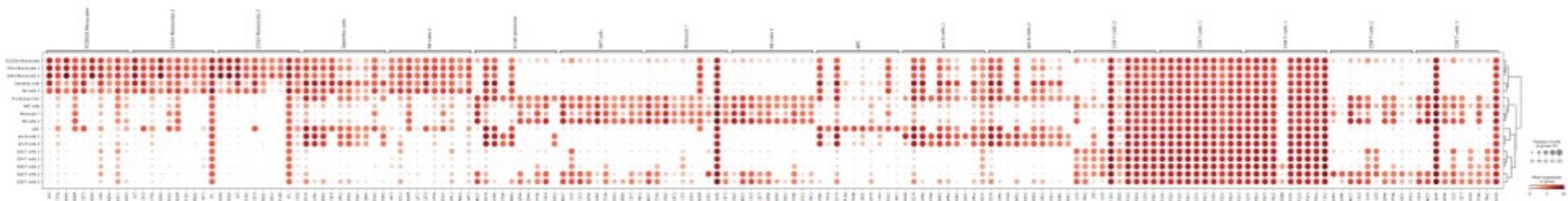
Cell type identification

- Marker based identification



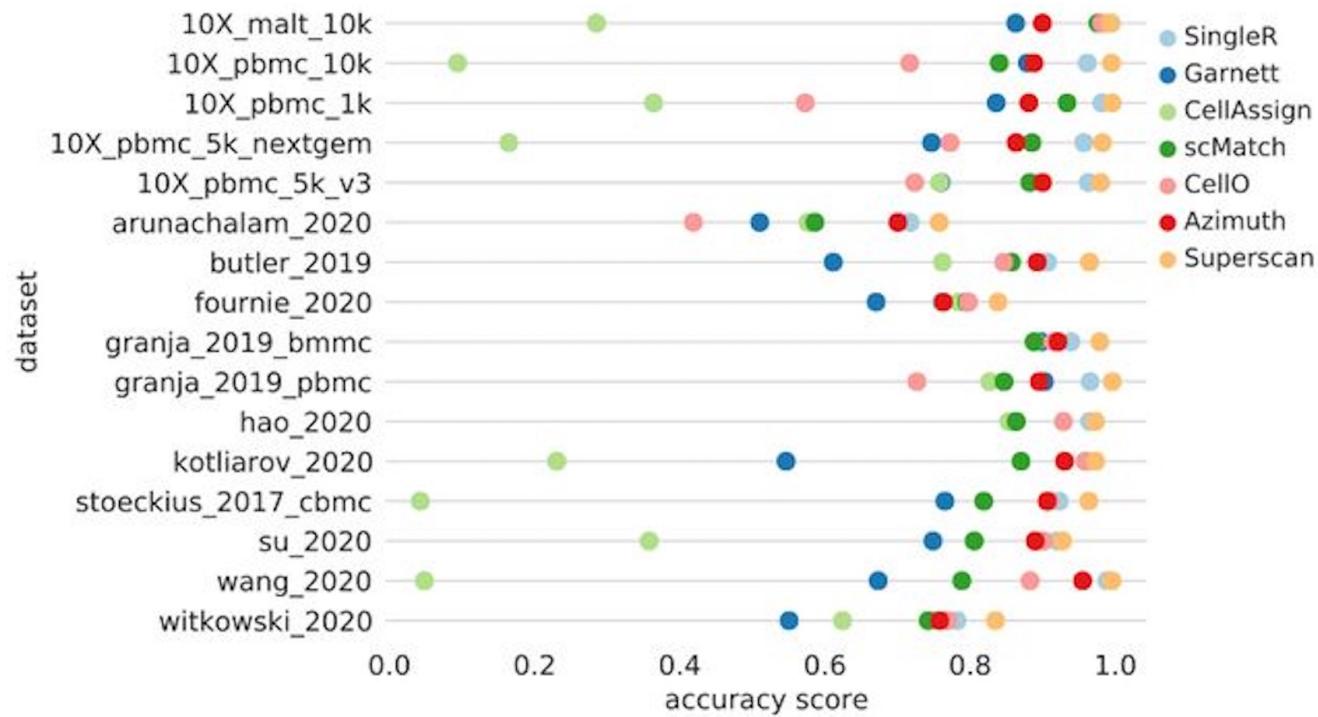
Cell type identification

- Differential expression



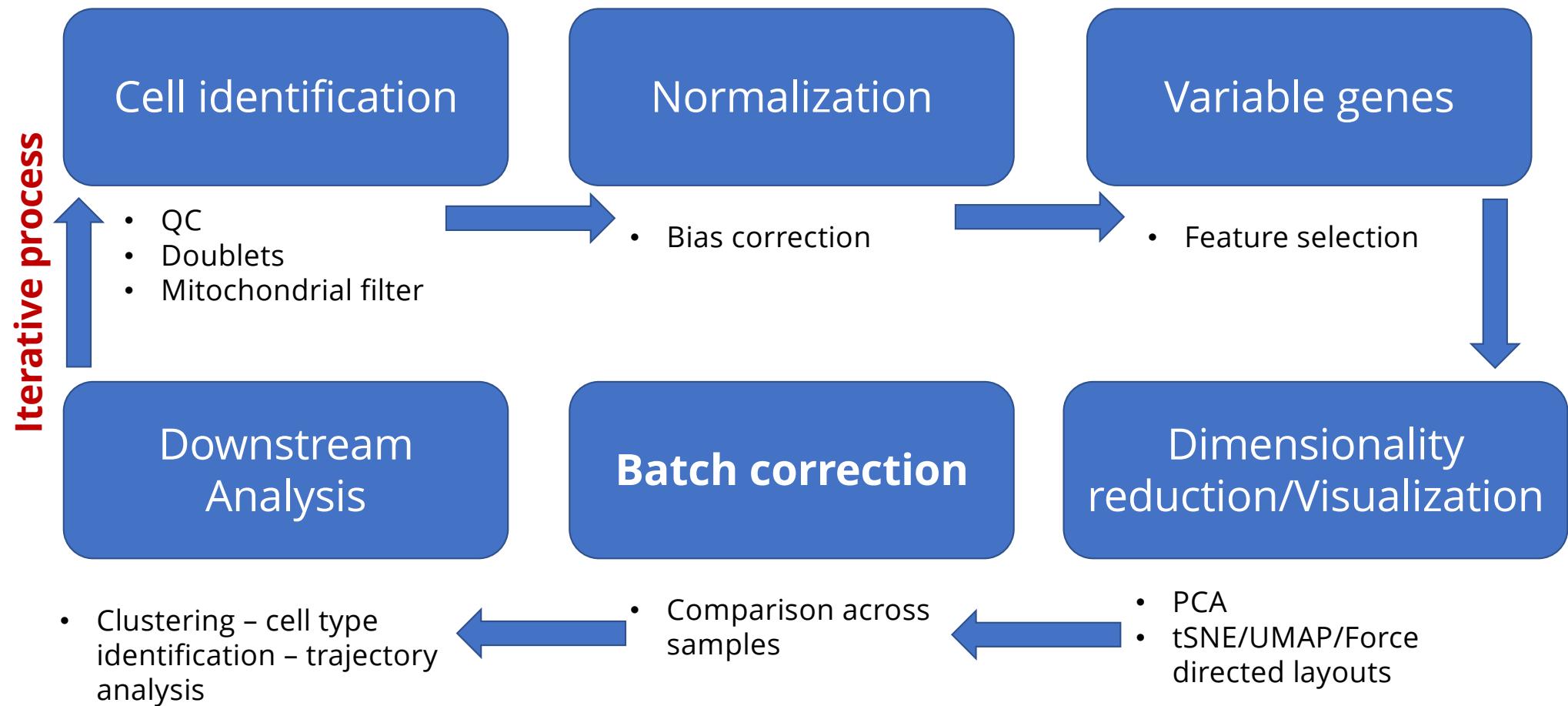
Cell type identification

- Supervised approaches: Train on manually labeled cells
- Superscan



Shasha et. al. 2021

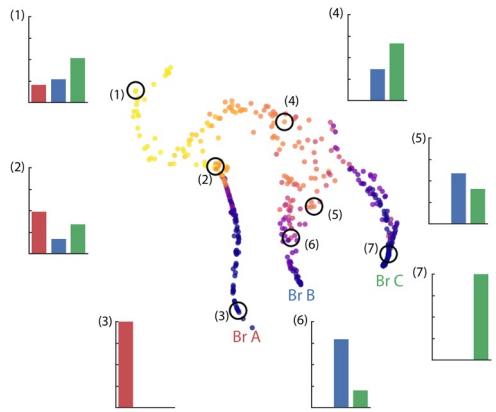
scRNA-seq analysis steps



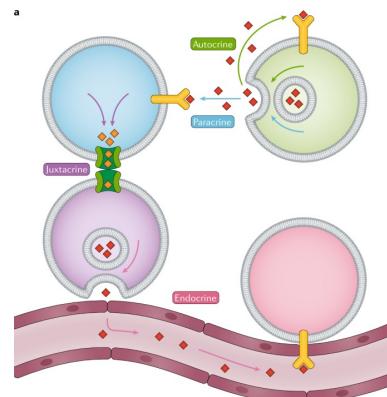
Interpreting single-cell data

Interpreting single-cell data

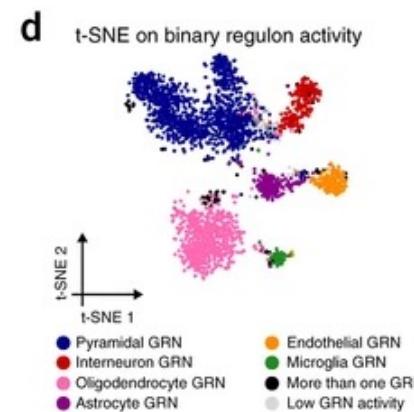
Trajectory analysis



Cell communication



Regulatory Networks

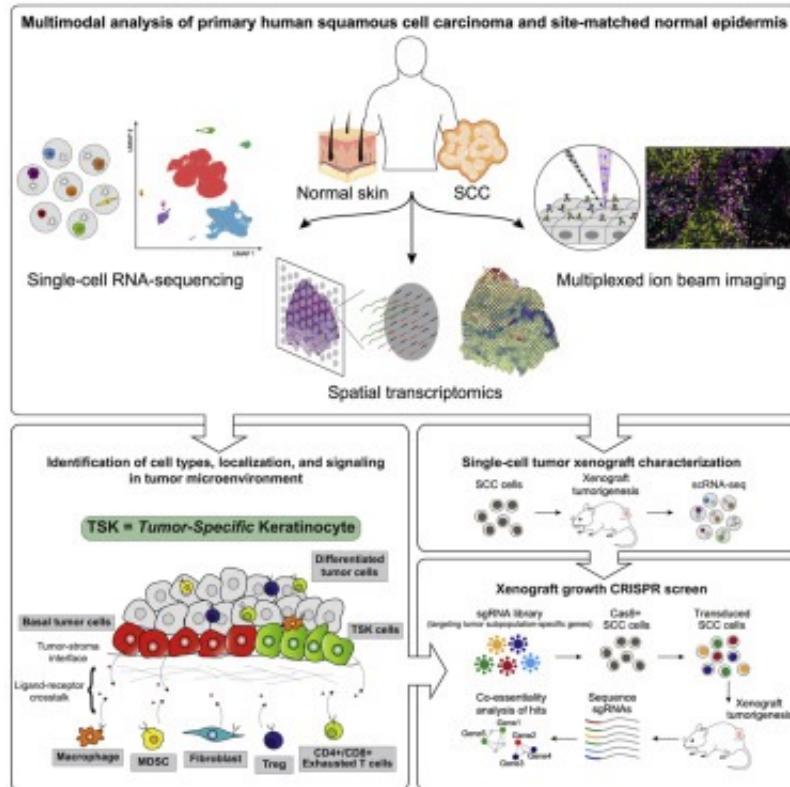


Where to find data?

Major data resources – Healthy/Normal

- Human Cell Atlas
 - Fetal Cell Atlas
- Tabular Muris

Major Data resources - GEO



Supplementary file	Size	Download	File type/resource
GSE144236_CAL27_counts.txt.gz	20.9 Mb	(ftp)(http)	TXT
GSE144236_CAL27_vitro_counts.txt.gz	20.0 Mb	(ftp)(http)	TXT
GSE144236_SCC13_counts.txt.gz	11.2 Mb	(ftp)(http)	TXT
GSE144236_XG_TME_counts.txt.gz	3.4 Mb	(ftp)(http)	TXT
GSE144236_cSCC_counts.txt.gz	127.2 Mb	(ftp)(http)	TXT
GSE144236_patient_metadata_new.txt.gz	648.8 Kb	(ftp)(http)	TXT

SRA Run Selector [?](#)

Processed data are available on Series record

Raw data are available in SRA

Count matrices (post QC) and metadata are typically made available

Interactive browsers

Mouse endoderm atlas



endoderm-explorer.com

A large number of studies set up webapps for interacting with the data