

# An Introduction to Single-Cell Genomics

*(Actually, mostly transcriptomics)*

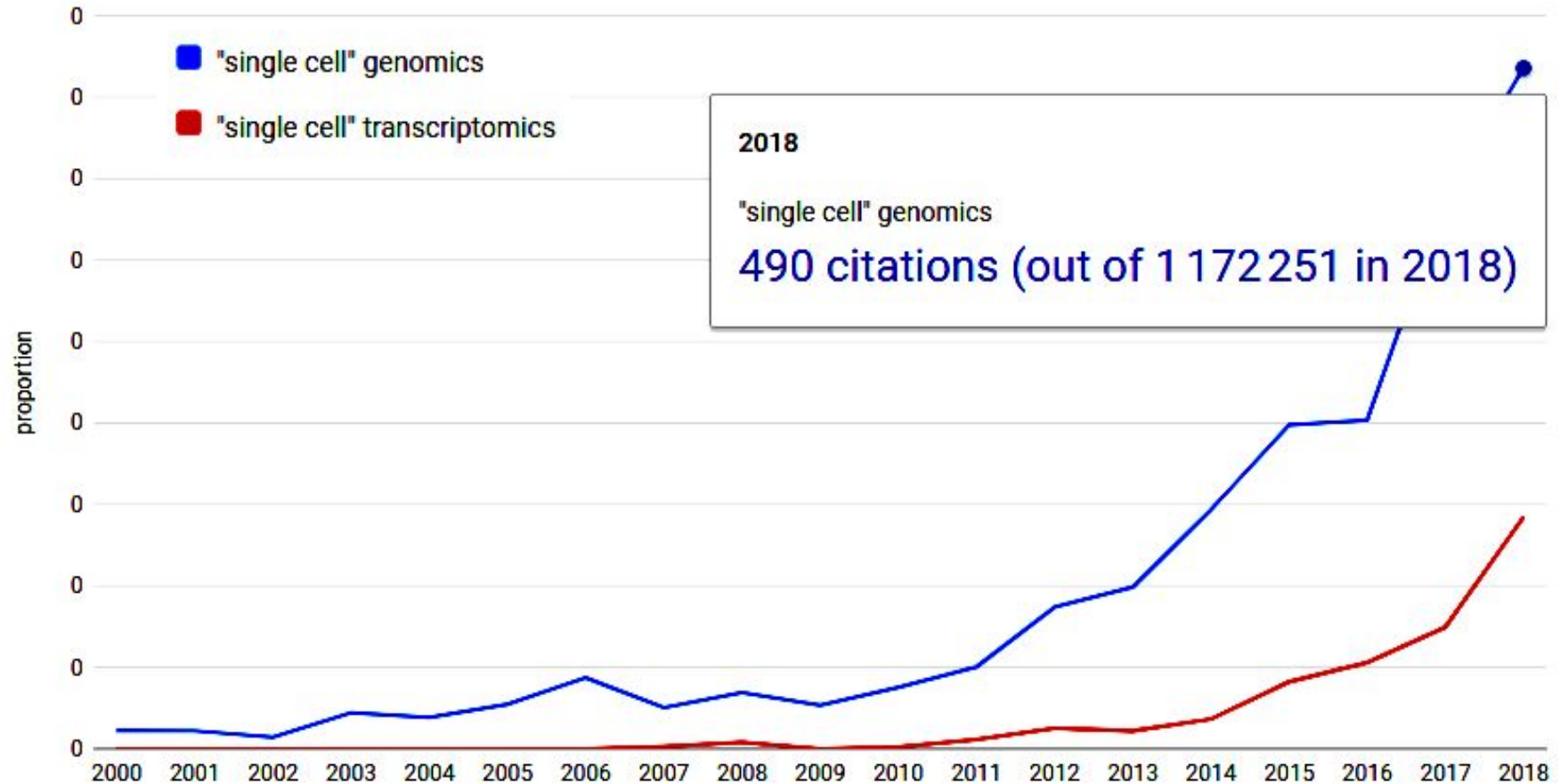
**Bastien JOB**

bastien.job@gustaveroussy.fr

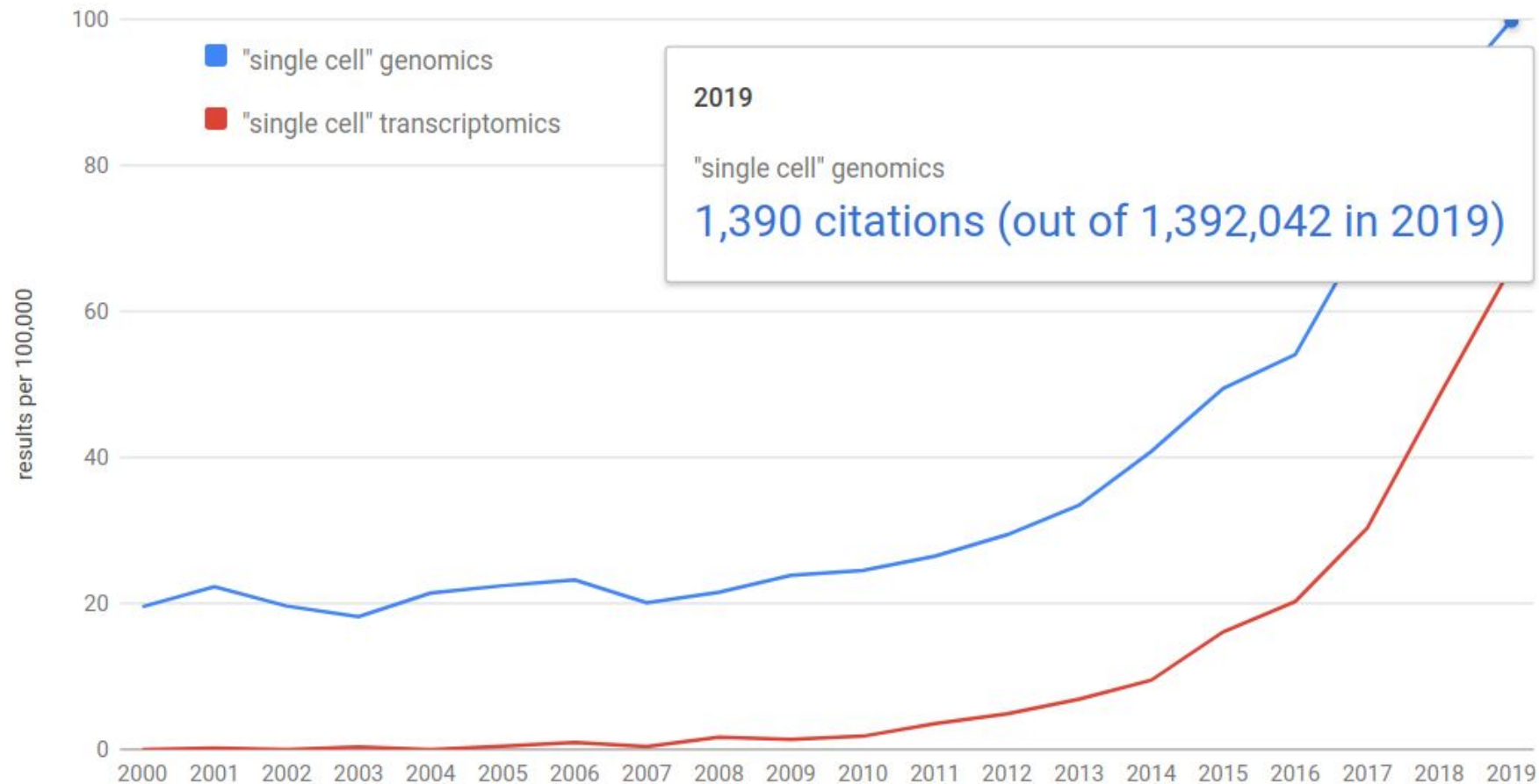
INSERM / Gustave Roussy

*So you say you've heard about single cell ?*

# Single cell in peer-reviewed publications (2018)

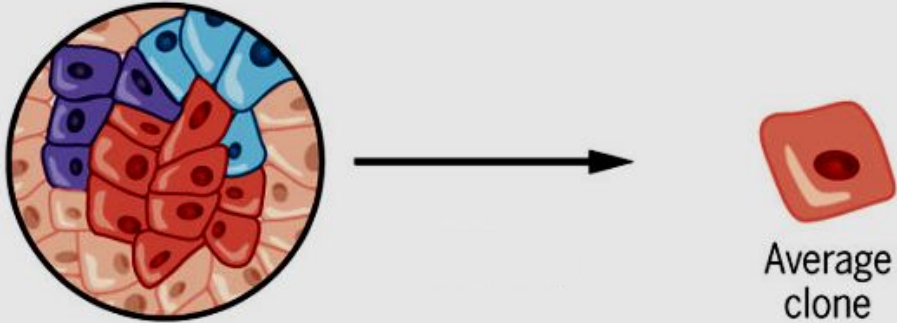


# Single cell in peer-reviewed publications (2019)

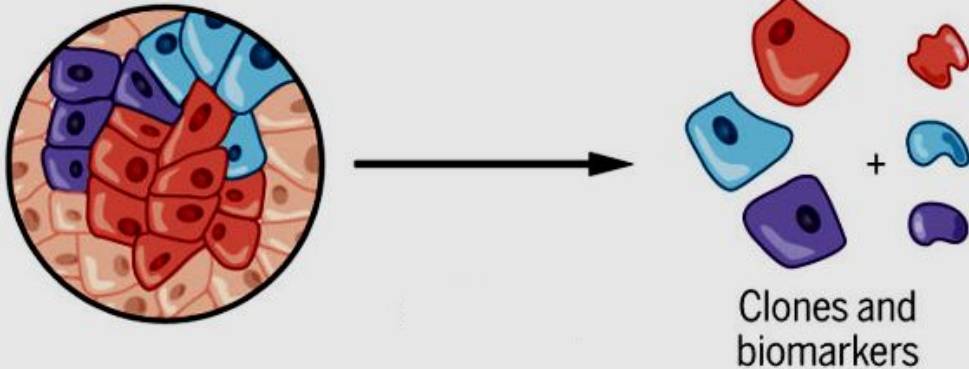


# Why so much hype ?

## A Bulk analysis



## B scRNA analysis

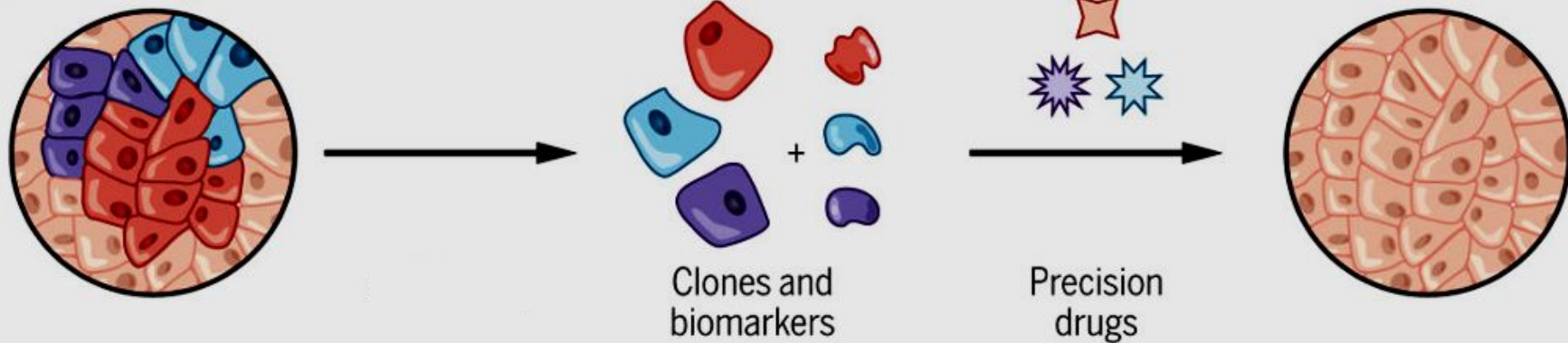


# Why so much hype ? (pathology)

## A Bulk analysis



## B scRNA analysis



# Why so much hype ?

**Bulk**



**Single cell**



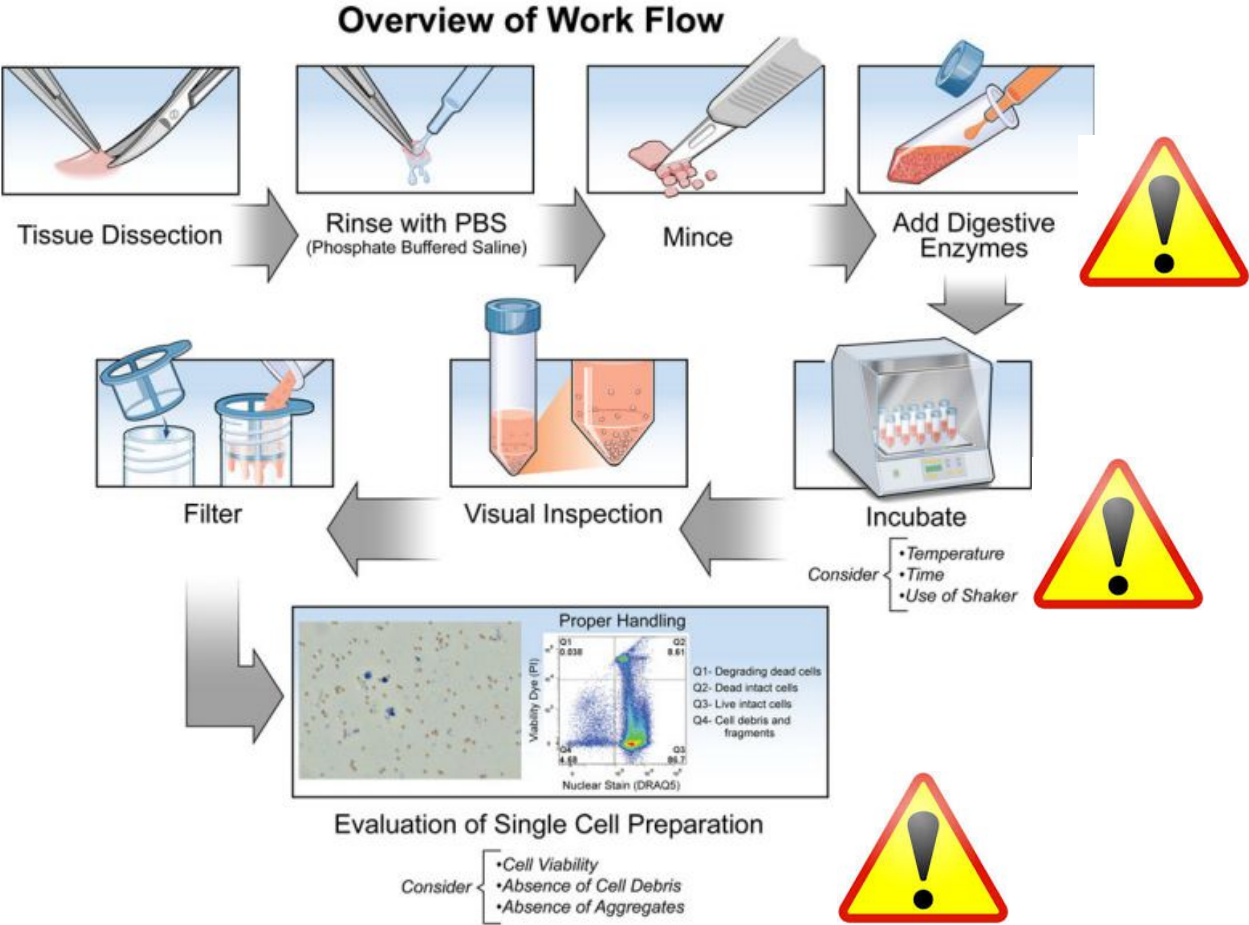
**Spatial single cell**



*From broad tissue to isolated cells*

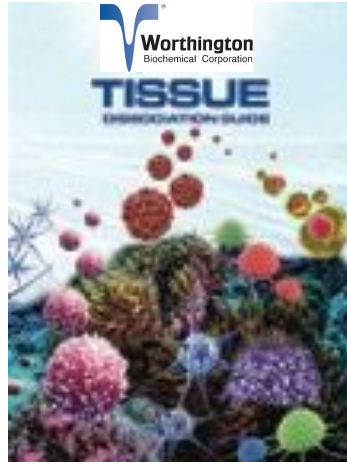


# Cells health and dissociation



# Cells health and dissociation : Worthington helpdesk

1. Type of tissue
2. Species of origin
3. Age of the animal
4. Genetic modification(s) (knockouts, etc.)
5. Dissociation medium used
6. Enzyme(s) used
7. Impurities in any crude enzyme preparation used
8. Concentration(s) of enzyme(s) used
9. Temperature
10. Incubation times



## II. Cell Isolation Theory

- Tissue Types
  - Epithelial Tissue
  - Connective Tissue
- Dissociating Enzymes
  - Collagenase
  - Trypsin
  - Elastase
  - Hyaluronidase
  - Papain
  - Chymotrypsin
  - Deoxyribonuclease I
  - Neutral Protease (Dispase)
  - Trypsin Inhibitor
  - Animal Origin Free (AOF) Enzymes
  - Celase® GMP

## III. Cell Isolation Techniques

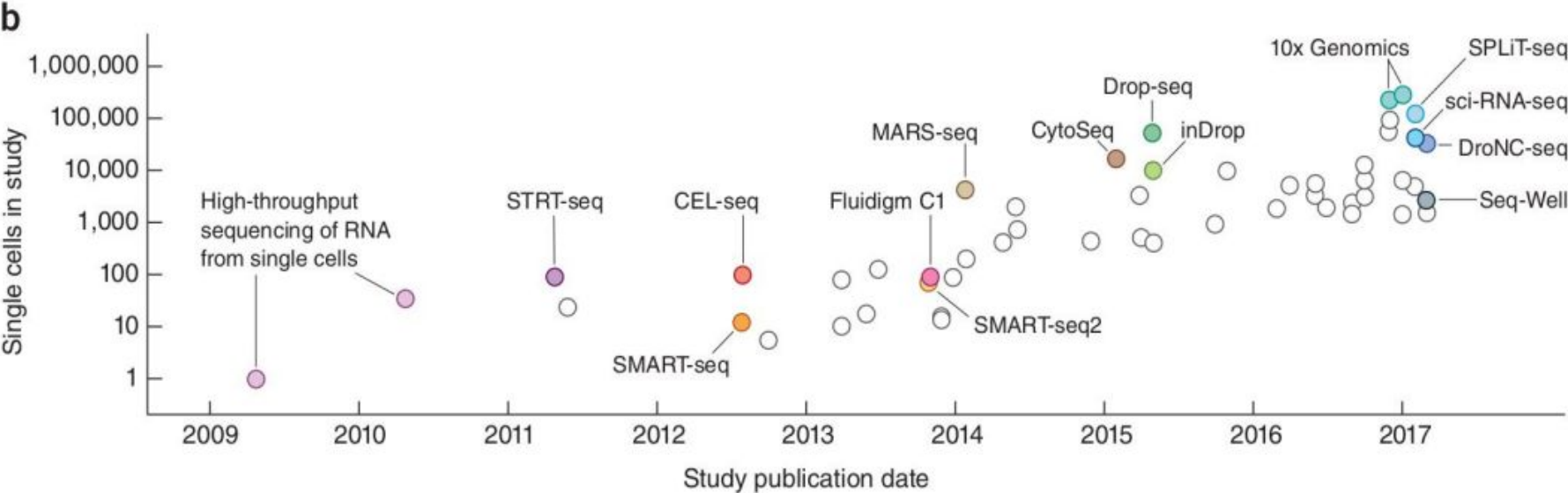
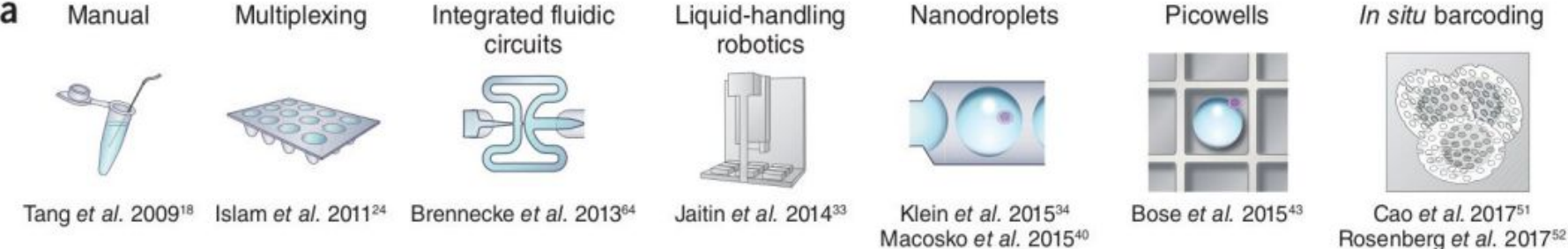
- Methods & Materials
  - Working With Enzymes
  - Basic Primary Cell Isolation
  - Equilibration with 95%O<sub>2</sub>:5%CO<sub>2</sub>
  - Trituration
  - Enzymatic Cell Harvesting
  - Cell Adhesion and Harvesting
  - Trypsin for Cell Harvesting
  - Cell Release Procedure
- Optimization Techniques
  - General Guidelines
  - Optimization Strategy
  - Cell Quantitation
  - Measure of Viability

## IV. Use-Tested Cell Isolation Systems

### Tissue Tables (references, grouped by tissue type and species)

Adipose/Fat	Adrenal	Bone	Brain
Cartilage	Colon	Endothelial	Epithelial
Eye	Heart	Intestine	Kidney
Liver	Lung	Lymph nodes	Mammary
Miscellaneous	Muscle	Neural	Pancreas
Parotid	Pituitary	Prostate	Reproductive
Scales	Skin	Spleen	Stem
Thymus	Thyroid/Parathyroid	Tonsil	Tumor

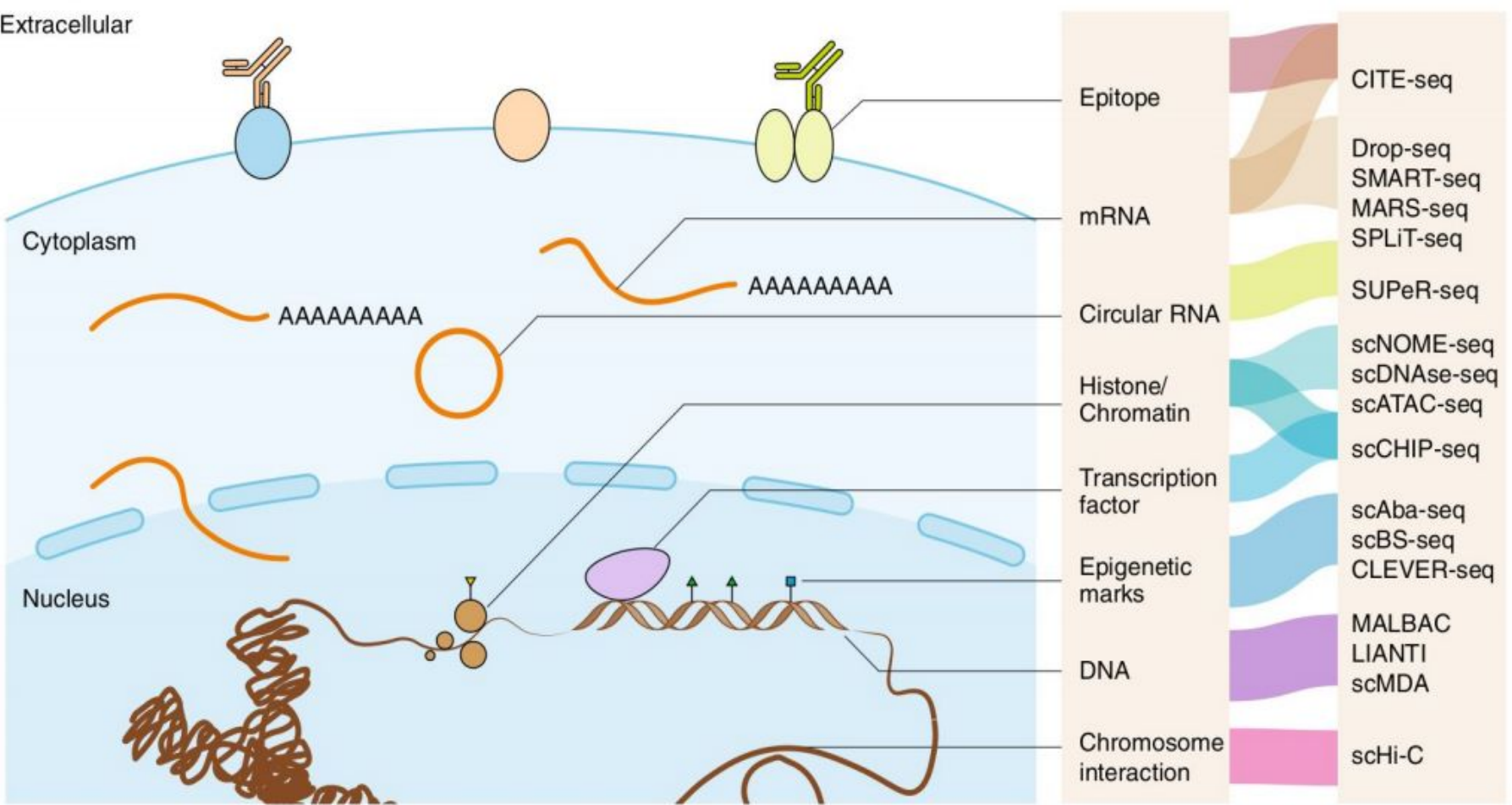
# Cells isolation : technologies over the last decade



Svensson et al. Nature protocols (2018)

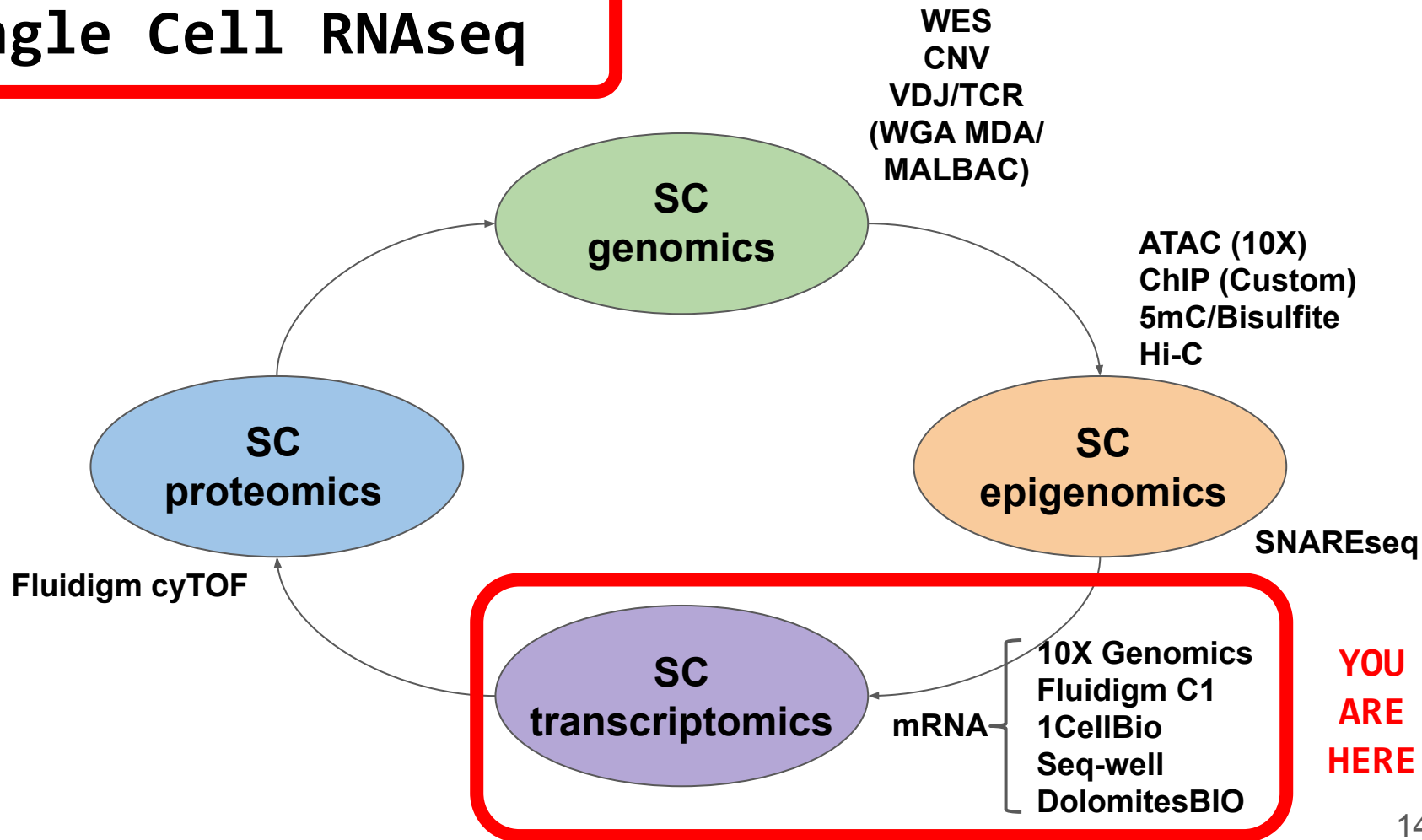
*From isolated cells to nucleotide sequences*

# Several protocols for several purposes

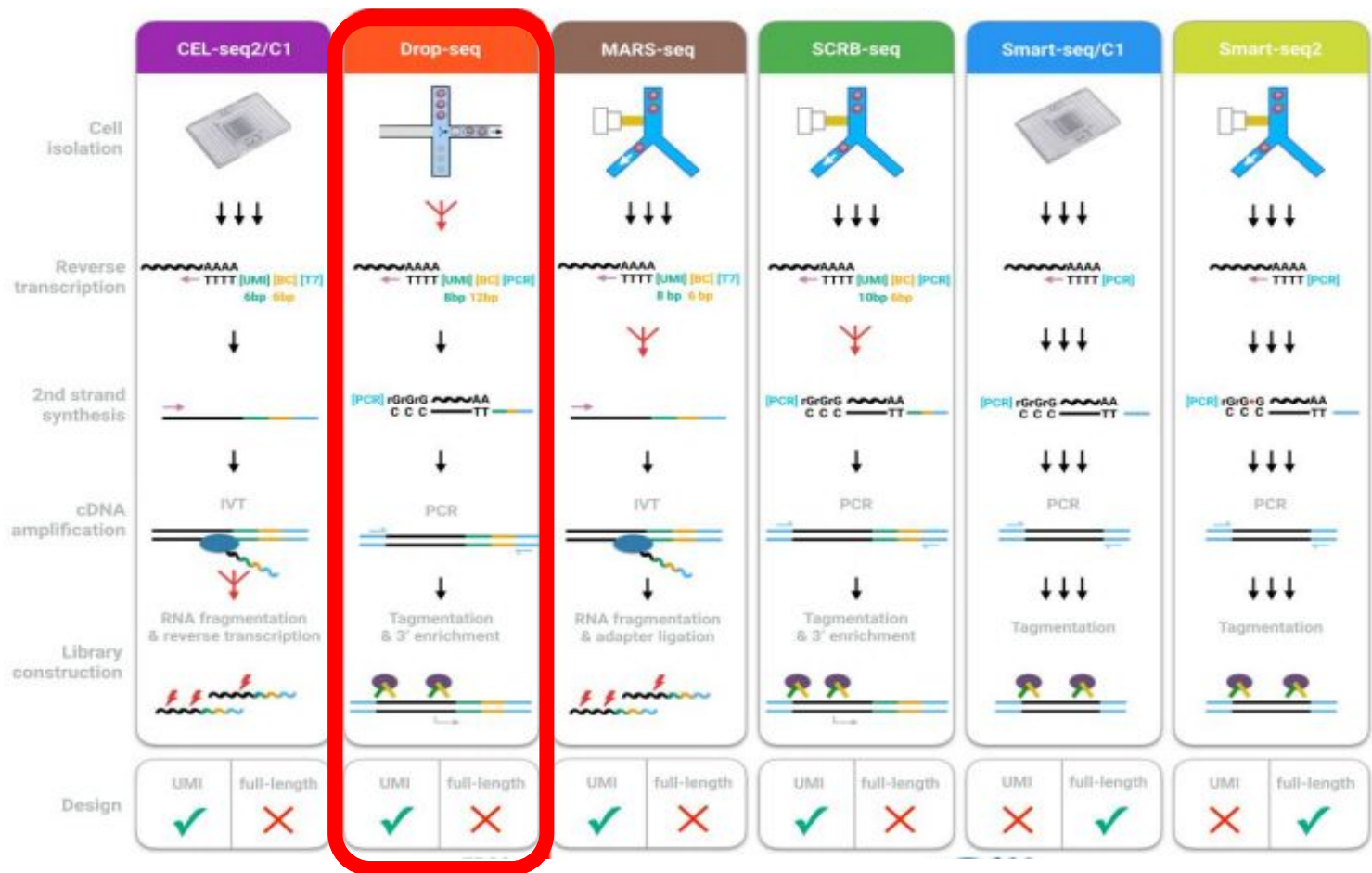


Ren et al. Genome Biology (2018)

# Single Cell RNAseq



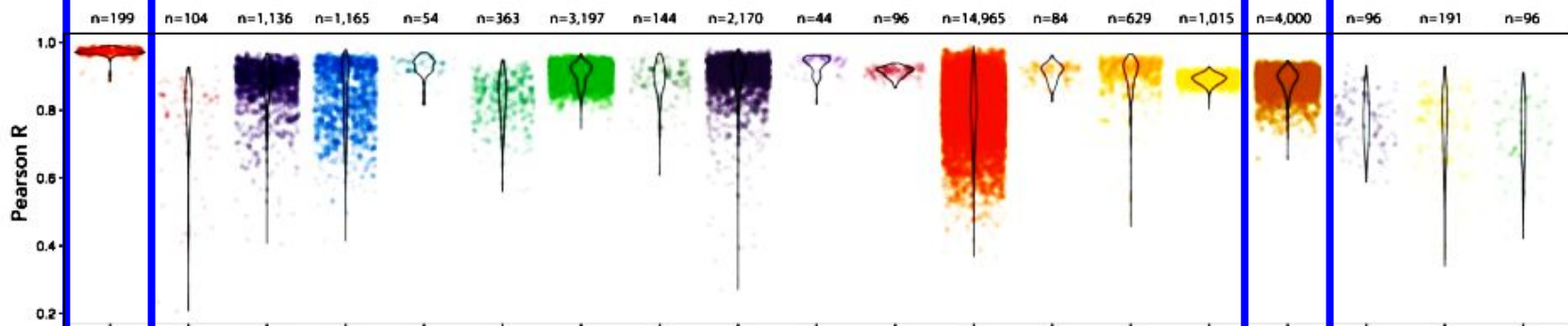
# From isolated cells to sequences (Drop-seq / 10X)



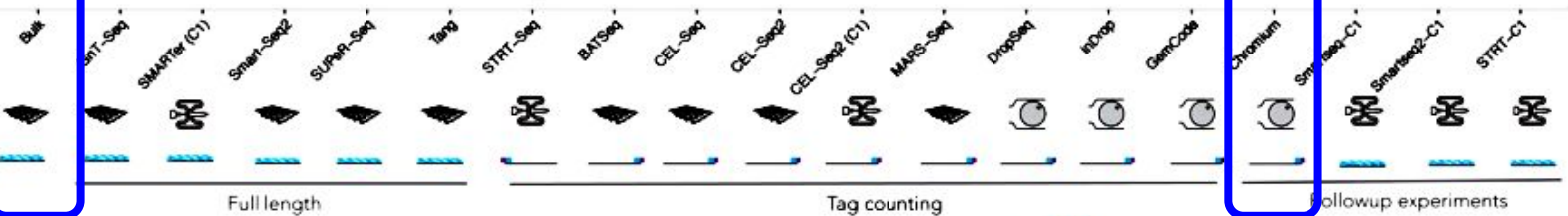
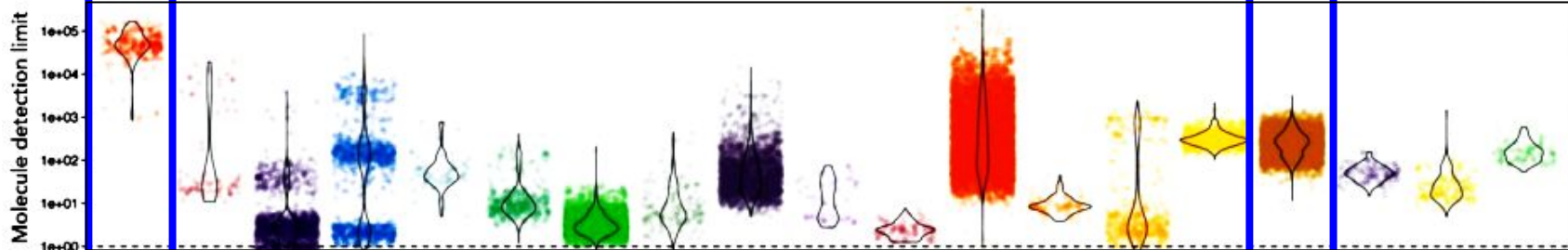
Bulk

10X

Accuracy



Sensitivity





# Drop-seq

**A**

1. Cells from suspension

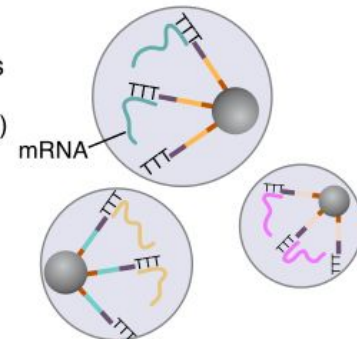
2. Microparticle and lysis buffer

3. Oil

Cell    Microparticle

5. RNA hybridization

4. Cell lysis  
(in seconds)



7. Reverse transcription  
with template switching

STAMPs

8. PCR  
(STAMPs  
as template)

9. Sequencing and analysis

- Each mRNA is mapped to its cell-of-origin and gene-of-origin
- Each cell's pool of mRNA can be analyzed

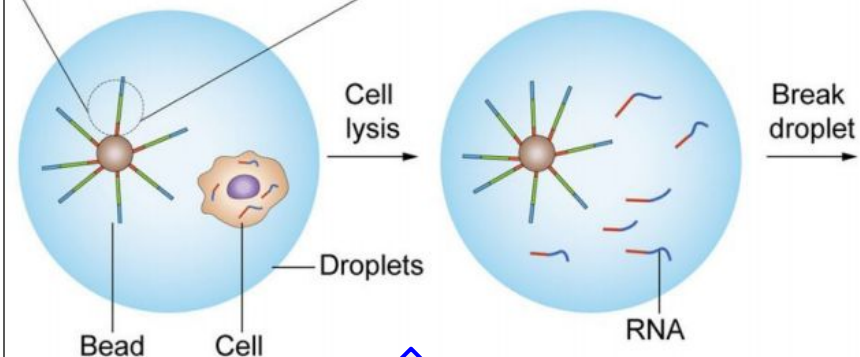
Macosko 2015

STAMPs = single-cell transcriptomes attached to microparticles

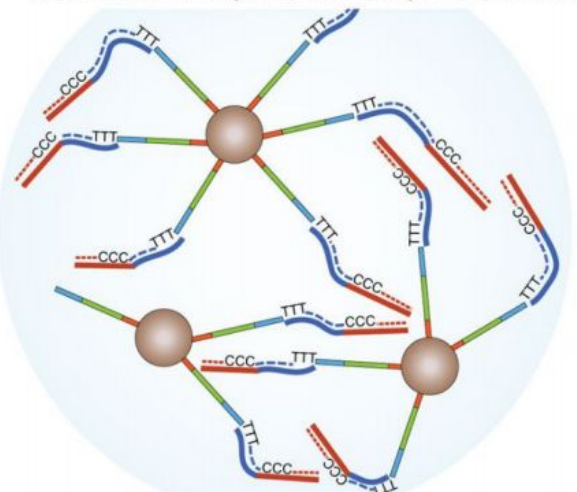
# 10X Chromium (3')

## Structure of the barcode primer bead

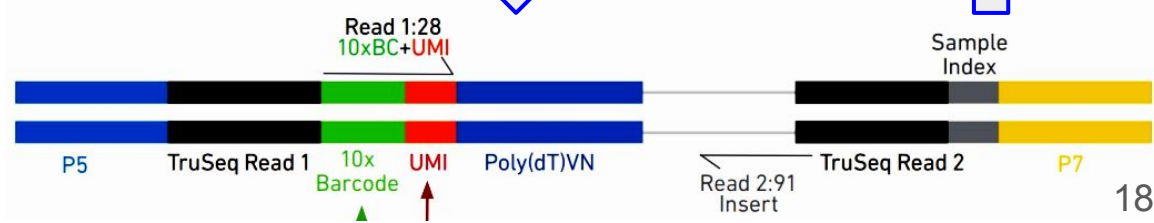
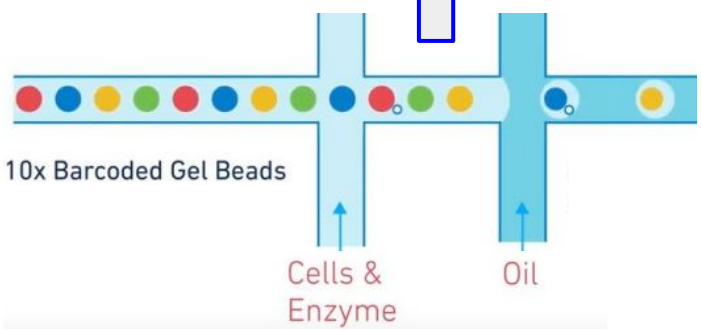
PCR handle Cell barcode UMI



## Reverse transcription with template switching



Sequencing



*From nucleotide sequences (reads)  
to count matrix*

CERTIFIED



BIOINFORMAGICIAN

# Reads QC

## FastQC Report

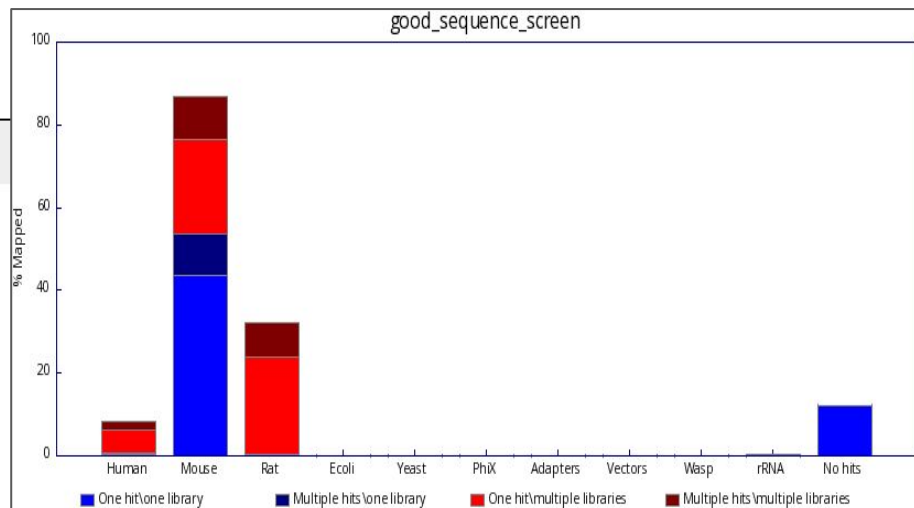
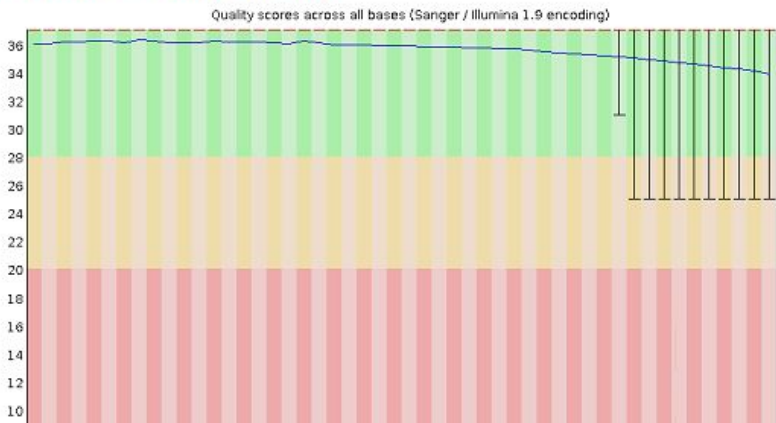
### Summary

- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✗ Per tile sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✓ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✓ Adapter Content

### Basic Statistics

Measure	Value
Filename	BC_392_1_529_R2_001.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	109443265
Sequences flagged as poor quality	0
Sequence length	91
%GC	43

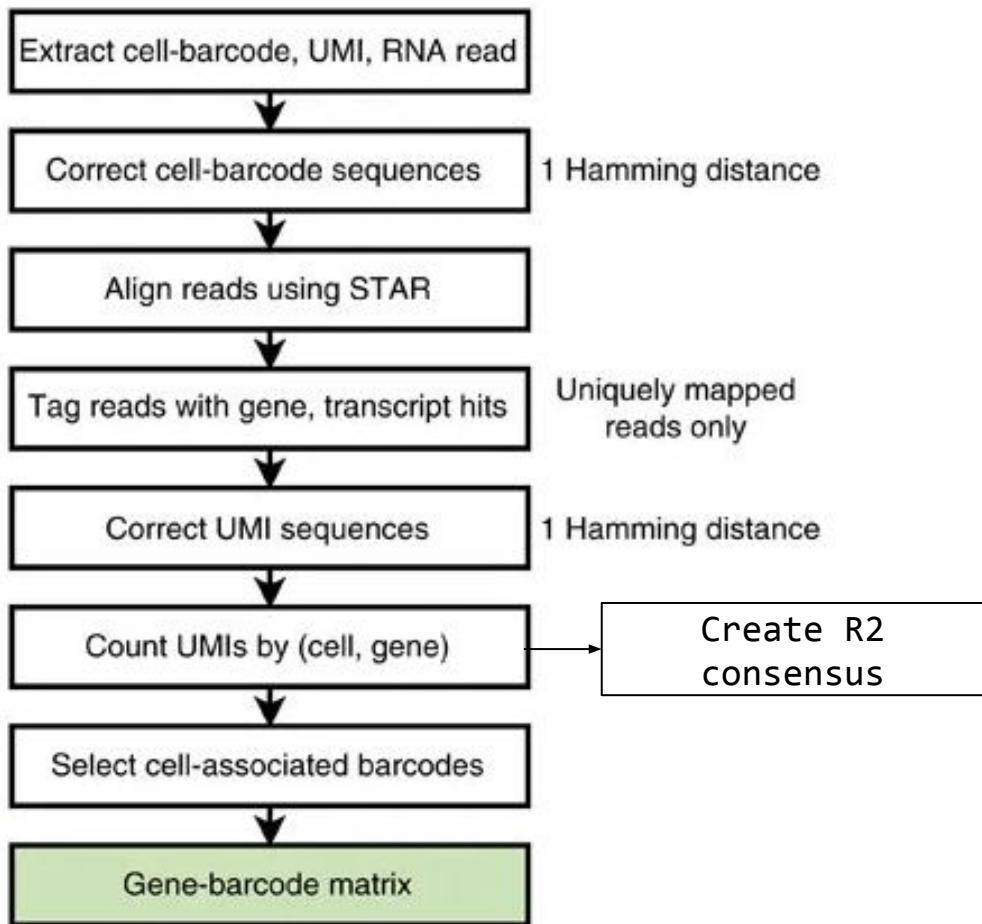
### Per base sequence quality



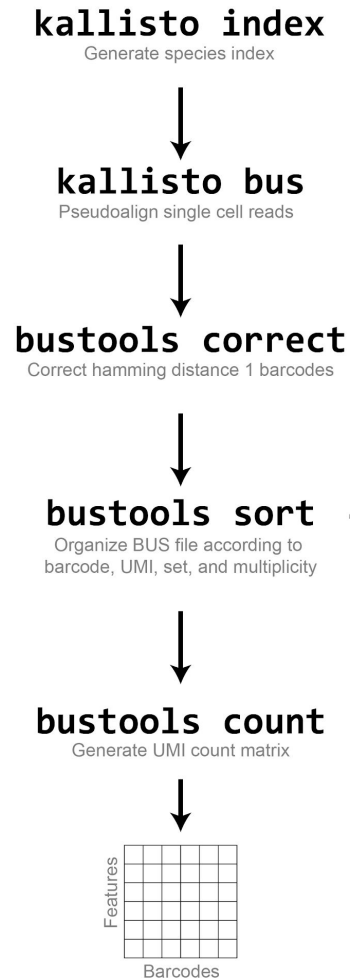
- As usual : FASTQC, FastqScreen, ...
- 10x :
  - Special attention to R1 : cell barcode + UMI (no N)
  - Control of the 4 sample libraries

# Reads processing workflows

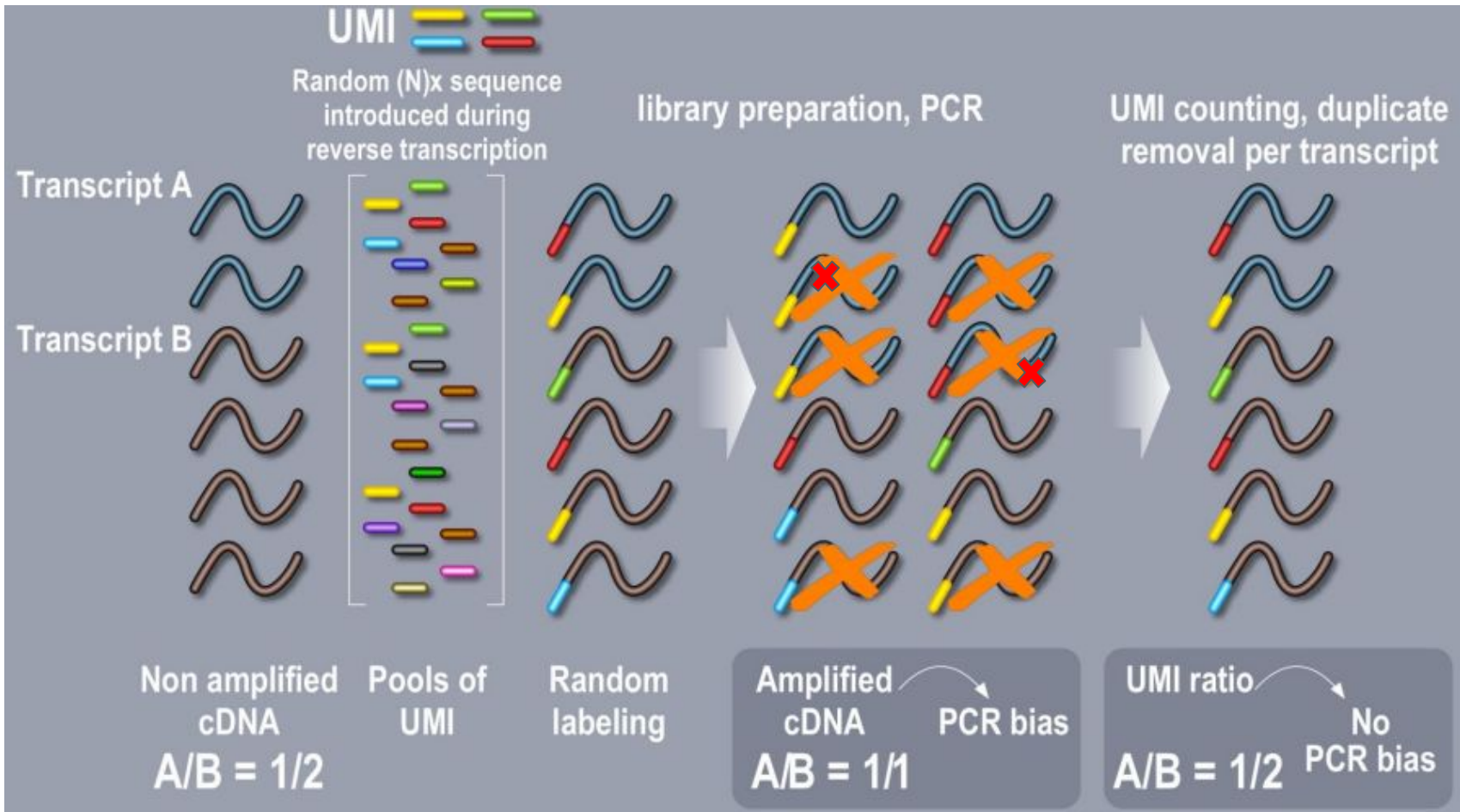
Mapping-based (STAR)



Pseudomapping-based (kallisto bustools)

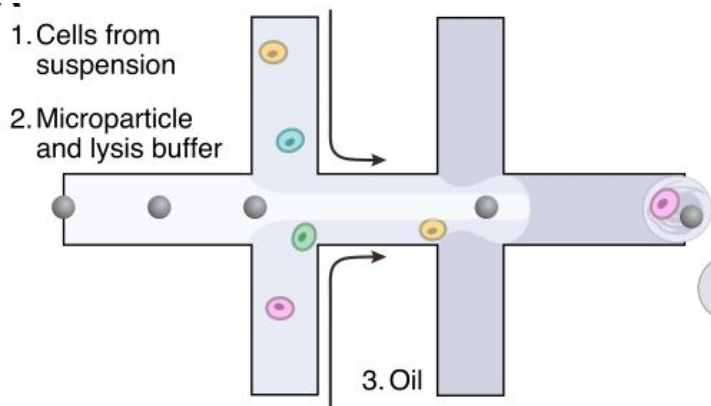


# Focus on : Unique Molecule Identifiers



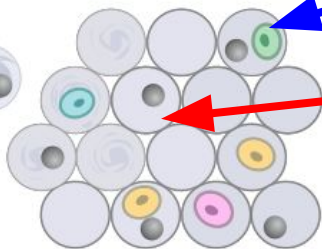
Islam et al., Nature Methods (2014)  
Study by Agnès Paquet

# Focus on : Empty droplets filtering

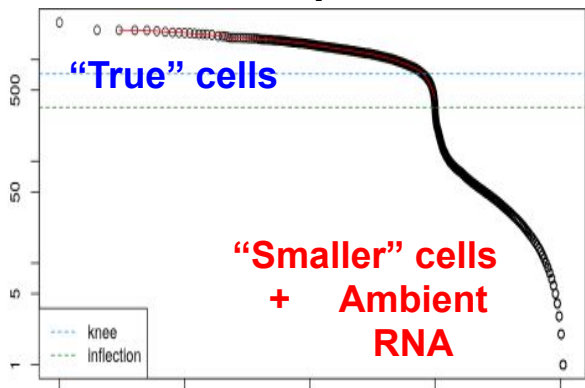


**THERE IS RNA HERE  
(CELL IN GEM)**

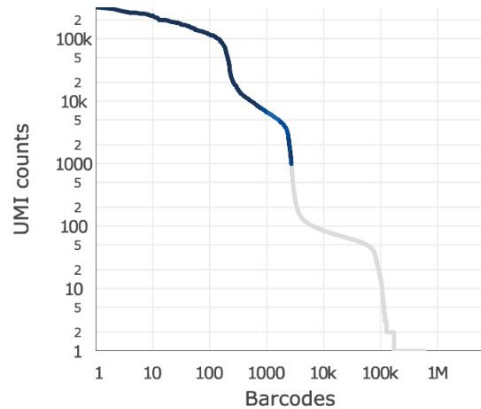
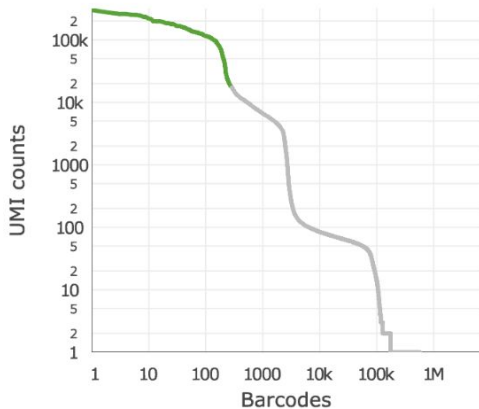
**THERE IS RNA HERE TOO !  
(NO CELL = AMBIENT)**



## KneepLOT



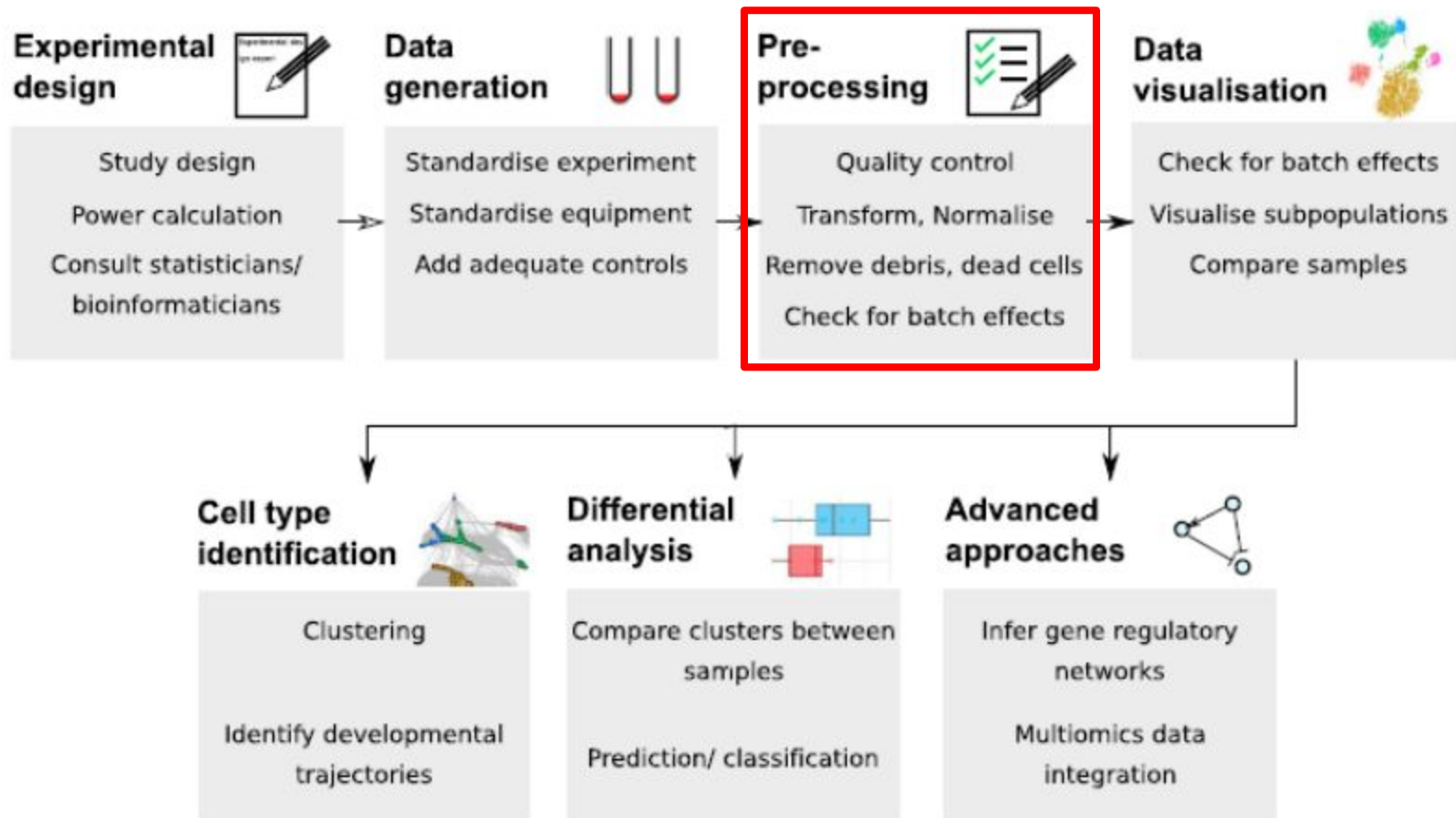
## EmptyDrops





*From raw count matrix to normalized data*

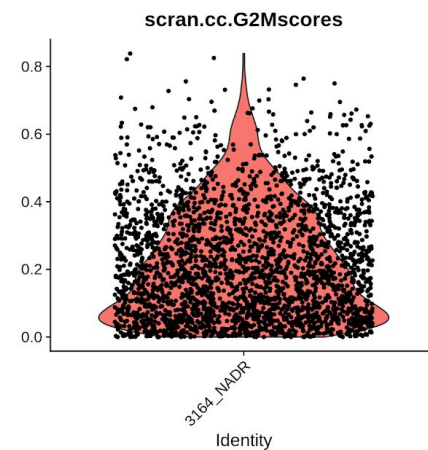
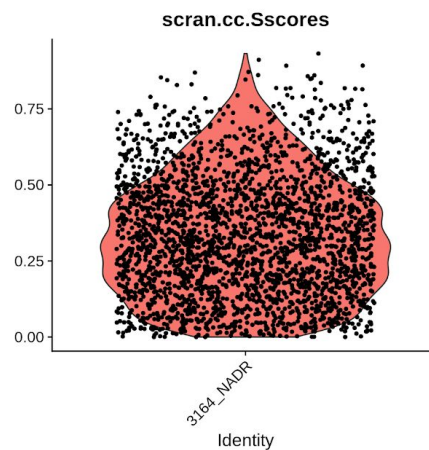
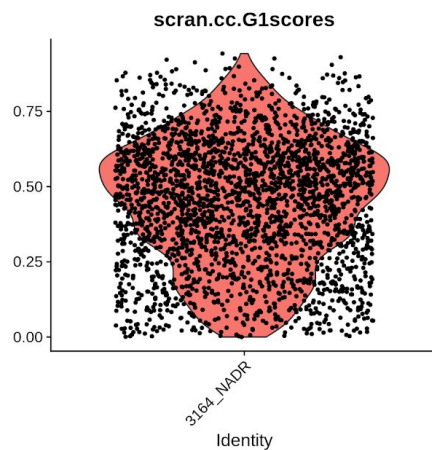
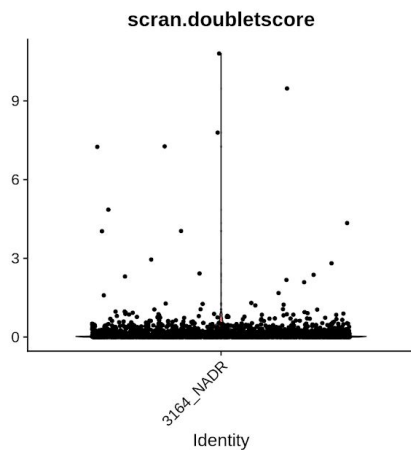
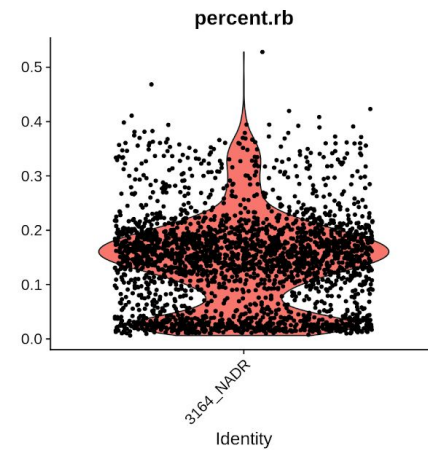
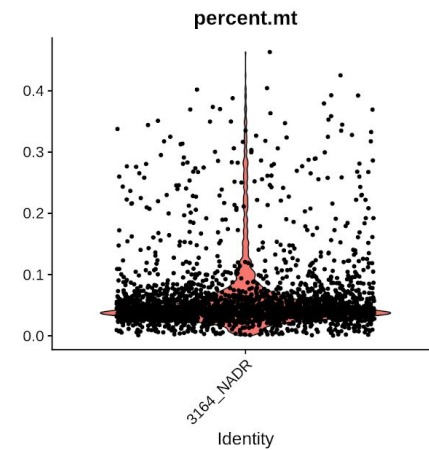
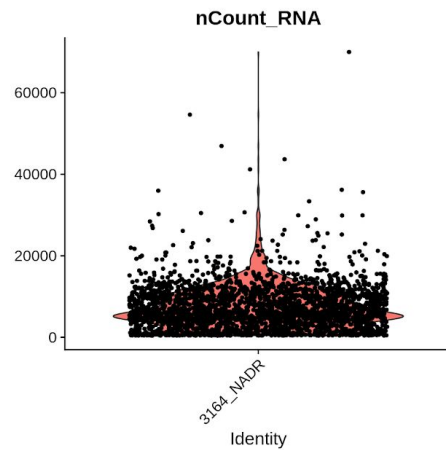
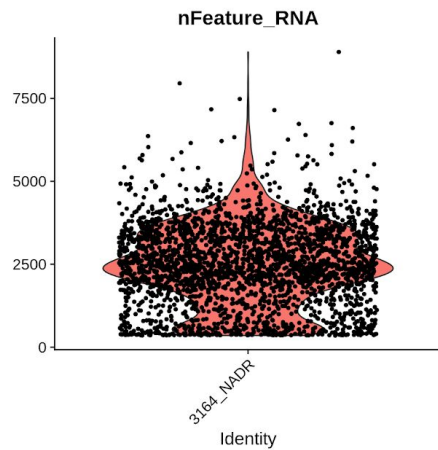
# Standard analysis pipeline



# Cell QC considerations

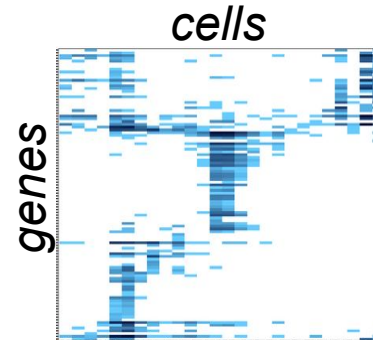
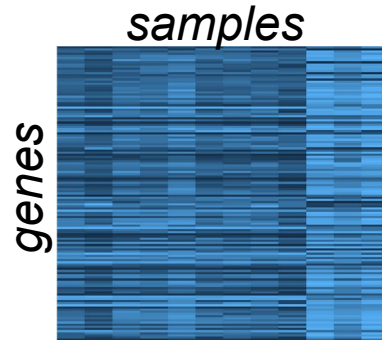
- The number of unique genes detected in each cell :
  - Low-quality cells or empty droplets will often have very few genes
  - Cell doublets or multiplets may exhibit an aberrantly high gene count
- Similarly, the total number of molecules detected within a cell (correlates strongly with unique genes)
- The percentage of reads that map to the mitochondrial genome :
  - Low-quality / dying cells often exhibit extensive mitochondrial contamination
- Other QC criteria to measure :
  - Cell cycle phase / score
  - Nuclear riboprotein-coding genes expression

# Cell QC : metrics



# Matrix normalization : Houston, we have a problem...

	<b>BULK</b>	<b>SINGLE-CELL</b>
Total RNA	100 ng (~10.000 cells)	10 pg (per cell)
mRNA	~ 5 ng (~10.000 cells)	<< 1 pg (per cell)
Reads	~100 million	~ 50 k (per cell)

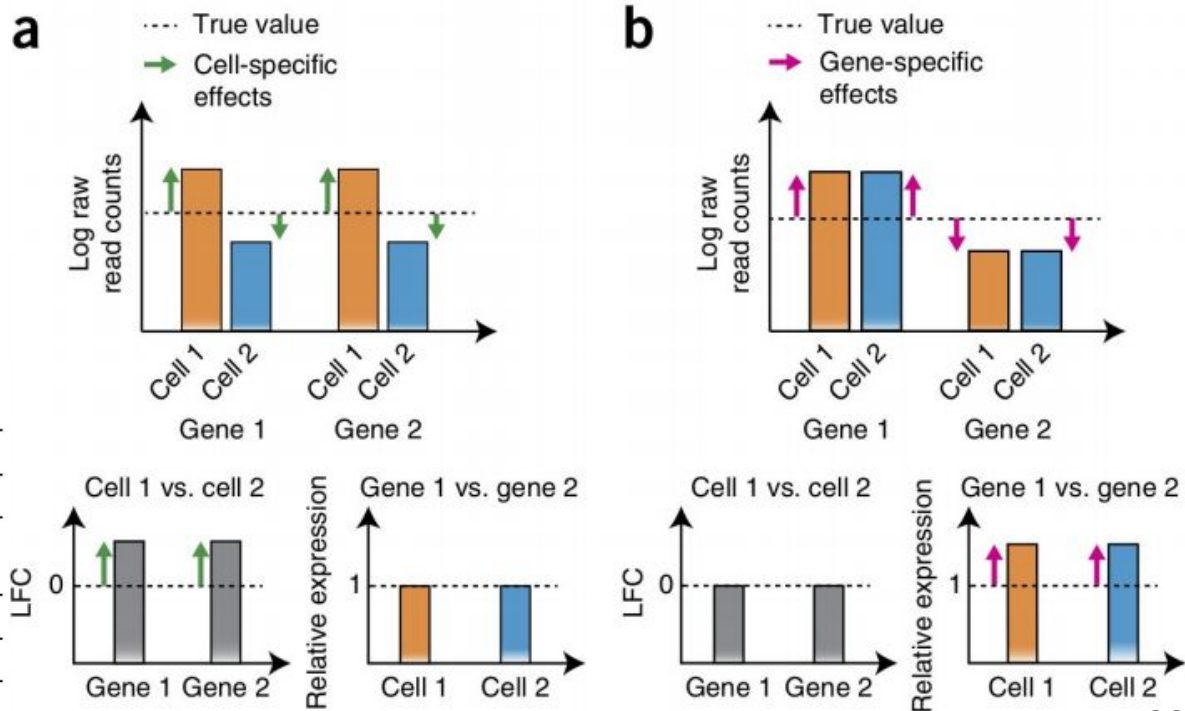


**SC MATRIX IS  
SPARSE !  
(ie, mostly  
filled with  
zeros)**

# Matrix normalization : different levels

- Process of **identifying** and **removing** systematic variation not due to real differences between RNA treatments i.e. differential gene expression.

- Cell-specific effects
- Gene-specific effects



# Bulk normalization methods are **KO**

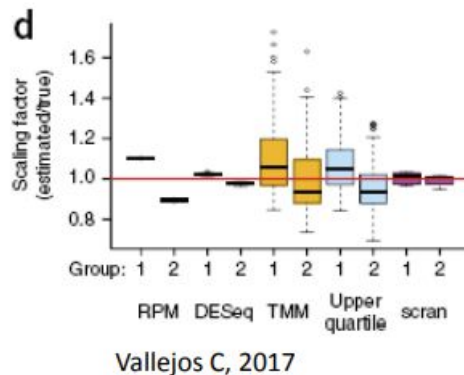
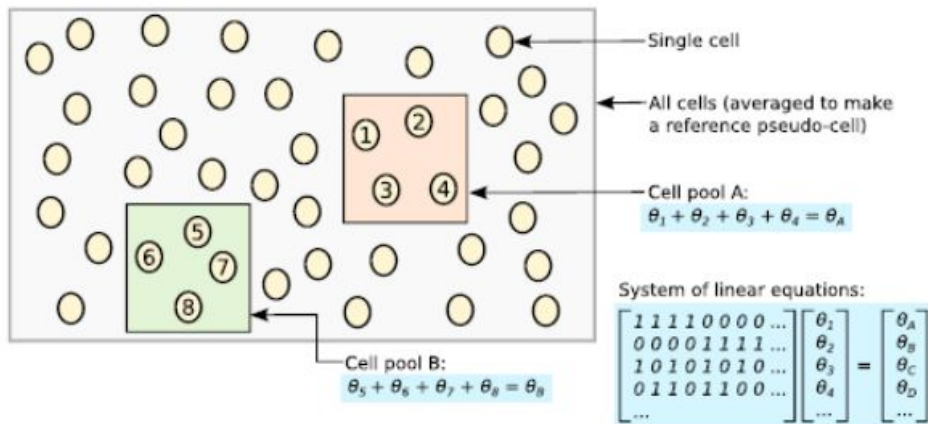
- RPKM/FPKM (Reads/Fragments per kilobase of transcript per million reads of library) : Normalize for sequencing depth and transcript length at the same time => **KO** if you **DO NOT** have full-length data
- Global scaling (eg: Upper Quartile) : **KO** if you have too many zeros
- Size factors calculation (eg: Estimation of library sampling depth) :
  - DESeq2, edgeR suppose that  $\geq 50\%$  of genes are **NOT** DE
  - **KO** if you have too many zeros
- TPM/CPM : **KO** if a small number of genes carry most of the signal

=> Rough solution : global log-normalization / Z-scoring



# Matrix normalization : scaling by factors

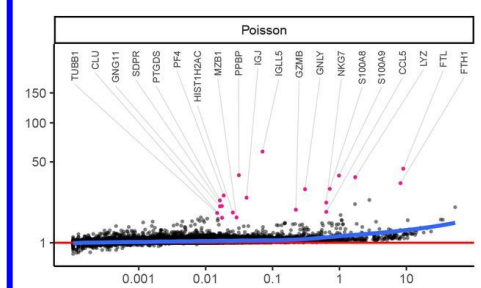
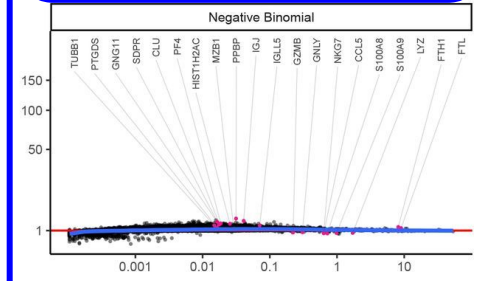
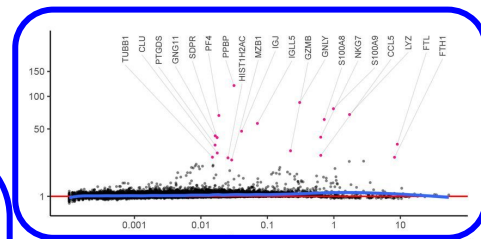
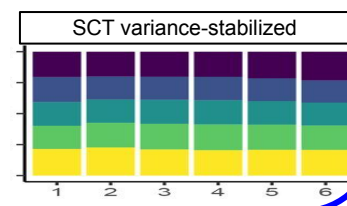
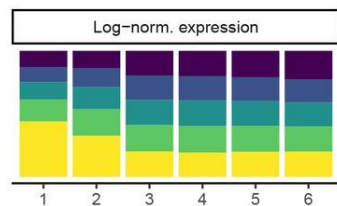
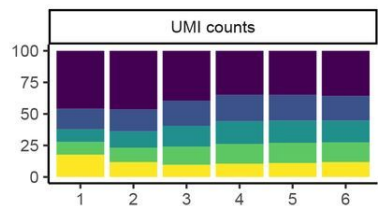
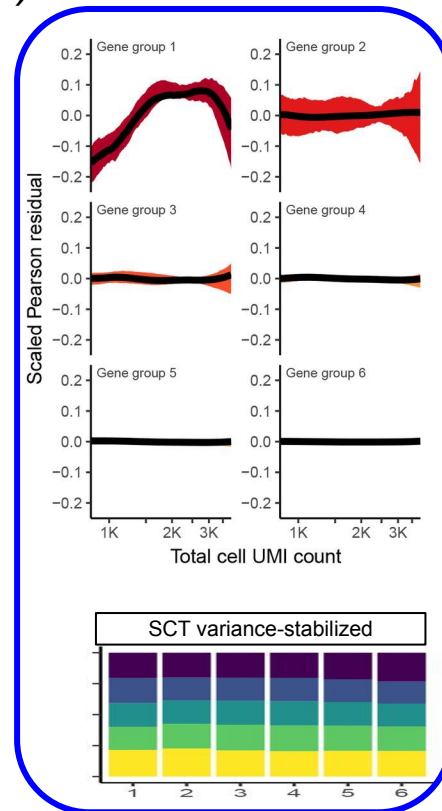
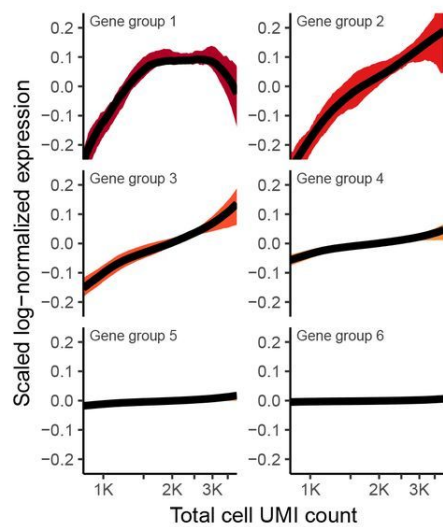
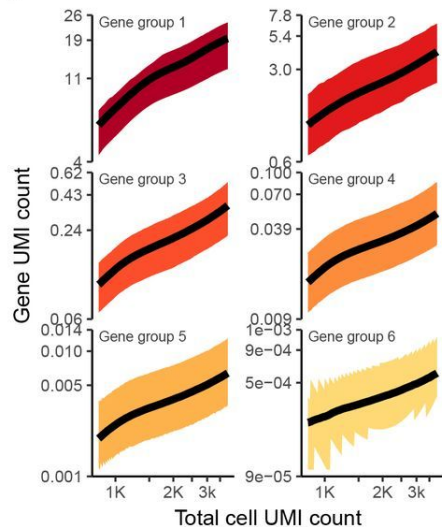
- Alternative method to compute the size factors
- Pool cells to reduce the number of zeros
- Estimate the size factors for the pool
- Repeat many time and use deconvolution to estimate each cell size factor
- Implemented in **scater**/**scran** packages





# Matrix normalization : variance stabilization

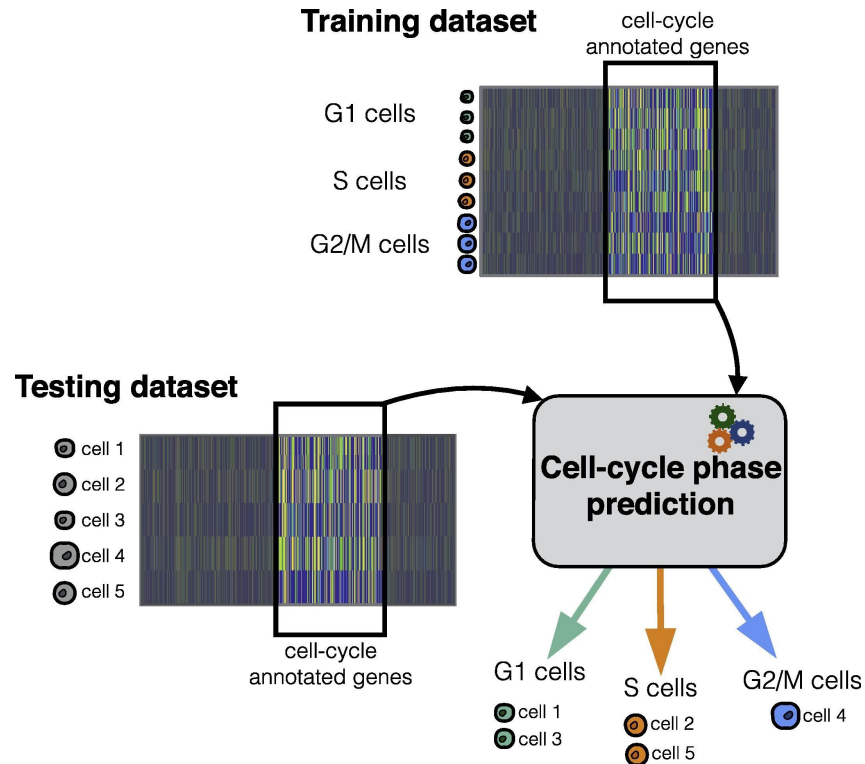
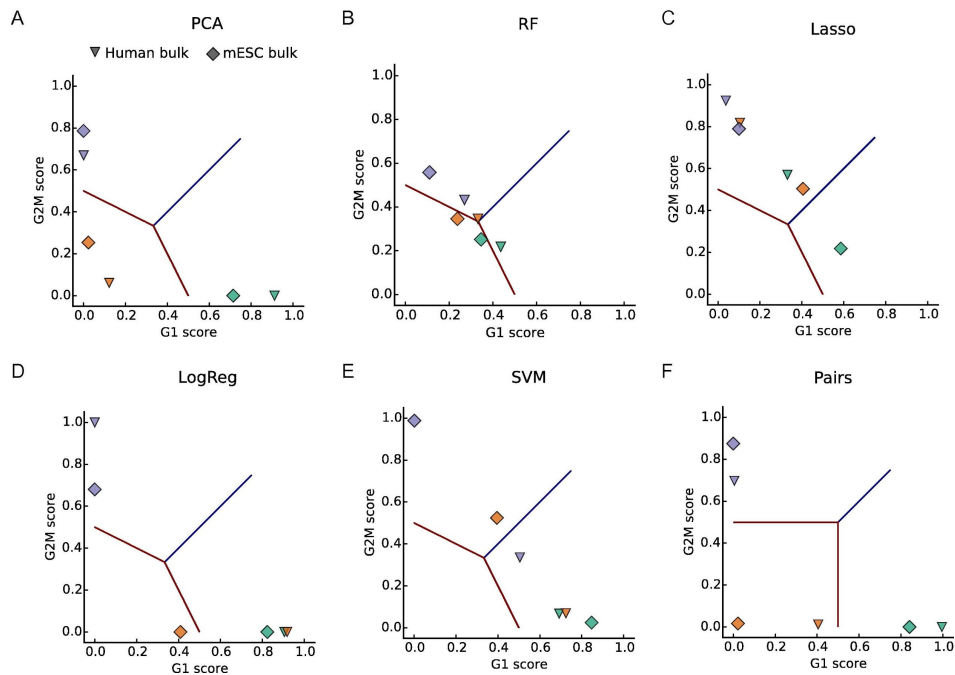
- Regularized negative binomial regression
- Implemented in **sctransform** (*Seurat*)



Cell group		Gene group ID, size			
1	2	3	4	5	6
1	2	3	4	5	6
55	171	1687	5942	4694	4260

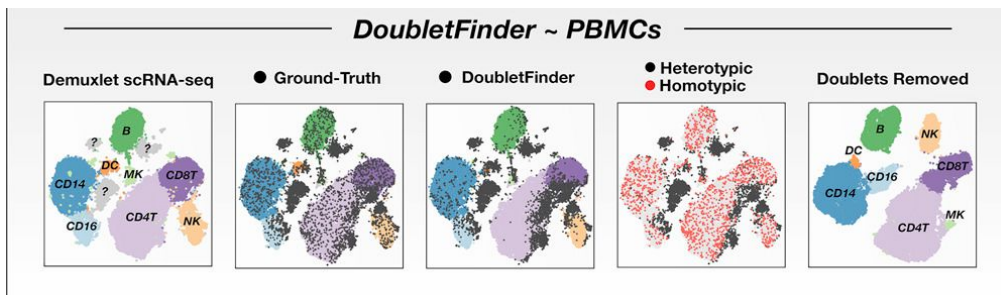
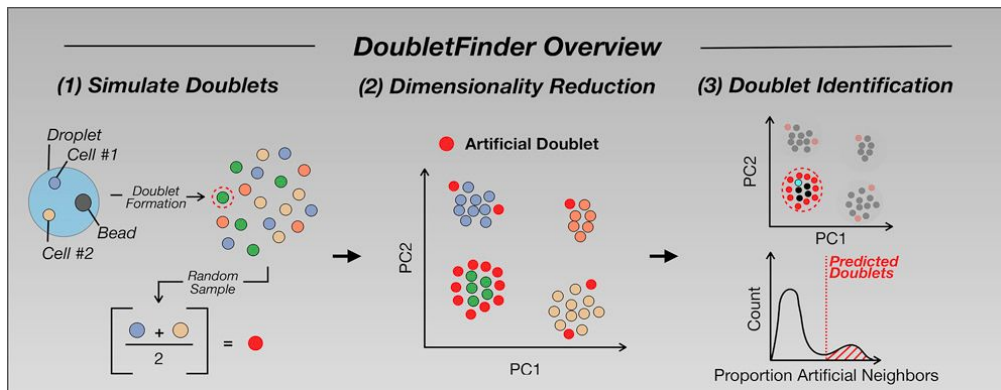
# Other factors : Cell cycle phase

- Training on reference set with the 3 phases identified
- Use pairs of differential genes
- Apply model pairs to new dataset and assign phases
- Implemented in **cyclone** (*scrn*)



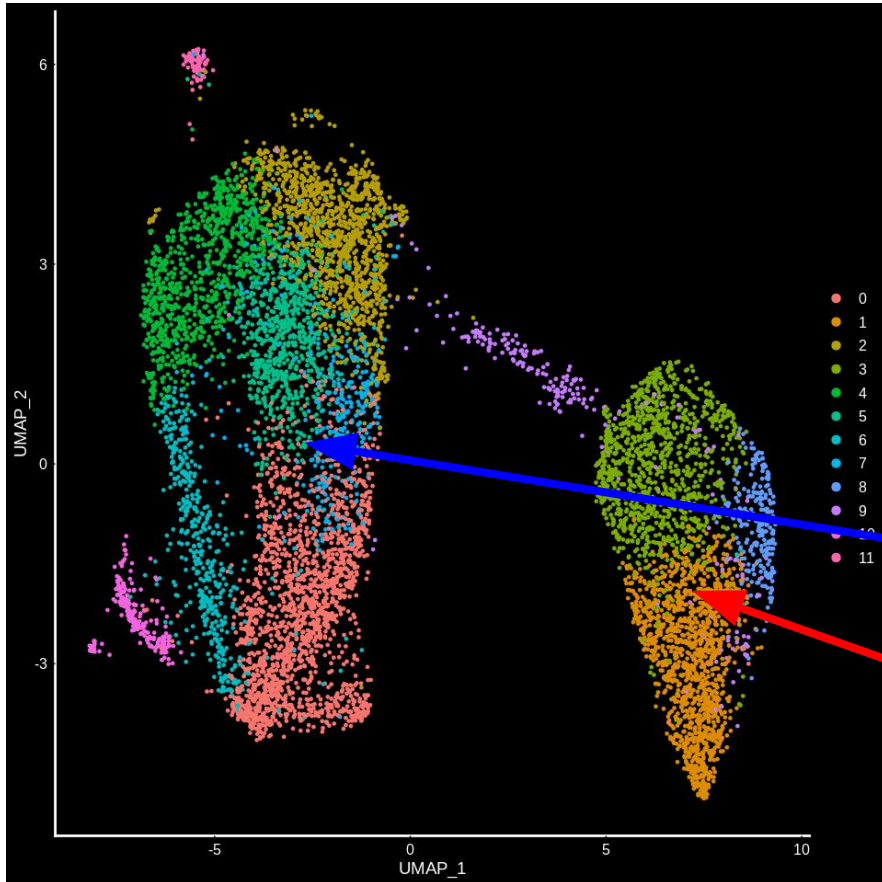
# Other factors : Cell doublets

- Two types of doublets :
  - Cells of the same type => higher global expression
  - Cells of the different types => artificial hybrid
- Methods : generate random artificial doublets, capture all



	AUC	pAUC90	pAUC95	pAUC97.5	AUPRC
<b>ch_cell-lines</b>					
● libsize	0.60	0.54	0.53	0.52	0.17
● features	0.60	0.55	0.54	0.53	0.19
● dblCells	0.64	0.62	0.61	0.60	0.37
● cxdx	0.65	0.59	0.57	0.55	0.26
● dblDetection	0.66	0.66	0.65	0.65	0.44
● scrublet	0.69	0.65	0.64	0.63	0.41
● dblFinder	0.69	0.66	0.65	0.65	0.45
● hybrid	0.70	0.64	0.63	0.61	0.40
● bcxs	0.70	0.66	0.64	0.62	0.43
<b>ch_pbmc</b>					
● dblCells	0.63	0.57	0.56	0.54	0.31
● libsize	0.78	0.63	0.57	0.54	0.44
● scrublet	0.78	0.67	0.63	0.59	0.52
● cxdx	0.78	0.69	0.65	0.61	0.54
● features	0.79	0.62	0.57	0.54	0.45
● bcxs	0.81	0.71	0.66	0.60	0.58
● hybrid	0.82	0.73	0.67	0.62	0.61
● dblDetection	0.82	0.75	0.69	0.62	0.63
● dblFinder	0.84	0.74	0.68	0.62	0.64
<b>demuxlet</b>					
● dblCells	0.79	0.70	0.65	0.60	0.46
● libsize	0.81	0.58	0.55	0.53	0.30
● features	0.85	0.62	0.57	0.55	0.37
● scrublet	0.87	0.74	0.68	0.62	0.53
● cxdx	0.89	0.71	0.63	0.57	0.49
● hybrid	0.91	0.78	0.68	0.58	0.57
● dblDetection	0.91	0.79	0.69	0.58	0.57
● bcxs	0.91	0.79	0.71	0.62	0.61
● dblFinder	0.92	0.79	0.70	0.63	0.62
<b>hg-mm</b>					
● libsize	0.87	0.66	0.59	0.54	0.27
● features	0.89	0.68	0.60	0.55	0.30
● dblCells	0.93	0.88	0.84	0.79	0.73
● bcxs	0.96	0.87	0.80	0.71	0.64
● hybrid	0.98	0.94	0.90	0.87	0.88
● scrublet	0.99	0.96	0.94	0.91	0.91
● cxdx	0.99	0.98	0.98	0.97	0.97
● dblDetection	0.99	0.99	0.98	0.98	0.97
● dblFinder	1.00	0.99	0.99	0.99	0.99

# Normalization : other biological factors

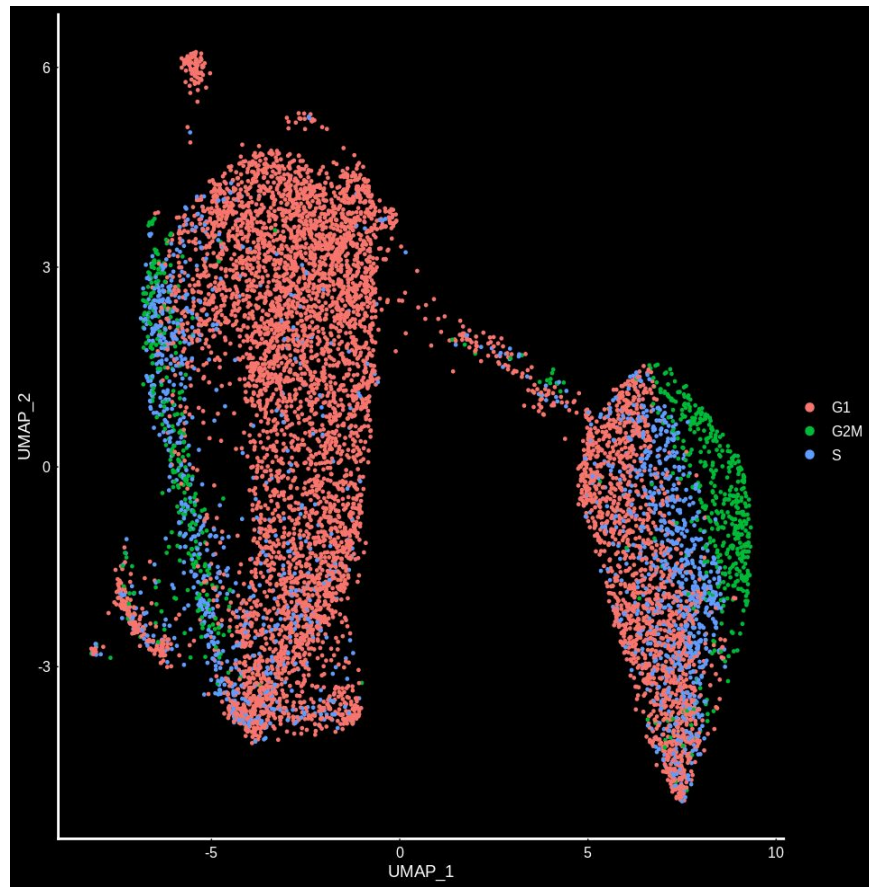
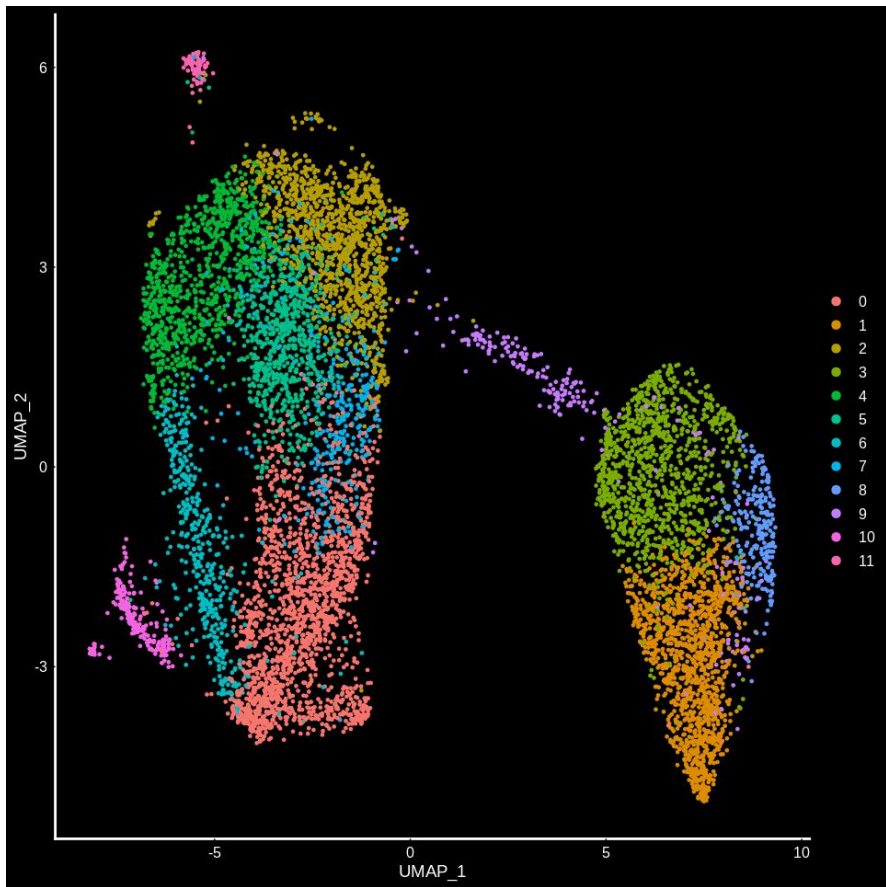


- 10X 3' scRNAseq
- Osteosarcoma metastasis
- 8911 cells x 18613 genes
- PCA (109 PCs retained)
- Louvain clustering
  - 12 clusters
- uMAP representation

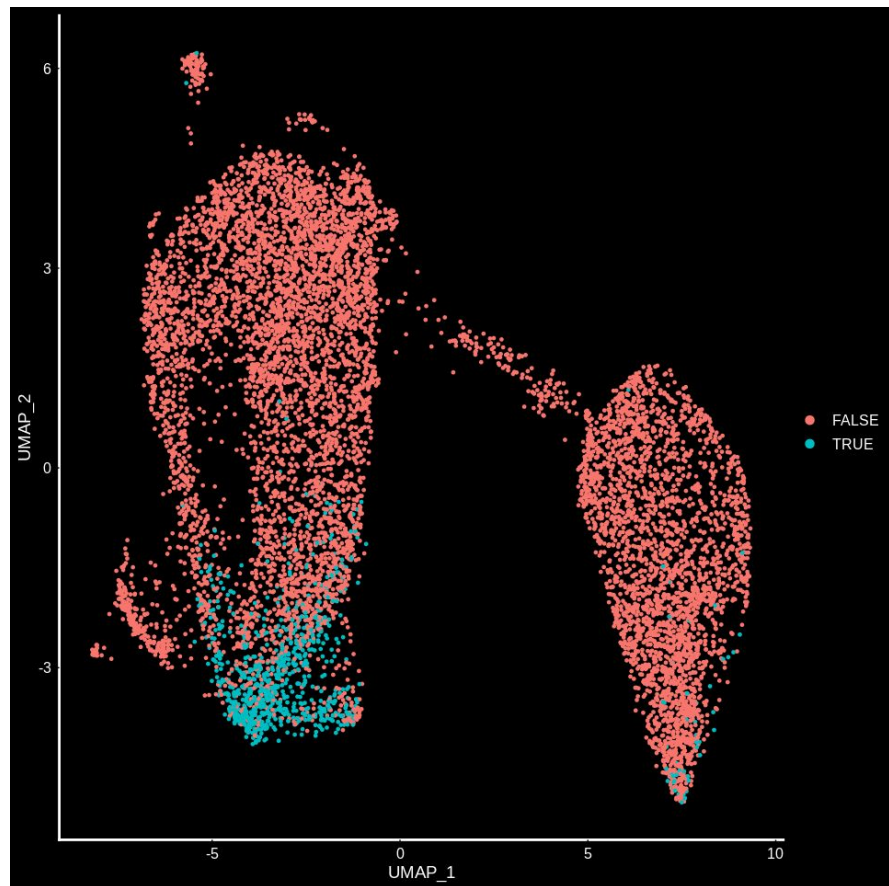
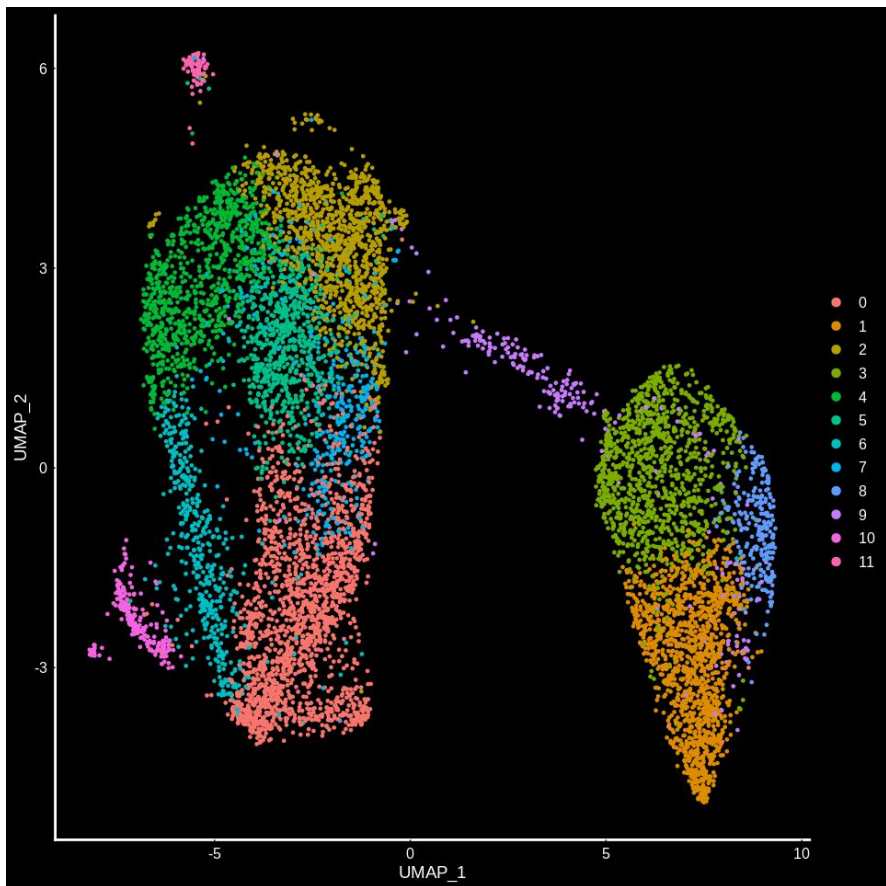
**Osteoblasts**

**Osteoclasts**

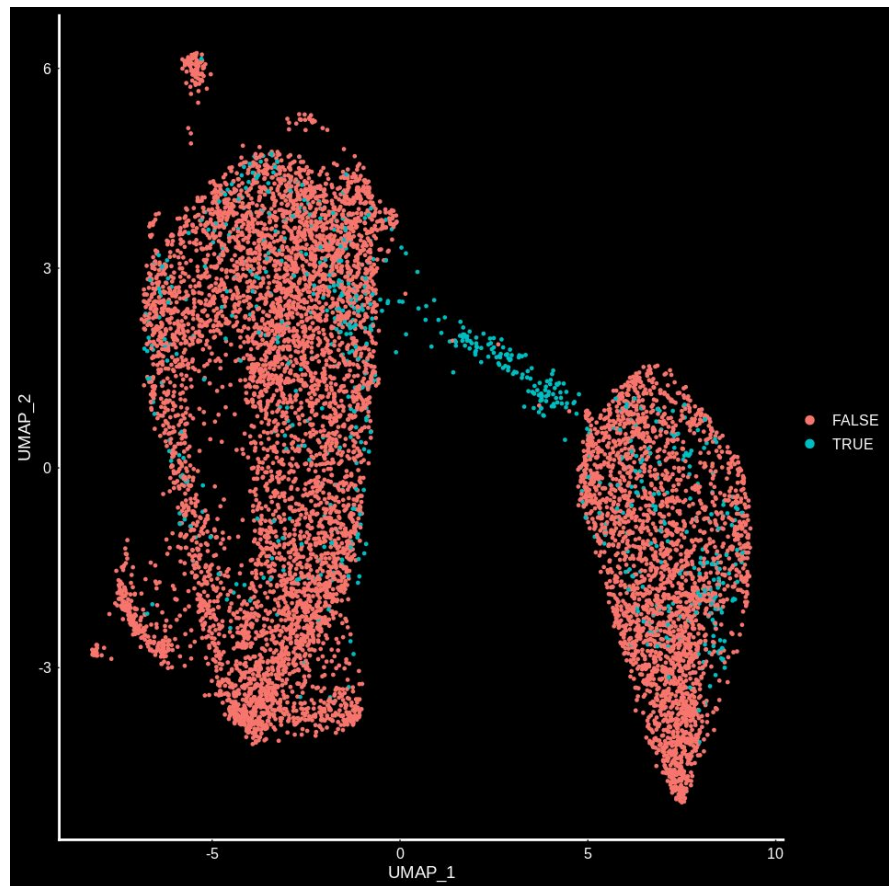
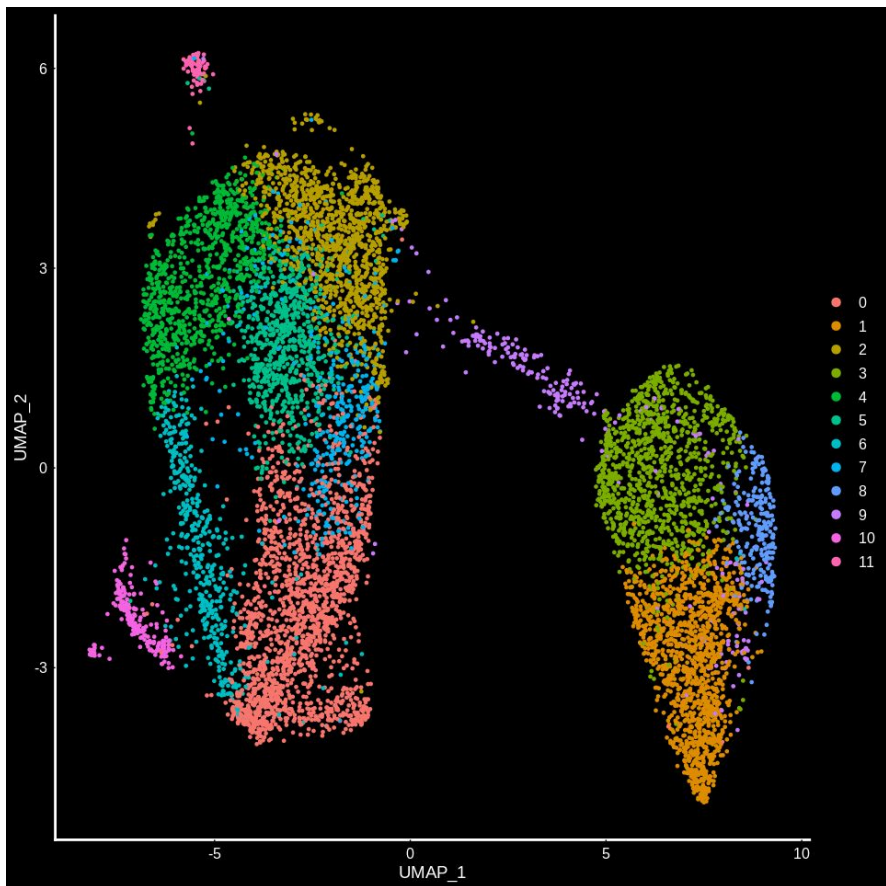
# Bias : Cell cycle phases / scores



# Bias : Dying cells status / score

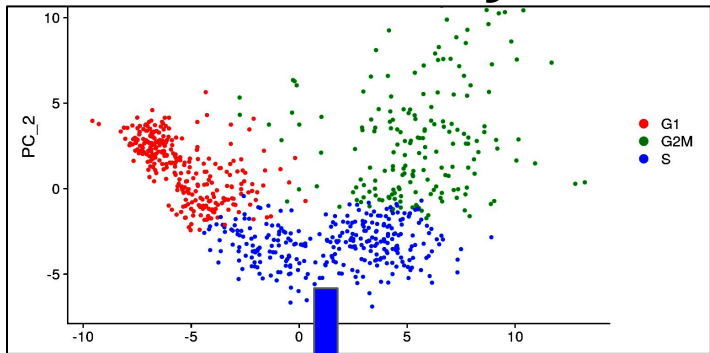


# Bias : Cell doublet status / score

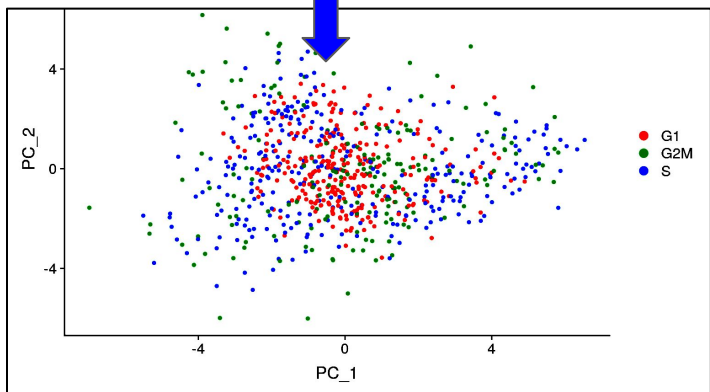


# Bias normalization : regression / deblocking

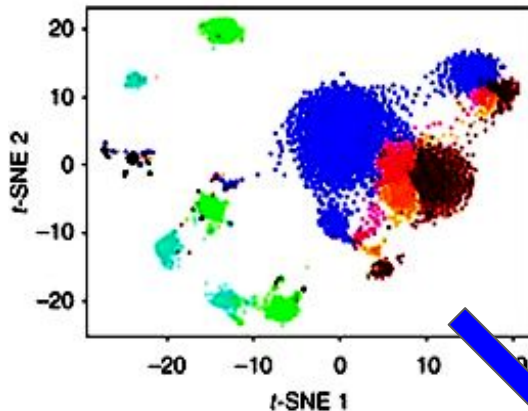
## Ex : Cell cycle



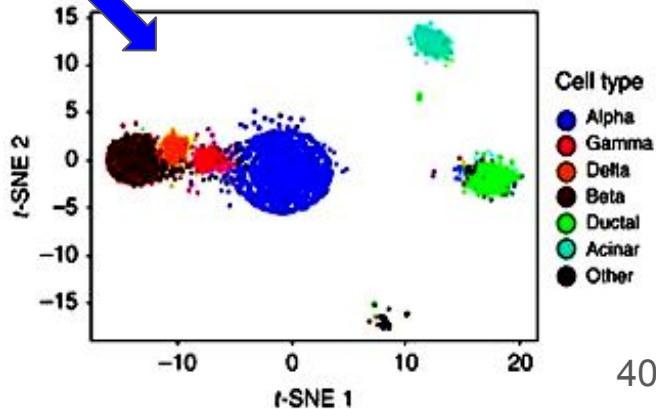
Regression



## Ex : Batch effect



Deblocking



Seurat tutorial

Haghverdi et al. Nature  
Biotech (2018) (MNN)

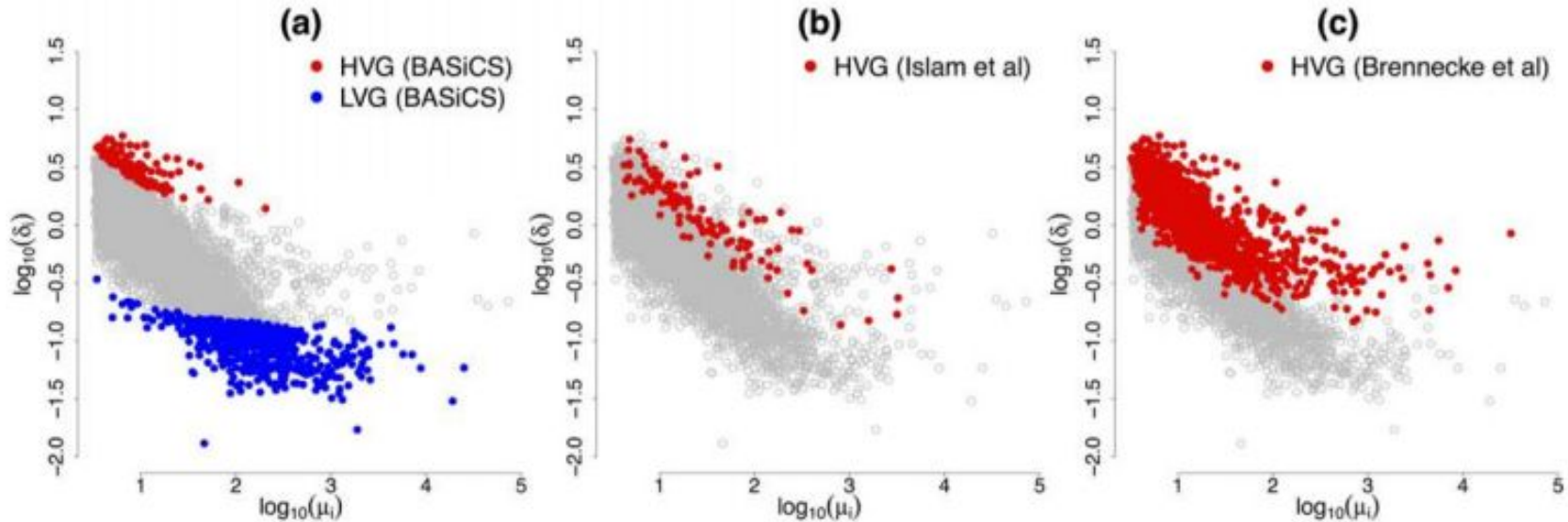


*From a normalized matrix to a reduced space*

# Feature selection : Highly variable genes (HVGs)

Postulate : genes with the highest variability should be the most useful to

1. Assess effect of unwanted sources of variation (cell to cell variation)
2. Quantify true biological differences (population to population variation)



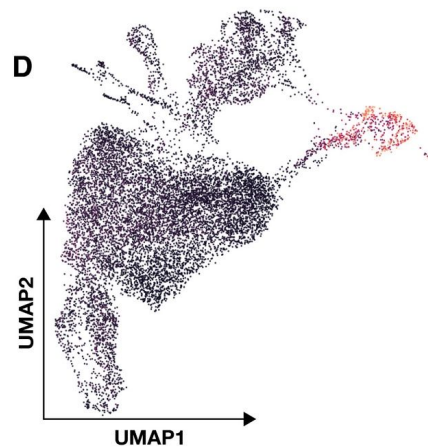
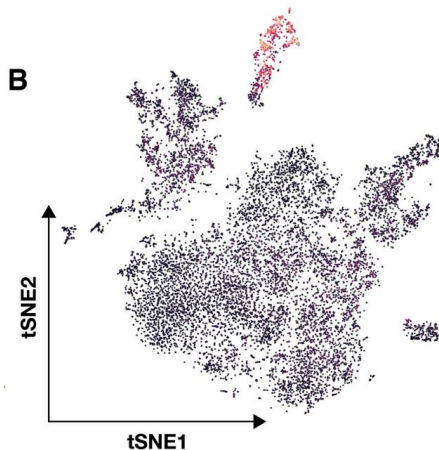
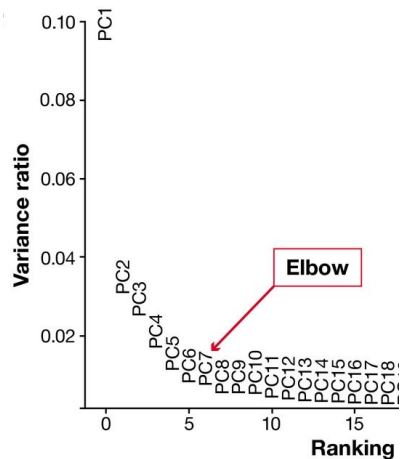
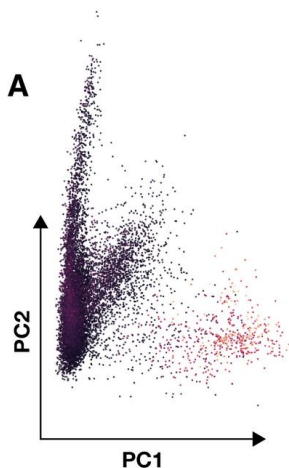
**Fig 8. Comparison of HVG detection among different methods.** For each of the 7,895 biological genes, posterior medians of biological cell-to-cell heterogeneity term  $\delta_i$  (log scale) against posterior medians of expression level  $\mu_i$  (log scale). While the methods described in [16] and [5] only provide a characterisation of HVG, BASiCS is able to detect those genes whose expression rates are stable among cells.

# Dimensionality reduction : simplification

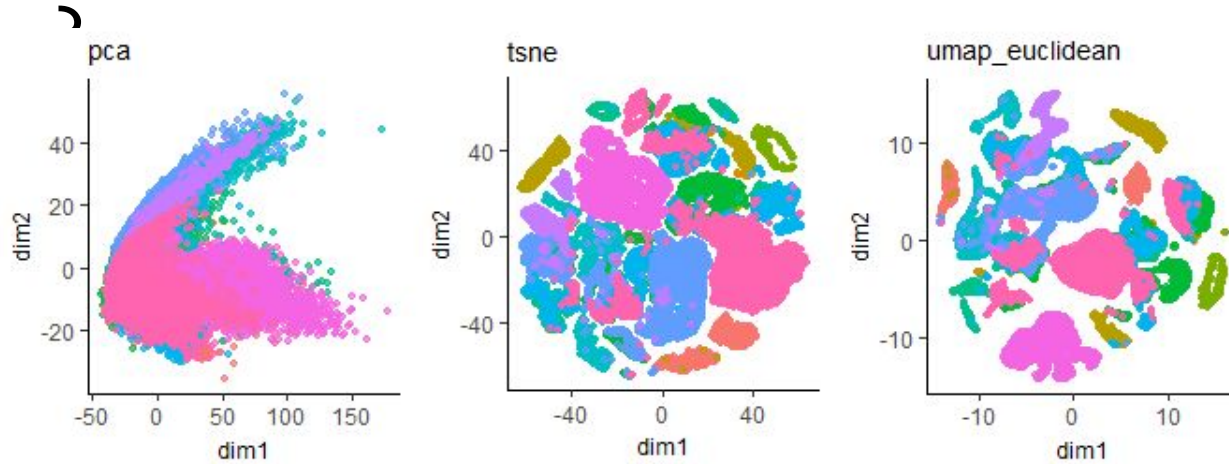
1. Need of an orthogonal space
2. Minimize curse of dimensionality
3. Filter out noise
4. Allow visualization
5. Reduce computational load

Popular methods used for single-cell data analysis:

1. PCA
2. ICA
3. tSNE
4. UMAP
5. Others : Diffusion map, Isomap



# Dimensionality reduction : PCA / tSNE / uMAP

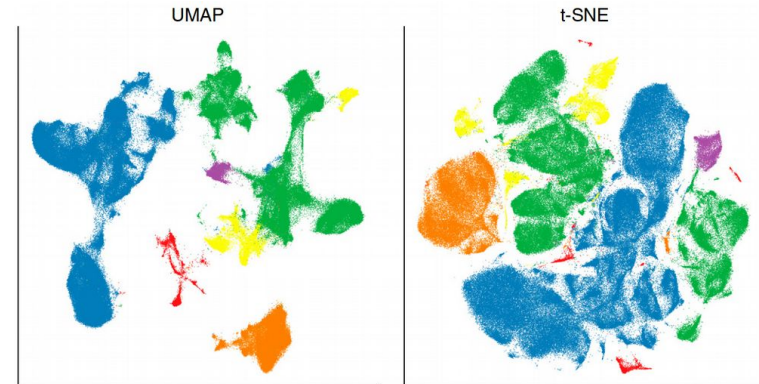


Reduction :

- PCA (on single cell data) is unable to concentrate relationships in 2-3 dimensions only

Visualization : uMAP > tSNE

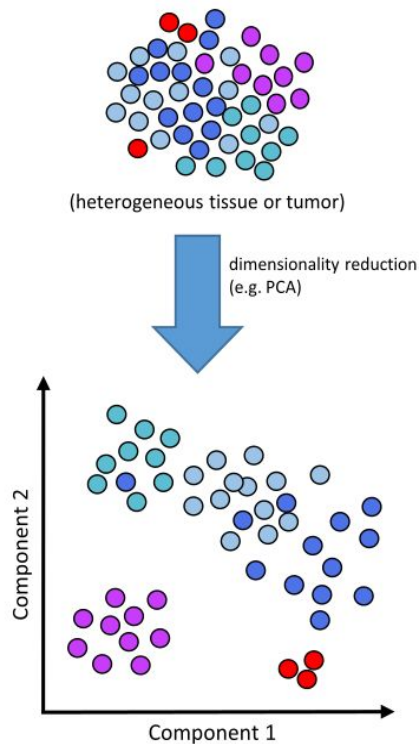
- Better compaction
- Mostly retains inter-cluster distances
  - Subpopulations
  - Trajectory
- More robust to parameter modifications
- (Slightly faster to generate)



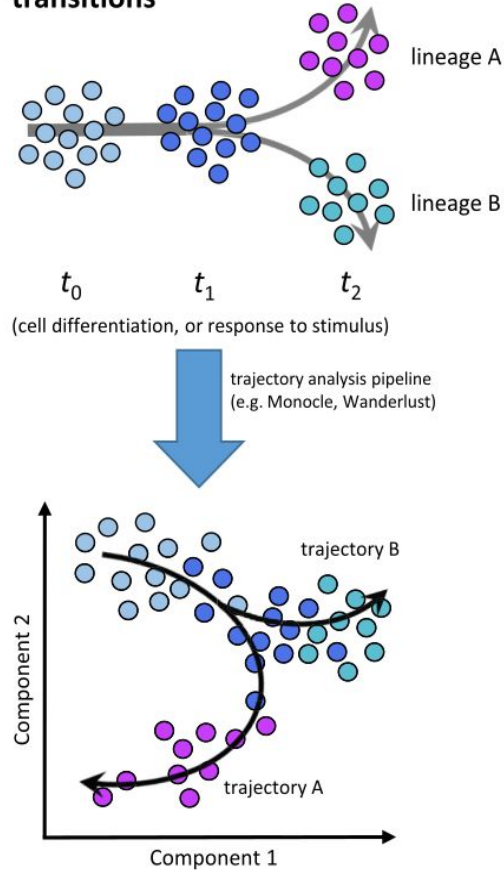
Cell types  
● Contaminant (including B) ● CD4 T ● CD8 T ● MAIT ● NK/ILC ●  $\gamma\delta$  T

*From a reduced space to ...  
... actually what you initially wanted !*

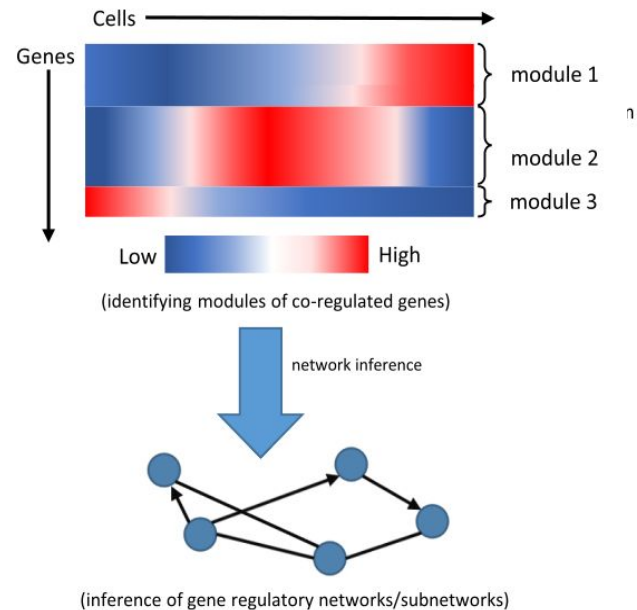
### a) Deconvolving heterogeneous cell populations



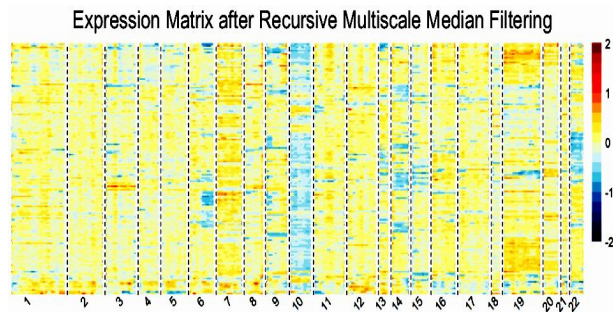
### b) Trajectory analysis of cell state transitions



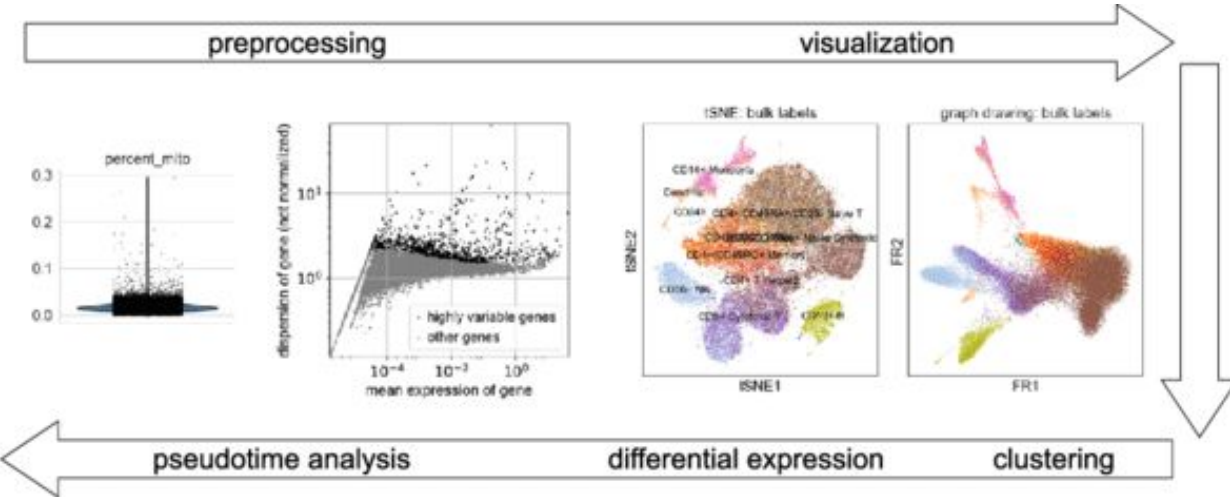
### d) Network inference



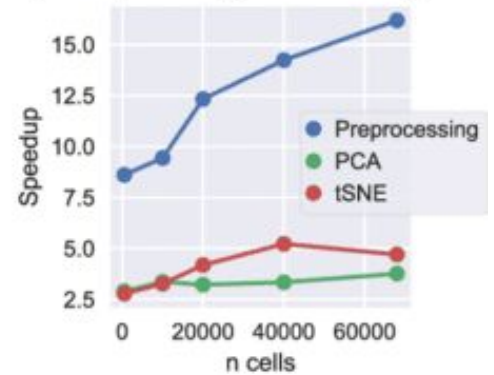
### e) Copy number estimation



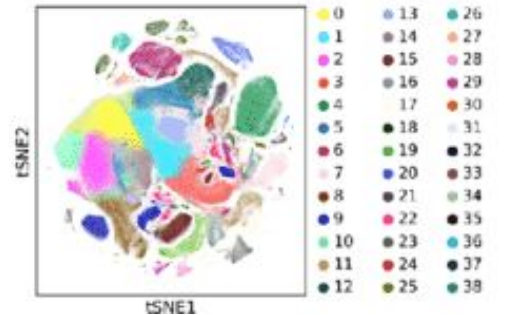
# The all-in-one Python toolbox : Scanpy



**b** Speedup: Scanpy vs. Cell Ranger R



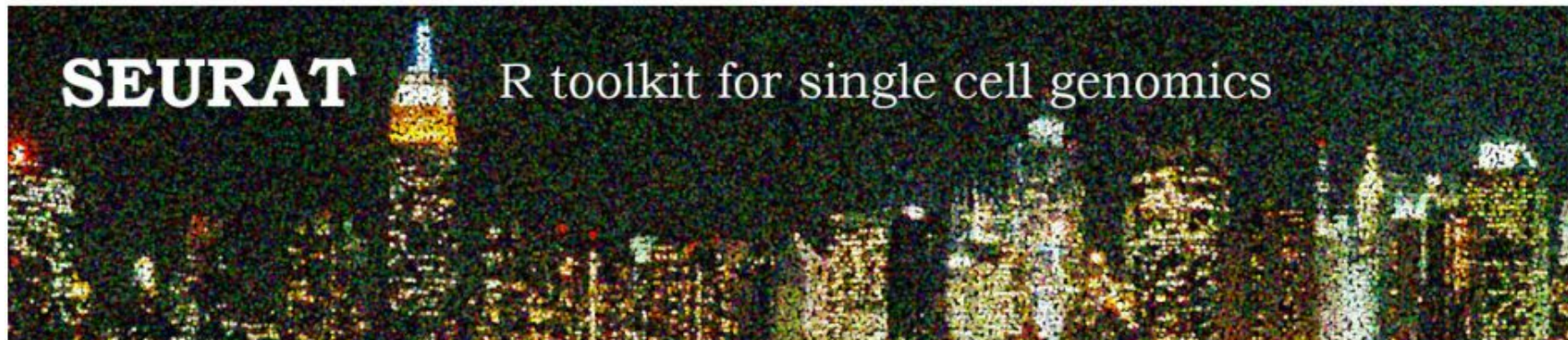
**c** tSNE of clustered 1.3 million cells



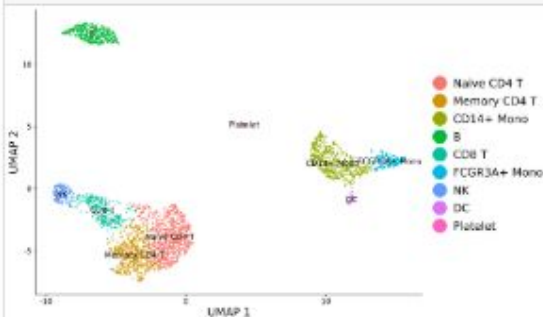
# The all-in-one R toolbox : Seurat (v3)

HOME NEWS PEOPLE RESEARCH PUBLICATIONS SEURAT JOIN/CONTACT

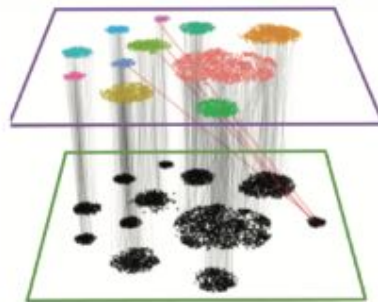
SINGLE CELL  
GENOMICS DAY



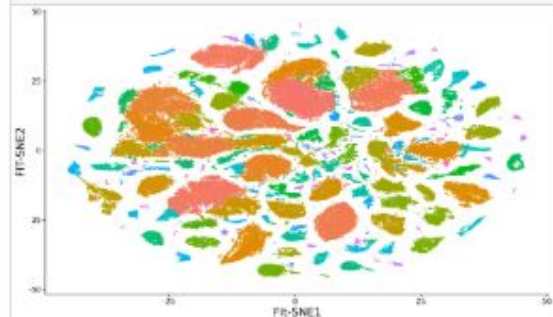
Guided tutorial --- 2,700 PBMCs



Multiple Dataset Integration and Label Transfer

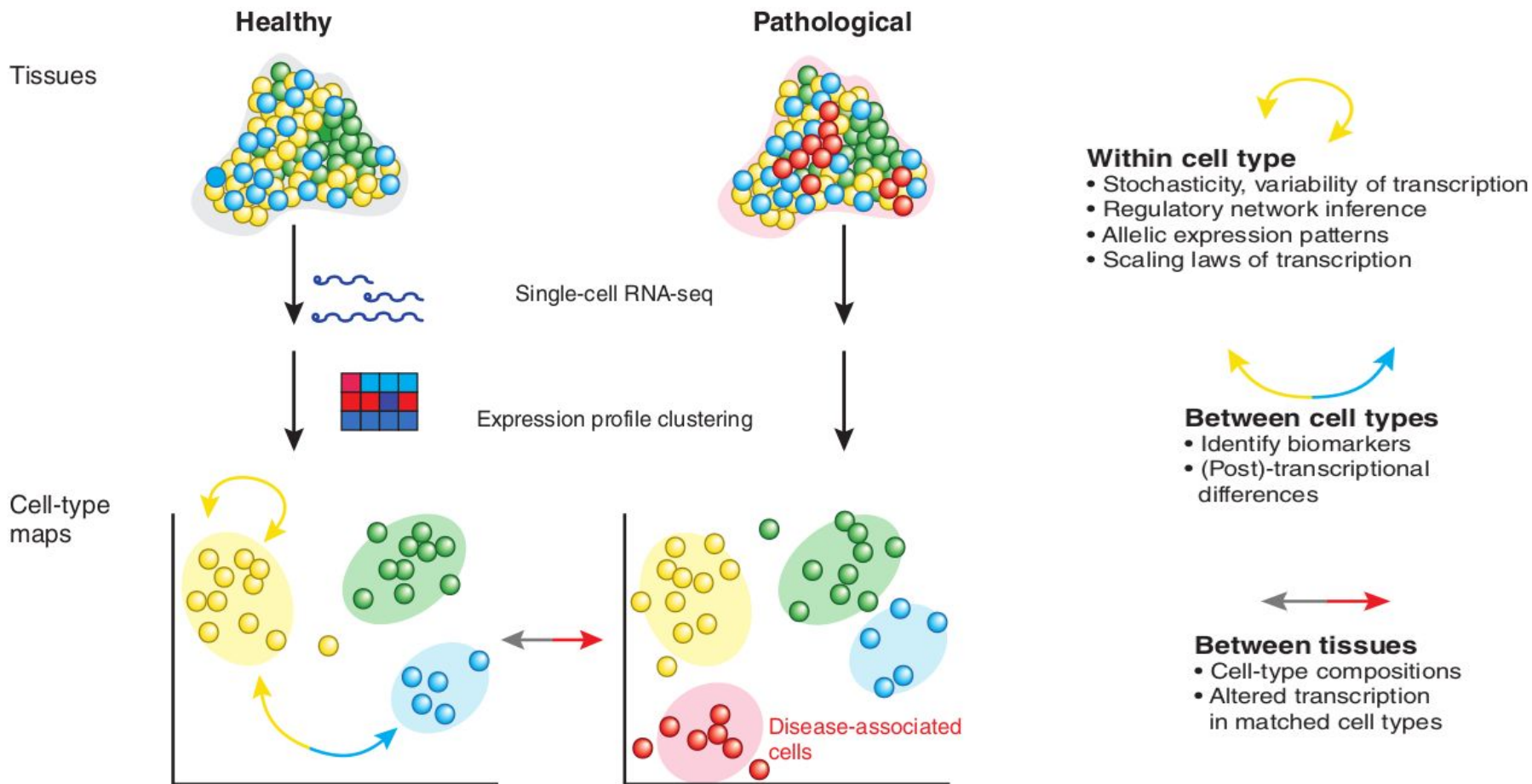


Mouse Cell Atlas, 250K cells



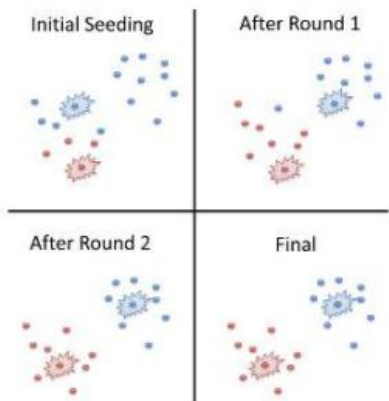


# Cell clustering

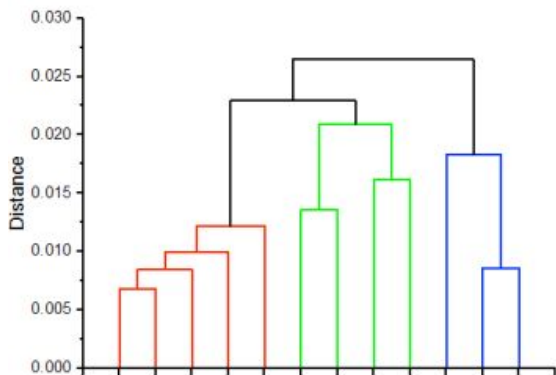


# Cell clustering : methods

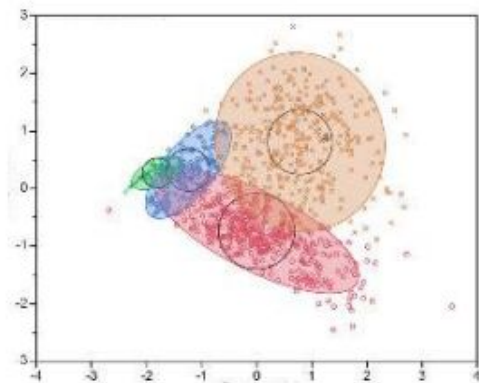
## 1) K-means based



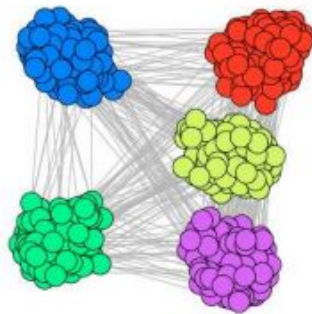
## 2) Hierarchical clustering



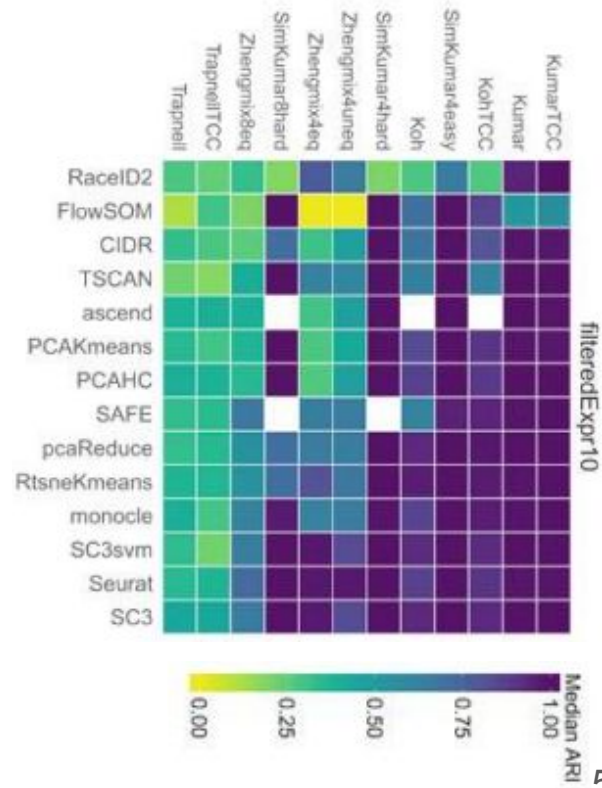
## 3) Model-based clustering (Mclust)



## 4) Graph-based clustering



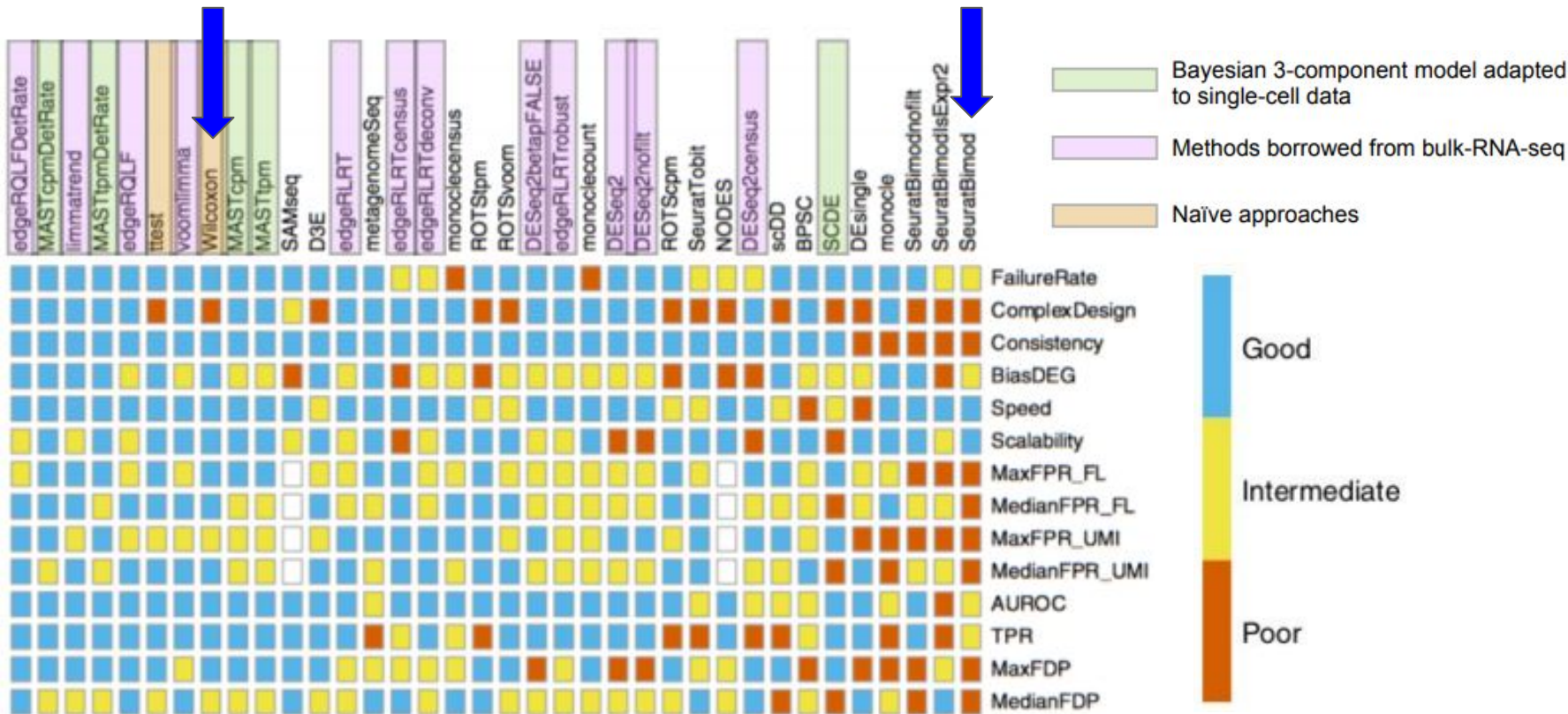
## 5) Single-cell specific methods



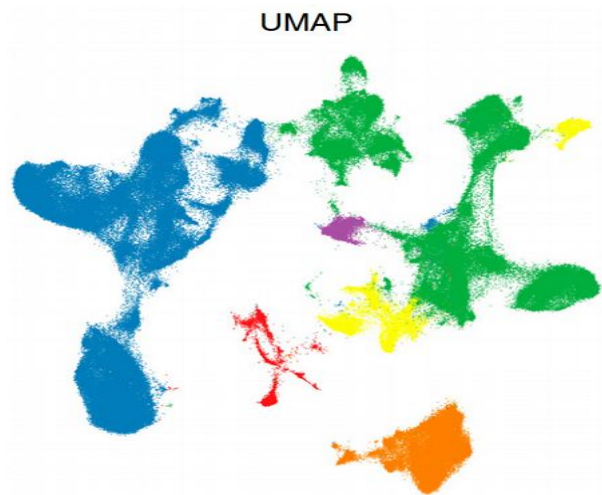
# Differential expression analysis

Seurat v3

Seurat v2



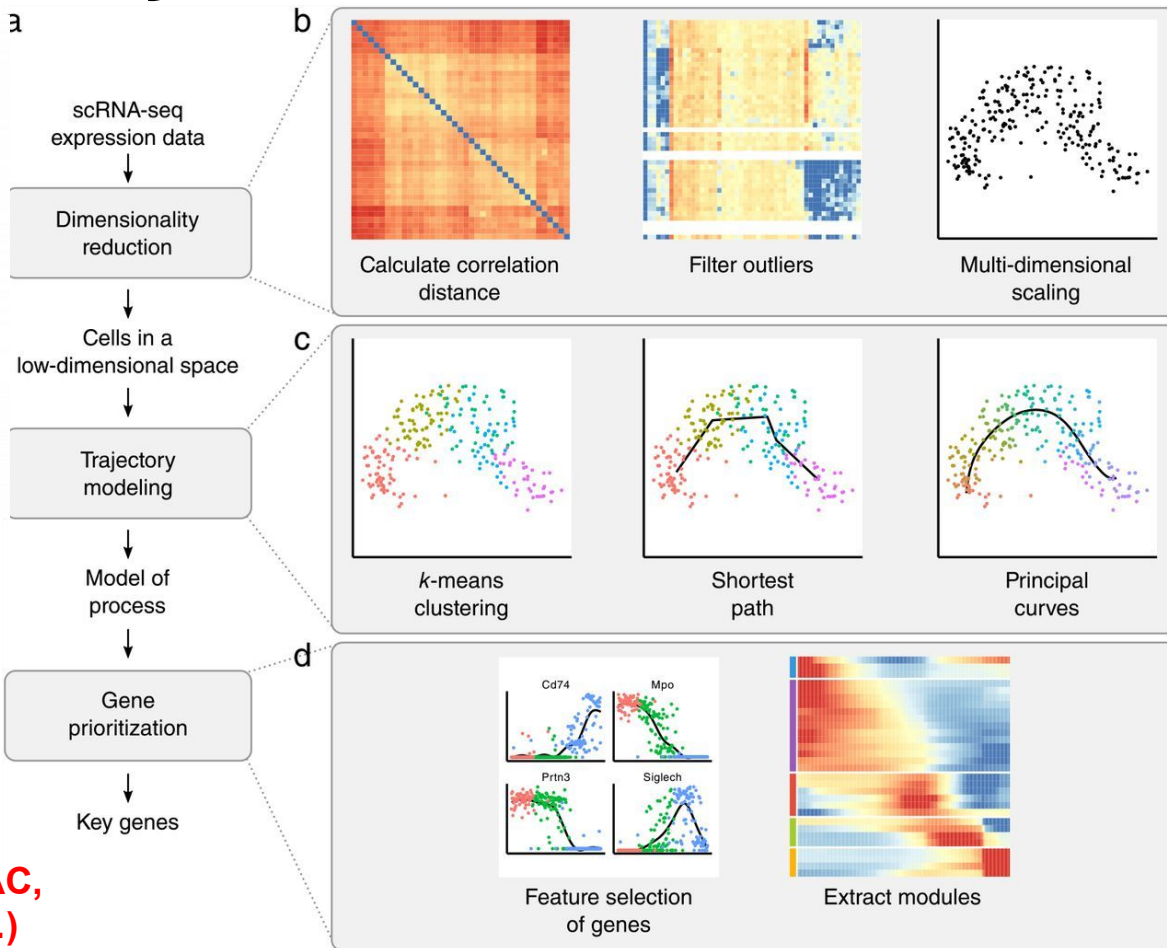
# Cell trajectory : methods



Most adopted tools :

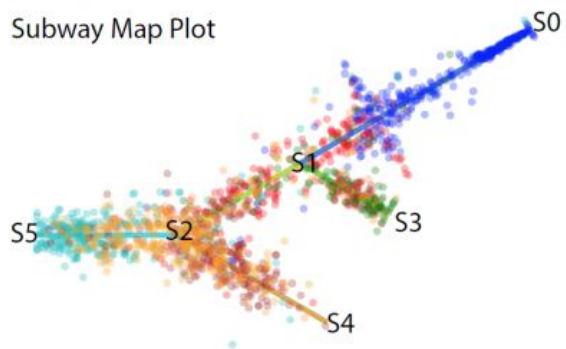
- Monocle 3
- PAGA
- STREAM
- Scorpis
- Slingshot

Not limited to scRNAseq ! (ATAC, CITE, multiomics, imagery-based ...)

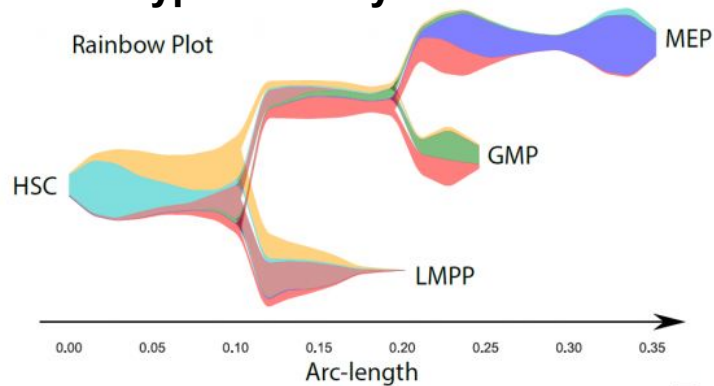


# Cell trajectory : visualization

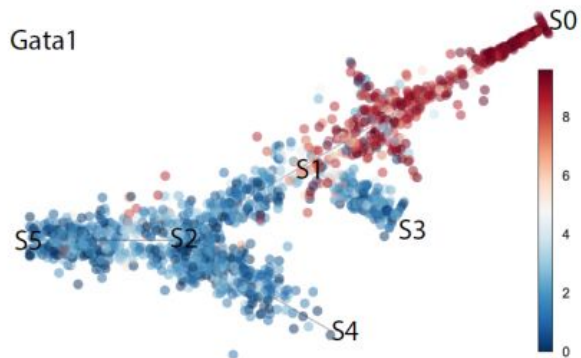
## Cell distance to path + cell types



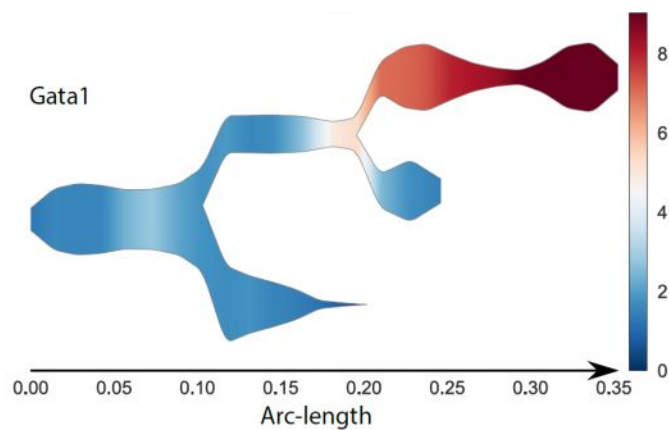
## Cell types density



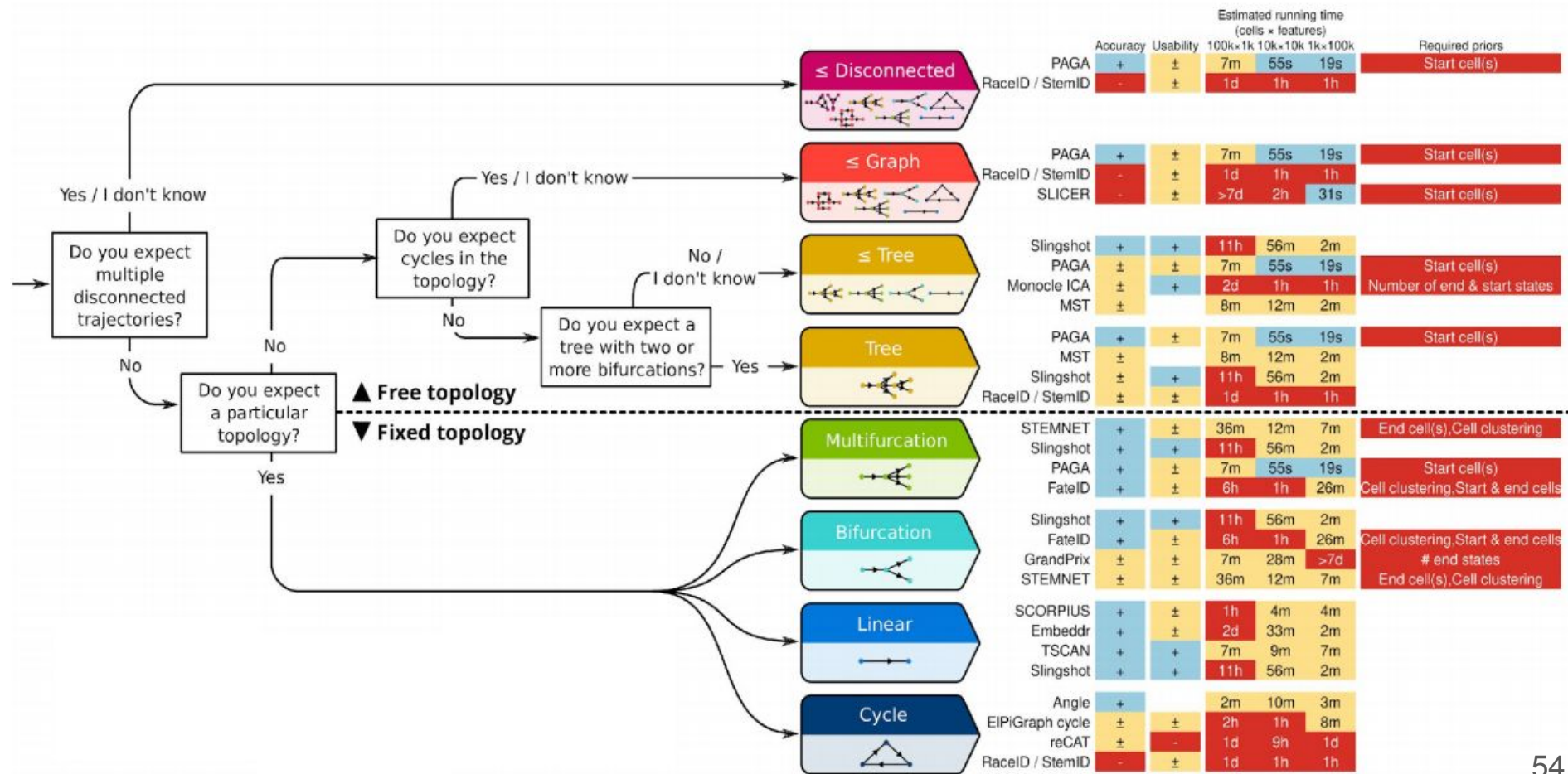
## Cell distance to path + gene expression



## Cells density + gene expression

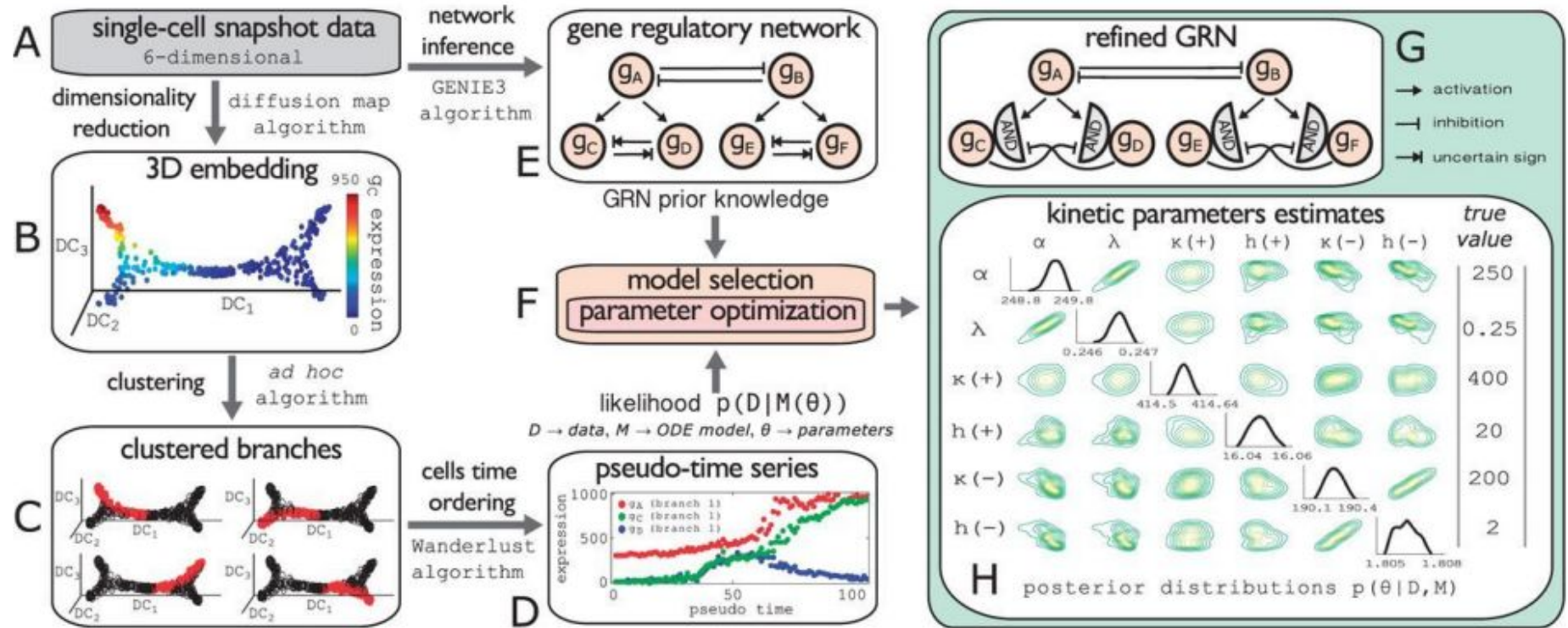


# Cell trajectory : Contexts



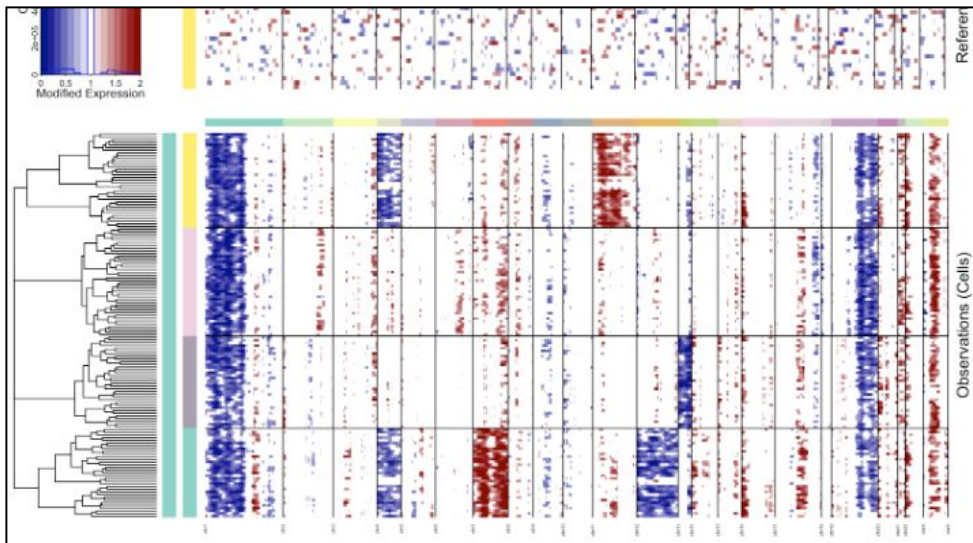
# Network inference

Using cell ordering from trajectory analysis + co-occurring / correlated genes

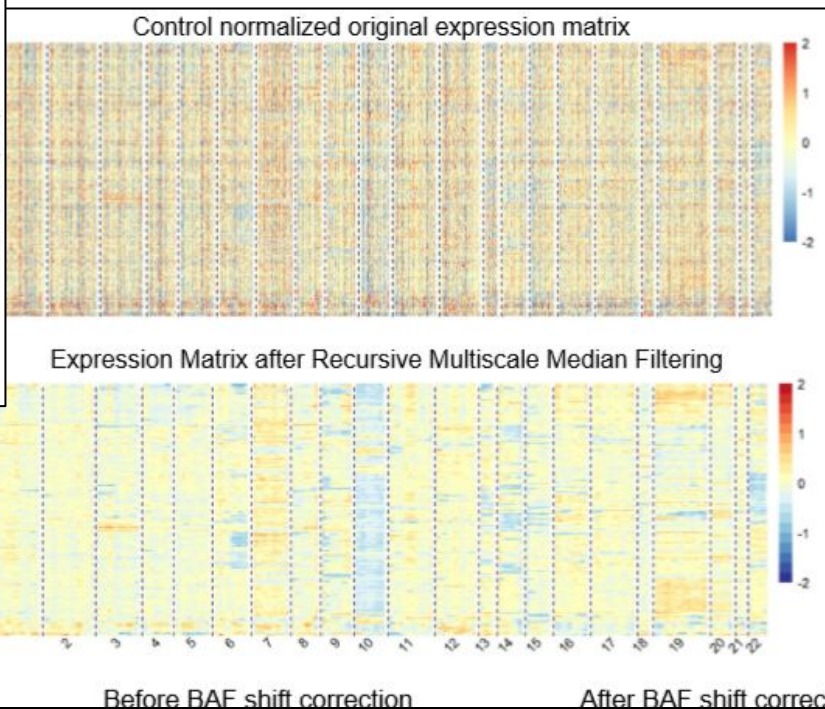


# Copy number estimation from scRNAseq

## InferCNV (Broad Institute)



## CaSpER (Armani et al, BioRxiv 2019)

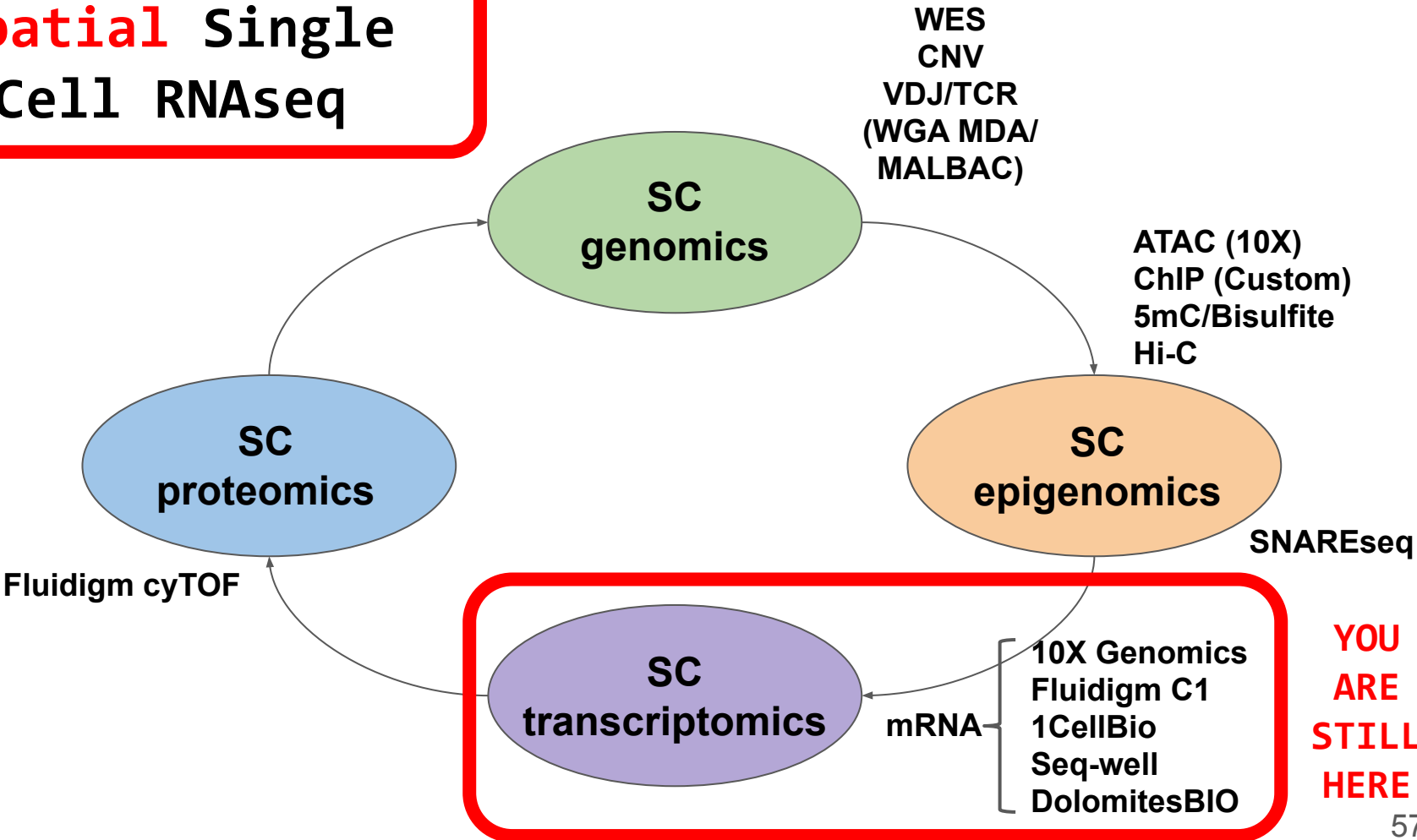


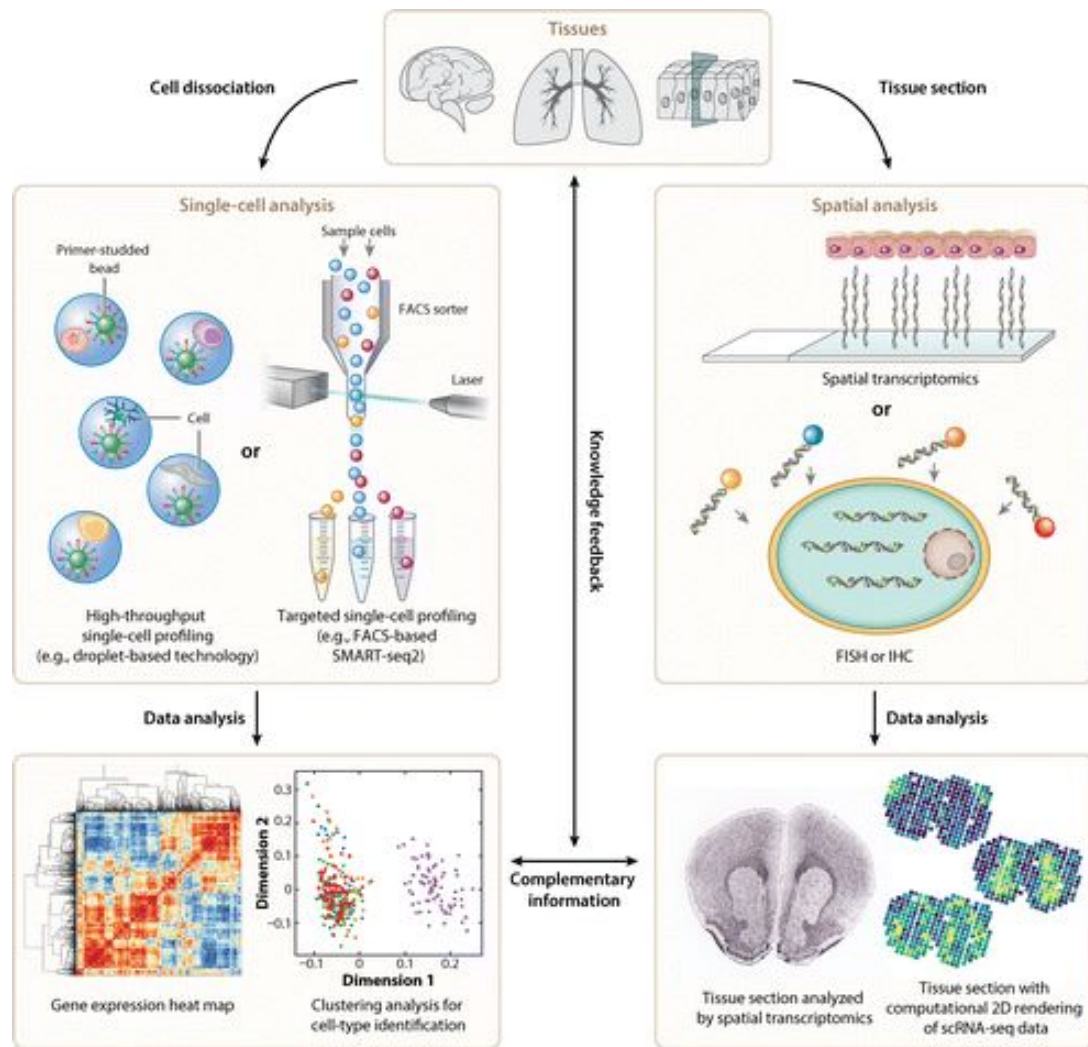
### WARNING :

- Coarse grain ( $> 10$  Mb)
- Requires  $> 75,000$  reads / cell

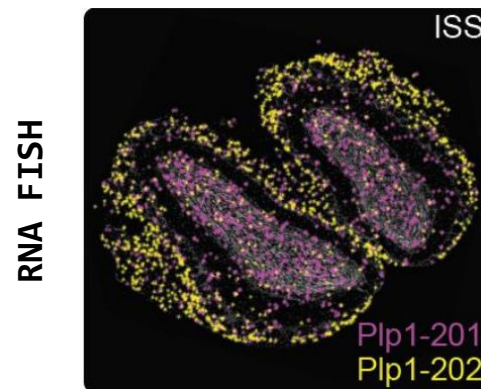
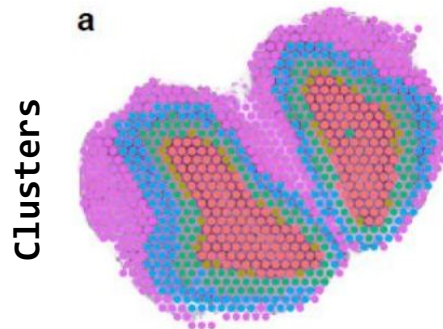
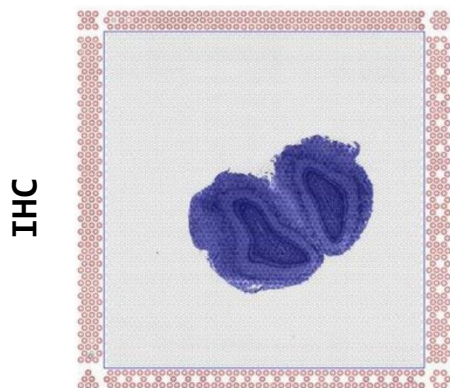
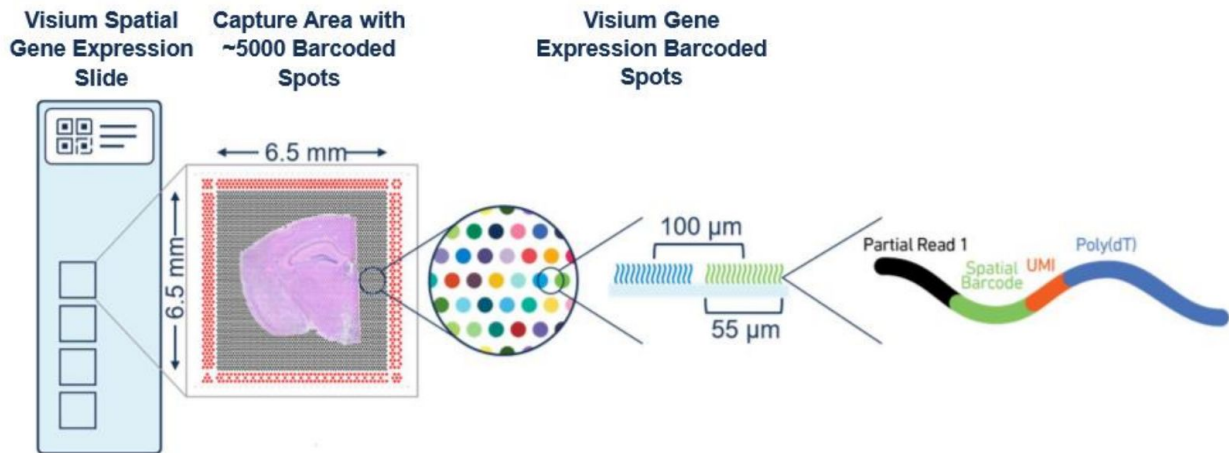


# Spatial Single Cell RNAseq



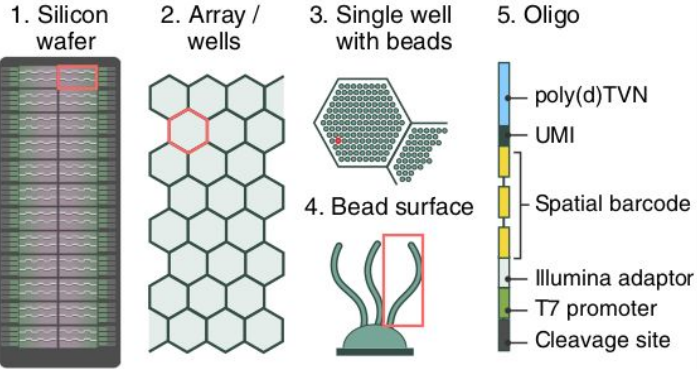


# 10x Genomics Visium

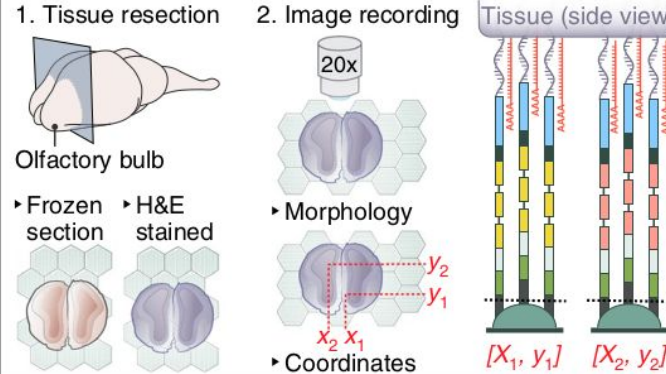


# Illumina “HD Spatial Transcriptomics”

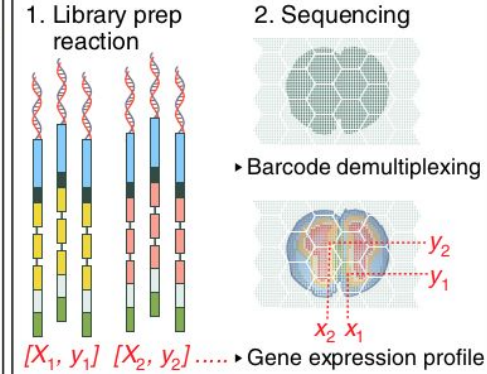
## Methodology



## Sample preparation

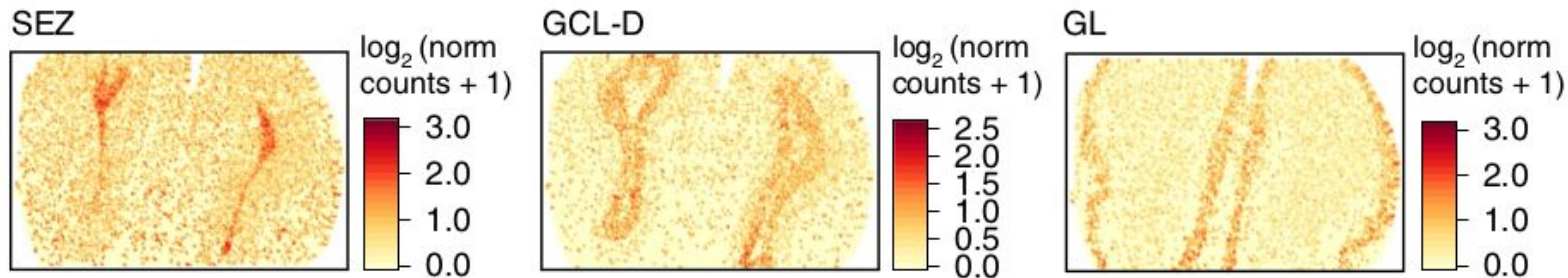


## Sequencing and analysis

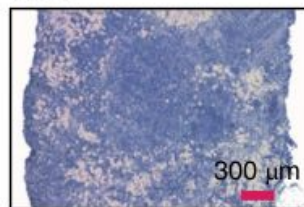


- 2,893,865 individual barcoded beads
- 1,4 M wells
- Well diameter ~ 2  $\mu\text{m}$ 
  - $\ll$  median cell diameter (20  $\mu\text{m}$ )
  - ~ 1,400 x higher resolution than “standard” ST
  - ~ 25 x compared to SLIDE-seq
- Array reading time ~ 3 H
- Challenging analysis strategy (low capture rate) ...
- Commercially available in 2020

# Illumina HDST



H&E



Annotations



- Fatty tissue, immune/lymphoid
- Fatty tissue, invasive cancer
- Fibrous tissue, invasive cancer
- Fibrous tissue, immune/lymphoid
- Invasive cancer, immune/lymphoid
- Immune/lymphoid
- Fatty tissue, fibrous tissue, invasive cancer
- Fibrous tissue
- Fibrous tissue, invasive cancer, immune/lymphoid
- Fatty tissue
- Fatty tissue, fibrous tissue, invasive cancer, immune/lymphoid
- Fatty tissue, invasive cancer, immune/lymphoid
- Invasive cancer

**c**

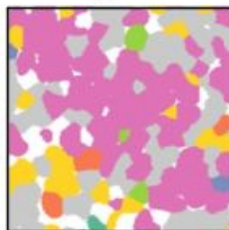
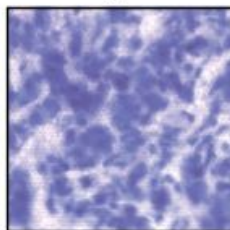
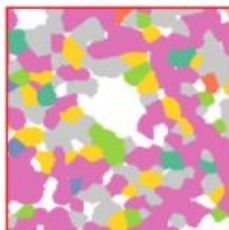
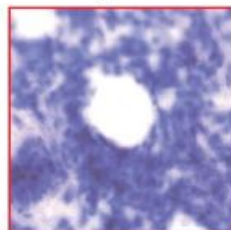
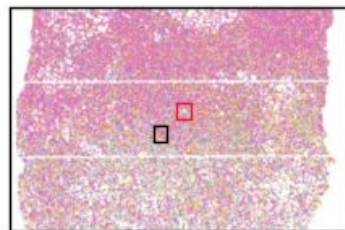
Cell types  
in sn-like data

H&E  
enlargement

sn-like  
enlargement

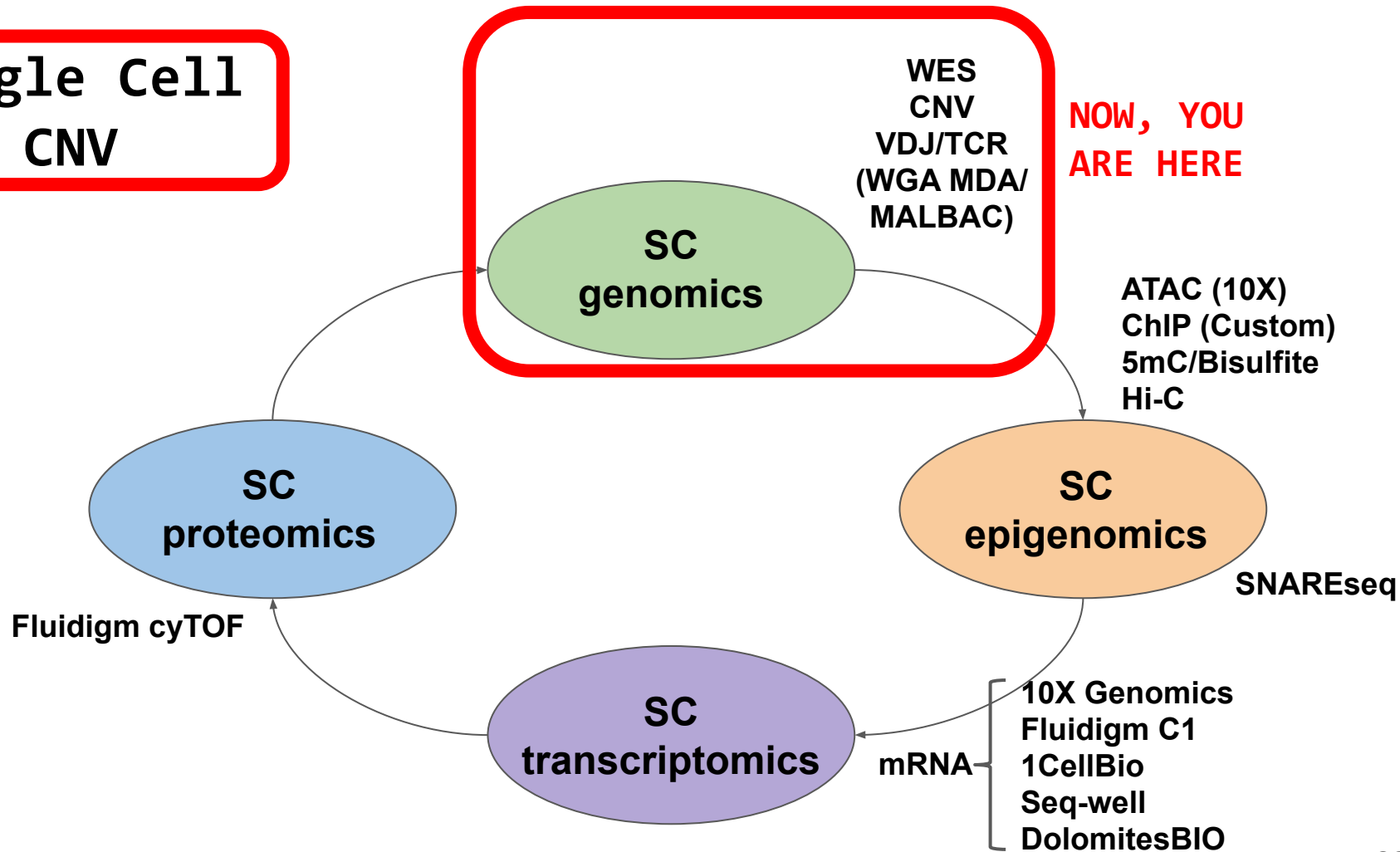
H&E  
enlargement

sn-like  
enlargement

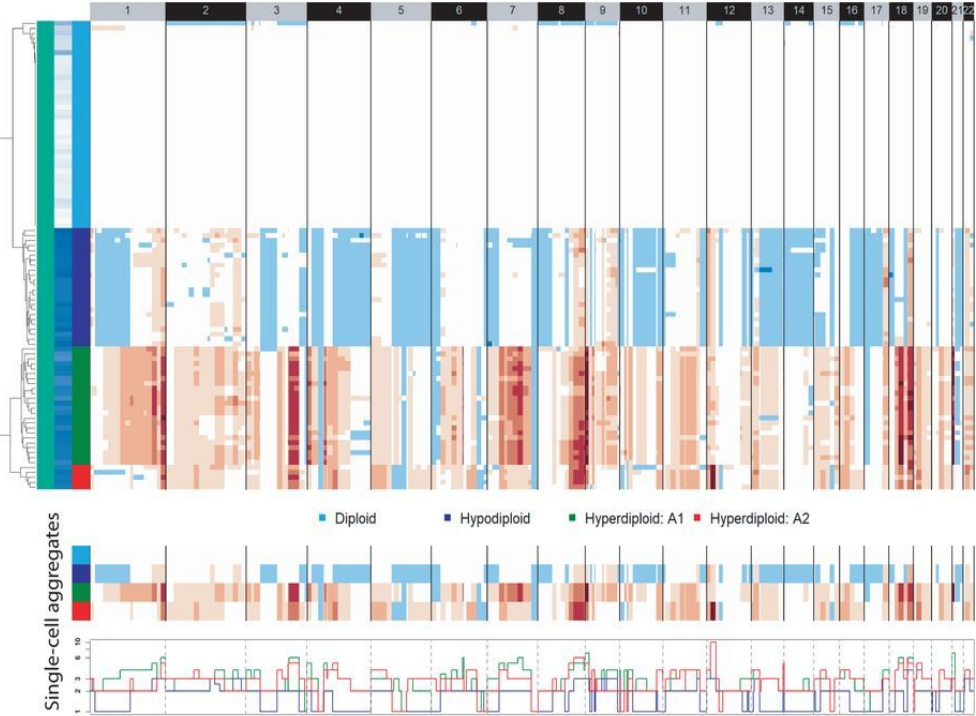
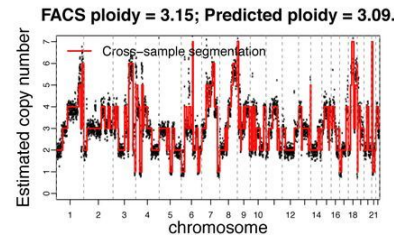
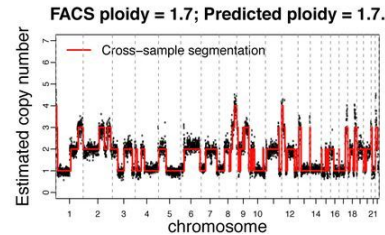
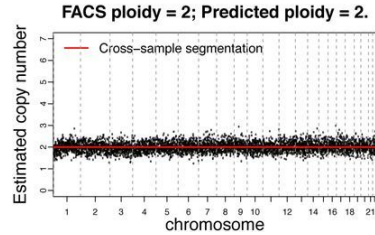
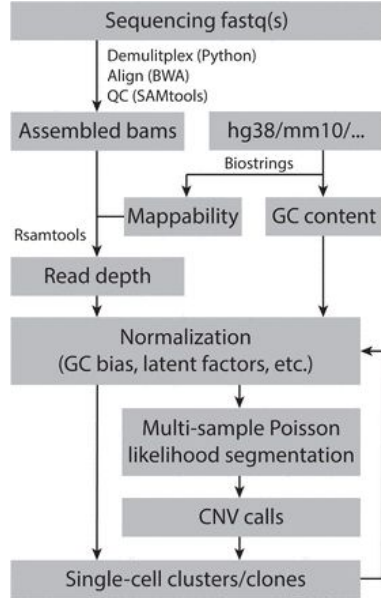


- T cells
- B cells
- Endothelial cells
- Epithelial cells
- Macrophages
- Stroma
- Unassigned nucleus

# Single Cell CNV



# scCNV results (SCOPE)

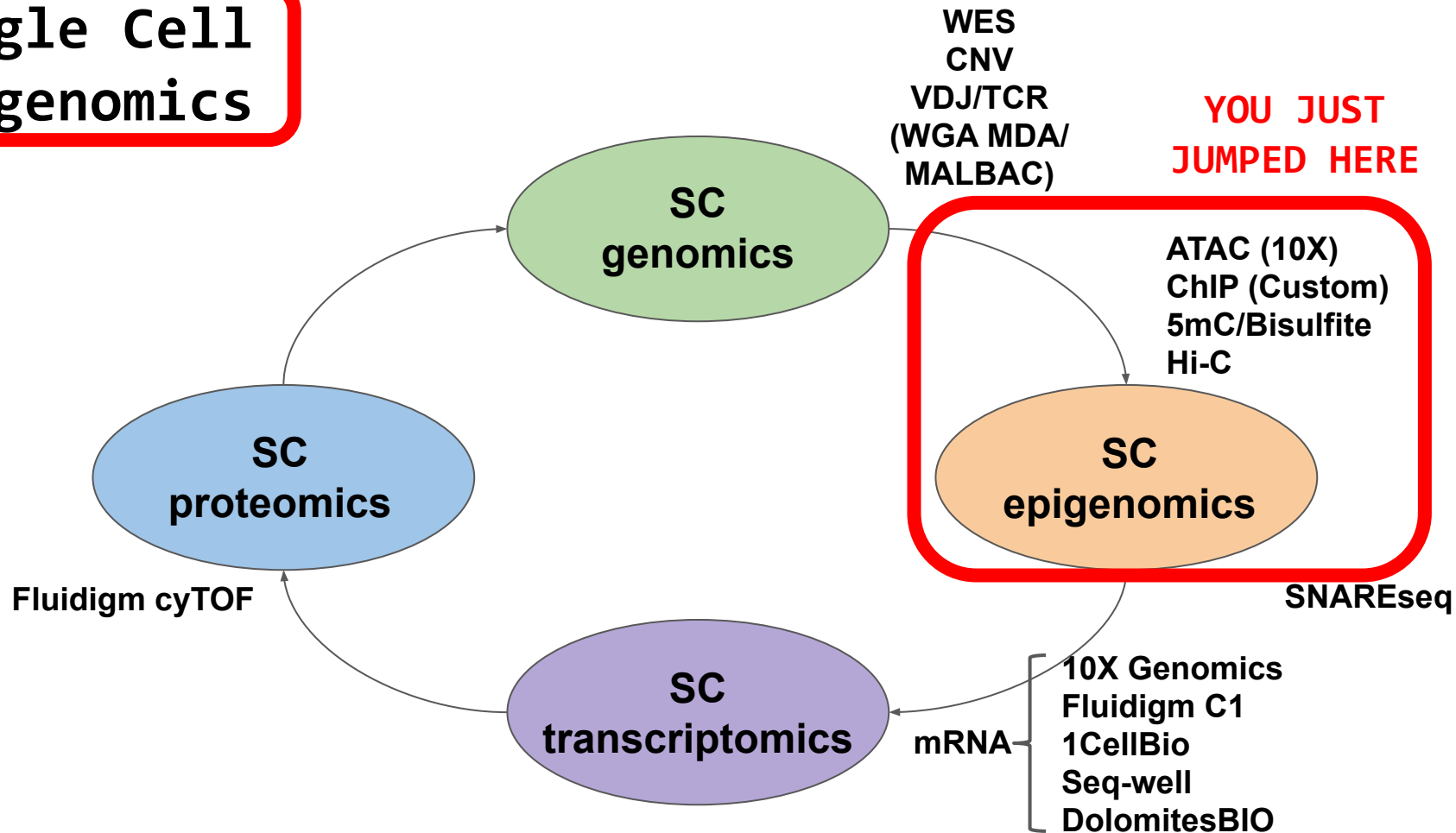


## WARNING :

- Limited resolution : > 2 Mb (binning)
- Requires > 750,000 reads / cell

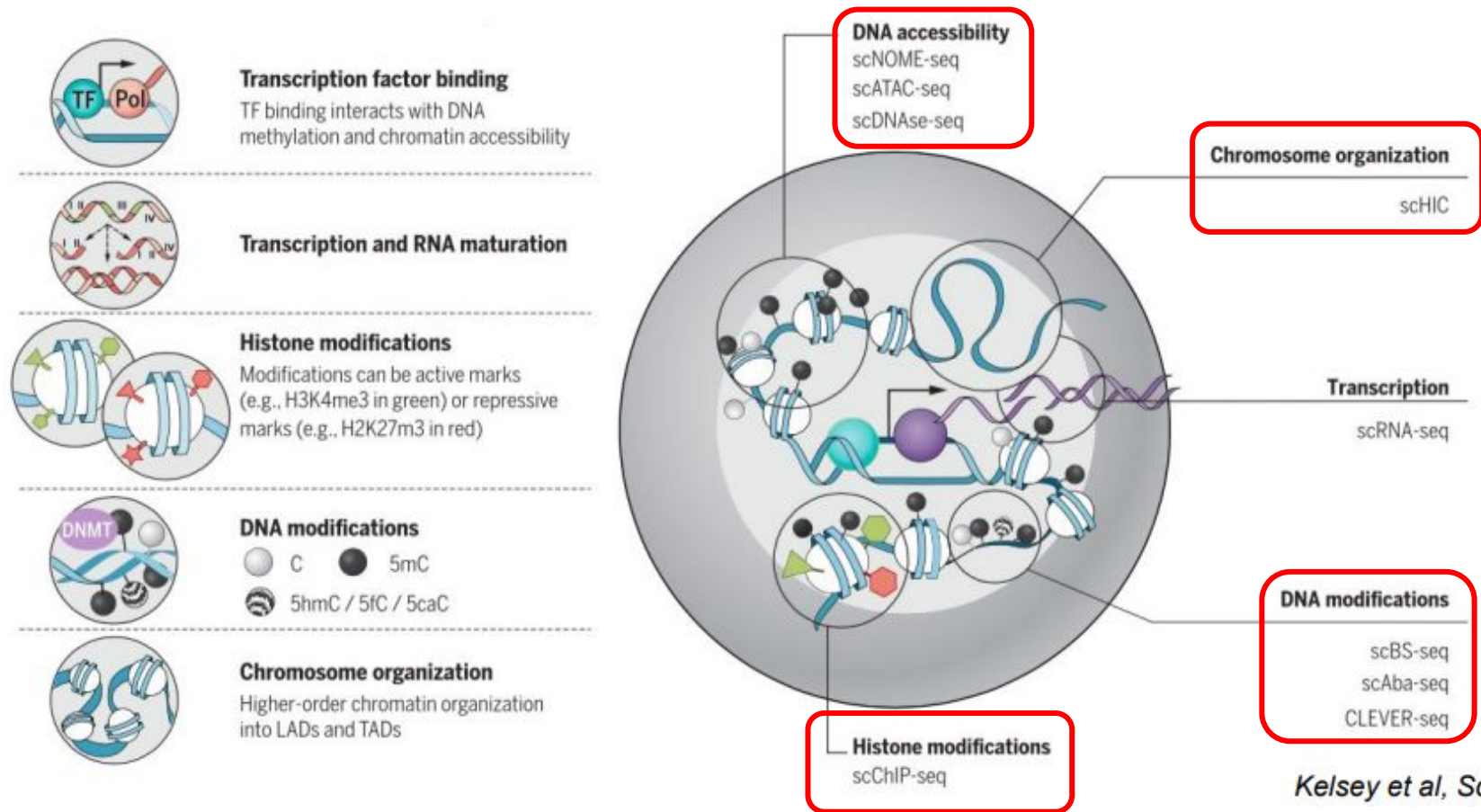
ALSO : SCYN

# Single Cell Epigenomics

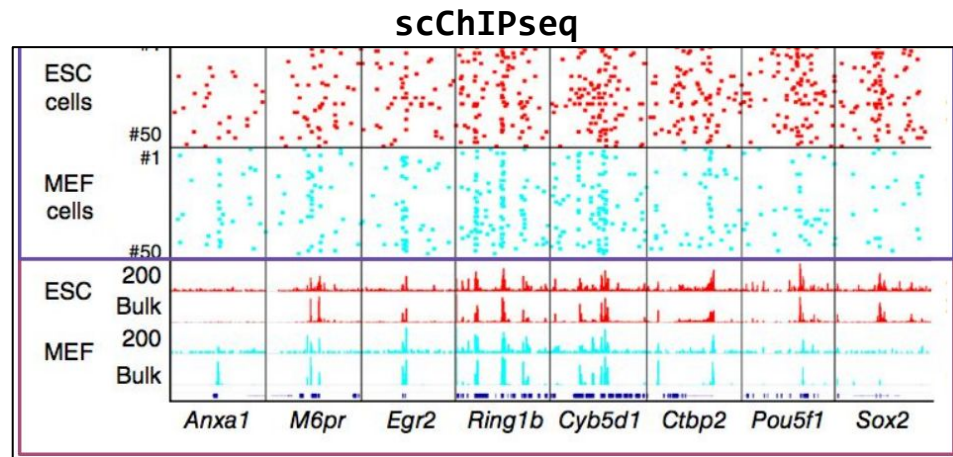
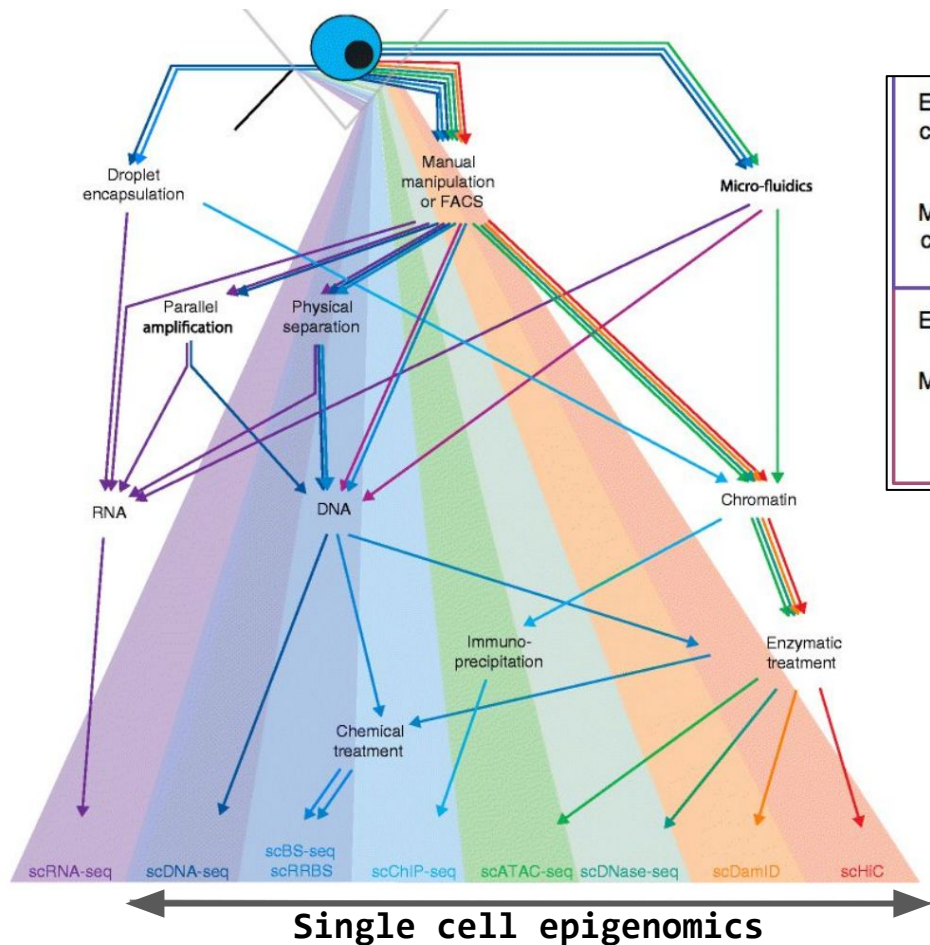




# Overview of scEpigenomics techniques



# Overview of scEpigenomics techniques



- scChIP : improvements in 2019
- scMeth : low coverage, low sensitivity (<20% CpG read)
- scHi-C : stable protocol & analysis still needed
- scATAC : most popular technology, numerous tools available

***Single Cell (RNAseq) Resources***  
*(some)*

# Tabula Muris

## ARTICLE

<https://doi.org/10.1038/s41586-018-0590-4>

## Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

The Tabula Muris Consortium\*

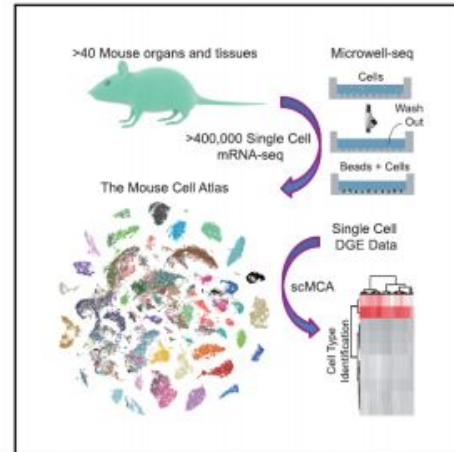
- ~100k cells
- 20 organs
- 2 techniques :
  - Droplet 3', short reads
  - FACS, long reads

Resource

Cell

### Mapping the Mouse Cell Atlas by Microwell-Seq

Graphical Abstract



Authors

Xiaoping Han, Renying Wang,  
Yincong Zhou, ..., Guo-Cheng Yuan,  
Ming Chen, Guoji Guo

Correspondence

xhan@zju.edu.cn (X.H.),  
ggj@zju.edu.cn (G.G.)

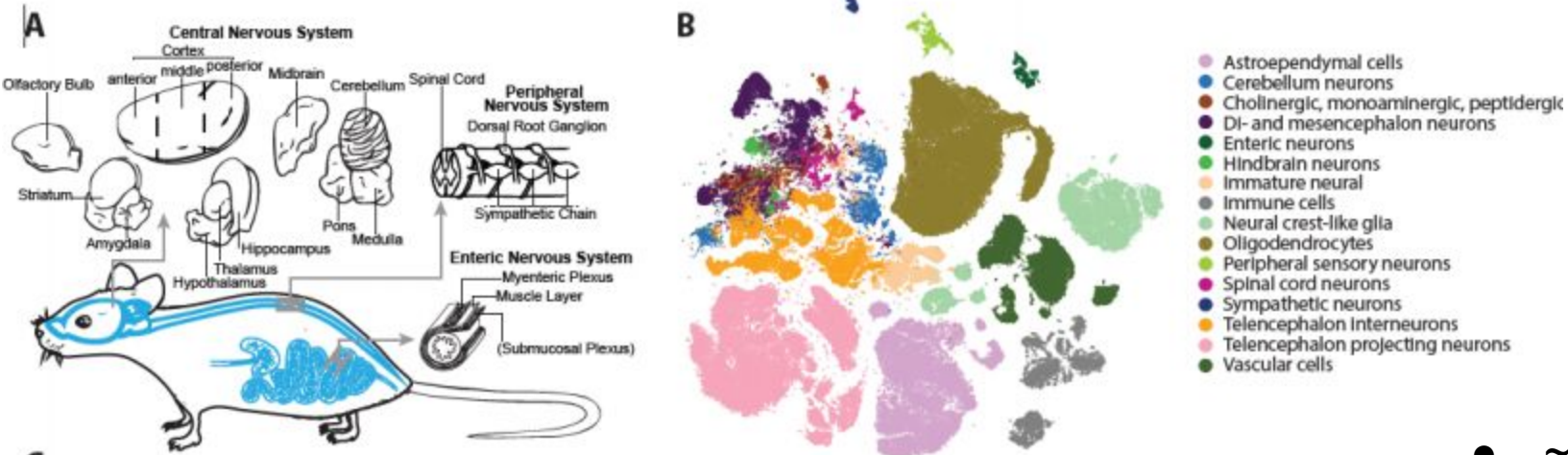
In Brief

Development of Microwell-seq allows construction of a mouse cell atlas at the single-cell level with a high-throughput and low-cost platform.

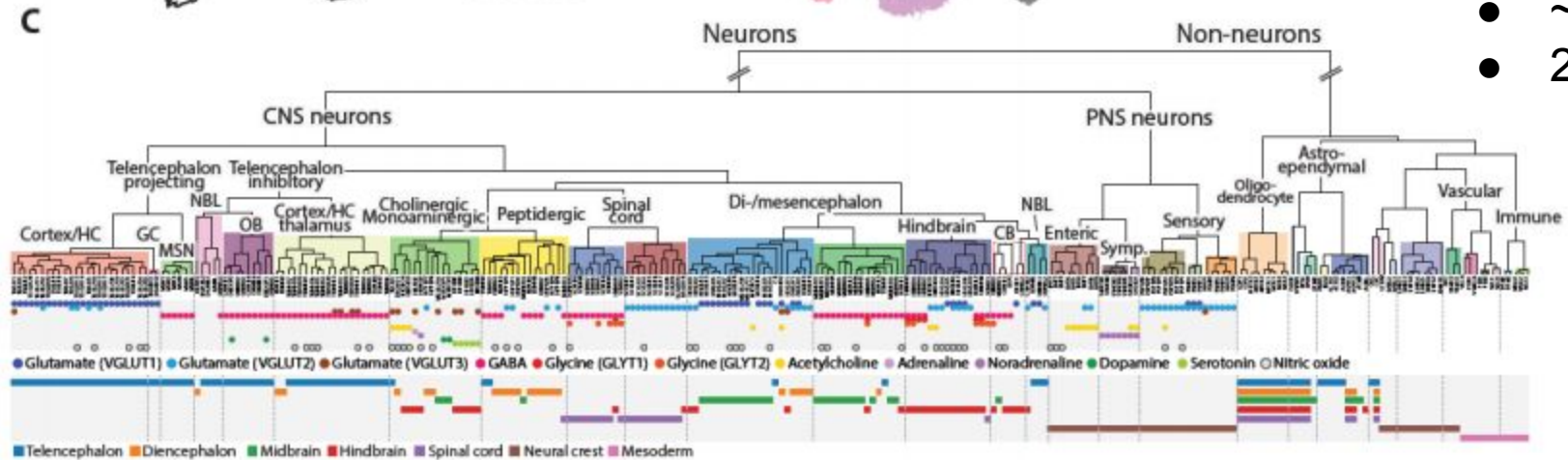
MCA browser

<http://bis.zju.edu.cn/MCA/>

# The Mouse Brain Atlas ([mousebrain.org](http://mousebrain.org))



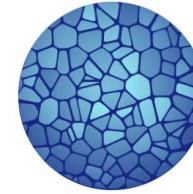
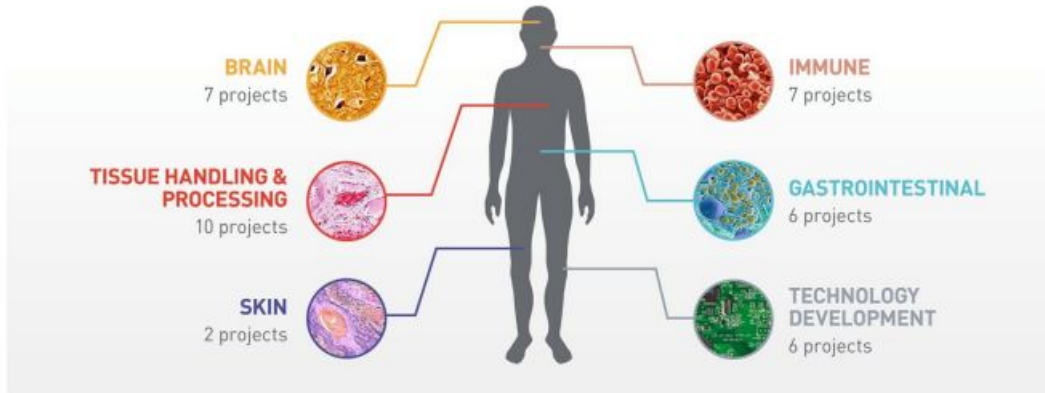
- ~160k cells
- 20 subtypes



# The Human Cell Atlas ([humancellatlas.org](http://humancellatlas.org))

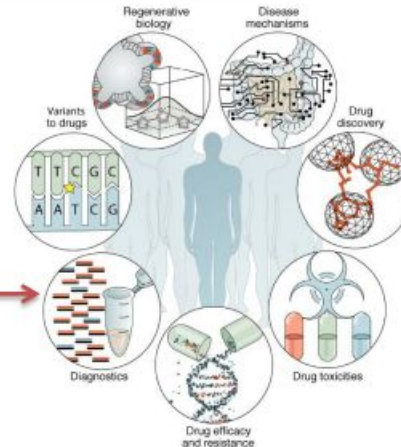
## MAPPING THE BASIC UNITS OF LIFE

CZI proudly supports **38 new projects** in these six areas for the Human Cell Atlas.



**HUMAN  
CELL  
ATLAS**

- Every cell type in the body
- First: define how to proceed
  - Best experimental practice / organ
  - Best bioinformatics methods
- Data will be made available to all



# PanglaoDB

PanglaoDB is a database for the scientific community interested in exploration of single cell RNA sequencing experiments from mouse and human. We collect and integrate data from multiple studies and present them through a unified framework.

## Usage examples

- Run a gene search for [SOX2](#), [PECAM1](#) or [ACE2](#)
- Browse the full list of [samples](#)
- Explore the list of cell type markers for [Schwann cells](#)
- Browse cell types of the mouse [retina](#)
- Look at the expression of [CRX](#) in photoreceptor cells
- Find cell clusters where [both](#) [PECAM1](#) and [VCAM1](#) are expressed using a [boolean search](#) with the 'and' operator
- Find [quiescent neural stem cells](#) using AND+NOT

## How to cite

Oscar Franzén, Li-Ming Gan, Johan L M Björkegren, *PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data*, **Database**, Volume 2019, 2019, baz046, doi:10.1093/database/baz046

## What is single cell RNA sequencing?

Adapted from the [Wikipedia](#) article on the topic: *Single cell RNA sequencing examines the transcriptomes from individual cells with*

## Database statistics

	<i>Mus musculus</i>	<i>Homo sapiens</i>
Samples	1063	305
Tissues ?	184	74
Cells ?	4,459,768	1,126,580
Clusters ?	8,651	1,748

## Dataset of the day

Take a closer look at the cellular composition of [Subventricular zone](#), using a dataset which consists of 1150 cells. Clustering of this dataset resulted in 8 cell clusters, containing among others, [Endothelial cells](#).

## News

21-05-2020 Ongoing work to move to new hosting.

30-01-2020 A corrupted MySQL table caused dysfunction in the search function, the problem has now been fixed.

***WYSIWYG Analysis Frameworks***  
*(mainly for scRNAseq)*



# SCHNAPPS : A R-shiny app for biologists



**SCHNAPPS**

**Input**

Parameters

General QC

Cell selection

Gene selection

Co-expression

Data Exploration

Expression

Panel plot

Subcluster analysis

Summary statistics of this dataset:

scEx.RData\_  
No. of cells: 200  
No. of genes: 958  
Median UMIs per cell: 76.5  
Total number of reads: 14930  
Memory used: 493 Mb  
Normalization used:  
DE\_logNormalization

Generate report

Download counts.csv

Download Rds

**Information:**

- **Clustering:** Clustering was performed with t-SNE followed by identification using DBSCAN
- **Cluster 0:** Cells that cannot be assigned to any cluster
- **3D Plot:** Enter gene name to visualize expression in a single cell
- **2D Plot:** Pick a cluster, highlight cells of interest to download gene expression matrix

Enter gene:

comma separated list of genes for UmiCountPerGenes:

comma separated list of genes for UmiCountPerGenes2:

X:  Y:  color:

!-Cluster1 CD52-Cluster2 CD52-Cluster3 CD52-Cluster4

show more options

CD52

Expression

Cluster

Cluster	cells
0	4
1	32
2	19
3	30
4	72
5	26

**SCHNAPPS**

**input**

Parameters

General QC

Cell selection

Gene selection

Co-expression

Data Exploration

Subcluster analysis

DGE analysis

Summary statistics of this dataset:

scEx.RData\_  
No. of cells: 200  
No. of genes: 958  
Median UMIs per cell: 76.5  
Total number of reads: 14930  
Memory used: 493 Mb  
Normalization used:  
DE\_logNormalization

Generate report

Download counts.csv

Download Rds

- **Subclustering:** Select a group of cells in plot1 and a different group of cells in plot2 for identifying differential features between these subclusters
- **colors:** colored by cluster identity
- **selection hint:** you can also select by groups you have defined in other plots.
- **selection hint:** also check out "Gene.count" to verify that number genes per cell.

Cluster:  X:  Y:

tsne2

tsne1

tsne2

tsne1

**Method to use for differential gene expression analysis**

**Method to use**

Chi-square test of an estimated binomial distribution

t-test

**Differentially Expressed Genes**

Selected items to be copied

Download table

**Cells**

Select all rows

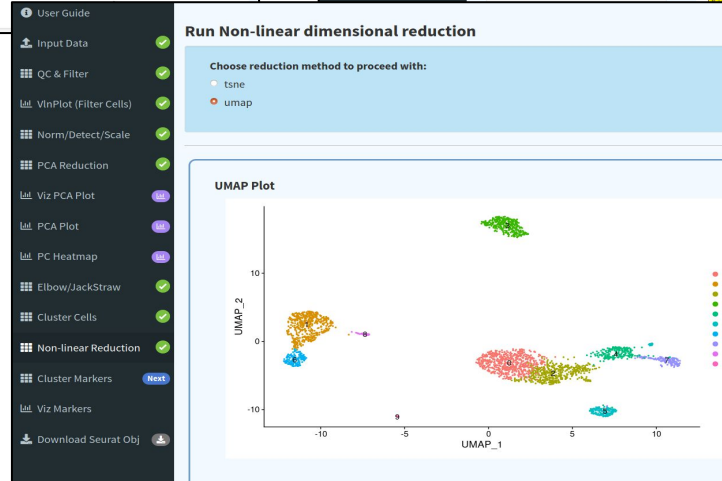
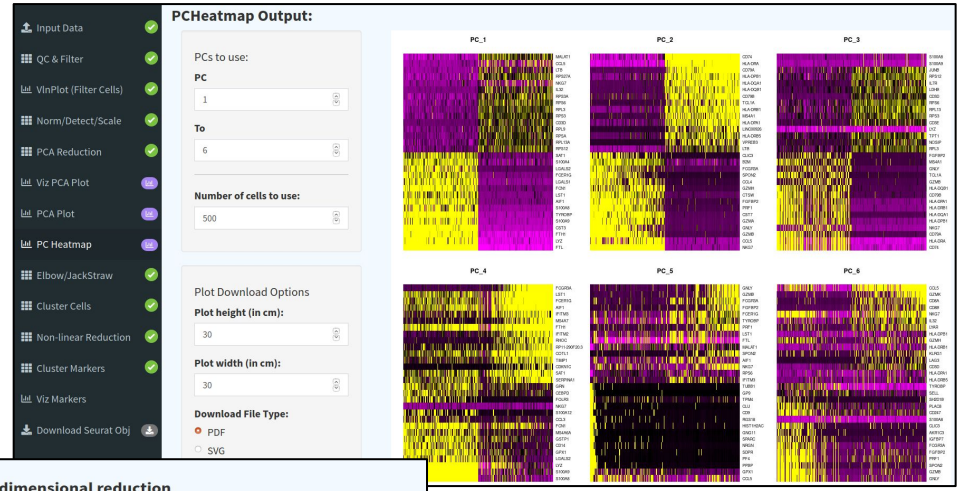
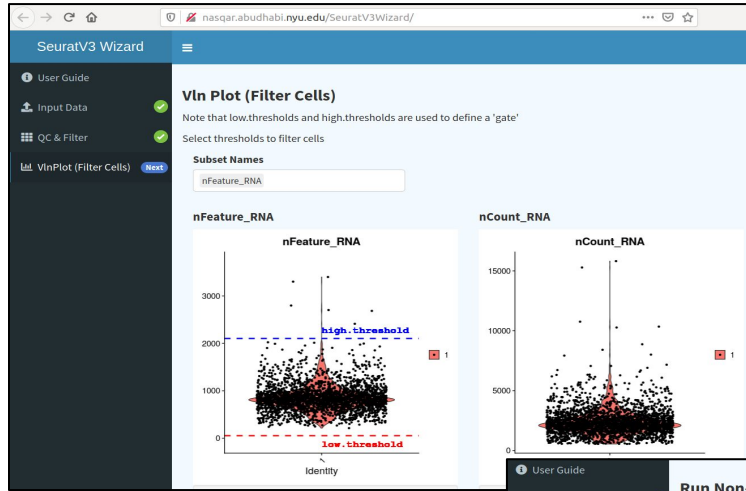
reorder cells by sum of selected genes

By Bernd Jagla (Pasteur Paris)

<https://c3bi-pasteur-fr.github.io/UTechSCB-SCHNAPPS>

<https://github.com/C3BI-pasteur-fr/UTechSCB-SCHNAPPS>

# SeuratV3Wizard

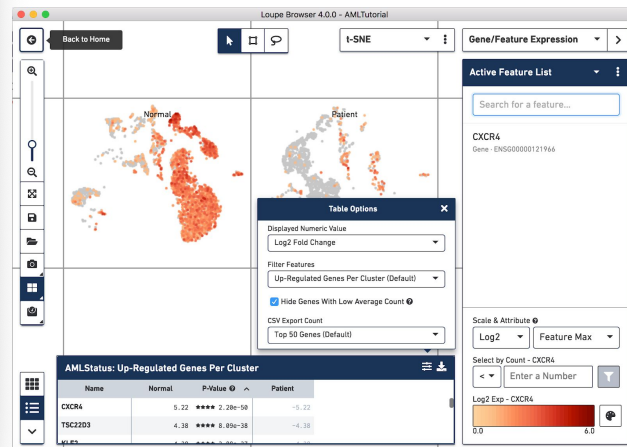


***Visualization tools***  
*(mainly for scRNAseq)*

# 10x Genomics Loupe Browser



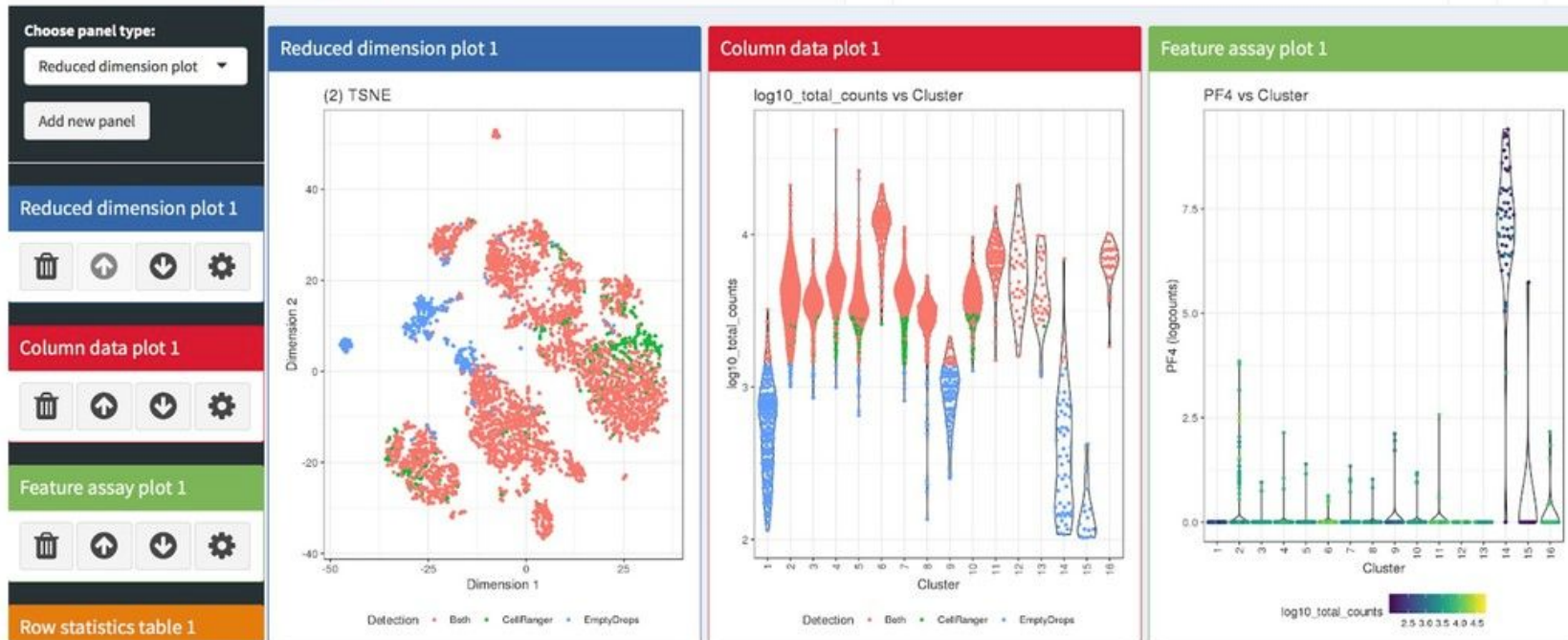
- Compatible with output from 10x Cell Ranger (“cloupe” files)
- Linux / OSX
- Supports Visium (Spatial)



## Interactive Data Visualization (iSEE)

**Creators:** Federico Marini,  
Aaron Lun, Charlotte Soneson,  
and Kevin Rue-Albrecht

Running emptyDrops on the PBMC 4K dataset



[https://marionilab.cruk.cam.ac.uk/iSEE\\_pbmc4k/](https://marionilab.cruk.cam.ac.uk/iSEE_pbmc4k/)

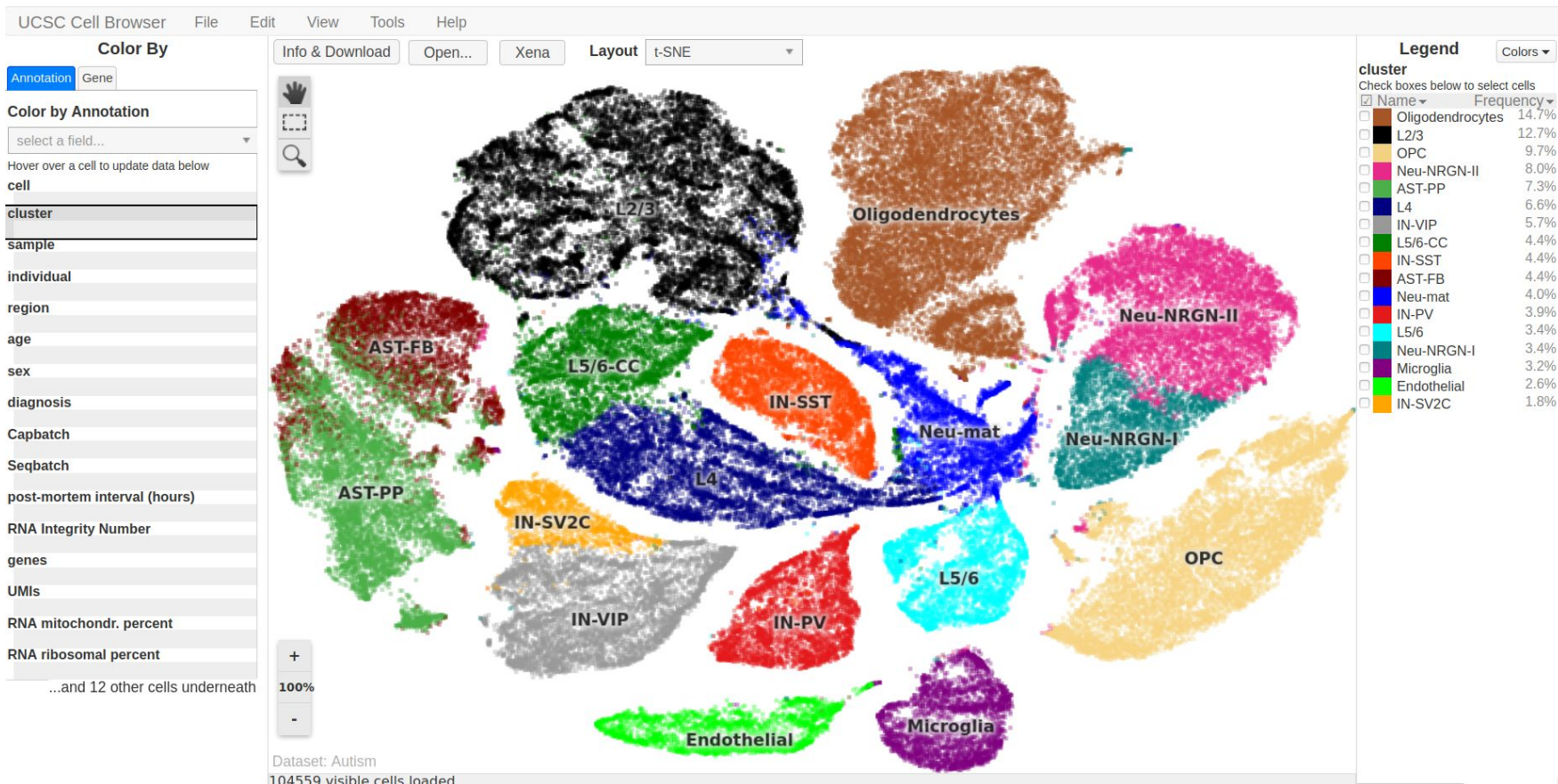
# CerebroApp



- ShinyApp (web GUI over some R)
- Binary format (CRB), converted from SeuratObject / SCE
- From QC to trajectory

*(my favorite one)*

# UCSC Cell Browser



# Acknowledgements

Marc Deloger  
Morgane Thomas-Chollier  
Agnès Paquet  
Marine Aglave  
Antonio Rausell  
Wouter Saelens  
Nathalie Gaspar



SINGLe-cellING in the RAINaseq (1952)

© Jacques van Helden



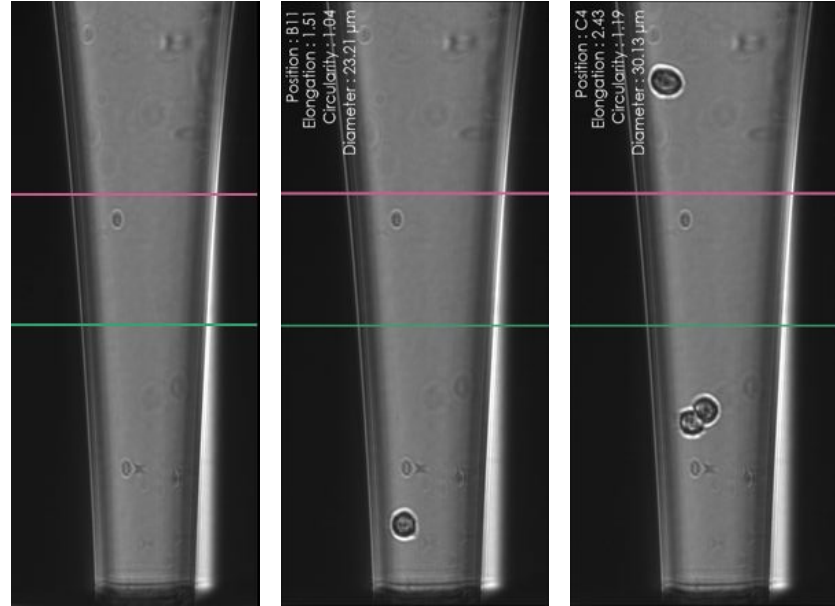
# ***APPENDIX***

# Some things I muted

- scRNAseq :
  - RNA velocity
  - Protein activity modelization
  - Stemness scoring
  - Variants detection
  - Integration:
    - Multiple samples
  - Multiple omics data
  - All non-droplet methods !
- sc Epigenomics : quite everything !
- Other :
  - Genomic :
    - Long reads
  - Non-genomic :
    - Imagery
- ERCCs ...

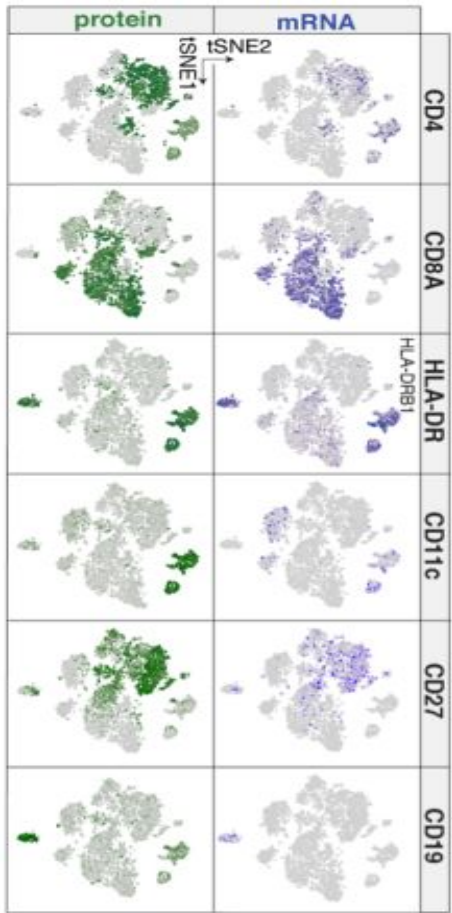
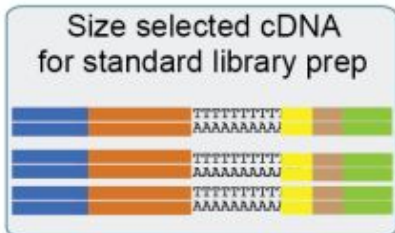
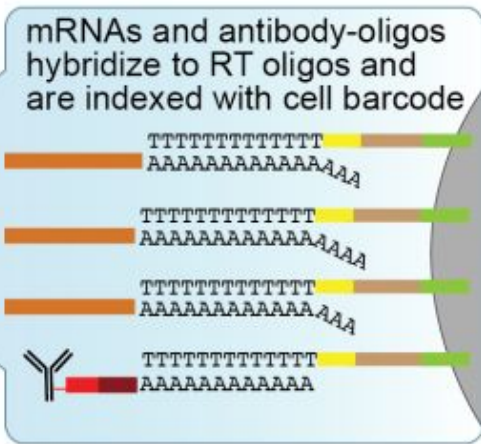
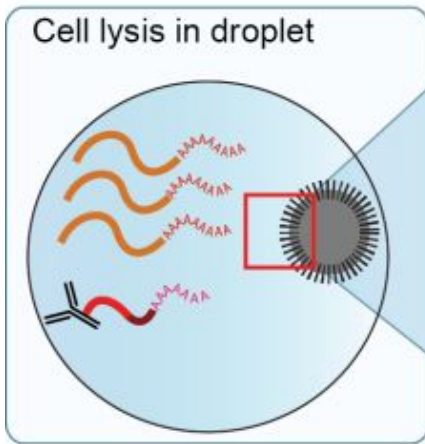
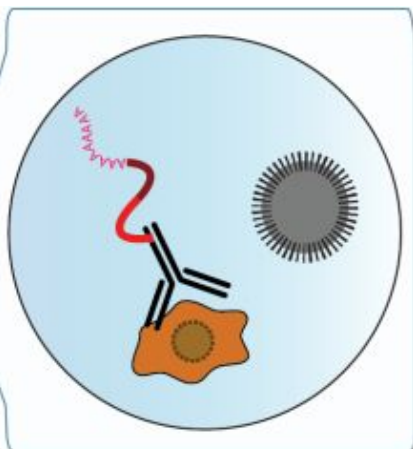
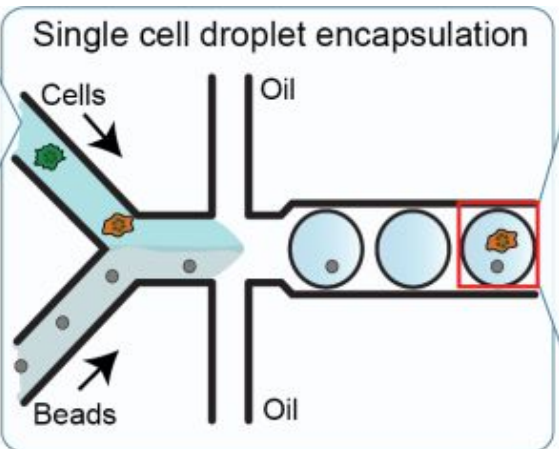
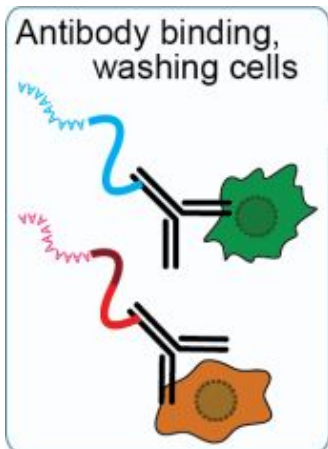
# Alternative isolation method : Cellenion IBSCI™ (Image Based Single Cell Isolation)

- Capillary real-time video recording :
  - Cell or no cell ?
  - More than 1 cell ?
  - Cell size ?
- Acoustic dispersion (more gentle)
- Middle scale :
  - Plate-based
  - Up to 1532 cells
- Cell recovery rate over 95%
- Open platform
  - Scalable, compatible
  - Custom reaction kits



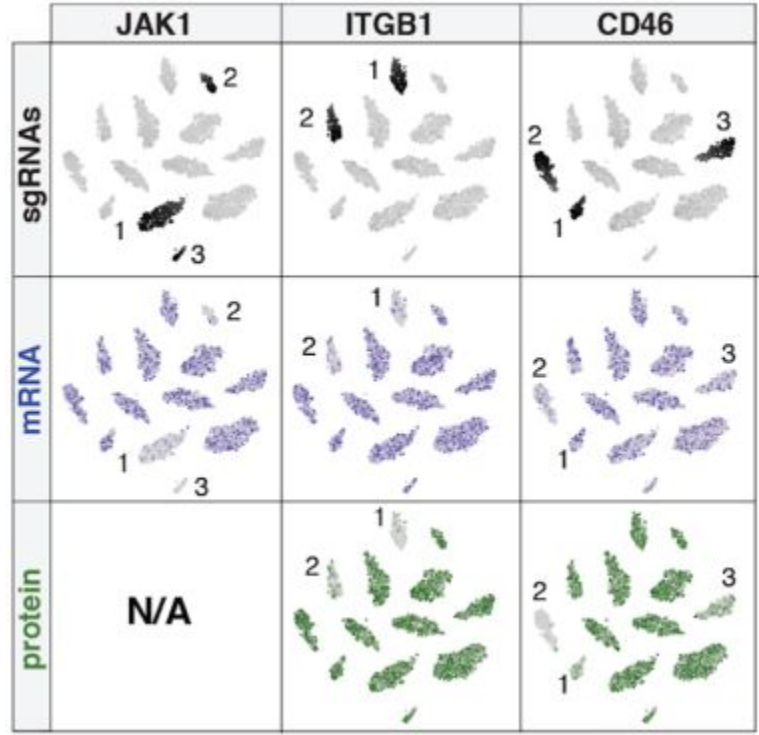
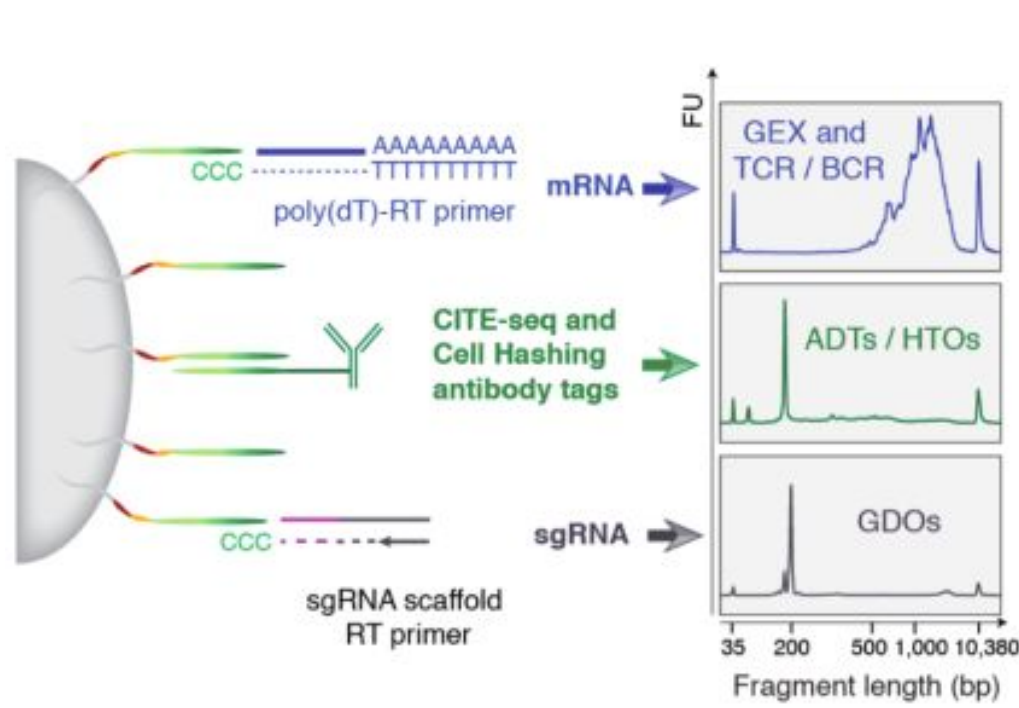
***CITE-seq***  
*(scRNAseq + proteins)*

# Cellular Indexing of Transcriptomes and Epitopes by Sequencing



***ECCITE-seq***  
(*scRNAseq + proteins + CRISPR gRNA*)

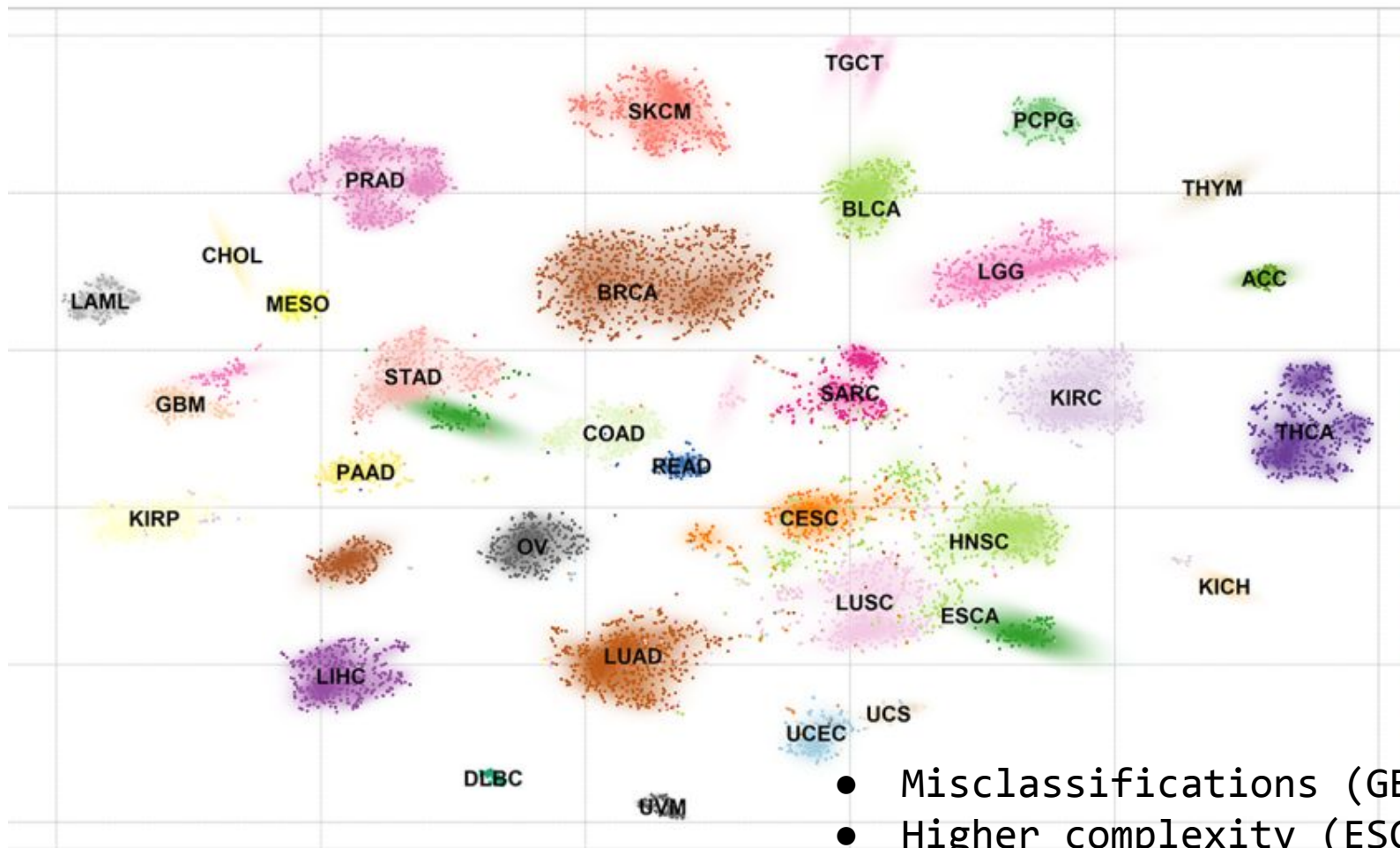
# Extended CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes by Sequencing (5')



*Some sweeties*

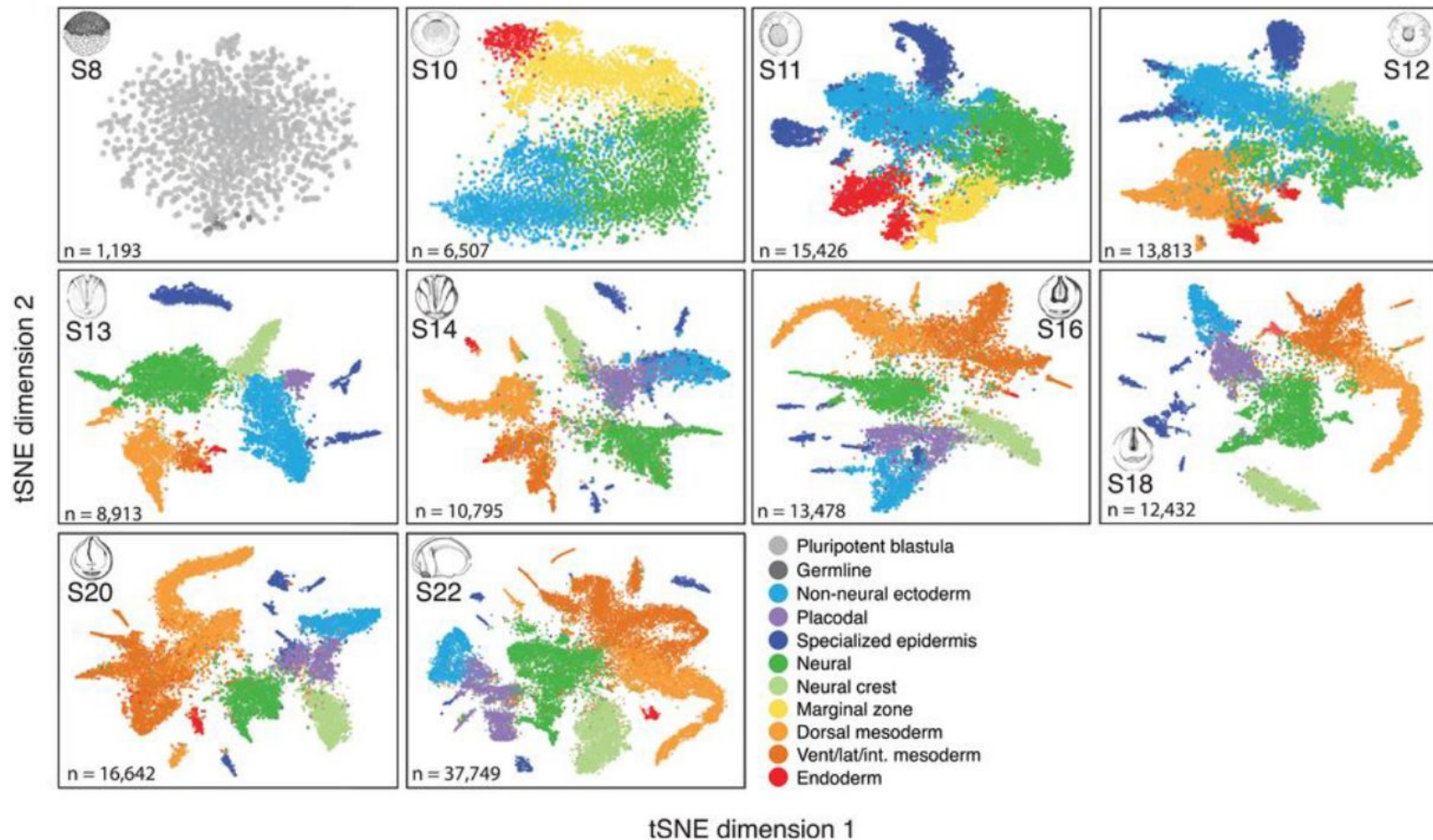


# t-SNE of the whole TCGA project (not SC)

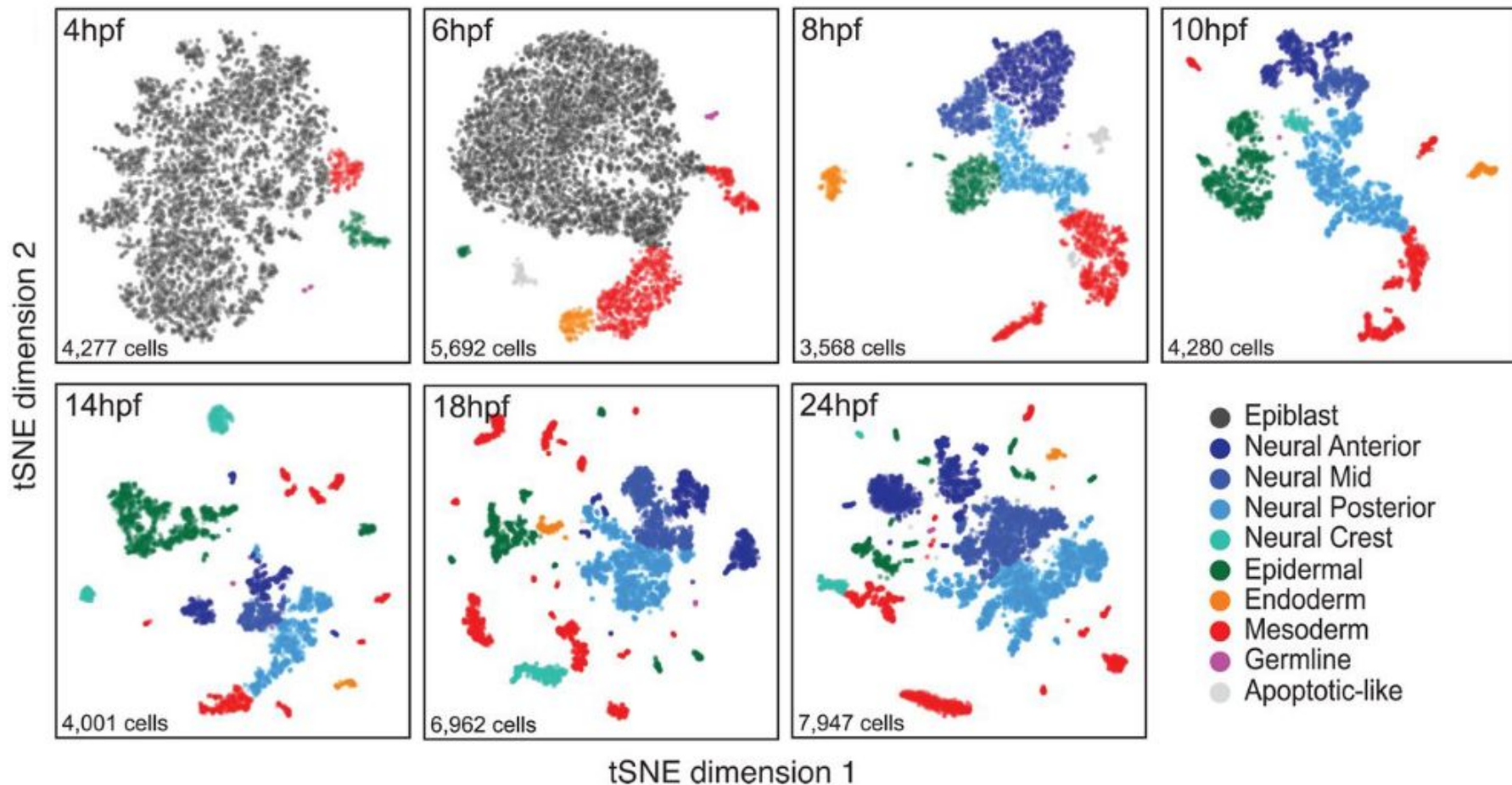


- Misclassifications (GBM-LGG)
- Higher complexity (ESCA-STAD)

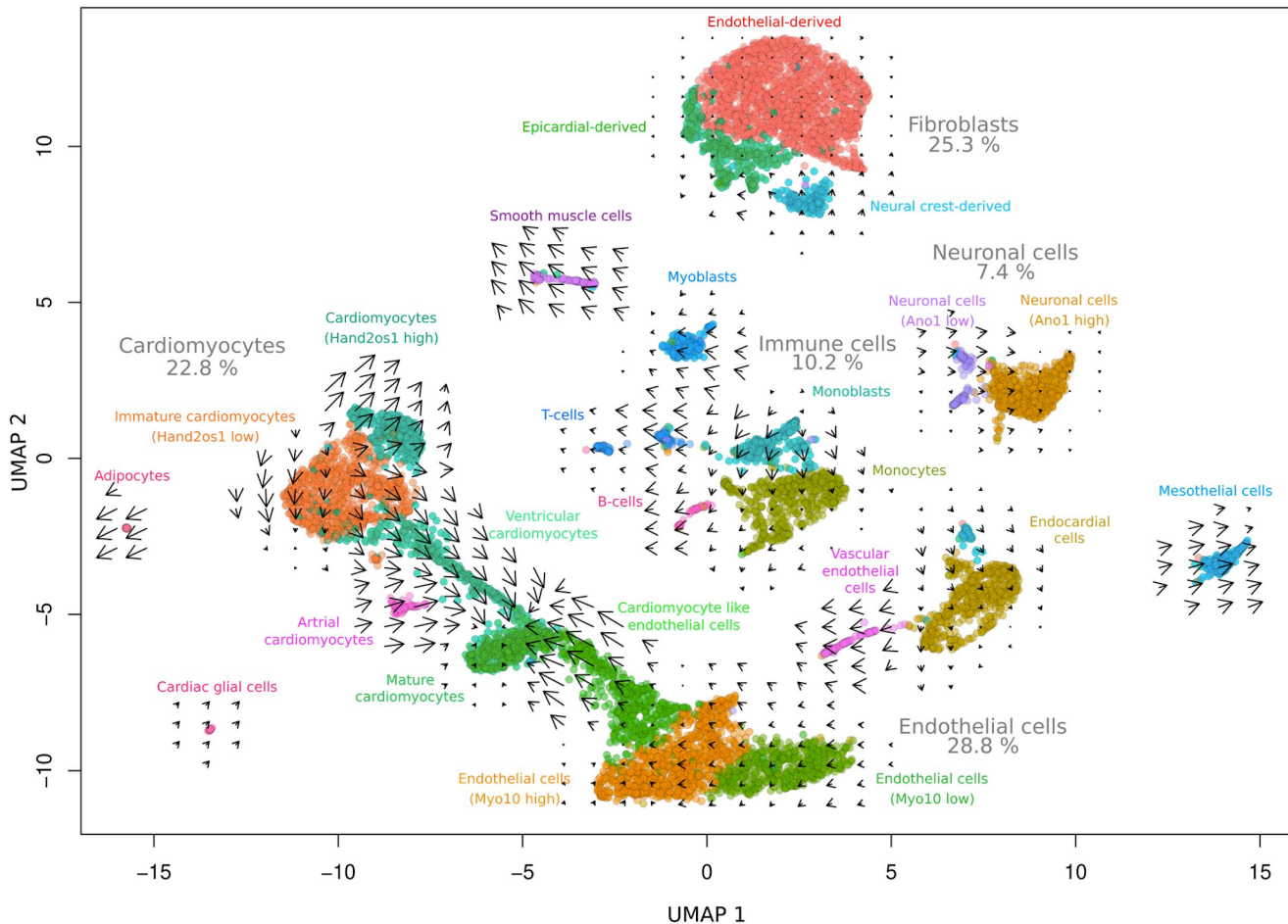
# Xenopus embryo development






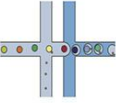

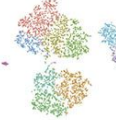
# Zebrafish embryo development



# Entire mouse heart : expression & velocity



# A little single scRNAseq cheatsheet

	<p><b>I. Tissue Procurement</b></p> <p><i>Source:</i></p> <ul style="list-style-type: none"> <li>- Primary human</li> <li>- Model organism</li> <li>- Cell culture</li> </ul>	<p><i>Key considerations:</i></p> <ul style="list-style-type: none"> <li>- Biological variation</li> <li>- Sampling/handling variation</li> <li>- Duration of sourcing</li> </ul>	<p><i>Study design:</i></p> <ul style="list-style-type: none"> <li>- Biological replicates</li> <li>- Technical replicates</li> <li>- Cell number calculation</li> <li>- Workflow optimization</li> </ul>
	<p><b>II. Tissue Dissociation</b></p> <p><i>Method:</i></p> <ul style="list-style-type: none"> <li>- Mechanical mincing</li> <li>- Enzymatic digestion</li> <li>- Automated blending</li> <li>- Microfluidics devices</li> </ul>	<p><i>Key considerations:</i></p> <ul style="list-style-type: none"> <li>- Experimental consistency</li> <li>- Shortest duration</li> <li>- Highest cell/nucleus quality</li> <li>- Representation of all cell types</li> </ul>	<p><i>Quality control:</i></p> <ul style="list-style-type: none"> <li>- FACS analysis</li> <li>- qPCR for marker genes</li> <li>- Imaging of cell integrity</li> <li>- RNA quality (RIN)</li> </ul>
	<p><b>III. Cell Enrichment (optional)</b></p> <p><i>Method:</i></p> <ul style="list-style-type: none"> <li>- Differential centrifugation, sedimentation, filtration</li> <li>- Antibody labeling for positive/negative selection</li> <li>- Flow cytometry or bead-based enrichment</li> <li>- Dead cell removal</li> </ul>	<p><i>Key considerations:</i></p> <ul style="list-style-type: none"> <li>- Additional handling</li> <li>- Longer duration</li> <li>- Loss of RNA quality</li> <li>- Transcriptome changes</li> </ul>	
	<p><b>IV. Single Cell RNAseq Platform</b></p> <p><i>Method:</i></p> <ul style="list-style-type: none"> <li>- Droplet-based</li> <li>- Tube-based after FACS</li> <li>- Microwell-based</li> <li>- Microfluidics-enabled</li> </ul>	<p><i>Key considerations:</i></p> <ul style="list-style-type: none"> <li>- Cell throughput and handling time</li> <li>- Gene coverage and cell type detection</li> <li>- Whole transcript versus 3' end counting</li> <li>- Imaging capability for doublet detection</li> </ul>	
	<p><b>V. Library Sequencing</b></p> <p><i>Method:</i></p> <ul style="list-style-type: none"> <li>- Illumina NGS</li> <li>- Compatible with cDNA library</li> </ul>	<p><i>Sequencing depth considerations:</i></p> <ul style="list-style-type: none"> <li>- 3' end counting: low depth ~50K RPC</li> <li>- Whole transcript: high depth ~1M RPC</li> <li>- Alternative splicing: ~20-30M RPC</li> <li>- Iterative optimization for biological system</li> </ul>	
	<p><b>VI. Computational Analysis</b></p> <p><i>Key considerations:</i></p> <ul style="list-style-type: none"> <li>- Separation of <i>batch</i> and <i>condition</i></li> <li>- Technical vs. biological variation</li> </ul>	<p><i>Sample Batch correction approaches:</i></p> <ul style="list-style-type: none"> <li>- Cell Hashing</li> <li>- Demuxlet</li> <li>- Canonical correlation analysis (CCA)</li> <li>- MAST</li> </ul>	