

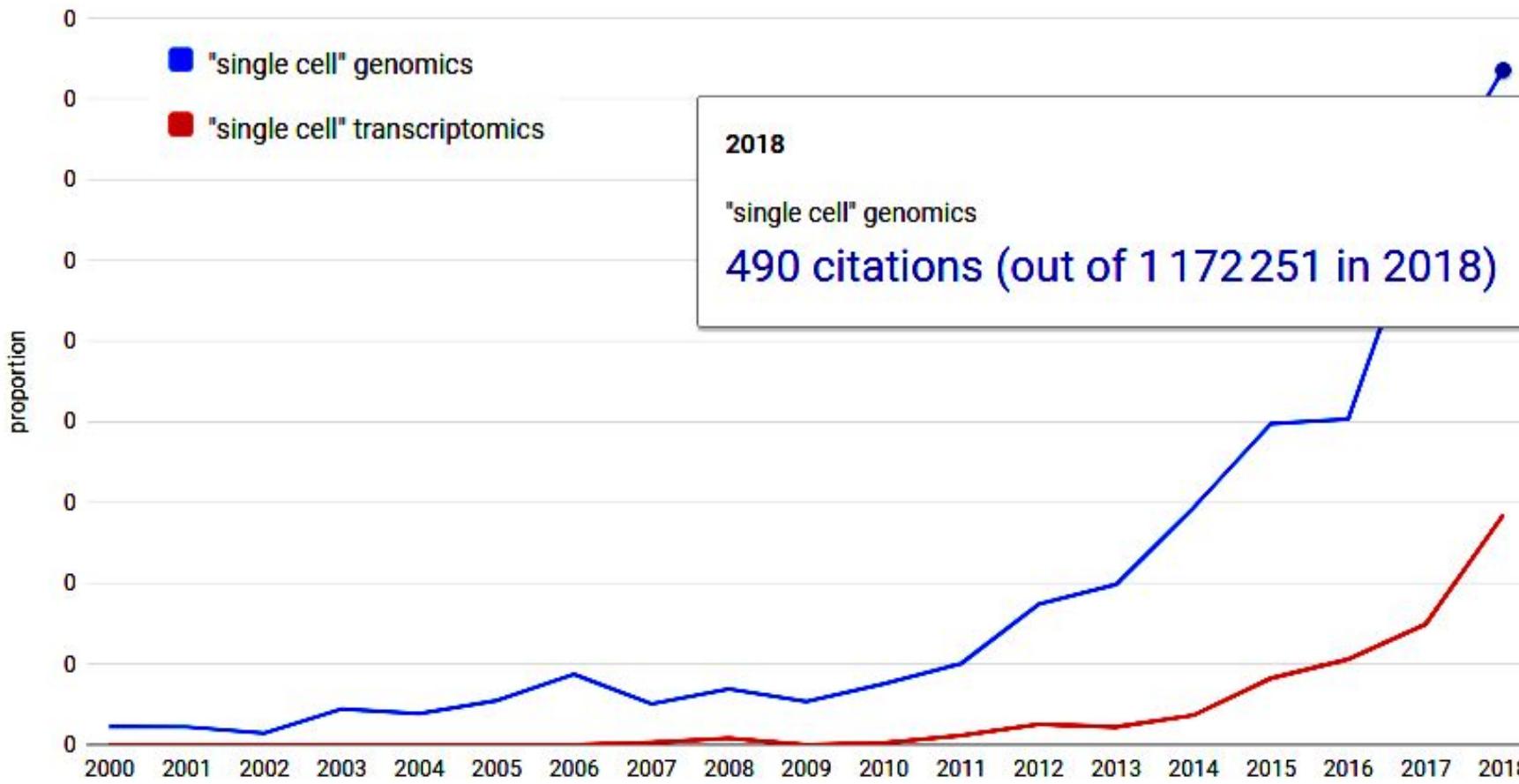
An Introduction to Single-Cell Genomics

(Actually, mostly transcriptomics)

Bastien JOB
INSERM / Gustave Roussy

So you say you've heard about single cell ?

Single cell in peer-reviewed publications (2018)

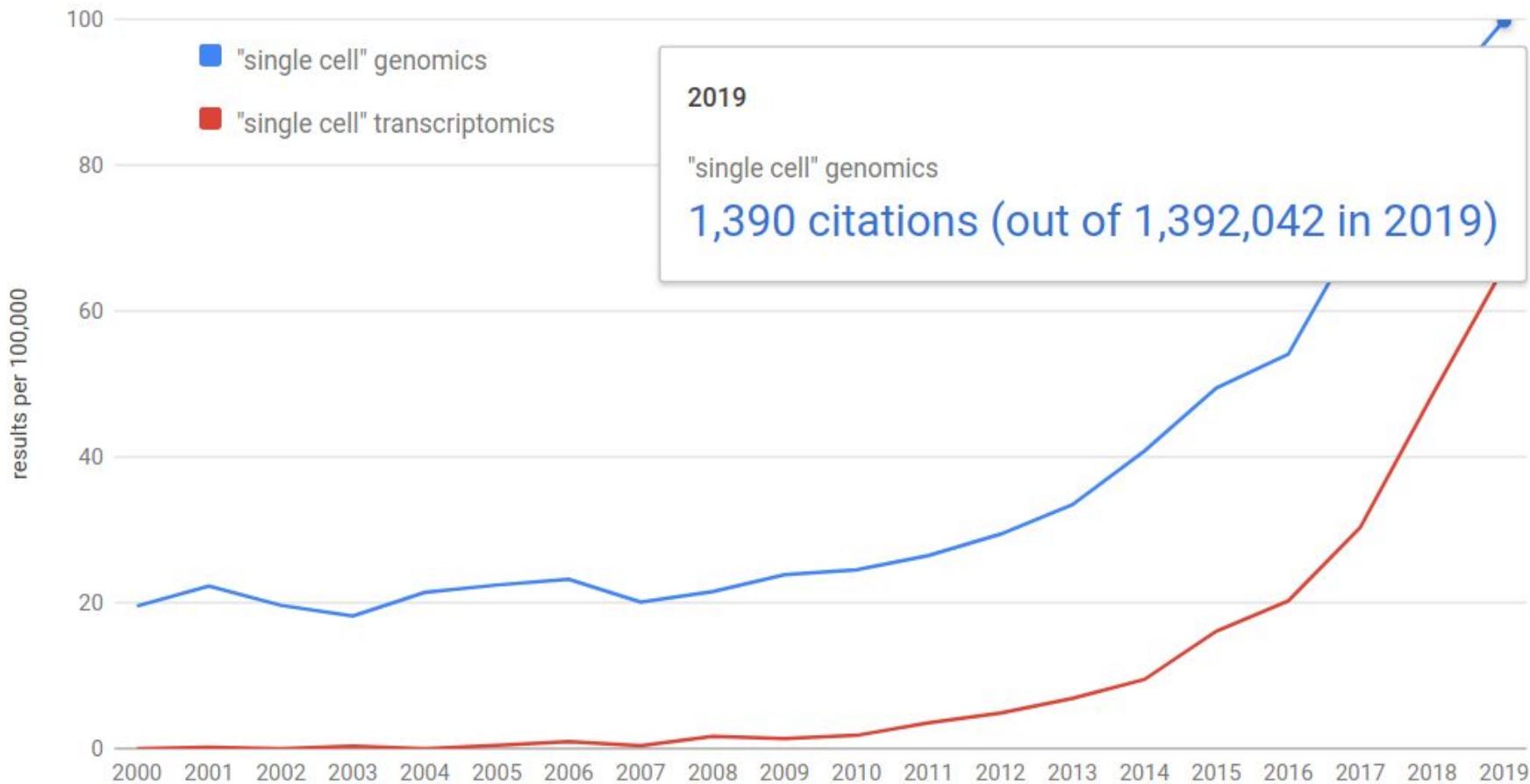


2018

"single cell" genomics

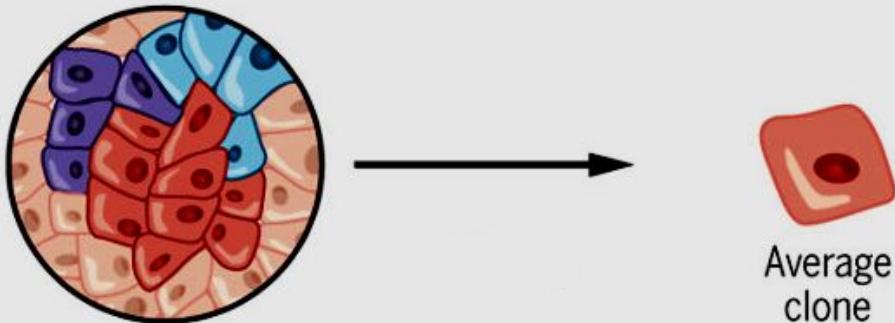
490 citations (out of 1 172 251 in 2018)

Single cell in peer-reviewed publications (2019)

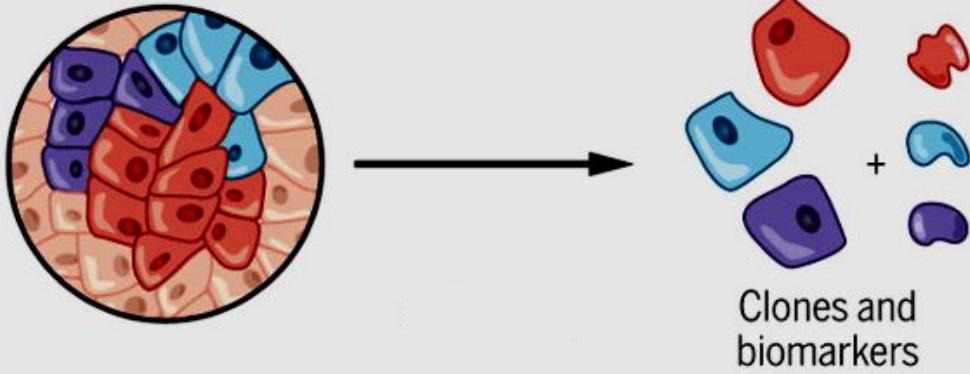


Why so much hype ?

A Bulk analysis



B scRNA analysis

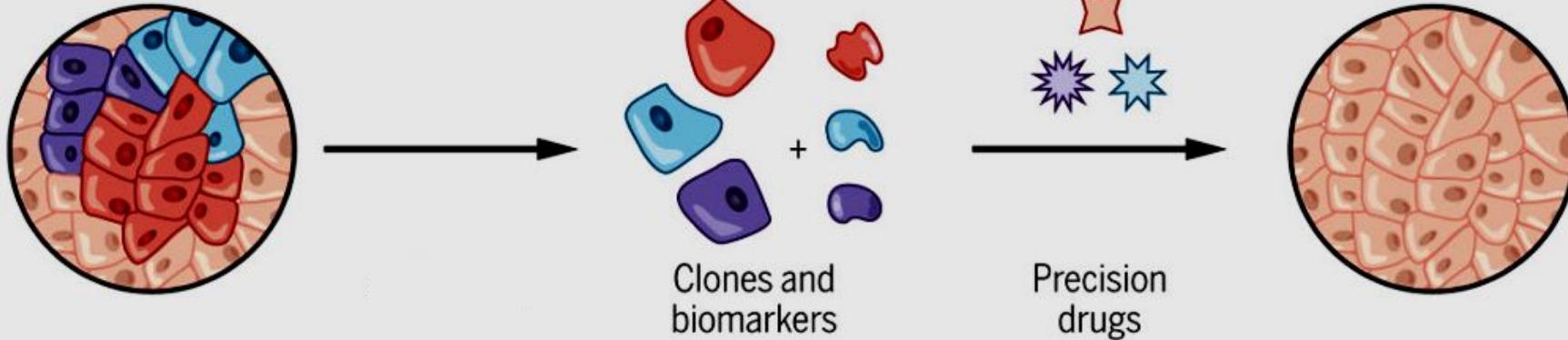


Why so much hype ? (pathology)

A Bulk analysis



B scRNA analysis



Why so much hype ?

Bulk



Single cell

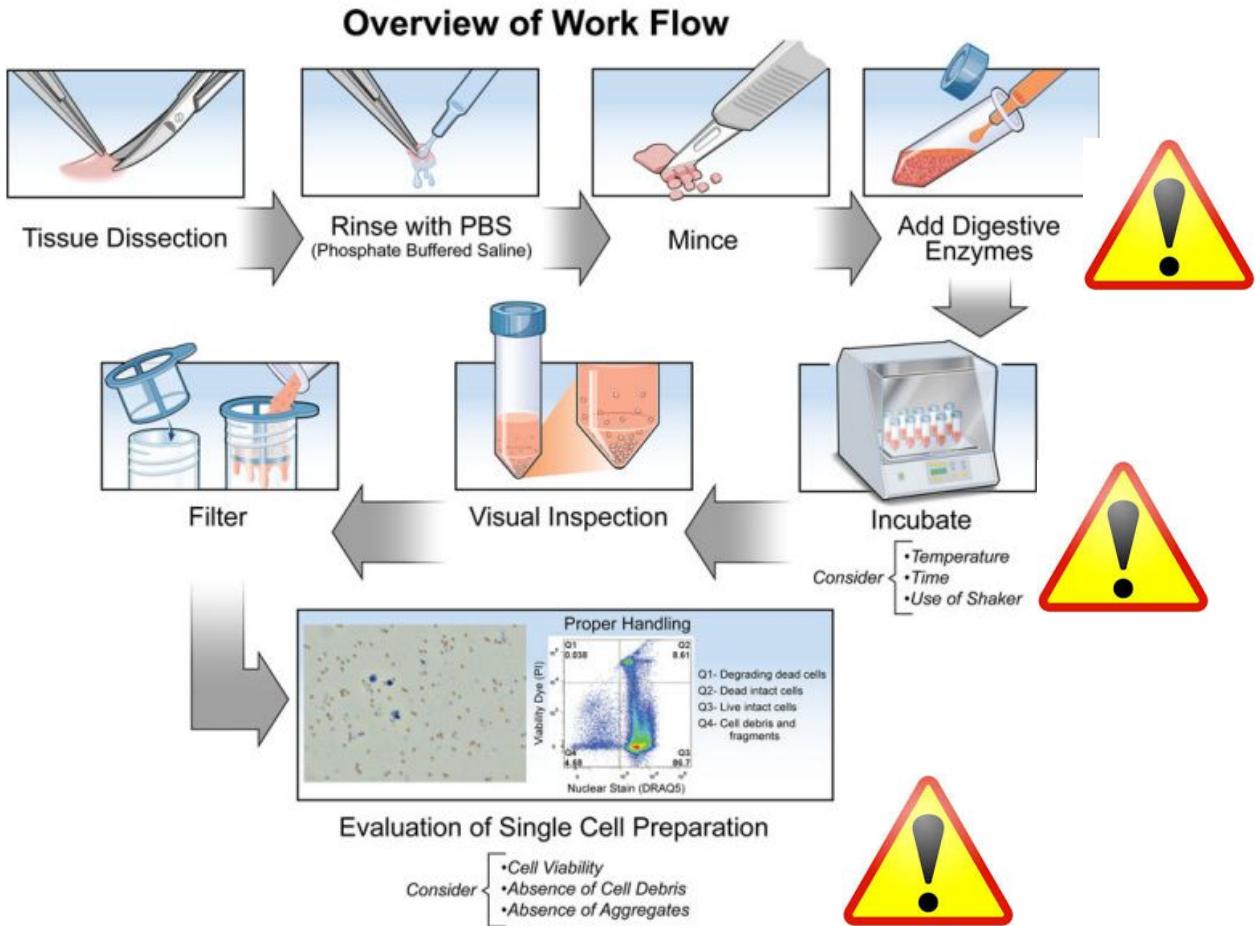


Spatial single cell



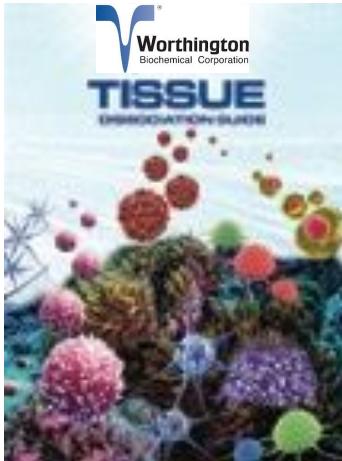
From broad tissue to isolated cells

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



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

II. Cell Isolation Theory

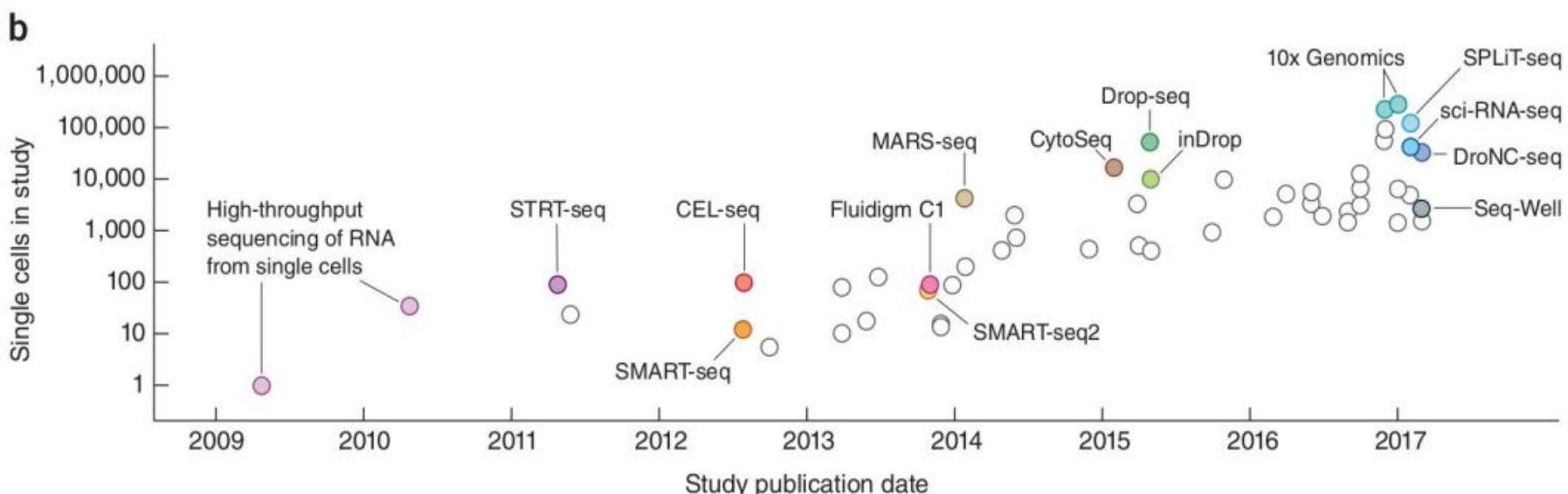
- Tissue Types
 - Epithelial Tissue
 - Connective Tissue
- Dissociating Enzymes
 - Collagenase
 - Trypsin
 - Elastase
 - Hyaluronidase
 - Papain
 - Chymotrypsin
 - Deoxyribonuclease I
 - Neutral Protease (Dispase)
 - Trypsin Inhibitor
 - Animal Origin Free (AOF) Enzymes
 - Celase® GMP

III. Cell Isolation Techniques

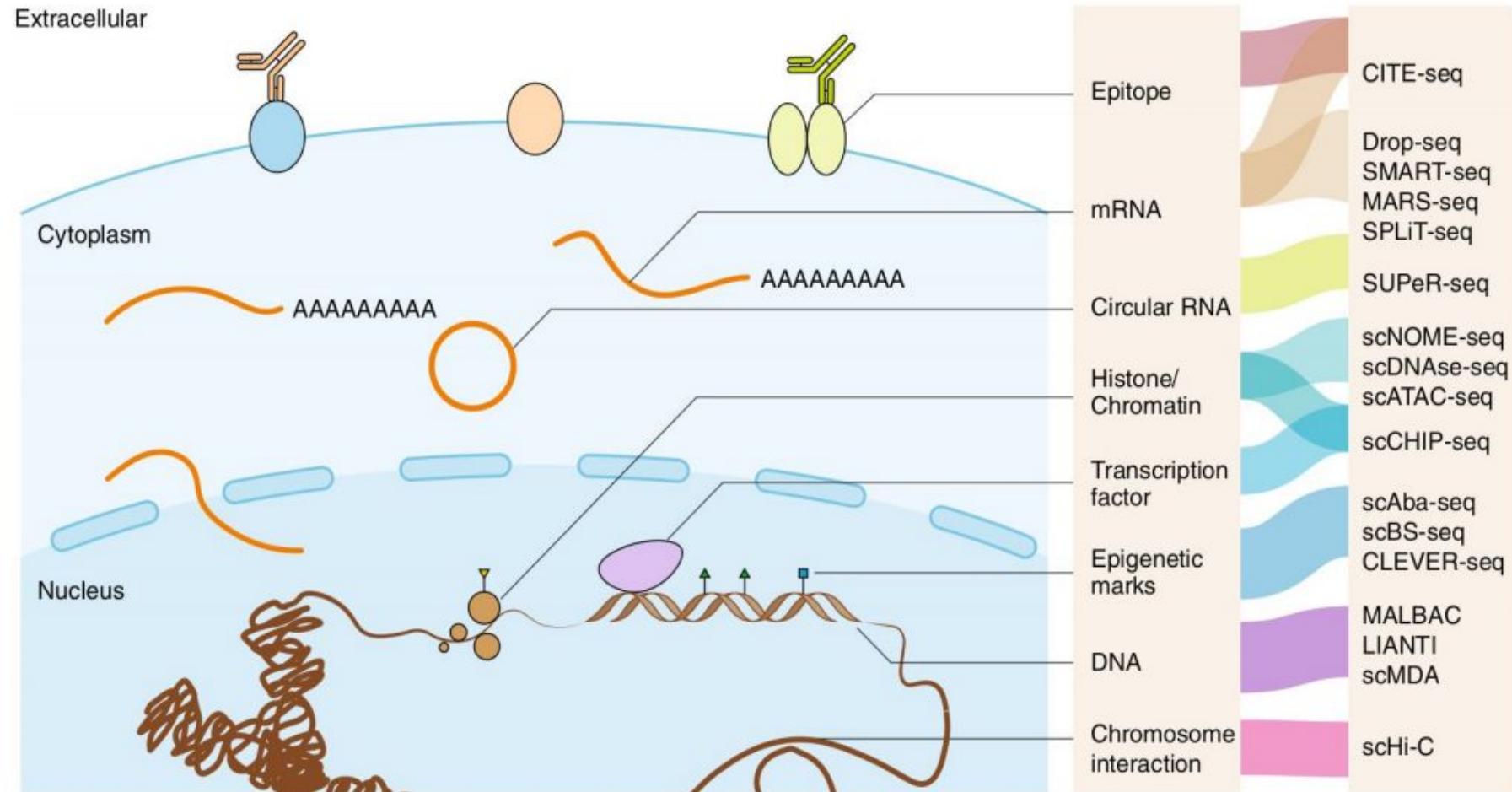
- Methods & Materials
 - Working With Enzymes
 - Basic Primary Cell Isolation
 - Equilibration with 95%O₂:5%CO₂
 - Trituration
 - Enzymatic Cell Harvesting
 - Cell Adhesion and Harvesting
 - Trypsin for Cell Harvesting
 - Cell Release Procedure
- Optimization Techniques
 - General Guidelines
 - Optimization Strategy
 - Cell Quantitation
 - Measure of Viability

IV. Use-Tested Cell Isolation Systems

Cells separation : technologies over the last decade

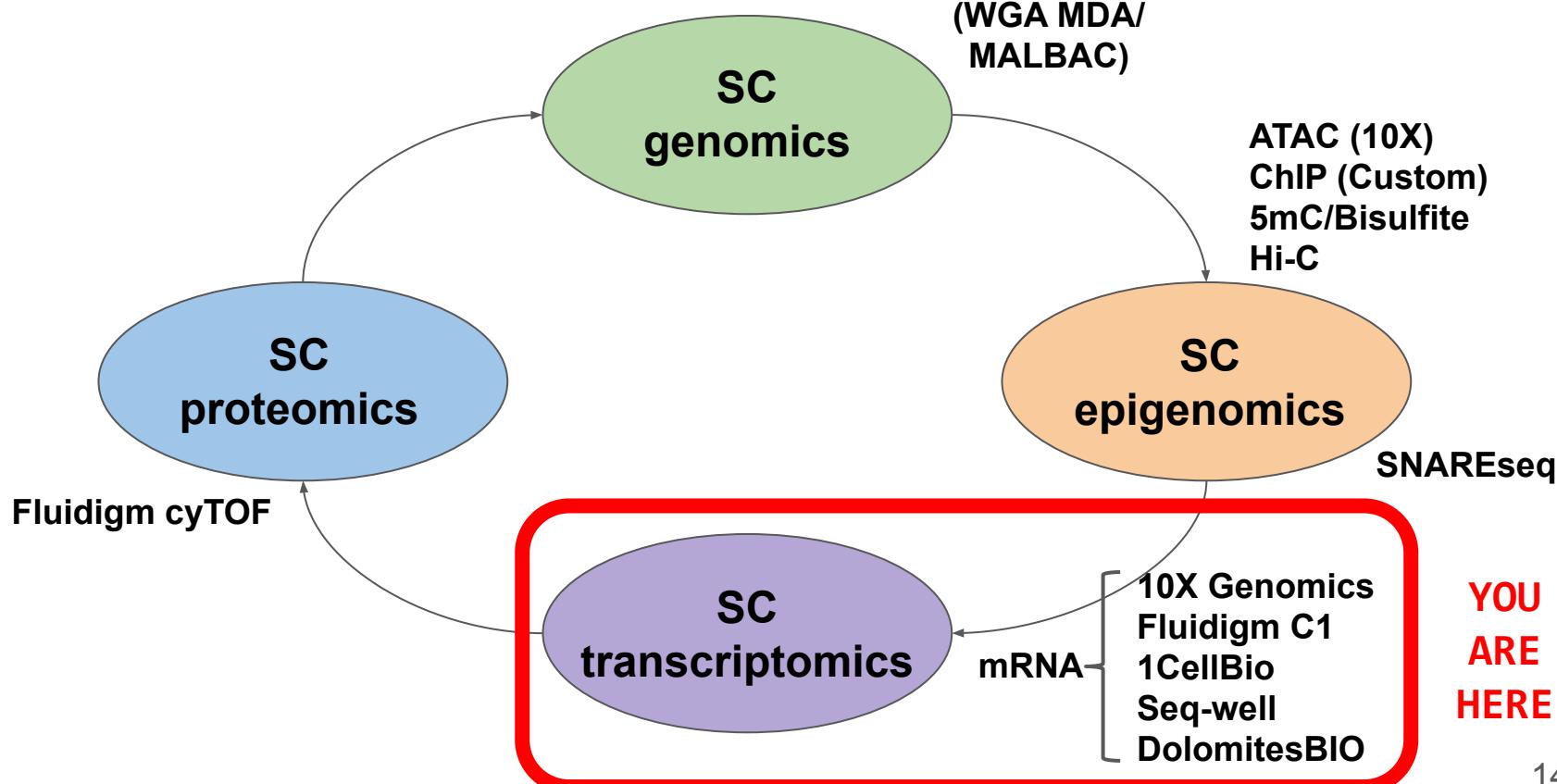


Several protocols for several purposes

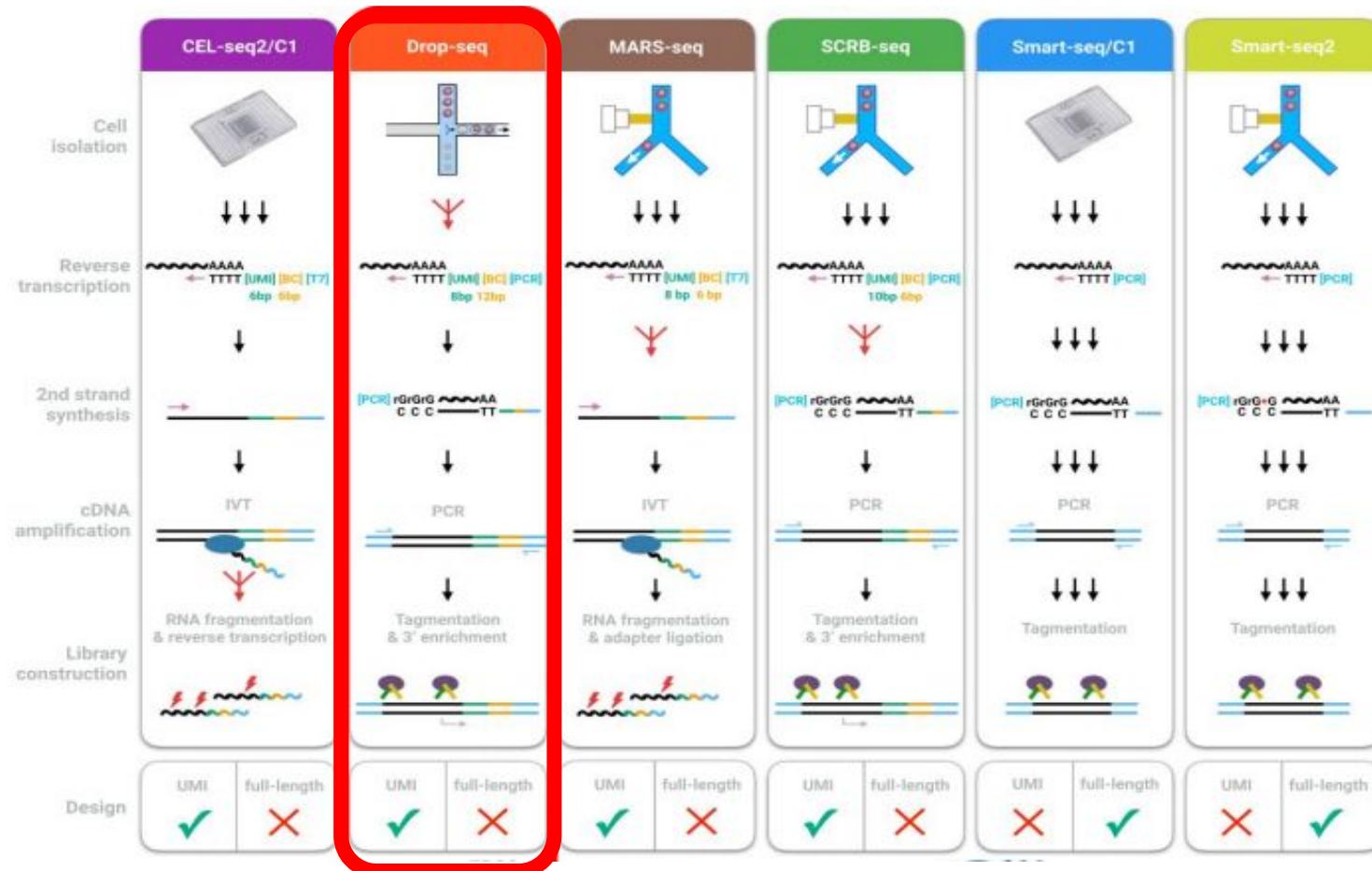


From isolated cells to nucleotide sequences

Single Cell RNaseq



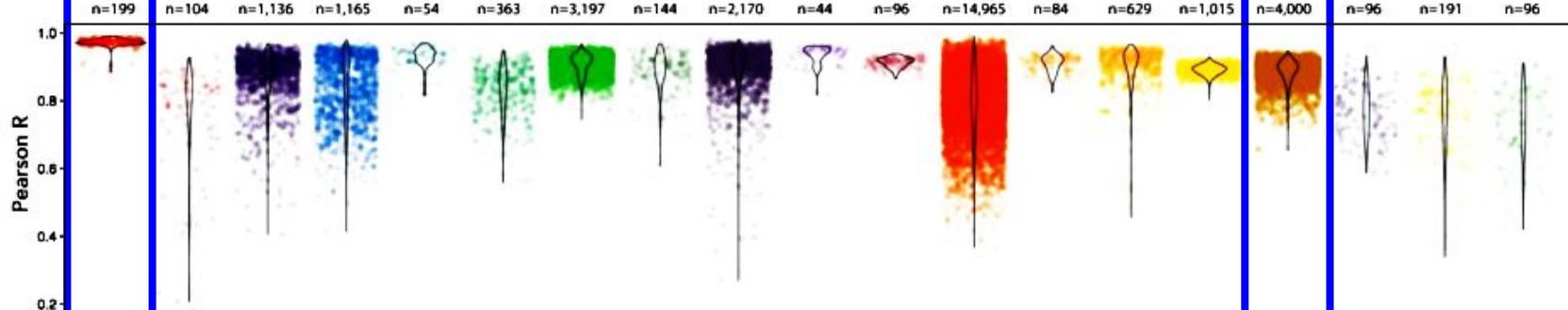
From isolated cells to sequences (Drop-seq / 10X)



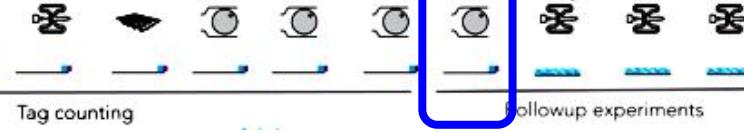
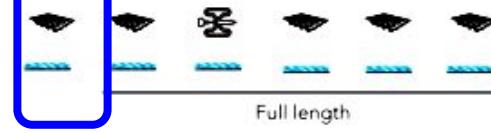
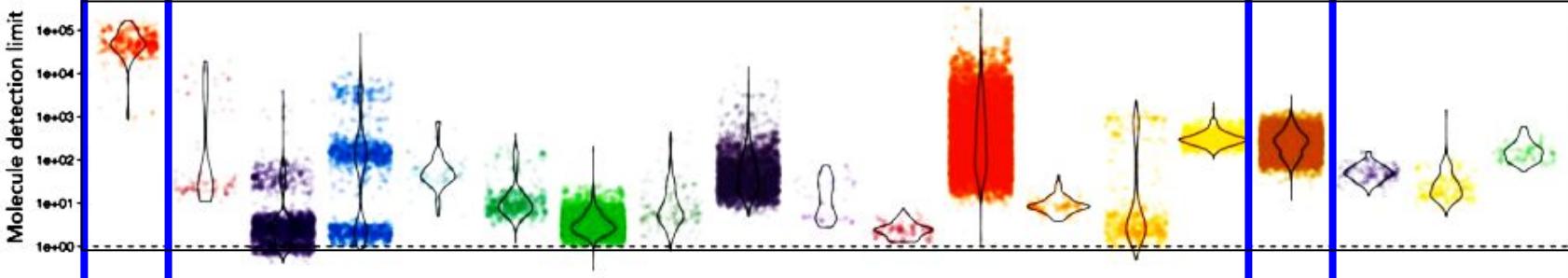
Bulk

10X

Accuracy

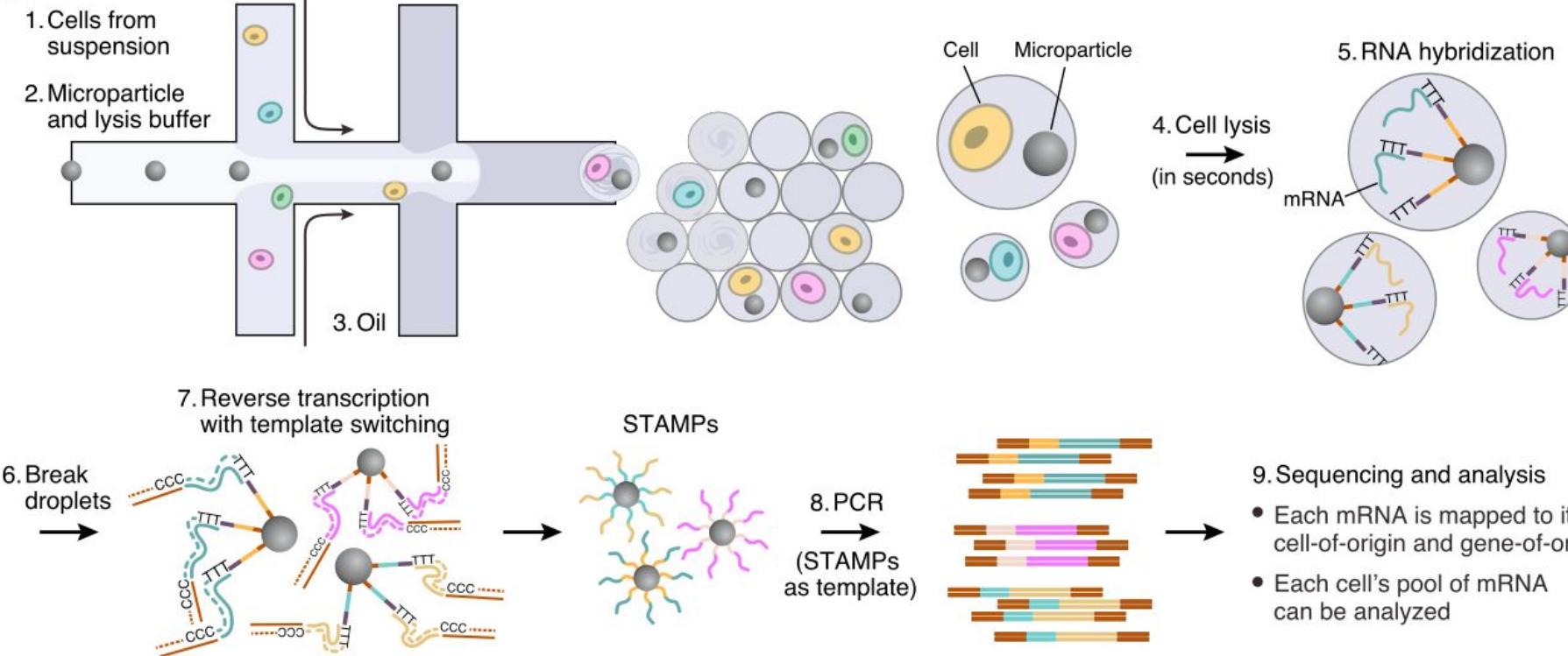


Sensitivity



Drop-seq

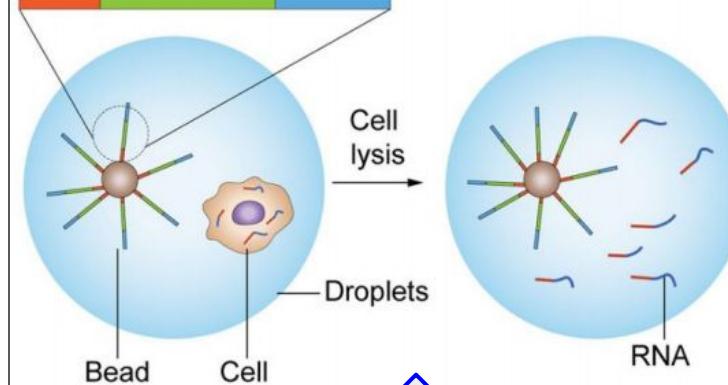
A



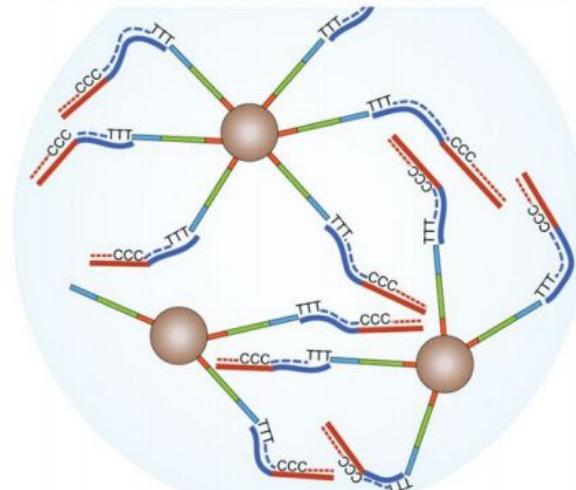
10X Chromium (3')

Structure of the barcode primer bead

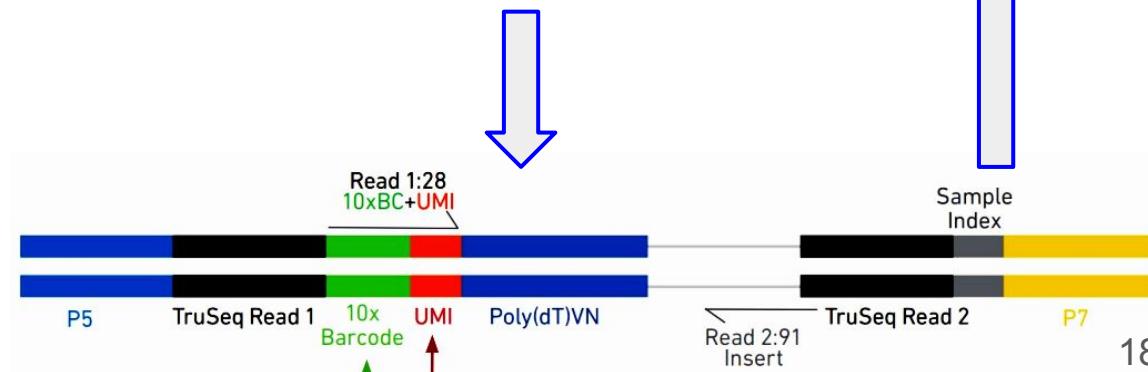
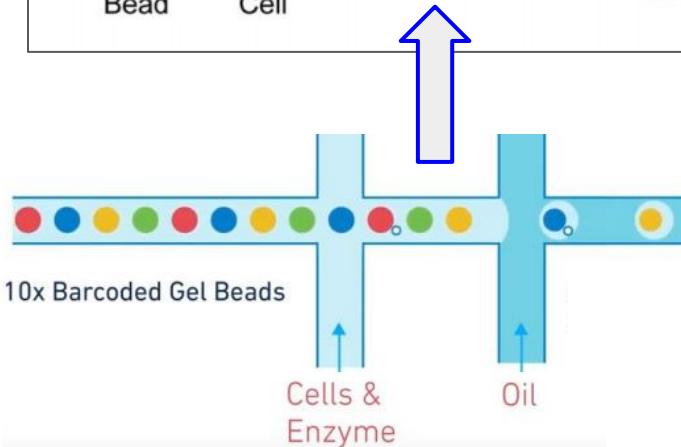
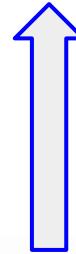
PCR
handle Cell barcode UMI



Reverse transcription with template switching



Sequencing



*From nucleotide sequences (reads)
to count matrix*

CERTIFIED



BIOINFORMAGICIAN

Reads QC

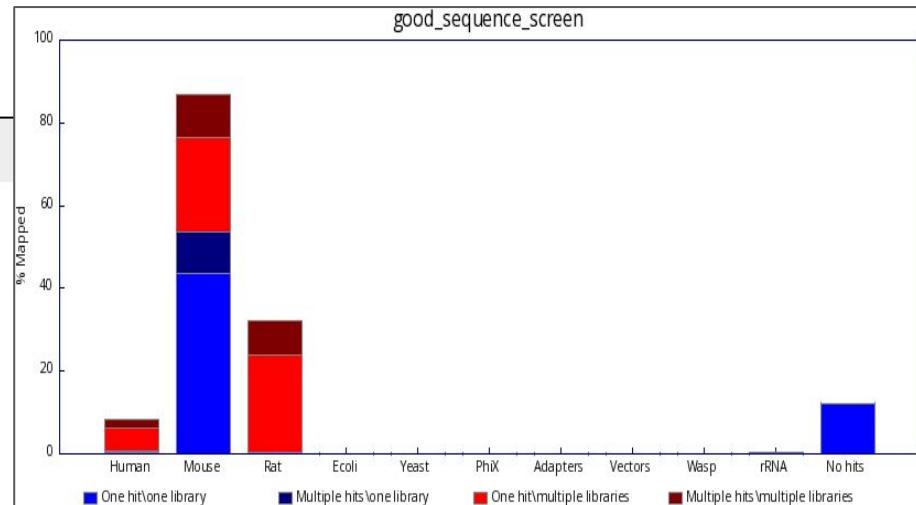
FastQC Report

Summary

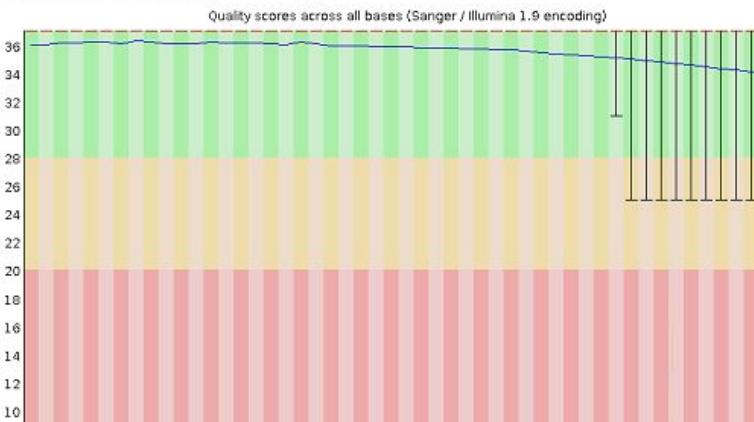
- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✗ Per tile sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✓ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✓ Adapter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



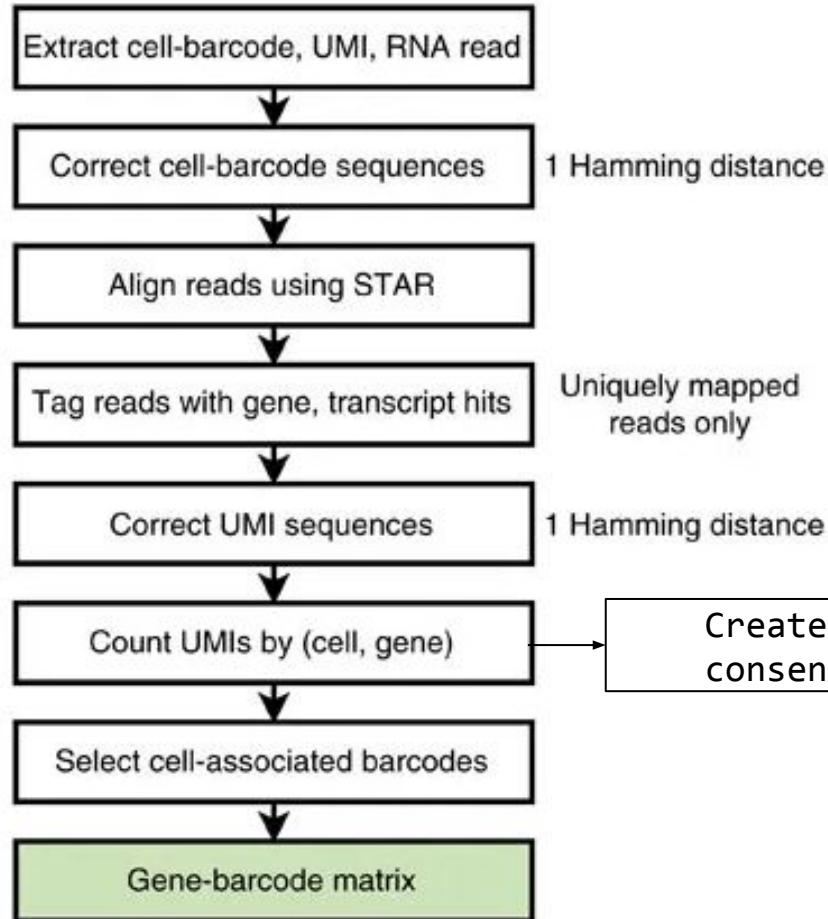
Per base sequence quality



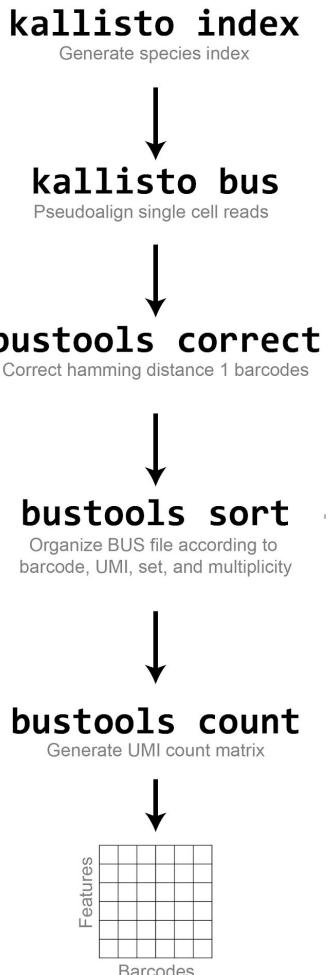
- As usual : FASTQC, FastqScreen, ...
- 10x :
 - Special attention to R1 : cell barcode + UMI (no N)
 - Control of the 4 sample libraries

Reads processing workflows

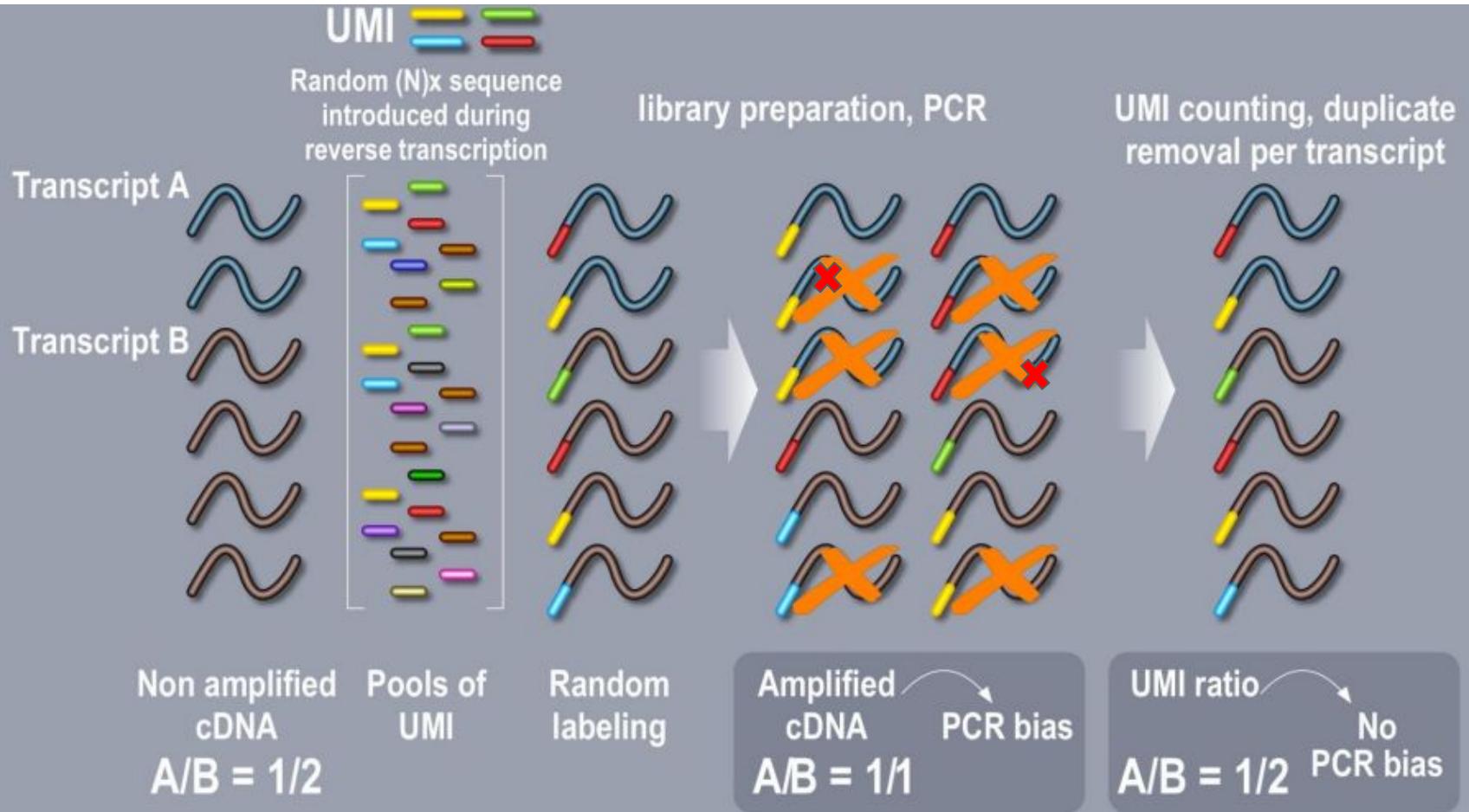
Mapping-based (STAR)



Pseudomapping-based (kallisto bustools)

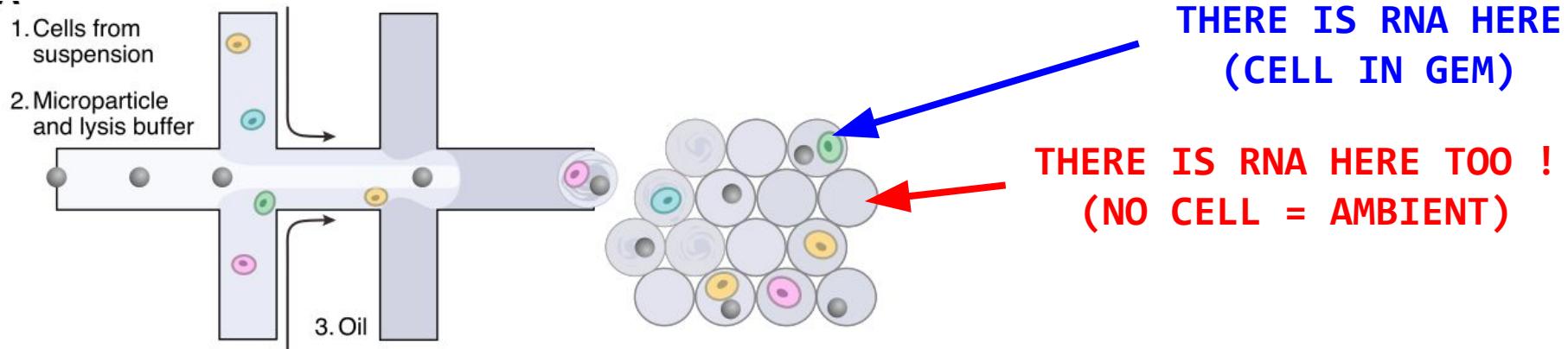


Focus on : Unique Molecule Identifiers

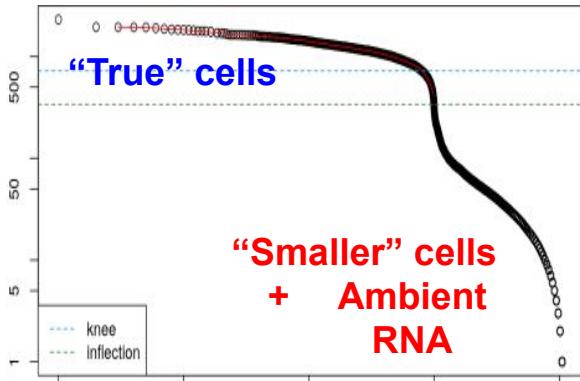


Islam et al., Nature Methods (2014)
Study by Agnès Paquet

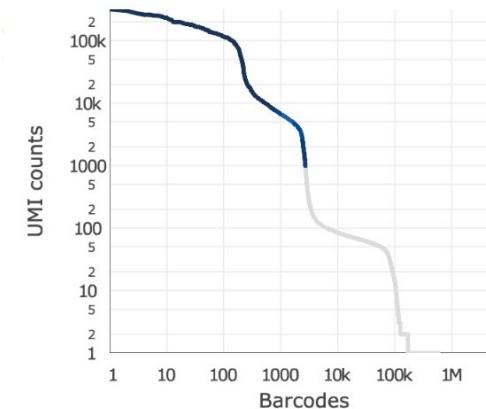
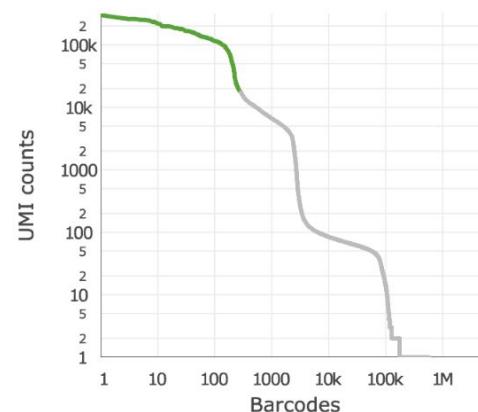
Focus on : Empty droplets filtering



Kneeplot

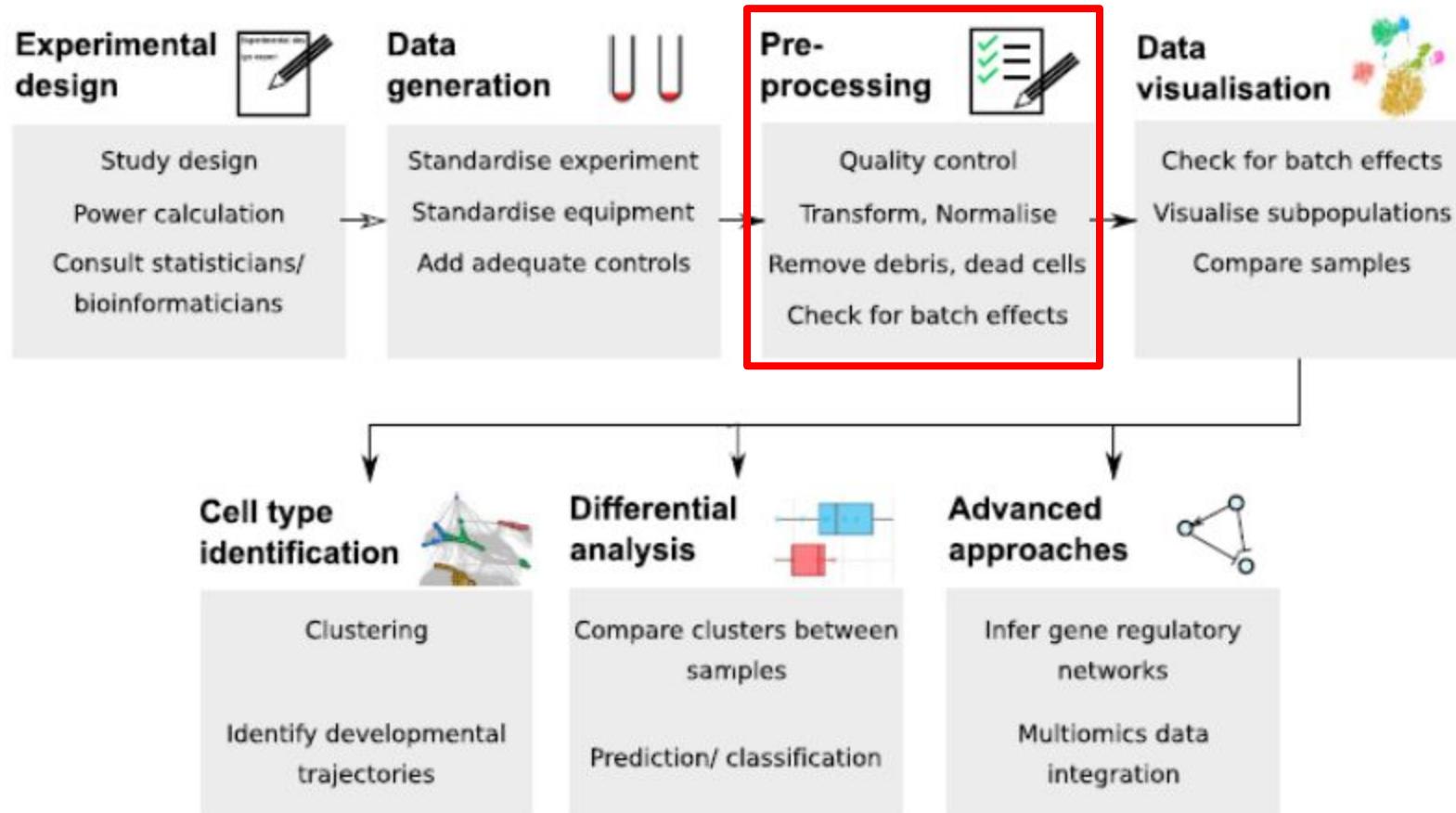


EmptyDrops



From raw count matrix to normalized data

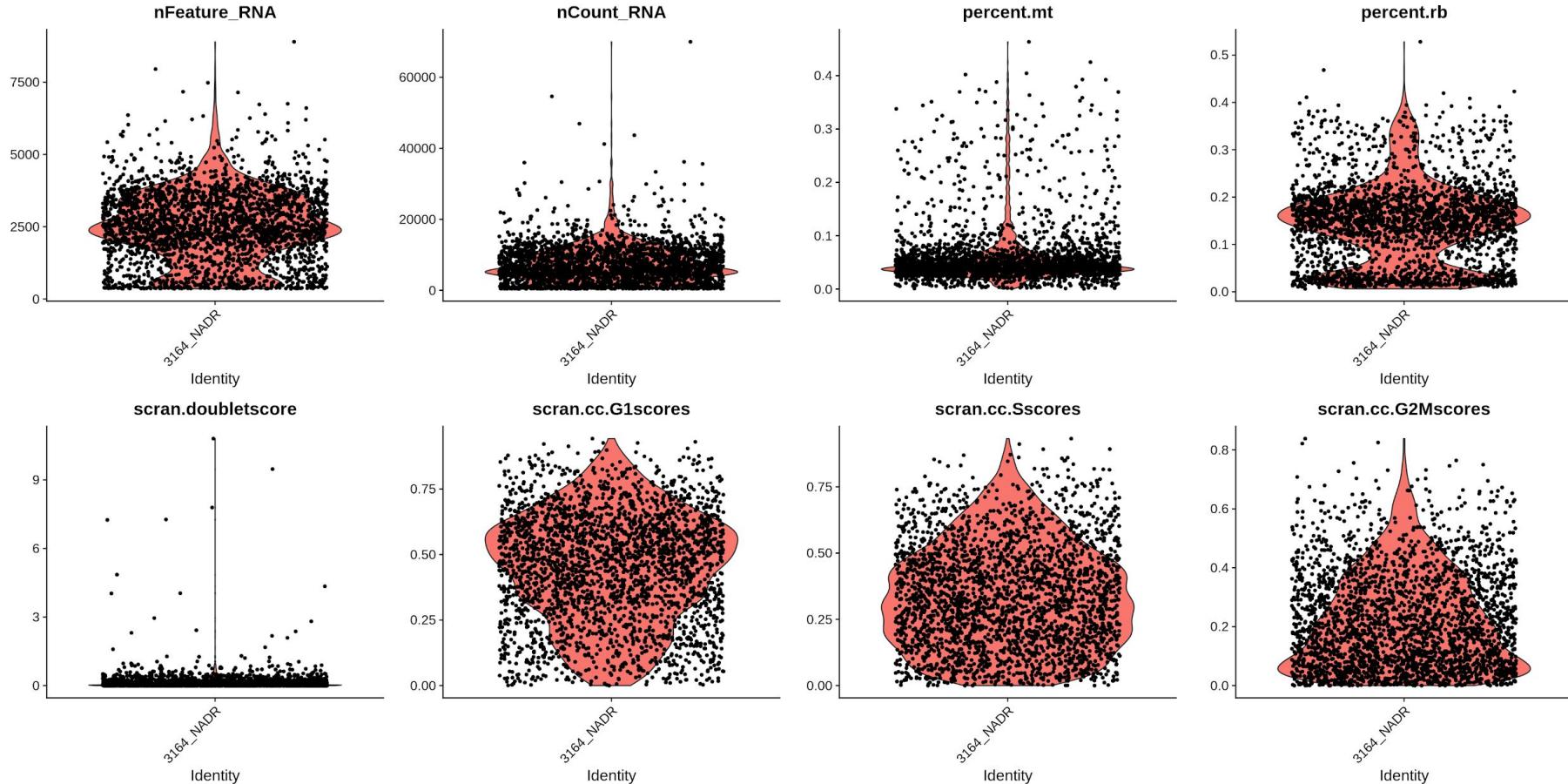
Standard analysis pipeline



Cell QC considerations

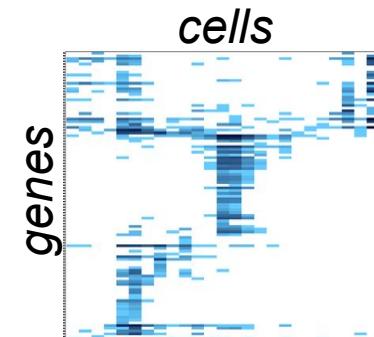
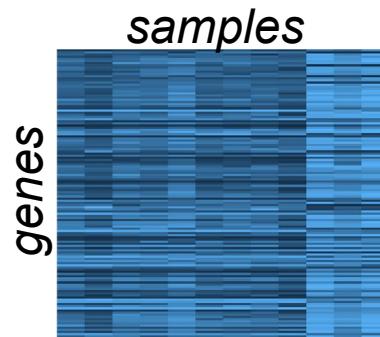
- The number of unique genes detected in each cell :
 - Low-quality cells or empty droplets will often have very few genes
 - Cell doublets or multiplets may exhibit an aberrantly high gene count
- Similarly, the total number of molecules detected within a cell (correlates strongly with unique genes)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

Cell QC : metrics



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

	BULK	SINGLE-CELL
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)

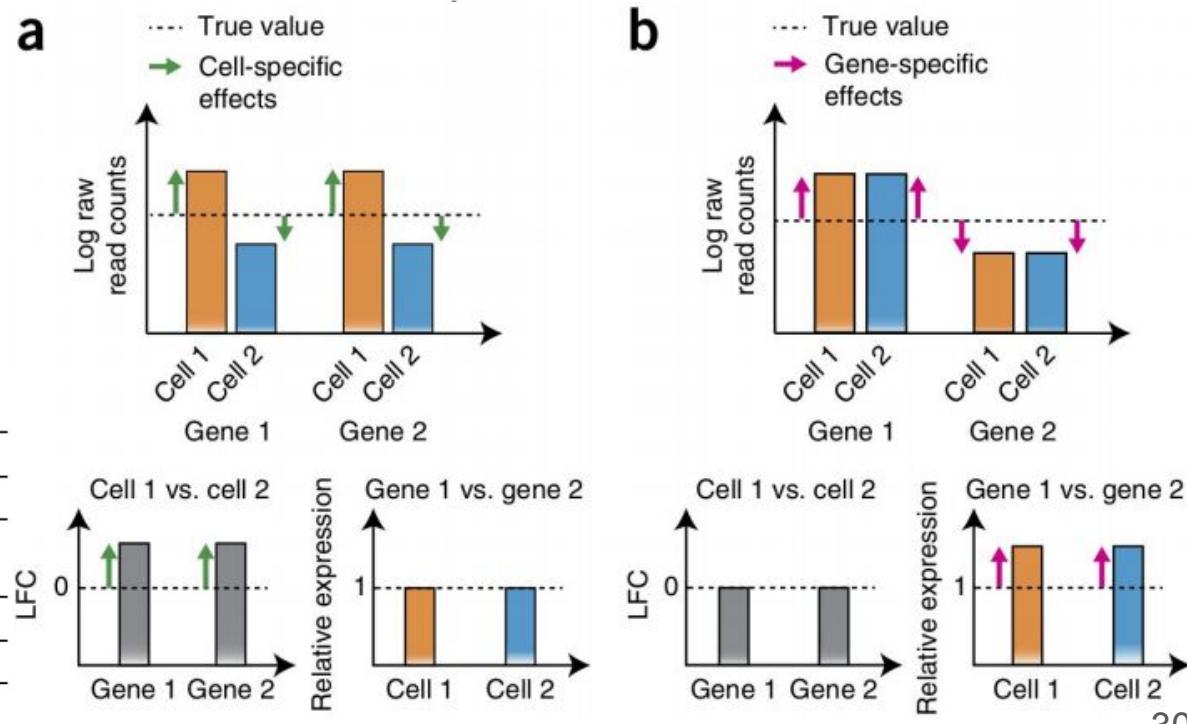


**SC MATRIX IS SPARSE !
(ie, mostly filled with zeros)**

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**

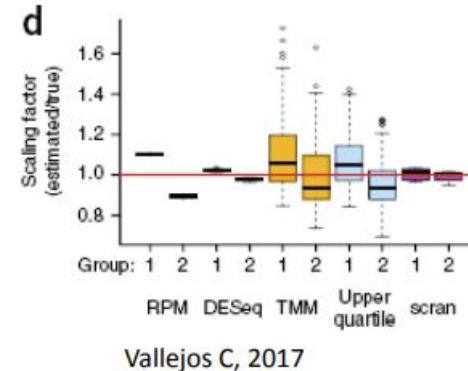
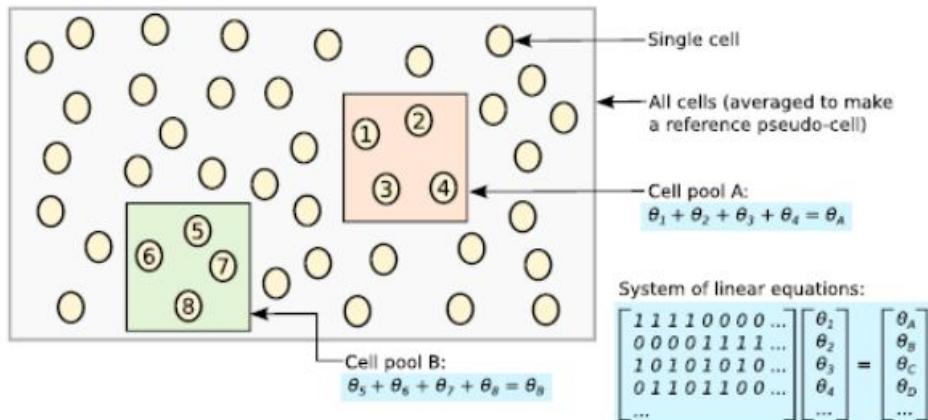
- RPKM/FPKM (Reads/Fragments per kilobase of transcript per million reads of library) : Normalize for sequencing depth and transcript length at the same time => **KO if you DO NOT have full-length data**
- Global scaling (eg: Upper Quartile) : **KO if you have too many zeros**
- Size factors calculation (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

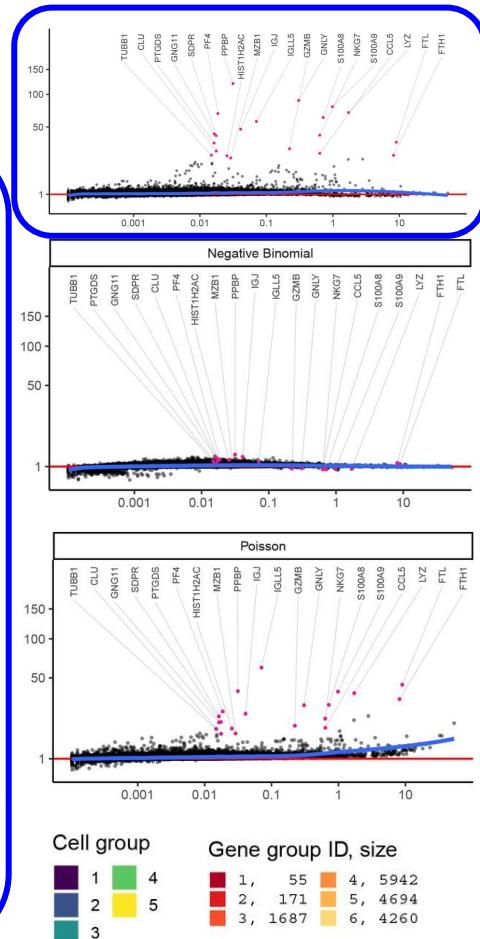
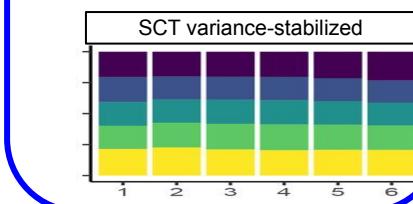
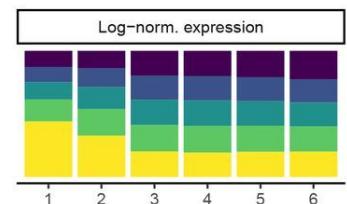
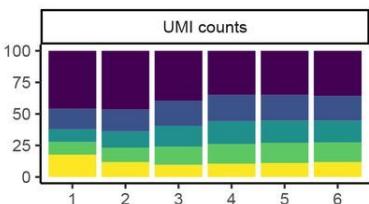
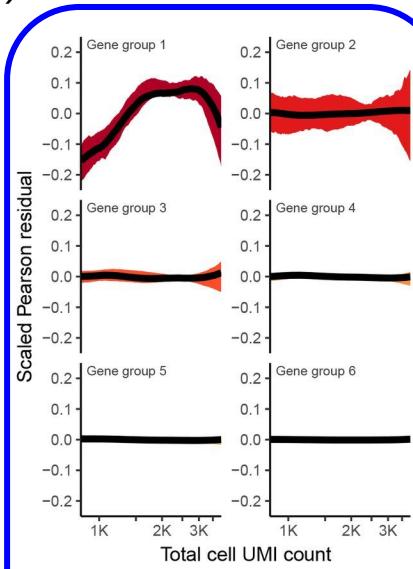
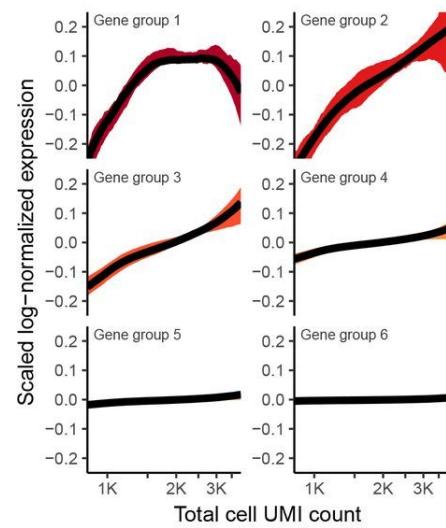
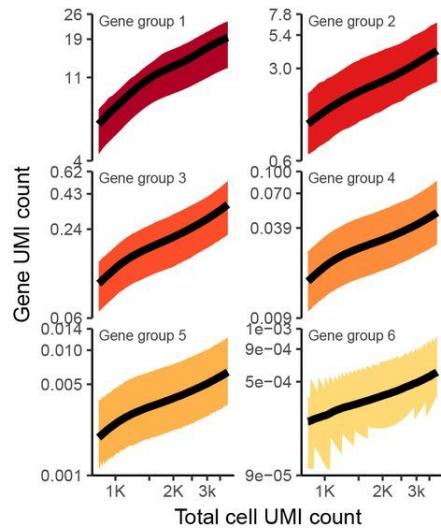
- Alternative method to compute the size factors
- Pool cells to reduce the number of zeros
- Estimate the size factors for the pool
- Repeat many time and use deconvolution to estimate each cell size factor
- Implemented in **scater/scran** packages



Vallejos C, 2017

Matrix normalization : variance stabilization

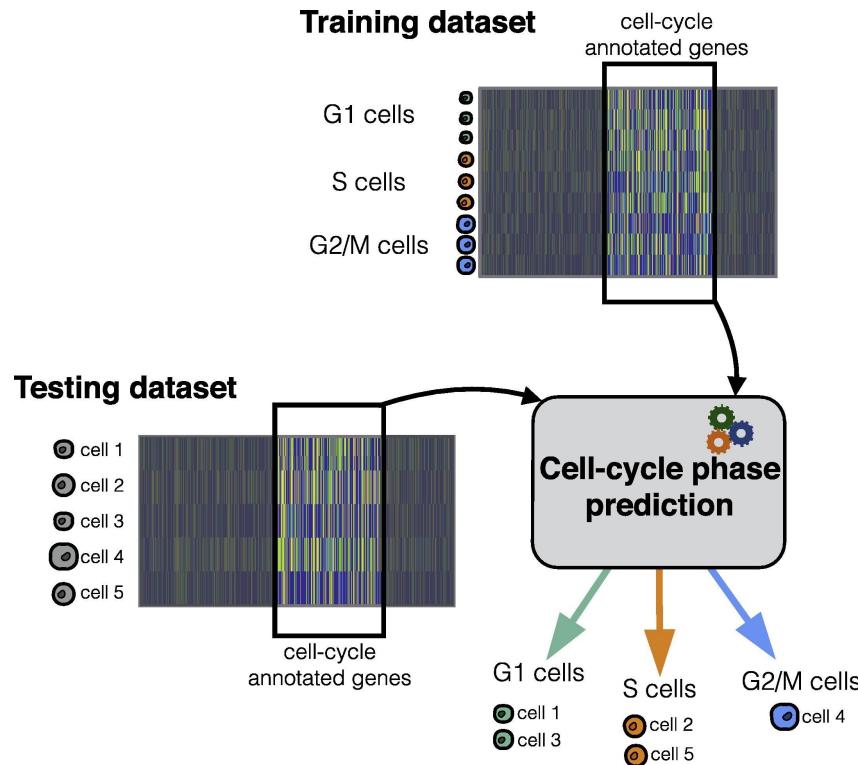
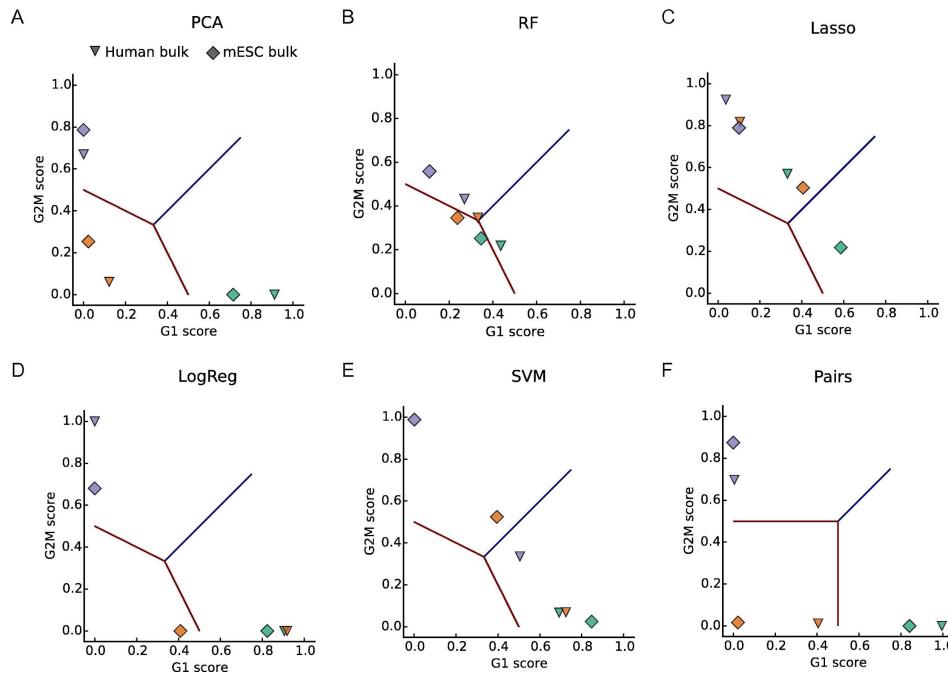
- Regularized negative binomial regression
- Implemented in **sctransform** (*Seurat*)



Hafemeister et al, BioRxiv (2019)

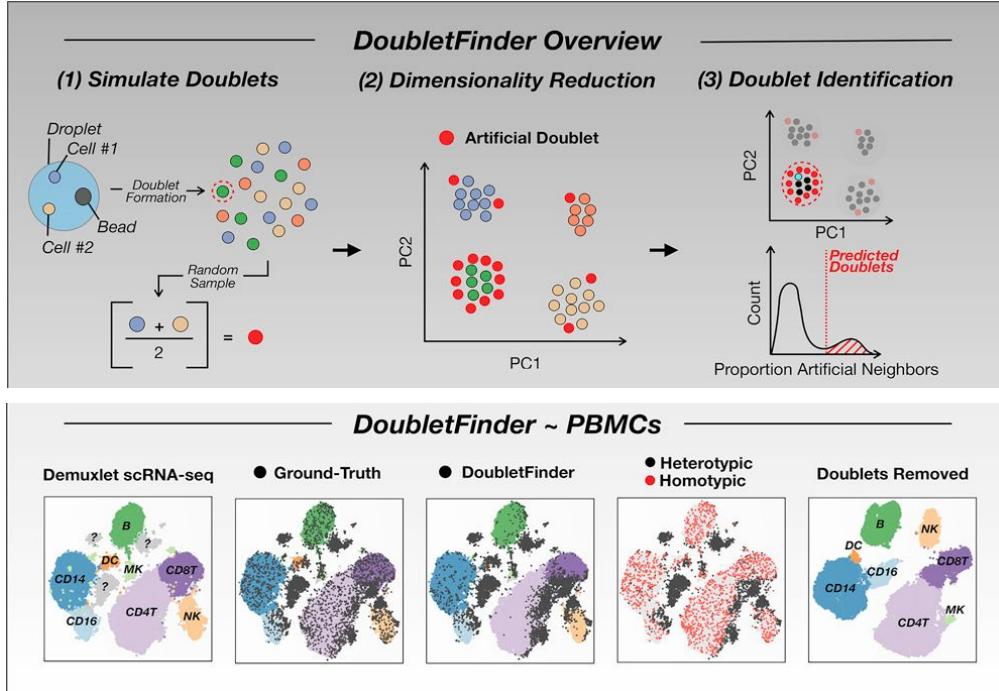
Other factors : Cell cycle phase

- 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** (*scran*)

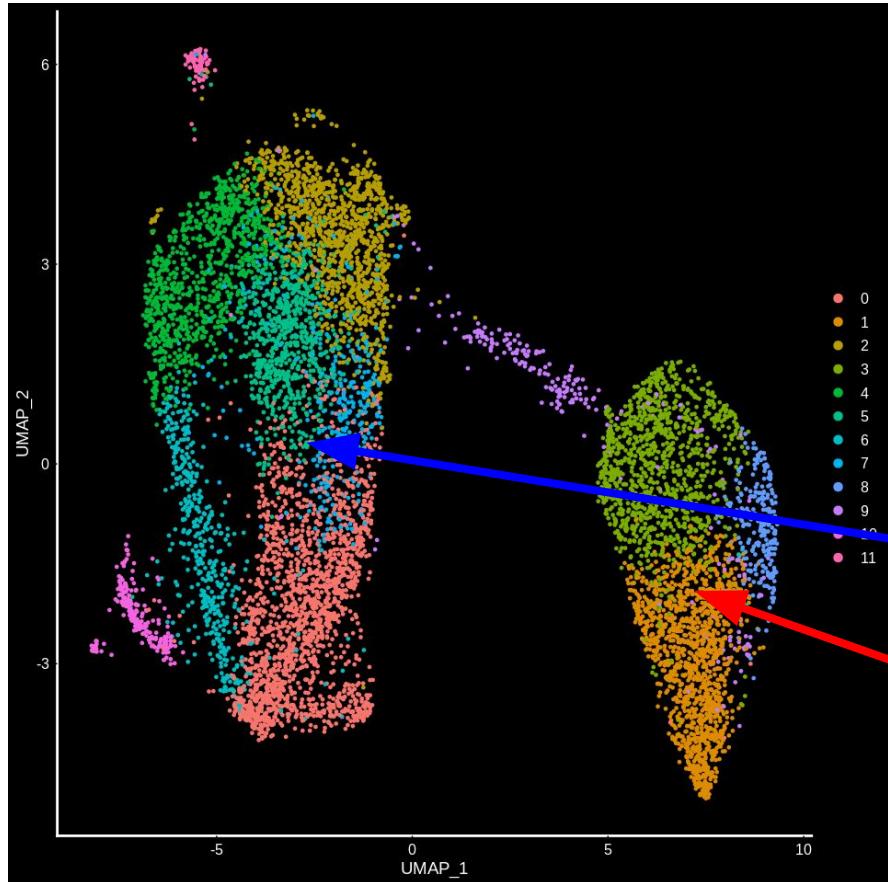


Other factors : Cell doublets

- 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



Normalization : other biological factors

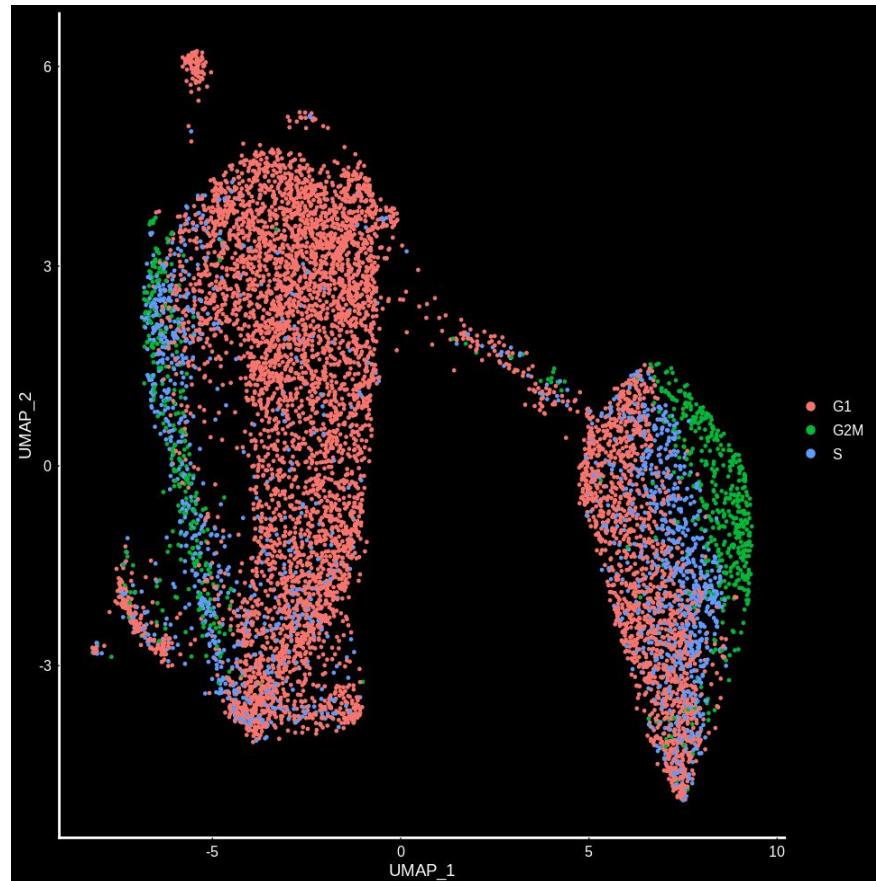
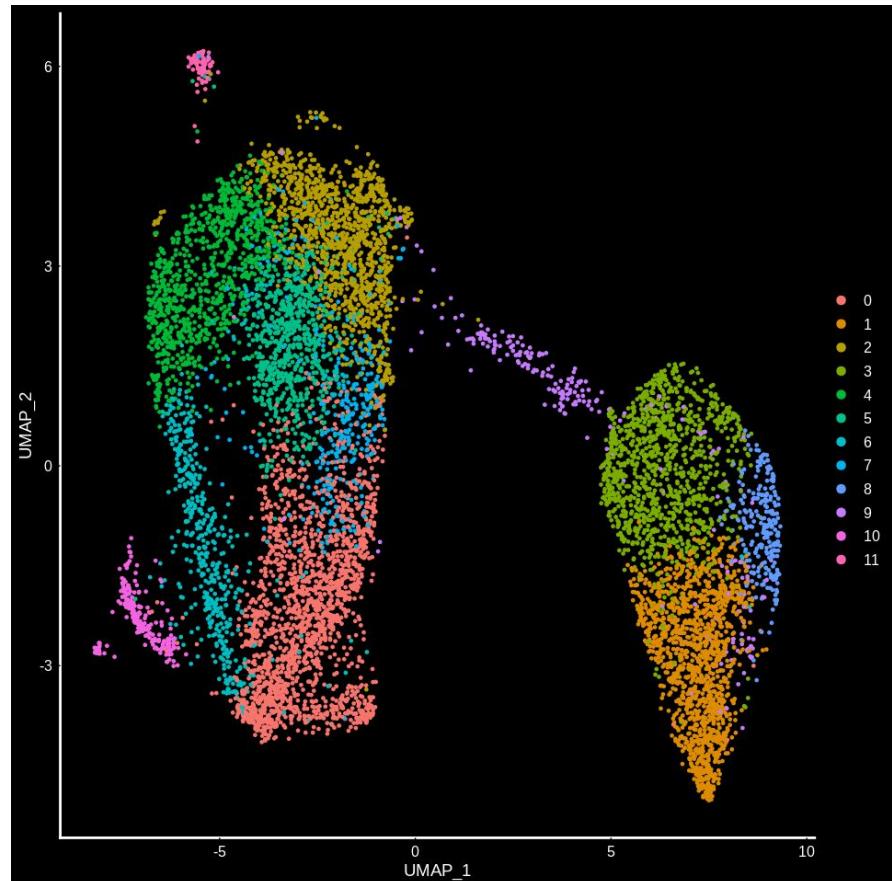


- 10X 3' scRNAseq
- Osteosarcoma metastasis
- 8911 cells x 18613 genes
- PCA (109 PCs retained)
- Louvain clustering
 - 12 clusters
- uMAP representation

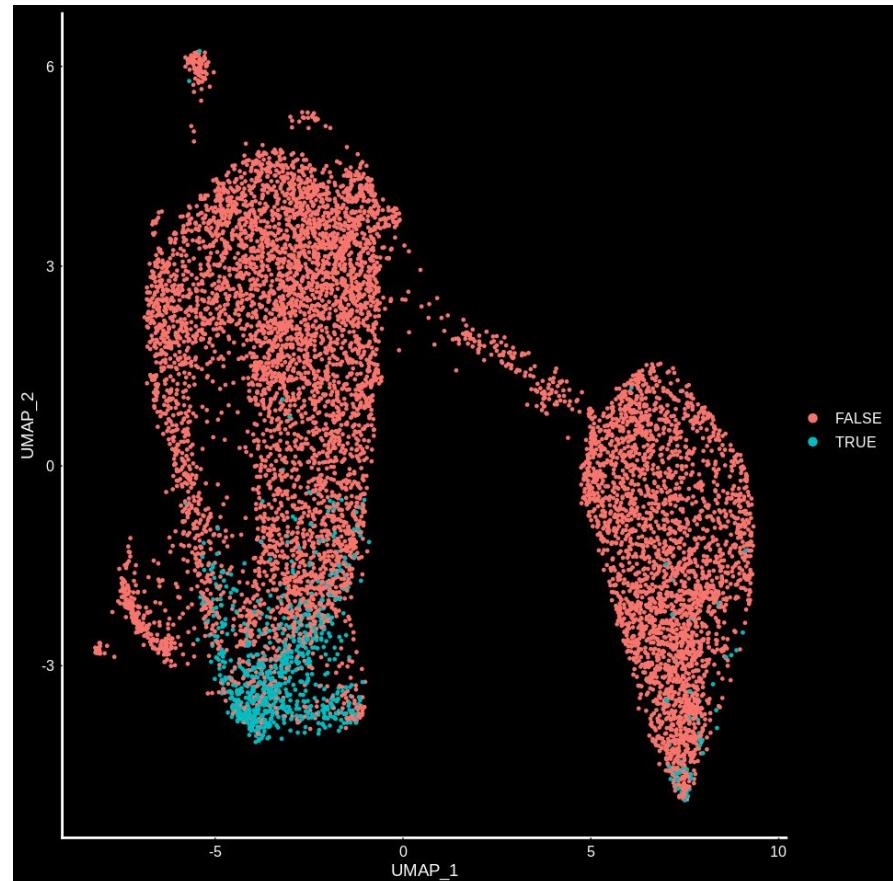
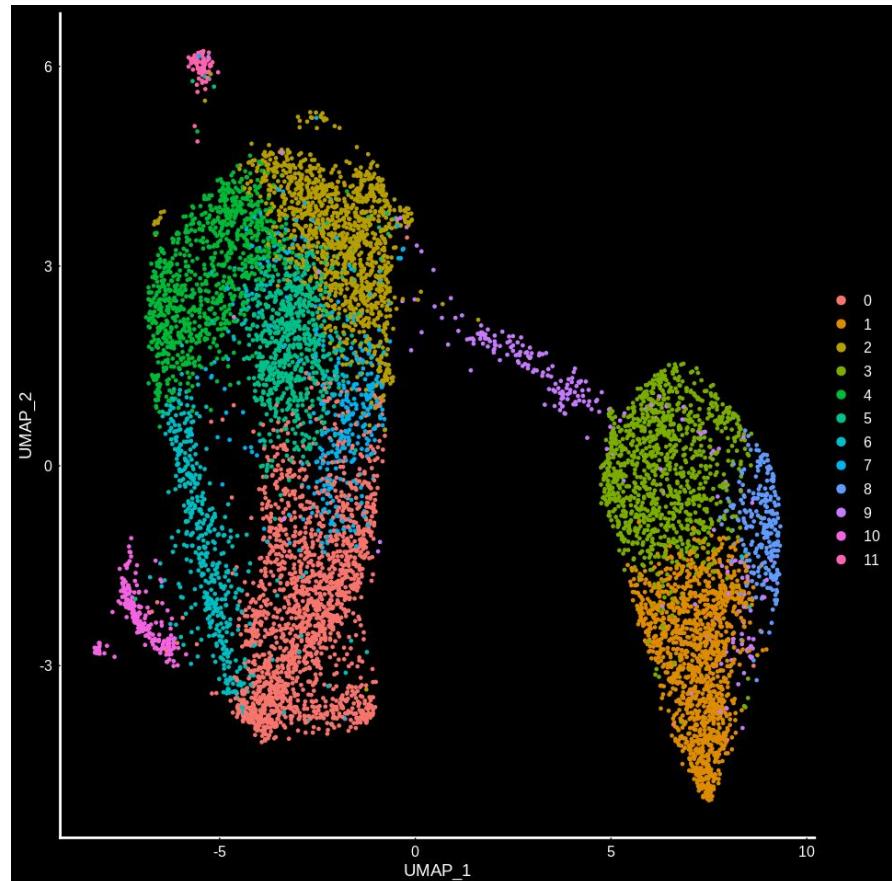
Osteoblasts

Osteoclasts

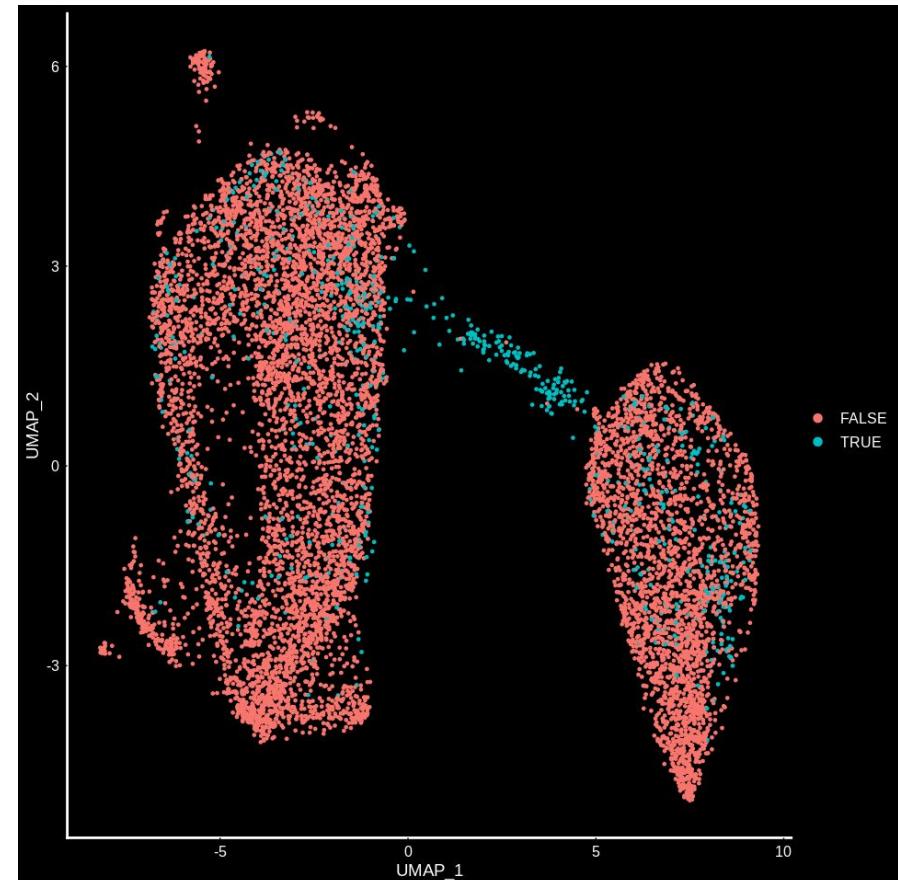
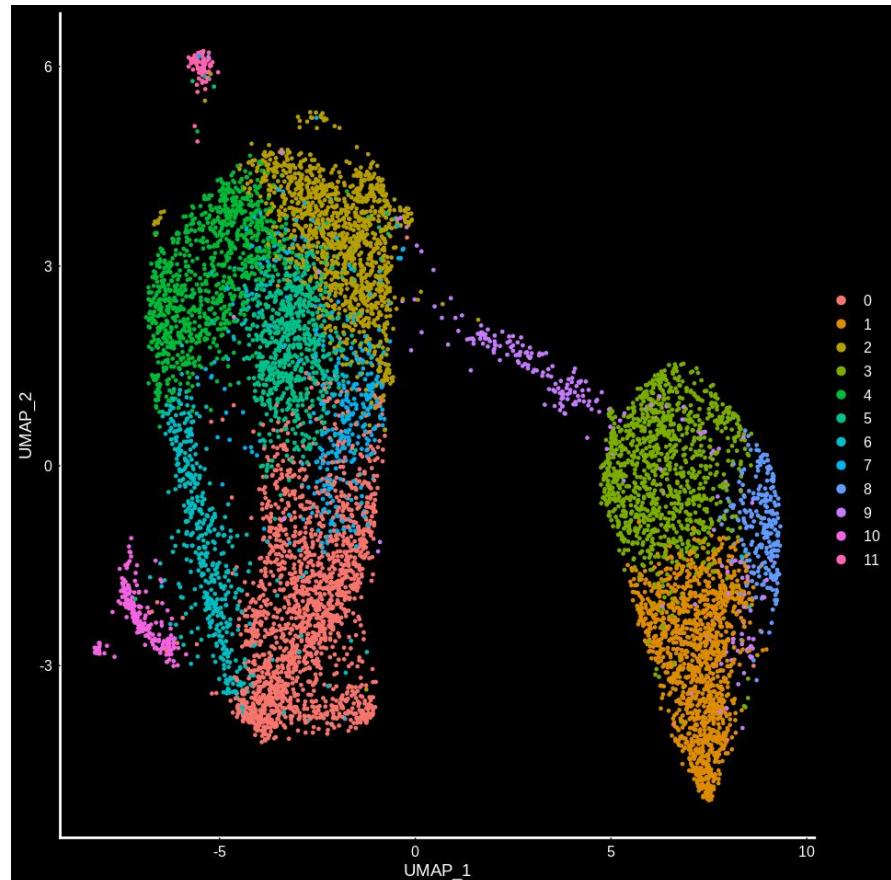
Bias : Cell cycle phases / scores



Bias : Dying cells status / score

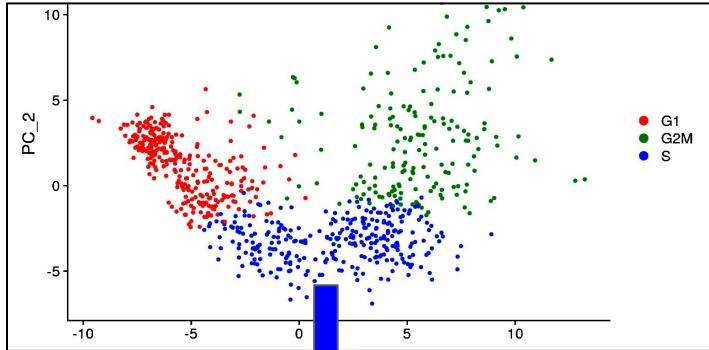


Bias : Cell doublet status / score

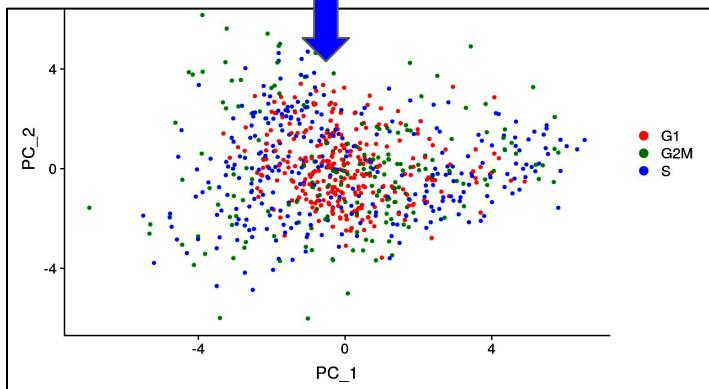


Bias normalization : regression / deblocking

Ex : Cell cycle

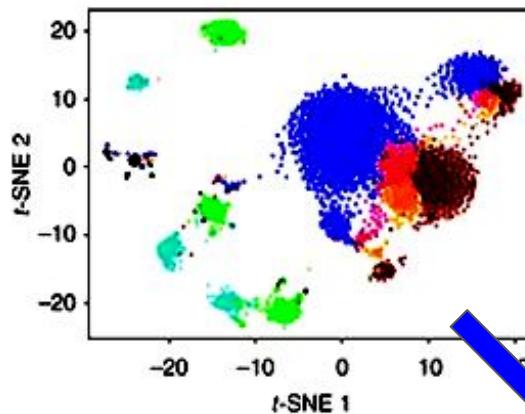


Regression

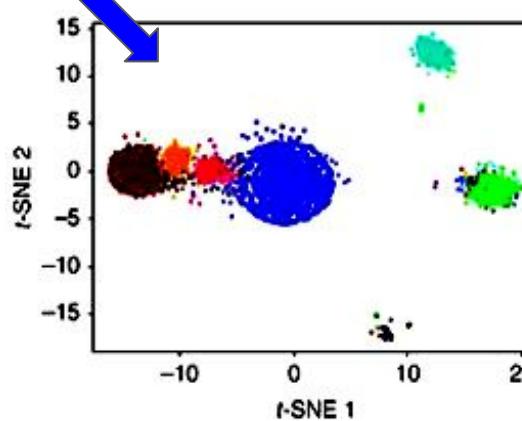


Seurat tutorial

Ex : Batch effect



Deblocking



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

Cell type

- Alpha
- Gamma
- Delta
- Beta
- Ductal
- Acinar
- Other

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)

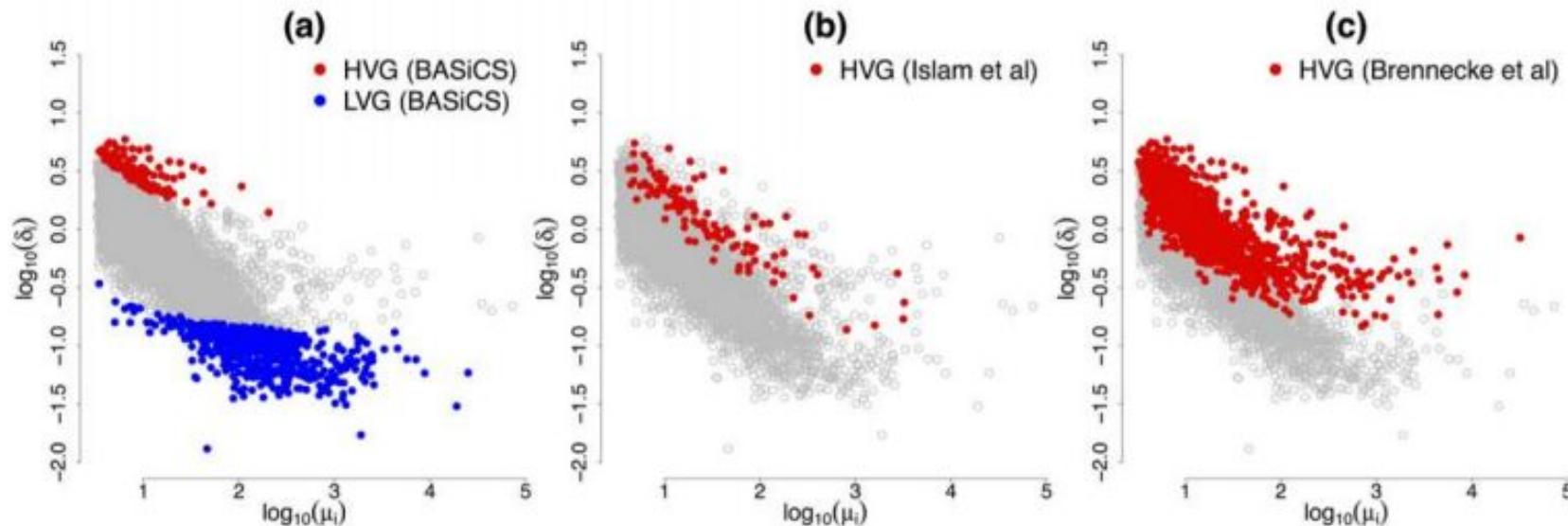


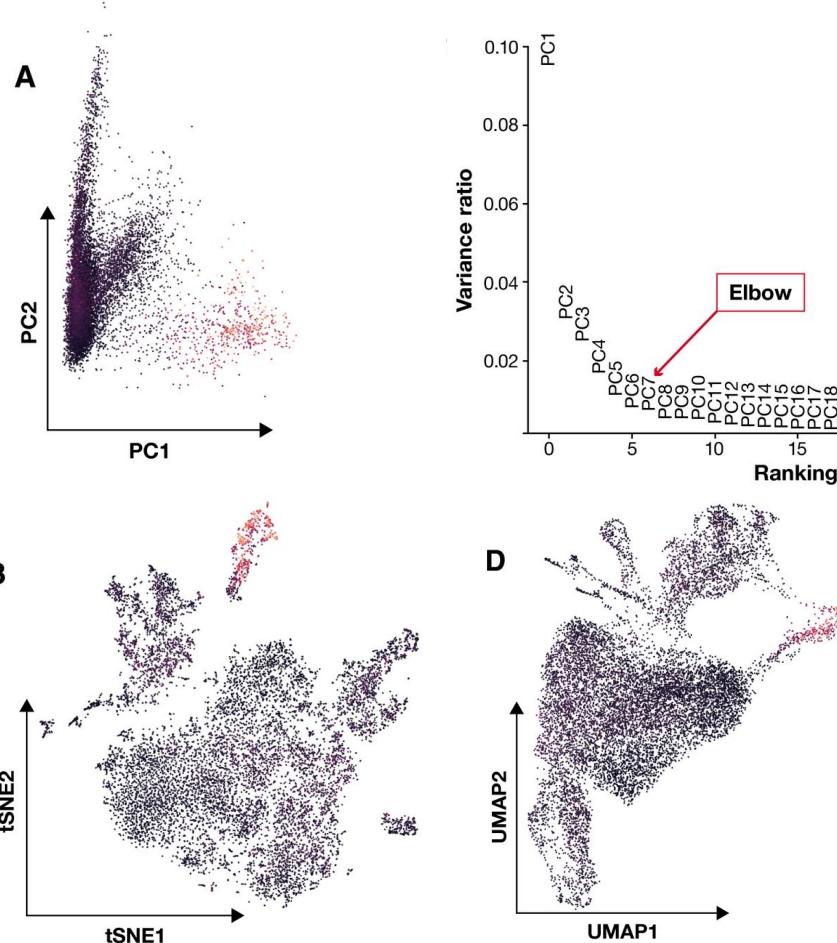
Fig 8. Comparison of HVG detection among different methods. For each of the 7,895 biological genes, posterior medians of biological cell-to-cell heterogeneity term δ_i (log scale) against posterior medians of expression level μ_i (log scale). While the methods described in [16] and [5] only provide a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

Dimensionality reduction : simplification

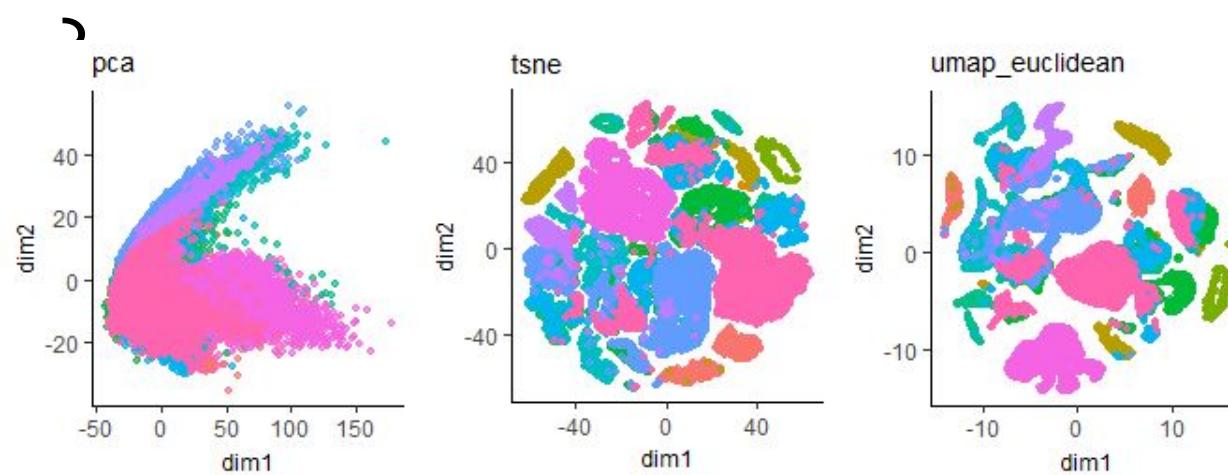
1. Need of an orthogonal space
2. Minimize curse of dimensionality
3. Filter out noise
4. Allow visualization
5. Reduce computational load

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

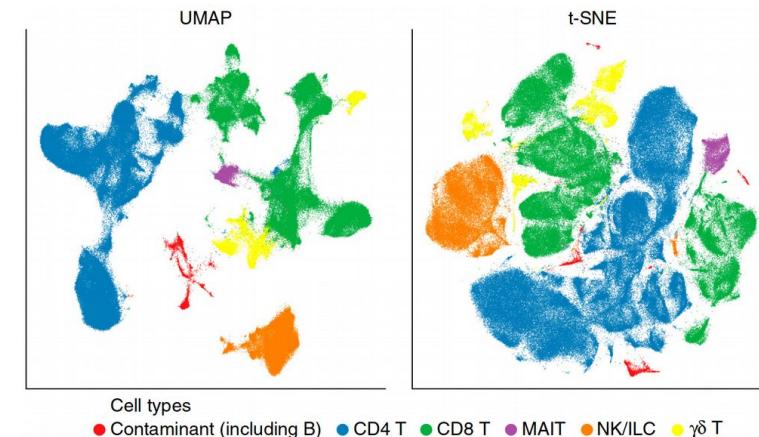


Reduction :

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

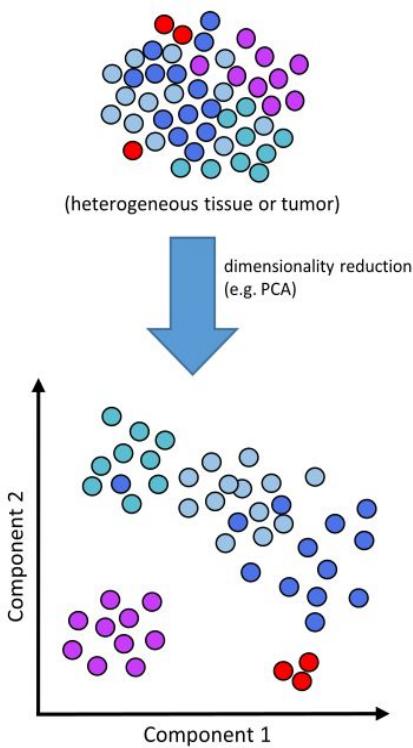
Visualization : uMAP > tSNE

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

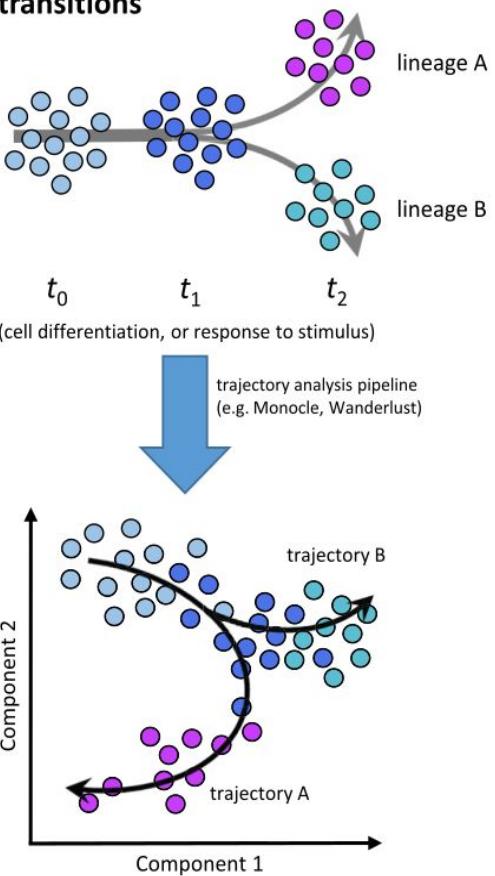


*From a reduced space to ...
... actually what you initially wanted !*

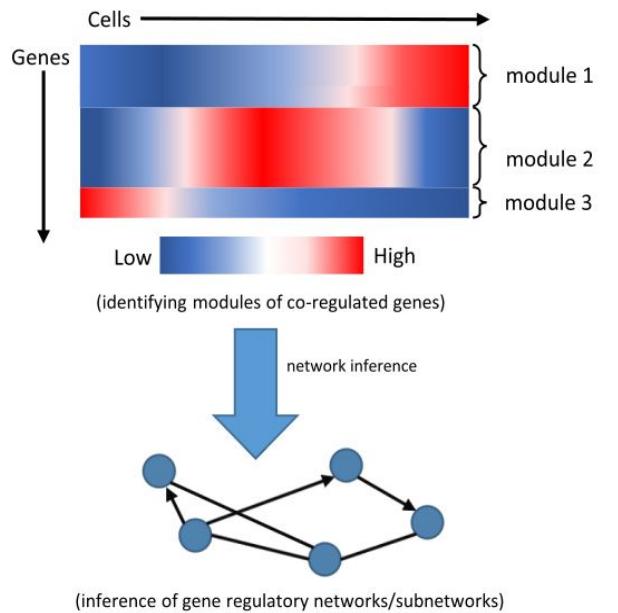
a) Deconvolving heterogeneous cell populations



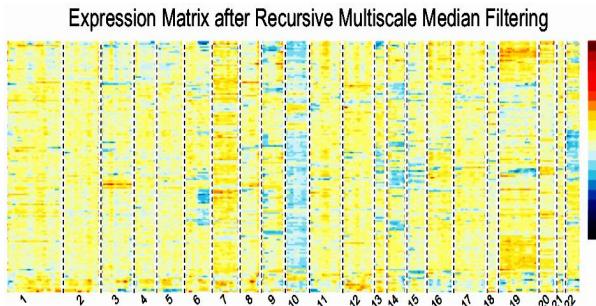
b) Trajectory analysis of cell state transitions



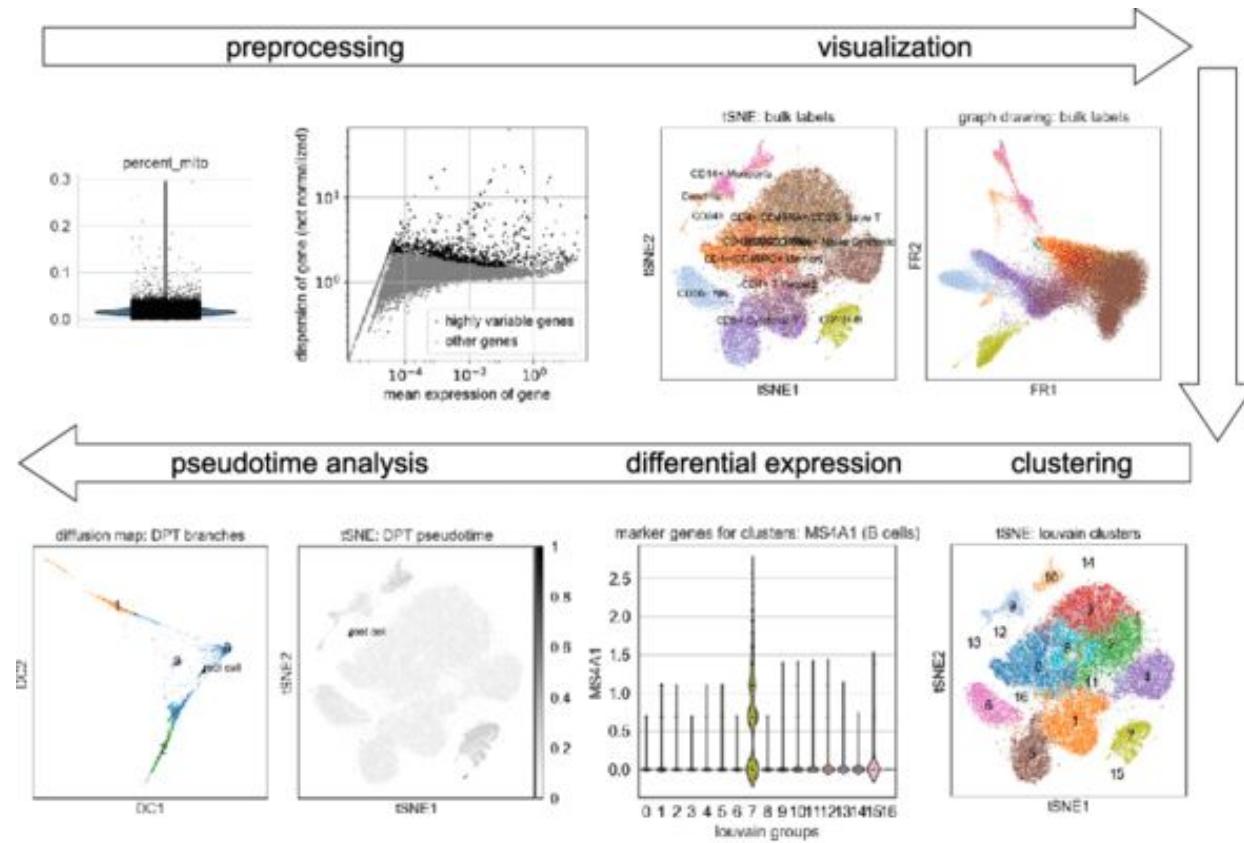
d) Network inference



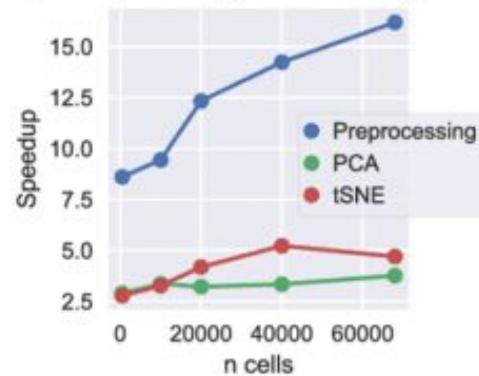
e) Copy number estimation



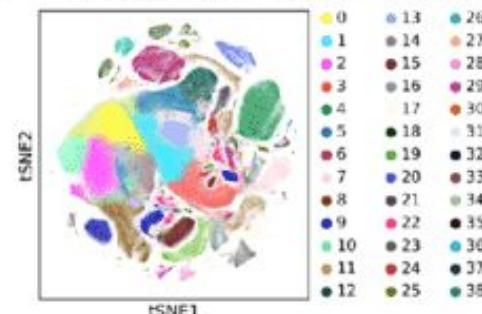
The all-in-one Python toolbox : Scanpy



b Speedup: Scanpy vs. Cell Ranger R



c tSNE of clustered 1.3 million cells



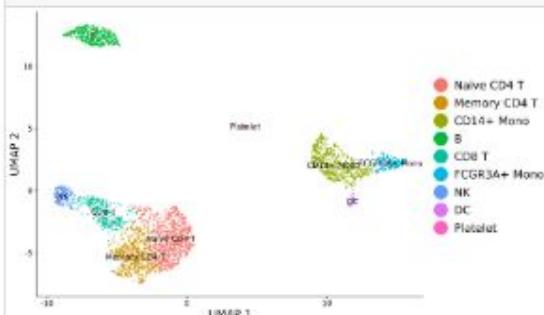
The all-in-one R toolbox : Seurat (v3)

HOME NEWS PEOPLE RESEARCH PUBLICATIONS SEURAT JOIN/CONTACT

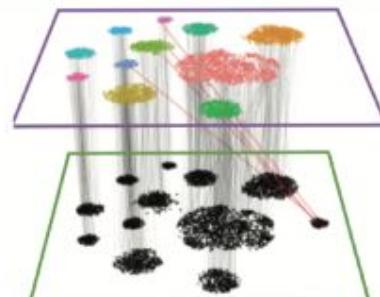
SINGLE CELL
GENOMICS DAY



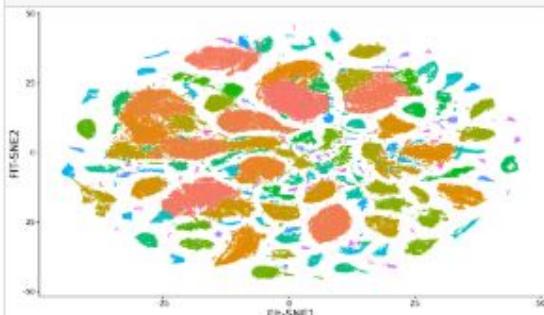
Guided tutorial --- 2,700 PBMCs



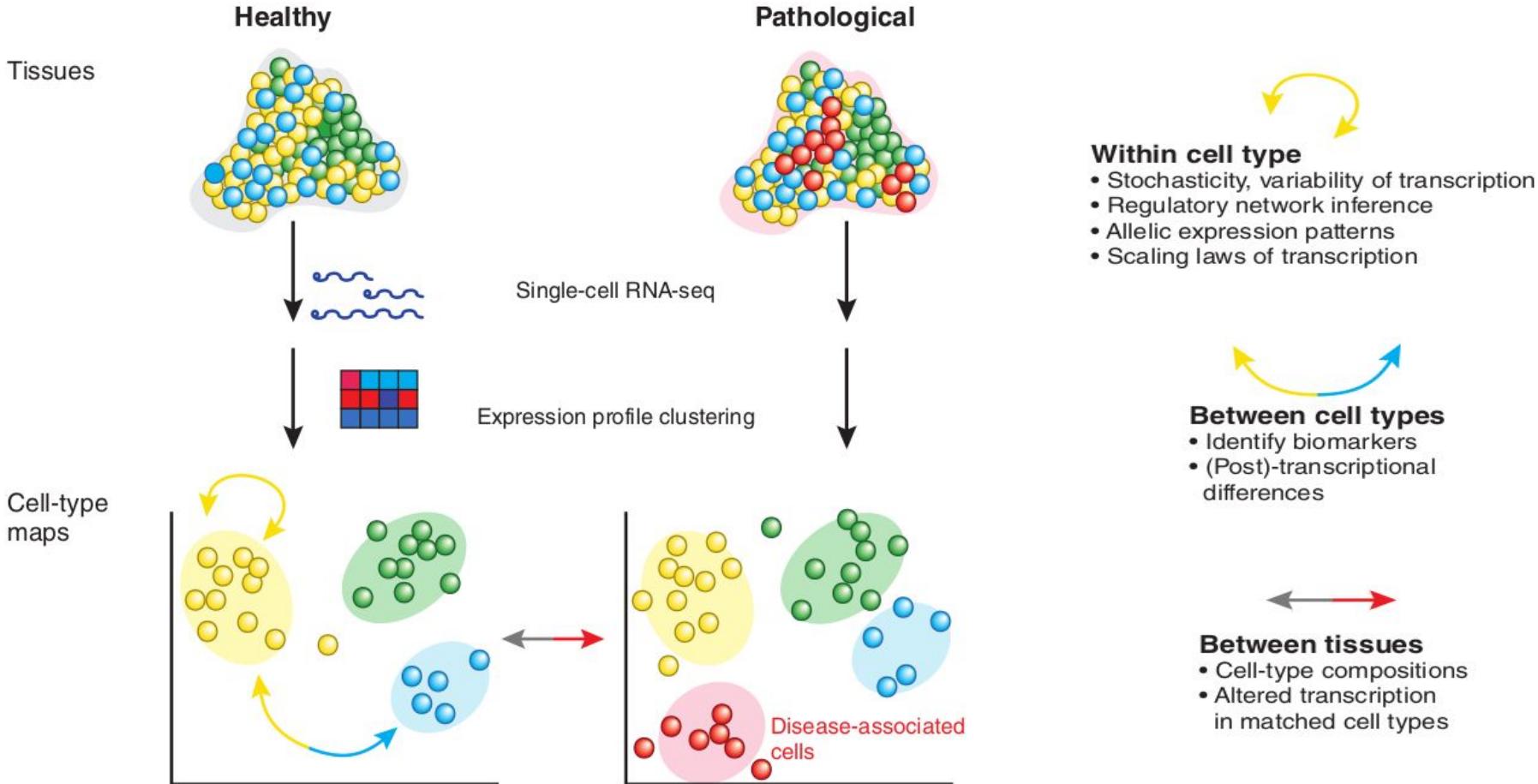
Multiple Dataset Integration and Label Transfer



Mouse Cell Atlas, 250K cells

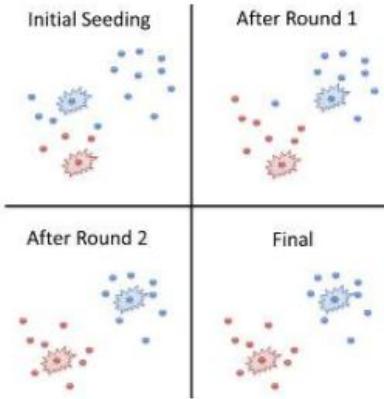


Cell clustering

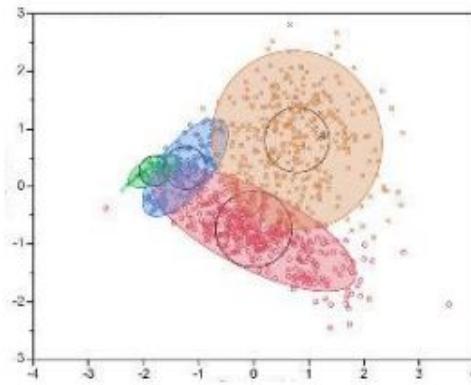


Cell clustering : methods

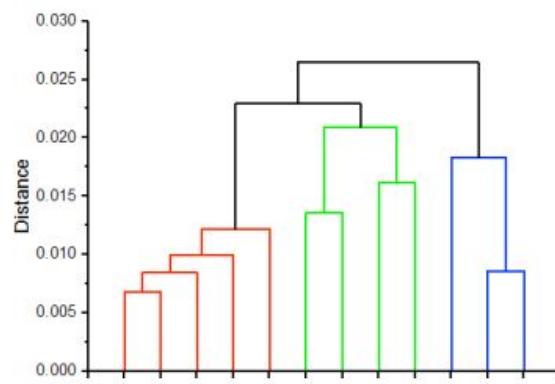
1) K-means based



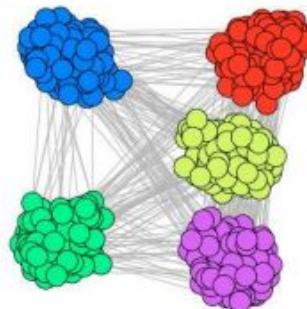
3) Model-based clustering (Mclust)



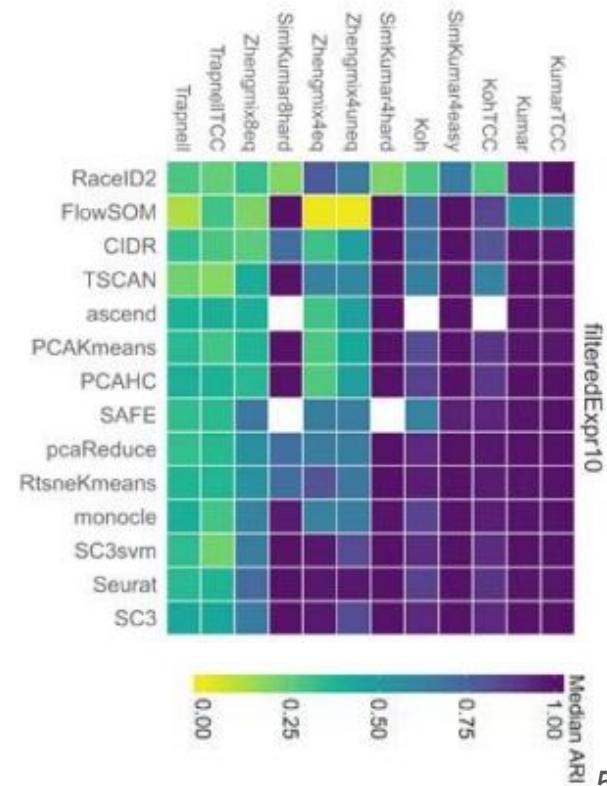
2) Hierarchical clustering



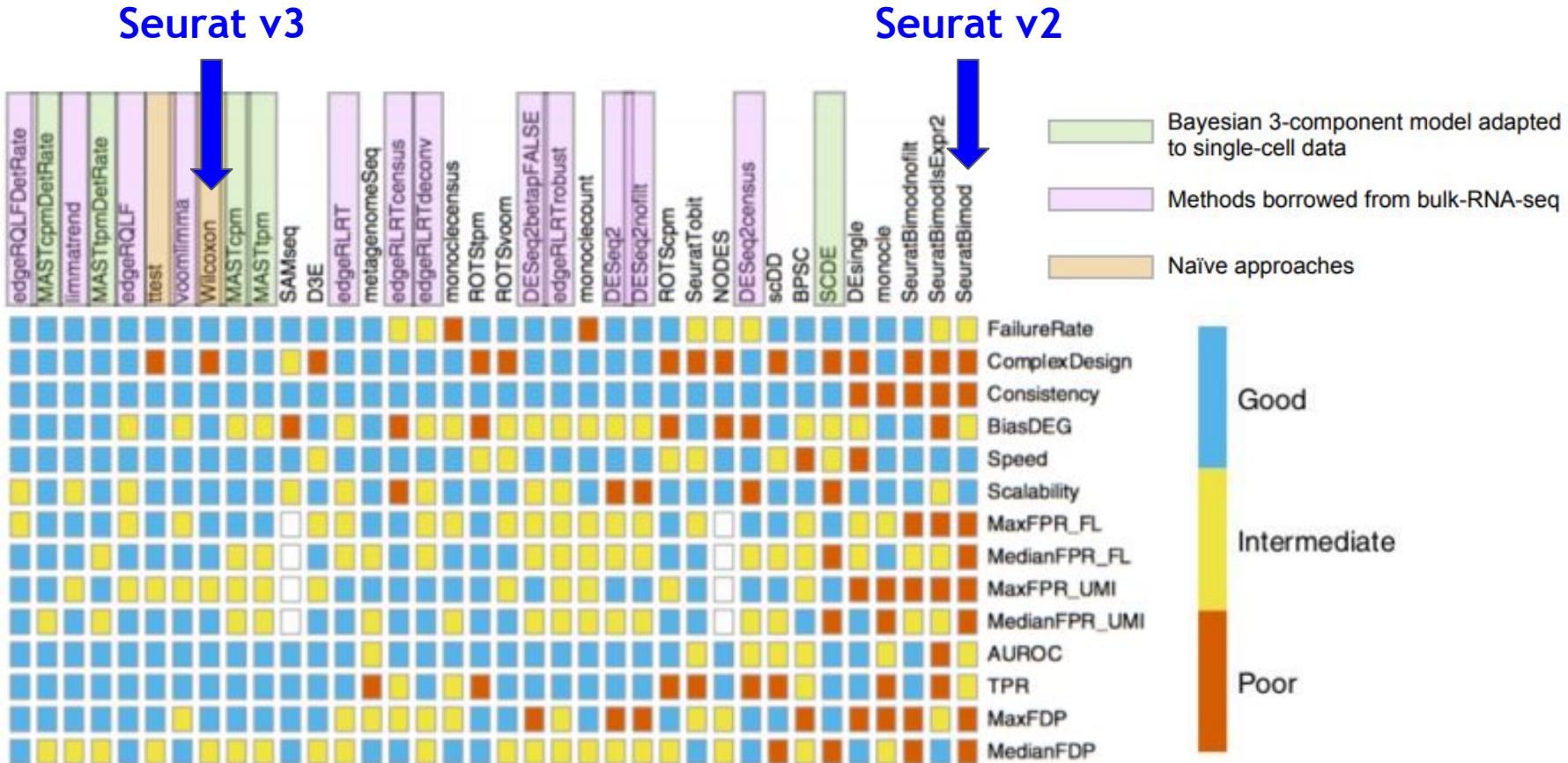
4) Graph-based clustering



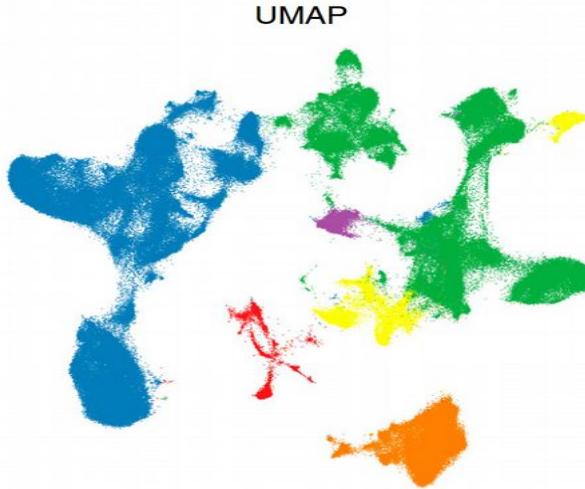
5) Single-cell specific methods



Differential expression analysis



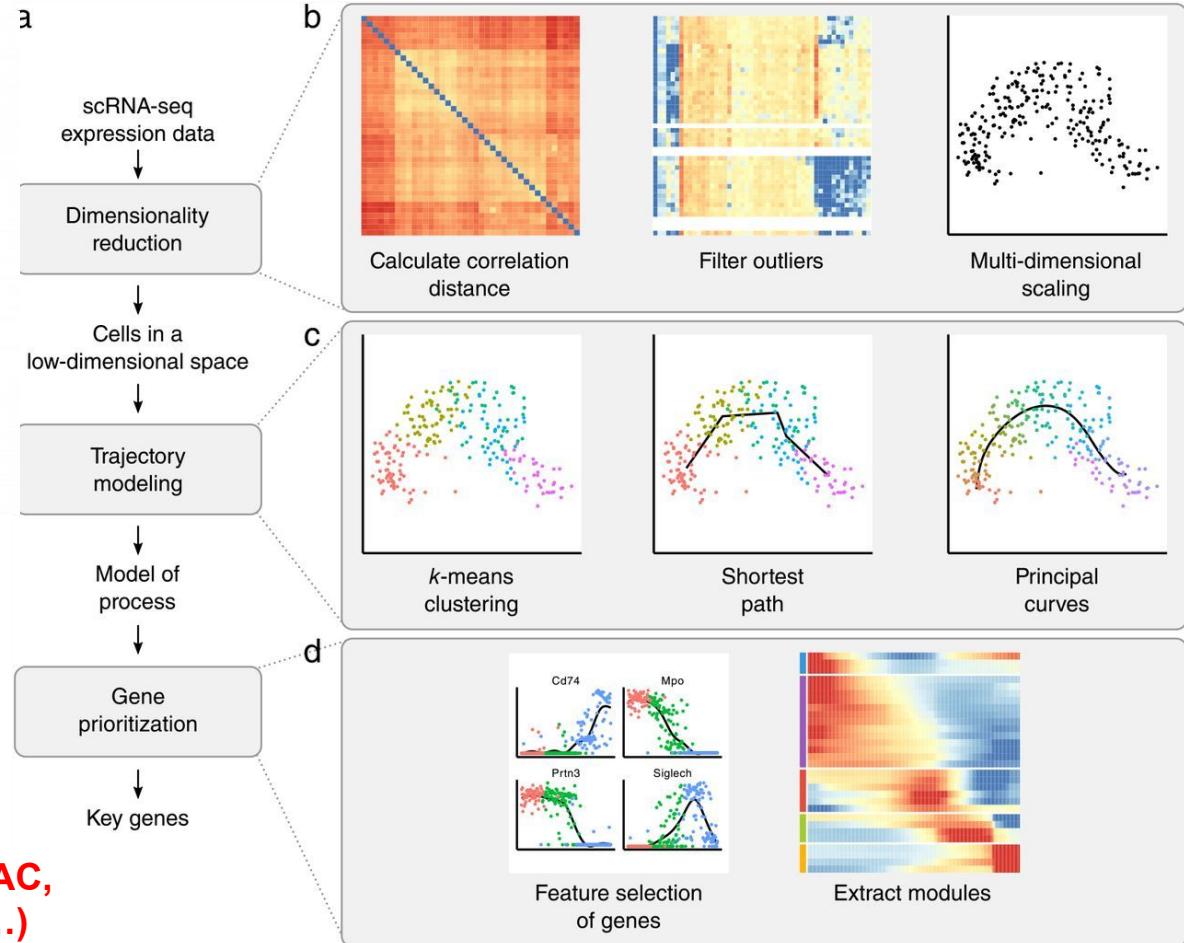
Cell trajectory : methods



Most adopted tools :

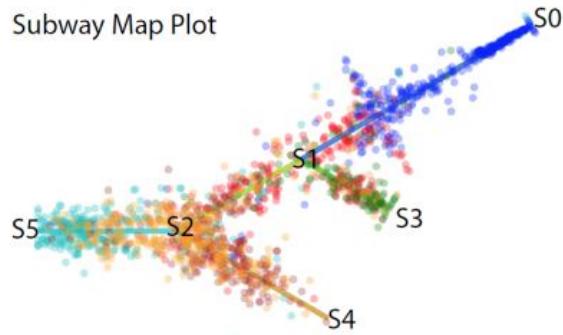
- Monocle 3
- PAGA
- STREAM
- Scorpius
- Slingshot

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

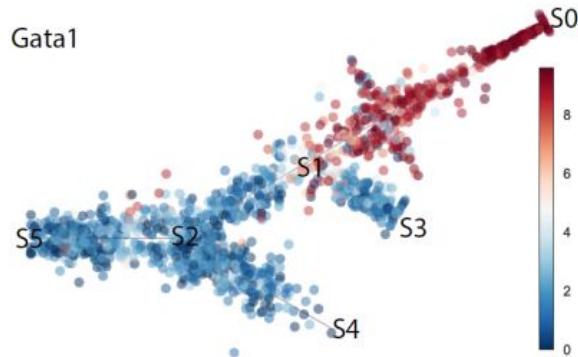


Cell trajectory : visualization

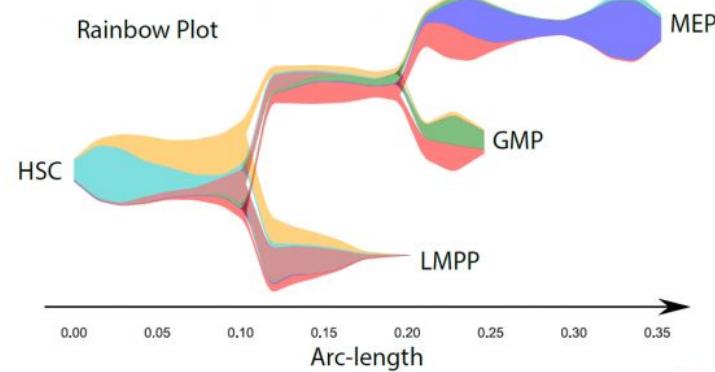
Cell distance to path + cell types



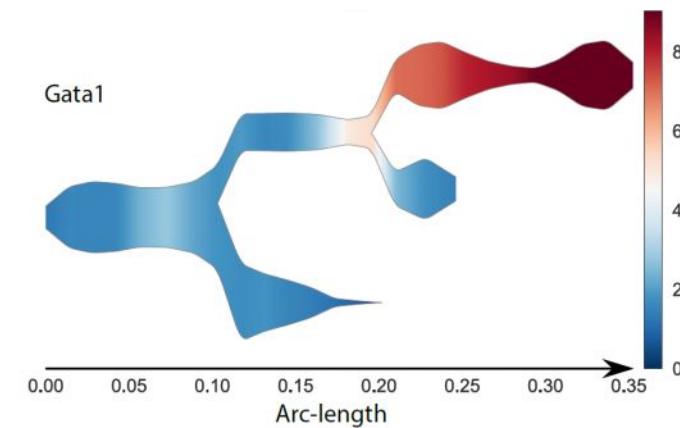
Cell distance to path + gene expression



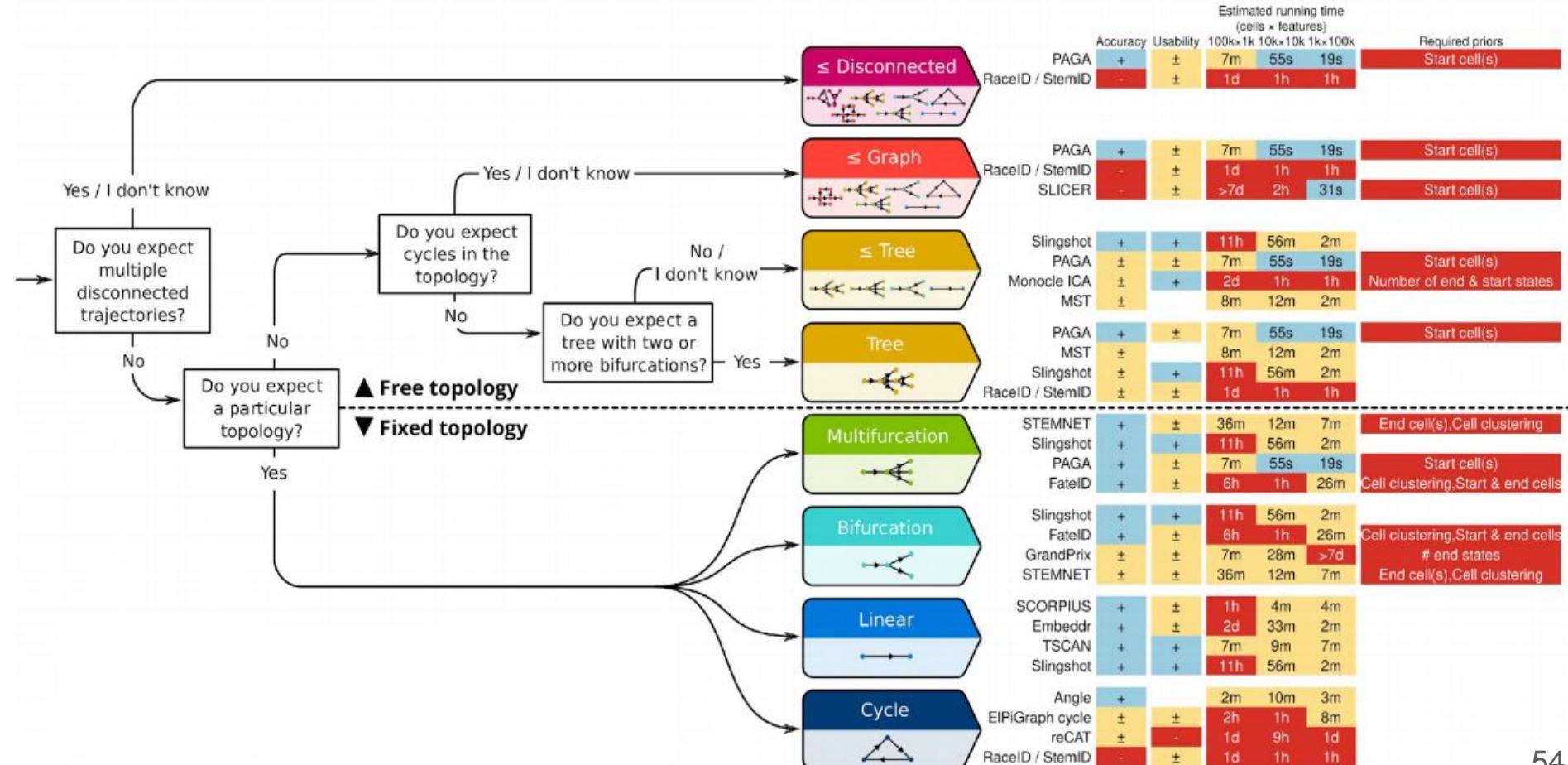
Cell types density



Cells density + gene expression

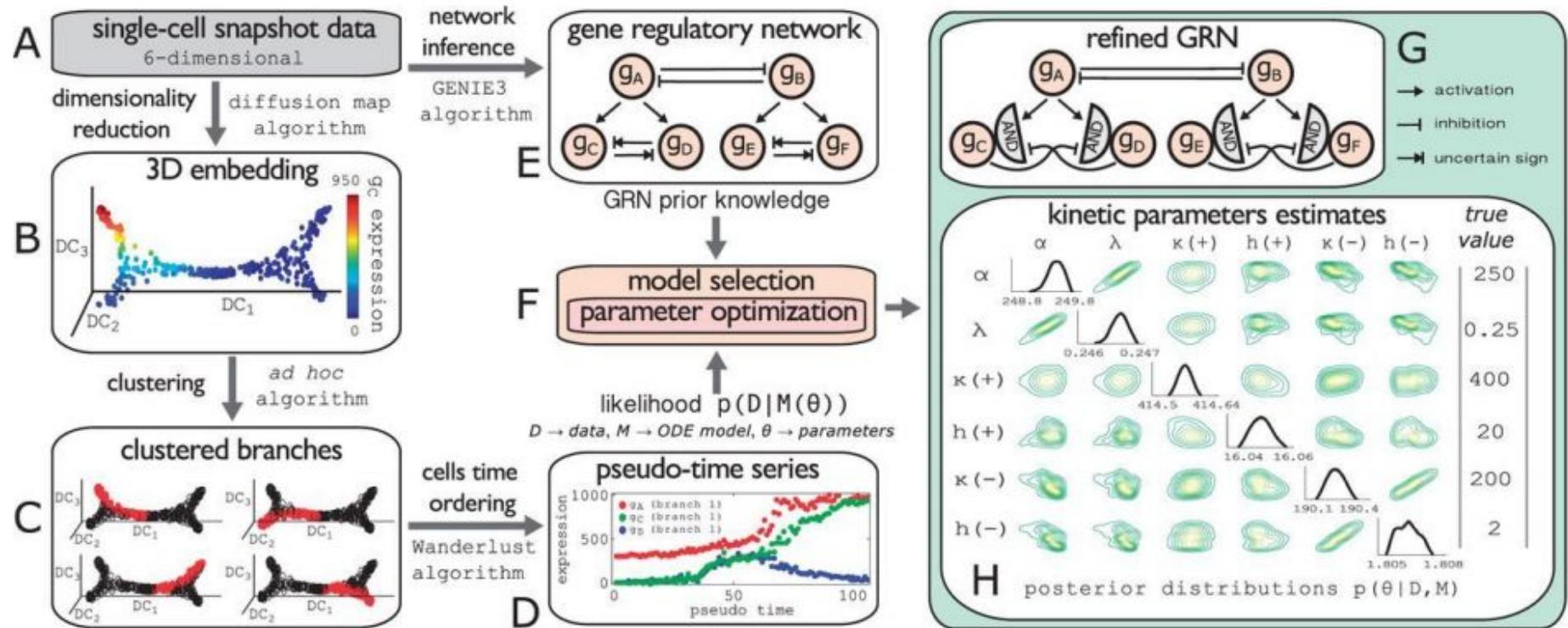


Cell trajectory : Contexts



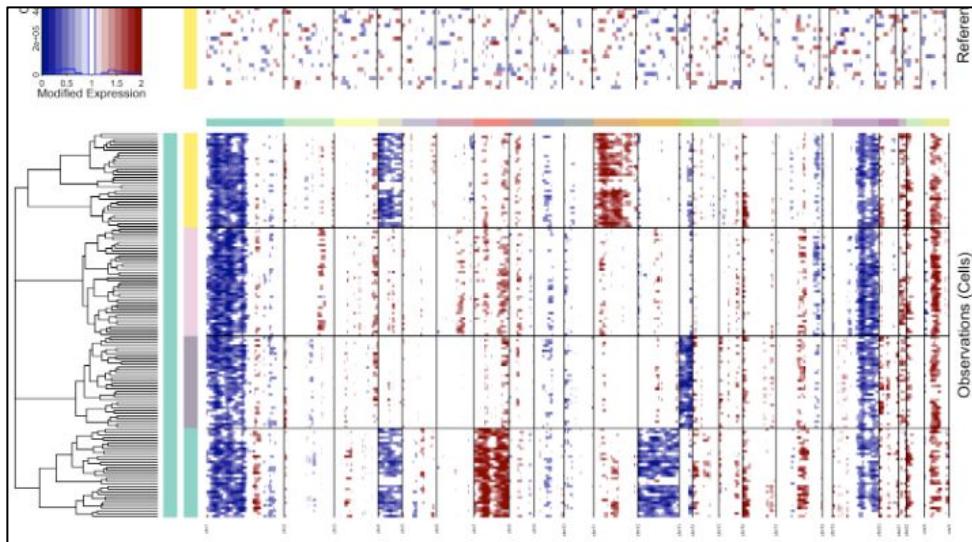
Network inference

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

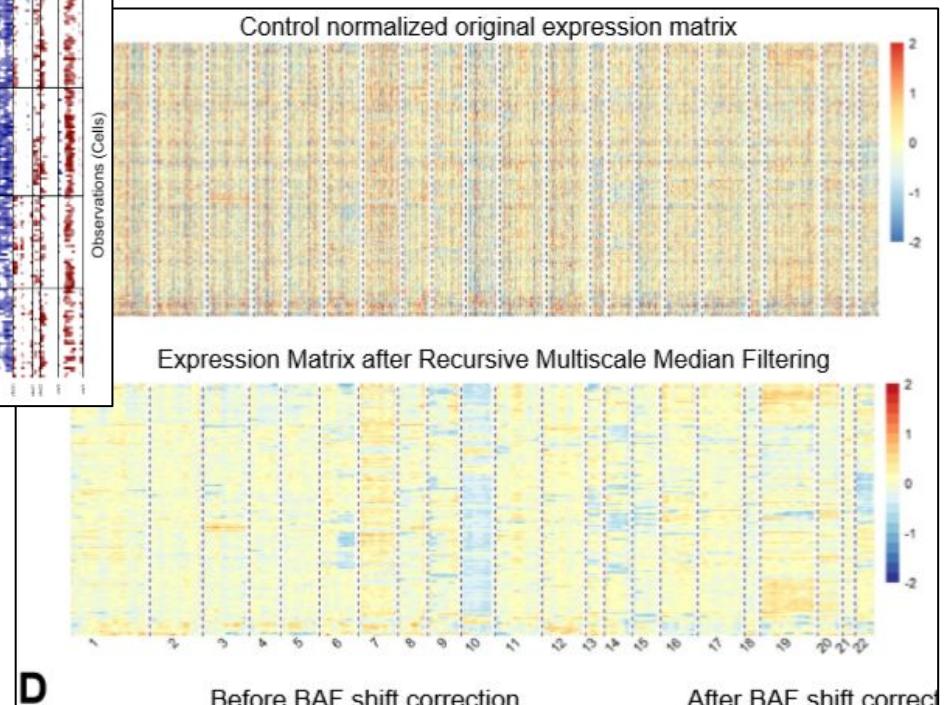


Copy number estimation from scRNASeq

InferCNV (Broad Institute)



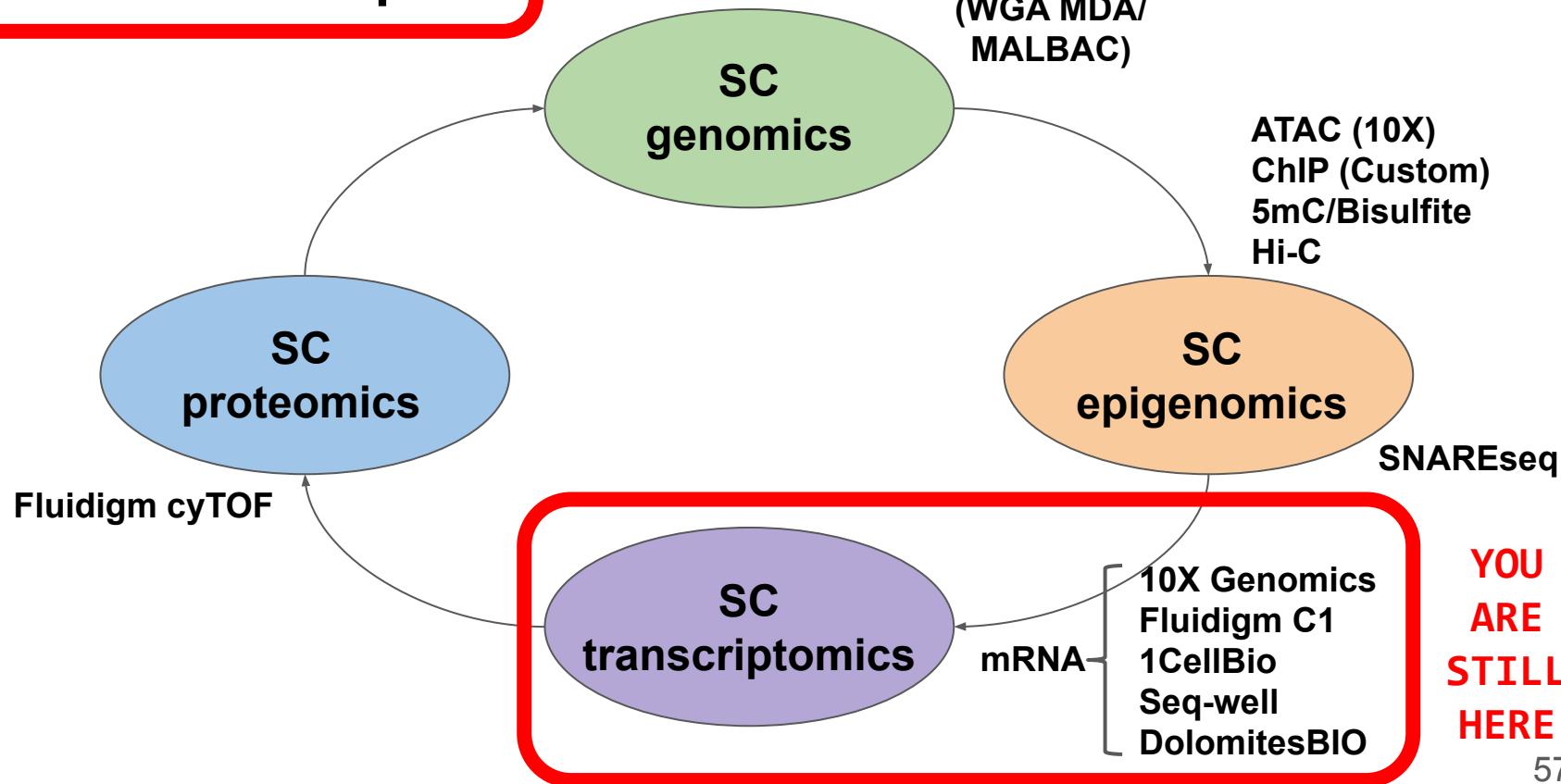
CaSpER (Armanci et al, BioRxiv 2019)



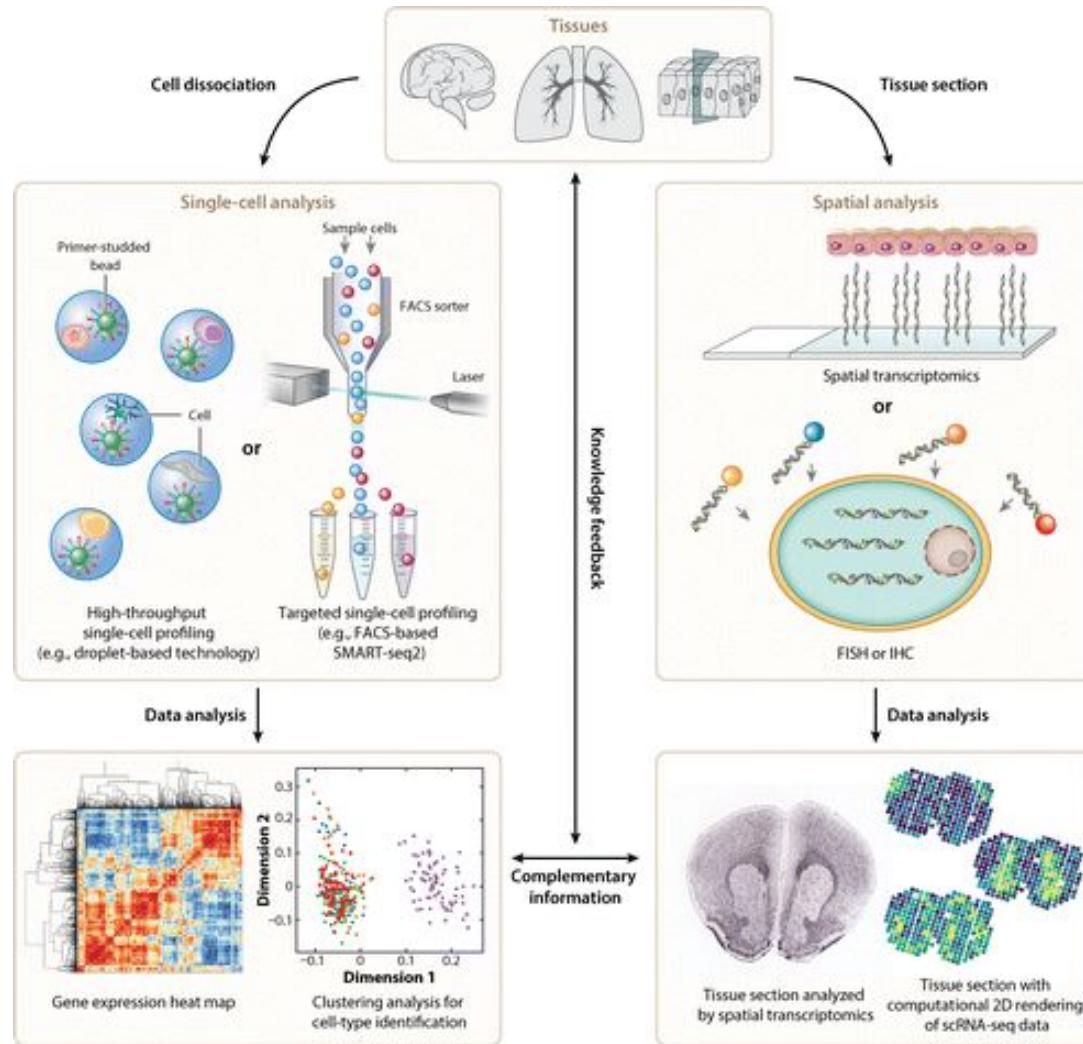
WARNING :

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

Spatial Single Cell RNAseq

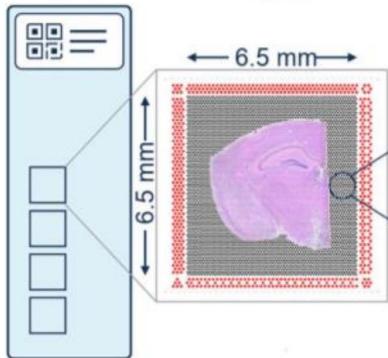


YOU
ARE
STILL
HERE



10x Genomics Visium

Visium Spatial
Gene Expression
Slide



Capture Area with
~5000 Barcoded
Spots

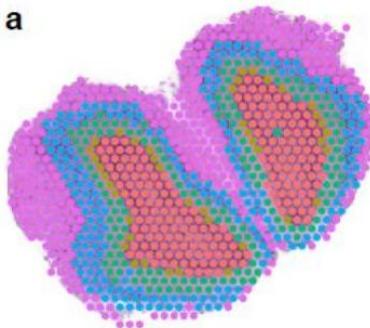
Visium Gene
Expression Barcoded
Spots



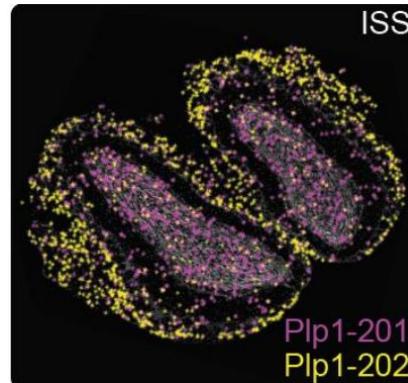
IHC



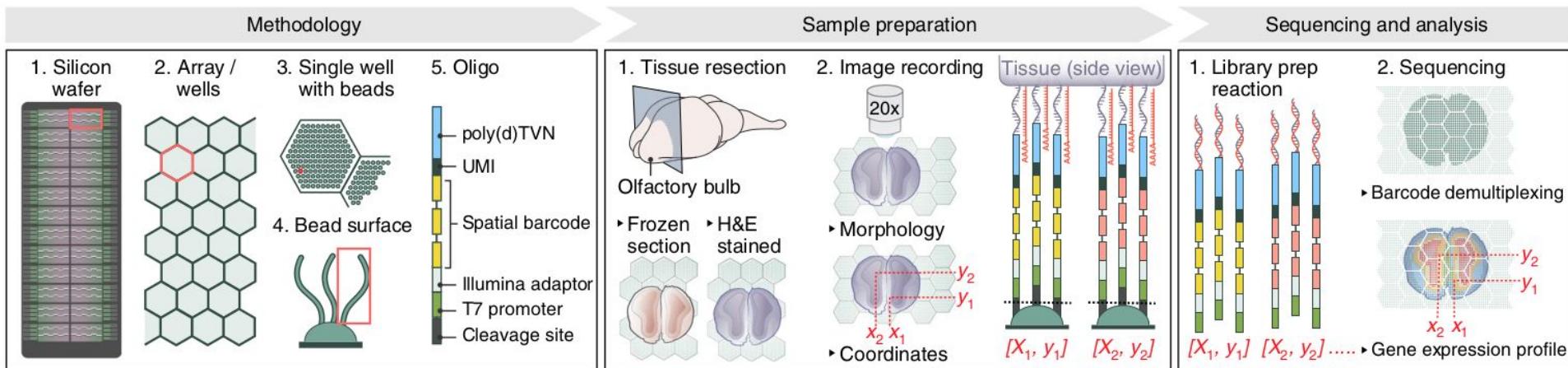
Clusters



RNA FISH

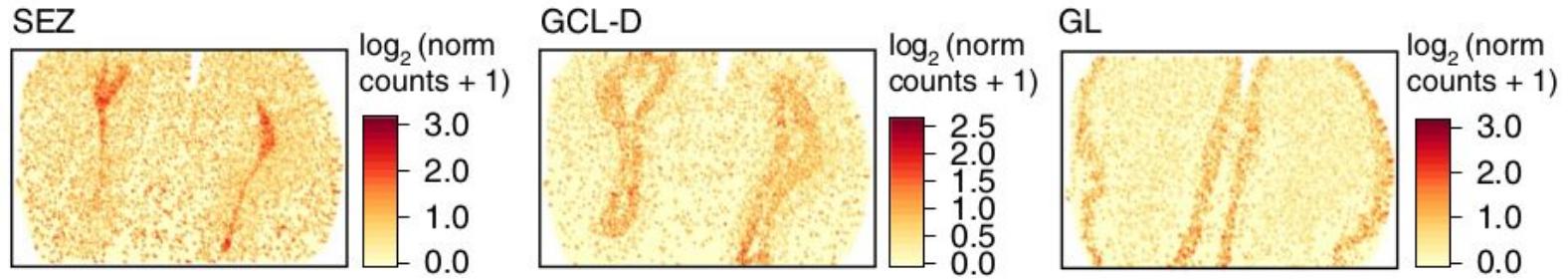


Illumina “HD Spatial Transcriptomics”

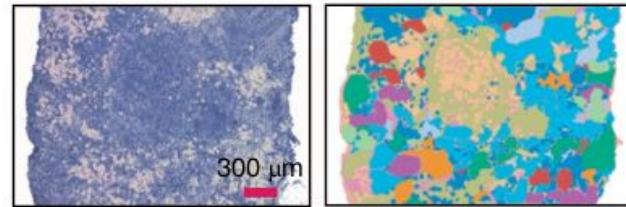


- 2,893,865 individual barcoded beads
- 1.4 M wells
- Well diameter $\sim 2 \mu\text{m}$
 - << median cell diameter (20 μm)
 - $\sim 1,400 \times$ higher resolution than “standard” ST
 - $\sim 25 \times$ compared to SLIDE-seq
- Array reading time $\sim 3 \text{ H}$
- Challenging analysis strategy (low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

Illumina HDST



H&E Annotations



- Fatty tissue, immune/lymphoid
- 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

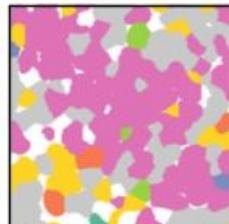
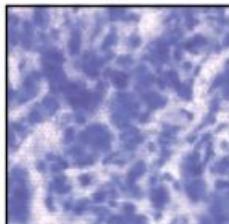
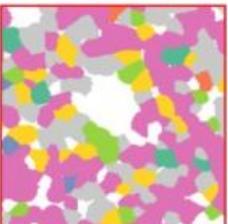
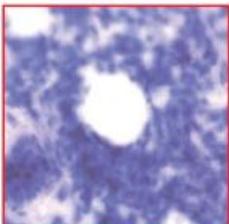
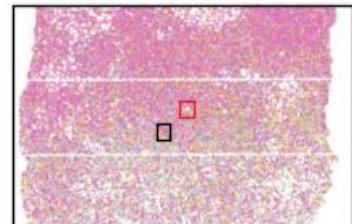
Cell types
in sn-like data

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

CNV

SC
proteomics

Fluidigm cyTOF

SC
genomics

WES
CNV
VDJ/TCR
(WGA MDA/
MALBAC)

NOW, YOU
ARE HERE

SC
transcriptomics

mRNA

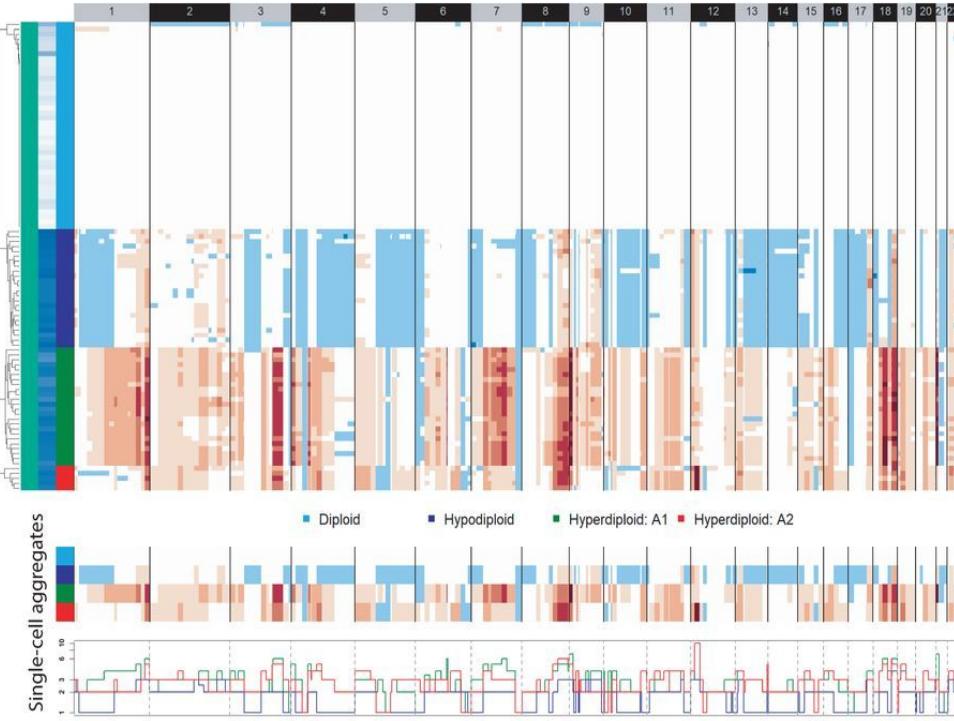
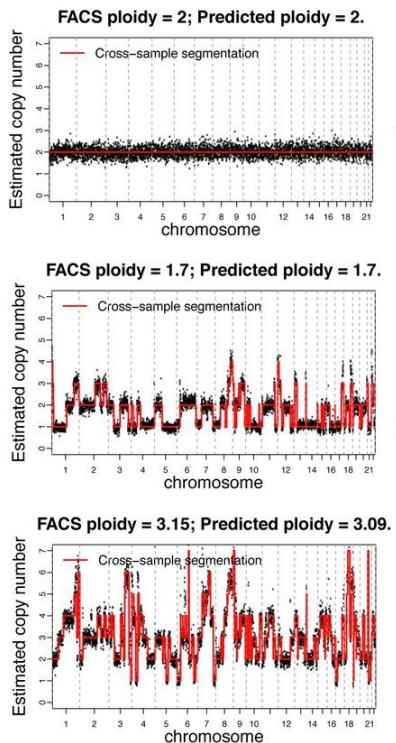
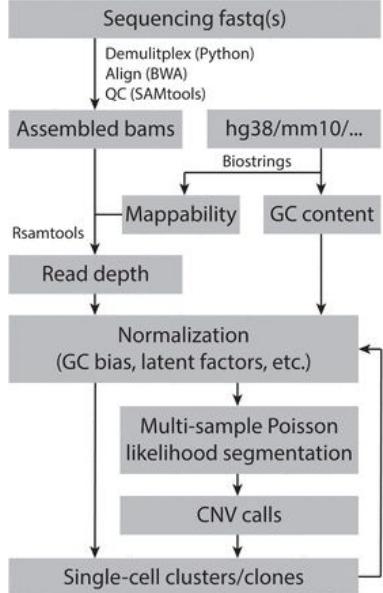
- 10X Genomics
- Fluidigm C1
- 1CellBio
- Seq-well
- DolomitesBIO

SC
epigenomics

SNAREseq

ATAC (10X)
ChIP (Custom)
5mC/Bisulfite
Hi-C

scCNV results (SCOPE)

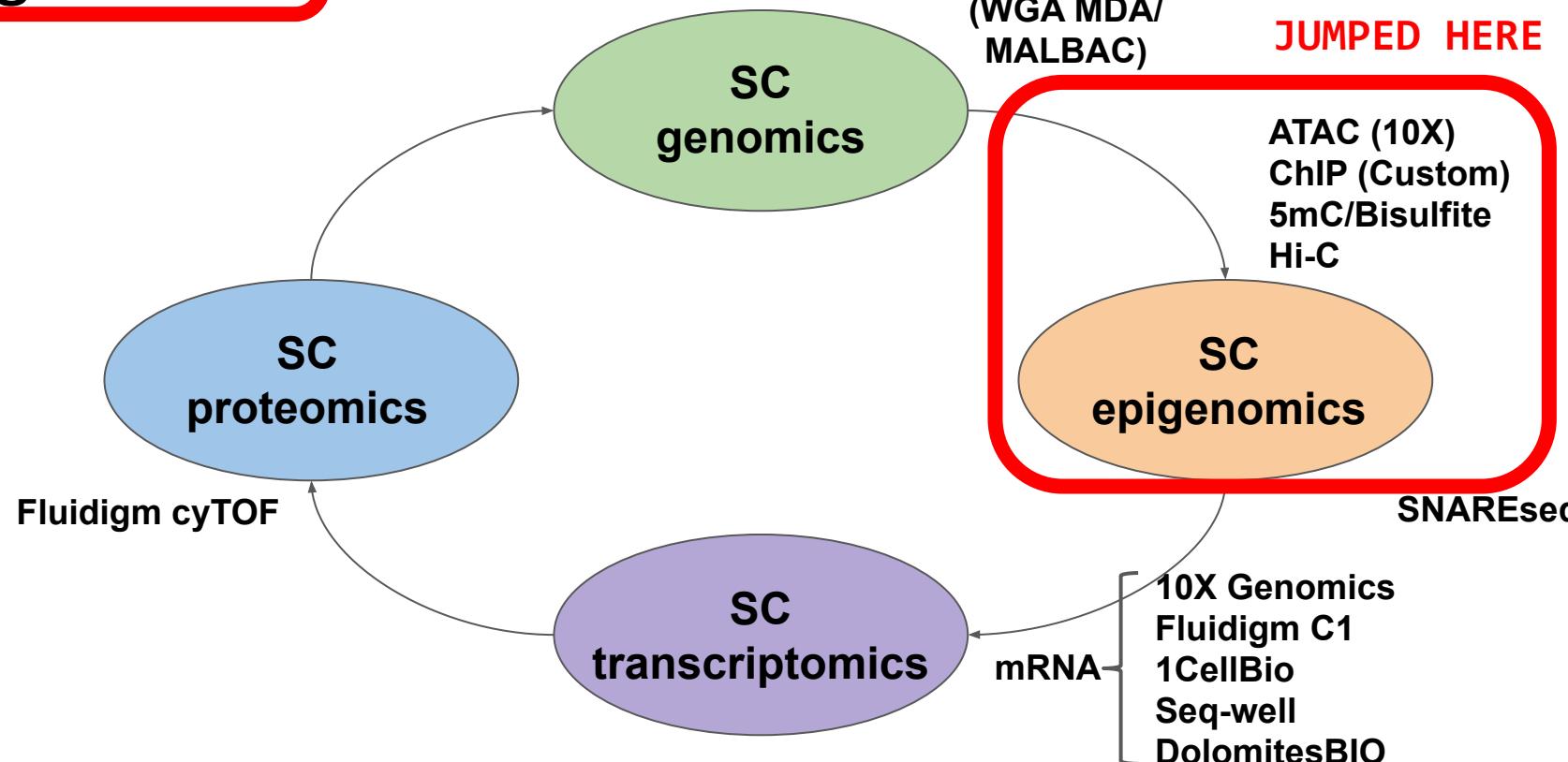


WARNING :

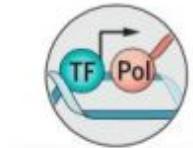
- Limited resolution : > 2 Mb (binning)
- Requires > 750,000 reads / cell

ALSO : SCYN

Single Cell Epigenomics

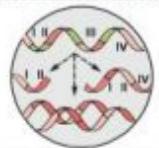


Overview of scEpigenomics techniques

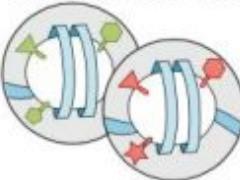


Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility



Transcription and RNA maturation



Histone modifications

Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27me3 in red)



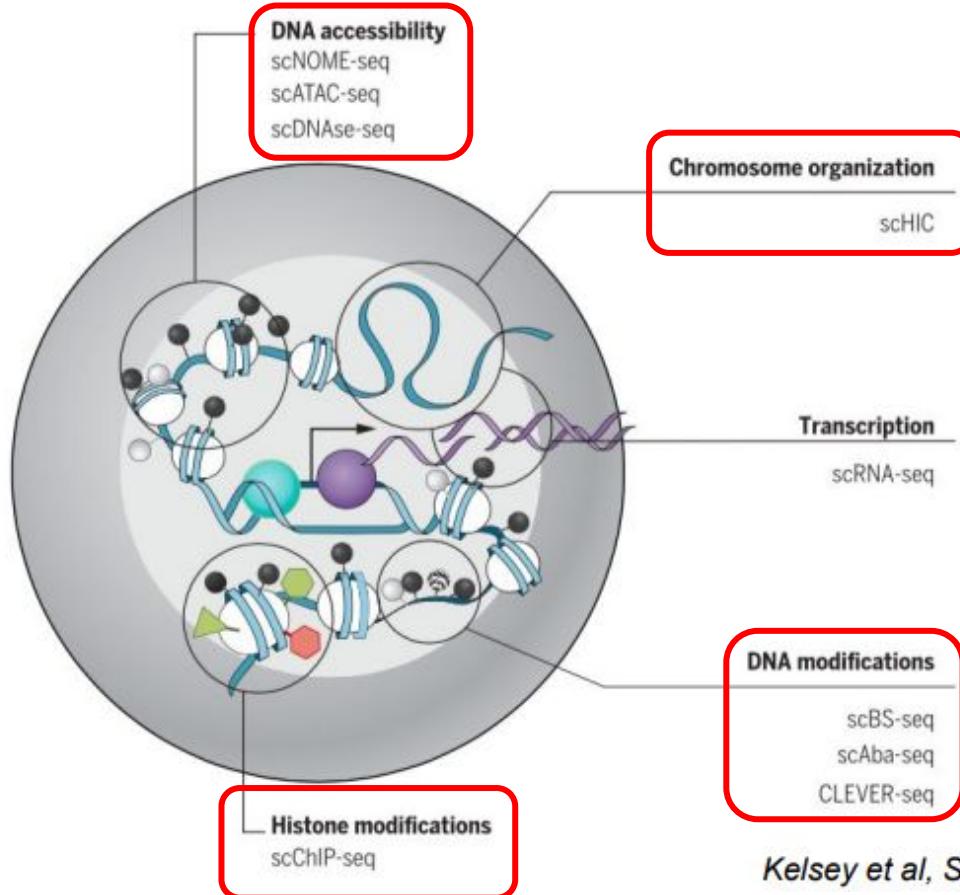
DNA modifications

● C ● 5mC
● 5hmC / 5fC / 5caC

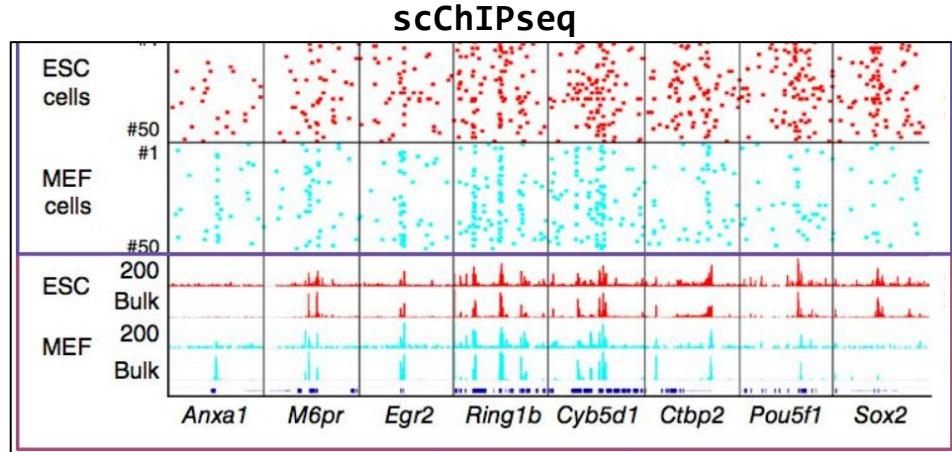
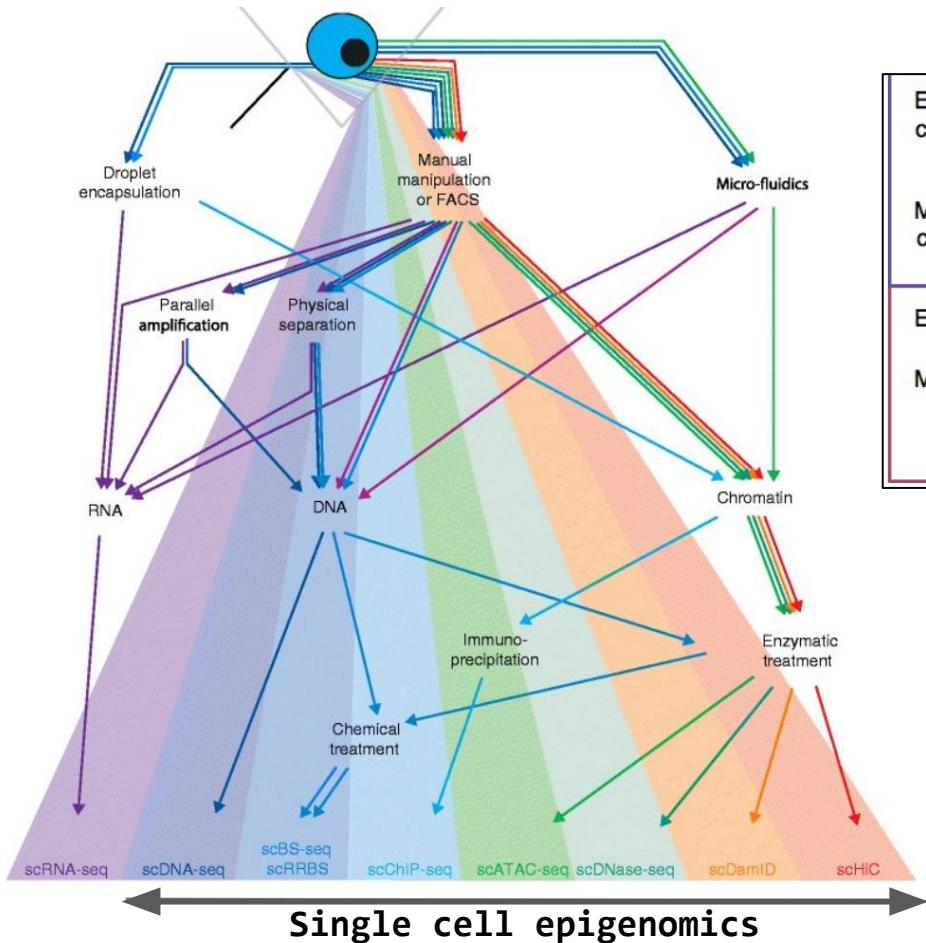


Chromosome organization

Higher-order chromatin organization into LADs and TADs



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>

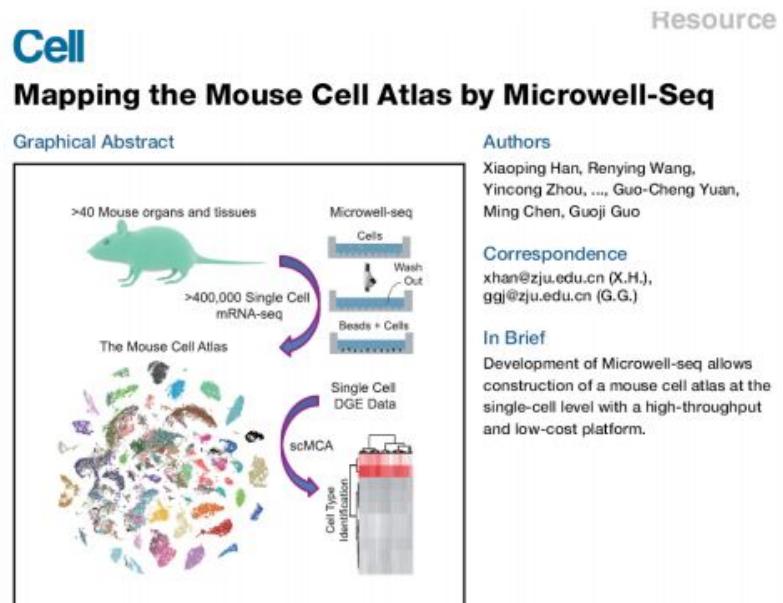
Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

The Tabula Muris Consortium*

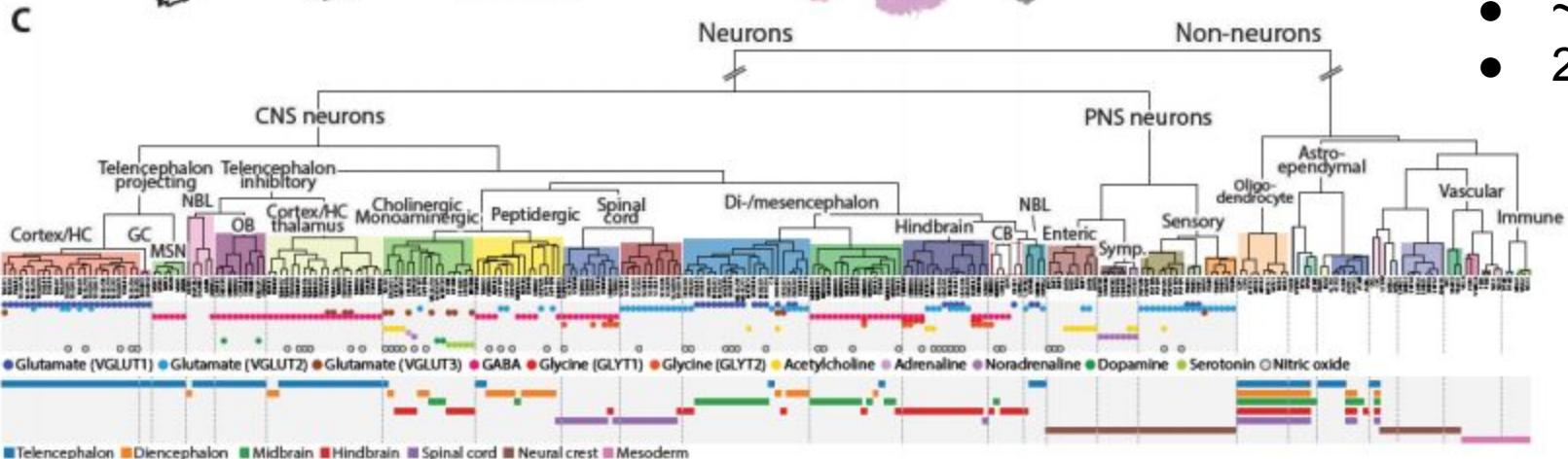
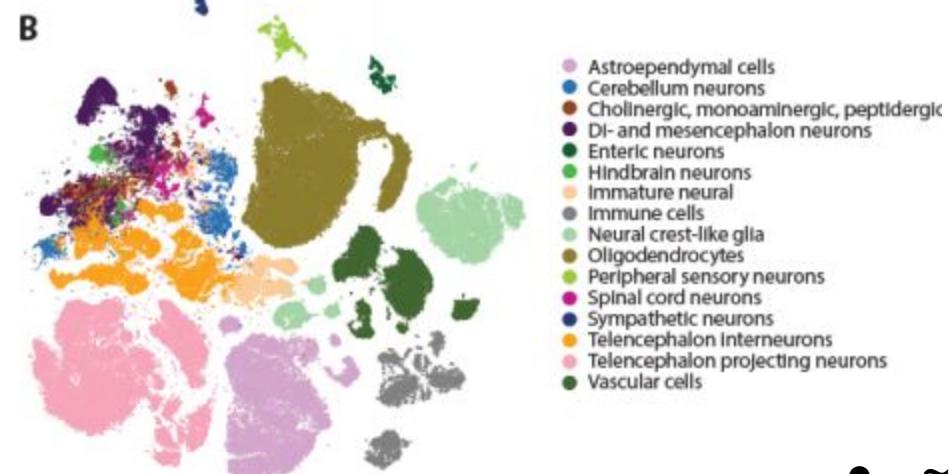
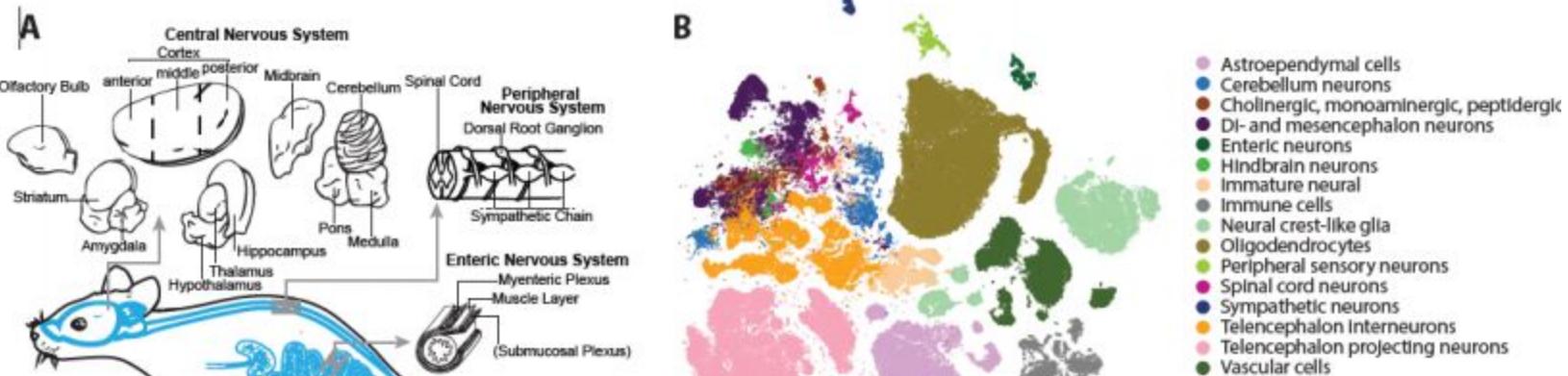
- ~100k cells
- 20 organs
- 2 techniques :
 - Droplet 3', short reads
 - FACS, long reads

MCA browser

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



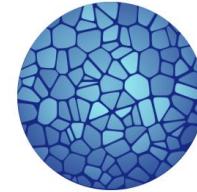
The Mouse Brain Atlas (mousebrain.org)



The Human Cell Atlas (humancellatlas.org)

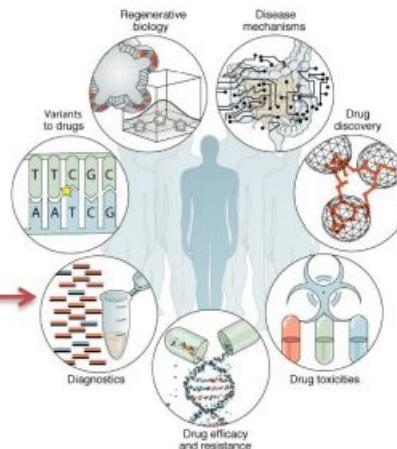
MAPPING THE BASIC UNITS OF LIFE

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

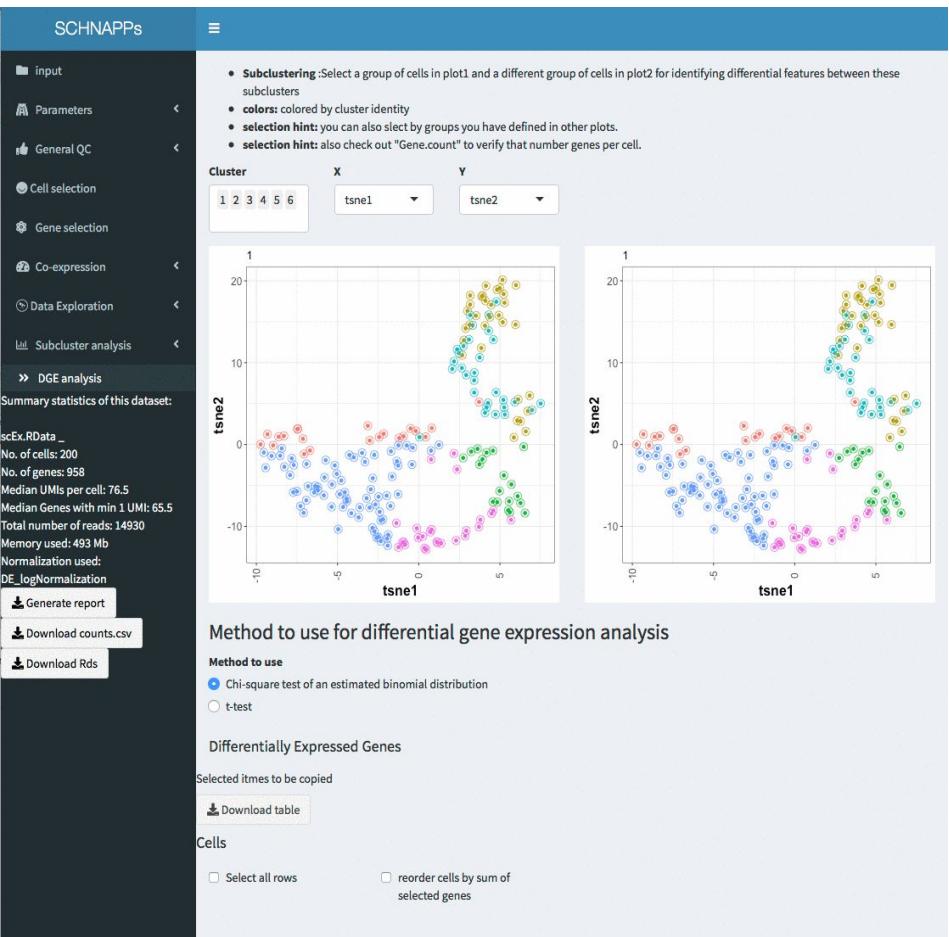
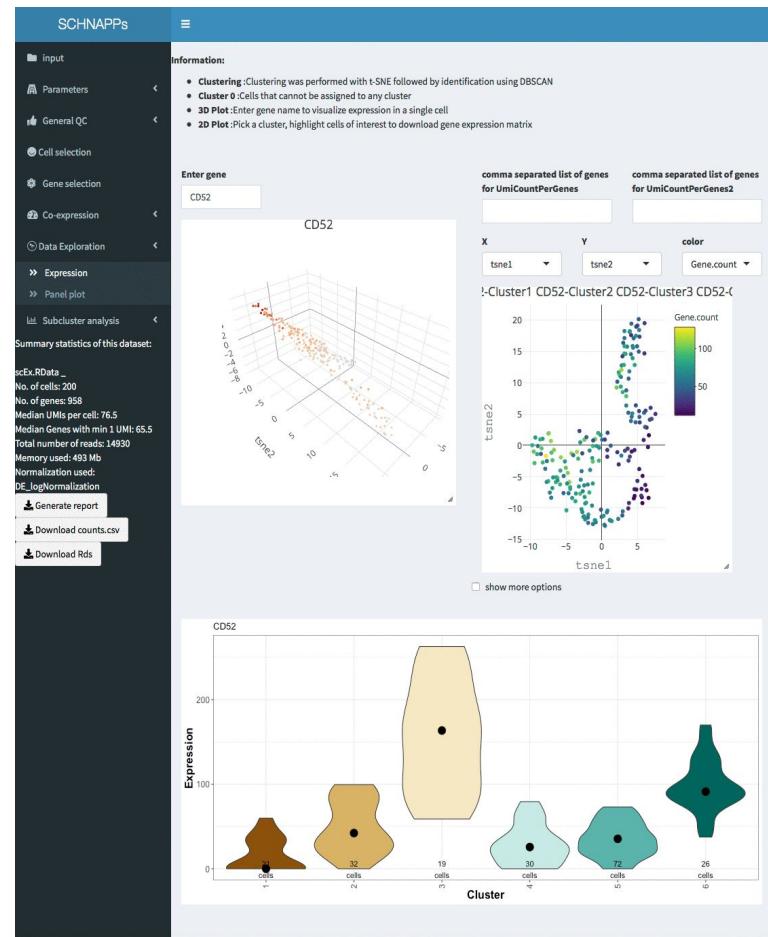
21-05-2020 Ongoing work to move to new hosting.

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



By Bernd Jagla (Pasteur Paris)

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

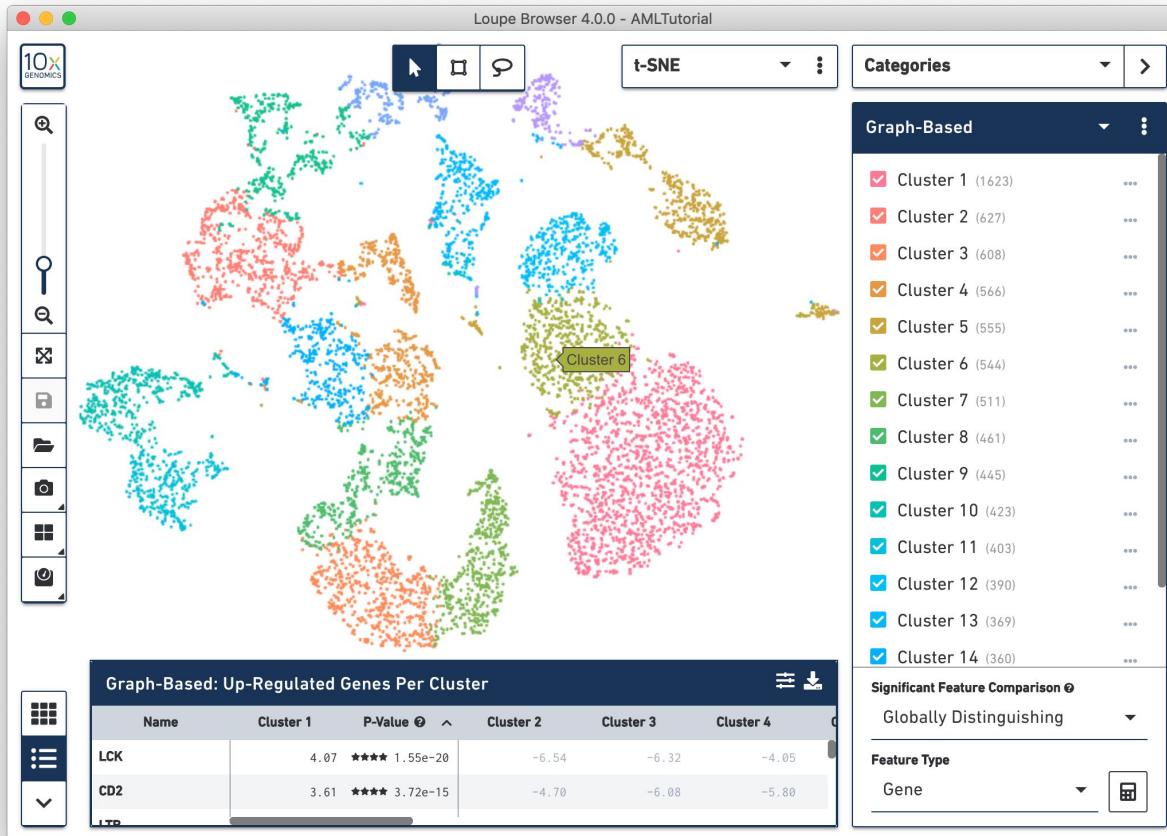
SeuratV3Wizard

The SeuratV3Wizard interface consists of three main panels:

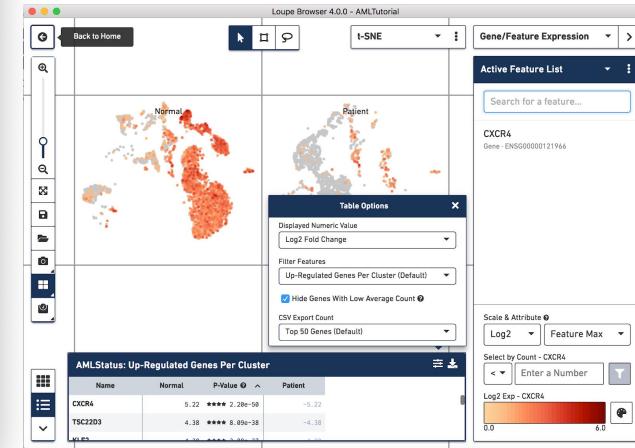
- Vin Plot (Filter Cells) Panel:** Shows two scatter plots: "nFeature_RNA" and "nCount_RNA". The "nFeature_RNA" plot has a red dashed horizontal line labeled "low.threshold" at approximately 100 and a blue dashed horizontal line labeled "high.threshold" at approximately 2000. The "nCount_RNA" plot shows a sharp peak at 1. A subset named "nFeature_RNA" is selected.
- PCHeatmap Output Panel:** A configuration panel for PCA heatmaps. It includes:
 - Input Data: QC & Filter, VinPlot (Filter Cells), Norm/Detect/Scale, PCA Reduction, Viz PCA Plot, PCA Plot, PC Heatmap, Elbow/JackStraw, Cluster Cells, Non-linear Reduction, Cluster Markers, Viz Markers, Download Seurat Obj.
 - PCs to use: PC 1 to 6.
 - Number of cells to use: 500.
 - Plot Download Options: Plot height (in cm): 30, Plot width (in cm): 30.
 - Download File Type: PDF (selected).
- Run Non-linear dimensional reduction Panel:** A configuration panel for non-linear dimensionality reduction:
 - Input Data: QC & Filter, VinPlot (Filter Cells), Norm/Detect/Scale, PCA Reduction, Viz PCA Plot, PCA Plot, PC Heatmap, Elbow/JackStraw, Cluster Cells, Non-linear Reduction, Cluster Markers, Viz Markers, Download Seurat Obj.
 - Choose reduction method to proceed with:
 - tsne (radio button selected)
 - umap
- UMAP Plot Panel:** A UMAP plot showing clusters of cells. The x-axis is UMAP_1 and the y-axis is UMAP_2. Cells are colored by cluster index from 0 to 9.
- Heatmap Panels:** Six PCA heatmap panels (PC_1 to PC_6) showing gene expression across samples. The samples are color-coded by cluster.

Visualization tools
(mainly for scRNAseq)

10x Genomics Loupe Browser

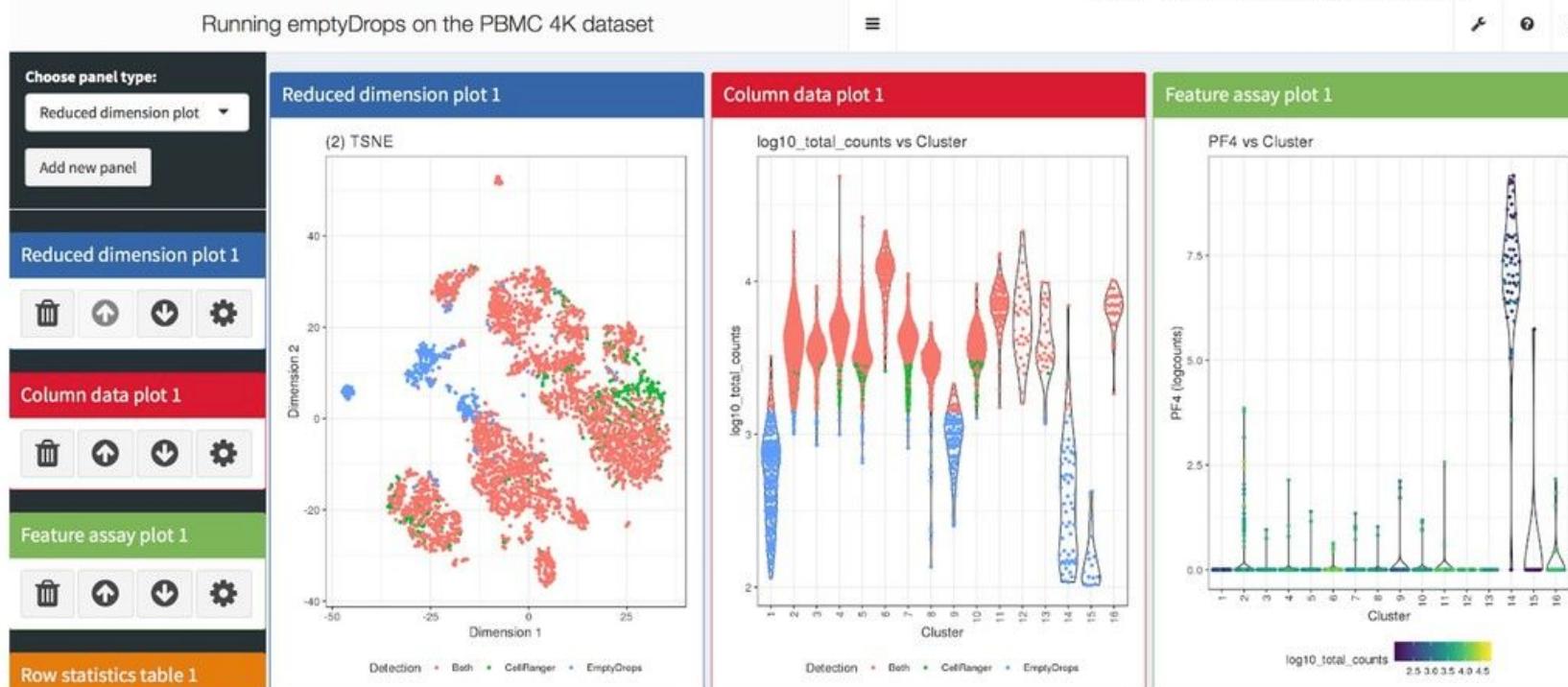


- Compatible with output from 10x Cell Ranger (“cloupe” files)
- Linux / OSX
- Supports Visium (Spatial)



Interactive Data Visualization (iSEE)

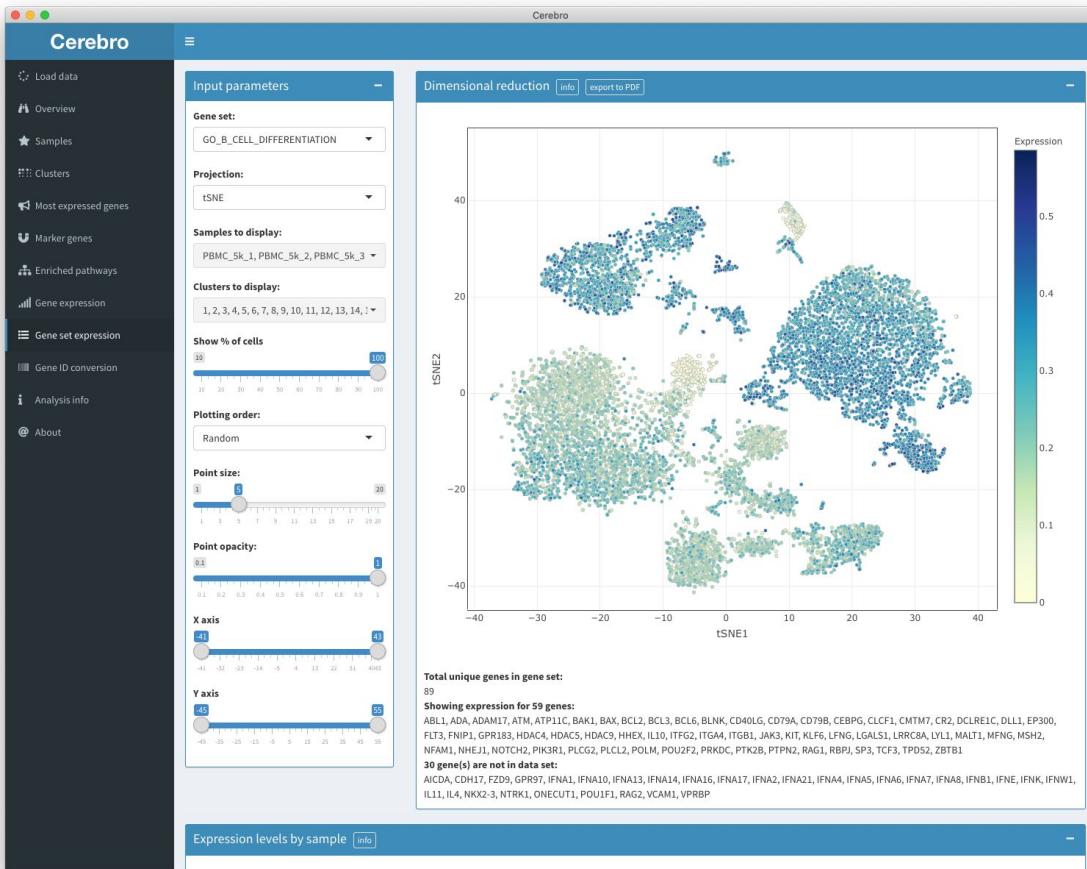
Running emptyDrops on the PBMC 4K dataset



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

https://marionilab.cruk.cam.ac.uk/iSEE_pbmc4k/

CerebroApp

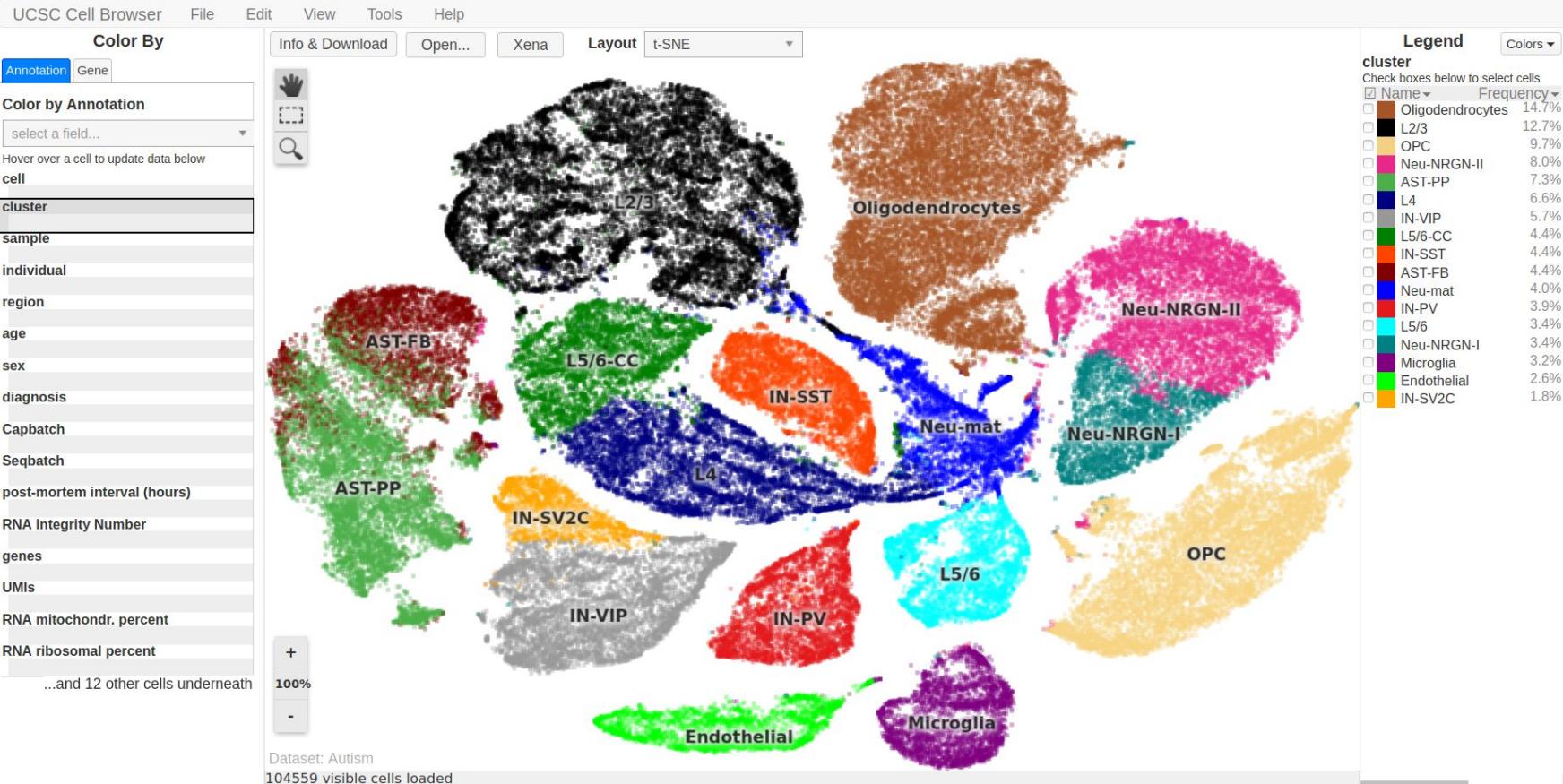


- ShinyApp (web GUI over some R)
- Binary format (CRB), converted from SeuratObject / SCE
- From QC to trajectory

(my favorite one)

<https://romanhaa.github.io/projects/cerebro/>

UCSC Cell Browser



<https://github.com/maximilanh/cellBrowser> (demo : <https://cells.ucsc.edu>)

Acknowledgements

Marc Deloger

Morgane Thomas-Chollier

Agnès Paquet

Marine Aglave

Antonio Rausell

Wouter Saelens

Nathalie Gaspar

... and you !



SINGle-cellING in the RAINaseq (1952)

© Jacques van Helden

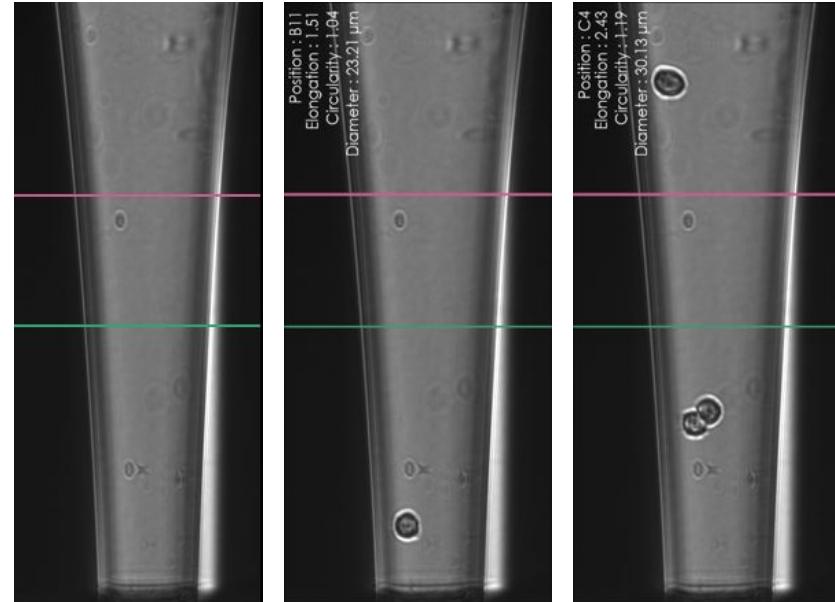
APPENDIX

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
- ...

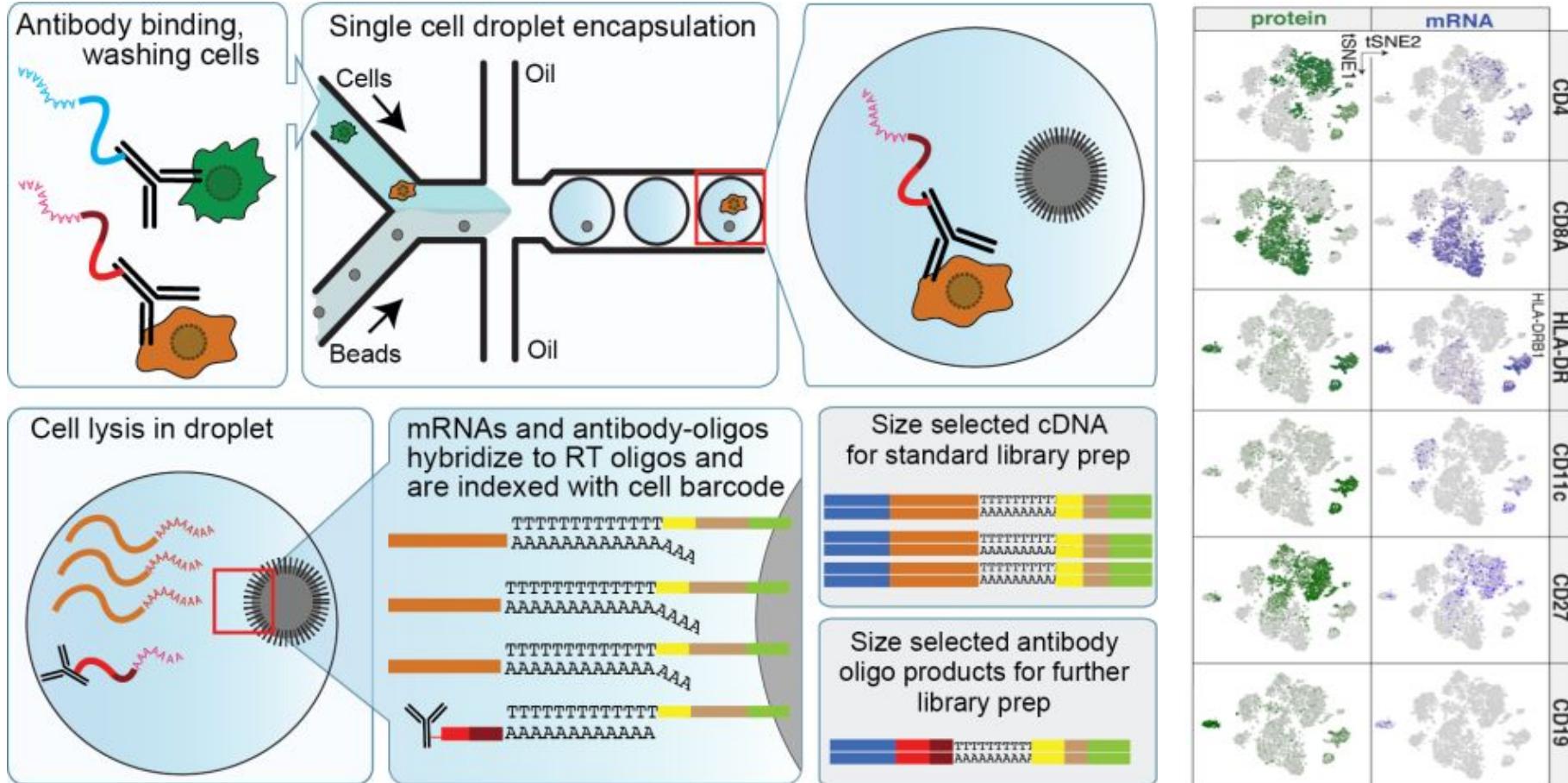
Alternative isolation method : Cellenion IBSCI™ (Image Based Single Cell Isolation)

- 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



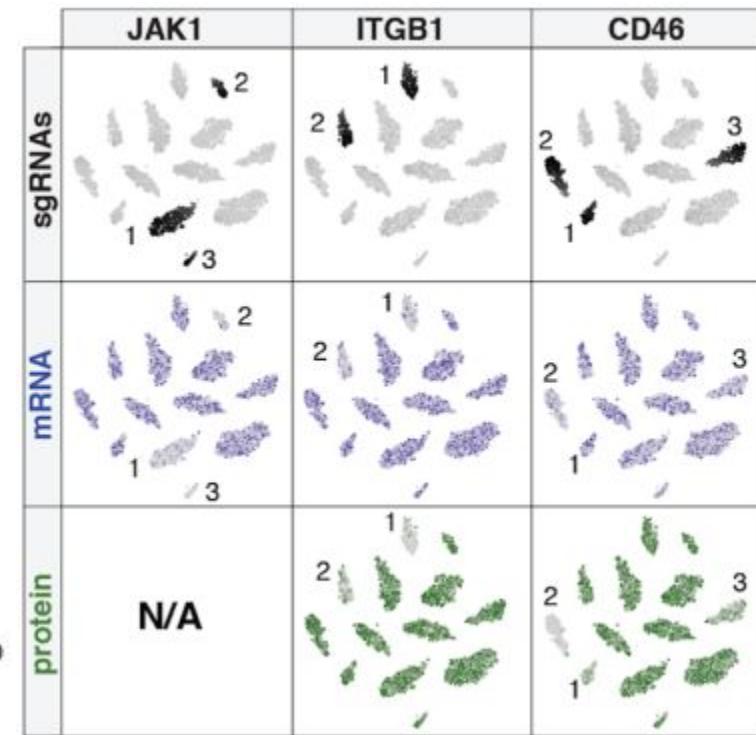
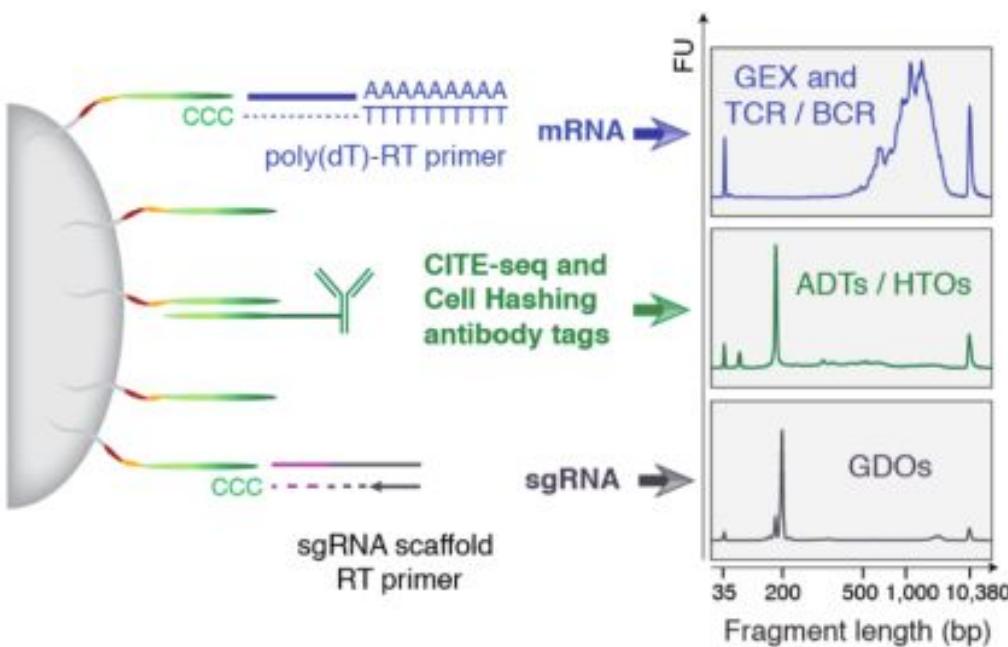
CITE-seq
(scRNAseq + proteins)

Cellular Indexing of Transcriptomes and Epitopes by Sequencing



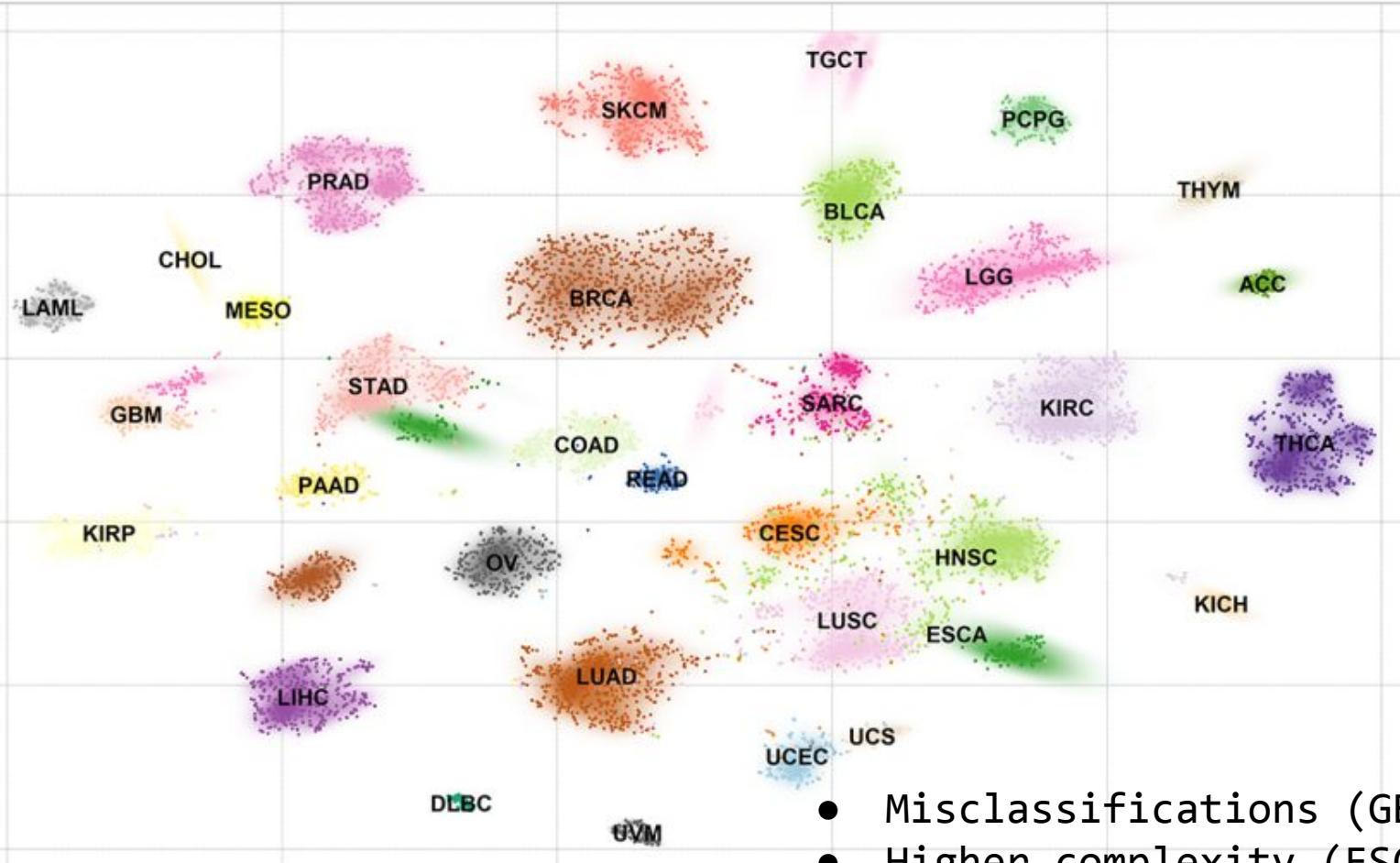
ECCITE-seq
(scRNAseq + proteins + CRISPR gRNA)

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



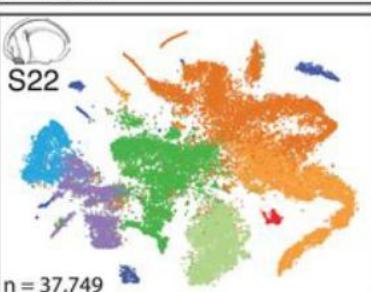
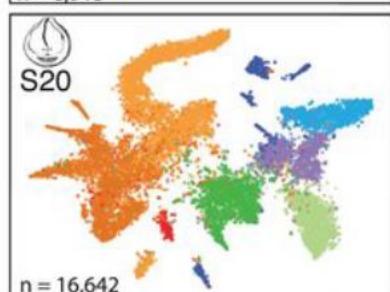
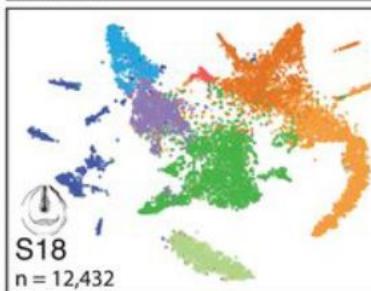
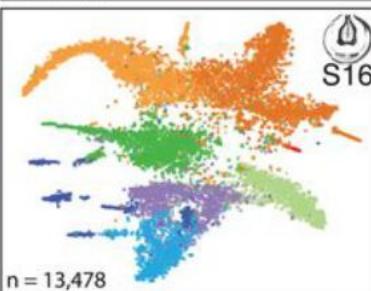
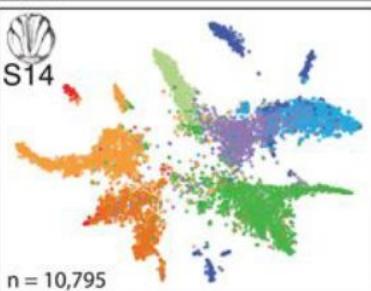
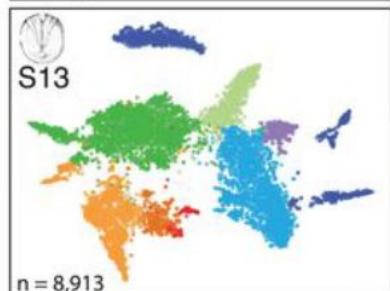
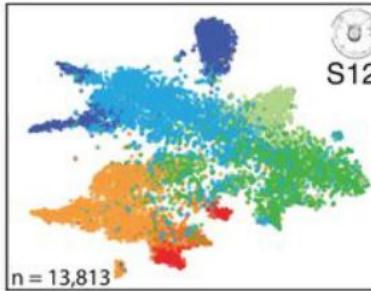
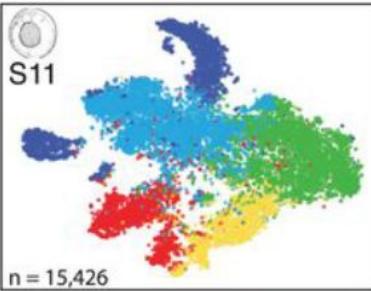
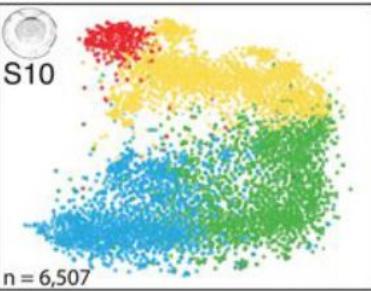
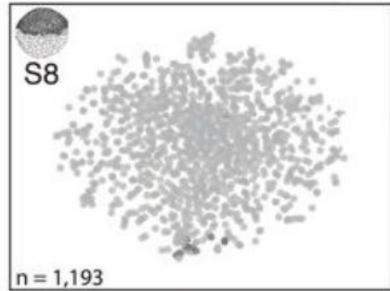
Some sweeties

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



Xenopus embryo development

tSNE dimension 2

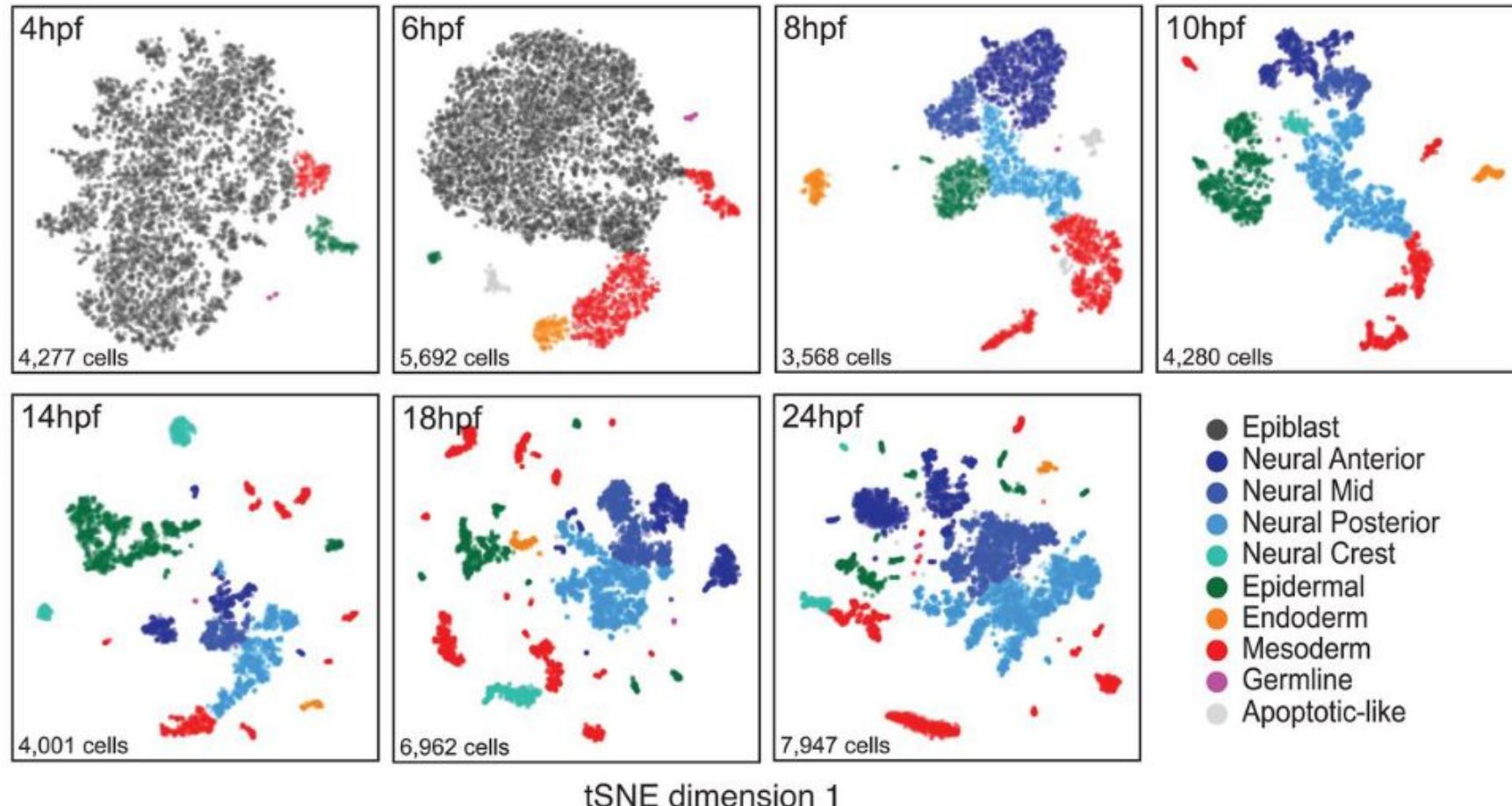


- Pluripotent blastula
- Germline
- Non-neural ectoderm
- Placodal
- Specialized epidermis
- Neural
- Neural crest
- Marginal zone
- Dorsal mesoderm
- Vent/lat/int. mesoderm
- Endoderm

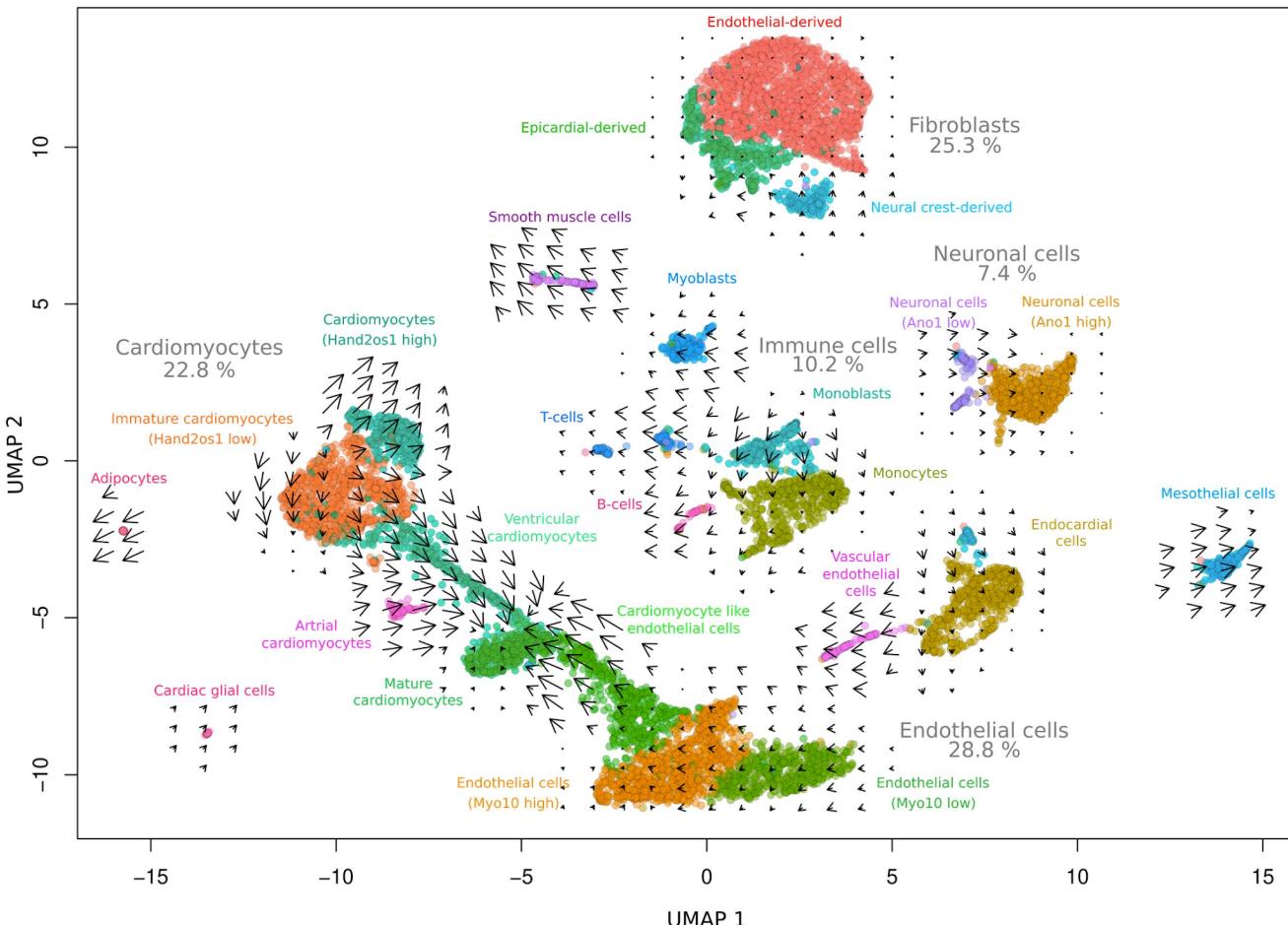
tSNE dimension 1

Zebrafish embryo development

tSNE dimension 2



Entire mouse heart : expression & velocity



<https://doi.org/10.3390/cells9020318>

A little single scRNAseq cheatsheet

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