

An Introduction to Single-Cell Genomics

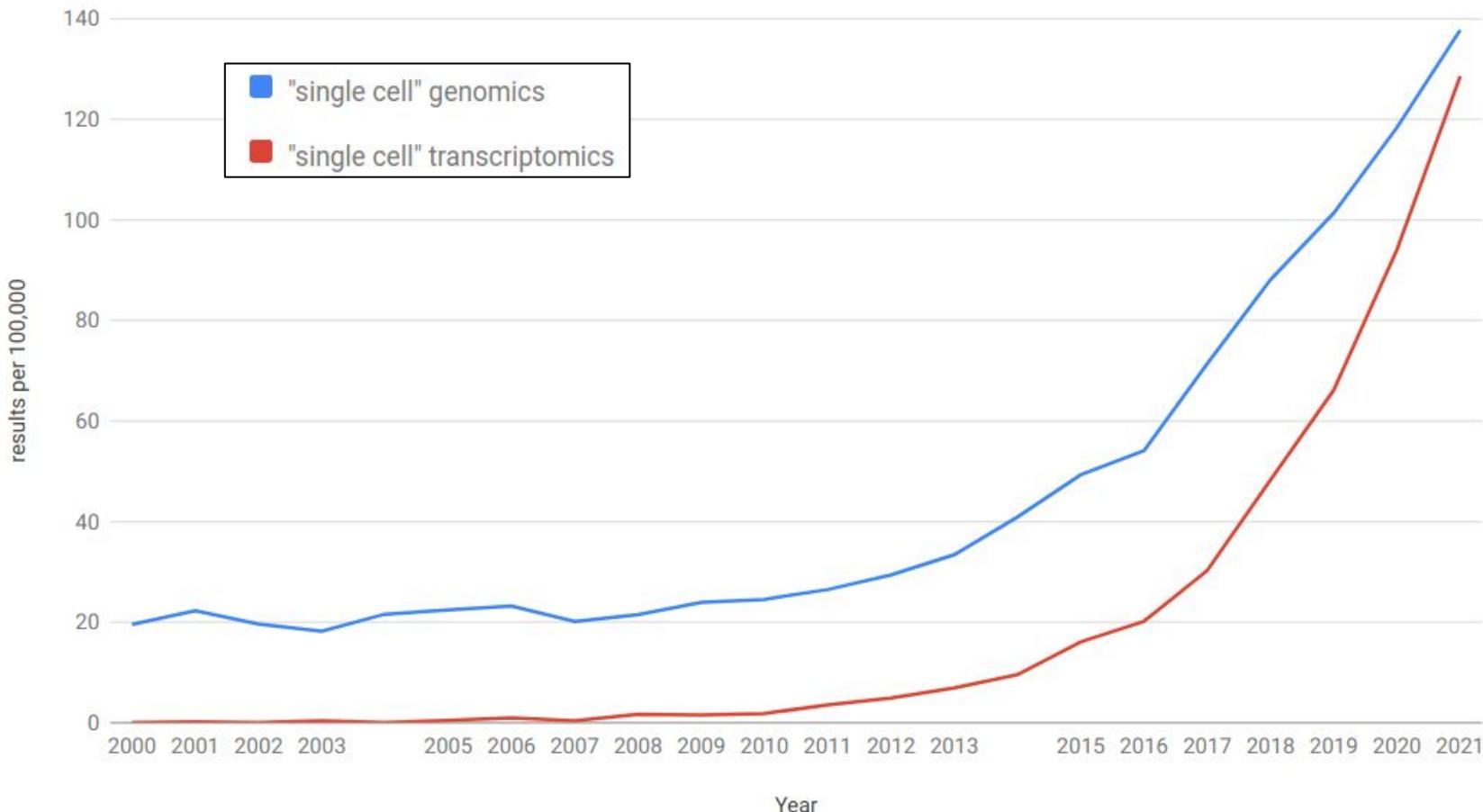
(Actually, mostly transcriptomics)
(Actually, mostly from 10X Genomics)

Bastien JOB
bastien.job@gustaveroussy.fr
INSERM / Gustave Roussy

*So you say you've heard
About
Single cell ?*

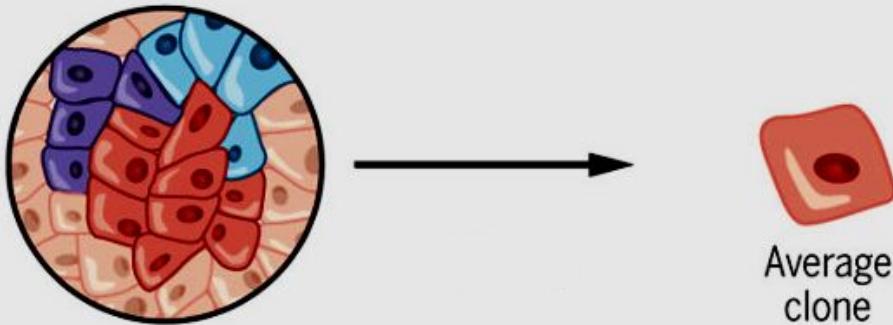


Single cell in peer-reviewed publications (2021)

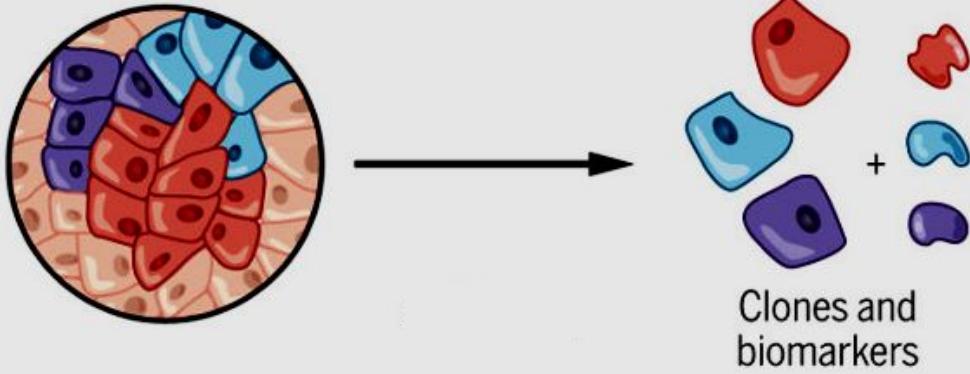


Why so much hype ?

A Bulk analysis



B scRNA analysis

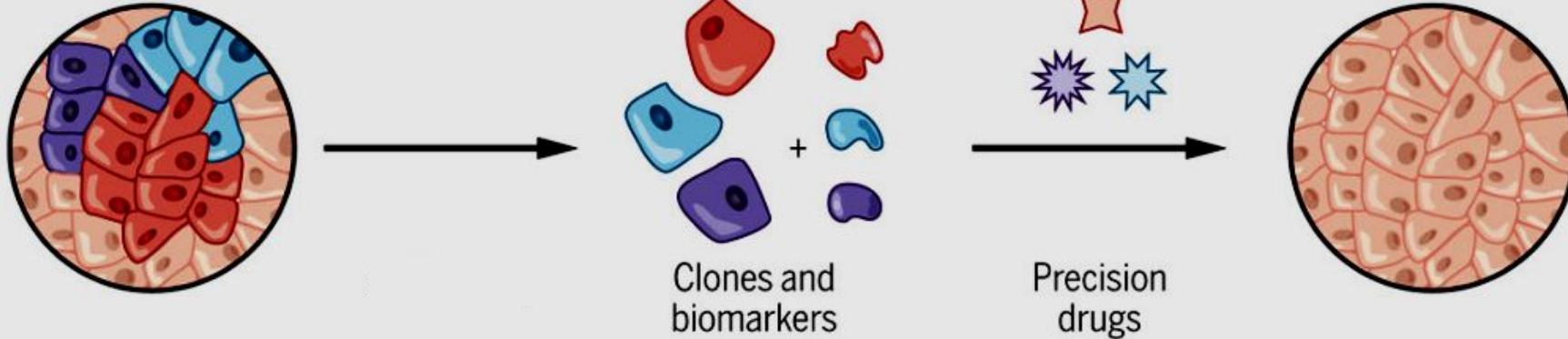


Why so much hype ? (pathology)

A Bulk analysis



B scRNA analysis



Why so much hype ?

Bulk



Single cell



Spatial single cell



Why so much hype ?

Bulk



Single cell



Spatial single cell



Full-length spatial single cell

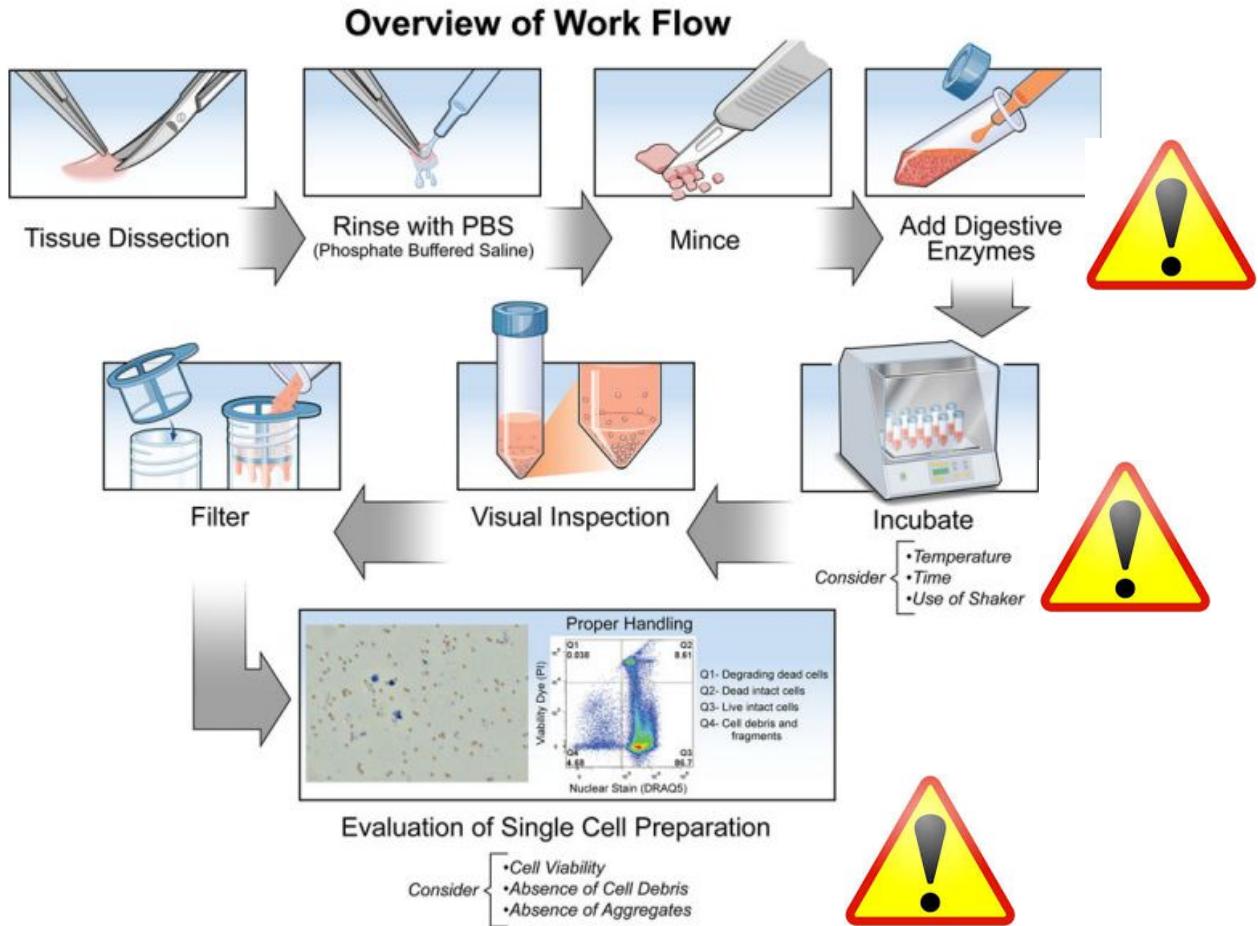




*From
Broad tissue
To
Isolated cells*

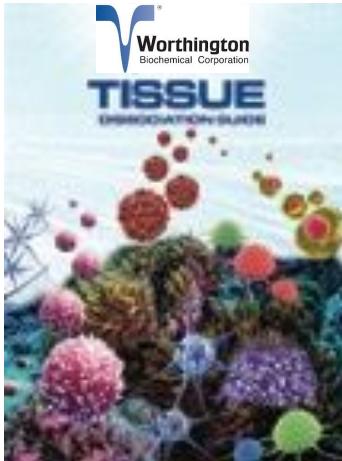


Cells health and dissociation



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

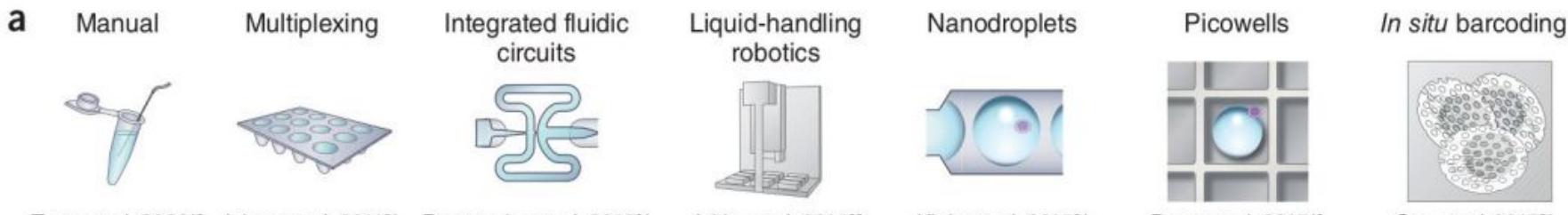
- Tissue Types
 - Epithelial Tissue
 - Connective Tissue
- Dissociating Enzymes
 - Collagenase
 - Trypsin
 - Elastase
 - Hyaluronidase
 - Papain
 - Chymotrypsin
 - Deoxyribonuclease I
 - Neutral Protease (Dispase)
 - Trypsin Inhibitor
 - Animal Origin Free (AOF) Enzymes
 - Celase® GMP

III. Cell Isolation Techniques

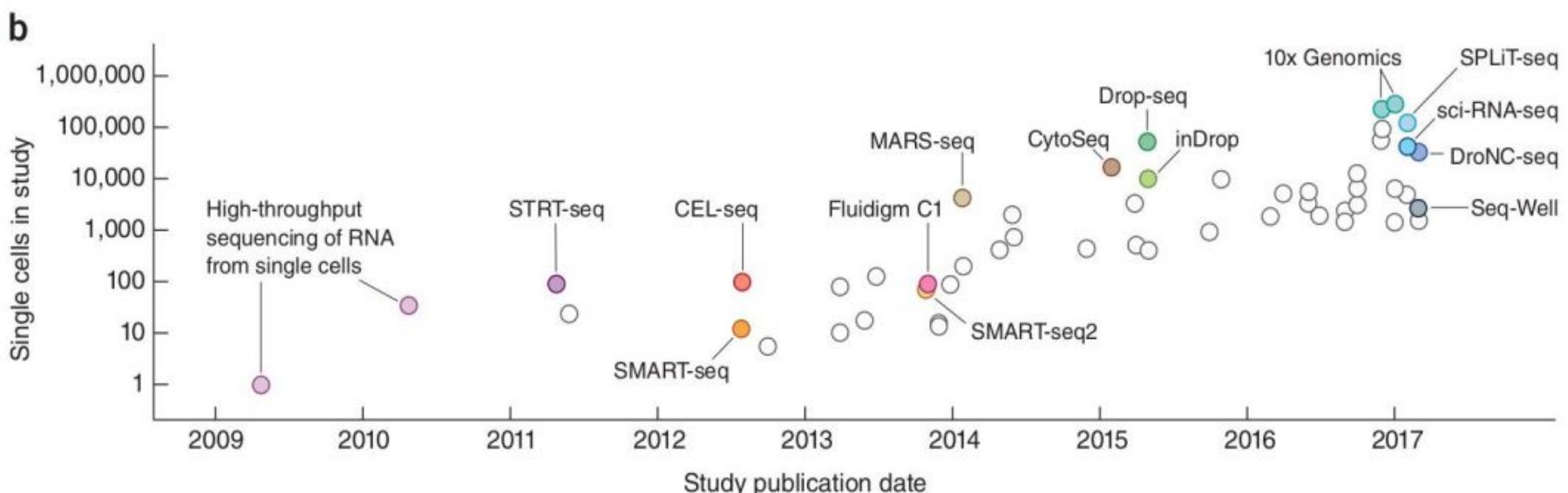
- Methods & Materials
 - Working With Enzymes
 - Basic Primary Cell Isolation
 - Equilibration with 95%O₂:5%CO₂
 - Trituration
 - Enzymatic Cell Harvesting
 - Cell Adhesion and Harvesting
 - Trypsin for Cell Harvesting
 - Cell Release Procedure
- Optimization Techniques
 - General Guidelines
 - Optimization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

IV. Use-Tested Cell Isolation Systems

Cells isolation : technologies over the last decade



Tang *et al.* 2009¹⁸ Islam *et al.* 2011²⁴ Brennecke *et al.* 2013⁶⁴ Jaitin *et al.* 2014³³ Klein *et al.* 2015³⁴ Macosko *et al.* 2015⁴⁰ Bose *et al.* 2015⁴³ Cao *et al.* 2017⁵¹ Rosenberg *et al.* 2017⁵²

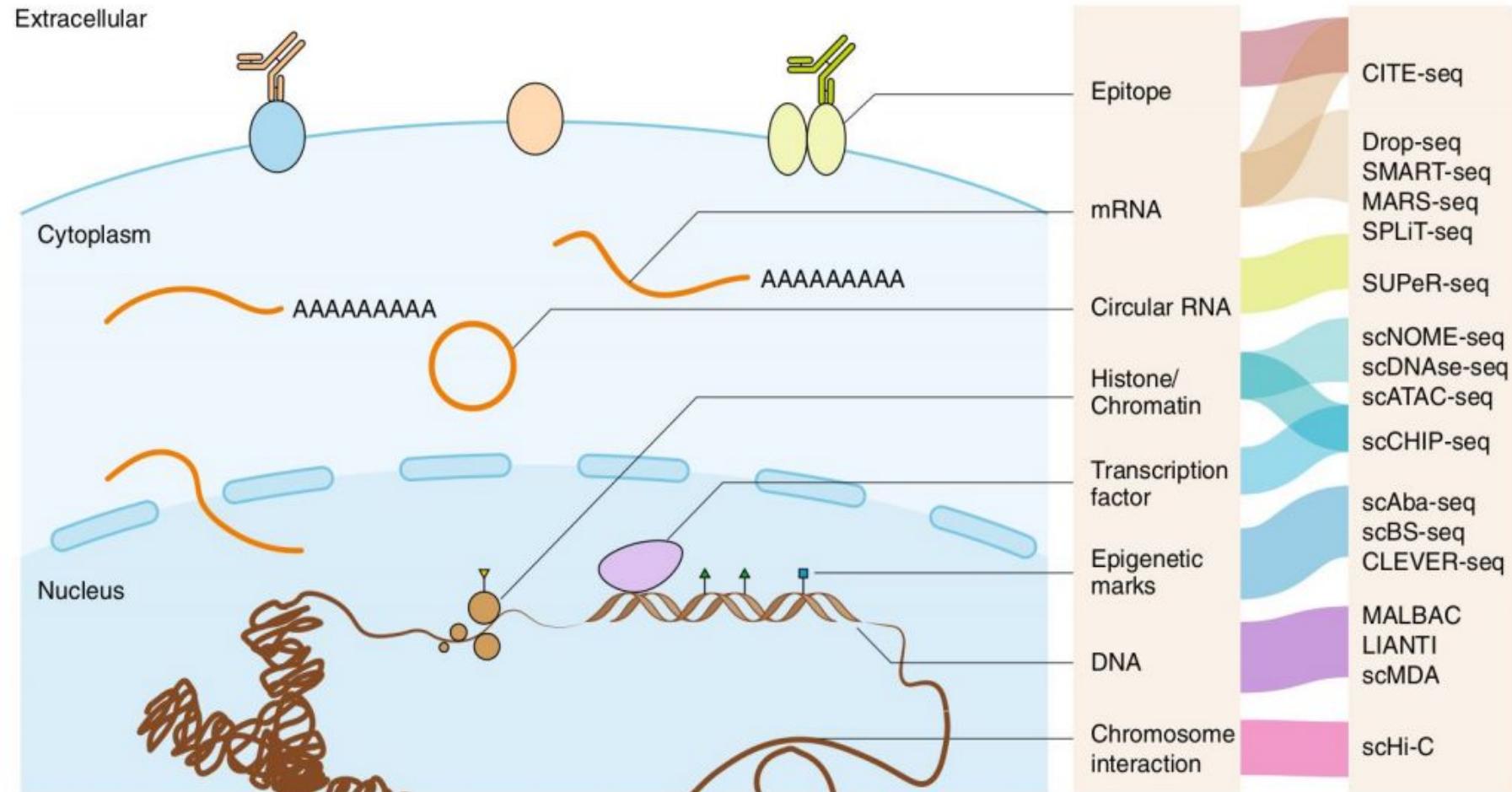




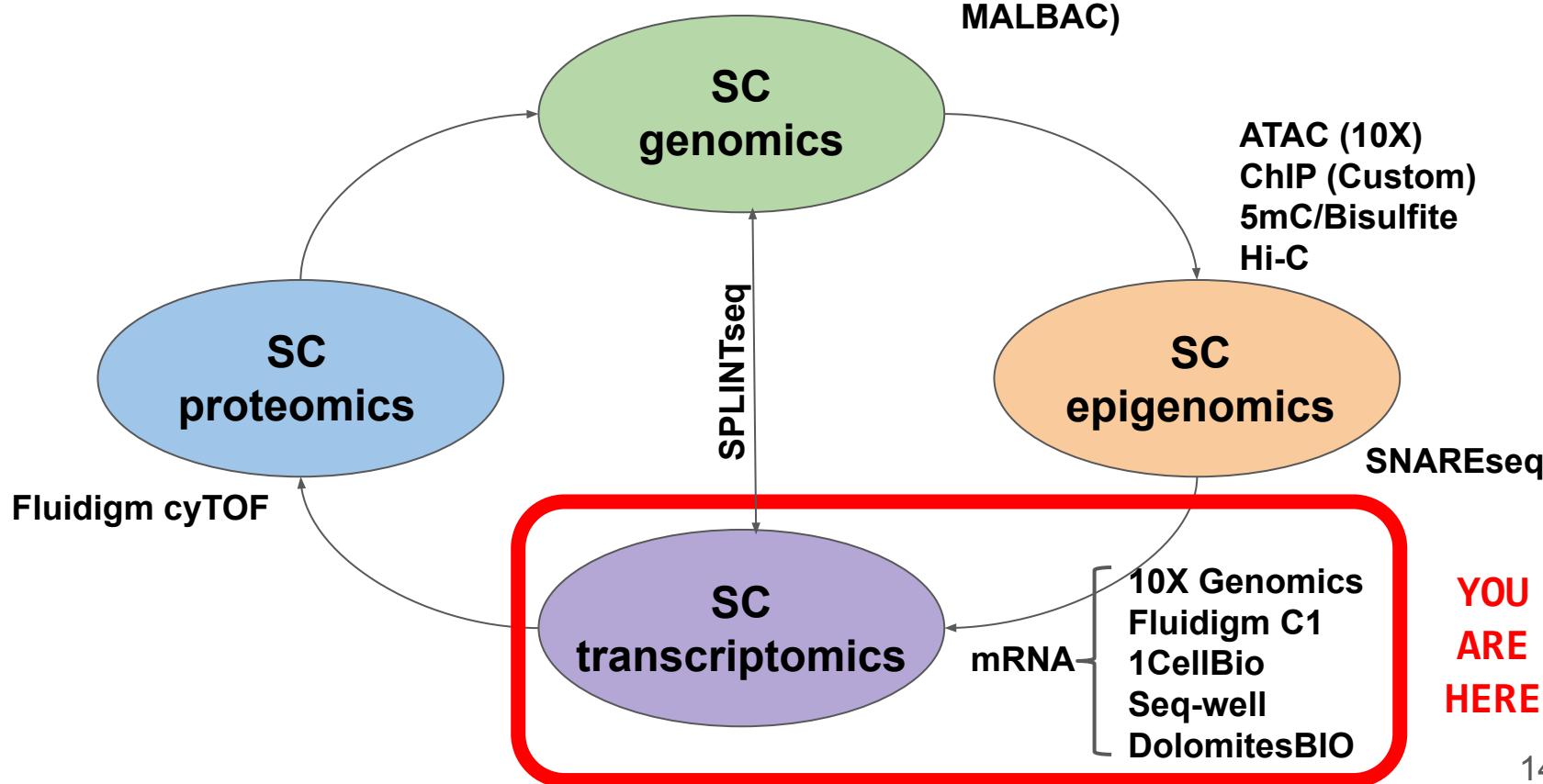
*From
Isolated cells
To
Nucleotide sequences (reads)*



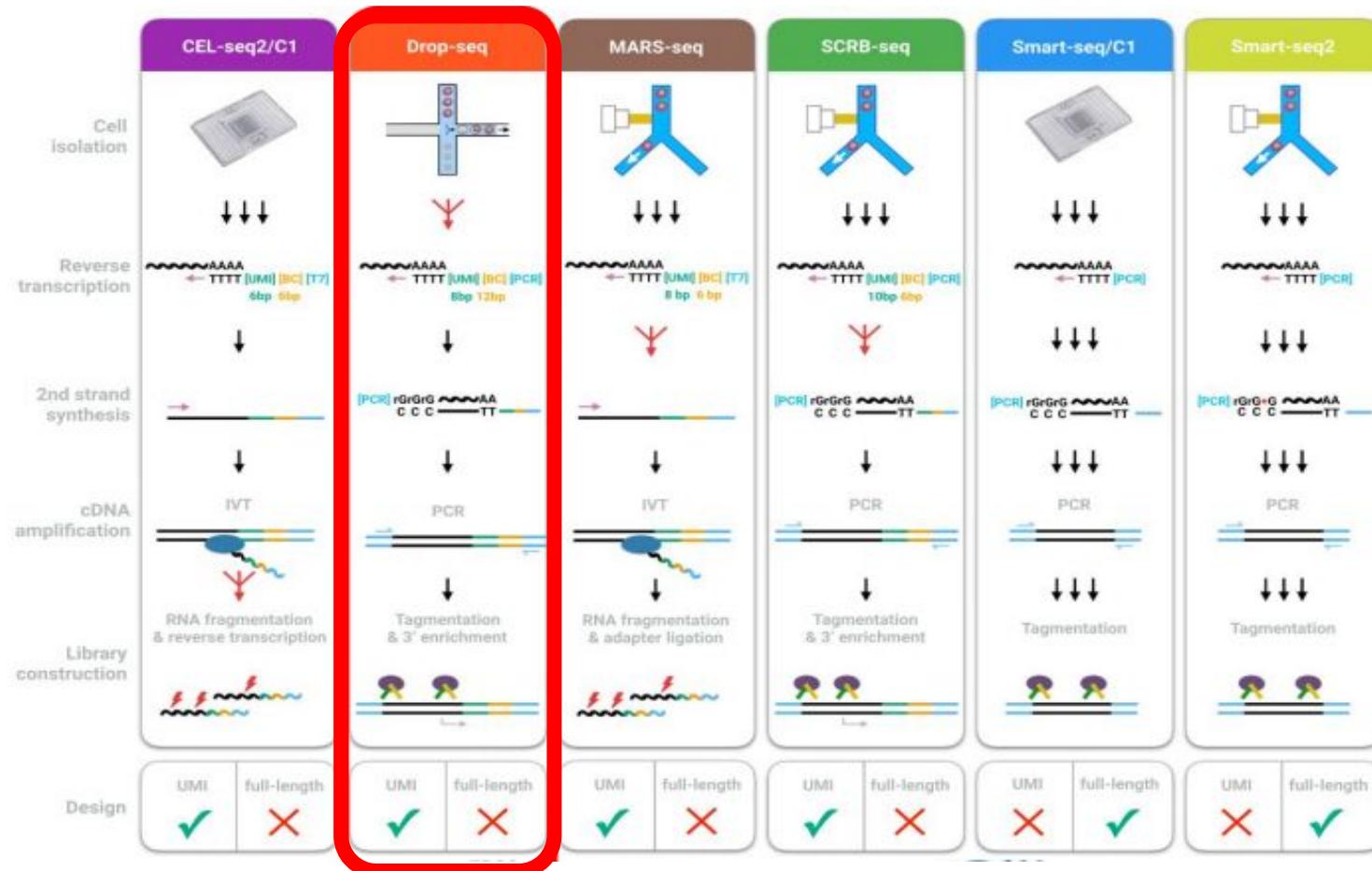
Several protocols for several purposes



Single Cell RNaseq



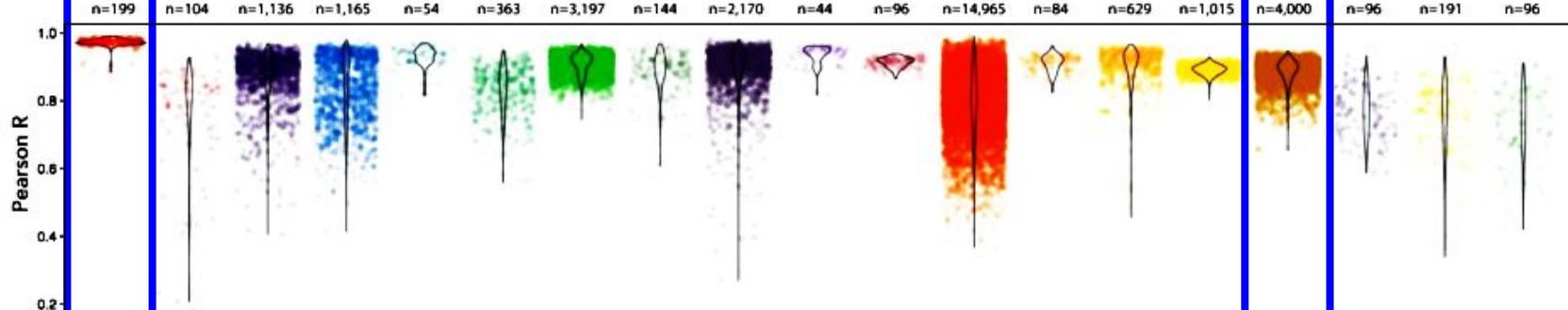
From isolated cells to sequences (Drop-seq / 10X)



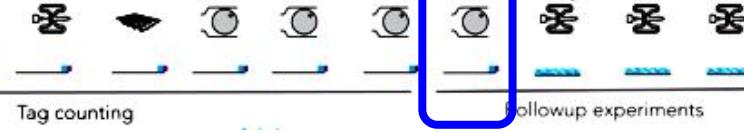
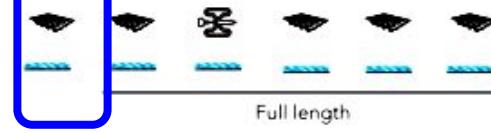
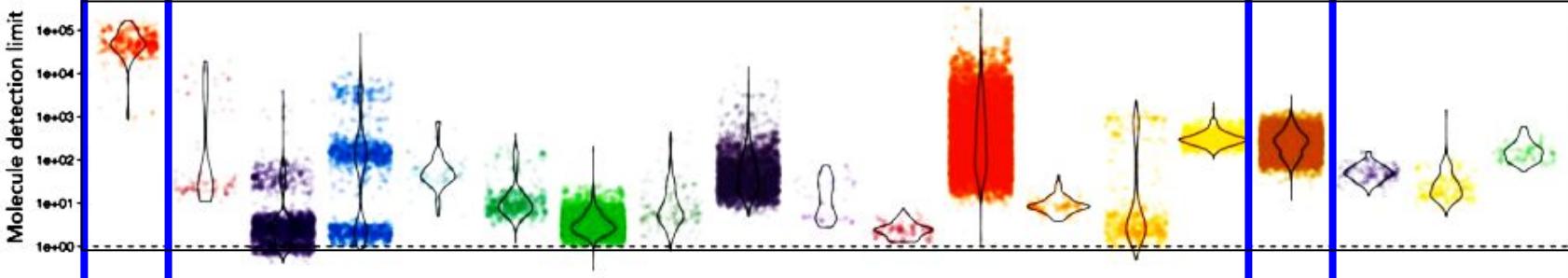
Bulk

10X

Accuracy

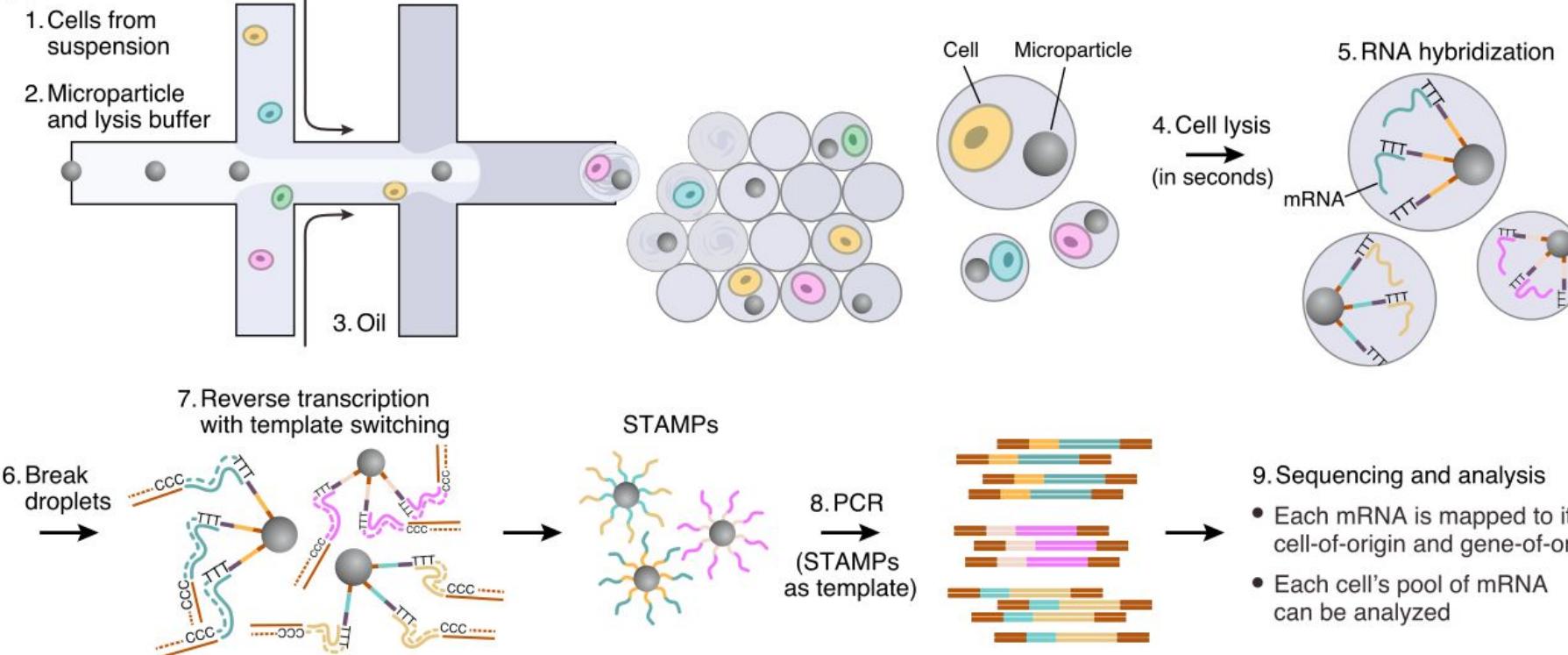


Sensitivity



Drop-seq

A

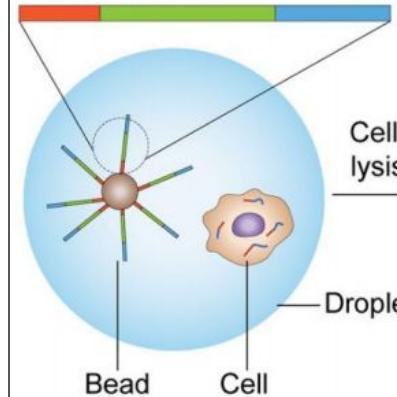


STAMPs = single-cell transcriptomes attached to microparticles

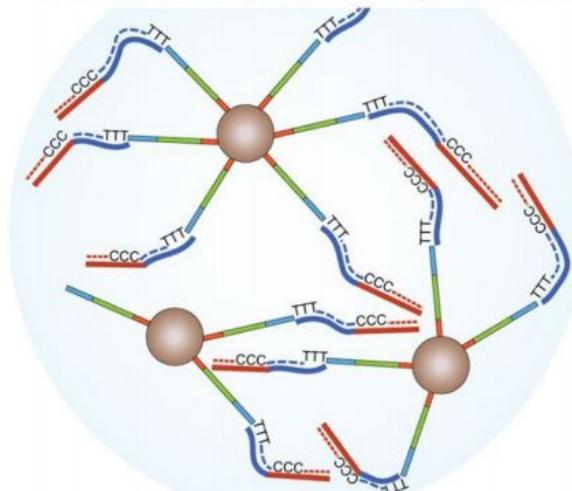
10X Chromium (3')

Structure of the barcode primer bead

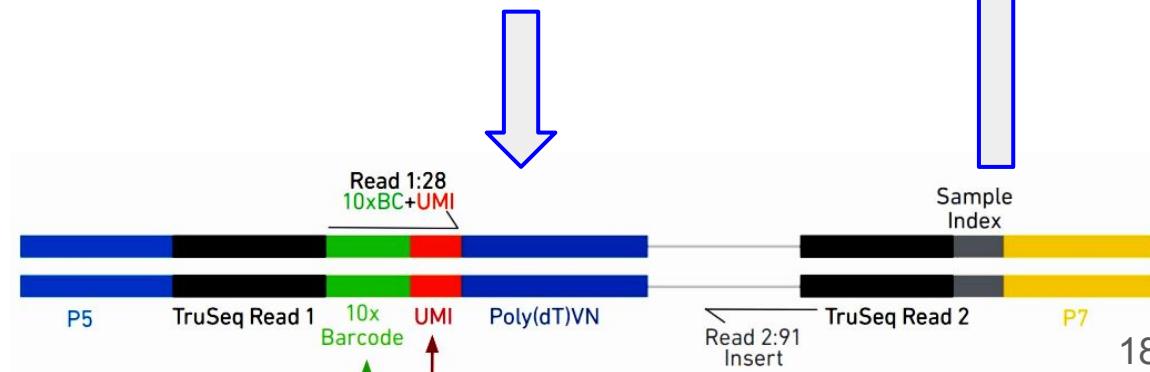
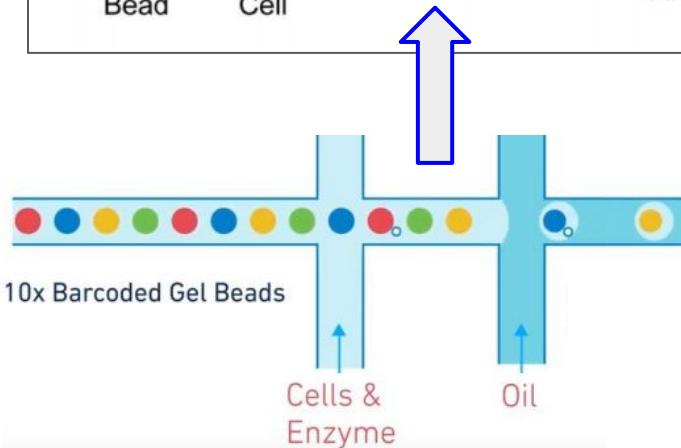
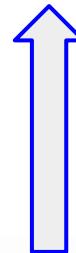
PCR
handle Cell barcode UMI



Reverse transcription with template switching



Sequencing





*From
Nucleotide sequences
To a
Raw count matrix*

Sample	Base	Count
S1	A	100
S1	T	100
S1	G	100
S1	C	100
S2	A	100
S2	T	100
S2	G	100
S2	C	100
S3	A	100
S3	T	100
S3	G	100
S3	C	100
S4	A	100
S4	T	100
S4	G	100
S4	C	100
S5	A	100
S5	T	100
S5	G	100
S5	C	100
S6	A	100
S6	T	100
S6	G	100
S6	C	100
S7	A	100
S7	T	100
S7	G	100
S7	C	100
S8	A	100
S8	T	100
S8	G	100
S8	C	100
S9	A	100
S9	T	100
S9	G	100
S9	C	100
S10	A	100
S10	T	100
S10	G	100
S10	C	100
S11	A	100
S11	T	100
S11	G	100
S11	C	100
S12	A	100
S12	T	100
S12	G	100
S12	C	100
S13	A	100
S13	T	100
S13	G	100
S13	C	100
S14	A	100
S14	T	100
S14	G	100
S14	C	100
S15	A	100
S15	T	100
S15	G	100
S15	C	100
S16	A	100
S16	T	100
S16	G	100
S16	C	100
S17	A	100
S17	T	100
S17	G	100
S17	C	100
S18	A	100
S18	T	100
S18	G	100
S18	C	100
S19	A	100
S19	T	100
S19	G	100
S19	C	100
S20	A	100
S20	T	100
S20	G	100
S20	C	100
S21	A	100
S21	T	100
S21	G	100
S21	C	100
S22	A	100
S22	T	100
S22	G	100
S22	C	100
S23	A	100
S23	T	100
S23	G	100
S23	C	100
S24	A	100
S24	T	100
S24	G	100
S24	C	100
S25	A	100
S25	T	100
S25	G	100
S25	C	100
S26	A	100
S26	T	100
S26	G	100
S26	C	100
S27	A	100
S27	T	100
S27	G	100
S27	C	100
S28	A	100
S28	T	100
S28	G	100
S28	C	100
S29	A	100
S29	T	100
S29	G	100
S29	C	100
S30	A	100
S30	T	100
S30	G	100
S30	C	100
S31	A	100
S31	T	100
S31	G	100
S31	C	100
S32	A	100
S32	T	100
S32	G	100
S32	C	100
S33	A	100
S33	T	100
S33	G	100
S33	C	100
S34	A	100
S34	T	100
S34	G	100
S34	C	100
S35	A	100
S35	T	100
S35	G	100
S35	C	100
S36	A	100
S36	T	100
S36	G	100
S36	C	100
S37	A	100
S37	T	100
S37	G	100
S37	C	100
S38	A	100
S38	T	100
S38	G	100
S38	C	100
S39	A	100
S39	T	100
S39	G	100
S39	C	100
S40	A	100
S40	T	100
S40	G	100
S40	C	100
S41	A	100
S41	T	100
S41	G	100
S41	C	100
S42	A	100
S42	T	100
S42	G	100
S42	C	100
S43	A	100
S43	T	100
S43	G	100
S43	C	100
S44	A	100
S44	T	100
S44	G	100
S44	C	100
S45	A	100
S45	T	100
S45	G	100
S45	C	100
S46	A	100
S46	T	100
S46	G	100
S46	C	100
S47	A	100
S47	T	100
S47	G	100
S47	C	100
S48	A	100
S48	T	100
S48	G	100
S48	C	100
S49	A	100
S49	T	100
S49	G	100
S49	C	100
S50	A	100
S50	T	100
S50	G	100
S50	C	100
S51	A	100
S51	T	100
S51	G	100
S51	C	100
S52	A	100
S52	T	100
S52	G	100
S52	C	100
S53	A	100
S53	T	100
S53	G	100
S53	C	100
S54	A	100
S54	T	100
S54	G	100
S54	C	100
S55	A	100
S55	T	100
S55	G	100
S55	C	100
S56	A	100
S56	T	100
S56	G	100
S56	C	100
S57	A	100
S57	T	100
S57	G	100
S57	C	100
S58	A	100
S58	T	100
S58	G	100
S58	C	100
S59	A	100
S59	T	100
S59	G	100
S59	C	100
S60	A	100
S60	T	100
S60	G	100
S60	C	100
S61	A	100
S61	T	100
S61	G	100
S61	C	100
S62	A	100
S62	T	100
S62	G	100
S62	C	100
S63	A	100
S63	T	100
S63	G	100
S63	C	100
S64	A	100
S64	T	100
S64	G	100
S64	C	100
S65	A	100
S65	T	100
S65	G	100
S65	C	100
S66	A	100
S66	T	100
S66	G	100
S66	C	100
S67	A	100
S67	T	100
S67	G	100
S67	C	100
S68	A	100
S68	T	100
S68	G	100
S68	C	100
S69	A	100
S69	T	100
S69	G	100
S69	C	100
S70	A	100
S70	T	100
S70	G	100
S70	C	100
S71	A	100
S71	T	100
S71	G	100
S71	C	100
S72	A	100
S72	T	100
S72	G	100
S72	C	100
S73	A	100
S73	T	100
S73	G	100
S73	C	100
S74	A	100
S74	T	100
S74	G	100
S74	C	100
S75	A	100
S75	T	100
S75	G	100
S75	C	100
S76	A	100
S76	T	100
S76	G	100
S76	C	100
S77	A	100
S77	T	100
S77	G	100
S77	C	100
S78	A	100
S78	T	100
S78	G	100
S78	C	100
S79	A	100
S79	T	100
S79	G	100
S79	C	100
S80	A	100
S80	T	100
S80	G	100
S80	C	100
S81	A	100
S81	T	100
S81	G	100
S81	C	100
S82	A	100
S82	T	100
S82	G	100
S82	C	100
S83	A	100
S83	T	100
S83	G	100
S83	C	100
S84	A	100
S84	T	100
S84	G	100
S84	C	100
S85	A	100
S85	T	100
S85	G	100
S85	C	100
S86	A	100
S86	T	100
S86	G	100
S86	C	100
S87	A	100
S87	T	100
S87	G	100
S87	C	100
S88	A	100
S88	T	100
S88	G	100
S88	C	100
S89	A	100
S89	T	100
S89	G	100
S89	C	100
S90	A	100
S90	T	100
S90	G	100
S90	C	100
S91	A	100
S91	T	100
S91	G	100
S91	C	100
S92	A	100
S92	T	100
S92	G	100
S92	C	100
S93	A	100
S93	T	100
S93	G	100
S93	C	100
S94	A	100
S94	T	100
S94	G	100
S94	C	100
S95	A	100
S95	T	100
S95	G	100
S95	C	100
S96	A	100
S96	T	100
S96	G	100
S96	C	100
S97	A	100
S97	T	100
S97	G	100
S97	C	100
S98	A	100
S98	T	100
S98	G	100
S98	C	100
S99	A	100
S99	T	100
S99	G	100
S99	C	100
S100	A	100
S100	T	100
S100	G	100
S100	C	100

CERTIFIED



BIOINFORMAGICIAN

Reads QC

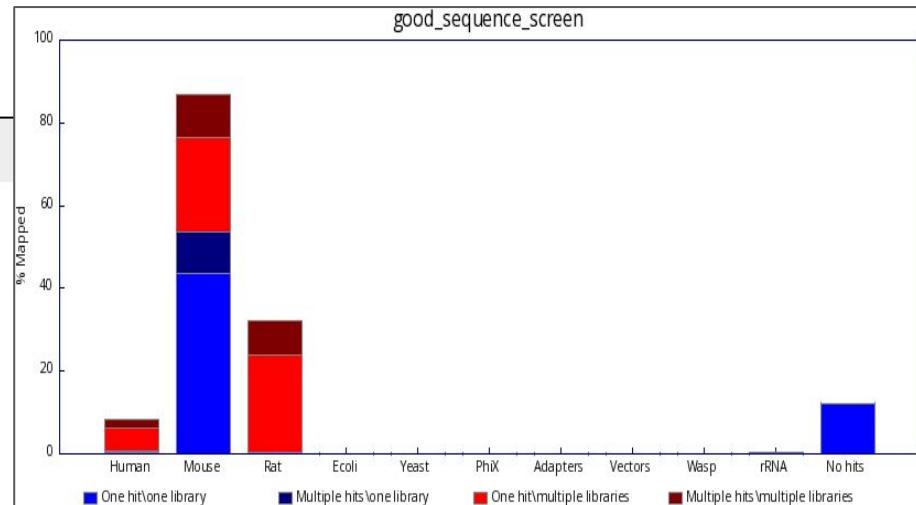
FastQC Report

Summary

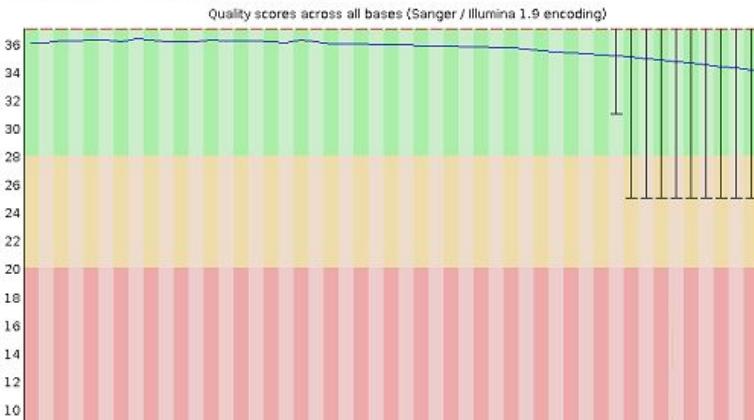
- ✓ Basic Statistics
- ✓ Per base sequence quality
- ✗ Per tile sequence quality
- ✓ Per sequence quality scores
- ✗ Per base sequence content
- ✓ Per sequence GC content
- ✓ Per base N content
- ✓ Sequence Length Distribution
- ✗ Sequence Duplication Levels
- ✗ Overrepresented sequences
- ✓ Adapter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



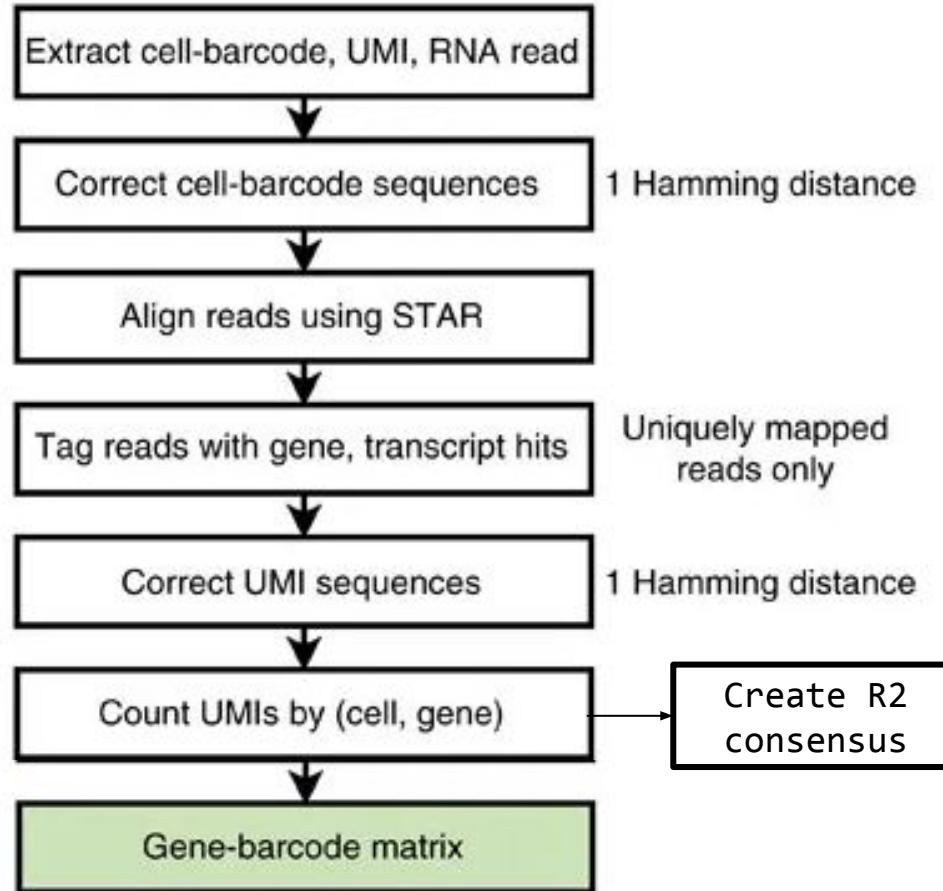
Per base sequence quality



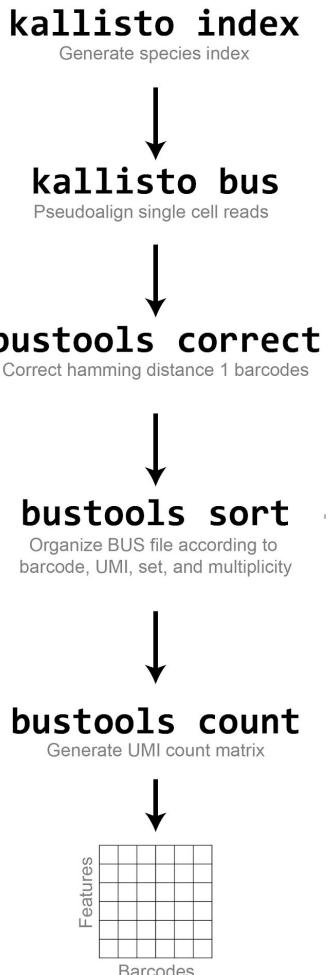
- As usual : FASTQC, FastqScreen, ...
- 10x :
 - Special attention to R1 : cell barcode + UMI (no N)
 - Control of the 4 sample libraries

Reads processing workflows

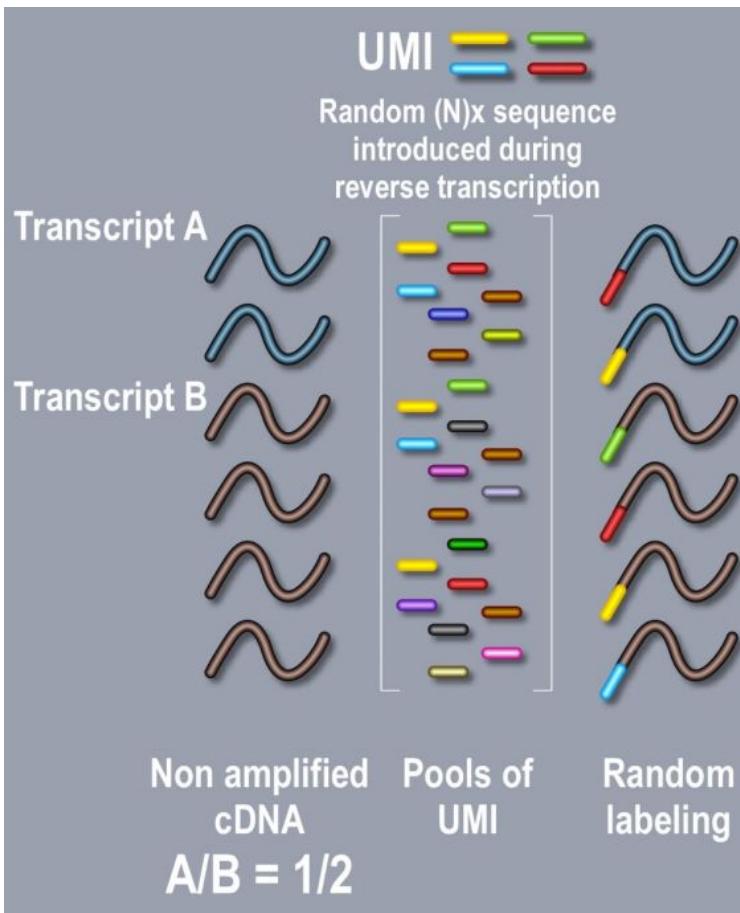
Mapping-based (STAR)



Pseudomapping-based (kallisto bustools)

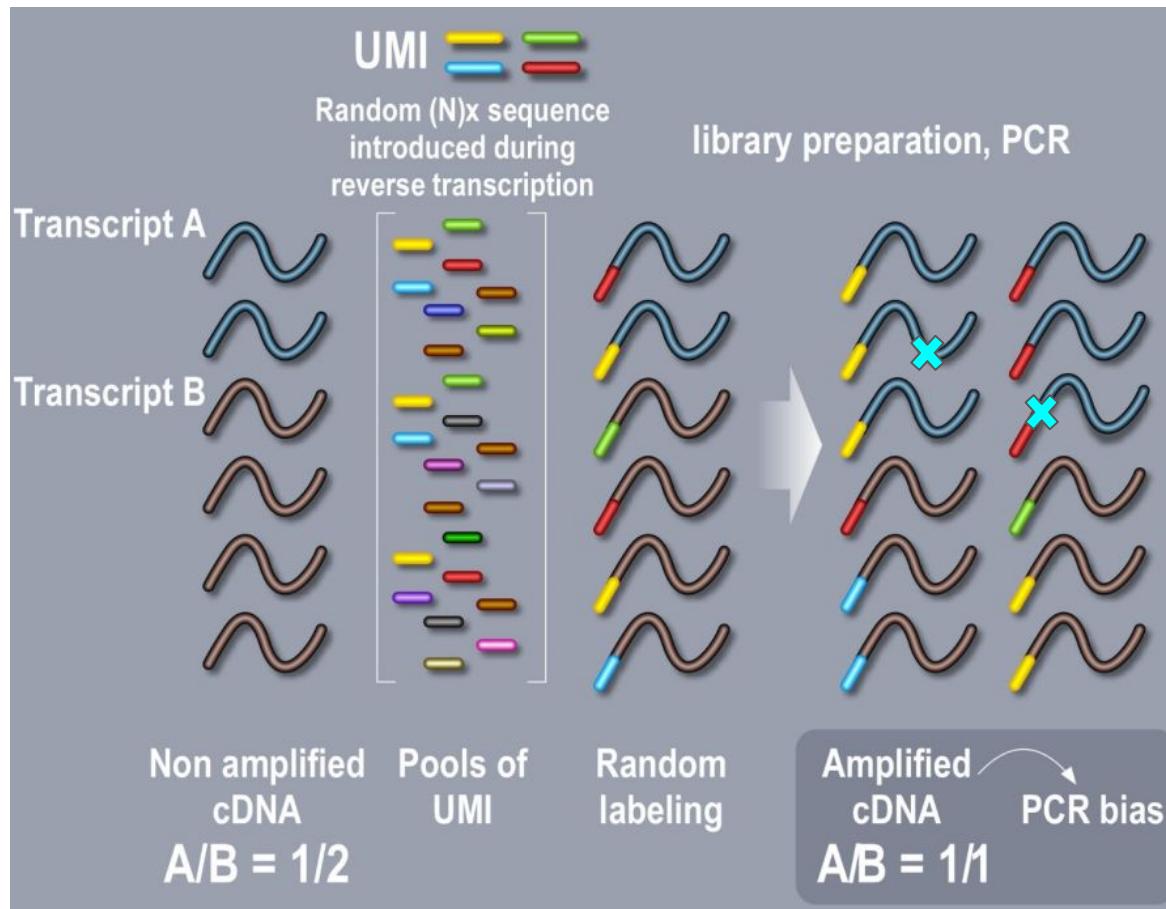


Focus on : Unique Molecule Identifiers



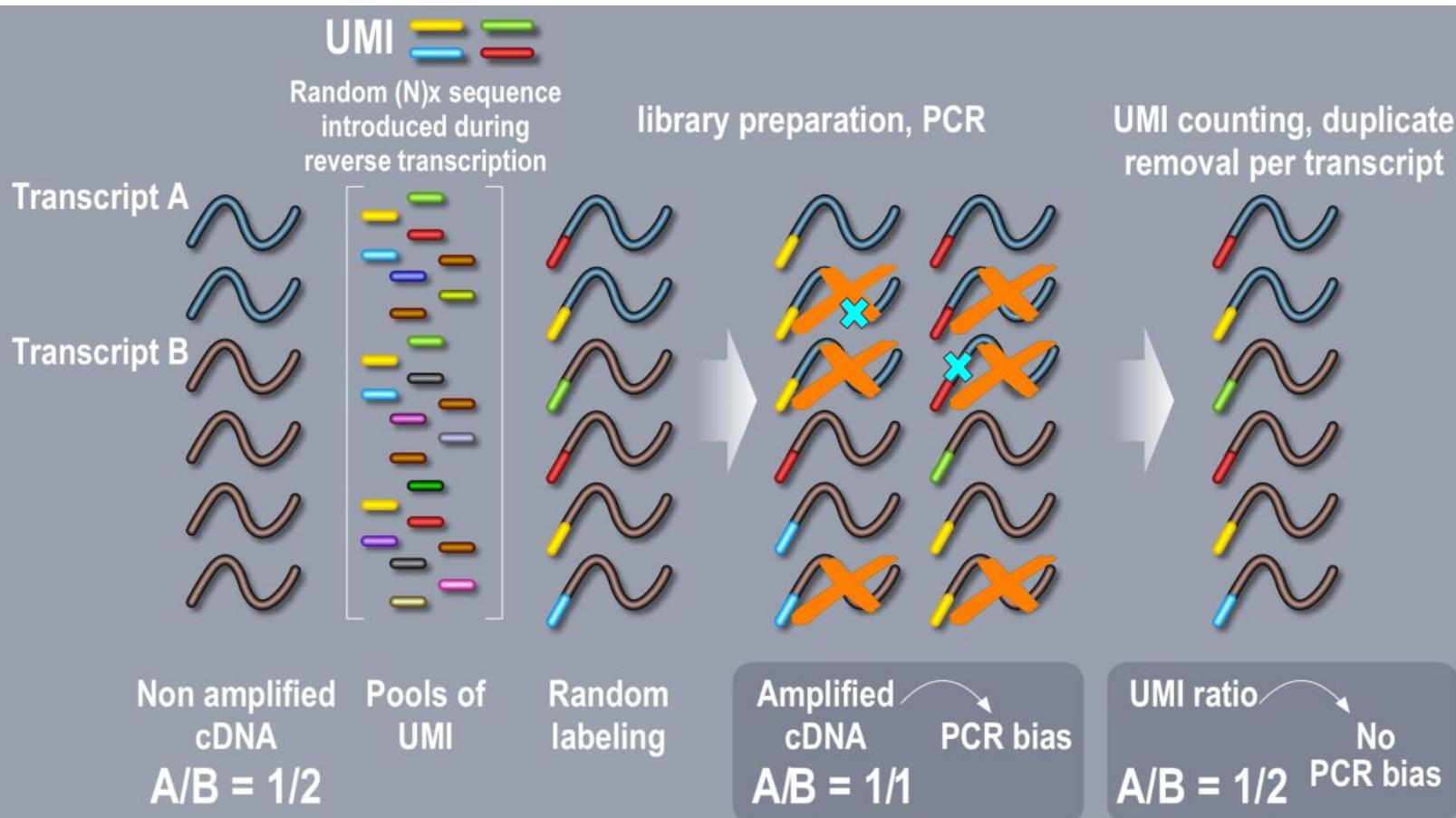
Islam et al., Nature Methods (2014)
Study by Agnès Paquet

Focus on : Unique Molecule Identifiers



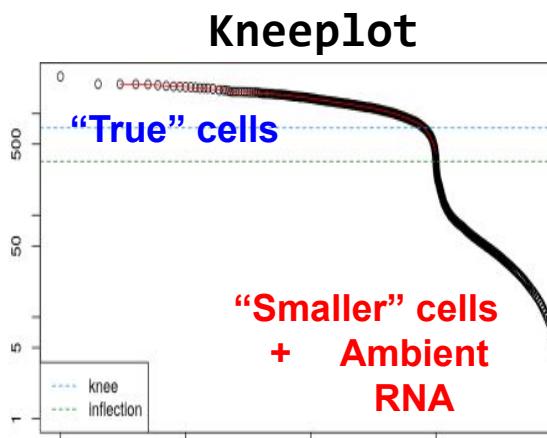
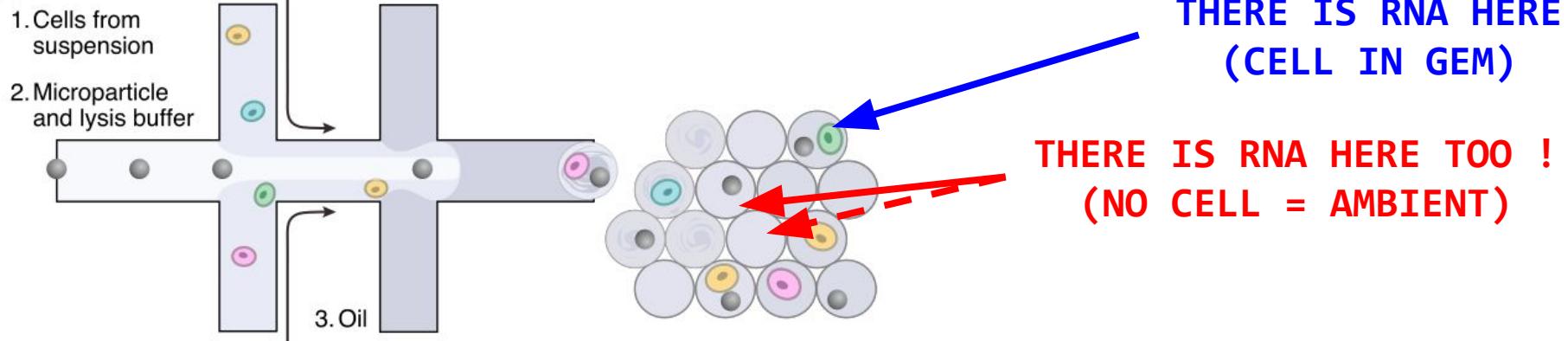
Islam et al., Nature Methods (2014)
Study by Agnès Paquet

Focus on : Unique Molecule Identifiers

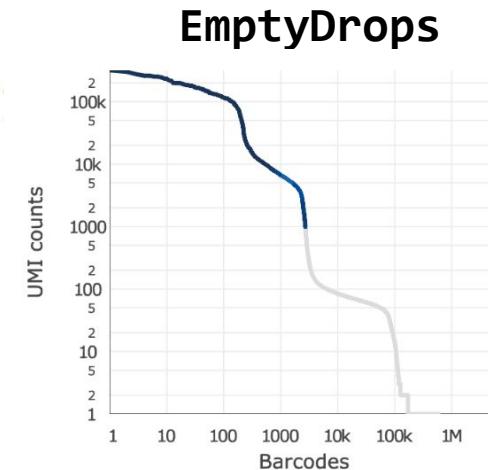
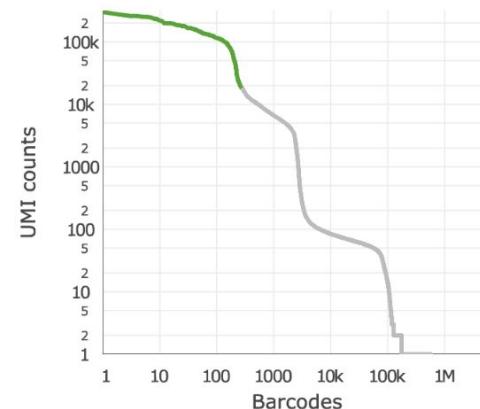


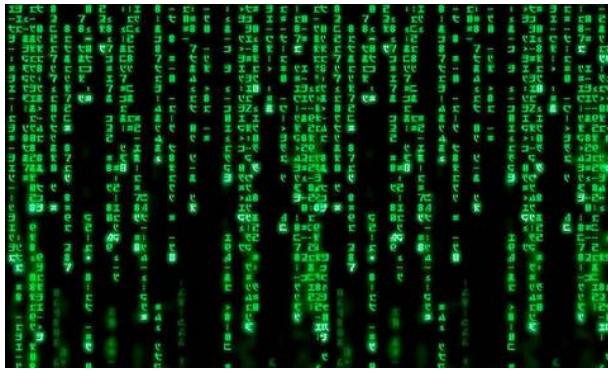
Islam et al., Nature Methods (2014)
Study by Agnès Paquet

Focus on : Empty droplets filtering



<<





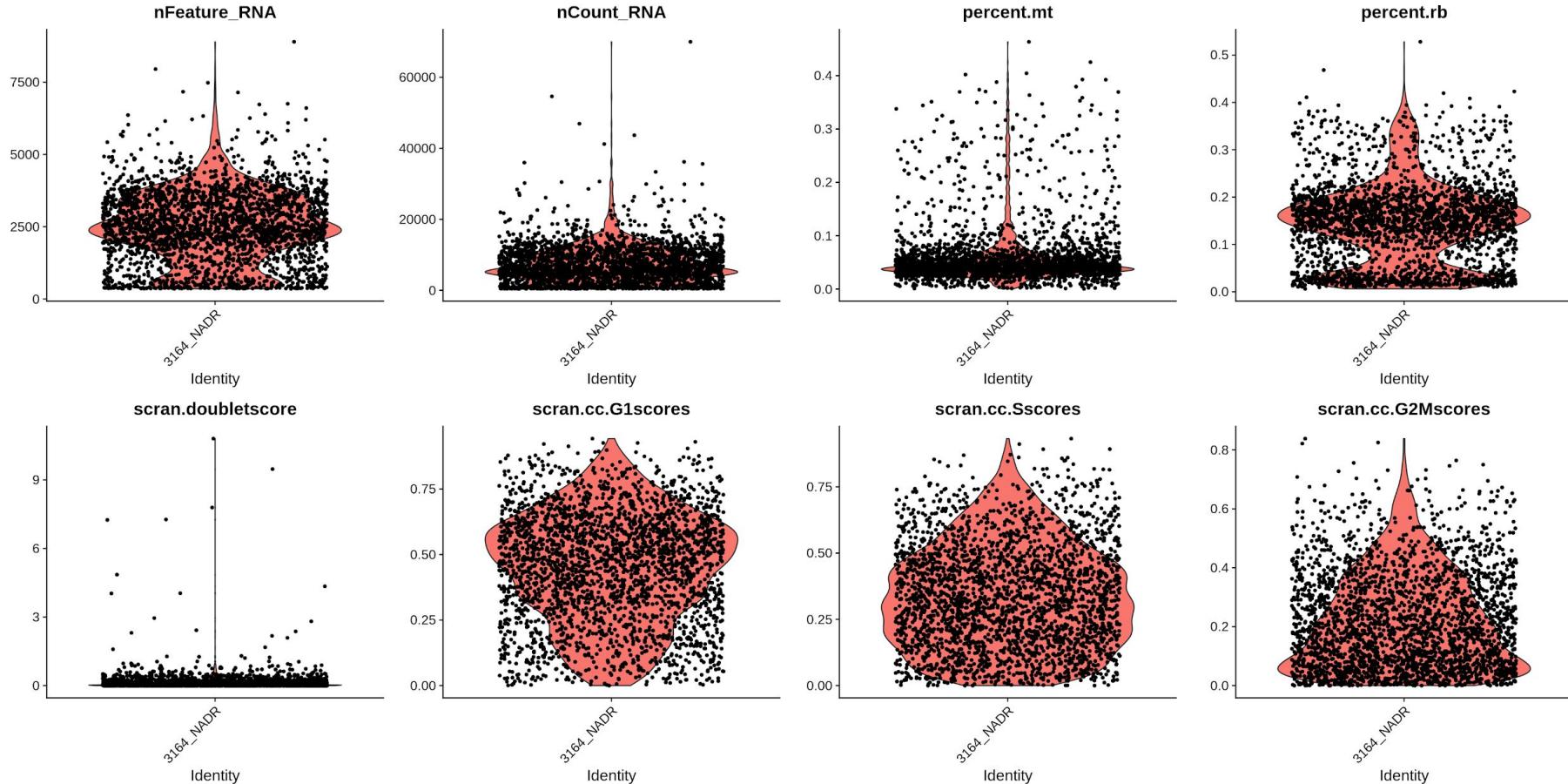
*From a
Raw count matrix
To a
Normalized matrix*



Cell QC considerations

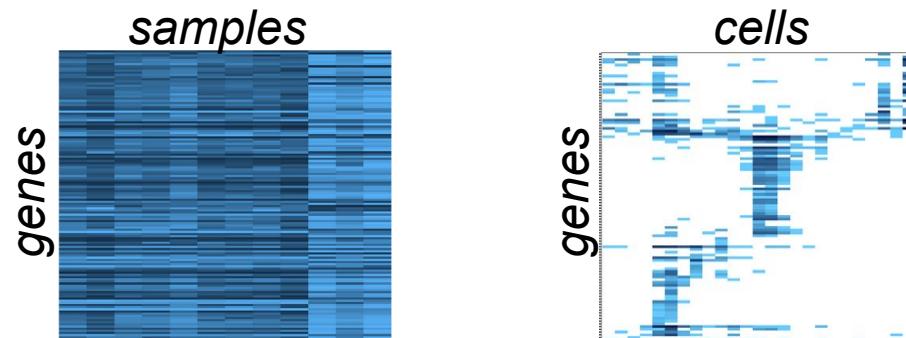
- The number of unique genes detected in each cell :
 - Low-quality cells or empty droplets will often have very few genes
 - Cell doublets or multiplets may exhibit an aberrantly high gene count
- Similarly, the total number of molecules detected within a cell (correlates strongly with unique genes)
- 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...

	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)

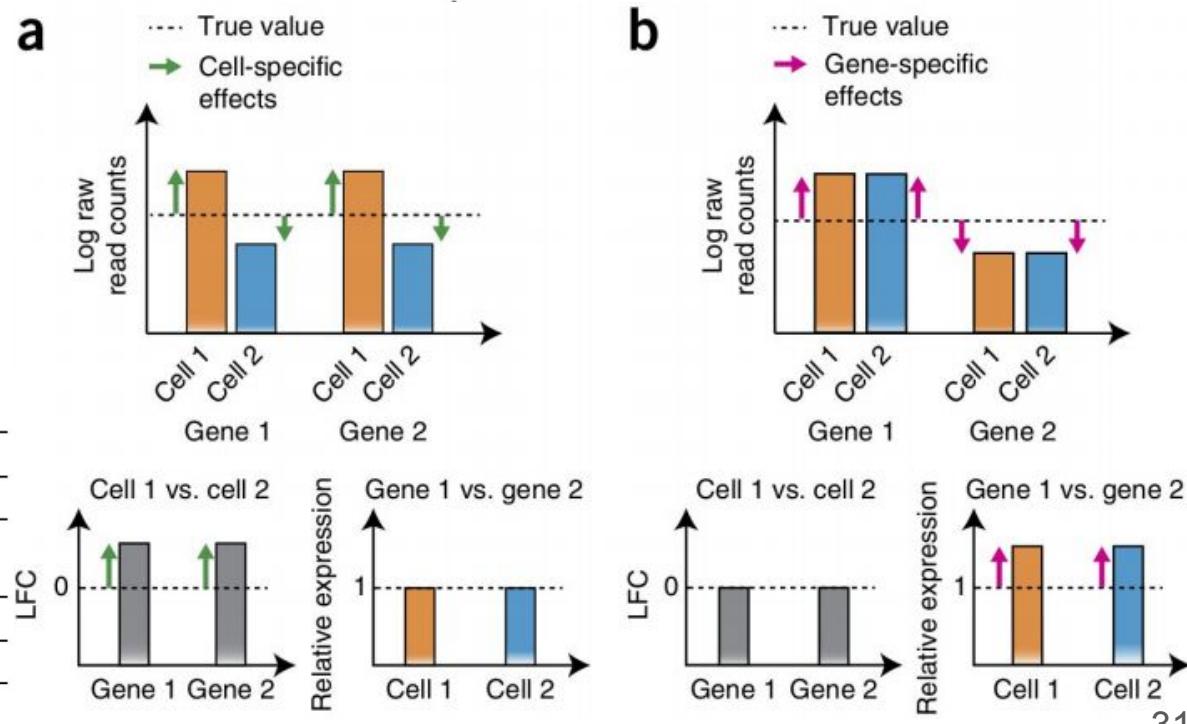


**SC MATRIX IS SPARSE !
(ie, mostly filled with zeros)**

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**

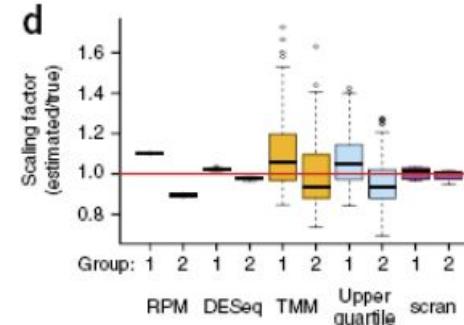
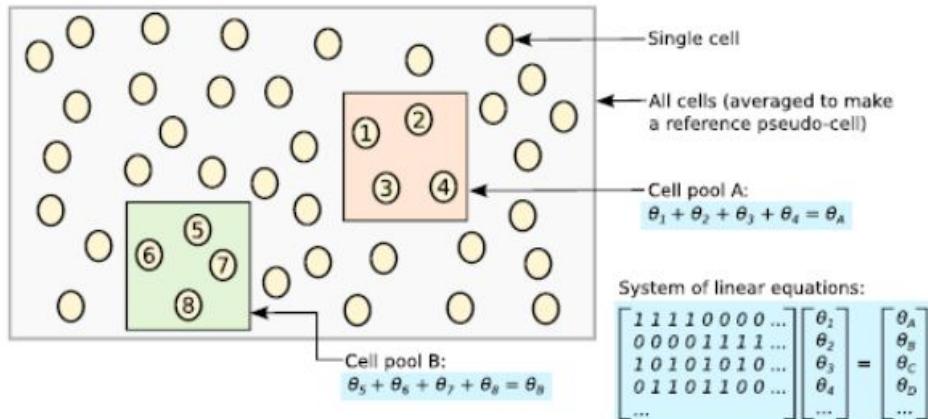
- RPKM/FPKM (Reads/Fragments per kilobase of transcript per million reads of library) : Normalize for sequencing depth and transcript length at the same time => **KO if you DO NOT have full-length data**
- Global scaling (eg: Upper Quartile) : **KO if you have too many zeros**
- Size factors calculation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

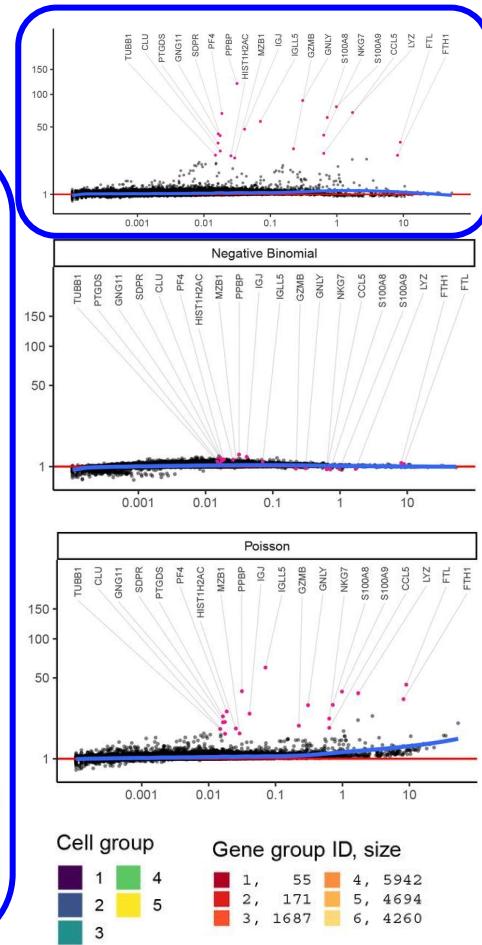
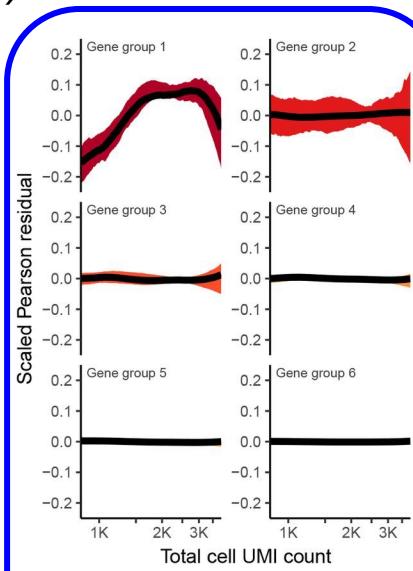
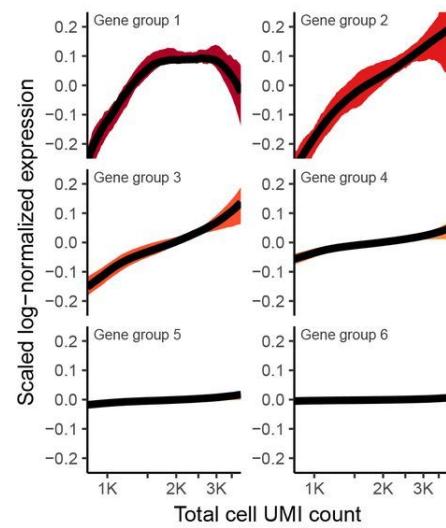
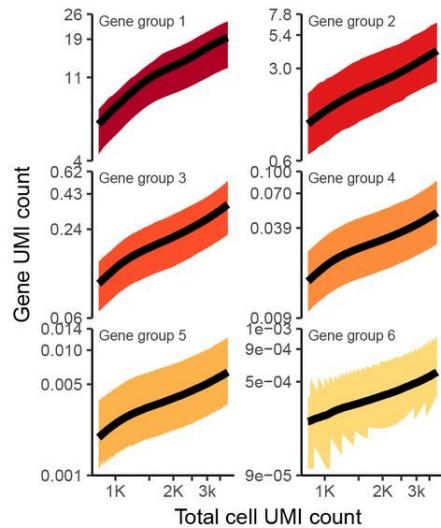
- Alternative method to compute the size factors
- Pool cells to reduce the number of zeros
- Estimate the size factors for the pool
- Repeat many time and use deconvolution to estimate each cell size factor
- Implemented in **scater/scran** packages



Vallejos C, 2017

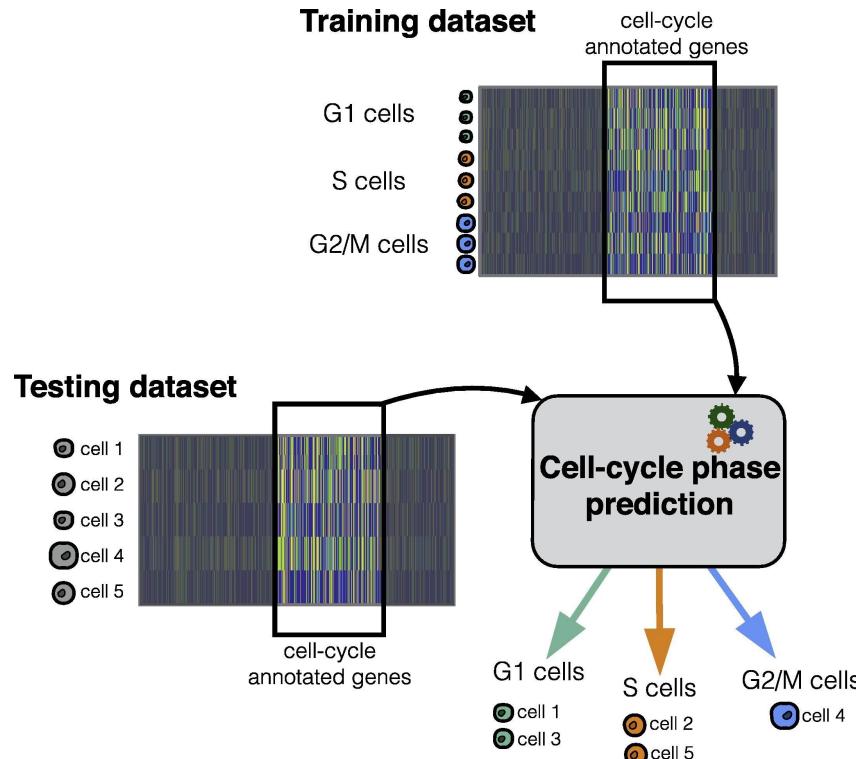
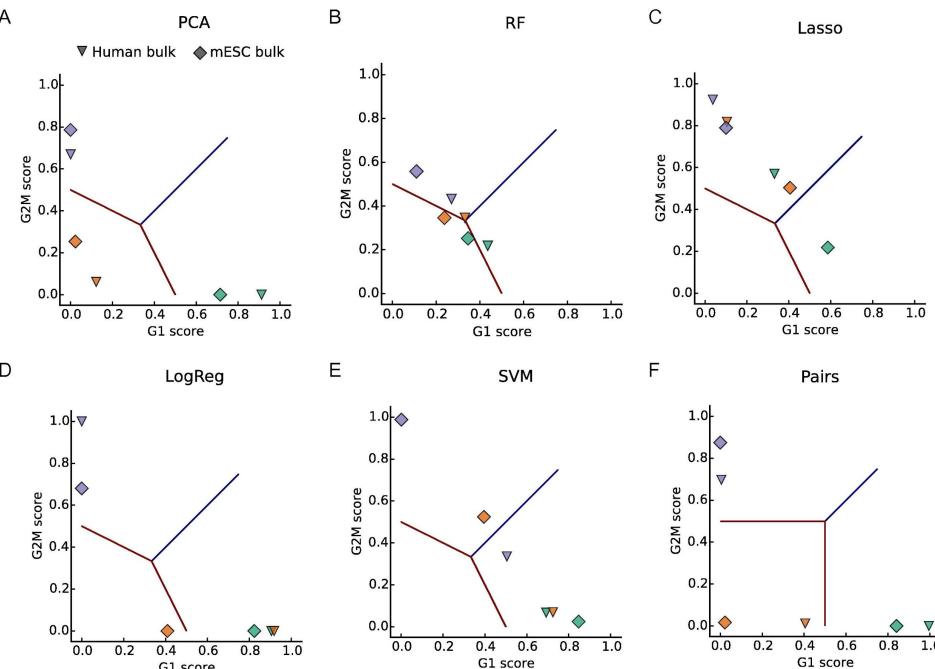
Matrix normalization : variance stabilization

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



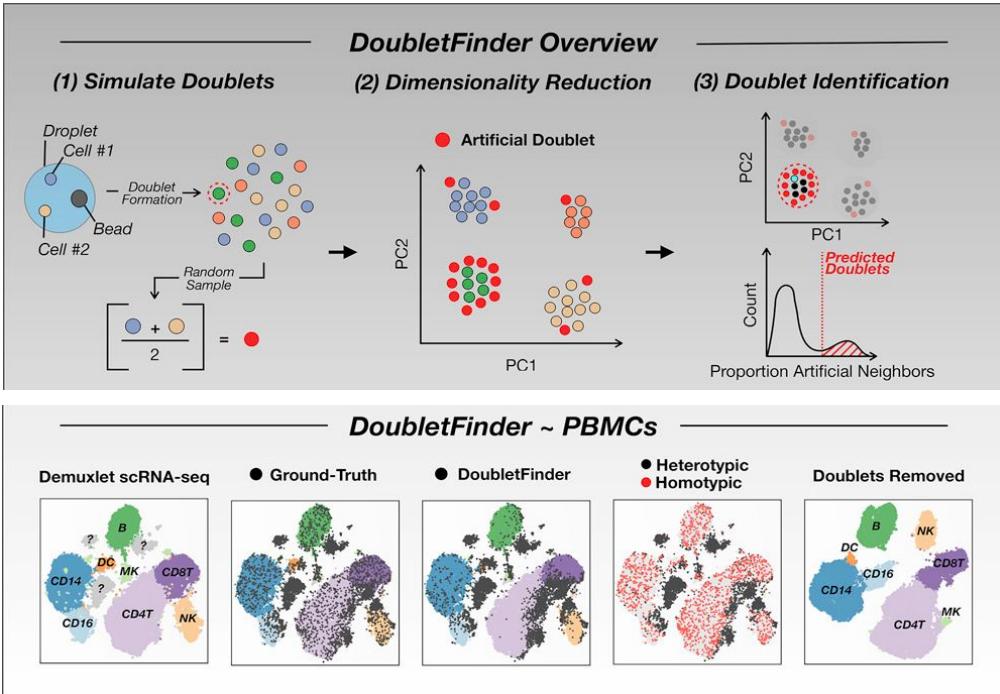
Known biases : Cell cycle phase

- 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*)



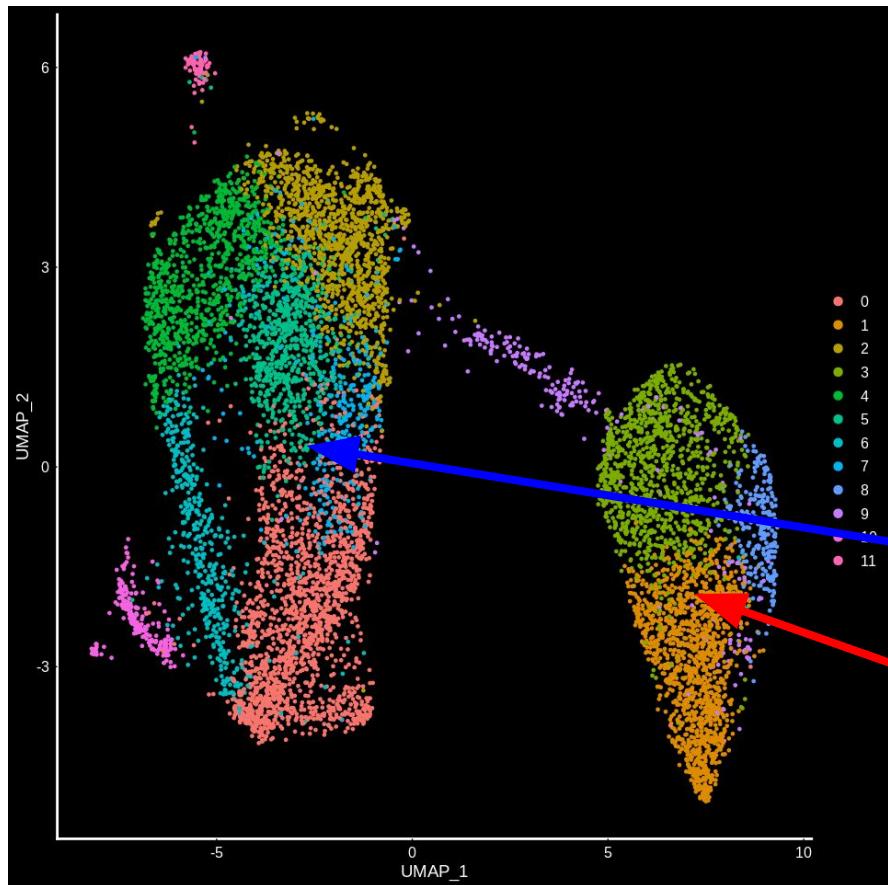
Known biases : Cell doublets

- 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
ch_cell-lines					
● 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
● cxds	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
● bcds	0.70	0.66	0.64	0.62	0.43
ch_pbmc					
● 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
● cxds	0.78	0.69	0.65	0.61	0.54
● features	0.79	0.62	0.57	0.54	0.45
● bcds	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
demuxlet					
● 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
● cxds	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
● bcds	0.91	0.79	0.71	0.62	0.61
● dblFinder	0.92	0.79	0.70	0.63	0.62
hg-mm					
● 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
● bcds	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
● cxds	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

Known biases : an IRL example

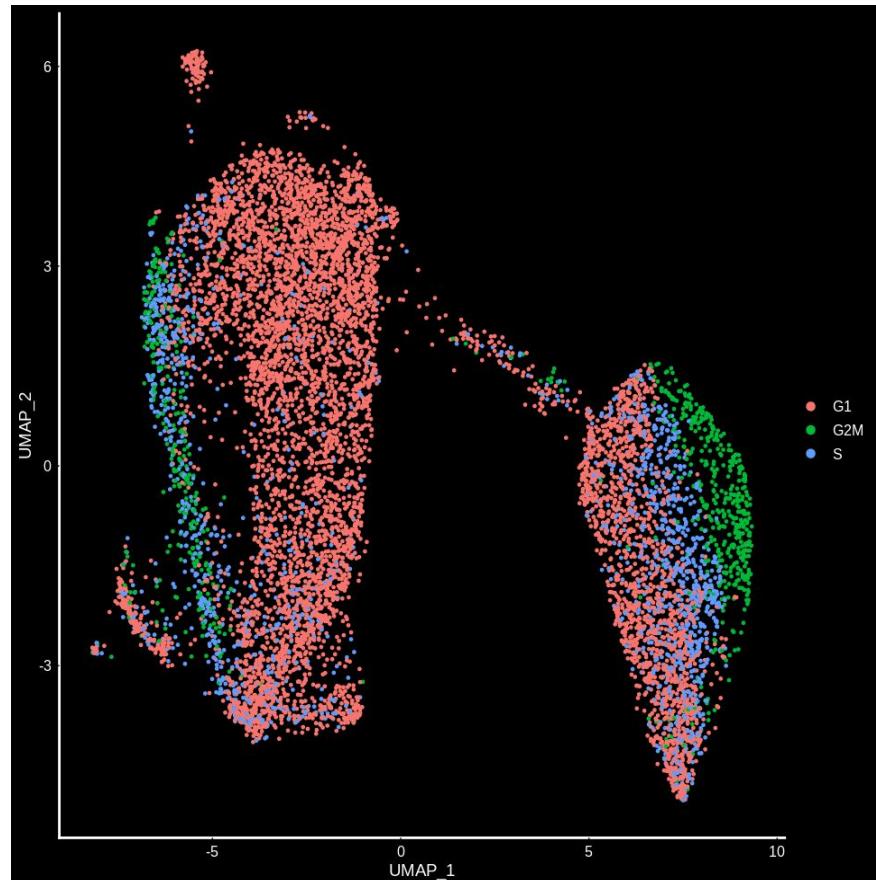
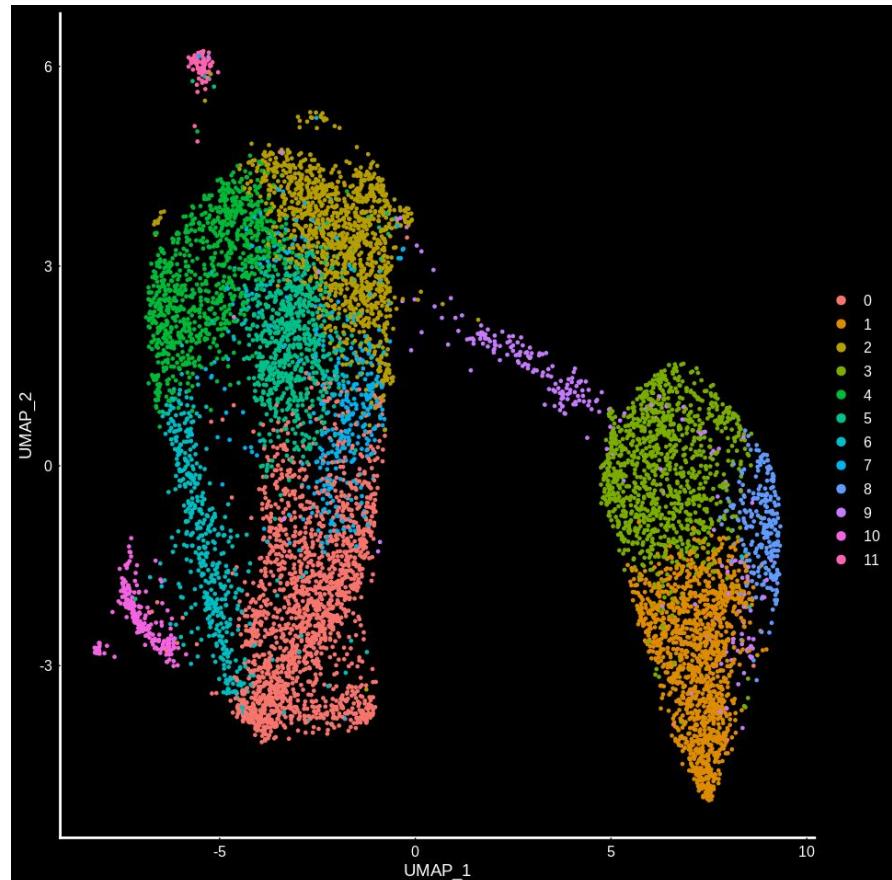


- 10X 3' scRNAseq v2
- Osteosarcoma metastasis
- 8911 cells x 18613 genes
- PCA (109 PCs retained)
- Louvain clustering
 - 12 clusters
- uMAP representation

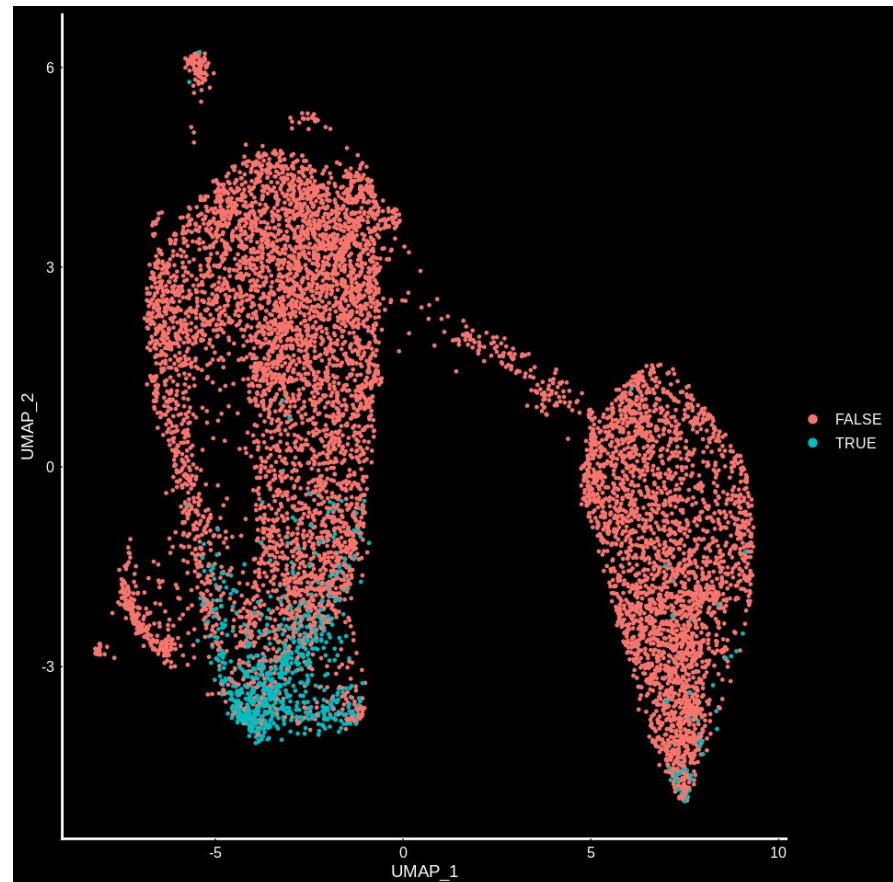
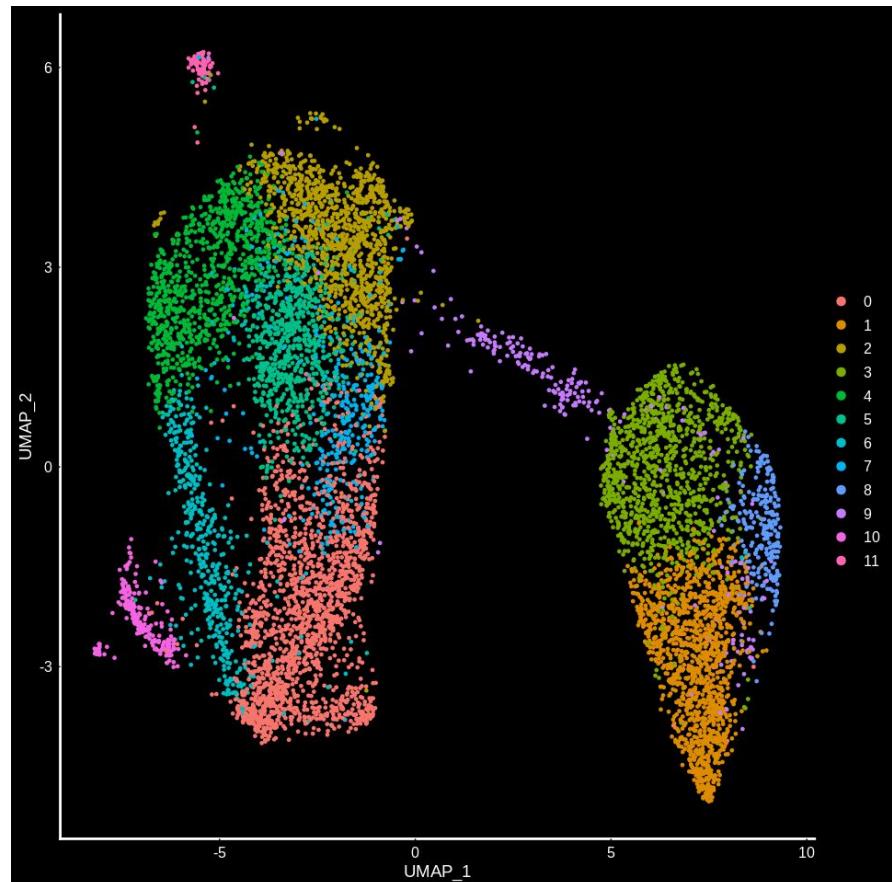
Osteoblasts

Osteoclasts

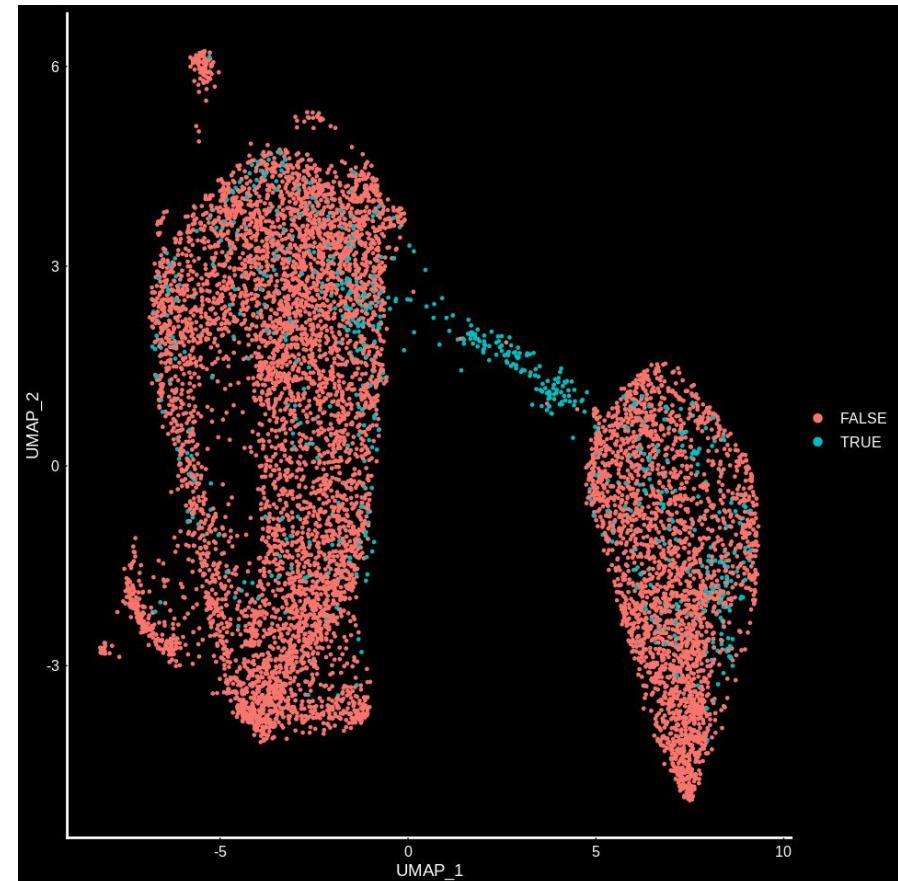
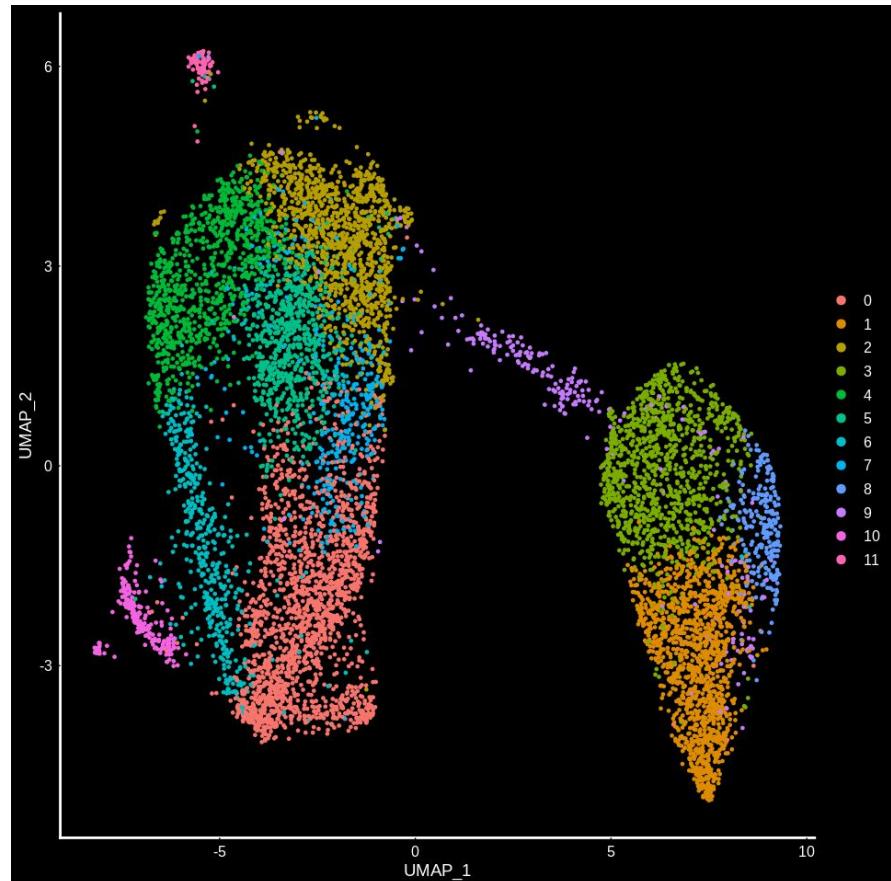
Bias : Cell cycle phases / scores



Bias : Dying cells status / score

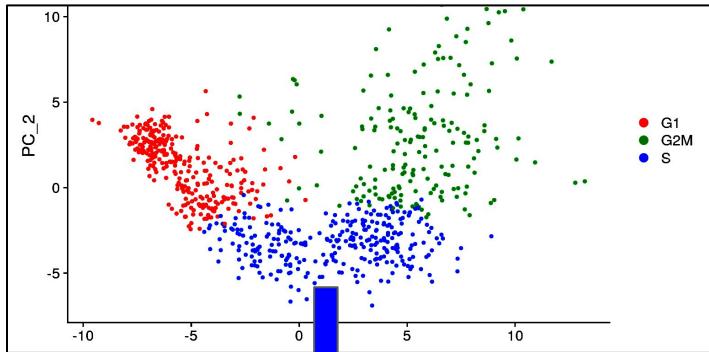


Bias : Cell doublet status / score

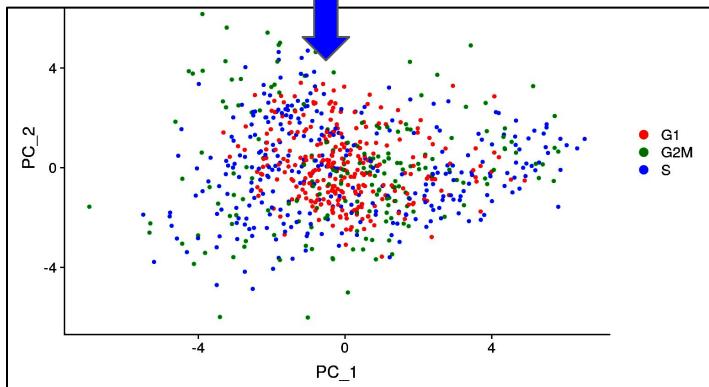


Bias normalization : regression / deblocking

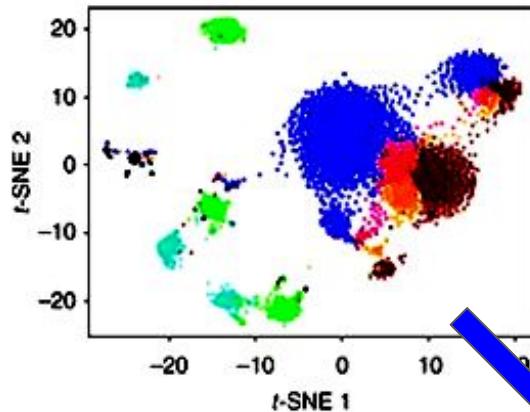
Ex : Cell cycle score



Regression

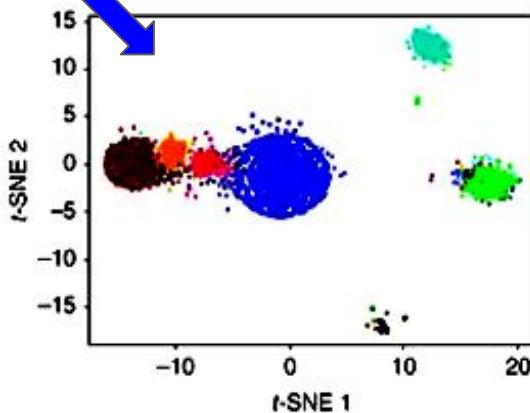


Ex : Batch effect



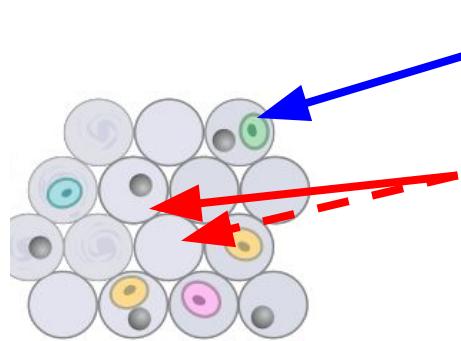
Deblocking

Seurat tutorial



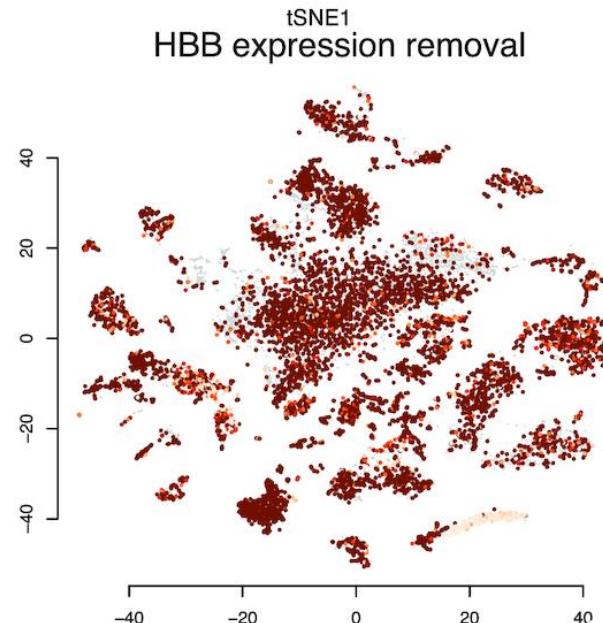
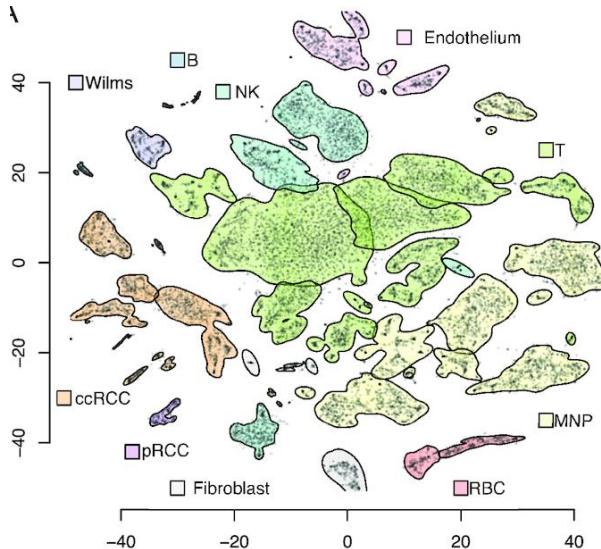
Ambient RNA filtering (soupX)

- emptyDrops : removed empty droplets (contained only ambient RNA)
- **BUT** non-empty droplets **ALSO** have ambient RNA !
- **soupX** determines the amount of ambient RNA in counts, removes it



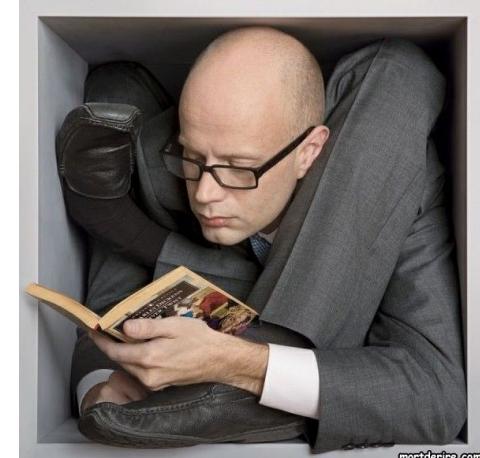
THERE IS RNA HERE
(CELL IN GEM RNA
+ AMBIENT)

THERE IS RNA HERE TOO !
(NO CELL = 100% AMBIENT)





*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)

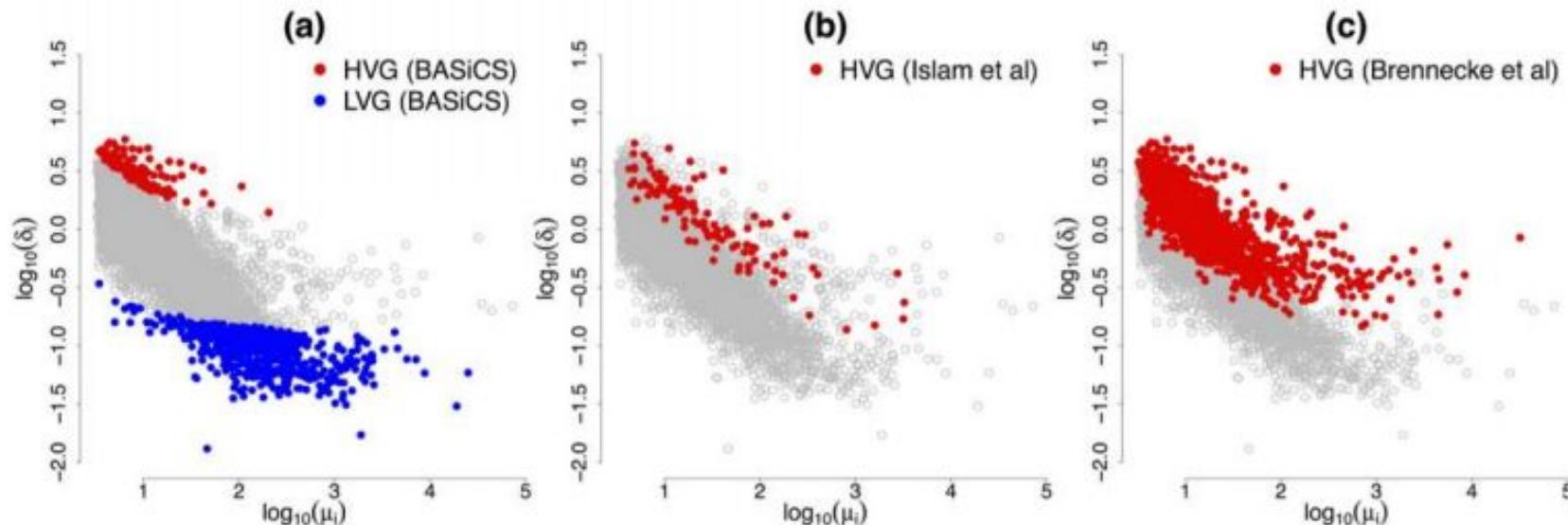


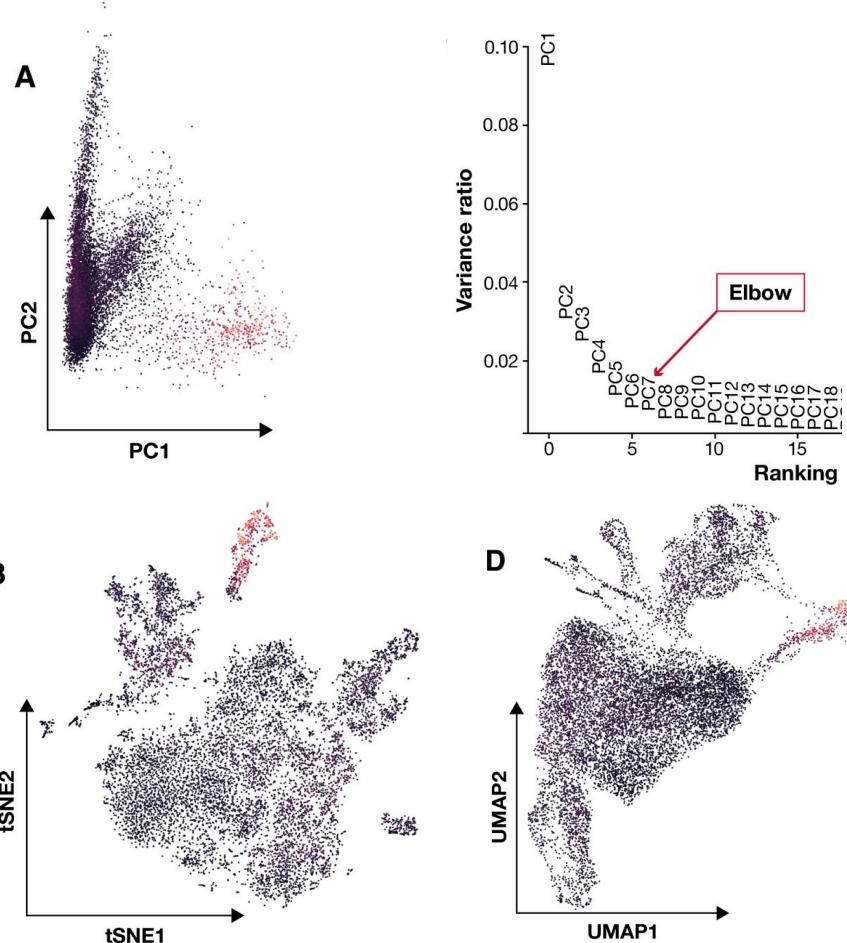
Fig 8. Comparison of HVG detection among different methods. For each of the 7,895 biological genes, posterior medians of biological cell-to-cell heterogeneity term δ_i (log scale) against posterior medians of expression level μ_i (log scale). While the methods described in [16] and [5] only provide a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

Dimensionality reduction : simplification + selection

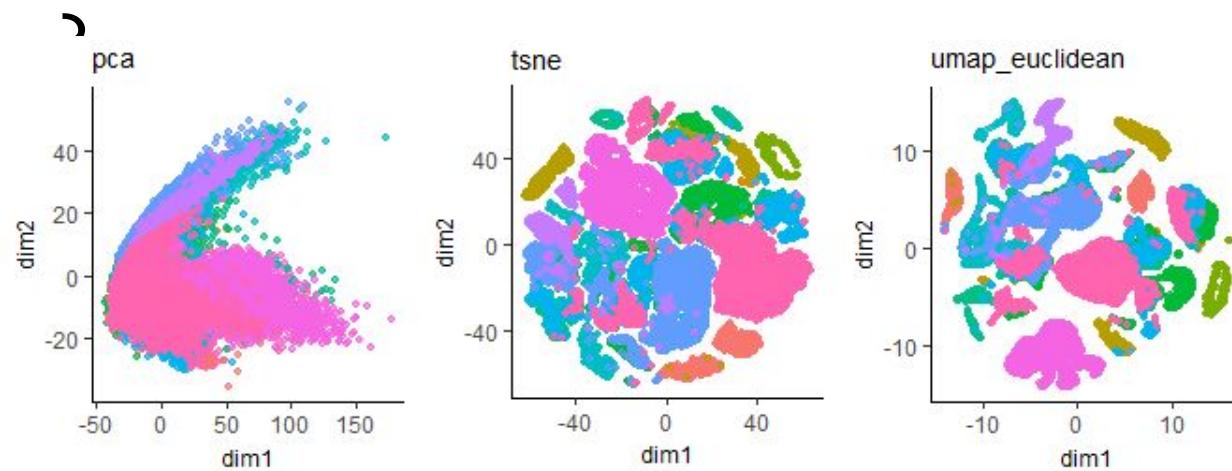
1. Need of an orthogonal space
2. Minimize curse of dimensionality
3. Filter out noise
4. Allow visualization
5. Reduce computational load

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

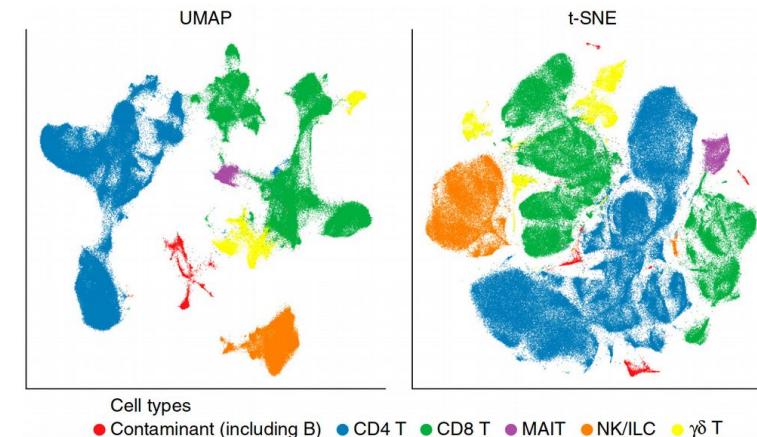


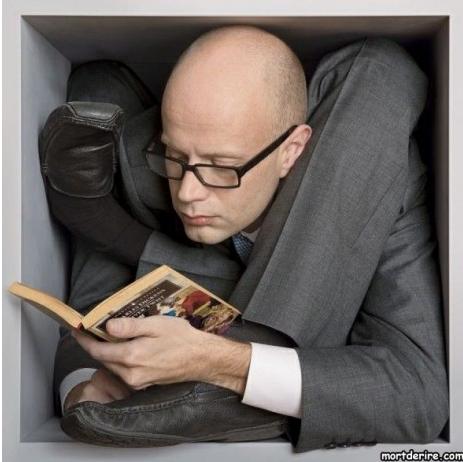
Reduction :

- 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 parameters modification
- (Slightly faster to generate)

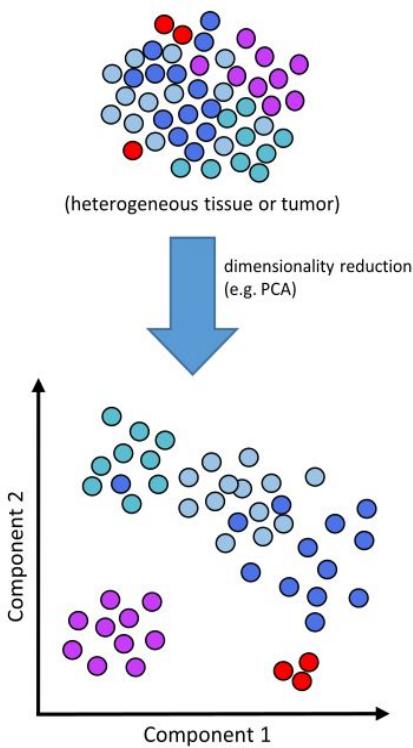




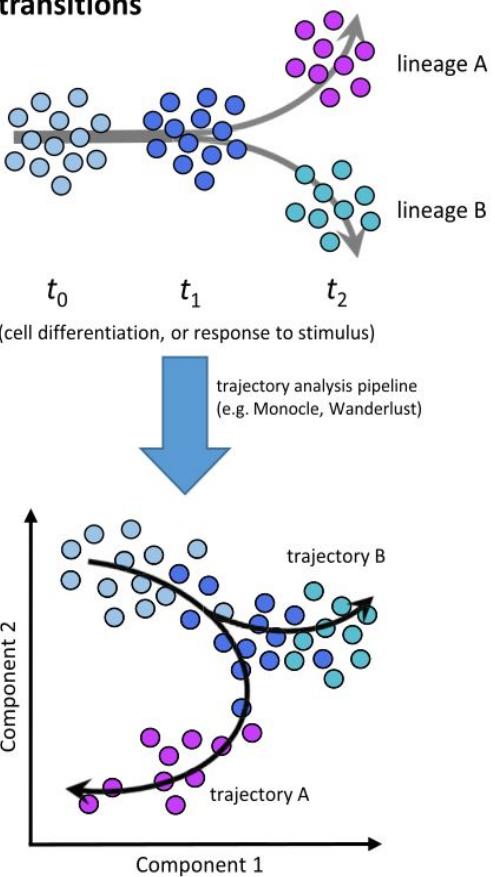
*From a
Reduced space
To ...
... finally what you wanted !*



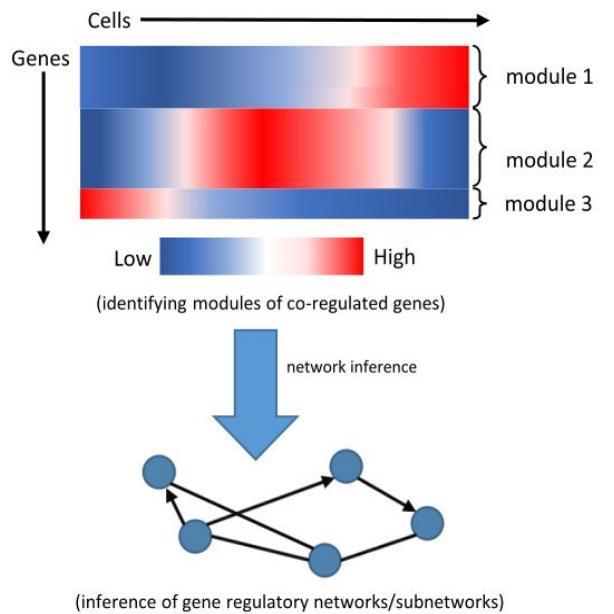
a) Deconvolving heterogeneous cell populations



b) Trajectory analysis of cell state transitions

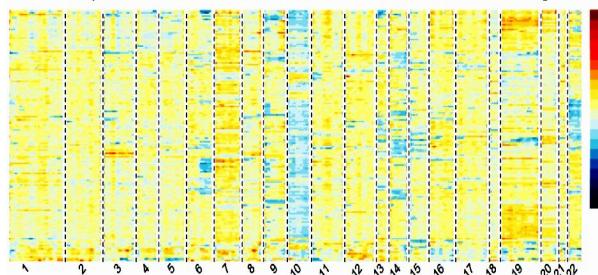


d) Network inference

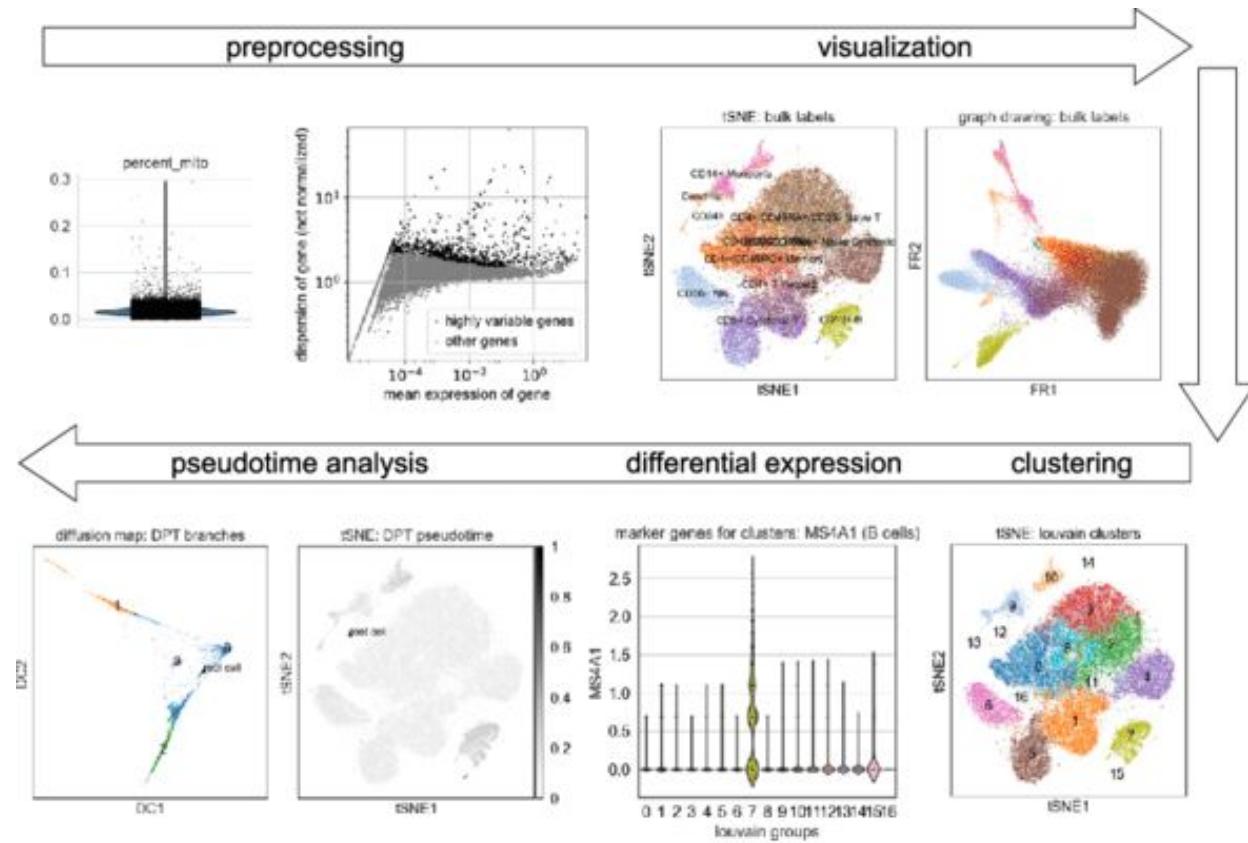


e) Copy number estimation

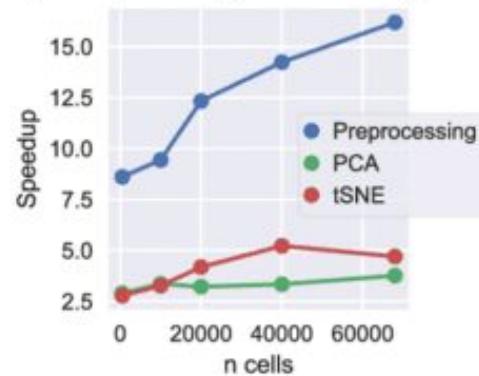
Expression Matrix after Recursive Multiscale Median Filtering



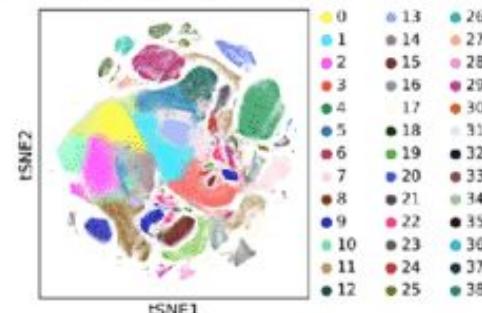
The all-in-one Python toolbox : Scanpy



b Speedup: Scanpy vs. Cell Ranger R



c tSNE of clustered 1.3 million cells



The all-in-one R toolbox : Seurat

Seurat 4.0.4

Install

Get started

Vignettes ▾

Extensions

FAQ

News

Reference

Archive



Official release of Seurat 4.0

We are excited to release Seurat v4.0! This update brings the following new features and functionality:

- **Integrative multimodal analysis.** The ability to make simultaneous measurements of multiple data types from the same cell, known as multimodal analysis, represents a new and exciting frontier for single-cell genomics. In Seurat v4, we introduce weighted nearest neighbor (WNN) analysis, an unsupervised strategy to learn the information content of each modality in each cell, and to define cellular state based on a weighted combination of both modalities. In our new paper, we generate a CITE-seq dataset featuring paired measurements of the transcriptome and 228 surface proteins, and leverage WNN to define a multimodal reference of human PBMC. You can use WNN to analyze multimodal data from a variety of technologies, including CITE-seq, ASAP-seq, 10X Genomics ATAC +

Links

Download from CRAN at
[https://cloud.r-project.org/
package=Seurat](https://cloud.r-project.org/package=Seurat)

Browse source code at
[https://github.com/satijalab/
seurat/](https://github.com/satijalab/seurat/)

Report a bug at
[https://github.com/satijalab/
seurat/issues](https://github.com/satijalab/seurat/issues)

License

[GPL-3](#) | file [LICENSE](#)

Community

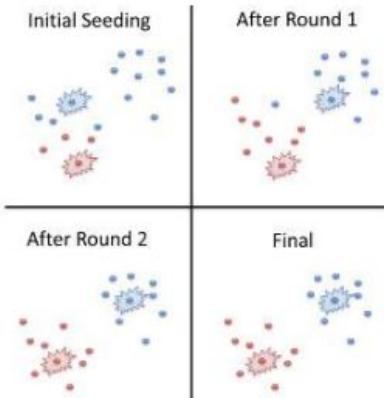
[Code of conduct](#)

Citation

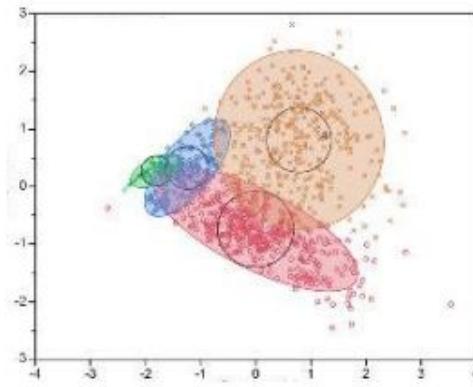
[Citing Seurat](#)

Cell clustering : methods

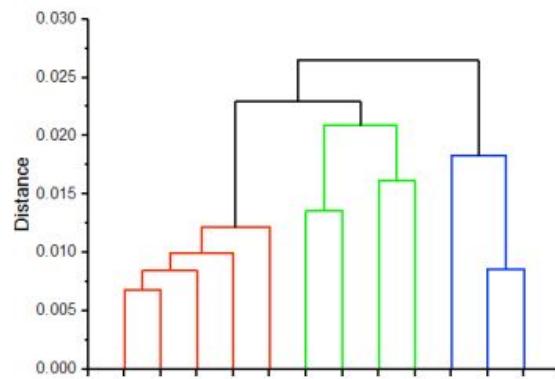
1) K-means based



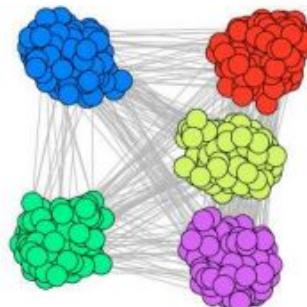
3) Model-based clustering (Mclust)



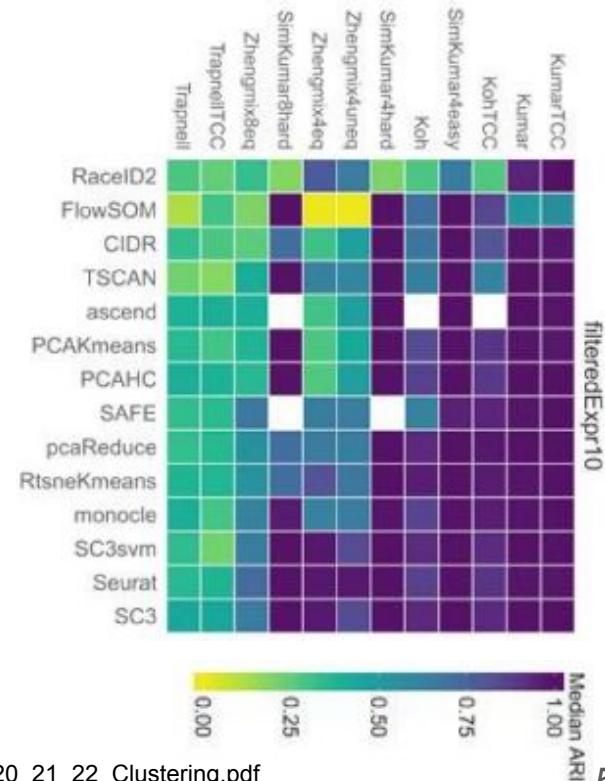
2) Hierarchical clustering



4) Graph-based clustering

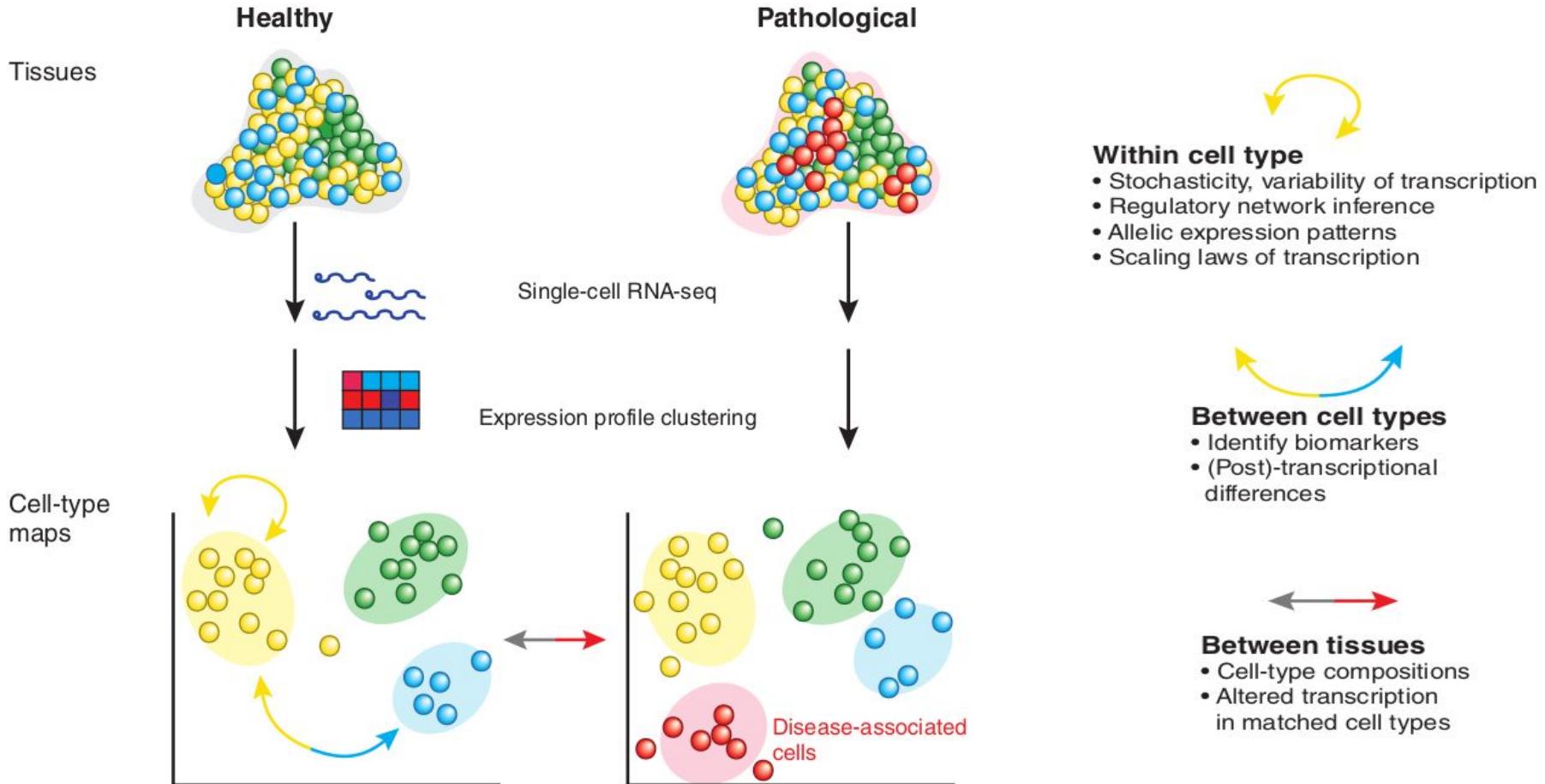


5) Single-cell specific methods



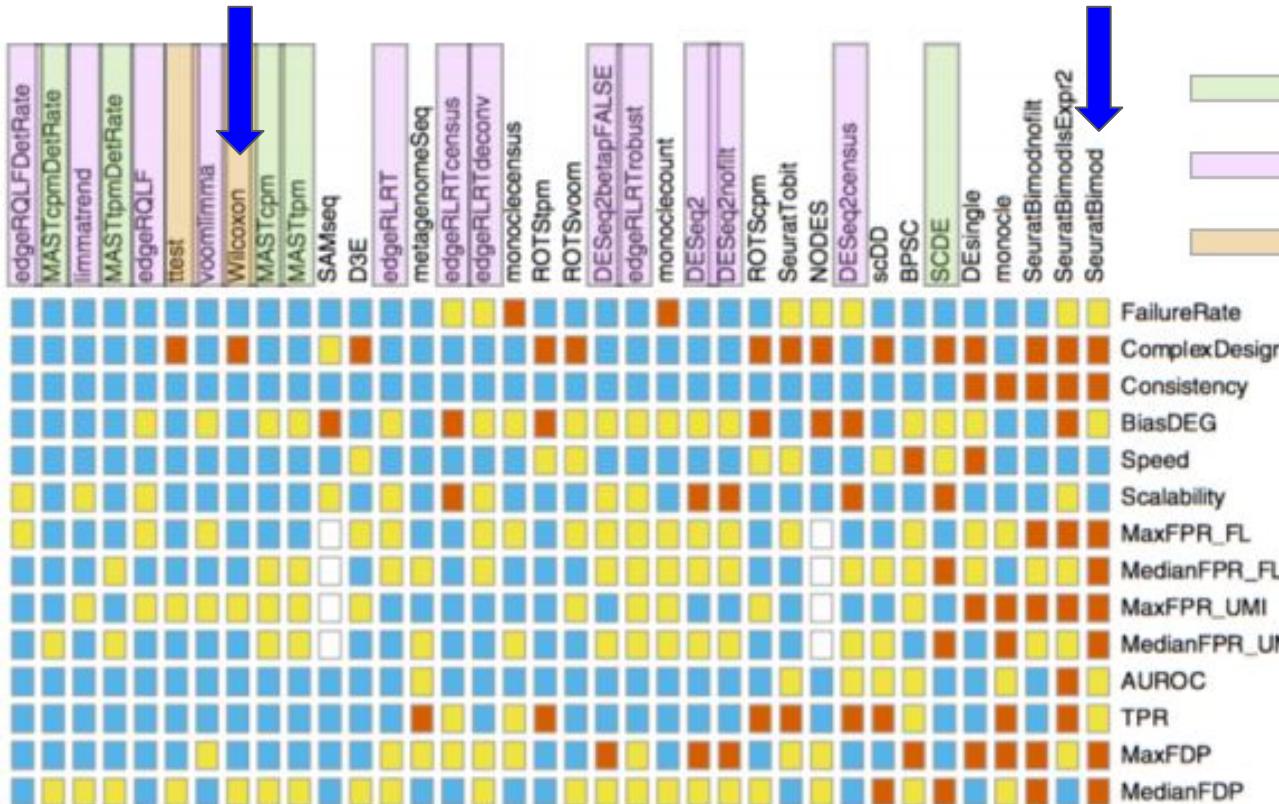
See : https://www.cse.msu.edu/~cse802/S17/slides/Lec_20_21_22_Clustering.pdf

Differential expression analysis



Differential expression analysis : methods

Seurat v3



Seurat v2

Bayesian 3-component model adapted to single-cell data

Methods borrowed from bulk-RNA-seq

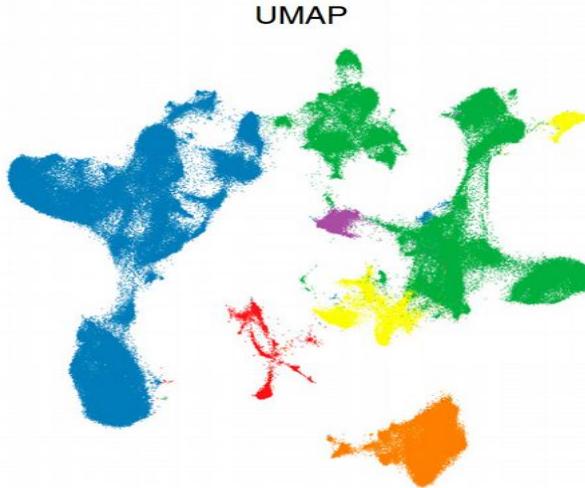
Naïve approaches

Good

Intermediate

Poor

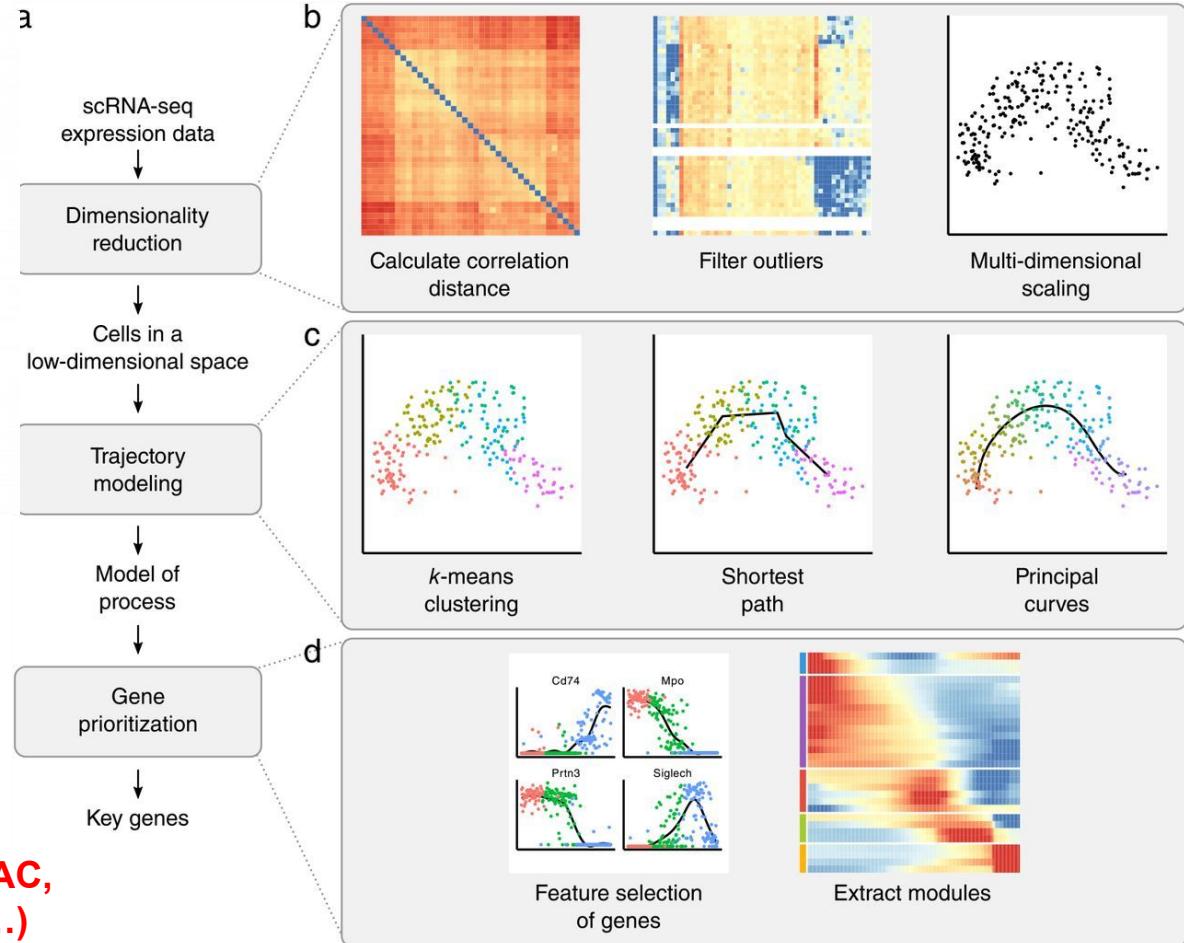
Cell trajectory : methods



Most adopted tools :

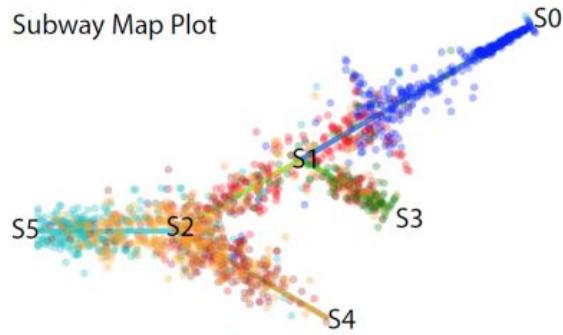
- Monocle 3
- PAGA
- STREAM
- Scorpius
- Slingshot

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

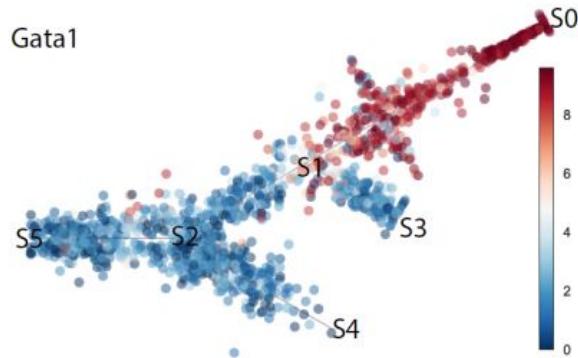


Cell trajectory : visualization

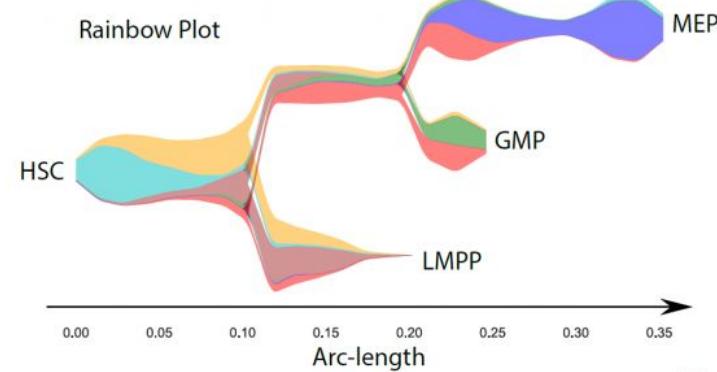
Cell distance to path + cell types



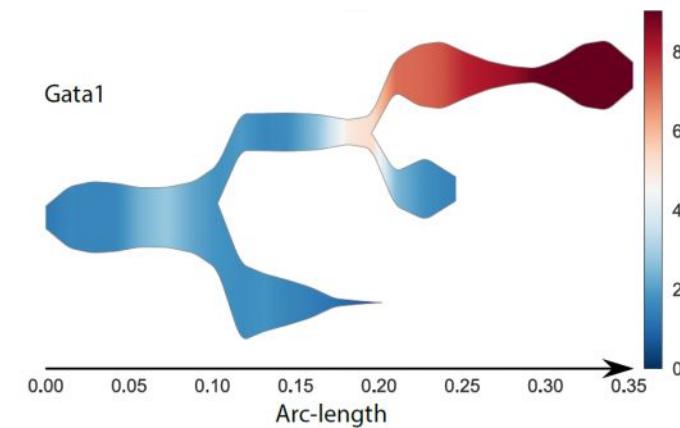
Cell distance to path + gene expression



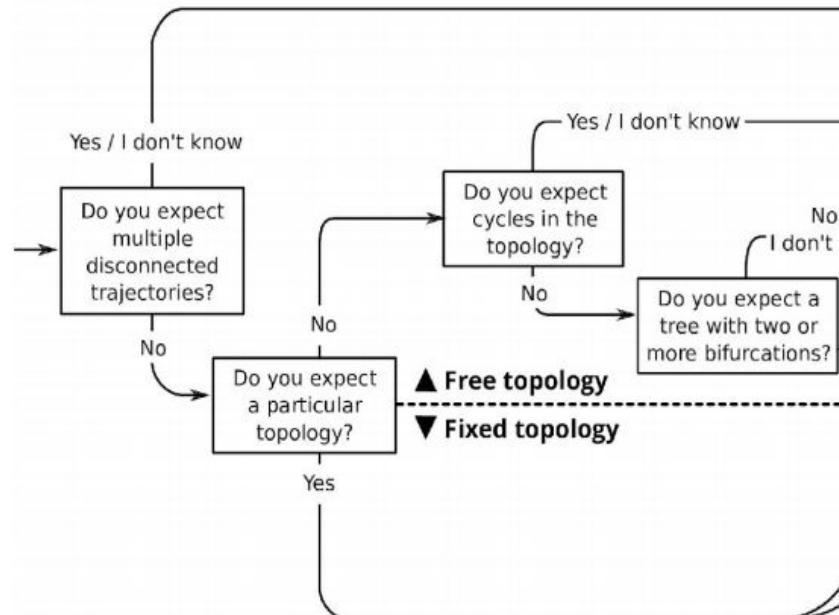
Cell types density



Cells density + gene expression



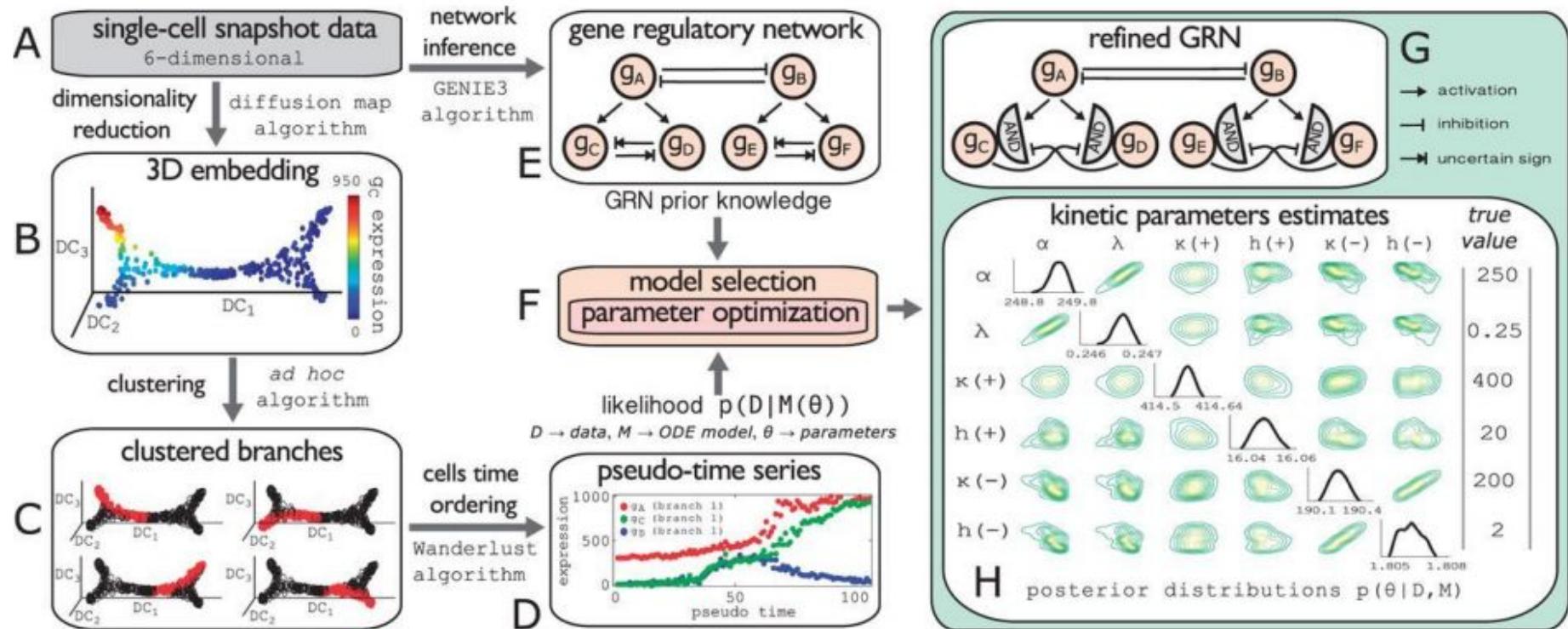
Cell trajectory : Contexts



	Estimated running time (cells × features)					
	Accuracy	Usability	100k×1k	10k×10k	1k×100k	Required priors
PAGA RacelID / StemID	+	±	7m	55s	19s	Start cell(s)
	-	±	1d	1h	1h	
PAGA RacelID / StemID SLICER	+	±	7m	55s	19s	Start cell(s)
	-	±	>7d	2h	31s	Start cell(s)
Slingshot PAGA Monocle ICA MST	+	+	11h	56m	2m	Start cell(s)
	±	±	7m	55s	19s	Number of end & start states
PAGA MST Slingshot RacelID / StemID	+	±	7m	55s	19s	Start cell(s)
	±	+	8m	12m	2m	
STEMNET Slingshot PAGA FateID	+	±	36m	12m	7m	End cell(s), Cell clustering
	+	+	11h	56m	2m	
Slingshot FateID GrandPrix STEMNET	+	±	7m	55s	19s	Start cell(s)
	+	±	6h	1h	26m	Cell clustering, Start & end cells
SCORPIUS Embeddr TSCAN Slingshot	+	±	7m	28m	>7d	# end states
	+	+	36m	12m	7m	End cell(s), Cell clustering
Angle EIPiGraph cycle reCAT RacelID / StemID	+	±	1h	4m	4m	
	±	±	2d	33m	2m	
Angle EIPiGraph cycle reCAT RacelID / StemID	+	+	7m	9m	7m	
	±	+	11h	56m	2m	

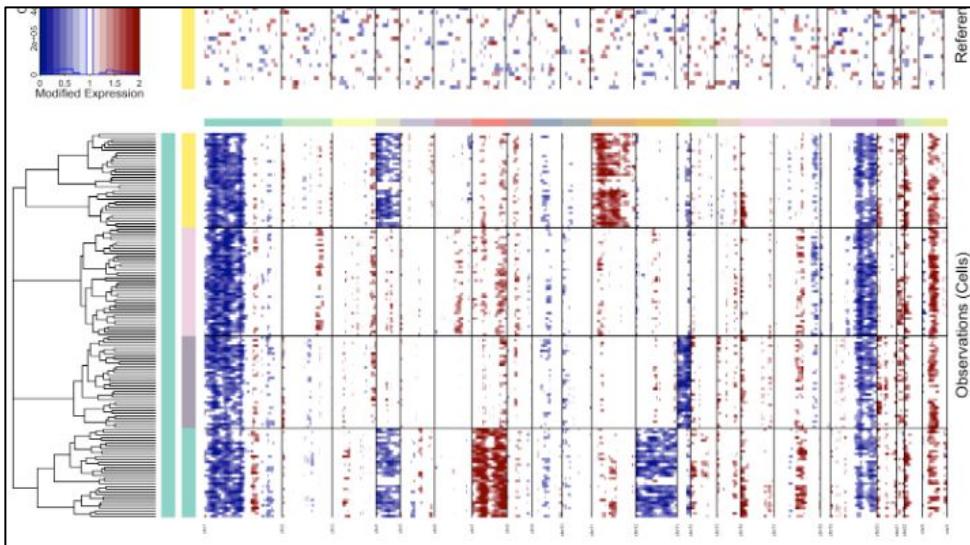
Network inference

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

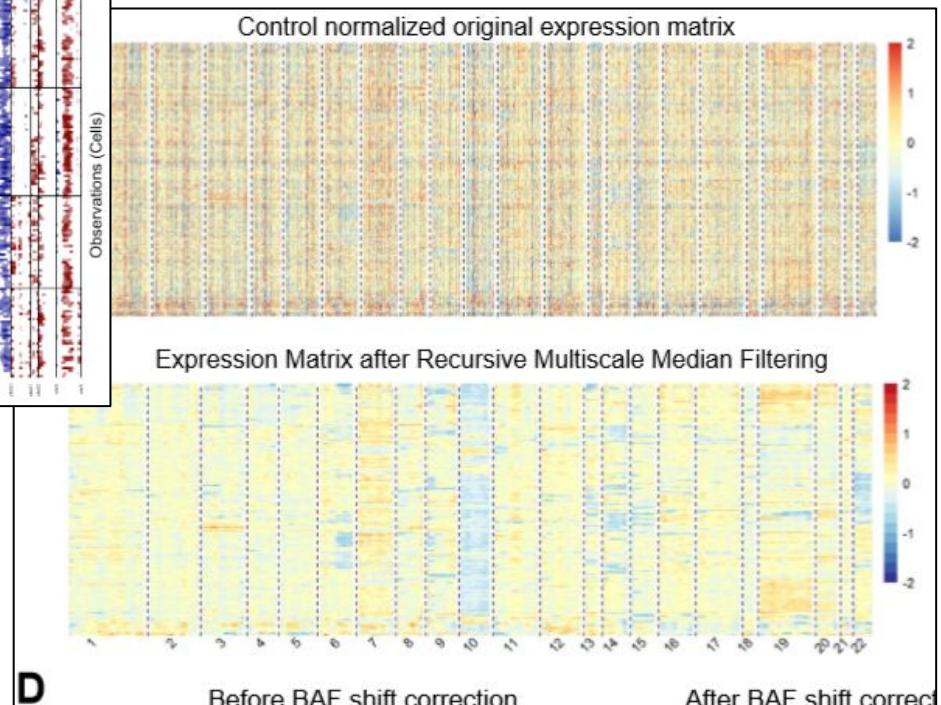


Copy number estimation from scRNASeq

InferCNV (Broad Institute)



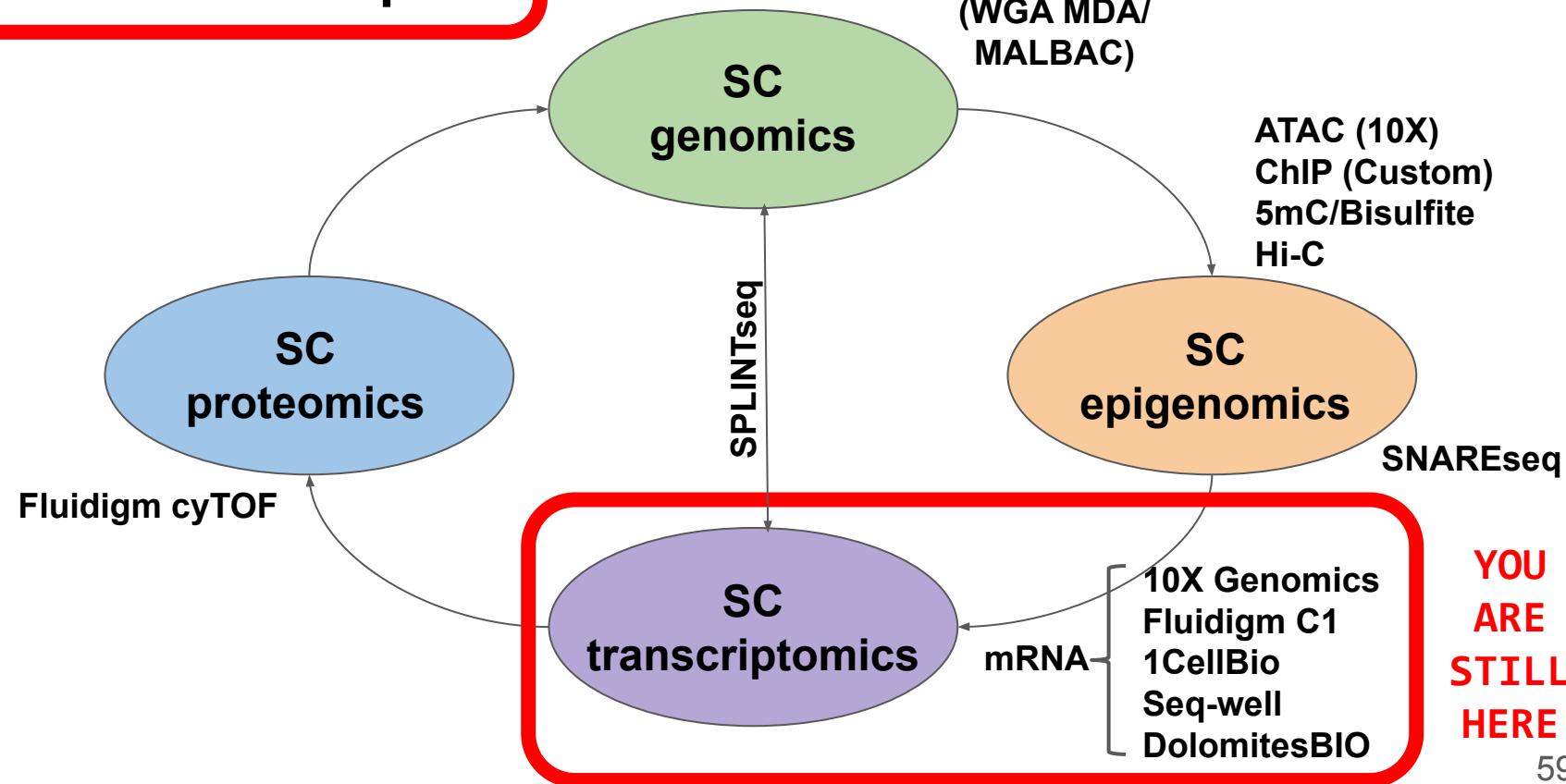
CaSpER (Armanci et al, BioRxiv 2019)

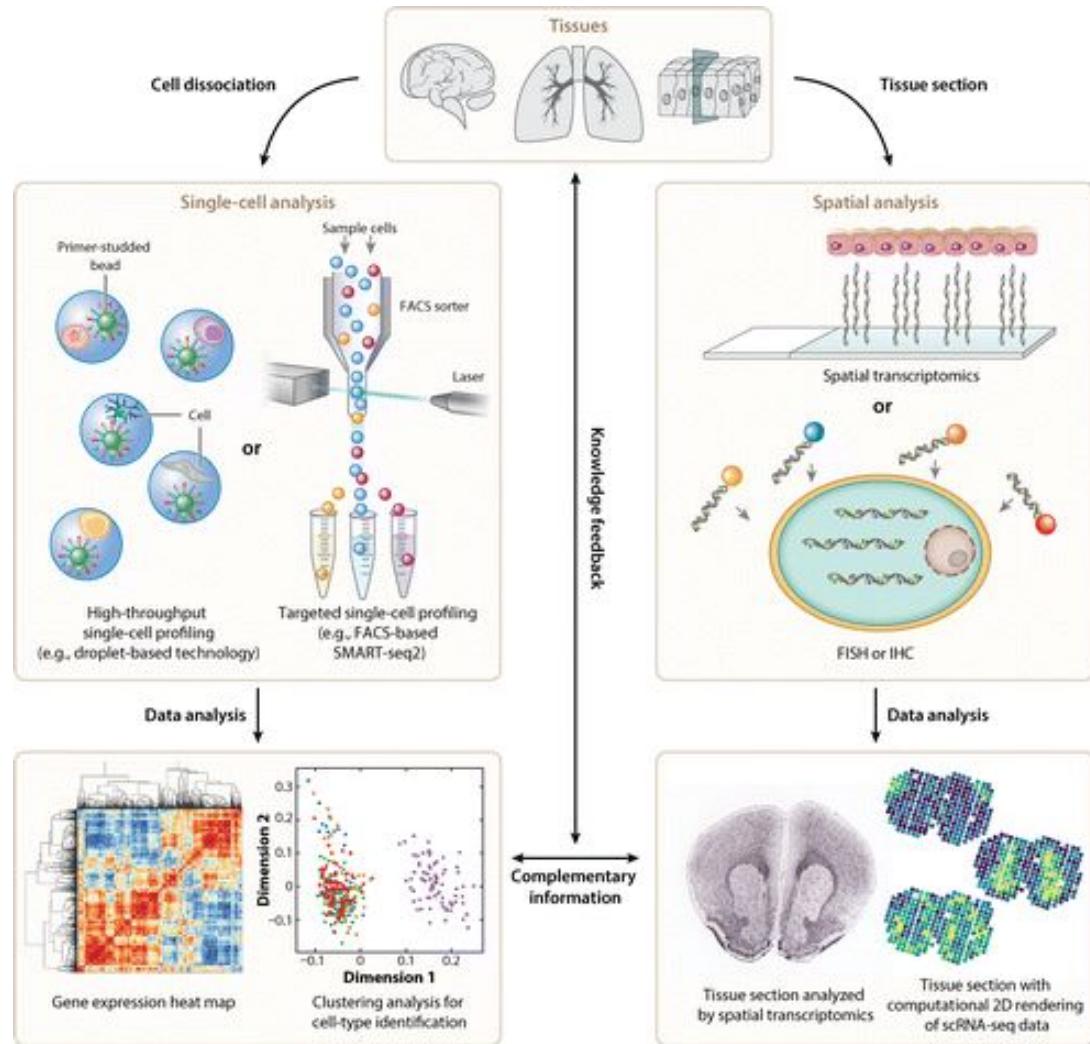


WARNING :

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

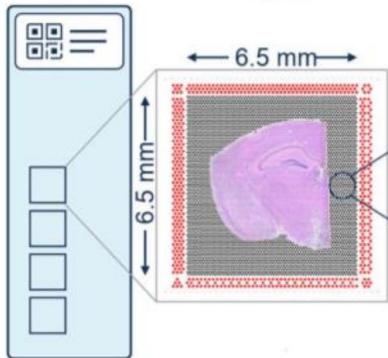
Spatial Single Cell RNAseq





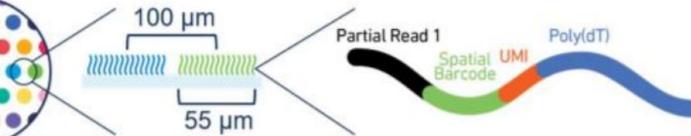
10x Genomics Visium

Visium Spatial
Gene Expression
Slide



Capture Area with
~5000 Barcoded
Spots

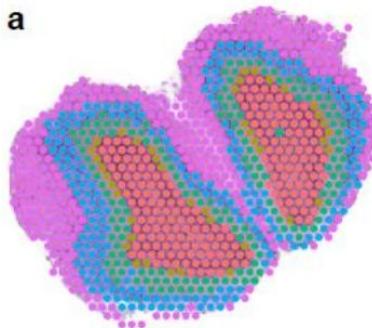
Visium Gene
Expression Barcoded
Spots



IHC

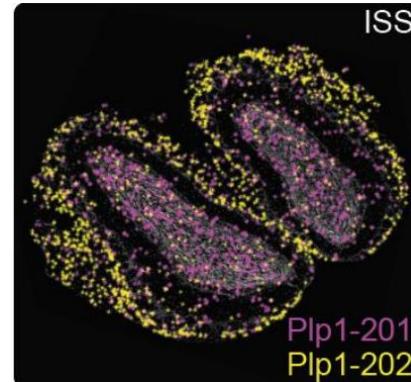


Clusters

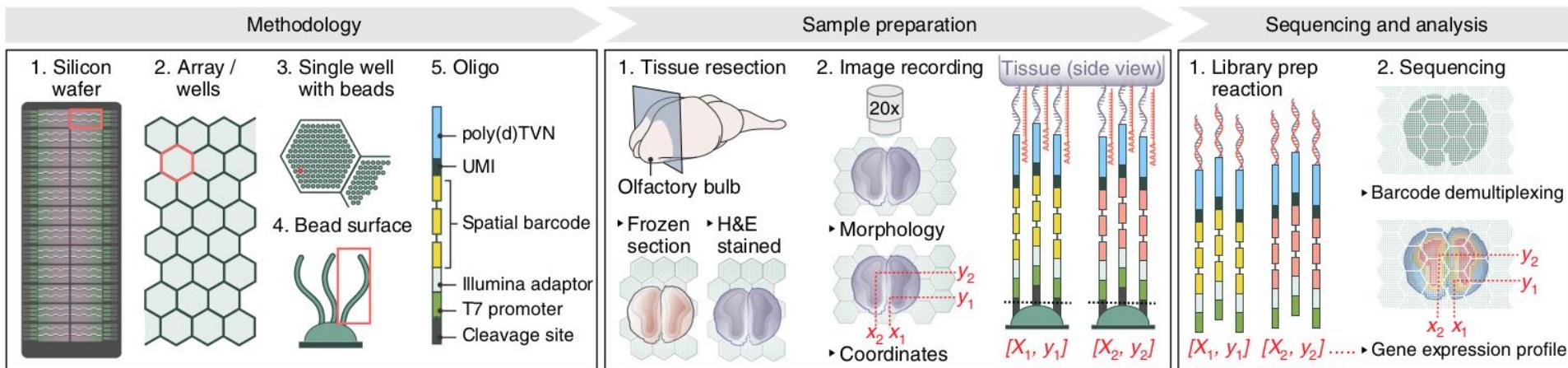


a

RNA FISH

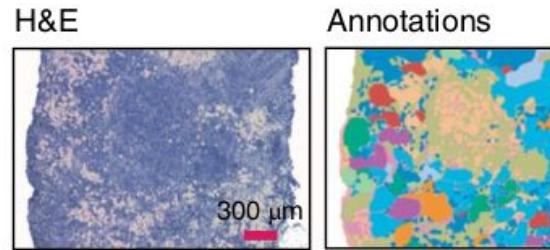
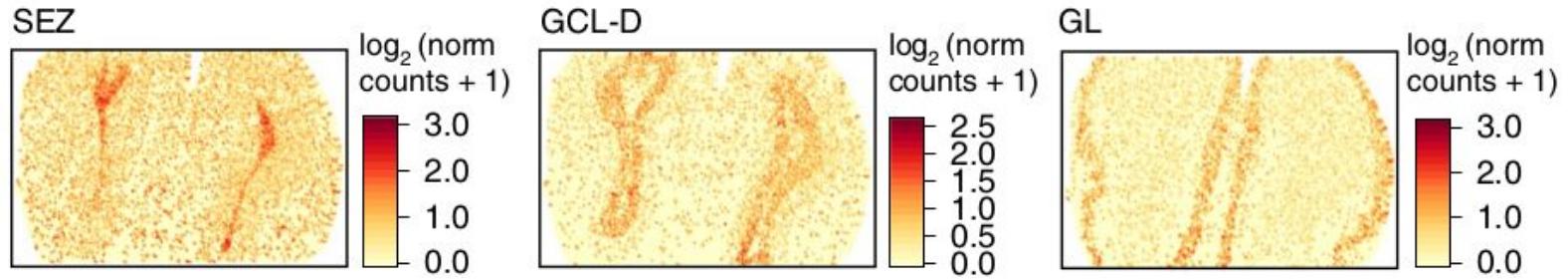


Illumina “HD Spatial Transcriptomics”

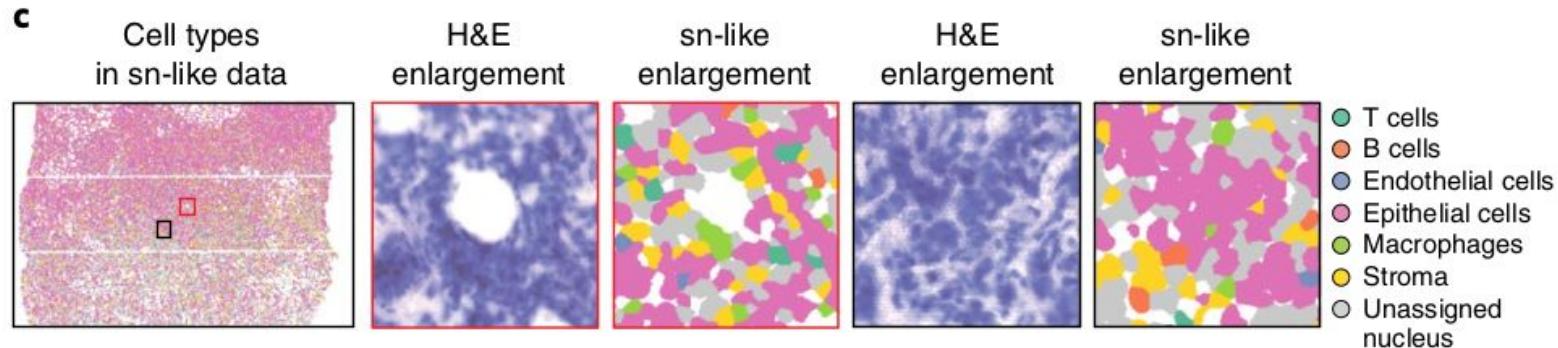


- 2,893,865 individual barcoded beads
- 1.4 M wells
- Well diameter ~ 2 μm
 - << median cell diameter (20 μm)
 - ~ 1,400 x higher resolution than “standard” ST
 - ~ 25 x compared to SLIDE-seq
- Array reading time ~ 3 H
- Challenging analysis strategy (low capture rate) ...
- Commercially available in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

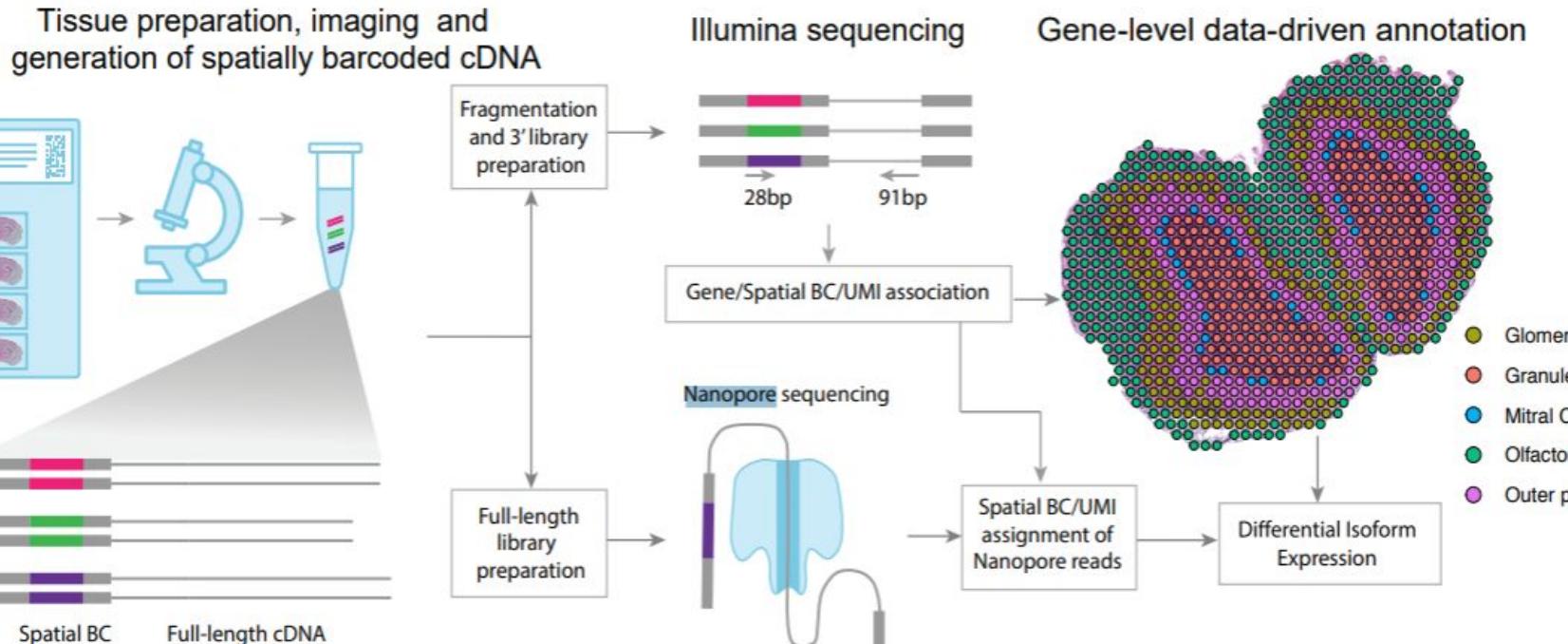
Illumina HDST



- 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



Spatial long reads



Single Cell CNV

SC
proteomics

Fluidigm cyTOF

SC
genomics

WES
CNV
VDJ/TCR
(WGA MDA/
MALBAC)

NOW, YOU
ARE HERE

SC
transcriptomics

mRNA

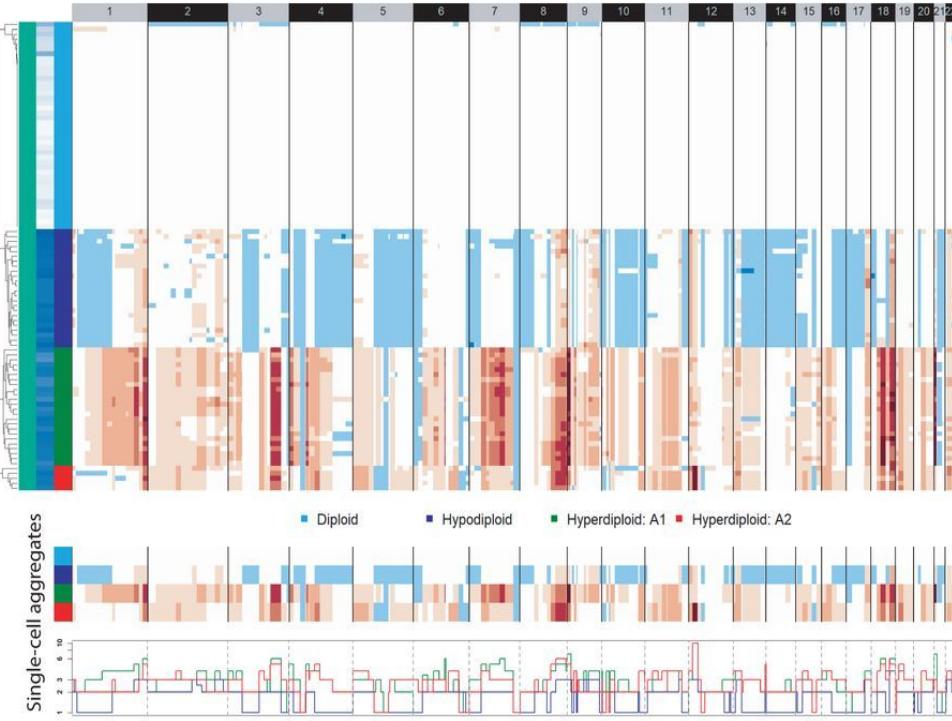
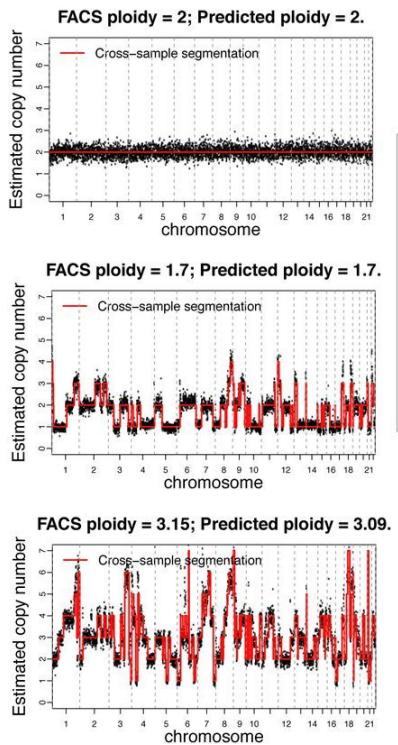
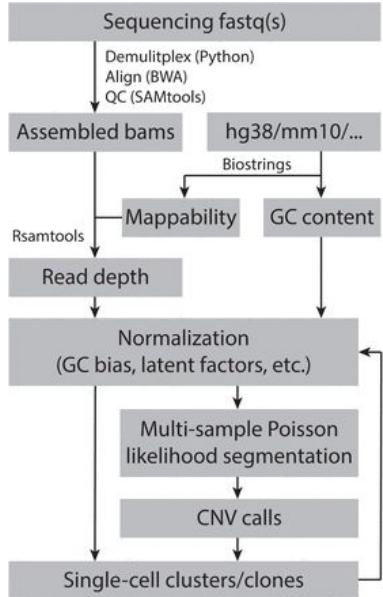
- 10X Genomics
- Fluidigm C1
- 1CellBio
- Seq-well
- DolomitesBIO

SC
epigenomics

SNAREseq

SPLINTseq

scCNV results (SCOPE)

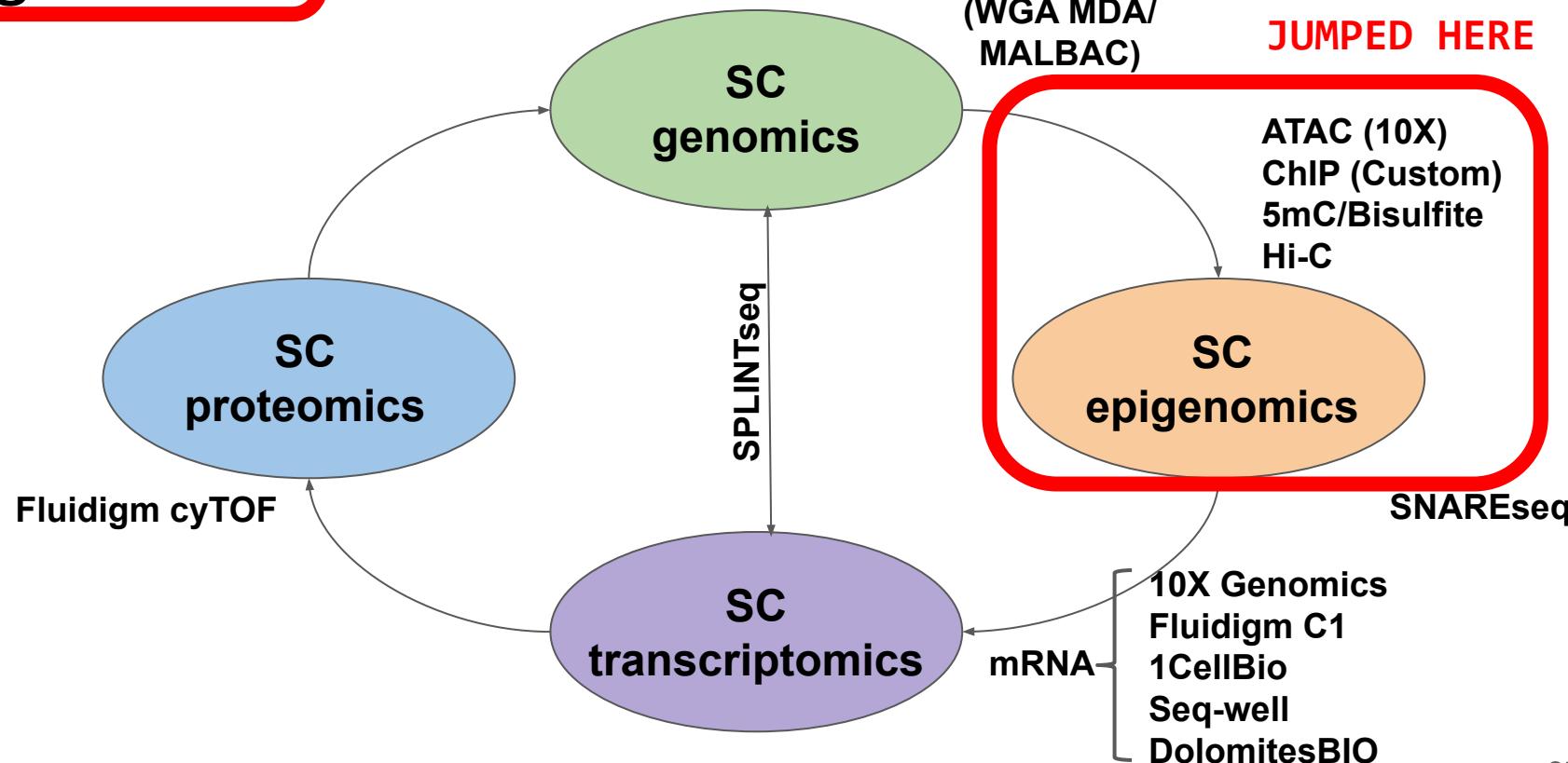


WARNING :

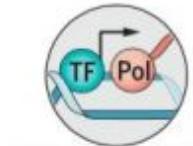
- Limited resolution : > 2 Mb (binning)
- Requires > 750,000 reads / cell

ALSO available : SCYN

Single Cell Epigenomics

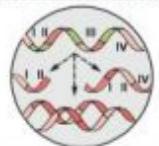


Overview of scEpigenomics techniques

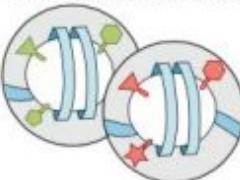


Transcription factor binding

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)



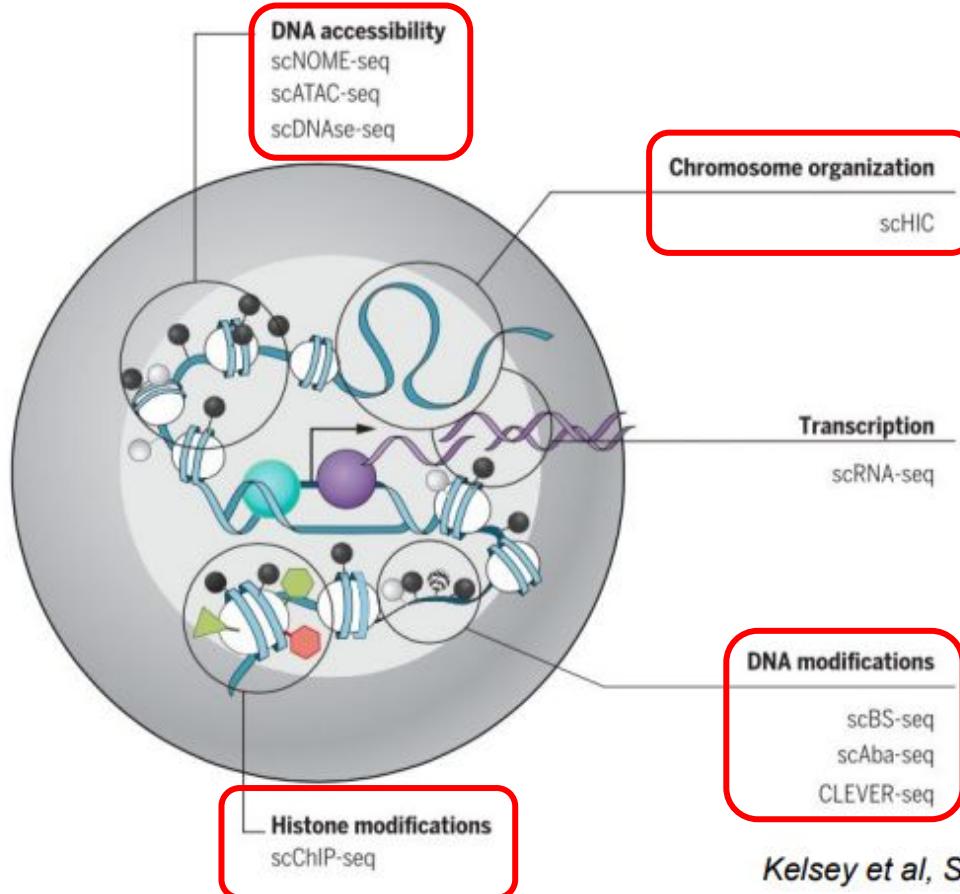
DNA modifications

- C
- 5mC
- 5hmC / 5fC / 5caC

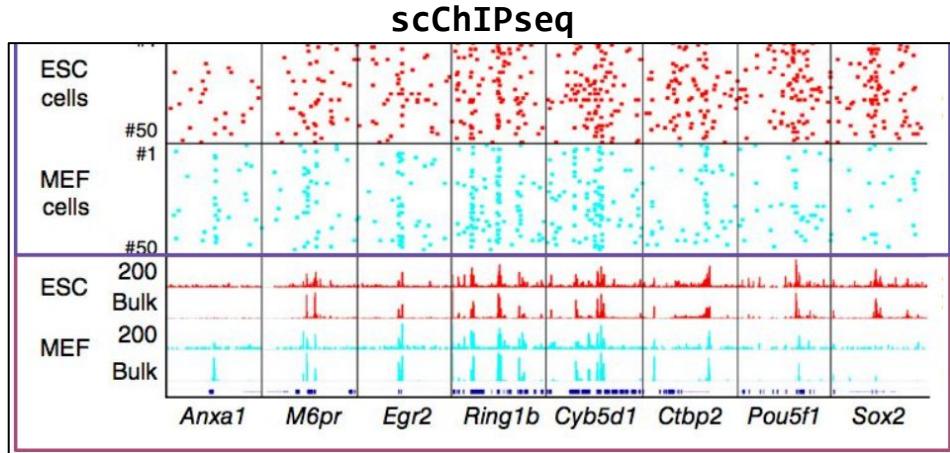
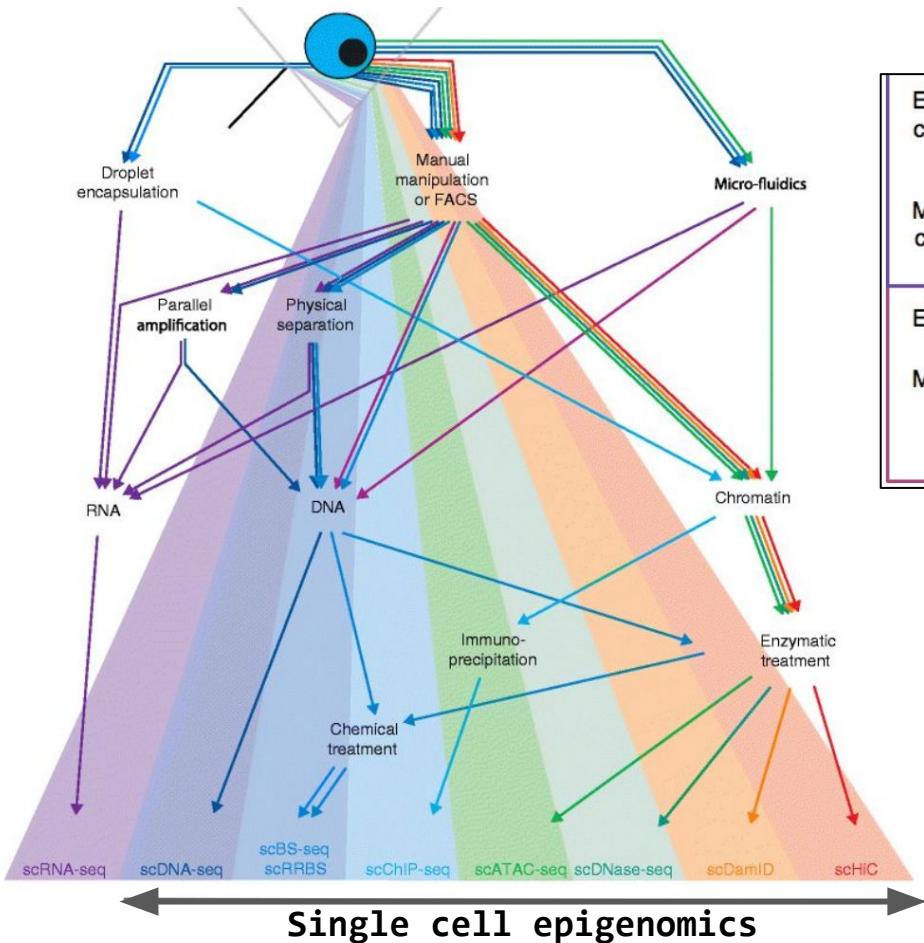


Chromosome organization

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>

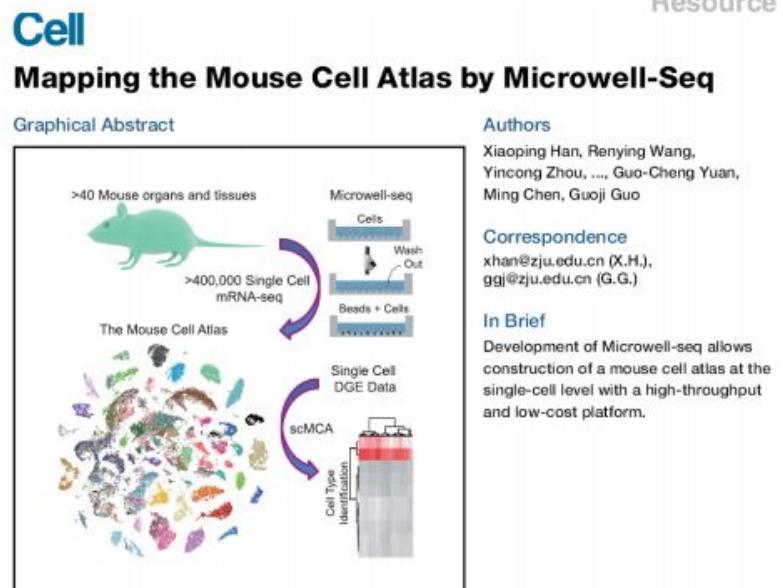
Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*

The Tabula Muris Consortium*

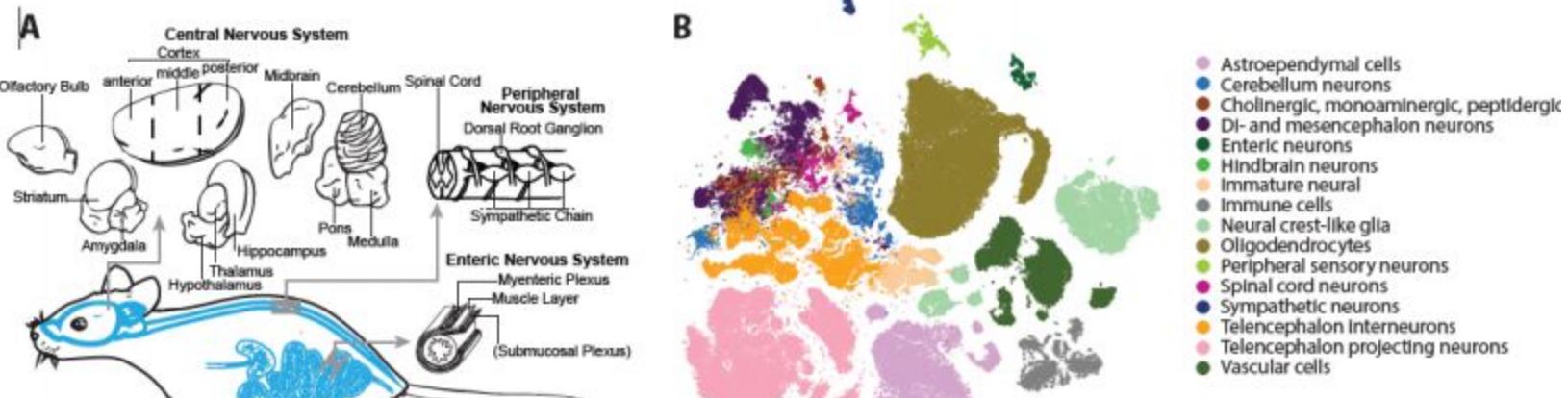
- ~100k cells
- 20 organs
- 2 techniques :
 - Droplet 3', short reads
 - FACS, long reads

MCA browser

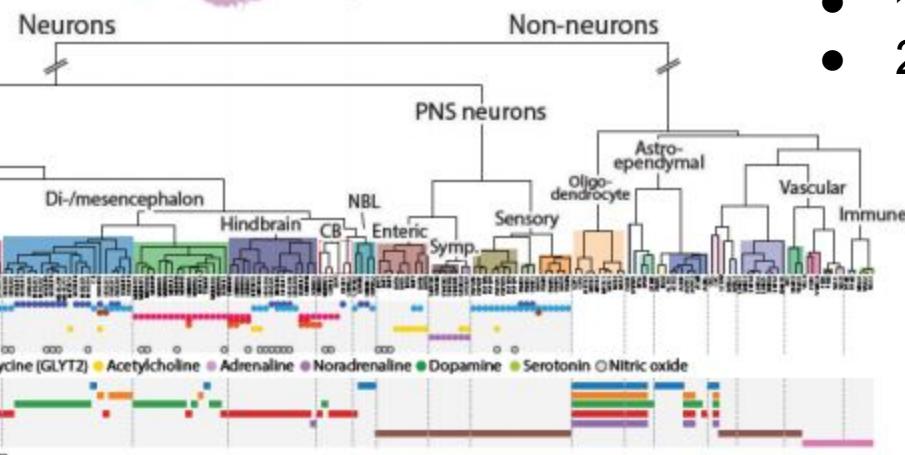
<http://bis.zju.edu.cn/MCA/>



The Mouse Brain Atlas (mousebrain.org)



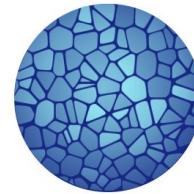
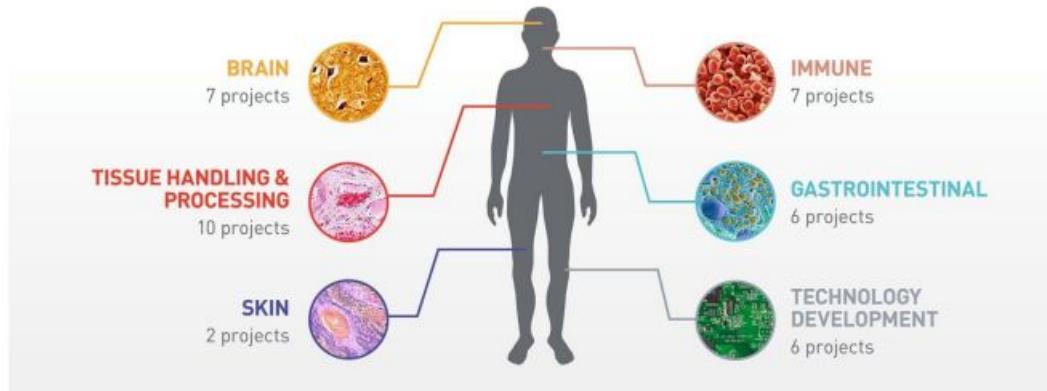
C



The Human Cell Atlas (humancellatlas.org)

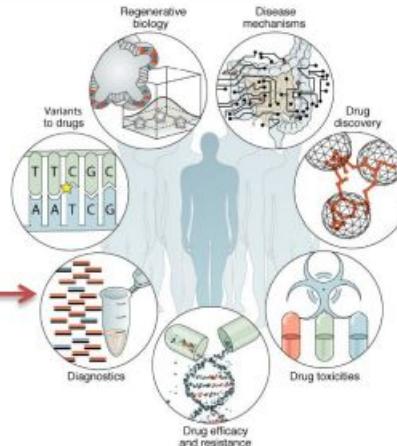
MAPPING THE BASIC UNITS OF LIFE

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

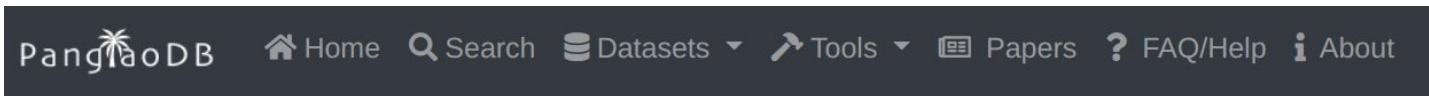


HUMAN
CELL
ATLAS

- 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

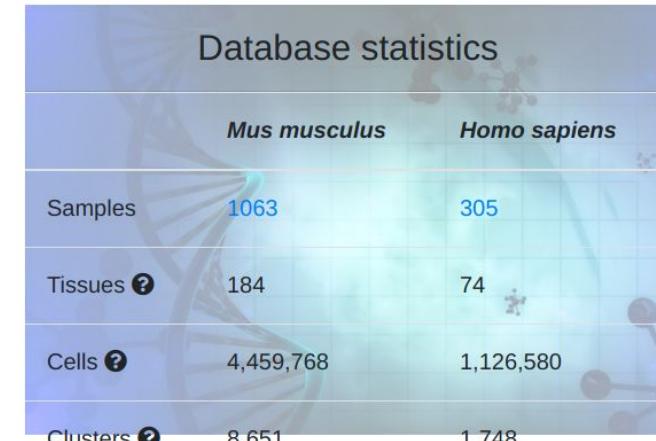


The navigation bar at the top of the page includes the PanglaoDB logo, Home, Search, Datasets, Tools, Papers, FAQ/Help, and About links.

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



A table comparing database statistics for *Mus musculus* and *Homo sapiens*.

	<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

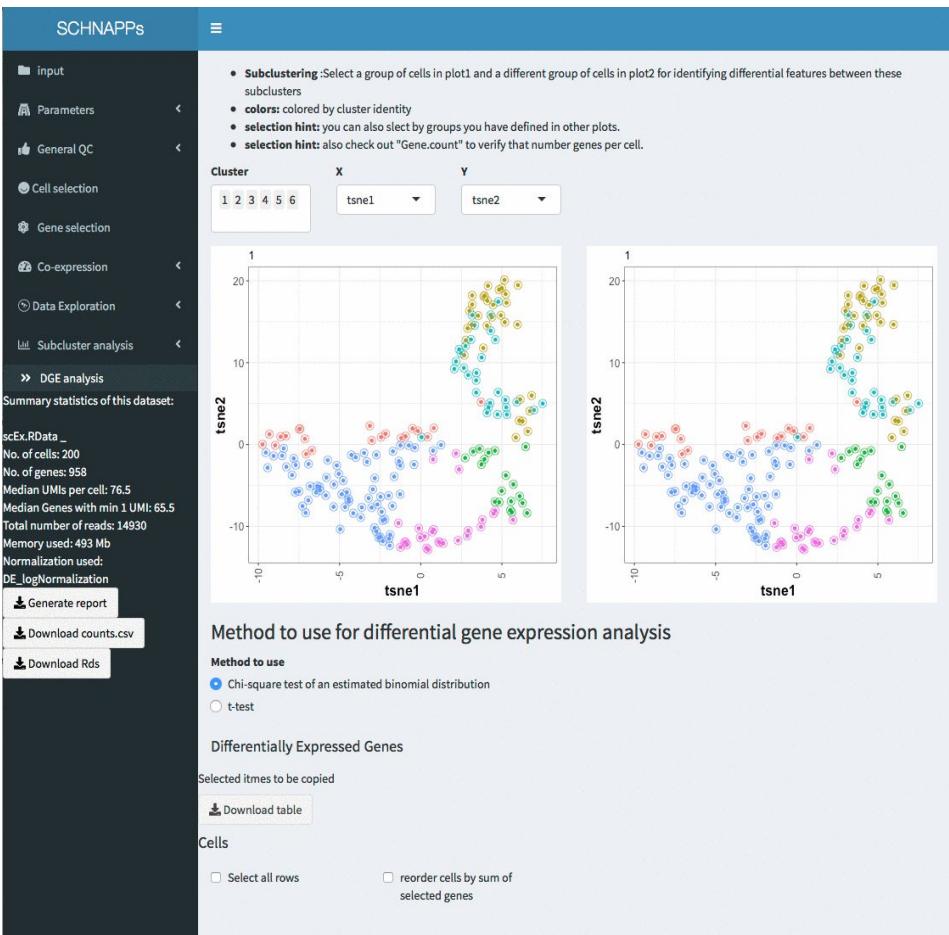
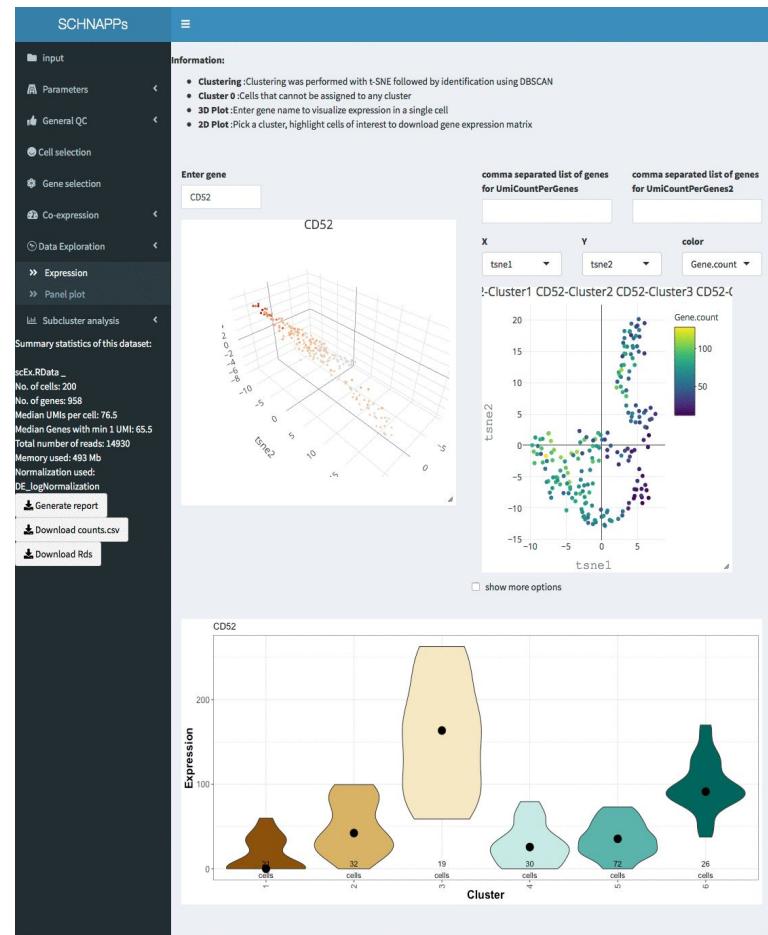
Dataset of the day

Take a closer look at the cellular composition of [Heart](#), using a dataset which consists of 871 cells. Clustering of this dataset resulted in 7 cell clusters, containing among others, [Smooth](#)

WYSIWYG analysis frameworks
(mainly for scRNAseq / scATACseq)



SCHNAPPS : A R-shiny app for biologists



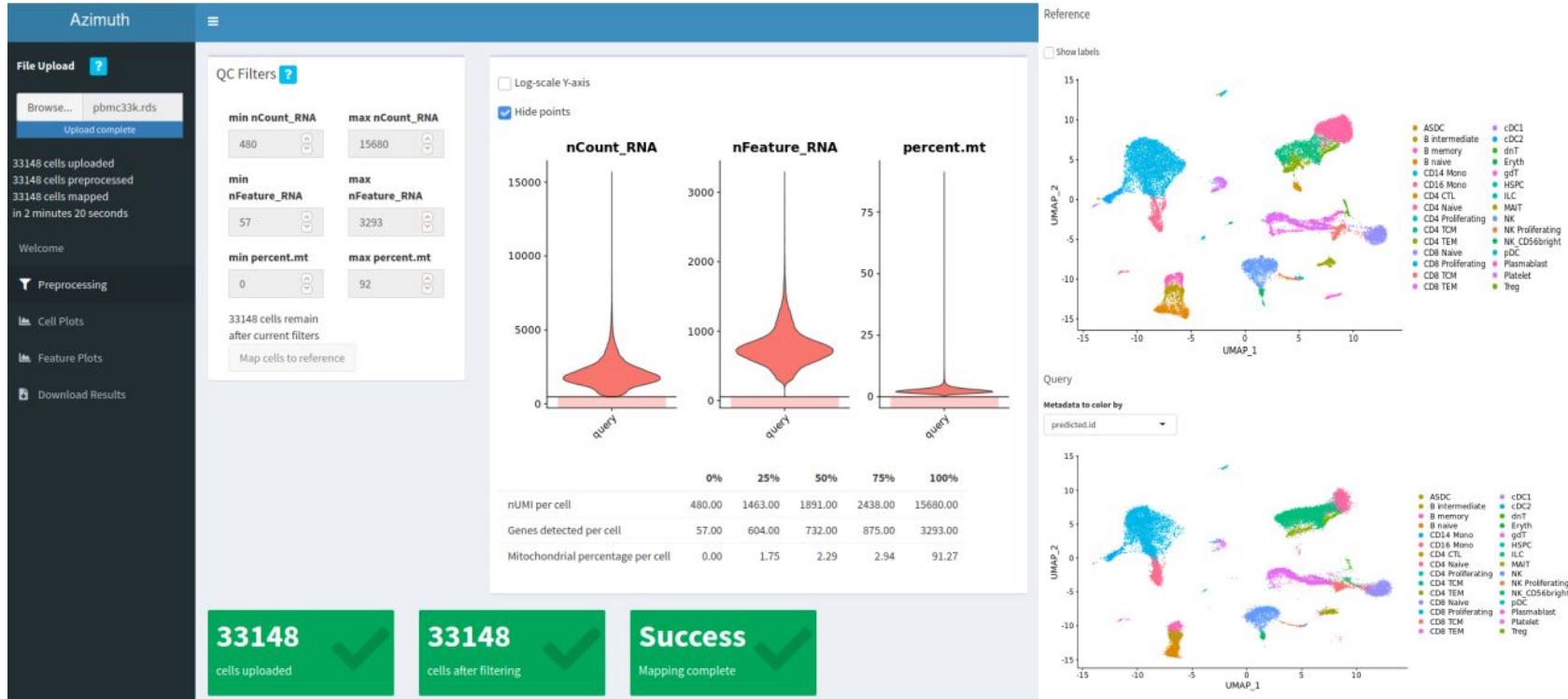
SeuratV3Wizard

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 name "nFeature_RNA" is selected in the dropdown.
- 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: A scatter plot showing clusters of cells in UMAP space (UMAP_1 vs UMAP_2). The plot includes a color key from 0 to 9.

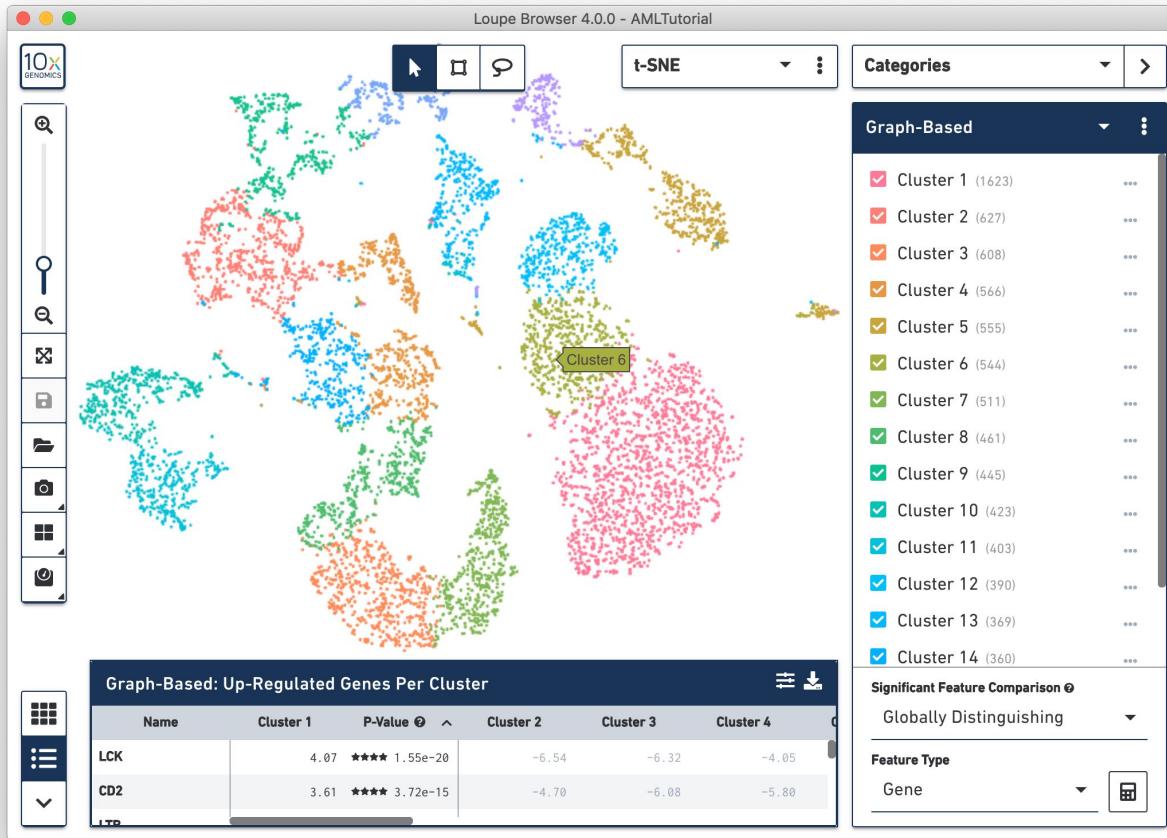
<http://nasqar.abudhabi.nyu.edu/SeuratV3Wizard/>

Azimuth (from Satija Lab)

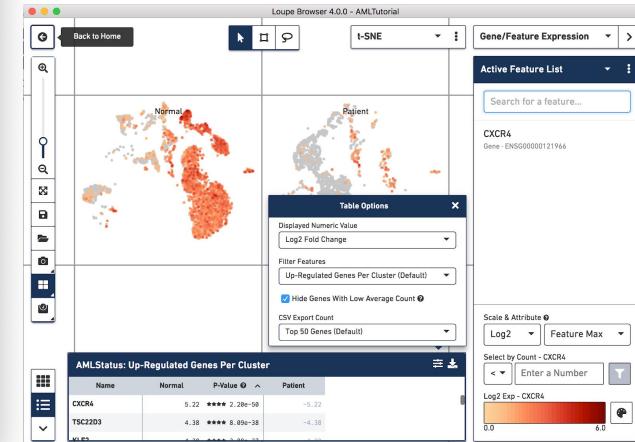


Visualization tools
(mainly for scRNAseq)

10x Genomics Loupe Browser

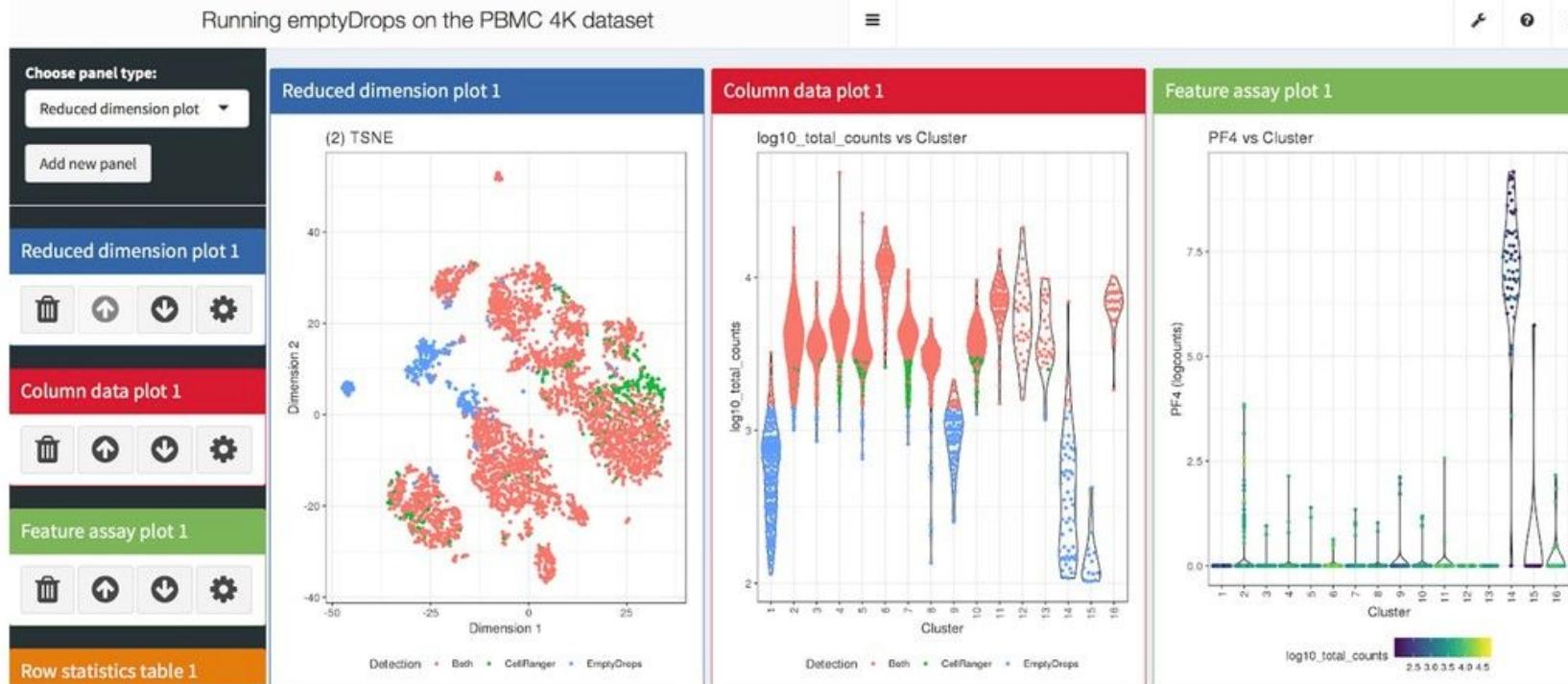


- Compatible with output from 10x Cell Ranger (“cloupe” files)
- Linux / OSX
- Supports Visium (Spatial)



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)

UCSC Cell Browser



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

BREAKING NEWS !

Announced by 10x on twitter in January

New High Throughput Instrument

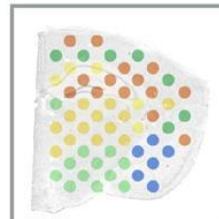


Chromium X
Coming 2H 2021

Making 1 million cell experiments routine



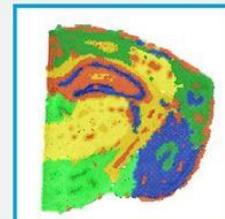
Visium at Single Cell Resolution



First Generation



Visium
Resolution 4x



Visium HD
Resolution 1,500x

Unlocking More Samples for Spatial



Visium CytAssist
Coming 1H 2022

Accessing archived and pre-mounted samples

Acknowledgements

Marc Deloger

Morgane Thomas-Chollier

Agnès Paquet

Marine Aglave

Antonio Rausell

Wouter Saelens

Nathalie Gaspar

... and you !



SINGle-cellING in the RAINaseq (1952)
A joke © Jacques van Helden

APPENDIX

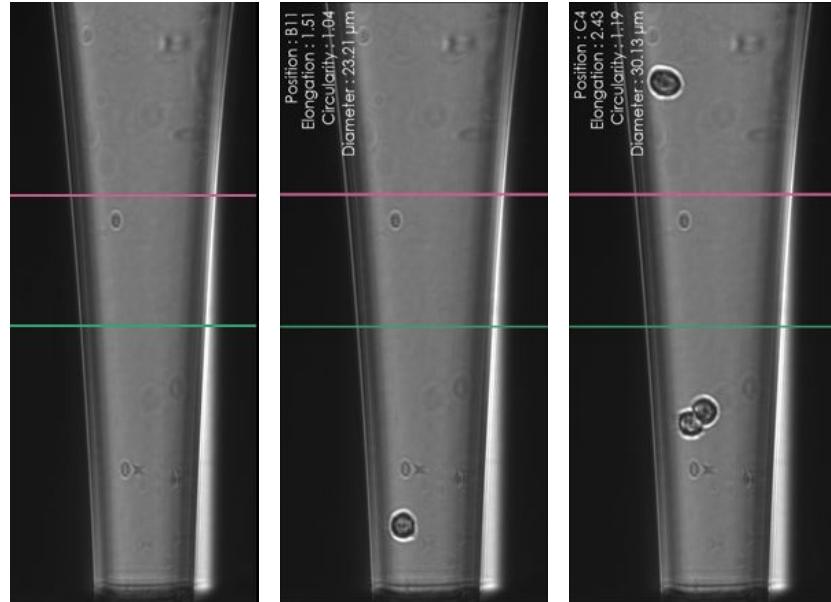
Some things I muted (*and many, many more*)

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

Alternative isolation method :

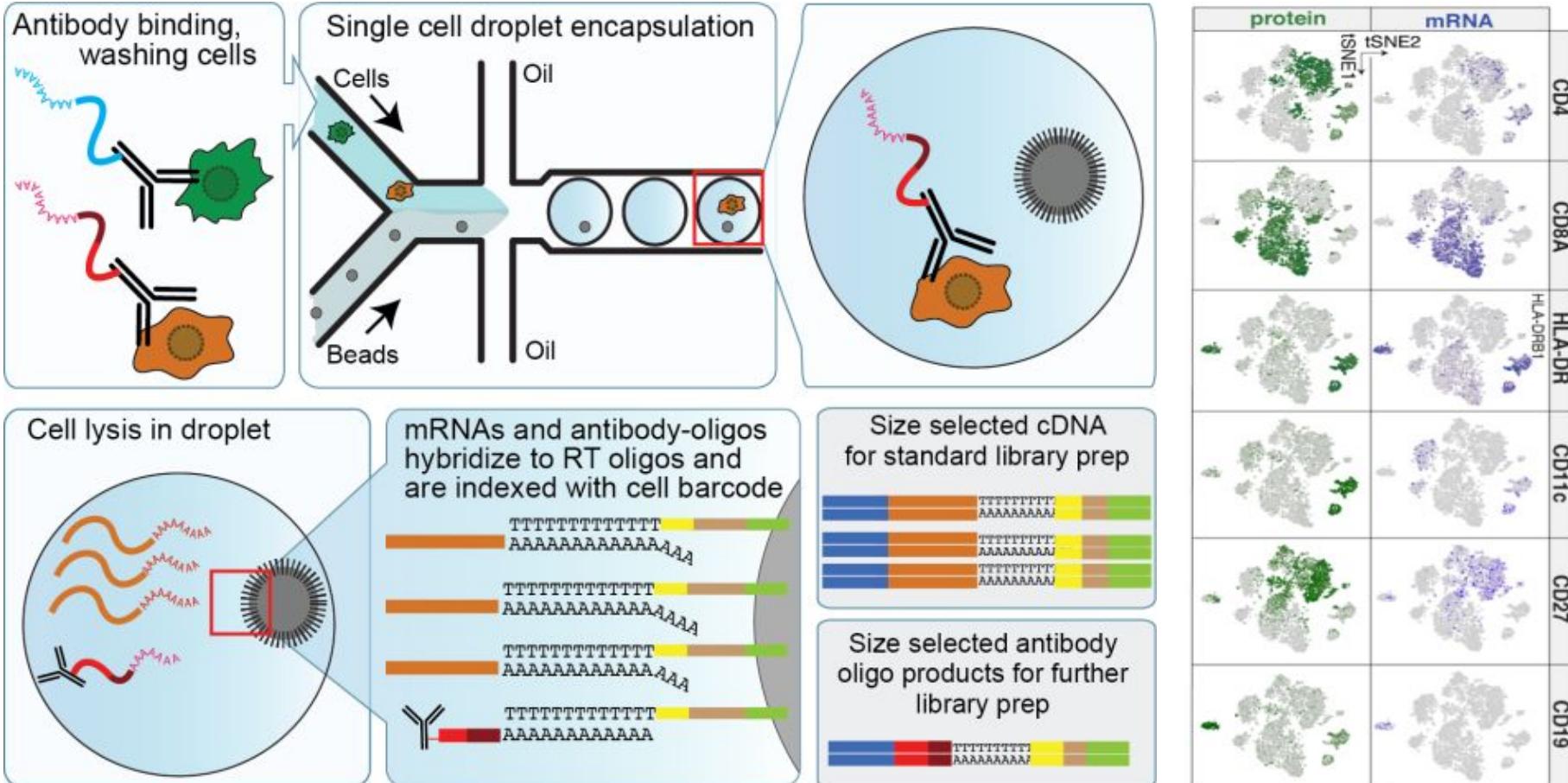
IBSCI™ (Image Based Single Cell Isolation)

- 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 1536 cells
- Cell recovery rate over 95%
- Open platform
 - Scalable, compatible
 - Custom reaction kits



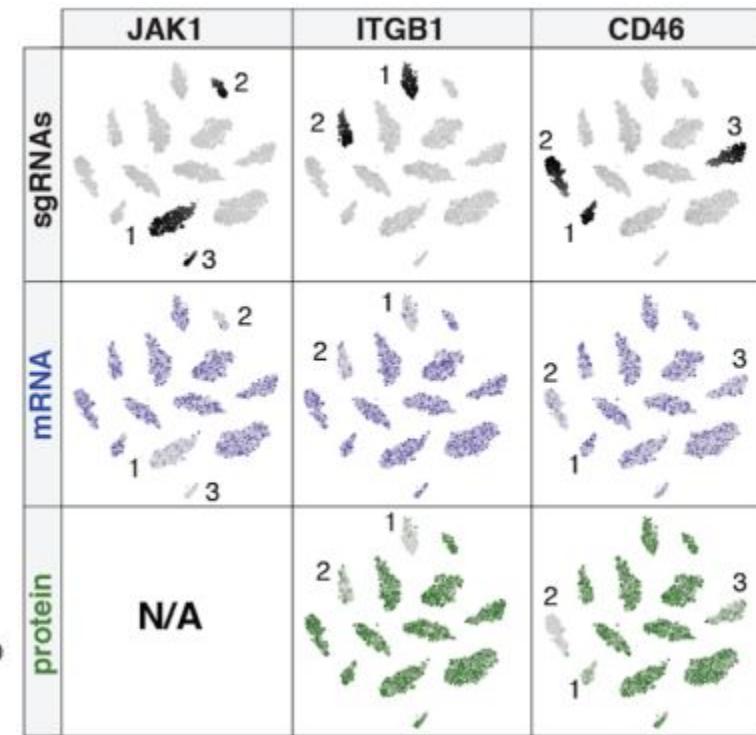
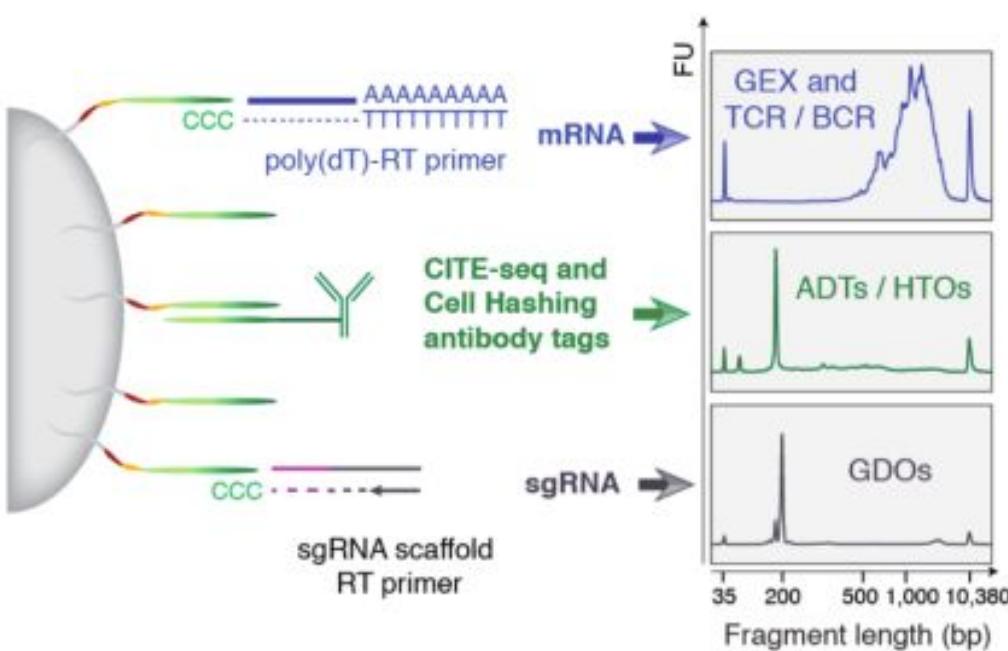
CITE-seq
(scRNAseq + proteins)

Cellular Indexing of Transcriptomes and Epitopes by Sequencing



ECCITE-seq
(scRNAseq + proteins + CRISPR gRNA)

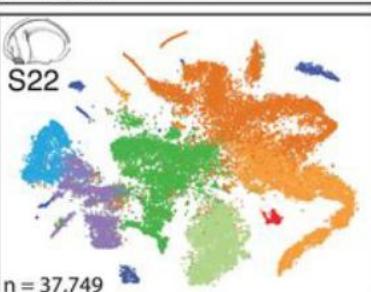
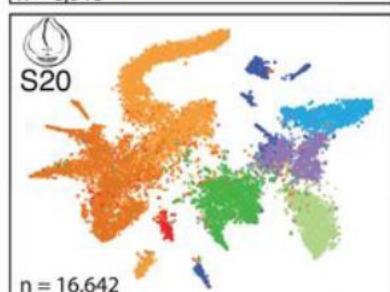
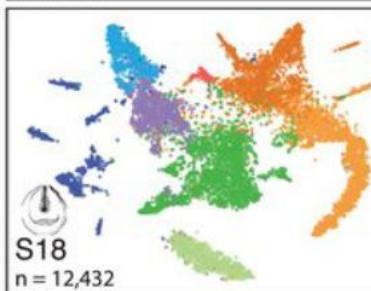
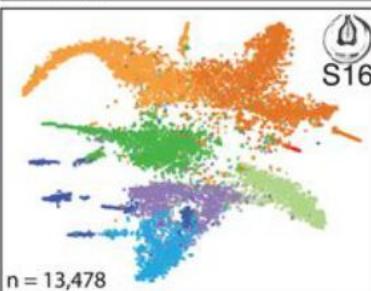
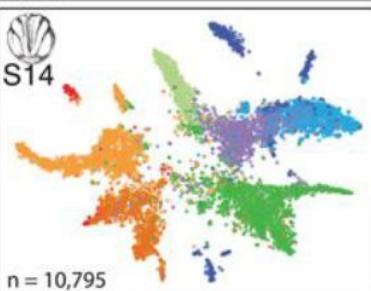
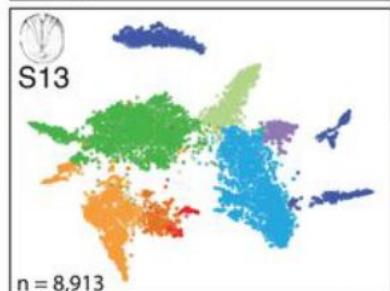
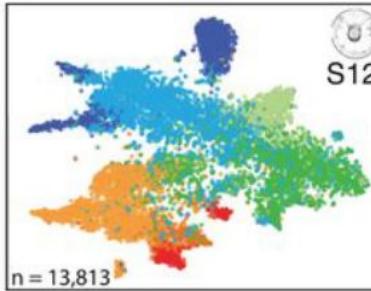
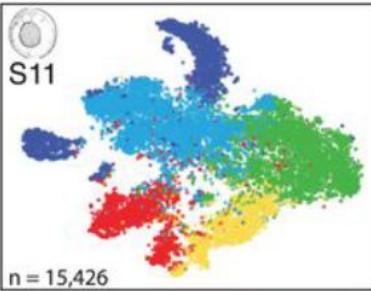
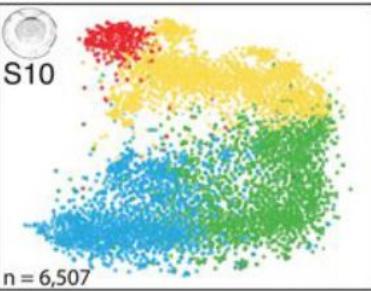
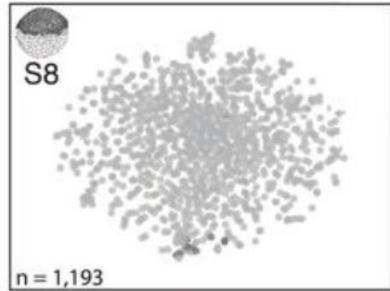
Extended CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes by Sequencing (5')



Some sweets

Xenopus embryo development

tSNE dimension 2

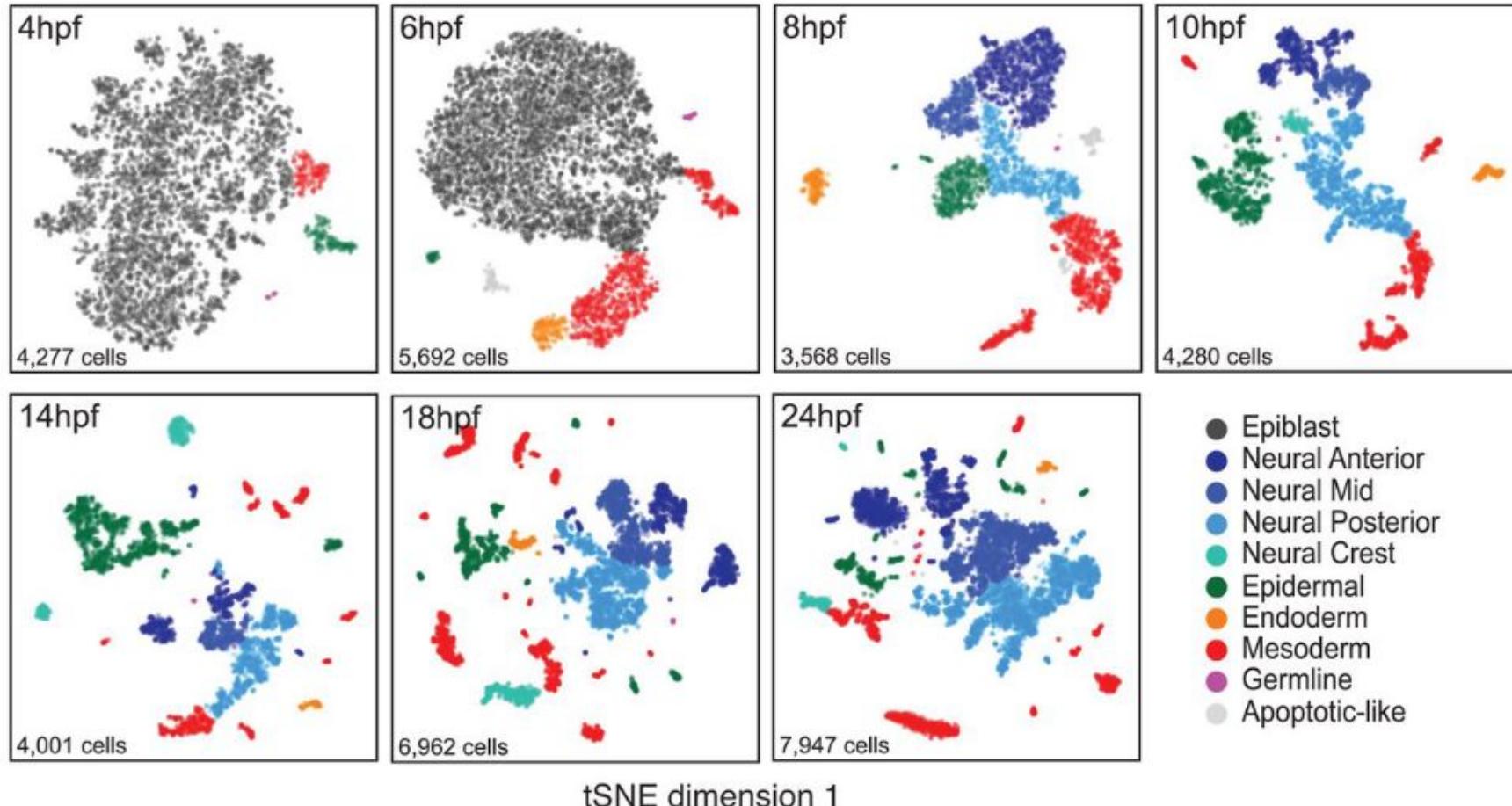


- Pluripotent blastula
- Germline
- Non-neural ectoderm
- Placodal
- Specialized epidermis
- Neural
- Neural crest
- Marginal zone
- Dorsal mesoderm
- Vent/lat/int. mesoderm
- Endoderm

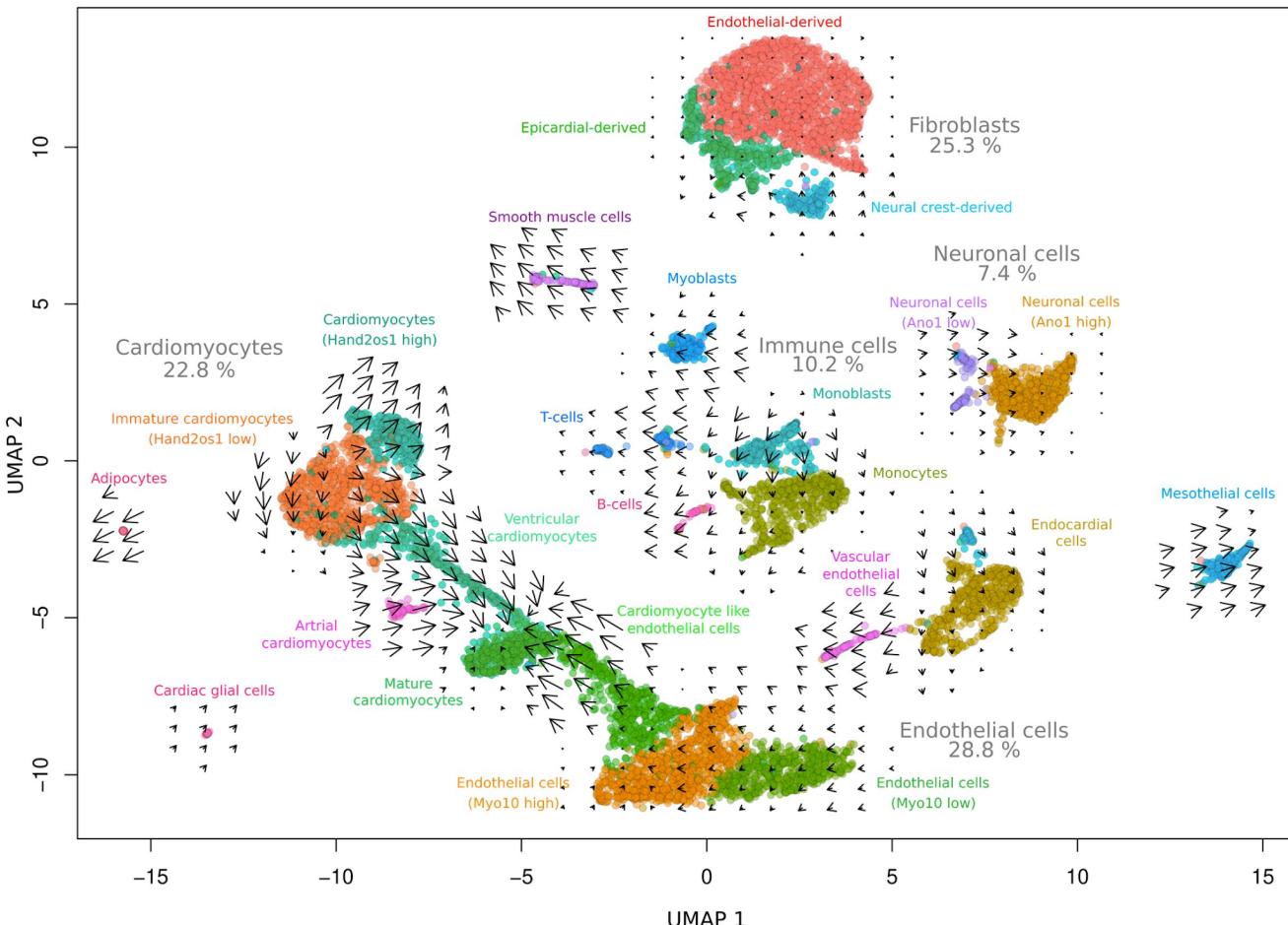
tSNE dimension 1

Zebrafish embryo development

tSNE dimension 2

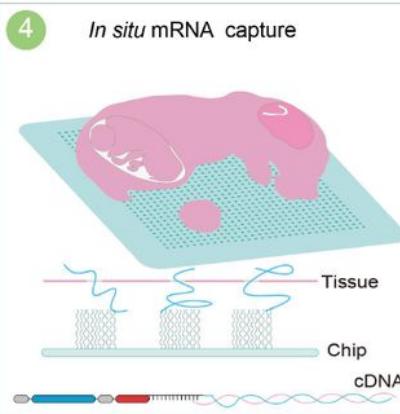
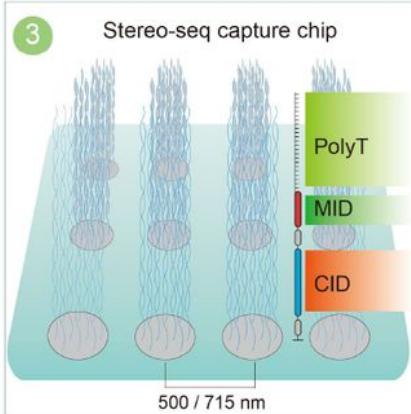
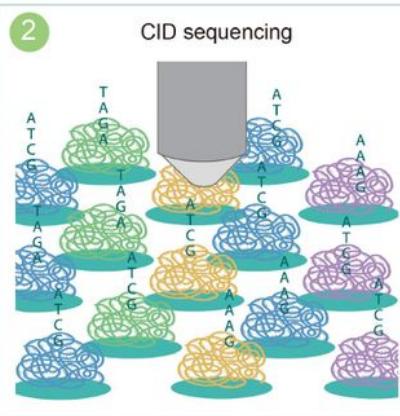
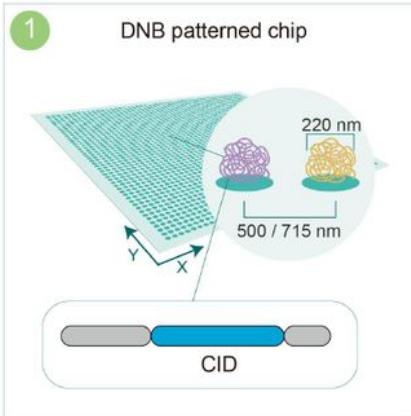


Entire mouse heart : expression & velocity



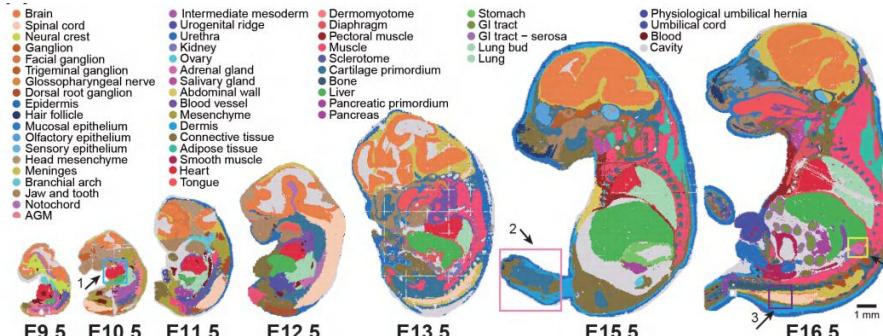
<https://doi.org/10.3390/cells9020318>

“StereoSeq” : Ultra HD spatial long reads

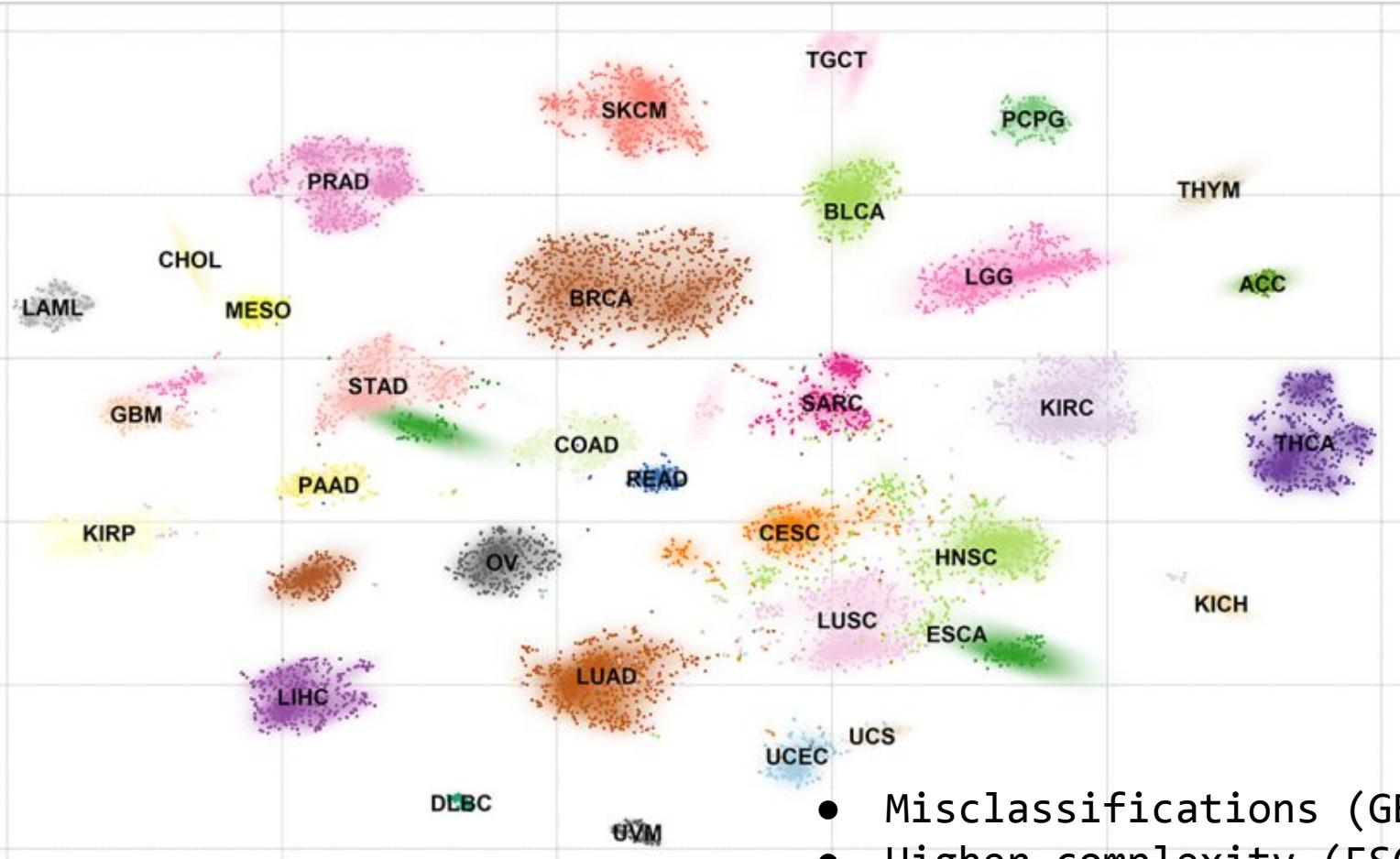


- Based upon “DNB” (DNA Balls) arrays by MGI/BGI
- Claimed resolution of 500-715 nm
- Active surface of 200 mm²
- Performed a developmental analysis of FULL mice embryos !

Cheng et al, BiorXiv 2021.01

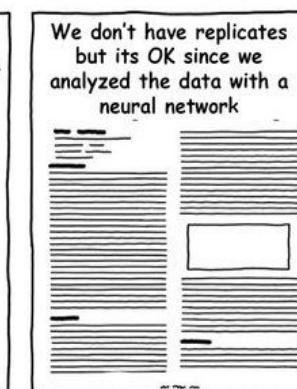
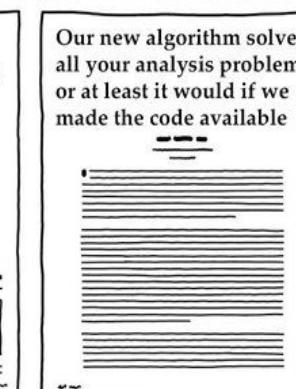
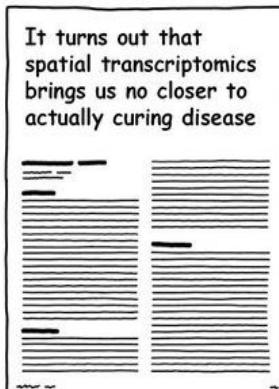
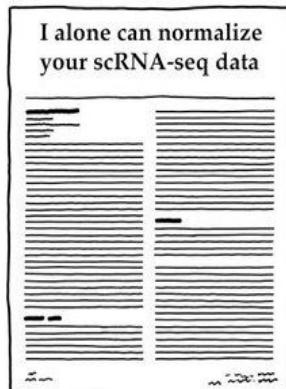
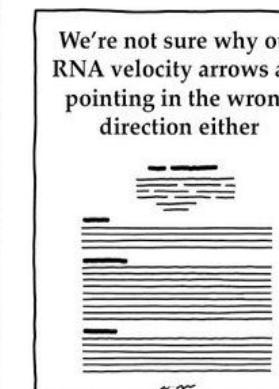
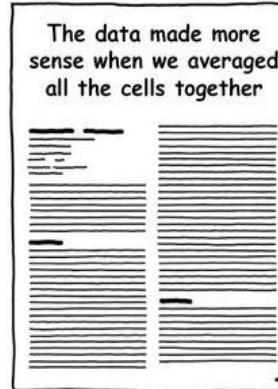
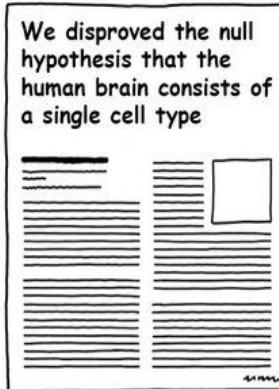
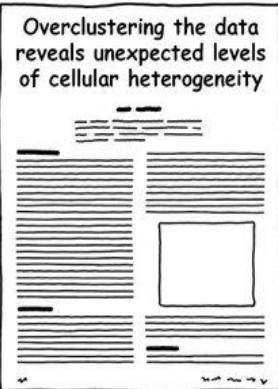


t-SNE of the whole TCGA project (*not SC*)



Single cell results and the community

TYPES OF SINGLE-CELL SEQUENCING PAPER



tSNE / uMAP plots *are art* !

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