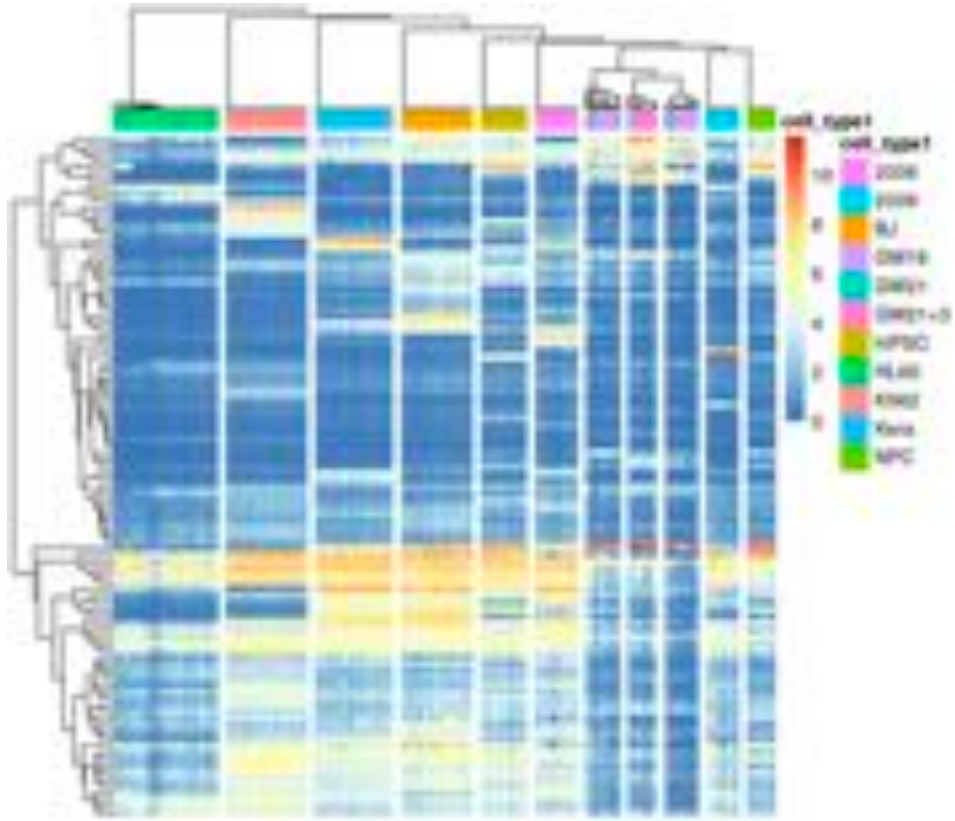
Analysis of Single-Cell RNA-Seq Data: Experimental Design

Orr Ashenberg, Jacques Serizay

June 2021

Physalia

Courses



Overall goals

Introduction to the rapidly expanding world of single-cell transcriptomics

Focus less on specific software tools but more on underlying concepts - so down the line, you can make informed choices

Hands on lab exercises analyzing single-cell heterogeneity

Create a fun, learning, collaborative, and interactive environment over the next week

A few organizational notes

https://github.com/js2264/scRNAseq_Physalia_2021

Write course notes and questions in a shared Google document

Raise hand in Zoom (Participants) to ask questions or use chat

Please use video and mute microphone when not in use

Please be patient with technical issues (network, Zoom, etc…)

Mouse organogenesis studied by single-cell RNA sequencing

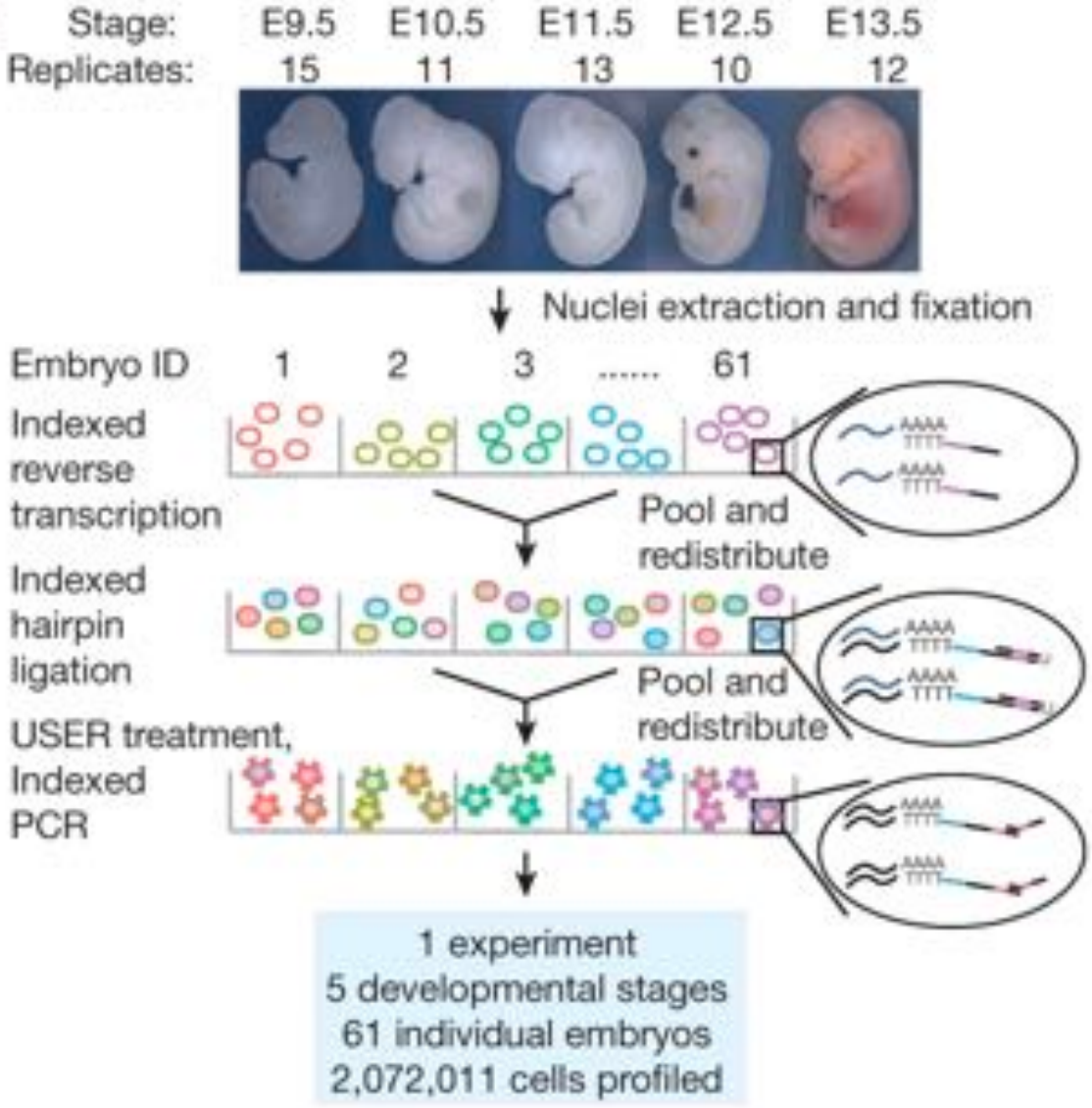
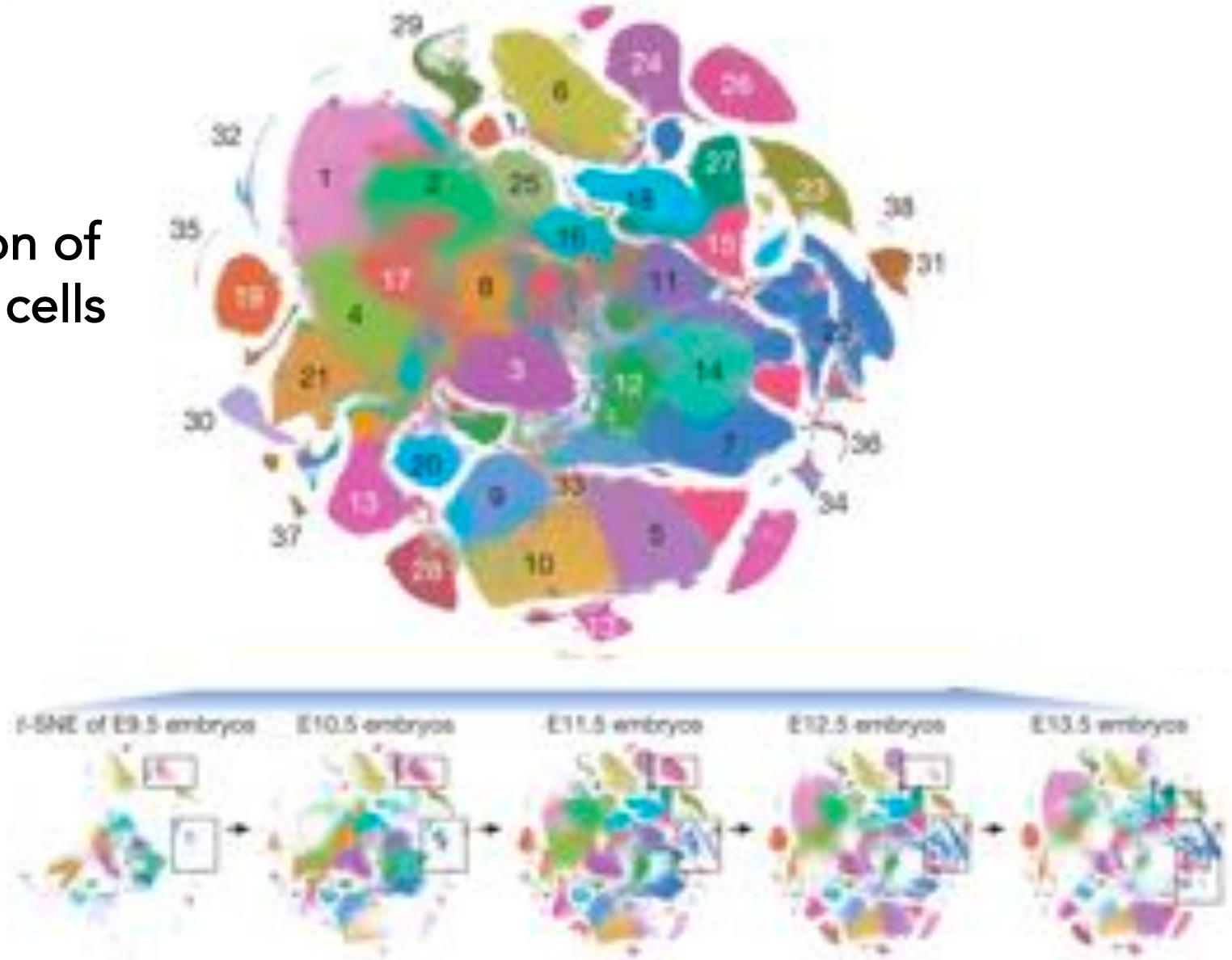


Fig. 1: sci-RNA-seq3 enables profiling of 2,072,011 cells from 61 mouse embryos across 5 developmental stages in a single experiment.

Cao, J., at al. *Nature* 2019

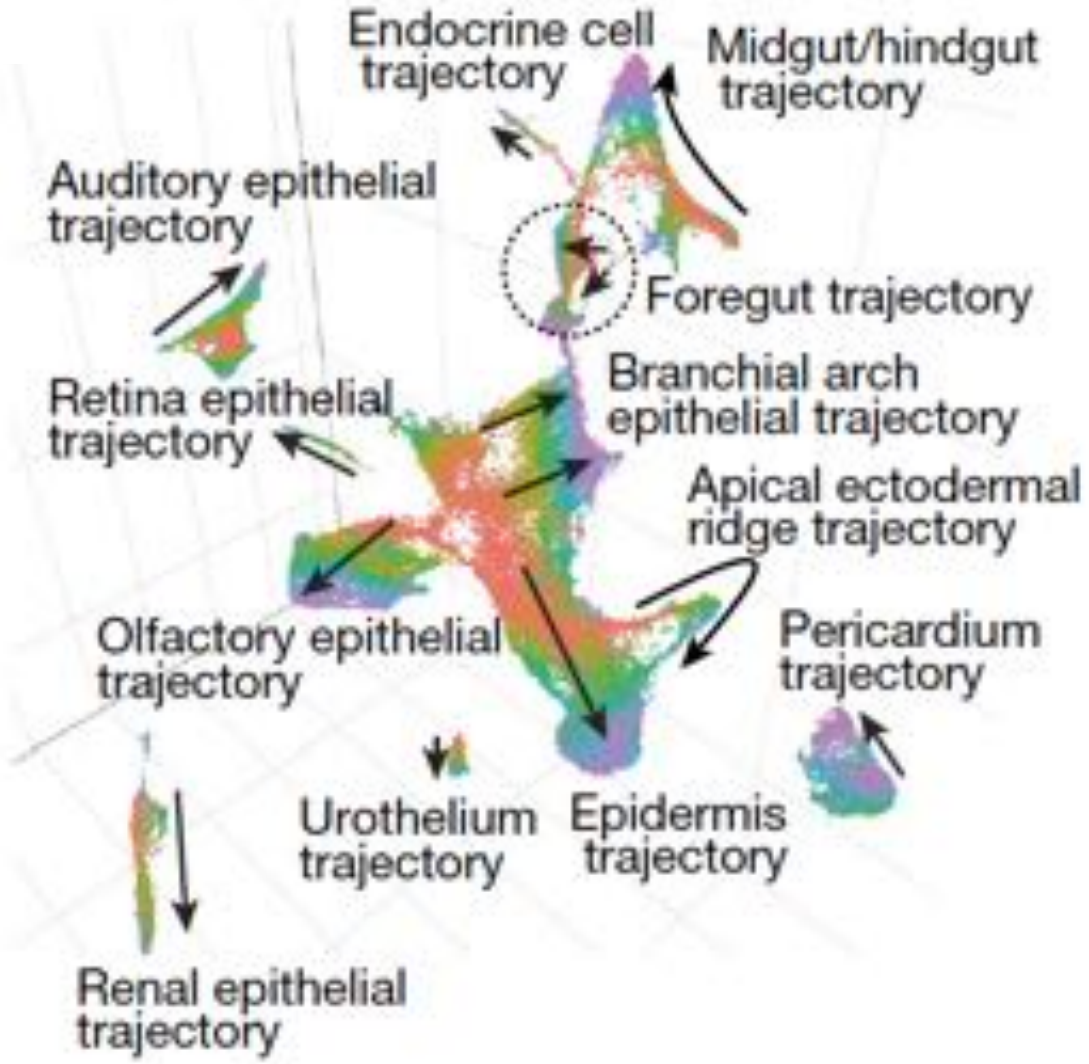
Mouse organogenesis studied by single-cell RNA sequencing

Clustering and visualization of 2,026,641 mouse embryo cells



Cao, J., at al. *Nature* 2019

Mouse organogenesis studied by single-cell RNA sequencing

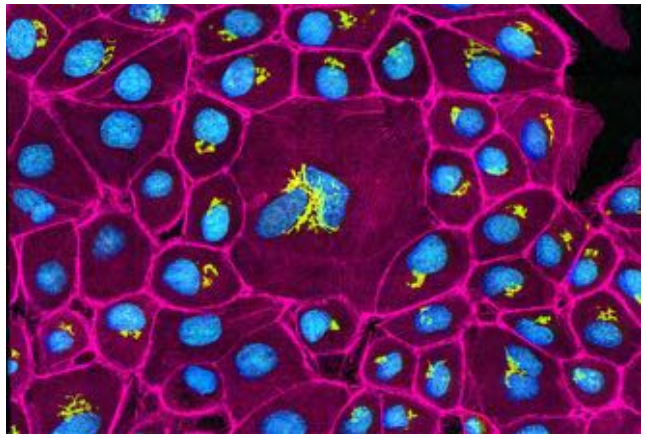


Cao, J., at al. *Nature* 2019

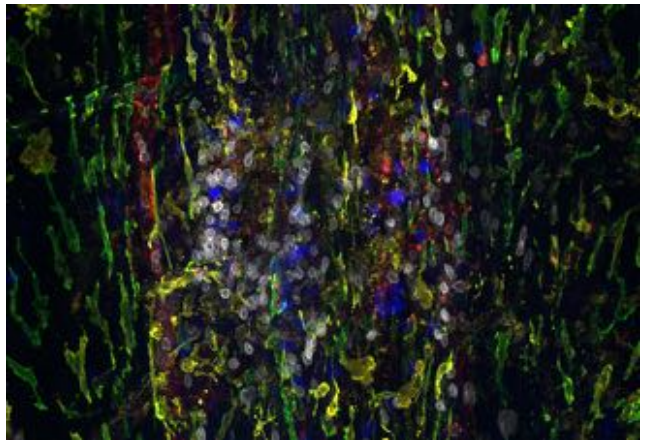
https://www.youtube.com/watch?v=cWG2CkkDcrM

Incredible diversity in cell types, states, and interactions across human tissues

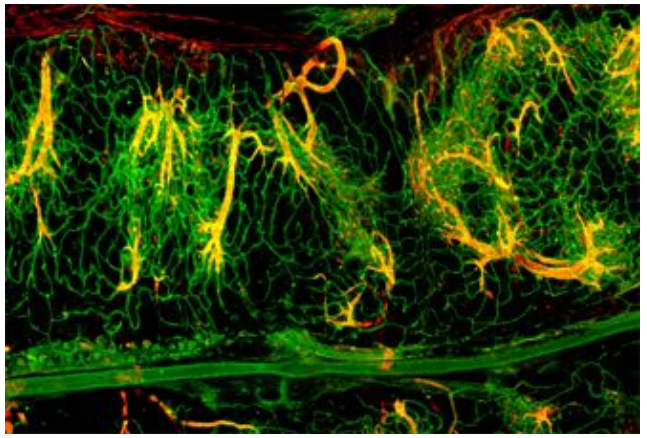
Skin epithelium



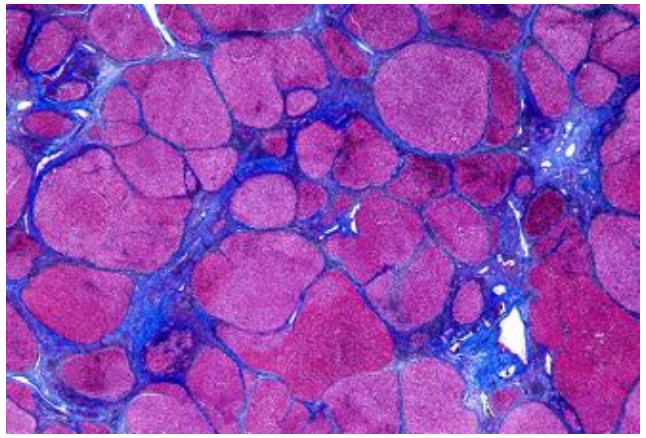
Brain meninges



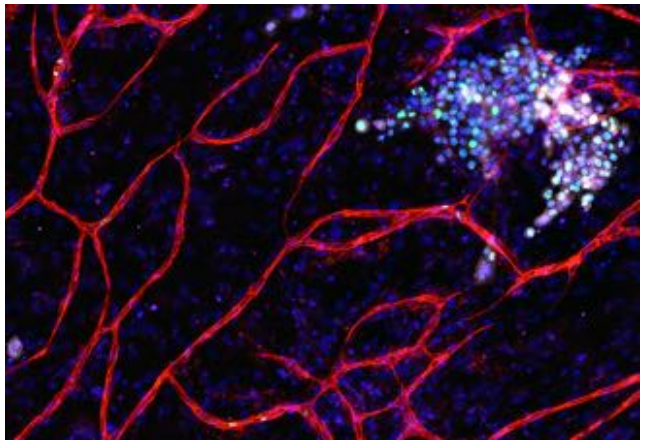
Blood vessels



Small intestine



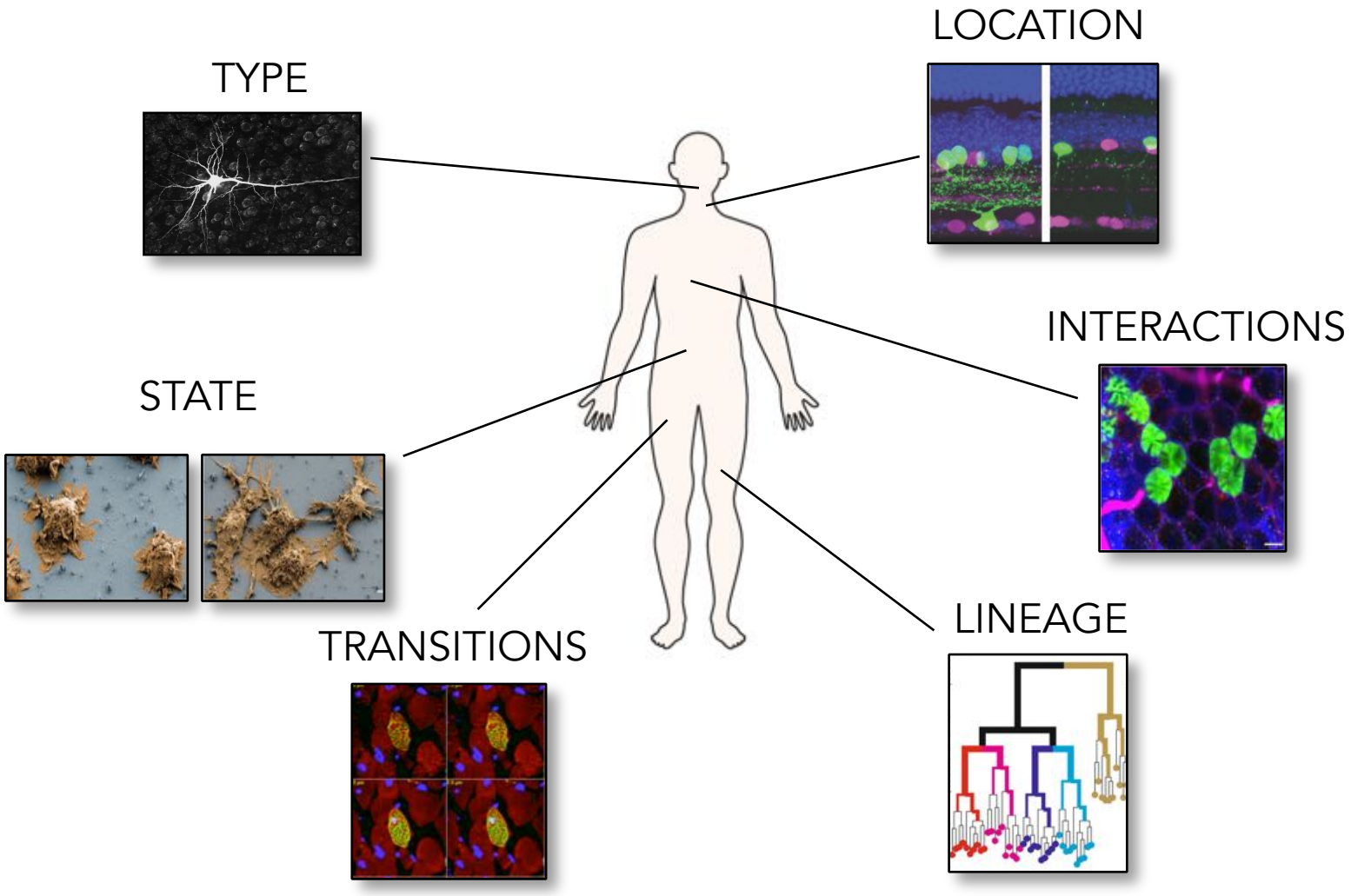
Liver cirrhosis



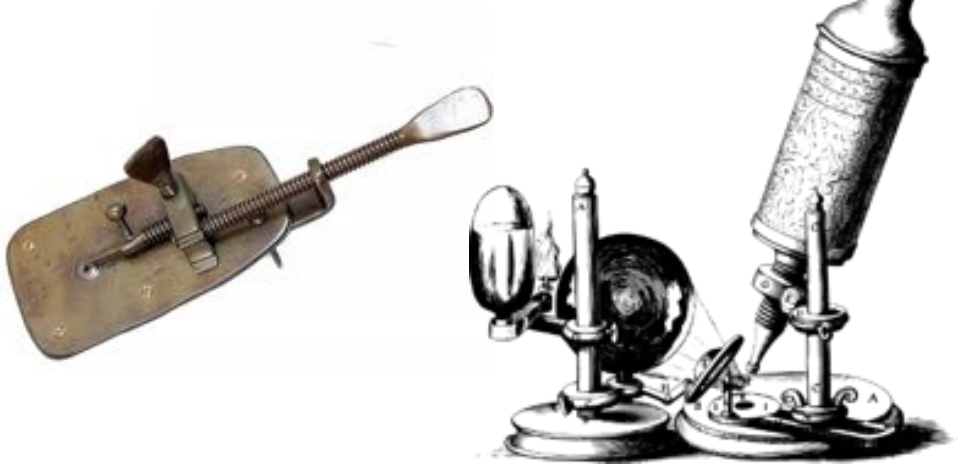
Breast cancer

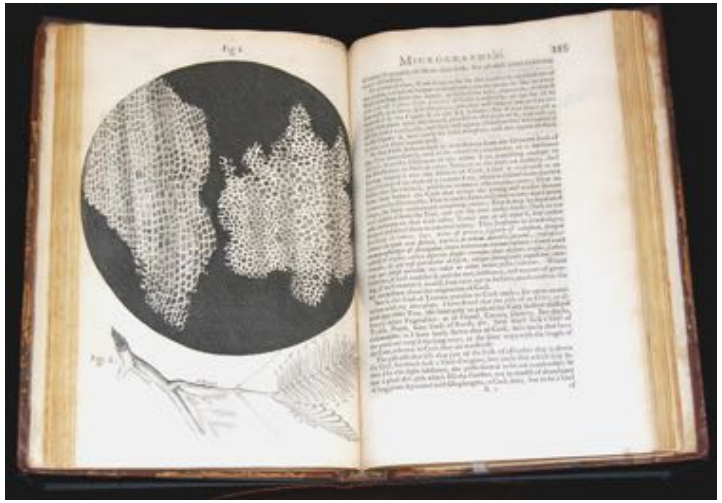
http://www.cell.com/pictureshow/skin
https://library.med.utah.edu/WebPath/webpath.html

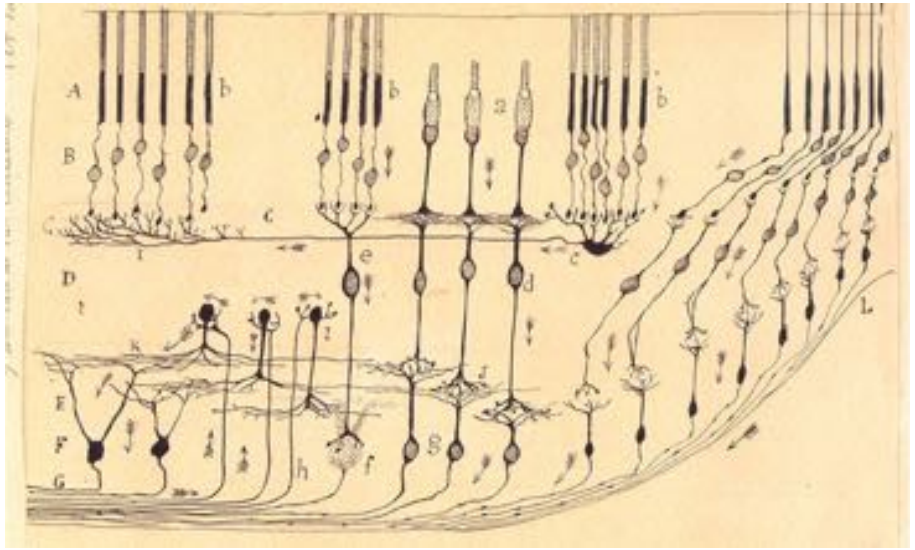
A cell's identity and fate are shaped by many features



Technological innovations allow observation at increasing resolution



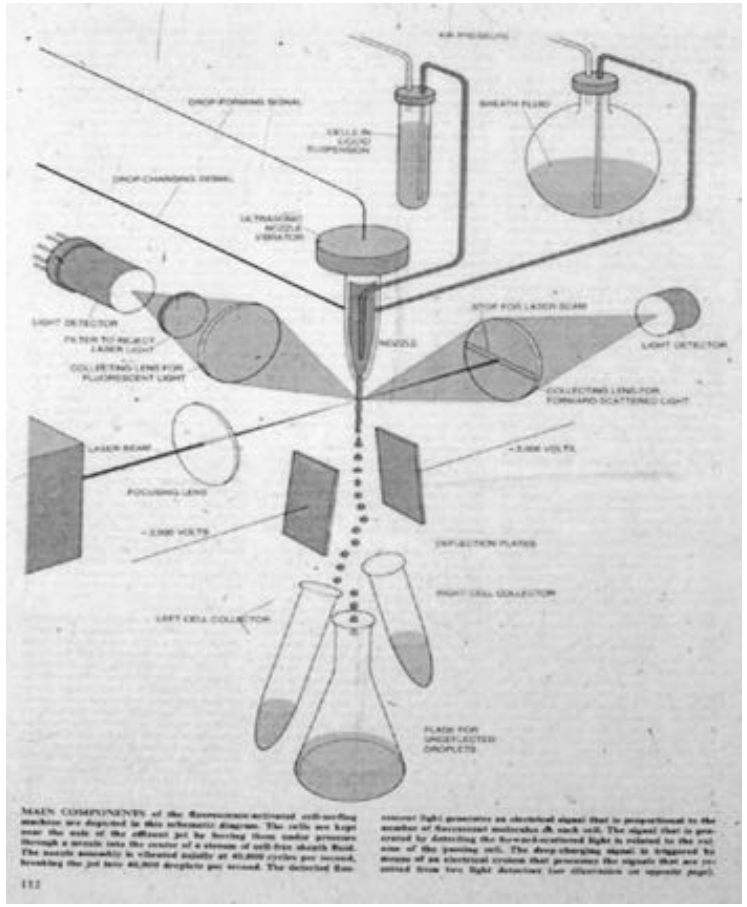


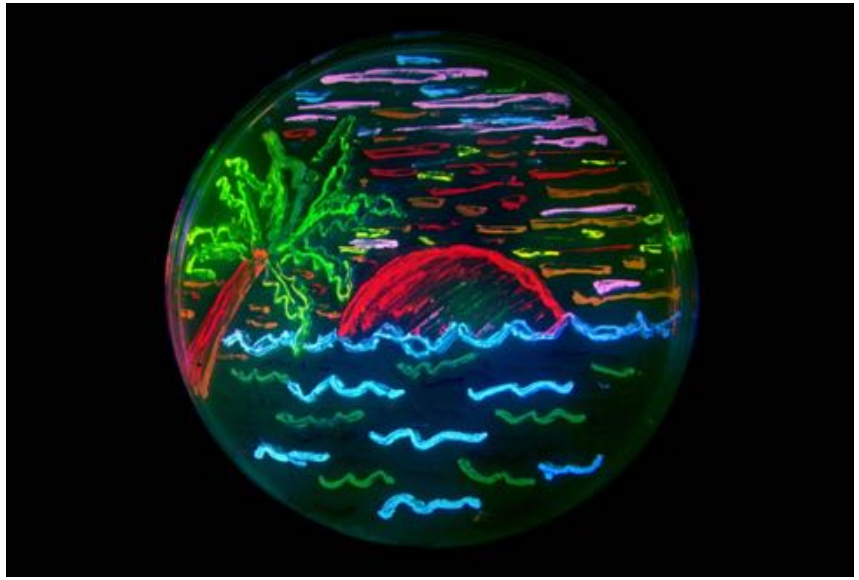


"by the help of Microscopes, there is nothing so small as to escape our inquiry"
Robert Hooke

www.microscopehistory.com
nikoninstruments.com
http://braintour.harvard.edu

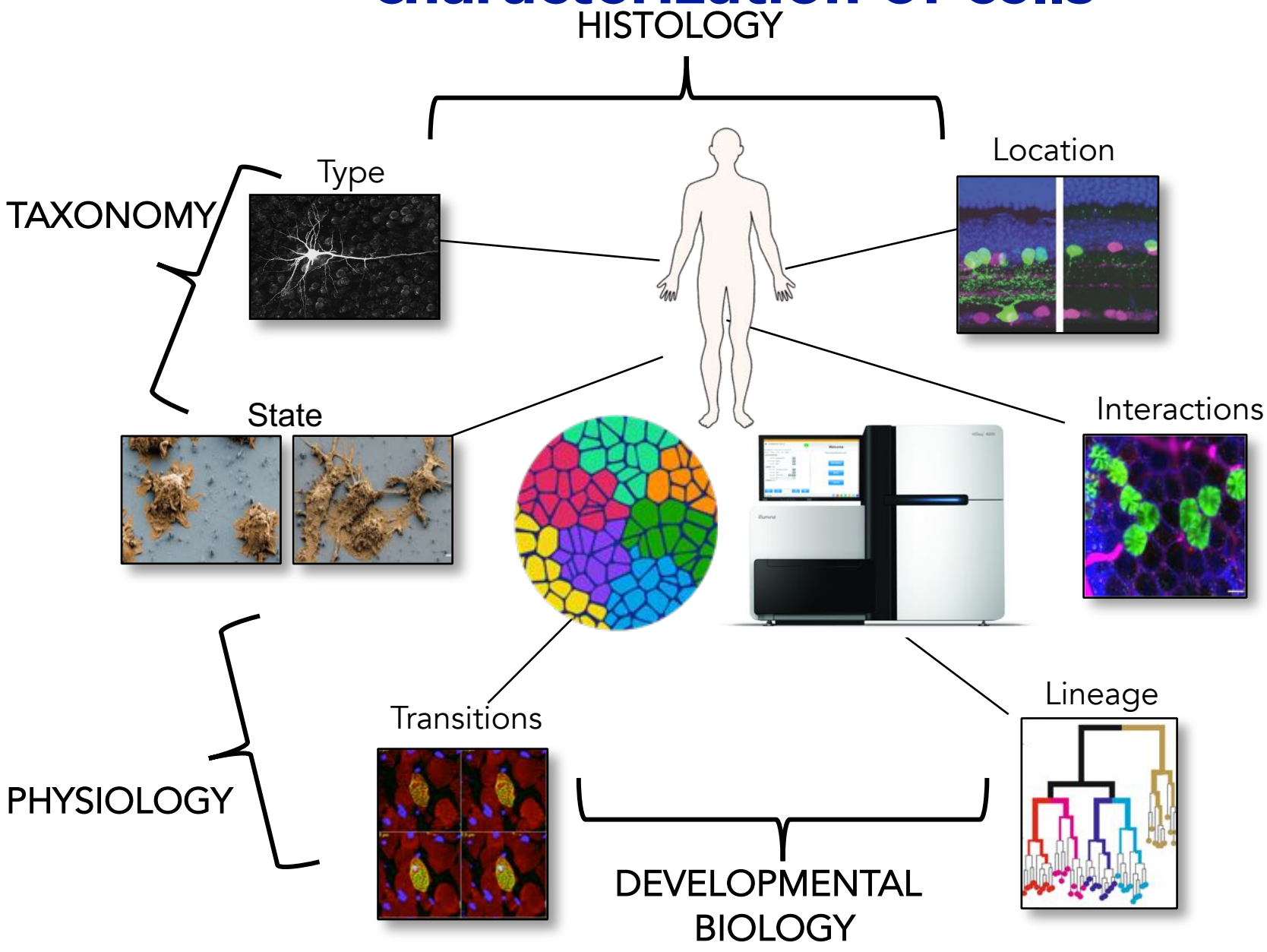
Technological innovations allow observation at increasing resolution



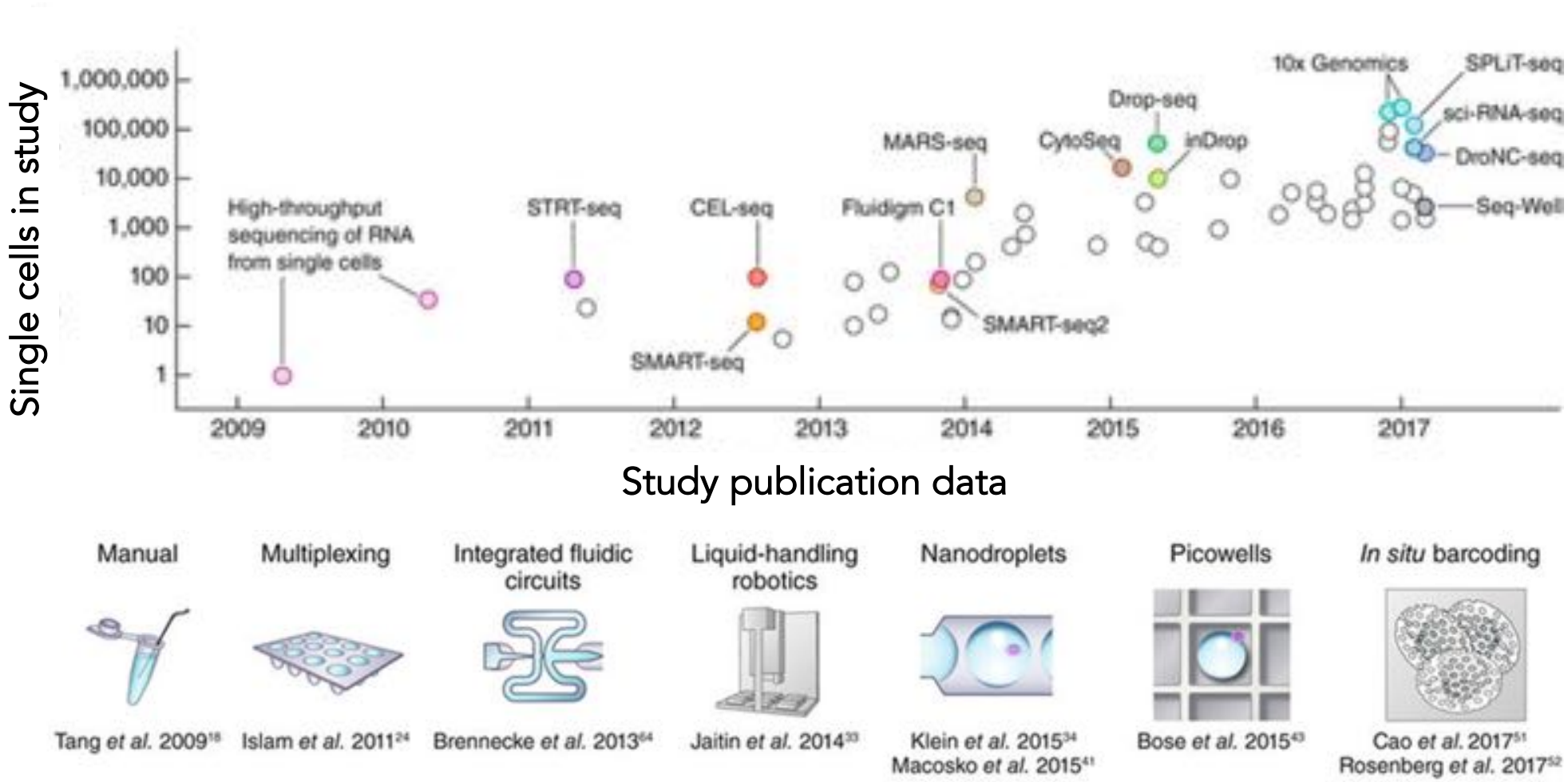


Herzenberg LA, at al. *Sci Am* 1976;**234**:108-117.
Nature doi:10.1038/nature.2016.20532

Single-cell genomics makes possible high-resolution characterization of cells



Single-cell RNA sequencing has grown exponentially

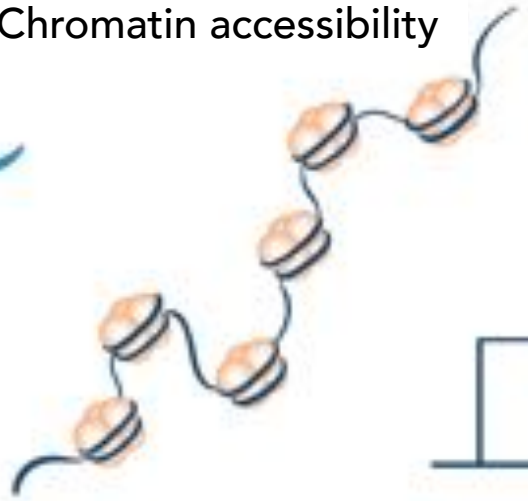


Svensson V et al. (2018) *Nature Protocols* 13:599–604.

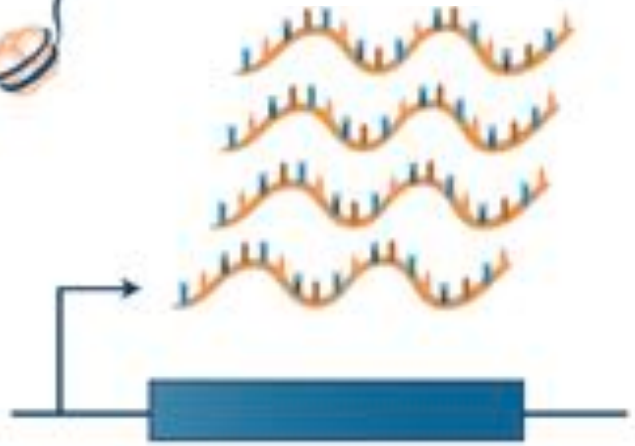
Ongoing developments in single-cell genomics: Many other molecules from single cells may be profiled

Genome

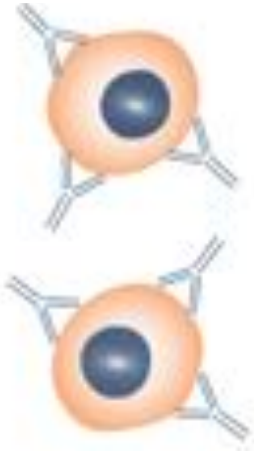
Chromatin accessibility



Gene expression

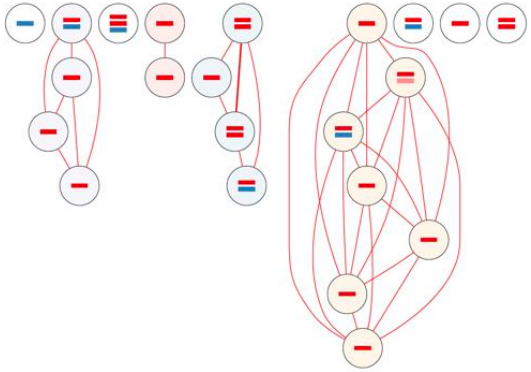


Protein abundance



Spatial transcriptomics

Immune repertoire



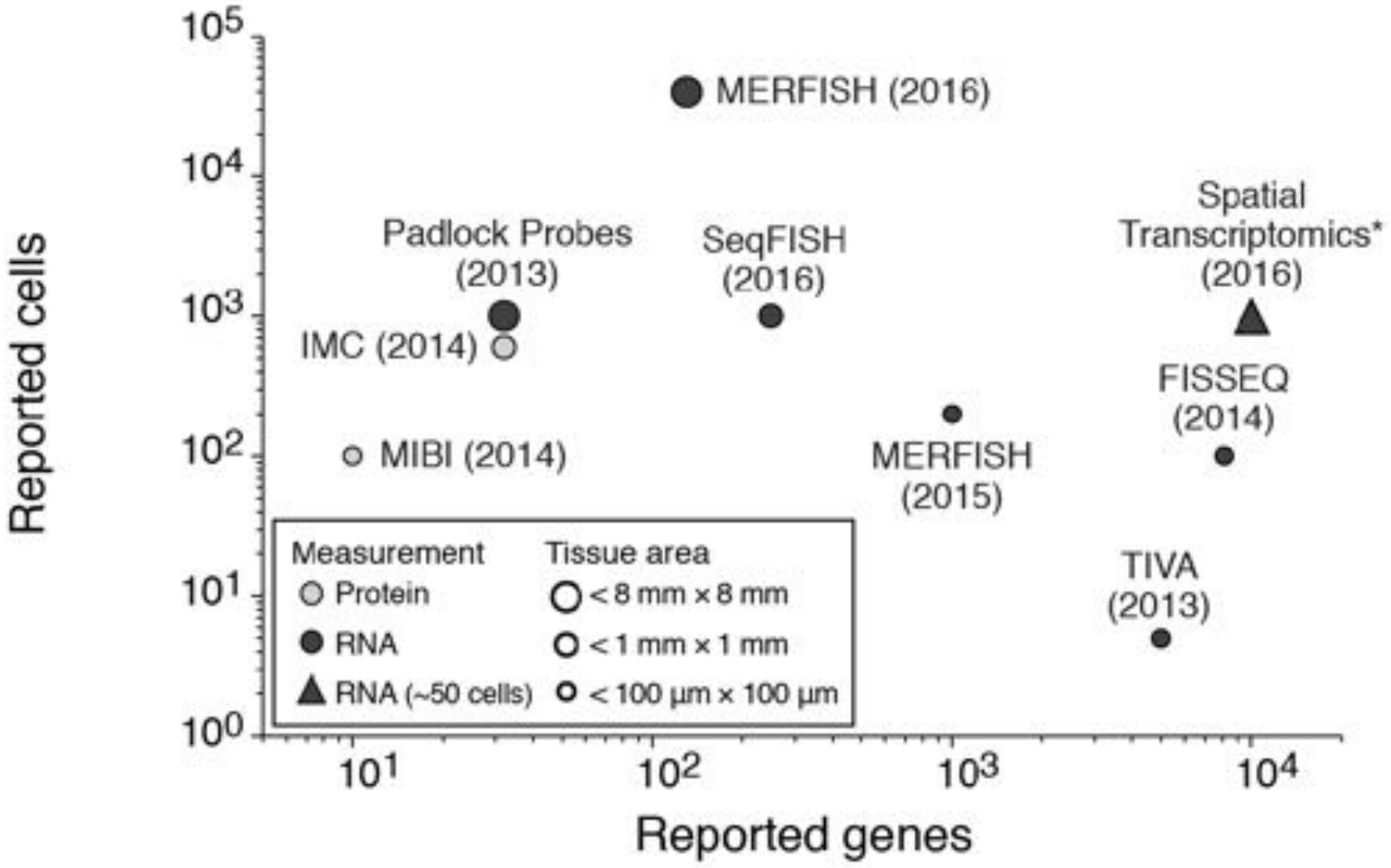
DNA and epigenome	RNA	Proteins	'multi-omics'
Single cell genomes (WES, WGS)	Full length (mRNA, total RNA)	Multi-parameter flow	DNA+RNA (G+T)
Single cell epigenomics (HiC, ChIP, ATAC, mC)	5' and 3' end counting	Mass cytometry	RNA+protein (T+P)
		Single cell proteomics	Epigenome + RNA

Efremova M. et al. (2020) *Nature Methods.* 17:11-20.

Ongoing developments in single-cell genomics: Growing toolbox for spatial genomics

Spatial genomics

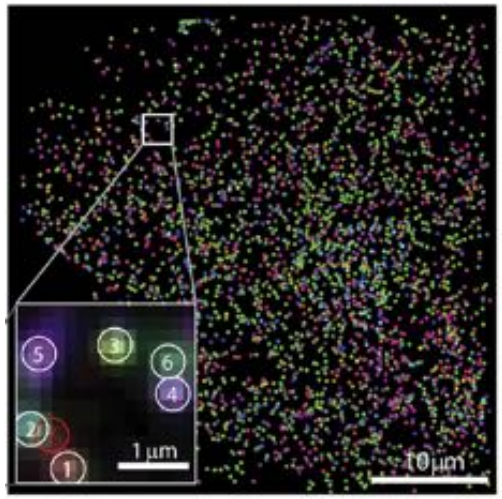
profiling single cells
in their *in situ* context



Human Cell Atlas Consortium White Paper, Sections 2 and 3

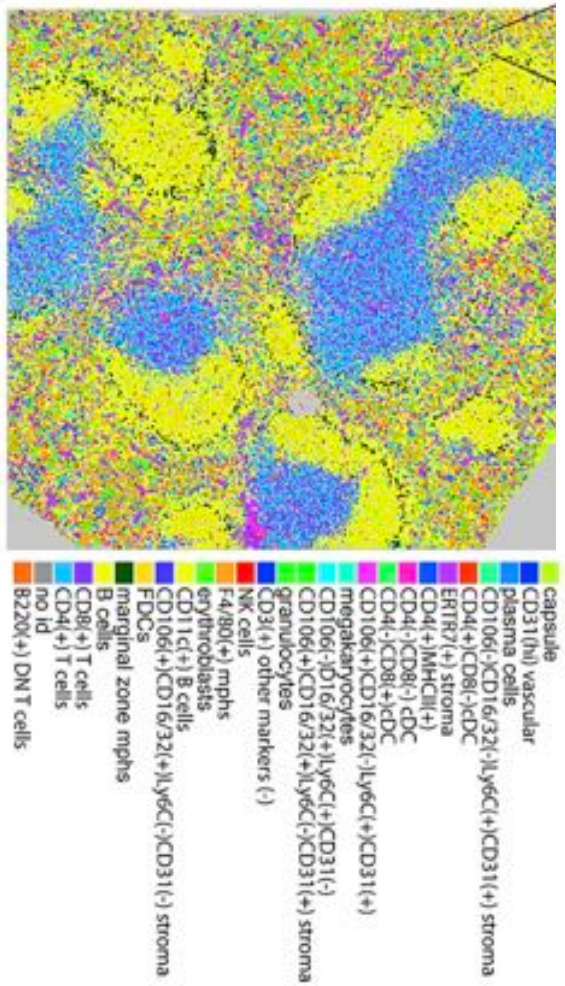
Ongoing developments in single-cell genomics: Growing toolbox for spatial genomics

MERFISH (RNA)



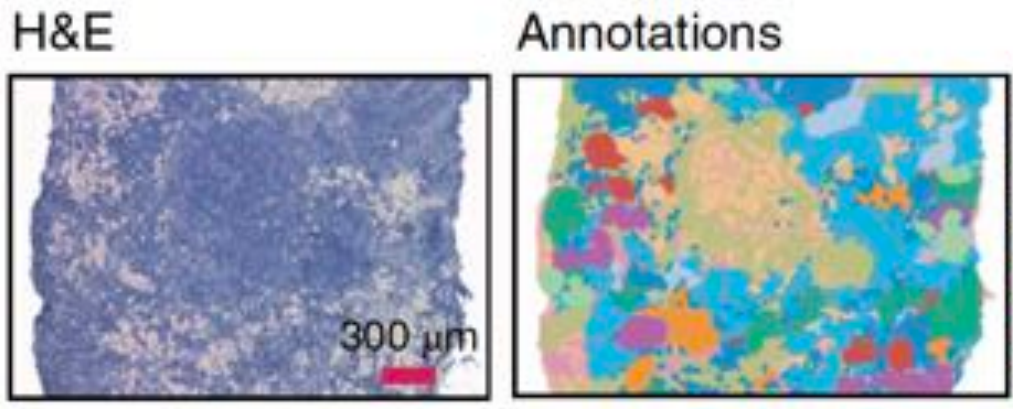
CODEX (protein)

mouse spleen



High Density Spatial Transcriptomics (RNA)

breast cancer



Moffitt J.R. et al. (2016) *PNAS* 113: 11046-11051.
Goltsev Y. et al. (2018) *Cell.* 174: 968-981.
Vickovic S. et al. (2019) *Nature Methods.* 16: 987–990.

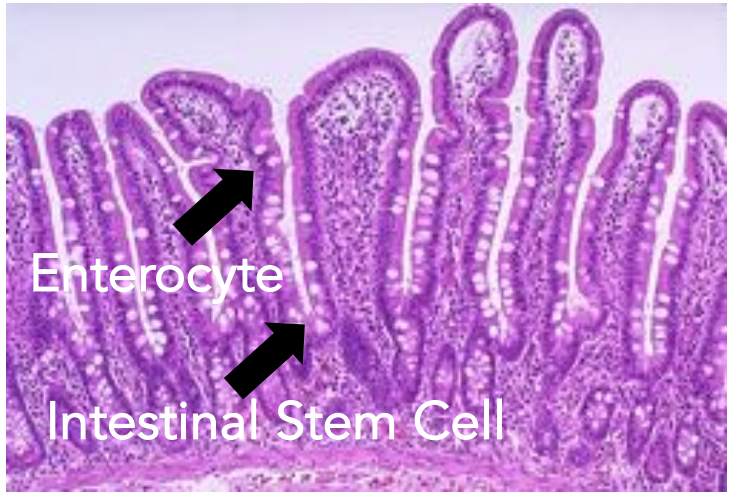
Human Cell Atlas mission

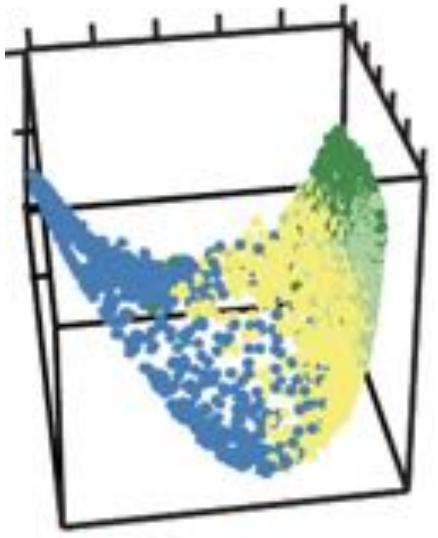
To create a comprehensive reference map of the types and properties of all human cells, the fundamental unit of life, as a basis for understanding, diagnosing, monitoring, and treating health and disease.

"The vestiges of the rupture reveal themselves, if someone brings forward a map of the world and considers carefully the coasts of the three [continents]."
Dutch map maker Abraham Ortelius (1596)

Human Cell Atlas mission

To create a comprehensive reference map of the types and properties of all human cells, the fundamental unit of life, as a basis for understanding, diagnosing, monitoring, and treating health and disease.

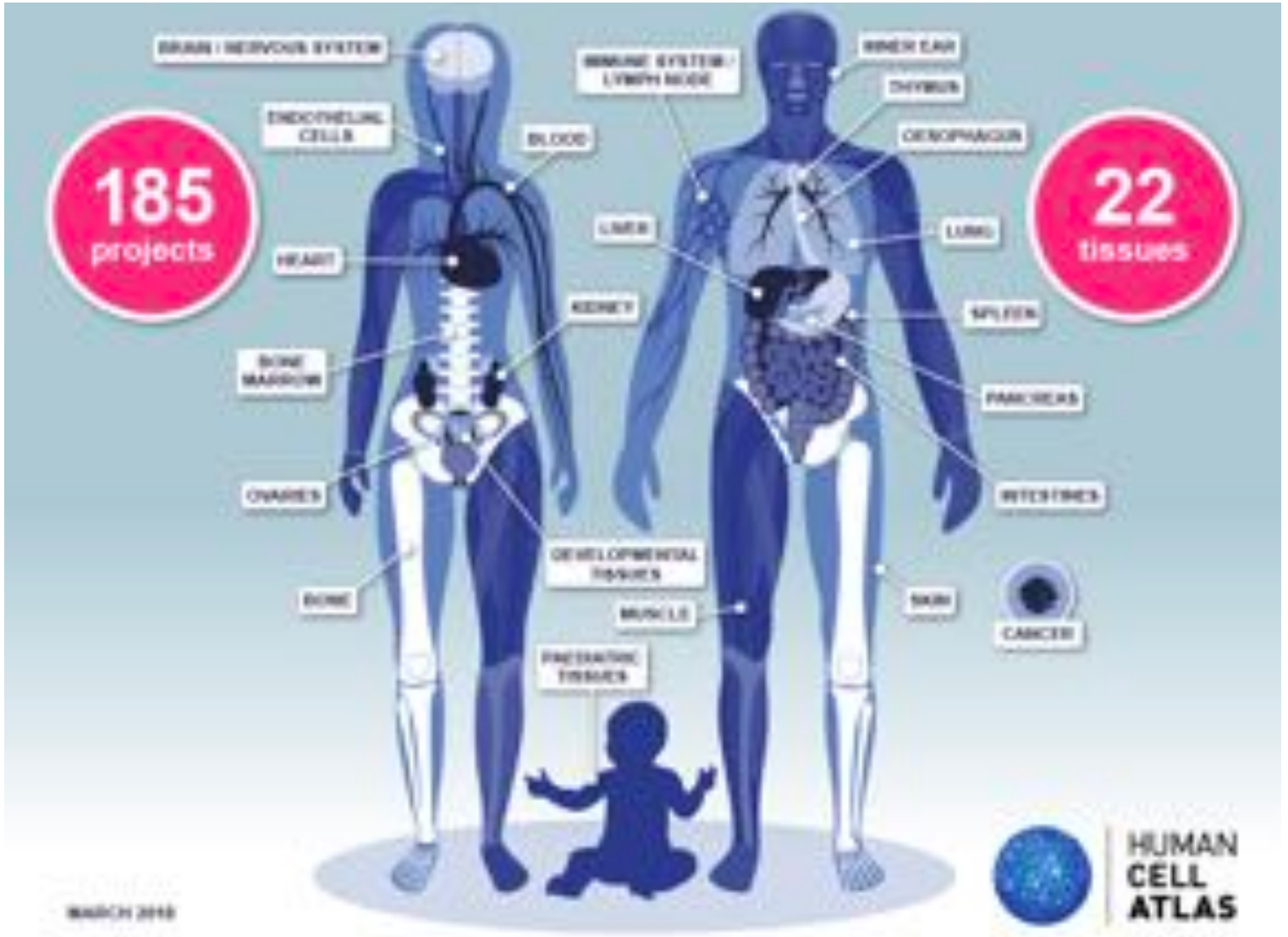




- Intestinal Stem Cell
- Enterocyte Progenitor
- Enterocyte

Haber, A.L., Biton, M., Rogel, N. et al. (2017) *Nature* 551: 333-339.

The Human Cell Atlas will sample most major tissues

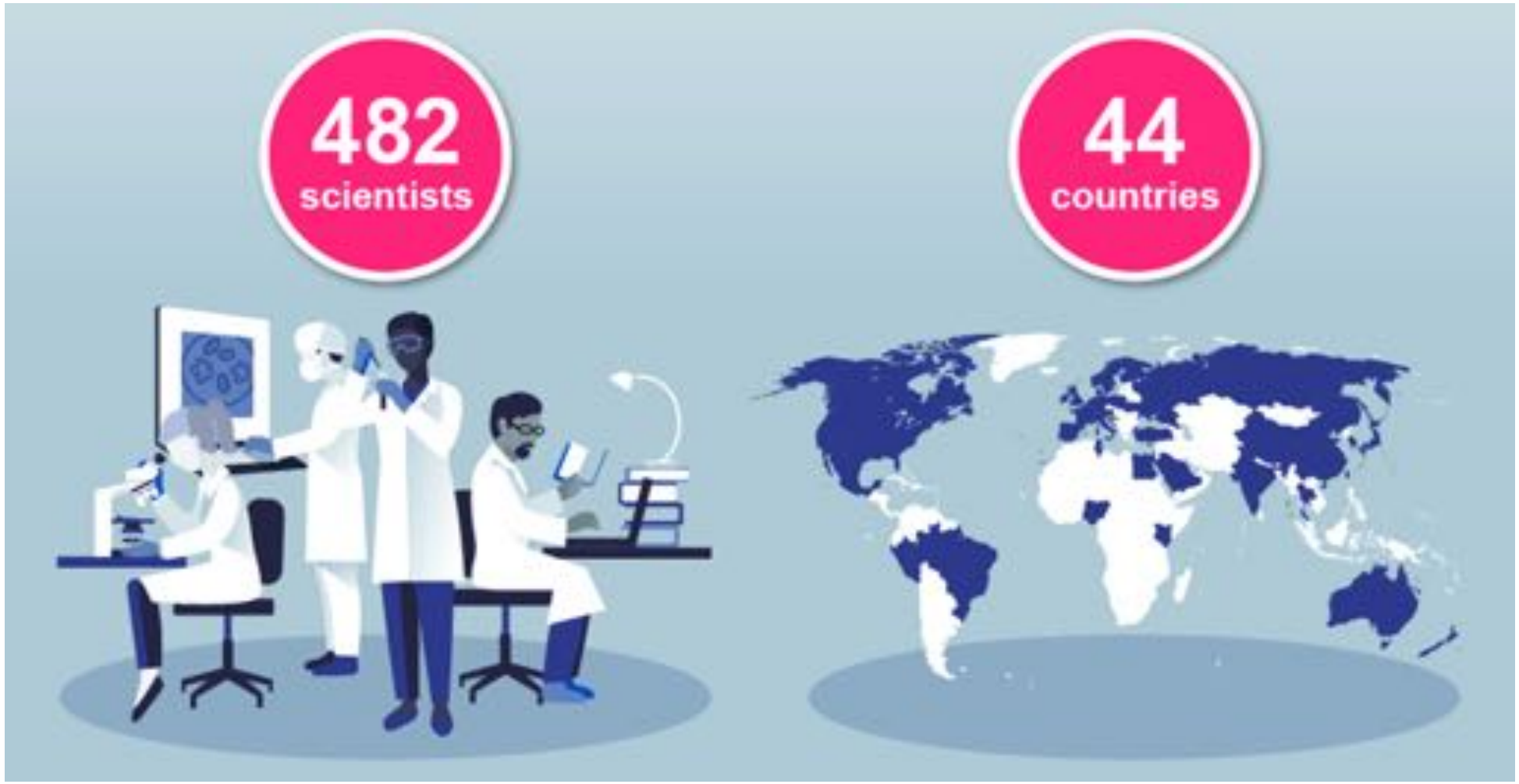


Tissue Sources

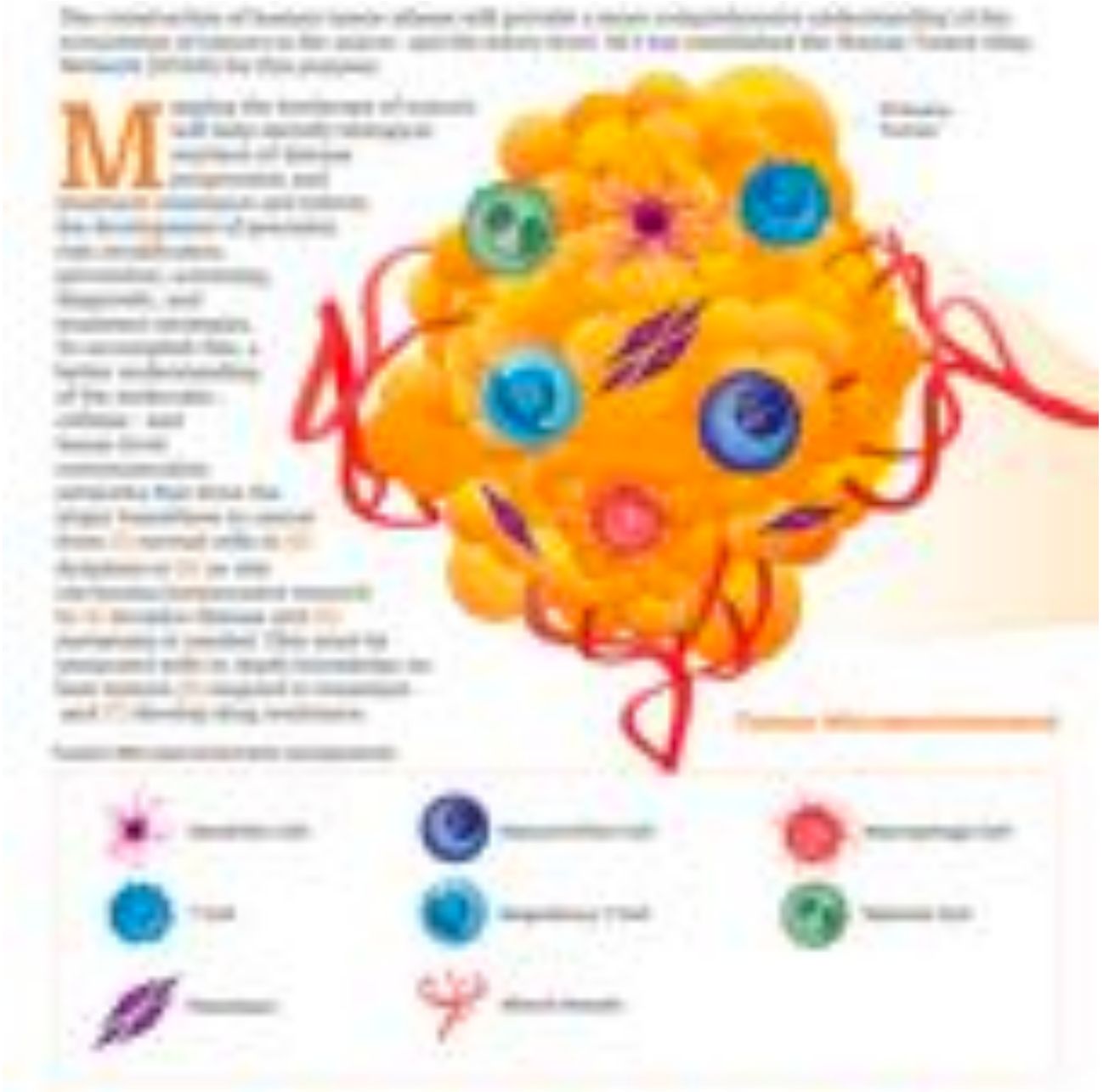
- 20-60 adults
- living donors, organ donors, post-mortem
- fresh, frozen, or fixed

First draft: 10 M cells → Comprehensive atlas: 10 B cells

The Human Cell Atlas is a global effort



Human Tumor Atlas Network



https://www.cancer.gov

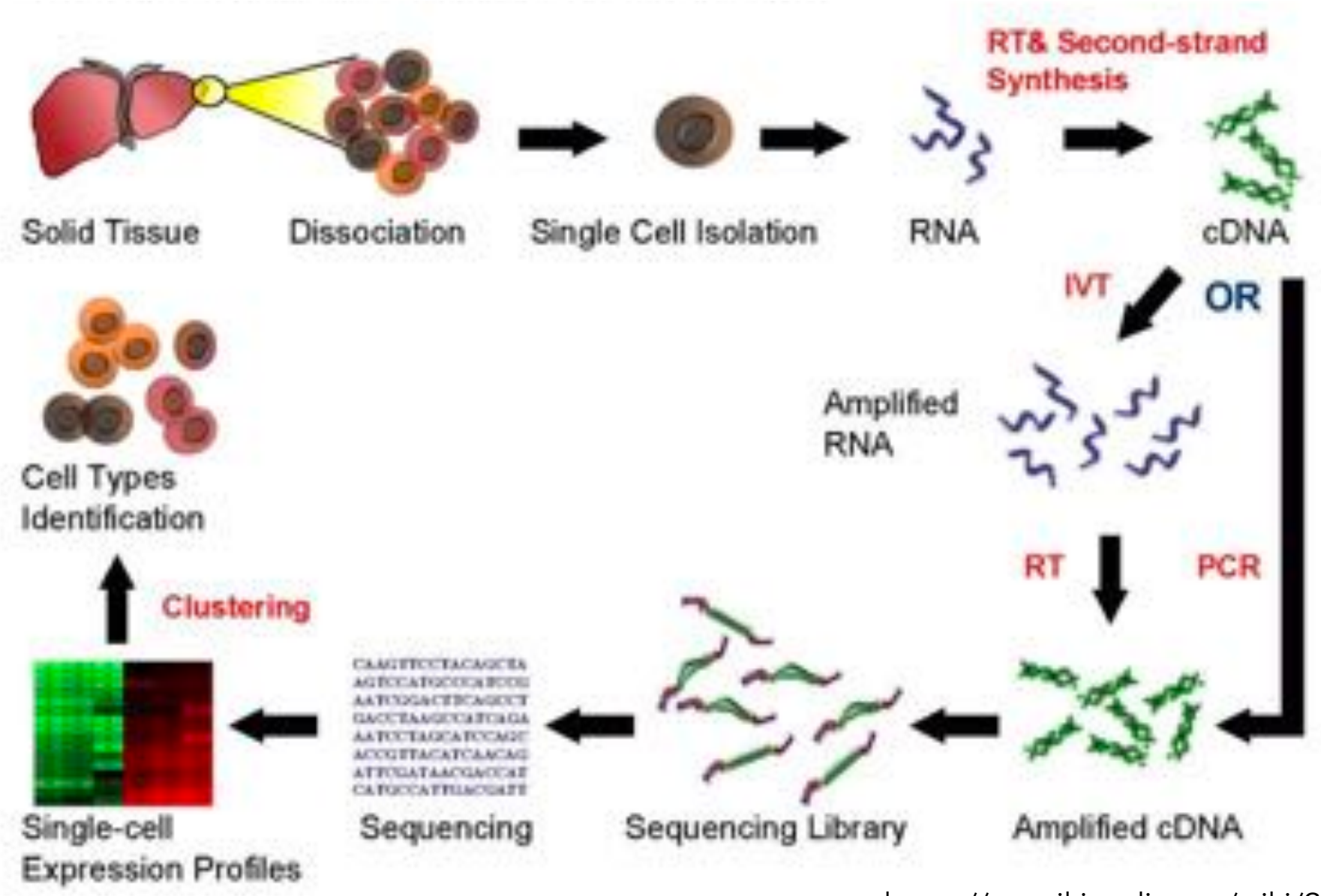
Typical RNA-Seq workflow

mRNA from a source

Slide courtesy of Karthik Shekhar

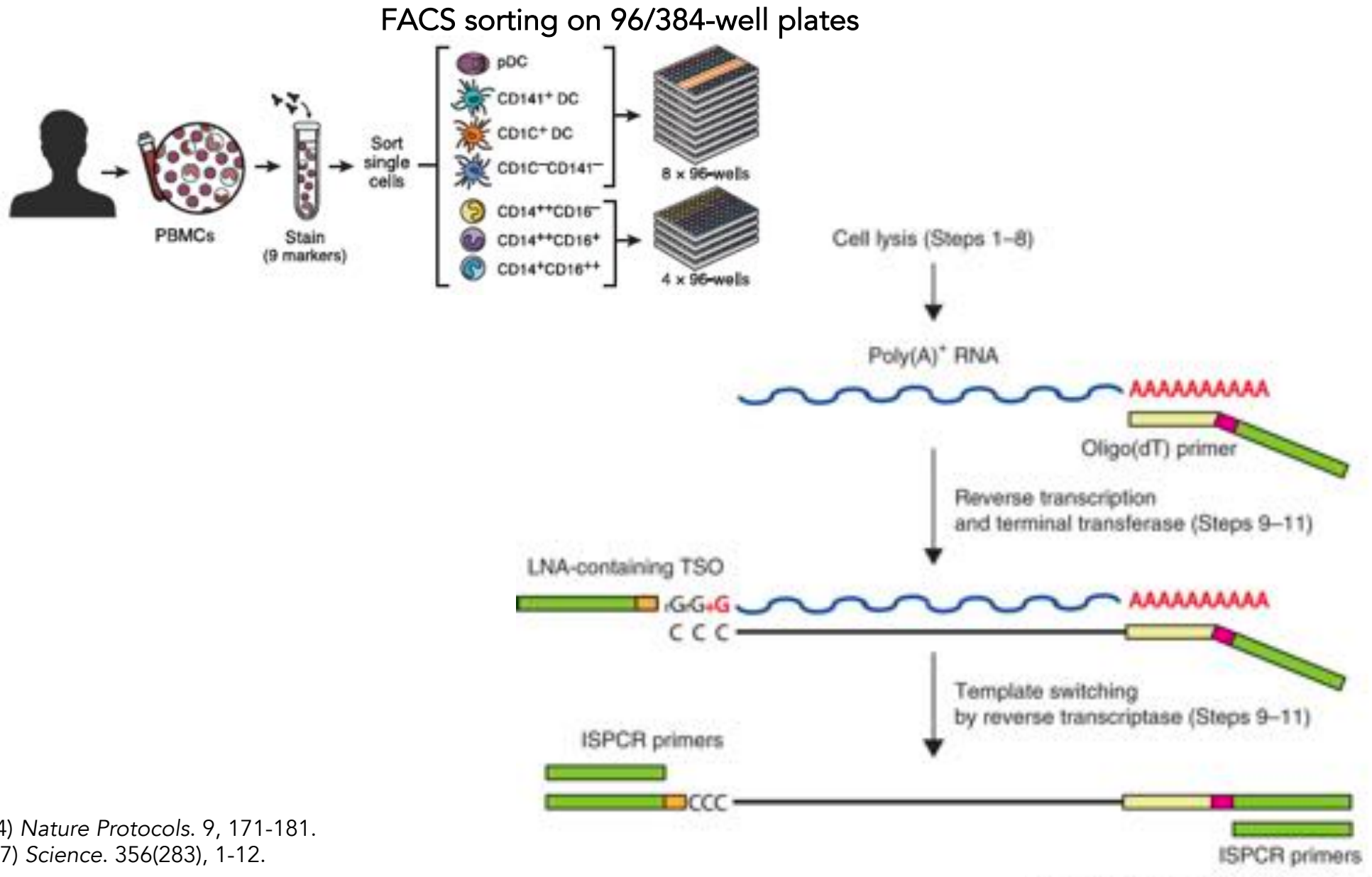
Wang, Gerstein, and Snyder, Nat. Rev. Genetics, 2009

Experimental design: single cell RNA-Seq



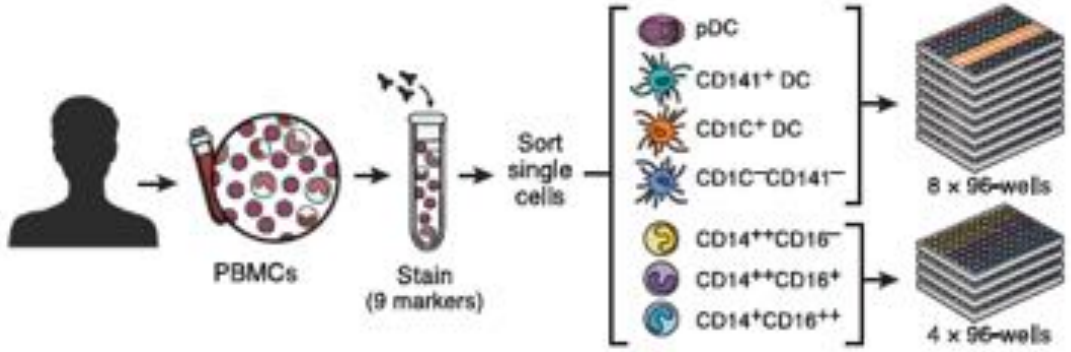
https://en.wikipedia.org/wiki/Single_cell_sequencing

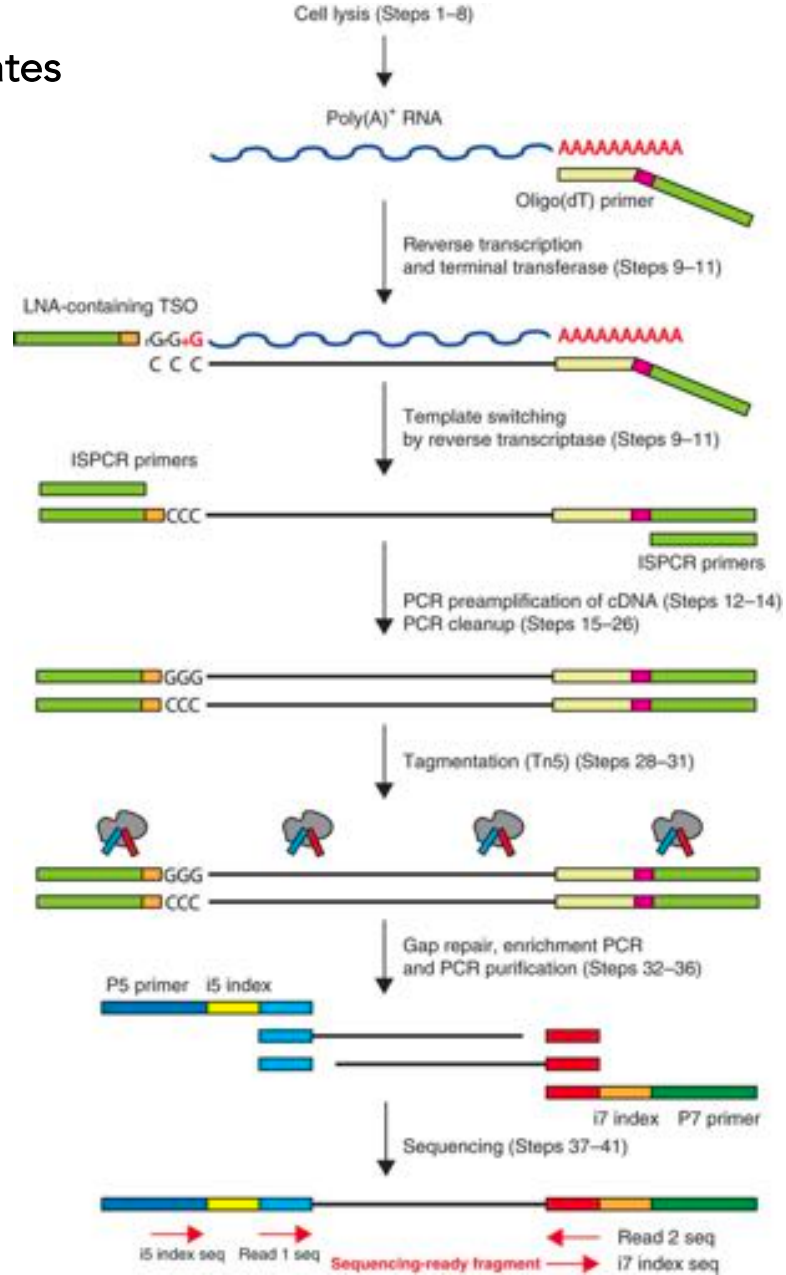
Single cell transcriptomics using SMART-Seq2



Picelli S. et *al*. (2014) *Nature Protocols*. 9, 171-181.
Villani A. et *al*. (2017) *Science*. 356(283), 1-12.

Single cell transcriptomics using SMART-Seq2

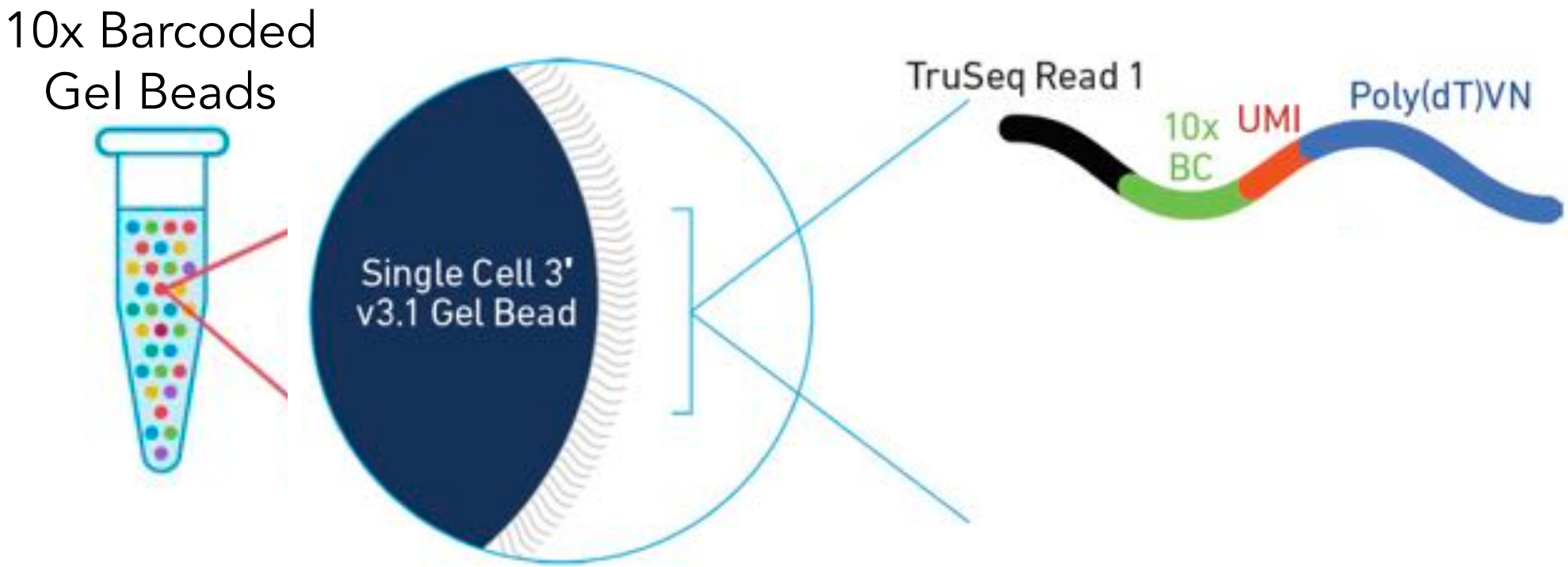




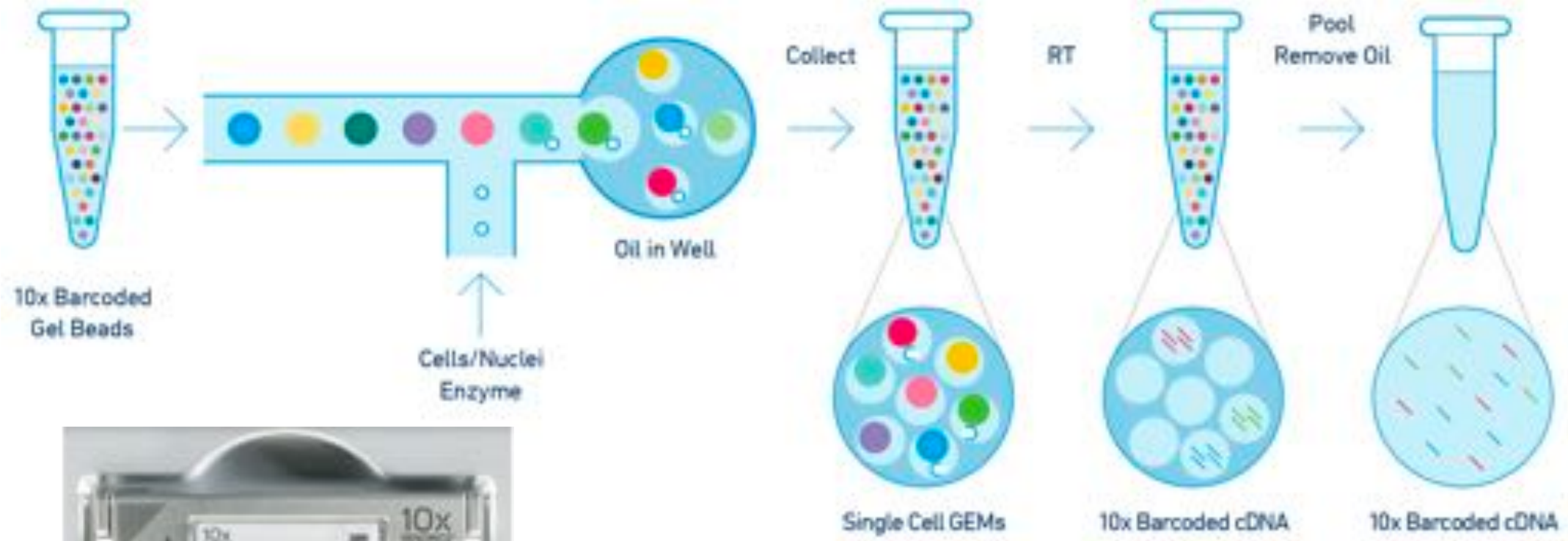
Picelli S. et *al*. (2014) *Nature Protocols*. 9, 171-181.
Villani A. et *al*. (2017) *Science*. 356(283), 1-12.

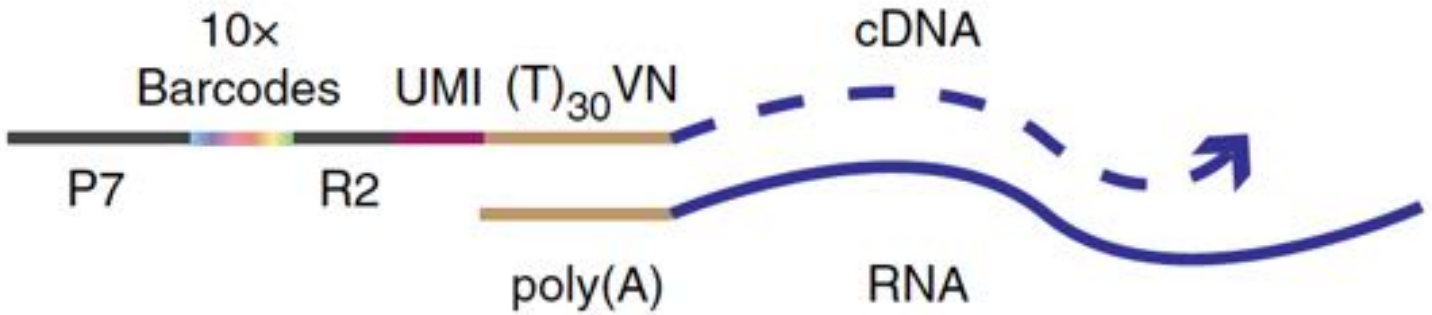
Single cell transcriptomics using droplets and microfluidics

10x Next GEM samples a pool of ~3,500,000 10x Barcodes to separately index each cell's transcriptome



Single cell transcriptomics using 10x Chromium system



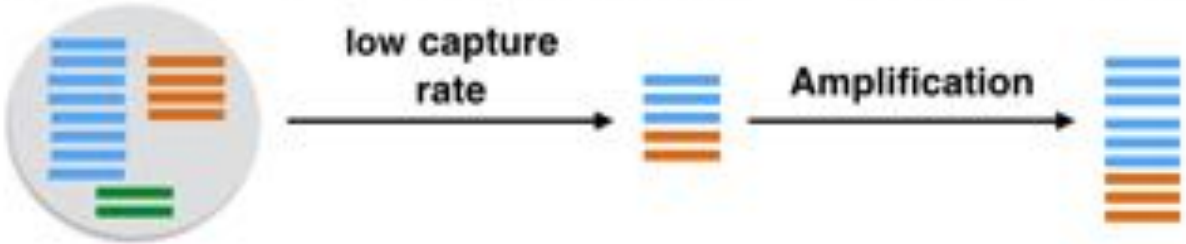


Visit https://liorpachter.wordpress.com/2019/02/07/sub-poisson-loading-for-single-cell-rna-seq/

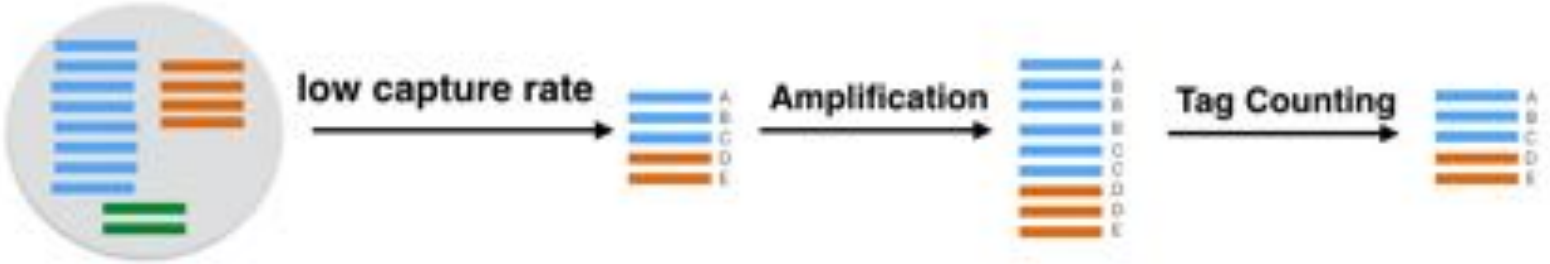
UMI : Unique Molecular Identifiers (Random Molecular Tags)

Early labeling of mRNA molecules with random nucleotide tags enables amplification biases to be corrected

- Low input amount -> transcript dropout + PCR amplification bias



- Unique Molecular Identifiers (UMIs) can correct for PCR bias



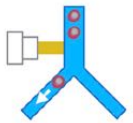
Remember : UMIs do not correct for low-capture rates, which leads to an abundance of false negatives. Capture rates are estimated to 5-20% across various protocols

Slide courtesy of Karthik Shekhar

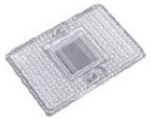
Single-cell RNA-Seq pipeline

Cell capture and Barcoding → Library Prep → Sequencing

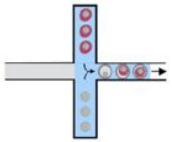
FACS



Micro-capture



Droplet



Nanowell

Split/pool barcoding

PolyA vs random priming

3′/5′ end tagging vs full-length

PCR vs *in vitro* transcription

…

Miseq

~ 20M reads total

Nextseq

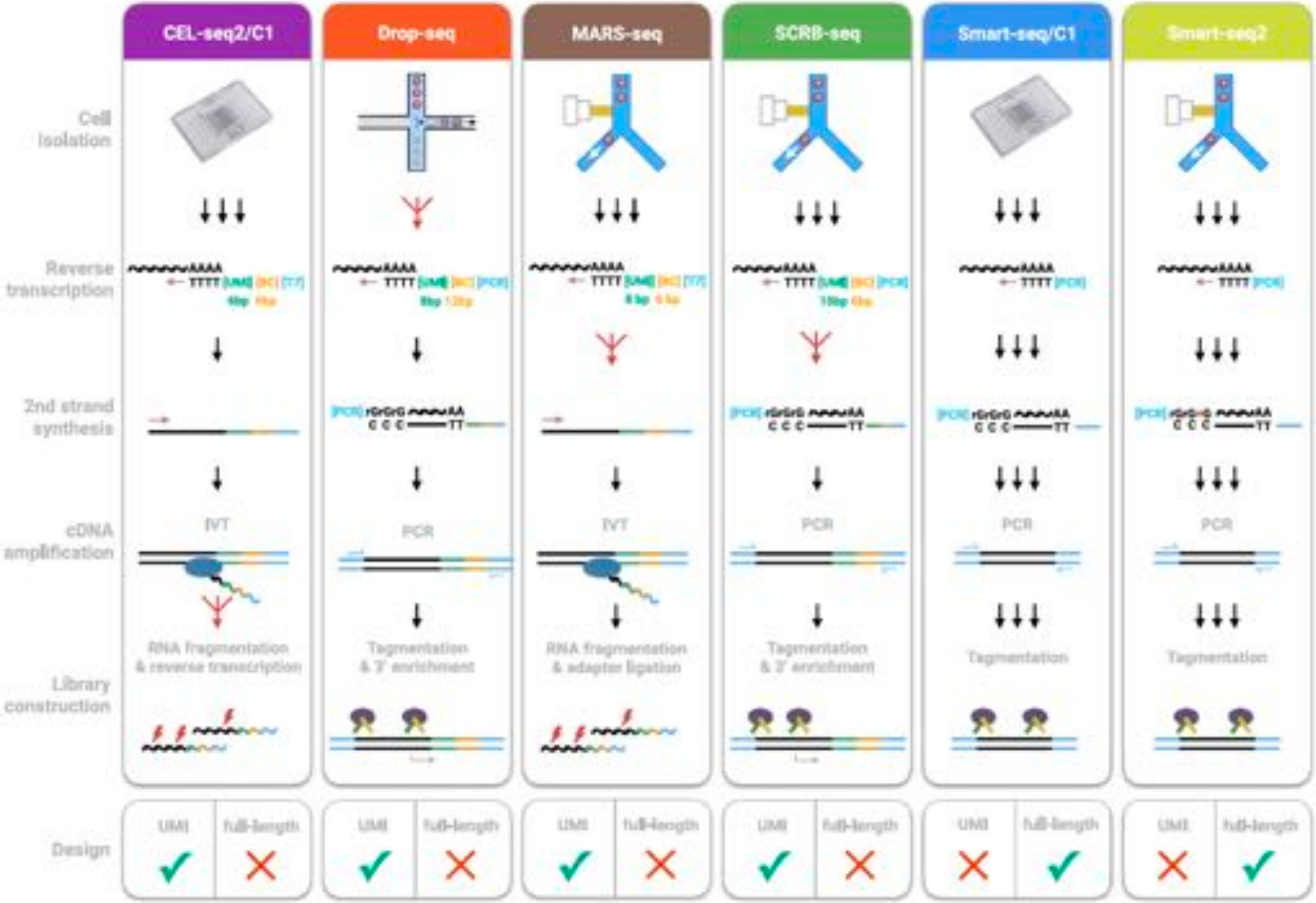
~ 500M reads total

HiSeq 4000

4 billion reads

Slide courtesy of Karthik Shekhar

There are many single-cell RNA sequencing methods



Ziegenhain C et *al.* (2017) *Mol Cell.* 65(4):631-643.

There are many single-cell RNA sequencing methods

	SMART-seq2	CEL-seq2	STRT-seq	Quartz-seq2	MARS-seq	Drop-seq	inDrop	Chromium	Seq-Well	sci-RNA-seq	SPLIT-seq
Single-cell isolation	FACS, microfluidics	FACS, microfluidics	FACS, microfluidics, nanowells	FACS	FACS	Droplet	Droplet	Droplet	Nanowells	Not needed	Not needed
Second strand synthesis	TSO	RNase H and DNA pol I	TSO	PolyA tailing and primer ligation	RNase H and DNA pol I	TSO	RNase H and DNA pol I	TSO	TSO	RNase H and DNA pol I	TSO
Full-length cDNA synthesis?	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Barcode addition	Library PCR with barcoded primers	Barcoded RT primers	Barcoded TSOs	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers and library PCR with barcoded primers	Ligation of barcoded RT primers
Pooling before library?	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Library amplification	PCR	In vitro transcription	PCR	PCR	In vitro transcription	PCR	In vitro transcription	PCR	PCR	PCR	PCR
Gene coverage	Full-length	3′	5′	3′	3′	3′	3′	3′	3′	3′	3′
Number of cells per assay											

Each protocol has **advantages** and **limitations**. What one ends up using is often dictated by multiple features - the **biological context, cost, objective** etc.

Chen X. et *al.* (2018) *Annual Review of Biomedical Data Science* Vol. 1:29-51

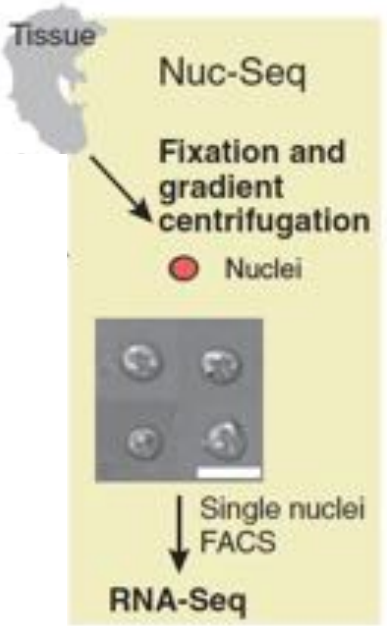
Considerations for single cell RNA-Seq

Choose protocol based on :

- Throughput (# of cells / reaction)
- Sample of origin
- Cost / Labor / Time limitations
- Gene body coverage - 5', 3' biased, or full-length?
- UMI vs no-UMI
- Sequencing depth / cell

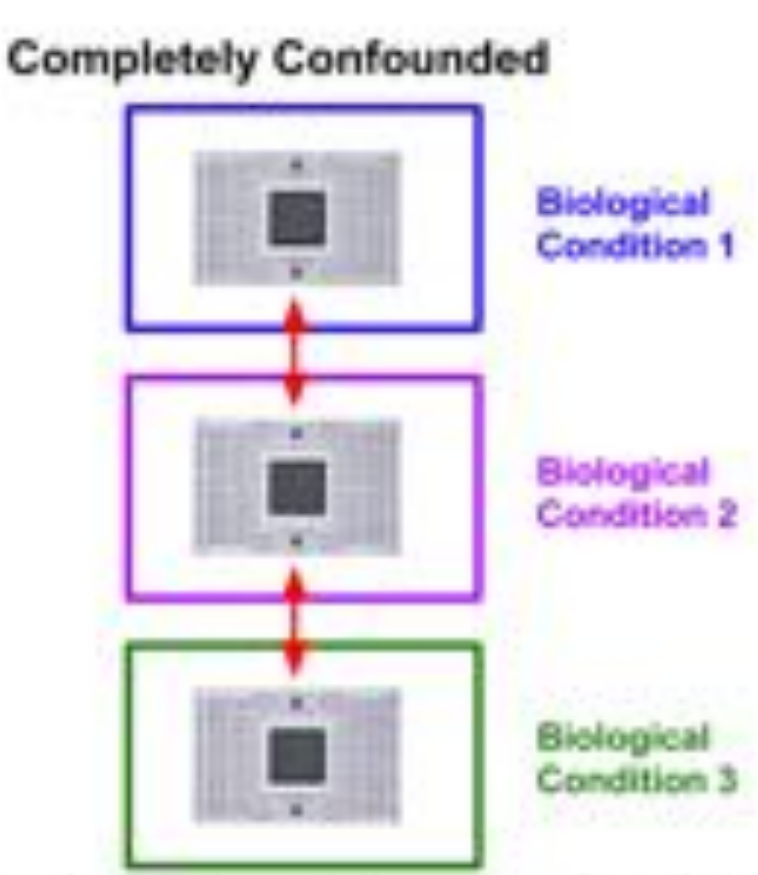
For example :

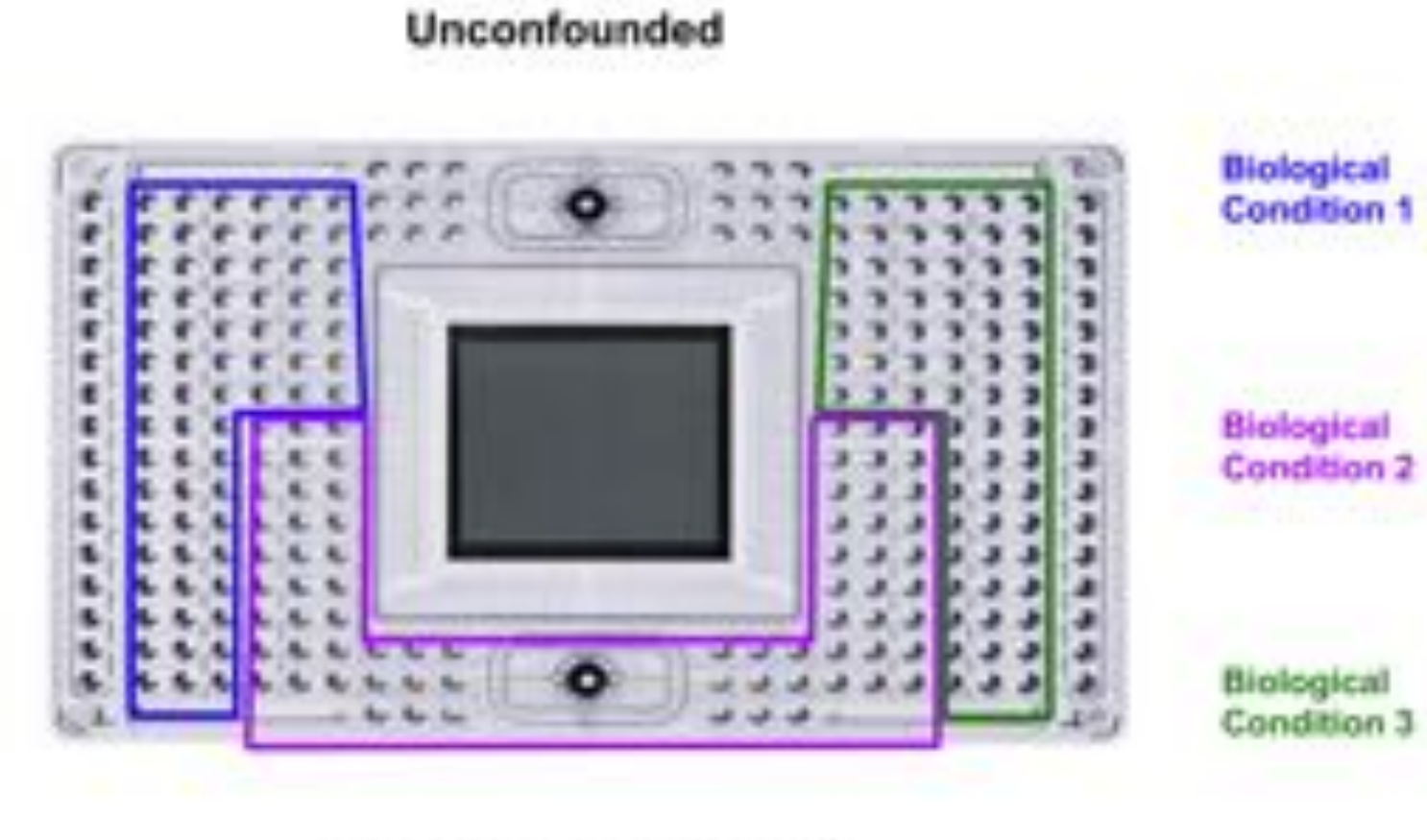
- If I want to classify all cell types in a diverse tissue (e.g. brain), I need high throughput
- If I want to re-annotate the transcriptome and discover new isoforms, I need full-length coverage
- If I only have access to archival human samples, I will need to use a method that permits fixed cells (or nuclei)



Slide modified from Karthik Shekhar
Habib N. et *al.* (2016) *Science*. 353, 925-928.

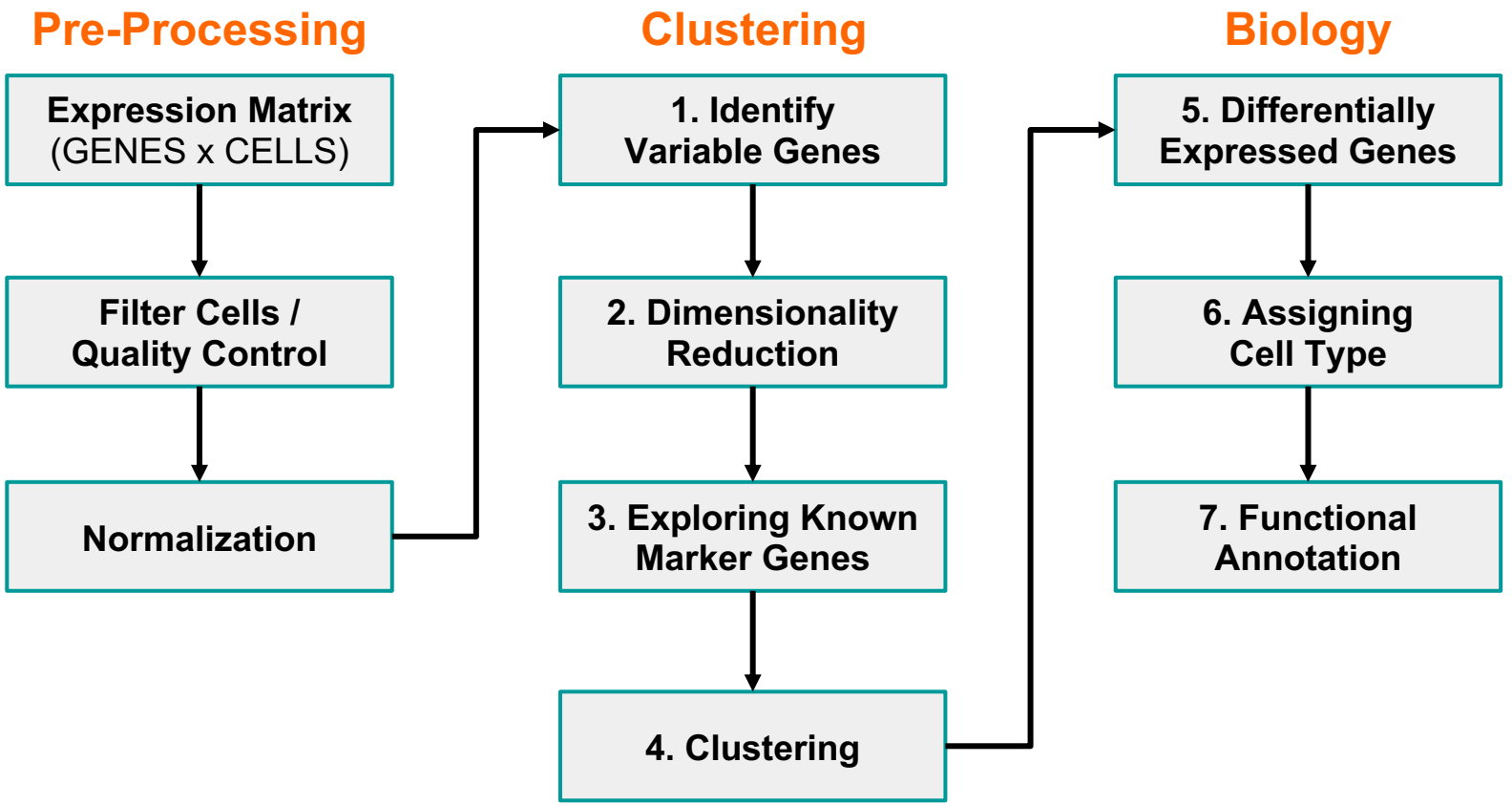
The most powerful way to control batch effects is with careful experimental design



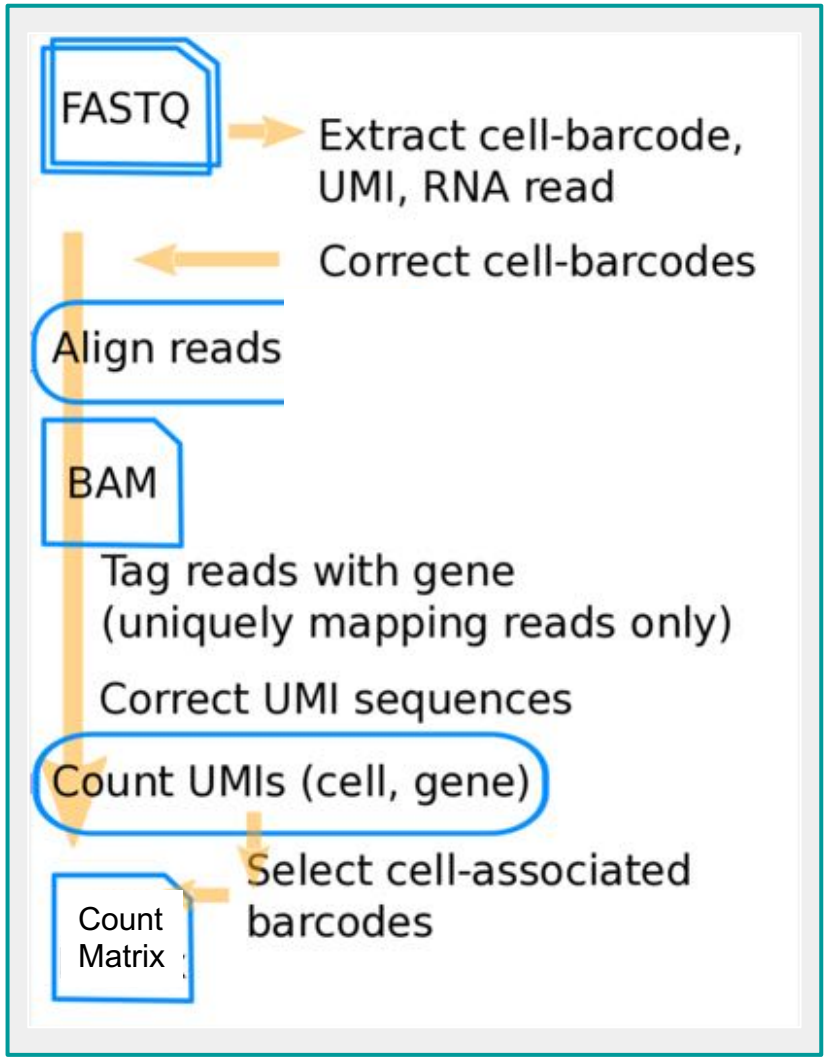


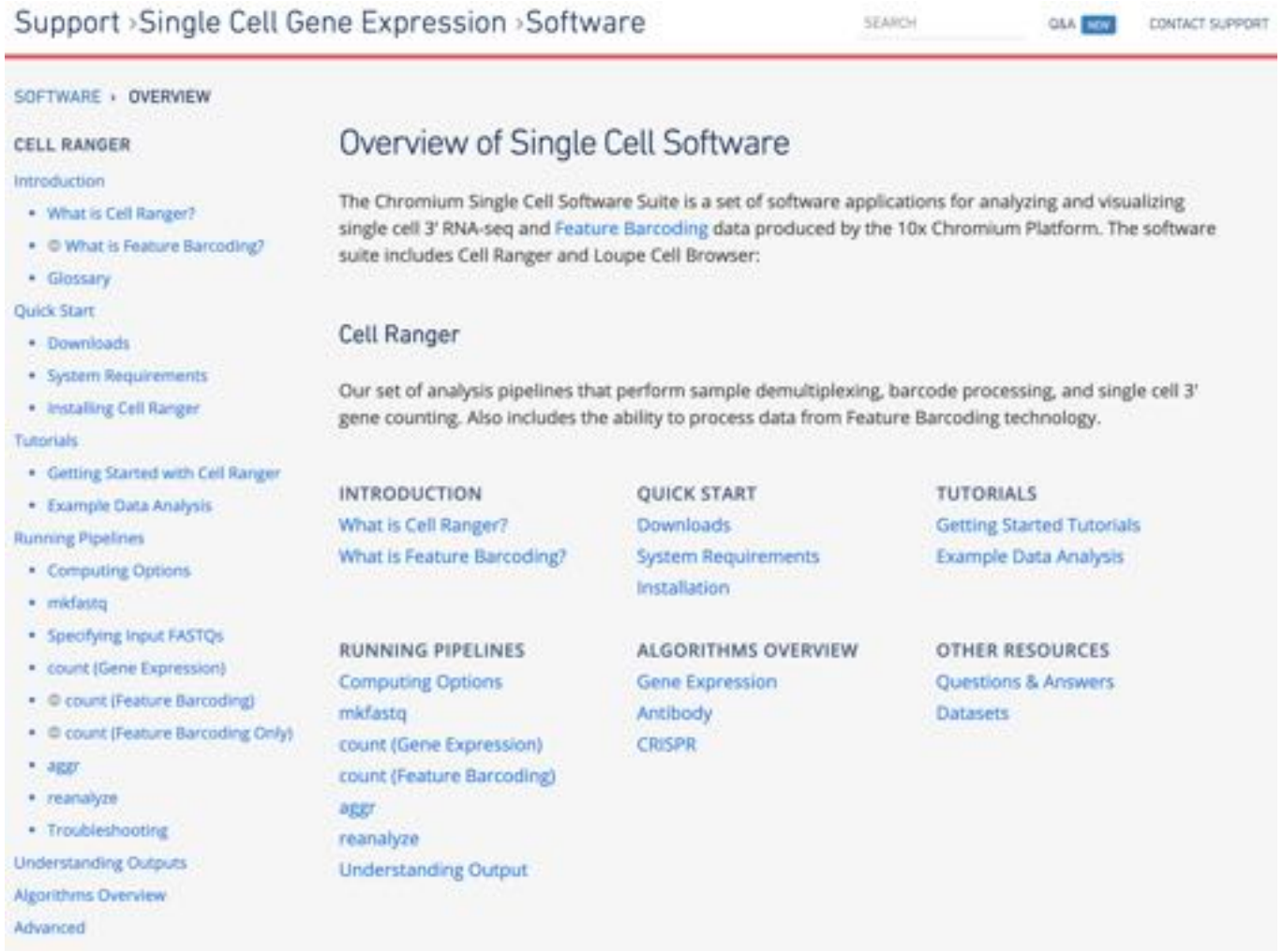
https://scrnaseq-course.cog.sanger.ac.uk/website/ideal-scrnaseq-pipeline-as-of-oct-2017.html

Single-cell RNA-Seq analysis pipeline
Analyzing the expression data



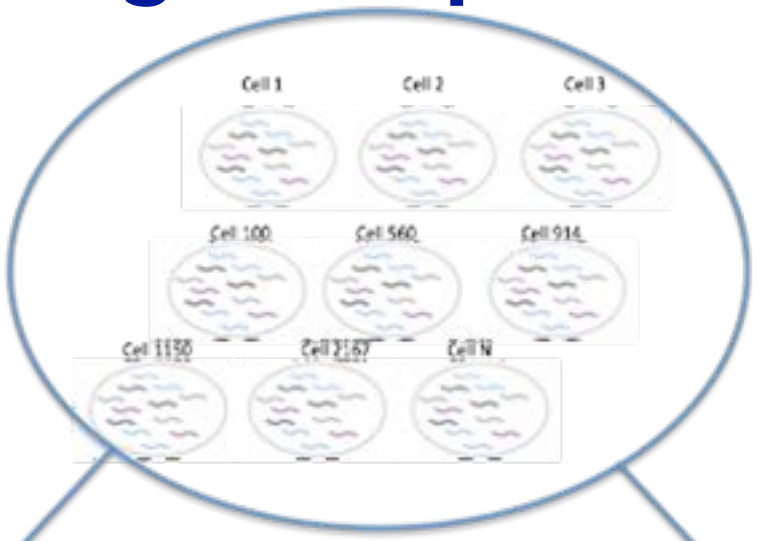
Single-cell RNA-Seq analysis pipeline
Generating the count matrix





https://support.10xgenomics.com/single-cell-gene-expression

Single-cell gene expression distributions are very different from bulk gene expression distributions



Gene Expression Matrix

	Exp
Gene 1	100
Gene 2	3.5
Gene 3	10
...	
Gene K	95
...	
Gene 25K	0.4

	Cell 1	Cell 2	Cell 3	....	Cell 5K
Gene 1	3	0	1		2
Gene 2	0	2	0		1
Gene 3	1	0	3		5
...					
Gene K	14	7	1		0
...					
Gene 25K	0	13	1		0

Population Average

Bulk

Cellular resolution

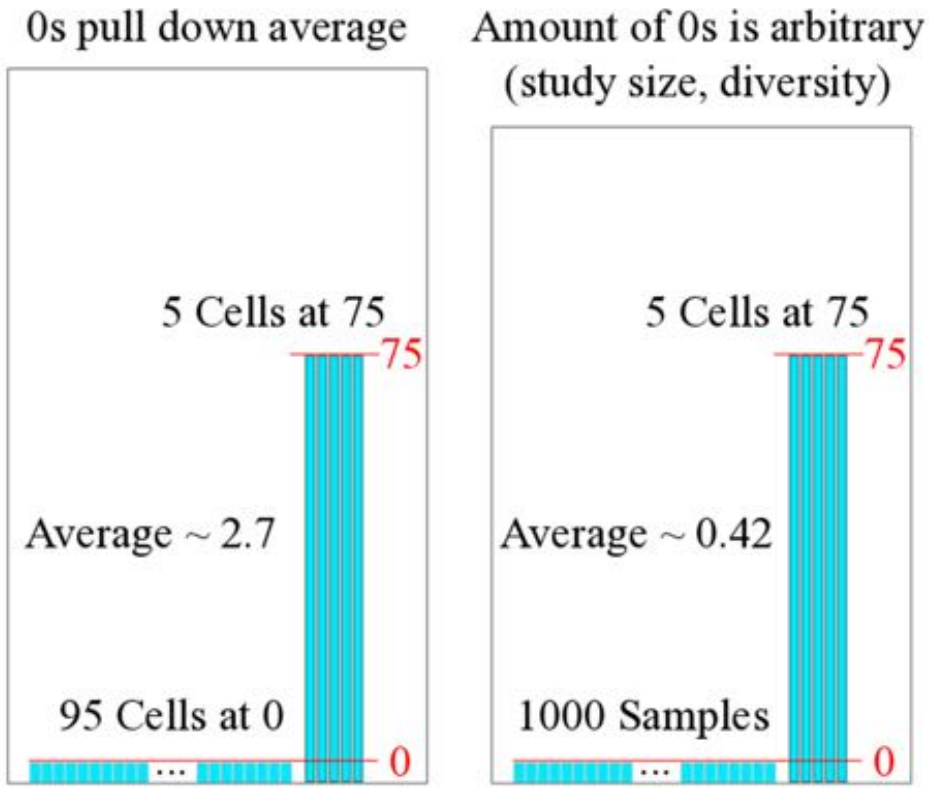
Single cell

Slide courtesy of Karthik Shekhar

Some single-cell RNA-Seq data challenges to remember

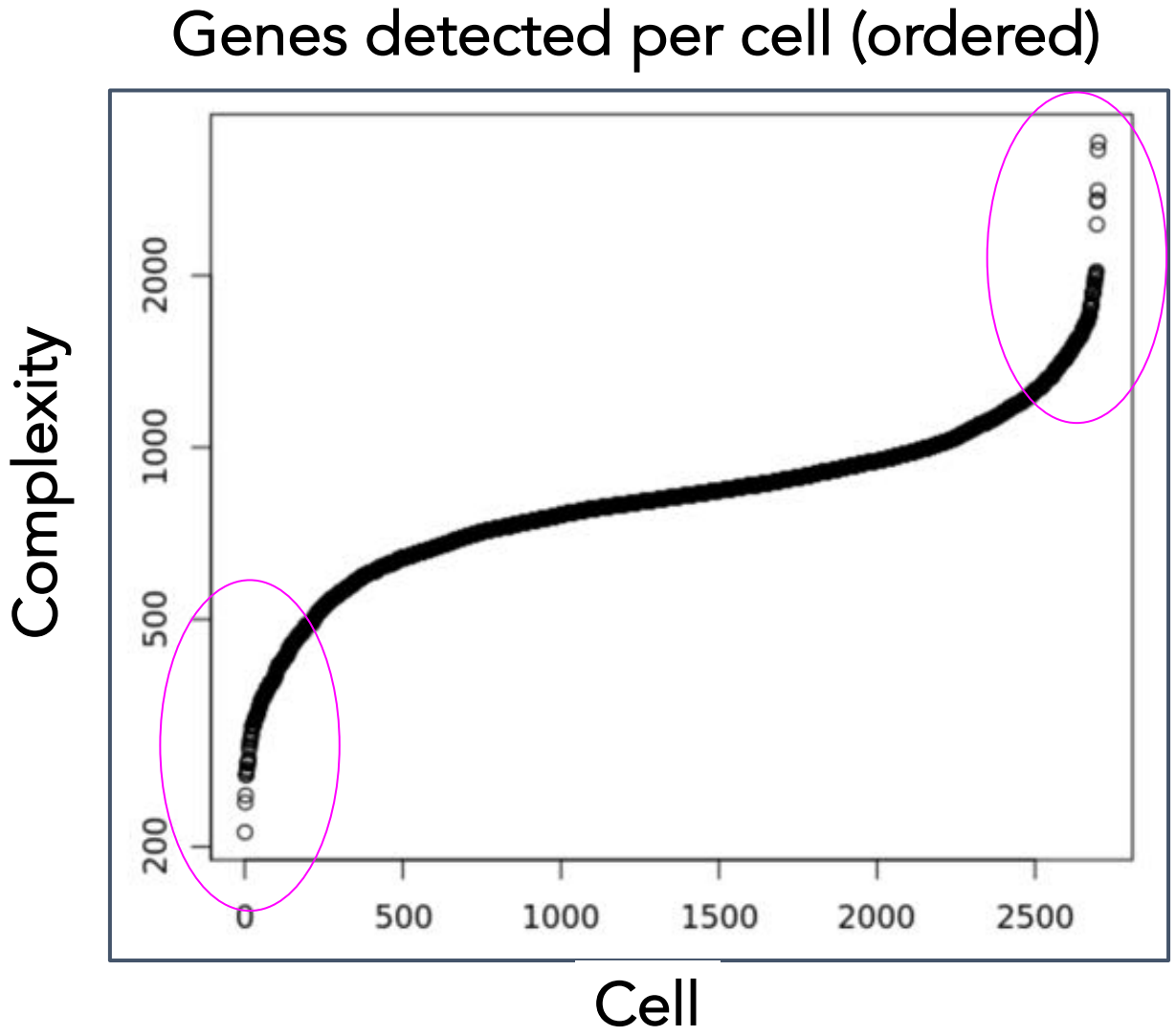
- **Drop out**: data has an excessive amount of zeros due to limiting mRNA

  Zero expression doesn't mean the gene isn't on



There are many quality control filters for genes and cells

Complexity =
Number of genes
detected in a cell



Some unique features and challenges of single cell RNA-Seq

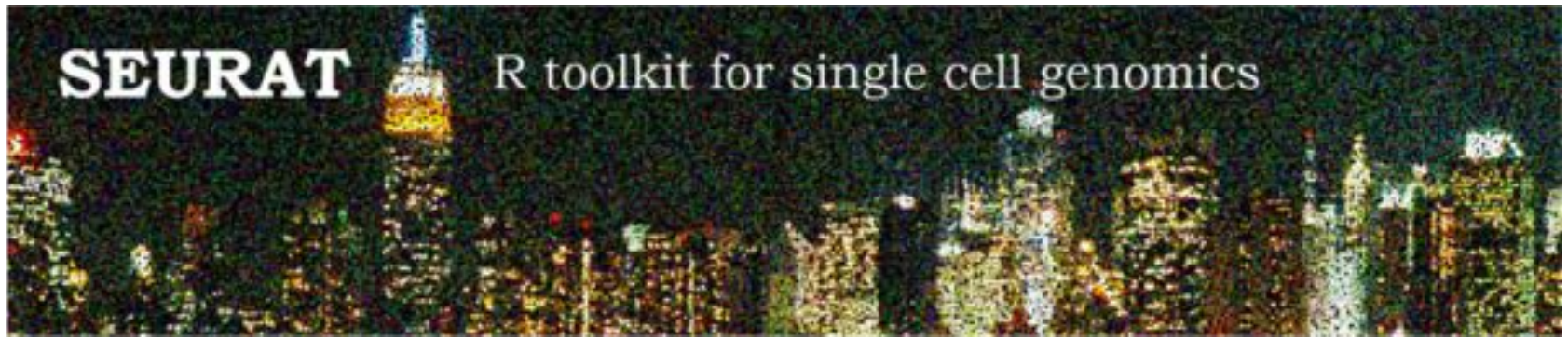
Features

- measures the distribution of expression levels for each gene across a population of single cells
- can study 1e2-1e6 cells in an experiment

Challenges

- amplification during library preparation
- gene dropout
- experimental design and computational analysis

Seurat and scanpy: single cell analysis toolkits

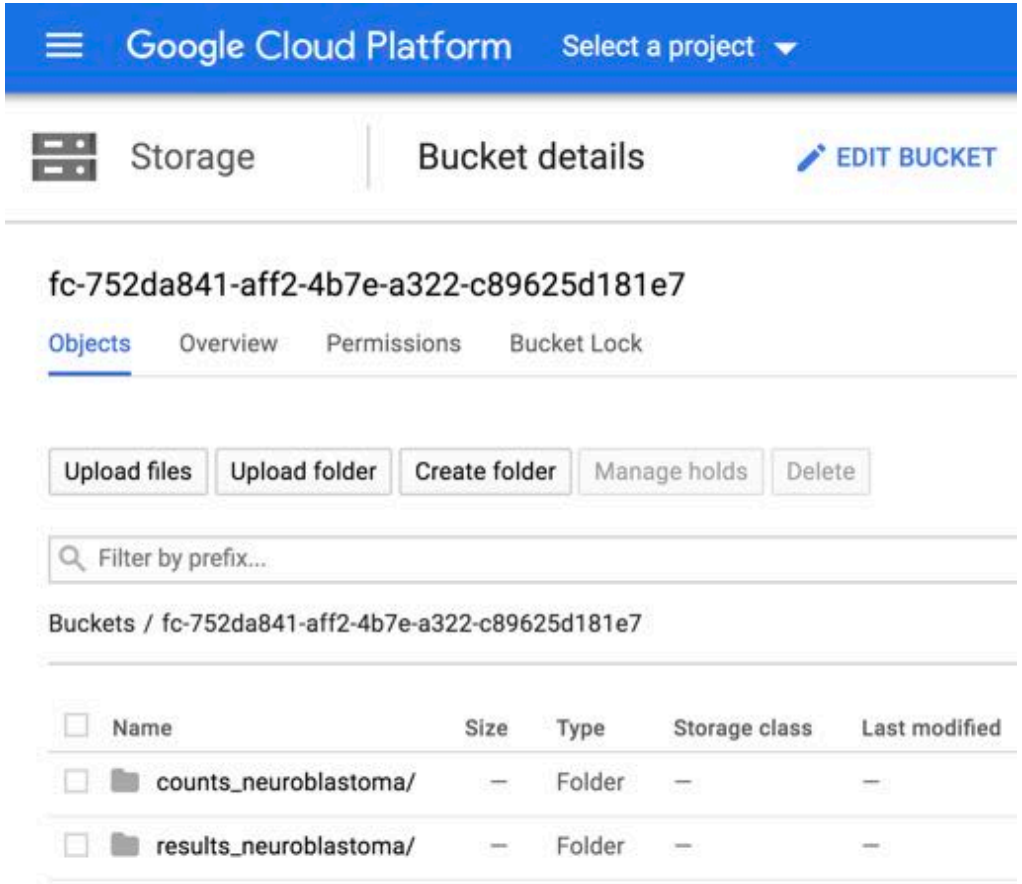


scanpy

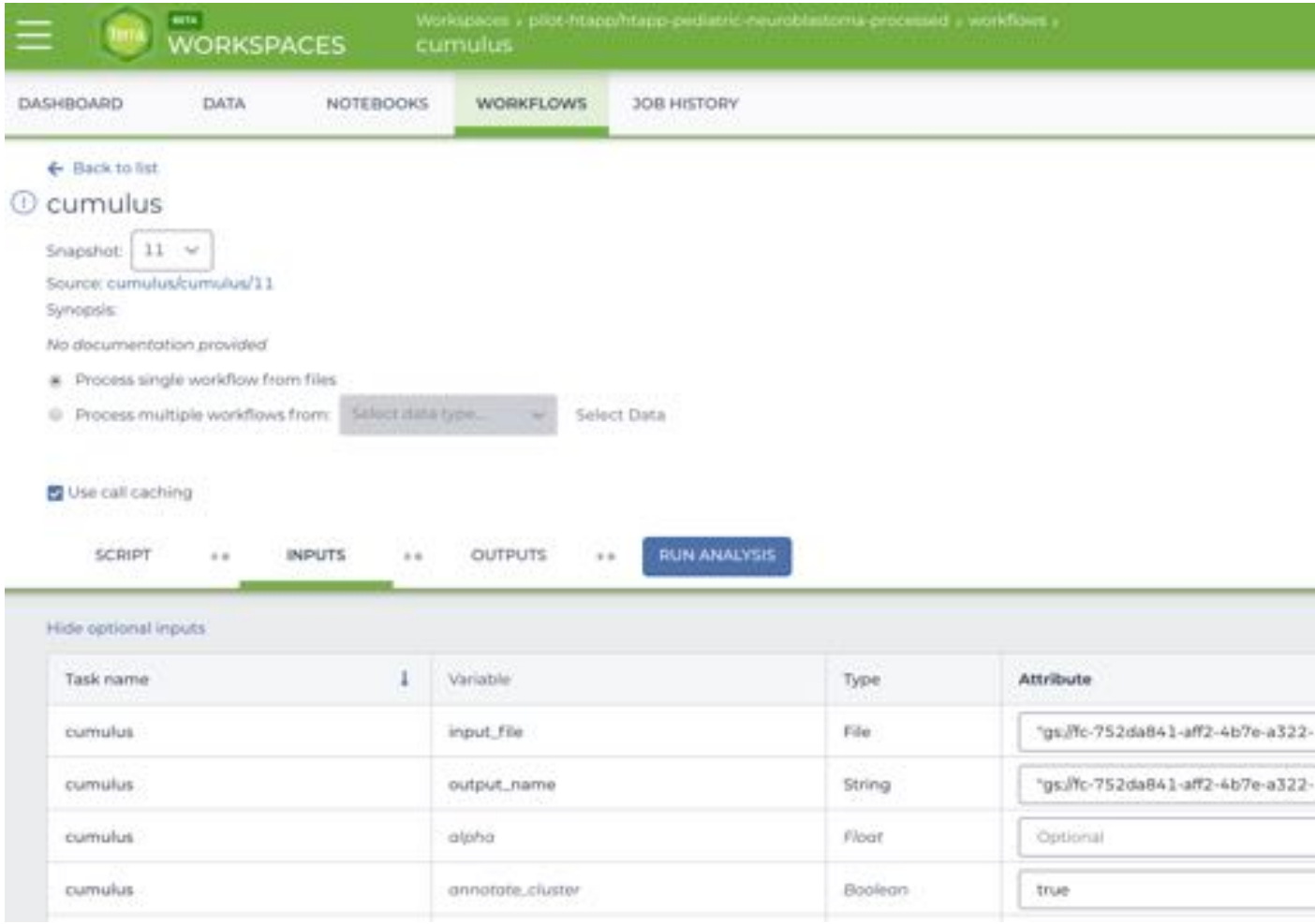
https://satijalab.org/seurat/
https://scanpy.readthedocs.io/en/stable/

Single-cell analysis computations in the Cloud

Inputs and outputs are stored in Google Cloud storage buckets

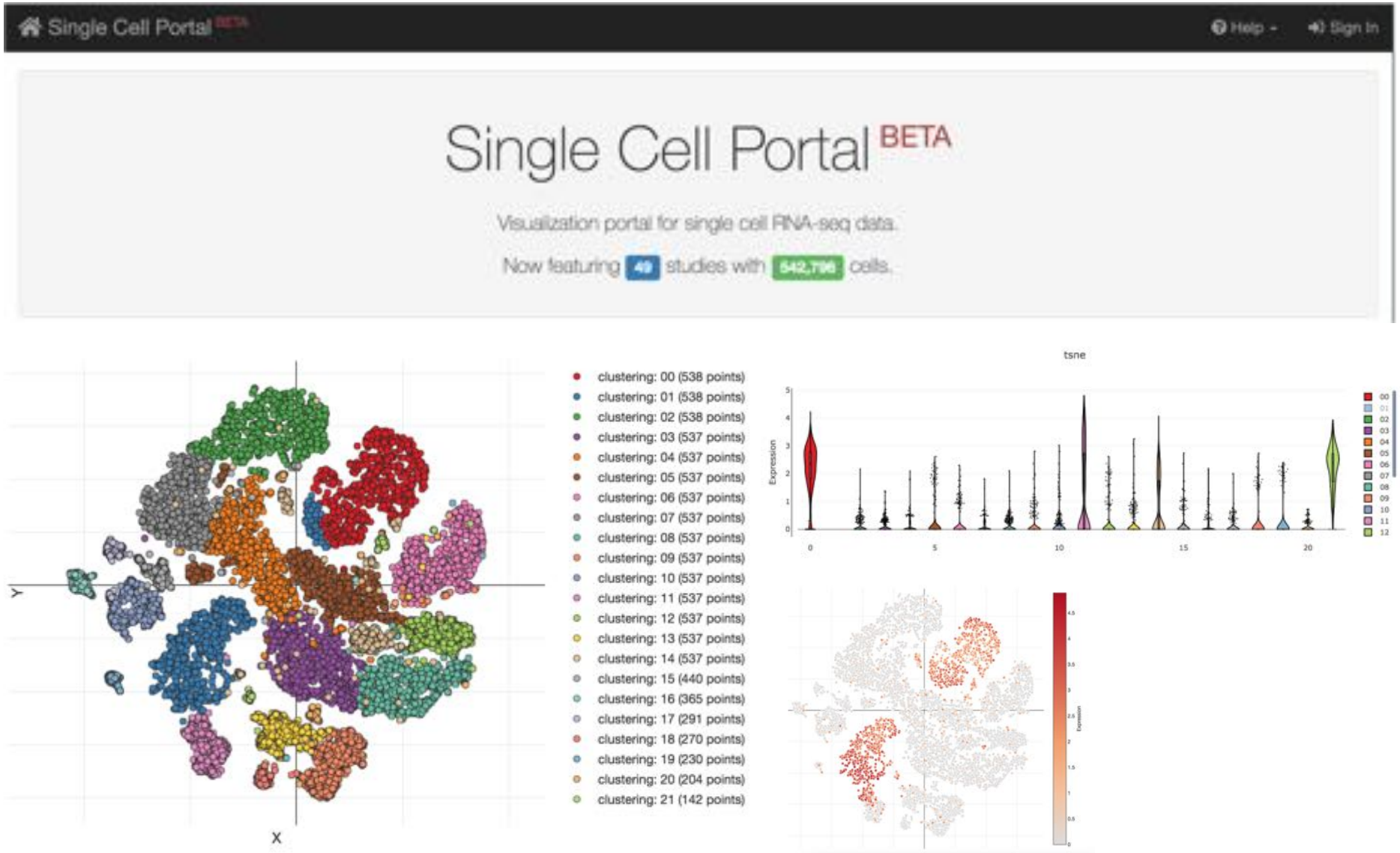


Data processing and analysis are performed in Terra



https://cumulus.readthedocs.io/en/latest/index.html

Single Cell Portal facilitates sharing of single-cell studies



A few single-cell resources

Comprehensive list of single-cell resources
https://github.com/seandavi/awesome-single-cell
https://www.scrna-tools.org/

Computational packages for single-cell analysis
https://satijalab.org/seurat/
https://scanpy.readthedocs.io/
http://bioconductor.org/packages/devel/workflows/html/simpleSingleCell.html

eLife Commentary on the Human Cell Atlas https://elifesciences.org/articles/27041

Nature Commentary on the Human Cell Atlas

https://www.nature.com/news/the-human-cell-atlas-from-vision-to-reality-1.22854

Online courses
https://scrnaseq-course.cog.sanger.ac.uk/website/index.html
https://osca.bioconductor.org/

Single cell data repositories
http://jinglebells.bgu.ac.il/
www.nxn.se/single-cell-studies/