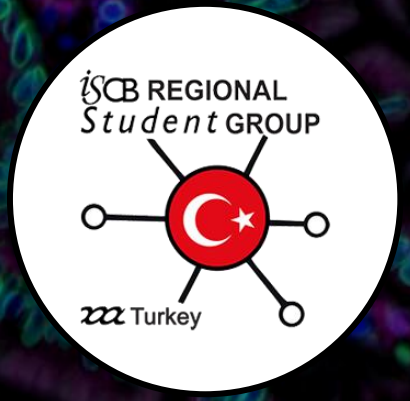
Introduction to Single-cell RNA sequencing (scRNA-seq)

Single-cell Workshop

19-20 December 2020



2013
METHOD
of the YEAR

Single-cell sequencing

nature methods

Editorial, N. M. "Method of the year 2013." Nat Methods 11 (2014): 1.

2018
BREAKTHROUGH
of the YEAR

"The single-cell revolution is just starting."

Development cell by cell

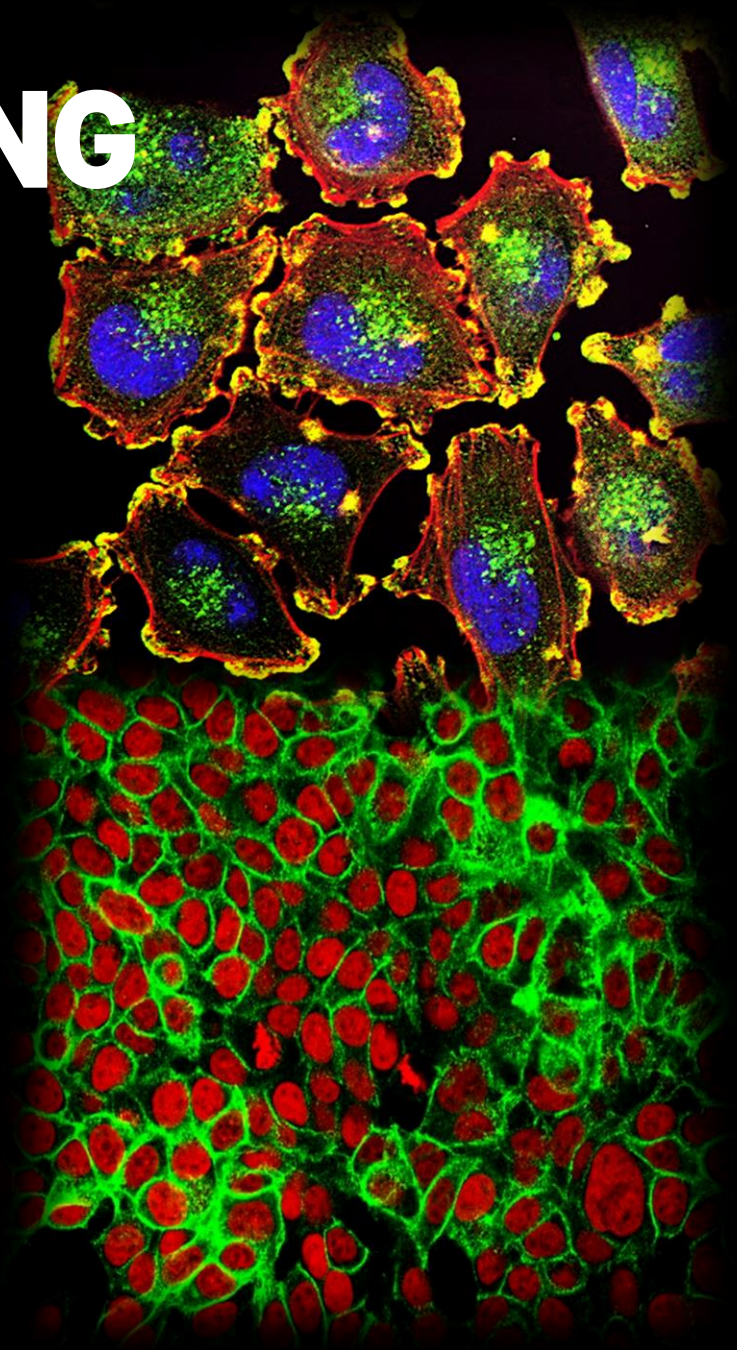
Science

Pennisi, Elizabeth. "Development cell by cell." (2018): 1344-1345.

SINGLE-CELL SEQUENCING

- examines the sequence information from individual cells with optimized NGS technologies.

- higher resolution of cellular differences
- better understanding of the function of an individual cell



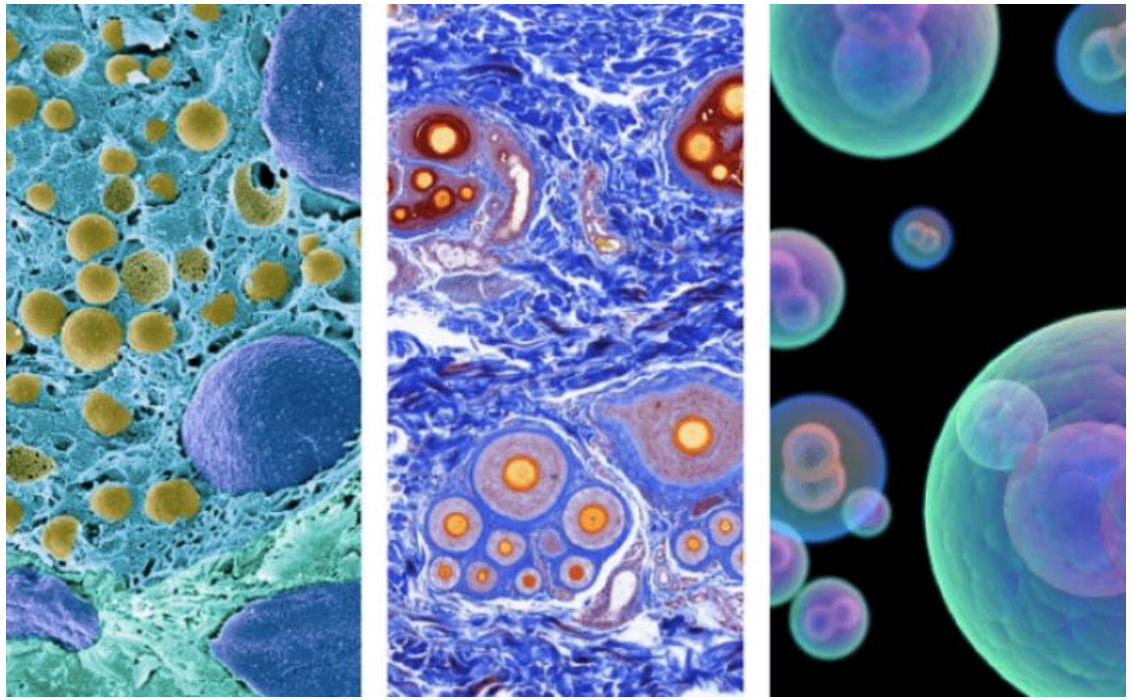
Photos by National Cancer Institute on Unsplash

Bulk vs single-cell

HUMAN CELL ATLAS

MISSION

To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.



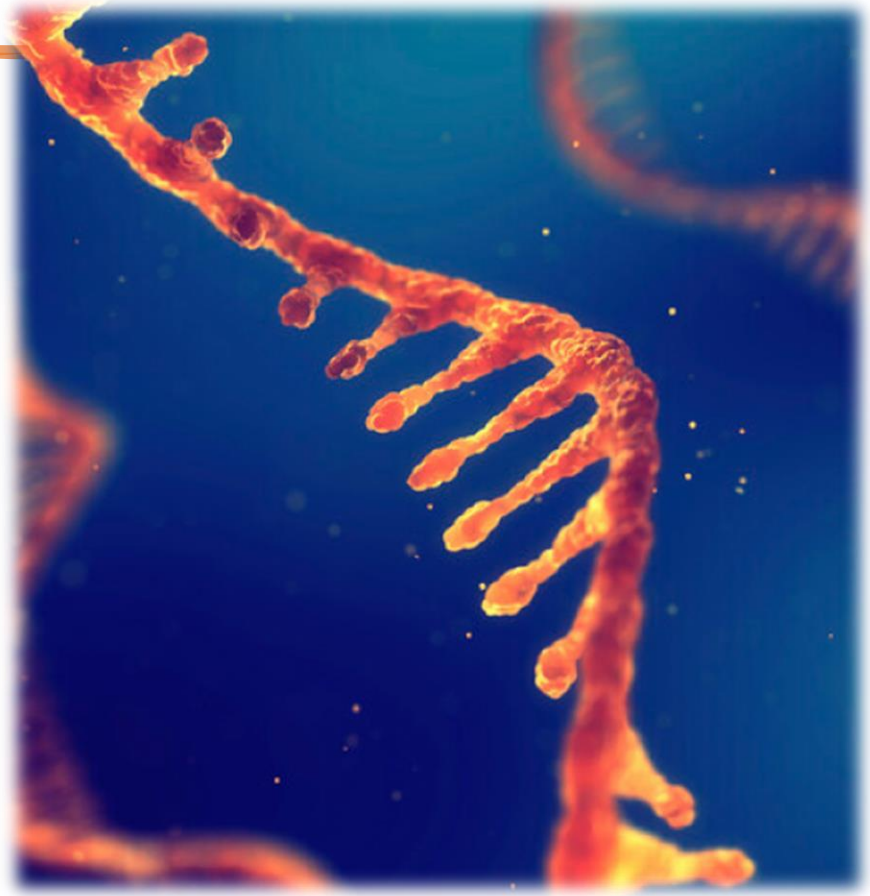
More specifically, a human cell atlas could:

- catalog all cell types and sub-types in the human body;
- map cell types to their location within tissues and within the body;
- distinguish cell states;
- capture the key characteristics of cells during transitions, such as activation or differentiation;
- trace the history of cells through a lineage

Transcriptomics

Transcriptome:
- *the sum of all of the RNA transcripts*
- *mostly refers to mRNAs*

- The study of the transcriptome using high-throughput techniques.
- A snapshot in time of the total transcripts present in a cell.

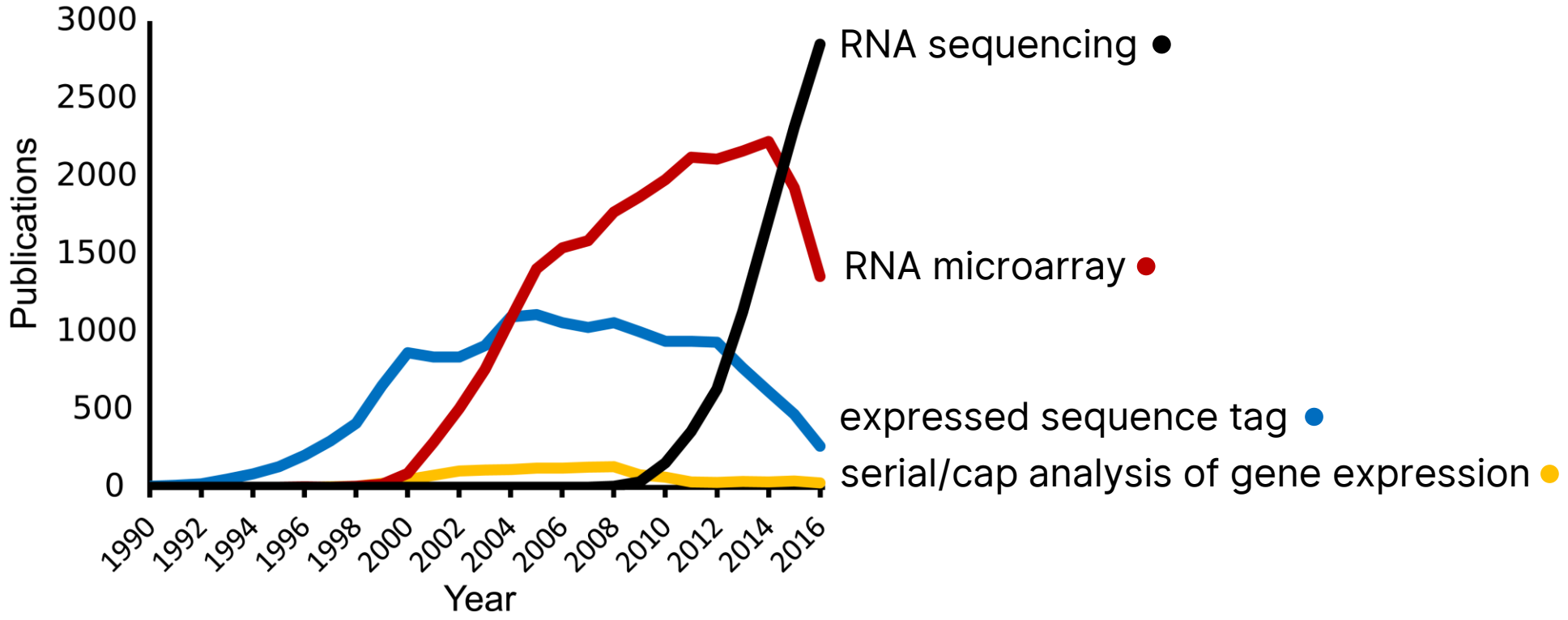


Main aims of transcriptomics

- to classify all types of transcripts (mRNAs, ncRNAs, sRNAs...)
- to determine the transcriptional structure of genes (e.g. splicing patterns and post-transcriptional modifications)
- to quantify the changing expression levels of each transcript in different organisms/conditions/timepoints

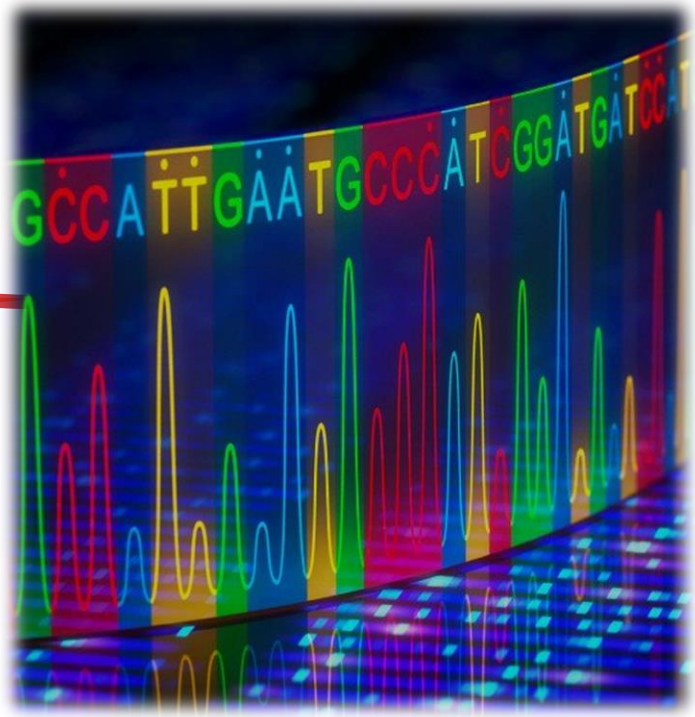
Lowe, Rohan, et al. "Transcriptomics technologies." PLoS computational biology 13.5 (2017): e1005457.

Transcriptomics method use over time



Lowe, Rohan, et al. "Transcriptomics technologies." *PLoS computational biology* 13.5 (2017): e1005457.

RNA sequencing (RNA-seq)



- A technique to measure the quantity of RNA molecules by using next-generation sequencing (NGS) technology.

Advantages

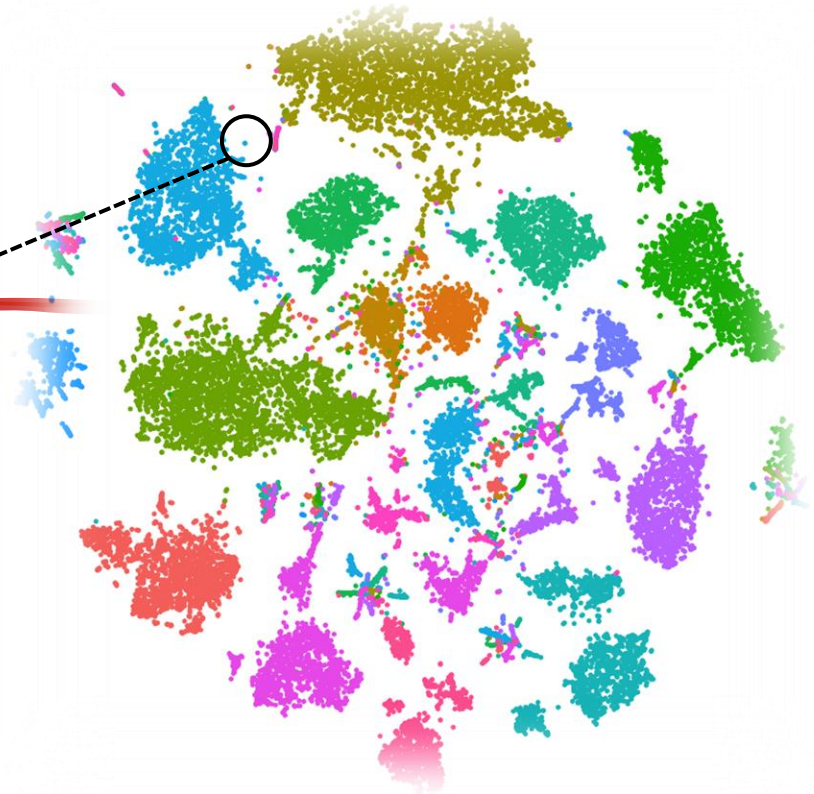
- wider dynamic range,
- more accuracy,
- higher sensitivity,
- ability to discover SNPs and rare mutations, previously unrecognized gene isoforms, microbial RNAs, and regulatory micro-RNAs

Disadvantages

- biases introduced during cDNA library construction and sequence alignment,
- lack of standardization between sequencing platforms and read depth,
- higher start-up costs

Whitley, Sarah K., William T. Horne, and Jay K. Kolls. "Research techniques made simple: methodology and clinical applications of RNA sequencing." *Journal of Investigative Dermatology* 136.8 (2016): e77-e82.

Single-cell RNA sequencing (scRNA-seq)



- A promising and highly trending approach to study the transcriptomes of individual cells.

Advantages

- can detect heterogeneity,
- can distinguish a small number of cells,
- can portray cell maps

Disadvantages

- bias of transcript coverage,
- low capture efficiency,
- higher level of technical noise,
- dropout events

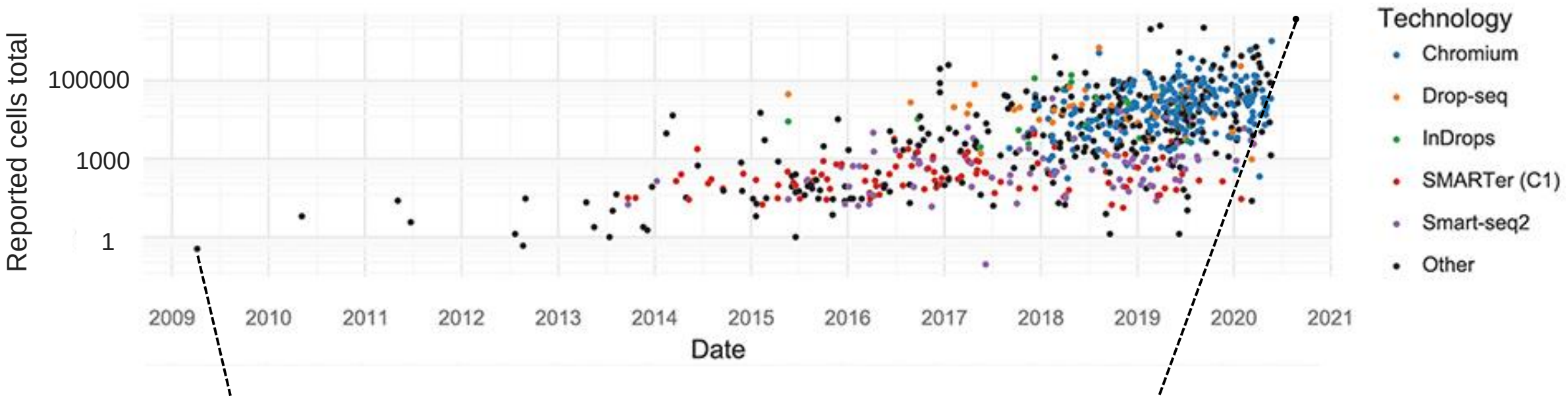
Common Applications

- describing population diversity
- tracing cell lineages
- classifying cell types
- genomic profiling of rare cells

Wang, Yong, and Nicholas E. Navin. "Advances and applications of single-cell sequencing technologies." Molecular cell 58.4 (2015): 598-609.

The number of cells measured in scRNA-seq experiments

according to scRNA-seq protocols



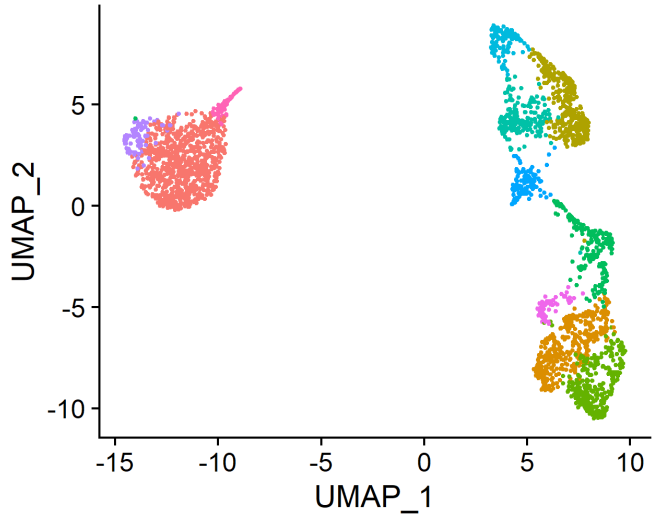
April 2009: 1 cell by Tang et al.

Tang, Fuchou, et al. "mRNA-Seq whole-transcriptome analysis of a single cell." *Nature methods* 6.5 (2009): 377-382.

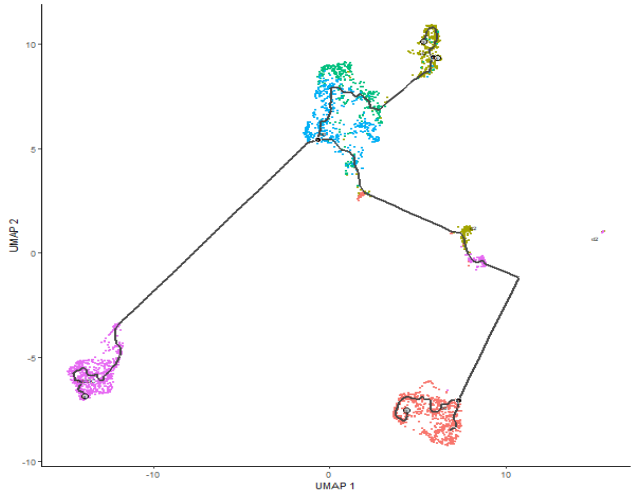
November 2020: ~4 million cells by Shendure Lab

Cao, Junyue, et al. "A human cell atlas of fetal gene expression." Science 370.6518 (2020).

Svensson, Valentine, Eduardo da Veiga Beltrame, and Lior Pachter. "A curated database reveals trends in single-cell transcriptomics." *Database* 2020 (2020).



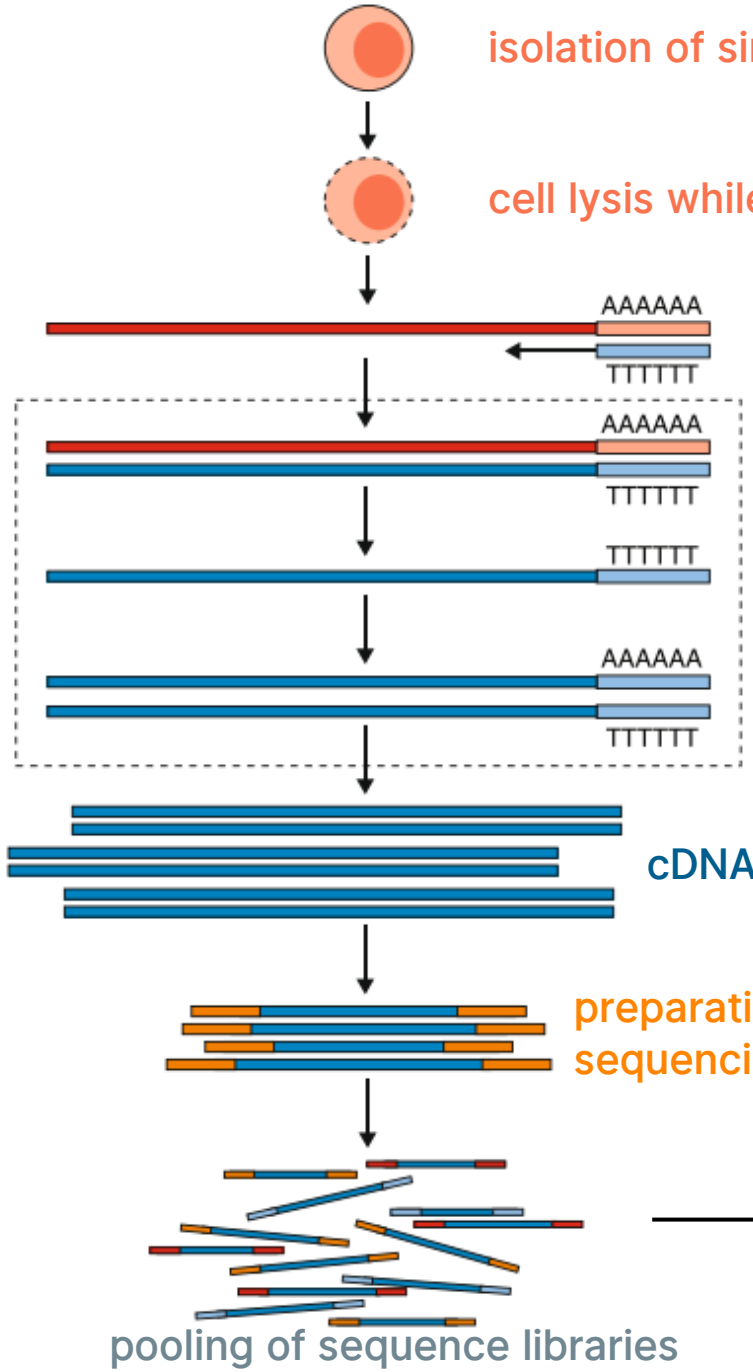
Why to perform scRNA-seq?



- to assess transcriptional similarities and differences within a population of cells
  - heterogeneity analysis
  - identification of rare cell populations
- to examine the single cells where each one is unique
  - individual T lymphocytes expressing highly diverse T-cell receptors
  - neurons within the brain
  - cells within an early-stage embryo
- to trace lineage and developmental relationships between heterogeneous cellular states
  - embryonal development
  - differentiation of cancer cells, myoblasts or stem cells

Haque, Ashraful, et al. "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications." *Genome medicine* 9.1 (2017): 1-12.

General workflow of scRNA-seq experiments

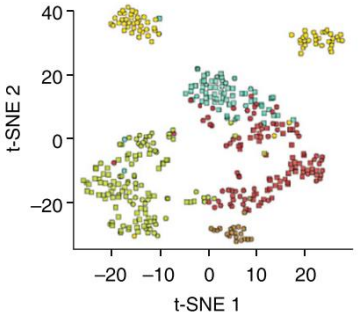


sequencing library

ATACGATAATTCCGA
CATTGATATGTCTAAT
GCCTTACAATCTTT
ATACGAGCAAAGGAA
GCCTTACTAATTATA
CATTGAGATTGGGTA

use of bioinformatic tools to assess quality and variability

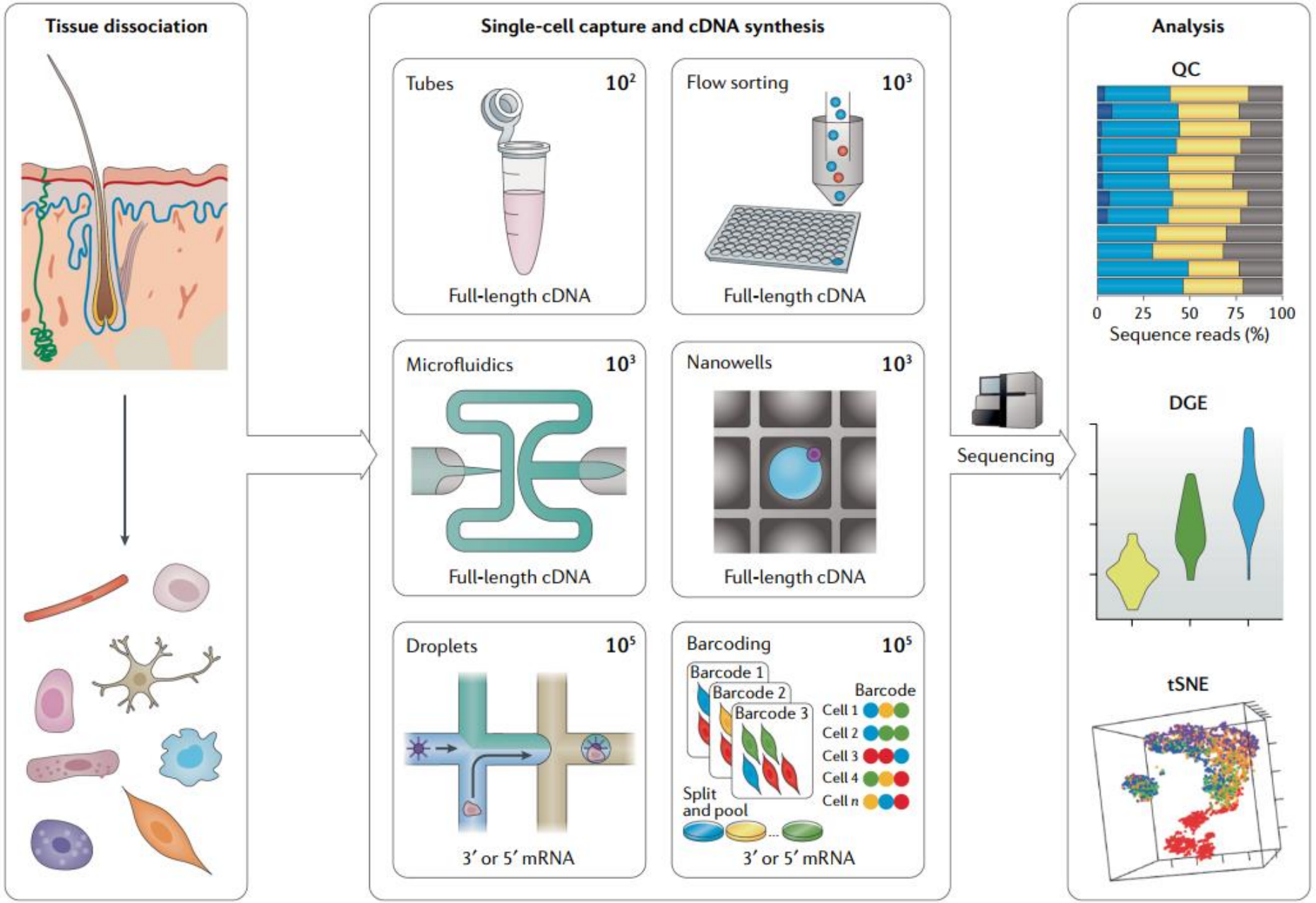
use of specialized tools to analyse and present the data



COUNT MATRIX

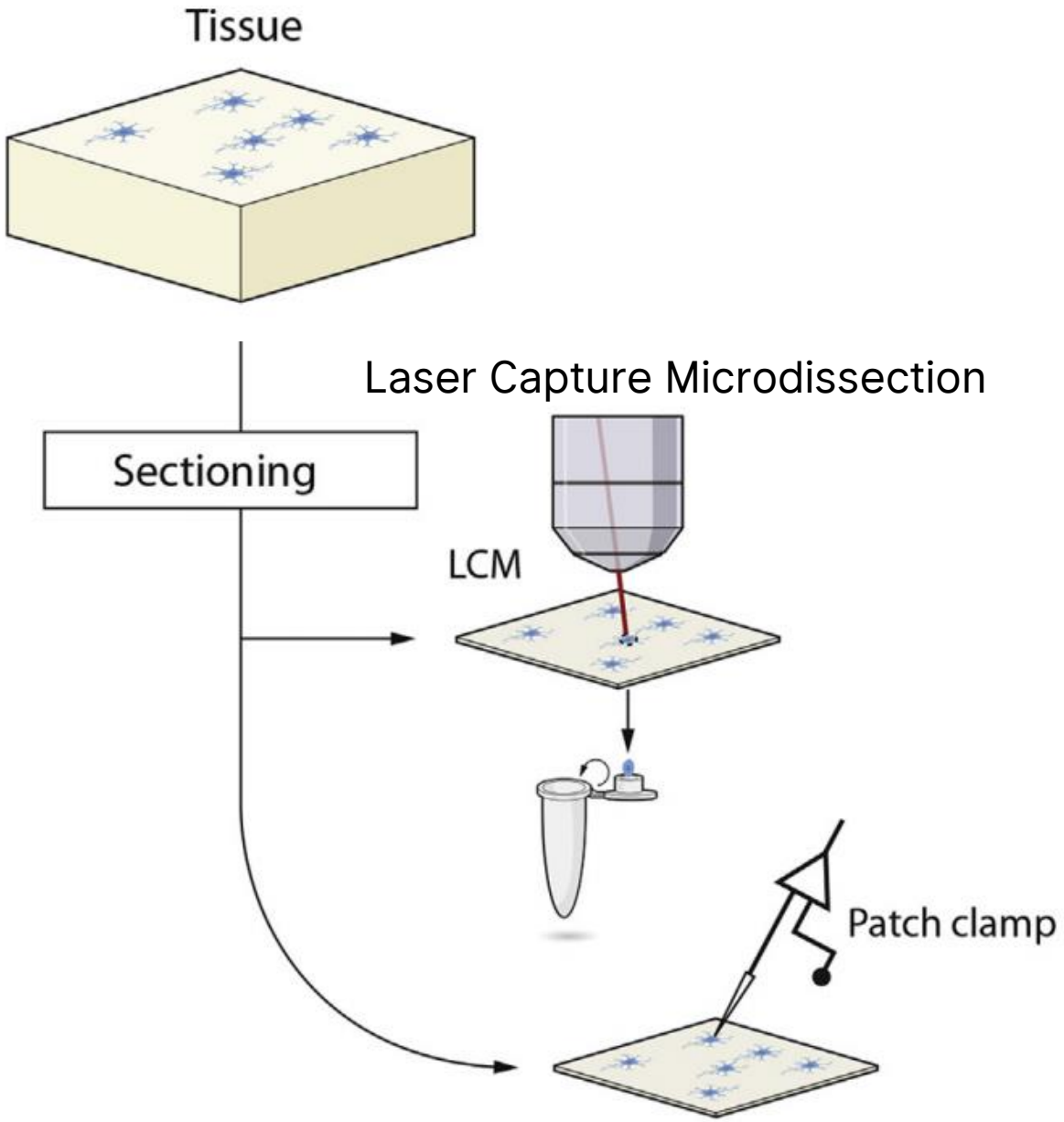
	Cell1	Cell2	...	CellN
Gene1	3	2	.	13
Gene2	2	3	.	1
Gene3	1	14	.	18
...	.	.	.	.
...	.	.	.	.
...	.	.	.	.
GeneM	25	0	.	0

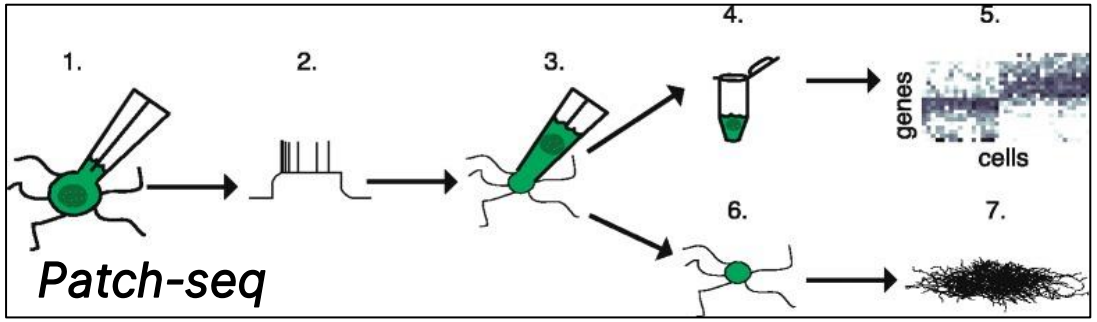
Haque, Ashraful, et al. "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications." *Genome medicine* 9.1 (2017): 1-12.



Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. Nat Rev Genet 20, 631–656 (2019). https://doi.org/10.1038/s41576-019-0150-2

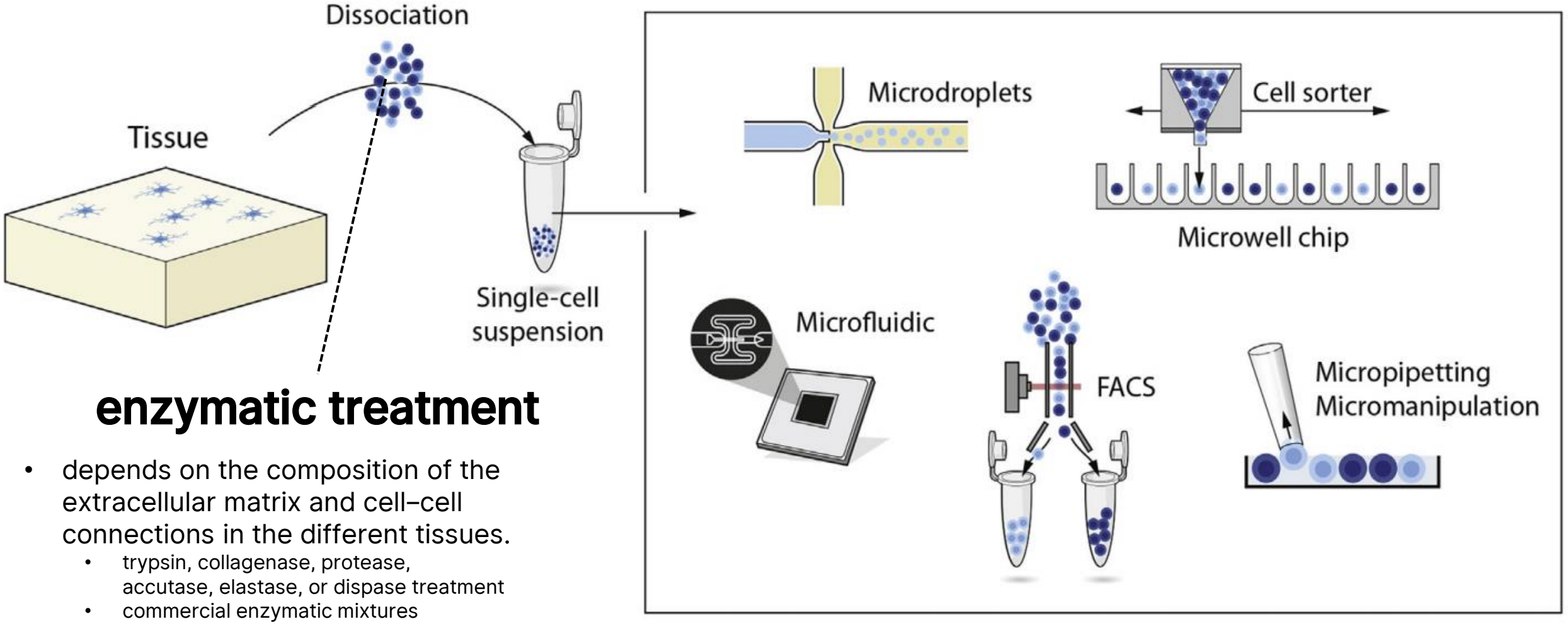
Isolation of single cells





- an unbiased strategy to characterize and classify cell types
  - by integrating information about each cell's morphology, physiology, and gene expression into a common framework.
- can be used as a complementary method to 'annotate' cell type classification based primarily on scRNA-seq of dissociated neurons
  - 30–40 Patch-seq samples per day with 2-3 people
    - ~$21/cell

Hedlund, Eva, and Qiaolin Deng. "Single-cell RNA sequencing: technical advancements and biological applications." *Molecular aspects of medicine* 59 (2018): 36-46.

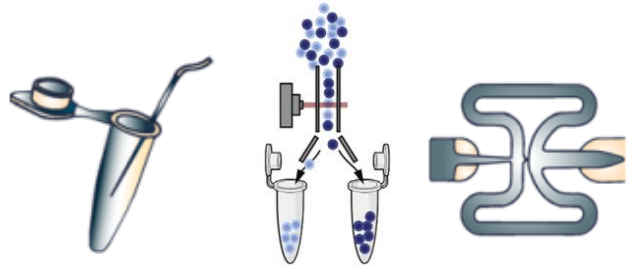


enzymatic treatment

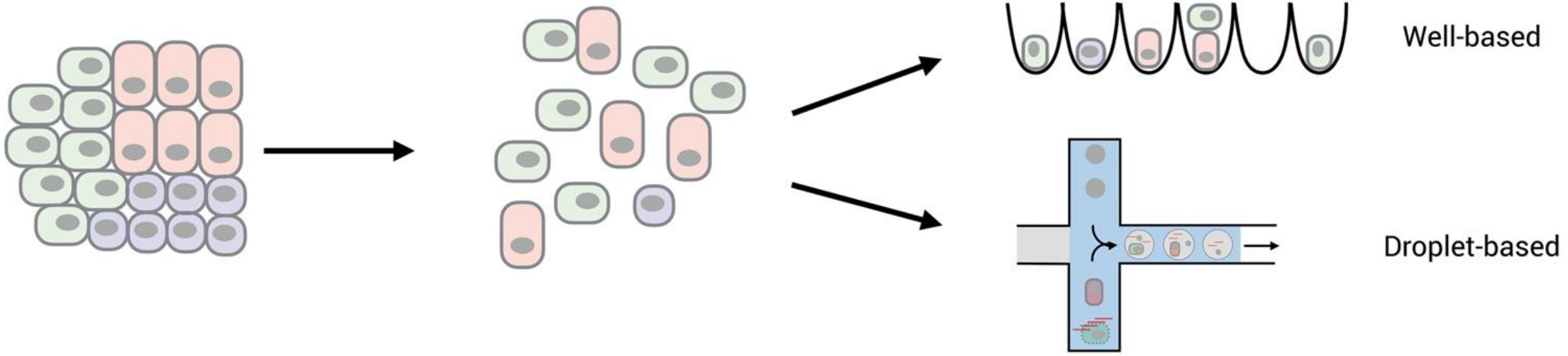
- depends on the composition of the extracellular matrix and cell–cell connections in the different tissues.
  - trypsin, collagenase, protease, accutase, elastase, or dispase treatment
  - commercial enzymatic mixtures (TrypLE Express and Liberase Blendzyme 3)
- cellular stressor for mature neurons
  - can degrade surface proteins
  - may alter gene expression.

Isolation of single cells

Hedlund, Eva, and Qiaolin Deng. "Single-cell RNA sequencing: technical advancements and biological applications." *Molecular aspects of medicine* 59 (2018): 36-46.



more controlled, more flexible,
lower-throughput, more expensive

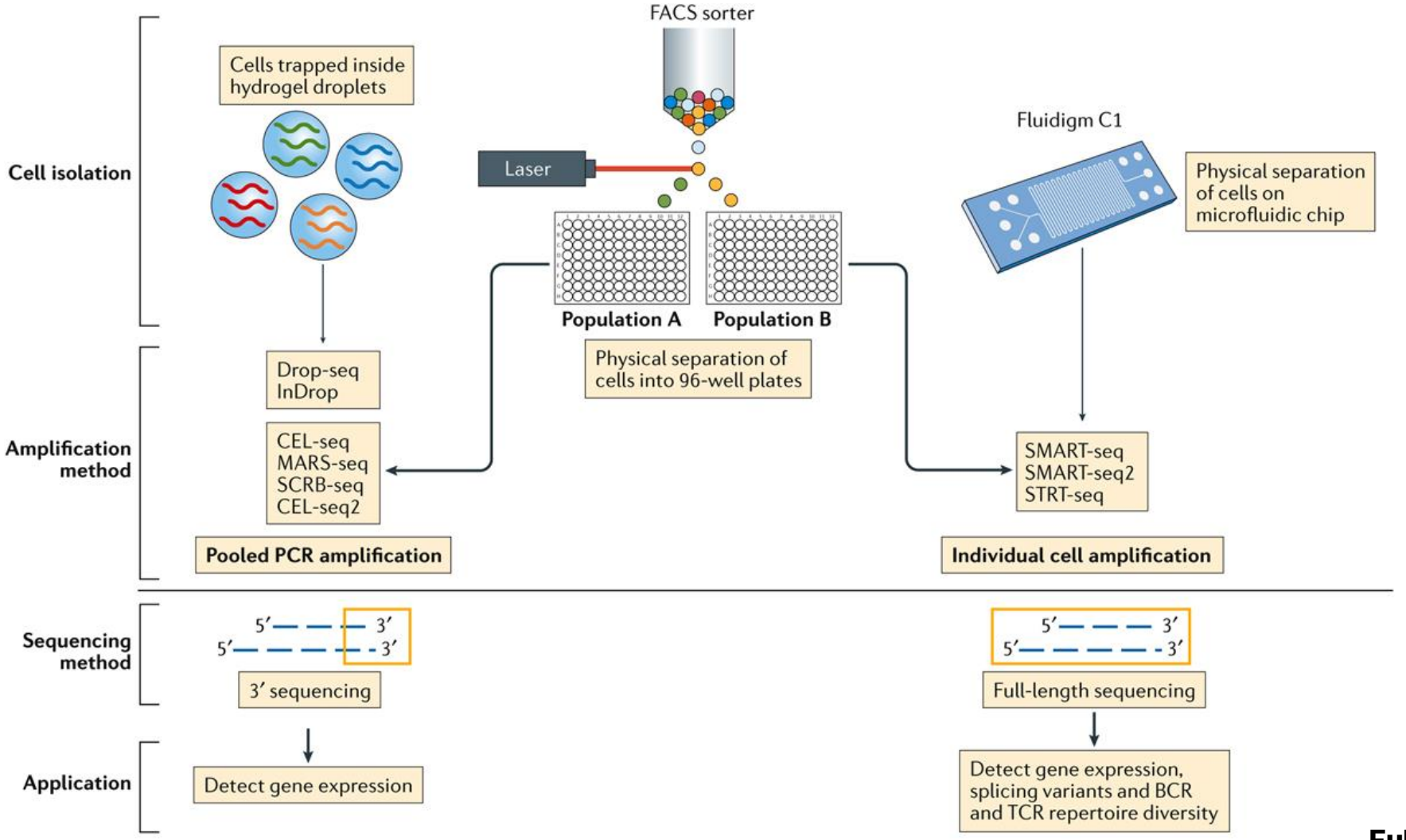


higher possibility of doublets,
efficient, high-throughput, cheaper

Isolation of single cells

Ziegenhain, Christoph, et al. "Quantitative single-cell transcriptomics." Briefings in functional genomics 17.4 (2018): 220-232.

Single-cell Methods

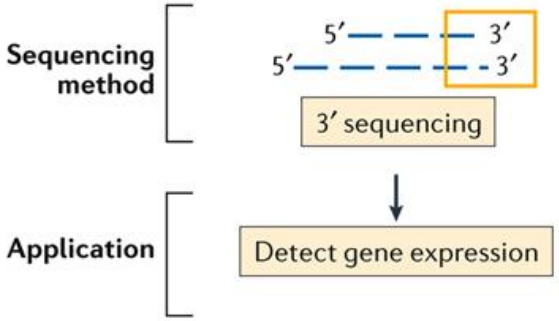


Tag-based (3' (*or 5'*)-end) sequencing:

Full-length sequencing

Nature Reviews | Immunology

Papalexi, Efthymia, and Rahul Satija. "Single-cell RNA sequencing to explore immune cell heterogeneity." *Nature Reviews Immunology* 18.1 (2018): 35.



Tag-based (3' (*or 5'*)-end) sequencing:

- **cell-specific barcodes:** non-random nucleotide sequences
  - determines which cell the read originated from
  - allowing for multiplexing the cDNA amplification
  - increasing the throughput of scRNA-seq library generation by one to three orders of magnitude

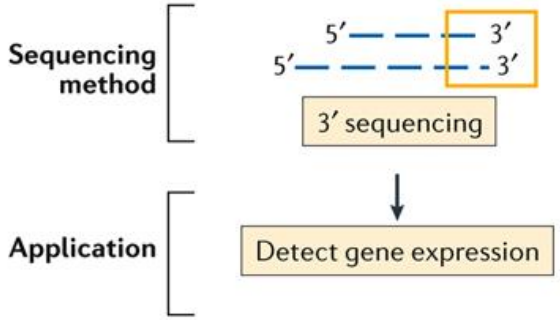
- **unique molecular identifiers (UMIs):** random nucleotide sequences
  - tag individual mRNA molecules
  - allow for the distinction between original molecules and amplification duplicates that derive from the cDNA or library amplification
  - allow more accurate quantification
  - used to collapse PCR duplicates

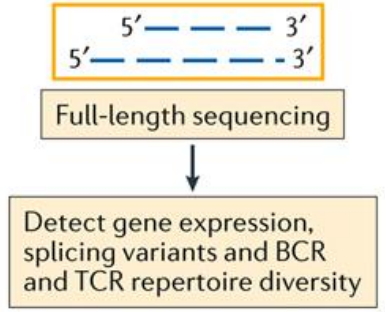
- **Barcode:**
  identifies independent cells

- **UMI:**
  identifies independent molecules

https://github.com/hbctraining/scRNA-seq

Papalexi, Efthymia, and Rahul Satija. "Single-cell RNA sequencing to explore immune cell heterogeneity." *Nature Reviews Immunology* 18.1 (2018): 35.
Hedlund, Eva, and Qiaolin Deng. "Single-cell RNA sequencing: technical advancements and biological applications." *Molecular aspects of medicine* 59 (2018): 36-46.





Tag-based (3' (*or 5'*)-end) sequencing:

- enables the incorporation of cell-specific barcodes and **unique molecular identifiers (UMIs)**
- Larger number of cells sequenced allows better identity of cell type populations
- Cheaper per cell cost
- Best results with > 10,000 cells

- **ERCC spike-ins:** broad concentration of synthetic RNAs of various lengths and GC content,
  - often used to measure biases in RNA-seq experiments and for quality control or normalization.
  - to better account for technical variability due to random dropout events during library preparation.

Full-length sequencing:

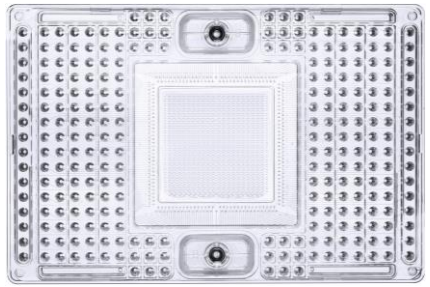
- tries to achieve a uniform gene body read coverage and increases the number of mappable reads.
  - detection of isoform-level differences
  - identification of allele-specific differences
- deeper sequencing of a smaller number of cells
- best for samples with low number of cells
- not possible to multiplex and pool all samples for one tube preparation of Illumina sequenceable libraries
- more expensive and troublesome
- not compatible with high-throughput methods
- no UMIs

https://github.com/hbctraining/scRNA-seq

Papalexi, Efthymia, and Rahul Satija. "Single-cell RNA sequencing to explore immune cell heterogeneity." *Nature Reviews Immunology* 18.1 (2018): 35.
Hedlund, Eva, and Qiaolin Deng. "Single-cell RNA sequencing: technical advancements and biological applications." *Molecular aspects of medicine* 59 (2018): 36-46.

FLUIDIGM®

Fluidigm C1 System



- Developed in early 2000s
- 96/800 cells at once

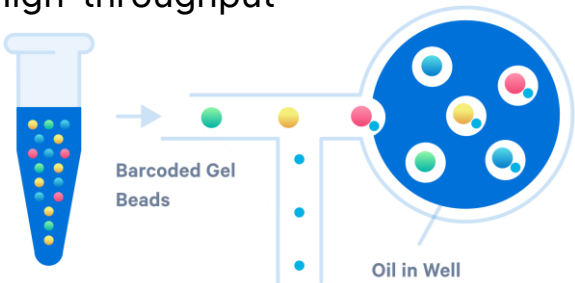
- difficult to integrate with experiments that involve long processing times
- high cost of the microfluidic cartridges

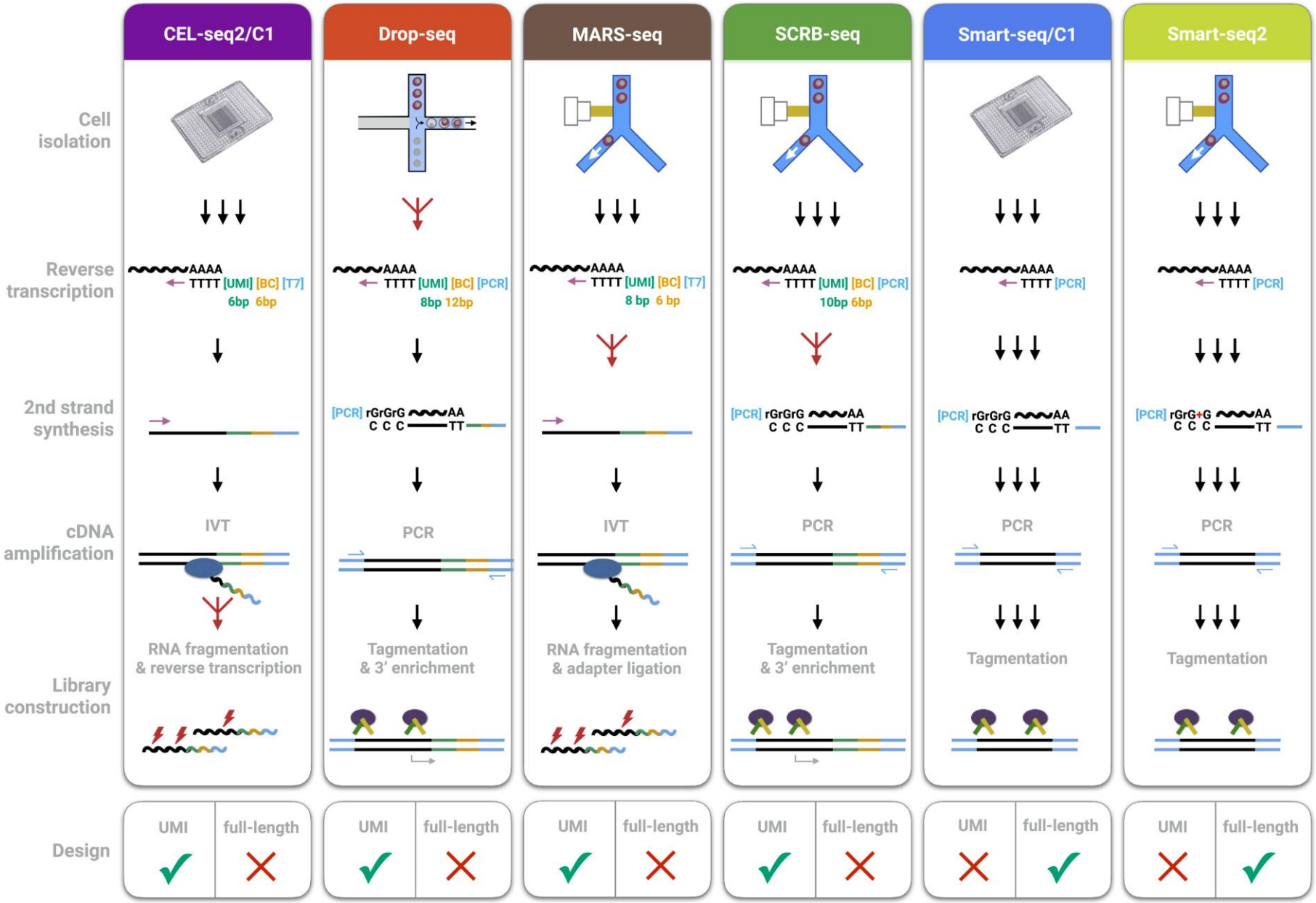
10x GENOMICS®

Chromium Single Cell Gene Expression Solution

- Developed in 2016.
- 1–8 samples in one run,
  - 10,000 cells per sample.

- High-throughput



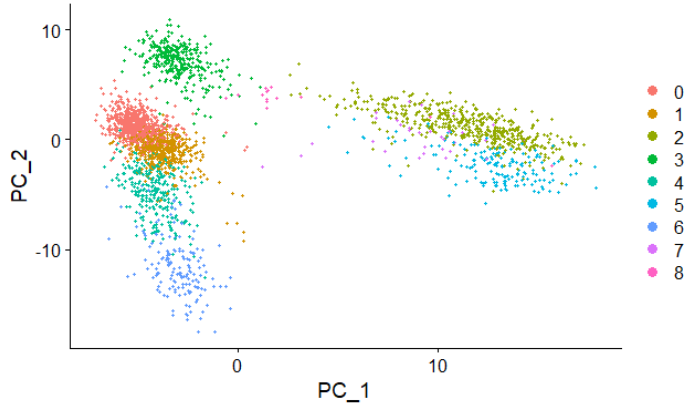


Ziegenhain, Christoph, et al. "Comparative analysis of single-cell RNA sequencing methods." *Molecular cell* 65.4 (2017): 631-643.

Methods	Transcript coverage	UMI possibility	Strand specific	References
Tang method	Nearly full-length	No	No	Tang et al., 2009
Quartz-Seq	Full-length	No	No	Sasagawa et al., 2013
SUPeR-seq	Full-length	No	No	Fan X. et al., 2015
Smart-seq	Full-length	No	No	Ramskold et al., 2012
Smart-seq2	Full-length	No	No	Picelli et al., 2013
MATQ-seq	Full-length	Yes	Yes	Sheng et al., 2017
STRT-seq and STRT/C1	5′-only	Yes	Yes	Islam et al., 2011, 2012
CEL-seq	3′-only	Yes	Yes	Hashimshony et al., 2012
CEL-seq2	3′-only	Yes	Yes	Hashimshony et al., 2016
MARS-seq	3′-only	Yes	Yes	Jaitin et al., 2014
CytoSeq	3′-only	Yes	Yes	Fan H.C. et al., 2015
Drop-seq	3′-only	Yes	Yes	Macosko et al., 2015
InDrop	3′-only	Yes	Yes	Klein et al., 2015
Chromium	3′-only	Yes	Yes	Zheng et al., 2017
SPLiT-seq	3′-only	Yes	Yes	Rosenberg et al., 2018
sci-RNA-seq	3′-only	Yes	Yes	Cao et al., 2017
Seq-Well	3′-only	Yes	Yes	Gierahn et al., 2017
DroNC-seq	3′-only	Yes	Yes	Habib et al., 2017
Quartz-Seq2	3′-only	Yes	Yes	Sasagawa et al., 2018

Chen, Geng, Baitang Ning, and Tieliu Shi. "Single-cell RNA-seq technologies and related computational data analysis." Frontiers in genetics 10 (2019): 317.

Dimension Reduction Methods for scRNA-seq

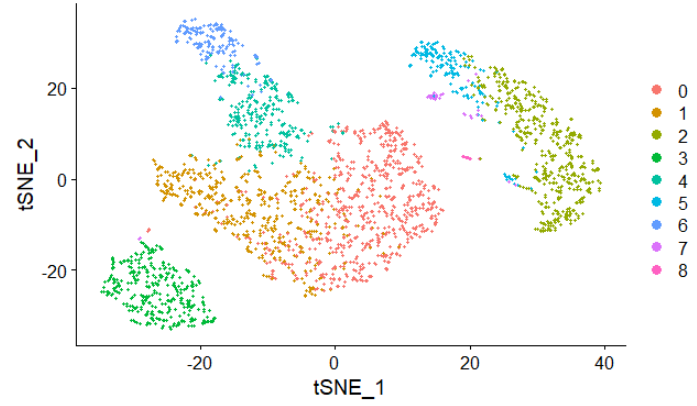


PCA (Principal Component Analysis) **LINEAR**

- performs an orthogonal transformation of the original dataset to create a set of new, uncorrelated variables or principal components (linear combinations of variables in the original dataset.)

Pros:
- highly interpretable and computationally efficient

Cons:
- very inappropriate due to highly non-linear structure

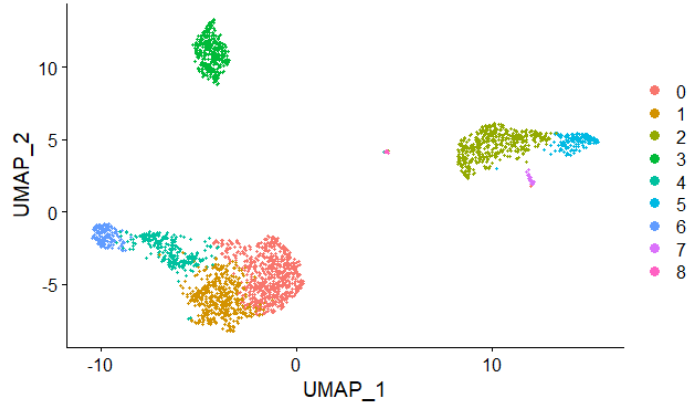


tSNE (t-Stochastic Neighbourhood Embedding) **LINEAR**

- graph based and non-linear technique, projects high dimensional data onto 2D or 3D components

Pros:
- designed to preserve the local structures

Cons:
- stochastic method
- global structure not preserved
- only limited to data exploration or visualisation.
- computationally expensive



UMAP (Uniform Manifold Approximation and Projection) **NON-LINEAR**

- graph based and principally similar to t-SNE where it constructs a high dimensional graph representation of the data, then optimizes a low-dimensional graph to be as structurally similar as possible.

Pros:
- computationally efficient
- better preservation of the global structure
- can be used as a pre-processing step

Cons:
- false detection of noise
- accuracy of global structure

https://towardsdatascience.com/guide-to-dimensionality-reduction-in-single-cell-rna-seq-analysis-1d77284eed1c

END OF THE SESSION

Introduction to
Single-cell RNA sequencing
(scRNA-seq)

Single-cell Workshop

19-20 December 2020