

# Single cell genomics I

Genomics Bio5488

Guoyan Zhao

Associate Professor of Genetics & Neurology



# Class Information

---



- Lectures: Mon, Wed 10:00-11:30 am
  - February 23: Single cell genomics I
  - February 25: Single cell genomics II
- Lab: February 27, 2025, Fri 10:00-11:30 am (Bring your laptop)
- Questions, suggestions and feedbacks: [gzhao@wustl.edu](mailto:gzhao@wustl.edu)

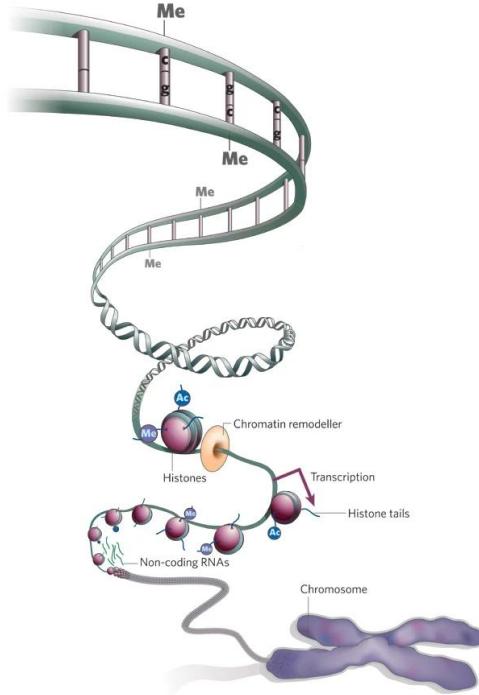


# Single cell genomics

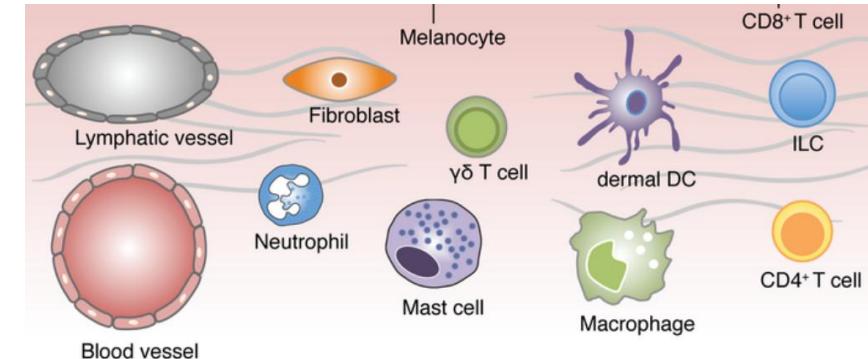
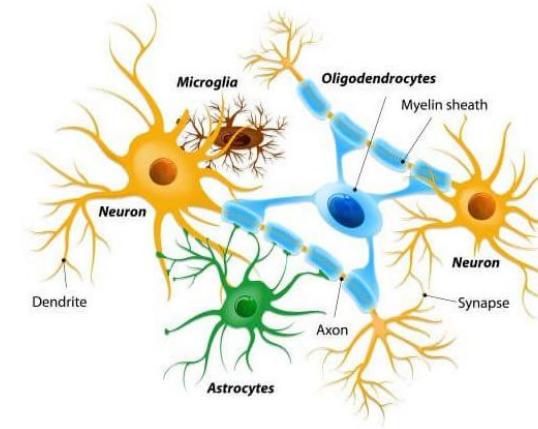
---

- Single cell genomics I
  - History of Single cell technology
  - Single cell RNA-sequencing (scRNA-seq) technology
  - Basic scRNA-seq data analysis workflow
  - Unlocking biological insights
  - Other unimodal single-cell technology
- Single cell genomics II
  - Single-cell multiomics
    - Transcriptome + Epigenome
    - Transcriptome + Protein
    - Transcriptome + CRISPR screening
    - Transcriptome + TCR/BCR
    - Transcriptome + Antigen specificity
  - Spatial genomics
    - Spatial transcriptomics
    - Spatial proteomics
    - Spatial multiomics
    - Spatial metabolomics
- Single cell genomics Lab
  - scRNA-seq data analysis

# Incredible cellular diversity encoded in the genome



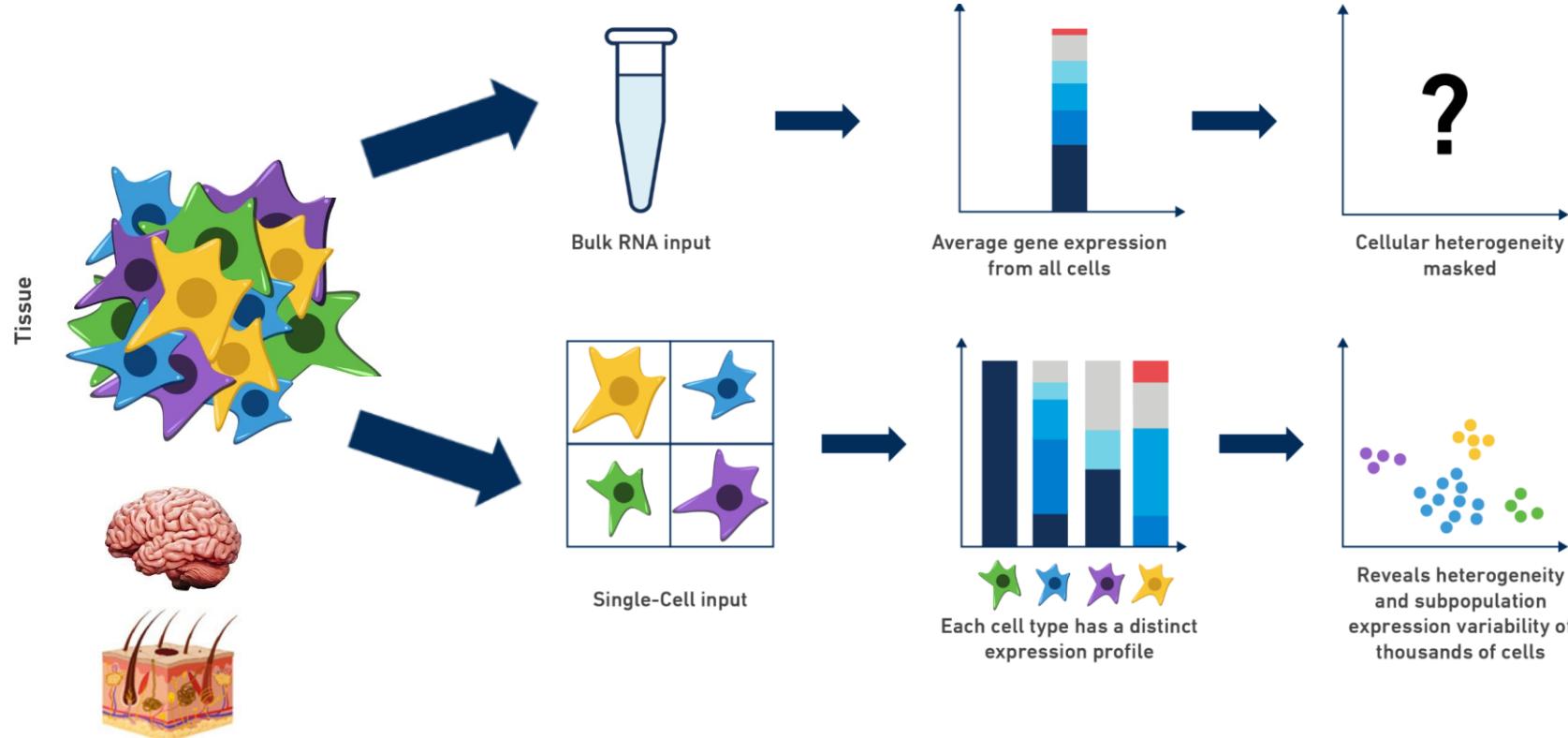
<https://doi.org/10.1038/454711a>



≈ 36 trillion cells in the male, ≈ 28 trillion in the female  
 ~ 400 cell types  
 ~ 1,200 cell groups

The human cell count and size distribution, Ian A. Hatton et al., PNAS 2023

# Bulk sequencing vs. Single cell sequencing



Gene ID	sample1
Gene 1	1
Gene 2	23
Gene 3	539
...	4648
Gene n	2

Gene ID	Cell1	Cell2	Cell3
Gene 1	1	1	0
Gene 2	1	0	0
Gene 3	5	2	0
...	0	0	1
Gene n	2	9	0

This can be hundreds to millions of cells

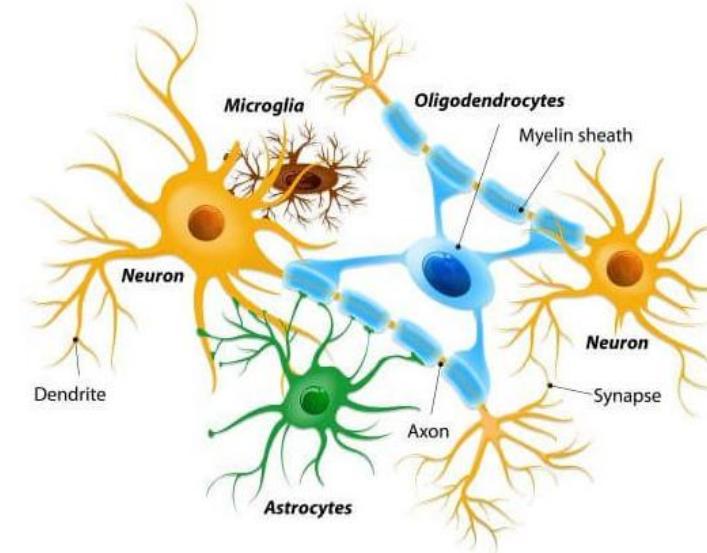
# Advantages of single-cell sequencing over bulk sequencing

---

bulk sequencing



single-cell sequencing



- Obtain genomic information (e.g. gene expression) for every cell
- Understand cell heterogeneity within the tissue
- Distinguish cell population changes vs. gene expression changes



# Application of scRNA-seq data

---

- Resolving cellular heterogeneity
  - Hundreds of new cell types in the brain
  - Reveal different cell states (different activation states of microglia)
- Identify rare cell populations
  - hyper-responsive immune cells within a seemingly homogeneous group
  - Examine cells where each one is essentially unique, such as individual T lymphocytes expressing highly diverse T-cell receptors
- Trace lineage and developmental relationships between heterogeneous, yet related, cellular states
  - embryonal development
  - cancer progression
  - microglia activation
- Mechanism of heterogenous drug response
- .....



# The challenge of single cell technology

---

- The amount of RNA present in a single cell ranges 1–50 pg depending on cell type.
- Successful signal detection with RNA-seq needs 0.1–1.0  $\mu$ g of total RNA
- The application of sequencing-based expression profiling to single cells required either a considerable increase in the sensitivity of those assays or the amplification of the input RNA

Exponential scaling of single-cell RNA-seq in the past decade



# History of single cell technology

---

- 1990: Single cell PCR
- Exponential amplification of cDNA from single hemopoietic cells
- Followed by PCR analysis on specific targets

## Representative in Vitro cDNA Amplification From Individual Hemopoietic Cells and Colonies

GERARD BRADY, MARY BARBARA, and NORMAN N. ISCOVE\*  
Ontario Cancer Institute, Toronto, Canada M4X 1K9

*Nucleic Acids Research*, 2006, Vol. 34, No. 5 e42  
doi:10.1093/nar/gkl050

- 2006: Single cell Microarray

## An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis

Kazuki Kurimoto<sup>1</sup>, Yukihiko Yabuta<sup>1</sup>, Yasuhide Ohinata<sup>1</sup>, Yukiko Ono<sup>1,4</sup>, Kenichiro D. Uno<sup>2</sup>, Rikuhiro G. Yamada<sup>3</sup>, Hiroki R. Ueda<sup>2,3</sup> and Mitinori Saitou<sup>1,5,6,\*</sup>



# 2009: single-cell RNA-seq was born

---

## **mRNA-Seq whole-transcriptome analysis of a single cell**

[Fuchou Tang](#), [Catalin Barbacioru](#), [Yangzhou Wang](#), [Ellen Nordman](#), [Clarence Lee](#), [Nanlan Xu](#),  
[Xiaohui Wang](#), [John Bodeau](#), [Brian B Tuch](#), [Asim Siddiqui](#), [Kaiqin Lao](#)✉ & [M Azim Surani](#)✉

[Nature Methods](#) **6**, 377–382 (2009) | [Cite this article](#)

- A modified version of the previous protocol developed for sc-Microarray
  - Time for RT incubation and PCR extension were increased
  - Modified the PCR primers eliminating end bias during sequencing
- Full length first strand cDNAs (0.5–3 kb)
- Used SOLiD sequencing system from Applied Biosystems
- Allowed the detection of thousands of genes and hundreds of new splice junctions more than a standard microarray experiment
- Compared two wild-type mature oocytes vs. two Dicer1<sup>-/-</sup> single oocytes



## Limitations of the Tang method

---

- Pronounced 3' bias, with the majority of the reads mapping to the 3' terminal portion of the transcripts
- Severe limitation for the study of transcriptional start sites (TSS) as well as in the analysis of the different splice variants
- Inefficiencies in the enzymatic reactions resulted in decreased sensitivity with consequent loss of lowly expressed transcripts
- Throughput: only sequenced 6 cells!!

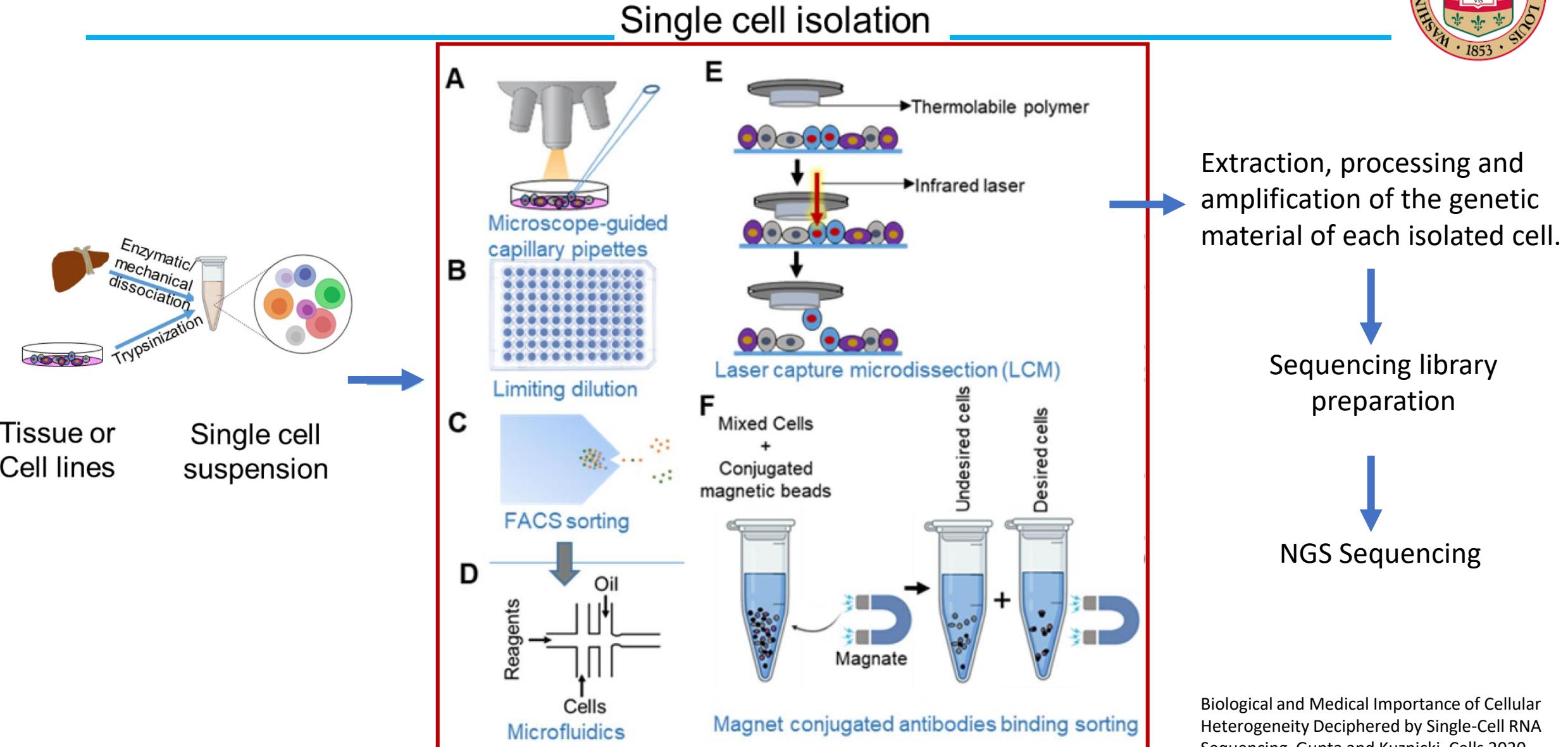
# Key technologies that have allowed jumps in experimental scale



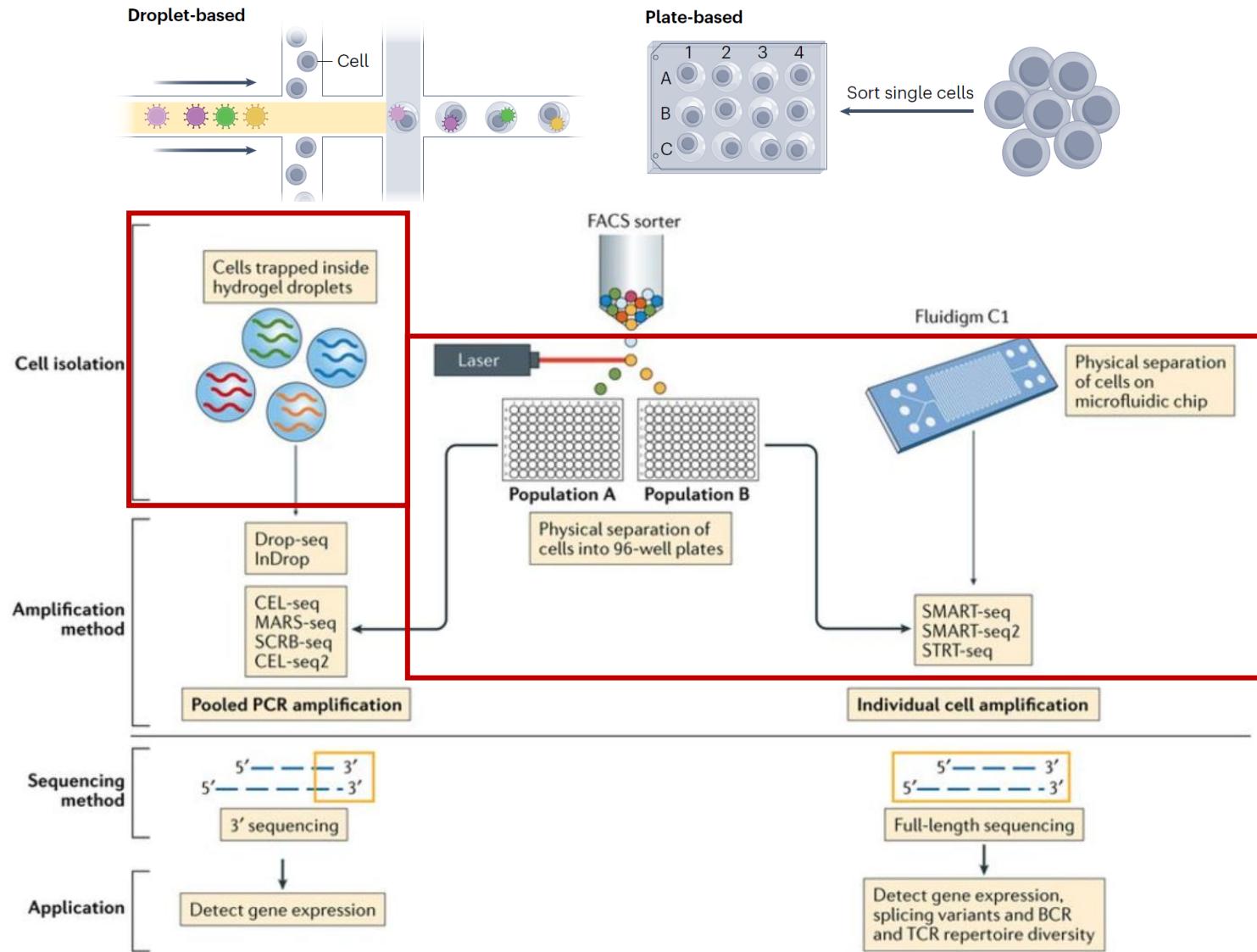
	Manual	Multiplexing	Integrated fluidic circuits	Liquid-handling robotics	Nanodroplets	Picowells	<i>In situ</i> barcoding
	Tang <i>et al.</i> 2009 <sup>18</sup>	Islam <i>et al.</i> 2011 <sup>24</sup>	Zeisel <i>et al.</i> 2015	Jaitin <i>et al.</i> 2014 <sup>33</sup>	Klein <i>et al.</i> 2015 <sup>34</sup> Macosko <i>et al.</i> 2015 <sup>40</sup>	Bose <i>et al.</i> 2015 <sup>43</sup>	Cao <i>et al.</i> 2017 <sup>51</sup> Rosenberg <i>et al.</i> 2017 <sup>52</sup>
6 cells	STRT-seq 85 cells	UMIs 3005 cells	MARS-Seq ~4000 cells	inDrop Drop-seq 10X Genomics tens of thousands	Seq-Well	Sci-RNA-seq hundreds of thousands	Split-seq

Exponential scaling of single-cell RNA-seq in the past decade, Svensson, Vento-Tormo & Teichmann, *Nature protocols*, 2018

# General workflow of single cell sequencing technologies



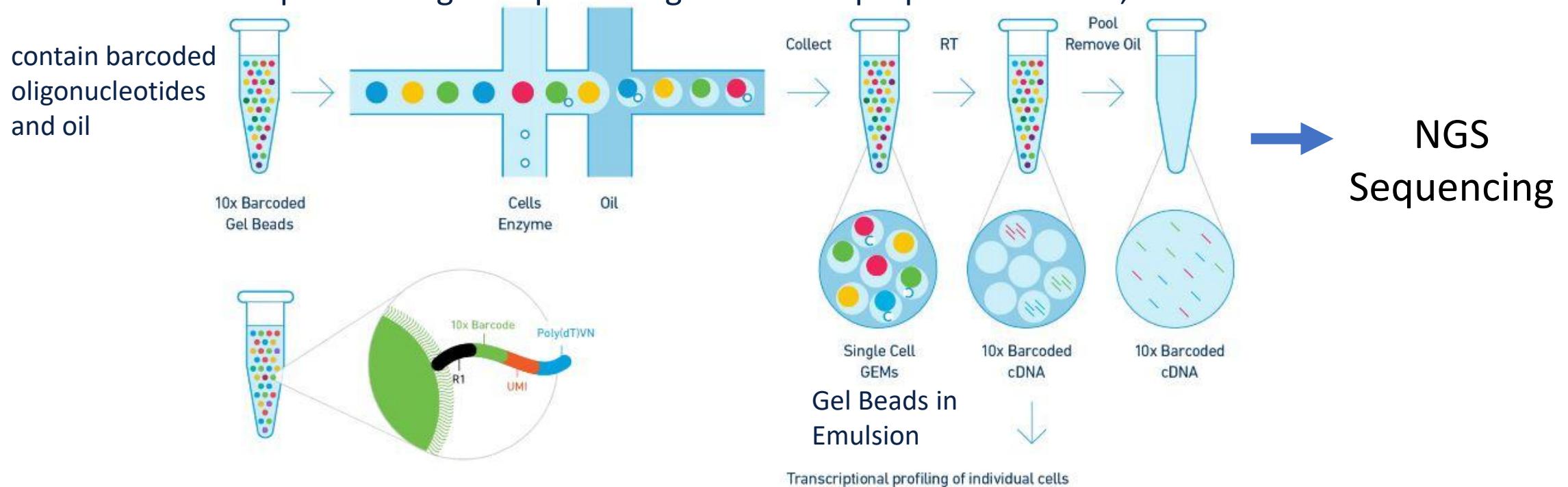
# Summary of major scRNA-seq technologies





# Successful commercialization: 10x Genomics

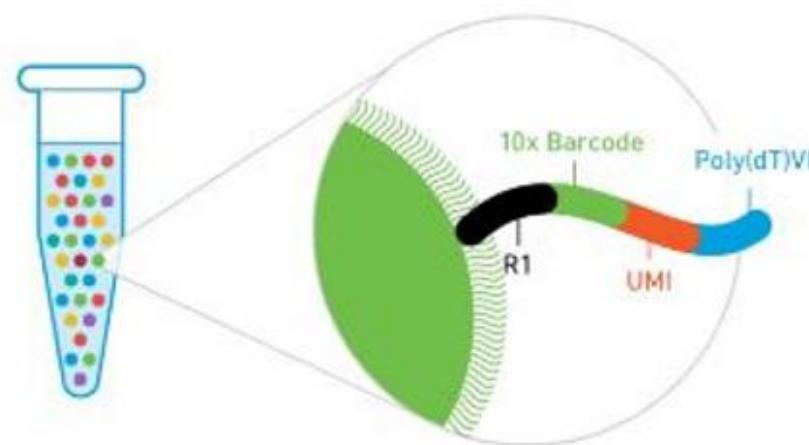
- 10X Genomics the Chromium™ Single Cell 3' Solution: 3' end sequencing
- Use microfluidic partitioning to capture single cells and prepare barcoded, NGS cDNA libraries.



- cells are loaded at a limiting dilution in order to maximize the number of GEMs containing a single cell (doublet formation )
- Next-GEM: High-throughput (10,000 cells per sample, 80,000 cells per run)
- GEM-X technology : 80% cell recovery, 20,000 cells/sample, 2 times more genes

# UMIs (Unique Molecular Identifiers)

- UMIs are random short nucleotide sequences with a very low likelihood of a duplicate UMI within a single bead.
- Each molecule has a unique UMI



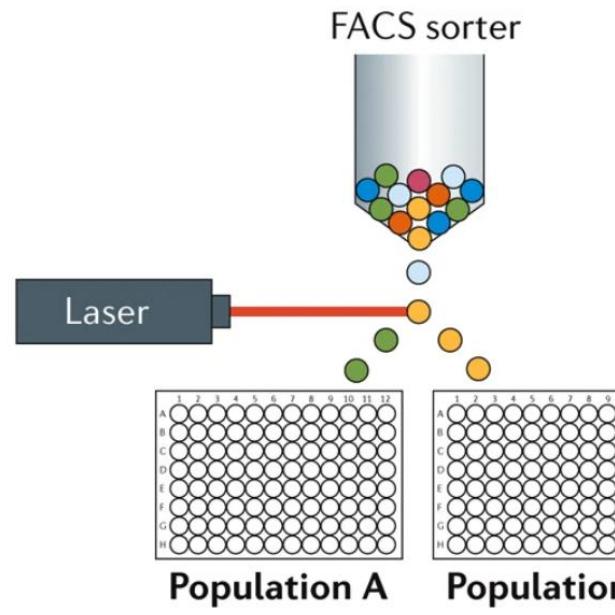
<https://mccarrolllab.org/dropseq/>

distinguish PCR duplicates vs. true biological diversity

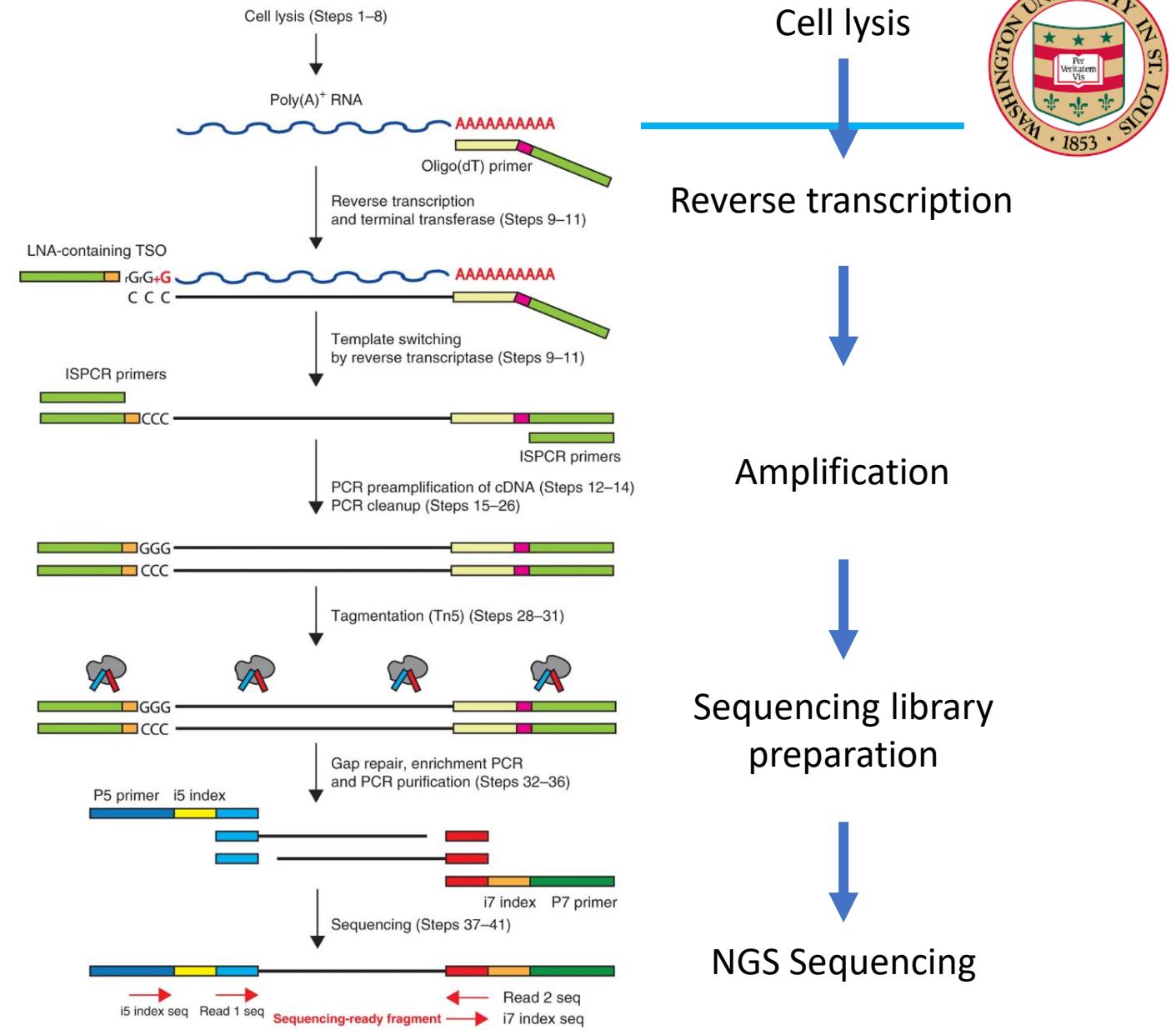
	Cellular barcode	UMI	
Cell 1	TTGCCGTGGTGT GGCGGGGA.....	CGGTGTTA	DDX51
	TTGCCGTGGTGT TATGGAGG.....	CCAGCACC	NOP2
	TTGCCGTGGTGT TCTCAAGT.....	AAAATGGC	ACTB
Cell 2	CGTTAGATGGCA GGGCCGGG.....	CTCATAGT	LBR
	CGTTAGATGGCAACGTTATA.....	ACGCGTAC	ODF2
	CGTTAGATGGCATCGAGATT.....	AGCCCTTT	HIF1A
Cell 3	AAATTATGACGA AGTTTGTA.....	GGGAATTAA	ACTB
	AAATTATGACGA AGTTTGTA.....	AGATGGGG	
	AAATTATGACGA TGTGCTTG.....	GACTGCAC	RPS15
Cell 4	GTAAACGTACCCTAGCTGT.....	GATTTTCT	GTPBP4
	GTAAACGTACC GCAGAAGT.....	GTTGGCGT	GAPDH
	GTAAACGTACC AAGGCTTG.....	CAAAGTTC	
	GTAAACGTACC TTCCGGTC.....	TCCAGTCG	ARL1
(Thousands of cells)			

[https://hbctraining.github.io/scRNA-seq/lessons/02\\_SC\\_generation\\_of\\_count\\_matrix.html](https://hbctraining.github.io/scRNA-seq/lessons/02_SC_generation_of_count_matrix.html)

# Smart-seq2 Technology



- Low-throughput (96 or 384 well plates)
- More expensive
- full-length cDNA sequencing





# Moving beyond physical cell isolation

---

- 2017: Sci-RNA-seq (single-cell combinatorial indexing RNA sequencing)

RESEARCH ARTICLE

## Comprehensive single-cell transcriptional profiling of a multicellular organism

Junyue Cao<sup>1,2,\*</sup>, Jonathan S. Packer<sup>1,\*</sup>, Vijay Ramani<sup>1,†</sup>, Darren A. Cusanovich<sup>1,‡</sup>, Chau Huynh<sup>1</sup>, Riza Daza<sup>1</sup>, Xiaojie Qiu<sup>1,2</sup>, Choli Lee<sup>1</sup>, Scott N. Furlan<sup>3,4,5</sup>, Frank J. Steemers<sup>6</sup>, Andrew Adey<sup>7,8</sup>, Robert H. Waterston<sup>1,‡</sup>, Cole Trapnell<sup>1,‡</sup>, Jay Shendure<sup>1,9,‡</sup>

- 2018: Split-pool ligation-based transcriptome sequencing (SPLiT-seq)

RESEARCH ARTICLE

## Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding

Alexander B. Rosenberg<sup>1,\*†</sup>, Charles M. Roco<sup>2,\*</sup>, Richard A. Muscat<sup>1</sup>, Anna Kuchina<sup>1</sup>, Paul Sample<sup>1</sup>, Zizhen Yao<sup>3</sup>, Lucas T. Graybuck<sup>3</sup>, David J. Peeler<sup>2</sup>, Sumit Mukherjee<sup>1</sup>, Wei Chen<sup>4</sup>, Suzie H. Pun<sup>2</sup>, Drew L. Sellers<sup>2,5</sup>, Bosiljka Tasic<sup>3</sup>, Georg Seelig<sup>1,4,6,†</sup>

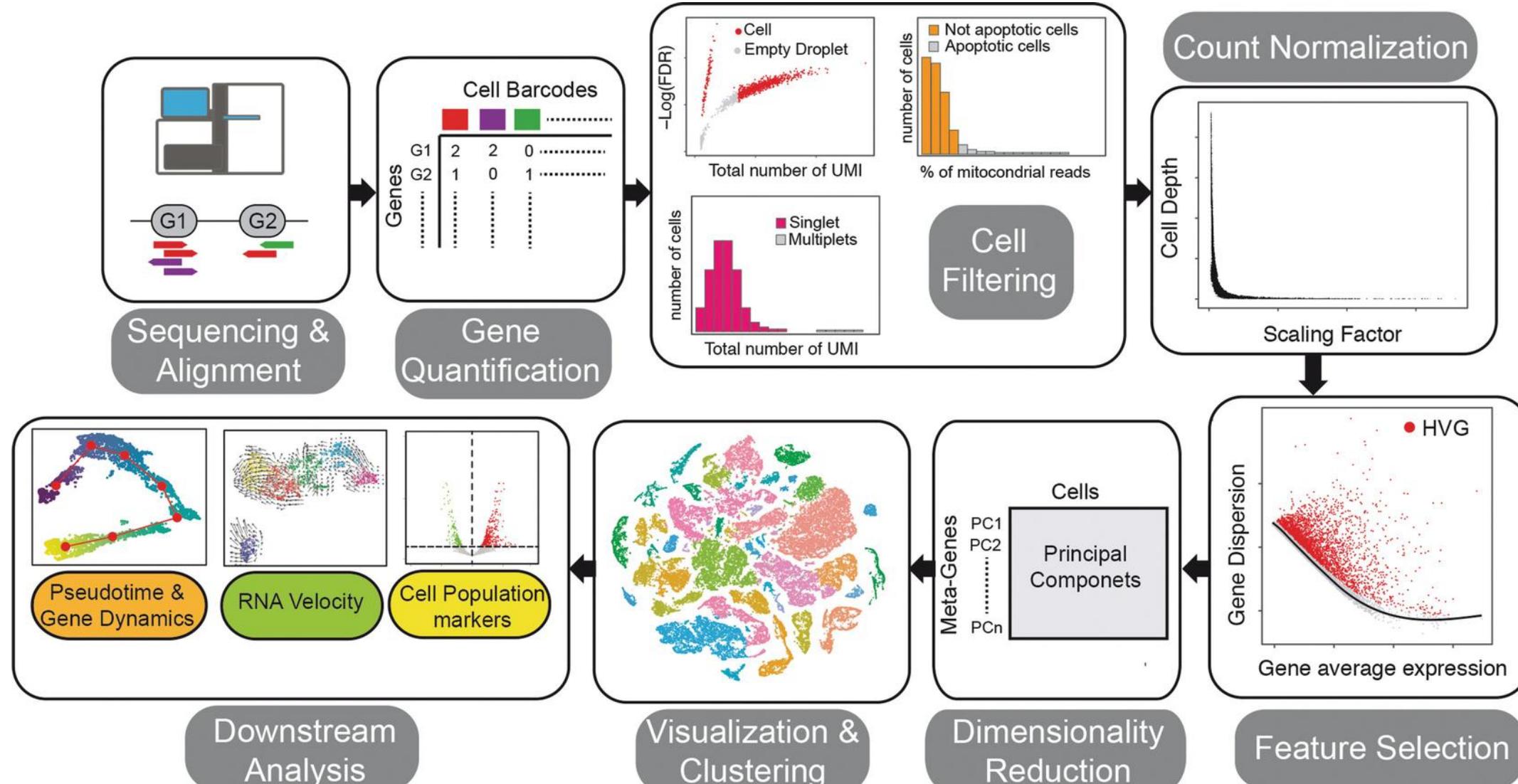


## Some limitations to consider

---

- Single cell datasets are sparse and suffer dropout
  - Dropout: a gene is observed at a low or moderate expression level in one cell but is not detected in another cell of the same cell type
- Low sensitivity: lowly expressed transcripts (e.g. those of transcription factors) are more likely to escape detection
- It is very VERY easy to produce poor quality datasets, resulting from poor cell handling
- It's easy to create artifacts e.g. cell multiplets, activated microglia states
- Loss of spatial, temporal and lineage information

# scRNA-seq data analysis - workflow





# Cell Ranger (10x Genomics)

---

- Inputs for cellranger software
  - Raw data in FASTQ format
  - Reference transcriptome
    - Downloadable pre-built references transcriptomes for human and mouse
    - Build custom reference for any species using reference genome sequence (FASTA file) and gene annotations (GTF file).

For primary analysis, the `cellranger count`, `cellranger vdj`, and `cellranger multi` pipelines will output the following types of files:

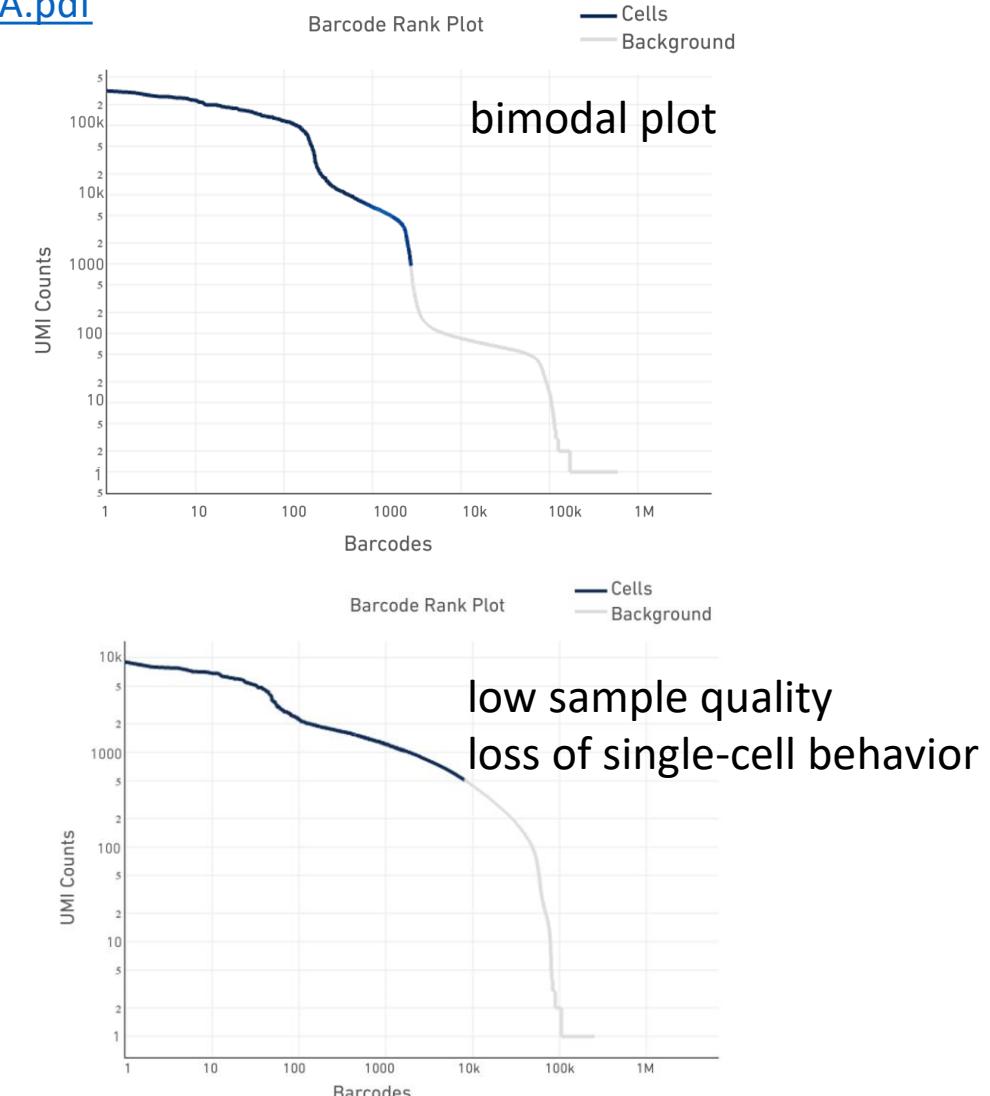
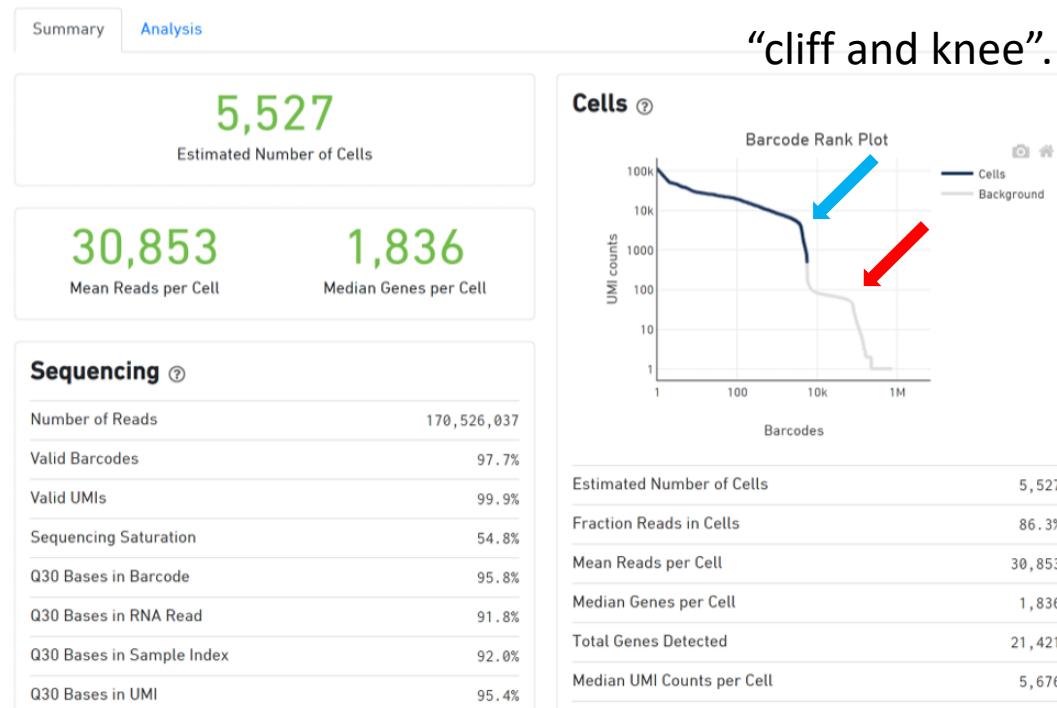
- [web summary \(HTML\) \(count, multi, vdj\)](#)
- [metrics summary CSV](#)
- [BAM](#)
- [raw and filtered feature-barcode matrices \(MEX, H5\)](#)
- [secondary analysis files \(CSV\)](#)
- [molecule info \(H5\)](#)
- [Loupe files \(cloupe and v loupe\)](#)

# Web Summary

## [CG000329\\_TechnicalNote\\_InterpretingCellRangerWebSummaryFiles\\_RevA.pdf](#)

### Interpreting Web Summary File Metrics

Representative summary files for Chromium Single Cell Gene Expression libraries and other Cell Ranger output files are available for [download](#) on the 10x Genomics Support website. Several Metrics in the web summary file can be used to assess the overall success of an experiment, including sequencing, mapping, and cell metrics. A representative web summary file for a Chromium Single Cell 3' Gene Expression library is shown below (Figure 1).



## Additional tools for read alignment and quantification



- Read Alignment
    - STAR (Dobin and Gingeras, 2015), and HISAT (Kim et al., 2015)
  - Expression Quantification of scRNA-Seq Data
    - Full length sequencing
      - Cufflinks (Trapnell et al., 2010), RSEM (Li and Dewey, 2011), and Stringtie (Pertea et al., 2015)

STAR / docs / STARsolo.md 

---

 **alexdobin** Fixed issues with documentation. fb84afe · 2 years ago 

**Preview** [Code](#) [Blame](#) 458 lines (387 loc) · 30.4 KB Raw    

---

# STARsolo: mapping, demultiplexing and quantification for single cell RNA-seq

## Gene X Cell count matrix

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

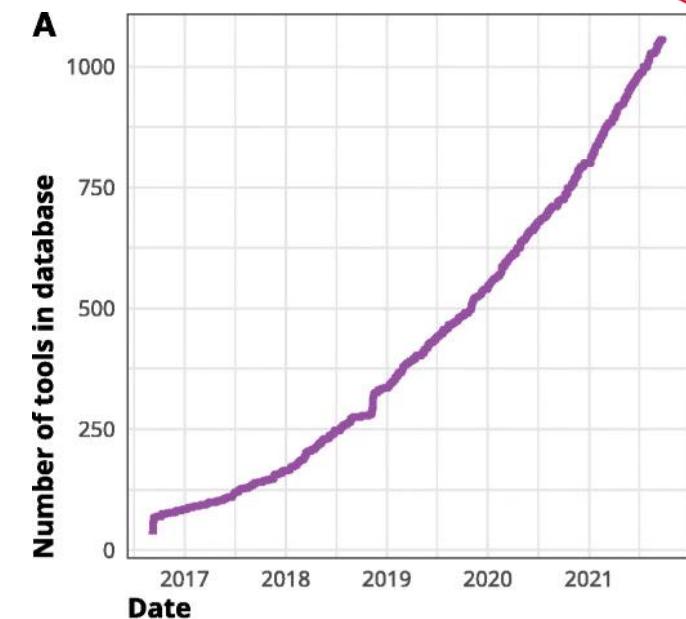


# scRNA-seq data analysis tools

- 1059 tools 2016- 2021 (15 new tools/month) (Zappia & Theis, 2021)

Most widely used general analysis toolboxes :

- Seurat (Satija et al., 2015): R
- Scanpy (Wolf et al., 2018): Python





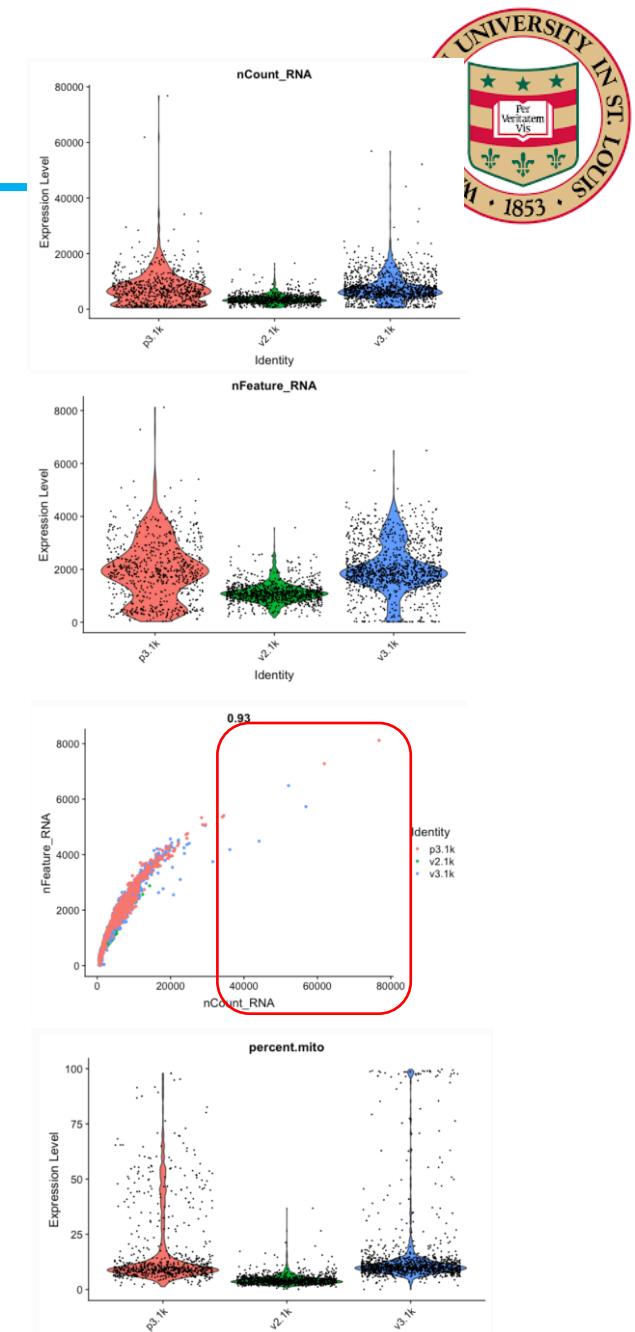
# Quality control of scRNA-Seq data is crucial

---

- Source of technical noise
  - bias of transcript coverage,
  - low capture efficiency
  - sequencing depth differences
  - dropout events
- Low-quality cells
  - Cells that are broken or dead
  - Doublets and multiple cells
- These technical noise and low-quality cells will hinder the downstream analysis and may lead to misinterpretation of the data.
  - Forming distinct, misleading clusters.
  - Inflating variance estimates in dimensionality reduction.

# Key QC Metrics

- Library size: total RNA counts per cell.
  - Low counts indicate RNA loss or prep inefficiencies.
  - Vary by protocol and dataset (commonly > 1000 for 10X Genomics)
- Expressed genes: Number of genes with non-zero expression.
  - Low values suggest incomplete transcript capture.
  - Vary by protocol and dataset (commonly > 500 for 10X Genomics)
- Mitochondrial percentage: Proportion of reads from mitochondrial genes.
  - High values point to cytoplasmic RNA loss during cell damage.
  - Commonly <10% for human and < 5% for mouse mitochondrial reads
- Spike-in percentage: Proportion of reads mapping to spike-ins.
  - Elevated levels indicate endogenous RNA loss.





# Normalization of scRNA-Seq data

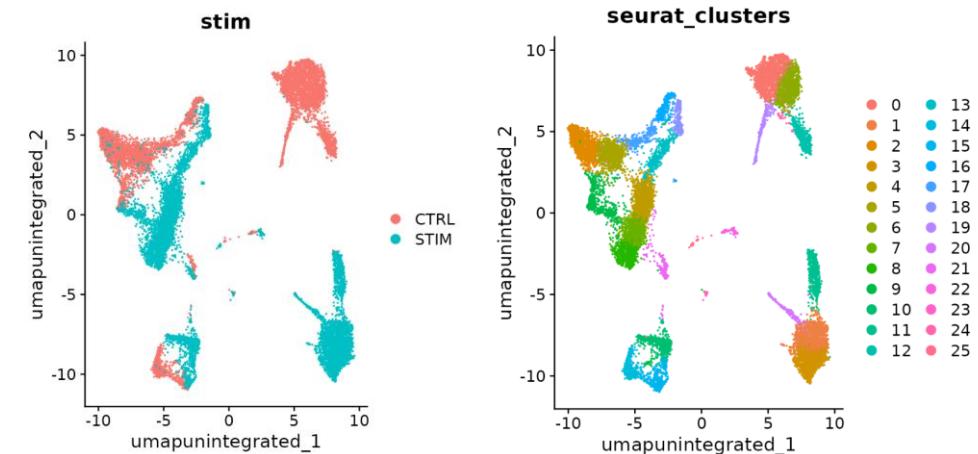
---

- Source of unwanted biases such as capture efficiency, sequencing depth, dropouts etc.
- Technical noise because of low starting material and challenging experimental protocols
- Method for scRNA-seq data normalization
  - SCnorm (Bacher et al., 2017)
  - SAMstrt (Katayama et al., 2013)
  - Sctransform (Hafemeister et al., 2019)
- The primary goal of single-cell normalization is to remove the influence of technical effects in the underlying molecular counts, while preserving true biological variation.
- Normalization of scRNA-seq data will benefit the downstream analyses including cell subpopulation identification and differential expression calling.

# What is Batch Effect?

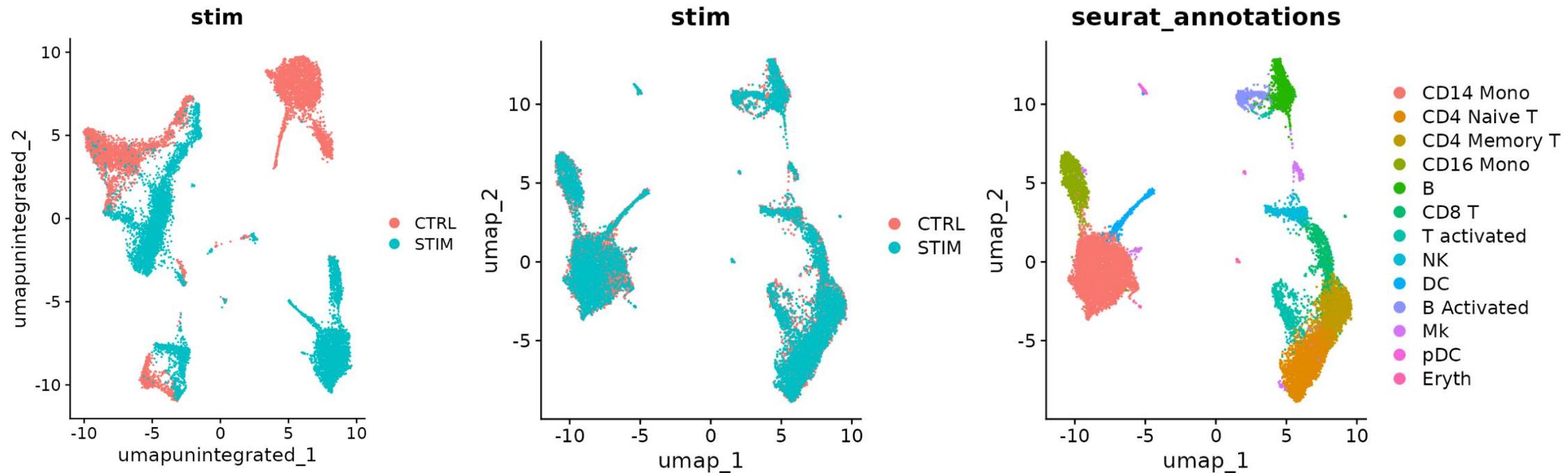
**Batch effects:** Variations in reagents, supplies, treatments, instruments and operators may introduce random or systematic differences between samples.

- Source of batch effect
  - difference cell dissociation protocols
  - different library preparation protocols
  - different sequencing platforms
  - data produced in multiple laboratories
  - data sets generated by different operators
- gene expression profile in one batch systematically differs from that in another

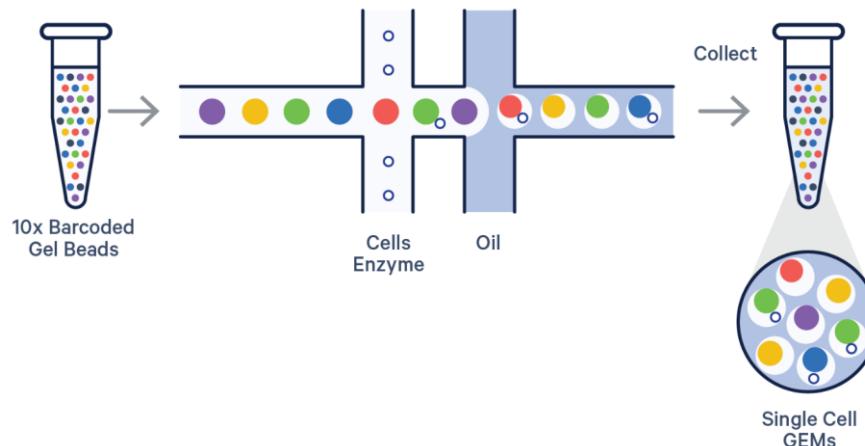


# Batch Effect Correction

- Batch effect correction tools for scRNA-seq data
  - MNN (mutual nearest neighbor) (Haghverdi et al., 2018)
  - kBET (k-nearest neighbor batch effect test) (Buttner et al., 2019).
  - Data integration (Stuart, Butler et al, 2019)



# Doublet removal



- Doublets exist in every dataset!
- Manual vs. software  
(DoubletFinder, Scrublet, etc.)  
doublet identification

## Multiplet Rate Table for the Single Cell 3' Gene Expression v3.1 assay

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.4%	~825	~500
~0.8%	~1,650	~1,000
~1.6%	~3,300	~2,000
~2.4%	~4,950	~3,000
~3.2%	~6,600	~4,000
~4.0%	~8,250	~5,000
~4.8%	~9,900	~6,000
~5.6%	~11,550	~7,000
~6.4%	~13,200	~8,000
~7.2%	~14,850	~9,000
~8.0%	~16,500	~10,000



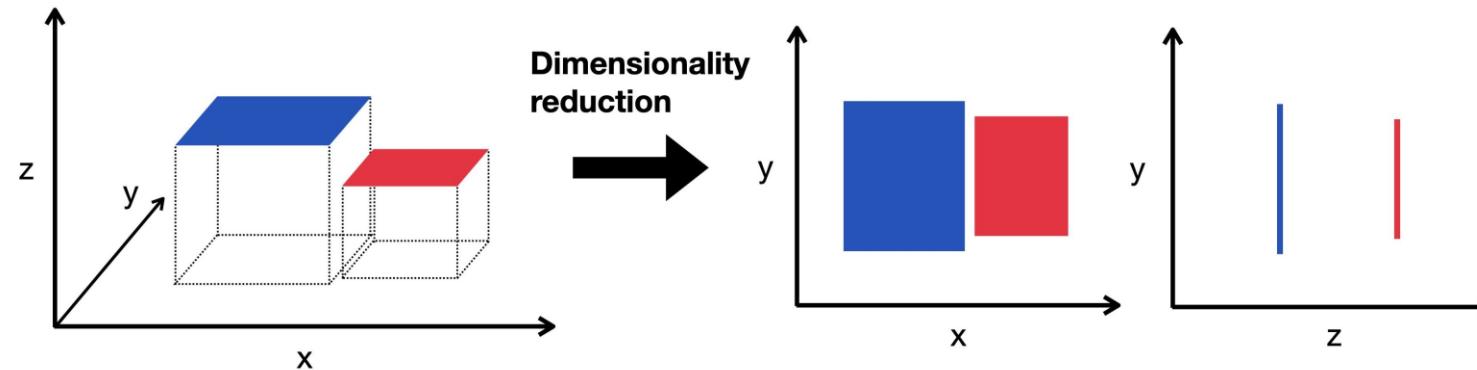
# Feature Selection

- snRNA-seq data are high dimensionality data:
  - tens of thousands of genes X thousands/millions of cells
  - To identify genes that contribute meaningful biological variation to be used in downstream analysis.
  - Improve computational efficiency for downstream analysis
  - Alleviate issues such as model overfitting in downstream analysis.
- highly variable genes (HVG) based method
- spike-in based method for data with spike-ins
- dropout-based method

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

# Dimensionality reduction

- The transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains as much meaningful properties of the original data as possible.



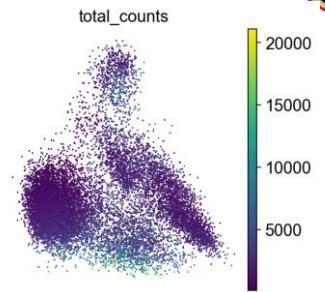
- Goal:
  - To learn the underlying manifold of the data in order to place similar cells together in low-dimensional space.
  - Reduce noise, computational complexity, and enable visualization

# Dimensionality reduction methods



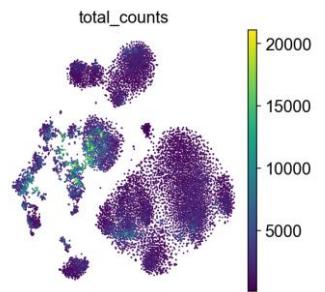
# PCA: a linear dimensional reduction algorithm

- creates a new set of uncorrelated variables (principal components, or PCs), via an orthogonal transformation of the original dataset.



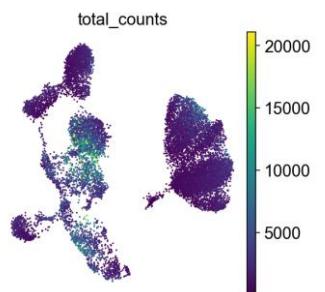
**t-SNE** (T-distributed stochastic neighbor embedding)

- graph based, non-linear dimensionality reduction technique



**UMAP** (uniform manifold approximation and projection)

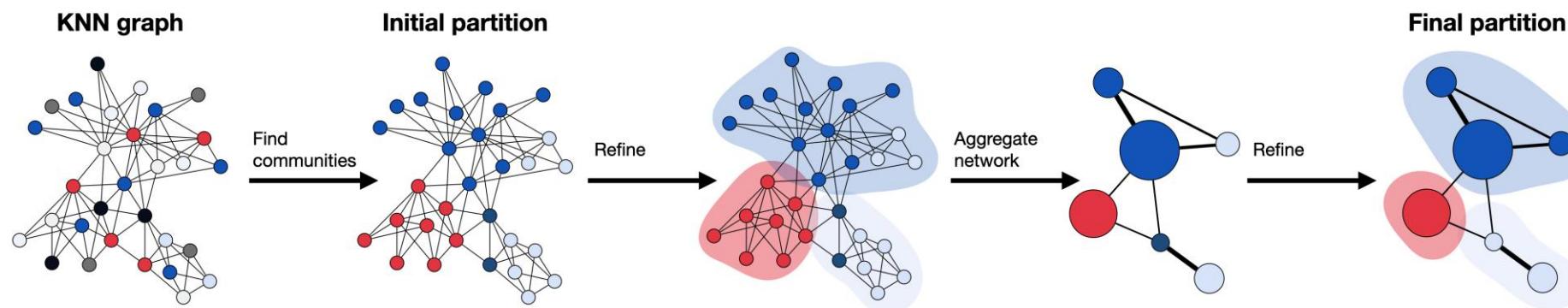
- graph based, non-linear dimensionality reduction technique
  - the fastest run times, the highest reproducibility and the most meaningful organization of cell clusters than other dimensionality reduction approaches (Becht et al., 2018).





# Cell clustering

- To group cells based on similarity in gene expression profiles, and to identify biologically meaningful groups in high-dimensional scRNA-seq data (e.g., cell types, states, or developmental stages).
- Supervised clustering methods: use a set of known markers in clustering
- Unsupervised clustering methods: for *de novo* identification of cell populations.
- k-means, hierarchical clustering, density-based clustering, graph-based clustering (K-nearest neighbor (KNN) graph)

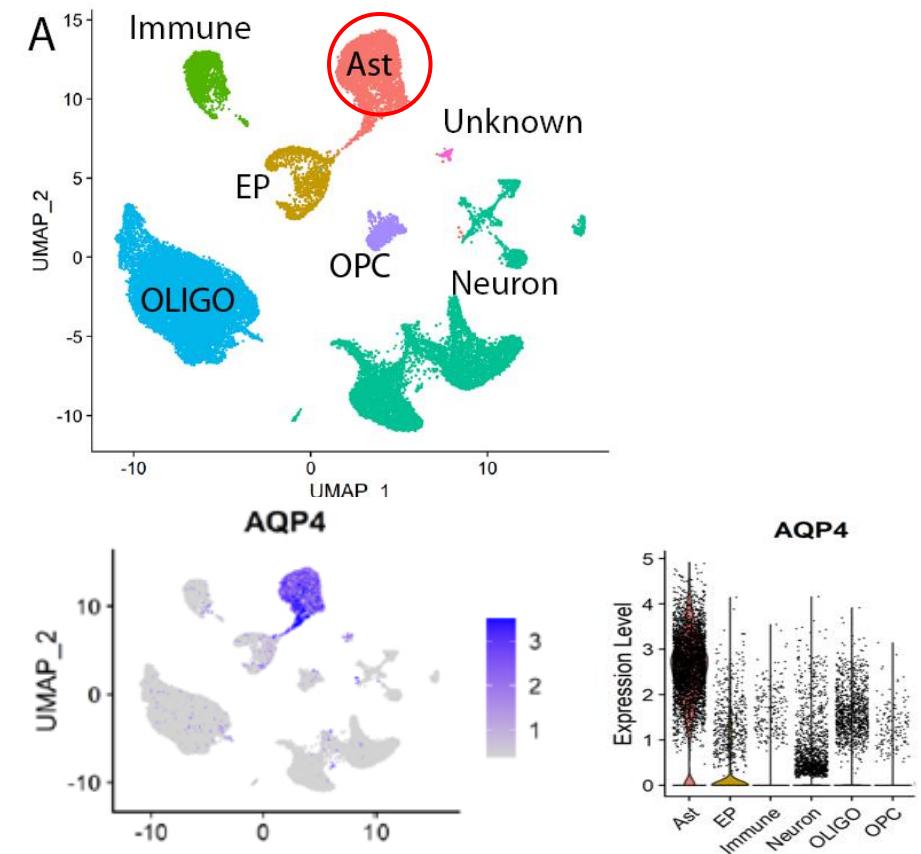


# Cell annotation

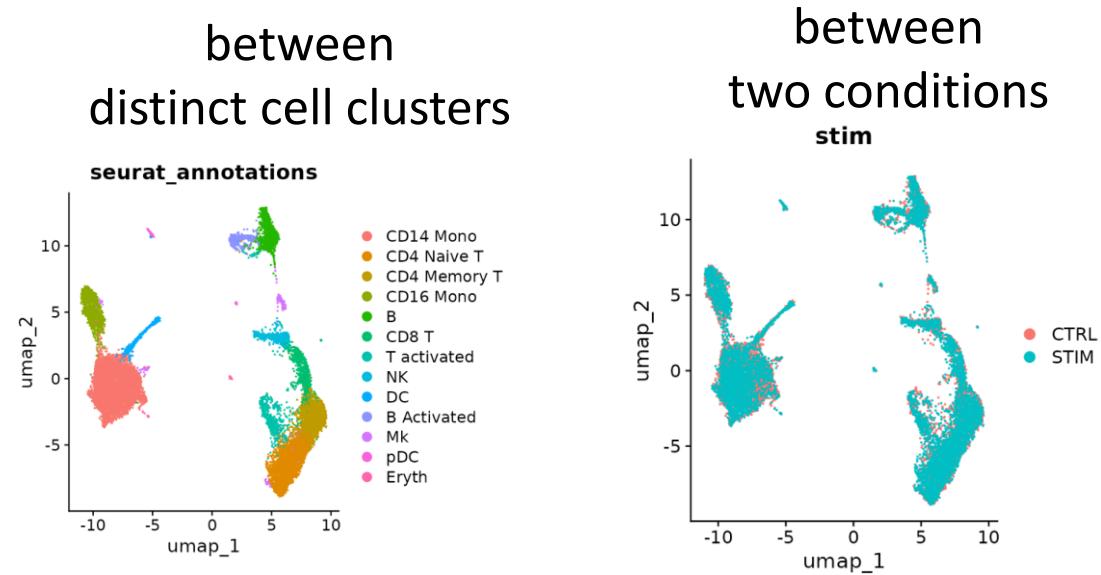
- Annotation is the process of assigning biological meaning to clusters of cells, usually based on their gene expression profiles.
- The goal is to map clusters to known cell types, states, or lineages based on expression patterns and marker genes.

## Why is annotation important?

- Provides biological context for identified clusters.
- Helps interpret cellular diversity in tissues or developmental stages.
- Facilitates comparative analysis across datasets, conditions or diseases.



# Differential Expression Analysis

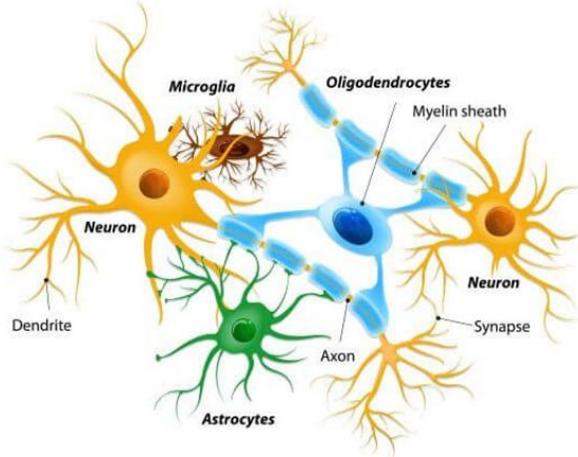


- Wilcoxon Rank-Sum Test, Pseudobulk using DESeq2, EdgeR, Mast

Why is DEG important in scRNA-seq?

- To distinguish different cell types or subpopulations.
- Helps to understand developmental processes
- Reveal biological insights into cell-type-specific responses to treatment or disease conditions
- Essential for identifying biomarkers or therapeutic targets.

# Single-cell RNA-seq vs. single-nucleus RNA-seq



	scRNA-seq	snRNA-seq
Input	whole cell	nuclei
Tissue	fresh tissue	fresh, lightly fixed, or frozen tissues, hard-to-dissociate tissues (brain, heart)
Cells	Easy-to-isolate cells (immune cells, microglia)	Difficult-to-isolate cells (mature neurons, cardiomyocytes)
Dissociation protocol	extended incubations and processing	quick and mild
Measurement	both cytoplasmic and nuclear transcripts	nuclear transcripts
Cons	Technical artifacts from heating, protease digestion	Can not capture RNA in the cytoplasm (gene isoforms, RNA in mitochondria and chloroplast etc.)

- Some transcripts might be attached to the rough endoplasmic reticulum and partially preserved in nuclear preps.
- snRNA-seq accomplishes an equivalent gene detection rate and cell type classification to that of scRNA-seq in adult kidney, human lung, and adult mammalian heart tissue.

Single-nucleus and single-cell transcriptomes compared in matched cortical cell types, Bakken et al., 2018, PLOS One.

Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis, Haojia Wu et al., 2019

Systematic Comparison of High-throughput Single-Cell and Single-Nucleus Transcriptomes during Cardiomyocyte Differentiation, Selewa et al., 2020



# General advice for analysis from CSHL

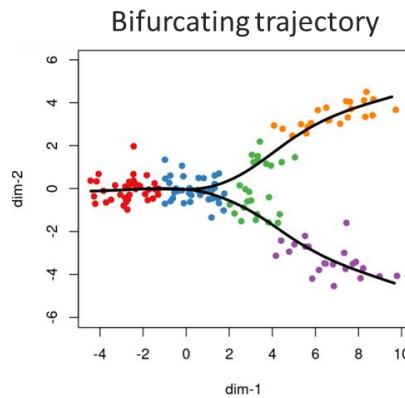
---

- Someone just spent a bunch of time, money, and effort generating this data - don't skimp or rush the analysis
- Set realistic expectations for everyone involved. The analysis will take time. A lot of time. (Sometimes years.)
- The analysis will require some programming skills. Point and click software is severely limiting.
- Every dataset is unique. Default thresholds and parameters are often not appropriate. There is no one-size-fits-all approach.
- Explore your data. Become the expert of it. Ask yourself if results make sense. Just because a program gave you an answer does not make it true.
- Flag it rather than delete it (e.g. low quality cells, doublets, etc.)
- Try multiple methods and look for concordance
- Don't try to interpret non-linear dimensionality reductions of your data (e.g. tSNE, UMAPs)
- Don't limit yourself to one software program!

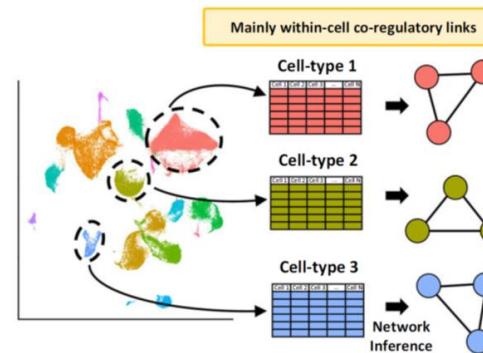
# Additional snRNA-seq data analysis

- Depending on the biological question to address, other existing tools that may provide other levels of information.

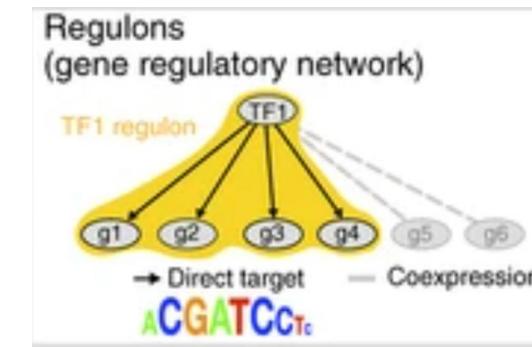
## Trajectory analysis



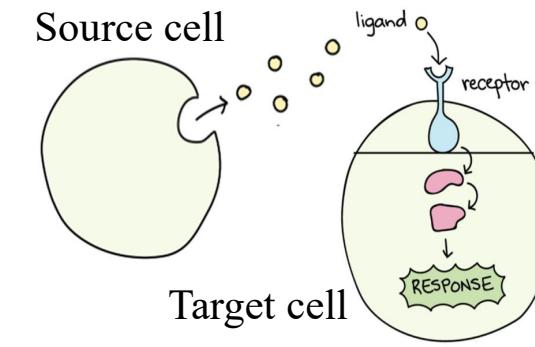
## Gene co-expression network analysis



## Gene regulatory network analysis



## Ligand-receptor network analysis



- Detection of eQTLs from scRNA-seq data
- Copy number variation detection
- Monoallelic gene expression
- Gene alternative splicing



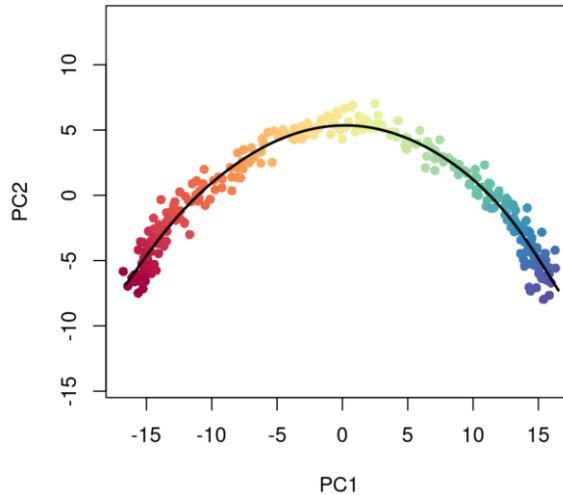
## 1) Trajectory/Pseudotime analysis

---

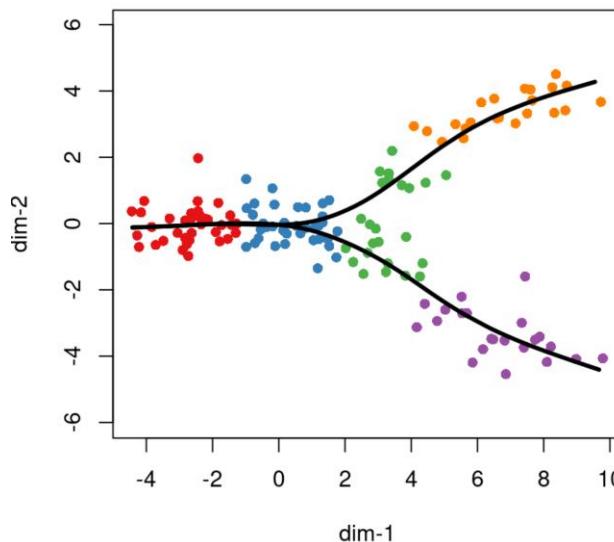
- The cells in many biological systems exhibit a continuous spectrum of states and involve transitions between different cellular states.
- Such dynamic processes within a portion of cells can be computationally modeled by reconstructing the cell trajectory/pseudotime based on scRNA-seq data.
- Pseudotime is an ordering of cells along the trajectory of a continuously process in a system, which allows the identification of the cell types at the beginning, intermediate, and end states of the trajectory (Griffiths et al., 2018).

# Trajectory topology

Linear trajectory



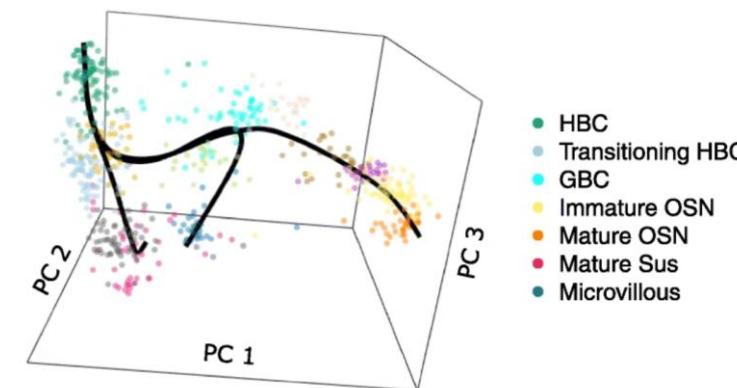
Bifurcating trajectory



Multi-branched trajectory

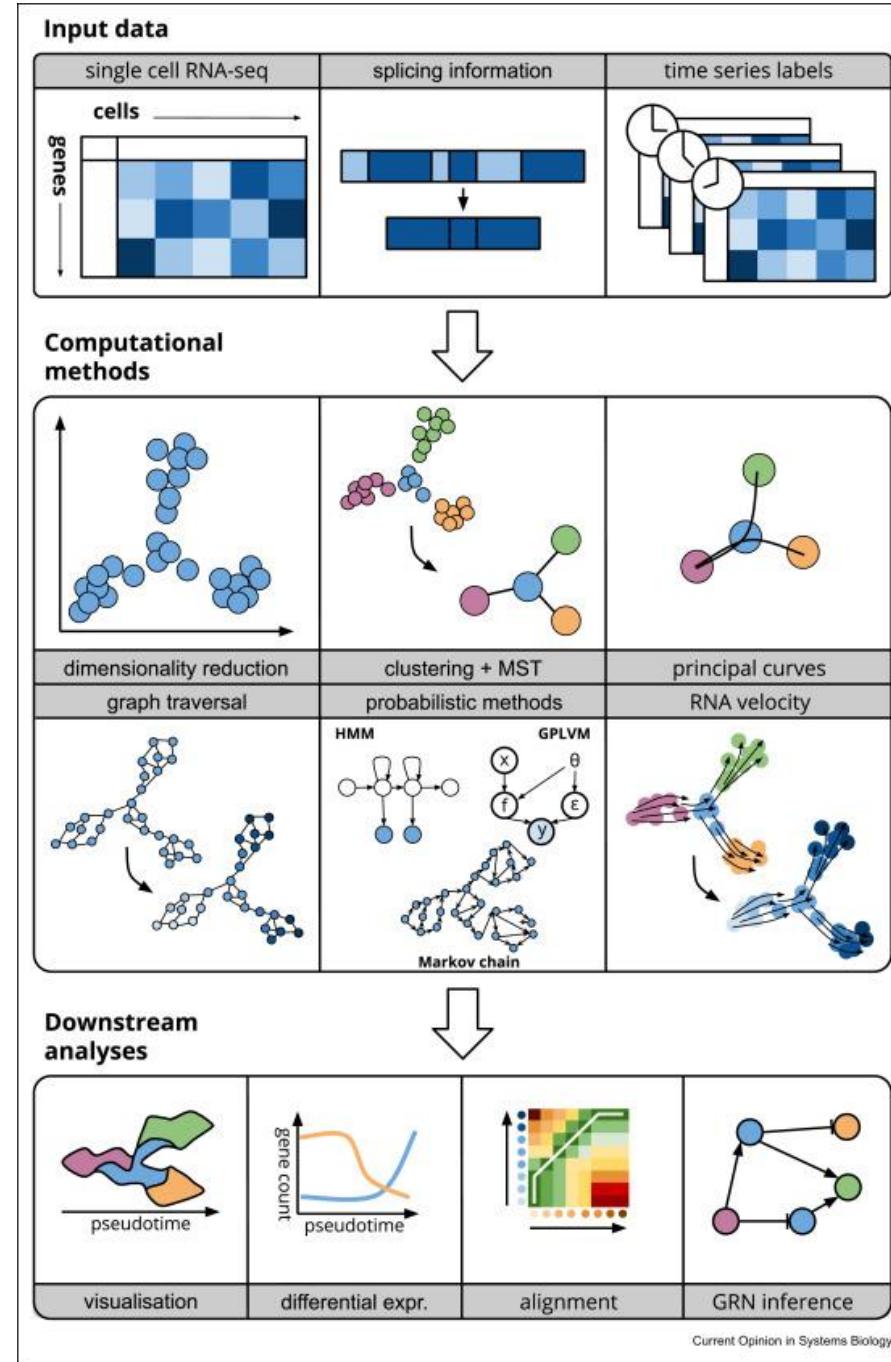


- Mature sustentacular cells (mSus) are produced via direct conversion of horizontal basal cells (HBC)
- Microvillous (MV) and mature olfactory sensory neurons (mOSN) cells require an intermediate, proliferative state (GBC).



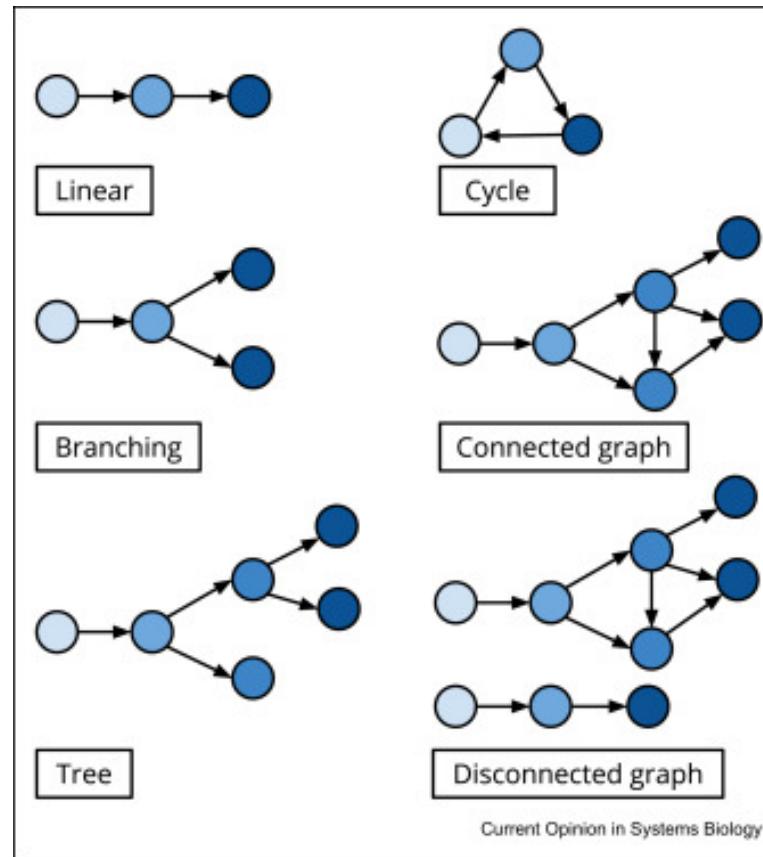
- reveal the gene expression dynamics across cell states
- identification of the factors triggering state transitions

# Trajectory topology



- Clustering-based approaches
- Principal curves
- Graph-based approaches
  - graph decomposition to reveal connected and disconnected components,
  - graph diffusion or traversal methods to construct the trajectory topology.
- Probabilistic approaches
- Include additional information, e.g. RNA velocity

# Which tool to use?



**Linear trajectories:** e.g. cell maturation, Wanderlust, MATCHER, and SCORPIUS

**Cyclical trajectories**, e.g. cell cycle, ElPiGraph and reCAT.

**Branched trajectories:** Slingshot and Monocle

**Complicated graphs:** PAGA, RacelID/StemID, and TinGa, allow the inclusion of loops or even multiple separated trajectories.

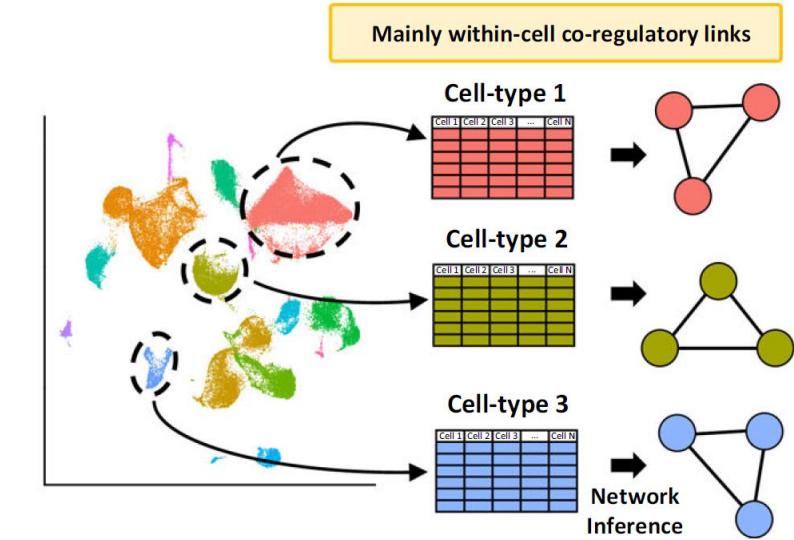
The best trajectory analysis tools (Saelens et al. 2019):

- TSCAN (Ji and Ji, 2016)
- Slingshot (Street et al., 2018)
- Monocle (Qiu et al., 2017)

- The true underlying process is unknown.
- Best practices: comparing multiple methods to confirm the general topology structure

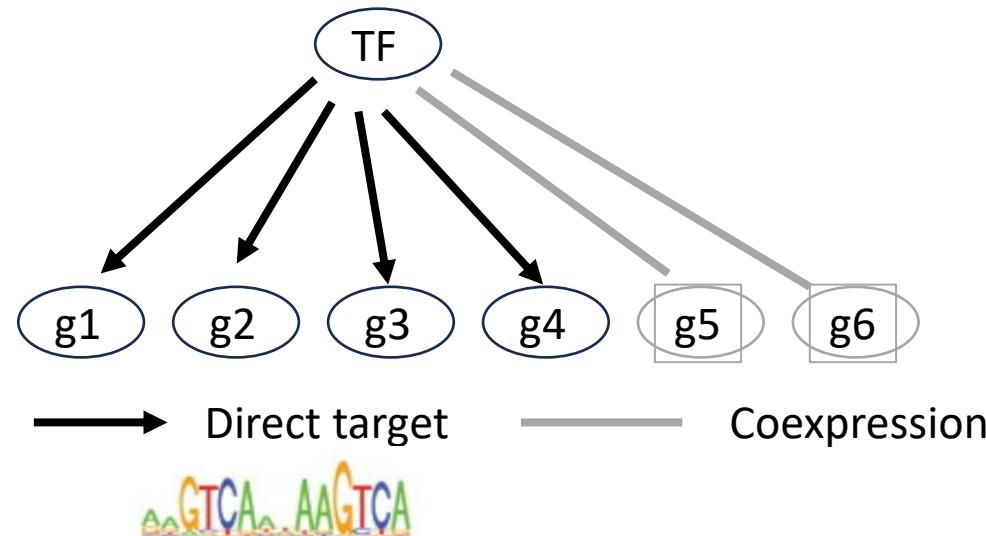
## 2) Gene co-expression network inference

- A gene co-expression network (GCN) is an undirected graph, where each node corresponds to a gene, and a pair of nodes is connected with an edge if there is a significant co-expression relationship between them.
- Can be constructed by looking for pairs of genes which show a similar expression pattern across samples, since the transcript levels of two co-expressed genes rise and fall together across samples.
- Co-expressed genes could be controlled by the same transcriptional regulatory program, functionally related, or members of the same pathway or protein complex.
- SCINET (Mohammadi et al., 2019)



### 3) Gene regulatory network inference

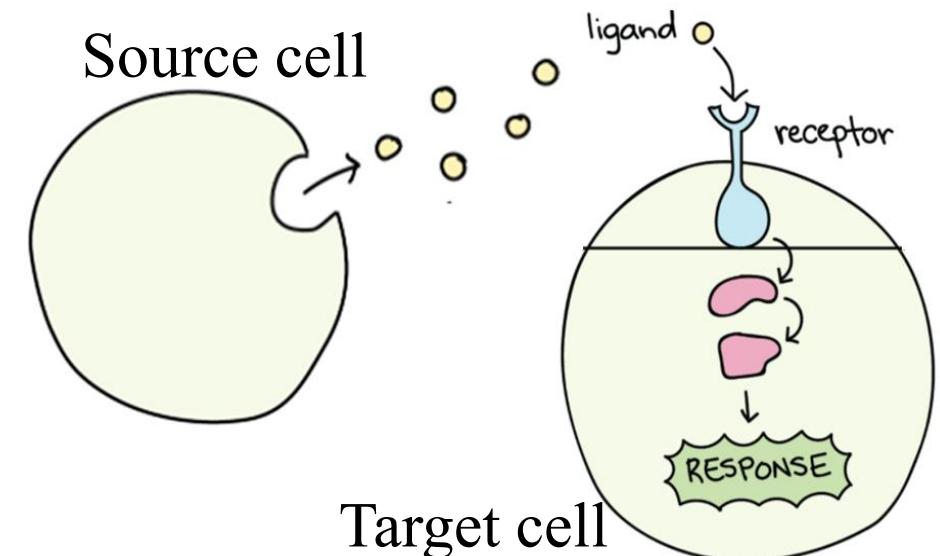
- Assumption: the genes highly correlated in expression could be co-regulated.
- SCENIC (Aibar et al., 2017)



## 4) Ligand-receptor network analysis



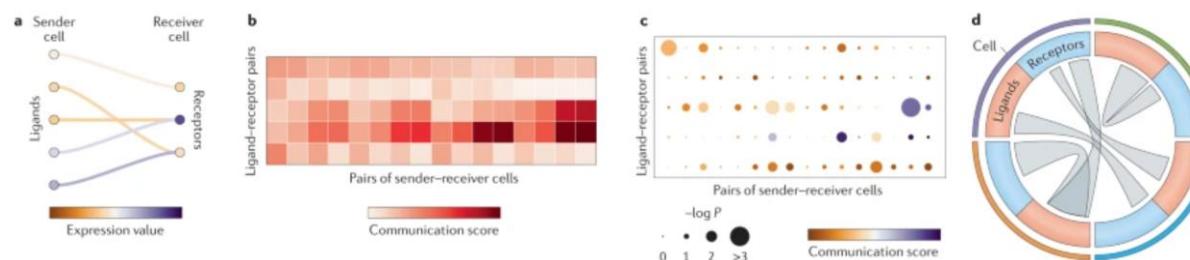
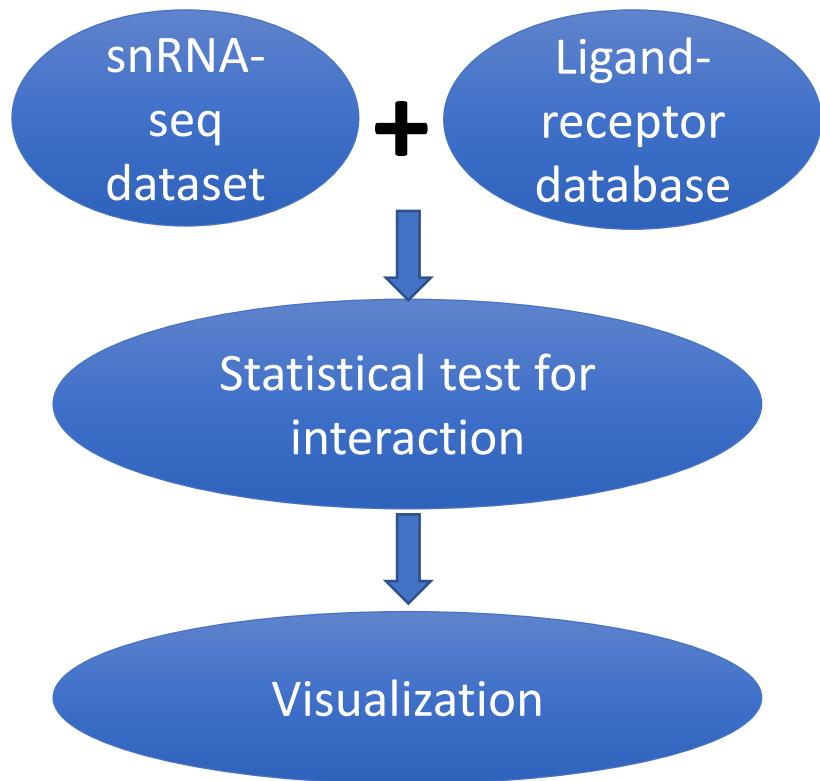
- Cell-cell communication is vital for multicellular organisms to coordinate cellular activities across diverse cell types and tissues in response to internal and external stimuli.
- Cell-cell communication is primarily mediated by ligand (L)-receptor (R) interactions
  - Ligands: hormones, growth factors, chemokines, cytokines and neurotransmitters
- Ligand-receptor network analysis
  - To identify the protein messages passed between cells and their associated pathways
  - To understand the directionality, magnitude and biological relevance of cell-cell communication



# Software of ligand-receptor network analysis



- **Differential combination-based tools:** significantly differentially expressed genes between cell clusters are identified and analyzed for ligand–receptor pairs. e.g. iTALK, CellTalker, PyMINEr
- **Network methods:** evaluate the effect of ligand–receptor co-expression on downstream signalling genes in the receiver cell. e.g. CCCExplorer, SoptSCm and NicheNet.
- **Expression permutation-based tools:** compute a communication score for each ligand–receptor pair and evaluate its significance through permutation. e.g. CellPhoneDB, CellChat, and ICELLNET



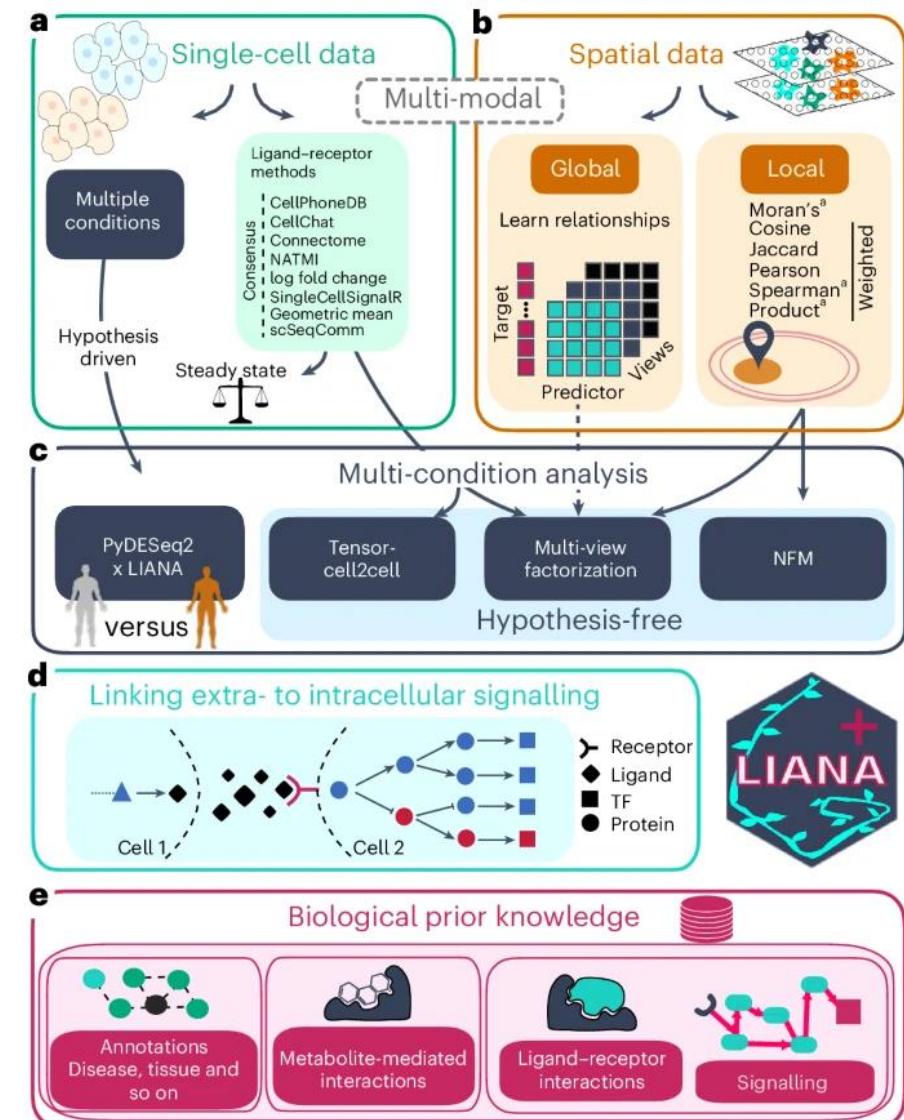


## LIANA+ provides an all-in-one framework for cell-cell communication inference

Daniel Dimitrov, Philipp Sven Lars Schäfer, Elias Farr, Pablo Rodriguez-Mier, Sebastian Lobentanzer, Pau Badia-i-Mompel, Aurelien Dugourd, Jovan Tanevski, Ricardo Omar Ramirez Flores & Julio Saez-Rodriguez

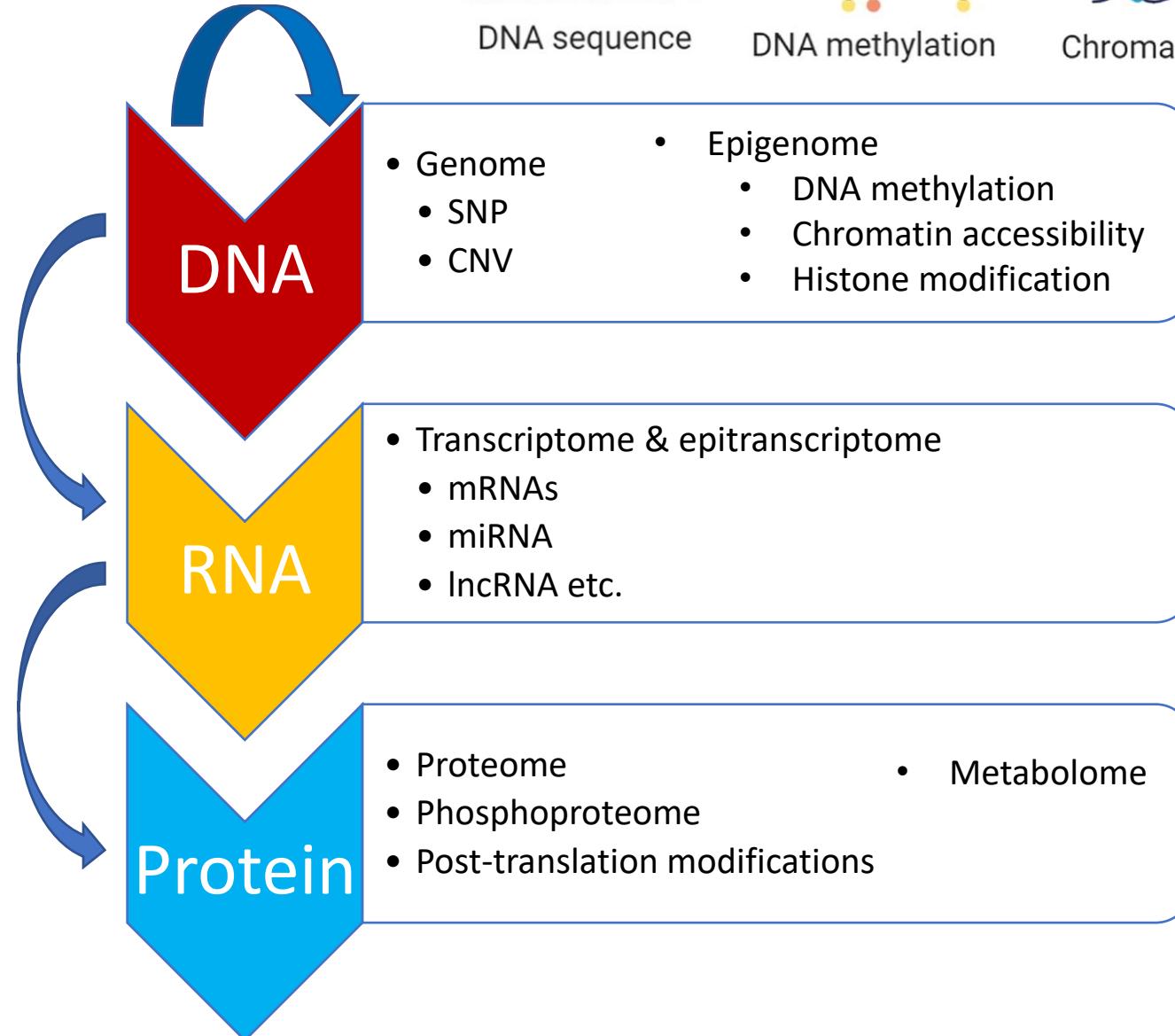
*Nature Cell Biology* 26, 1613–1622 (2024) | [Cite this article](#)

- LIANA+ re-implements and adapts eight ligand–receptor methods to infer interactions from single-cell data, along with a flexible consensus that can integrate any combination of these methods.



# Unimodal single-cell technology

## Central dogma



## Bulk version

BS-seq  
DNase-seq  
ATAC-seq  
MNase-seq  
ChIP-seq  
Cut&Tag  
...

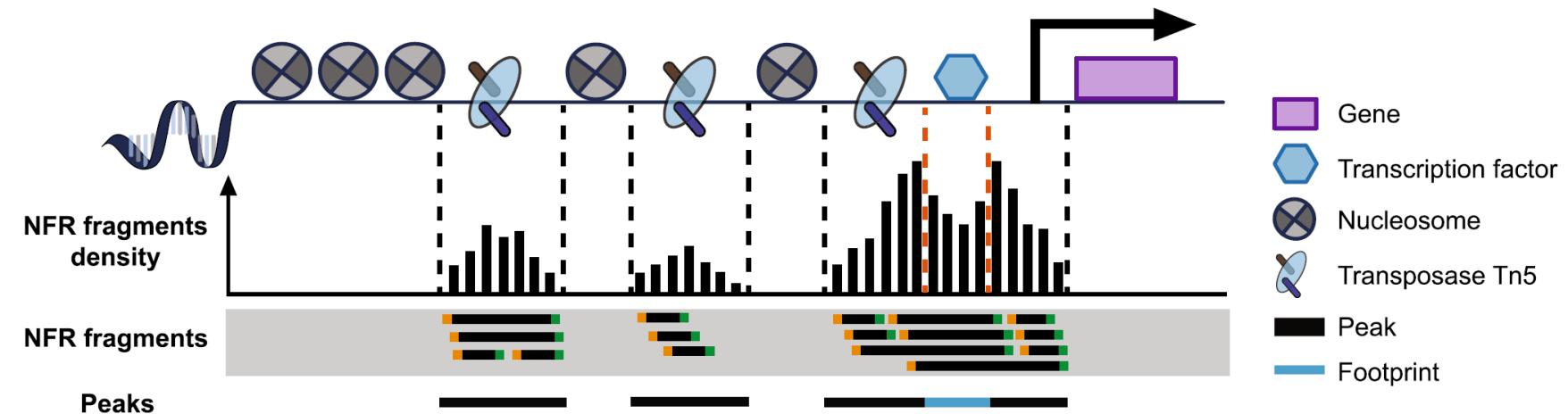
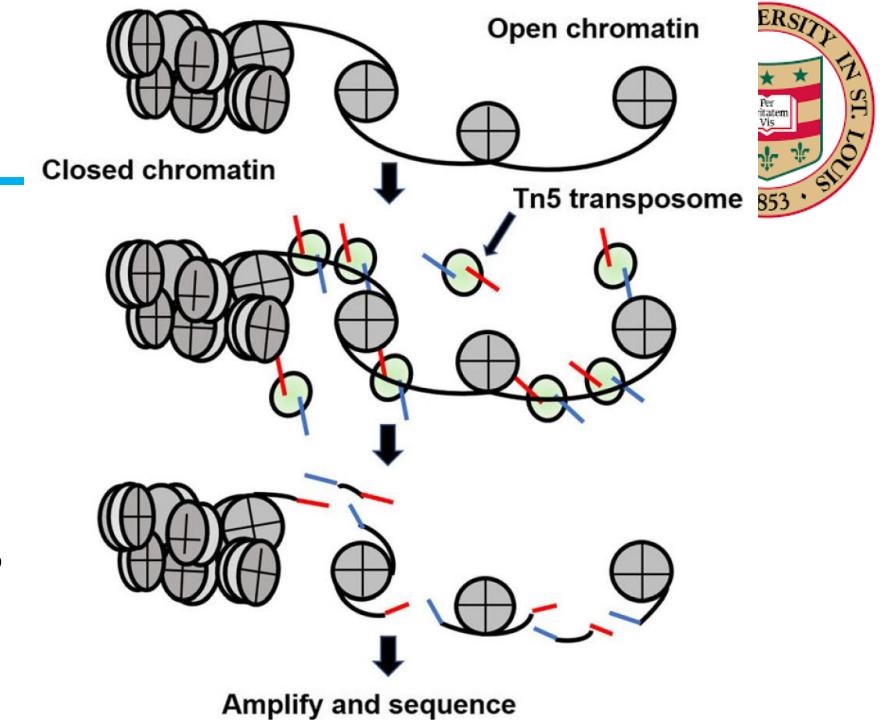
## Single-cell version

BS-seq  
DNase-seq  
ATAC-seq  
MNase-seq  
ChIP-seq  
Cut&Tag  
...

- 5mC
- 5hmC
- Open chromatin
- Nucleosome
- Nucleosome with modified histone
- Transcription factor
- RNA polymerase
- Cohesin

# Single Cell ATAC-seq

- ATAC-seq: Assay for transposase-accessible chromatin using sequencing
- Buenrostro et al., Nature Methods, 2013
- Uses a hyperactive Tn5 transposase to insert sequencing adaptors into accessible chromatin regions
- Measuring chromatin accessibility = potential regulatory sequences





# Single-cell proteomics

---

- Proteins are the functional actors of cells
- Proteomic differences poorly correlate with corresponding transcriptomic differences between biological states
- Post-translational modifications (PTMs)
- mass spectrometry (MS)-based single-cell proteomics

► *Mol Syst Biol.* 2022 Feb 28;18(3):e10798. doi: [10.1525/msb.202110798](https://doi.org/10.1525/msb.202110798)

Sadly, at this point, only a few specialized labs have true access to single-cell proteomics-based cutting-edge research.

- fairly deep expertise needed in cell sorting
- specialized cell lysis techniques
- sophisticated liquid chromatography (LC) mass spectrometry instrumentation
- advanced statistical analysis techniques

**Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation**

[Andreas-David Brunner](#)<sup>1</sup>, [Marvin Thielert](#)<sup>1</sup>, [Catherine Vasilopoulou](#)<sup>1</sup>, [Constantin Ammar](#)<sup>1</sup>, [Fabian Coscia](#)<sup>2</sup>, [Andreas Mund](#)<sup>2</sup>, [Ole B Hoerning](#)<sup>3</sup>, [Nicolai Bache](#)<sup>3</sup>, [Amalia Apalategui](#)<sup>4</sup>, [Markus Lubeck](#)<sup>4</sup>, [Sabrina Richter](#)<sup>5,6</sup>, [David S Fischer](#)<sup>5,6</sup>, [Oliver Raether](#)<sup>4</sup>, [Melvin A Park](#)<sup>7</sup>, [Florian Meier](#)<sup>1,8</sup>, [Fabian J Theis](#)<sup>5,6</sup>, [Matthias Mann](#)<sup>1,2,✉</sup>

- 10-fold improved sensitivity
- True single-cell proteomics

The Current State of Single-Cell Proteomics Data Analysis, Vanderaa and Gatto, 2023  
Initial recommendations for performing, benchmarking and reporting single-cell proteomics experiments, Gatto et al., 2023  
A review of the current state of single-cell proteomics and future perspective, Ahmad and Budnik, 2023

# Commercially available unimodal single cell technology

---



10X Chromium IX



Single cell transcriptome  
Single Cell Methylation



Tapestri



BD Rhapsody™ System

single nucleotide variation (SNV)  
copy number variation (CNV)



# Single cell genomics

---

- Single cell genomics I
  - History of Single cell technology
  - Single cell RNA-sequencing (scRNA-seq) technology
  - Basic scRNA-seq data analysis workflow
  - Unlocking biological insights
  - Other unimodal single-cell technology
- Single cell genomics II
  - Single-cell multiomics
    - Transcriptome + Epigenome
    - Transcriptome + Protein
    - Transcriptome + CRISPR screening
    - Transcriptome + TCR/BCR
    - Transcriptome + Antigen specificity
  - Spatial genomics
    - Spatial transcriptomics
    - Spatial proteomics
    - Spatial multiomics
    - Spatial metabolomics
- Single cell genomics Lab
  - scRNA-seq data analysis