

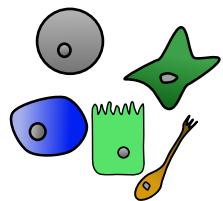
Introduction to single-cell multiomics

“Single-cell multiomic data analysis”
Summer Semester 2025
Antonio Scialdone

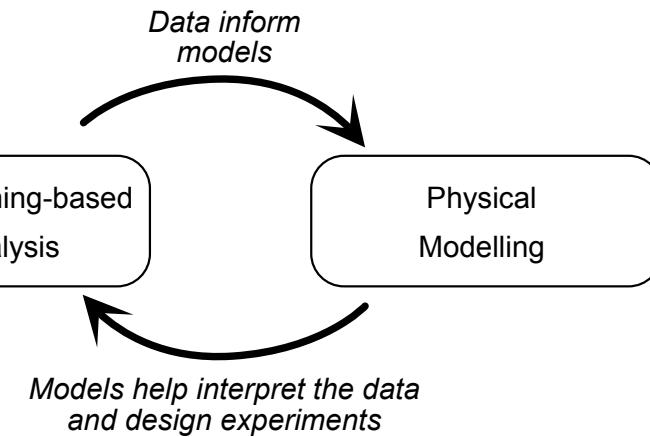
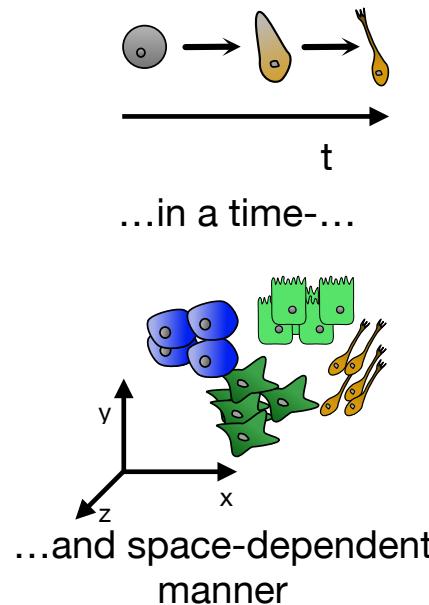
Scialdone Lab - Physics and data-based modelling of cellular decision making



Joshua Clauss
Sara De Benedetti
Veronica Finazzi
Kim Job
Gabriele Lubatti
Martin Miranda
Marco Stock



The same set of instructions
can generate many **different**
cellular identities...



How?

Overview of the course

- 8 days in total : 27.5 - 30.5 and 3.6 - 6.6

- (Typical) day schedule:

10:00 - 11:00	Lecture
11:00 - 11:15	Coffee Break
11:15 - 12:00	Lecture / Practical
12:00 - 13:00	Lunch Break
13:00 - 14:30	Practical
14:30 - ...	Homework assignment

- Scientific Seminars:

- Wednesday 4th, 10:00, Gizem Mansour / Ali Oguz Can
- Thursday 5th, 10:00, Antonio Scialdone
- Friday 6th, 10:00, Maria Colomé-Tatché

- Evaluation based on homework assignments

- Communication via Slack Channel

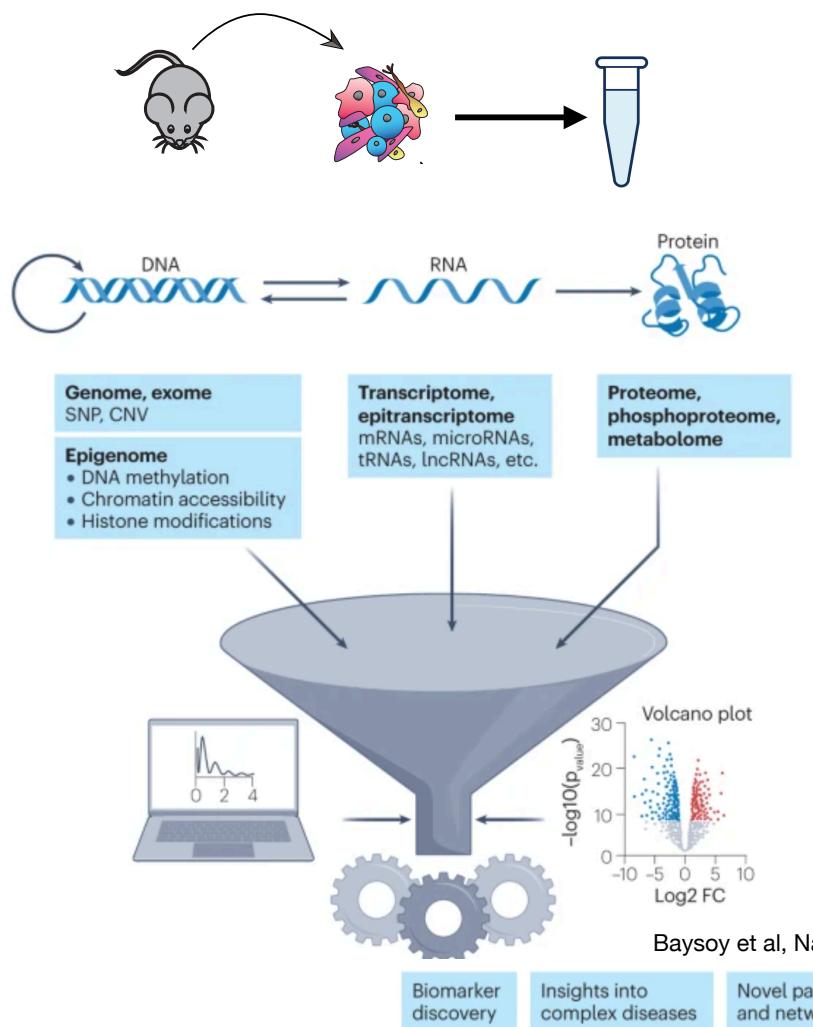
By the end of this course you should...

- Understand the opportunities and challenges that single-cell multiomics data present
- Know the first steps in the analysis of a scRNA-seq/ATAC-seq data
- Perform basic analyses of scRNA-seq/ATAC-seq data with popular R packages (Seurat, Signac, etc)

Lecture outline

- General introduction
- Single-cell RNA- and ATAC-seq
- First step: getting count matrices

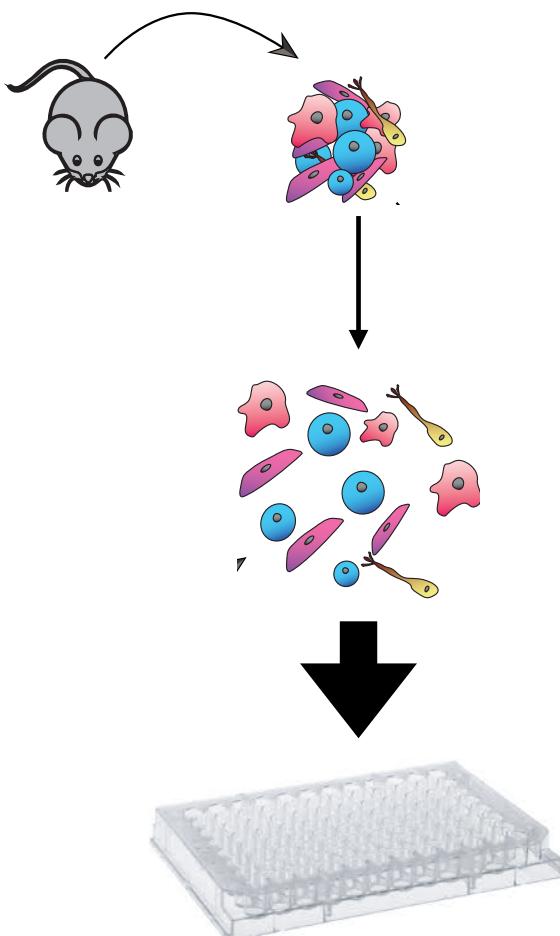
“Omics” technologies allow the molecular profiling of biological samples



- Collective, **unbiased** characterisation of specific group of biological molecules
- Several types of molecular features can be mapped: mRNA quantity, mRNA modification, DNA copy number, DNA modification, chromatin accessibility, chromatin physical contacts, etc

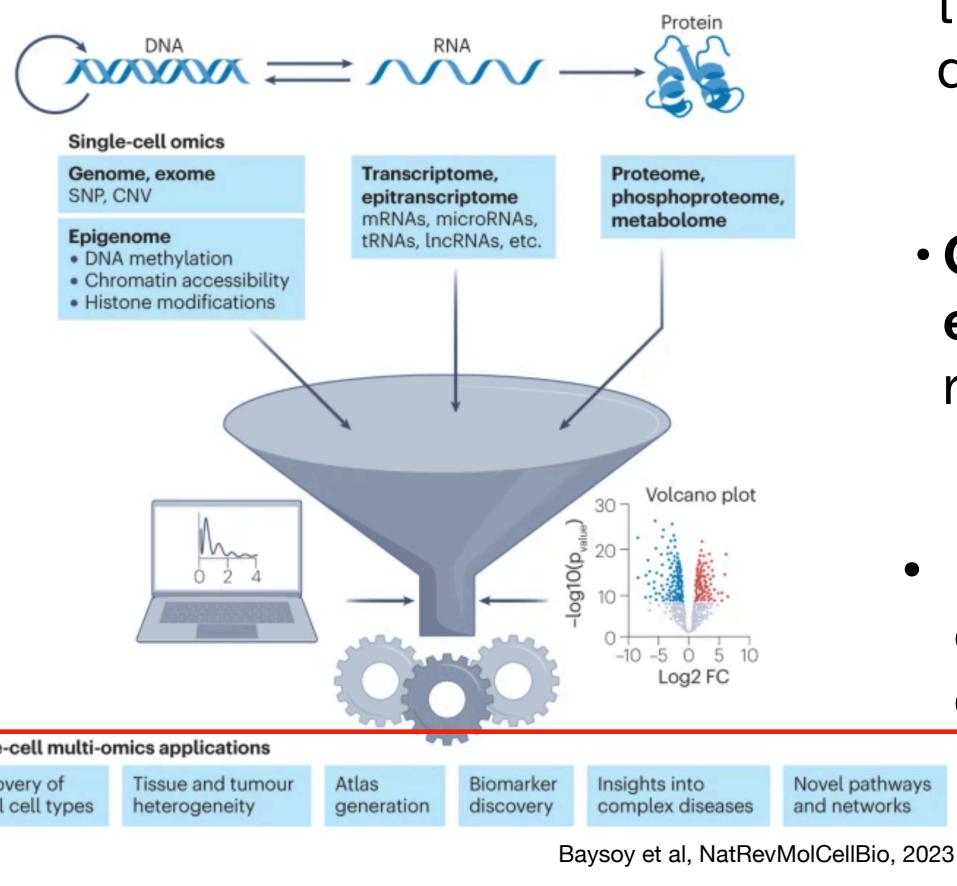
Baysoy et al, NatRevMolCellBio, 2023

Single-cell-level analyses are needed to explore cellular heterogeneity



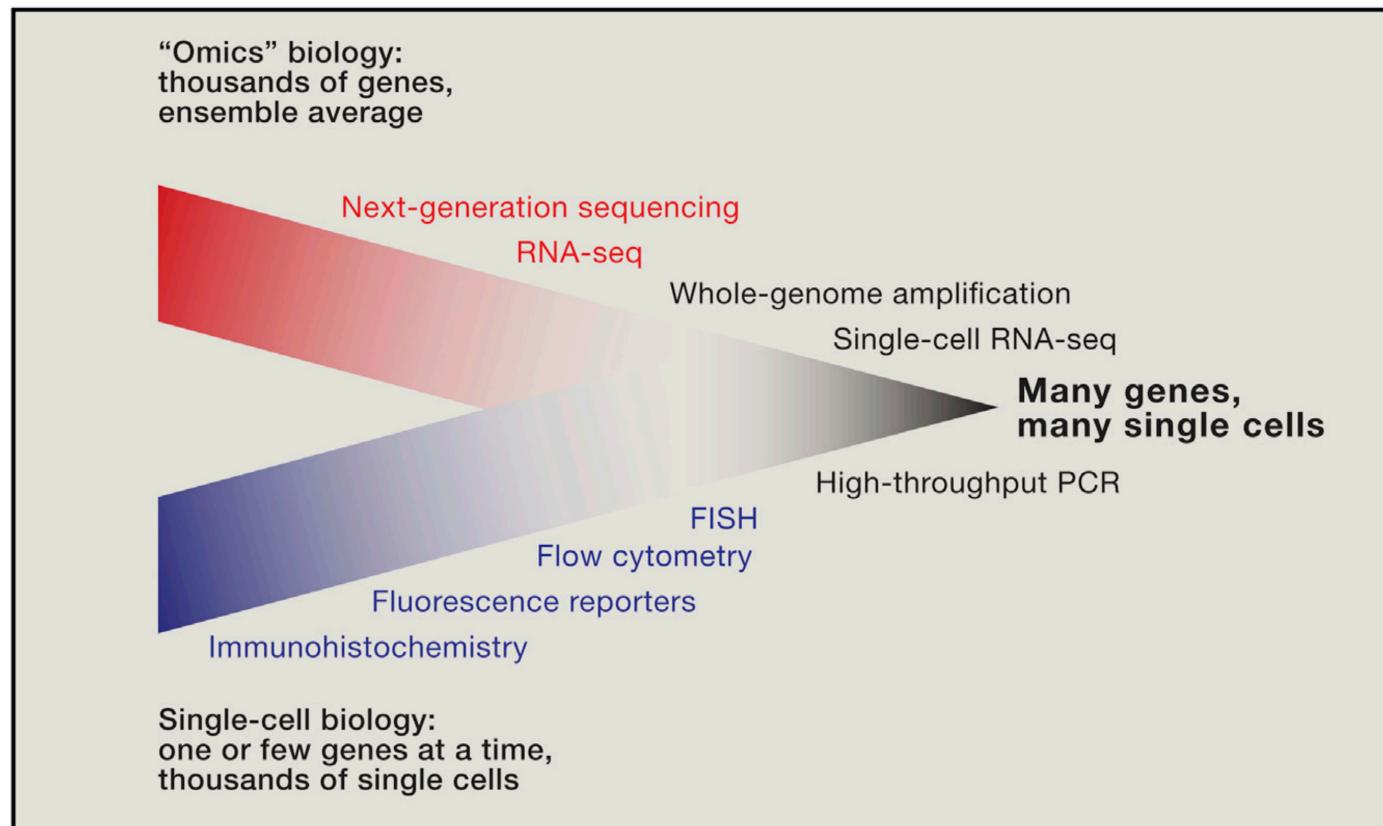
- **Cells are heterogeneous...**in their response to the environment, in their differentiation dynamics, etc
- **Cell heterogeneity can have important effects on the biology.** For example: drug response, cell differentiation, etc
- **Rare cells:** in some cases there's just not enough material to apply bulk techniques. For example, embryos, stem cells, etc

Single-cell-level analyses are needed to explore cellular heterogeneity



- **Cells are heterogeneous...**in their response to the environment, in their differentiation dynamics, etc
- **Cell heterogeneity can have important effects on the biology.** For example: drug response, cell differentiation, etc
- **Rare cells:** in some cases there's just not enough material to apply bulk techniques. For example, embryos, stem cells, etc

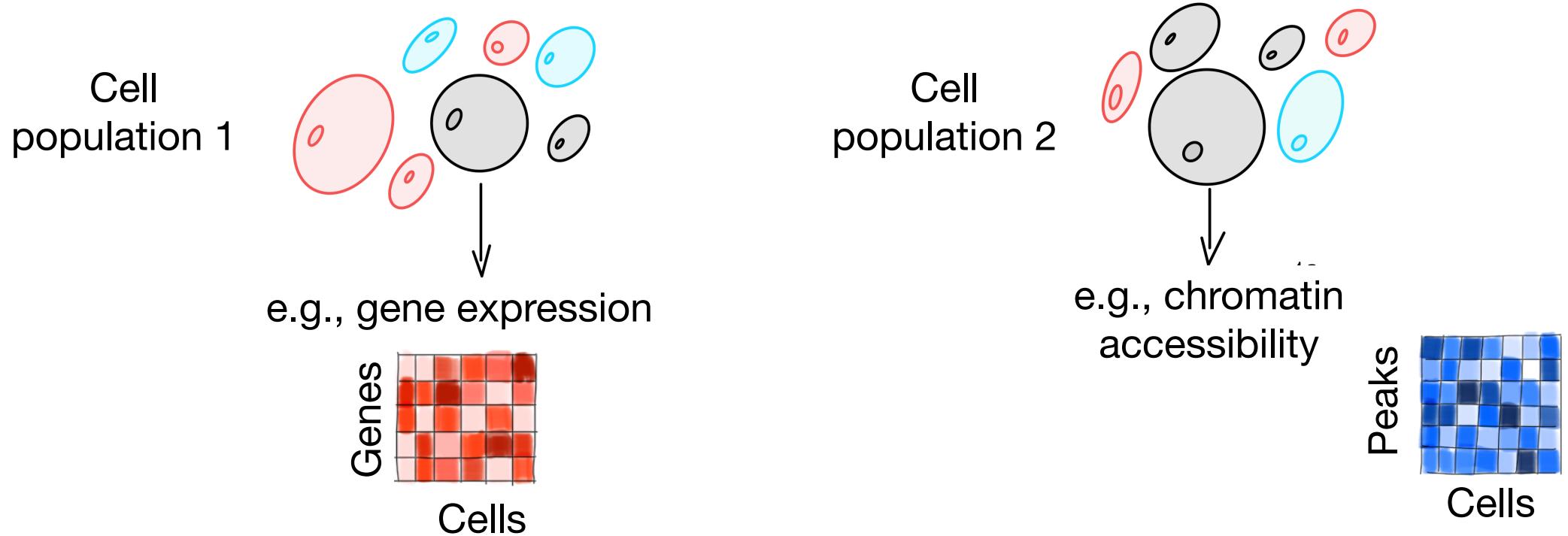
Single-cell omics enables full molecular characterisation of single cells



Convergence of “omics” biology and single-cell biology

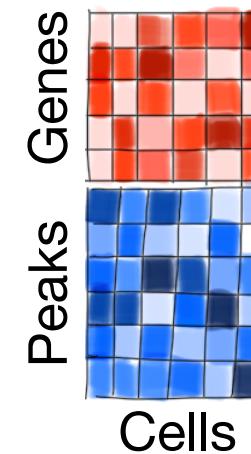
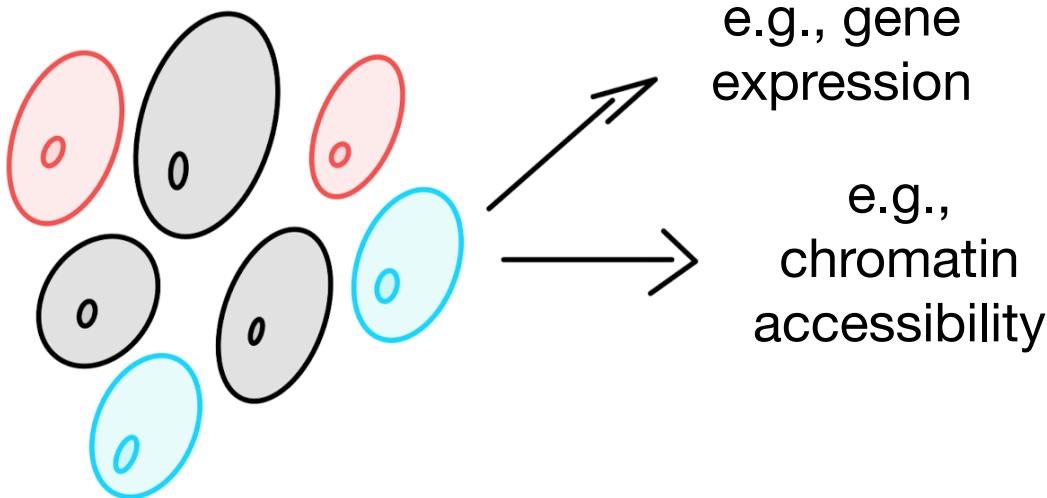
Junker and Van Oudenaarden, Cell, 157:8, 2014

Combining multiple omics modalities from single cells



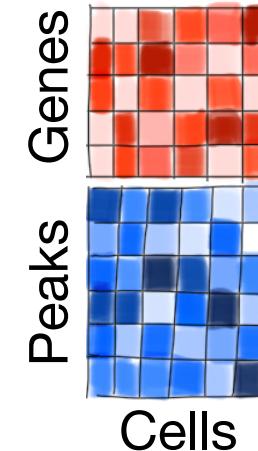
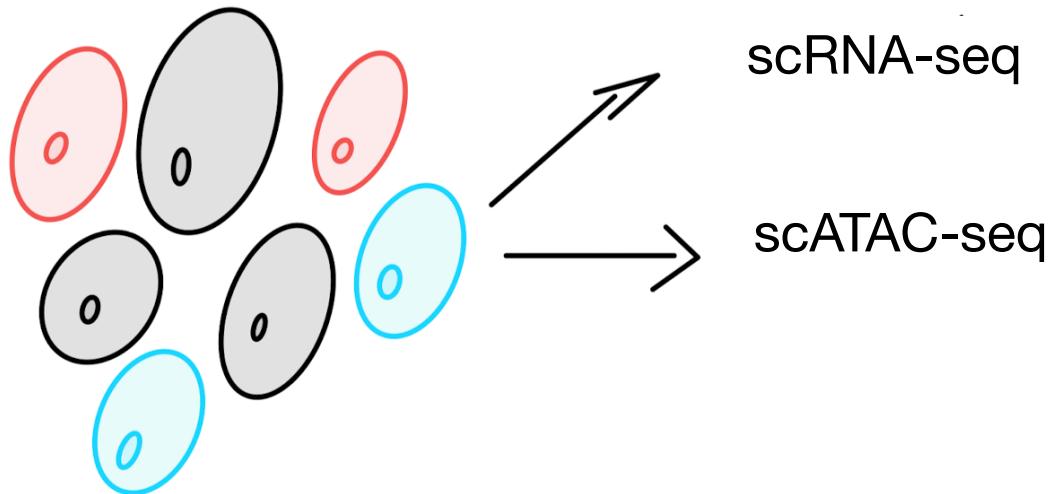
- Example: two different populations of cells, for each we can profile a different omic layer
- We can only combine these two layers computationally, under some strong-ish assumptions (e.g., same cell states present in all populations, etc)

Combining multiple omics modalities from single cells



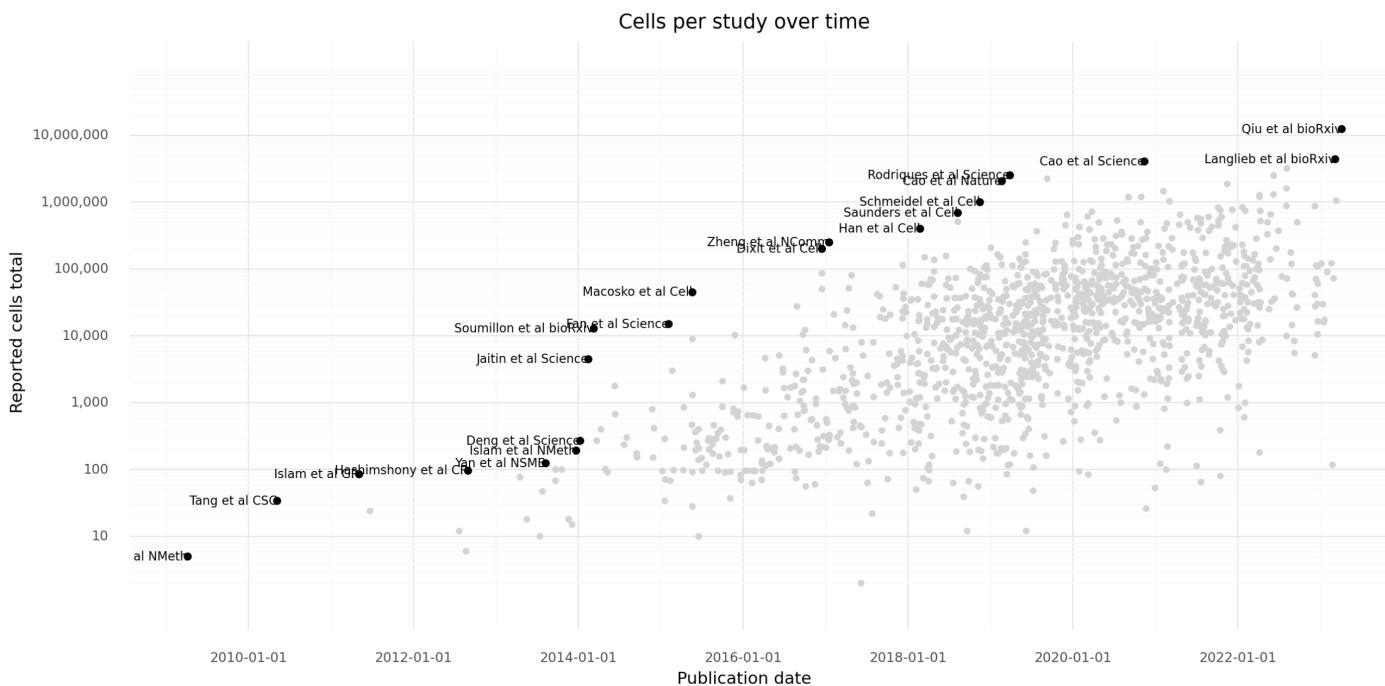
- Single-cell multiomics: quantify different omic layers from the **same single cells**
- Possibility to study the relationship between the different layers and infer causal relationships (example: gene regulatory networks)

Focus on scRNA/ATAC-seq



- Some of the most commonly used omics techniques
- Many open challenges from the computational point of view: e.g., handling of very large datasets, combination of multiple “layers”, etc

Single-cell RNA-seq

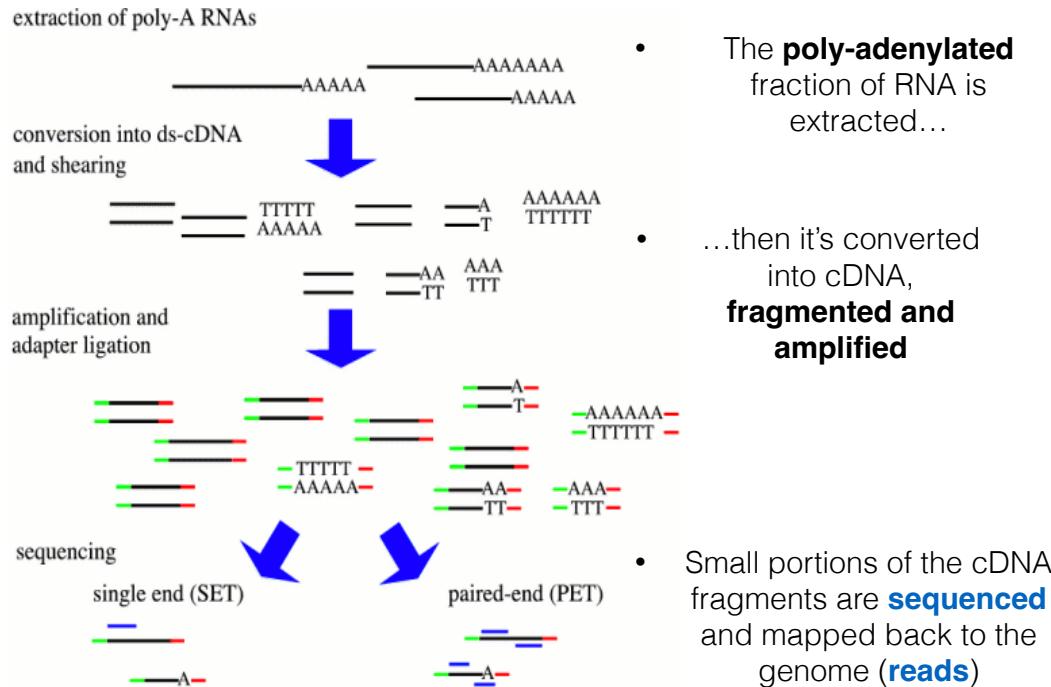


- First publication in 2009
- Fast growth in throughput
- Has sparked the creation of consortia aiming to transcriptomically profile every cell type (e.g., Human Cell Atlas)

Svensson et al, Database, 2020

<https://vals.github.io/single-cell-studies/README.html#with-r>

Single-cell RNA-seq



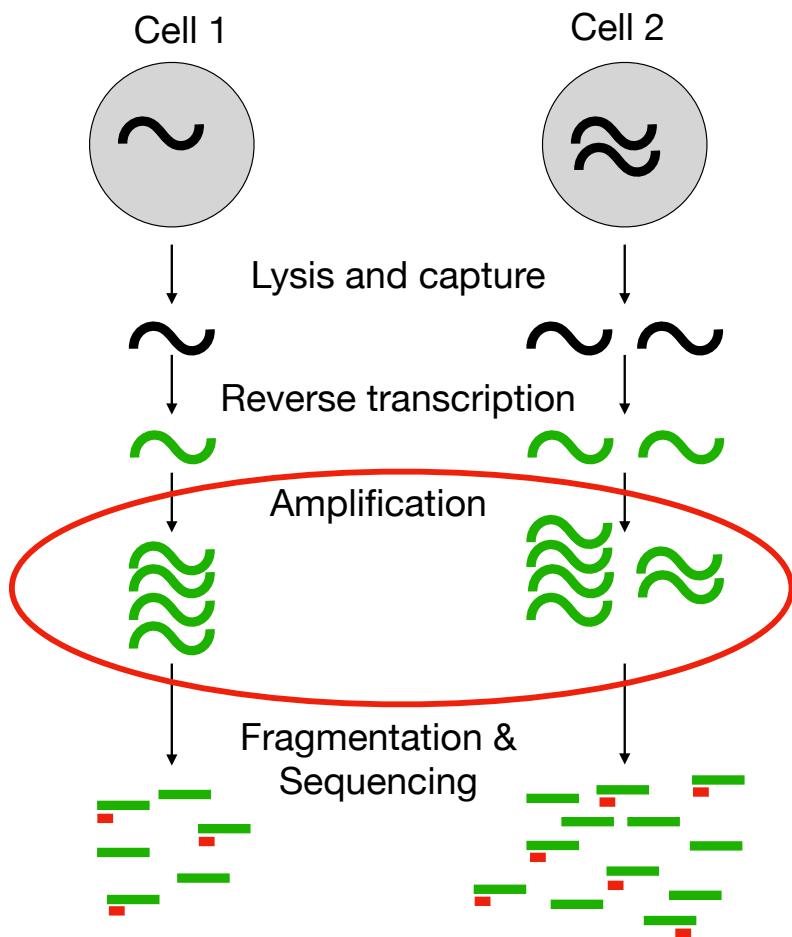
- The **polyadenylated** fraction of RNA is extracted...

- ...then it's converted into cDNA, **fragmented and amplified**

- Small portions of the cDNA fragments are **sequenced** and mapped back to the genome (**reads**)

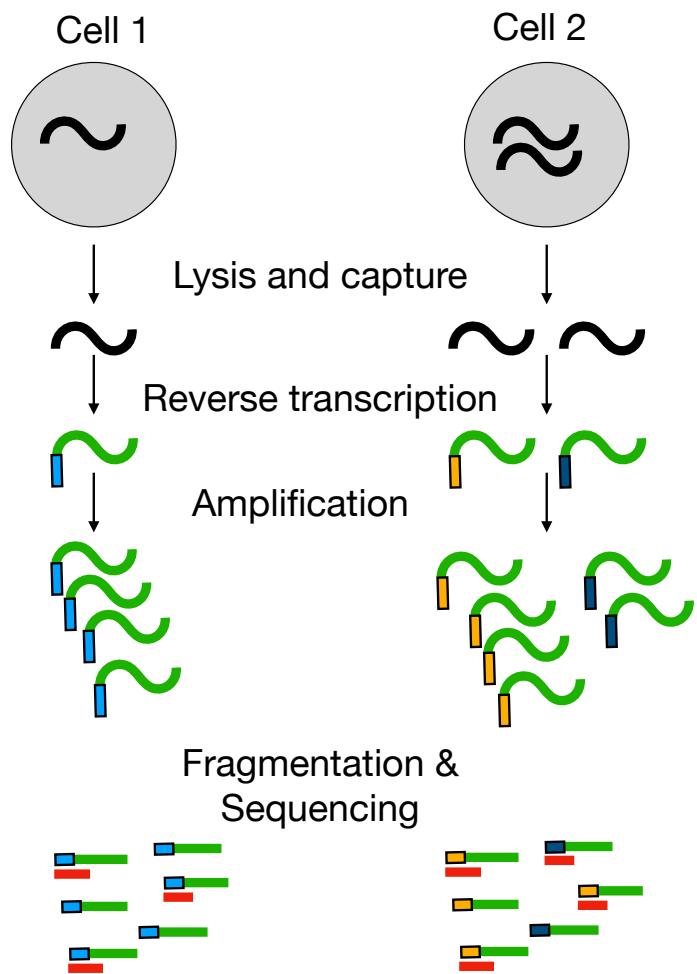
- Unbiased measure of mature mRNA molecules in single cells
- Lots of protocols available
- Possibility to use **Unique Molecular Identifiers**

Single-cell RNA-seq: Unique Molecular Identifiers



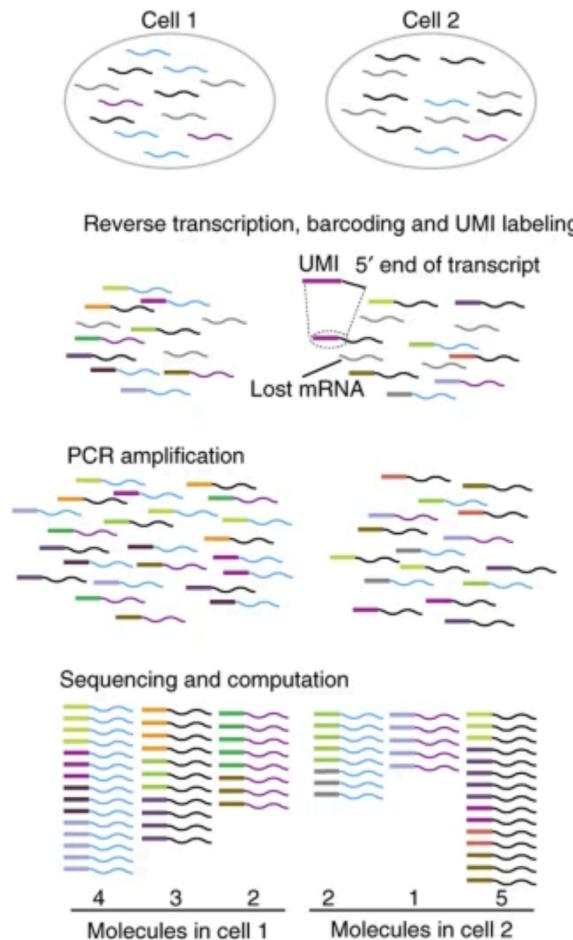
- Each of the steps introduces noise in the system!
- One source of noise is the **amplification step**

Single-cell RNA-seq: Unique Molecular Identifiers



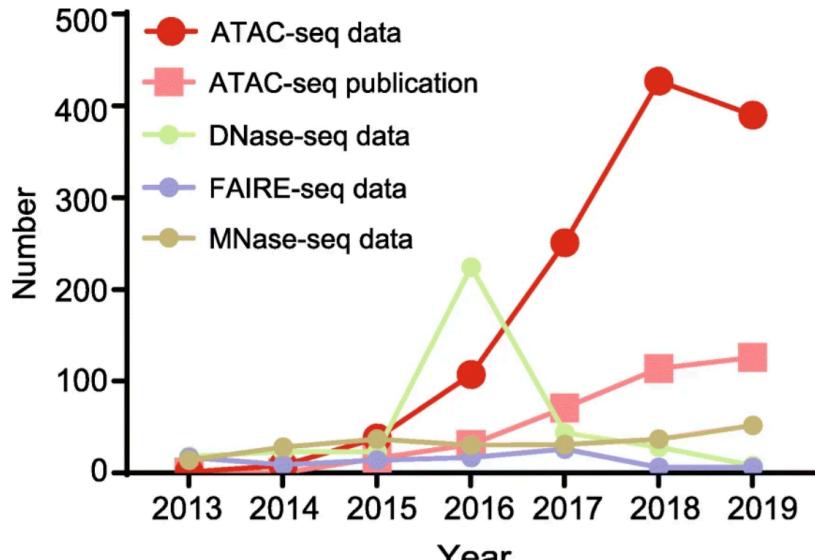
- Unique Molecular Identifiers (UMIs) are random sequences labelling individual molecules
- They reduce **amplification noise**

Single-cell RNA-seq: Unique Molecular Identifiers



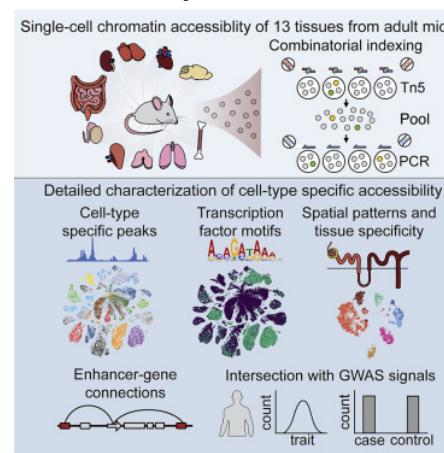
- First published in 2013
- Now many protocols include them

Single-cell ATAC-seq



Yan et al, *Genome Biology*, 2020

- Bulk ATAC-seq first described in 2013, single-cell ATAC-seq in 2015
- Allows for a direct assessment of chromatin accessibility
- Fast (doesn't require overnight incubation)
- Examples of datasets:



Cusanovich et al, *Cell*, 2018

<https://doi.org/10.1038/s41467-021-21583-9> OPEN
Comprehensive analysis of single cell ATAC-seq data with SnapATAC

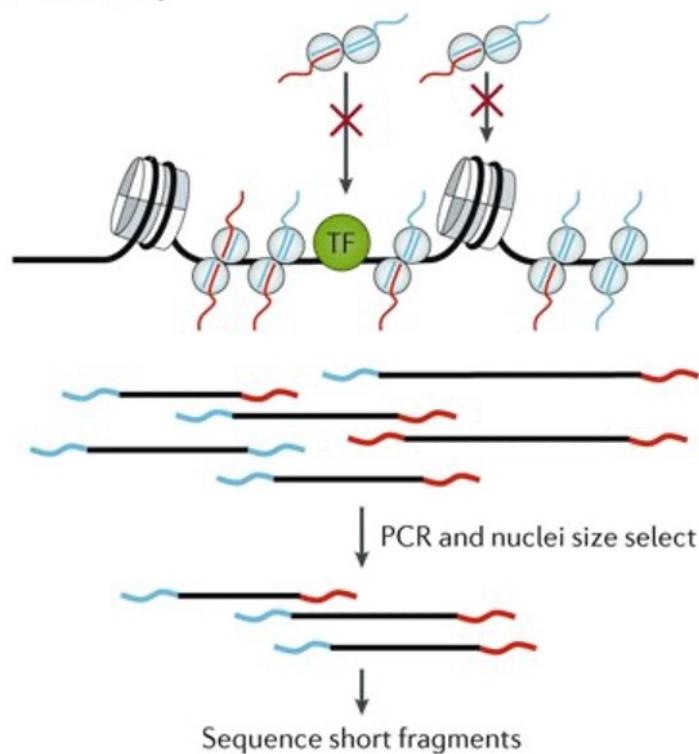
Profiling of ~55k cells from the mouse secondary motor cortex

Fang et al, *NatComm*, 2021

Single-cell ATAC-seq

ATAC = Assay for Transposase Accessible Chromatin

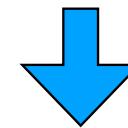
b ATAC-seq



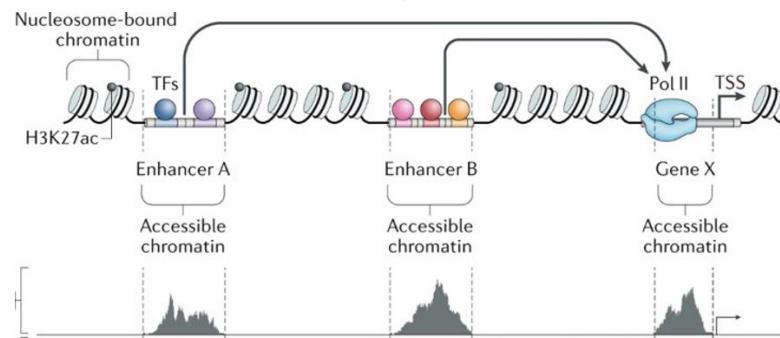
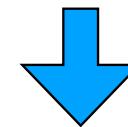
Klemm et al, NatRevGenetics, 2019

Main steps:

- Isolate nuclei
- Expose genomic DNA to hyperactive Tn5 transposes pre-loaded with sequencing adapters
- Isolate and amplify transposed fragments
- Sequencing

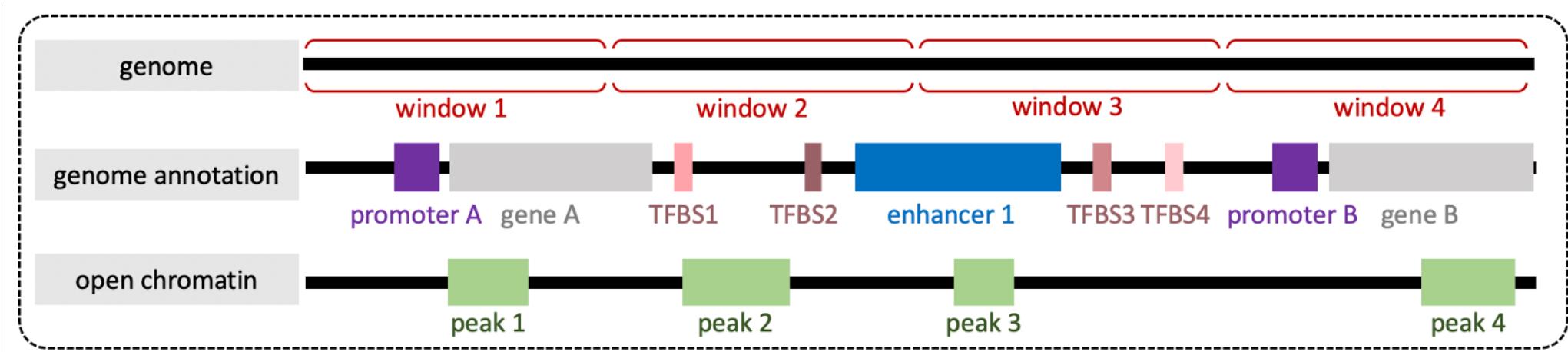


Peaks align to genomic regions with highly accessible chromatin



Single-cell ATAC-seq: feature of interest

- Window based approach
- Annotation based approach



Droplet-based methods for single-cell omics

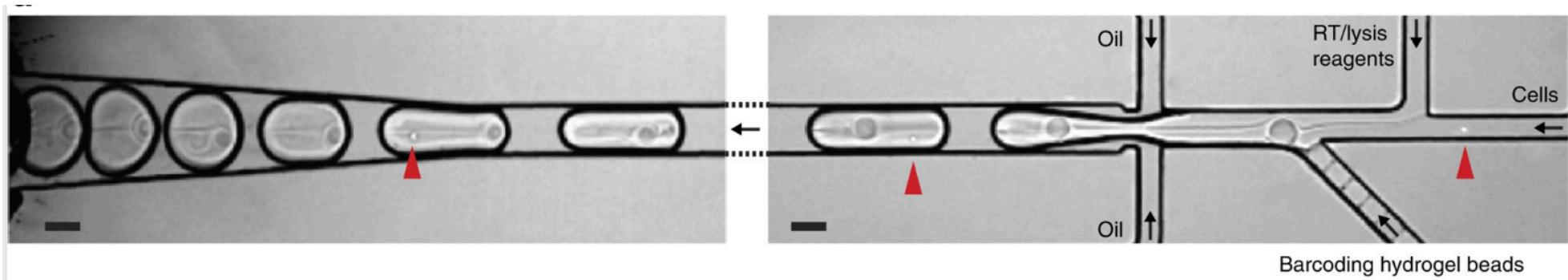
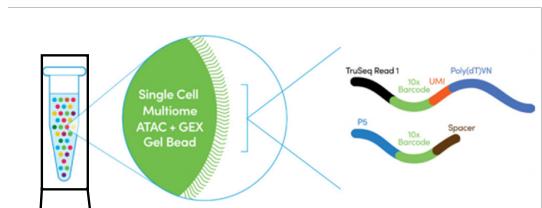


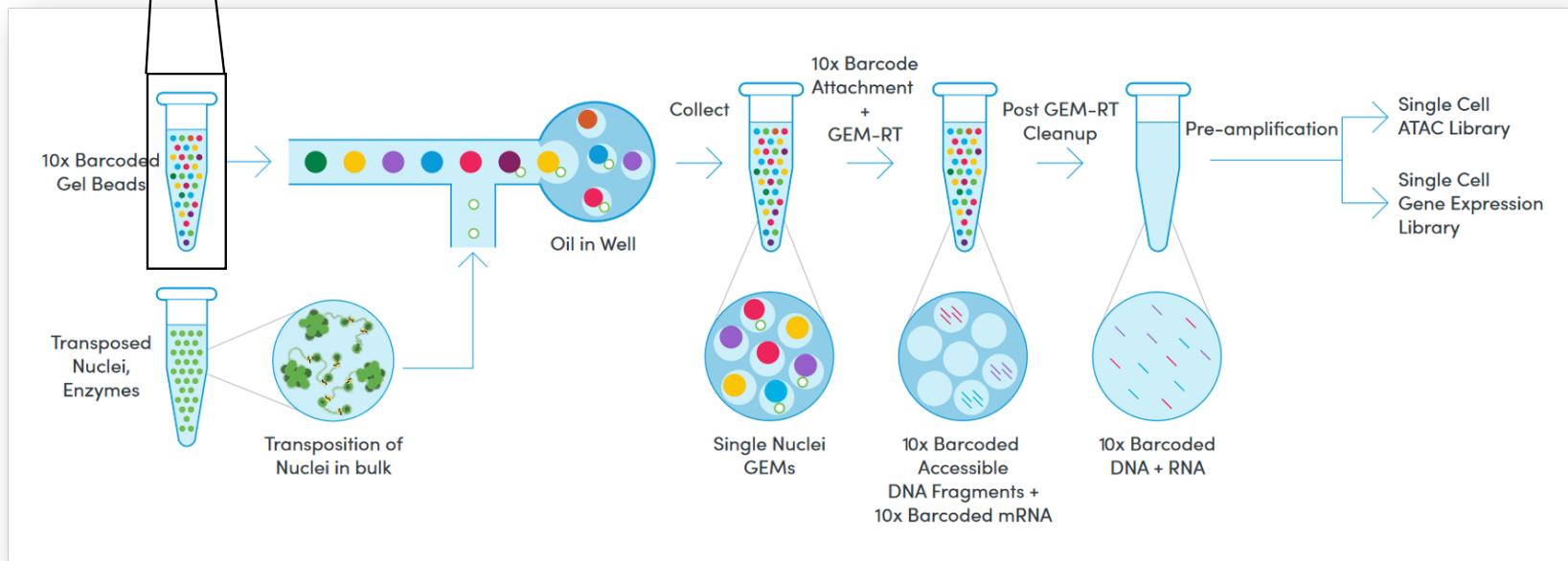
Image from inDrops protocol, Zilionis et al, NatProtocols, 2016

- Encapsulating cells in (sub-microlitre) droplets, alongside **hydrogel beads** with barcodes and reagents
- **Highly scalable** technology to perform high-throughput single-cell omics experiments
- Possibility to perform **single-cell multiomics**

Chromium Single-cell Multiome ATAC + Gene expression



Note: this is single-nuclei RNA-seq (only mRNA in the nuclei are captured)



The dataset we'll analyse

The screenshot shows the 10X Genomics website. At the top, there's a dark blue header with the 10X Genomics logo on the left and navigation links for Products, Research Areas, Resources, Support, and Company on the right. Below the header, a breadcrumb navigation shows 'Support > Single Cell Multiome ATAC + Gene Exp. > Datasets'. To the right of the breadcrumb are links for 'SEARCH', 'Q&A', and 'CONTACT SUPPORT'. The main content area features a section titled 'Frozen human healthy brain tissue (3k)' with a subtitle 'Single Cell Multiome ATAC + Gene Exp. Dataset by Cell Ranger ARC 1.0.0'. Below this, there's a paragraph about the source of the tissue and a detailed description of the isolation and processing protocol. Another paragraph describes the library generation and sequencing process.

Frozen human healthy brain tissue (3k)

Single Cell Multiome ATAC + Gene Exp. Dataset by Cell Ranger ARC 1.0.0

Flash frozen human healthy brain tissue (cerebellum) was acquired from BioVT Asterand®.

Nuclei were isolated from a 2605mg section of flash frozen healthy human brain following the Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + Gene Expression Sequencing demonstrated protocol (CG000375). The isolated nuclei were stained with 7-AAD and then flow sorted using the BD FACSMelody™ cell sorter to clean up the nuclei from the debris. The 7-AAD positive sorted nuclei were then permeabilized using lysis buffer containing digitonin as per the demonstrated protocol.

Paired ATAC and gene expression libraries were generated from the isolated nuclei as described in the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide (CG000338 Rev A) and sequenced on Illumina Novaseq 6000 v1 Kit (Forward Strand Dual-Index Workflow).

Single-cell multiome data from frozen human healthy brain tissue (~3k cells)

The software we will use



For scRNA-seq:



Hao et al, NatBiotech, 2023

For scATAC-seq:



Stuart et al, NatMethods, 2021

- Storing data in well-structured objects
- Pre-made functions to perform many analyses, make plots, etc

Other popular choices:



Packages should not be used as “black boxes”!



New Results

[Follow this preprint](#)

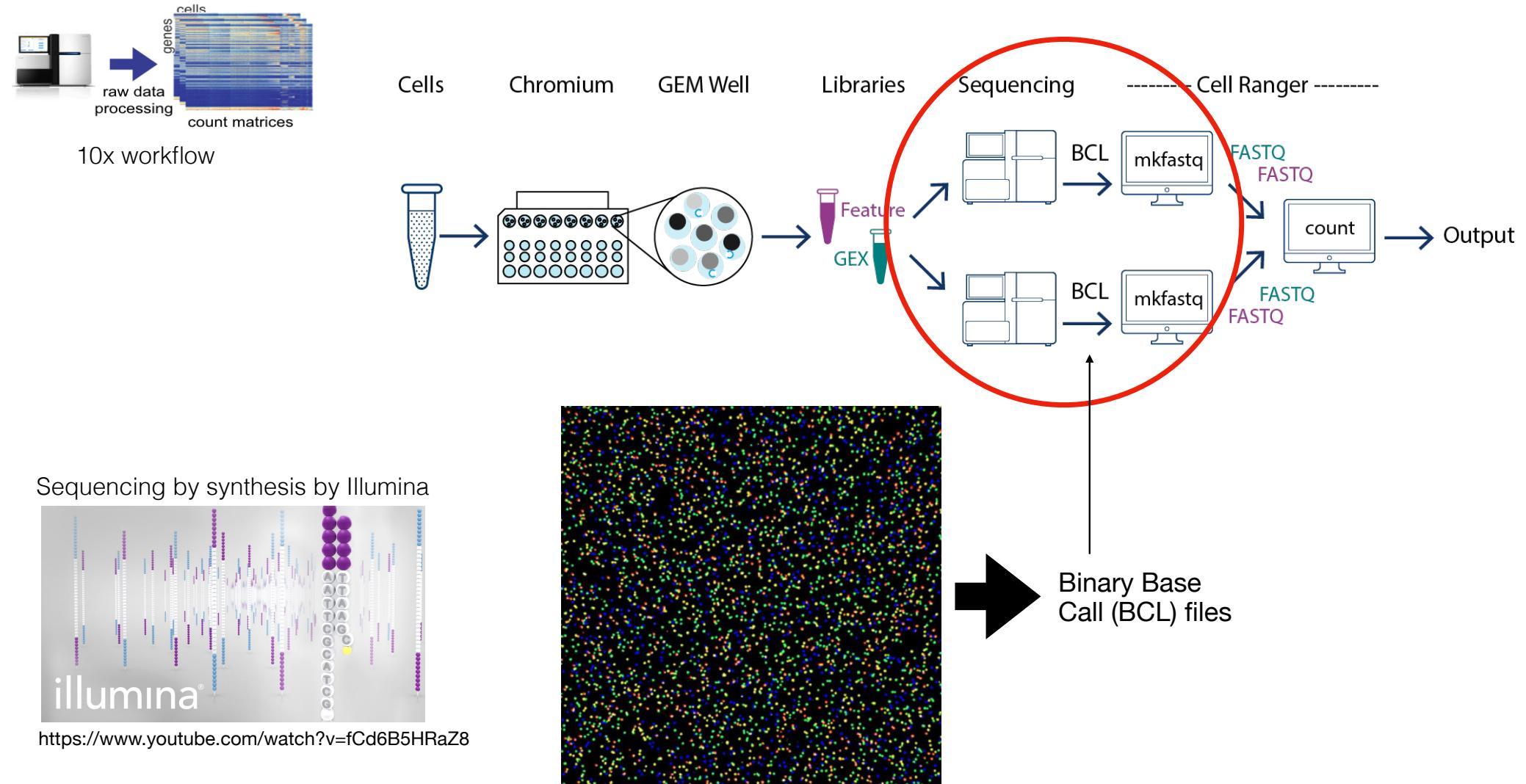
The impact of package selection and versioning on single-cell RNA-seq analysis

Joseph M Rich, Lambda Moses, Pétur Helgi Einarsson, Kayla Jackson, Laura Luebbert, A. Sina Booeshaghi, Sindri Antonsson, Delaney K. Sullivan, Nicolas Bray, Pál Melsted, Lior Pachter

<https://doi.org/10.1101/2024.04.04.588111>

- Always have full control and awareness of what the function does to your data and what is plotted
- Any function you use typically has many parameters: never stop at “default” values

First step: library sequencing

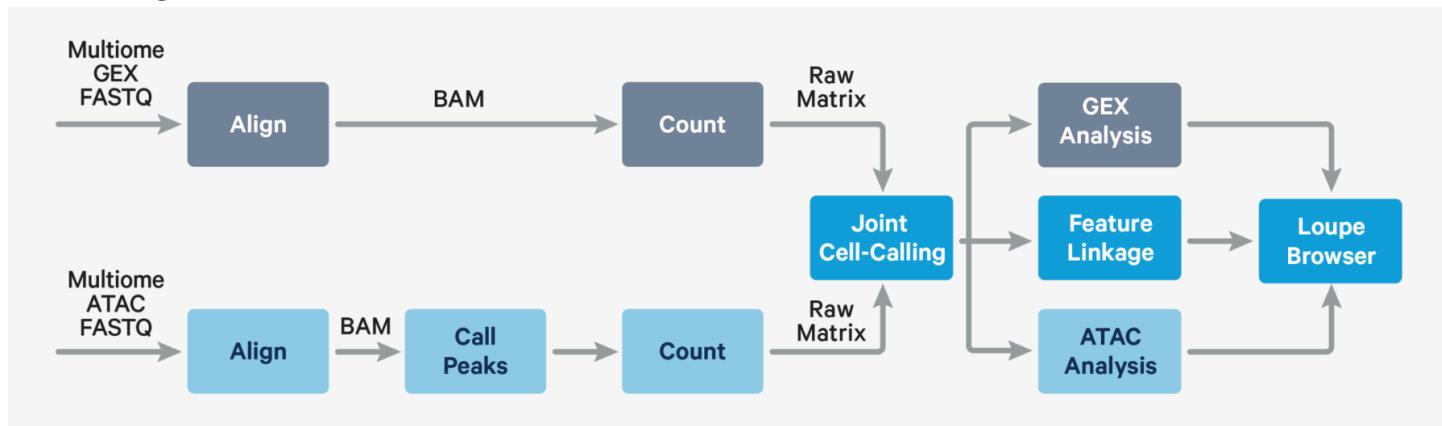


Processing FASTQ files with cellRanger

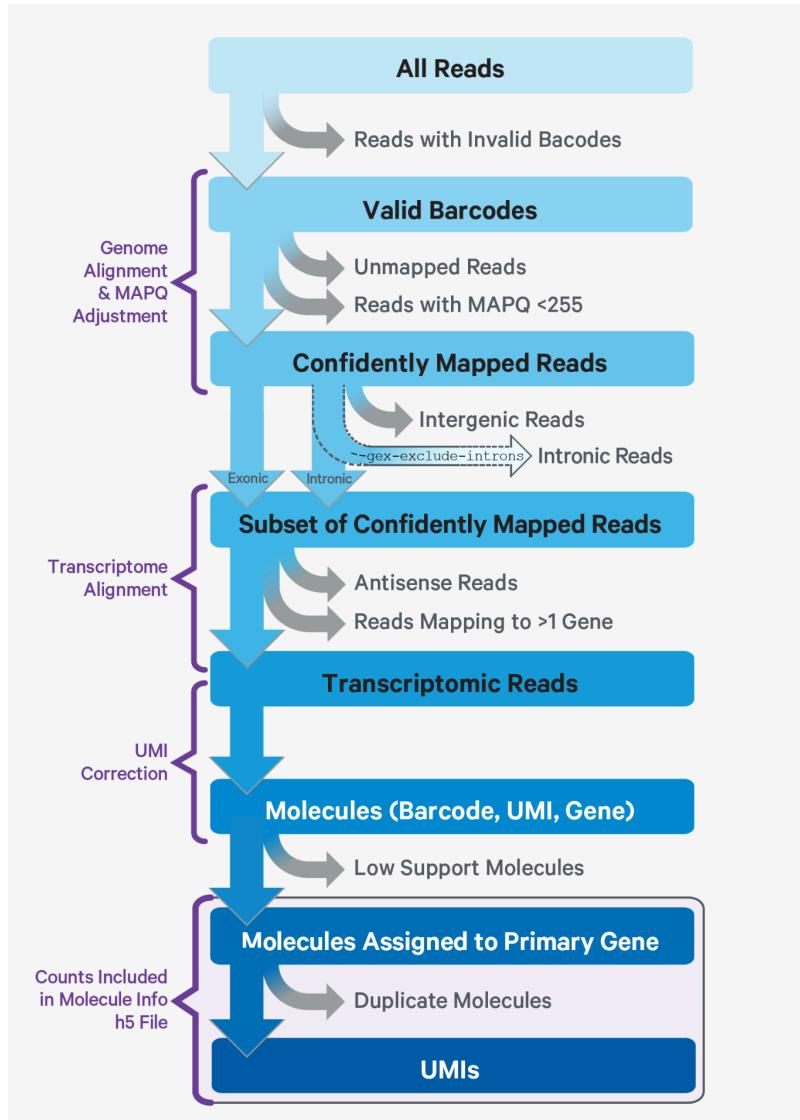
Starting point for both RNA-seq and ATAC-seq computational data analysis: FASTQ files

- Cell Barcodes
 - UMI (for RNA-seq)

“cellRanger-arc count” workflow

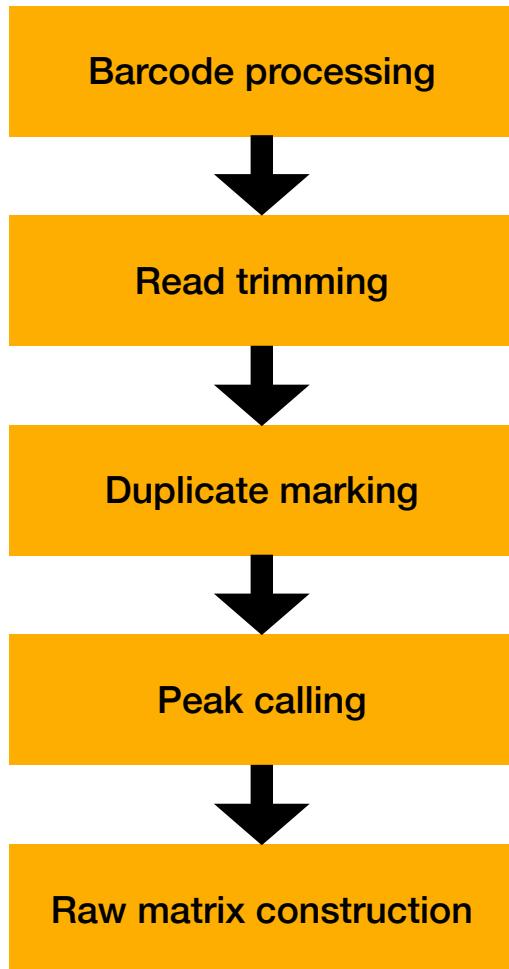


RNA-seq pipeline

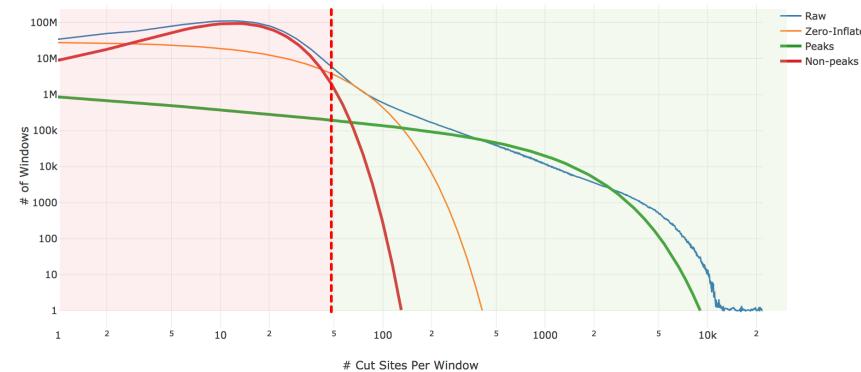


Output is a file including rows with:
Barcode; UMI; gene; count

ATAC-seq pipeline

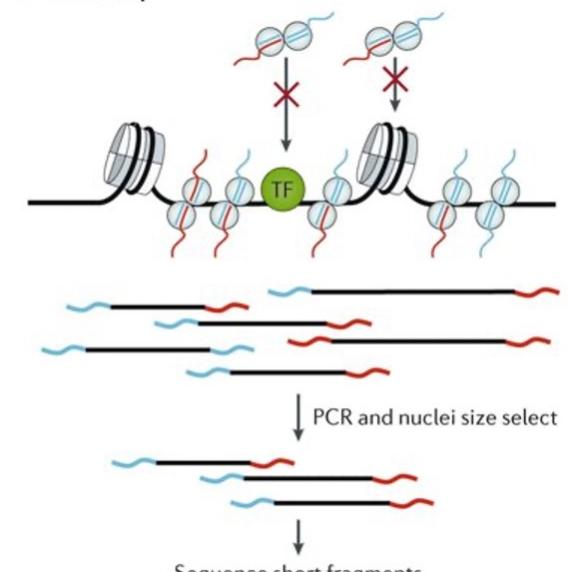


- Goal: identify which genomic regions are likely to be the key features (=peaks)
- The patterns of chromatin accessibility depend on the cells present in the sample; hence, **the peaks have to be called directly from the data**
- First step is **pooling reads from all barcodes** to increase signal, then counting “cut-sites” in a given window on the genome



Output is a file where rows are peaks, columns are barcodes, and each entry is the number of cut-sites within the corresponding peak for a given barcode

b ATAC-seq



Klemm et al, NatRevGenetics, 2019

- Fitting of the signal to identify candidate peaks
- Further filtering based on a locally measured signal-to-noise ratio

Joint cell calling

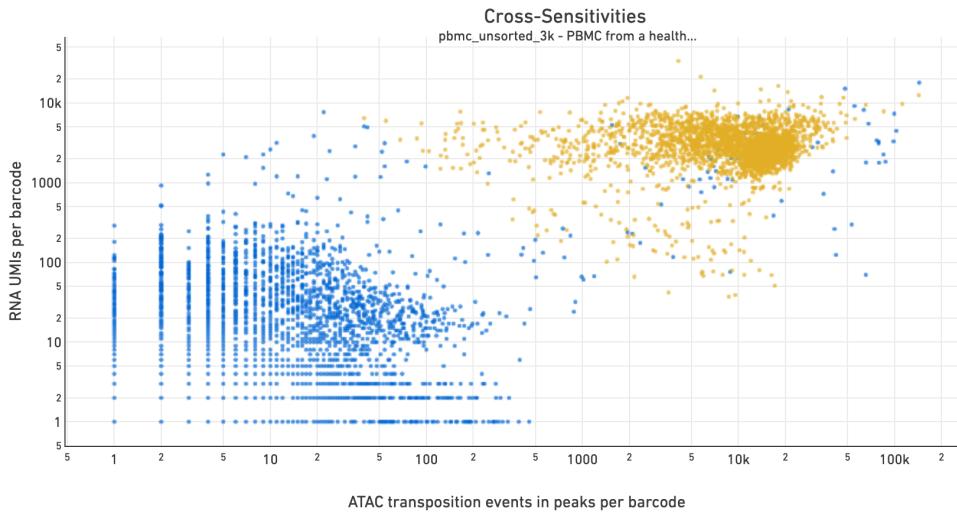
For each barcode, we have an associated count of transposition events in peaks (ATAC-seq) and gene expression UMIs (RNA-seq)

Which barcodes correspond to cells and which are background signal?

Filtering (e.g., to eliminate putative gel bead doublets, or barcodes not observed in both assays)

Cell calling

Joint Cell Calling



- First estimation with “ordmag” algorithm
- Refining with K-means clustering to split the set of barcodes into two groups: “cells” and “non-cells”

Joint Cell Calling



Finally, count matrices!

RNA-seq: Gene count table

Gene ID	Cell_1	Cell_2	Cell_3
ENSMUSG000000000001	1	0	3
ENSMUSG000000000003	2	0	0
ENSMUSG000000000028	1	0	10
ENSMUSG000000000031	0	2	3

...

ATAC-seq: Peak counts

Peak position	Cell_1	Cell_2	Cell_3
chr1:180995-181723	0	1	0
chr1:191025-191934	0	0	0
chr1:629469-630395	0	0	0
chr1:633551-6344751	2	0	0

...

- #genes ~25k - 50k
- #peaks ~ 10^5
- #cells typically ~10k-100k
- Very sparse: ~90% zeros (“dropouts”)