# Introduction to Single-Cell RNA-seq

The CCDL

# What can bulk RNA-seq vs single-cell RNA-seq help us determine?

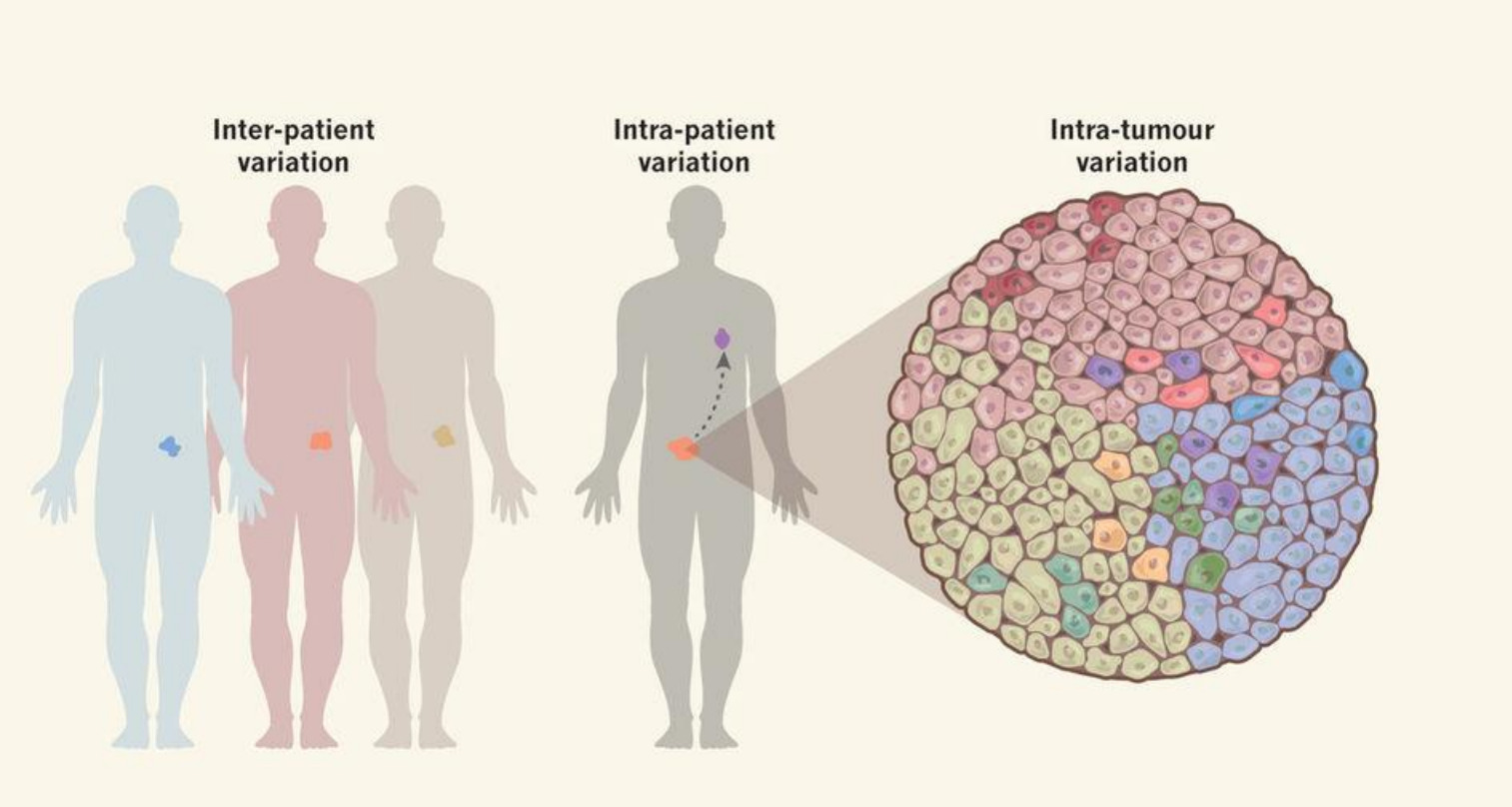



Image from Tanaka *et al*. 2018 https://doi.org/10.1038/s41551-017-0162-1.

# What can bulk RNA-seq vs single-cell RNA-seq help us determine?

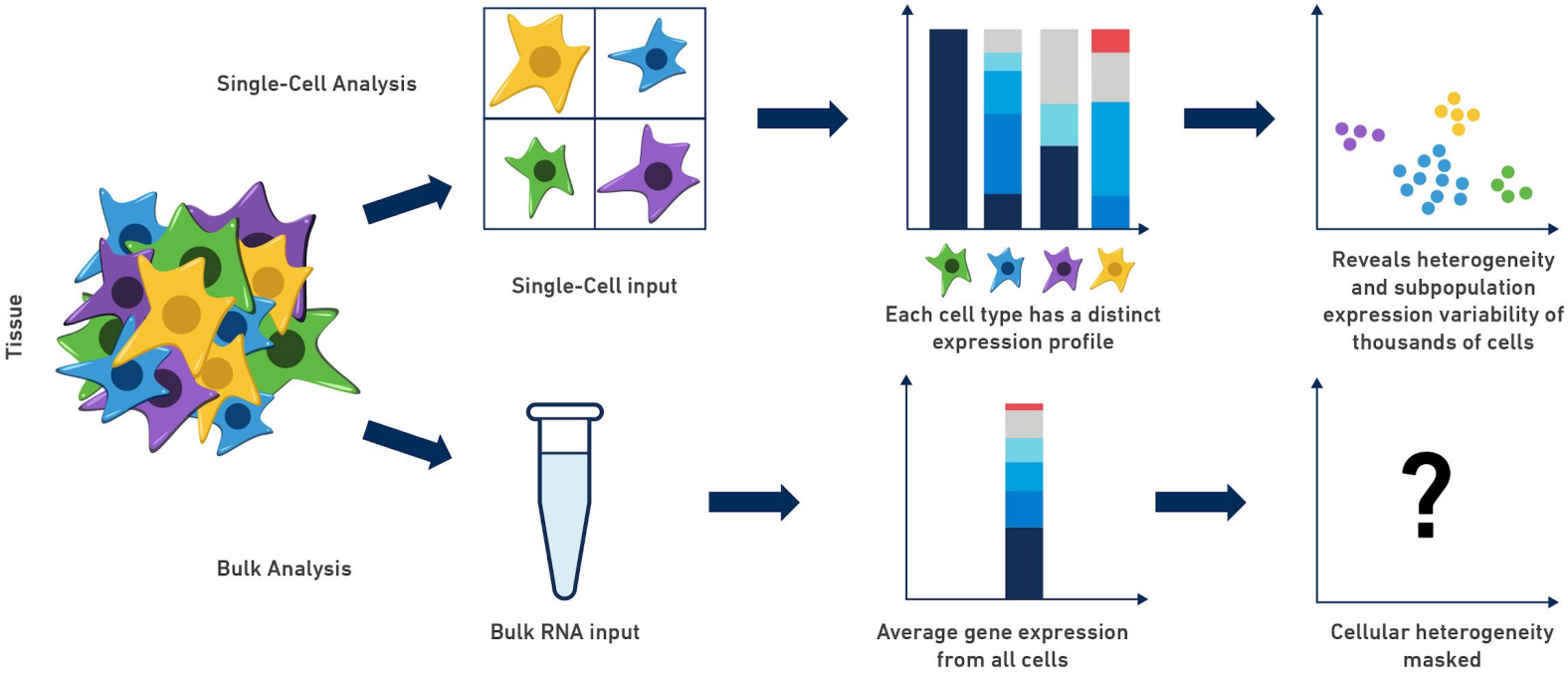



Image from 10X Genomics blog:
https://community.10xgenomics.com/t5/10x-Blog/Single-Cell-RNA-Seq-An-Introductory-Overview-and-Tools-for/ba-p/547

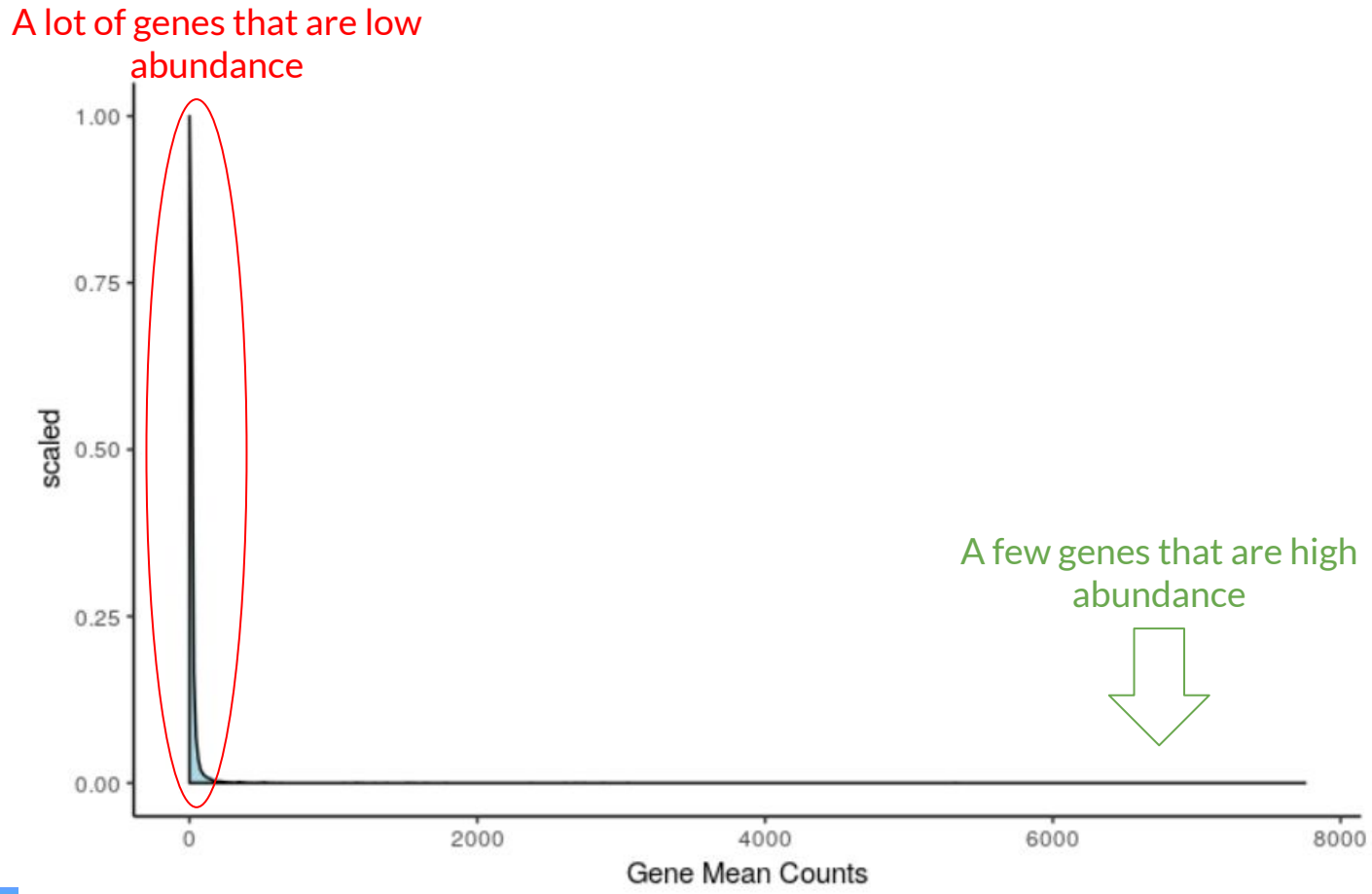
# Single-cell RNA-seq quirks

Less starting material means:

- More PCR amplification (*and its associated biases*)
- More zero counts
  - Biology - Not every gene is expressed in every cell
  - Technical - Biased capture methods, Sequencing every RNA in every cell requires a lot more sequencing

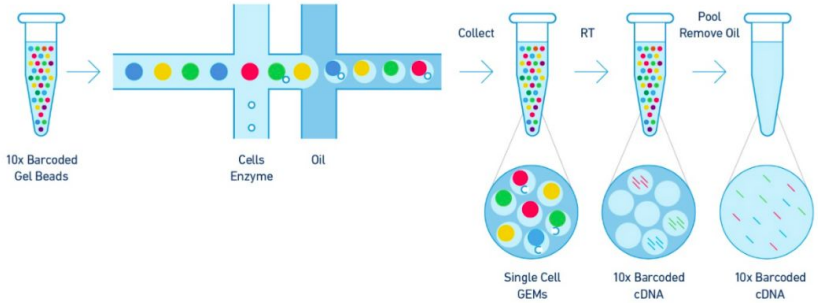
Choi *et al.* (Genome Biology, 2020) https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02103-2

# Single-cell gene mean density graph



# Single Cell Basic Set-ups

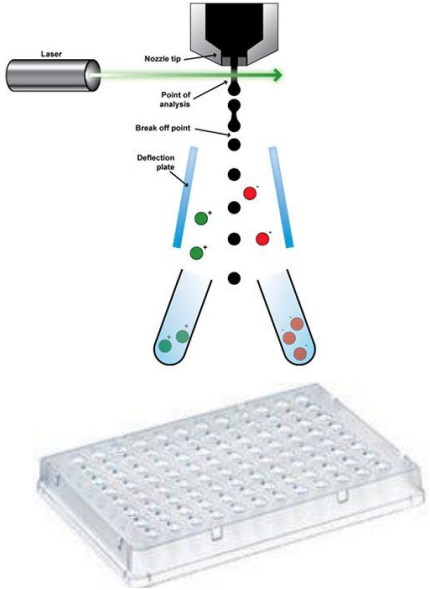
## 1. Tag-based scRNA-seq




Tag-based separation of cells' data *after* sequencing

**Example: 10X Genomics Chromium**
Zheng *et al.* 2017
https://www.ncbi.nlm.nih.gov/pubmed/28091601

## 2. Full-length scRNA-seq




Physical separation of cells *before* sequencing

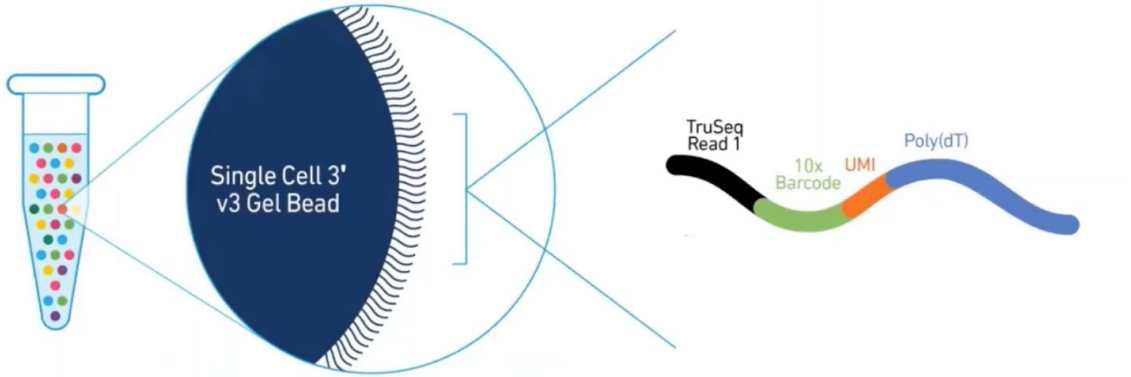
**Example: Smart-seq2**
Picelli *et al.* 2014
https://www.nature.com/articles/nprot.2014.006

# Cell Barcodes + Unique Molecular Identifiers (UMIs) are used to label individual transcripts

Each droplet contains 1 cell, all with the same cell barcode

Within each droplet, each bead contains millions of distinct UMIs

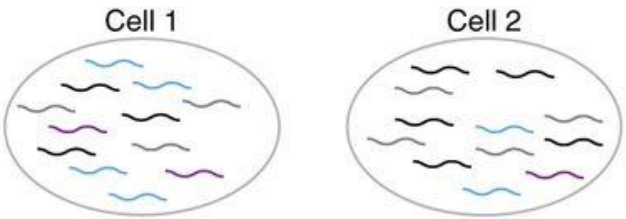
Each transcript within a cell is tagged with a cell barcode and unique molecular identifier (UMI)



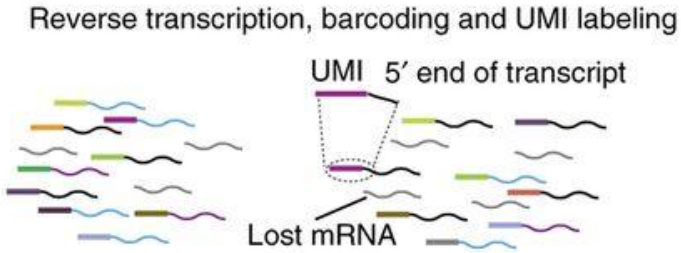
# Unique Molecular Identifiers (UMIs):

a '*snapshot*' of the original molecules in the *pre-amplified* cell

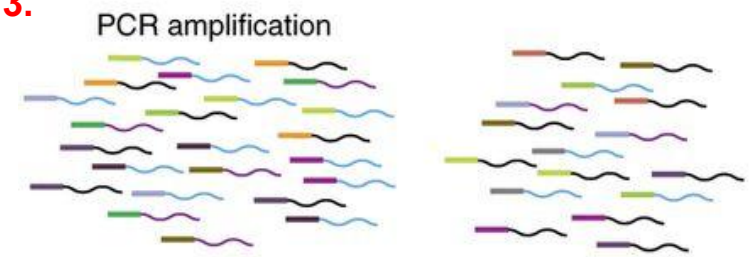
**1.**



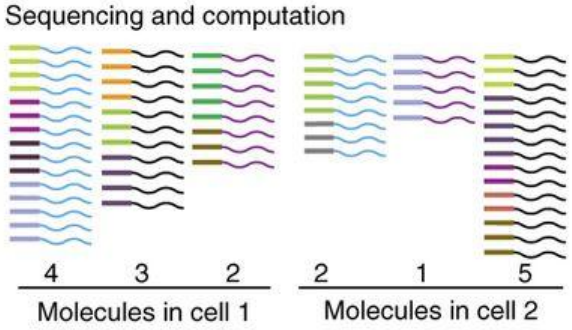

**2.**




**3.**

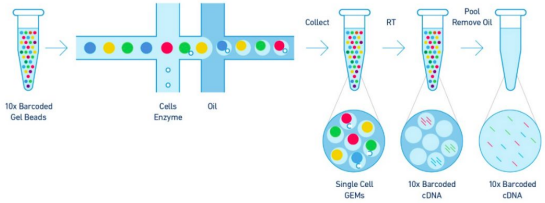



**4.**




Original image from: Islam *et al.* 2014 https://doi.org/10.1038/nmeth.2772

# **Tag-Based** scRNA-seq




Pros:

- Can profile up to millions of cells.
- Takes less computing power.
- File storage requirements are smaller.
- Much less expensive.

Cons:

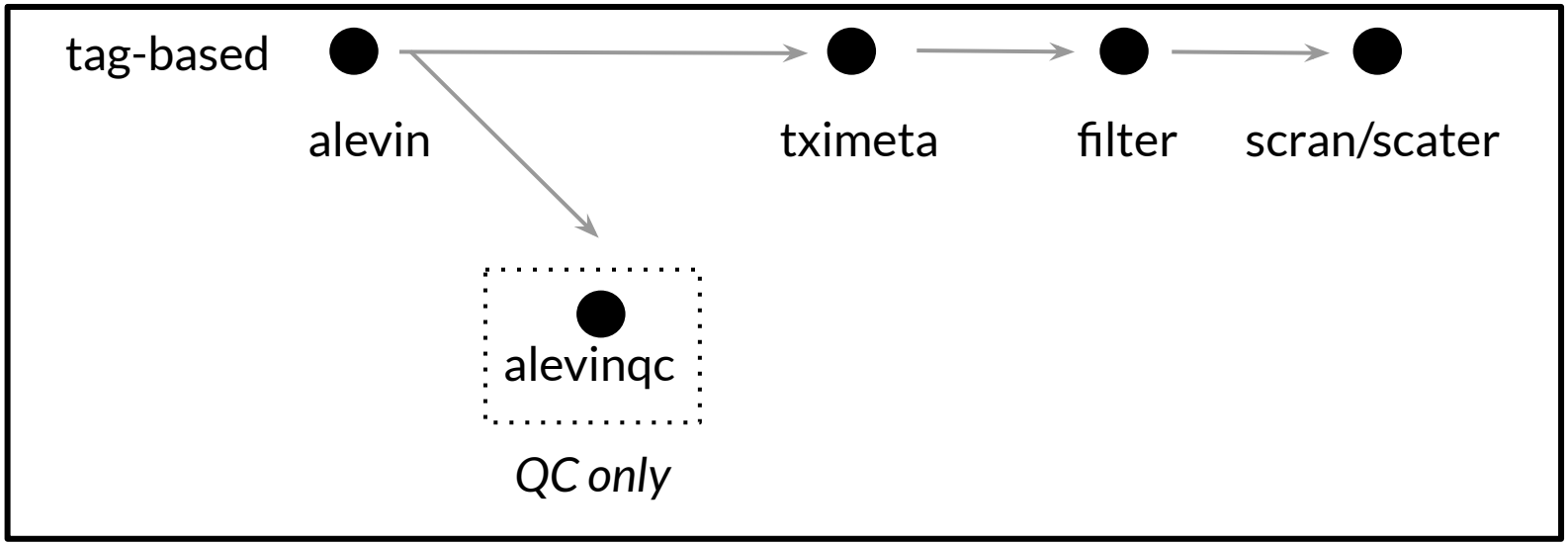
- More intense 3' bias because sequencing is not bidirectional.
- Coverage is generally not as deep as full-length scRNA-seq.

# Pre-processing scRNA-seq


Step 1: Separate data by cell barcode and UMI
Step 2: Align reads to determine genes present in each cell
Step 3: Collapse duplicate UMI's to create gene x cell count matrix

- Many different pre-processing tools are available
- Cell Ranger, 10X supported tool, is popular mainly for being user friendly, but is very slow (aligns to the entire genome)
- Alevin is a faster salmon based pre-processing tool (aligns to the transcriptome)

Comparison of common alignment tools: https://www.biorxiv.org/content/10.1101/2021.02.15.430948v2


tag-based
alevin
tximeta
filter
scran/scater
alevinqc
QC only

# Resources for you in `00-scRNA-seq_introduction.md`

- Hemburg lab scRNA-seq training course
- ASAP: Automated Single-cell Analysis Pipeline is a web server that allows you to process scRNA-seq data.
- Smith. Unique Molecular Identifiers – the problem, the solution and the proof - article on background of UMIs
- Literature on technologies