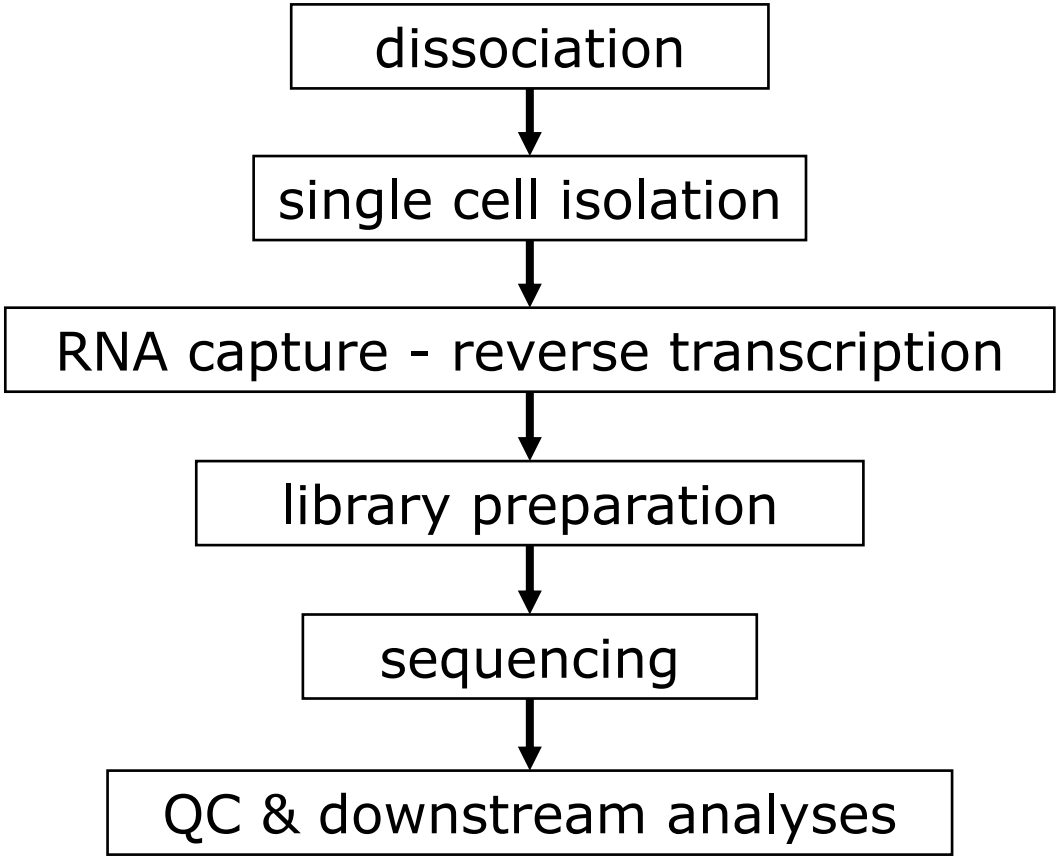
Single cell transcriptomics

Introduction to single cell RNA-seq

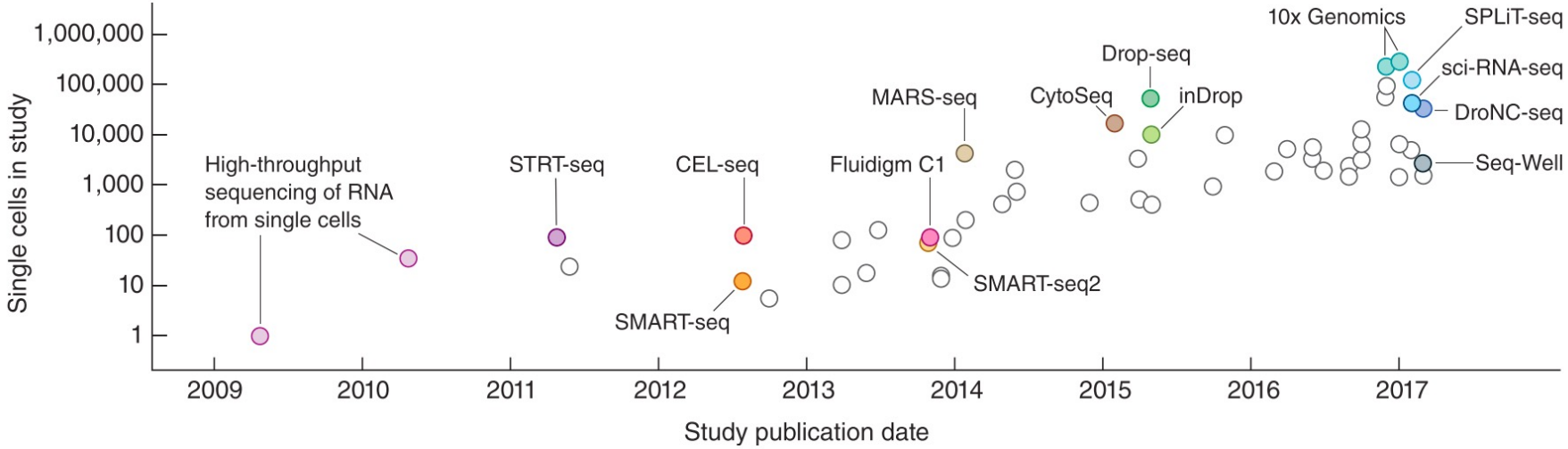
Why single cell RNAseq?

- Cell specific effects of treatments
- Differential gene expression between and within cell types
- Understand:
  - drug action/delivery
  - cell fate
  - cell types affected by (viral) infection
  - regulation through co-expression
  - differences in cell populations (between tissues)
  - …

scRNA-seq workflow



Technologies



- Plate-based: SMART-seq
- Droplet-based: 10x genomics (3’ kit)

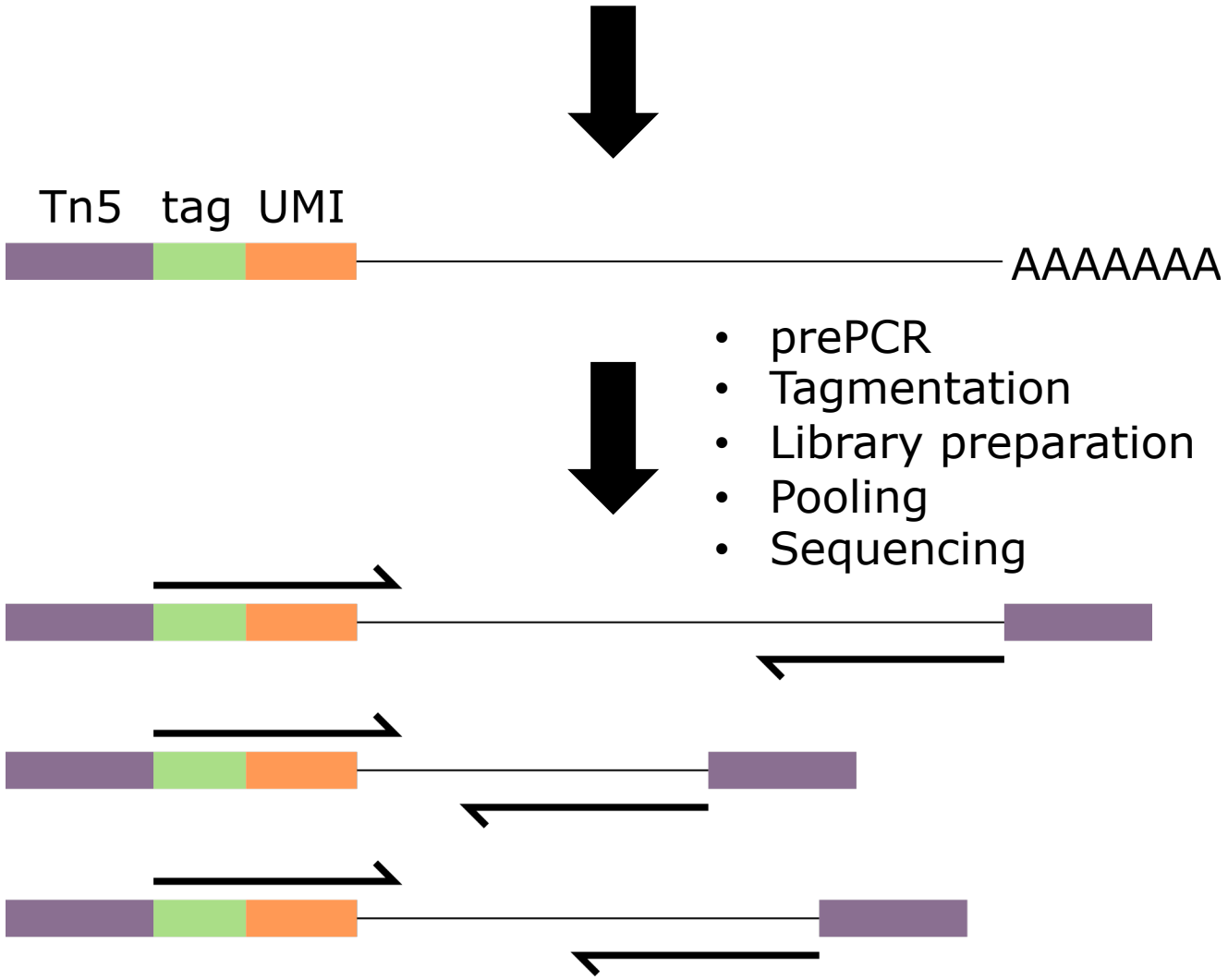
Svensson V et al., Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc. 2018;13:599–604.

SMART-seq

- Use FACS for: 1 well = 1 cell
- Library preparation per cell
- Whole gene can be sequenced

Hagemann-Jensen M et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nat Biotechnol. 2020;38:708–14.

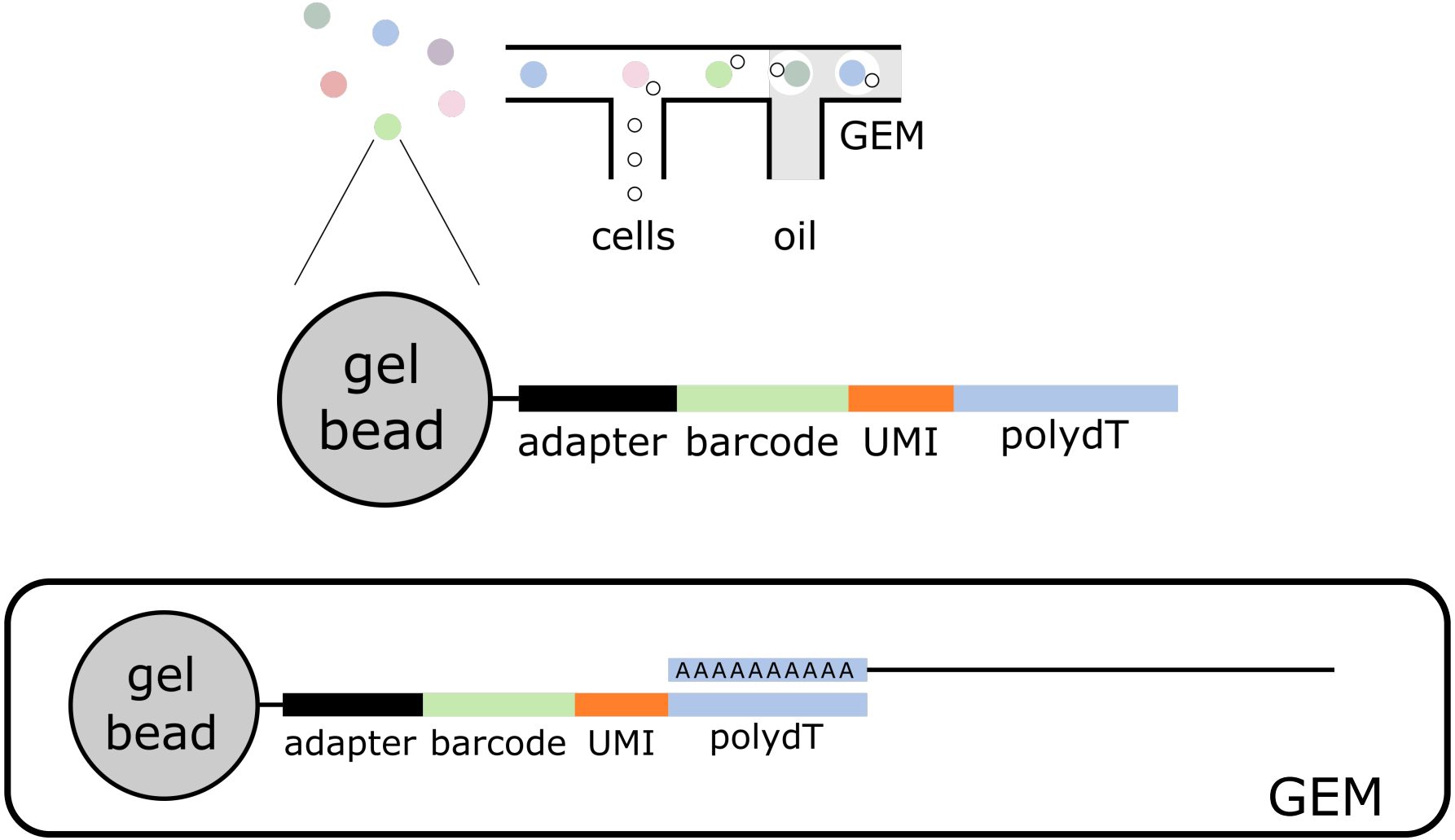
cell lysing + reverse transcription + template switching



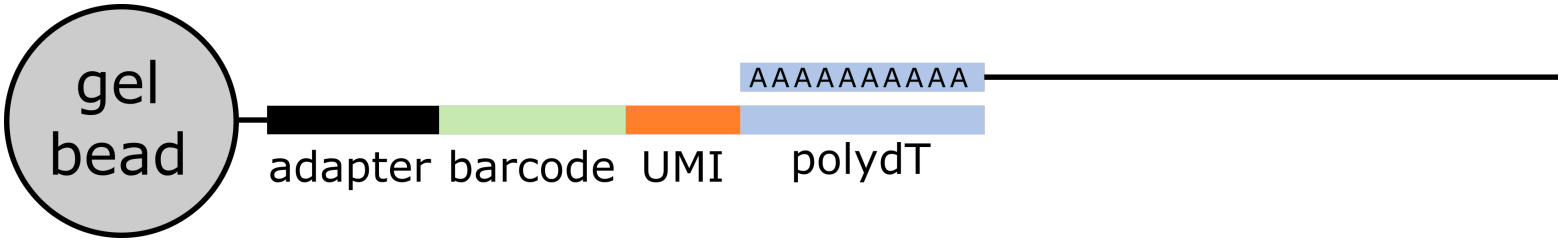
10x genomics (3’ kit)

- Cell isolation within oil droplet
- RNA capture with 1 bead/cell
- Sequencing from 3’ end
- Only expression

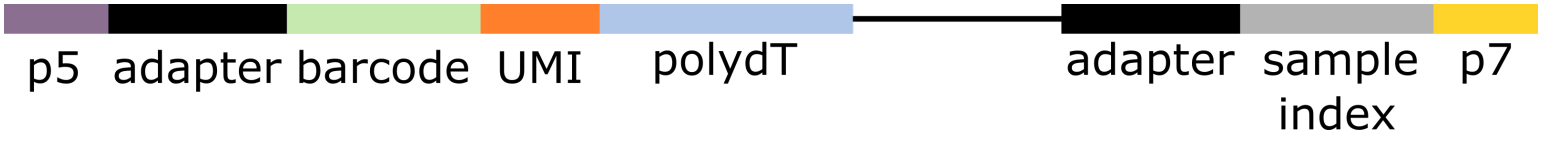
Zheng GXY et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun; 2017;8.

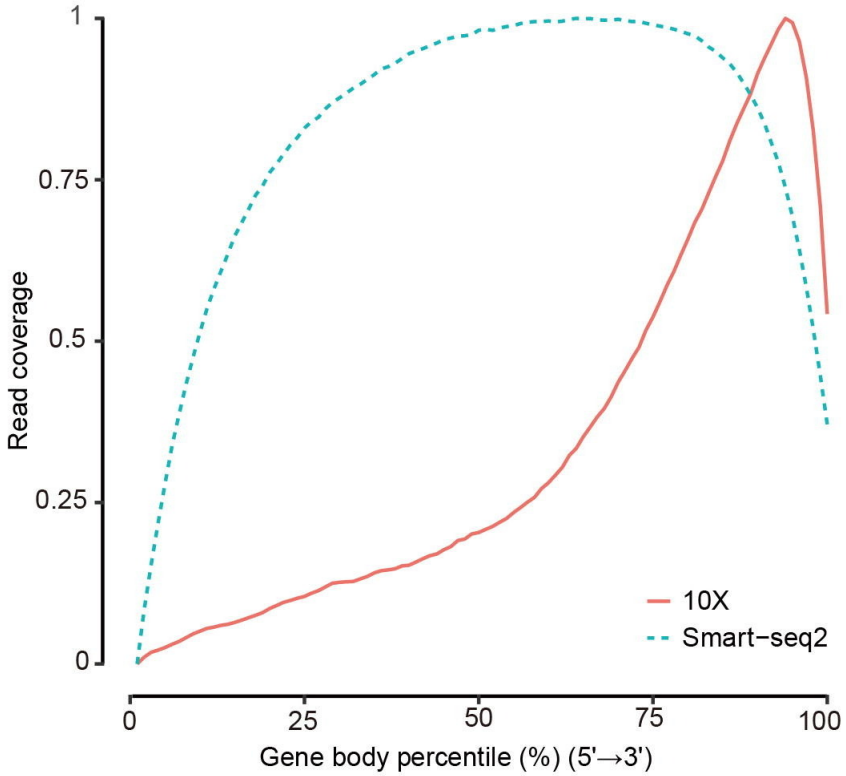


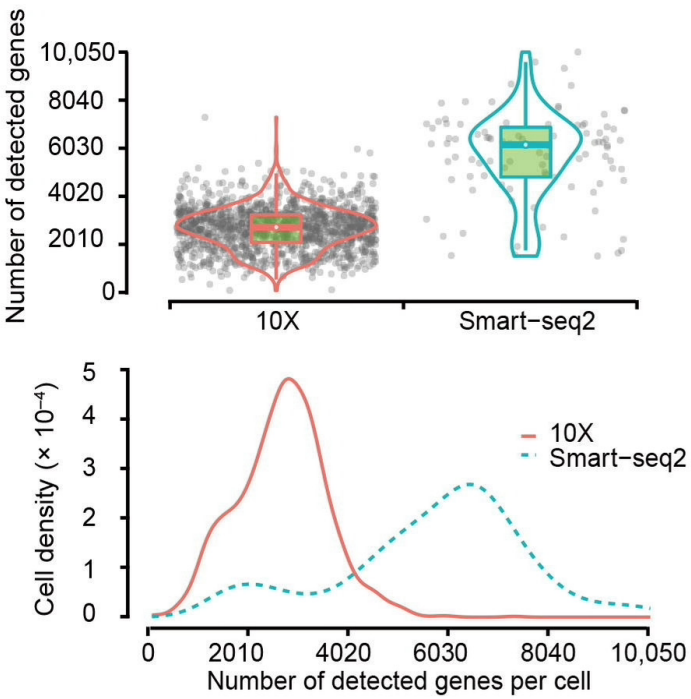
All captured **transcripts** from **single** cell:
identical + **unique** barcode



- reverse transcription
- breaking GEMs
- fragmentation
- primer ligation
- index PCR







Wang X, et al. Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. Genomics Proteomics Bioinformatics; 2021

Droplet (10x genomics)	**SMART-seq**
3’ poly-A bead capture	polydT reverse transcription
Strong bias 3’ end	Whole transcript coverage
Expression analysis	Expression + isoform analysis
Low # transcripts/cell	High # transcripts/cell
Investment (cell sorter)	Only FACS needed
10-100k cells	up to 1k cells
1 sample ≈ 1 library	1 cell = 1 library
Isolation by droplets - doublets	Isolation by FACS – bias to large cells
Low cost/cell	High cost/cell

Quiz Question 2

Experimental design

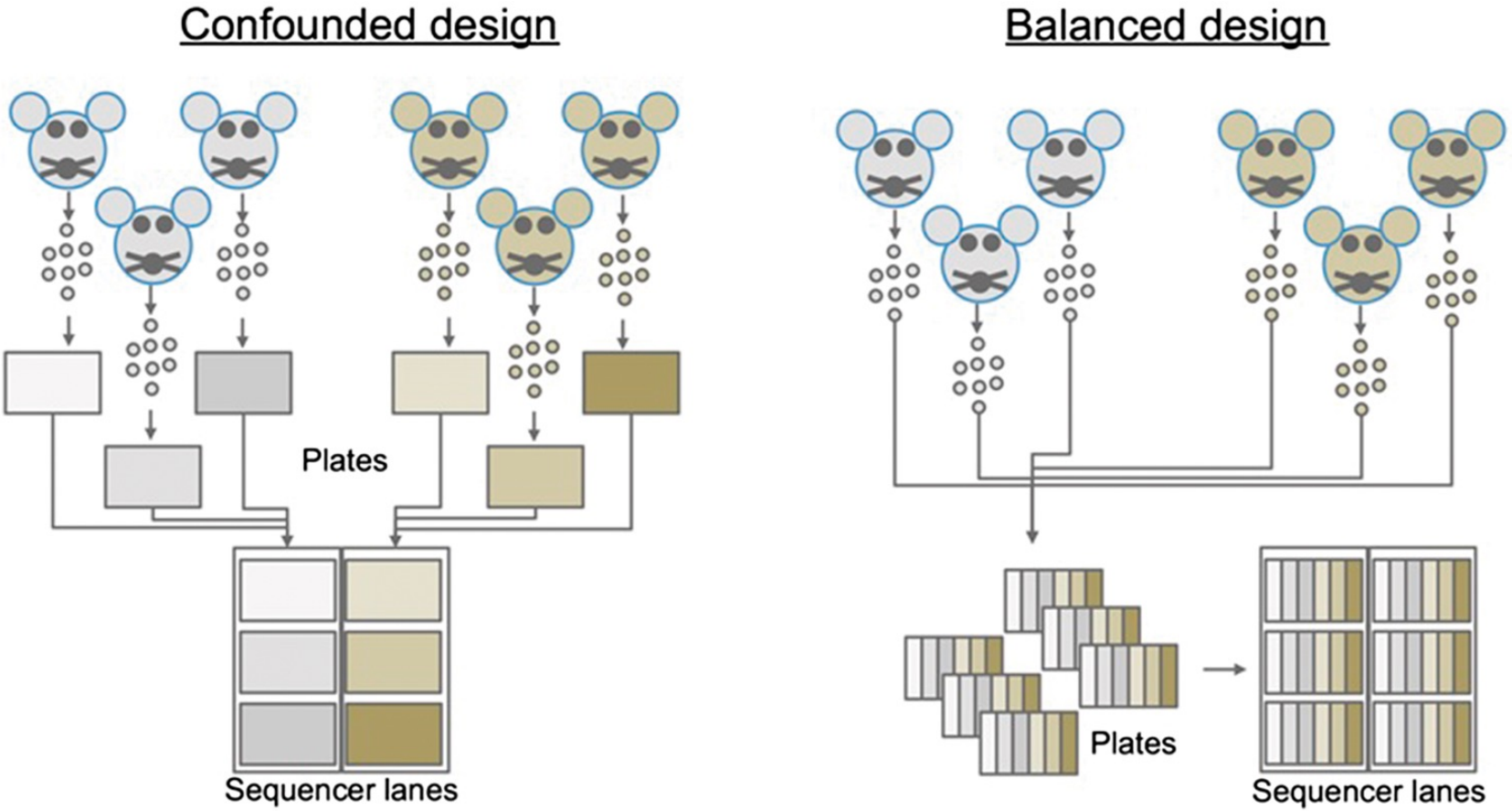
- Replication, randomization and blocking
- Be aware of confounding factors, e.g.:
  - Person performing handling
  - Reagents
  - Sequencing lane/library

- Record any factor for downstream correction

Further reading:

- https://doi.org/10.3389/fcell.2018.00108
- https://doi.org/10.1093/bib/bby007
- https://doi.org/10.1093/bfgp/elx035

Experimental design

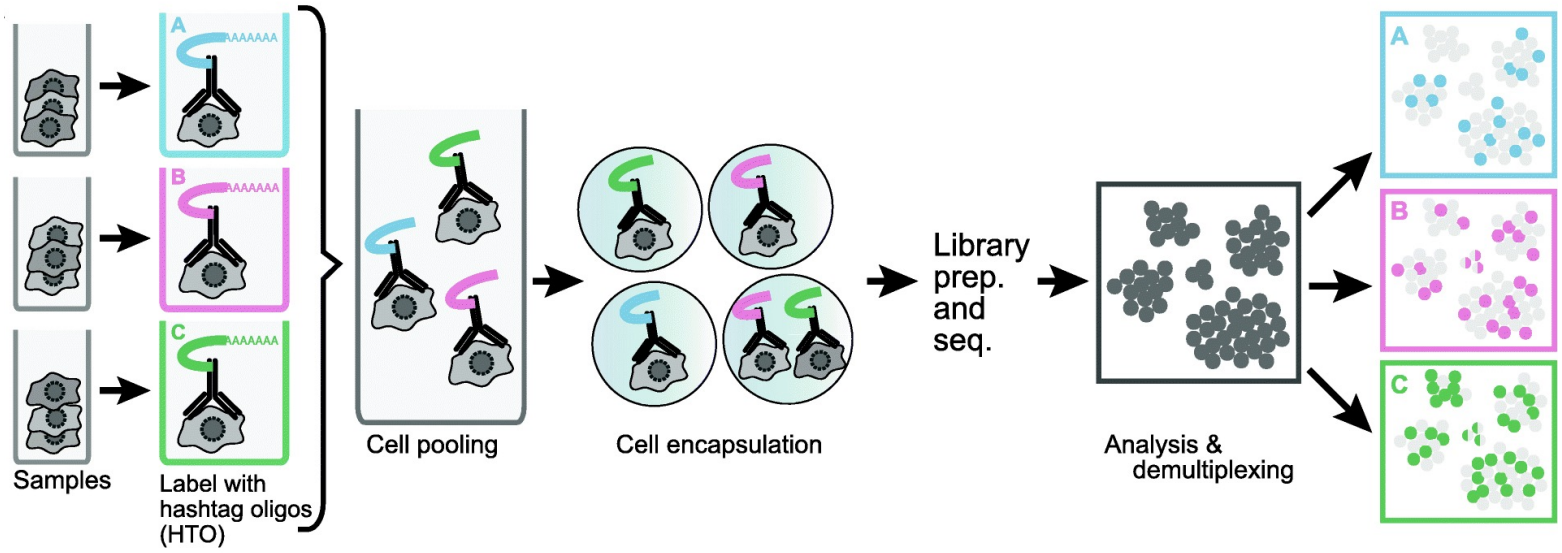


Baran-Gale J, et al. Experimental design for single-cell RNA sequencing. Brief Funct Genomics. 2018;17:233–9.

Question 3

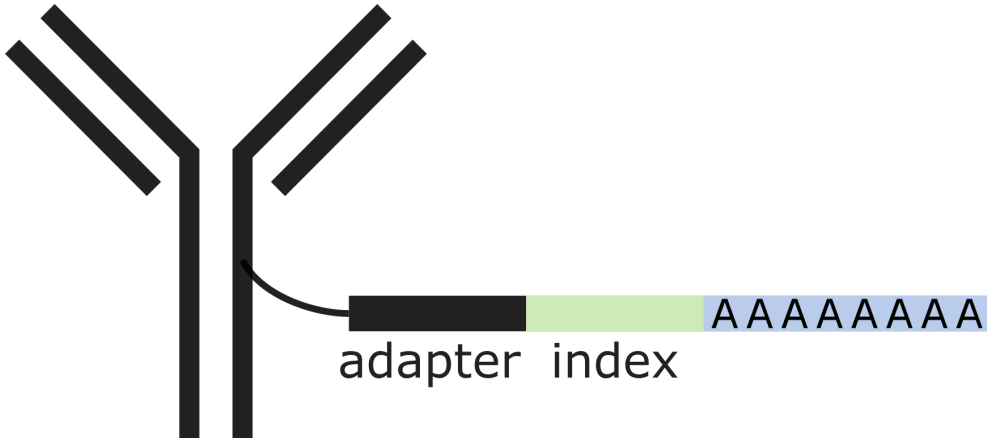
Cell hashing

Solution to 1 sample = 1 library



Stoeckius M et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. Genome Biology; 2018;19:1–12.

Quantify proteins



- Quantification of (cell-surface) proteins
- Together with transcriptome

Stoeckius M et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14:865–8.

Question 4

single-nucleus RNA-seq

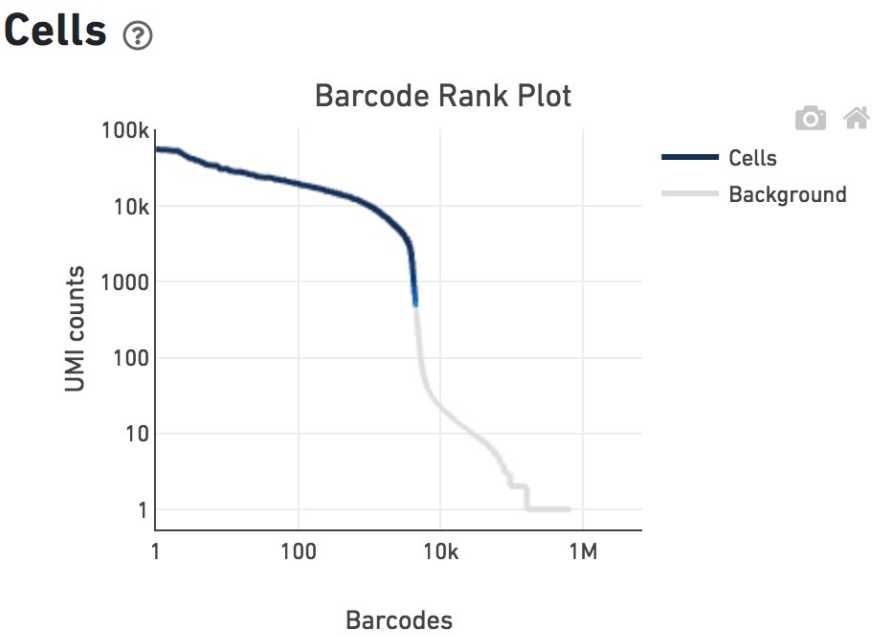
- Alternative to scRNA-seq
- For tissues difficult to dissociate
- No ribosomes -> no translation of transcription factors during processing
- Lower representation of immune cells + surface proteins

Denisenko E, et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol.; 2020;21:1–25.

After sequencing (pre-processing)

1. Demultiplexing
2. Alignment
3. Quantification: # UMI/gene
4. Cell calling

For 10x all with `cellranger count`



Notes on cellranger count

- Human & mouse: download pre-built from 10x website
- Other organisms: custom reference with **`cellranger mkref`**
- Exogenous marker genes (e.g. GFP): add sequence to both fasta and gtf
- Features (e.g.) hashing or surface-proteins: feature barcode reference csv

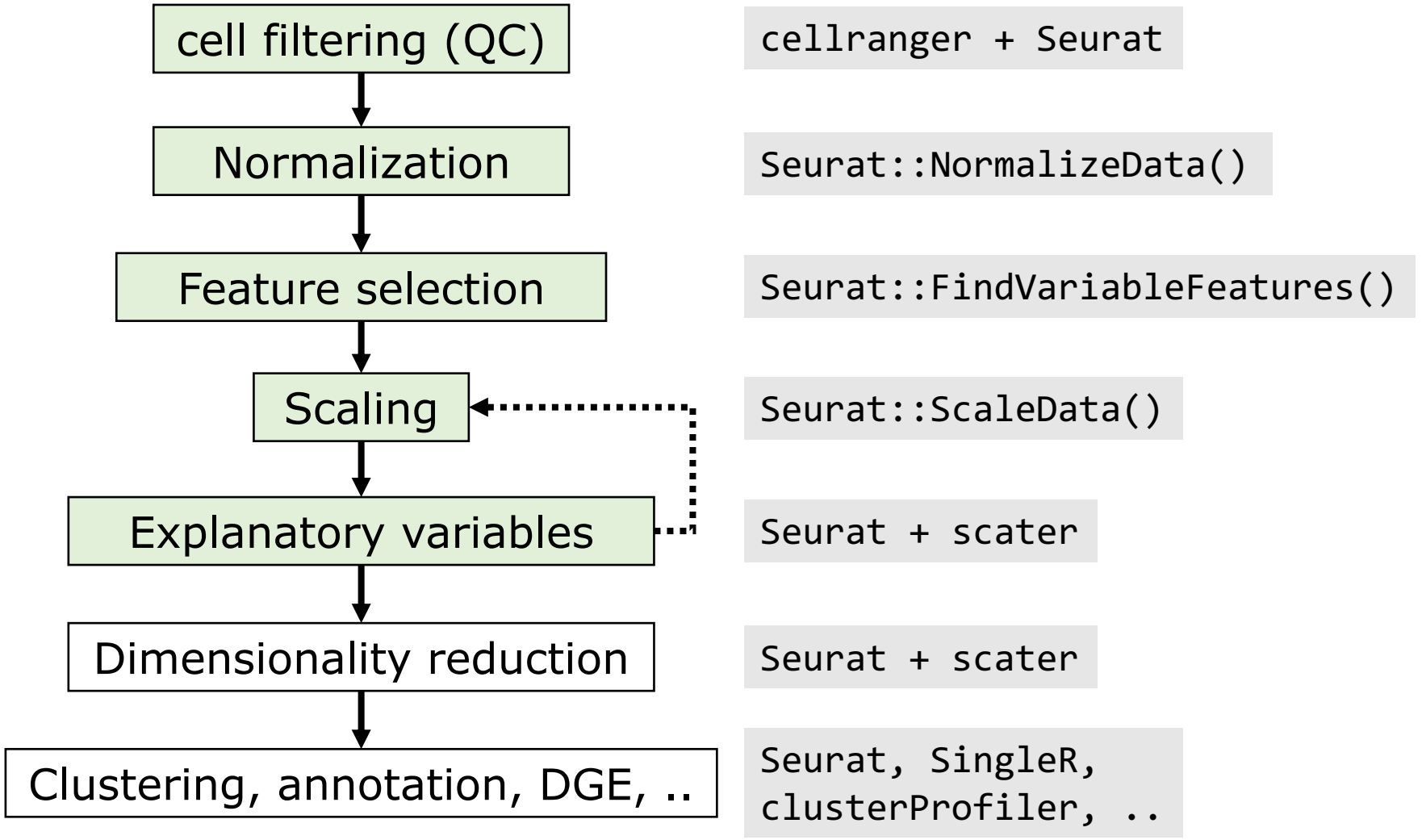
extensive documentation:
https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger

cells ->

genes ->

	ATAC-1	CCG-1	GCGA-1
RPL22	5	13	3
PARK7	0	9	3
ENO1	1	35	0
PLA2G2A	0	0	0
CAMK2N1	0	6	0
CDC42	0	8	1
C1QA	0	0	25
C1QC	0	0	25
C1QB	0	0	29
ID3	0	35	0
RPL11	18	29	16
CLIC4	0	4	0

Downstream analysis



Cell filtering

- Find doublets (e.g. with `DoubletFinder`)
- % UMI in mitochondrial genes:
  - Points to dying/stressed cells
  - Filter by threshold (e.g. 20%)
- Number of detected genes
  - Can point to dropouts/doublets

McGinnis CS et al. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst; 2019;8:329-337

Normalization & scaling

- **Normalization** (per cell): remove technical effects (i.e. library size)
- **Scaling** (per gene): standardize range, mean and variance

Both are mainly for the purpose of dimensionality reduction