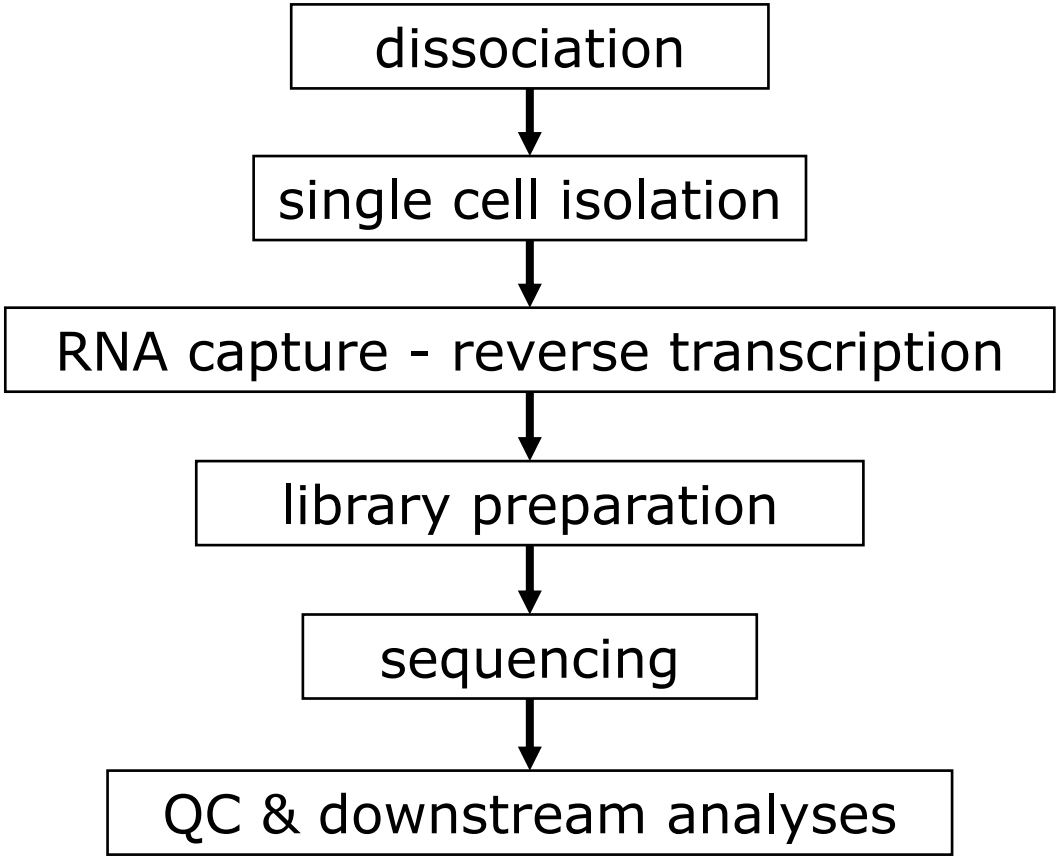
Single cell transcriptomics

Introduction to single cell RNA-seq

Why single cell RNAseq?

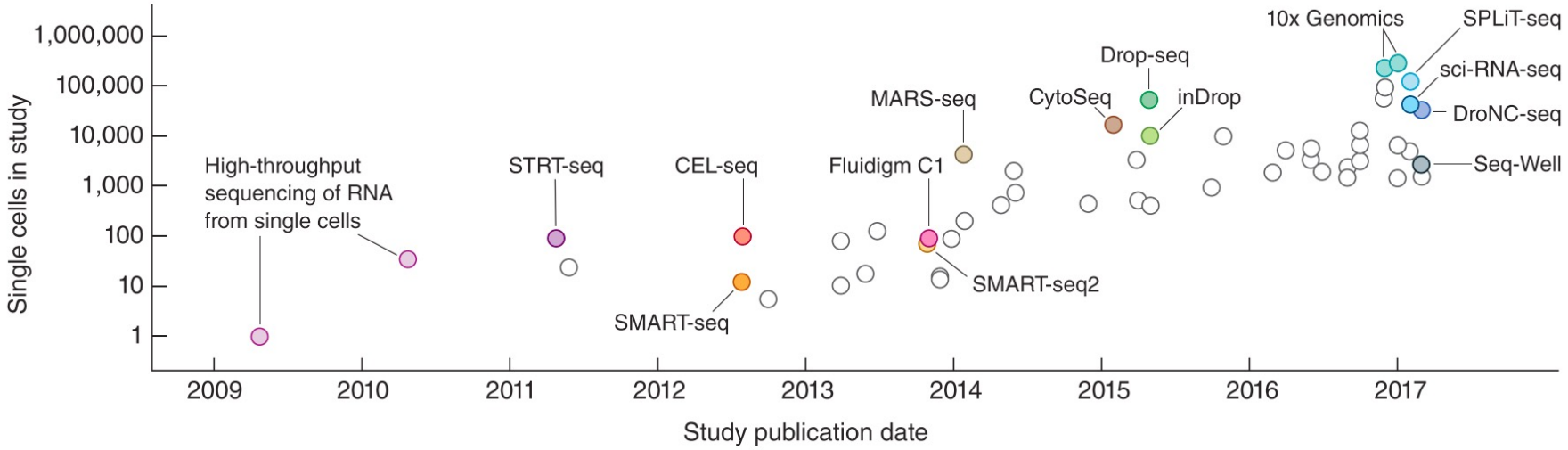
- Discover rare cell populations
- Determine gene regulatory relationships
- Understand drug delivery
- Cell specific effects of treatments
- Understand differences in cell populations
- Screen for CRISPR targets
- …

scRNA-seq workflow

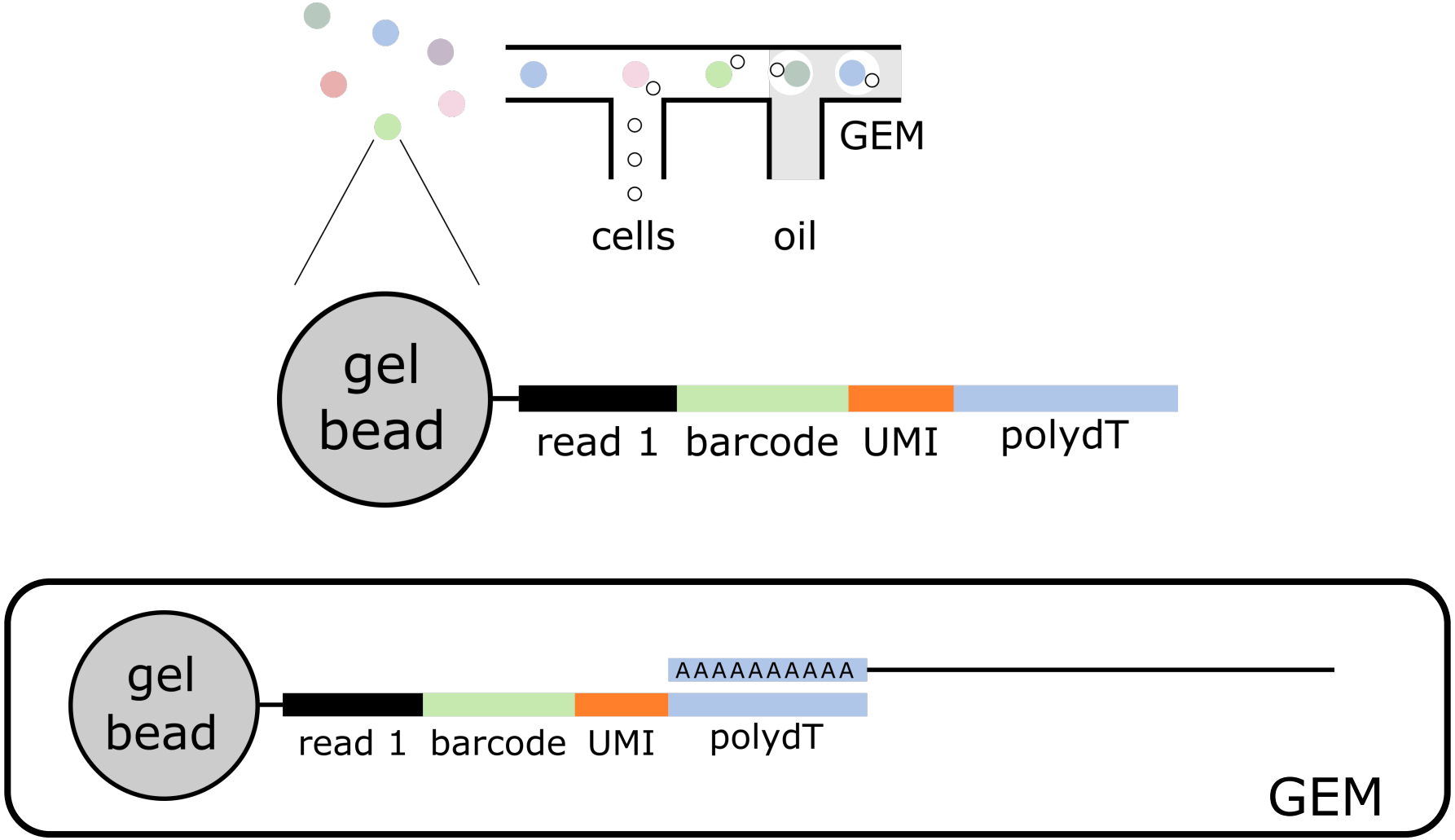


Frequently used technologies

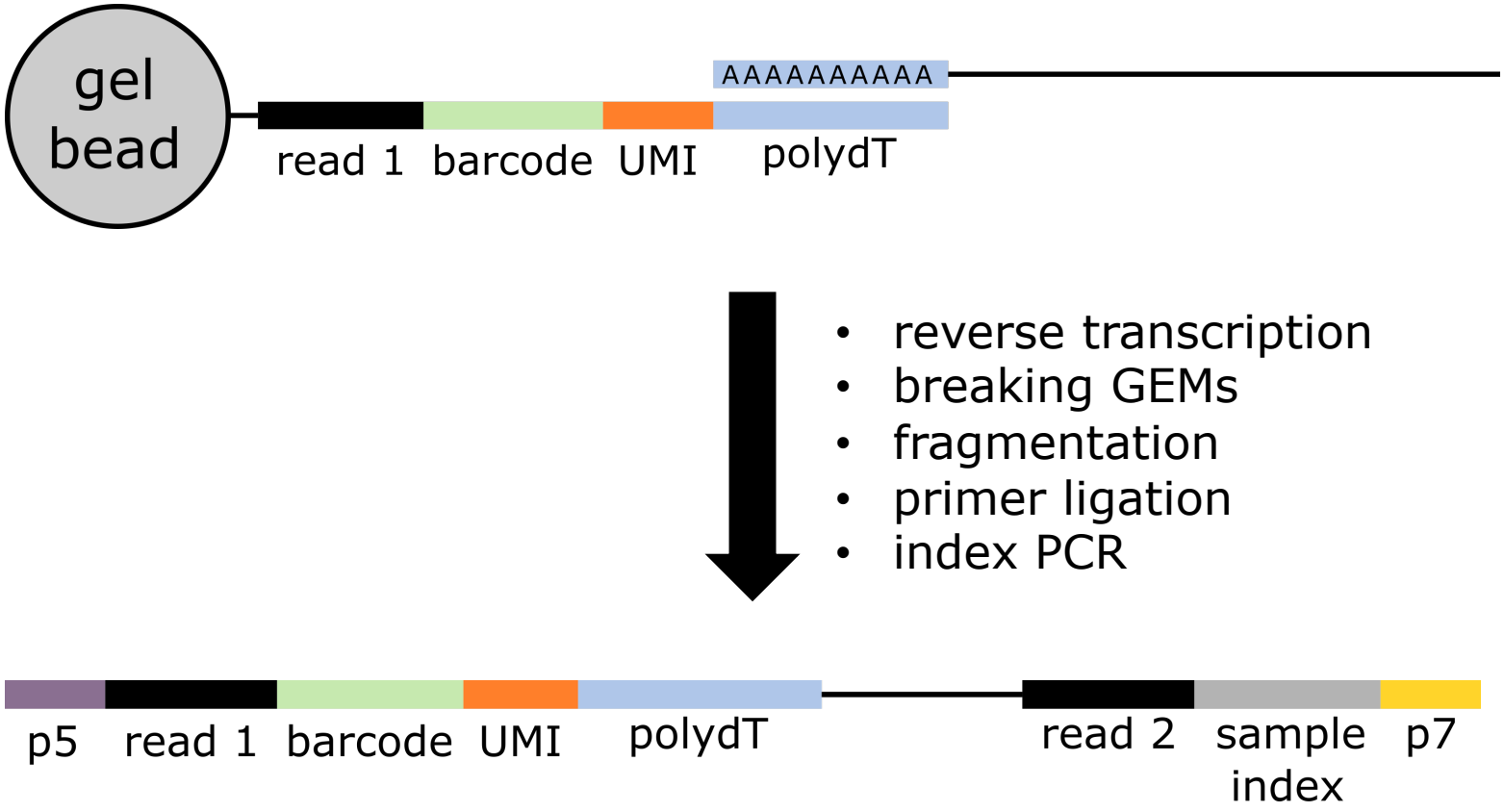
- Droplet-based: 10x genomics
- Plate-based: SMART-seq



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



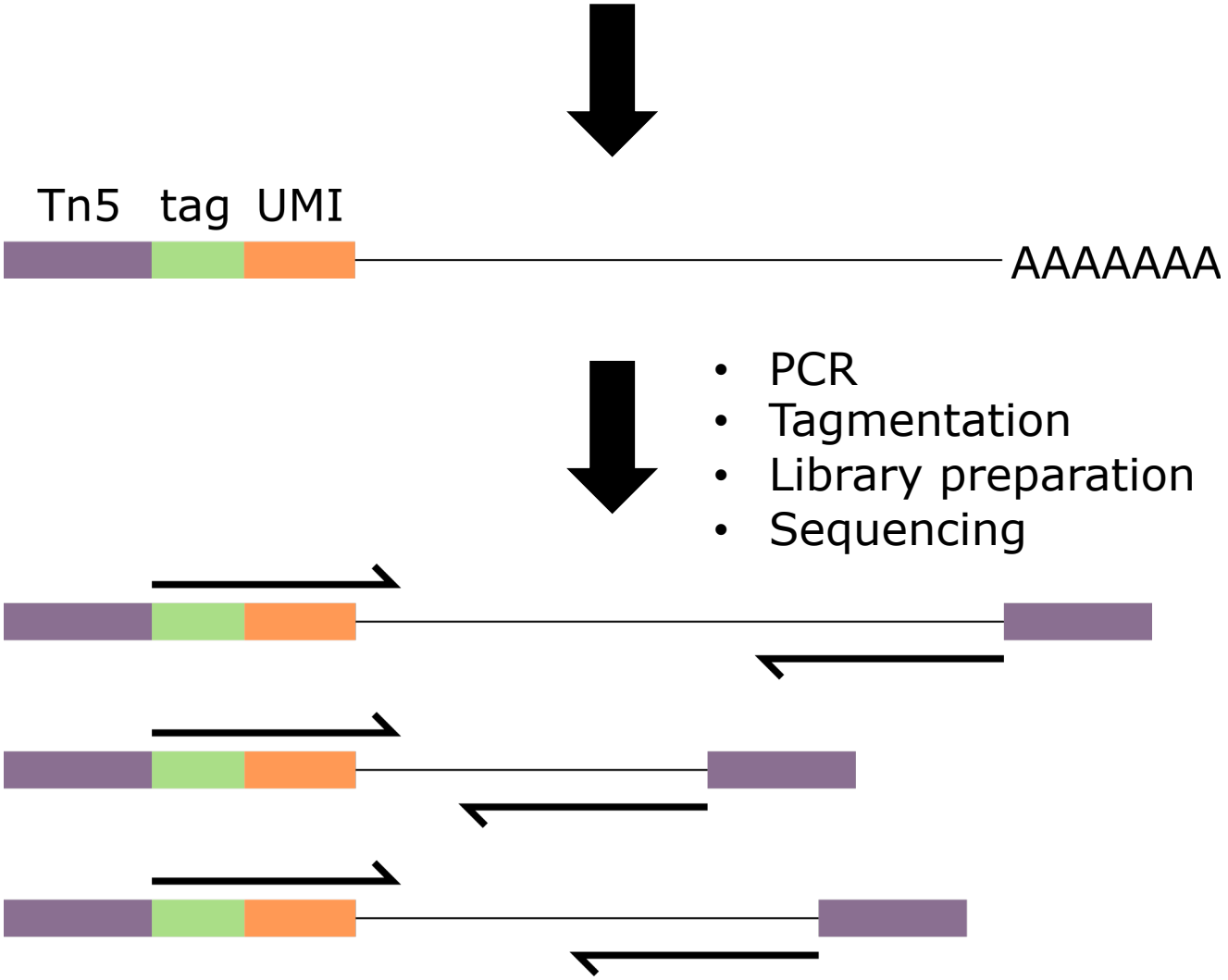
All captured **transcripts** from **single** cell:
identical + **unique** barcode


gel bead
AAAAAAAAAA
read 1
barcode
UMI
polydT
• reverse transcription
• breaking GEMs
• fragmentation
• primer ligation
• index PCR
p5
read 1
barcode
UMI
polydT
read 2
sample index
p7

SMART-seq3

- Cells added to plates with FACS
- Template switching
- Sequencing from 5’-end
- Fragmentation through tagmentation
- Paired-end sequencing enables isoform reconstruction

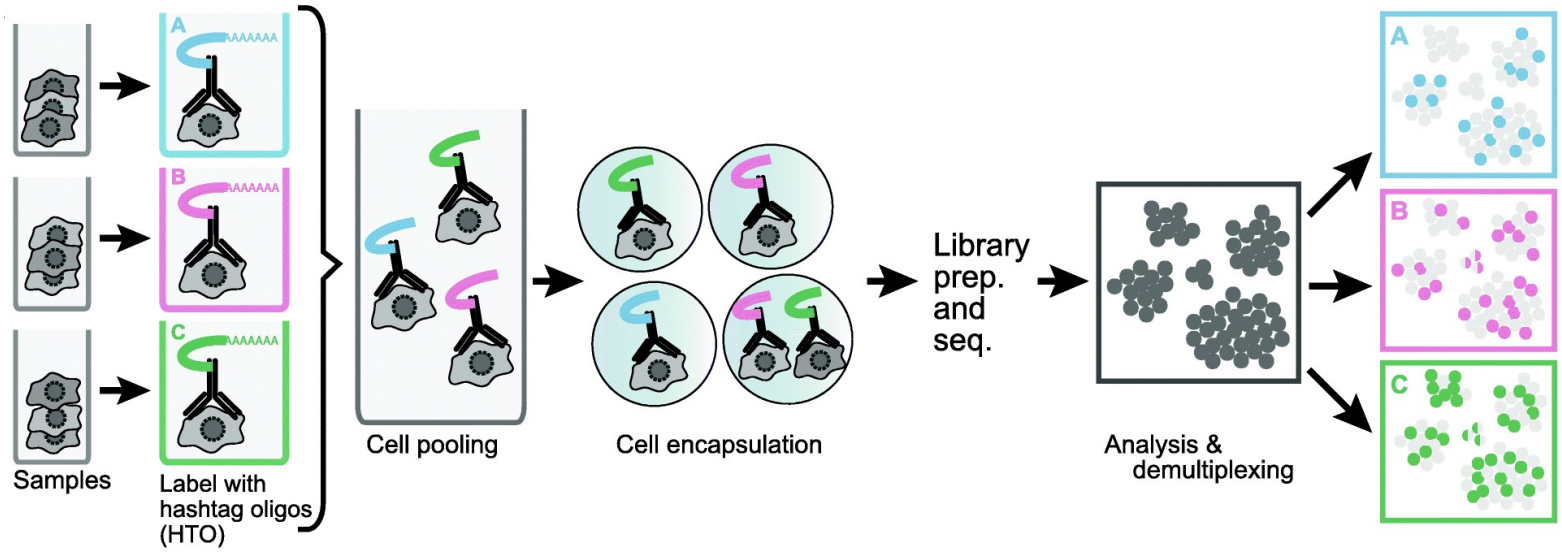
Reverse transcription + template switching



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

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.

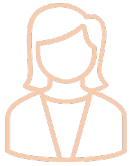
Experimental design

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

- Record possible confounding factor for downstream batch correction

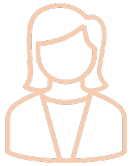
Further reading:

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



treated 1	treated 2
treated 3	treated 4

control 1	control 2
control 3	control 4



control 3	control 2
treated 1	treated 4

control 1	treated 3
treated 2	control 4

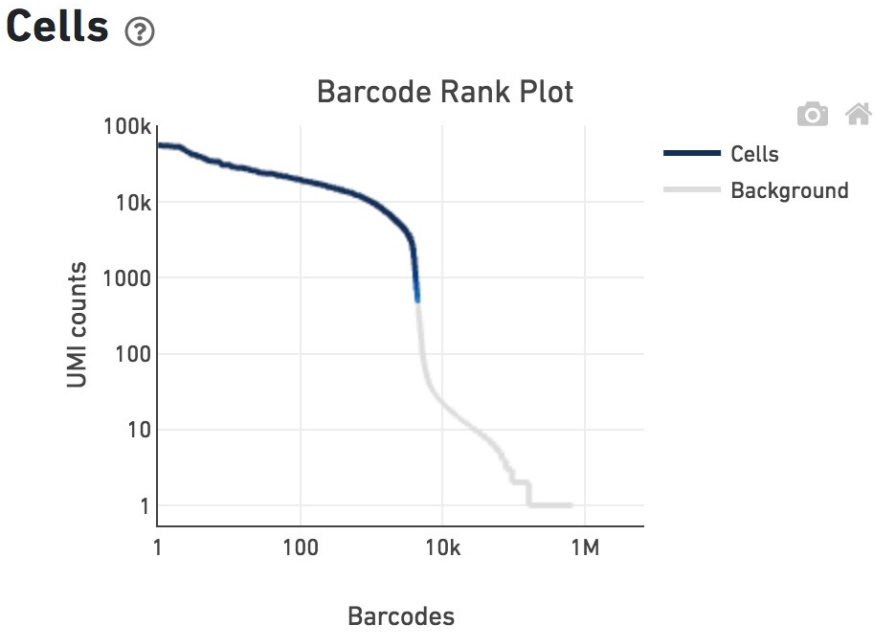
After sequencing (pre-processing)

1. Demultiplexing
2. Alignment
3. Quantification: # UMI/gene
4. Cell calling: i.e. detection of doublets and empty droplets/wells

For 10x (human/mouse/rat): all with `cellranger`

Cell calling

- Doublets: high #UMI/cell
- Dropouts: low #UMI/cell



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

Further cell filtering

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

Normalization & scaling

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