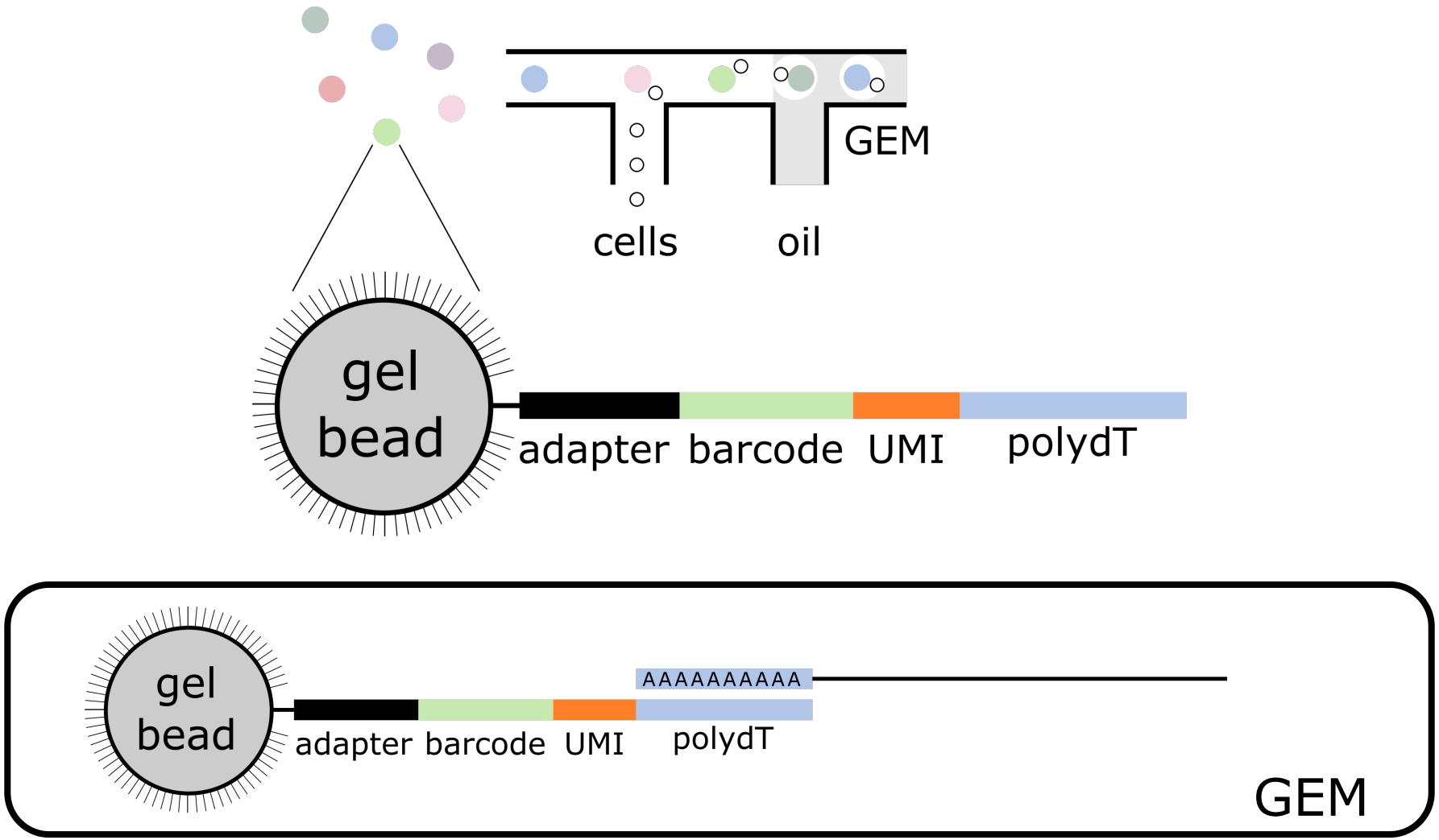
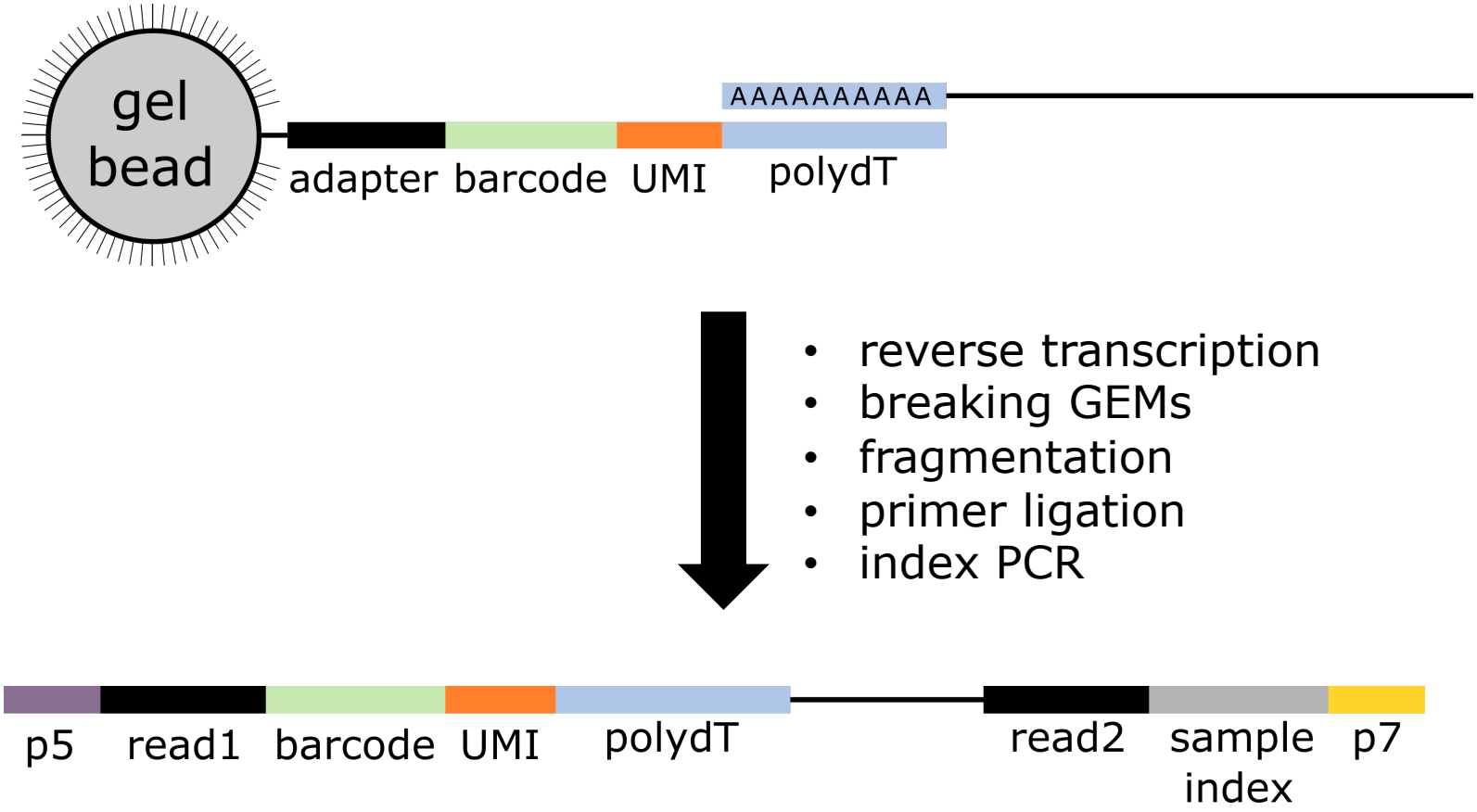
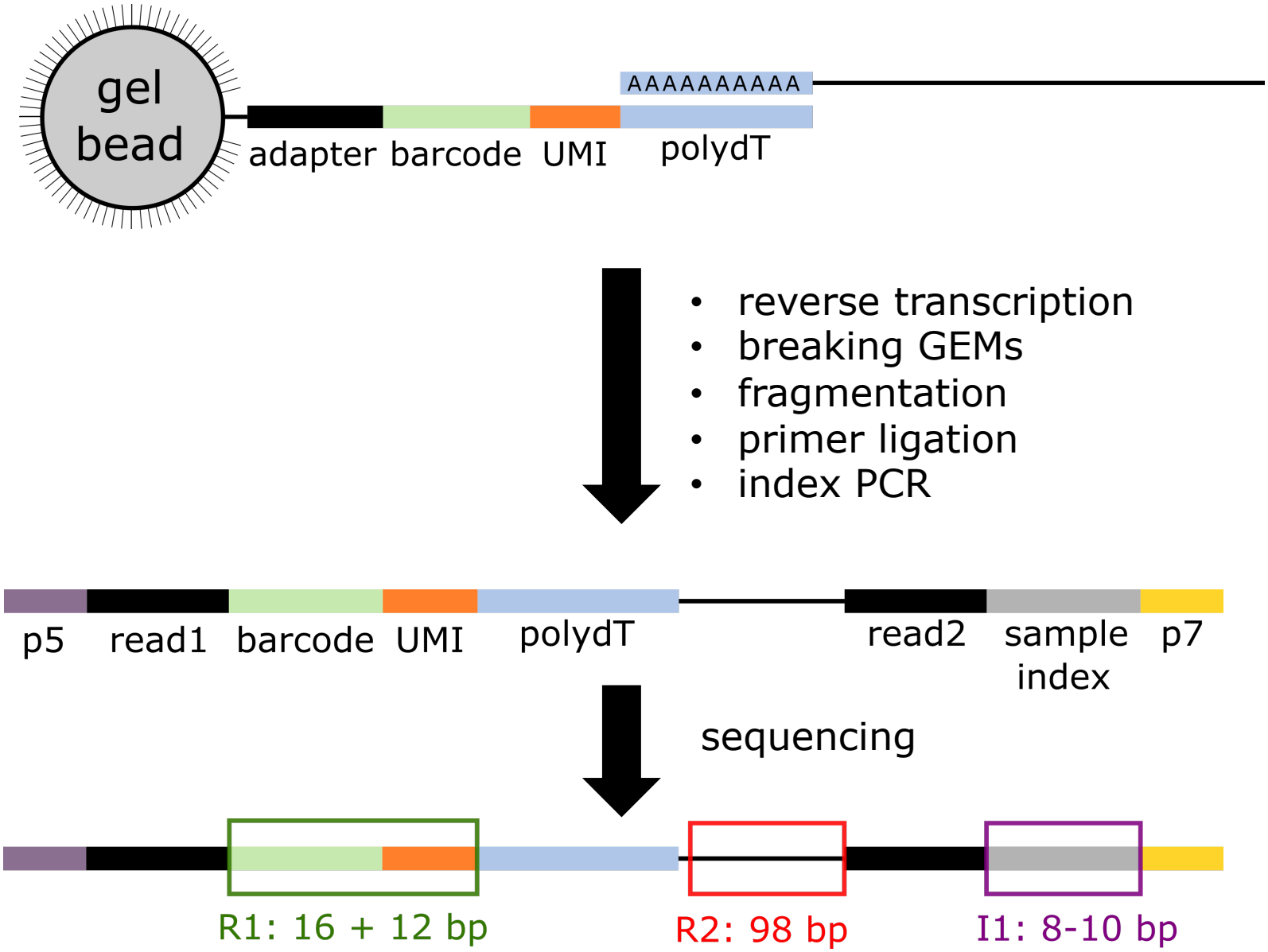
Single cell transcriptomics

10x genomics Chromium



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


gel bead
AAAAAAAAAA
adapter
barcode
UMI
polydT
reverse transcription
breaking GEMs
fragmentation
primer ligation
index PCR
p5
read1
barcode
UMI
polydT
read2
sample index
p7


gel bead
AAAAAAAAAA
adapter
barcode
UMI
polydT
reverse transcription
breaking GEMs
fragmentation
primer ligation
index PCR
p5
read1
barcode
UMI
polydT
read2
sample index
p7
sequencing
R1: 16 + 12 bp
R2: 98 bp
I1: 8-10 bp

Sequencing output

sample# read type

```
ETV6-RUNX1_1_S1_L001_I1_001.fastq.gz
ETV6-RUNX1_1_S1_L001_R1_001.fastq.gz
ETV6-RUNX1_1_S1_L001_R2_001.fastq.gz
```

sample ID lane

- Dual indexing: second index in I2
- Indexes can also be added to fastq titles

After sequencing (pre-processing)

1. Assign reads to cell
2. Alignment
3. Quantification: # UMI/gene
4. Cell calling

For 10x all with `cellranger count`

Alternatives:
STARSolo
Alevin

cellranger references

- 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

Why count UMI (and not read alignments?)

- UMI: Unique Molecular Identifier:
  - Identifies each molecule (i.e. sequence) uniquely
- Molecules from a common PCR template -> carry the same UMI
- By counting UMI: correct for PCR duplicates

Cellranger report

ETV6-RUNX1_1

Alerts

The analysis detected ⚠ 1 warning.

	Alert	Value	Detail
⚠	Fraction of RNA read bases with Q-score >= 30 is low	59.4%	Fraction of RNA read bases with Q-score >= 30 should be above 65%. A lower fraction might indicate poor sequencing quality. This is Read 1 for the Single Cell 3' v1 chemistry and Single Cell 5' paired end, Read 2 for the Single Cell 3' v2/v3 chemistry and Single Cell 5' R2-only)

Summary | Analysis

3,091
Estimated Number of Cells

68,259
Mean Reads per Cell

1,717
Median Genes per Cell

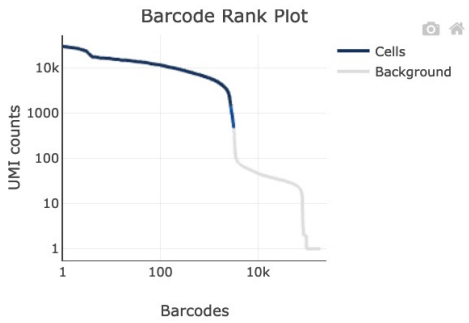
Sequencing

Number of Reads	210,987,037
Number of Short Reads Skipped	0
Valid Barcodes	98.2%
Valid UMIs	100.0%
Sequencing Saturation	84.4%
Q30 Bases in Barcode	96.4%
Q30 Bases in RNA Read	59.4%
Q30 Bases in UMI	96.5%

Mapping

Reads Mapped to Genome	95.8%
Reads Mapped Confidently to Genome	92.9%
Reads Mapped Confidently to Intergenic Regions	5.2%
Reads Mapped Confidently to Intronic Regions	25.5%
Reads Mapped Confidently to Exonic Regions	62.2%
Reads Mapped Confidently to Transcriptome	58.2%
Reads Mapped Antisense to Gene	1.2%

Cells

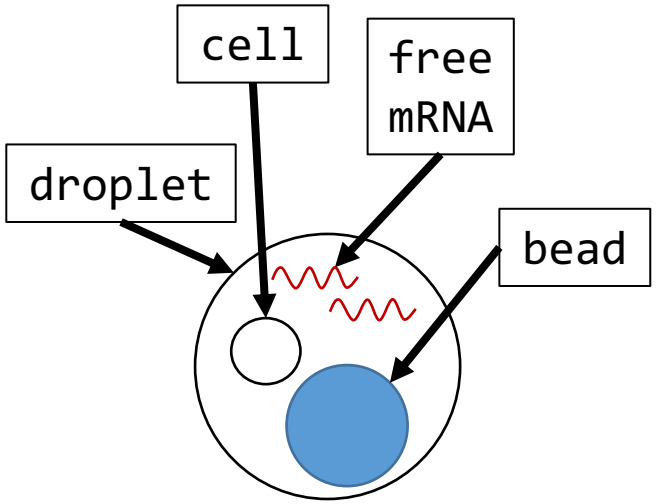


Estimated Number of Cells	3,091
Fraction Reads in Cells	91.1%
Mean Reads per Cell	68,259
Median Genes per Cell	1,717
Total Genes Detected	18,334
Median UMI Counts per Cell	4,825

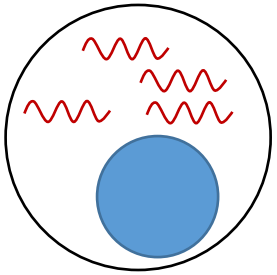
Sample

Sample ID	ETV6-RUNX1_1
Sample Description	
Chemistry	Single Cell 3' v2
Include introns	False
Reference Path	…nger/refdata-cellranger-GRCh38-3.0.0
Transcriptome	GRCh38-3.0.0
Pipeline Version	cellranger-6.0.1

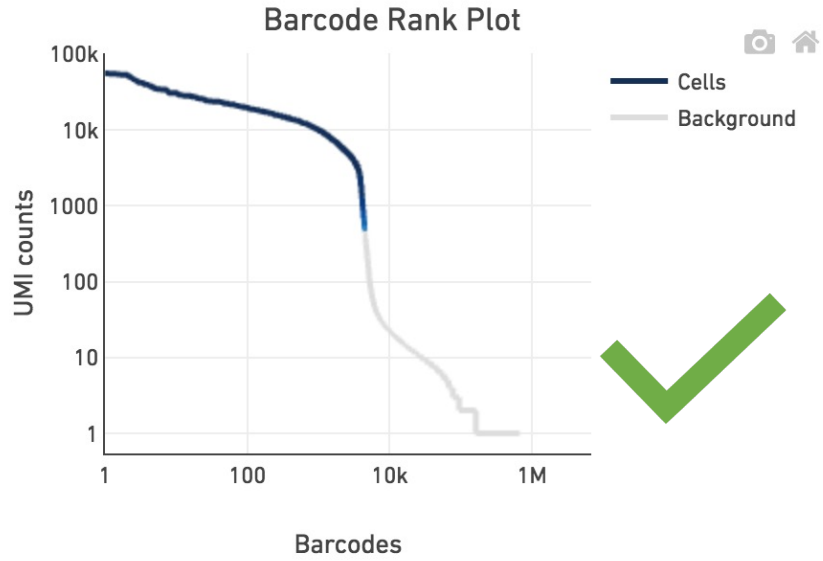
Cell calling

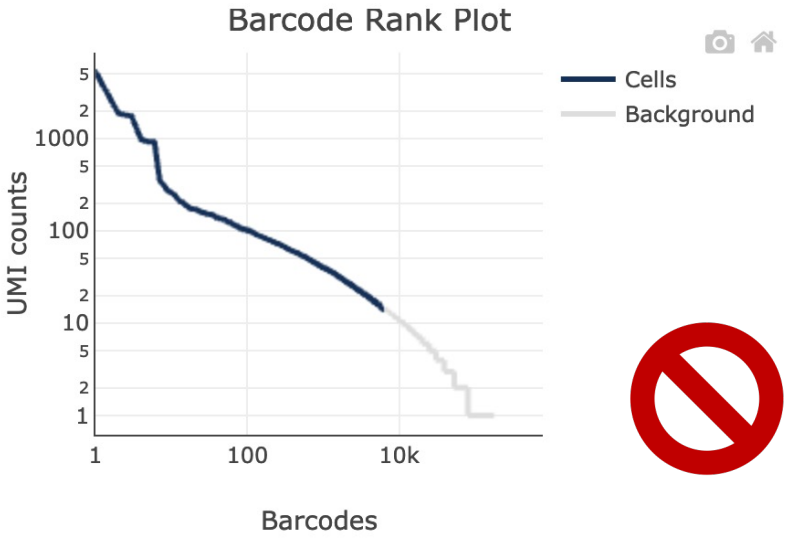


Cell captured



Only background captured





Background ‘cells’: low #UMI/cell

Other parameters

- Captured cells: 1,000-8,000
- Reads/cell: 30,000-100,000 (or more)
- Sequencing saturation
- Reads mapped to genome/transcriptome

$$saturation = 1 - \frac{\#\ unique\ reads}{\#\ reads}$$