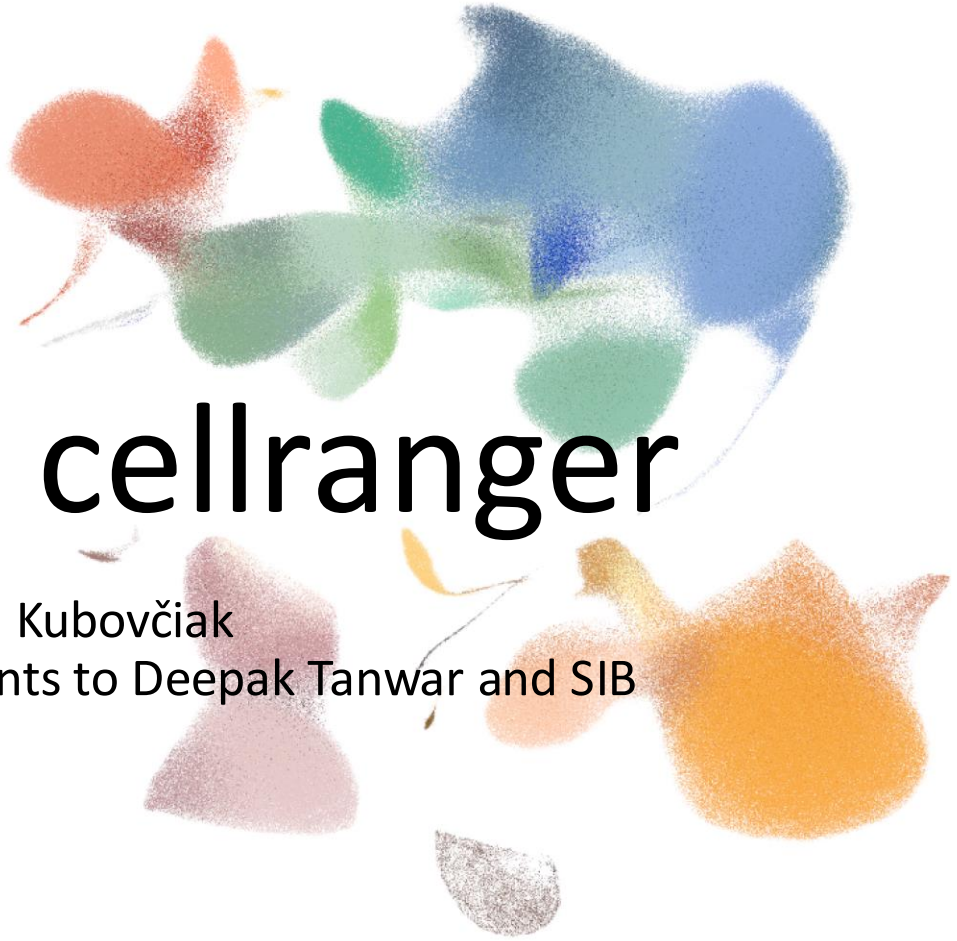


10x and cellranger

Jan Kubovčiak

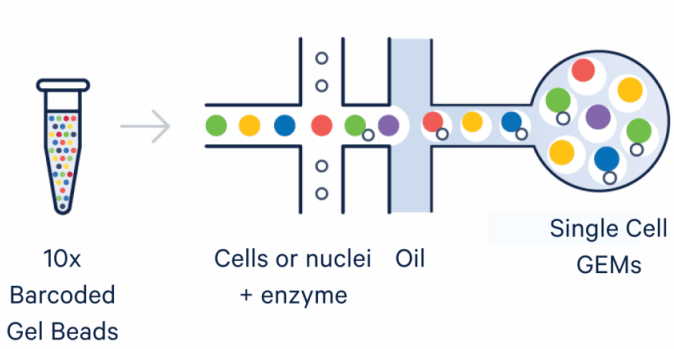
with acknowledgements to Deepak Tanwar and SIB

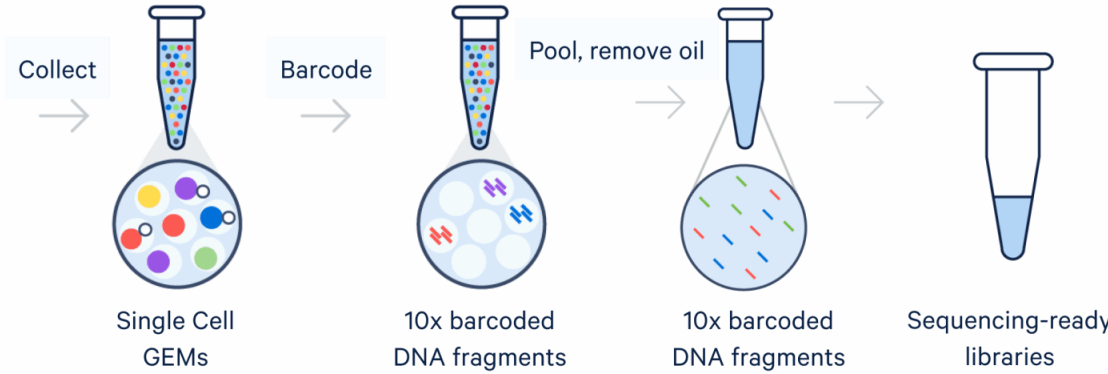
elixir CZECH REPUBLIC

IMG

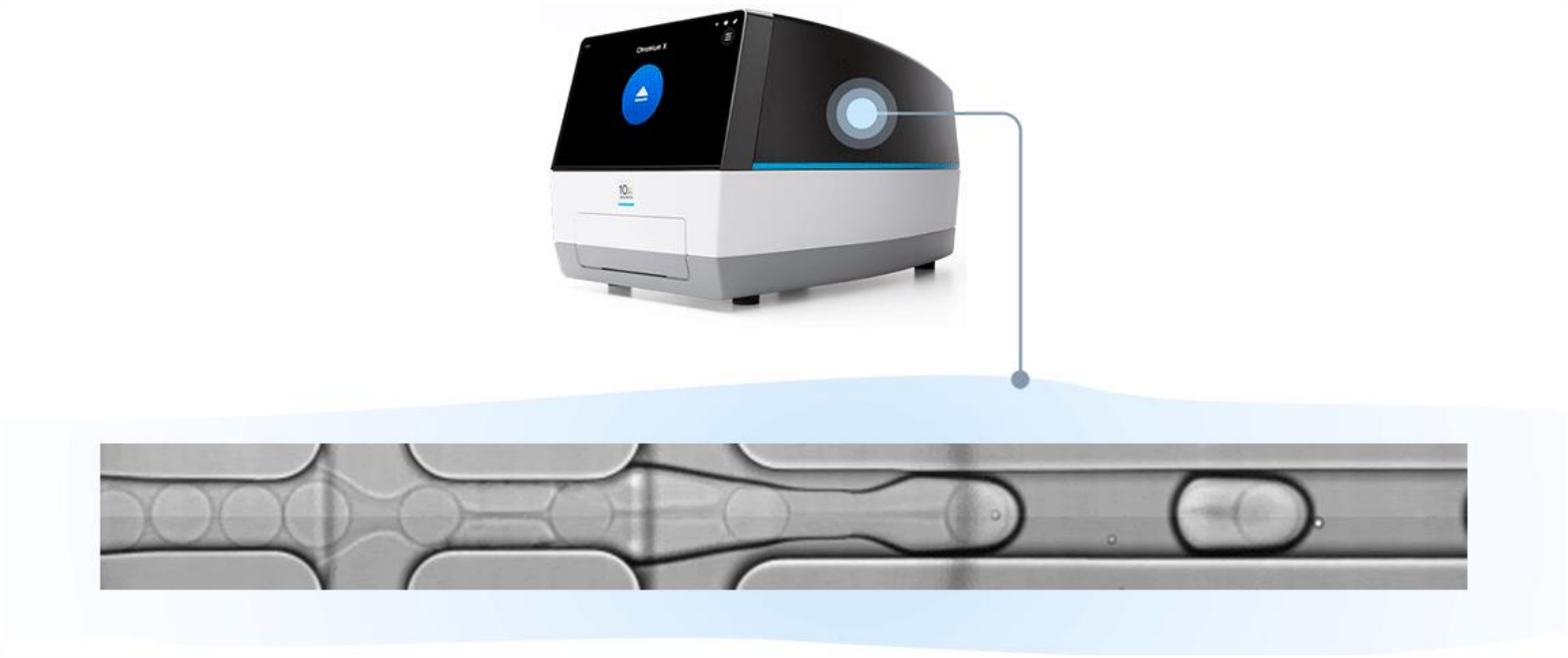
GEMs

- way to label cells separately
- Gel Beads-in-emulsion

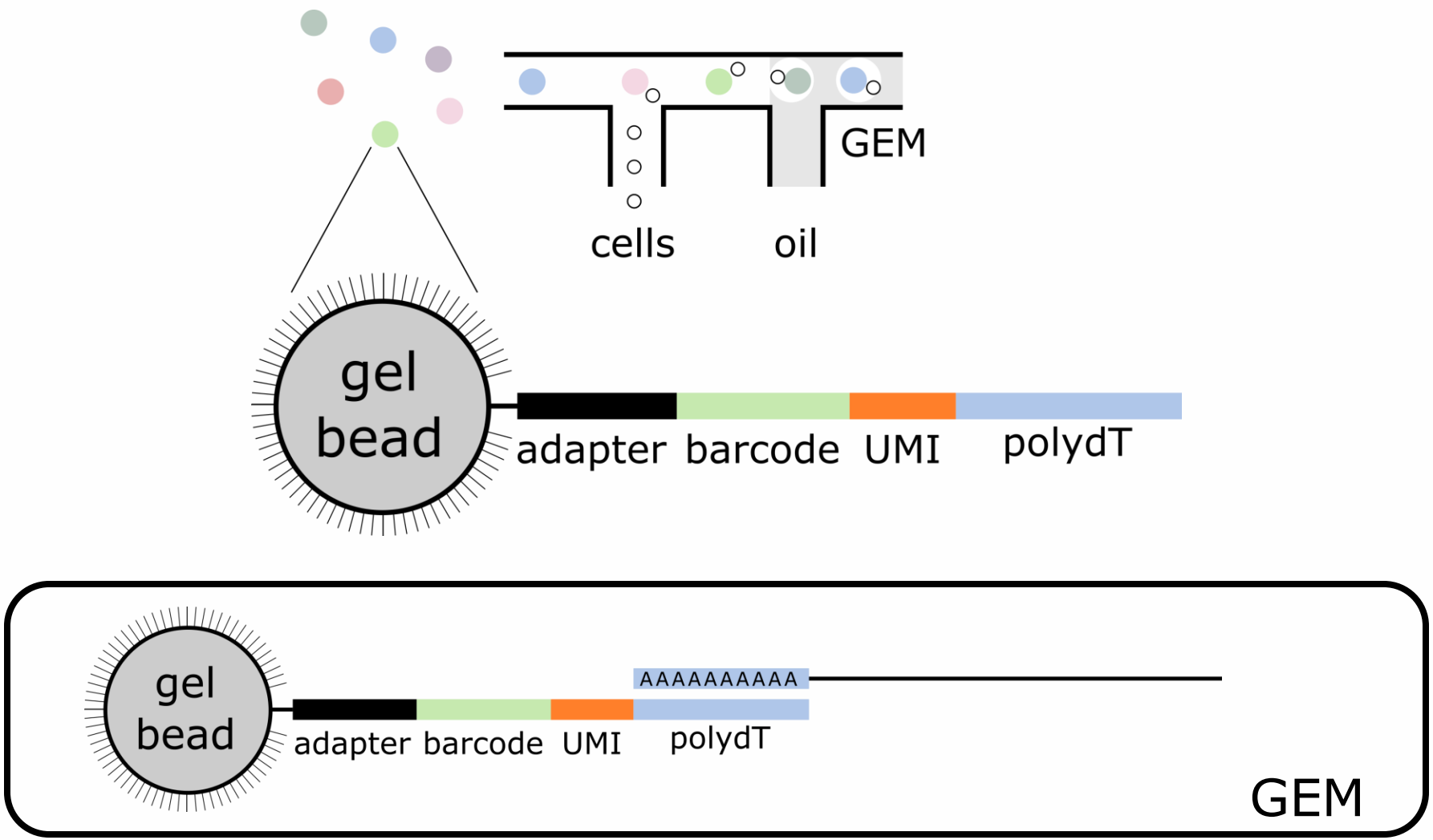




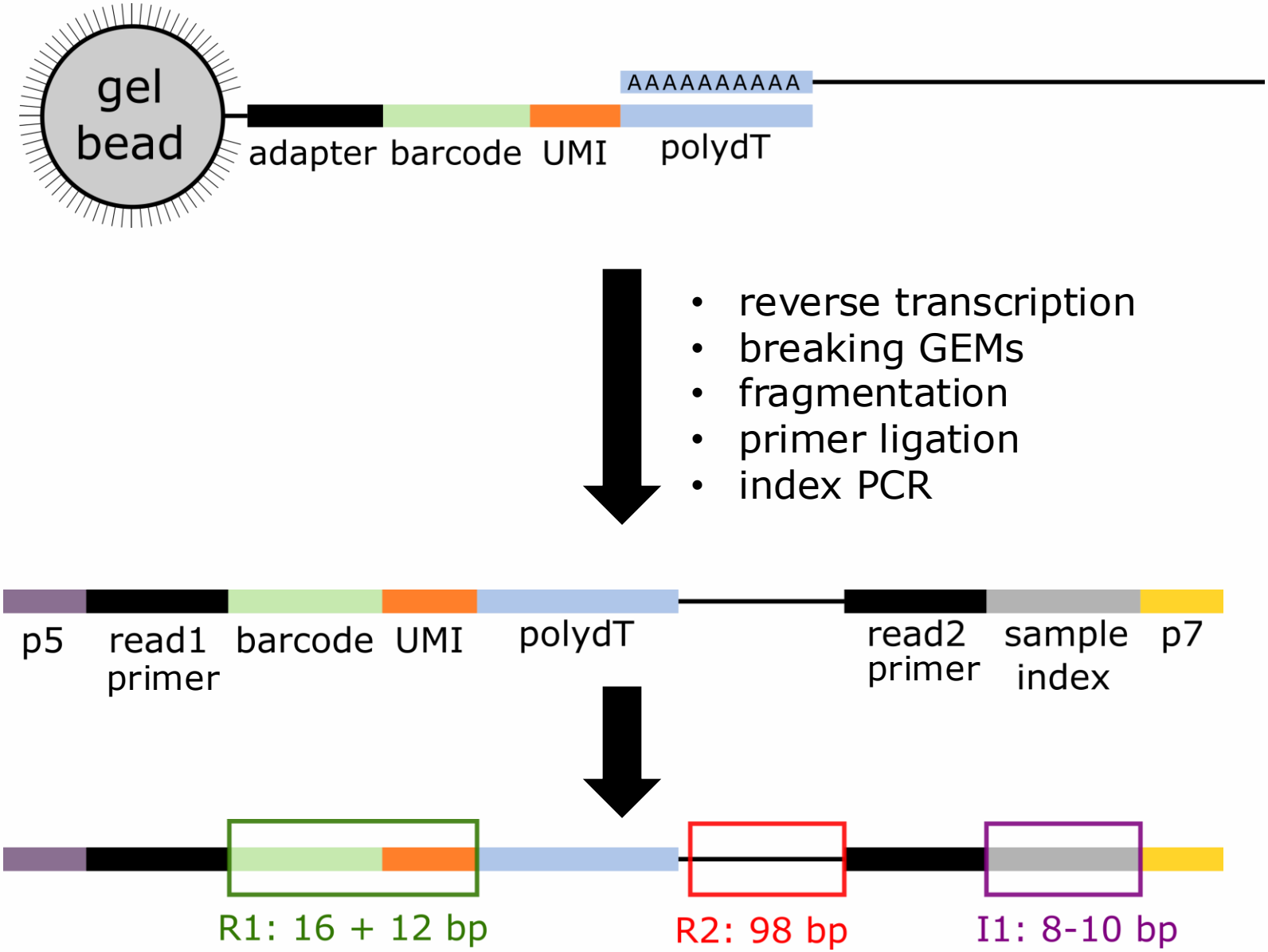
Chromium instrument



10,000+ GEMs in 4 minutes.



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



Illumina sequencing

Sequencing output

sample# read type

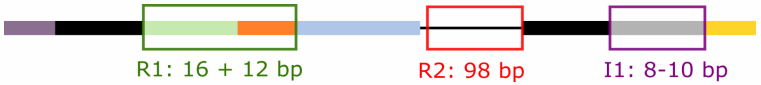
```
ETV6-RUNX1_1_S1_L001_R1_001.fastq.gz
ETV6-RUNX1_1_S1_L001_R2_001.fastq.gz
ETV6-RUNX1_1_S1_L001_I1_001.fastq.gz
```

sample ID lane

For 10x all with `cellranger`

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

IMG



ETV6-RUNX1_1_S1_L001_R1_001.fastq.gz

@SRR9264343.187191690 K00271:135:HK3J7BBXX:3:2110:12591:47964 length=26
GGCAATTCATTGGGCCCCCGAAGCTC
+
AAF-FJFJJJJJJJJJJJJJJJJJ
@SRR9264343.5519124 K00271:135:HK3J7BBXX:3:1112:2950:8295 length=26
GGCAATTCATTGGGTGAGGTTGTAGG
+
AAFFFJJJJJFFJFJJFJJJJJJJJJ
@SRR9264343.149664941 K00271:135:HK3J7BBXX:3:2217:29579:19267 length=26
TCGGATGTCTGAAAGAGTGCTAGAAT
...

ETV6-RUNX1_1_S1_L001_R2_001.fastq.gz

@SRR9264343.187191690 K00271:135:HK3J7BBXX:3:2110:12591:47964 length=98
GTAGTAAAAGACTGGTTAATGATAACAATGCATCGTAAAACCTTCAGAAGGAAAGGAGAATGTTTTGTGGACCACTTTCGTTTTCTTT
TTTGCGTGTG
+
-777<<A-<FF<<A7-<FAFAFFJJJJJFFFFFFFJFFJJJFFJ7JFFJJJ<FJJAFJJFJJ<A-<AA<FFF)<A--7-7--AFFJ--AFAFA7<AA-
@SRR9264343.5519124 K00271:135:HK3J7BBXX:3:1112:2950:8295 length=98
AGACTTTCTACCTGGTCATATACTCTGCAGCTGTTAGAATGTGCAAGCACTTGGGGACAGCATGAGCTTGCTGTTGTACACAGGGTA
TTTCTAGAAGC
+
--77-77--7-77<-7AFAJJJJJ7A7AFF7-A-7FFJJFA-AF77FF<-A<FFJJF-)FF7F<7FA7FAAAFJ-A--A<AAFJJ<F----7--<F77
@SRR9264343.149664941 K00271:135:HK3J7BBXX:3:2217:29579:19267 length=98
GGTCCGCAAGCAGGTGGTGCACATCCCGTCCTTCATTGTCCGCCTGGCTTCCCAGAAGCGCATCGACTTCTCTCTGCGCTCTCCCTA
CGGGGGTGGCC
...

ETV6-RUNX1_1_S1_L001_I1_001.fastq.gz

@SRR9264343.187191690 K00271:135:HK3J7BBXX:3:2110:12591:47964 length=8
CCTAGACC
+
AAAFFJJJ
@SRR9264343.5519124 K00271:135:HK3J7BBXX:3:1112:2950:8295 length=8
CCTAGACC
+
AAAFFFJF
@SRR9264343.149664941 K00271:135:HK3J7BBXX:3:2217:29579:19267 length=8
CCTAGACC
...

After sequencing (pre-processing)

1. Read mapping
2. Quantification: # UMI/gene
3. Cell calling

For 10x all with cellranger

Alternatives:
STARSolo
Alevin

`cellranger references`

- Human & mouse: download pre-built from 10x website
- Other organisms: custom reference with **`cellranger mkref`**
  - Use **ensembl.org**
- Exogenous marker genes (e.g. GFP): add sequence to both fasta and gtf

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

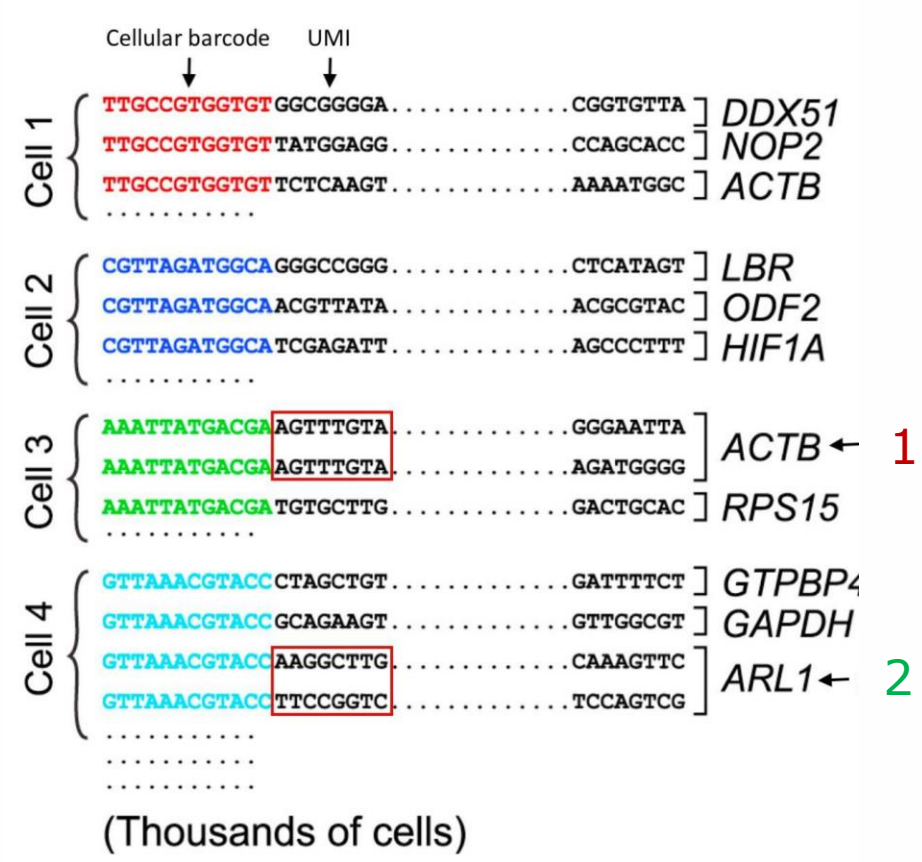
IMG

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 replicates

Why UMIs

- Reads with different UMIs were derived from different molecules and are biological duplicates
- Reads with the same UMI originated from the same molecule and are technical duplicates – the UMIs should be collapsed to be counted as 1



Cellranger count 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

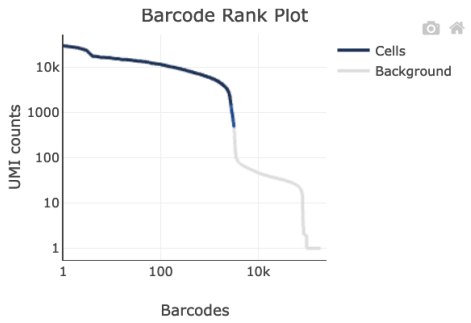
Sequencing

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

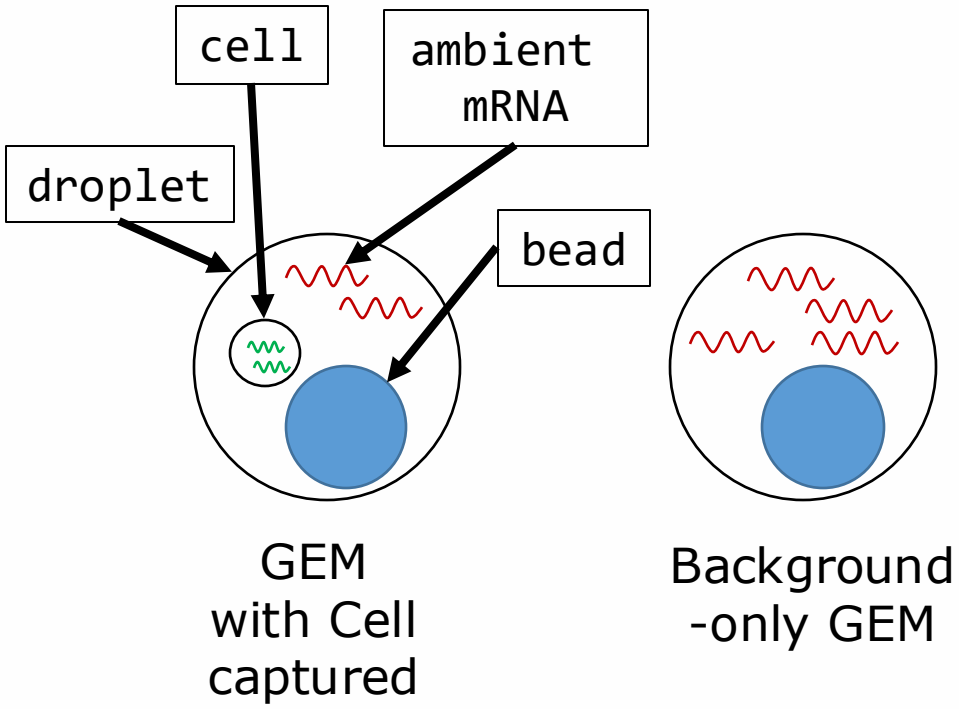


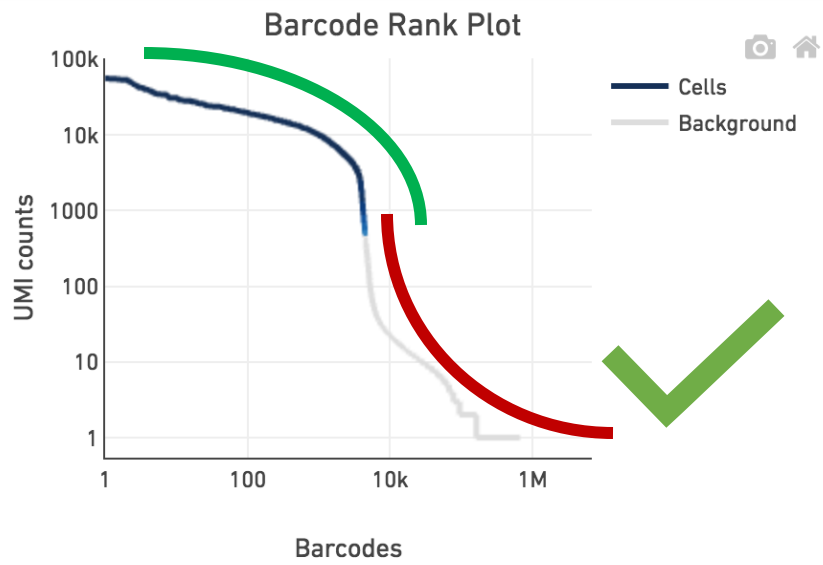
Estimated Number of Cells	3,091
Fraction Reads in Cells	91.1%
Mean Reads per Cell	68,259
Median Genes per Cell	1,717
Total Genes Detected	18,334
Median UMI Counts per Cell	4,825

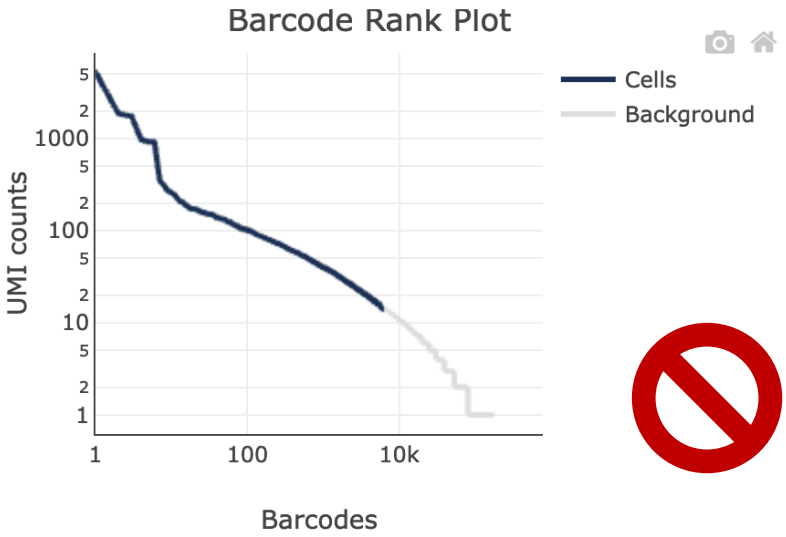
Sample

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







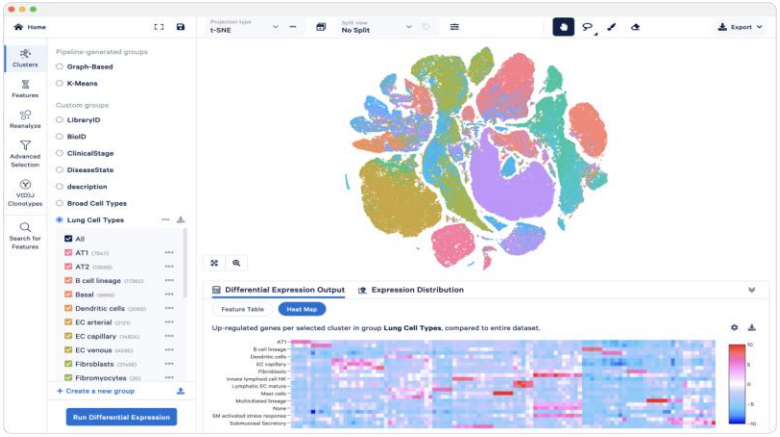
Background GEMs: low #UMI/barcode

Other parameters

scanpy

SEURAT

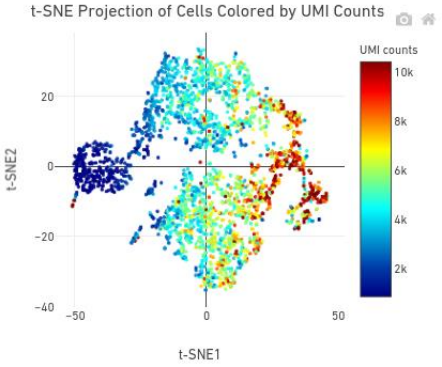
Loupe browser

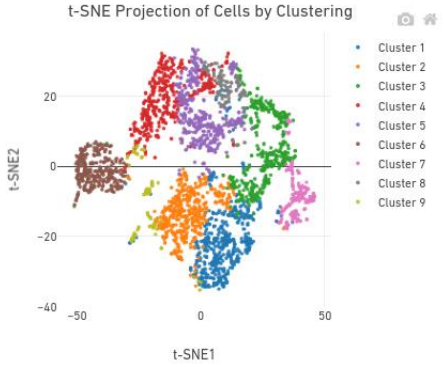


Summary | Analysis

t-SNE Projection

Clustering Type: Graph-based



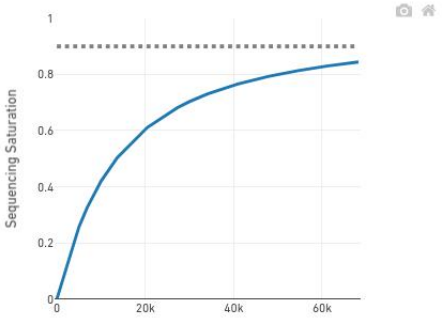


Top Features by Cluster (Log2 fold-change, p-value)

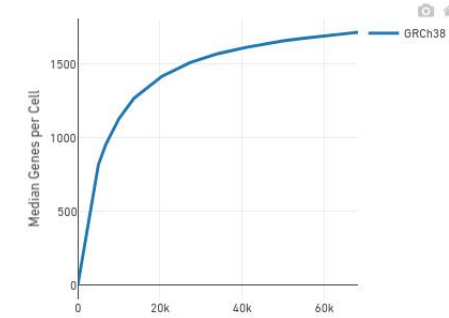
Feature		Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6		C
ID	Name	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2FC	p-value	L2F
ENSG00000277734	TRAC													
ENSG00000008517	IL32													
ENSG00000124785	NRN1	1.02	2e-2											
ENSG00000104870	FCGRT	0.92	5e-2											
ENSG00000185650	ZFP36L1	0.84	1e-1											
ENSG00000136929	HEMGN													
ENSG00000214113	LYRM4													
ENSG00000101197	BIRC7													
ENSG00000172889	EGFL7													
ENSG00000096384	HSP90AB1													

Previous | Page 1 of 37 | 10 rows | Next

Sequencing Saturation



Median Genes per Cell



IMG

10x single cell portfolio

Flex

Fix, batch, and run on your schedule.

Protein coding gene coverage

Performs well with low-quality and FFPE samples

Multiomic readouts from the same cell

Gene expression
Protein
CRISPR

Throughput options

Up to 8M cells (1–3,072 samples) per run

Products

Flex >

Universal

Species agnostic. Maximum versatility.

Whole transcriptome coverage

Delivers broadest set of information, including isoforms, SNPs, etc.

Multiomic readouts from the same cell

3' or 5' gene expression
TCR/BCR
Protein
CRISPR

Throughput options

Up to 160,000 cells (1–8 samples) per run

Products

Universal 3' >
Universal 5' >

Epi Chromatin

Unmask epigenomic profiles.

ATAC-seq chemistry

Explore open chromatin regions
Link directly to 3' gene expression (Multiome kit)

Multiomic readouts from the same cell

Chromatin accessibility
3' gene expression

Throughput options

Up to 80,000 nuclei (1–8 samples) per run

Products

Epi ATAC >
Epi Multiome >

IMG

Thanks for your attention!