Cancer High Throughput Data Analysis Workshop

15 Ordibehest, Spring 1402

Session 3 :

An Introduction To Single Cell RNA-Seq Data Analysis

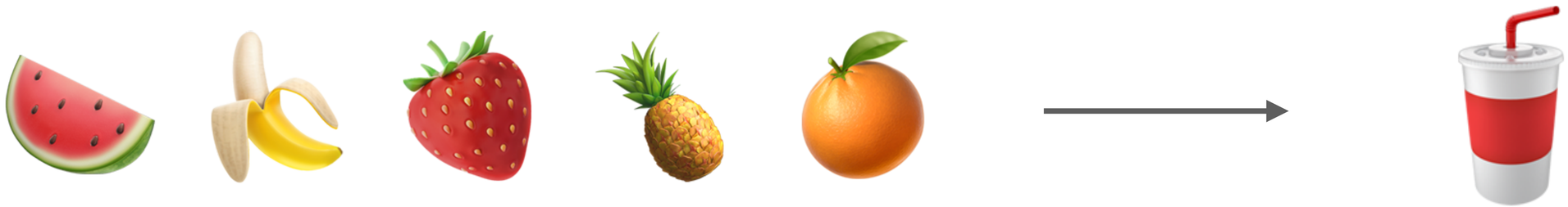
Arsham Mikaeili Namini

Section 1

Introducing the Single Cell Sequencing

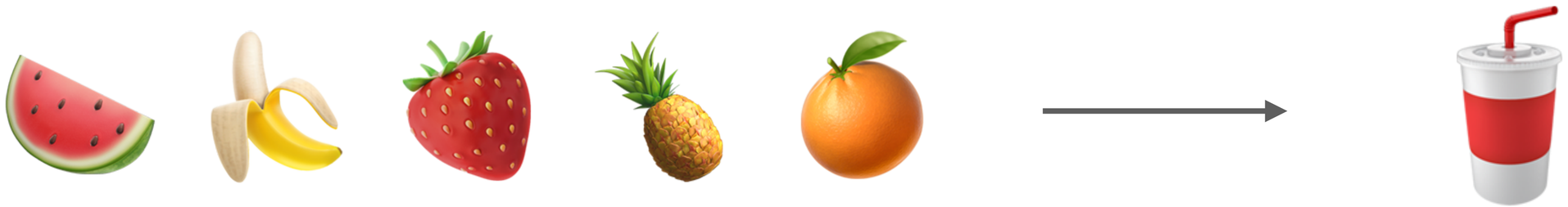
Bulk and Single Cell Sequencing

Fruit Smoothie Analogy



Bulk and Single Cell Sequencing

Fruit Smoothie Analogy



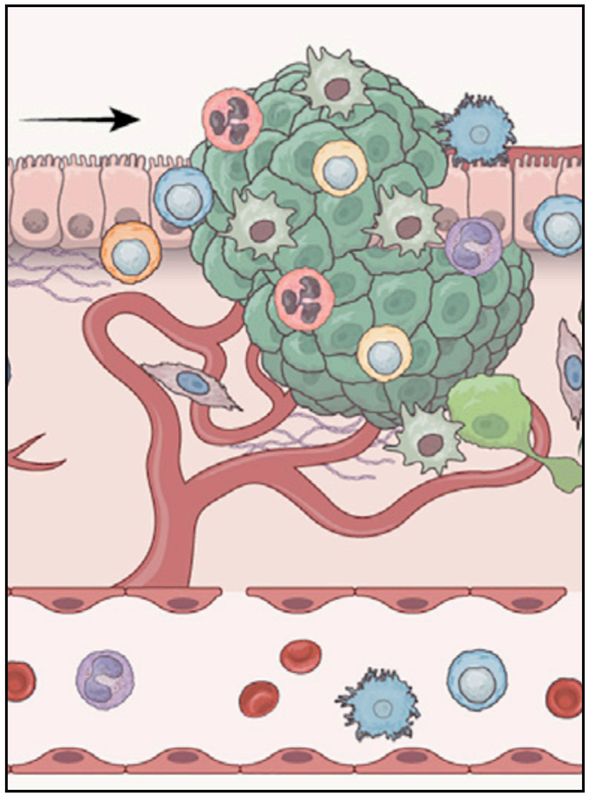
Single Cells

Bulk

Bulk and Single Cell Sequencing

Fruit Smoothie Analogy

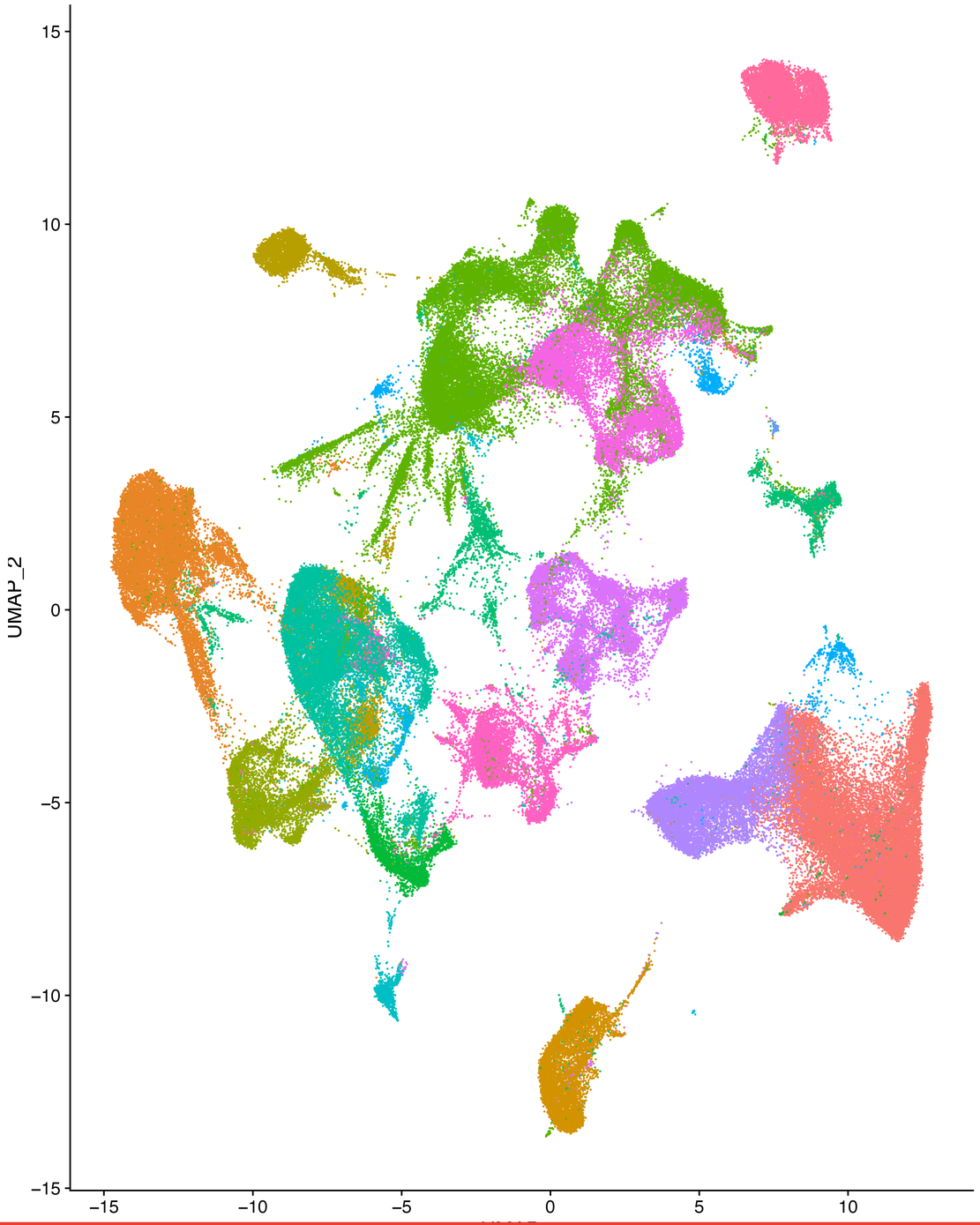
Single Cells



https://doi.org/10.1016/j.ccell.2023.02.016

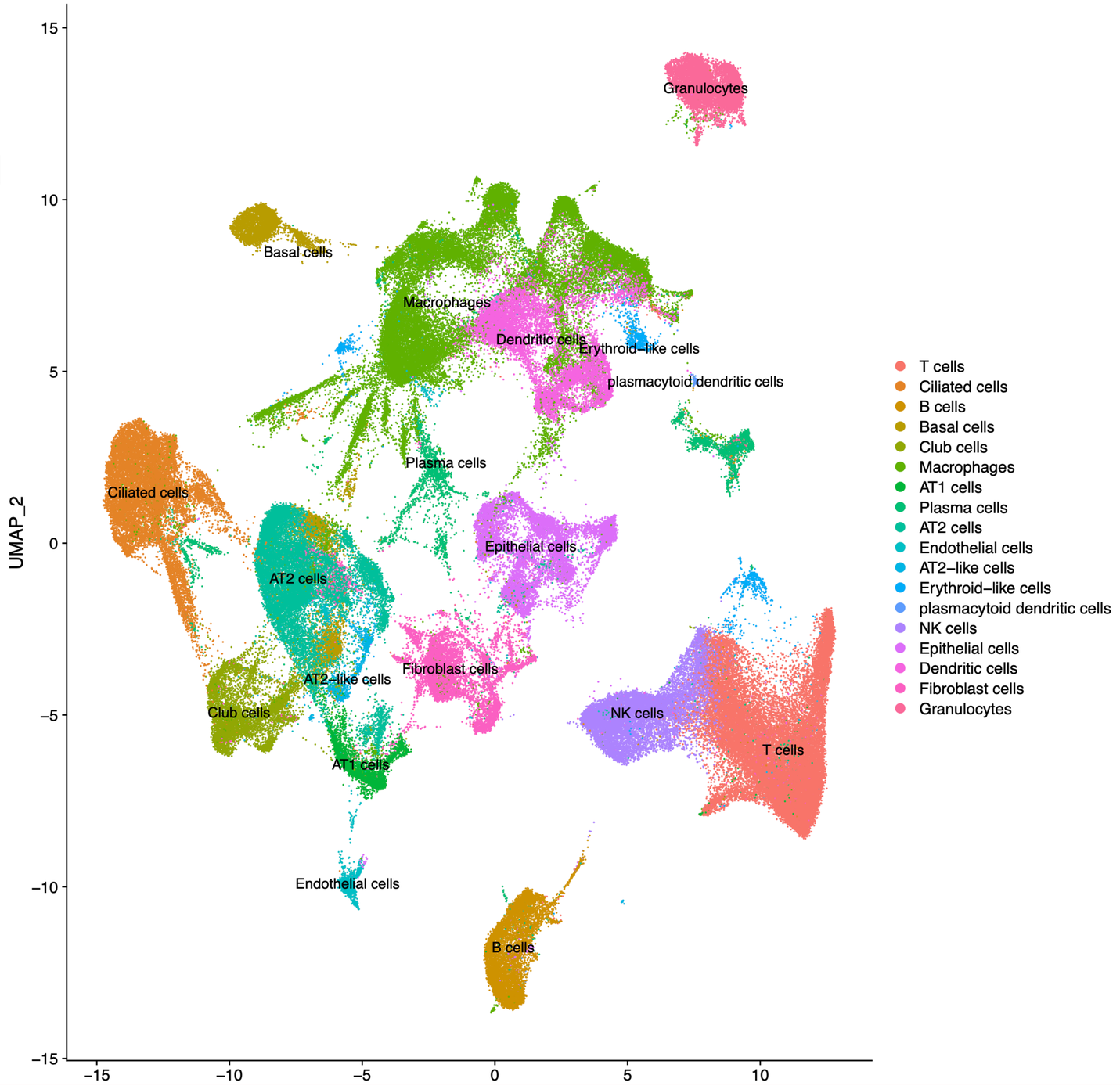
Bulk and Single Cell Sequencing

Cell Type Identification in Single cell Sequencing



Bulk and Single Cell Sequencing

Cell Type Identification in Single cell Sequencing



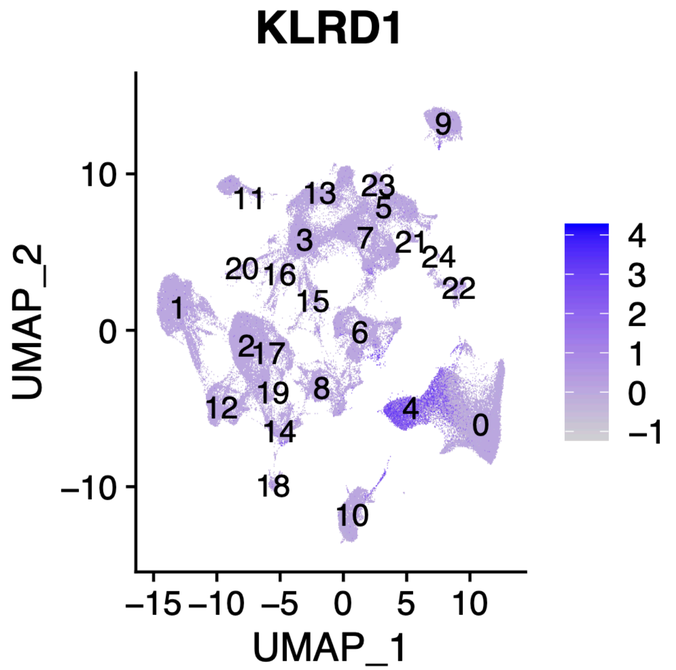
Stop and Think

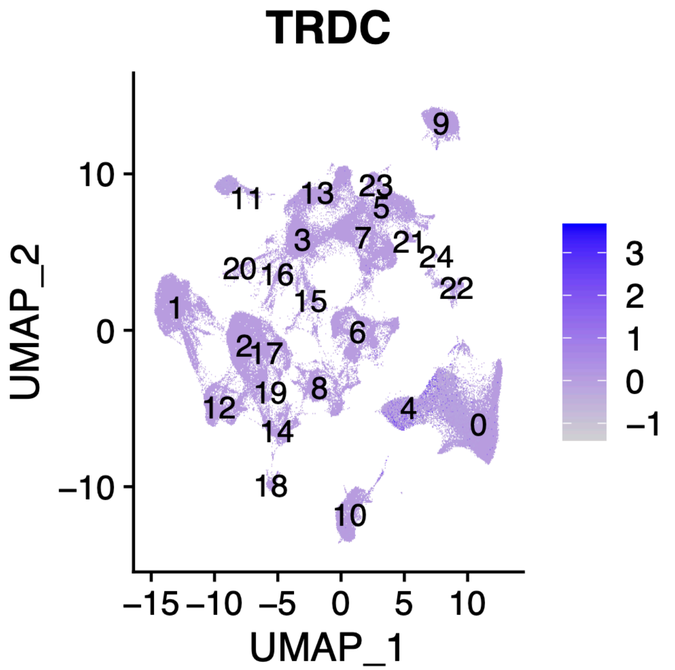
We can have same parallel information by performing Flow Cytometry on our data,

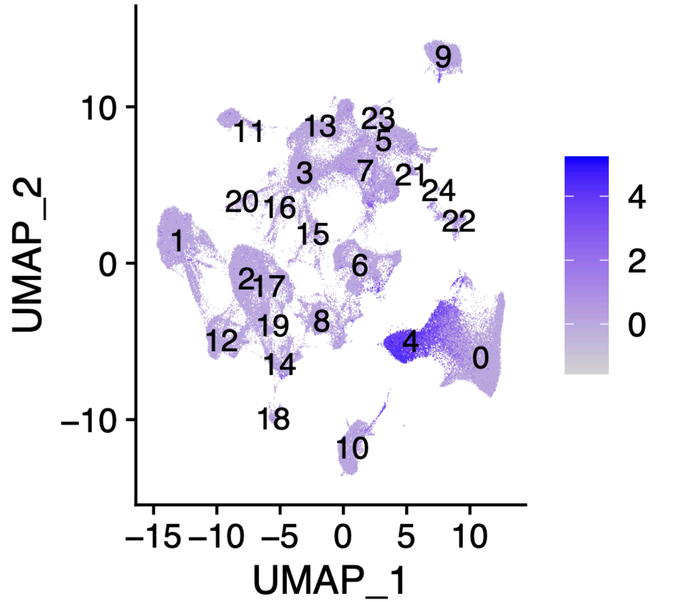
So, why we need to use scRNA sequencing?

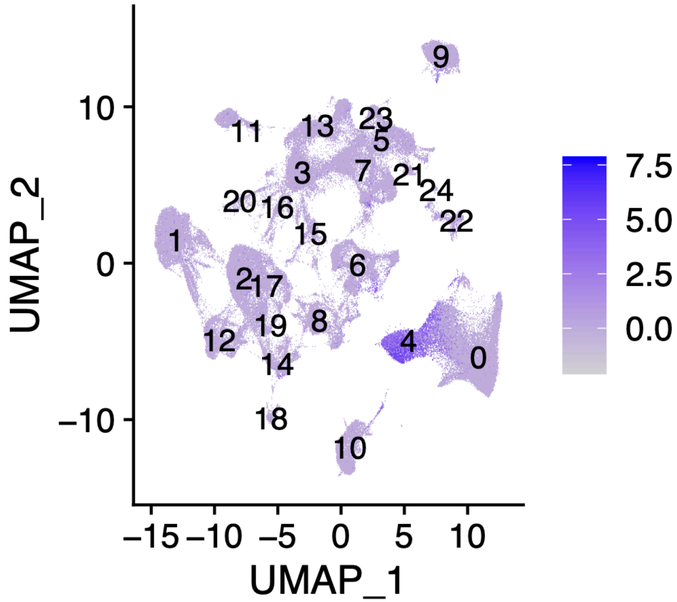
Bulk and Single Cell Sequencing

Cell Type Identification in Single cell Sequencing



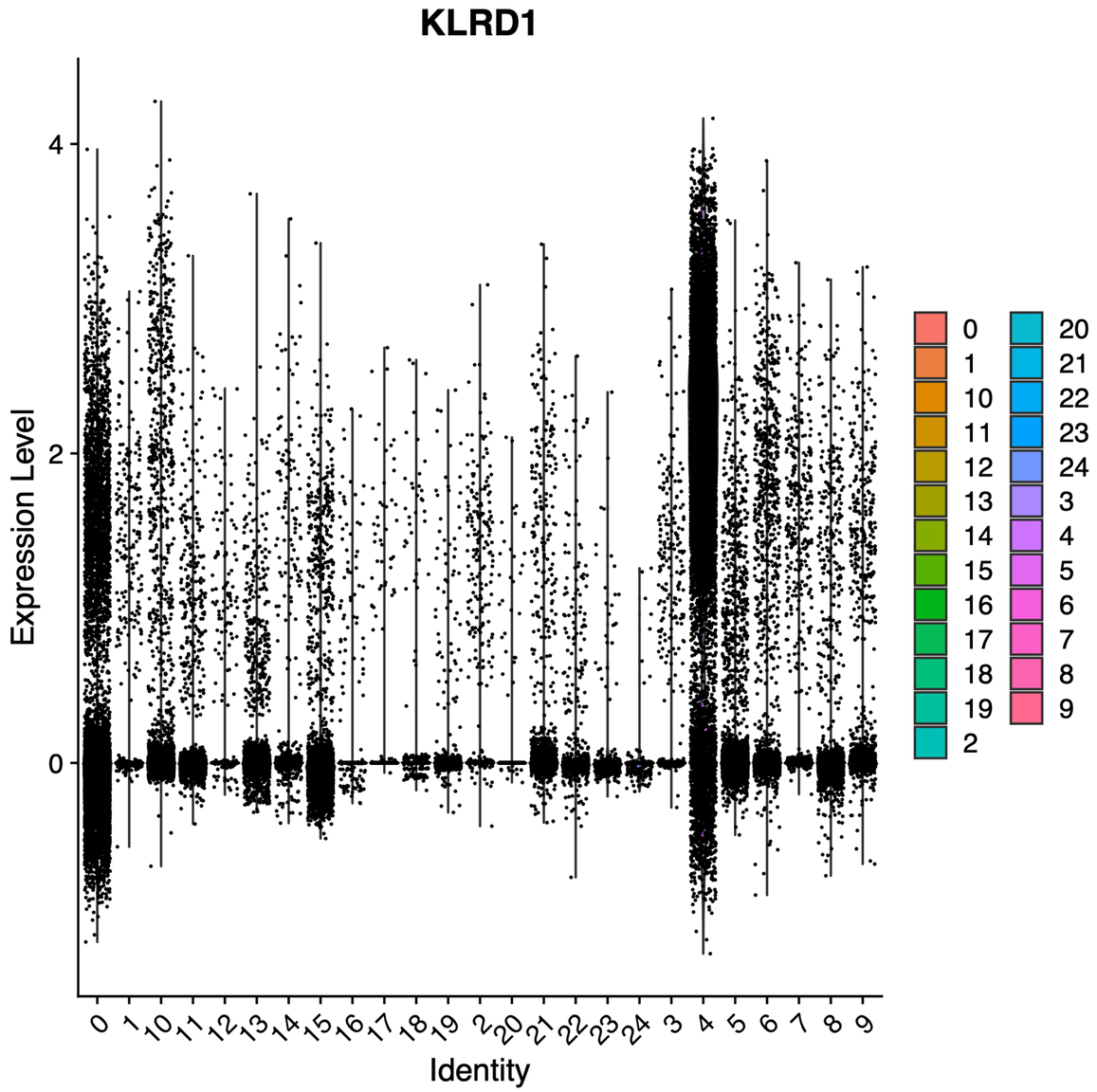






Bulk and Single Cell Sequencing

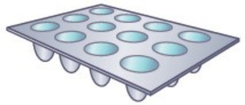
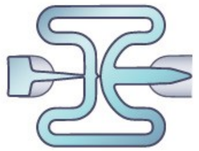
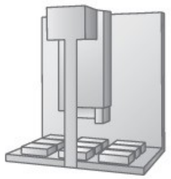
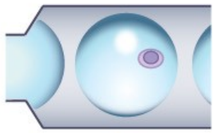
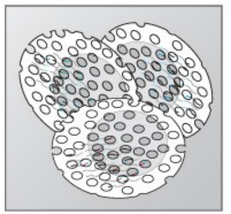
Cell Type Identification in Single cell Sequencing



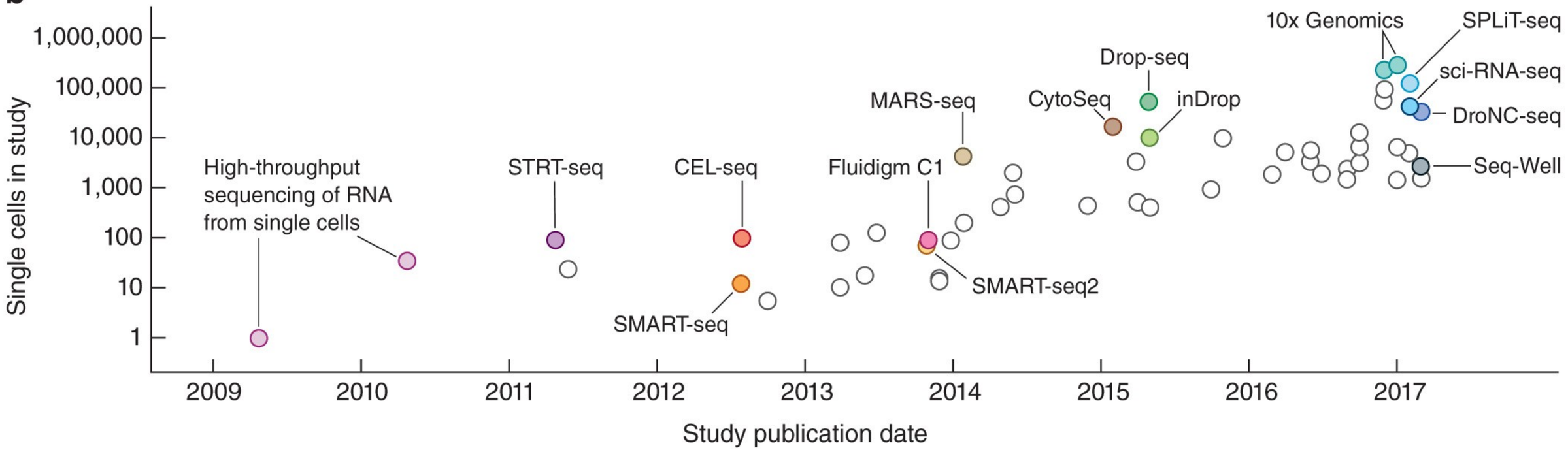
Section 2

Introducing the Single Cell Sequencing and 10X Genomics Platform

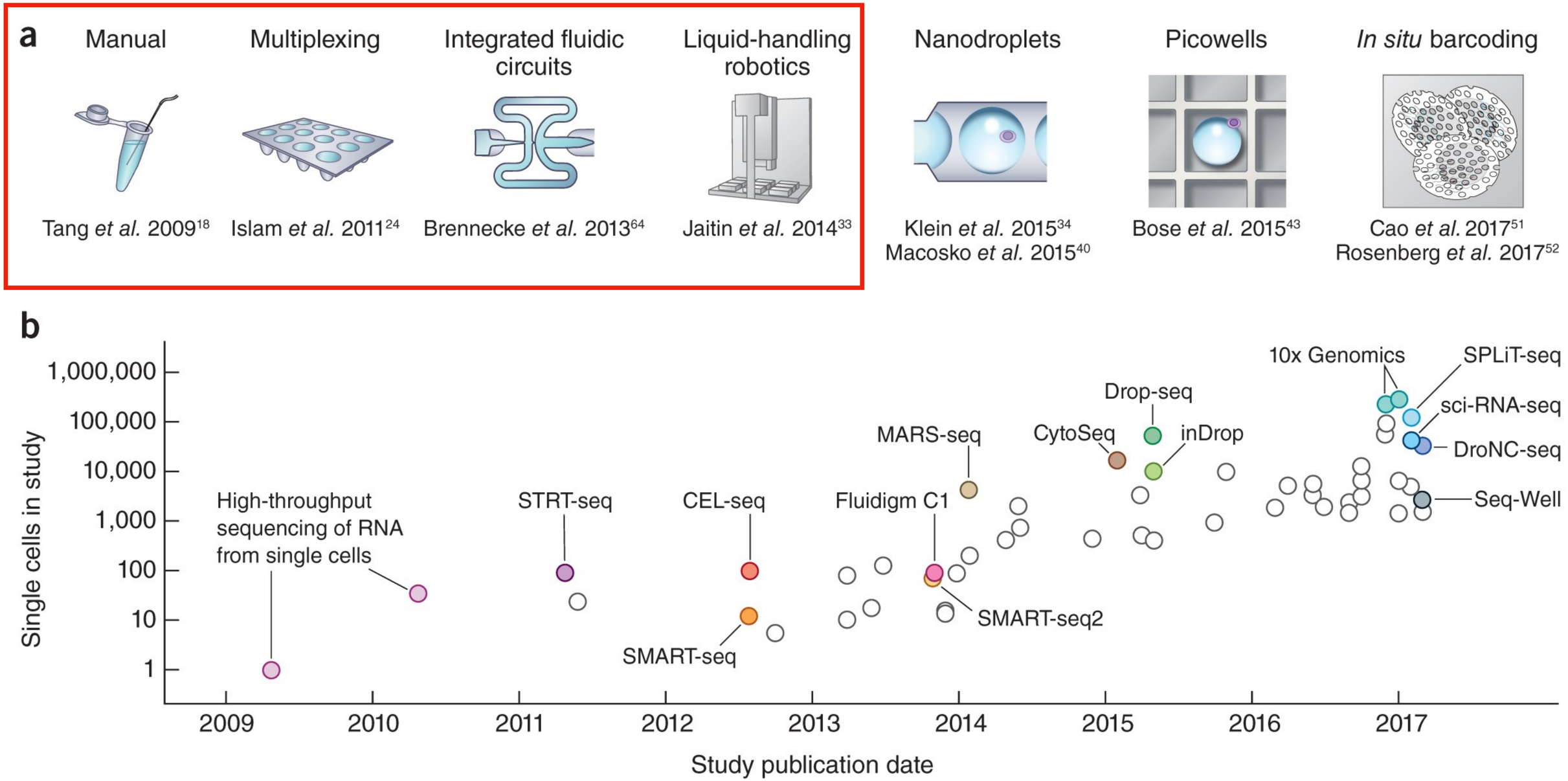
a

Manual	Multiplexing	Integrated fluidic circuits	Liquid-handling robotics	Nanodroplets	Picowells	*In situ* barcoding
						
Tang *et al.* 2009[18]	Islam *et al.* 2011[24]	Brennecke *et al.* 2013[64]	Jaitin *et al.* 2014[33]	Klein *et al.* 2015[34] Macosko *et al.* 2015[40]	Bose *et al.* 2015[43]	Cao *et al.* 2017[51] Rosenberg *et al.* 2017[52]

b



https://doi.org/10.1038/nmeth.2714



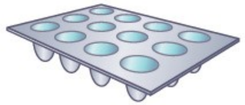
https://doi.org/10.1038/nmeth.2714

a

Manual

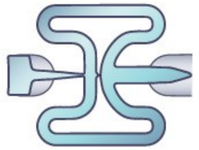
Tang *et al.* 2009[18]

Multiplexing



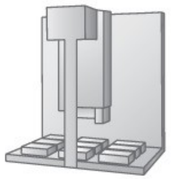
Islam *et al.* 2011[24]

Integrated fluidic circuits



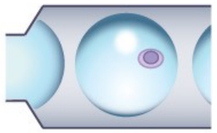
Brennecke *et al.* 2013[64]

Liquid-handling robotics



Jaitin *et al.* 2014[33]

Nanodroplets

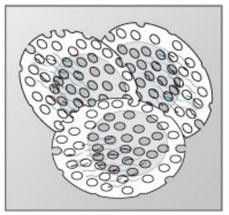


Klein *et al.* 2015[34]
Macosko *et al.* 2015[40]

Picowells

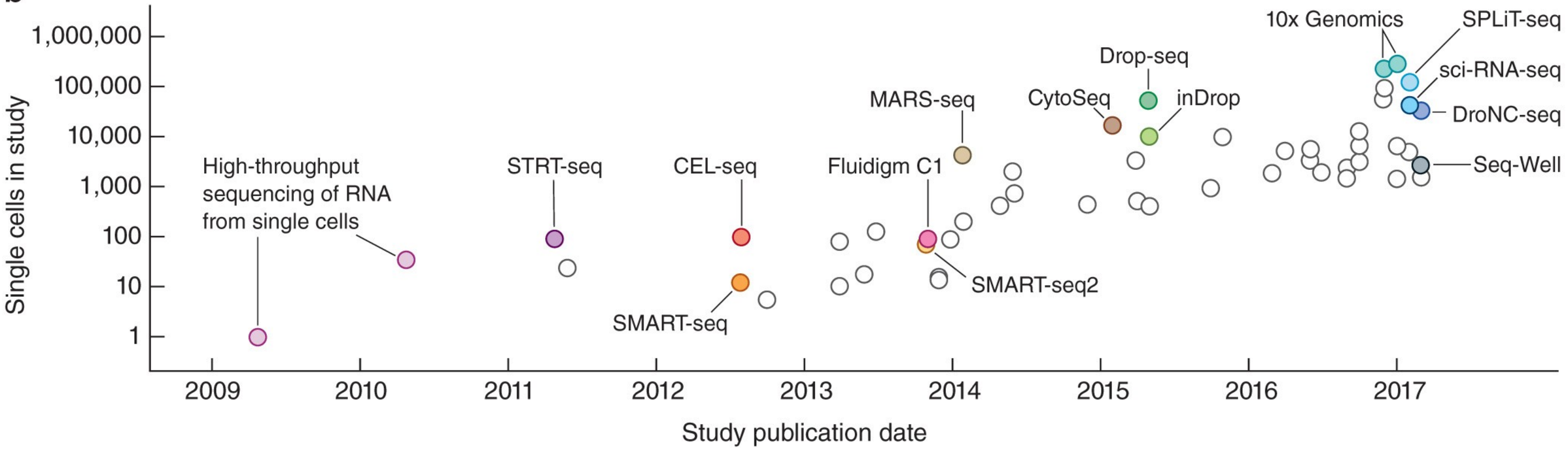
Bose *et al.* 2015[43]

In situ barcoding



Cao *et al.* 2017[51]
Rosenberg *et al.* 2017[52]

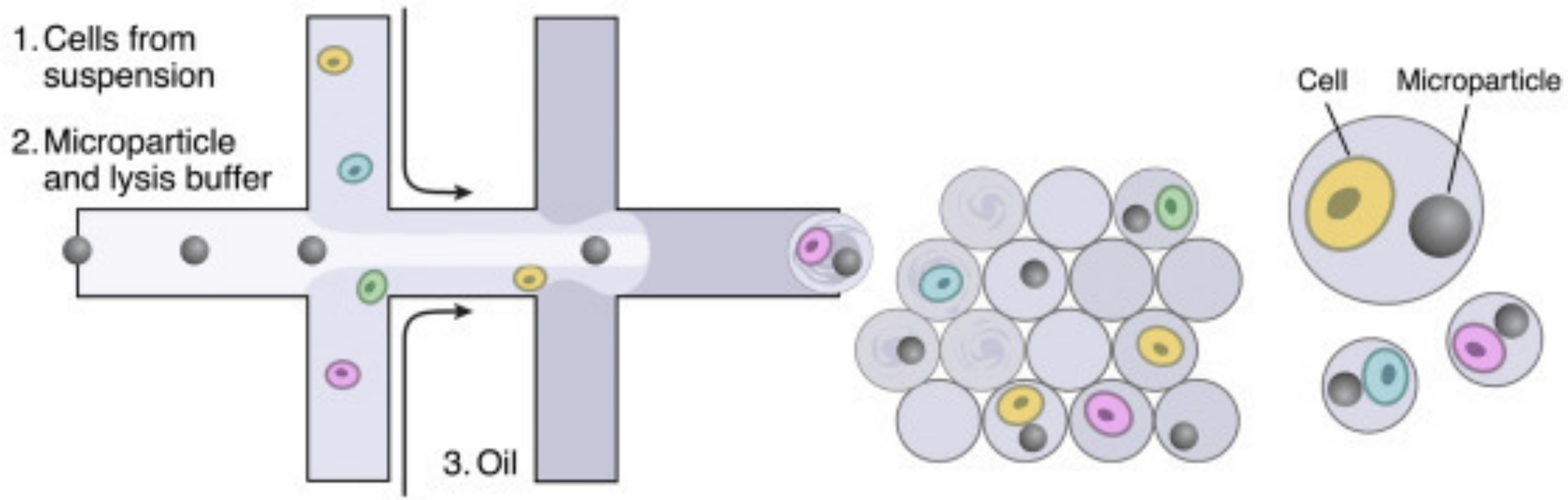
b



https://doi.org/10.1038/nmeth.2714

Novel Single Cell RNA Sequencing Techniques

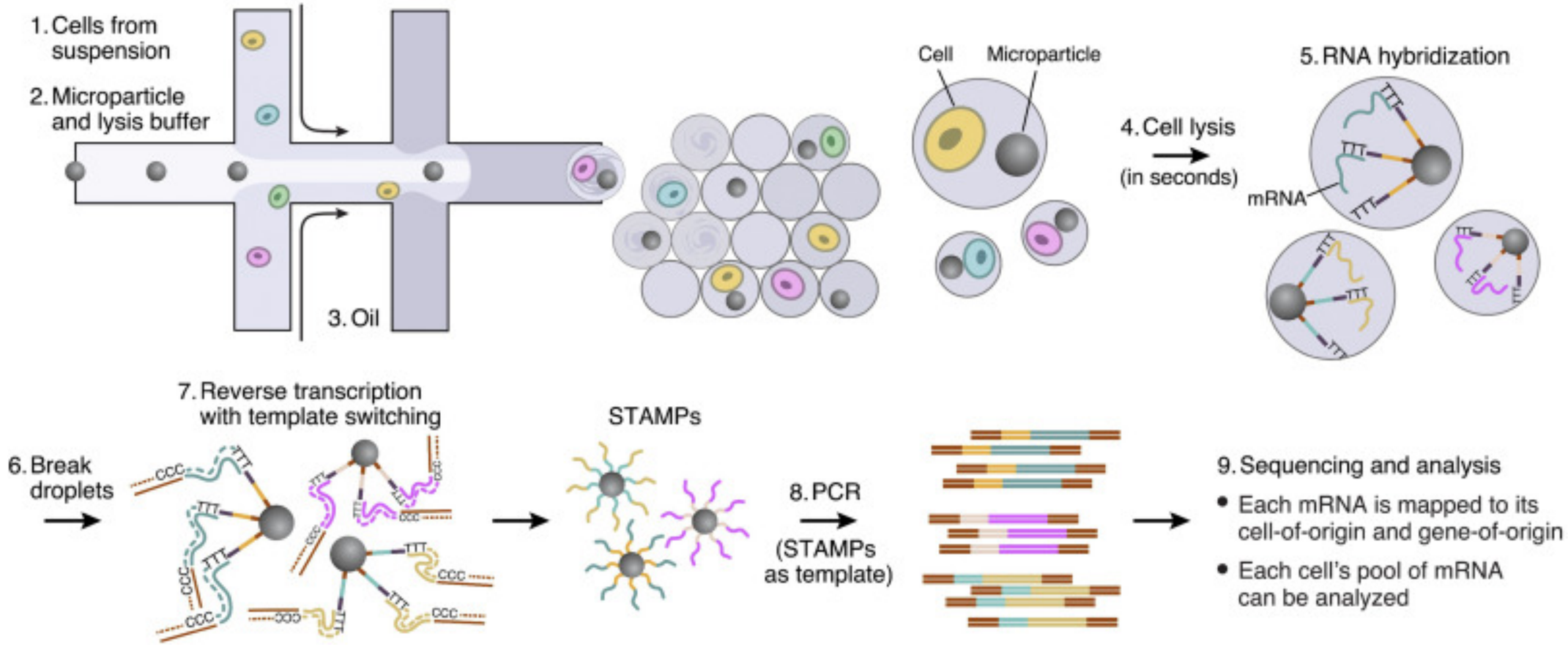
Drop Seq



https://doi.org/10.1016/j.biotechadv.2020.107537

Novel Single Cell RNA Sequencing Techniques

Drop Seq



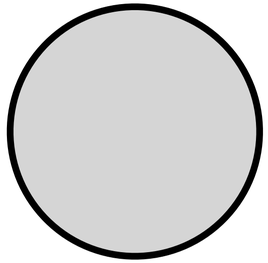
https://doi.org/10.1016/j.biotechadv.2020.107537

Novel Single Cell RNA Sequencing Techniques

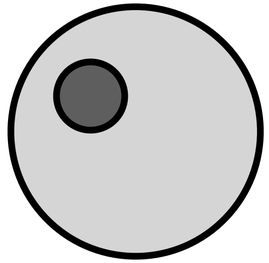
Drop Seq

Cell

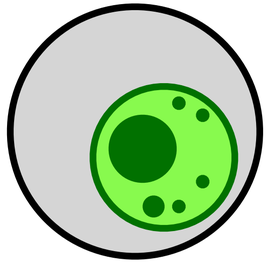
Bead



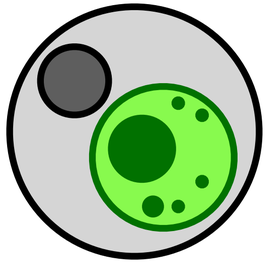
Empty Droplet



Contain a bead

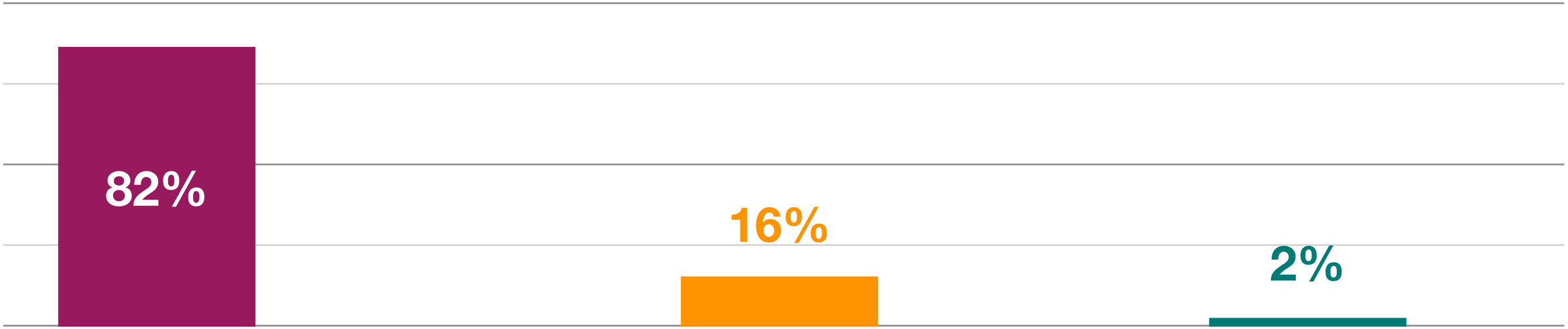


Contain a cell



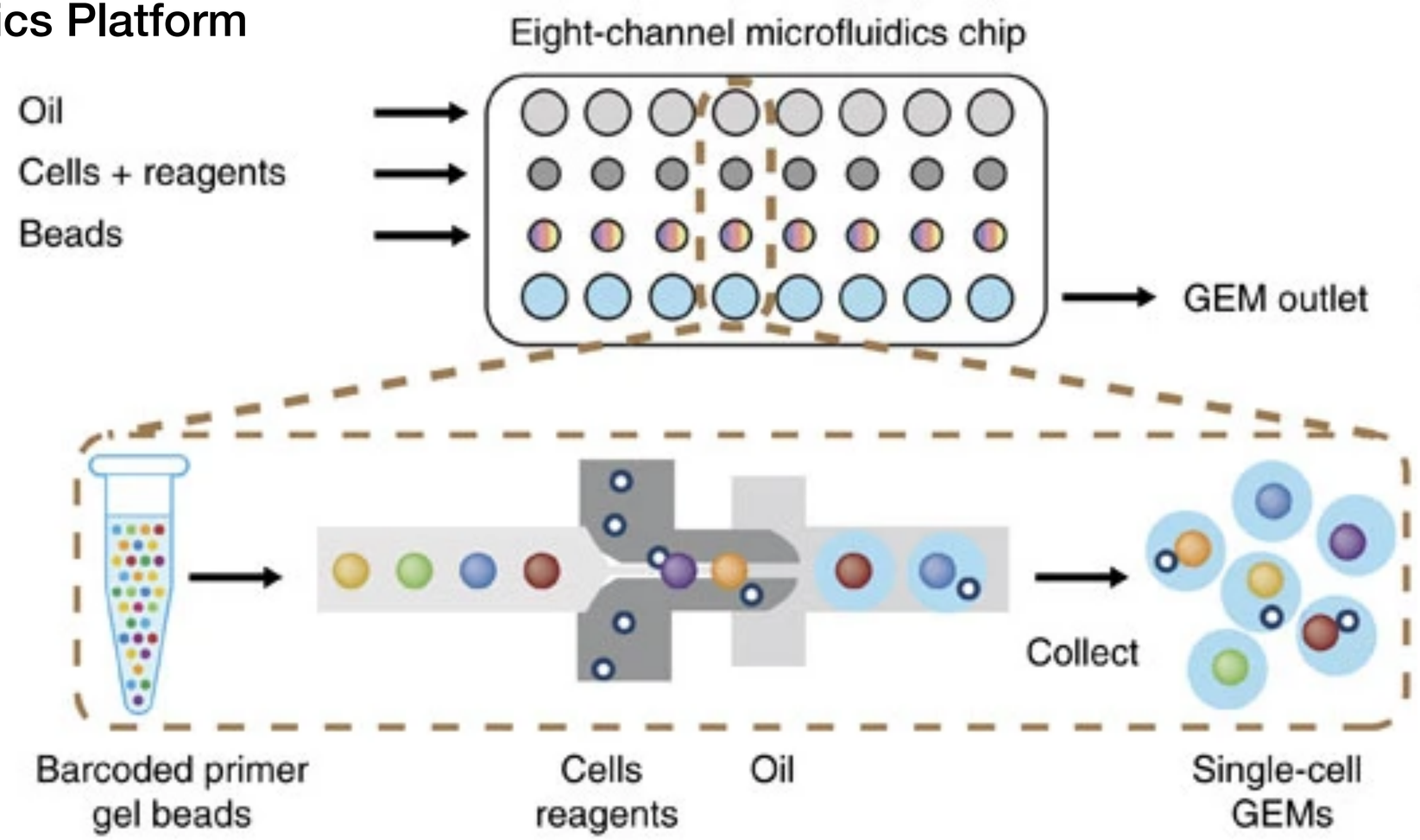
Contain a cell and a bead

Optimal



Novel Single Cell RNA Sequencing Techniques

10X Genomics Platform



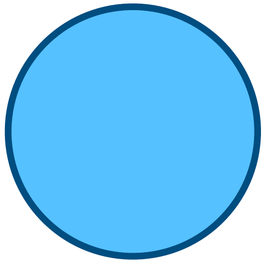
https://doi.org/10.1038/ncomms14049

Novel Single Cell RNA Sequencing Techniques

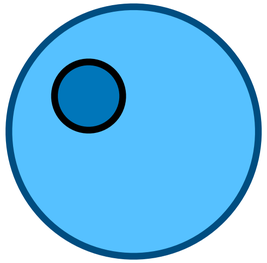
10X Genomics Platform

Cell

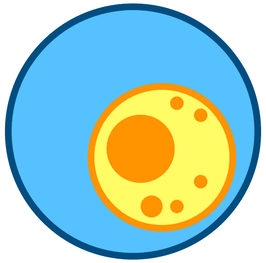
Bead



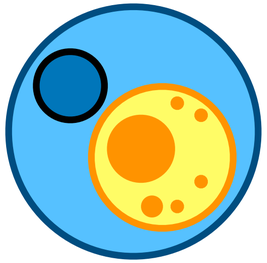
Empty Droplet



Contain a bead

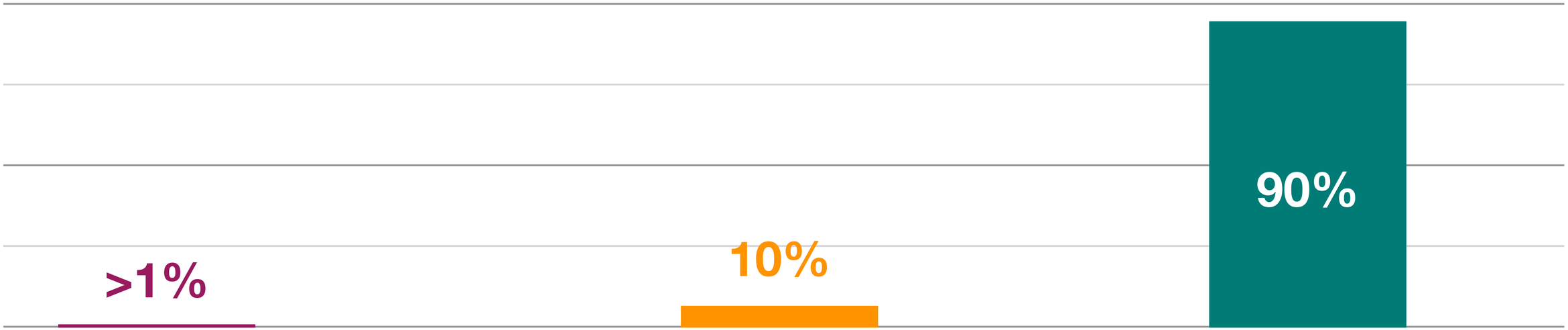


Contain a cell



Contain a cell and a bead

Optimal



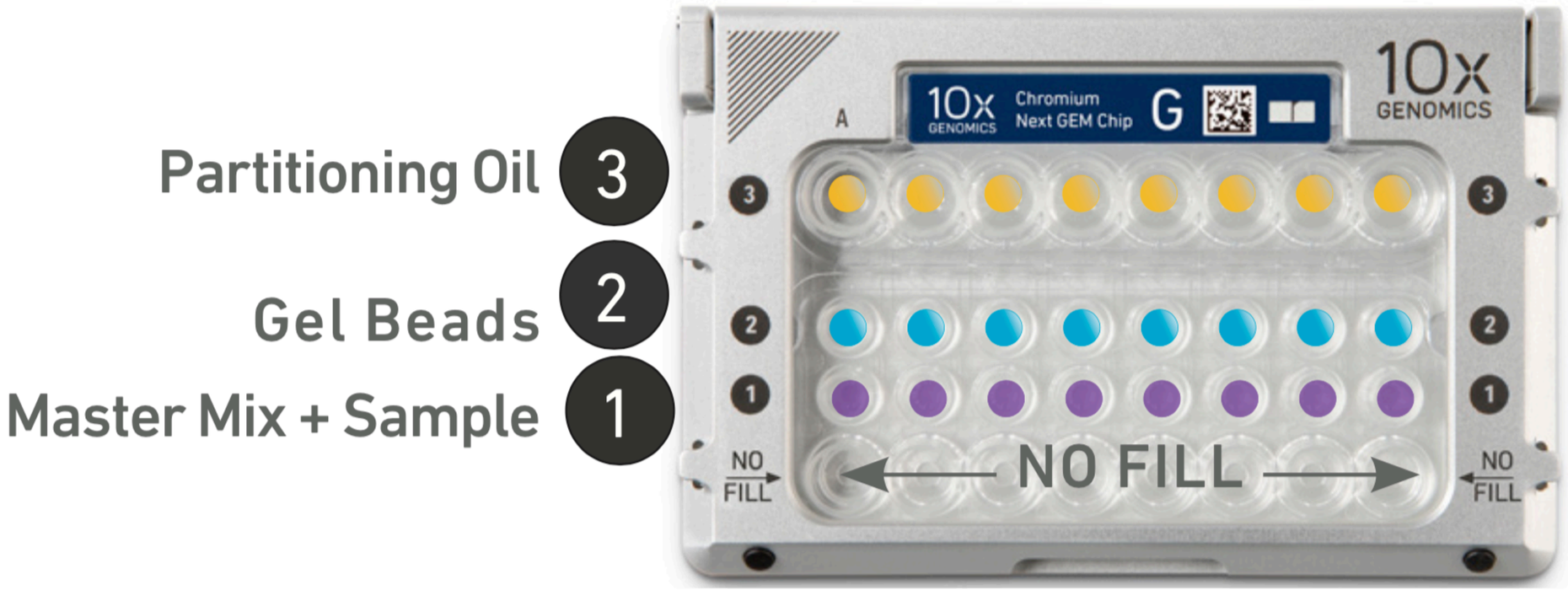
Novel Single Cell RNA Sequencing Techniques

10X Genomics Platform

- Each lanes of the chip is capable to sequence up to 25,000 cells
- Each chip has 8 lanes
- In one single run you can process 200,000 cells
- The process time is 15 minute per chip

Novel Single Cell RNA Sequencing Techniques

10X Genomics Platform

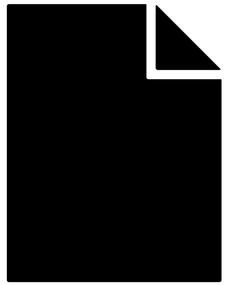


Section 3

Introducing to the 10X Genomics Data Structure

10X Genomics Data Structure

We have 3 Files



barcodes.tsv

features.tsv

matrix.mtx

	AAACCCAAGACCTTTG.1
1	AAACCCAAGCCTGAAG-1
2	AAACCCAAGCTGTGCC-1
3	AAACCCAAGGCCACTC-1
4	AAACCCAAGGGTTAAT-1
5	AAACCCAAGTCGAATA-1
6	AAACCCAAGTTCCTGA-1
7	AAACCCACAACACGAG-1
8	AAACCCACAACTGGTT-1
9	AAACCCACAGGGATAC-1
10	AAACCCACAGTAACAA-1
11	AAACCCACAGTCAGAG-1
12	AAACCCAGTCCAAATC-1
13	AAACCCAGTCGCGTCA-1
14	AAACCCAGTGCTCTCT-1
15	AAACCCAGTTGGCTAT-1
16	AAACCCATCGGAGCAA-1
17	AAACGAAAGAATCTAG-1
18	AAACGAAAGCGGACAT-1
19	AAACGAAAGGTGCTGA-1
20	AAACGAAAGTCACTGT-1
21	AAACGAACAATTGGTC-1
22	AAACGAACAATTGTGC-1
23	AAACGAACACCTTCCA-1
24	AAACGAACATCACCAA-1
25	AAACGAACATGACAAA-1
26	AAACGAAGTGTTACAC-1
27	AAACGAATCCACTGGG-1
28	AAACGAATCTAGTCAG-1
29	AAACGCTAGCCATGCC-1

Showing 1 to 29 of 15,215 entries, 1 total columns

barcodes.tsv

1	AAACCCAAGCCTGAAG-1
2	AAACCCAAGCTGTGCC-1
3	AAACCCAAGGCCACTC-1
4	AAACCCAAGGGTTAAT-1
5	AAACCCAAGTCGAATA-1

Size : 289 kb
Dimension : [1] 15215   1

features.tsv

	ENSG00000243485	MIR1302.2HG	Gene.Expression
1	ENSG00000237613	FAM138A	Gene Expression
2	ENSG00000186092	OR4F5	Gene Expression
3	ENSG00000238009	AL627309.1	Gene Expression
4	ENSG00000239945	AL627309.3	Gene Expression
5	ENSG00000239906	AL627309.2	Gene Expression
6	ENSG00000241599	AL627309.4	Gene Expression
7	ENSG00000236601	AL732372.1	Gene Expression
8	ENSG00000284733	OR4F29	Gene Expression
9	ENSG00000235146	AC114498.1	Gene Expression
10	ENSG00000284662	OR4F16	Gene Expression
11	ENSG00000229905	AL669831.2	Gene Expression
12	ENSG00000237491	AL669831.5	Gene Expression
13	ENSG00000177757	FAM87B	Gene Expression
14	ENSG00000225880	LINC00115	Gene Expression
15	ENSG00000230368	FAM41C	Gene Expression
16	ENSG00000272438	AL645608.7	Gene Expression
17	ENSG00000230699	AL645608.3	Gene Expression
18	ENSG00000241180	AL645608.5	Gene Expression
19	ENSG00000223764	AL645608.1	Gene Expression
20	ENSG00000187634	SAMD11	Gene Expression
21	ENSG00000188976	NOC2L	Gene Expression
22	ENSG00000187961	KLHL17	Gene Expression
23	ENSG00000187583	PLEKHN1	Gene Expression
24	ENSG00000187642	PERM1	Gene Expression
25	ENSG00000272512	AL645608.8	Gene Expression
26	ENSG00000188290	HES4	Gene Expression
27	ENSG00000187608	ISG15	Gene Expression
28	ENSG00000224969	AL645608.2	Gene Expression
29	ENSG00000188157	AGRN	Gene Expression

Showing 1 to 29 of 33,537 entries, 3 total columns

13	ENSG00000177757	FAM87B	Gene Expression
14	ENSG00000225880	LINC00115	Gene Expression
15	ENSG00000230368	FAM41C	Gene Expression
16	ENSG00000272438	AL645608.7	Gene Expression
17	ENSG00000230699	AL645608.3	Gene Expression

```
Size : 1.3 MB
Dimension : [1] 33537     3
```

features.tsv

	ENSG00000243485	MIR1302.2HG	Gene.Expression
1	ENSG00000237613	FAM138A	Gene Expression
2	ENSG00000186092	OR4F5	Gene Expression
3	ENSG00000238009	AL627309.1	Gene Expression
4	ENSG00000239945	AL627309.3	Gene Expression
5	ENSG00000239906	AL627309.2	Gene Expression
6	ENSG00000241599	AL627309.4	Gene Expression
7	ENSG00000236601	AL732372.1	Gene Expression
8	ENSG00000284733	OR4F29	Gene Expression
9	ENSG00000235146	AC114498.1	Gene Expression
10	ENSG00000284662	OR4F16	Gene Expression
11	ENSG00000229905	AL669831.2	Gene Expression
12	ENSG00000237491	AL669831.5	Gene Expression
13	ENSG00000177757	FAM87B	Gene Expression
14	ENSG00000225880	LINC00115	Gene Expression
15	ENSG00000230368	FAM41C	Gene Expression
16	ENSG00000272438	AL645608.7	Gene Expression
17	ENSG00000230699	AL645608.3	Gene Expression
18	ENSG00000241180	AL645608.5	Gene Expression
19	ENSG00000223764	AL645608.1	Gene Expression
20	ENSG00000187634	SAMD11	Gene Expression
21	ENSG00000188976	NOC2L	Gene Expression
22	ENSG00000187961	KLHL17	Gene Expression
23	ENSG00000187583	PLEKHN1	Gene Expression
24	ENSG00000187642	PERM1	Gene Expression
25	ENSG00000272512	AL645608.8	Gene Expression
26	ENSG00000188290	HES4	Gene Expression
27	ENSG00000187608	ISG15	Gene Expression
28	ENSG00000224969	AL645608.2	Gene Expression
29	ENSG00000188157	AGRN	Gene Expression

Showing 1 to 29 of 33,537 entries, 3 total columns

13	ENSG00000177757	FAM87B	Gene Expression
14	ENSG00000225880	LINC00115	Gene Expression
15	ENSG00000230368	FAM41C	Gene Expression
16	ENSG00000272438	AL645608.7	Gene Expression
17	ENSG00000230699	AL645608.3	Gene Expression

```
Size : 1.3 MB
Dimension : [1] 33537     3
```

features.tsv

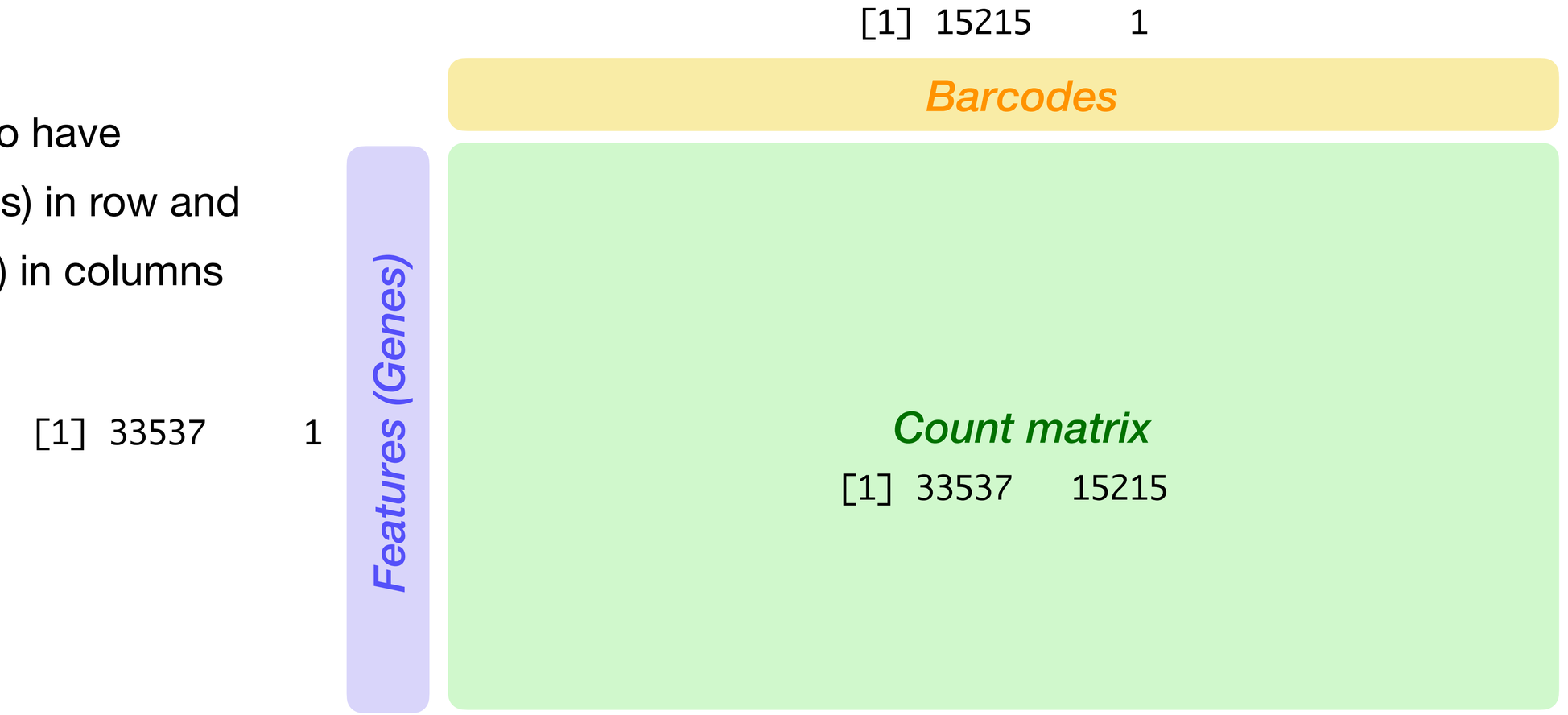
```
$ gzip -cd filtered_feature_bc_matrix/features.tsv.gz
ENSG00000141510       TP53          Gene Expression
ENSG00000012048       BRCA1         Gene Expression
ENSG00000139687       RB1           Gene Expression
CD3_GCCTGACTAGATCCA   CD3           Antibody Capture
CD19_CGTGCAACACTCGTA  CD19          Antibody Capture
```

https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/output/matrices

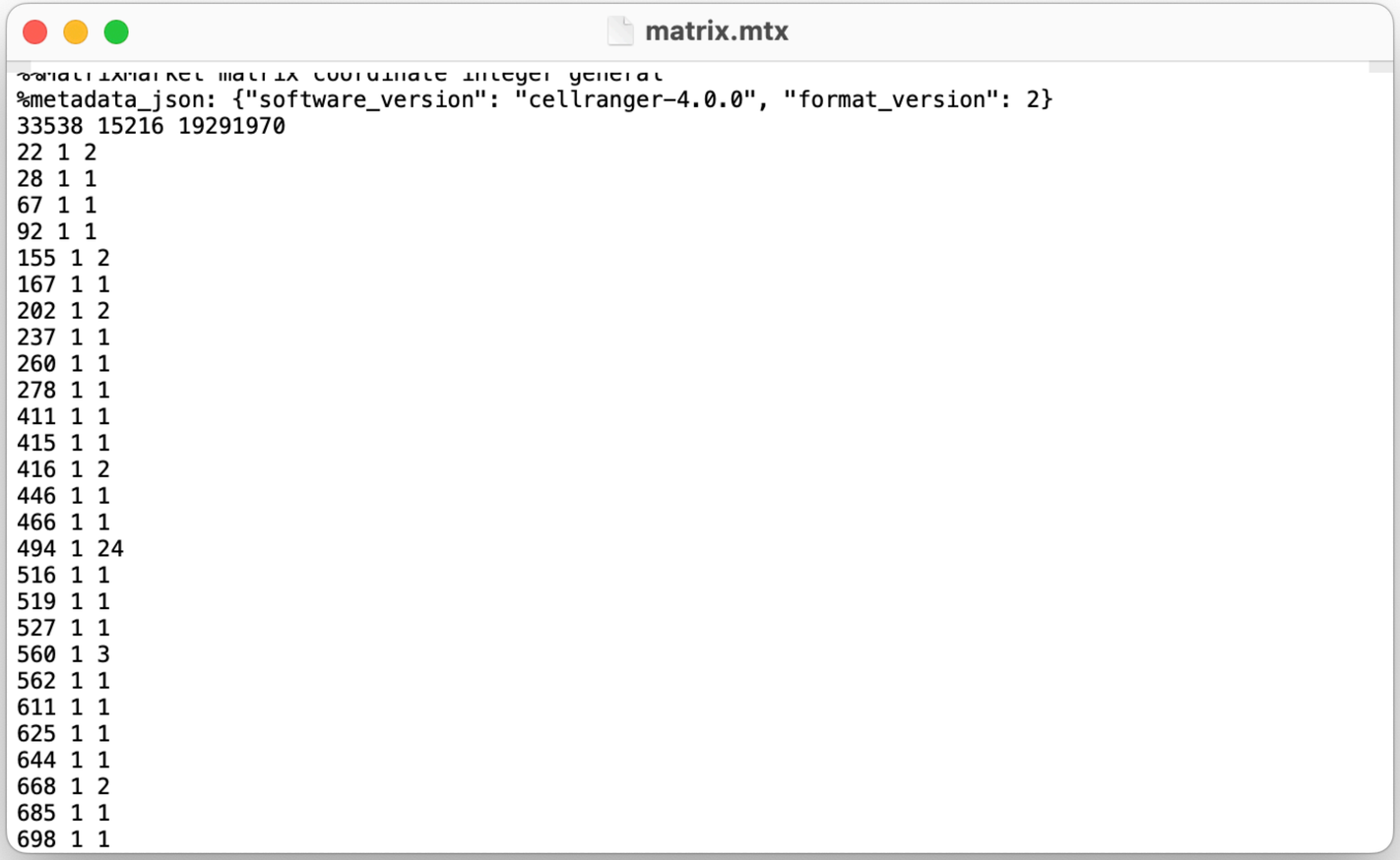
10X Genomics Data Structure

We have 3 Files

It is common to have features (Genes) in row and barcode (Cells) in columns

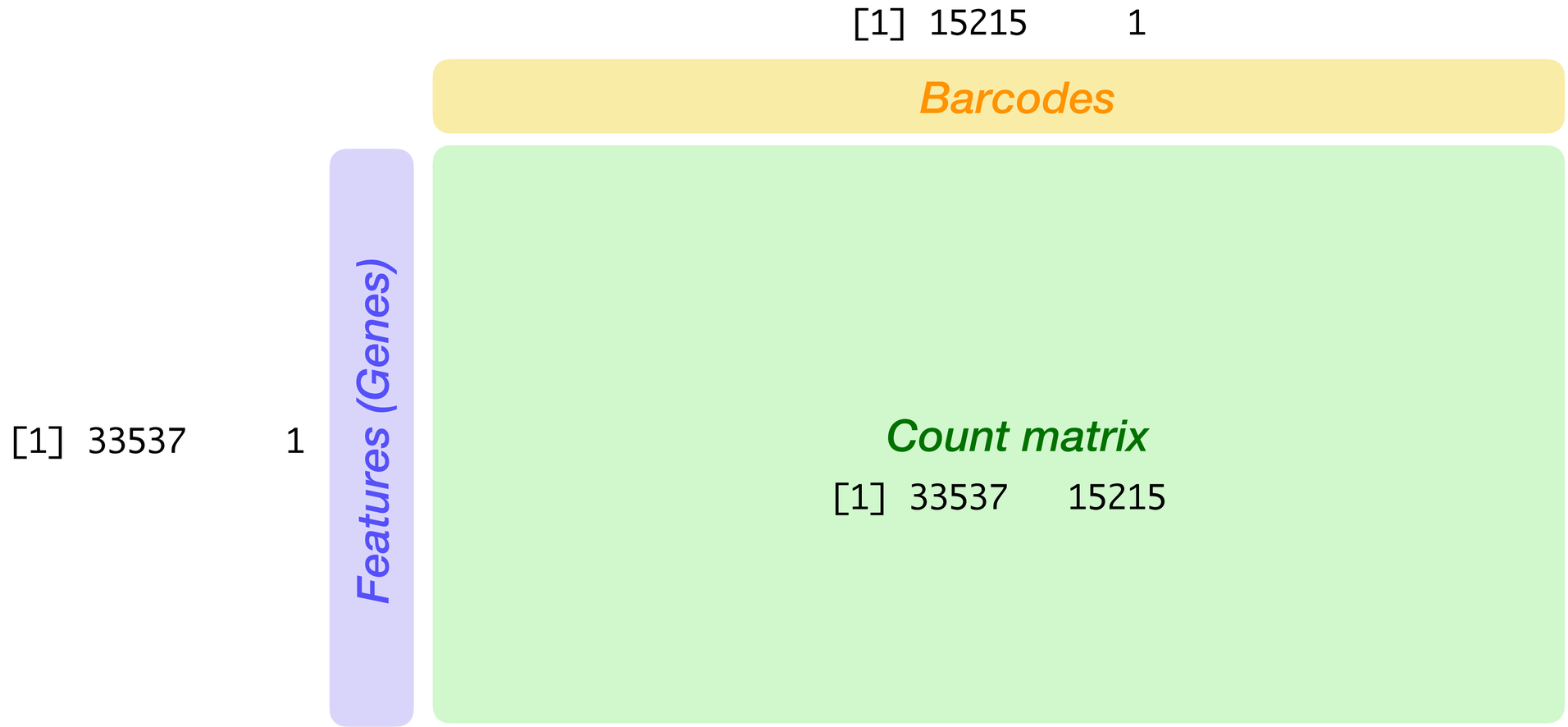


matrix.mtx



Size : 250.3 MB

Dimension : [1] 19291972 1

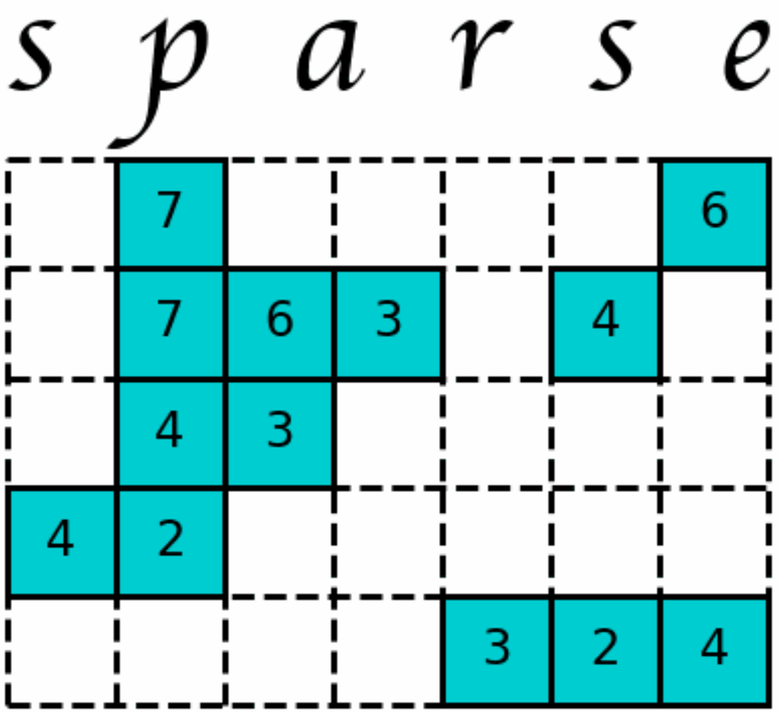


33537 * 15215 = 510,314,208

19,291,972

10X Genomics Data Structure

We have 3 Files



DENSE

0	7	0	0	0	0	6
0	7	6	3	0	4	0
0	4	3	0	0	0	0
4	2	0	0	0	0	0
0	0	0	0	3	2	4

10X Genomics Data Structure

We have 3 Files

```r
pbmc.data[c("CD3D", "TCL1A", "MS4A1"), 1:30]
```

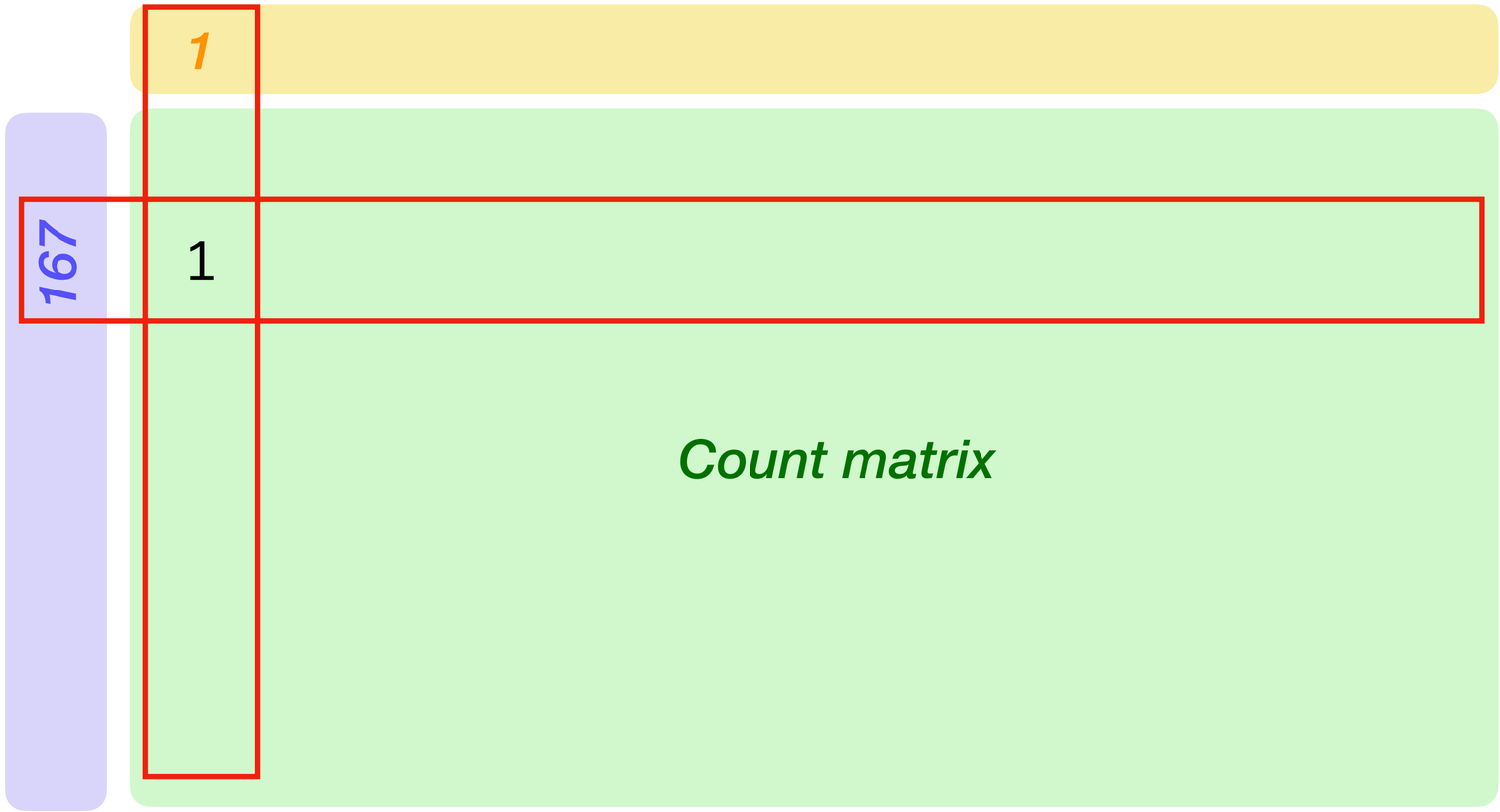
```
3 x 30 sparse Matrix of class "dgCMatrix"
  [[ suppressing 30 column names ‘AAACCCAAGCTCGTGC-1’, ‘AAACCCAAGTCAGCGA-1’,
‘AAACCCACAACGGTAG-1’ ... ]]

CD3D  . . . . . 2 . . . . 2 3 1 . . . 3 . . 1 2 . 2 5 . 1 . . . .
TCL1A . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
MS4A1 . . . . . . . . . . . . . 1 . . . . . . . . . . . 1 . . . . 2
```

	featurs	barcodes	expression
1	33538	15216	19291970
2	22	1	2
3	28	1	1
4	67	1	1
5	92	1	1
6	155	1	2
7	167	1	1
8	202	1	2
9	237	1	1
10	260	1	1
11	278	1	1
12	411	1	1
13	415	1	1
14	416	1	2
15	446	1	1
16	466	1	1
17	494	1	24
18	516	1	1
19	519	1	1
20	527	1	1
21	560	1	3
22	562	1	1
23	611	1	1
24	625	1	1
25	644	1	1
26	668	1	2
27	685	1	1
28	698	1	1
29	782	1	1

Showing 1 to 29 of 99 entries, 3 total columns

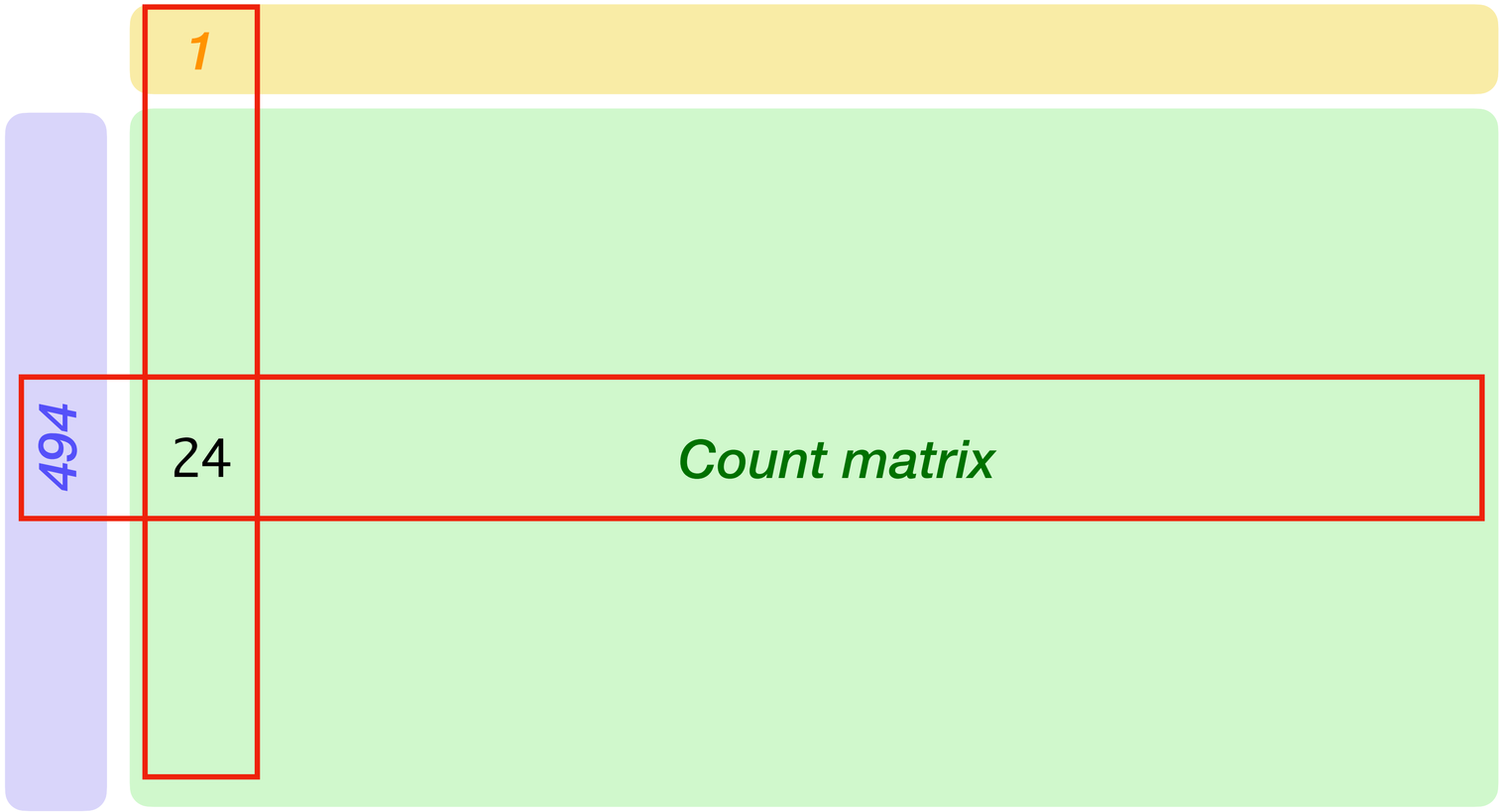
	Feature	Barcode	Raw expression count
7	167	1	1



	featurs	barcodes	expression
1	33538	15216	19291970
2	22	1	2
3	28	1	1
4	67	1	1
5	92	1	1
6	155	1	2
7	167	1	1
8	202	1	2
9	237	1	1
10	260	1	1
11	278	1	1
12	411	1	1
13	415	1	1
14	416	1	2
15	446	1	1
16	466	1	1
17	494	1	24
18	516	1	1
19	519	1	1
20	527	1	1
21	560	1	3
22	562	1	1
23	611	1	1
24	625	1	1
25	644	1	1
26	668	1	2
27	685	1	1
28	698	1	1
29	782	1	1

Showing 1 to 29 of 99 entries, 3 total columns

	Feature	Barcode	Raw expression count
17	494	1	24



The data we are using today in our lab

GSM5699781

Sample GSM5699781 Query DataSets for GSM5699781

Status	Public on Nov 30, 2021
Title	TD5 scRNA-seq
Sample type	SRA
Source name	Lung adenocarcinoma
Organism	Homo sapiens
Characteristics	histolgical type: AIS radiological type: SSN gender: Female
Extracted molecule	total RNA
Extraction protocol	A human tumour dissociation kit enzyme solution (Miltenyi Biotec; 200 µl of H-enzyme, 100 µl of R-enzyme, 25 µl of A-enzyme, 4.7 ml of DMEM) was added for enzymatic digestion for 30 minutes at 37°C, and the sample was filtered through a Miltenyi 70-mm sieve. After centrifugation, the granular cells were suspended in erythrocyte lysis buffer. Finally, the cells were mixed with 1 ml of PBS, and the numbers of live cells and aggregated cells were measured with an automatic cell counter For droplet-based scRNA-seq, GemCodeGel beads, chips and library kits (10x Genomics) were used to process single cells according to the gel beads in emulsion (GEMs) protocol (Genergy Inc., Shanghai, China). Based on the manufacturer's instructions, the Chromium Single Cell 3' V2/V3 Kit was used to construct the scRNA-seq reagent library. Then, the cell suspension generated from each sample was reacted with certain reagents to generate single-cell GEMs in the Chromium Controller for sample and cell barcoding, with a target output of 8,000 to 12,000 cells per sample. The amplified cDNA and final library were evaluated with an Agilent Bioanalyzer using a high-sensitivity DNA kit (2100, Agilent Technologies). Then, the libraries were assembled and sorted on a NovaSeq 6000; approximately 400 M readings were obtained for each sample.
Library strategy	RNA-Seq
Library source	transcriptomic
Library selection	cDNA
Instrument model	Illumina NovaSeq 6000
Data processing	The cellranger software suite was used to perform sample de-multiplexing. The final library pool was sequenced on the NovaSeq 6000 instrument using 150-base-pair paired-end reads. Genome_build: hg19 Supplementary_files_format_and_content: raw counts

Please Download these three files

Supplementary file	Size	Download	File type/resource
GSM5699781_TD5_barcodes.tsv.gz	90.6 Kb	(ftp)(http)	TSV
GSM5699781_TD5_features.tsv.gz	297.6 Kb	(ftp)(http)	TSV
GSM5699781_TD5_matrix.mtx.gz	86.7 Mb	(ftp)(http)	MTX

SRA Run Selector

Raw data are available in SRA

Processed data provided as supplementary file

Thanks For Your Attention!

If you have any further questions please feel free to email me

arshammikaeili [at] gmail [dot] com