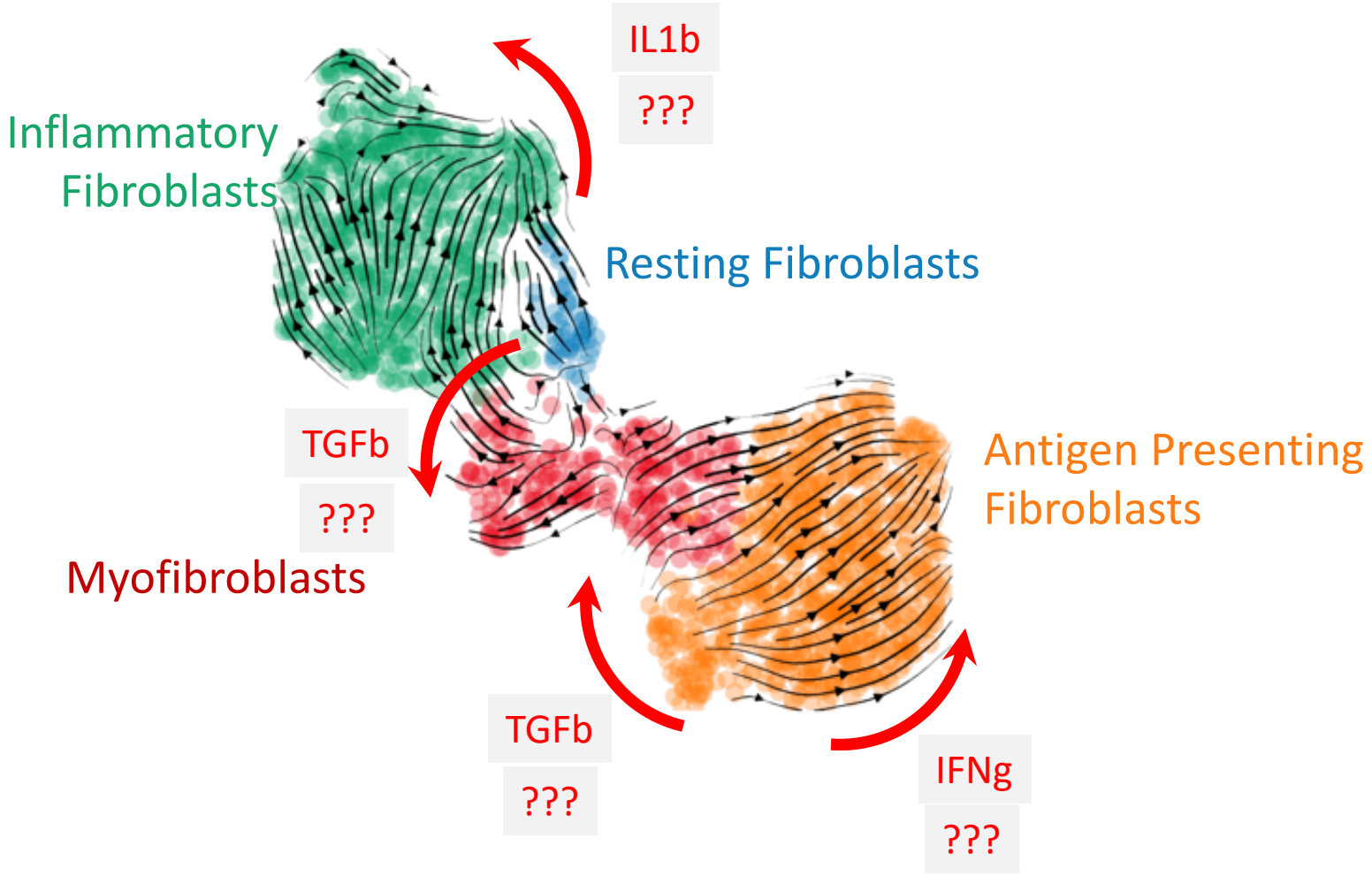
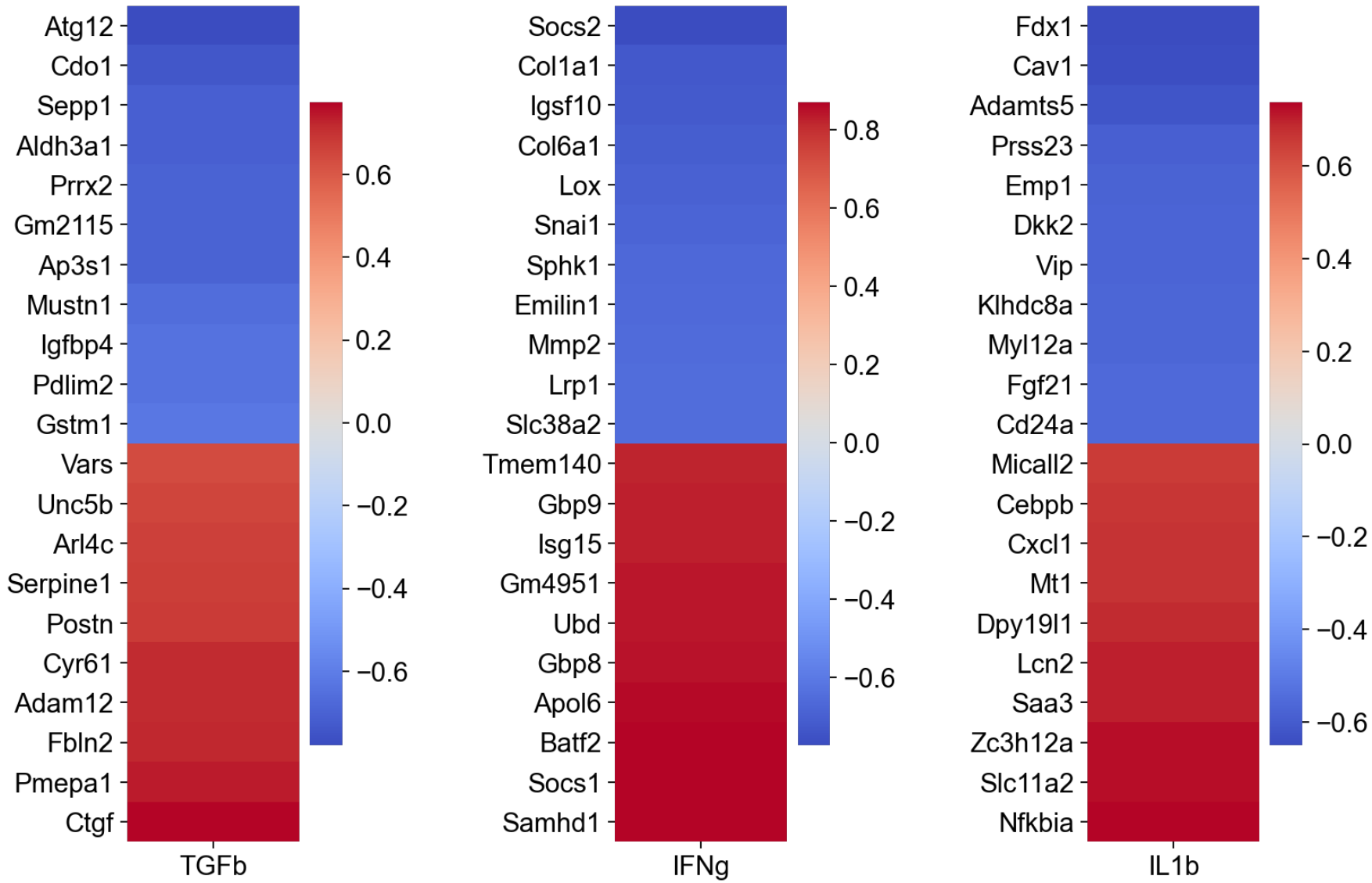
SeqTec 2021 – Multiome experiment

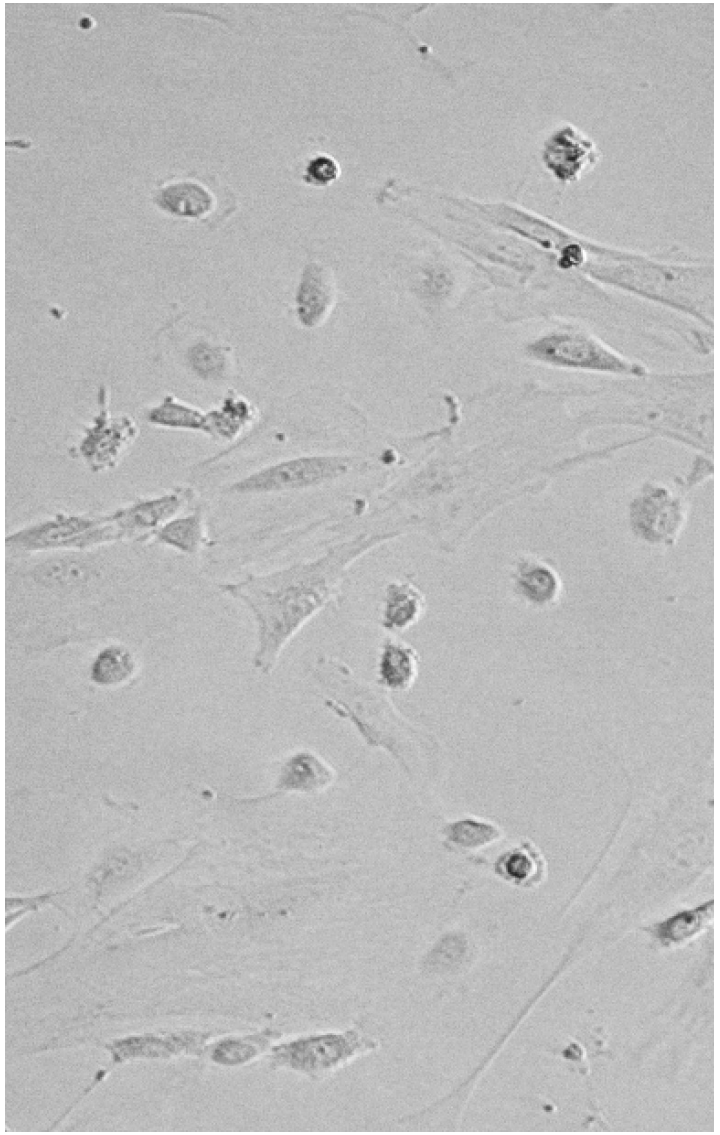
- **Goal:** Use a sample barcoding strategy called ‘SnuBar’ to run a 10X Genomics Multiome (RNA + ATAC) experiment with multiple biological samples
- Cell Line: NIH-3T3 fibroblasts
  - Treatments: IL-1b, TGFb, IFNg, LIF, TNFa
  - Pairwise combinations of several factors (12 samples total)

- Demultiplex samples with CITE-seq-Count and Hashsolo

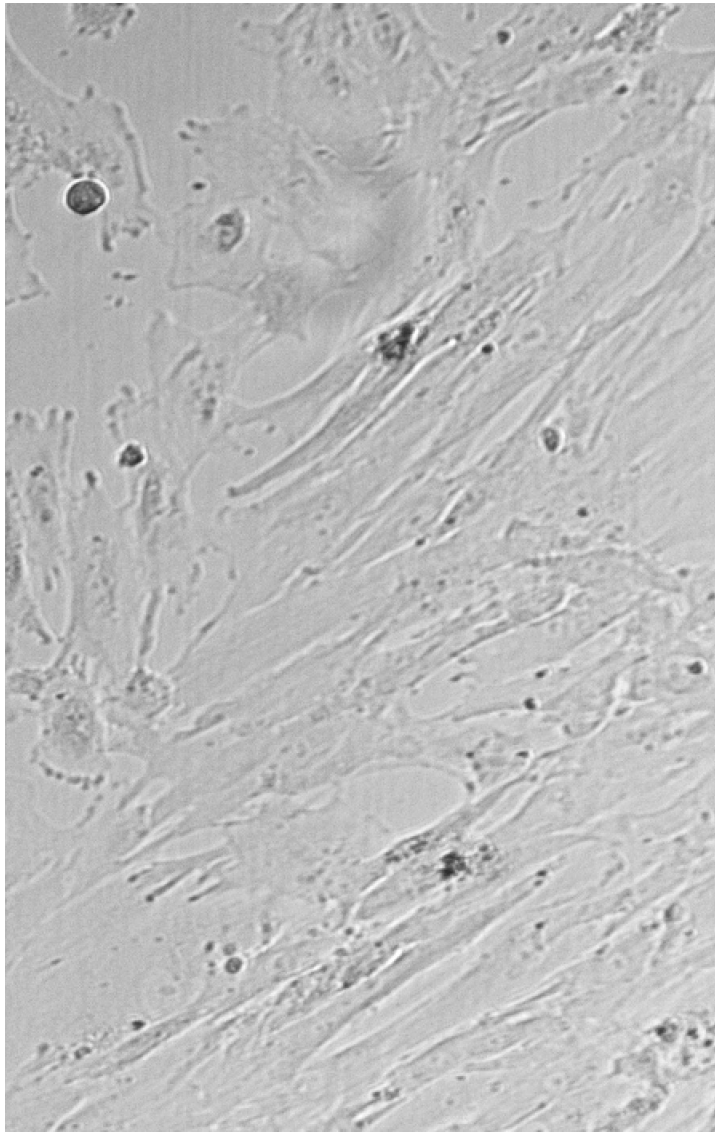

IL1b
???
Inflammatory
Fibroblasts
Resting Fibroblasts
TGFb
???
Myofibroblasts
Antigen Presenting
Fibroblasts
TGFb
???
IFNg
???


Atg12
Cdo1
Sepp1
Aldh3a1
Prrx2
Gm2115
Ap3s1
Mustn1
Igfbp4
Pdlim2
Gstm1
Vars
Unc5b
Arl4c
Serpine1
Postn
Cyr61
Adam12
Fbln2
Pmepa1
Ctgf
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
TGFb
Socs2
Col1a1
Igsf10
Col6a1
Lox
Snai1
Sphk1
Emilin1
Mmp2
Lrp1
Slc38a2
Tmem140
Gbp9
Isg15
Gm4951
Ubd
Gbp8
Apol6
Batf2
Socs1
Samhd1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
IFNg
Fdx1
Cav1
Adamts5
Prss23
Emp1
Dkk2
Vip
Klhdc8a
Myl12a
Fgf21
Cd24a
Micall2
Cebpb
Cxcl1
Mt1
Dpy19l1
Lcn2
Saa3
Zc3h12a
Slc11a2
Nfkbia
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
IL1b

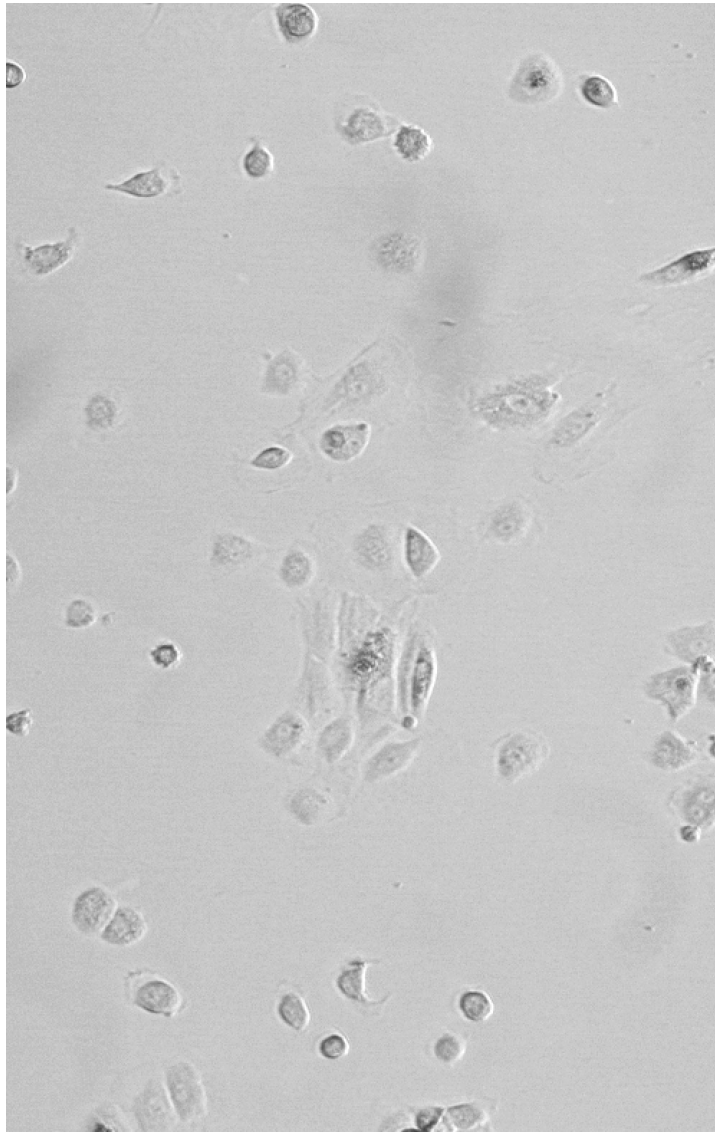
Control

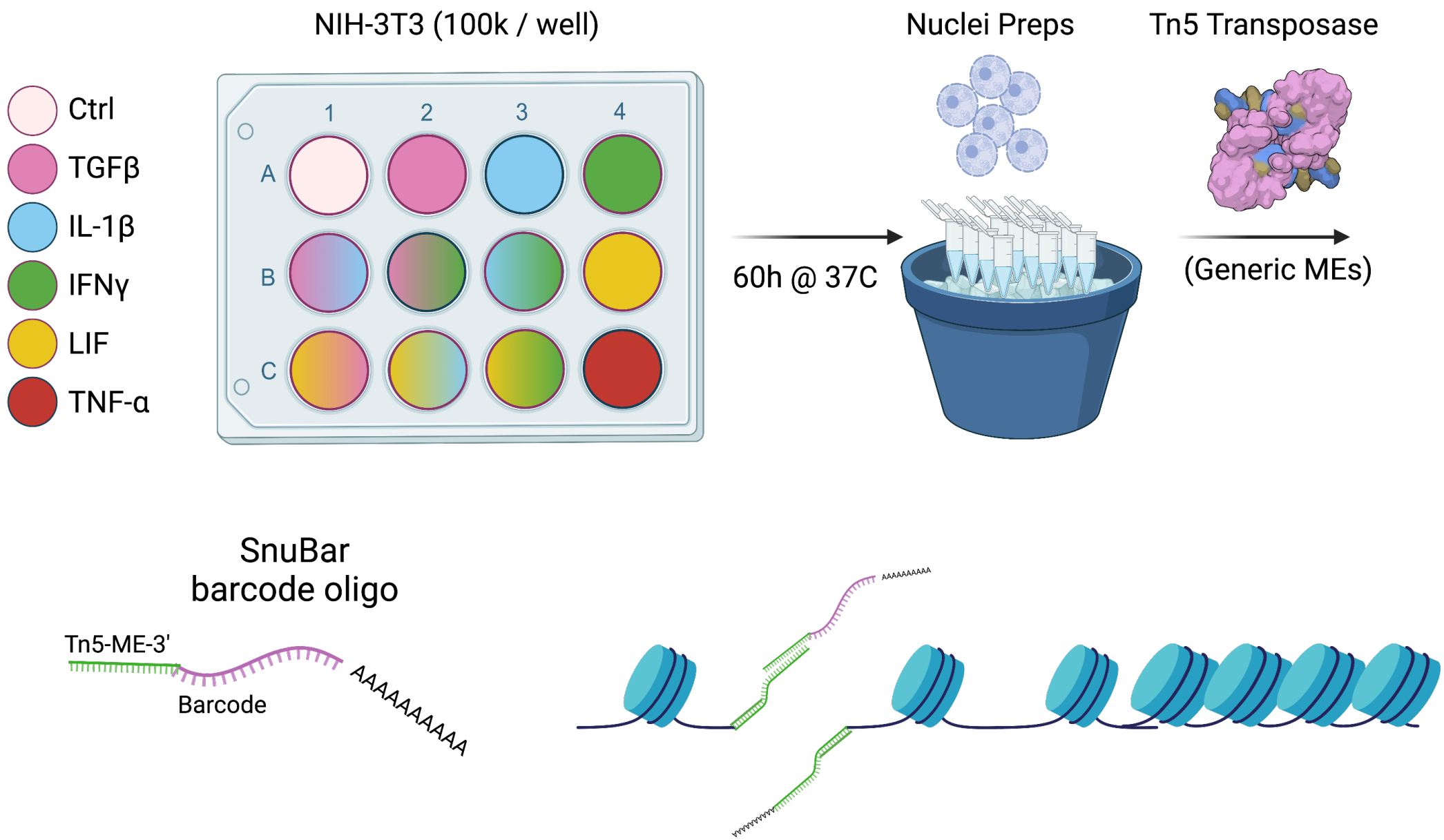


TGFβ

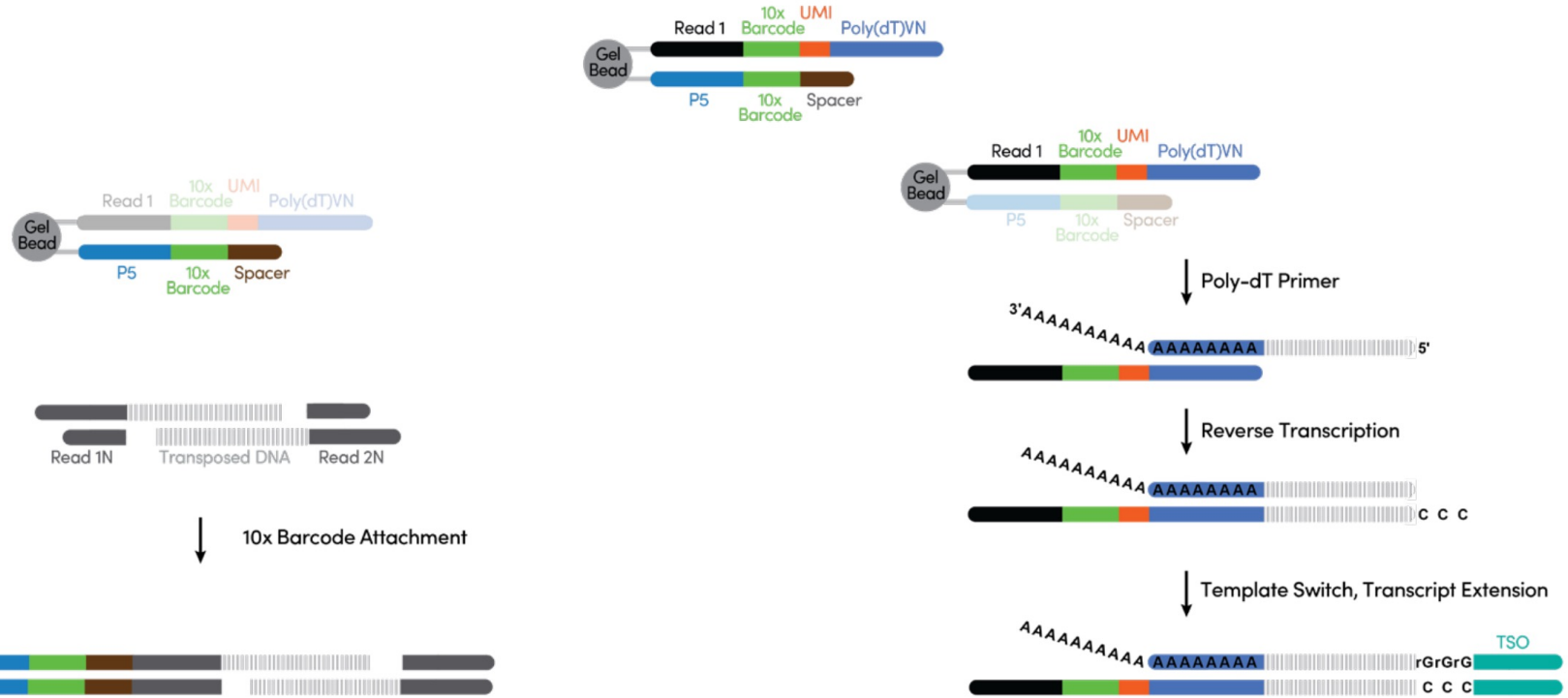


IFNγ




NIH-3T3 (100k / well)
Nuclei Preps
Tn5 Transposase
Ctrl
TGFβ
IL-1β
IFNγ
LIF
TNF-α
1
2
3
4
A
B
C
60h @ 37C
(Generic MEs)
SnuBar
barcode oligo
Tn5-ME-3'
Barcode
AAAAAAAAA
AAAAAAAAA

GEM Generation & Barcoding



Snu-Bar

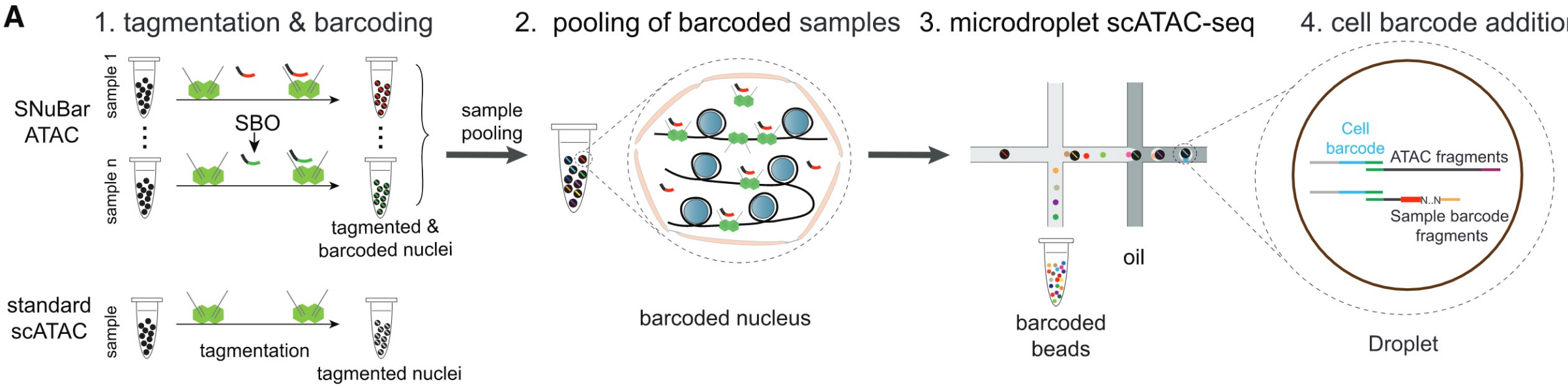
Molecular Cell

Technology

Simple oligonucleotide-based multiplexing of single-cell chromatin accessibility

Authors

Kaile Wang, Zhenna Xiao, Yun Yan, ...,
Bora Lim, Steven H. Lin,
Nicholas E. Navin

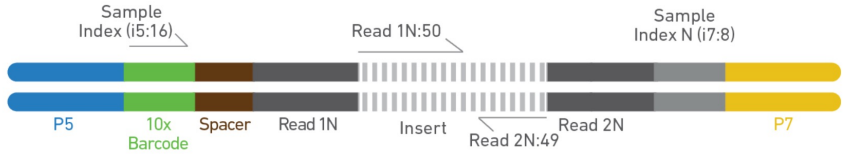


Sequencing Requirements

	Read 1	I7	I5	Read 2
10X. Rec	50	8	16	49
Min.*	30	8	16	30

*NextSeq500 SE75 compatible

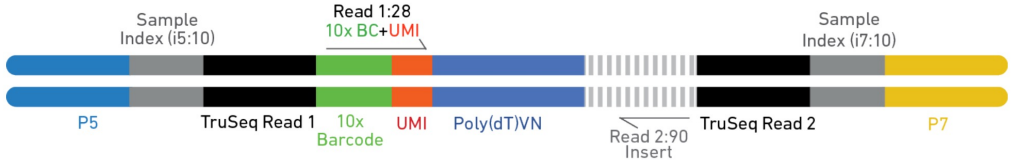
Chromium Single Cell Multiome ATAC Library



	Read 1	I7	I5	Read 2
10X. Rec	28	8	n/a	49
Min. *	28	8	n/a	56

*NextSeq500 SE75 compatible

Chromium Single Cell Multiome Gene Expression Library

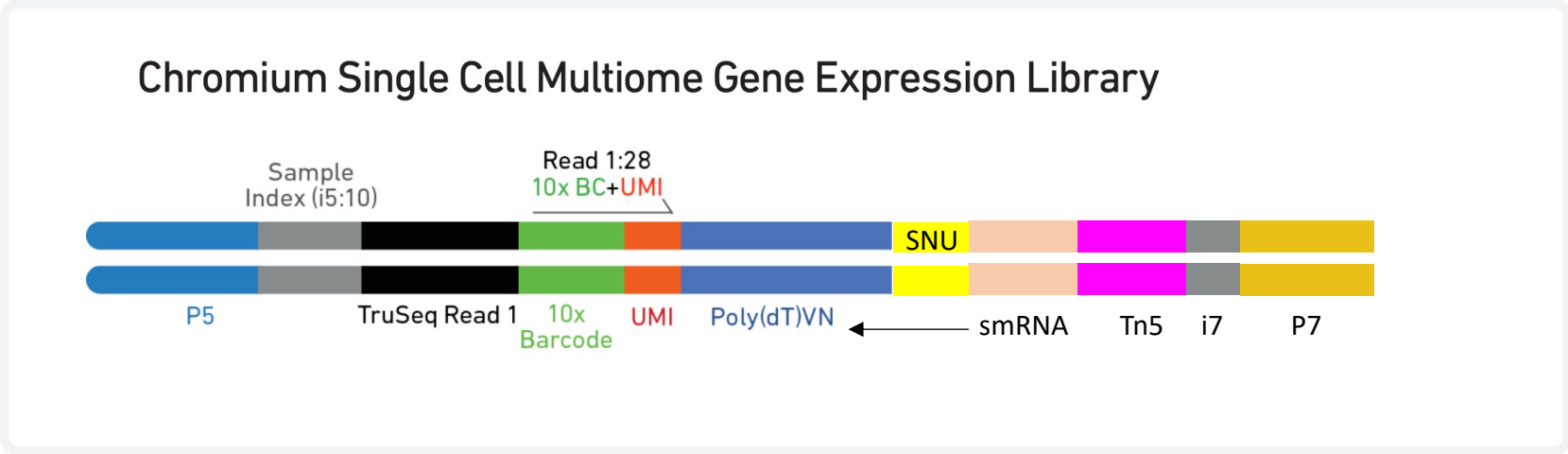


```
5′              AGATGTGTATAAGAGACAG 3′
3′ AAAGAATATACCCTCTACACATATTCTCTGTC-/5phos/


5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG 3′
3′               TCTACACATATTCTCTGTC-/5phos/
```

Nextera Tn5

Tn5 handle	Illumina Small RNA primer	10bp Barcode		
CGAGCCCACGAGAC	CCTTGGCACCCGAGAATTCCA	GCCTCTTTGT	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	Snubar1
CGAGCCCACGAGAC	CCTTGGCACCCGAGAATTCCA	CAGGGATGGT	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	Snubar2
CGAGCCCACGAGAC	CCTTGGCACCCGAGAATTCCA	CTACTGCCGA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	Snubar3



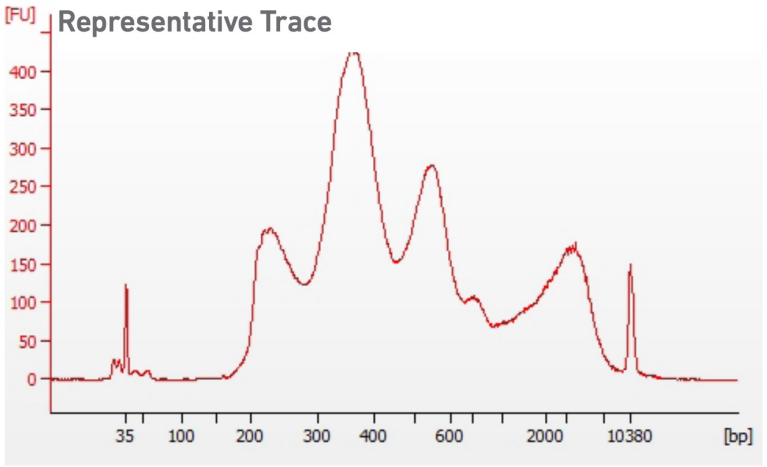
Multiome Library QC

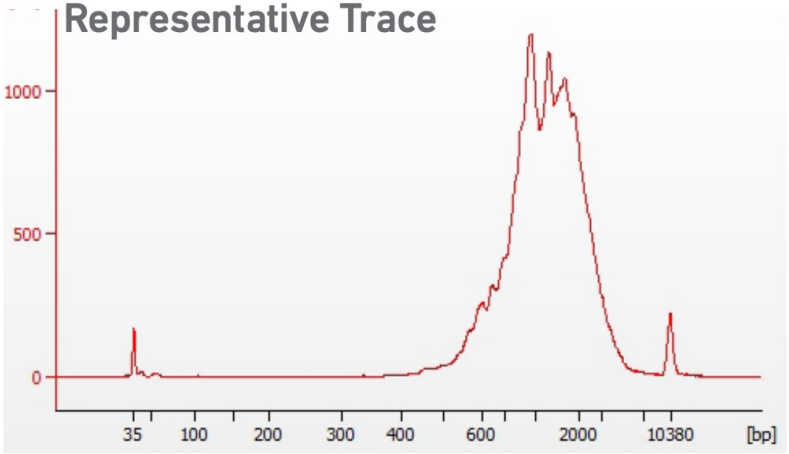
Cycle Number Optimization Table

Targeted Nuclei Recovery	Total Cycles
≤2,000	9
2,001-6,000	8
6,001-10,000	7

Example: Library Construction Input Mass & SI PCR Cycles

Cell Type	Targeted Nuclei Recovery	Total cDNA Yield (ng)	cDNA Input into Fragmentation		SI PCR Cycle Number
			Volume (µl)	Mass (ng)	
High RNA Content	Low	150 ng	10 µl	37.5 ng	14
	High	400 ng	10 µl	100 ng	13
Low RNA Content	Low	1 ng	10 µl	0.25 ng	16
	High	100 ng	10 µl	25 ng	14





Preall_CSHLCourse2021_8k_RNA

Alerts

The analysis detected ⚠ 3 warnings.

	Alert	Value	Detail
⚠	Low Fraction Reads Confidently Mapped To Transcriptome	22.6%	Ideal > 30%. This can indicate use of the wrong reference transcriptome, a reference transcriptome with overlapping genes, poor library quality, poor sequencing quality, or reads shorter than the recommended minimum. Application performance may be affected.
⚠	High Fraction of Reads Mapped Antisense to Genes	25.2%	Ideal < 10%. This can indicate use of an unsupported chemistry type (e.g. using Single Cell V(D)J for gene counting). Application performance is likely to be affected.
⚠	Low Fraction Reads in Cells	63.0%	Ideal > 70%. Application performance may be affected. Many of the reads were not assigned to cell-associated barcodes. This could be caused by high levels of ambient RNA or by a significant population of cells with a low RNA content, which the algorithm did not call as cells. The latter case can be addressed by inspecting the data to determine the appropriate cell count and using --force-cells.

Summary | Analysis

20,377
Estimated Number of Cells

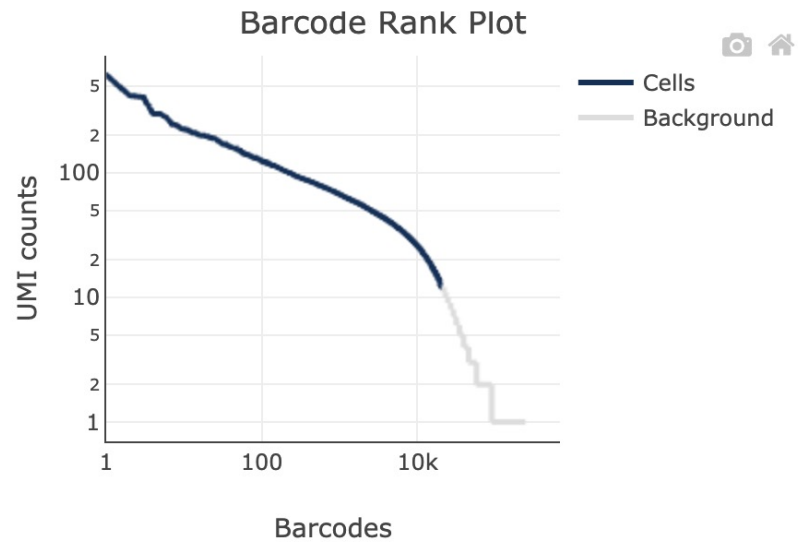
6,400
Mean Reads per Cell

24
Median Genes per Cell

Sequencing

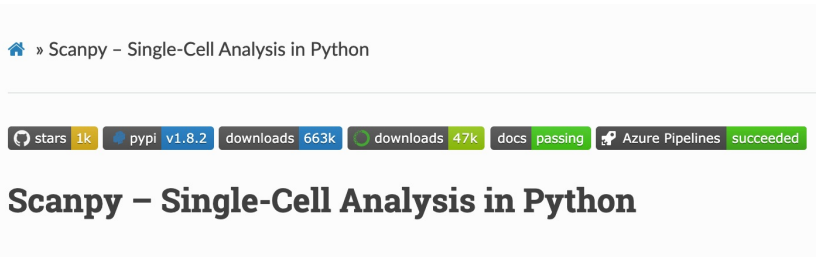
Number of Reads	130,418,393
Number of Short Reads Skipped	0
Valid Barcodes	87.9%
Valid UMIs	100.0%
Sequencing Saturation	96.2%
Q30 Bases in Barcode	97.2%
Q30 Bases in RNA Read	91.5%
Q30 Bases in UMI	96.2%

Cells



Estimated Number of Cells	20,377
Fraction Reads in Cells	63.0%
Mean Reads per Cell	6,400
Median Genes per Cell	24
Total Genes Detected	18,417
Median UMI Counts per Cell	26

Count Snubars and import into Scanpy




```
hashdir=/Volumes/mccombie/Illumina_runs/NextSeqData/NextSeqOutput/211117_
NB551387_0662_AHLK7WBGXK/NGS-CR-2140/

SAMPLE=Preall_CSHLCourse2021_16k_snubar

R1_FASTQ=$(ls $hashdir/$SAMPLE*_R1_*)
R2_FASTQ=$(ls $hashdir/$SAMPLE*_R2_*)

CITE-seq-Count \
-R1 $R1_FASTQ \
-R2 $R2_FASTQ \
-t hash_tags.csv  \
-cbf 1 \
-cbl 16 \
-umif 17 \
-umil 28 \
-o "$SAMPLE"_Result.tsv \
-cells 16000
```

features.tsv.gz
barcodes.tsv.gz
matrix.mtx.gz

SnuBar Barcode	Snubar1_control-GCCTCTTTGT	Snubar2_TGFB-GCCAGGATAC	Snubar3_TGFB+IL-1B-CAGGGATGGT	Snubar4_IL-1B-CTACTGCCGA	Snubar5_IFN-TAGTTCGGAG	Snubar6_IFN+TGFB-AGCATAAGTA	Snubar7_IFN+IL-1B-AAGGTAAACT
GCAATGTTCCTCAGTC-1	23	20	11	45	18	12	13
ATTTGCAAGCTTTGGG-1	41	36	9	68	20	15	12
CACCAACCAGGCTAGA-1	17	15	8	50	15	13	4
AATTGACGTAATCACG-1	22	18	13	80	18	10	9
ACCGGCTAGTAACCAC-1	28	24	7	55	23	7	10
...	...	...	...	...	...	...	...
GTTGGCGGTGAGACTC-1	48	32	12	73	31	29	19
AGAAAGGCAATGCGCT-1	27	18	12	47	17	12	5
GGCTTAAGTACCGGAT-1	46	30	14	77	29	18	15
GTTACGCCATGCTCCC-1	26	11	7	38	22	11	13
GTGCACGGTTGTCCAT-1	26	26	2	43	13	10	11

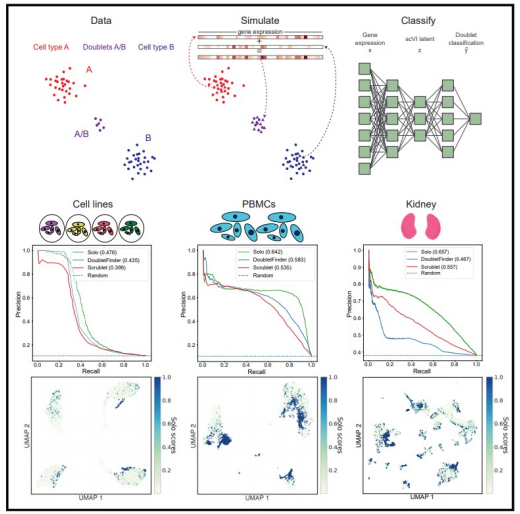
16000 rows × 13 columns

Read in Snubar counts into scanpy -> create Anndata object

Demultiplex with Hashsolo

Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning

Graphical Abstract



Authors

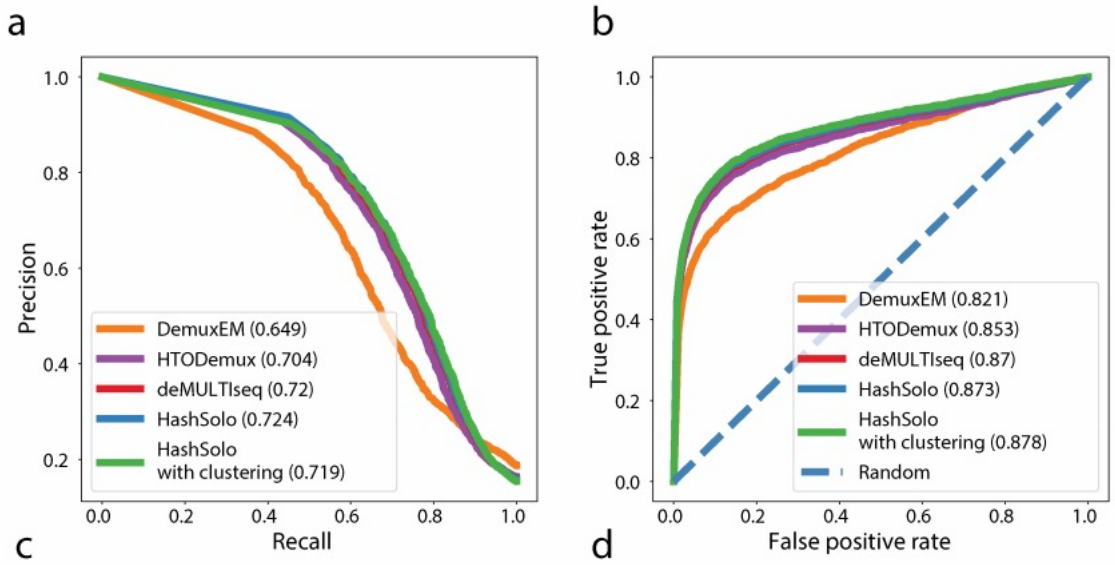
Nicholas J. Bernstein, Nicole L. Fong, Irene Lam, Margaret A. Roy, David G. Hendrickson, David R. Kelley

Correspondence

dgh@calicolabs.com (D.G.H.), drk@calicolabs.com (D.R.K.)

In Brief

Current single-cell RNA sequencing technologies occasionally allow multiple cells to be combined into a single profile, which challenges downstream analyses. Bernstein et al. introduce a semi-supervised deep learning method called Solo that identifies these "doublet" cells with greater accuracy than existing methods.




```
>>> from solo import hashsolo
>>> import anndata
>>> cell_hashing_data = anndata.read("cell_hashing_counts.h5ad")
>>> hashsolo.hashsolo(cell_hashing_data)
>>> cell_hashing_data.obs.head()
                  most_likeli_hypothesis  cluster_feature  negative_hypothesis_probability  singlet_hypothesis_probability  doublet_hypothesis_probability
Classification
index
CCTTTCTGTCCGAACC                       2                0                     1.203673e-16                        0.000002                        0.999998
Doublet
CTGATAGGTGACTCAT                       1                0                     1.370633e-09                        0.999920                        0.000080
BatchF-GTGTGACGTATT_x
AGCTCTCGTTGTCTTT                       1                0                     2.369380e-13                        0.996992                        0.003008
BatchE-GAGGCTGAGCTA_x
GTGCGGTAGCGATGAC                       1                0                     1.579405e-09                        0.999879                        0.000121
BatchB-ACATGTTACCGT_x
AAATGCCTCTAACCGA                       1                0                     1.867626e-13                        0.999707                        0.000293
BatchB-ACATGTTACCGT_x
>>> demultiplex.plot_qc_checks_cell_hashing(cell_hashing_data)
```