

Single Cell Sequencing



CSHL Course: Advanced Sequencing Technologies & Applications

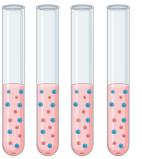
Jon Preall
Research Associate Professor
Cold Spring Harbor Laboratory

Single Cell Biology Core at CSHL

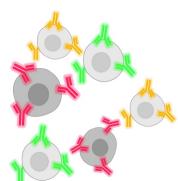
Study Design



Tissue Dissociation



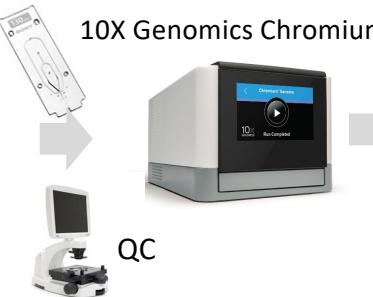
Flow panel optimization



Sony SH800 FACSorter



10X Genomics Chromium



Preali, Jonathan
Resch. Assoc. Profess. & Head
of Genomics Tech Development

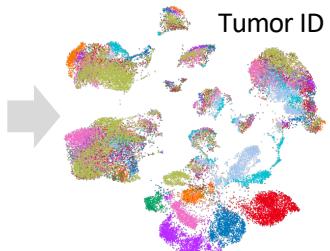


Regan, Claire
Computational Science
Developer II

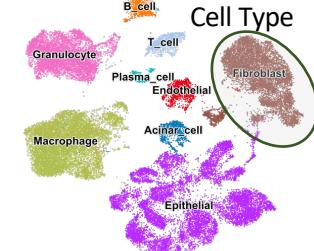


Zhang, Evan
Research Technician I

Batch Correction Data Integration

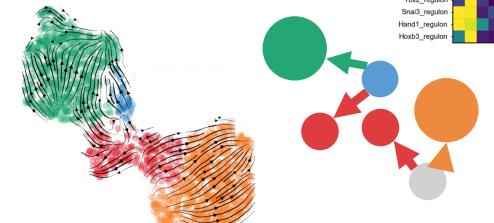


Clustering Cell Type Identification

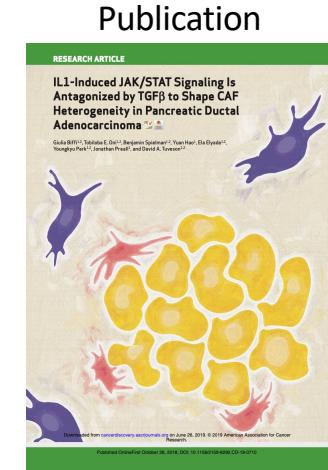


Focused Analysis

RNA Velocity Trajectory inference Transcription Factor Modules



Lineage Analysis





www.singlecellbio.org

10X Genomics Xenium



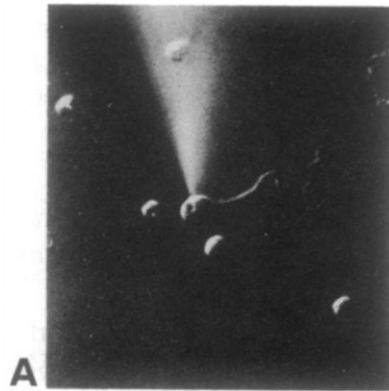
Visium CytAssist

Analysis of gene expression in single live neurons

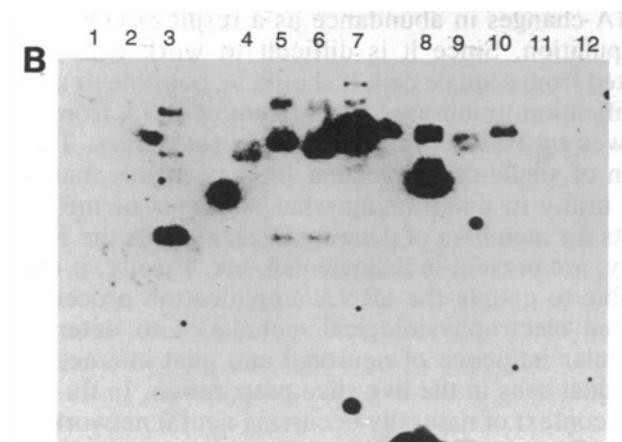
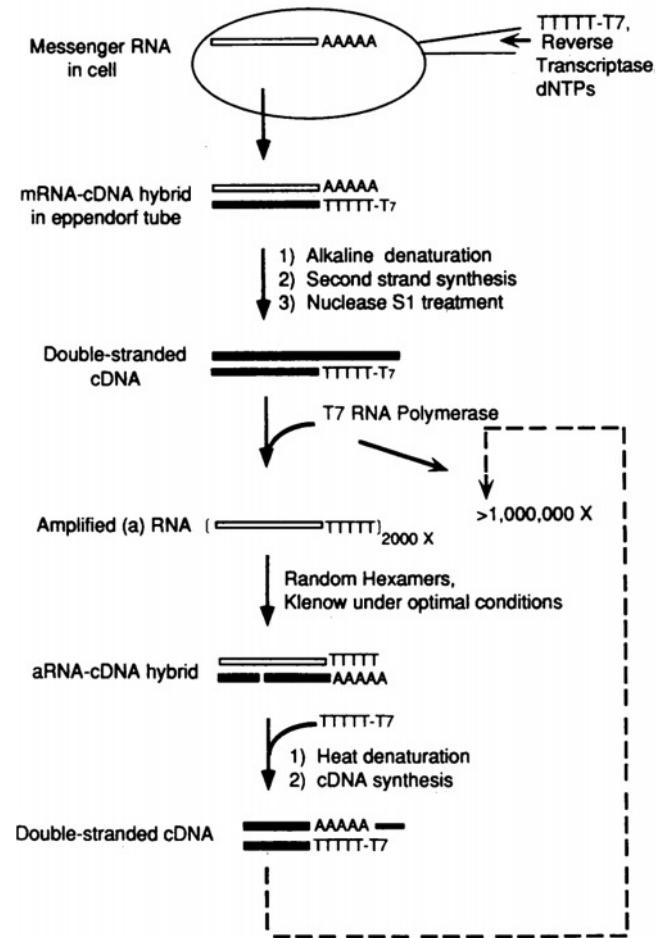
(amplified, antisense RNA/expression profile/mRNA complexity/pyramidal cell)

JAMES EBERWINE*†‡, HERMES YEH§, KEVIN MIYASHIRO*, YANXIANG CAO*, SURESH NAIR*,
RICHARD FINNELL*¶, MARTHA ZETTEL§, AND PAUL COLEMAN§

Departments of *Pharmacology and †Psychiatry, University of Pennsylvania Medical School, Philadelphia, PA 19104; and Department of §Neurobiology and Anatomy, University of Rochester Medical Center, Rochester, NY 14642



Microinjection of cDNA synthesis reagents directly into single neurons



Southern Blot
Plasmid standards
containing gene of interest
Probed with aRNA

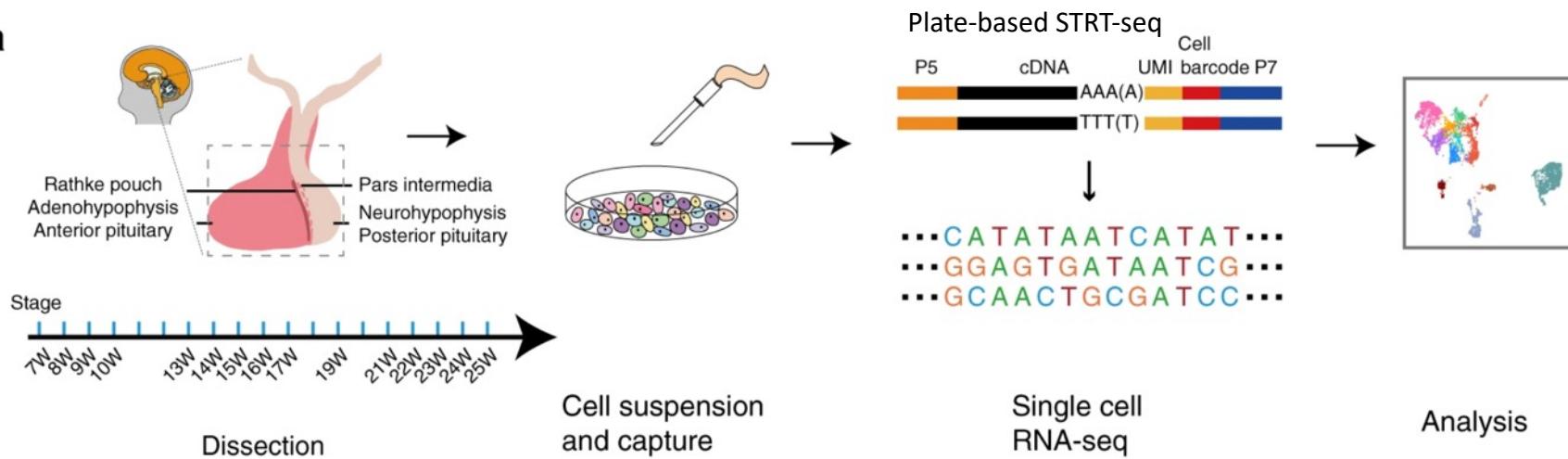
1992

Single-cell transcriptomics identifies divergent developmental lineage trajectories during human pituitary development

Shu Zhang, Yueli Cui, Xinyi Ma, Jun Yong, Liying Yan, Ming Yang, Jie Ren, Fuchou Tang, Lu Wen✉ & Jie Qiao✉

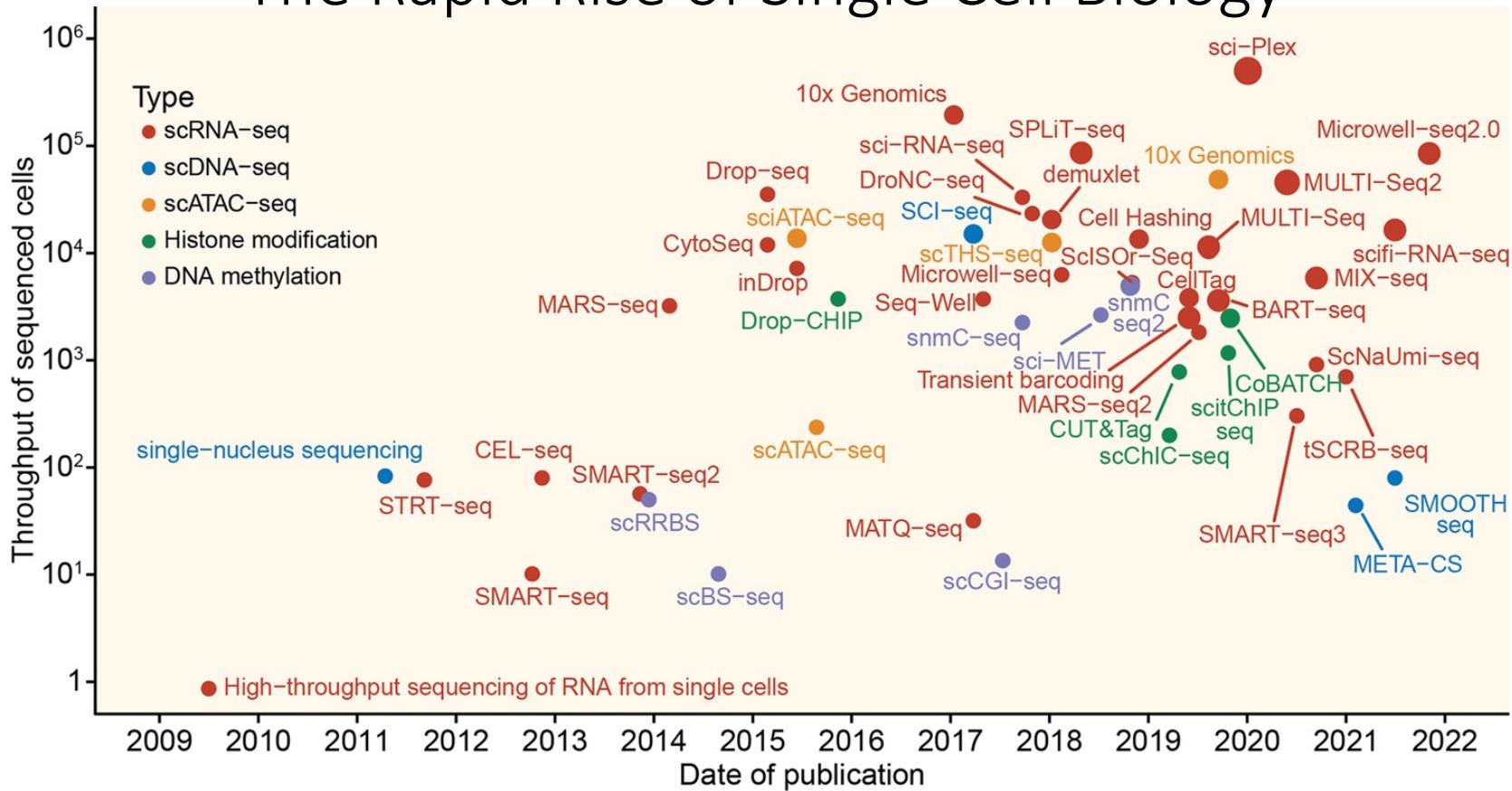
Nature Communications 11, Article number: 5275 (2020) | Cite this article

a



4,113 mouth-pipetted cells!

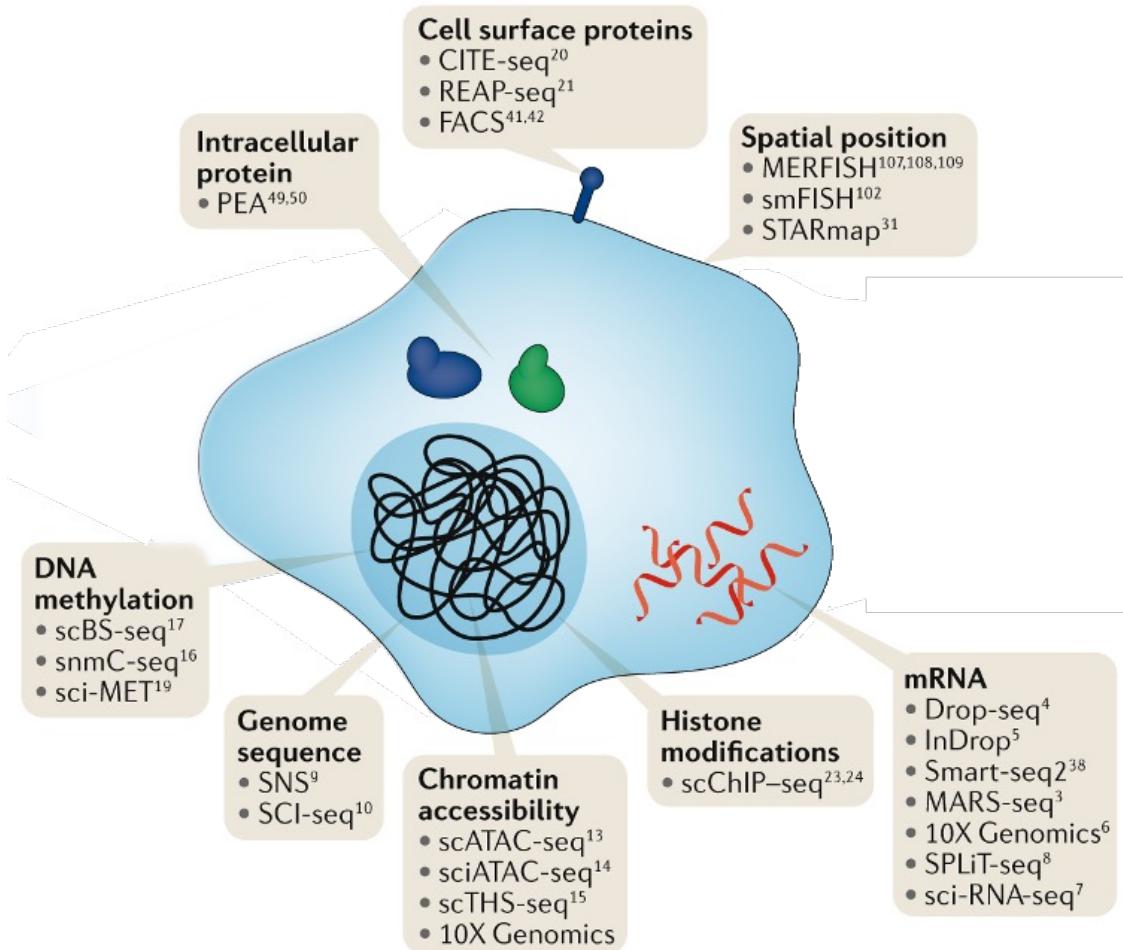
The Rapid Rise of Single Cell Biology



Jia, Q., Chu, H., Jin, Z. et al. High-throughput single-cell sequencing in cancer research. *Sig Transduct Target Ther* 7, 145 (2022)

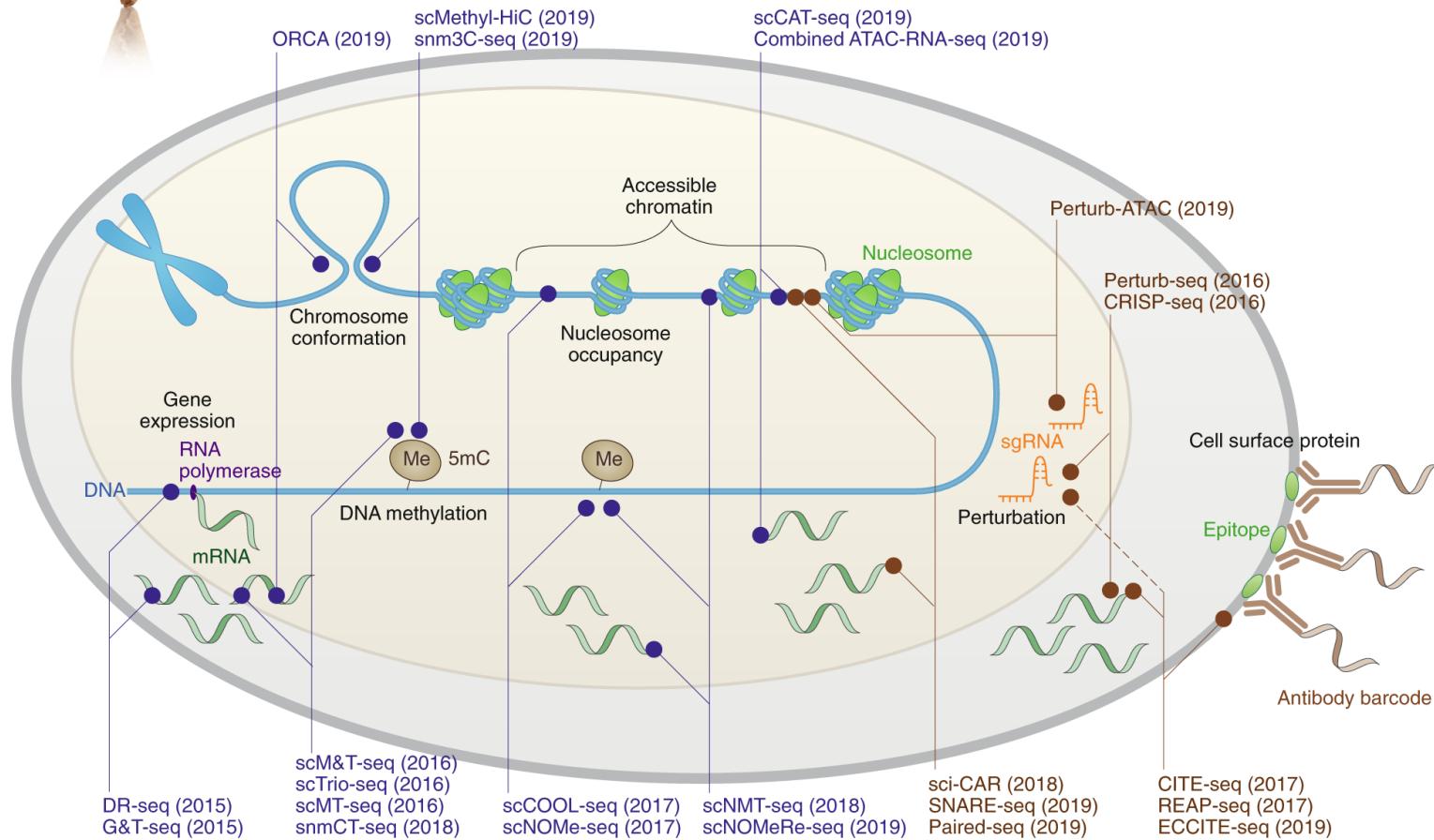


Many Flavors of Single cell 'Omics

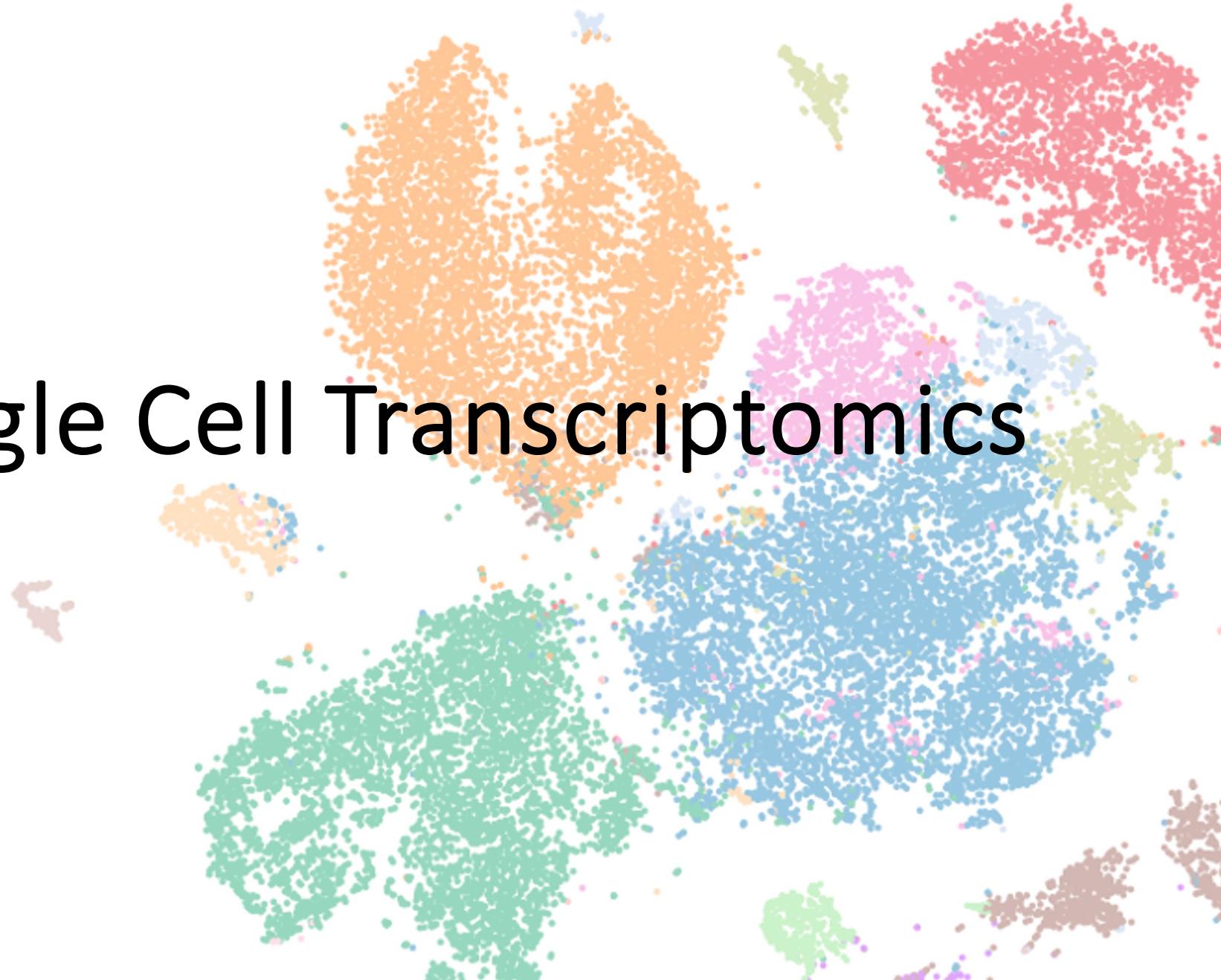




Many Flavors of Multi- 'Omics



Single Cell Transcriptomics



scRNAseq Platforms

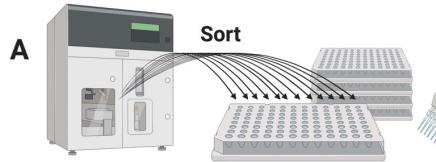
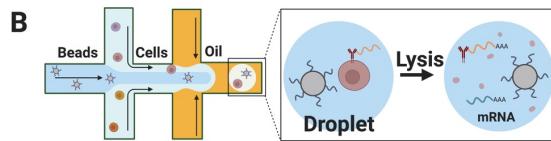


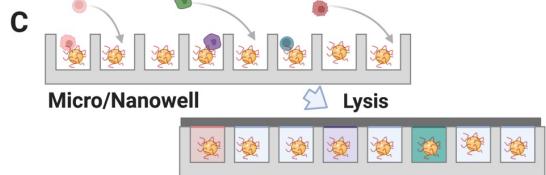
Plate-based / Sorted

- Smart-Seq
- CEL-seq
- Quartz-Seq



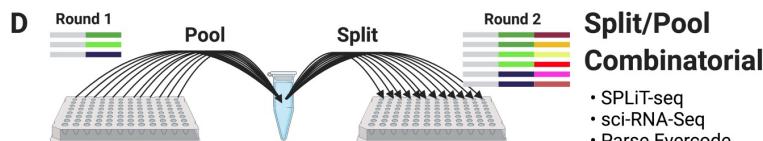
Droplet Microfluidics

- Drop-Seq
- 10X Genomics
- Bio-Rad ddSeq



Micro / Nanowell

- Microwell-Seq
- BD Rhapsody
- Singleron Matrix



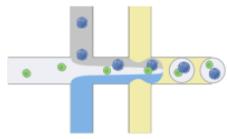
Split/Pool Combinatorial

- SPLIT-seq
- sci-RNA-Seq
- Parse Evercode

Throughput (cost/labor per cell)	Flexibility	Sensitivity / Max Depth	Protocol Simplicity / Accessibility	Adoption / Available public datasets
+	+++	+++	++	++
++	+	++	+++	+++
++	++	++	+	+
+++	++	++	++	++

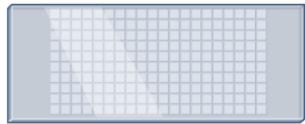
Comparative Analysis of Commercial Single-Cell RNA Sequencing Technologies

Single Cell RNA Sequencing



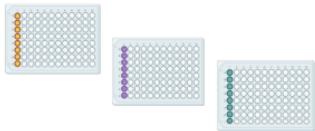
Emulsion:

- Chromium Single Cell 3' Reagent Kit, v3.1 Chemistry (10X Genomics)
- Chromium Single Cell 5' Reagent Kit, v2 Chemistry (10X Genomics)
- Chromium Fixed RNA Profiling (FRP) Reagent Kit (10X Genomics)
- PIPseq T20 3' Single Cell RNA Kit v4.0 (Fluent Biosciences)



Microwell:

- BD Rhapsody WTA Reagent Kit (Becton Dickinson)
- HIVE CLX scRNA Seq Kit (Honeycomb Biotechnologies)
- GEXSCOPE® Single Cell RNA Library Kit (Singleron Biotechnologies)



In plate combinatorial indexing:

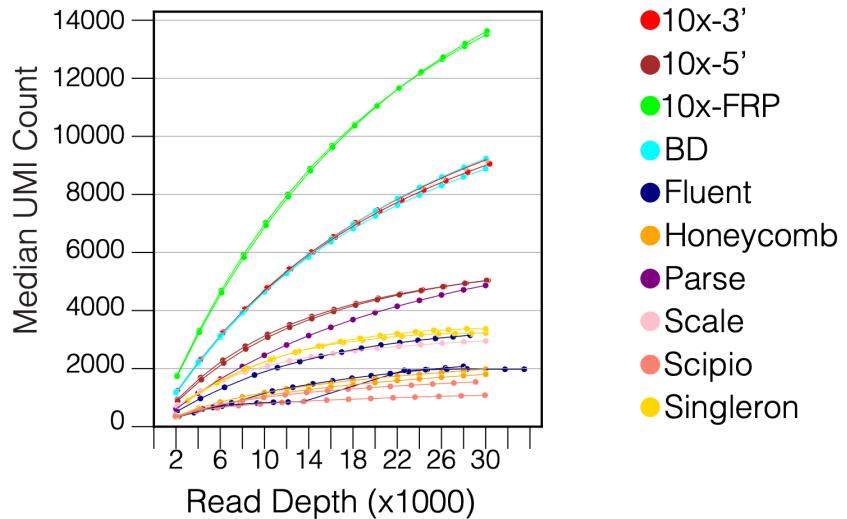
- Evercode WT v2 (Parse Biosciences)
- Single Cell RNA Kit (Scale Biosciences)



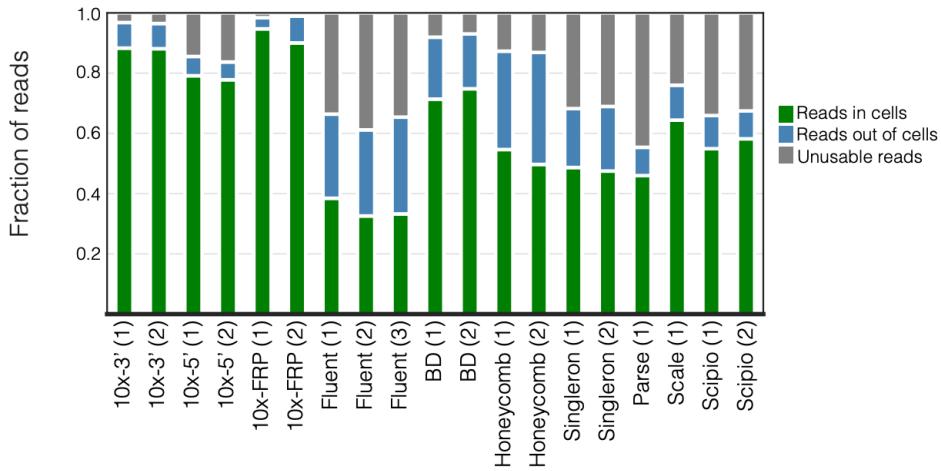
Matrikel:

- ASTERIA Single-cell RNASeq Kit (Scipio Bioscience)

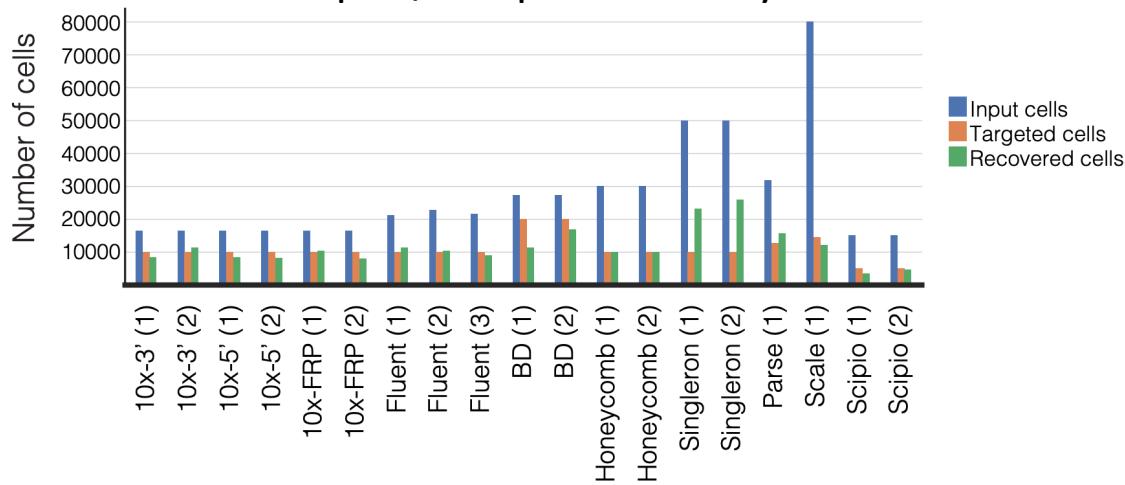
RNA sensitivity



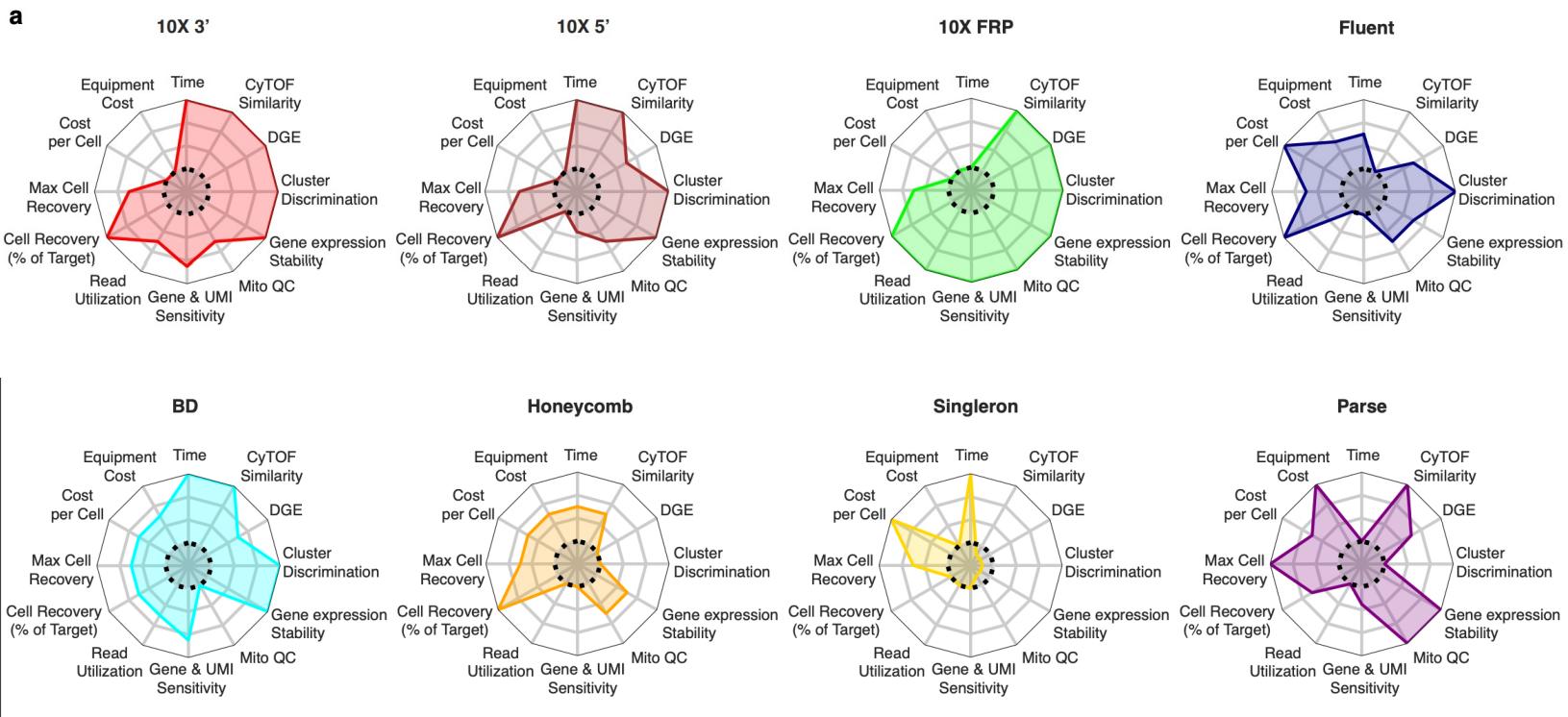
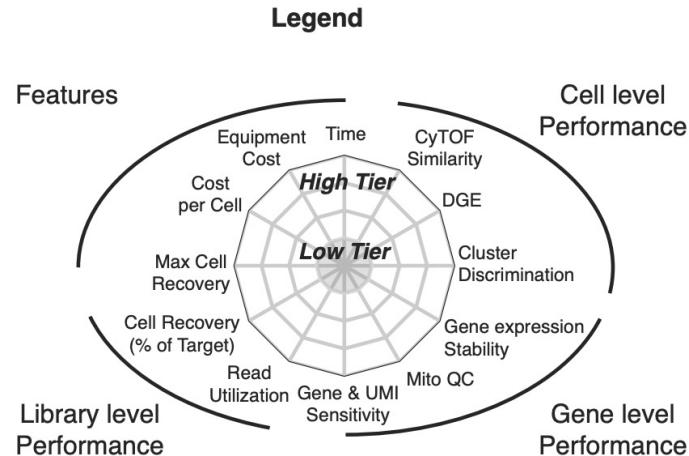
Sequencing efficiency



Cell Input / Output Efficiency



Multi-Dimensional Performance Summary



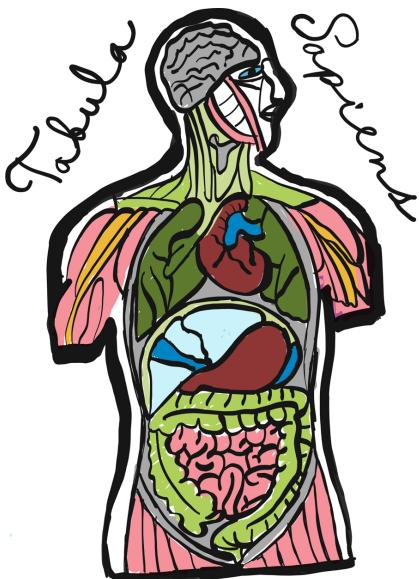
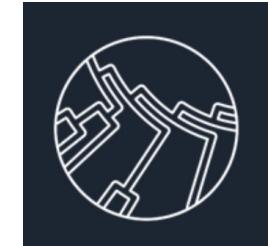
10X Genomics: the *lingua franca* of the single-cell age



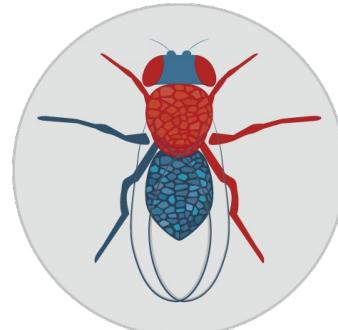
- Easy
- Robust
- Expensive.



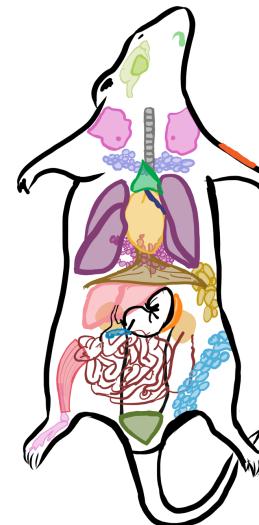
Allen Brain Map



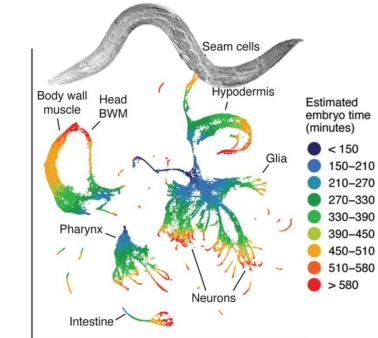
Fly Cell Atlas



Tabula Muris



C elegans



Packer et al (2019) Science

Applications

- RNA-seq
- V(D)J profiling
- ATAC-seq
- RNA + ATAC
- Surface Proteins
- CRISPR guides
- "FLEX" Fixed Cell RNA profiling



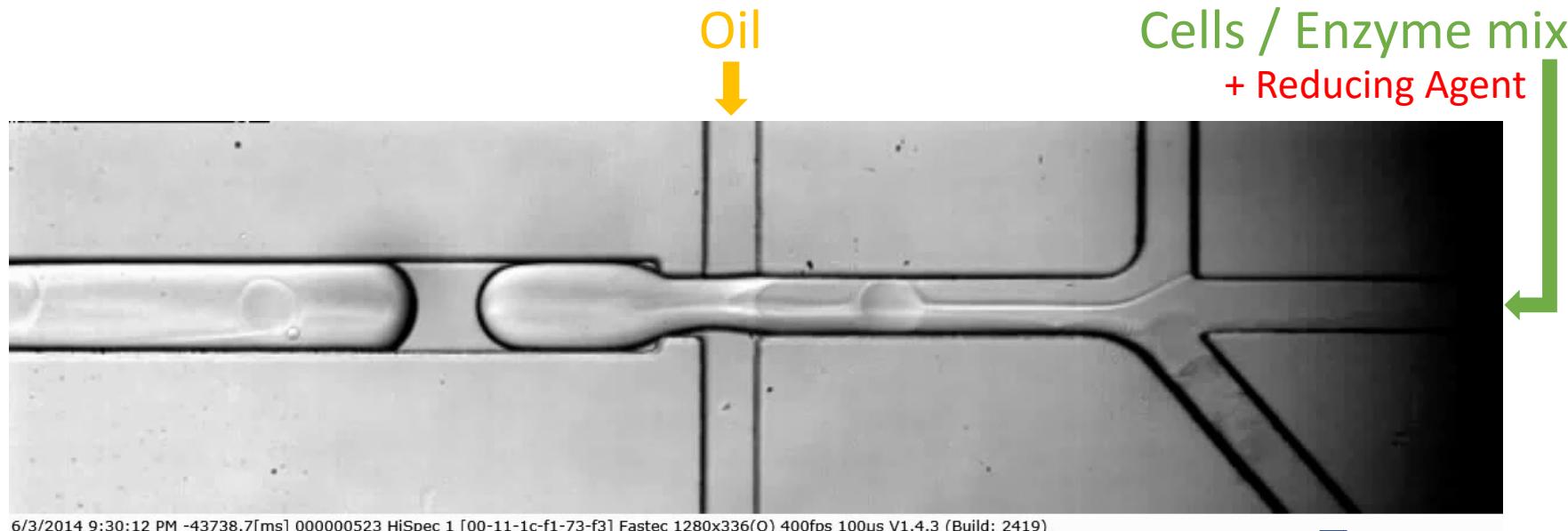
Chromium Xo

Chromium iX

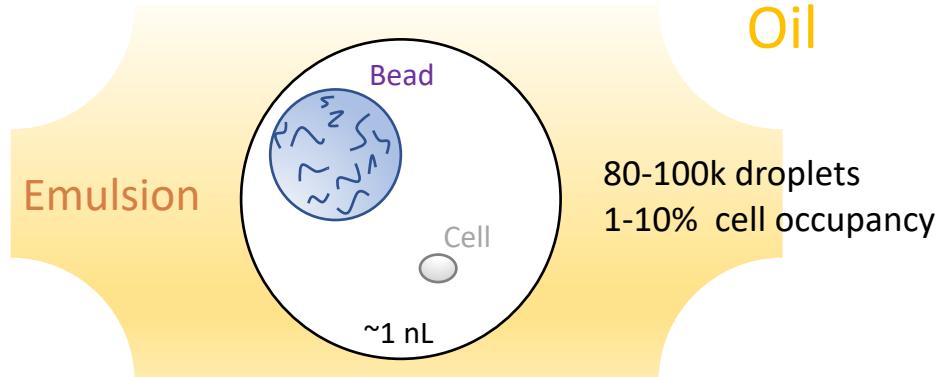
Chromium X

Description	Enables 3' single cell gene expression only	Core assays and capabilities	Full portfolio of Chromium assays and capabilities
3' Gene Expression	✓	✓	✓
5' Gene Expression	✓	✓	✓
FLEX fixed RNA		✓	✓
ATAC		✓	✓
Multiome ATAC+RNA		✓	✓
<i>On-Chip Multiplexing</i>			
3' Gene Expression			✓
5' Gene Expression			✓

Under the Hood



6/3/2014 9:30:12 PM -43738.7[ms] 000000523 HiSpec 1 [00-11-1c-f1-73-f3] Fastec 1280x336(Q) 400fps 100µs V1.4.3 (Build: 2419)

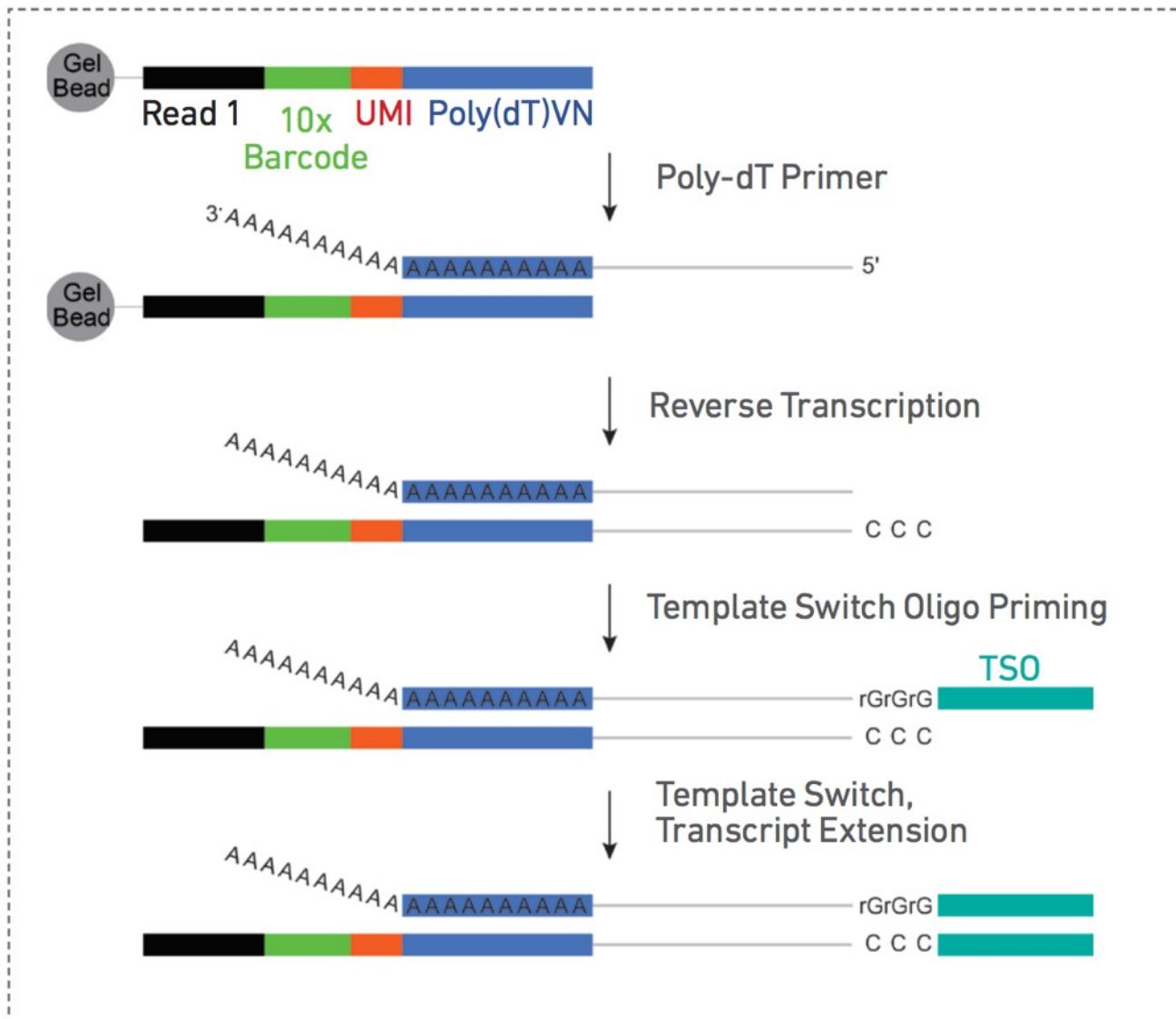


- -S-S- crosslinked hydrogel
- 10^9 primer molecules per bead
- 1 barcode sequence per bead
- Pool of 3M total possible barcodes

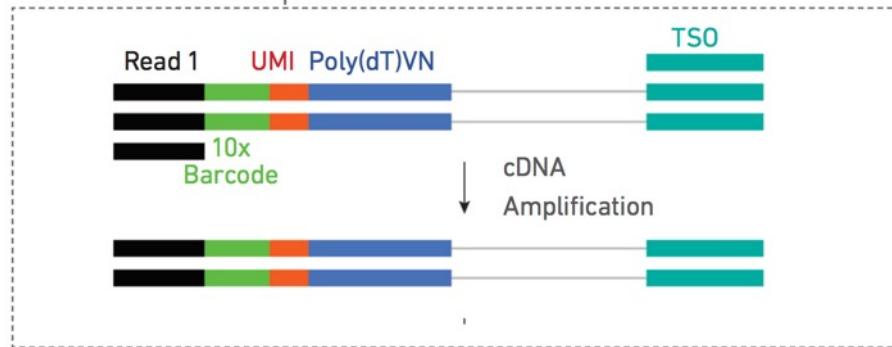


Debris Filters

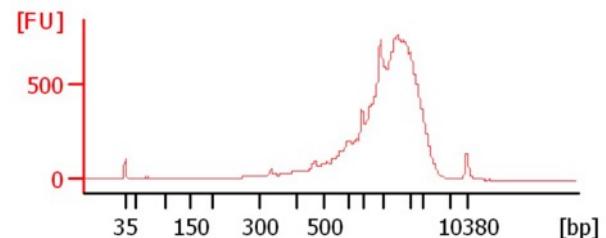
Single Cell 3' Chemistry Overview



Pooled cDNA amplification

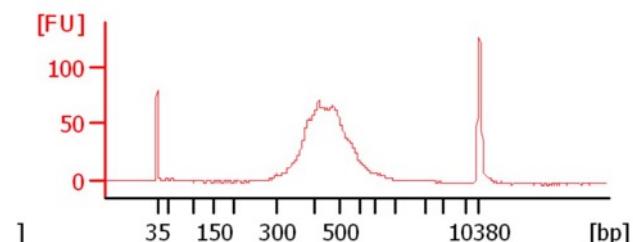
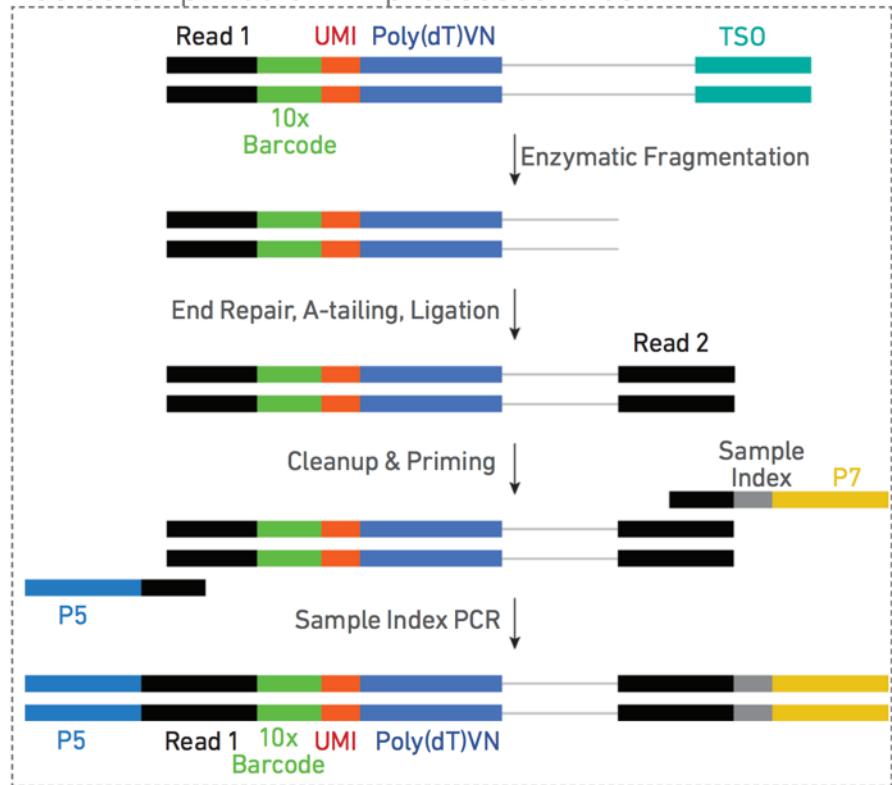


Bioanalyzer



Amplified cDNA

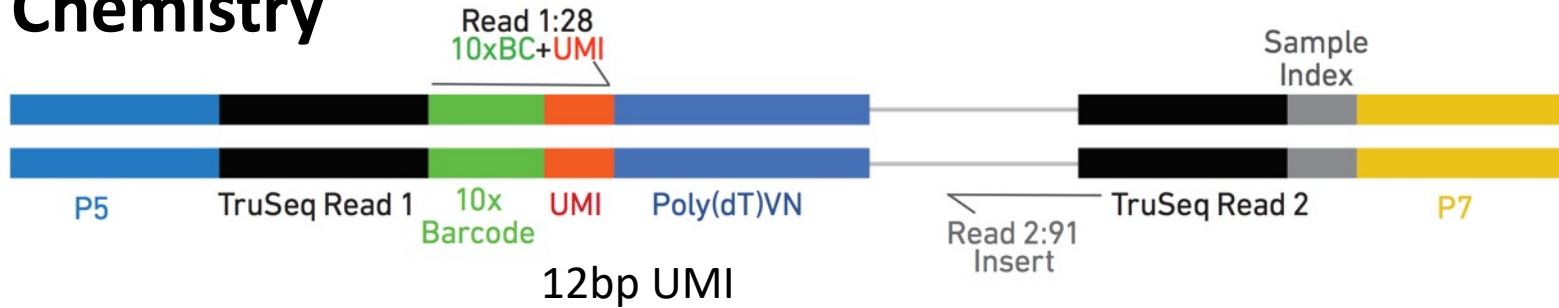
Pooled amplified cDNA processed in bulk



Final Library

Anatomy of a 10X 3'-Single Cell Amplicon

V3 Chemistry



Unique Molecular Identifier (UMI)

Random ~8-10bp sequence incorporated during oligo synthesis

Cell barcode UMI

CCCCCCCCXXXXXXTTTTTTTVN
AAAAAAAAAAABN---IFNgamma---

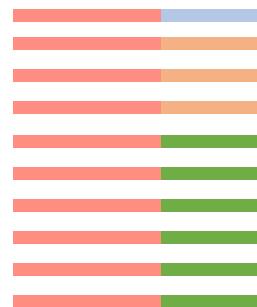
CCCCCCCCXXXXXXTTTTTTTVN
AAAAAAAAAAABN---IFNgamma---

CCCCCCCCXXXXXXTTTTTTTVN
AAAAAAAAAAABN---IFNgamma---

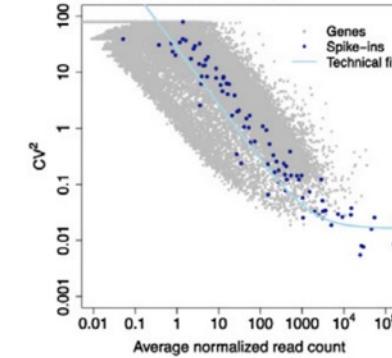
Before PCR



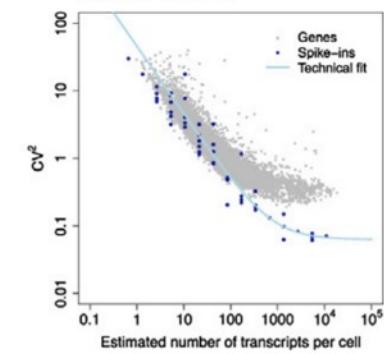
After PCR



without UMIs

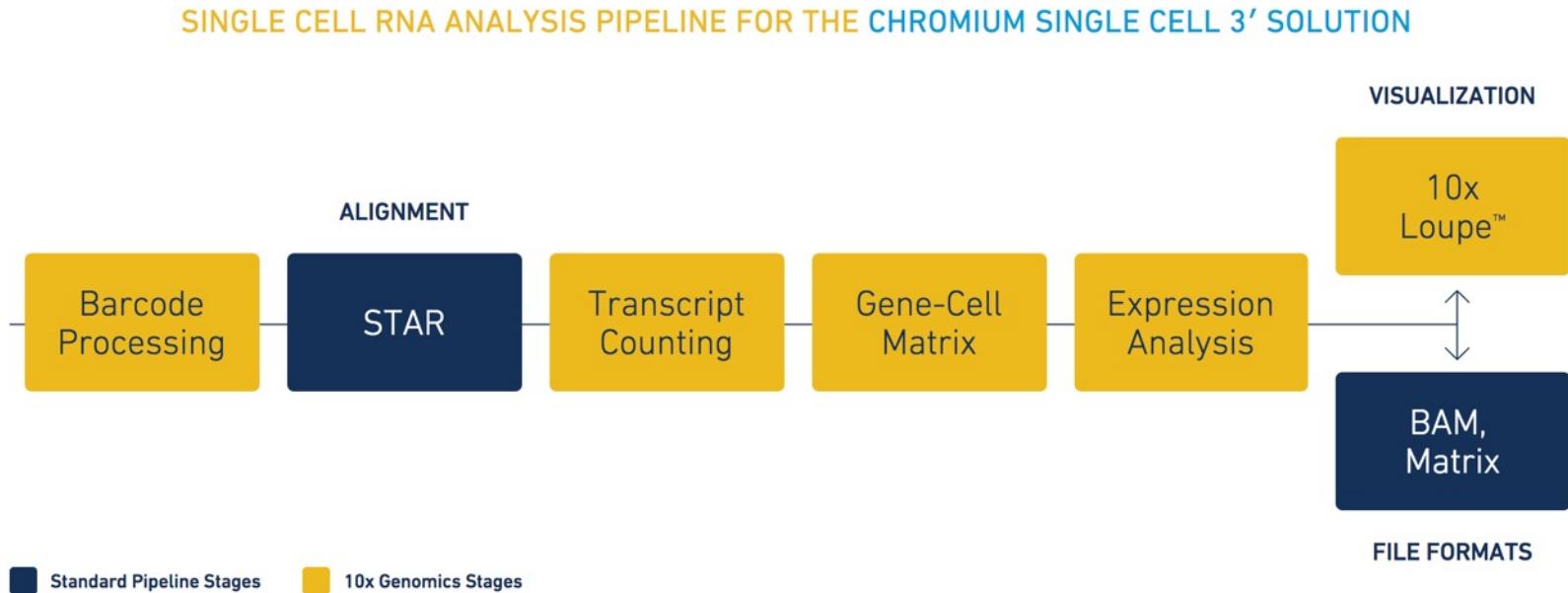


with UMIs



Mapping and Transcript Quantification

Cellranger Count pipeline: [10X Genomics support page](#)



Digital Gene Expression, Not Coverage

“Deep” Single Cell Libraries

Well-based, eg. SmartSeq
Fluidigm C1

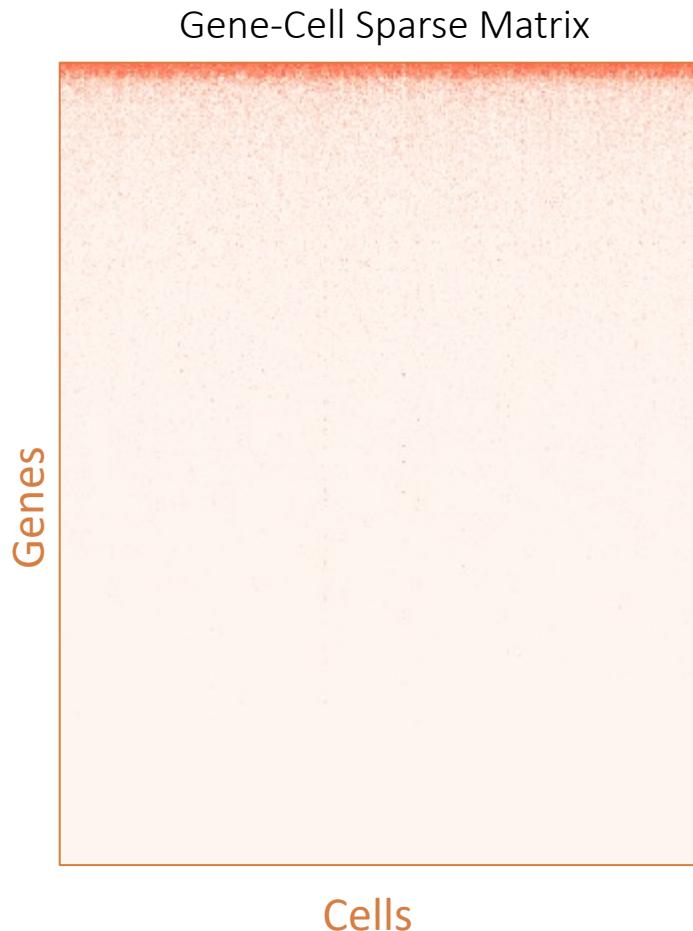


Droplet – Based DGE libraries

Drop-Seq
10X Genomics
Seq-Well



Sparse sampling of gene expression



Top	Gene Expression	US Wealth
1%	15%	35%
10%	55%	73%
20%	73%	86%

The Best Site On the Internet. Probably.

- <https://github.com/Teichlab/scg> lib structs

Detailed visual guides to dozens of single-cell genomics methods

Adapter and primer sequences:

Barcoded Tn5 sequence s5: 5'- TCGTCGGCAGCTCTCCACG [8-bp Tn5 index] GCGATCGAGGACGGCAGATGTGTATAAGAGACAG

Barcoded Tn5 sequence s7: 5'- GTCTCGTGGGCTCGGTGTCCCTGTCC [8-bp Tn5 index] CACCGTCTCCGCCTCAGATGTGTATAAGAGACAG -3'

Tn5 binding site 19-bp Mosaic End (ME) bottom: 5' - /Phos/AGATGTGTATAAGAGACAG -3'

P5 index primer entry point (s5): 5'- TCGTCGGCAGCGTCTCCACGC -3'

P7 index primer entry point (s7): 5'- GTCTCGTGGGCTCGGCTGTCCCTGTCC -3'

P5 index primer: 5'- AATGATACTGGCACCACCGAGATCTACAC[i5]TCGTGGCAGCGTCT

P7 index primer: 5'- CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGCTCGCTGCCGTGCC -3'

Read 1 sequencing primer: 5'- GCGATCGAGGACGGCAGATGTGTATAAGAGACAG -3'

Index 1 sequencing primer (i7): 5' - CTGTCTCTTATACACATCTGAGGCAGAGACGGT

Product 1 (*s5* at both ends, not amplifiable due to semi-suppressive PCR):

5'- TCGTCGGCAGGTCTCCACGC[8-bp Tn5 index]GCGATCGAGGACGGCAGATGTGTATAAGACAGXXXXXXXXXXXX...XXX CTGCTCTTATACACATCT
TCTACACATATTCTCTGTC XXX...XXXXXXXXXXXXCACAGAATATGTGAGACGGCAGGGACTAGCG[8-bp Tn5 index]CGCACCTCTGCGACGGCTGCT -5'

Product 2 (s7 at both ends, not amplifiable due to semi-suppressiv PCR):

5' - **GTC**TGTCGGCTCGGCTGTCCTGTCC[8-bp Tn5 index]CACC~~GTC~~CGCTCAAGATGTGATAAGACAGXXXXXXXXXXXX...XXX CTGTCTCTTACACATCT
TCTACACATATTCTGTCT XXX...XXXXXXXXXXGACAGAGAAATGTGTAGA~~CTCCG~~CCCTG~~CAC~~[8-bp Tn5 index]CCTGTCCTGT~~CGG~~T~~CGG~~TGCTCTG -5'

“What I cannot create, I do not understand.” --Feynman

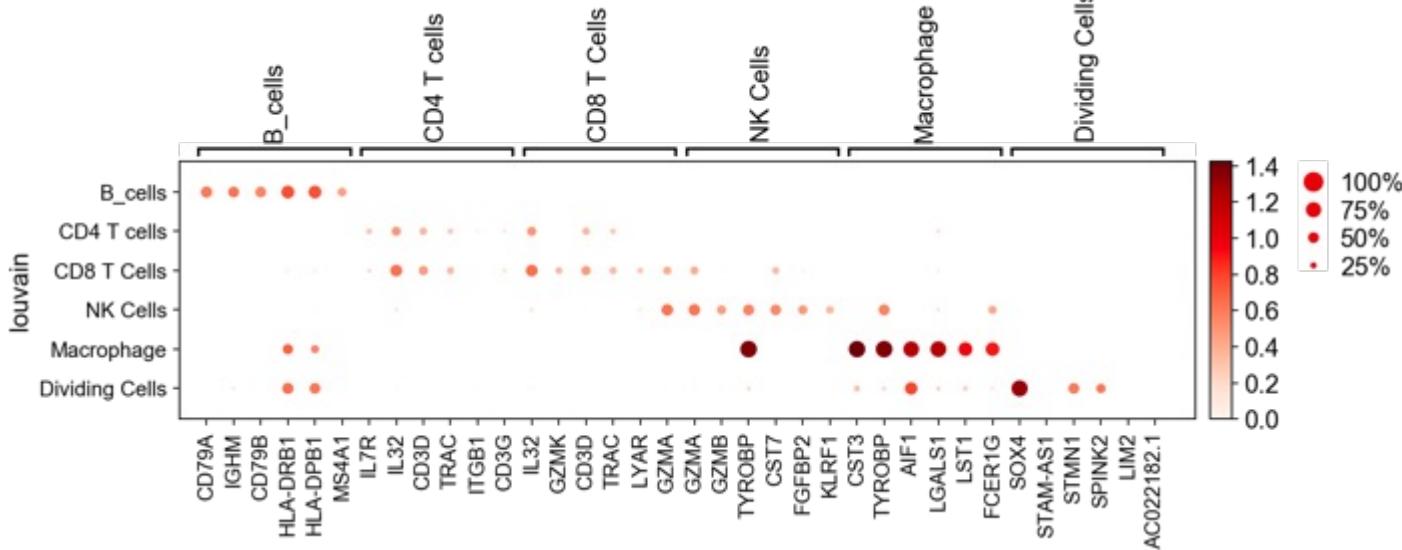
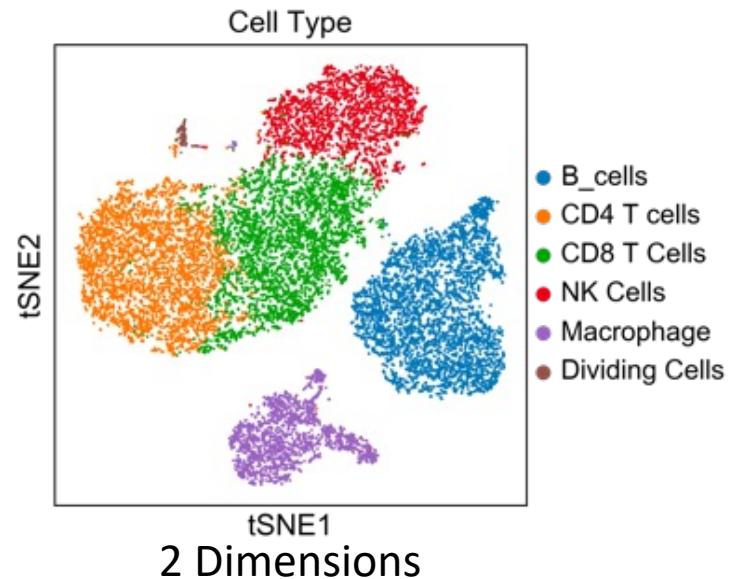
Basic scRNAseq pipeline

	Gene 1	Gene 2	Gene 3	Gene 4
Cell1	0	0	4	1
Cell2	0	1	0	12
Cell3	0	0	0	11
Cell4	5	0	0	2
Cell5	20	1	0	0

25,000+ Dimensions →

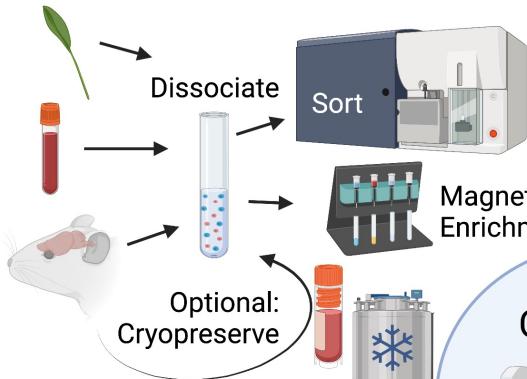
Dimensionality Reduction

Feature Selection
Normalization
PCA / NMF / ICA
tSNE / UMAP



Single Cell Workflow Overview

A. Sample Prep



- Collect sample
- Digest / Dissociate
- Enrich for cell types
- Deplete dead cells
- Remove debris

Resources Committed \$ (⌚ ?)

Opportunities to Abort and Retry

B. Barcoding / Library Prep

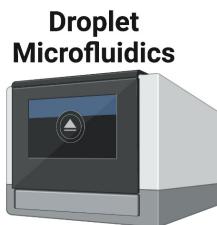
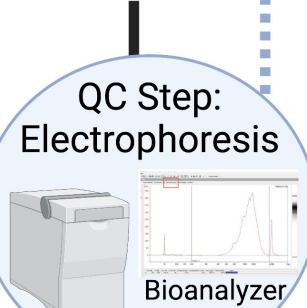


Plate-based
Nanowell

- Capturing and barcoding cells
- Amplifying barcoded cDNA or other library
- Adding sequencing adapters



Optional QC Step: Skim Sequencing

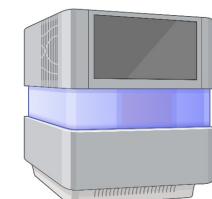


\$\$\$

(\$)
optional

C. Sequencing

Short Reads:
(Gene Exp., ATAC, CNV, Immune Profiling)



Illumina

and / or

Long Reads:
(immune profiling, isoform seq.)

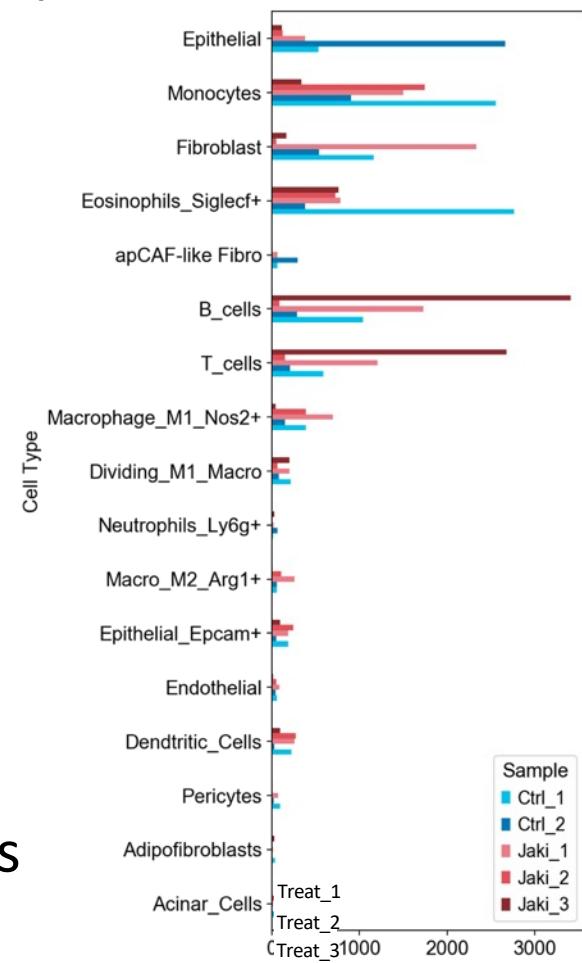
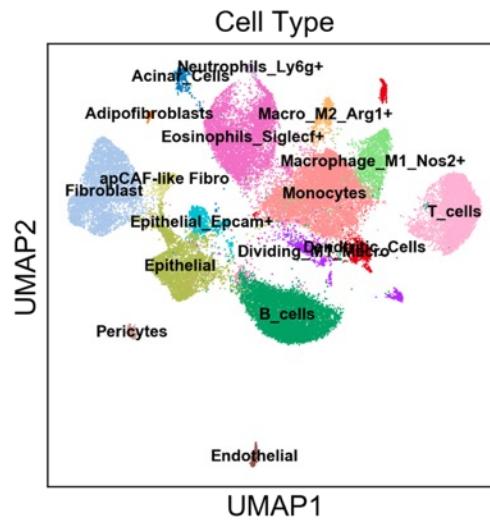
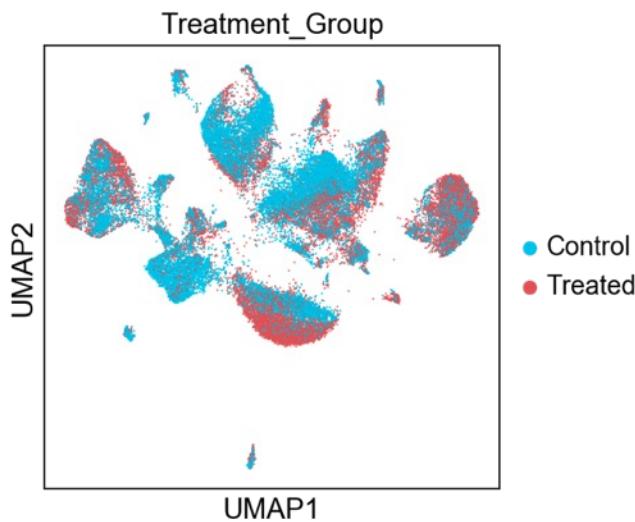


Oxford Nanopore

PacBio

\$\$\$

scRNAseq is a poor cytometry tool



- Unreliable – highly sensitive to conditions
- Expensive
- Low throughput

... But if you must, at least use some statistics:

propeller: testing for differences in cell type proportions in single cell data

<https://www.biorxiv.org/content/10.1101/2021.11.28.470236v1.full>

scDC: single cell differential composition analysis

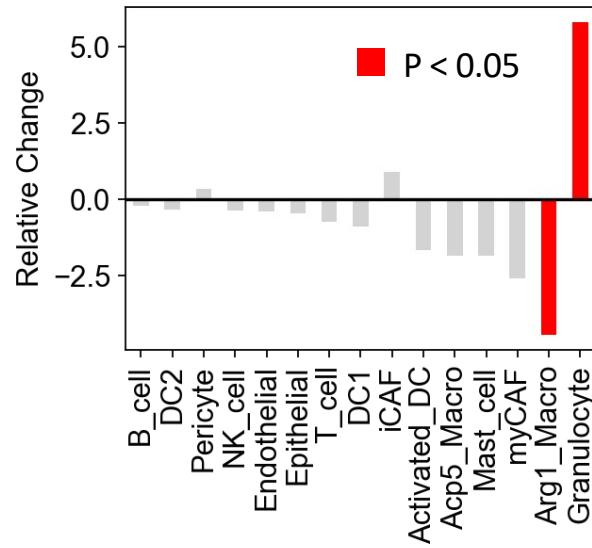
<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3211-9>

CTDS: Cell Type Diversity Statistic

<https://www.frontiersin.org/articles/10.3389/fgene.2022.855076/full>

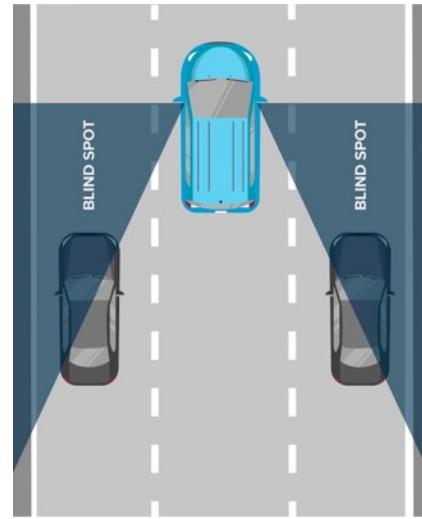
DA-seq: Detecting differentially abundant (DA) subpopulations

<https://www.pnas.org/doi/10.1073/pnas.2100293118>

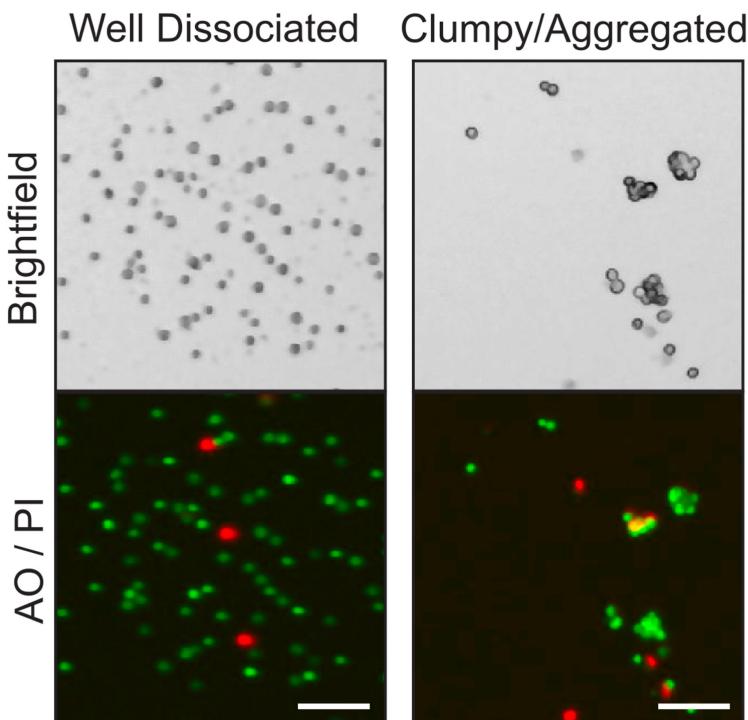
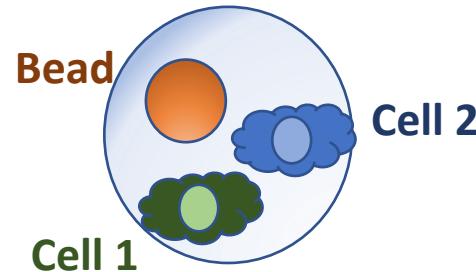
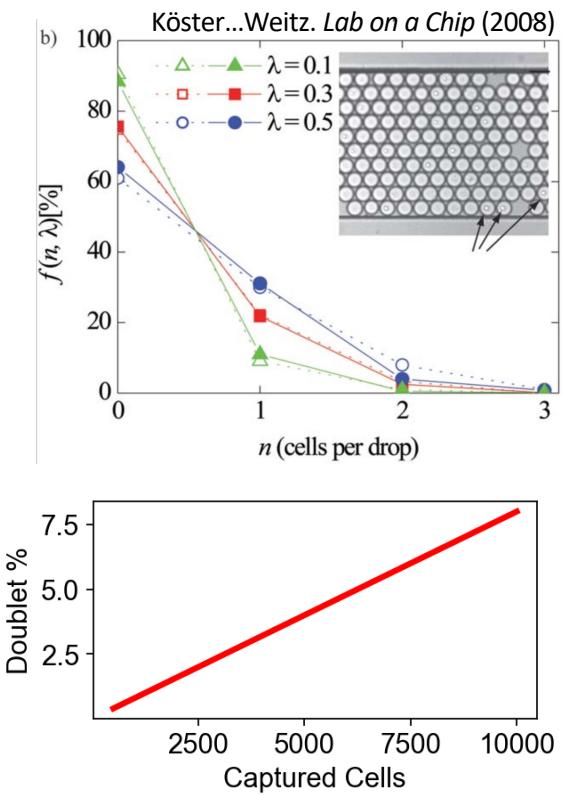


Blind Spots

- Some cell types might be missed
 - Low mRNA count – filtered from matrix
 - Early 10X Genomics Software (v2)
 - Defaulted to exclude lots of lymphocytes
 - Hard to dissociate from tissue
 - Fibroblasts
 - Cells might die quickly during prep
 - Stem cells
 - Fragile: (Acinar cells, Plasma cells)
 - High RNase / protease content (Acinar, Neutrophils)
 - Peripheral blood neutrophils especially!!!
 - Doublets / Multiplets



Doublets / Multiplets



Doublet Filtering

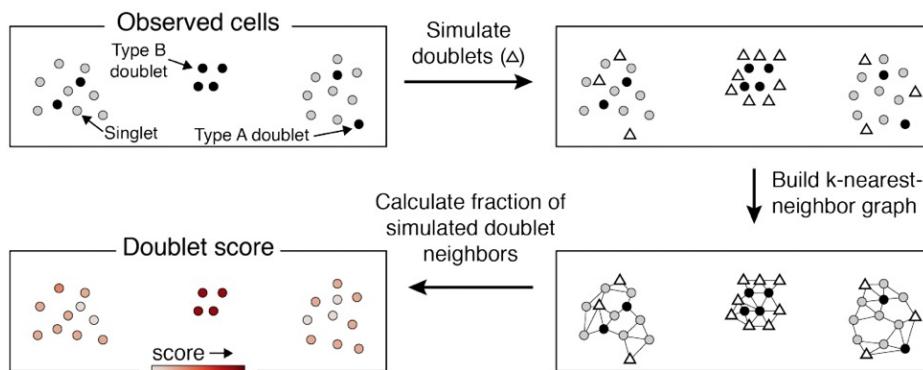
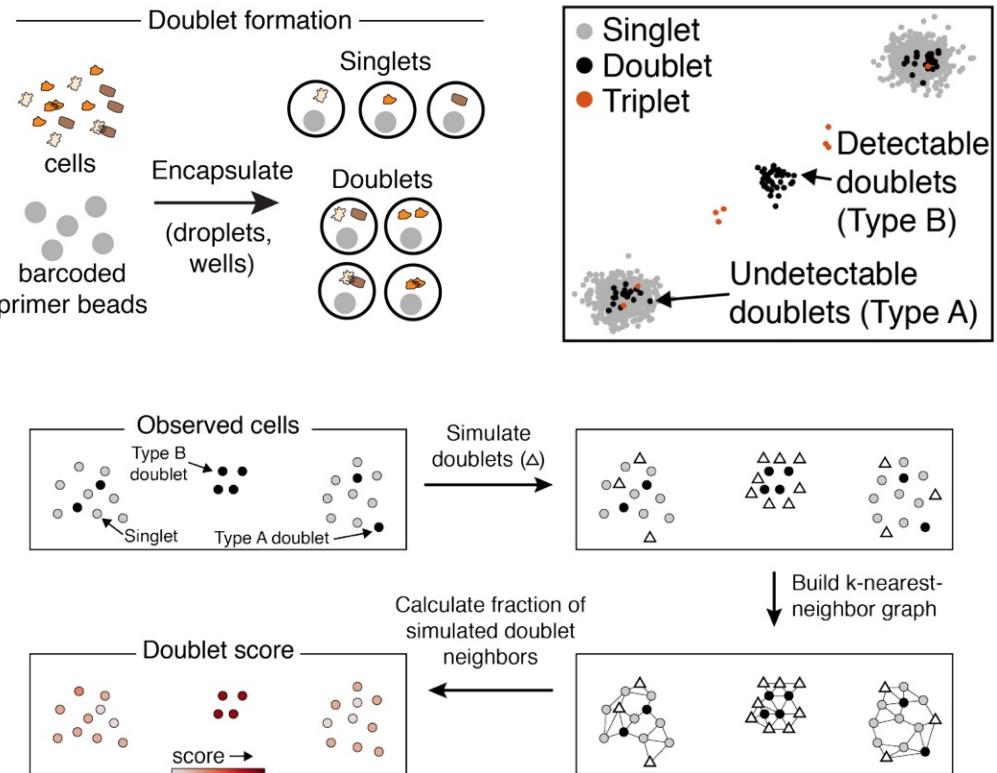
Scrublet

• [DoubletFinder](#) - [R] - Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. [BioRxiv](#)

• [DoubletDecon](#) - [R] - Cell-State Aware Removal of Single-Cell RNA-Seq Doublets. [\[BioRxiv\]](#) ([DoubletDecon: Cell-State Aware Removal of Single-Cell RNA-Seq Doublets](#))

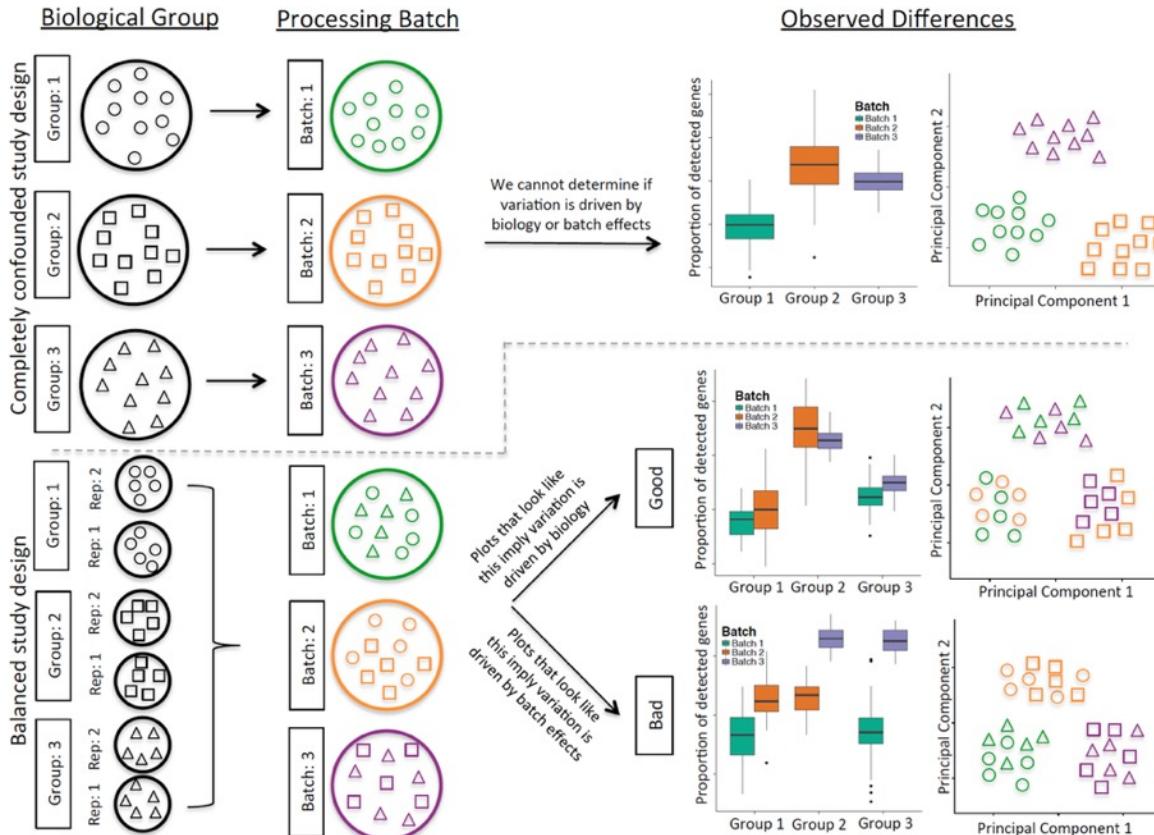
• [DoubletDetection](#) - [R, Python] - A Python3 package to detect doublets (technical errors) in single-cell RNA-seq count matrices. An [R implementation](#) is in development.

• [Scrublet](#) - [Python] - Computational identification of cell doublets in single-cell transcriptomic data. [BioRxiv](#)

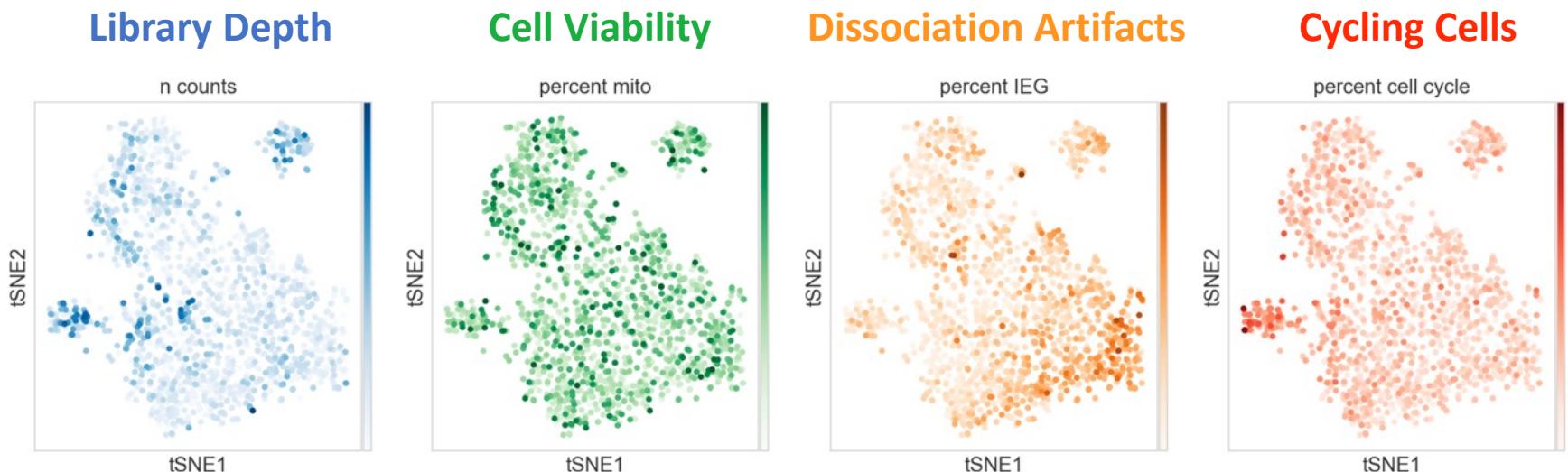


Wolock et al. (2018) bioRxiv

Batch effects and study design



Example Sources of Unwanted Variation & “Batch Effects”



Sex – matched studies are helpful!

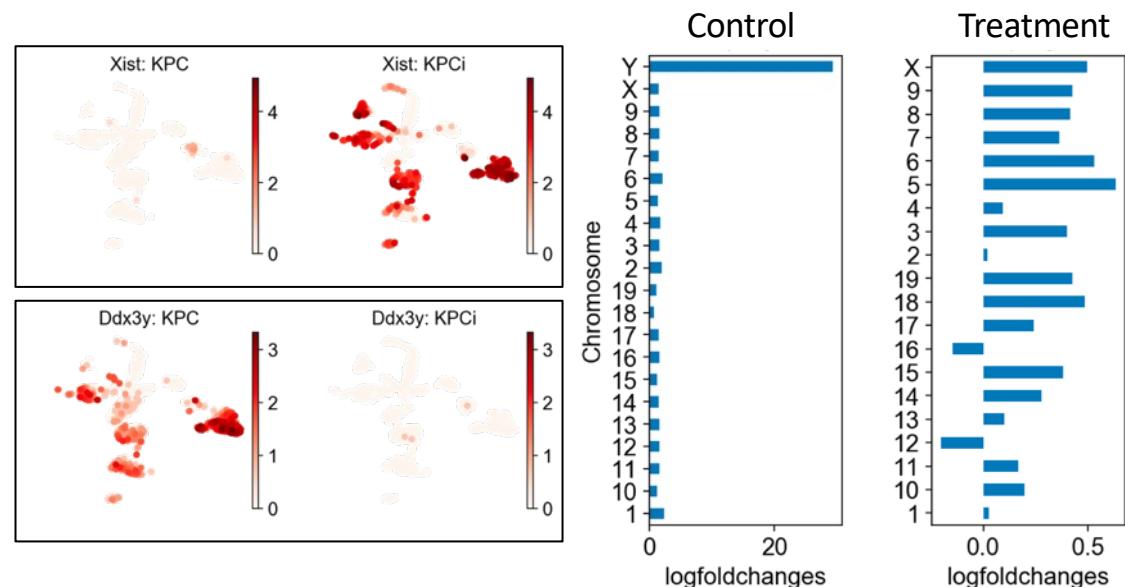
Major confounder: Male / Female

Treatment: Female

Control: Male

Consequence:

Unsupervised differential gene expression calling will be dominated by sex-specific expression. No way of separating this variable from the treatment variable



Batch Correction

Confounded Study Example:

WT and KO mice

Prepared on same day

Same colony

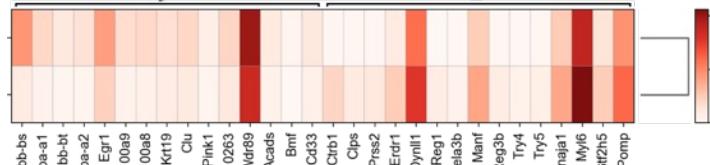
Same set of hands

Diffex dominated by same genes within every cluster

! major batch effect issues

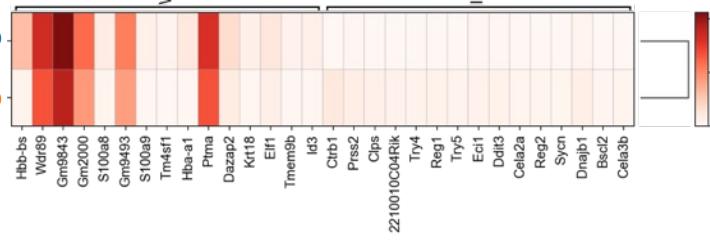
Myeloid

Ctrl
KO



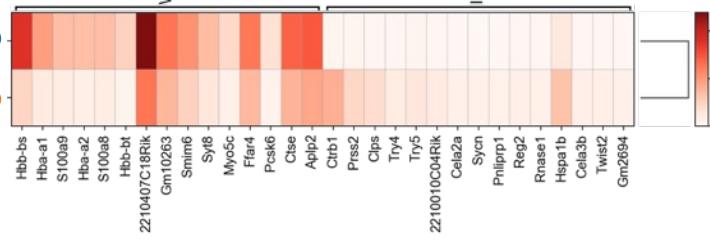
B cells

Ctrl
KO



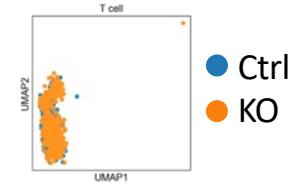
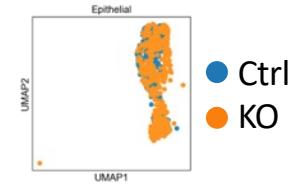
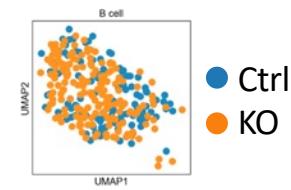
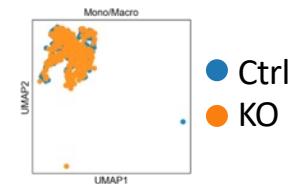
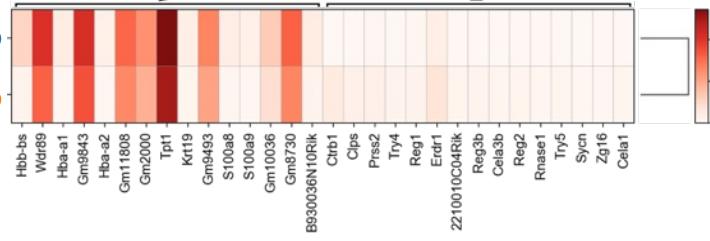
Epithelial

Ctrl
KO

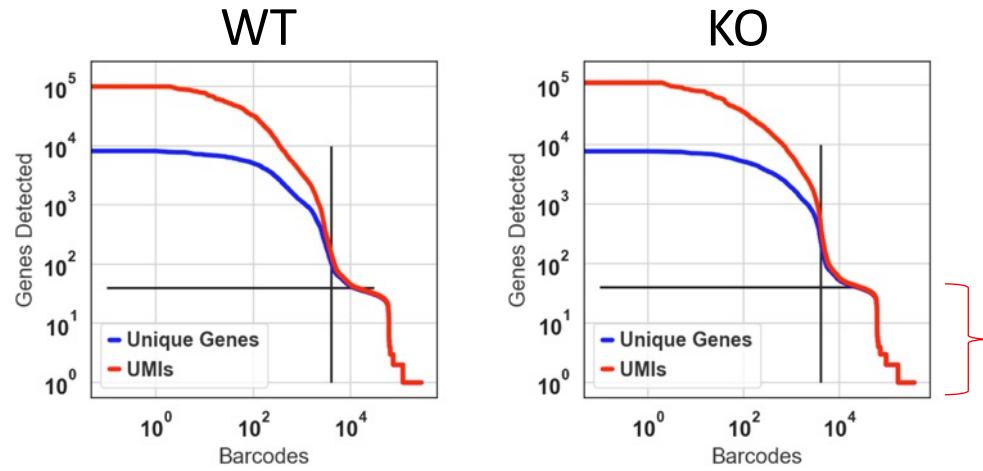


T cell

Ctrl
KO

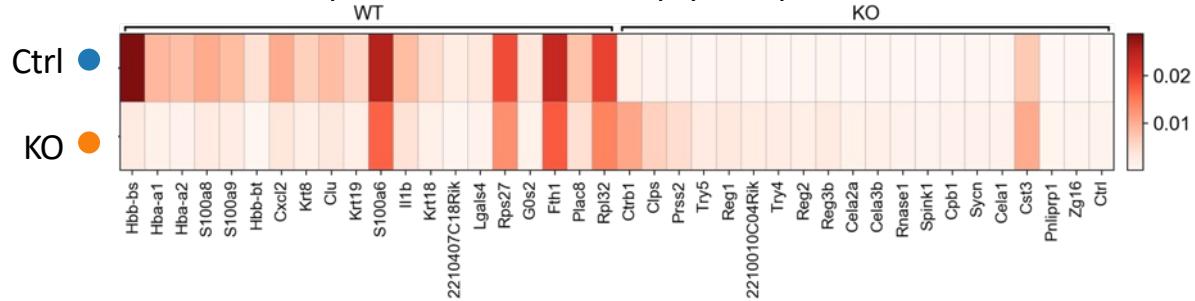


Controlling for batch effects



Ambient RNA
in droplets

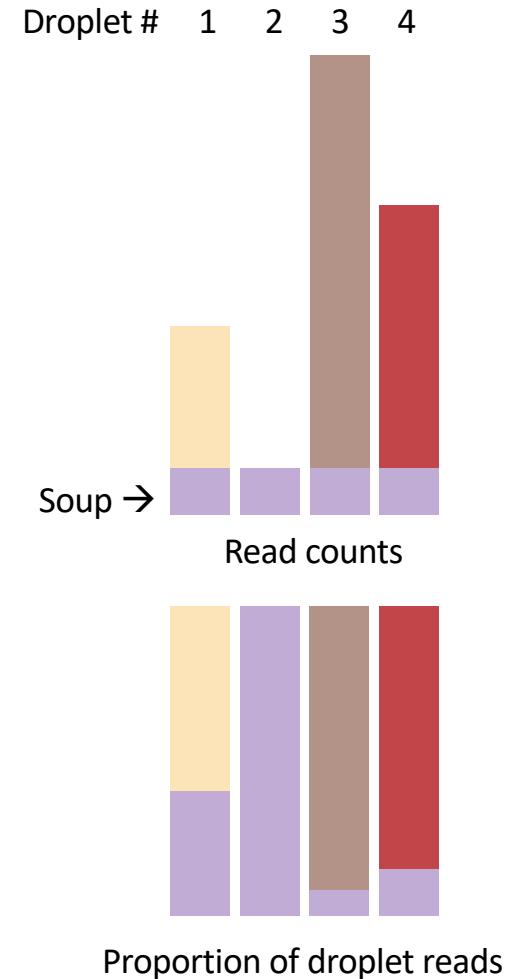
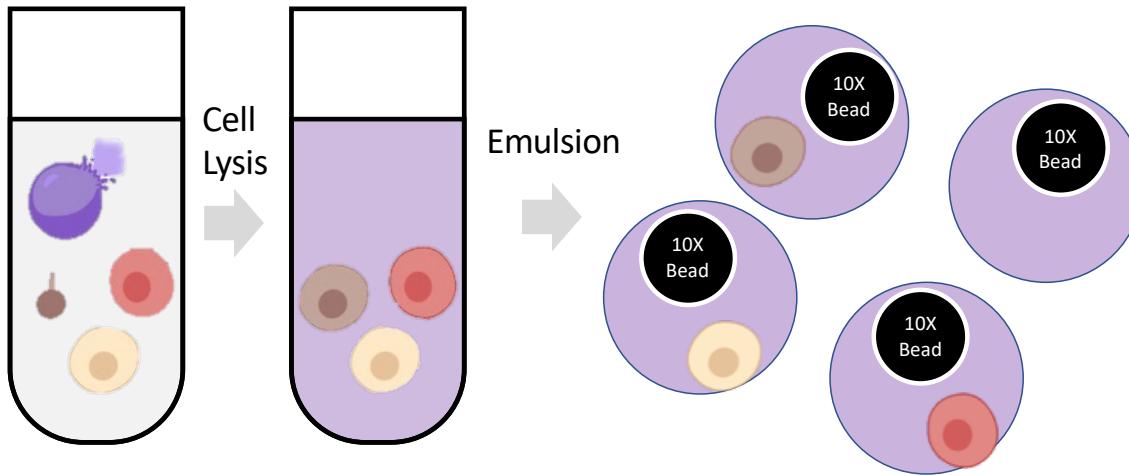
Differential expression between "empty" droplets:



Significant sources of
contaminating mRNA:

WT:	Erythrocytes
WT:	Epithelial
WT:	Granulocytes
KO:	Acinar cells

Ambient RNA: "SOUP"

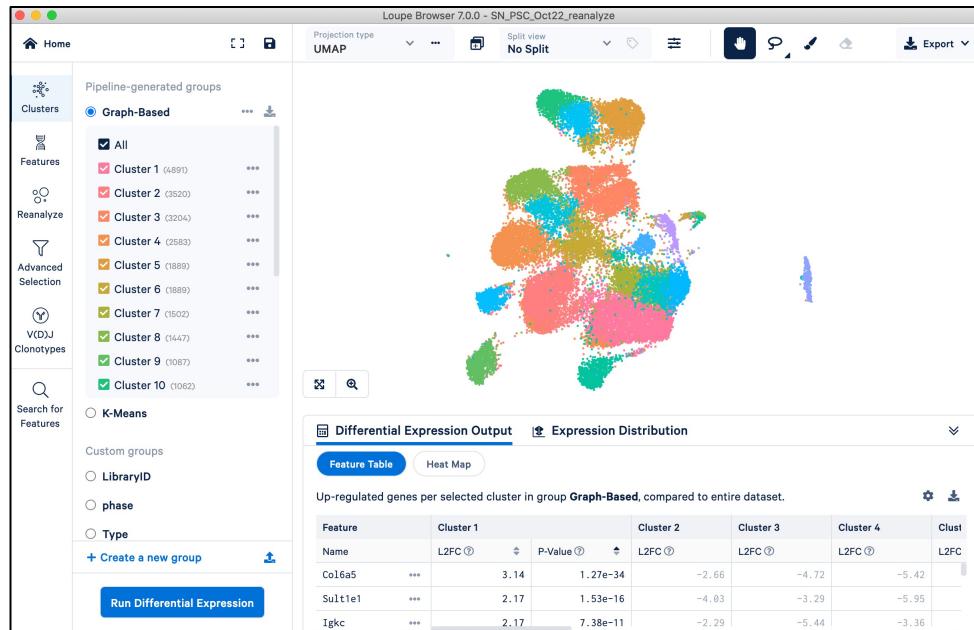


Computational methods to mitigate (but not totally fix) Soup:

- *SoupX*
- *DecontX*
- *Cellbender*

The best time to fix Ambient RNA problems is at the bench!

Loupe Cell Browser



<https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest>

Can:

- Quickly visualize genes
- Do guided clustering via marker genes / hand-drawn selections
- Calculate Differential Expression
- Subset cells and re-run PCA / UMAP / Clustering

Can't

- Pseudotime
- Transcription factor analysis
- Sequence-level analysis
- other fancy things

Getting started with your own analyses

Rahul Satija -

R

<https://satijalab.org/seurat>

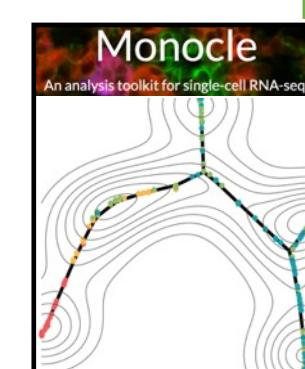


Fabian Theis - München



<https://scanpy.readthedocs.io/en/latest/>

Python



Cole Trapnell – WashU



Macosko lab

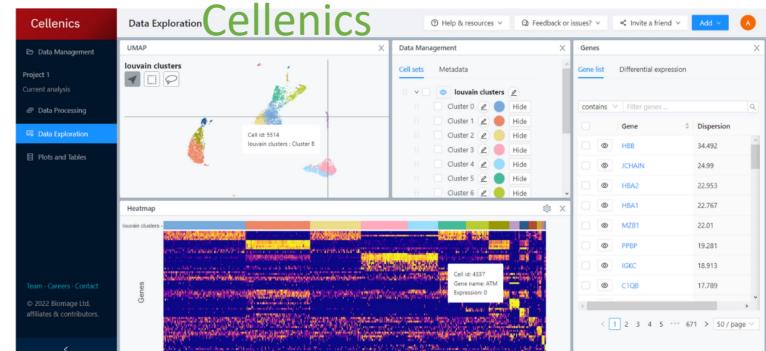
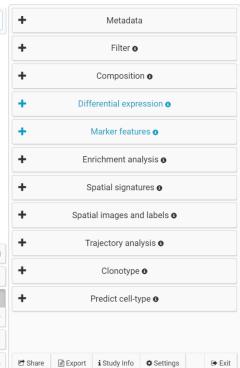
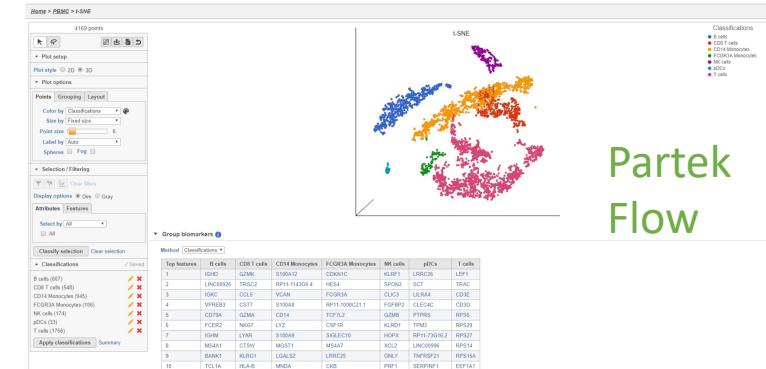
AWESOME SINGLE CELL RESOURCE

<https://github.com/seandavi/awesome-single-cell>

Or if you fear the command line and have \$\$\$...

Feature	Cellenics®	BioTuring Browser	10X Loupe Browser	Partek® Flow®	Cellxgene	Rosalind
Type of application	Web	Desktop	Desktop	Desktop/ AWS	Desktop/ Web	Web
Open-source?	Yes	No	No	No	Yes	No
Free for academia	Yes	No	Yes	No	Yes	No
Intuitive	Yes	Yes	Yes	No	Partly	Yes
Data import: count matrices	Yes	Yes	No	Yes	No	Yes
Data import: Fastq and h5ad files	No	Yes	No	Yes	Only h5ad	Yes
Data import: Seurat or Scanpy object	No	Yes	No	Yes	No	No
Multiple single-cell technologies supported	Yes	Yes	10X only	Yes	10x only	10x only
Multi-omics technologies supported	No	Yes	Yes	Yes	No	Yes
Species supported	All	All	Limited/All	All	Limited	Limited
Data sharing between researchers	Limited	Yes	No	Yes	Limited	Yes
Data processing: QC and filtering	Yes	Yes	No	Yes	No	Limited
Data integration: support for several algorithms	Yes	Yes	No	Yes	No	Limited
Data normalization: support for several methods	No	No	No	Yes	No	Limited
In-depth data exploration	Yes	Yes	Yes	Yes	Limited	Yes
Cell set prediction	Yes	Yes	No	No	No	Yes
Range of plots: trajectory	Yes	Yes	No	Yes	No	No
A variety of plots available	Yes	Yes	Limited	Yes	Limited	Yes
Plots are fully customizable to publishable quality	Yes	Limited	No	Limited	Limited	Yes

Rosalind

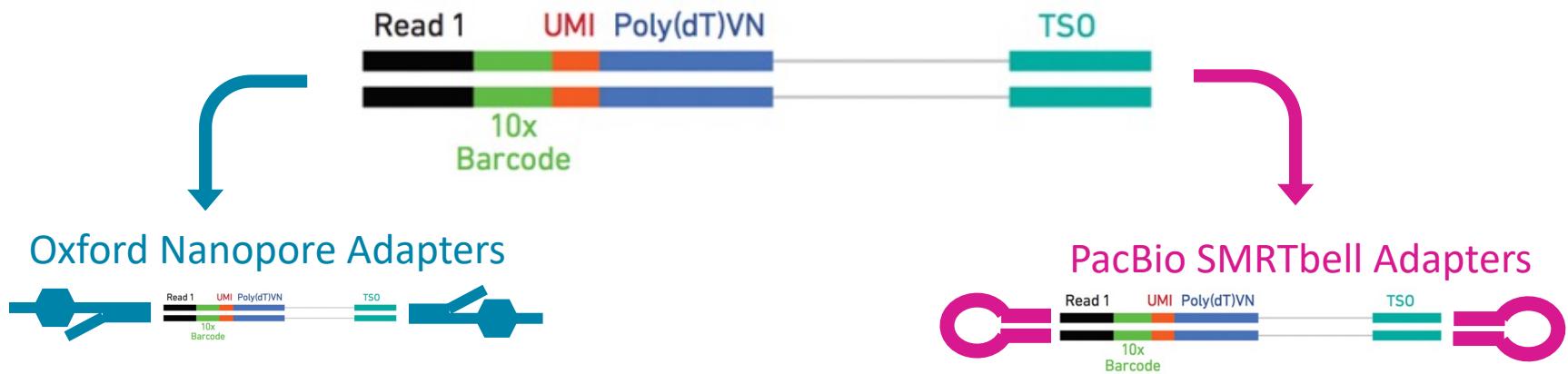


Hacking Droplets



Single cell: Long Read Sequencing

Sequencing full length cDNAs

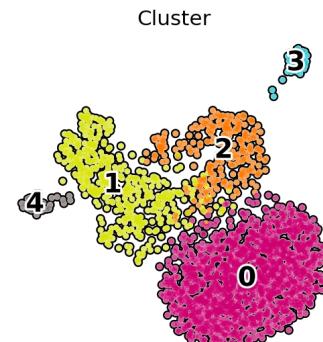
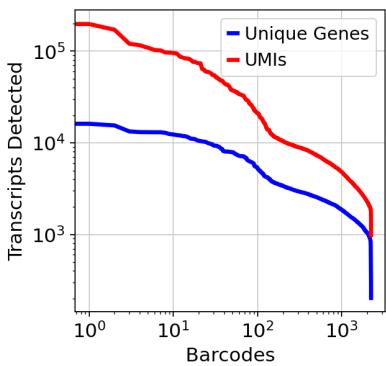
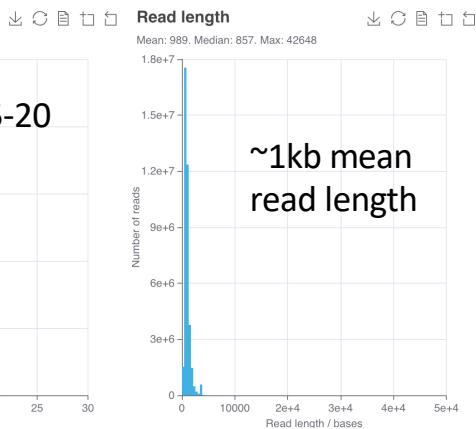
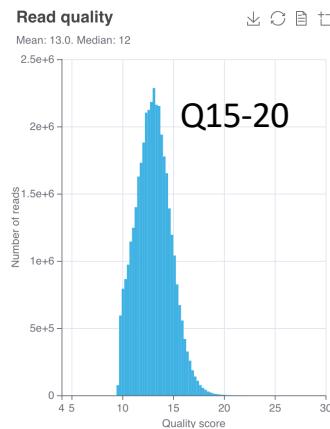


10X on the Promethion

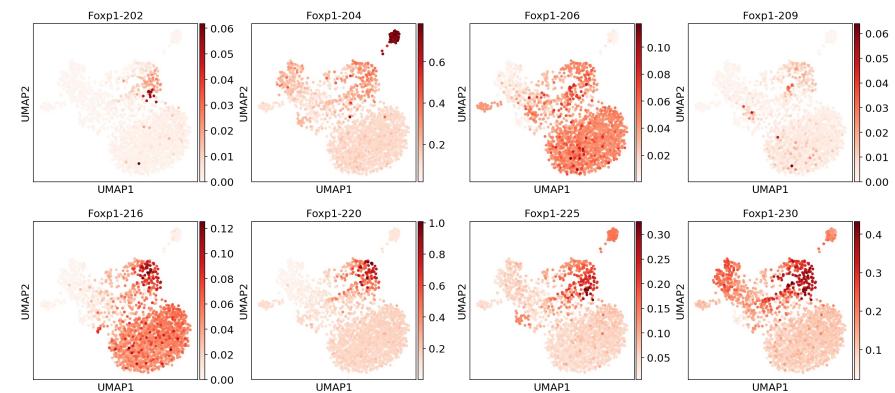
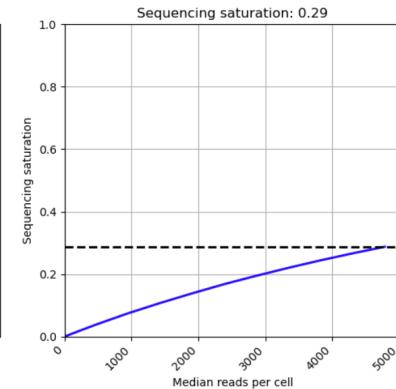
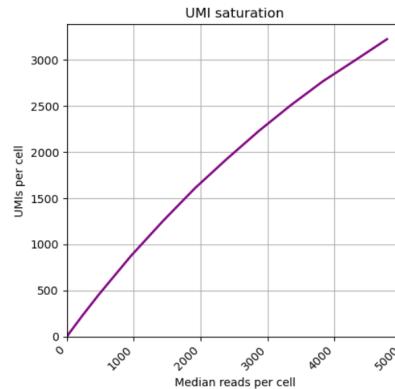


Orthotopic Tumor model MLM-3: Lung met-derived
HPV16+ Oropharyngeal Cancer

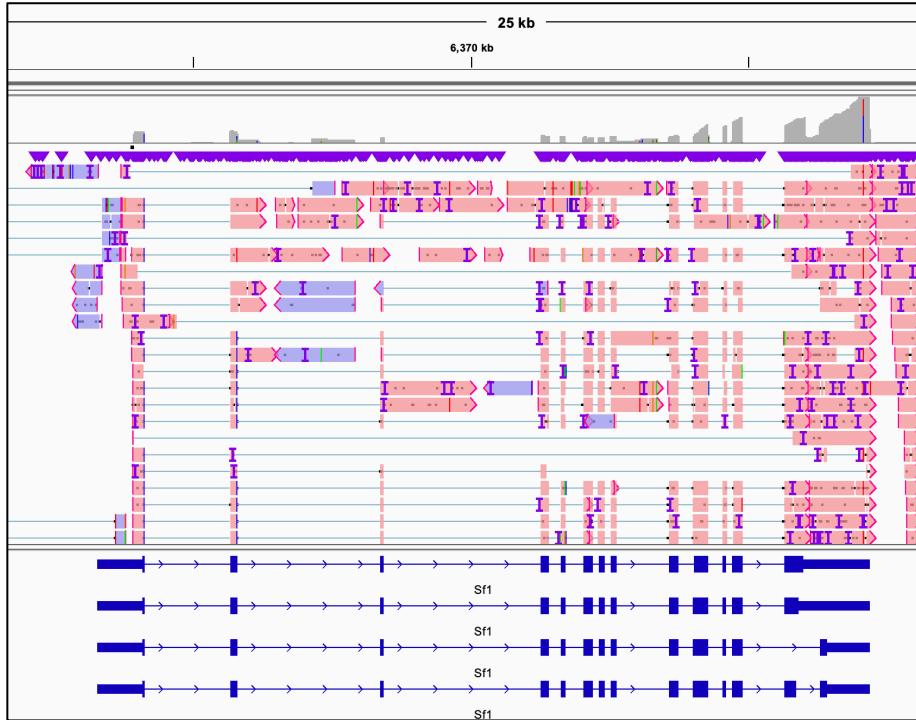
R10 Flow cell: 38M fastq-pass reads



Saturation curves comparable to Illumina



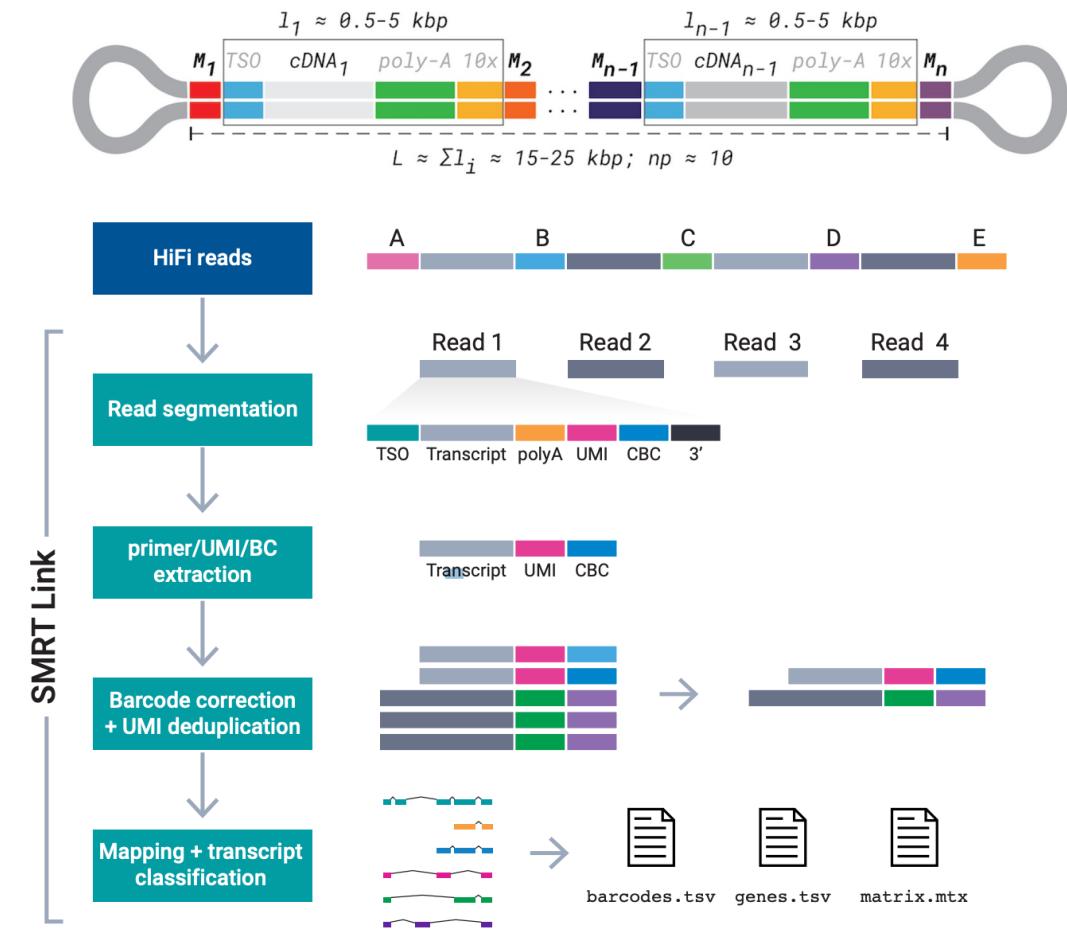
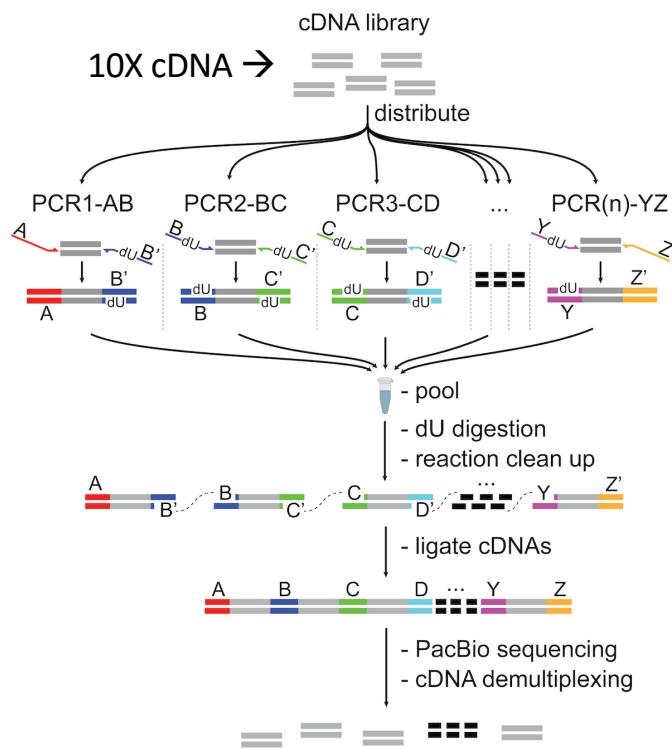
10X on the Promethion



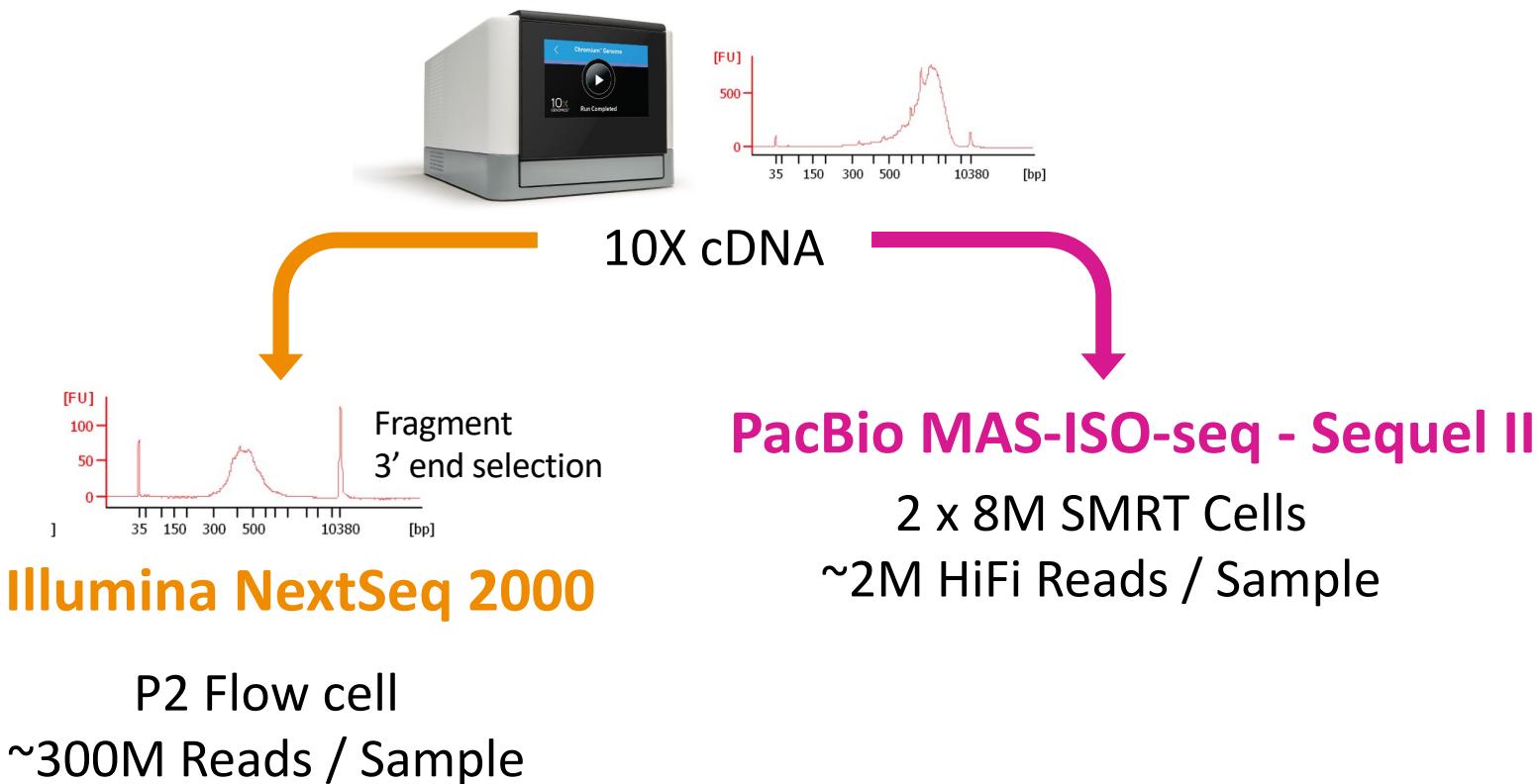
- 35-100M Reads / Flow Cell
- ~Full Length cDNA
- Passable Barcode Identification
- ???? UMI Counting Accuracy
- ???? De-Novo Transcript Assembly

Long Read Sequencing with PacBio

MAS-ISO-Seq / "KINNEX (new version)

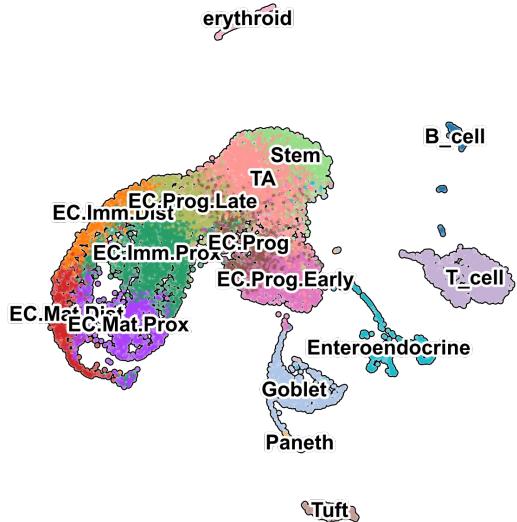


2 Mouse Small Intestine Samples



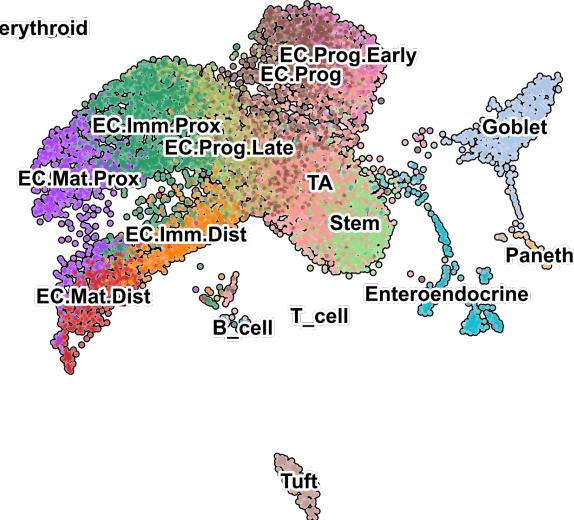
Illumina vs PacBio: Mouse Intestine sample

Cell Types: Illumina

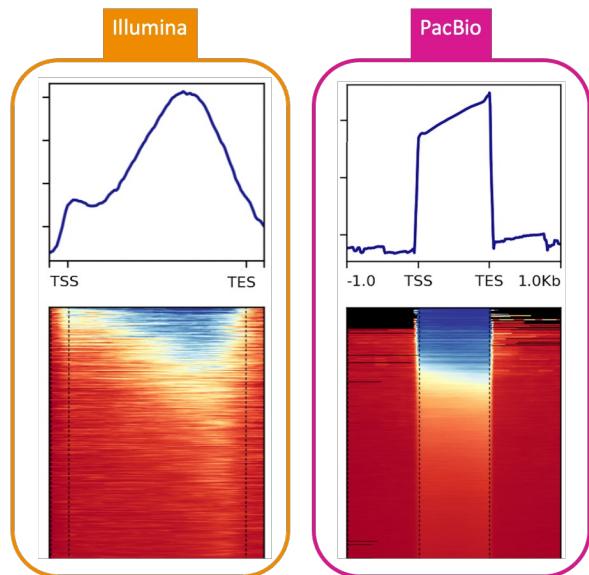


(Gene level data)

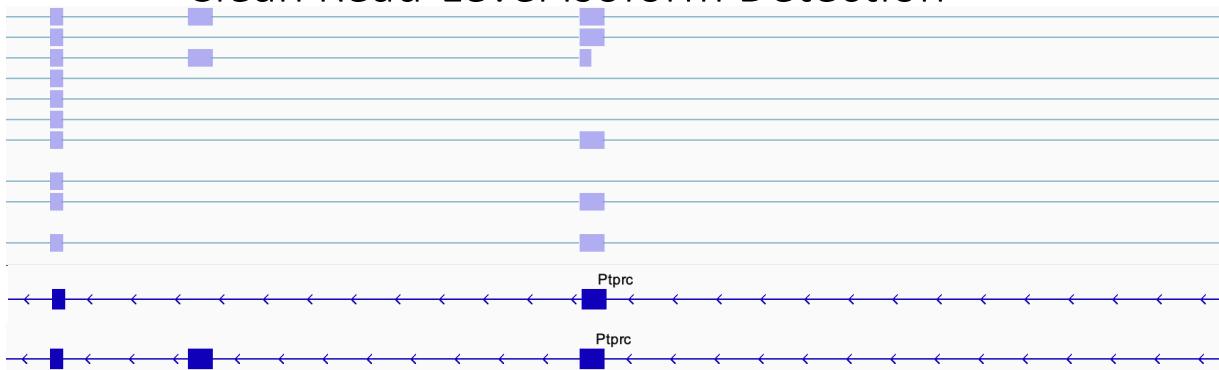
Cell Types: PacBio



Transcript Coverage



Clean Read-Level Isoform Detection

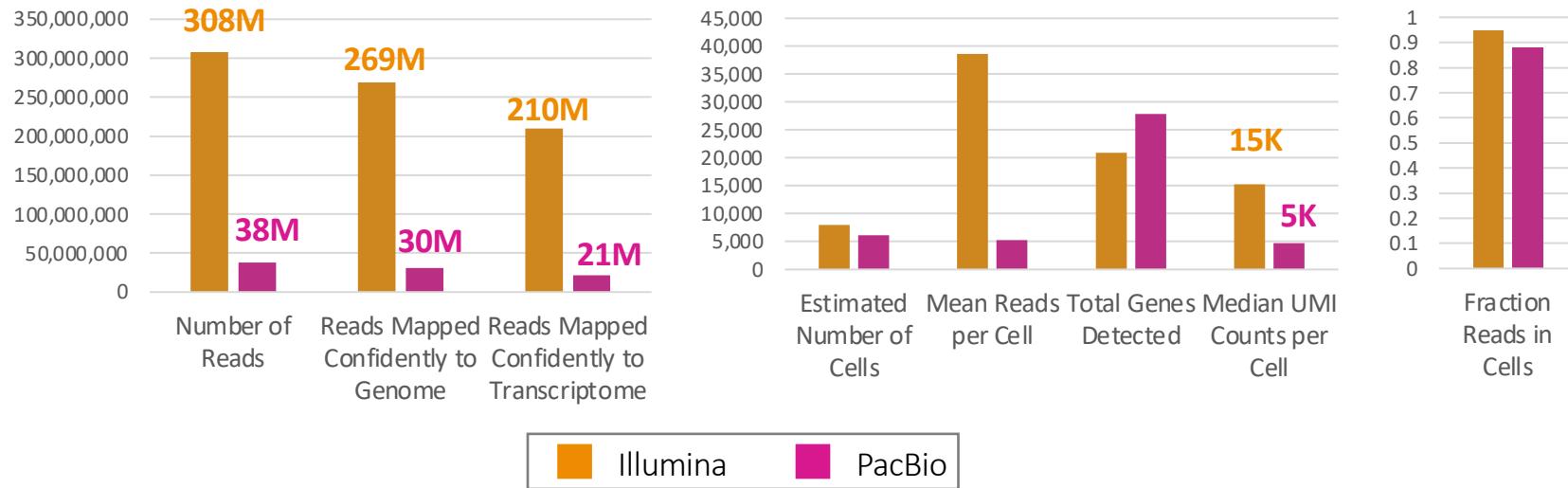


Reference: Haber, Regev (2017)

Illumina vs PacBio: Performance Comparison

At roughly equal sequencing \$\$ spent

RSII



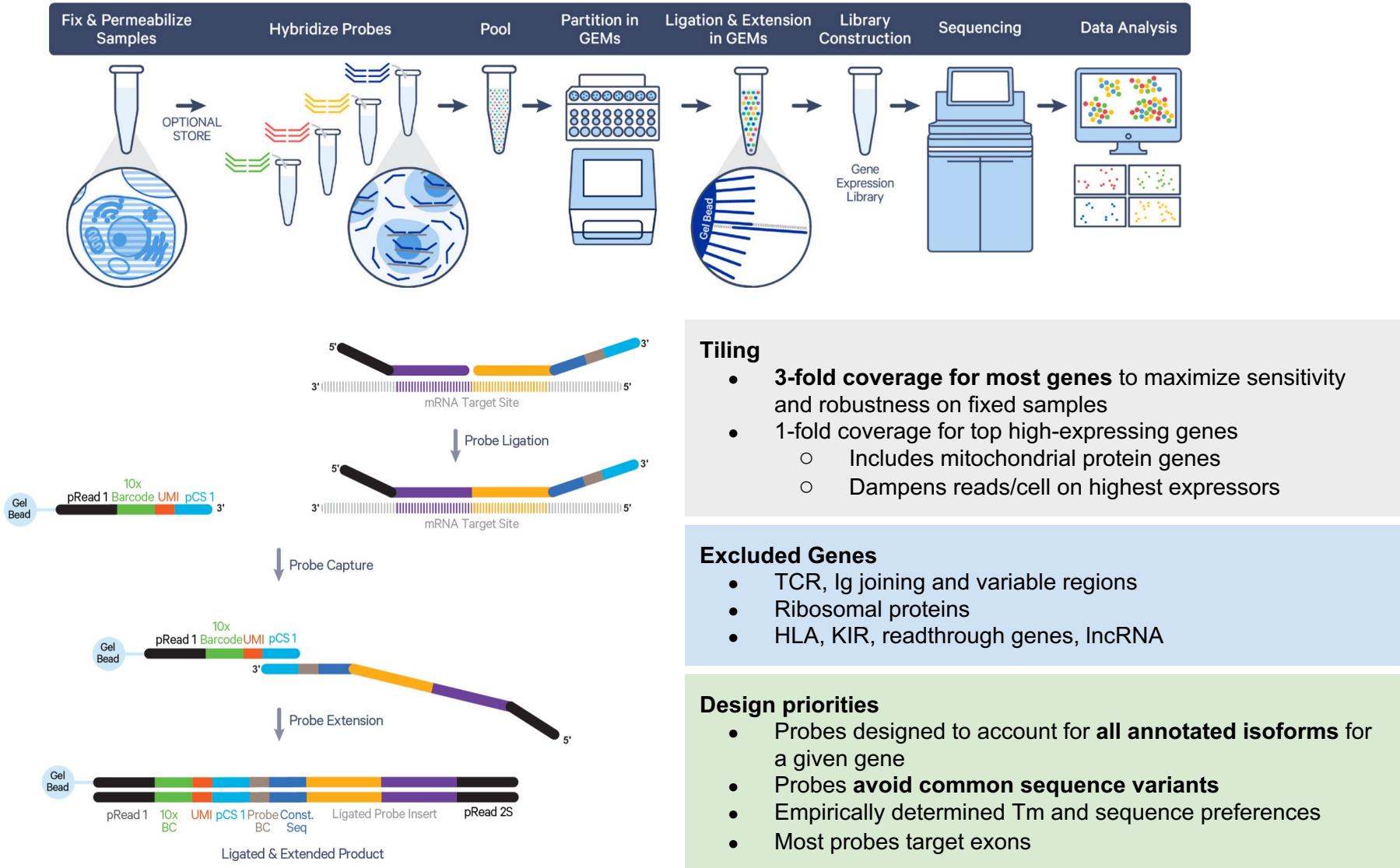
4X Improved
Price / Read:

PacBio Revio



- 200bp-40kb reads
- 25M ZMW flowcells
- Reads: ~12M per flowcell
- 80-100M cDNA molecules
- 4 Cells in parallel
- Quality: single pass accuracy ~85%
- HiFi >99.99%

Fixed RNA Profiling



snPATHO-seq – isolation of nuclei from FFPE cells

 CSH
Cold Spring Harbor Laboratory

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results [View current version of this article](#) [Follow this preprint](#)

snPATHO-seq: unlocking the FFPE archives for single nucleus RNA profiling

Andres F Vallejo, Kate Harvey, Taopeng Wang, Kellie Wise, Lisa M Butler, Jose Polo, Jasmine T Plummer, Alexander Swarbrick, Luciano G Martelotto

doi: <https://doi.org/10.1101/2022.08.23.505054>

This article is a preprint and has not been certified by peer review [what does this mean?].



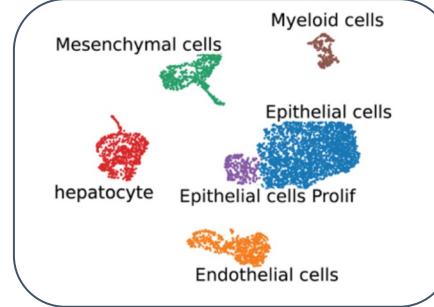
[Abstract](#) [Info/History](#) [Metrics](#) [Preview PDF](#)

Abstract

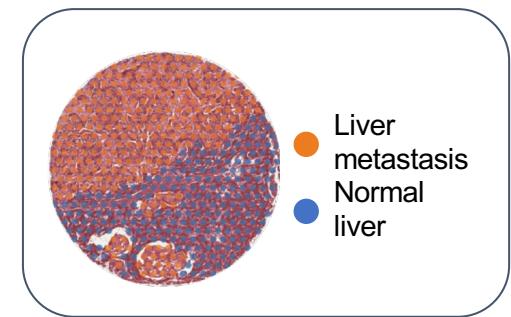
FFPE (formalin-fixed, paraffin-embedded) tissue archives are the largest repository of clinically annotated human specimens. Despite numerous advances in technology, current methods for sequencing of FFPE-fixed single-cells are slow, labour intensive, insufficiently sensitive and have a low resolution, making it difficult to fully exploit their enormous research and clinical



Chromium Single Cell FFPE

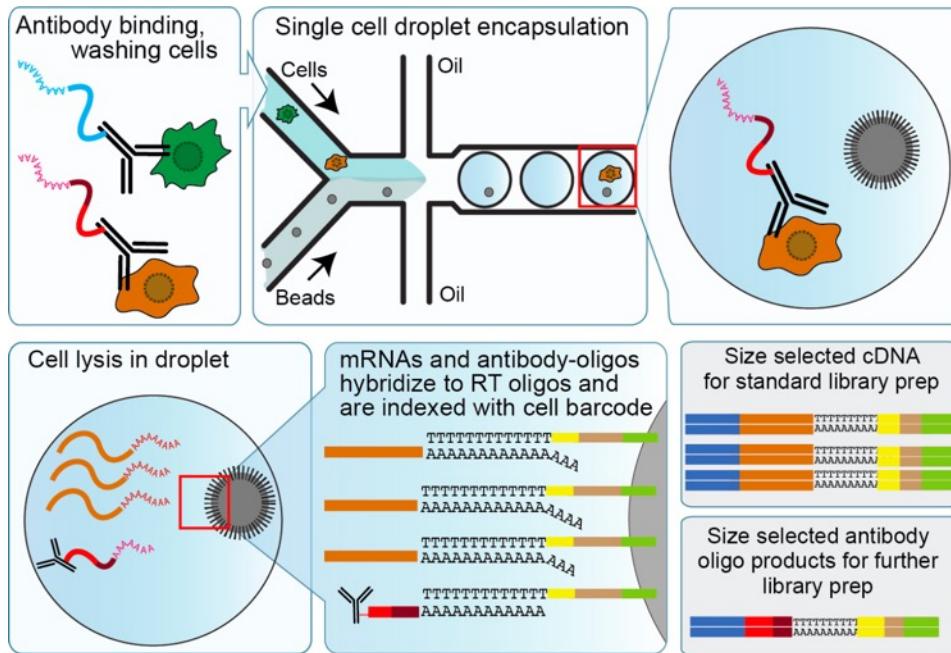


Visium Spatial FFPE



doi: <https://doi.org/10.1101/2022.08.23.505054>

CITE-Seq / REAP-Seq



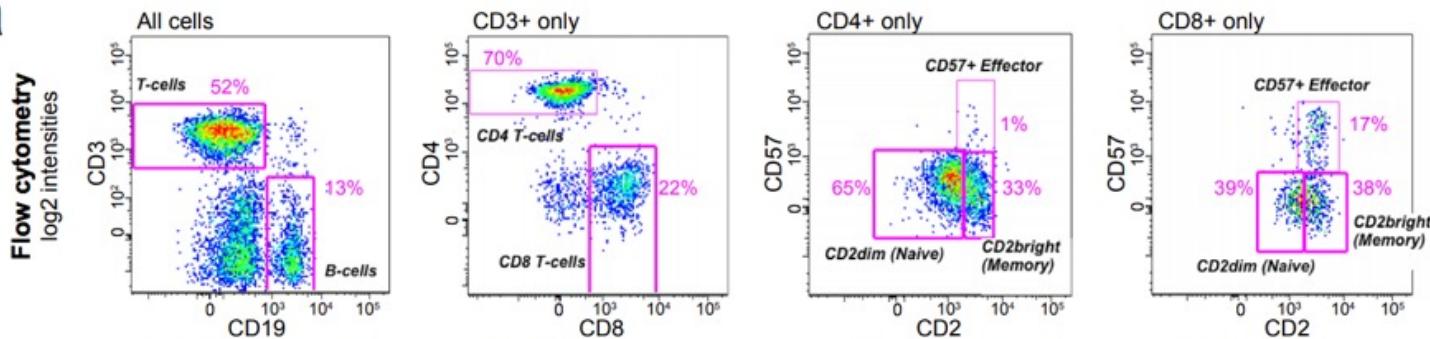
Antibody Derived Tag (ADT)
sequenced as part of normal 10X run

Enables:

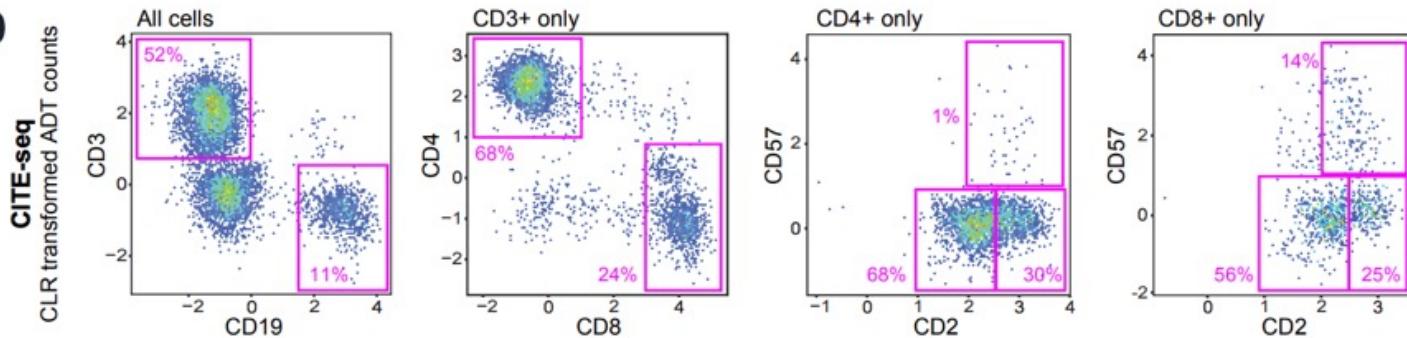
- Simultaneous mRNA + Protein Abundance
- Increased sensitivity to individual targets
- ‘Superloading’

CITE-Seq / REAP-Seq

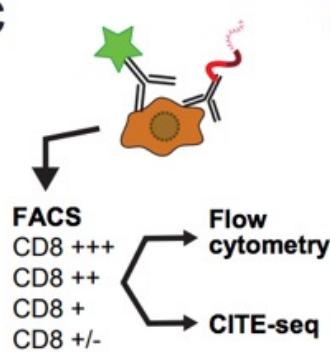
a



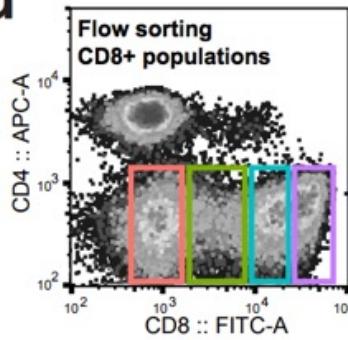
b



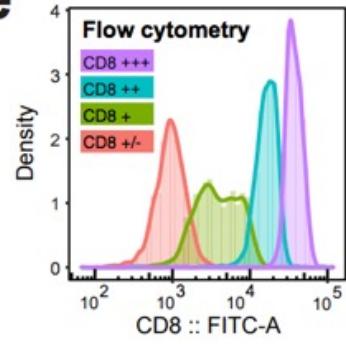
c



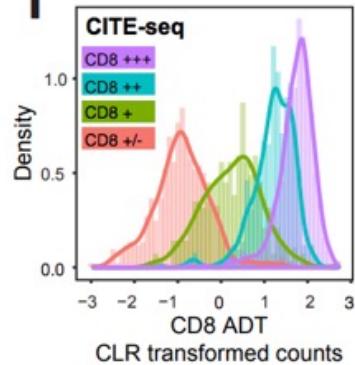
d



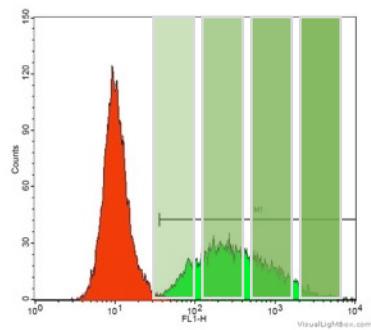
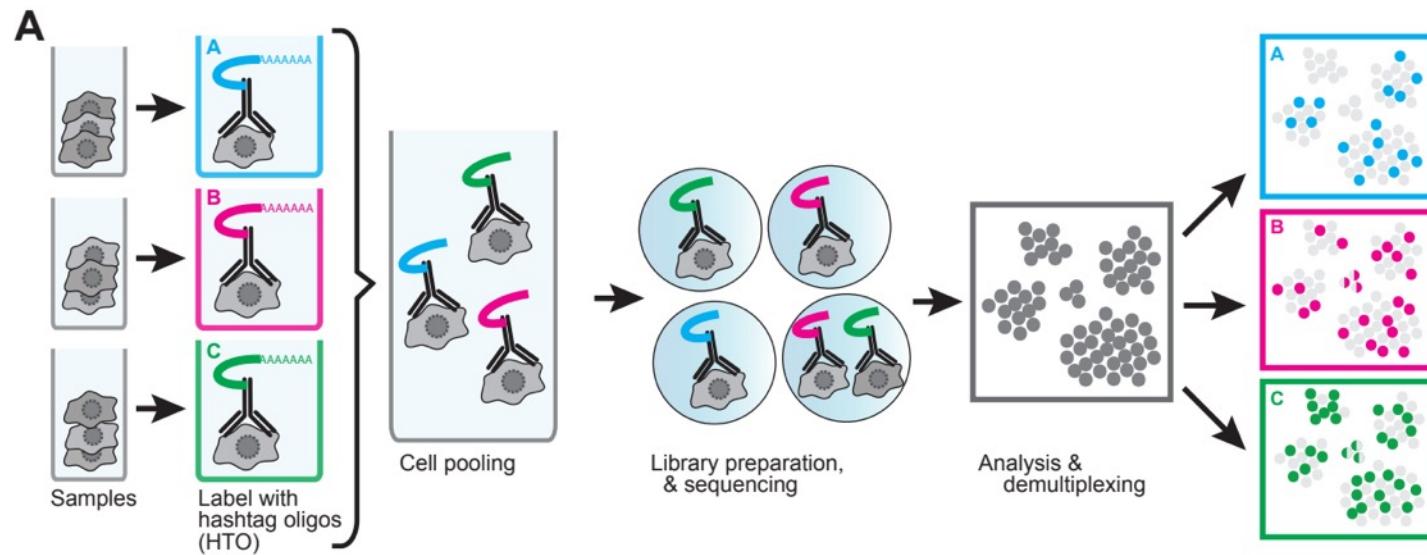
e



f

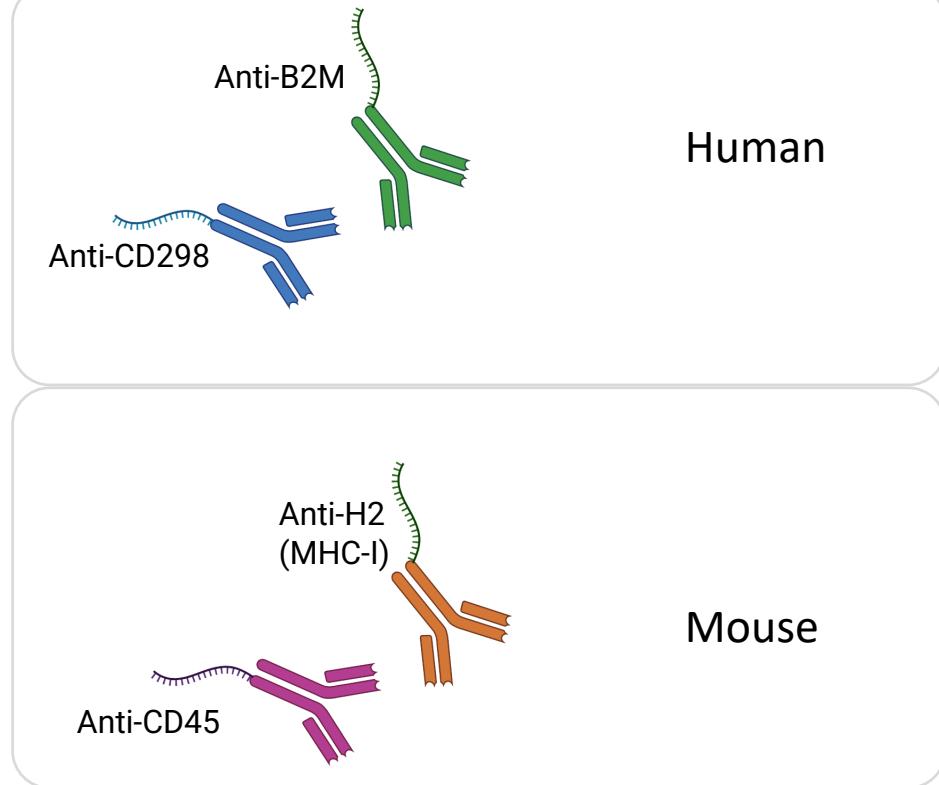


Multiplexing with ADTs: “Cell Hashing”



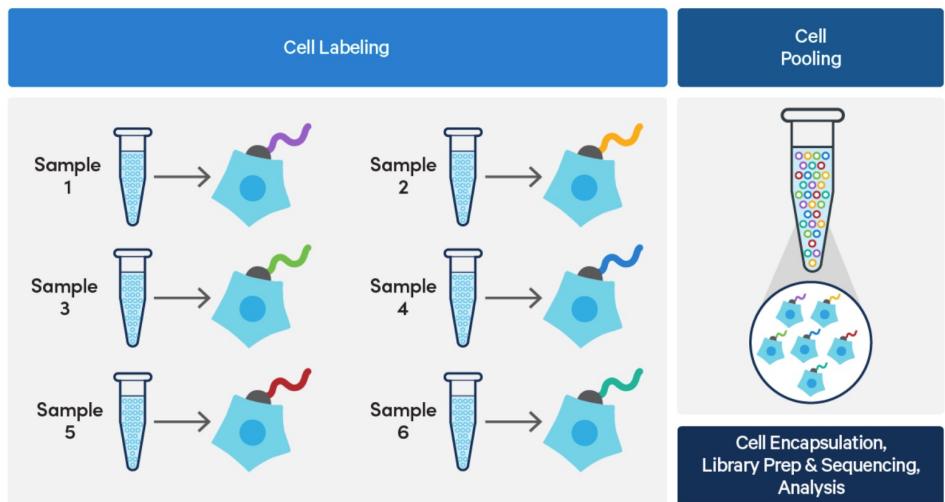
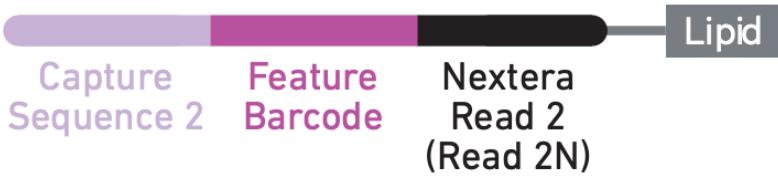
Sort multiple bins → HTO Label → Repool & Capture

Total-Seq (Biolegend)

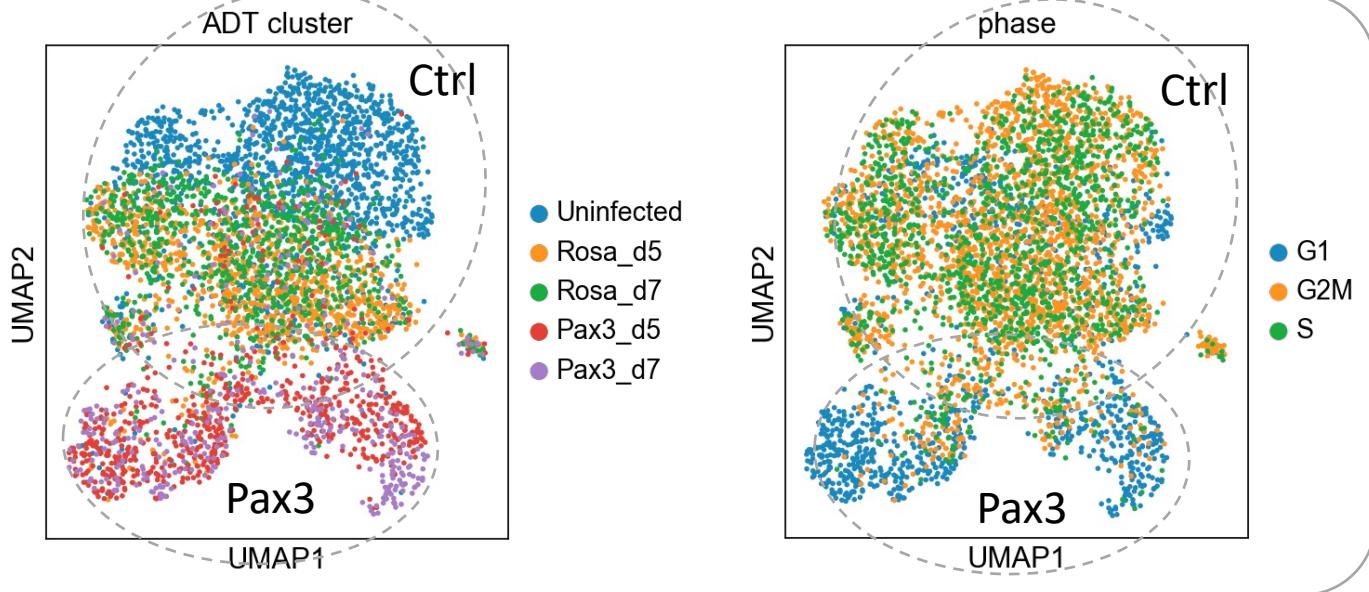


CellPlex (10X Genomics)

Cholesterol / Lipid anchor

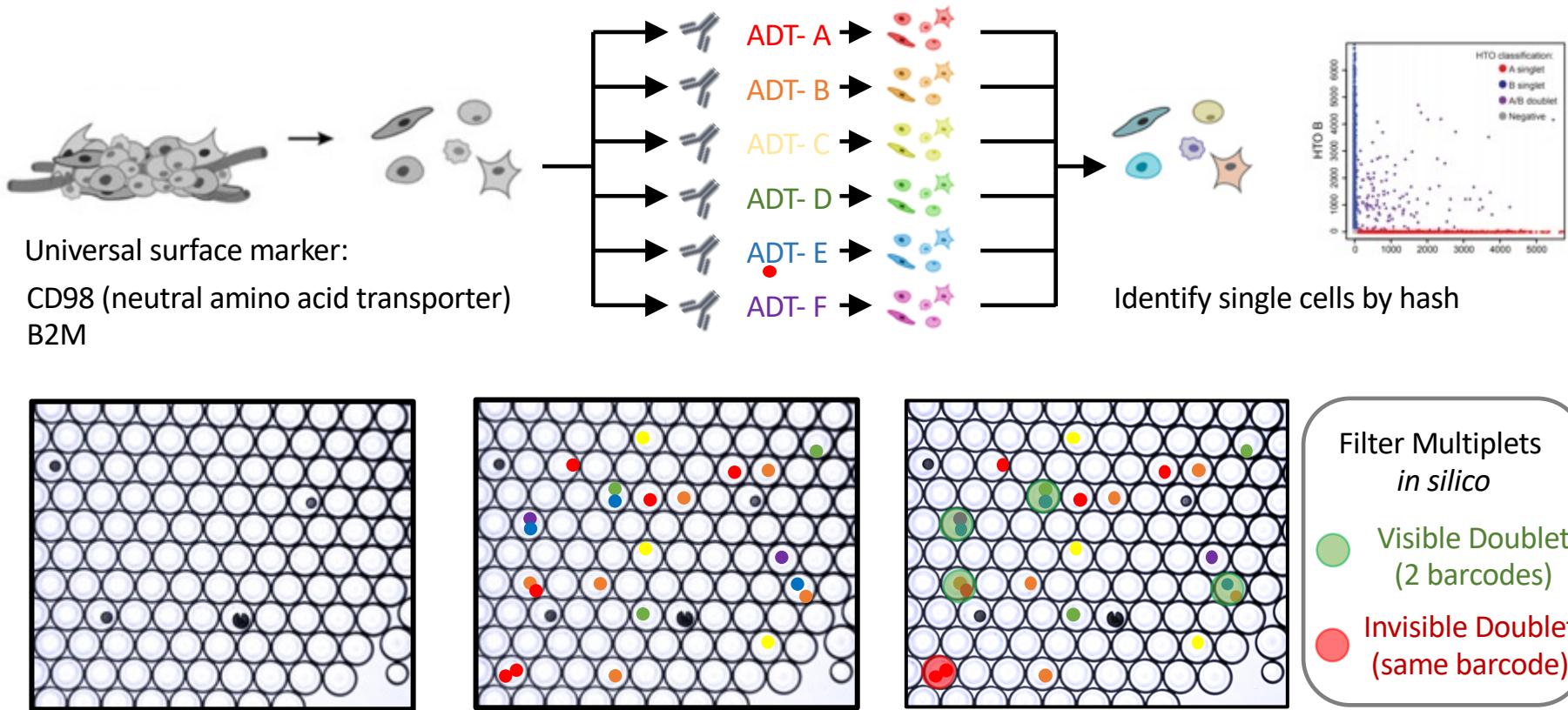


Hashed 1 Lane

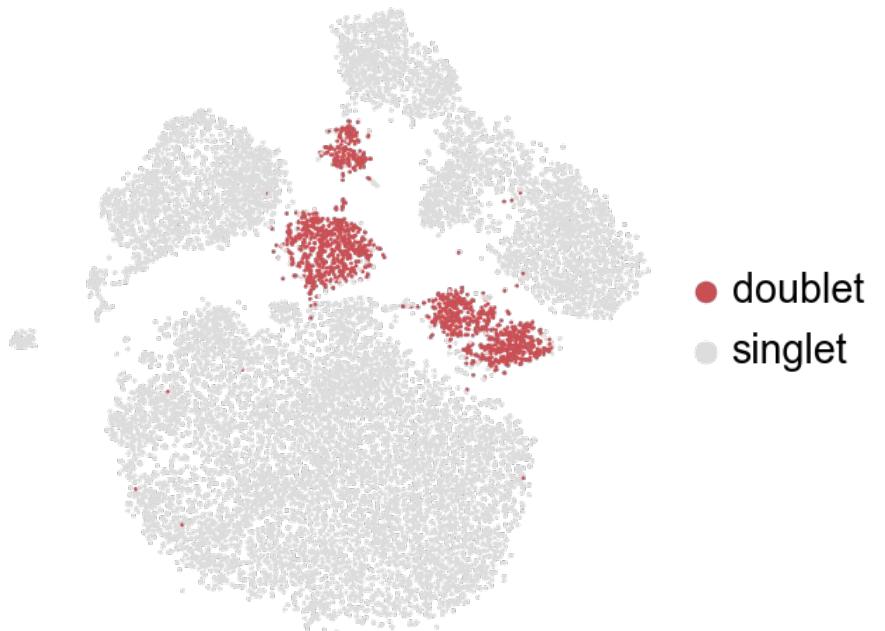


4 Lanes

Superloading with ADTs: “Cell Hashing”

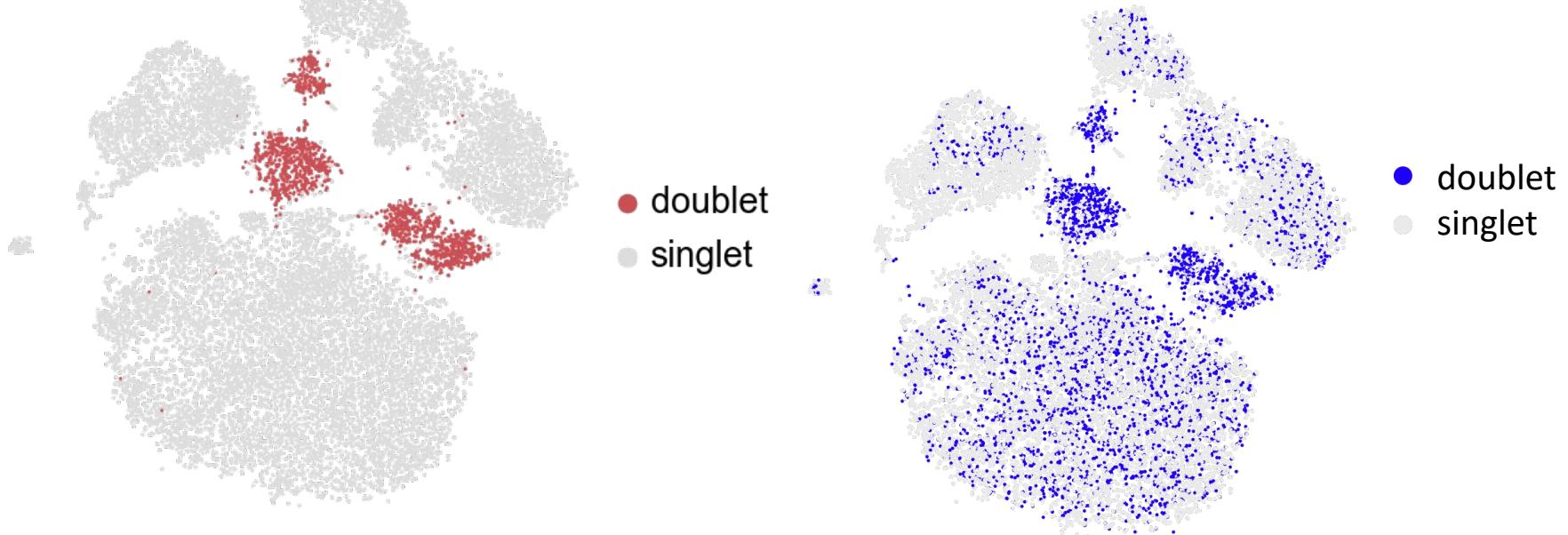


Doublet Detection by Cell Cluster



singlet	17,110
doublet	1,668

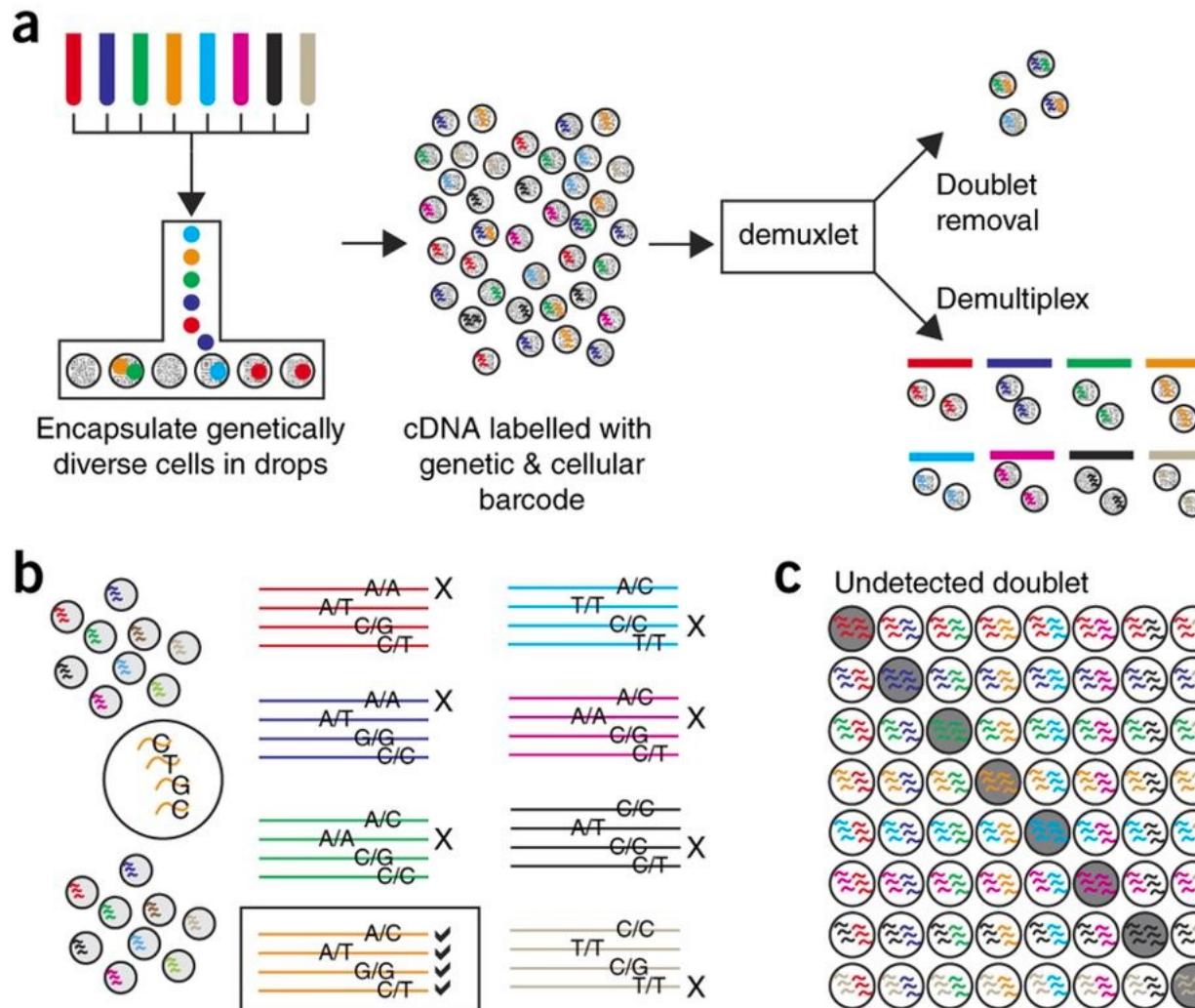
Doublet Detection by Hash Tag



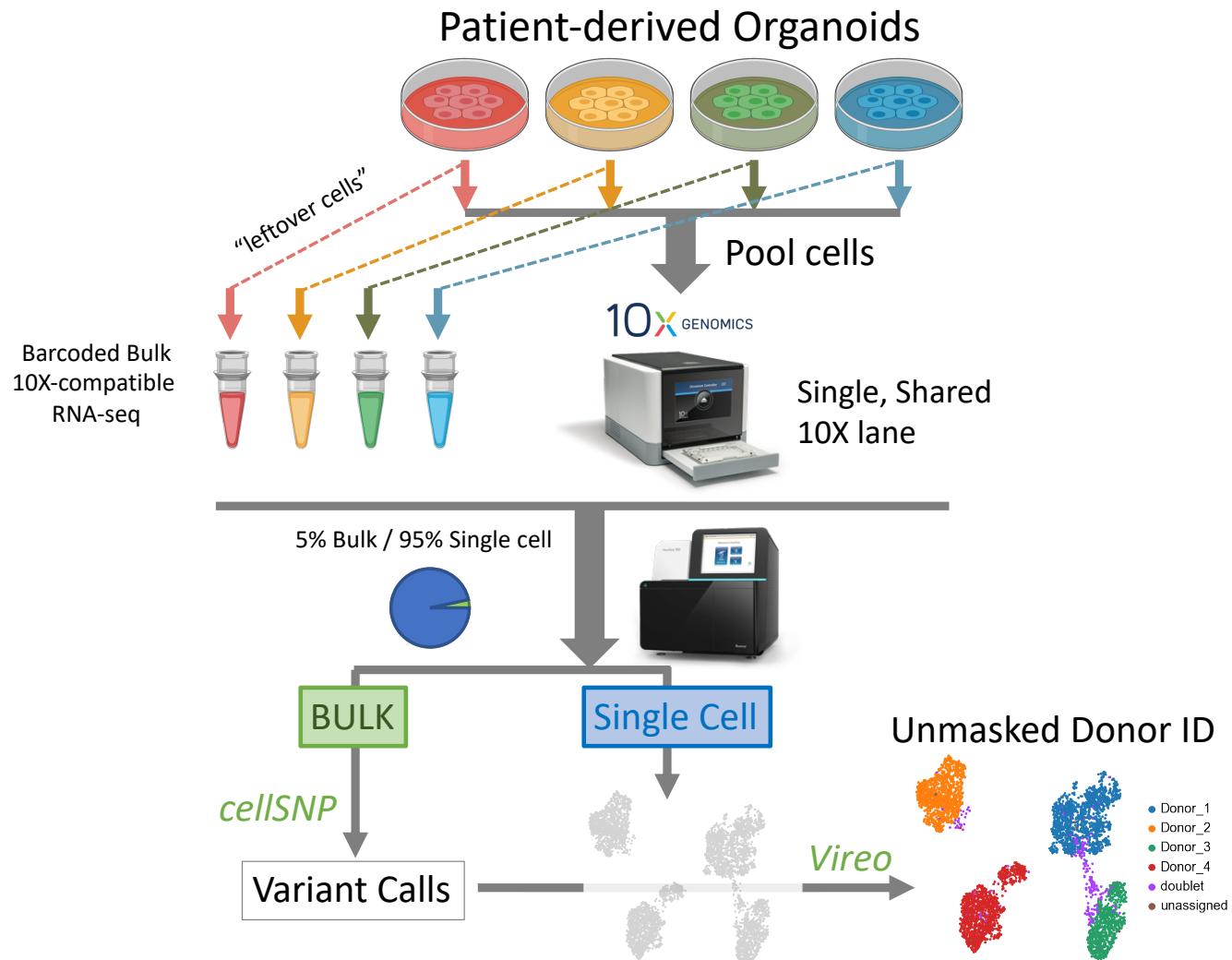
singlet	15,148
doublet	3,630

Multiplexing Using Natural Genetic Variation

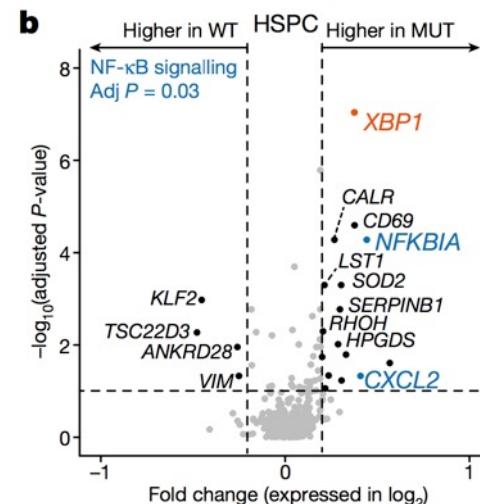
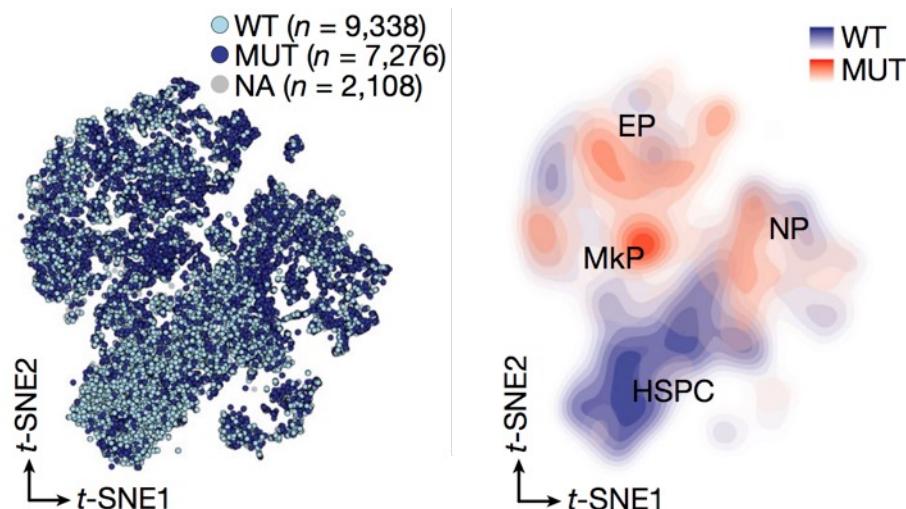
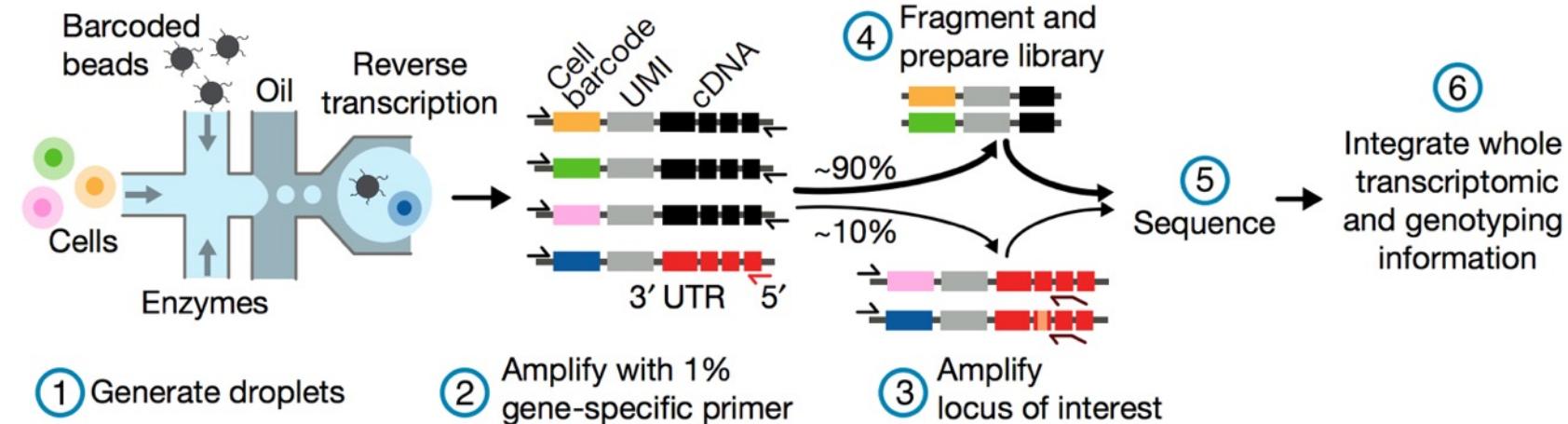
Demuxlet



Jimmie Ye lab

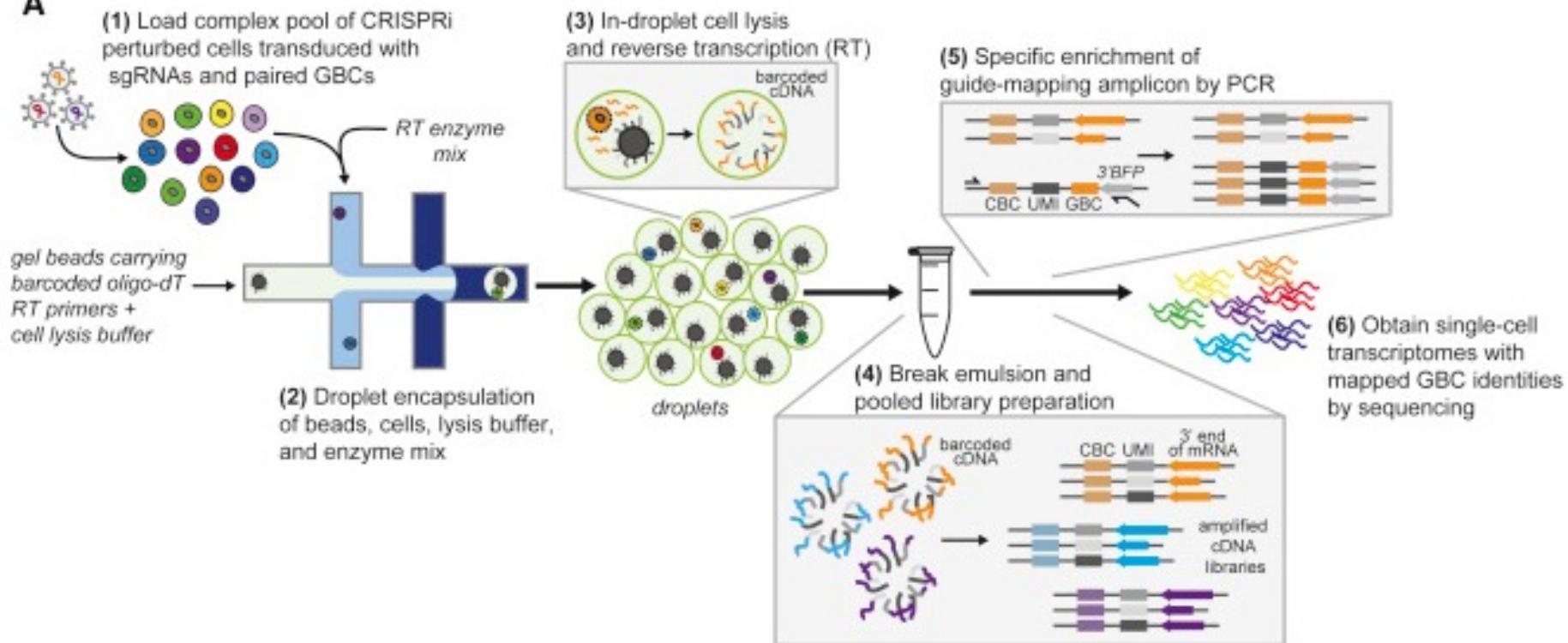


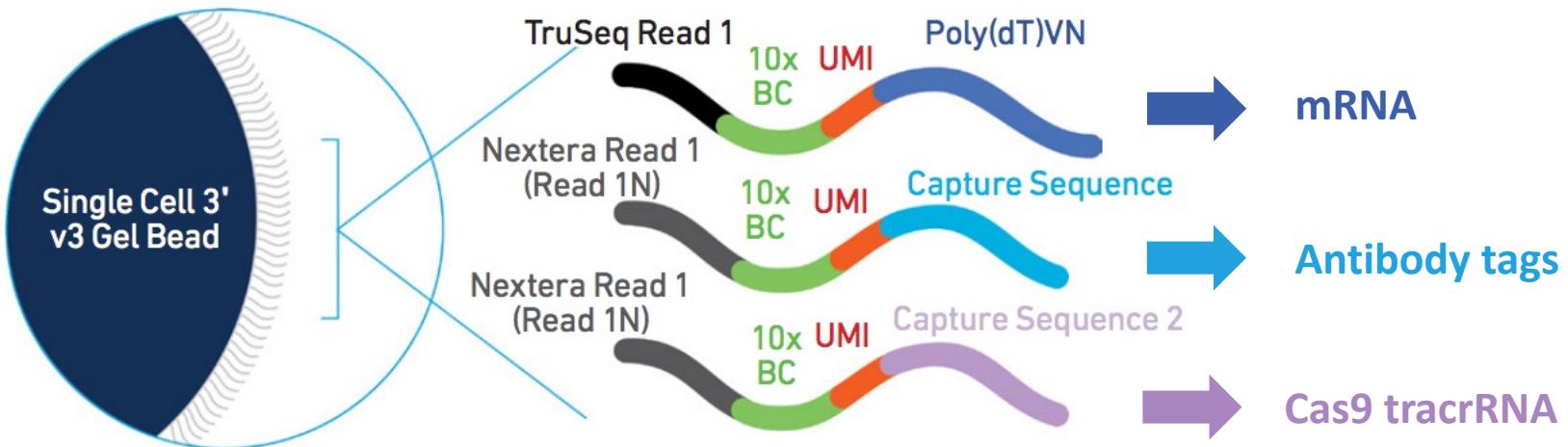
GENOTYPING OF TRANSCRIPTOMES



Perturb-Seq

A

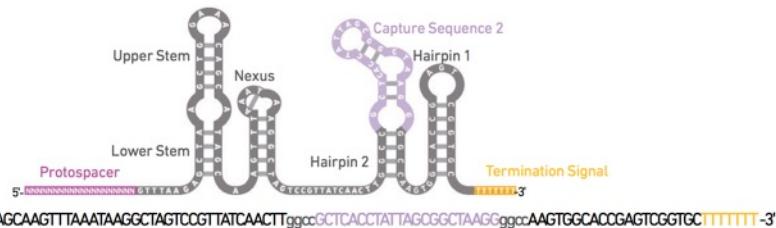




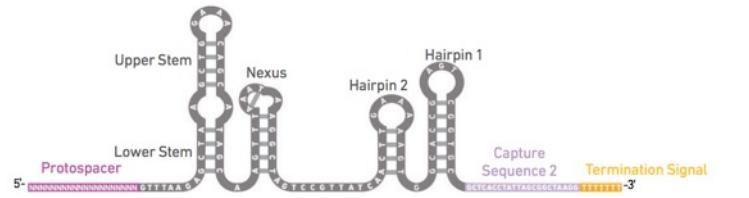
- Multiple RT primer sequences per bead
 - High efficiency capture of antibody tags, CRISPR guides

Capture Sequence 2 on Gel Bead: 5'-CCTTAGCCGCTAATAGGTGAGC-3'

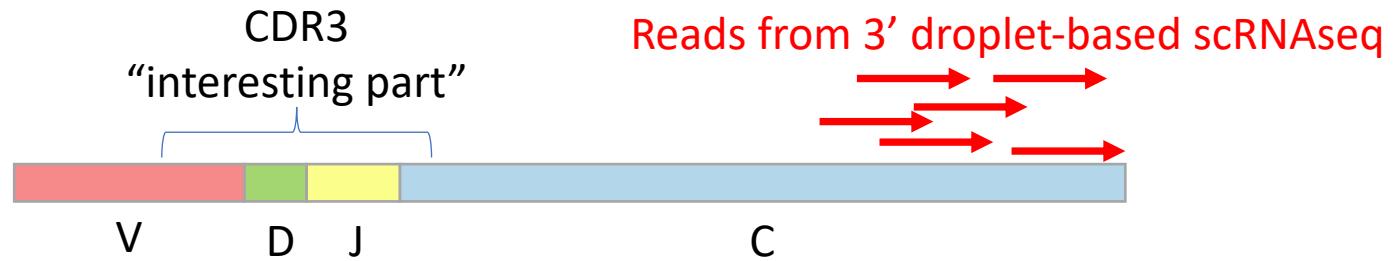
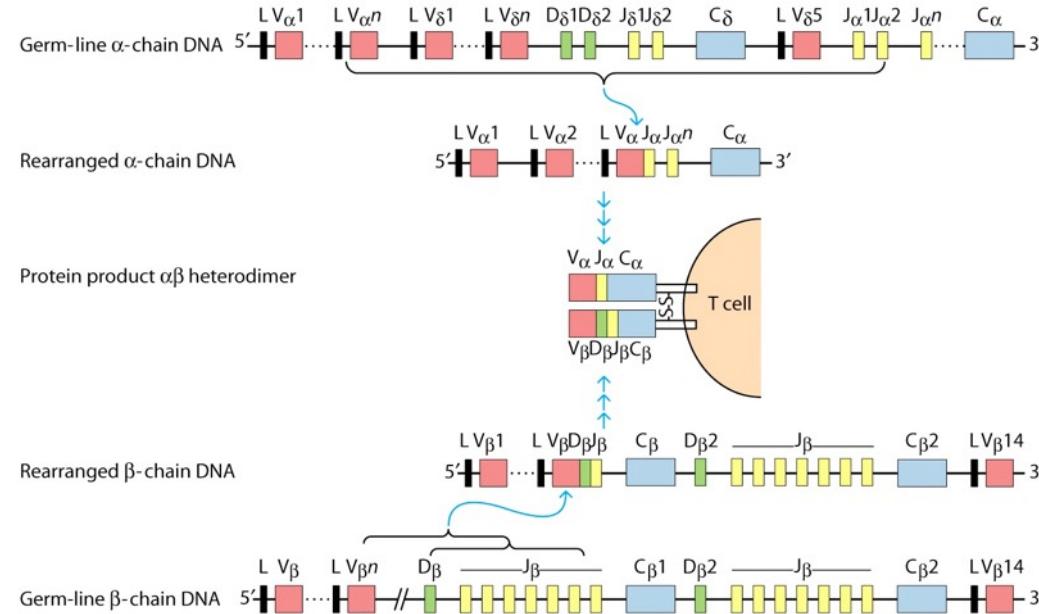
Capture Sequence 2 integrated in sgRNA hairpin



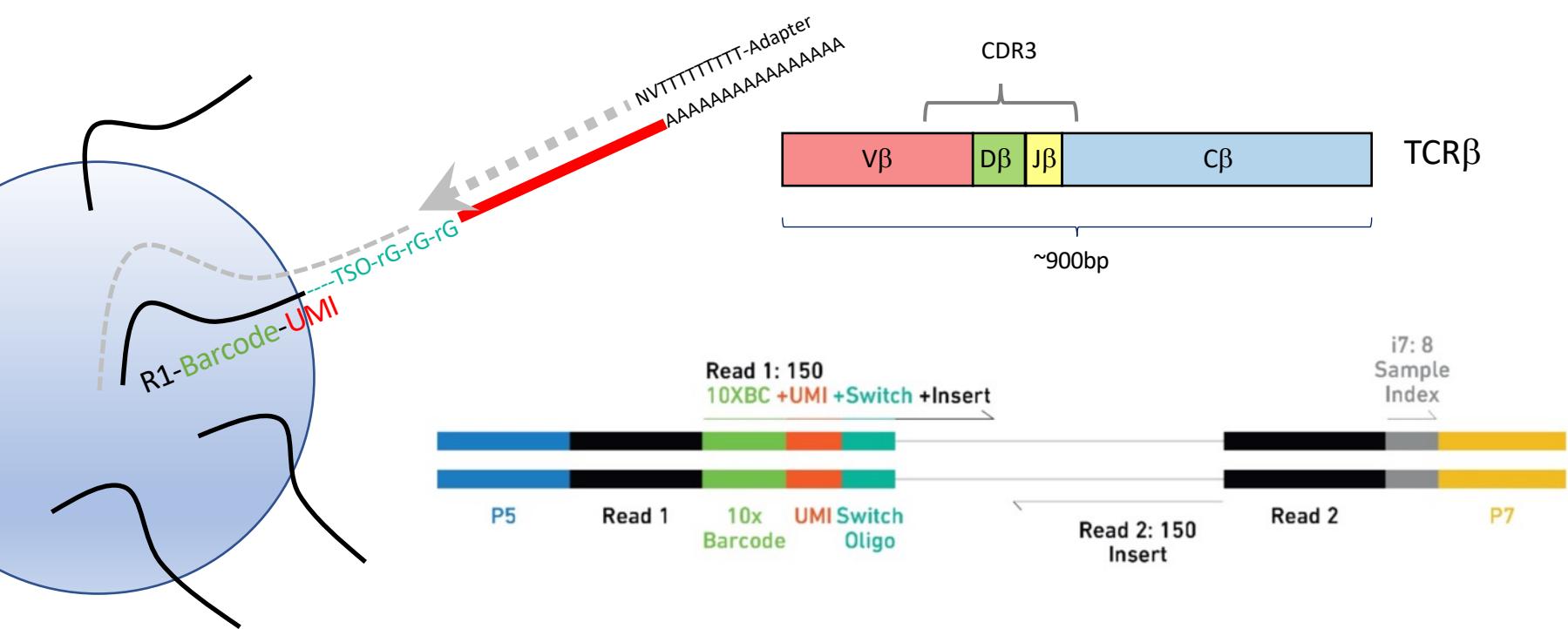
Capture Sequence 2 integrated in sgRNA 3'-end



TCR/BCR Profiling



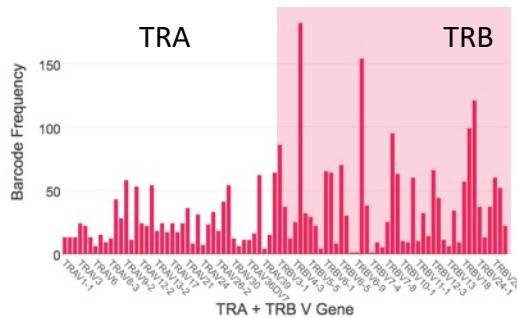
5'-Barcoded Libraries



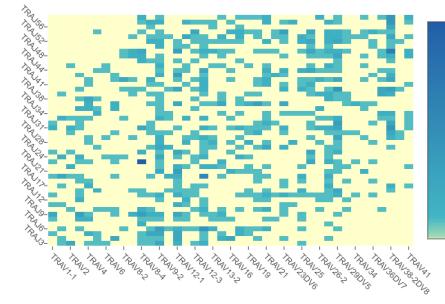
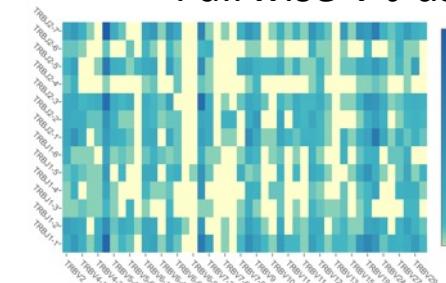
- Problem: standard transcriptome libraries have strong 3'-bias
- CDR3 mapping requires 5'-Barcoded library
- Random fragmentation to sample different 3'-ends of reads
- Require much longer reads (300bp) at a depth of 5,000X / cell

10X VDJ output example

Individual V / J usage



Pairwise V-J usage



Independent
T-cells

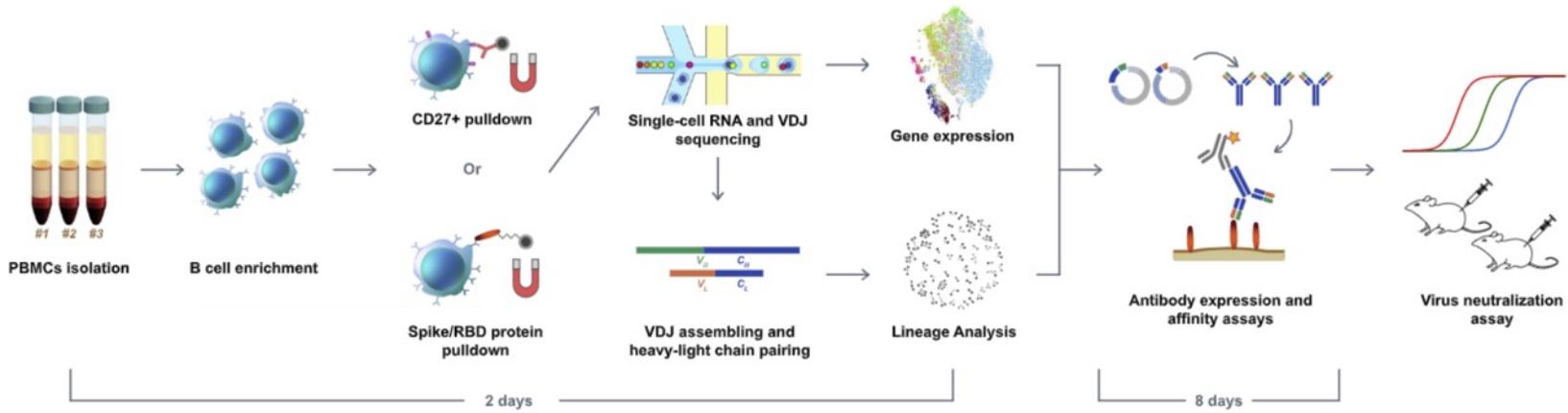
	5'UTR	TRBV7-2		TRBJ1-6	TRBC1
Reference					
Consensus					
CTAGAGTAGGGAGTC-1					
GGGAGATGTGTTGGGA-1					
GTGTTAGCATTTGTCGA-1					

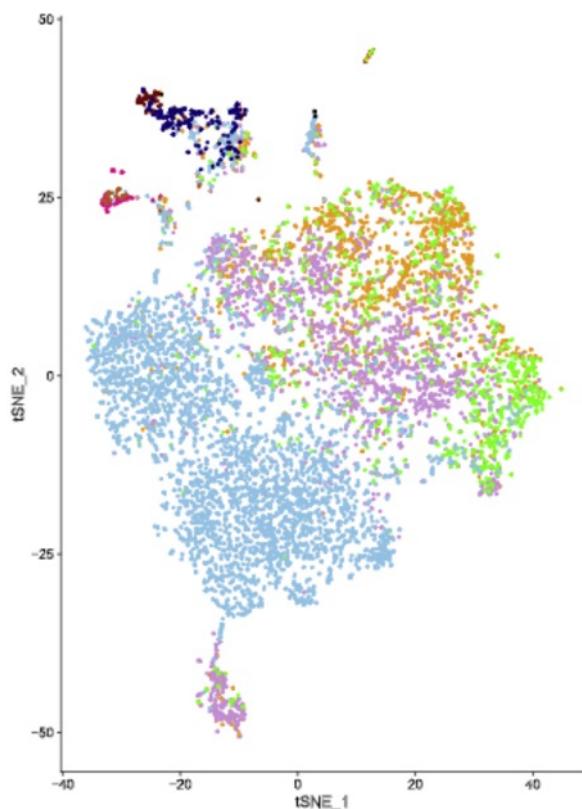
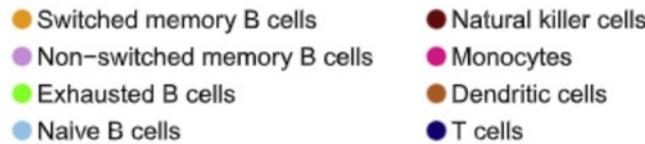
Example Rearranged TCR beta chain

CDR3 AA: CASRRGGGKTYEQYF
NT: TGTGCCAGCCGCCGGGGCGGGGGAAAACCTACGAGCAGTACTTC

Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells

A





Rapid Filter for Neutralizing antibody candidates:

VDJ sequencing:

1. Select only IgG1 isotypes
2. Clones with multiple observed cells
3. Clones with somatic hypermutation

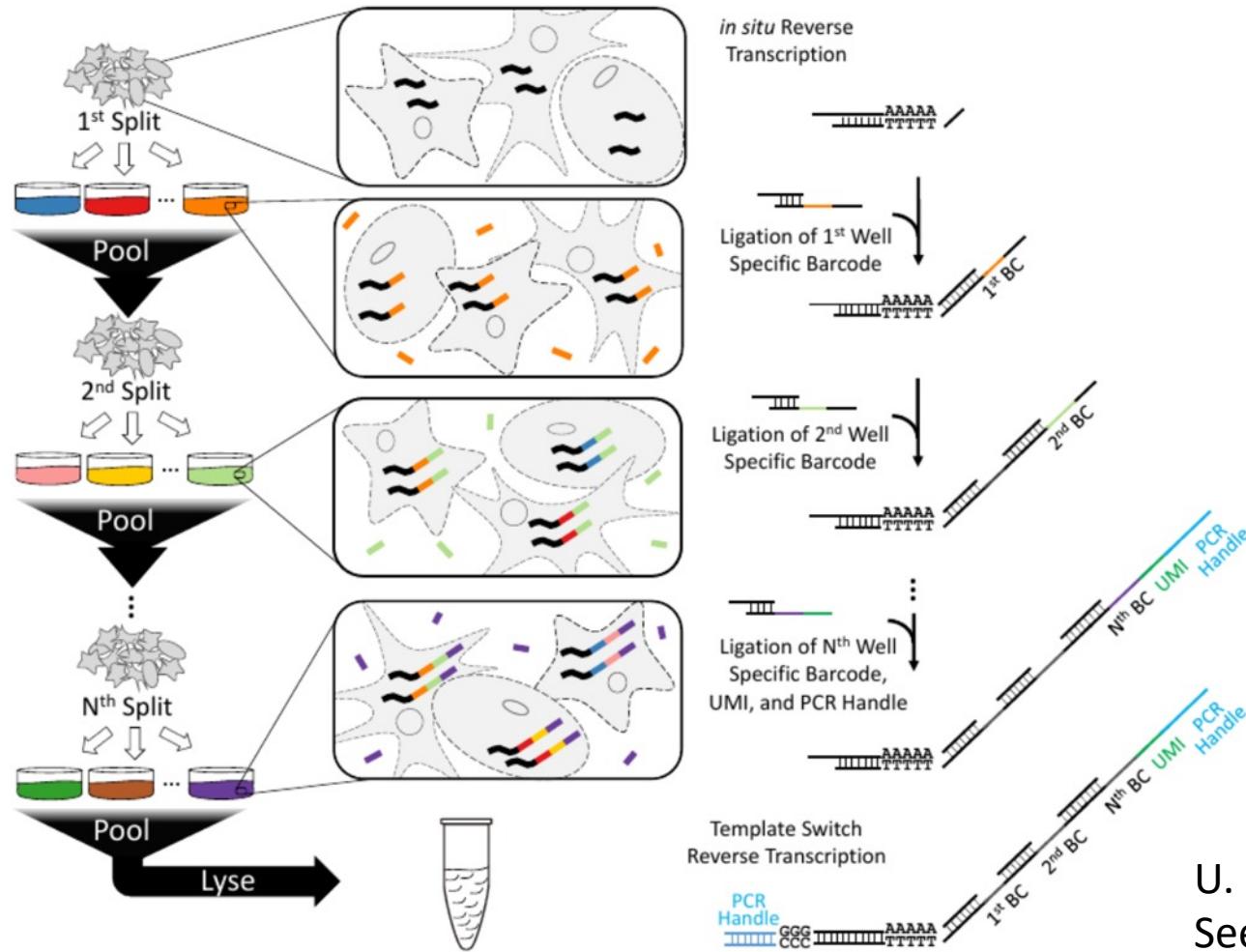
Gene expression analysis

1. Exclude exhausted and naïve phenotypes
2. Favor memory and plasma phenotypes

Other high-throughput platforms

Combinatorial Indexing

Split-Seq



U. Washington
Seelig Lab



Evercode™ Whole Transcriptome

WT Mini

<10k cells

1-12 samples

WT

10k-100k cells

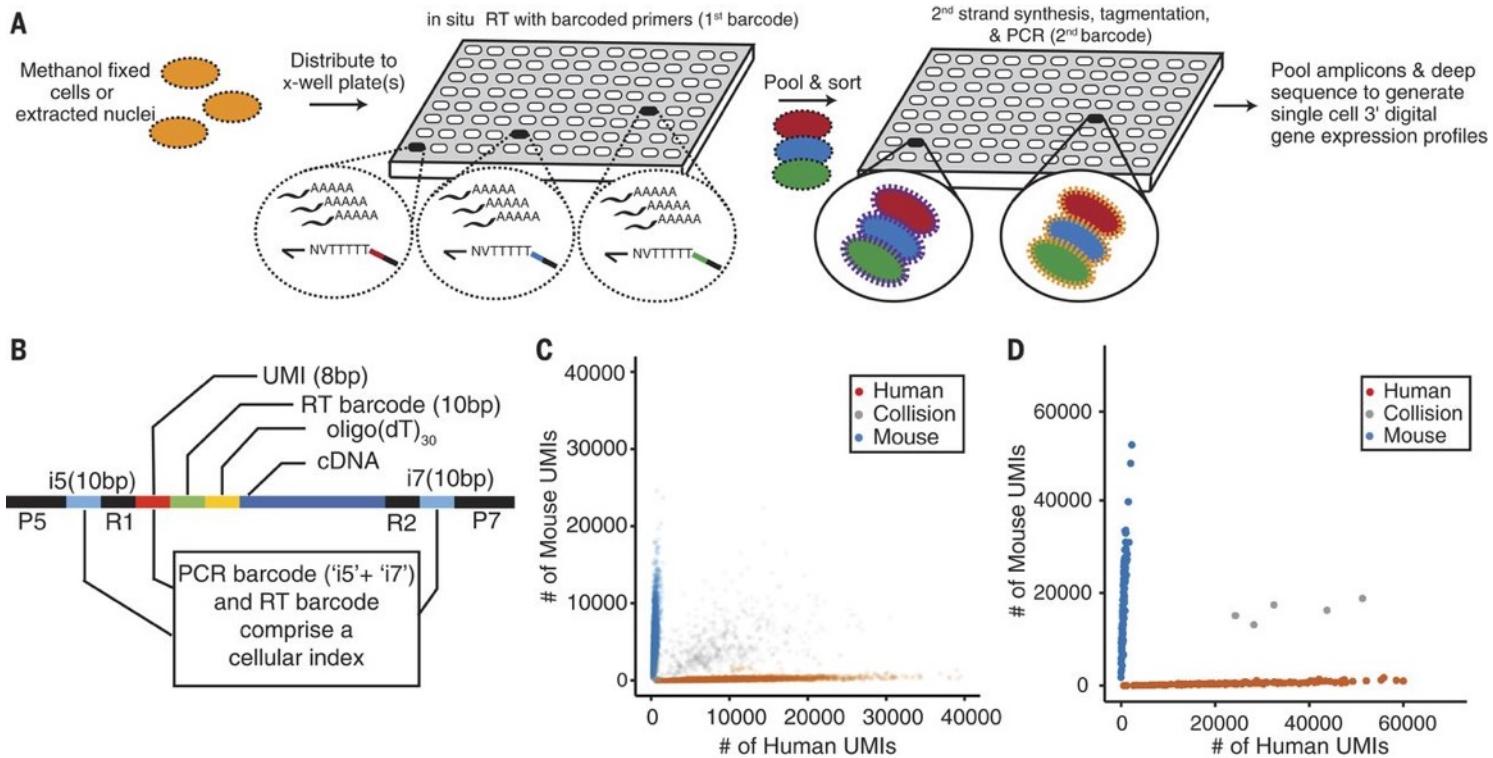
1-48 samples

WT Mega

100k-1M cells

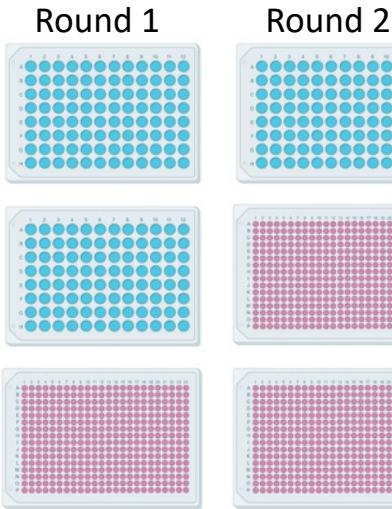
1-96 samples

Combinatorial Indexing sci-Seq



U. Washington
Shendure Lab

Combinatorial Scaling

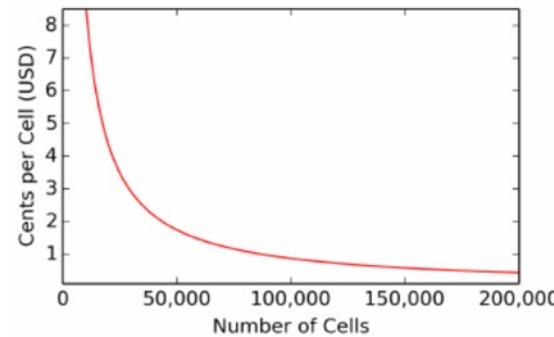
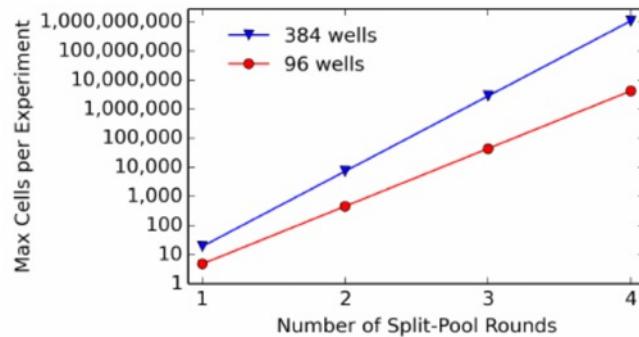


$$96 \times 96 = 9,216 \div 10 = \sim 921 \text{ cells}$$

$$96 \times 384 = 36,864 \div 10 = \sim 3,684 \text{ cells}$$

$$384 \times 384 = 147,456 \div 10 = 14,746 \text{ cells}$$

- To avoid random sampling of same barcode combinations, use ~10% of total theoretical combinations as input



Enormously scalable
Can achieve <\$0.01 per cell

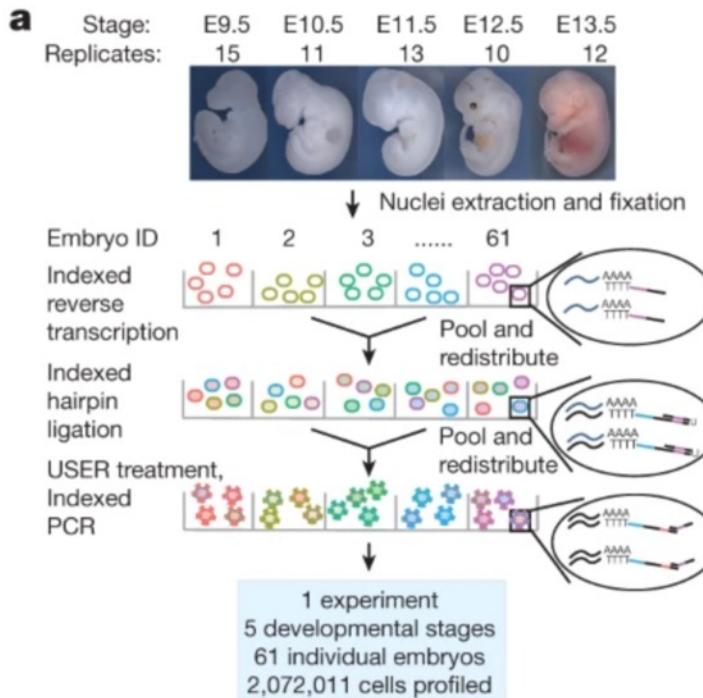
Labor intensive
Significant 'boot-up' cost
Significant validation cost
Who can afford that much sequencing, anyway?

The single-cell transcriptional landscape of mammalian organogenesis

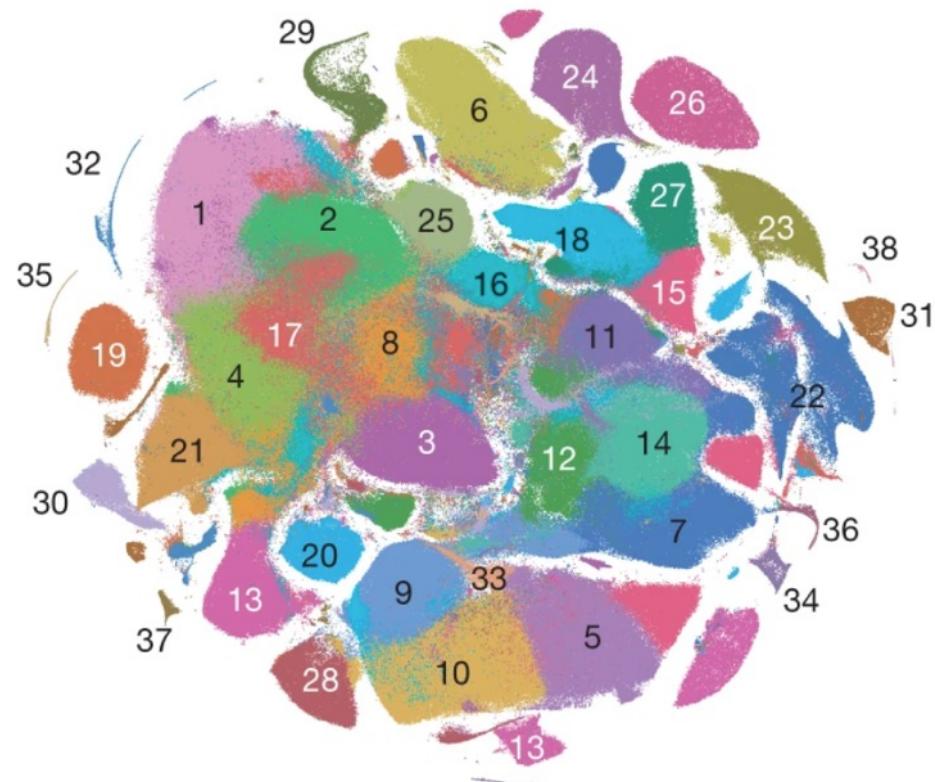
Junyue Cao, Malte Spielmann, Xiaojie Qiu, Xingfan Huang, Daniel M. Ibrahim, Andrew J. Hill, Fan Zhang,
Stefan Mundlos, Lena Christiansen, Frank J. Steemers, Cole Trapnell & Jay Shendure

Nature 566, 496–502(2019) | Cite this article

sci-RNA-seq3



Mouse embryonic development



2,058,652 single-cell transcriptomes

A human cell atlas of fetal gene expression

Junyue Cao^{1,*}, Diana R. O'Day², Hannah A. Pliner³, Paul D. Kingsley⁴, Mei Deng², Riza M. Daza¹, Michael A. Zager^{3,6}, Kimberly A. Aldinger^{2,5}, Ronnie Blecher¹, Fan Zhang⁷, Malte Spielmann^{8,9}, James Palis⁴, Dan Doherty^{2,3,5}, Frank J. Steemers⁷, Ian A. Glass^{2,3,5}, Cole Trapnell^{1,3,10,#}, Jay Shendure^{1,3,10,11,#}

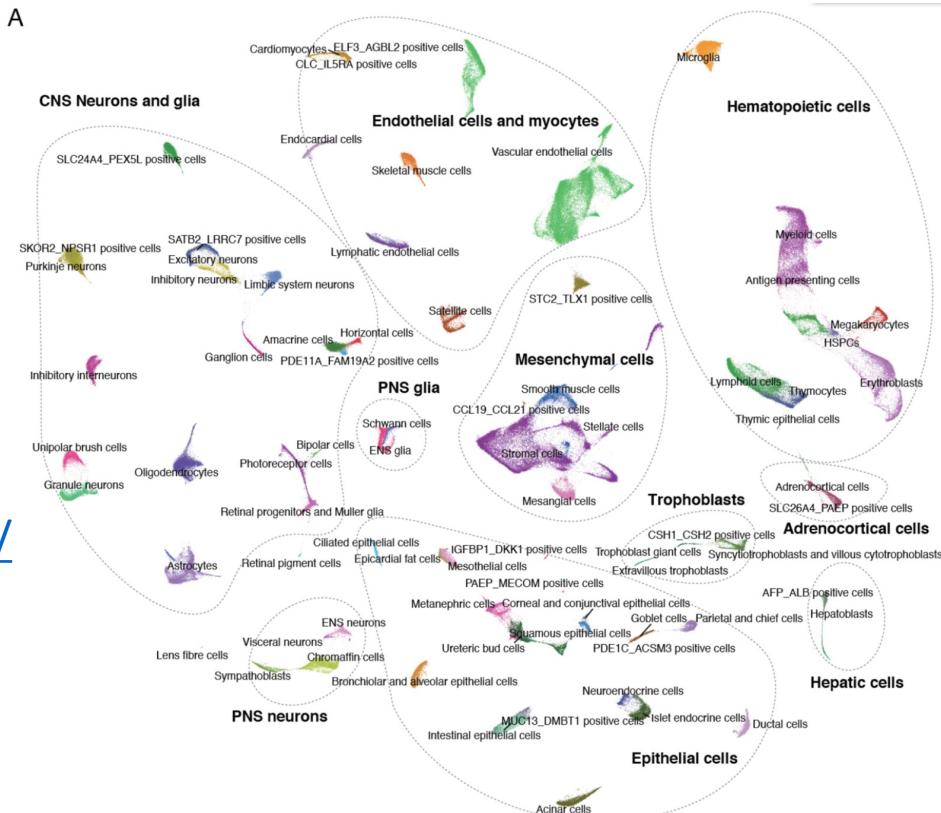
Published in final edited form as:

Science. 2020 November 13; 370(6518): . doi:10.1126/science.aba7721.

sci-RNA-seq3

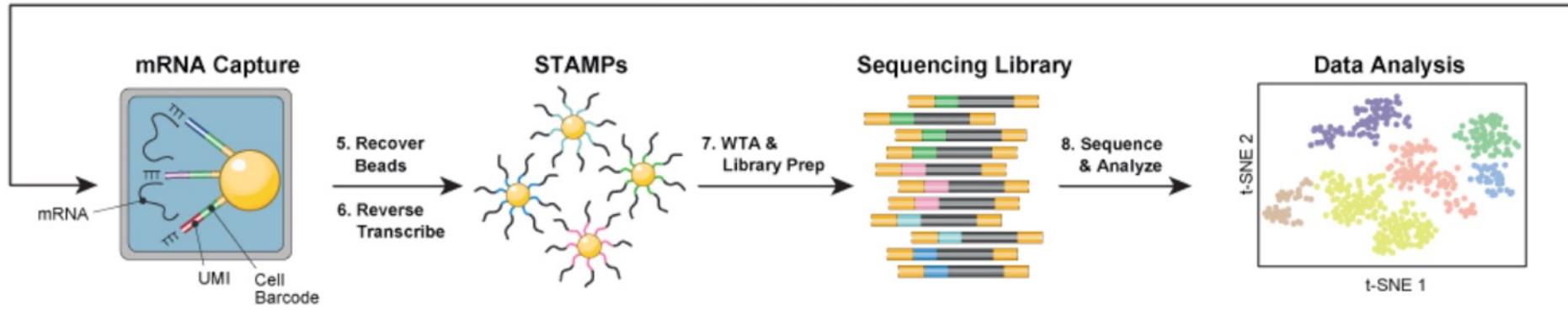
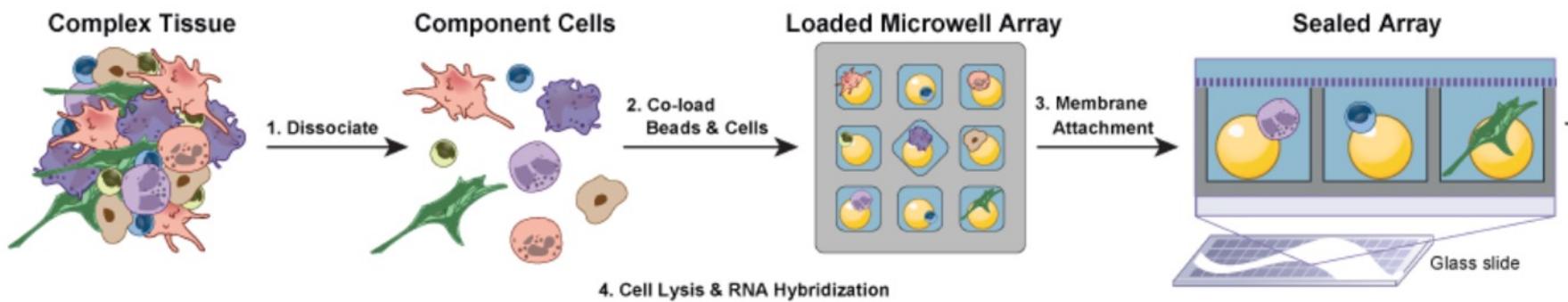
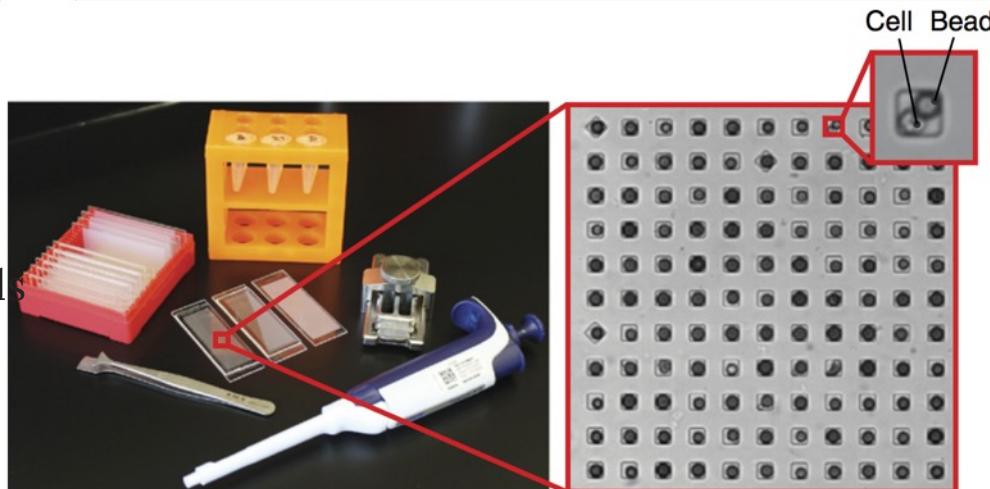
> 4M cells
15 organs
110 samples

<https://descartes.brotmanbaty.org/>



Seq-Well

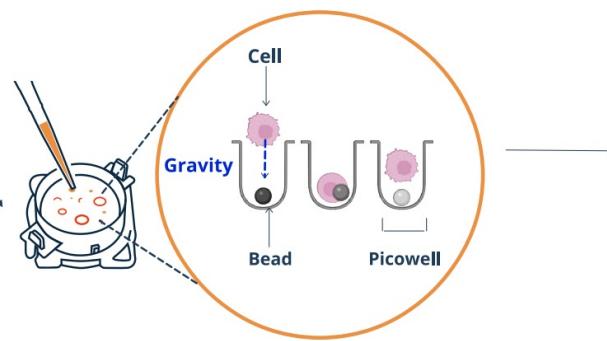
PDMS array of ~86,000 subnanoliter wells
Sized to fit 1 bead per well
Drop-Seq style barcoded beads
Sealed chamber for each cell





THE HIVE

HONEYCOMB

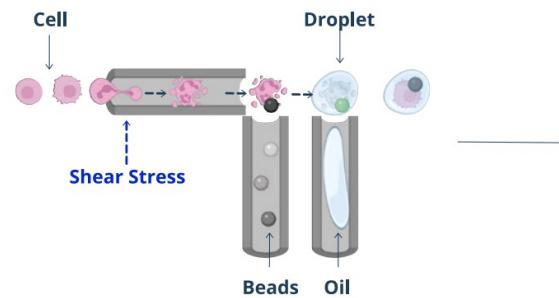


HIVE CELL CAPTURE: Gentle, by gravity or 30g spin.

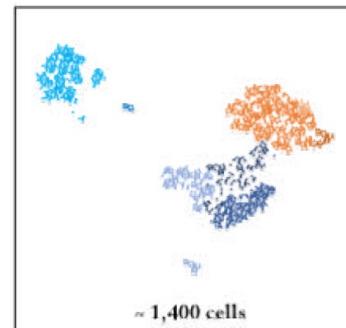
HIVE scRNASeq Method



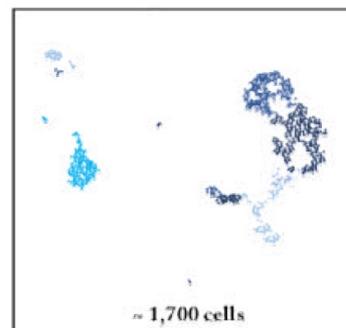
Droplet Based Method



DROPLET CELL CAPTURE: Cells subjected to shear stress.



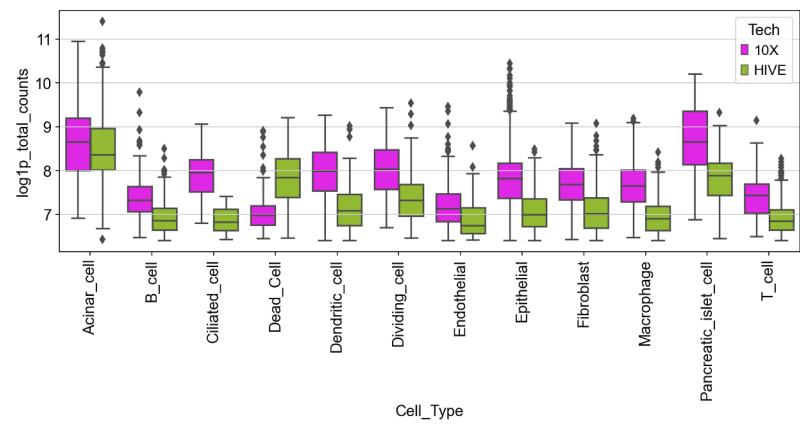
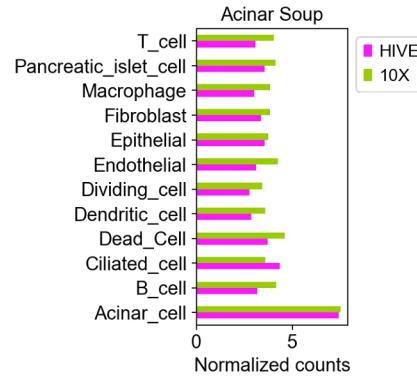
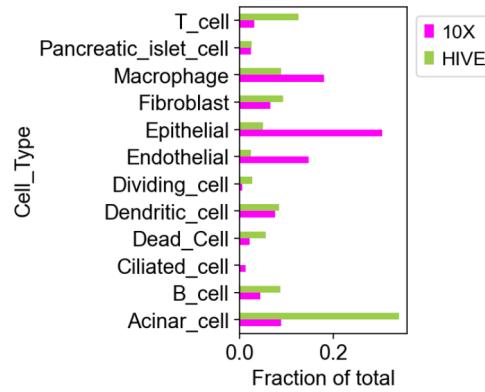
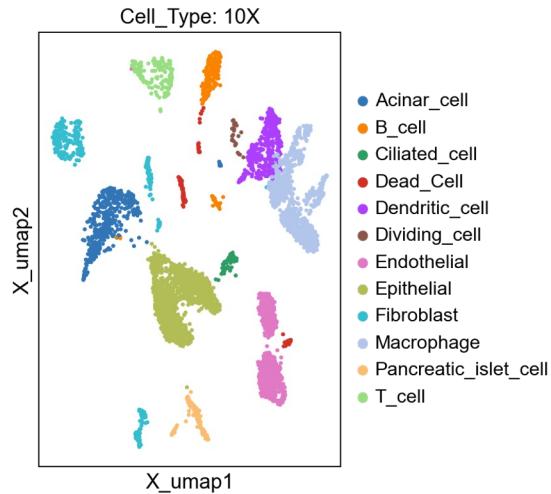
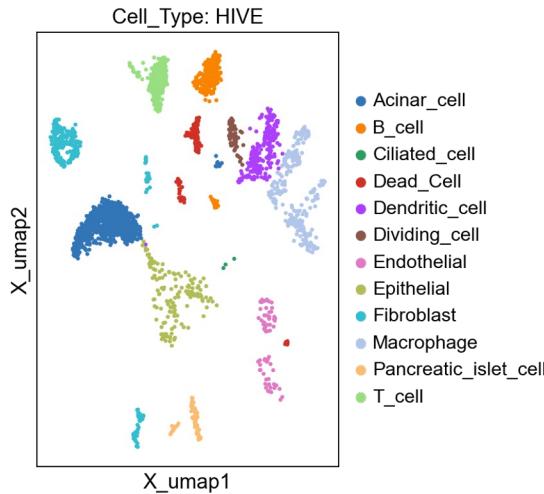
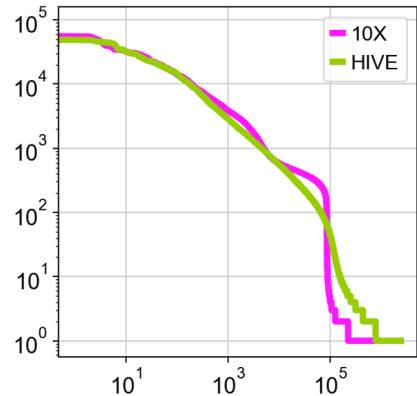
FULL BIOLOGY RECOVERED, including fragile cells like **granulocytes** (neutrophils, eosinophils, and basophils).



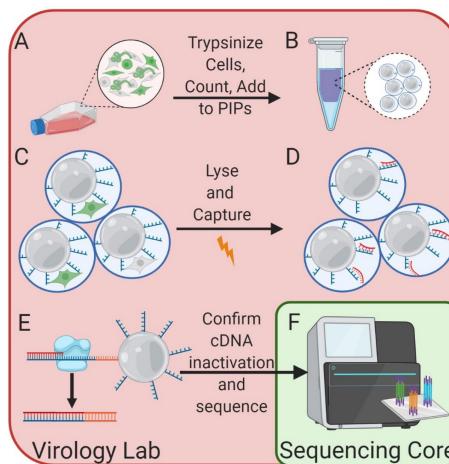
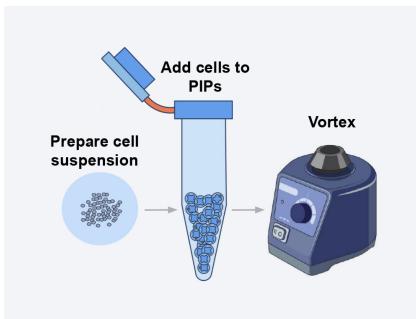
INCOMPLETE BIOLOGY RECOVERED, with fragile **granulocytes** missing.

Cold-dissociated Normal Mouse Pancreas

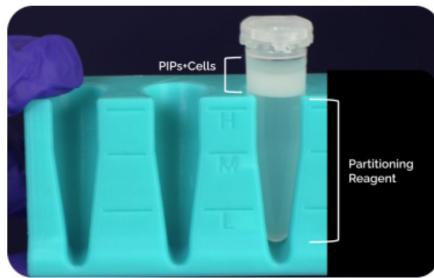
10X = 132,845,676 reads
 HIVE = 171,905,206 reads

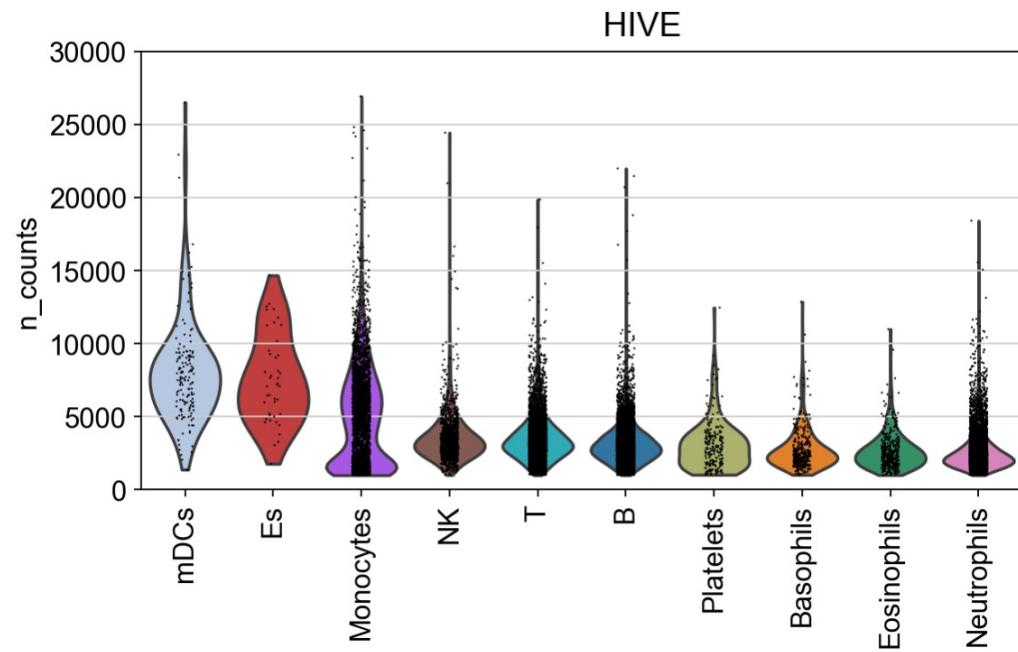
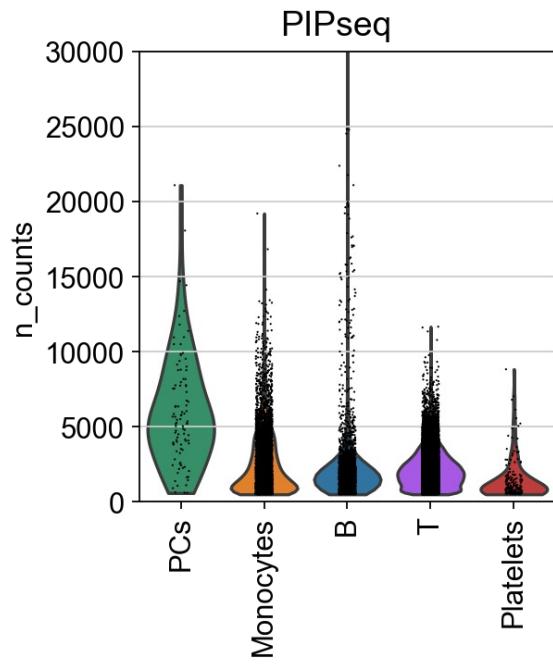
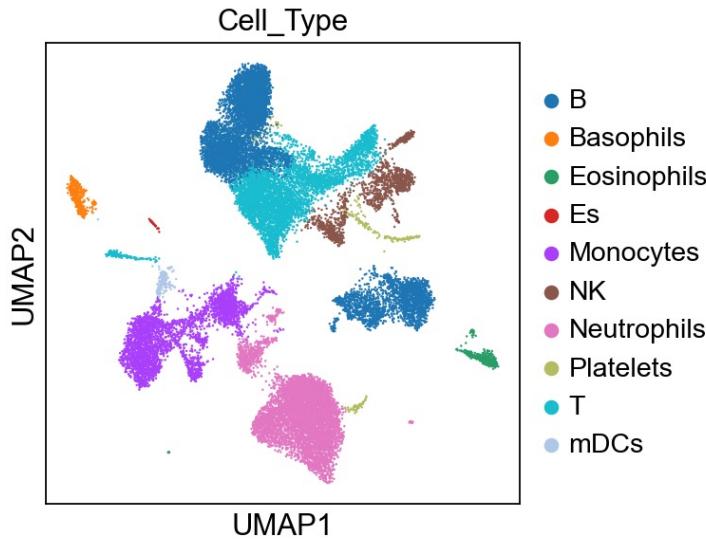
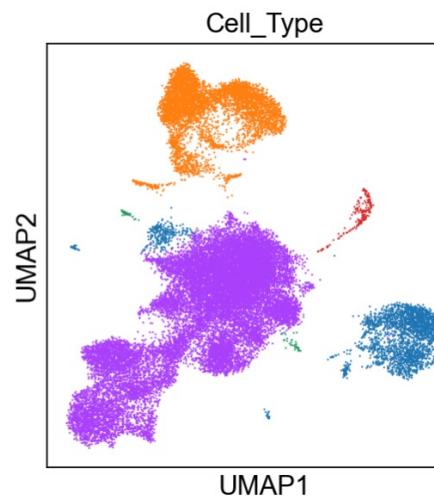


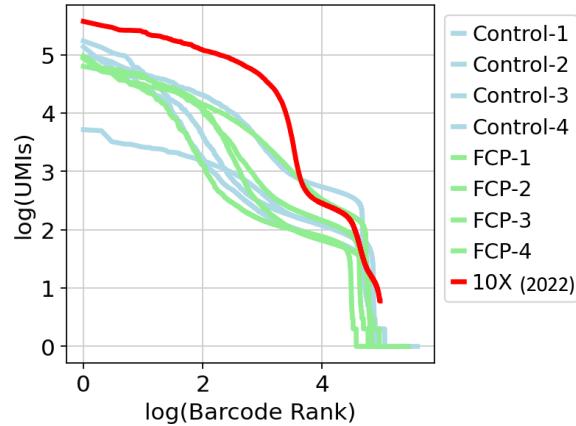
Fluent Biosciences PIPseq



- PIPseq T2:
~2,000 cells
\$300
- PIPseq T20:
~20,000 cells
\$900
- PIPseq T100:
~100,000 cells
\$??? (need quote)

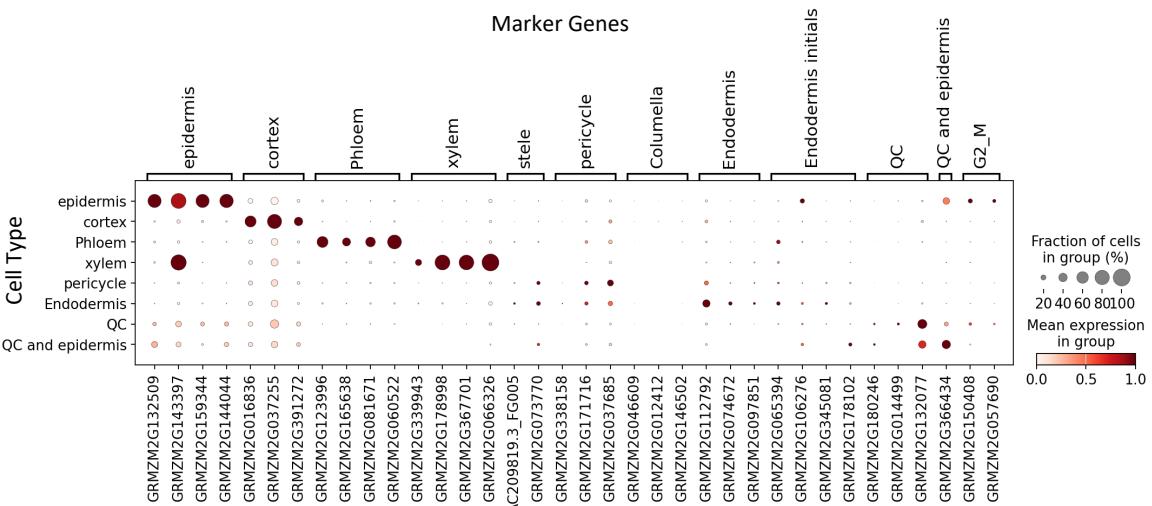
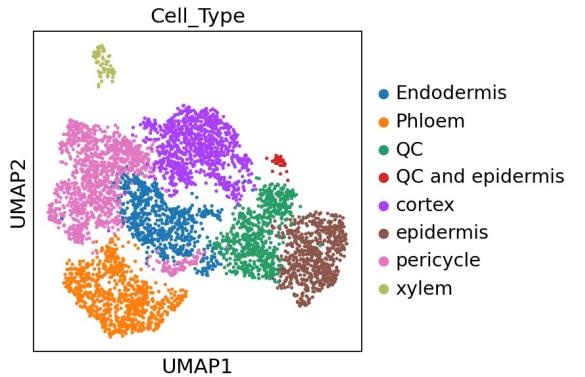
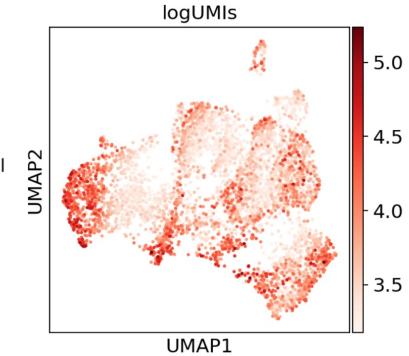
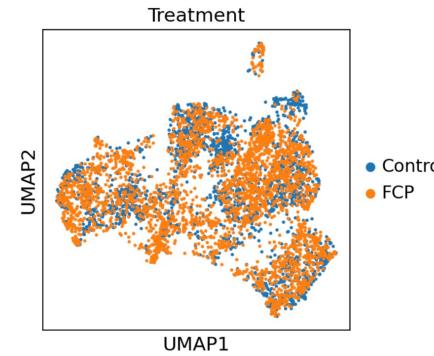






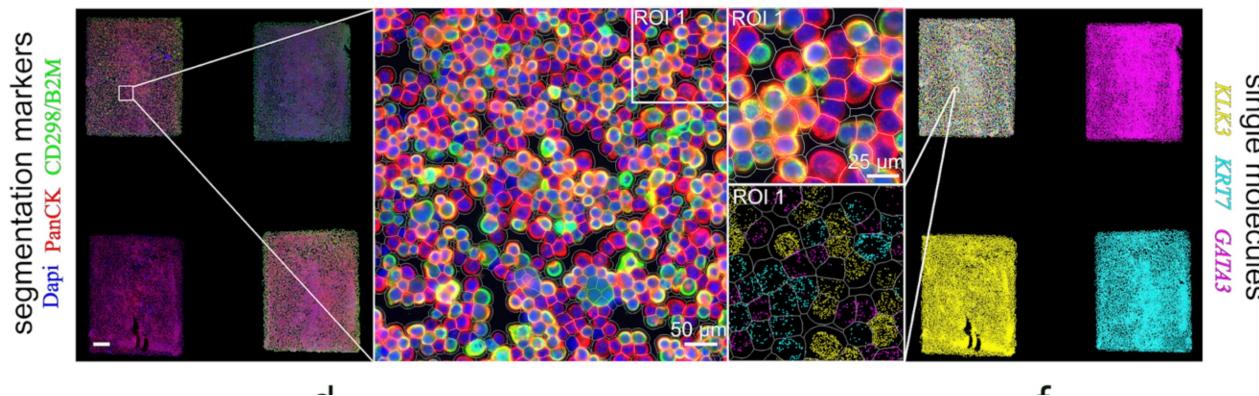
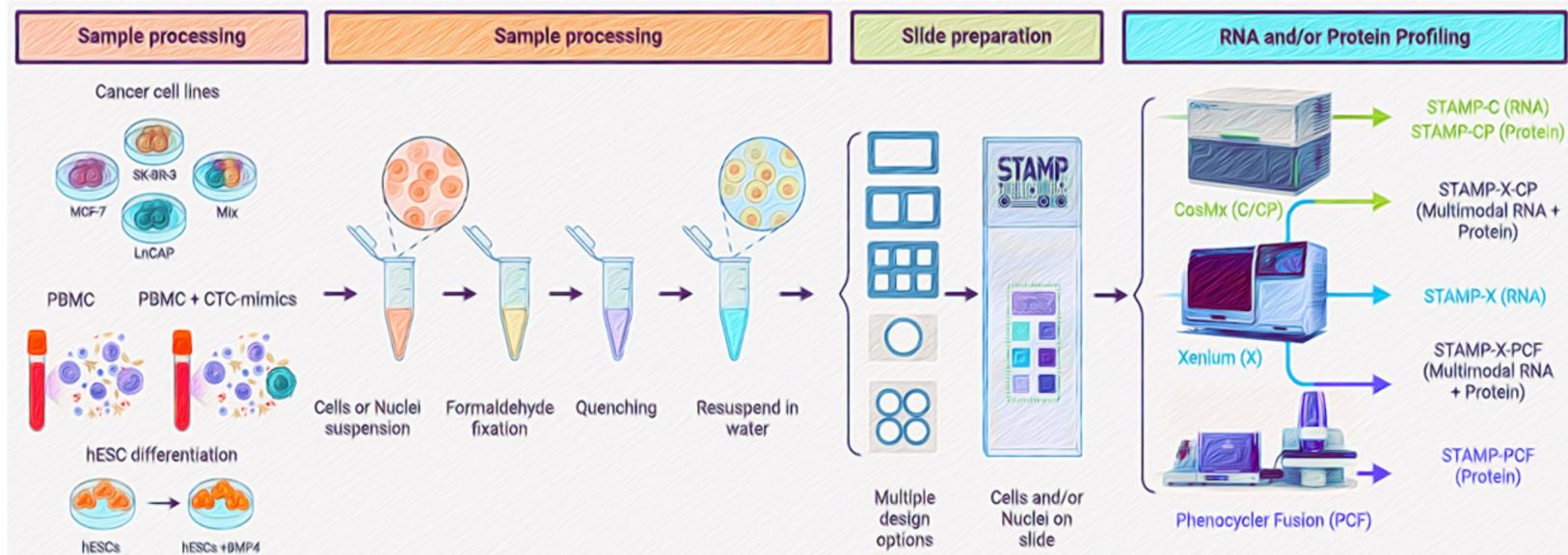
UMIs > 1,500

Sample	Filtered Cells
FCP-3	2134
Control-4	1678
FCP-1	471
FCP-4	341
Control-3	193
Control-2	107
Control-1	103
FCP-2	87



STAMP: Single-Cell Transcriptomics Analysis and Multimodal Profiling through Imaging

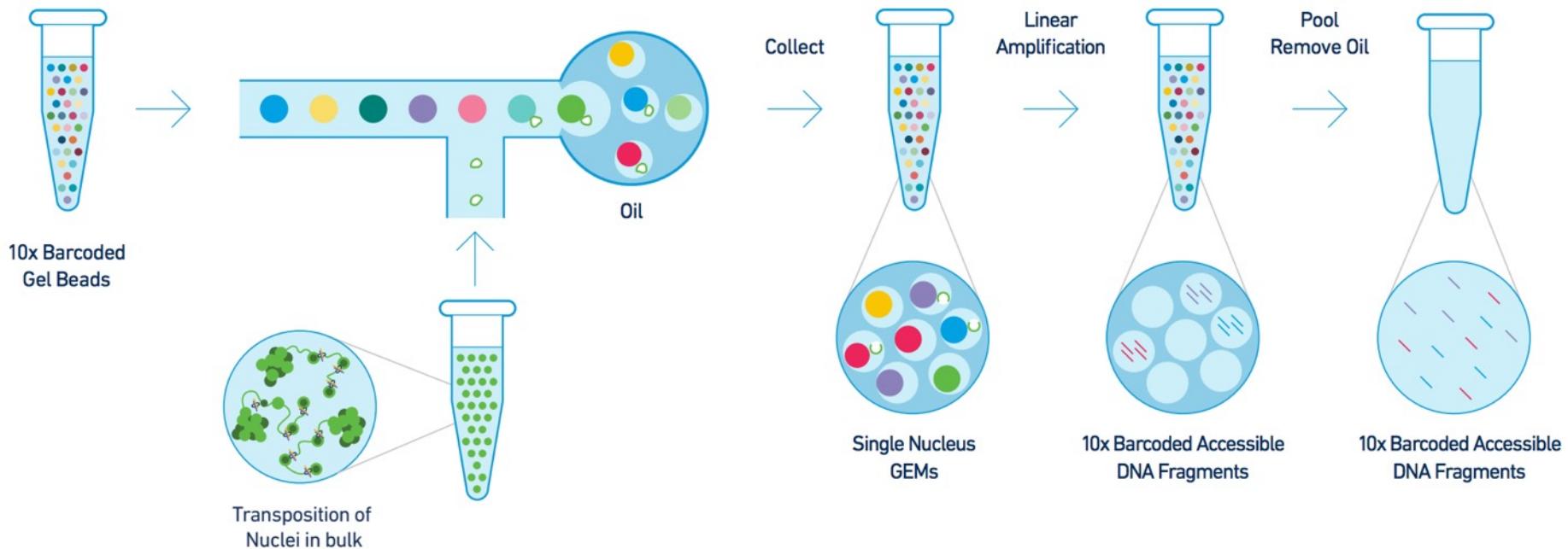
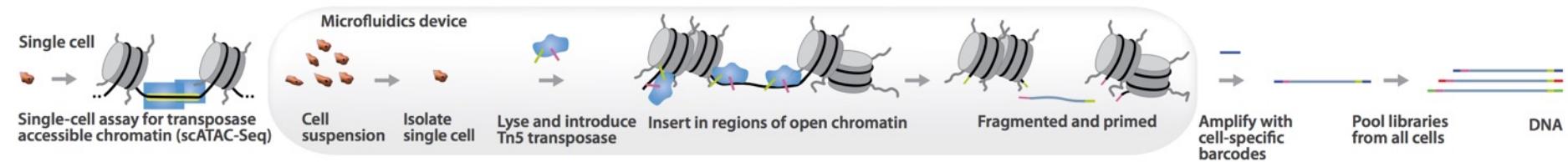
Emanuele Pitino, ... Luciano Martelotto. *bioRxiv*





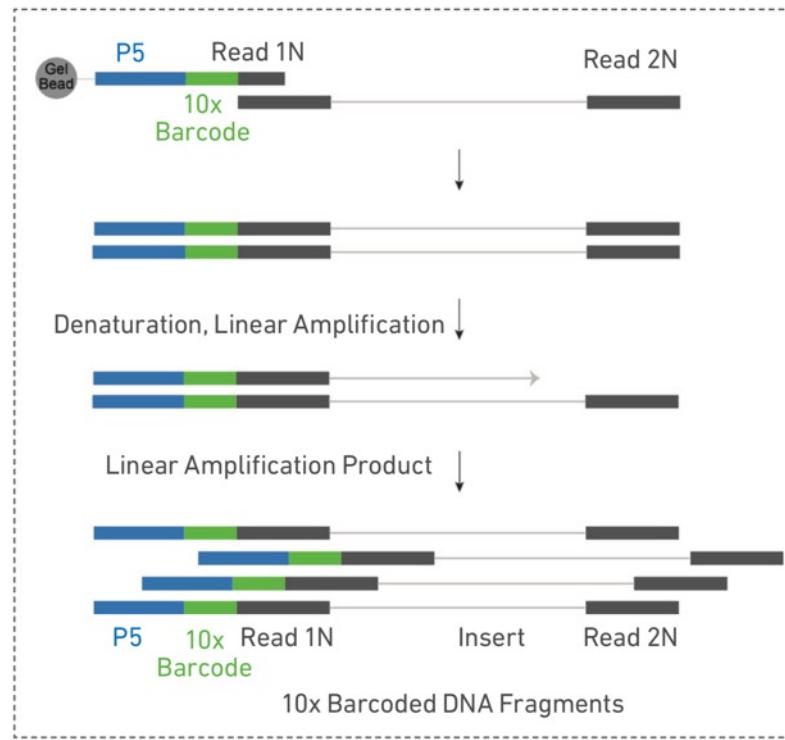
The Other
Nucleic Acid

10X Genomics Single Cell ATAC

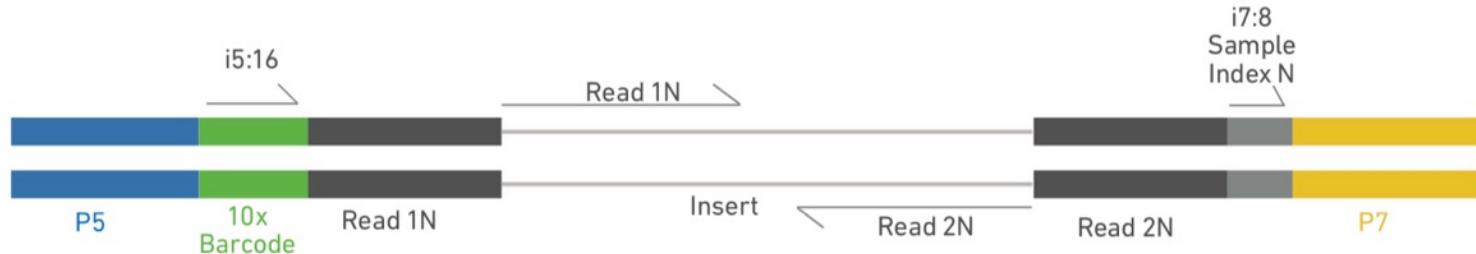


10X Genomics Single Cell ATAC

Inside Individual GEMs

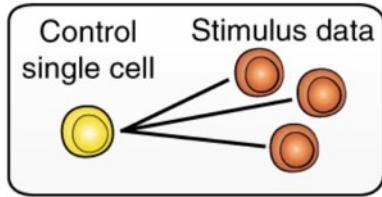


Chromium Single Cell ATAC Library



a

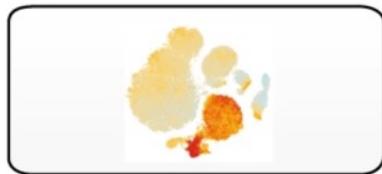
Find k -nearest neighbors ($k = 20$)



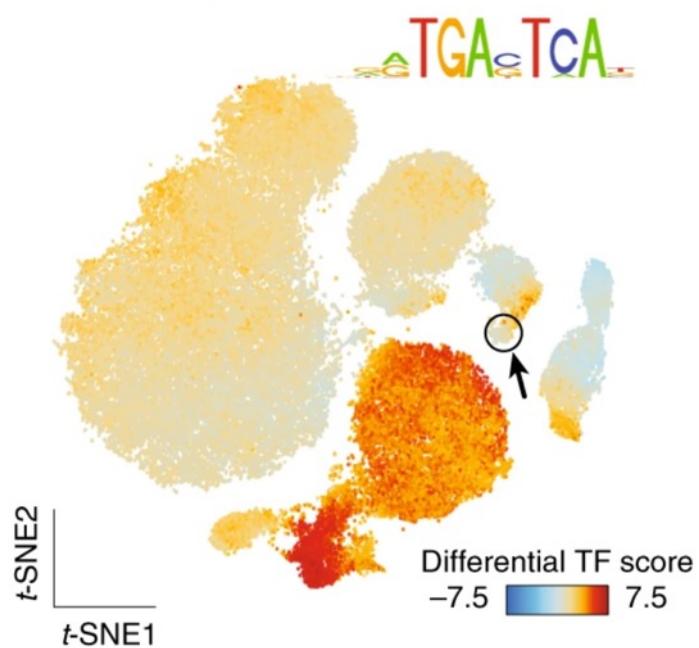
Compute differential TFs

Stimulation – control

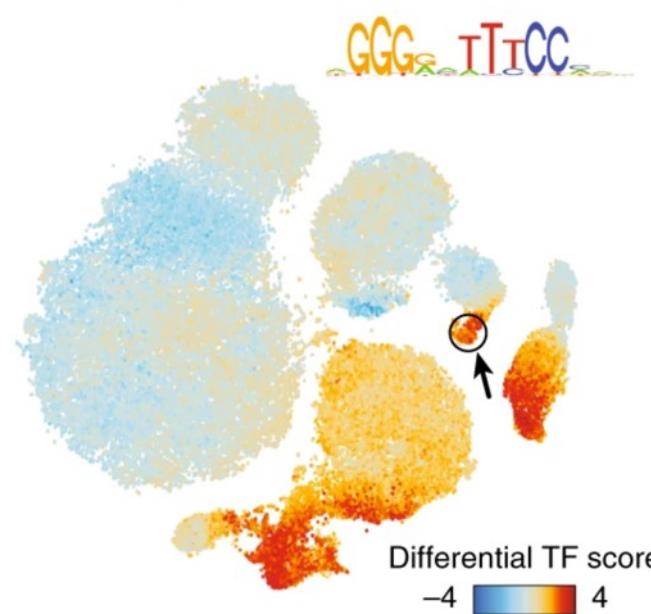
Smooth and display TFs on t-SNE

**b**

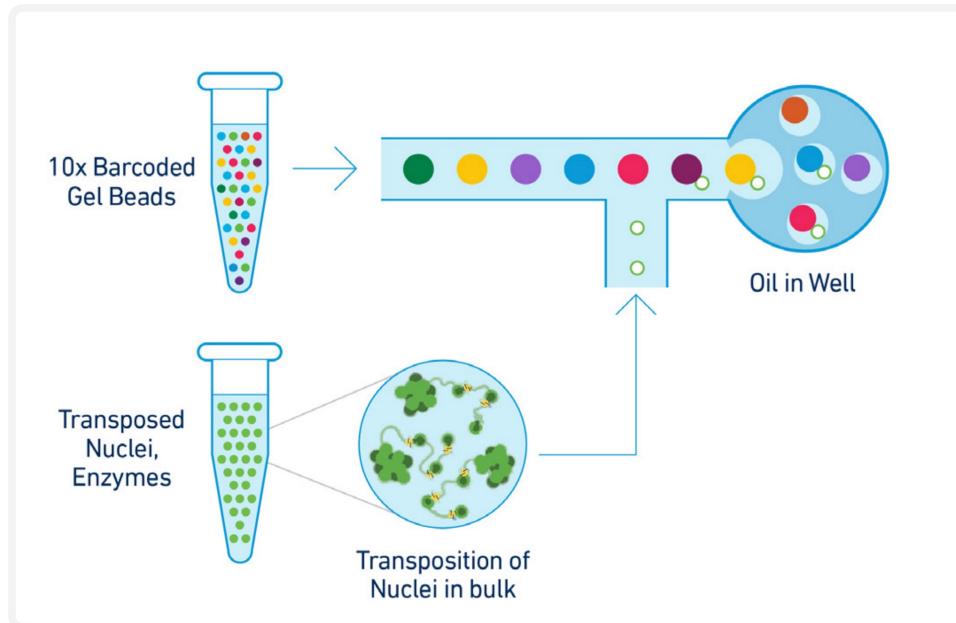
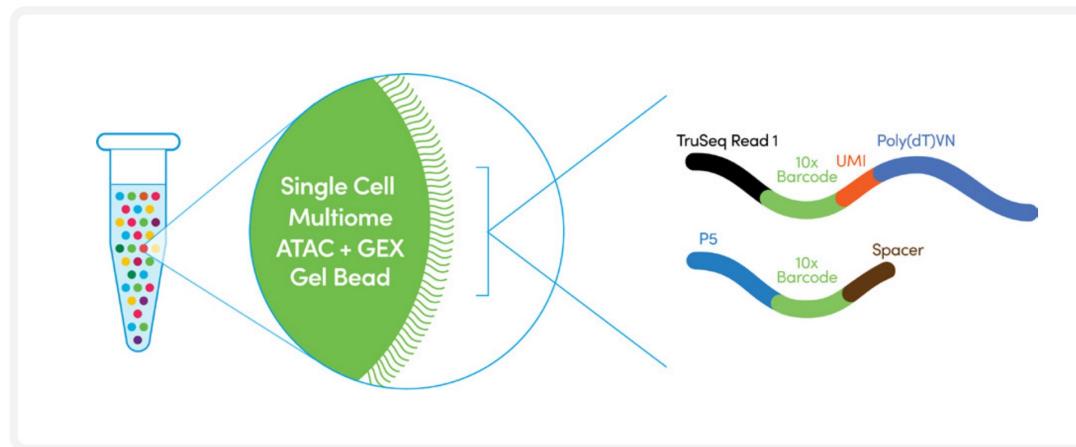
Jun TF motif
Stimulation – control

**c**

NF- κ B TF motif
Stimulation – control



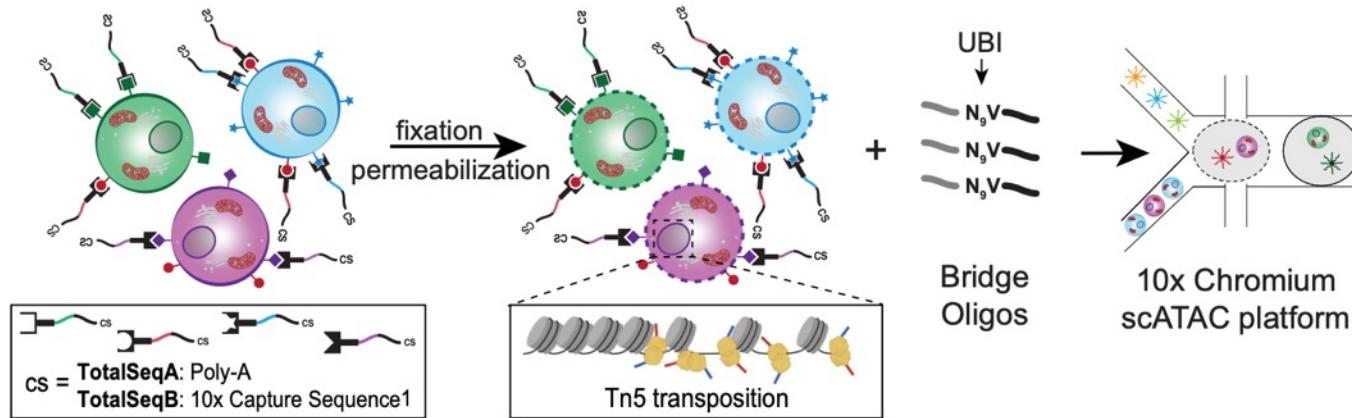
10X Genomics Multiome



ASAP-seq

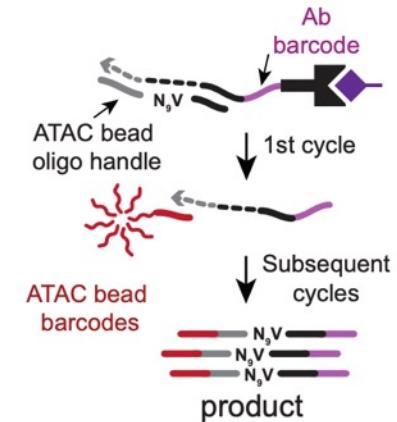
a

ASAP-seq = ATAC with Select Antigen Profiling by sequencing



b

Inside droplets



b

I. Annealing of antibody tag with BOA and extension in droplets



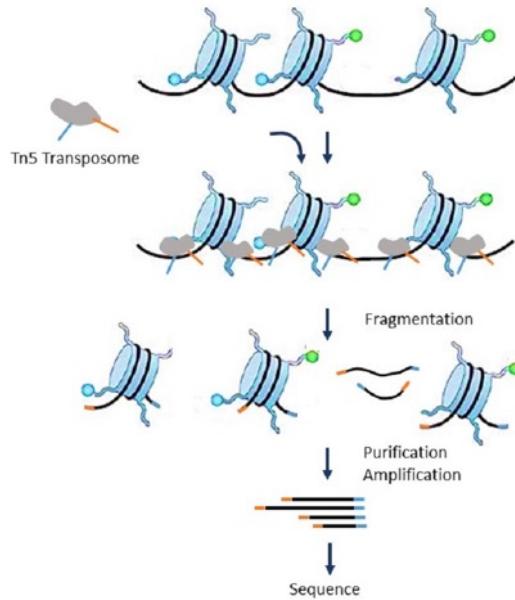
II. Annealing of extended antibody tag with barcoded oligo



III. Extension of barcoded oligo and amplification for ≤11 cycles



sci-ATAC



Round 1: Internally Barcoded Tn5 transposomes

Round 2:

Barcoded PCR primers

FANS 2,500
per well

Index 1: Transposase-based FANS 22 per well

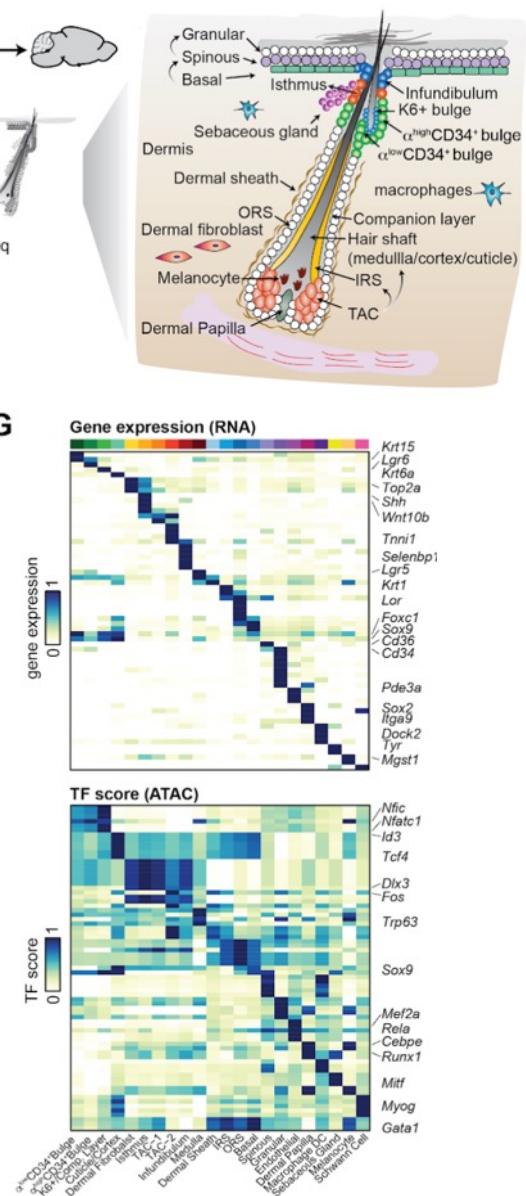
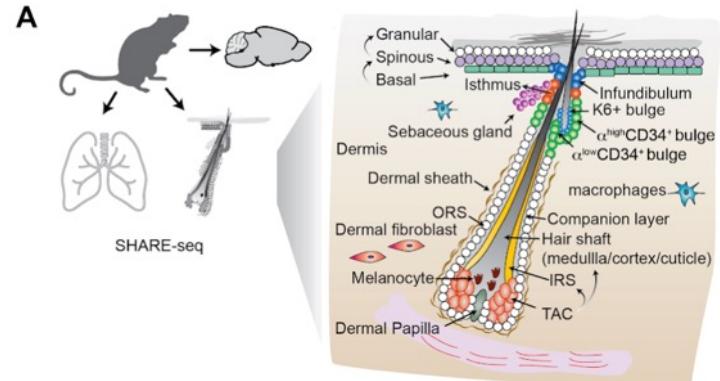
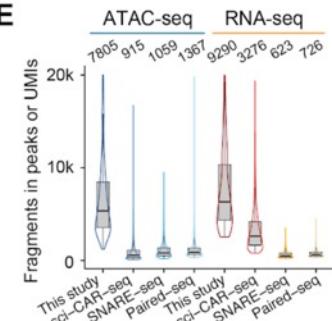
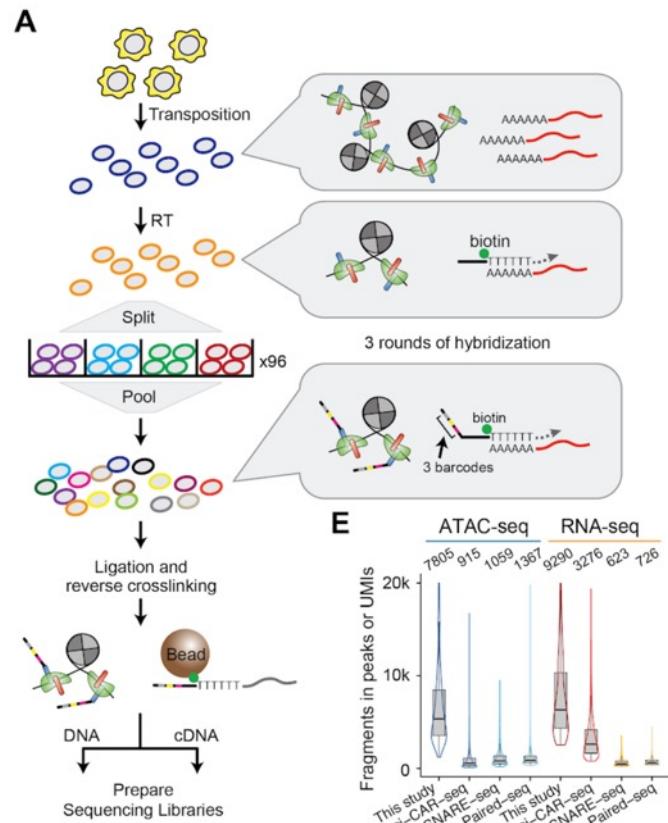
The figure consists of four separate network diagrams, each enclosed in a vertical black bar. Each diagram depicts a cluster of nodes connected by lines representing edges. The colors of the nodes and edges vary between the four clusters.

Index 2: PCR-based

5' - AATGATACCGGCACCCAGATCTACACNNNNNNNTCTGGCGCAGCGTCTCCACGNNNNNNNGCATGAGGACGCGAGATGTATAAGAGACAGXXXXXX...XXXXXXCTGTCCTTACACATCTGGGGCGGAGACGGTGNNNNNNNGGACAGGGACAGCGCAGGCCACGAGACNNNNNNNNTCTGTATGCCCTCTCTGCTG-3'
 3' - TTACTATGCCGTGTCCTAGATGTGNNNNNNNAGACGCCGTCGAGCGGTGCGNNNNNNNGCTAGCTCTGCCGTACACATTTCTGTGXXXXXX...XXXXXXGACAGAGATATGTGTAGACTCCGGCTCTGCCACNNNNNNNCTGTCCTGCGGTGCTCTNNNNNNNTAGAGCATACGGCAGAACGAAAC-5'
 Illumina P5 i5 s5 8 bp ME gDNA ME 8 bp s7 i7 Illumina P7
 Tn5 barcode Tn5 barcode

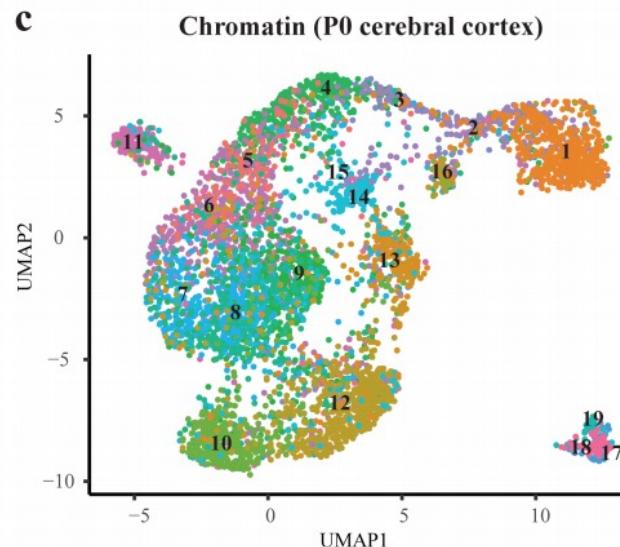
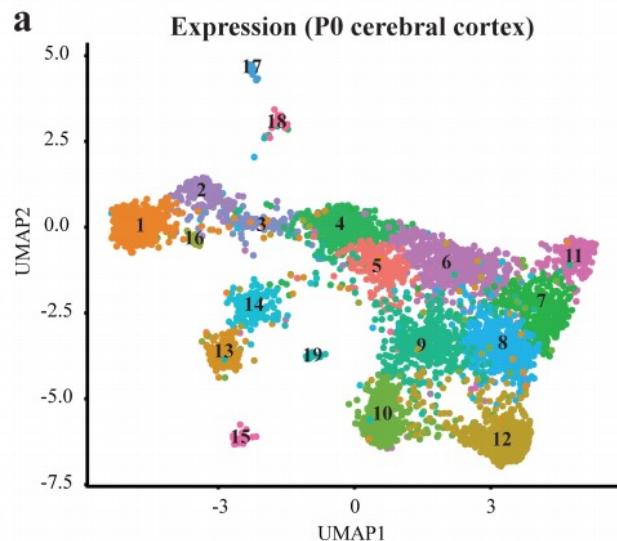
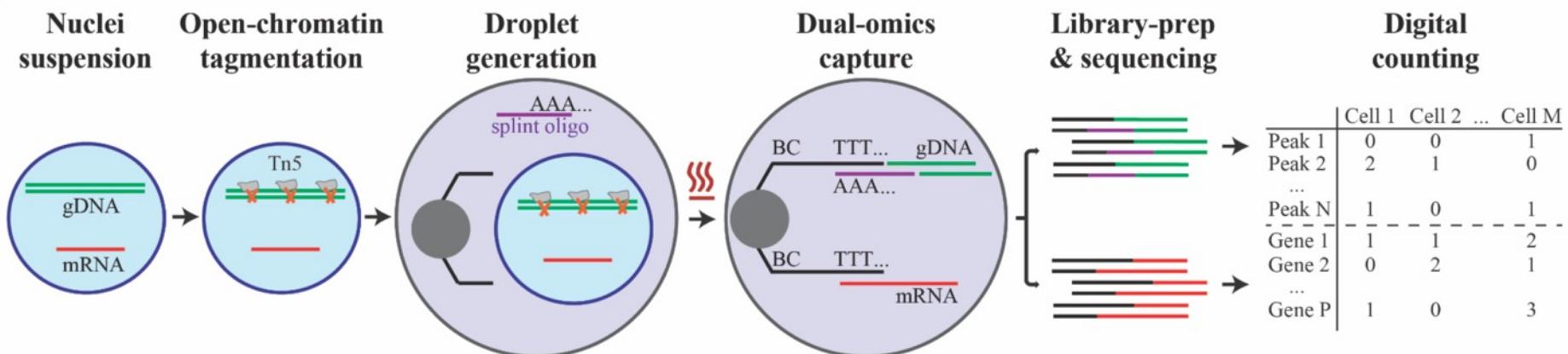
SHARE-Seq

- Same-cell scRNA/ATAC
- Combinatorial split-pool barcoding of adapters



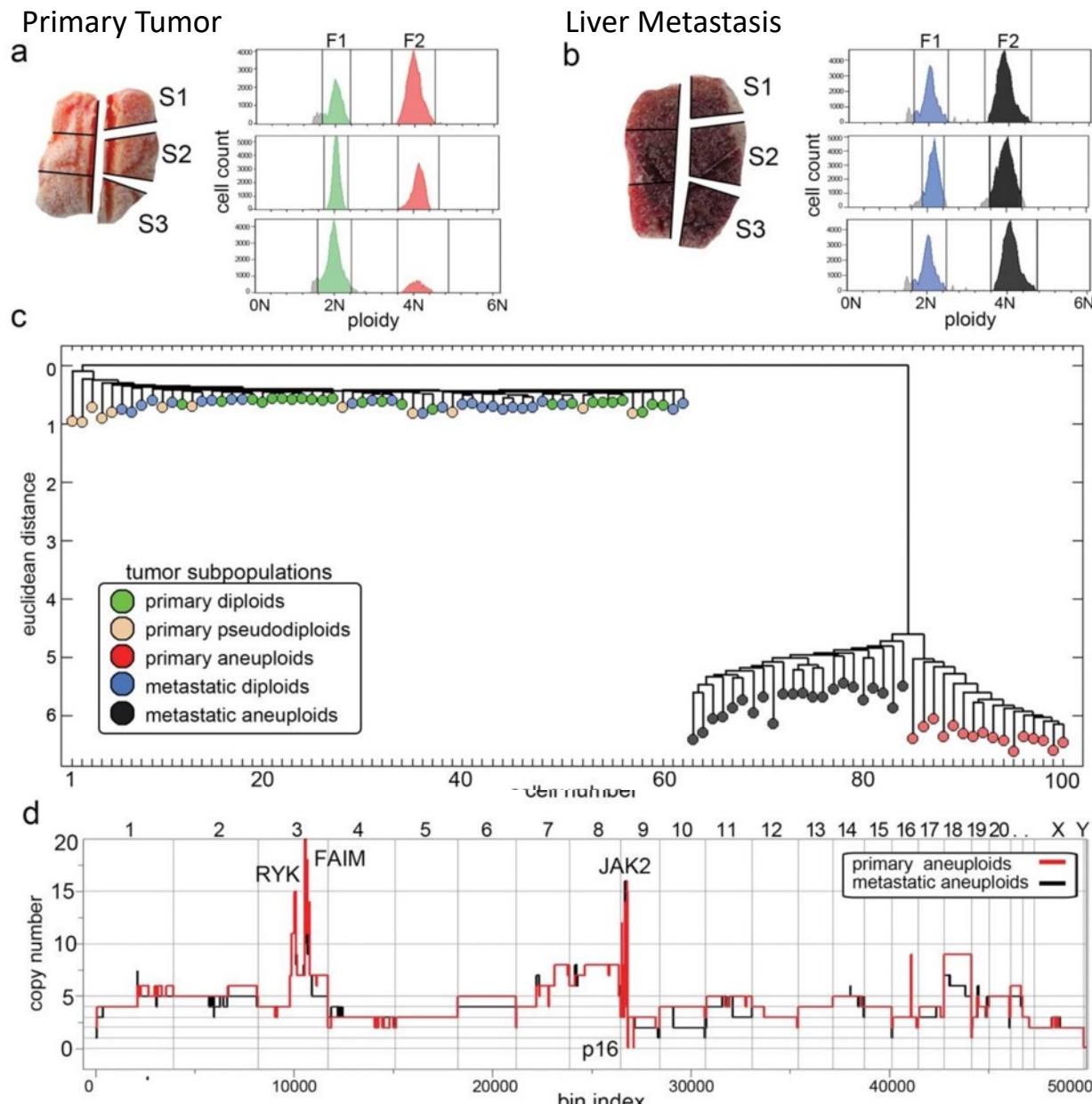
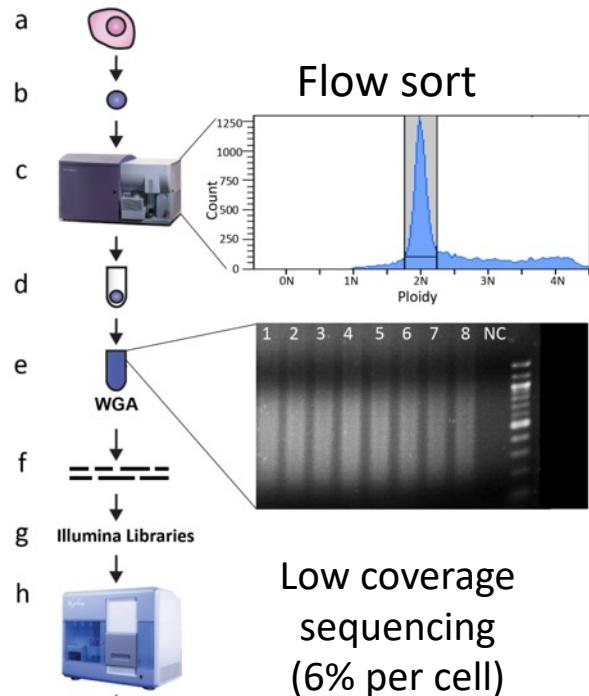
SNARE-seq

a

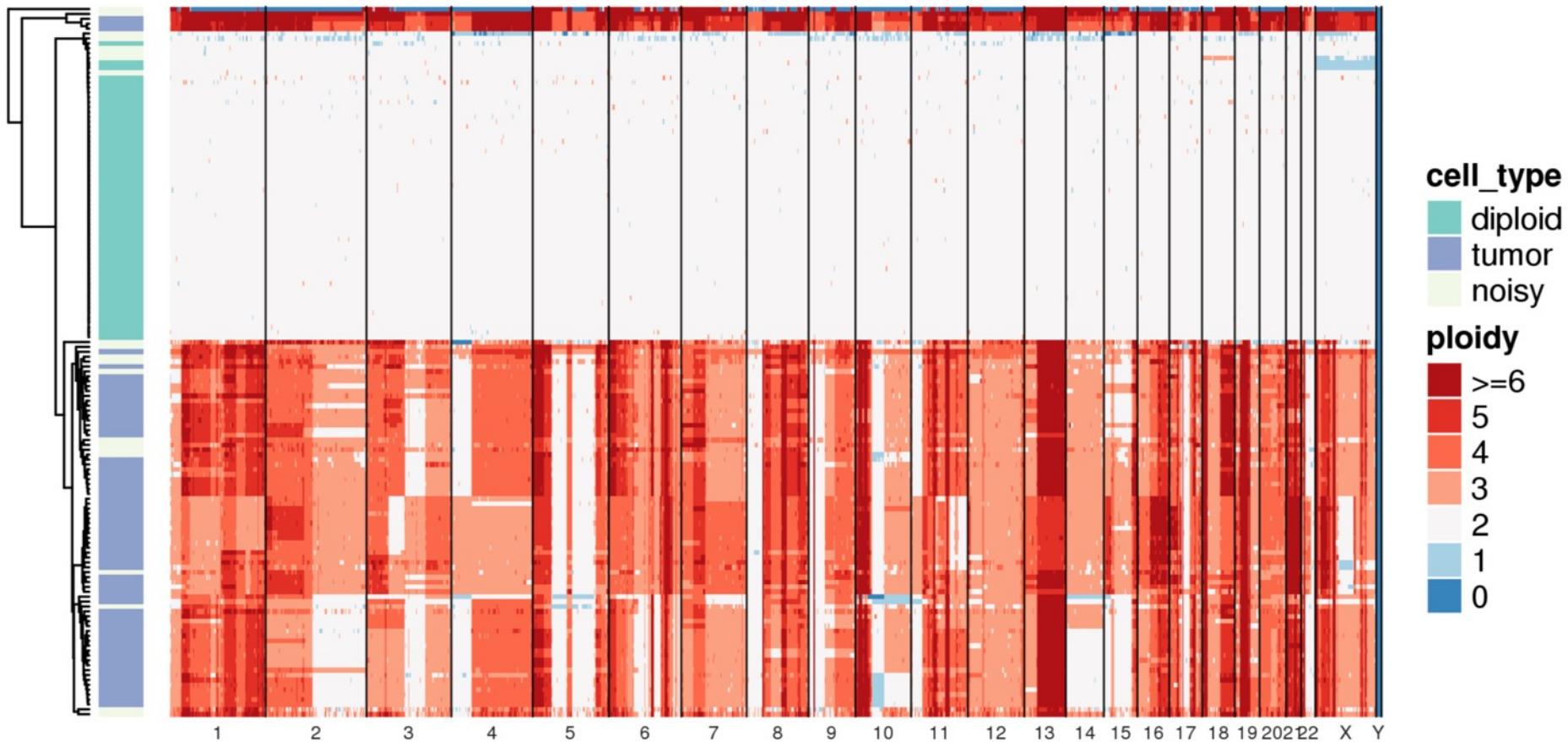


Single cell CNV

Nick Navin, Mike Wigler
CSHL



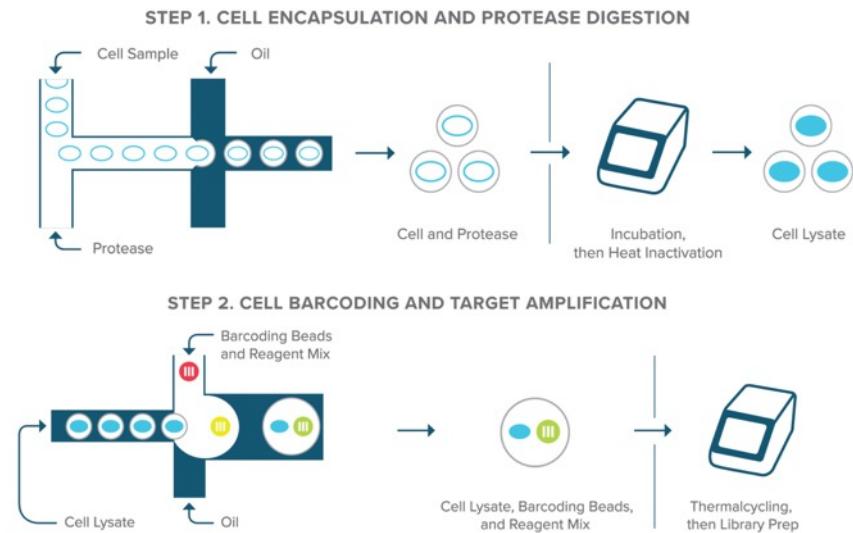
Droplet-based Single Cell CNV



Mission Bio Tapestri

DNA-focused microfluidic platform

For SNV & CNV



Mission Bio Tapestri

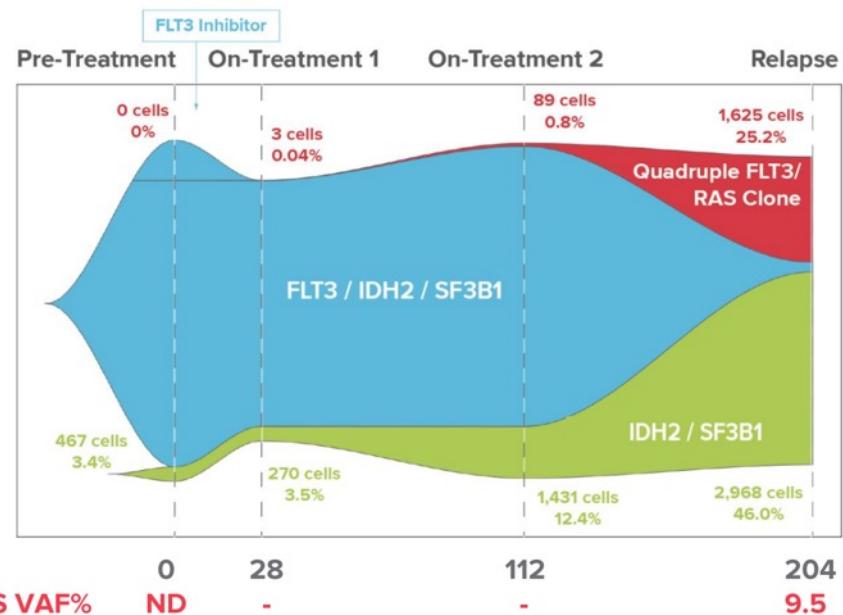
59 GENES - TUMOR HOTSPOT PANEL

ABL1	CSF1R	FGFR1	IDH2	MLH1	RB1
AKT1	CTNNB1	FGFR2	JAK1	MPL	RET
ALK	DDR2	FGFR3	JAK2	MTOR	SMAD4
APC	EGFR	FLT3	JAK3	NOTCH1	SMARCB1
AR	ERBB2	GNA11	KDR	NRAS	SMO
ATM	ERBB3	GNAQ	KIT	PDGFRA	SRC
BRAF	ERBB4	GNAS	KRAS	PIK3CA	STK11
CDH1	ESR1	HNF1A	MAP2K1	PTEN	TP53
CDK4	EZH2	HRAS	MAP2K2	PTPN11	VHL
CDKN2A	FBXW7	IDH1	MET	RAF1	

45-GENE MYELOID PANEL

ASXL1	ERG	KDM6A	NRAS	SMC1A
ATM	ETV6	KIT	PHF6	SMC3
BCOR	EZH2	KMT2A	PPM1D	STAG2
BRAF	FLT3	KRAS	PTEN	STAT3
CALR	GATA2	MPL	PTPN11	TET2
CBL	GNAS	MYC	RAD21	TP53
CHEK2	IDH1	MYD88	RUNX1	U2AF1L5
CSF3R	IDH2	NF1	SETBP1	WT1
DNMT3A	JAK2	NPM1	SF3B1	ZRSR2

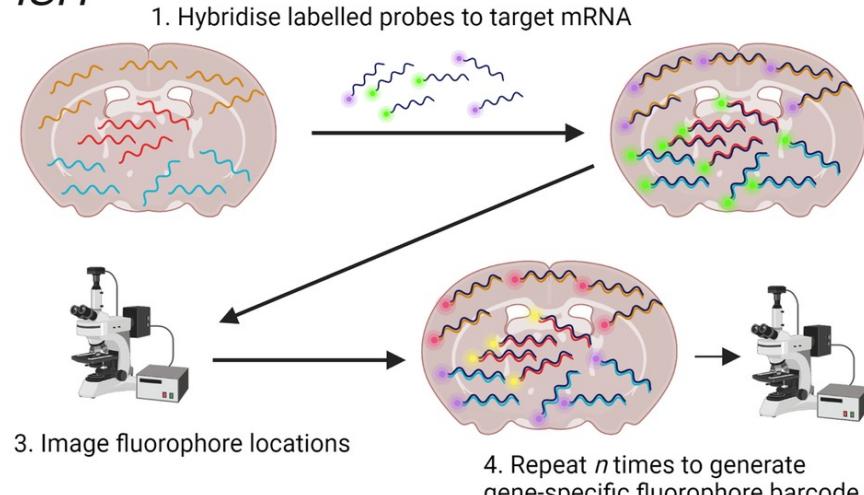
Clonal Architecture Resolved Over Time



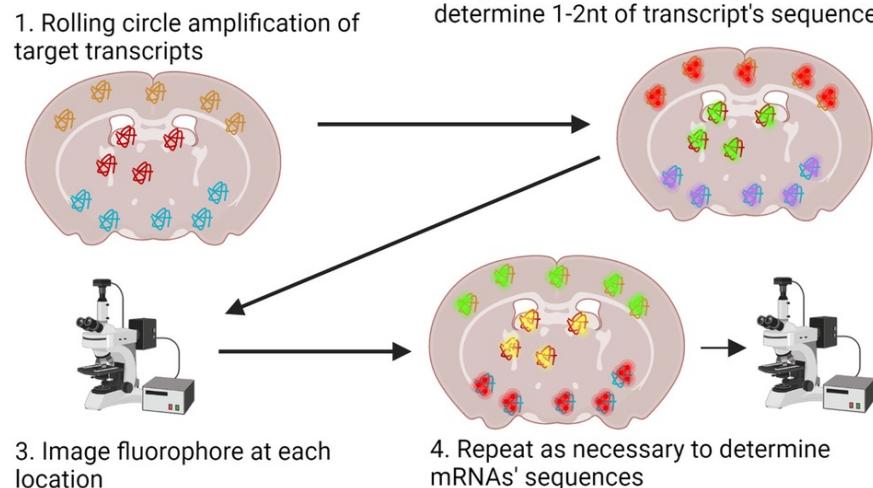
Spatial Biology

Imaging methods

ISH

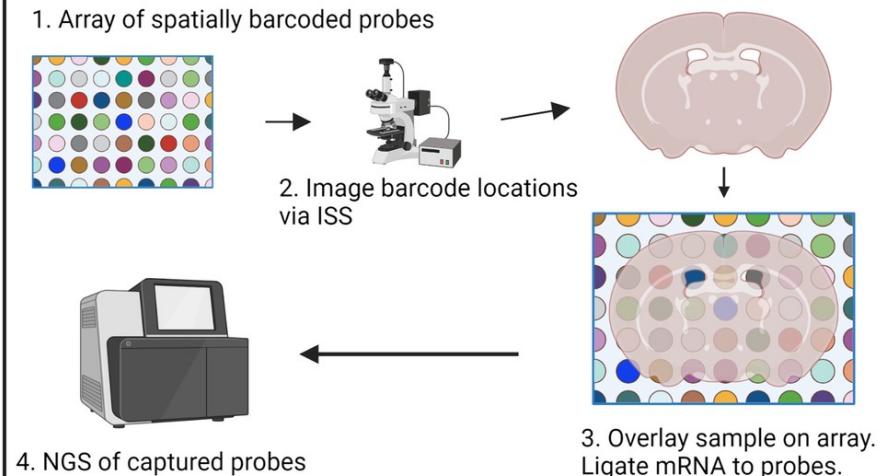


ISS

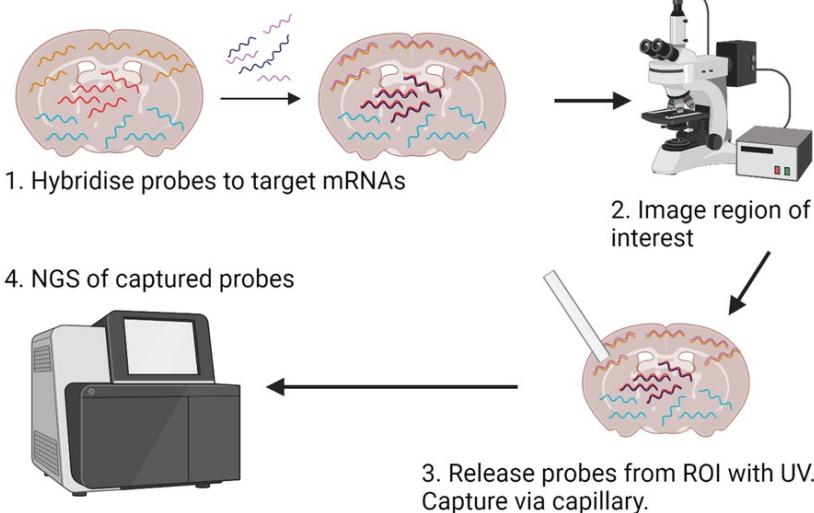


Sequencing methods

Arrays



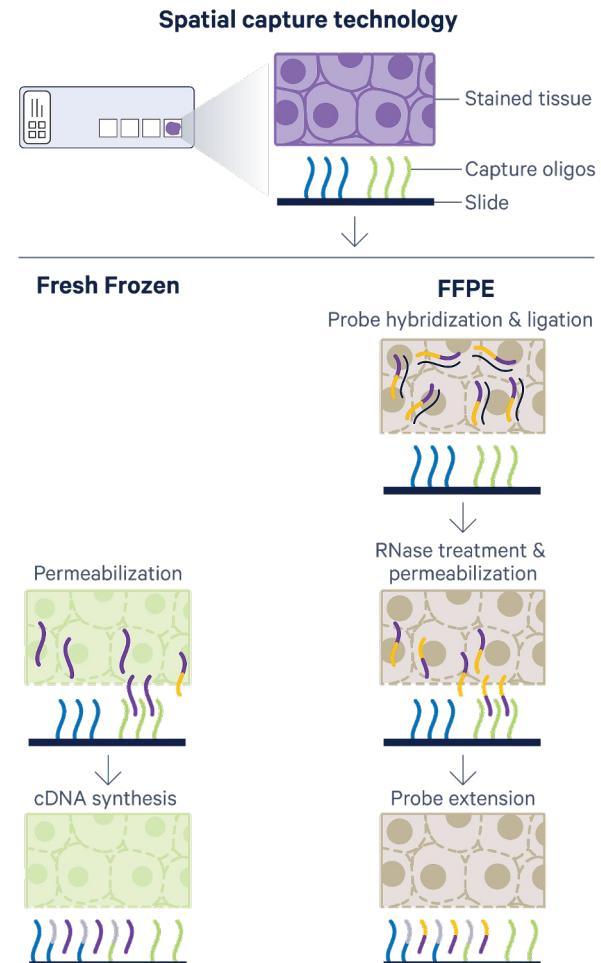
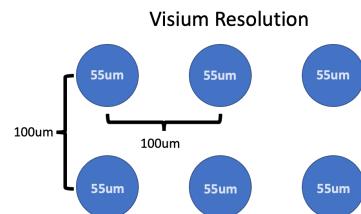
Microdissection

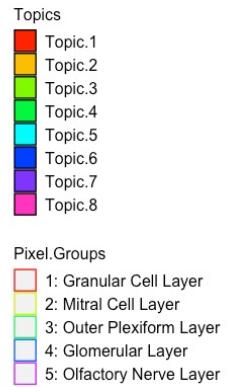
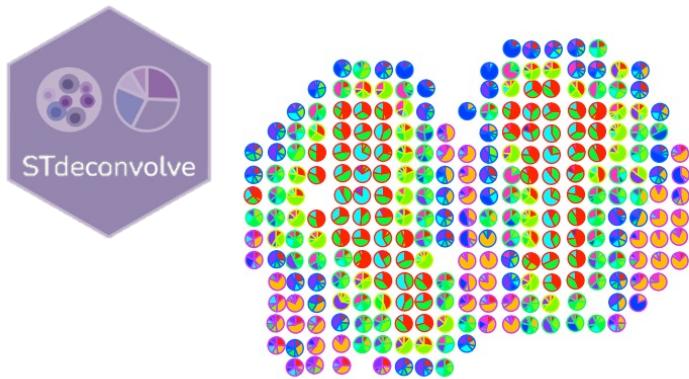




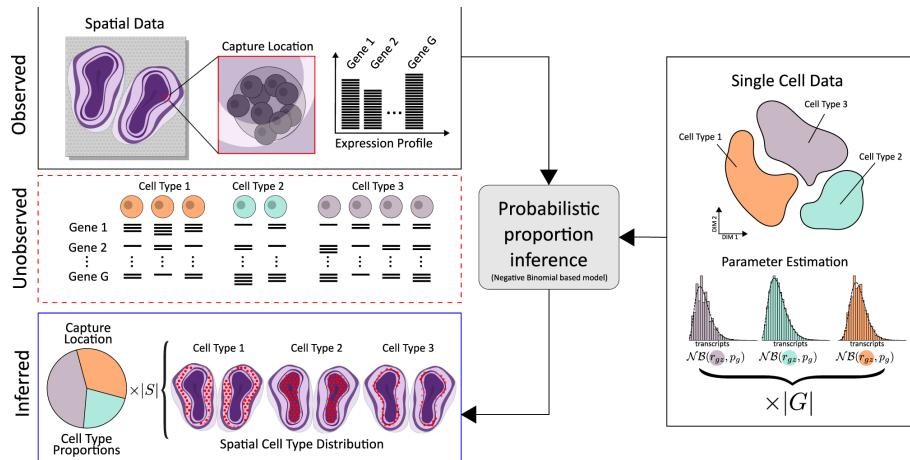
Visium

- 5,000 spots / capture area
- 55um spots
- 100um spot-to-spot distance
- Whole transcriptome (dT priming, Fresh/Frozen)
or
20K+ gene probes (FFPE)
- H&E or 4-color IF co-staining options
- Proteomics via barcoded antibodies
- **FFPE** and Fresh Frozen options
 - \$5,000 / slide (4 capture areas)
 - \$3500 Sequencing (Fresh Frozen)
 - \$1800 Sequencing (FFPE)

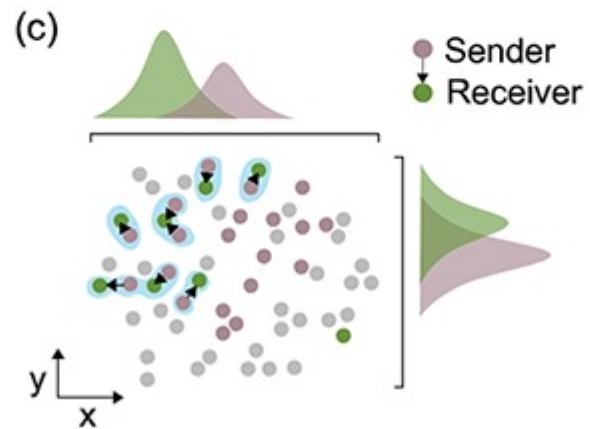




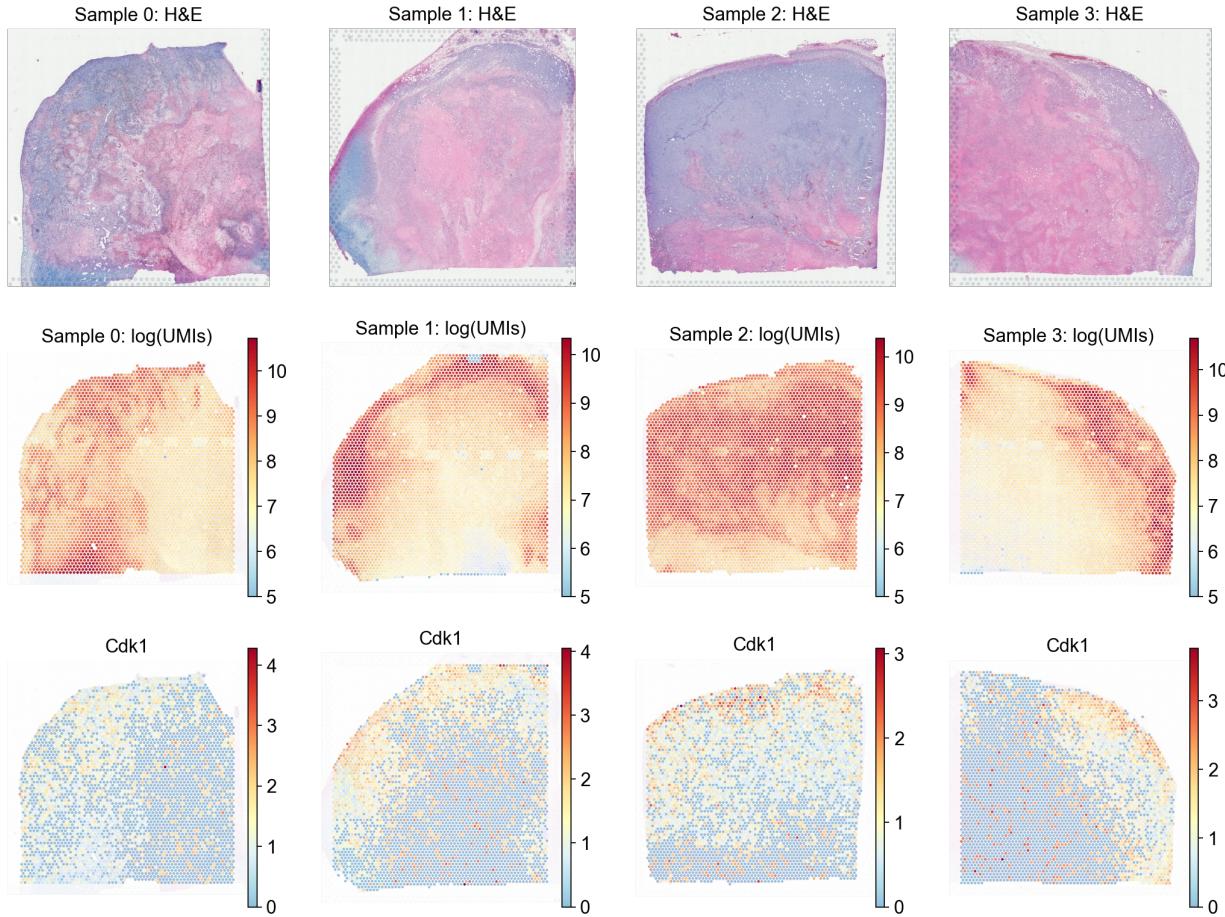
StereoScope – Probabilistic Cell Type Deconvolution



SpaTalk Ligand / Receptor Analysis



Mouse Tumor Tissue: Visium FFPE

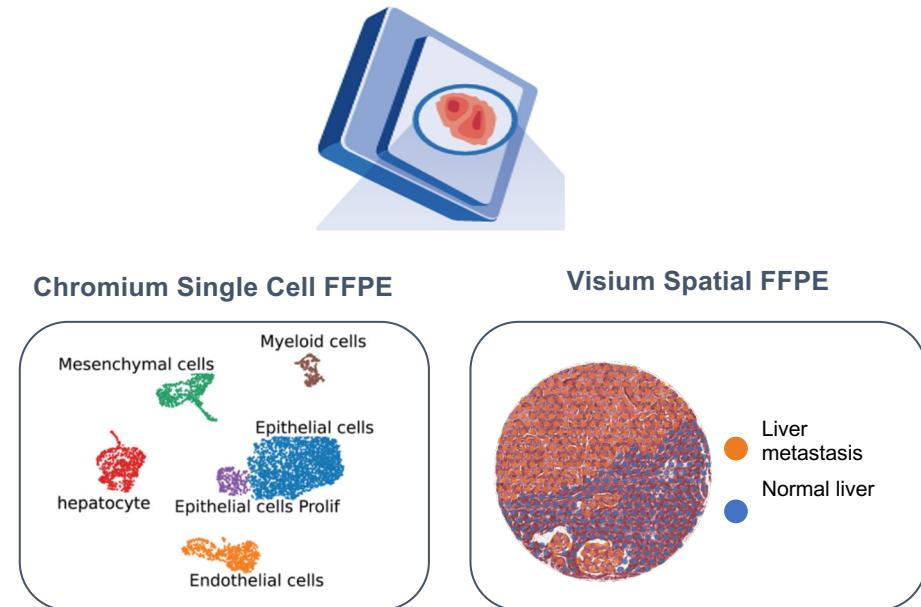


snPATHO-seq – isolation of nuclei from FFPE cells

Enables **retrospective** studies with spatial & single-cell components

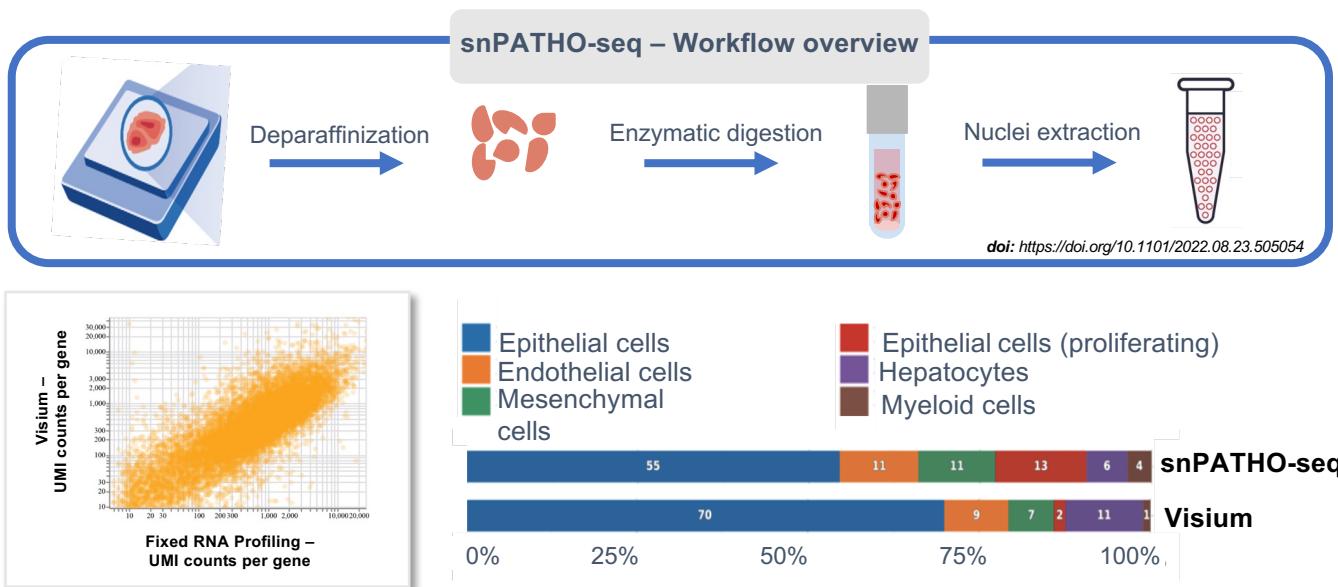


The image shows a screenshot of a bioRxiv preprint page. At the top left is the CSHL logo and the bioRxiv logo with the tagline "THE PREPRINT SERVER FOR BIOLOGY". Below the header is a yellow box containing a reminder: "bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive." The main title of the article is "snPATHO-seq: unlocking the FFPE archives for single nucleus RNA profiling". It lists authors: Andres F Vallejo, Kate Harvey, Taopeng Wang, Kellie Wise, Lisa M Butler, Jose Polo, Jasmine T Plummer, Alexander Swarbrick, and Luciano G Martelotto. The DOI is provided as <https://doi.org/10.1101/2022.08.23.50504>. A note below states: "This article is a preprint and has not been certified by peer review [what does this mean?].". Below the abstract are social media sharing icons (0 shares, 0 comments, 0 downloads, 0 likes, 83 tweets) and links to "Abstract", "Info/History", "Metrics", and "Preview PDF". The abstract text discusses the challenges of sequencing FFPE tissue archives due to their formalin-fixation and paraffin embedding, which makes them slow, labour intensive, and less sensitive than single-cell sequencing methods.



doi: <https://doi.org/10.1101/2022.08.23.505054>

FLEX and Visium show strong correlation on FFPE samples



Spatial Profiling @ Subcellular Resolution



10X Xenium

100s – 1000s of genes, pre-designed panels, FFPE compatible

<https://www.10xgenomics.com/products/xenium-in-situ/human-breast-dataset-explorer>

Or....

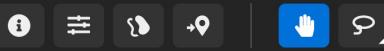


Vizgen MERSCOPE

Custom gene panels, FFPE compatible

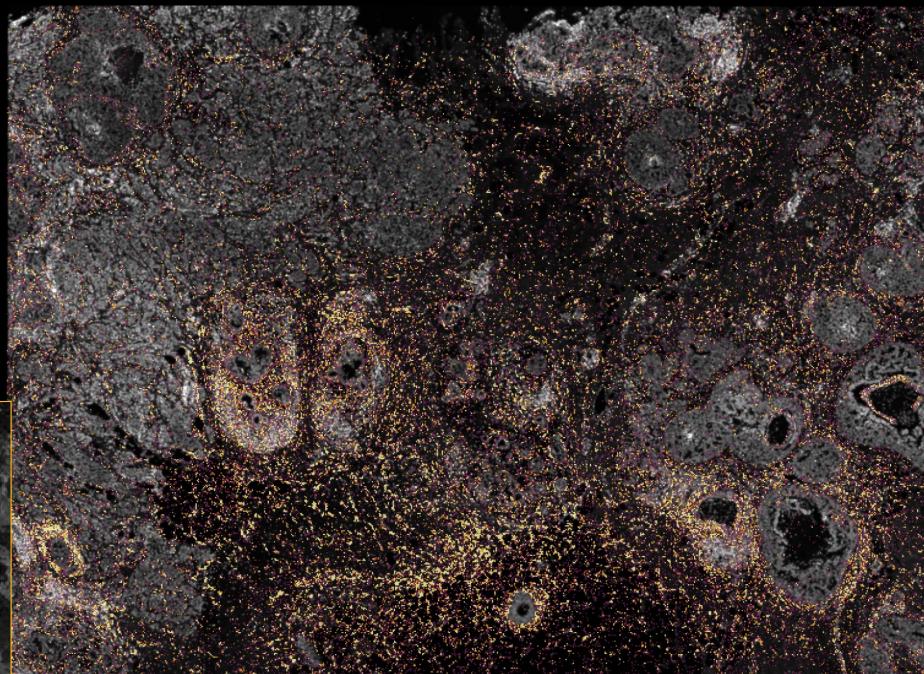
Proteogenomics

<https://vizgen.com/data-release-program/>



FFPE
313 Gene panel
0.4 cm²

PDGFRA

5242.5 2605.4 μm

Acknowledgements

Preall Lab:

Claire Regan
Brian He
Caleb Mallery
Rad Utama

Tuveson Lab

Ela Elyada
Giulia Biffi

Dick McCombie / NGS
Sara
Goodwin
Elena
Ghiban

Beyaz Lab

Funding:

Northwell Health
CSHL Cancer Center Support

Grant

Fearon Lab

Rena Yan

