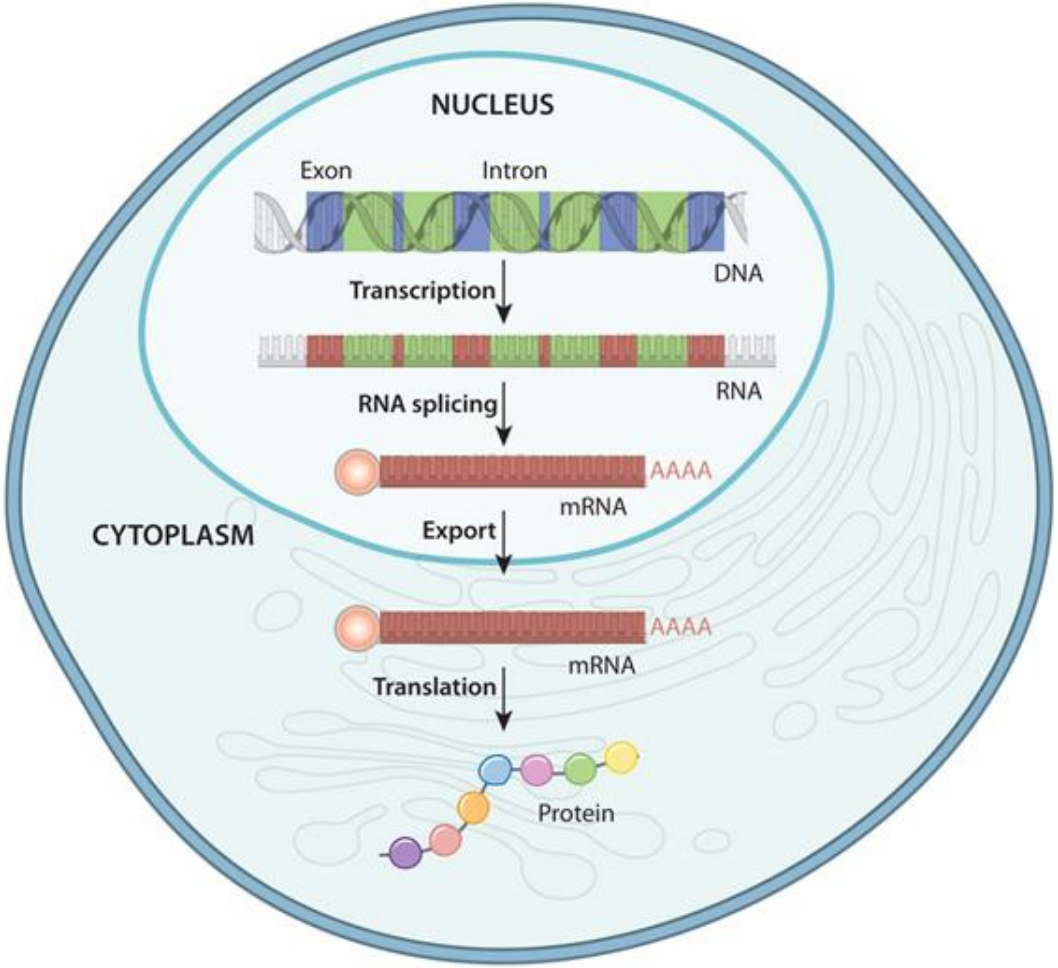
THE POWER OF SINGLE CELL ANALYSIS: SINGLE CELL SEQ AND SPATIAL TRANSCRIPTOMICS

ADVANCED CELL EXPLORATION (ACE) CORE
MIKE WONG

UNIVERSITY
OF ALBERTA

Eukaryotic Gene Expression

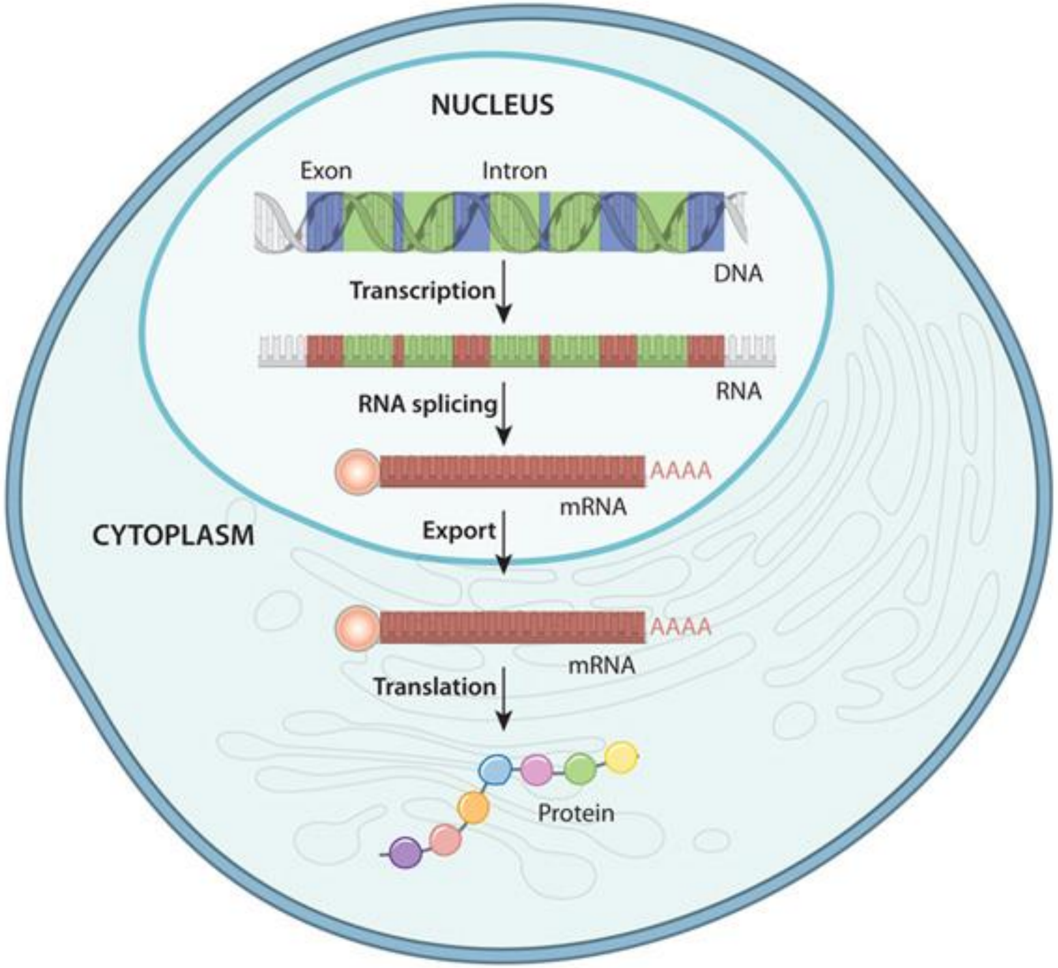


www.nature.com

Eukaryotic Gene Expression

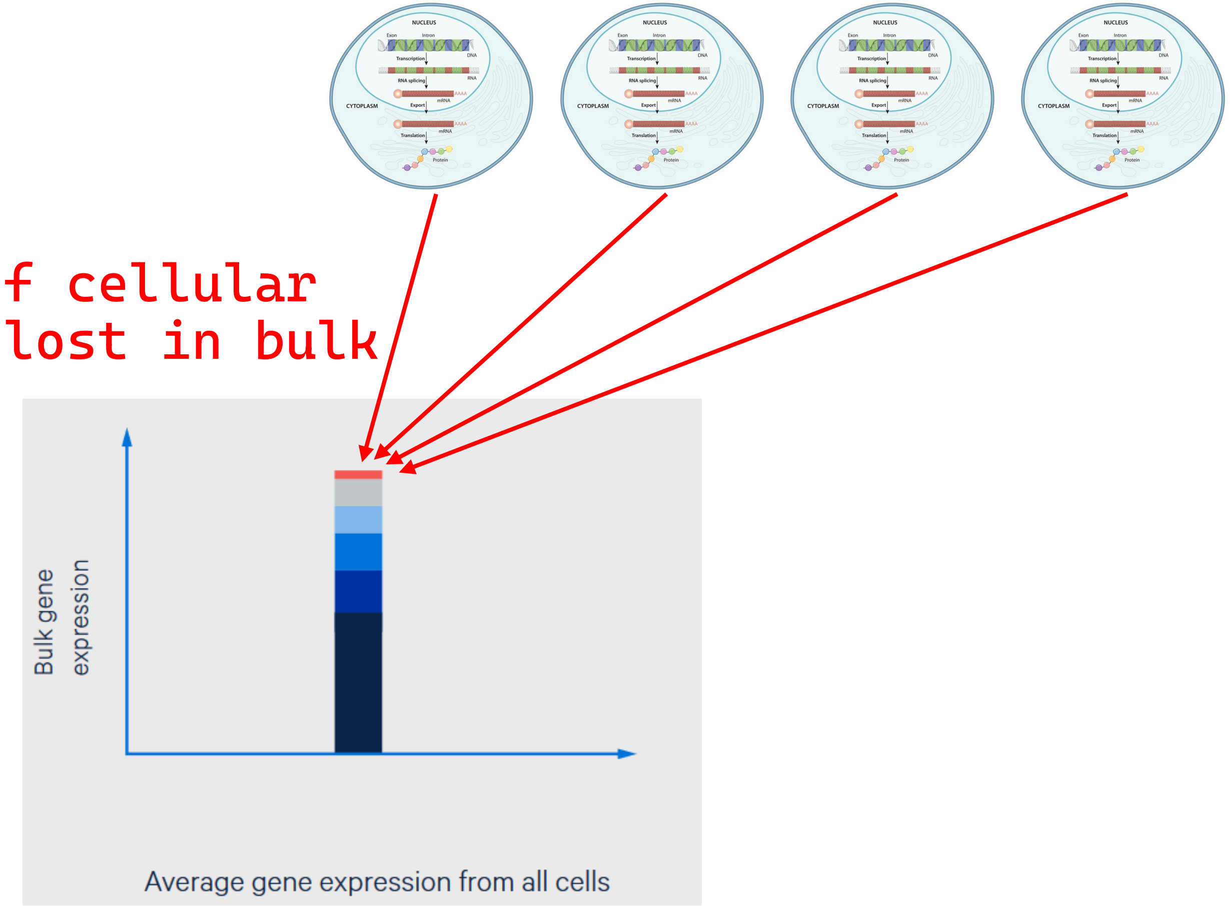
Central Dogma of Biology

- DNA -> mRNA -> Protein
- Gene expression previously measured in ‘bulk’ methods where all RNA from a batch of cells is collected to analyze
  - Quantitative RTPCR
    - Target must be known
  - Bulk RNA Sequencing
    - Can only trace transcripts to whole input tissue or cell suspension
    - All population context lost



www.nature.com

Eukaryotic Gene Expression

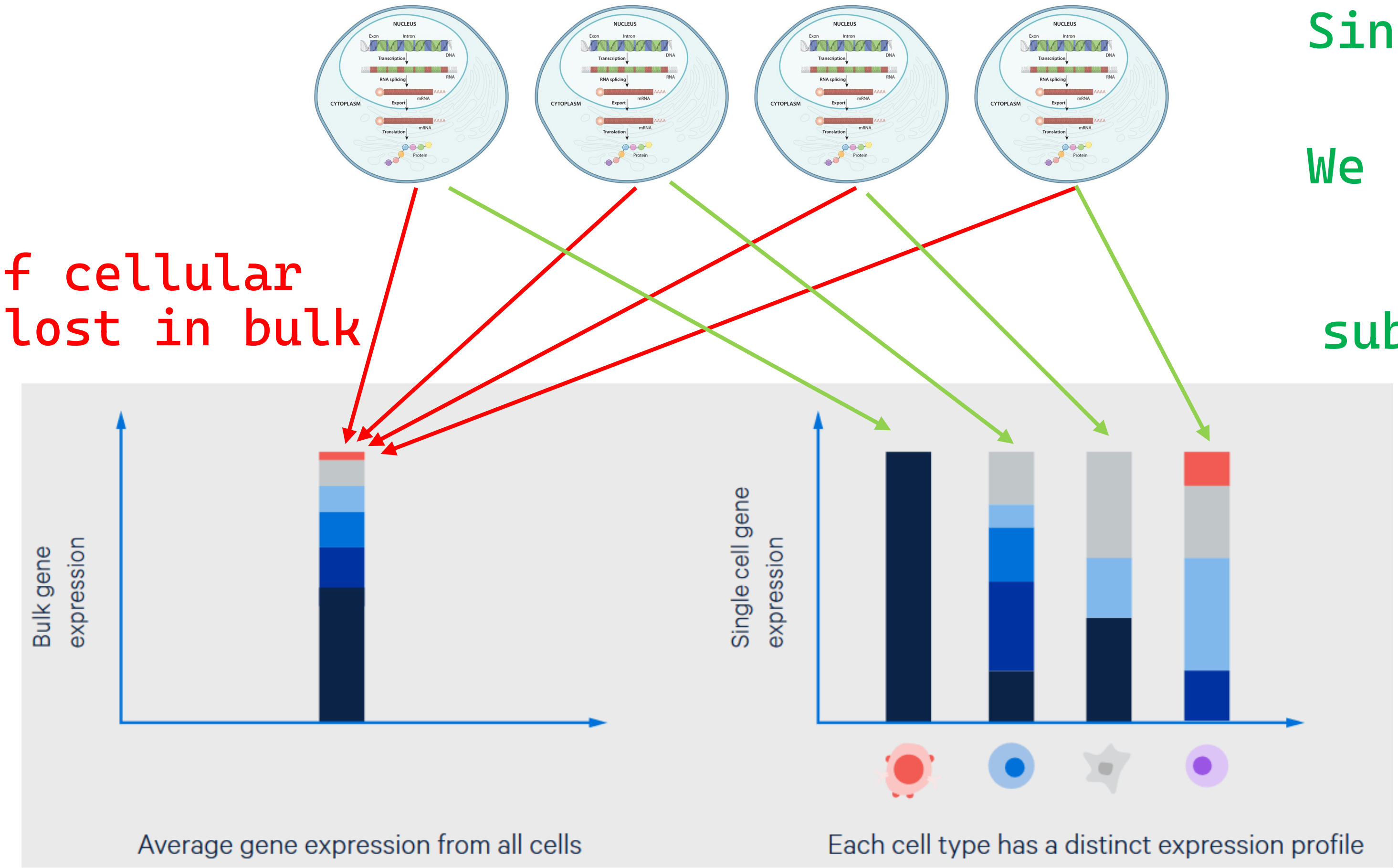


Complexity of cellular populations lost in bulk analysis

www.nature.com

Eukaryotic Gene Expression

Complexity of cellular populations lost in bulk analysis



Average gene expression from all cells

Each cell type has a distinct expression profile

Single Cell Gene Expression

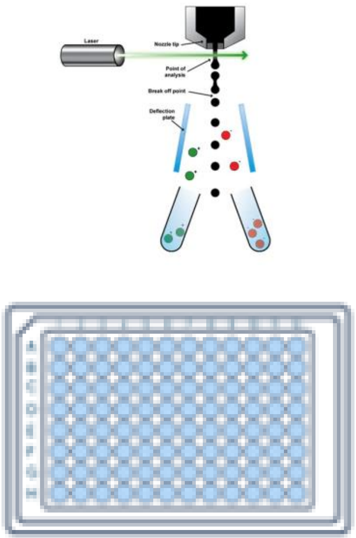
We can now start looking at individual subpopulations!

www.nature.com

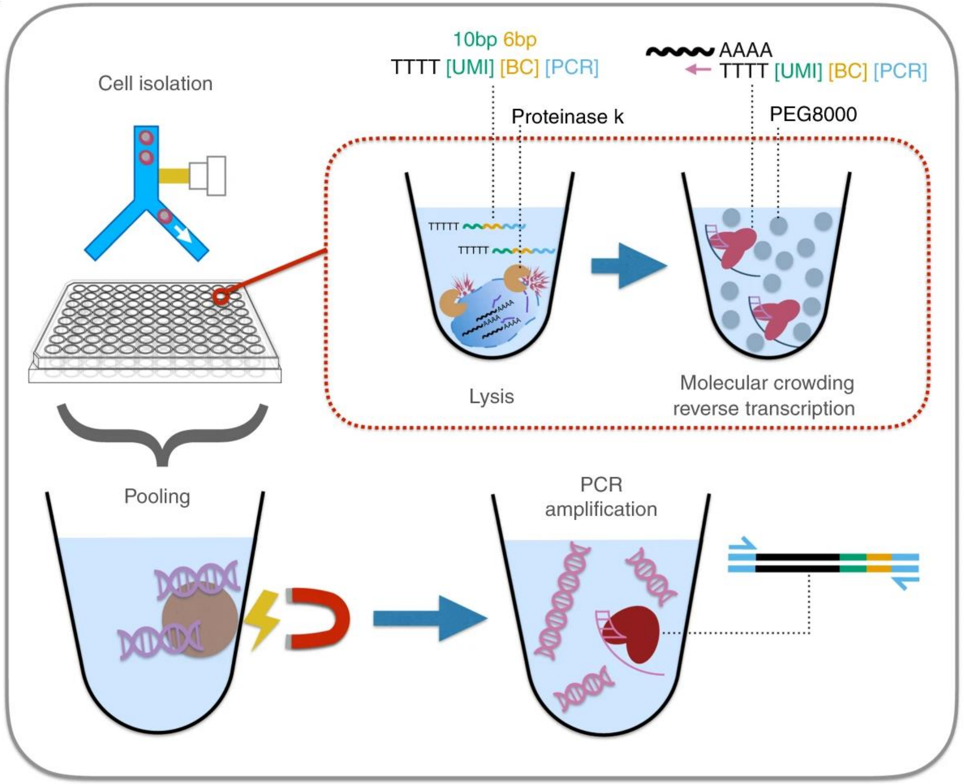
Single Cell Sequencing Techniques

True Single Cell Sequencing

- **Glass pipette picked or FACS-Sorted Single Cells**
- **Similar processing steps to regular bulk RNA seq, but with ultra-low input**
- **Low throughput, high cost per cell**
- **Currently used for very specific purposes**



1-384 Cells

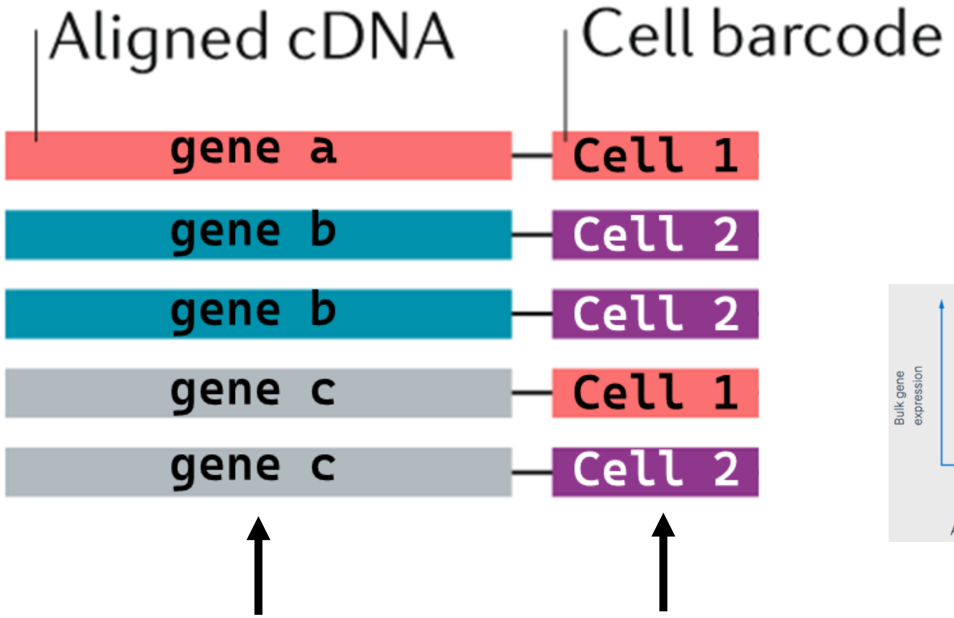


Single Cell Sequencing Techniques

Single Cell Barcoding techniques

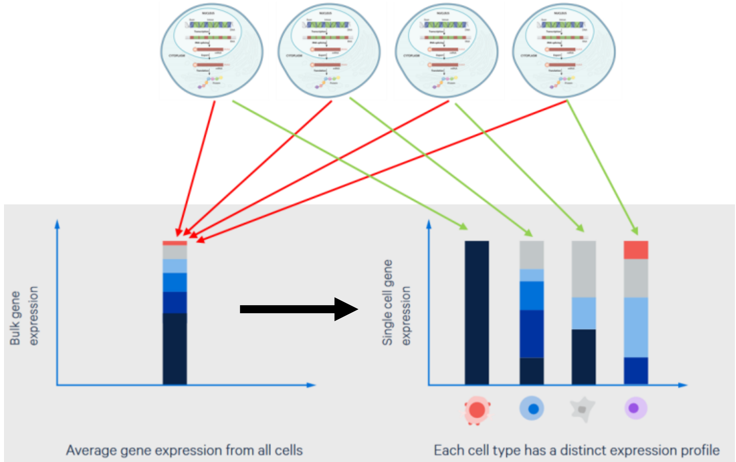
- <u>**Barcode the cDNAs from each cell with a cell identifier**</u>
- **Process as bulk RNA**
- <u>**Use the identifier to assign transcript counts to individual cells**</u>

- **Moderate to high throughput**
- **Expensive, but low cost/cell**



cDNA created from mRNA transcripts

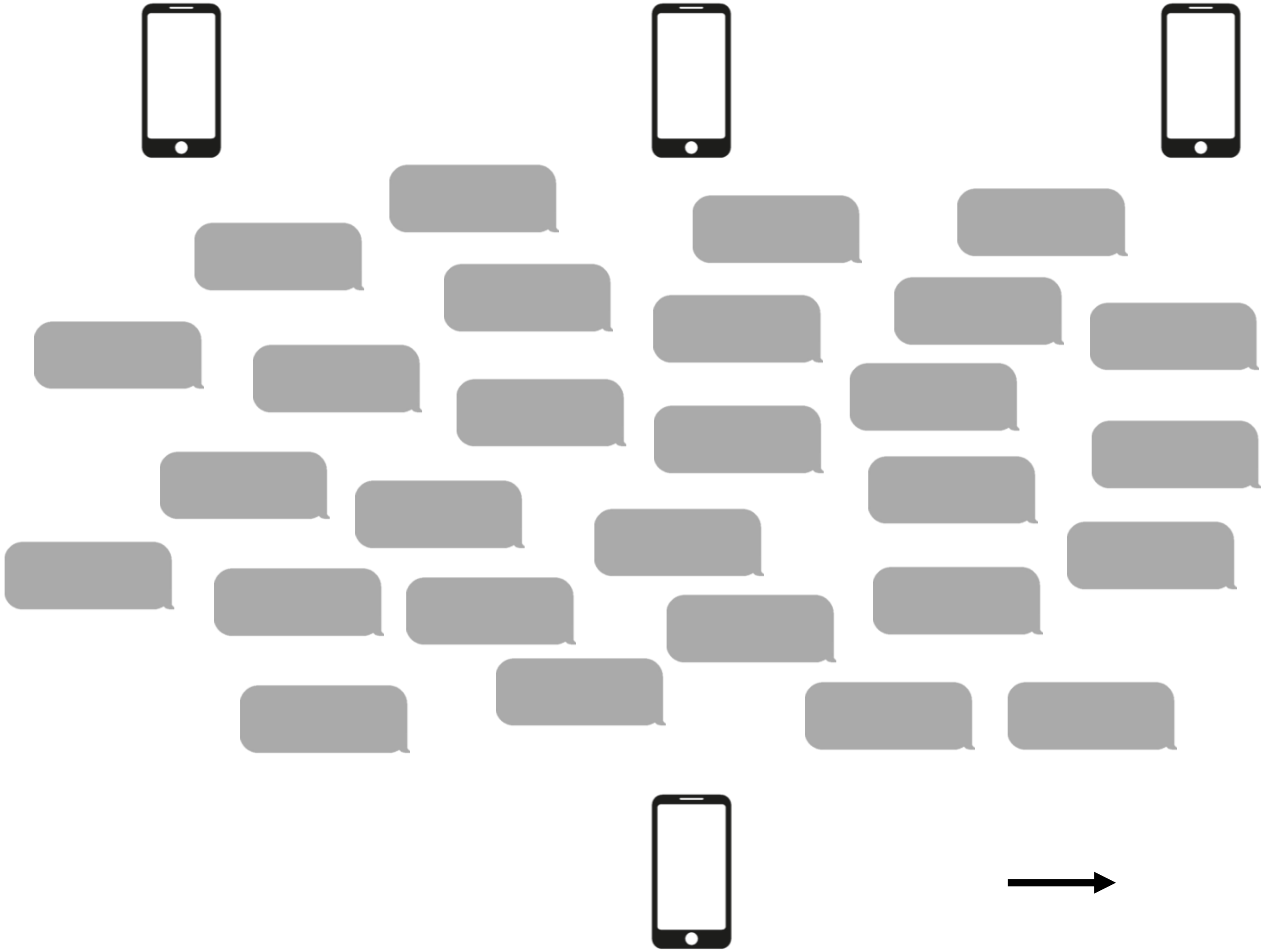
Cell identifier barcode



[10X Genomics]

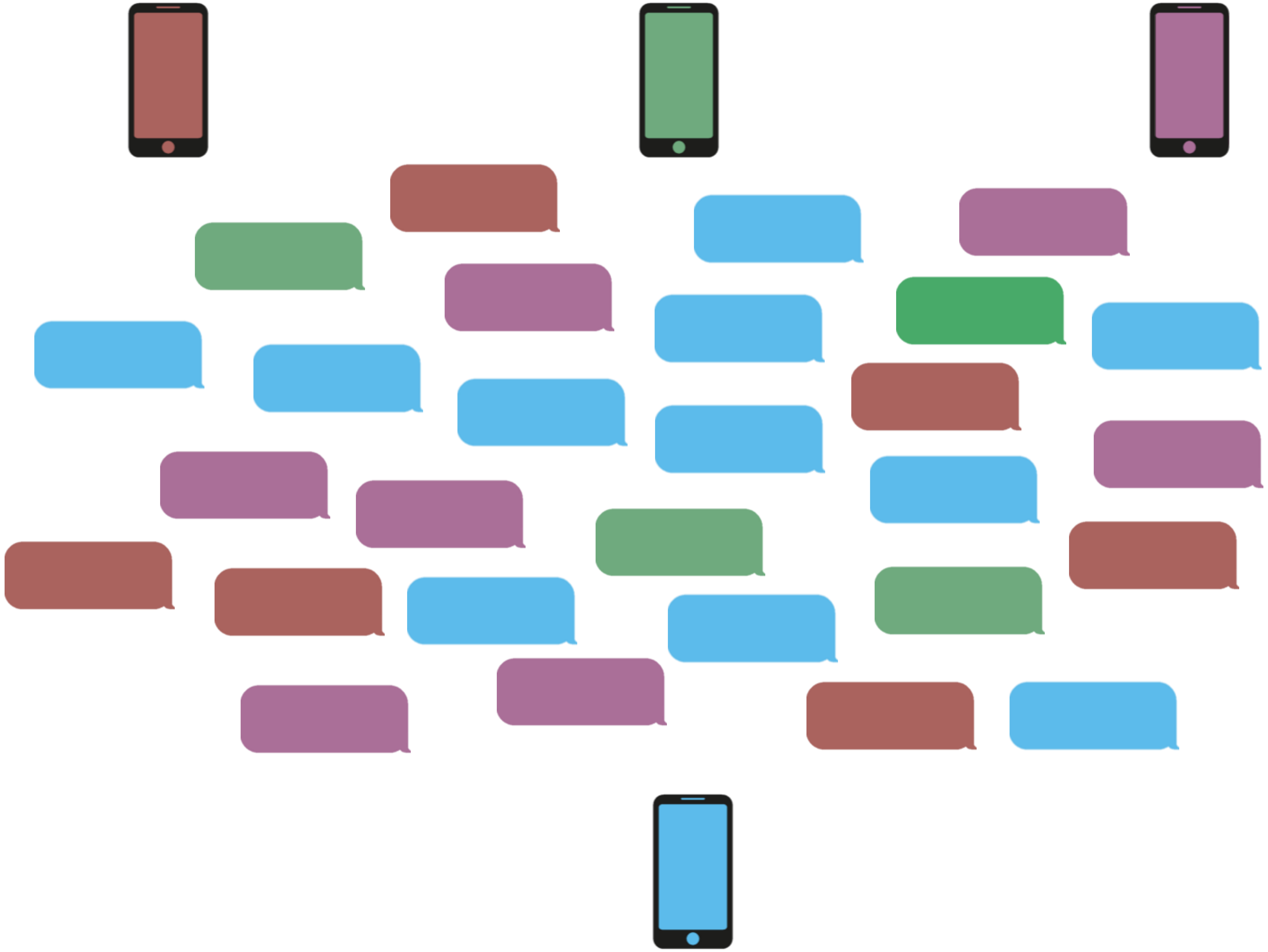
Single Cell Sequencing

- **With bulk RNA sequencing, all messages are mixed together!**



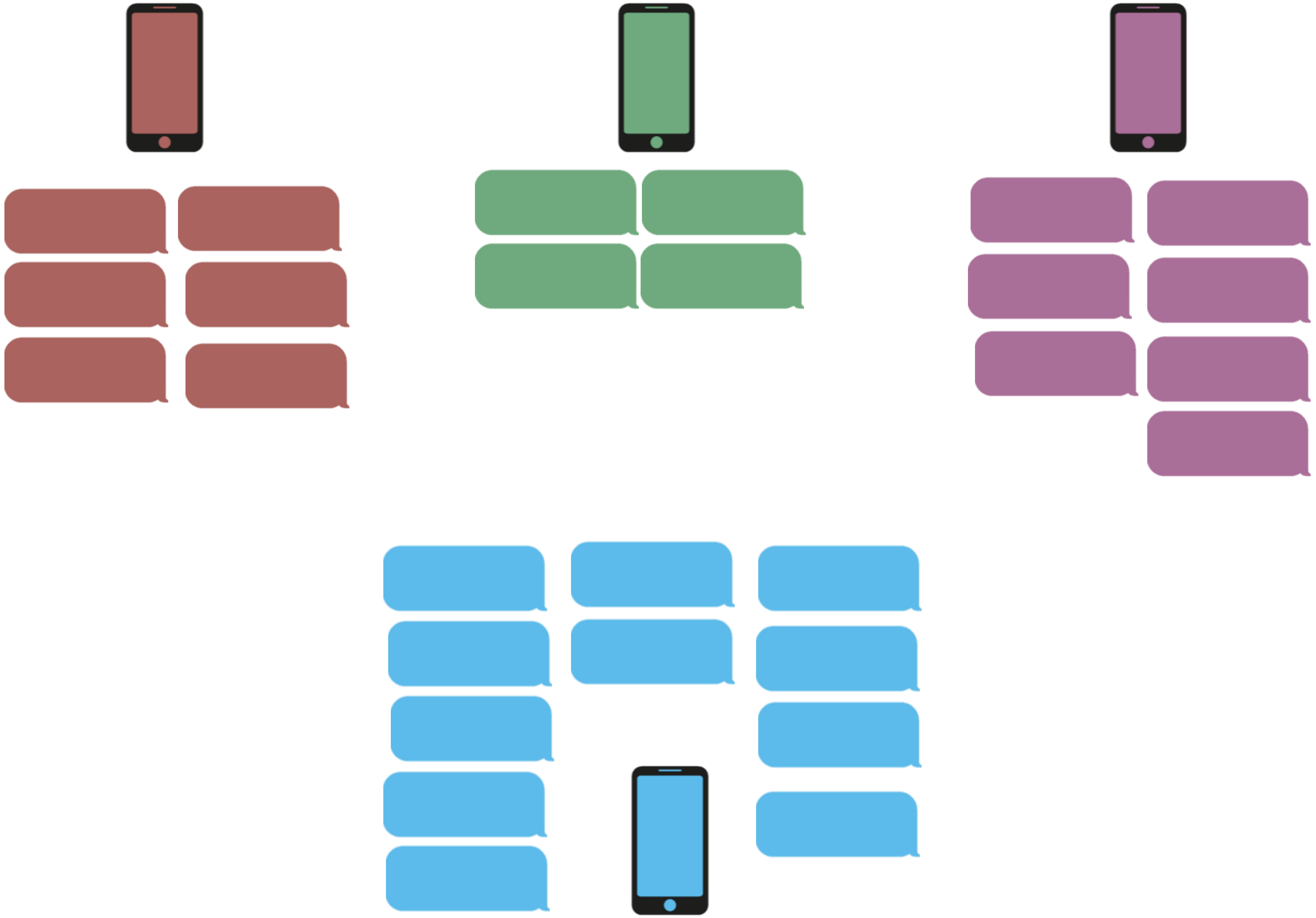
Single Cell Sequencing

- With bulk RNA sequencing, all messages are mixed together!
- If we add handles, we can now figure out the source of each message



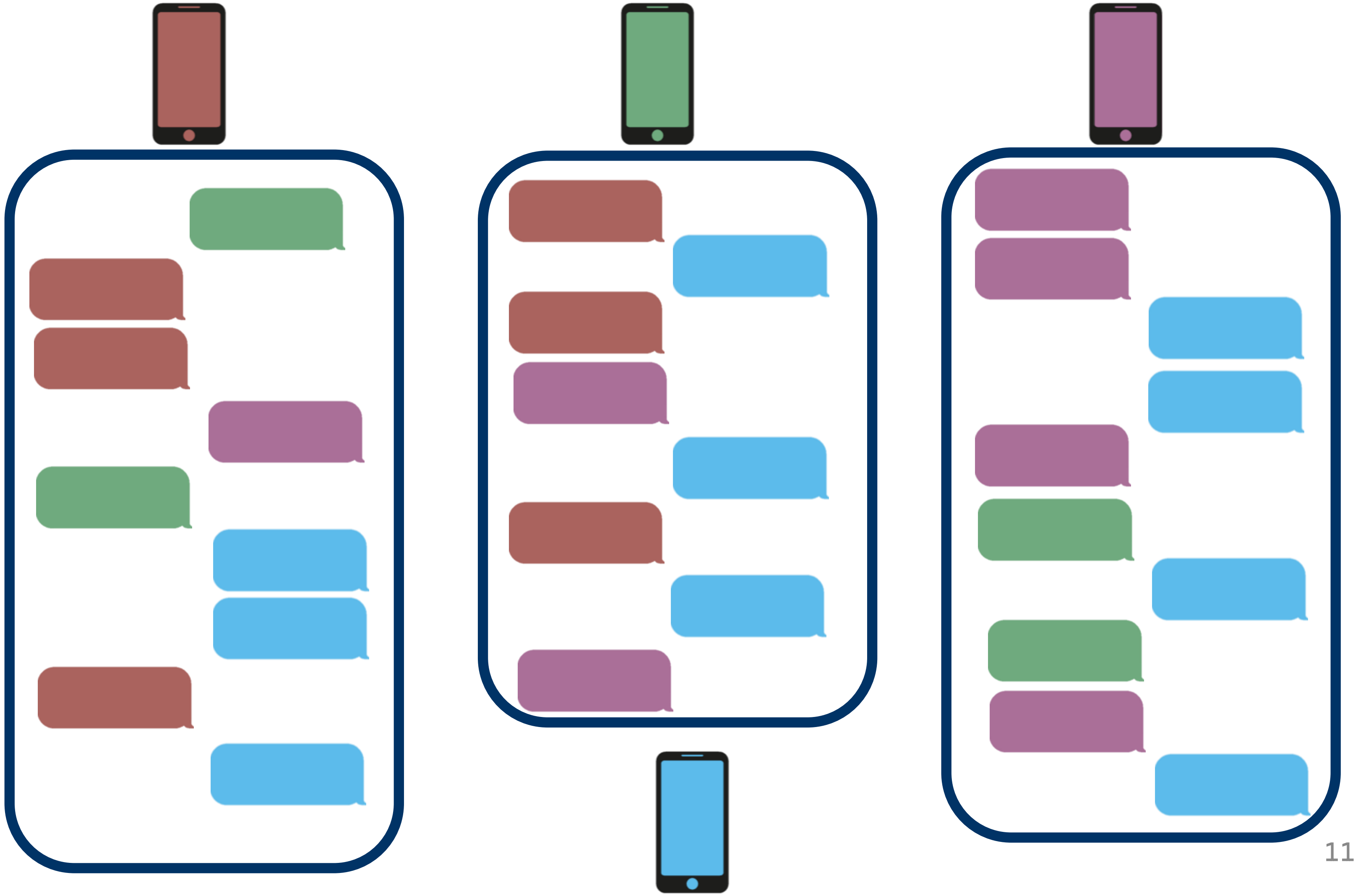
Single Cell Sequencing

- **With bulk RNA sequencing, all messages are mixed together!**
- **If we add handles, we can now figure out the source of each message**



Single Cell Sequencing

- With bulk RNA sequencing, all messages are mixed together!
- If we add handles, we can now figure out the source of each message
- Provide context to the groups of messages
- Determine contributions to each conversation

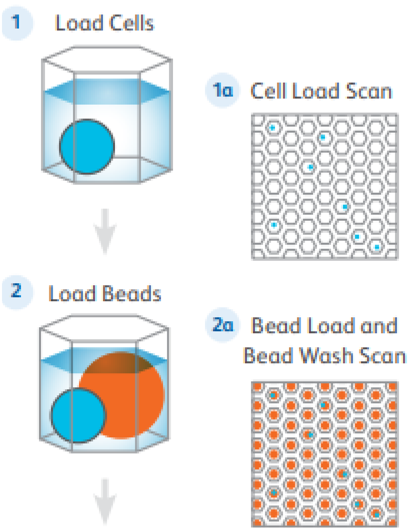


Single Cell Sequencing Techniques

Single Cell Barcoding techniques

- **Use various methods to give cells unique barcodes when mRNA is captured**
- **Moderate to high throughput**
- **Expensive, but low cost/cell**

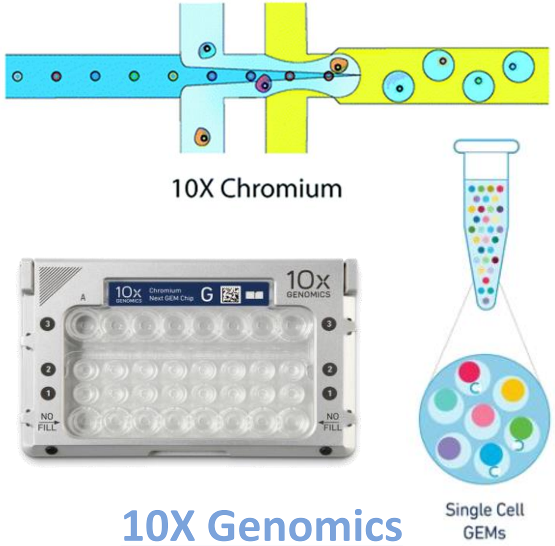
Gravity Loaded Nano-well Arrays



BD Rhapsody

Up to 100 000 cells/sample

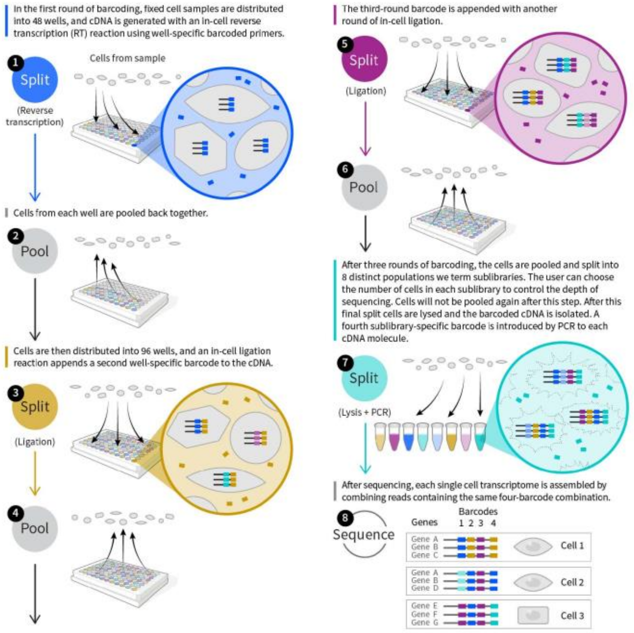
Microfluidic Droplet Based Partitioning



10X Genomics Chromium

Up to 10 000 cells/ sample

Combinatorial Barcoding


In the first round of barcoding, fixed cell samples are distributed into 48 wells, and cDNA is generated with an in-cell reverse transcription (RT) reaction using well-specific barcoded primers.

Cells from each well are pooled back together.

Cells are then distributed into 96 wells, and an in-cell ligation reaction appends a second well-specific barcode to the cDNA.

The third-round barcode is appended with another round of in-cell ligation.

After three rounds of barcoding, the cells are pooled and split into 8 distinct populations we term sublibraries. The user can choose the number of cells in each sublibrary to control the depth of sequencing. Cells will not be pooled again after this step. After this final split cells are lysed and the barcoded cDNA is isolated. A fourth sublibrary-specific barcode is introduced by PCR to each cDNA molecule.

After sequencing, each single cell transcriptome is assembled by combining reads containing the same four-barcode combination.

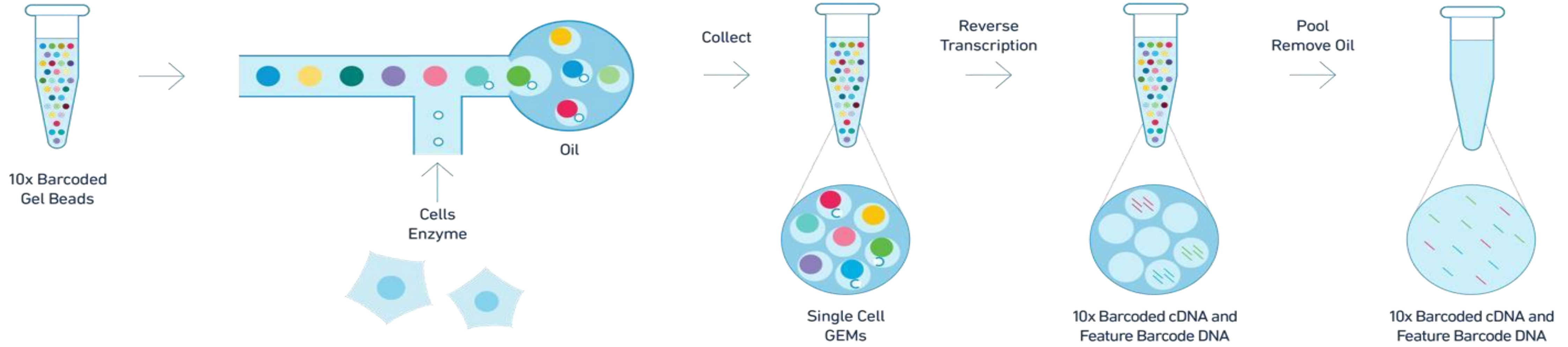
Parse Biosciences

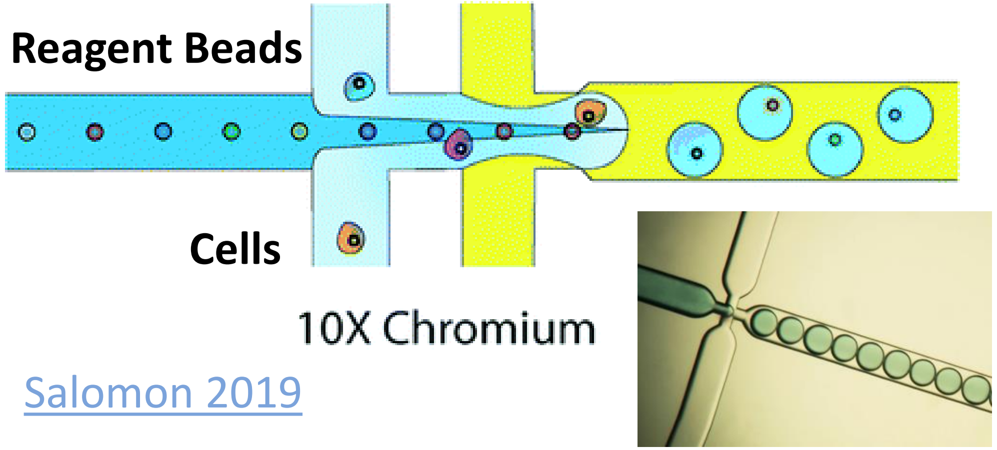
Up to 1 000 000 cells/sample

Droplet Microfluidic Partitioning scRNASeq

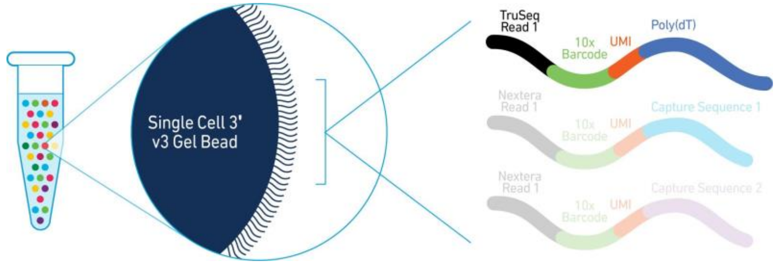
Fluidic Partitioning

Create oil droplet ‘chambers’ that will contain single cell reactions





Salomon 2019



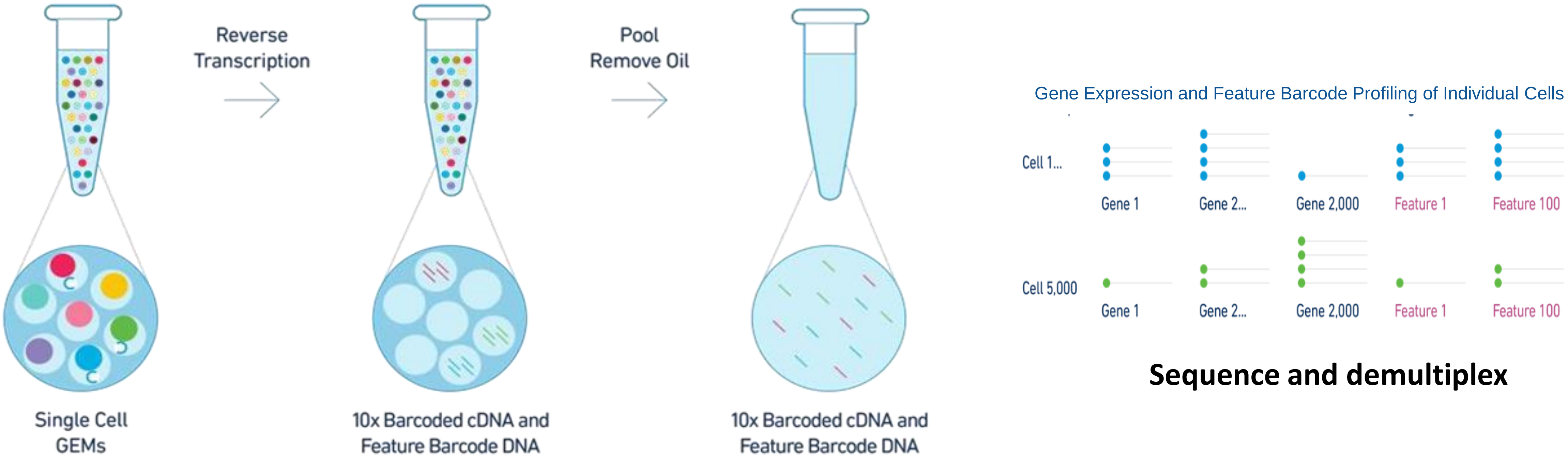
Beads contain:

- Poly-dT capture sequence
- TruSeq for Illumina Sequencing
- 10x Single Cell Barcode
- Unique Molecule Identifier (UMI)
- Polymerases and buffer components

10X Genomics

Droplet Microfluidic Partitioning scRNASeq

Fluidic Partitioning Reagent Beads with Individual Cells



Sequence and demultiplex

Emulsion Partitioning

Melting of beads
And cell lysis
Single cell barcoding

Remove partition process the sample the same as bulk RNA

[10X Genomics]

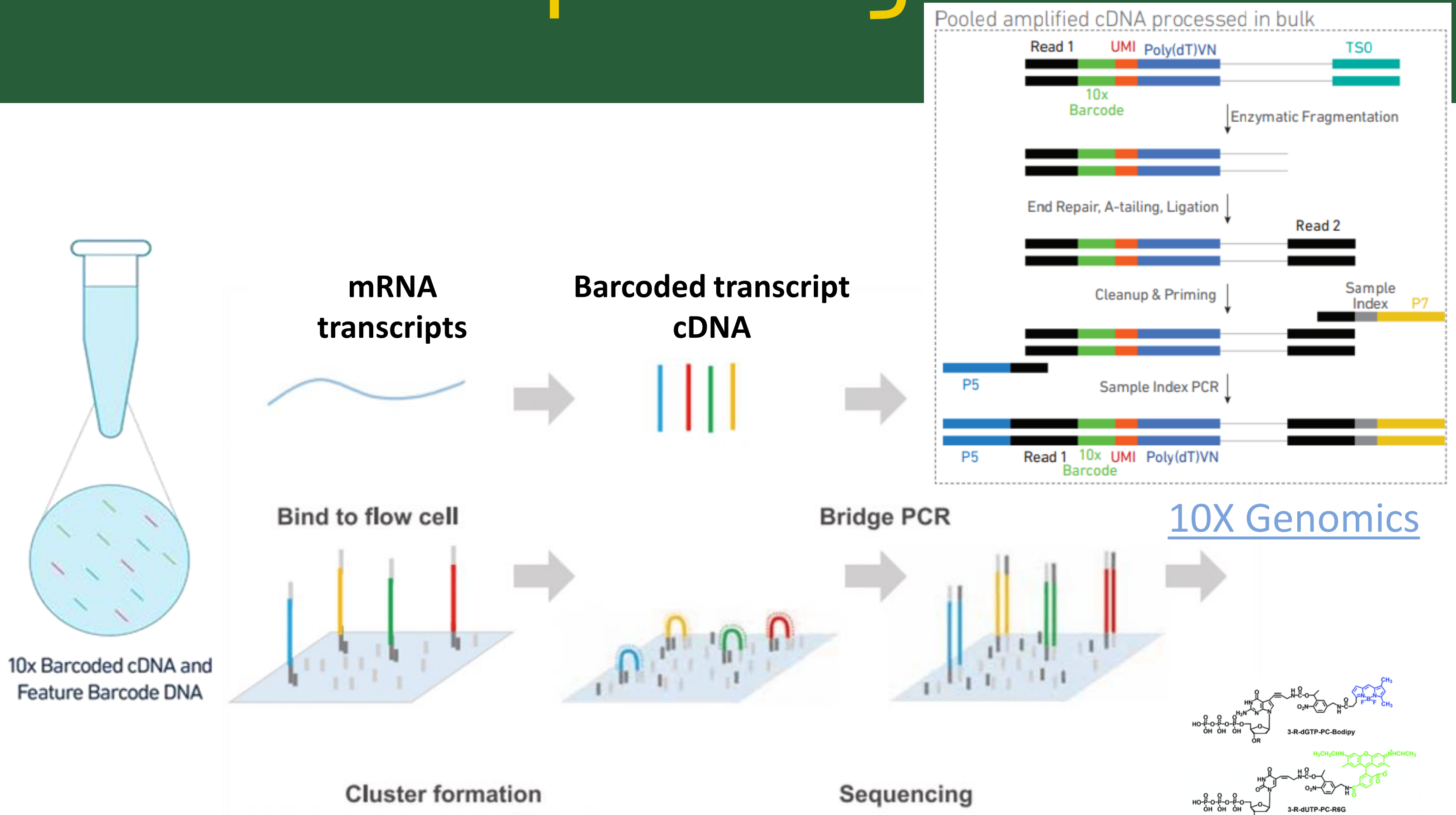
Library Preparation and Sequencing

Library prep:

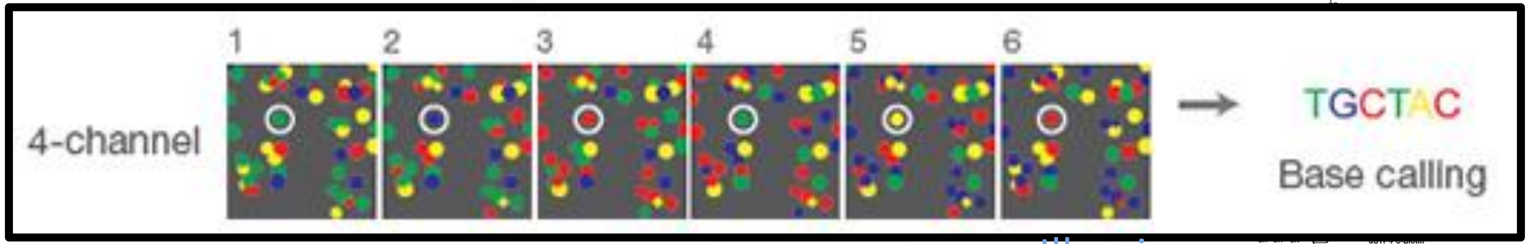
- Fragment and prepare the cDNA for sequencing.

Sequencing by synthesis

- Adapters to support bridge amplification
- Build clusters of identical fragments
- Change the nucleotides to fluorescent tagged versions with chemical stops
    - Clusters emit the fluorescence of the current nucleotide
    - Chemically cleave the fluorescent and repeat



10X Genomics



Primers on Next Gen Sequencing (illumina)
https://www.youtube.com/watch?v=fCd6B5HRaZ8

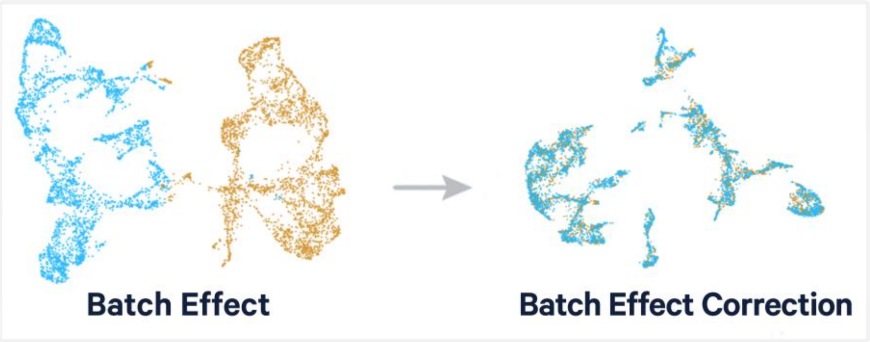
Illumina
National Center for Genome Analysis Support Blog

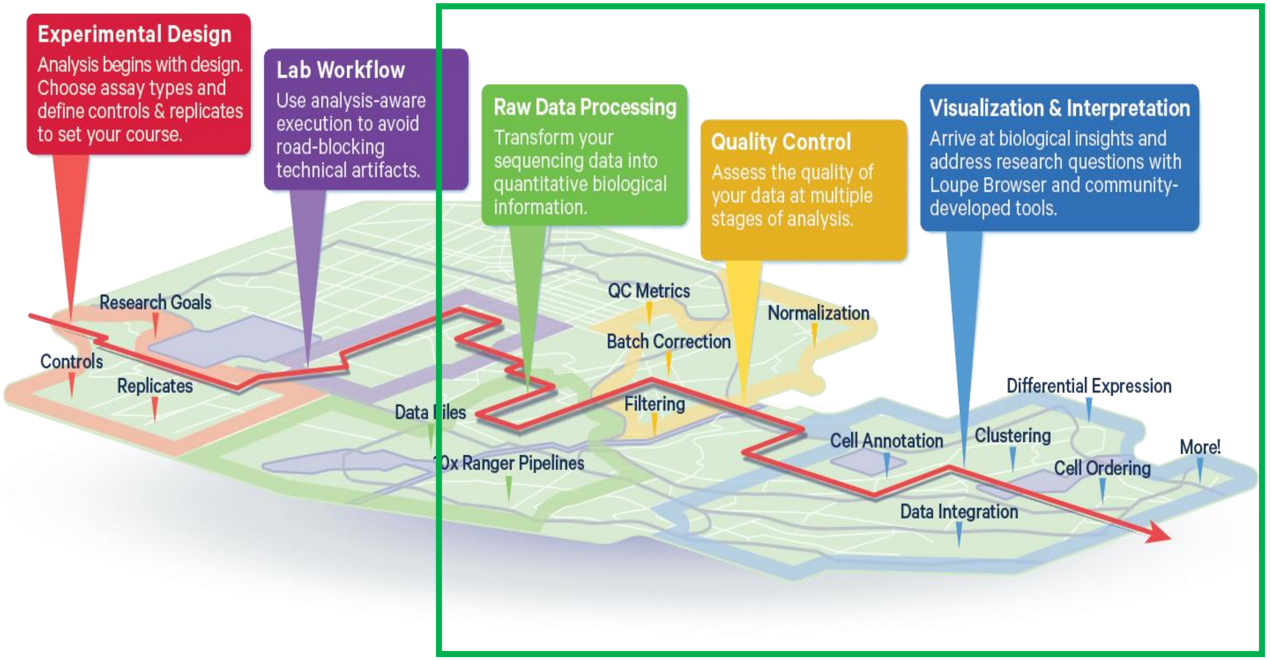
Data Processing and Quality Control

Once Sequencing is complete, still a long way to go!

Quality control and downstream processing is a huge part of scRNAseq

- Align sequences to identify genes
- De-multiplex all barcodes (samples, cells, UMI) and to create gene expression matrices
- Remove dead cells
- Empty Droplet Detection
- Adjust for ambient RNA
- Correct batch effects
  - Unsupervised clustering can be heavily impacted by batch effects before correction and aggregation





Experimental Design Analysis begins with design. Choose assay types and define controls & replicates to set your course.

Lab Workflow Use analysis-aware execution to avoid road-blocking technical artifacts.

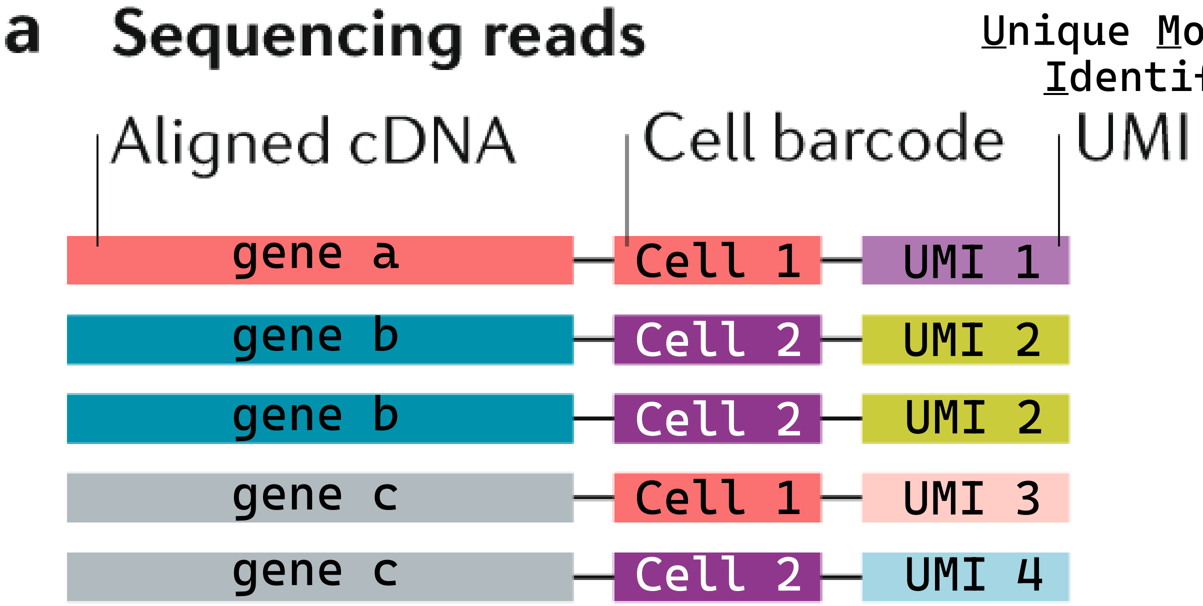
Raw Data Processing Transform your sequencing data into quantitative biological information.

Quality Control Assess the quality of your data at multiple stages of analysis.

Visualization & Interpretation Arrive at biological insights and address research questions with Loupe Browser and community-developed tools.

10X Genomics

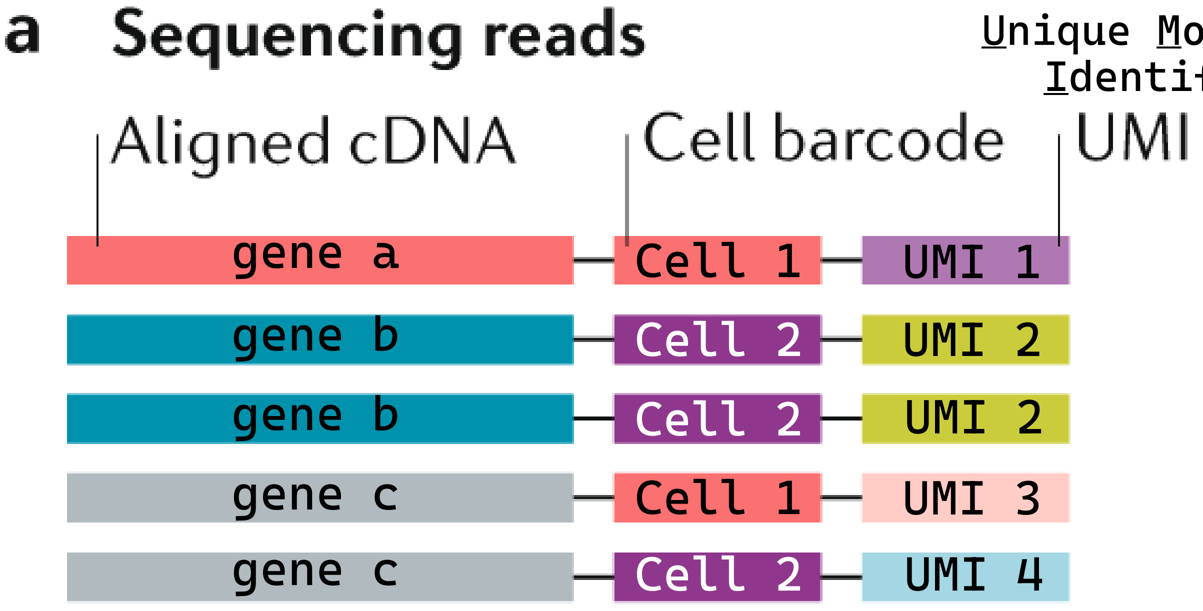
Sequence Alignment and Binning



Unique Molecule Identifier

Wu & Zhang Nature 2020

Sequence Alignment and Binning



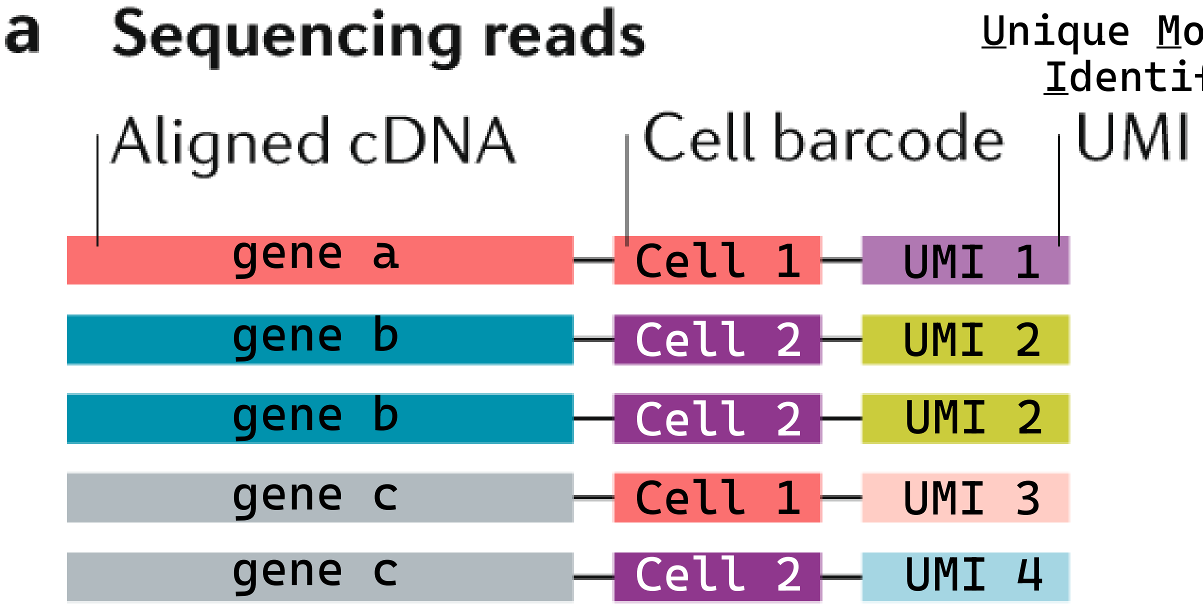
Unique Molecule Identifier

cDNA fragment is aligned against a reference genome

- **3 genes detected**

Wu & Zhang Nature 2020

Sequence Alignment and Binning



Unique Molecule Identifier

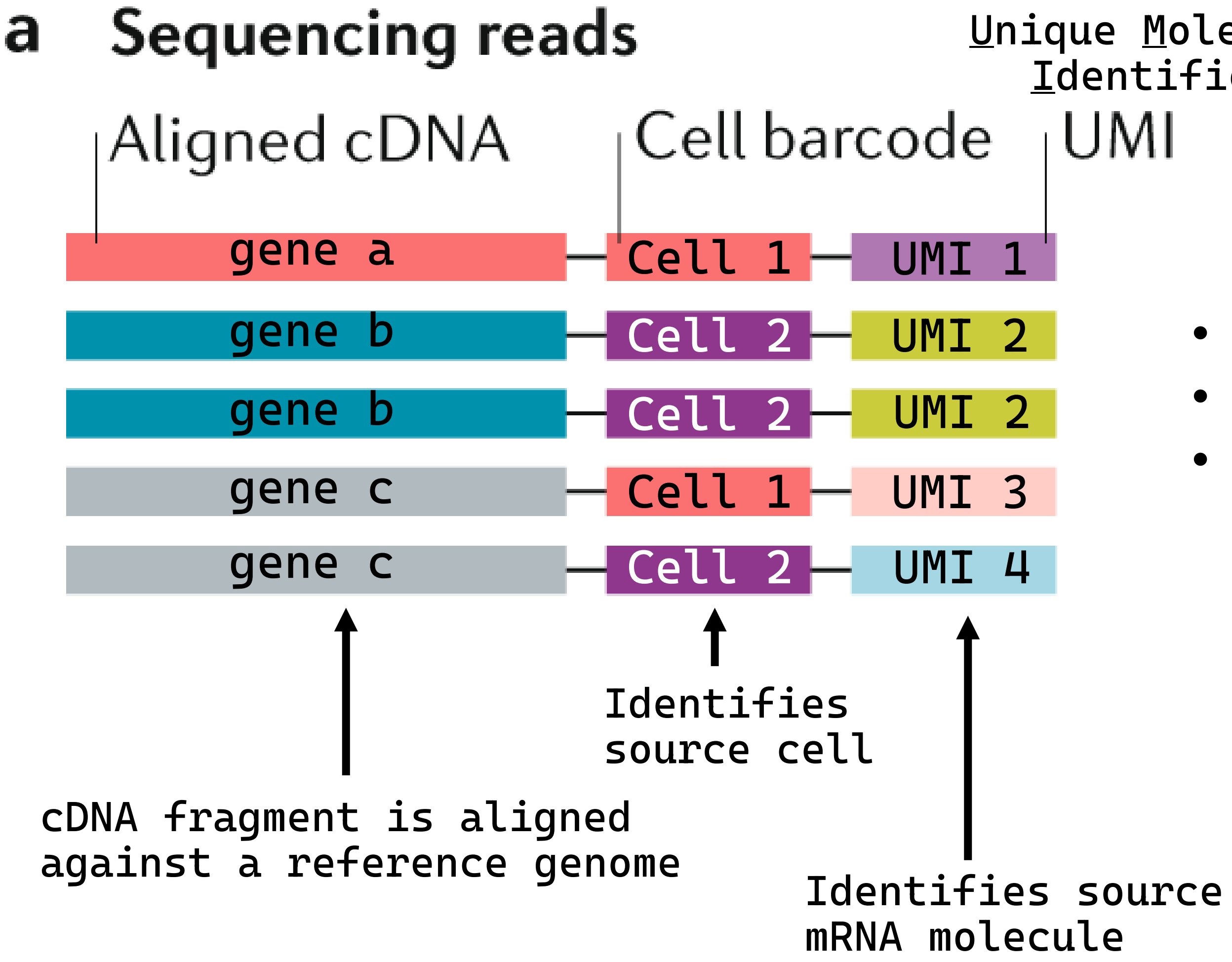
Identifies source cell

cDNA fragment is aligned against a reference genome

- **3 genes detected**
- **2 cells detected**

Wu & Zhang Nature 2020

Sequence Alignment and Binning



Unique Molecule Identifier

Identifies source cell

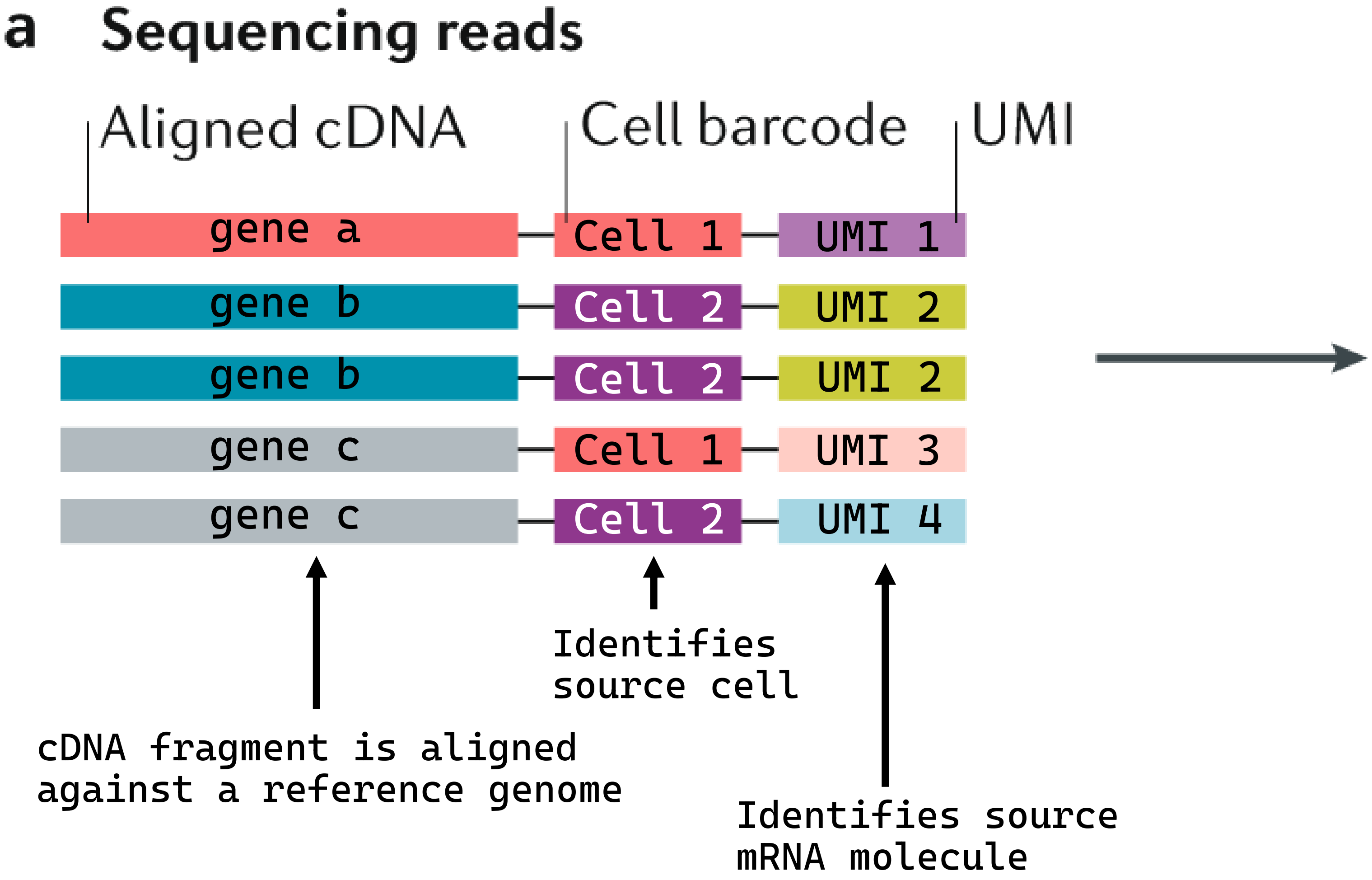
cDNA fragment is aligned against a reference genome

Identifies source mRNA molecule

- **3 genes detected**
- **2 cells detected**
- **4 unique original transcript molecules**

Wu & Zhang Nature 2020

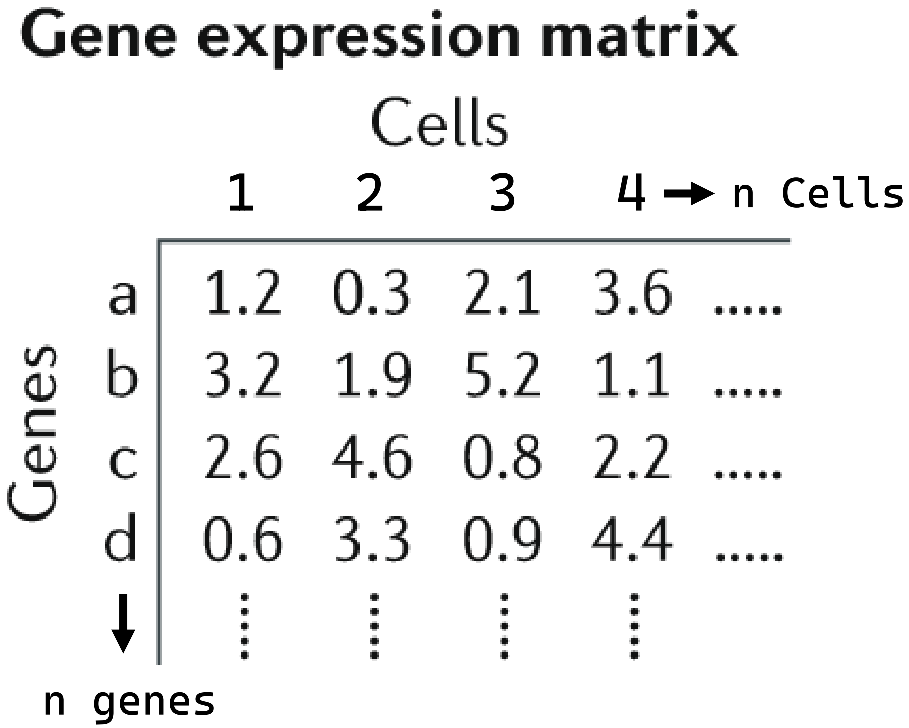
Sequence Alignment and Binning



cDNA fragment is aligned against a reference genome

Identifies source cell

Identifies source mRNA molecule



Gene expression matrix

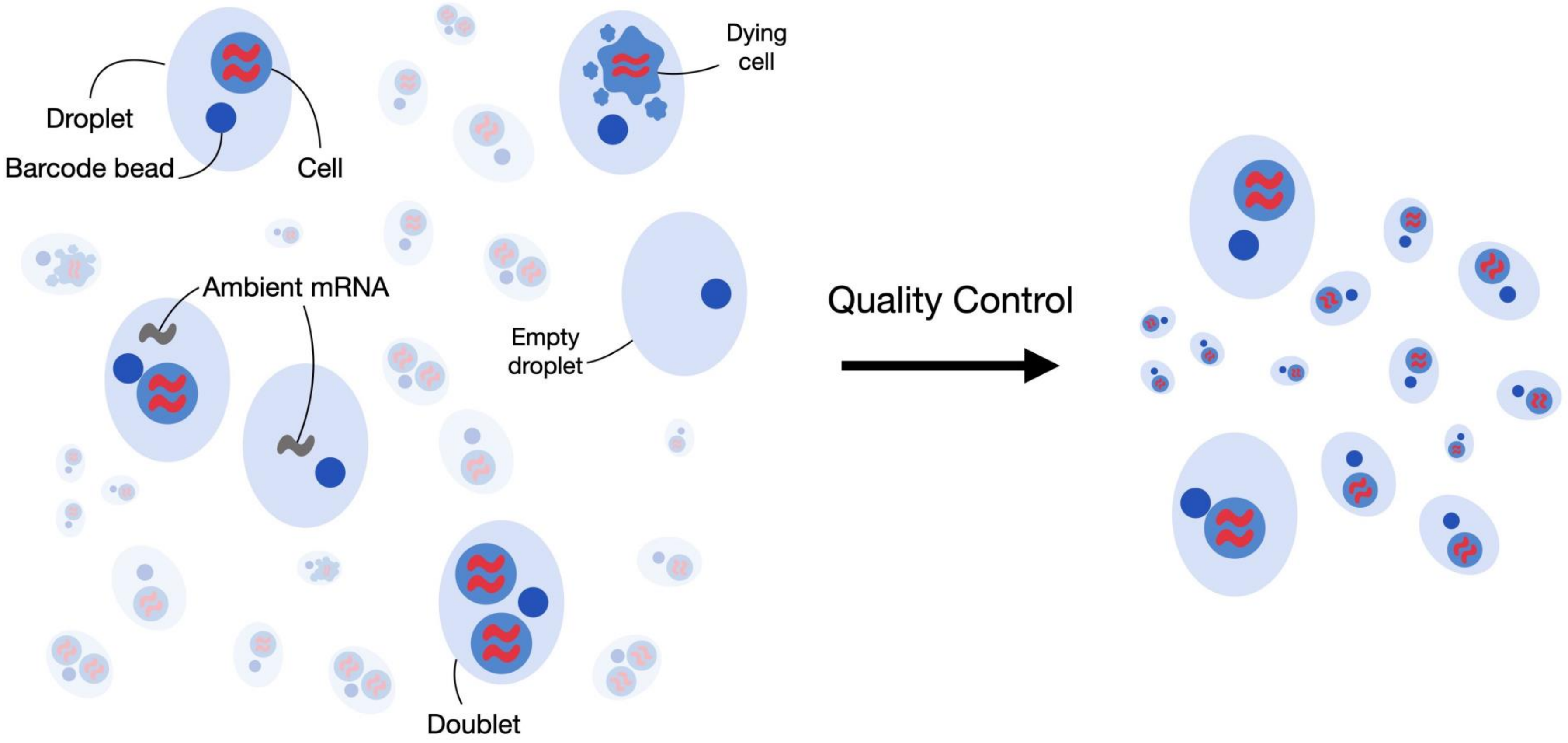
Genes \ Cells	1	2	3	4 → n Cells
a	1.2	0.3	2.1	3.6 .....
b	3.2	1.9	5.2	1.1 .....
c	2.6	4.6	0.8	2.2 .....
d	0.6	3.3	0.9	4.4 .....
↓ n genes	⋮	⋮	⋮	⋮

20 000 gene dimensions!!!
At 10 000 cells/ sample:
500 million data points per sample!!!

Wu & Zhang Nature 2020

Quality Control

Many scenarios of imperfect droplets that can affect your data!



sc-best-practices.org

Single Cell Gene Expression

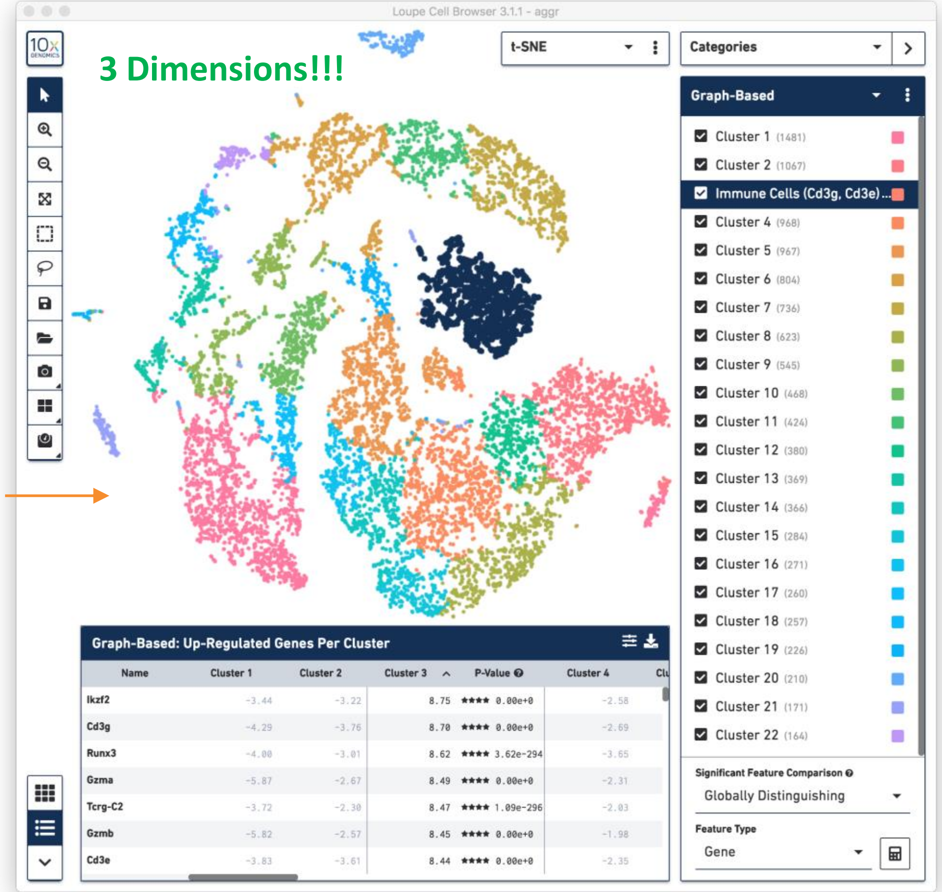
Dimension Reduction

- Very difficult to visualize thousands of dimensions of data at once
- Use fancy stats and data science techniques (clustering) to find patterns and associations within the data to group
- Supervised (Semi) Clustering
    - Takes user input on cluster number
- Unsupervised
    - Will make as many clusters as it thinks exists, depending on variance limits for specific algorithms

Gene Count Matrices

	Cell 1	**Cell 2**	**Cell *x* -> *10000***
Gene 1	**3**	**170**	**…**
Gene 2	**500**	**30**	**…**
Gene *x* -> *20000*	**…**	**…**	**…**

20 000 genes *10 000 cells/ sample: 500 million data points per sample!!!



Loupe Cell Browser 3.1.1 - aggr

3 Dimensions!!!

☑ Cluster 1 (1481)
☑ Cluster 2 (1067)
☑ Immune Cells (Cd3g, Cd3e)...
☑ Cluster 4 (968)
☑ Cluster 5 (967)
☑ Cluster 6 (804)
☑ Cluster 7 (736)
☑ Cluster 8 (623)
☑ Cluster 9 (545)
☑ Cluster 10 (468)
☑ Cluster 11 (424)
☑ Cluster 12 (380)
☑ Cluster 13 (369)
☑ Cluster 14 (366)
☑ Cluster 15 (284)
☑ Cluster 16 (271)
☑ Cluster 17 (260)
☑ Cluster 18 (257)
☑ Cluster 19 (226)
☑ Cluster 20 (210)
☑ Cluster 21 (171)
☑ Cluster 22 (164)

Significant Feature Comparison: Globally Distinguishing

Feature Type: Gene

Graph-Based: Up-Regulated Genes Per Cluster

Name	Cluster 1	Cluster 2	Cluster 3	P-Value	Cluster 4
Ikzf2	-3.44	-3.22	8.75	★★★★ 0.00e+0	-2.58
Cd3g	-4.29	-3.76	8.70	★★★★ 0.00e+0	-2.69
Runx3	-4.00	-3.01	8.62	★★★★ 3.62e-294	-3.65
Gzma	-5.87	-2.67	8.49	★★★★ 0.00e+0	-2.31
Tcrg-C2	-3.72	-2.30	8.47	★★★★ 1.09e-296	-2.03
Gzmb	-5.82	-2.57	8.45	★★★★ 0.00e+0	-1.98
Cd3e	-3.83	-3.61	8.44	★★★★ 0.00e+0	-2.35

Not necessary to understand the underlying data science mathematics in order to understand what the algorithm functions. This is an excellent resource for learning about data science and machine learning techniques without needing any coding or advanced mathematic knowledge:

https://machinelearningmastery.com/start-here/#algorithms

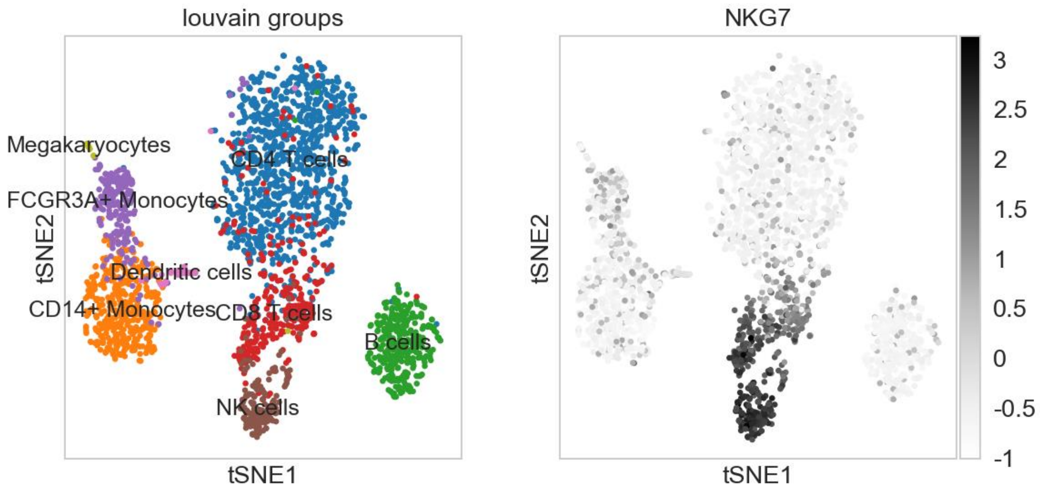
10X Genomics

Single Cell Gene Expression

Now can do some really interesting analyses!

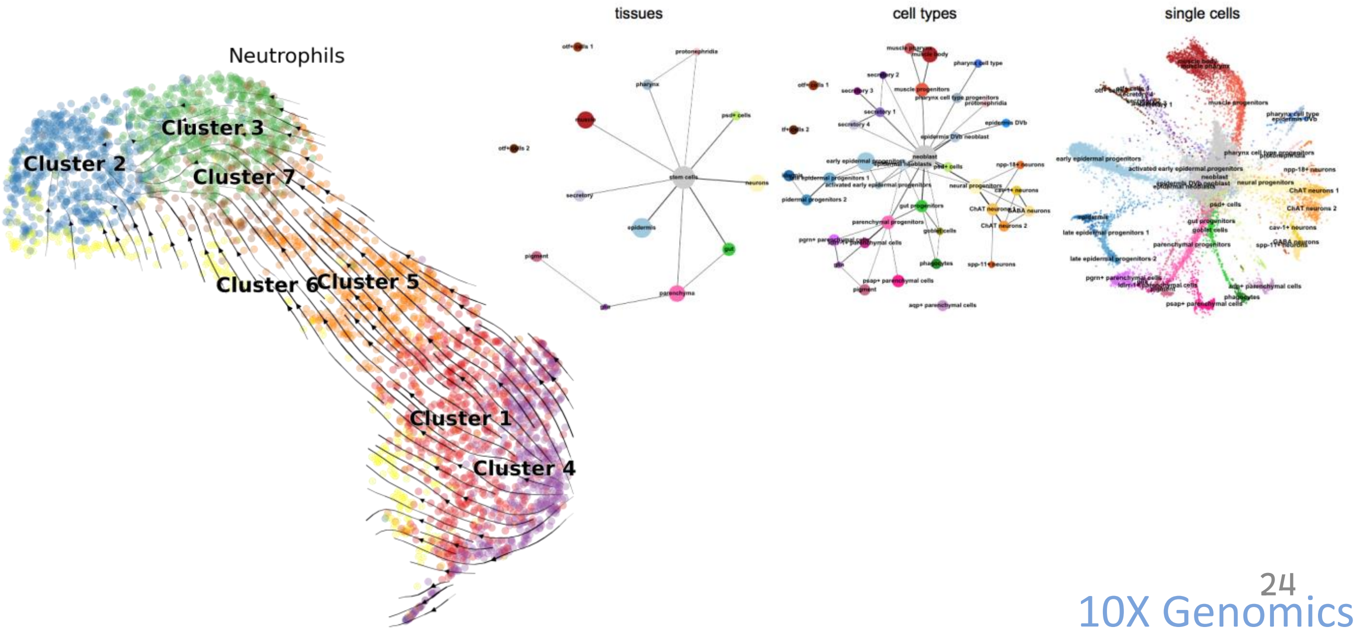
Analysis

- Clustering
- Differential expression
- Associated genes
- Endless data mining
    - Revisit previous experiments with new genes of interest



Trajectory Analysis

- What direction are progenitor cells differentiating to, in what proportions? How is this affected by experimental conditions?



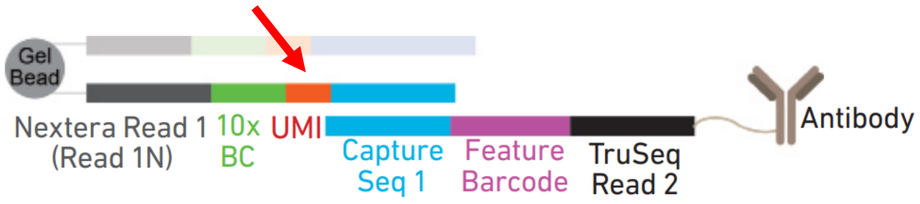
[10X Genomics](#)

Additional Single Cell Assays

Protein Expression

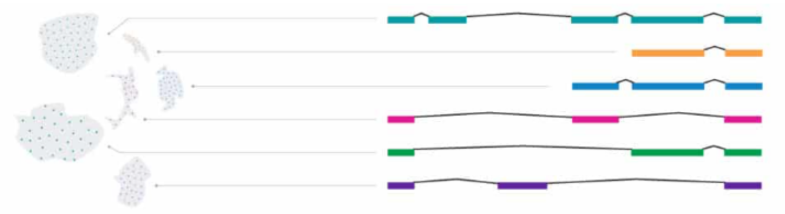
- Tag cell surface moieties with antibodies bound to a Capture Sequence and Feature Barcode CITE-Seq)
- Acts in place of a transcript once in the partitioned barcoding reaction

UMI – Allows Quantification



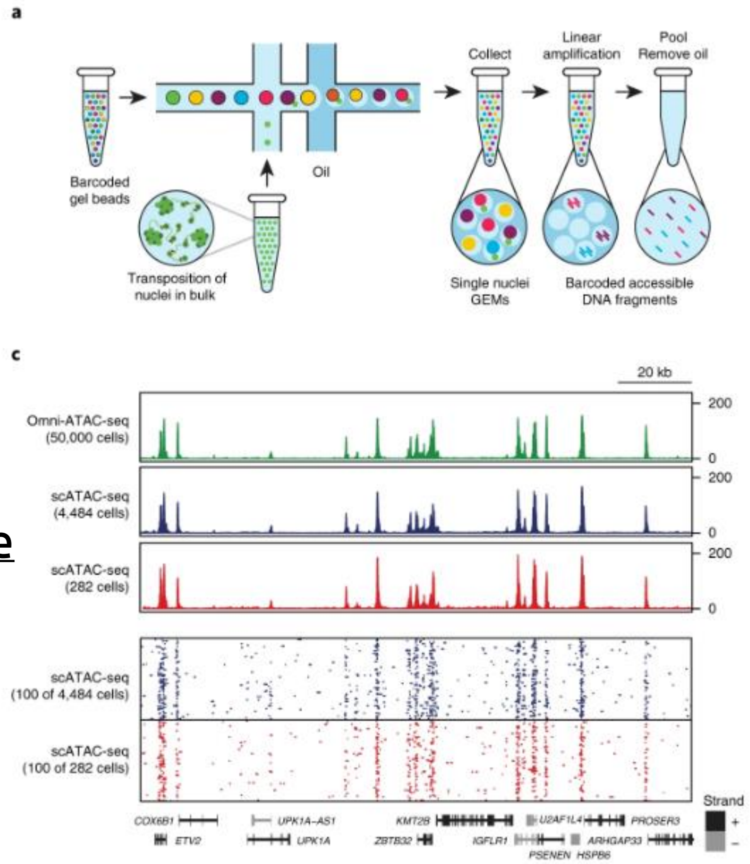
B Cell and T Cell receptor sequencing

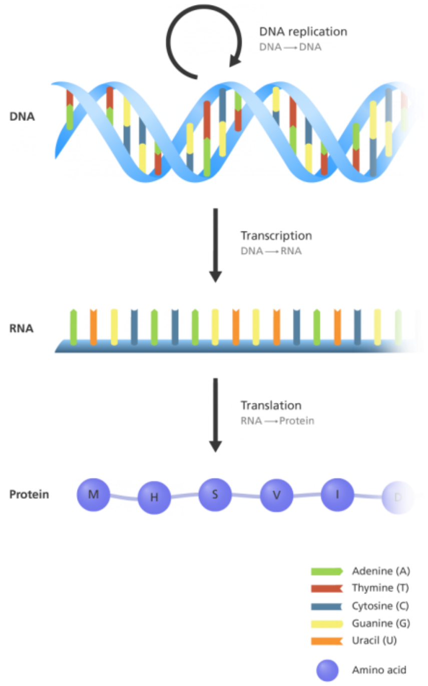
Full length transcript sequencing for splice variants (PacBio long read sequencing)



DNA availability

Transposase-accessible chromatin with sequencing (ATAC-Seq)

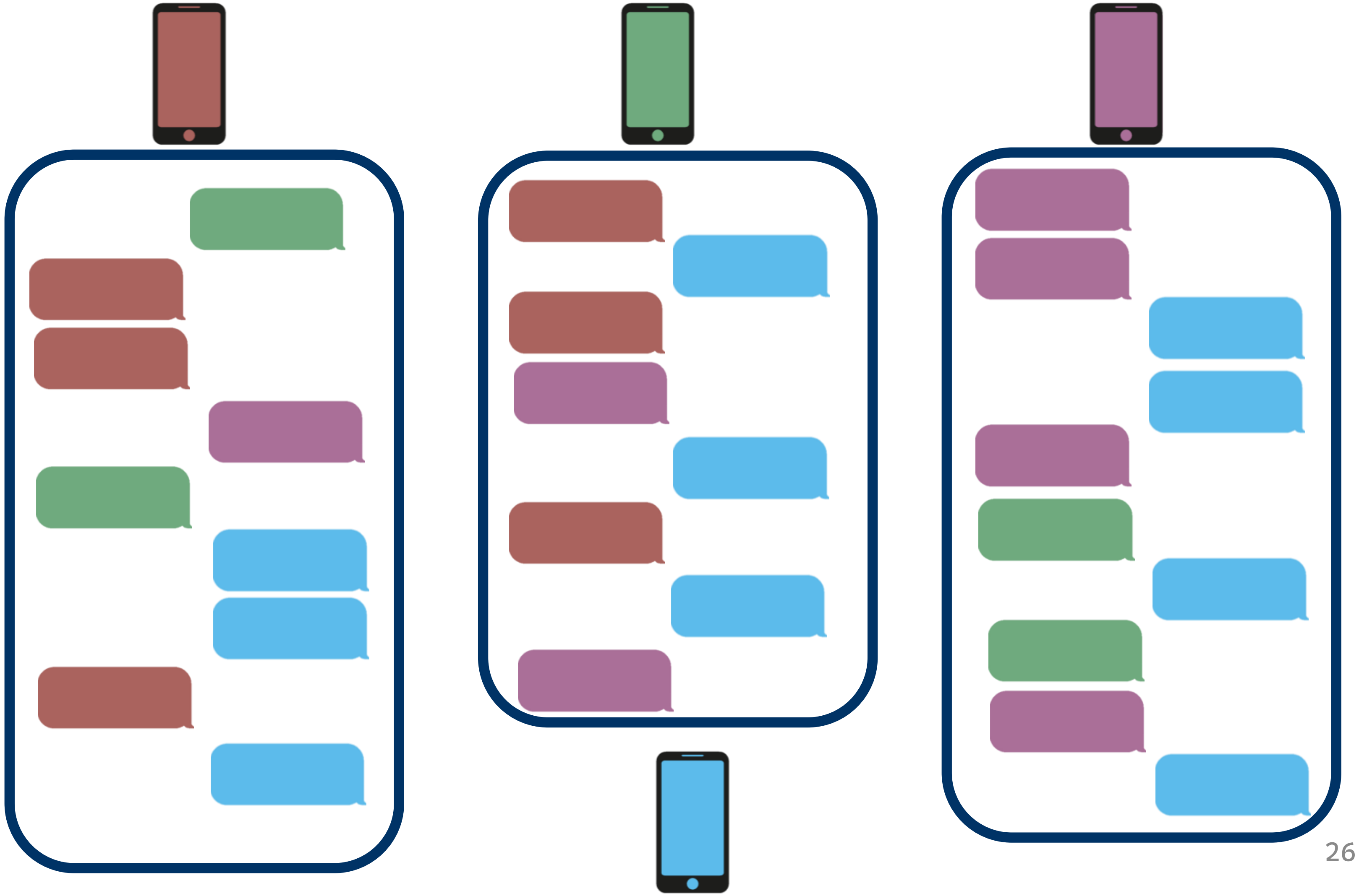




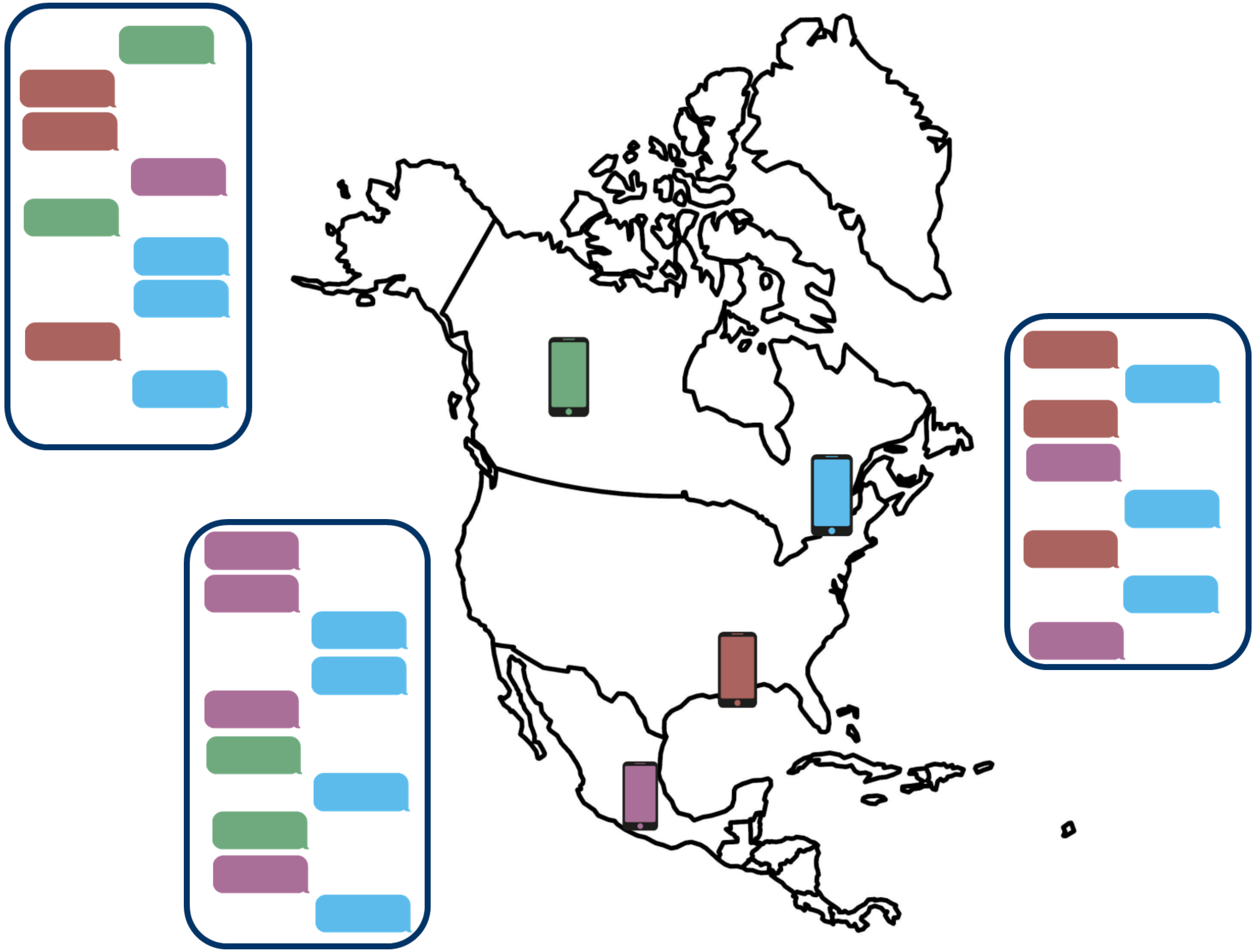
10X Genomics

Single Cell Sequencing

Now we can see which cells are contributing to the total gene expression conversation!



Single Cell Sequencing

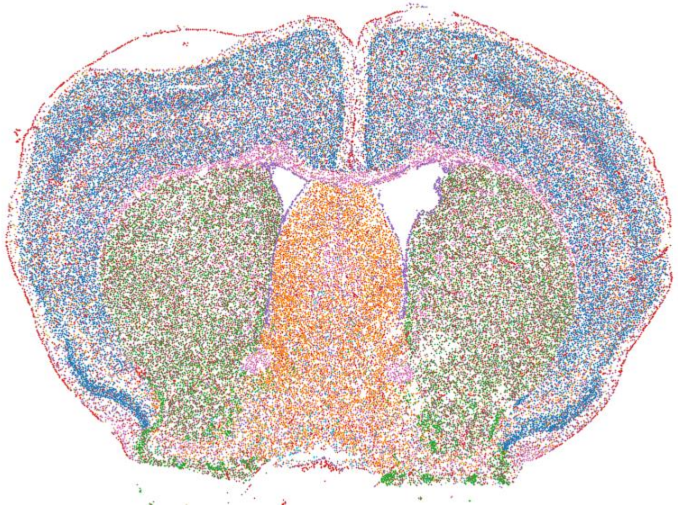


Now we can see which cells are contributing to the total gene expression conversation!

But what if we add spatial context as well!?

Spatial Transcriptomics

Bulk and single cell sequencing

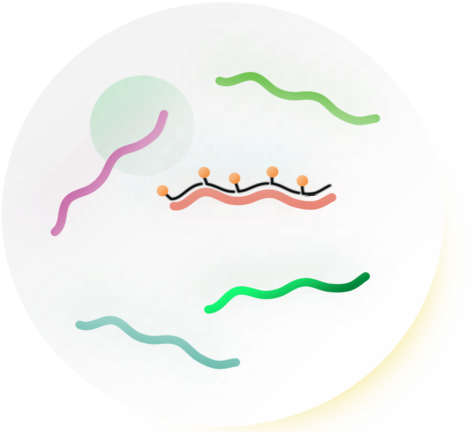


Spatial genomics with true single cell resolution

Next big field in genomics

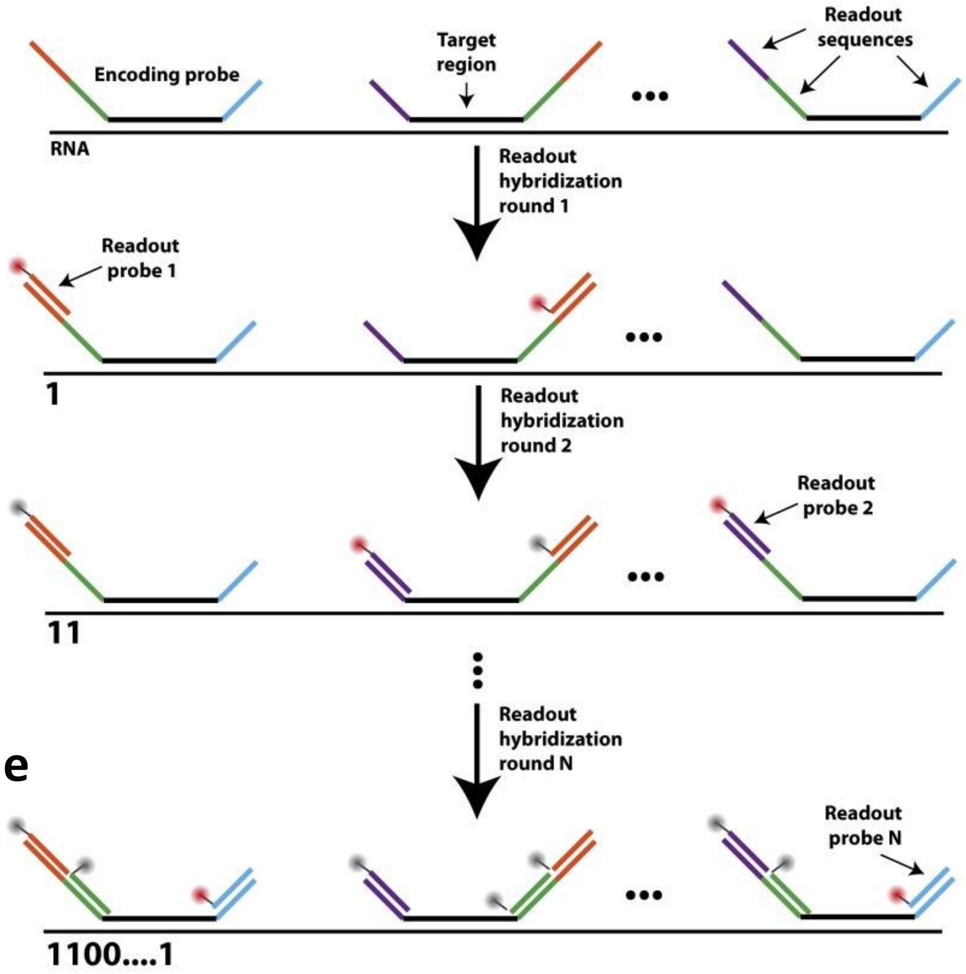
Many technologies available

Fluorescence In-Situ Hybridization



- Single Molecule In-Situ Hybridization (smFISH)
- 1 gene transcript with a specific fluorophore captured with a microscope

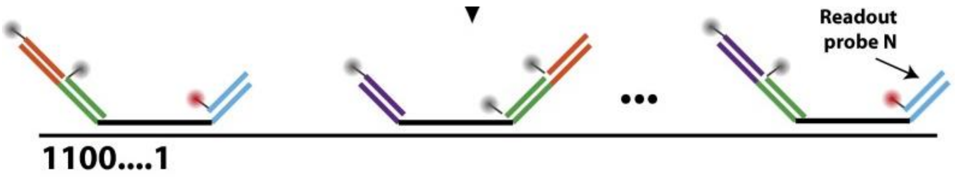
- Multiplexed Error-Robust (merFISH)
- Hundreds (thousands) of gene targets
- Probes designed along the length of a gene with multiple readout sequences

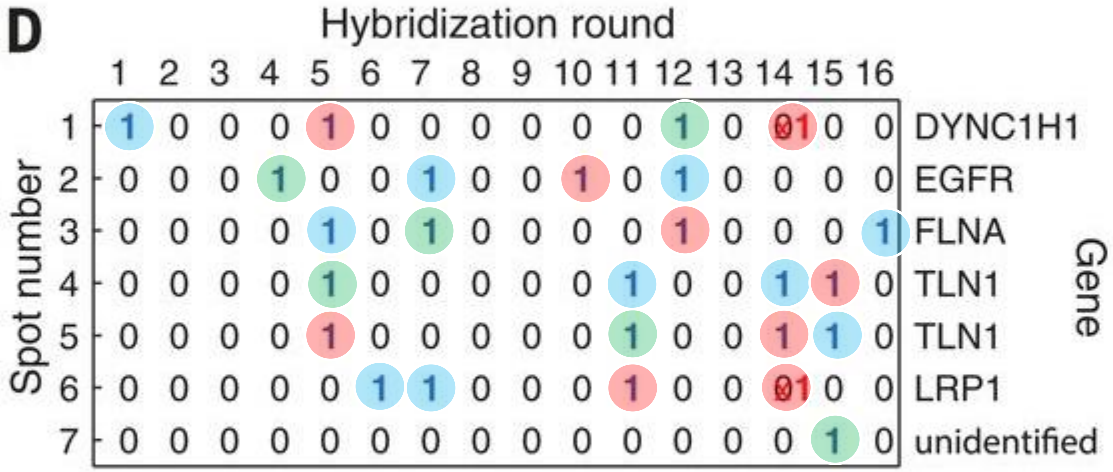


https://europepmc.org/article/MED/27625426

merFISH – How it works

- **Multiplexed Error-Robust Fluorescences In-Situ Hybridization**
- **Spots are read in sequence**

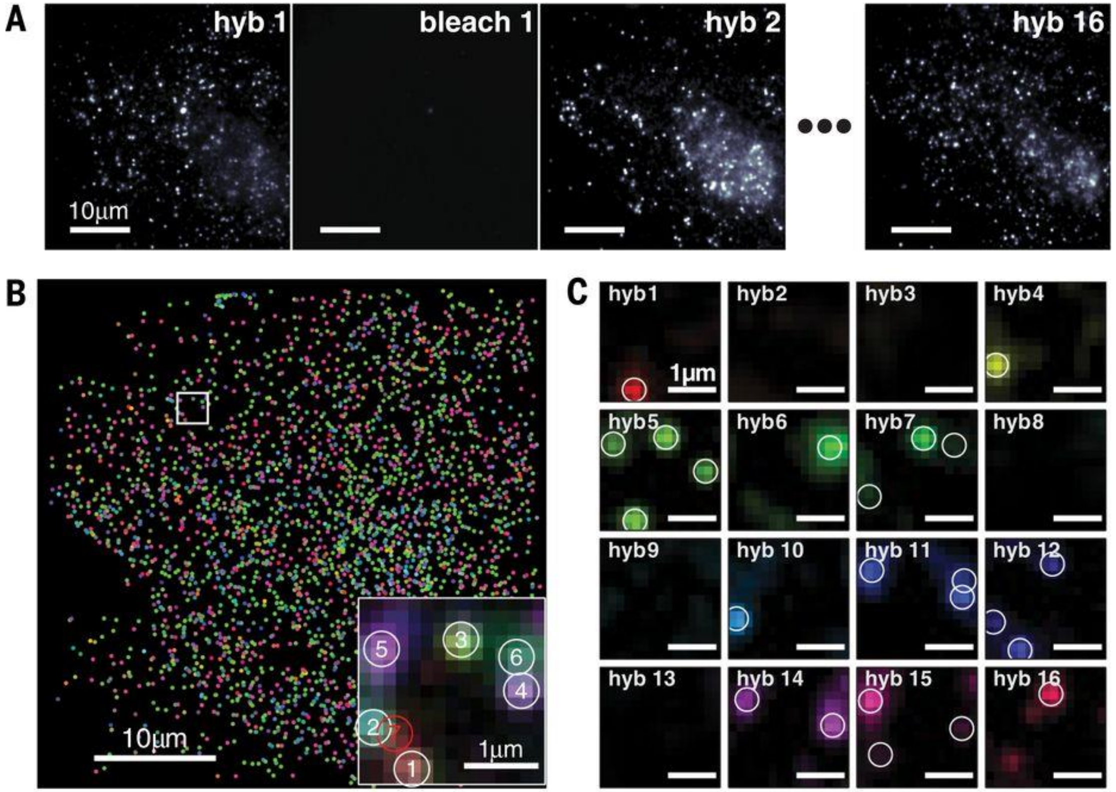




Hybridization round

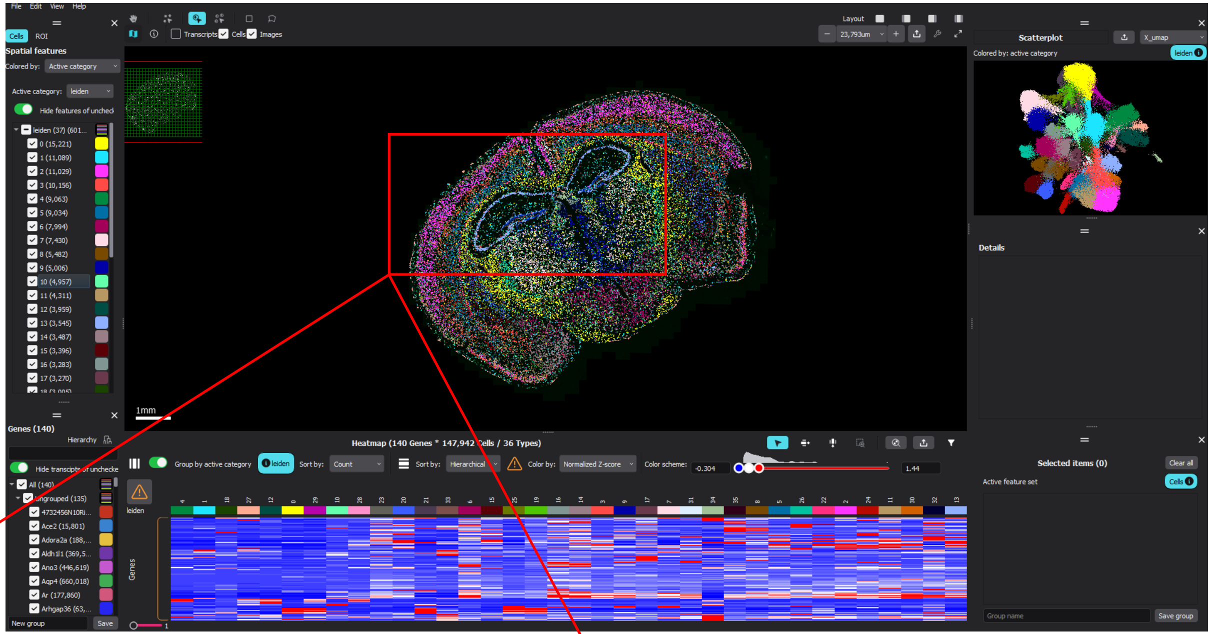
Spot number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Gene
1	1	0	0	0	1	0	0	0	0	0	0	1	0	~~0~~1	0	0	DYNC1H1
2	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	EGFR
3	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	FLNA
4	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	TLN1
5	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	TLN1
6	0	0	0	0	0	1	1	0	0	0	1	0	0	~~0~~1	0	0	LRP1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	unidentified

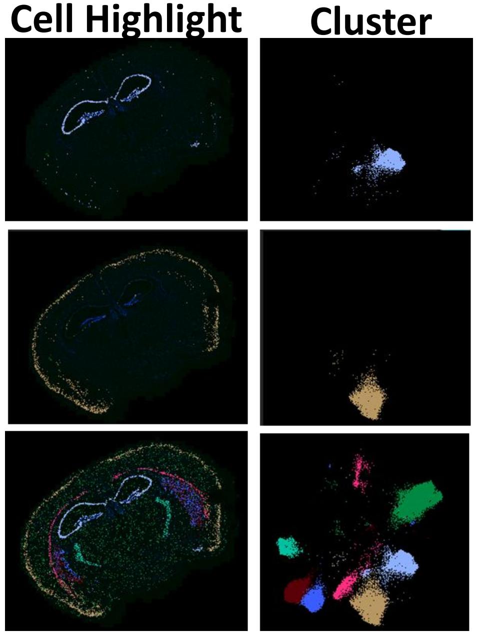
- 1 miss-call can be corrected

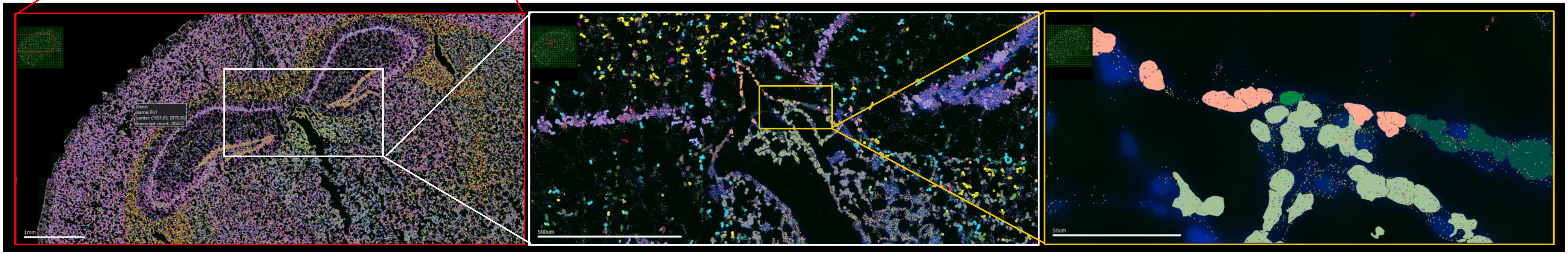


Chen et al, Science, 2015

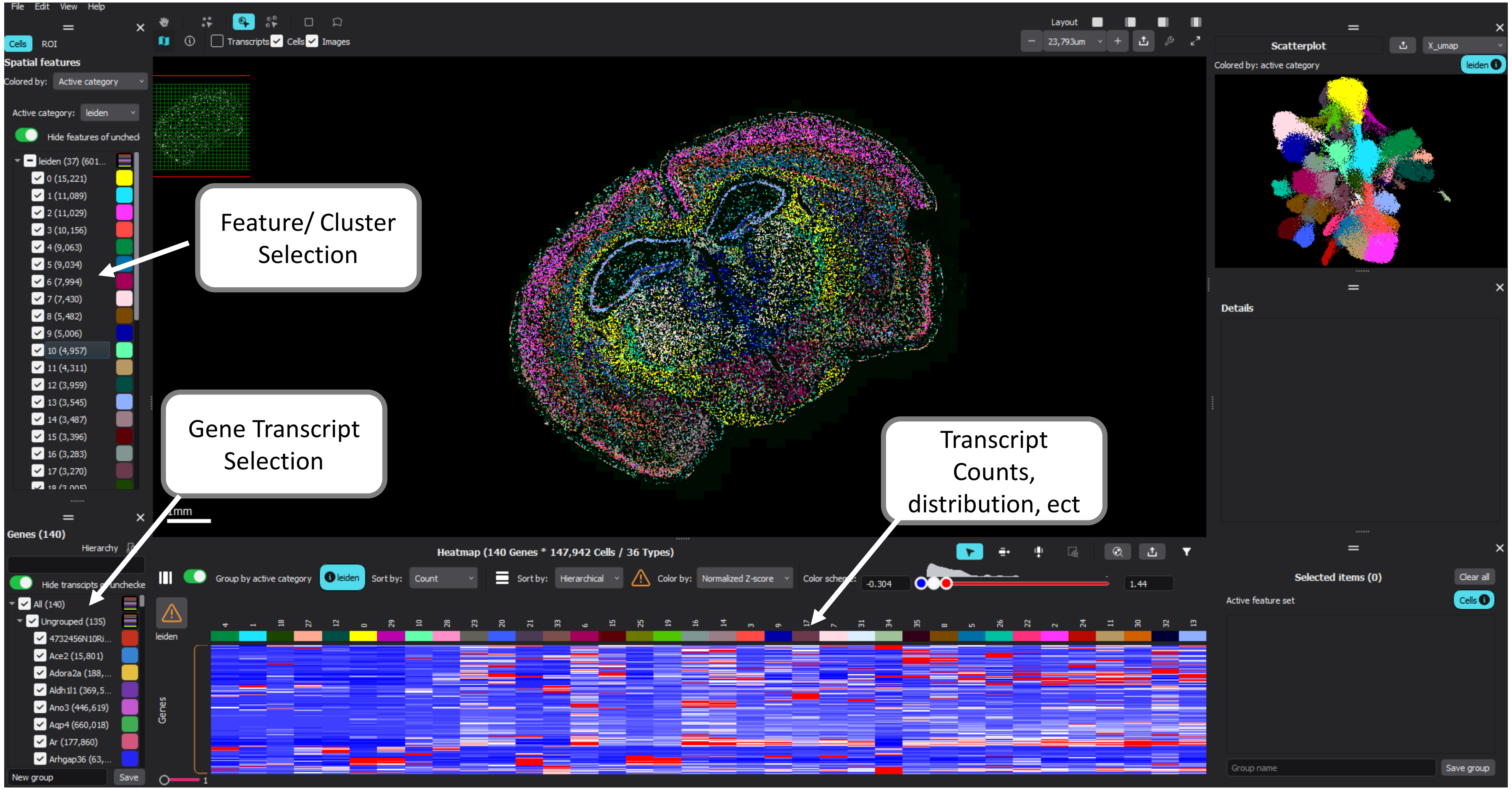
Spatial Transcriptomics







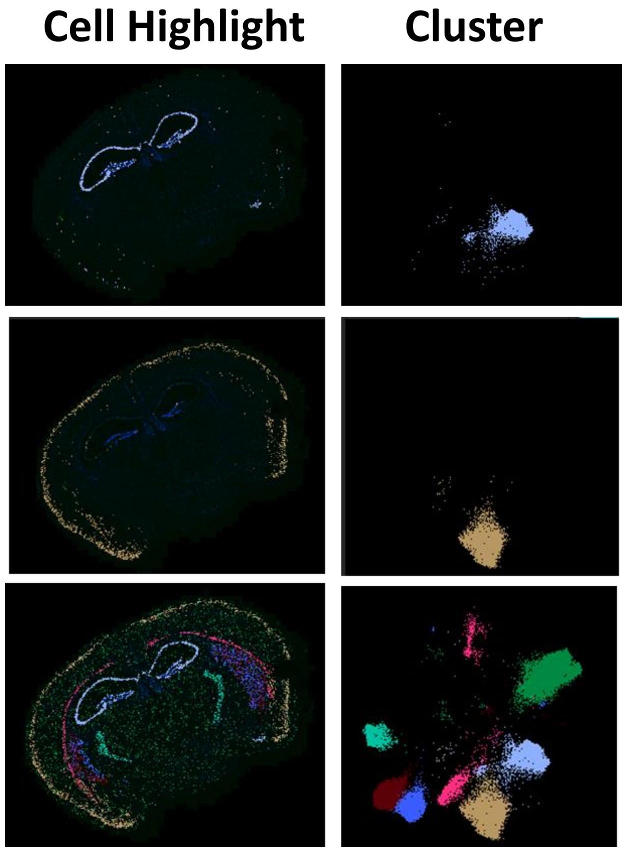
Spatial Transcriptomics



Feature/ Cluster Selection

Gene Transcript Selection

Transcript Counts, distribution, ect



THANKS!

ADVANCED CELL EXPLORATION CORE

MIKE WONG

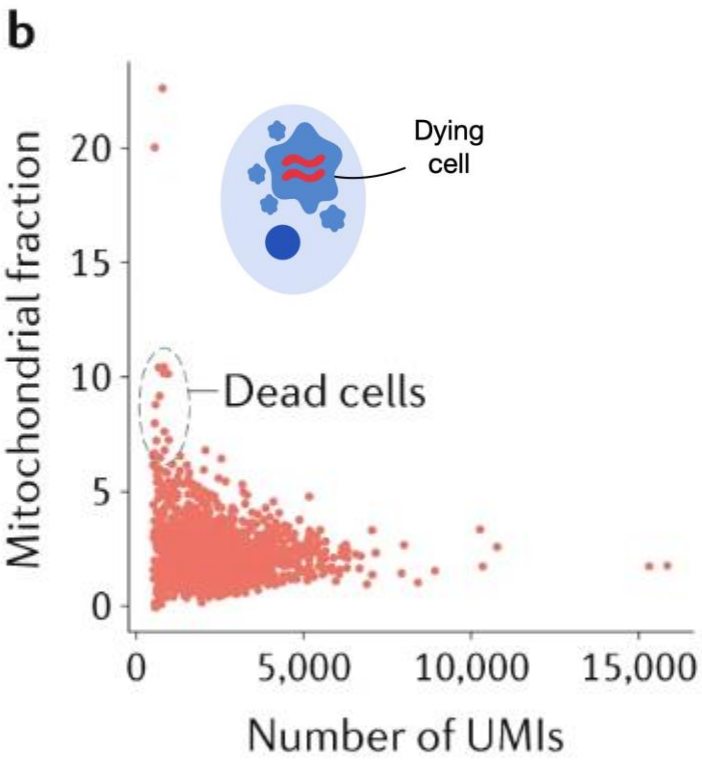
MEWONG@UALBERTA.CA

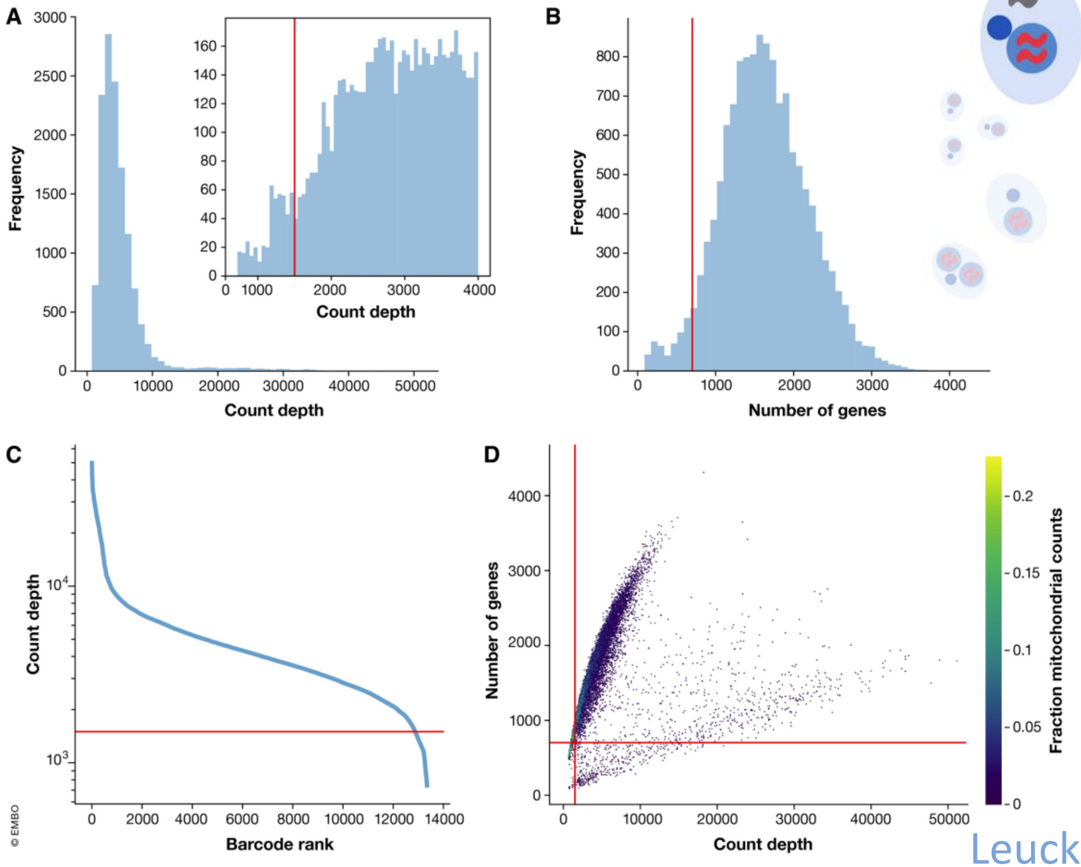
UNIVERSITY
OF ALBERTA

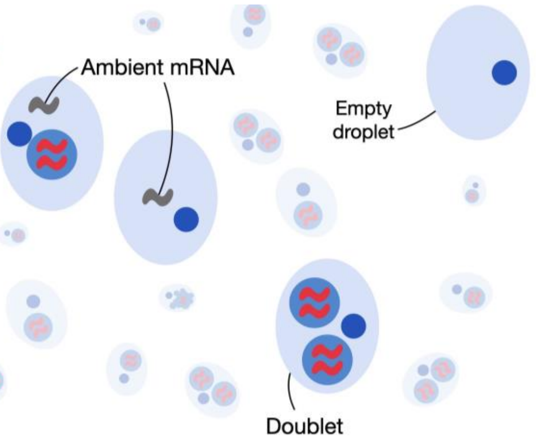
Doublet and Empty Droplet Removal

Gene counts, identified genes and mitochondrial gene expression all important

*Can be heavily impacted by specific biology of cell populations







Leucken and Theis 2019