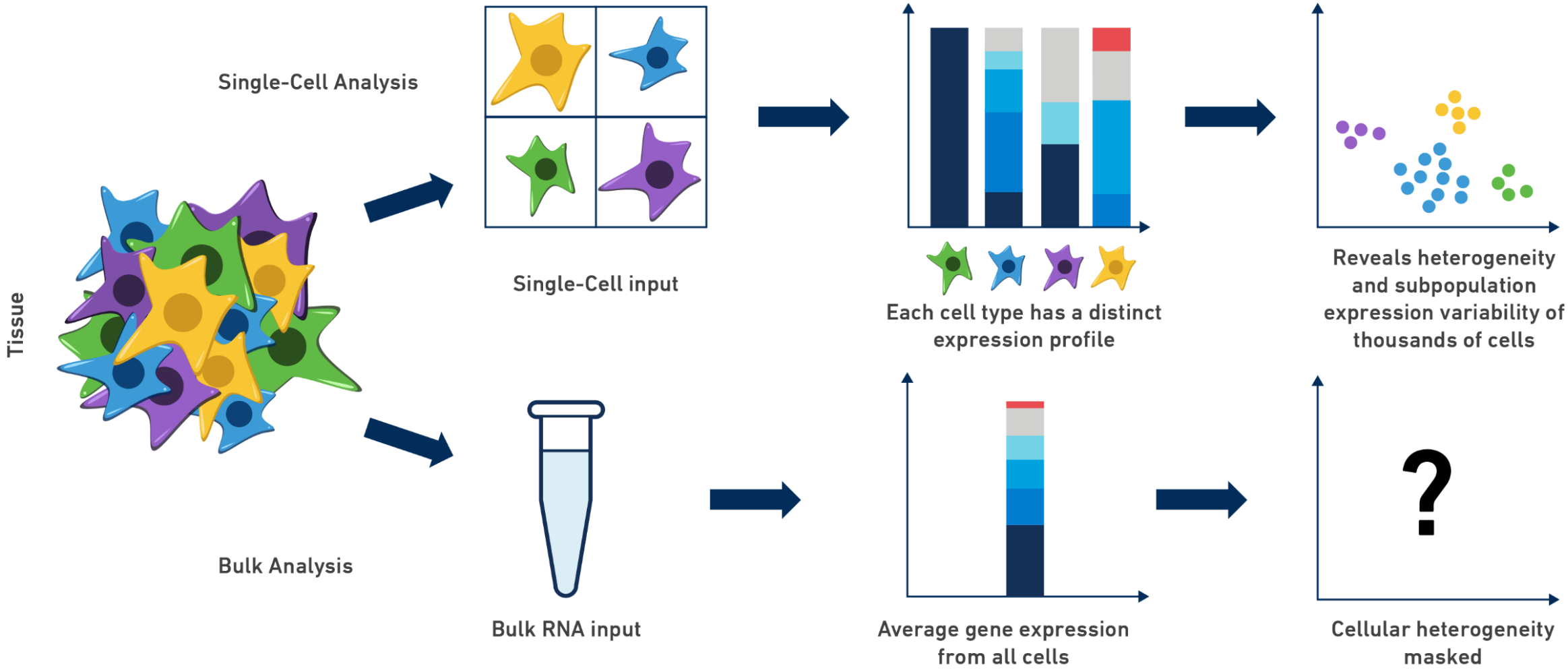
Single Cell Transcriptomics – a simple walkthrough

Adwait Sathe, PhD

McDermott Center Bioinformatics Lab

adwait.sathe@utsouthwestern.edu

RNASeq

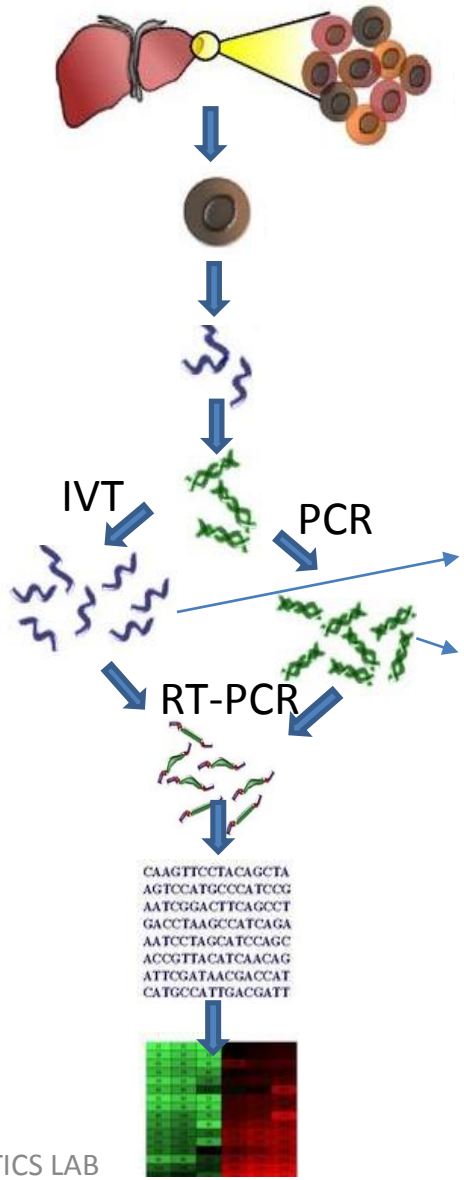


https://www.10xgenomics.com/solutions/single-cell/

RNASeq

Single cell RNASeq

Single cell RNA Sequencing Workflow



Dissociation of solid tissue

Isolate single cells
DROP-Seq, Smart-Seq etc.

RNA

cDNA

Amplified RNA

Amplified cDNA

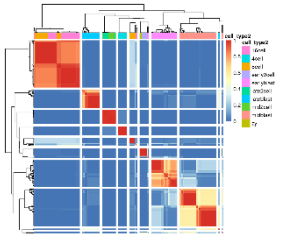
Sequencing library

Sequencing

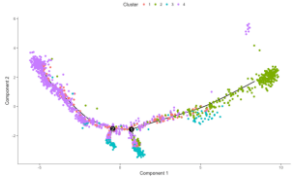
Single cell
expression profile

Novel applications!

Clustering and
Cell type identification



novel markers
for cell types



Constructing Single
Cell Trajectories

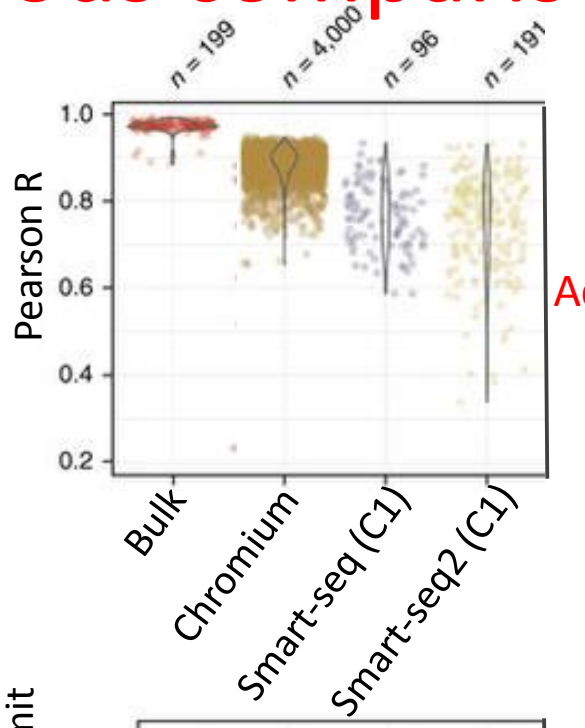
Single cell RNA Sequencing methods comparison

RNASeq
- Single cell RNASeq
- ScRNAseq methods

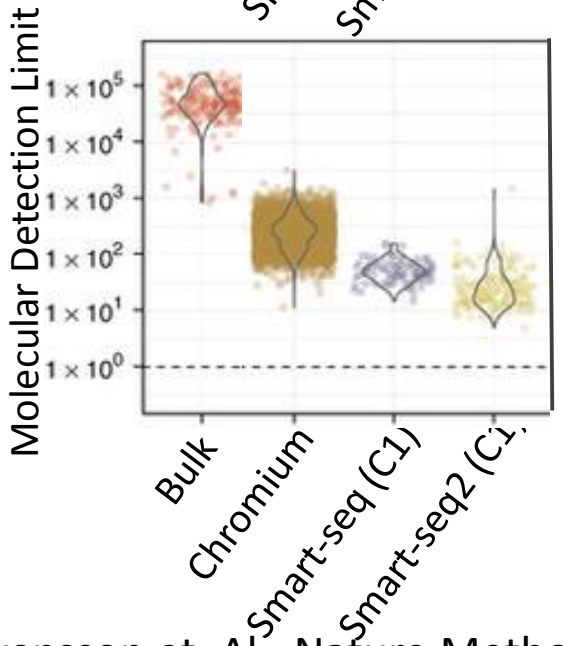
Chromium, 10X UMI – unique molecular identifiers
- 3’ counting (only 3’ of transcript sequenced)
- 8-10 bp UMI
- 10X barcoded gel beads
- Throughput (number of cells) – 10^4–10^5

Smart-seq/C1 microfluidic platform
- Full length
- No UMI
- Fluidigm C1
- Throughput (number of cells) – 10^2–10^3

Smart-seq2 most genes per cell among methods
- Full-length
- No UMI
- FACS
- Throughput (number of cells) – 10^2–10^3



Accuracy



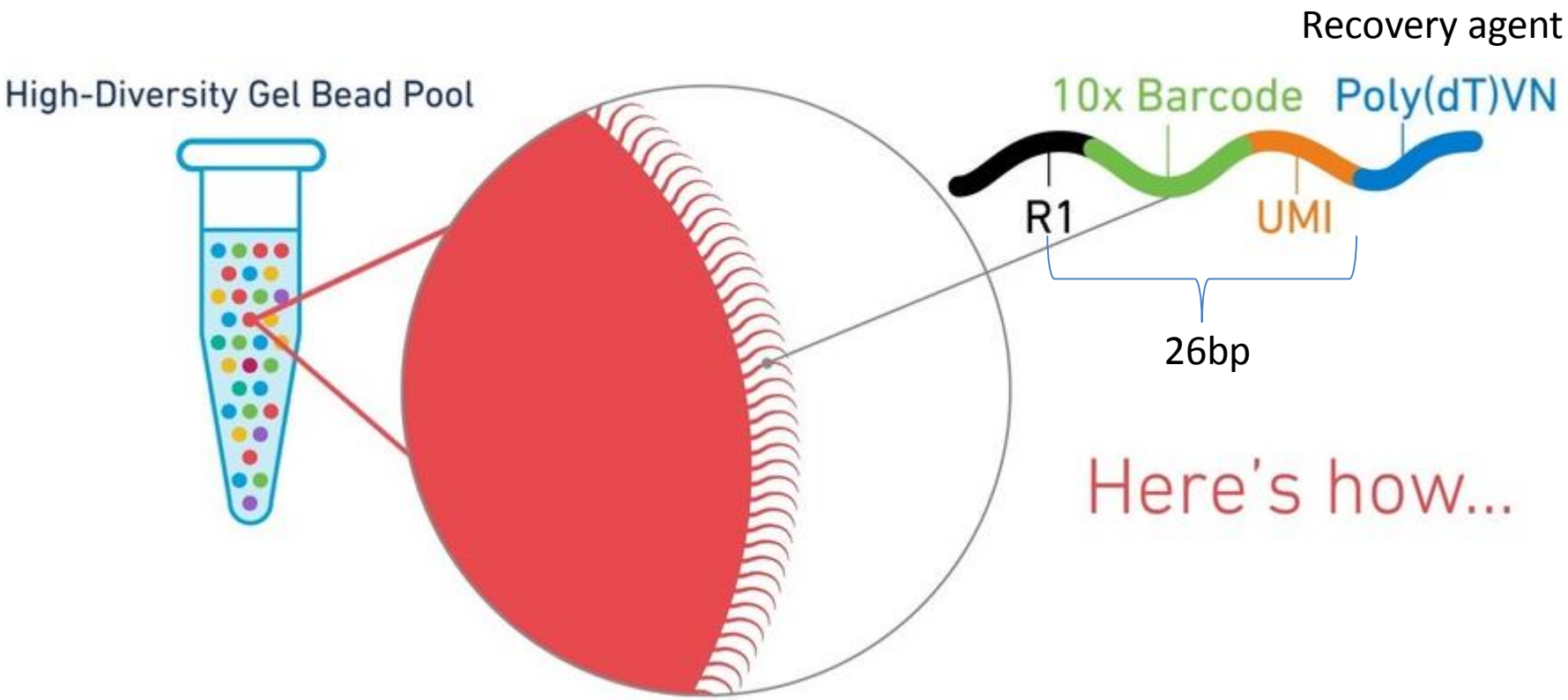
Sensitivity

Svensson et. Al., Nature Methods (2017)

RNASeq

- Single cell RNASeq
- ScRNAseq methods
- 10X Chromium

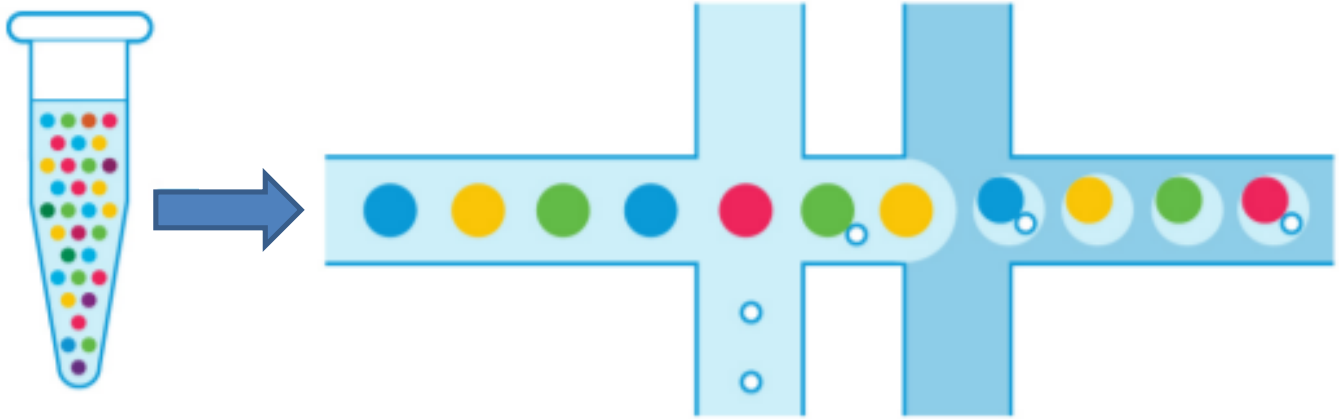
10X genomics protocol



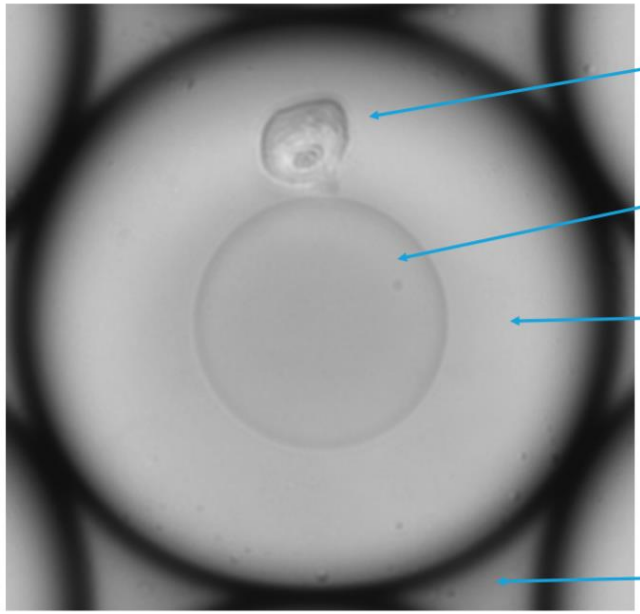
https://www.10xgenomics.com/solutions/single-cell/

10X genomics protocol

- RNASeq
  - Single cell RNASeq
  - ScRNAseq methods
  - 10X Chromium



Cells Lysis & RT

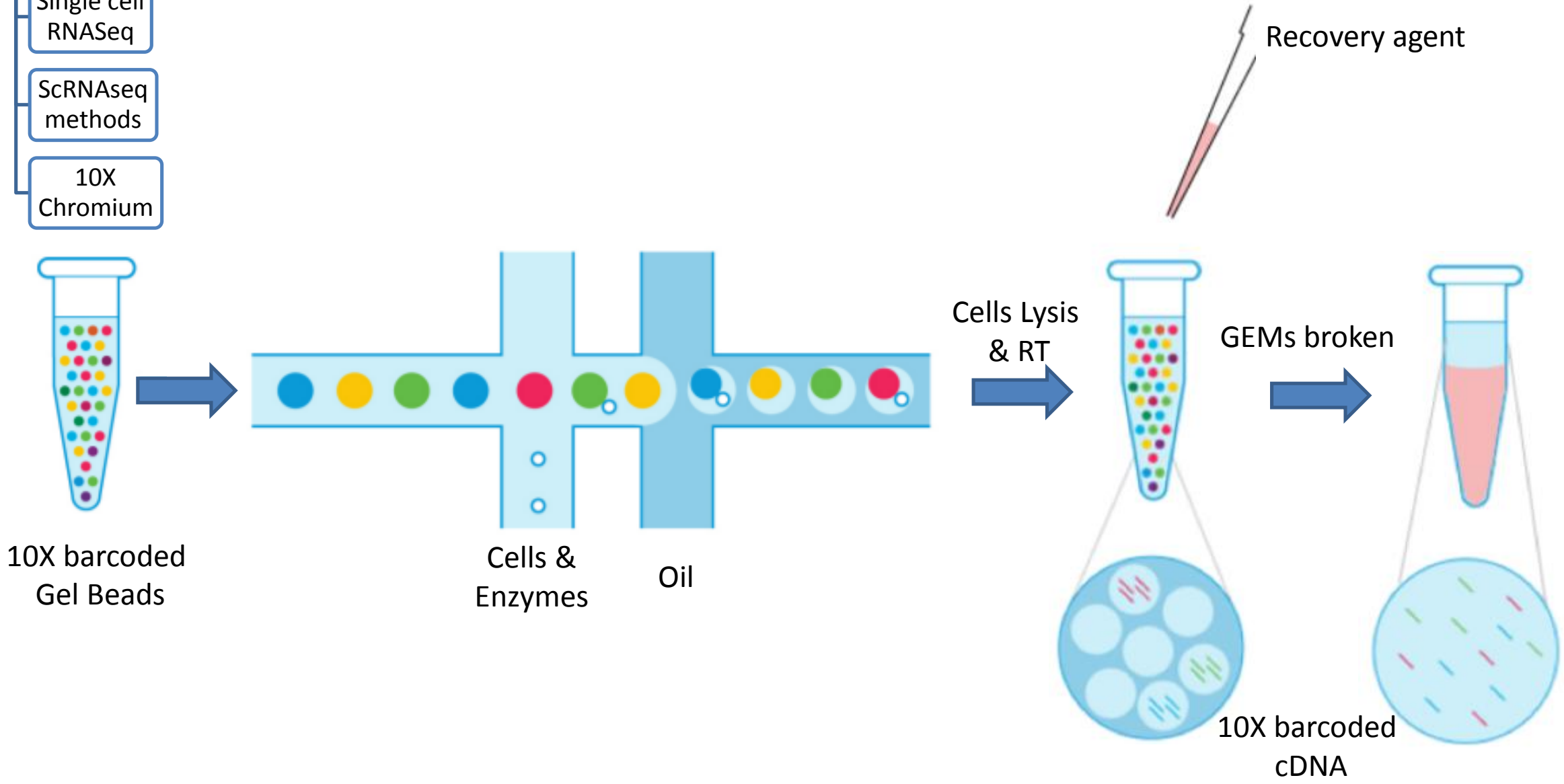


Gel Bead-in-Emulsion (GEM)

https://www.10xgenomics.com/solutions/single-cell/

10X genomics protocol

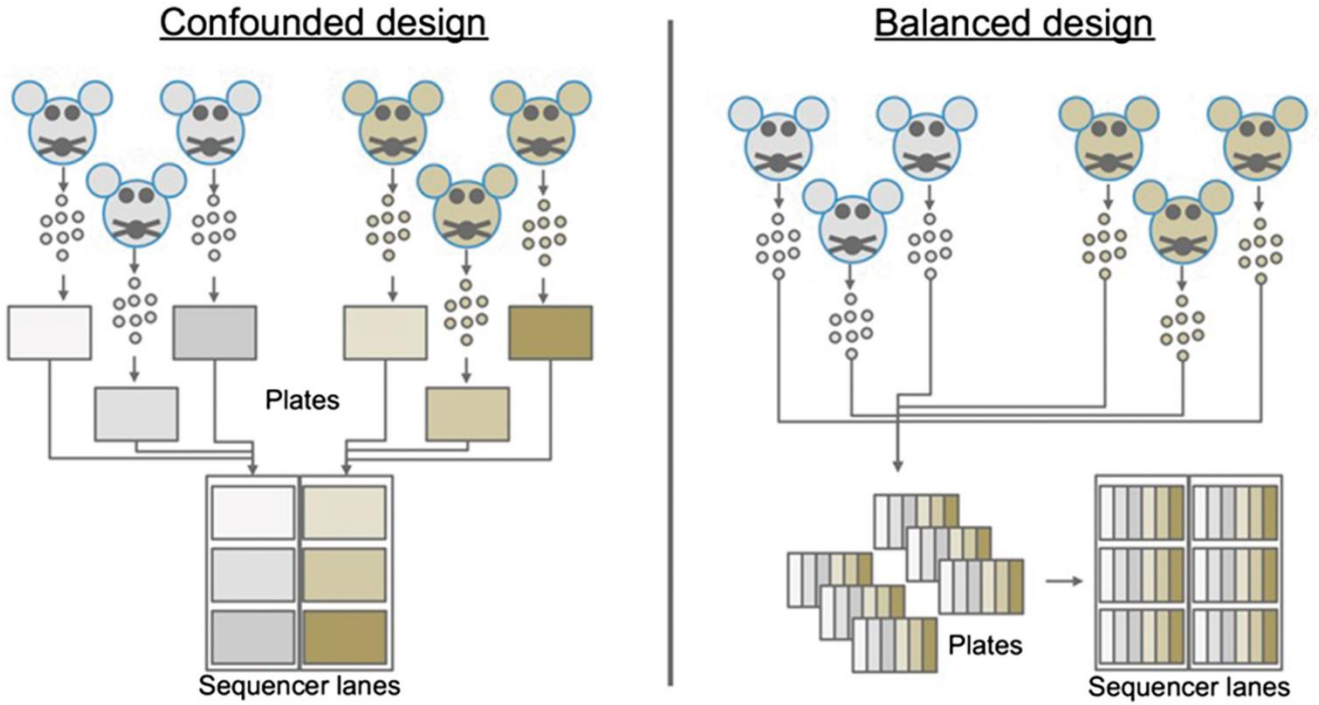
- RNASeq
  - Single cell RNASeq
  - ScRNAseq methods
  - 10X Chromium



https://www.10xgenomics.com/solutions/single-cell/

- RNASeq
  - Single cell RNASeq
  - ScRNAseq methods
  - 10X Chromium
  - DESIGN

10X genomics protocol



(1)

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.4%	~870	~500
~0.8%	~1700	~1000
~2.3%	~5300	~3000
~3.9%	~8700	~5000
~7.6%	~17400	~10000

65%

Determine starting cells:
http://satijalab.org/howmanycells

QC metrics for your cell type
http://10xqc.com/

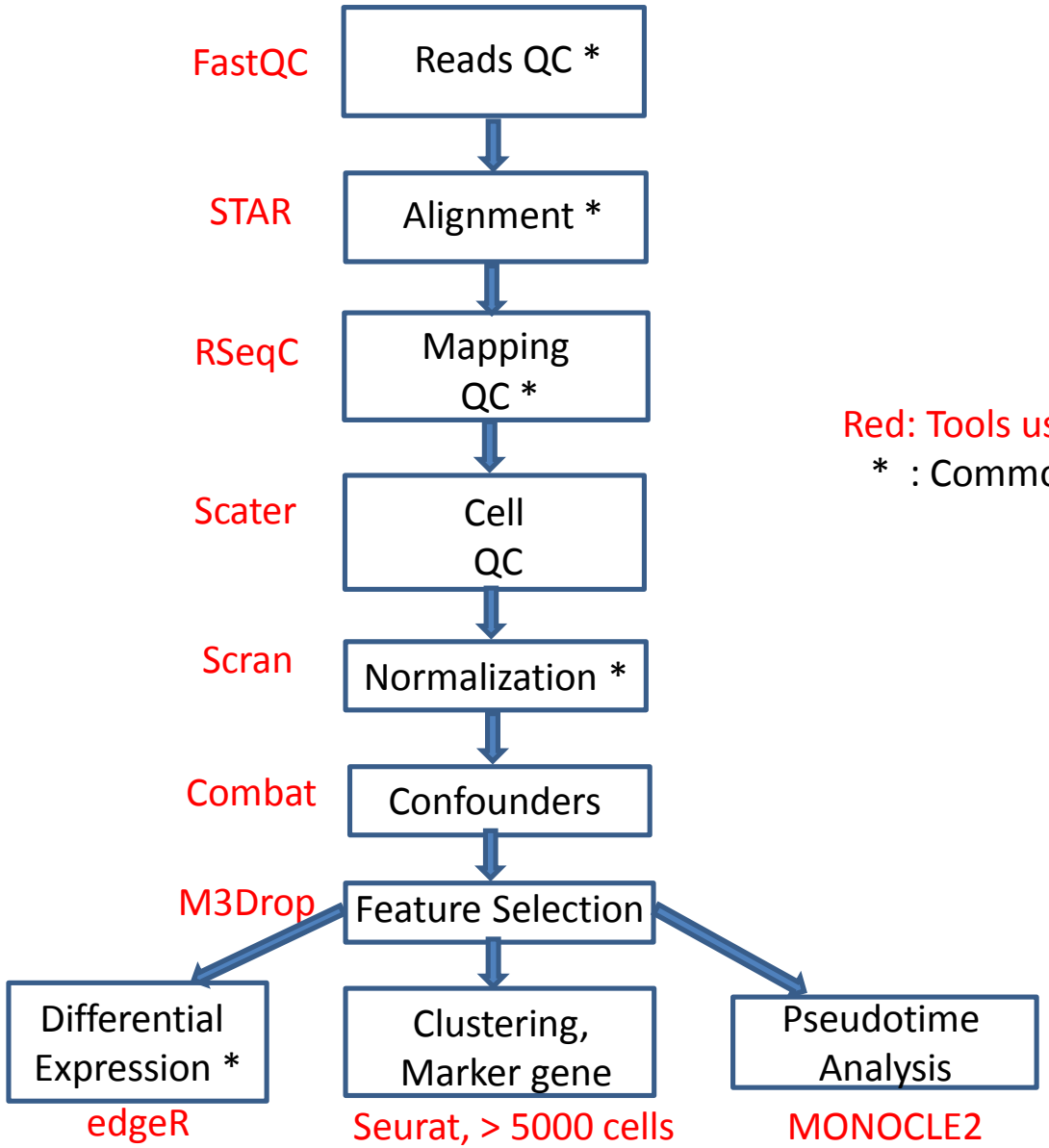
(2)

(1) Jeanette Baran-Gale et. al., Brief Funct Genomics, 2017. doi:10.1093/bfgp/elx035
(2) https://www.10xgenomics.com/solutions/single-cell

RNASeq

- Single cell RNASeq
- ScRNAseq methods
- 10X Chromium
- DESIGN
- Analysis pipeline

Single cell RNA Sequencing computational pipeline



Red: Tools used for each step

* : Common to Bulk RNASeq

RNASeq

Single cell RNASeq

ScRNASeq

Let us carry out a single cell RNASeq experiment

- Peripheral blood mononuclear cells (**PBMCs**) from a healthy donor. PBMCs are primary cells with relatively small amounts of RNA (~1pg RNA/cell).
- Sequenced on Illumina Hiseq4000
- 26bp read1 (16bp Chromium barcode and 10bp UMI), 98bp read2 (transcript), and 8bp I7 sample barcode

Dataset: https://support.10xgenomics.com/single-cell-gene-expression/datasets/2.0.1/pbmc8k

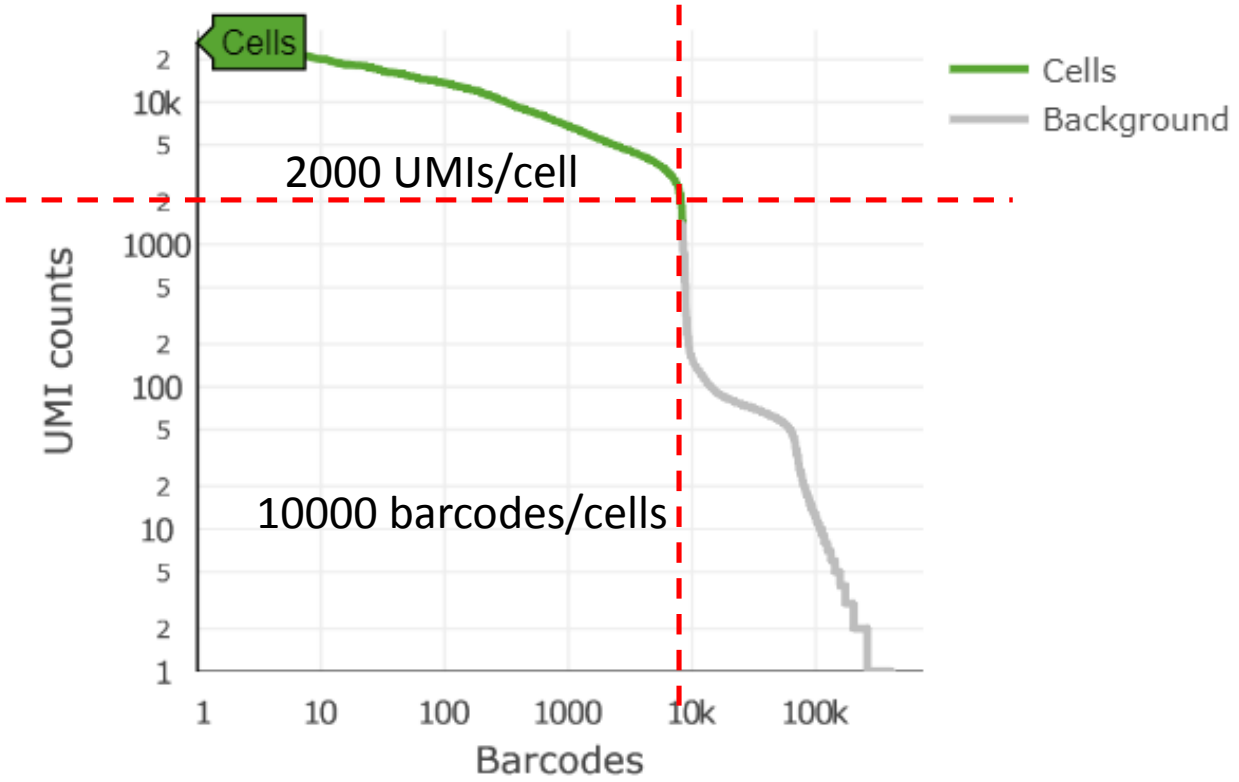
RNASeq

Single cell RNASeq

ScRNASeq

SUMMARY

10X genomics summary



Parameter	PBMC Observed	PBMC Recommended
Estimated Number of Cells/Barcodes	8381	500-10,000
Fraction Reads in Cell	93.10%	
Mean Reads per Cell	93,552	50,000
Median Genes per Cell	1297	1000
Total Genes Detected	21,425	
Median UMI counts per cell	4084	

RNASeq

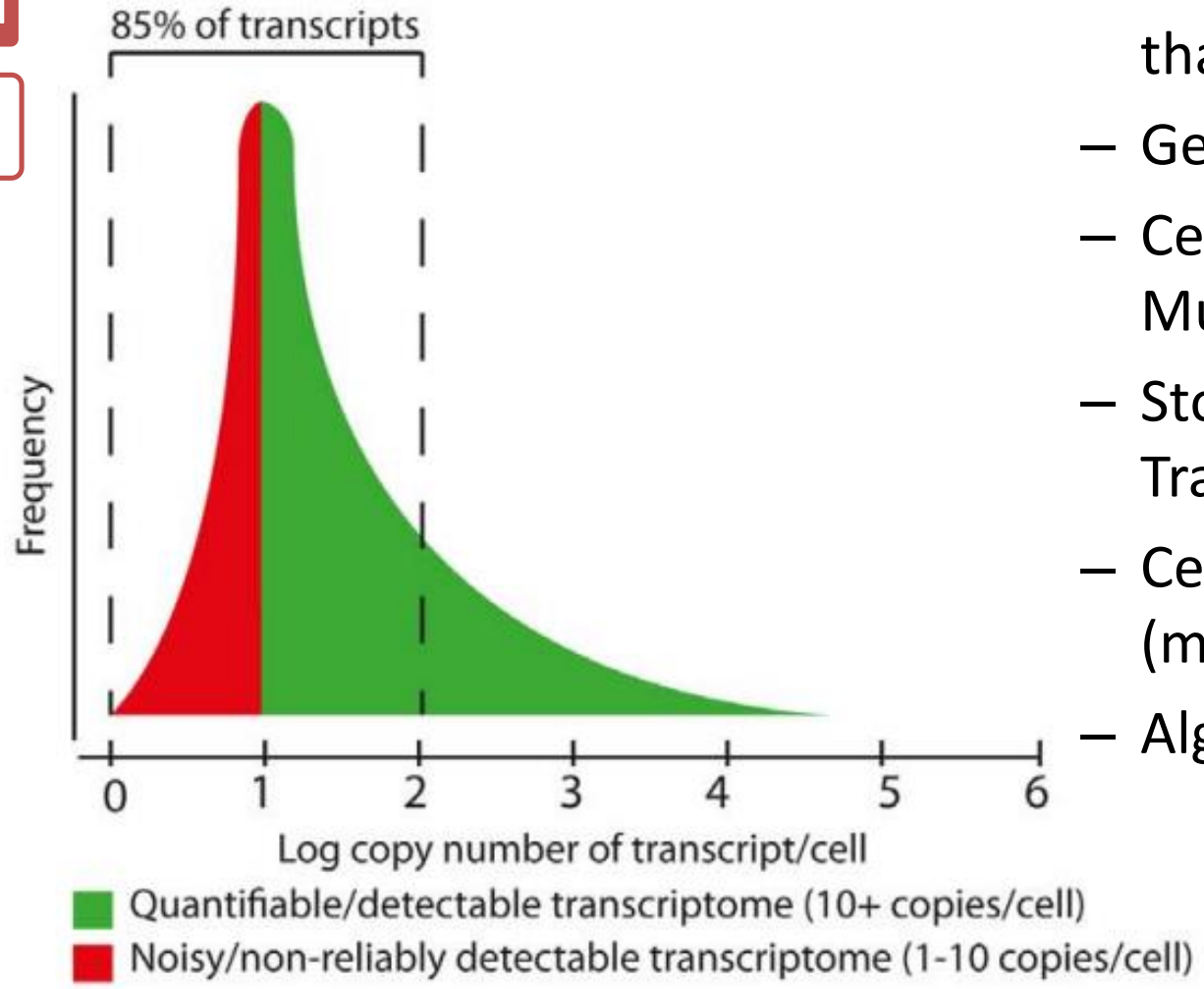
Single cell RNASeq

ScRNASeq

SUMMARY

PITFALLS

Potential problems



- Low amplification efficiency, typically less than 10%
- Gene dropout rates
- Cell quality: Live/dead, Missing cells, Multiple cells
- Stochastic: cell cycle phases, Transcriptional bursting
- Cell Capture rate ≠ population frequency (multiples/gel beads, empty beads)
- Algorithm development

Macaulay and Voet, PLOS Genetics, 2014. Computational Methods for Analysis of Single Cell RNA-Seq Data, Ion Măndoiu, University of Connecticut.

RNASeq

Single cell RNASeq

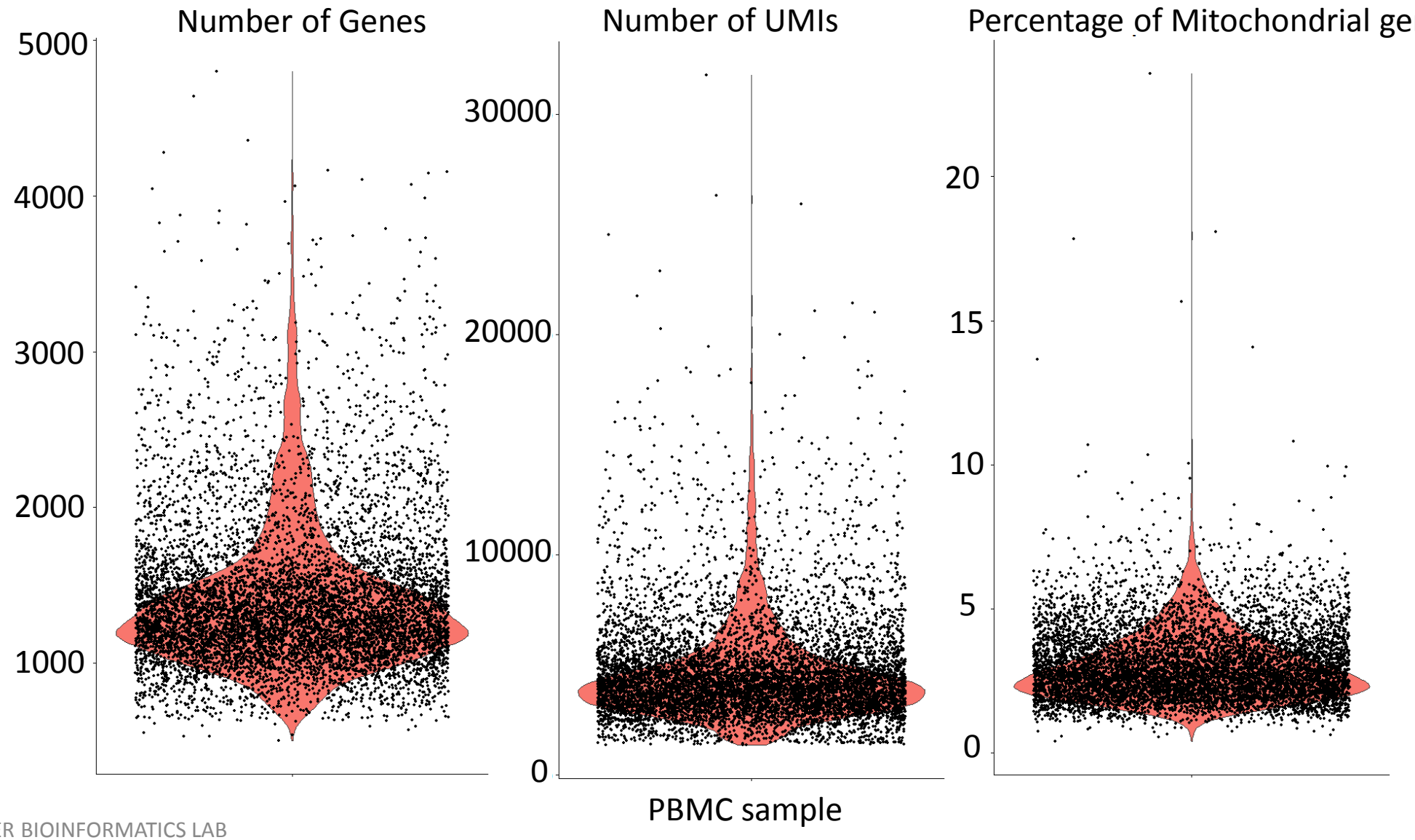
ScRNASeq

SUMMARY

PITFALLS

QC

Exploring the Quality Control required



Clustering

- Determine a subset of genes to use for clustering; this is because not all genes are informative, such as those that are lowly expressed.
- The approach is to select gene based on their average expression and variability across cells
- We scale the data and remove unwanted sources of variation (technical, cell cycle stage, batches etc.)

RNASeq

Single cell RNASeq

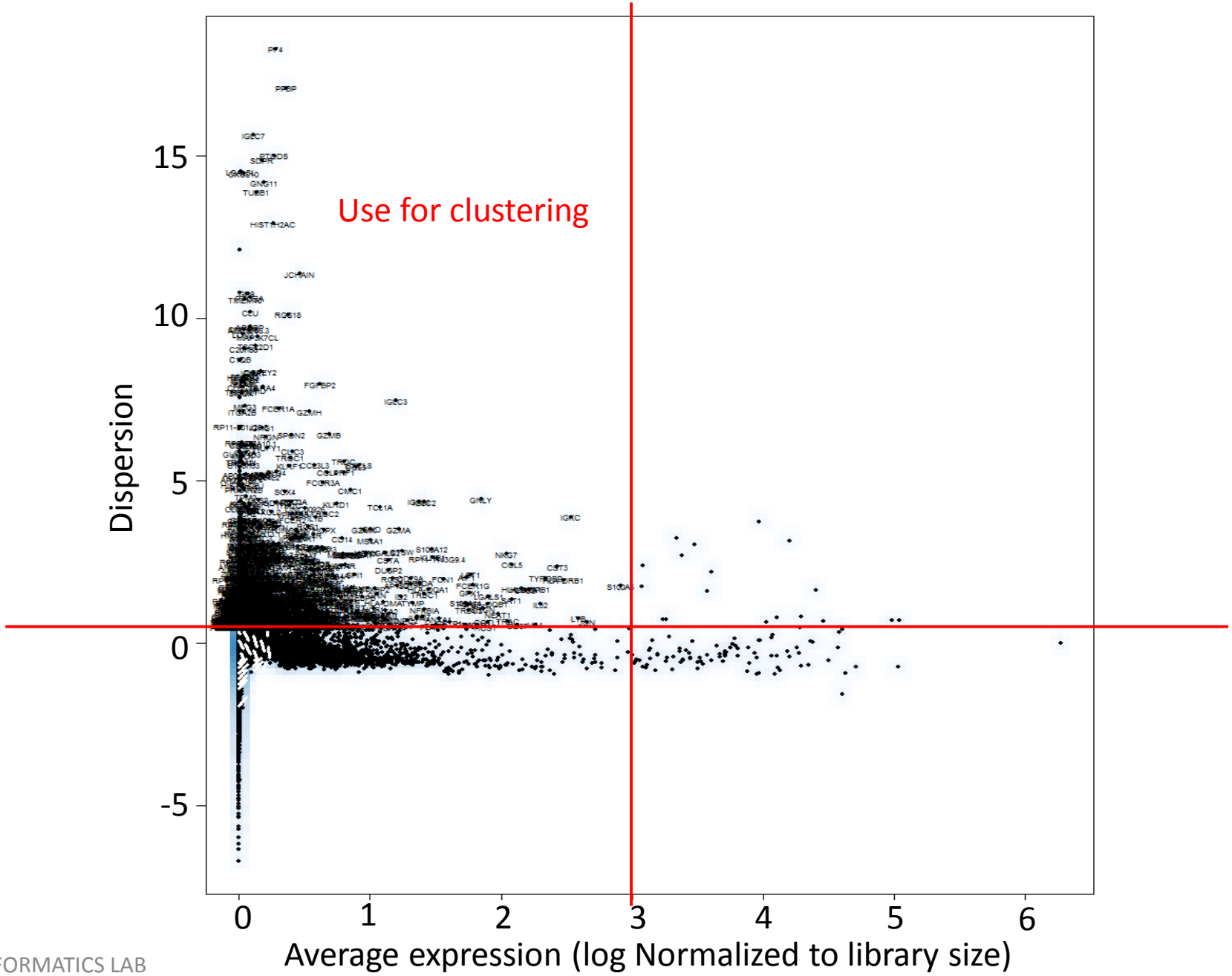
ScRNASeq

SUMMARY

PITFALLS

QC

Mean vs Variation: Detection of variable genes



RNASeq

Single cell RNASeq

ScRNASeq

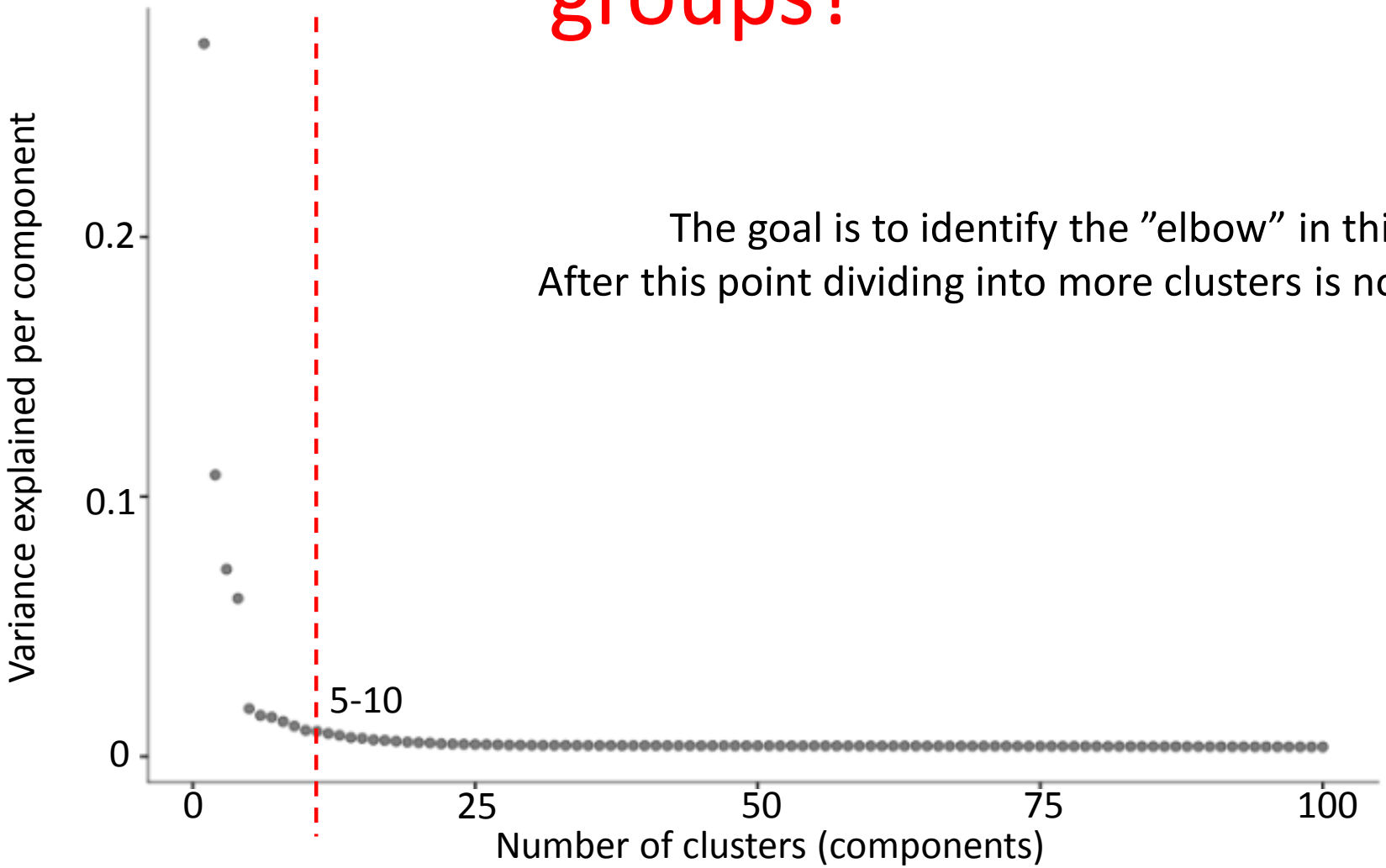
SUMMARY

PITFALLS

QC

CLUSTERING

How many clusters are enough to divide the data into meaningful groups?



The goal is to identify the ”elbow” in this plot,
After this point dividing into more clusters is not informative

RNASeq

Single cell RNASeq

ScRNASeq

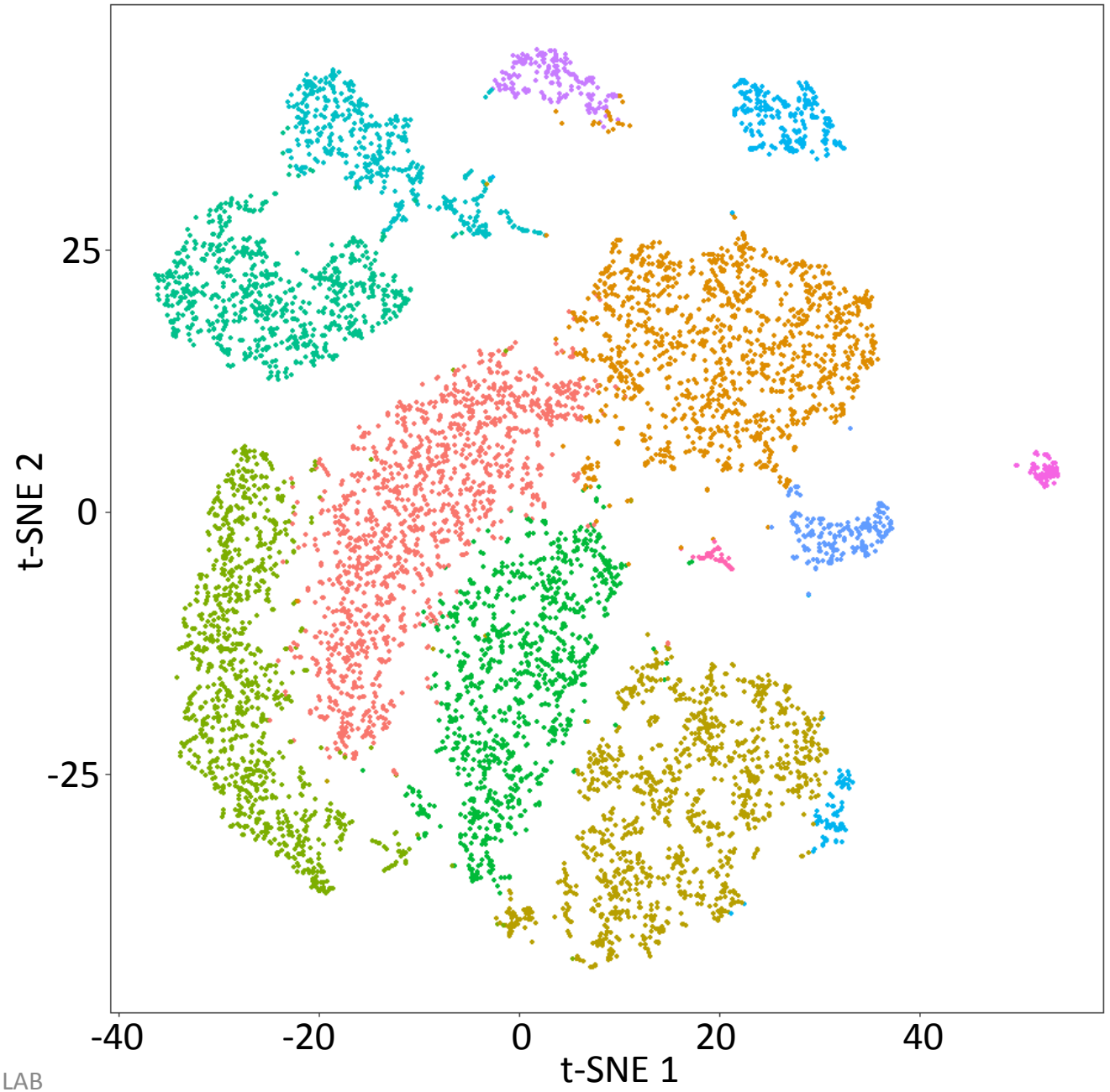
SUMMARY

PITFALLS

QC

CLUSTERING

Graph Cluster results



T-SNE plot visualization
t-Distributed
Stochastic Neighbor Embedding

0
1
2
3
4
5
6
7
8
9
10
11

RNASeq

Single cell RNASeq

ScRNASeq

SUMMARY

PITFALLS

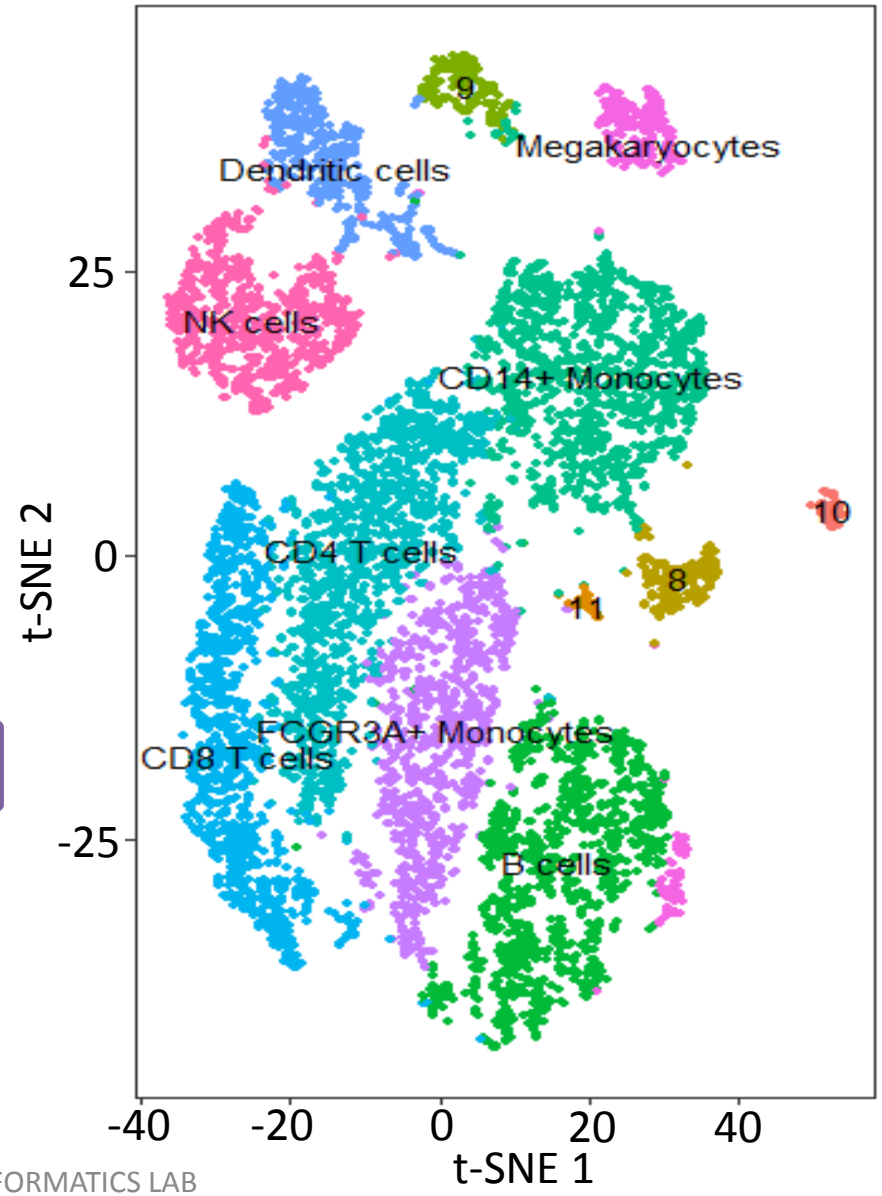
QC

CLUSTERING

HETEROGENIETY

KNOWN MARKER

Differential expression
Identification of markers for cell types



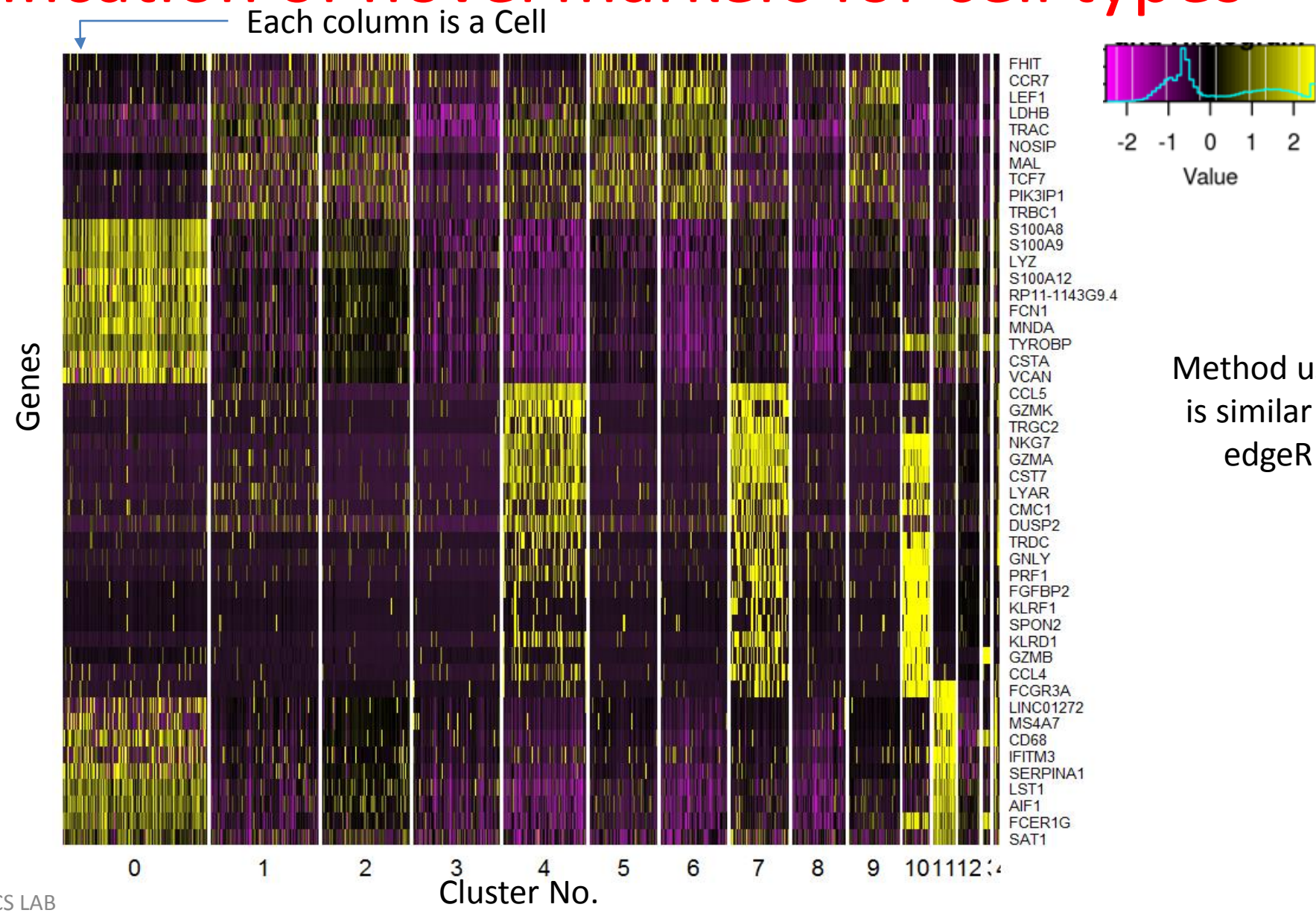
Supervised – PBMC known markers

Markers	Cell Type
IL7R	CD4 T cells
CD14, LYZ	CD14+ Monocytes
MS4A1	B cells
CD8A	CD8 T cells
FCGR3A, MS4A7	FCGR3A+ Monocytes
GNLY, NKG7	NK cells
FCER1A, CST3	Dendritic Cells
PPBP	Megakaryocytes

- RNASeq
  - Single cell RNASeq
- ScRNASeq
  - SUMMARY
  - PITFALLS
  - QC
  - CLUSTERING
- HETEROGENIETY
  - KNOWN MARKER
  - NOVEL MARKER

Differential expression

Identification of novel markers for cell types



Method used is similar to edgeR

RNASeq

Single cell RNASeq

ScRNASeq

SUMMARY

PITFALLS

QC

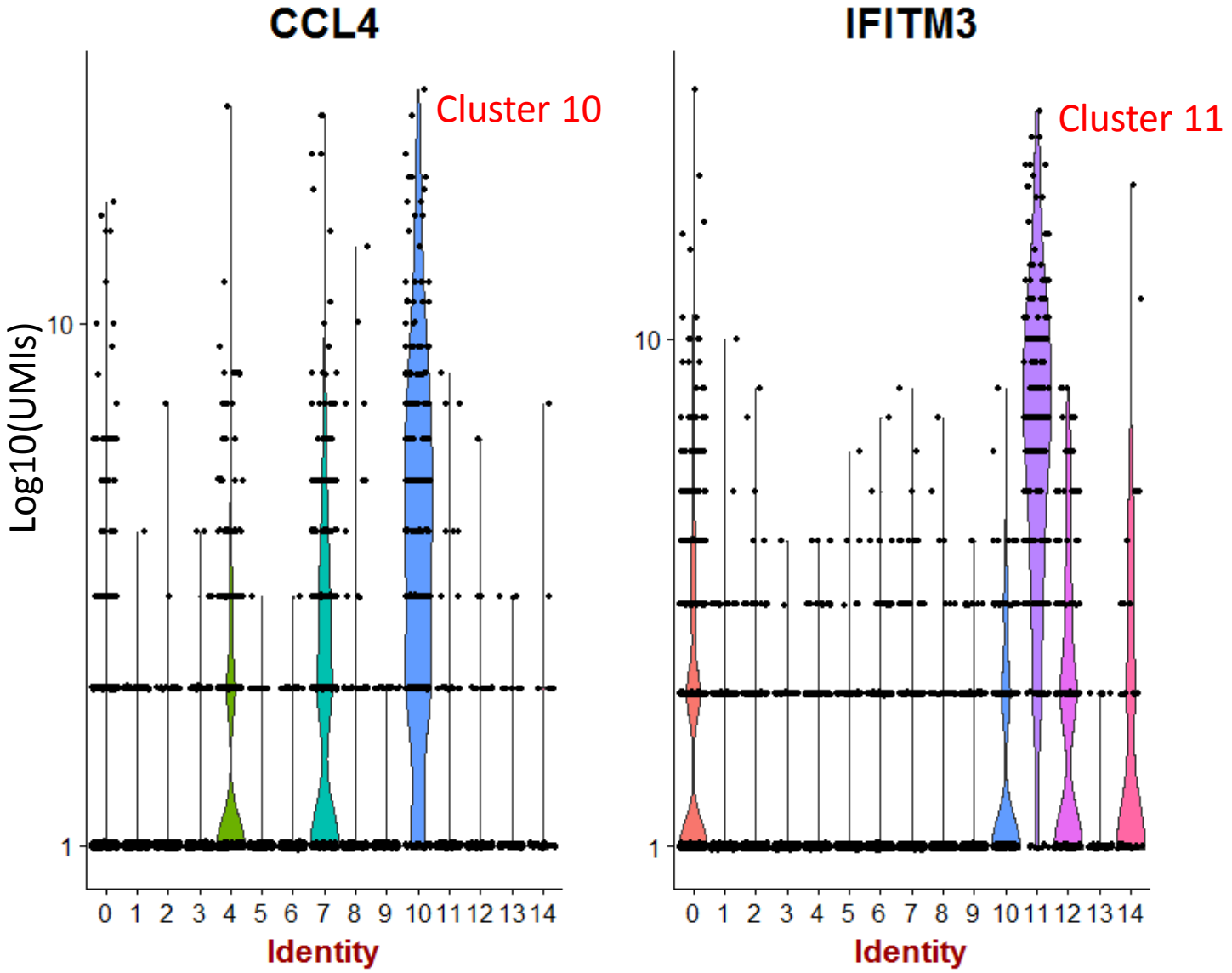
CLUSTERING

HETEROGENIETY

KNOWN MARKER

NOVEL MARKER

Differential expression
Identification of novel markers for cell types



RNASeq

Single cell RNASeq

ScRNASeq

SUMMARY

PITFALLS

QC

CLUSTERING

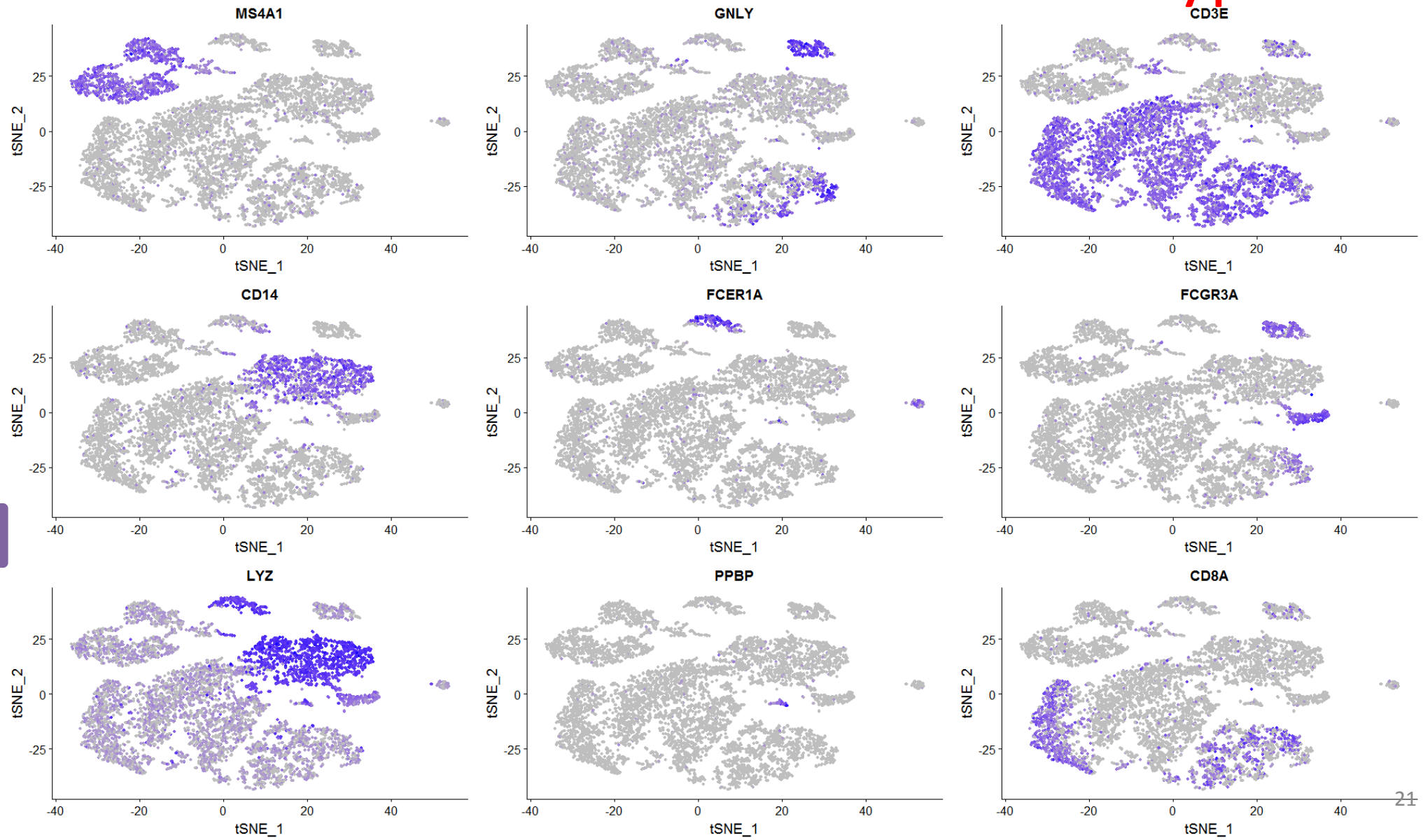
HETEROGENIETY

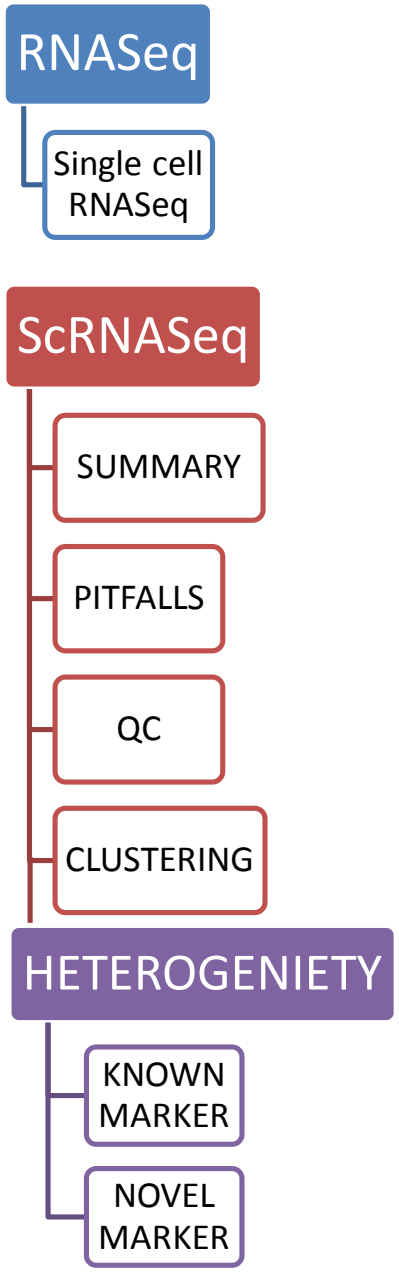
KNOWN MARKER

NOVEL MARKER

Differential expression

Identification of novel markers for cell types





Cloupe browser further exploration

Acknowledgements

❖**McDermott Center Bioinformatics Lab**

Chao Xing, PhD - Director

Mohammad Kanchwala

Ashwani Kumar

❖**McDermott Center NGS core team**

Workflow used is based on:

❖Seurat: Macosko, Basu, Satija et al., Cell, 2015 (Updated approach: Combining dimensional reduction with graph-based clustering)

❖Monocle: Xiaojie Qiu, Andrew Hill, Cole Trapnell et al (2017)

❖Computational Methods for Analysis of Single Cell RNA-Seq Data, Ion Măndoiu, University of Connecticut

THANK YOU