Multi-Omics in SC

sc DNA methylation using scRRBS

sc chromatin accessibility using scATACseq

sc Proteomics using CITEseq

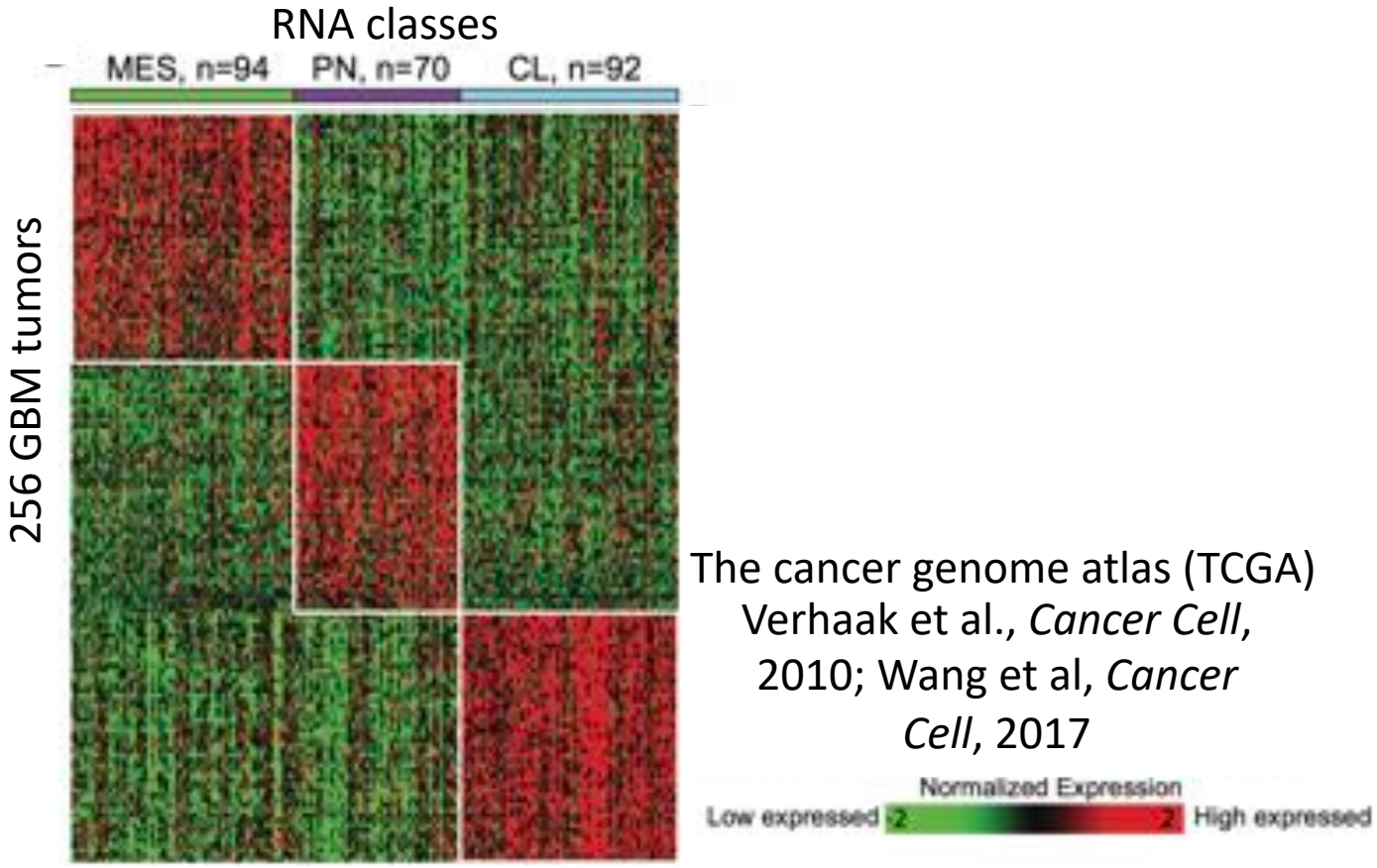
scDNAmethylation

Why do we need scDNA methylation?

Glioma tumors exhibit inter-tumor heterogeneity

Glioma brain tumors:

- Main types:
  - Glioblastoma (GBM) ~90%
  - IDH-mutant glioma 8-9%
- Life expectancy hardly changed over the last 40 years of research
- Highly heterogenous



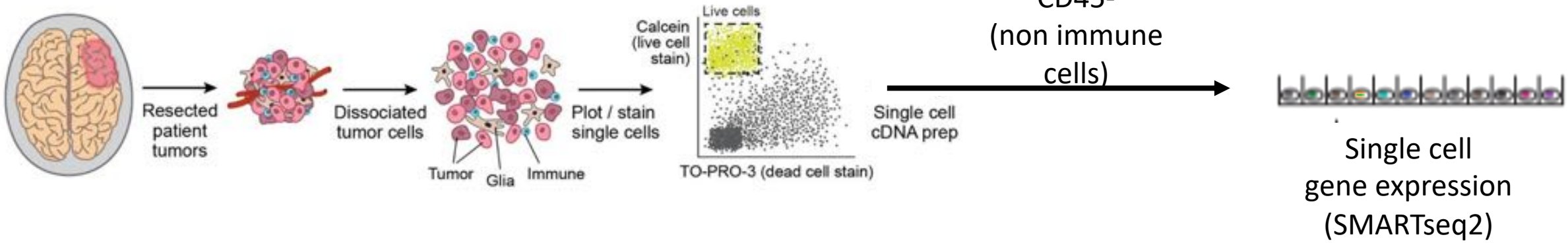
The cancer genome atlas (TCGA) Verhaak et al., *Cancer Cell*, 2010; Wang et al, *Cancer Cell*, 2017

Is the heterogeneity driven by:

the tumors are composed of different cell types?

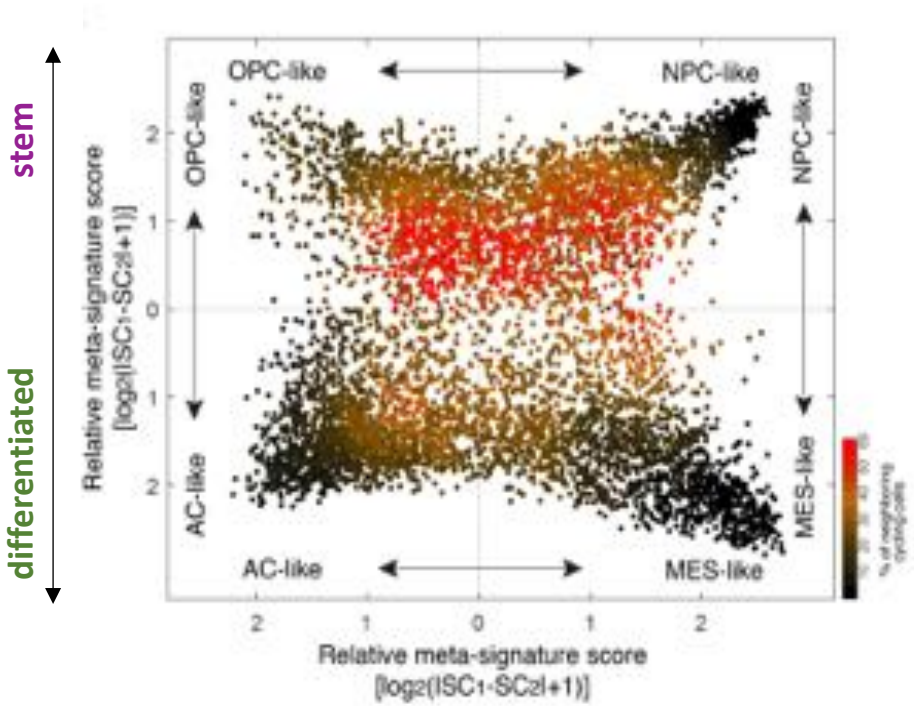
the tumors have the same cell types mixed in different proportions?

Revealing glioma compositions via single cell RNA sequencing

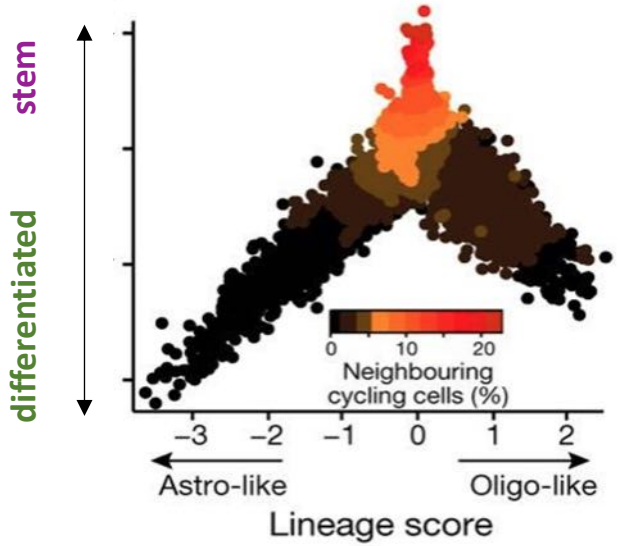


single cell RNAseq reveals intra-tumor heterogeneity

Glioblastoma



IDH-mutant glioma



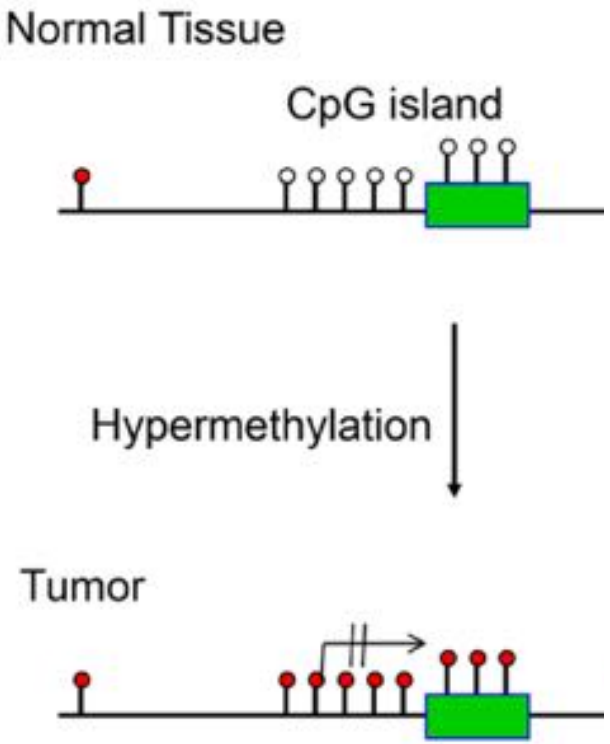
Patel et al. 2014;
Tirosh et al. 2016;
Venteicher et al. 2017;
Neftel et al. 2019

Is the heterogeneity driven by:

~~the tumors are composed of different cell types~~

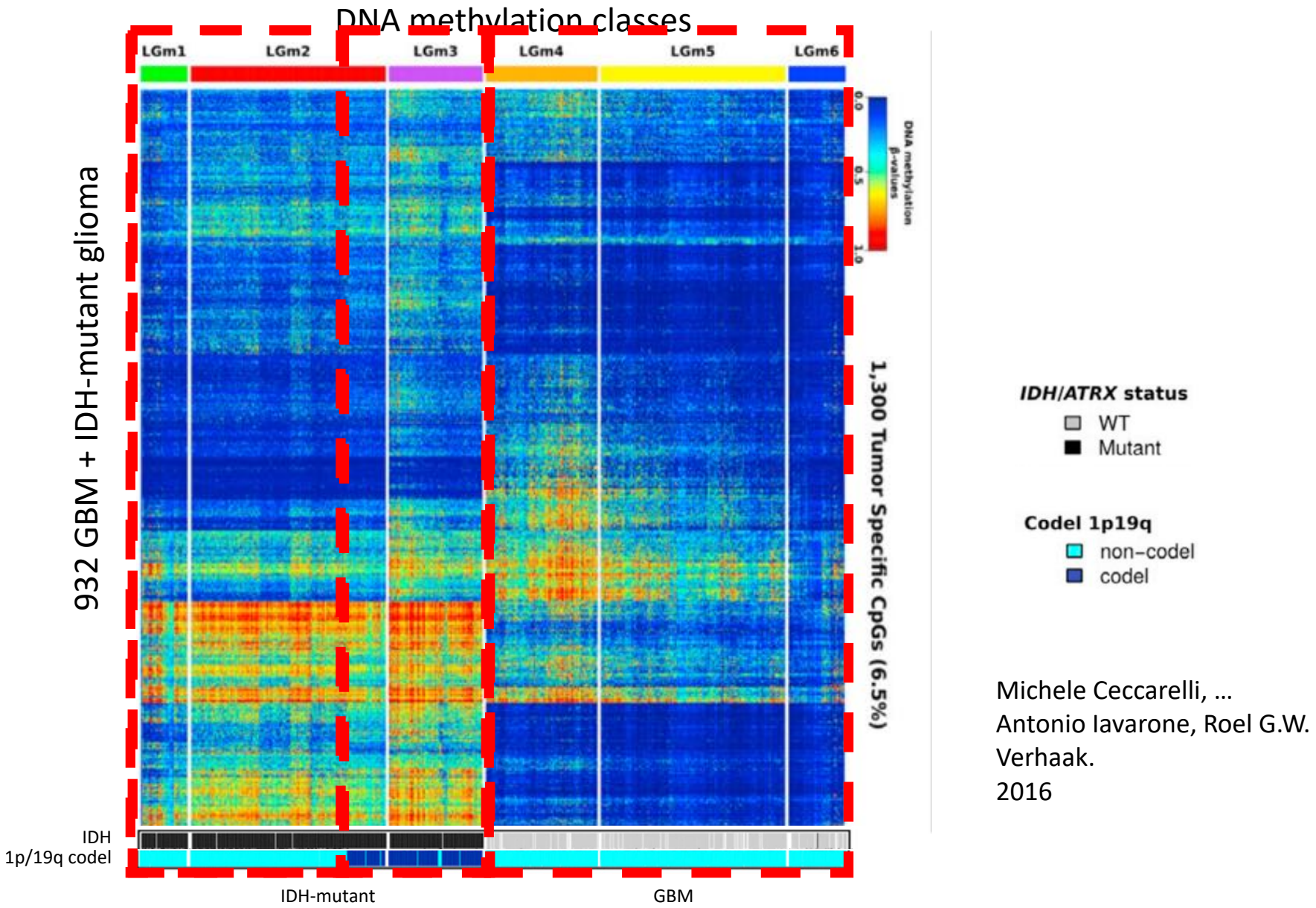
the tumors have the same cell types mixed in different proportions!

DNA methylation in healthy tissue and disease



From Pfeifer et. al., 2018

Inter-tumor DNA methylation heterogeneity



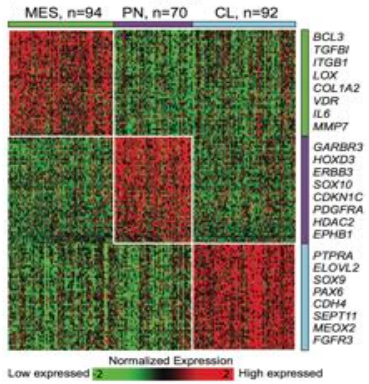
Michele Ceccarelli, …
Antonio Iavarone, Roel G.W. Verhaak.
2016

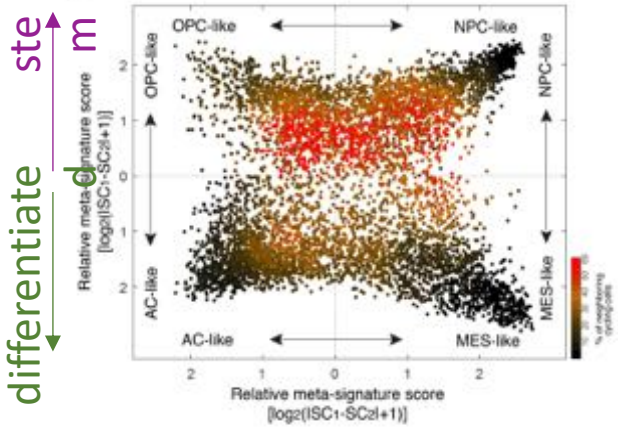
The goal: dissecting epigenetic identity of cellular states underlying glioma evolution

inter-tumor heterogeneity

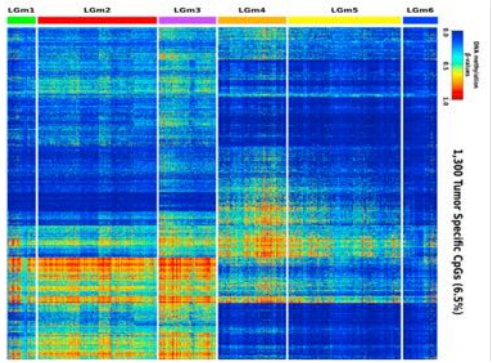
intra-tumor heterogeneity

RNA





DNA methylation

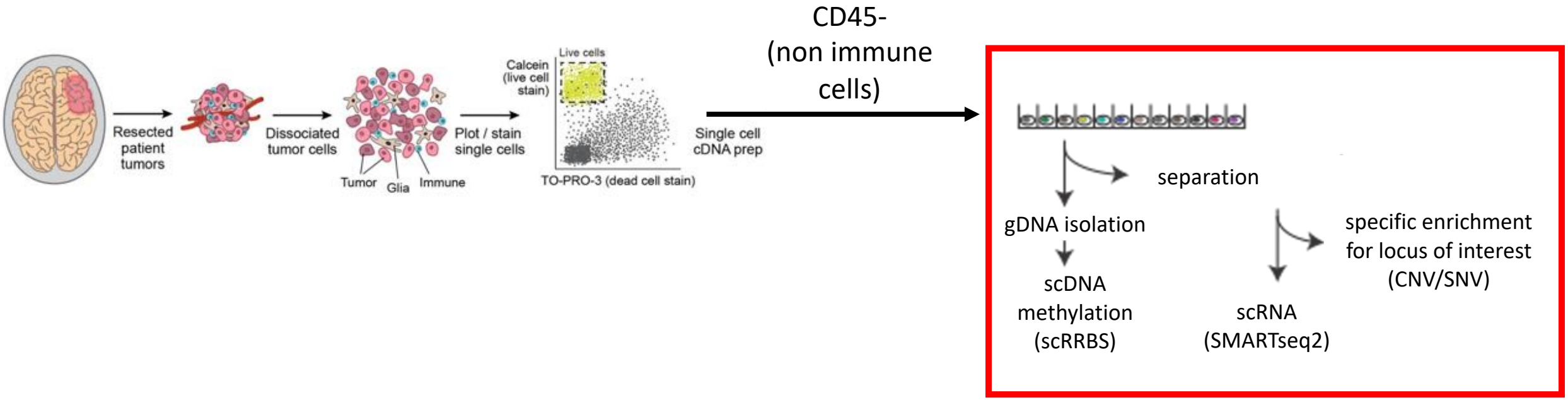


Is there a link between the RNA function and DNA methylation?

What is driving the DNA methylation heterogeneity?

Using scDNA methylation to dissect epigenetic events driving cellular states

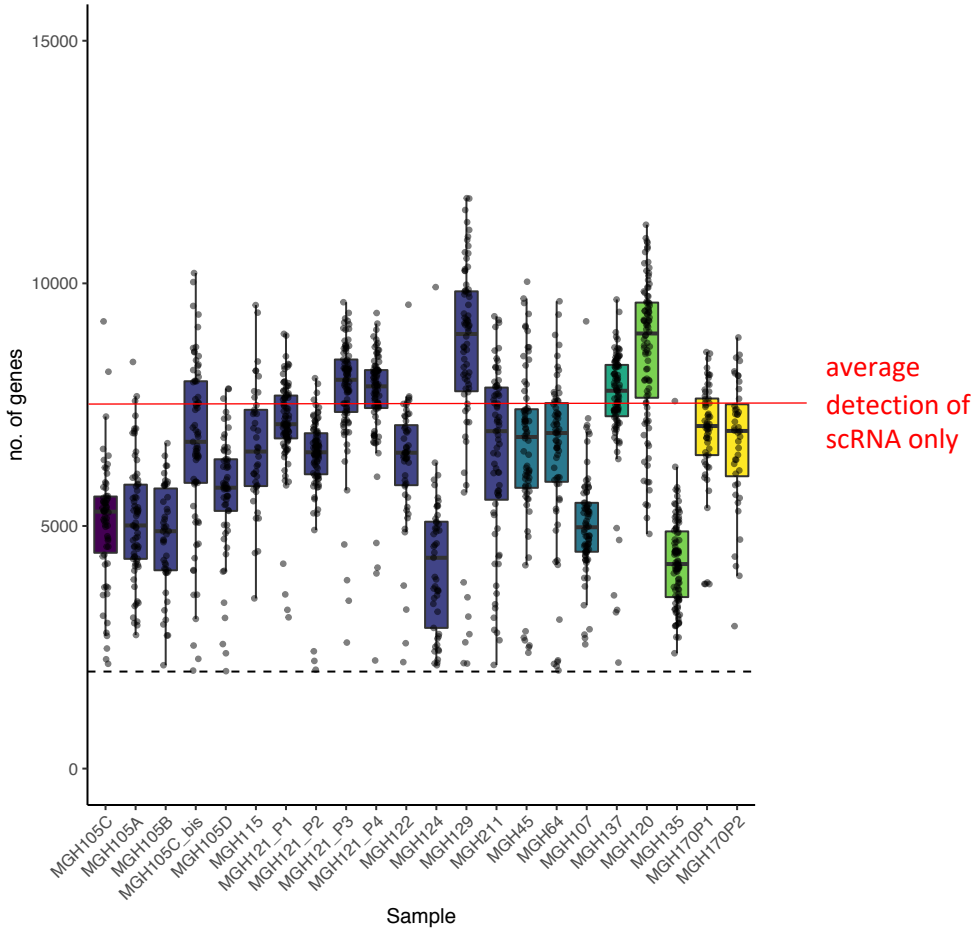
Simultaneously interrogating DNA methylation and RNA in the **same single cells**



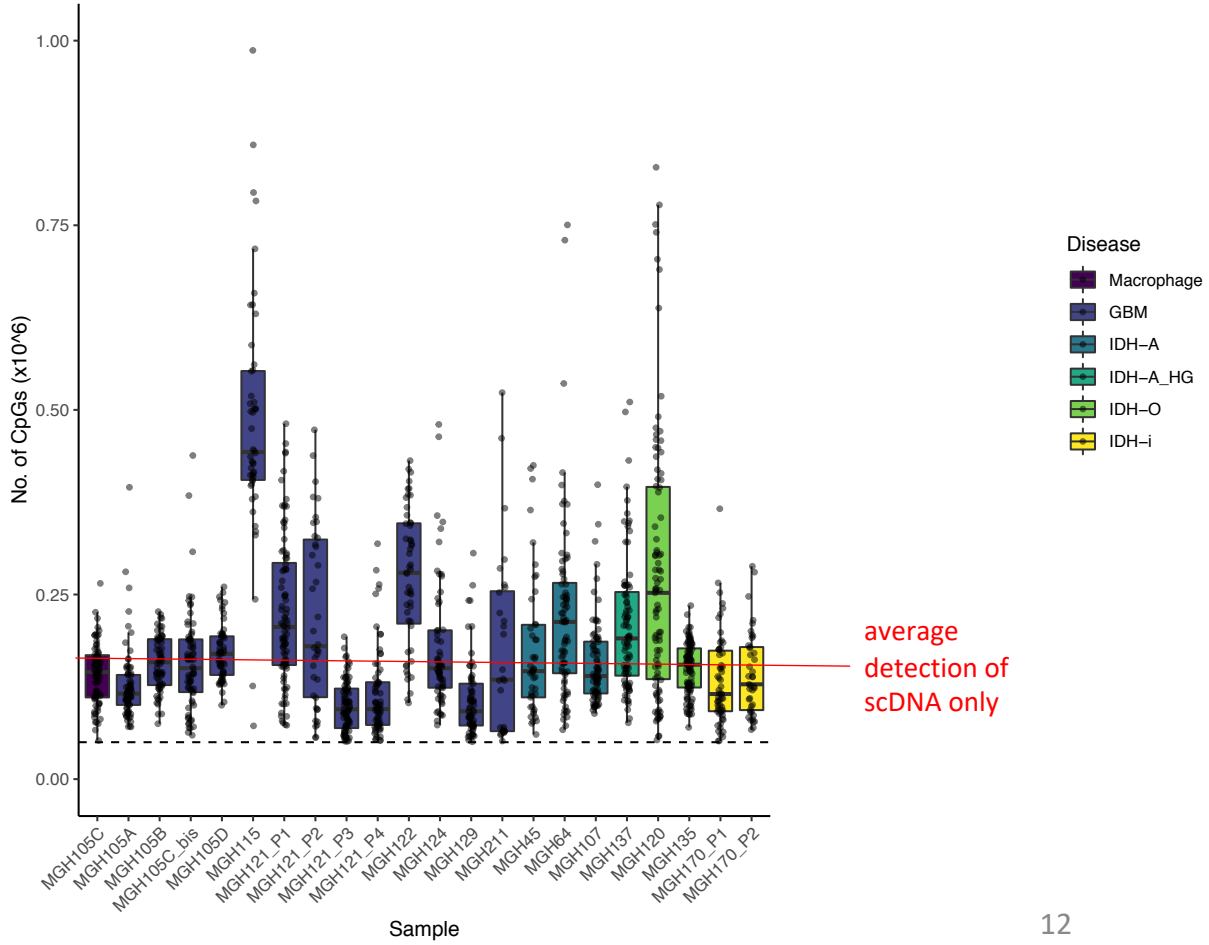
11

Number of features using scDNAseq compared to number of features using scRANseq

scRNA

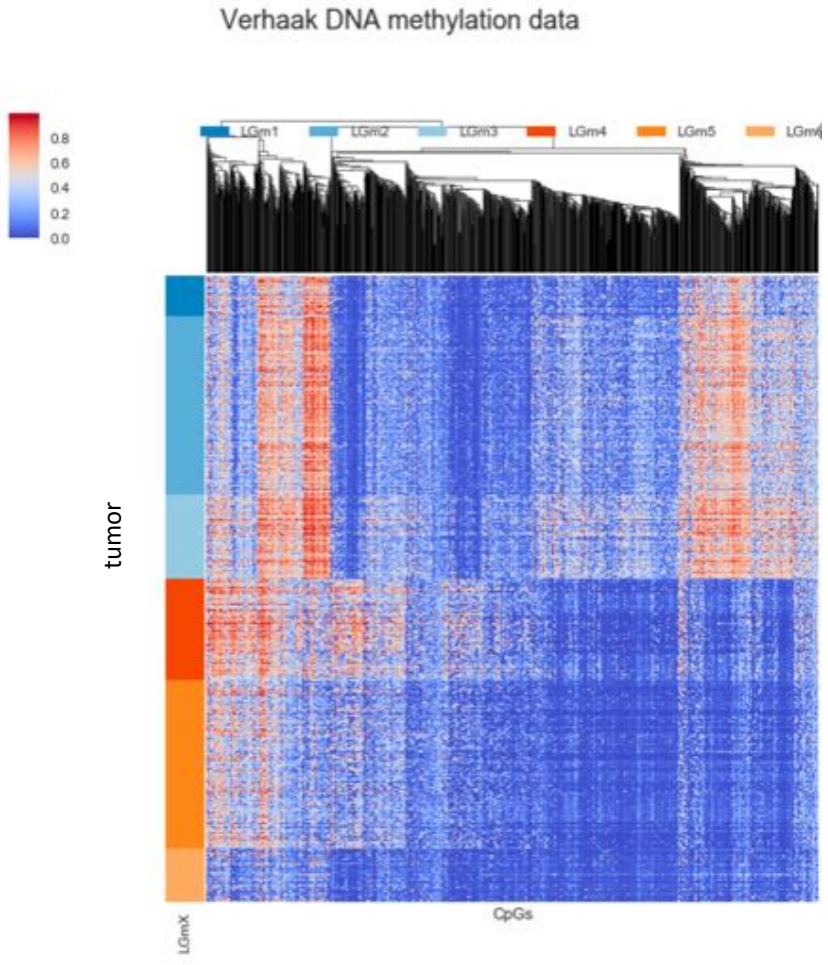


scDNA methylation

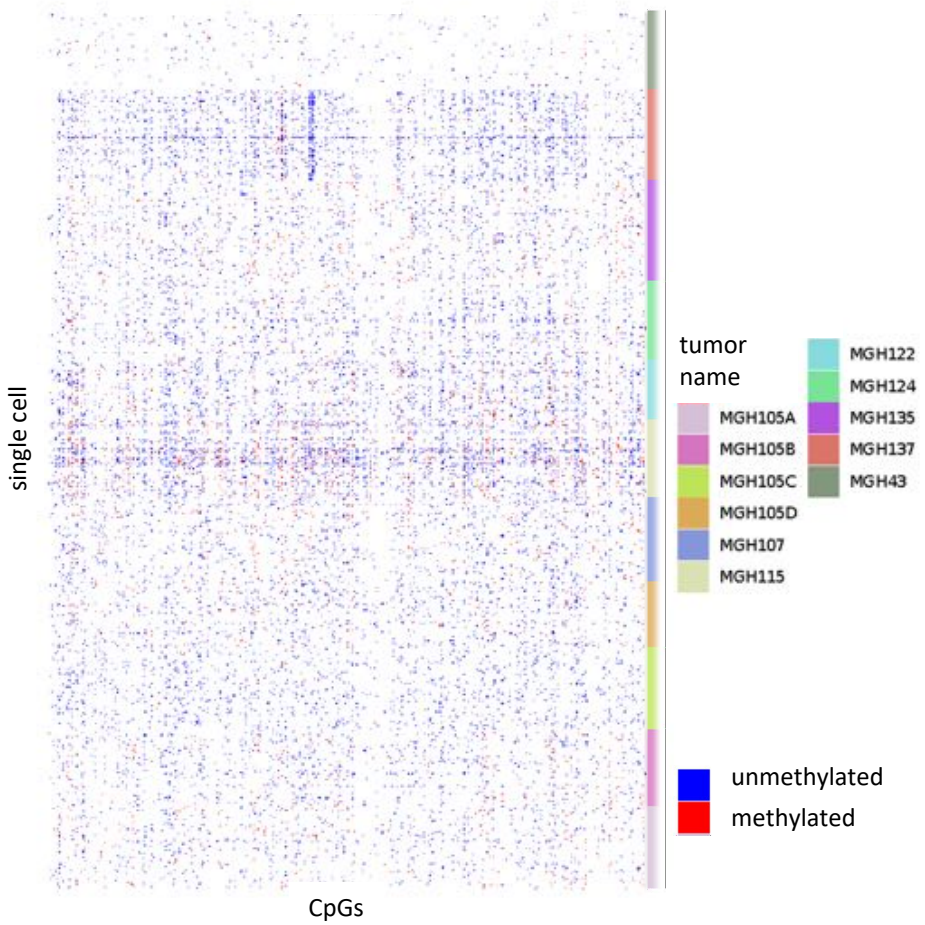


scDNA methylation is **really** sparse

Bulk DNA methylation

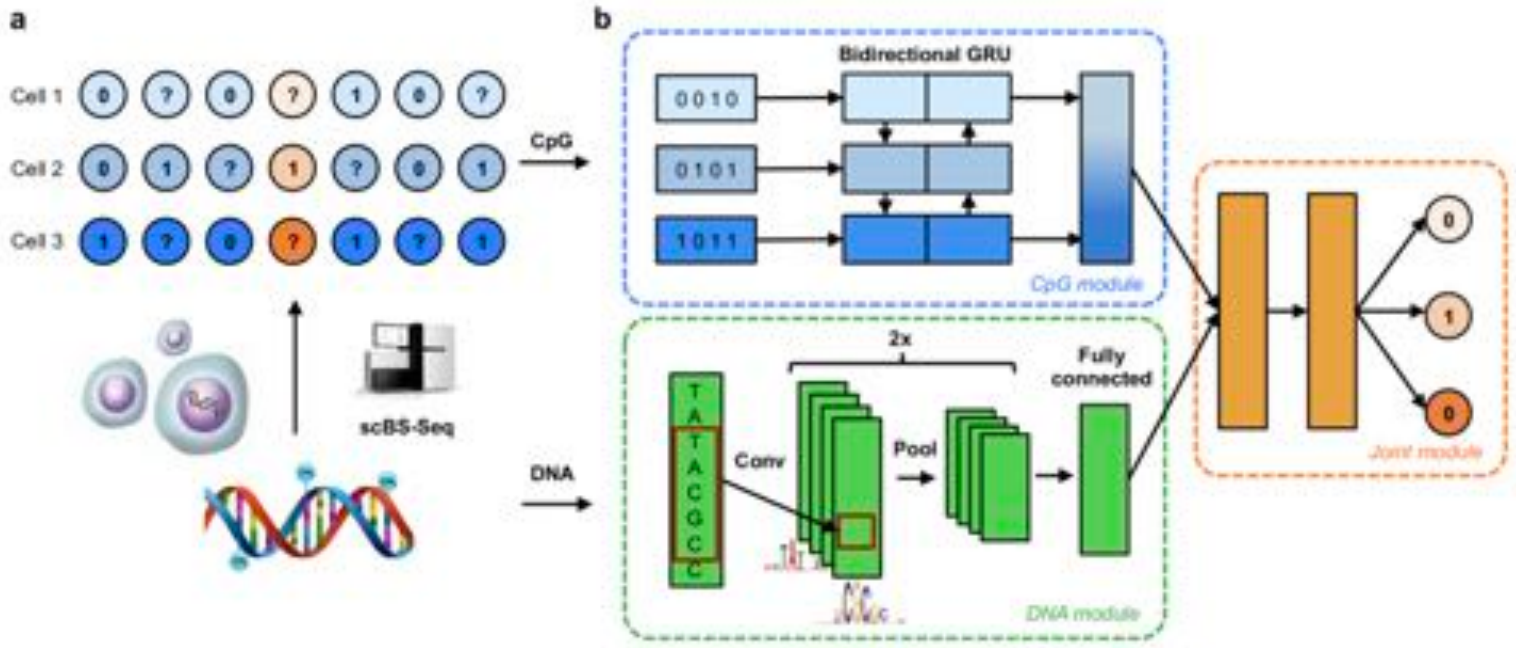


scDNA methylation

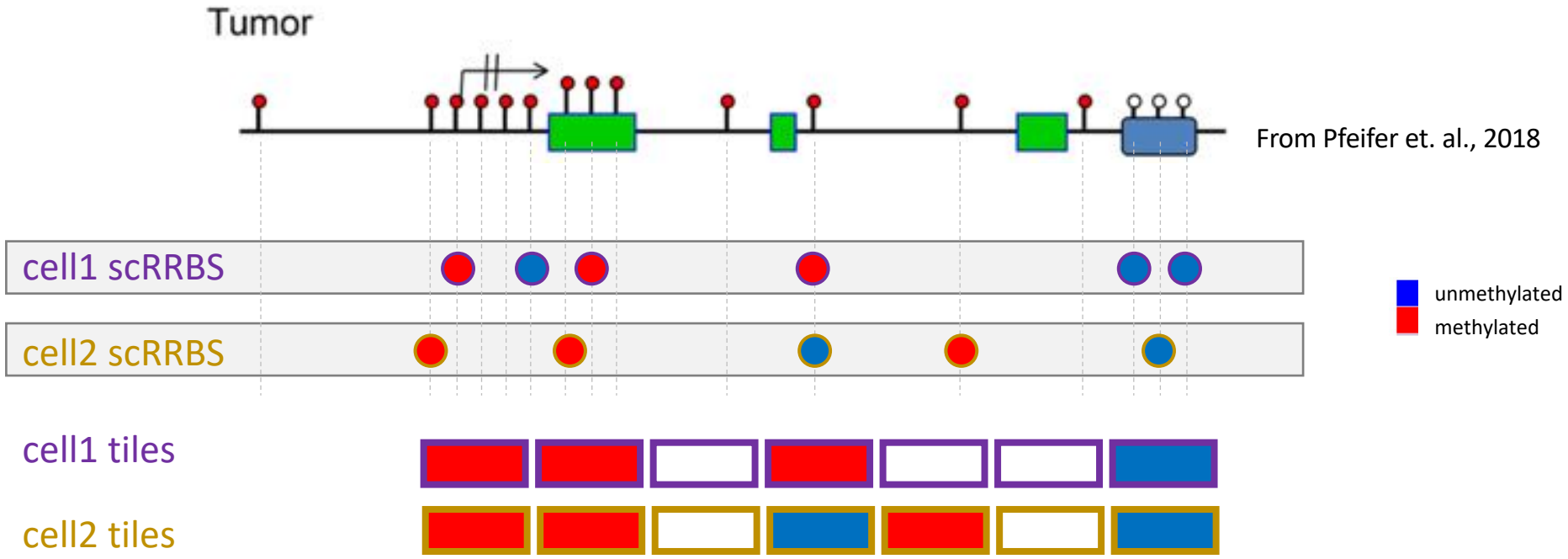


Impute sc DNA data: option1. machine learning techniques

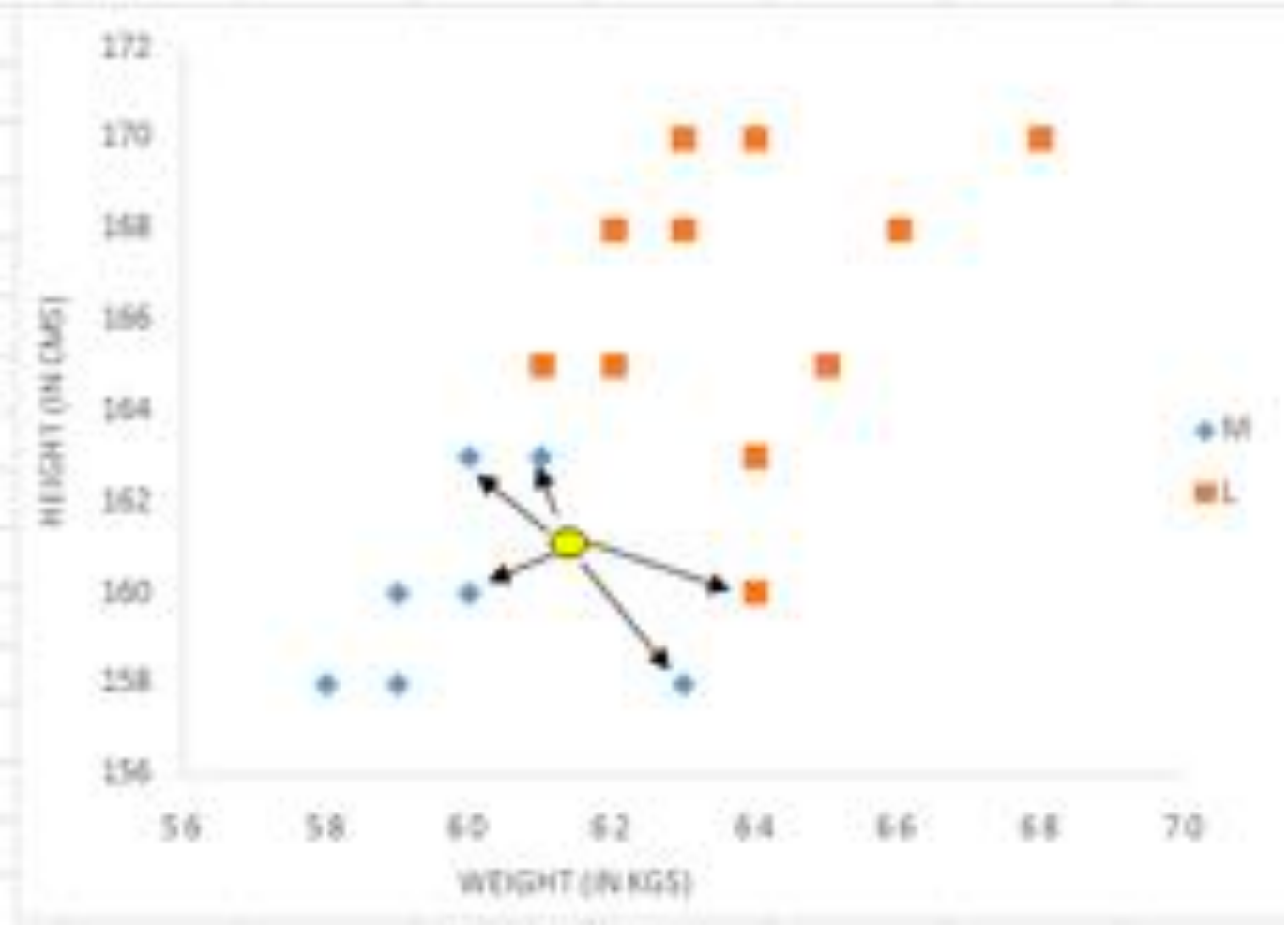
DeepCpG: Deep neural networks for predicting single-cell DNA methylation



Impute sc DNA data: option2. create tiles with average methylation in the tile



Impute sc DNA data: option3. KNN



scDNA methylation clustering separates non-malignant cells from malignant cells

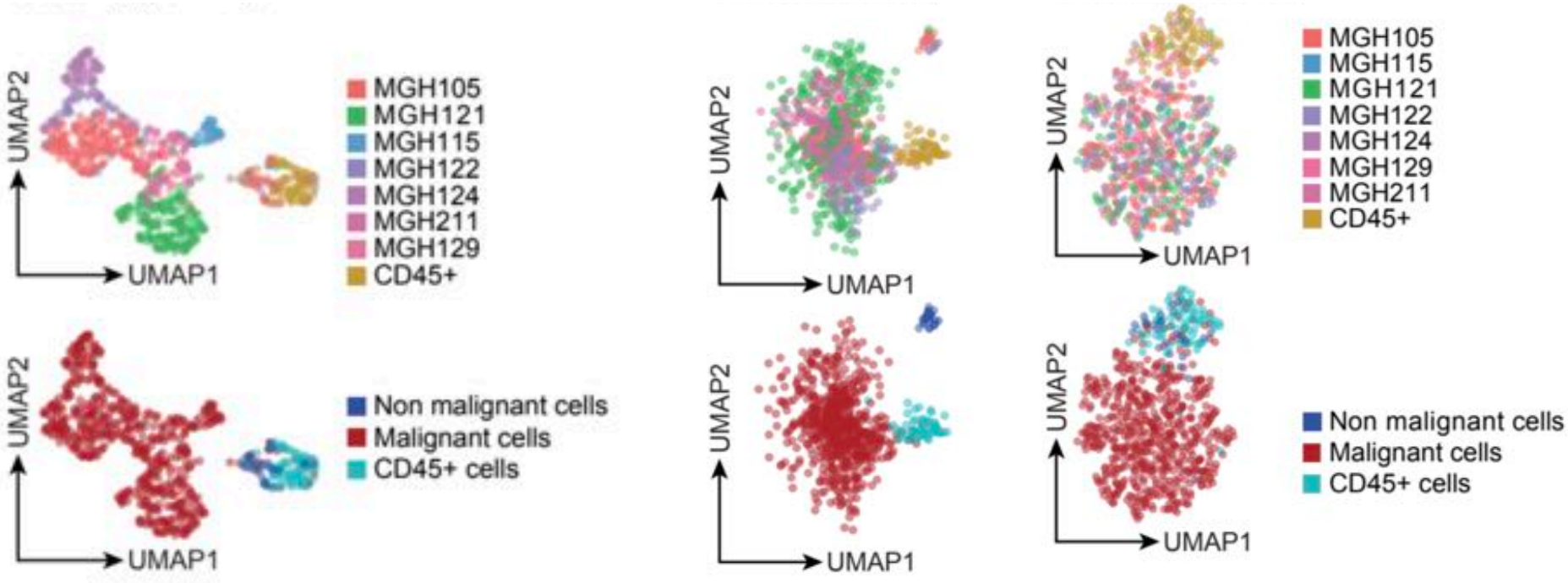
without batch correction

with batch correction

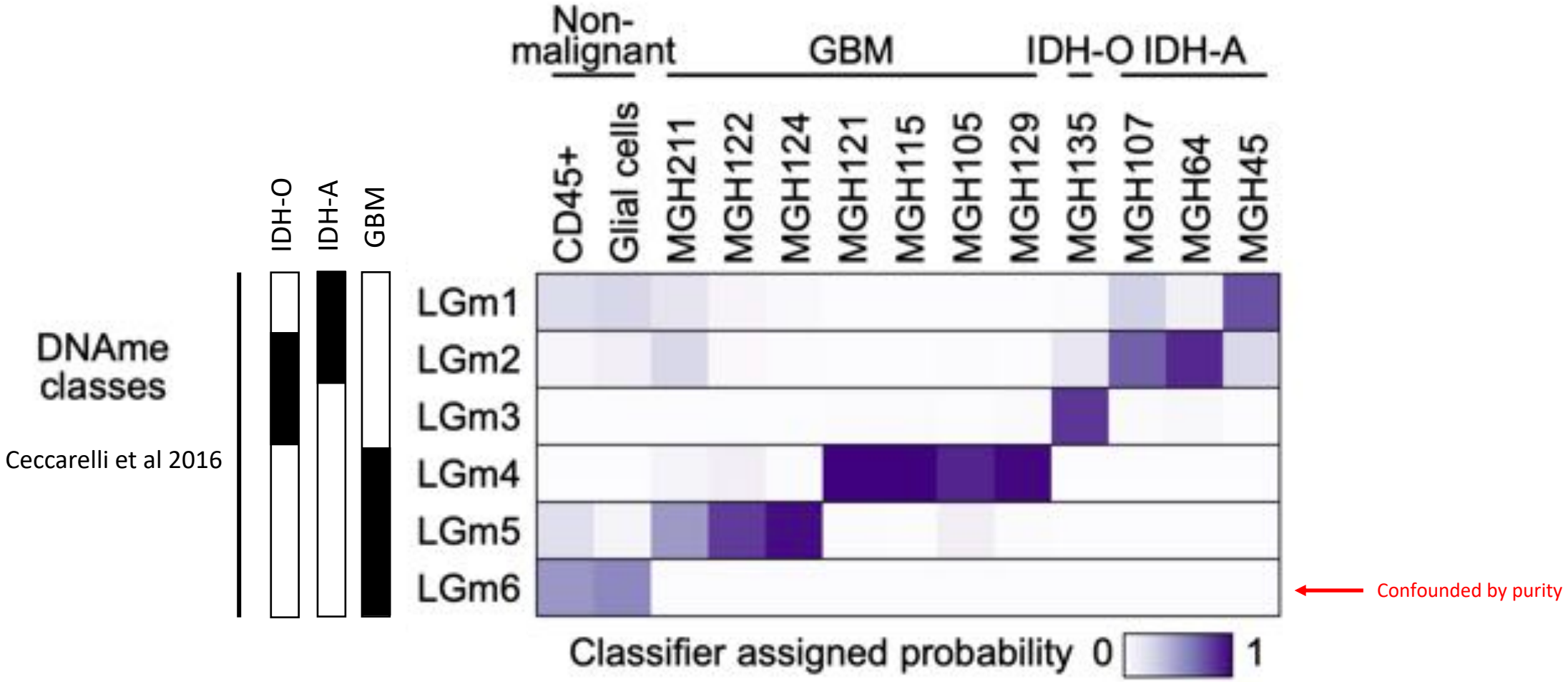
scDNA methylation

scRNA

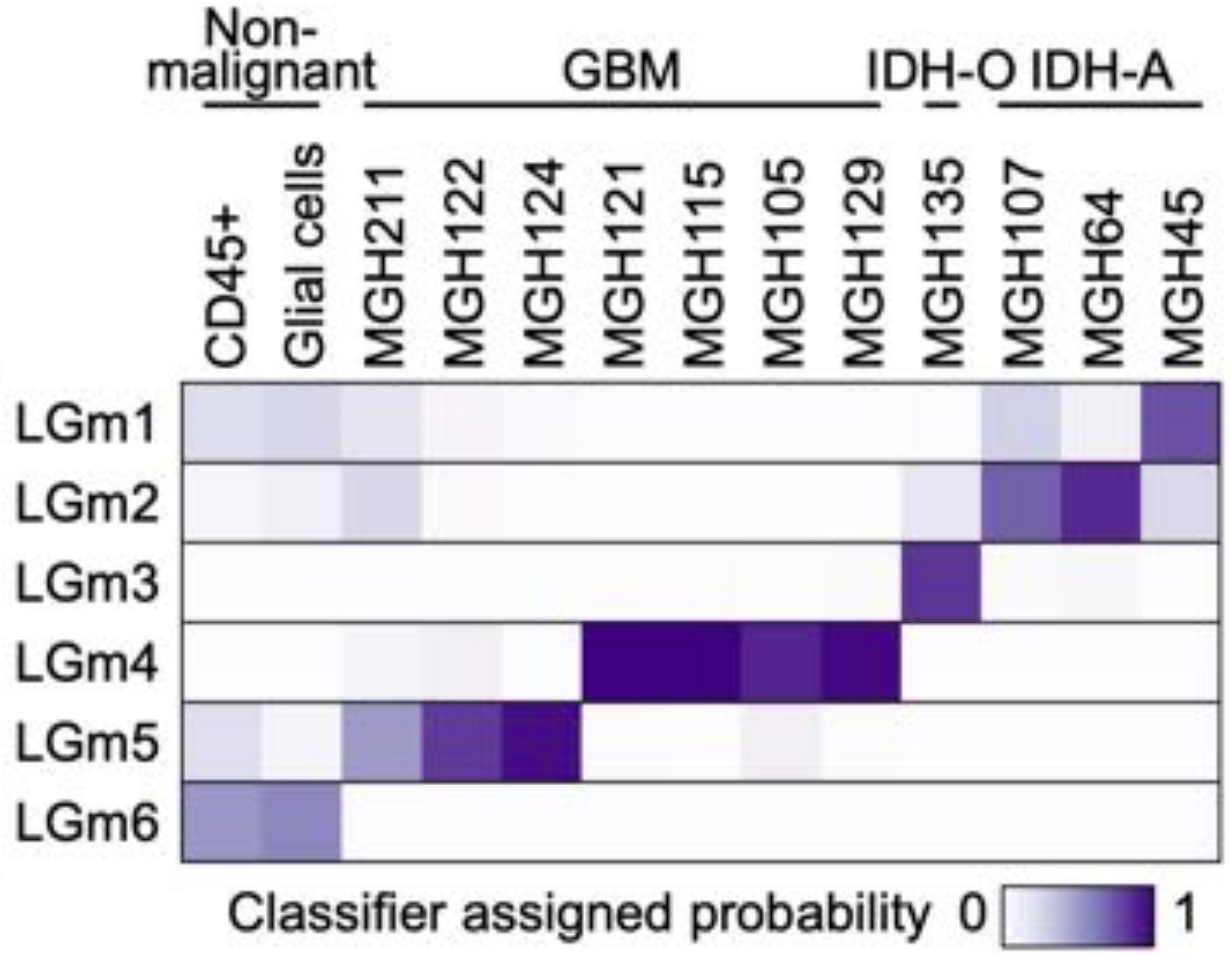
scDNA methylation



Mapping the tumors to previously defined classes recapitulates and refines previous classification



Mapping the tumors to previously defined classes recapitulates and refines previous classification

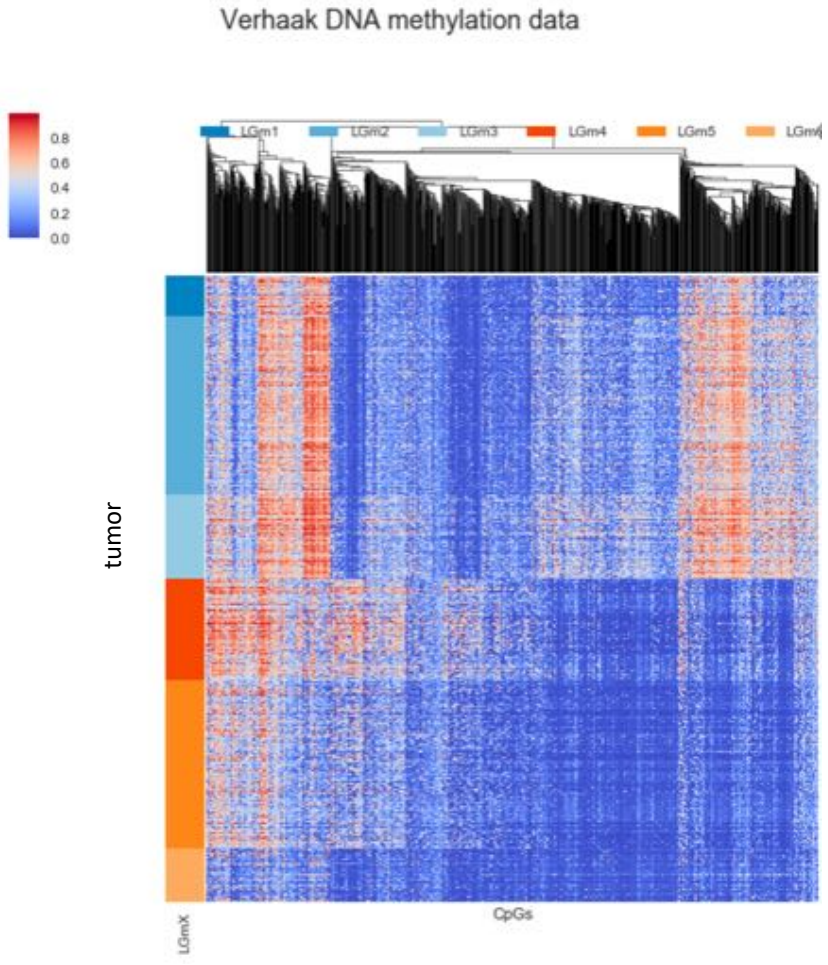


Translating single cell RRBS data to bulk 450K data is possible via machine learning techniques:

- We map scRRBS probes to the 450K probes
- Train a logistic regression classifier to translate from scRRBS to 450K

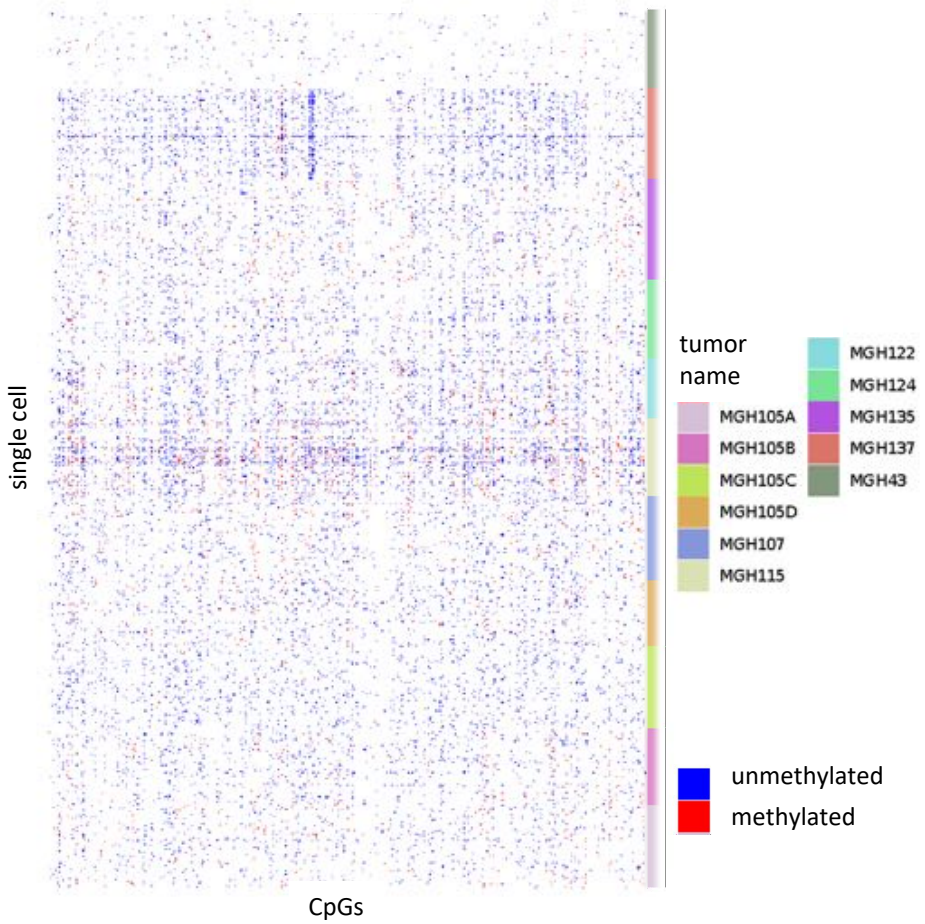
Mapping the tumors to previously defined classes recapitulates and refines previous classification

Bulk DNA methylation



Classifier
F(sc)->LGm

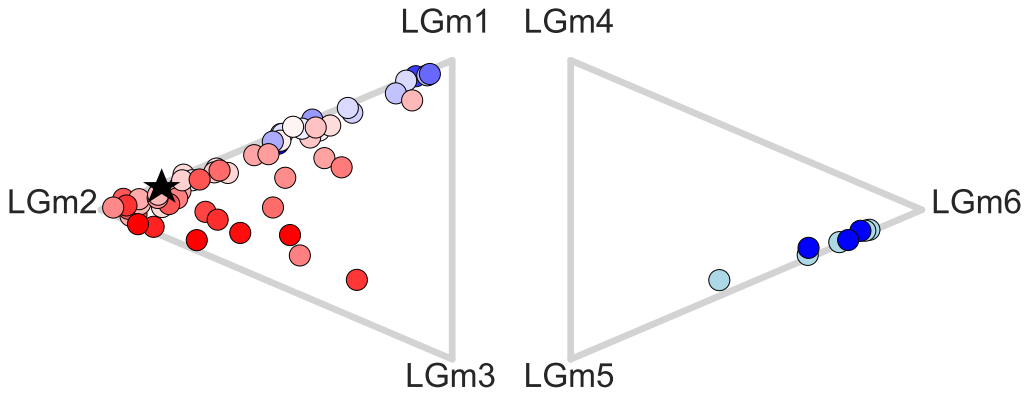
scDNA methylation



Gliomas are heterogenous on the DNA methylation level

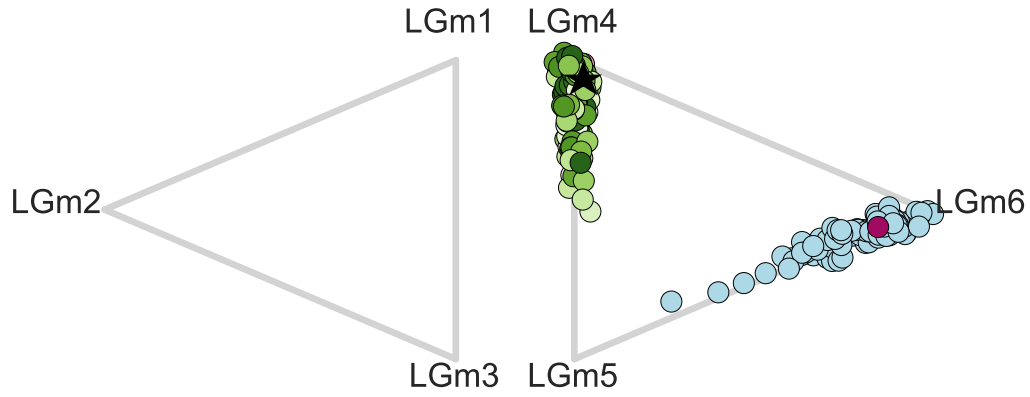
IDH-mutant

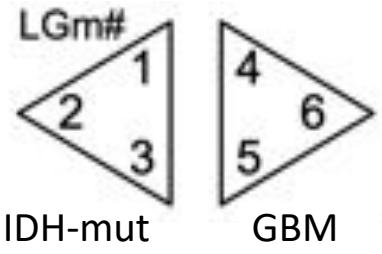
MGH64 n = 67



GBM

MGH105 n = 321





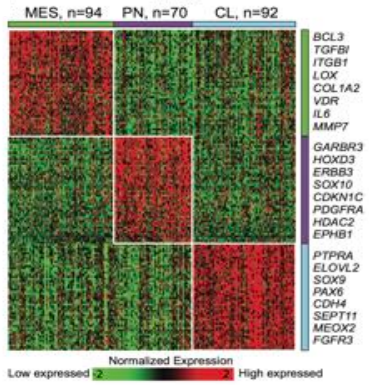
★ pseudo bulk tumor ○ single cell

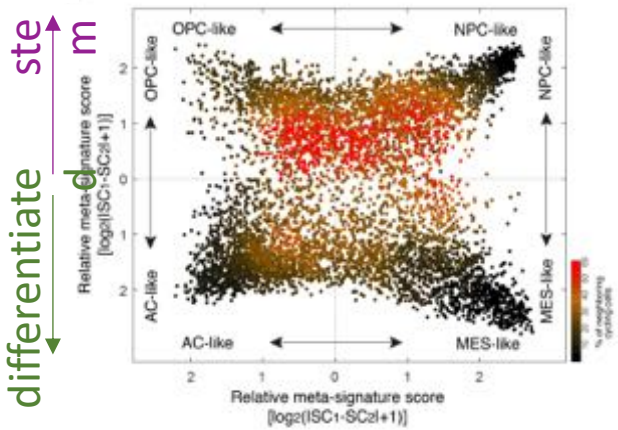
The goal: dissecting epigenetic identity of cellular states underlying glioma evolution

inter-tumor heterogeneity

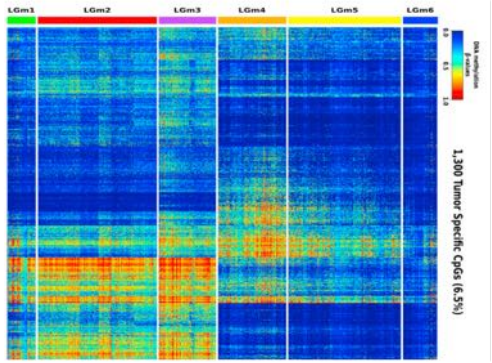
intra-tumor heterogeneity

RNA

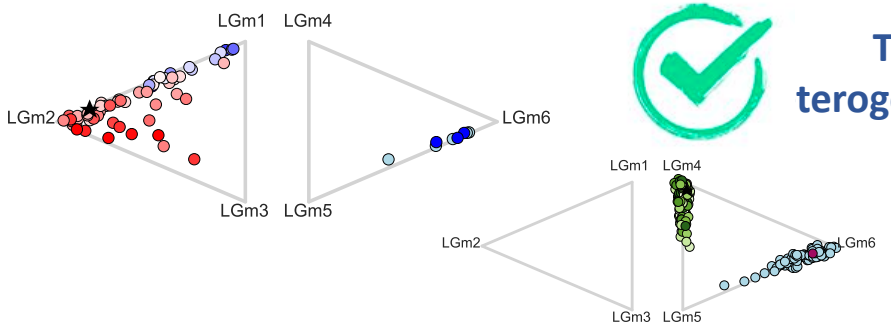




DNA methylation

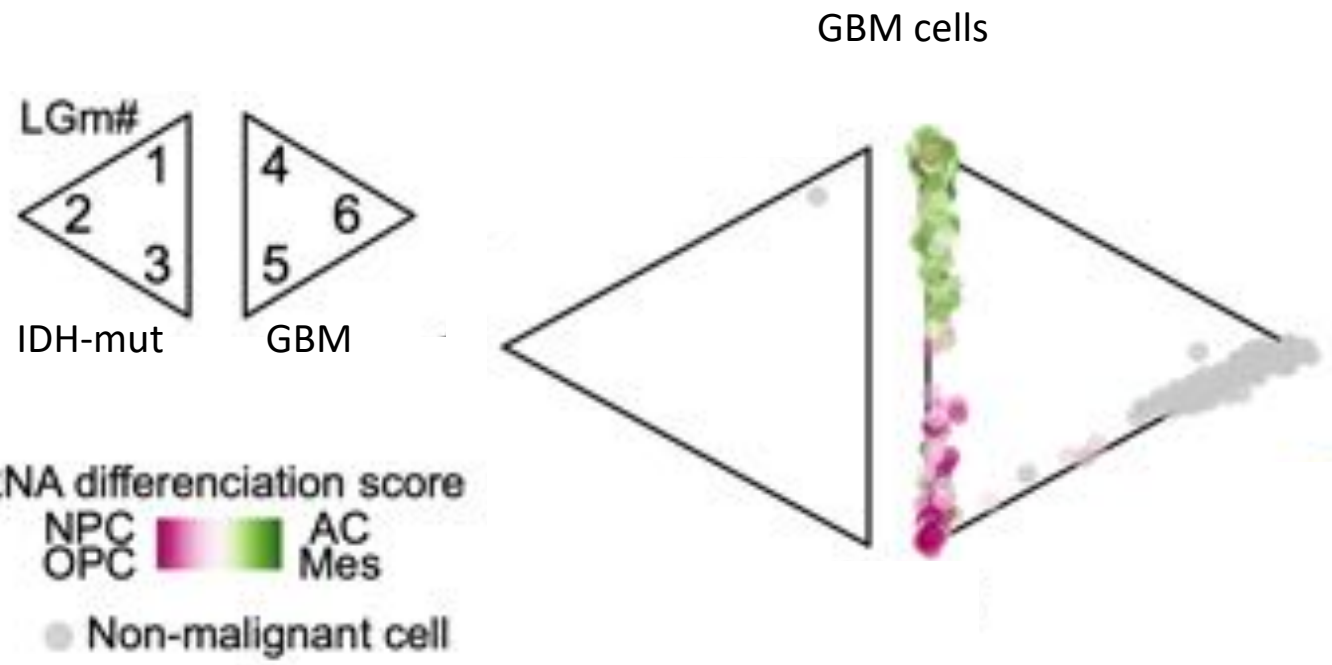


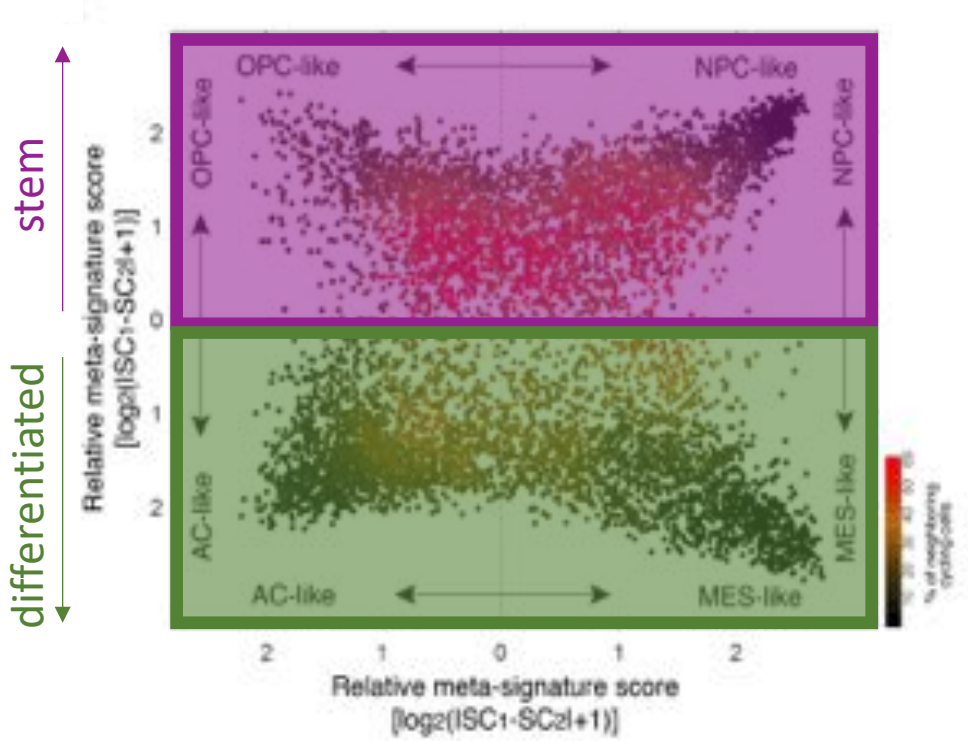
Is there a link between the RNA function and DNA methylation?



The DNA methylation terogeneity stems from different cells composition

Finding association between DNAmet state to RNA state





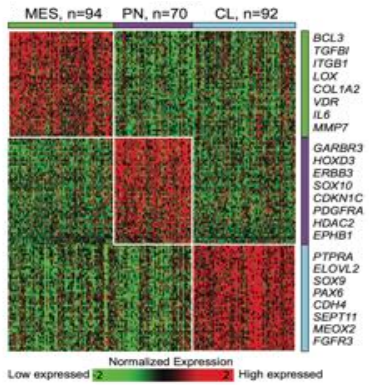
The position of each single cell represents the DNA methylation class, the color represents the RNA type

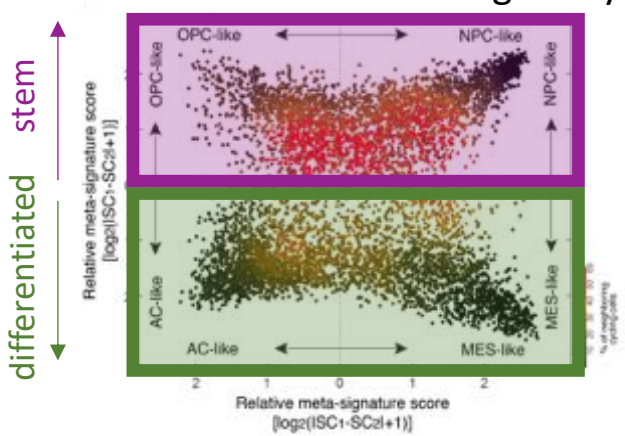
The goal: dissecting epigenetic identity of cellular states underlying glioma evolution

inter-tumor heterogeneity

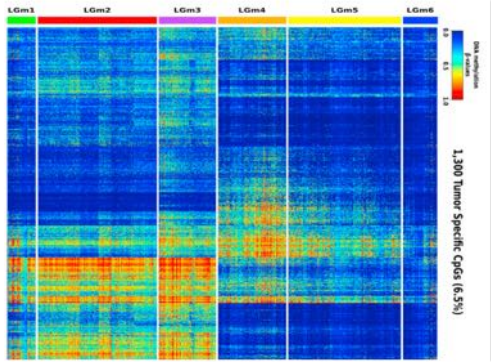
intra-tumor heterogeneity

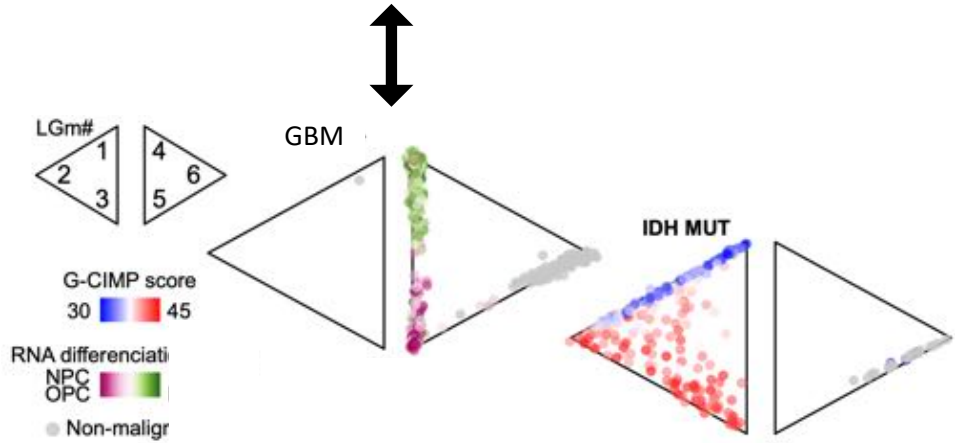
RNA



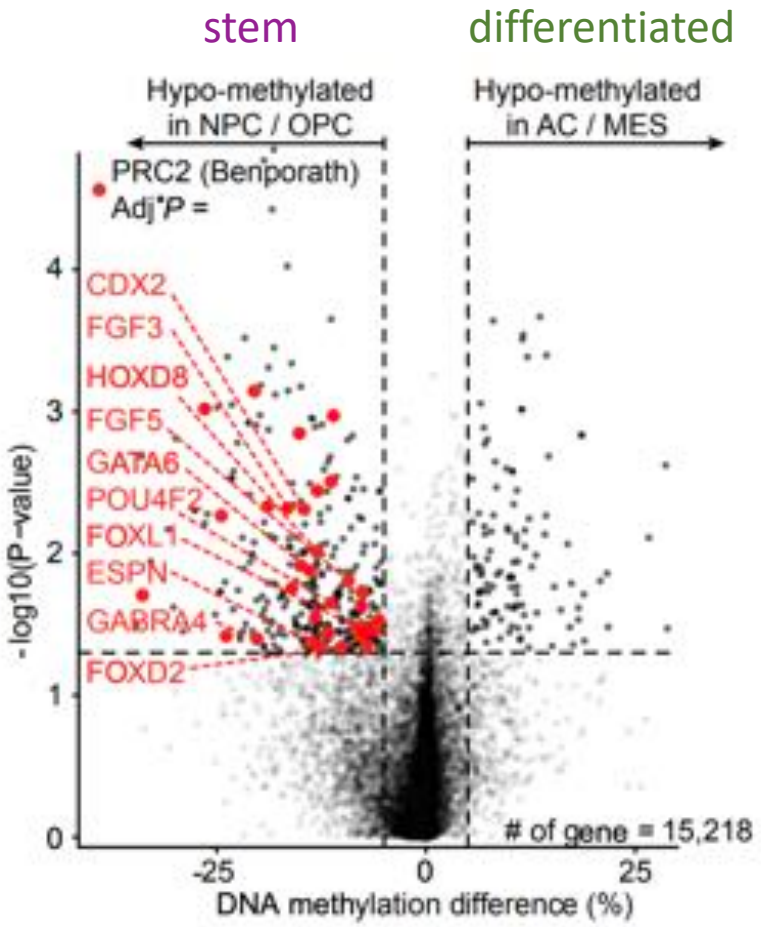


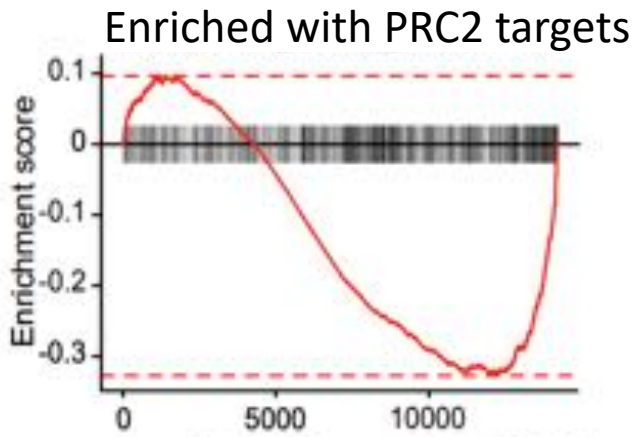
DNA methylation





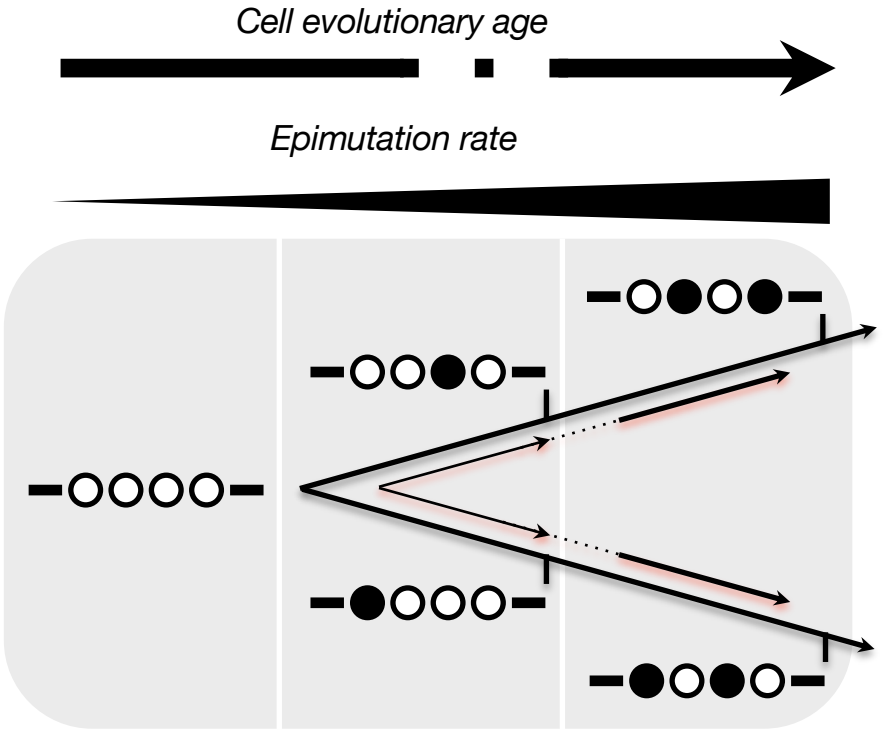
Differentially methylated events in single cells identifies PRC2 targets as hypo methylated in stem states





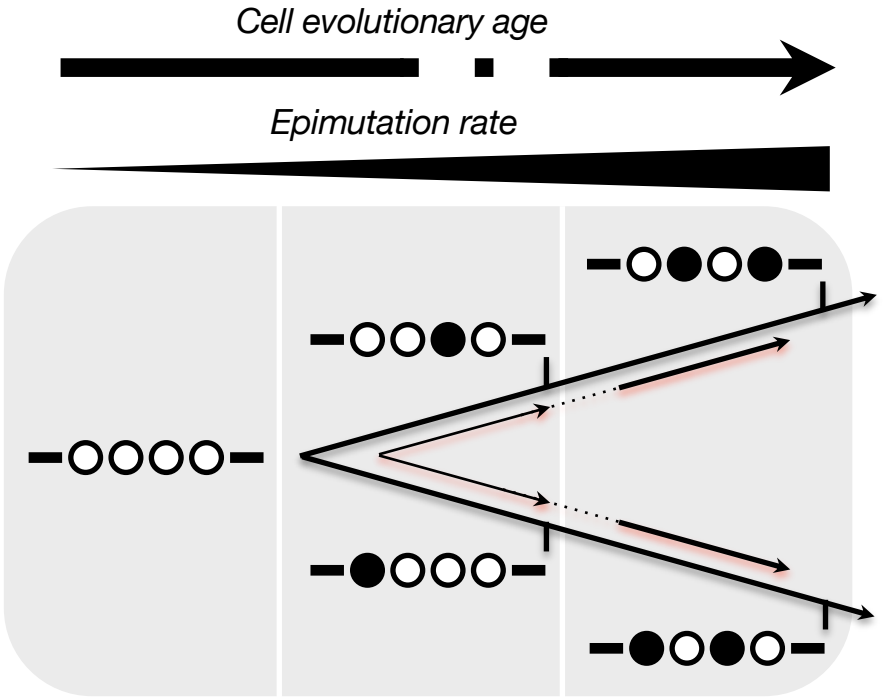
Analysis and figure by Ronan Chaligne and Federico Gaiti

Epimutation rate - a molecular clock



Slide by Ronan Chaligne

Motivation for epi-phylogenetic trees

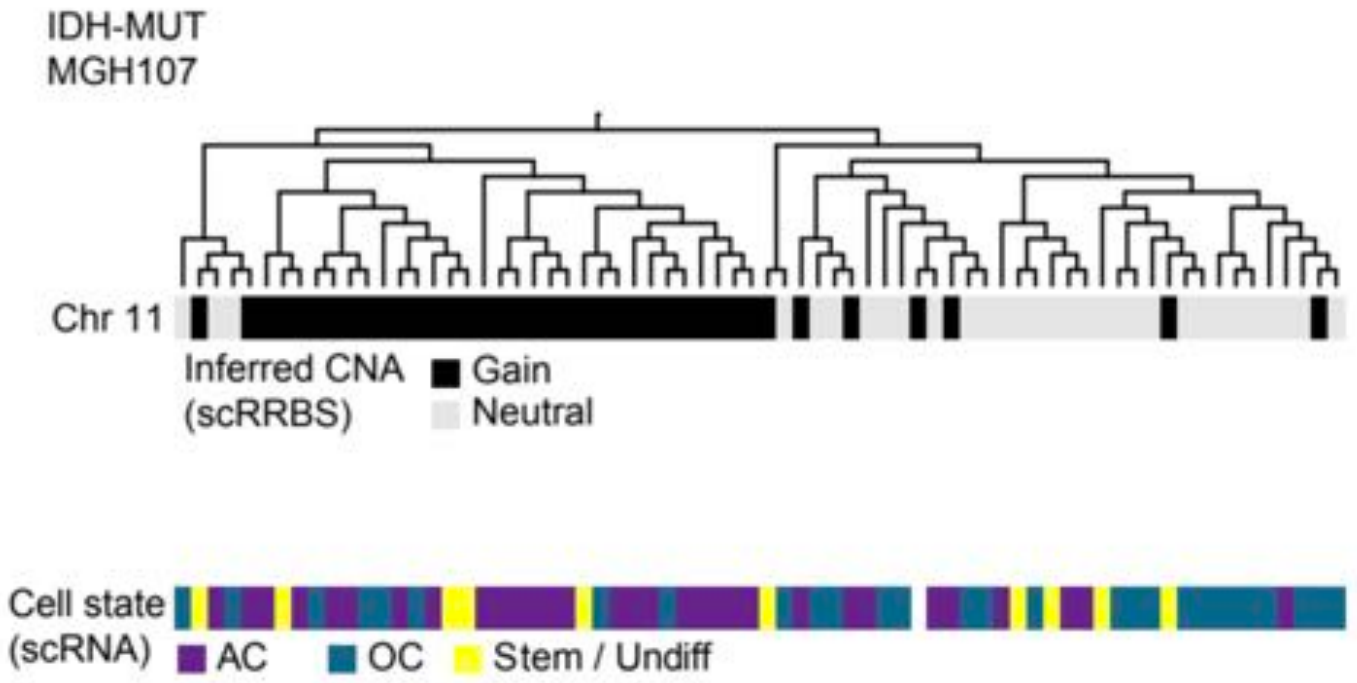


Epimutation rate >> mutation rate

Lineage reconstruction at low sequencing cost and despite single-cell data sparsity

Slide by Ronan Chaligne

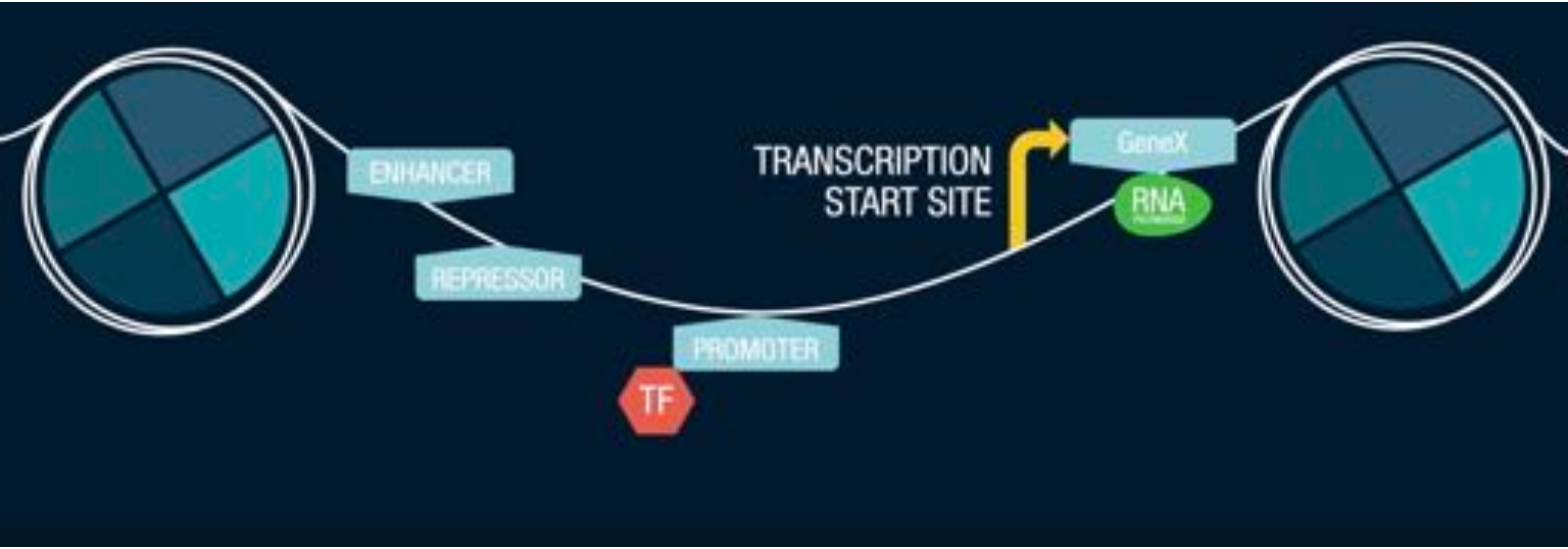
epi-phylogenetic trees reveal tumorigenesis in glioma



scATACseq

ATACseq measures accessible chromatin

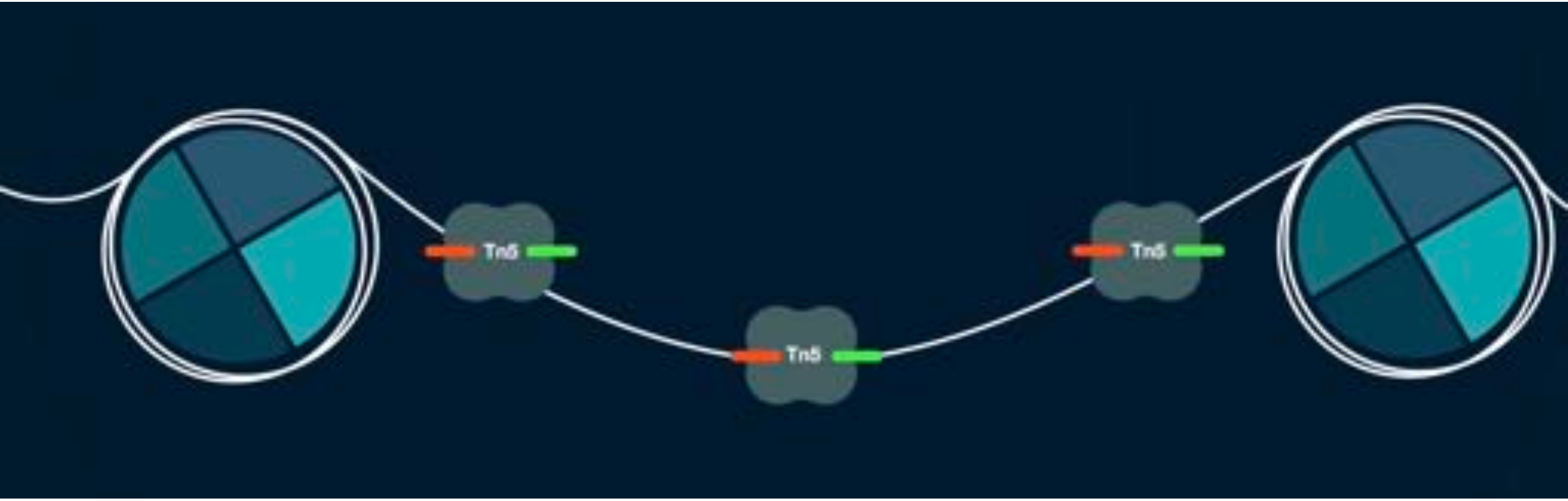
In the cell nucleus, the chromosomes contain tightly packed chromatin material. Part of the chromatin is open and accessible to many regulatory factors who control the expression and suppression of a variety of genes.



"How Single-Cell ATAC-Seq Works", Bio-Rad Laboratories

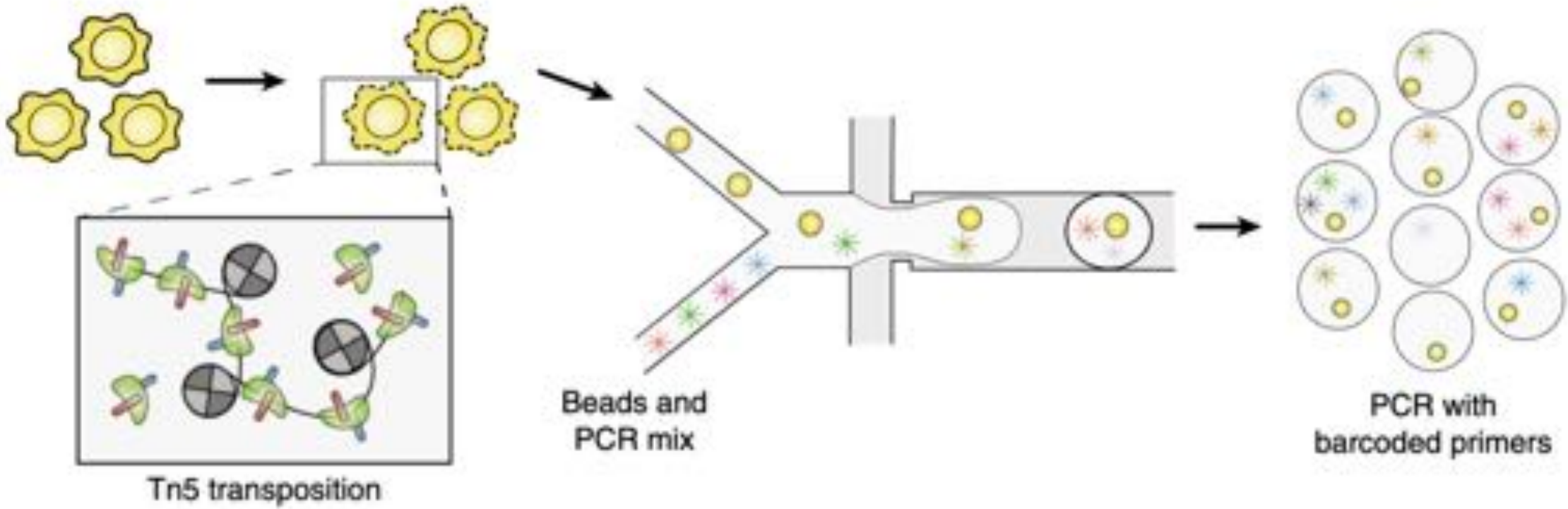
ATACseq measures accessible chromatin

ATACseq (as well as scATACseq) measures how open this piece of DNA is. This openness is a proxy of how easily a transcription factor can bind these parts of the genome. ATACseq measures by using an enzyme called Tn5 transposase which binds open chromatin and inserts DNA sequencing adapters.



"How Single-Cell ATAC-Seq Works", Bio-Rad Laboratories

Each cell is captures in a droplet with cell barcode



Lareau, Duarte, Chew et al. 2019, Nature Biotechnology

single cell resolution is crucial

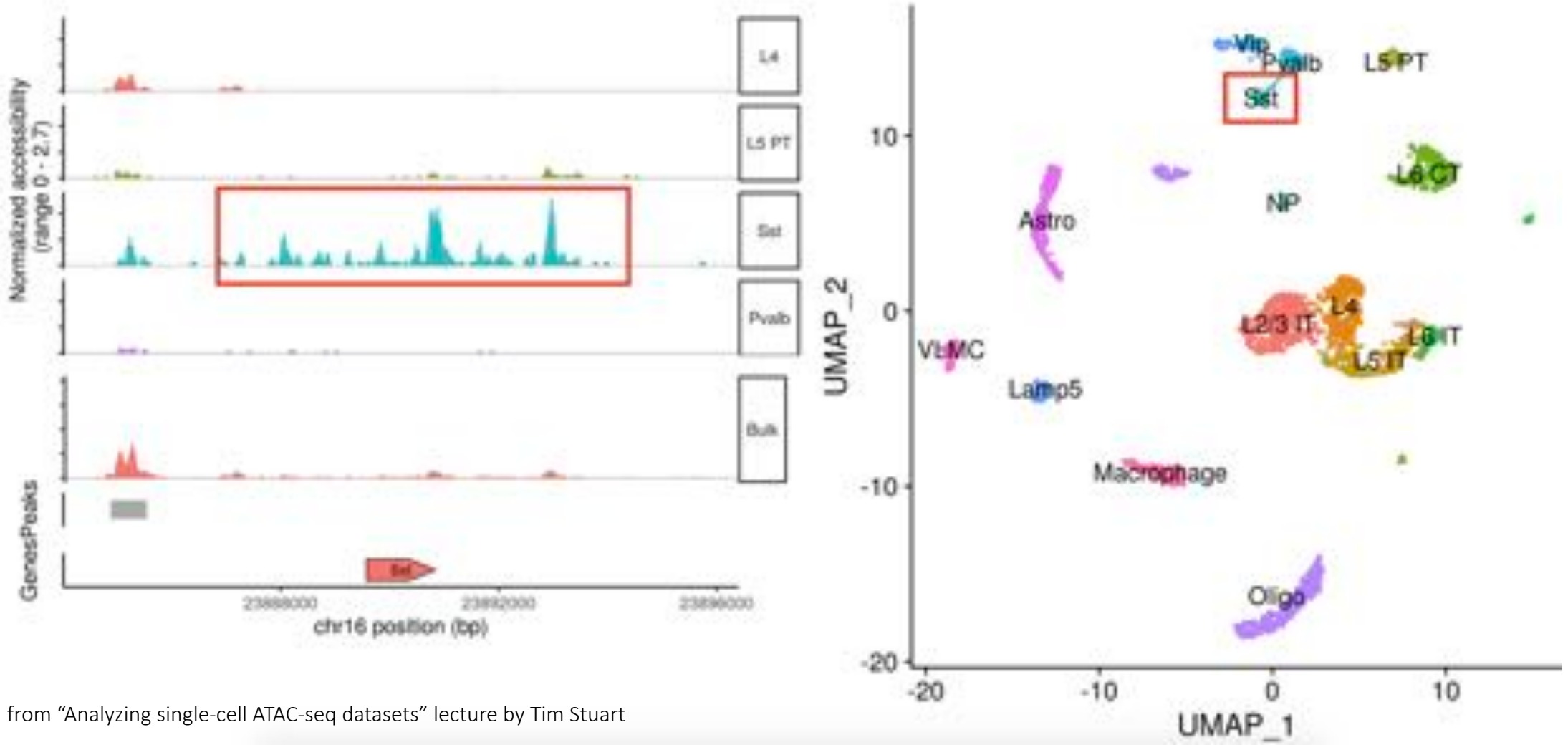


figure adopted from “Analyzing single-cell ATAC-seq datasets” lecture by Tim Stuart

scATACseq data

A full list of **all** unique fragments across all single cells, as opposed to only reads that map to peaks.

1. Indexed fragment file

chrom	start	stop	barcode	reads
chr1	3000141	3000517	GGTTGCGAGCCGCAAA-1	3
chr1	3000159	3000373	CTCAGCTAGTGTCACT-1	1
chr1	3000431	3000621	GAAGTCTGTAACACTC-1	1

2. Large sparse matrix

	AAACGAAAGAGTTTGA-1	AAACGAAAGCGAGCTA-1
chr1:565107-565550	.	.
chr1:569174-569639	.	.
chr1:713460-714823	.	2
chr1:752422-753038	.	.
chr1:762106-763359	.	4

Each value in the matrix represents the number of Tn5 cut sites for each single barcode (i.e. cell) that map within each peak

figure adopted from “Analyzing single-cell ATAC-seq datasets” lecture by Tim Stuart

Defining features for scATACseq

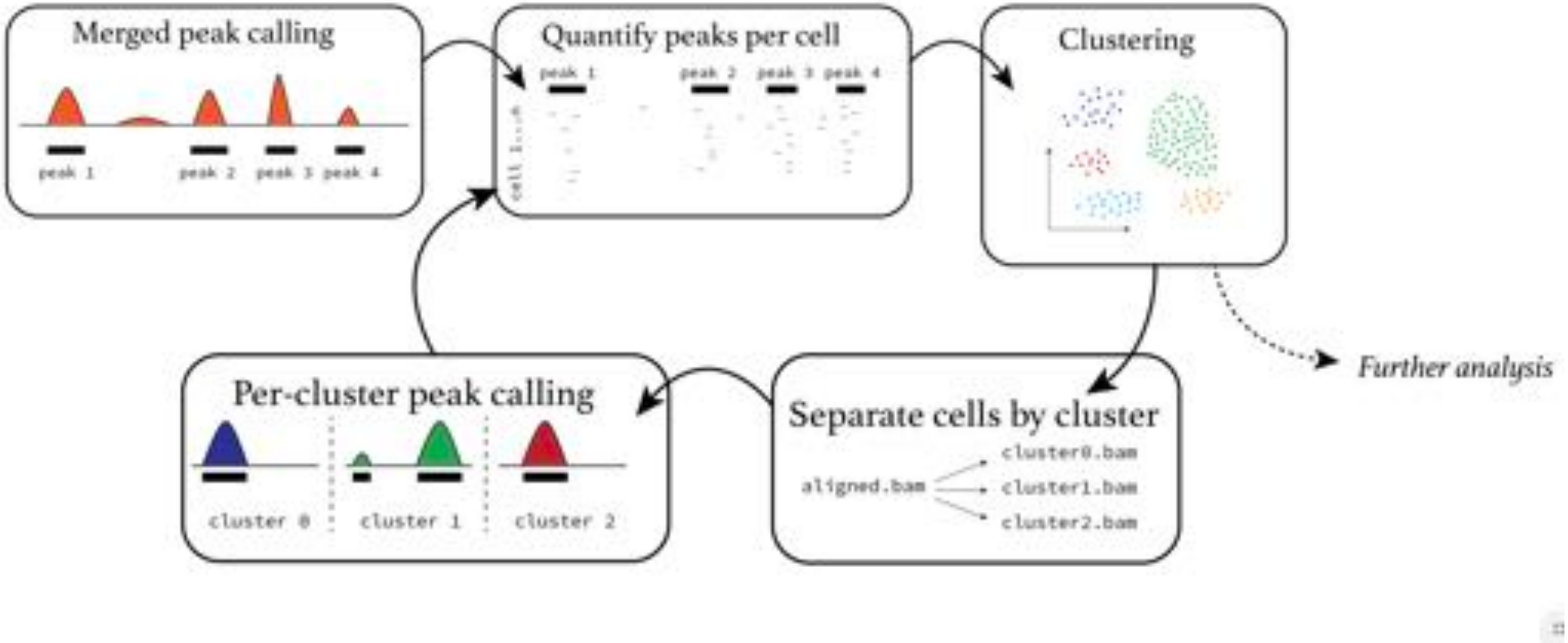


figure adopted from “Analyzing single-cell ATAC-seq datasets” lecture by Tim Stuart

scATACseq data

1. Indexed fragment file

chrom	start	stop	barcode	reads
chr1	3000141	3000517	GGTTGCGAGCCGCAAA-1	3
chr1	3000159	3000373	CTCAGCTAGTGTCACT-1	1
chr1	3000431	3000621	GAAGTCTGTAACACTC-1	1

2. Large sparse matrix

	AAACGAAAGAGTTTGA-1	AAACGAAAGCGAGCTA-1
chr1:565107-565550	.	.
chr1:569174-569639	.	.
chr1:713460-714823	.	2
chr1:752422-753038	.	.
chr1:762106-763359	.	4

figure adopted from “Analyzing single-cell ATAC-seq datasets” lecture by Tim Stuart

Use LSI instead of PCA

Originally developed for topic modeling / natural language processing (Deerwester et al. 1990)

First applied to scATAC-seq in 2015 (Cusanovich et al. *Science*)

Cell = document, peak = word

LSI combines local (within-cell) and global (across-population) weighting functions then applies singular value decomposition

Multiple different local and global weighting functions that can be used

figure adopted from “Analyzing single-cell ATAC-seq datasets” lecture by Tim Stuart

scATAcseq yields well separated clusters

```
pbmc <- RunUMAP(object = pbmc, reduction = 'lsi', dims = 2:30)
pbmc <- FindNeighbors(object = pbmc, reduction = 'lsi', dims = 2:30)
pbmc <- FindClusters(object = pbmc, verbose = FALSE, algorithm = 3)
DimPlot(object = pbmc, label = TRUE) + NoLegend()
```

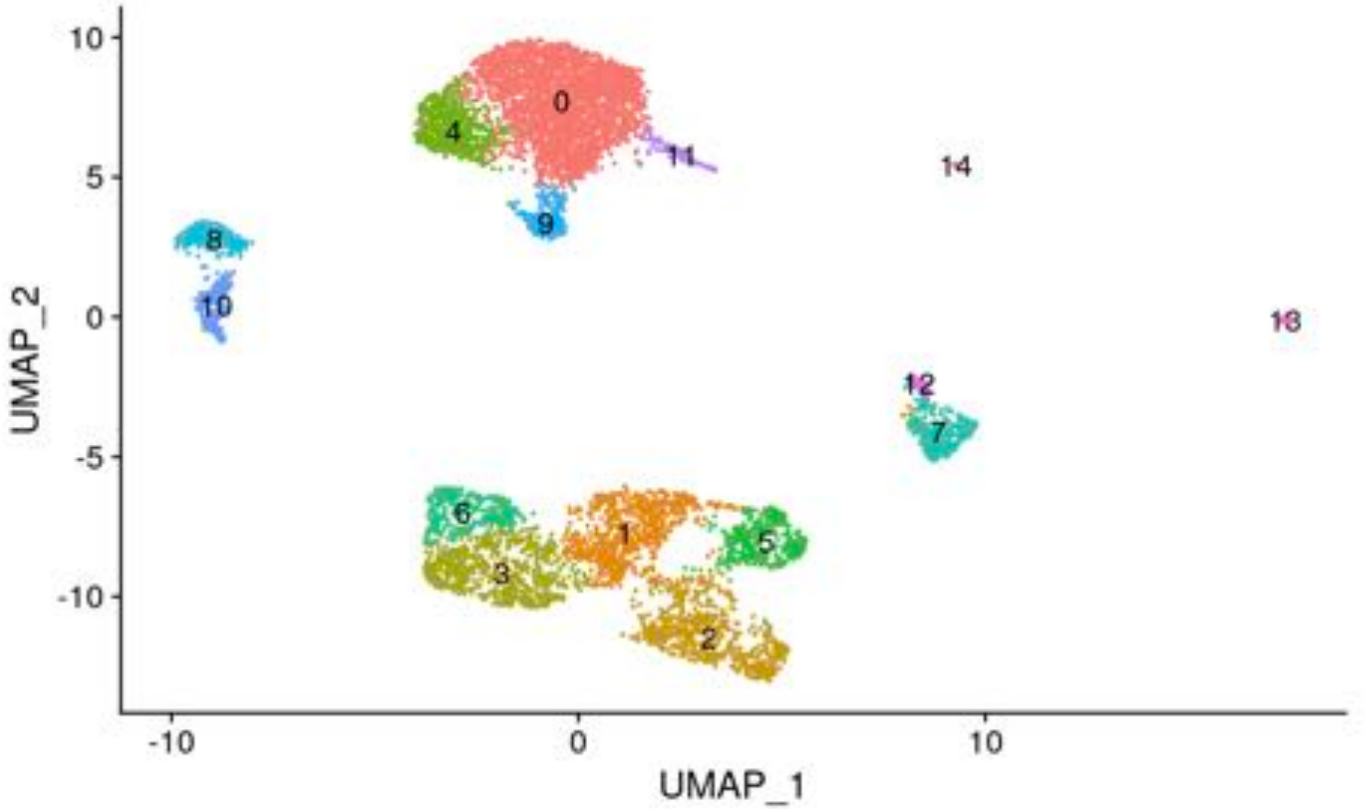


figure adopted from signac tutorial, https://satijalab.org/signac/articles/pbmc_vignette.html

Use fragments file to infer gene activity and annotate clusters

Quantify the activity of each gene in the genome by assessing the chromatin accessibility associated with each gene: count the number of fragments for each cell that map to the promoter + gene body

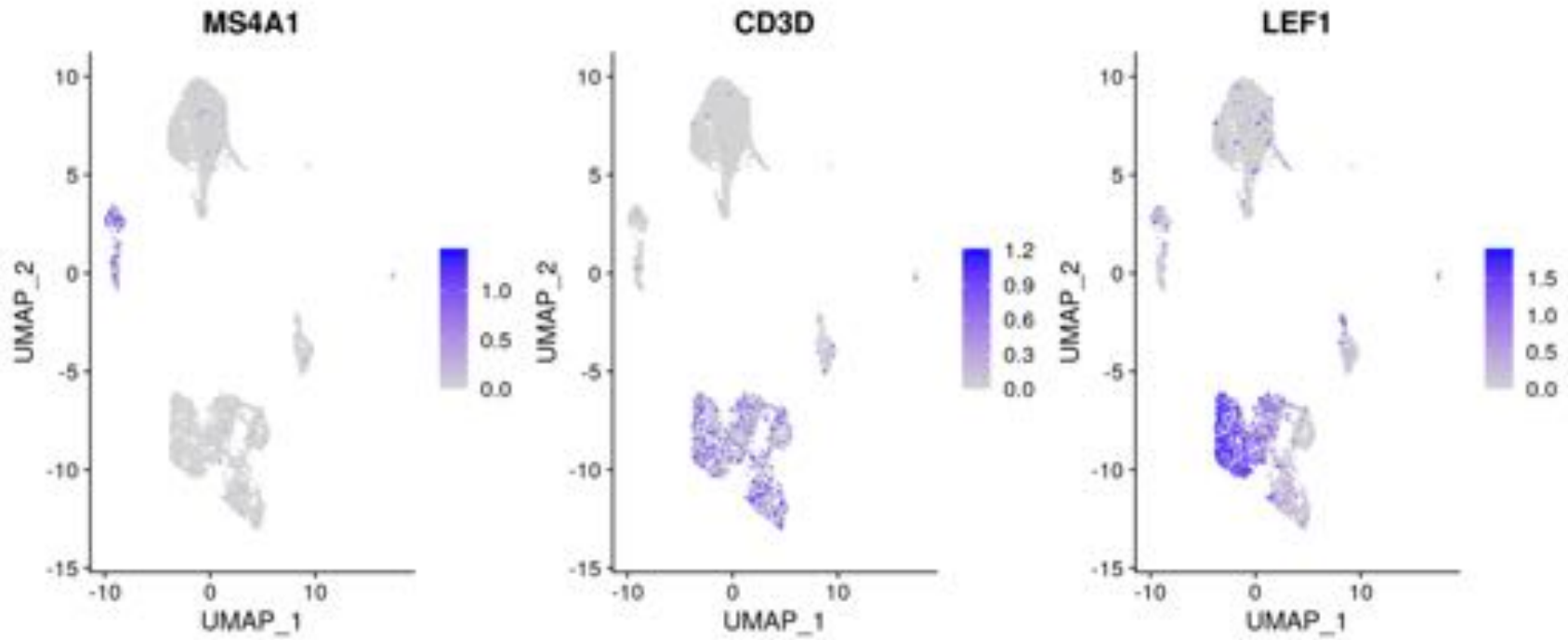


figure adopted from signac tutorial, https://satijalab.org/signac/articles/pbmc_vignette.html

Integrating with scRNA-seq data using CCA

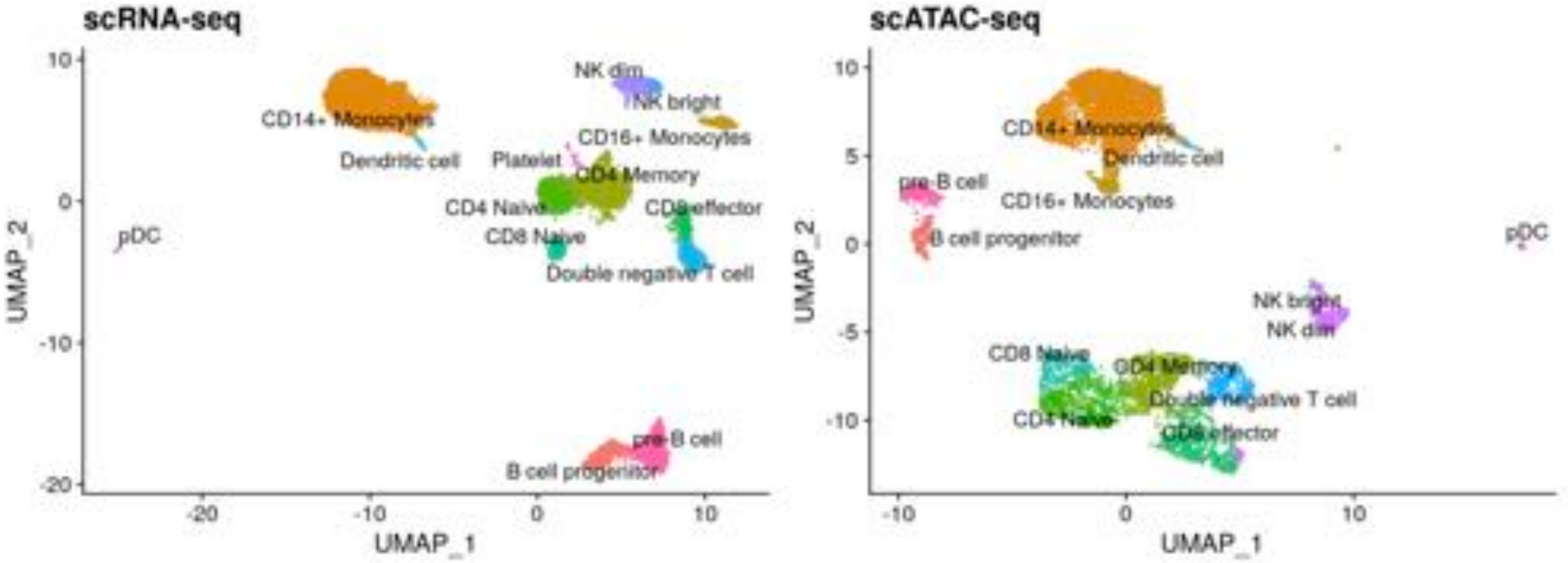


figure adopted from signac tutorial, https://satijalab.org/signac/articles/pbmc_vignette.html

Finding overrepresented motifs

To identify potentially important cell-type-specific regulatory sequences, signac searches for DNA motifs that are overrepresented in a set of peaks that are differentially accessible between cell types.

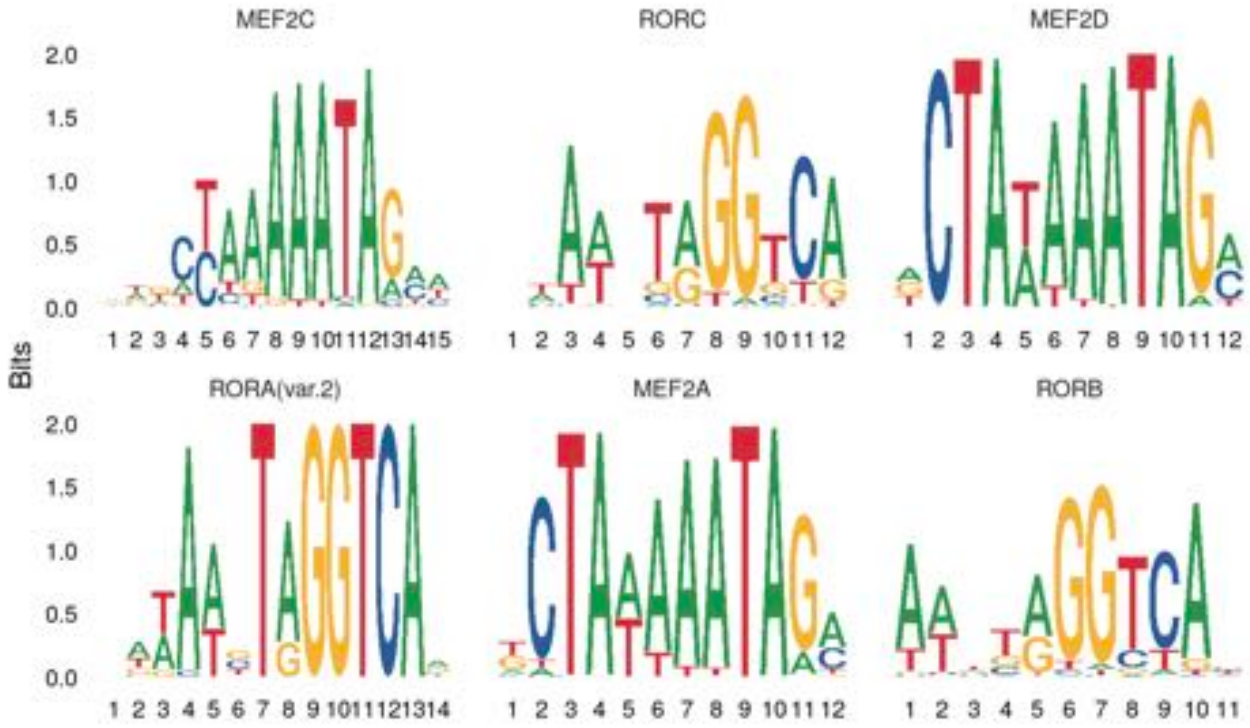


figure adopted from Motif analysis with Signac tutorial, https://satijalab.org/signac/articles/motif_vignette.html

Computing motif activities

ChromVAR identifies motifs associated with variability in chromatin accessibility between cells.

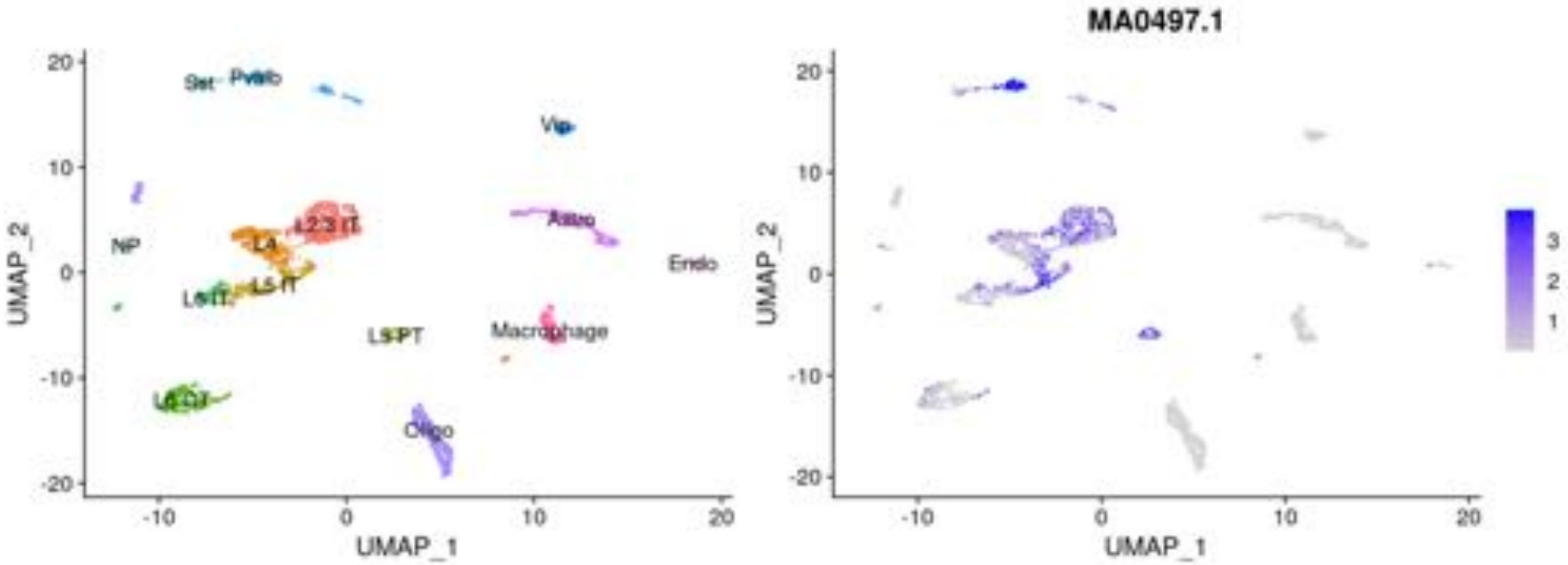


figure adopted from Motif analysis with Signac tutorial, https://satijalab.org/signac/articles/motif_vignette.html

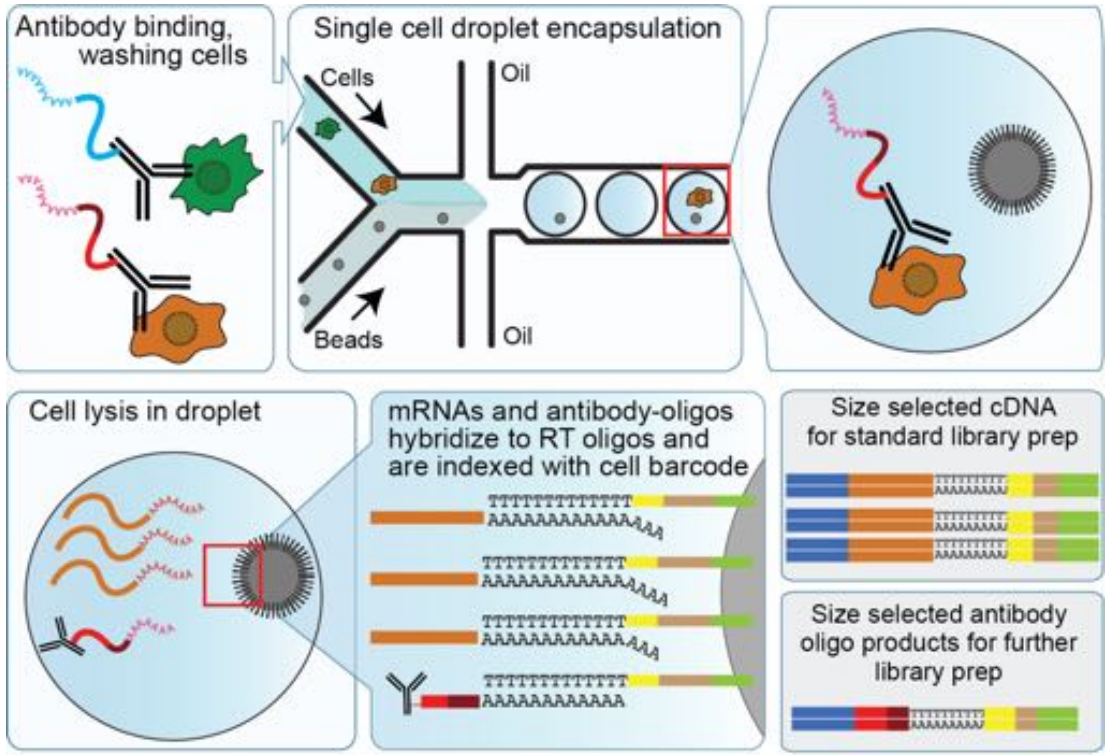
Software solutions

Signac	Extension of Seurat of single-cell chromatin data	https://github.com/timoast/signac
SnapATAC	Analysis pipeline for single-cell ATAC-seq	https://github.com/r3fang/SnapATAC
cisTopic	Probabilistic modeling of cis-regulatory topics	https://github.com/aertslab/cisTopic
chromVAR	Chromatin variability across regions	https://github.com/GreenleafLab/chromVAR https://github.com/caleblareau/gchromVAR
CICERO	Predicting the cis-regulatory landscape	https://github.com/cole-trapnell-lab/cicero-release

figure adopted from “Analyzing single-cell ATAC-seq datasets” lecture by Tim Stuart

CITEseq

CITEseq protocol

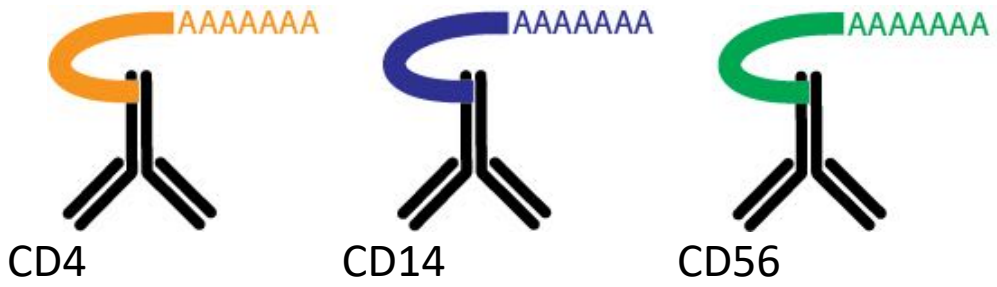


Count matrix1: scRNAseq

Count matrix2: CITEseq

Input

- List of cell surface proteins of interest:
- For example, I chose: CD4, CD14, CD56



Output

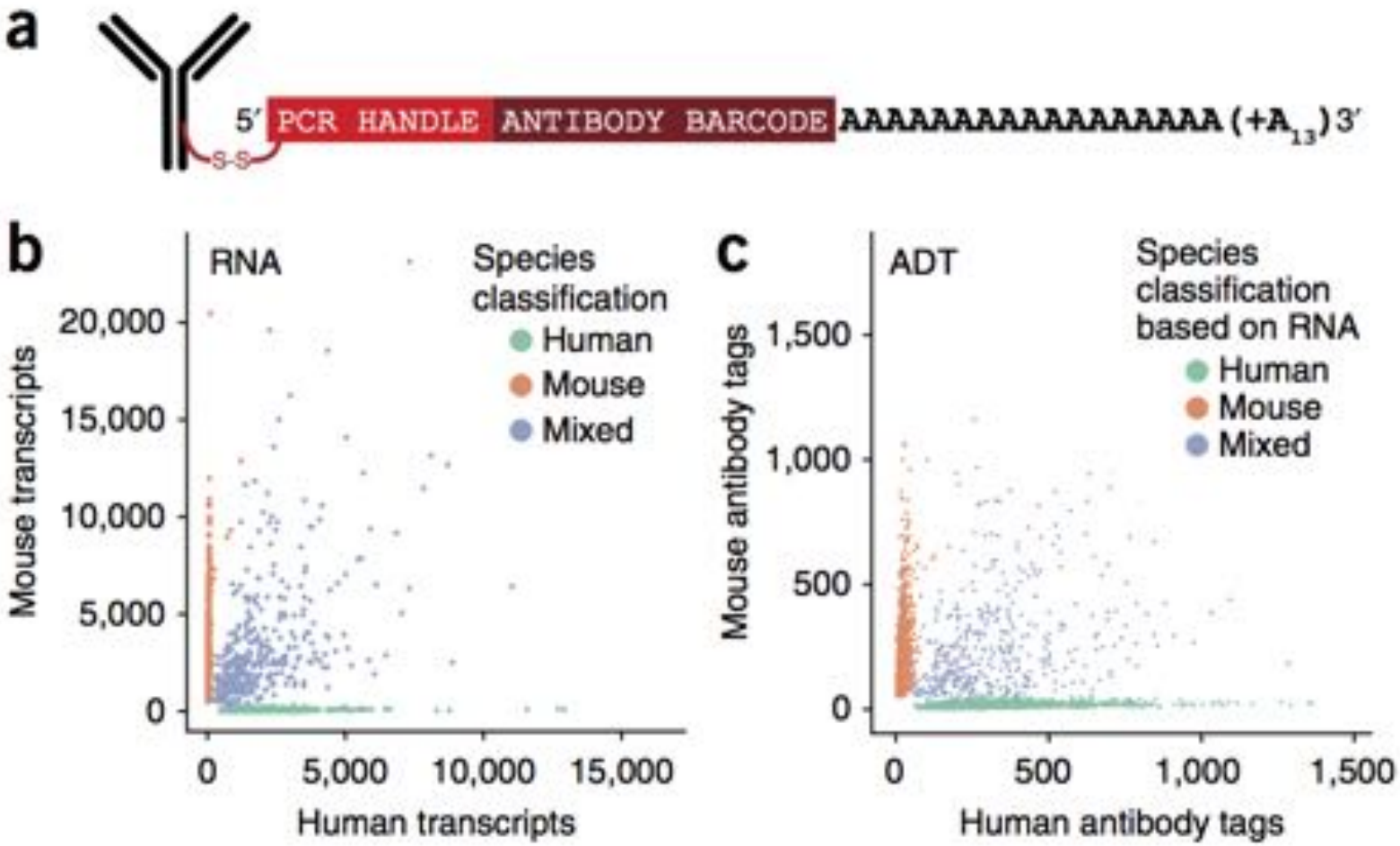
scRNAseq

	cell1	cell2	cell3	cell4
Gene1	0	0	0	4
Gene2	0	2	0	0
Gene3	0	0	0	0

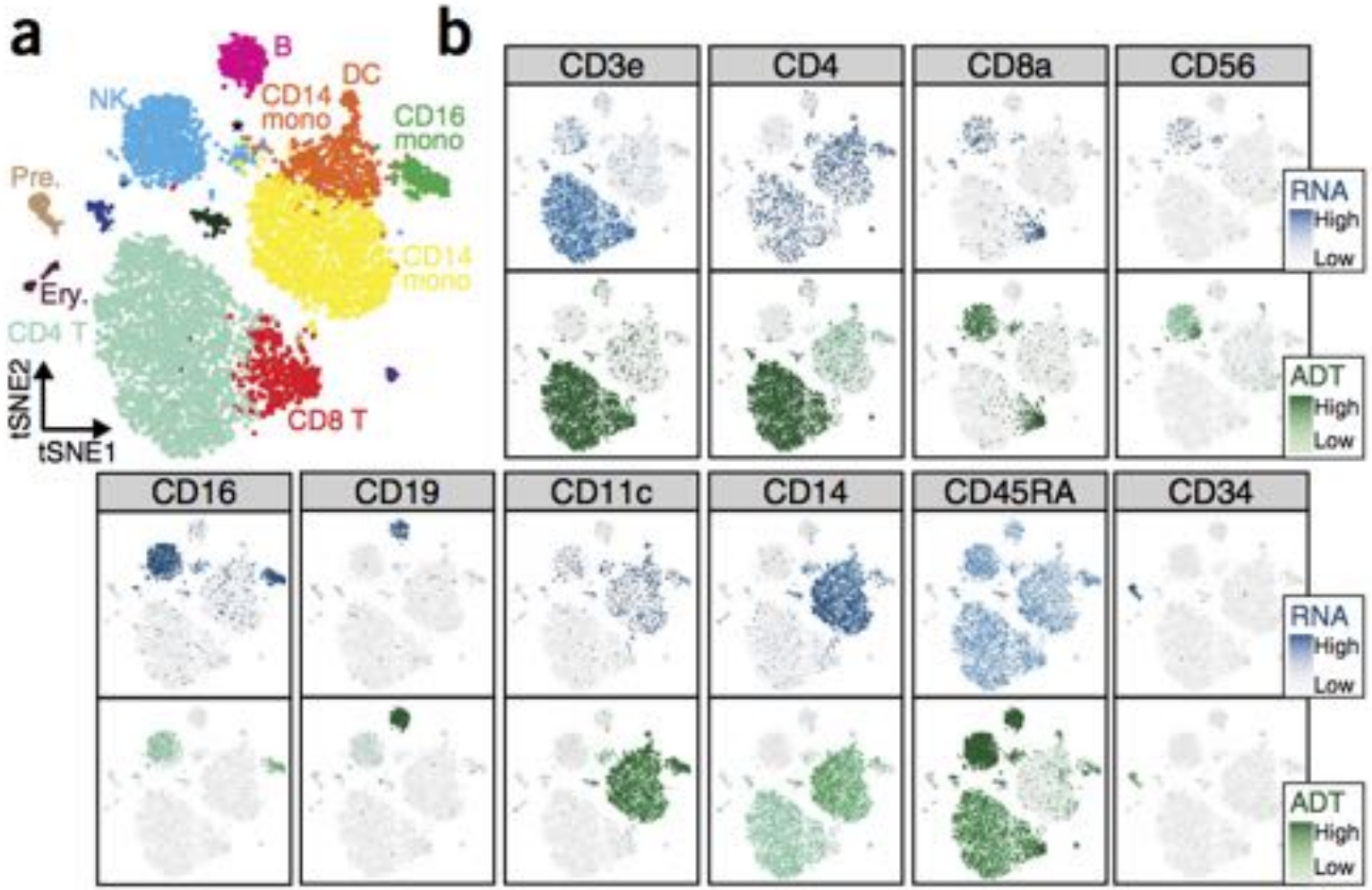
CITEseq

	cell1	cell2	cell3	cell4
Antibody for CD4	100	0	0	38
Antibody for CD14	0	0	45	0
Antibody for CD56	2	0	0	0

Simultaneous protein and RNA measurement in single cells

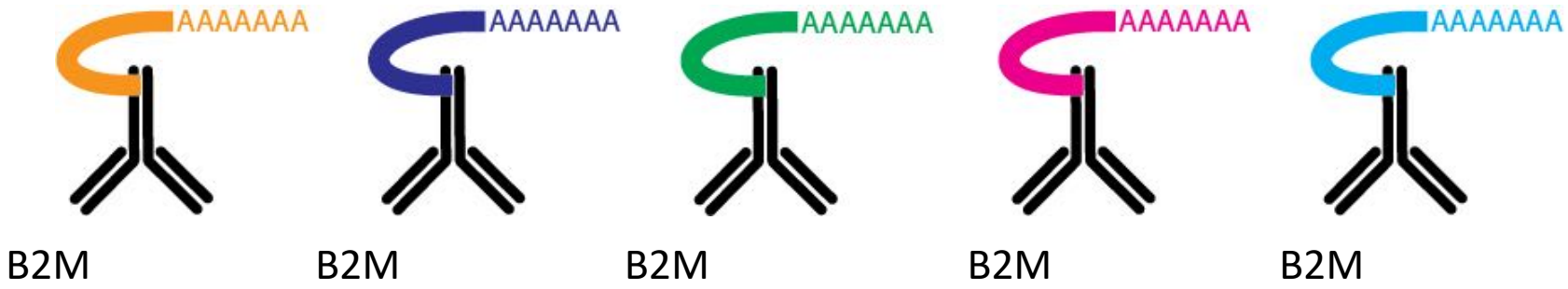


Simultaneous protein and RNA measurement in single cells



Cell hashing using CITEseq

- Cell Hashing is a method that enables sample multiplexing and super-loading on single cell RNA-sequencing platforms.
- Cell Hashing uses a series of oligo-tagged antibodies against ubiquitously expressed surface proteins with different barcodes to uniquely label cells from distinct samples, which can be subsequently pooled in one scRNA-seq run.



Overloading

By sequencing these tags alongside the cellular transcriptome, we can assign each cell to its sample of origin, and robustly identify doublets originating from multiple samples.