An Overview of single-cell ATAC-seq and Multiome analysis

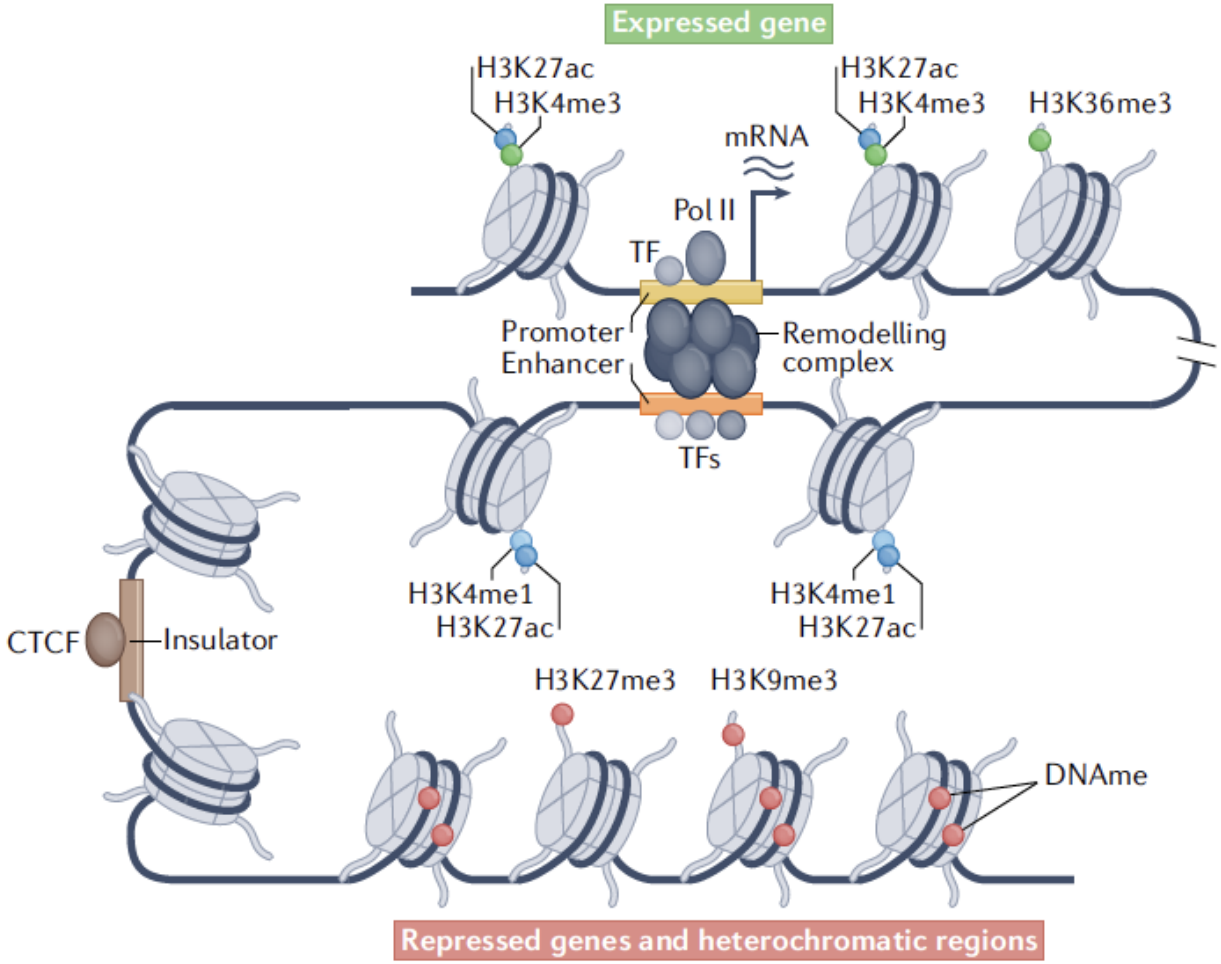
CDN Workshop

June 9, 2025

Dongwon Lee, PhD
Division of Nephrology
Boston Children's Hospital &
Harvard Medical School
Broad Institute of MIT and Harvard

cis-regulatory elements & Open Chromatin

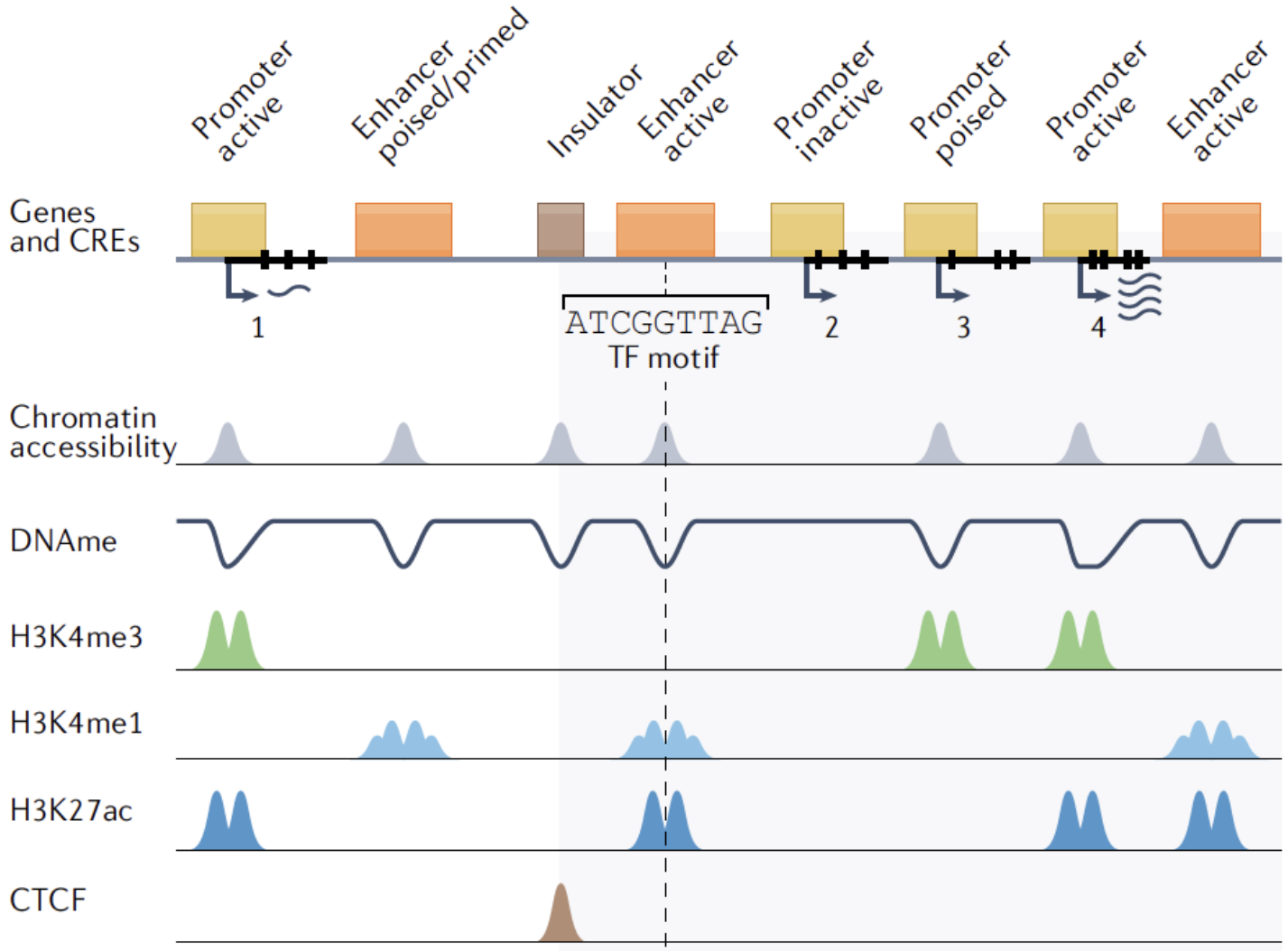
cis-regulatory elements (CREs)



- Regions of non-coding DNA that regulate the transcription of genes
- Promoter, Enhancer, Insulator, Silencers
- Different epigenetic marks are associated with different types of CREs
- Clustering of transcription factor binding sites (TFBSs)
- Sequence-specific DNA binding of Transcription Factors (TFs)

(Preissl et al., Nat Rev Genet, 2023)

Open chromatin is a general marker of CREs

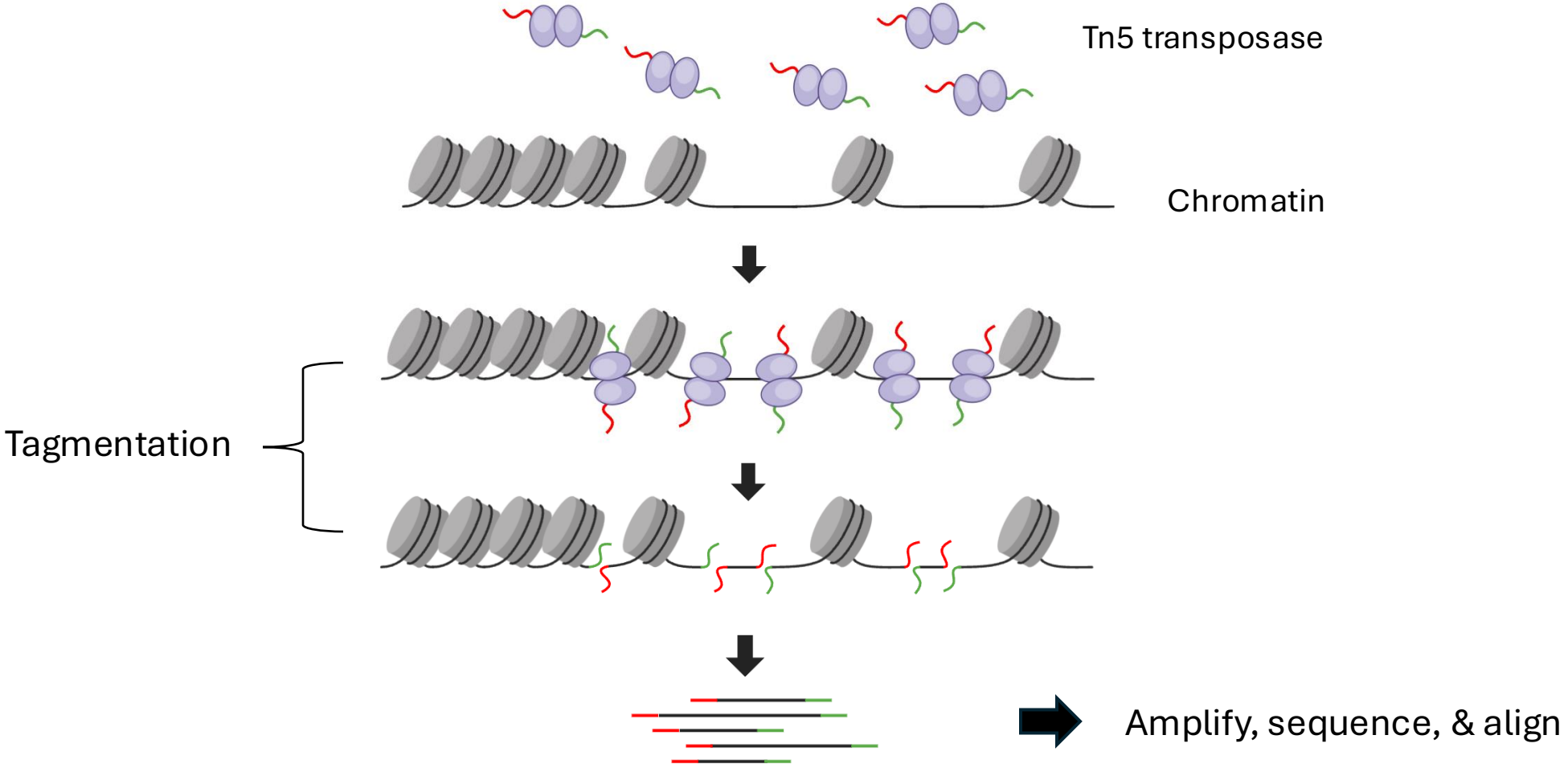


(Preissl et al., Nat Rev Genet, 2023)

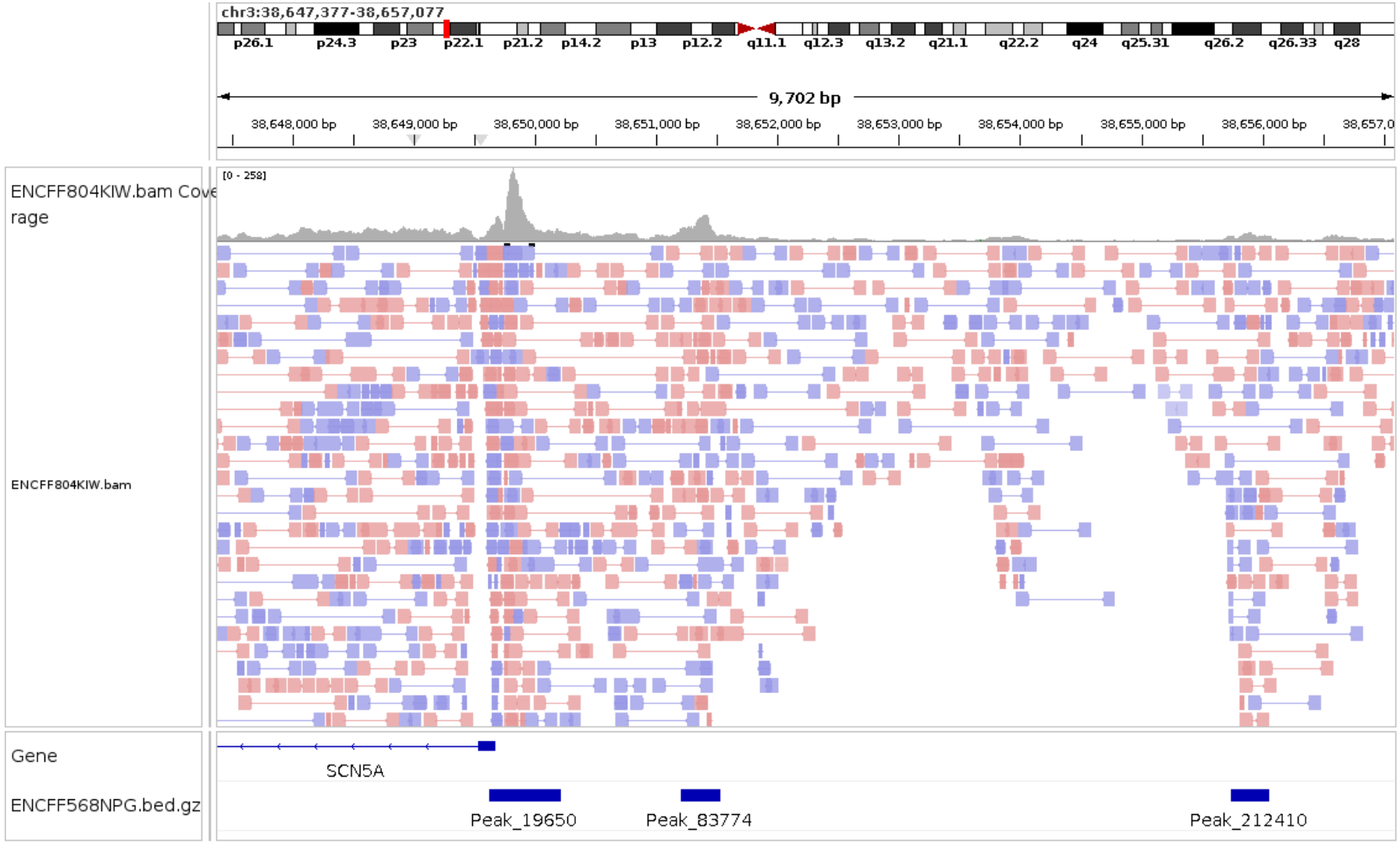
Open chromatin and ATAC-seq

ATAC-seq

Assay for Transposase-Accessible Chromatin (ATAC) with high-throughput sequencing
- Buenrostro et al., Nat Methods, 2013



ATAC-seq: alignment example

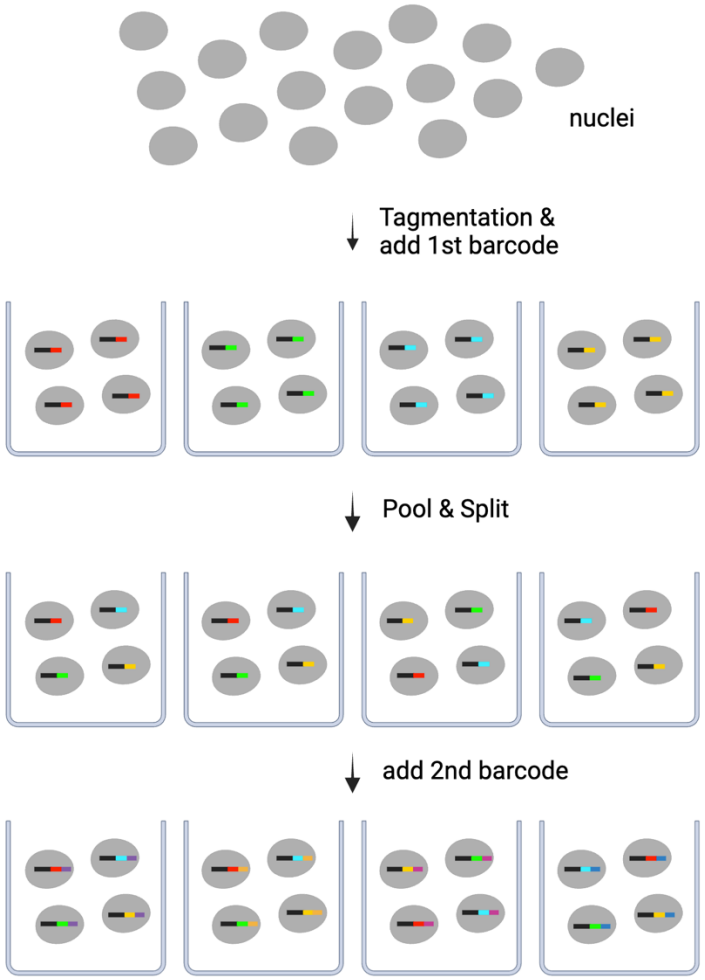


IGV Browser (v.2.16.0)

Single-cell (nucleus) ATAC-seq

- In single-cell ATAC-seq, a unique cell ”barcode” is assigned to every fragment
- Also **single-nucleus ATAC-seq, or snATAC-seq**
- Approaches
  - Combinatorial indexing-based techniques
  - Microfluidics-based approaches

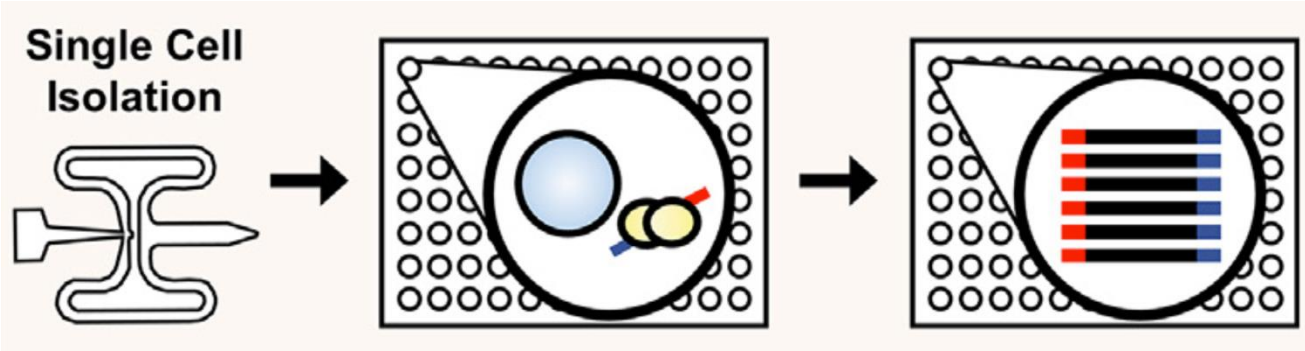
scATAC-seq technologies (1)



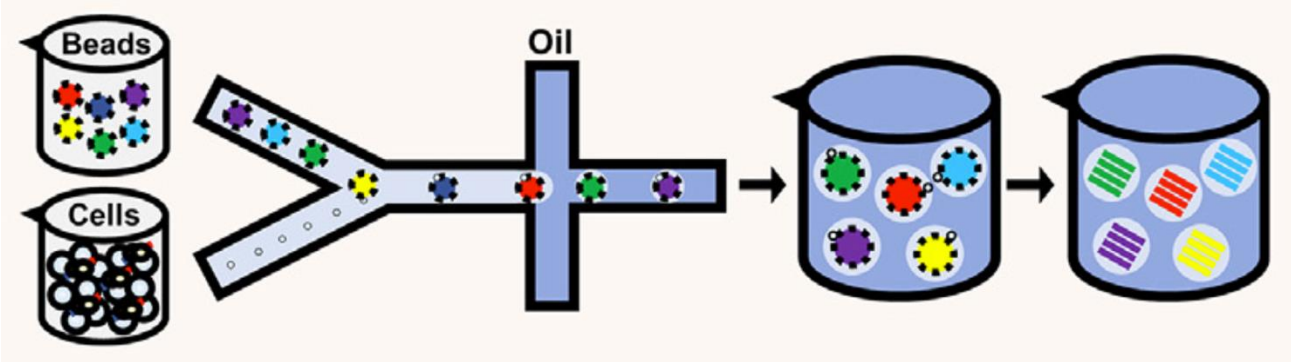
- Combinatorial indexing based approaches
  - Sci-ATAC-seq (Cusanovich et al., Science 2015)
  - Paired-seq (Zhu et al., Nat Struct Mol Biol 2019)
  - dsciATAC-seq (Lareau et al., Nat Biotech 2019)
  - Sci-ATAC-seq3 (Domcke et al., Science 2020)
  - …
- Highly scalable & Cost effective
  - Barcodes can be added multiple times
  - ~ 1 million cells can be analyzed in one experiment
- Potentially less sensitive than other methods

scATAC-seq technologies (2)

- Microfluidics-based approaches
  - Plate-based: Fluidigm's integrated fluidics circuit (IFC)



  - Droplet-based: 10X Genomics Chromium



(Baek & Lee, Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *CSBJ*, 2020)

ATAC-seq data analysis workflow (1): Raw reads → Cell x peak matrix

1. **Preprocessing:** Demultiplexing, adapter trimming, and read alignment to the reference genome.
2. **Quality Control (QC):** Assessing data quality, filtering low-quality reads, and nuclei, and removing doublets
3. **Peak Calling:** Identifying statistically significant open chromatin regions (peaks).
4. **Quantifying Accessibility:** Generating a cell-by-peak matrix.

ATAC-seq data analysis workflow (2): Cell x peak matrix → Clusters

1. **Dimensionality Reduction:** Transforming high-dimensional data into lower dimensions for visualization (e.g., UMAP, t-SNE).
2. **Clustering:** Grouping nuclei based on similar chromatin accessibility profiles to identify distinct cell populations.
3. **Cell Type Annotation:** Assigning biological identities to clusters using known markers or atlases.

ATAC-seq data analysis workflow (3): Downstream analysis & Biological Interpretation

1. **Differential Accessibility (DA) Analysis:** Identifying regions with significant accessibility changes between cell types or conditions.
2. **Motif Enrichment Analysis:** Discovering transcription factor binding motifs enriched in open chromatin regions.
3. **Gene Activity Scores / Peak-to-Gene Linking:** Estimating gene expression potential from chromatin accessibility and linking regulatory elements to target genes.

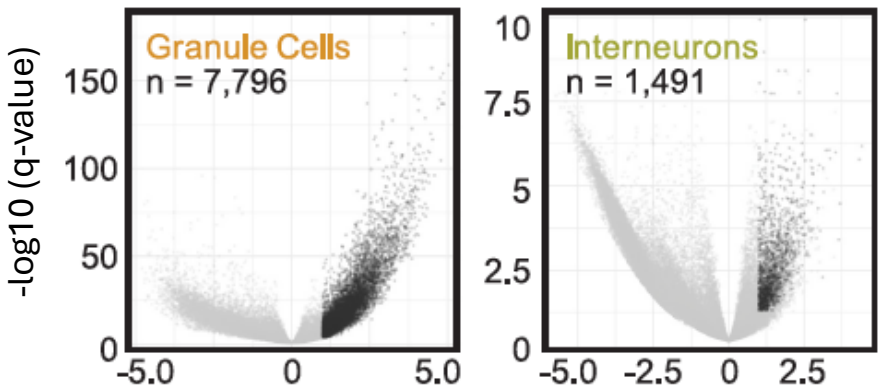
The entire analysis process is **iterative**, and **visualization** at every step is critical !

Resource

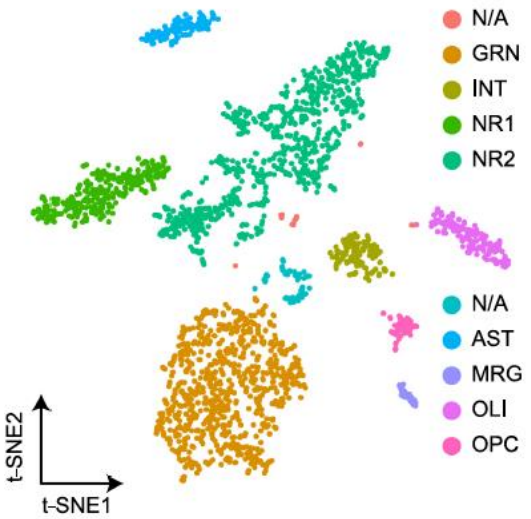
The accessible chromatin landscape of the murine hippocampus at single-cell resolution

John R. Sinnamon,[1,8] Kristof A. Torkenczy,[2,8] Michael W. Linhoff,[1] Sarah A. Vitak,[2] Ryan M. Mulqueen,[2] Hannah A. Pliner,[3] Cole Trapnell,[3] Frank J. Steemers,[4] Gail Mandel,[1] and Andrew C. Adey[2,5,6,7]

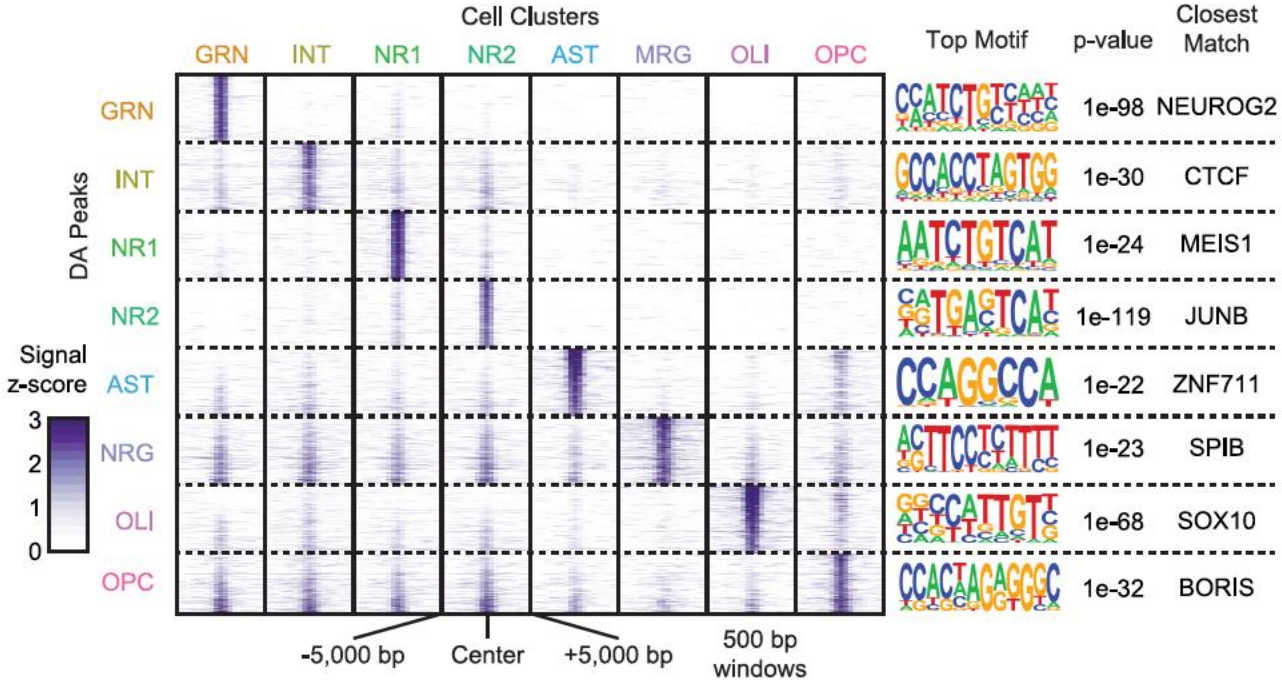
[1]The Vollum Institute, Oregon Health and Science University, Portland, Oregon 97239, USA; [2]Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon 97239, USA; [3]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA; [4]Illumina, Incorporated, San Diego, California 92122, USA; [5]Knight Cardiovascular Institute, Oregon Health and Science University, Portland, Oregon 97239, USA; [6]Knight Cancer Institute, Oregon Health and Science University, Portland, Oregon 97239, USA; [7]Knight Center for Early Detection Advanced Research, Oregon Health and Science University, Portland, Oregon 97239, USA



sci-ATAC-seq of 2,346 cells from mouse hippocampus



GRN: granule cells
INT: interneurons
NR1: pyramidal neurons (1)
NR2: pyramidal neurons (2)
AST: astrocytes
MRG: microglia
OLI: oligodendrocytes
OPC: OLI progenitor cells



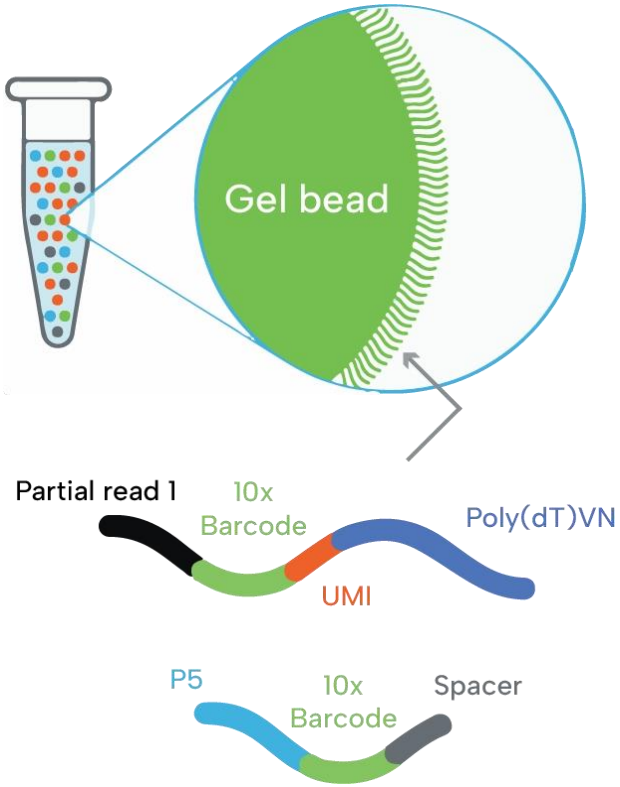
(Sinnamon et al., Genome Res, 2019)

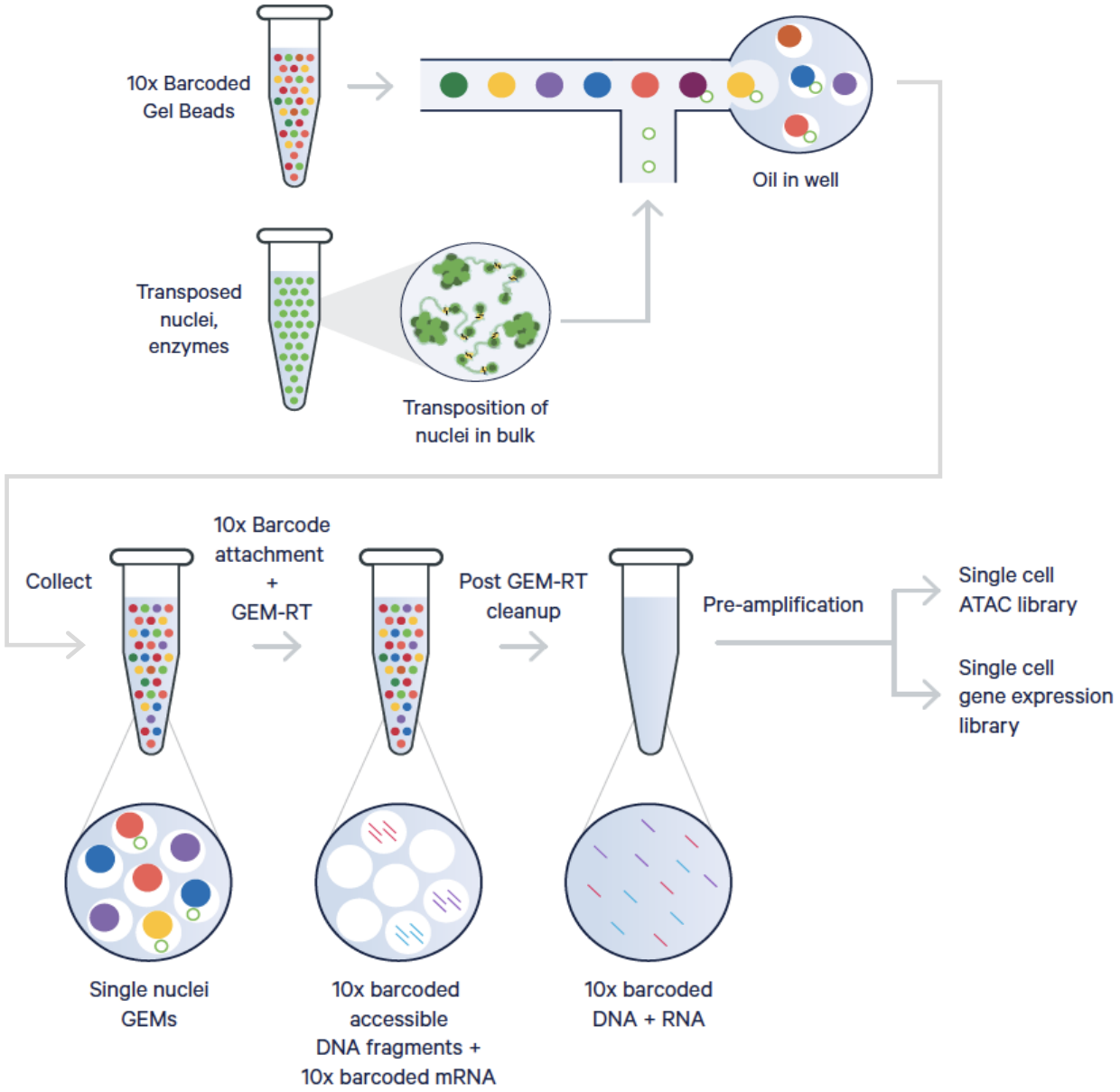
Single-Nucleus Multiome (RNA + ATAC)

Why Multiome? The Power of Integration:

- Simultaneous measurement of chromatin accessibility (ATAC) and gene expression (RNA) from the *same* single nucleus.
- Address limitations of inferring relationships from separate datasets.
- Directly connect regulatory elements to gene activity in specific cell types.
- Gain a more comprehensive understanding of cellular states and regulatory mechanisms.

10X Multiome Experimental Overview



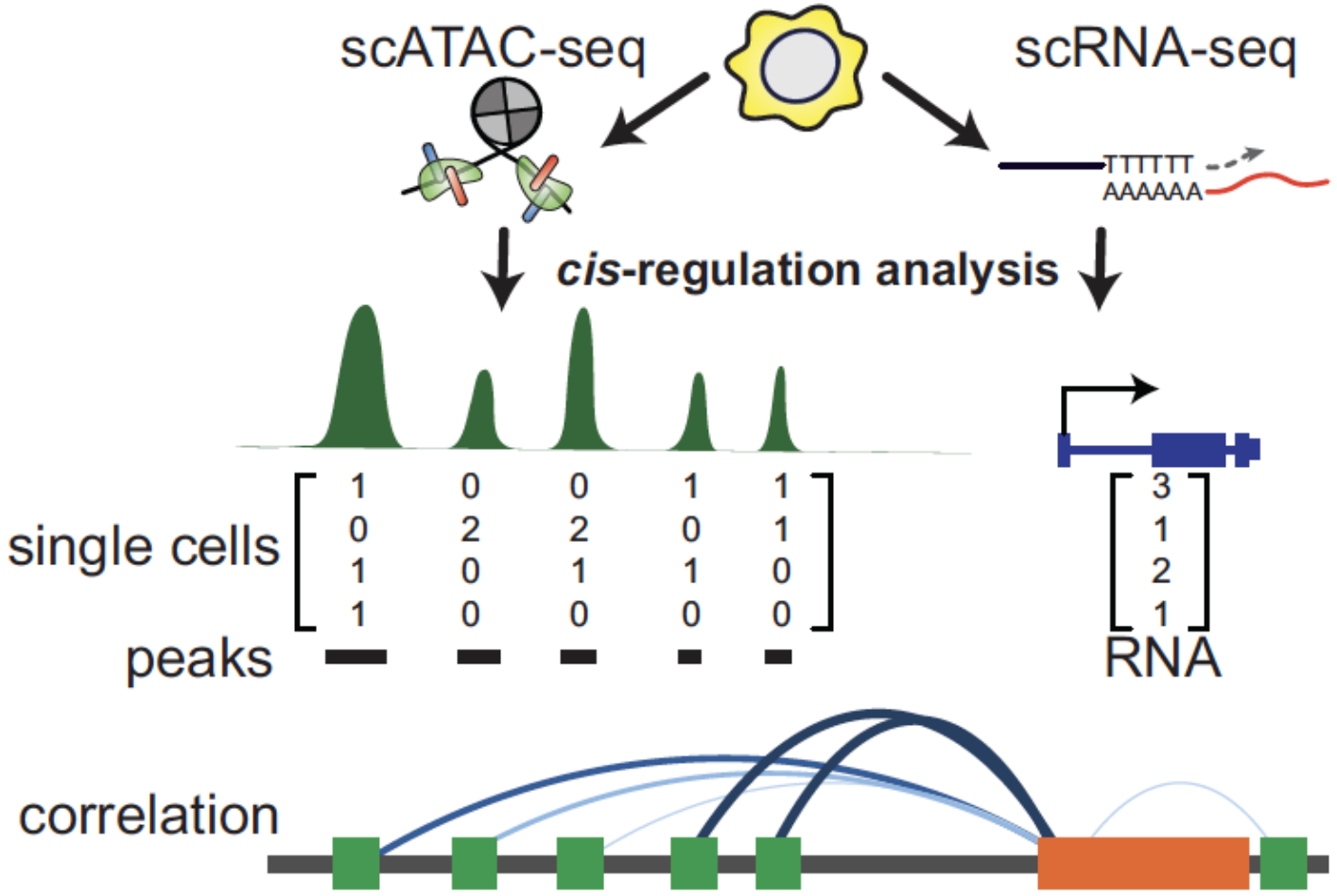


(Source: 10X Genomics)

Key Concepts in Multiome Data Analysis

- **Independent Preprocessing:** Initial processing of ATAC and RNA modalities separately.
- **Joint Dimensionality Reduction & Clustering:** Integrating both modalities to identify more robust cell types/states (e.g., Weighted Nearest Neighbor (WNN) analysis).
- **Enhanced Cell Annotation:** Using both RNA and ATAC features for more precise cell type identification.
- **Direct Regulatory Linkages:**
  - Correlating gene expression with accessibility of promoters and enhancers.
  - Building gene regulatory networks.

Linking CREs to their target genes



(Ma et al., Cell 2020)

Key Considerations & Future Directions

- **Computational Demands:** High-throughput data requires robust computational infrastructure.
- **Data Sparsity:** Challenges in analyzing sparse single-cell epigenomic data.
- **Tool Selection:** Overview of popular software packages and frameworks (e.g., Seurat/Signac, ArchR, CellRanger Multiome).
- **Biological Interpretation:** The ultimate goal is to derive meaningful biological insights!

scATAC-seq analysis tools

Tool	Platform	Feature Matrix	Preprocessing	Clustering	DAR	Motif/k-mer	Gene activity	Co-accessibility	Trajectory	Pathway
ChromVAR	R	TF motifs, k-mer	O	O	X	O	X	X	X	X
SCRAT	R/Web	Selectable feature	O	O	O	X	X	X	X	X
scABC	R	Peak	O	O	X	O (ChromVAR)	X	X	X	X
Cicero	R	TSS	O	O	O	X	O	O	O	X
Scasat	Python/R	Peak	O	O	O	X	X	X	X	O (GREAT)
cisTopic	R	Peak	O	O	X	X	O	X	X	O
snapATAC	Python/R	Bin, peak	O	O	O	O (ChromVAR, Homer)	O	X	X	O (GREAT)
epiScanpy	Python	Peak	O	O	X	X	X	X	X	X
Destin	R	Peak	O	O	O	X	X	X	X	X
SCALE	Python	Peak	O	O	O	O (ChromVAR)	X	X	X	X
scATAC-pro	Python/R	Peak	O	O	O	O (ChromVAR)	O	O (Cicero)	X	O (GREAT)
Signac	R	Peak	O	O	O	O (ChromVAR)	O	X	X	X
ArchR	R	Bin, peak	O	O	O	O (ChromVAR), TF footprinting	O	O	O	X

(Baek & Lee, Single-cell ATAC sequencing analysis: From data preprocessing to hypothesis generation. *CSBJ*, 2020)

Questions?