ATAC-seq analysis

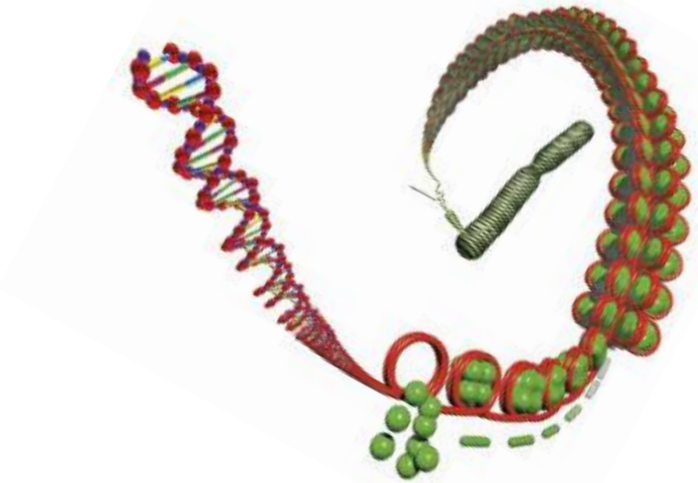
Bareket Dassa, Bioinformatics Unit

Introduction to Deep Sequencing Analysis course

2019-2020

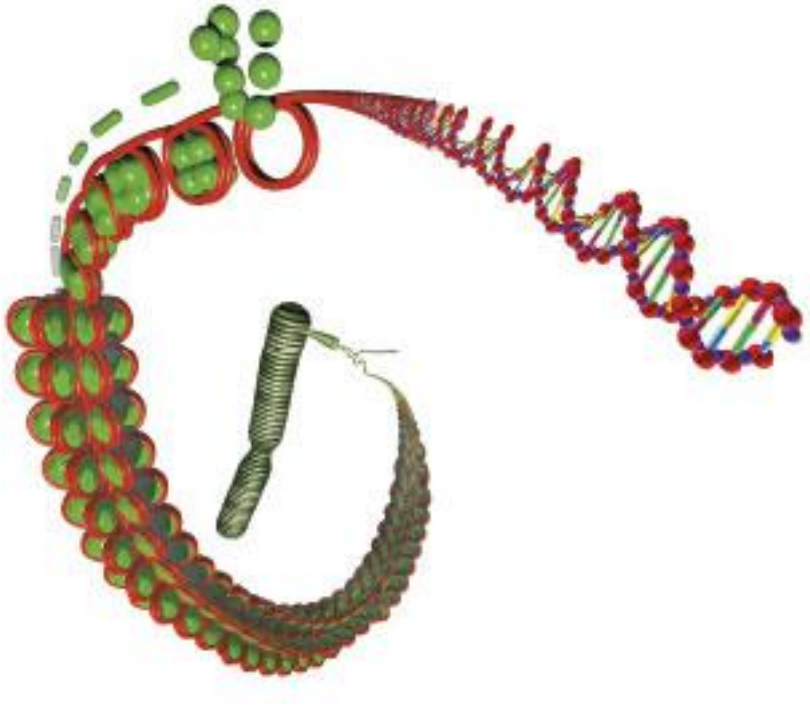
ATAC-seq lecture outline

- Why studying epigenomics?
- What is ATAC-seq?
- How is an ATAC-seq experiment designed?
- How is ATAC-seq analyzed bionformatically?
- Which are the available applications of ATAC-seq



What is epigenomic regulation?

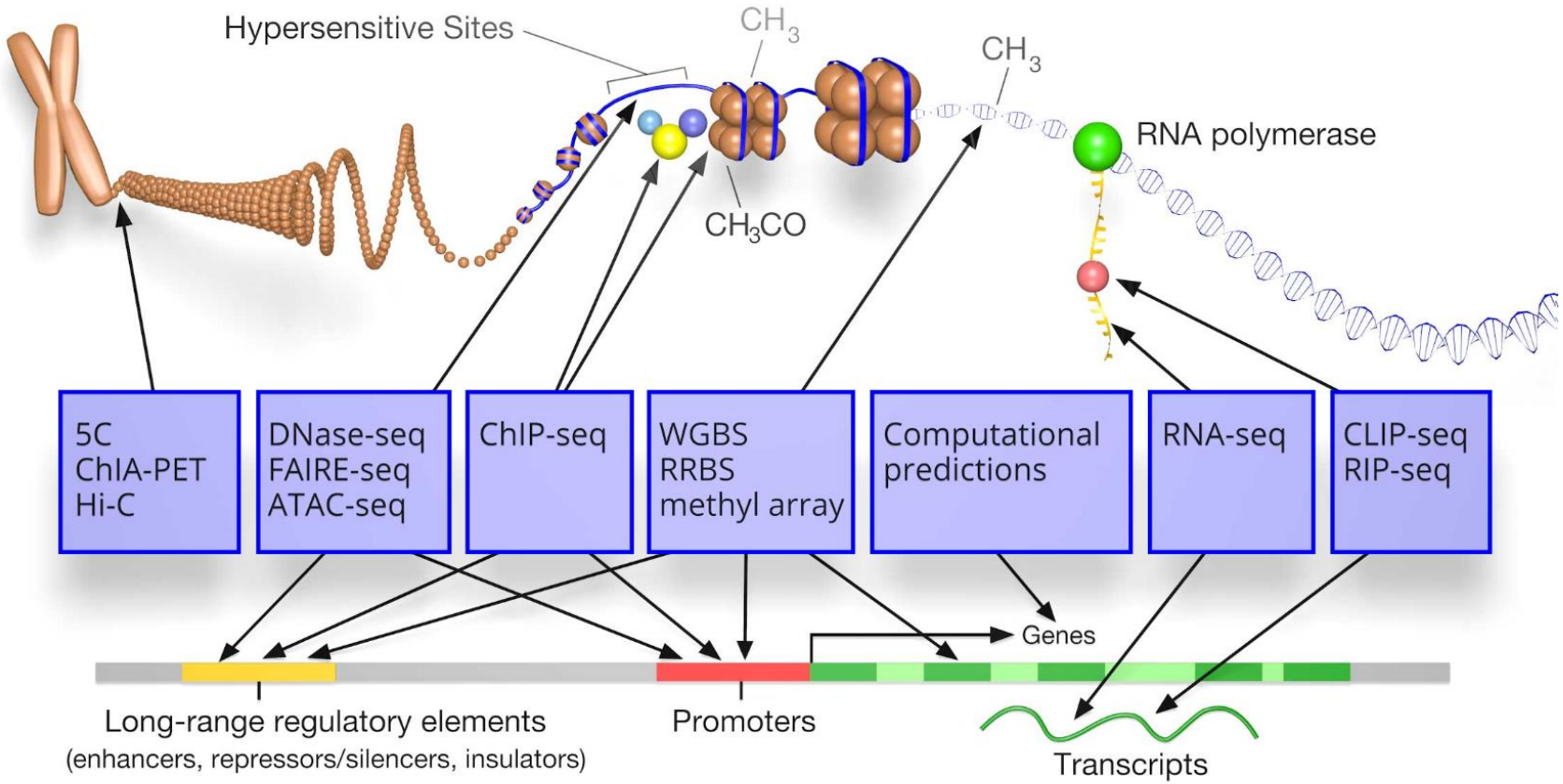
- Chromatin remodeling is highly dynamic
- Epigenetics involves genetic control by factors other than the DNA sequence
- Epigenetic **regulation** can switch genes on /off, and determine which and when genes are transcribed



Epigenetics in animation:
https://www.youtube.com/watch?v=JMT6oRYgkTk
Min 0:36, 1:50

Why study epigenomic regulation?

- Combining NGS assays with specialized biochemical protocols, to profiles genome-wide epigenetic modifications

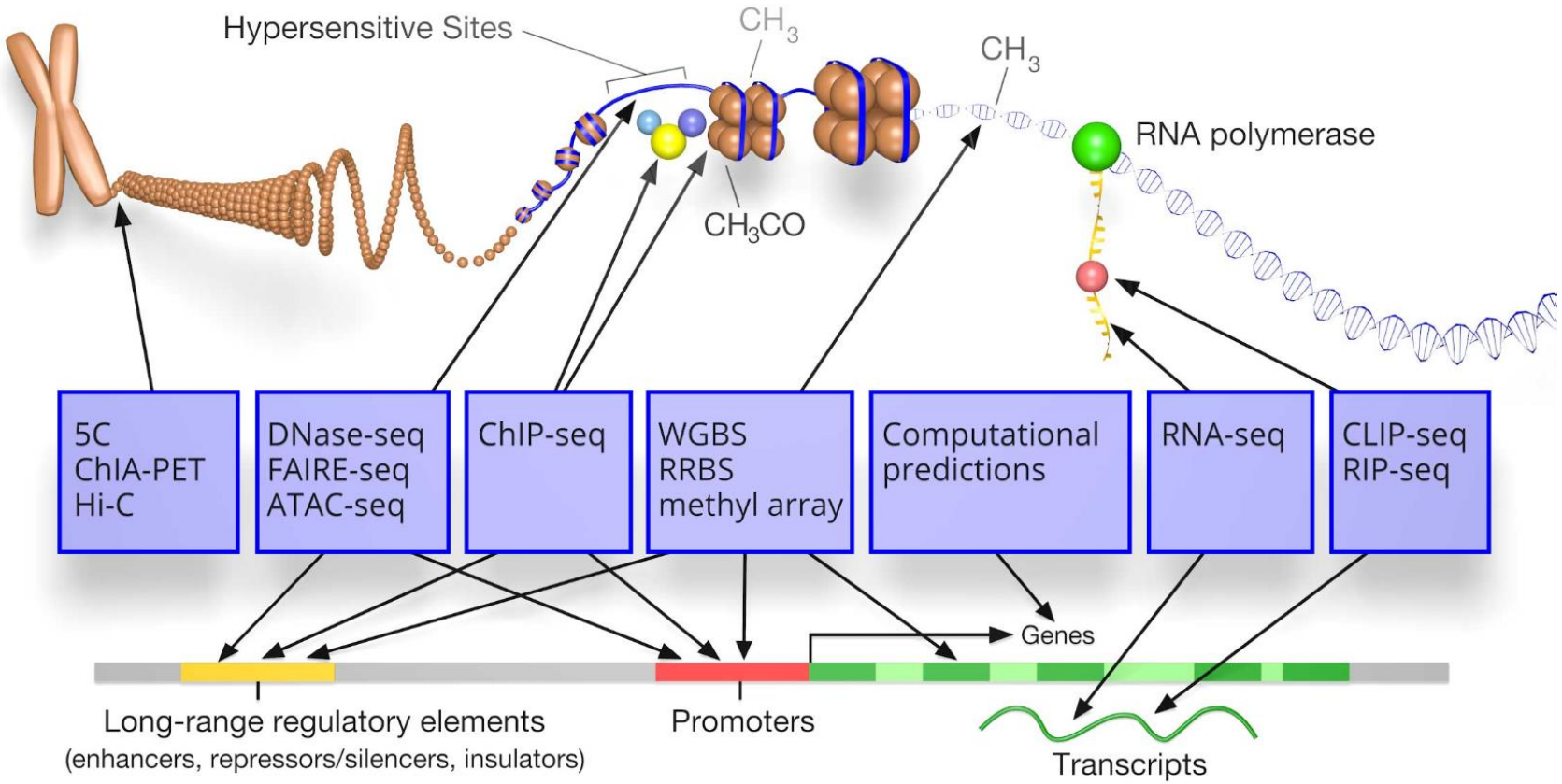


ENCODE

Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

Why study epigenomic regulation?

binding of proteins to specific regions of the genome
target sequences of transcription factors
histones positions and specific **modification**

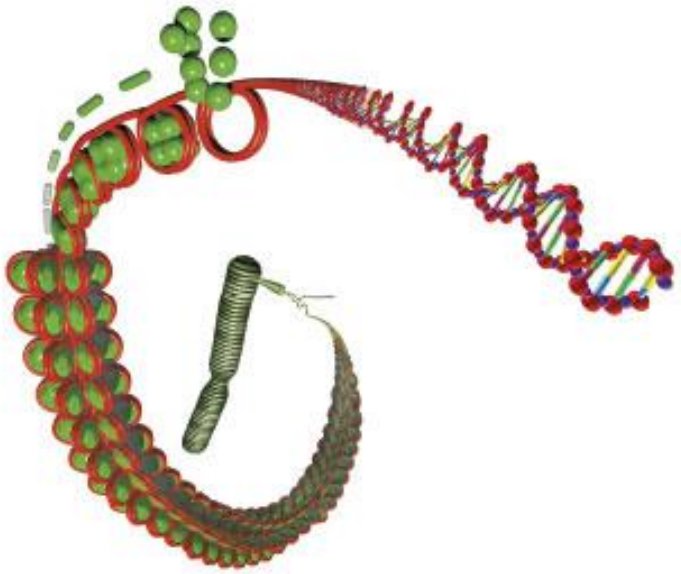


ENCODE

Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

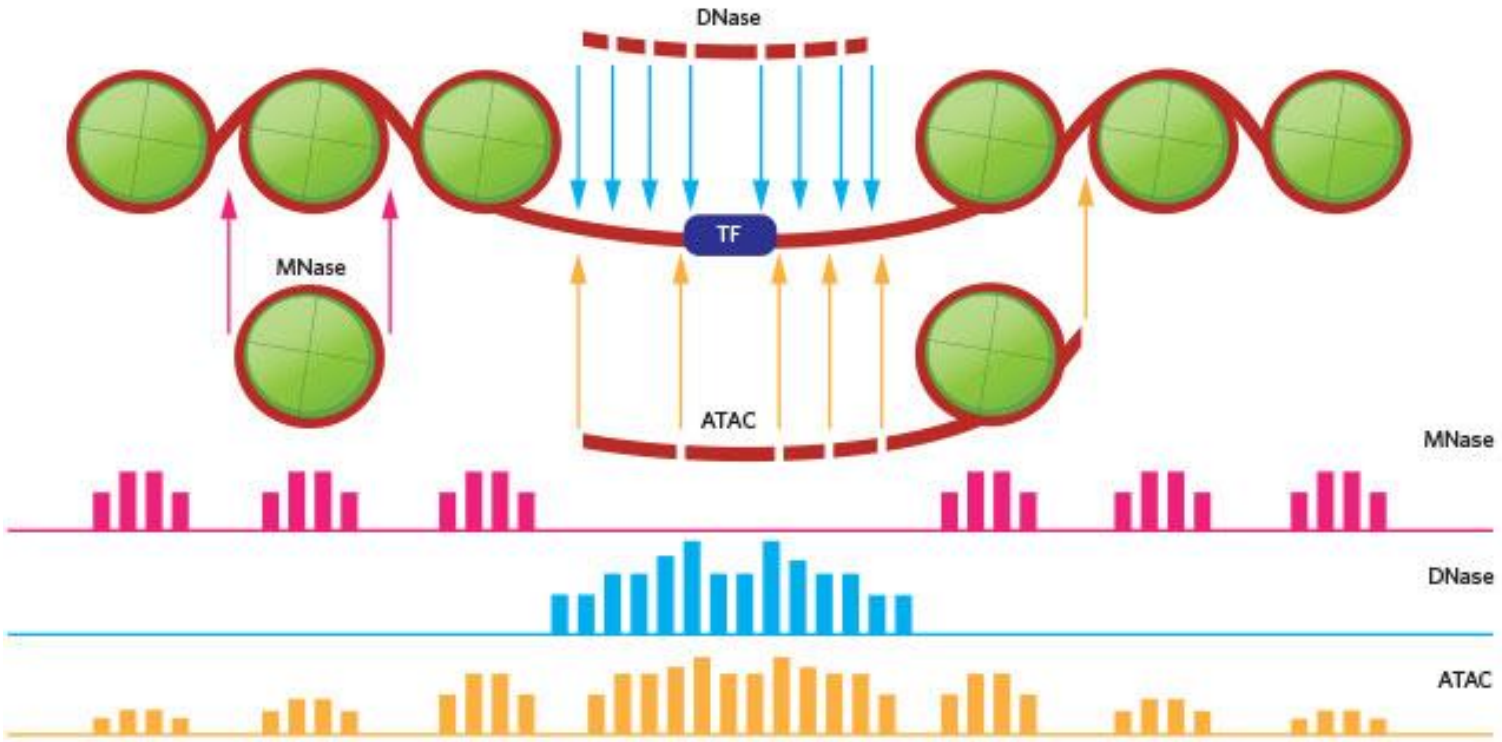
ATAC-seq lecture outline

- Why studying epigenomics?
- What is ATAC-seq?
- How is an ATAC-seq experiment designed?
- How is ATAC-seq analyzed bionformatically?
- Which are the available applications of ATAC-seq



ATAC-seq =
Assay for transposase-accessible chromatin using sequencing

- ATAC-seq captures open and accessible regions of chromatin ("openness")
- Provides genome-wide information on chromatin compaction



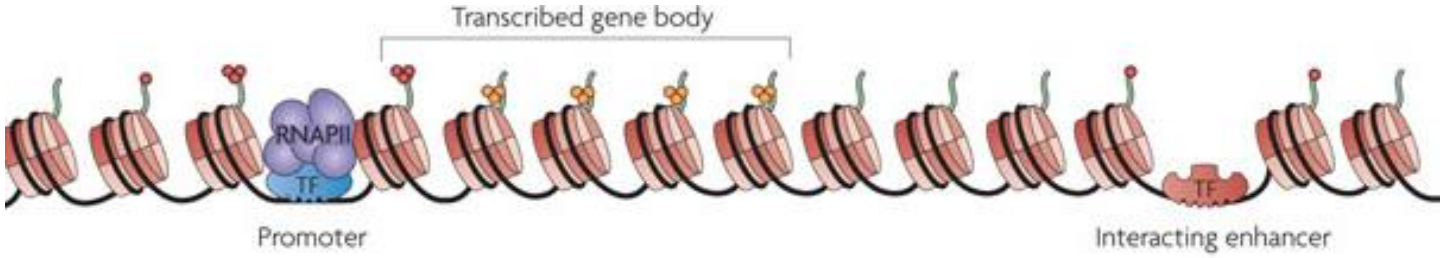
From: https://www.the-scientist.com/lab-tools/reveling-in-the-revealed-34261

Why using ATAC-seq?

ATAC-seq provides genome-wide information on **open chromatin regions** at nucleotide resolution

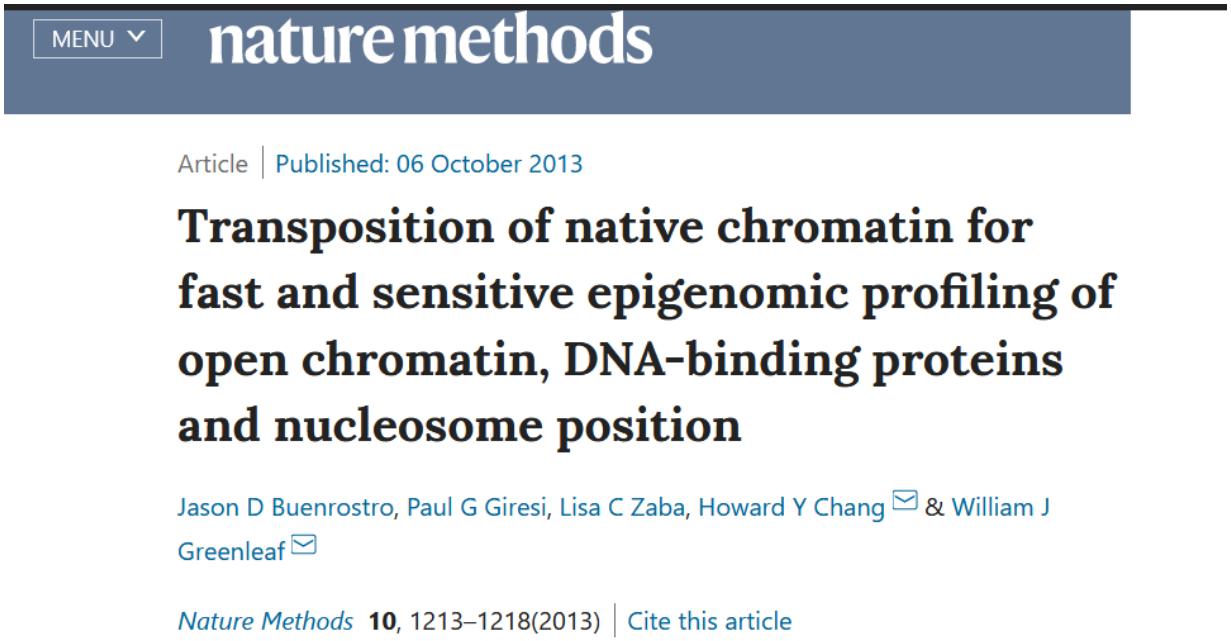
What information?

- **Profile regulatory elements** (promoters, enhancers), which are accessible to transcription machinery
- **Nucleosome positioning** and chromatin compaction
- Characterize genome wide **DNA-protein interactions** (TF, RNA polymerase)



Nature Reviews | Genetics

ATAC-seq was first described at the Greenleaf lab:



MENU nature methods

Article | Published: 06 October 2013

Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position

Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang & William J Greenleaf

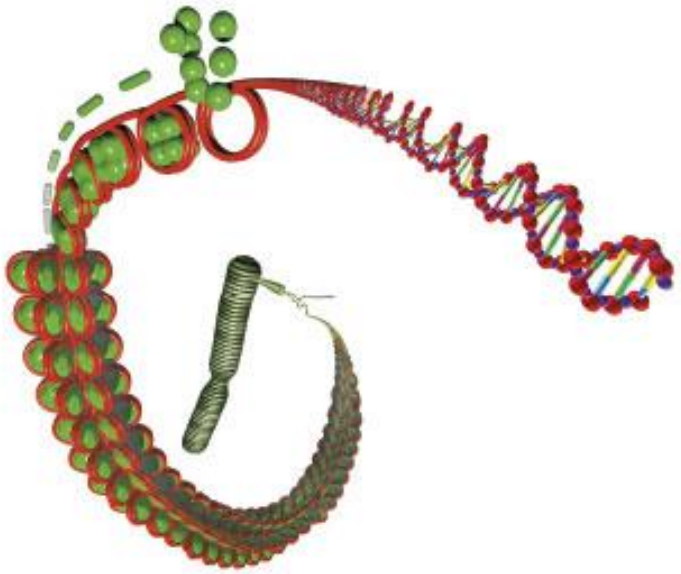
Nature Methods **10**, 1213–1218(2013) | Cite this article

- **Rapid** assay preparation time
- Protocol requires a **small input** (500-50,000 cells)
- **Quantifies** differences in cellular response to treatment or disease

Greenleaf lab

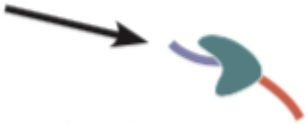
ATAC-seq lecture outline

- Why studying epigenomics?
- What is ATAC-seq?
- How is an ATAC-seq experiment designed?
- How is ATAC-seq analyzed bionformatically?
- Which are the available applications of ATAC-seq



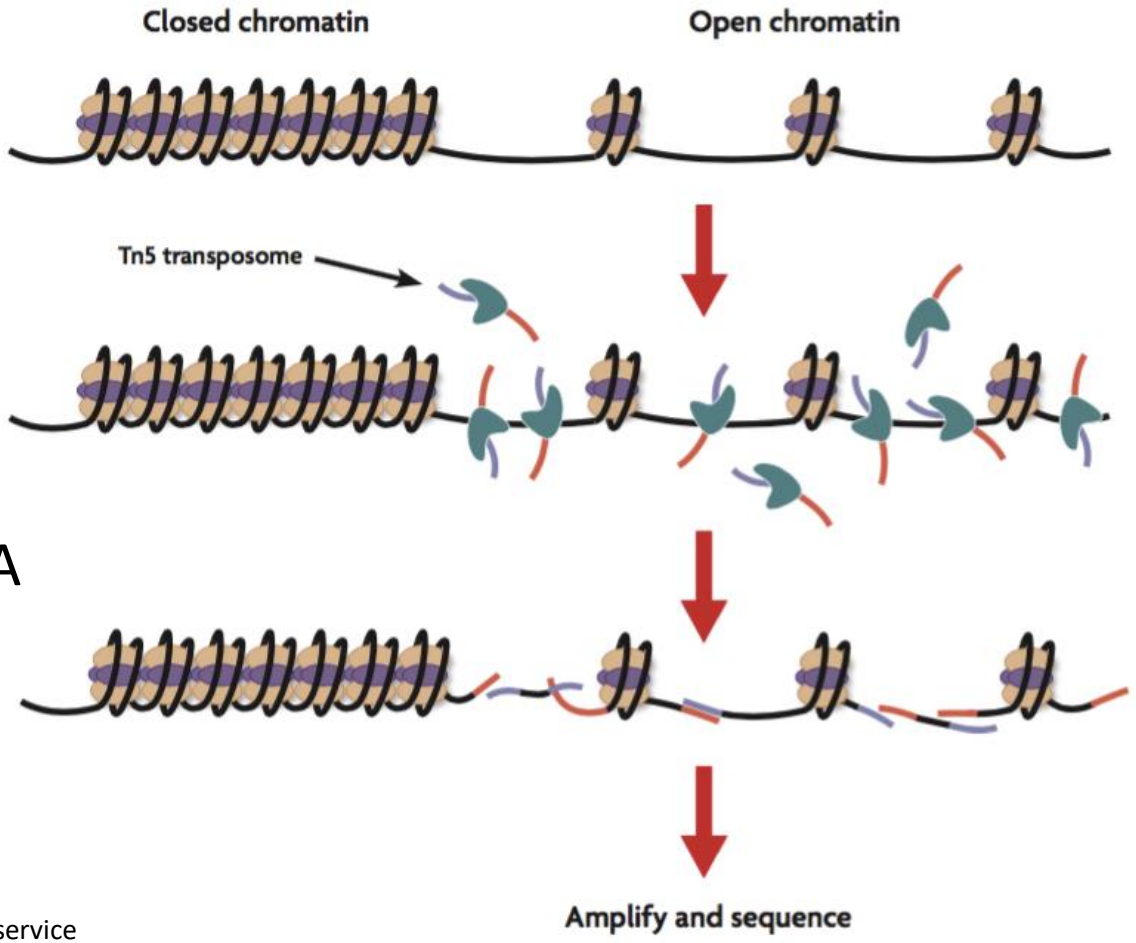
ATAC-Seq assay

- An engineered **transposase** is loaded with sequencing adapters (red and blue)



- **Tn5 tagmentation** simultaneously fragments DNA only in regions of open chromatin

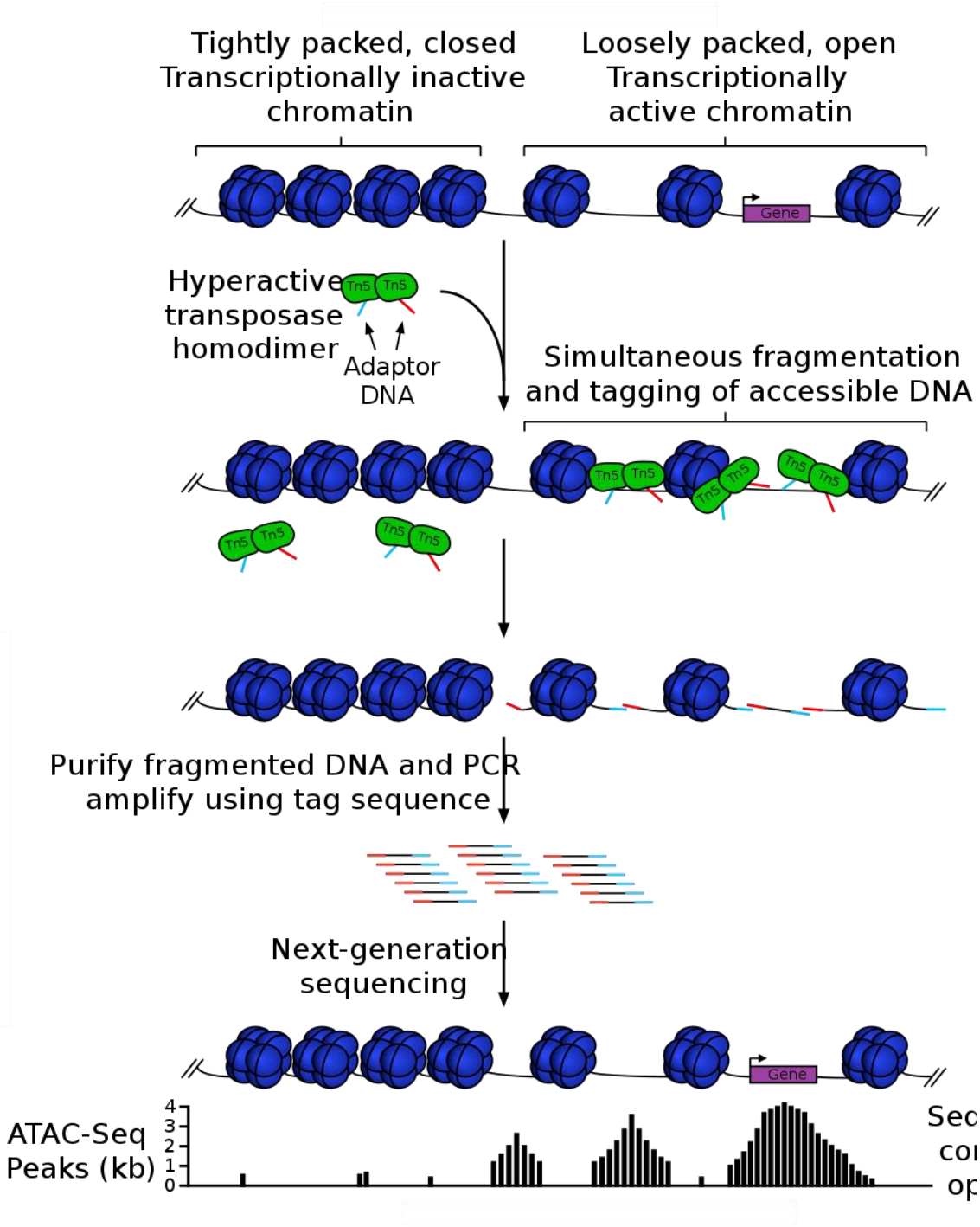
- And tags the resulting DNA with sequencing adapters to generate a library for PCR amplification



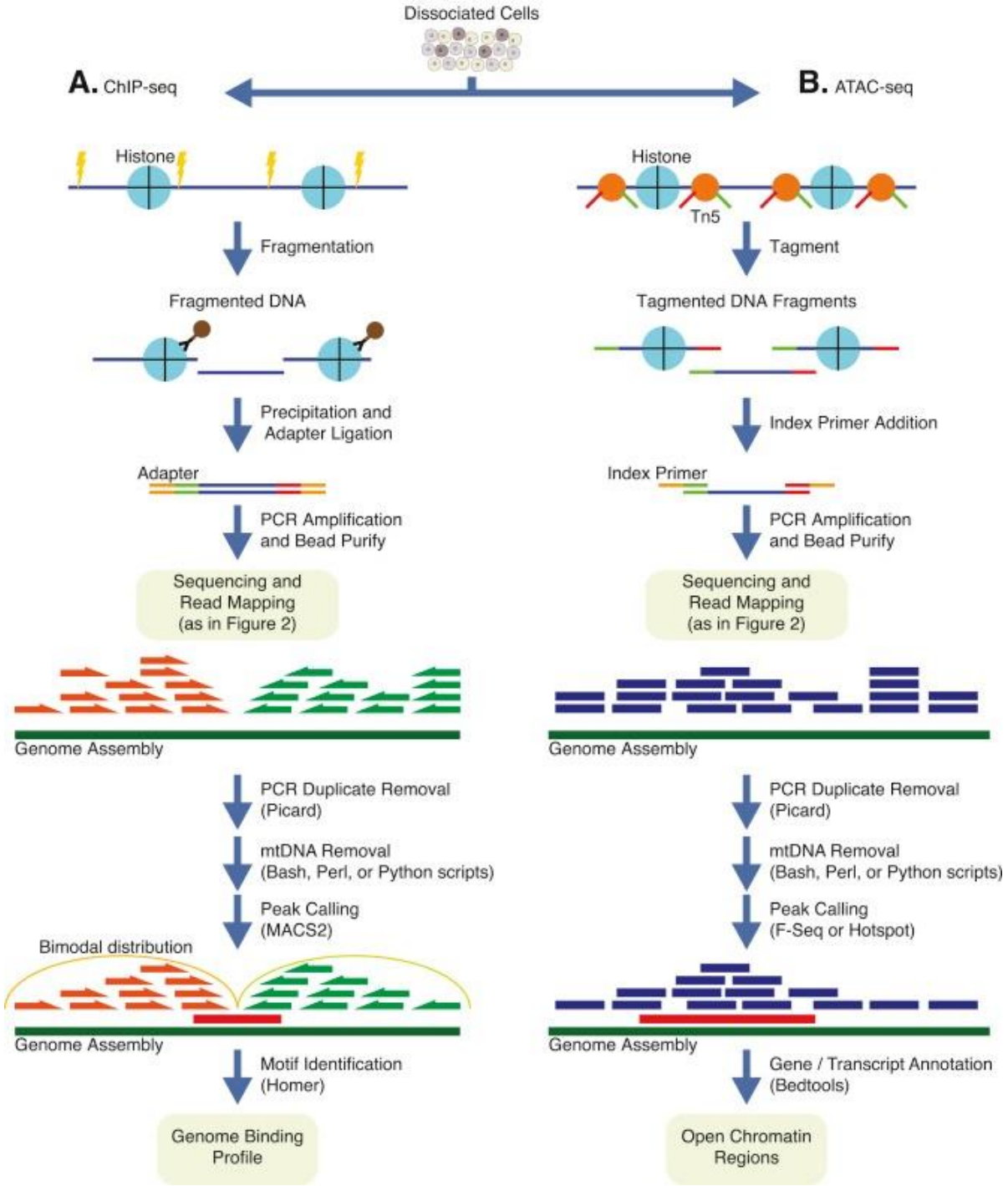
From: https://www.activemotif.com/catalog/1233/atac-seq-service

ATAC-Seq assay steps:

- Lyse cells
- Transposase reaction
- Purification of tagments
- Library amplification
- Illumina sequencing
- Bioinformatic analysis



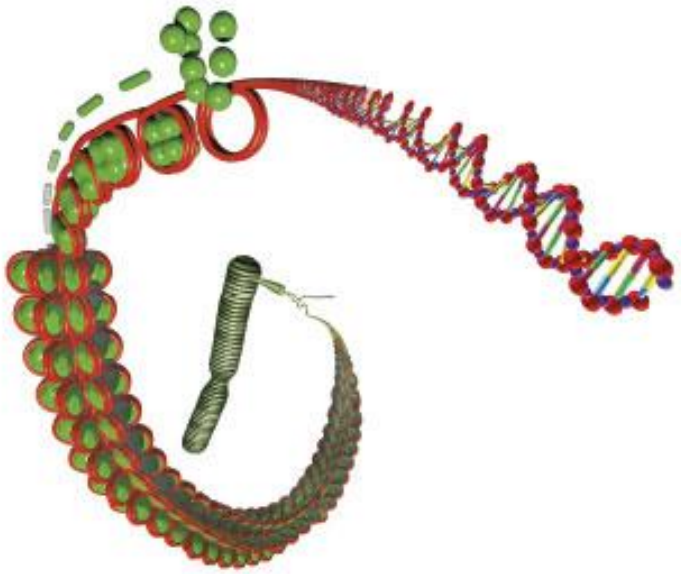
ATAC-seq compared with ChIP-seq



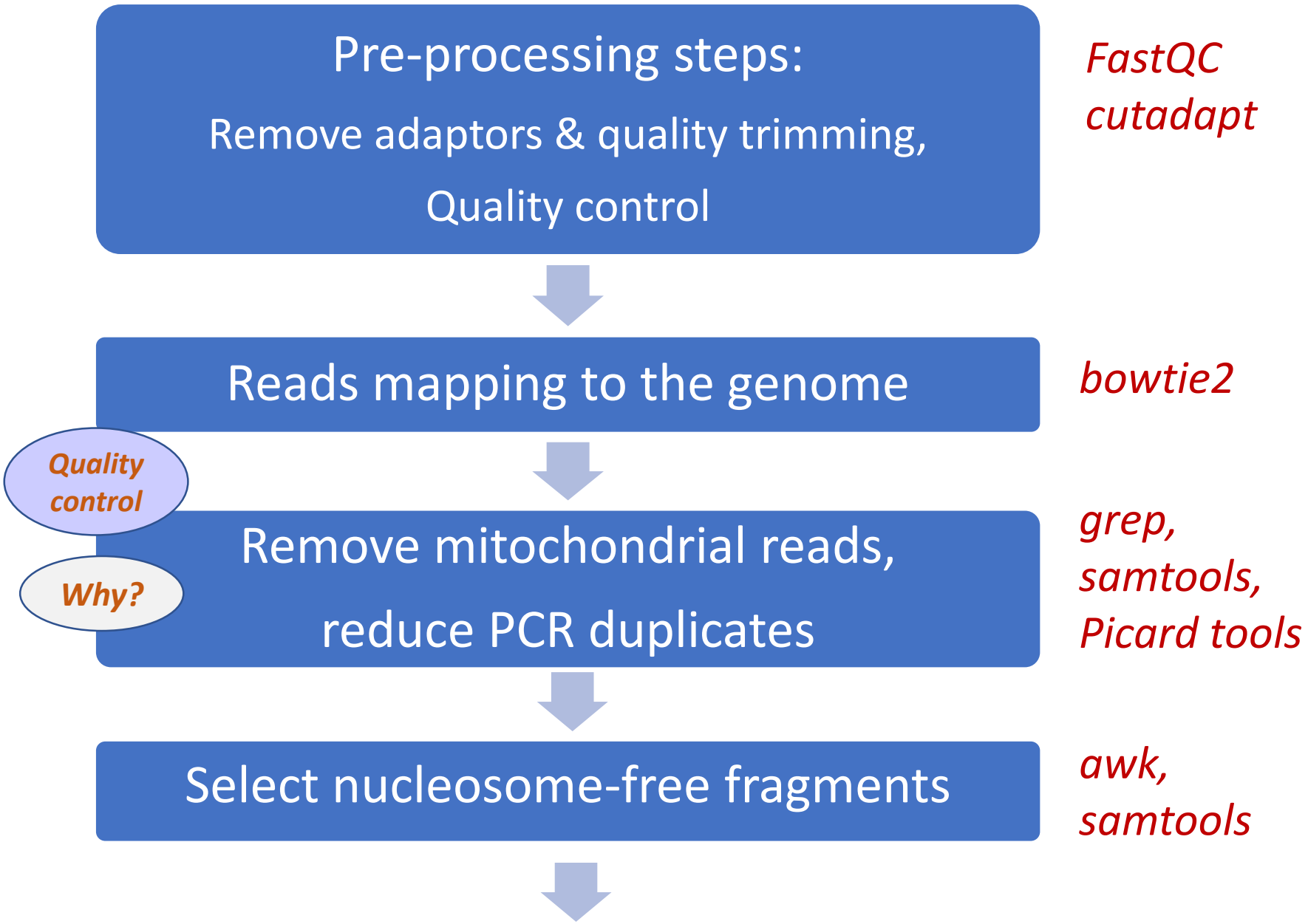
https://www.sciencedirect.com/science/article/pii/S1350946216300301#fig6

ATAC-seq lecture outline

- Why studying epigenomics?
- What is ATAC-seq?
- How is an ATAC-seq experiment designed?
- How is ATAC-seq analyzed bionformatically?
- Which are the available applications of ATAC-seq



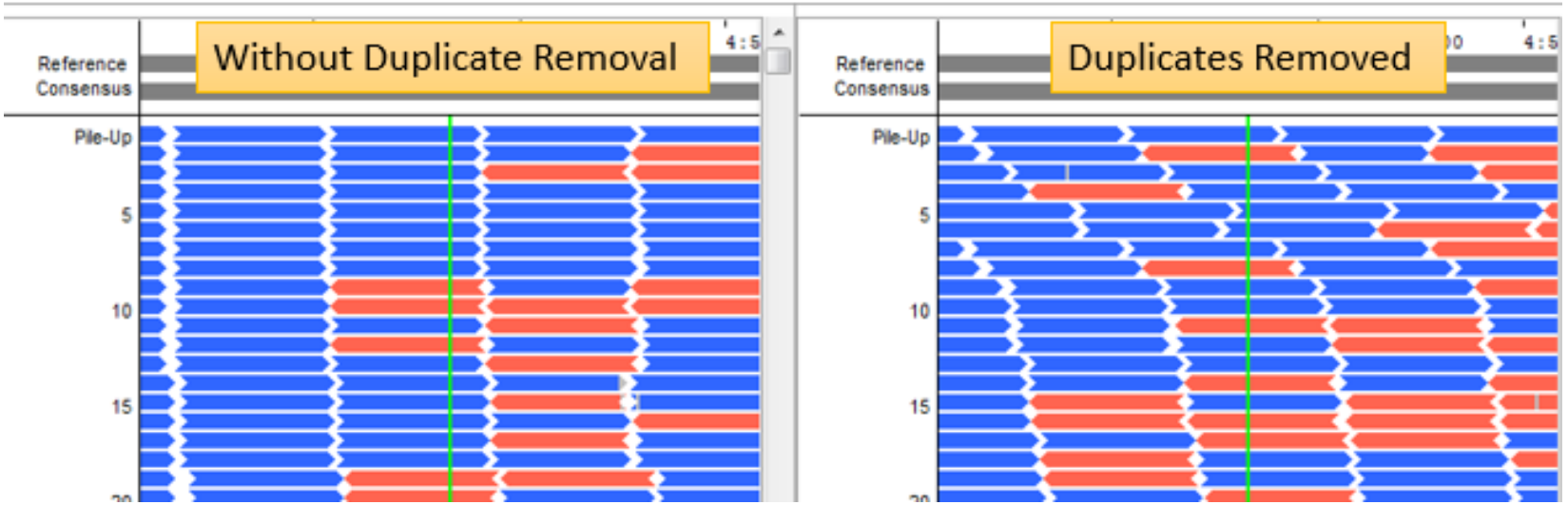
ATAC-seq analysis workflow



Why should we reduce reads, which map on mitochondrial DNA?

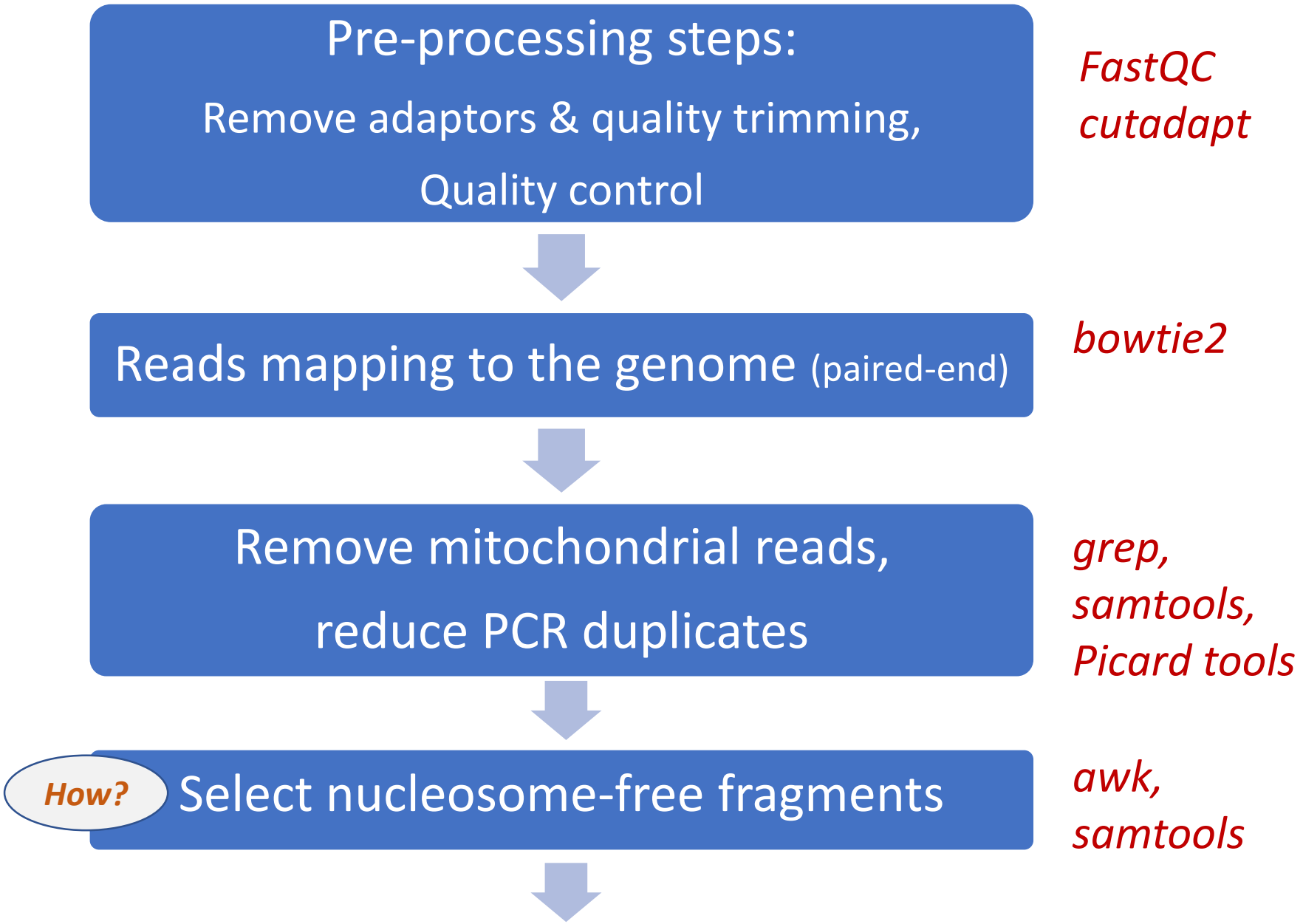
Mitochondrial DNA, unlike the nuclear genome, is not compacted in nucleosomes.
ATAC-seq samples may contain ~20–80% of mitochondrial sequencing reads, depending on the cell type

Reduce PCR duplicates



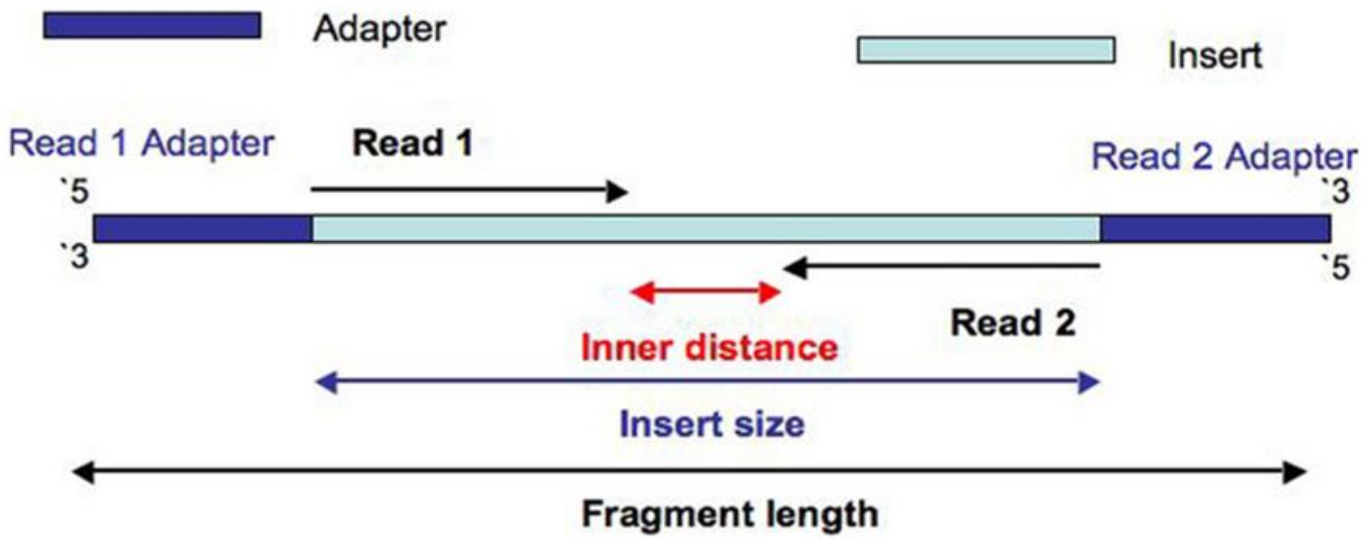
From: softgenetics.com

ATAC-seq analysis workflow



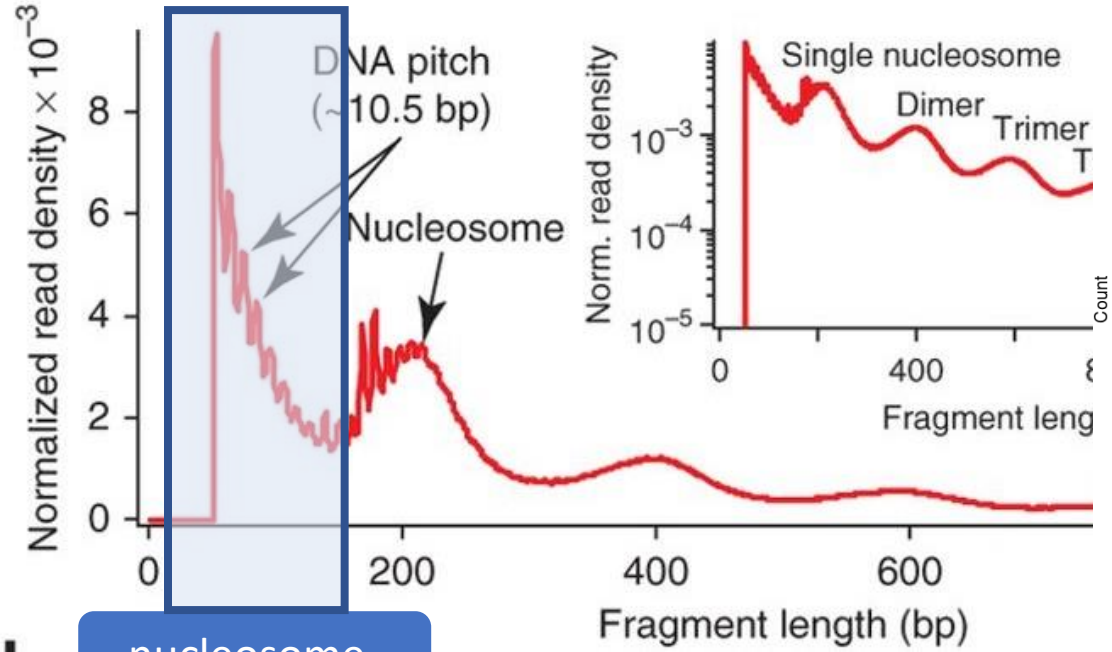
Selecting nucleosome-free fragments

- Insert size = distance between the R1 and R2 read pairs
- We wish to select reads that are shorter than the length generally protected by a nucleosome



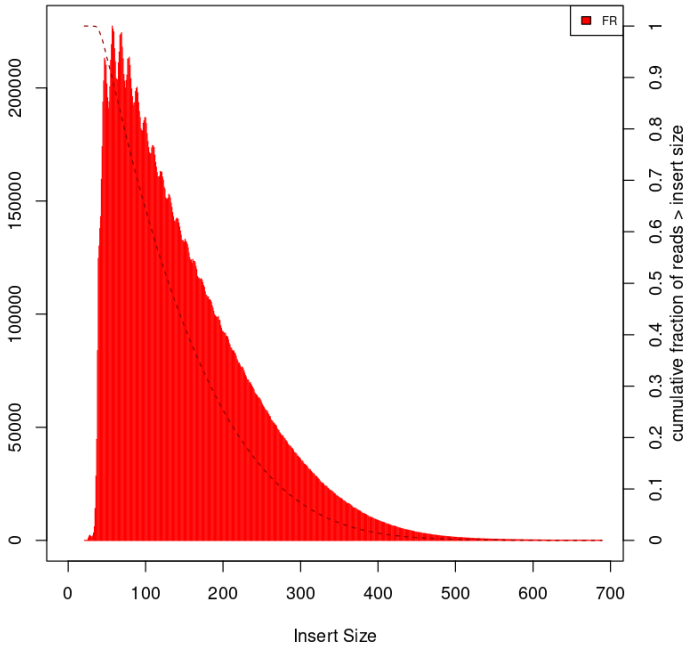
Insert size distribution

- Distribution of sequenced fragments should show a **periodicity** of ~200bp
- Select reads that are **shorter** than the length generally protected by a nucleosome
- Insert size distribution is a good indication on the quality of your experiment



nucleosome-free fragments

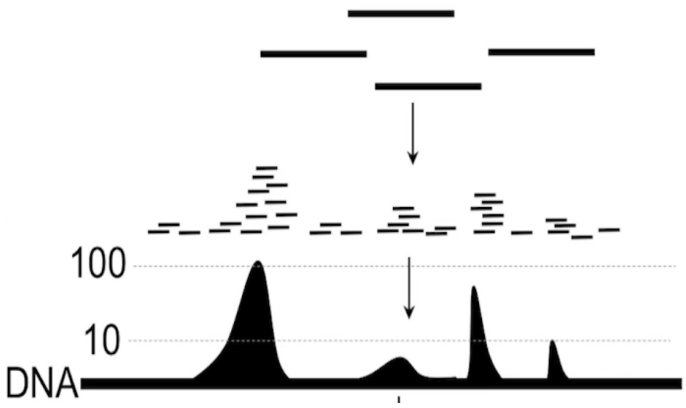
The profile of "naked DNA" without nucleosomes



From: Galaxy Training

Buenrostro (2013) Nature Methods

ATAC-seq analysis workflow (2)



Peak calling — *macs2*

Peaks visualization — *IGV*

Peak annotation, Motif discovery, Functional enrichment — *Homer, GREAT, CEAS*

Known ATAC-seq bias

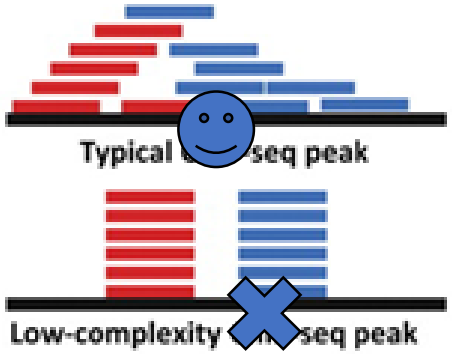
- **Controlling for the enzymatic cleavage bias** with “naked DNA” control

The Tn5 transposase is known to cleave DNA in a sequence-dependent manner, because of its tendency to cleave some DNA sequences more efficiently than others.

Chung, H.-R. *et al. The effect of micrococcal nuclease digestion on nucleosome positioning data*. *PLoS ONE* (2010).

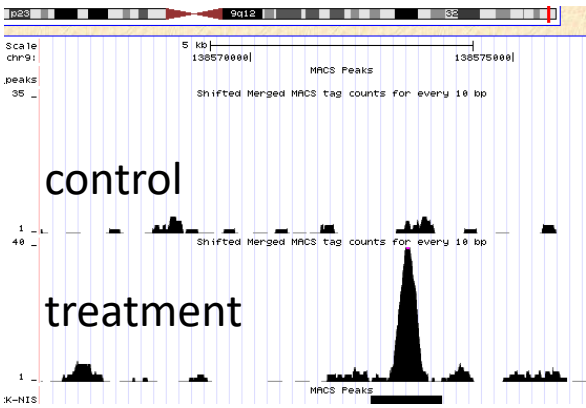
- **Avoiding high read redundancy**:

Filter out duplicate reads to avoid calling false peaks (i.e. reads at the exact same genome location and the same strand if their number exceeds the expected redundancy).



- **Adjusting for sequencing depth:**

High levels of coverage are needed for an informative experiment



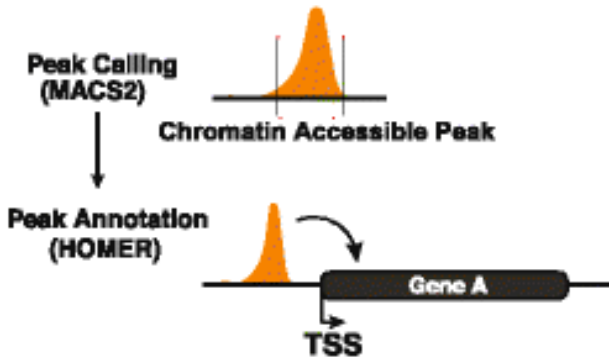
ENCODE consortium’s Standards, Guidelines and Best Practices:
https://www.encodeproject.org/atac-seq/

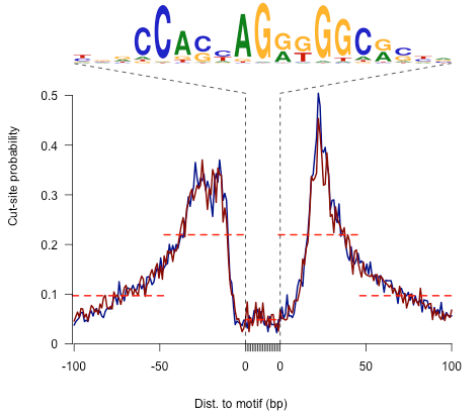
ATAC-seq analysis workflow (2)

Peak calling — *macs2*

Peaks visualization — *IGV*

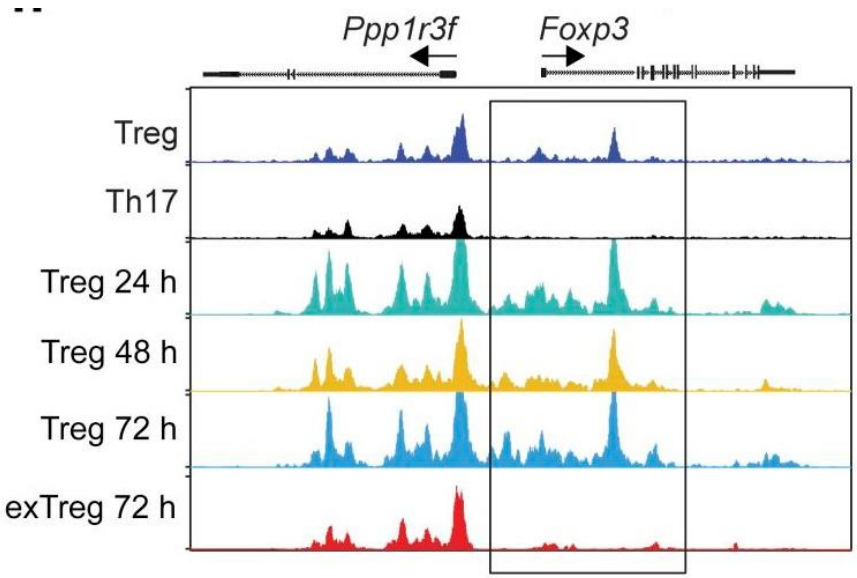
Peak annotation, Motif discovery, Functional enrichment — *Homer, GREAT, CEAS*



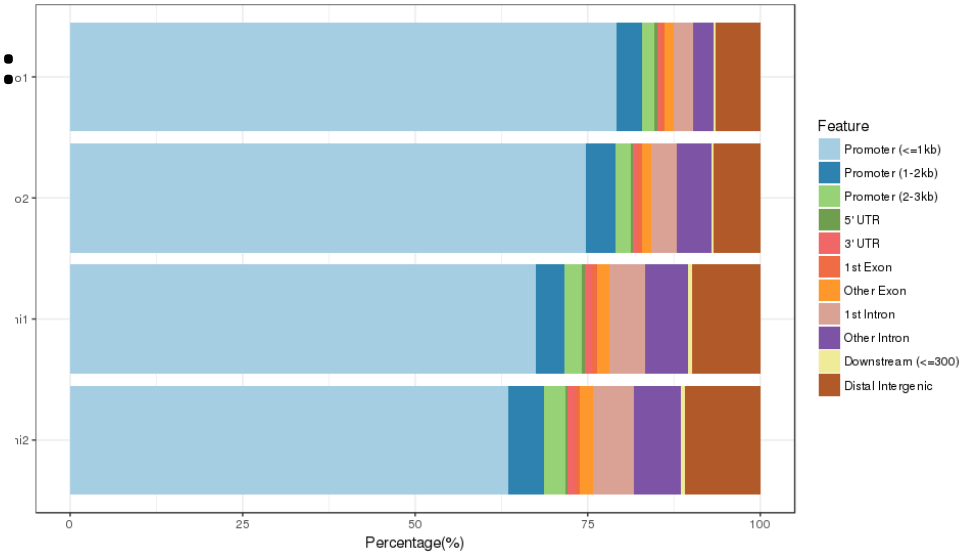


Downstream analysis of ATAC-seq

A. Peak calling and visualization on a genome browser (IGV, UCSC):



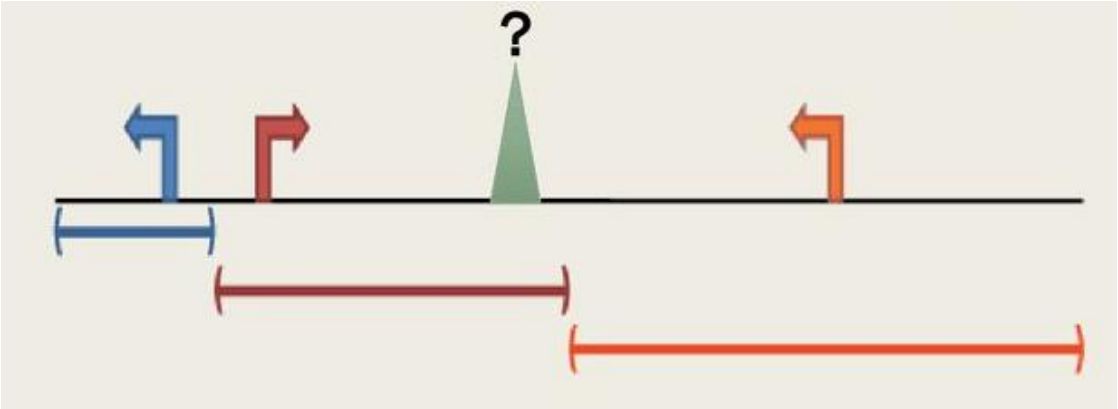
B. Assign peaks to genomic regions:



Downstream analysis of ATAC-seq

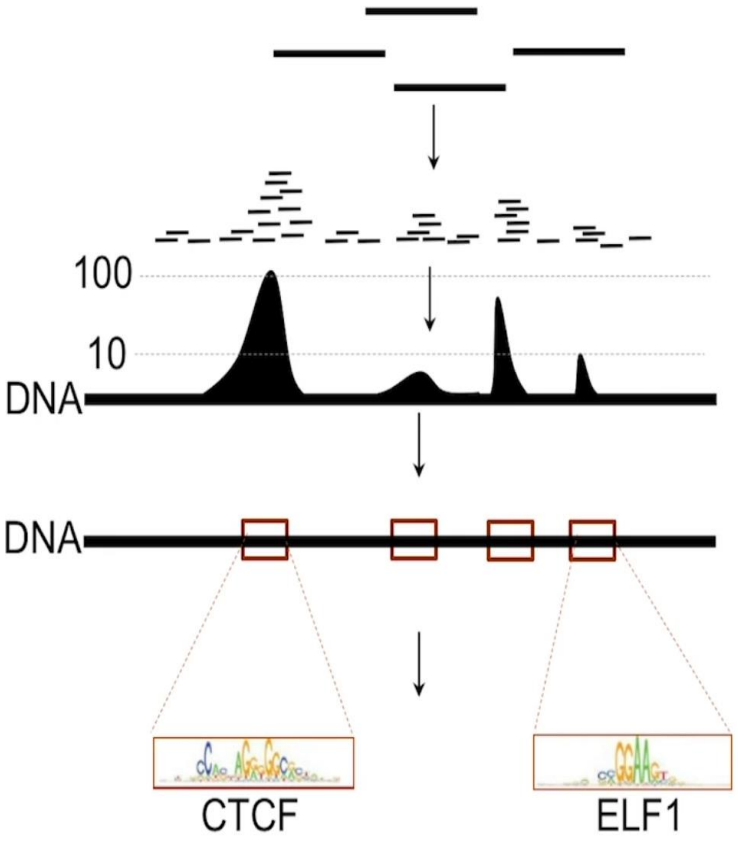
C. Peaks annotation and functional enrichment

- Assign peaks to nearest genes (using GREAT, HOMER)
- Quantification of peaks (DiffBind)
- Functional enrichment analysis
- Motif enrichment and TF footprint



Applications of ATAC-seq

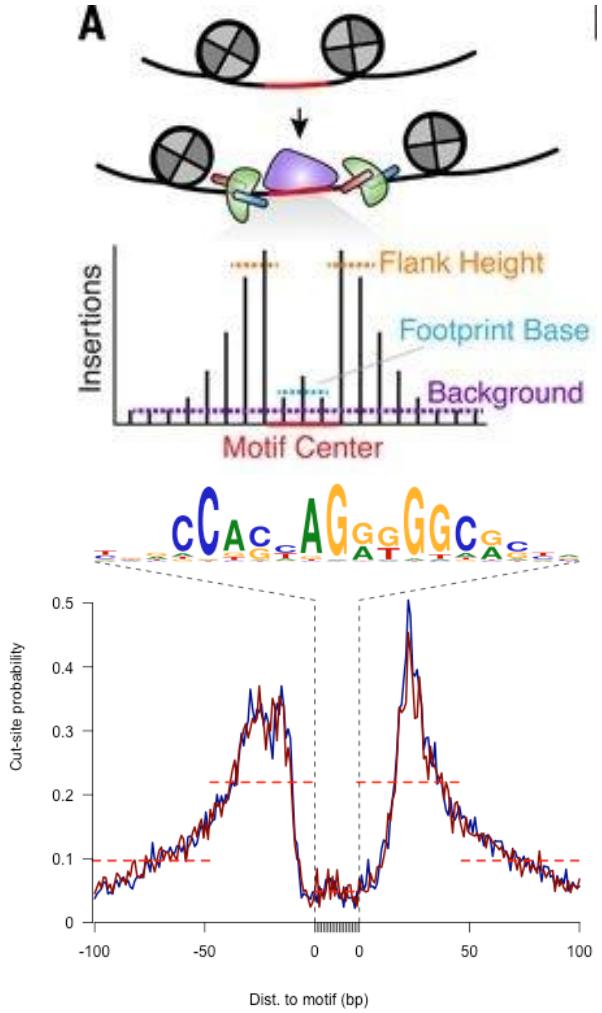
Motif enrichment



Infer footprints of DNA-protein binding

(genome-wide factor occupancy)

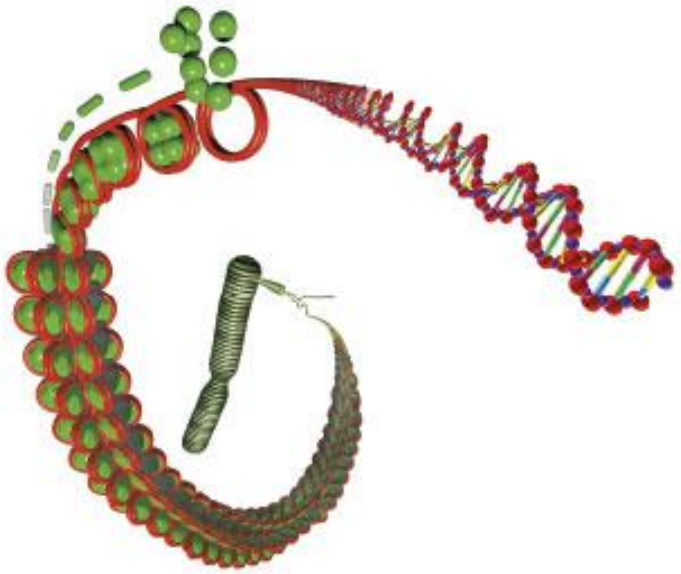
Requires deeper sequencing



Ricardo N Ramirez, Harvard Medical School

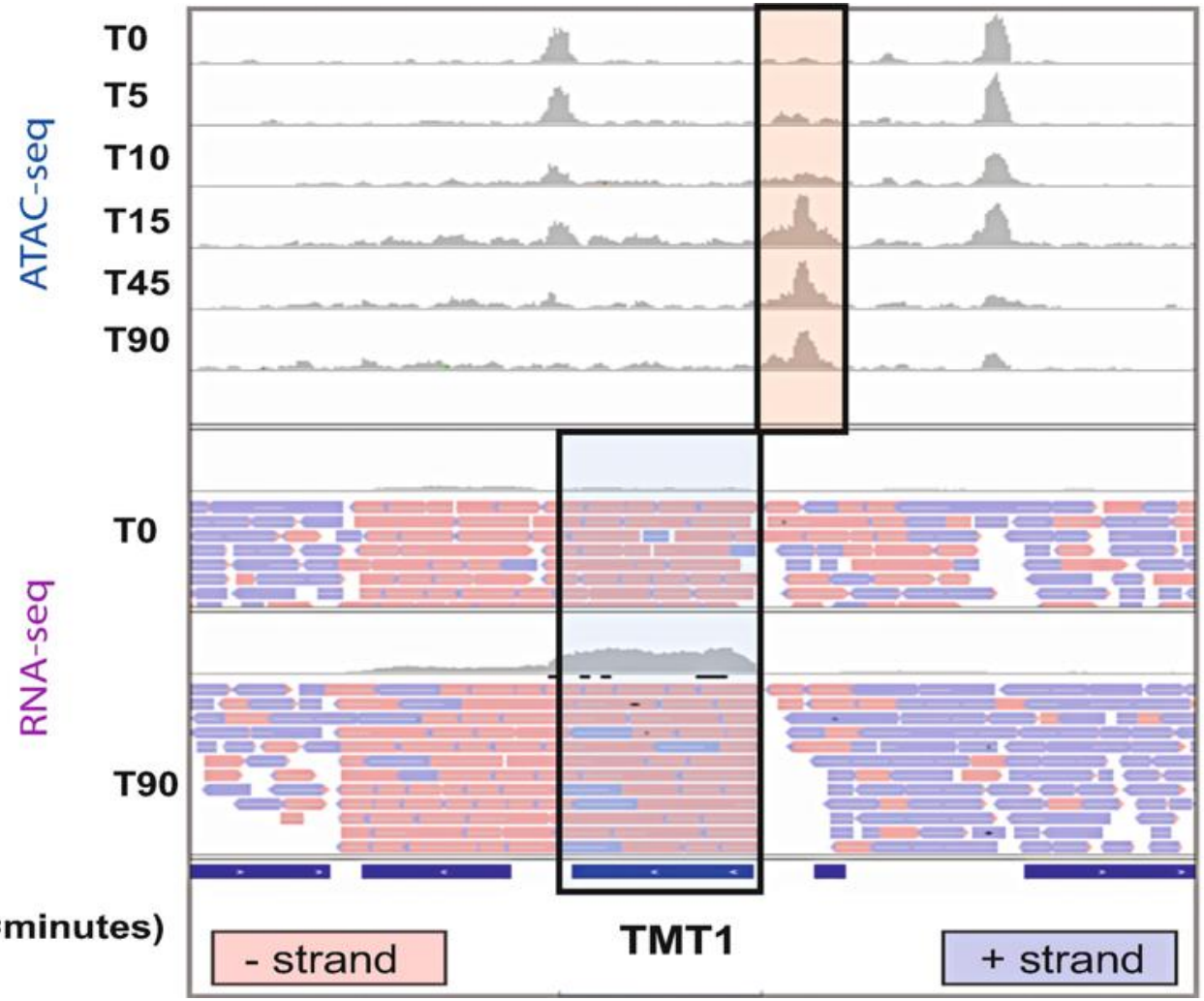
ATAC-seq lecture outline

- Why studying epigenomics?
- What is ATAC-seq?
- How is an ATAC-seq experiment designed?
- How is ATAC-seq analyzed bionformatically?
- Which are the available applications of ATAC-seq



Linking ATAC-seq with RNA-seq

Complementing open chromatin with gene expression for studying the relationship between genome structure and changes in regulation/function



For more reading:

Integration of ATAC-seq and RNA-seq to generate dynamic gene regulatory networks:

A Transcriptional Time Course of Myeloid Differentiation

(Ramirez et al., 2017, Cell Systems)

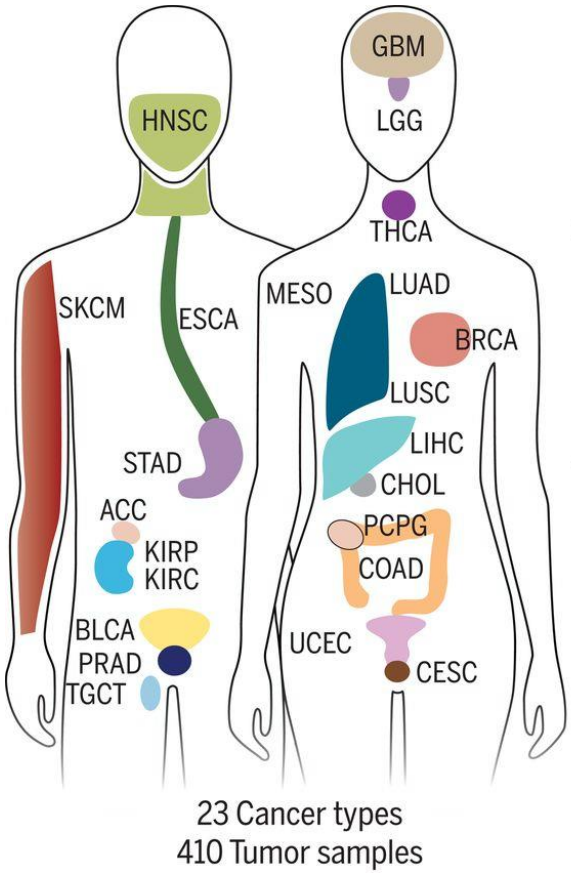
Combining ATAC-Seq with RNA-Seq:

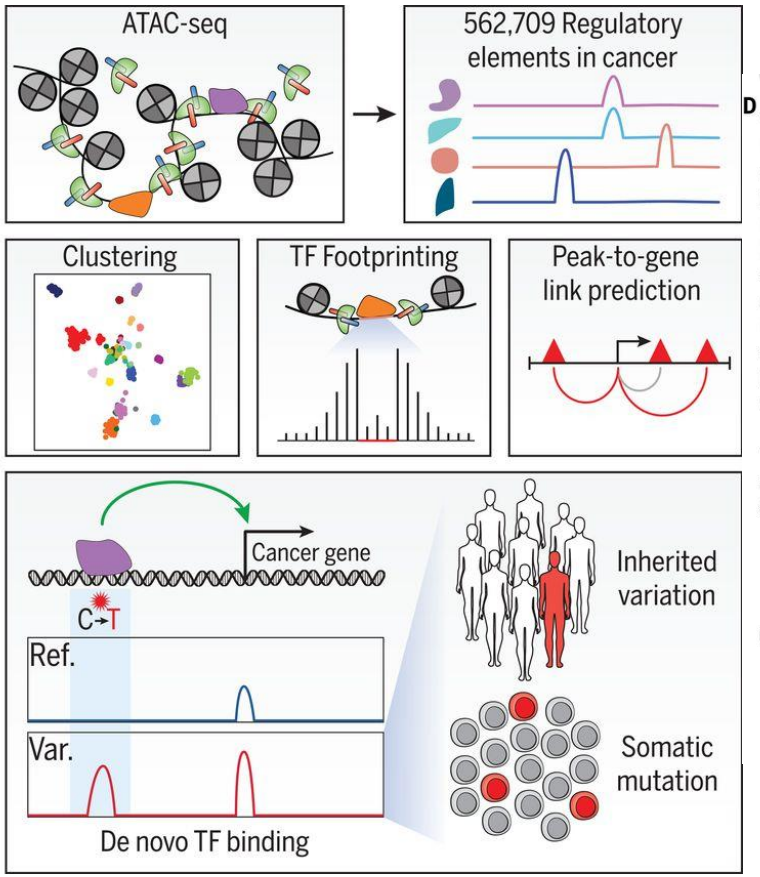
https://link.springer.com/protocol/10.1007%2F978-1-4939-8618-7_15

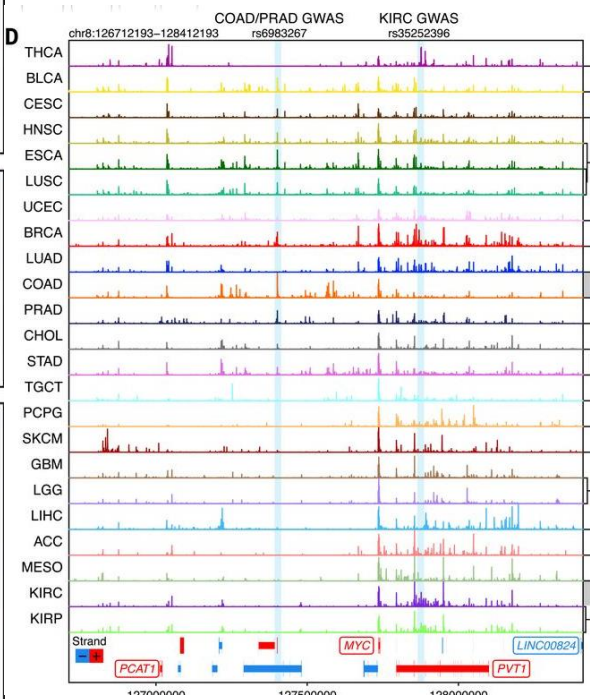
“The chromatin accessibility landscape of primary human cancers”

Science (2018)

- Generated ATAC-seq data in 410 tumor samples from TCGA across 23 cancer types.
- Identify distinct **TF-DNA** interactions in cancer
- Predicted **interactions** between distal regulatory elements from genome-wide correlation of gene expression and chromatin accessibility
- Linking regulatory interactions to cancer-linked genetic variants







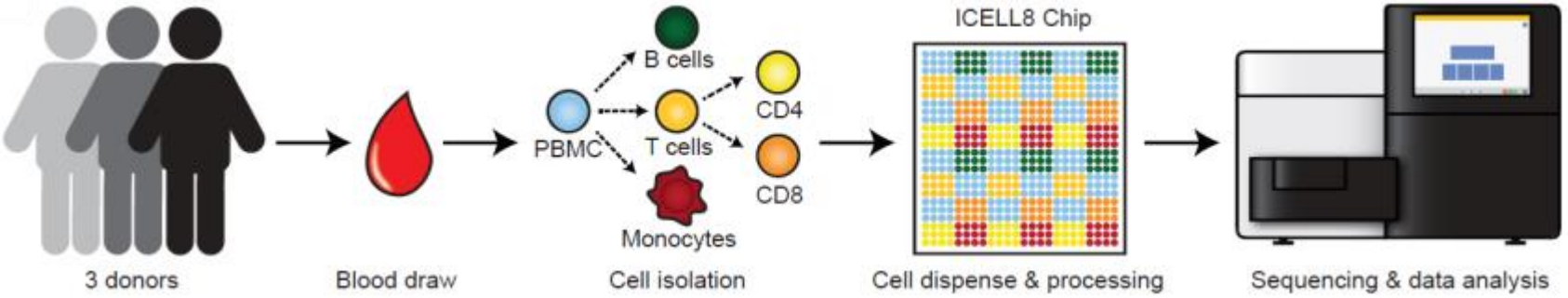
High-throughput single-cell ATAC-seq
Toward Single-cell “Regulomics”?

- ATAC-seq methods provide snapshots of a dynamic process that is **averaged** across thousands of cells (bulk)

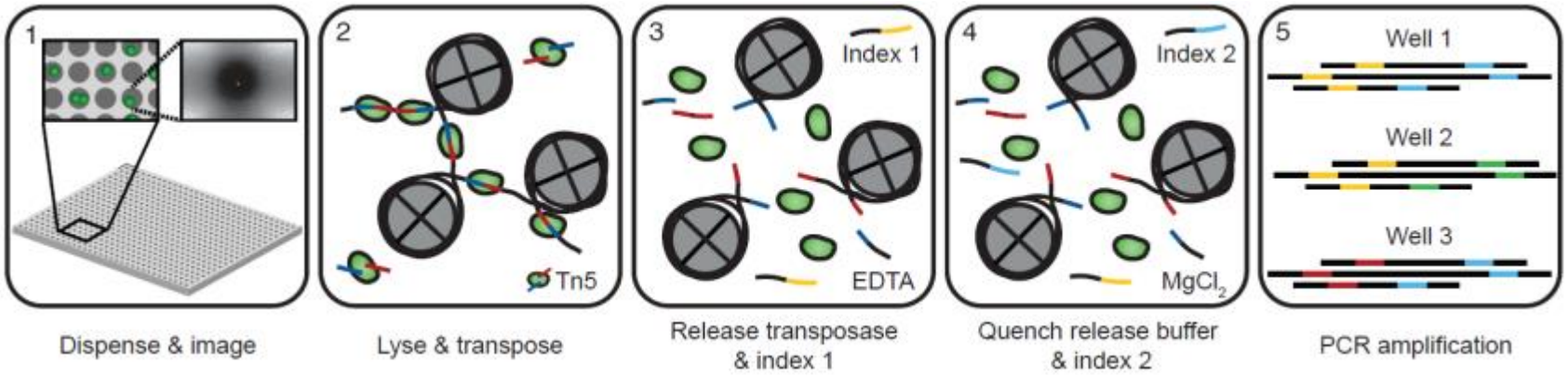
- Single-cell chromatin accessibility can potentially reveal **cell-type-specific** epigenomic variability

High-throughput single-cell ATAC-seq
Toward Single-cell "Regulomics"?

Example: using ATAC-seq to identify epigenomic states of multiple cell types from human donors:



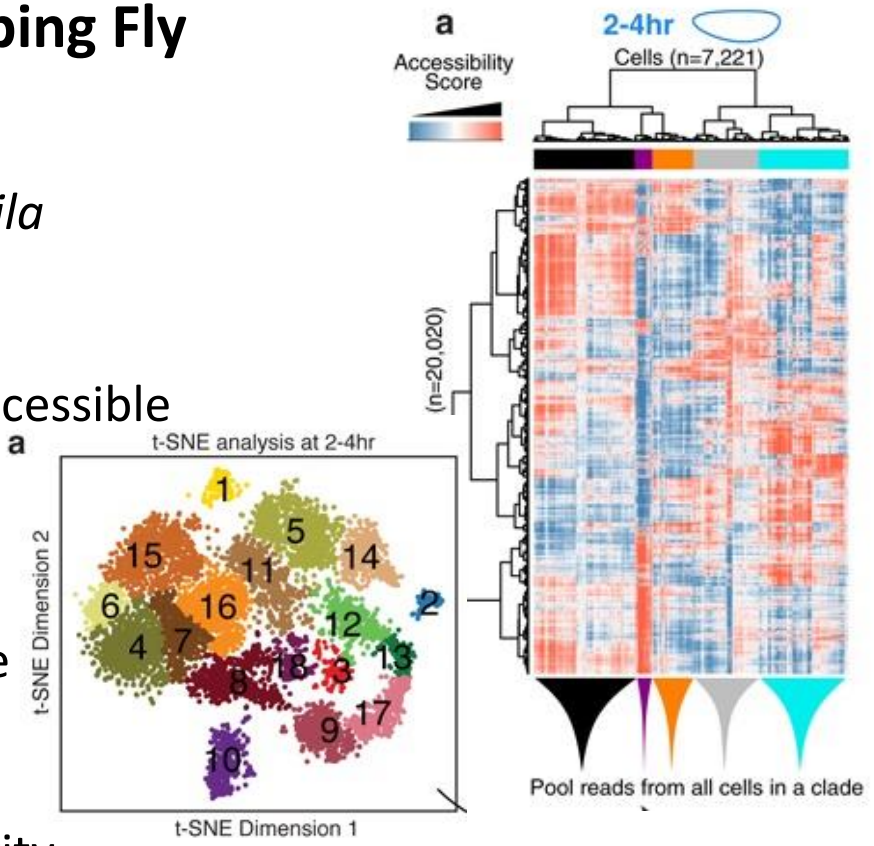
Implemented scATAC-seq workflow on the ICELL8:

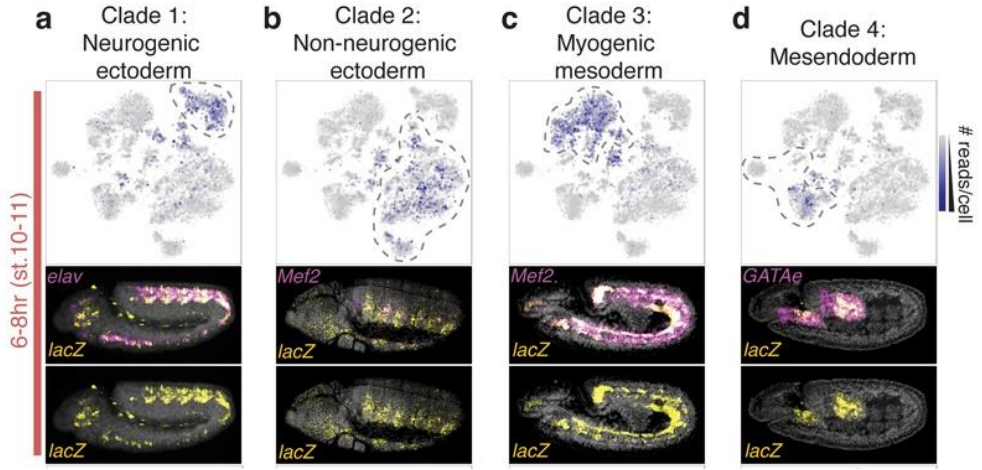


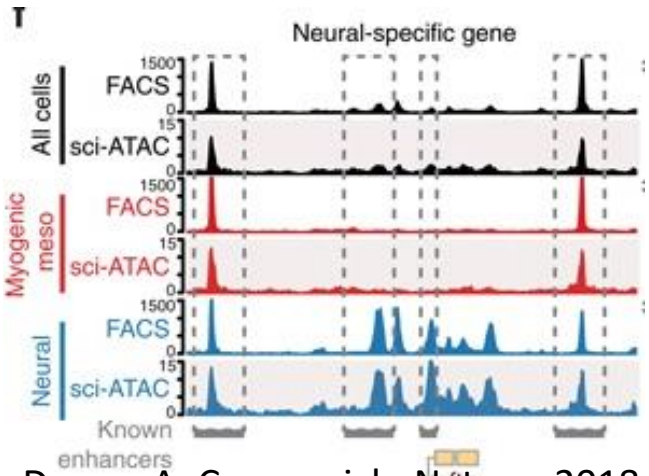
Mezger et al, Nature Communications, 2018 and takarabio

A Chromatin Cell Atlas of the Developing Fly Embryo using sci-ATAC-seq

- Profiled >20,000 single nuclei from *Drosophila* embryos
- Identify sites that were significantly more accessible in a specific cell type
- Intersect clade-specific peaks of chromatin accessibility with enhancer activity and gene expression
- Validation by *In situ* image of enhancer activity







Darren A. Cusanovich, Nature. 2018

ATAC-seq lecture summary

- ATAC-seq **captures** and **quantifies** open and accessible regions of chromatin
- ATAC-seq profiles genome-wide information on **nucleosome positioning in regulatory regions** (promoters, enhancers, or other regulatory elements accessible to transcription machinery)
- A transposase Tn5 cuts an exposed DNA region and simultaneously ligates sequencing adapters
- A bioinformatic workflow based on nucleosome-free fragments is available
- Available downstream applications and new methods (single-cell ATAC-seq)

Is chromatin accessibility indicative of active/functional regulatory regions?

"Patterns of reads in open chromatin regions result from a **complex interplay** of experimental effects with TF binding and nucleosome occupancy, among other biological factors"

He, H. H. et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. Nature Methods (2014).