ATAC-Seq analysis

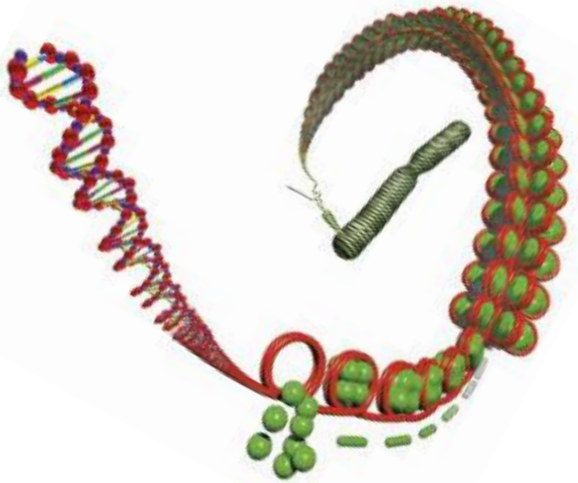
Bareket Dassa, Bioinformatics Unit

Introduction to Deep Sequencing Analysis course

2021

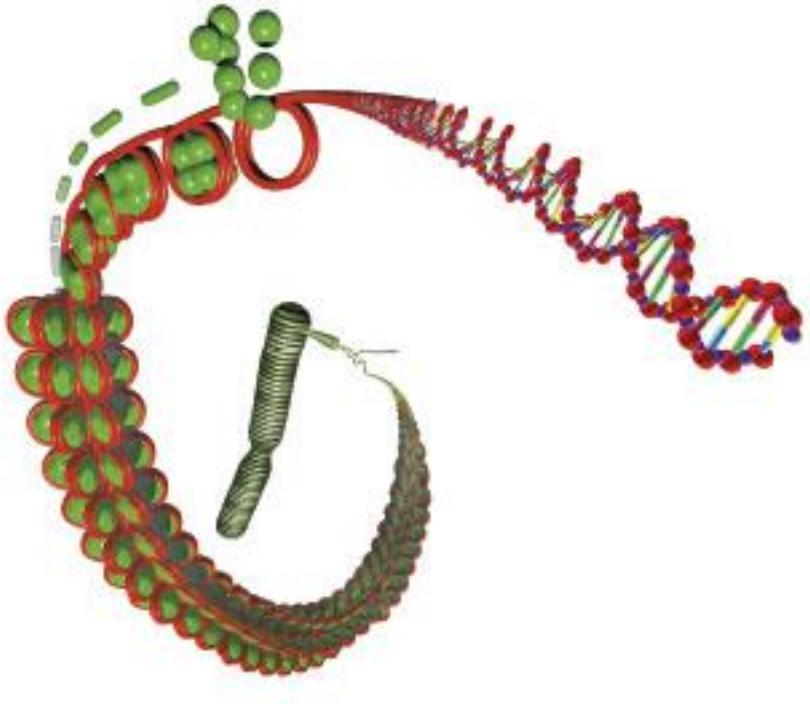
ATAC-Seq lecture outline

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



What is epigenomic regulation?

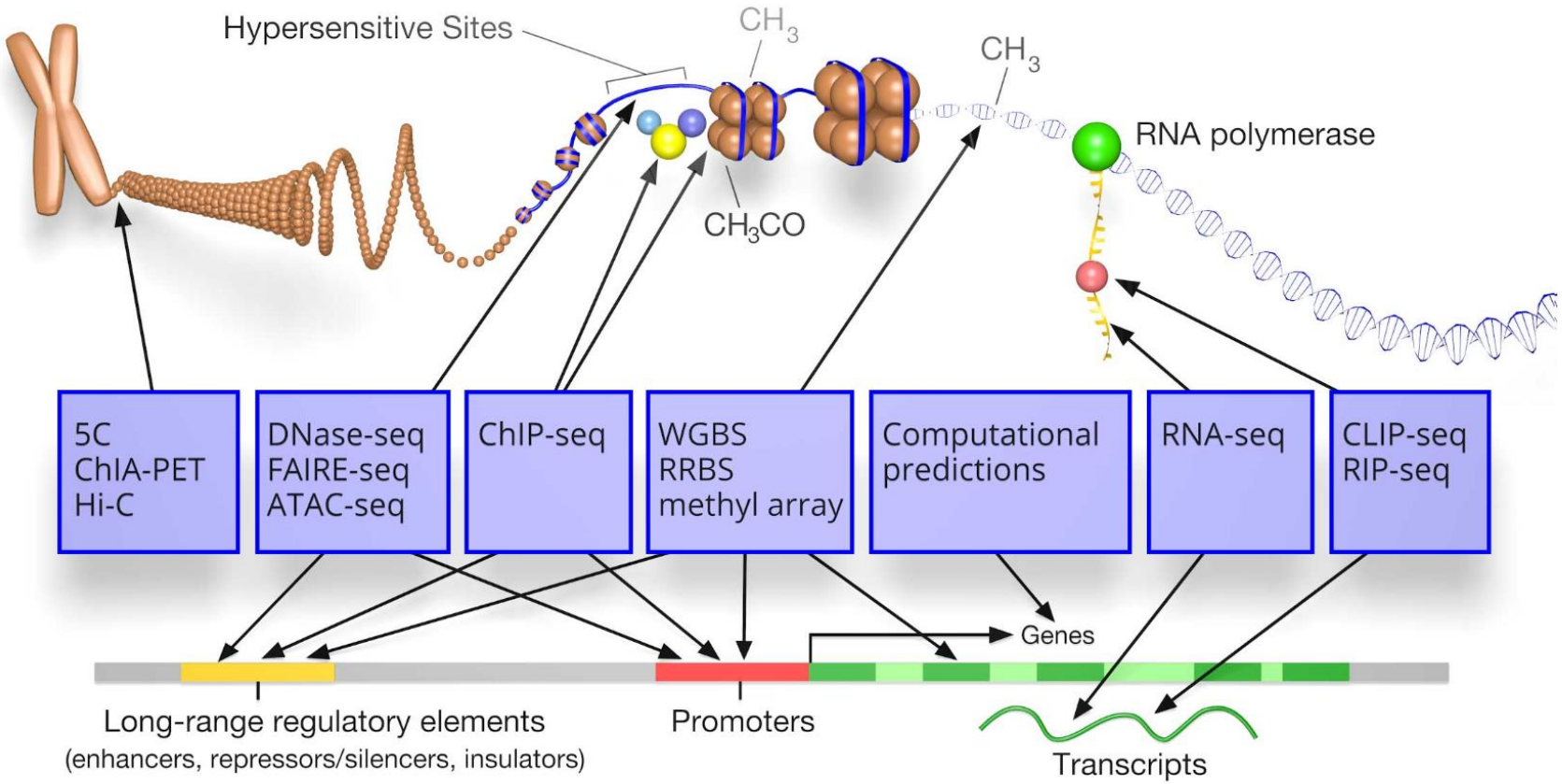
- Chromatin remodeling is highly dynamic
- Epigenetics involves genetic control by factors other than the DNA sequence
- Epigenetic **regulation** can switch genes on /off, and determine which and when genes are transcribed



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

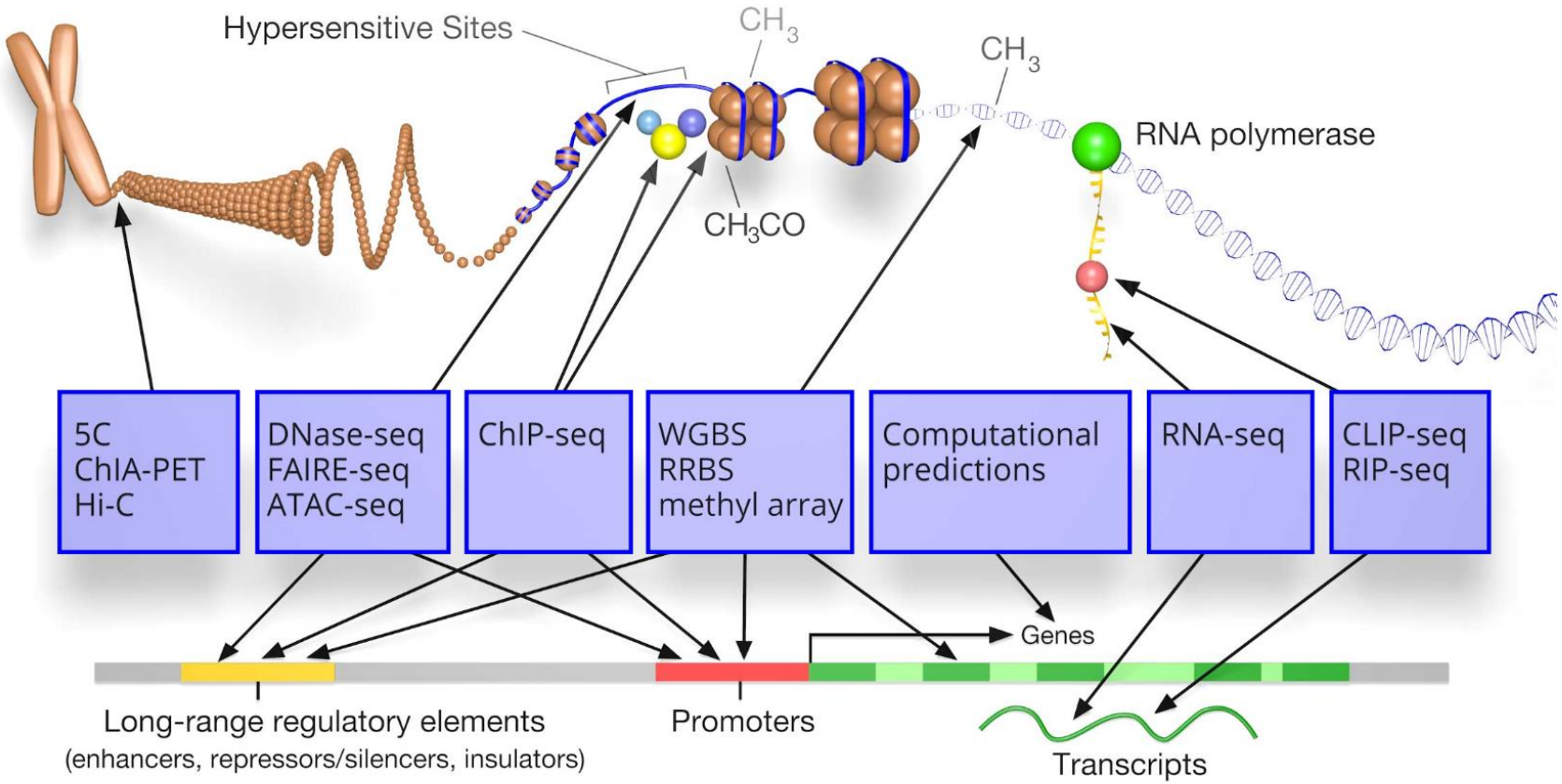


ENCODE

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**

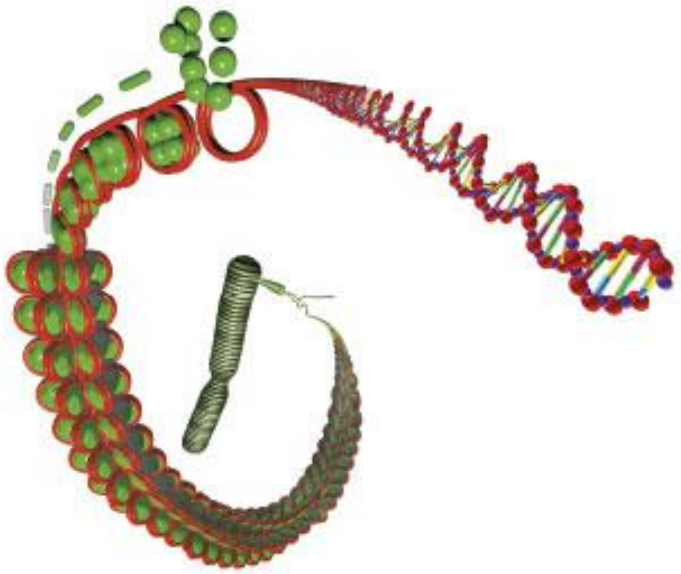


ENCODE

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 bioinformatically?
- Which are the available applications of ATAC-Seq



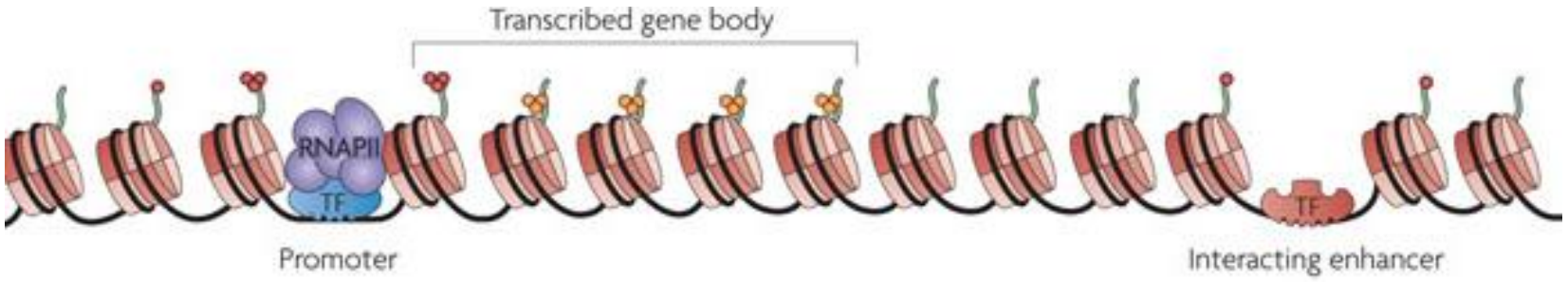
ATAC-Seq =
Assay for Transposase-Accessible Chromatin using sequencing

"open"

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

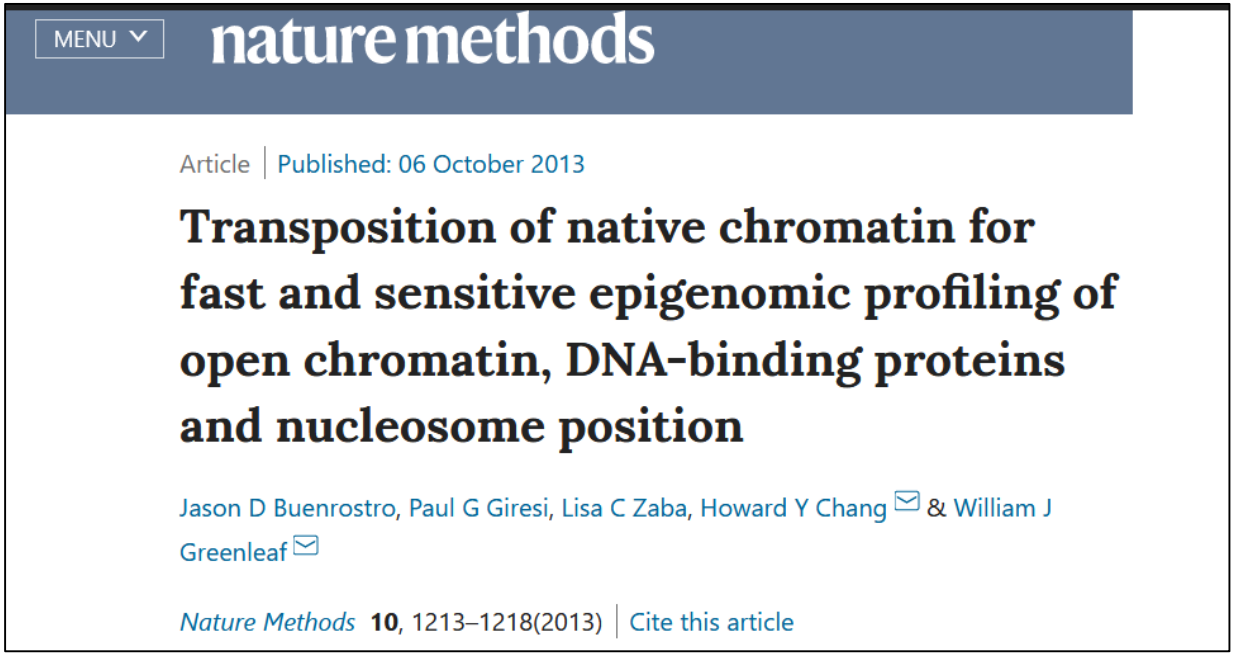
What information ATAC-Seq provides?

- **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

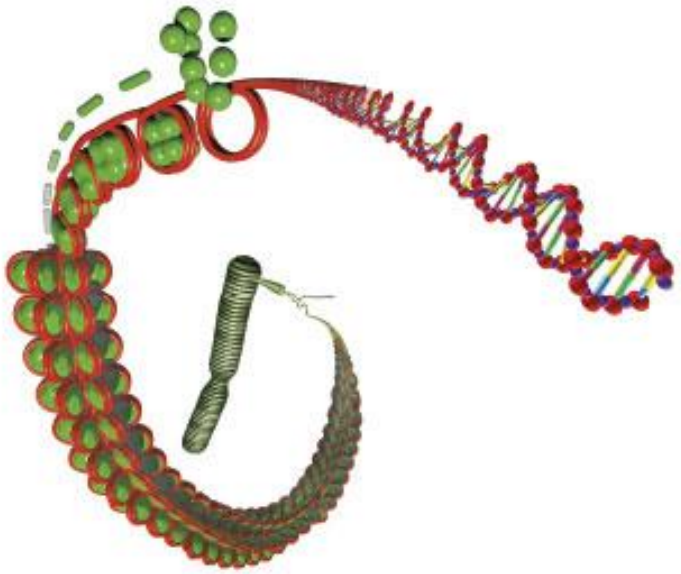
Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang & William J Greenleaf

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

ATAC-Seq lecture outline

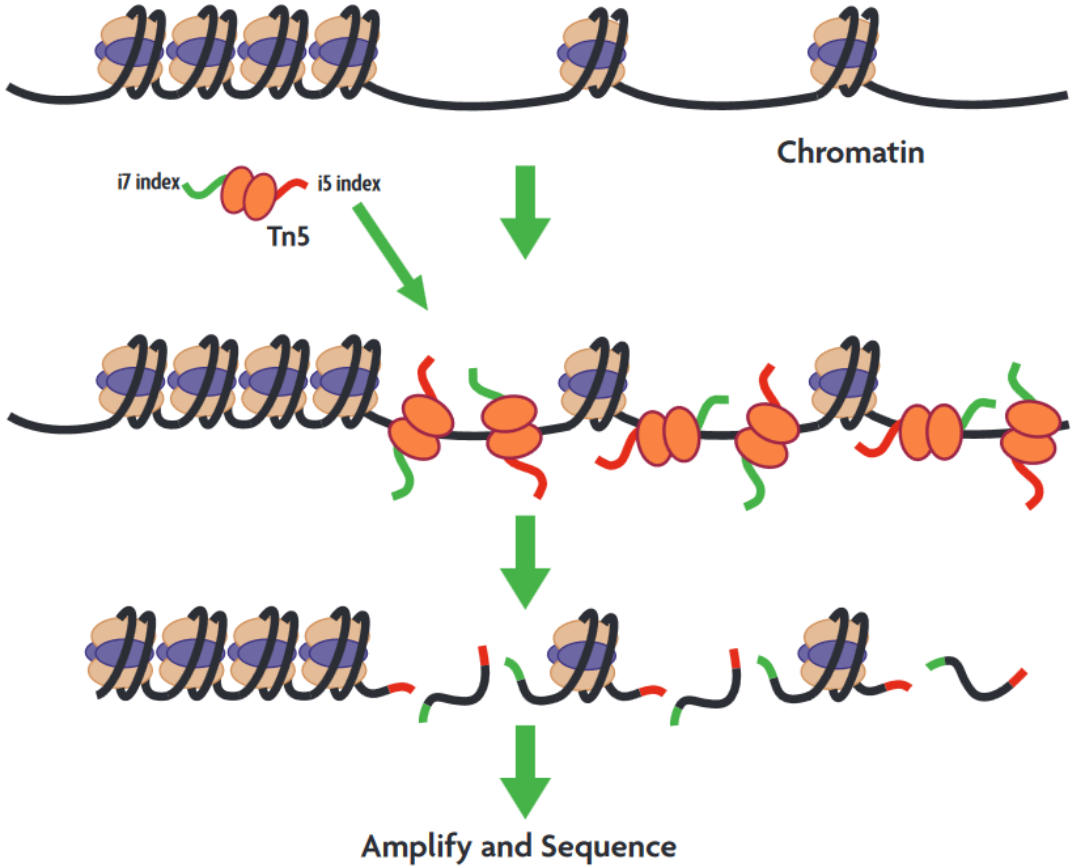
- Why studying epigenomics?
- What is ATAC-Seq?
- How is an ATAC-Seq experiment designed?
- How is ATAC-Seq analyzed bionformatically?
- Which are the available applications of ATAC-Seq



ATAC-Seq assay

- An engineered **transposase** is loaded with sequencing adapters

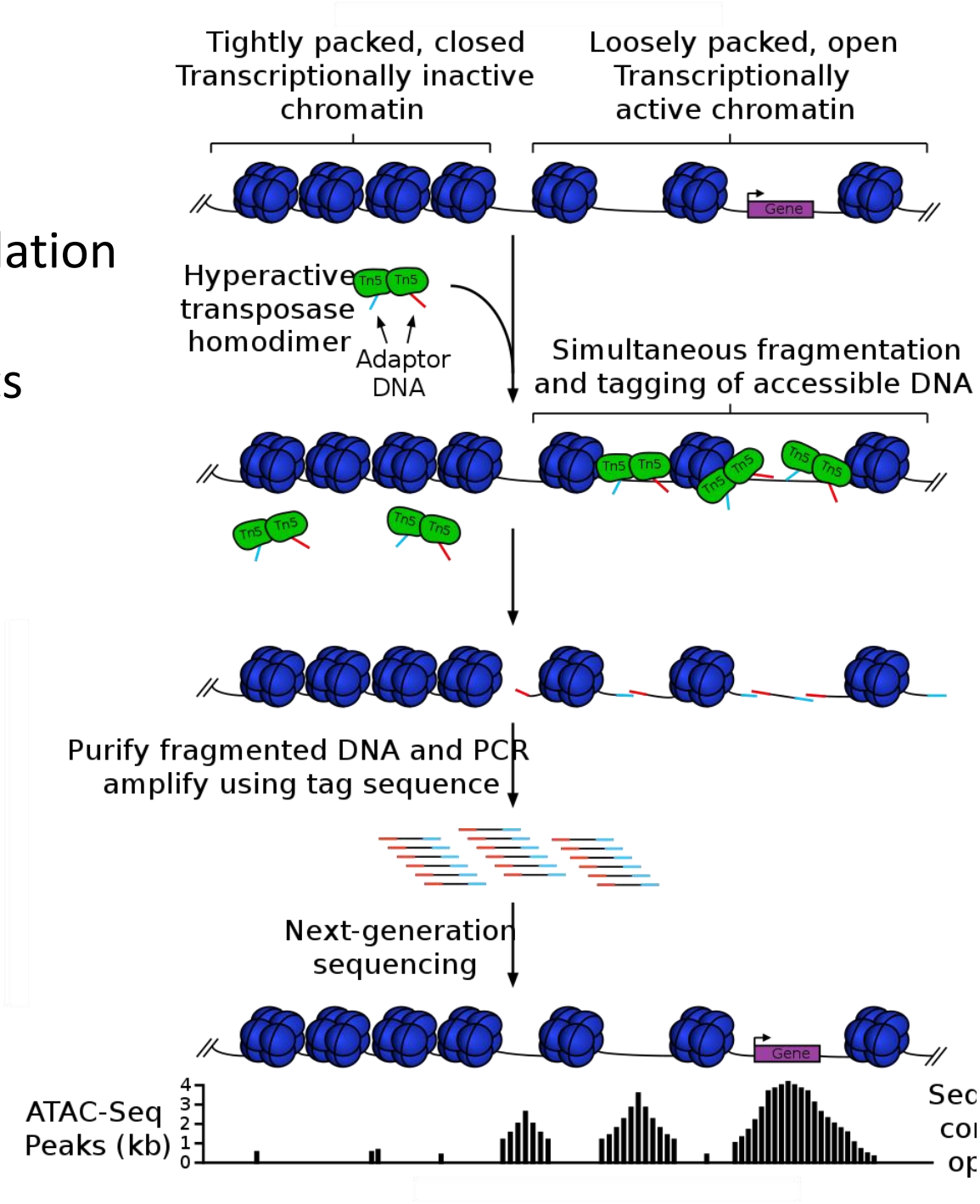
- **Tn5 tagmentation:** simultaneously **fragments** DNA only in regions of open chromatin
  AND
  adds tags (sequencing adapters) to generate a library for PCR amplification



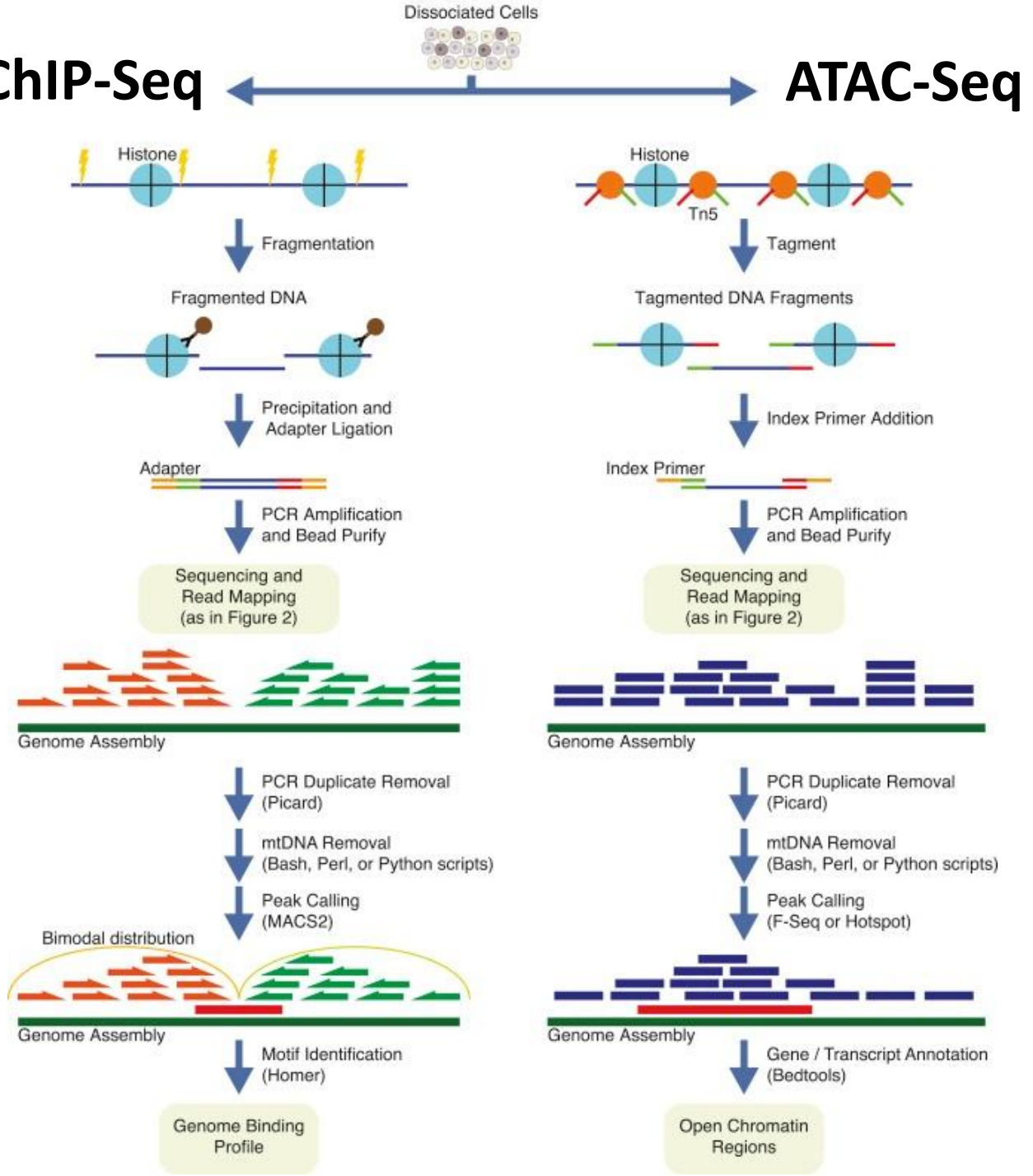
From: https://www.activemotif.com/catalog/1233/atac-seq-service

ATAC-Seq assay steps:

1. Cell lysis and nuclei isolation
2. Transposase reaction
3. Purification of tagments
4. Library amplification
5. Illumina sequencing
6. Bioinformatic analysis



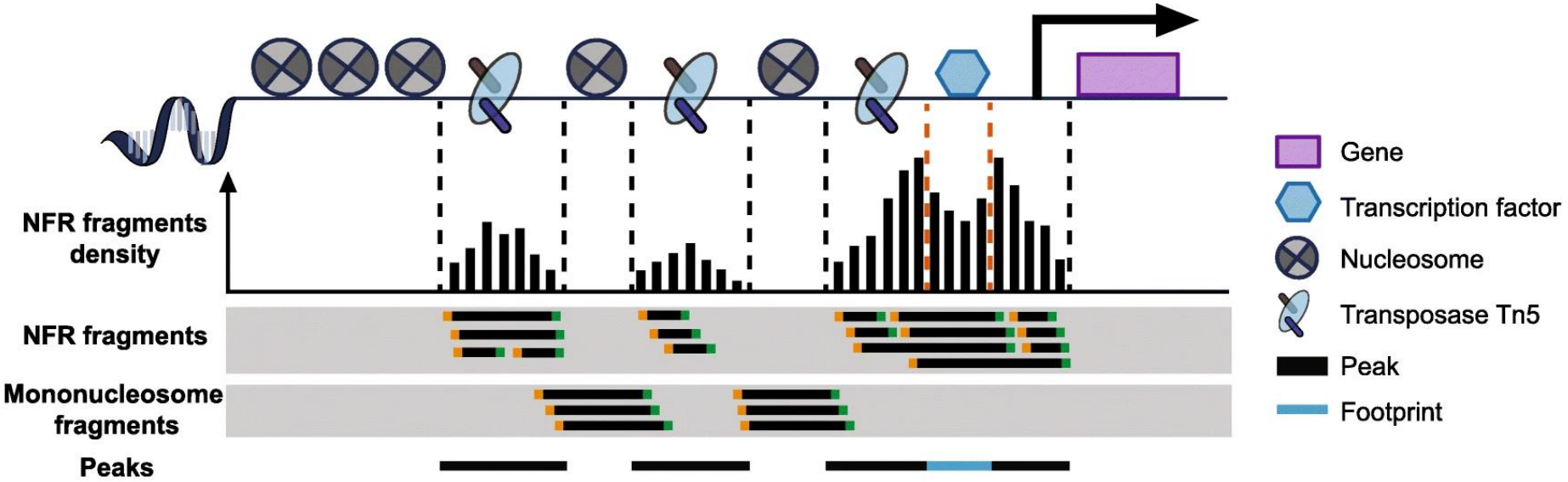
ATAC-Seq compared with ChIP-Seq



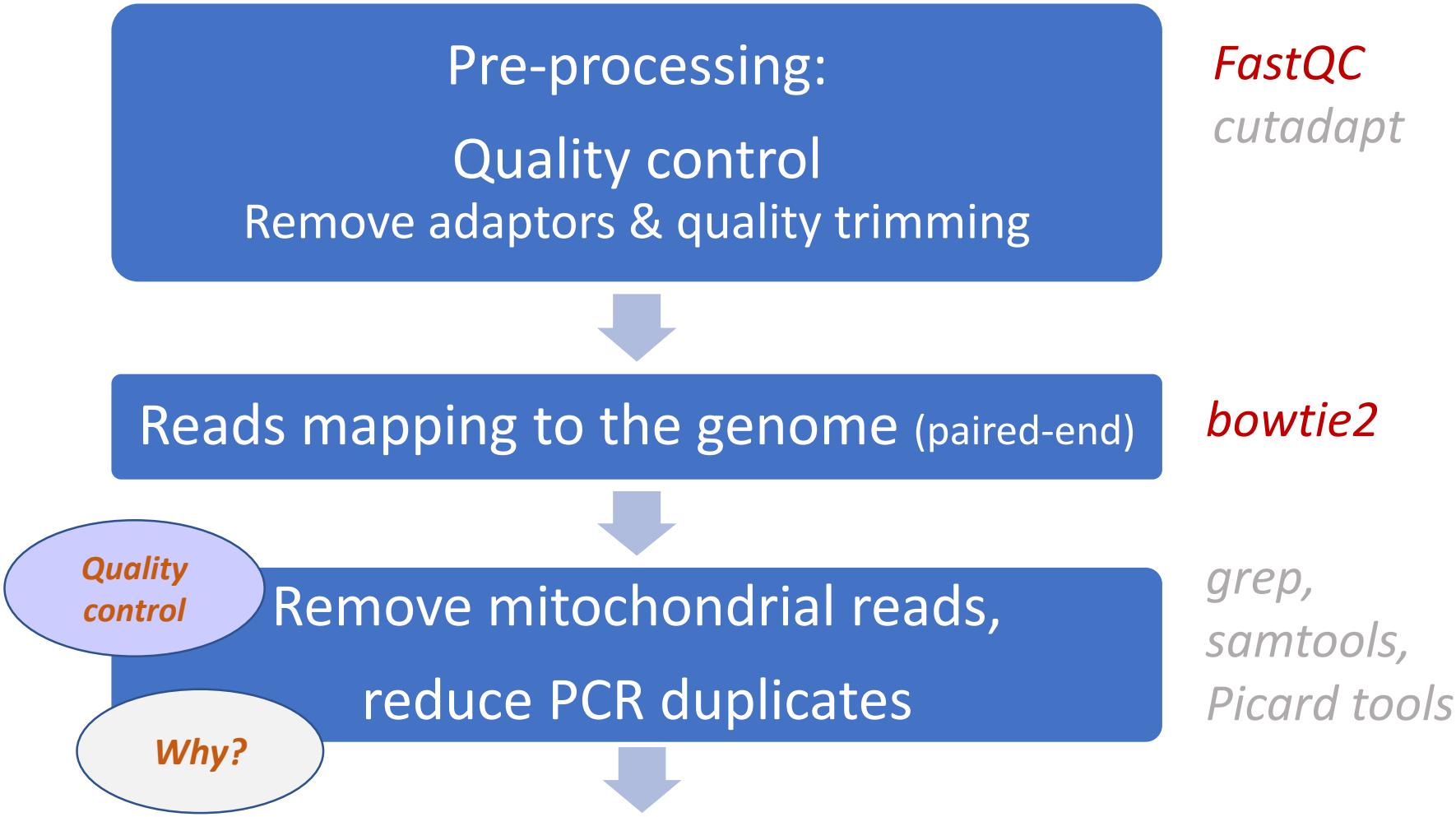
https://www.sciencedirect.com/science/article/pii/S1350946216300301#fig6

ATAC-Seq lecture outline

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



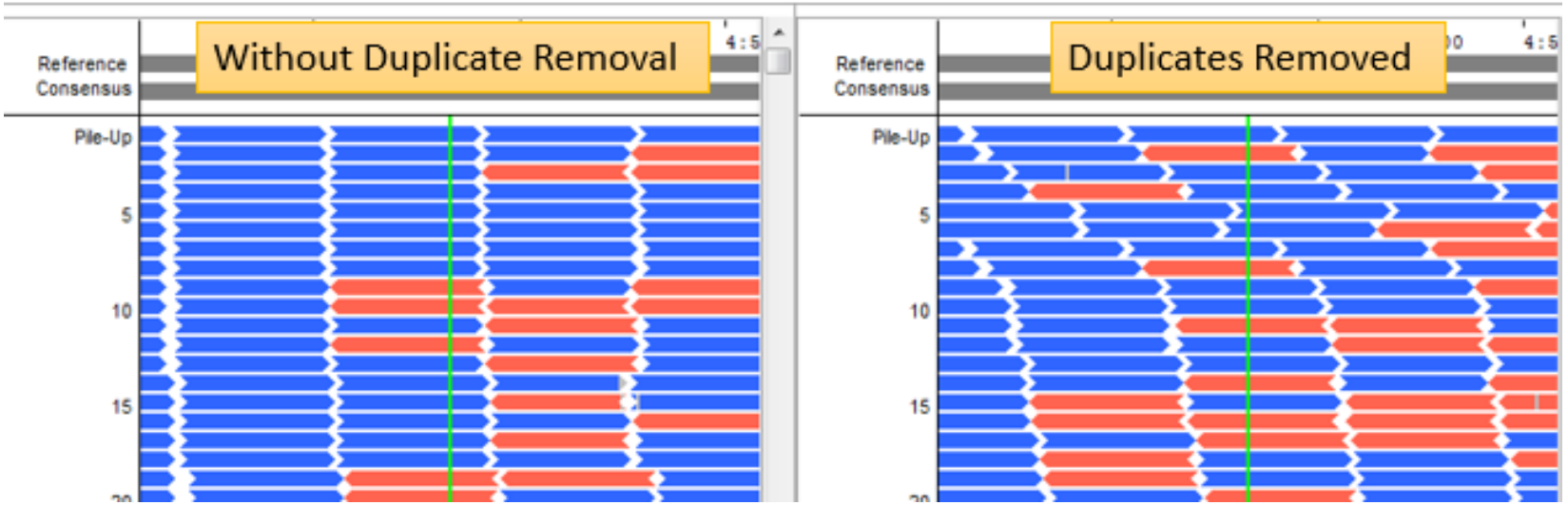
ATAC-Seq analysis workflow



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 (cont.)

Pre-processing steps:
Remove adaptors & quality trimming,
Quality control

FastQC
cutadapt

Reads mapping to the genome (paired-end)

bowtie2

Remove mitochondrial reads,
reduce PCR duplicates

grep,
samtools,
Picard tools

How?

Select nucleosome-free fragments

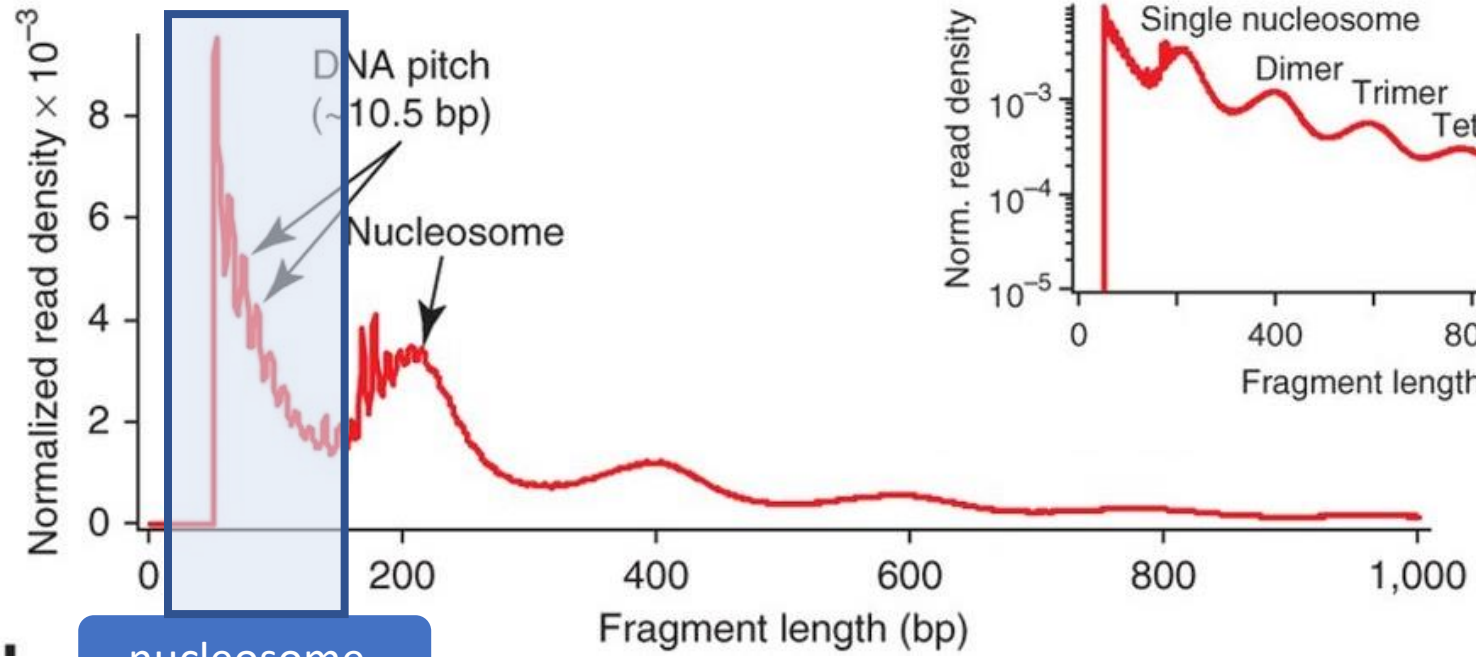
awk,
samtools

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

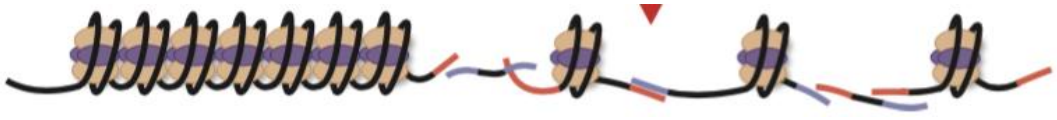


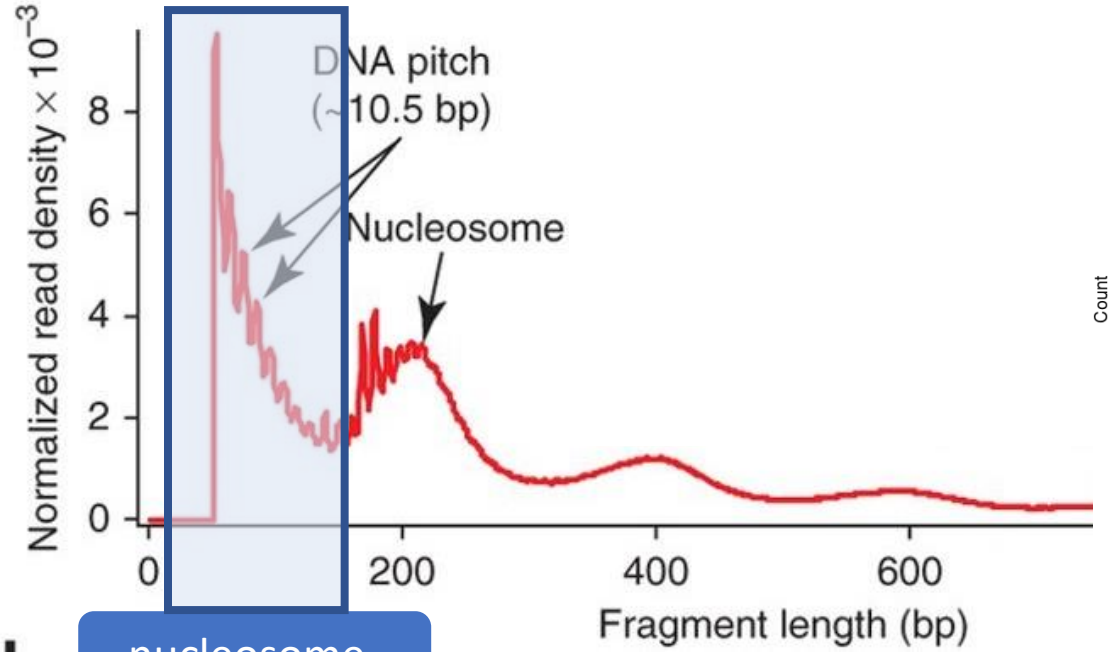
nucleosome-free fragments

Buenrostro (2013) Nature Methods

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

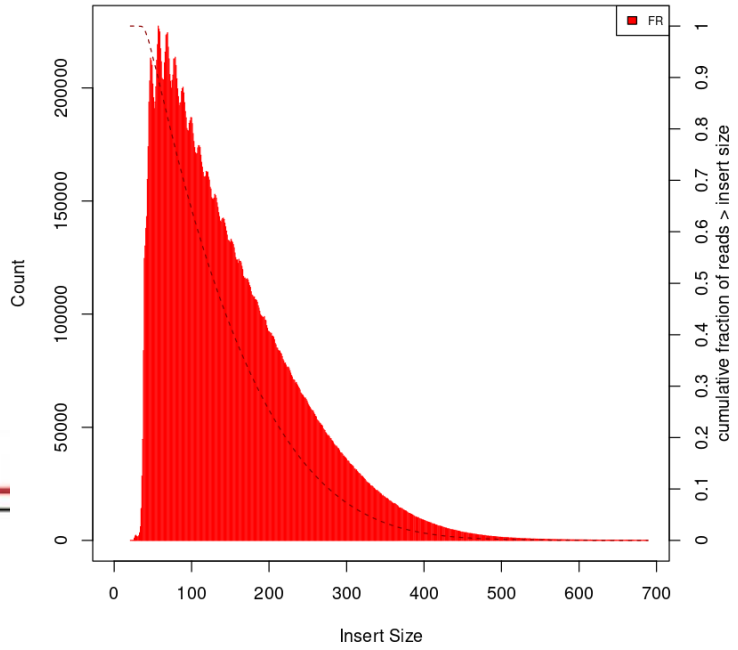




nucleosome-free fragments

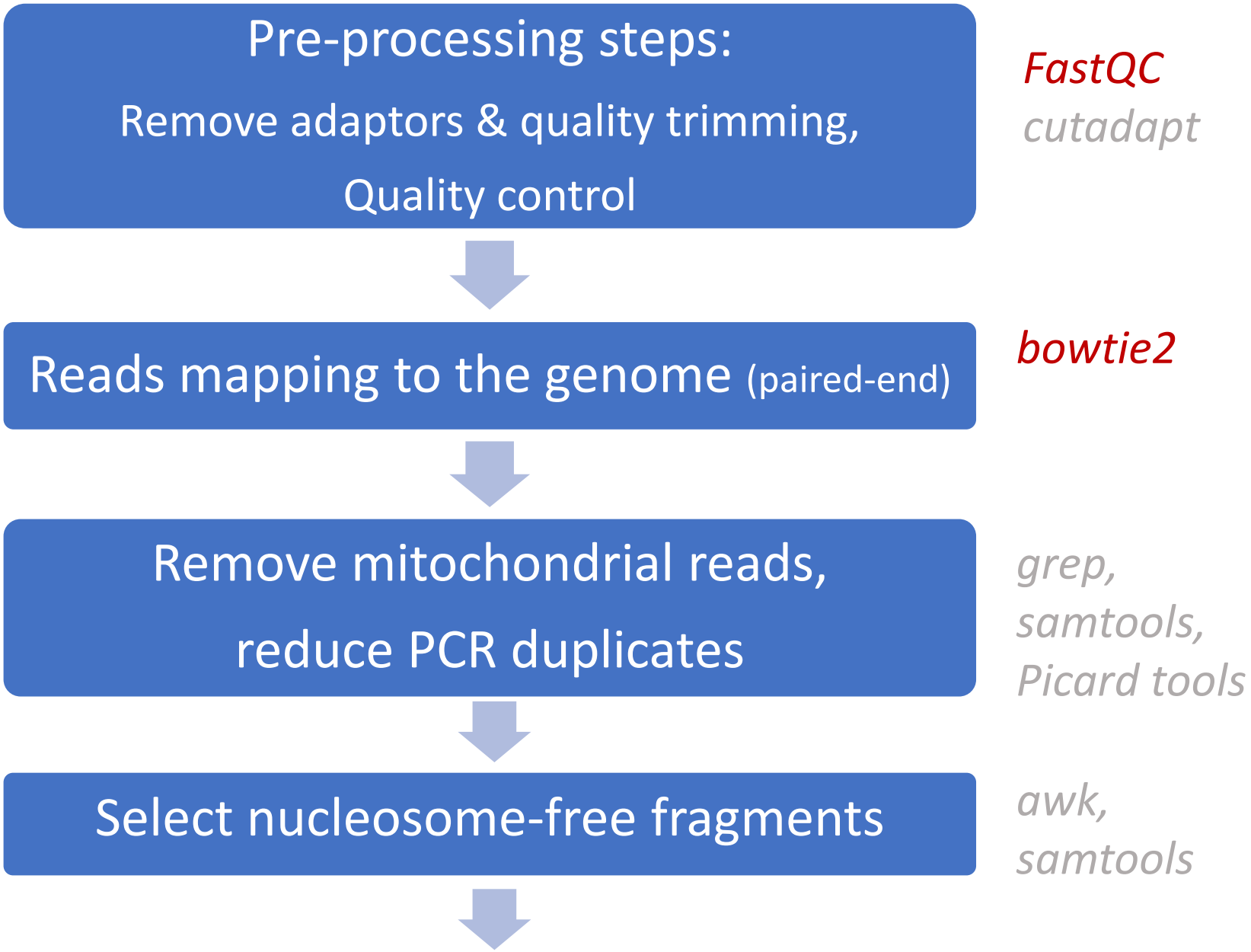
Buenrostro (2013) Nature Methods

The profile of "naked DNA" without nucleosomes



From: Galaxy Training

ATAC-Seq analysis workflow (cont.)



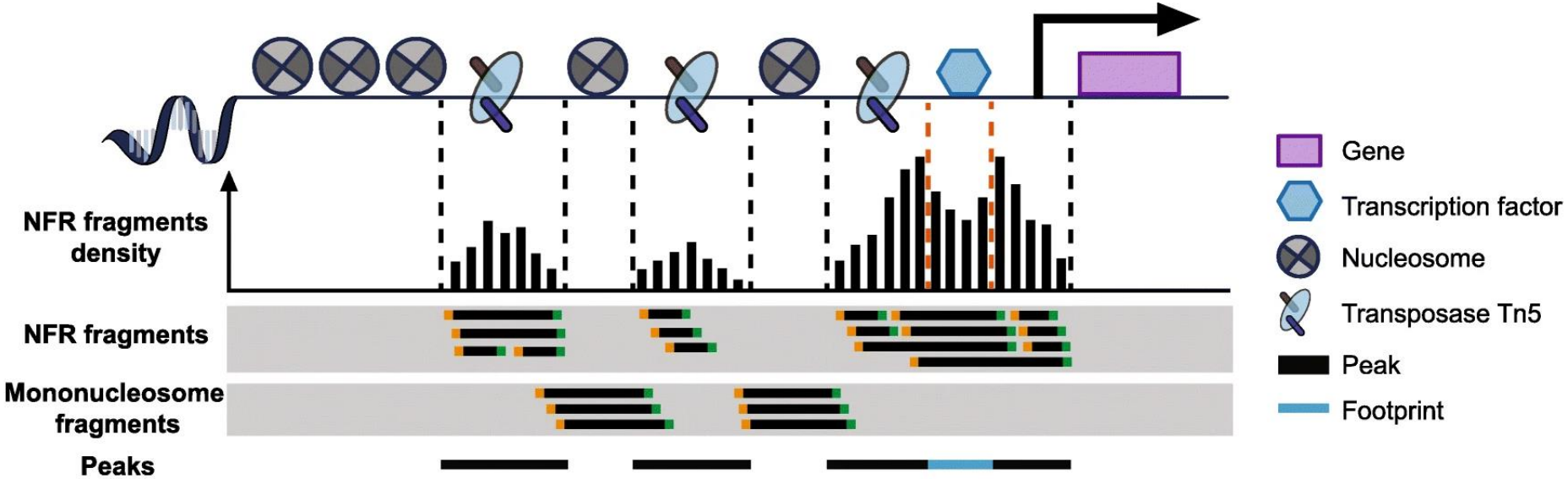
ATAC-Seq analysis workflow (cont.)

Peak calling (without a control)

macs2

Peaks visualization

IGV

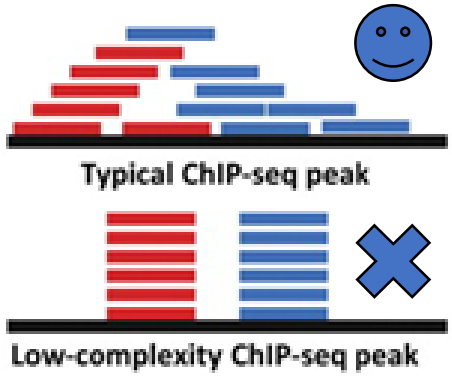


Known ATAC-Seq bias

- **Cleavage bias of the transposase**
  The Tn5 transposase can cleave DNA in a sequence-dependent manner, because of its tendency to cleave some DNA sequences more efficiently than others.

- **Avoiding high read redundancy**:

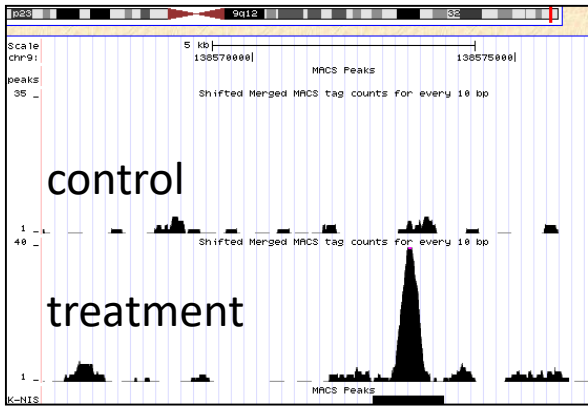
Filter out duplicate reads to avoid calling false peaks (reads at the exact same genome location and the same strand if their number exceeds the expected redundancy).



- **Sequencing depth, replicates**

An informative experiment should have >25 M aligned, non-duplicate, non-mitochondrial fragments

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

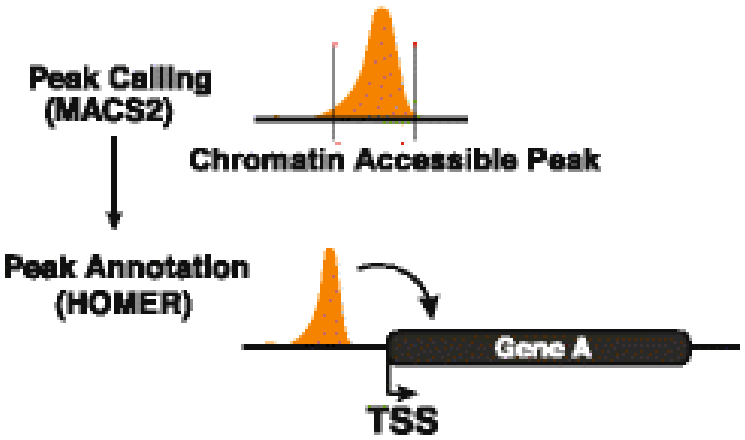


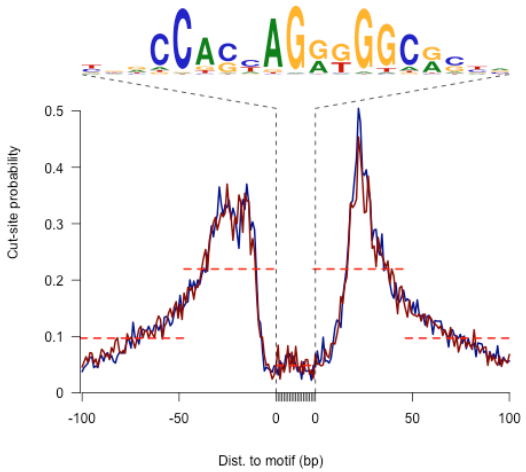
ATAC-Seq analysis workflow (2)

Peak calling — *macs2*

Peaks visualization — *IGV*

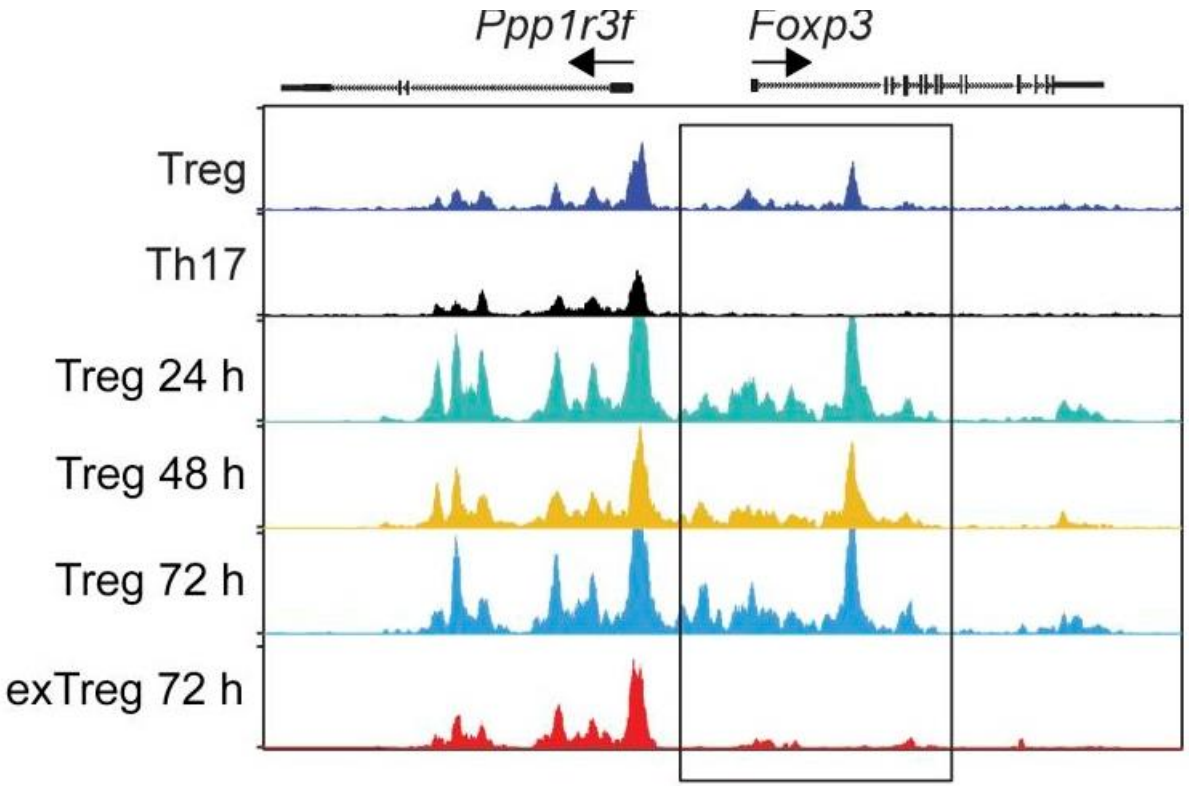
Peak annotation, Motif discovery, Functional enrichment — *Homer, GREAT, CEAS*





Downstream analysis of ATAC-Seq

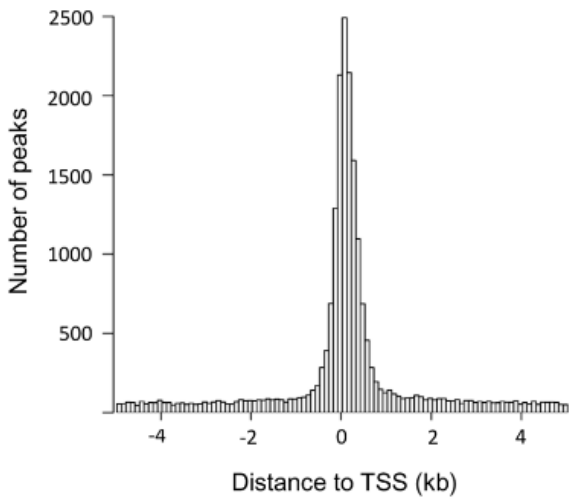
A. Peak calling and visualization on a genome browser (IGV, UCSC):

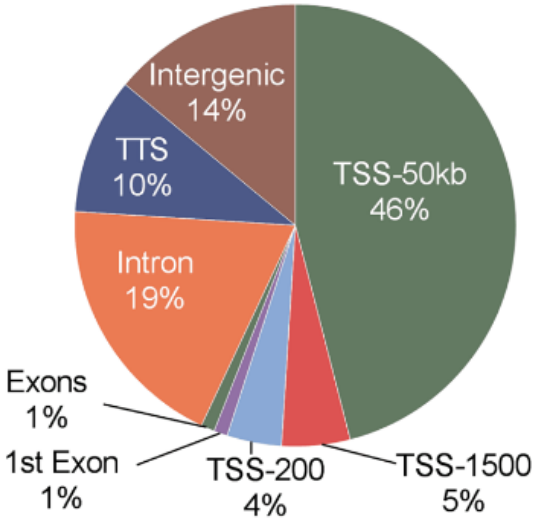


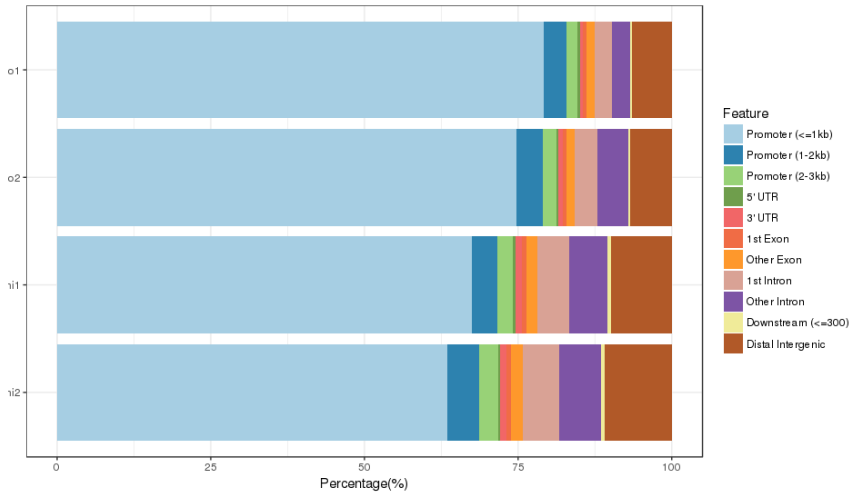
igv

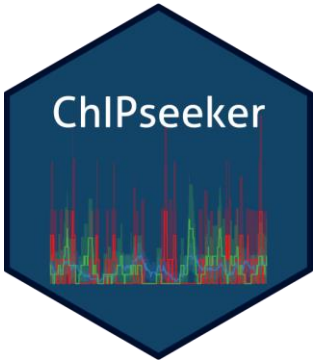
UCSC Genome Browser

B. Assign peaks to genomic regions:





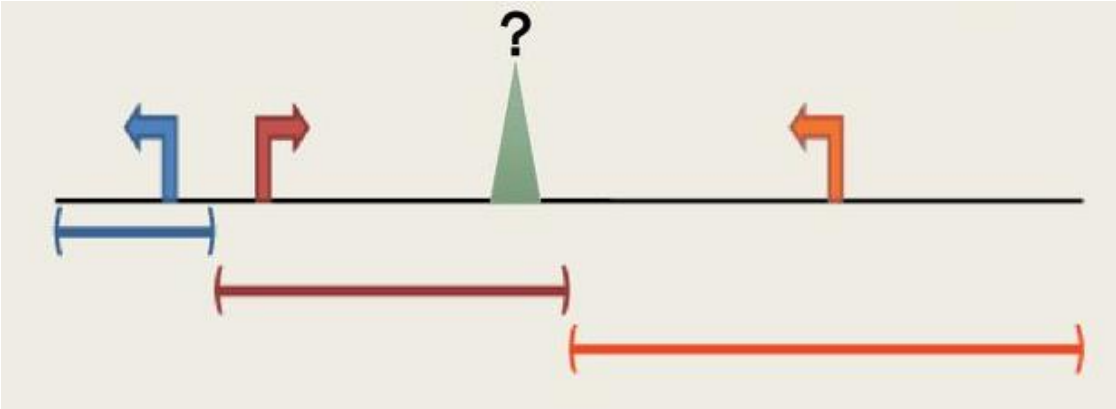




Downstream analysis of ATAC-Seq

C. Peaks annotation and functional enrichment

- Assign peaks to nearest genes (using GREAT, HOMER)
- Gene functional enrichment analysis
- Quantification of peaks (DiffBind)



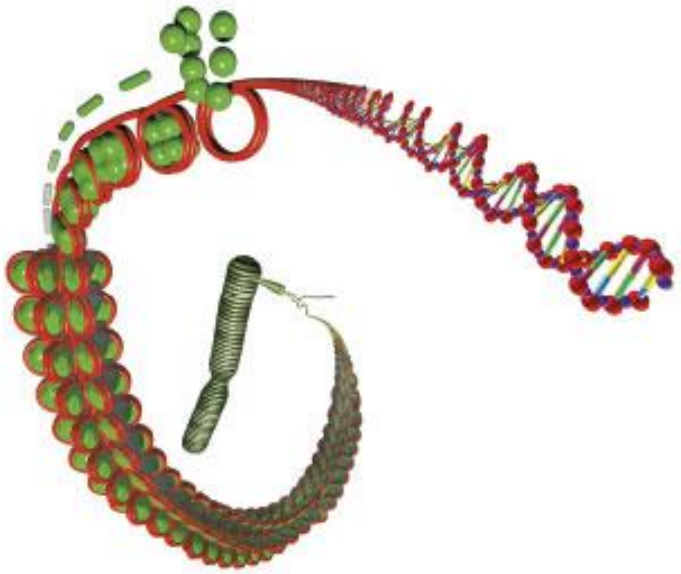
HOMER
Software for motif discovery and next-gen sequencing analysis

MEME-ChIP
Motif Analysis of Large Nucleotide Datasets

GREAT

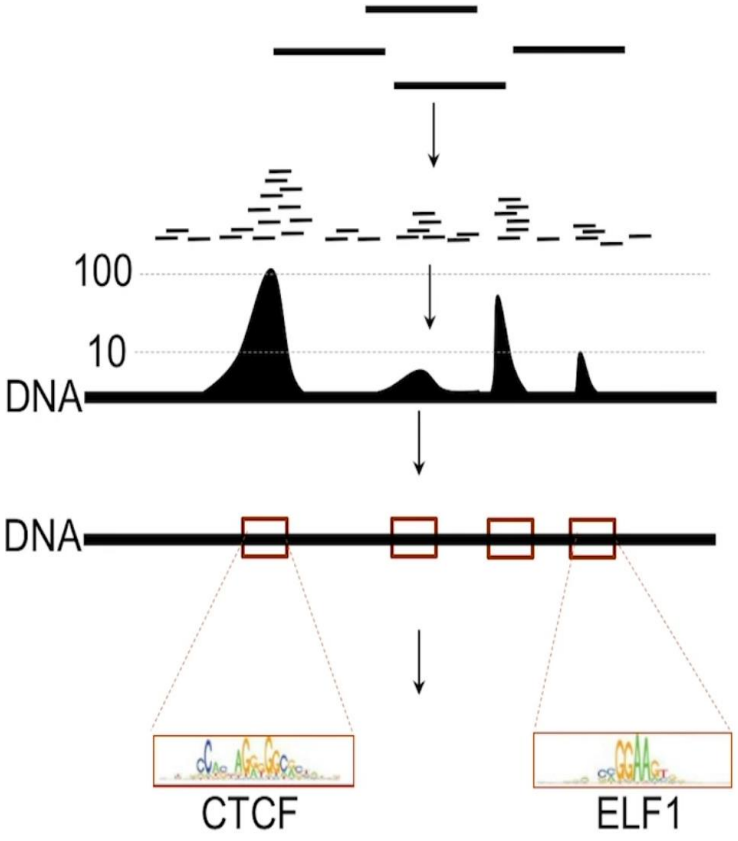
ATAC-Seq lecture outline

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



Applications of ATAC-Seq

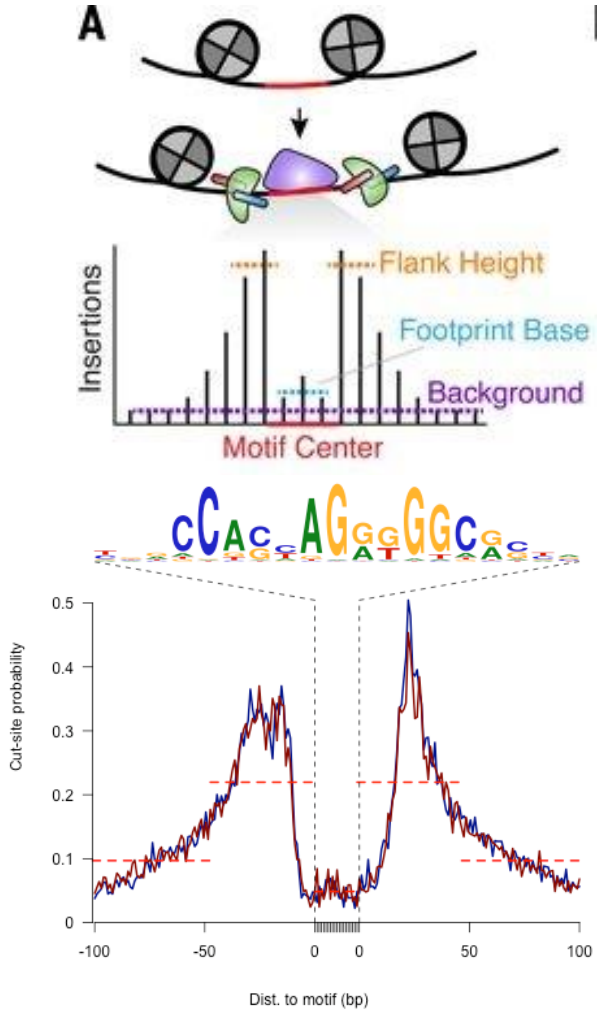
Motif discovery/enrichment



Infer footprints

of DNA-protein binding

Requires deeper sequencing

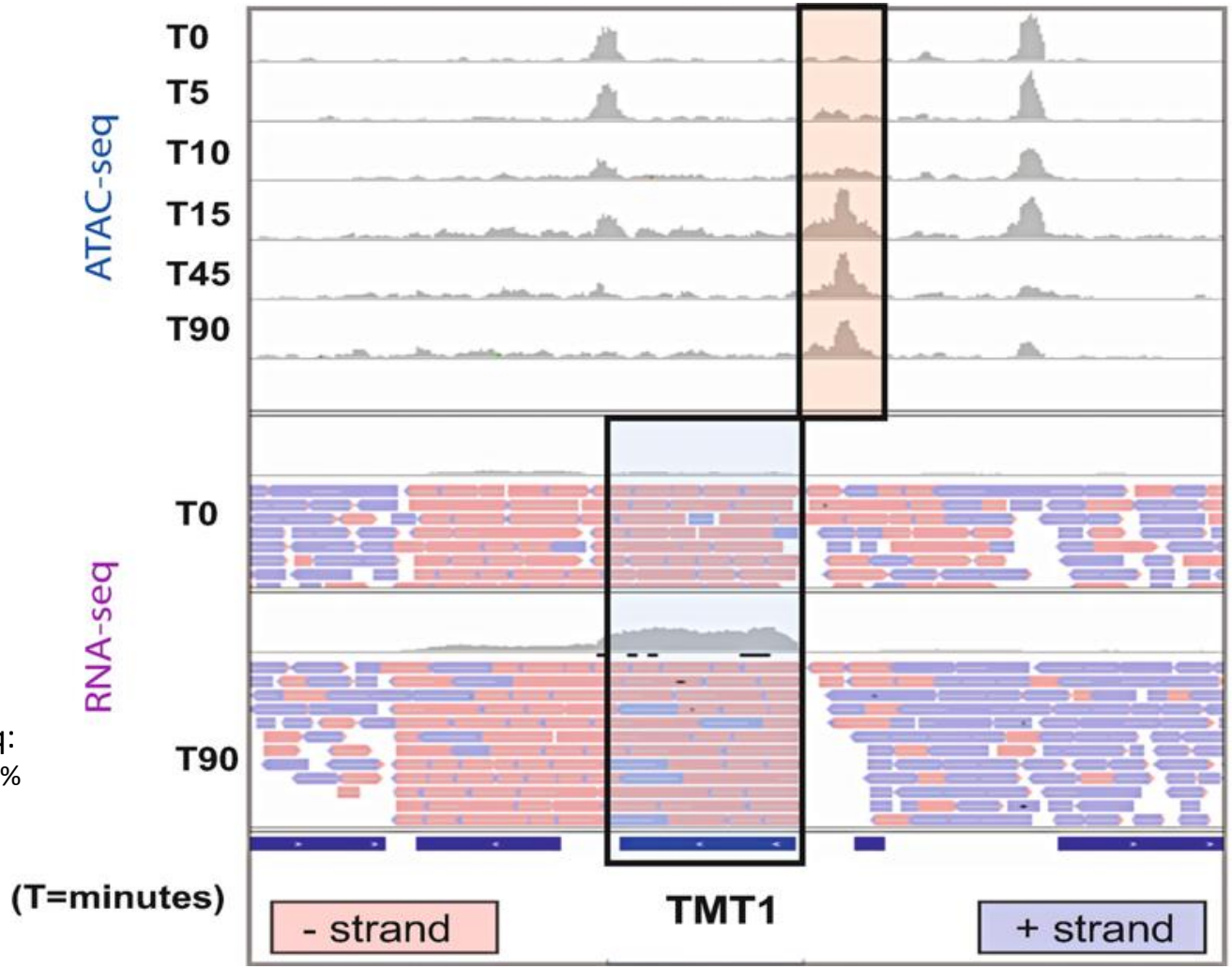


Ricardo N Ramirez, Harvard Medical School

Is chromatin accessibility indicative of active/functional genes?

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:

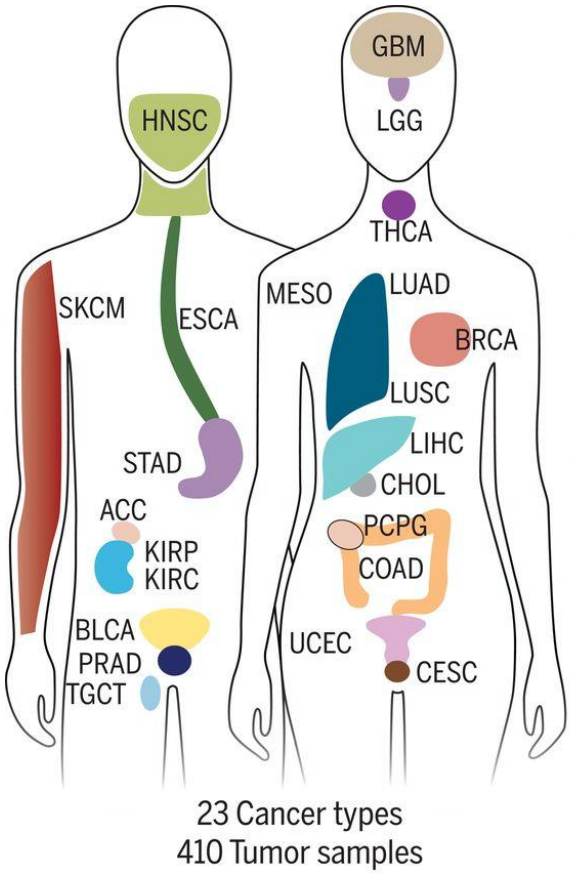
1. Integration of ATAC-Seq and RNA-seq to generate dynamic gene regulatory networks:
A Transcriptional Time Course of Myeloid Differentiation
(Ramirez et al., 2017, Cell Systems)

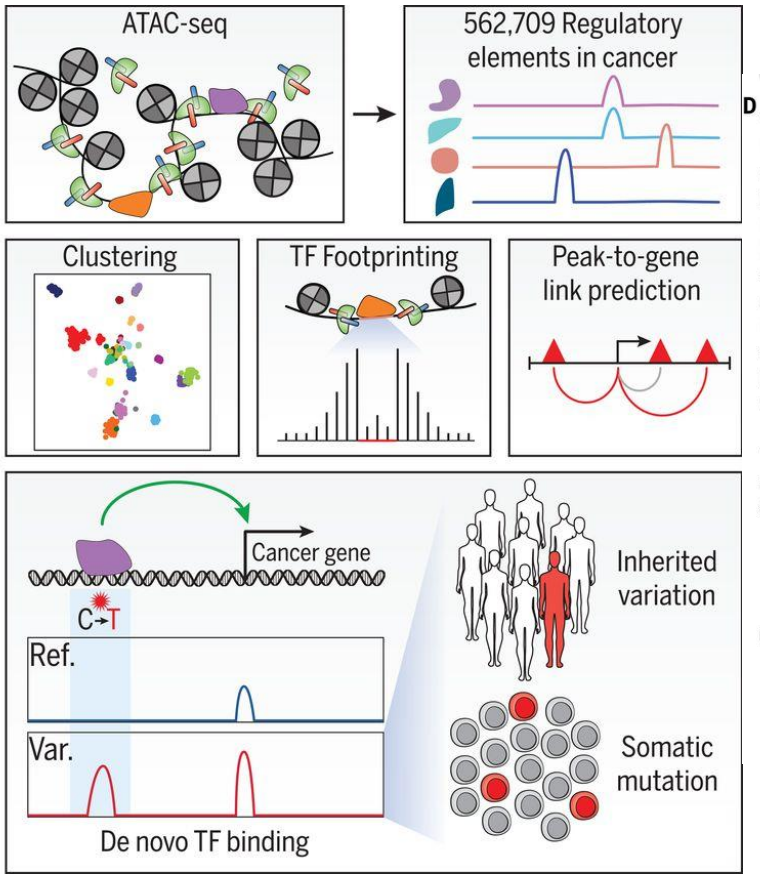
2. Combining ATAC-Seq with RNA-Seq:
https://link.springer.com/protocol/10.1007%2F978-1-4939-8618-7_15

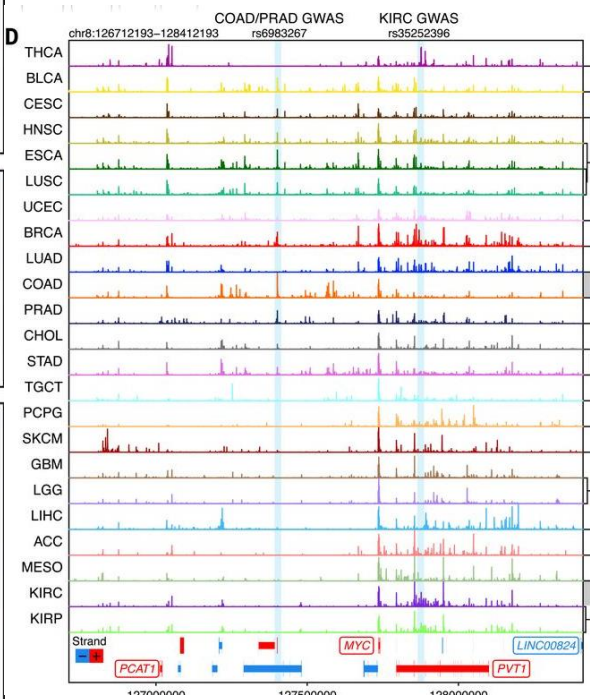
“The chromatin accessibility landscape of primary human cancers”

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



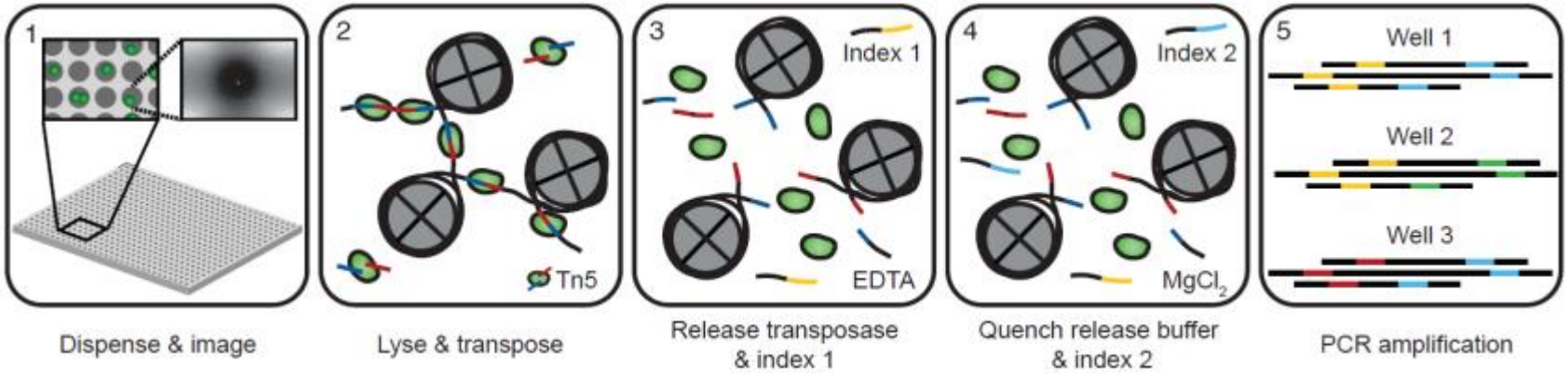




High-throughput *single-cell* ATAC-Seq
Toward Single-cell "Regulomics"?

- 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



Mezger et al, Nature Communications, 2018 and takarabio

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 (available in UTAP)
- Available downstream applications and new methods (single-cell ATAC-Seq)