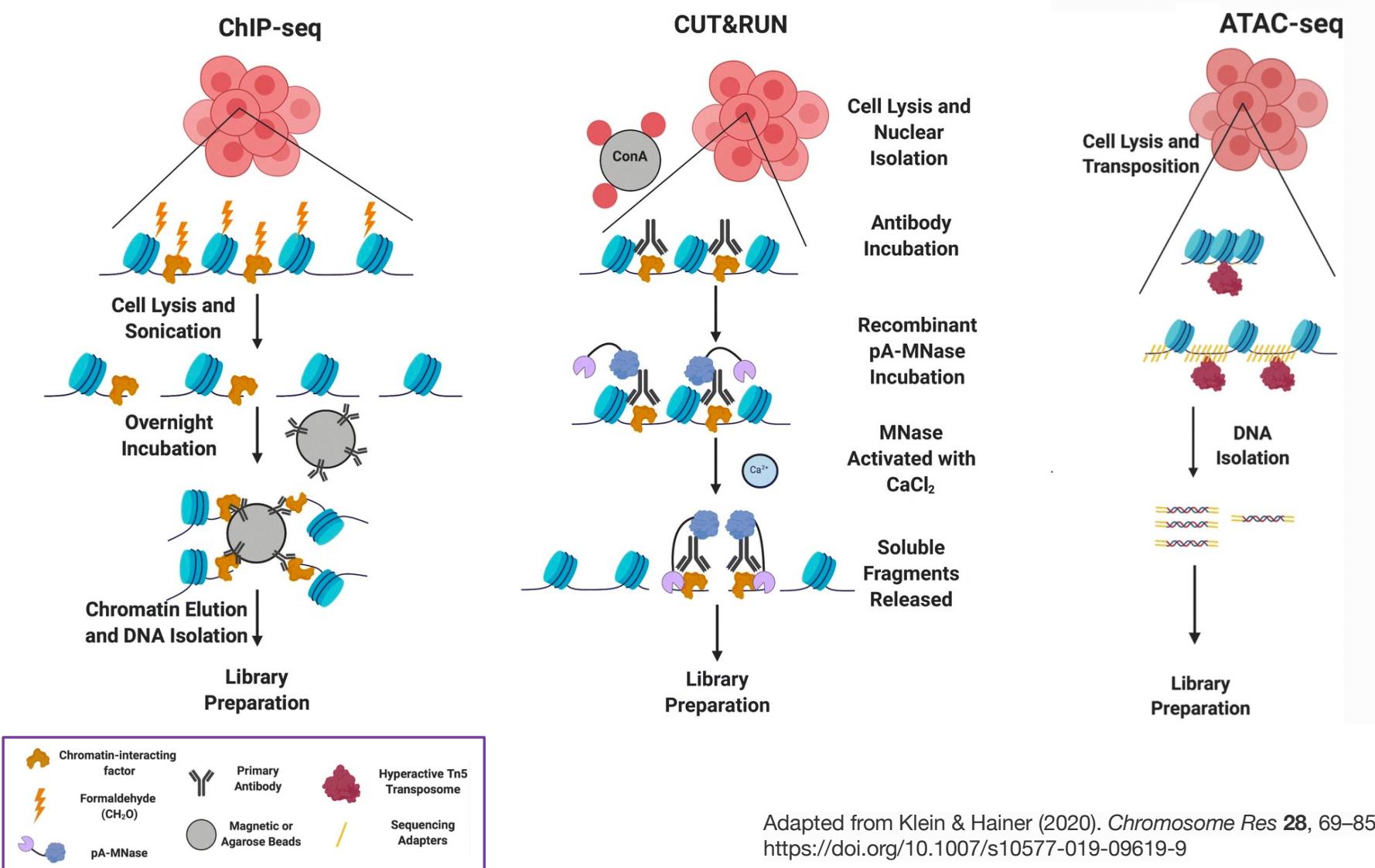
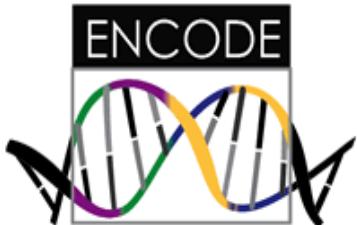
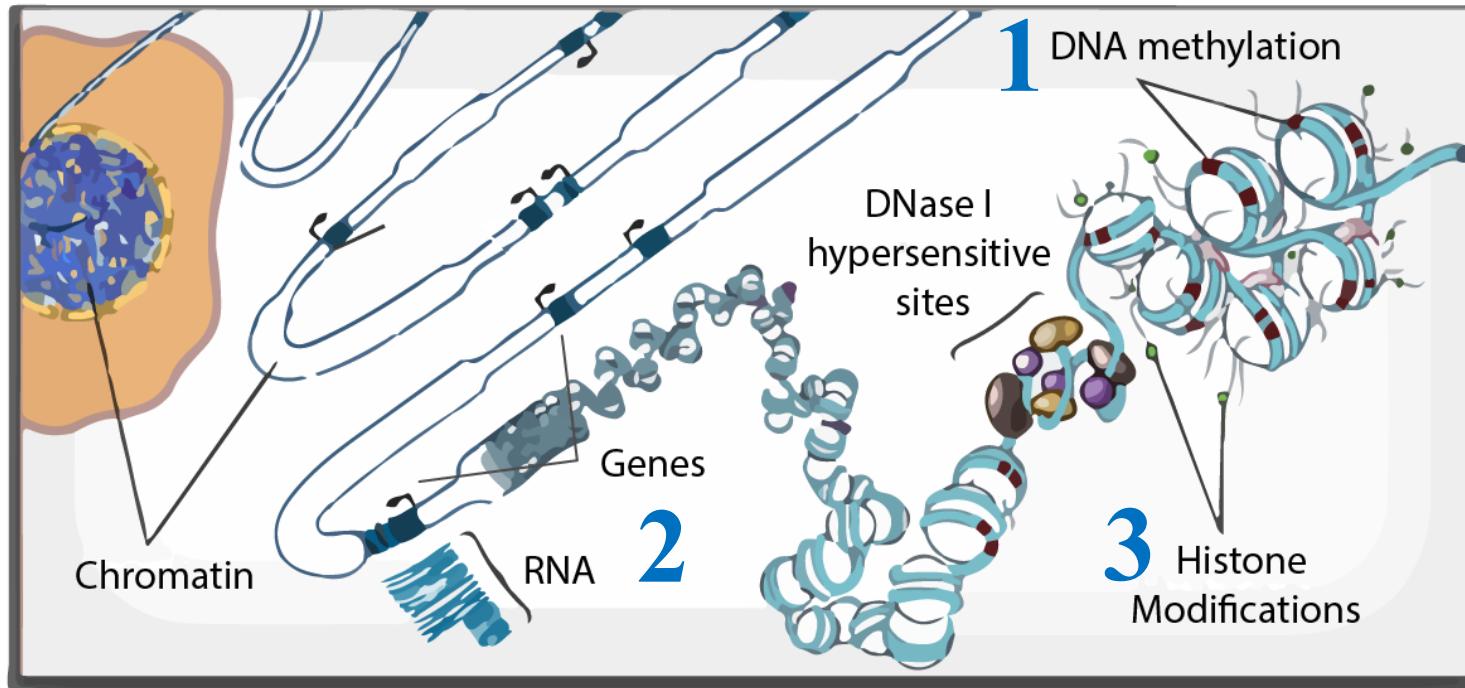


Chromatin Biology: ChIP-Seq, ATAC-Seq, and CUT&RUN



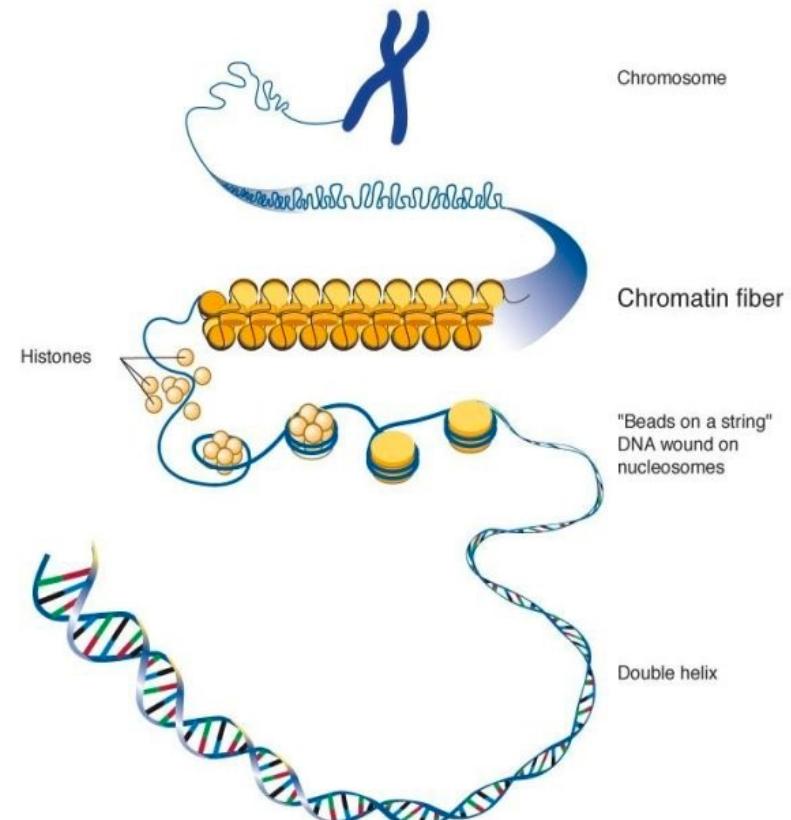
Adapted from Klein & Hainer (2020). *Chromosome Res* 28, 69–85.
<https://doi.org/10.1007/s10577-019-09619-9>

Epigenetic Regulatory Mechanisms



What is chromatin?

- **Chromatin:** a mixture of DNA and proteins that form the chromosomes found in the cells of humans and other higher organisms
- **Nucleosome:** 147 bp of DNA wound around 8 histone proteins (octamer) consisting of 2 copies each of the core histones (H2A, H2B, H3, H4)
- **Heterochromatin:** condensed chromatin
- **Euchromatin:** extended chromatin

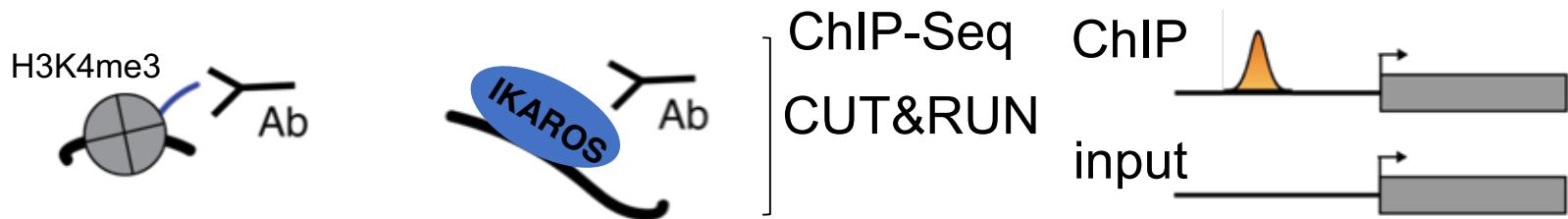


[https://www.creative-diagnostics.com/blog/index.php/
the-structure-and-function-of-chromatin/](https://www.creative-diagnostics.com/blog/index.php/the-structure-and-function-of-chromatin/)

But how do we study this?

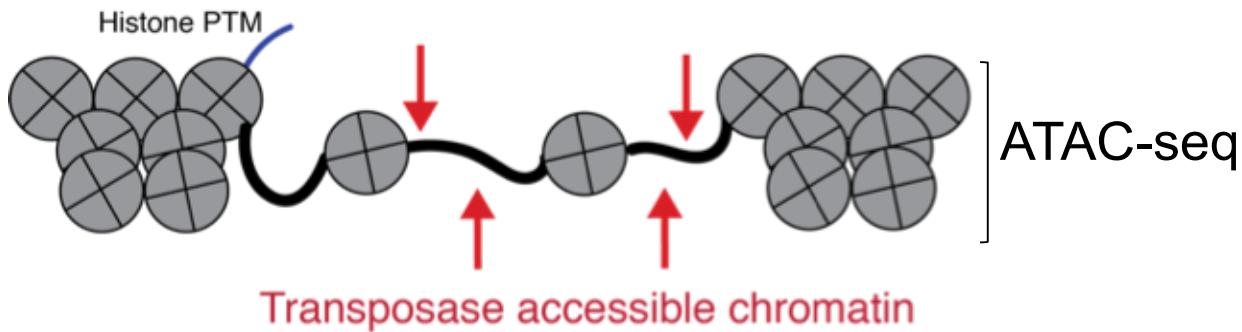
- Study the different states in chromatin
(euchromatin vs heterochromatin)
- Study epigenetic regulatory mechanisms that
can lead to altered gene expression

Today's Menu

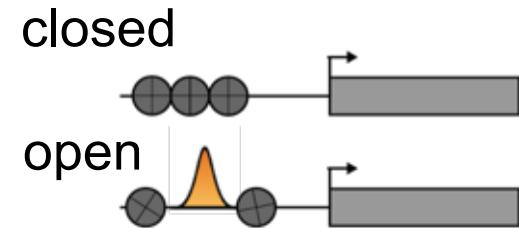


Where across the genome does this TF bind?

Is this gene “active” or in a “repressive” state?



Is this region “open” or “closed” for transcription?

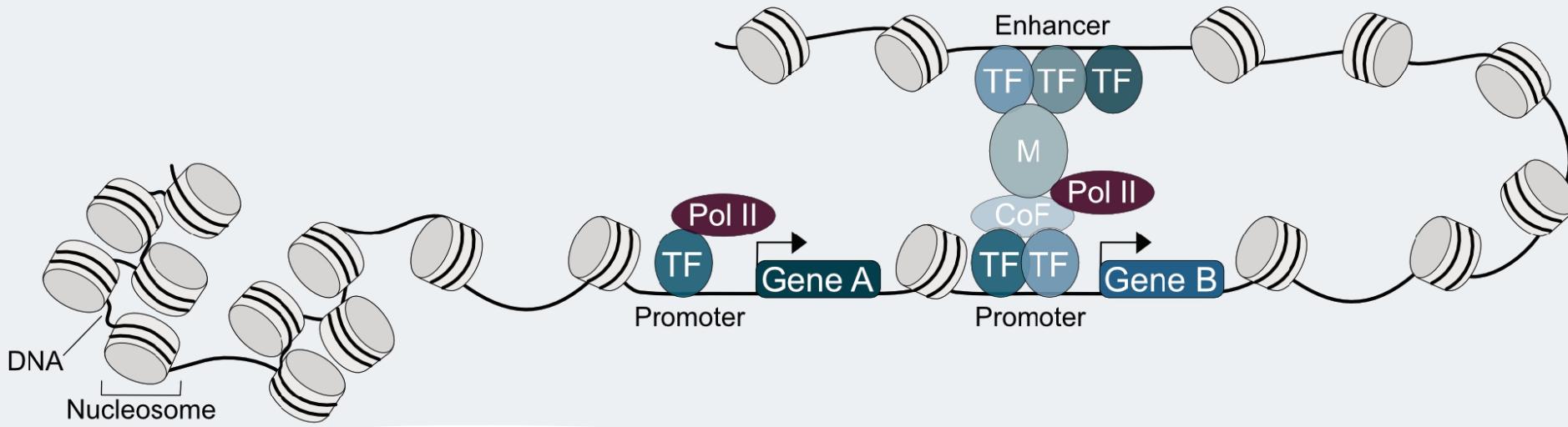


CLOSED

Transcription OFF

OPEN

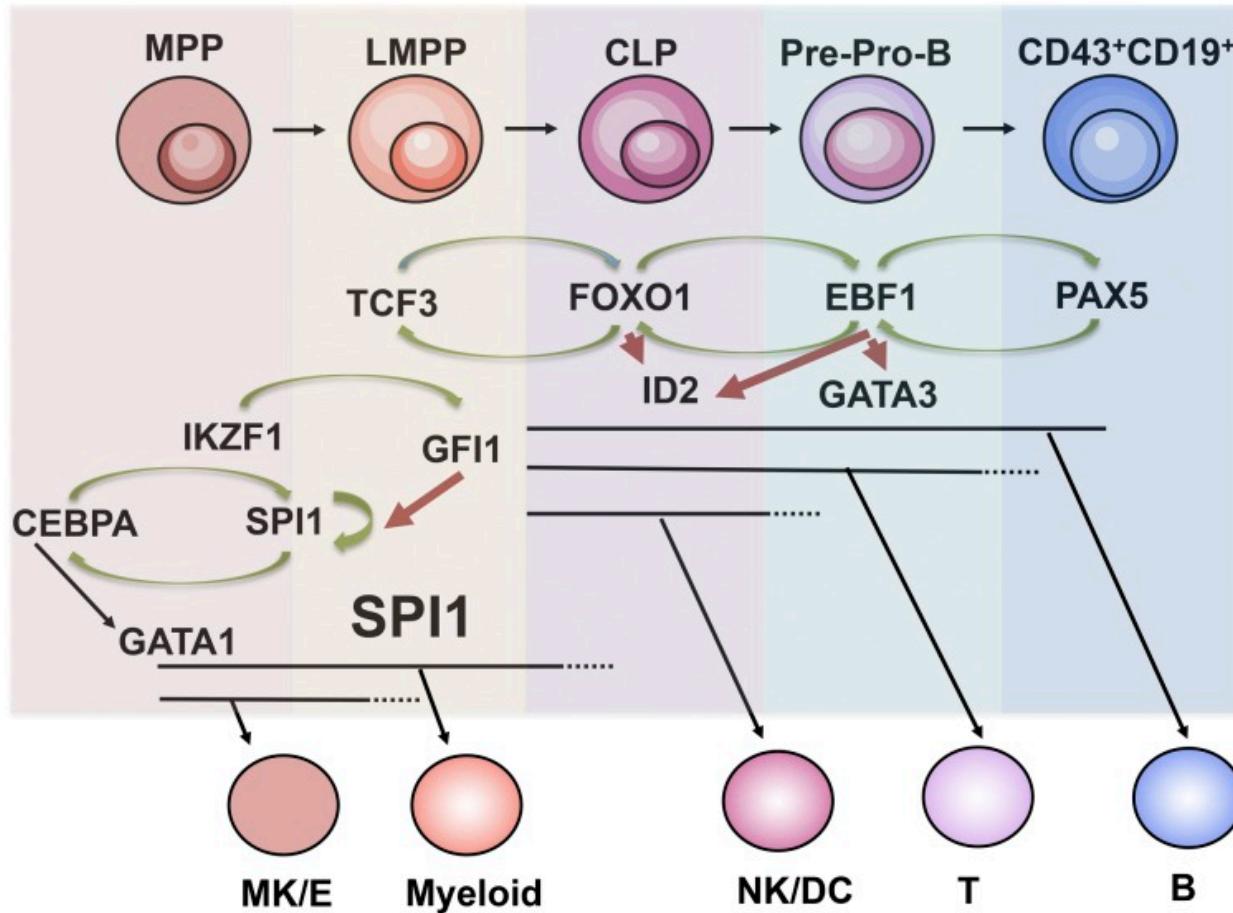
Transcription ON



Aim of ChIP- Seq application

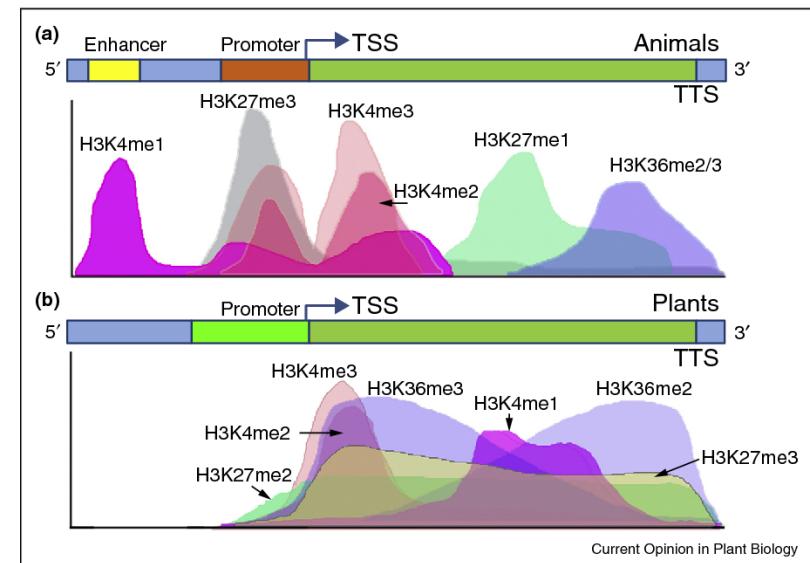
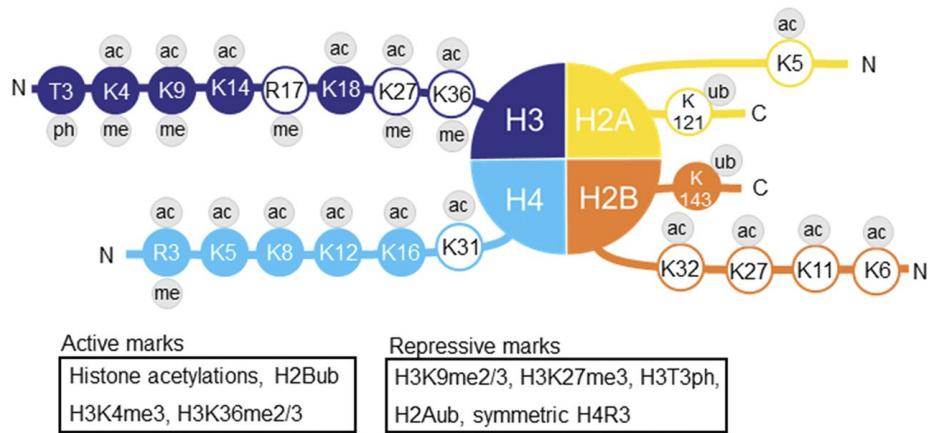
- Chromatin Immuno-Precipitation followed by Sequencing
- Characterize genome-wide DNA-protein interactions *in vivo*, for factors such as:
 - Binding of Transcription Factors (TFs) to regions in the genome
 - Post-translations modifications occurring on histone tails

TF binding during immune cell development



Complex transcriptional regulation

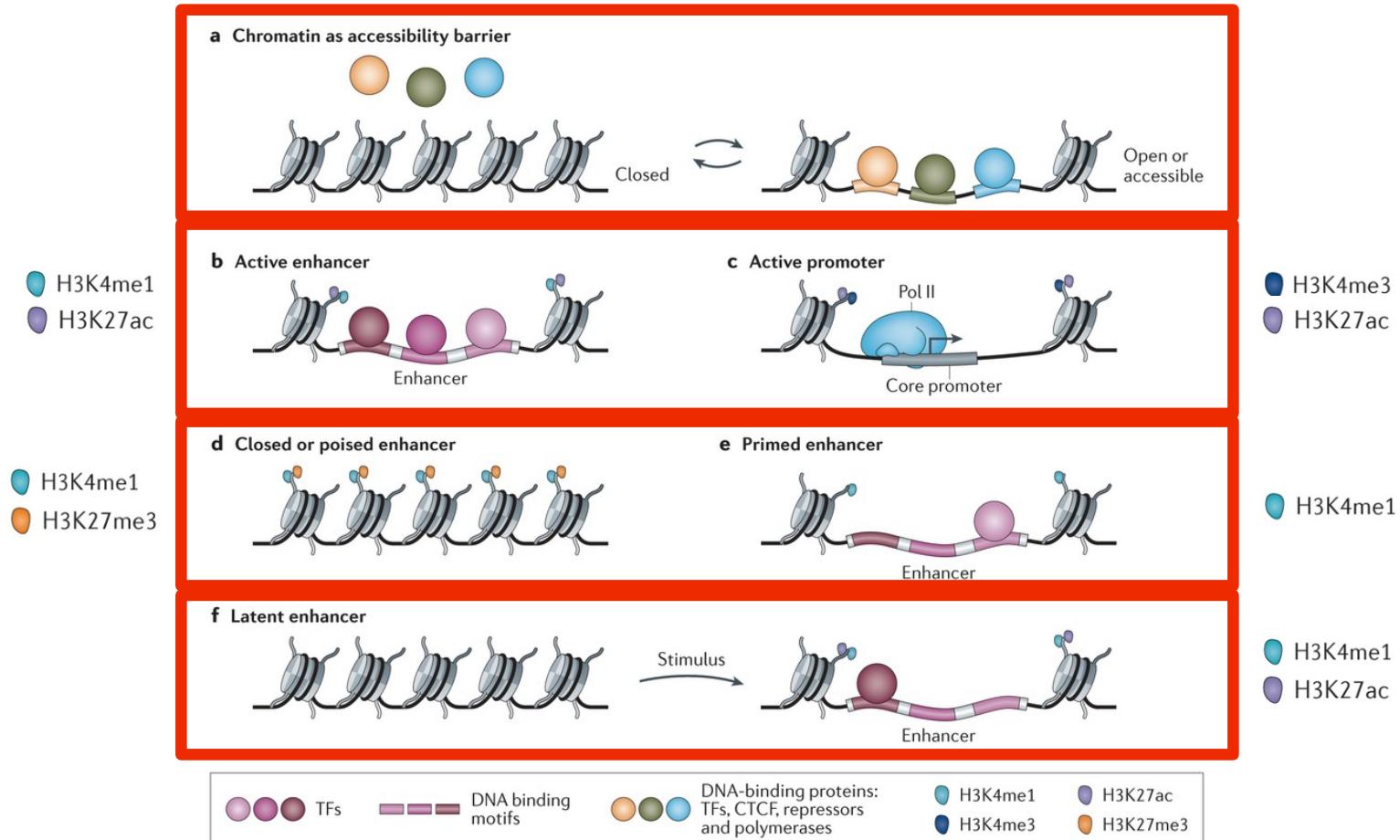
PTMs have different implications for gene transcription and are distributed at different regions across the gene



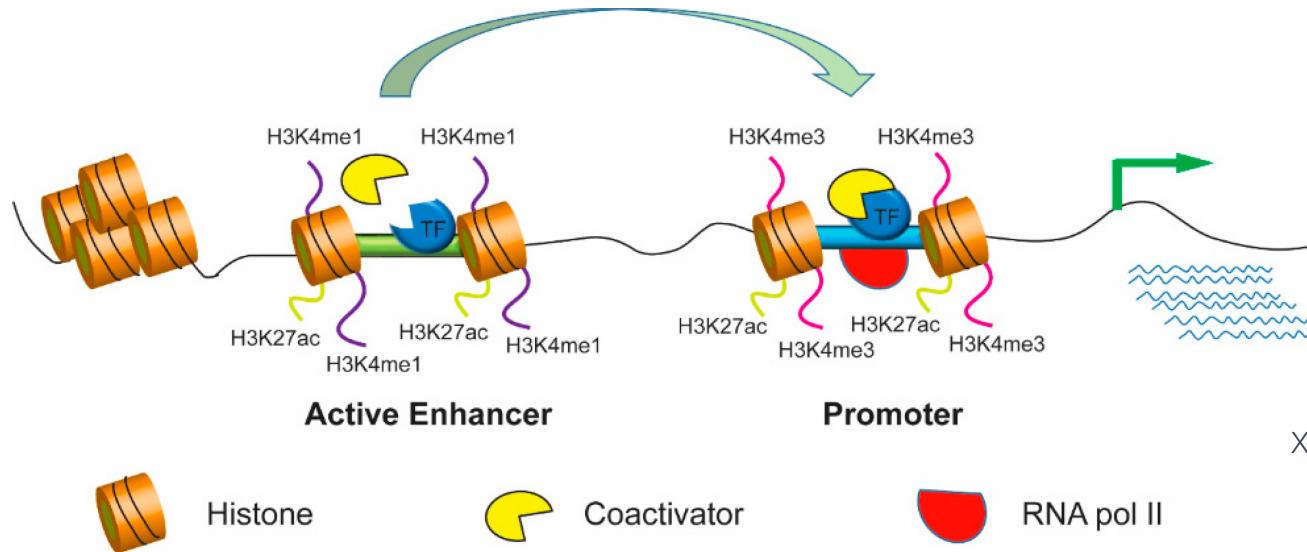
Histone proteins undergo post-translational modification (PTM) which impacts their interactions with DNA

There are at least nine different types of histone modifications have been discovered. In my work we focus on acetylation & methylation

Examples of common histone PTMs



Enhancers, repressors and cofactors can all regulate transcription

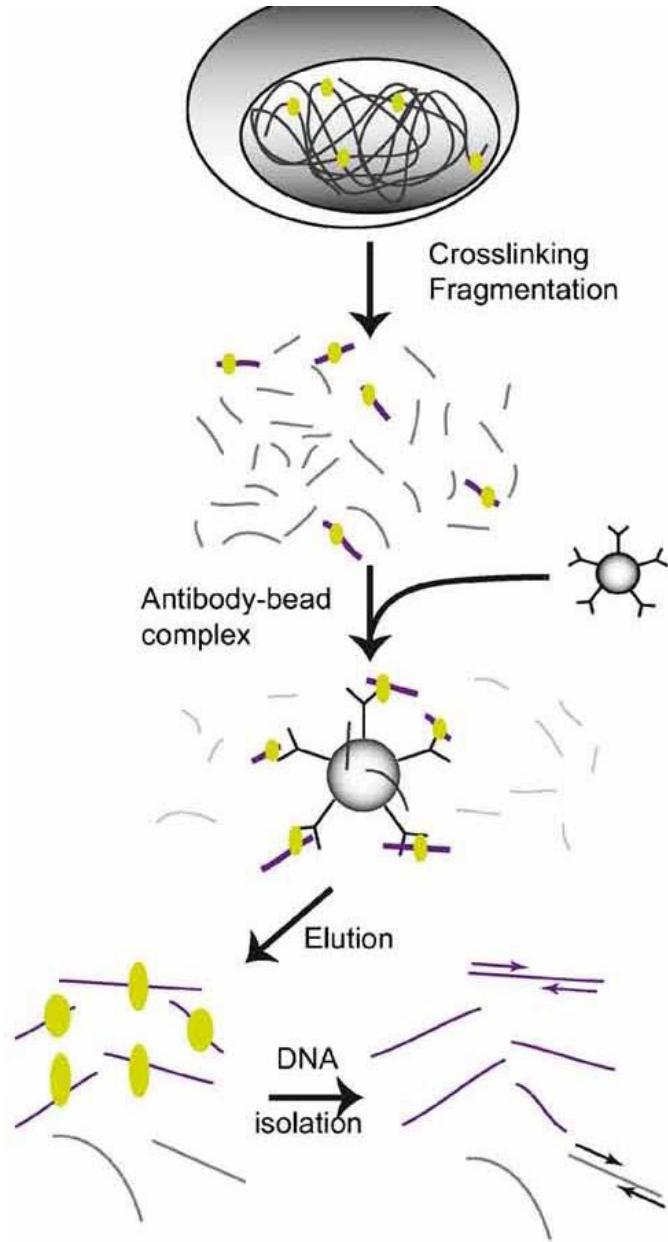


- Enhancers are DNA regulatory elements that activate transcription to a higher level
- Operate from a distance by forming chromatin loops that bring the enhancer and target gene into proximity
- Silencers reduce transcription from their target promoters

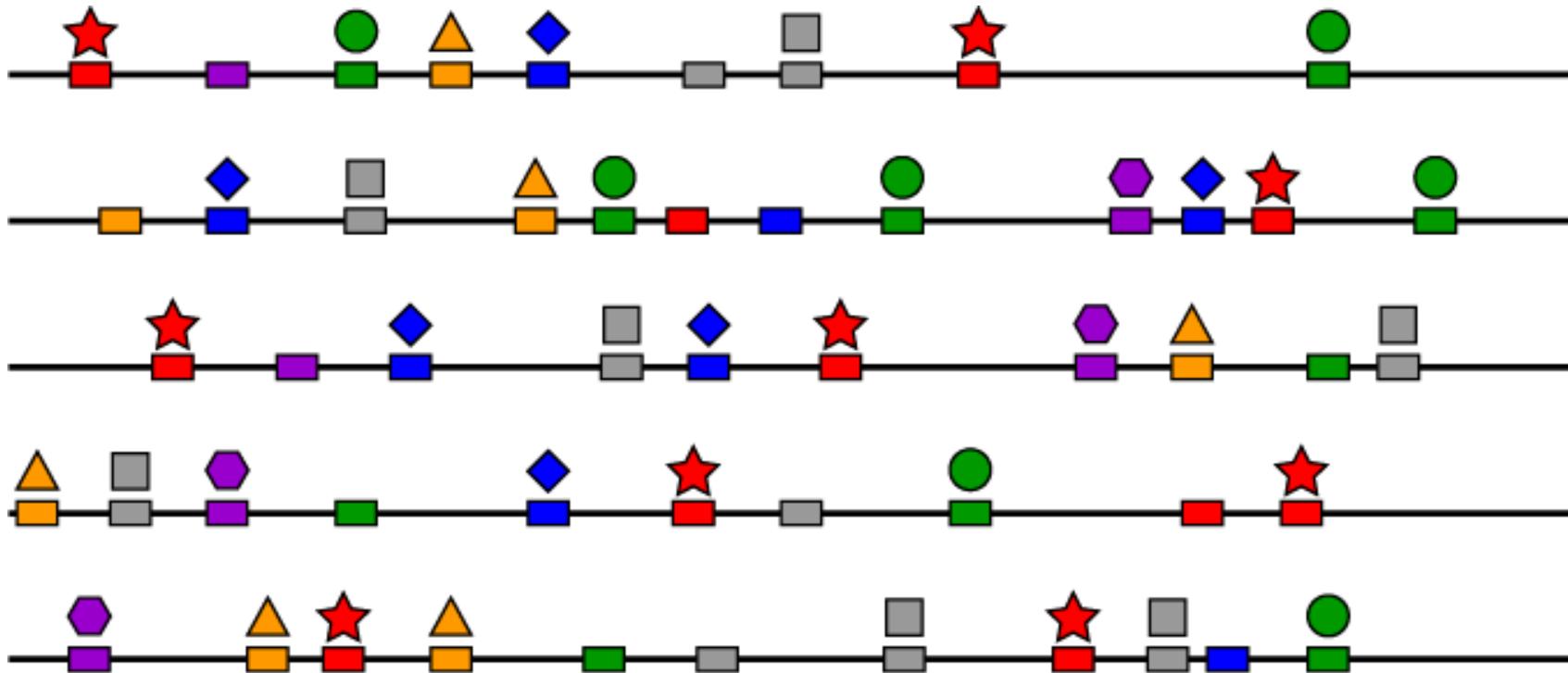
ChIP – How is it done?

ChIP is a technique that permits to

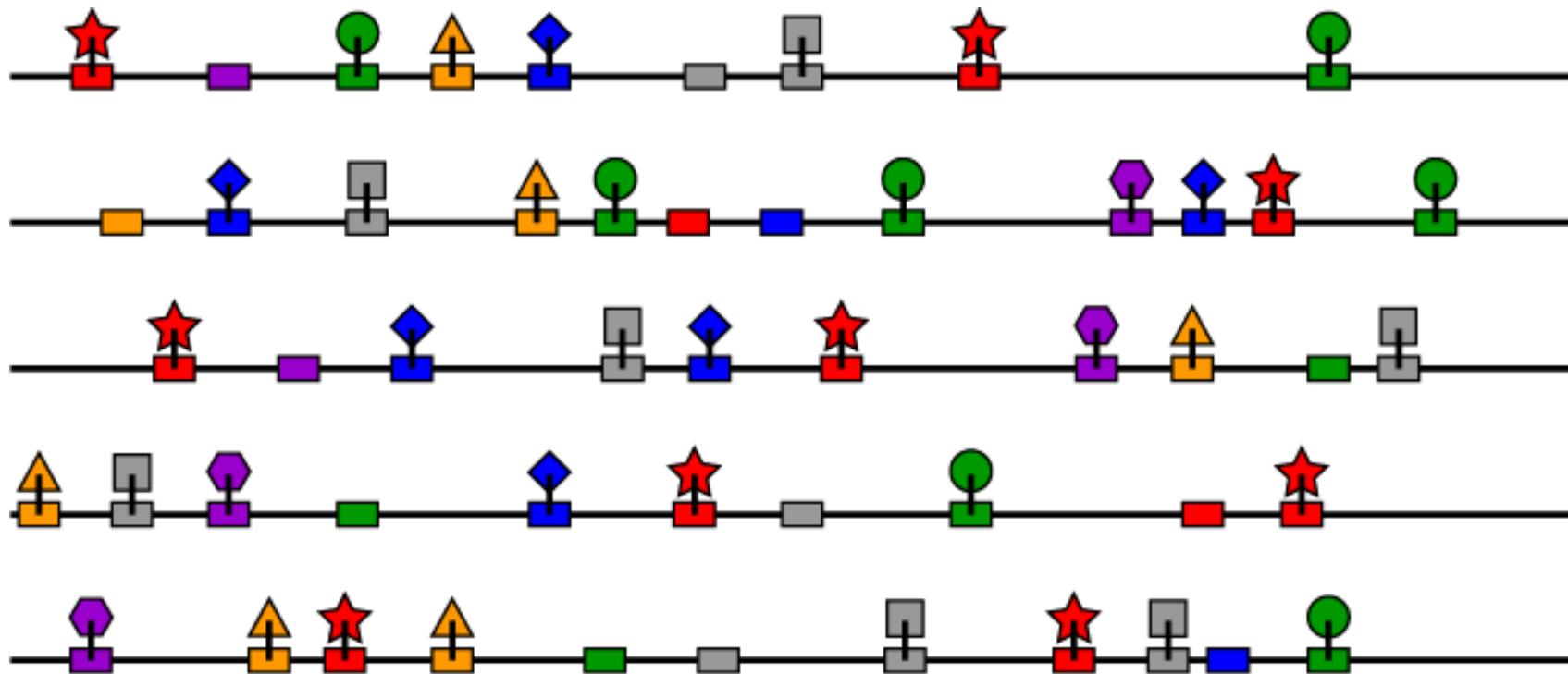
- “Freeze” the protein-DNA binding events inside the cell nucleus
- Use antibodies to extract the DNA bound by a specific protein



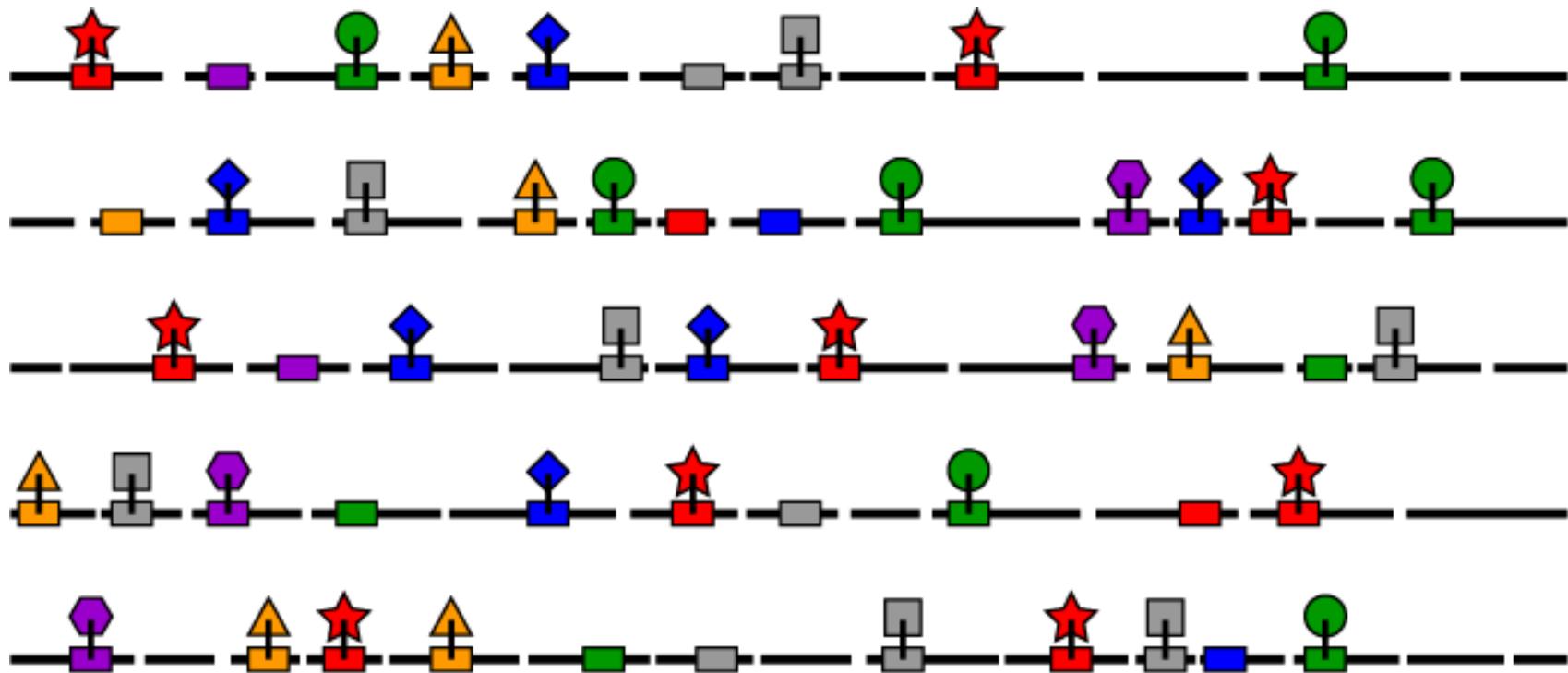
Chromatin ImmunoPrecipitation (ChIP)



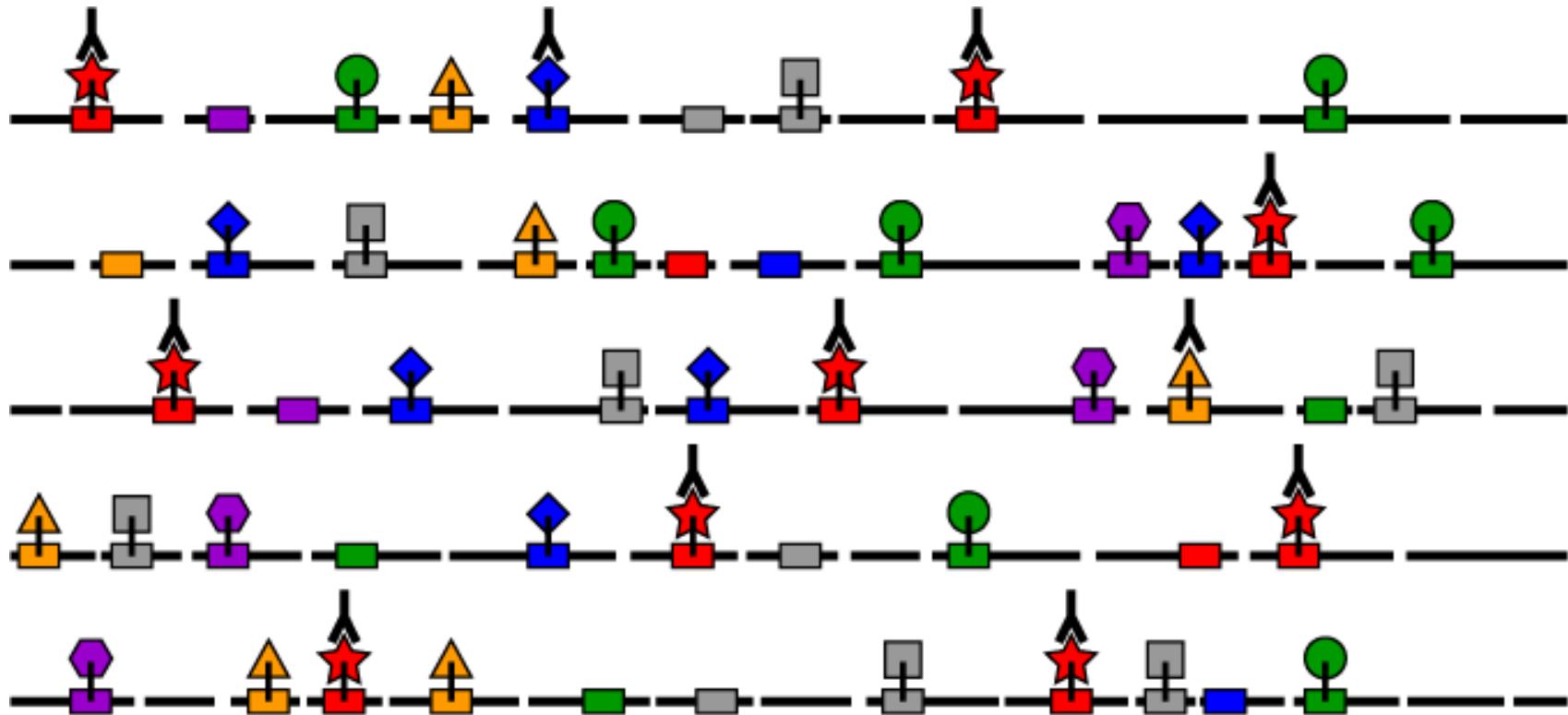
TF/DNA Crosslinking *in vivo*



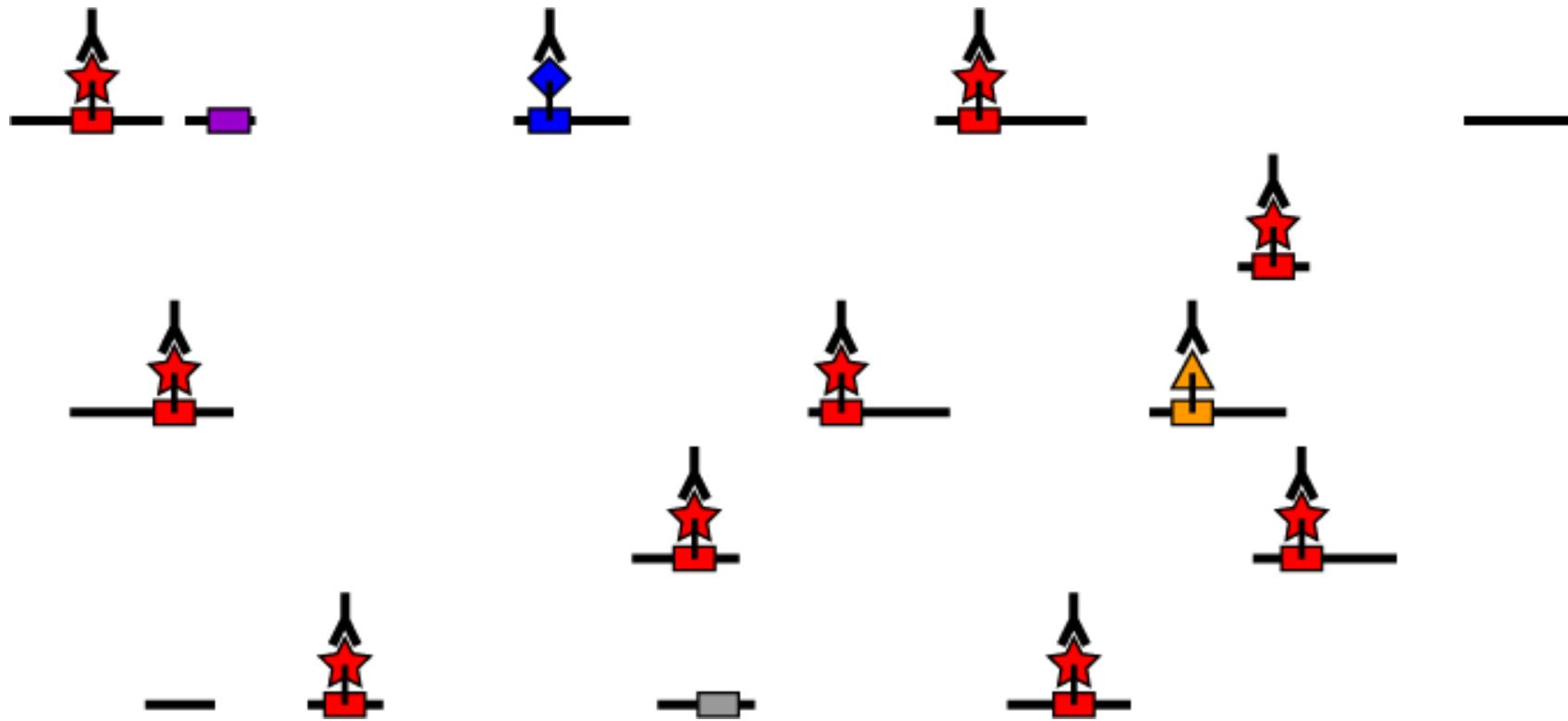
Sonication (~200bp)



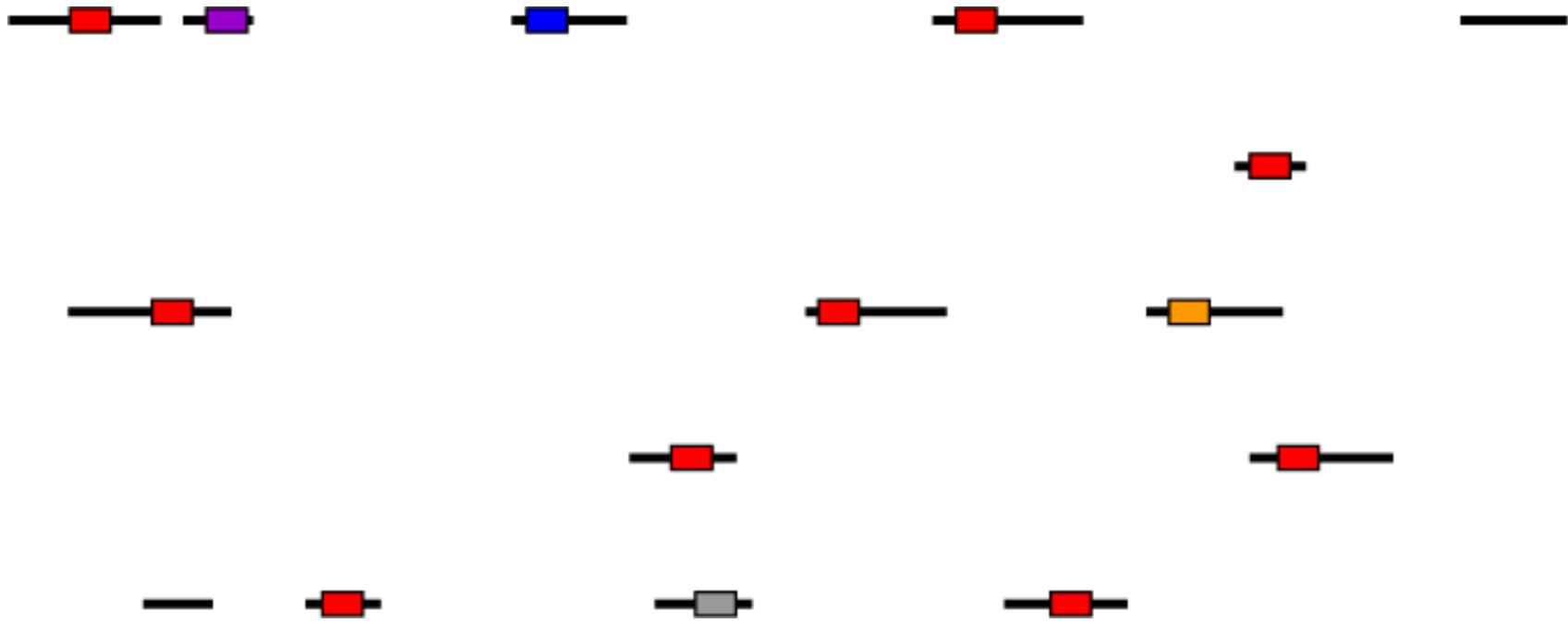
TF-specific Antibody



Immunoprecipitation



Reverse Crosslink and DNA Purification



Followed by library preparation (PCR and adapter ligation), then sequencing

CUT&RUN

- Cleavage Under Targets & Release Using Nuclease (CUT&RUN) is a revolutionary genomic mapping strategy developed by the group of Dr. Steven Henikoff.
- A fusion of Protein A to Micrococcal Nuclease (pA-MNase) is recruited to selectively cleave antibody-bound chromatin

CUT&RUN is also used to study various targets including histone PTM and other protein/DNA interactions

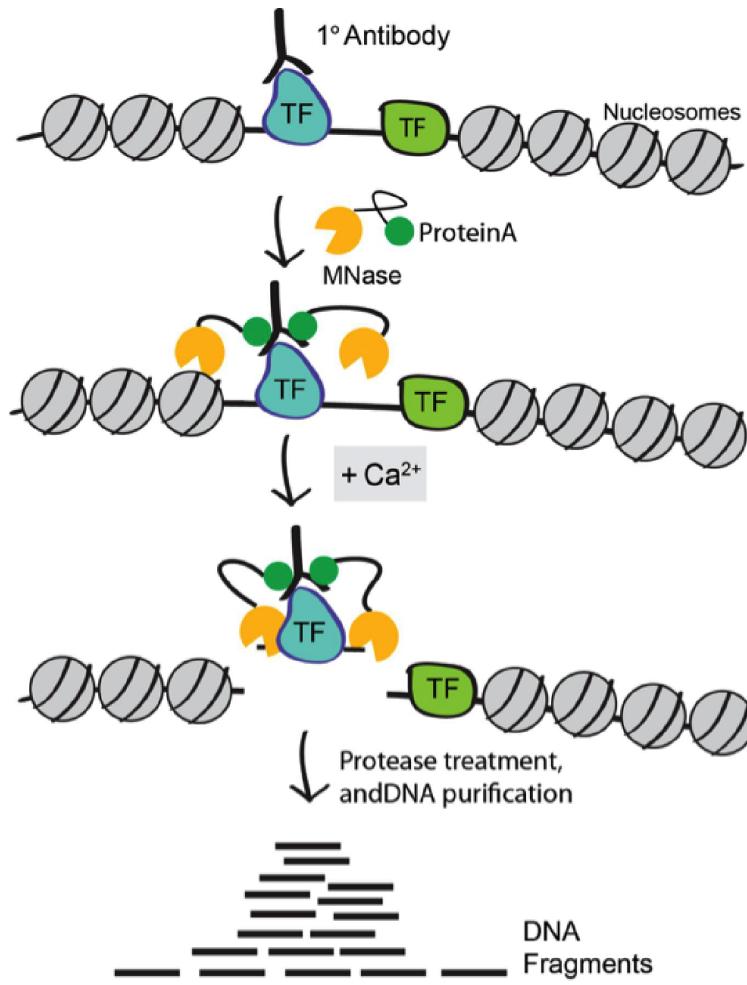


Figure 1. CUT&RUN schematic (see text for details).

Result of ChIP-Seq (and CUT&RUN): genome-wide profiles

Active promoters

H3K4me3, H3K9Ac

Active enhancers

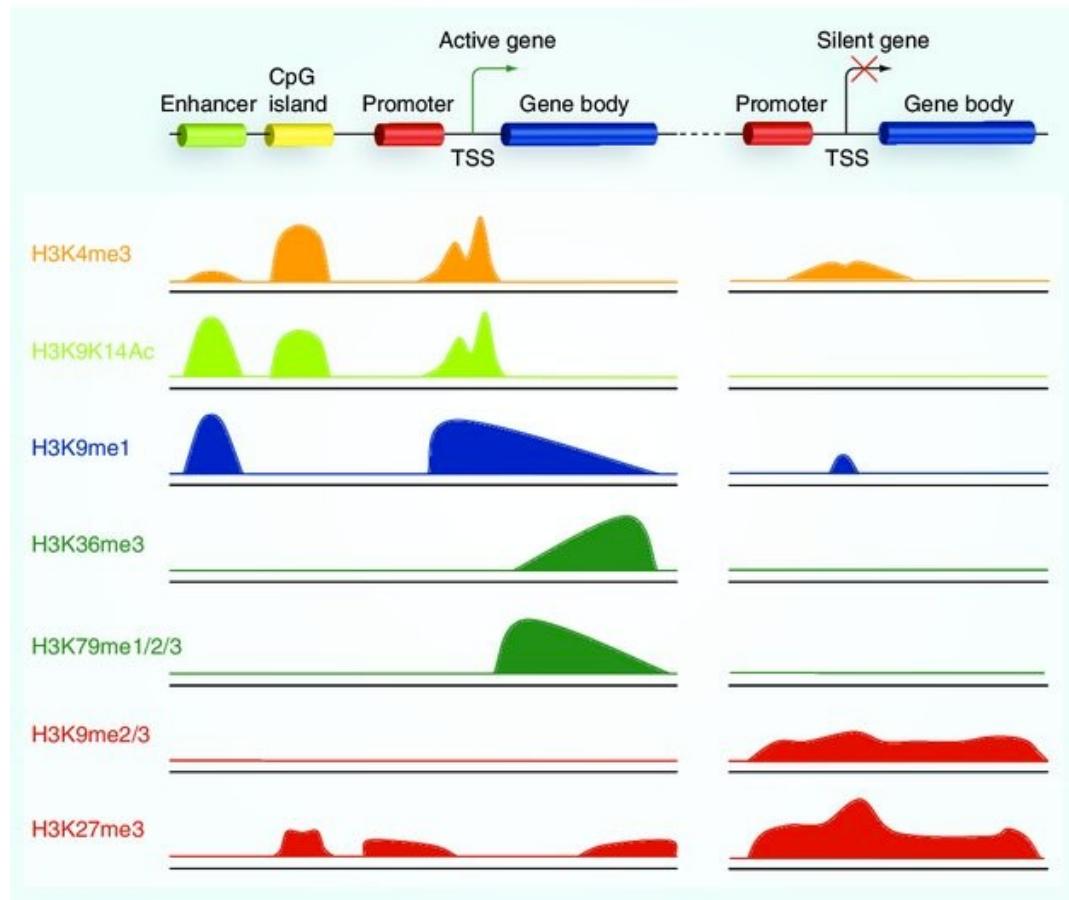
H3K27Ac, H3K4me1

Repressors

H3K9me3, H3K27me3

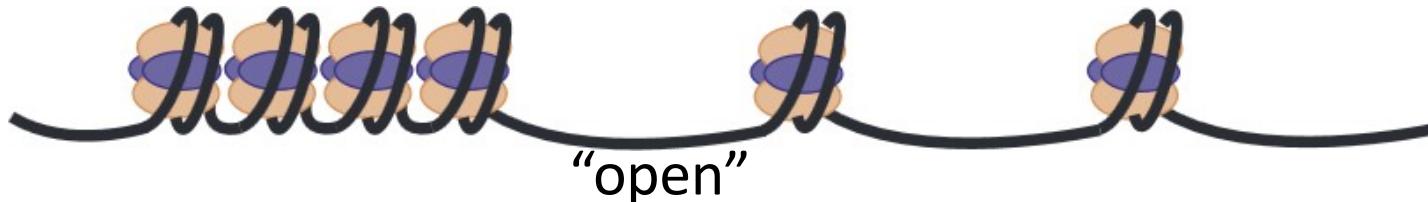
Transcribed gene bodies

H3K36me3



ATAC-Seq =

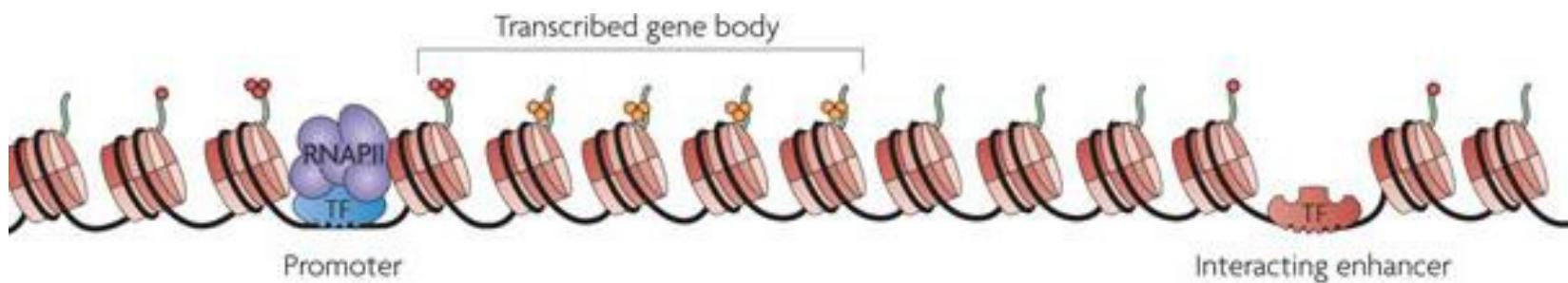
Assay for Transposase-Accessible Chromatin using sequencing



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

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)



ATAC-Seq was first described at the



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

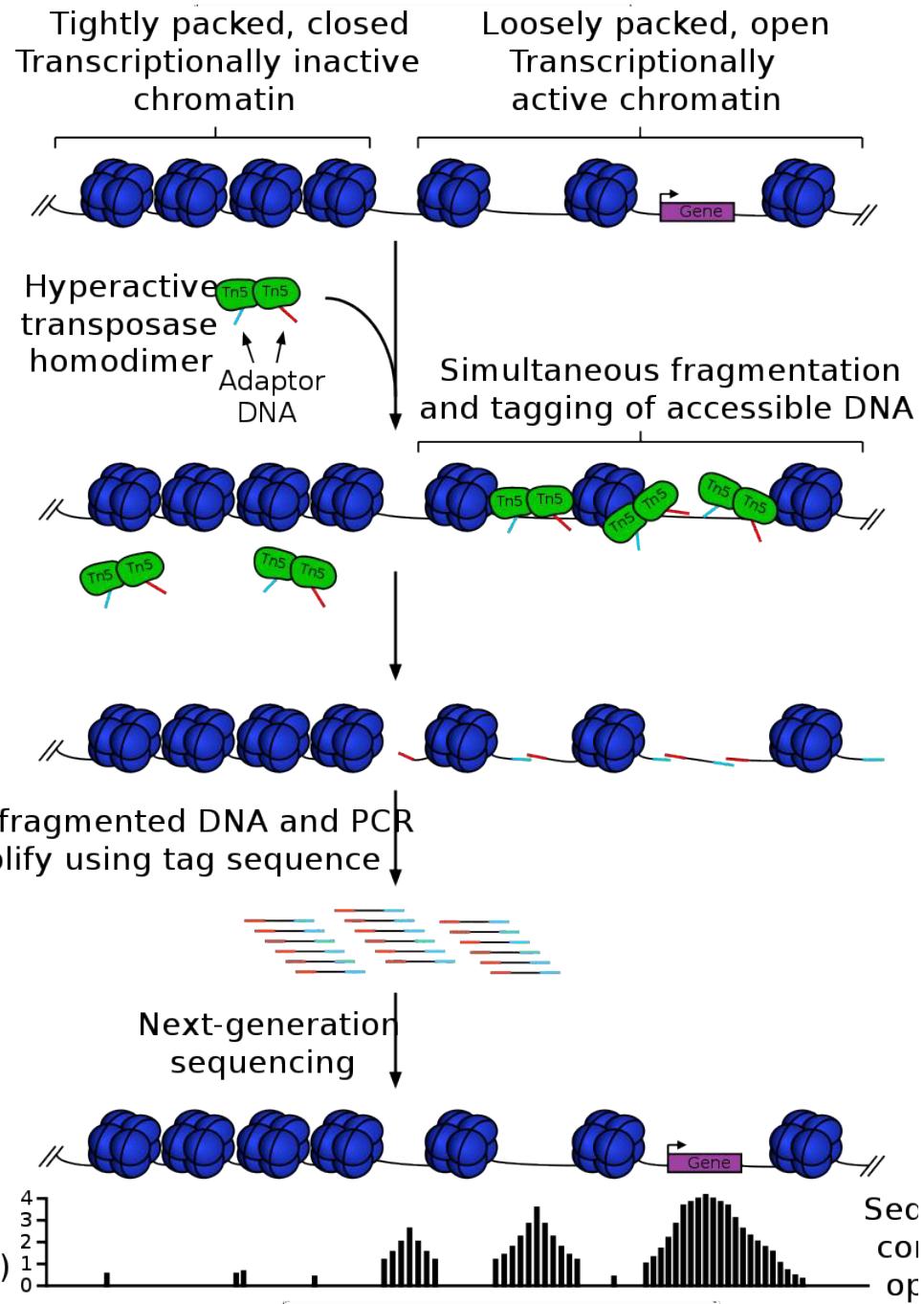
Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang & William J Greenleaf

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 assay steps

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



Profiles of open and closed chromatin



Maps open chromatin TF occupancy
nucleosome occupancy

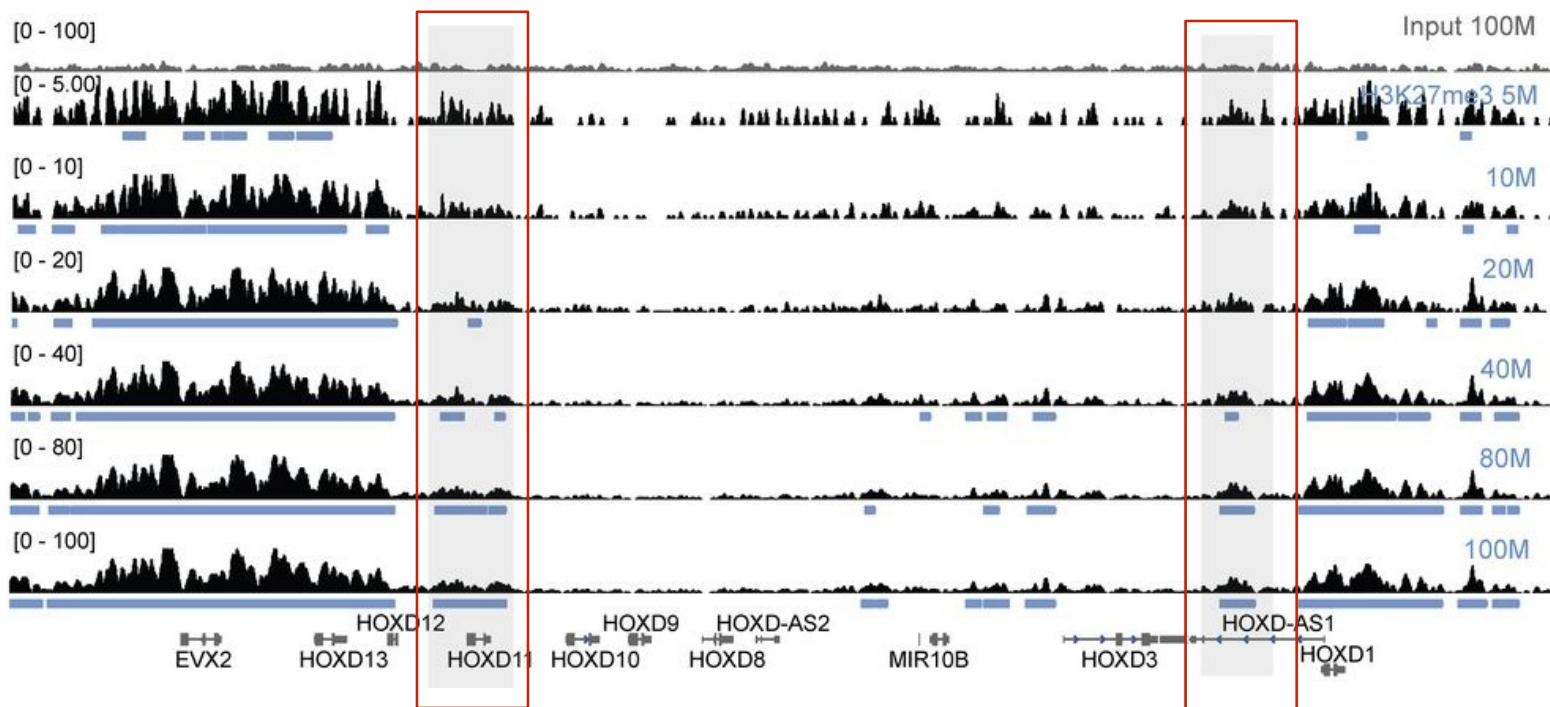
Many considerations when generating genome-wide datasets

Many differences!

- i. Cell # / sample prep / amount of starting material
- ii. Time to complete
- iii. Antibody quality
- iv. Addition of positive/negative controls
- v. Sequencing depth
- vi. Variation in data processing steps and programs used

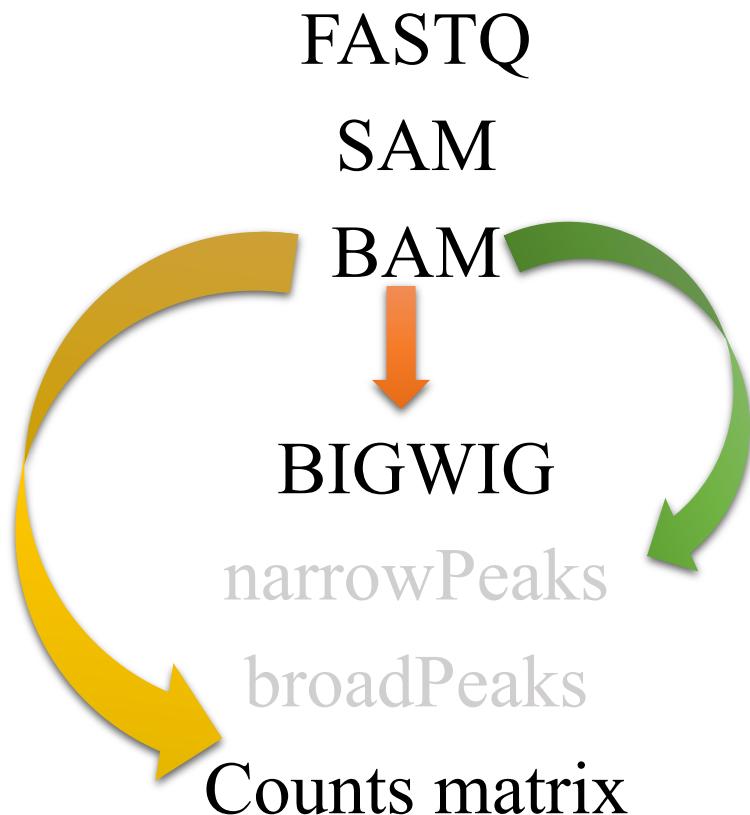
Impact of sequencing depth

H3K27me3



Adapted from Jung et al (2014). NAR.

File outputs of Genomic Datasets



Count versus genomic coordinate

- ❖ RNA-Seq – the idea here is to understand the gene counts per sample

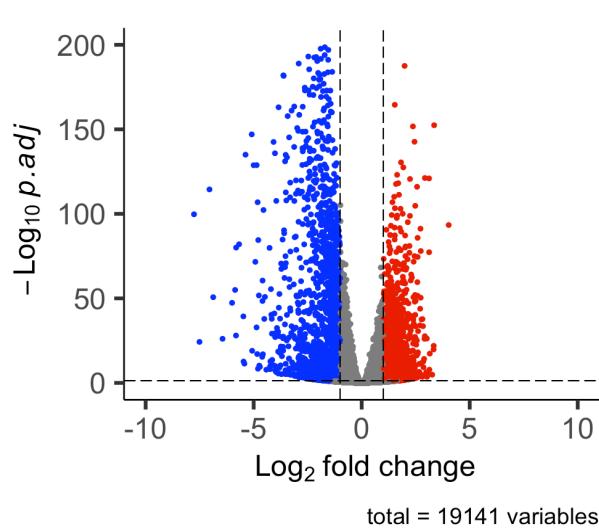
countData

gene	ctrl_1	ctrl_2	exp_1	exp_1
geneA	10	11	56	45
geneB	0	0	128	54
geneC	42	41	59	41
geneD	103	122	1	23
geneE	10	23	14	56
geneF	0	1	2	0
...
...
...

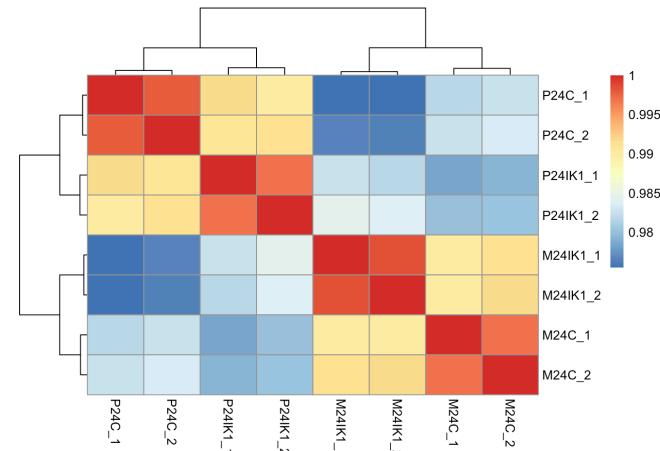
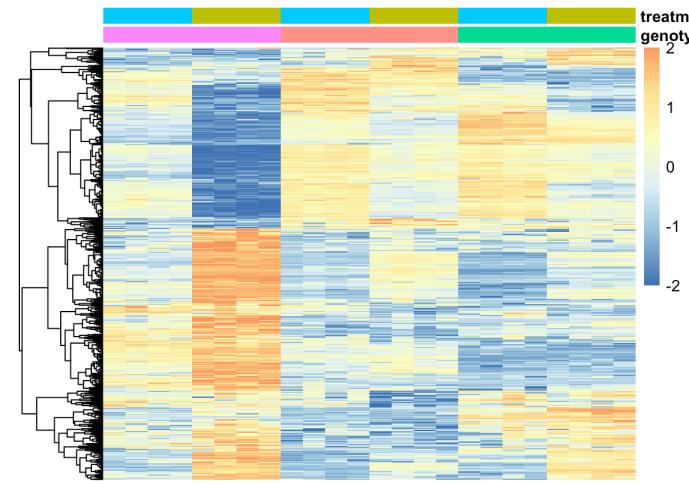
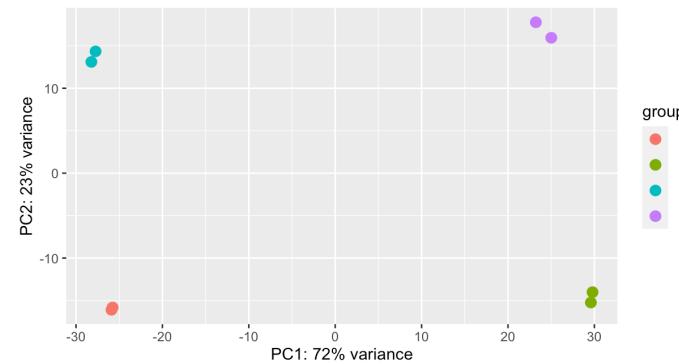
colData

id	treatment	sex
ctrl_1	control	male
ctrl_2	control	female
exp_1	treatment	male
exp_2	treatment	female

Sample names:
ctrl_1, ctrl_2, exp_1, exp_2

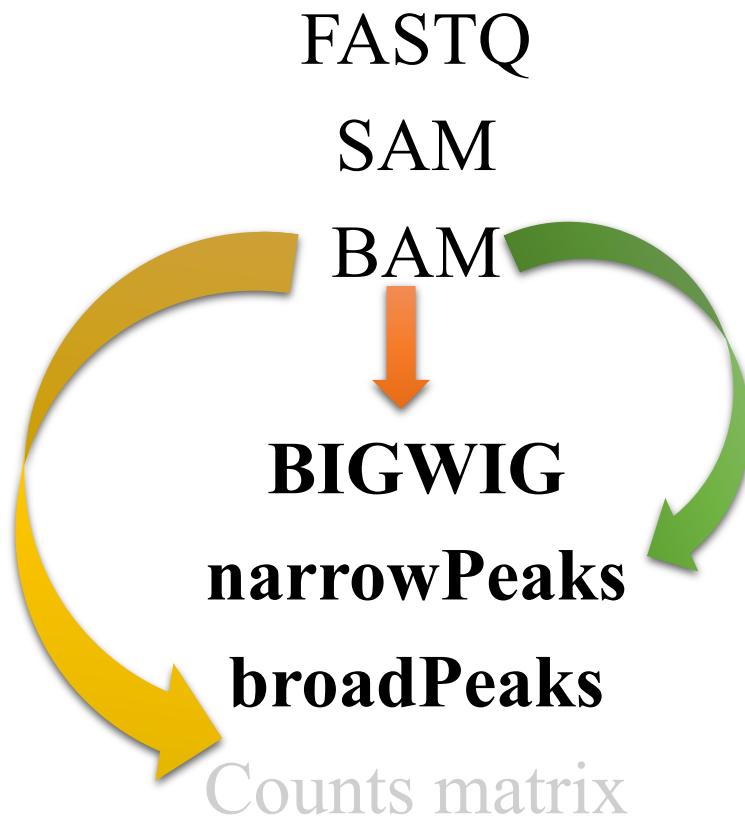


NS
-Log10Q
Signif. down-regulated
Signif. up-regulated
log2FoldChange



Typical Outputs for RNA-Seq

File outputs of Genomic Datasets



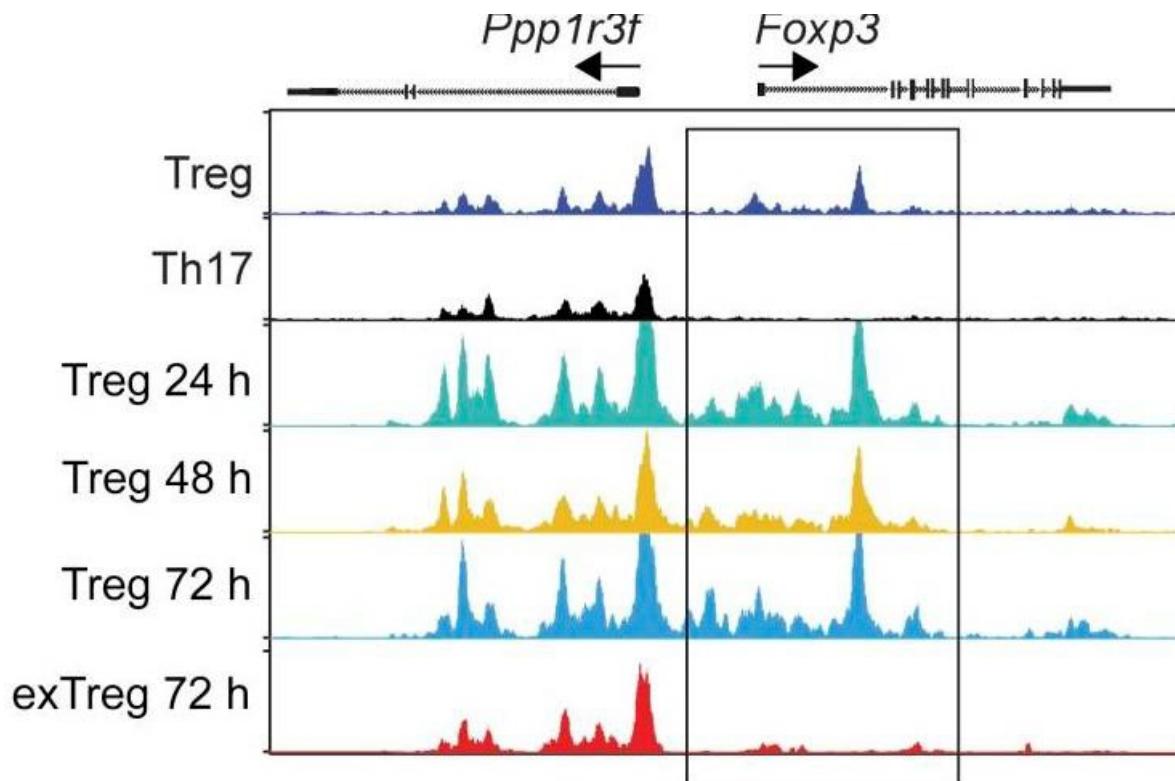
Count versus genomic coordinate

- ❖ CUT&RUN, ChIP-Seq, ATAC-Seq – the idea is to learn where is this target (TF or PTM) located across the genome? What is the location?
- Fundamentally, this location as **chromosome:start:end (peaks files)**

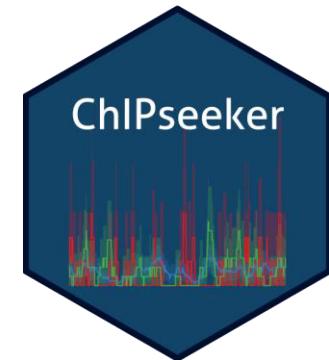
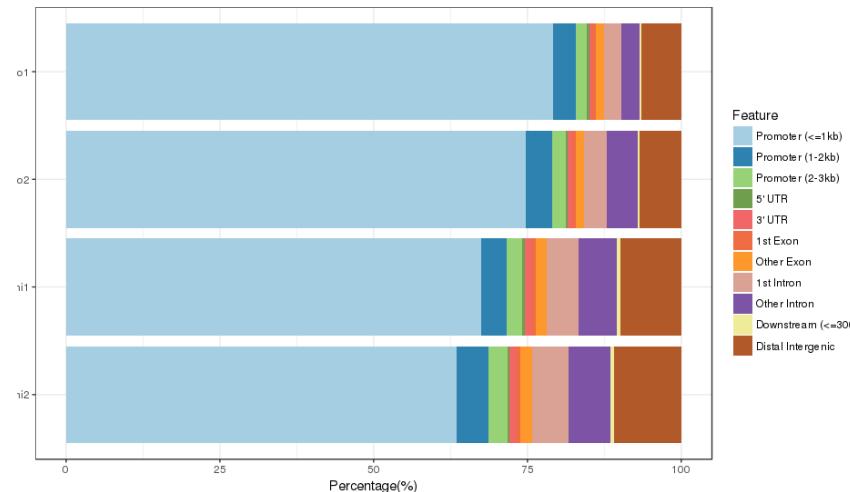
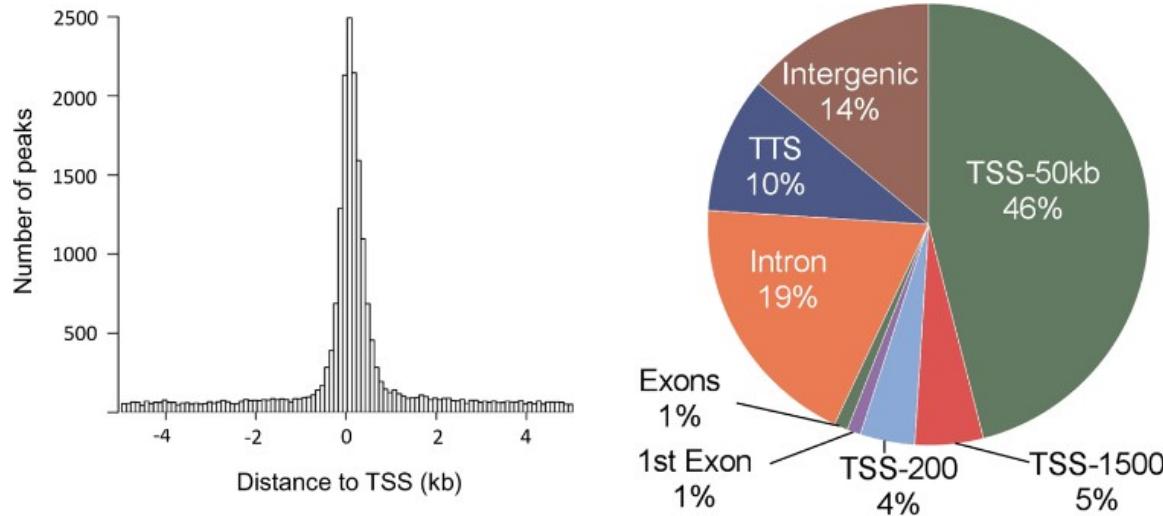
chr7	157103058	157103313	.	1000	.
chr2	149335745	149335997	.	1000	.
chr2	45030052	45030295	.	1000	.
chr20	55807184	55807446	.	1000	.
chr6	31139711	31139982	.	1000	.
chr15	71496031	71496283	.	1000	.
chr12	130255896	130256146	.	1000	.
chr14	42805182	42805428	.	1000	.
chr1	203488575	203488821	.	1000	.

Downstream analysis

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

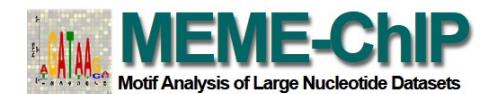
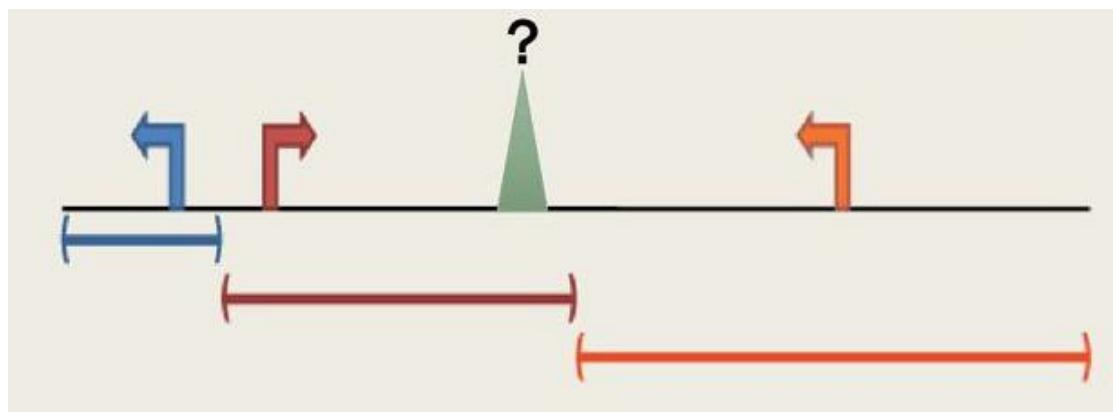


B. Assignment of peaks to genomic regions:

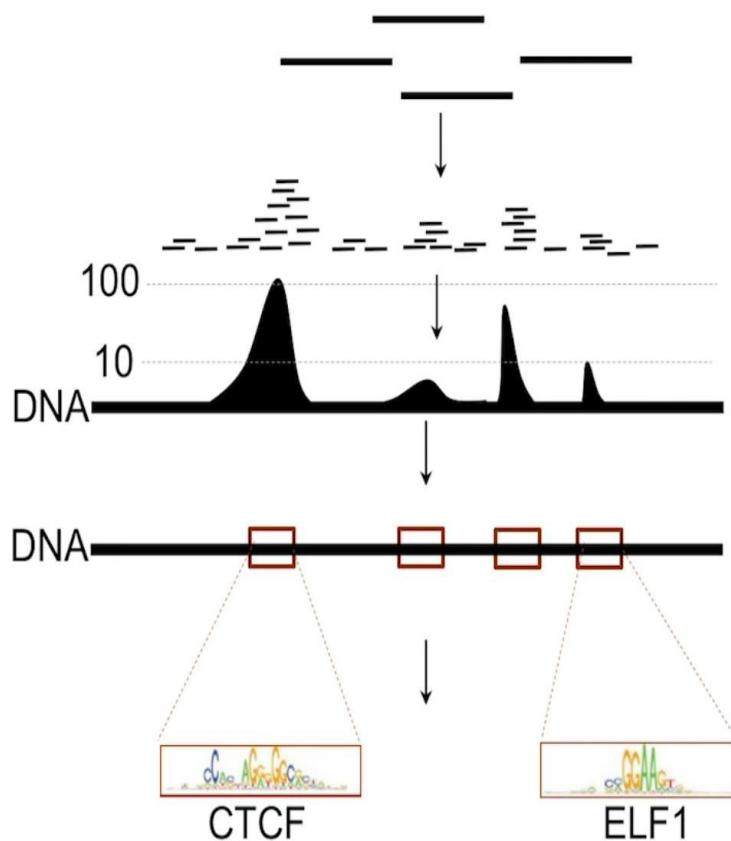


C. Peaks annotation and functional enrichment

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

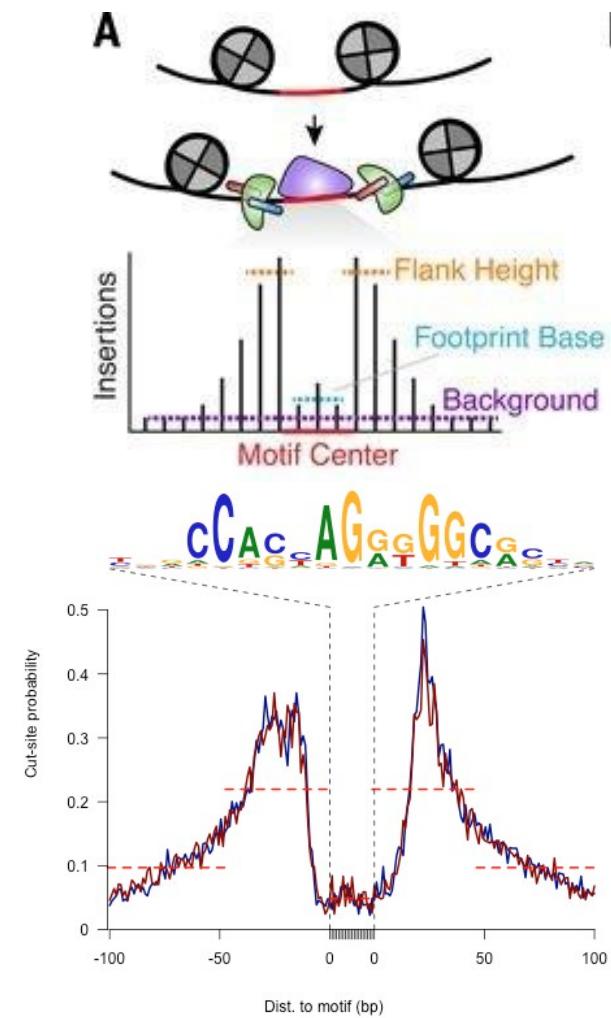


Motif discovery/enrichment

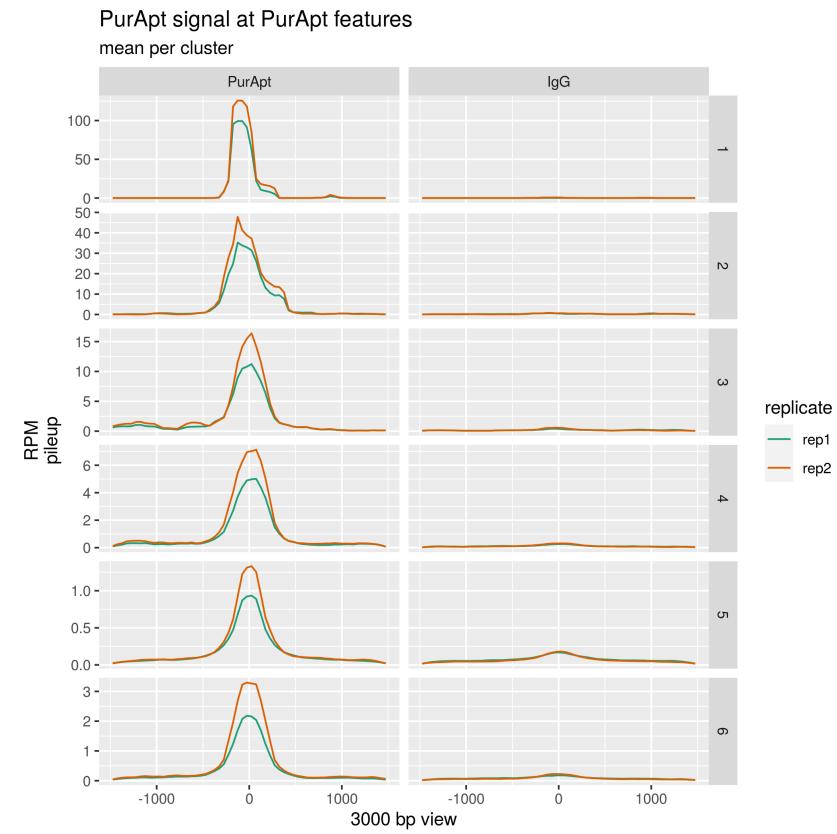
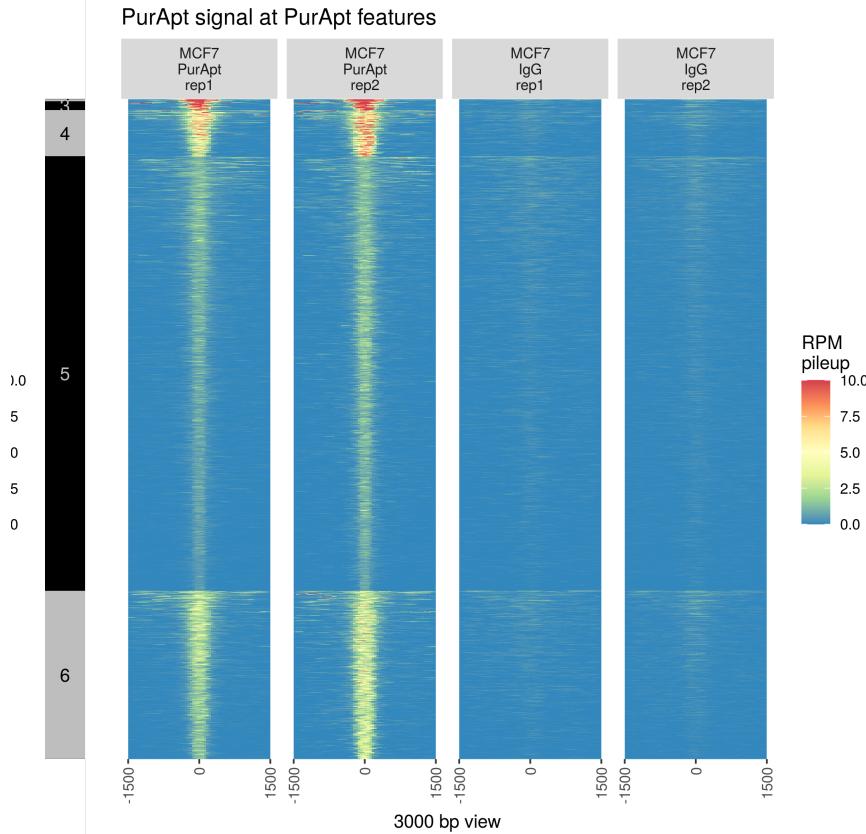


Ricardo N Ramirez, Harvard Medical School

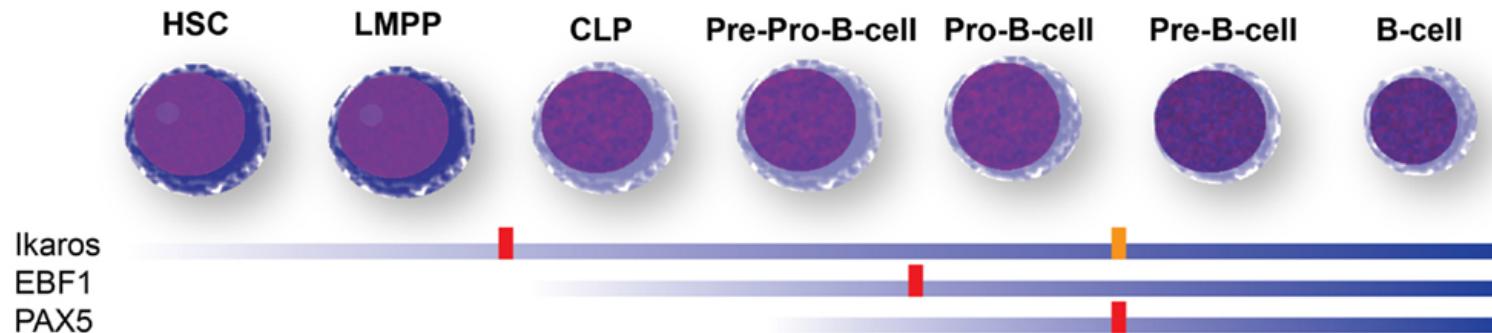
**Infer footprints
of DNA-protein binding
Requires deeper sequencing**



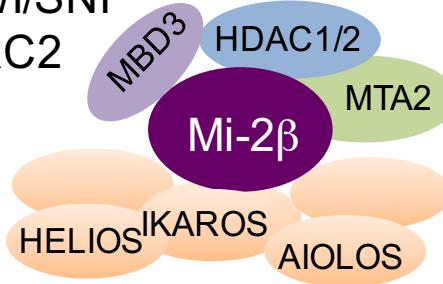
Signal heatmaps



Now let's look at some data!



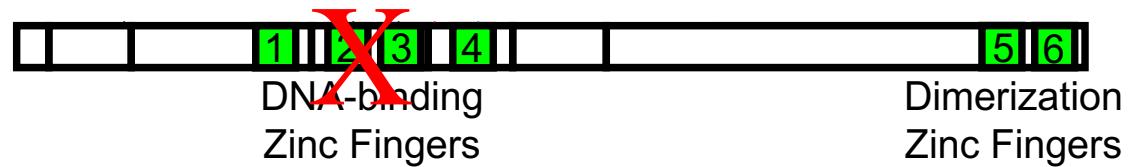
- Oligomerization
- Chromatin modifier complexes
 - Mi-2/NuRD
 - SWI/SNF
 - PRC2



Let's look at some data!



Mutant

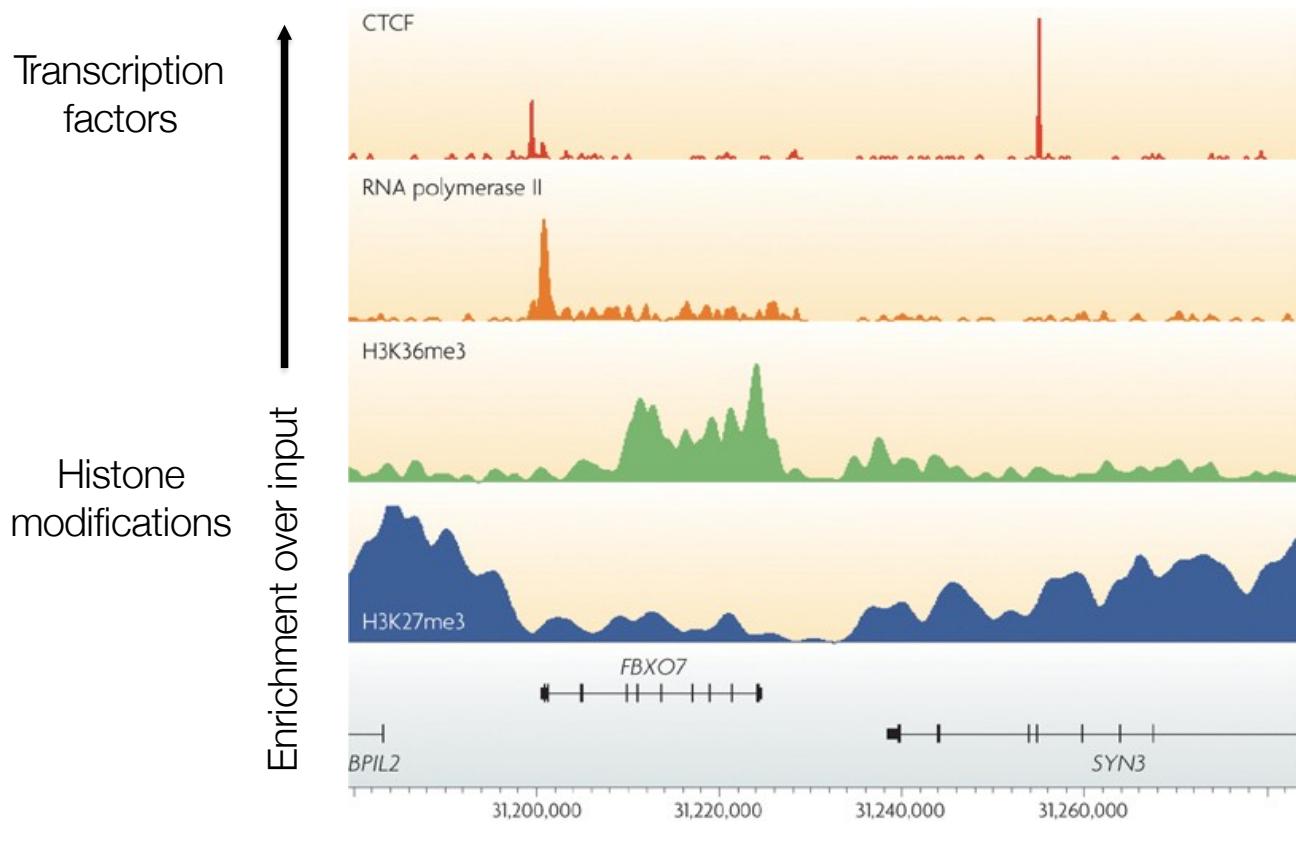


Link to dataset

<https://genome.ucsc.edu/s/pdrodrig/mm10%2Dmmg232>

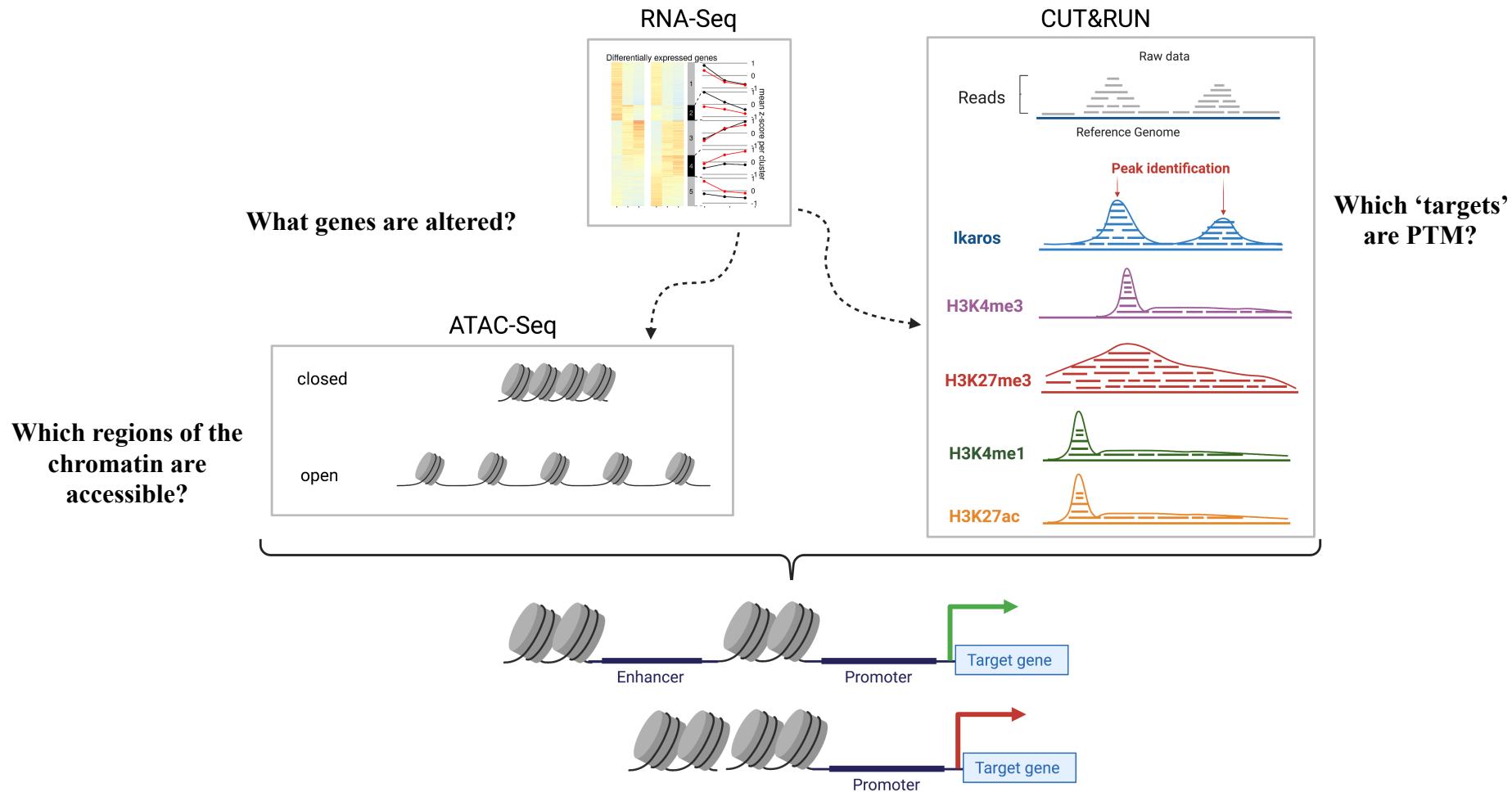


Types of signals



Adapted from Park (2009). Nature Reviews Genetics.

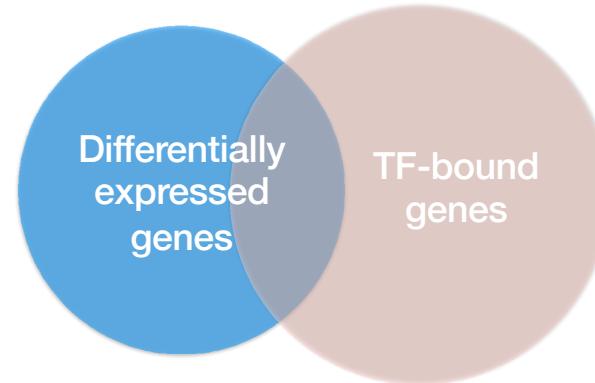
These assays are complementary and are often integrated



Purpose: Of gaining an understanding of gene regulation

Downstream analysis

- Integrative analysis of RNA-seq and ChIP-seq
 - Which of the regulated genes are direct targets of the TF?
 - Is the TF an activator, repressor, or both?
 - Does the TF have different binding partners depending on the direction of regulation?



BETA
Binding and Expression Target Analysis

[Introduction](#) | [Citation](#) | [Run on Webserver](#) | [Download](#) | [Installation](#) | [Tutorial](#) | [Contact](#)

Summary

Binding and Expression Target Analysis (BETA) is a software package that integrates ChIP-seq of transcription factors or chromatin regulators with differential gene expression data to infer direct target genes. BETA has three functions: (1) to predict whether the factor has activating or repressive function; (2) to infer the factor's target genes; and (3) to identify the motif of the factor and its collaborators which might modulate the factor's activating or repressive function. Here we describe the implementation and features of BETA to

Announcements

❖ STAR aligner – intro & script

❖ Overview of MULTIQC

1. Homework #8: FASTQC now due March 10th
 - ✓ Guidelines will be posted on Thursday, March 2nd
2. Homework #9: FASTQC + **alignment** stats now due March 24th
 - ✓ Guidelines will be posted on Monday, March 6th

❖ Other Assignment

- Primary research article summary – due April 7th
 - ✓ Guidelines will be posted on Monday, March 6th

❖ Homework #7 – These will be graded by the end of the week

hisat2_align.sh sections

- I. Part 1: Provide the job submission parameters
- II. Part 2: Section to keep the naming convention for each file output by script
- III. Part 3: Load the modules required to run the command
- IV. Part 4: Provide the actual commands to be executed
- V. **Part 5: Run MULTIQC after script has completed running**

Part 5: Run MULTIQC

Activate conda first

```
conda activate multiqc
```

then run the command

```
multiqc .
```

Citation

- “Understanding chromatin biology using high throughput sequencing (HTS)” – HSPH Bioinformatics Core w/ Dr. Meeta Mistry