Assays to measure chromatin accessibility

NGS analysis for gene regulation and epigenomics

Physalia 2021

Jacques Serizay

How to measure chromatin accessibility: originally with nucleases

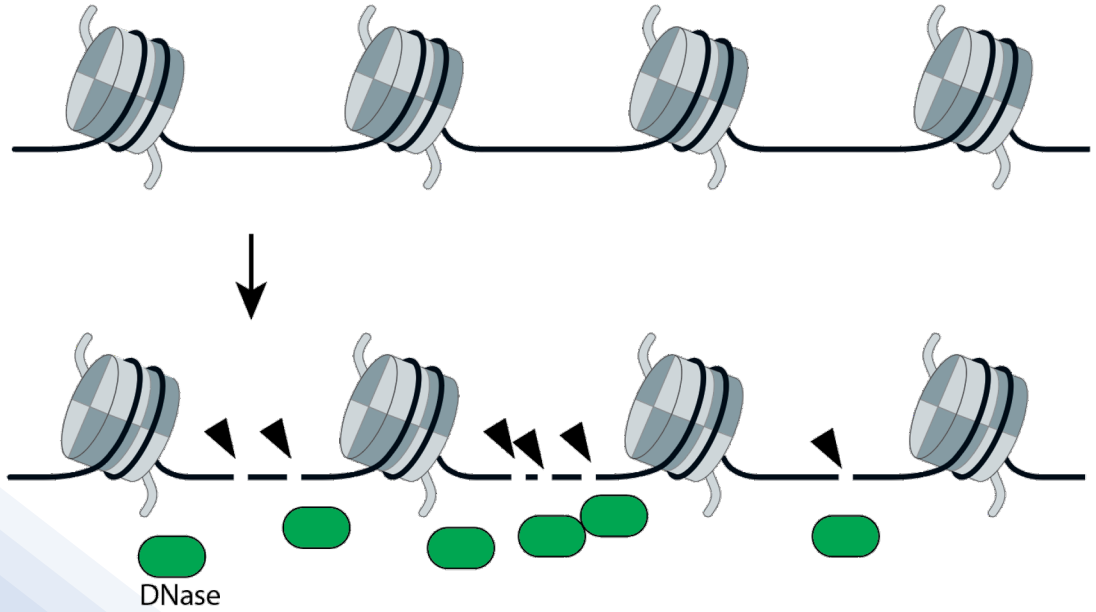
- The rationale to profile chromatin accessibility is that loci sensitive to enzymatic activity must be accessible to such enzymes.

How to measure chromatin accessibility: originally with nucleases

- Nuclease enzymes were historically used to profile chromatin accessibility

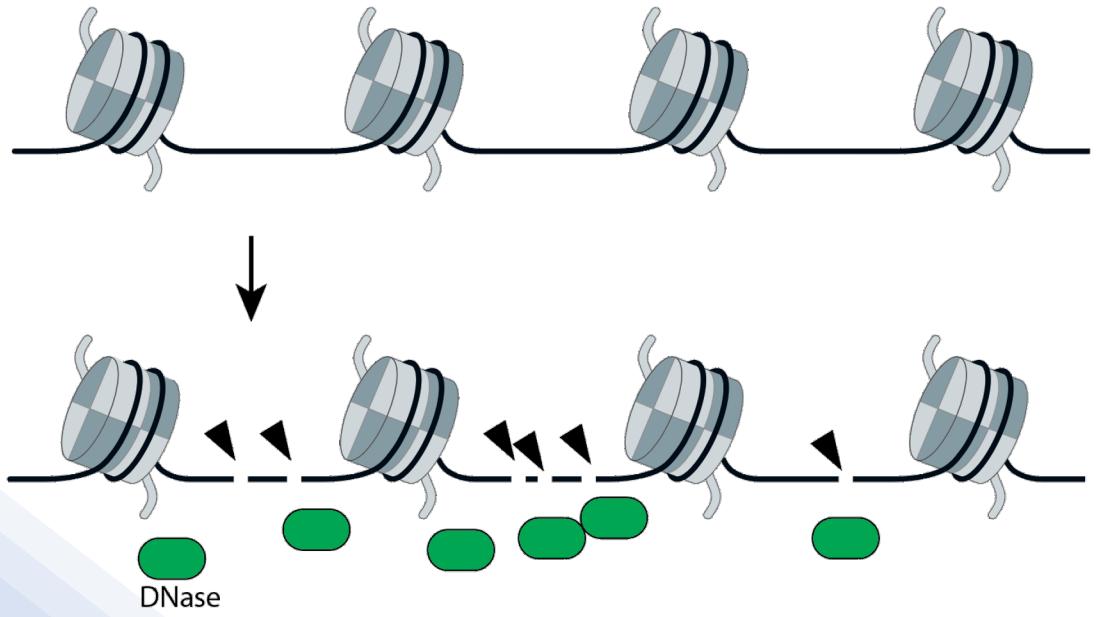
How to measure chromatin accessibility: originally with nucleases

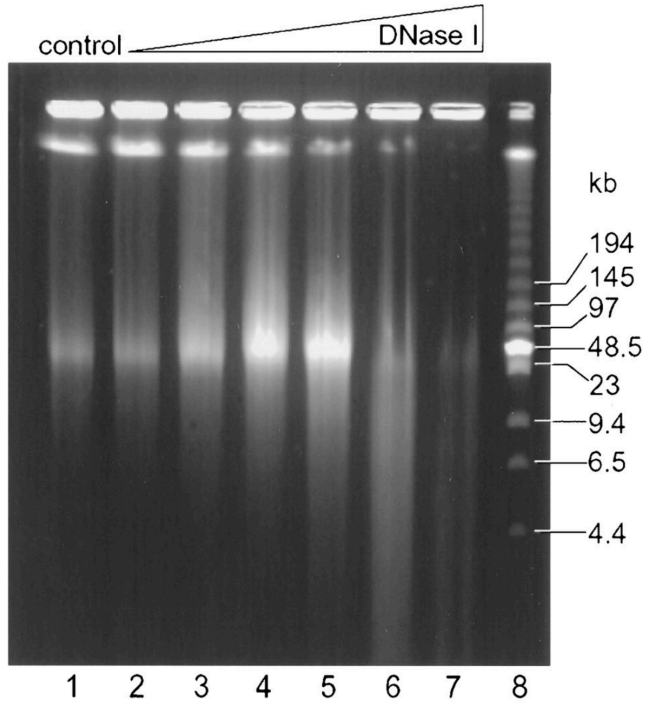
- Nuclease enzymes were historically used to profile chromatin accessibility
  - Dnase I (deoxyribonuclease I): a non-specific double-strand endonuclease



How to measure chromatin accessibility: originally with nucleases

- Nuclease enzymes were historically used to profile chromatin accessibility
  - Dnase I (deoxyribonuclease I): a non-specific double-strand endonuclease

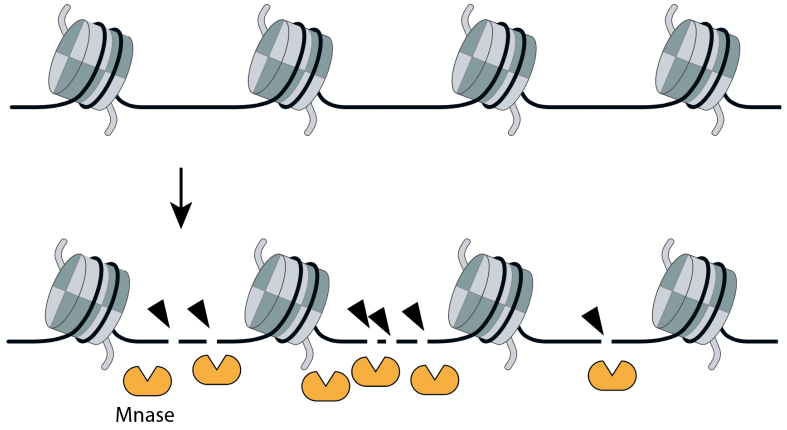




Paul & Ferl, Plant Cell 1998

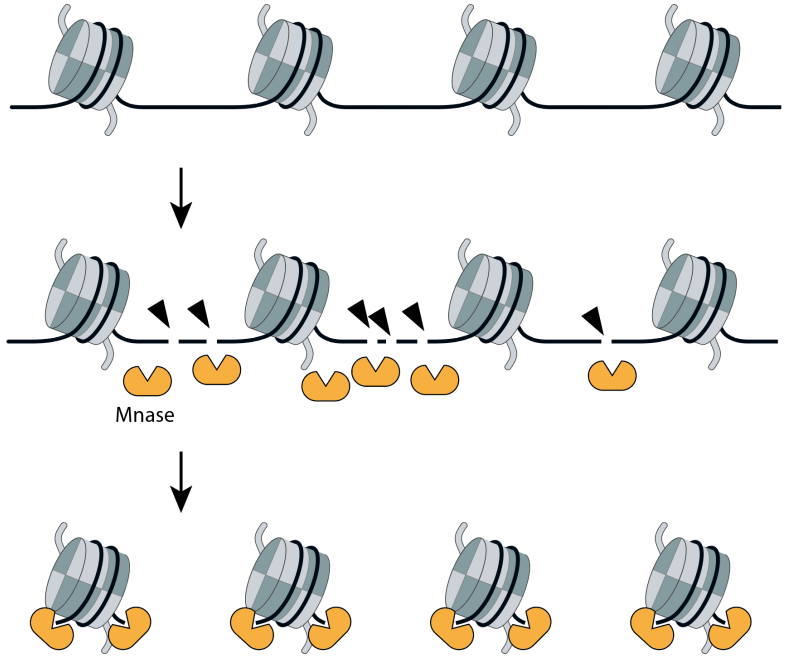
How to measure chromatin accessibility: originally with nucleases

- Nuclease enzymes were historically used to profile chromatin accessibility
  - Mnase (Micrococcal nuclease): a endo-exonuclease with preference to single-strand DNA but with also double-strand nuclease activity



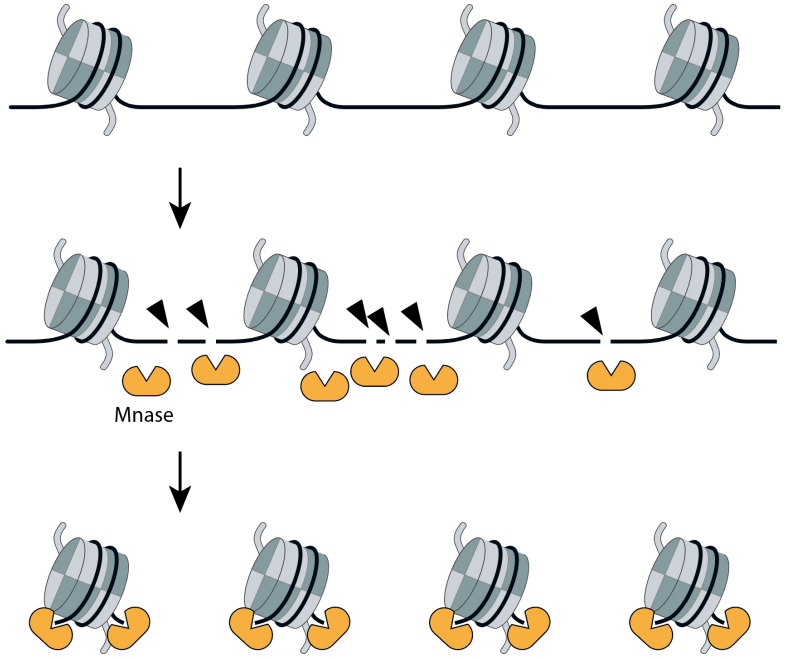
How to measure chromatin accessibility: originally with nucleases

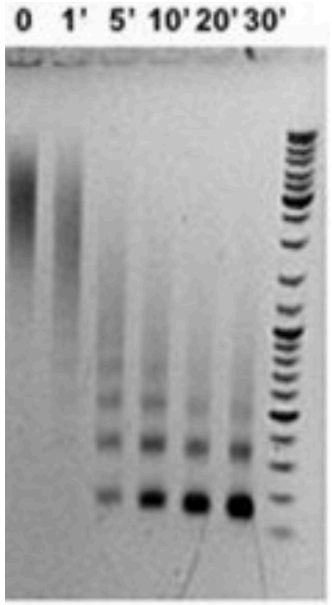
- Nuclease enzymes were historically used to profile chromatin accessibility
  - Mnase (Micrococcal nuclease): a endo-exonuclease with preference to single-strand DNA but with also double-strand nuclease activity



How to measure chromatin accessibility: originally with nucleases

- Nuclease enzymes were historically used to profile chromatin accessibility
  - Mnase (Micrococcal nuclease): a endo-exonuclease with preference to single-strand DNA but with also double-strand nuclease activity

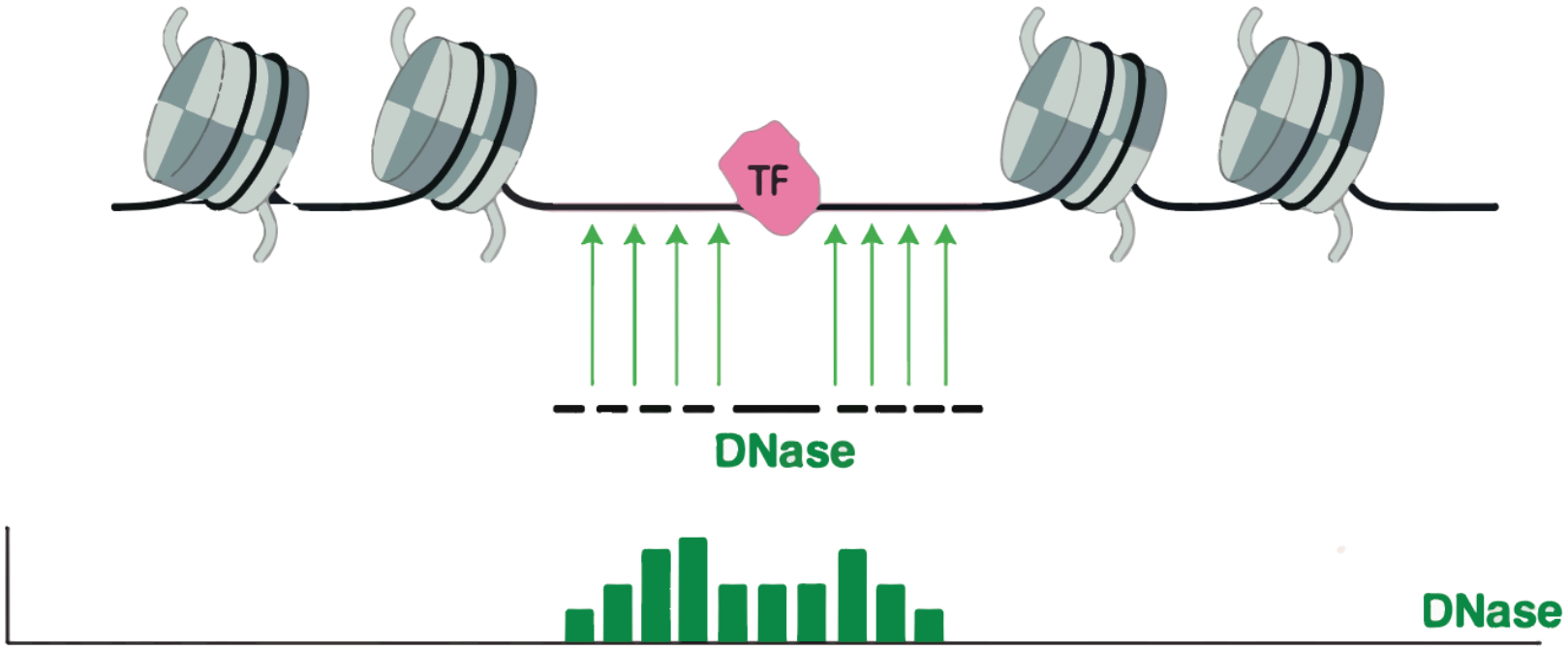




Rodríguez-Campos & Azorín, PLoS One 2007

Making an NGS library from nuclease-based accessibility assays

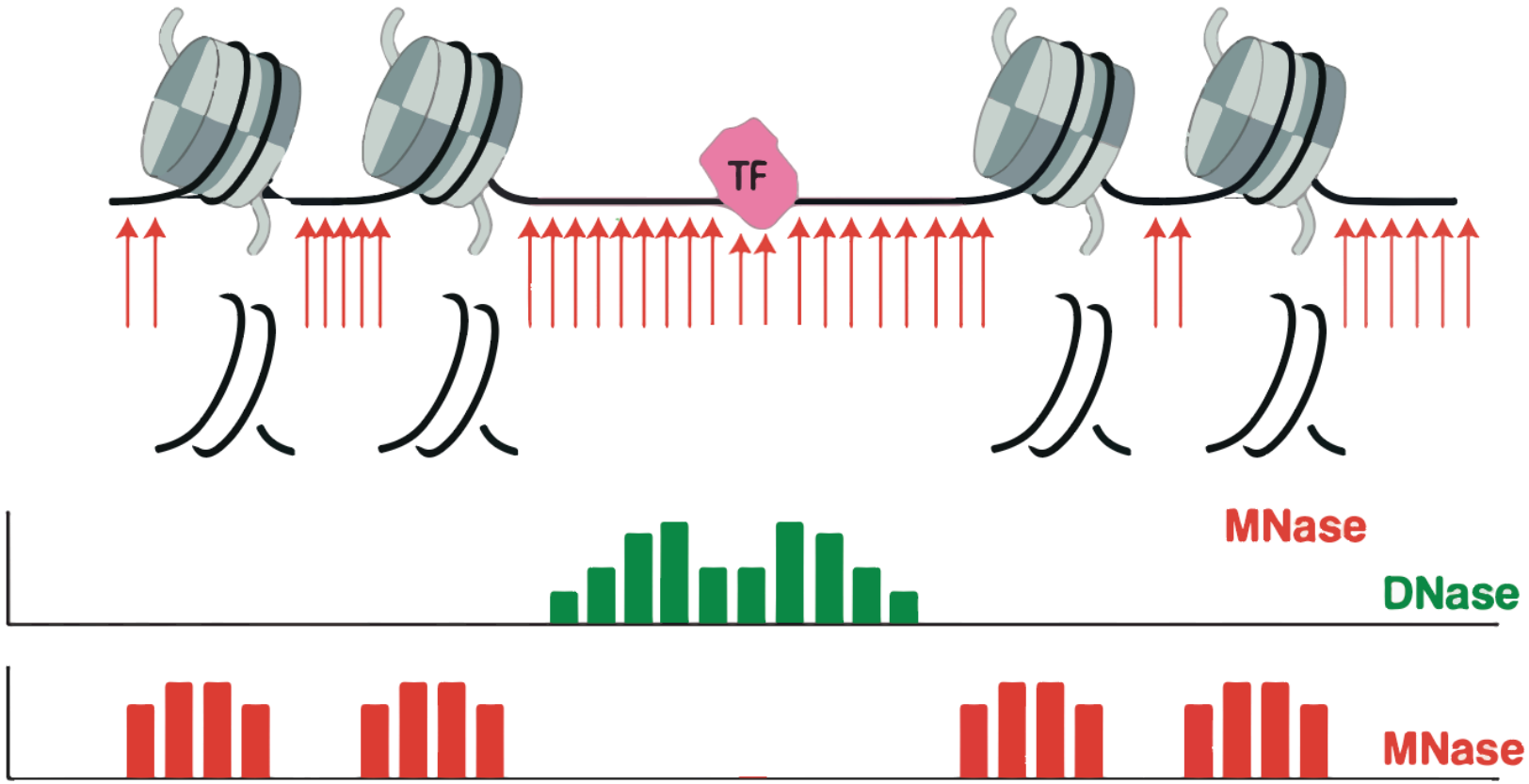
Chromatin accessibility is almost universally measured by quantifying the susceptibility of chromatin to enzymatic cleavage of its constituent DNA



Jacques Serizay

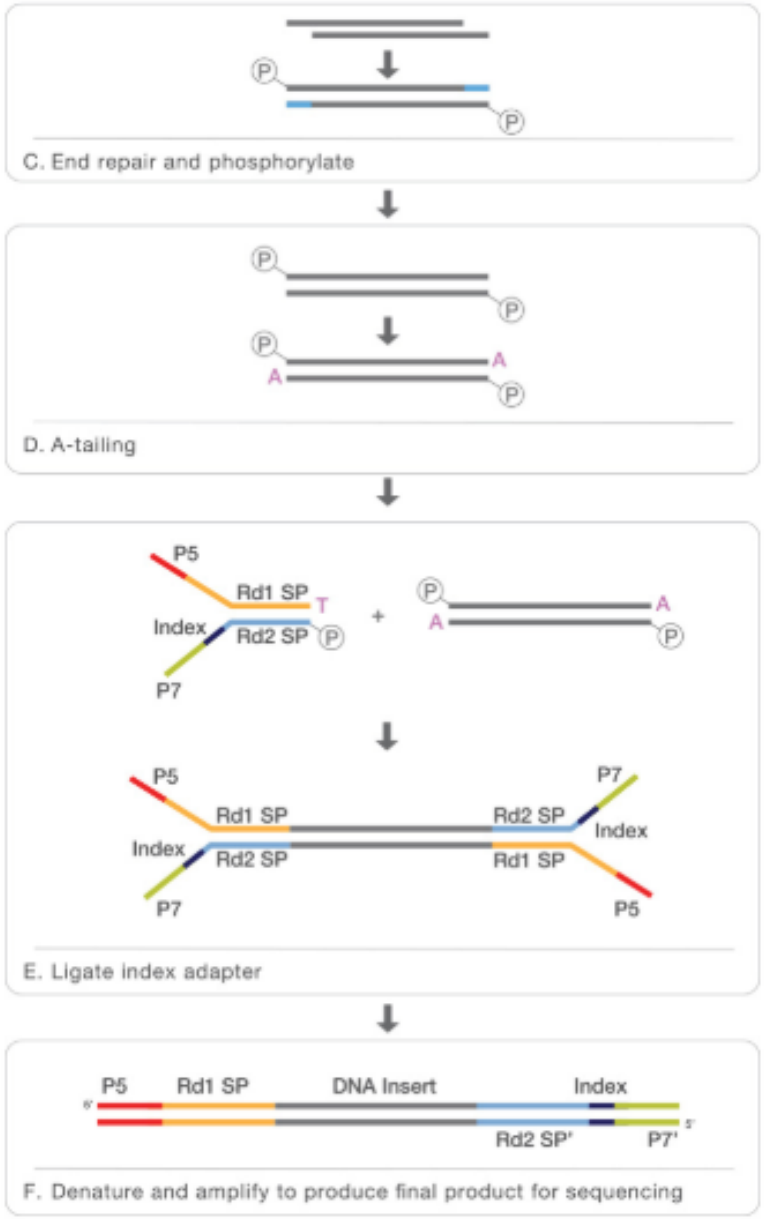
Making an NGS library from nuclease-based accessibility assays

Chromatin accessibility is almost universally measured by quantifying the susceptibility of chromatin to enzymatic cleavage of its constituent DNA



Making an NGS library from nuclease-based accessibility assays

- DNase I- or Mnase-based digestion of the chromatin both result in the isolation of dsDNA fragments.
- From there: classic NGS library preparation tedious steps…



ATAC-seq: a “new” enzyme for a new assay

- The most important technique that emerged since DNase-seq / Mnase-seq is **ATAC-seq**

ATAC-seq: a “new” enzyme for a new assay

- The most important technique that emerged since DNase-seq / Mnase-seq is **ATAC-seq**

- ATAC-seq stands for Assay for Transposase-Accessible Chromatin using sequencing

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

ATAC-seq: use of Tn5 to profile accessibility

- In ATAC-seq, Tn5 transposome is used instead of DNase I or Mnase

ATAC-seq: use of Tn5 to profile accessibility

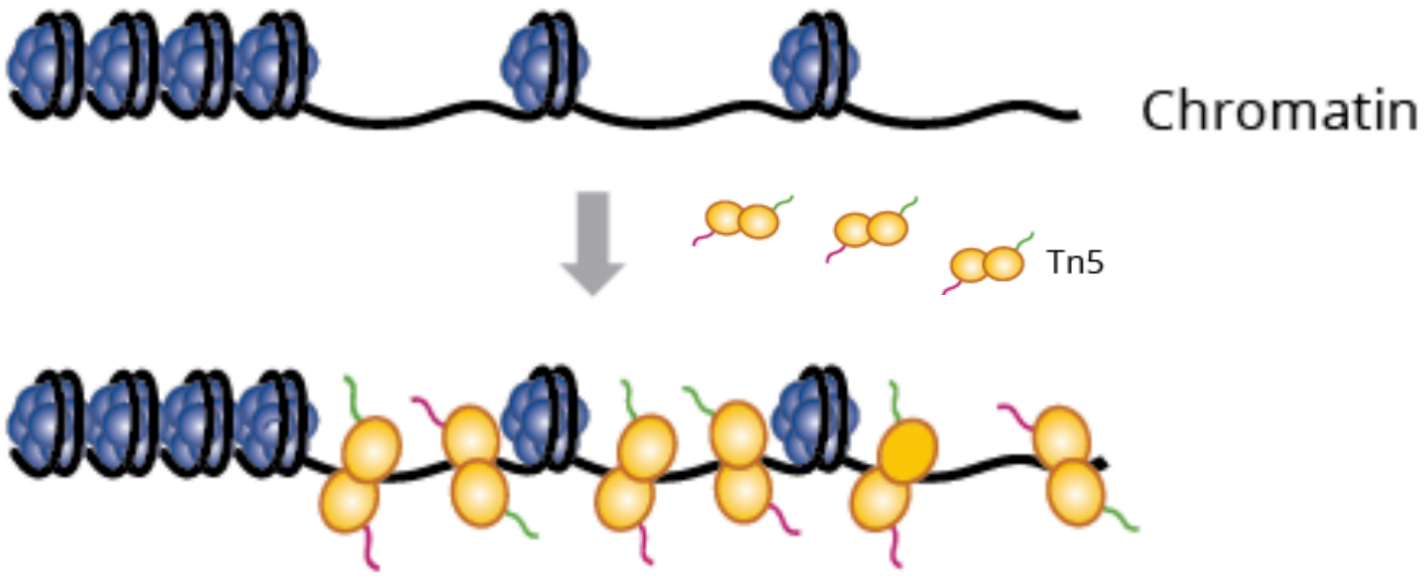
- In ATAC-seq, Tn5 transposome is used instead of DNase I or Mnase

- Tn5 transposome consists of:

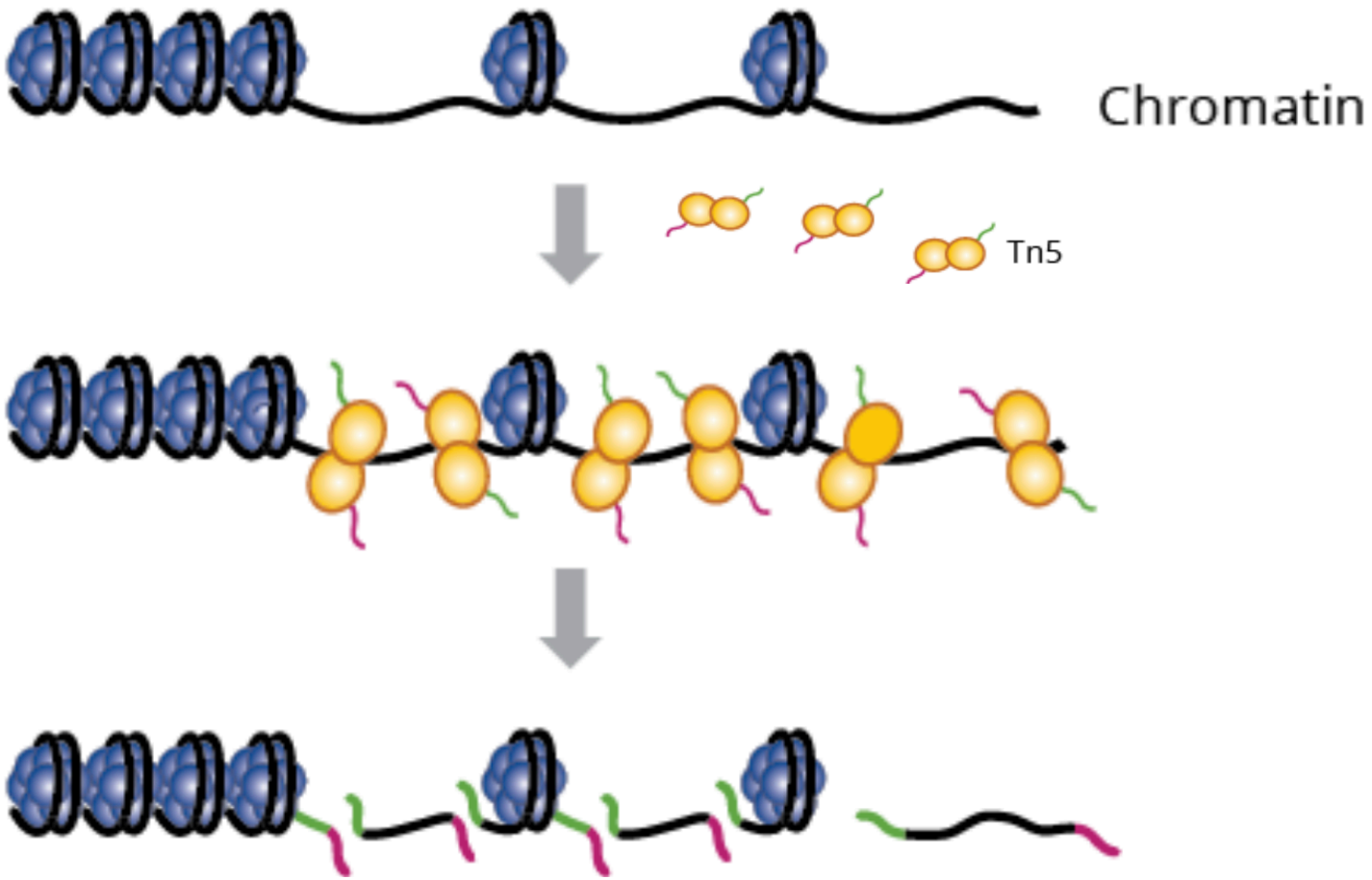
  - The Tn5 transposase: an enzyme that can integrates transposons in foreign DNA
  - Tn5 transposons: usually minimal oligosequences loaded onto the transposase

ATAC-seq: use of Tn5 to profile accessibility

ATAC-seq: use of Tn5 to profile accessibility



ATAC-seq: use of Tn5 to profile accessibility

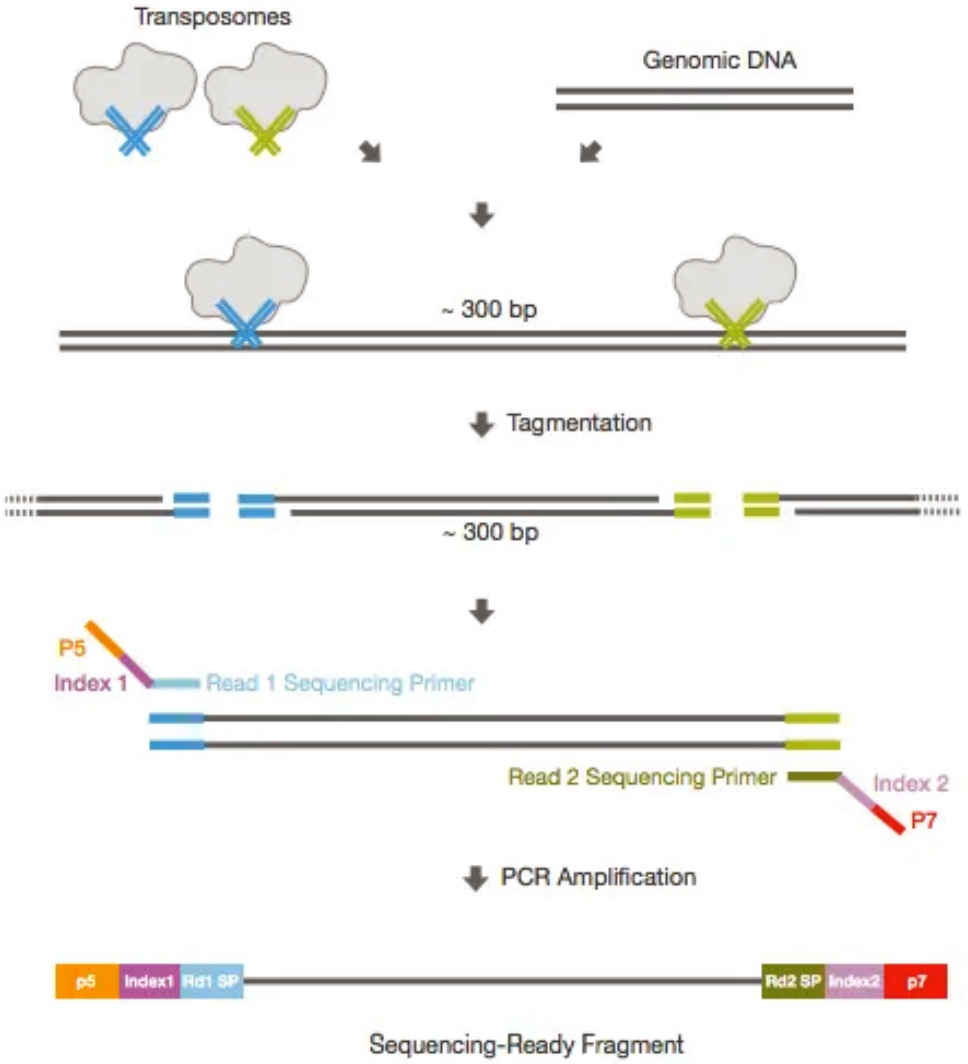


ATAC-seq: from chromatin to NGS library

- Since the sequence of the transposons loaded on the Tn5 transposome is known, one can use them to run a PCR

→ "Tagmented" DNA will be amplified

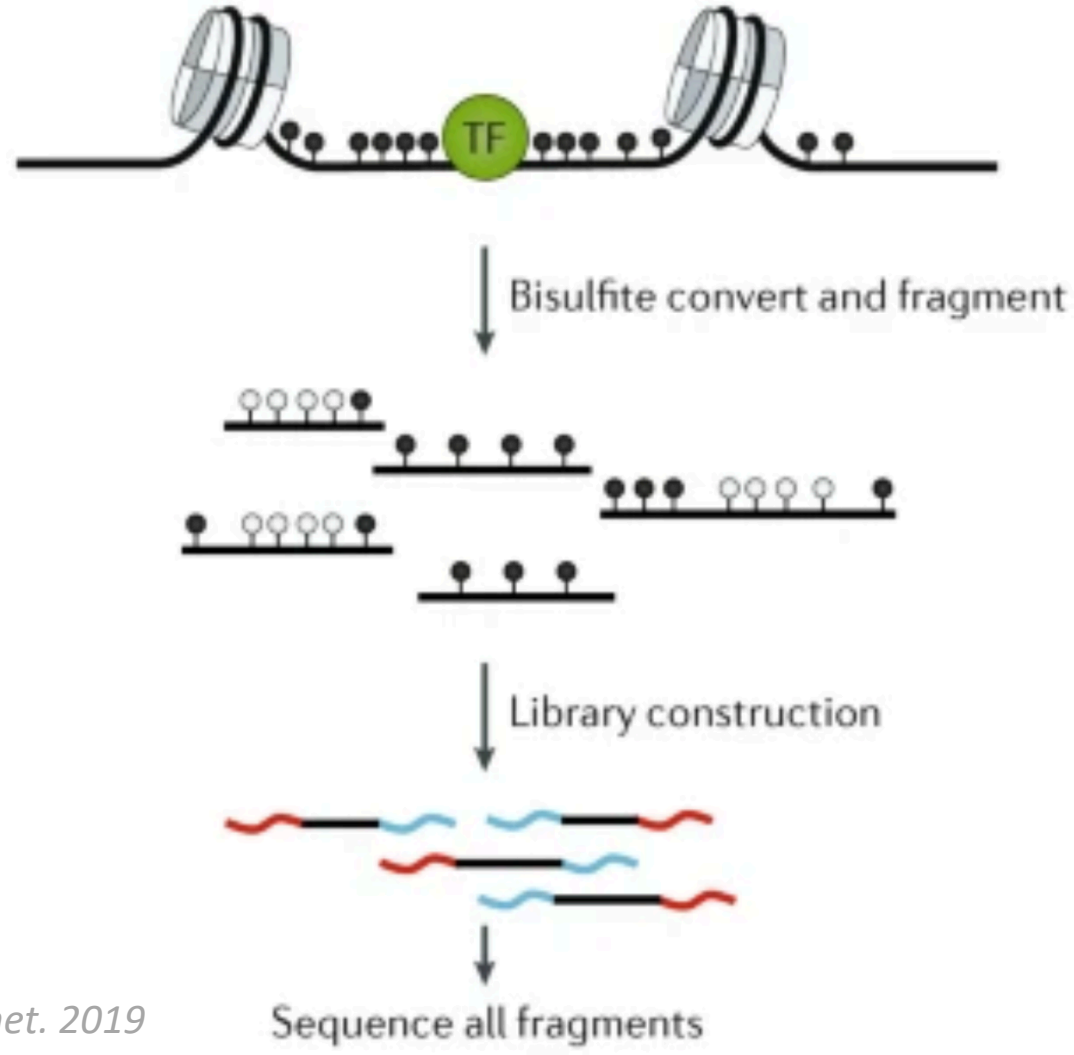
→ Each end of a fragment corresponds to a transposition event



Emergence of other enzymatic or mechanical approaches

- Enzymatic approaches
  - **NOMe-seq**: Nucleosome Occupancy and Methylome sequencing

→ Uses a GpC methyltransferase to methylate accessible DNA

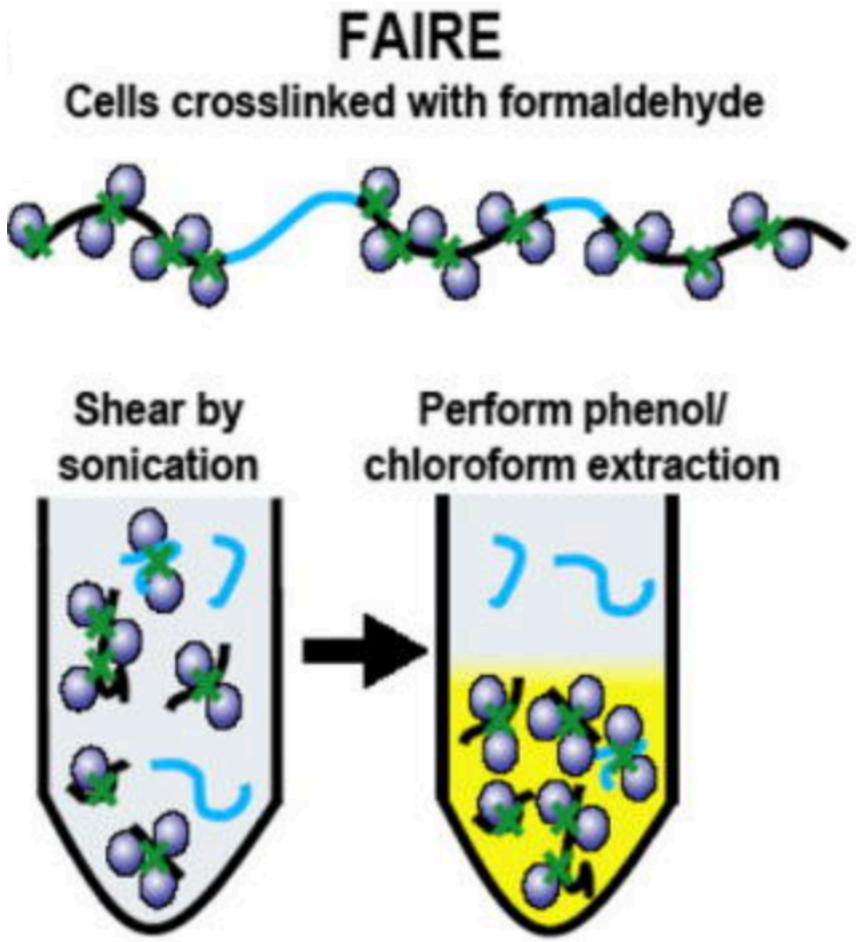


Klem et al., Nat. Rev. Genet. 2019

Emergence of other enzymatic or mechanical approaches

- Mechanical approaches
  - **FAIRE-seq**: Formaldehyde-Assisted Isolation of Regulatory Elements

→ Uses a crosslinking + sonication + phenol extraction to isolate nucleosome-depleted chromatin

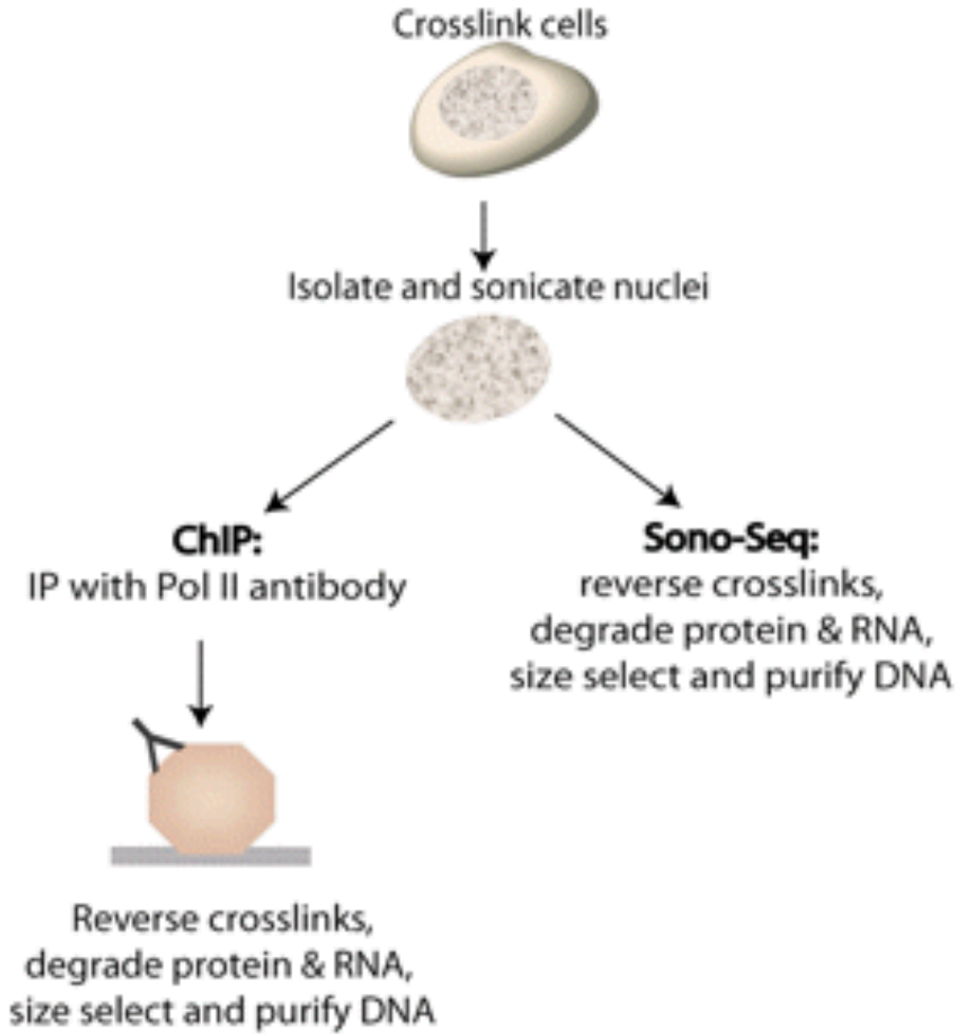


Giresi & Lieb, Methods 2009

Emergence of other enzymatic or mechanical approaches

- Mechanical approaches
  - **SONO-seq**: chromatin Sonication followed by sequencing

→ Uses a crosslinking + sonication + size selection to isolate small fragmented chromatin



Auerbach et al., PNAS 2009

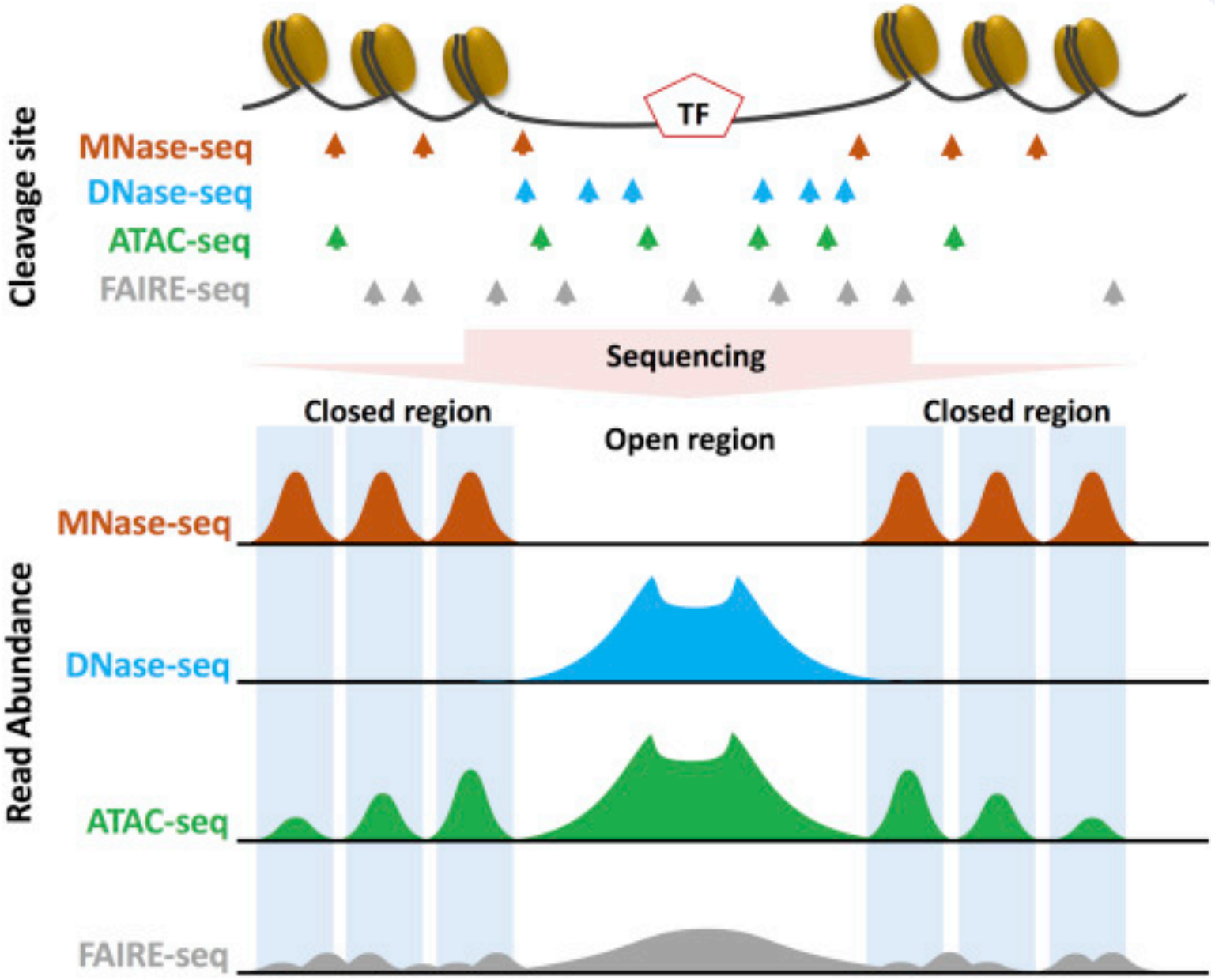
Comparison of the main experimental approaches

	Cell type/Number	Sequencing type	Traditional approach	Genomic target	Experimental considerations
MNase-seq	Any cell type 1 to 10 million cells	Paired-end or Single-end	MNase digests unprotected DNA	Maps the total nucleosome population in a qualitative and quantitative manner	1. Requires many cells.
					2. Laborious enzyme titrations.
					3. Probes total nucleosomal population, not active regulatory regions only.
					4. Degrades active regulatory regions, making their detection possible only *indirectly*.
					5. Requires 150 to 200 million reads for standard accessibility studies of the human genome.
DNase-seq	Any cell type 1 to 10 million cells	Paired-end or Single-end	DNase I cuts within unprotected DNA	Maps open chromatin	1. Requires many cells.
					2. Time-consuming and complicated sample preparations.
					3. Laborious enzyme titrations.
					4. Requires 20 to 50 million reads for standard accessibility studies of the human genome.
FAIRE-seq	Any cell type 100,000 to 10 million cells	Paired-end or Single-end	Based on the phenol-chloroform separation of nucleosome-bound and free sonicated areas of a genome, in the interphase and aqueous phase respectively	Maps open chromatin	1. Low signal-to-noise ratio, making computational data interpretation very difficult.
					2. Results depend highly on fixation efficiency.
					3. Requires 20 to 50 million reads for standard accessibility studies of the human genome.
ATAC-seq	500 to 50,000 freshly isolated cells	Paired-end or Single-end	Unfixed nuclei are tagged *in vitro* with adapters for NGS by purified Tn5 transposase. Adapters are integrated into regions of accessible chromatin	Maps open chromatin, TF and nucleosome occupancy	1. Contamination of generated data with mitochondrial DNA.
					2. Immature data analysis tools.
					3. Requires 60 to 100 million reads for standard accessibility studies of the human genome.

Tsompana & Buck, Epigenetics and Chromatin 2014

Comparison of the main experimental approaches

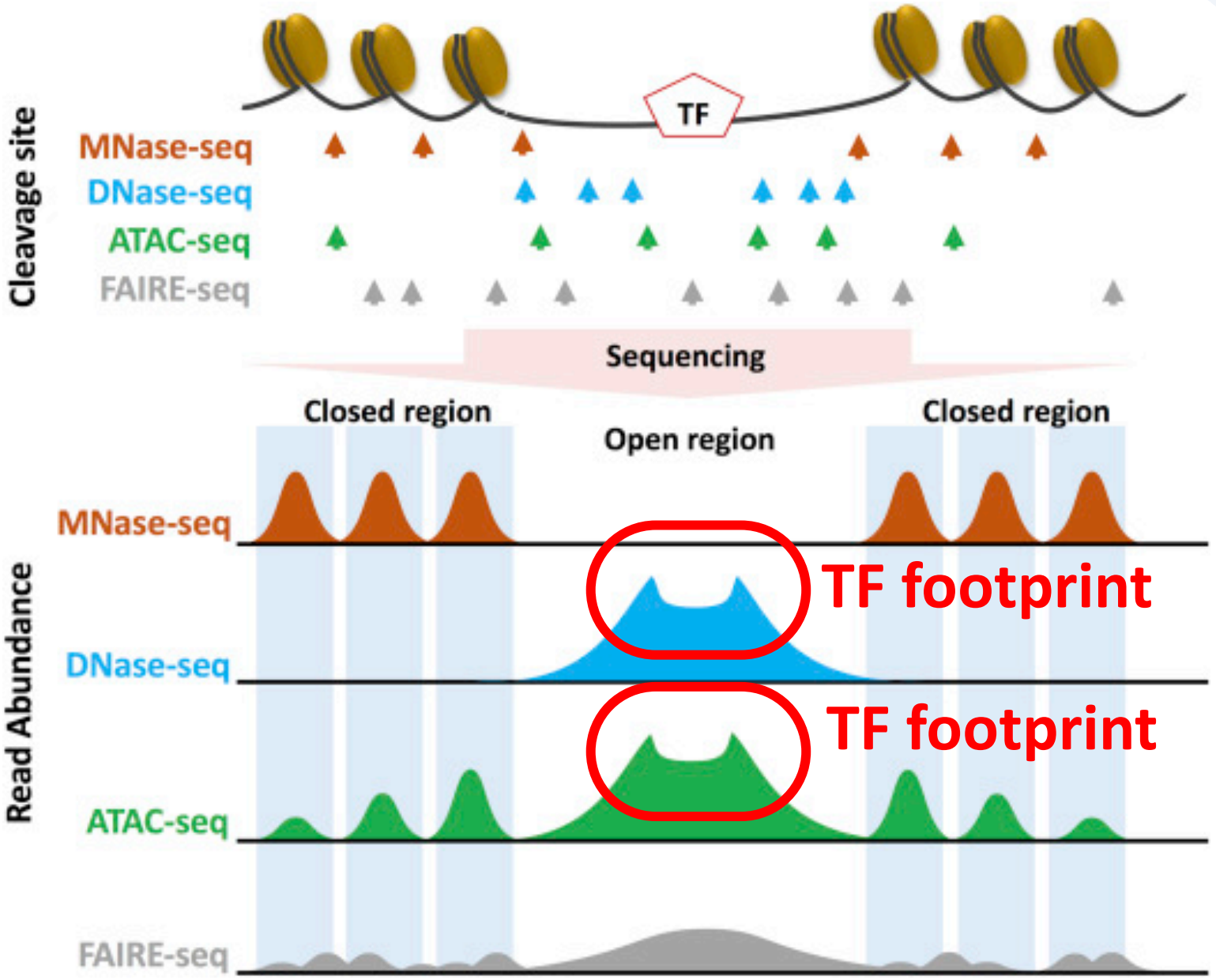
- Each assay gives a different answer



Hsu et al., Epigenetics in Human Diseases 2018

Comparison of the main experimental approaches

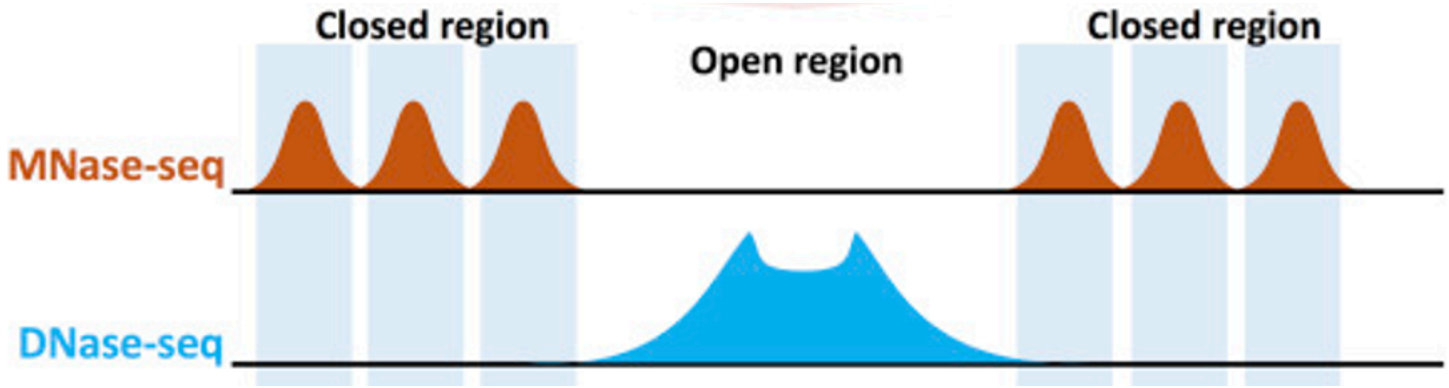
- Each assay gives a different answer



Positive vs. negative measurements

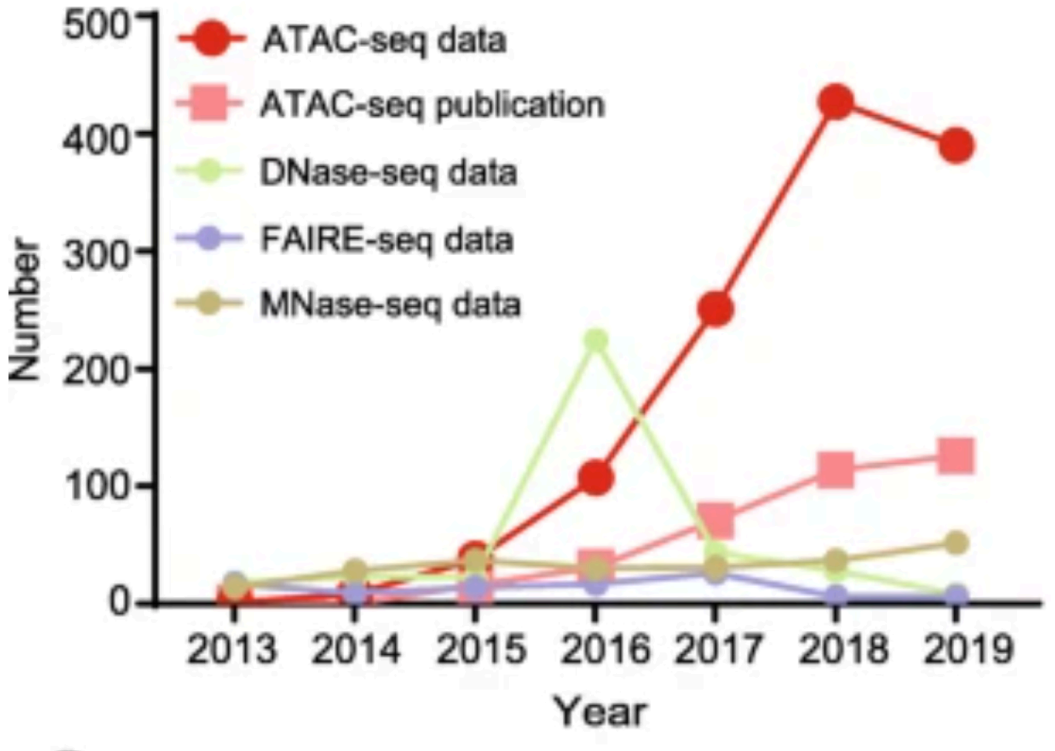
- DNAse-seq relies on presence of signal (**positive measurements**) to map accessible regulatory elements

- Mnase-seq relies on absence of signal (**negative measurements**) to map accessible regulatory elements



ATAC-seq is an increasingly used approach

- ATAC-seq signal: intermediate between DNase-seq and Mnase-seq signals:
  - TF footprints are visible in ATAC-seq signals
  - Nucleosomes flanking an accessible region can also be detected



Yan et al., Genome Biol. 2020

Emerging single-cell approaches to profile chromatin accessibility

- scATAC-seq:
  - Essentially performing ATAC-seq within tiny droplets, each containing a single nucleus.

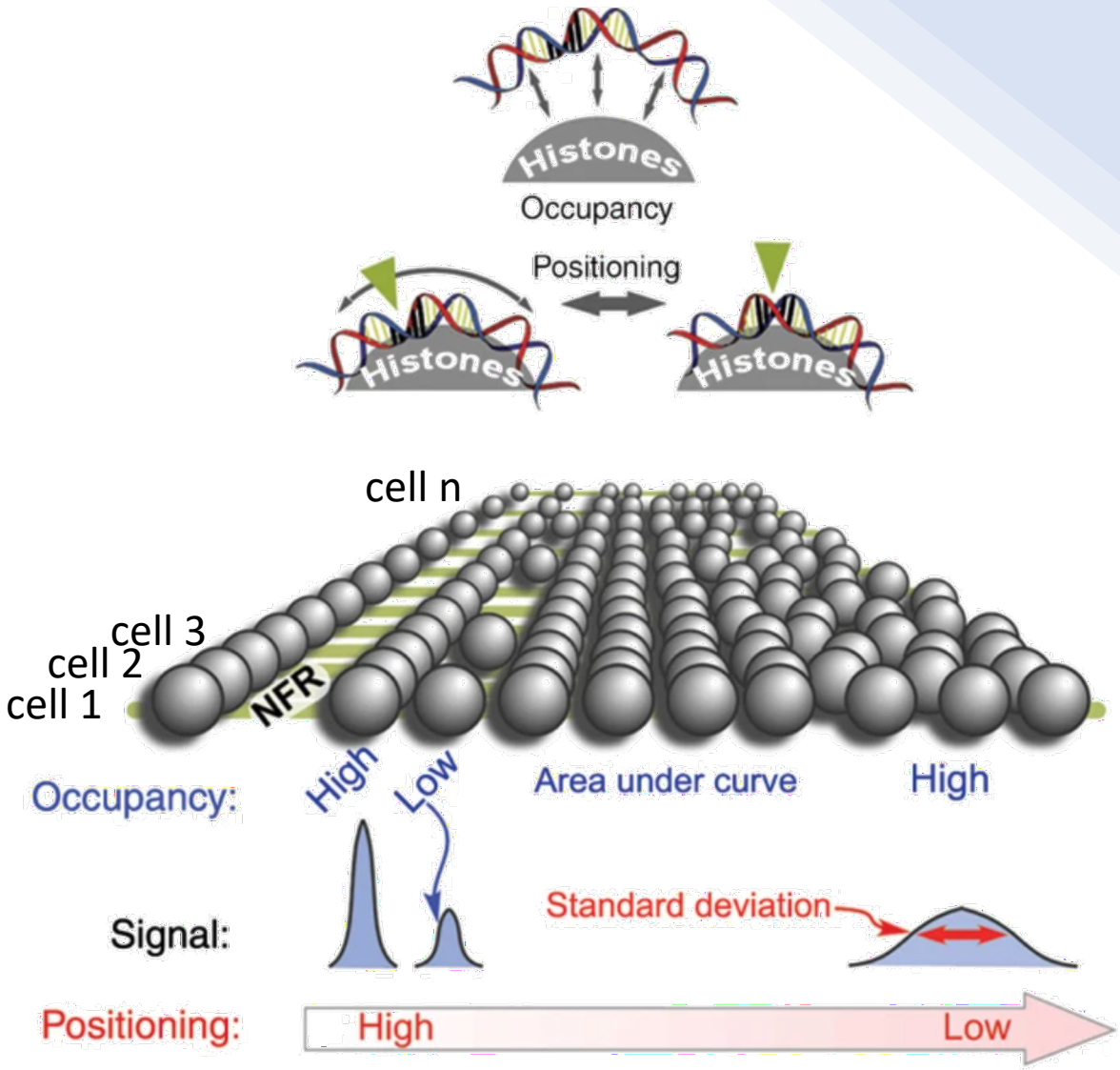
https://www.10xgenomics.com/products/single-cell-atac

- sci-CAR:
  - Randomly sorting cells in 384-well plates and indexing cells. Pool and repeat. After several rounds, each cell has a unique combination of indices. Both RNA-seq and ATAC-seq are performed on pooled cells and demultiplexing is done after sequencing.

Cao et al., Science 2018

Some semantics

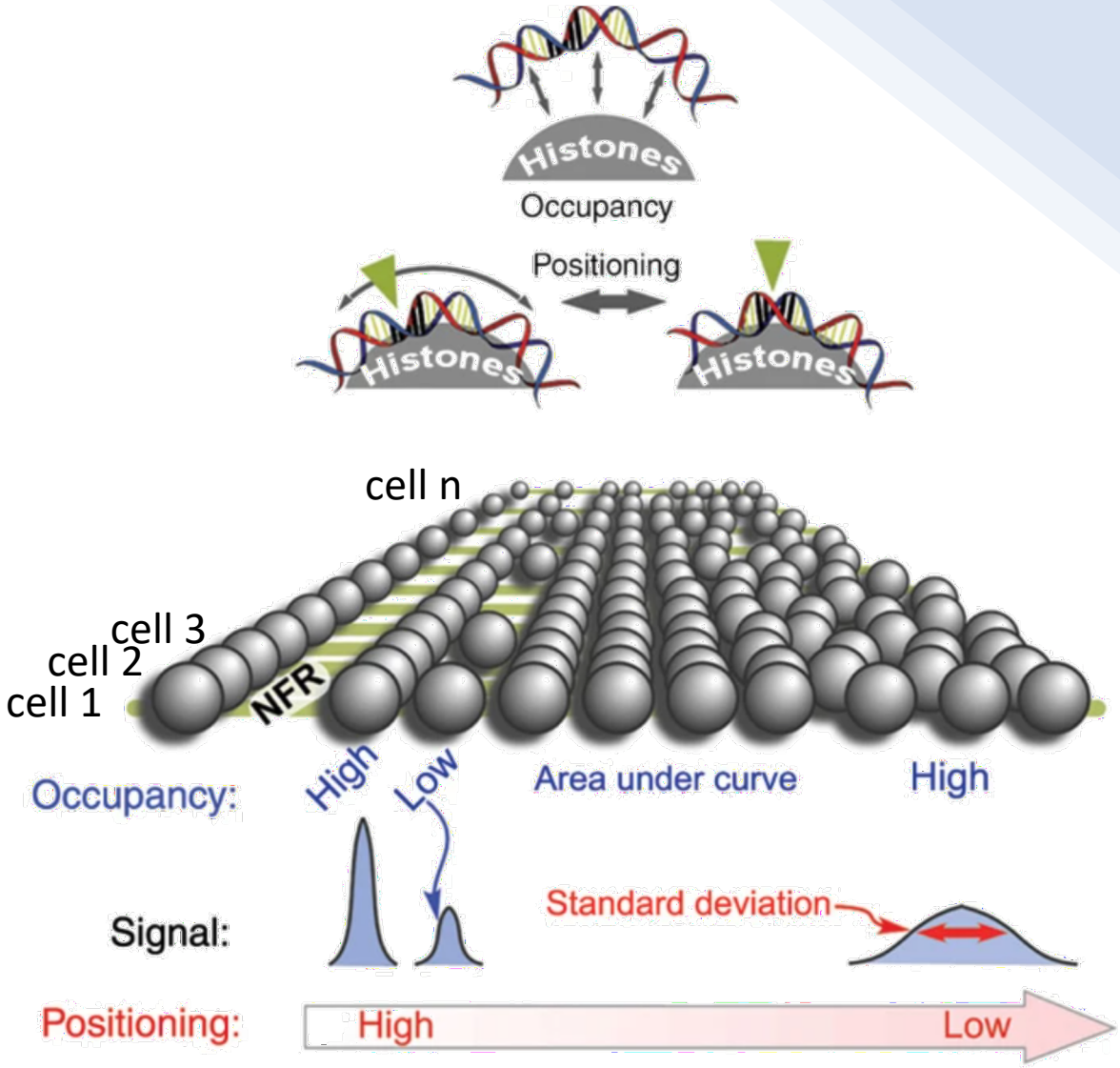
- Accessibility
- nucleosome occupancy
- nucleosome positioning



Pugh, Nature Structural & Molecular Biology. 2010

Some semantics

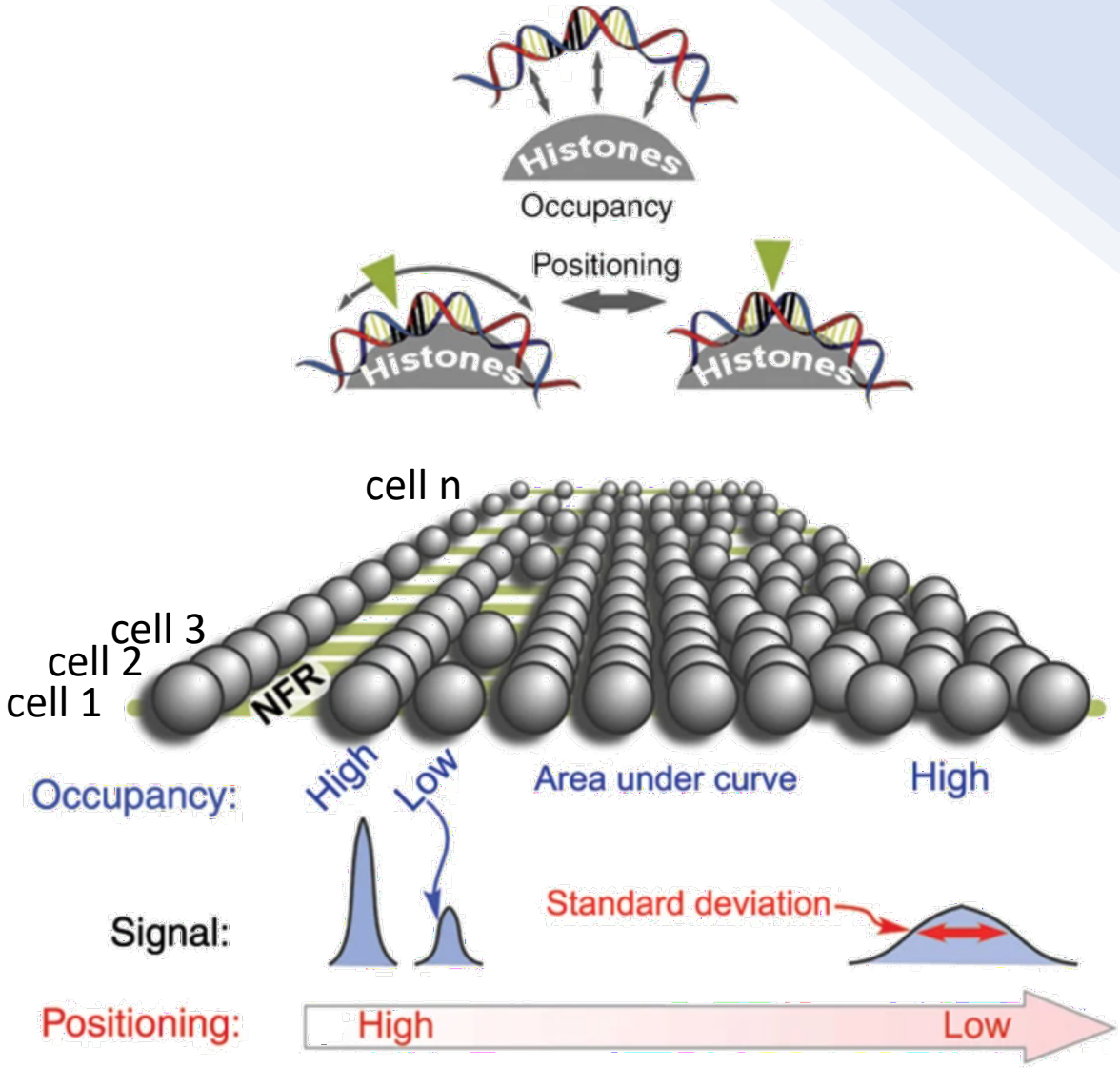
- Accessibility
  - Is the absence of proteins bound to a region of DNA
- nucleosome occupancy
- nucleosome positioning



Pugh, Nature Structural & Molecular Biology. 2010

Some semantics

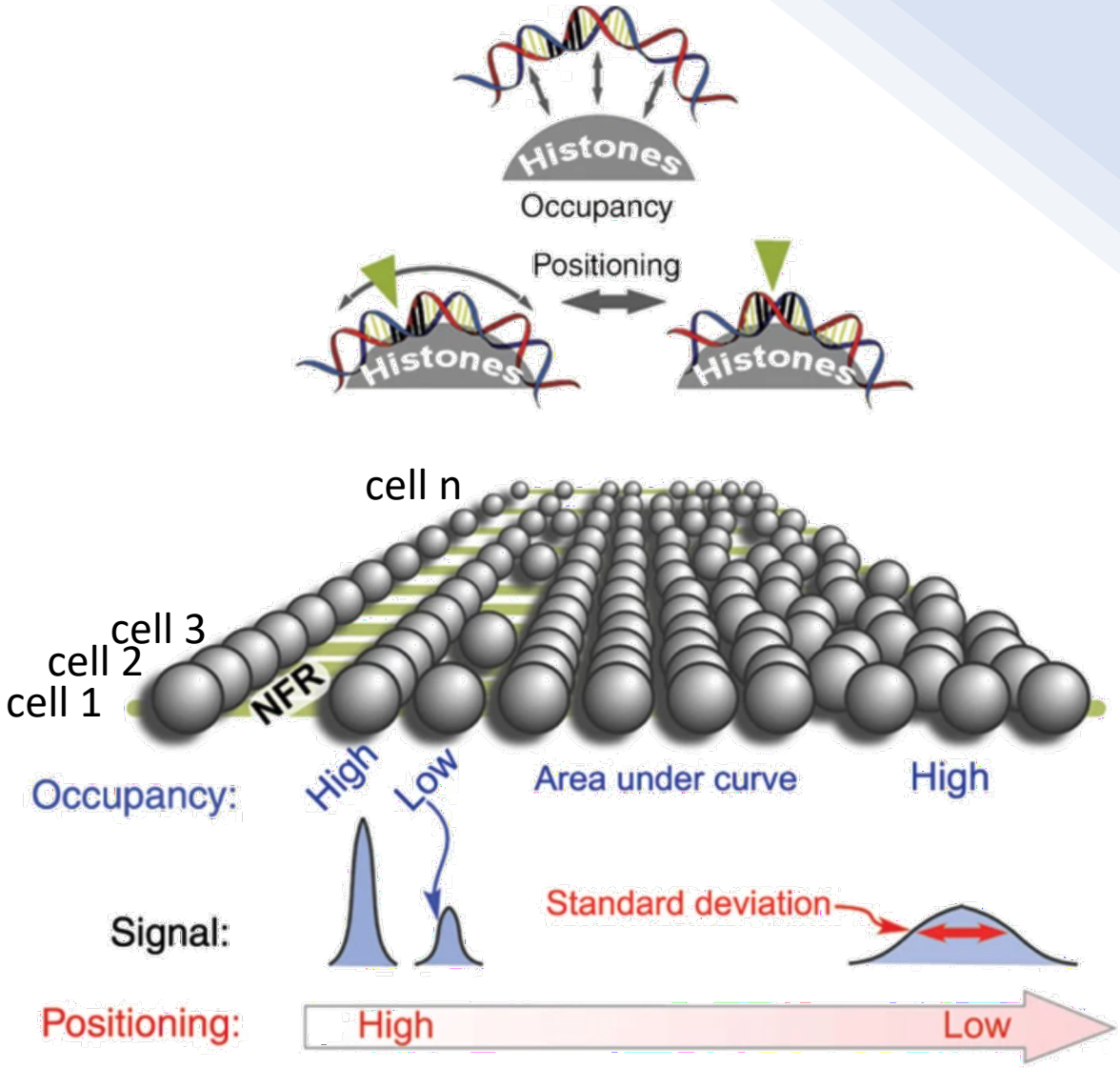
- Accessibility
  - Is the absence of proteins bound to a region of DNA
- nucleosome occupancy
  - Is the presence of nucleosomes bound to a region of DNA
- nucleosome positioning



Pugh, Nature Structural & Molecular Biology. 2010

Some semantics

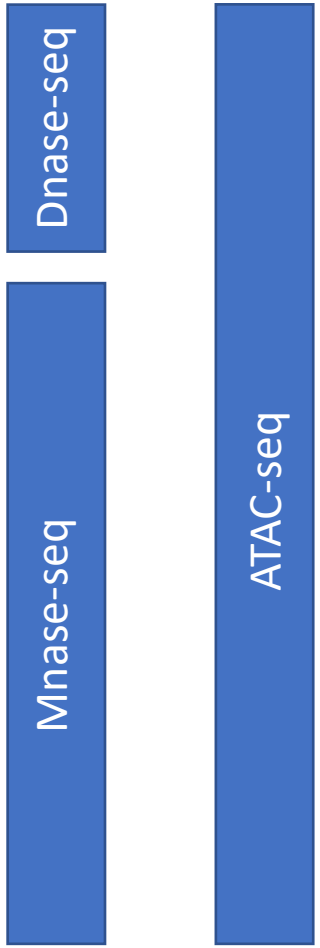
- Accessibility
  - Is the absence of proteins bound to a region of DNA
- nucleosome occupancy
  - Is the presence of nucleosomes bound to a region of DNA
- nucleosome positioning
  - Is the reproducible alignment of nucleosomes over a specific region of DNA



Pugh, Nature Structural & Molecular Biology. 2010

Some semantics

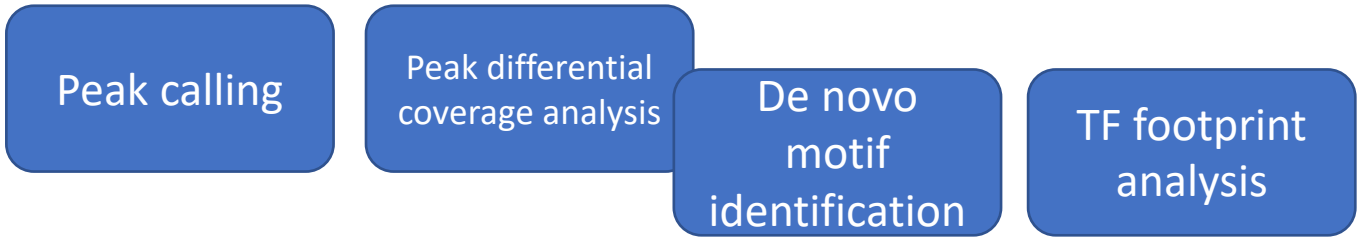
- Accessibility
  - Is the absence of proteins bound to a region of DNA
- nucleosome occupancy
  - Is the presence of nucleosomes bound to a region of DNA
- nucleosome positioning
  - Is the reproducible alignment of nucleosomes over a specific region of DNA



Pugh, Nature Structural & Molecular Biology. 2010

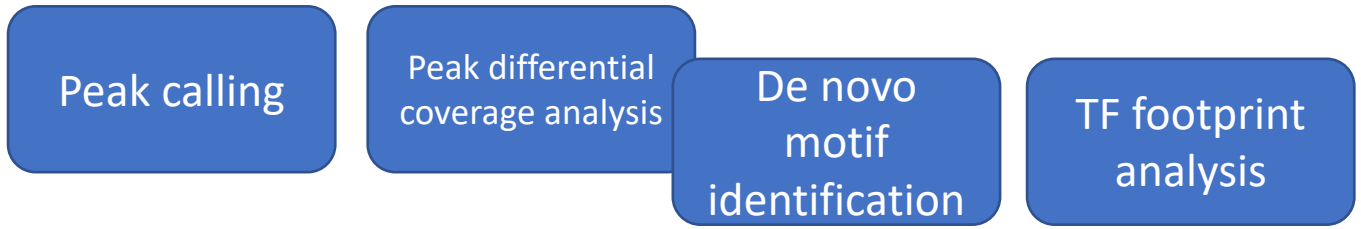
Different assays for different analyses

- DNase-seq



- Mnase-seq

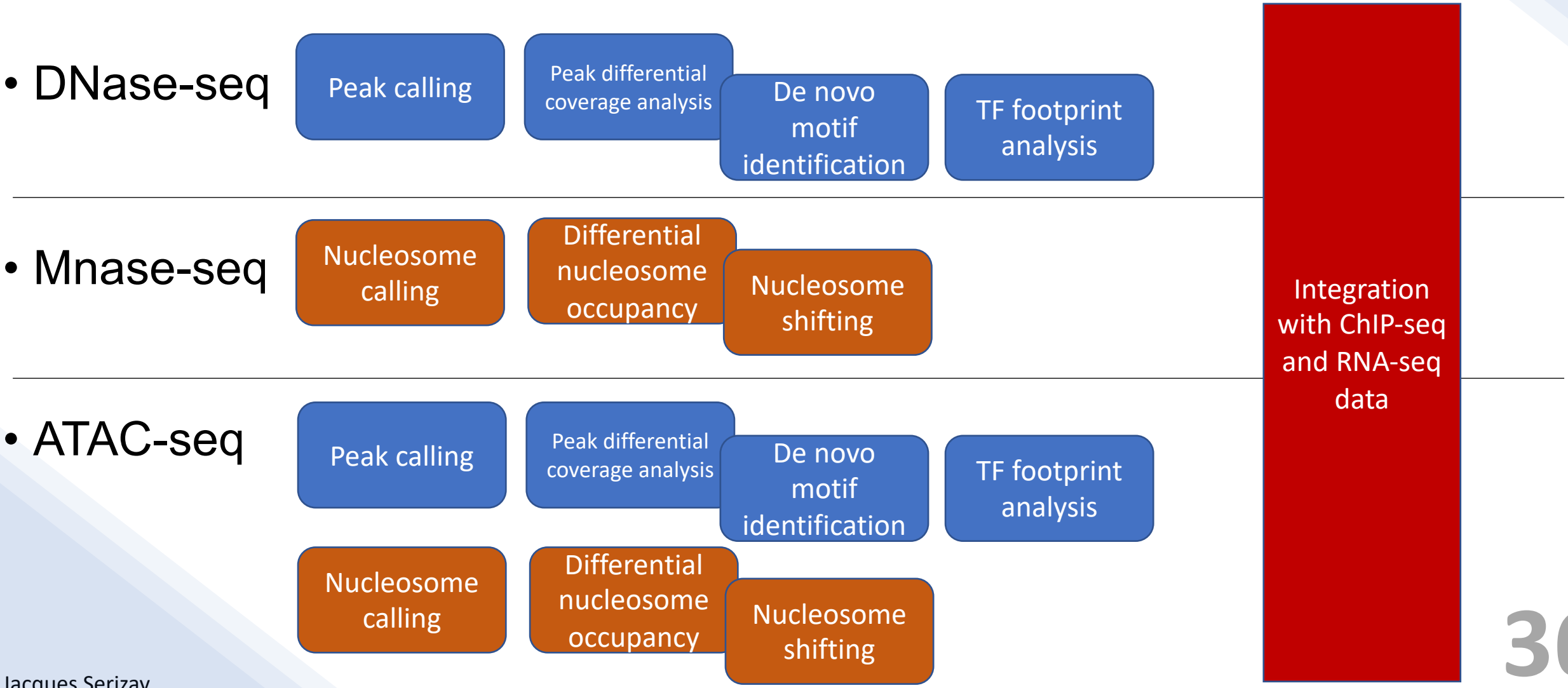
- ATAC-seq



Different assays for different analyses

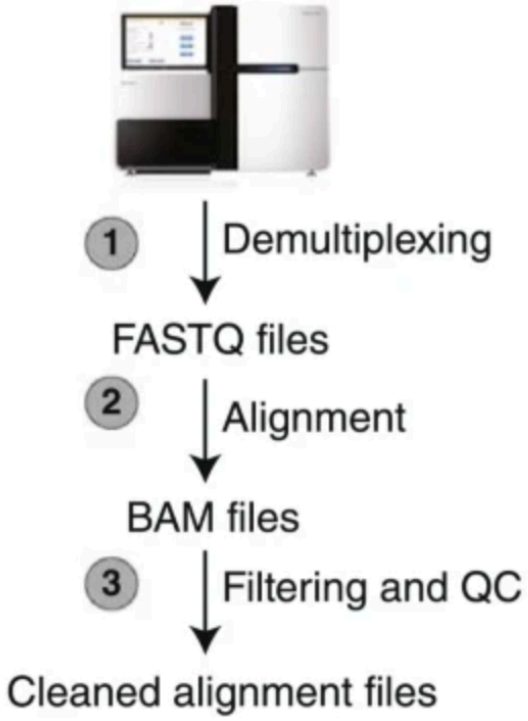
- DNase-seq
  - Peak calling
  - Peak differential coverage analysis
  - De novo motif identification
  - TF footprint analysis
- Mnase-seq
  - Nucleosome calling
  - Differential nucleosome occupancy
  - Nucleosome shifting
- ATAC-seq
  - Peak calling
  - Peak differential coverage analysis
  - De novo motif identification
  - TF footprint analysis
  - Nucleosome calling
  - Differential nucleosome occupancy
  - Nucleosome shifting

Different assays for different analyses



ATAC-seq downstream analysis

- First steps: just like any other “1D” NGS library (i.e. classic coverage enrichment assays, e.g. ChIP-seq, RNA-seq, …)



ATAC-seq downstream analysis

- Demultiplexing reads
- Adaptor trimming
- Fastq QC
- Alignment
- Filtering
- Alignment QC
- Peak calling
- Track generation

ATAC-seq downstream analysis

- Demultiplexing reads — Bcl → fastq
- Adaptor trimming — Fastq → fastq
- Fastq QC — Fastq → Fastqc
- Alignment — Fastq → bam
- Filtering — Bam → bam
- Alignment QC — Bam → txt
- Peak calling — Bam → bed *
- Track generation — Bam → bigiwg

ATAC-seq downstream analysis

• Demultiplexing reads	Bcl → fastq	Illumina pipelines
• Adaptor trimming	Fastq → fastq	trim_galore
• Fastq QC	Fastq → Fastqc	FastQC
• Alignment	Fastq → bam	bwa
• Filtering	Bam → bam	samtools
• Alignment QC	Bam → txt	samtools
• Peak calling	Bam → bed *	yapc
• Track generation	Bam → bigiwg	bedtools

Peak callers

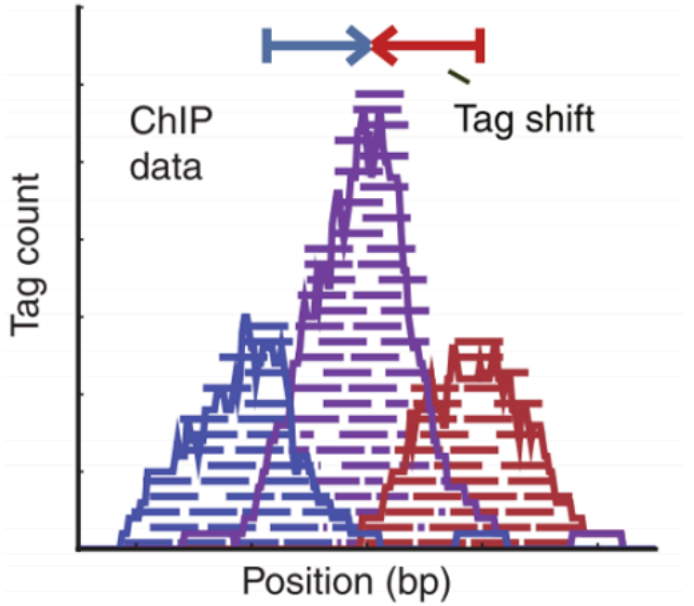
- Most peak callers were designed before the emergence of ATAC-seq

Peak callers

- Most peak callers were designed before the emergence of ATAC-seq
- Few of them directly aim at identifying peaks in chromatin accessibility signals

Peak callers

- For instance, MACS2 primary goal is to find a model to shift single-end reads toward the real TF's position

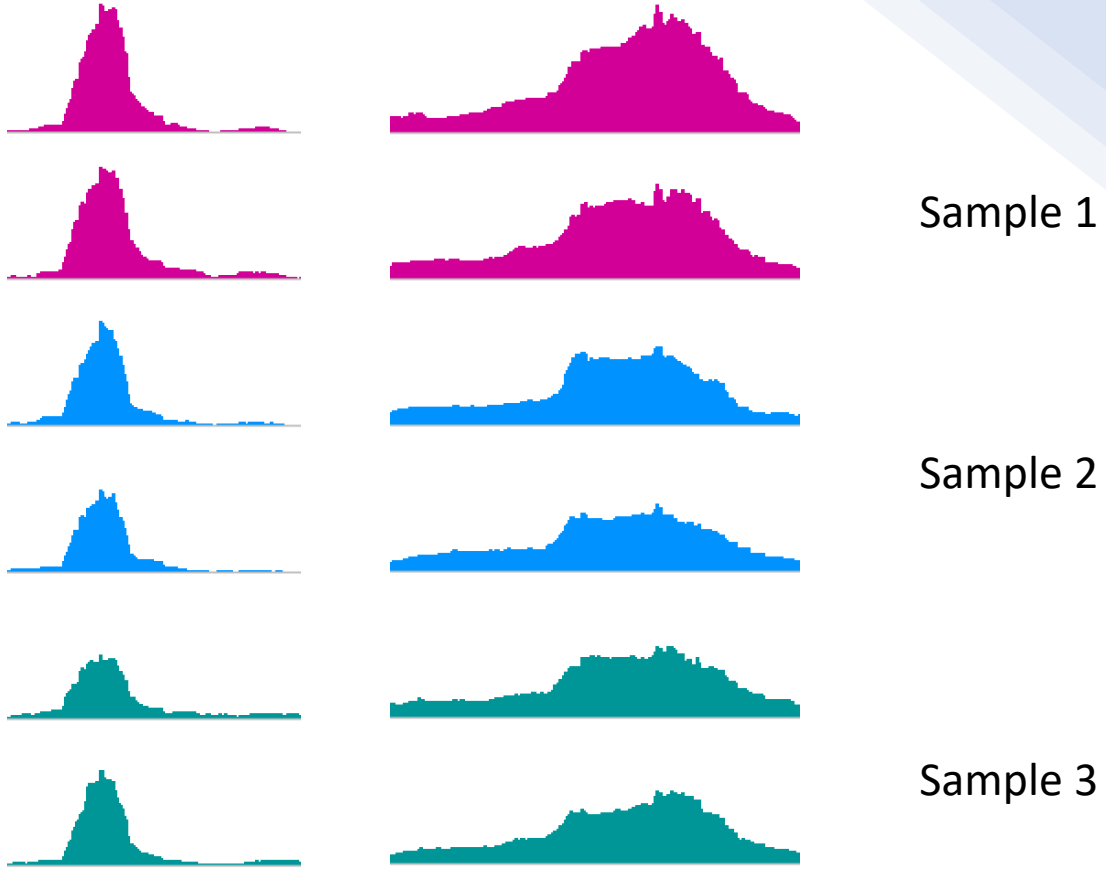


Peak callers

- Since the emergence of ATAC-seq, new peak callers were designed with the identification of peak “shape” in mind.

Peak callers

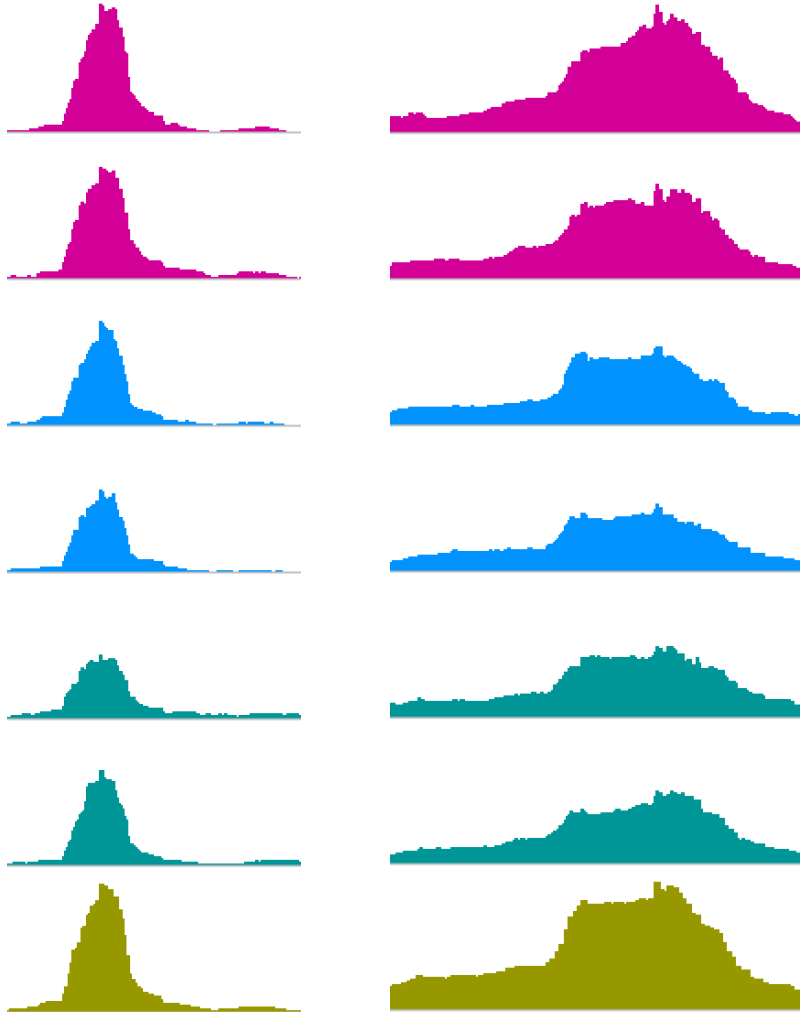
- YAPC (yet another peak caller) computes the second derivate of the aggregated signal from multiple bigwigs, then find concave regions (corresponding to peaks).



Jänes et al., eLife 2018

Peak callers

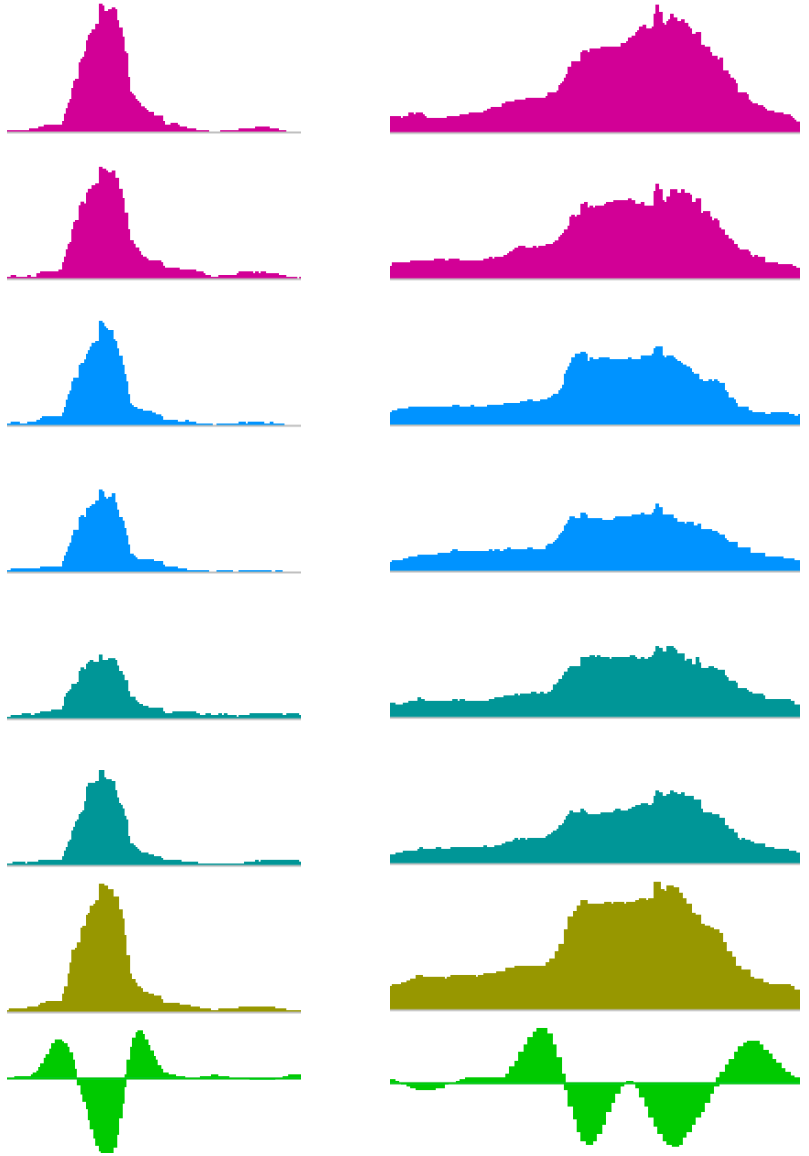
- YAPC (yet another peak caller) computes the second derivate of the aggregated signal from multiple bigwigs, then find concave regions (corresponding to peaks).



Jänes et al., eLife 2018

Peak callers

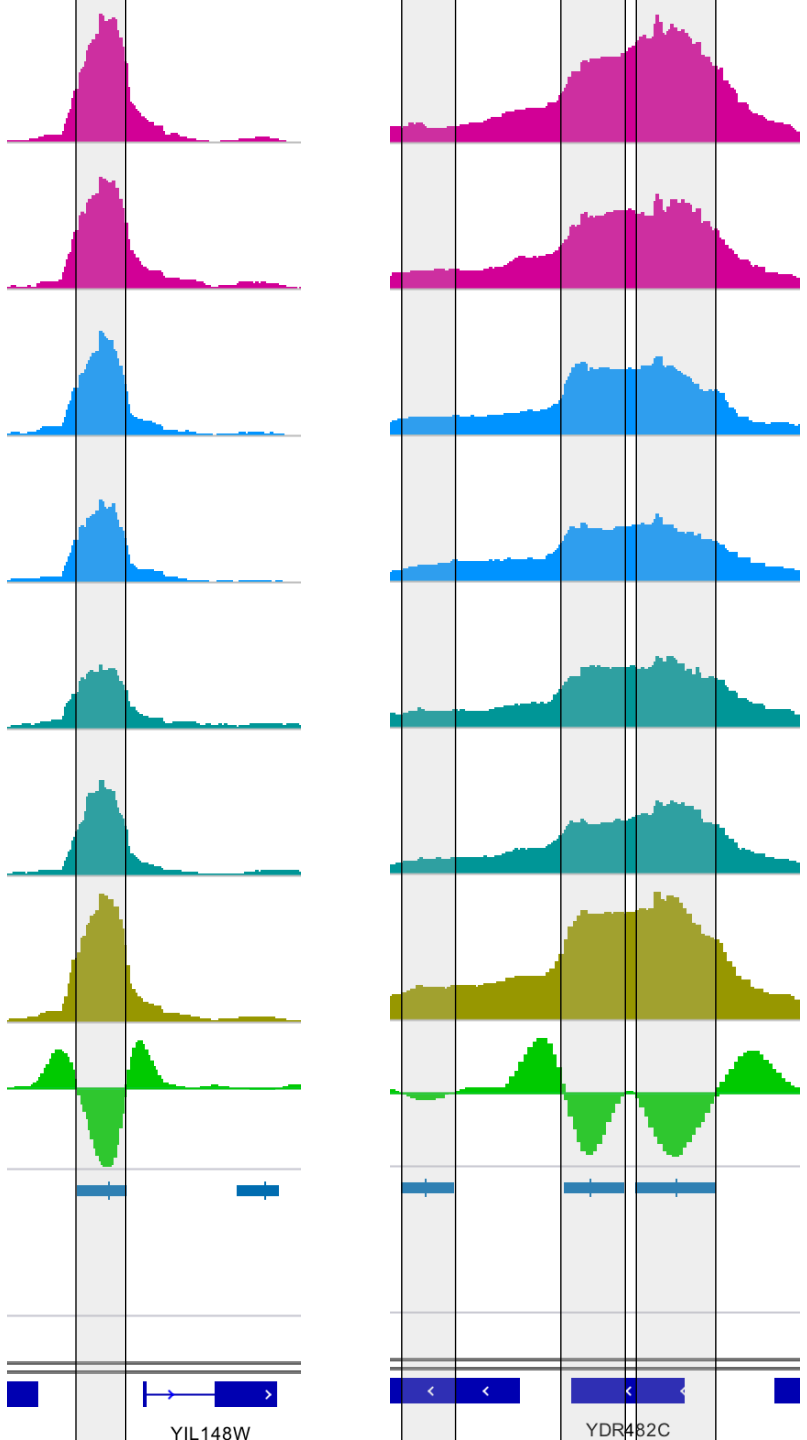
- YAPC (yet another peak caller) computes the second derivate of the aggregated signal from multiple bigwigs, then find concave regions (corresponding to peaks).



Jänes et al., eLife 2018

Peak callers

- YAPC (yet another peak caller) computes the second derivate of the aggregated signal from multiple bigwigs, then find concave regions (corresponding to peaks).



Sample 1

Sample 2

Sample 3

Aggregated signal

2nd derivative

Peaks

Gene annot.

Jänes et al., eLife 2018

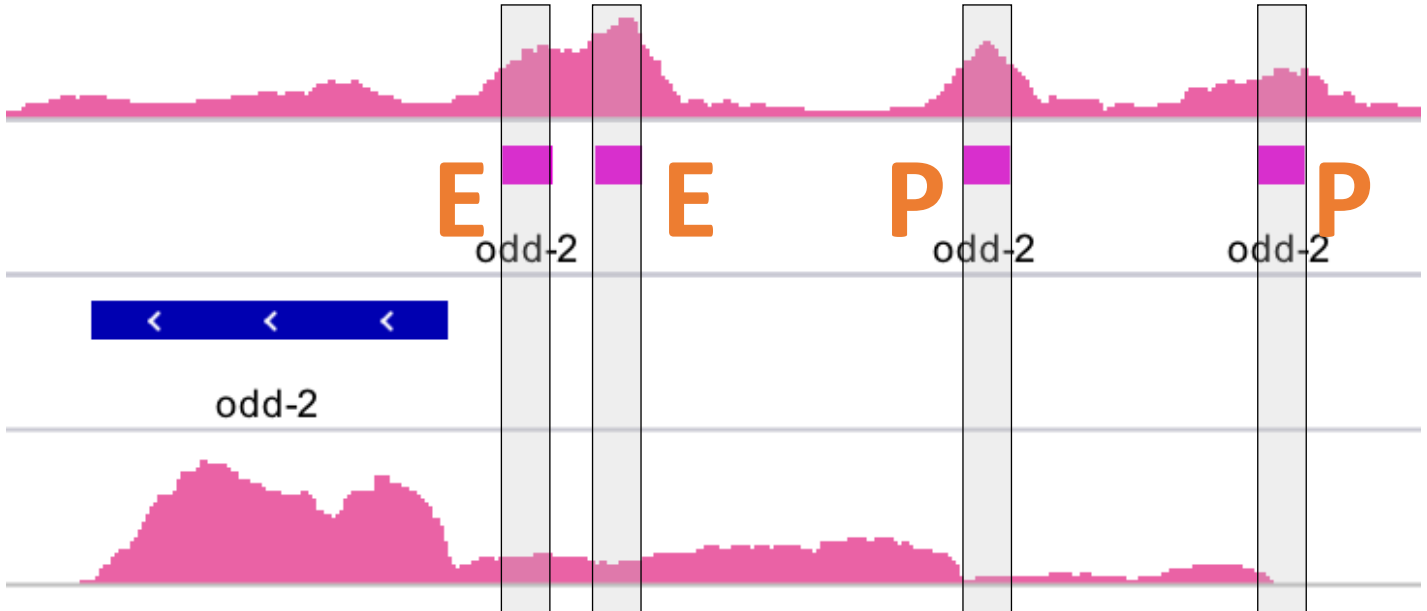
Annotating regulatory elements

- Annotating the function of accessible chromatin loci (e.g. promoters or enhancers) is not straightforward

Annotating regulatory elements

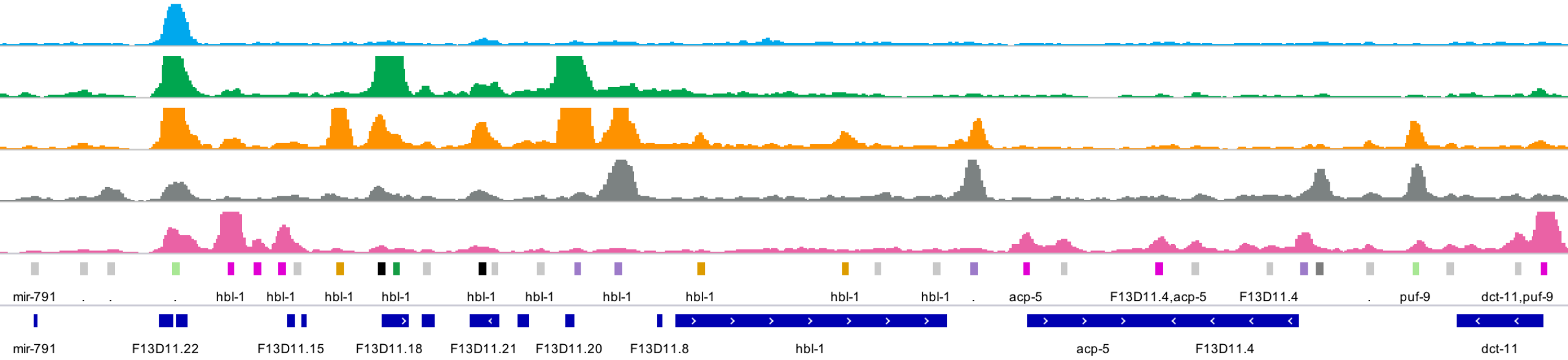
- Annotating the function of accessible chromatin loci (e.g. promoters or enhancers) is not straightforward

- Often, relying on complementary RNA-seq data is helpful



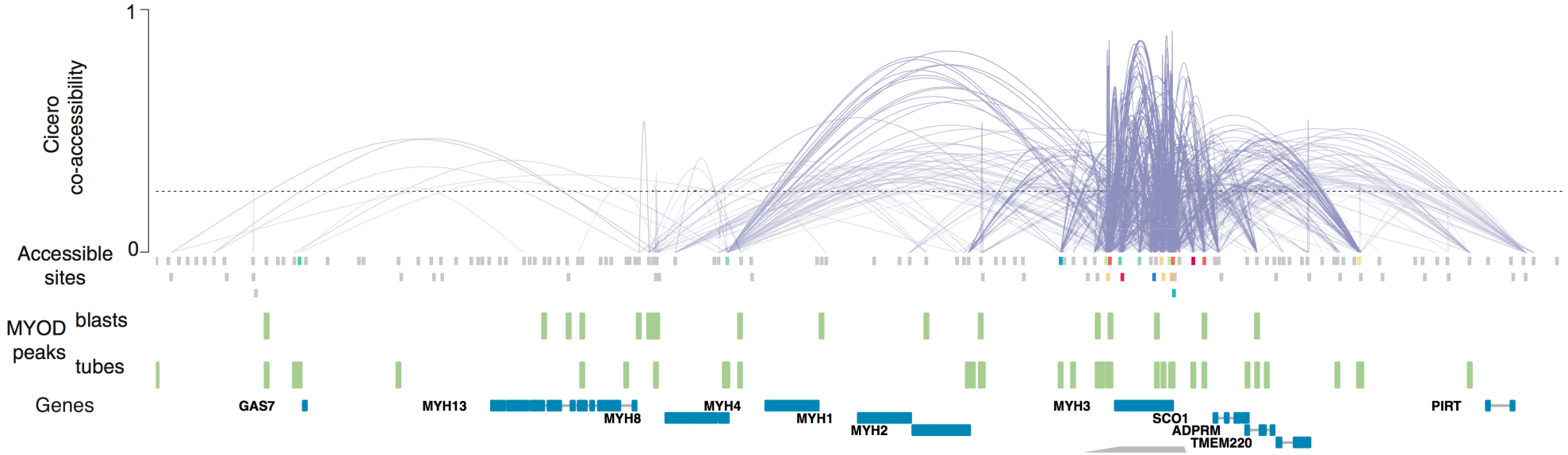
Associating distal regulatory elements to the gene(s) they regulate

- Linking a regulatory element to the gene it regulates could be harder that it might seem



Associating distal regulatory elements to the gene(s) they regulate

- Using single-cell ATAC-seq data, Cicero can be used to link distal elements to their regulated genes



Pliner et al., Mol. Cell 2018