



UNIVERSITY OF
CAMBRIDGE

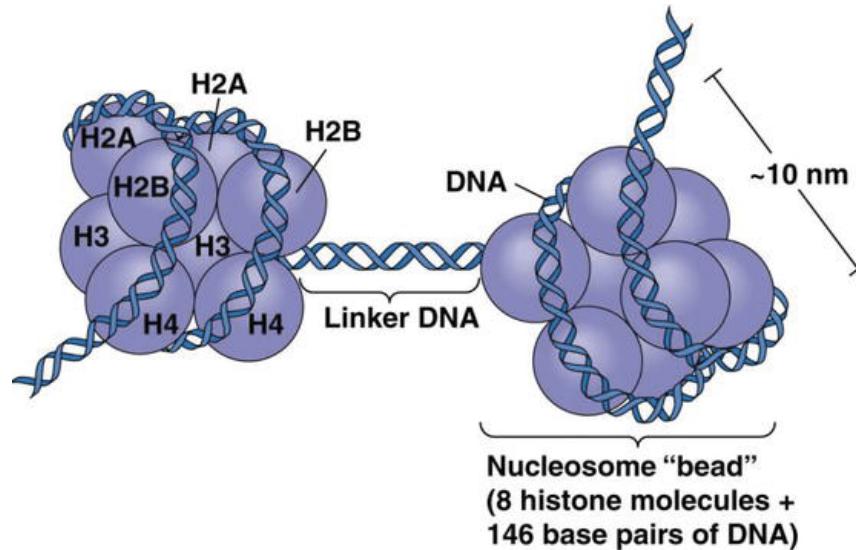
DETECTING AND
ANALYZING OPEN
CHROMATIN REGIONS

Denis Seyres, NHSBT (ds777@medschl.cam.ac.uk)

Chromatin accessibility

Chromatin accessibility

- Nucleosome - Spiral-shaped region in DNA consisting of 147bp tightly wrapped around a core octamer of histones.
- But ~200 bps of DNA in total are associated with the nucleosome.
- Not static – can slide on the genome and modify DNA exposure
- Linker DNA – Nucleosome-free regions (NFRs) or accessible chromatin



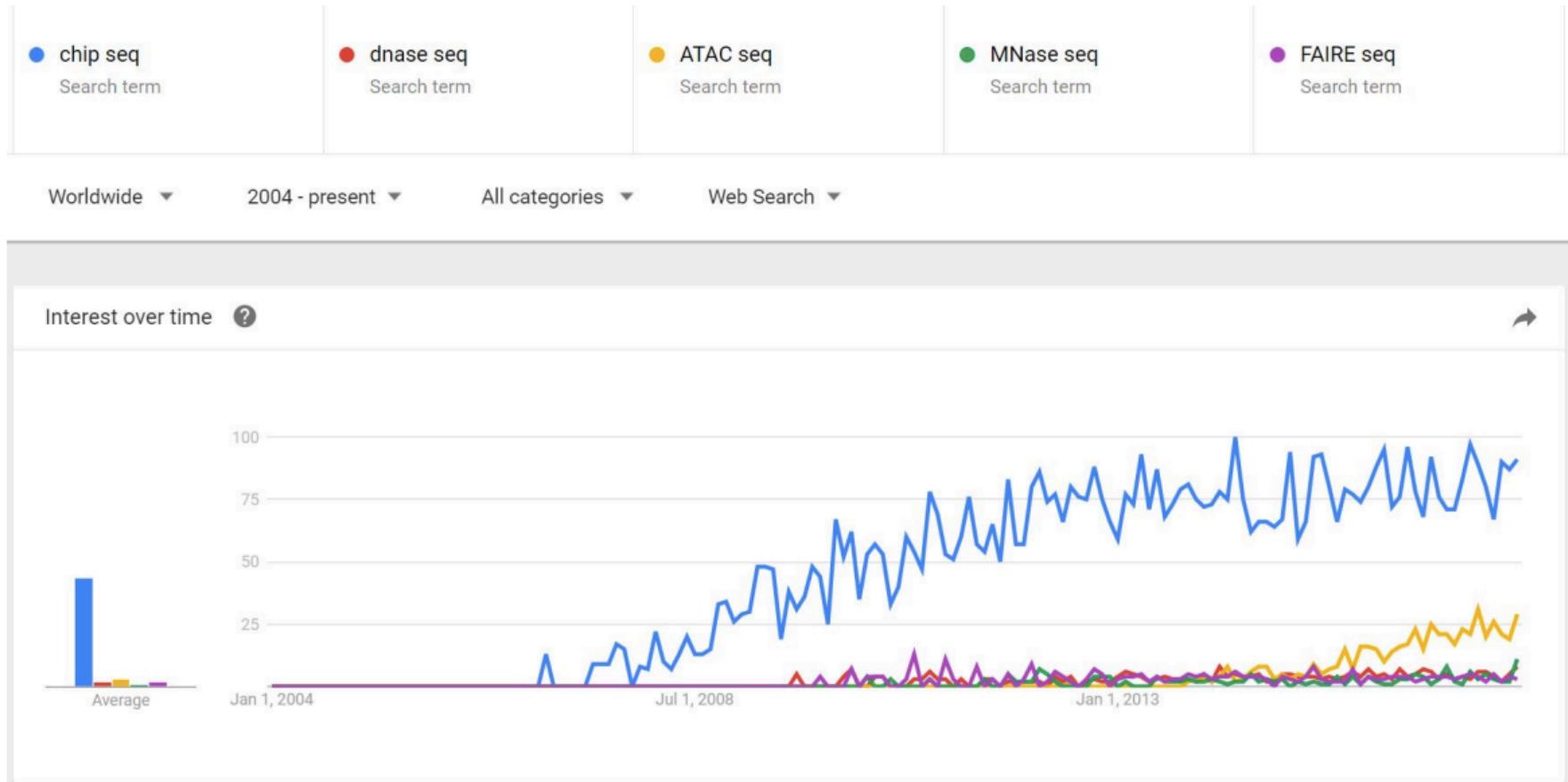
Why chromatin accessibility matters?

- Transcriptionally active chromatin is more “open”
- Direct assays show that it is more accessible to DNases.
- More accessible to components of the transcriptional apparatus:
 - Transcription factors
 - Histone modifiers
 - Chromatin proteins
 - ...

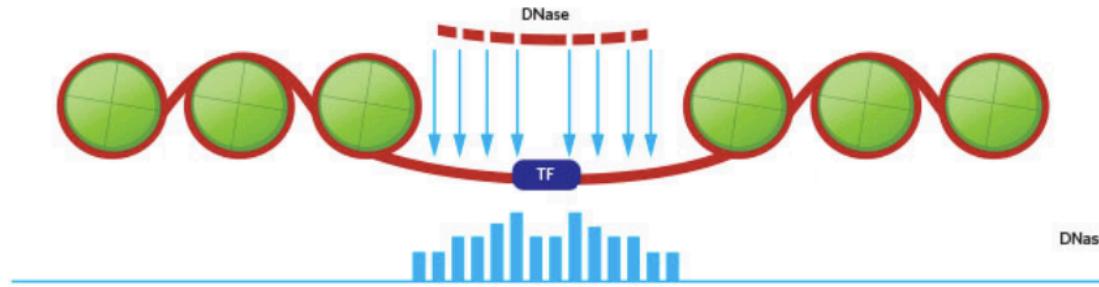
Methods to capture accessible chromatin

1. **DNase-seq** (DNase I hypersensitive sites Sequencing) uses the sensitivity of accessible regions in the genome to the DNase I
2. **FAIRE-seq** (Formaldehyde-Assisted Isolation of Regulatory Elements sequencing) uses the chemical properties of protein-bound DNA in a two-phase separation method
3. **ATAC-seq** (Assay for Transposable Accessible Chromatin sequencing) uses the Tn5 transposase to integrate (synthetic) transposons into accessible regions of the genome
4. **MNase-seq** (Micrococcal Nuclease sequencing) uses the micrococcal nuclease enzyme to identify nucleosome positioning throughout the genome

Popularity of the methods



DNase-seq



DNase-seq has historically been a valuable tool for identifying all different types of regulatory elements, including promoters, enhancers, silencers, insulators and locus control regions.

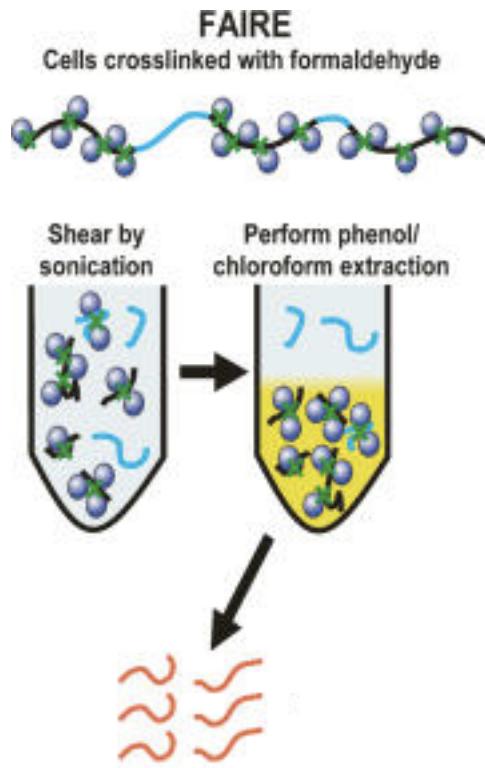
DNase-seq is better established than any of the other chromatin accessibility methods.

It was applied to a wide range of cell types and species, including plants.

Its cutting bias is better understood.

DNase-seq could be used for analysis of protected regions of the genome (DNase footprinting)

FAIRE-seq



Antibody and enzyme independency.

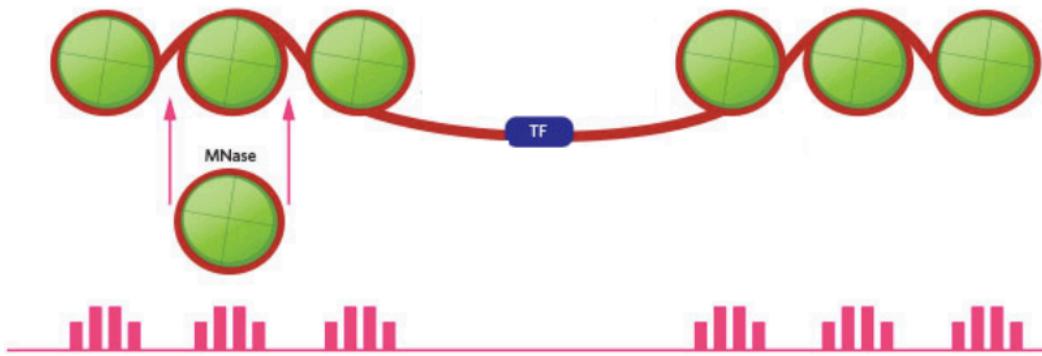
In contrast to DNase-seq, the FAIRE-seq protocol doesn't require the permeabilization of cells or isolation of nuclei, and can analyze any cell types

Analysis. Extensive amount of computational processing and analysis is required for comprehensive interpretation of genome-wide results.

Absence of transcription factor footprinting.
Low signal-to-noise ratio.
Fixation variation among tissues.

Adapted from Giresi et al., Genome Res. 2007

MNase-seq

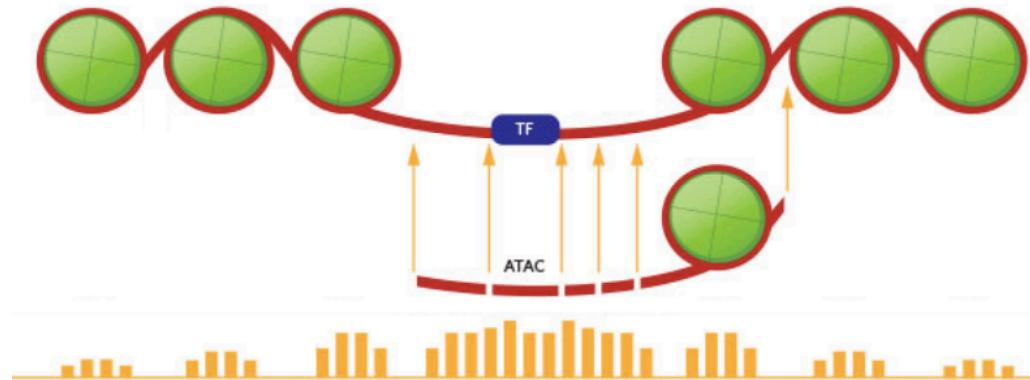


MNase works by cutting both exposed DNA strands of the genome; the DNA associated with nucleosomes is recovered and sequenced. Higher concentration of enzyme leads to a majority of mono-nucleosomes.

Conceptually, reverse of DNA-seq and ATAC-seq. So it indirectly identifies open-chromatin regions of the genome.

Enzymes used since 40 years!

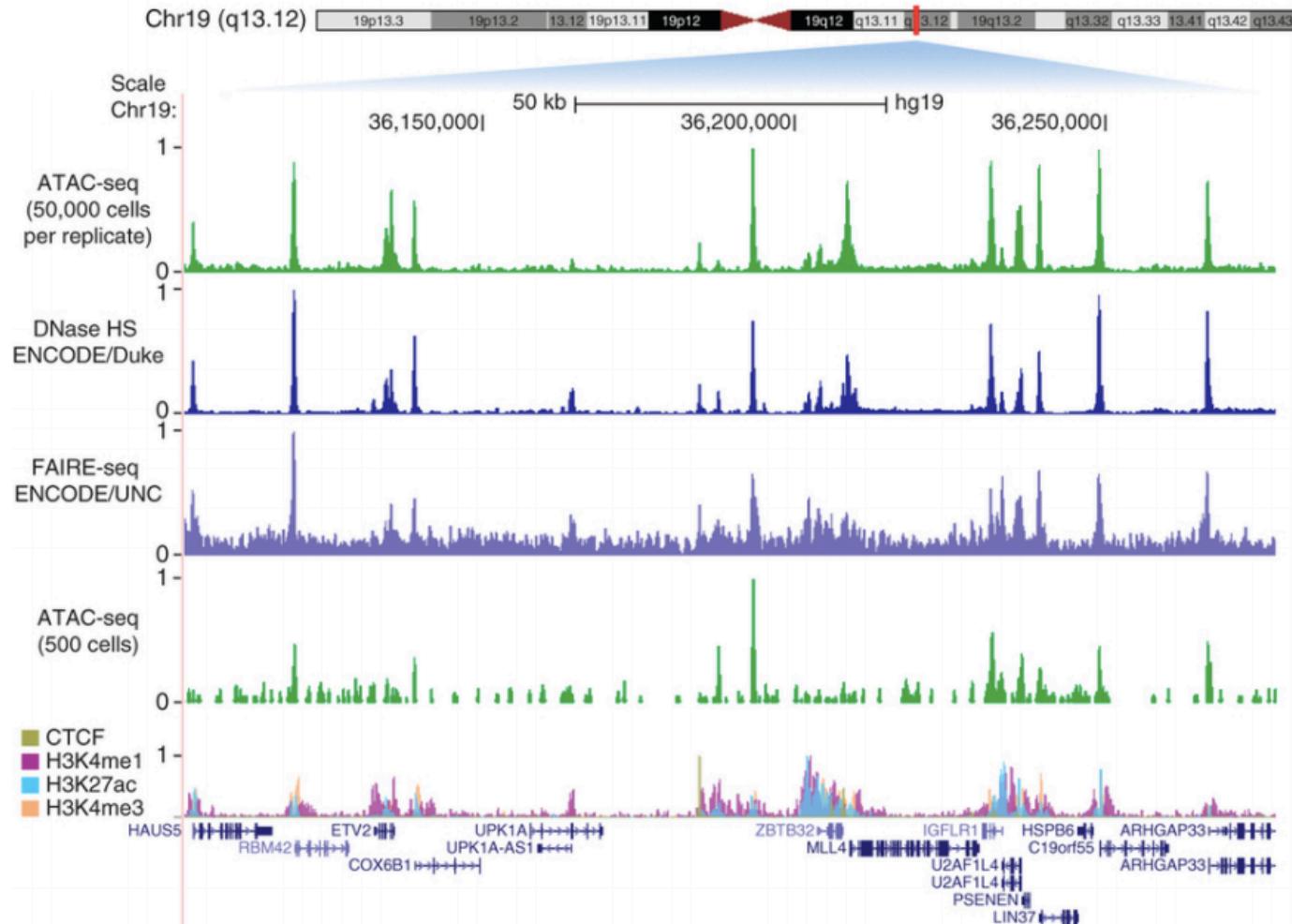
ATAC-seq



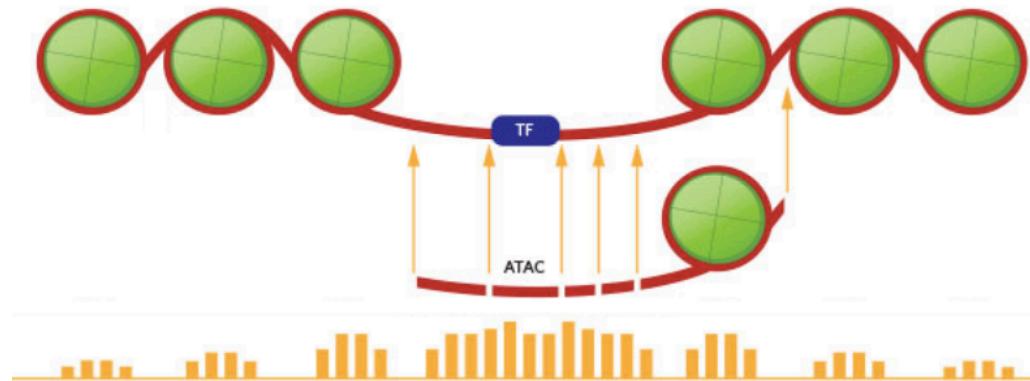
ATAC-seq simultaneously cuts and ligates sequencing adapters preferentially in nucleosomes-free regions using the enzyme Tn5 transposase. The DNA bits captured between the adapters are then amplified and sequenced.

50,000 cells are sufficient for this technique, as opposed to others like MNase-seq or DNase-seq that require at least 1,000-fold more material.

ATAC-seq



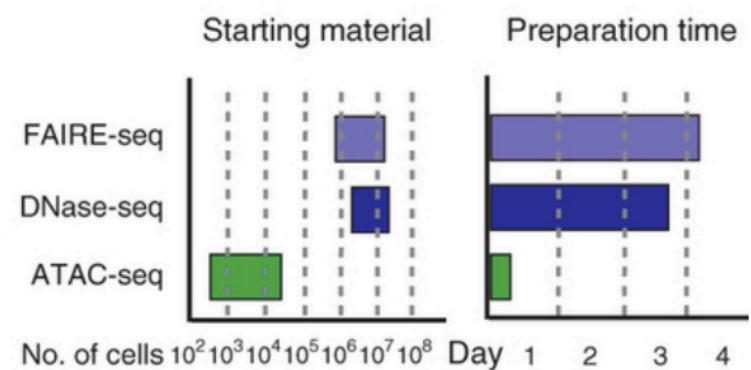
ATAC-seq



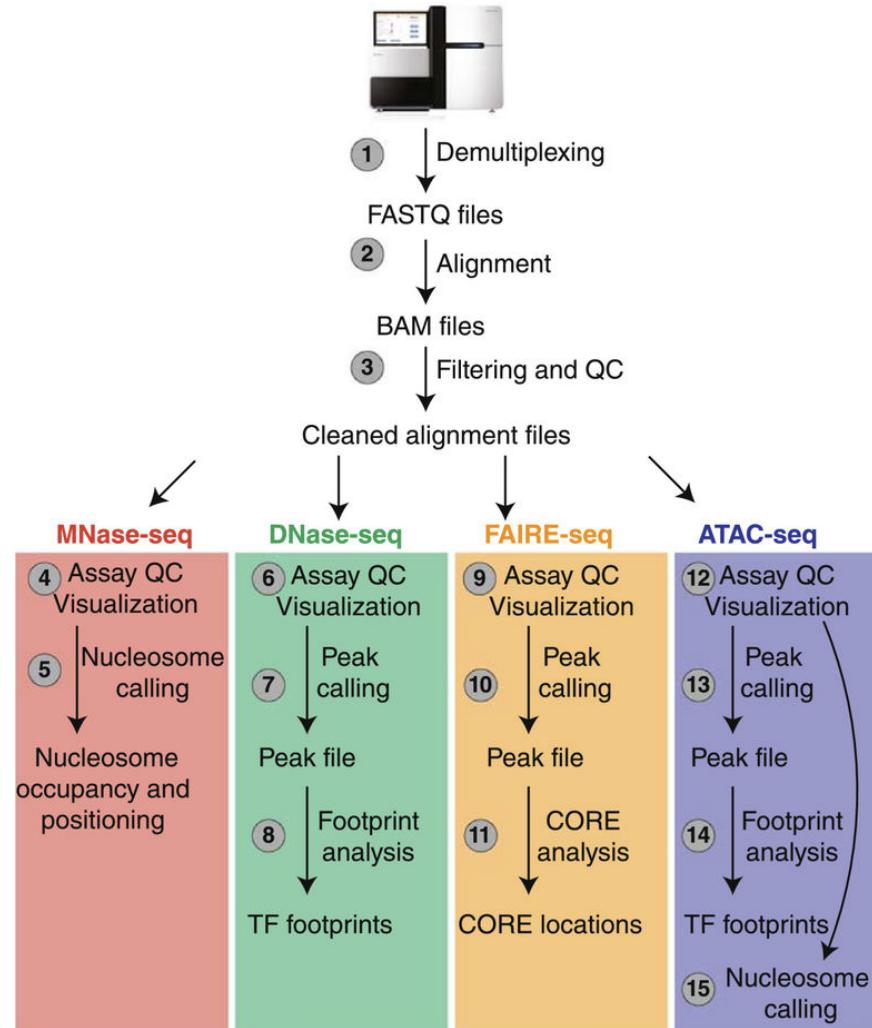
Low requirements on the amount of the biological sample.

Speed: The whole protocol requires 3 hours in total.

High signal-to-noise ratio

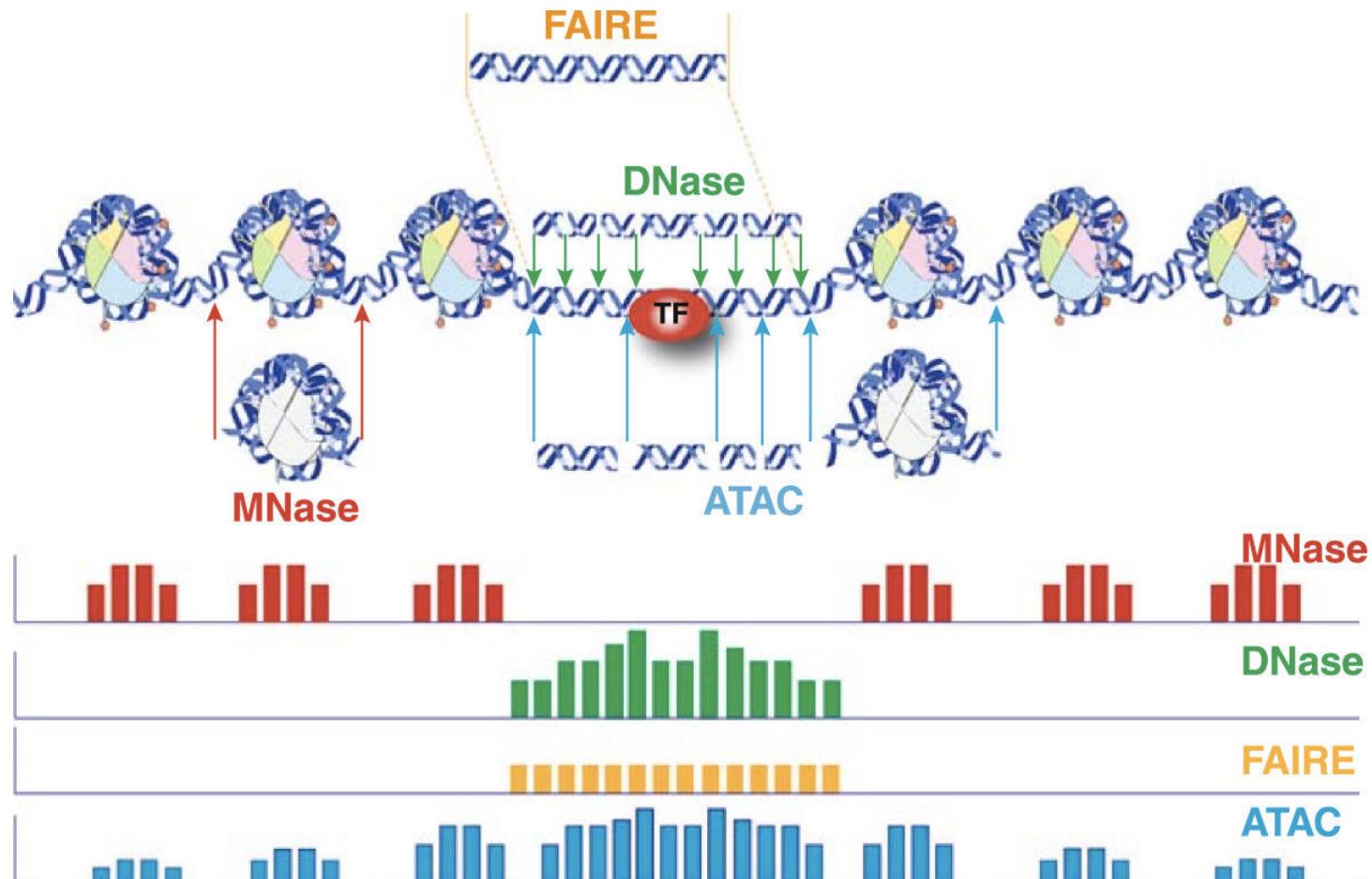


ChIP-seq-based workflows



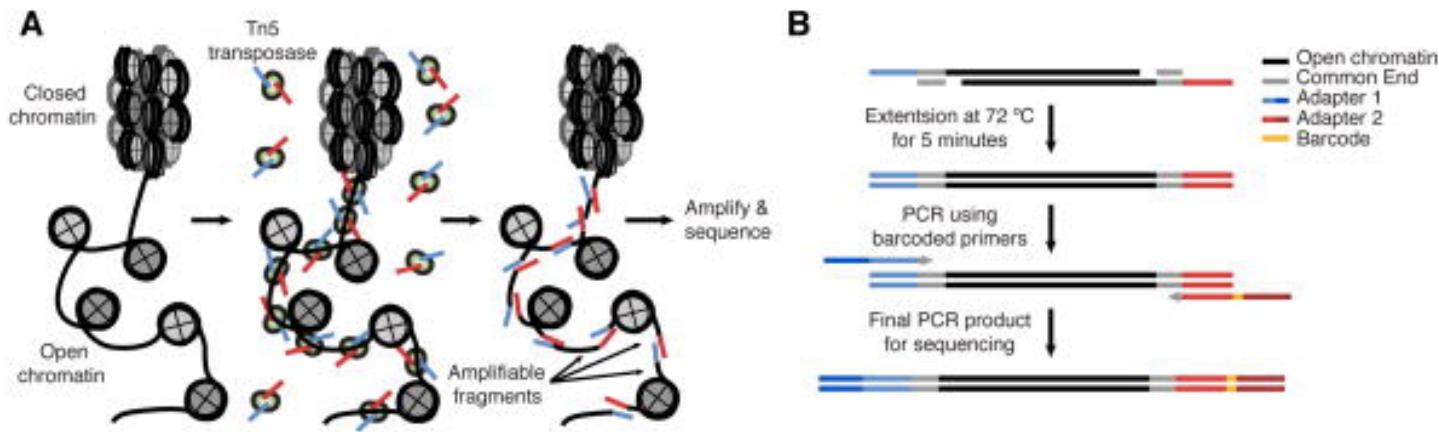
Adapted from Tsompana and Buck, Epigenetics & Chromatin 2014

Methods overview



ATAC-seq data processing

Experimental procedure and design



From Buenrostro et al. *Curr Protoc Mol Biol*. 2015

Replicates

To begin with, two replicates per experimental group are sufficient.

Controls

Not typically run, but it would be similar ChIP control (input). Useful to help define regions of the genome that are more challenging to sequence or to align reads unambiguously.

PCR amplification

As for ChIP-seq, as few PCR cycles as possible

Sequencing depth

In human, it is recommended to have 40-50 million mapped reads per sample. For transcription factor foot-printing >200M mapped reads is recommended (Neph et al., 2012)

Sequencing mode

Paired-end highly recommended but single-end works also well (with limitations!)

Read processing

ATAC-seq library preparation generally uses the Illumina Nextera library prep kit. Read trimming needs to be performed according to these adapters which appear clearly over-represented in FastQC step.

Illumina Nextera Adapters

Nextera Transposase Adapters

(Used for Nextera fragmentation)

Read 1

5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

Read 2

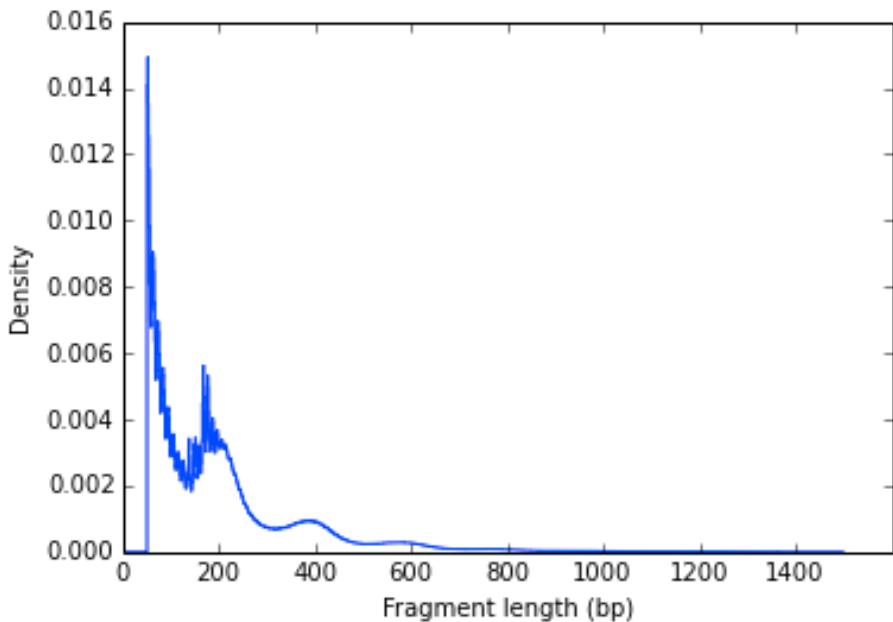
5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Read mapping to reference genome does not require particular options. It can be done with all tools available for ChIP-seq (Bowtie, BWA...). ATAC-seq analysis requires removal of PCR duplicates reads, as for ChIP-seq.

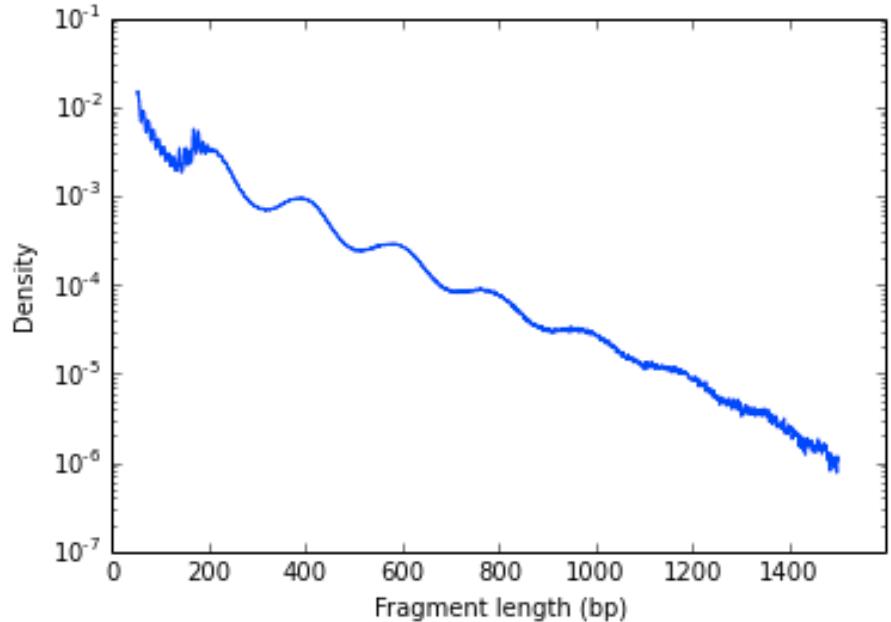
Quality control

Fragment size density of the libraries

Regular scale



Log scale



Sharp peak at <100 bp region (open chromatin), a peak at ~200bp region (mono-nucleosome), and other larger peaks (multi-nucleosomes)
Clear nucleosome phasing pattern indicates a good quality of the experiment

ATAC-seq data analysis

Peak-calling

Nucleosome positionning

Peak calling

ATAC-seq techniques doesn't provide properly speaking an 'input' as for ChIP-seq.

Peak callers like MACS2, which perform best when they work with an input file, may encounter some difficulties with ATAC-seq data.

Otherwise, MACS2 is the most used tool in ATAC-seq analysis. Playing with parameters like --shift and --extsize with --nomodel can provide more narrow peaks that fits best inter-nucleosomal regions.

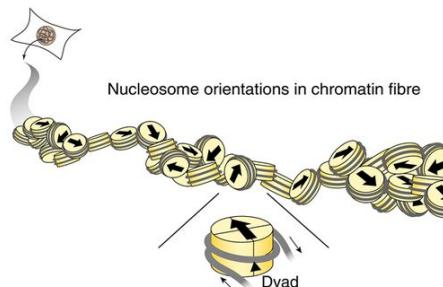
Other methods (Fseq, ZINBA) analyze density variation of the reads rather than trying to perform a comparison to input. These tools provide in general broad regions.

Best practices are to call peaks with different tools and compare.

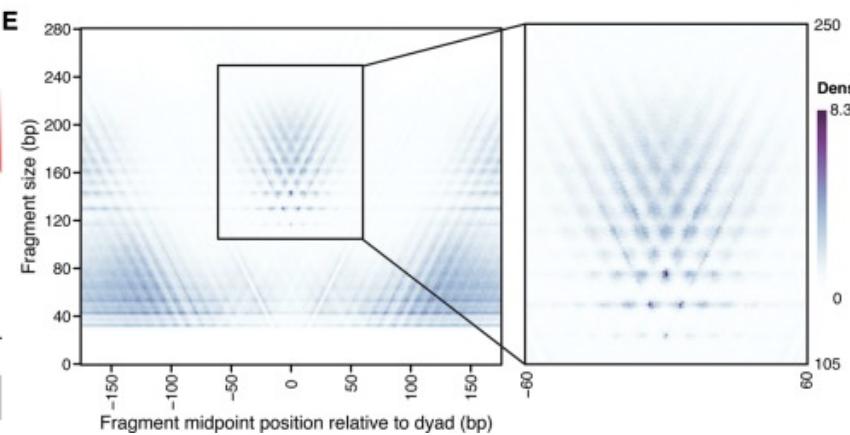
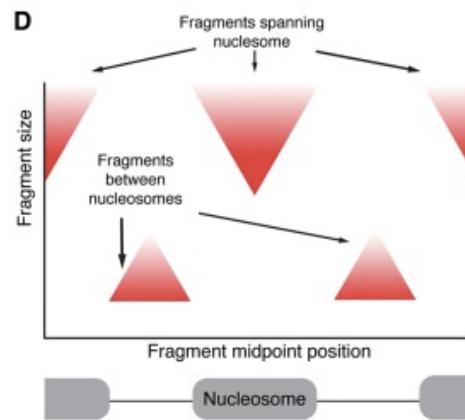
This 'naive' peak calling could be followed by a more accurate, refined, nucleosome-free regions identification based on nucleosome positioning analysis.

Nucleosome positioning – V plots

V-plots demonstrate structured ATAC-seq signal around nucleosomes. It plots the density of fragment sizes versus fragment center locations relative to a genomic feature of interest (in this case, nucleosome dyads)



Dyad position here means the center point of the DNA that would be wrapped around the nucleosome

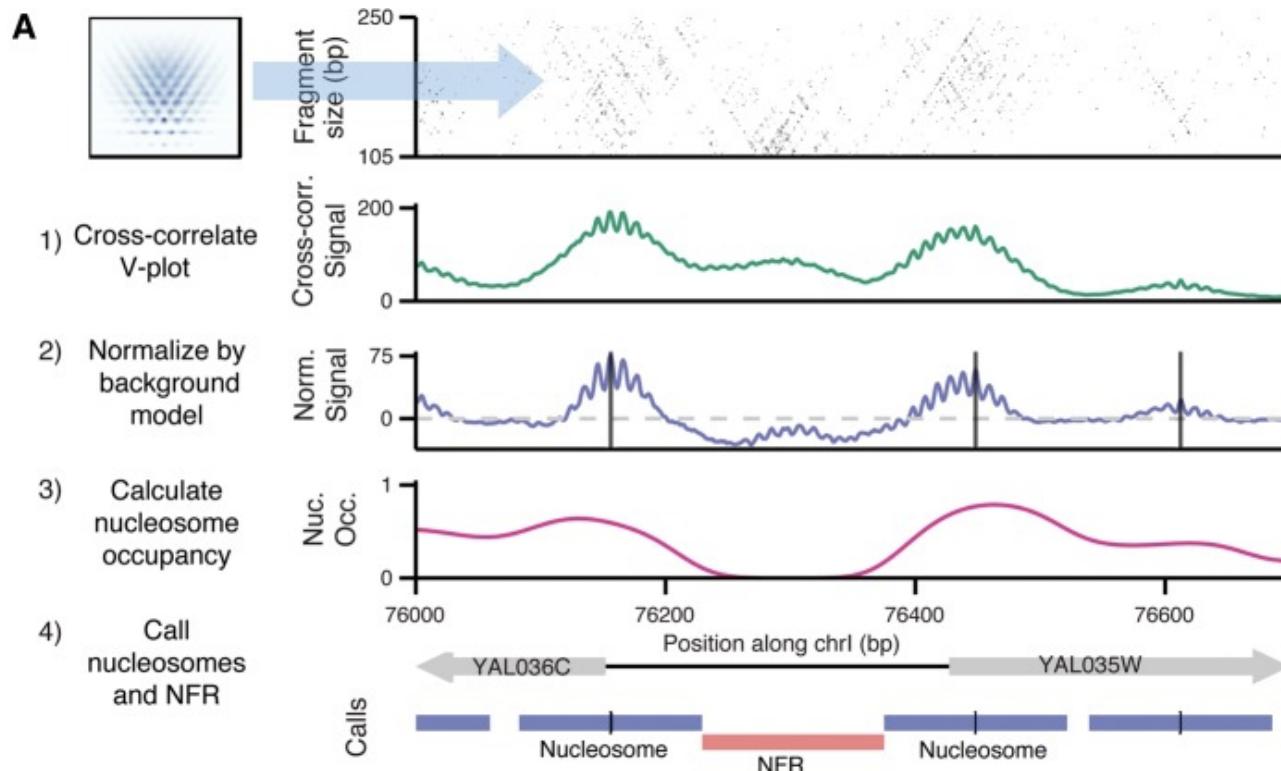


Adapted from Schep et al., Genome Res. 2015

The apex of the "V" represents the smallest possible fragment that spans the DNA protected by a nucleosome (117bp). The most frequent fragment size is ~143bp.

"Short ATAC-seq fragments are concentrated at nucleosome-free regions (NFR), whereas long fragments are enriched at nucleosome-associated DNA" (Buenrostro et al. 2013)

Nucleosome positioning



Hands-on!

Open ATAC-seq analysis tutorial and follow
carefully each steps

