

Assay for Transposase-Accessible Chromatin – sequencing (ATAC-seq)

Epigenomics Data Analysis Workshop

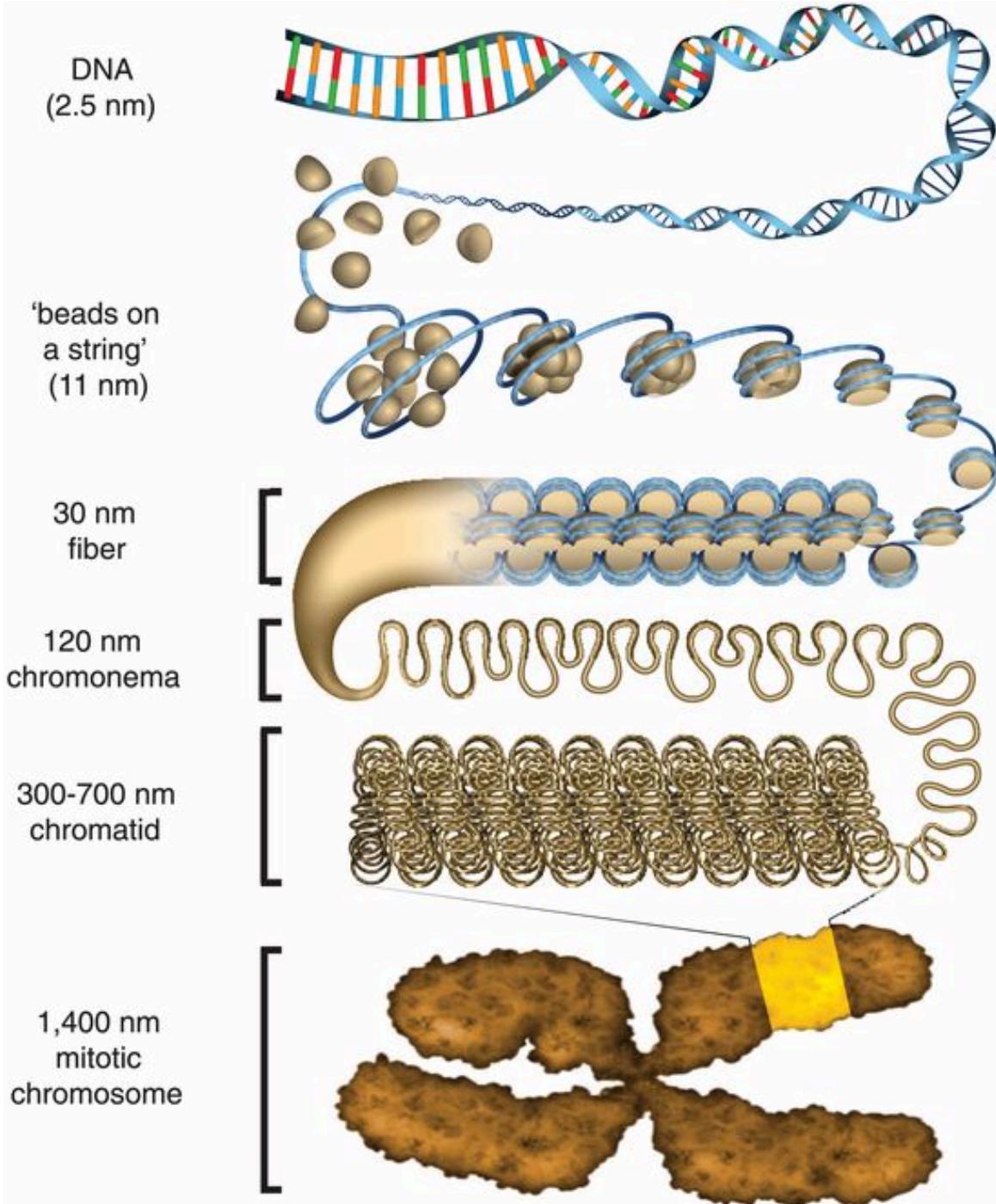
Stockholm, 17 September 2024

Agata Smialowska

NBIS, SciLifeLab, Stockholm University



Chromatin structure



Chromatin structure

ChIP-seq

ATAC-seq, DNaseI-seq,
MNase-seq

3D organisation
Hi-C, Dam-ID

DNA
(2.5 nm)

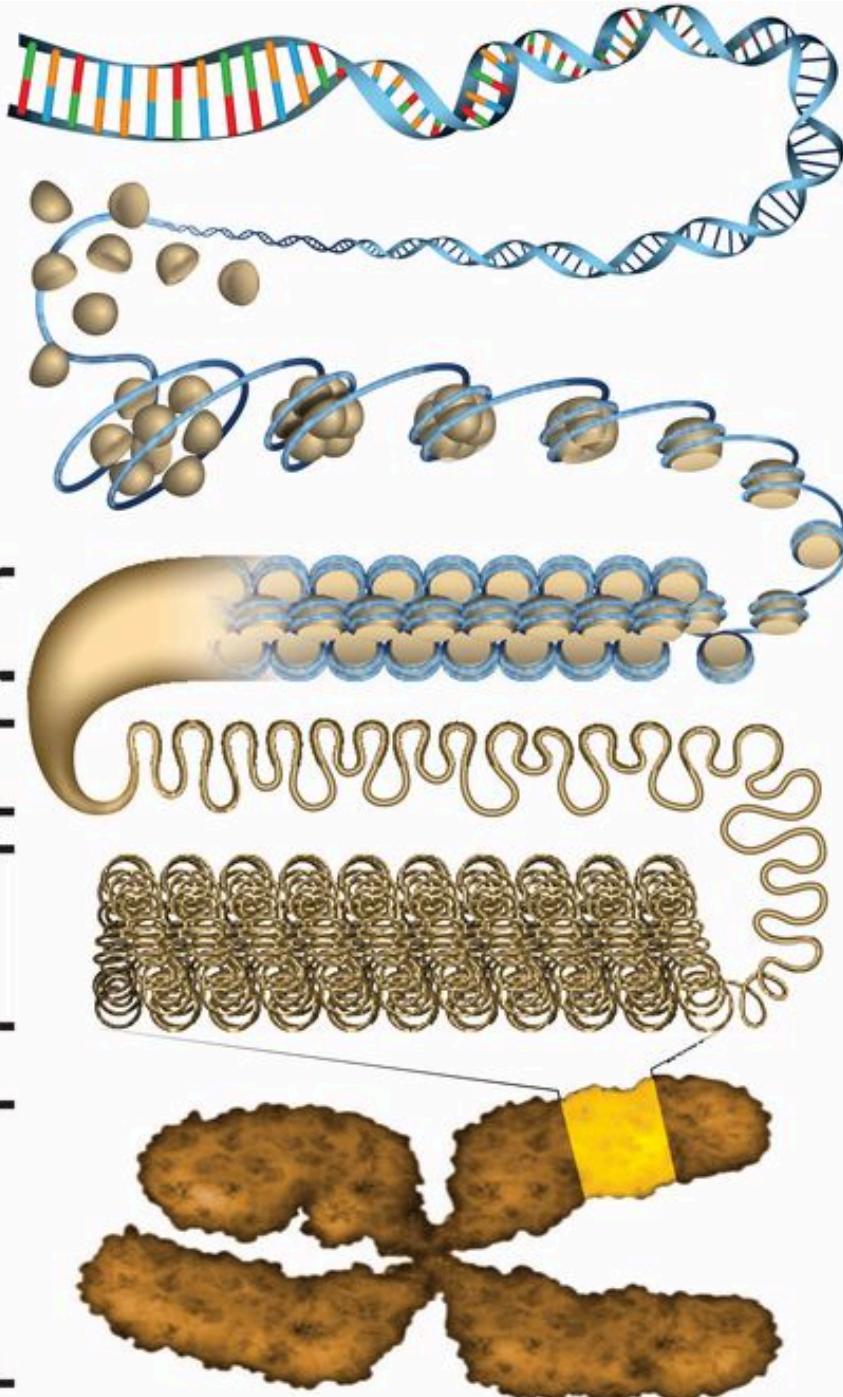
'beads on
a string'
(11 nm)

30 nm
fiber

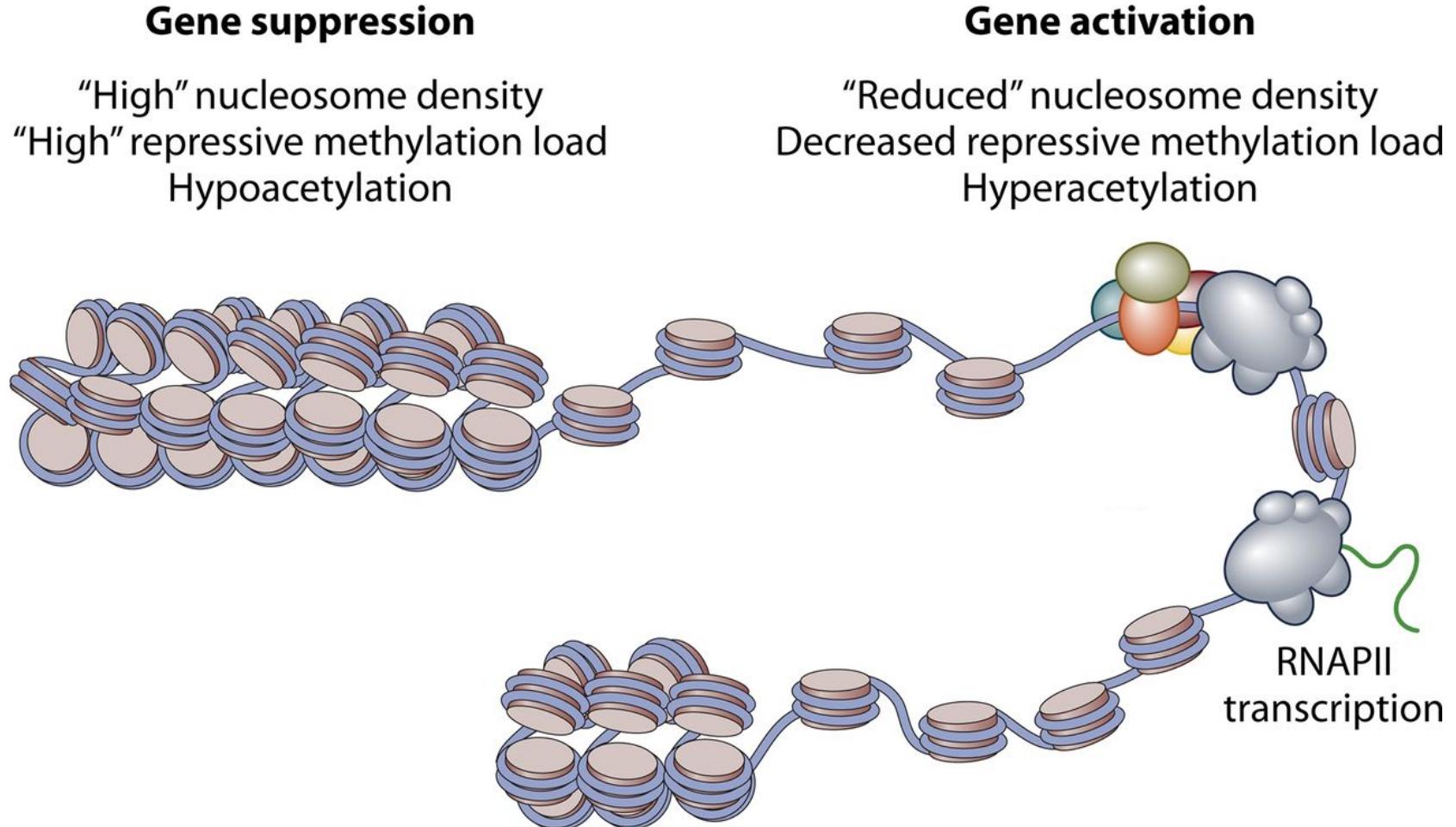
120 nm
chromonema

300-700 nm
chromatid

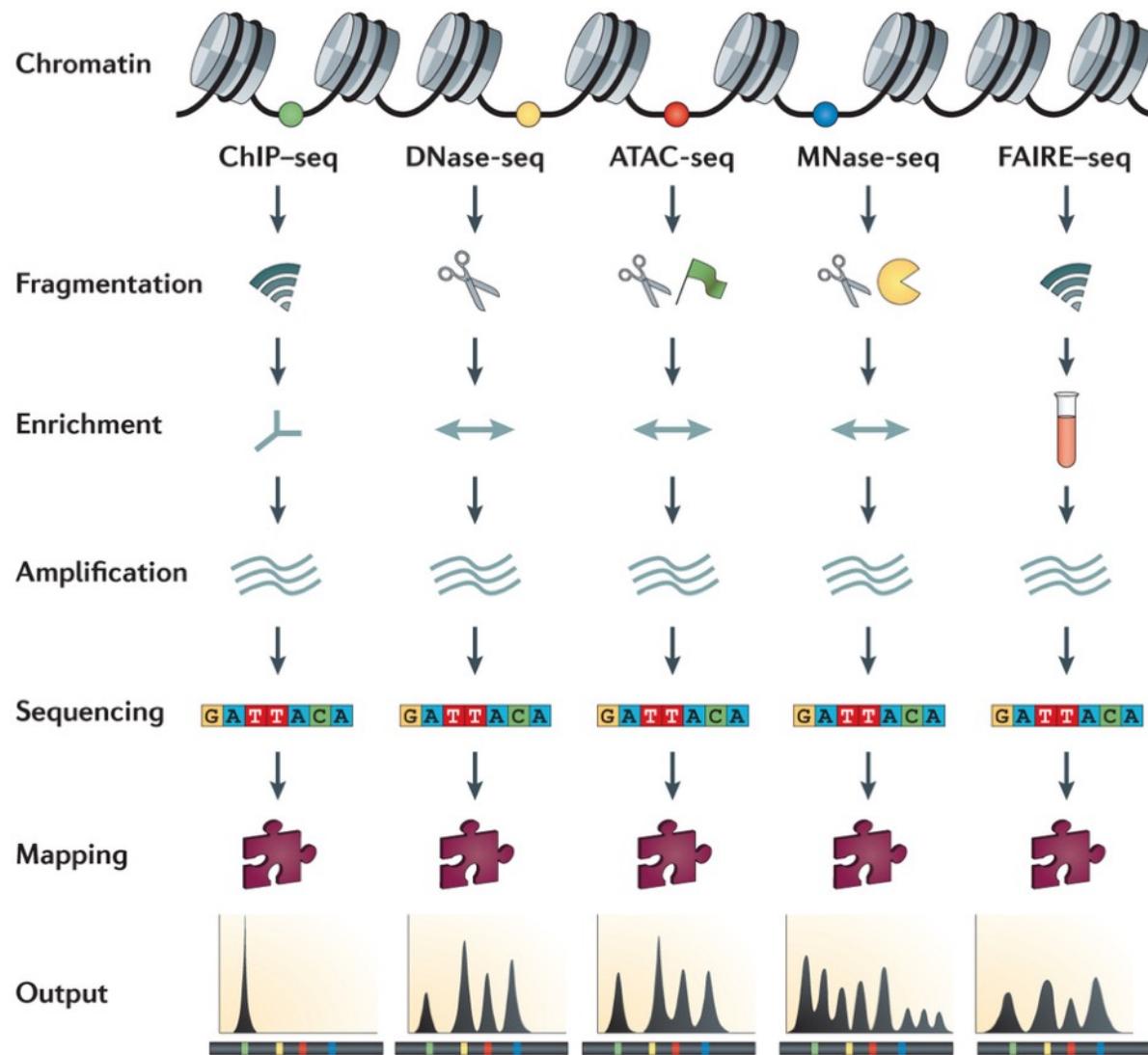
1,400 nm
mitotic
chromosome



Chromatin states and gene expression



Functional genomics techniques to probe chromatin states

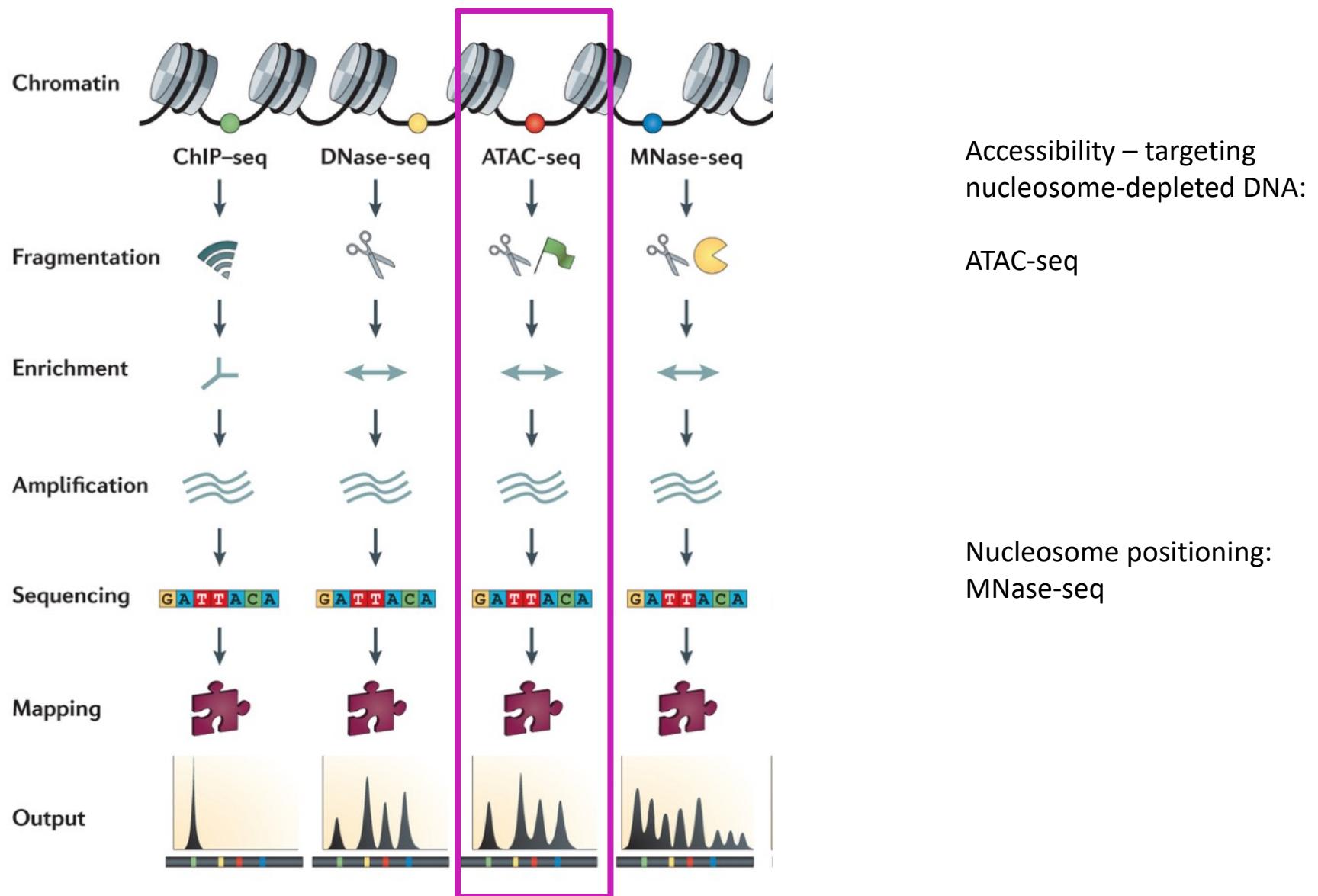


Accessibility – targeting nucleosome-depleted DNA:

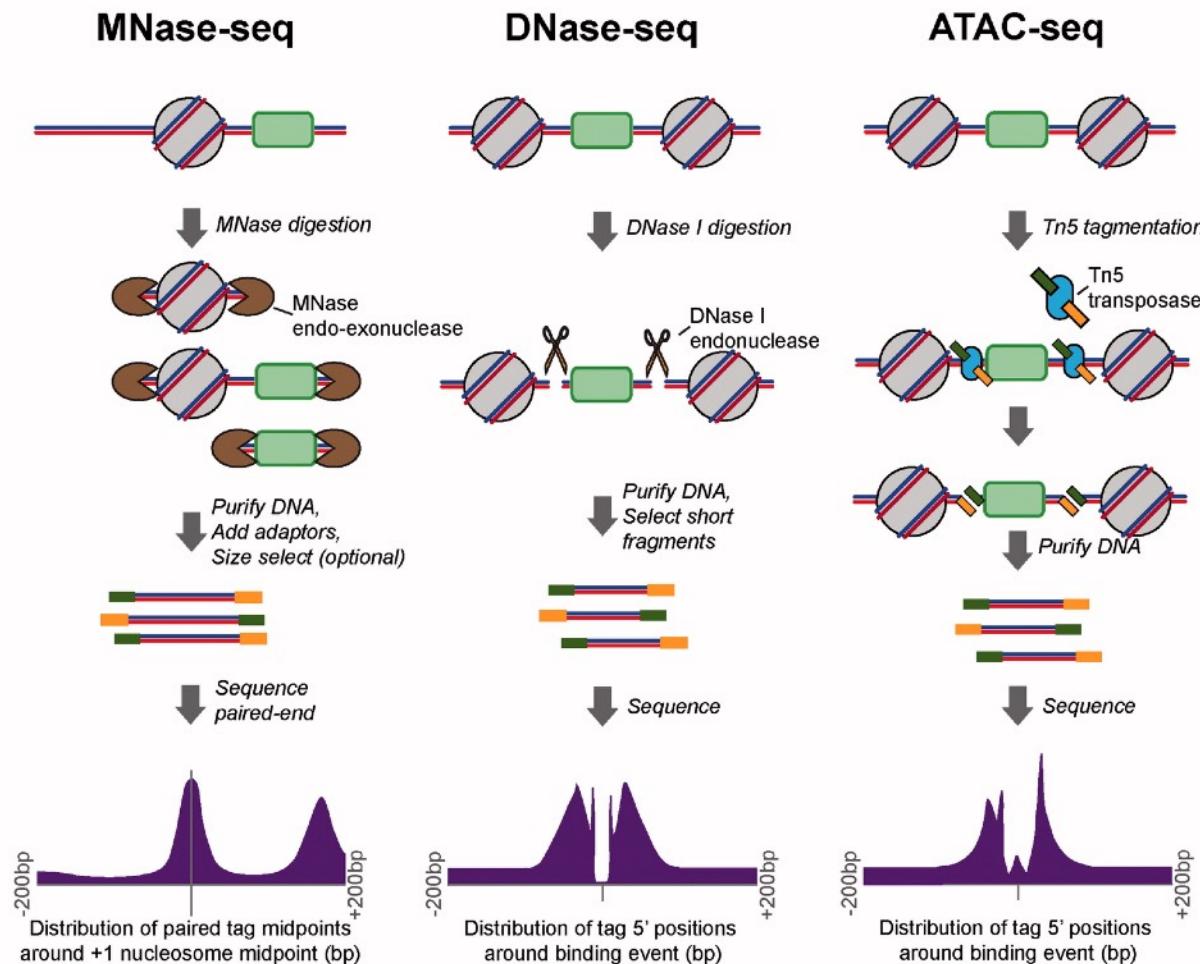
DNase-seq
ATAC-seq
FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements)

Nucleosome positioning:
MNase-seq

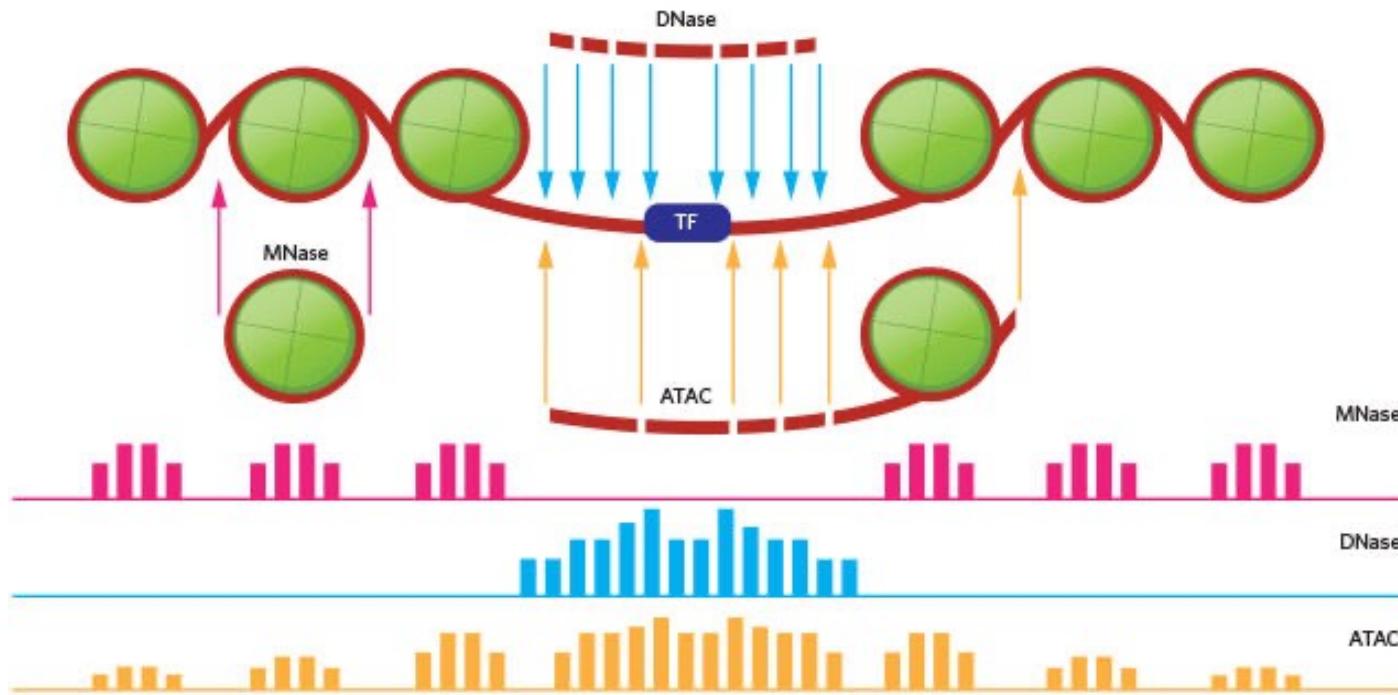
Functional genomics techniques to probe chromatin states



Functional genomics techniques to identify open chromatin regions



Functional genomics techniques to probe chromatin states



Assay for Transposase-Accessible Chromatin (ATAC)-seq

- The method published in bulk (Buenrostro et al., 2015) and single cell (Buenrostro et al., 2015)

Current Protocols in Molecular Biology / Volume 109, Issue 1

UNIT

ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide

Jason D. Buenrostro, Beijing Wu, Howard Y. Chang, William J. Greenleaf

First published: 05 January 2015

<https://doi.org/10.1002/0471142727.mb2129s109>

Citations: 696

Published: 17 June 2015

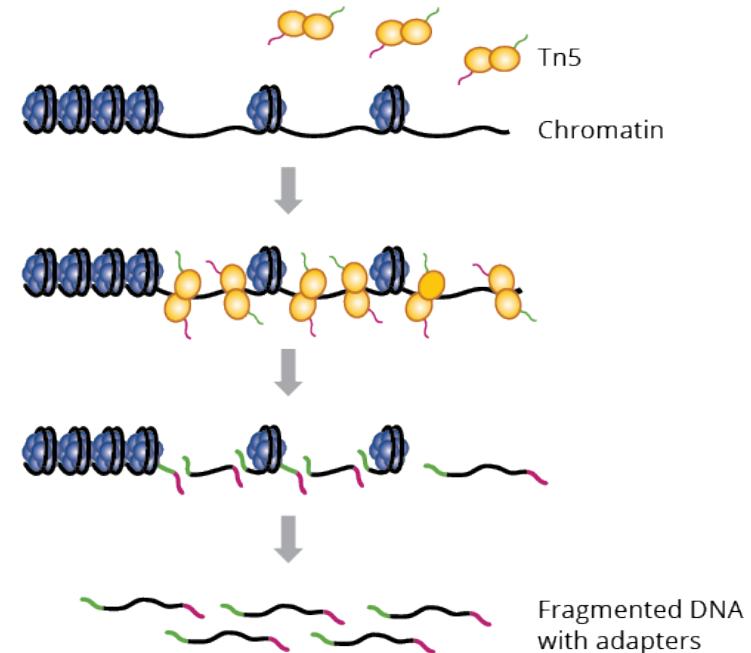
Single-cell chromatin accessibility reveals principles of regulatory variation

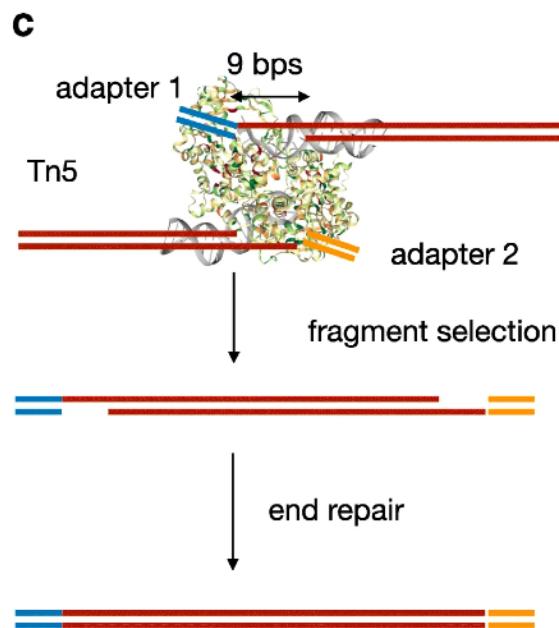
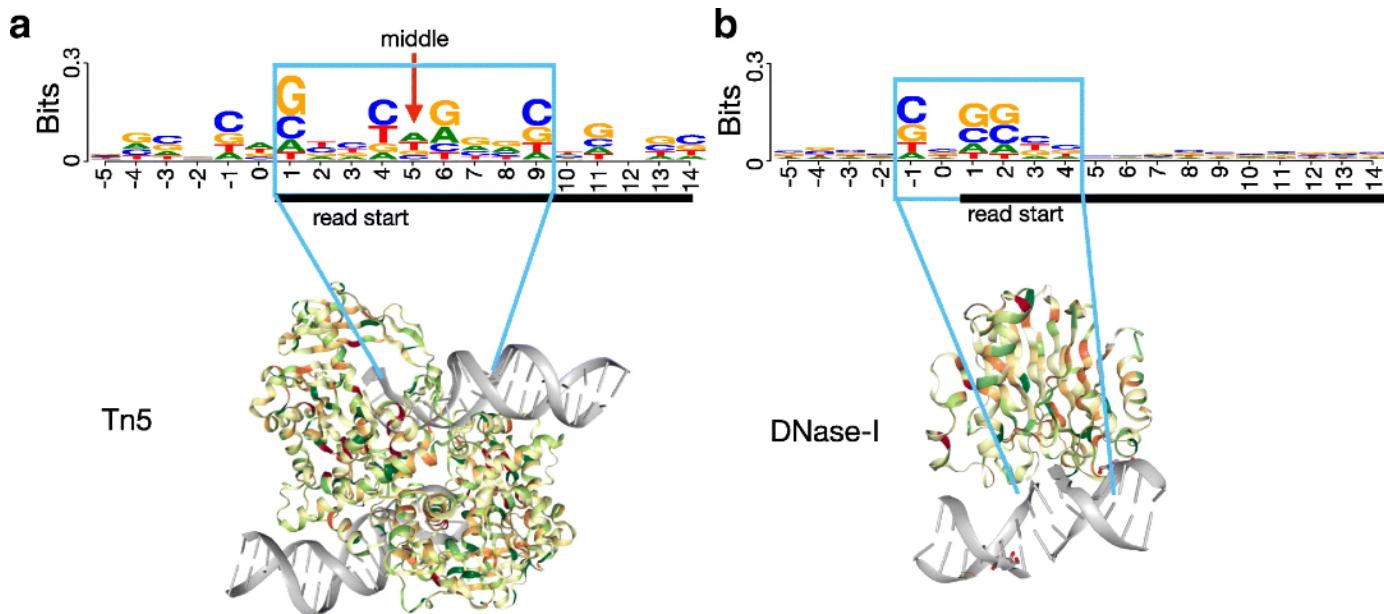
Jason D. Buenrostro, Beijing Wu, Ulrike M. Litzenburger, Dave Ruff, Michael L. Gonzales, Michael P. Snyder, Howard Y. Chang & William J. Greenleaf

Nature 523, 486–490(2015) | [Cite this article](#)

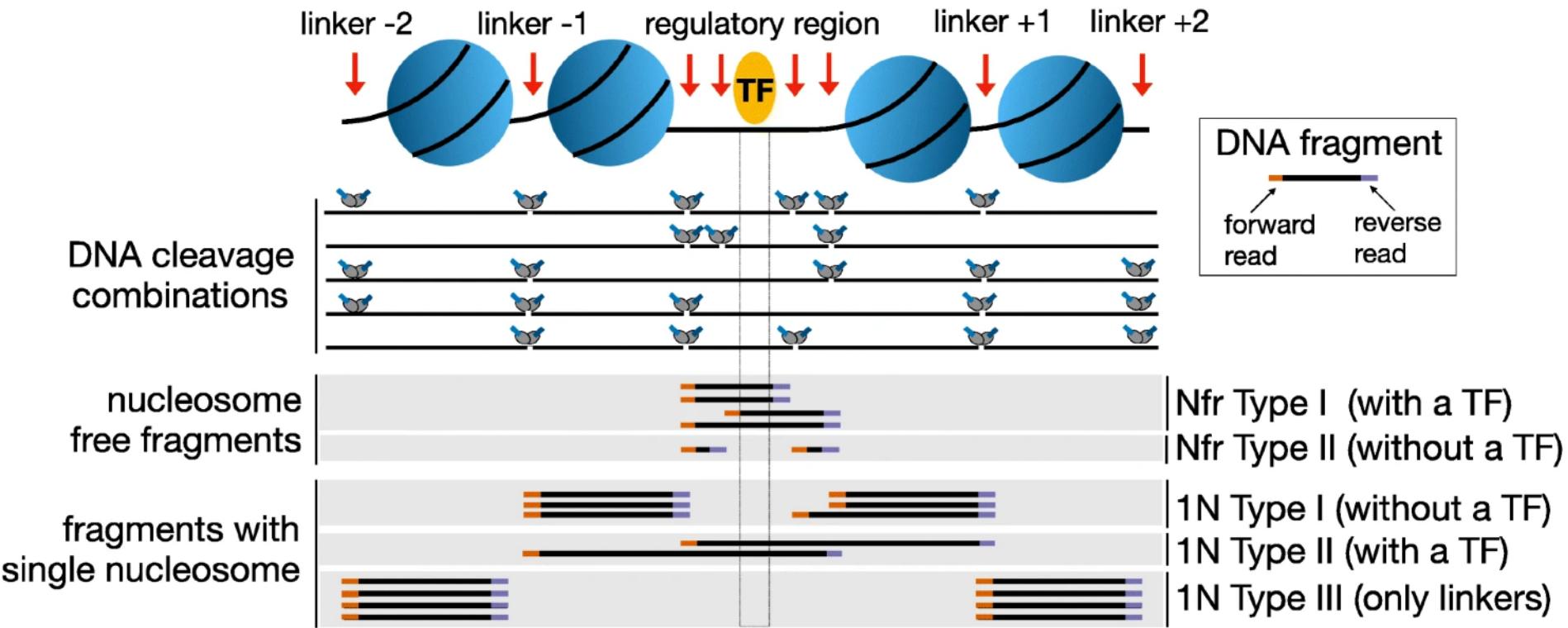
21k Accesses | 600 Citations | 100 Altmetric | [Metrics](#)

- It probes access to chromatin by using Tn5 transposase to insert sequencing adapters into DNA which allows simultaneous fragmentation of chromatin and integration of the adapters into open chromatin regions
- Significantly fewer cells needed (~ 50,000 cells for ATAC-seq compared to millions of cells for the other methods (DNase-seq or FAIRE-seq))
- Two step process, one day of work
- Applications: **accessibility**, nucleosome positioning at transcription start sites, transcription factor footprinting

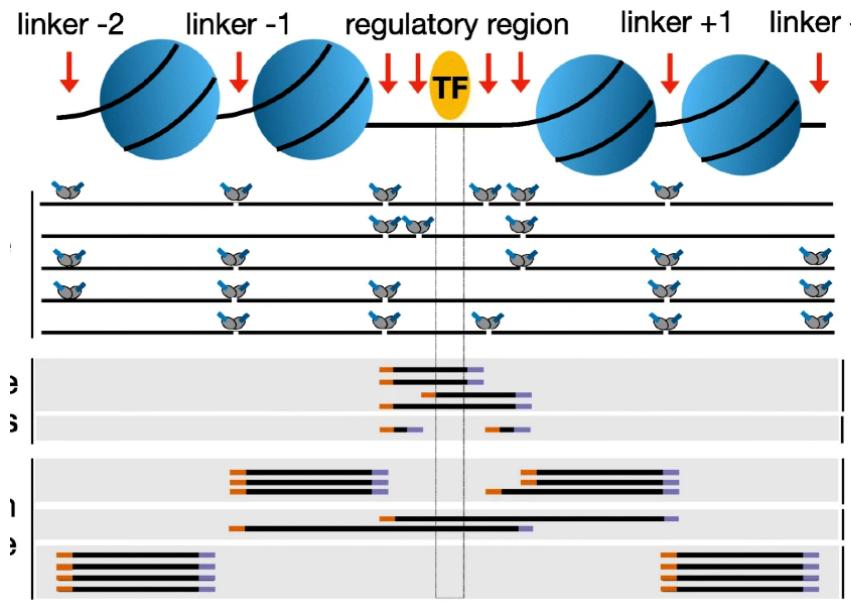




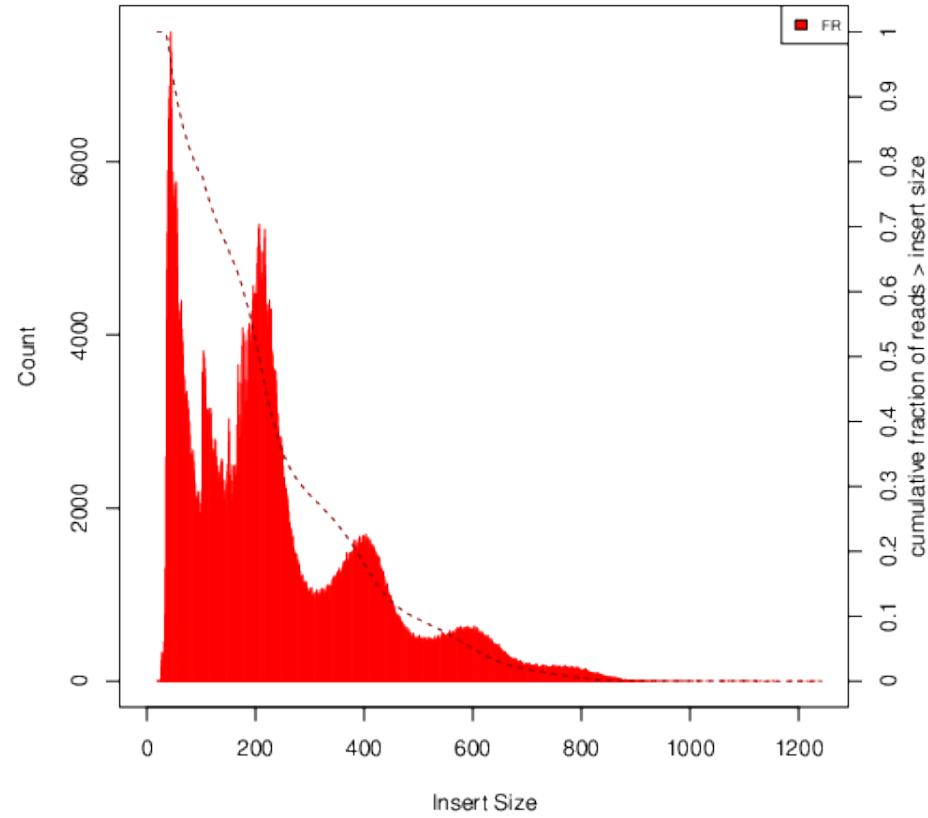
DNA fragments generated in ATAC-seq



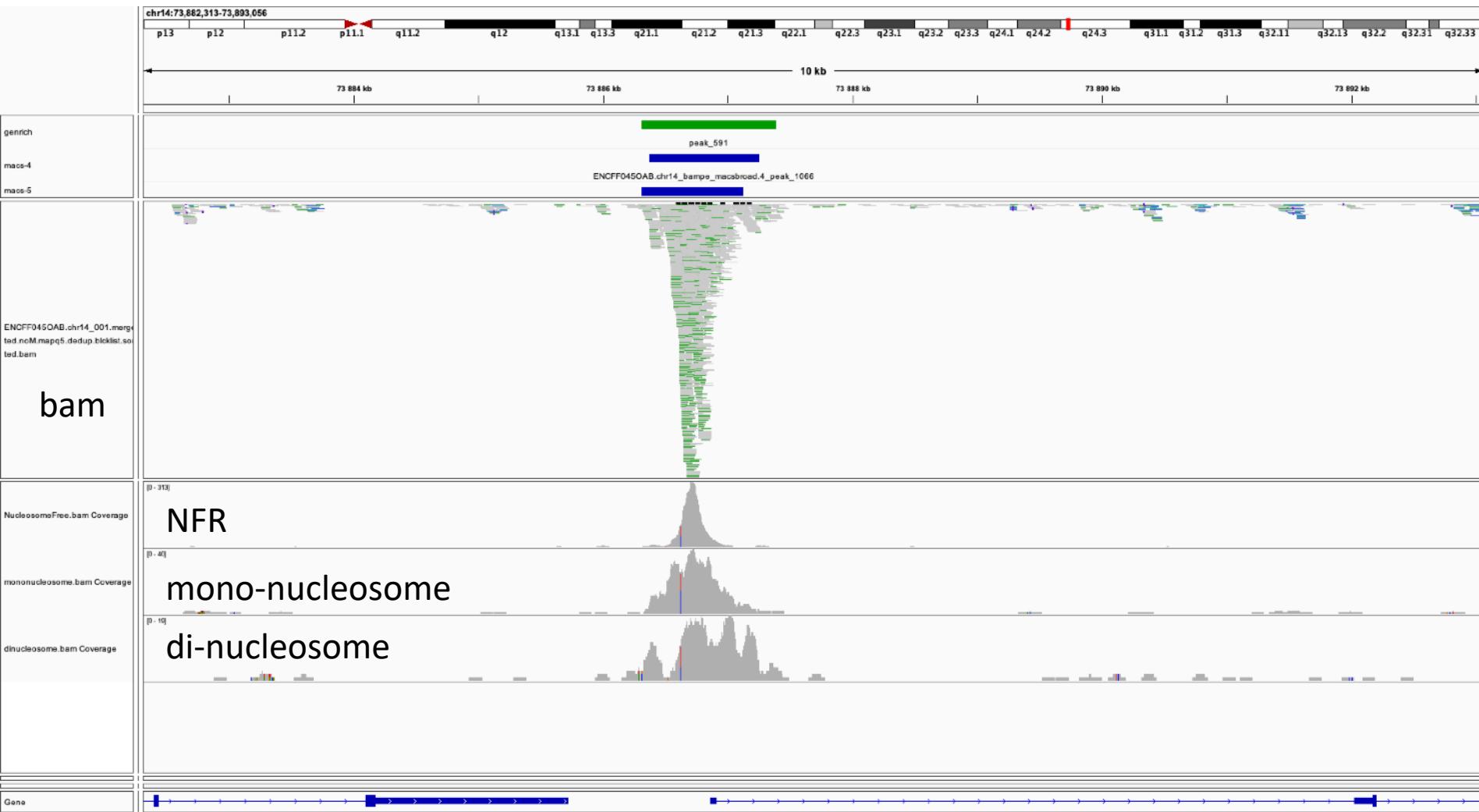
DNA fragments generated in ATAC-seq: QC



Insert Size Histogram for All_ Reads
in file ENCFF045OAB.chr14.blacklist_M_filt.mapq5.dedup.bam



Distribution of nucleosome-free and mono nucleosome signal at TSS



Distribution of nucleosome-free and mono nucleosome signal at TSS

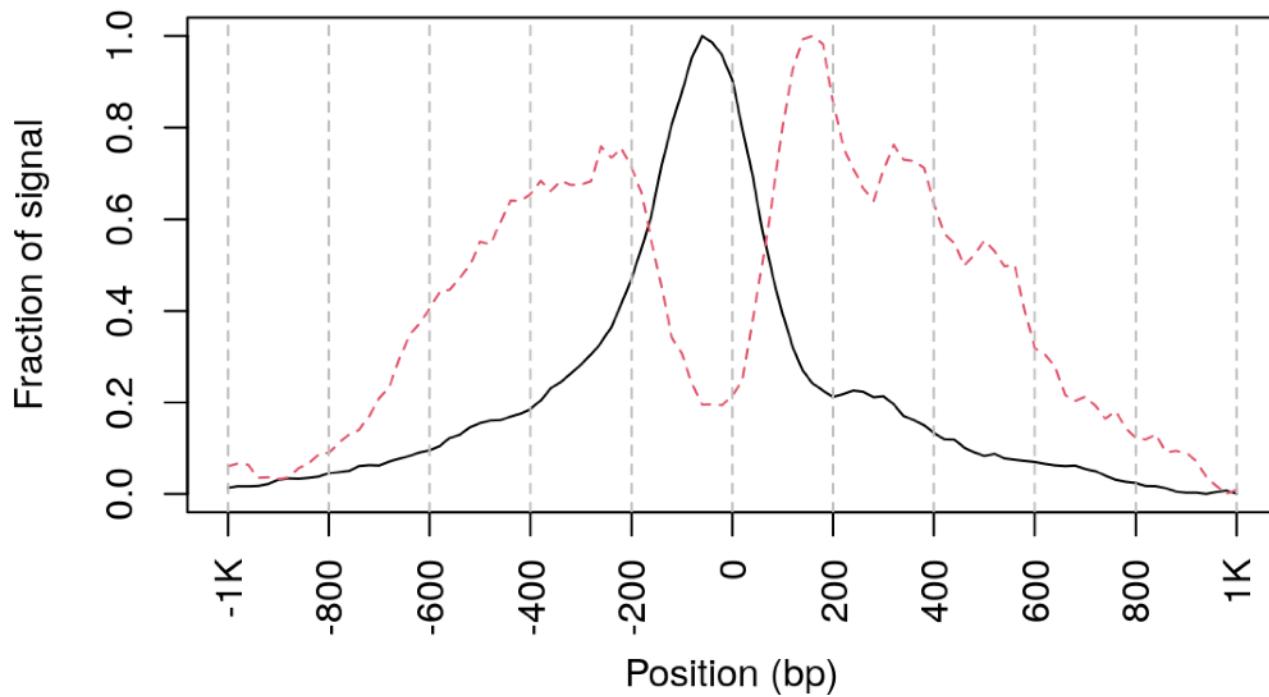
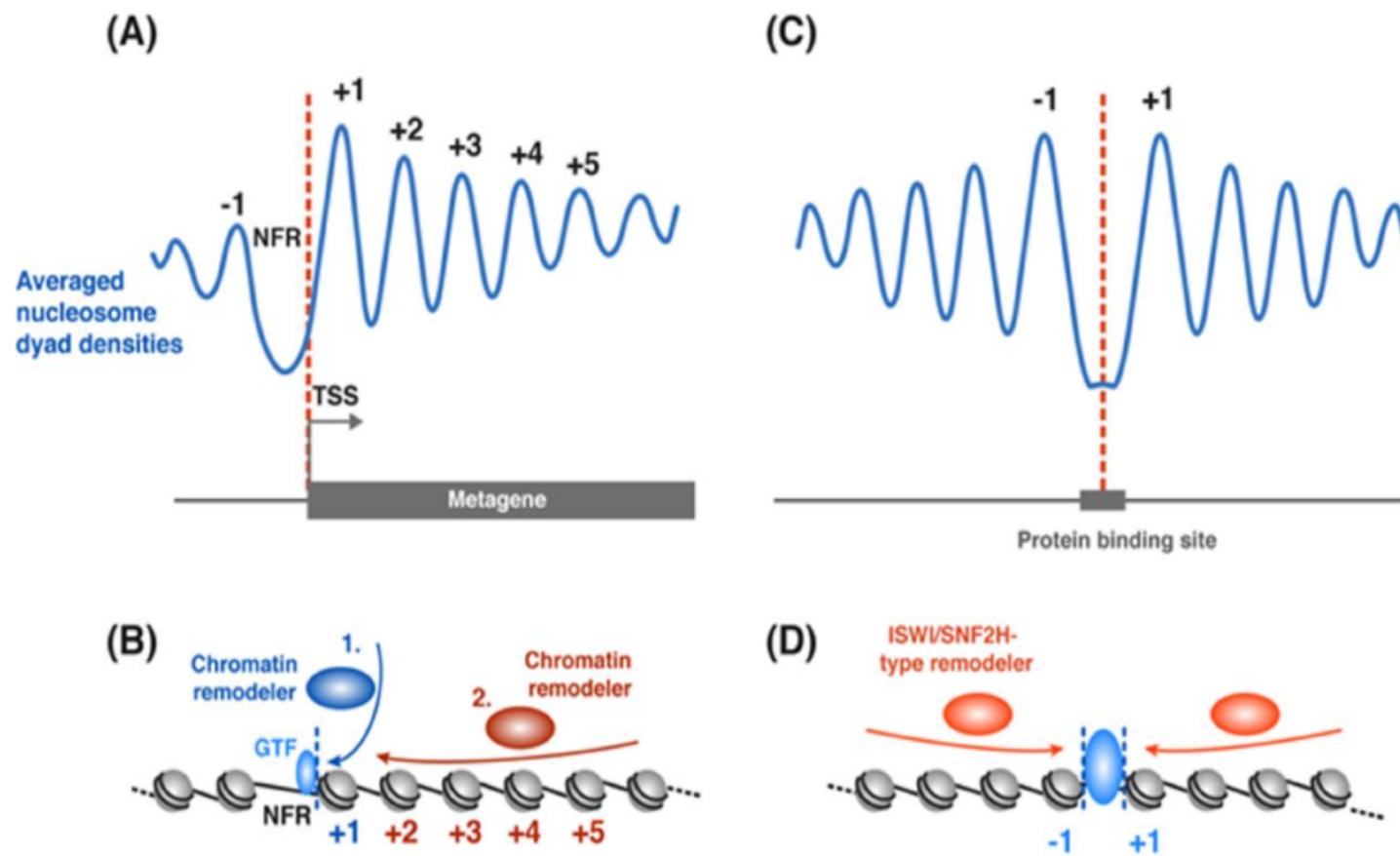


image: ATACseqQC

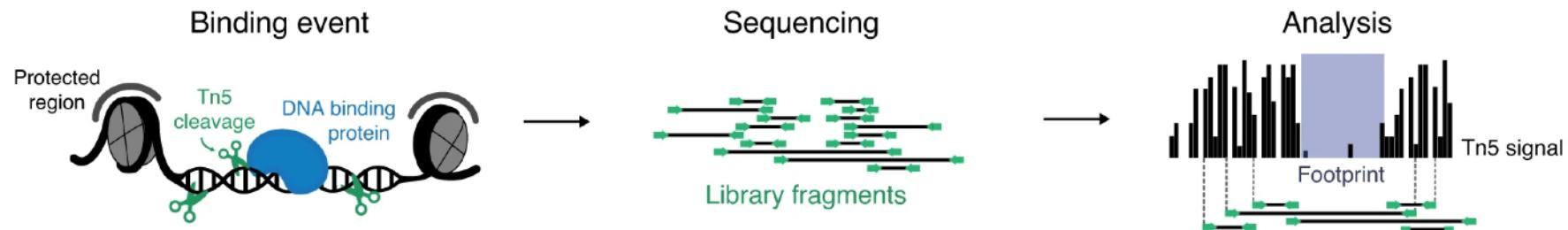
ATAC-seq: peaks



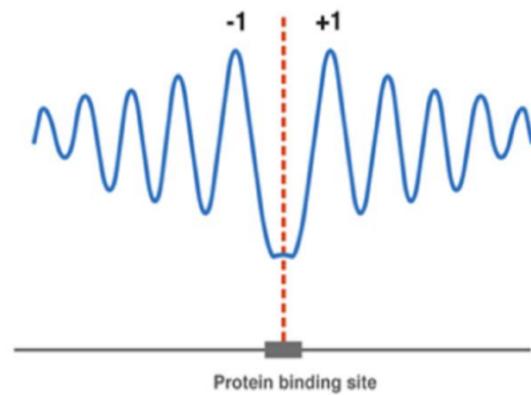
ATAC-seq: nucleosome resolution



Transcription Factor Footprinting: Principle



(C)



TF footprint

(D)

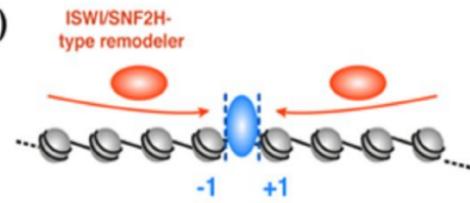
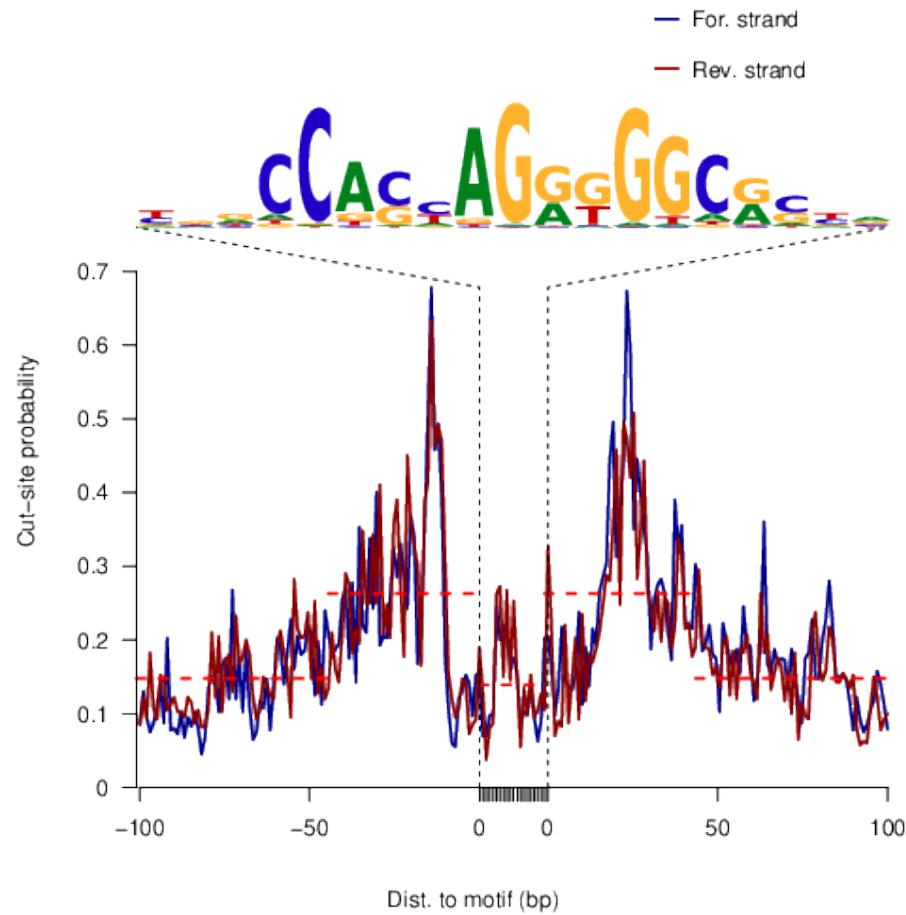


image:DOI:10.1038/s41467-020-18035-1

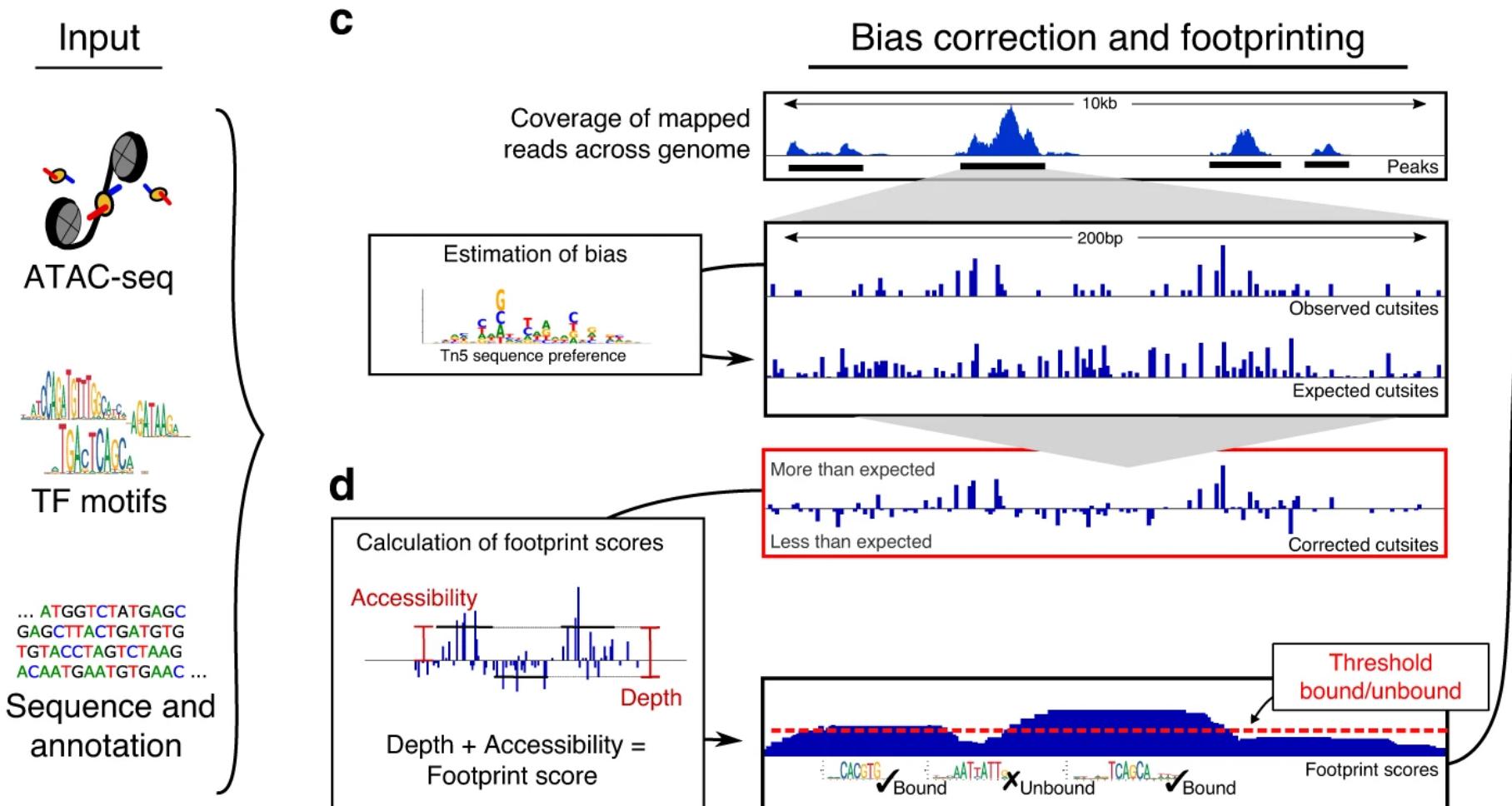
image: https://doi-org.ezp.sub.su.se/10.1042/EBC20180058

TF Footprinting - (a very simple) example (CTCF)



Transcription Factor Footprinting

The TOBIAS footprinting framework



Analysis workflow for ATAC-seq data

raw data *fastq



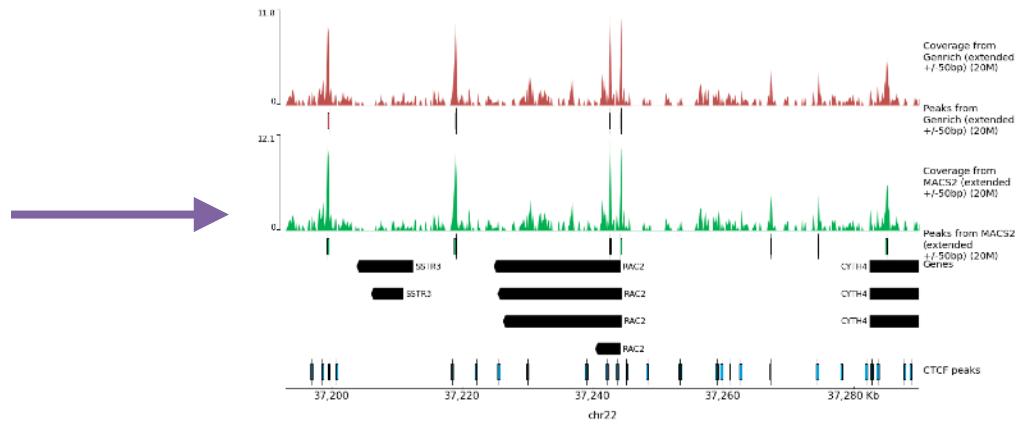
peaks & more

- fastq processing: quality and adapter trimming
- read mapping
- alignment filtering
- peak calling
- DA
- peak annotation
- visualisation
- TF footprinting

Analysis workflow

```
1 @ERR000589.41 EAS139_45:5:1:2:111/1
2 CTTTCCTCCCTGCTTCTGGCCCCACCATTCCAGGGAACATCTTGTCA
3 +
4 3IIIIIIIIIIII>1IIIFF9BG08E00I%IG+&?(4)%00646.C1#&(
5 @ERR000589.42 EAS139_45:5:1:2:1293/1
6 AGTTGTTAAAATCCAAGCCAATTAAGATAAGTCTTATCTTTAAAGAAAT
7 +
8 IIIIIGII.AIII=I9G-/II=+I=4?761BA2C9I+5A711+&>1$/I
```

raw data *fastq



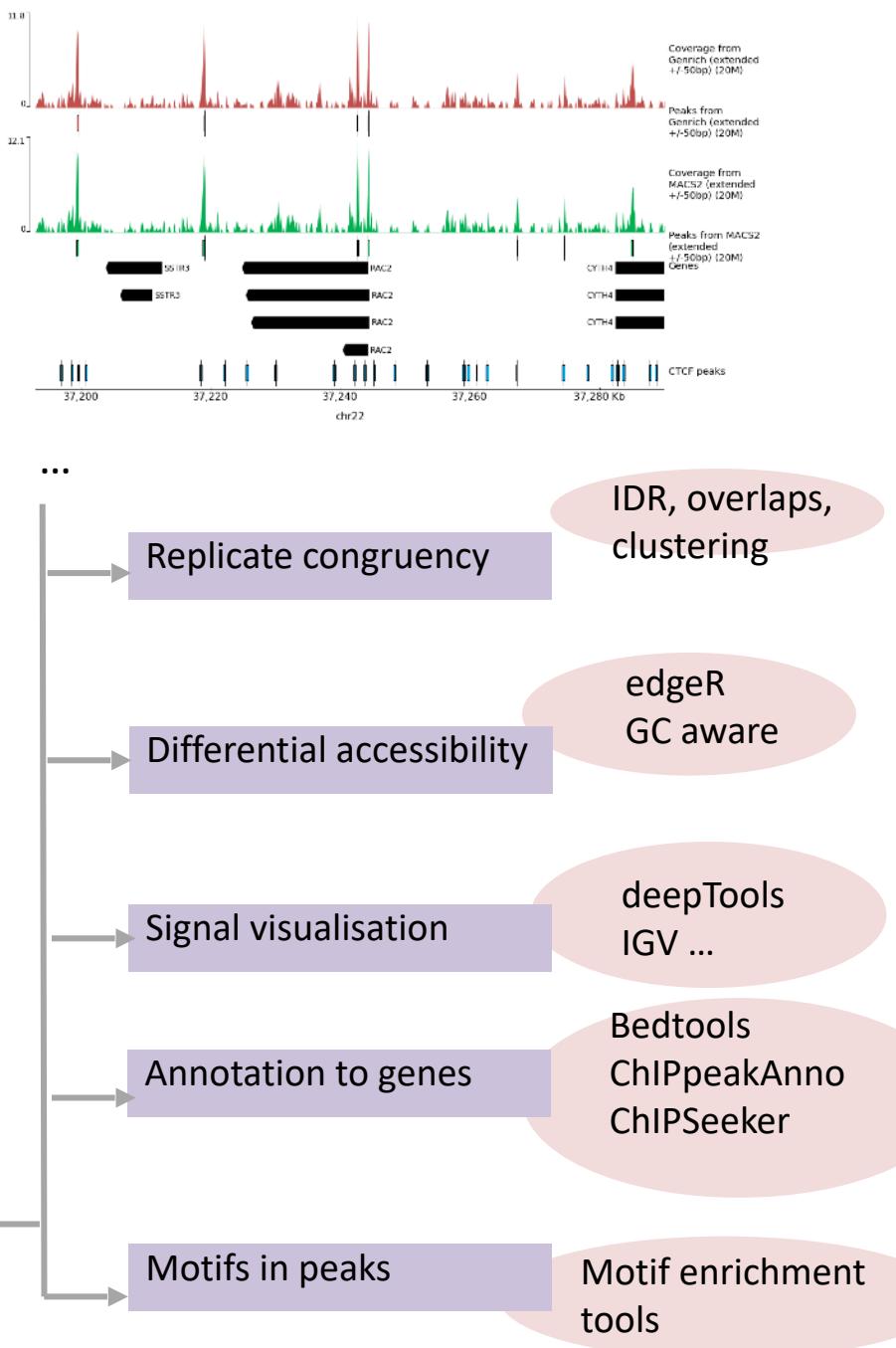
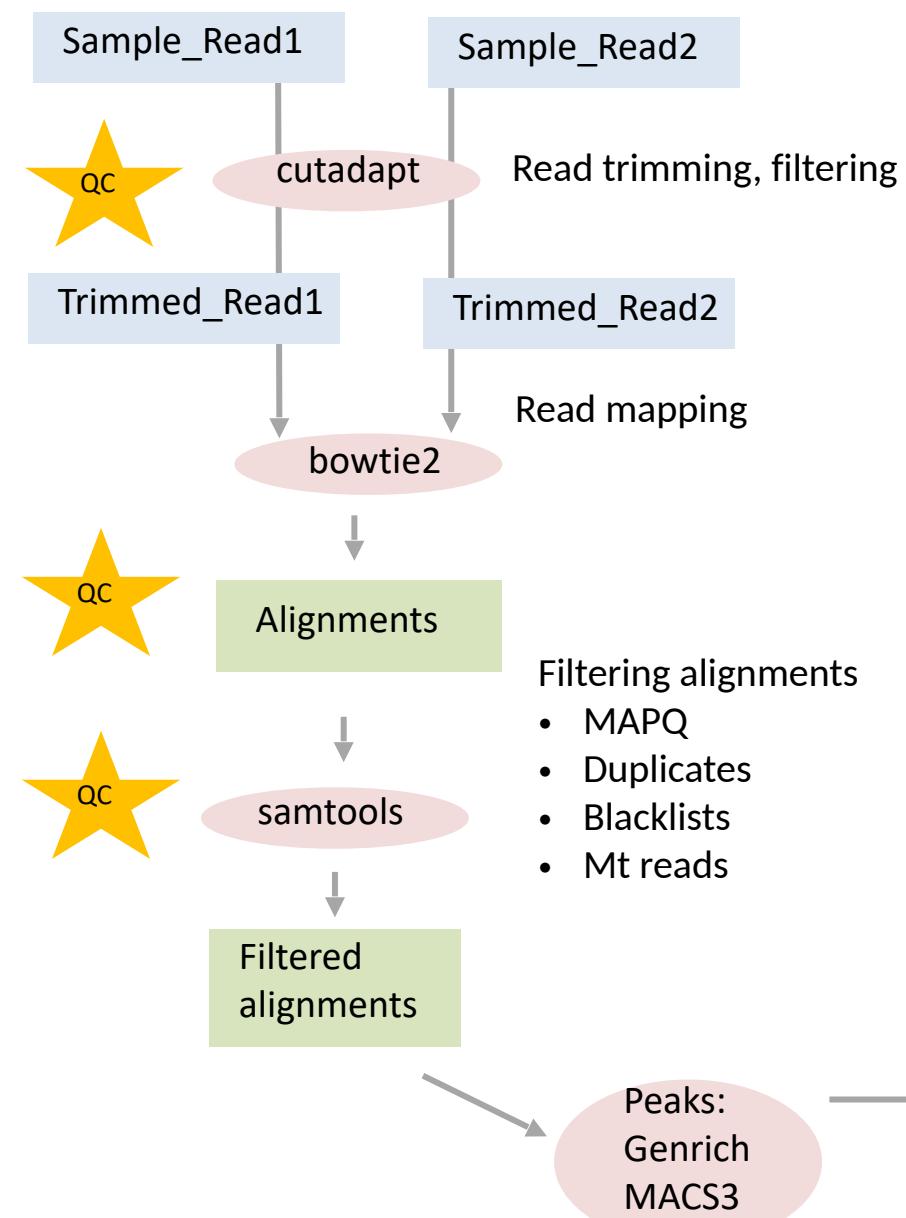
peaks & more

reads aligned to reference *bam



peaks

Analysis workflow



Special considerations for ATAC-seq data analysis

- Paired end (PE) sequencing is required
- QC: fragment length distribution – mononucleosome peak should be evident
- QC: fraction of Mt reads – it can be high (up to 40%) – calculate sequencing depth accordingly
- For current data quality standards, refer to ENCODE; currently 25 million non-duplicate, non-mitochondrial aligned read pairs (i.e. 50M PE reads); non-redundant fraction >0.9; fraction of reads in called peak regions (FRiP) >0.3; mononucleosome peak present; TSS enrichment observed
- Peak calling
 - Genrich – peak caller dedicated to ATAC-seq data (has an ATAC-seq mode); PE data only
 - hmmratac (MACS3) – learns the chromatin structure (from fragment length) and calls peaks based on the presence of the ATAC signature (a peak in NFR fraction flanked by peaks in mono-nucleosome fraction);
 - callpeak (MACS3) – can be used, in PE mode

Special considerations for ATAC-seq differential accessibility analysis: effect of normalisation

Methodology | Open Access | Published: 22 April 2020

ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation

Jake J. Reske, Mike R. Wilson & Ronald L. Chandler 

Epigenetics & Chromatin 13, Article number: 22 (2020) | [Cite this article](#)

doi:<https://doi.org/10.1186/s13072-020-00342-y>

Normalization benchmark of ATAC-seq datasets shows the importance of accounting for GC-content effects

 Koen Van den Berge, Hsin-Jung Chou,  Hector Roux de Bézieux,  Kelly Street,  Davide Risso,  John Ngai,  Sandrine Dudoit

doi: <https://doi.org/10.1101/2021.01.26.428252>

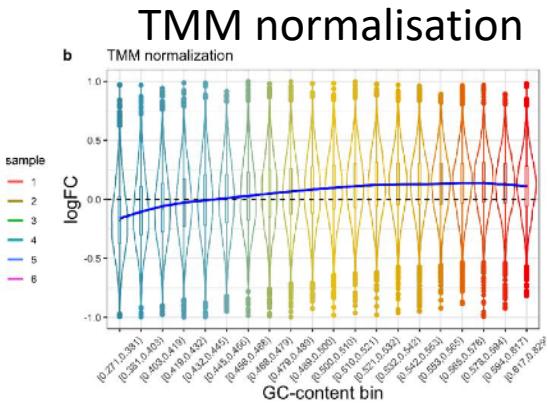
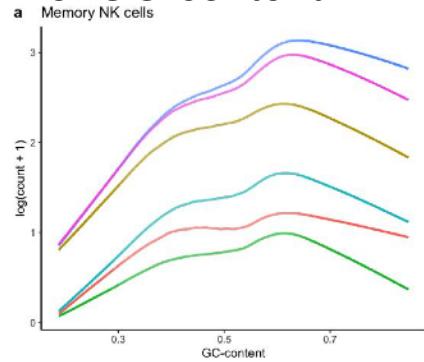
This article is a preprint and has not been certified by peer review [what does this mean?].

- GC-content effects are omnipresent in ATAC-seq datasets;
- Since the GC-content effects are sample-specific, they can bias downstream analyses such as clustering and differential accessibility analysis;
- We introduce a GC aware normalization method;
- Our work clearly shows that accounting for GC-content effects in the normalization is crucial for common downstream ATAC-seq data analyses.

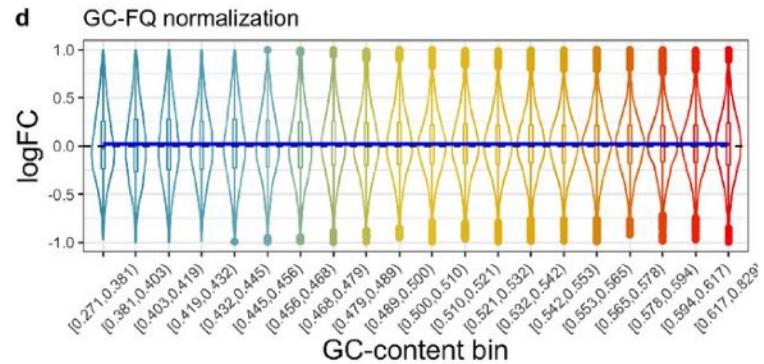
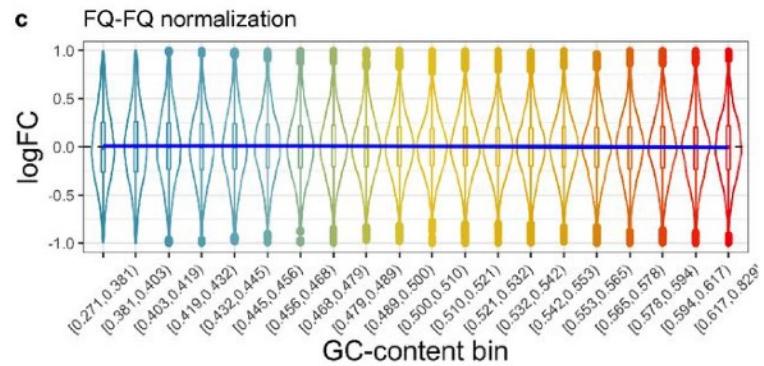
doi: <https://doi.org/10.1101/2021.01.26.428252>

Special considerations for ATAC-seq differential accessibility analysis: effect of normalisation

log-count as a function
of GC-content



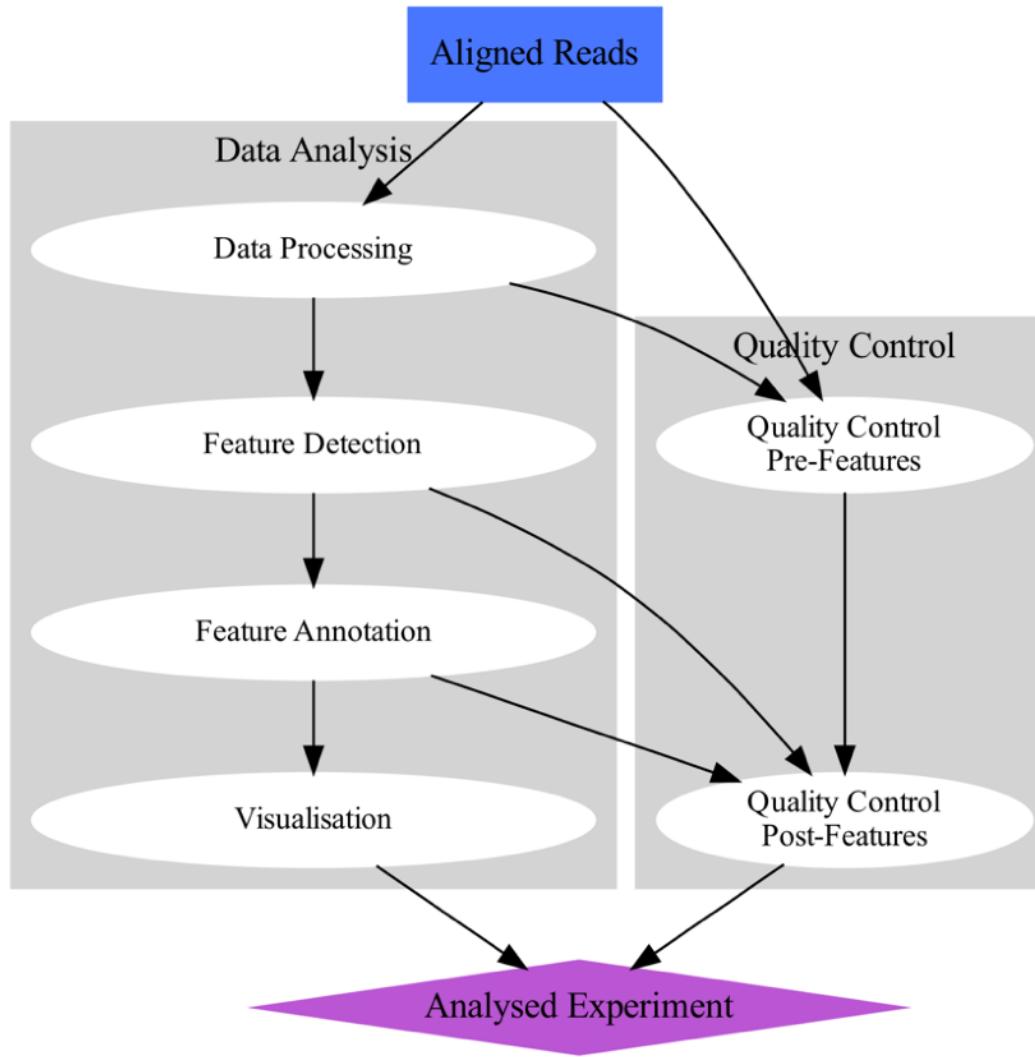
GC aware normalisation



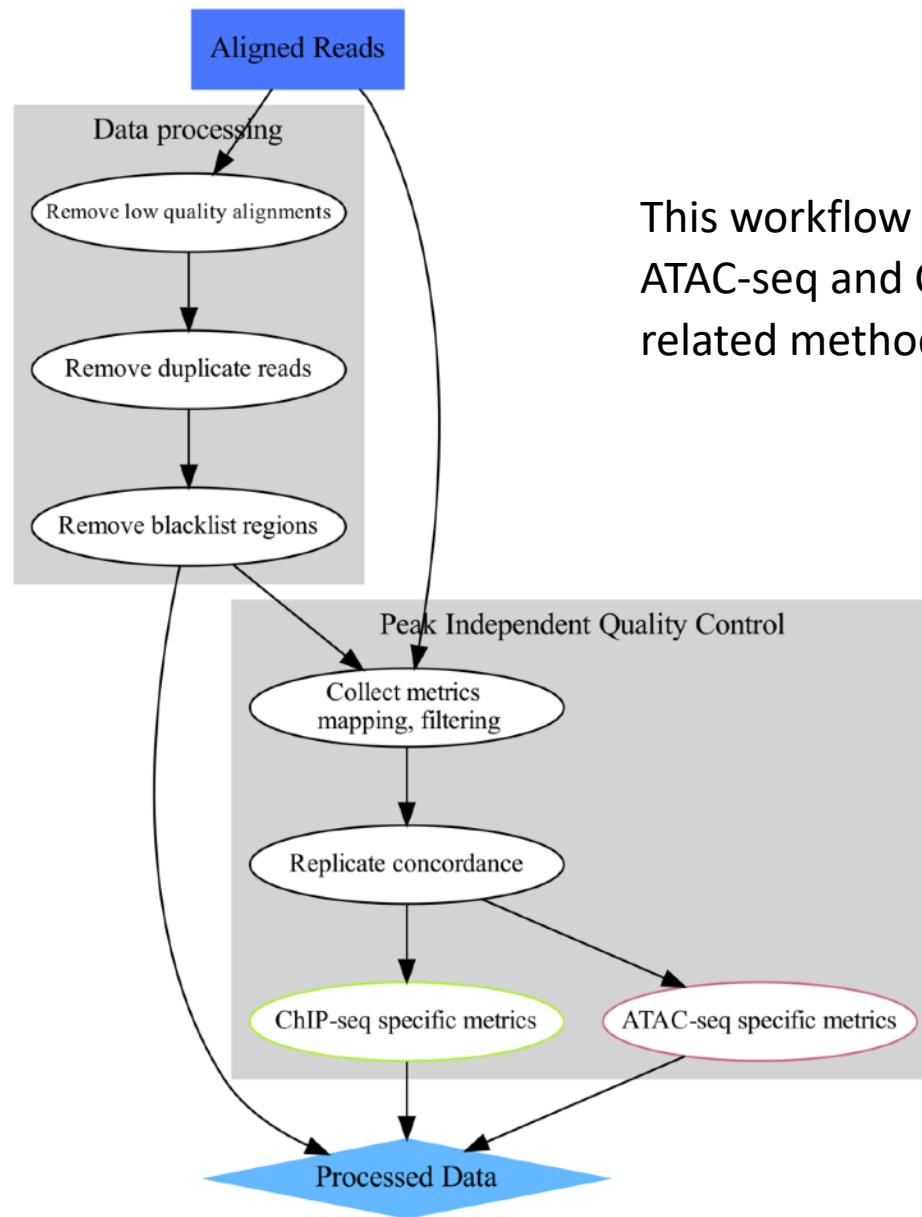
Differential accessibility
log-fold change in bins by GC content

A bias for peaks with low and high GC-content
(in a null setting, LFC should be centred on
zero)

Exercise Overview

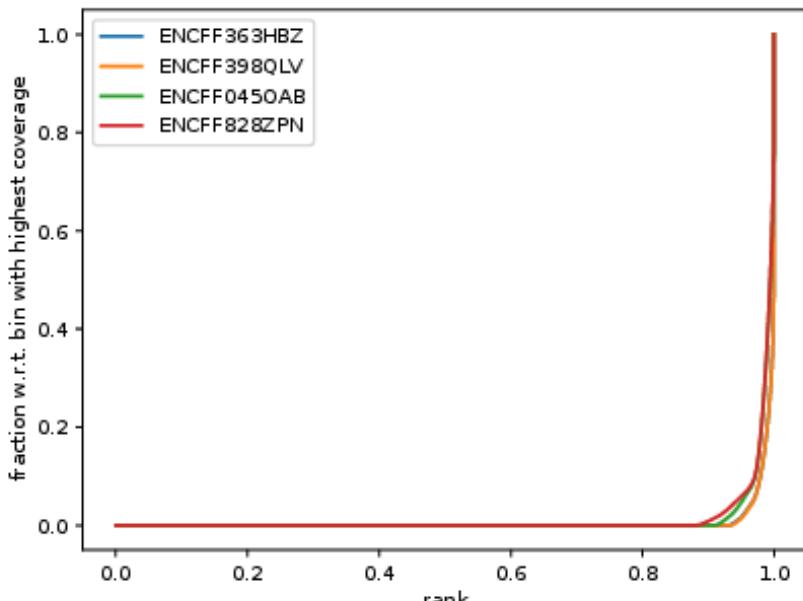


Exercise Overview: Data preprocessing and QC



This workflow is very similar for ATAC-seq and ChIP-seq and related methods

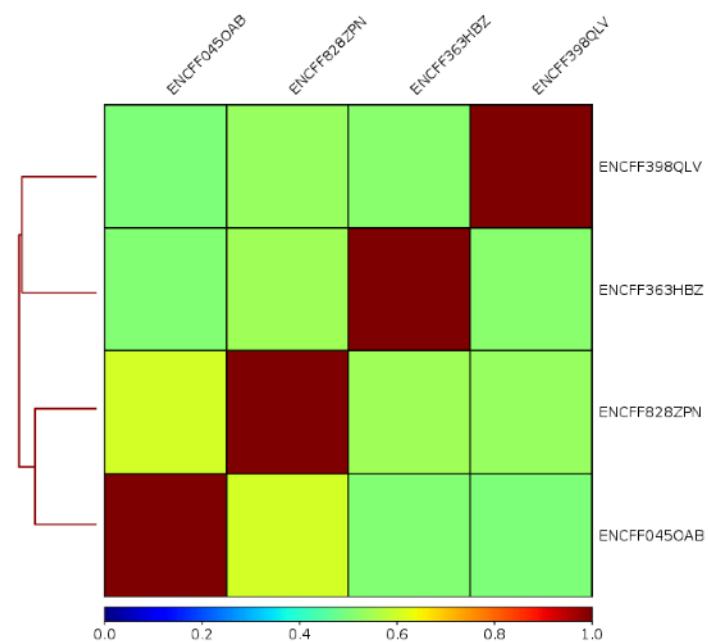
Exercise Overview: Data preprocessing and QC



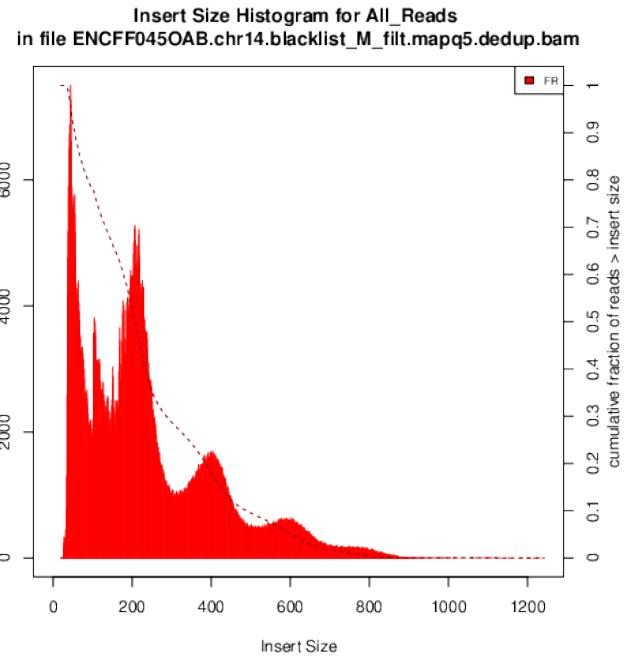
Cumulative enrichment ("fingerprint")

Replicate clustering

These QC steps are common for ATAC-seq and ChIP-seq and related methods

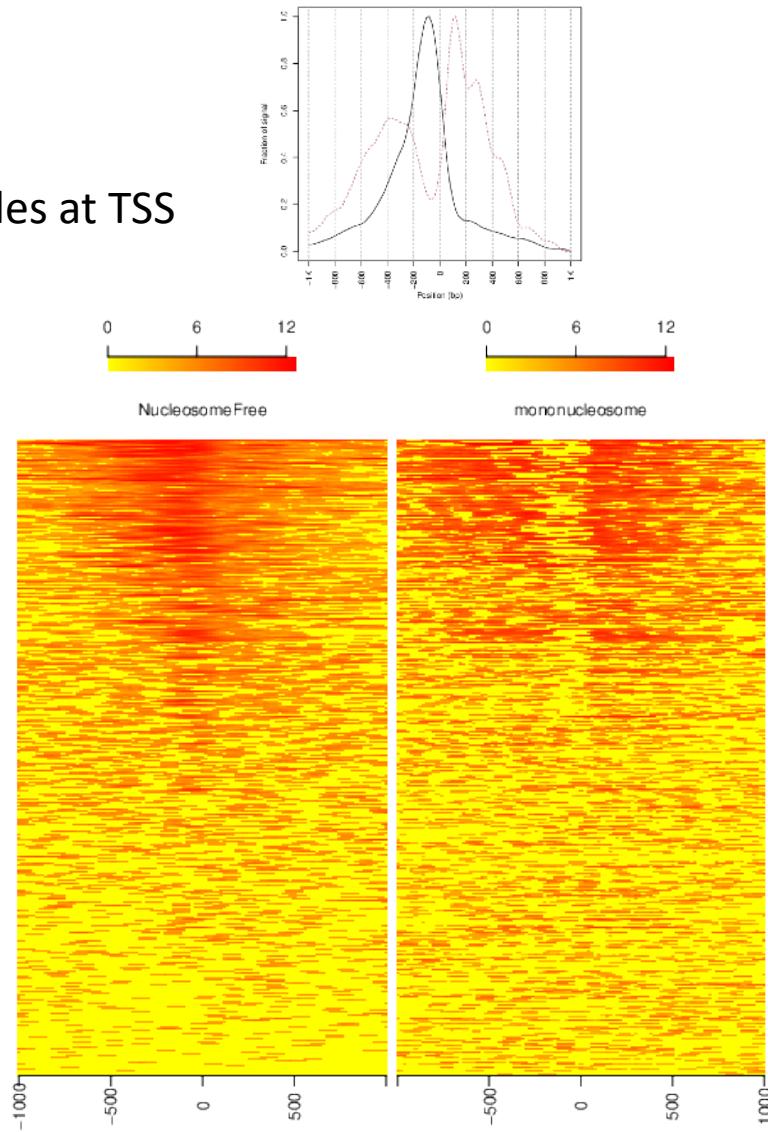


Exercise Overview: ATAC-seq specific QC

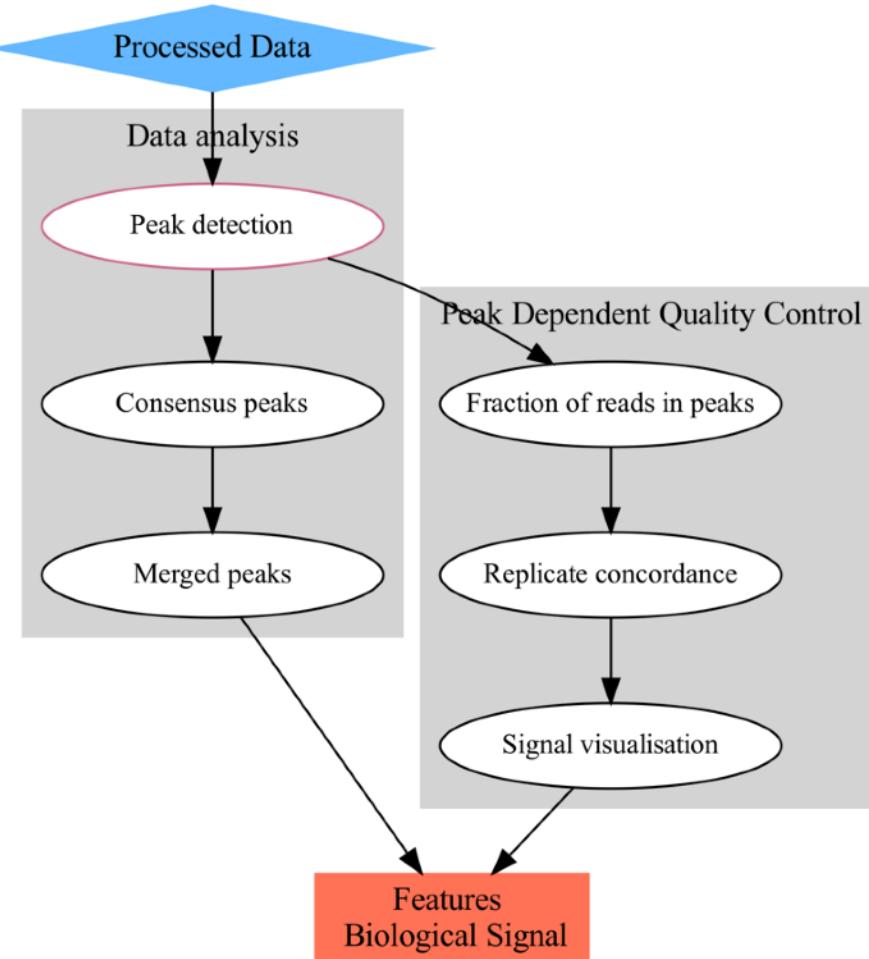


Fragment length distribution

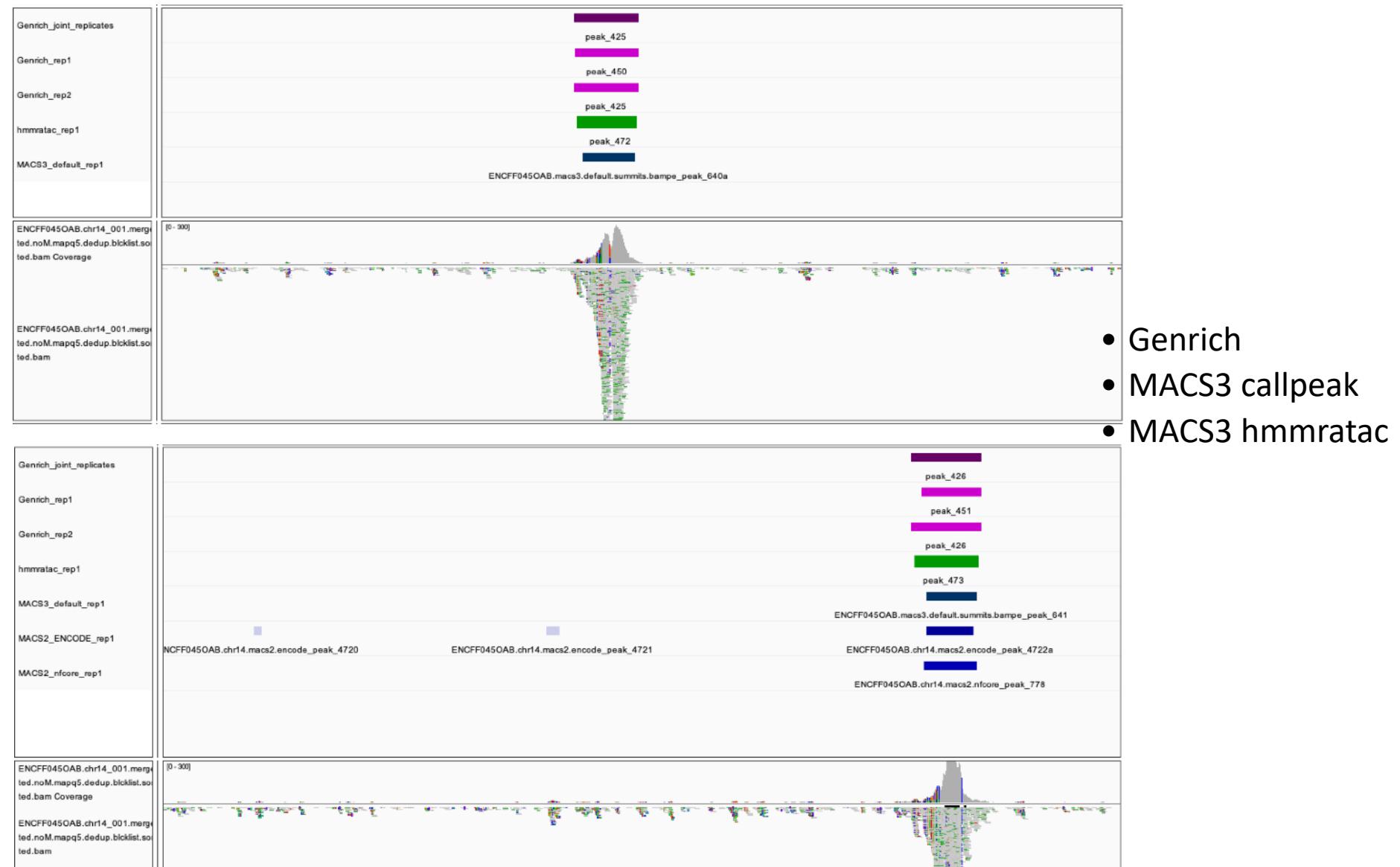
Signal profiles at TSS



Exercise Overview: ATAC-seq peak calling



Exercise Overview: ATAC-seq peak calling



Thank you for listening

Please follow the tutorials:

1. Data preprocessing and QC
2. ATAC-seq specific QC
3. ATAC-seq peak calling

agata.smialowska@nbis.se