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

Epigenomics Data Analysis Workshop

Stockholm, 19 September 2023

Agata Smialowska

NBIS, SciLifeLab, Stockholm University

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

UNIVERSITET STOCKHOLMS

Chromatin structure

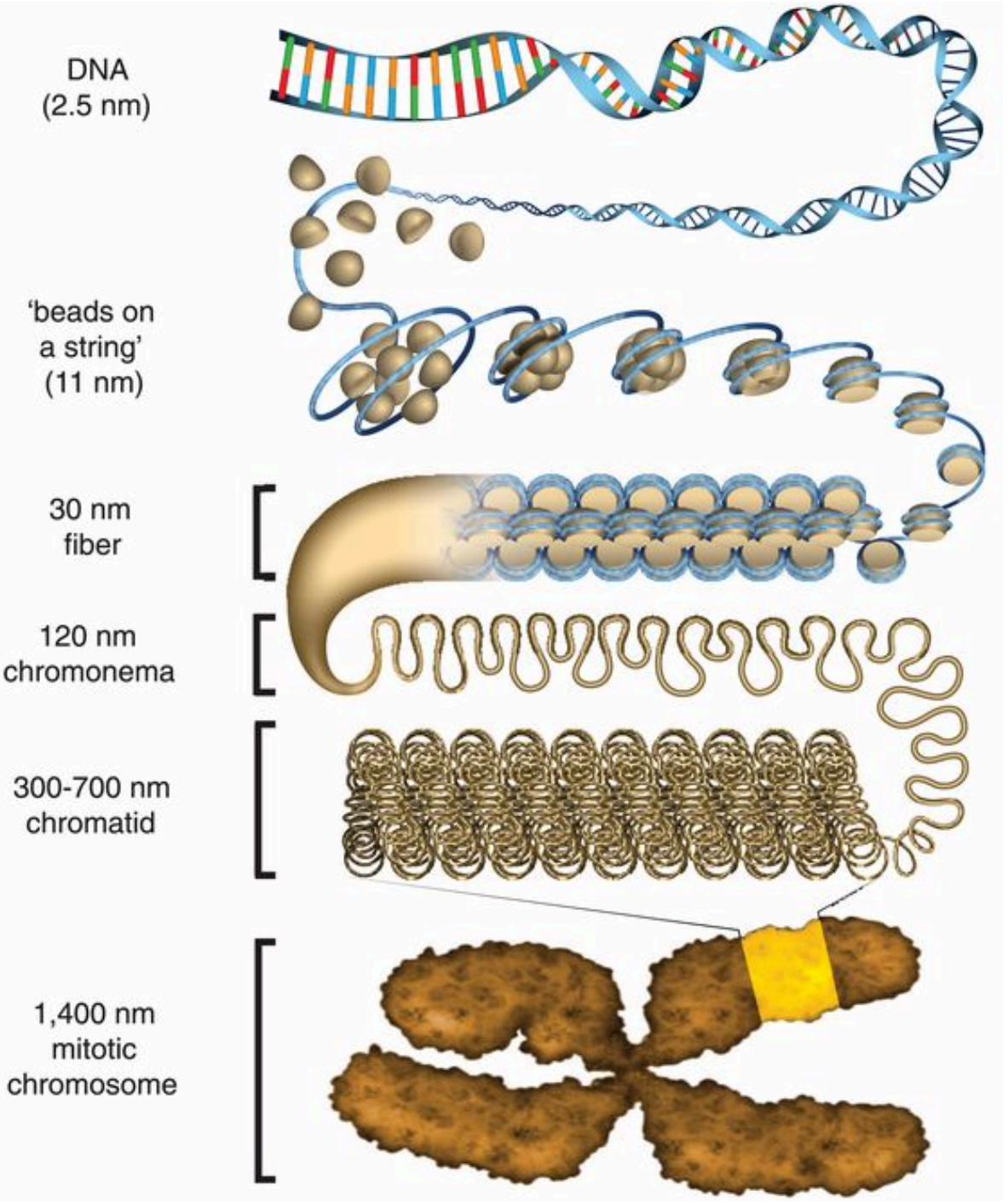


image: DOI: 10.1126/science.aag0025

Chromatin structure

ChIP-seq

ATAC-seq, DNaseI-seq,
MNase-seq

3D organisation
Hi-C, Dam-ID

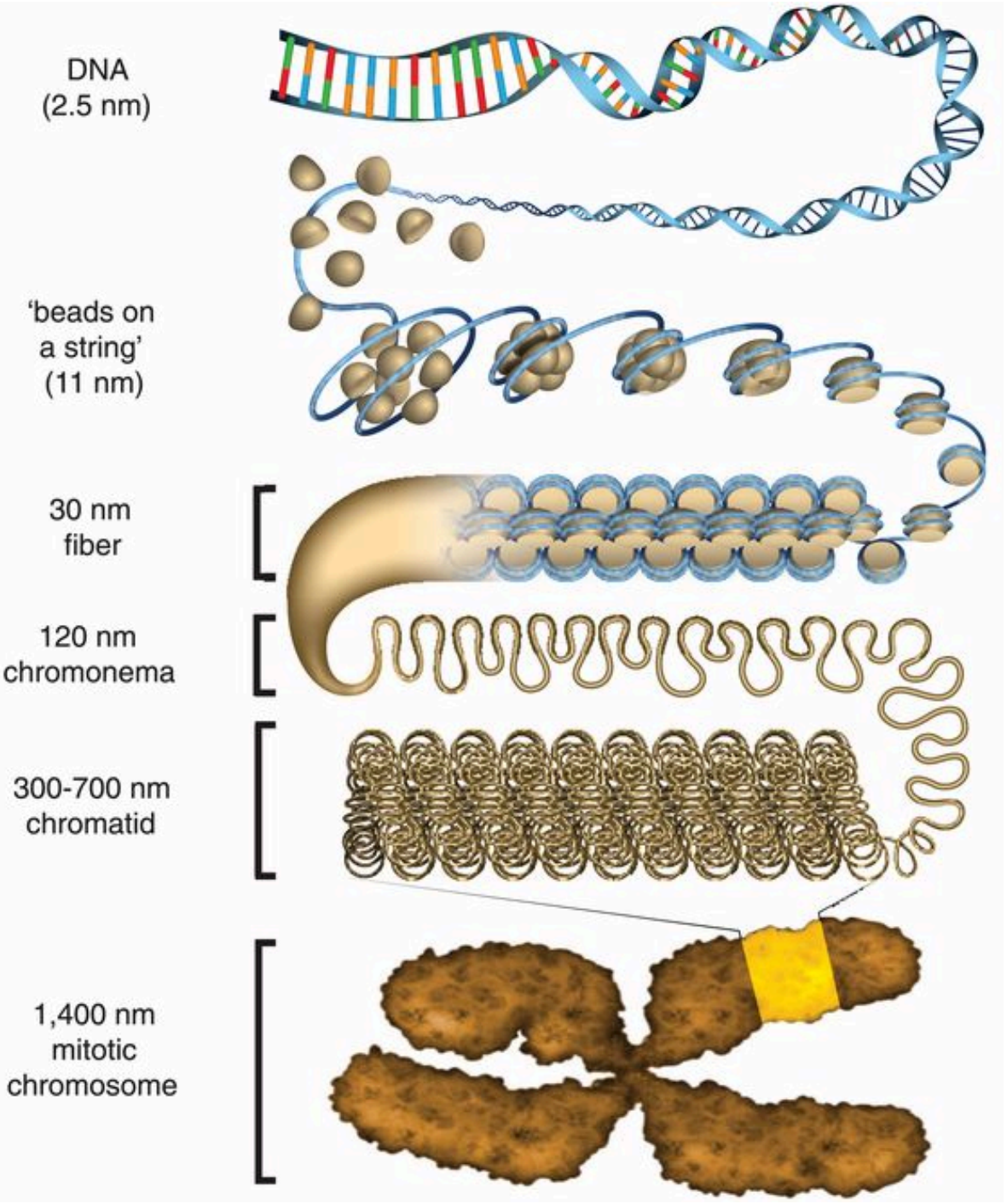


image: DOI: 10.1126/science.aag0025

Chromatin states and gene expression

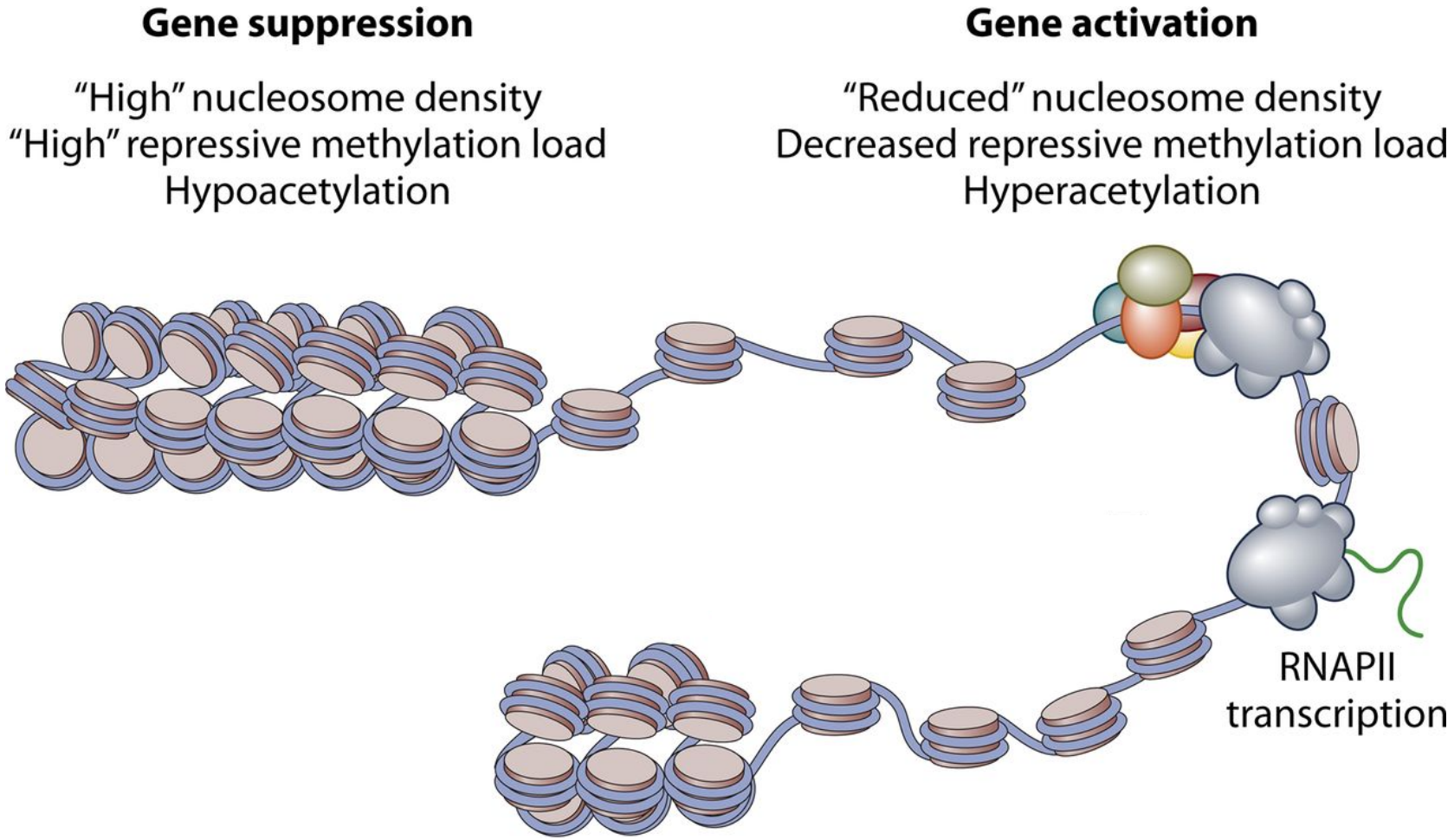
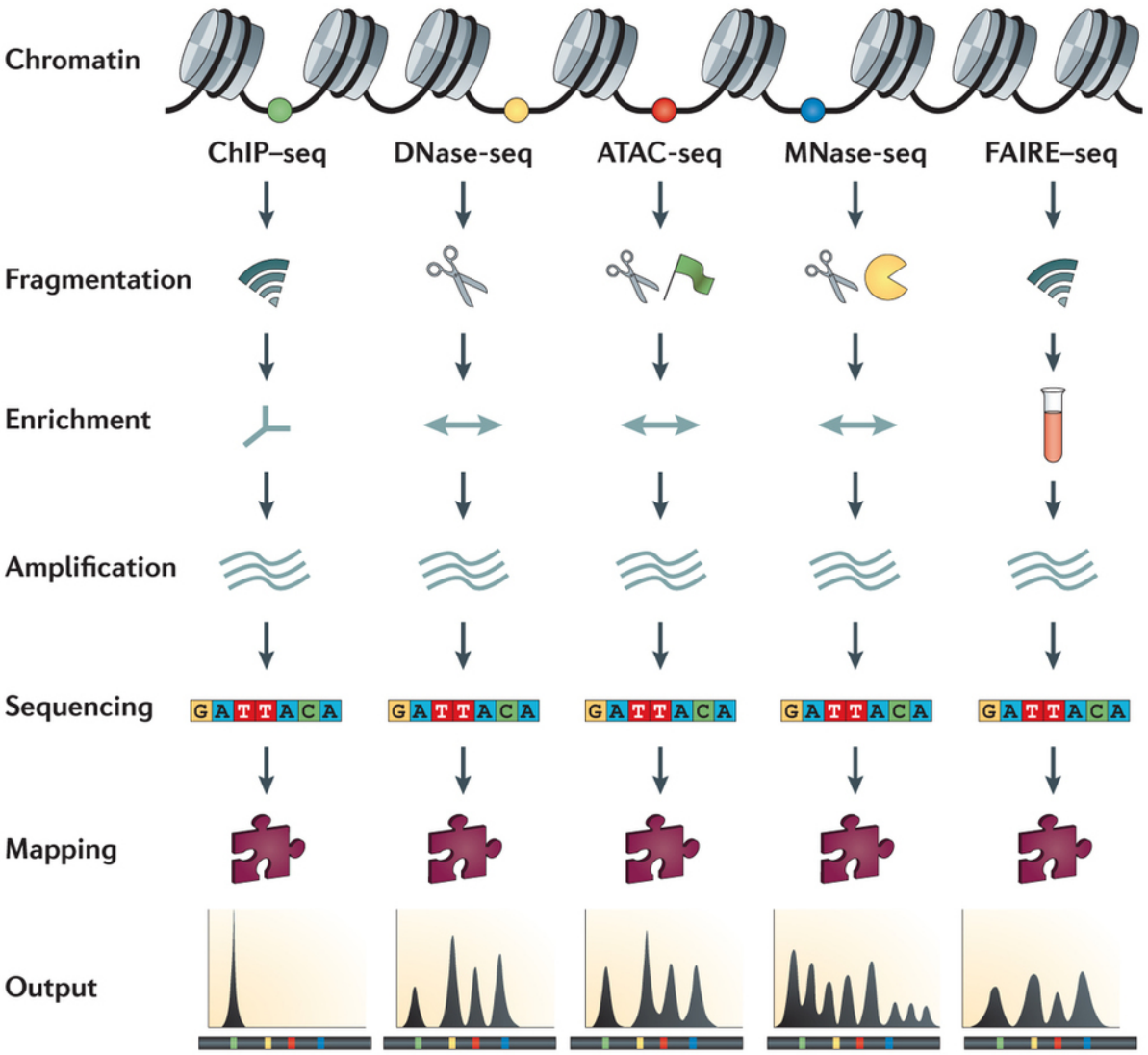


image: https://doi.org/10.1128/mBio.00098-16

Functional genomics techniques to probe chromatin states



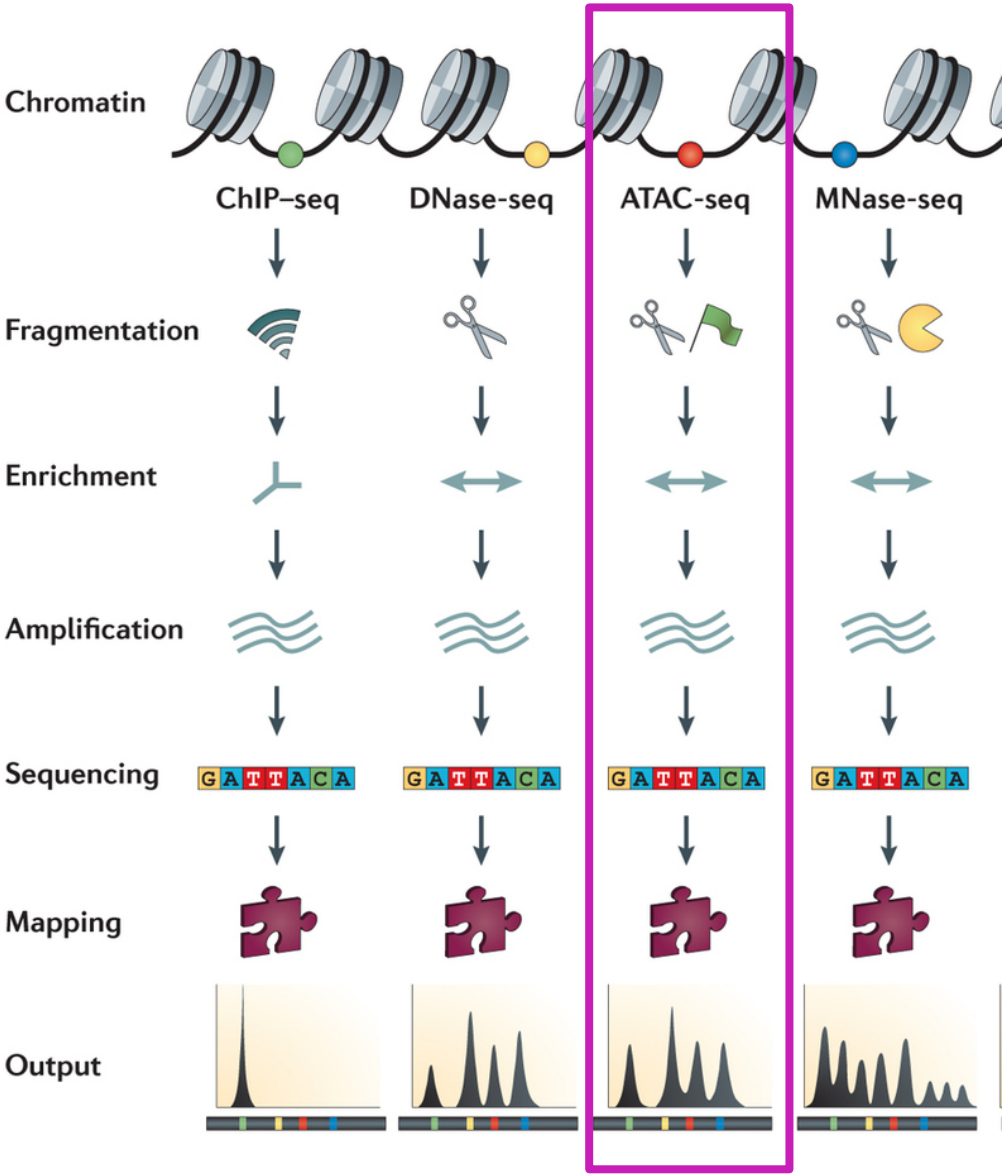
Accessibility – targeting nucleosome-depleted DNA:

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

Nucleosome positioning:
MNase-seq

Clifford et al, Nature Rev Genet, 2014

Functional genomics techniques to probe chromatin states



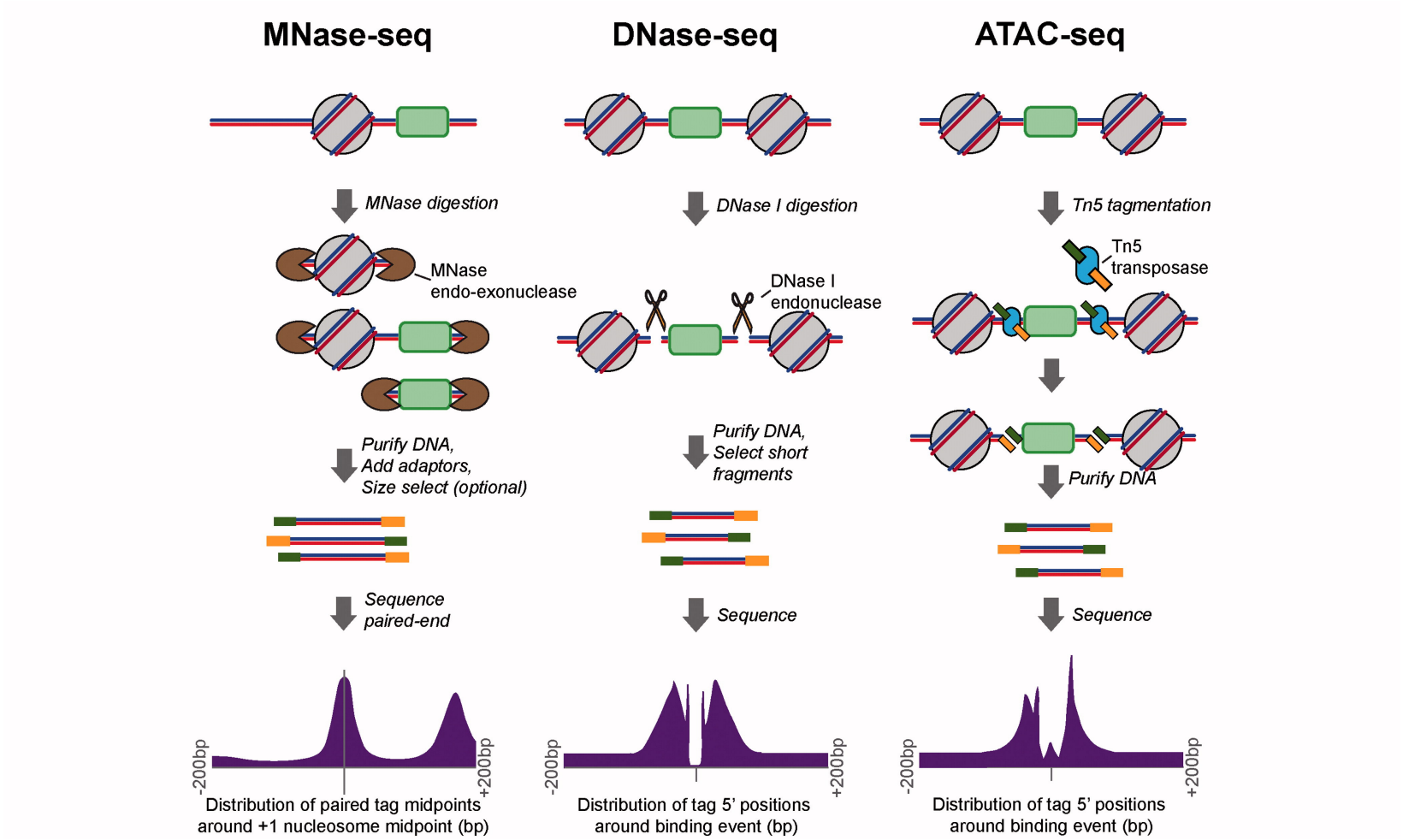
Accessibility – targeting nucleosome-depleted DNA:

ATAC-seq

Nucleosome positioning: MNase-seq

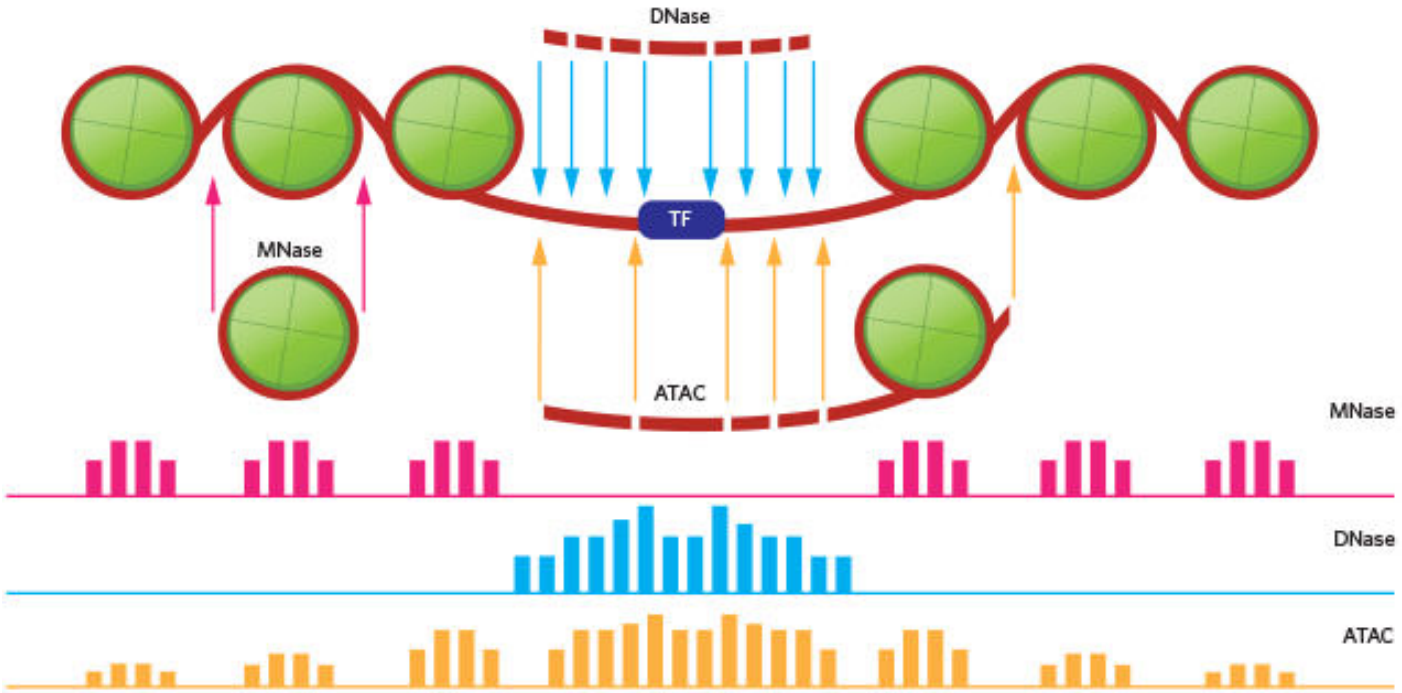
Clifford et al, Nature Rev Genet, 2014

Functional genomics techniques to identify open chromatin regions



Pugh 2015

Functional genomics techniques to probe chromatin states



http://www.the-scientist.com/?articles.view/articleNo/44772/title/Reveling-in-the-Revealed

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

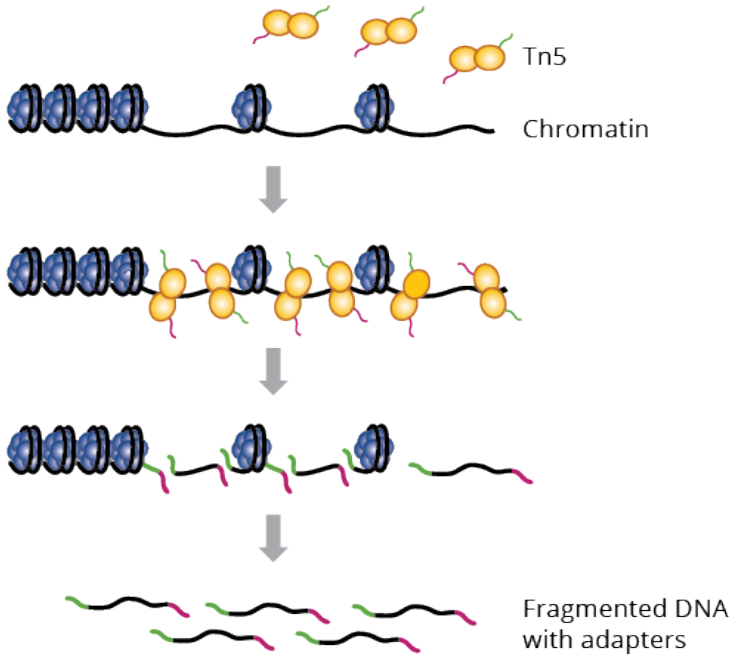
Single-cell chromatin accessibility reveals principles of regulatory variation

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



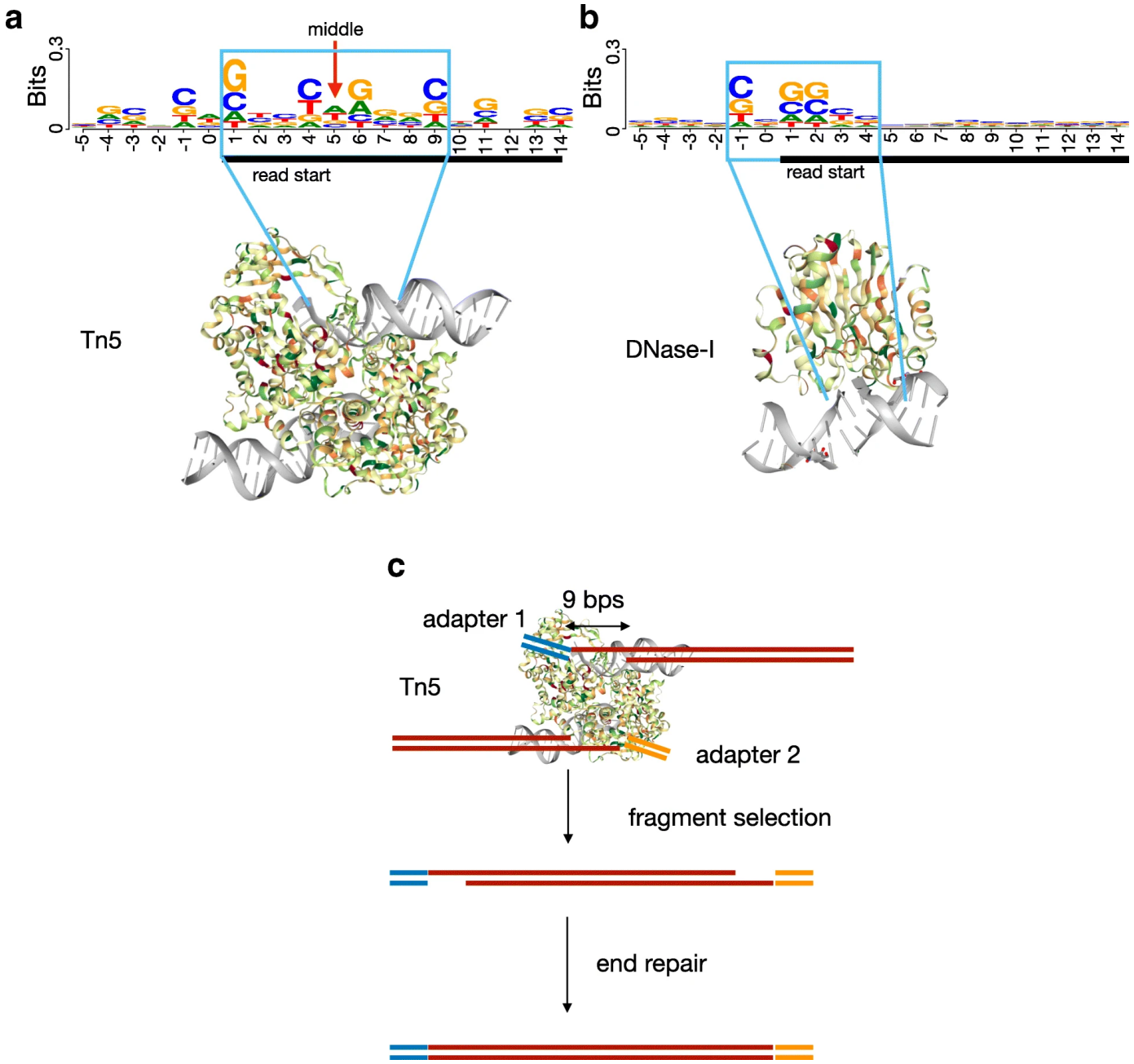


image: https://doi.org/10.1186/s13059-019-1642-2

DNA fragments generated in ATAC-seq

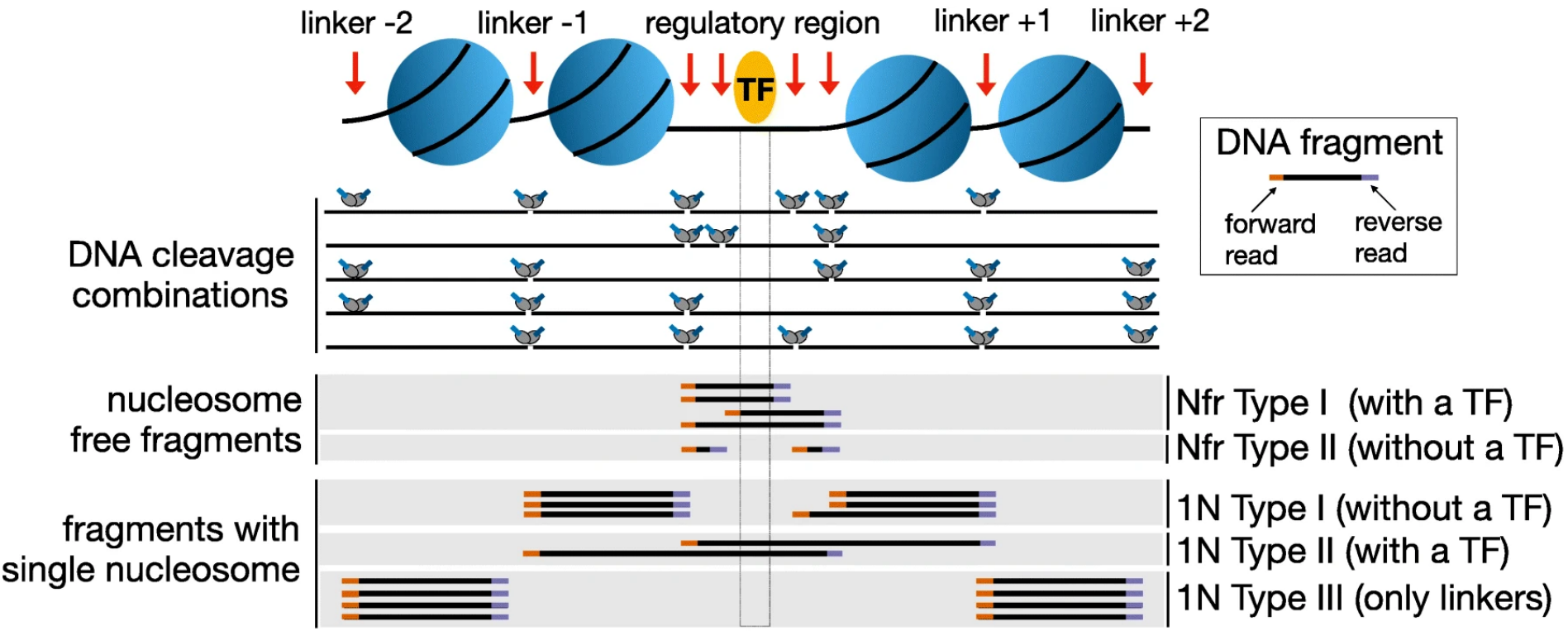
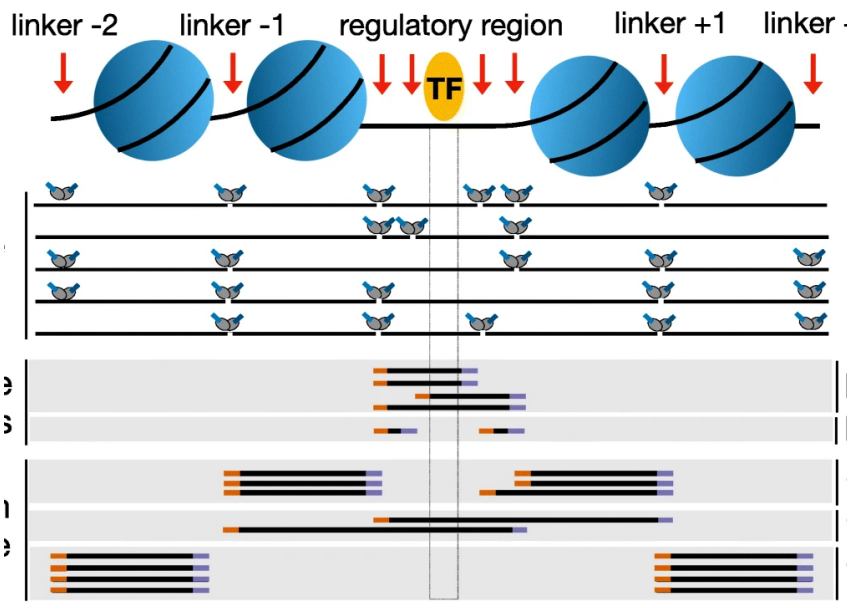
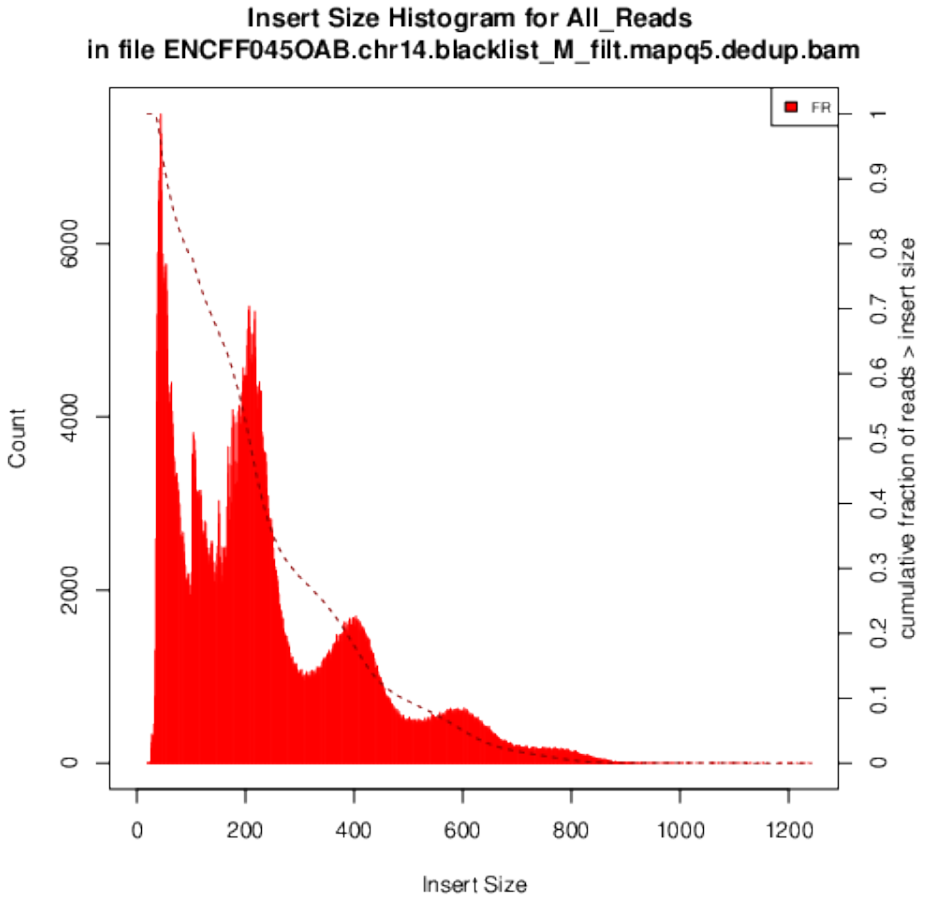


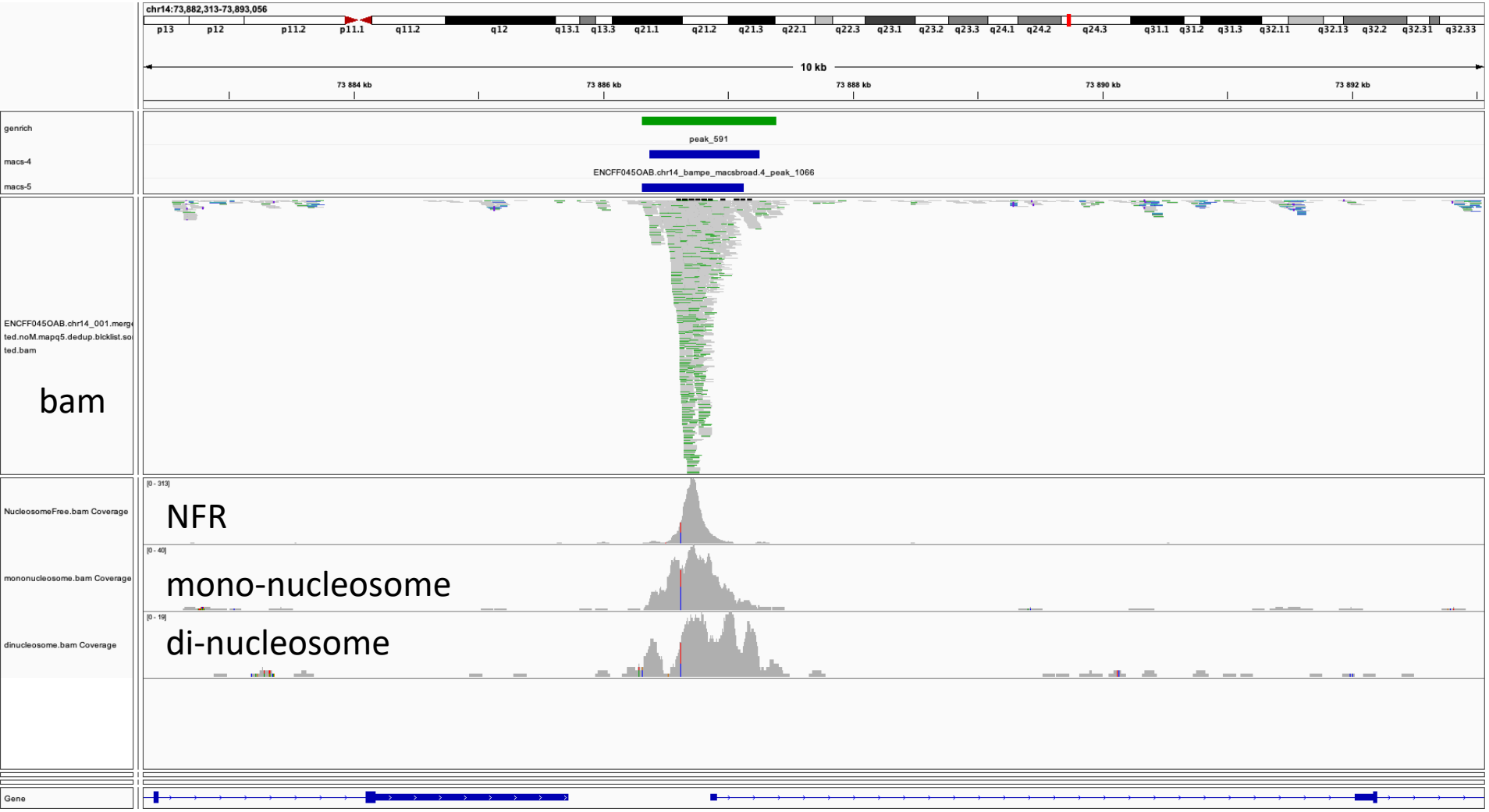
image: https://doi.org/10.1186/s13059-019-1642-2

DNA fragments generated in ATAC-seq: QC





Distribution of nucleosome-free and mono nucleosome signal at TSS



Distribution of nucleosome-free and mono nucleosome signal at TSS

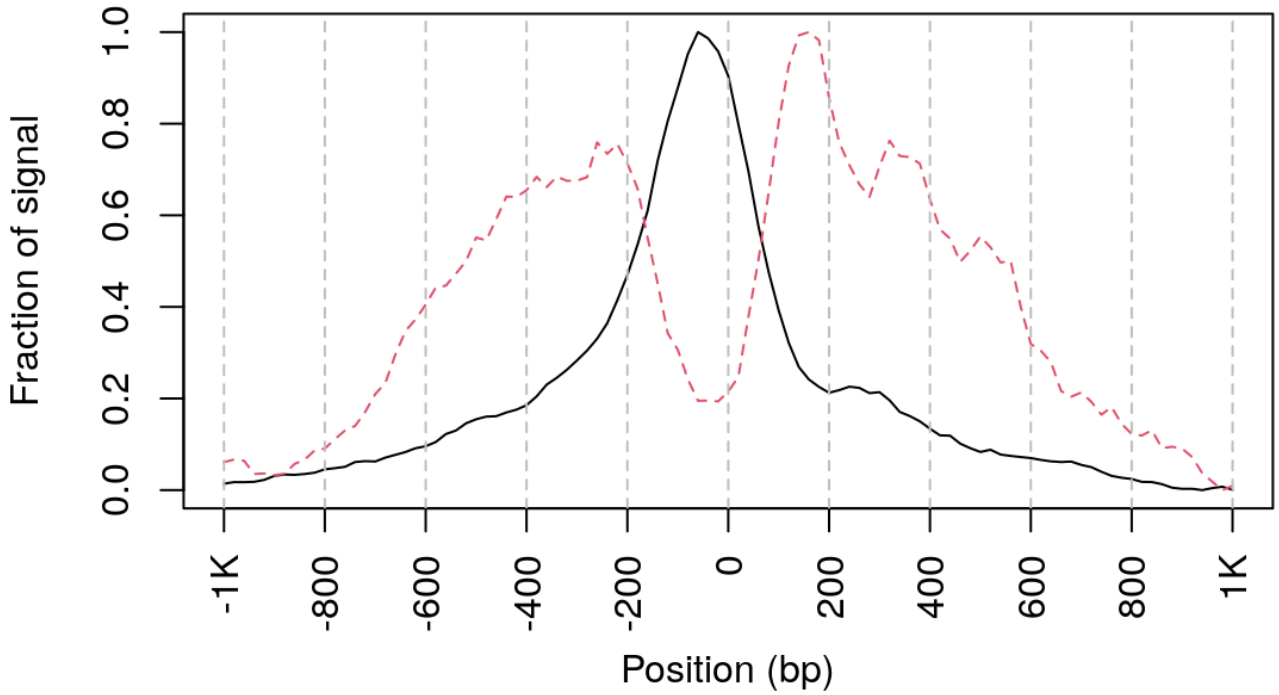


image: ATACseqQC

ATAC-seq: peaks

ATAC-seq: nucleosome resolution

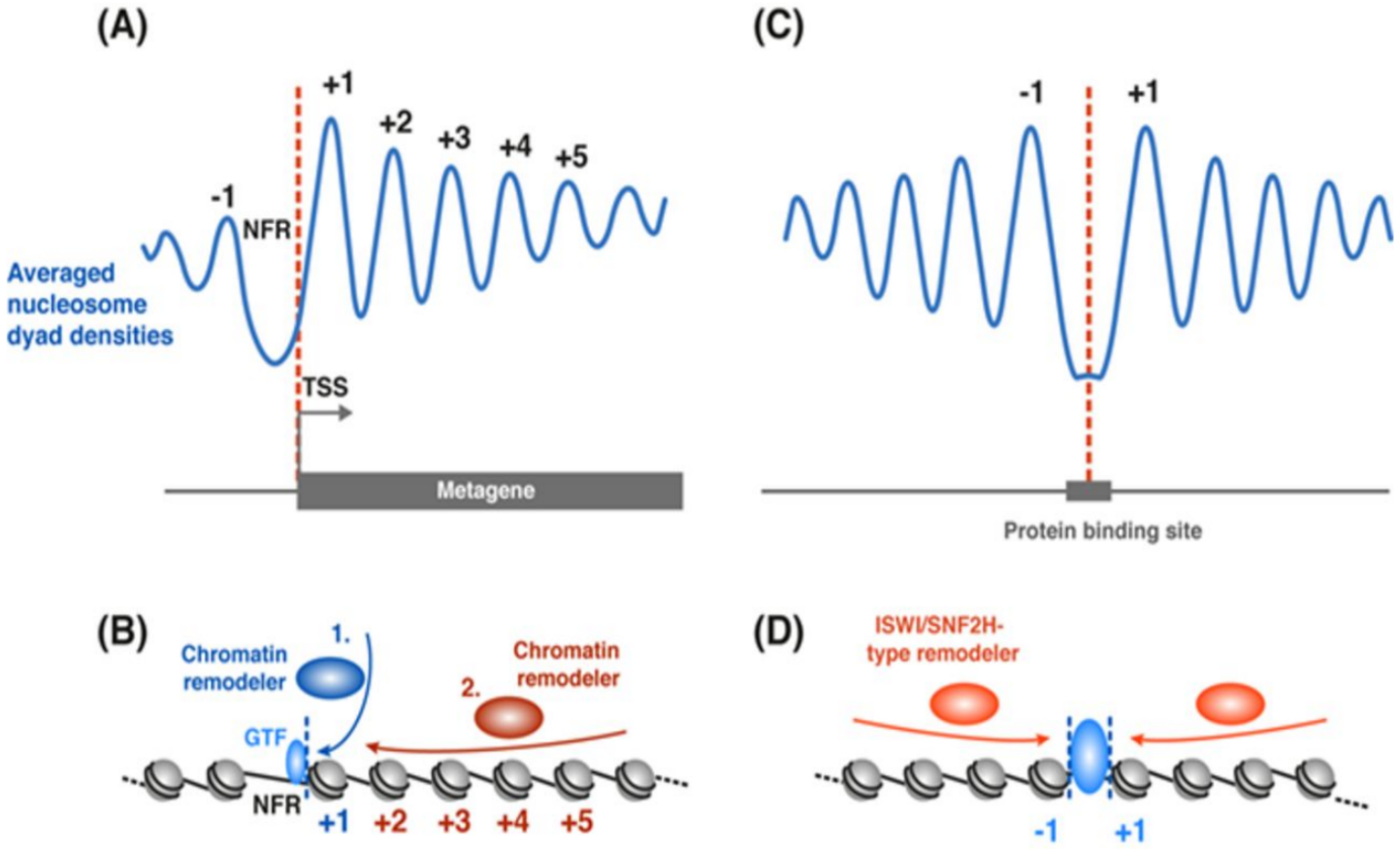
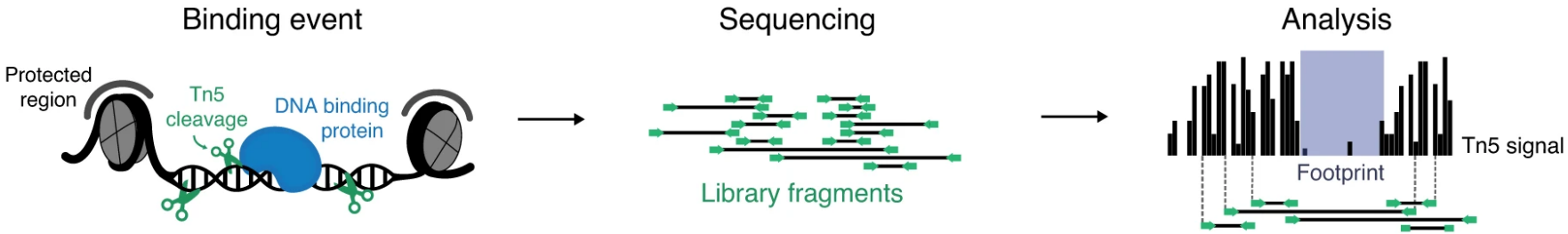
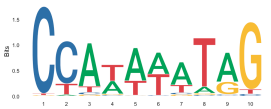


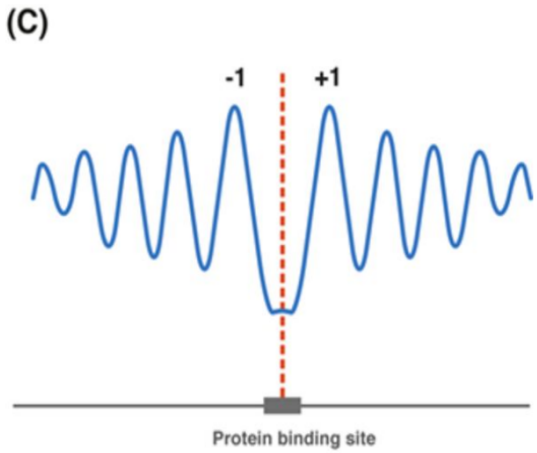
image: https://doi-org.ezp.sub.su.se/10.1042/EBC20180058

Transcription Factor Footprinting: Principle





TF footprint



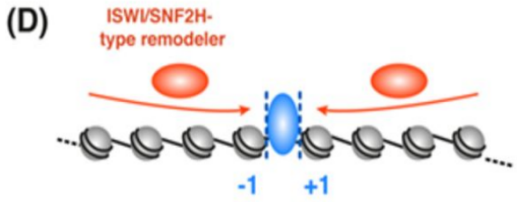
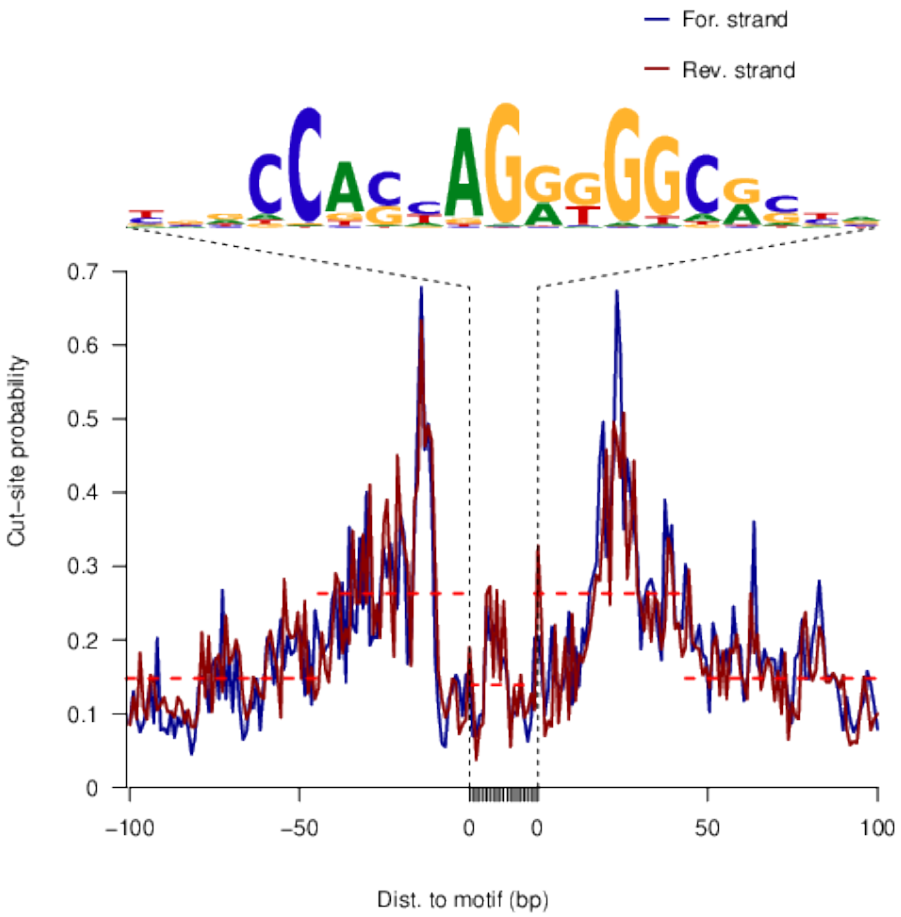


image:DOI:10.1038/s41467-020-18035-1

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

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



Transcription Factor Footprinting

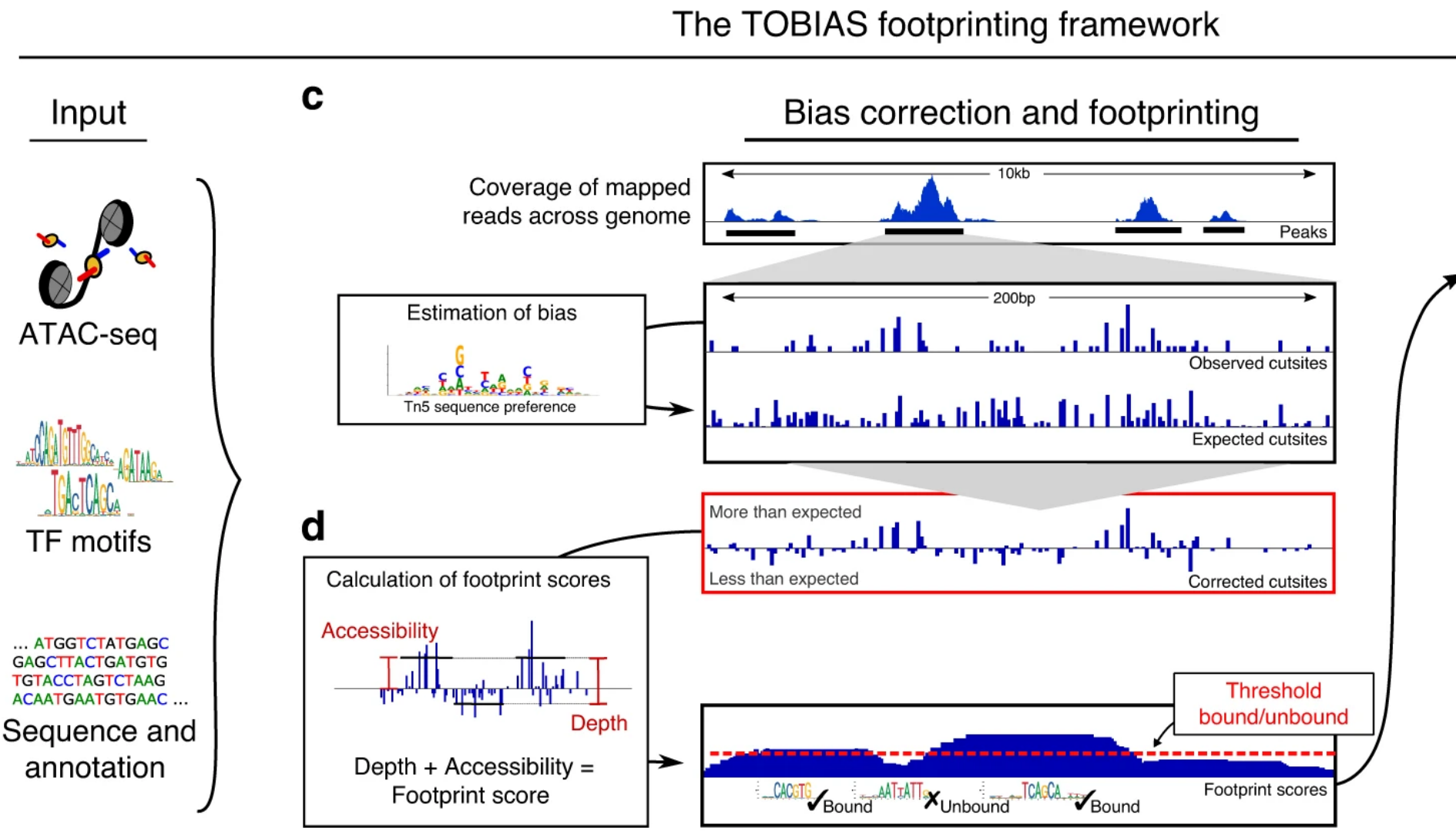
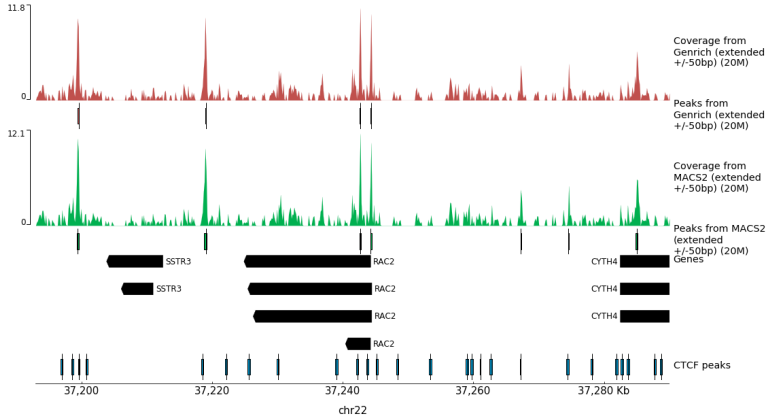


image:DOI:10.1038/s41467-020-18035-1

Analysis workflow

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

raw data *fastq

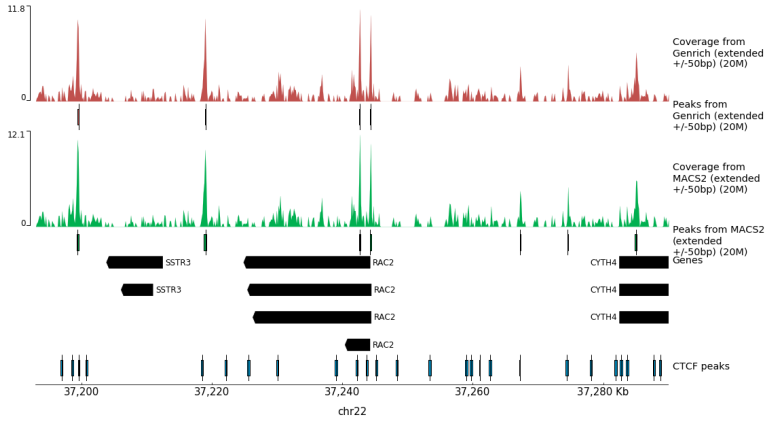


peaks & more

reads aligned to reference *bam

peaks

Analysis workflow



Sample_Read1

Sample_Read2

QC

cutadapt

Read trimming, filtering

Trimmed_Read1

Trimmed_Read2

Read mapping

bowtie2

QC

Alignments

Filtering alignments

- MAPQ
- Duplicates
- Blacklists
- Mt reads

QC

samtools

Filtered alignments

Peaks: Genrich MACS3

...

Replicate congruency

IDR, overlaps, clustering

Differential accessibility

edgeR
GC aware

Signal visualisation

deepTools
IGV ...

Annotation to genes

Bedtools
ChIPpeakAnno
ChIPSeeker

Motifs in peaks

Motif enrichment tools

Special considerations for ATAC-seq data analysis

- Paired end (PE) sequencing is recommended
- 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

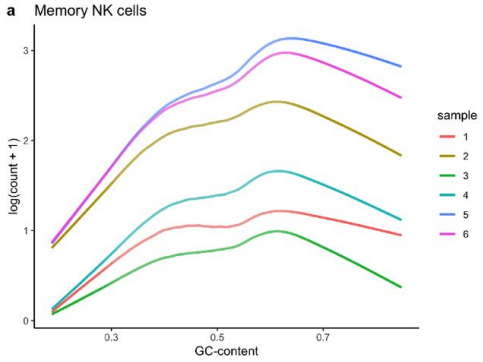
doi: https://doi.org/10.1101/2021.01.26.428252

- 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.

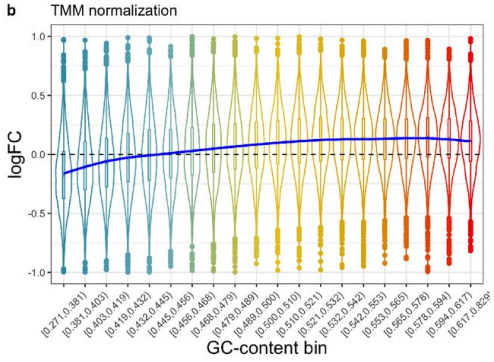
doi: https://doi.org/10.1101/2021.01.26.428252

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

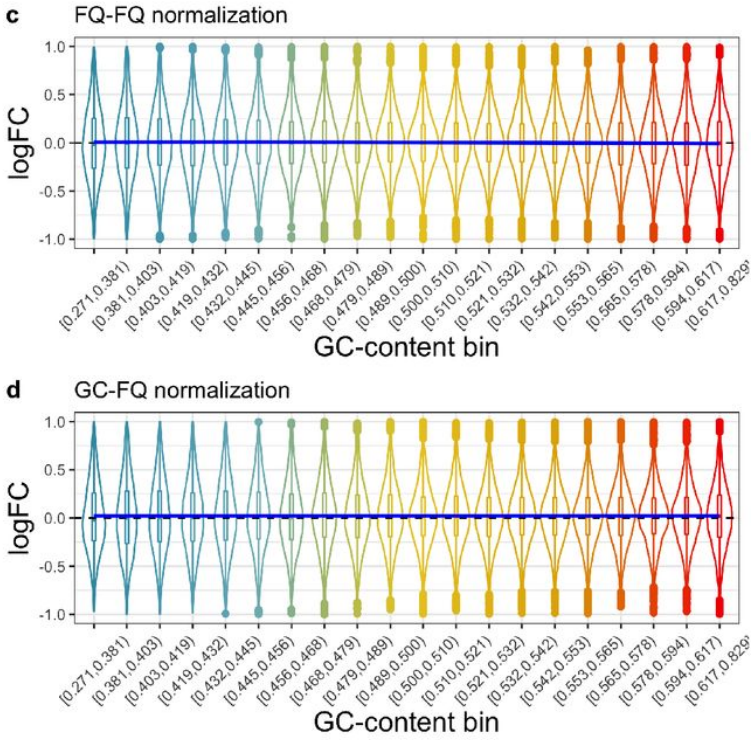
log-count as a function of GC-content



TMM normalisation



GC aware normalisation

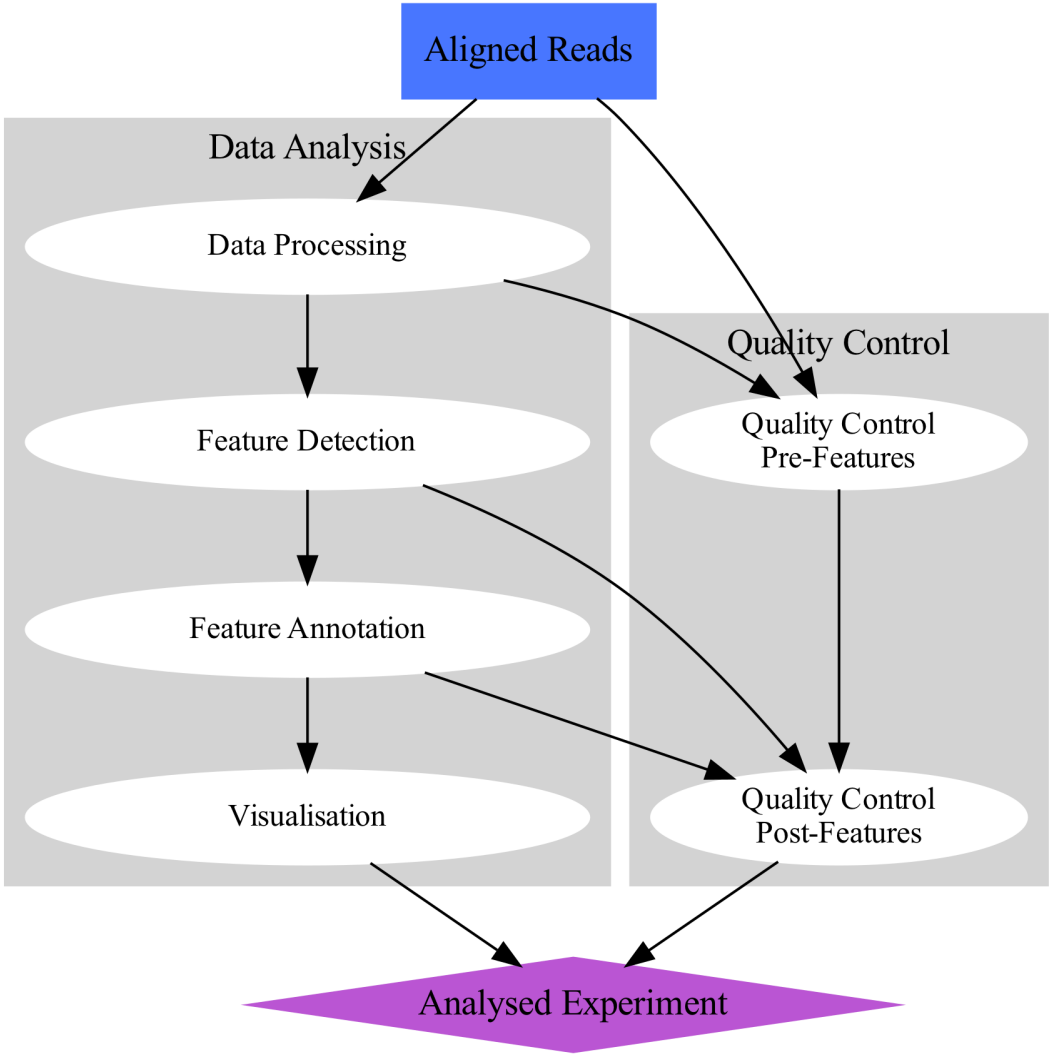


Differential accessibility
log-fold change in bins by GC content

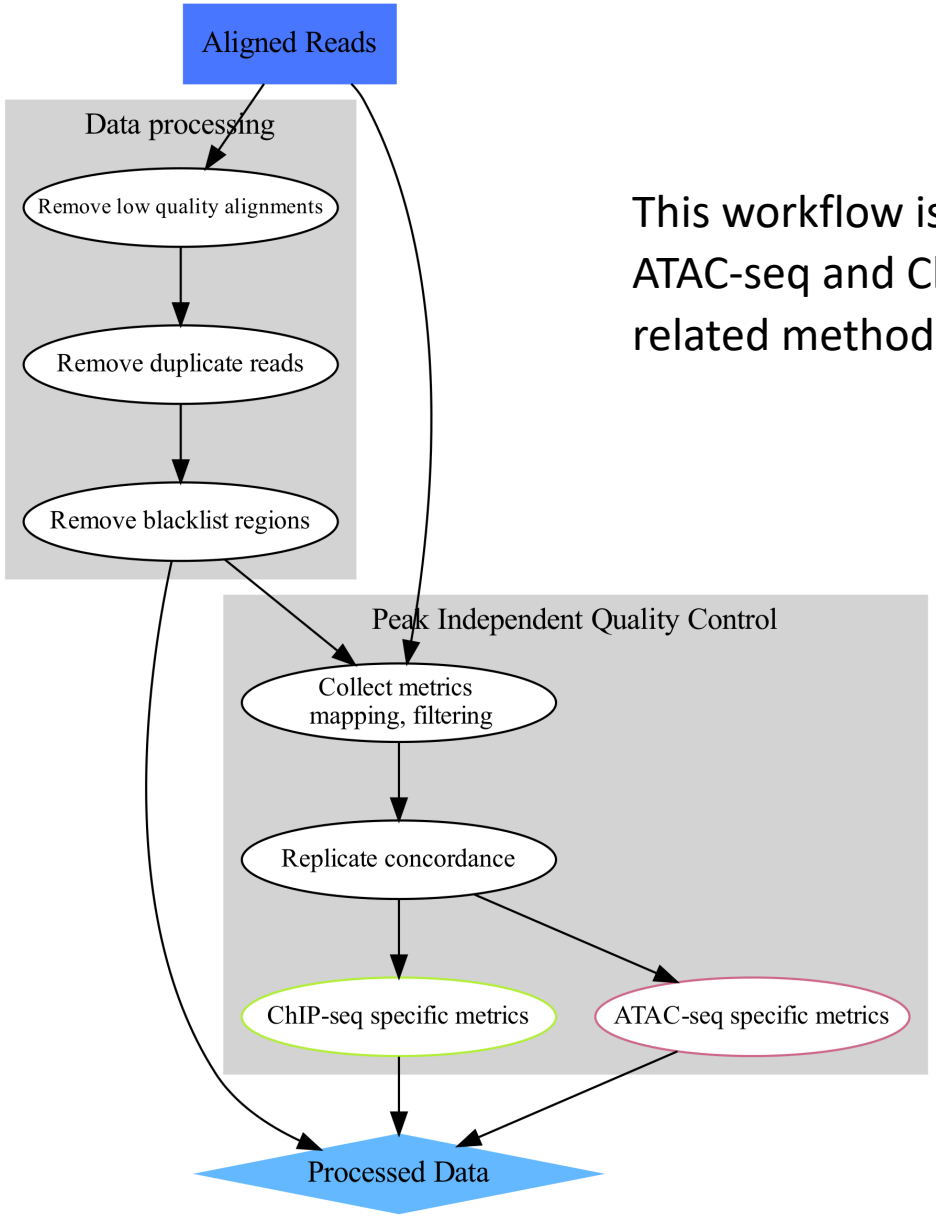
A bias for peaks with low and high GC-content (in a null setting, LFC should be centred on zero)

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

Exercise Overview

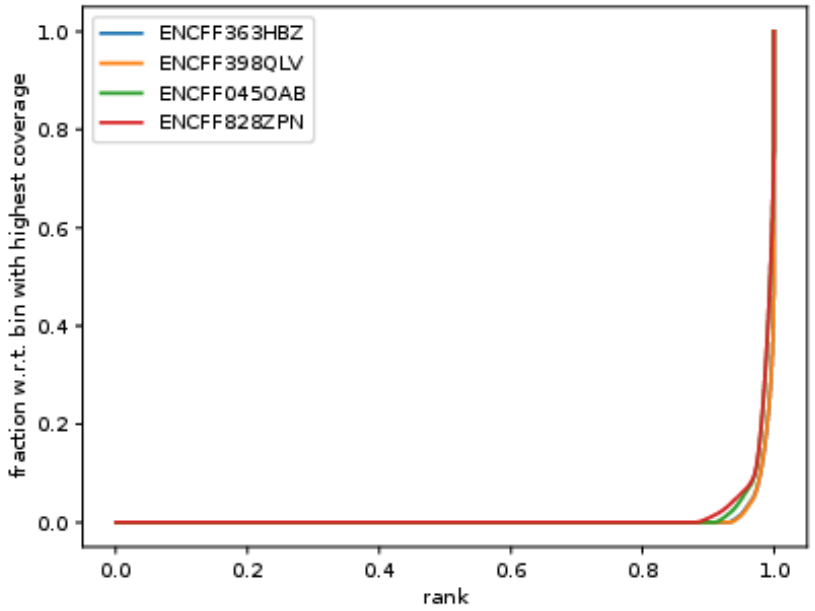


Exercise Overview: Data preprocessing and QC



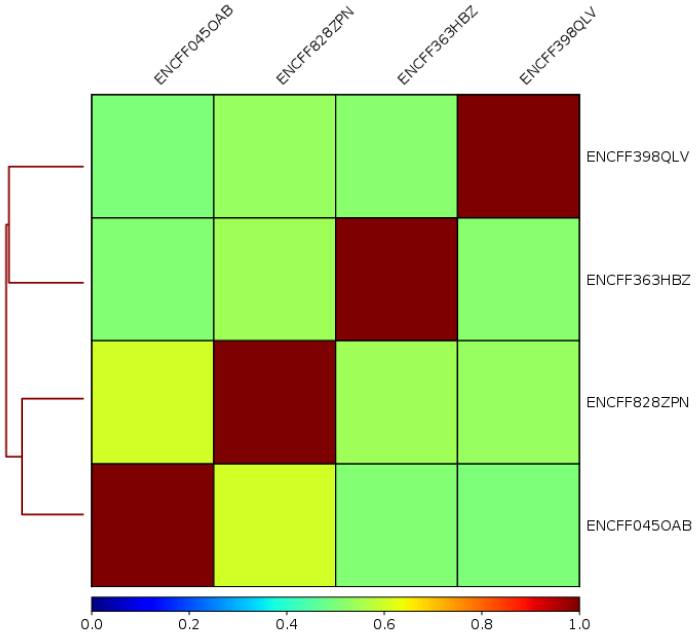
This workflow is very similar for ATAC-seq and ChIP-seq and related methods

Exercise Overview: Data preprocessing and QC



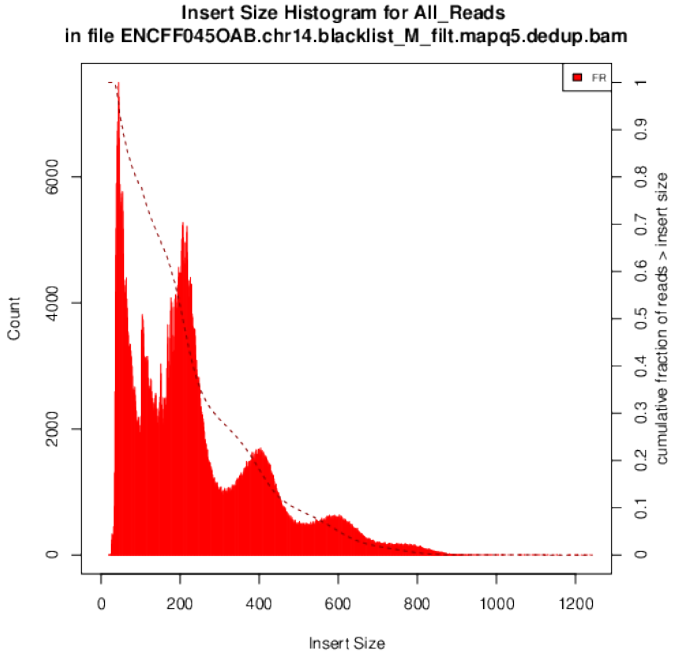
Cumulative enrichment ("fingerprint")

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



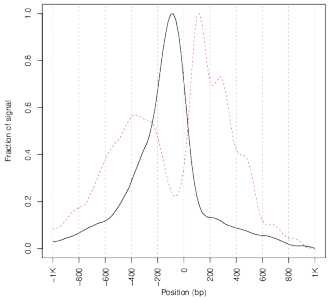
Replicate clustering

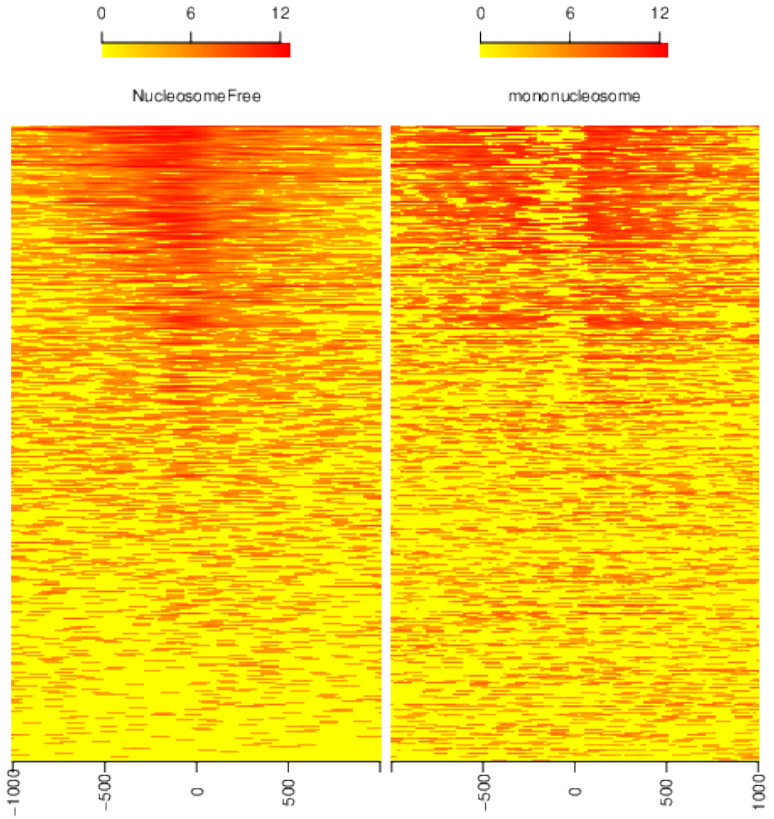
Exercise Overview: ATAC-seq specific QC



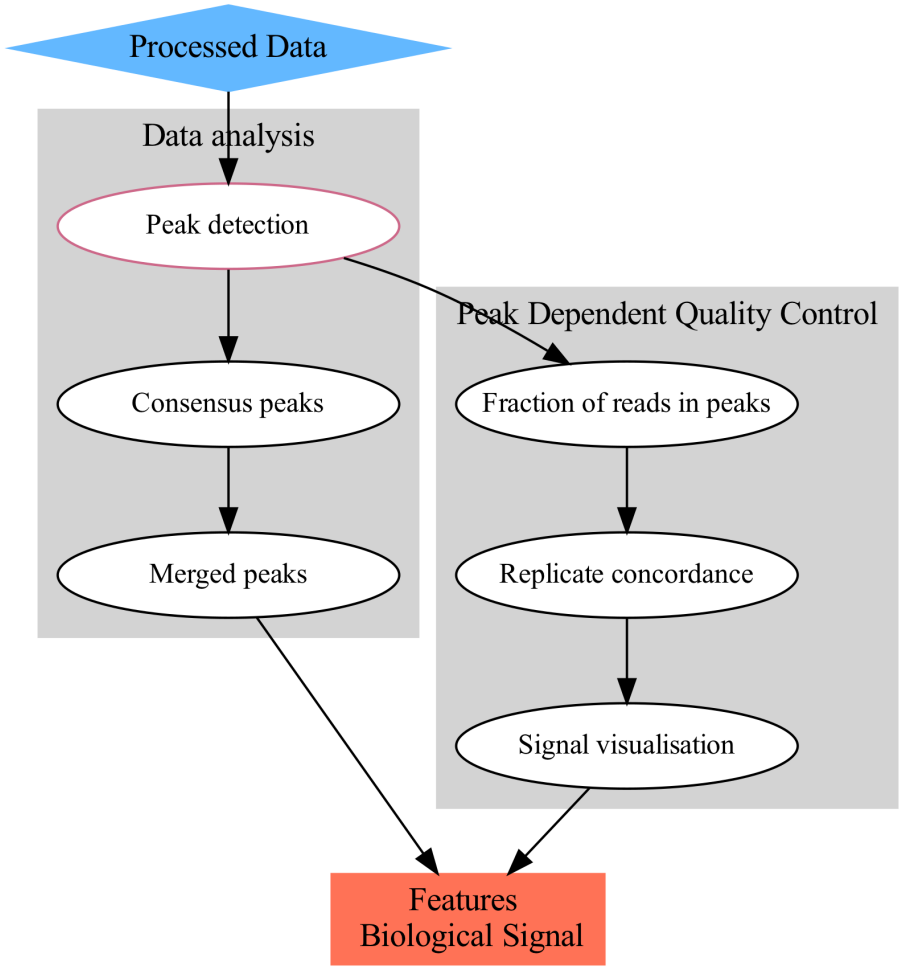
Fragment length distribution

Signal profiles at TSS

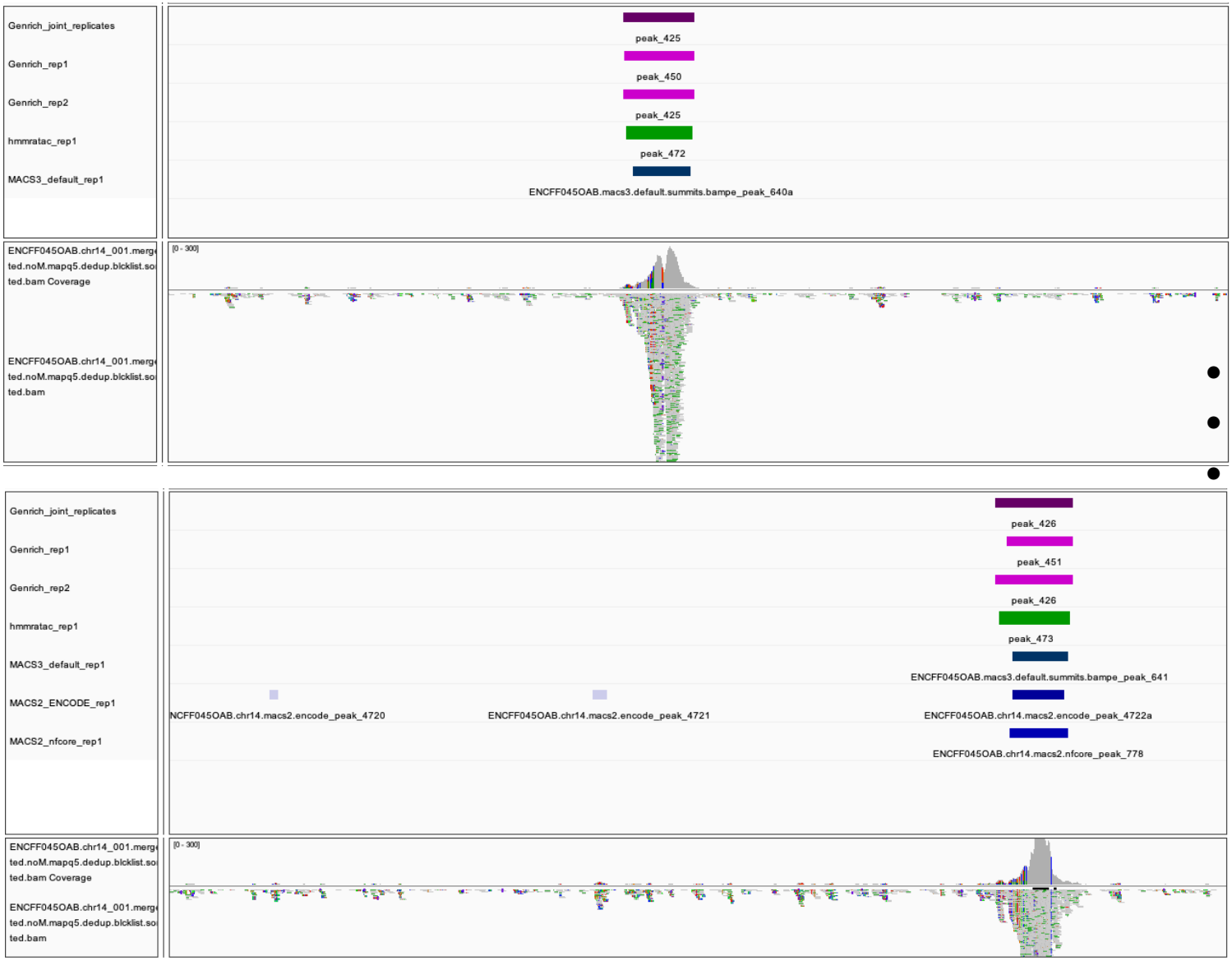




Exercise Overview: ATAC-seq peak calling



Exercise Overview: ATAC-seq peak calling



- Genrich
- MACS3 callpeak
- MACS3 hmmratac

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