Prediction of Regulatory Networks from Expression and Chromatin Data

Ivan G. Costa, RWTH Aachen University, Germany
Marcel Schulz, Saarland University & Max Planck Institute for Informatics, Germany
Matthias Heinig, Helmholtz Center Munich, Germany

RWTH AACHEN UNIVERSITY

max planck institut informatik

HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

Overview

Time	Topic	Who
2:30 - 2:45	Introduction / gene regulation / transcription / chromatin	IC
2:45 - 3:00	Introduction ChIP-seq peak calling	MH
3:00 - 3:50	Practical peak calling	MH & JH
4:15 - 4:30	Introduction Footprints	IC
4:30 - 4:45	Introduction Regulatory networks	MS
4:45 - 5:50	Practical Regulatory Networks	IG, MS & FS
5:50 - 6:00	Q & A session	all

Material - https://github.com/SchulzLab/EpigenomicsTutorial-ISMB2017

Team

Ivan Costa (IC)

Matthias Heinig (MH)

Johann Hawe

Marcel Schulz(MH)

Florian Schmidt (FS)

Introduction to Footprints

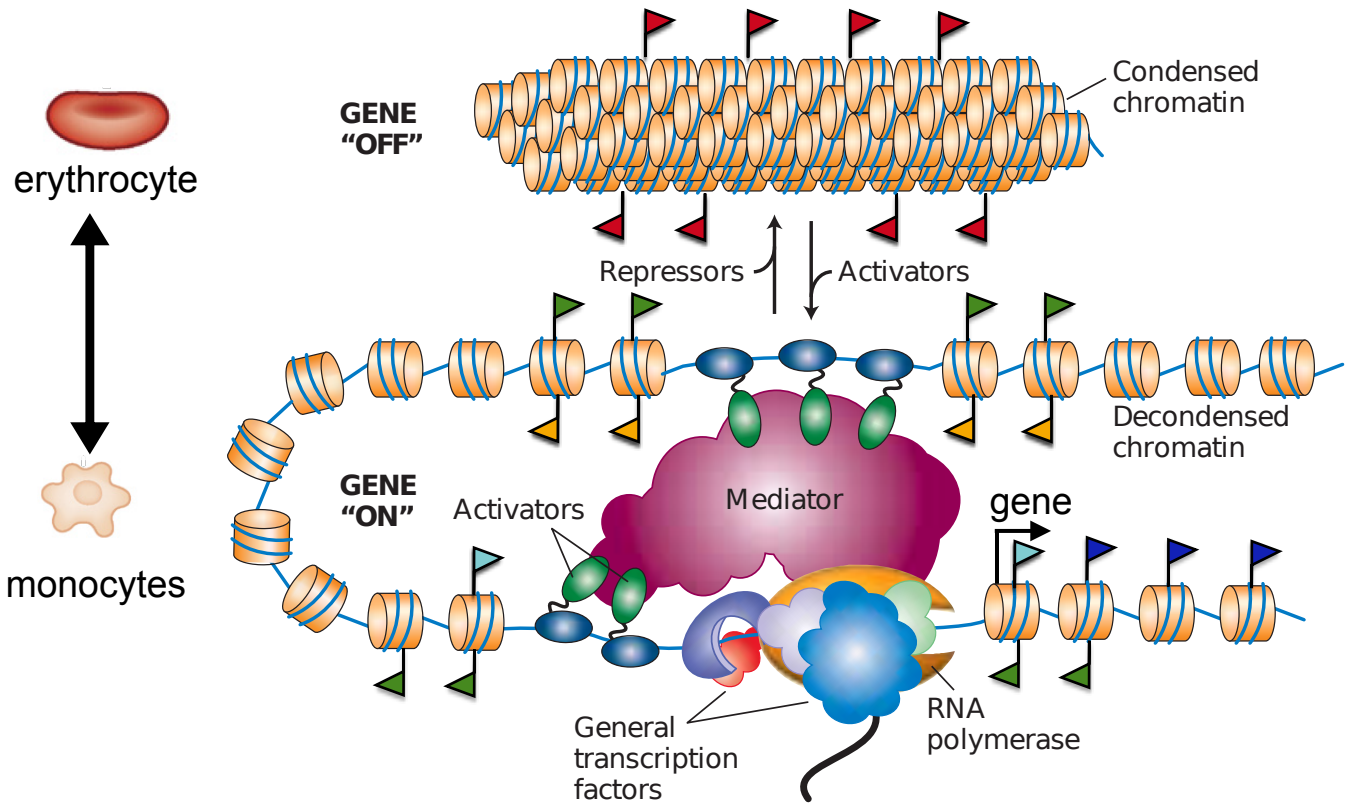
Ivan G. Costa
RWTH Aachen University, Germany

www.costalab.org

AI ces

IZKF Interdisziplinäres Zentrum für Klinische Forschung | RWTH AACHEN UNIVERSITY

Chromatin and Cell Memory/Plasticity



Histone Code

Transcription

H3K79me2, H3k36me3

Active Regions

H3K27ac, H3K9ac

Active Promoters

H3K4me3

Active Enhancers

H3K4me1

Repressed Prom.

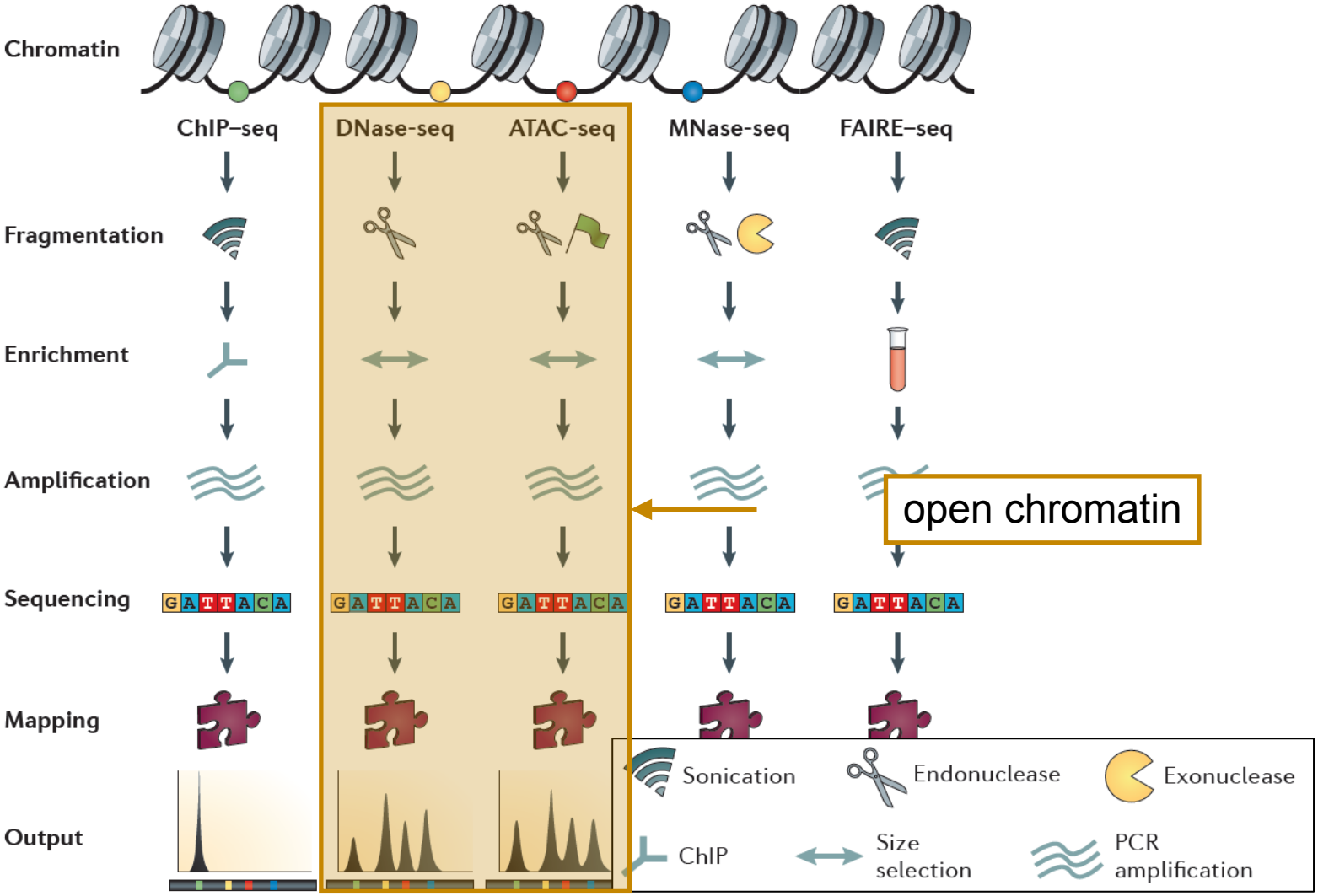
H3K27me3

Repressed Regions

H3K9me3

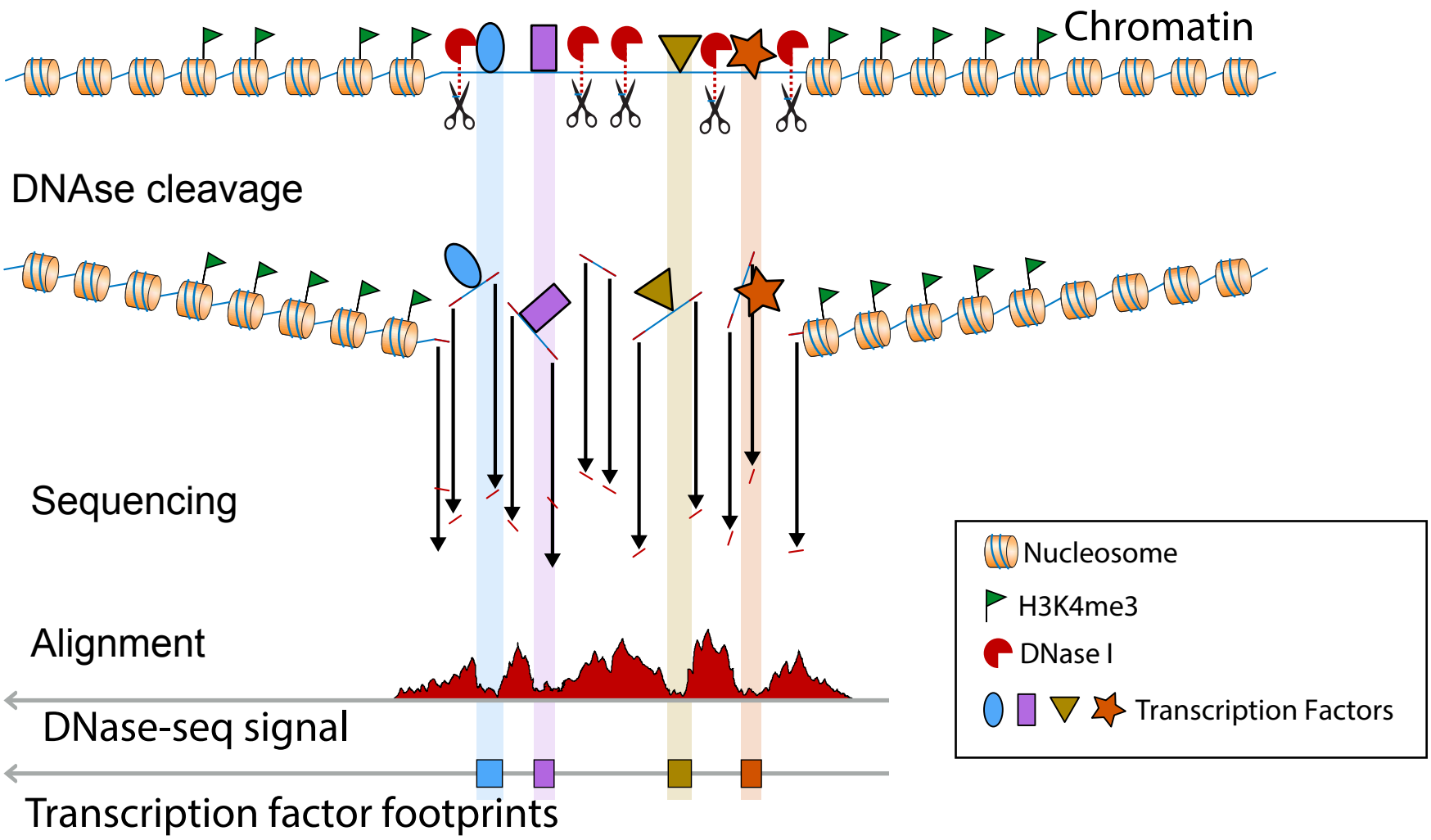
Adapted from Lodish, B. et al. (2004) 5th ed.

NGS and Chromatin

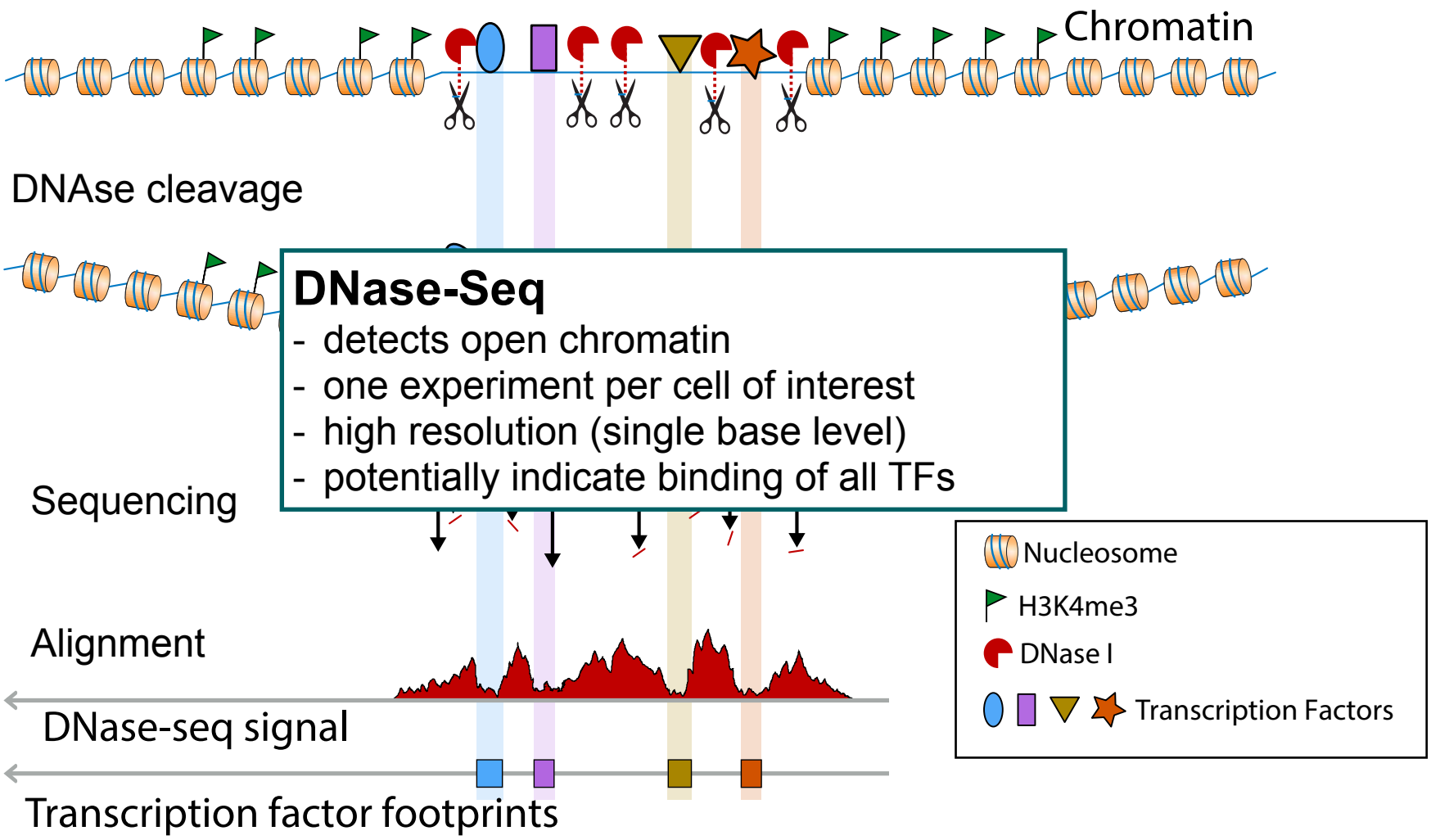


Source: Meyer, C.A. *and Liu X.S.* (2014). *Nature Reviews Genetics*.

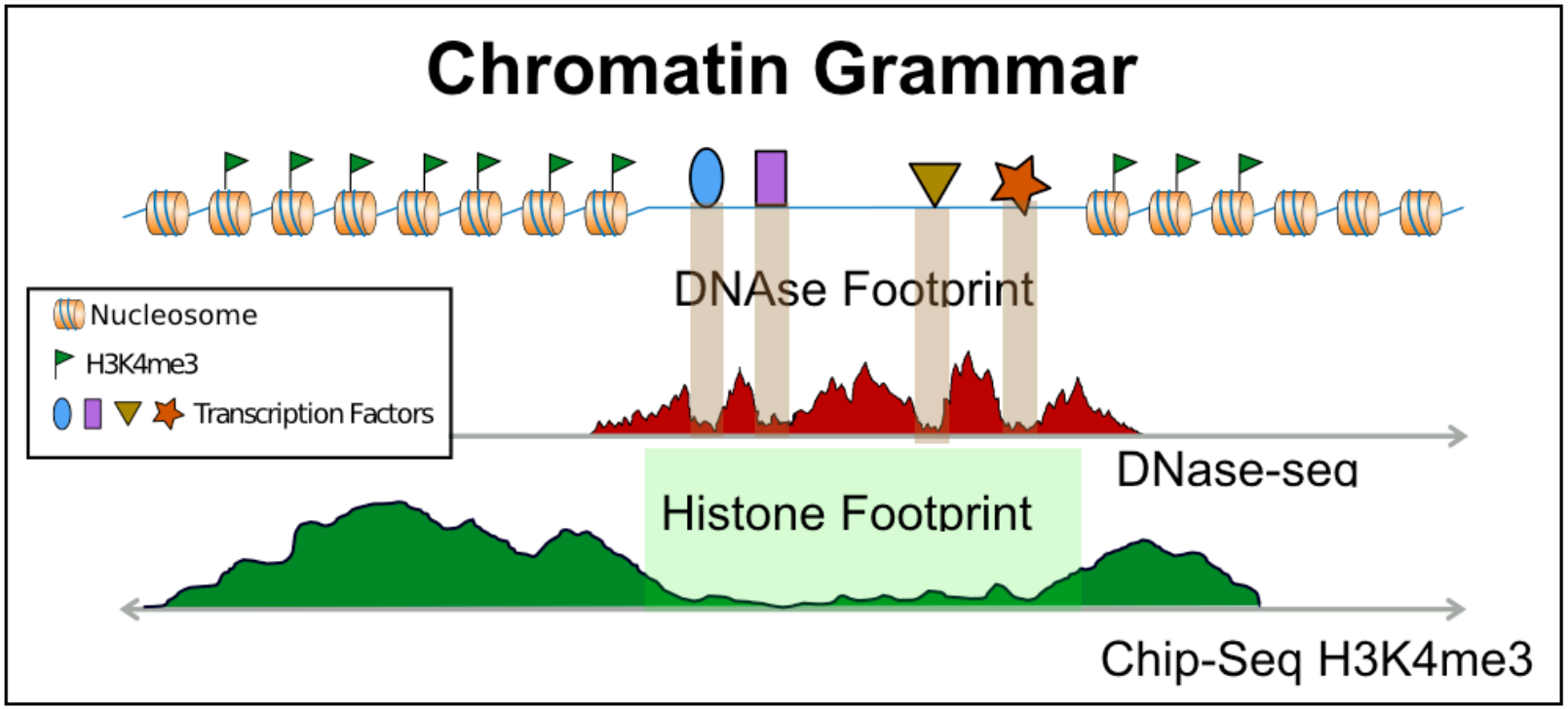
DNA - Protein interactions with DNase-seq



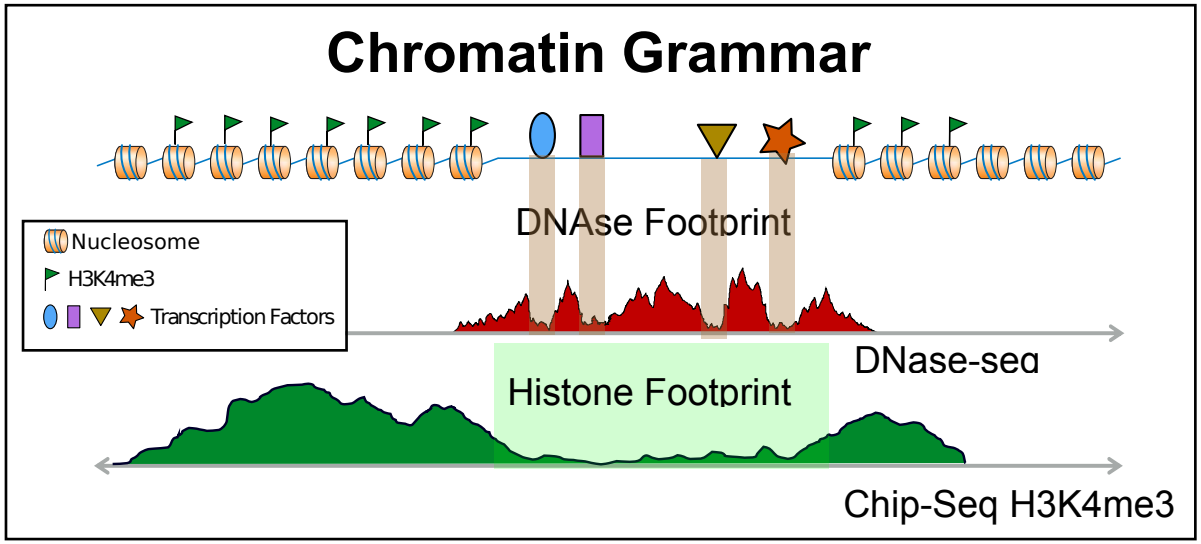
DNA - Protein interactions with DNase-seq



Detection of Active Binding Sites



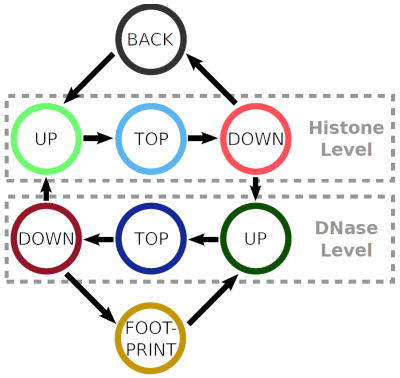
Detection of Active Binding Sites



HINT (Hmm-based IdeNtification of Transcription factor footprints)

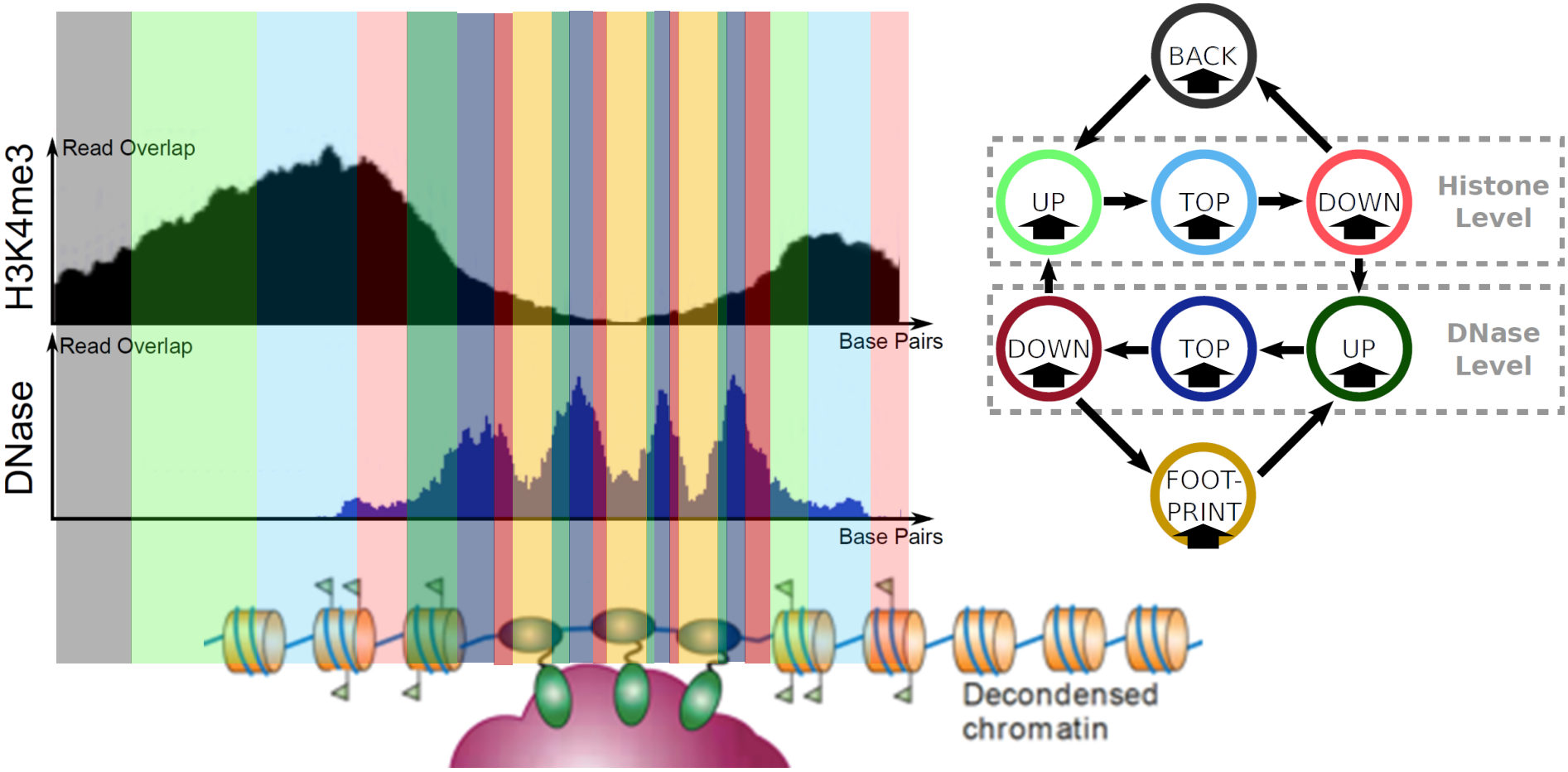
- scan DNase and/or ChIP-Seq (activating marks) to predict footprints
- normalization for cleavage bias and global artifacts
- obtain cell independent models

8 State HMM



Gusmao EG *et. al*, (2014), Bioinformatics, 30(22):3143-51.
Gusmao EG *et. al,* (2016) Nature Methods, 13, 303–309.

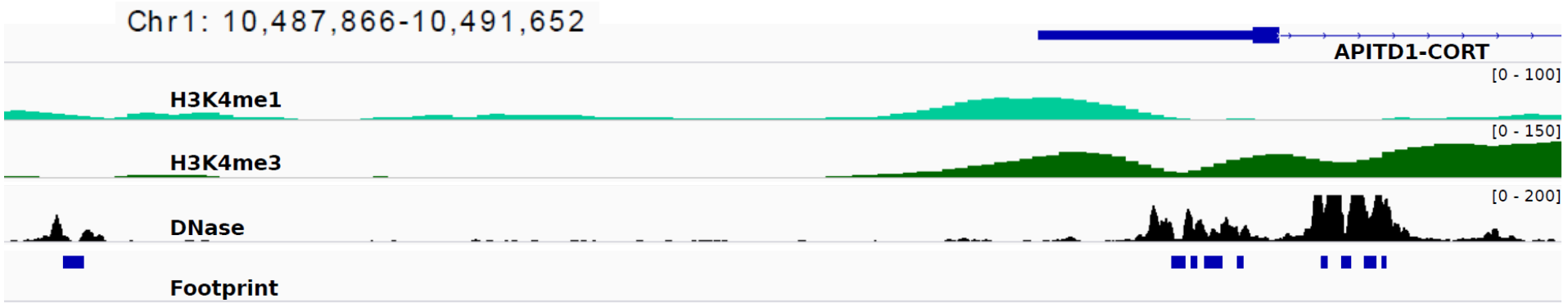
Method - HINT



- Emissions - multivariate Gaussian (signal and slope of histone and DNase)
- HMM trained on manually annotated region

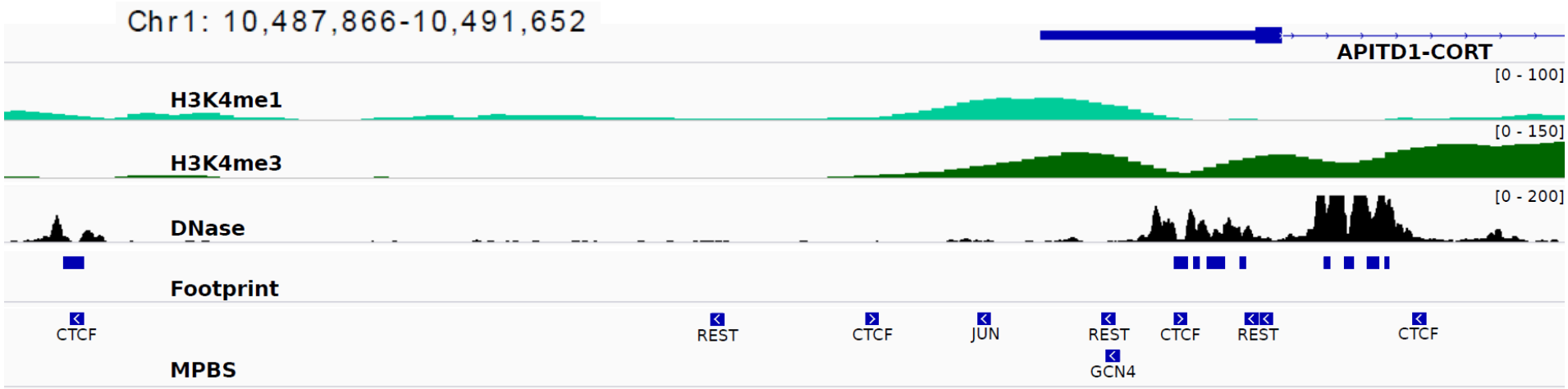
Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

Example - K562 cells



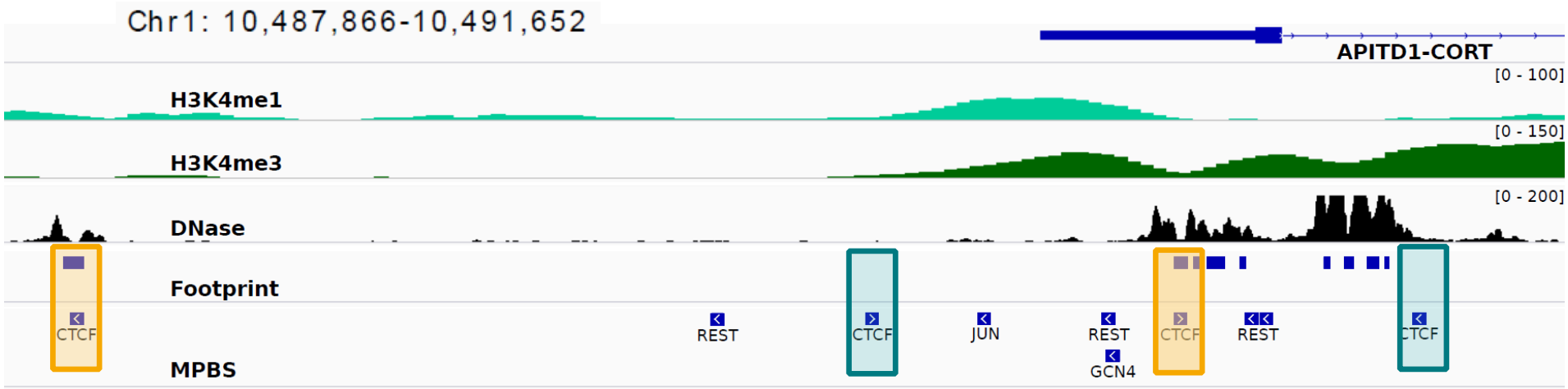
Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

Example - K562 cells



Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

Example - K562 cells



footprint (active)

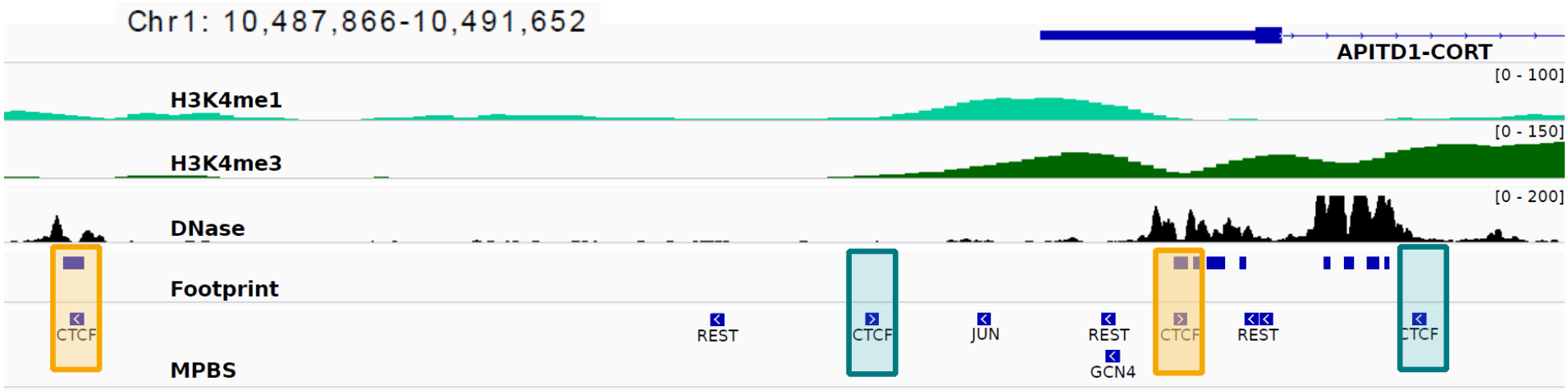
no footprint (inactive)

footprint (active)

no footprint (inactive)

Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

Example - K562 cells

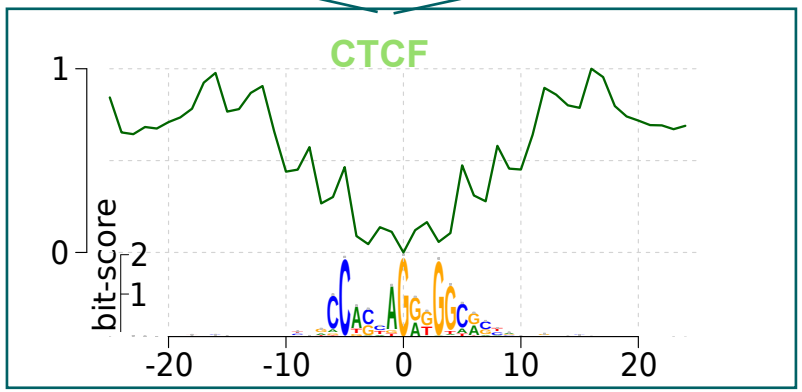


footprint (active)

no footprint (inactive)

footprint (active)

no footprint (inactive)



Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

DNase-seq Artifacts - Cleavage Bias

- DNase I prefers to bind (and cleave) some DNA regions …

...ACCGGG... ...CCCGGG...

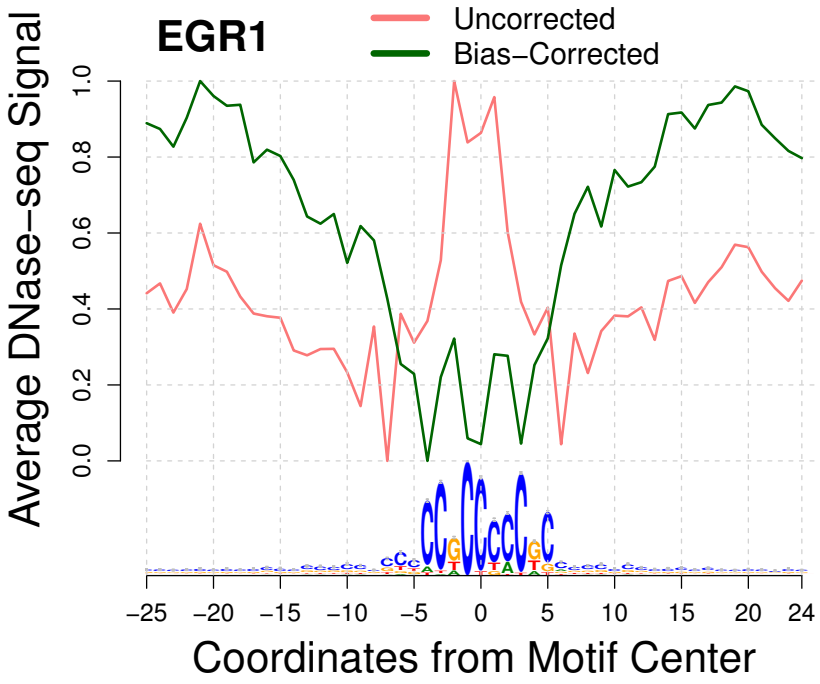
- … than other DNA regions.

...TCAATT... ...GCATTT...

- Example*: We observe a ~3.5 higher frequency of reads starting in **ACCGGG** than the frequency of **ACCGGG** in the genome .

- For a given position *i* around a *k-mer* with x_i reads

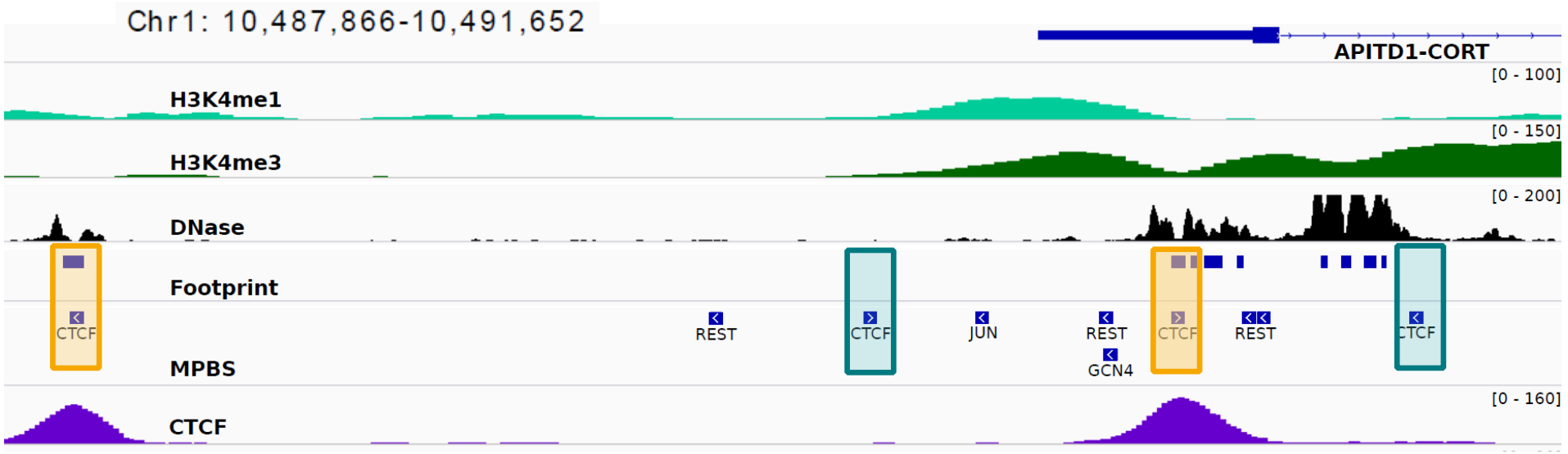
$$corrected\ x_i = x_i\ /\ bias_{k\text{-}mer}$$



*All examples are for cell type K562 (Single-hit prot. – Crawford's Lab).

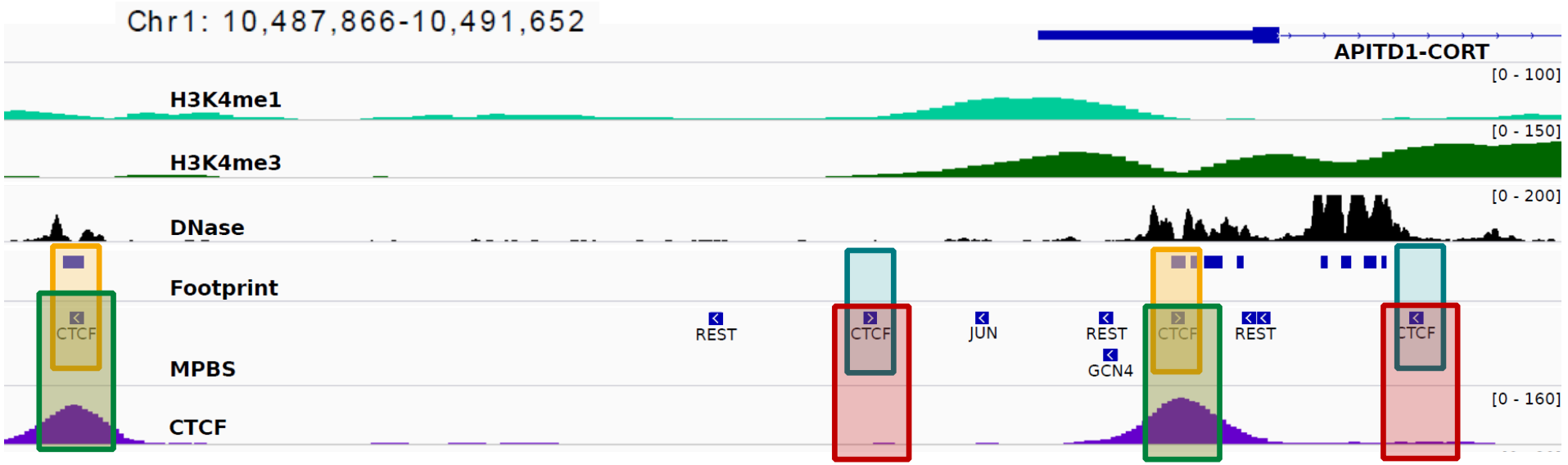
He, H.H. *et al.* (2014). *Nature Methods* 11(1):73-8.

Gold Standard - TF ChIP-Seq and motif search



Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

Gold Standard - TF ChIP-Seq and motif search



ChIP-Seq	Footprint	Outcome
with	*Active*	*True Positive*
with	*Inactive*	*False Negative*
without	*Active*	*False Positive*
without	*Inactive*	*True Negative*

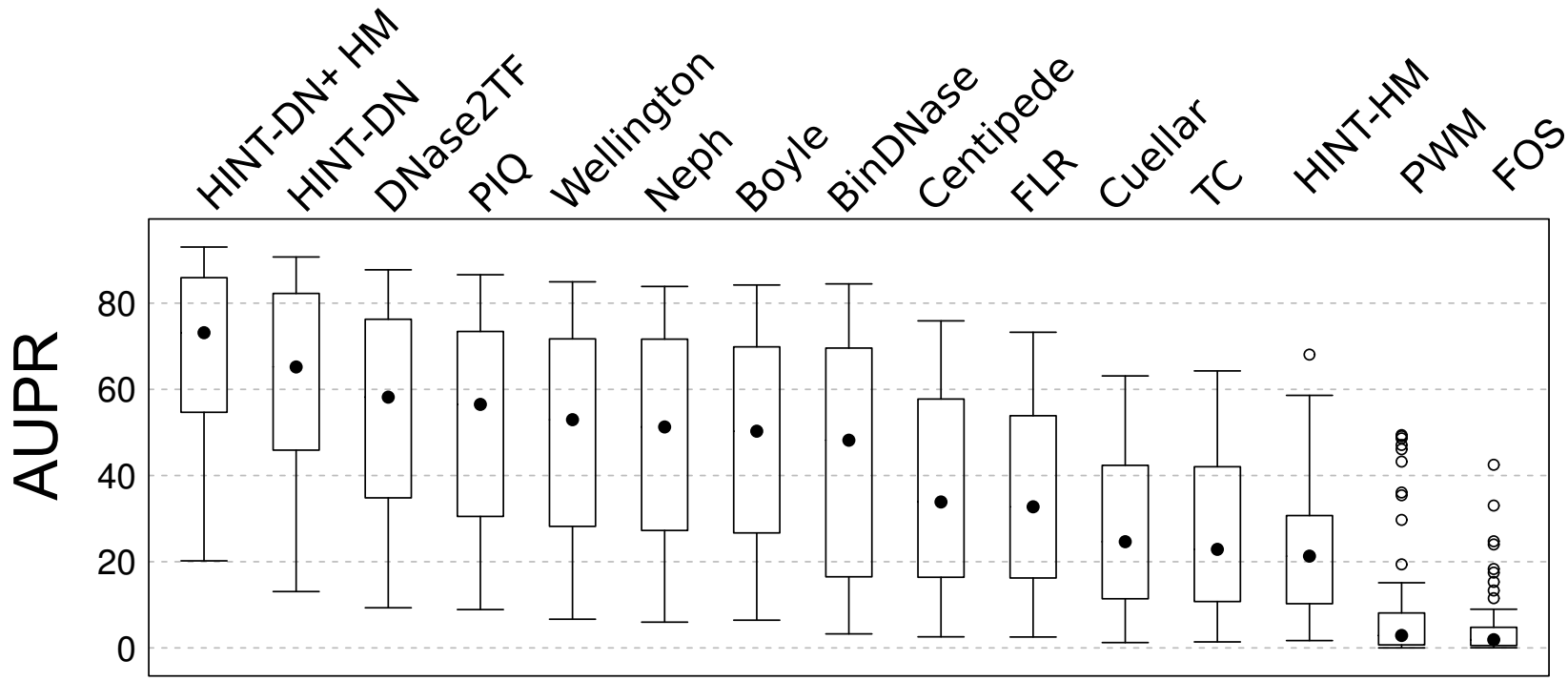
Sensitivity

Specificity

ROC
AUC
AUPR

Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

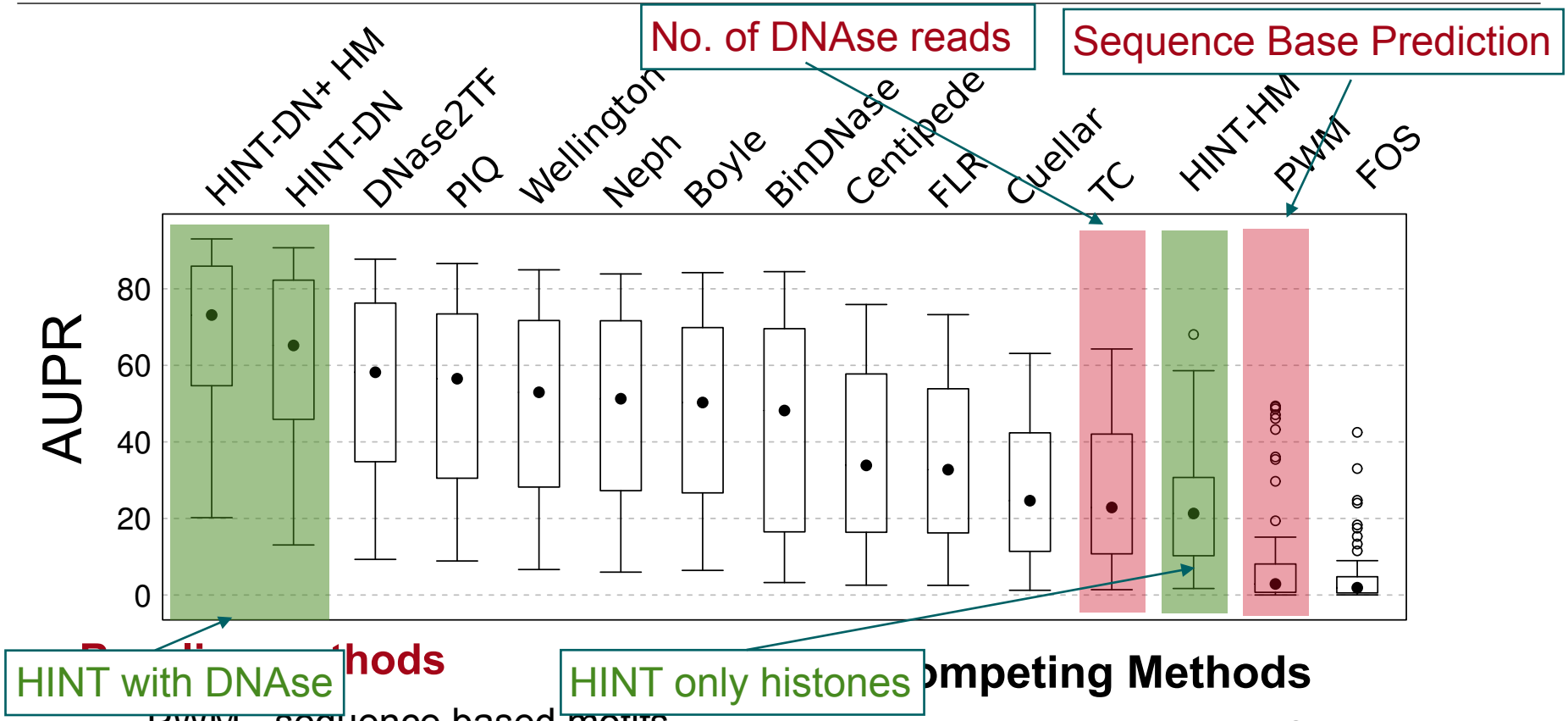
Evaluation on 88 Transcription factors H1-ESC and K562



- **Baseline methods**
  - PWM - sequence based motifs
  - TC - number of DNase reads
- **Proposed Methods**
  - HINT-DN (DNase), HINT-HM (H3K4me3 & H3K4me1)
- **Competing Methods**
  - BinDNAse, Boyle, Centipede, Cuellar, DNAse2TF, FLR. FOS, Neph, PIQ, Wellington

Gusmao EG *et. al,* (2016) Nature Methods, 13, 303–309.

Evaluation on 88 Transcription factors H1-ESC and K562



- **Proposed Methods**
  - HINT-DN (DNase), HINT-HM (H3K4me3 & H3K4me1)

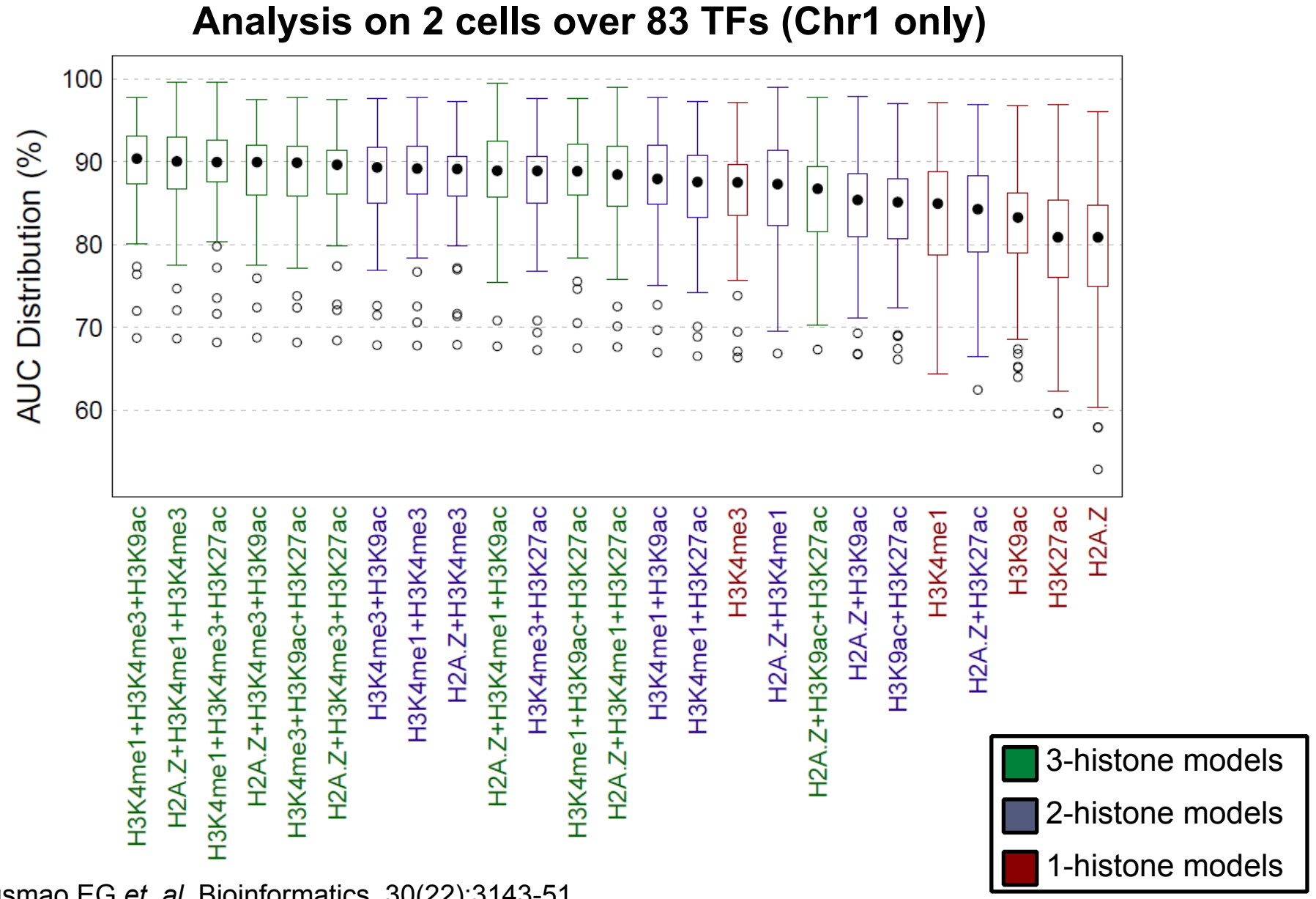
ompeting Methods

- BinDNAse, Boyle, Centipede, Cuellar, DNAse2TF, FLR. FOS, Neph, PIQ, Wellington

- PWM - sequence based motifs
- TC - number of DNase reads

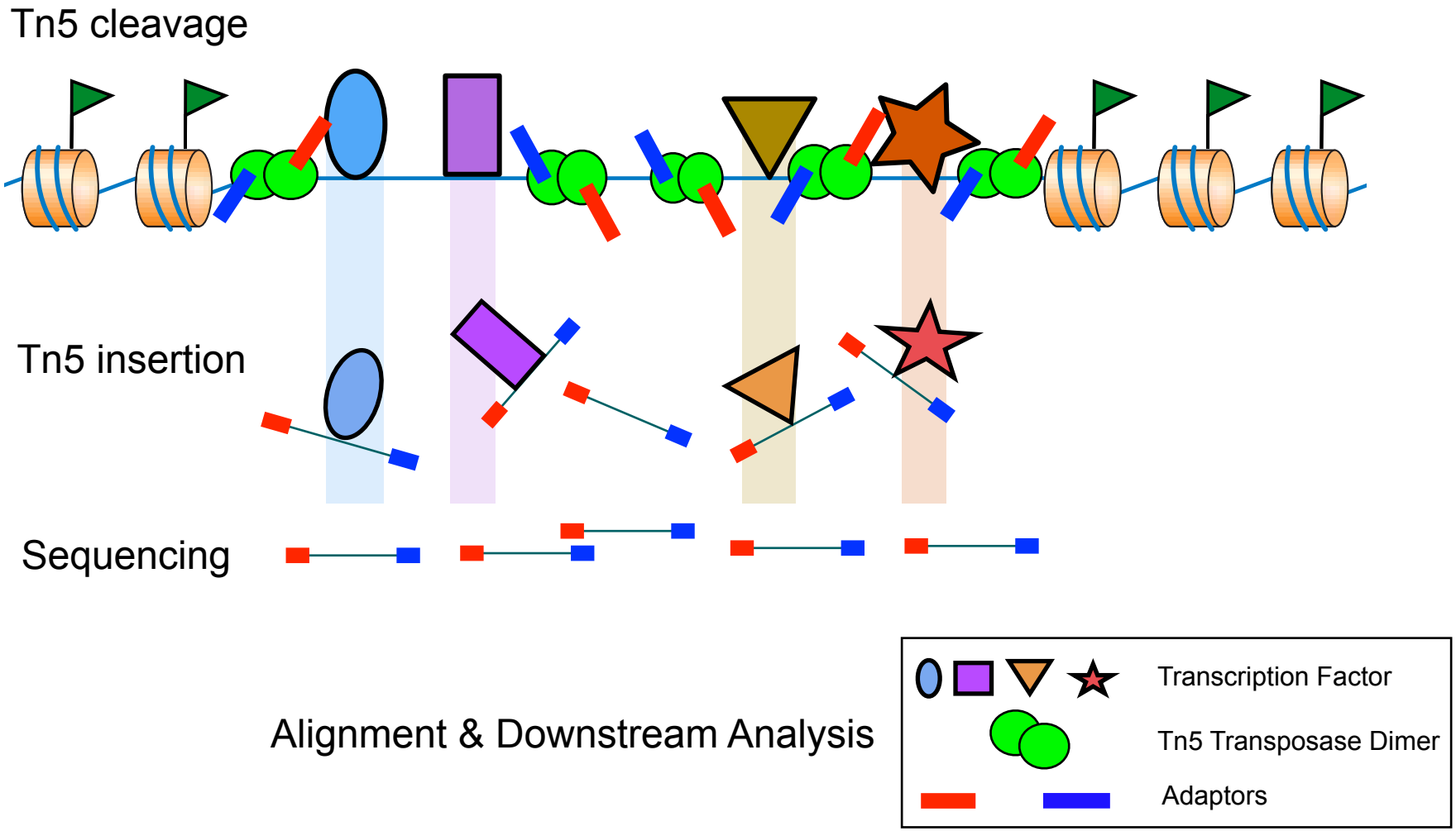
Gusmao EG *et. al,* (2016) Nature Methods, 13, 303–309.

Evaluations - Which histone modifications?



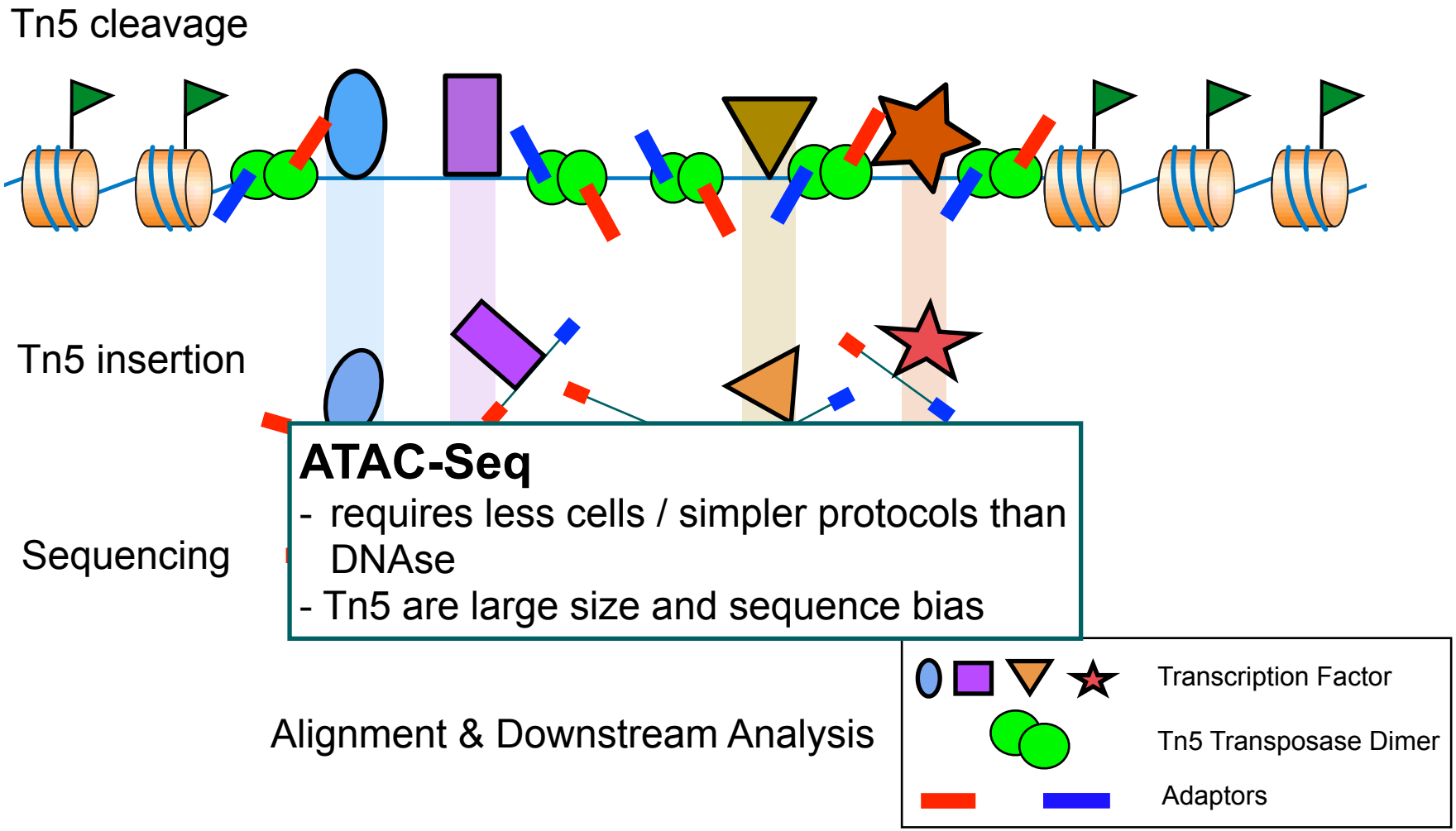
Gusmao EG *et. al*, Bioinformatics, 30(22):3143-51.

DNA - Protein interactions with ATAC-seq



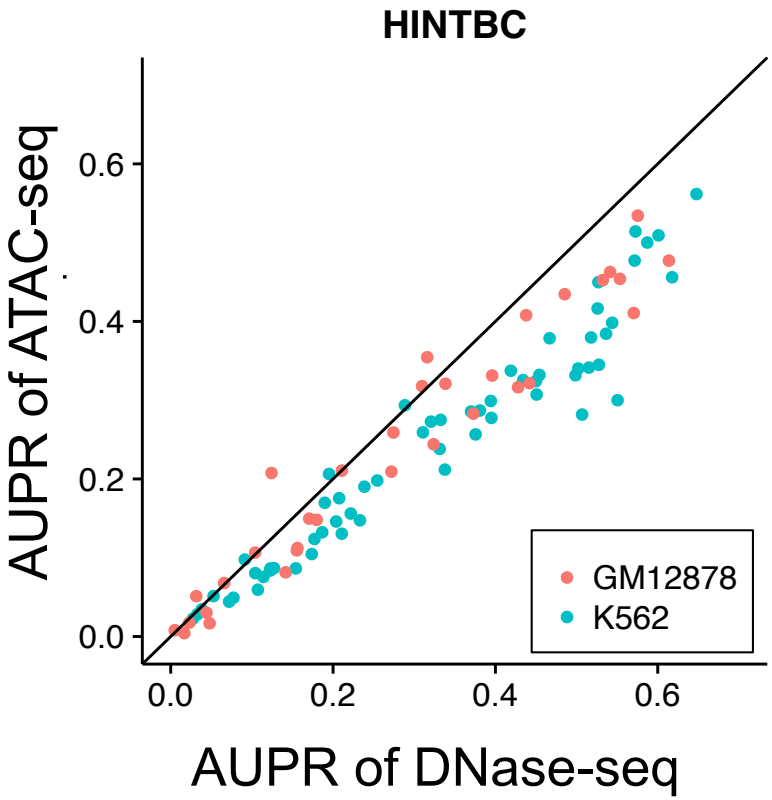
Li *et. al, unpublished*

DNA - Protein interactions with ATAC-seq



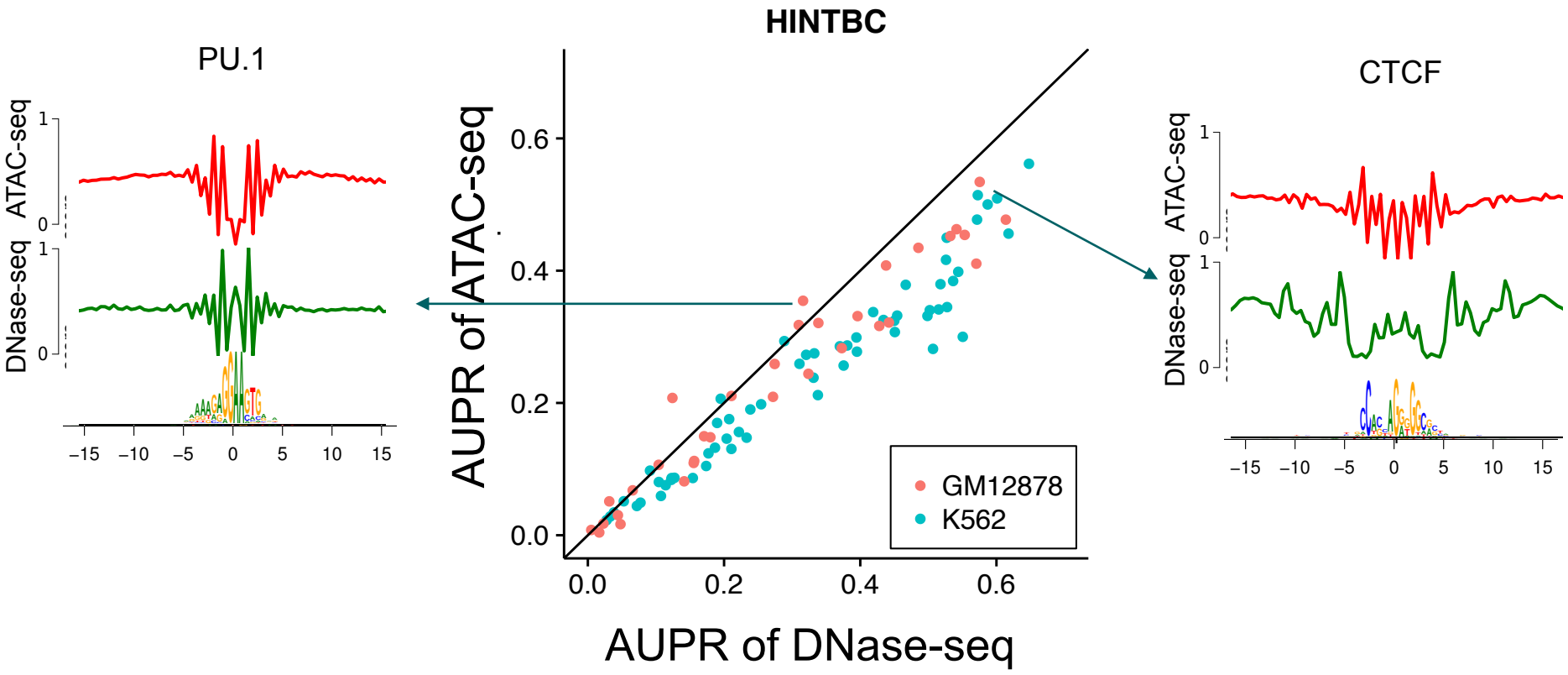
Li *et. al, unpublished*

Open Chromatin Protocols - Comparison



Li *et. al, unpublished*

Open Chromatin Protocols - Comparison



Li *et. al, unpublished*

Overview

Footprint analysis

- allow detection of cell specific binding sites
- cleavage bias correction is crucial in DNAse-seq

Alternatie chromatin protocols (ATAC-seq)

- alternative for experiments with low cell counts
- footprinting is comparable to DNase-seq
    - also requires bias correction

Practical Footprints

Ivan G. Costa
RWTH Aachen University, Germany

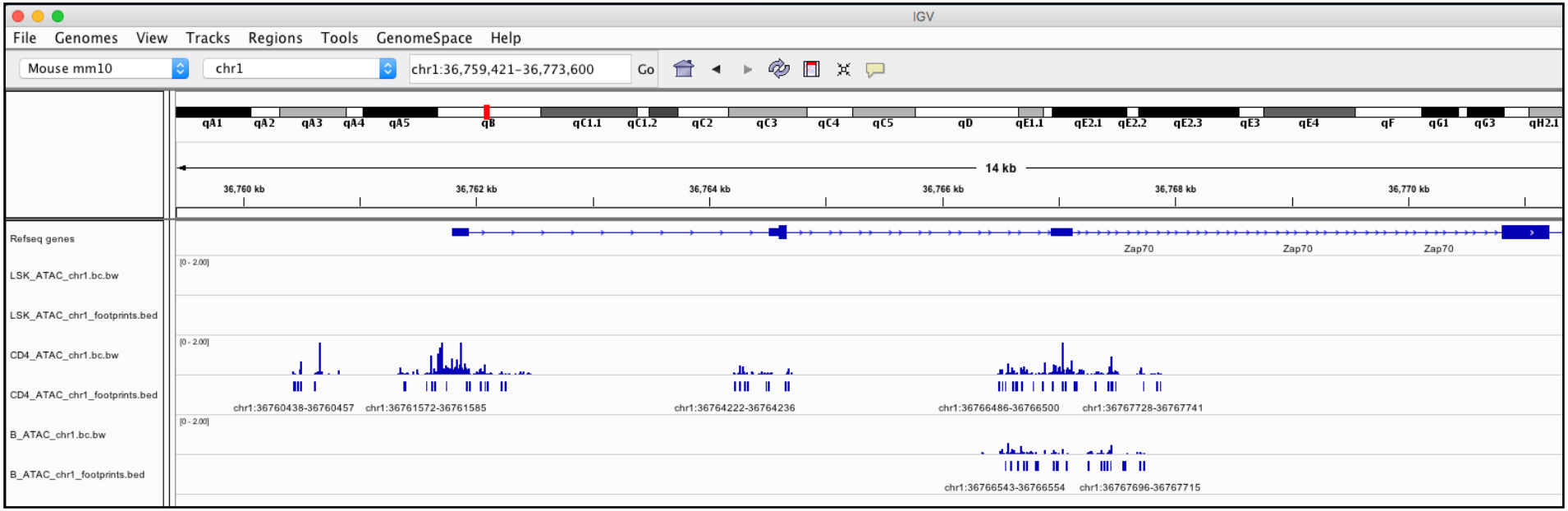
www.costalab.org

AICES

IZKF Interdisziplinäres Zentrum für Klinische Forschung | RWTH AACHEN UNIVERSITY

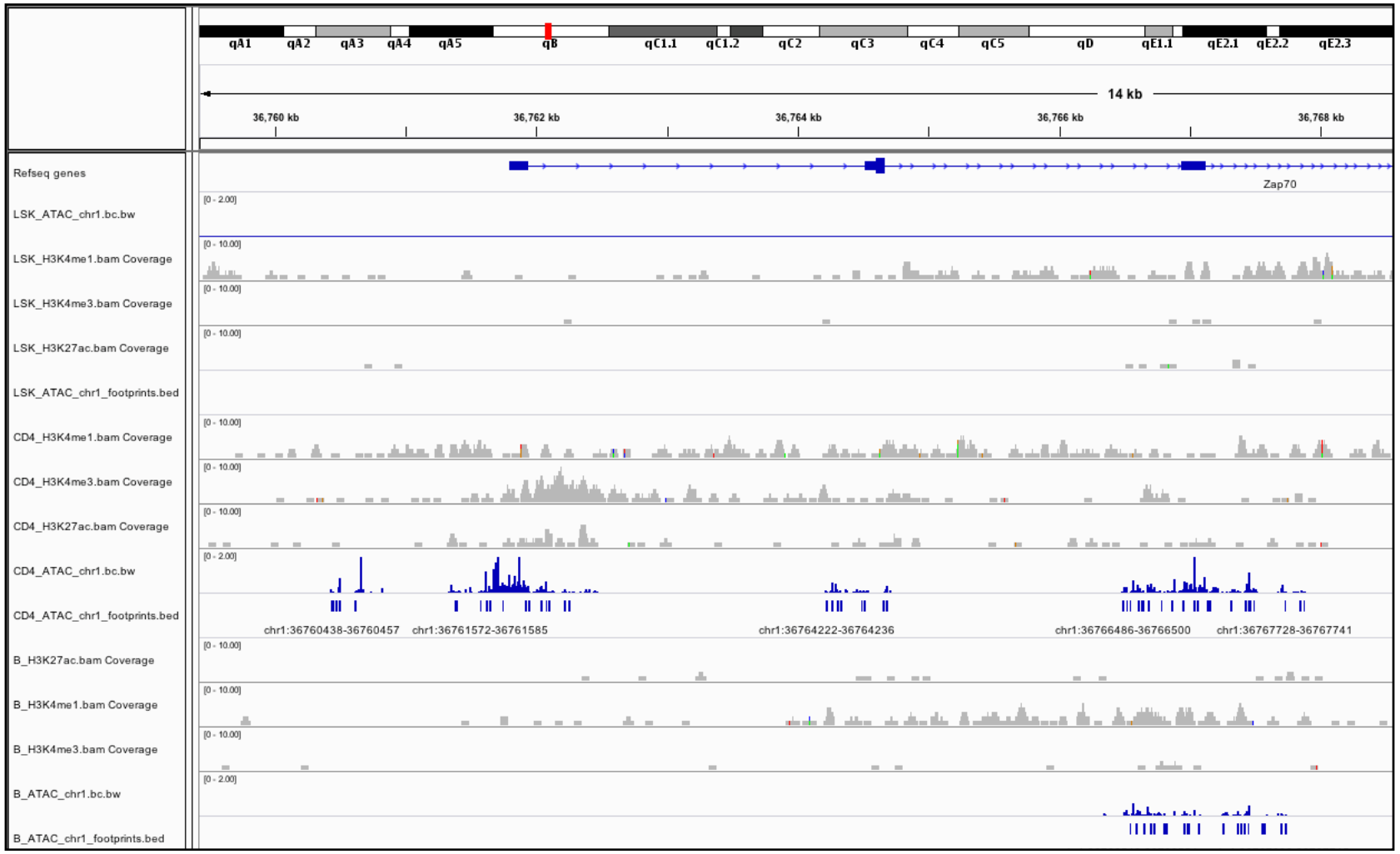
Tutorial / IGV Screenshots

ATAC-seq profiles (bigWig) and Footprints around Zap70 locus



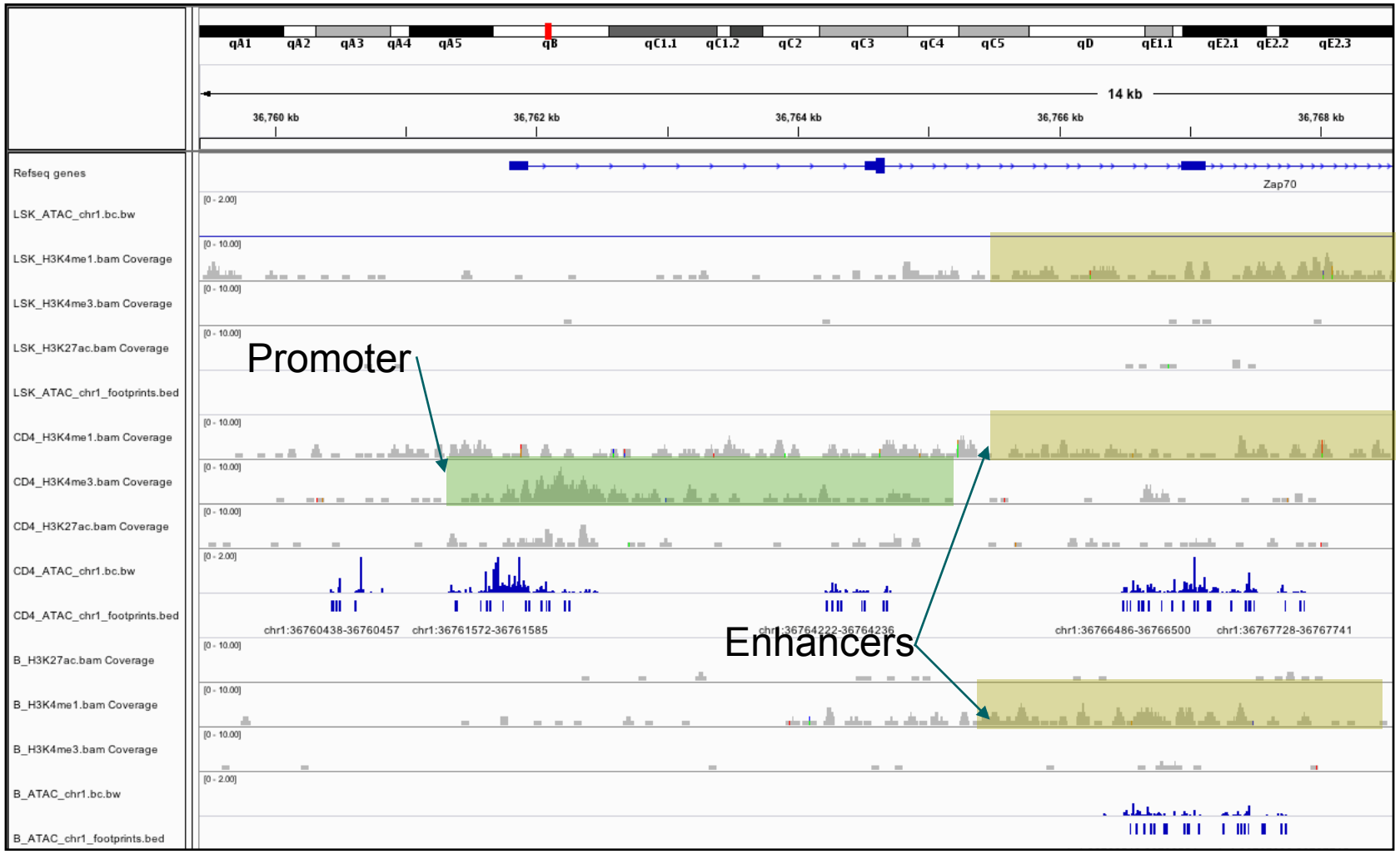
Tutorial / IGV Screenshots

ATAC-seq, Histones and Footprints around Zap70 locus



Tutorial / IGV Screenshots

ATAC-seq, Histones and Footprints around Zap70 locus

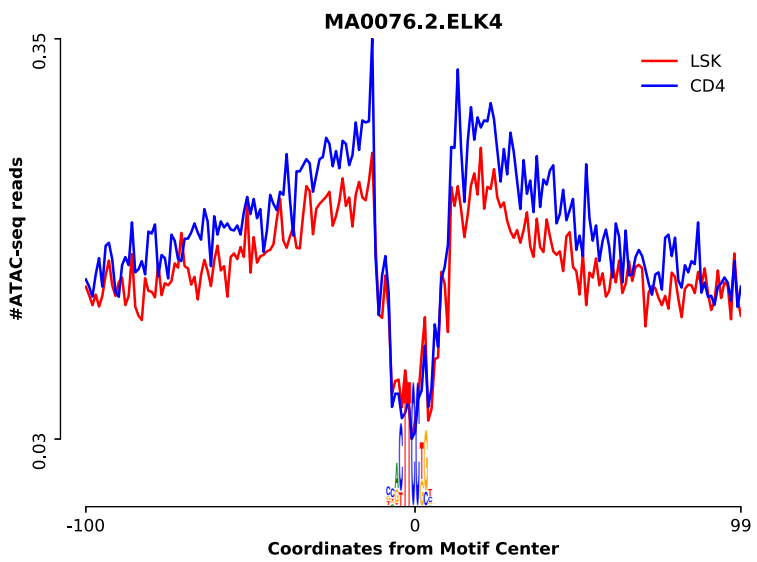


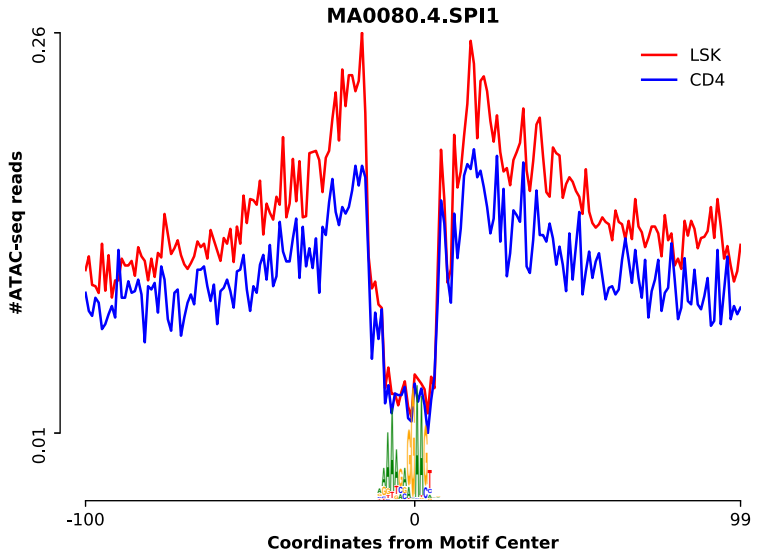
Tutorial / IGV Screenshots

Motif Matching of Pu.1 and Elk4

Tutorial / IGV Screenshots

ATAC-seq profiles around Pu.1(Spi1) and Elk4





www.costalab.org

IZKF Interdisziplinäres Zentrum für Klinische Forschung

RWTH AACHEN UNIVERSITY