Ch. 2

Establishment & maintenance of chromatin features

Ch 2 - L 1.1

- PTMs and other chromatin features distribution throughout the genome

Post-genomics

Genetics

Comparative (phylogenetic conservation indicates conserved function)

Human Genetic Variation (1000 Human Genomes - HapMap)

GWAS – Genome variations – phenotype correlation

Gene expression and phenotype

Functional Genomics

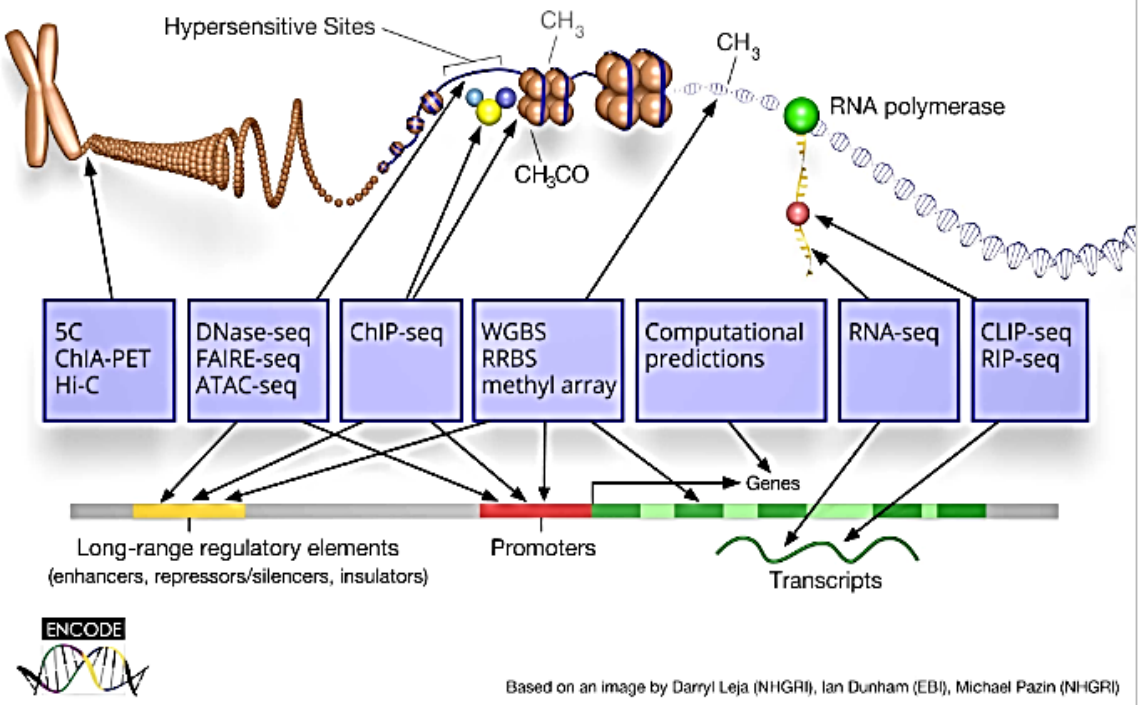
Epigenomics:	CpG methylation
	Histone modifications (PTMs)
	Chromatin status
	Protein-DNA mapping (e.g. Transcription Factors)
Transcriptomics:	Coding and noncoding RNAs

Sicuro | https://www.encodeproject.org

Home - PubMed - N | Università di Torino | Home Page - MedInT | Course: Advanced M | Innovative Training N | Dictionary | Definitio | Didattica | bioinformatics

ENCODE Data Encyclopedia Materials & Methods Help Search...

ENCODE: Encyclopedia of DNA Elements



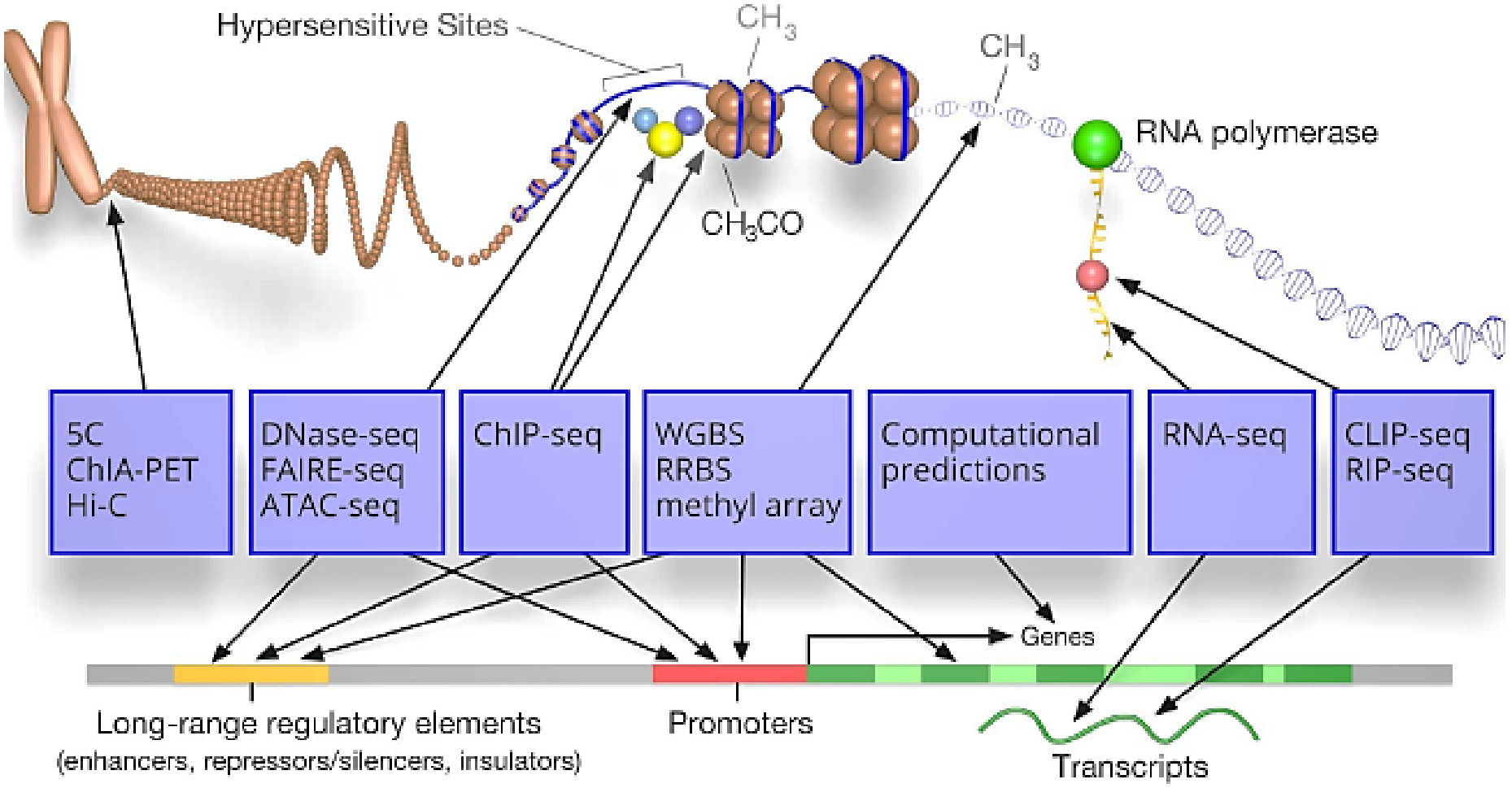
The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

Get Started

HUMAN MOUSE WORM FLY

[The ENCODE Project Website]

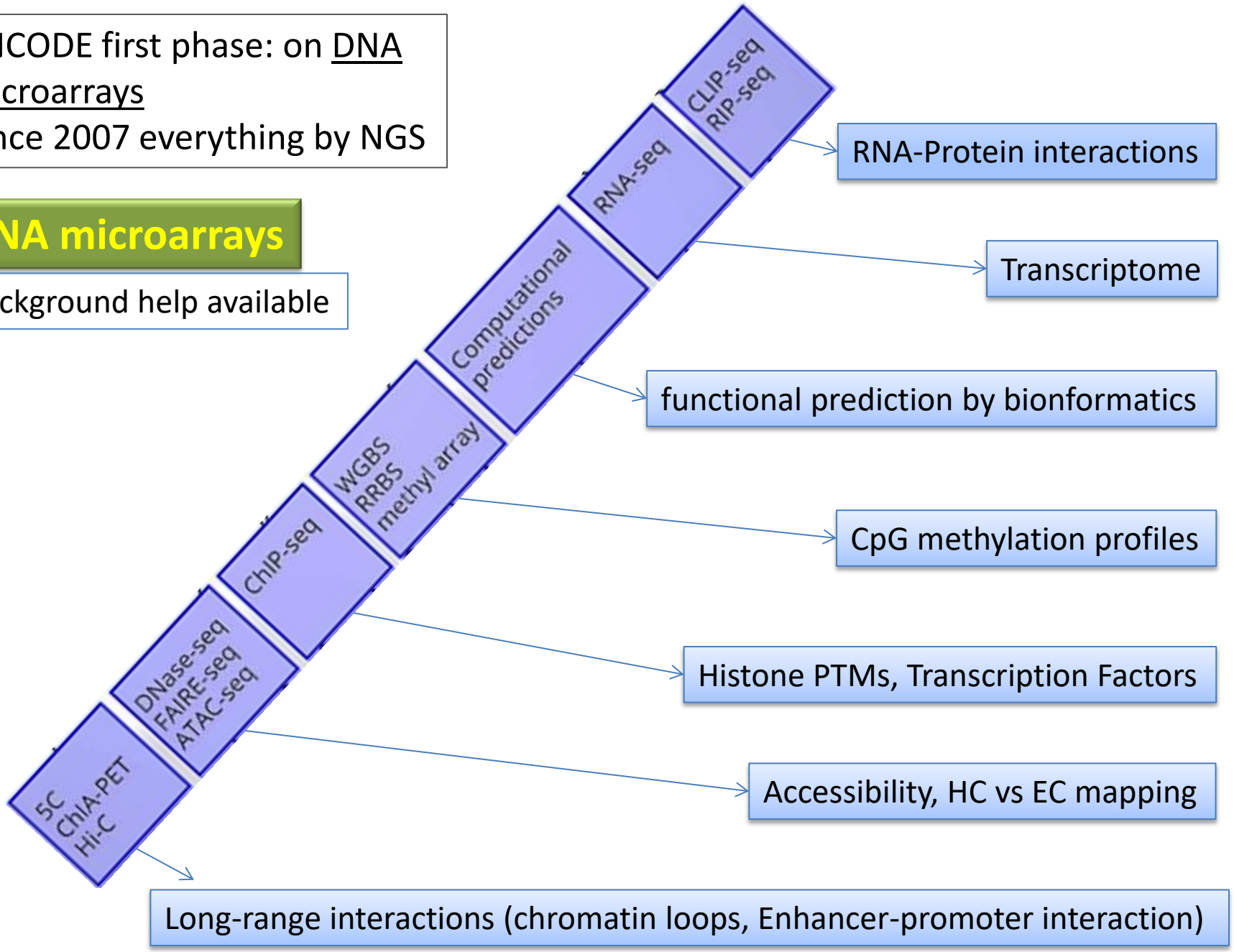
ENCODE: Encyclopedia of DNA Elements



ENCODE first phase: on DNA microarrays
Since 2007 everything by NGS

DNA microarrays

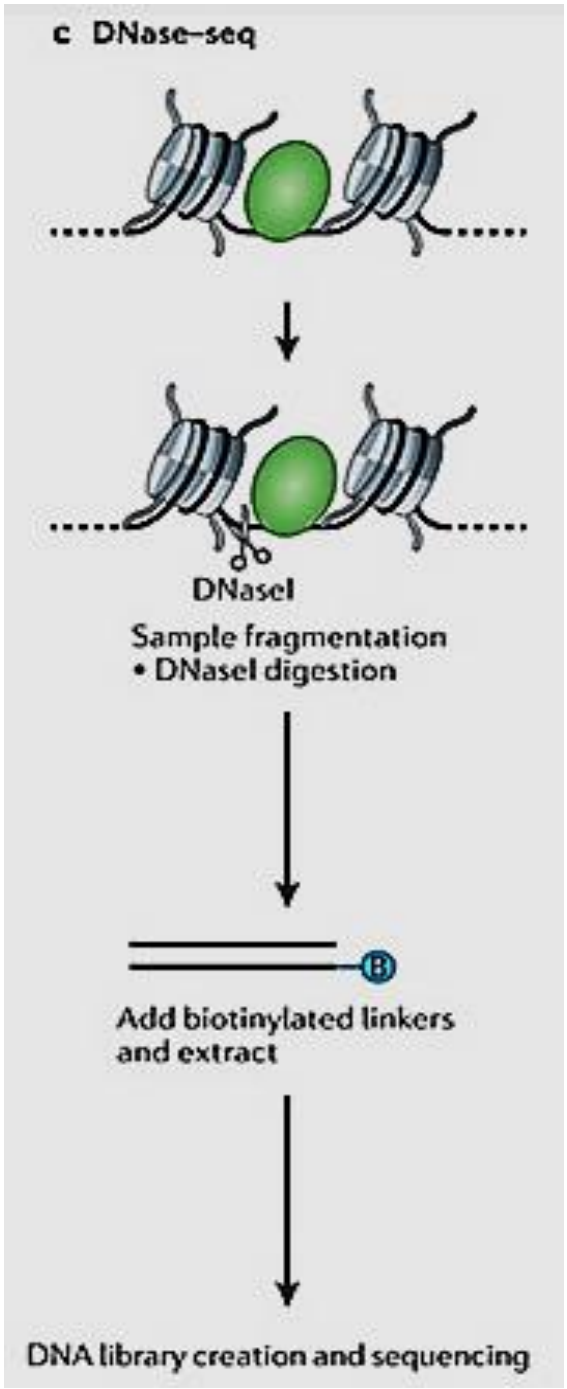
Background help available



Can we access information on chromatin **accessibility** **genome-wide** ?

Dnase-Seq
FAIRE-Seq (Formaldehyde Assisted Isolation of Regulatory Elements)
ATAC-Seq (Assay for Transposase-Accessible Chromatin)

All these methods are coupled to NGS, but they have been (and still are) also analyzed using microarrays

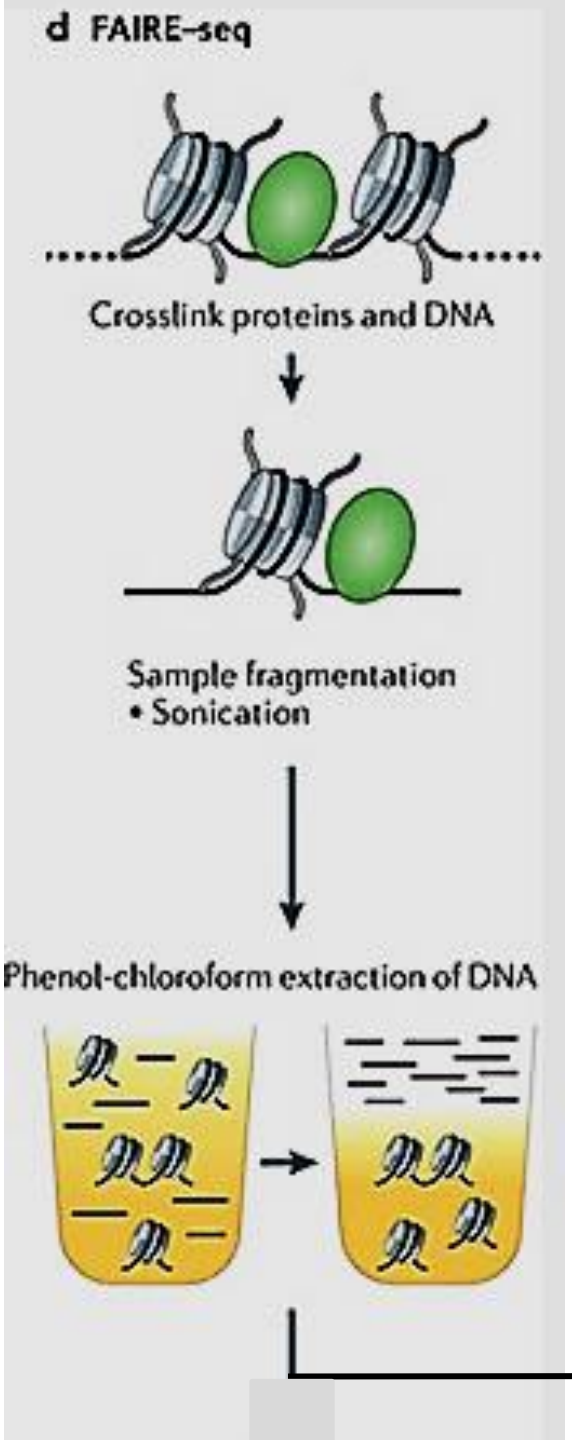


DNase I sensitivity assay

→ **DNase-Seq**

DNase–seq experiments combine traditional DHS assays with high-throughput sequencing to simultaneously identify all types of regulatory regions genome-wide.
The 5′ end of a sequence tag generated by DNase–seq indicates the site of a DNaseI digestion event, and regions of enrichment in digestion events are identified as DHS sites, each of which can contain binding sites of multiple factors.

Furey 2012, Nat Rev Genet. 13: 840-852

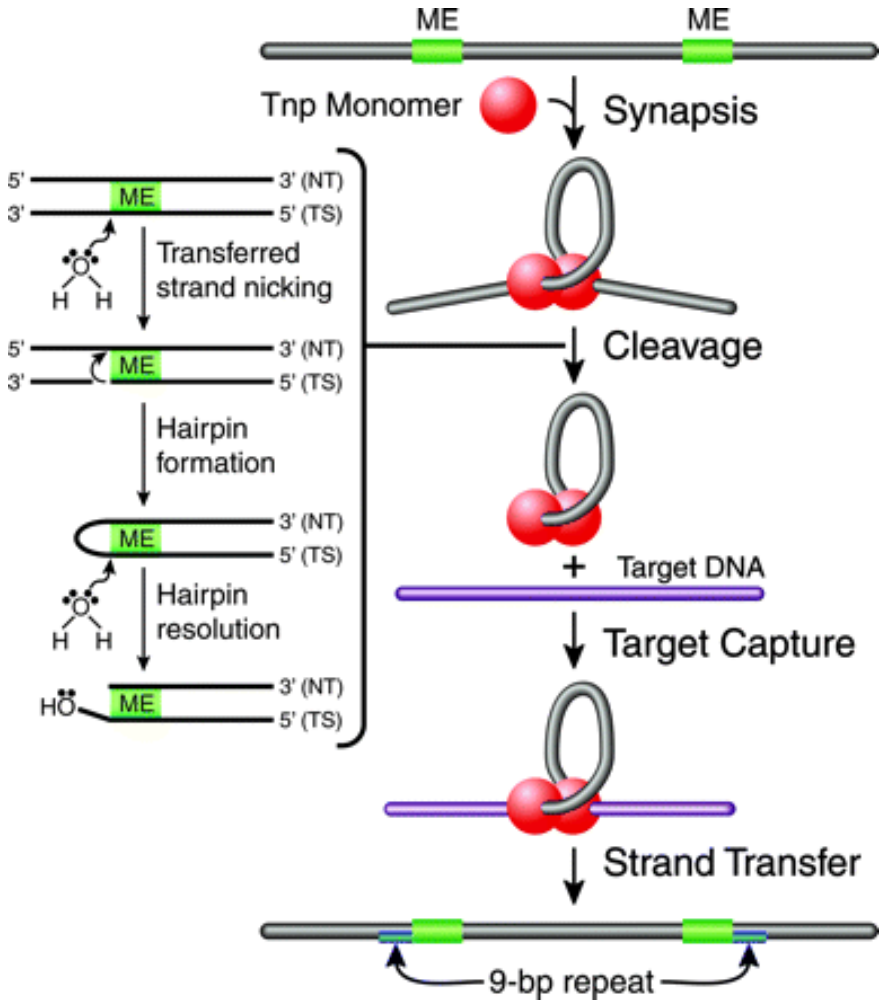


FAIRE–seq

The FAIRE–seq assay starts with formaldehyde crosslinking, similarly to ChIP, but then instead of using an antibody to target specific factors, DNA is sonicated and the extract is subjected to phenol-chloroform extraction. The nucleosome-depleted fraction of DNA is preferentially segregated to the aqueous phase. FAIRE-enriched DNA has been shown to correspond to regulatory regions.

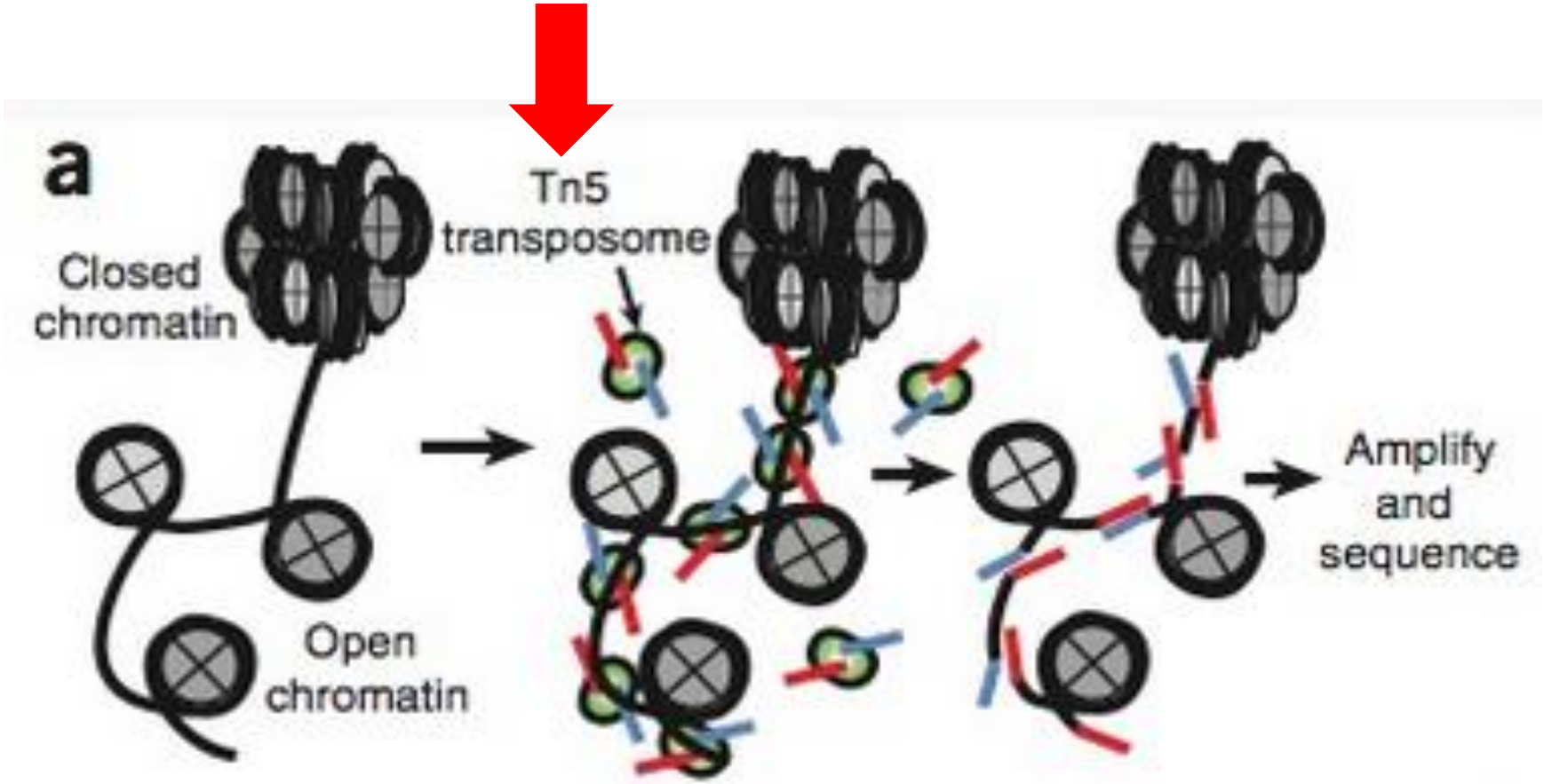
Furey 2012, Nat Rev Genet. 13: 840-852

DNA library creation and sequencing



How Tn5 transposase works

from Gradman et al. (2008) J. Bacteriol. 190: 1484-87. DOI: 10.1128/JB.01488-07



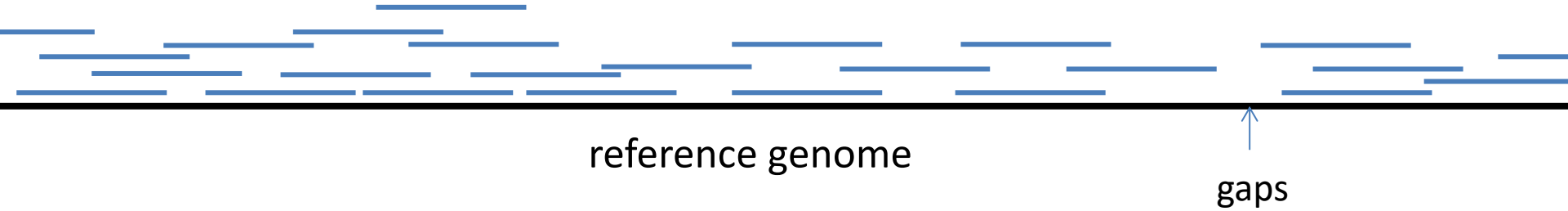
(a) ATAC-seq reaction schematic.
Transposase (green), loaded with sequencing adaptors (red and blue), inserts only in regions of open chromatin (between nucleosomes in gray) and generates sequencing-library fragments that can be PCR-amplified.

Buenrostro et al., Nat Methods. 2013, 10(12):1213-8.

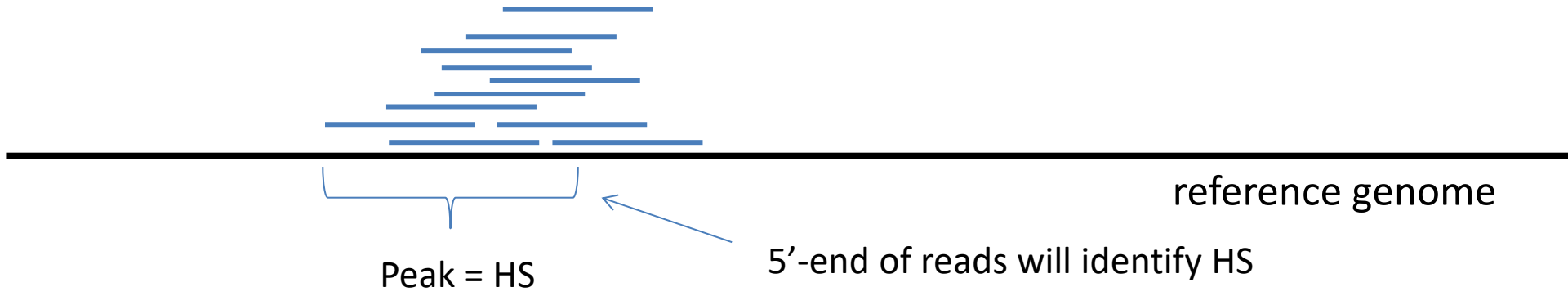
Transposase complex loaded
with sequencing linkers

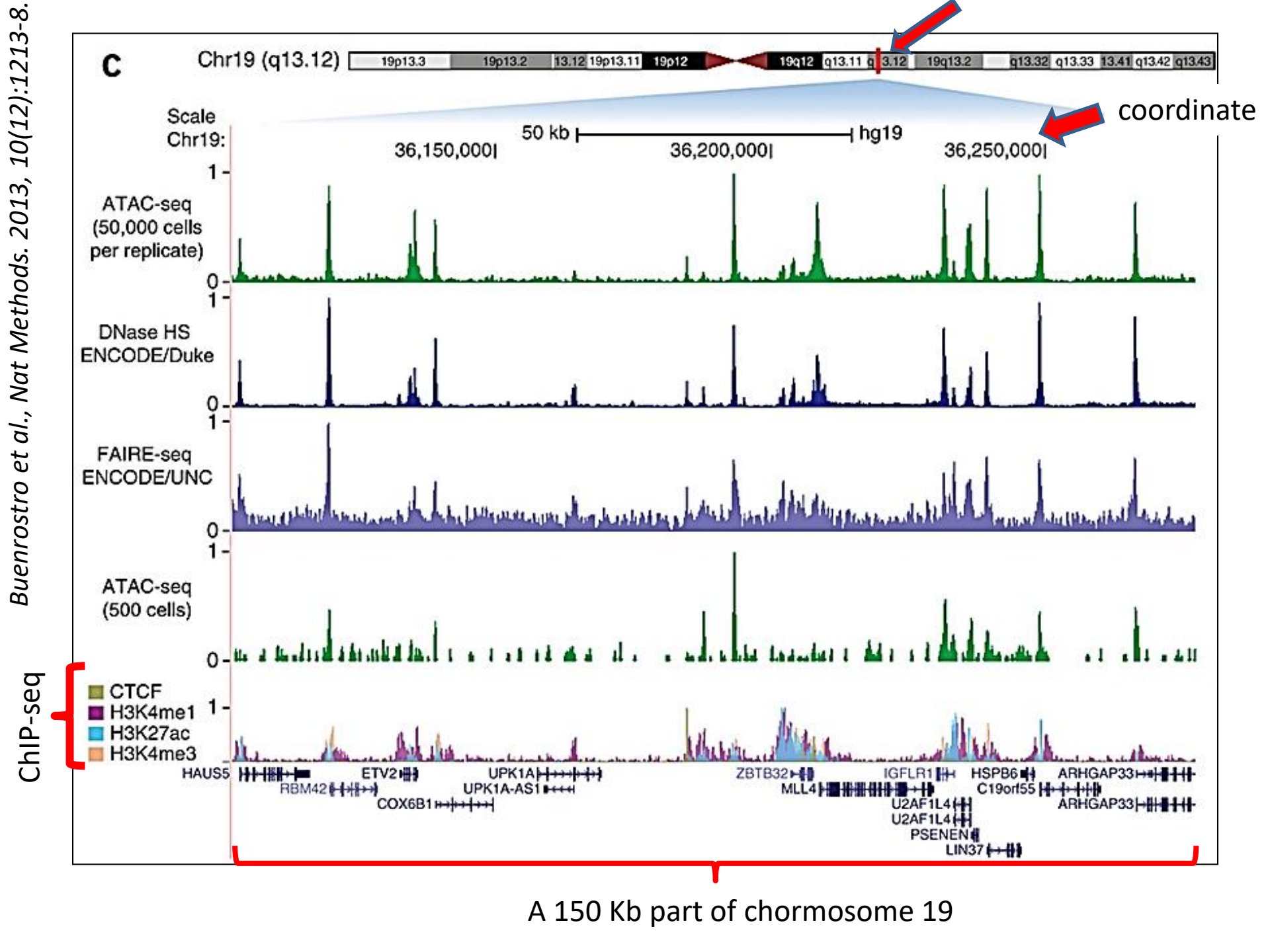
Mapping data.

Results are very different from those obtained from NGS on genomic DNA, which are a «probabilistic» distribution of fragments on the reference genome.



ATAC-Seq will give only fragments from open regions




Buenrostro et al., Nat Methods. 2013, 10(12):1213-8.
c
Chr19 (q13.12)
coordinate
Scale
Chr19:
50 kb
hg19
36,150,000
36,200,000
36,250,000
ATAC-seq
(50,000 cells
per replicate)
DNase HS
ENCODE/Duke
FAIRE-seq
ENCODE/UNC
ATAC-seq
(500 cells)
ChIP-seq
CTCF
H3K4me1
H3K27ac
H3K4me3
HAUS5
ETV2
RBM42
COX6B1
UPK1A
UPK1A-AS1
ZBTB32
MLL4
IGFLR1
HSPB6
C19orf55
ARHGAP33
U2AF1L4
U2AF1L4
PSENEN
LIN37
A 150 Kb part of chormosome 19

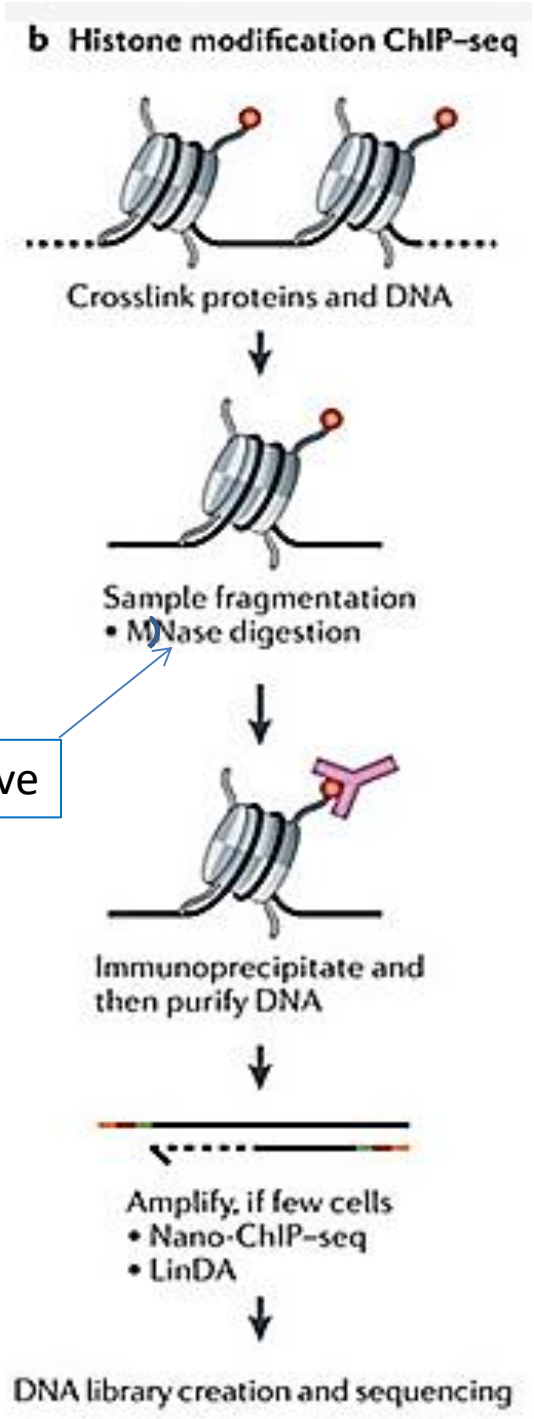
Histone PTMs were studied genome-wide using ChIP-Seq or **ChIP-on-chip**

What is this?

Very **specific antibodies** are available, which not only recognize subtle differences (e.g. symmetric *versus* nonsymmetric arginine dimethylation) but also the peptide context in with PTM occurs (i.e. the exact aminoacid that is modified), but

facultative

caution!
not all the available antibodies are good for ChIP. Often two or three different antibodies are used together.



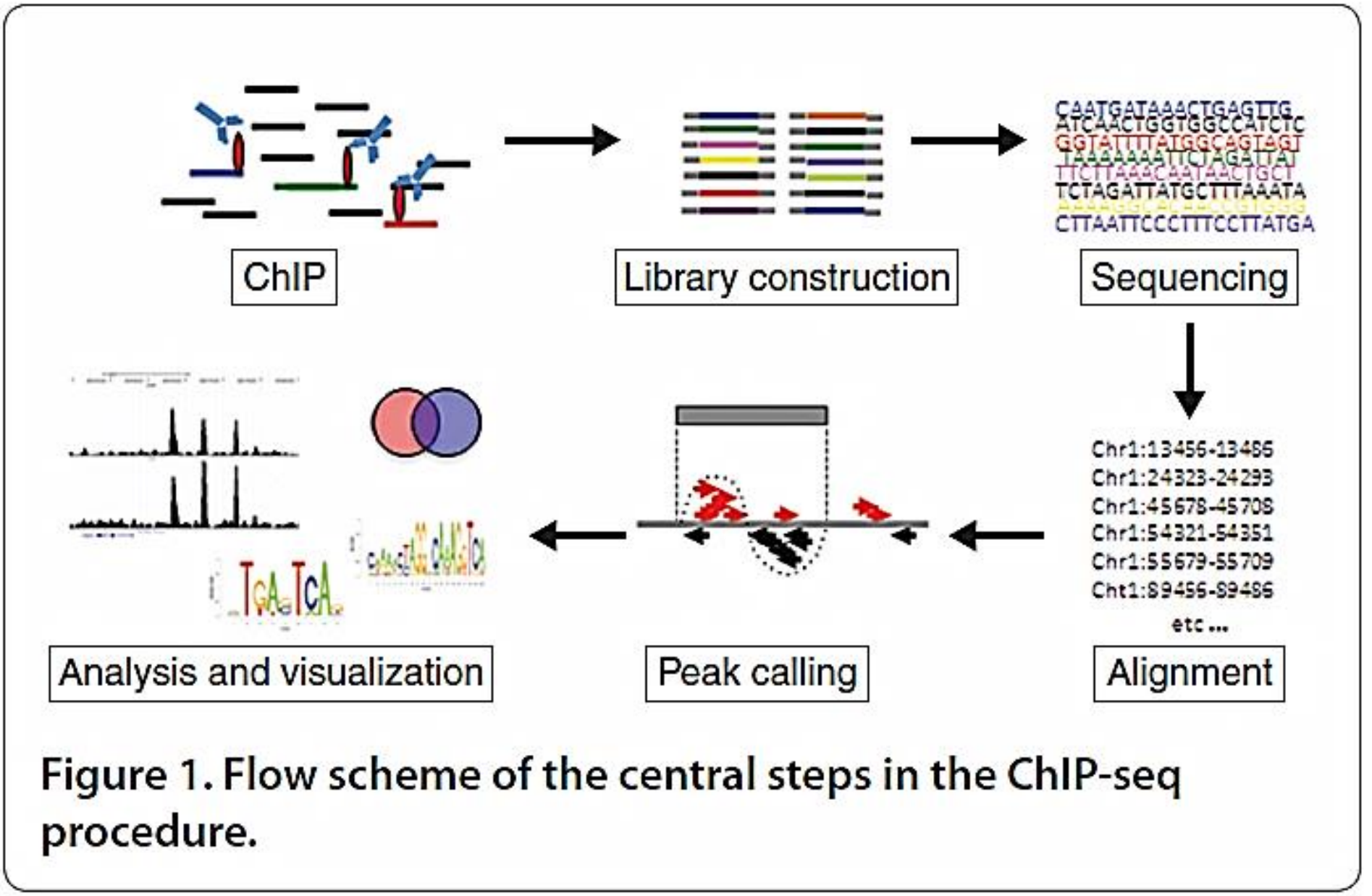


Figure 1. Flow scheme of the central steps in the ChIP-seq procedure.

Liu *et al*. *BMC Biology* 2010, **8**:56

Let us take a look of one of the first papers published, reporting PTMs analysis genome-wide using NGS (Illumina-Solexa first version)

Resource

Cell

High-Resolution Profiling of Histone Methylations in the Human Genome

Artem Barski,[1,3] Suresh Cuddapah,[1,3] Kairong Cui,[1,3] Tae-Young Roh,[1,3] Dustin E. Schones,[1,3] Zhibin Wang,[1,3] Gang Wei,[1,3] Iouri Chepelev,[2] and Keji Zhao[1,*]

[1] Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA

[2] Department of Human Genetics, Gonda Neuroscience and Genetics Research Center, University of California, Los Angeles, Los Angeles, CA 90095, USA

[3] These authors contributed equally to this work and are listed alphabetically.

*Correspondence: zhaok@nhlbi.nih.gov

DOI 10.1016/j.cell.2007.05.009

Analysis done using primary human CD4+ T cells

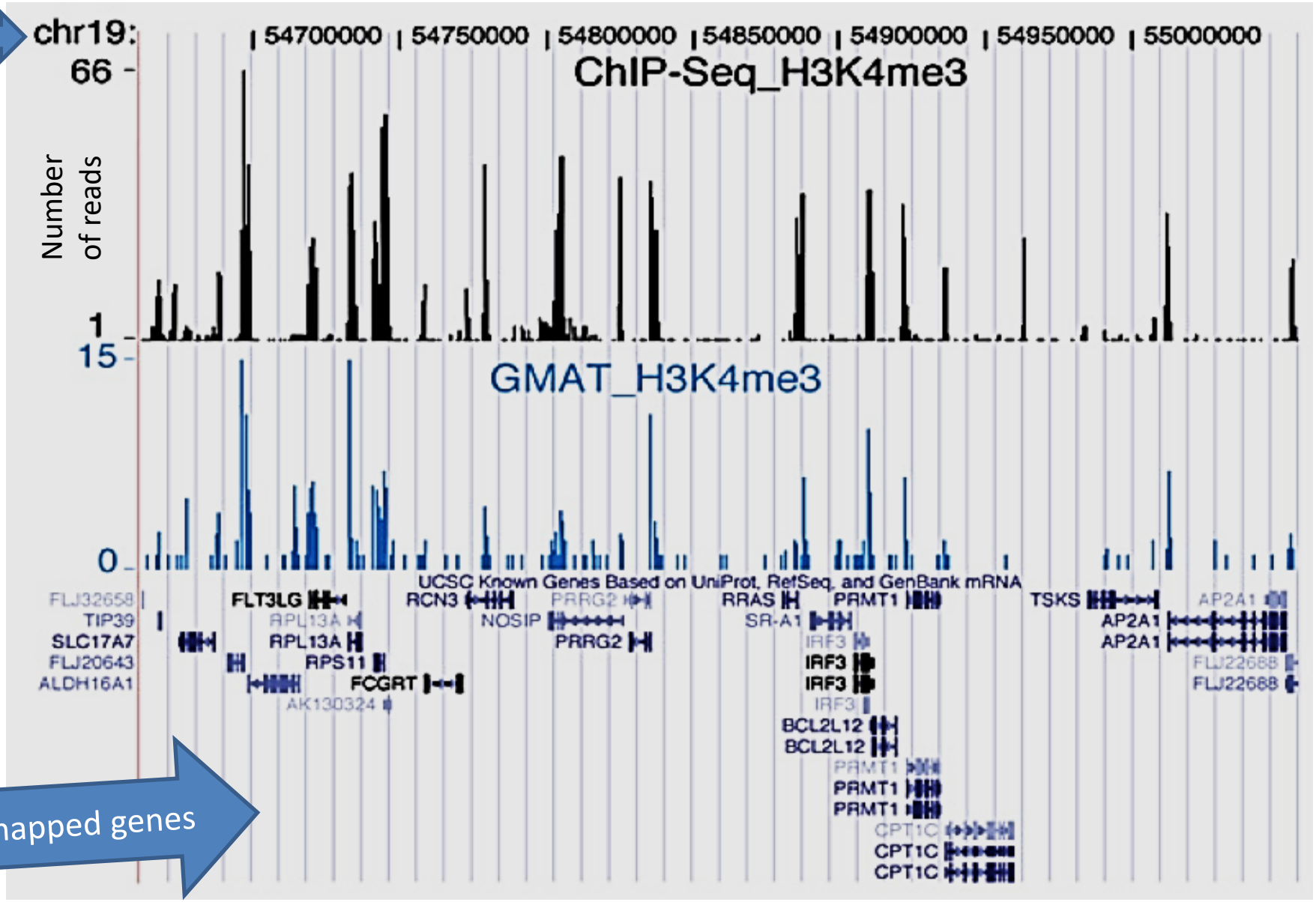
SUMMARY

Histone modifications are implicated in influencing gene expression. We have generated high-resolution maps for the genome-wide distribution of 20 histone lysine and arginine methylations as well as histone variant H2A.Z, RNA polymerase II, and the insulator binding protein CTCF across the human genome using the Solexa 1G sequencing technology. Typical patterns of histone methylations exhibited at promoters, insulators, enhancers, and transcribed regions are identified. The monomethylations of H3K27, H3K9, H4K20, H3K79, and H2BK5 are all linked to gene activation, whereas trimethylations of H3K27, H3K9, and H3K79 are linked to repression. H2A.Z associates with functional regulatory elements, and CTCF marks boundaries of histone methylation domains. Chromosome banding patterns are correlated with unique patterns of histone modifications. Chromosome breakpoints detected in T cell cancers frequently reside in chromatin regions associated with H3K4 methylations. Our data provide new insights into the function of histone methylation and chromatin organization in genome function.

Cell 129, 823–837 (2007)

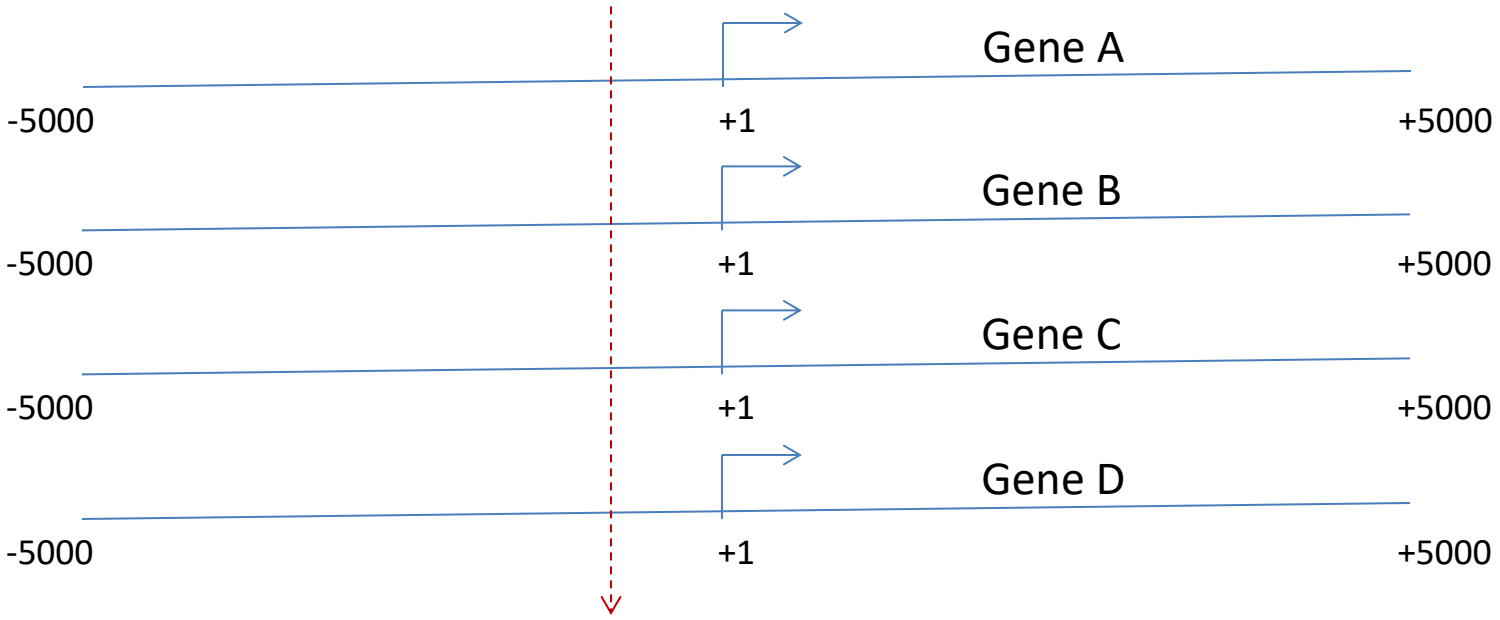
:

Snapshot of the genome browser Chr. 19: 400 Kb window



mapped genes

Data are then collected from many similar position. In the figure, all RefSeq genes are taken into account and frequencies aligned using the TSS as central point, in a window of 5Kbp.
Then, the frequency of counts for each position is calculated over the entire set of genes and plotted.



F(n) (the frequence of each nucleotide call)

Data were divided in 4 groups of genes characterized by different levels of expression in CD4+ T cells

High – Medium – Low - Silent

(based on Affimetrix microarray analysis)

Frequency representation

ChIP-Seq using anti-RNA Polymerase II (Pol II) antibody

ChIP-Seq using anti-H3 dimetyl-lysine 4 antibody

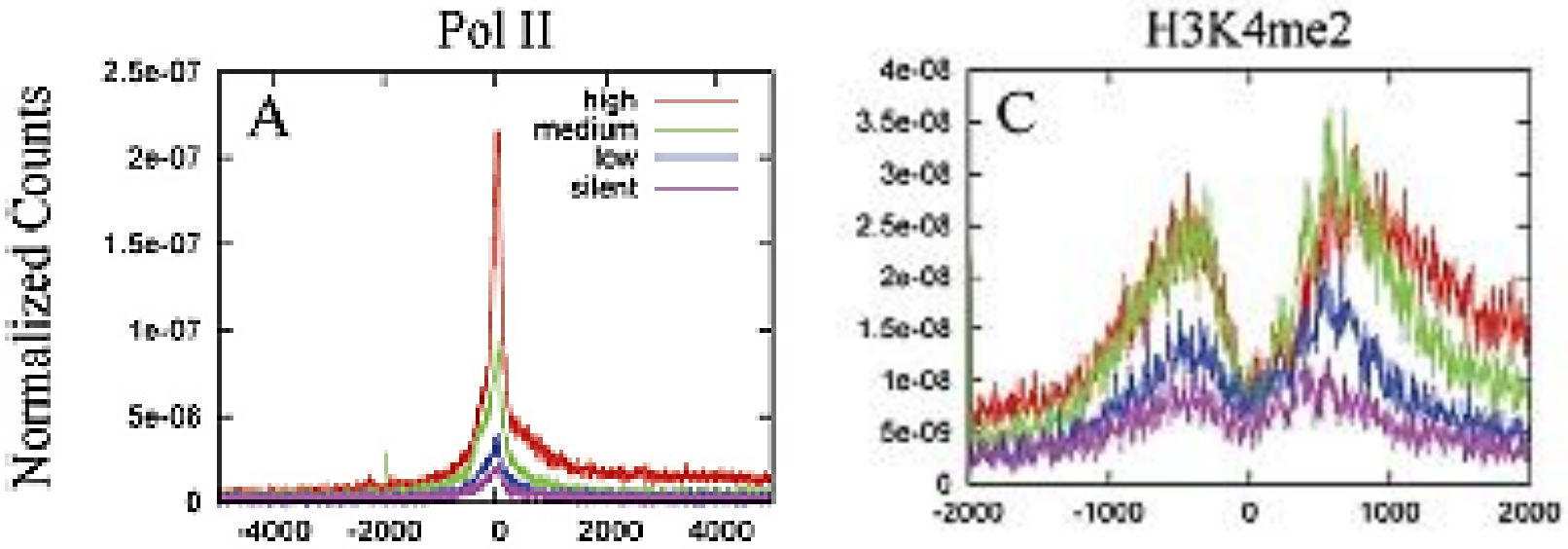
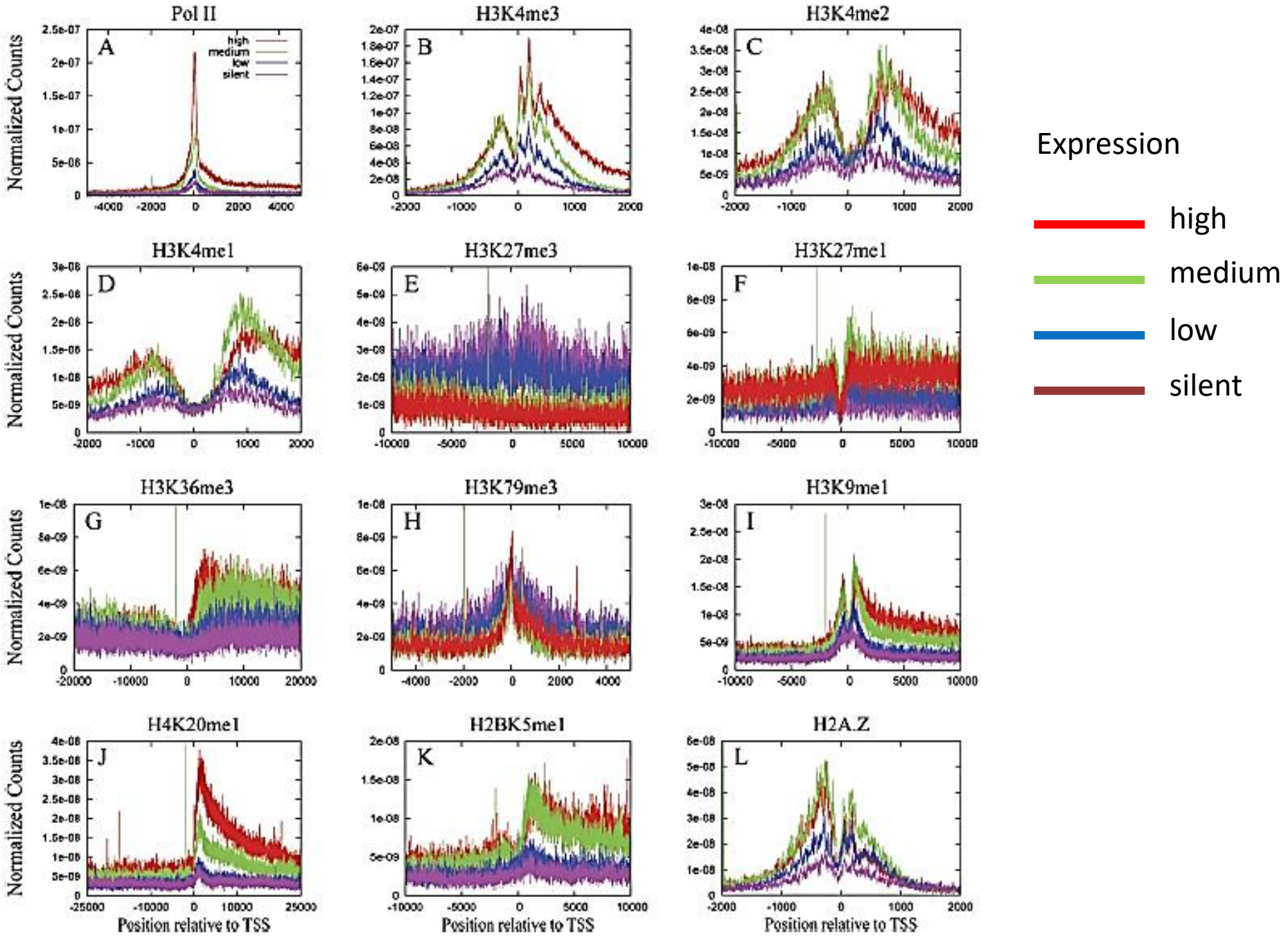
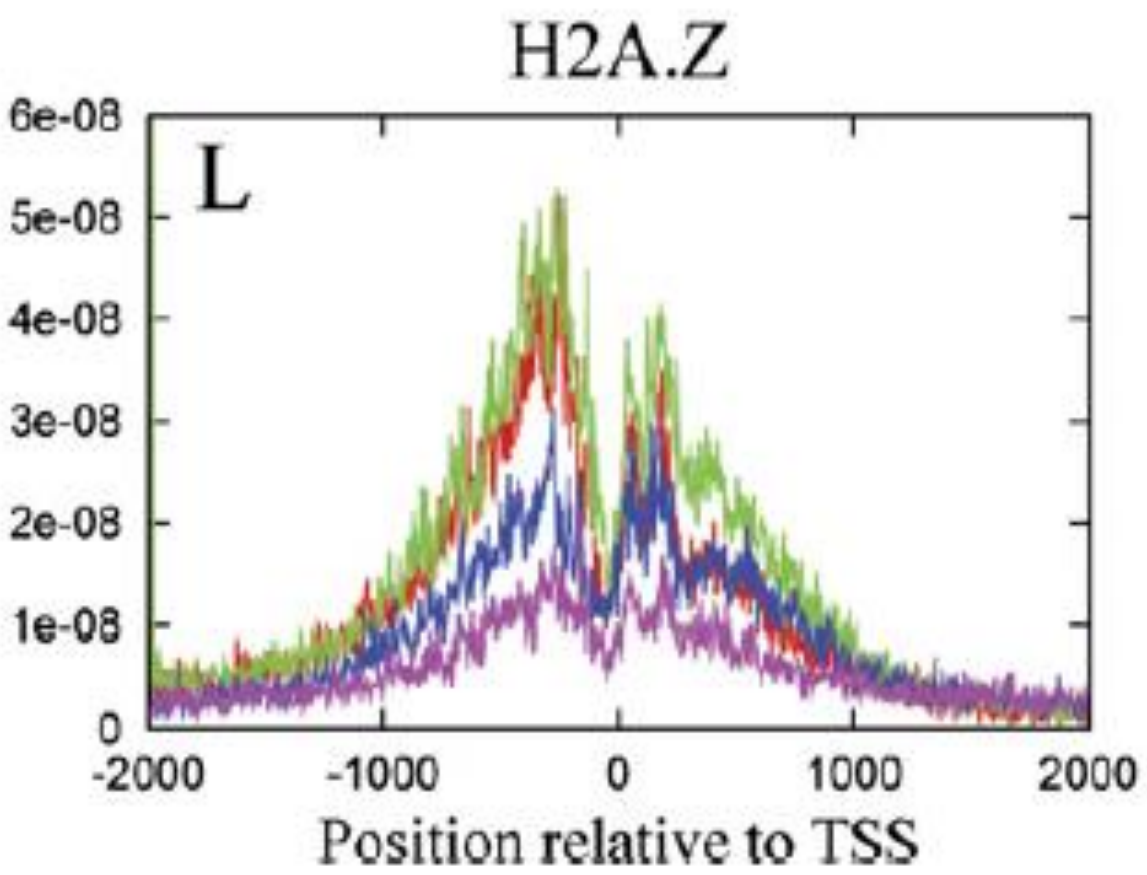


Figure 2. Histone Methylation near Transcription Start Sites
(A)–(L) Profiles of the histone methylation indicated above each panel across the TSS for highly active, two stages of intermediately active and silent genes are shown. Twelve thousand human genes were separated into twelve groups of one thousand genes according to their **expression level** (as measured on Affymetrix microarrays): **High – Medium – Low – Silent**


Pol II
H3K4me3
H3K4me2
H3K4me1
H3K27me3
H3K27me1
H3K36me3
H3K79me3
H3K9me1
H4K20me1
H2BK5me1
H2A.Z
A
B
C
D
E
F
G
H
I
J
K
L
high
medium
low
silent
Normalized Counts
Position relative to TSS
Expression
high
medium
low
silent



Expression

high

medium

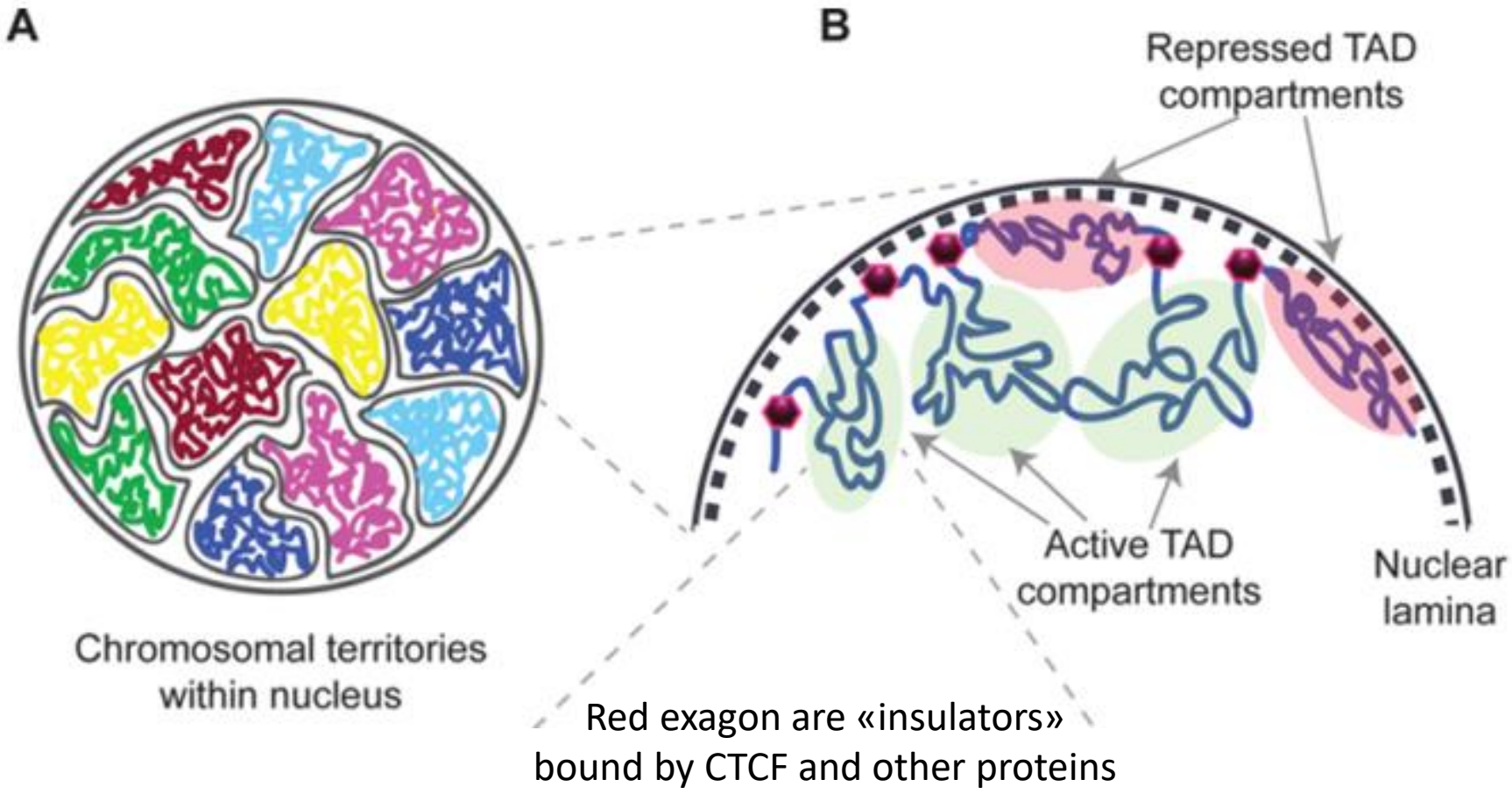
low

silent

Histone PTMs Distribution:

« gross distribution » - quite large «domains» that are separated from one another by «insulator» sequences (bound by proteins like CTCF)

« focused distribution » - specific to small functional parts, e.g. Promoters, Enhancers, Centromeres /Telomeres.. etc



Red exagon are «insulators» bound by CTCF and other proteins

from Matharu, et al. PLOS Genetics 11 (12): e1005640 (2015). DOI:10.1371/journal.pgen.1005640. PMID 26632825.

Features: Histone PTMs - CpG methylation

widely «repressive»

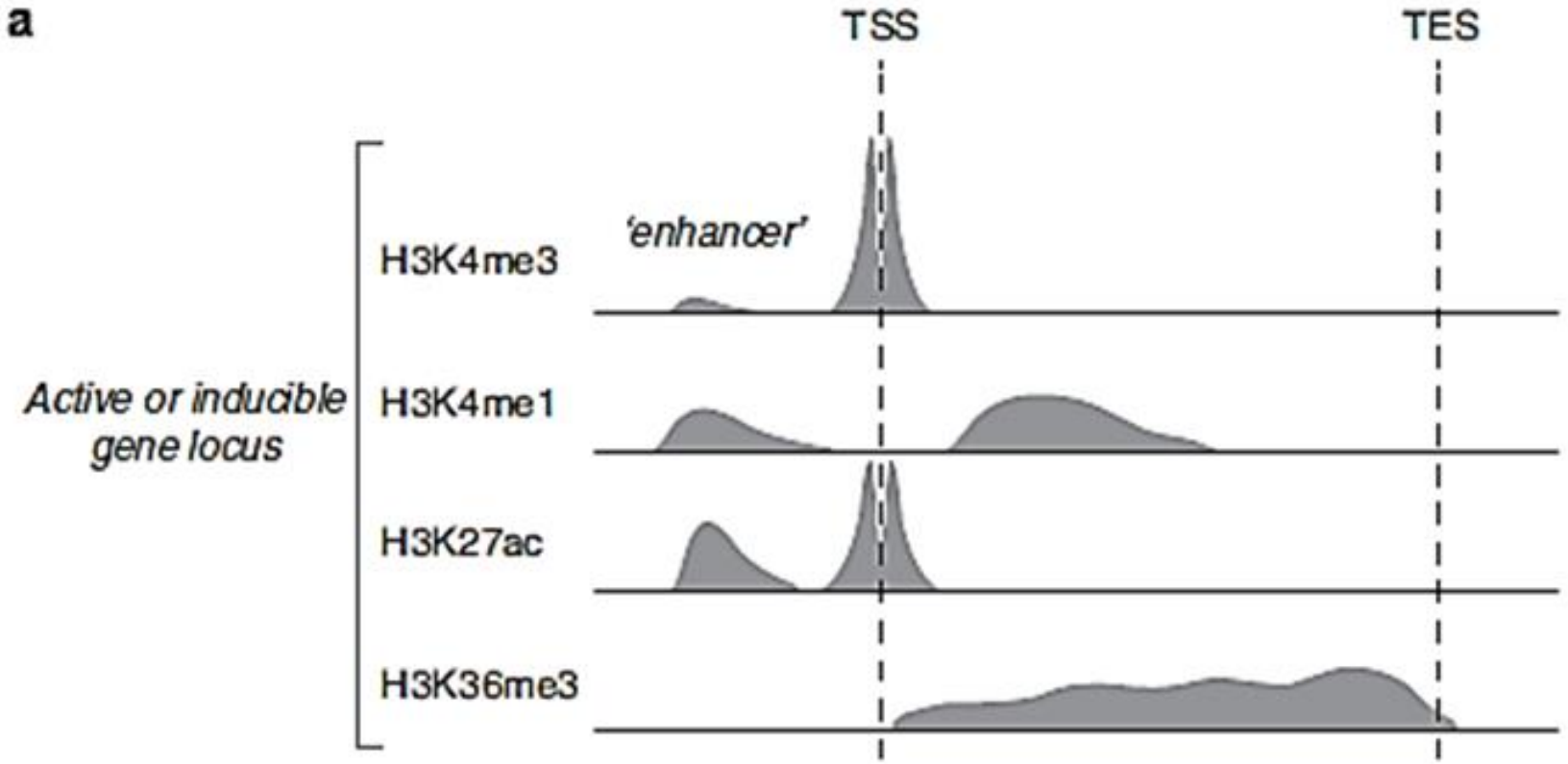
H3K9me3, H3K27me3, CpG island hypermethylation

widely «active»

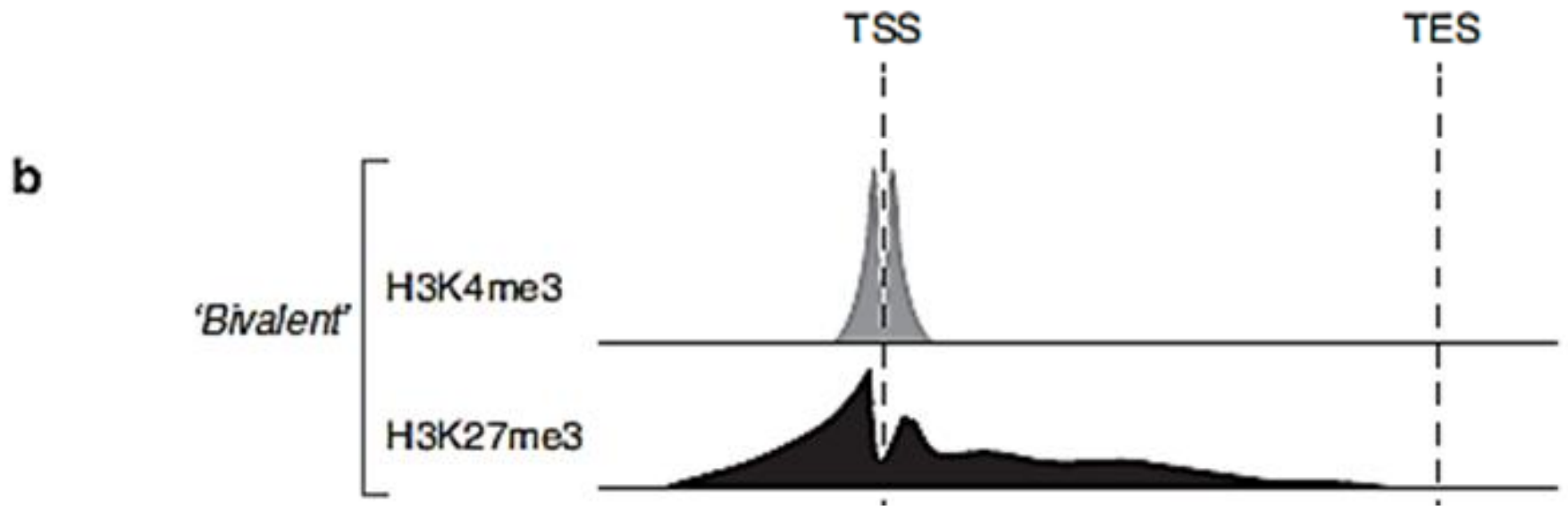
H3K4me(1,2,3), H3K27ac, H3K36me3,
H-hyperacetylation

bivalent domains (developmental)

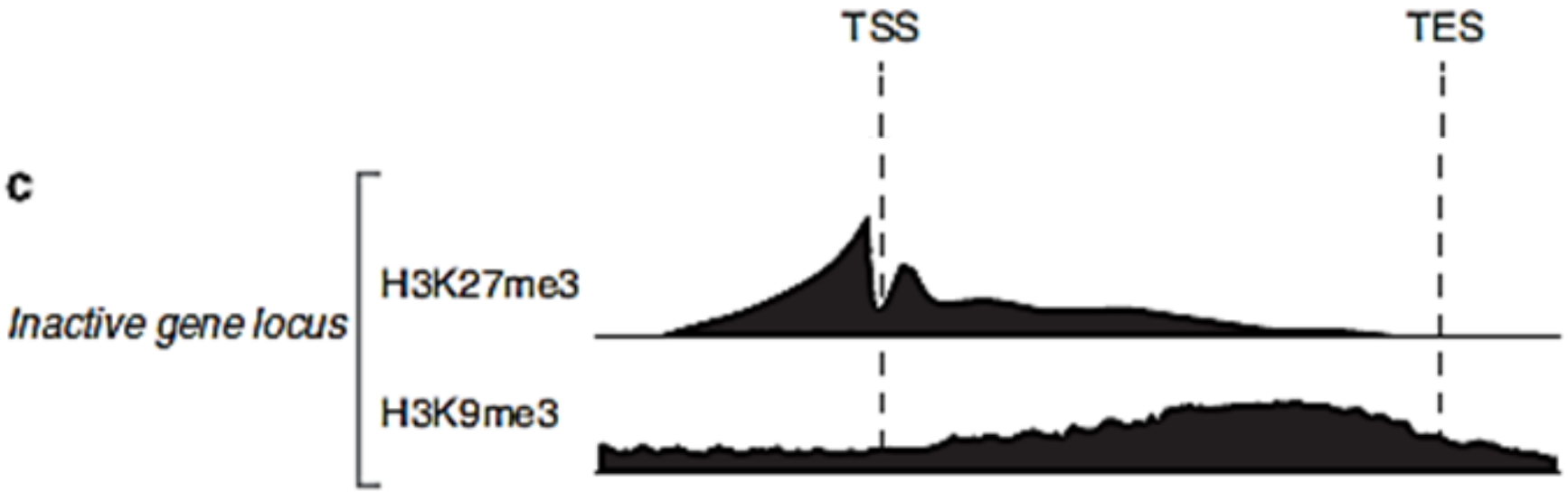
H3K4me(1,2,3), H3K27me3



Kimura et al, J Human Genet (2013) 58:439-445



Kimura et al, J Human Genet (2013) 58:439-445



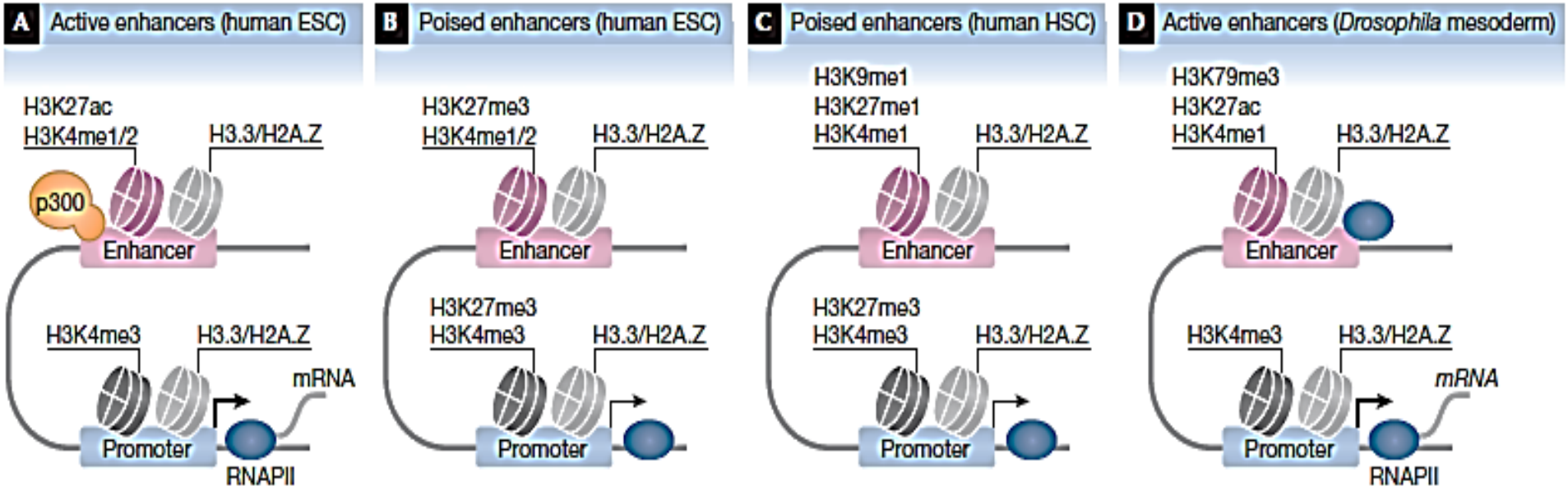
Kimura et al, J Human Genet (2013) 58:439-445

other specific points of PTMs are ENHANCERS

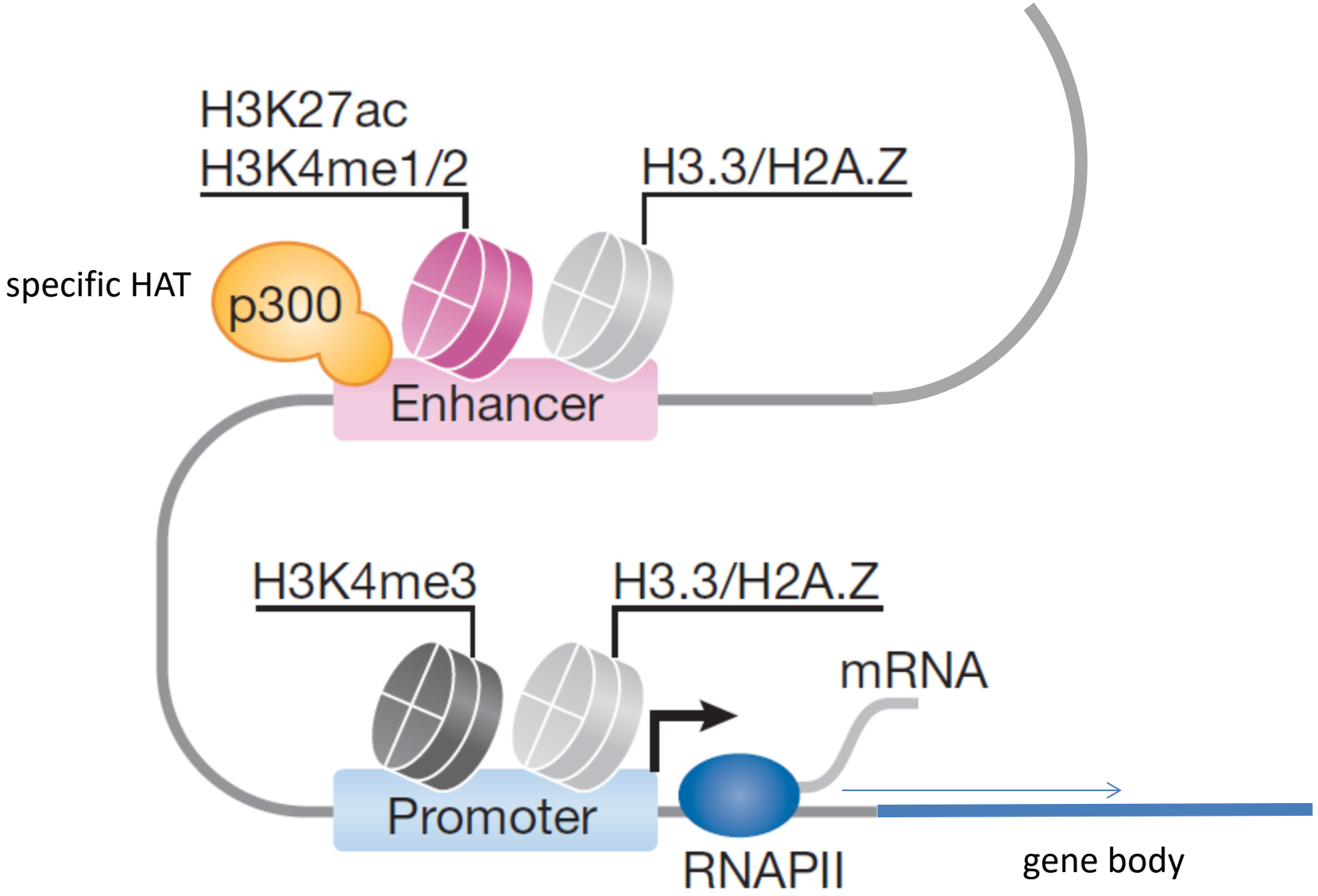
Most of eukaryotic regulation depends on gene-distal regulatory modules called Enhancers (that outnumber promoters by a factor of ten)

Enhancers posses multiple sequence elements recognized by Transcription Factors.

Active Enhancers, as well as Enhancers marked for activation, show specific PTMs at the nucleosomes flanking their sequence.

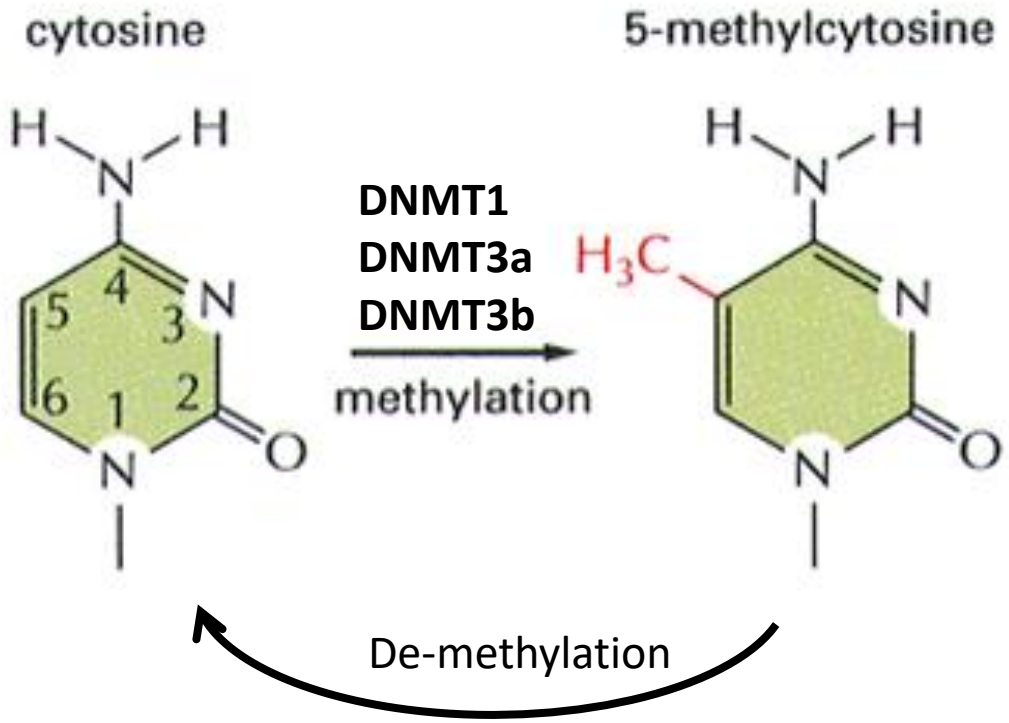


from Ong & Corces, 2012



modified from Ong & Corces, 2012

What about CpG methylation ?



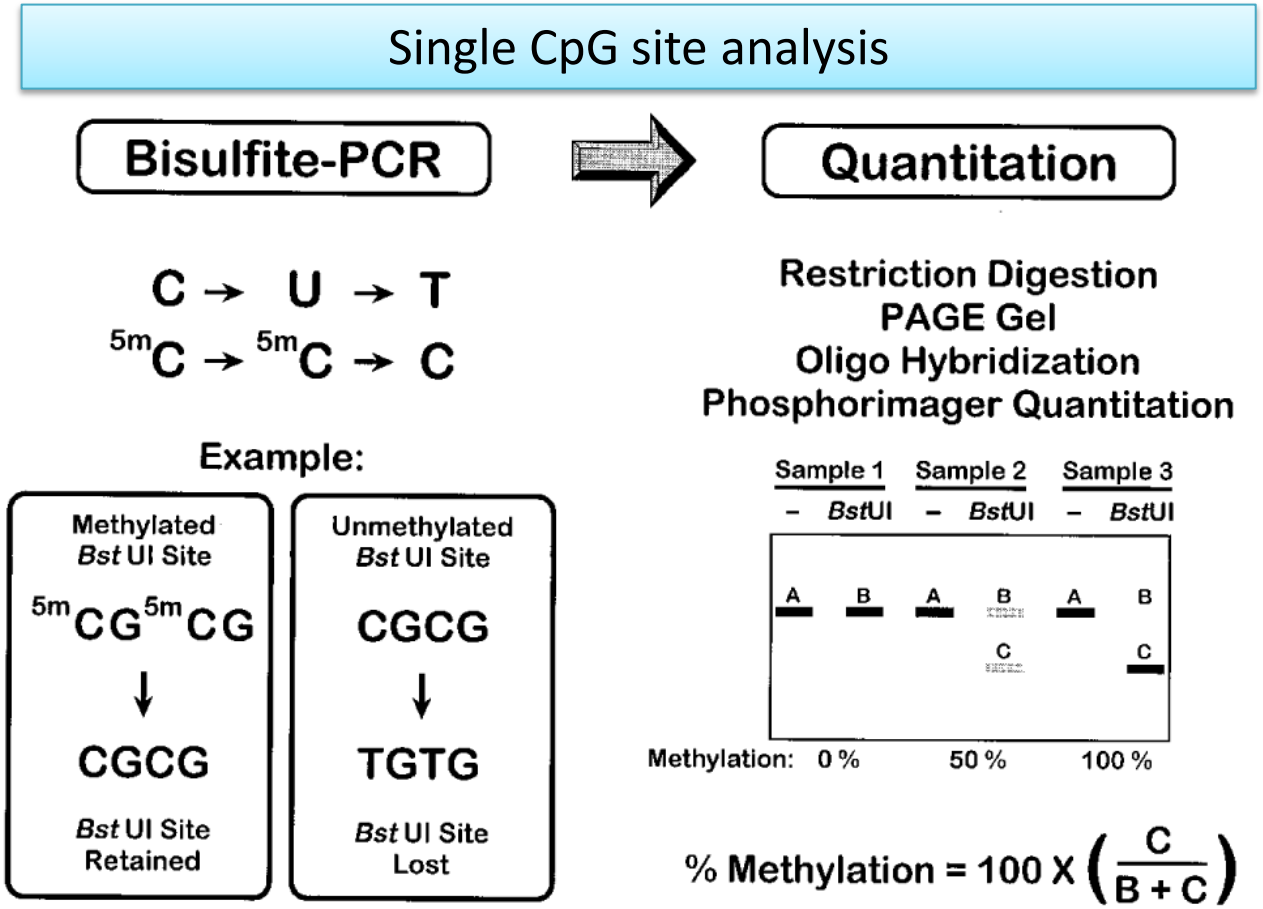
How to measure CpG methylation

At single locus level:

1. Methylation-sensitive restriction enzymes
    1. HpaII*/MspI (CCGG)
    2. SmaI*/XmaI (CCCGGG)
    3. McrBC – recognize 2 methylated (G/A)pC (50-1000bp apart)

2. Bisulfite conversion (Methyl-C is not modified , C is converted to U)
    1. Enzymatic analysis (e.g. Cobra-next slide)
    2. Cloning alleles and Sanger sequencing

Most common methods to detect CpC methylation are based on **bisulfite**, which converts C (but not methyl-C) to Uridine. Uridine in DNA is then replaced by T in the following PCR. Conservation of C (in the case of methyl-C) or change in T are subsequently detected (simplest) by restriction site analysis, as in the **example** that follows:



Is it possible to study DNA CpG methylation genome-wide ?

Approach I

1) DNA immunoprecipitation using an Ab against 5-methyl-cytosine

2) DNA immunoprecipitation using tagged-MBD (methyl-DNA binding)

followed by hybridization to microarrays or NGS

Approach II

Bisulfite conversion of the whole genomic DNA,

followed by NGS (WGBS)

On microarrays...

Anti-5—methylcytosine capture

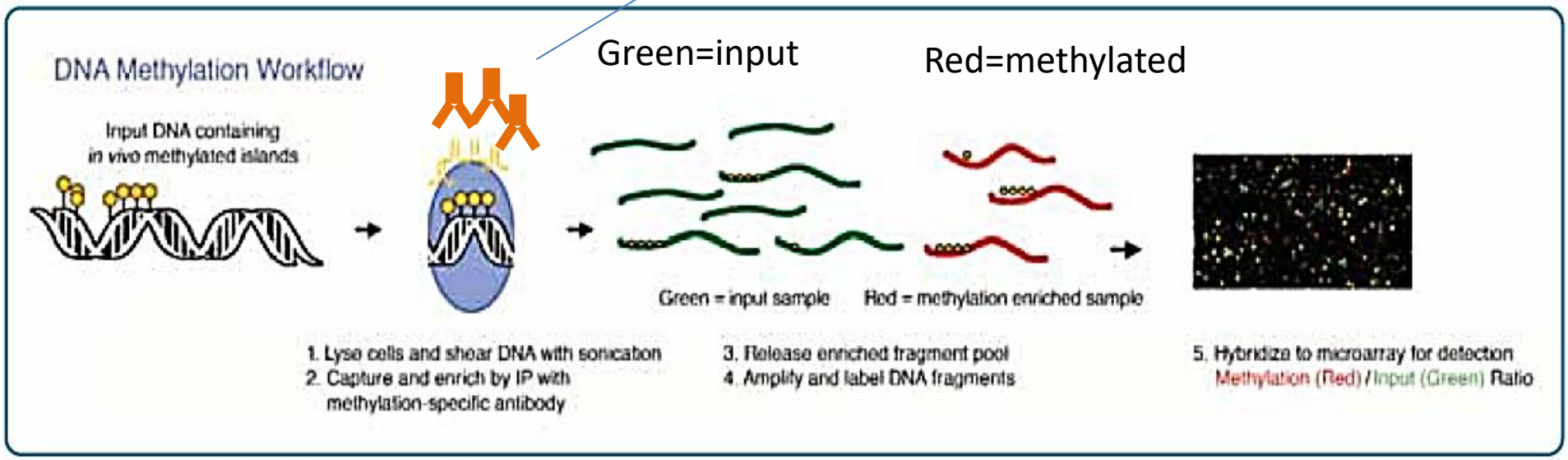
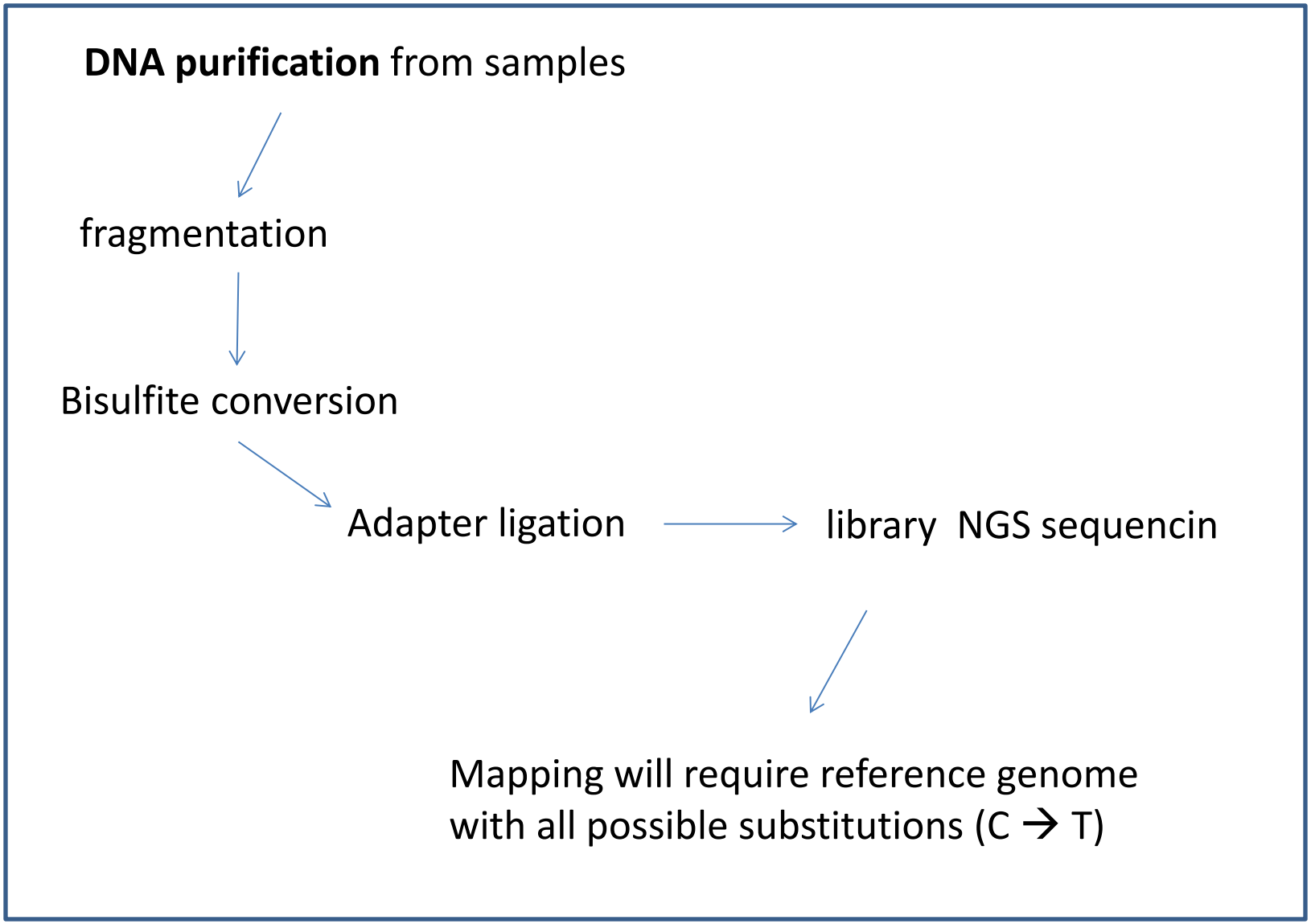


Figure 1. Affinity-based isolation of methylated DNA. A representative example of one method for isolation and enrichment of methylated DNA. Methylated regions of the genome (mDIP) from a genomic DNA sample are isolated with a monoclonal antibody to 5-methylcytosine. Isolated DNA is purified, Cyanine 5-labeled, and competitively hybridized against similarly Cyanine 3-labeled "input" genomic DNA onto a single microarray. Arrays are washed, scanned, and analyzed with Feature Extraction software. Relative DNA methylation levels for each probe/CpG Island are reflected in changes in Cyanine 5/Cyanine 3 ratios.

From the AGILENT website

Bisulfite

Whole genome



CpG hypermethylation observed mainly at heterochromatic, inactive regions.

Due to spontaneous deamination, «C» is converted to «T» at a quit high rate. If repair fails, this leads to mutations.

One of the Repair systems in cells is linked to transcription (Transcription-coupled Repair)

As evolutionary consequence, active and regulatory regions have kept CG islands, while the remaining genome has a «C» content lower than expected.

An important fraction of human genes are associated to «CpG islands» (especially constitutive, house-keeping genes).