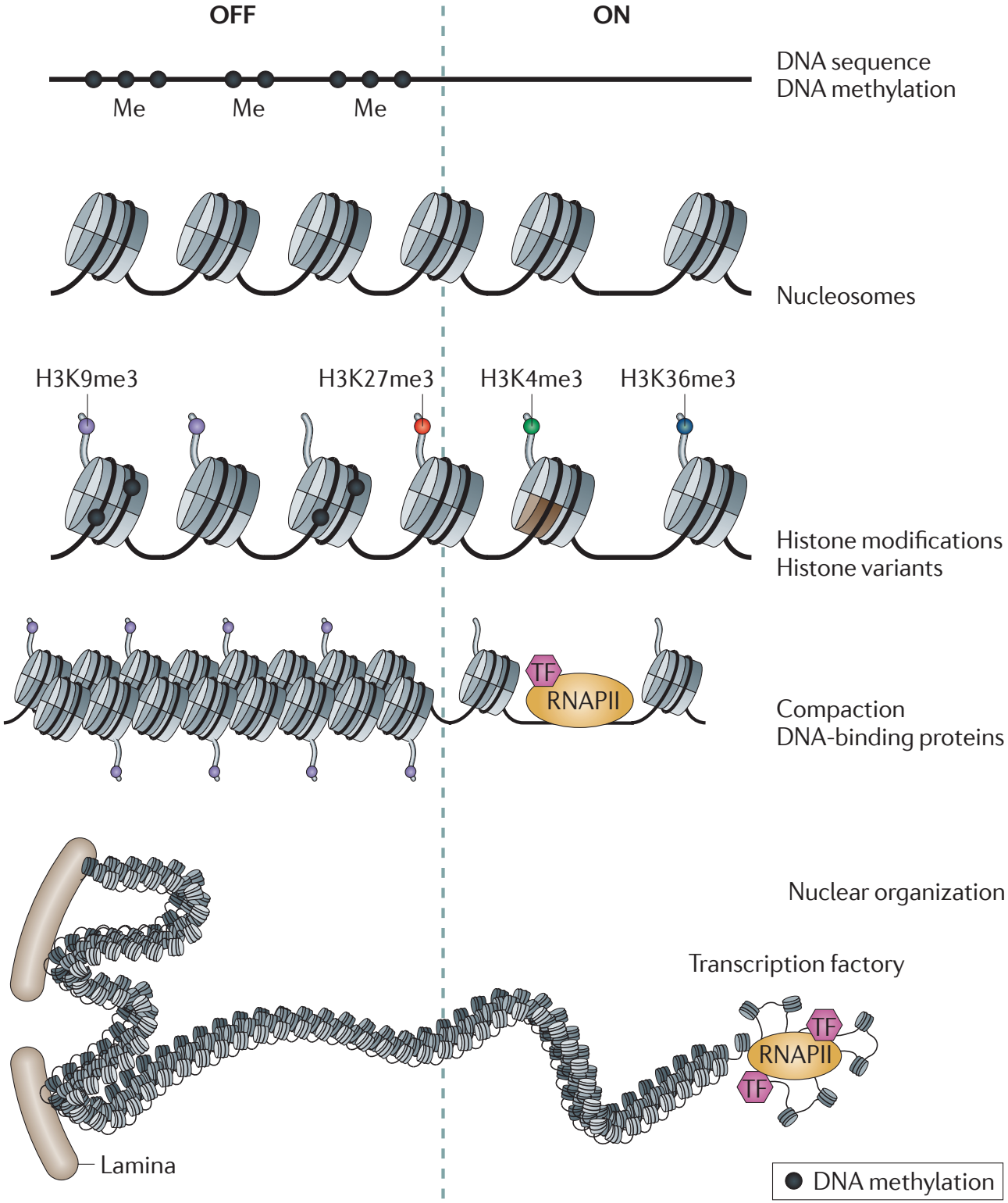
applied genomics

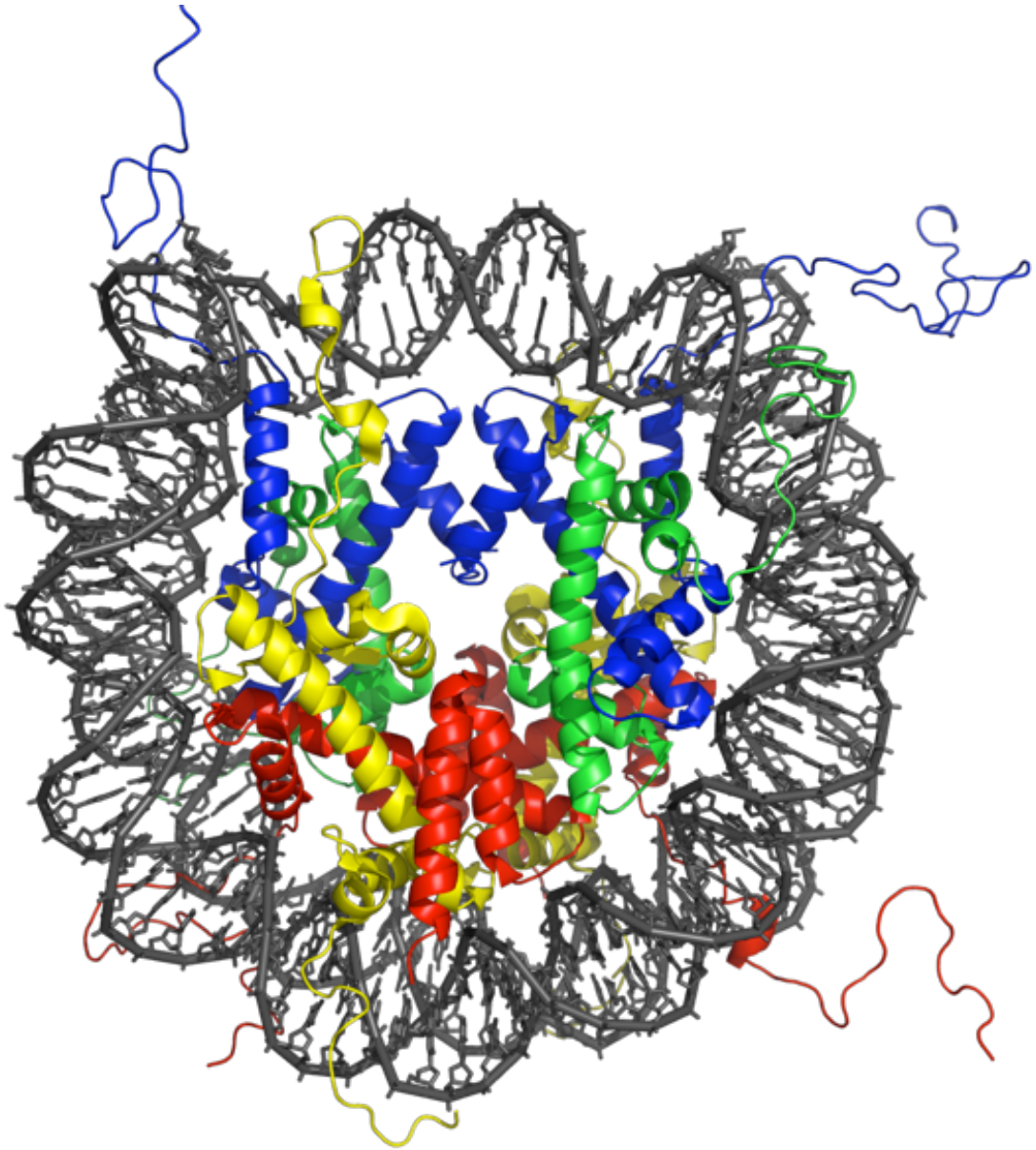
BIOL5V01 - Applied Genomics Laboratory Course

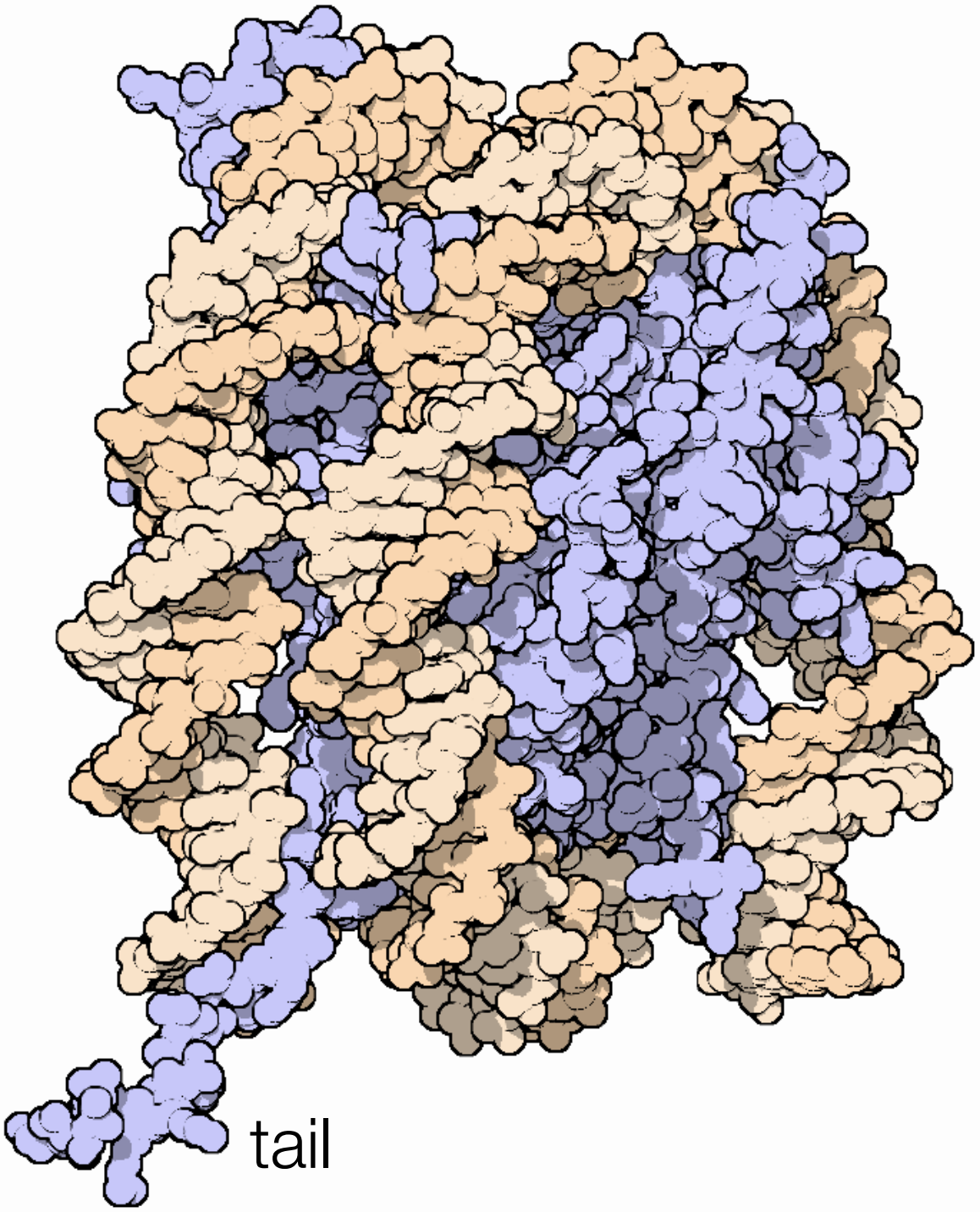
Tae Hoon Kim, Ph.D.
genome@utdallas.edu
http://taehoonkim.org


OFF
ON
DNA sequence
DNA methylation
Me
Me
Me
Nucleosomes
H3K9me3
H3K27me3
H3K4me3
H3K36me3
Histone modifications
Histone variants
TF
RNAPII
Compaction
DNA-binding proteins
Nuclear organization
Transcription factory
TF
RNAPII
TF
Lamina
DNA methylation

nucleosome - substrate for genetic processes

- *equal mass of histone to DNA in nucleus*
- *146bp DNA*
- *histones H2A , H2B , H3 and H4*
- *histone tails extend out from the nucleosome*
- *histone tails can be covalently modified at a number of specific residues*
- *nucleosomes can slide along the DNA or disassemble by ATP-dependent remodeling*
- *histone variants: H2AZ, H2AX, macroH2A, H3.3, CENPA*
- *linker histones*




tail

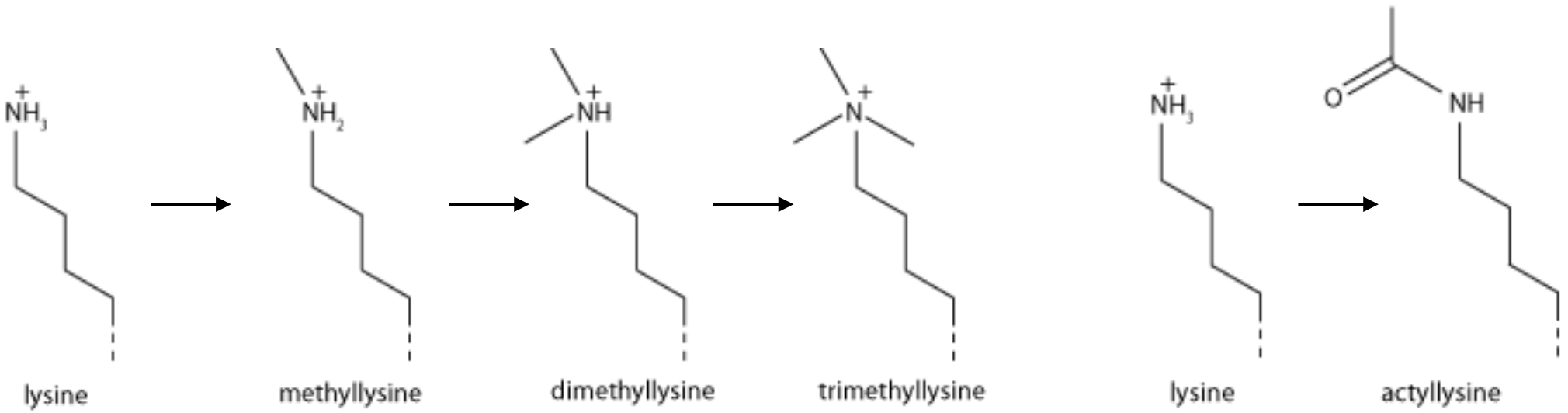
histone code

- *histones are an active component of epigenome*
- *histones compose equal amount of mass as DNA in the nucleus*

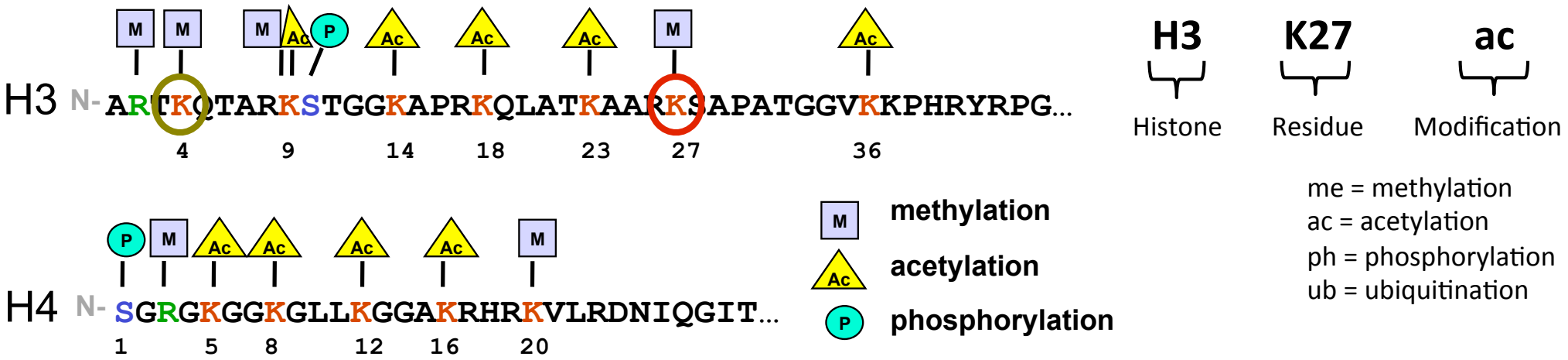
- *histone modifications single or combination serve as a code for regulation*
- *writers and readers of histone code*
  - *histone modifying enzymes are writers*
  - *modified histone binders are readers*

histone modifications and their functional associations

Type of modification	H3K4	H3K9	H3K14	H3K27	H3K79	H3K36	H4K20	H2BK5
mono-methylation	+	+		+	+		+	+
di-methylation		-		-	+			
tri-methylation	+	-		-	+, -	+		-
acetylation		+	+	+				



histone code

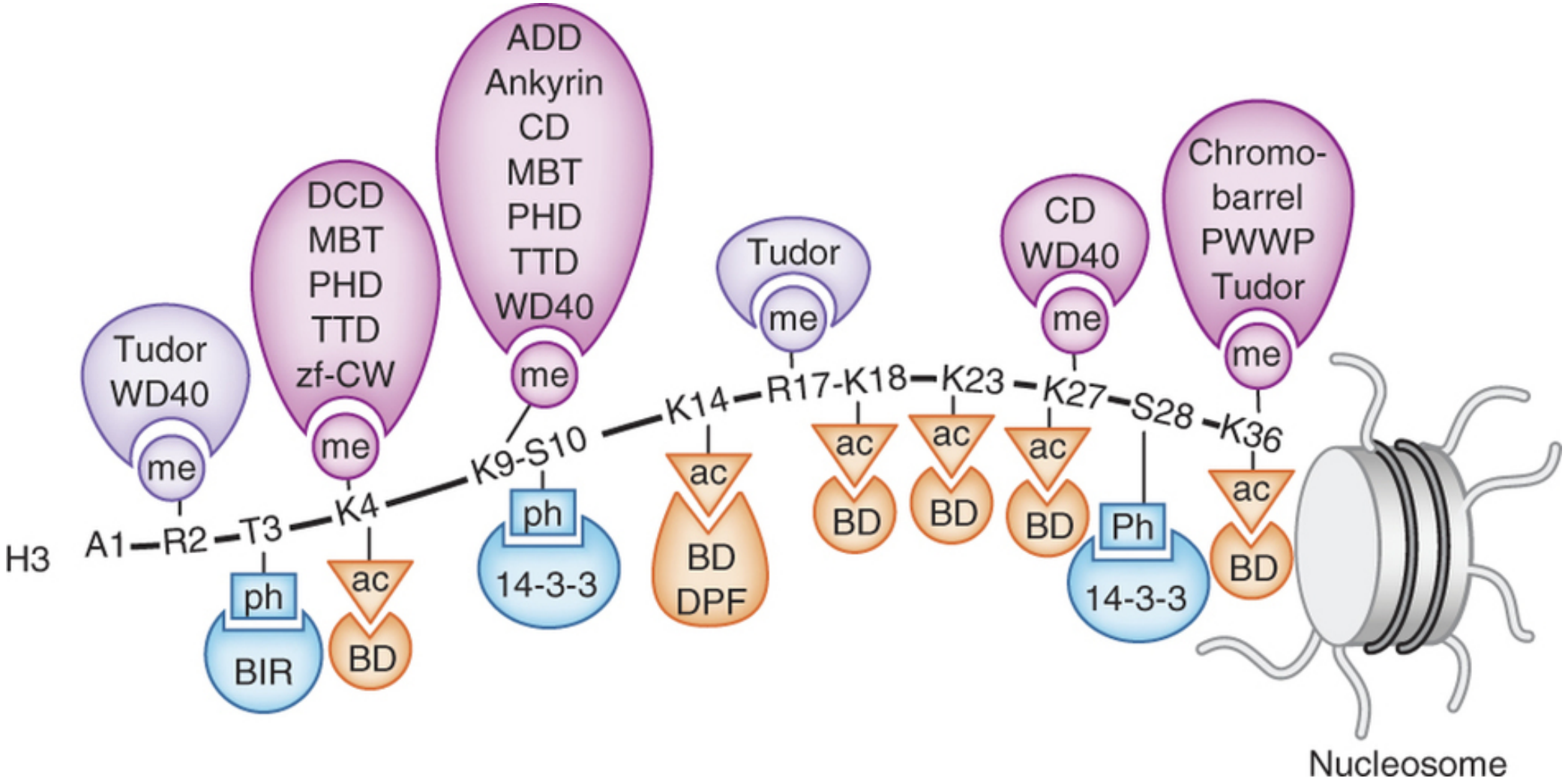


- *specific histone modifications serve as a docking site for other regulatory proteins that reads the marks*
- *specific histone modifications indicate gene activity and chromatin structures:*
  - *trimethylation of lysine **4** on histone H3 (H3K4me3) is associated with active genes (Trithorax)*
  - *trimethylation of lysine **27** on histone H3 (H3K27me3) is associated with silent genes (Polycomb)*
- *histone modifications are reversible and dynamic:*
  - *acetylated histones can be deacetylated*
  - *methylated histones can be demethylated*

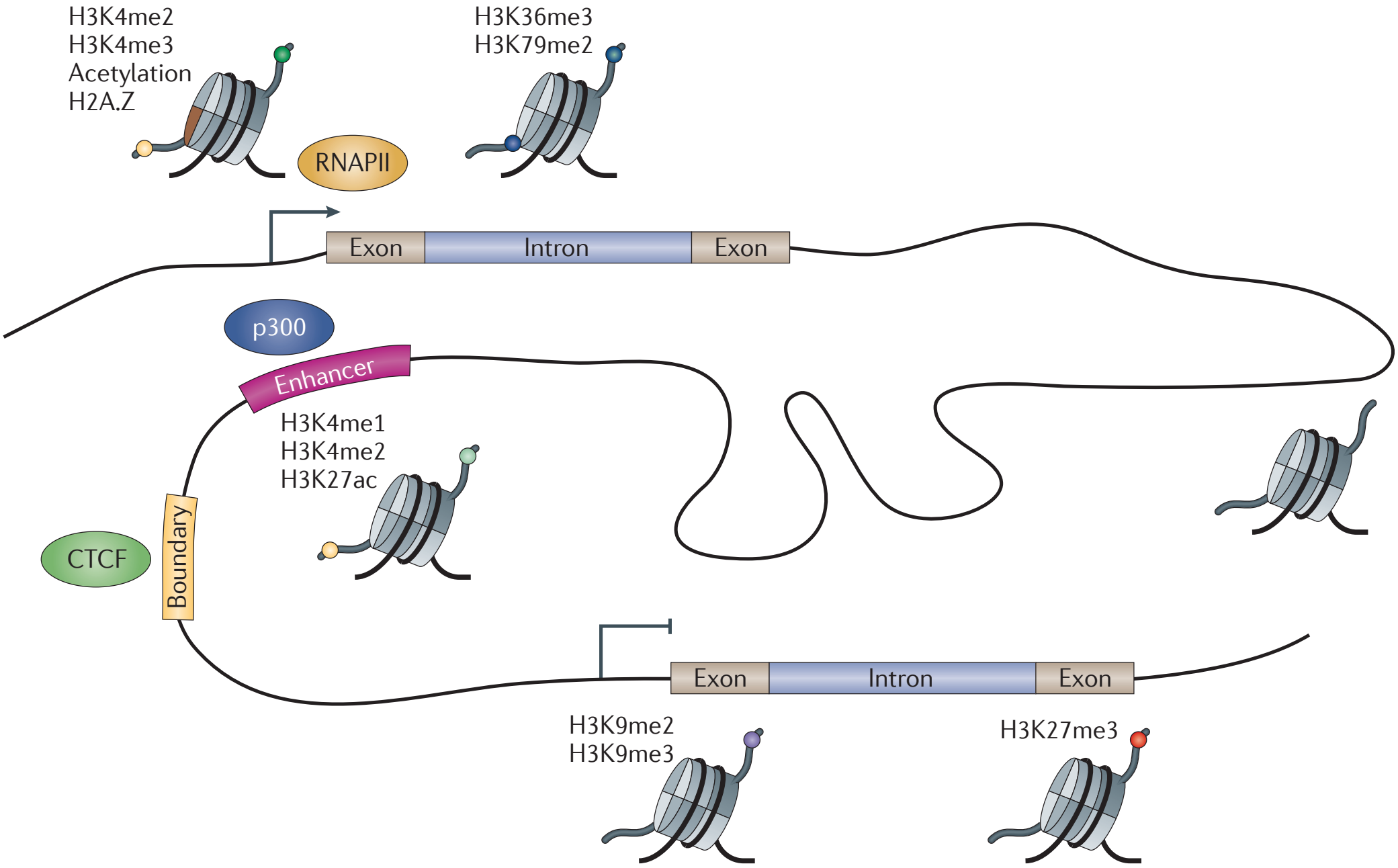
histone code writers/erasers

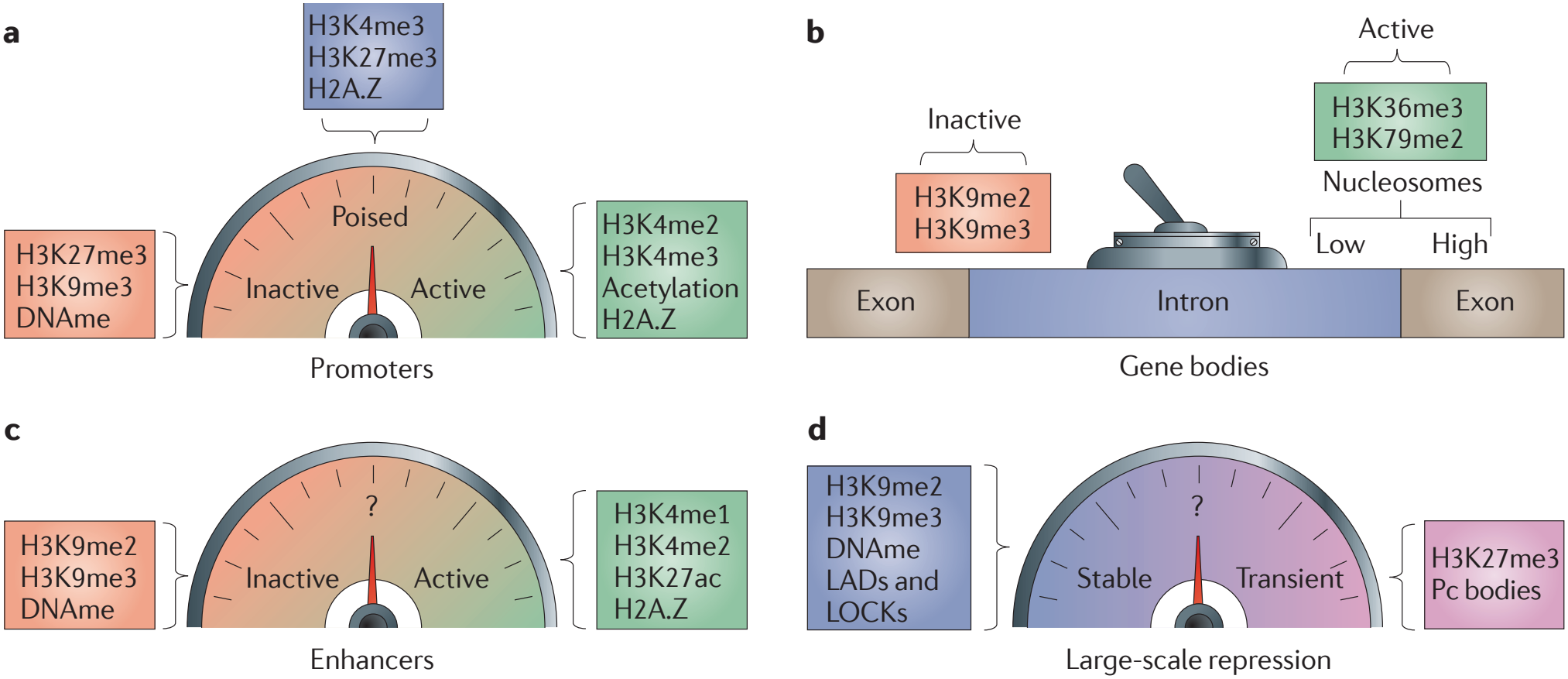
- *histone acetyl transferases (HAT)*
- *histone deacetylases (HDAC)*
- *lysine methyl transferases (KMT)*
- *lysine demethylases (KDM)*
- *kinases*
- *ubiquitin ligases*
- *deubiquitinating enzymes*
- *poly-ADP-ribose polymerase (PARP)*
- *poly-ADP-ribose glycohydrolase (PARG)*
- *…*

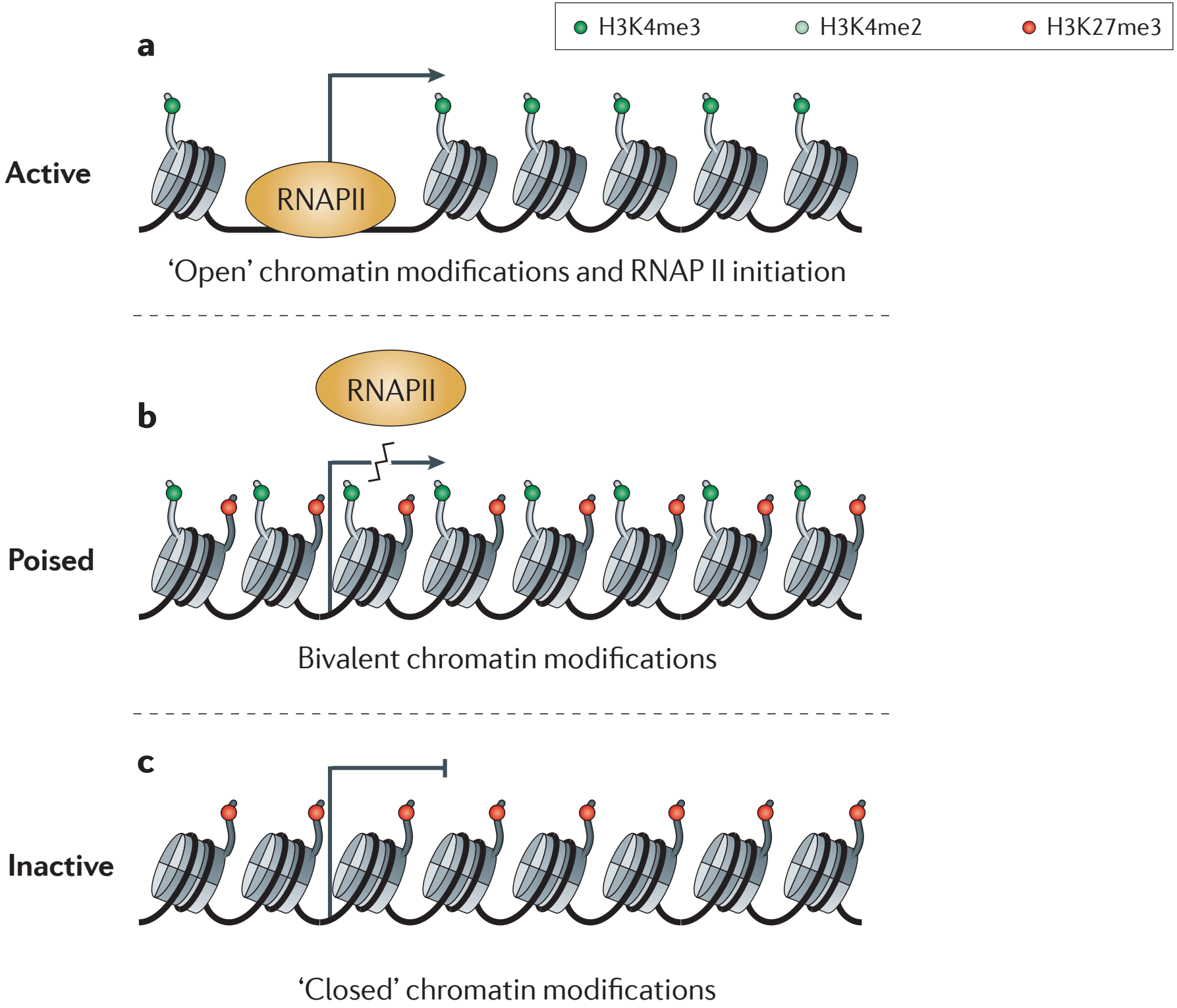
histone code readers

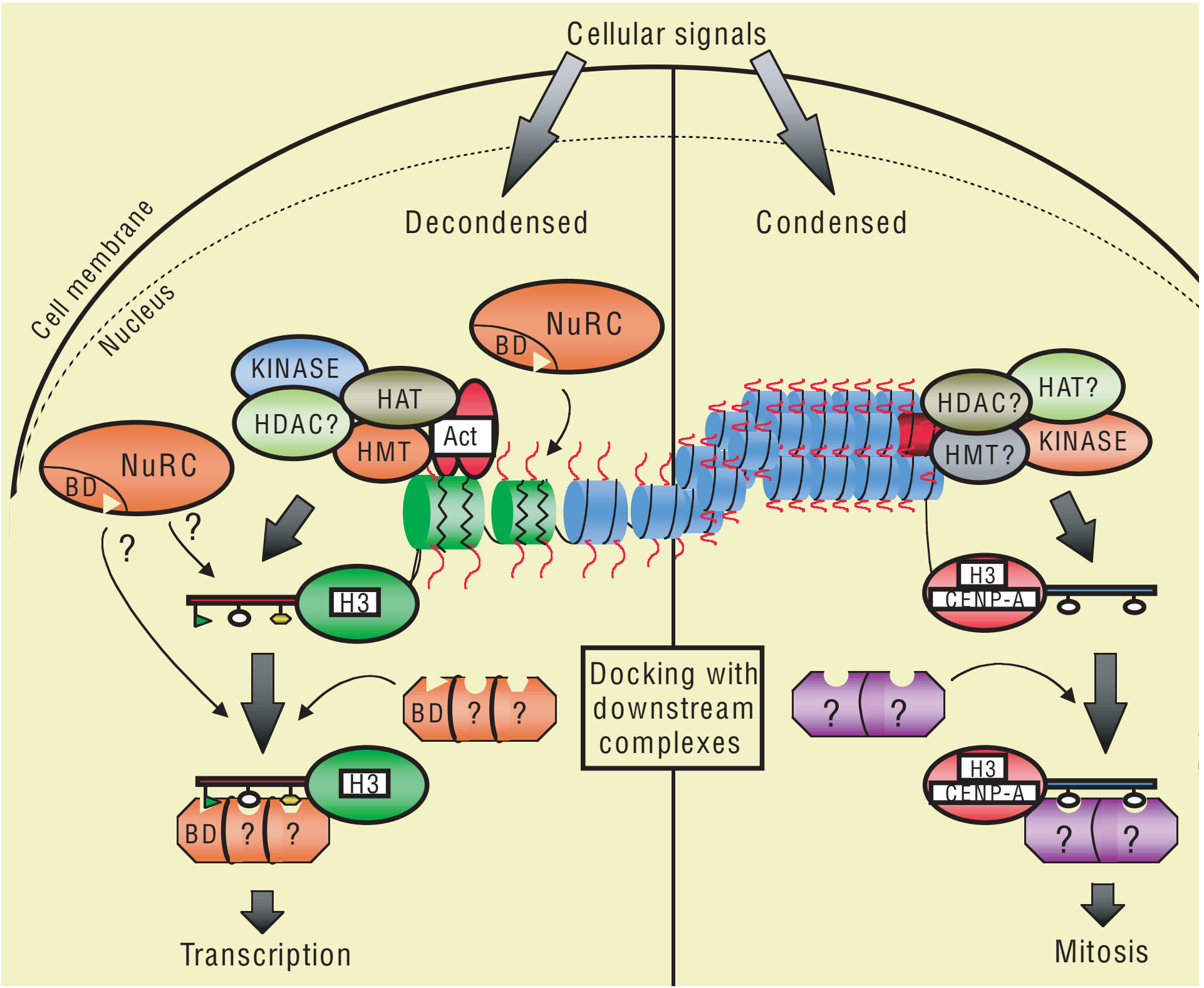


Musselman et al. (2012) Nat Struct Mol Biol 19:1218
Chi et al. (2010) Nat Rev Cancer 10:457

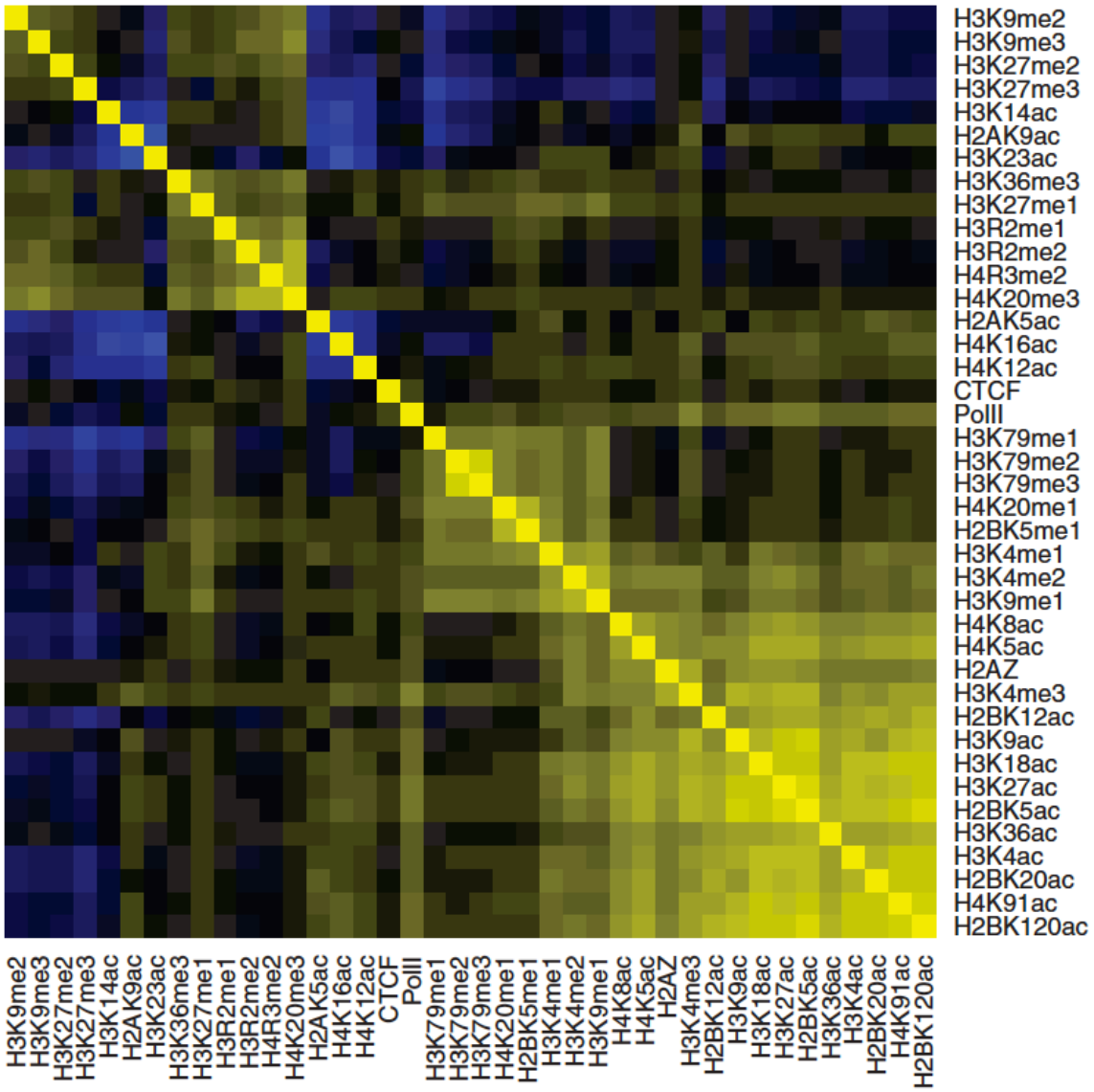

H3K4me2
H3K4me3
Acetylation
H2A.Z
RNAPII
H3K36me3
H3K79me2
Exon
Intron
Exon
p300
Enhancer
H3K4me1
H3K4me2
H3K27ac
CTCF
Boundary
Exon
Intron
Exon
H3K9me2
H3K9me3
H3K27me3


a
H3K4me3
H3K27me3
H2A.Z
Poised
H3K27me3
H3K9me3
DNAme
Inactive
Active
H3K4me2
H3K4me3
Acetylation
H2A.Z
Promoters
b
Active
H3K36me3
H3K79me2
Inactive
H3K9me2
H3K9me3
Nucleosomes
Low
High
Exon
Intron
Exon
Gene bodies
c
?
H3K9me2
H3K9me3
DNAme
Inactive
Active
H3K4me1
H3K4me2
H3K27ac
H2A.Z
Enhancers
d
H3K9me2
H3K9me3
DNAme
LADs and
LOCKs
?
Stable
Transient
H3K27me3
Pc bodies
Large-scale repression

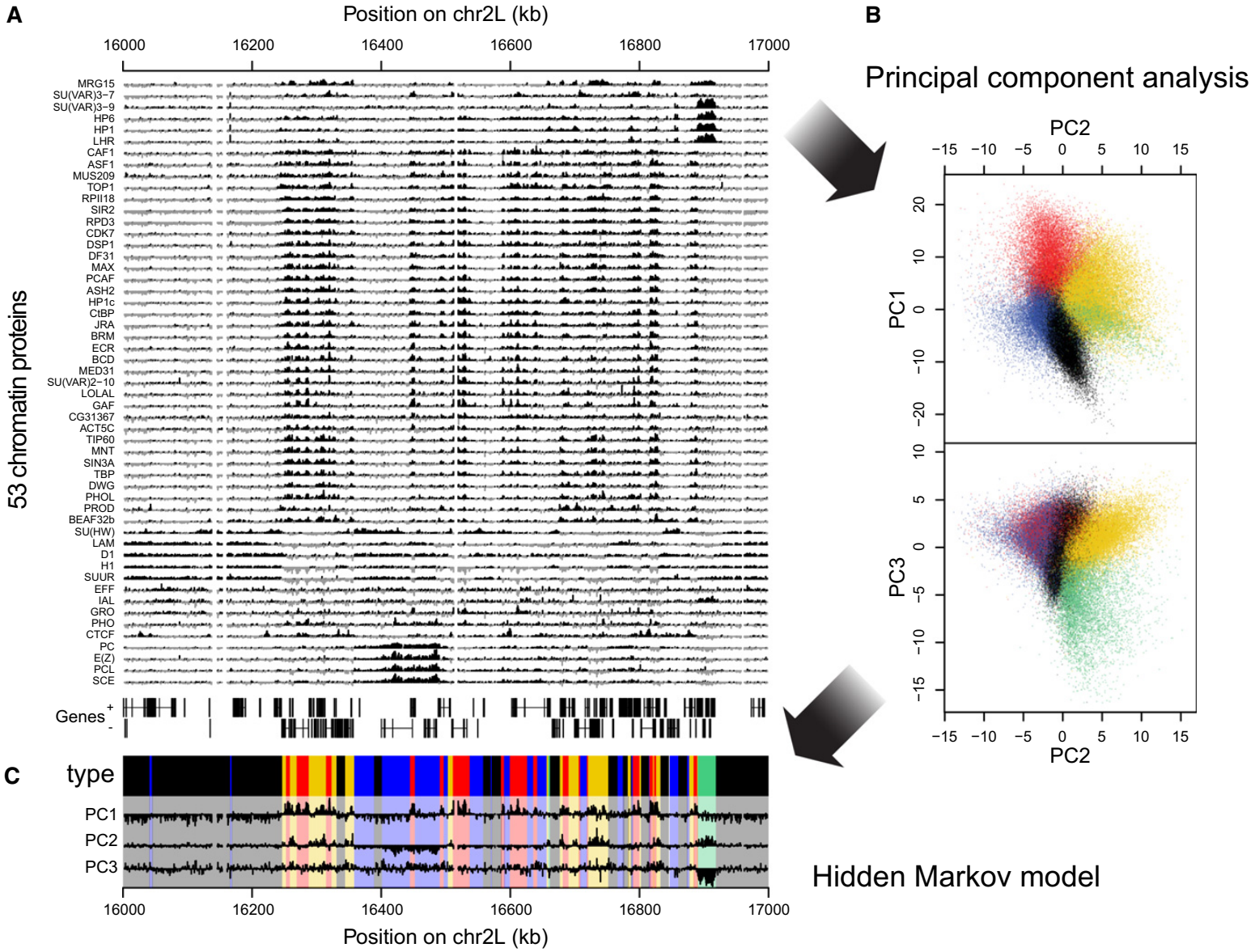

H3K4me3
H3K4me2
H3K27me3
a
Active
RNAPII
'Open' chromatin modifications and RNAP II initiation
RNAPII
b
Poised
Bivalent chromatin modifications
c
Inactive
'Closed' chromatin modifications


Cellular signals
Decondensed
Condensed
Cell membrane
Nucleus
NuRC
BD
KINASE
HDAC?
HAT
HMT
Act
?
?
H3
BD ? ?
Docking with downstream complexes
HDAC?
HAT?
HMT?
KINASE
H3
CENP-A
? ?
Transcription
Mitosis

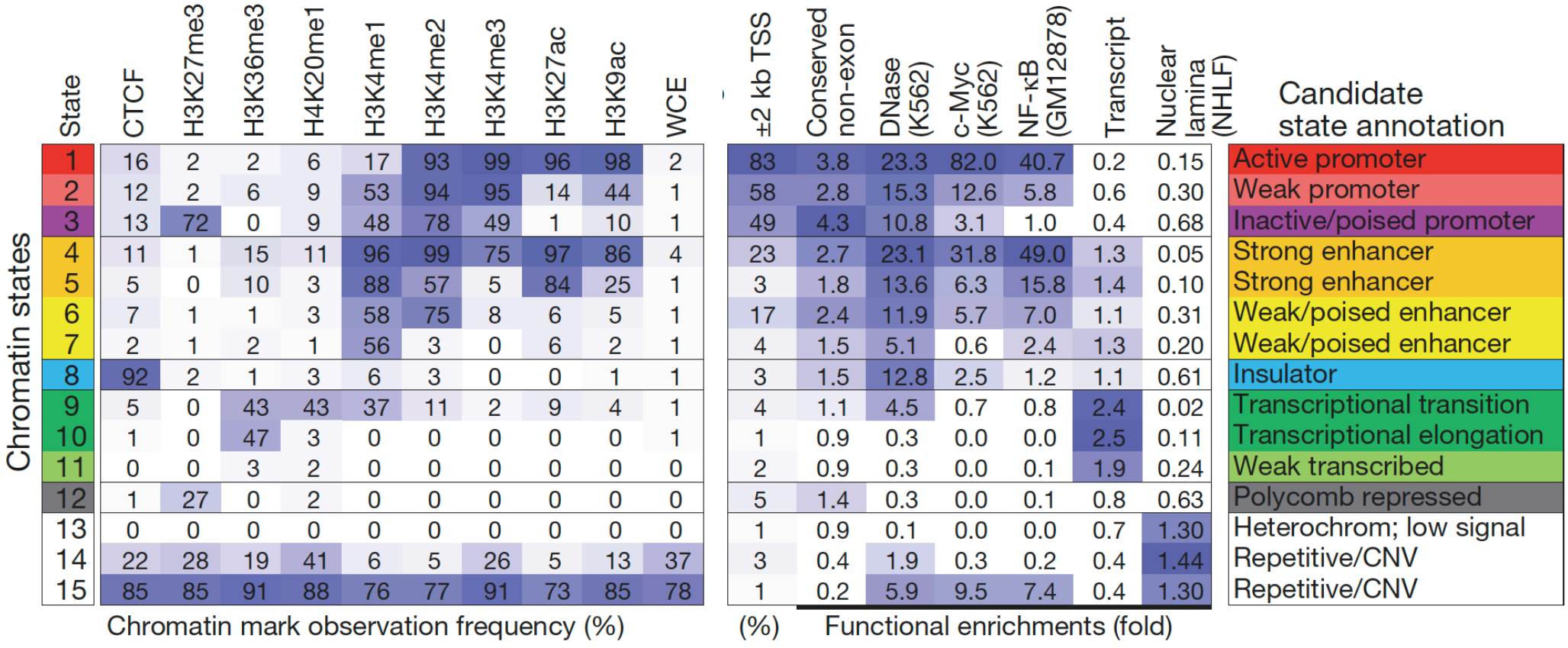
histone modifications are correlated



five colors of Drosophila chromatin

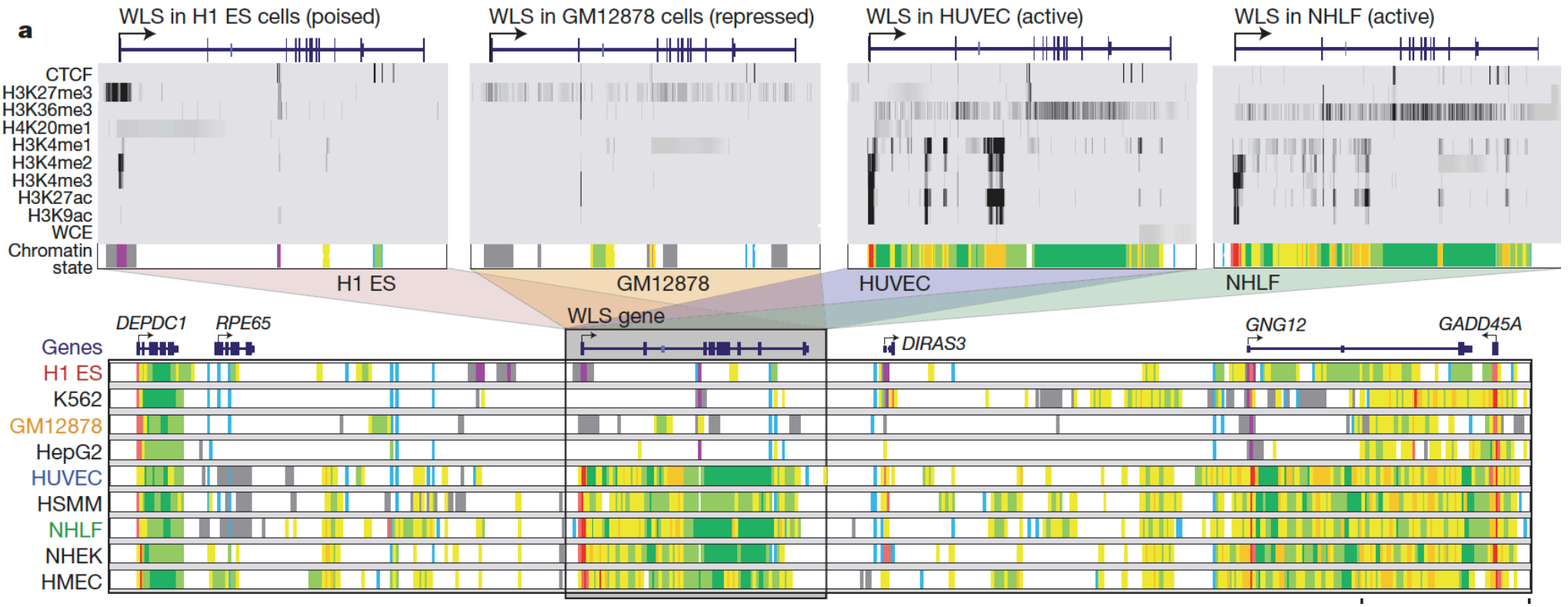


modeling chromatin states in the human genome



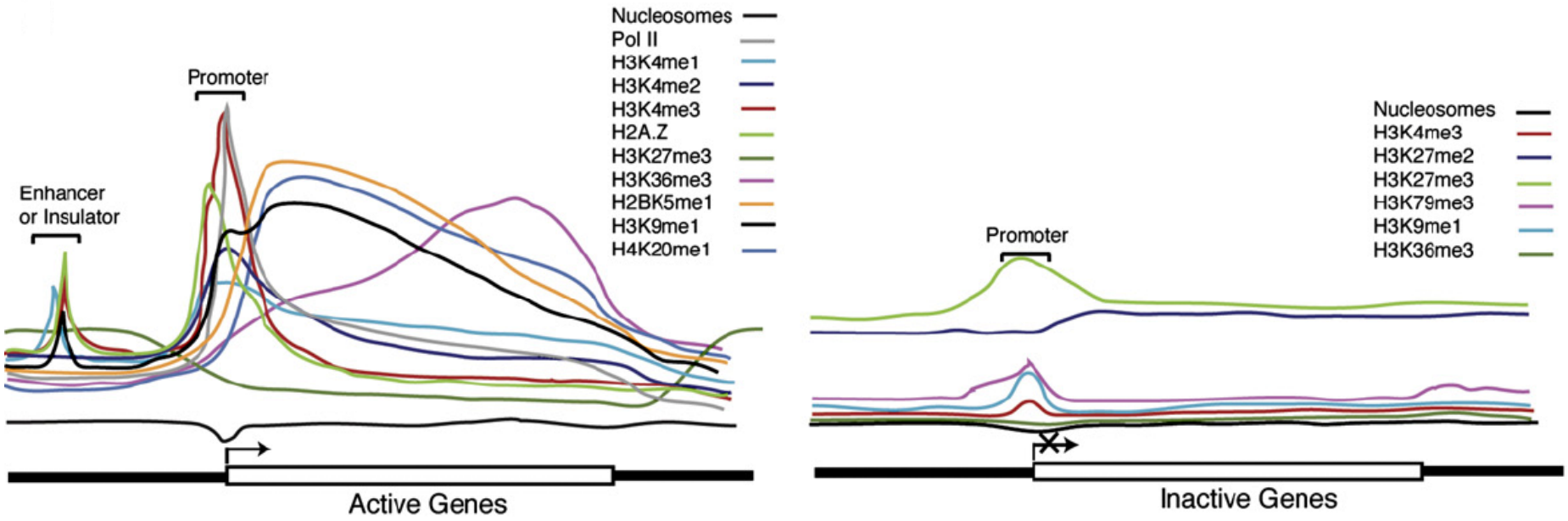
Ernst *et al.*, *Nature* 473:43 (2011)

chromatin states vary across cell types



histone methylation at promoters, enhancers and gene bodies

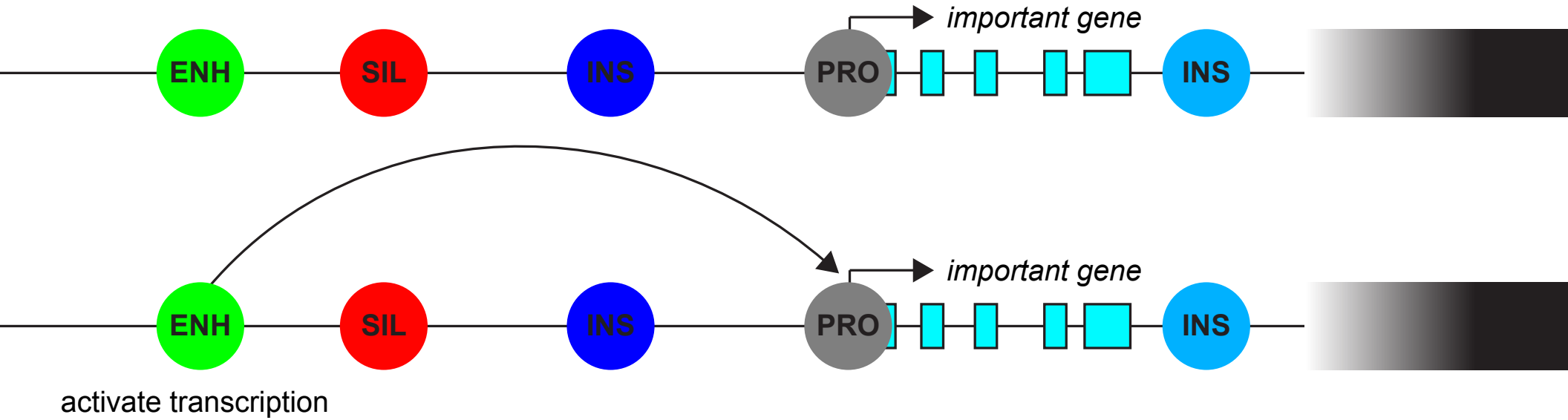
T-cells



- *gene activity and histone modification*
  - *genes activity can be predicted from histone modification signature*
  - *direction of transcription can be determined from histone modification pattern*
- *RNA polymerase is paused at the promoter - transcription elongation (not polymerase recruitment) might be the rate limiting step for gene expression*

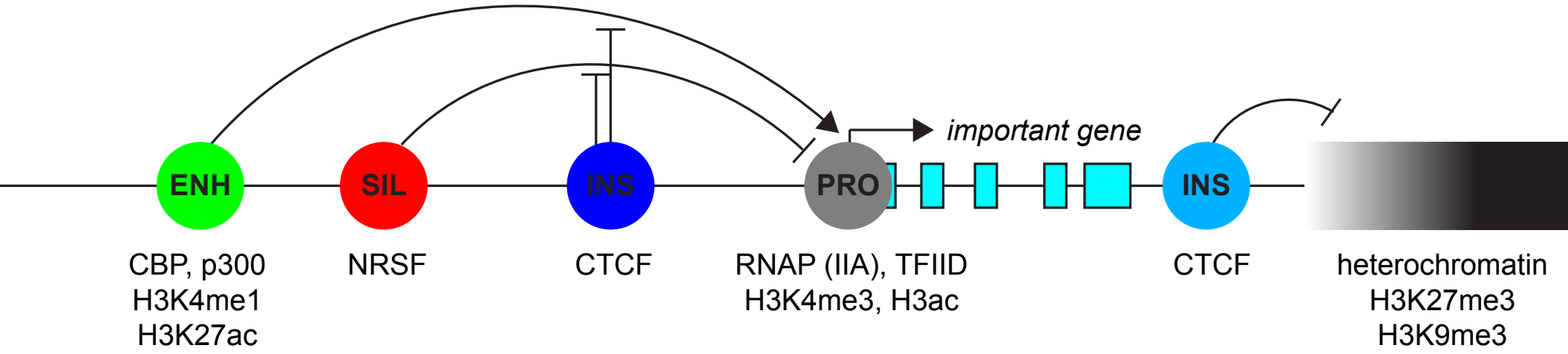
Barski et al Cell 129:823 (2007)

how do you regulate a gene?



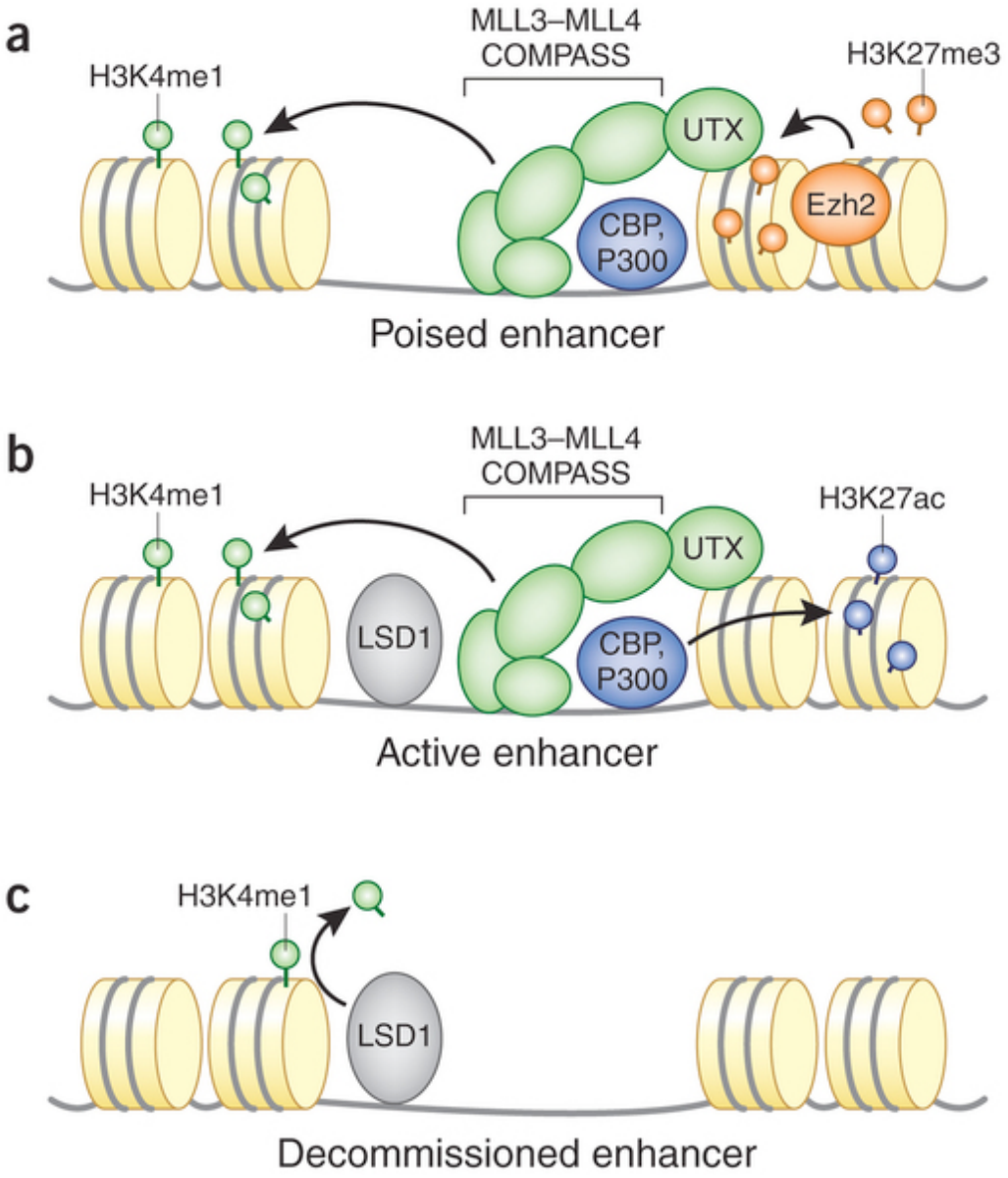
activate transcription

- *sequence specific transcription factors bind to enhancer and promoter to recruit the transcription machinery (general transcription factors and RNA polymerase) to synthesize mRNA*
- *how to specify regulation?*
  - *combinatorial regulation*
  - *coregulators (co-activators/co-repressor) serve to integrate cis-regulatory signal*


ENH
SIL
INS
PRO
INS
important gene
CBP, p300
H3K4me1
H3K27ac
NRSF
CTCF
RNAP (IIA), TFIID
H3K4me3, H3ac
CTCF
heterochromatin
H3K27me3
H3K9me3

enhancer regulation

- *poised enhancers are marked with H3K4me1 by MLL3/4 (TrxG) complex*
- *active enhancer recruits p300/CBP and acetylates histones resulting in H3K27ac*
- *when enhancer is decomissioned, histone demethylase LSD1 is recruited to remove H3K4me1*

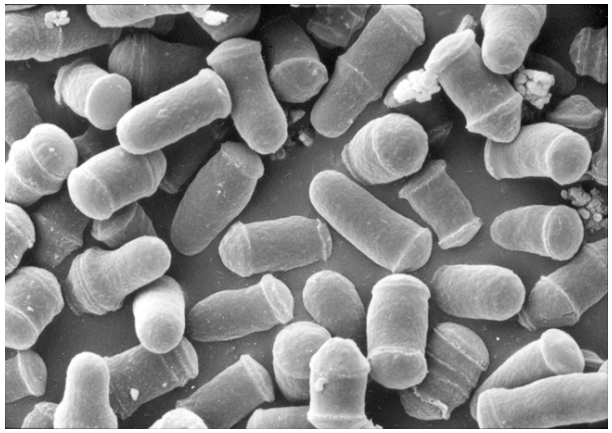


budding yeast

- *12 chromosomes*
- *12Mb*
- *no HMT for H3K9*
- *no HP1 homolog*

yeastgenome.org

unmodified chromatin



fission yeast

- *3 large chromosomes*
- *12.6 Mb*
- *many repeats and transposons*
- *H3K9me3*
- *HP1 homolog*

pombase.org

H3K9me3

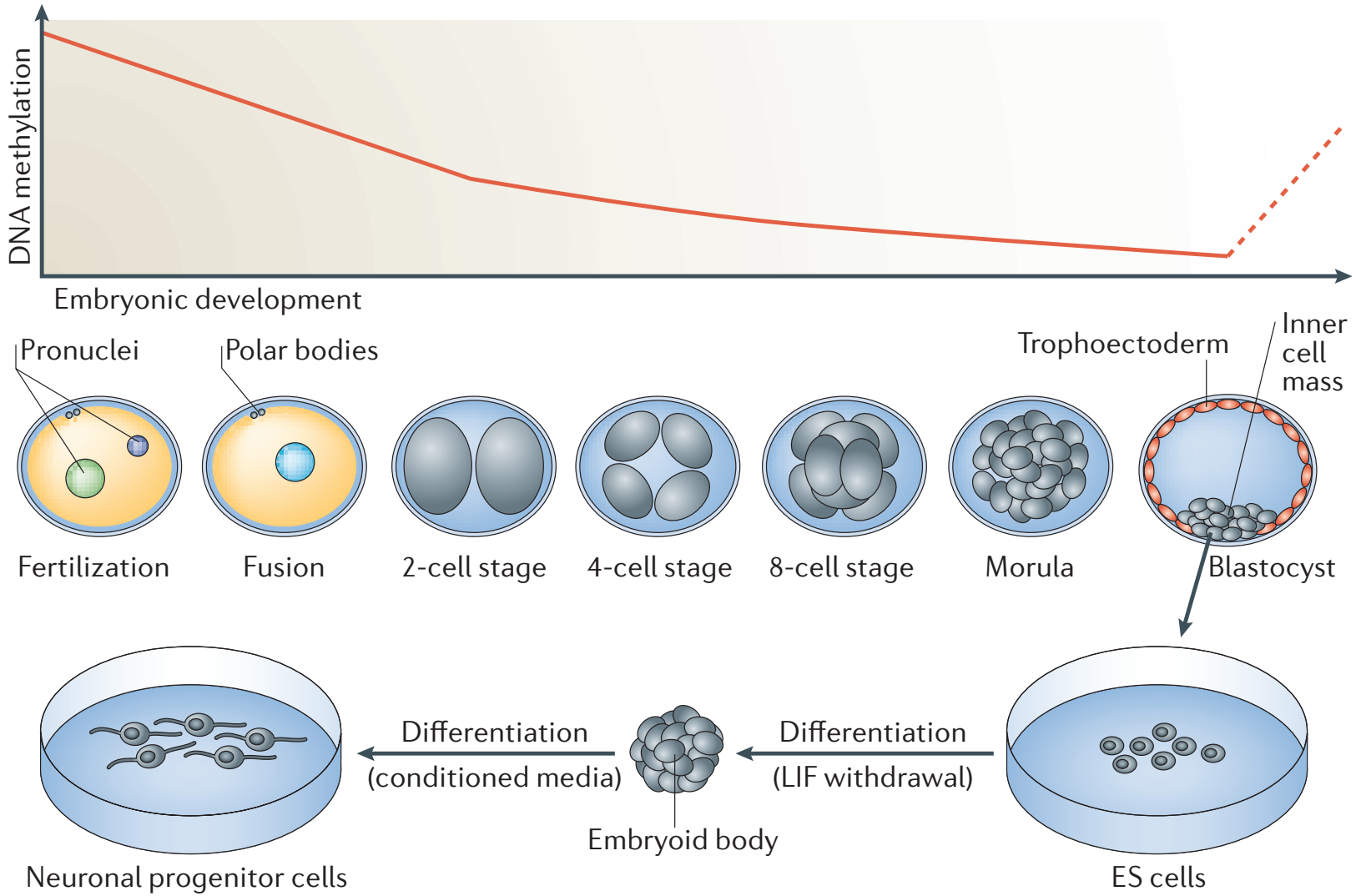
fruit fly

- *5 chromosomes*
- *118.4Mb*
- *many repeats and transposons*
- *H3K9me3*
- *HP1*
- *H3K27me3*
- *Polycomb*

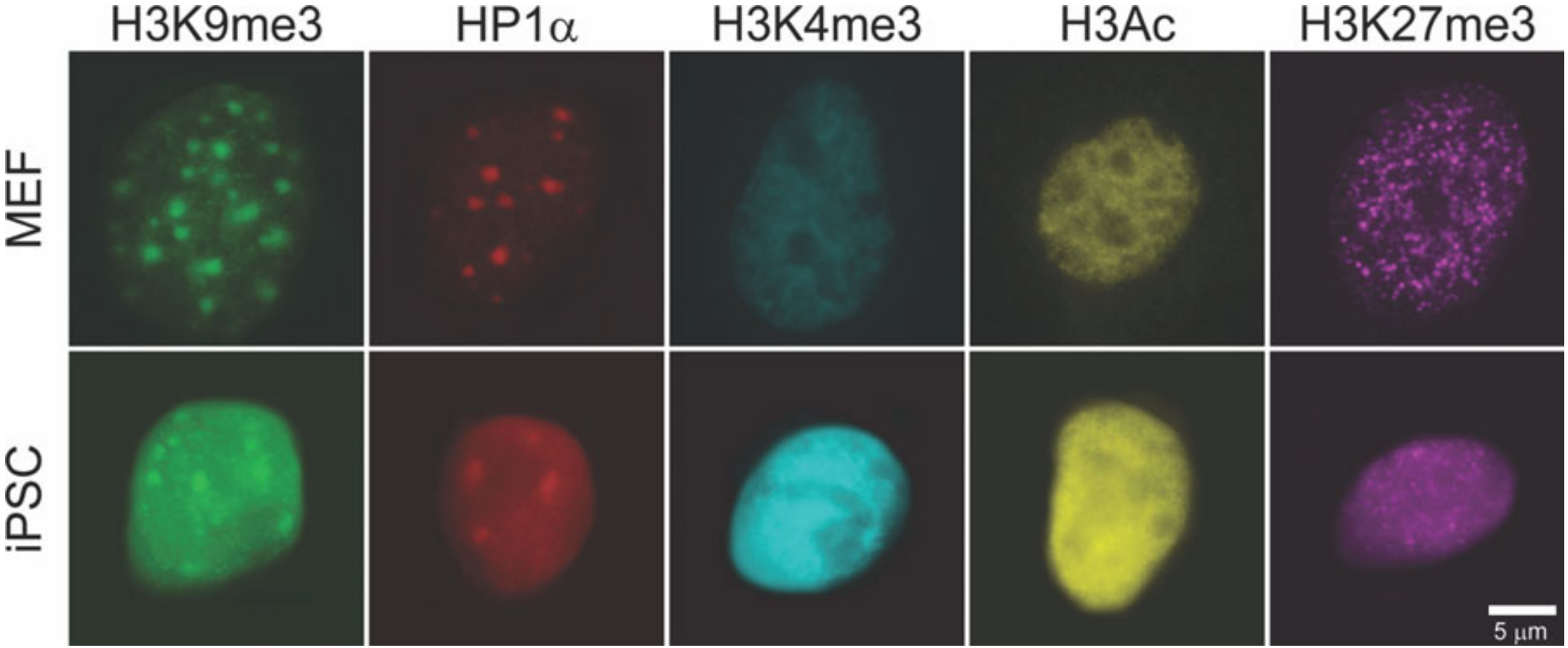
flybase.org

H3K9me3, H3K27me3

DNA methylation changes during development



nuclear differences between embryonic stem cells and differentiated cells



Biran A, Meshorer E. Concise Review: Chromatin and Genome Organization in Reprogramming. *STEM CELLS*. 2012;30(9):1793-1799.

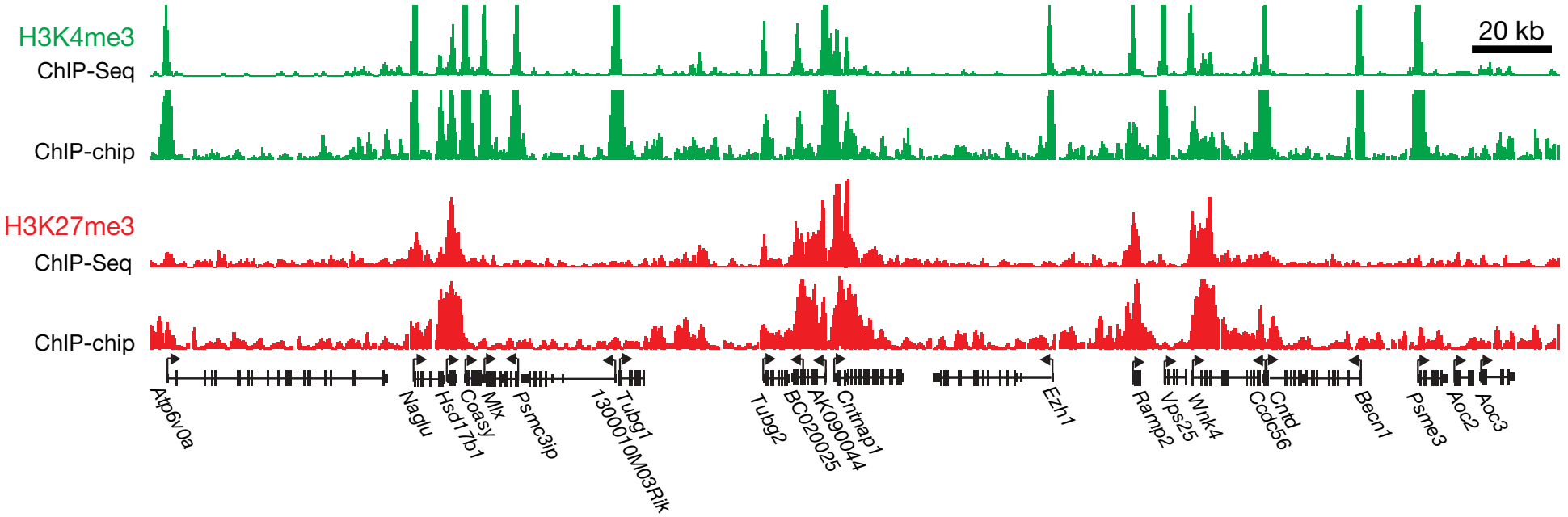
Vol 448|2 August 2007|doi:10.1038/nature06008

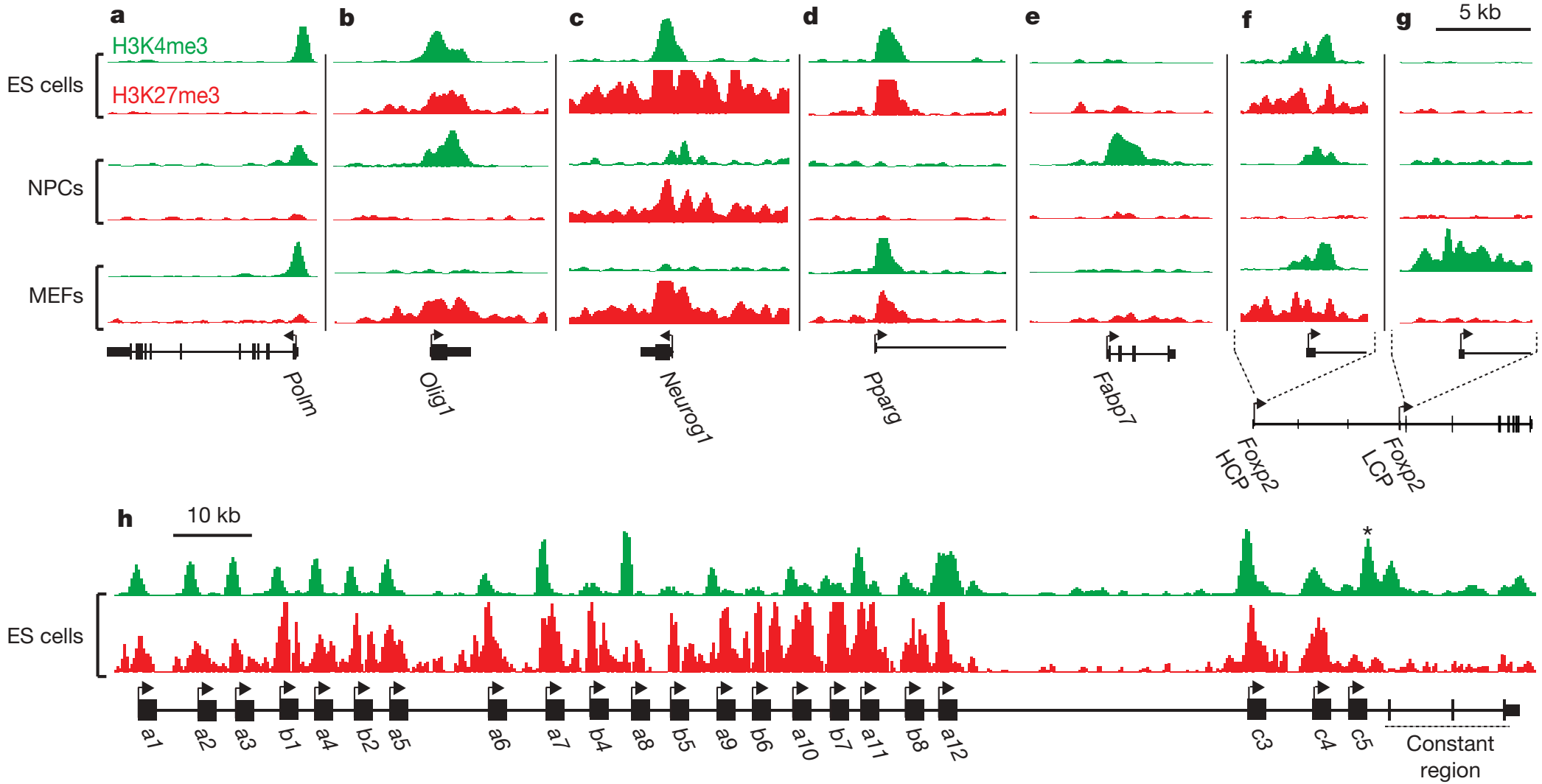
nature

ARTICLES

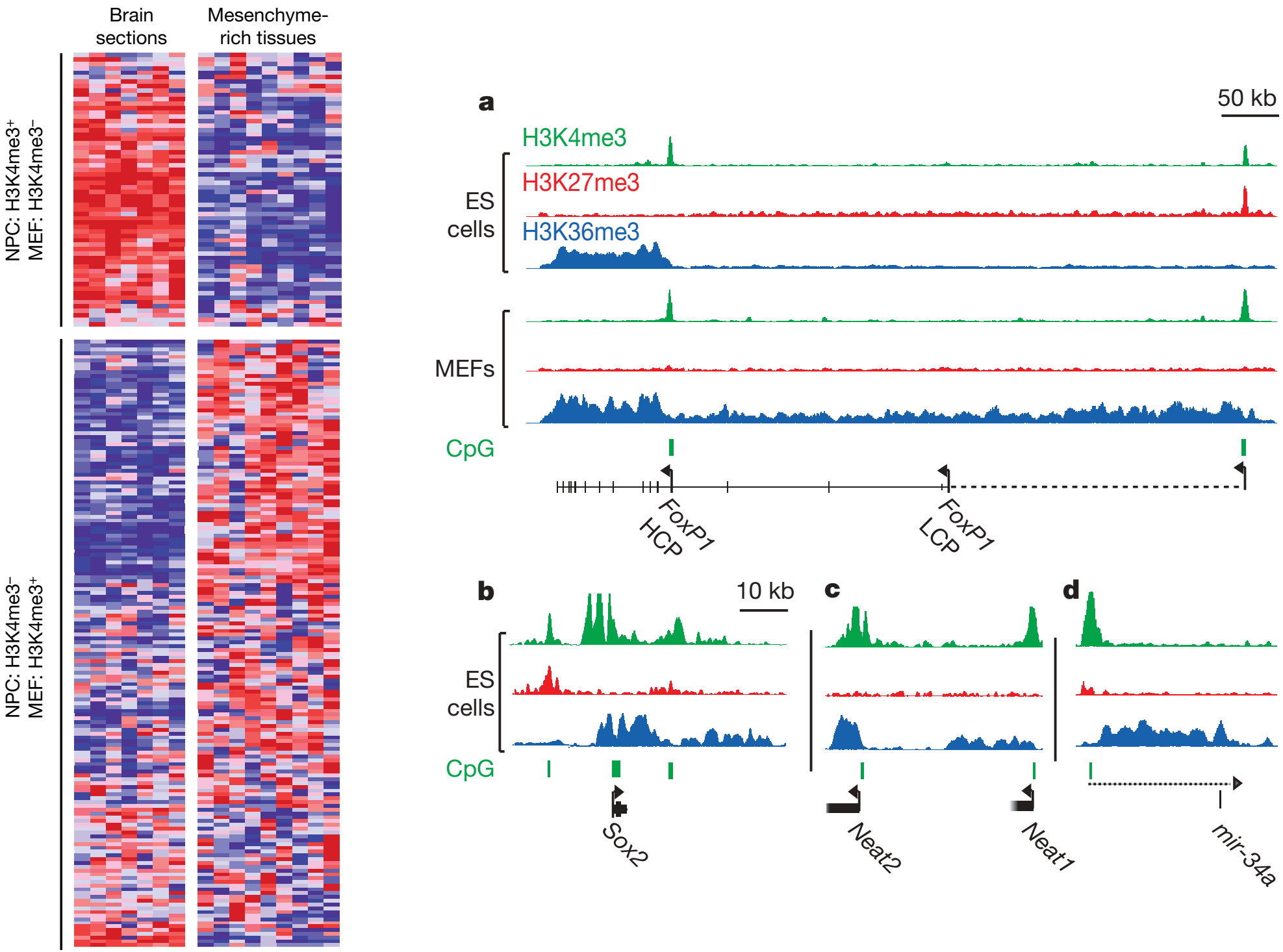
Genome-wide maps of chromatin state in pluripotent and lineage-committed cells

Tarjei S. Mikkelsen[1,2], Manching Ku[1,4], David B. Jaffe[1], Biju Issac[1,4], Erez Lieberman[1,2], Georgia Giannoukos[1], Pablo Alvarez[1], William Brockman[1], Tae-Kyung Kim[5], Richard P. Koche[1,2,4], William Lee[1], Eric Mendenhall[1,4], Aisling O'Donovan[4], Aviva Presser[1], Carsten Russ[1], Xiaohui Xie[1], Alexander Meissner[3], Marius Wernig[3], Rudolf Jaenisch[3], Chad Nusbaum[1], Eric S. Lander[1,3]* & Bradley E. Bernstein[1,4,6]*


H3K4me3
ChIP-Seq
ChIP-chip
H3K27me3
ChIP-Seq
ChIP-chip
20 kb
Atp6v0a
Naglu
Hsd17b1
Coasy
Mlx
Psmc3ip
Tubg1
1300010M03Rik
Tubg2
BC020025
AK090044
Cntnap1
Ezh1
Ramp2
Vps25
Wnk4
Cntd
Ccdc56
Becn1
Psme3
Aoc2
Aoc3



Mikkelsen et al, Nature 2007



Mikkelsen et al, Nature 2007

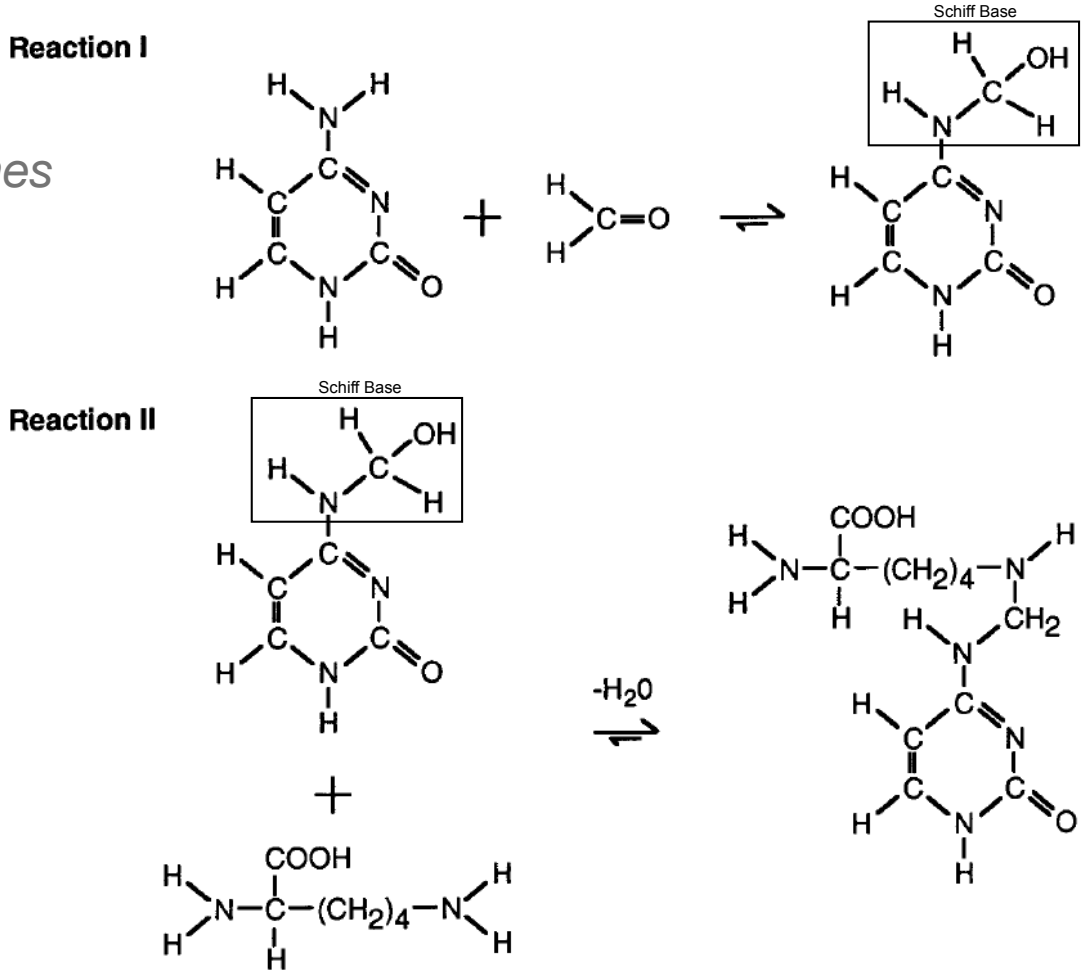
Encyclopedia of DNA
elements (ENCODE)
Project since 2003

http://www.nature.com/encode/#/threads

http://genome.ucsc.edu/ENCODE

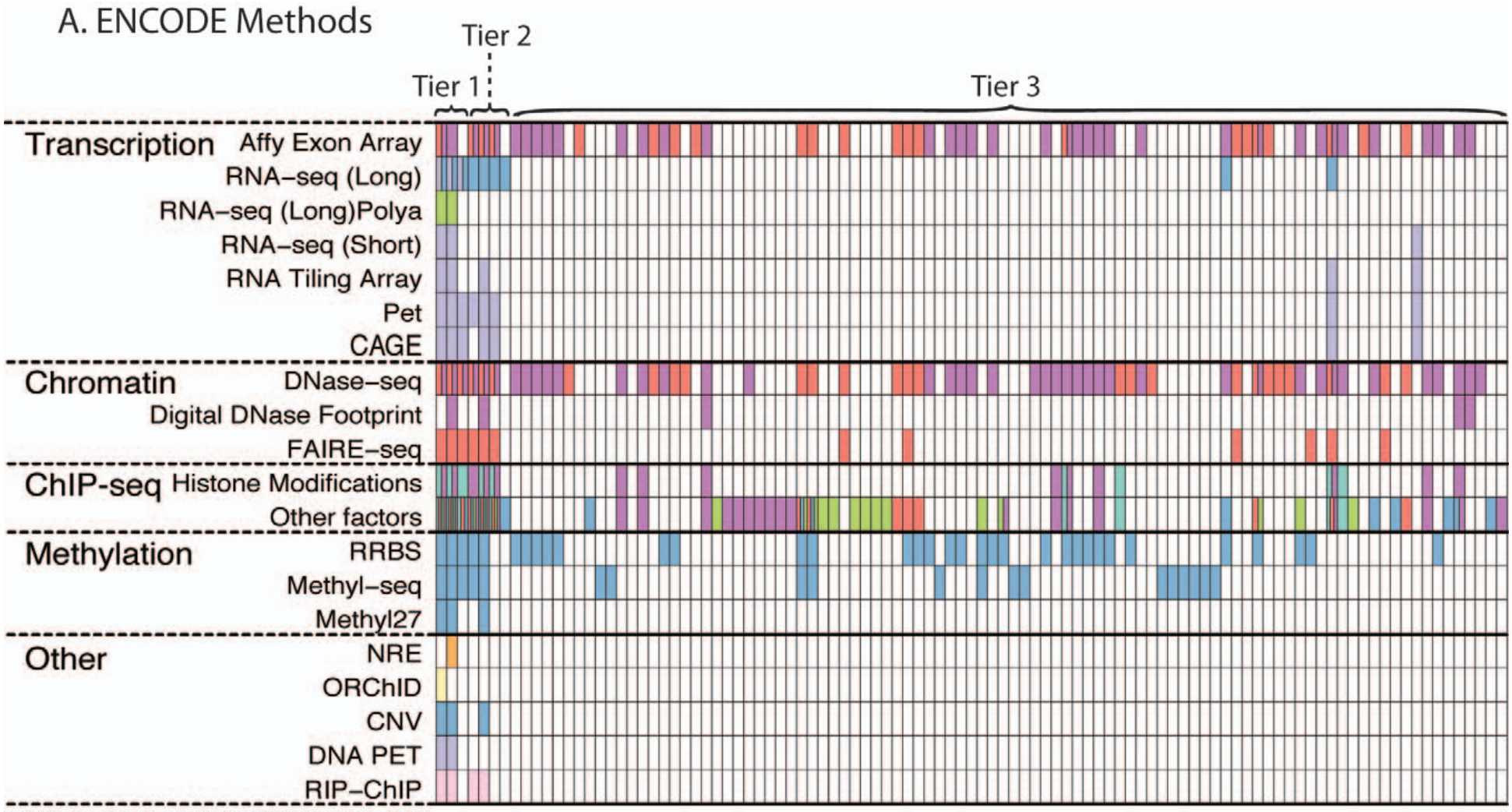
chemistry of formaldehyde crosslinking

- *versatile 2 Å crosslinker*
- *highly cell permeable*
- *limited by availability of primary amines in the vicinity*
- *a selective crosslinker - not general*
  - *requires close proximity of free amines*
  - *some proteins not crosslinked*
- *crosslinks protein-proteins, protein-DNA contacts*
- *reversible*



considerations

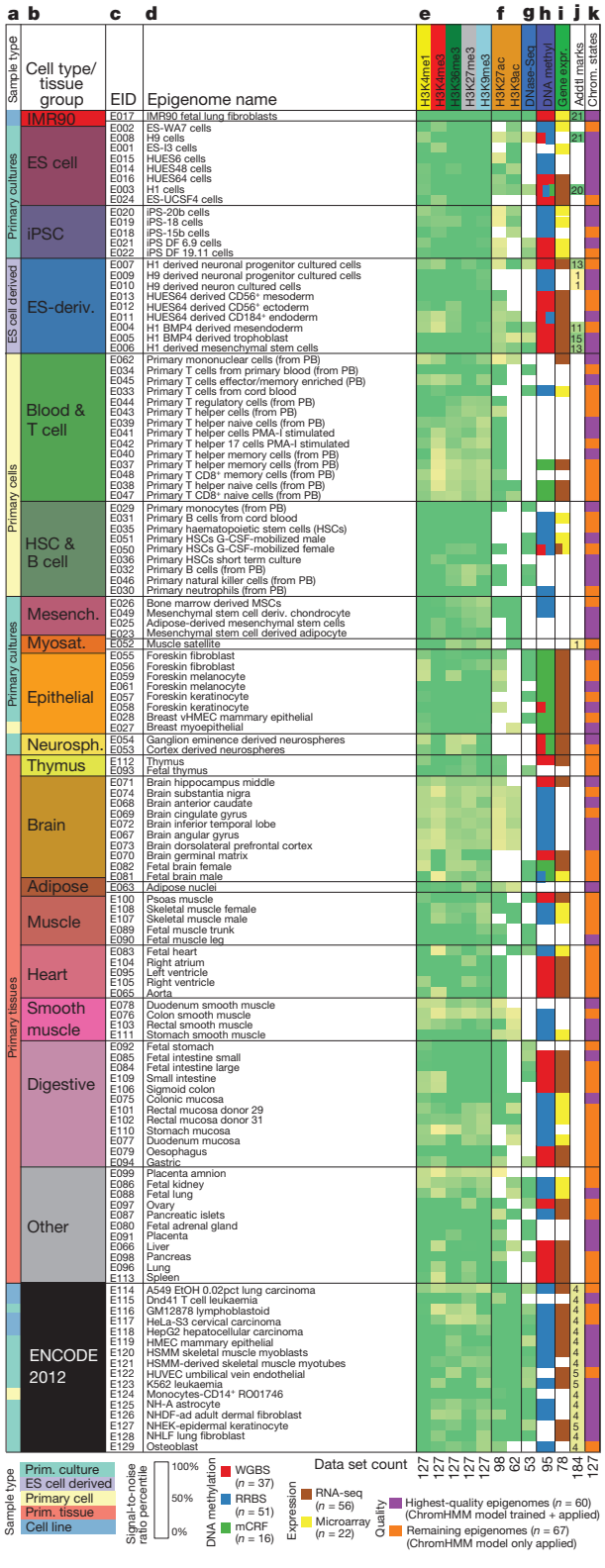
- *direct vs indirect associations*
- *ensemble measurement of association*
  - *can't discern binding at individual chromosomes*
- *requires specific antibodies*
- *tagging not always effective*


A. ENCODE Methods
Tier 1
Tier 2
Tier 3
Transcription
Affy Exon Array
RNA–seq (Long)
RNA–seq (Long)Polya
RNA–seq (Short)
RNA Tiling Array
Pet
CAGE
Chromatin
DNase–seq
Digital DNase Footprint
FAIRE–seq
ChIP-seq
Histone Modifications
Other factors
Methylation
RRBS
Methyl–seq
Methyl27
Other
NRE
ORChID
CNV
DNA PET
RIP–ChIP

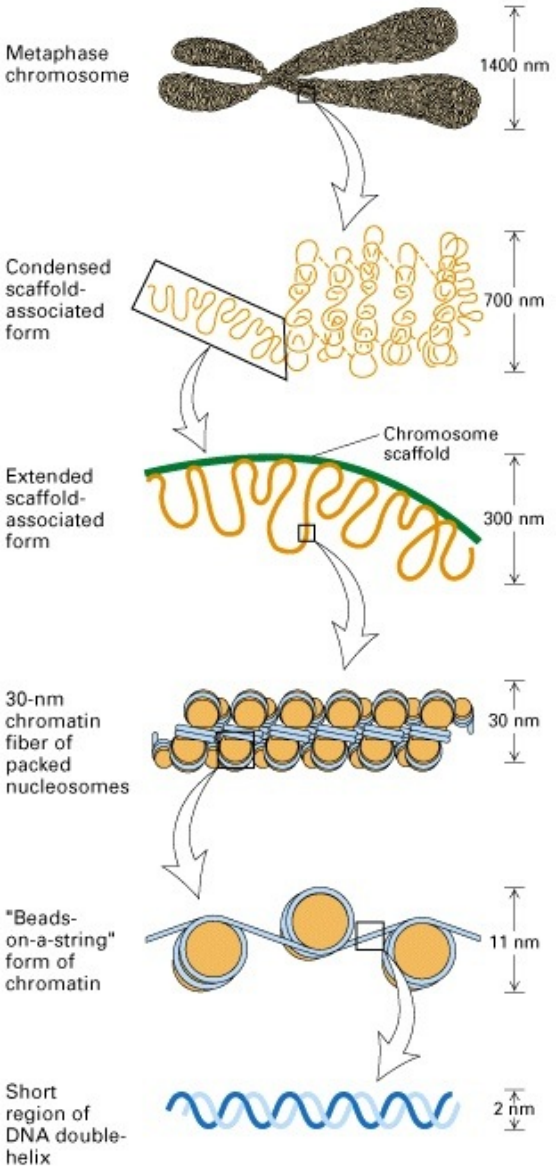
>80% of the genome is functional

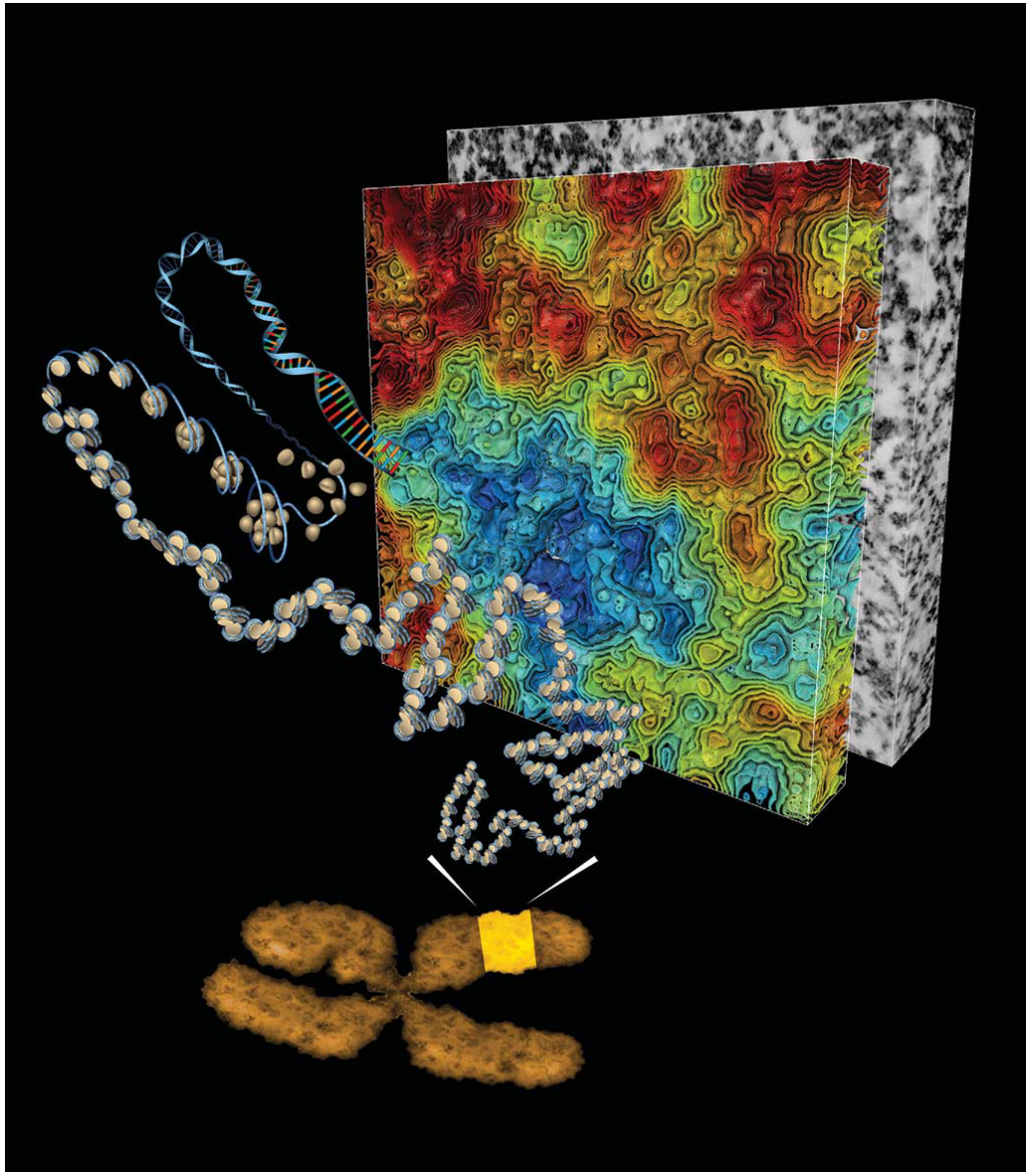
ENCODE defined product or reproducible biochemical activity

http://www.nature.com/epigenomeroadmap


a
b
c
d
e
f
g
h
i
j
k
Sample type
Cell type/ tissue group
EID
Epigenome name
H3K4me1
H3K4me3
H3K36me3
H3K27me3
H3K9me3
H3K27ac
H3K9ac
DNase-Seq
DNA methyl
Gene expr.
Add'l marks
Chrom. states
Primary cultures
ES cell derived
Primary cells
Primary tissues
Cell line
IMR90
ES cell
iPSC
ES-deriv.
Blood & T cell
HSC & B cell
Mesench.
Myosat.
Epithelial
Neurosph.
Thymus
Brain
Adipose
Muscle
Heart
Smooth muscle
Digestive
Other
ENCODE 2012
Prim. culture
ES cell derived
Primary cell
Prim. tissue
Cell line
Signal-to-noise ratio percentile
100%
50%
0%
DNA methylation
WGBS (n = 37)
RRBS (n = 51)
mCRF (n = 16)
Expression
RNA-seq (n = 56)
Microarray (n = 22)
Quality
Highest-quality epigenomes (n = 60) (ChromHMM model trained + applied)
Remaining epigenomes (n = 67) (ChromHMM model only applied)
Data set count
127 127 127 127 127 98 62 53 95 78 184 127

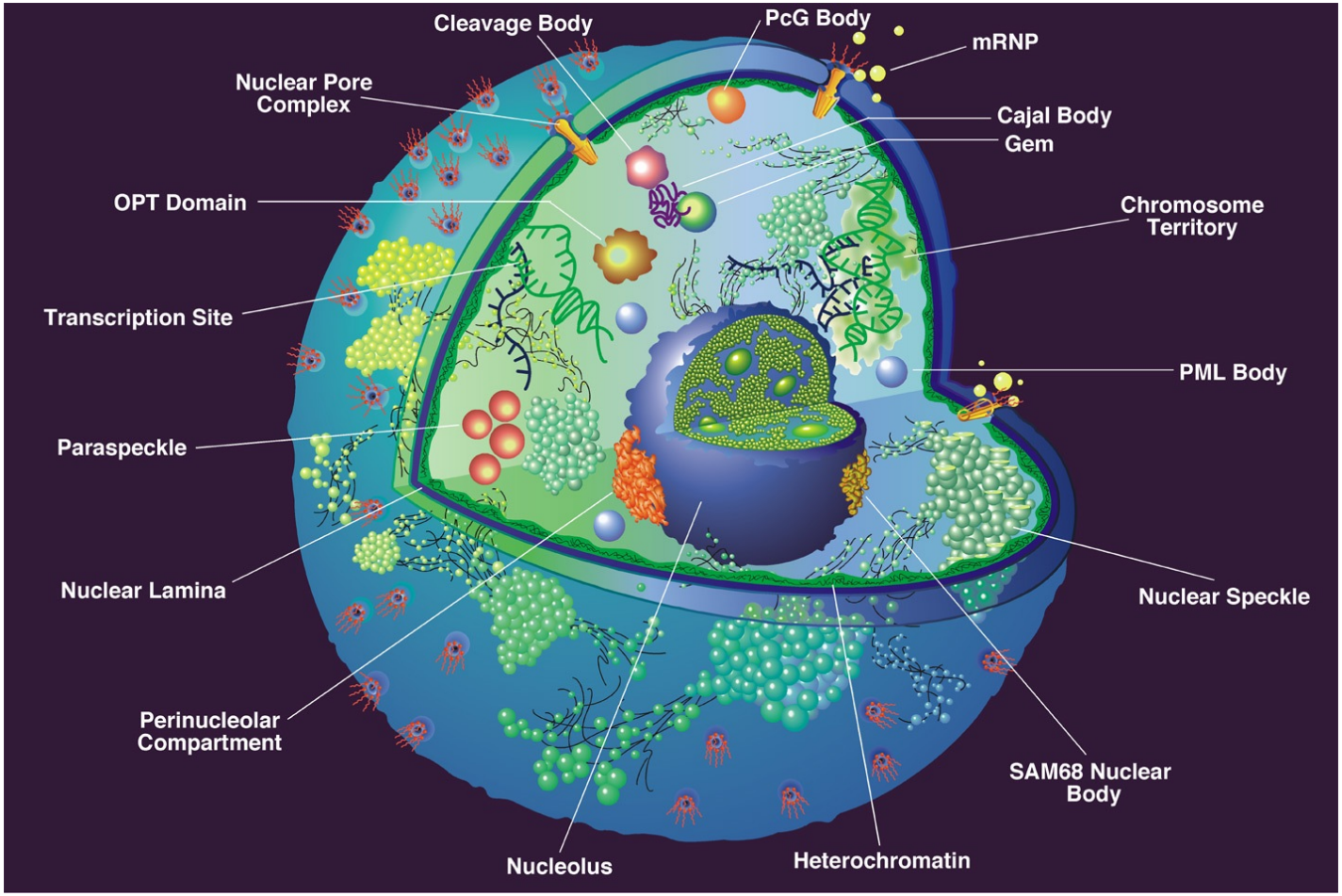
chromatin structure

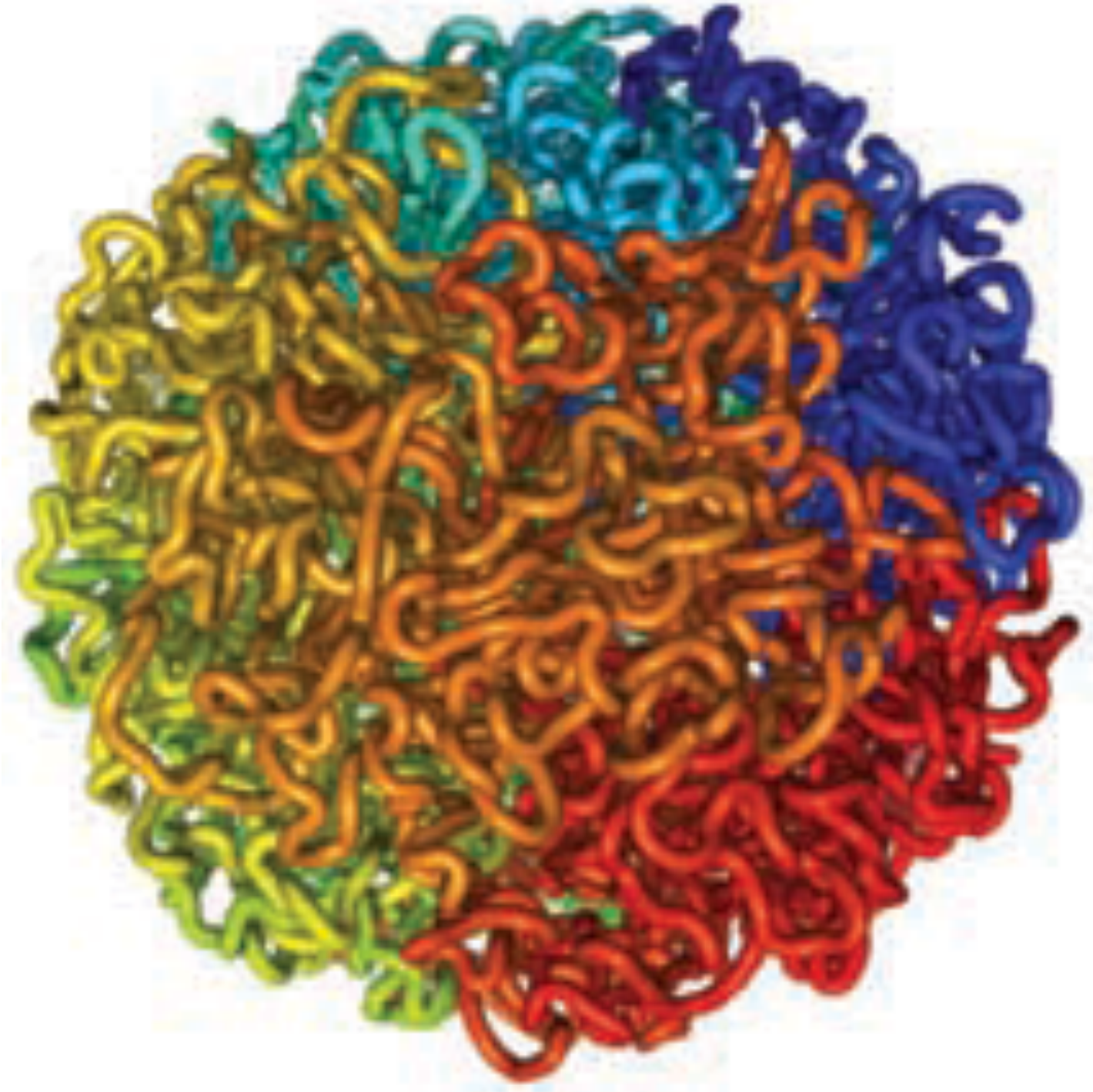




Lodish et al, MCB; Horng Science 2017

nuclear bodies



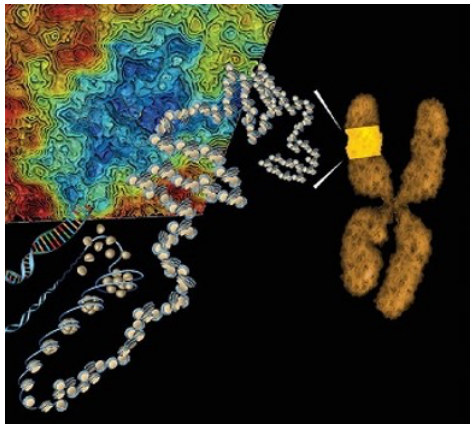


Lieberman-Aiden et al. Science 2009.

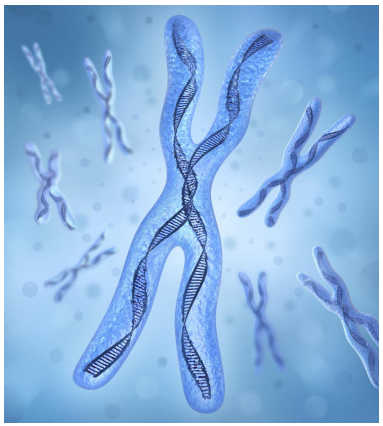
4D Nucleome

https://commonfund.nih.gov/4dnucleome

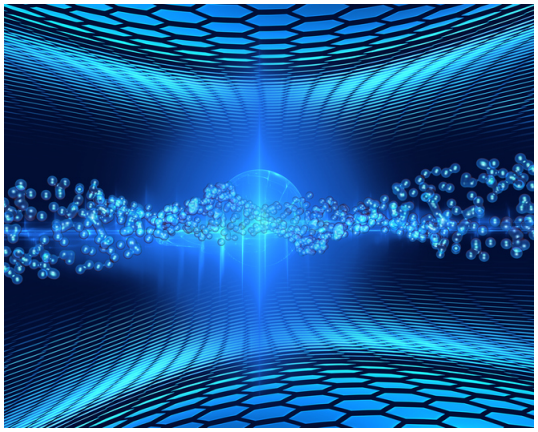
visualizing the nucleus in 3D



genes in motion



genome organization



genome organization in development