Epigenetics introduction

Mikhail Dozmorov

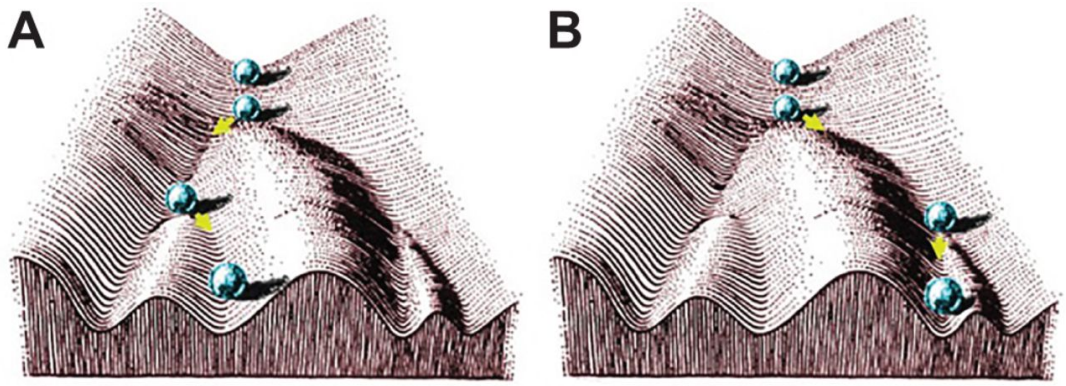
2021-04-07

Genome is more than genes

- One way to expand the regulatory capacity of mammalian genomes has been the addition of regulatory elements located at a distance from their target genes.
- Functionally, these elements can be subdivided into three categories:
  1. **Enhancers**, which provide an activating influence on transcription.
  2. **Silencers**, providing a repressive influence.
  3. **Insulators**, which prevent the transmission of regulatory influence between genomic regions.
- The regulatory elements greatly outnumber promoters and have, therefore, been proposed as a driving force behind transcriptional regulation
- These elements may have overlapping functions, they may play different roles in different cell types or developmental stages and they may overlap with promoters and genes

Epigenetics

- **Epigenetics** (*epi* - "over, on top of" + genetic) is the study of (heritable) changes in gene function that occur without a change in the sequence of the DNA.
- Critical for development and differentiation
- Epigenetic dysregulation is involved in developmental diseases, cancers, certain infectious diseases, and more



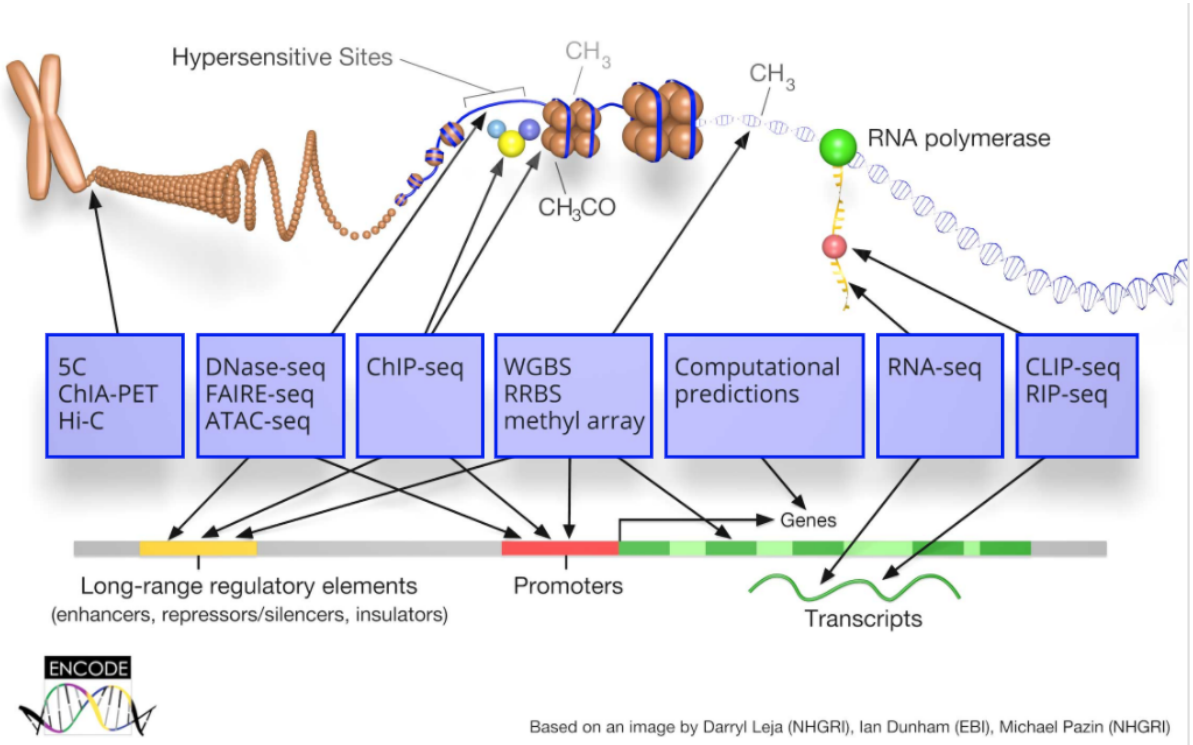
Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characters. Nature 150(3811):563–565.

ENCODE: Encylopedia Of DNA Elements

- Grand question: Where are the promoter, enhancer, and other regulatory regions of the human genome?
- 14 manually chosen and 30 randomly selected human genomic regions, in total ~30Mb (1%) of the human genome sequence
- Dozens of labs did ChIP-seq, under rigorous quality guidelines, for over 100 transcription factors and histone modifications, plus related assays for DNA methylation, chromatin accessibility etc.
- Published 14 June 2007, Nature, over 100 authors. "An integrated encyclopedia of DNA elements in the human genome."

Nature. 2012 Sep 6;489(7414):57-74. https://www.nature.com/articles/nature11247

The Encyclopedia of DNA Elements Project

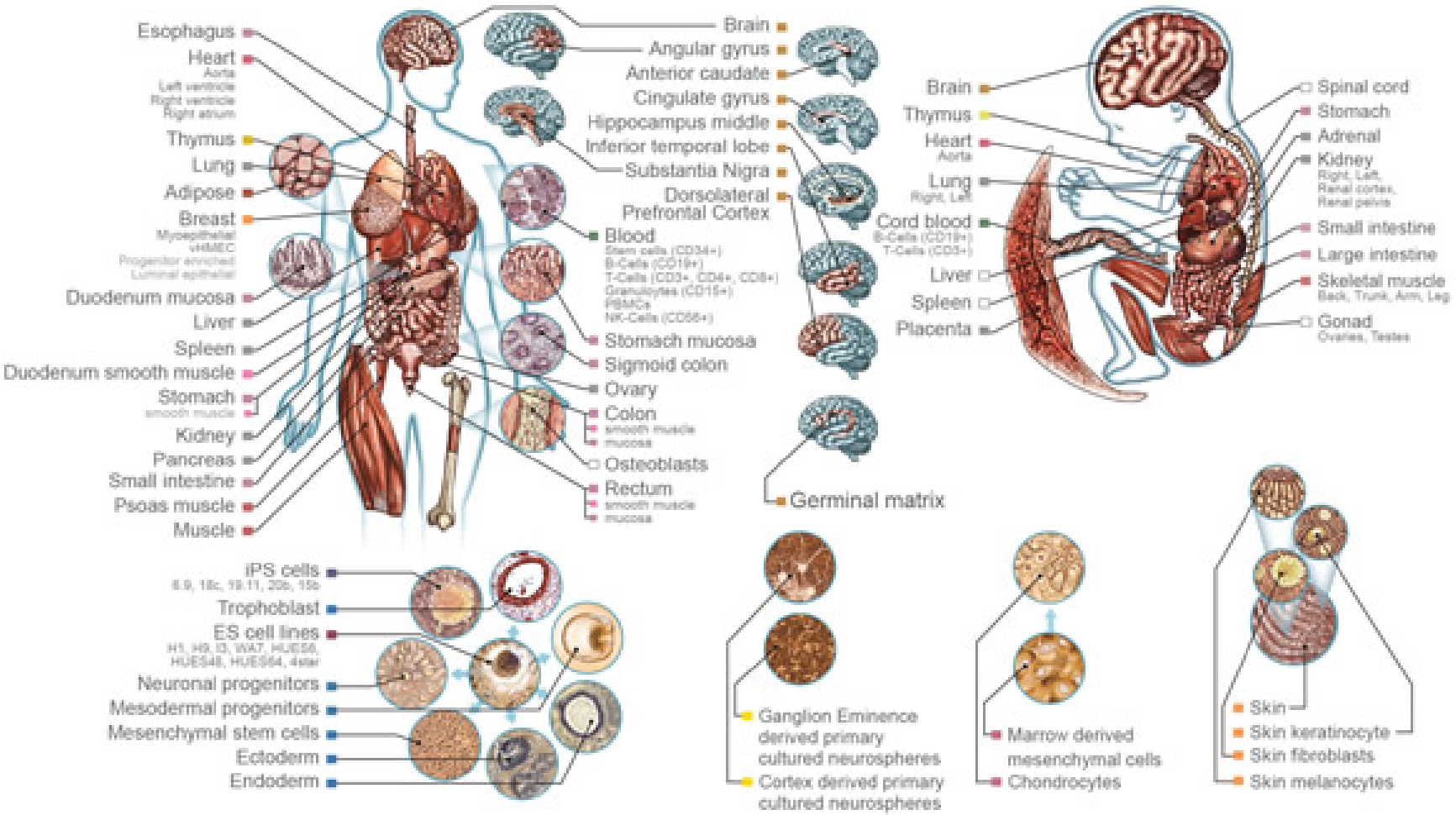


The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at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

Get Started

https://www.encodeproject.org/

Roadmap Epigenomics: chromatin regulation in normal cells



http://www.roadmapepigenomics.org/

IHEC

- Nowadays, the most recent and complete resource is IHEC, the International Human Epigenome Consortium
- International effort with several funding agencies
- Goal: Providing standardized reference epigenomes for a variety of normal and disease tissues
- Member groups take part in committees working on standards (assays, data/metadata distribution, ethics. . .)

http://ihec-epigenomes.org/

IHEC Data Portal

- Goal: Integrate epigenomic public datasets produced within the International Human Epigenome Consortium
- Raw data is in controlled access repositories
- Over 8,000 human datasets
- Datasets from 7 consortia, others coming
- Offers tools for datasets discovery, visualization and pre-analysis

http://epigenomesportal.ca/ihec/

IHEC Data Portal

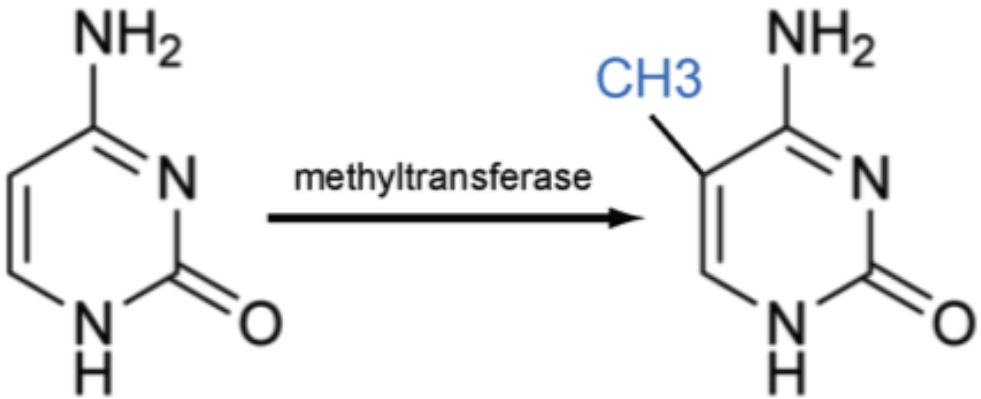
- Publicly accessible datasets
  - Datasets made available in the IHEC Data Portal is publicly accessible for everyone's own research
- Human data offered by such consortia usually falls in one of two categories:
  - Controlled access data
    - Raw data from sequencers
    - Clinical/sensitive information such as phenotypes
    - Archived at repositories such as EGA and dbGaP
  - Public data
    - Annotation tracks, to use in tools such as UCSC Genome Browser, Ensembl and IGV.
    - Some donor, sample and library metadata

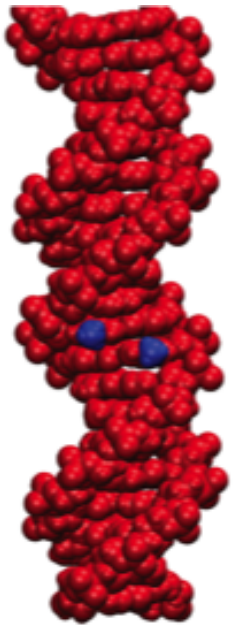
Functional Genomics

- **Functional genomics** is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and *interactions*.

DNA methylation

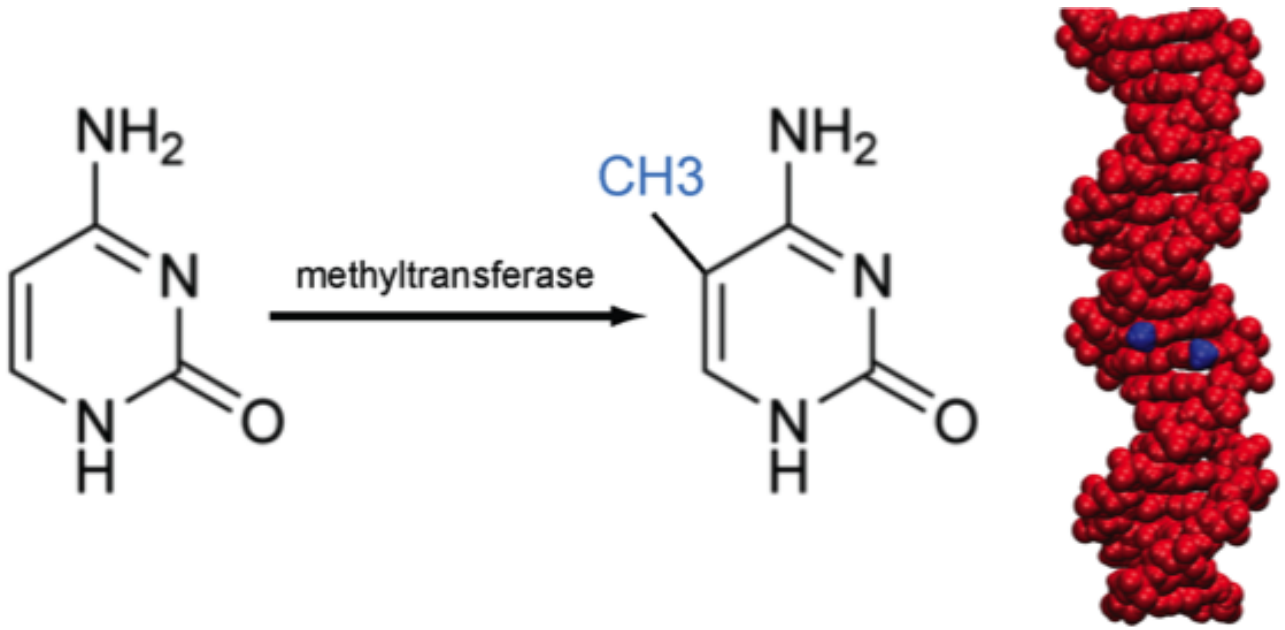
- DNA methylation is a type of chemical modification of DNA which involves the addition of a methyl group to the number 5 carbon of the cytosine (5C), to convert cytosine to 5-methylcytosine (5mC).
- The most well characterized epigenetic mechanism.
- In humans, DNA methylation occurs in cytosines that precede guanines (hence, CpG)
- Invertebrates (Drosophila, yeast) do not exhibit cytosine methylation





DNA modifications

- Other nucleotides (Adenine, Guanine, Thymine) can also be modified
- Mainly in bacteria, but genomes of eukaryotes may contain base modifications on bases other than cytosine, such as methylated adenine or guanine



Sood AJ, Viner C, Hoffman MM. 2016. "**DNAmod: the DNA modification database.**" bioRxiv 071712. https://www.pmgenomics.ca/hoffmanlab/proj/dnamod/

CpG Sites and CpG islands

- CpG sites are not randomly distributed in the genome - the frequency of CpG sites in human genomes is 1% (~28 million CpGs), which is less than the expected (~4-6%).
- Around 60-90% of CpGs are methylated in mammals.
- DNA methylation frequently occurs in repeated sequences, and may help to suppress transcription from these sequences, and aid chromosomal stability.

CpG Sites and CpG islands

- There are regions of the DNA that have a higher concentration of CpG sites ($> 60\%$), named the CpG islands, which tend to be located in the promoter regions of many genes.
- Between 200-1000 bp in length
- Usually not methylated.
- Less than 10% of CpGs occur in CG-dense regions

non-CG methylation

- Embryonic stem cells have ~25% of non-CG methylation (mCHG and mCHH, where H=A, C, T).
- non-CG methylation correlates with gene expression in ESCs.
- non-CG methylation is on anti-sense strand of gene bodies, and correlates with increased *intronic* transcription in ESCs.
- non-CG methylation is depleted in ehancers in ESCs.

Lister, Ryan, Mattia Pelizzola, Robert H. Dowen, R. David Hawkins, Gary Hon, Julian Tonti-Filippini, Joseph R. Nery, et al. "Human DNA Methylomes at Base Resolution Show Widespread Epigenomic Differences." Nature 462, no. 7271 (November 19, 2009): 315–22. https://doi.org/10.1038/nature08514.

Creation and maintenance of DNA methylation

- In humans, DNA is methylated by three enzymes, DNA methyltransferase DNMT1, DNMT3a, DNMT3b.
- DNMT1 is the maintenance methyltransferase that is responsible for copying DNA methylation patterns to the daughter strands during DNA replication.
- DNMT3a and 3b are the *de novo* methyltransferases that, in combination with DNMT3L, set up DNA methylation patterns early in development.

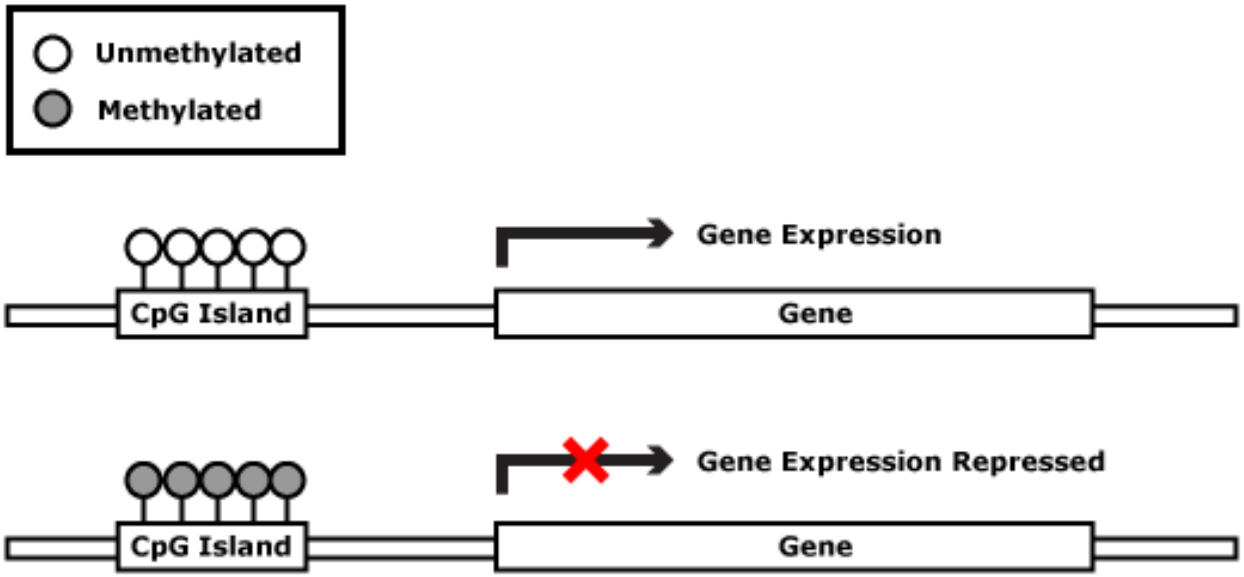
Removal of DNA methylation

- Loss of 5mC can be achieved **passively** by dilution during replication or exclusion of DNMT1 from the nucleus.
- Ten-eleven translocation (TET) family of proteins can **actively** convert 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC) in vertebrates - demethylation
- Iterative oxidations of 5hmC catalysed by TET result in 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5caC mark is excised from DNA by G/T mismatch-specific thymine-DNA glycosylase (TDG), which as a result returns cytosine residue back to its unmodified state

Guo, Junjie U., Yijing Su, Chun Zhong, Guo-li Ming, and Hongjun Song. "Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain." Cell 145, no. 3 (April 29, 2011): 423–34. https://doi.org/10.1016/j.cell.2011.03.022.

Roles of DNA methylation

- Transcriptional gene silencing
- Maintain genome stability
- Embryonic development
- Genomic imprinting
- X chromosome inactivation (females)



http://learn.genetics.utah.edu/content/epigenetics/imprinting/

Factors associated with changes in DNA methylation

- Aging (developmental stage)
- Diet
- Inflammatory patterns
- Environmental exposures
- Smoking
- Alcohol

DNA methylation and cancer

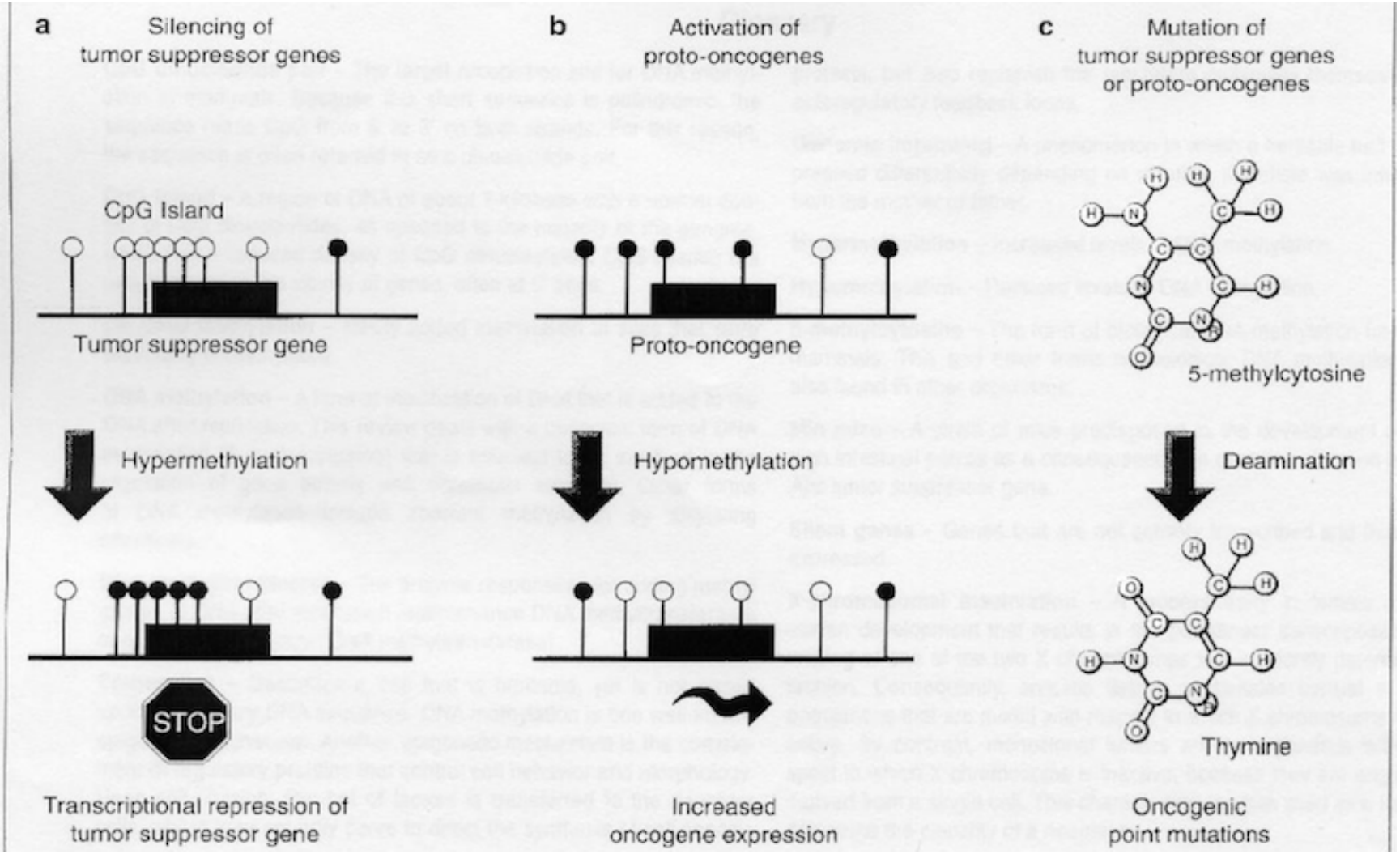
Hypomethylation – decrease methylation levels

- A lower level of DNA methylation in tumors was one of the first epigenetic alterations to be found in human cancer. (Feinberg AP, et al., 1983).
- Demethylation of the promoter region of proto-oncogenes will activate normally repressed gene expression
- Global hypomethylation of DNA sequences that are normally heavily methylated may result in:
  - Chromosomal instability
  - Increased transcription from transposable elements
  - An elevated mutation rate due to mitotic recombination

DNA hypermethylation

Hypermethylation – increase methylation levels

- Hypermethylation of the CpG islands in the promoter regions of tumor-suppressor genes is a major event in the origin of many cancers.
- Hypermethylation of promoters can inactivate tumor-suppressor genes, affect genes involved in the cell cycle, DNA repair, and the metabolism of carcinogens, all of which are involved in the development of cancer.
- The profiles of hypermethylation of the CpG islands in tumor-suppressor genes are specific to the cancer type.



Laird PW "**Oncogenic mechanisms mediated by DNA methylation.**" Mol Med Today. 1997
http://www.cell.com/moltod/pdf/S1357-4310(97)01019-8.pdf

Application of DNA methylation assays

Early diagnosis

- Detection of CpG-island hypermethylation in biological fluids and serum

Prognosis

- Hypemethylation of specific genes
- Whole DNA methylation profiles

Prediction

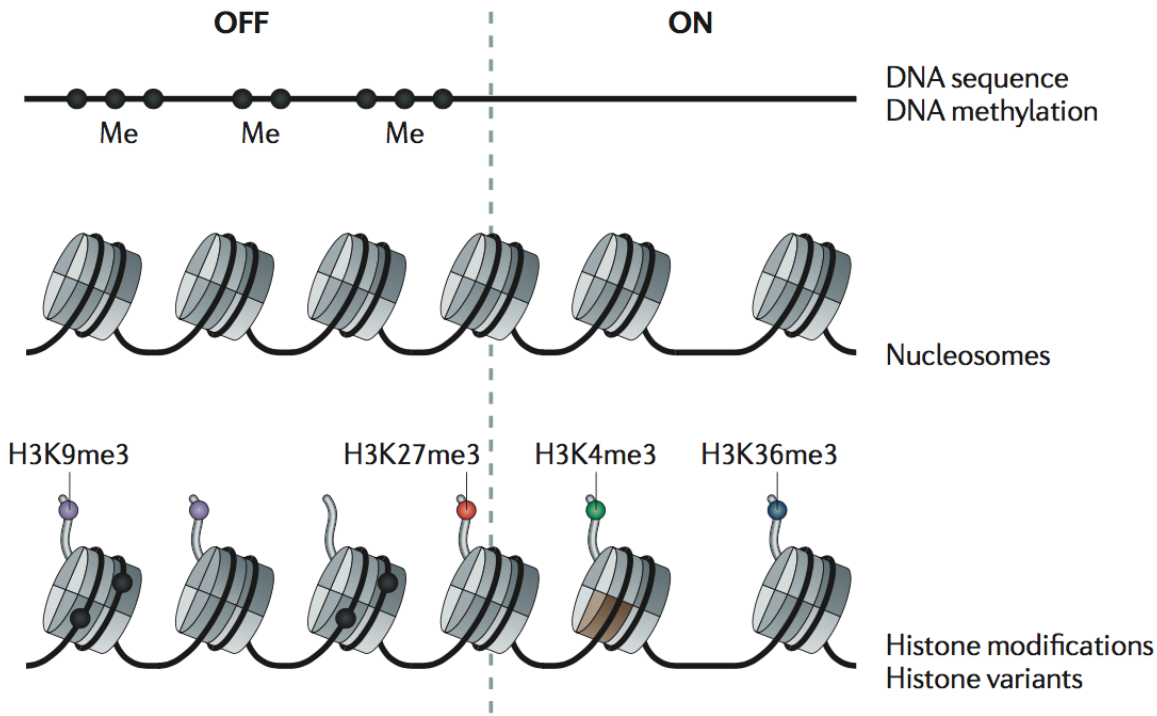
- CpG island hypermethylation as a marker of response to chemotherapy

Prevention

- Developing DNMTs inhibitors as chemopreventive drugs to reactive silenced genes

Chromatin structure

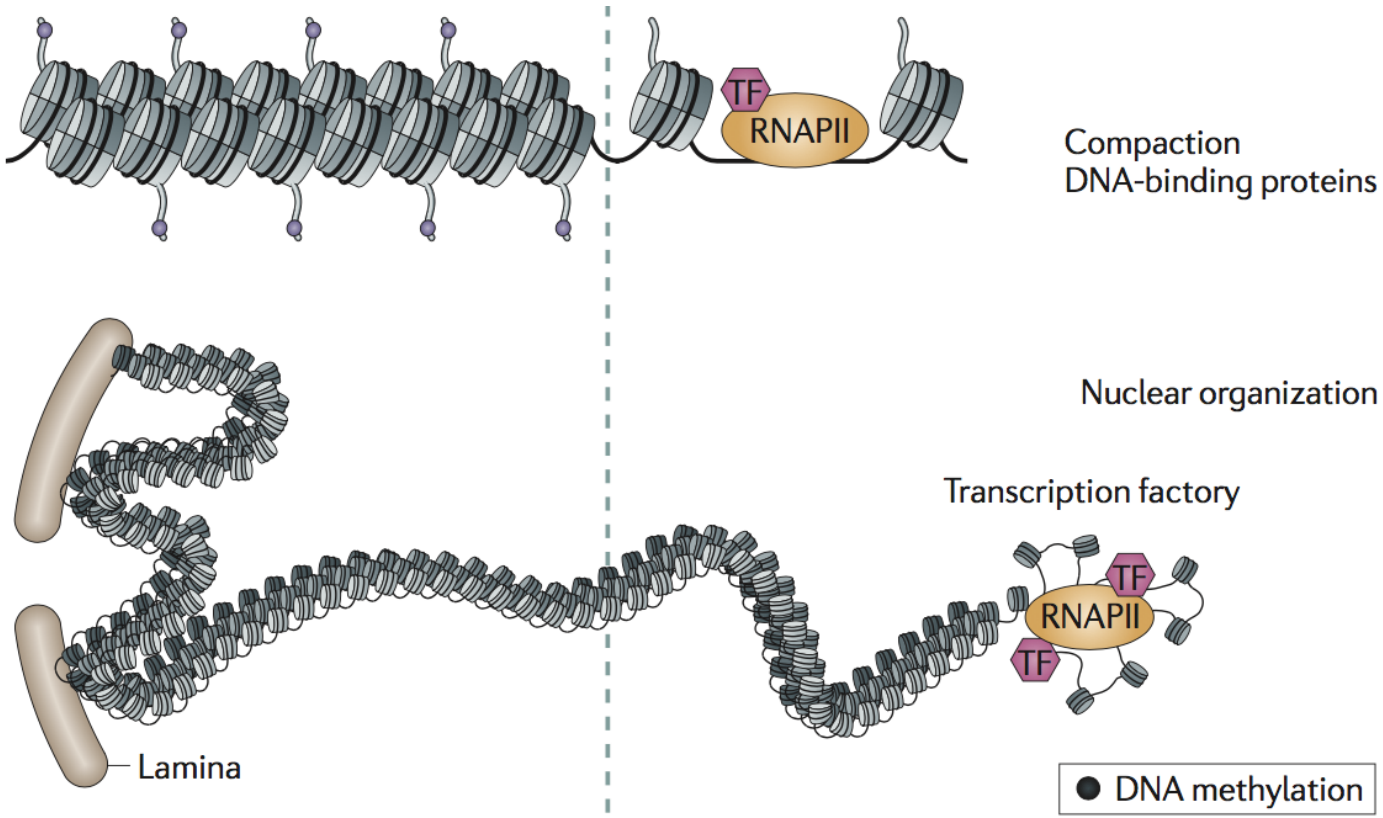
- Chromatin organization has multiple structural layers and organizes chromatin into "domains"
- Both DNA methylation and chromatin marks contain important functional information



https://www.ncbi.nlm.nih.gov/pubmed/21116306

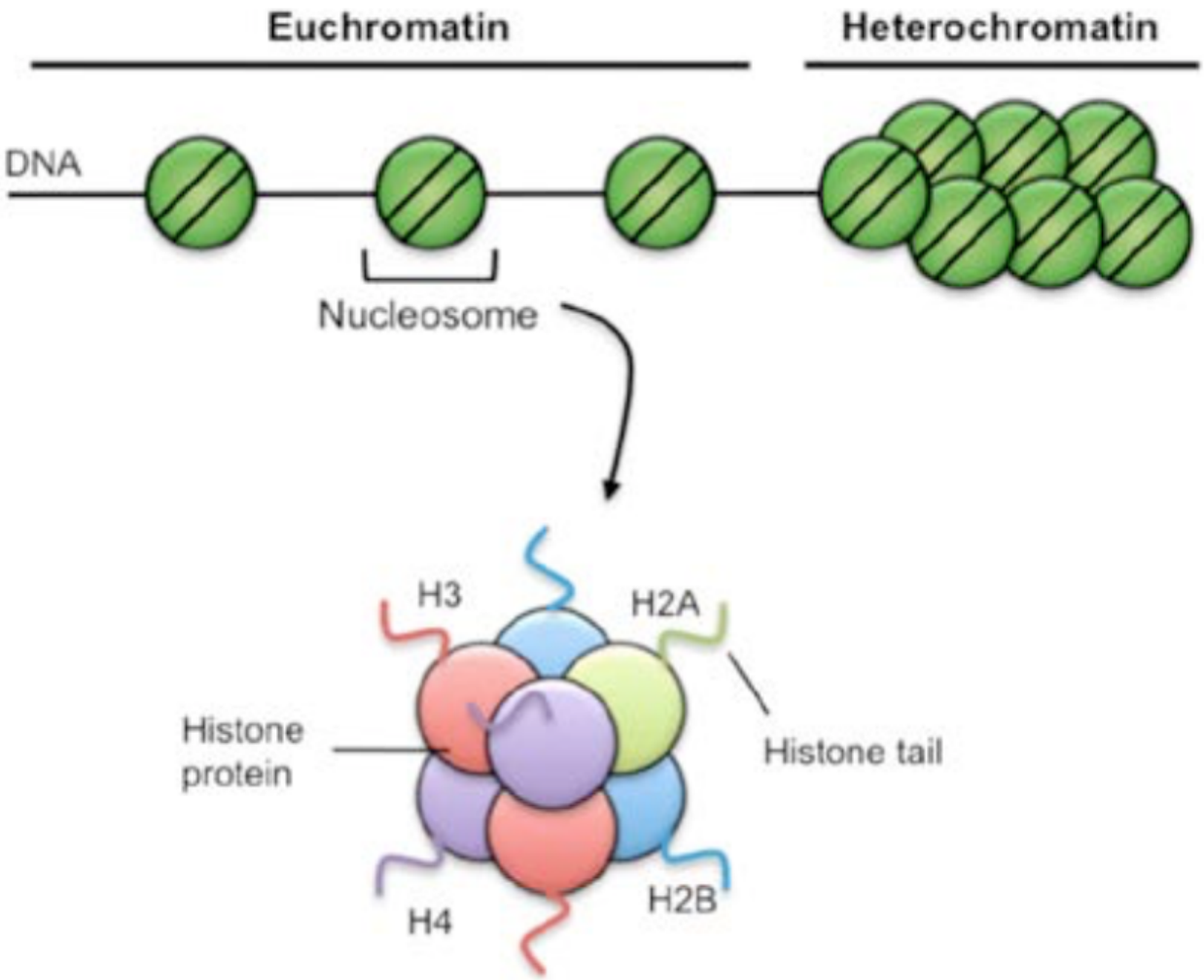
Chromatin structure

- Open chromatin - actively transcribed genes
- Heterochromatin - silent, non-transcriptionally active DNA

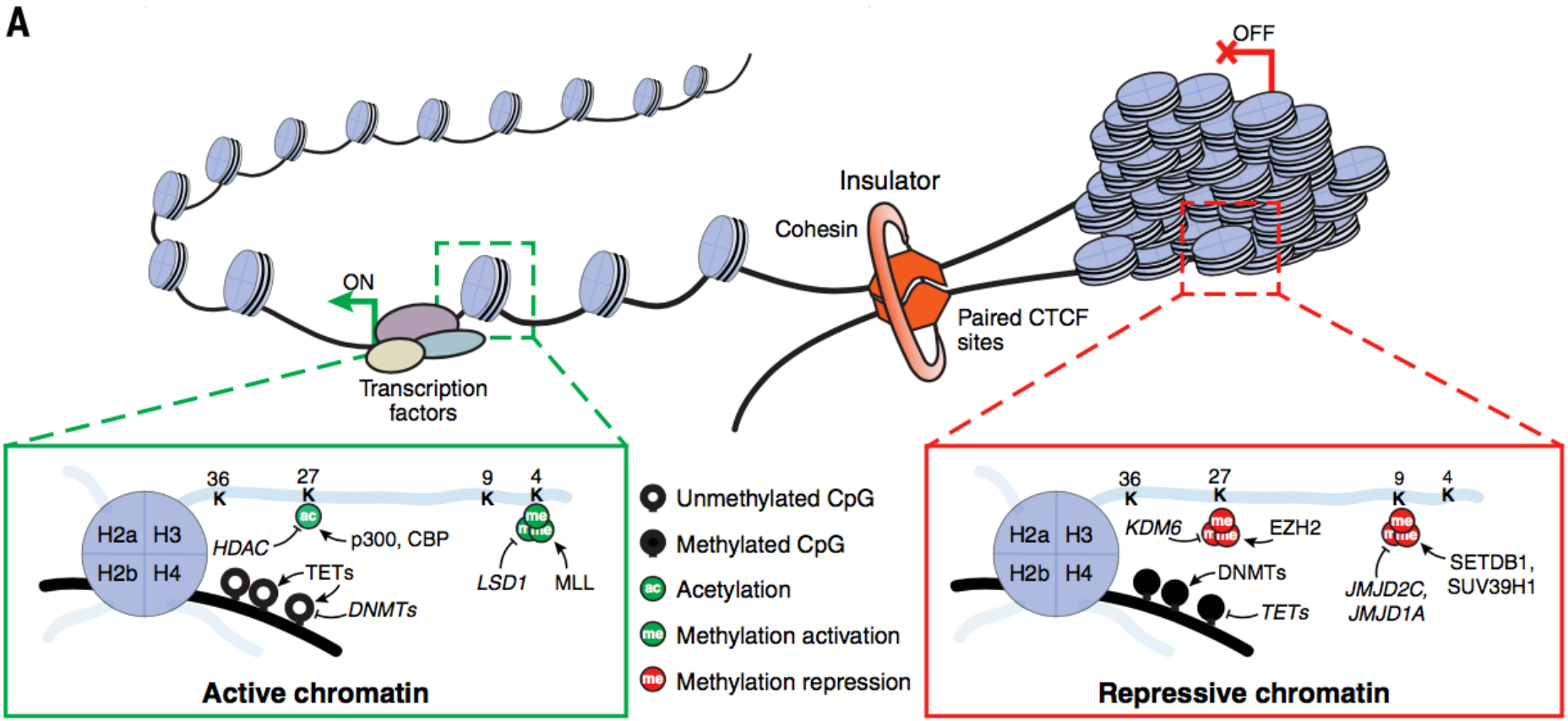


https://www.ncbi.nlm.nih.gov/pubmed/21116306

Active and inactive chromatin

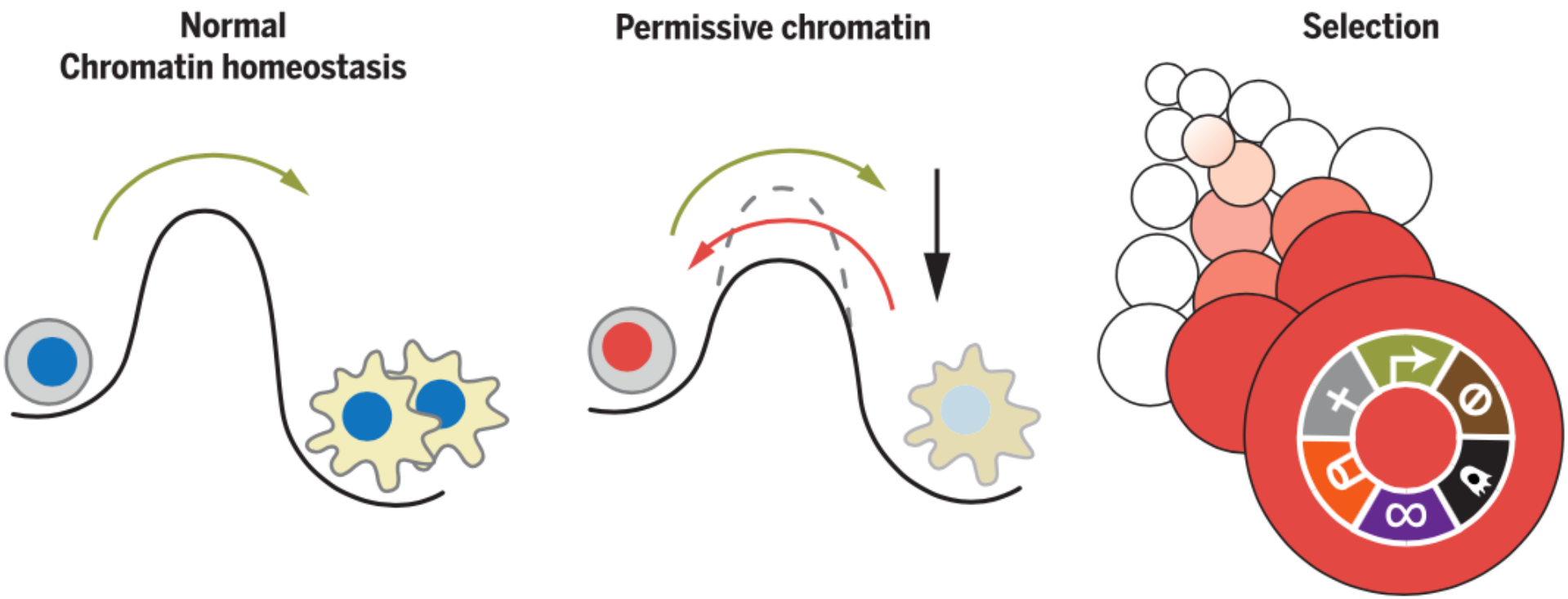


Chromatin structure affects cellular identity and state transitions



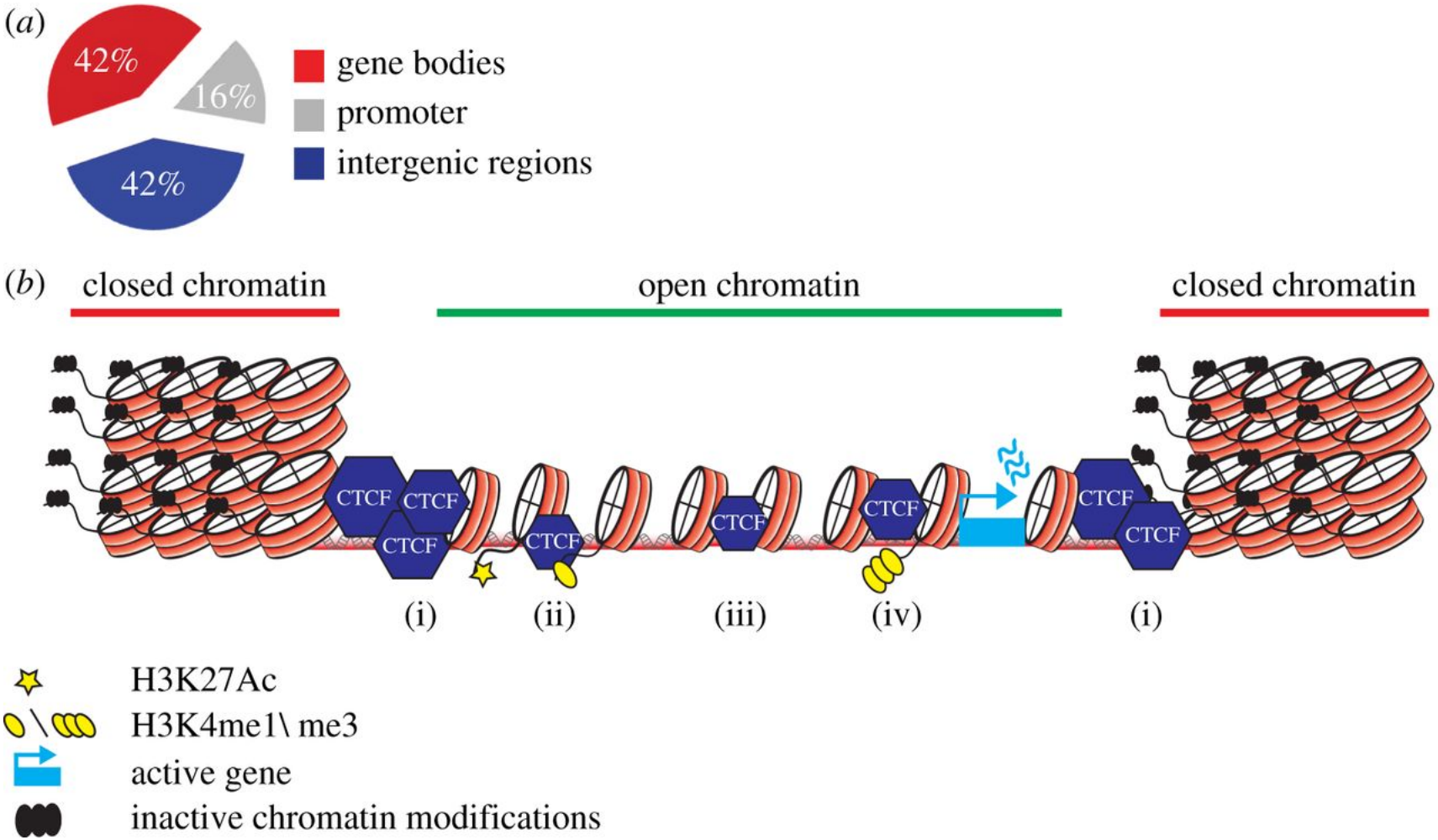
http://science.sciencemag.org/content/357/6348/eaal2380

Role of chromatin structure in cancer



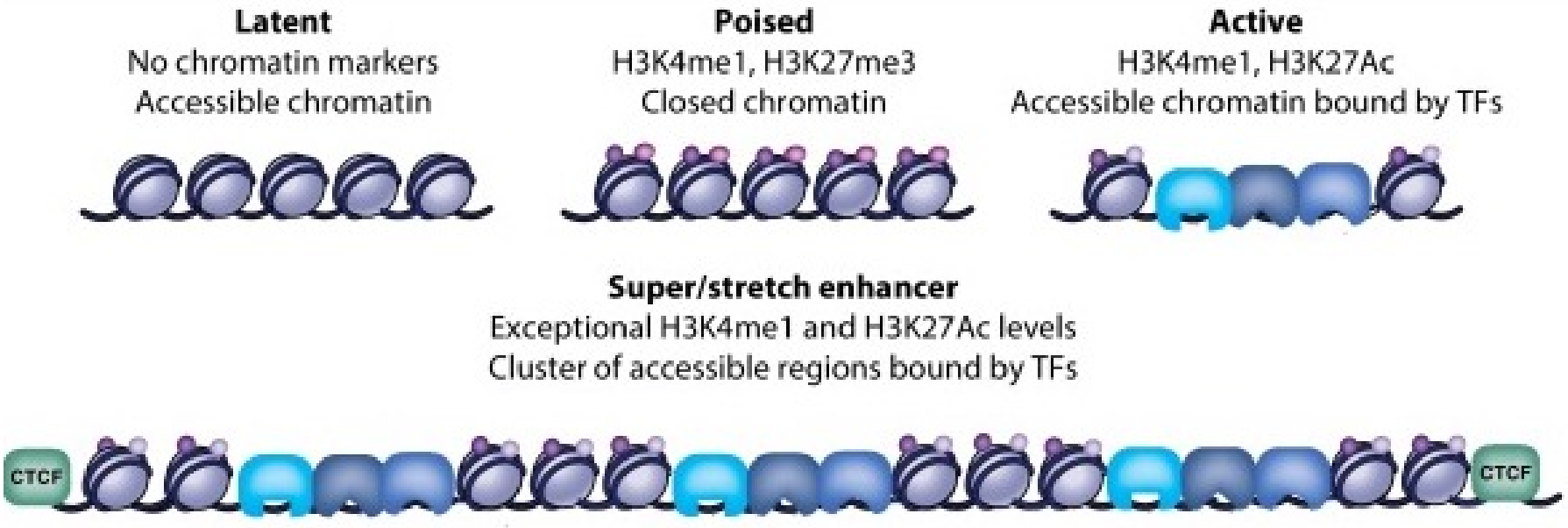
http://science.sciencemag.org/content/357/6348/eaal2380

Insulators/boundaries of open/closed chromatin



http://rstb.royalsocietypublishing.org/content/368/1620/20120369

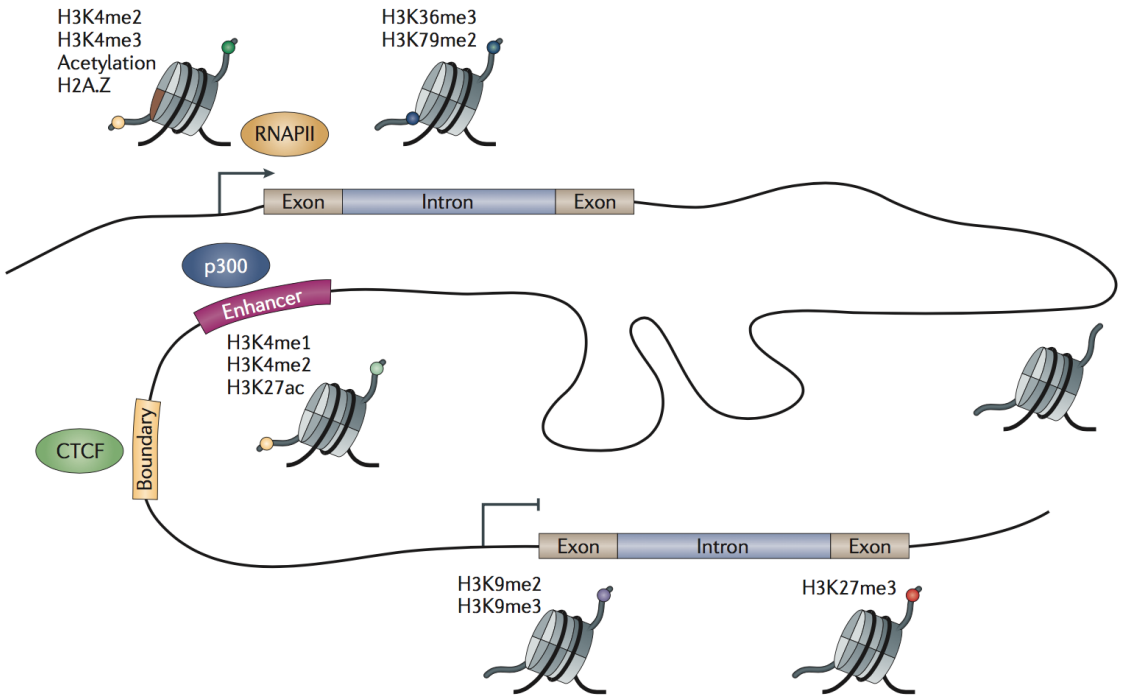
Super-enhancers



http://www.cell.com/trends/immunology/fulltext/S1471-4906(15)00172-6

Histone code

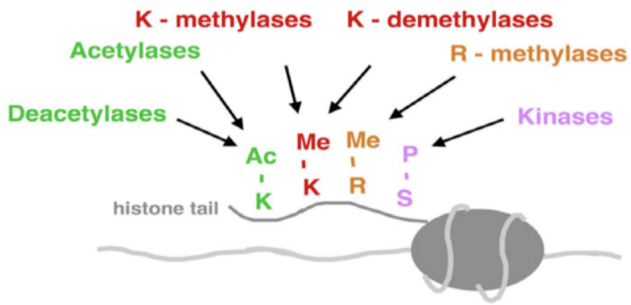
- Histone modifications demarcate functional elements in mammalian genomes.
- Can be broadly divided into activating and repressive histone marks.



https://www.ncbi.nlm.nih.gov/pubmed/21116306

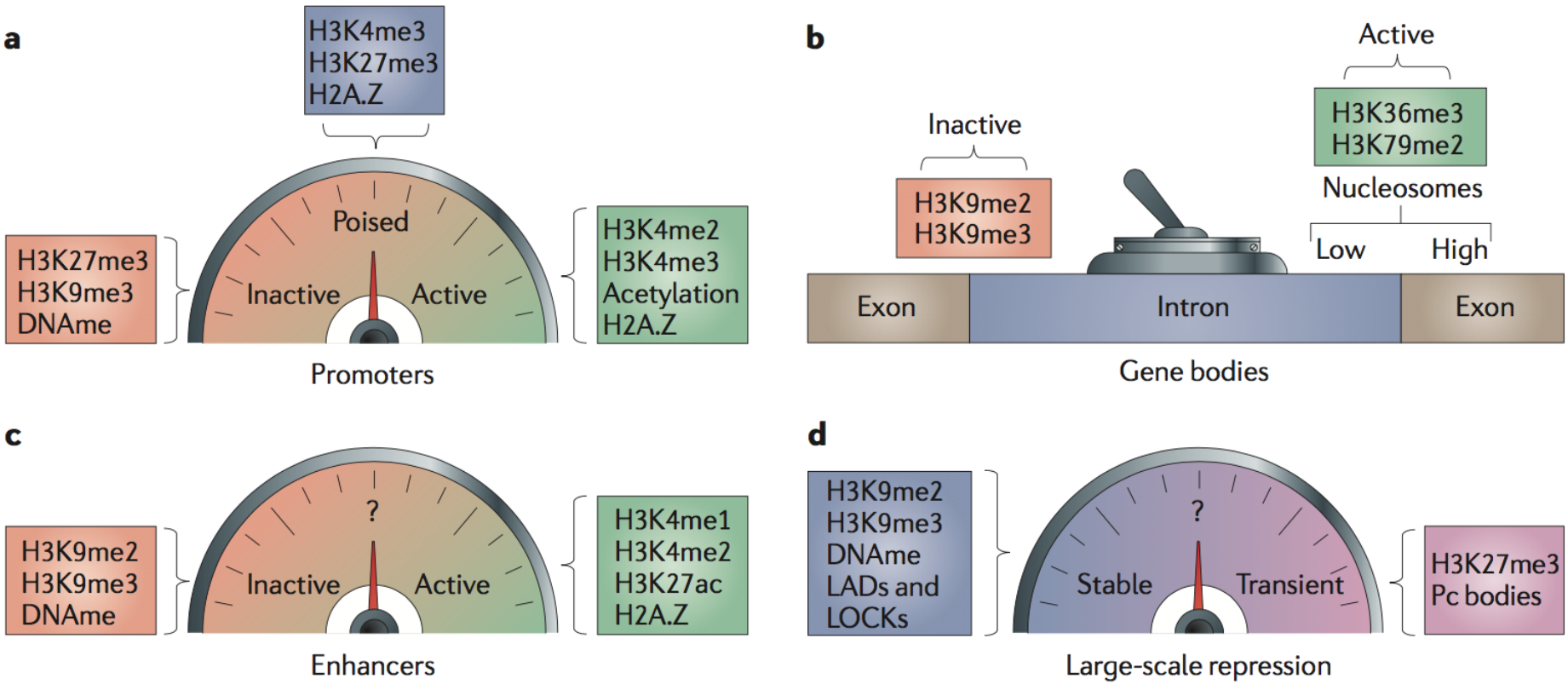
Histone modifications

- Methylation of lysine residuals
- Methylation of arginine residuals
- Acetylation,
- Ubiquitination,
- ADP-ribosylation
- Phospohrylation
- Mono/di/tri methylation



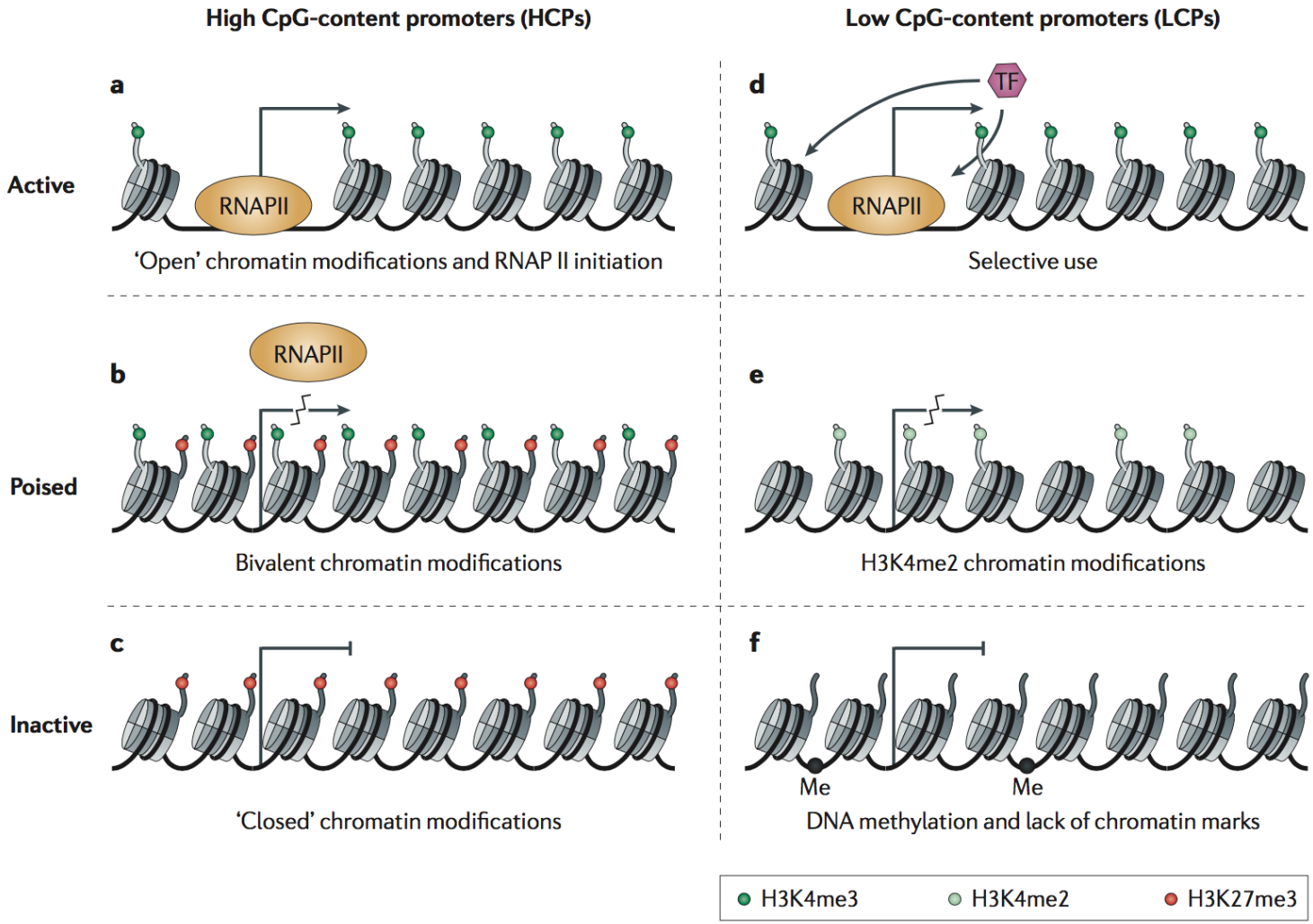
https://www.ncbi.nlm.nih.gov/pubmed/17320508

Histone code summary



https://www.ncbi.nlm.nih.gov/pubmed/21116306

Methylation and histone code interplay



https://www.ncbi.nlm.nih.gov/pubmed/21116306