Epigenetics introduction

Mikhail Dozmorov

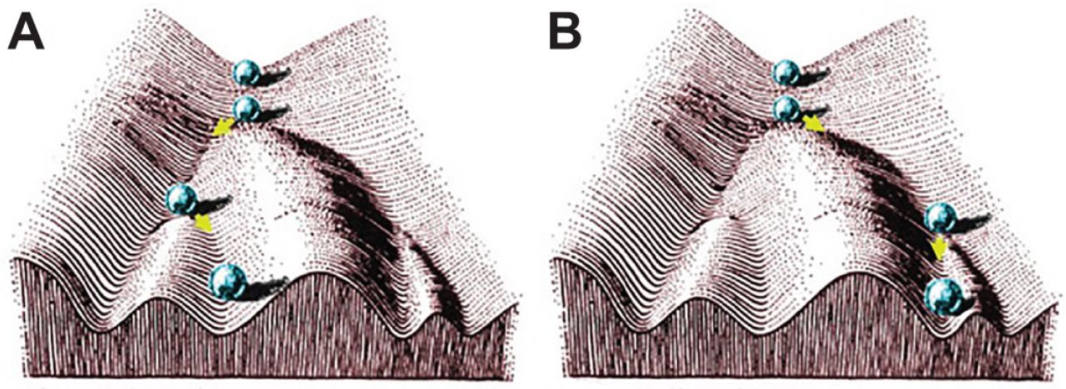
Spring 2018

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. **Enhancer sequences**, 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



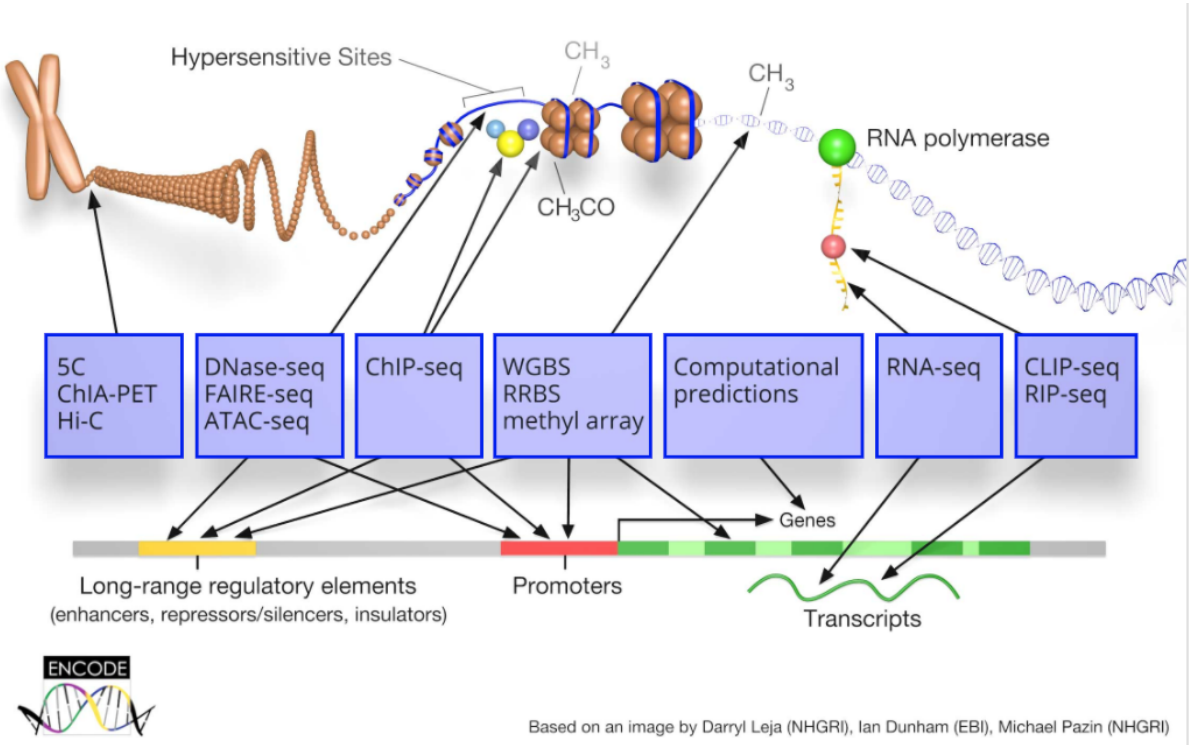
Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characters. Nature 150(3811):563–565.

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

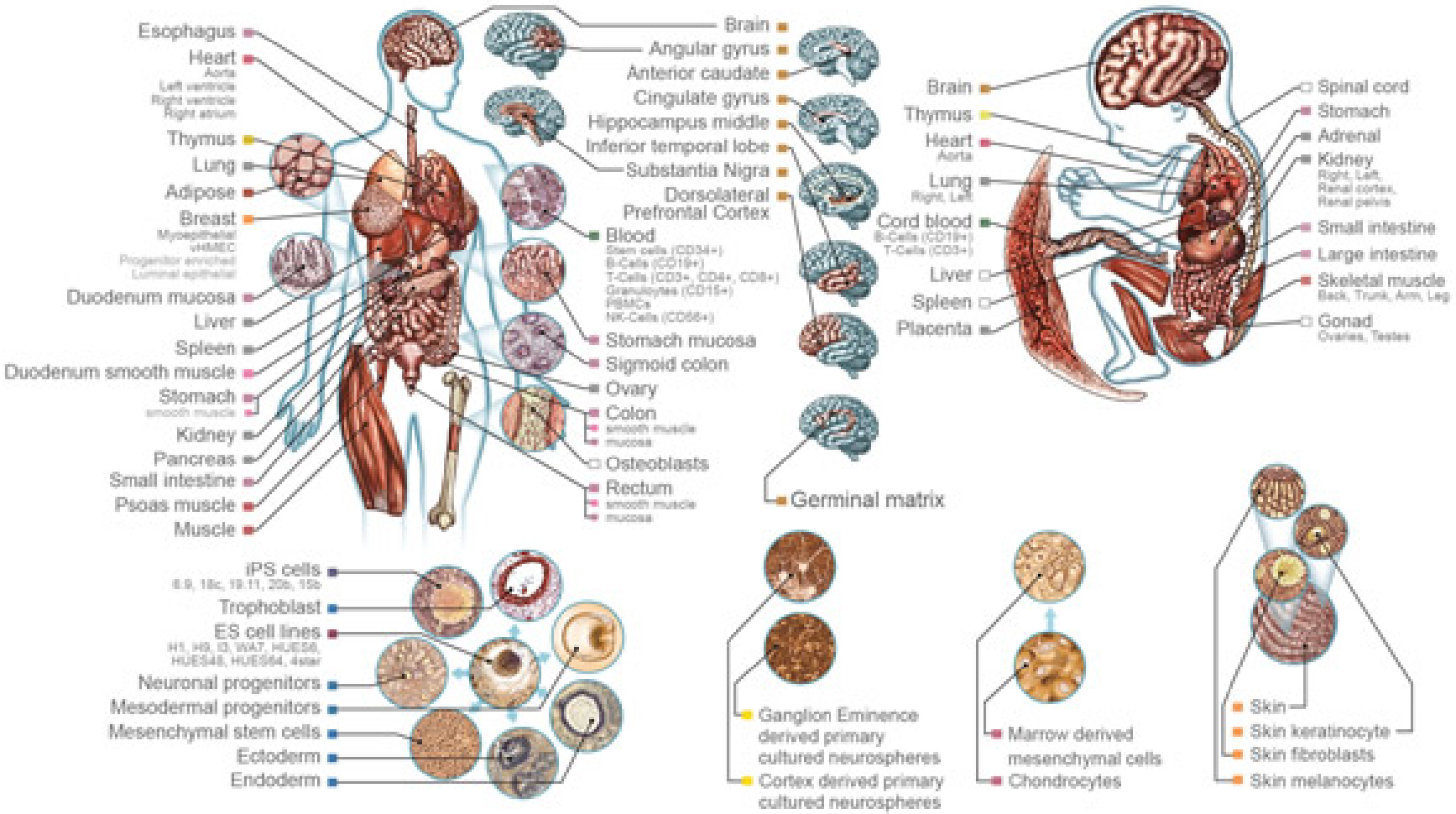


The ENCODE (Encyclopedia of DNA Elements) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at 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

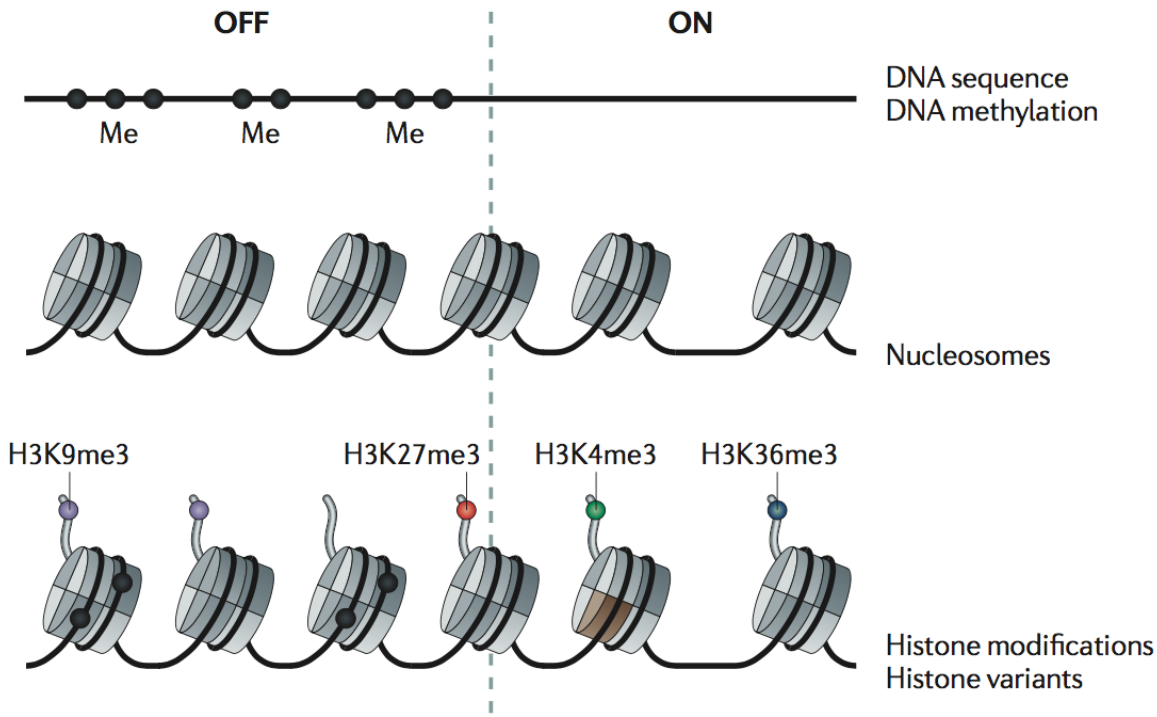
- Publicly accessible datasets
  - Datasets made available in the IHEC Data Portal is publicly accessible for everyone's own research
- Human data offered by such consortia usually falls in one of two categories:
  - Controlled access data
    - Raw data from sequencers
    - Clinical/sensitive information 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*.

Chromatin structure

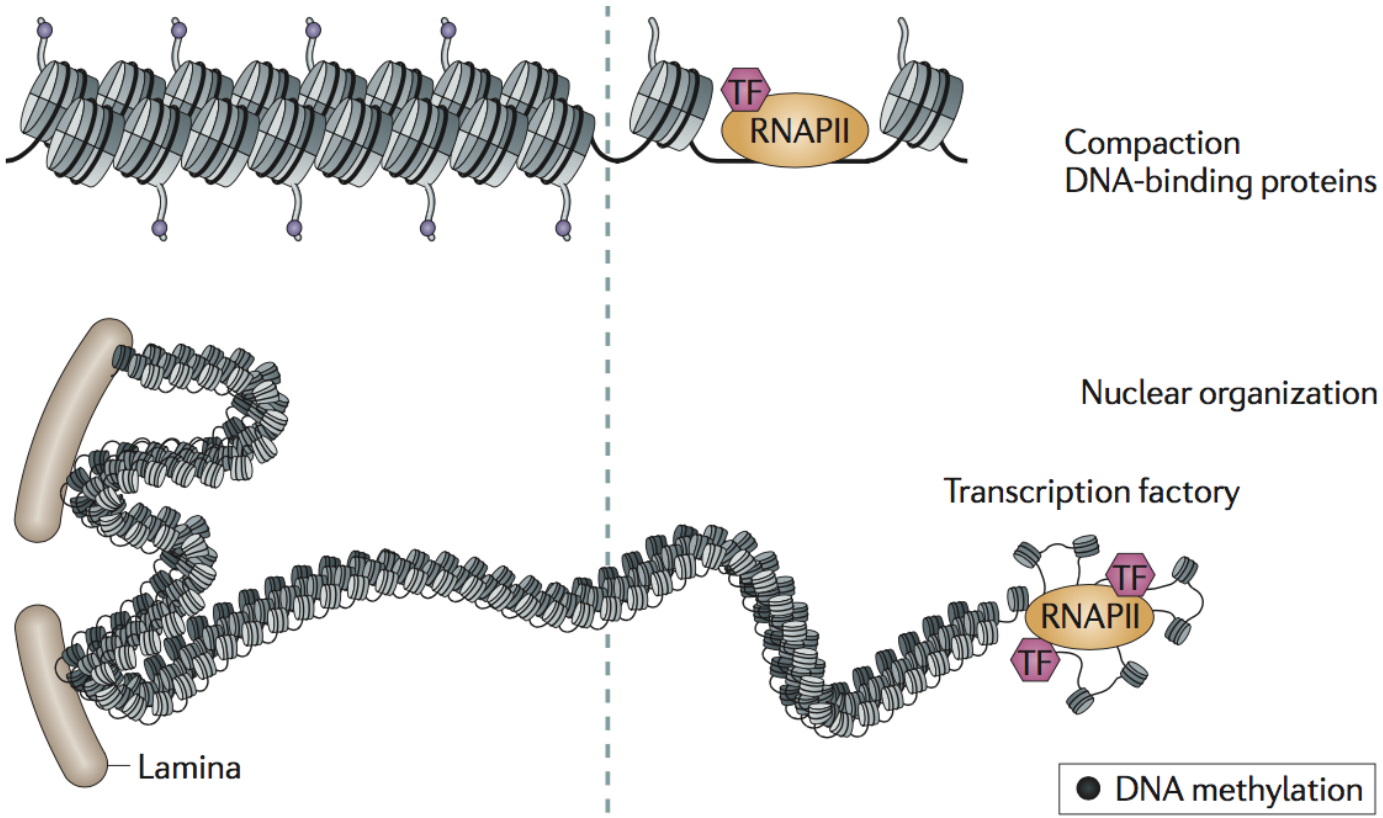
- Chromatin organization has multiple structural layers and organizes chromatin into "domains"
- Both DNA methylation and chromatin marks contain important functional information



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

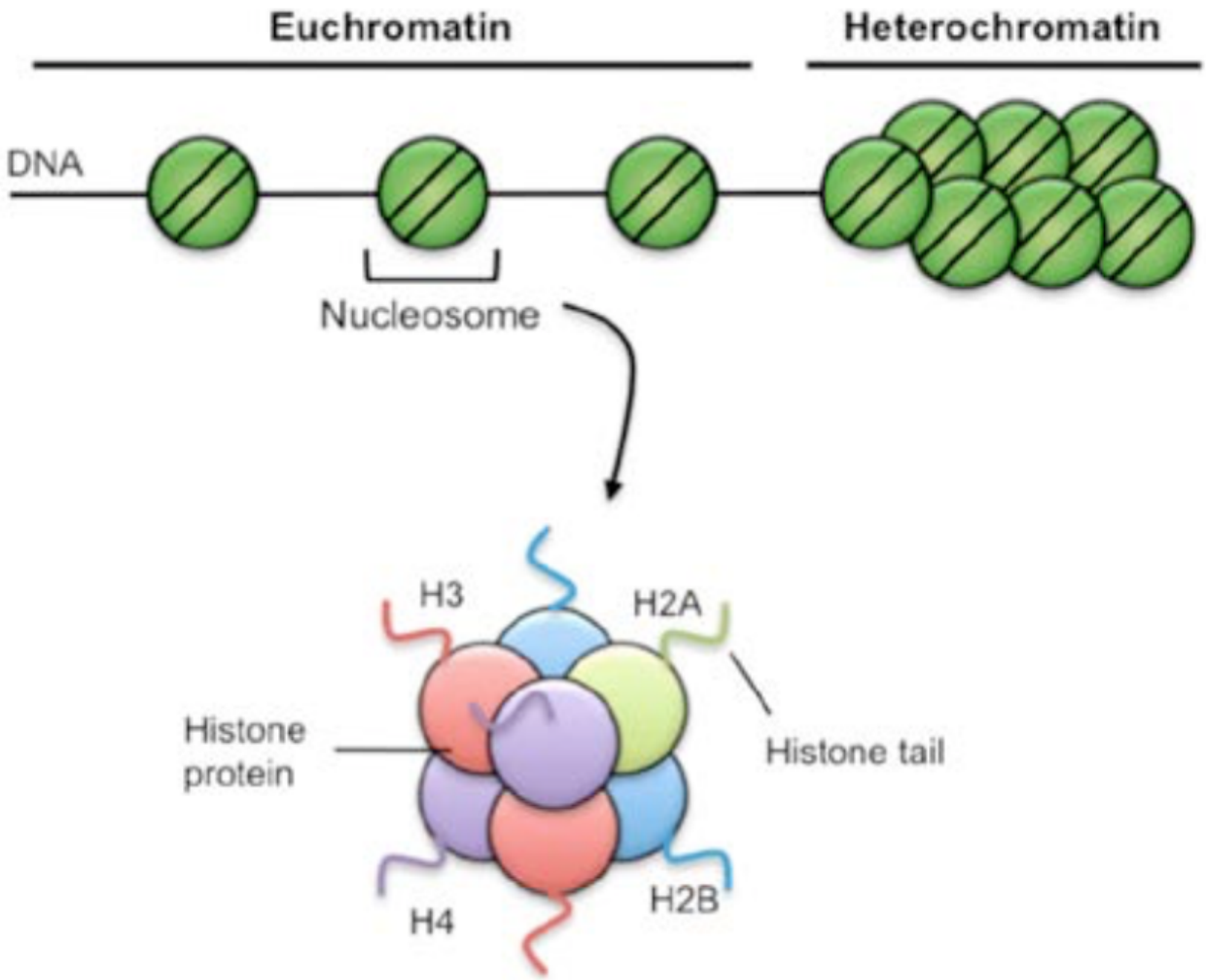
Chromatin structure

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

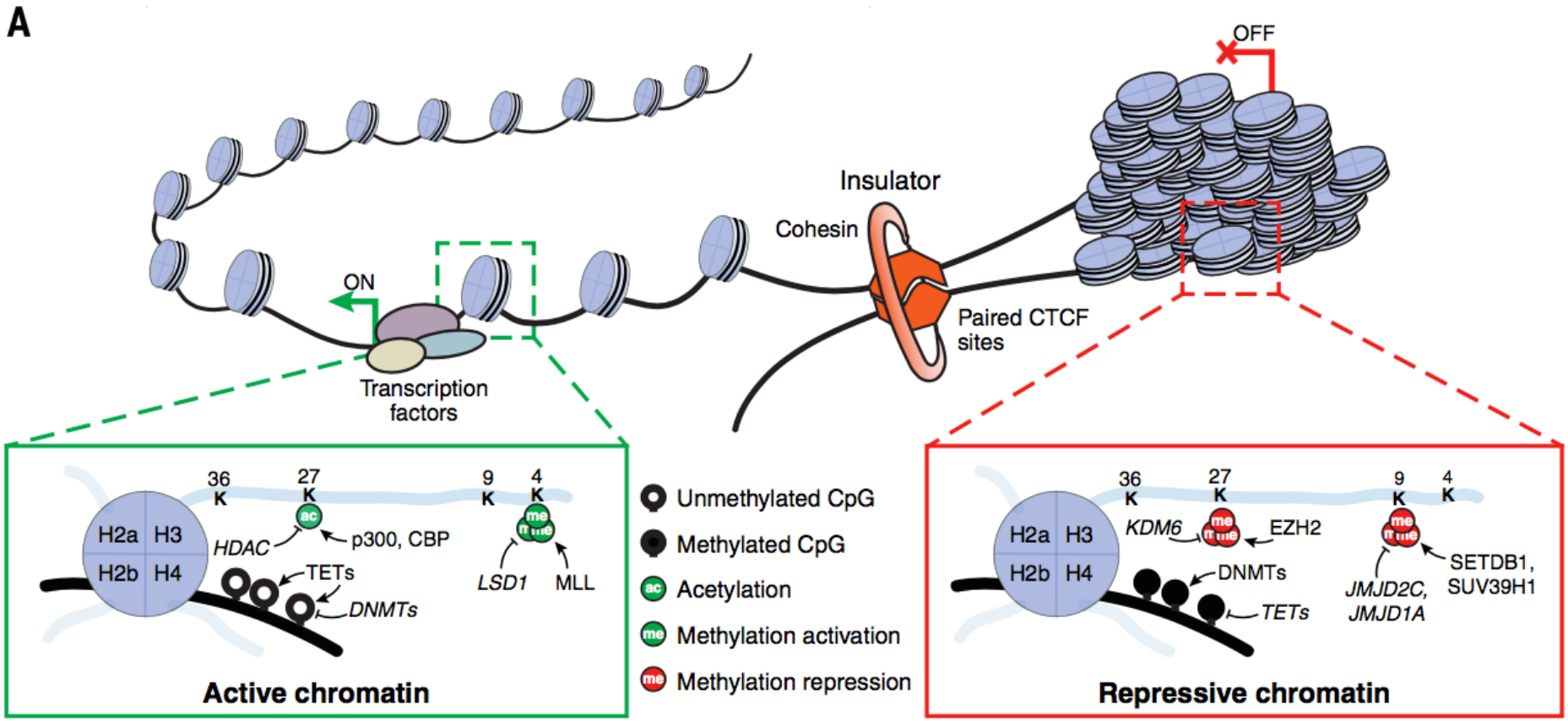


https://www.ncbi.nlm.nih.gov/pubmed/21116306

Active and inactive chromatin

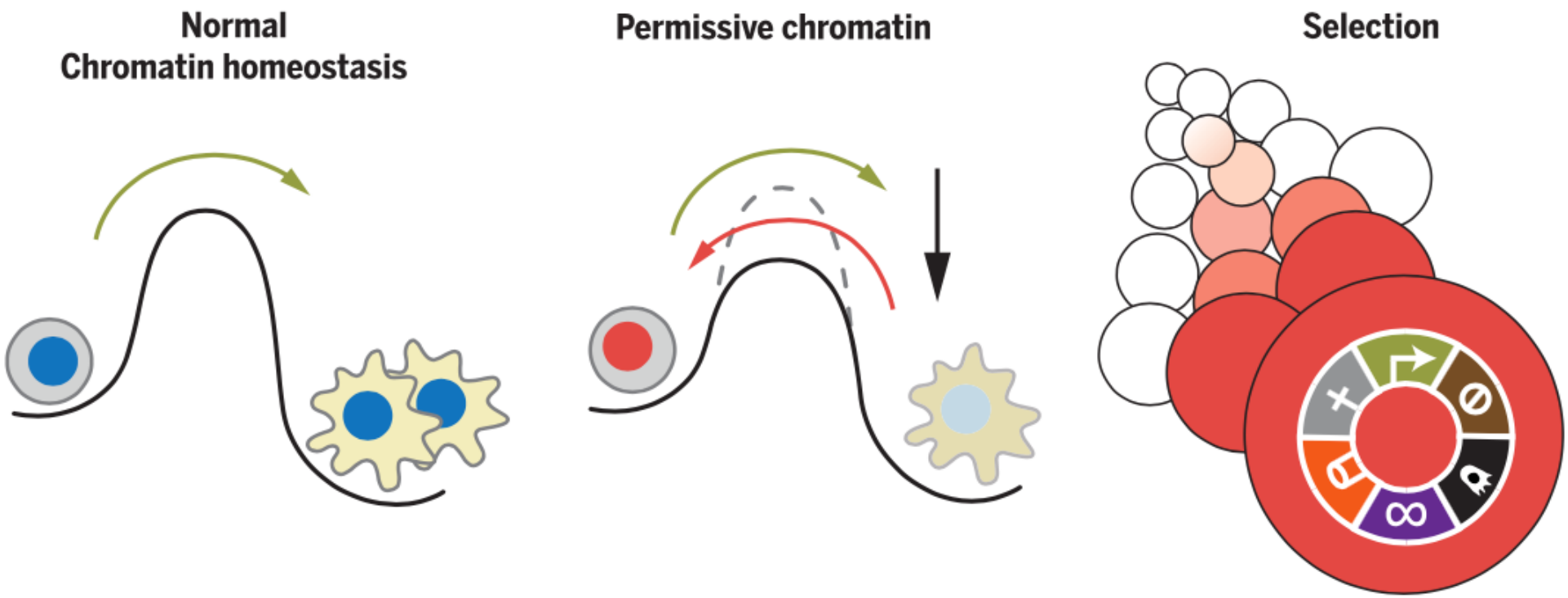


Chromatin structure affects cellular identity and state transitions



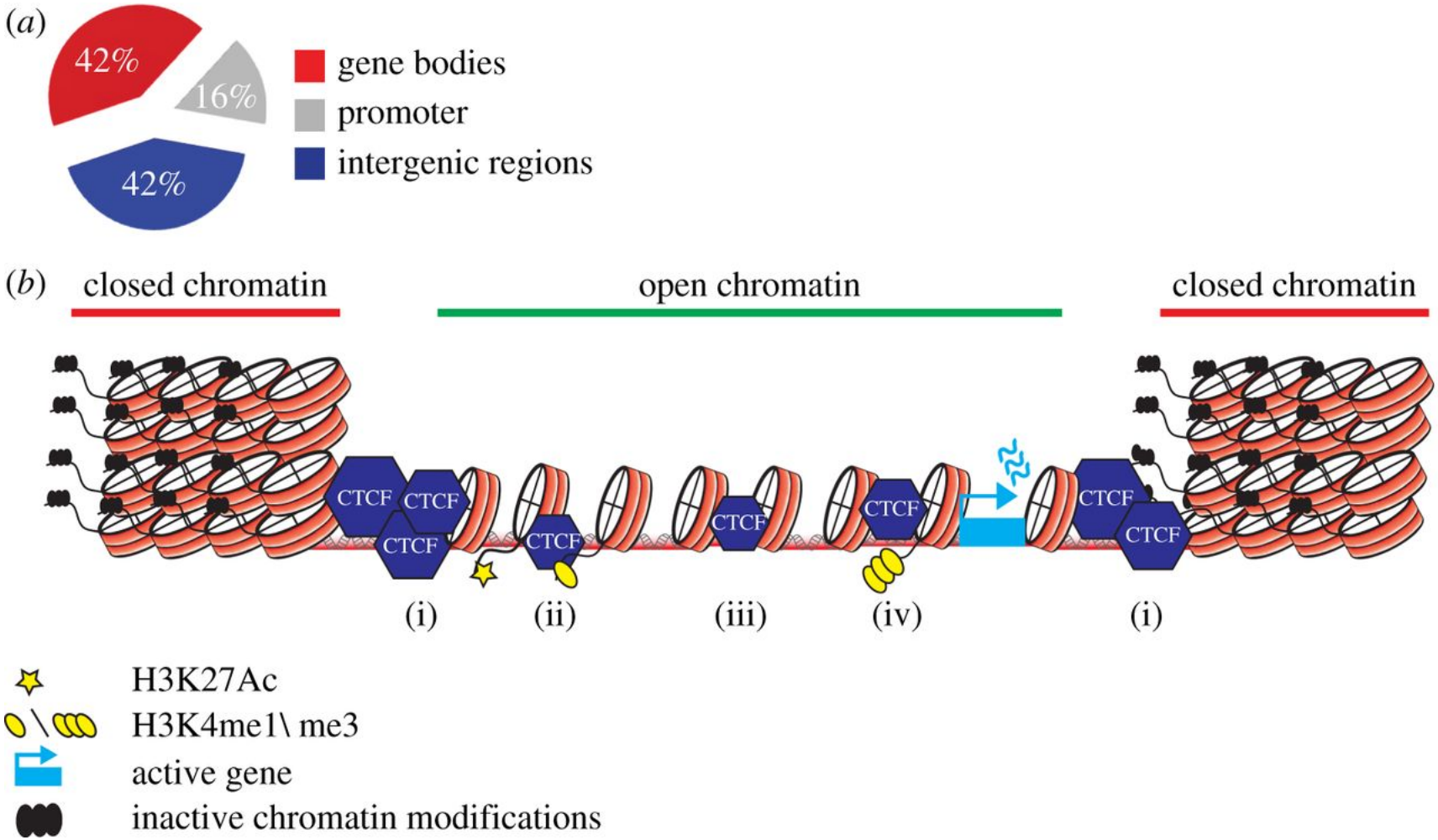
http://science.sciencemag.org/content/357/6348/eaal2380

Role of chromatin structure in cancer



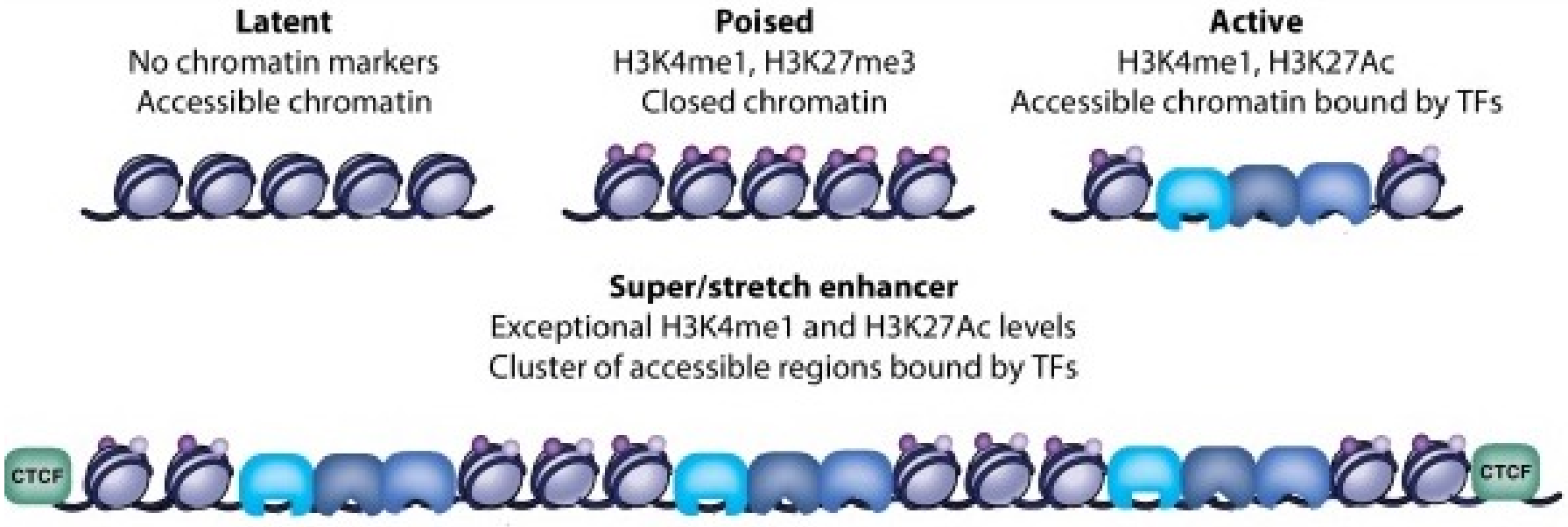
http://science.sciencemag.org/content/357/6348/eaal2380

Insulators/boundaries of open/closed chromatin



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

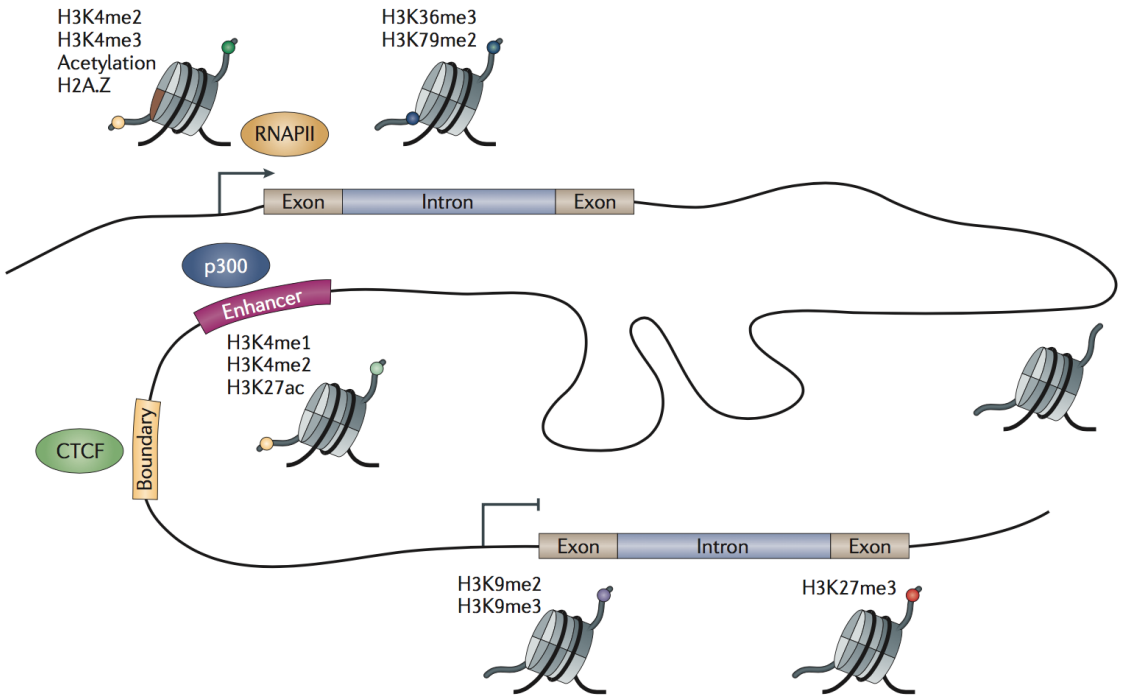
Super-enhancers



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

Histone code

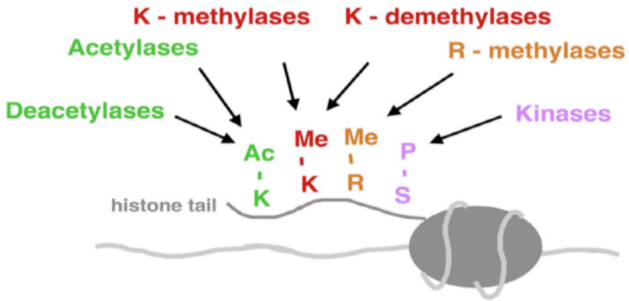
- Histone modifications demarcate functional elements in mammalian genomes.
- Can be broadly divided into activating and repressive histone marks.



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

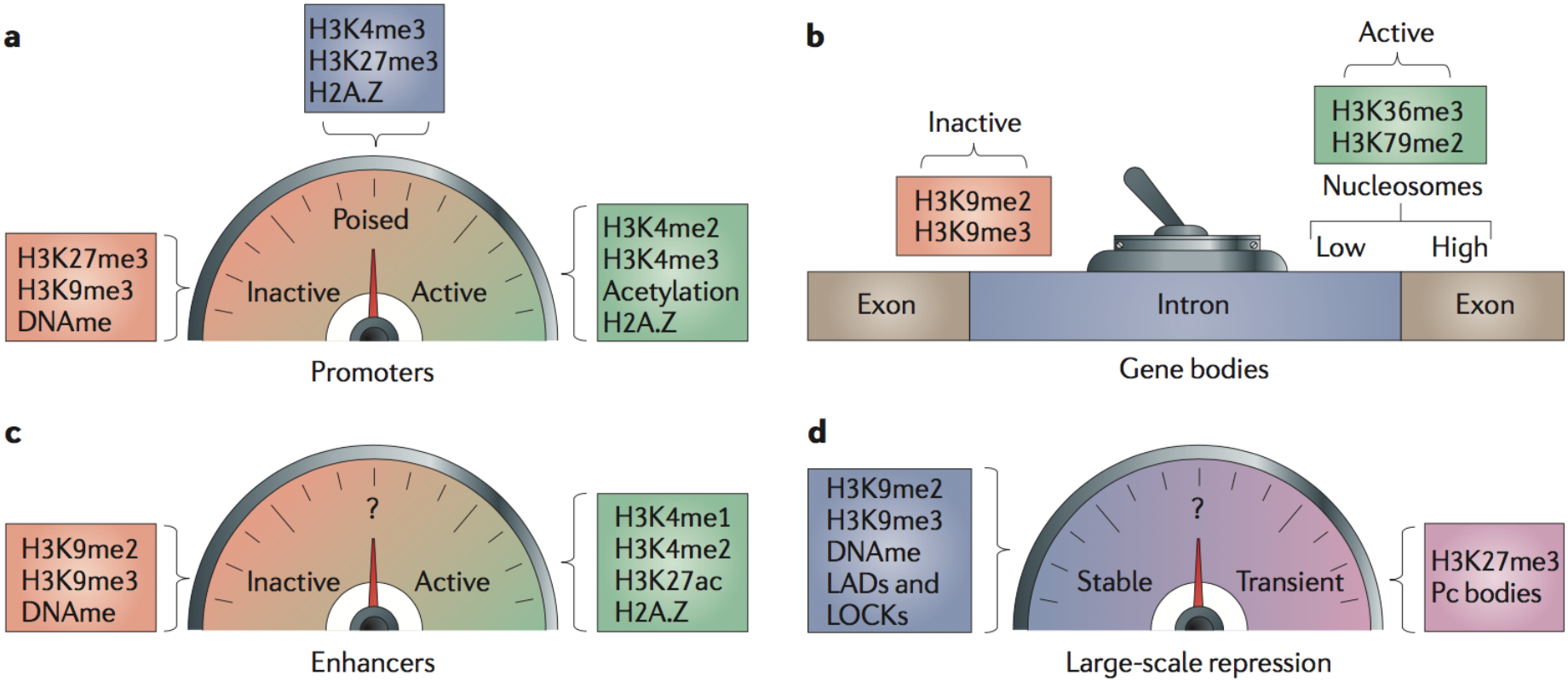
Histone modifications

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



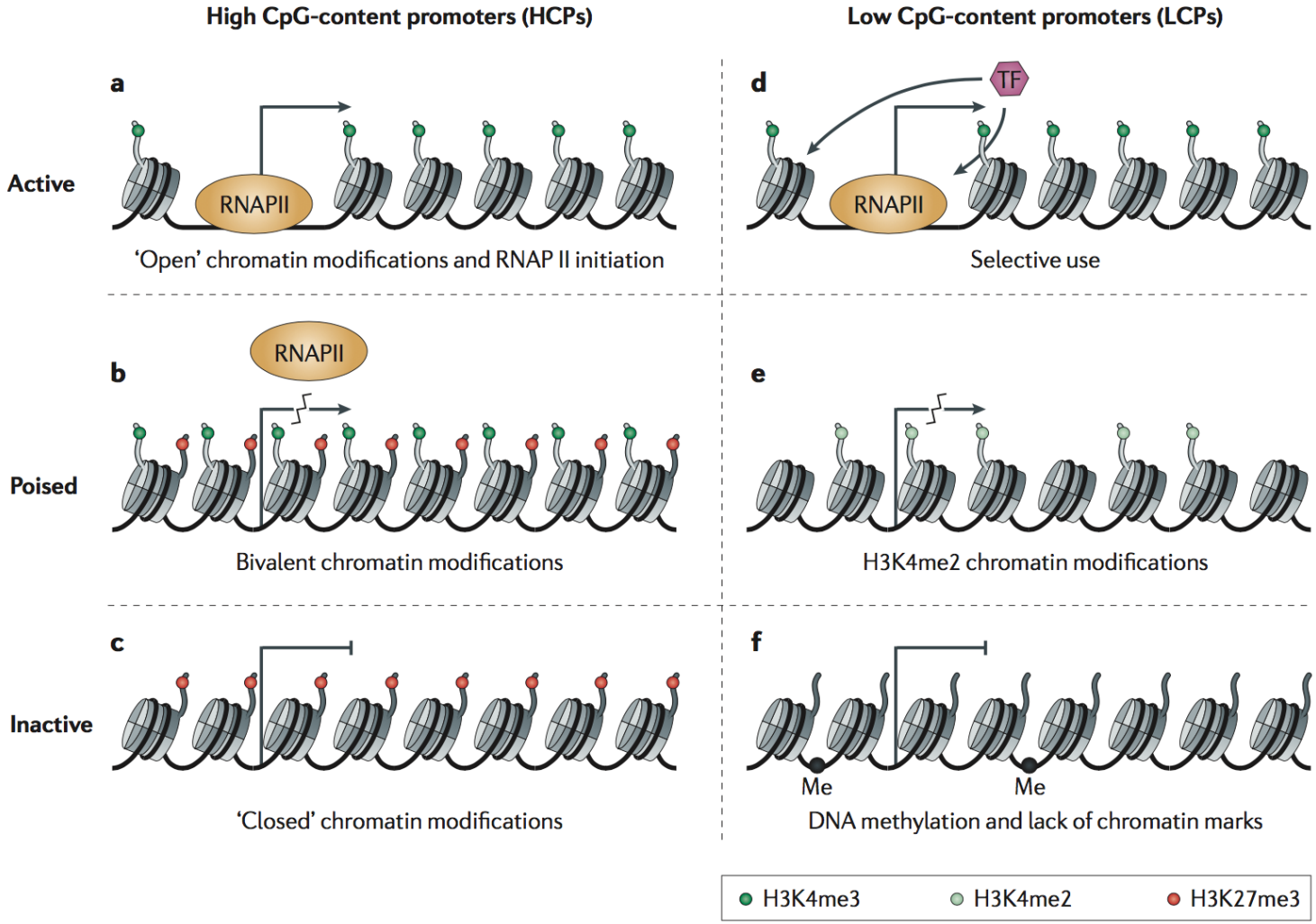
https://www.ncbi.nlm.nih.gov/pubmed/17320508

Histone code summary



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

Methylation and histone code interplay



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