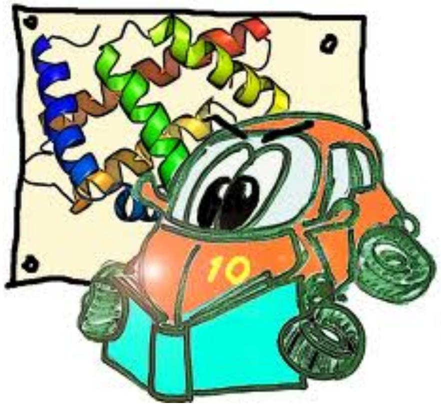
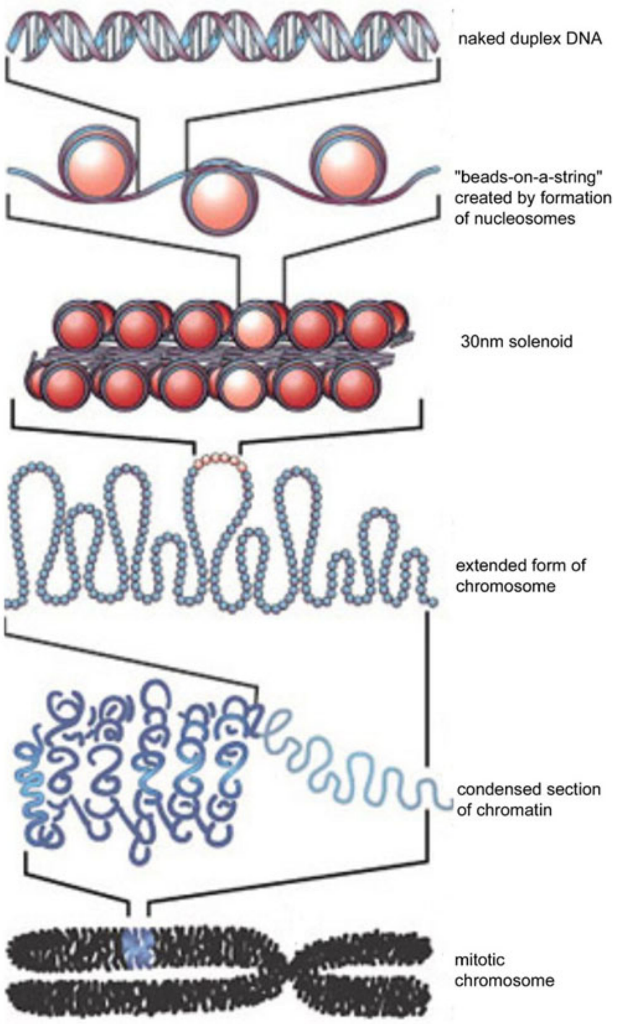
Machine Learning in Computational Biology CSC 2431



Lecture 4: Epigenetics

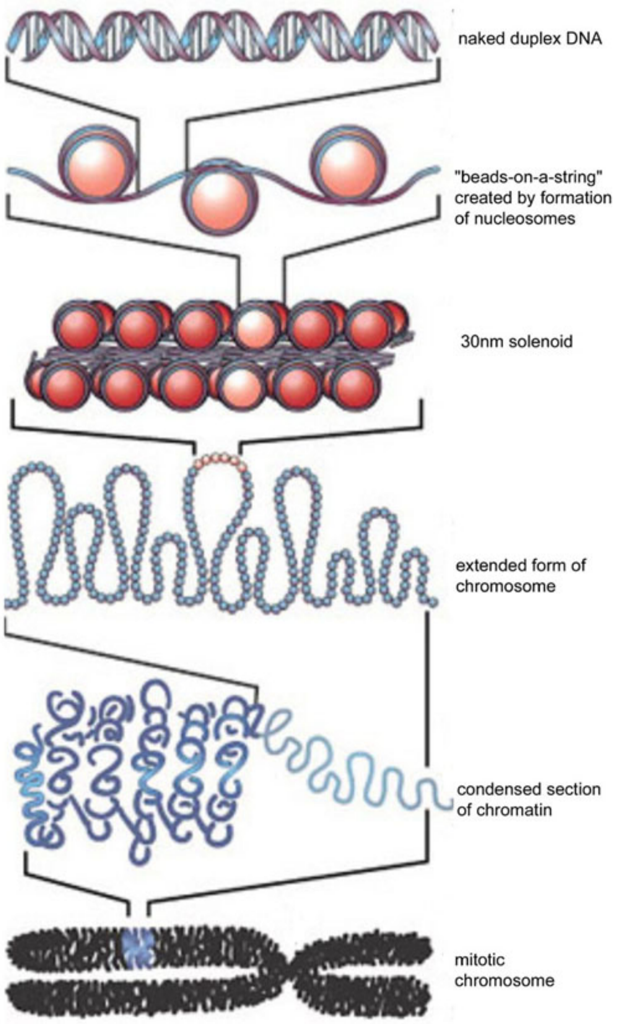
Instructor: Anna Goldenberg

Definitions



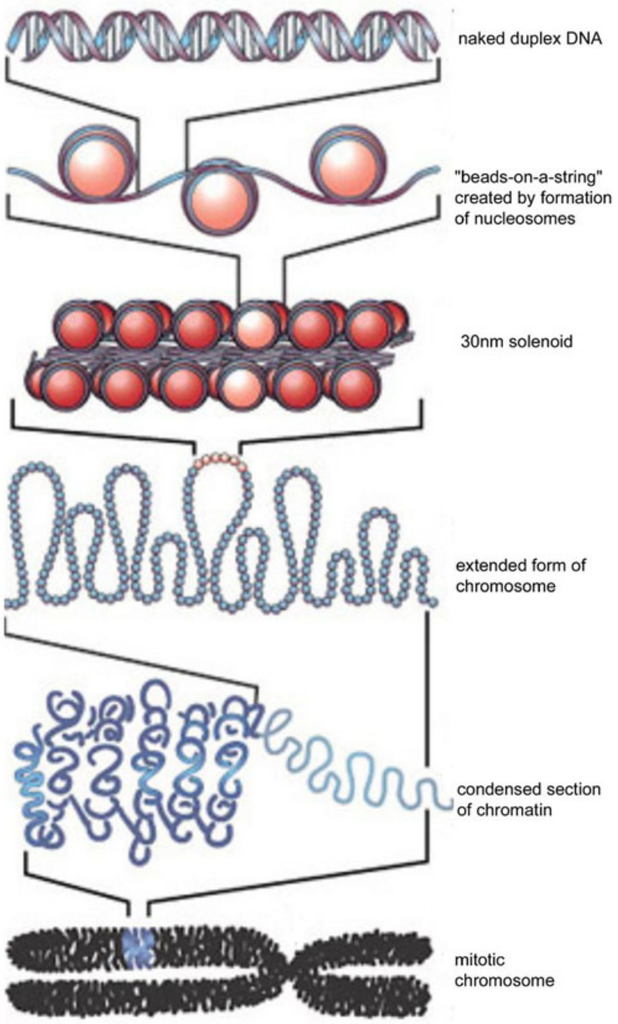
Definitions

- Histone: cluster of proteins



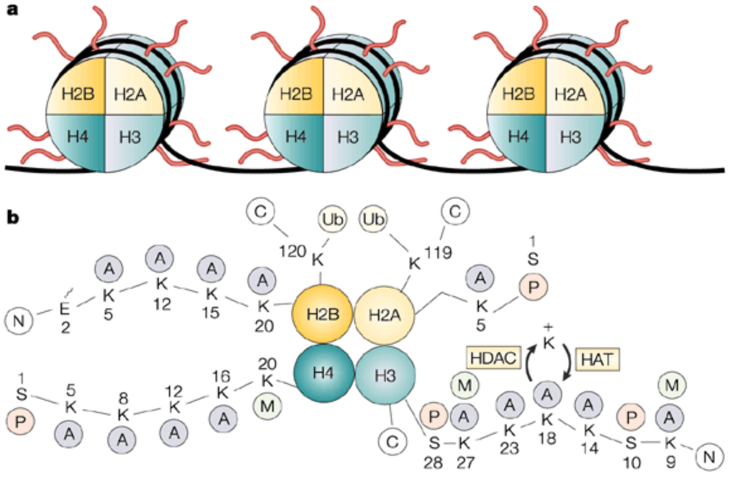
Definitions

- Histone: cluster of proteins
- Histone + DNA(146-7bp) = nucleosome

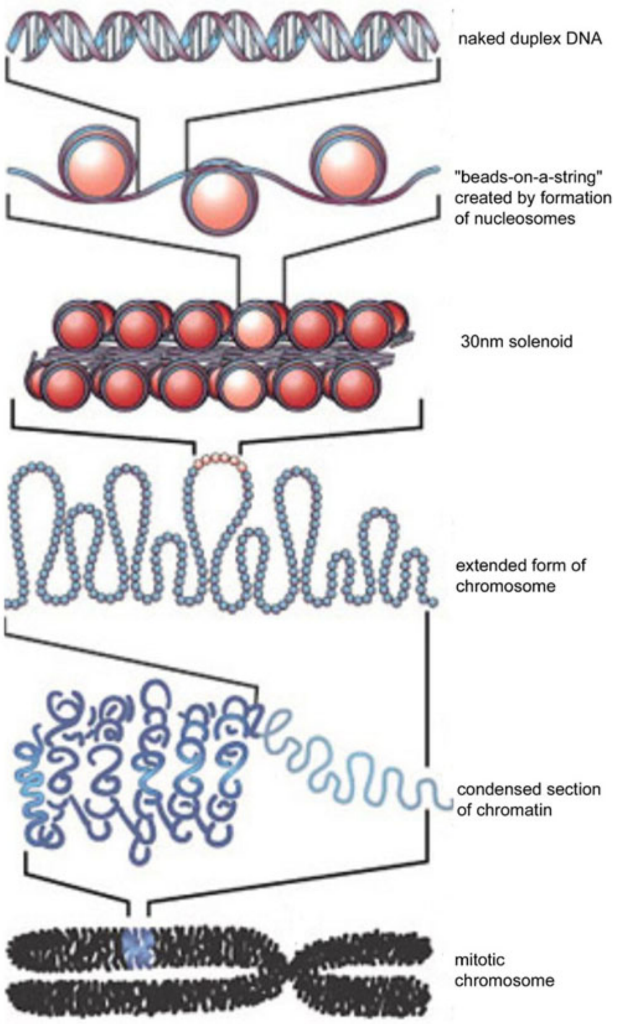


Definitions

- Histone: cluster of proteins
- Histone + DNA(146-7bp) = nucleosome



Nature Reviews | Cancer



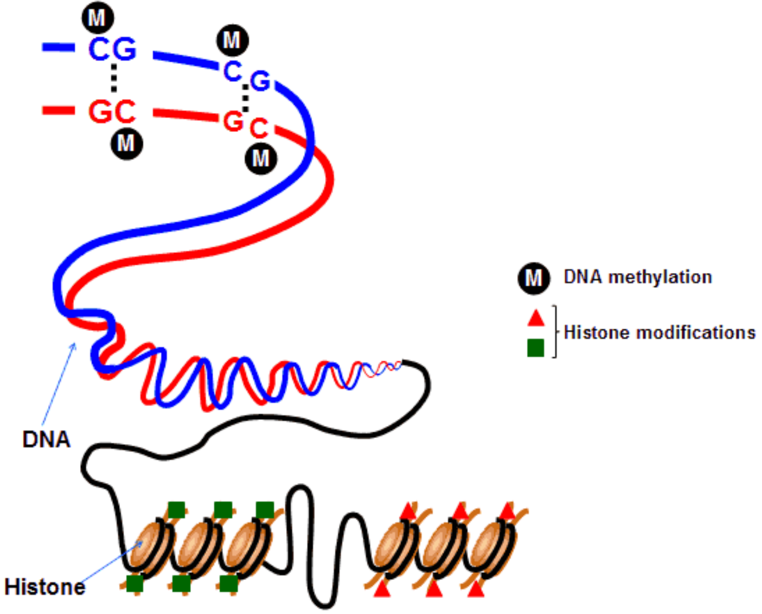
Definitions

- “Epi” – over, above, outer
- Epigenetics – stably heritable phenotype changes in a chromosome without alterations in the DNA sequence
  - Histone modifications
  - DNA methylation
- Epige*nomics* – refers to the study of the complete set of epigenetic alterations
- “Epigenetic code” – epigenetic features that maintain different phenotypes in different cells

Epigenetics

Modification to DNA
– DNA methylation

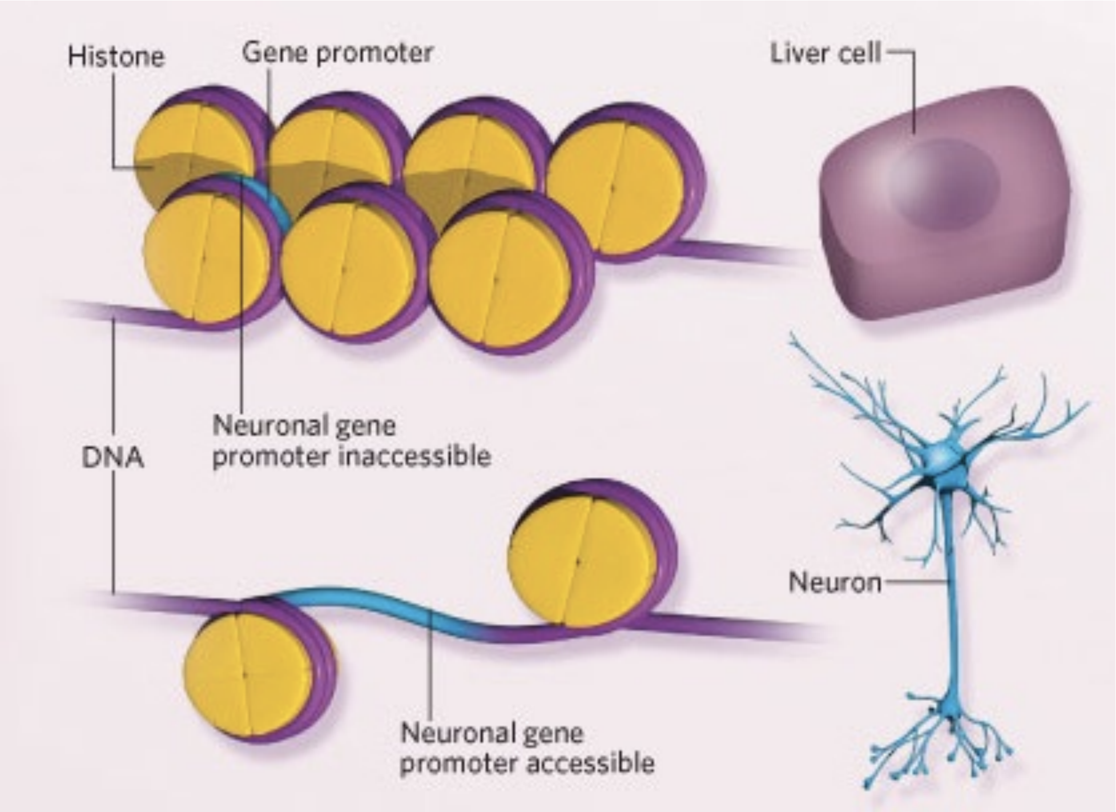
Modification to histones
(proteins around which DNA is wound)



These modifications change

- during differentiation
- as a response to environment

Example: differentiation



Tightly wound DNA – *heterochromatin*

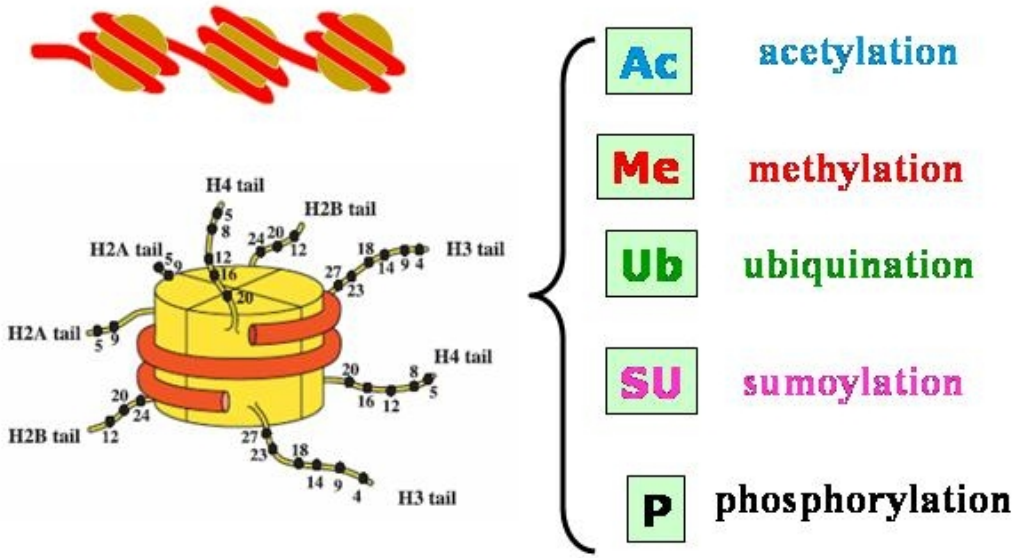
Loosely packed, open – *euchromatin*

Specific epigenetic processes

1. Imprinting (e.g. Angelman syndrome – maternally lost genes on chr15, paternally silenced)
2. Gene silencing
3. X chromosome inactivation
4. Paramutation (interaction between alleles at a single locus, e.g. maize)
5. Bookmarking (transmitting cellular pattern of expression during mitosis to the daughter cell)
6. Reprogramming
7. Transvection (interaction of alleles on diff. homologous chromosomes)
8. Maternal effects
9. Progress of carcinogenesis
10. Regulation of histone modifications and heterochromatin

Histone modifications (posttranslational)

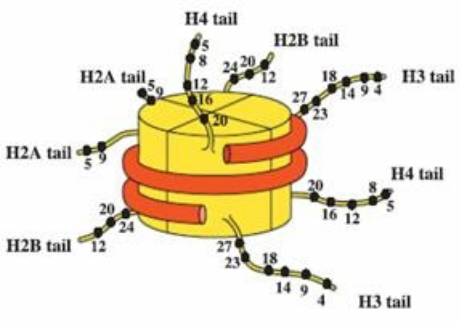
N-termini (tails) are particularly highly modified

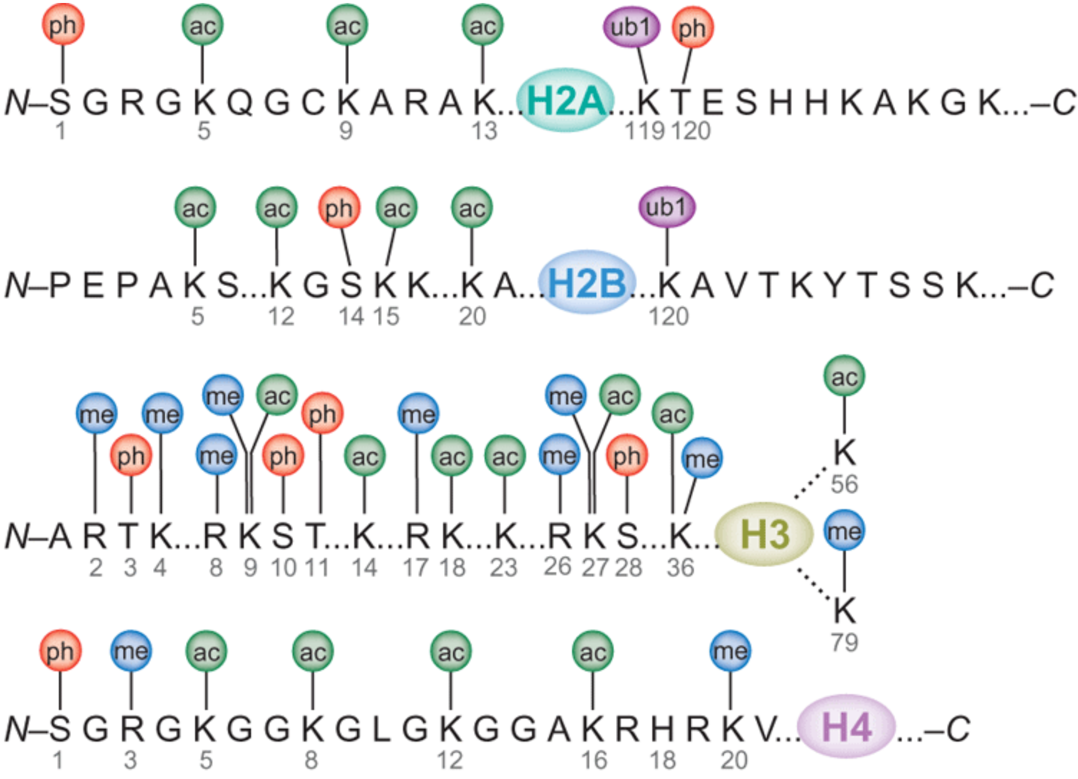


The figure illustrates nucleosome models and major posttranslational modifications which play essential roles in gene expression regulation and disease processes

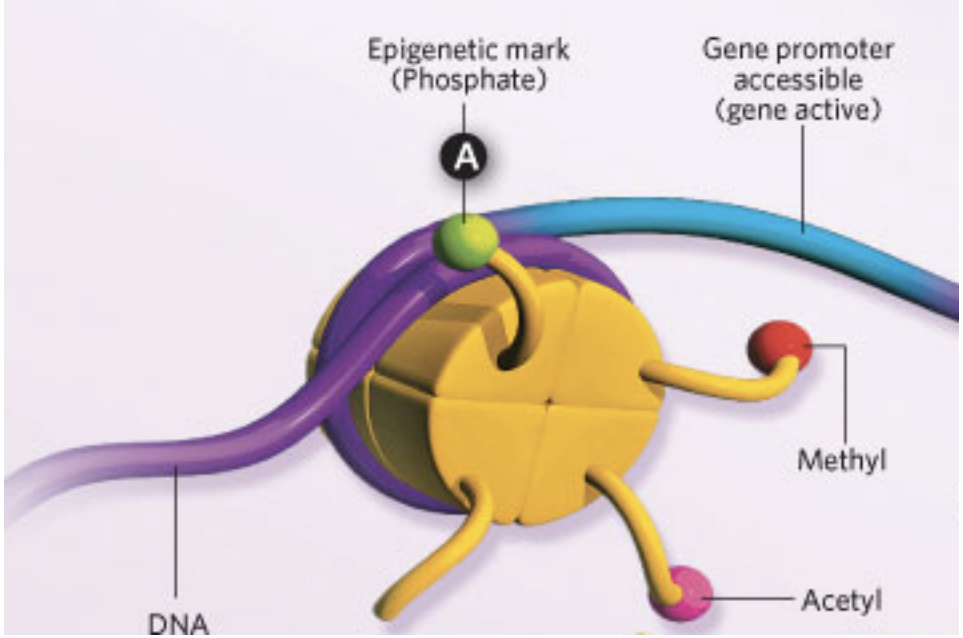
Histone modifications (posttranslational)

N-termini (tails) are particularly highly modified



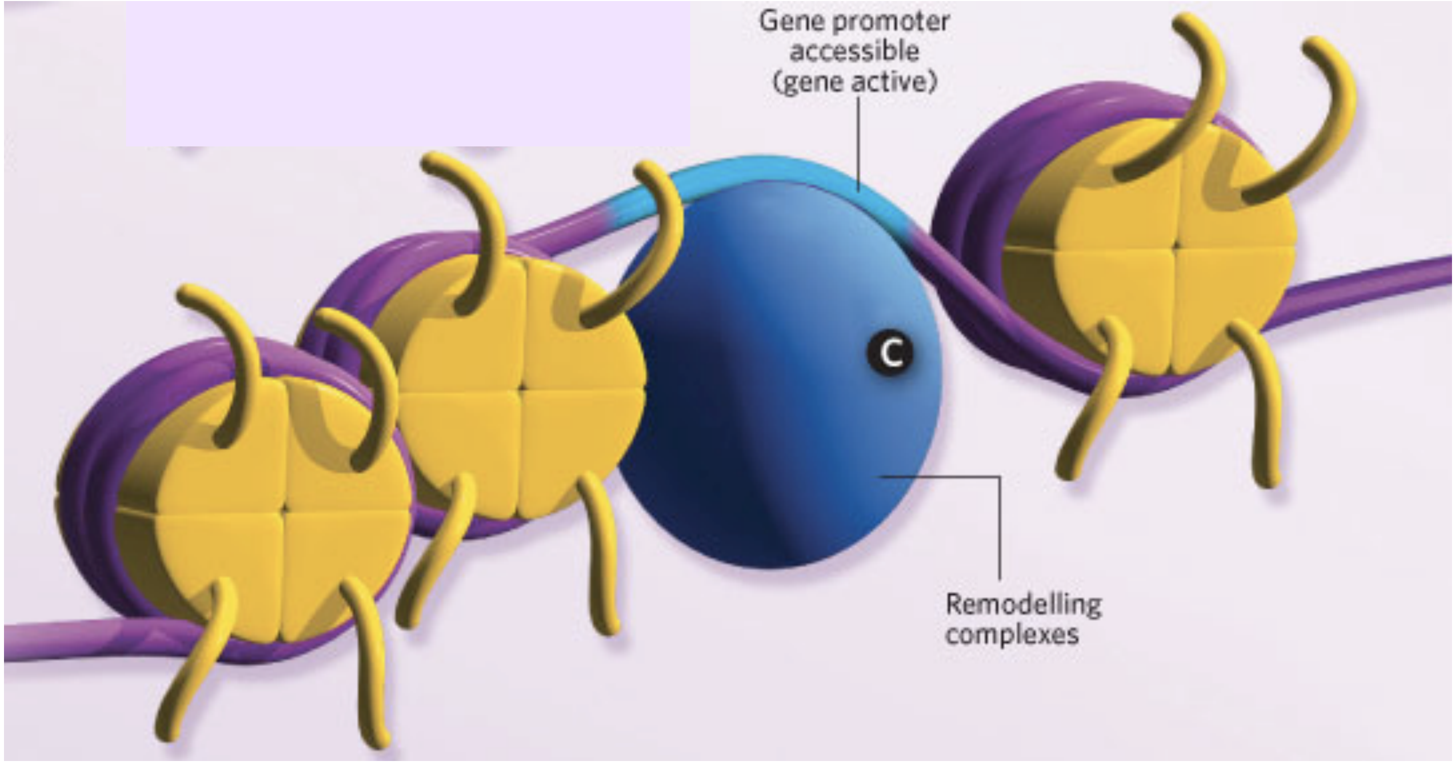


Acetylation and phosphorylation – help to open chromatin



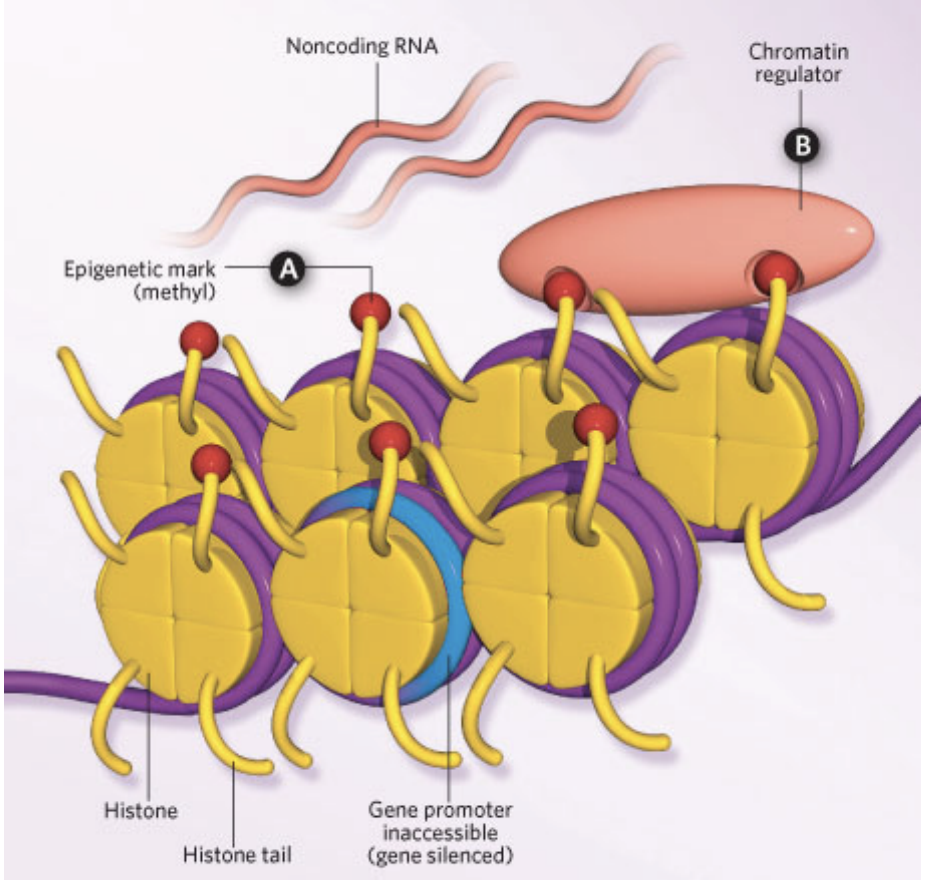
Another way to keep chromatin open

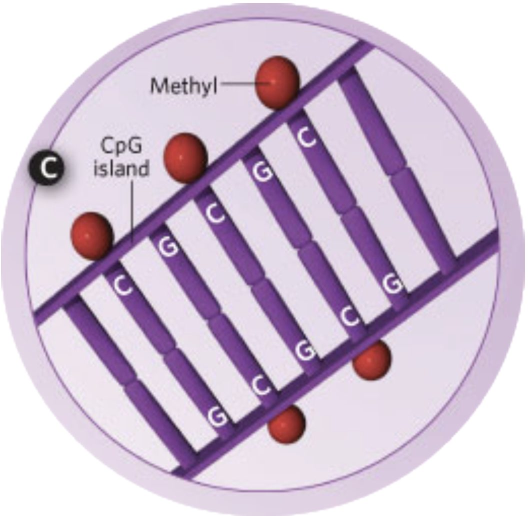
- Chromatin remodeling complex



Closed chromatin, gene silencing

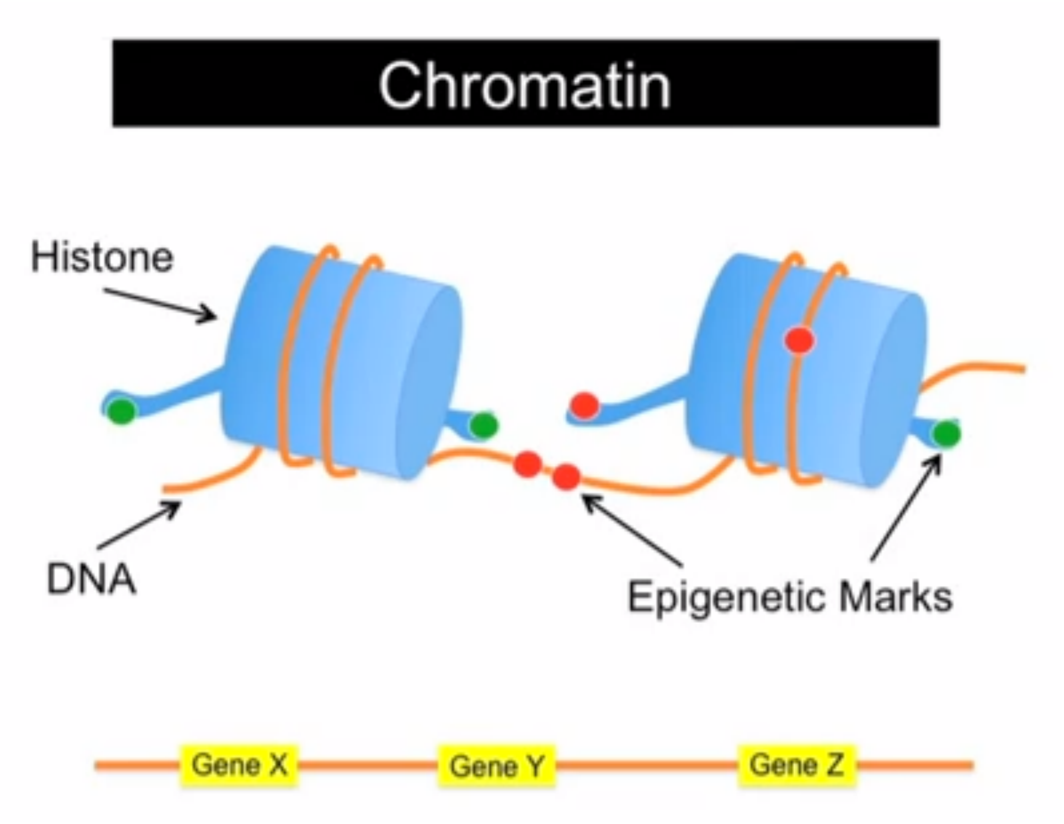
Histone and DNA methylation

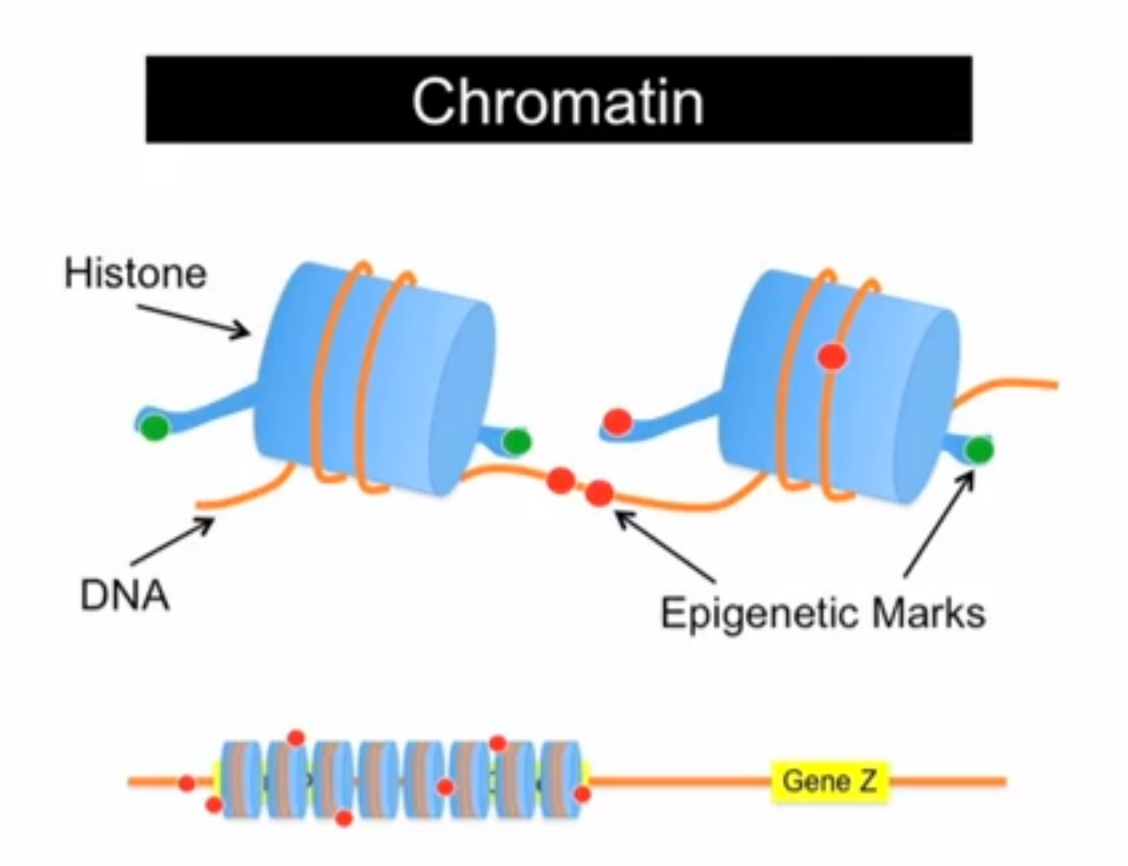




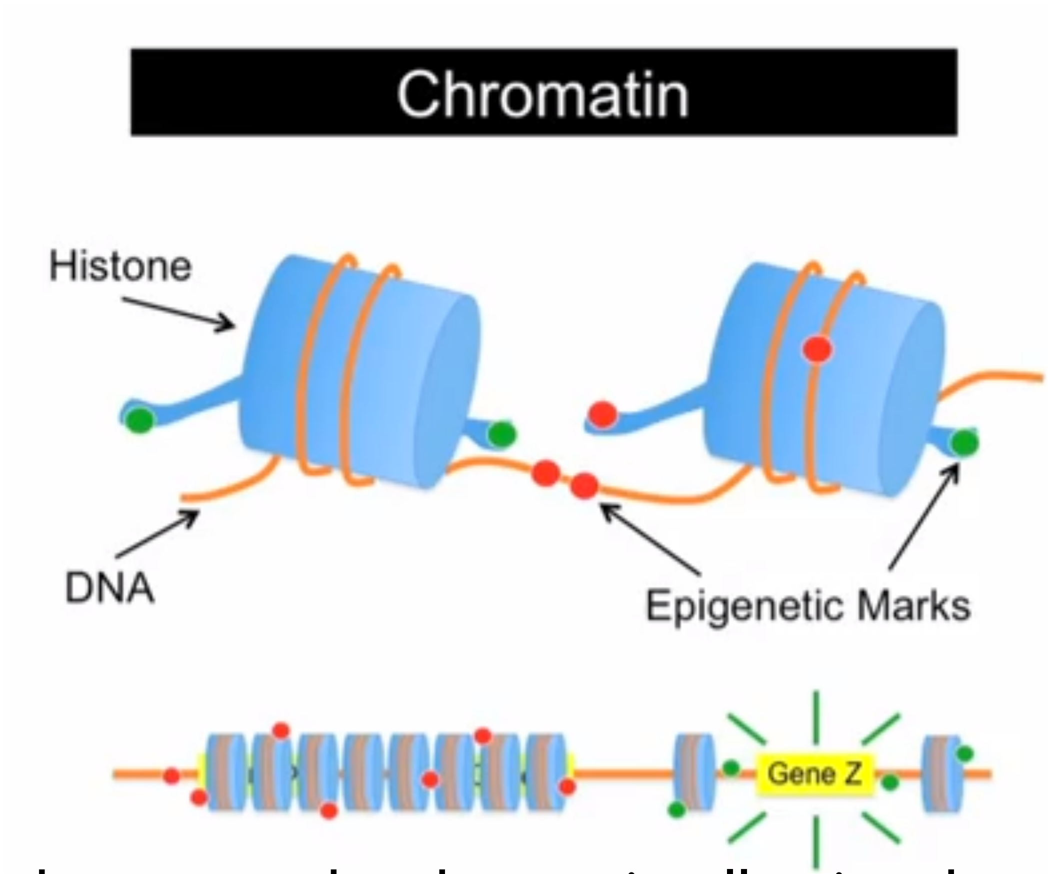
Epigenetic marks

Epigenetic marks – small chemical tags that sit on top of chromatin and help instruct it whether to open or to compact





Red marks – condense the chromatin, prevent the cell from being able to read the gene, turn the gene off (silencing)



Green marks – open the chromatin, allowing the gene to be read

DNA methylation

- 28 million of CpG regions in the genome
- 60-80% are heavily methylated
- CpG islands (100-2,000bp enriched for CpG often found at promoters) are un-methylated across cell types

- Modulation of DNA (de-)methylation is still unknown!

Typical computational analysis

- Statistical testing for differential DNA methylation at a single CpGs and/or large genomic regions
- Correction for multiple hypothesis testing
- Ranking based on statistical significance and effective size

Typical methods for DMR detection

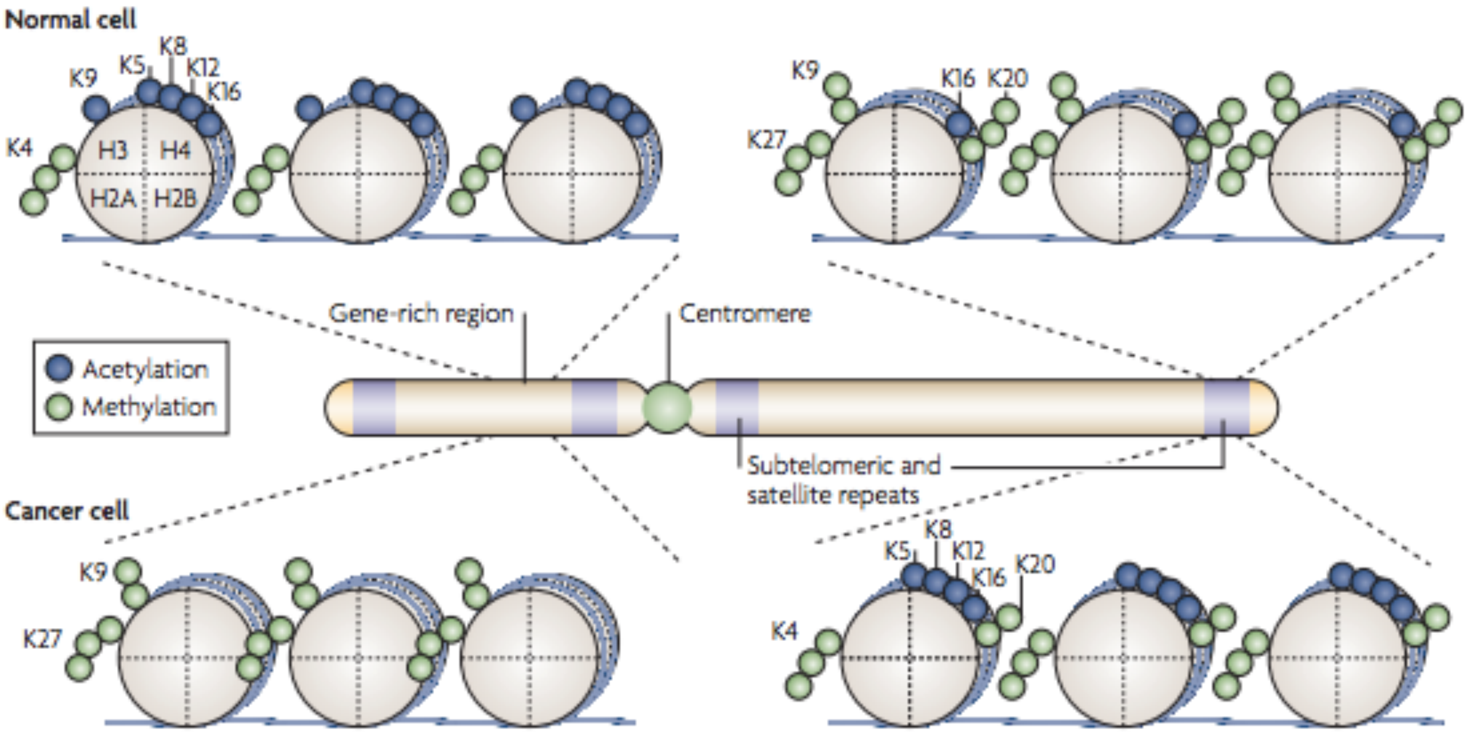
- T-test or Wilcoxon rank-sum (Wang et al, Bioinformatics, 2012)
- Mixture Models (Wang, Genetic Epidemiology, 2011)
- Information theoretic approaches (Zhang et al, NAR, 2011)
- Logistic M values (Du et al, BMC Bioinformatics 2010)
- Feature Selection (Zhuang et al, BMC Bioinformatics, 2012)
- Stratification of t-tests (Chen et al, Bioinformatics, 2012)
- Aggregation of genomic regions by type (Poage et al, Cancer Research, 2012)
- Correction for copy-number aberrations (Robinson et al, Genome Research, 2012)
- Linear regression with batch effect removal and peak detection (Jaffe et al, Int J Epidem, 2012)

C. Bock, Nature Reviews Genetics, v13, October 2012

Computational challenges

- Comprehensive mapping of histone modifications, nucleosome positioning, TF binding and chromosomal organization per tissue is going to be done on a smaller scale: detecting signal will be hard since the sample size will be small
- Integrating all the epigenetic data together and with other types of data
- Tools to help identify causes from consequences of the differences in DNA methylation
- New technologies: nanopore sequencing, new tools to address biases
- Functional relevance of the DNA methylation variants

Example:
Histone modification profiles
Normal vs Cancer



Key findings in cancer

1. Hypermethylation of CpG islands

   CpG islands in the promoters of tumor suppressor genes are methylated

   ↓

   Tumor suppressor genes are inactivated

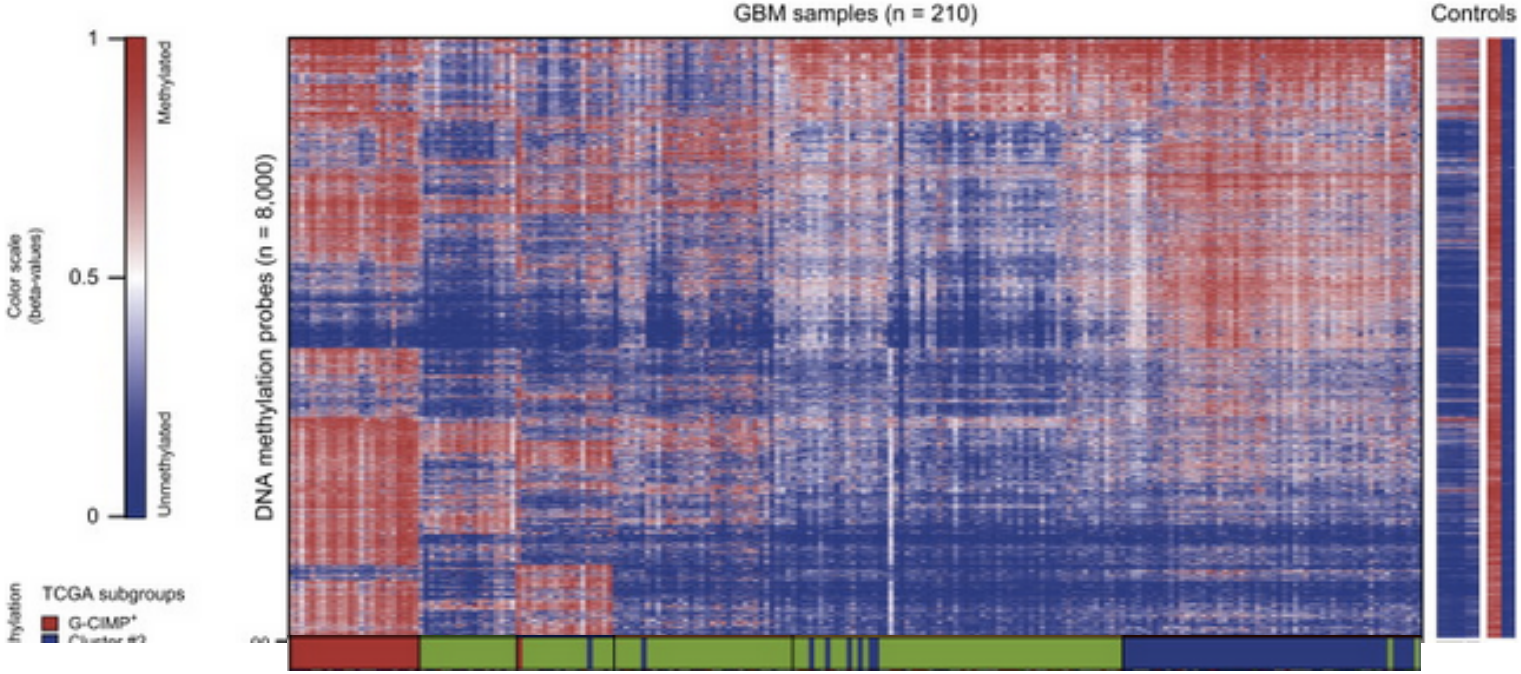
   ↓

   Tumors are able to grow

2. General Hypomethylation

Interesting case

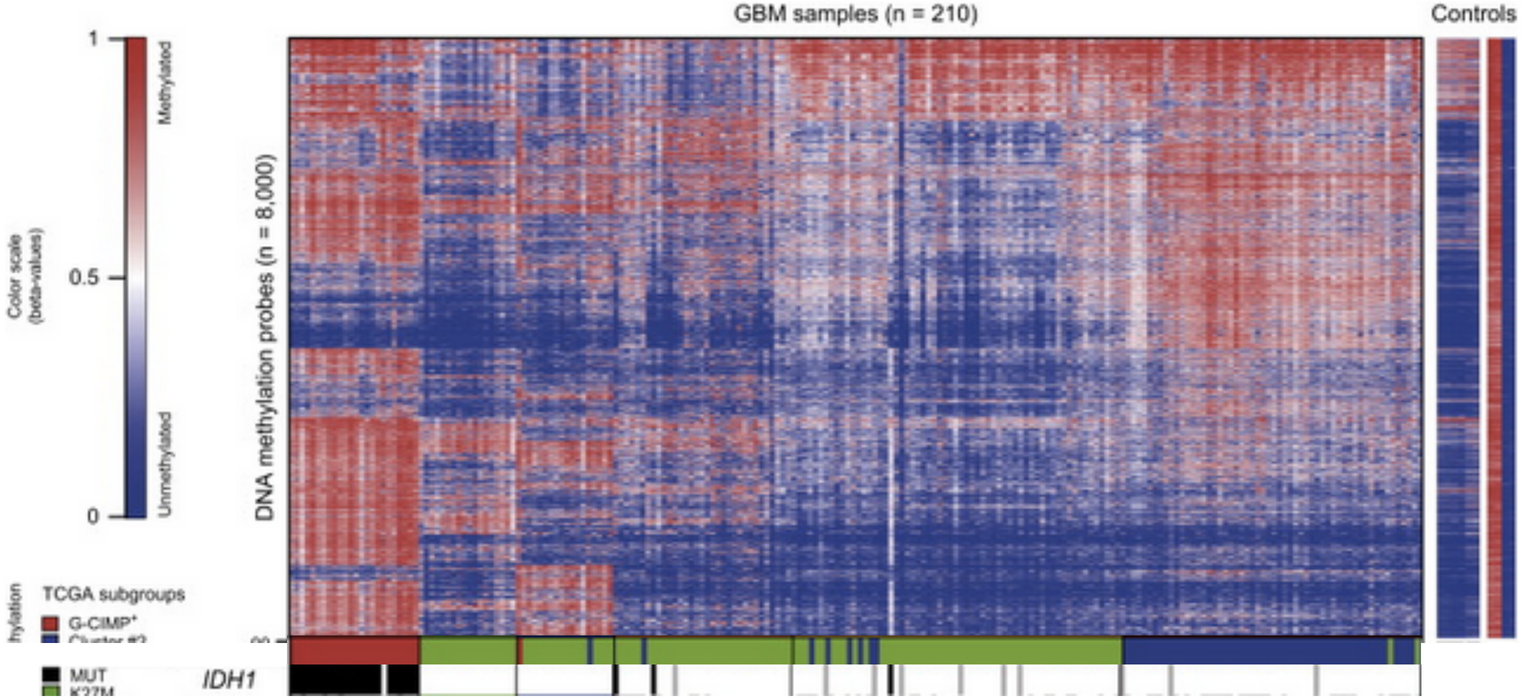
Glioblastoma Multiforme



Sturm et al, Cancer Cell, 2012

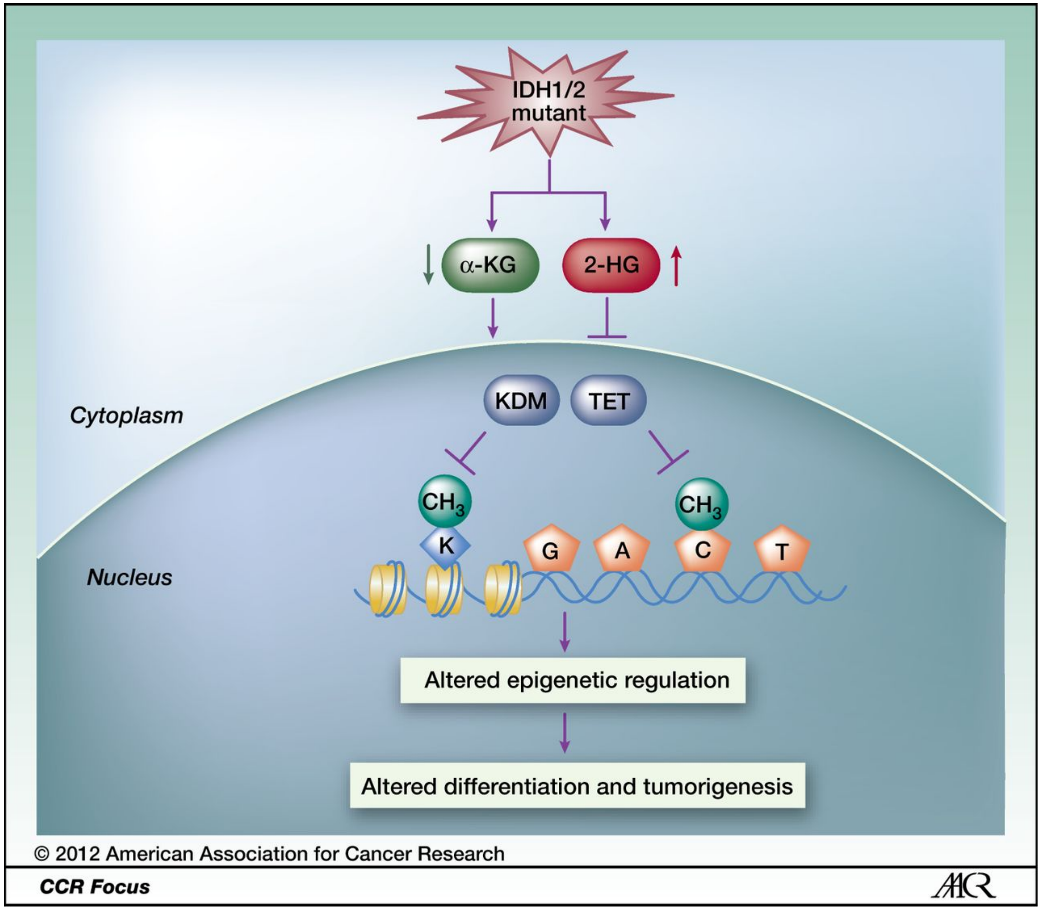
Interesting case

Glioblastoma Multiforme



Sturm et al, Cancer Cell, 2012

IDH1



CCR Focus

AACR

IHD1/2 mutations inhibit both histone and DNA demethylation and alter epigenetic regulation

Epigenetics Databases

- MethDB 5,382 methylation patterns, 48 species, 1151 individuals, 198 tissues and cell lines, 79 phenotypes
- PubMeth 5000+ records on methylated genes in cancers
- REBASE 22,000+ DNA methyltransferases genes derived from GenBank
- MeInfoText methylation information across 205 human cancer types
- MethPrimerDB 259 primer sets from human, mouse and rat for DNA methliation analysis
- ChromDB 9,341 chromatin association proteins
- The Histone Database – 254 sequences from histone H1, 383 from H2, 311 from H2B, 1043 from histone H3 and 198 from H4
- Epigenetic Roadmap (NIH project)

Papers

- DNA methylation across tissues:
  Ma, B., Wilker, E. H., Willis-Owen, S. A., Byun, H. M., Wong, K. C., Motta, V., ... & Liang, L. (2014). Predicting DNA methylation level across human tissues. Nucleic acids research, 42(6), 3515-3528.

- Inferring chromatin states
  Ernst, Jason, and Manolis Kellis. "Discovery and characterization of chromatin states for systematic annotation of the human genome." Nature biotechnology 28.8 (2010): 817-825.