3rd Lecture: Epigenetic Gene Regulation

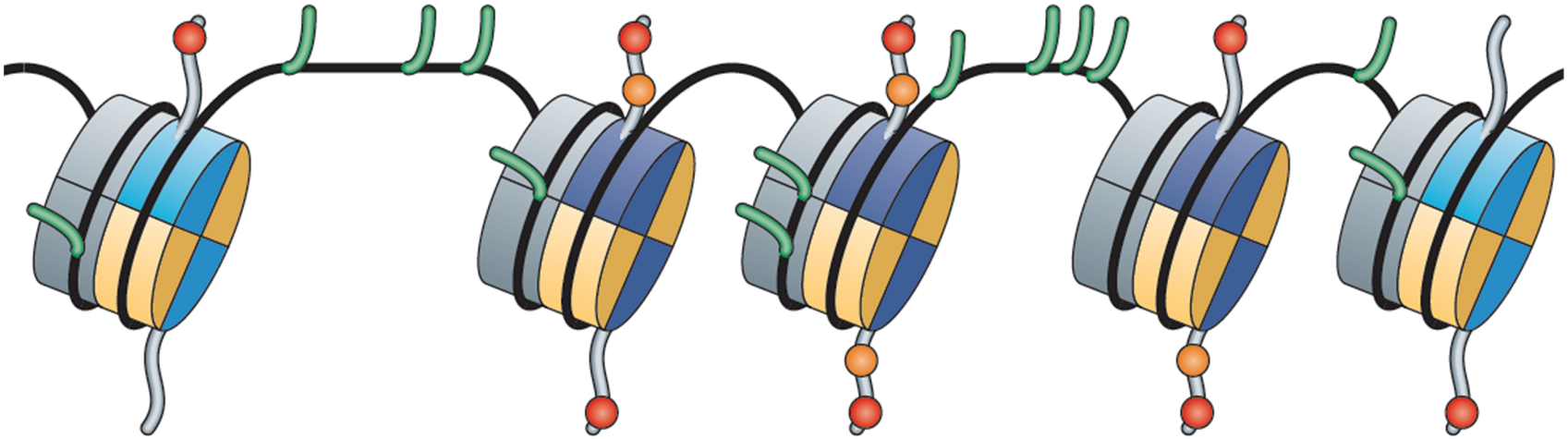
Christian Arnold

Computational EvoDevo
Institute of Computer Science
University of Leipzig

October 28, 2011

Lecture overview

1. Introduction to Epigenetic Gene Regulation



What is epigenetics, really?

What is epigenetics, really?

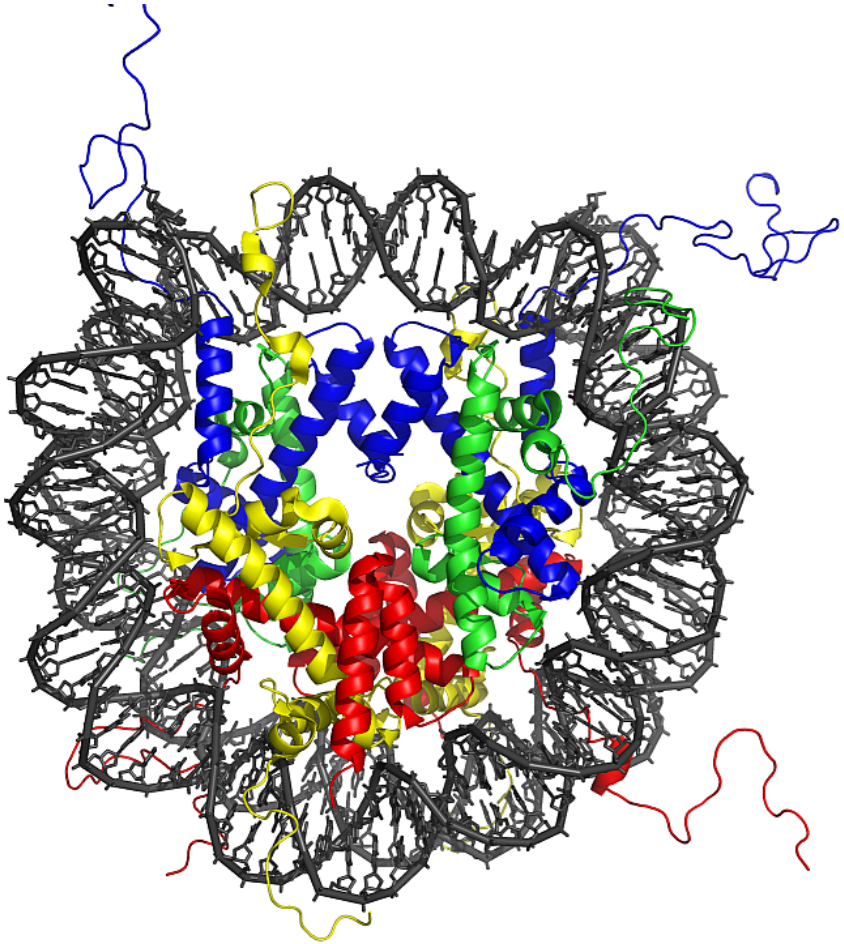
Original (and still used) definition:

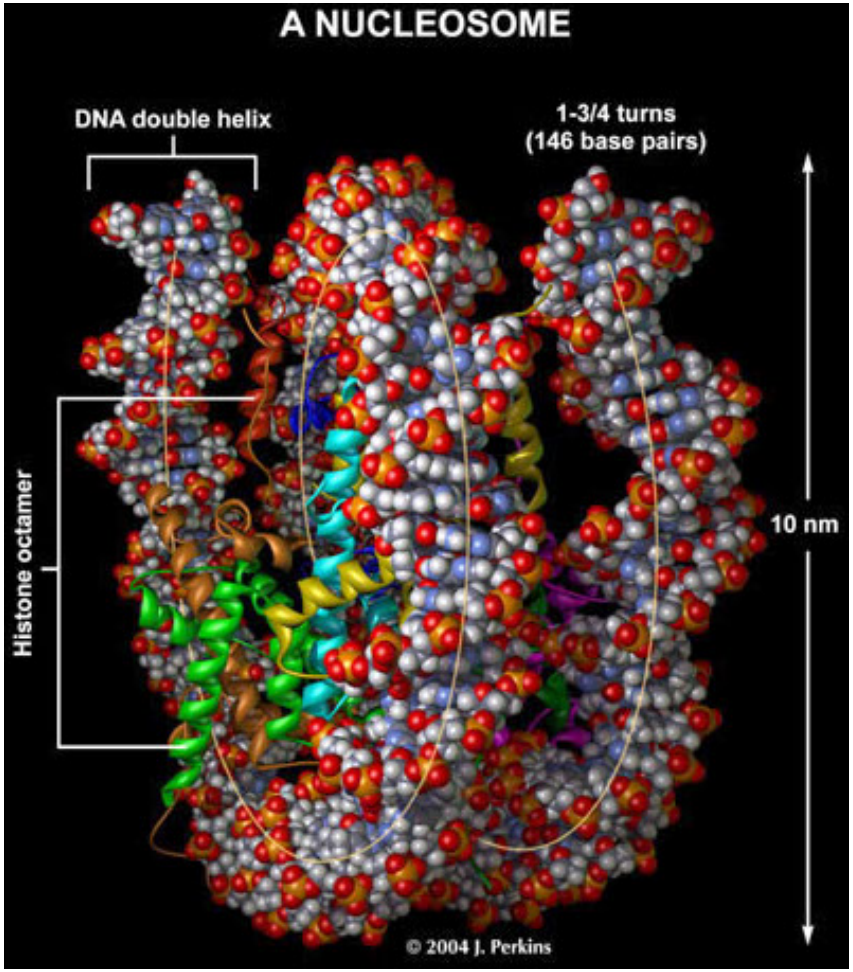
"The study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence." (Riggs et al. 1996)

Now:

Anything chromatin-related

Nucleosome-DNA structure

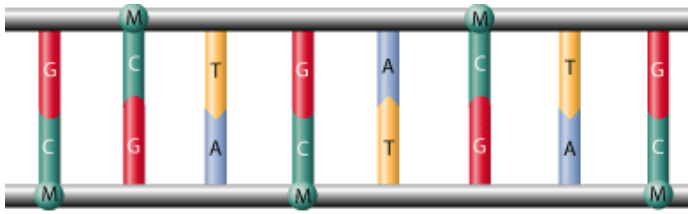


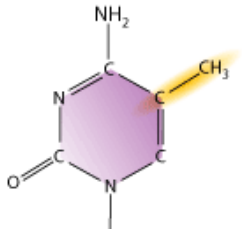


Epigenetic keyplayers

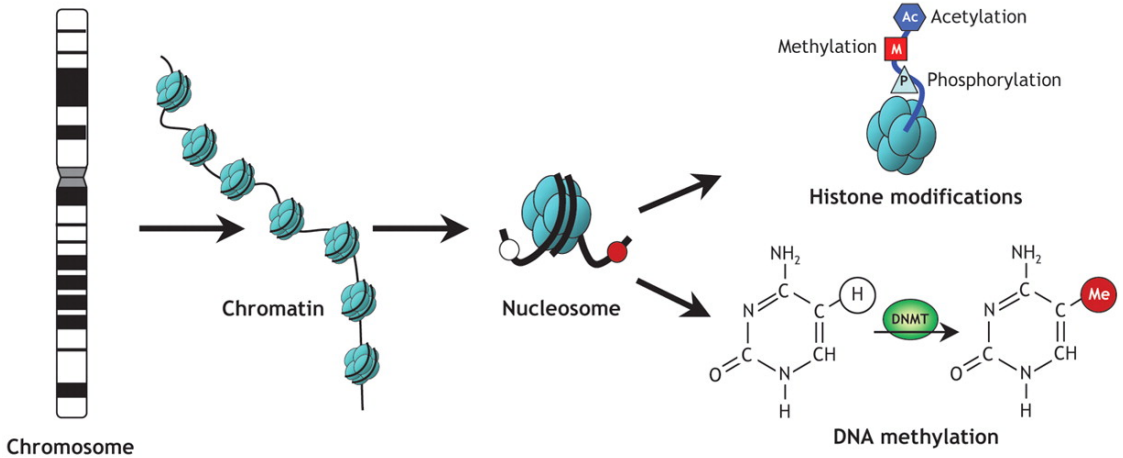
- Histone modifications: methylation, acetylation, phosphorylation, ubiquitylation, SUMOylation, propionylation, butyrylation, formylation, citrullination, proline isomerization, ADP ribosylation, tyrosine hydroxylation, lysine crotonylation, ...
- Histone variants: e.g., for H3: CenH3/CENP-A, H3.1, H3.2, and H3.3, H3.T
- Nucleosome occupancy and spacing
- Histone replacement (turnover) and nucleosome eviction
- Chromatin packaging
- DNA methylation (yielding Methylcytosine and subsequently Hydroxymethylcytosine)

Epigenetic keyplayers

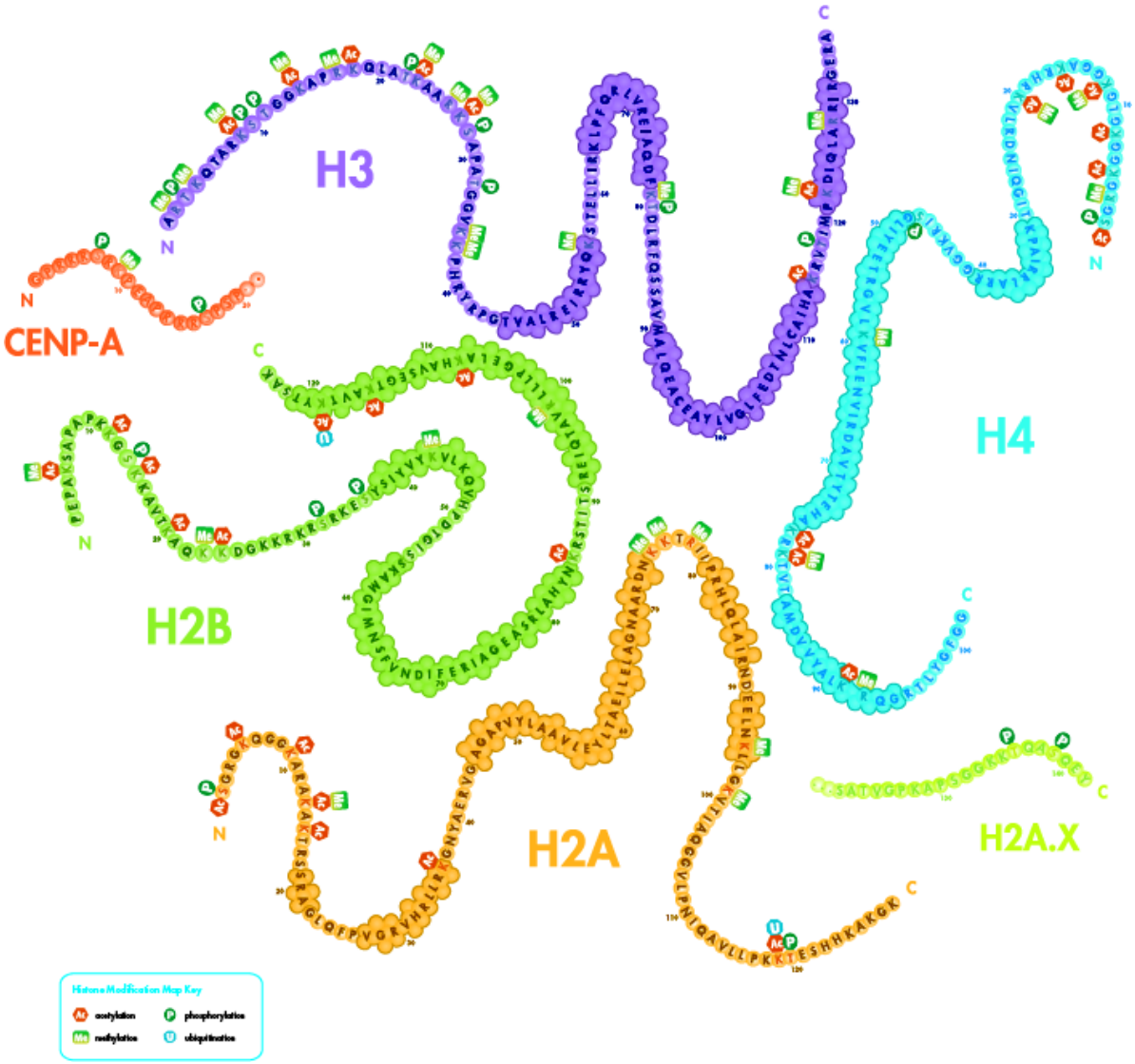




DNA methylation is the addition of a methyl group (M) to the DNA base cytosine (C).



Histones and their post-translational modifications



Histone modifications: Functions and Examples

- Histone methylations and acetylation are the best-studied histone modifications
- Functions
  - involved in a multitude of biological processes, such as transcription, DNA repair, DNA replication, DNA methylation, or possibly even epigenetic inheritance
  - control DNA accessibility by altering either the net charge of the histones or inter-nucleosomal interactions
  - attract or inhibit chromatin binding complexes, causing regulatory changes
- large differences in stability among modifications
- **modifications are reversible**

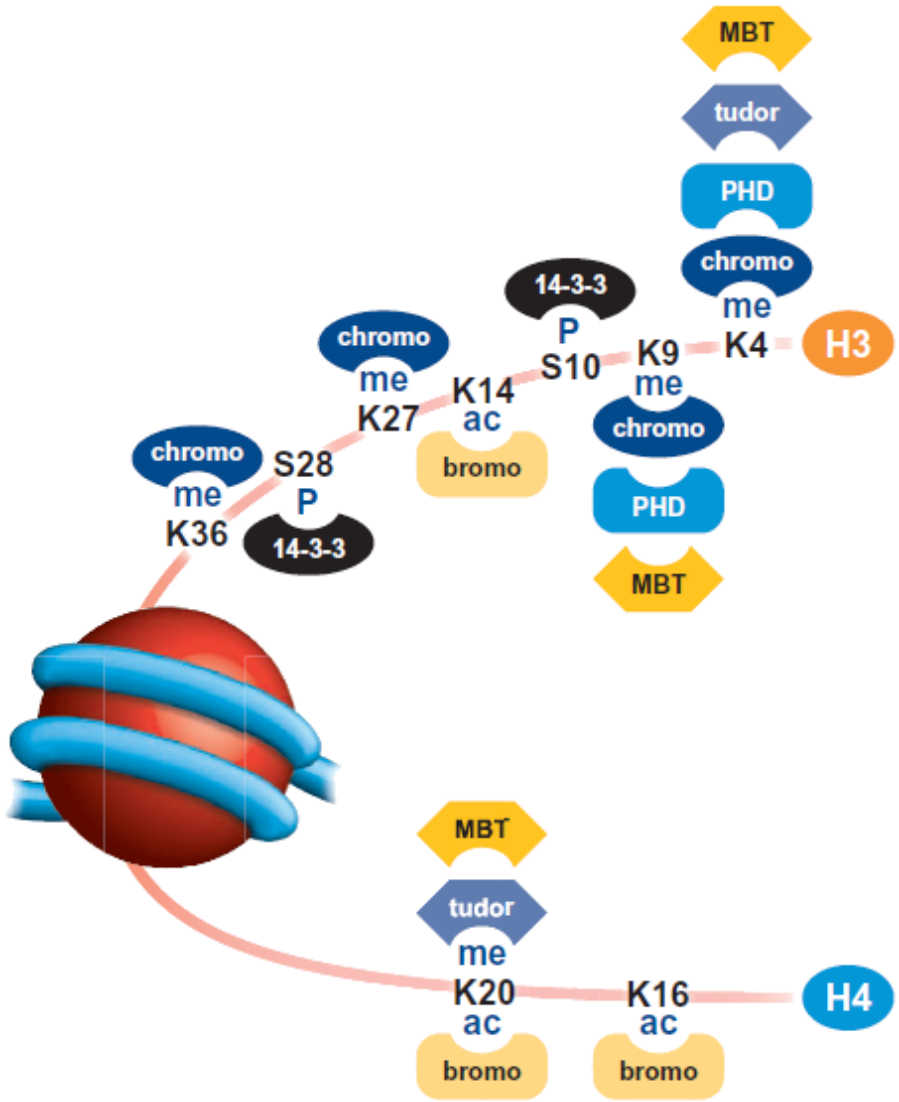
Histone methylations and acetylations

- Histone methylations
  - effect on transcription dependent on specific residue and methylation degree
  - at lysine (K) or arginine (R) residues
  - K can be mono-, di-, or trimethylated, R only mono- and dimethylated (symmetric and asymmetric)
- Histone acetylations

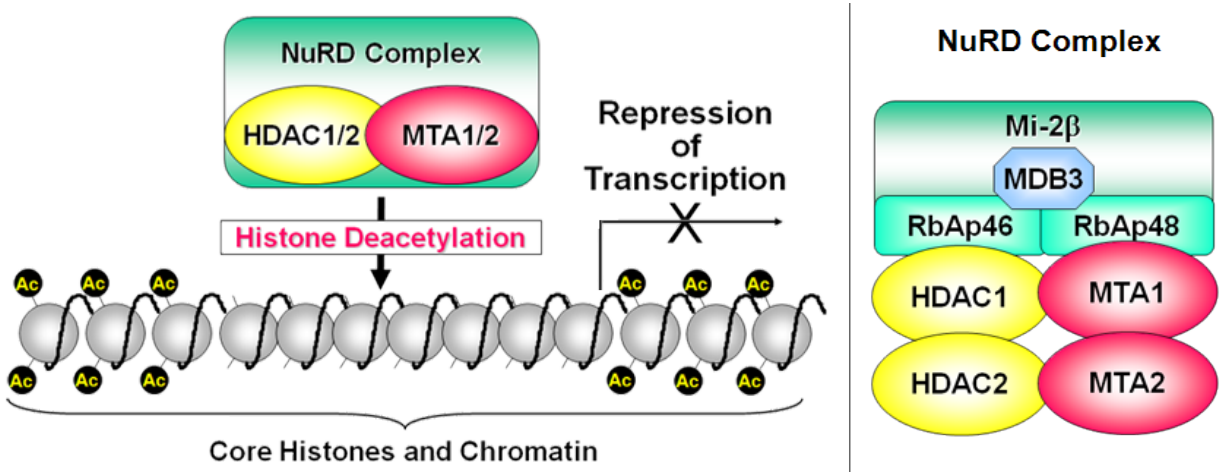
Histone methylations and acetylations

- Histone methylations
  - effect on transcription dependent on specific residue and methylation degree
  - at lysine (K) or arginine (R) residues
  - K can be mono-, di-, or trimethylated, R only mono- and dimethylated (symmetric and asymmetric)
- Histone acetylations
  - generally highly correlated with gene activity
  - only at lysine (K) residues

Histone binding domains



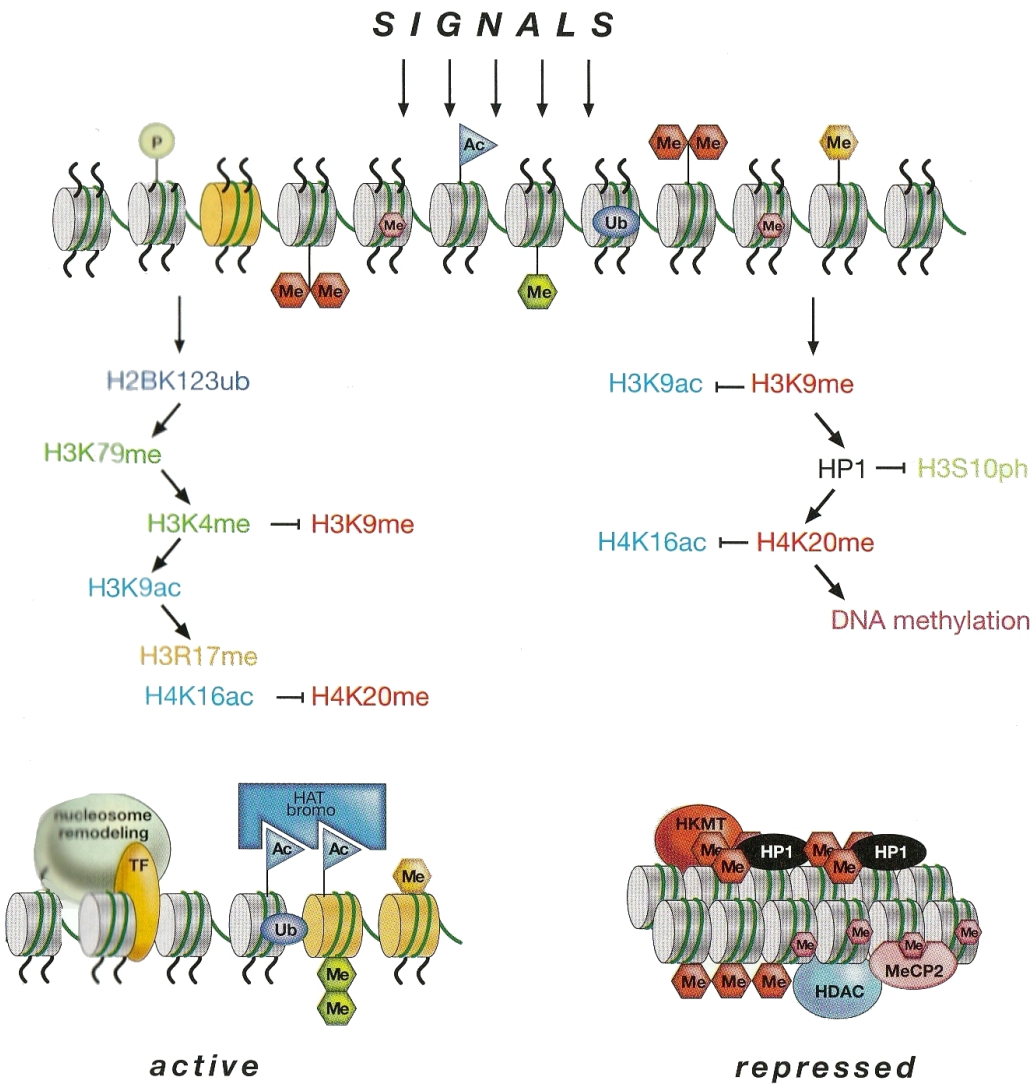
Chromatin remodeling example: The NuRD complex



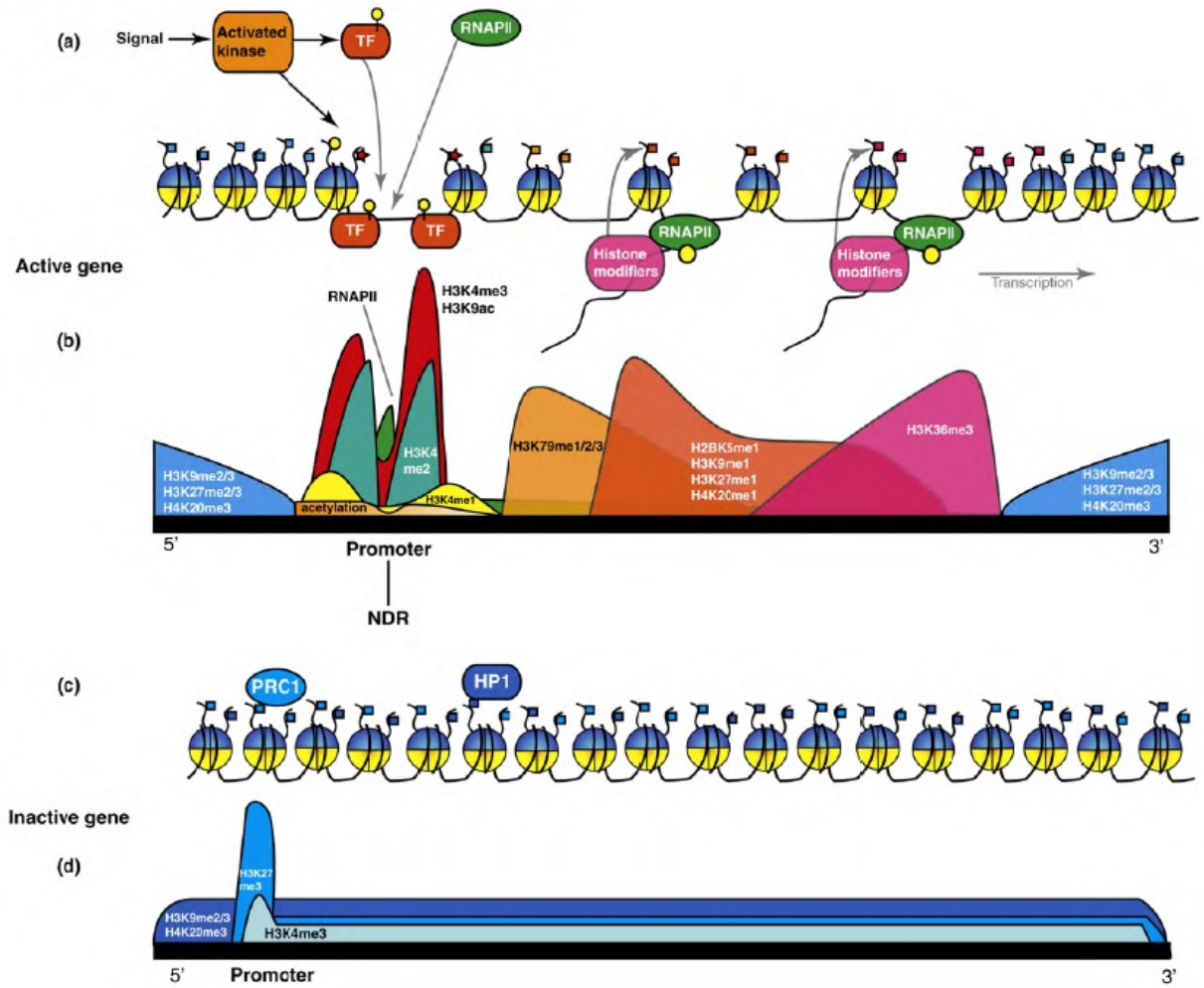
Source: Toh and Nicolson 2009. The role of the MTA family and their encoded proteins in human cancers: molecular functions and clinical implications.

- NuRD: **Nu**cleosome **R**emodeling and Histone **D**eacetylation complex
- ATP-dependent chromatin remodeling and histone deacetylase (HDAC) activity
- establishes transcriptional repression in vertebrates, invertebrates and fungi

Histone crosstalk and signaling pathways



Summary: Active and silent genes



Source: Barth and Imhof 2010. Fast signals and slow marks:the dynamics of histone modifications.

2. Epigenetics and Bioinformatics

Epigenetics deals with large amounts of data

- No epigenetic analysis without computers possible
- A single experiment may produce Terabytes of data
- Downstream analyses have to be automated and programmed efficiently

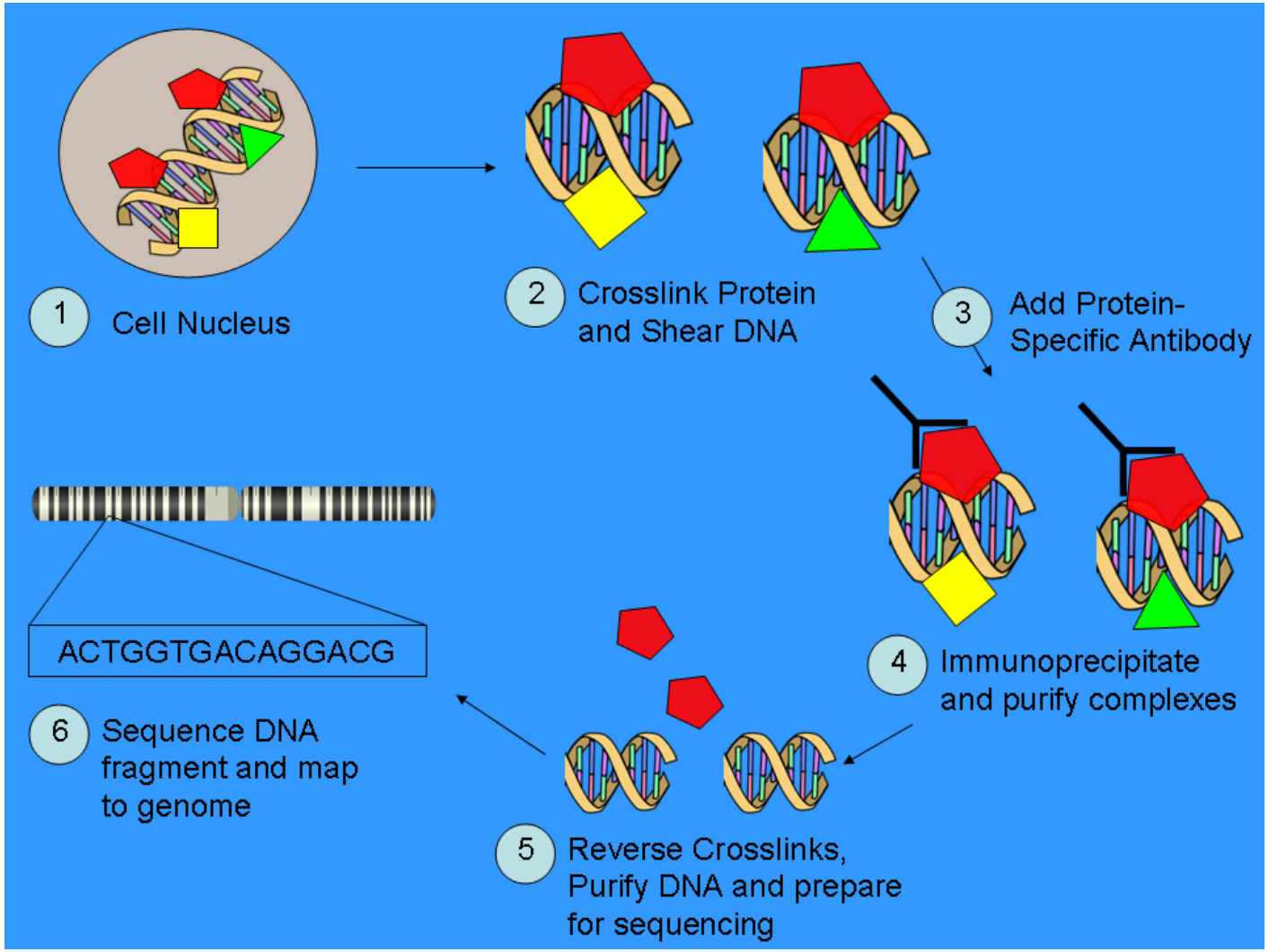
How can histone modification data be obtained?

ChIP-Seq

- Chromatin-Immunoprecipitation followed by (high throughput) sequencing[1]
- many advantages to ChIP-chip (e.g., base pair resolution, no hybdrization issues)
- generally used to identify interactions between proteins and DNA such as transcription factor binding sites
- traditional ChIP protocol can be modified to obtain genome-wide maps of histone modifications or nucleosome positions

[1] Park 2009 - ChIP-seq: advantages and challenges of a maturing technology

Traditional ChIP-Seq workflow



Source: Wikipedia

Main ChIP-Seq workflow for histone modifications

1. native chromatin is used as starting chromatin (no cross-linking needed, histone wrap around DNA naturally)
2. chromatin is sheared by micrococcal nuclease digestion and not by sonication (cuts DNA at the linker), yielding DNA fragments of one or a few intact nucleosomes
3. immunoprecipitating the protein of interest using specific antibodies (coupled to agarose or magnetic beads)
4. purifying and identifying the complex-associated DNA using either **high throughput sequencing (ChIP-Seq)**, microarrays (ChIP-on-chip or ChIP-chip), molecular cloning and sequencing, or PCR

ChIP-Seq background estimation

- no background
- model background using statisical methods (based on the sample mitself)
- negative controls
  - Input-DNA
  - Non-specific antibody (e.g., Immunoglobulin G, IgG)
  - Different tissue (more for transcription factor binding sites)

Steps for the analysis of ChIP-Seq data

1. ChIP-Seq and negative control
2. Bioinformatic analyses:
   - Mapping
   - Peak calling
   - correlation analysis (with respect to other modifications or gene annotation, for example)

Mapping

Important issues

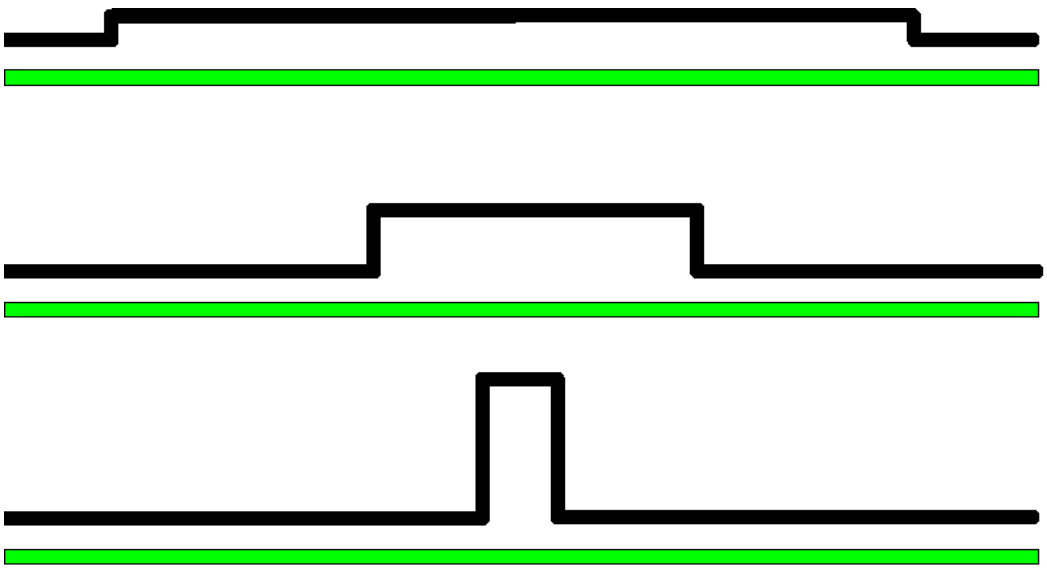
- How to handle non-uniquely mapping reads?
- How to handle reads that have no perfect match?
  - Sequencing error?
  - Quality of the reference genome?
  - Mapping algorithmus exact or a heuristic?
  - SNP?
- Which percentage of the genome is mappable at all with the given read length and the specific parameters?

Peak calling

- extract signal from background noise and identify enriched regions in the genome
- very important, because the majority of the signal may come from noise
- a lot of technical issues influence the results

Peak calling

A lot of bioinformatics programs are available, all with their own flavour and application area.

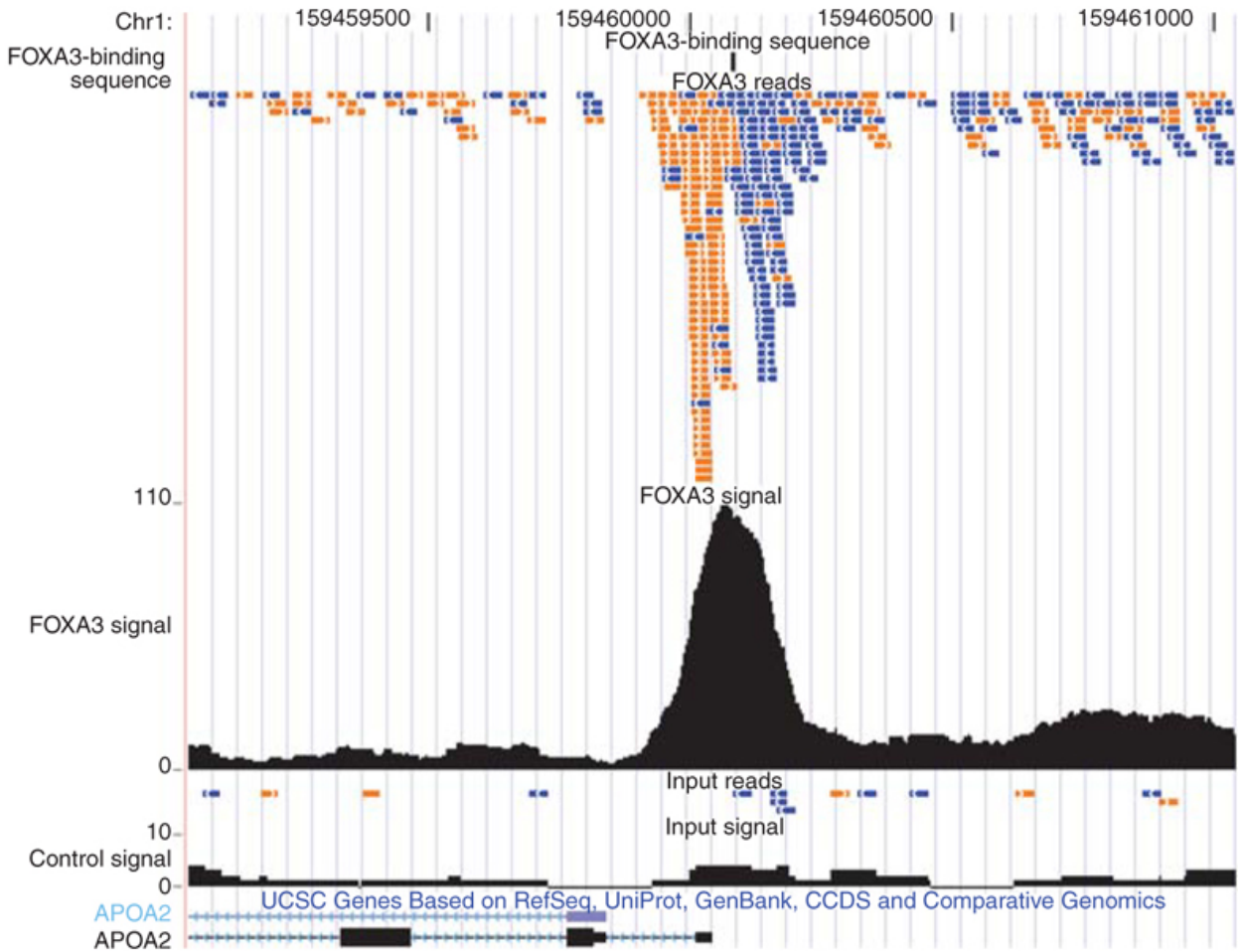


Long-range
(e.g., histone modifications)

Mid-range
(e.g. polymerase binding)

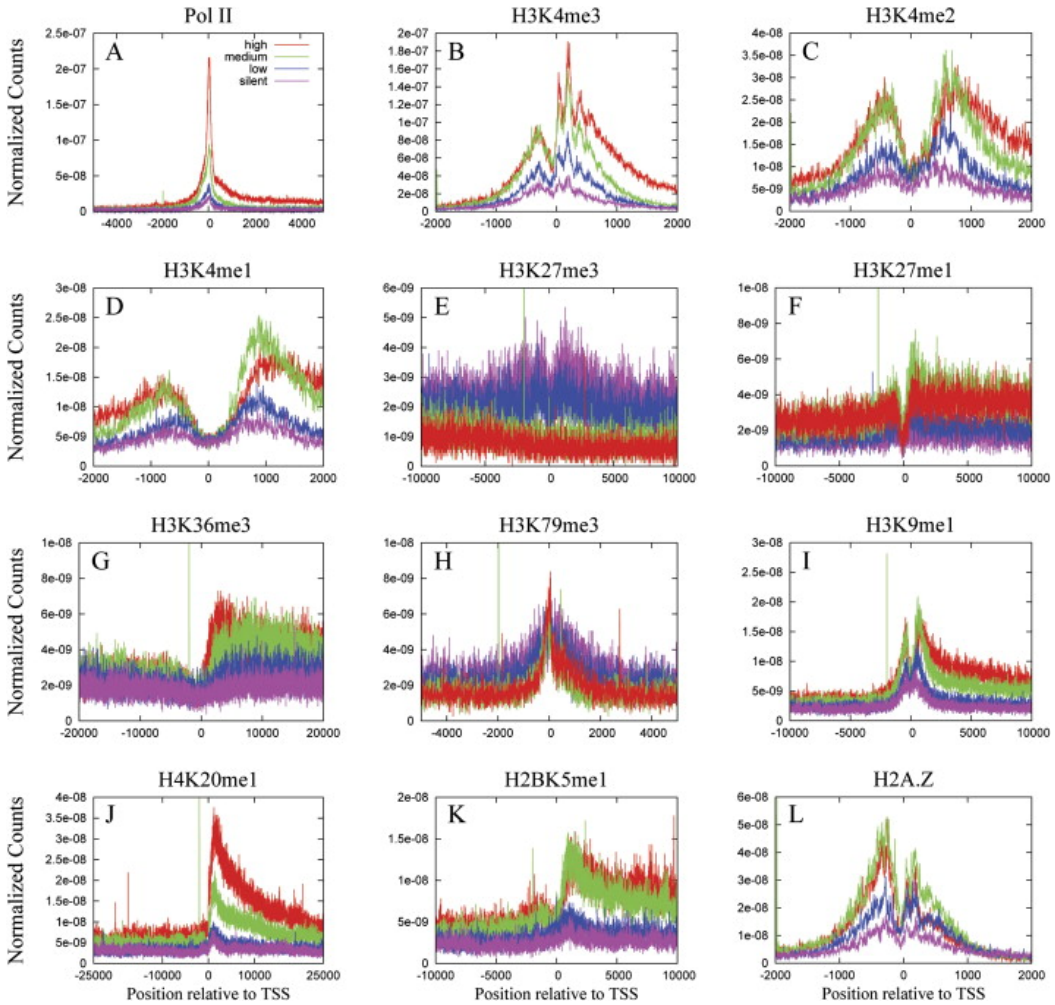
Punctate
(e.g. TF binding)

Peak calling example



Source: Shah 2009. Chromatin immunoprecipitation sequencing (ChIP-Seq) on the SOLiD system.

Bioinformatics analyses of histone modifications



Source: Barski et al. 2007. High-Resolution Profiling of Histone Methylations in the Human Genome.

3. Summary and Take-Home Message(s)

Summary and take-home message(s)

- Epigenetic mechanisms are inherently involved in gene regulation.
- It has yet to be determined if transcription is cause or consequence of epigenetic mechanisms.
- Bioinformatics methods are crucial for the analysis of epigenetic data, which is peppered with a multitude of issues and limitations

Thank you for your attention!

Are there any questions?

Enjoy the long weekend!