Epigenetics

02-715 Advanced Topics in Computational Genomics
Epigenome Overview

- Two-meter long DNA sequences are packaged into each cell in such a way that allows various transcription factors and enhancers to access the DNA for gene expression.
Epigenome Overview

• Epigenome: combination of all chromatin modifications in any given cell type
  – DNA methylation
  – Post-translational histone modifications primarily at N-terminal tails
    • Methylations, acetylation, phosphorylation, ubiquitylation, ADP-ribosylation
  – Nucleosome positioning

• Highly dynamic and governed by a complex interplay of genetic and environmental factors
Nucleosomes and Histones

• Nucleosome: a unit of DNA packaging

• Nucleosome core particle
  – a disc-shaped histone core around which the DNA is wrapped tightly
    1.7 turns in a left-handed coil
  – The structure was solved in 1997
  – Histones are among the most highly conserved eucaryotic proteins.
Nucleosomes and Histones
Structure of Core Histones
Assembly of a Histone Octamer on DNA
Nucleosomes and Histones

• Histone fold region and interaction between DNA and histones
  – 142 hydrogen bonds between DNA and the histone core in each nucleosome
  – More than one-fifth of the amino acids in each of the core histones are either lysine or arginine, and their positive charges neutralize the negatively charged DNA backbone

• Histone N-terminals, called tail, extends out from DNA-histone core
  – Subject to various modifications
Heterochromatin

- Heterochromatin: a tightly packed form of chromatin
- Genes located in heterochromatin are turned off
Histone Modification

- Serine Phosphorylation adds a negative charge to a histone
Histone Modification

• Lysine acetylation and methylation as competing reactions
  – Acetylation removes the plus charge on lysine
  – Different binding proteins recognize different modifications
Histone Modification

- Lysine (K) and arginine (R) can be methylated in several different ways
- Some positions can be modified by either methylation or acetylation but not both
Models of the Functions of Histone Modification

• Charge neutralization
  – The acetylation of histones neutralizes positive charges on DNA and phosphorylation adds a negative charge

• Histone code
  – Control of gene regulation
  – Multiple histone modifications can combinatorially or sequentially interact

• Signaling pathway
  – Histone modifications as signaling platform to facilitate binding of enzymes for their function on chromatin
DNA Methylation

• DNA itself can be covalently modified

• Methylation in cytosine (C)
  – pays a role in gene regulation
  – No impact on DNA base pairing
  – Observed in CG sequence (based-paired with GC on the opposite strand)

• How the DNA methylation pattern is established
  – Genome-wide demethylation, shortly after fertilization
  – Denovo DNA methyltransferases establish new methylation patterns during development
  – Once a methylation pattern is established, it is inherited through cell division within a tissue type by maintenance DNA methyltransferase
Methylation

- Methylation of cytosine (C) nucleotides in DNA
Methylation

- Methylation is maintained in DNA replication by maintenance methyltransferase – the same methylation pattern in the same tissue type
Methylation

- Associated with repressing gene expression
  - X-chromosome silencing
  - Promoter regions of expressed genes in a tissue are usually unmethylated
  - Different methylation patterns have been observed in Identical twins – environmental effects
CpG Islands and Methylation

• Methylated C nucleotides tend to be eliminated in the course of evolution (tends to mutate to T)
  – Methylated cytosines are mutated to T, but not repaired, while mutations of unmethylated cytosines are recognized by DNA repair enzymes and repaired
  – Many C nucleotides have been lost during evolution
-- CpG islands often occur in functional regions of DNA and stay unmethylated
-- CpG islands often occur in the promoters of the housekeeping genes
Methylation

- Black lines: CG dinucleotide
- Red lollipos: methylated Cs
- CG islands occur in transcribed regions and are conserved during evolution
Genome-wide Detection of DNA Methylation

**a** Restriction enzyme

1. 5' - 5' - HpaII
2. Msel digestion
3. Adaptor ligation
4. HpaII digestion

**b** Bisulphite treatment

1. Bisulphite treatment
2. PCR
3. Immunoprecipitation
4. Amplification

**c** mCIP

1. Sonication
2. Immunoprecipitation

Genomic coordinates
ChIP-Seq for Detecting Histone Marks

- Immunoprecipitation
- DNA purification
- End repair, adaptor ligation
- Cluster generation
- Sequence and map reads to reference genome
- Genomic coordinates
Linked DNA Methylation and Histone Modification

- Establishment of bimodal methylation
Linked DNA Methylation and Histone Modification

• Turning off pluripotent genes
Key Research Questions

• A complete catalogue of all possible histone modifications

• How does the epigenome influence gene regulation?

• How is the nucleosome positioning determined?
  – DNA sequence can influence nucleosome positions

• What is the role of epigenome in complex diseases, tumorigenesis?

• How are various epigenetic marks correlated? What are their combinatorial effects?

• Can we infer chromatin states given epigenetic marks?
  – Chromatin states may be associated with functions
Epigenome and GWAS

• Genome information explains only the small part of heritability.

• There is an increasing evidence that the epigenome plays an important role in complex diseases and cancer
Challenges in Epigenome Association Mapping

• The origin of epigenetic variation is mostly unknown
  – Heritable vs. non-heritable variation.
  – Early development vs. later in the development
    • Soma-wide vs. tissue-specific variability
  – Environmentally-induced variability (diet, smoking, etc.)
    • smokers have a greater epigenetic modification before lung diseases

• Hard to establish causality
  – Epigenetic aberration could be the cause of disease or the result of disease
    • Longitudinal study to disentangle the difference
  – Epigenetic aberration could result from the genome aberration
    • Twin studies can be useful but datasets are limited
### Study Designs for Epigenome Association Mapping

<table>
<thead>
<tr>
<th>Design</th>
<th>Key advantage</th>
<th>Key disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case versus control</td>
<td>Many cohorts exist</td>
<td>Cannot easily control for environmental and genetic confounders</td>
</tr>
<tr>
<td>Families</td>
<td>Can study potential inheritance</td>
<td>Few large cohorts of this type exist</td>
</tr>
<tr>
<td>Disease-discordant monozygotic twins</td>
<td>Can control for genetics</td>
<td>Few large cohorts of this type exist</td>
</tr>
<tr>
<td>Prospectively sampled, longitudinal</td>
<td>Can establish causality</td>
<td>Slow and difficult to establish</td>
</tr>
</tbody>
</table>
Combining Genetic and Epigenetic Association Mapping

• Genetic variations and epigenetic variations can influence the phenotype independently

• Genetic variations can induce phenotypic variations and epigenetic variations