Goals for lecture

Key concepts

• Importance of epigenetic data for understanding transcriptional regulation
• Use of epigenetic data for predicting transcription factor binding sites
Identical DNAs but identical fates?

Chromosomes

PNAS July 26, 2005 102 (30) 10604-10609; https://doi.org/10.1073/pnas.0500398102
Defining epigenetics

- Formally: attributes that are “in addition to” genetic sequence or sequence modifications
  - “Epigenetic code” (vs. genetic code)
- Informally: experiments that reveal the context of DNA sequence
  - DNA has multiple states and modifications

\[
\text{G A C T A G T G C G T T T A C T}
\]

vs.

\[
\text{G T G C G T T T A C T}
\]

inaccessible

Histones

modification
Chromatin packages DNA around Histones

Importance of epigenetics

Better understand

• DNA binding and transcriptional regulation
• Differences between cell and tissue types
• Development and other important processes
PWMs are not enough

- Genome-wide motif scanning is imprecise
- Transcription factors (TFs) bind < 5% of their motif matches
- Same motif matches in all cells and conditions
PWMs are not enough

- DNA looping can bring distant binding sites close to transcription start sites
- Which genes does an enhancer regulate?

Enhancer: DNA binding site for TFs, can be far from affected gene

Promoter: DNA binding site for TFs, close to gene transcription start site
Mapping regulatory elements genome-wide

- Can do much better than motif scanning with additional data
- ChIP-seq measures binding sites for one TF at a time
- Epigenetic data suggests where some TF binds

Shlyueva Nature Reviews Genetics 2014
DNase I hypersensitivity

- Regulatory proteins bind accessible DNA
- DNase I enzyme cuts open chromatin regions that are not protected by nucleosomes

Nucleosome: DNA wrapped around histone proteins

Wang *PLoS ONE* 2012
Histone modifications

- Mark particular regulatory configurations

Two copies of histone proteins H2A, H2B, H3, H4

- H3 (protein) K27 (amino acid) ac (modification)

Latham *Nature Structural & Molecular Biology* 2007; Katie Ris-Vicari

Shlyueva *Nature Reviews Genetics* 2014
DNA methylation

- Reversible DNA modification
- Represses gene expression

Gene “switched on”
- Active (open) chromatin
- Unmethylated cytosines (white circles)
- Acetylated histones

Gene “switched off”
- Silent (condensed) chromatin
- Methylated cytosines (red circles)
- Deacetylated histones
3D organization of chromatin

- Algorithms to predict long range enhancer-promoter interactions
- Or measure with chromosome conformation capture (3C, Hi-C, etc.)

Rao Cell 2014
3D organization of chromatin

- Hi-C produces 2D chromatin contact maps
- Learn domains, enhancer-promoter interactions

Rao Cell 2014
Next Generation Sequencing (NGS) for epigenomics

Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)
Large-scale epigenetic maps

- Epigenomes are condition-specific
- Roadmap Epigenomics Consortium and ENCODE surveyed over 100 types of cells and tissues
Genome annotation

- Combinations of epigenetic signals can predict functional state
  - ChromHMM: Hidden Markov Model
  - Segway: Dynamic Bayesian network
Genome annotation

- States are more interpretable than raw data

Ernst and Kellis *Nature Methods* 2012
Predicting TF binding with DNase-Seq
DNase I hypersensitive sites

- Arrows indicate DNase I cleavage sites
- Obtain short reads that we map to the genome

Wang PLoS ONE 2012
DNase I footprints

- Distribution of mapped reads is informative of open chromatin and specific TF binding sites

- TF binding prevents DNase cleavage leaving Dnase I “footprint”, only consider 5’ end
DNase I footprints to TF binding predictions

• DNase footprints suggest that some TF binds that location

• We want to know which TF binds that location

• Two ideas:
  – Search for DNase footprint patterns, then match TF motifs
  – Search for motif matches in genome, then model proximal DNase-Seq reads

We’ll consider this approach