How to make (more) sense of chromatin state data

Wouter Meuleman
Kellis Lab (MIT/Broad)
Although a great summary, can still be hard to interpret / work with, especially as we’re moving towards 1,000s of epigenomes.
Chromatin state calls across many epigenomes can be viewed as an alignment of sequences with a finite alphabet.

There are good ways of modeling such alignments.
Instead of letters, let's consider an alphabet of colors only.
Logos allow for the inclusion of background frequencies: the expected occurrence of each color can be different
- **surprisal** indicates for each state/color how unlikely/enriched
- Sum at each position == relative entropy (Kullback-Leibler div.)
Some applications of epilogos

interactive visualization

consensus epigenomes

comparative epigenomics

pattern discovery & search
Interactive visualization

Visualization tracks for the WashU epigenome browser

http://compbio.mit.edu/epilogos

Stand-alone as well as embeddable in your own website
Some applications of epilogos

- **interactive visualization**
  - `<image>`

- **consensus epigenomes**
  - `<image>`

- **comparative epigenomics**
  - `<image>`

- **pattern discovery & search**
  - `<image>`
Consensus epigenomes

For each position, the chromatin state with the highest information content forms the ‘consensus’.

\[ \text{CACGTTG} \rightarrow \text{CACGTTG} \]
Interpretation of large-scale (epi)genomic datasets through information-based dimensionality reduction

Thus far: mainly data representation/visualization. Can we also perform analyses using epilogos?
Some applications of **epilogos**

- **Interactive visualization**
- **Consensus epigenomes**
- **Comparative epigenomics**
- **Pattern discovery & search**
Select two subsets of epigenomes, build separate epilogos and compare them.
Application 2: Comparative epigenomics

Comparison: Male vs. Female donors

XIST genes

chromosomal position (chrX, kb)
Some applications of epilogos

- **interactive visualization**
- **consensus epigenomes**
- **comparative epigenomics**
- **pattern discovery & search**
Turning things upside down

In addition to studying single regions across many epigenomes, we can also use epilogos to study many regions in a single epigenome.
Prototype discovery & search: active TSSs

- Single epigenome, many regions
- Use Gibbs sampling / EM to discover ‘prototypes’ for regions
- Regions do not have to be aligned around some landmark
Prototype discovery & search: inactive TSSs

- Discovers multiple mechanisms of silencing (i.e., quiescent, heterochromatin, Polycomb)
- **Prototypes can be used to search in other epigenomes**
- Think of prototyping, e.g., enhancers in cell type X and finding instances in unrelated cell type Y
Some applications of epilogos

- Interactive visualization
- Consensus epigenomes
- Comparative epigenomics
- Pattern discovery & search
epilogos

Visualization and analysis of chromatin state models

- Information-based dimensionality reduction
- Comparative epigenomics
- Prototype discovery and search
  - Currently implementing search / scoring
- Use with custom data
  - Summer student will work on making this more accessible

http://compbio.mit.edu/epilogos
THANKS!

• Soheil Feizi (information theory)
• Nezar Abdennur (TAD boundaries)
• Xin Zhou & Daofeng Li (WashU browser)
• Anshul Kundaje (ENCODE pipeline / Roadmap)
• Jin Wook Lee (Roadmap supplement site)
• Members of Kellis Lab (discussions / ideas)

http://compbio.mit.edu/epilogos

meuleman@mit.edu