Spectacle: fast chromatin state annotation using spectral learning

Jimin Song and Kevin C. Chen
Rutgers University

Kamalika Chaudhuri and Chicheng Zhang
UC San Diego
Chromatin State Annotation via HMMs

• Input: a set of chromatin marks in a cell type
• Output: chromatin states (i.e. colors) representing biological features such as enhancers, promoters, etc.
Overview

- **Spectacle**: predicting chromatin states for a single cell type
  - *Song and Chen, Genome Biology, 2015*

- **Spectacle-Tree**: joint analysis of multiple samples
  - *Zhang, Song, Chaudhuri, Chen, NIPS 2015*
    - *Song, Zhang, Chaudhuri, Chen, in review*

- **Main idea**: Expectation-Maximization is slow and only finds a local optimum, so we use Spectral Learning instead

- Spectacle learns the model ~100 times faster than ChromHMM
Comparison of Chromatin States

(a) ChromHMM

(b) Spectacle
Enrichment of Disease-Associated Variants

• There is higher enrichment of disease SNPs in chromatin state 20 which was found only by Spectacle
Spectacle-Tree: analyzing multiple cell types
Summary of Spectacle-Tree Results

• Tree-HMM took 13 hours
• Spectacle-Tree took 22 min

• Tree-HMM found many null chromatin states
• Spectacle-Tree found biologically significant chromatin states (enhancers, promoters etc.)

• Spectacle-Tree had higher prediction accuracy for promoters than Tree-HMM or Spectacle

• Tree-HMM finds the same chromatin state in all samples
• Spectacle-Tree found poised enhancers in ES cells
Conclusions

• Spectacle analyzes a single cell type or tissue
• Spectacle-Tree analyzes multiple related samples

• Main advantages over ChromHMM and Segway are
  • Faster speed
  • Ability to jointly analyze multiple samples
  • Higher accuracy when a lot of the genome has no biochemical signal

• Software and chromatin state annotations for ENCODE and Roadmap Epigenomics samples are available on Kevin Chen’s web page

• For any questions, please email kcchen@dls.rutgers.edu