Comparison of Histone Protein Locating Algorithms
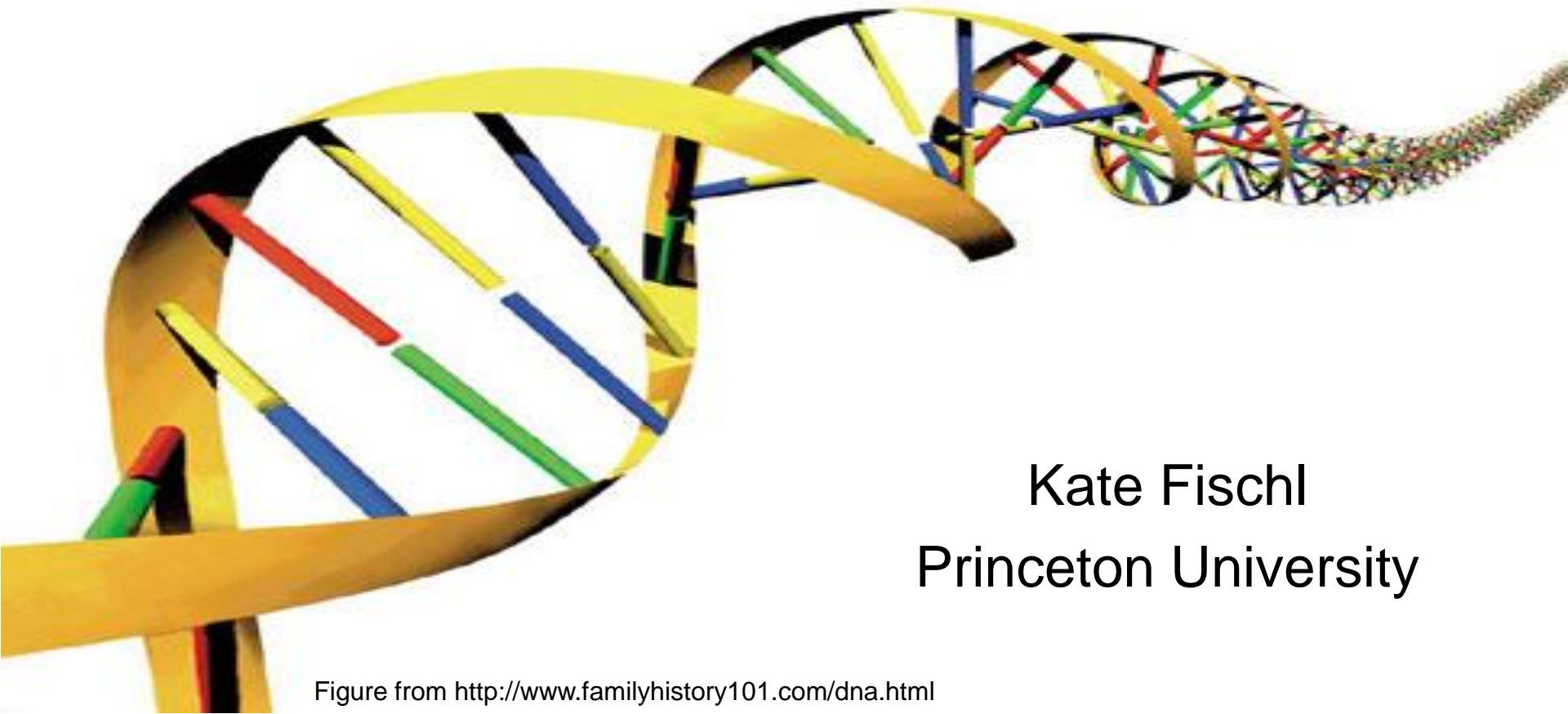
Kate Fischl

Princeton University

Figure from http://www.familyhistory101.com/dna.html

Epigenetics

Epigenetics: study changes in gene expression caused by factors outside of the DNA sequence

Histone Proteins: package and order DNA compactly, as well as play a role in gene regulation

Figure from Qiu 2006.
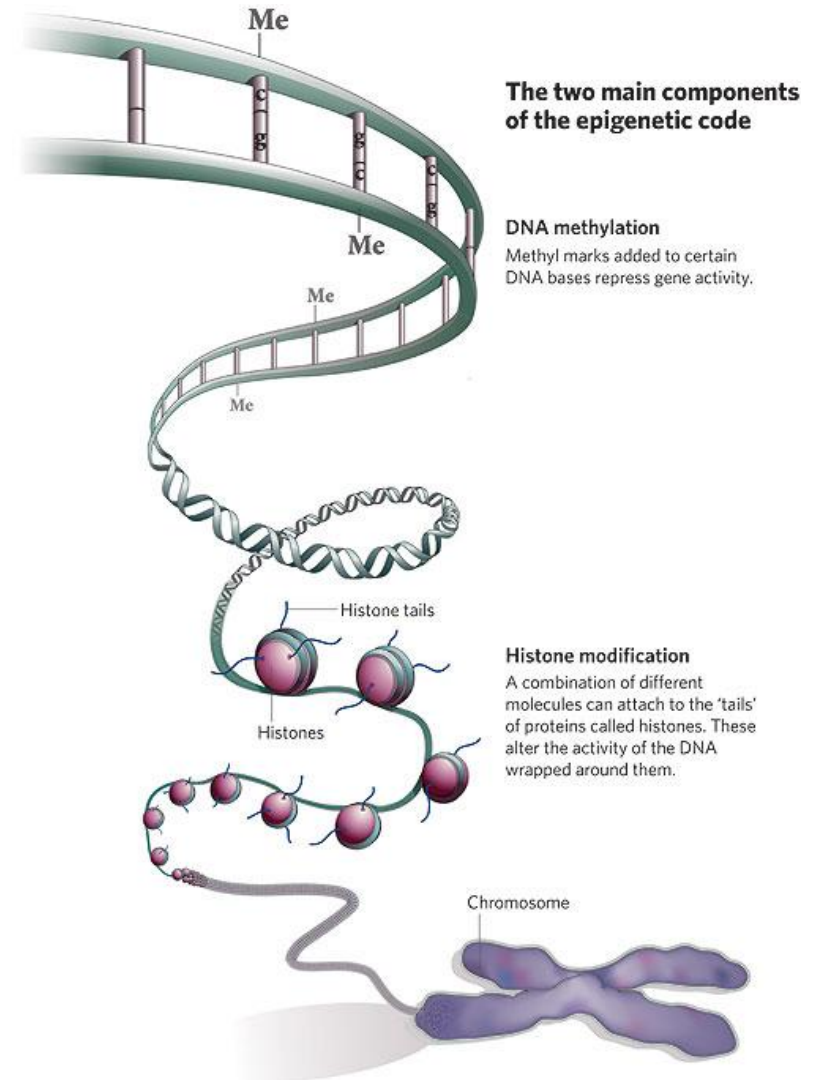
ChIP-Seq

ChIP: Chromatin **I**mmuno**p**recipitation

seq: High-throughput **seq**uencing techniques
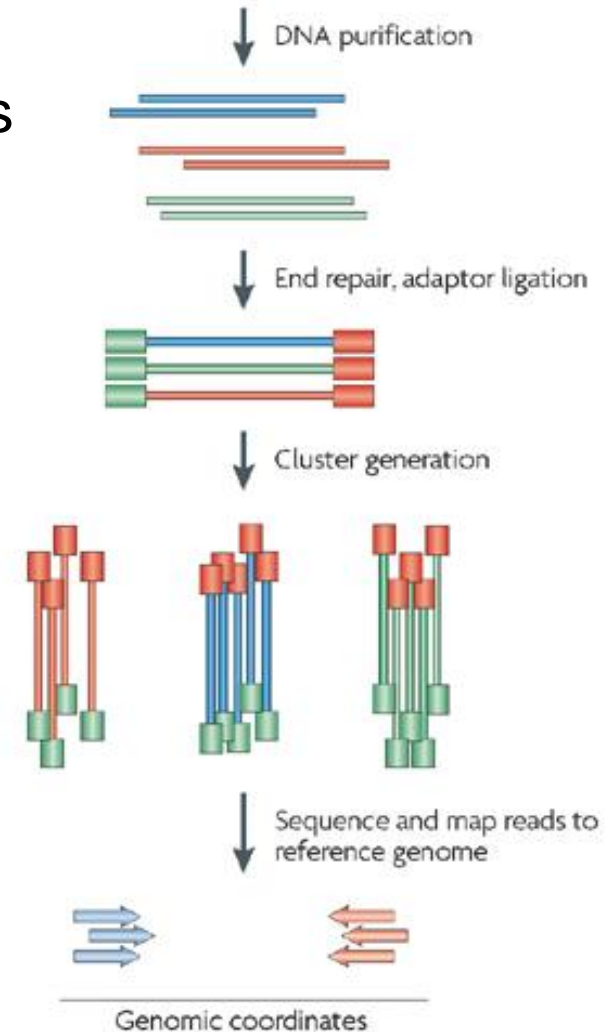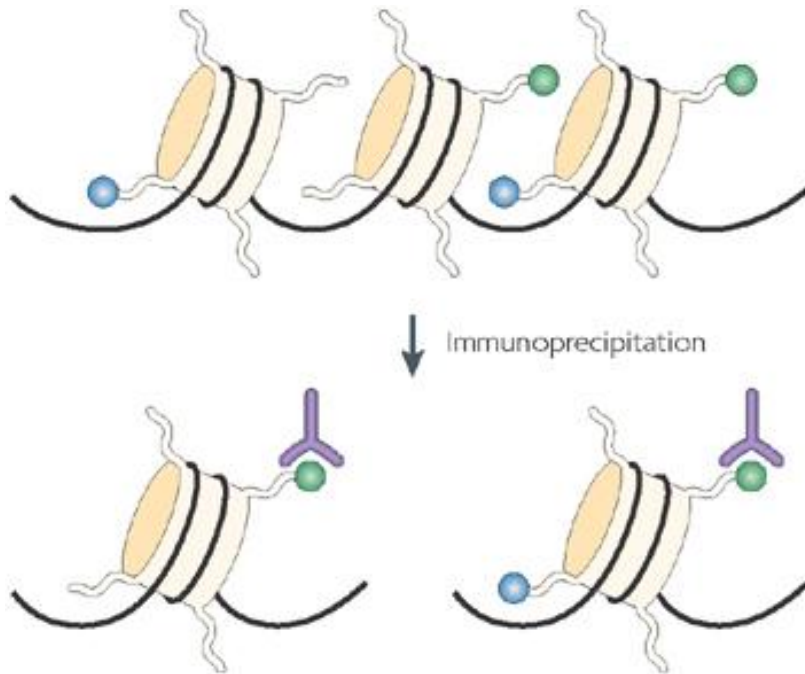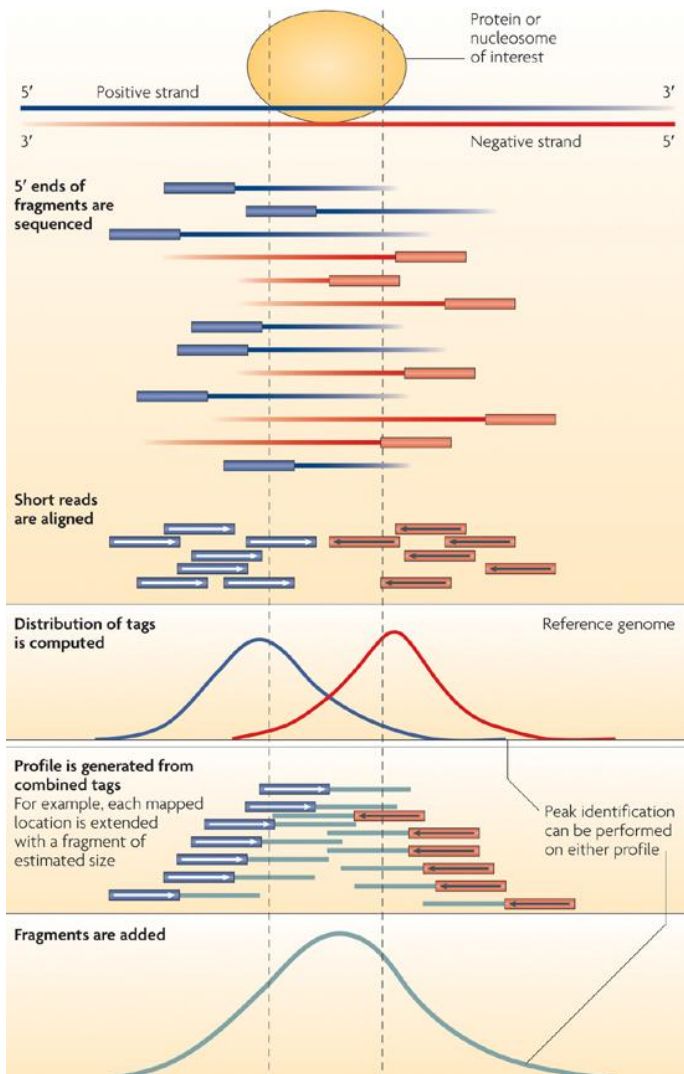

Figure from Schones and Zhao 2008.

Figure from P.J. Park. 2009.

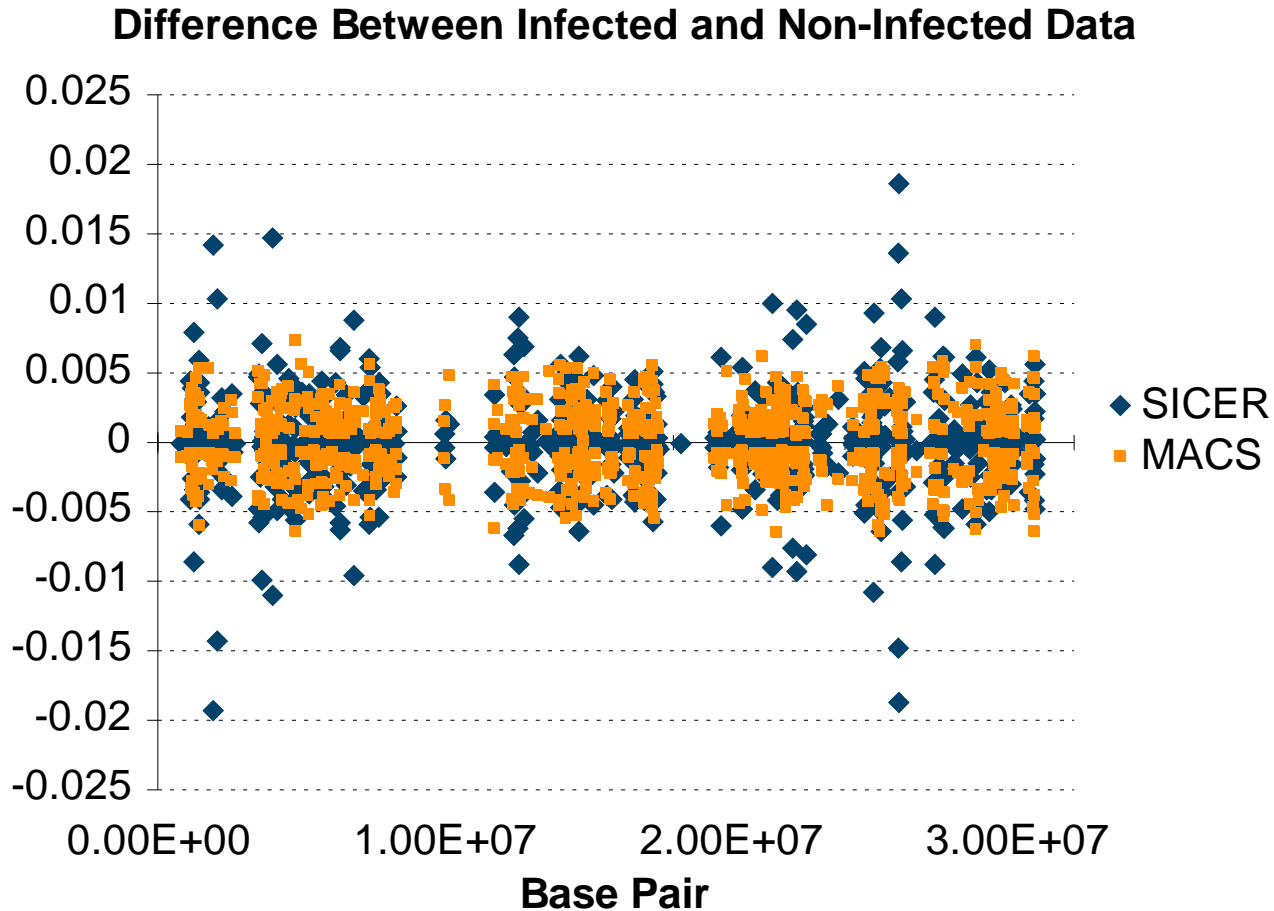• 30+ open source software packages for ChIP-Seq analysis:

 1) Form signal profile from read counts

 2) Identify locations of peaks in signal

 3) Determine significance of peaks

• Focused on understanding a few different algorithms:

 1) **SICER** - spatial clustering
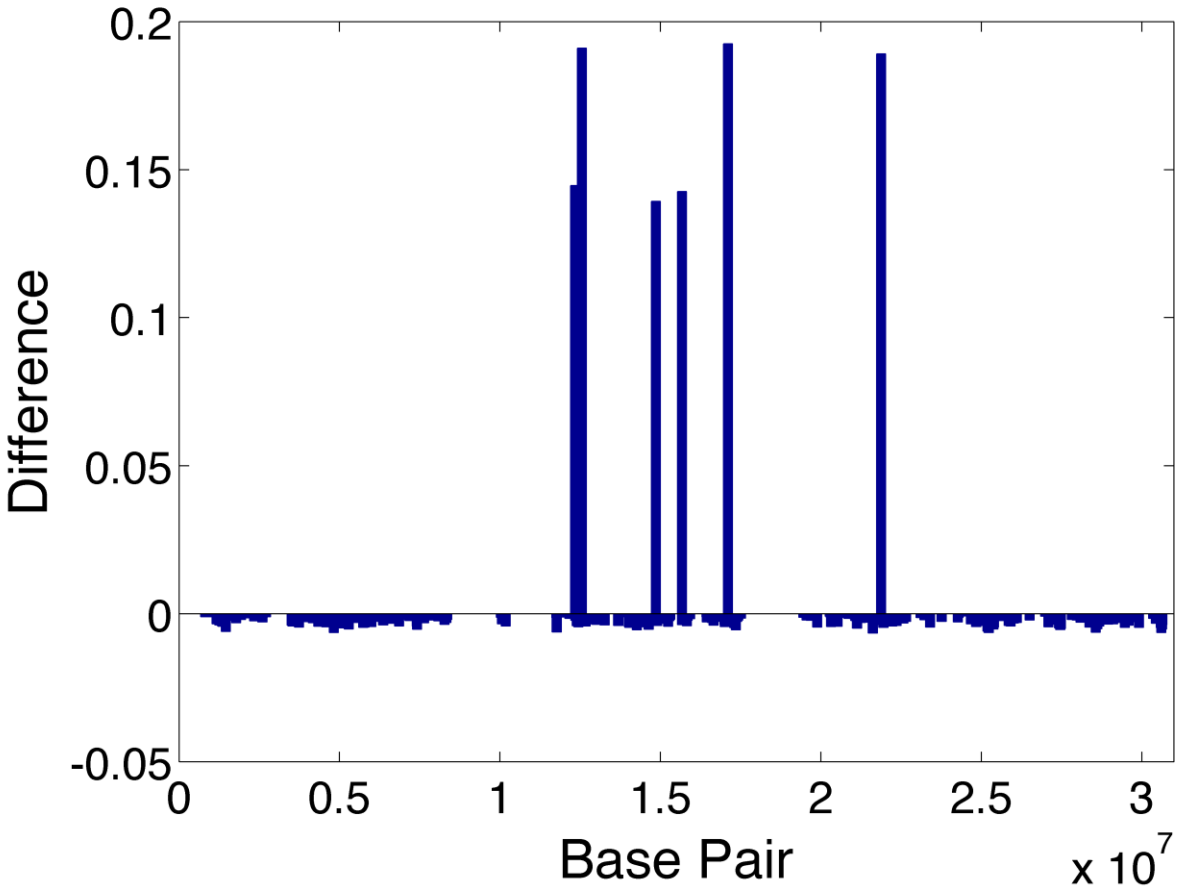
 2) **MACS** - local Poisson distribution

⋮

Results and Findings

Difference Between Infected and Non-Infected Data

• SICER outputs greater differences between non-infected and infected data than MACS

Results and Findings

Role of Control Data in MACS

• Infected data with no control subtracted from infected data with non-infected as control

• MACS identifies areas of difference between non-infected and infected data

• SICER unable to compare the role of control data in this fashion

• Develop more comprehensive measures to compare ChIP-Seq algorithms

• Understand the capabilities and limitations of other existing ChIP-Seq algorithms

• Apply this knowledge to the project here at the University of Maryland

Acknowledgments

- National Science Foundation CISE award #0755224

- BIEN-MERIT program at the University of Maryland

- All of my mentors: Prof. Min Wu, Prof. Jiuzhou Song, Avinash Varna, Wei-Hong Chuang, and Apratim Mitra