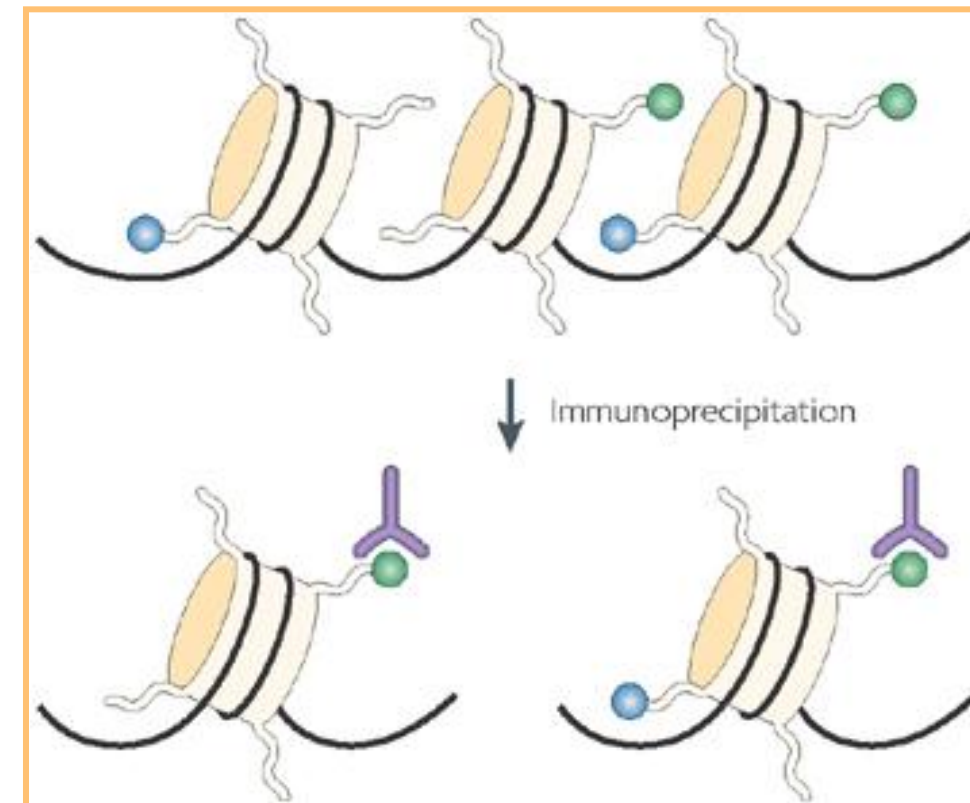


J. Qiu. 2006.

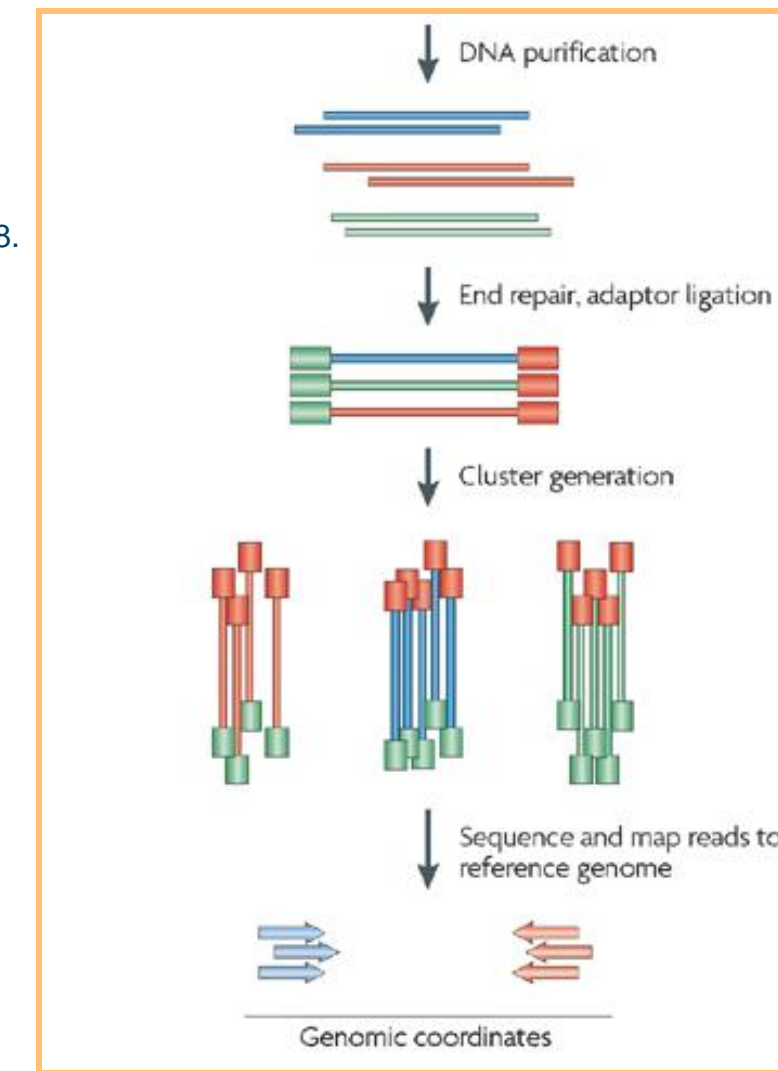
## Introduction

- **Epigenetics:** study changes in gene expression caused by factors outside of the DNA sequence
- Histone protein modifications linked to epigenetics

- ChIP-Seq coupled with processing algorithms locates histone protein modifications along genome



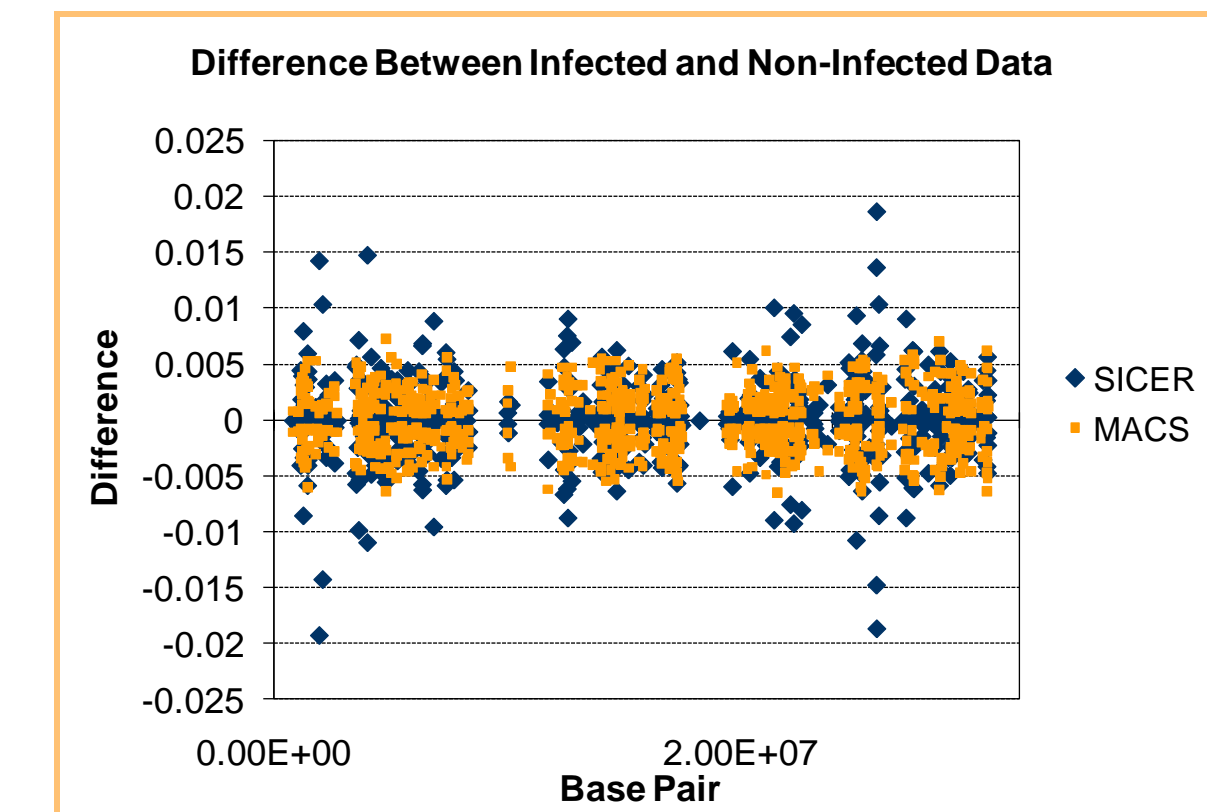
**ChIP: Chromatin Immunoprecipitation**



**seq: High-throughput sequencing techniques**

## Results and Findings

- Data from chickens inbred to have the same DNA
- When exposed to disease some chickens infected, some resistant. Related to histone protein modification location.



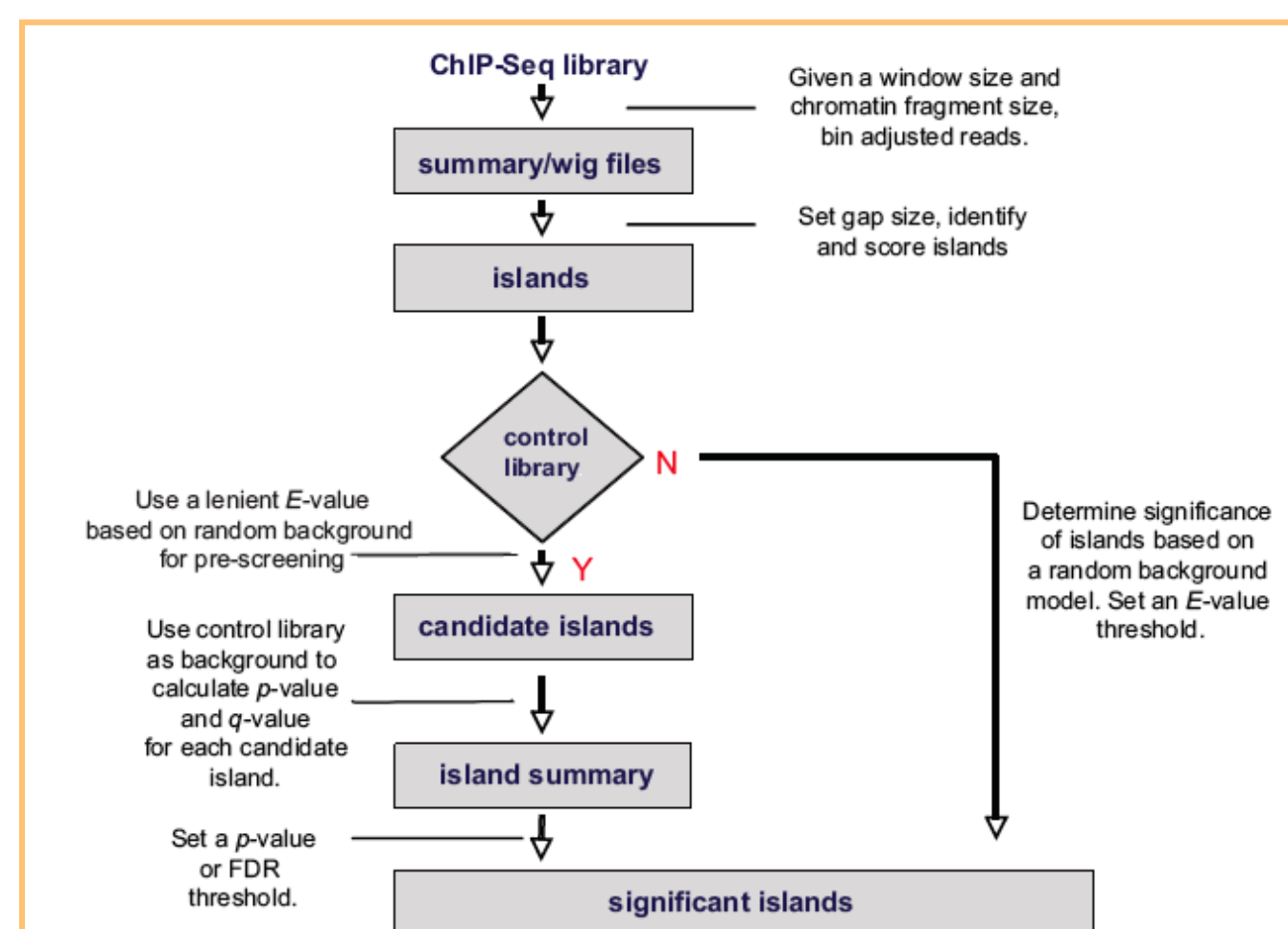
- SICER outputs greater differences between non-infected and infected data than MACS

## Methods

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

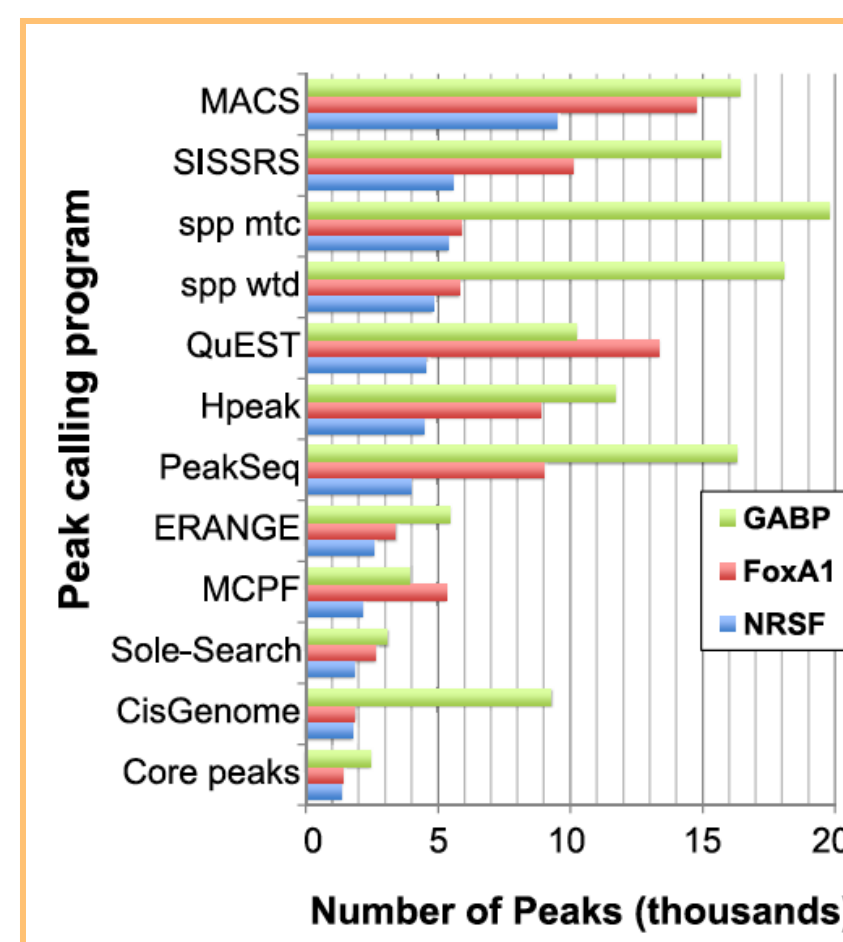
  - 1) Form signal profile from read counts
  - 2) Identify locations of peaks in signal
  - 3) Determine significance of peaks

### SICER Flow Chart



C. Zang, et al. 2009.

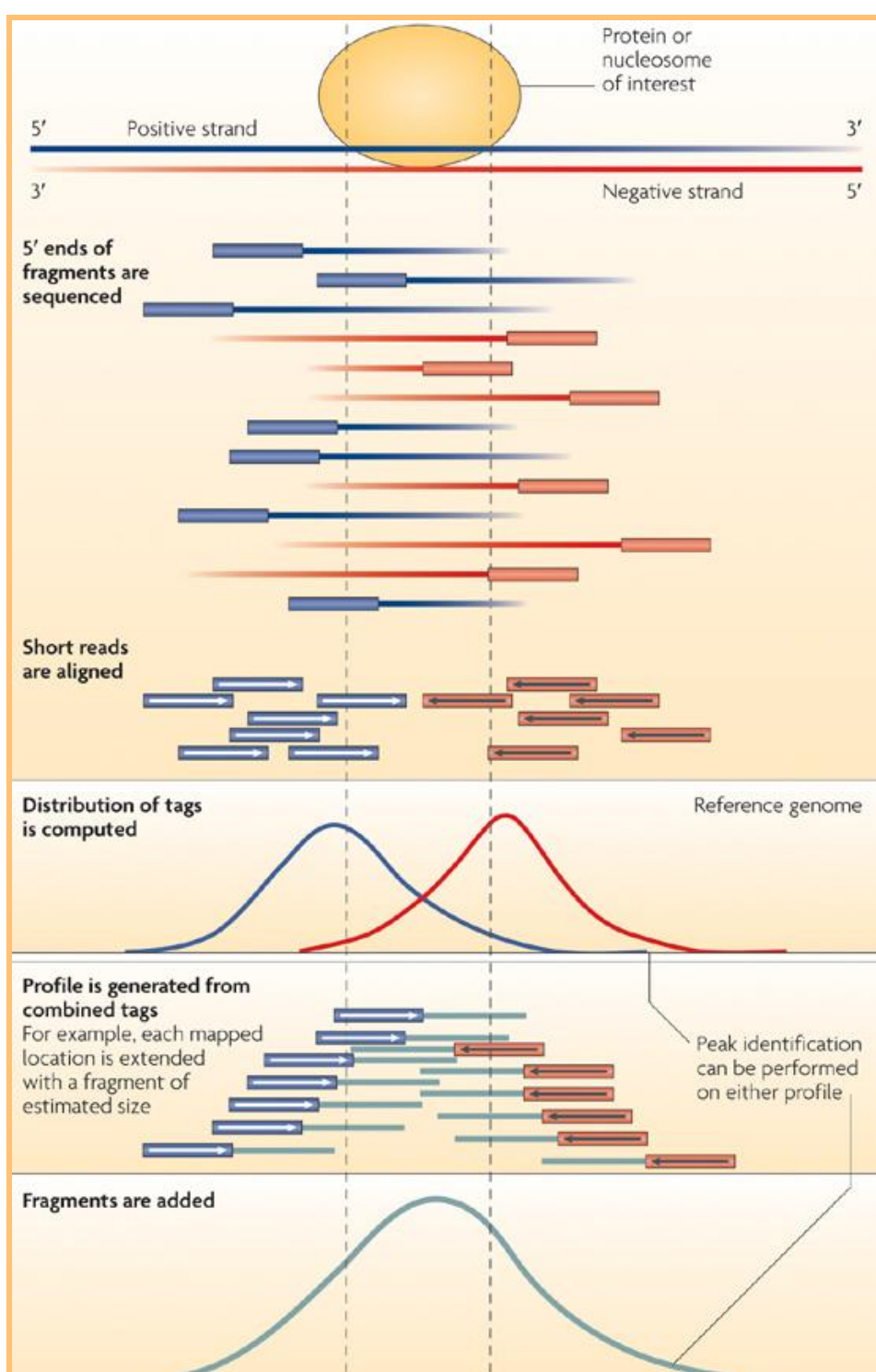
### Peak Identification



E.G. Wilbanks and M.T. Facciotti. 2010.

- Existing studies compare peak identification and significance prediction
- Focused on understanding three main algorithms:

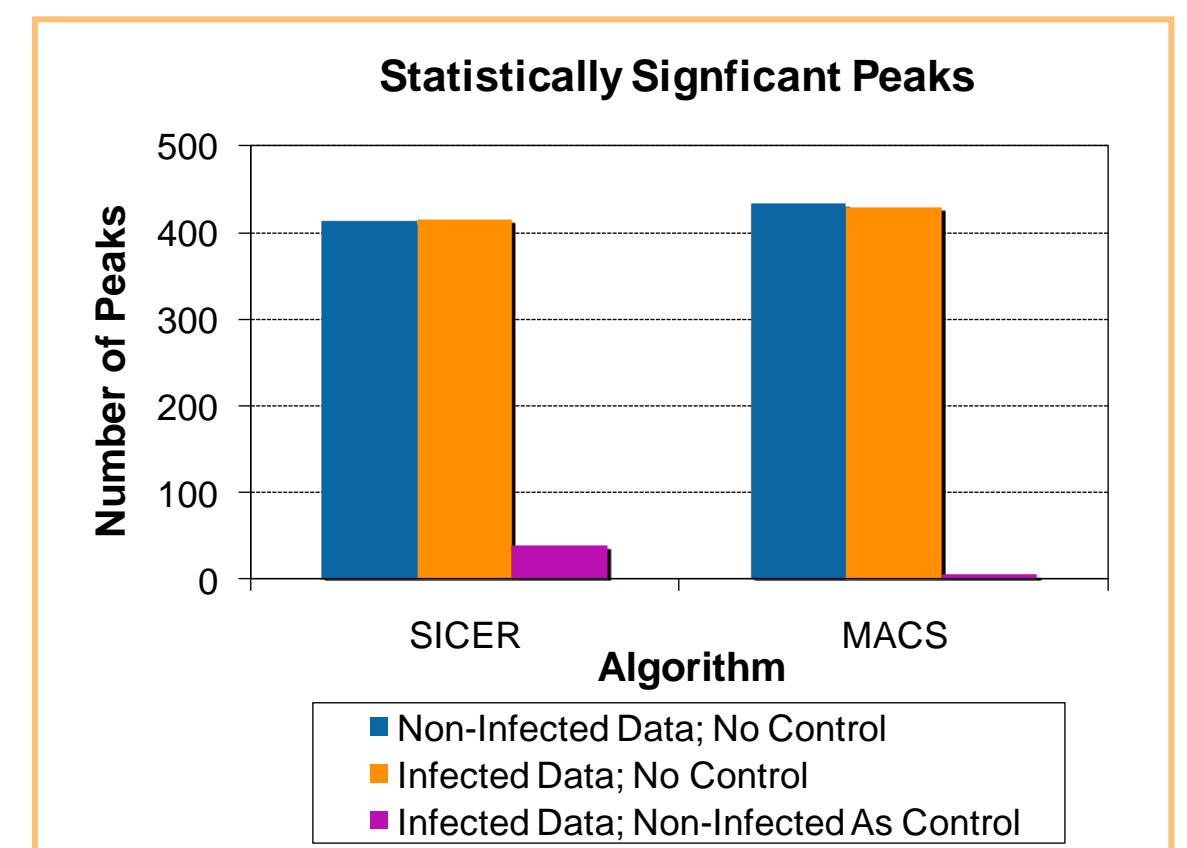
  - 1) SICER - spatial clustering
  - 2) MACS - local Poisson distribution
  - 3) F-Seq - kernel density estimation



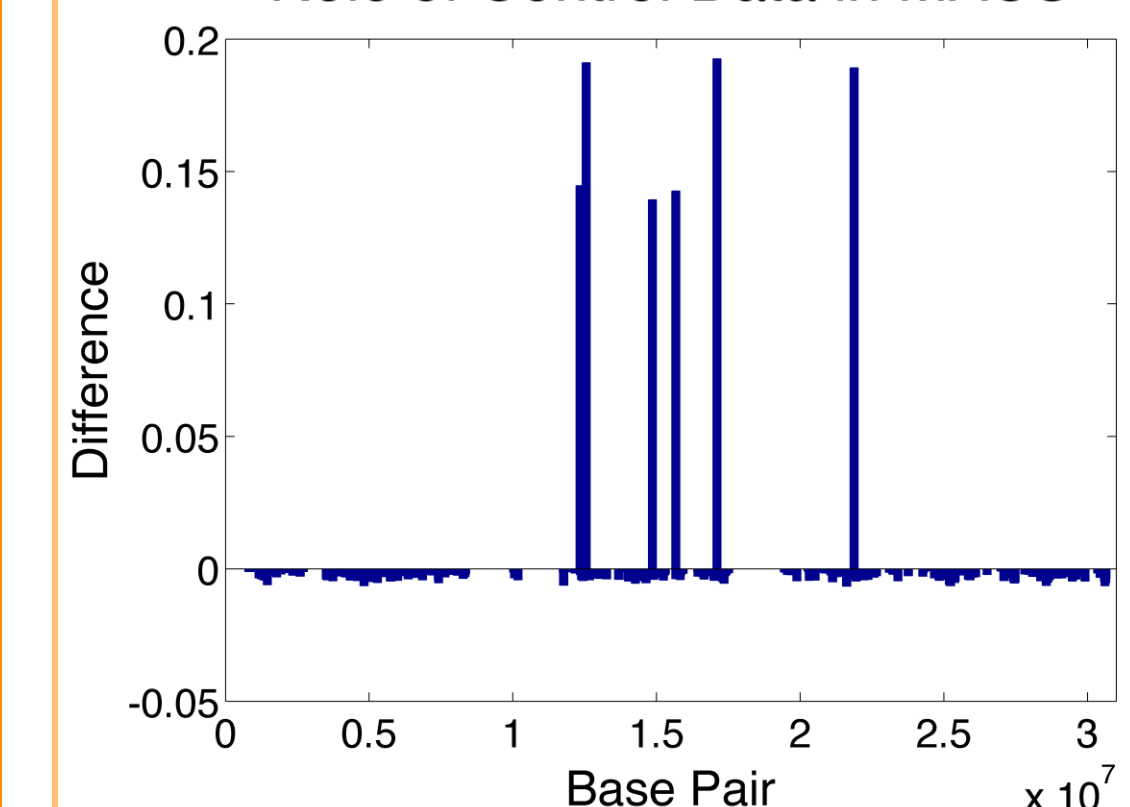
P.J. Park. 2009.

- Peak analysis algorithms use Poisson distributions, negative binomial models, t-distributions, Gaussian kernel density estimation, clustering techniques, Markov models, among others

- Without control data, MACS locates more significant peaks
- With control data, SICER locates more significant peaks



### 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