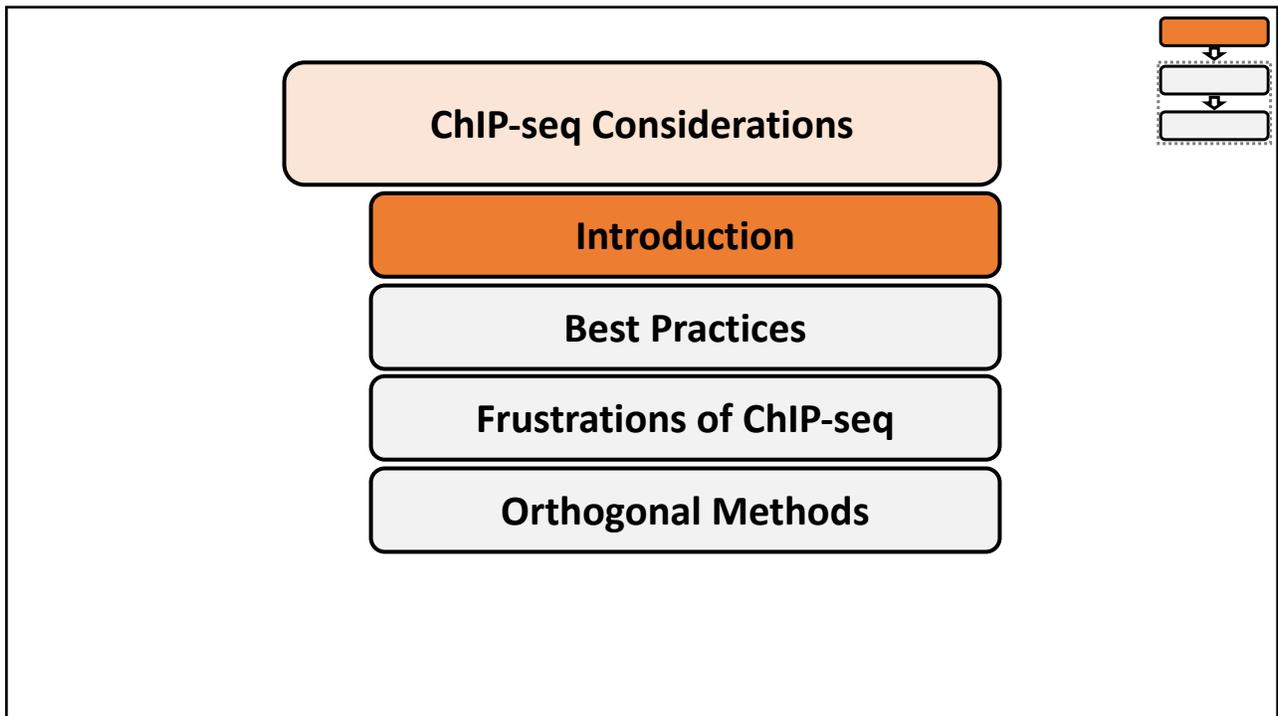
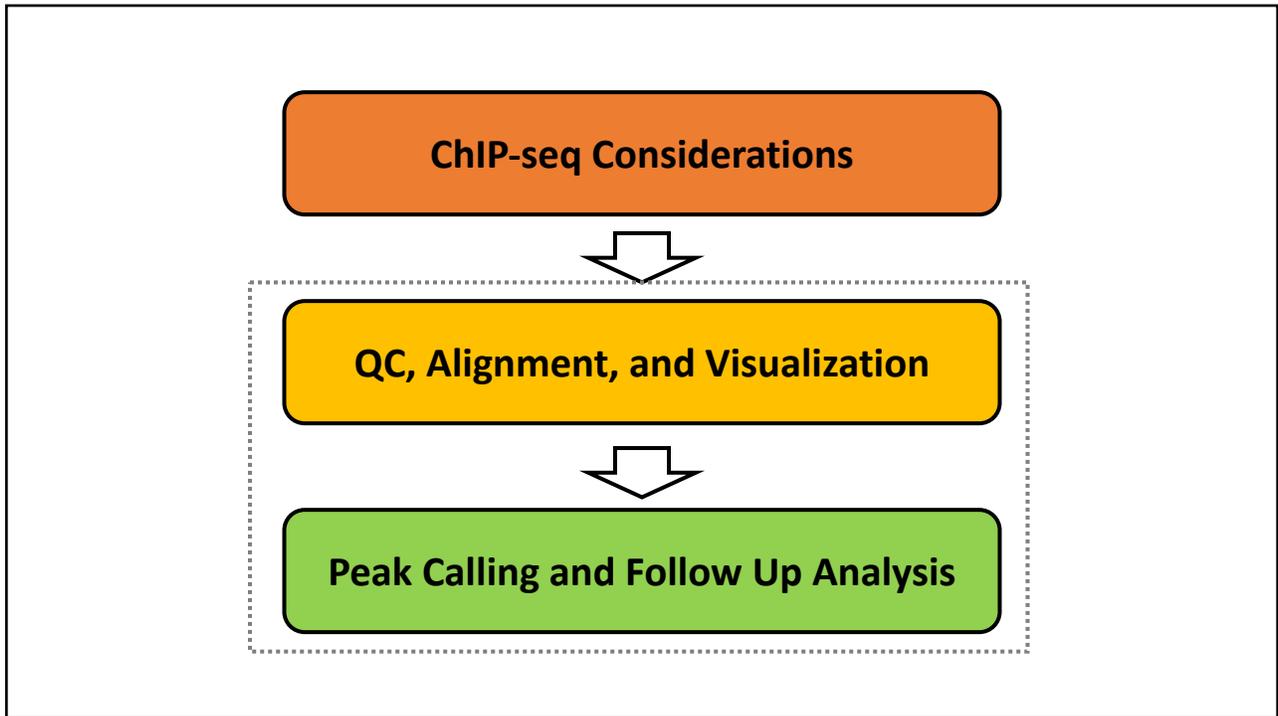


# ChIP-Seq Data Analysis: Probing DNA-Protein Interactions

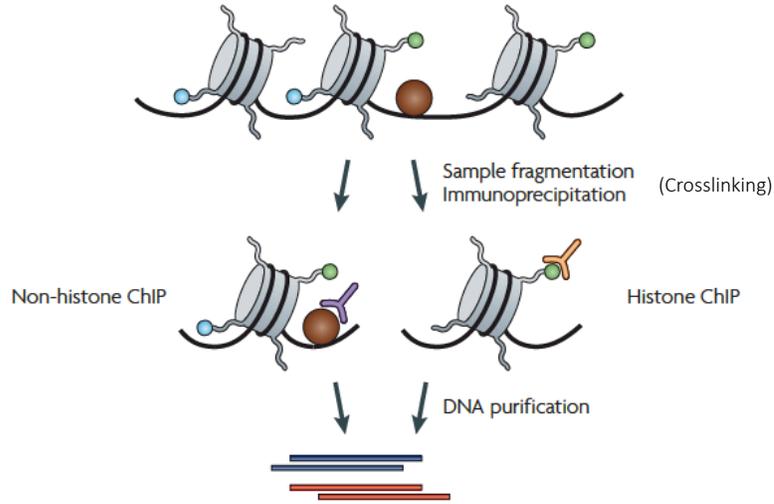
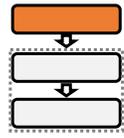
Who are we?

- Paul Schaugency<sup>1,2</sup>
- Tovah Markowitz<sup>1</sup>
- Alexei Lobanov<sup>3</sup>

1. NIAID Collaborative Bioinformatics Resource (NCBR)
2. Center for Cancer Research Sequencing Facility (CCR-SF)  
Bioinformatics
3. Center for Cancer Research Bioinformatics Resource (CCBR)

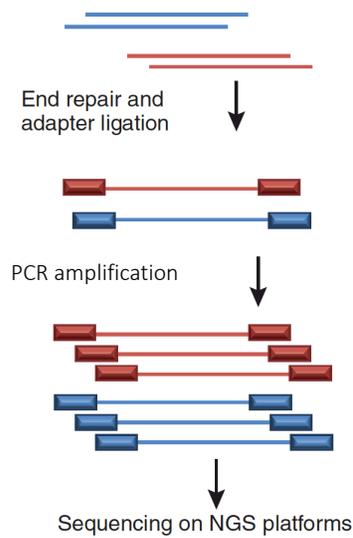
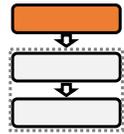


# What is Chromatin ImmunoPrecipitation (ChIP)?



Park, P. Nature Reviews. 2009.

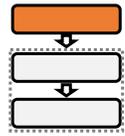
# What is ChIP-seq?



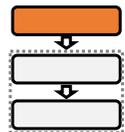
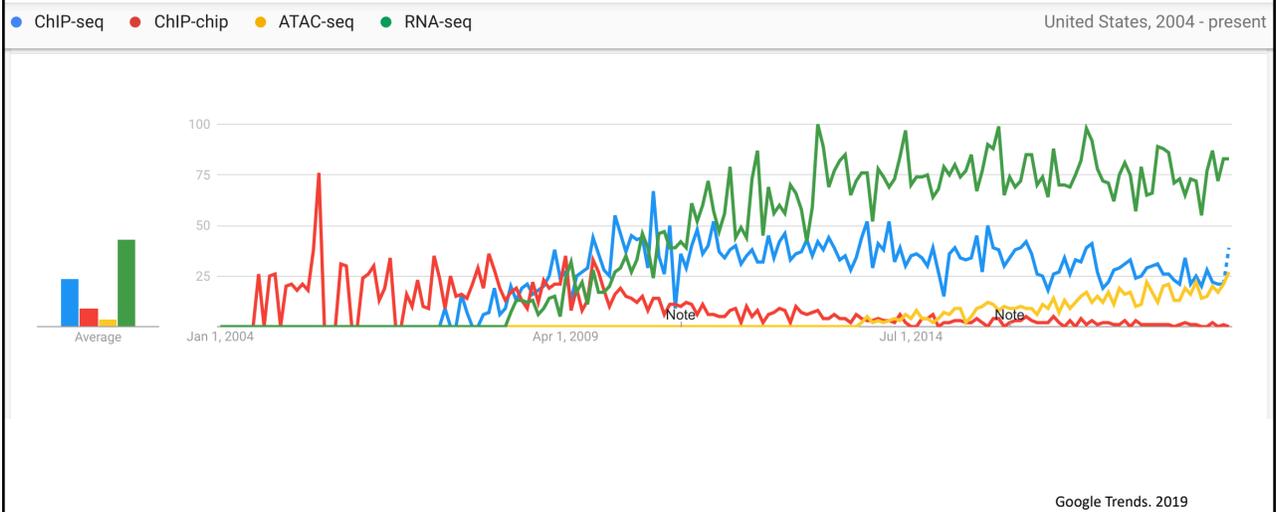
Kidder et. al. Nature Immunology. 2011

## Why do ChIP-seq?

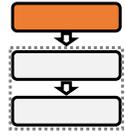
- Define Protein-DNA interactions/Histone modifications across the entire genome and different conditions.
- Define DNA binding sites for DNA-binding proteins.
- Reveal gene regulatory networks when combined with RNA-seq and/or Methylation data.



## Utilization of ChIP-seq

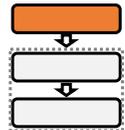


## Adaptations to ChIP-seq



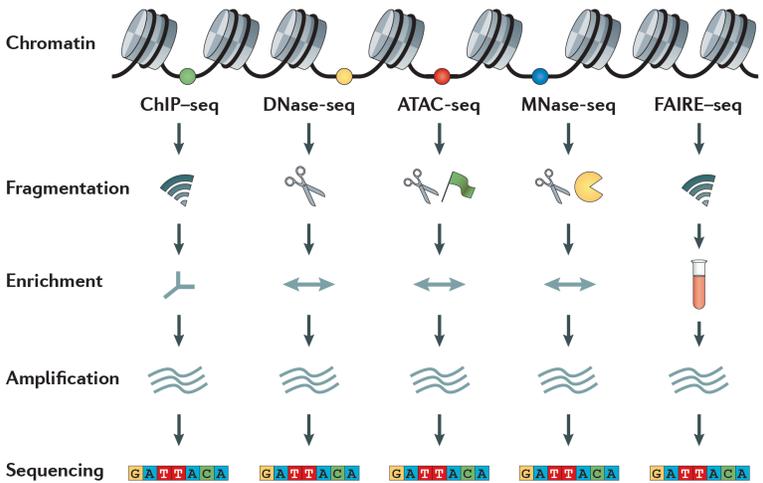
- ChIP-Exo
  - Adds two exonuclease steps to increase the resolution of ChIP
- X-ChIP-seq
  - ChIP-seq with MNase not sonication.
- Highthroughput ChIP (HT-ChIP)
  - ChIP for up to 96 antibodies at once.
- Single cell ChIP (scChIP)
  - ChIP on each individual cell.

## And others...



- AHT-ChIP-Seq
- BisChIP-Seq
- CAST-ChIP
- ChIP-BMS
- ChIP-BS-seq
- ChIPmentation
- Drop-ChIP
- Mint-ChIP
- PAT-ChIP
- reChIP-seq
- CUT & RUN

# Alternative methods to find open chromatin



Meyer and Liu. Nature Reviews. 2014.

## ChIP-seq Considerations

Introduction

Best Practices

Frustrations of ChIP-seq

Orthogonal Methods

## ENCODE best practices

- “ChIP-seq grade” antibodies
- Control samples created in parallel with matching ChIP samples
- At least 2 biological replicates
- Minimum useable reads/fragments
  - Useable reads: uniquely mapped after removal of PCR duplicates
  - For Transcription Factor or other narrow features: 10-15M
  - For Histone or other broad features: > 30M
- No preference between single-end or paired-end sequencing

<https://www.encodeproject.org/about/experiment-guidelines/>

Landt et al 2012. Genome Res

### ChIP-seq Considerations

Introduction

Best Practices

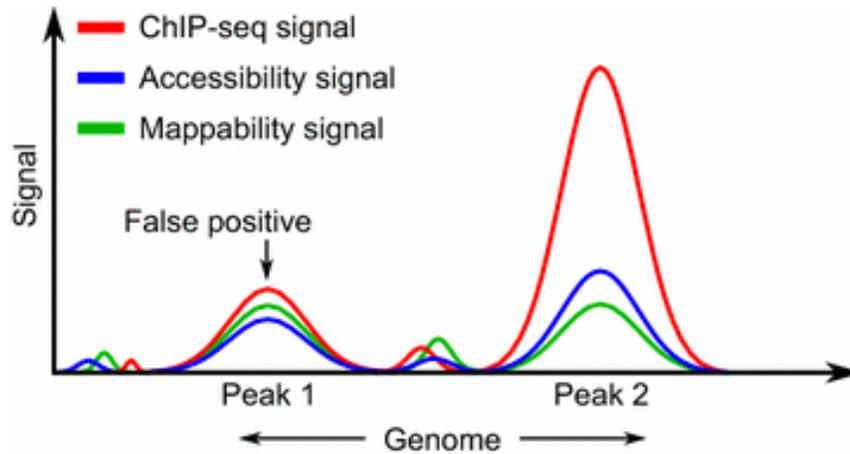
Controls

Replicates

Frustrations of ChIP-seq

Orthogonal Methods

## Why you need controls



Ramachandran et al 2015. BMC Epigenetics & Chromatin

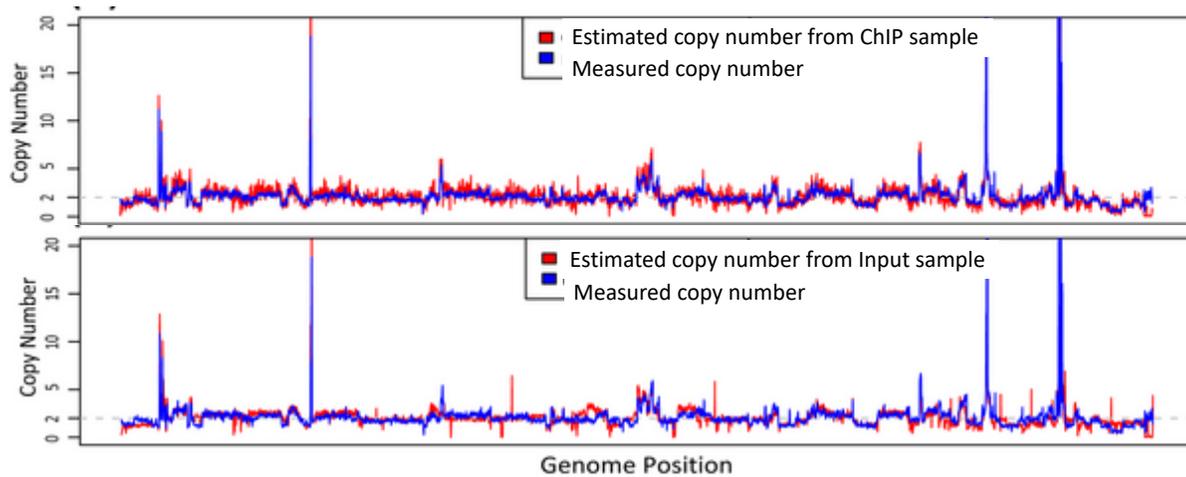
## Biases controls can correct for

- Copy number variation
- Incorrect mapping of repetitive regions
- GC bias
- Non-uniform fragmentation
- Non-specific pull-down

## Types of controls

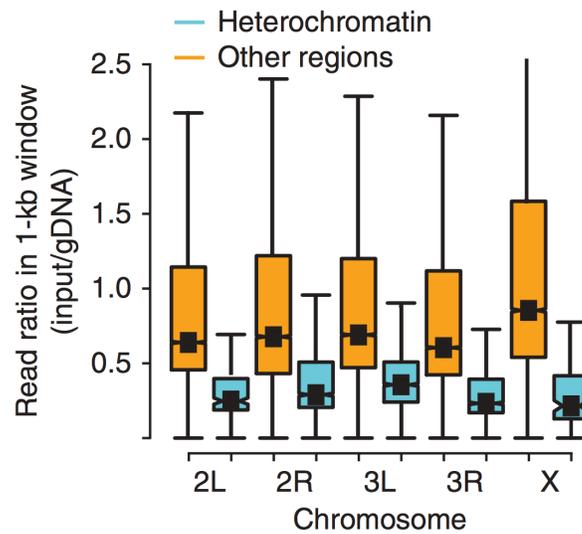
- Input:
  - Crosslink, lyse, and fragment like CHIP but no IP step
- Mock (sometimes also referred to as IgG):
  - Processed like a CHIP sample, but IP without an antibody (just the beads)
- IgG:
  - CHIP with an antibody that has no target within the nucleus of the cells of interest

## Copy number biases



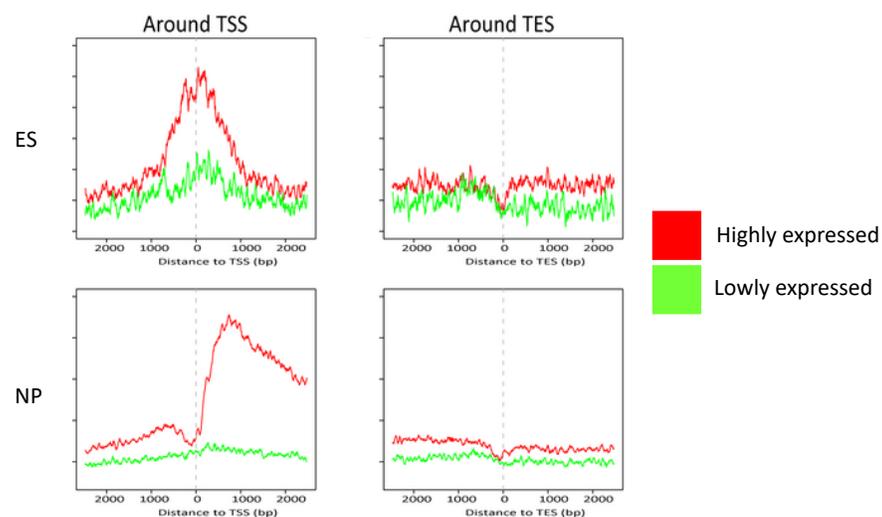
Vega et al 2009. PLOS One

## Heterochromatin biases



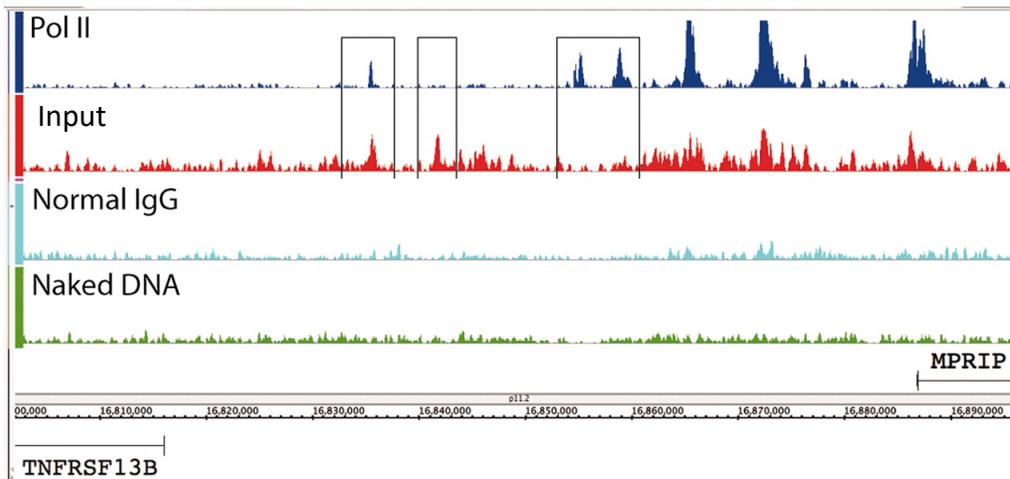
Chen et al 2012. Nature Methods

## Biases vary by cell type and are affected by gene expression



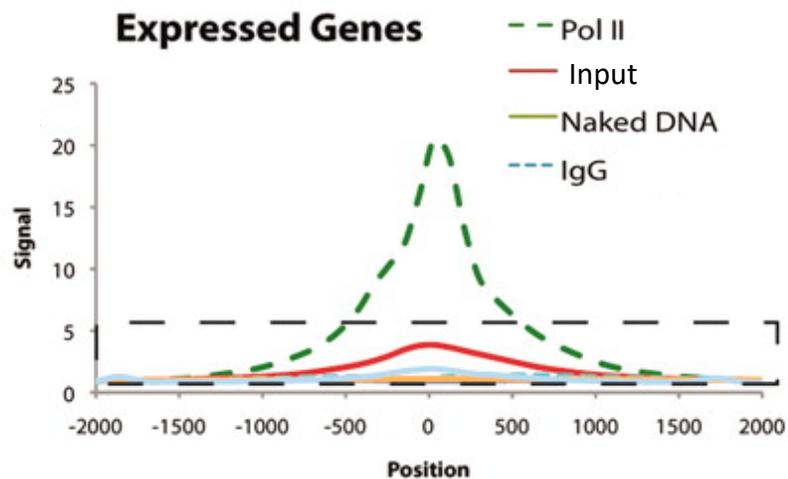
Vega et al 2009. PLOS One

## Different controls behave differently

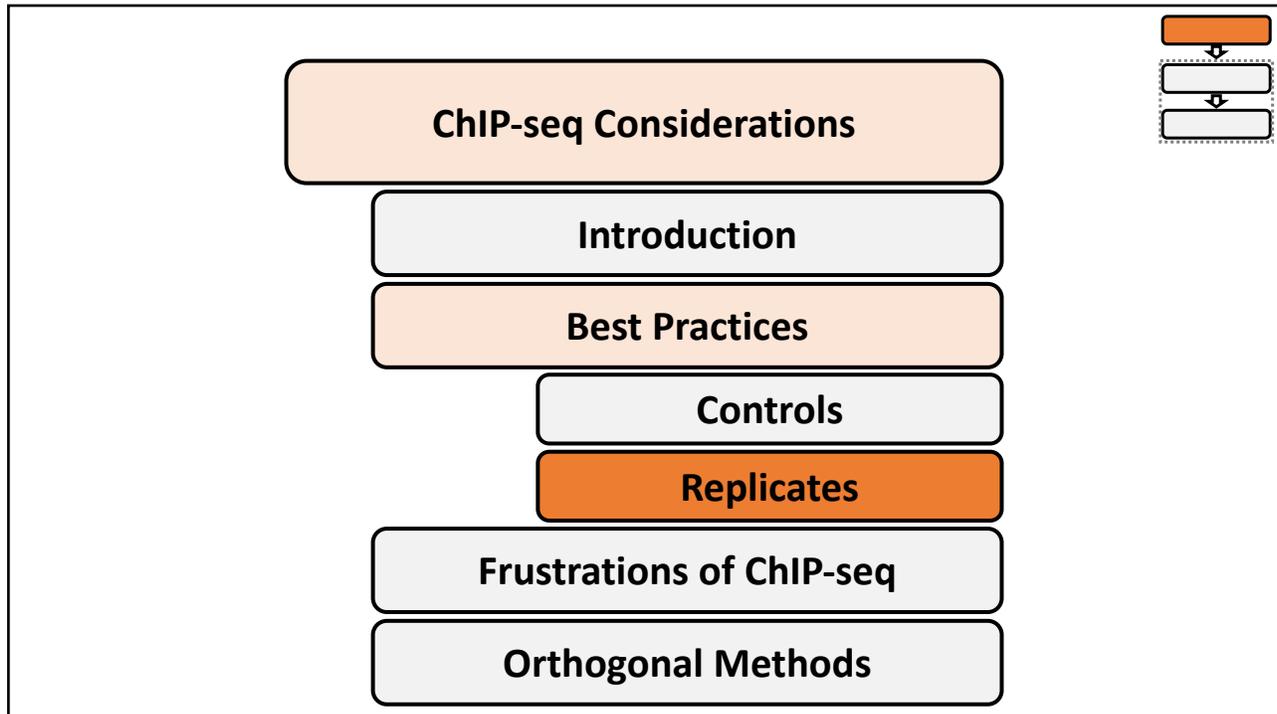


Auerbach et al 2009. PNAS

## Input is enriched at open chromatin while IgG is not



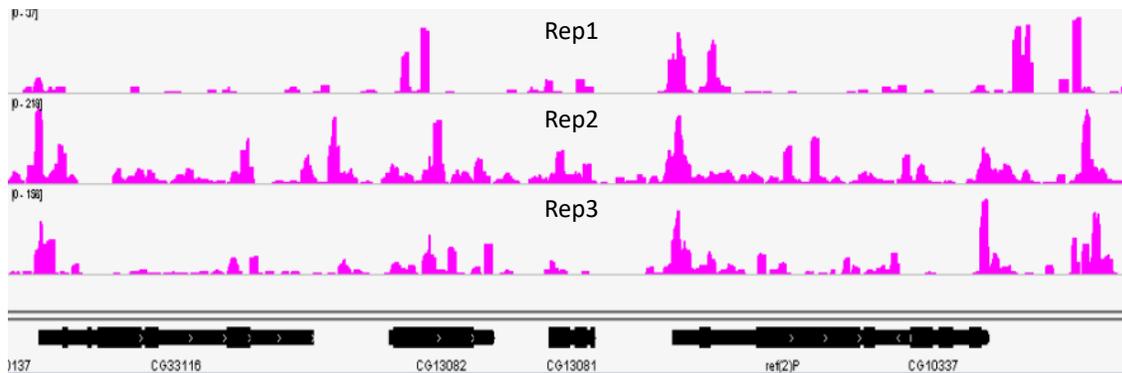
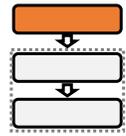
Auerbach et al 2009. PNAS



## Types of replicates

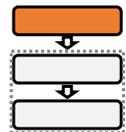
- **Biological/Experimental**
  - To capture variability between different runs
  - eg, repeating ChIP multiple times with the same antibody with cells from the same line grown separately, starting with a different passage of cells, or related samples with the same mutation of interest
- **Technical**
  - Often refers to resequencing the same libraries to deal with sequencing biases
  - Can also include replication of any/all steps following fixation

Biological/Experimental replicates are needed to differentiate between real peaks and background noise



Yang et al 2014. Comput Struct Biotechnol J

## Dealing with replicates



	Number of Samples	Information from individual replicate
Pooling all replicates	No limitation	Lost
Merge after peak calling	Pairwise combinations	Kept
Select one best replicate	No limitation	Lost
Majority rule	No limitation	Kept

Yang et al 2014. Comput Struct Biotechnol J

## ENCODE best practices

- “ChIP-seq grade” antibodies
- Control samples created in parallel with matching ChIP samples
- At least 2 biological replicates
- Minimum useable reads/fragments
  - Useable reads: uniquely mapped after removal of PCR duplicates
  - For Transcription Factor or other narrow features: 10-15M
  - For Histone or other broad features: > 30M
- No preference between single-end or paired-end sequencing

Landt et al 2012. Genome Res

## ChIP-seq Considerations

Introduction

Best Practices

Frustrations of ChIP-seq

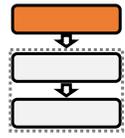
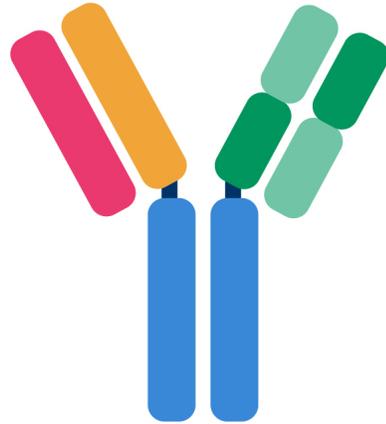
Technical Failure

Inherent Limitations

Orthogonal Methods

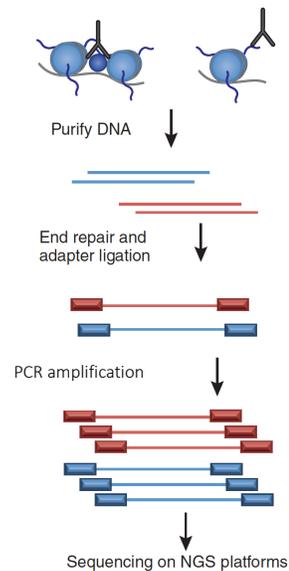
## Technical Failure: Antibodies

- Low binding affinity
- Insufficient antibody
- Non-specific binding
  - Monoclonal vs Polyclonal
- Protein is not a good ChIP candidate.

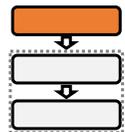


## Technical Failure: Methodological

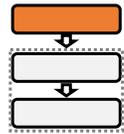
- Not enough Input DNA
- Improper shearing/digestion
- Improper size selection
  - Illumina has a limit on insert length size.
- Too high adaptor/ChIP fragment ratio
- Too many PCR cycles



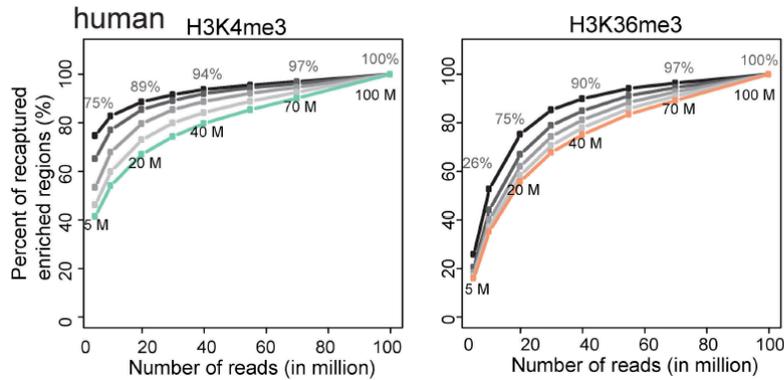
Kidder et. al. Nature Immunology, 2011



# Technical Failure: Low Sequencing Depth

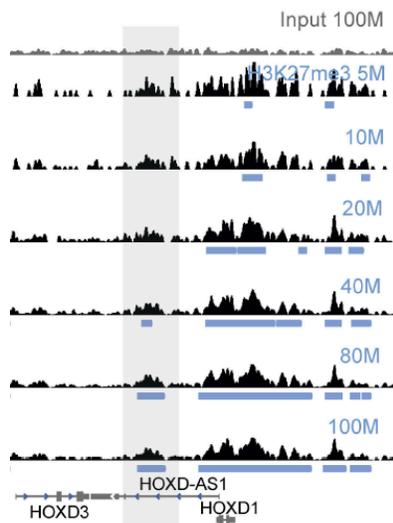
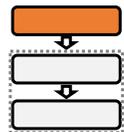


- Did you saturate the amount information coming from your library?

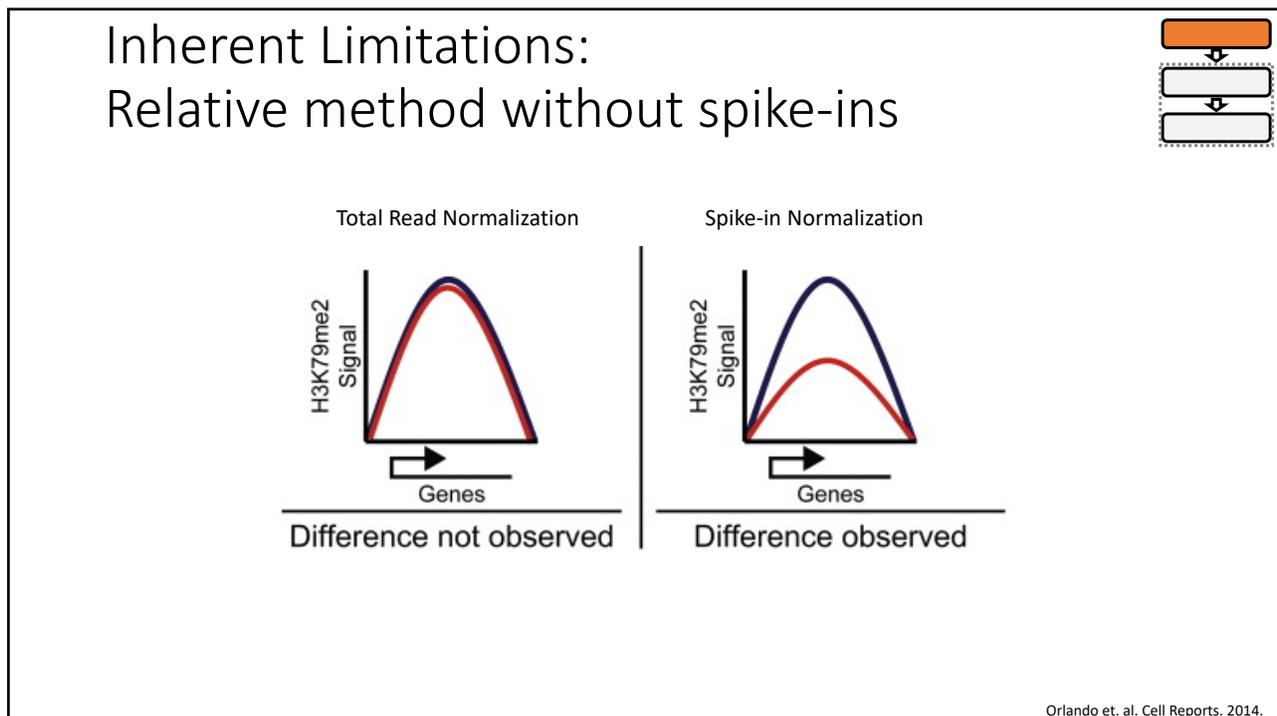
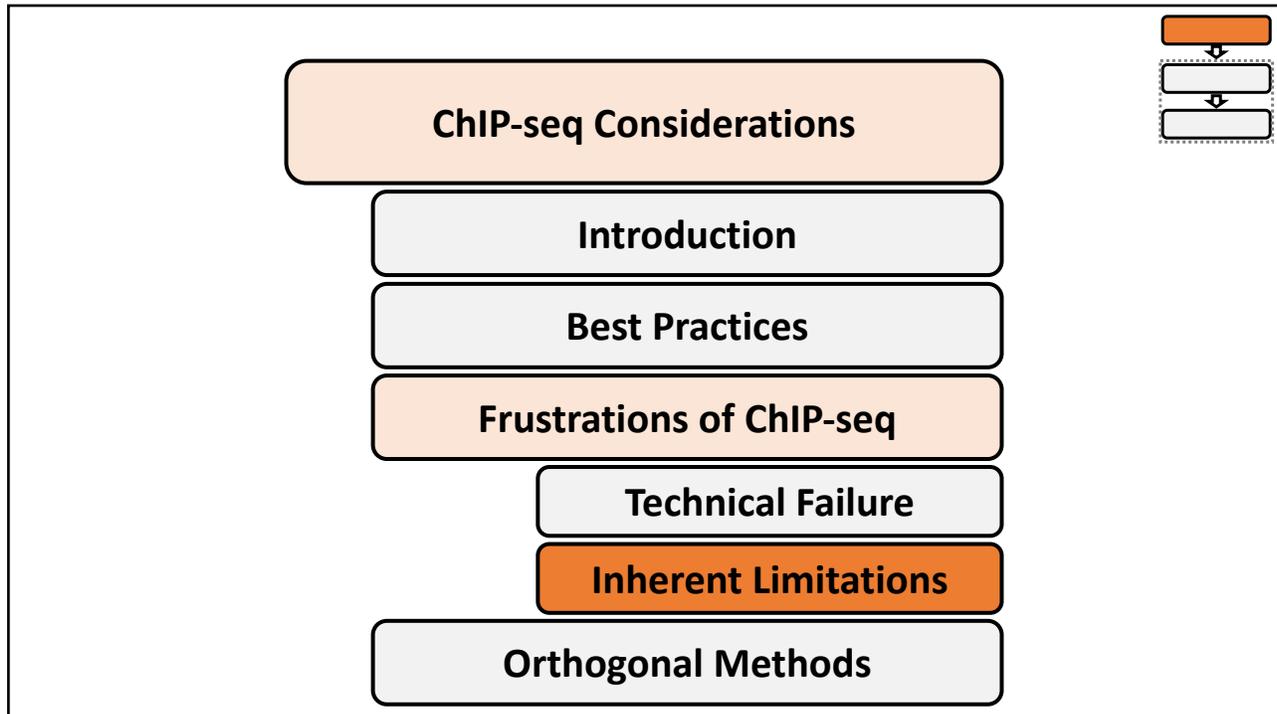


Jung et. al. Nucleic Acids Research. 2014

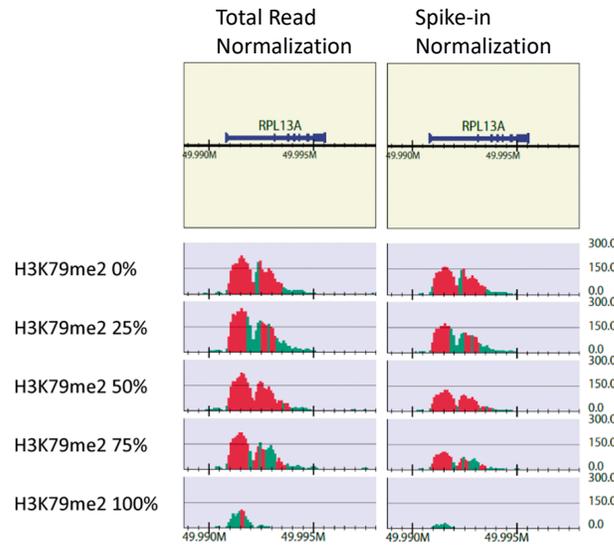
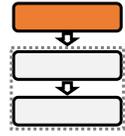
# Technical Failure: Saturation



Jung et. al. Nucleic Acids Research. 2014

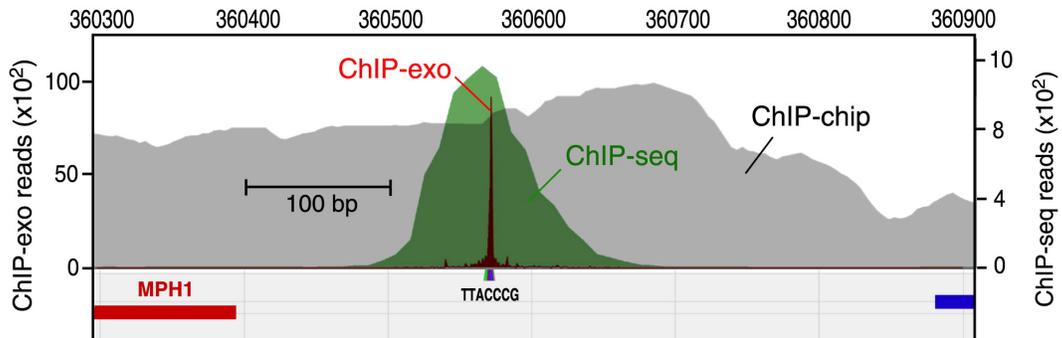
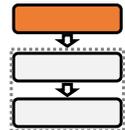


# Inherent Limitations: Relative method without spike-ins



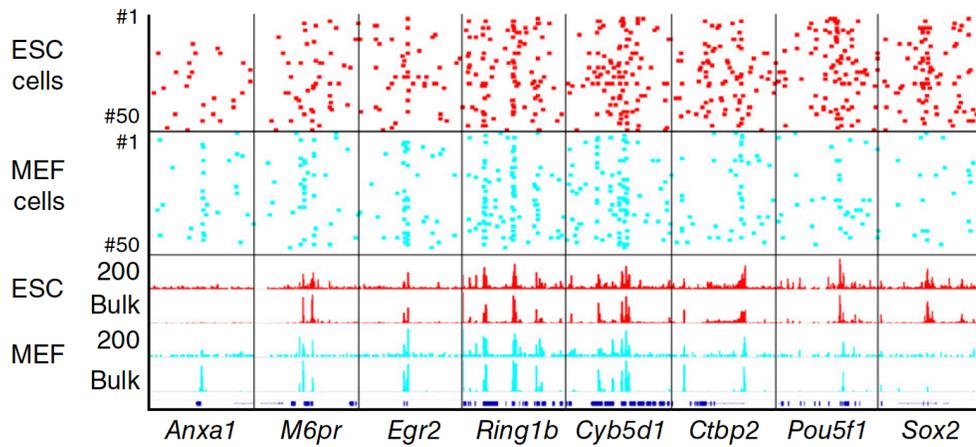
Nakato and Shirahige. Briefings in Bioinformatics. 2017

# Inherent Limitations: Resolution



Rhee and Pugh. Cell. 2011.

## Inherent Limitations: Bulk method not single cell



Rotem et. al. Nature Biotechnology. 2015.

## ChIP-seq Considerations

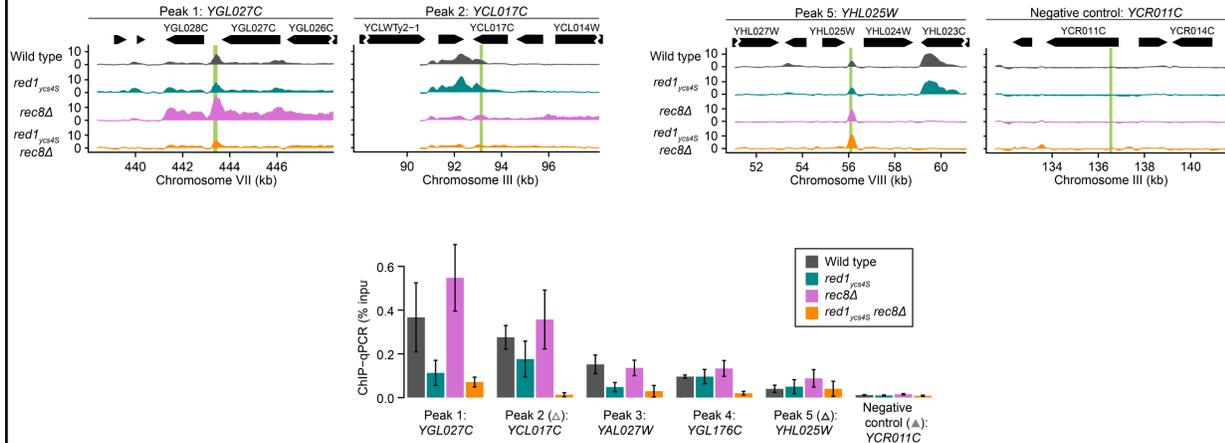
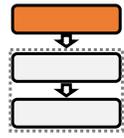
Introduction

Best Practices

Frustrations of ChIP-seq

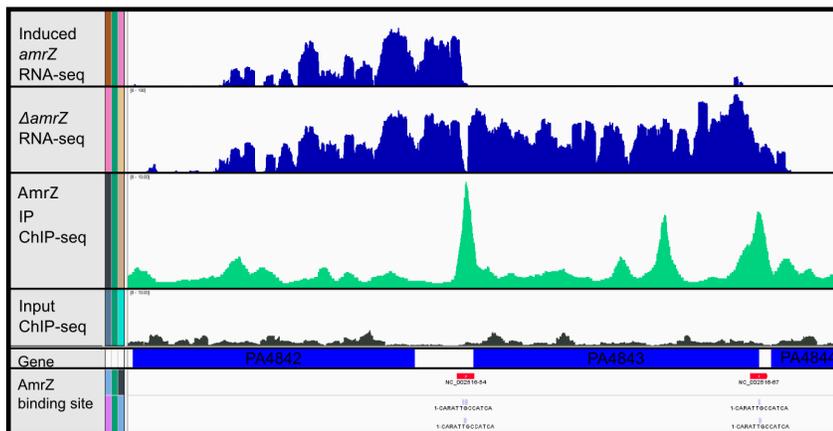
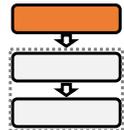
Orthogonal Methods

# Orthogonal Methods: ChIP-PCR/ChIP-QPCR



Markowitz et. al. PLoS Genetics. 2017.

# Orthogonal Methods: Integration with other sequencing methods



ChIP-seq identifies a binding site that changes transcription when deleted

Jones et. al. PLOS Pathogens. 2014

