

# High Resolution Identification of Protein-DNA Binding Events using ChIP-exo

Dongjun Chung<sup>6</sup>, René Welch<sup>1</sup>, Irene Ong<sup>3</sup>, Jeffrey Grass<sup>3,4</sup>, Robert Landick<sup>3,4,5</sup>, and  
Sündüz Keleş<sup>1,2</sup>

<sup>1</sup>Department of Statistics, University of Wisconsin - Madison

<sup>2</sup>Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison

<sup>3</sup>Great Lakes Bioenergy Research Center

<sup>4</sup>Department of Biochemistry, University of Wisconsin - Madison

<sup>5</sup>Department of Bacteriology, University of Wisconsin - Madison

<sup>6</sup>Department of Public Health Sciences, Medical University of South Carolina

## Abstract

Recently, ChIP-exo has been developed to investigate protein-DNA interaction in higher resolution compared to popularly used ChIP-Seq. Although ChIP-exo has drawn much attention and is considered as powerful assay, currently, no systematic studies have yet been conducted to determine optimal strategies for experimental design and analysis of ChIP-exo. In order to address these questions, we evaluated diverse aspects of ChIP-exo and found the following characteristics of ChIP-exo data. First, Background of ChIP-exo data is quite different from that of ChIP-Seq data. However, sequence biases inherently present in ChIP-Seq data still exist in ChIP-exo data. Second, in ChIP-exo data, reads are located around binding sites much more tightly and hence, it has potential for high resolution identification of protein-DNA interaction sites, hence the space to allocate the reads is greatly reduced. Third, although often assumed in the ChIP-exo data analysis methods, the peak pair assumption does not hold well in real ChIP-exo data. Fourth, spatial resolution of ChIP-exo is comparable to that of PET ChIP-Seq and both of them are significantly better than resolution of SET ChIP-Seq. Finally, for given fixed sequencing depth, ChIP-exo provides higher sensitivity, specificity, and spatial resolution than PET ChIP-seq.

In this article, we provide a quality control pipeline which visually assesses ChIP-exo biases and calculates a signal-to-noise measure. Also, we updated dPeak, which makes a striking balance in sensitivity, specificity, and spatial resolution for ChIP-exo data analysis.

# 1 Introduction

ChIP-exo (Chromatin Immunoprecipitation followed by exonuclease digestion and next generation sequencing) Rhee and Pugh ([2]) is the state-of-the-art experiment developed to attain single base-pair resolution of protein binding site identification and it is considered as a powerful alternative to popularly used ChIP-seq (Chromatin Immunoprecipitation coupled with next generation sequencing ) assay. ChIP-exo experiments first capture millions of DNA fragments (150 - 250 bp in length) that the protein under study interacts with using random fragmentation of DNA and a protein-specific antibody. Then, exonuclease is introduced to trim 5' end of each DNA fragment to a fixed distance from the bound protein. As a result, boundaries around the protein of interest constructed with 5' ends of fragments are located much closer to bound protein compared to ChIP-seq. This is the step unique to ChIP-exo that could potentially provide significantly higher spatial resolution compared to ChIP-seq. Finally, high throughput sequencing of a small region (25 to 100 bp) at 5' end of each fragment generates millions of reads or tags.

While the number of produced ChIP-exo data keeps increasing, characteristics of ChIP-exo data and optimal strategies for experimental design and analysis of ChIP-exo data are not fully investigated yet, including issues of sequence biases inherent to ChIP-exo data, choice of optimal statistical methods, and determination of optimal sequencing depth. However, currently, the number of available ChIP-exo data is still limited and their sequencing depths are still insufficient for such investigation.

## 2 Results

## 3 Discussion

## Methods

## References

- [1] Eric Mendenhall and Bradley Bernstein. Dna-protein interactions in high definition. *Genome Biology*, 2012.
- [2] Ho Sung Rhee and Franklin Pugh. Comprehensive genome-wide protein-dna interactions detected at single-nucleotide resolution. *Cell*, 2011.