THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

---

# HONGKAI JI

Department of Biostatistics
John Hopkins Bloomberg School of Public Health

# Single Cell Genomic Data Science Methods for Mapping Gene Regulatory Landscape

MONDAY, October 26, 2020 at 4:00 PM
Via Zoom (session information will be e-mailed to subscribers)
Virtual Reception After Colloquium

## ABSTRACT

Understanding how gene expression is controlled temporally and spatially is crucial for studying human development and diseases. Recent technological advances have made it possible to map genome-wide regulatory element activities in single cells. However, single-cell regulome data are sparse and noisy. It is challenging to use these data to accurately measure the activity of each individual regulatory element in each cell. Moreover, single-cell regulome data currently are only available for a limited number of sample types and conditions. In this talk, I will discuss how these challenges may be tackled from a data scientist's perspective. First, I will show that by leveraging massive functional genomic datasets in public databases, one can use single-cell RNA-seq (scRNA-seq) data to predict regulome. The prediction not only offers competitive accuracy but may also outperform single-cell ATAC-seq (scATAC-seq), the state-of-the-art single cell regulome mapping technology, when studying a single cell or a rare cell subpopulation. I will introduce BIRD, Big Data Regression for predicting DNase I hypersensitivity, to deal with this ultra-high dimensional problem. Second, I will show that publicly available regulome data can also be used to improve scATAC-seq data analysis. I will introduce SCATE, a statistical framework that adaptively integrates information from co-activated regulatory elements, similar cells, and publicly available regulome data to substantially increase the accuracy for estimating activities of individual regulatory elements using scATAC-seq. BIRD and SCATE provide two examples demonstrating new challenges and opportunities single-cell genomics brings and unique contributions data science can make in this rapidly evolving field.

---

For further information and inquiries about building access for persons with disabilities, please contact Jonathan Rodriguez at 773.702.8333 or send him an email at jgrodriguez@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following website: https://lists.uchicago.edu/web/subscribe/statseminars.