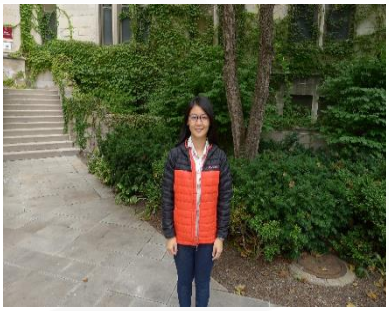
psd

DISSERTATION PROPOSAL PRESENTATION

Statistical challenges in high-throughput transcriptomics data

WHEN
March 7, 2022
1:00-2:00pm

WHERE
Zoom Meeting



Chih-Hsuan Wu, PhD candidate

This presentation consists of two parts.

In many existing pipelines for scRNA-seq, pre-processing steps are applied to account for excessive zeros before further data analysis. A novel framework HIPPO, on the other hand, leverages zero proportions to explain cellular heterogeneity and integrates feature selection with iterative clustering. We further show that the k-inflated genes can also be a useful indicator in feature selection. Besides Poisson assumption, we are also working on Negative Binomial distribution data. Our goal is to improve HIPPO and introduce a new hierarchical clustering procedure with more general settings.

The RNA modification plays a big role in regulating gene expressions. Especially, m6A is by far the most abundant mRNA modification. A clustering analysis suggests that m6A modification is highly cell type-specific. To better characterize the relationship between m6A and gene expressions, we propose a cell type-specific differential methylation analysis with mixture Poisson GLMM model based on the framework of RADAR. We incorporate the Poisson model with log-Gamma random effects and hence the estimates can be solved via an iteratively optimizing method. In the future, we will apply this model on T2D dataset in which the relative proportions of each cell type are known.

stat.uchicago.edu

THE UNIVERSITY OF CHICAGO | DEPARTMENT OF STATISTICS