The Hong Kong University of Science and Technology

Department of Mathematics

<u>Seminar on Statistics</u>

Removing Unwanted Variation Using both Control and Target Genes in Single Cell RNA Sequencing Studies
By

Dr. Mengjie CHEN
University of Chicago

Abstract

Single cell RNA sequencing (scRNAseq) technique is becoming increasingly popular for unbiased and high-resolutional transcriptome analysis of heterogeneous cell populations. Despite its many advantages, scRNAseq, like any other genomic sequencing technique, is susceptible to the influence of confounding effects. Controlling for confounding effects in scRNAseq data is thus a crucial step for proper data normalization and accurate downstream analysis. Several recent methodological studies have demonstrated the use of control genes (including spike-ins) for controlling for confounding effects in scRNAseq studies. However, these methods can be suboptimal as they ignore the rich information contained in the target genes. Here, we develop an alternative statistical method, which we refer to as scPLS, for more accurate inference of confounding effects. Our method models control and target genes jointly to better infer and control for confounding effects. To accompany our method, we develop a novel expectation maximization algorithm for scalable inference. Our algorithm is an order of magnitude faster than standard ones, making scPLS applicable to hundreds of cells and hundreds of thousands of genes. With simulations and studies, we show the effectiveness of scPLS in removing technical confounding effects as well as for removing cell cycle effects. Under the same framework, we will further discuss how to identify subpopulations using a Bayesian nonparametric approach.

Date: *Thursday,30 November, 2017*

Time: *11:00a.m.-12:00noon*

Venue: *Room 5508, Academic Building, (near Lifts 25&26), HKUST*

All are welcome!