Computer Science Defense

Statistical Modeling and Learning in Single-cell RNA Sequencing Data

Kenong Su Emory University

Abstract: Single-cell RNA sequencing (scRNA-seq) technologies have revolutionized biological research, and cell clustering becomes an important and commonly performed task in scRNA-seq data analysis. An essential step in scRNA-seq clustering is to select a subset of genes (referred to as features), whose expression patterns will further be adopted for downstream clustering analysis. It is noting that both the quality and quantity of the feature set will have significant impact on the clustering accuracy. However, almost all existing scRNA-seq clustering tools select features relying on some simple unsupervised methods, mostly based on statistical moments. Plus, these existing tools tend to choose random top (e.g., 1000 or 2000) features for cell clustering. In this talk, I will present a novel unsupervised algorithm named FEAST (Su et al., 2021) specifically designed for selecting most representative genes in scRNA-seq experiment is how to decide a proper number of cells in order to reach a desired power level, in the context of differential expression tests and marker gene detections. I will also present POWSC pipeline (Su et al., 2020), a simulation-based approach, to provide comprehensive power evaluation and sample size recommendation. The findings from applying POWSC can potentially guide scRNA-seq experimental designs.

Tuesday, April 13, 2021, 2:00 pm https://emory.zoom.us/j/94588103525

> Computer Science Emory University