Statistical learning for analyzing single-cell multi-omics data

Wei Chen

University of Pittsburgh
E-mail: wec47@pitt.edu

Abstract: The rapid advances in single cell technologies have been playing important roles in understanding the heterogeneity and dynamics of various cell populations in complex multicellular tissue or organs. The recently developed droplet-based single cell transcriptome sequencing (scRNA-seq) technology enables researchers to measure the gene expression of tens of thousands single cells simultaneously. More recently, coupling with droplet-based scRNA-seq, another revolutionary technology named Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE-seq) allows the detection of cell surface proteins and transcriptome profiling within the same cell simultaneously. Despite the rapid advances in technologies, novel statistical methods and computation tools for analyzing single-cell multi-omics data are still lacking. In this study, we developed a novel random effects model that jointly analyze the paired data from scRNA-seq and CITE-seq experiments under a Bayesian framework. In the simulation study and analysis of in-house real data sets, we demonstrated the validity and advantages of our method in understanding immune cells as well as facilitating novel biological discoveries.