Yale SCHOOL OF PUBLIC HEALTH
Biostatistics

SPRUCE and MAPLE: Bayesian Spatial Multivariate Mixture Model for High Throughput Spatial Transcriptomics Data

Dongjun Chung, PhD
Associate Professor
Department of Biomedical Informatics
The Ohio State University

12:00 Noon Eastern time, **Tuesday, November 30, 2021**

ABSTRACT

High-throughput spatial transcriptomics (HST) is a rapidly emerging experimental technology that allows for profiling gene expression in tissue samples at or near single-cell level while retaining the spatial location of each sequencing unit. Through analyzing HST data, we seek to identify sub-populations within a tissue sample that reflect biological cell types or states. Existing methods either ignore the spatial heterogeneity in gene expression profiles, fail to account for important statistical features such as skewness and heavy-tails, or are heuristic-based methods that lack the inferential benefits of statistical modeling. To address this gap, we developed SPRUCE (SPatial Random effects-based clUstering of single CEll data), a Bayesian spatial multivariate finite mixture model based on multivariate skew normal distributions, which is capable of identifying distinct cellular sub-populations in HST data. We further implement a novel combination of Polya-Gamma data augmentation and spatial random effects to infer spatially correlated mixture component membership probabilities. Our comprehensive simulation studies indicate that SPRUCE outperforms existing methods with respect to identifying tissue architectures. The real data application of SPRUCE indicates that SPRUCE can potentially provide novel insight that could be further confirmed with follow-up validation experiments. Finally, I will also discuss our recent work, namely MAPLE (Multi-sAmple sPatiaL transcriptomics model), a novel extension of SPRUCE for the multi-sample HST data analysis.