Wednesday, September 6, 2017
3:30 PM – 4:30 PM
Forum Hall, Palmer Commons Building

Robust and powerful
single cell RNA-seq studies by
leveraging natural genetic variation

Hyun Min Kang, Ph.D.
John G. Searle Assistant Professor of Biostatistics (until August 2017)
Associate Professor of Biostatistics (starting from September 2017)

Droplet-based single-cell RNA-sequencing (dscRNA-seq) has enabled rapid, massively parallel profiling of transcriptomes from tens of thousands of cells. For population-scale experiments to assess genetic and/or environmental effects on transcriptional regulation, sample size (or per-sample cost) and batch effect are limiting factors to identify associations at the resolution of individual cell or cell-type. Here we introduce a new experimental design and a statistical model to enable cost-effective multiplexing of dscRNA-seq experiment while avoid technical batch effects. Our statistical model implemented in the *demuxlet* software tool enables dscRNA-seq of cells from many individuals at a very high loading concentration in a single library preparation. Our algorithm harnesses natural genetic variation to deconvolute the sample identity of each droplet and while identifying and filtering out droplets containing two cells using a mixture model. These capabilities enable multiplexed dscRNA-seq experiments where cells from multiple individuals are pooled and captured at higher throughput than standard workflows. To demonstrate the performance of *demuxlet*, we sequenced 3 pools of peripheral blood mononuclear cells (PBMCs) from 8 lupus patients. Given array-based genotyping data for each individual, demuxlet correctly recovered the sample identity of >99% of singlets, and identified doublets at rates consistent with previous estimates. In PBMCs, we demonstrate the utility of multiplexed dscRNA-seq in two applications: characterizing cell type specificity and inter-individual variability of cytokine response from 8 lupus patients and mapping genetic variants associated with cell type specific gene expression from 23 donors. Demuxlet is fast, accurate, scalable and could be extended to other single cell datasets that incorporate natural or synthetic DNA barcodes.

MICHIGAN MEDICINE
UNIVERSITY OF MICHIGAN

Hosted by Dr. Alan Boyle