BIOSTATISTICS SEMINAR

Transfer learning and Data Alignment in Single Cell Transcriptomics

Cells are the basic biological units of multicellular organisms. The development of single-cell RNA sequencing (scRNA-seq) technologies have enabled us to study the diversity of cell types in tissue and to elucidate the roles of individual cell types in development and disease. Yet, scRNA-seq data are noisy and sparse, with only a small proportion of the transcripts that are present in each cell represented in the final data matrix. We propose a transfer learning framework to borrow information across related single cell data sets for de-noising and expression recovery. Our goal is to leverage the expanding resources of publicly available scRNA-seq data, for example, the Human Cell Atlas which aims to be a comprehensive map of cell types in the human body. Our method is based on a deep neural network, trained to extract transferable gene expression features across studies coming from different labs, generated by different technologies, and/or obtained from different species. Through this framework, we explore the limits of data sharing: How much can be learned across cell types, tissues, and species? How useful are data from other technologies and labs in improving the estimates from your own study? We also explore the implications of technical batch artifacts in the joint analysis of multiple data sets, and propose strategies for robust estimation in the presence of such technical artifacts.

Nancy R. Zhang, PhD

Professor of Statistics
The Wharton School
University of Pennsylvania

Thursday September 13, 2018   **3:30 pm - 4:30 pm**
Blue Cross and Blue Shield of North Carolina Foundation Auditorium

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH