

BIOSTATISTICS SEMINAR



Lynn Lin, PhD
Assistant Professor,
Dept. of Statistics,
Penn State

Validation and Annotation for Single-Cell Data Clustering

Clustering single-cell data is one of the central analysis components to identify and characterize different cell types, especially towards the novel ones. The cell types identified often help to substantiate existing hypotheses or motivate new ones. However, it is generally known that clustering results obtained computationally can vary drastically depending on different samples, algorithms used, or even initializations. Yet the combinatorial nature of the clustering result, which is a partition rather than a set of parameters or a function, blurs notions of mean, and variance. This intrinsic difficulty hinders the development of methods to improve clustering by aggregation or to assess the uncertainty of individual clusters generated. In this talk, I will present a new approach to overcome that barrier by aligning clusters via optimal transport. Equipped with this technique, set relationships between clusters such as one-to-one match, split, and merge can be revealed. New algorithm can be developed to enhance clustering by any baseline method using bootstrap samples. I will also briefly talk about how the same technique be used to provide biological annotation to the clusters identified to achieve human interpretable clustering results.

Thursday March 5, 2020

3:30 pm - 4:30 pm

Blue Cross and Blue Shield of North Carolina Foundation Auditorium



UNC
GILLINGS SCHOOL OF
GLOBAL PUBLIC HEALTH