

Computational methods to decipher human disease using GWAS and functional genomics



Martin Jinye Zhang, PhD
Research Associate
Harvard School of Public Health

Integrating genome-wide association studies (GWAS) with functional genomics can aid biological interpretation of disease-associated variants. First, I will introduce single-cell disease relevance score (scDRS), a method that identifies disease-critical cell populations by integrating GWAS with single-cell RNA-sequencing (scRNA-seq). Previous works can identify disease associations for predefined cell types, while scDRS can detect heterogeneous disease associations within classical cell types without predefined cell type annotations. Applying scDRS to 74 diseases/traits and 1.3 million single-cell gene-expression profiles, we identified subpopulations of disease-associated cells not captured by existing cell-type labels, such as T cell subpopulations associated with inflammatory bowel disease, partially characterized by their effector-like states. Second, I will discuss optimizing parameters in scRNA-seq experiments to maximally extract biological information. An underlying question is how to allocate the limited sequencing budget between the number of cells and sequencing depth. Instead of focusing on results of a specific analysis, we information-theoretically evaluated the fundamental limit in learning the underlying biological ground truth. We determined that, for estimating many important gene properties, the optimal allocation was to sequence at a depth of around one read per cell per gene, slightly deeper than most existing data sets. Interestingly, the corresponding optimal estimator was different from the commonly-used plug-in estimator, which produced severely biased estimates.

Thursday January 19, 2023, 3:30-4:30 PM Eastern

133 Rosenau Hall

Virtual using link and info below.

<https://unc.zoom.us/j/91249030964?pwd=UXloTWlHajdObkRqd1d5TnRaMitYdz09>

Meeting ID: 912 4903 0964

Passcode: 884852