

Statistical methods for integrative analysis of large-scale whole-genome sequencing studies



Xihao Li, PhD
Post-Doctoral Researcher
Harvard T.H. Chan School of Public Health
Department of Biostatistics

Large-scale whole genome sequencing (WGS) studies have enabled the analysis of rare variants (RVs) associated with complex human traits. There are several challenges in analyzing WGS data, including computation scalability, limited scope to integrate variant biological functions, and lack of ability to leverage summary statistics across multiple studies. In this talk, Dr. Li will present two recent methods to address these challenges. First, we propose STAAR (variant-Set Test for Association using Annotation infoRmation), a scalable and powerful RV association test method that effectively integrates both variant functional categories and multiple complementary annotations using a dynamic weighting scheme. For the latter, we introduce ‘annotation principal components’, multidimensional summaries of in silico variant annotations. STAAR accounts for population structure and relatedness and is scalable for analyzing biobank-scale whole-genome sequencing studies of continuous and dichotomous traits. Second, we propose MetaSTAAR, a resource-efficient RV meta-analysis framework, for large-scale WGS association studies. By storing the linkage disequilibrium (LD) information of rare variant score statistics in a new sparse matrix format, MetaSTAAR is storage efficient and computationally scalable for analyzing large-scale WGS data, producing results comparable to using pooled individual-level data. We applied STAAR and MetaSTAAR to identify RVs associated with four lipid traits in a total of 30,138 ancestrally diverse and related samples from 14 studies of the Trans-Omics for Precision Medicine (TOPMed) Program. We discovered and replicated several conditionally significant RV associations with lipid traits, including disruptive missense RVs of *NPC1L1* associated with low-density lipoprotein cholesterol.

Thursday, February 2, 2023, 3:30-4:30 PM Eastern

133 Rosenau Hall

Virtual using link and info below.

<https://unc.zoom.us/j/91249030964?pwd=UXloTWIHajdQbkRqd1d5TnRaMitYdz09>

Meeting ID: 912 4903 0964

Passcode: 884852