Ancestry Estimation and Control of Population Structure in Genetic Association Studies
Chaolong Wang, Genome Institute of Singapore

Estimating individual ancestry is important in genetic association studies where population structure leads to false positive signals. With the increasing of sample size in disease association studies based on next-generation sequencing, control of population structure and cryptic relatedness within the study sample becomes more challenging. In this talk, I will first introduce our recent works on a unified statistical framework, called LASER, for estimating an individual's ancestry based on either low-coverage sequencing data or array genotyping data. LASER enables accurate estimation of both continental ancestry and fine-scale ancestry using small amounts of genetic data, and can facilitate integrative analysis of genetic data from different sources. A successful application of our LASER method has led to discovery of a novel rare variant associated with age-related macular degeneration (AMD). In the second part of this talk, I will briefly introduce our on-going work for control of both population stratification and family relatedness using generalized linear mixed models (GLMM). We demonstrate that linear mixed models (LMM), when applying to binary traits, can result in incorrect p values for testing association between disease and genetic variants in the presence of population stratification. In contrast, GLMM controls both population structure and family relatedness very well. We develop a computationally efficient software package for applying GLMM in large-scale genetic datasets.