



THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

Summer Research Program 2022

PG STUDENT SEMINAR

A fast and scalable statistical framework for genetic prediction with large-scale dataset

By

Dr. Mingxuan CAI

Abstract

The rich resources of massive genetic data offer an unprecedented chance for individualized disease risk prediction. Through statistical modelling, the risk scores derived from genetic variants can effectively identify the individuals with higher disease risk from general population. However, multiple challenges arise when constructing risk prediction from massive data. First, the massive genetic data usually is comprised of hundreds of thousands of samples with millions of variants. Computational cost for standard statistical analysis becomes unfordable. Second, the individual-level genetic data are usually of restricted access due to privacy protection. Third, due to the large difference of genetic architectures between populations and the limited sample size from non-European populations, risk prediction has been less accurate for the non-European individuals. To improve the prediction accuracy in non-European populations, we propose a cross-population analysis framework for genetic risk prediction with both individual-level (XPA) and summary-level (XPASS) genetic data. By leveraging trans-ancestry genetic correlation, our methods can borrow information from the Biobank-scale European population data to improve risk prediction in the non-European populations. In a Chinese cohort, our methods achieved 7.3%-198.0% accuracy gain for height and 19.5%-313.3% accuracy gain for body mass index (BMI) in terms of predictive R^2 compared to existing prediction approaches.

Date : 3 August 2022 (Wednesday)

Time : 3:00pm

Zoom Meeting : <https://hkust.zoom.us/j/8097069142> (Passcode: hkust)

All are Welcome!