The Hong Kong University of Science and Technology

Department of Mathematics

PhD THESIS EXAMINATION

Scalable Statistical Methods for Cross-population Risk Prediction and Integrative Analysis of Multi-omics Data

By

Mr. Mingxuan CAI

ABSTRACT

The large-scale genome-wide association studies (GWASs) have detected tens of thousands of risk variants underlying complex phenotypes. However, there are still outstanding challenges that hamper the clinical translation and biological interpretation of GWAS discoveries.

In clinical applications, the development of polygenic-risk-scores (PRSs) has proved useful to stratify the general population into different risk groups for the European population. However, PRS is less accurate in non-European populations due to genetic differences across different populations. To improve the prediction accuracy in non-European populations, we propose a cross-population analysis framework (XPA) for PRS construction with large-scale GWAS 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. With innovations in data structure and algorithm design, our methods provide a substantial saving in computational time and memory usage. Our method achieved 7.3%-198.0% accuracy gain for the prediction of height in a Chinese cohort in terms of predictive R2 compared to existing PRS approaches.

For a better biological interpretation of GWAS discoveries, integrative analysis of multi-omics data have been conducted to implicate biological insights. By leveraging existing GWAS and transcriptomic information, transcriptome-wide association studies (TWAS) have achieved many successes in identifying trait-associations of genetically-regulated expression (GREX) levels. Considering the increasing availability of transcriptomic data from different conditions and the often unknown trait-relevant cell/tissue-types, we propose a method and tool, IGREX, for precisely quantifying the proportion of phenotypic variation attributed to the GREX component. Using transcriptomic data of 48 tissue types from the GTEx project as a reference panel, we evaluated the tissue-specific GREX impact on a wide spectrum of phenotypes and provided biological insights.

Date : **17 May 2022, Tuesday**

Time : **10:00 a.m.**

Venue : **Online via ZOOM**
ID: 965 7691 8873 (Passcode: hkust)
https://hkust.zoom.us/j/96576918873

***Thesis Examination Committee*:**

Chairman	**:**	**Prof. Yusong GUO, LIFS/HKUST**
Thesis Supervisors	**:**	**Prof. Can YANG, MATH/HKUST**
Member	**:**	**Prof. Dong XIA, MATH/HKUST**
Member	**:**	**Prof. Zhigang BAO, MATH/HKUST**
Member	**:**	**Prof. Jiguang WANG, LIFS /HKUST**
External Examiner	**:**	**Prof. Ming-Yen CHENG, Department of Mathematics/ Hong Kong Baptist University**

(Open to all faculty and students)

The student's thesis is now being displayed on the reception counter in the General Administration Office (Room 3461).