ON-SITE BIOLOGY COLLOQUIUM

Friday, 27 Oct 2023 | 4 pm | LT32

Hosted by Prof Greg Tucker-Kellogg
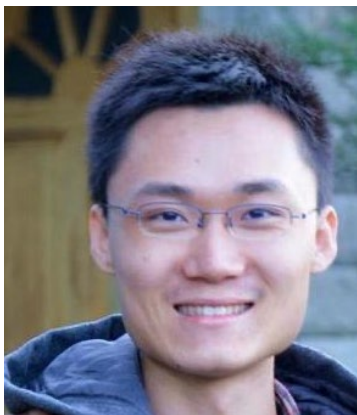
Map to Block S1A

Identifying risk genes for complex diseases with genomics and data science

By **Liu Boxiang**

Departments of Pharmacy & Biomedical Informatics, NUS

About the Speaker

Boxiang Liu obtained BA degree in Biophysics from Illinois Wesleyan University, a MS degree in Statistics and a PhD degree in Bioinformatics from Stanford University. He was a research leader at Baidu Research USA and joined the National University of Singapore as an Assistant Professor in 2022. His research group specializes in genetic regulation of molecular traits (QTLs) and single-cell multi-omics. He is the winner of President's Award in Natural Sciences and Mathematics, Stanford University CEHG fellowship, Charles B. Carrington Memorial Award, and the Chinese Government Outstanding Overseas Ph.D. Students. His research group focuses on using computational and statistical tools to understand the genetics of complex human diseases, with the long-term goal of validating known and identifying novel drug targets.

High-throughput genotyping and sequencing have led to the discovery of thousands of disease-associated variants. Because most of these variants lie in non-coding regions, their functional mechanisms remain unclear. To uncover target genes and pathways regulated by disease-associated variants, it is critical to perform functional genomic assays and develop novel computational approaches to analyze assayed information. We demonstrate the power of functional genomic assays with two diseases. Coronary artery disease (CAD) is the leading cause of death globally. It is estimated that 40 - 60% of CAD severity can be attributed to genetic factors. GWAS meta-analyses have uncovered more than 100 significant loci, but most are difficult to interpret because they reside in non-coding regions. We generated transcriptomic and whole-genome sequencing datasets in human coronary artery smooth muscle cells (HCASMC) from 52 unrelated donors and an ATAC-seq dataset on 8 donors and found that HCASMC-specific genetic regulatory mechanisms are highly enriched in CAD GWAS signals. By jointly analyzing eQTL and GWAS datasets, we identified five risk genes. TCF21 and SMAD3 has subsequently been validated by single-cell analysis in atherosclerotic mouse models. Age-related macular degeneration is one of the leading causes of blindness in elderlies. It has been estimated that genetic factors explain 45% - 70% of variation in severity of age-related macular degeneration. Retinal pigment epithelium (RPE) serves vital roles in ocular development but is underrepresented in genetic regulation studies. We performed RNA-seq in RPE from 24 donors under two metabolic conditions and discovered hundreds of shared and condition-specific eQTLs. By jointly analyzing RPE eQTL and AMD GWAS, we identified several risk genes including RDH5. In particular, we found that the eQTL regulatory SNP also regulates splicing. Experimental validation confirms that the minor allele leads to aberrant splicing and subsequently RNA non-sense-mediated decay. This result revealed the genetic mechanism of RDH5 regulation and confirmed RDH5 as a risk gene for age-related macular degeneration, making it a potential target for drug development. To facilitate functional interpretation of GWAS, we perform colocalization analysis across hundreds of traits and make all results publicly available.