Indiana University-Purdue University Indianapolis
Department of Mathematical Sciences

Statistics Seminar

12:15pm—1:15pm, Tuesday, January 24, 2023
Zoom Meeting: Meeting ID: 845 0989 4694

Speaker: **Zilin Li**
Department of Biostatistics and Health Data Science,
Indiana University School of Medicine

Title: **STAARpipeline: an all-in-one rare-variant tool for biobank-scale whole-genome sequencing data**

Abstract:

Large-scale whole-genome sequencing (WGS) studies have enabled the analysis of rare variant associations with complex human diseases and traits. Variant set analysis is a powerful approach to studying rare variant associations. However, existing methods have limited ability to define the variant set in the genome, especially for the noncoding genome. We propose a computationally efficient and robust rare variant association-detection framework, STAARpipeline, to automatically annotate a WGS study and perform flexible rare variant association analysis, including gene-centric analysis and fixed-window and dynamic-window-based non-gene-centric analysis by incorporating variant functional annotations. In gene-centric analysis, STAARpipeline groups coding and noncoding variants based on functional categories of genes and incorporate multiple functional annotations. In non-gene-centric analysis, in addition to fixed-size sliding window analysis, STAARpipeline provides a data-adaptive-size dynamic window analysis. All these variant sets could be automatically defined and selected in STAARpipeline. STAARpipeline also provides analytical follow-up of dissecting association signals independent of known variants via conditional analysis. We applied the STAARpipeline to analyze the total cholesterol in 30,138 samples from the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. All analyses scale well in computation time and memory. We discover several potentially new significant associations with lipids, including a finding of rare variants in an intergenic region near JKAMPP1 associated with total cholesterol. In summary, the STAARpipeline is a powerful and resource-efficient tool for association analysis of biobank-scale WGS studies.

Bio:
Dr. Zilin Li is an Assistant Professor in the Department of Biostatistics and Health Data Science at Indiana University School of Medicine. Before being an assistant professor, he was a research scientist, research associate, and postdoctoral research fellow in Professor Xihong Lin's lab in the Department of Biostatistics at Harvard T.H. Chan School of Public Health. He received his Ph.D. from Tsinghua University in 2016, supervised by Professor Xihong Lin. His research interests lie in statistical genetics and high-dimensional statistics with applications for analyzing massive health data, especially developing statistical methods for scalable analysis of large-scale genetics and genomics data.