

Big Data Training for Cancer Research

Special Lecture Series

Microbiome Data Science - Phylogenetic Tree, Bacterial Growth and Biosynthetic Gene Clusters

Dr. Hongzhe Li

June 14, 2021, 1:00 – 2:30 PM (EDT)



Speaker Bio: Dr. Hongzhe Li is Perelman Professor of Biostatistics, Epidemiology and Informatics at the Perelman School of Medicine at the University of Pennsylvania. He is Vice Chair of Research Integration, Director of Center of Statistics in Big Data and former Chair of the Graduate Program in Biostatistic at Penn. Dr. Li has been elected as a Fellow of the American Statistical Association (ASA), a Fellow of the Institute of Mathematical Statistics (IMS) and a Fellow of AAAS. Dr. Li served on the Board of Scientific Counselors of the National Cancer Institute of NIH and regularly serves on various NIH study sections. He served as Chair of the Section on Statistics in Genomics and Genetics of the ASA. He is currently an Associate Editor of Biometrics, Statistica Sinica and also co-Editor-in-Chief of Statistics in Biosciences. Dr. Li's research focuses on developing statistical and computational methods for analysis of large-scale genetic, genomics and metagenomics data and theory on high dimensional statistics. He has published papers in Science, Nature, Nature Genetics, Nature Methods, Nature Microbiology, Science Translational Medicine, Cell Host & Microbe, JASA, JRSS, Biometrika, Annals of Statistics and Annals of Applied Statistics etc.

Abstract: The gut microbiome plays an important role in maintenance of human health. High-throughput shotgun metagenomic sequencing of a large set of samples provides an important tool to interrogate the gut microbiome. Besides providing footprints of taxonomic community composition and genes, these data can be further explored to study the bacterial growth dynamics and metabolic potentials via generation of small molecules and secondary metabolites. Everything from microbiome diagnosis to microbiome-based therapy will rely on vast amounts of data analysis. In this talk, I will present several computational and statistical method for analysis of data measured on phylogenetic tree and methods for estimating the bacterial growth rate for metagenome-assembled genomes (MAGs) and for predicting all biosynthetic gene clusters (BGCs) in bacterial genomes. The key statistical and computational tools used include optimal permutation recovery based on low-rank matrix projection and improved LSTM deep learning methods to improve prediction of BGCs. I will demonstrate the application of these methods using several ongoing microbiome studies of inflammatory bowel disease and cancer at the University of Pennsylvania.

Register at: https://purdue-edu.zoom.us/webinar/register/WN_f10ZkBd-Sd2A1I2Ydur3Mg