NIH NATIONAL CANCER INSTITUTE | PURDUE UNIVERSITY CENTER FOR CANCER RESEARCH | INDIANA UNIVERSITY MELVIN AND BREN SIMON COMPREHENSIVE CANCER CENTER | BIG CARE 2022

Big Data Training for Cancer Research

Special Lecture Series

Scalable Analysis of Large Biobanks and Whole Genome Sequencing Studies

Dr. Xihong Lin

May 24, 2022, 1:00 – 2:30 PM (EDT)

Abstract: Big data from genome, exposome, and phenome are becoming available at a rapidly increasing rate. Examples include Whole Genome Sequencing data, smartphone data, wearable devices, and Electronic Health Records (EHRs). A rapidly increasing number of large scale national and institutional biobanks have emerged worldwide. Biobanks integrate genotype, electronic health records, and lifestyle data, and is the trend of health science research. In this talk, I will discuss several analytic issues in analysis of large scale biobanks and population-based Whole Genome Sequencing (WGS) studies of common and rare genetic variants and EHRs. I will discuss scalable mixed model analysis using sparse genetic related matrix to account for relatedness and population structure; estimation of the number of ancestry principal components using Bulk Eigenvalue Matching Analysis (BEMA); and geometric differences in PCA of multiple phenotypes and PCA of multiple genotypes. The discussions are illustrated using ongoing large scale whole genome sequencing studies of the Genome Sequencing Program of the National Human Genome Research Institute and the Trans-Omics Precision Medicine Program from the National Heart, Lung and Blood Institute, and the UK Biobank and FinnGen.

Speaker Bio: Xihong Lin, PhD is Professor and former Chair of the Department of Biostatistics, Coordinating Director of the Program in Quantitative Genomics at the Harvard T. H. Chan School of Public Health, and Professor of the Department of Statistics at the Faculty of Arts and Sciences of Harvard University, and Associate Member of the Broad Institute of MIT and Harvard. Dr. Lin's research interests lie in development and application of scalable statistical and machine learning methods for analysis of massive high-throughput data from genome, exposome and phenome, as well as complex epidemiological, biobank and health data. Dr. Lin has been active in COVID-19 research. Dr. Lin received the MERIT Award (R37) (2007-2015) and the Outstanding Investigator Award (OIA) (R35) (2015-2022) from the National Cancer Institute (NCI). She is the contact PI of the Harvard Analysis Center of the NHGRI Genome Sequencing Program. Dr. Lin is an elected member of the National Academy of Medicine. She has received several prestigious awards including the 2002 Mortimer Spiegelman Award from the American Public Health Association, and the 2006 Presidents' Award of the Committee of Presidents of Statistical Societies (COPSS). She is an elected fellow of American Statistical Association, Institute of Mathematical Statistics, and International Statistical Institute. Dr. Lin is the former Chair of the COPSS (2010-2012) and a former member of the Committee of Applied and Theoretical Statistics of the National Academy of Science. She is the founding chair of the US Biostatistics Department Chair Group, and the founding co-chair of the Young Researcher Workshop of East-North American Region (ENAR) of International Biometric Society. She is the former Coordinating Editor of Biometrics and the founding co-editor of Statistics in Biosciences. She has served on a large number of committees of many statistical societies, and numerous NIH and NSF review panels.

Register at: https://purdue-edu.zoom.us/webinar/register/WN_KHX8EpXURR2ufJ-7ifdfuw