

Analysis of Large-Scale Biobanks and Whole Genome Sequencing Studies

Xihong Lin

Departments of Biostatistics and Department of Statistics
Harvard University and Broad Institute

Big data from genome, exposome, and phenome are becoming available at a rapidly increasing rate with no apparent end in sight. 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 discuss opportunities, analytic tools and resources, and challenges presented by large scale biobanks and population-based Whole Genome Sequencing (WGS) studies of common and rare genetic variants and EHRs. 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.