



## CBDS Distinguished Speaker Seminar Series

# “Scalable Integrative Analysis of Large Biobank and Population-Based Whole Genome Sequencing Studies with Multi-Omics Data”

**Xihong Lin, MS, PhD, Professor of Biostatistics**

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

Host: **Steven Kleinstein, PhD**

Anthony N. Brady Professor of Pathology; Co-Director of Graduate Studies,  
Computational Biology and Bioinformatics

Wednesday, December 7, 2022  
4 p.m. to 5 p.m. Zoom Seminar

**Zoom:** <https://yale.zoom.us/j/93538715108>

Telephone: 203-432-9666 or 646-568-7788

Meeting ID: 935 3871 5108

In person Zoom Gathering: TAC N203

RSVP to [cbds@yale.edu](mailto:cbds@yale.edu) if you plan to attend.



Whole Genome/Exome Sequencing (WGS/WES) data and Electronic Health Records (EHRs), such as large scale national and institutional biobanks, have emerged rapidly worldwide. In this talk, I will discuss the analytic tools and resources for scalable analysis of large scale biobank- and population-based Whole Genome Sequencing (WGS) association studies of common and rare variants by integrating WGS data with multi-faceted functional annotation data. Discussions include fitting mixed models for continuous and discrete and survival phenotypes using sparse GRM in population and biobank based studies, and rare variant association tests and meta-analysis by incorporating multi-faceted variant functional annotations using individual level data

and WGS summary statistics. I will also provide a demo of FAVOR ([favor.genohub.org](http://favor.genohub.org)), a variant functional annotation online portal and resource that provides multi-faceted functional annotations of genome-wide 9 billion variants, and FAVORAnnotator, a tool to functionally annotate any WGS/WES studies. Cloud-based platforms for these resources will be discussed.

The presentation will be illustrated using ongoing large scale population-based whole genome sequencing studies and biobanks of quantitative, case-control, and time-to-event phenotypes, including the Genome Sequencing Program (GSP) of the National Human Genome Research Institute and the Trans-Omics Precision Medicine Program (TOPMed) from the National Heart, Lung and Blood Institute, and the UK Biobank and FinnGen, which have collectively sequenced about 1 million genomes.

**Bio:** Xihong Lin, PhD is Professor 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 received the MERIT Award (R37) (2007-2015) and the Outstanding Investigator Award (OIA) (R35) (2015-2029) from the National Cancer Institute (NCI). She is the contact PI of the Harvard Analysis Center of the NHGRI Genome Sequencing Program, and the multiple PI of one of the Predictive Modeling Centers of the NHGRI Impact of Genomic Variation on Function (IGVF) 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, the 2006 Presidents' Award of the Committee of Presidents of Statistical Societies (COPSS), and the 2022 Marvin Zelen Leadership in Statistical Science Award. 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.