

**University of California, Irvine
Statistics Seminar**

***Scalable Statistical Inference for Analysis
of Massive Health Data: Challenges and Opportunities***

**Xihong Lin
Professor, Biostatistics
Harvard University**

**Thursday, October 11, 2018
4 p.m., 6011 Bren Hall
(Bldg. #314 on campus map)**

Massive data from genome, exposome, and phenome are becoming available at an increasing rate with no apparent end in sight. Examples include Whole Genome Sequencing data, smartphone data, wearable devices, and Electronic Medical Records. The emerging field of Health Data Science presents statisticians with many exciting research and training opportunities and challenges. Success in health data science requires scalable statistical inference integrated with computational science, information science and domain science. In this talk, I discuss some of such challenges and opportunities, and emphasize the importance of incorporating domain knowledge in health data science method development. I illustrate the key points using several use cases, including high-dimensional testing of dense and sparse signals for Whole Genome Sequencing (WGS) association studies, integrative analysis of different types and sources of data using causal mediation analysis, analysis of multiple phenotypes (pleiotropy) using biobanks and Electronic Medical Records (EMRs), reproducible and replicable research, and cloud computing.

For directions/parking information, please visit <https://uci.edu/visit/maps.php> and <http://www.ics.uci.edu/about/visit/index.php>