Abstract: The biological data deluge thanks to recent advances, has fundamentally transformed life sciences and biomedical research into a data science frontier. We witness an era of data acquisition on a broader scale, with finer accuracy, higher dimensionality, and higher throughput than ever. The unprecedented accumulation of genomic data presents a unique challenging opportunity to dive deep into understanding biology. To fully exploit big genomic data and enable translation of genomic analytics to precision medicine, we have developed a suite of machine learning methods toward predictive modeling of genomics data. In this talk, I will first review the current status of population sequencing of human genomes to chart human genetic variation map and assess the functional impact of genetic variants of various types in the 1000 Genomes Project and the Human Genome Structural Variation Consortium. Second, I will present our recent work of integrating vast genomic data for quantitative trait locus network analysis, epistatic analysis and phenotype prediction. Third, I will introduce emerging concerns of genetic privacy and summarize our recent investigation in this area. Finally, I will conclude the talk with discussions of future directions in infrastructure support that transforms beyond current high performance computing for big data genomics.

Bio: Xinghua (Mindy) Shi is an assistant professor at the Department of Bioinformatics and Genomics in the College of Computing and Informatics, the University of North Carolina at Charlotte. She received her Ph.D. and M.S. degrees in Computer Science from the University of Chicago, and M. Eng. and B. Eng. degrees in Computer Science from Beijing Institute of Technology, China. She was trained as a postdoctoral researcher and medical genetics fellow at Harvard Medical School and Broad Institute of Harvard and MIT. Her research areas include bioinformatics, machine learning, complex network analysis, data privacy and security. Aimed at solving large-scale computational problems in life sciences, her current research is focused on integrating multi-modal genomic data to study how genetic architecture affects biological processes and complex phenotypes at the systems level. Her research has been supported by NSF (including NSF CAREER award), NIH and DoD.