

**Speaker:** Yufeng Shen, Columbia University

**Date:** Wednesday, April 6, 2022

**Time:** 11:30 AM to 1:00 PM

**Host:** Bonnie Berger

**Title:** Predicting genetic effect of missense variants

**Abstract:** Accurate prediction of functional impact of missense variants is critically important for interpreting genome sequence in clinical diagnosis and new risk gene discovery by rare variants. Recent progress in machine learning and availability of large-scale population genomic sequencing data provide new opportunities to significantly improve computational predictions. In this talk I will present gMVP, a new method based on graph attention neural networks. Its main component is a graph with nodes encoding features of amino acids and edges encoding features based on coevolution, which enables pooling of information from both local protein sequence context and functionally correlated distal positions. We train the model using known pathogenic variants as positives and random missense variants as negatives. Objective evaluation of classification performance and generalizability is difficult in genomics, as labeled data are noisy and biased. I will go through performance evaluation based on data sets that have entirely different nature compared to training data: deep mutational scan data of a few genes and de novo mutation data from large-scale developmental disorder studies. Finally, I will show the model supports transfer learning to optimize gain- and loss-of-function predictions in sodium and calcium channels.