Speaker: Victoria Popic, Broad Institute
Date: Wednesday, April 20, 2022
Host: Bonnie Berger
Title: Structural variant discovery using deep learning

Abstract: Structural variants (SV) are the greatest source of genetic diversity in the human genome and play a pivotal role in diseases such as Alzheimer's, autism, autoimmune and cardiovascular disorders, and cancer. Current approaches rely on hand-crafted features and heuristics to model different SV classes, which cannot easily scale to the vast diversity of SV types nor fully harness all the information available in sequencing datasets. As a result, general SV discovery still remains an open problem. In this work I will motivate the use of deep learning for SV discovery, which allows us to shift method development away from ad hoc hand-engineered models and heuristics-based pipelines to scalable and sustainable models that can learn complex patterns of variation automatically from the data. I will lay out how SV calling can be formulated as a deep learning task and present our framework, Cue, to call and genotype SVs of diverse size and type, including complex and subclonal SVs. At a high level, Cue reduces SV discovery to a keypoint localization task in multi-channel images constructed from sequence alignments that capture multiple SV-informative signals and uses a stacked hourglass convolutional neural network to predict the type, genotype, and genomic locus of the SVs captured in each image.