Anomaly Detection for Precision Medicine

This talk will discuss the anomaly detection machine learning paradigm as a model for precision medicine. The idea is to characterize molecular data from an individual patient in a clinically-meaningful way by comparing it to a large population of control samples. Such one-sided learning allows for the identification of rare disorders or the characterization of common but molecularly heterogeneous ones. However, existing anomaly detection methods, particularly distance-based methods, fare poorly on problems with the dimensions and characteristics of genomic data.

We therefore introduce our results from developing a feature prediction approach that works well on both traditional machine learning data sets and gene expression data. We will discuss method evaluation and describe further work extending the anomaly detection model to incorporate functional interpretation. We will show how this approach has led to the discovery of new information about developmental disorders. Finally, we will discuss the relevance of such methods for characterizing sequence variation, which requires revisiting the issue of scalability, and present some recent results suggesting the feasibility of this approach.