Rethinking Experimental Design in High-dimensional Biology

In classical settings, a well-designed randomized study allows one to avoid confounding and potentially infer causality among variables of interest. High-dimensional biology brings a new set of challenges to this paradigm because thousands of highly related variables are considered simultaneously. In this talk, I will revisit the idea of well-designed randomized studies in the context of high-dimensional biology. First, I will demonstrate some problems that arise when a traditional randomized experiment is applied, and I will present a new technique called "surrogate variable analysis" for overcoming these problems. Second, I will show that a new experimental technique provides a powerful means for providing many local randomizations of "hardwired" components of the genome so that causal relationships among genes may be inferred on a large scale. I apply this technique to an experiment in yeast, showing how transcriptional regulatory networks may be inferred.