Abstract: High throughput biological data often contains signals from multiple unobserved latent variables in addition to the signal of primary interest. In a classification analysis, some of these latent variables may be partially correlated with the phenotype of interest and therefore helpful, some may be uncorrelated and thus merely contribute additional noise, while more perniciously, some may be spuriously correlated with the phenotype in the training set but not in the target population, leading to poor generalized predictive performance. Moreover, whether potentially helpful or not, these latent variables may obscure weaker direct effects that capture the signal of primary interest. It is therefore desirable to separate out these latent variables. This talk has two parts. The first outlines a classification algorithm that first isolates the signal of primary interest from other latent variables, but then exploits both to improve prediction, leading to sometimes substantial gains. The second discusses how to remove uncorrelated, non-stationary, or otherwise harmful latent variables.