

STATISTICS COLLOQUIUM

MONDAY, NOVEMBER 23, 2009

TALK: 4:00 PM — SCIENCE CENTER, ROOM 309

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

“Simple modeling strategies for micro-array and association studies”

**James Booth
Department of Biological Statistics
and Computational Biology
Cornell University**

ABSTRACT

Modern data sets often consist of large numbers of predictors and all samples sizes. This fact has generated a vast literature on statistical methods for the so-called large p small n problem. Examples in biology include gene expression data from micro-arrays, and association studies involving a large number of genetic markers and a given phenotype. In this talk I will discuss these of three simple modeling and computational strategies for such problems: (1) random effects to induce shrinkage and for model parsimony; (2) mixtures for empirical Bayes prediction and classification; and (3) computation via the EM algorithm. A manuscript giving details of the micro-array portion of the presentation is available online at <http://www.stat.cornell.edu/lemma>.