

STANFORD UNIVERSITY  
DEPARTMENT OF STATISTICS  
DEPARTMENTAL SEMINAR

4:15 p.m., Tuesday, November 2, 1999  
Sequoia Hall Rm. 200  
(Cookies at 3:45 in 1st Floor Lounge)

*Trevor Hastie & Rob Tibshirani*  
*Stanford University*

**New Statistical methods for DNA microarrays**

It is now possible to simultaneously measure the expression of thousands of genes during cellular differentiation and response, through the use of DNA microarrays. A major statistical task is to understand the structure of the data that arise from this technology. A typical data set consists of a few thousand gene expression measurements on each of about 50 samples. Hierarchical clustering has proven to be useful for describing the groupings of different samples and different genes. In this talk we describe a new method called "gene shaving" which tries to find clusters of genes that show large variation over the samples. The technique can be unsupervised, that is, treat the samples as unlabelled, or partially or fully supervised by known class labels for the samples. We illustrate the technique on some cancer tumor data.

This work is joint with David Botstein (Genetics, Stanford), Pat Brown (Biochemistry, Stanford) and Mike Eisen (Berkeley).