

**Statistics For Genomics – Reading Assignment 2**  
**Due: April 11<sup>th</sup> At the Beginning of Class**

**Read the papers (available from the class website):**

- (1) “Cluster analysis and display of genome-wide expression patterns”, by Eisen et al.
- (2) “Significance analysis of microarrays applied to the ionizing radiation response” by Tusher et al.

**Briefly Answer Each of These Questions:**

- (1) What are the differences between hierarchical clustering and differential expression analysis? When would you apply one method versus the other?
- (2) What would happen if you performed a differential expression analysis and then performed clustering? Would the patterns be informative? Why or why not?
- (3) What statistic does equation [1] in the Tusher paper look like? What is the purpose of  $s_0$ ?
- (4) Suppose there is a cluster of 20 tightly co-regulated genes in an experiment. Suppose the rest of the genes are just random noise. After hierarchical clustering, would the cluster of 20 genes look different if there were only 20 noise genes measured versus 100 versus 1,000? How so? What does this imply for clustering microarray measurements on 20,000 genes?