

Reading assignment on Differentially expressed genes

The two papers for this assignment describe methods for identifying differentially expressed genes. They both use the same affy data set (response to ionizing radiation) from an experiment that comprises replication for two conditions of interest -- see Reading List.

- Efron *et al.*, notwithstanding the availability of replication, actually model the data distribution as a mixture of two classes to estimate a posterior probability curve (“Empirical Bayes” refers to the fact that empirical distributions, as opposed to postulated Gaussians, are used in the mixture).
- Thusher *et al.* introduce the now well-known SAM approach (significance analysis of microarrays; software for this is available).

Both papers discuss the FDR (false discovery rate) as a measure to control the performance of multiple comparisons. This is a less conservative alternative to FEWER (family-wise type I error rate). As usual, pay attention to the criteria used to assess performance of a method. Also, ask yourself whether/how methods can be generalized to many conditions of interest.

Instructions:

Work in groups. Read the two papers carefully and critically, and discuss them among yourselves. Write up a report of at most 3 pages (typed), highlighting features that you found interesting, points that you found confusing and open questions. Due date: **Thur March 18th, in class.**