Due Friday April 11

1 Instructions

1. The goal for this part of the project is to run SAM on your data and then to evaluate your results using the multiple comparison ideas from class.

2. Just like with limma, I want you to run at least 2 comparisons. Here you'll need to run the comparisons separately. That is, run the entire SAM model twice with difference "cl" vectors (notice that's "l" as in class, not the number 1).

3. For the R code, you should use the siggenes userguide as well as the code / examples I gave in class.

2 Things to put on your web site (next)

- The significant genes from your SAM comparisons (you decide how many). Remember (from above), I want at least two comparisons, so you need at least two lists of genes.

- You should give a description of the "significance" of the genes. That is, I don't just want FDR, I want you to tell me what it is that these genes are able to discriminate. Give the action (reject or not reject), and then state the conclusion is words of the problem.

- For **both** the limma and the SAM comparisons, give some information about controlling for multiple comparisons. That is, choose a method of control (one of the FWER controls or the FDR), and update your list of genes to the leftover significant subset. Be clear about how you are controlling for multiple comparisons.

- A big picture discussion of the differential expression for your experiment. I want you to discuss the overlaps you saw in the previous assignment and this assignment. Are the gene lists identical? Similar? Why? What are the differences in the methods? What lists do you trust? What cutoffs do you trust? Just the level of significance? Or should you control FWER? Or maybe the FDR?