

# GENOME 560, Spring 2012

## Problem Set #5

(Will be done in class 5/31!)

---

### 1. Multiple Hypothesis Testing and Bootstrapping

Let's revisit the multiple hypothesis testing problem we discussed in class this Tuesday. In the gene expression data 'RMA\_Filtered.txt', we performed the t-test for each gene, comparing its expression levels in the first 16 columns with those in the next 16 columns. Here is what we did.

```
> a <- read.table(header = T,
file="http://www.cs.washington.edu/homes/suinlee/genome560/RMA_Filtered.txt")
> b <- a[,2:33]
>
> fun <- function(d){return(t.test(d[1:16],d[17:32])$p.value)}
> p <- apply(b, 1, fun)
```

Now, instead of applying p-value correction methods such as Bonferroni or B&H correction, let's use a permutation test-based method to figure out which p-value is significant.

- Permute the columns of the matrix  $b$  (above) and redo the t-test. (Hint: use the command 'sample' when permuting the columns.) Plot the distribution of the p-values using 'hist'. What is the minimum p-value out of all 5,194 p-values? Show the R commands you used.
- Now, we want to repeat the permutation test you did in part (a) 100 times and generate the distribution of the minimum p-values from the 100 tests. Unless you want to type the same commands 100 times, you should use the for loop. Here is how to use it.

```
> pmin <- rep(1,100)
> for (i in 1:100) {
> pmin[i] <- [DO SOMETHING]
> }
```

Show the distribution of the 100 minimum p-values from 100 using 'hist' and submit the R commands you used.

- Describe how you will use the 100 minimum p-values from part (b) to decide which p-value should be considered significant. How many genes have significant p-values based on that criteria?