

Genome 560  
Homework 1  
Due April 8, 2008

1. [10 pts] An opinion paper published in Nature Reviews Cancer titled “Bias as a threat to the validity of cancer molecular-marker reseach” is on the course web-site. Read it before answering this question.

Dr. Michael J. Maccoss was recently profiled on King 5 news, along with Clem Furlong, for their work on identifying a jet oil additive that may be causing illness in commercial airline crews. Dr. Maccoss has decided to only pursue high profile projects because he likes the way he looks on TV. Next up on the agenda, is a comparative proteomics experiment to identify possible protein biomarkers for early diagnosis of Type 2 Diabetes. The basic study design is to analyze blood samples obtained from 50 cases and 50 controls using mass spectrometry (or tandem ms-ms, or whatever funny terms protein people use). Assume that 10 samples can be analyzed in a single run of the mass spec. Describe an experimental design that will allow valid statistical inferences to be made. Comment on potential confounding variables and how the study design you propose minimizes their affect.

2. [15 pts] Download the data file “RMA\_Filtered.txt” from the course web-site. Using R, do the following:

- A. In class, we calculated simple descriptive statistics for the first probeset, including the overall mean, the CEU specific mean, and the YRI specific mean. Repeat these calculations for all 5194 probesets (first average the replicates as we did in class; thus each probeset will consist of 8 CEU expression levels and 8 YRI expression levels). As a hint, there are many ways to do this including a simple for loop or more efficiently with the R command `tapply`. Submit the commands that you used to obtain the desired means.
- B. Make a boxplot consisting of the 5194 overall mean expression levels, CEU specific means, and YRI specific means. Submit the commands you used and the figure.
- C. For each probeset calculate the difference in average gene expression levels between the CEU and YRI samples. Submit the commands that you used to generate the result.
- D. Make a histogram of the results obtained in part C above. Modify the histogram’s x-label, y-label, and title to be more descriptive than the default values. Submit the figure.