Bios 555: High throughput data analysis using R and Bioconductor

Homework 2

The written assignment is due on **Sep 26, Wednesday before class at 3pm**.

I. Short answer questions, 10 points each. Be creative in answering the questions.

1. What does a gene expression microarray measure? Why is it important?
2. What is hybridization? The amount of hybridization on different probes are extracted from the images and called "fluorescent intensities". What do the fluorescent intensities of each probe represent?
3. Why does one have to normalize the microarray data?
4. What is quantile normalization? Can you think of any pitfall of the procedure?
5. What is summarization of gene expression? Can you design a simple method to do the summarization?
6. In DE (differential expression) detection, why a simple gene-by-gene t-test is not ideal? What do people usually do to avoid this problem?
7. Empirical Bayes (EB) methods are widely used for microarray DE test. The main idea is shrinkage estimator under a hierarchical model. Describe what a shrinkage estimator is and why it is useful.

II. Based on the results obtained from the lab, write a short report to present the exploratory analysis results of comparing microarray generated from different platforms and the gold standards (30 pts). You can read https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3026357/ to get extra information about the experiment.

III. (15 Bonus points) Using Taqman data as gold standard, compare the performance of SAM and limma for the U133A array data. Comparison can be performed in various ways, such as ROC, FDR, power, etc. Be creative.