Exam III Outline – Gene Expression Analysis

1. Relationship between two-sample t-test and a linear model with coded variables.
2. Reading in raw data (.CEL files) from GEO and processing the data using robust multi-array average (RMA). *(omitted)*
3. Reading in processed data from the Gene Expression Omnibus (GEO) using the *getGEO* function from the *GEOquery* library. **(omitted)**
4. Evaluating whether a single gene is differentially expressed, using the two-sample t-test, constructing a boxplot and calculating its fold-change and *p*-value.
5. Identifying differentially expressed genes using the *limma* package and understanding false discovery rates (FDRs)
6. Generating heatmaps
7. Converting probe name to gene symbol and vice versa using the platform (GPL) data provided by GEO
8. Classification using *k*-nearest neighbors *(knn)*, including leave-one-out classification, optimization, and making predictions in a test dataset.
9. Perform a gene set enrichment analysis using DAVID, based on a list of genes
10. Additional data analysis and R programming concepts (based on your Piazza questions)