CSC 314, Bioinformatics Lab #11   Name:_____
The Gene Expression Omnibus

GEO Questions

1. Go to the GEO homepage (http://www.ncbi.nlm.nih.gov/geo/), search for *lung adenocarcinoma* (do not use quotes), and look at the results in the GEO DataSets database (which includes the 4 kinds of records discussed previously).

    a. How many GEO series (studies) are there *for humans*?

    b. How many GEO samples are there *for humans*?

2. Look at the GEO dataset entry, *Cigarette smoking effect on lung adenocarcinoma*, (Dataset Record GDS3257).

    a. How many samples are in this dataset?

    b. What is the platform the samples are profiled on? (Name and Platform ID)

3. Look at the cluster heatmap using the default settings. Based on gene expression profiles alone, the samples separate (cluster) predominantly into what two biological groups?

4. Find genes that are differentially expressed between males and females. How many probes are significant (at 0.05)?

5. Next to *Display Settings* near the top of the page, click on *Sorted By* and change the default sort preference to *Sorted by Subgroup Effect.* For the top two unique differentially expressed <u>genes</u> (not probes), identify the gene symbol, whether it has higher expression in males or females, and the chromosome the gene is on (you can click on the Gene link for this information)

6. Does it make sense that these genes would be differentially expressed? Why or why not?

7. Look at the record for the corresponding GSE Series for this dataset. Was the sample from the patient corresponding to GSM254628 a male or female? How old was this person when the sample was profiled?

8. For this individual, what is the expression value for probe 1773_at (round to 2 decimal places)? What is the gene symbol that this probe corresponds to (you can give the gene symbol)?

9. Find the top 250 differentially expressed genes that are differentially expressed between "Normal Lung Tissue" and "Adenocarcinoma of the Lung".

    a. How many normal samples are there? How many lung samples are there?

    b. Which probe <u>and</u> gene is the most significantly differentially expressed?

    c. The adjusted p-value indicates that the probability that this probe is a false positive is _____?

    d. The p-value for this gene corresponds to the following probability: if there really was no difference between the tumor and normal samples, the probability of observing a log2 fold change of at least _____ is equal to _____?

    e. Is the probe up-regulated or down-regulated in tumor samples?