CSC 314**, Bioinformatics Lab #7 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).
   1. How many GEO series (studies) are there *for humans*?
   2. How many GEO samples are there *for humans*?
2. Look at the GEO dataset entry, *Cigarette smoking effect on lung adenocarcinoma*, (Dataset Record GDS3257).
   1. How many samples are in this dataset?
   2. 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 genes (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".
   1. How many normal samples are there? How many lung samples are there?
   2. Which probe and gene is the most significantly differentially expressed?
   3. The adjusted p-value indicates that the probability that this probe is a false positive is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_?
   4. 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 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_?
   5. Is the probe up-regulated or down-regulated in tumor samples?

**Probability and Statistics Questions**

1. Suppose that a bag contains 10 green marbles, 10 blue marbles, and 5 red marbles.
   1. If one marble is randomly selected, what is the probability that it is blue?
   2. If one marble is randomly selected, what is the probability that it is red?
   3. If two marbles are randomly selected, what is the probability that they are both green?
2. The answer to the Birthday Problem is that in a room with 23 people, there is more than a 50% chance that at least 2 people share the same birthday (assuming all birthdays are equally likely and 365 days in a year). However, the probability that someone has the same birthday as *you* is much lower. You will calculate this probability by completing the following steps.
   1. In a room with 2 people (you and one other person), what is the probability that the other person has a *different* birthday than you (Hint: how many birthdays are different than yours? How many possible birthdays are there?)?
   2. In a room with 3 people (including you), what is the probability that the other two people have a *different* birthday than you.
   3. In a room with 23 people, what is the probability that *nobody* has the same birthday as you?
   4. In a room with 23 people, what is the probability that at least one other person has the *same* birthday as you?
3. Hypothesis testing:
   1. An analysis for whether a gene is differentially expressed between males and females finds that expression is 4 times higher in males than females (P = 0.20). Is there evidence that this gene is differentially expressed? Why or why not?
   2. An analysis for whether a gene is differentially expressed between individuals with and without the flu finds that expression is 1.3 times higher in those with the flu. Is there evidence that this gene is differentially expressed? Why or why not?