Bio 311: Group Clustering HW

Due date: March 30, 2017

Intsructions: Answer the following questions according to your analysis of the data so far. Choose a representative from your group and submit your group's answers to the representative's Sakai dropbox by the *beginning of class* on the due date. Email the instructors and TA to inform them whose dropbox is being used for your group.

- Describe your biological question and project scope. For this clustering assignment, did you choose to work with the entire data set, or a subset of the data? How did you make this choice?

- Carry out clustering analyses of the gene expression data you have been provided with, exploring multiple methods and distance metrics as you see fit.

- Describe your clustering methods. What clustering method(s) did you use? What distance metrics? If you used hierarchical clustering, which algorithm did you use? When determining clusters from a hierarchical dendrogram, which cut height did you use? What was your rationale for using that cut height? If you used k-means clustering, what value of k did you choose and why? Were you able to come up with a clustering that you though worked "best" for your data? What criteria did you use to judge utility of the clusterings you generated?

- What is the distribution of sizes of clusters in your analysis (few clusters with many genes or many clusters with few genes)?

- If your focal paper(s) described cluster analyses of the expression data, how do there results compare to yours?

- Generate an overview figure for your clustering analysis that you can use for your poster. In your figure, include a heatmap and dendrogram (if you used heirarchical clustering), or other graphical representation of your clustering analysis.

- Generate some figures illustrating the dynamical behavior of genes within some of the clusters you delineated.