Bio 311: Group Transcription Factor (TF) Network HW

Due date: April 11, 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 Drs. Magwene and Schmid to inform them whose dropbox is being used for your group.

1. State briefly the TF(s) your group is working on and the biological questions of focus for your group. Has your biological question and/or project scope changed after your clustering analysis? If so, how and why did you make this choice?
2. In what form is the information about TF binding provided for your TFs of interest (e.g. p-values, enrichment over background, etc)? What thresholds or criteria did the authors of your focal paper(s) use to determine which genes were bound by your TFs of interest?
3. Briefly describe your analysis pipeline you used to generate your network. For example, what p-value cutoff did you choose for the ChIP-chip data and why? If you have TF knockout data, how did you determine which genes are differentially expressed in response to the TF knockout(s)? If you have time series data without TF knockouts, how did you determine which TFs were contained in which cluster?
4. How many genes do each of your TFs of interest appear to regulate, based on the TF-DNA (ChIP-chip or -seq) binding data you've been provided?
5. Carry out a TF network analysis relevant to your project scope using igraph and/or ggraph as demonstrated in class. Provide a visual representation of one or major connected components of your TF network as a figure (hint: igraph can make visualizations, but Cytoscape http://www.cytoscape.org/ is also a useful program for visualizing and analyzing iGraph results).
6. Describe your network: What do the edges of your network represent? The nodes? Is your graph directed? Undirected? How many total nodes and edges are present in your graph? How many TF(s)? How many target genes?
7. Analyze your network: What is the out-degree for your TF(s)? What is the distribution of in-degree across the target genes (e.g. a histogram could represent this result)? (Hint: The Lee et al. paper did such an analysis in one of the figures...).
8. Synthesize information from your network to think about the function of the TF(s) of interest. How many target genes are direct targets of the TF? Indirect? Which dataset(s) did you use to determine this and how? What are the functions of the target genes? What functional categories are enriched among the target genes? Are any of the target genes themselves also TFs?

9. Do sets of genes bound by your TFs fall into clusters, as determined by your gene expression clustering analysis?