Name: ______________________________

BIOINF 525 Module 3
Lab #1
3/23/2017

Please complete the exercises below. Throughout the lab sessions for this module, we will use the following notation:

Plain text indicates actions that should be taken
Italicized text indicates explanatory material
Bold text indicates a point where a written response is required

---

Exercise 1 – Network design

Working with an assigned group of peers, design a BioBrick-based construct that would yield a transient burst of GFP expression when E. coli cells bearing the plasmid undergo cold shock (20 C) while growing in glucose minimal media. **Indicate the part numbers to be assembled (in order), draw a schematic of the resulting mini-network, and explain why your construct will implement the desired function.**

BBa_J45503 (cold inducible promoter) – BBa_J61100 (standard RBS) -- BBa_C0051 (lambda cI protein) -- BBa_B0010 (standard terminator) --

BBa_I12007 (lambda cI inducible promoter) – BBa_J61100 (standard RBS) – Bba)C0056 (434 cI protein) -- BBa_B0010 (standard terminator) --

BBa_I12040 (lambda cI induced/434 cI repressed promoter) -- BBa_J61100 (standard RBS) -- BBa_J97001 (GFP) -- BBa_B0010 (standard terminator)

The key to this exercise is recognizing that the needed logic is an incoherent FFL; thus, cold shock should trigger production of a transcription factor (TF) that directly activates GFP expression, but that TF should also cause synthesis of a second TF that represses GFP expression. The above set of parts implements that logic: the cold shock causes production of lambda cI, which activates both GFP and 434 cI production. 434 cI represses production through the particular logic of promoter BBa_I12040. The network looks roughly like

```
Cold → lambda cI → GFP
        \         --
         \        /
          \      /
           434 cI
```

Exercise 2 – Network analysis with Cytoscape

We will use the Cytoscape package to view and analyze a variety of biological networks, and to understand their properties.

First, we introduce different methods to find and import data in Cytoscape.

Start a new cytoscape session with an empty network

Install the WikiPathways add-on using the Apps->App Manager window

WikiPathways provides simple import of a variety of publicly curated biological networks. First we will use it to examine the network regulating apoptosis in human cells.

Import the human apoptosis regulatory network using File -> Import -> Network -> Public Databases. Select WikiPathways as the data source, and search for apoptosis. Select the entry for Homo sapiens, choose "Import as Pathway" from the pulldown menu at the bottom left, and then push that button.

After visually examining the network, follow the same steps as above, but push the pulldown button next to "Import as Pathway" and instead choose "Import as Network". Compare the two representations of this network.

The pathway view is a better overview of the network, but the Network view is more suitable for analysis of connectivity and network properties.

Identify at least one feed-forward loop manually using the Network view, and list the components and type of that loop here:

Coherent type 1 FFL:
CASP11 activates CASP1 and CASP3; CASP1 activates CASP3

Next, we will perform simple analysis of graph properties for this regulatory network in Cytoscape, and compare the results with those for a C. elegans *neural network.*

With the apoptosis network from the previous example shown in Network view, choose Tools -> Network Analyzer -> Network Analysis -> Analyze network. Acknowledge the warning regarding directed and undirected edges.

Examine the results in the "Simple Parameters" tab – a description of these values can be found at
http://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/index.html#simple

Record here the average distance between two nodes in the network:
4.87 (this is the characteristic path length)

To provide a comparison, download a connectivity map of the *C. elegans* brain from http://bit.ly/2ndU09X
(network from D. J. Watts and S. H. Strogatz, *Nature* **393**, 440-442 (1998)).

To load the data, choose File -> Import -> Network -> File
Choose the celegansneural.gml file that you downloaded
Choose "Create new network collection" from the pulldown menu and press ok

Initially the network will be displayed using a highly uninformative layout. To re-arrange the nodes to make them more visible, select Layout -> yFiles Layouts -> Organic.

Now we will compare some basic properties of the apoptosis regulatory network and C. elegans neural network.

Run the "Analyze network" command on the *C. elegans* data exactly as you did for the apoptosis regulatory network, treating the network as undirected, and compare the basic parameters of the networks. Note that you may find it useful to use the bar with an arrow at the top of the "Results Panel" to switch between network results.

Which network is more densely connected?

The C. elegans neural network (higher network density)

Which network has a shorter average distance between nodes?

The C. elegans neural network (characteristic path length is <2.5)

As you should recall from lecture, scale-free layouts are common in some types of biological networks, and are characterized by a node degree distribution that follows the equation $P(k) \sim k^{-\gamma}$. This also means that a plot of log(frequency) vs. log(node degree) will be linear.

Inspect the <u>node degree distributions</u> for both the *C. elegans* neural network and apoptosis network. These are accessible by clicking the right arrow near the top of the network analysis results, near where it says "Simple Parameters".

Do either or both of the networks that you are considering show scale-free layouts? If so, which ones? Why might this be the case?

Only the apoptosis network shows a scale-free layout. Likely this is because of the physical constraints in the neural network limiting connectivity.

The constant gamma in the equation of the node degree distribution equation above can be obtained by fitting a power-law type equation to the degree distribution. This can be done by using the "Fit power law" button.

If either or both of your networks showed likely scale-free behavior, calculate the value of the coefficient γ in the power law equation ('b' in the fit from Cytoscape).

The fit for the apoptosis network gives b=-1.759

Now we will use Cytoscape to reproduce some classic results on the S. cerevisiae galactose utilization network. We will consider a mixed network containing protein-DNA and protein-protein interactions regulating galactose metabolism in yeast. This example is adapted with some variations from a longer tutorial given in Nat. Protoc. 2:2366-2382 (2007).

Download the bundle of data files containing the galactose utilization network and some associated gene expression data from http://go.nature.com/2mRFXTO

Exit your Cytoscape session and then begin a new session. On startup, choose to initialize "from network file" and select the galFiltered.sif file from the bundle that you downloaded.

Once the network is loaded, since it contains directed edges and information on edge types, it is useful to go to the "Style" tab in the top left of the Cytoscape interface, and switch from 'default' to 'directed'. Zooming in will reveal both the directionality of each regulatory interaction and whether it represents a protein-protein (pp) or protein-DNA (pd) interaction.

Furthermore, you can make the nodes appear with more meaningful names than the standard systemic nomenclature. To do so, go to File->Import->Table->File, select galGeneNames.csv, and load it into the galFiltered.sif network using default settings. Then, back in the Style menu, you can change "Label" to use the column called "Gene Name".

At this point, we'd like to see whether any of the classical network motifs are apparent in the galactose utilization network. We will use the NetMatch plugin, described in detail at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4642848/*

Using the Cytoscape App Manager, install the NetMatchStar App, and then choose Apps->NetMatch* from the main menu to activate it.

To test for whether a particular motif is enriched in your target network, a query must be constructed and matched against the target. To generate a FFL query, choose the NetMatch* tab of the control panel, and then push the + button to generate a query network.

Right clicking on the (initially empty) query network will open a menu that allows you to add nodes or edges. Using those tools, draw a feed-forward loop in the query network window. Once you are done, save the ffl query network using the disk icon in the left panel.

With the FFL query saved, go to the main NetMatch* menu and select the FFL that you drew as the Query and galFiltered.sif as the target network. Make sure that "labeled" is unchecked and "directed" is checked under Graph Properties, and then click "Match" to identify all FFLs in the network.

How many feed-forward loops are present in the galactose utilization network? Give the gene names involved in one of them.

22
One example contains YGL153W,YLR191W,YDR244W

To check whether there is a significant enrichment of FFLs in this network, we need to compare to randomly generated data. In the bottom left of the main NetMatch* window, select the "Significance" tab.

Netmatch contains several possible methods for generating random networks, detailed in the paper cited above. For now, we will use the simple Shuffling model, which randomizes the edges in the network while preserving the degree distribution.*

Make sure that the "Shuffling" radio button is selected, and press "Start" under the list of models to generate a set of 50 random networks and compare the abundance of FFLs in them to that found in the galactose network.

Are feed-forward loops over-represented in the galactose network?

Yes; only an average of 2 occurred in the random networks

What Z-score did you obtain for your sampling? (the Z-score is the difference of an observed value from the mean, divided by the standard deviation)

9.997 (note that this will vary somewhat depending on the random sampling)

Give an R command that would calculate the p-value of this observation, under the null hypothesis that the number of occurrences of FFLs in the network is normally distributed with a mean and standard deviation equal to those observed in the random sampling.[1]

```
1-pnorm(9.997)
```

As a final step in analyzing the galactose data, we will see how Cytoscape can be used to map gene expression data onto a network and analyze the network behavior in light of that information.

To load the gene expression data, go to File->Import->Table->File, and select the galExpData.pvals file that you downloaded. Load it into the galFiltered.sif network.

The data table contains log fold changes and corresponding p values for expression levels of genes in the network upon knockout of the central regulators Gal1, Gal80, or Gal4.

To color nodes in the network by their log fold change values, go to the “Style” tab, choose “Fill Color”, and change the Column entry to one of the expression columns and the Mapping Type to Continuous Mapping. Change the Column to look at the expression data set in each of the three knockout conditions. *(note that positive log fold changes indicate increases in transcription in the knockout)*

GAL4 is a transcription factor labeled (also labeled YPL248C) in the network shown here. Based on the log fold changes seen for GAL4 targets in the gal4 knockout data set, do you think it is a transcriptional activator or repressor? Why? (“Because yeastgenome.org told me so” is not a valid answer)

Activator. Most of its targets drop in abundance when it is knocked out.

Finally, we want to identify clusters of genes that are most similarly affected by a particular perturbation. Install the jActiveModules App, and select the corresponding tab in the control panel. Highlight the gal4 expression data set, and then click Search.

[1] This is a terrible assumption because the number of FFLs is a small, non-negative integer, but we use it here for the sake of simplicity

Once the search is completed, you can see the set of identified modules by pressing 'g' (show grid) and then selecting the jActiveModules search result. Clicking on each of the modules will display a Z score at the bottom (Z-scores greater than 3 in this analysis indicate a strong module).

Go back to the grid view (press g again) and choose the highest-scoring module that was identified, then double click on it to see the details. (You may need to change the Style back to 'Directed' to get the view that you are used to).

What is the most highly connected node in the identified subnetwork? Do you recognize any classic network motifs? If so, what genes are involved, and what is the organization of the motif?

YPL248C/GAL4 is at the center of the network. The identified module contains a feed forward loop containing YPL248C -> (YML051W) -> YBR020W