GCBA/MGCB/BMI 815, Fall 2017 **Instructor: Babu Guda**

Due date: 8am, 12/11/17 (Monday)

Final Project Questions: Total 40 points

Question-1 (8 points)

CYTOSCAPE

Download and install Cytoscape as instructed in class and finish the following tasks. Working on a 64-bit machine is highly recommended, choose Cytoscape version accordingly.

-Load data from public databases using Data source-Universal Interaction Database Client; Search condition-Import Interactome; and Species-as assigned to you.

a) How many nodes and edges are present in the initial network? Remove the duplicate edges and self-loops, and provide the updated numbers on the nodes and edges.

-Search the node by the gene name assigned to you in the network. If there are multiple entries, select using the Uniprot ID provided corresponding to your gene. Select the first neighbors of the selected gene. Create a sub network with only the first neighbors of the gene assigned to you.

b) How many nodes and edges are present in the resulting sub-network?

-Change the style of the sub-network using the control panel on the left side. Do not remove the previous style (PSIMI 25 Style). Make the following changes to the network and paste the network image in your report.

a. By default the network layout is in 'Prefuse type'. Change the layout to "Organic". There are chances that you may not see any changes in the layout because of the size of the sub network.
b. Lock the node width and height
c. Change node label font size to 10.
d. Change edge color to red. Make sure you check the box for the "Edge color to arrows" before changing the edge color.

Question-2 (7 points)

Using GDC website: https://portal.gdc.cancer.gov/
Go to Repository link, select 'Cases' tab and select 'TCGA' (under 'Program' option)

Perform all the tasks below using ONLY the cancer type assigned to you.
1. How many cases exist in the cancer type assigned to you (use 'Cases' tab).
2. List the number of files in each 'Sample Type' (Primary Tumor, Blood Derived Normal, Solid Tissue Normal) in the cancer type assigned to you
3. List the number of files in each 'Data Category' (Clinical, Copy number, transcriptome profiling, etc.) in the cancer type assigned to you. Note that not every cancer type has data in all categories.
4. How many files are available for download with RNA-Seq data in aligned format (.bam files) in the cancer type assigned to you.

5. From the 'Exploration link', Identify the gene that has mutations in most number of patients in the cancer type assigned to you.

Question-3 (8 points)

MG-RAST http://metagenomics.anl.gov/

Each student has been assigned a separate State/Location for the project. Use MG-RAST web portal to answer the following questions. Provide screenshots of your answers where required.

1. How many projects are there for your location?
2. Choose the first project in the list. Provide Project-ID and state how many metagenomes are available in this project?
3. Select the metagenome with the highest number of sequence count (you can sort if there are many).
    a. How many sequences failed QC? Submit the Pie-chart showing the QC
    b. What was the 'Sequencing Method' used for the metagenome?
    c. Submit the 'Source Hit Distribution' pie chart
    d. Submit the 'Taxonomic Hits distribution' at the 'Genus' level pie chart
    e. What is the most abundant 'Phylum' in the metagenome. Include figure.
    f. What is the Alpha diversity of the metagenome? Include figure

Question-4 (8 points)

Ingenuity Pathway Analysis

Login into IPA (Ingenuity Pathway Analysis) and carryout analysis using your assigned gene list and species, and share the following files (with babu.guda@unmc.edu) using the 'Share Manager'.

a) Upload data and do the 'Core analysis'. Share the Core analysis file.
b) Pick the 'Top canonical pathway' identified in your Core analysis and retrieve its image from the top search bar under Pathways tab. Use 'Overlay' button to see the left panel and select 'Biomarkers' option from the pull-down menu. Select preferably 'prostate cancer' biomarkers, if any; if not, select 'any cancer' biomarkers. It will highlight the biomarkers on the pathway image in purple color (Note: Make sure to select the 'highlight' option at the bottom as shown here. 

Save the image with the pathway name. It will be save under 'My pathways' subfolder in your working folder. Share this file using Share manager.

Question-5 (9 points)

Perform the following tasks using Vector NTI software
Select a protein of your choice and find out the following. You may add your favorite protein from any source or from the local database. Please include screen shots from the program that support your answers, where applicable. Images may be resized to fit into your document.

1. Provide the following details on the protein you selected
    a. Protein accession number or UniProt ID
    b. Descriptive name of the protein
    c. Length (how many amino acids)
    d. Molecular weight
    e. What's the most frequent amino acid in your protein (from composition)

2. Perform Pfam database search and list the unique Pfam domain names (NOT Pfam Ids or numbers) present in your protein sequence
3. Run a gel for the protein along with a marker and find out how many fragments are generated from the protein and approximately how long the gel should run to separate all fragments
4. Back-translate the protein sequence and create a new DNA sequence. Find ORFs in the new DNA sequence, select an ORF and add a mutation (single-base insertion, deletion or substitution) to the sequence. Submit the screenshot of the mutated sequence.
5. Design primers for the selected ORF sequence from above (Your sequence of selection must be longer than the target to design primers) and submit the screenshot of the primer sequence details.

How to submit your project?

- Your final project should contain a single Word file containing your responses to all questions. Two files from IPA analysis (Question-4) must be shared directly from IPA.
- All the results (tables, screenshots or other images) should be inserted into your Word document. The Word file name MUST start with your last name.
- Your description of results should be brief and to the point (1-2 sentences)
- Please email your project file to babu.guda@unmc.edu by 8 AM on 12/11/17 (Monday)