Bioinformatics Lab: Part 2

This is part 2 of the bioinformatics lab, which will guide you through using your DEGs that you found from part 1 of this lab. You'll start with importing the list of your DEGs that you found into String-db (functional protein associated networks) String-db, then exporting this network created in String-db into Cytoscape (Cytoscape) for enrichment analysis. For this assignment you will be answering questions, but also including figures of what you have generated (these will be marked in bold). You can answer these questions on a separate document.

Exporting DEGs from R

1.) From part 1 of the lab, you should have generated individual results for each comparison between control and treatment conditions. (Total of three lists, LPS.results, BaP.results, LPS.BaP.results). For part 2, you'll need only the DEGs from these lists.

You can generate these in a similar fashion, except you need to specify **p.value=0.05, lfc=1.2, adjust.method="none"**, where p.value and lfc indicate the cutoffs for DEGs and adjust.method specifies not to use any p-value adjustment. Here, we're using no p-value adjustment so we can get more DEGs, but keep in mind that there could be quite a few false positives (whereas an adjustment method like Benjamini-Hochberg would have fewer false positives, but more false negatives; your choice of adjustment method depends on how conservative you would like to be, and in this case we don't want to be conservative at all).

Below is the line of code for LPS vs control. Repeat this with BaP vs control and LPS.BaP vs control for a total of 3 DEG lists (LPS.DEGs, BaP.DEGs, and LPS.BaP.DEGs).

```
LPS.DEGs <- topTable(fit, coef="factorsLPS", number=Inf,
                                    p.value=0.05, lfc=1.2, adjust.method="none")
```

Once you've created your 3 lists, export them via the code below. They will save as .csv files in your working directory which you can then open in Excel or Google Sheets. Moving forward in this lab, you can copy the list of DEGs by copying the leftmost column of each .csv file, which will consist of numeric gene IDs.

```
write.csv(LPS.DEGs, "LPS_DEGs.csv")
write.csv(BaP.DEGs, "BaP_DEGs.csv")
write.csv(LPS.BaP.DEGs, "LPS_BaP_DEGs.csv")
```

You will no longer need R for the remainder of the lab, so knit your R markdown file for submission.

Using String-db

2.) For the lists of DEGs for each comparison that you just exported, you will be running them through String-db. Link to string-db here String-db.

These lists of DEGs you have can be copy-pasted directly into the search engine under the "Multiple Proteins section", one gene per line. You should also specify the organism that you are searching for, consider the model we are using and search for the scientific name. We previously touched in class how string-db creates the protein-interaction network. However, for your additional reference and reading you can see the help section for String-db that covers some aspects of how String-db creates these networks.

You can export any of the networks you create in string-db under the "Exports" sections as a PNG image or other images. **Export them and include them as images in this document. Should have an image for each list (total of three images)**

Try clicking through your network, clicking on nodes (circles) and edges (connecting lines) and see what information you are getting. This will help with the further questions.

Q.1) Do all your networks look like each other or are there differences?
Q.2) Look at some of the nodes (circles) in your networks, are all the nodes connected within the network? If some are unconnected what does this tell you about these nodes (genes)?
Q.3) Look at your networks, is it one entire big network or is there multiple small network clusters? If there are any smaller clusters, click on those nodes within the cluster and look at them. What do all the nodes in the smaller cluster (s) have in common? It is okay if you don't know many of the biological processes or functions, but is there common keywords/themes that you see pop up?
Q4.) Choose any node of interest to you within your networks, give the name of the gene in your node and some information you found from it in String-db. From the information you found from this gene, why do you think this gene is being affected (depending on what network you are looking at) from the treatment itself.

Using Cytoscape

3.) String-db is a helpful tool to establish an initial protein-interaction network for your DEGs, in addition there is another tool named Cytoscape where additional annotation and analysis can be done for your networks. Importing networks you made in string-db is as easy as clicking on "send network to Cytoscape" option under the "Exports" section in String-db.

You can download Cytoscape here at the download link. To import your string-db into Cytoscape you will also be using an app called stringApp. To download apps for Cytoscape you can do it internally through the Apps tab → Apps manager → Type the name of the app in the search bar and install. You can also install it externally through the Cytoscape app page here here. Once downloaded you can go ahead and open Cytoscape and import your network in by clicking the send network button in String-db. When importing multiple networks, you should see them pop up in Cytoscape (pay attention to what they are named, you can change their names to anything). Take some time to get familiar with the program, don't be afraid to click around and try things out. When you network is imported in Cytoscape it should look like your network in String-db, however some of the orientation could be changed.

In Cytoscape you can zoom in and out of your network (you will see a blue box over what is currently in view in the smaller window). Once your whole network is in view, export an image of your networks in Cytoscape by clicking on file on the top left and exporting your network as an image.

We will be using Cytoscape to perform enrichment analysis, where biological pathways that are enriched in a gene list can be identified. Cytoscape's app BiNGO can be used for this. The link to download this app for Cytoscape is here. Here you can also read up on how this app is performing this enrichment analysis. If you want to read up on BiNGO interpretation some additional resources can be found here in the user's manual **(I strongly advise you go through this manual before making your interpretations).** Once downloaded you can select this app from the Apps drop-down menu (inside Cytoscape) and continue.

You can input a name for your cluster, and you can also select paste genes from text to paste your entire network list in the text box. You can leave all settings on default besides the organism which you can choose our organism's scientific name from the dropdown list and move forward (it will also make you choose a file location for the output). Once you have all your information you can then click start BiNGO. BiNGO will open a separate tab for data output, use this to answer the following questions. Tip: You can also change the resolution of your screen to fit the BinGO network (try different resolutions).

Answer the following questions using all your networks (just make sure to mention what each network is titled).

Q.5) Take a look at your GO descriptions, what are the top terms if you sort by cluster frequency (double-click the column header)? These terms are meant to be informative and help you

understand some patterns in your data, are some of these terms unspecific or not helpful? Out of the top ten terms which one seems the most informative? What is the cluster frequency of this term and how do you think this term helps give an understand of what is going on in the condition response?

Q.6) Try sorting by Corrected p-value, look at the top 20 terms. Out of these which three seems the most informative to you and why (tie it back in to condition response). Sometimes sorting by different columns can give you a different outlook on your data.

Q.7) When sorting by Corrected p-value, are your terms similar across your networks that you generated? If not, what are some differences that are apparent in which terms are present? If there are some differences, how do you think this can inform us about the dataset and the response differences and/or similarities occurring between conditions?

Include a screenshot of your enrichment analysis for your networks (don't have to include all terms just a screenshot of what is visible at first on your output).

BiNGO also generates a network of your enrichment analysis, include a screenshot of this network for all three of your networks. You don't have to screenshot the entire network; this can be done on only a segment of the network (maybe a segment you found most informative or interesting).

Run BiNGO on a smaller cluster of nodes in your dataset (a smaller branching network that you see). You can do this by Control + Dragging or Command + Dragging (for macs) your cursor over your network to highlight a group of nodes. Once you do this, your selected nodes will show up in your node tables. You can then copy the column of gene names that you have selected (you can paste this into your BiNGO). Do this for all three of your networks.

Q.8) How does this BiNGO analysis differ from the full network BiNGO analysis that you did (sort by corrected-p value), what are some differences in terms that you see? What does this BiNGO analysis tell you about the smaller cluster within your network?

Q.9) What major systems to you believe are being affected in the mice from their treatment conditions after looking at this enrichment analysis?

Applications?

This is the end of your bioinformatics lab. You have explored a tiny bit of what Cytoscape can do, but as you have seen from the number of apps there are many possible ways you can use Cytoscape to interpret your data. Answer the following questions from your brief usage of Cytoscape as a tool.

Q.10) What benefit do you think a tool like Cytoscape has for looking at datasets such as yours?

Q.11) Do you feel like Cytoscape helped you gain some insight into your data? Would you like to use more tools like this to analyze your data?

In your final lab submission, make sure you include BOTH parts of your lab: One knitted R Markdown with part 1 questions answered, and a file containing figures and answered questions from part 2.