Protocol for Protein Protein Interaction Network Scripts:

Created by Sam Altshuler

5/15/20

Files Needed:

1. R Scripts
   1. Interaction\_Networking.R
   2. Edit\_enrichment\_encore.R
   3. Legend\_Creator.R
   4. Network\_graphing.R
2. String Files (Can be found on String-DB.org: <https://string-db.org/cgi/download.pl?sessionId=2tGyg923YEJA&species_text=Neurospora+crassa>)
   1. These will be downloaded as zipped files, use 7zip or other unzipping software to extract the text files
   2. "5141.protein.links.full.v11.0.txt"
      1. For the interaction links
   3. "5141.protein.aliases.v11.0.txt"
      1. For mapping from String ID to other gene names
3. Background file:
   1. 5141\_background.csv
      1. Background file for *Neurospora crassa*, courtesy of Hannah de los Santos
   2. Copy\_5141\_MasterEntrezAliasing.txt
      1. Table with the master aliasing doc of the sting ID, NCU, entrez ID, and protein name for every *N. crassa* protein in STRINGdb
   3. .csv file with experimental data
      1. In two columns: one with the Uniprot IDs and one with the NCU numbers
      2. Make sure that the number of Uniprot IDs is the same as the number of NCU numbers
         1. Contaminations may result in mismatching lengths of columns
         2. SCRIPT WILL NOT RUN IF THE # OF NCU NUMBERS =/= # UNIPROT IDS
      3. Save it as a .csv, not an excel file

Procedure:

1. Make sure that all the files above are properly stored in the same folder
2. Make sure that the files with the experimental data (Uniprot ID and NCU numbers) are correct
   1. The script will remove contaminants, but the list of non-contaminants needs to be the same length as the list of NCU numbers
3. Open up Interactions\_Networking.R
   1. In line 52: replace the placeholder name with the name of the file with experimental data:
      1. Example: “placeholder.csv” 🡪 “DD8.csv”
   2. In line 86: read in the file from String that has all of the links
      1. “5141.protein.links.full.v11.0.txt”
   3. In line 90: read in the master aliasing doc
      1. “Copy\_5141\_MasterEntrezAliasing.txt”
   4. In line 348: replace the placeholder file name with the chosen name for the data frame with the interaction data to be stored in
      1. Once again, stored as a csv, but can be opened in Excel to few and manually change the information as needed
   5. Once all of the file names are changed, save the script
   6. Hit control+shift+enter to run all of the code
      1. Or click “Source” in the top right corner of the script window
      2. The script will take 20-30 minutes to run
      3. When done, the final line in the console should say “Protein Protein Interaction Network Saved”
4. At this point, you can manually curate the data in the csv file outputted by the script
   1. Fix missing NCU or gene names due to outdated Uniprot IDs in the aliasing file
   2. Look up string IDs that did not have Uniprot IDs
   3. Fixing these NA values will ensure for a better overall picture, but is not necessary to run the following scripts
5. Open up the “edit\_enrichment\_encore.R” script
   1. Make sure that the background file “5141\_background.csv” is saved in the same folder as the script
   2. On line 273: change the placeholder name to the output file from the previous script
   3. On line 676: change the placeholder name to the chosen name for the output data from this script
      1. Make sure it stays as a .RData file to preserve the formatting in the interactions data frame
   4. On line 678: change the placeholder name to the chosen name for the output .csv from this script
      1. This is the final analytical output, the .Rdata gets transformed more to make the figures
   5. Hit control+shift+enter to run all of the code
      1. Or click “Source” in the top right corner of the script window
6. Open up the “Legend\_Creator.R” script
   1. Make sure that all datasets being compared have been run through the first two scripts
   2. On line 127: change the placeholder name to the output file from the previous script
      1. Repeat this every 6 lines for the amount of data sets being compared, comment out the sections not being used
   3. On line 170-175: comment out the variables not being used by a dataset
   4. On line 196-226: comment out the variables not being used
   5. On line 236: change the placeholder name to the chosen name for the legend data from this script
   6. On line 240-255: comment out the lines for variables not used, change placeholder file to the chosen name for the output data from this script
   7. Hit control+shift+enter to run all of the code
      1. Or click “Source” in the top right corner of the script window
7. Open up the “Network\_Graphing.R” script:
   1. In line 13: change the placeholder name to the name of the output data from the edit\_enrichment\_encore script
   2. In line 14: change the placeholder name to the name of the legend data .Rdata file saved in the previous script
   3. In line 124, 159, and 194: choose the name for the graph/network figure
      1. It will be saved as a pdf
      2. If you are tinkering with the sizing of the nodes, make sure the pdf is closed before rerunning the script or else it will not run
      3. The pdf will also be converted to a .tiff at the end
         1. THESE FILES TAKE UP A LOT OF SPACE, PUT IN ZIPPED FOLDER FOR EMAILING OR UPLOADING TO SERVER
   4. Hit control+shift+enter to run all of the code
      1. Or click “Source” in the top right corner of the script window
   5. Once the script is done running
      1. The pdf of the network figure will be in the same folder as the rest of the files
   6. If the network does not look ideal, here are a few areas that you can change (in order of priority):
      1. Line 143: Vertex.size: changes the size of the vertices, it is defaulted to 30, but if the vertices are overlapping, change it to a lower value (for a group of ~40 vertices, vertex.size = 15 works). This will be the most effective way to prevent overlap and make the figure look better.
         1. Line 145: Vertex.label.cex: changes the label size on the vertices. It is defaulted to 1 and vertex.label.cex = 0.5 works for vertex.size = 15. However this is dependent on the length of the label, smaller label lengths can allow for larger sizing
         2. Make sure to change it in all three pdf repetitions
      2. Line 132: repulse.rad: change the repulsion radius: this is defaulted to the number of vertices in the graph cubed, but a slightly lower value may allow for better spacing.
      3. Line 131 area: the area can be changed, the bigger the area, the more vertices it can hold, however changing the vertex size is more effective than this