Proteinarium GUI User Guide

1. Getting Started

- Create the Gene Set File 1. This must be a text file with one line per sample. Each line has the format <Sample Identifier> = <HGNC Symbol 1>, <HGNC Symbol 2>, ..., <HGNC Symbol n> for a sample with n seed genes.

- (Optional) If you wish to analyze a dichotomous phenotype, a second gene set file can be created for Gene Set 2. This file must be formatted the same way as done for Gene Set 1.

2. Running Proteinarium

Selecting Files

- In the GUI, select File > New analysis

- Choose the data files for upload. If you are only analyzing one group or a single sample, leave the Geneset File 2 file field empty. NOTE: If you would like to select new Gene Set Files, click on the "Clear Gene Files" button before selecting new Gene Set Files.

- Enter a Project Name into the field "Project Name". If left blank, the program default to the project name "SIM"

Setting Parameters – Configuration File

- Before running Proteinarium, confirm or make changes to the program's configurable parameters. A table with all possible parameters, their definitions and their default settings are provided by selecting Help > Available Configurations.

- Select File > Settings or "Settings" button to view and change parameters. If any changes are made click "Apply Changes" before exiting. *NOTE: The values of the parameters for a New Analysis are not all set to the Default values. These are the values used in our testing and validation with bootstrapping turned off (ie 0 iterations).*

- For Advance Settings Select File > Settings and click on the "Advance Settings" button. Once changed click on "Set Settings" and the "Apply Changes" buttons.

Running Proteinarium

- Once files are uploaded and parameters are set, click "Run Proteinarium"

- If more than one sample is contained within the input file (s), the dendrogam will be displayed. All output files generated by Proteinarium will be saved to the folder specified by the *outputDirectory* configuration option. By default, this goes to a folder called "Output" in the same directory as the Proteinarium.jar file.

    - **<projectName>_ClusterAnalyses.csv**: cluster analysis files
    - **<projectName>_Dendrogram.png**: dendrogram image
    - **<projectName>_Dendrogram.txt**: representation of the dendrogram in Newick tree format

3. Viewing Clusters

- When Proteinarium is done running, the GUI will navigate to a new window in which you can input the Cluster/Sample ID for which you are interesting in viewing or obtaining more information.

- Enter sample ID or branch number (ex: *C12*) in the space provided.

- Select either "View Cluster" or "Get Cluster Information".

- If "View Cluster" is selected, the corresponding the output files will be available in the \<outputDirectory\>/<cluster or sample ID\> folder. And the following files are generated:
    - **<cluster or sample ID>_Dendrogram.txt**
    - For each of the five possible output networks--Group 1, Group 2, [Group 1 + Group 2], [Group 1 - Group 2], [Group 2 - Group 1], three files are generated to summarize that network. For example:
        1. **<cluster or sample ID>_Group1_GeneSet.txt**: list of genes in the network and information about which input set they originated from (i.e. from Group 1, Group 2, or imputed from the interactome) and on how many samples the gene was found
        2. **<cluster or sample ID>_Group1_Interactions.txt**: network interaction matrix
        3. **<cluster or sample ID>_Group1.png**: image of the network

- If "Get Cluster Information" is selected, the following information will be appended to the file "SystemsOutput.txt" as indicated by the text box of the GUI.
    - *Average Distance (Height)*
    - *Bootstrapping Confidence*
    - *Total Number of Samples*
    - *Number in Group 1* (number of samples)
    - *Number in Group 2* (number of samples)
    - *p-value (Fisher Exact test for Group 1 and Group 2)*
    - *Group 1 and Group 2 Clustering Coefficient*

- *Group 1 Clustering Coefficient*
- *Group 2 Clustering Coefficient*
- *Group 1 minus Group 2 Clustering Coefficient*
- *Group 2 minus Group 1 Clustering Coefficient*
- *Group 1 Patients* (Sample IDs of the individuals)
- *Group 2 Patients* (Sample IDs of the individuals)

- Output files can be opened from the GUI: select File > Open to open a file explorer window.

4. Change the Parameter Configuration and Re-run Proteinarium:

- Select File > Home to return to main screen for running Proteinarium
- Select File > Settings
- Change desired parameter values
- Click Apply Changes
- Click “Run Proteinarium” to re-run Proteinarium with new configuration. NOTE: Change Project Name otherwise the previous data will be overwritten.