PHONEMeS How To

This document contains information about how to run a PHONEMeS analysis, based on the scripts and files in the PHONEMeS_example folder.

This is based on PHONEMeS 0.2.3, run locally using R 3.1.0 (in RStudio) and on a cluster using R 2.15.2 and LSF 8. Running these scripts also require the igraph package (version 0.7.1) and the BioNet package (version 1.24.0). The networks are visualized using Cytoscape 2.8.0 although we are working on a plugin for Cytoscape 3 that automatically imports the contents of a results folder and produces the annotated networks.

1. Prepare the data
Run prepOptim.R and produce data4cluster_n.RData
prepOptim.R goes from a network object and a data object (resulting from data normalization, summarization and Gaussian mixture modeling) to the objects needed to run PHONEMeS on a cluster.

2. Move to a cluster
Copy on cluster: data4cluster_n.Rdata, processGx_n.R, runScriptGx_n.sh, scriptGxopt_50models_n.R, import_n.R

3. Install the package and make a results directory
Make sure PHONEMeS is installed on R on your cluster, and you have, in the directory where you will run the scripts, a folder called Results_n (that is where the functions will put the results of each generation).

In R: *install.packages("PHONEMeS_0.2.2.tar.gz",repos=NULL)*
In Terminal: *mkdir Results_17*

4. Run one independent optimization (repeat multiple times with different results folder and *n* indices)
Run runScriptGx_n.sh

./runScriptGx_17.sh

5. Copy the results locally
Copy back locally the resulting file (pn_imported.RData) as well as the diagnostic plots in optim_n.pdf

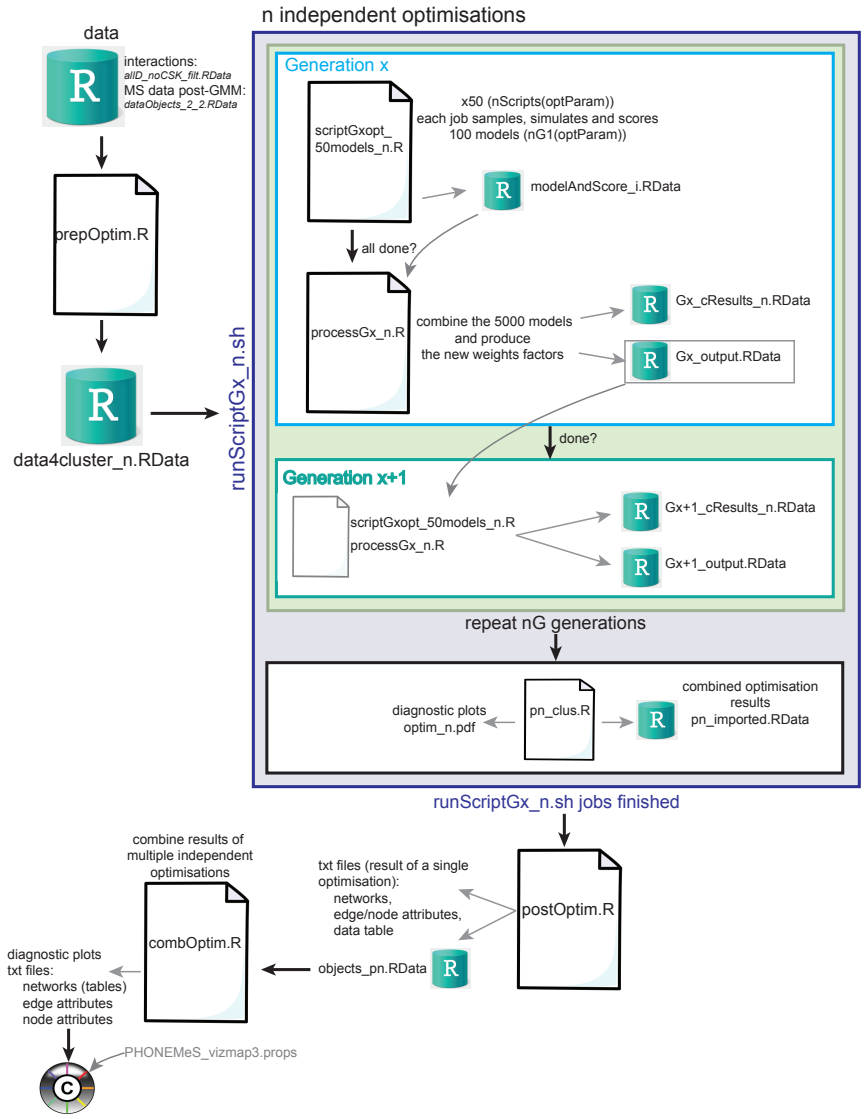
6. Process each independent optimisation
Run postOptim.R
This processes the results of a single optimisation (stop after this step if you are not running multiple independent optimisations). Repeat steps 3-6 with a different Results folder and different n indices, for multiple independent optimizations.
N.B. This file needs a few elements to be changed depending on the optimisation (such as the location of files etc).

7. Combine multiple independent optimisations
Once postOptim.R has been run on all independent optimisations and you have a series of objects_pn.RData files, run comb_optim.R.

This produces the combined plots, as well as the final resulting networks (max input averaged frequencies across independent optimisations, max scoring paths by averaged frequencies, etc).

8. Visualise resulting networks

The resulting txt files (either individual ones from step 6 or combined ones from step7) can be imported (as tables) into cytoscape (using the vizmap property file PHONEMeS_vizmap3.props). The main network files are the maximum input based on frequency, and the columns to be read as source and targets are *K.ID* and *S.cc*, respectively. The column *SID* should be read in as an edge attribute as it is used for mapping. The edge weight is fixed by the edge attribute *_EA* file, column *f50*. The edge mapping is done using the *SID* column. The node attribute used to color the nodes is produced by step 7, and it is called "AllNodes_nodesP_NA_pn.txt" (it is identical for all independent optimisations). The data table (used to get the cluster/status/FC information) is imported as a node attribute, it is called "AllNodes_DA.txt".



Note: in the script *prepOptim.R* that is run locally, you need to change the parameter *resN* (match the *n* in the workflow above), as well as the drugs, treatments, and possibly optimization parameters that you would like to change. In the scripts that run on a cluster (*processGx_n.R*, *import_n.R*, *scriptGxopt_50models_n.*R) you only need to change the index in the load("*data4cluster_n.R*") string. In the masterscript *runScriptGx_n.sh* you need to change the *genM=pngen$j* and *genRes=pnres$j* strings, the *n* index in the names of the R scripts called, and the *n* index in the output file *pnout*.