PHENOMESCAPE v1.0

PhenomeScape is a Cytoscape app to run the PhenomeExpress algorithm in the Cytoscape environment [1].

Requirements:
PhenomeScape requires Cytoscape v3.3 or above

A detailed manual on the use of Cytoscape can be found :
http://manual.cytoscape.org/en/stable/index.html

FAQ

What is PhenomeExpress?

PhenomeExpress is a method for the analysis of disease-related gene expression data. It produces *de novo* pathways/sub-networks enriched in differentially expressed genes and that are likely to be related to the disease under study.

What is the motivation behind using network based tools?

Large lists of differentially expressed genes are hard to interpret. Often pathway based methods using KEGG or Reactome pathways are used, but these are limited to very well studied canonical pathways that have arbitrary defined boundaries. Network based methods make use of protein-protein interaction data to find pathways using the expression data itself.

What makes PhenomeExpress different?

It makes use of phenotype to gene associations. Many genes have been perturbed in mouse and zebrafish and the corresponding phenotype recorded. Often this information will be manually used to aid understanding of lists of differentially expressed genes. It makes sense to use this information to find the pathways and highlight these genes of particular interest.

How does it work?

PhenomeExpress works in two steps. First it ranks proteins in the protein-protein interaction network on the basis of the gene expression data and also the chosen phenotypes. The aim is that proteins with relevant phenotype associations that are differentially expressed and that surrounded by other such proteins will be very highly ranked. PhenomeExpress uses the chosen “seed” phenotypes as the starting point to find other closely related phenotypes and also uses their known associations to aid the analysis.

The second step is connect the highly ranked proteins together to form multiple sub-networks that if both differentially regulated and related to the chosen phenotypes. All the networks are tested to assess whether you would expect to see the total level of differential expression in a sub-network of that size by chance alone.

What do I need to use PhenomeExpress?

It takes as input:
A table of expression data i.e gene names, fold changes and p-values.
A protein-protein interaction network - HumanConsensusPathDB and String Networks are provided
UberPheno phenotype terms that describe the phenotypes of the disease.

Using PhenomeScape

Below gives step-by-step instructions to run PhenomeExpress on the example expression data provided in github repository. Perform the steps within the black boxes to follow along with the analysis of the expression data.

TUTORIAL

Download Example Expression data

Browse to and download the expression text file:
https://raw.githubusercontent.com/soulj/PhenomeScape/master/ExampleData/Human_KneeOsteoarthritis.txt

Save this file as "Human_KneeOsteoarthritis.txt".

Overview of Steps:

1. Load Network
2. Add expression data to network
    2a. Import expression data from file
    2b. Merge network and expression data by gene symbol column
3. Choose phenotypes
4. Select Table Columns
5. Set network parameters
6. Starting the Analysis
7. Visualise Results
8. Export the results

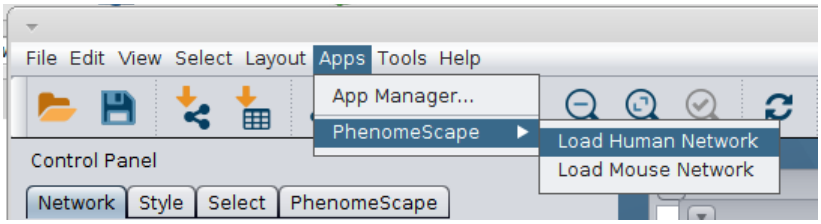
Step 1: Load a network

PhenomeExpress requires a protein-protein interaction network where the names are either human or mouse official gene symbols. Example networks from HumanConsensusDB and StringDB are provided for convenience.

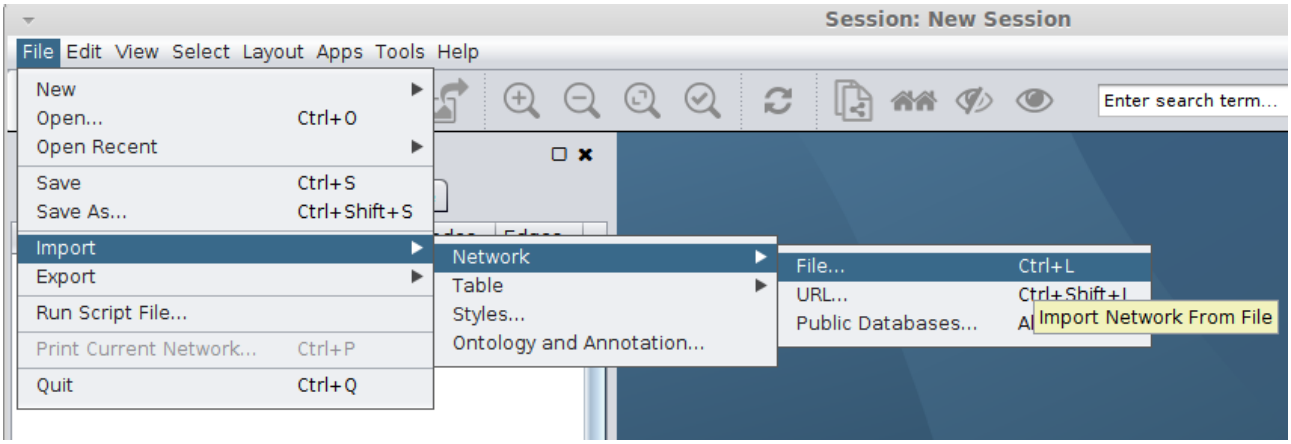
To load the provided mouse network for example:
Apps->PhenomeScape->Load Mouse Network

The example expression data is from human cartilage so we load the human protein-protein interaction network:

Apps->PhenomeScape->Load Human Network



A custom network can be easily imported by:
File->Import->Network->Select the file



Edges may be weighted by a confidence score representing the probability that the observed edge is true. If provided then this will be used in the calculation of the activity scores. This edge attribute must to be in a column named “Confidence”.

Step 2: Add expression data

2a. Import expression data from file

Gene expression data should be analysed to give fold changes and (adjusted) p-values. The expression data should contain gene symbols that match nodes in the loaded network.

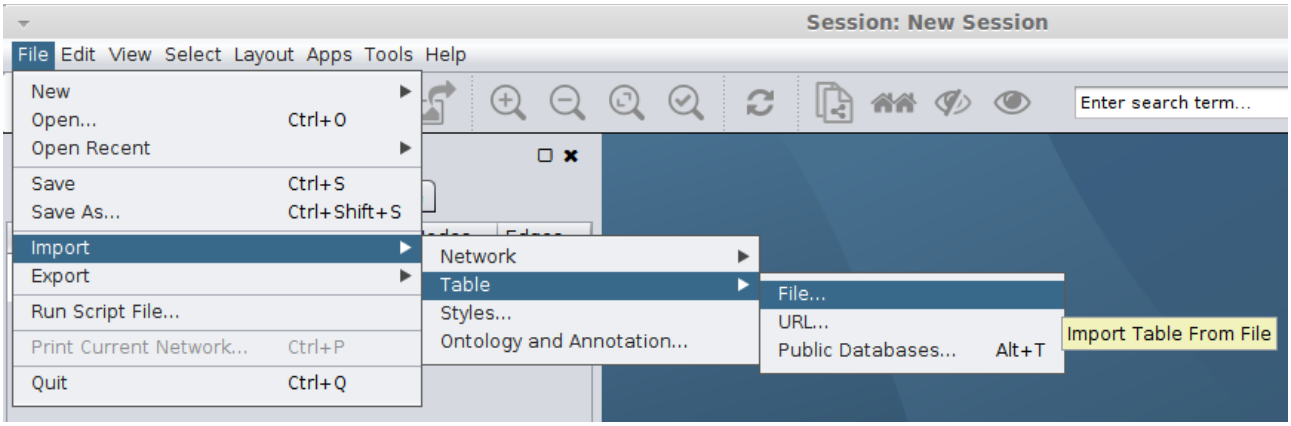
The fold change should be in log2 units.

A expression score for each node in the network will be calculated as:

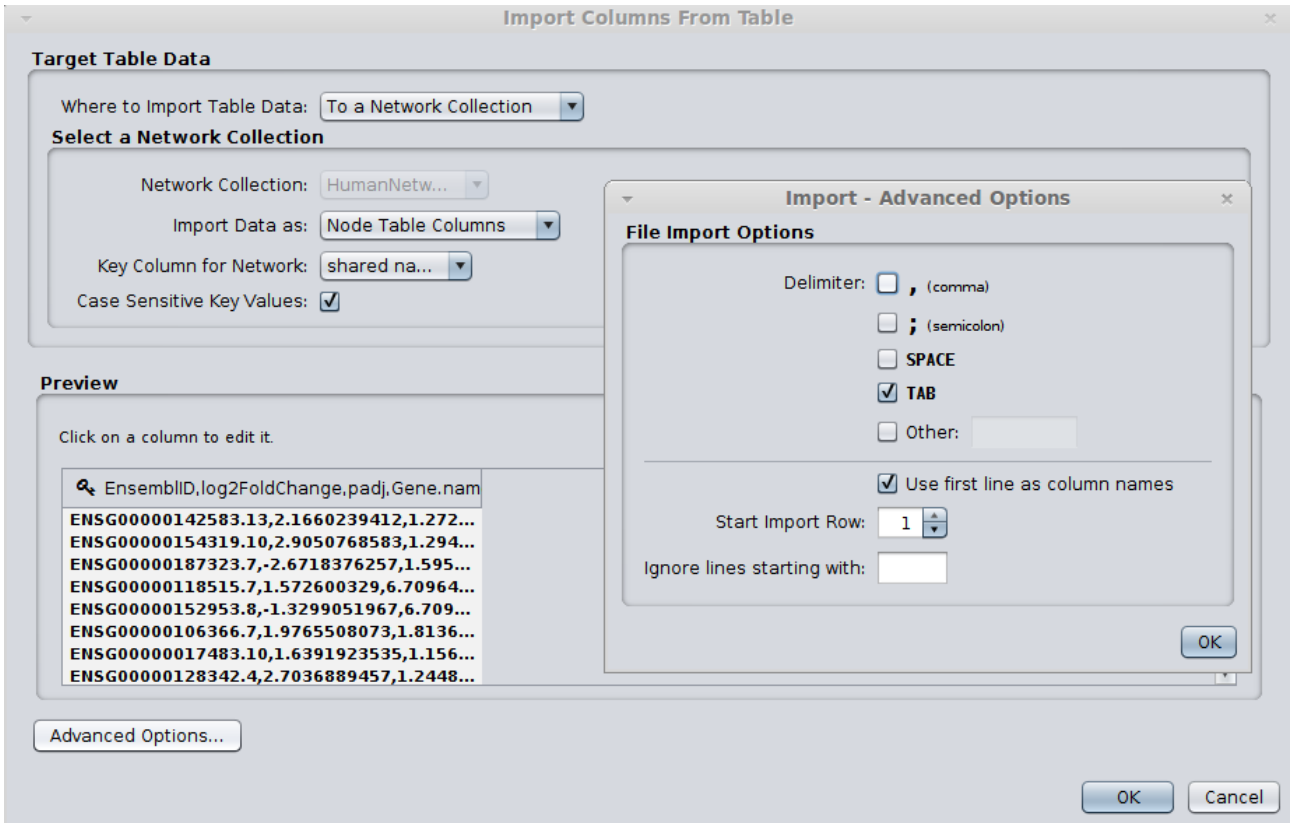
log2foldchange * -log10(pvalue)

To load the expression data into Cytoscape:

File->Import->Table ->File
Select the "Human_KneeOsteoarthritis.txt" file from your computer and click Open.



Select the required expression data file from your computer and and choose the network collection matching the loaded network.



The file is comma separated so click advanced options in the bottom left of the panel, unclick the tab check box, click the comma check box and click OK.

Choose the network collection to match the name of the loaded human network "HumanNetwork".

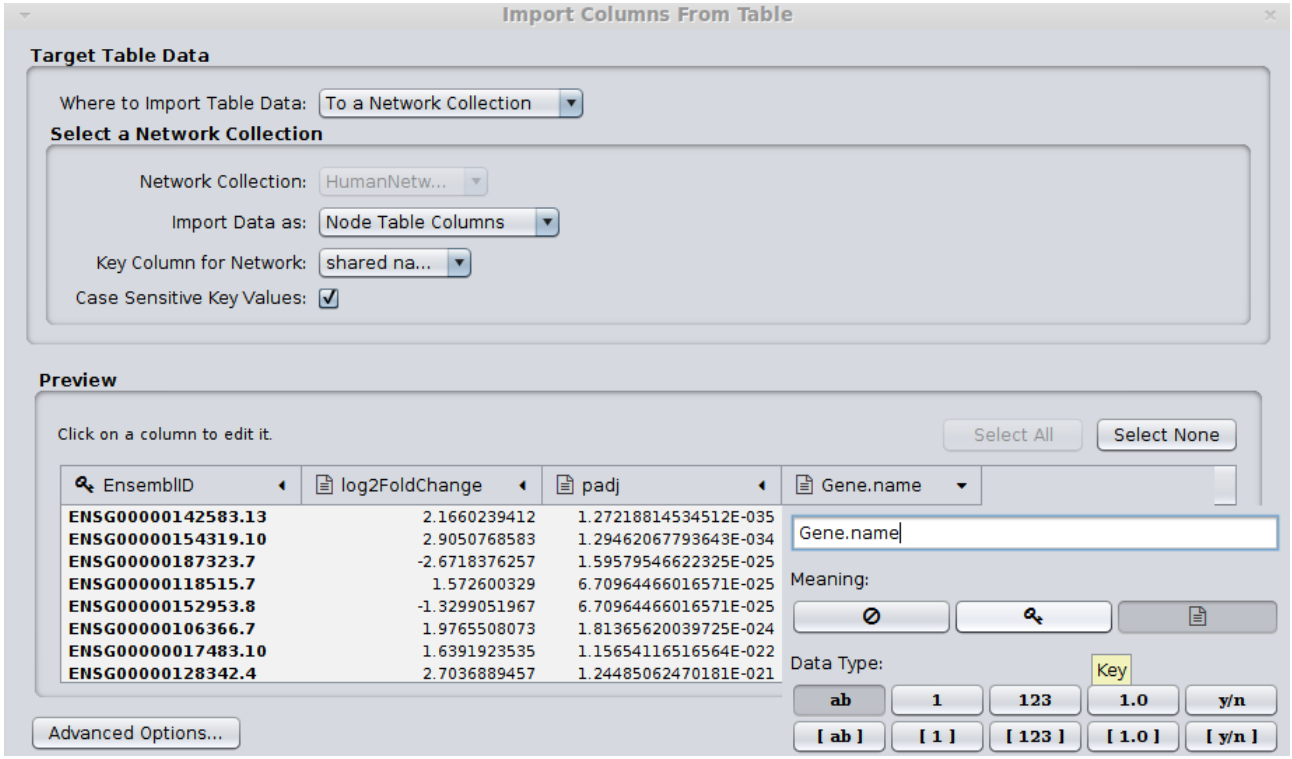
Don't click OK just yet!

2b. Merge network and expression data by gene symbol column

In the example figure below the key symbol is next to the Ensembl Gene ID column in the table. This symbol indicates the column that will be used to match the nodes in the table to the nodes in the network.

It is necessary to change the column with the key symbol to the column in your table with the Human/Mouse Official Gene Symbols.

Click on the column name "Gene.name". In the pop up box, under Meaning, click the key symbol.

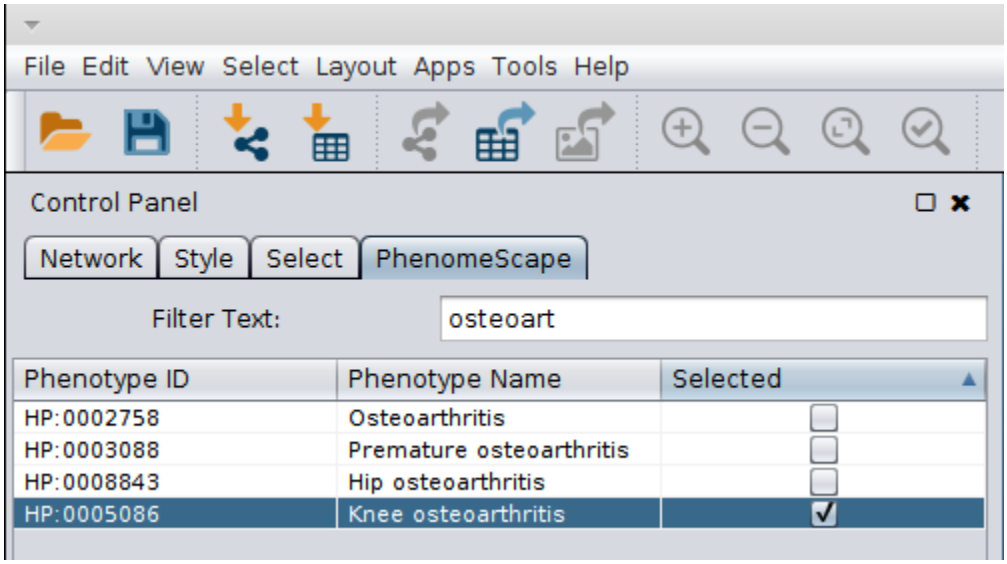


Note: not all the nodes in the network need expression data. Nodes without expression data will be removed to make a condition specific network. The largest connected component of the network will then be used to find sub-networks.

Step 3. Selecting Phenotypes

Phenotypes guide the algorithm to the most relevant parts of the network to the phenotypes of interest. The aim is to find groups of differentially expressed genes that when perturbed are likely to give rise to the observed phenotypes or related phenotypes.

Typing in the text box will allow you to search for phenotypes that describe the features of the disease under study. Click on the “selected” box to choose a phenotype. You must select at least one phenotype. Clicking on a column name will sort that column. To see all selected phenotypes sort by the selected column.



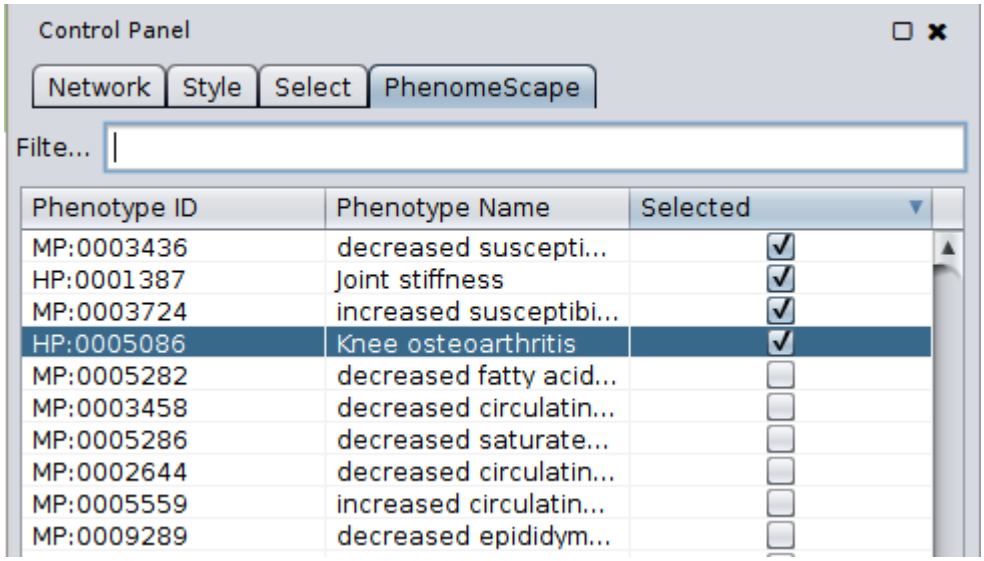
Choosing phenotypes:

Select most specific phenotypes relevant to your disease. The Phenotype-Phenotype network allows the algorithm to infer related phenotypes to ones selected.

For the example expression data we wish to select phenotypes related to osteoarthritis. The Phenotype-Phenotype network allows the algorithm to infer related phenotypes to ones selected.

In the search box type “knee osteoarthritis”, click the checkbox under the selected column
In the search box type “joint stiffness”, click the checkbox under the selected column
In the search box type “increased susceptibility of induced arthritis”, click the checkbox under the selected column
In the search box type “decreased susceptibility of induced arthritis”, click the checkbox under the selected column

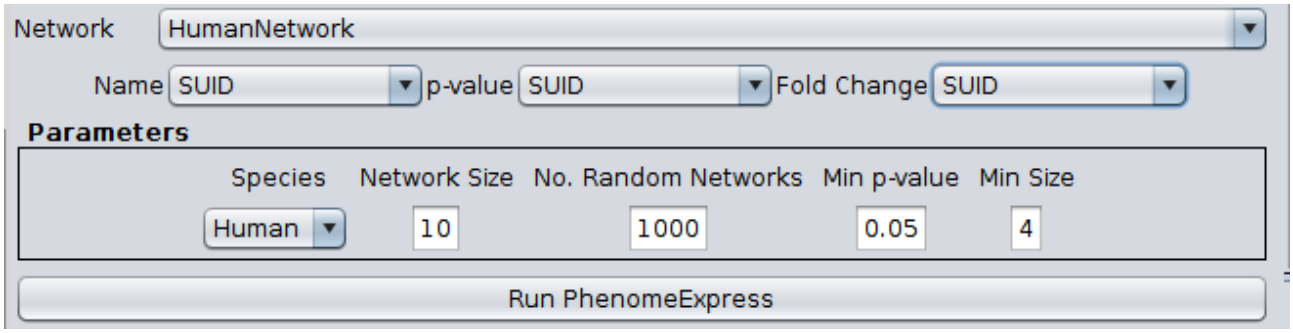
Delete all text in the search box and click on the selected column name to sort by that column. You should see the four chosen phenotypes with a tick in the selected box.



Step 4. Select Network and the Table Columns

Select the network to be used as input for PhenomeScape.

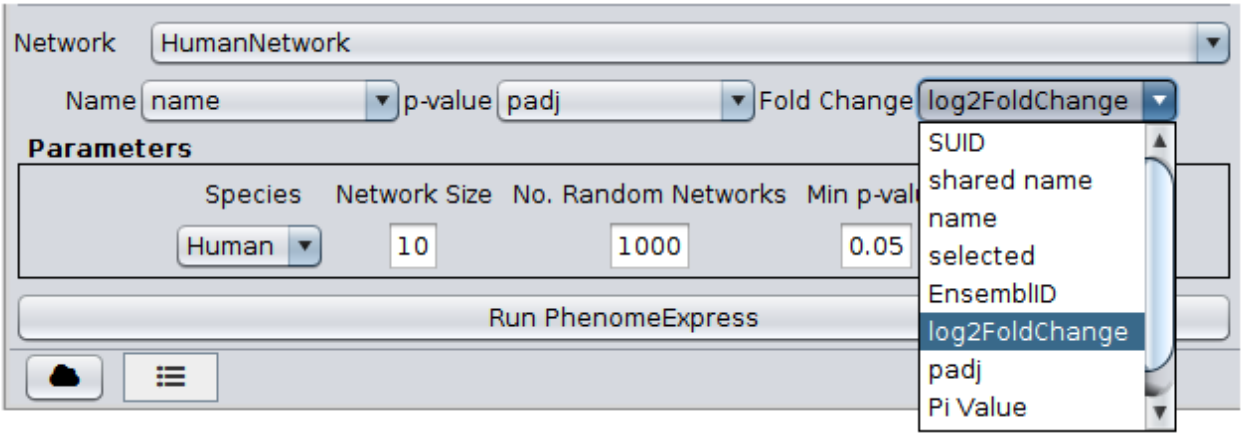
Select the “HumanNetwork” to be used as input



The available columns in the select network's node table will available as choices for the gene name, p-value and fold change boxes. Choose the appropriate columns for input.

Choose the corresponding columns in the network table for the gene name, p-value and fold change.

Under Name select the corresponding column "name"
Under p-value select the corresponding column "padj"
Under Fold Change select the corresponding column "log2FoldChange"



Step 5. Setting parameters

There are several parameters that control the final sub-networks produced.

Species - It is important to choose the correct species so the correct gene – phenotype associations can be used.

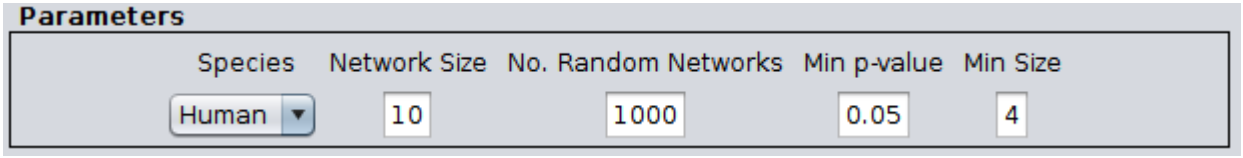
Network size parameter determines the maximum initial network size prior to identifying consensus sub-networks. A value of between 10-20 usually produces sub-networks that are small enough to be readily interpretable.

No. Random Networks - the sub-network significance is determined by random sub-sampling of the filtered protein interaction network. The maximum possible p-value is 1/ num permutations.

Min p-value - which p-value threshold should be used to filter the sub-networks?

Min size - what is the minimum final sub-network size allowed?

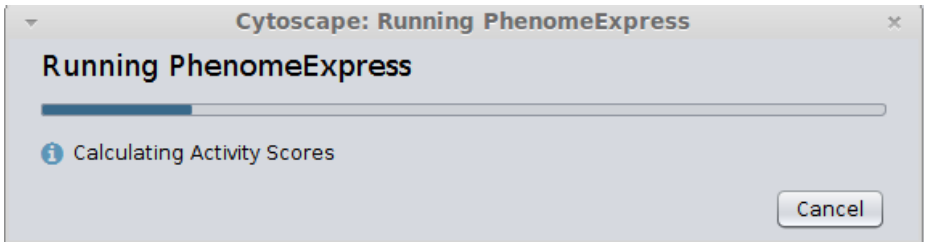
For the analysis of the example expression data will leave them as their defaults. Note that Human is set as the default species.



Step 6: Starting the Analysis

Click the "Run PhenomeExpress" button, to start PhenomeExpress.

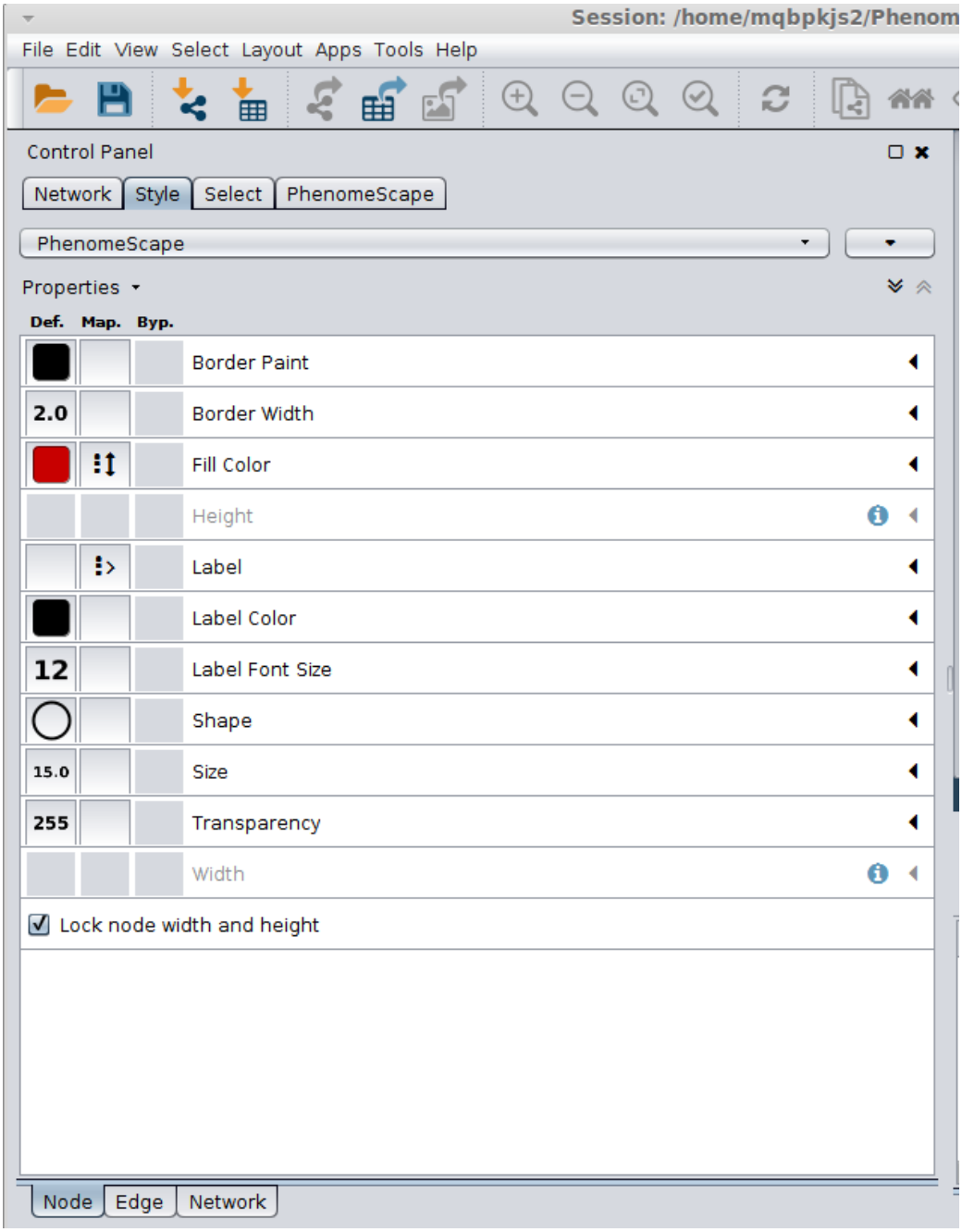
If the input looks acceptable you should see a loading bar to keep you up to date on the progress. The process should take under 5 mins.

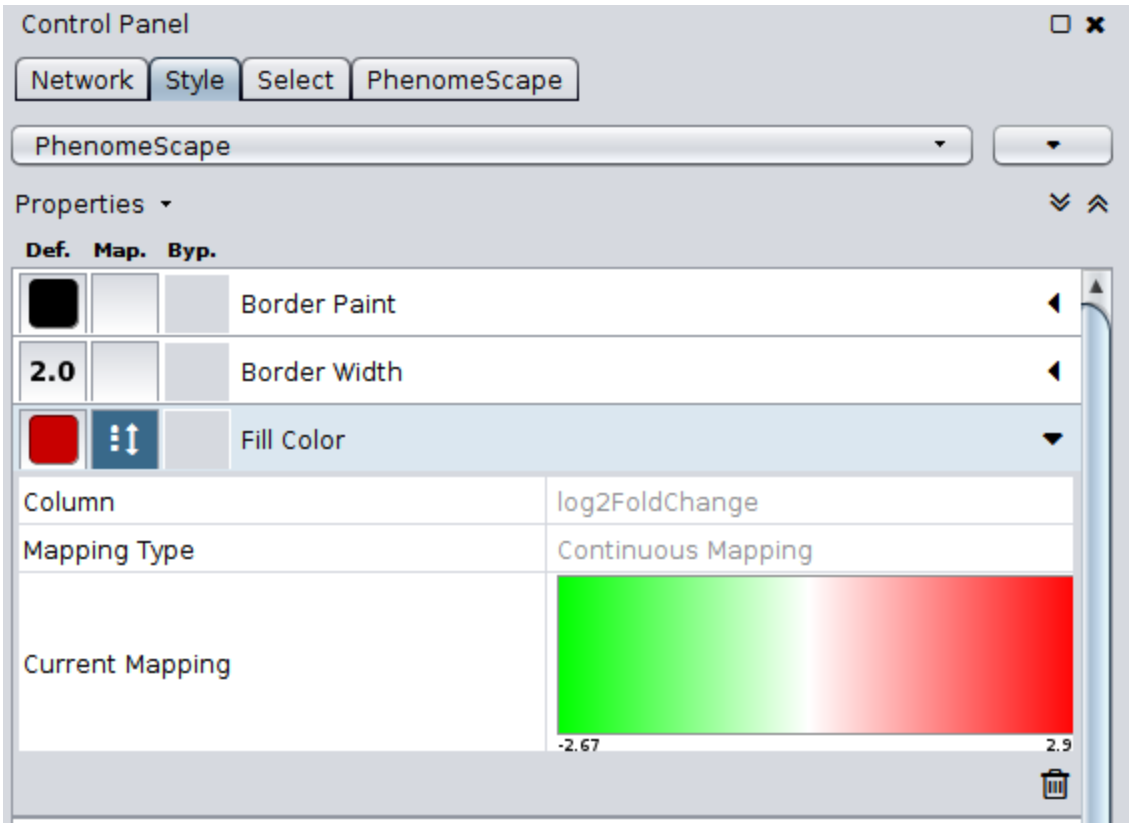


Step 7: Viewing the results

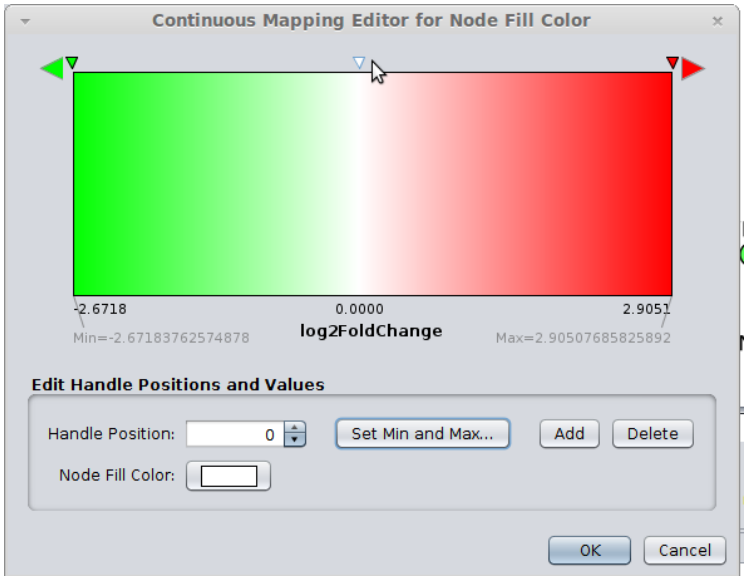
A visual style will be automatically created for the identified sub-networks. To set the correct colours for the nodes in the network.

Select the Style tab in the Control Panel. Click on Fill Colour arrow to show the gradient box.


Session: /home/mqbpkjs2/Phenom
File Edit View Select Layout Apps Tools Help
Control Panel
Network Style Select PhenomeScape
PhenomeScape
Properties
Def. Map. Byp.
Border Paint
2.0
Border Width
Fill Color
Height
Label
Label Color
12
Label Font Size
Shape
15.0
Size
255
Transparency
Width
Lock node width and height
Node Edge Network



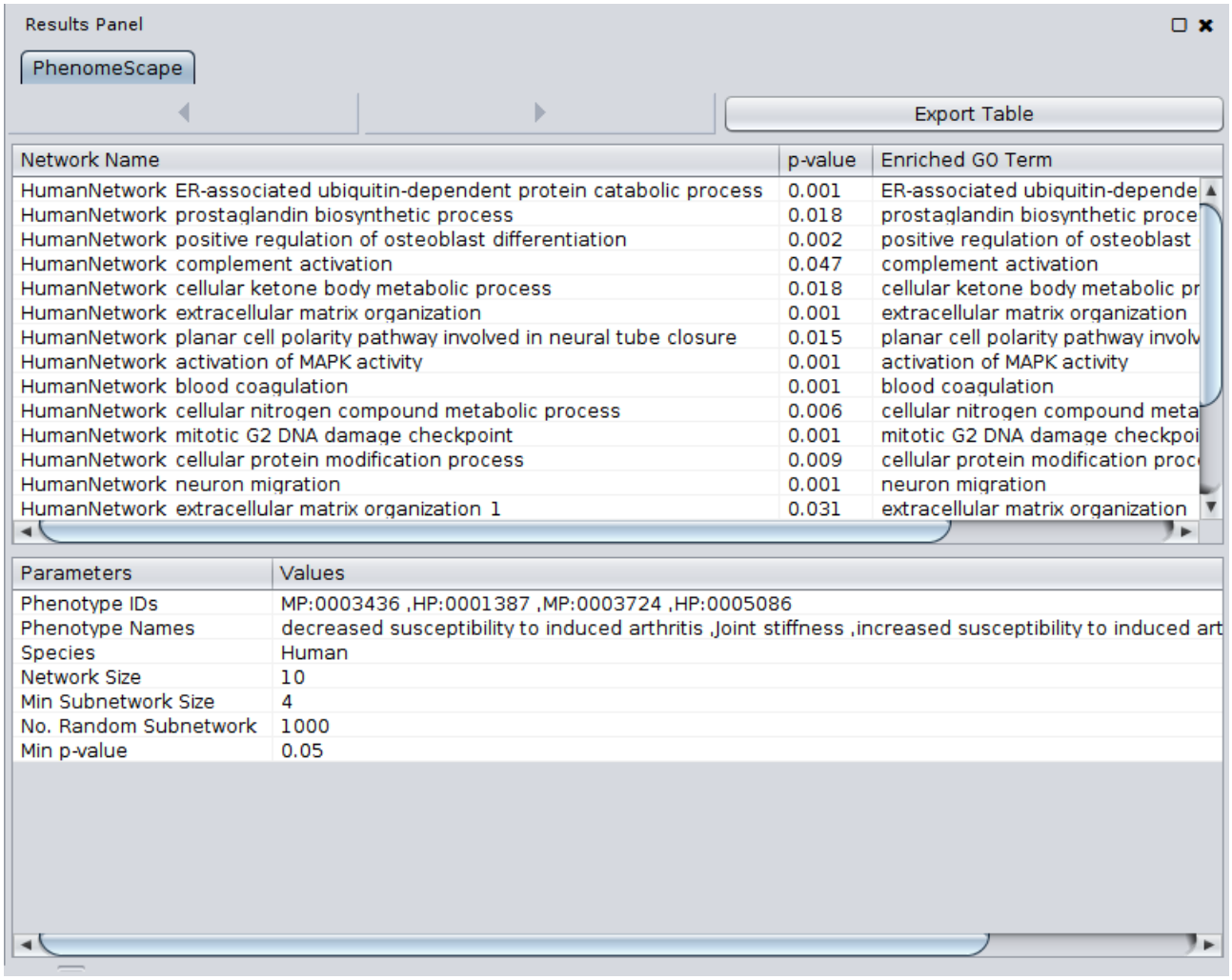
Double click on the Gradient box under Current Mapping to open the editor.



Click on the white triangle on the top the gradient box and click OK. This will set the colour of the sub-networks to the correct green-red gradient for negative to positive fold change.

The results panel will automatically open upon completion of the analysis. It gives an overview of the results – the empirical sub-network p-value, network size, the top enriched GO term are shown. The sub-networks are shown for visualisation and further analysis in Cytoscape i.e what transcription factor targets are significantly enriched in the subnetworks? If more than one run of PhenomeExpress has been performed, then the arrow buttons allow scrolling through the results.

By default the networks are laid out with the force-directed network layout and coloured by fold change. Phenotype nodes are shown in as blue rectangles. The blue edges represent the associations between the chosen phenotypes and the genes in the sub-network.



8. Export the results

The export button allows saving of the results table and parameter choices as a text file.

In the results panel, click on Export Table
Type “Osteoarthritis_Subnetworks” as the file name in the save box and choose the save location.
Click Save to save the results table.

Figure of the subnetworks can be created using standard cytoscape image export: http://manual.cytoscape.org/en/stable/Publish_Your_Data.html

References

1. Soul J, Hardingham TE, Boot-Handford RP, Schwartz J-M: **PhenomeExpress: a refined network analysis of expression datasets by inclusion of known disease phenotypes.** *Sci Rep* 2015, **5**:8117.