PRINCIPLE: PRINCE Plugin (v1.0)

User Guide

September 21, 2011

Introduction

This is a user guide for the PRINCIPLE Cytoscape plugin, implementing the PRINCE algorithm [1]. PRINCE is a method for prioritizing disease associated genes. Given a query disease, a disease-disease similarity measure and disease-gene associations, PRINCE provides a smooth score over the Human Protein-Protein Interactions (PPI) network.

Installation

Copy the PRINCE plug-in (PrincePlugin.jar) to the plug-ins directory under the Cytoscape installation directory, e.g:
Windows: C:\Program Files\Cytoscape_v2.8.2\plugins
Linux/mac: ~/Cytoscape/plugins

PRINCIPLE was verified on Cytoscape versions 2.7.0-2.8.2.

Usage

Running Cytoscape

Windows: Cytoscape.exe. Newer versions might require running the Cytoscape.bat file instead, adding the flag -Xmx512m to the java command line.

Linux/Mac - ~\cytoscape\cytoscape.sh

Important note: Due to space limitations of the plugins pane, the plugin might not be apparent on the initial loading of Cytoscape. Use the arrows near the plugin tabs or expand the pane to view the plugin. Using the menu Plugins->Manage plugins allows for verification that PRINCIPLE was indeed loaded (Fig. 1)
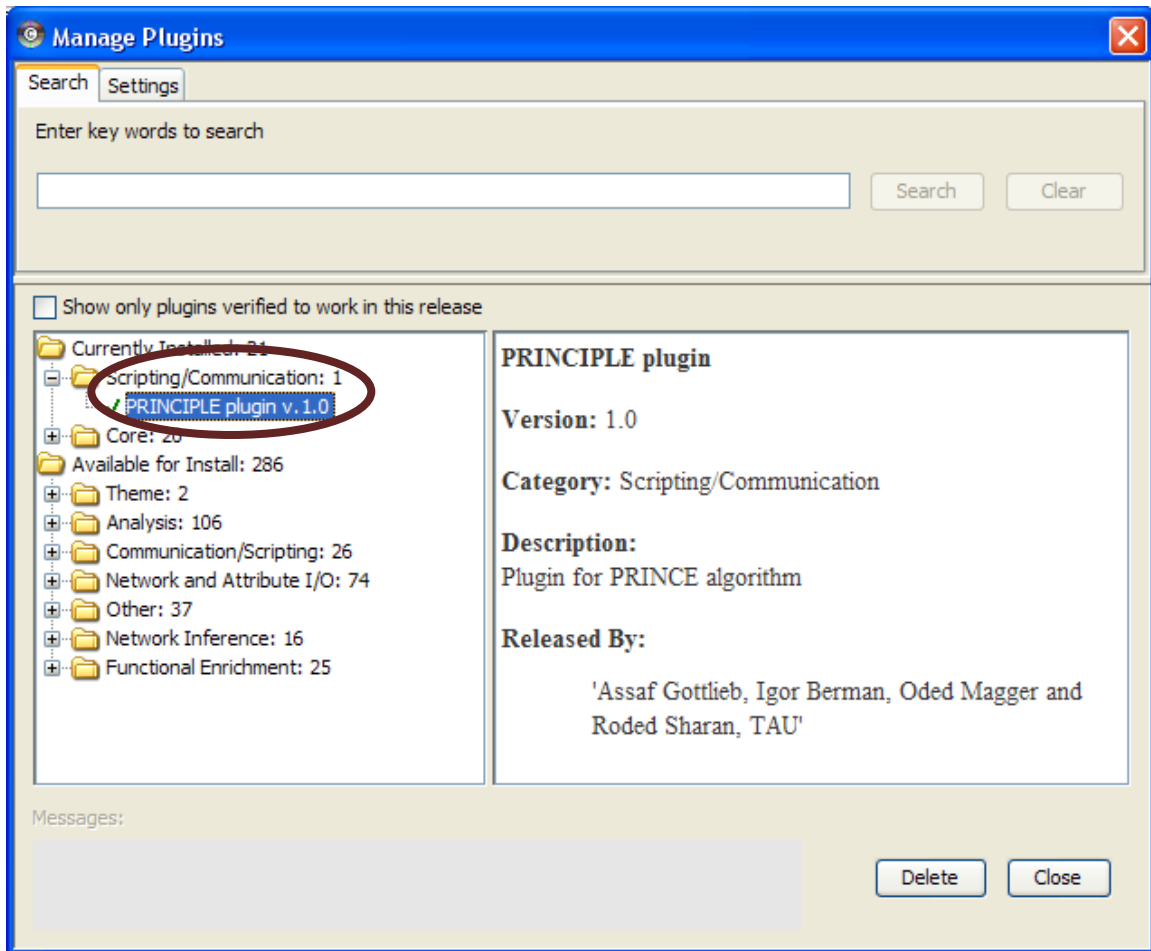
Figure 1. PRINCIPLE plugin viewable in plugin manager.

Sections

PRINCIPLE plug-in contains four sections (Fig. 2):

1. Configuring the input files (Fig. 2, section 1)
2. Configuring parameters (Fig. 2, section 2)
3. Specifying advanced options, currently supporting only an optional output file (Fig. 2, section 3)
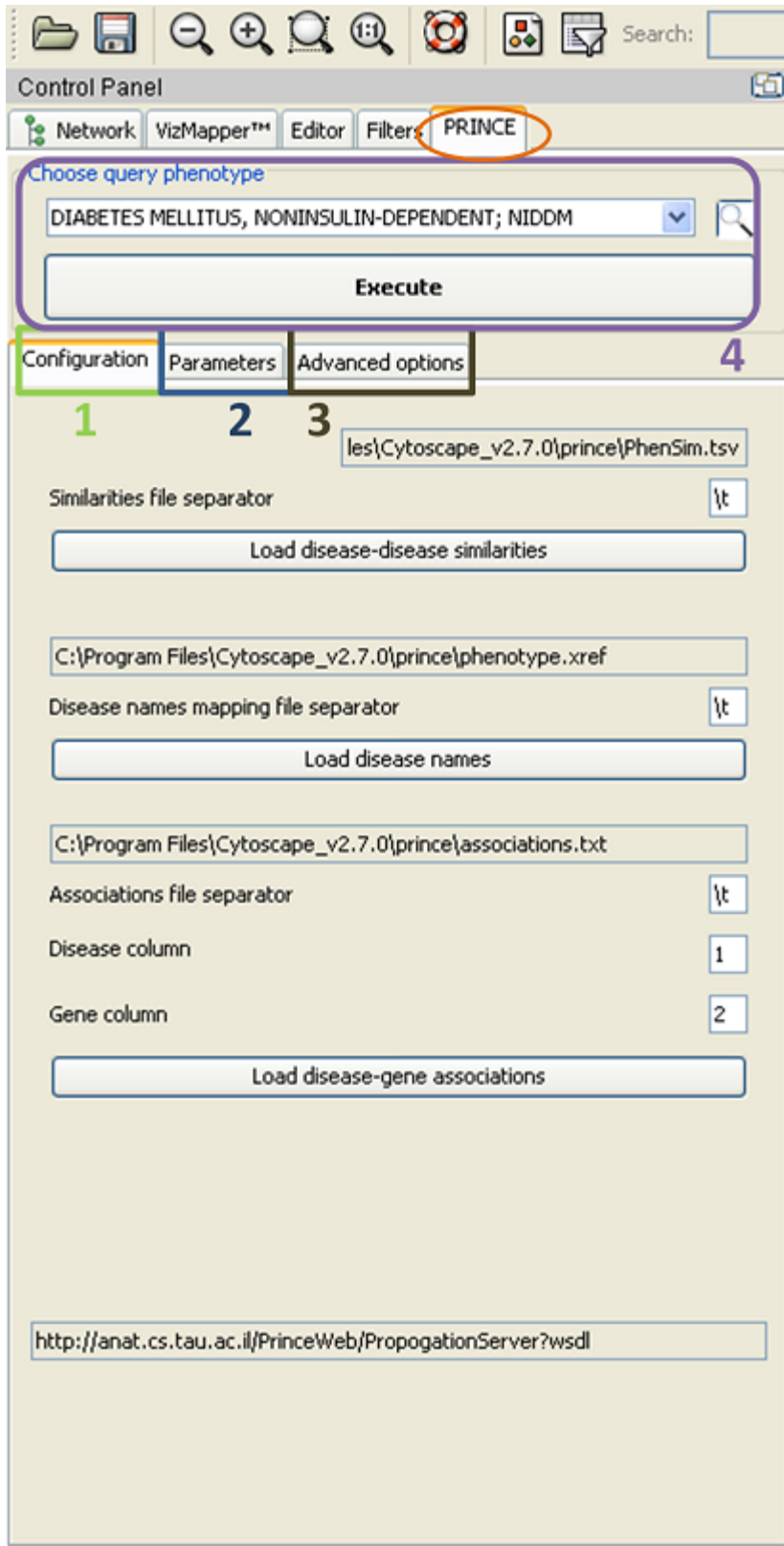4. Execution against a query disease (Fig. 2, section 4)

Figure 2. PRINCIPLE plugin sections.

1. Uploading input files

PRINCIPLE requires uploading three input files (Fig. 3):

1. Disease-disease similarity measure file, size Nx(N+1) (N- number of diseases). The first column is the MIM codes of the diseases. (Fig. 3, Rectangle 1).
2. Disease names and MIM codes, size Nx2 (Fig. 3, Rectangle 2) (Optional).
3. Disease-gene associations, size Mx2 (M<=N) (Fig. 3, Rectangle 3) (Optional).

The plugin loads by default a Disease names and MIM codes input file and Disease-gene associations input file, extracted from OMIM (#2 and #3 above).

The user can specify a different delimiter for each input file (default tab - \t). Use \s for specifying a space delimiter.

Additionally, the user can specify the columns where the disease MIM codes and the Entrez Gene ID reside in the disease-gene association file (default first and second, respectively). The disease-gene association file does not require all diseases to be represented.

The displayed MIM codes and diseases in the query box are the intersection of the diseases having similarity information in the disease-disease similarity measure input file and the MIM codes specified in the Disease names and MIM codes input file.
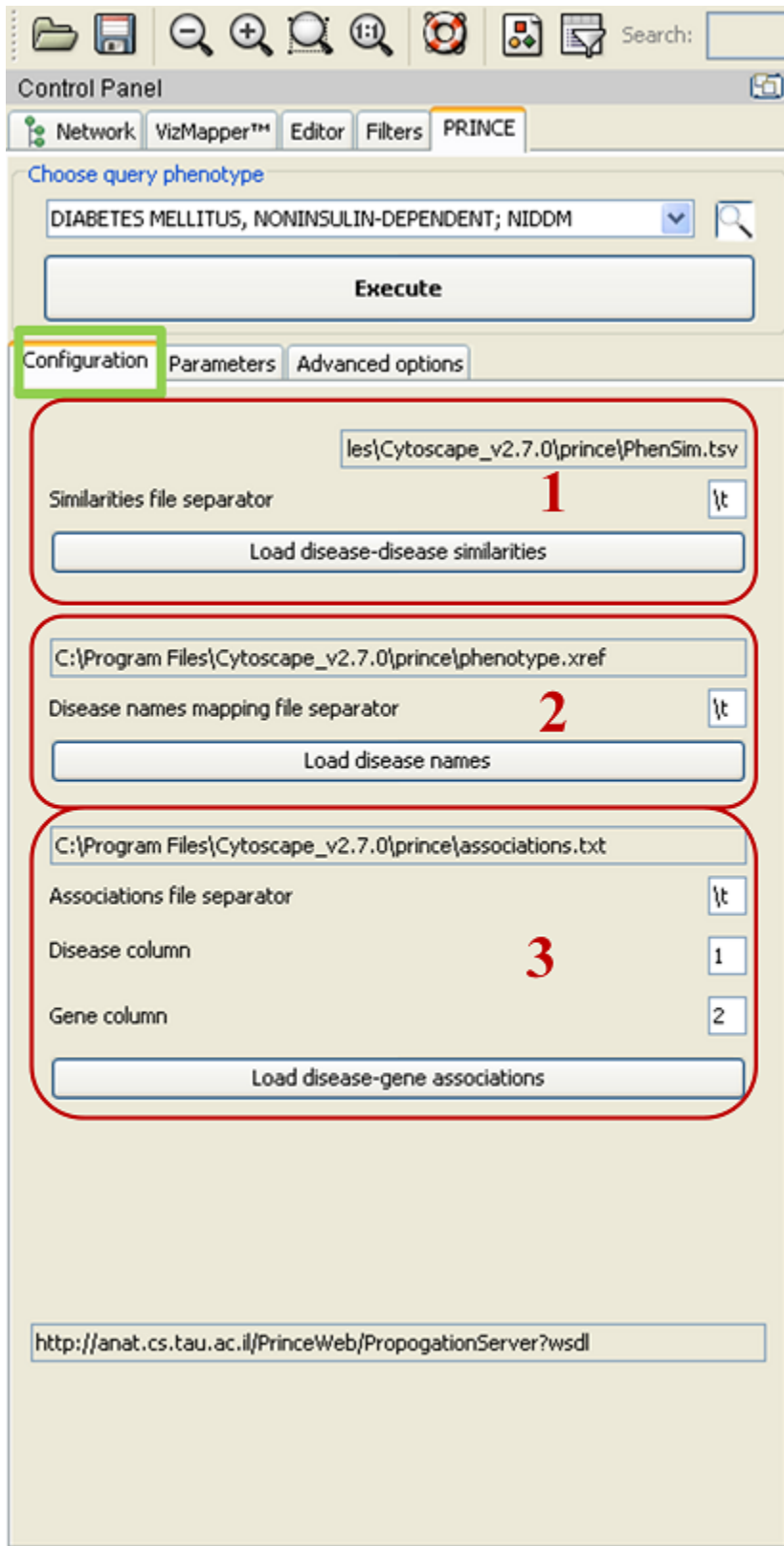
Figure 3. Configuring input files.

2. Selecting parameters

PRINCIPLE allows adjustment of three parameters (Fig. 4):

1. Alpha parameter, weighting the relative importance of the prior-based versus the network-based components of the score of each protein (default 0.9)(Fig. 4, Rectangle 1).

   The weighting follows the formula: $F(v) = \alpha \left[\sum_{u \in N(v)} F(u) w(v,u) \right] + (1-\alpha) Y(v)$

   Where $F(v)$ is the score of protein v, $w(u,v)$ is the normalized weights of the protein-protein interactions between proteins u and v and $Y(v)$ is the prior score of the protein.

2. The number of top ranked proteins to display (default 10) (Fig. 4, Rectangle 2). Their immediate interacting proteins are also displayed.

3. Number of iterations to perform, propagating the scores over the network. While more iterations provide more accurate results, the default 10 typically converges to the steady-state solution [1] (Fig. 4, Rectangle 3).
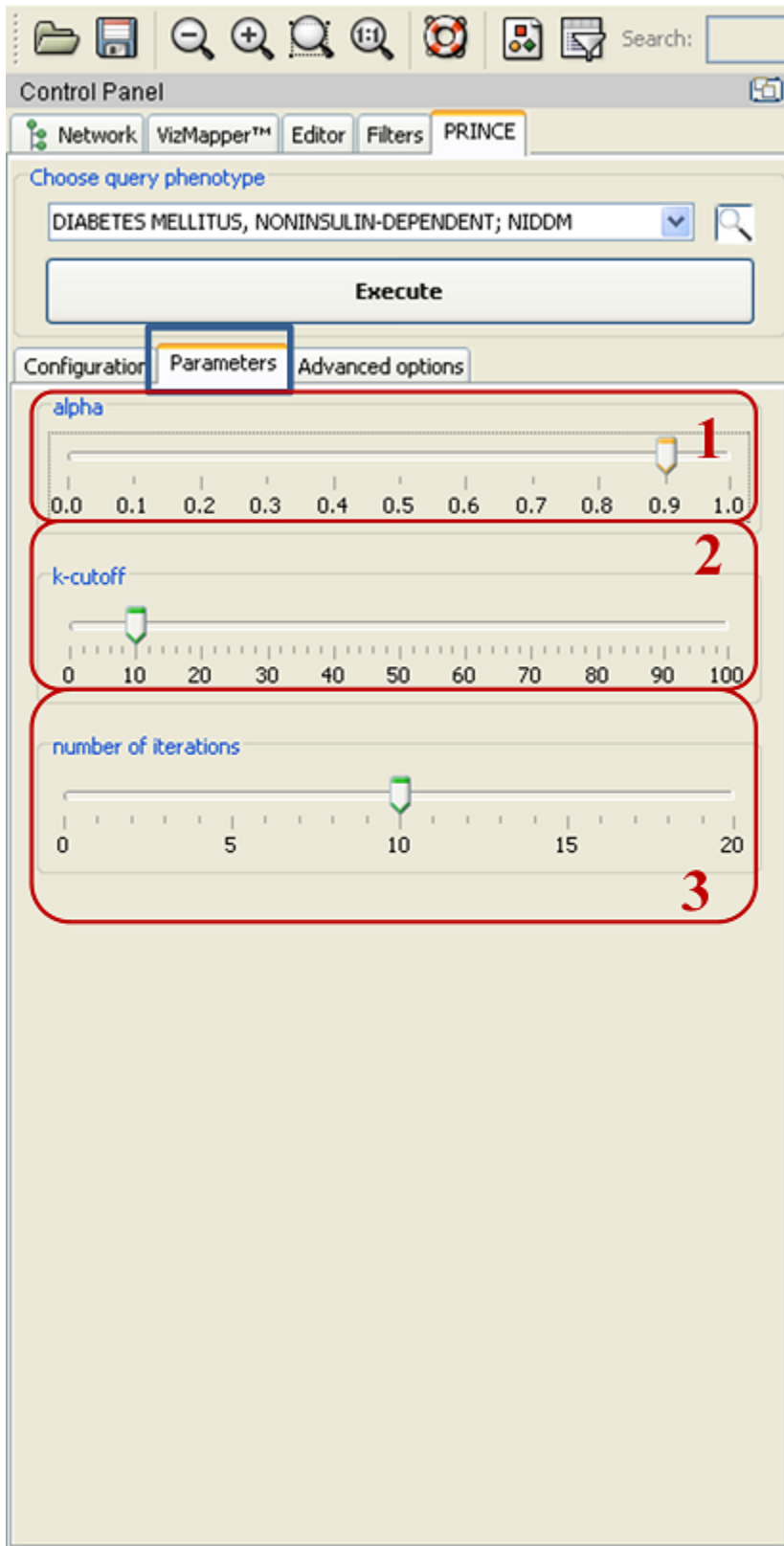
Figure 4. Configuring parameters.

3. Specifying an output file

PRINCIPLE allows specifying an output file (Fig. 5). This file lists the top k-cutoff ranked proteins, specifying their scores and the set of diseases associated with each protein.
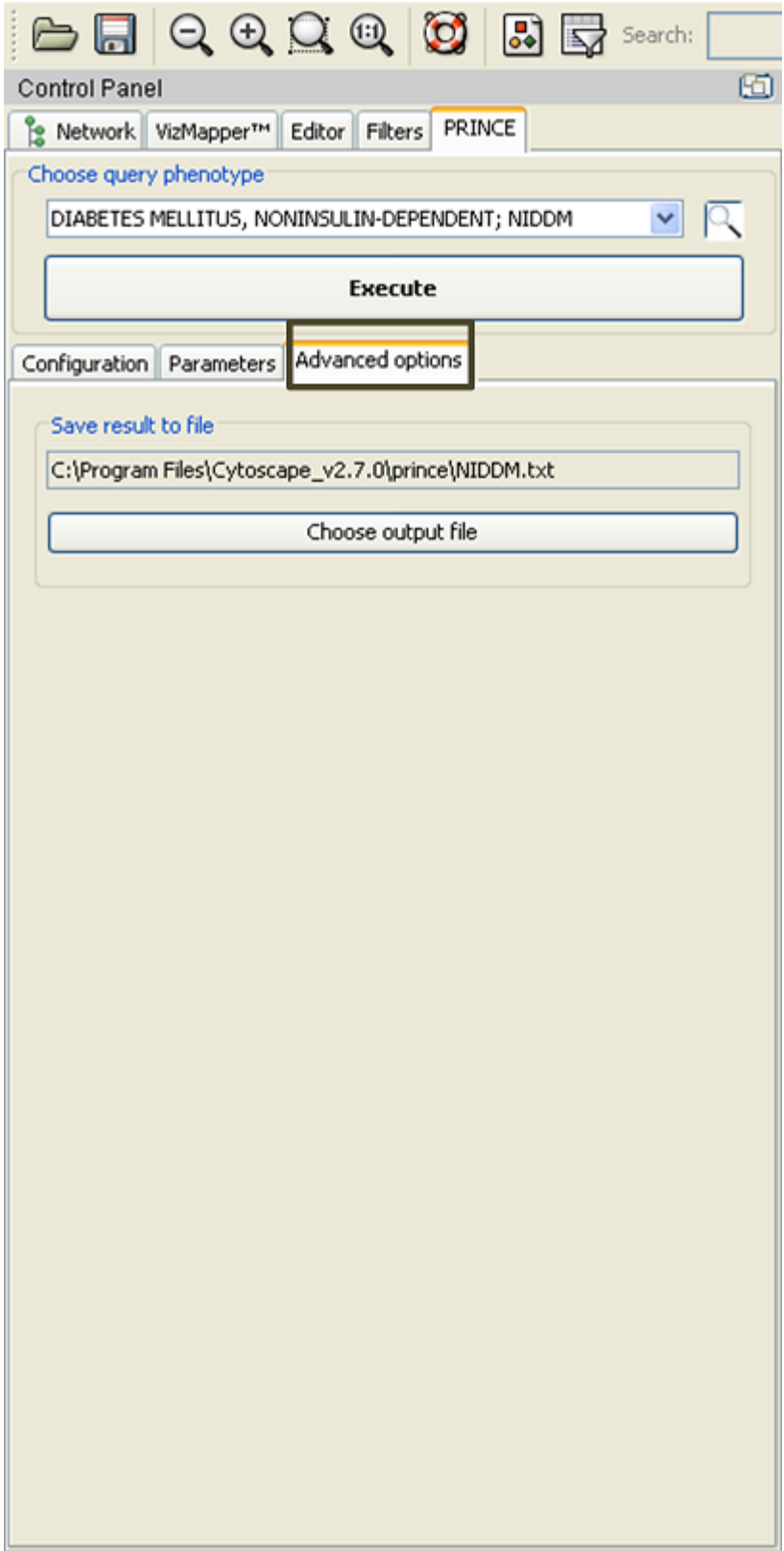
Figure 5. Specifying an output file.

4. Executing a query

After uploading of the input files and adjusting the parameters, select the query disease using either the disease name or the MIM code. Selection can be done through either (i) browsing the ordered list; (ii) typing the prefix of the requested disease or (iii) using the search box.

5. Output

The PRINCIPLE plug-in outputs the subnetworks formed from the top *k* nodes and their direct neighbors (Fig. 6).
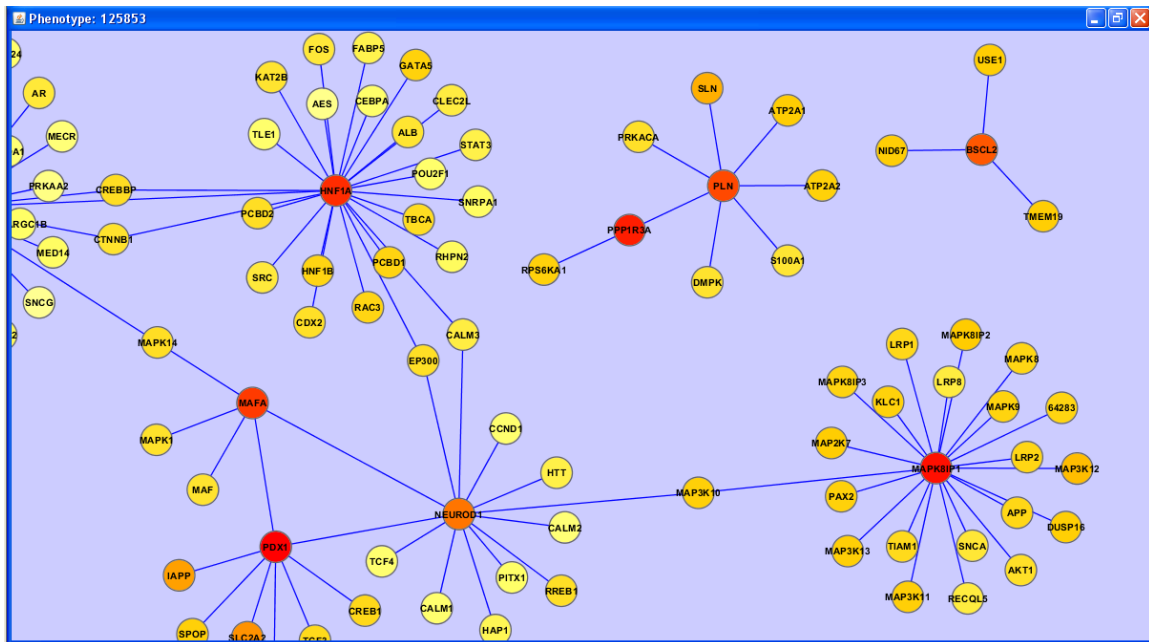The title of the network displays the MIM code and the red intensity is proportional to the protein rank.

Figure 6. Example output subnetwork (extract from the total subnetwork).

Upon choosing a set of proteins form the output subnetworks, the node attribute browser tab displays the rank and score of each protein in the chosen set (Fig. 7).

Right clicking on any protein introduces additional sources for information on the protein.
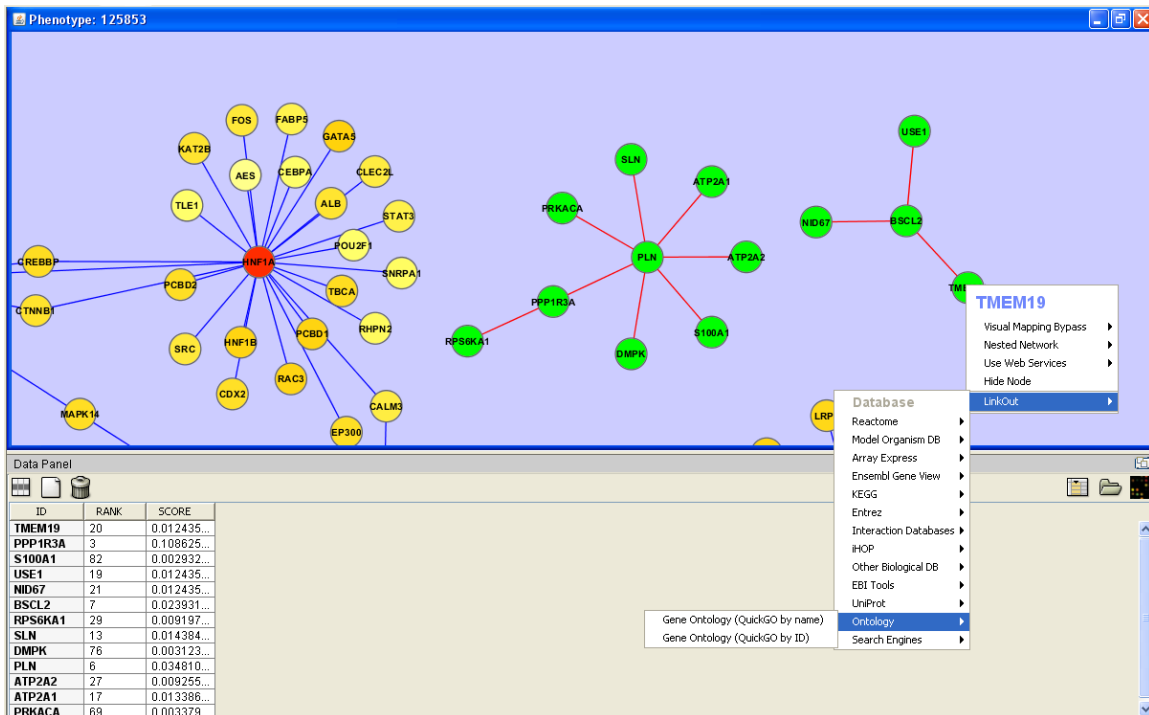
Figure 7. Example output network showing properties of selected proteins (green nodes).

References

1. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R: **Associating genes and protein complexes with disease via network propagation**. *PLoS Comput Biol* 2010, **6**(1):e1000641.