OVERVIEW

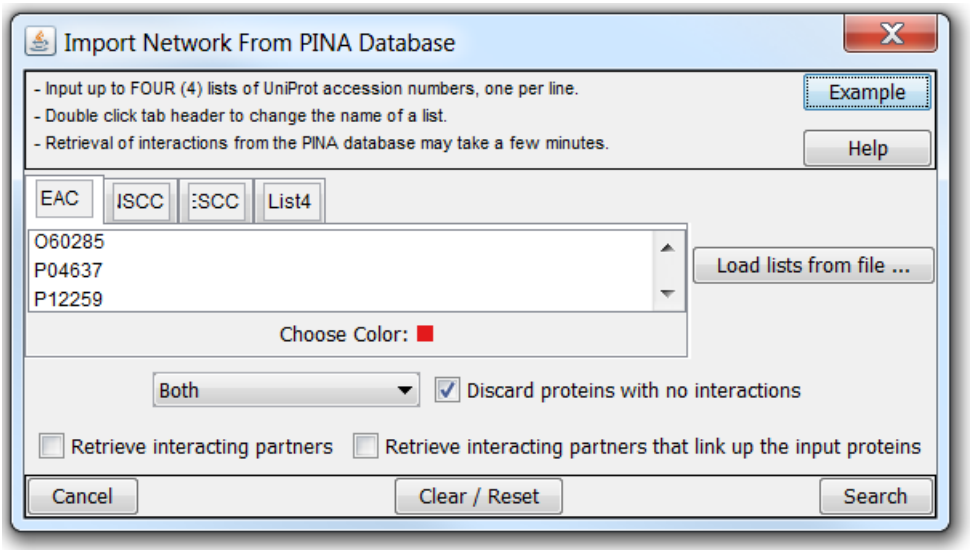
The **P**rotein **I**nteraction **N**etwork **A**nalysis For **M**ultiple **S**ets (PINA4MS) is a visualization and analysis tool to study interactions between multiple sets of proteins. It uses the curated protein-protein interaction data from **PINA** and kinase-substrate relationships from PhosphoSitePlus to link both intra- and inter- multiple sets of proteins. Human Protein Atlas (HPA) information on protein expression has been incorporated. A number of unique features have been developed to facilitate the analysis including: 1) **Customized node pie chart** allowing users to easily identify common and distinct proteins in the input sets, as well as interactions between different sets; 2) A **clickable Venn diagram** is provided on the side panel to support quick selection and highlight of exclusive set(s) of proteins in the network; 3) A customized layout algorithm, **Venn-Galaxy** layout, is implemented to provide an elaborate view on the localization of the protein sets and highlight the interface and connections in between; 4) Customized sub-menus and check box panels to support quick selection and highlight of the protein expression data from the Human Protein Atlas.

Protein-Protein Interaction SEARCH

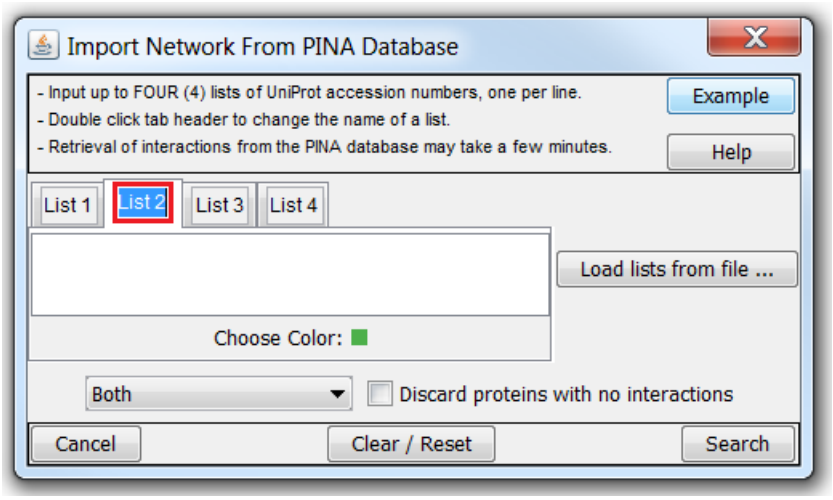
After starting PINA4MS from Cytoscape, an input dialog would pop up.

Enter up to FOUR lists of UniProt ACs

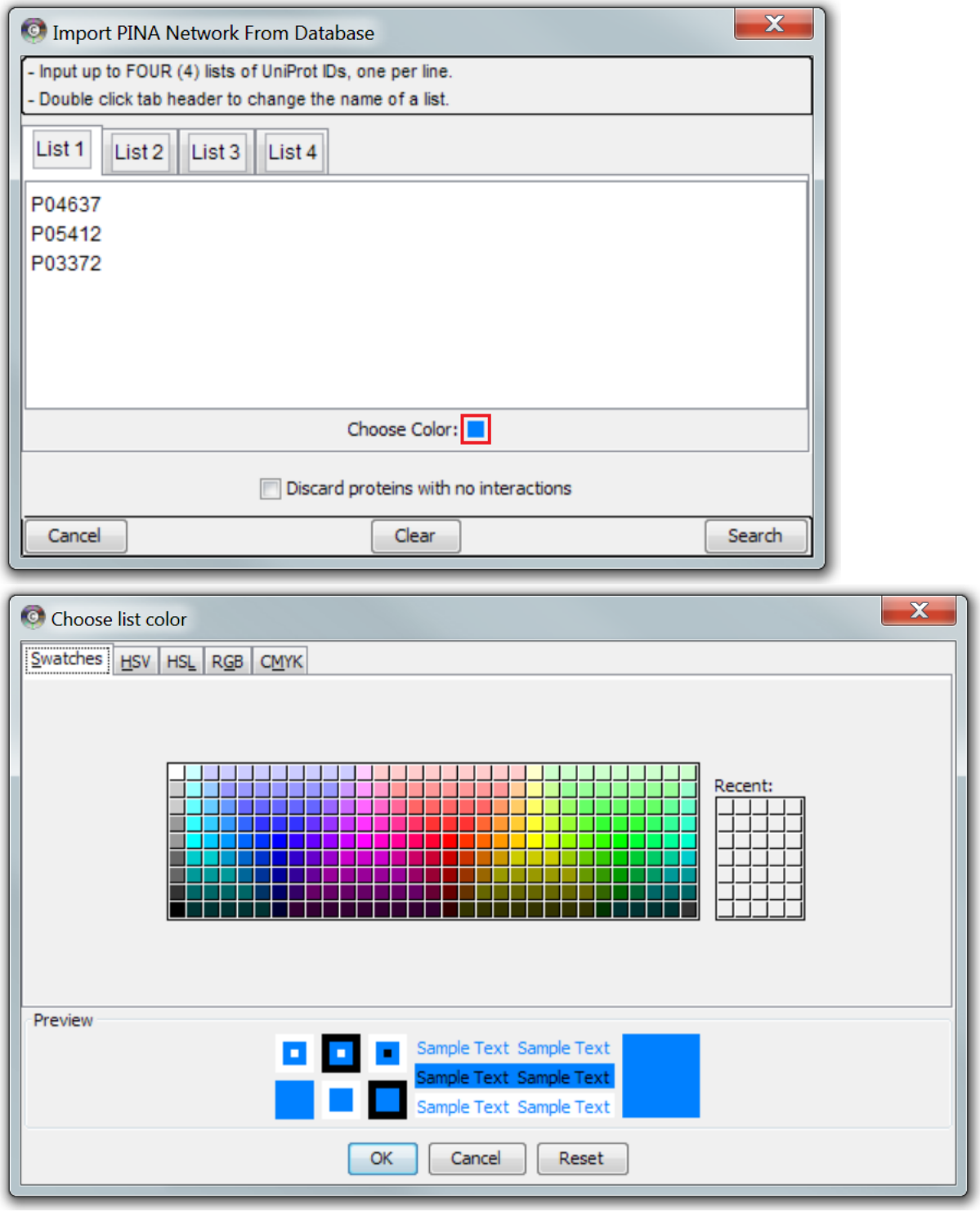
1. To search the PINA database, enter UniProt ACs into the tab text box, each on its own line:



2. To name a list of ACs, **double-click the tab header** and enter the desired identifier:

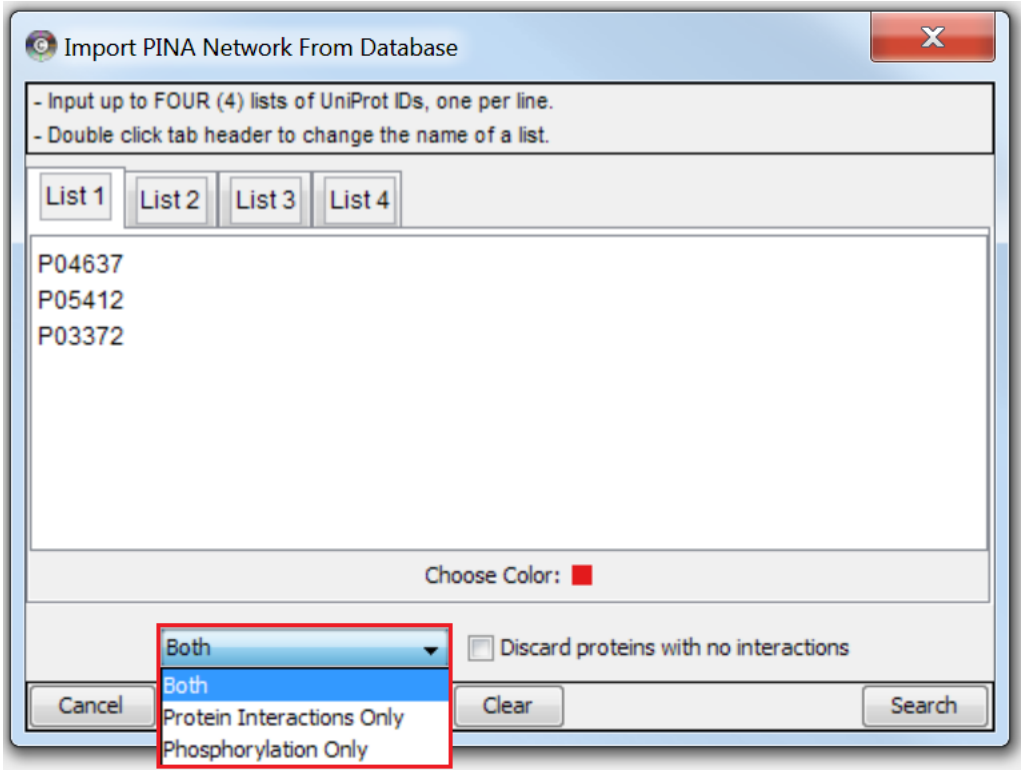

Import Network From PINA Database
- Input up to FOUR (4) lists of UniProt accession numbers, one per line.
- Double click tab header to change the name of a list.
- Retrieval of interactions from the PINA database may take a few minutes.
Example
Help
List 1
List 2
List 3
List 4
Load lists from file ...
Choose Color:
Both
Discard proteins with no interactions
Cancel
Clear / Reset
Search

3. To change the representative colour for a list, click on the **Choose Color** box at the bottom of the text box and choose the desired colour:

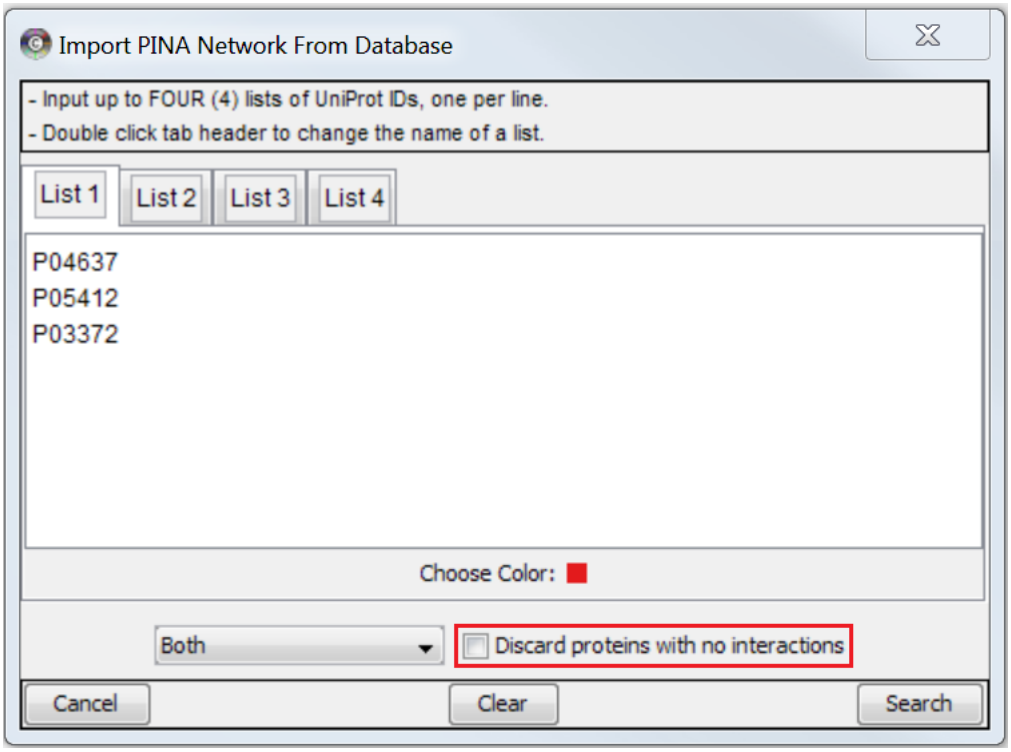


4. Repeat steps 1-3 for the next list(s).

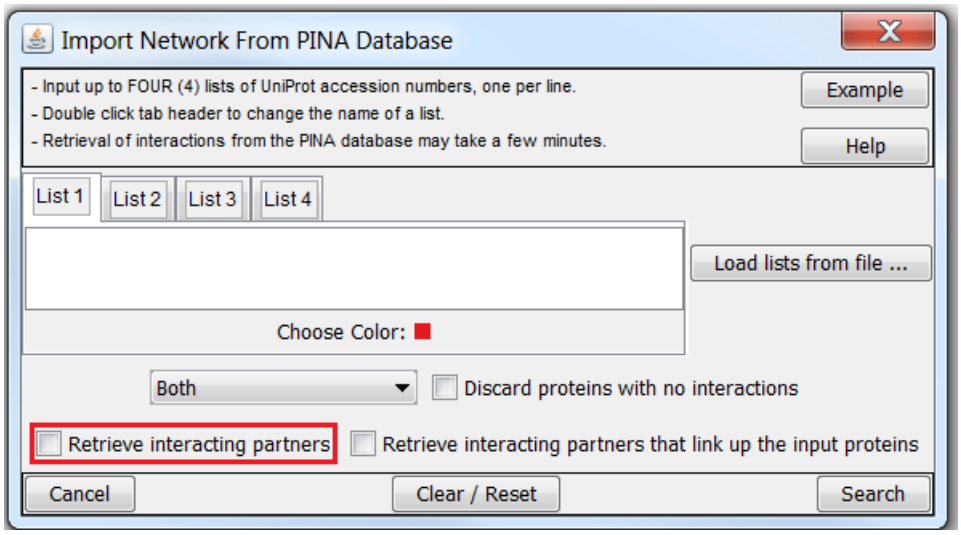
5. To choose the type of interactions to query, click the **drop-down menu** beneath the color chooser and select **Protein Interactions Only**, **Phosphorylation Only** or **Both**:



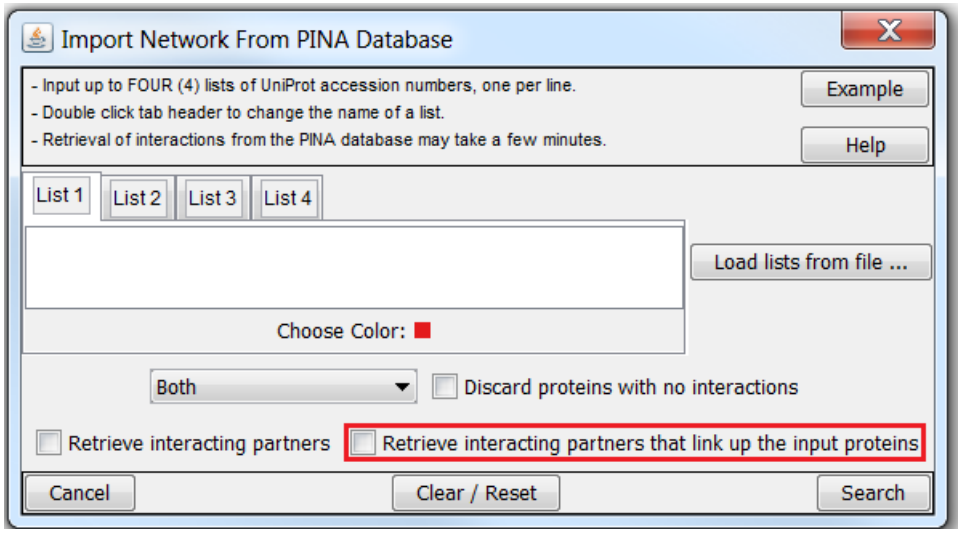
6. With the **Discard proteins with no interactions** option, the network will consist of only the proteins with identified PPIs (including those with only self-interactions) in PINA:



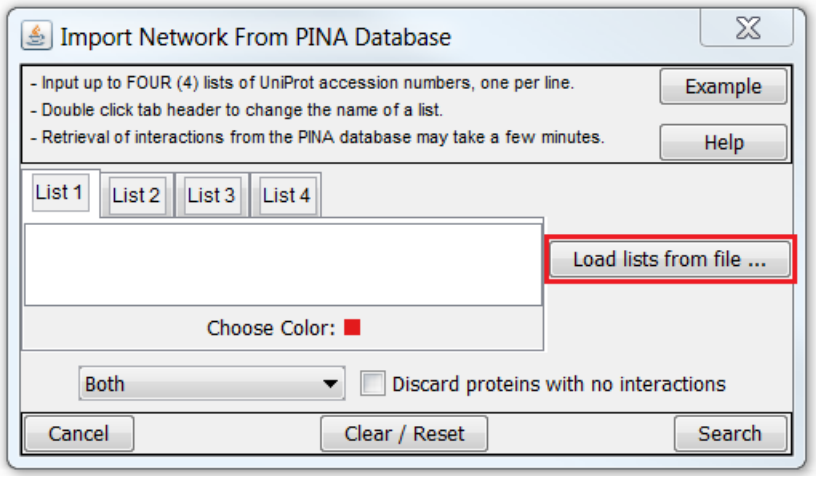
7. With the **Retrieve Interacting Partners** option, an additional search will be performed which finds any proteins interacting with the query proteins in the database:



8. With the **Retrieve interacting partners that link up the input proteins** option, an additional search will be performed which finds any proteins in the PPI database linking up two or more of the query proteins. That is, a dangling protein which only connects to one of the query proteins will not be retrieved:



9. Click **Search** to query the PINA database and create the network.
10. Alternatively, the list(s) and header(s) can be loaded from a file directly via the **Load lists from file …** button. The file should be formatted into tab-separated columns, in which the first column represents the first list of UniProts, the second column contains the second list, and so on. The first line of the file must contain the header(s) of the list(s):




```
1   EAC   →  ESCC  →  HNSCC
2   Q92556 → P42771 → P01112
3   Q13485 → Q8N726 → Q8WXH0
4   O43897 → P42336 → Q15648
5   Q13023 → P60484 → Q8NF91
6   Q8NF91 → Q16236 → Q14790
7   Q8N0X7 → Q9UKF5 → P42771
8   P42771 → P06400 → Q8N726
9   Q8N726 → O15550 → P42336
10  Q76N89 → Q14517 → P60484
11  P42336 → P46531 → Q15910
12  Q92794 → Q09472 → O14896
13  Q92608 → Q49AJ0 → P57078
14  O14497 → O14686 → P06400
15  P51532 → P04637 → Q14517
16  Q9UKB5 → Q32MQ0 → P46531
17  Q5T1H1 → Q9NYQ8 → Q9H3D4
18  Q68CP9 →     →    Q9UM47
19  O00206 →     →    Q04721
20  Q9Y5Y9 →     →    P04637
```

11. To set the representative colours via file loading, a line can be added after the header line to indicate the colour for each list using the “Color:” flag and HTML colour hash codes:

```
1   EAC   →  ESCC  →  HNSCC
2   Color:#FF0000  →  Color:#00FF00  →  Color:#000000
3   Q92556 → P42771 → P01112
4   Q13485 → Q8N726 → Q8WXH0
5   O43897 → P42336 → Q15648
6   Q13023 → P60484 → Q8NF91
7   Q8NF91 → Q16236 → Q14790
8   Q8N0X7 → Q9UKF5 → P42771
9   P42771 → P06400 → Q8N726
10  Q8N726 → O15550 → P42336
11  Q76N89 → Q14517 → P60484
12  P42336 → P46531 → Q15910
13  Q92794 → Q09472 → O14896
14  Q92608 → Q49AJ0 → P57078
15  O14497 → O14686 → P06400
```

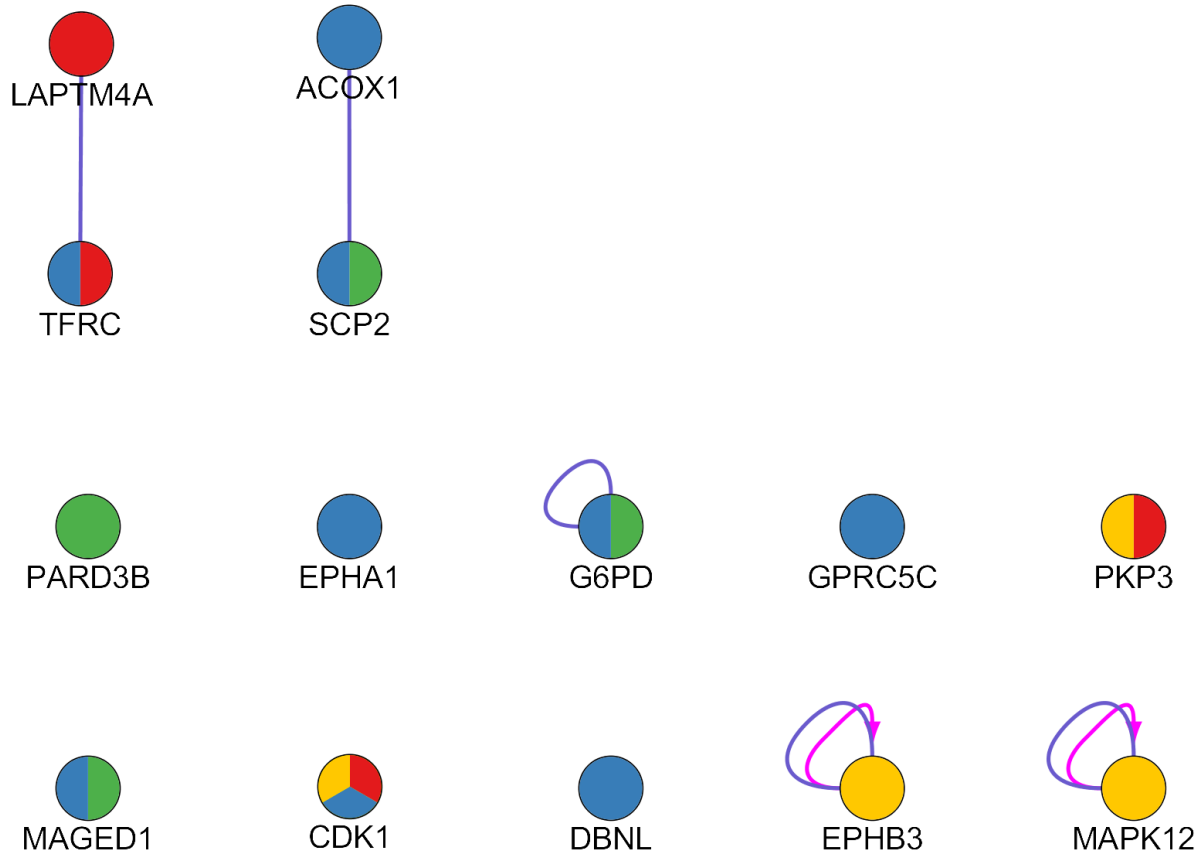
The first list will be red (#FF0000), second list green (#00FF00), and third list black (#000000).

NETWORK VIEW

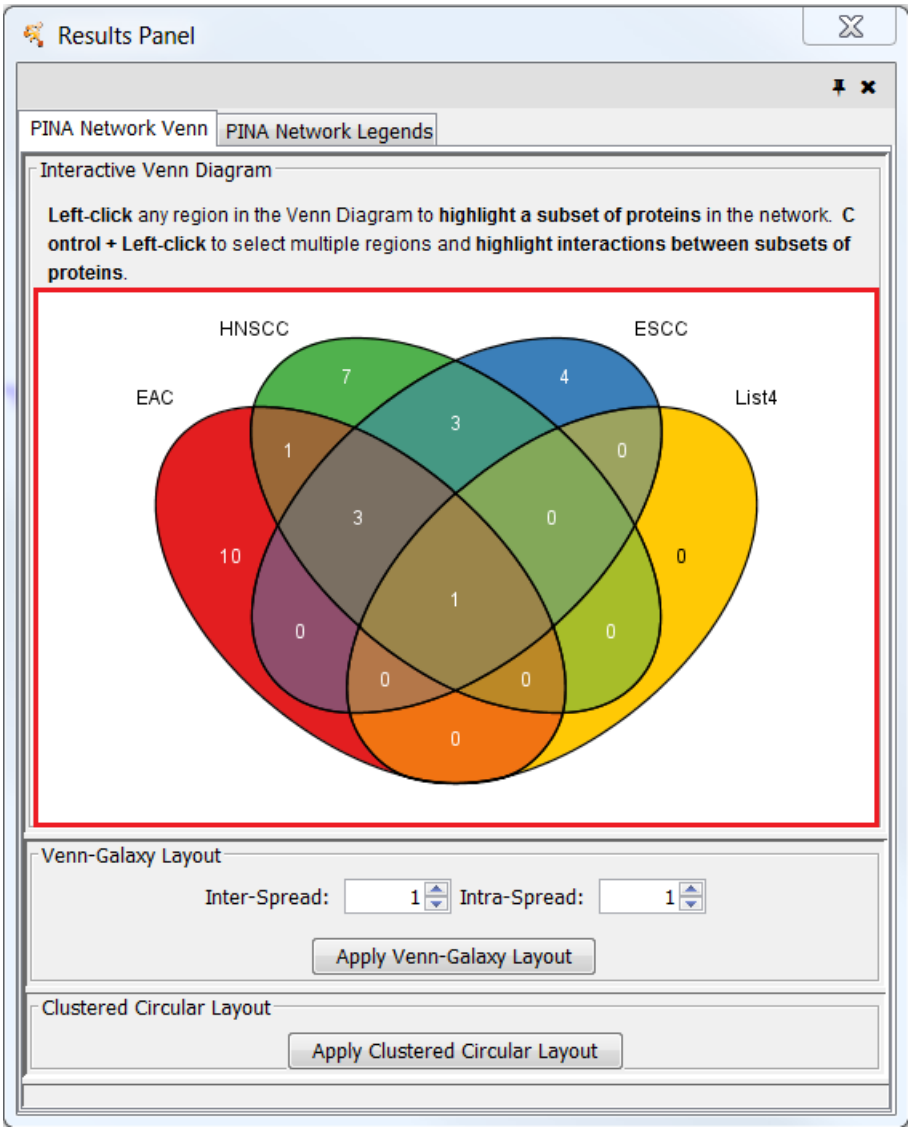
Generic protein-protein interactions (PPIs) are displayed as ***blue lines***. Kinase-substract interactions (KSs) are displayed as ***magenta arrows***.

Custom Node Pie Chart

Protein nodes are shown using custom node pie charts to highlight the presence / absence of each protein in each of the input lists. If a protein is present in a particular list or lists, the corresponding section of the node pie chart is painted with the list colour(s):



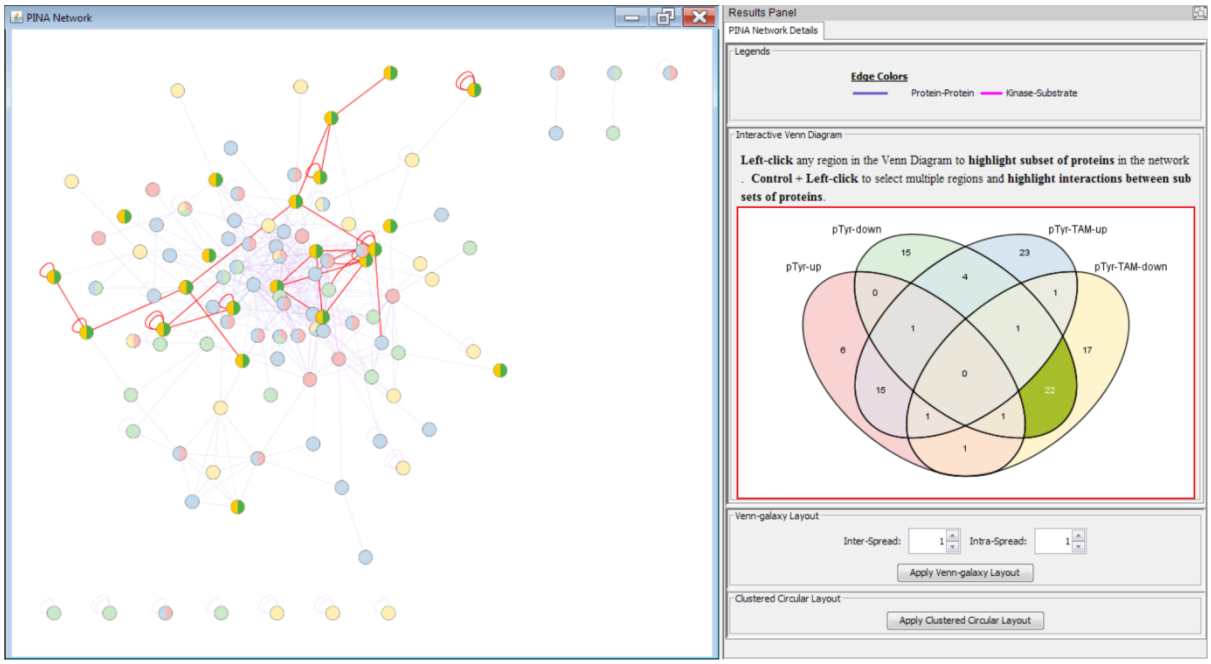
Venn Diagram



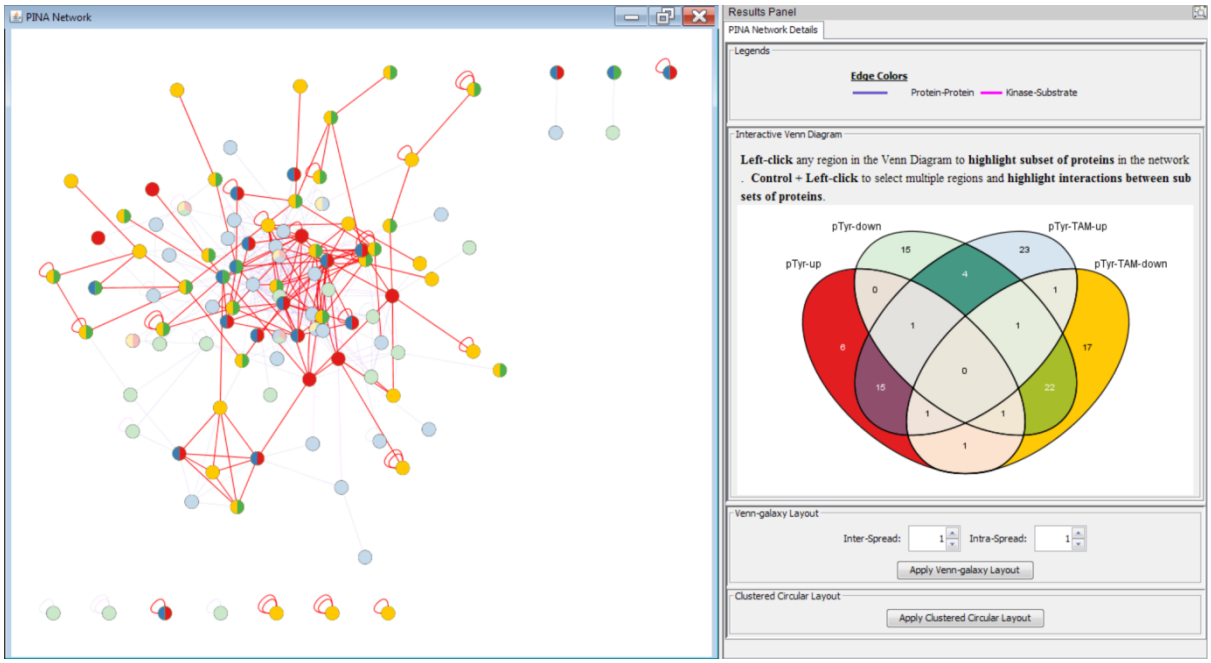
A clickable Venn diagram is provided on the side panel to support quick selection and highlight of exclusive set(s) of proteins in the network. The number of nodes belonging to each exclusive Venn set is shown in the centre of the corresponding Venn area:

1. Left-click an exclusive Venn area to select and highlight the corresponding nodes and edges in the network. A generic PPI is highlighted if both interacting nodes are in the exclusive Venn set. A kinase-substract interaction is highlighted if the source node is in the exclusive Venn set, regardless of whether the target node's presence or absence in the exclusive Venn

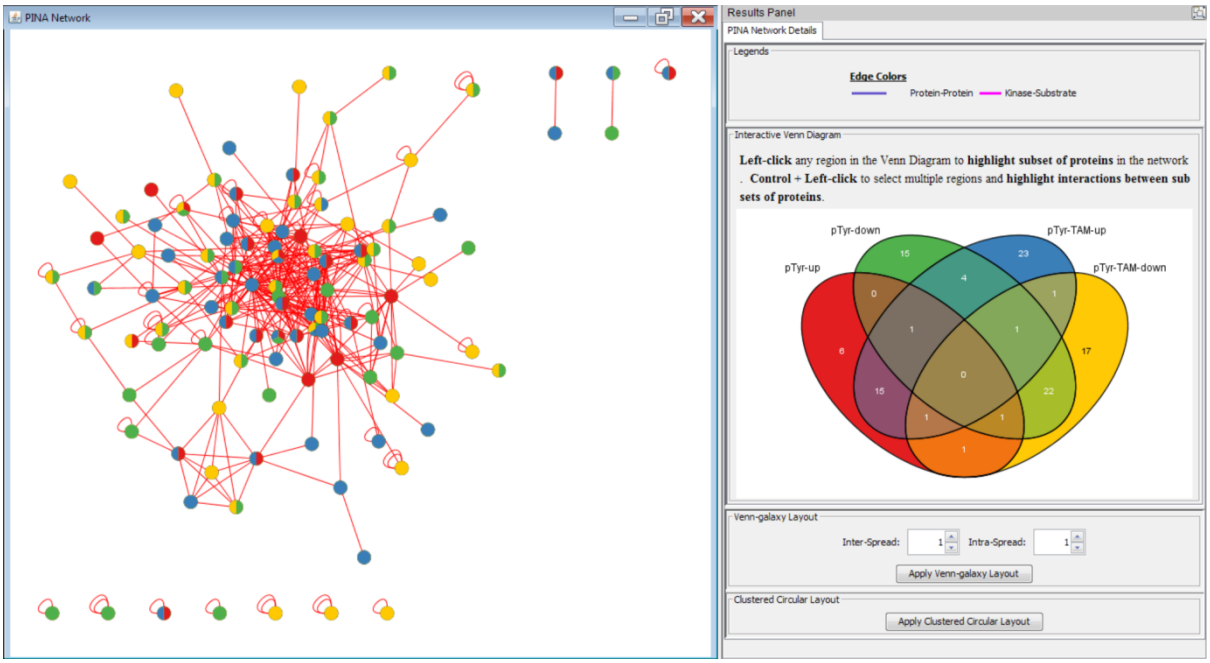
set. All remaining nodes and edges are dimmed:



2. Ctrl-click multiple exclusive Venn areas to select and highlight multiple exclusive Venn sets in the network:

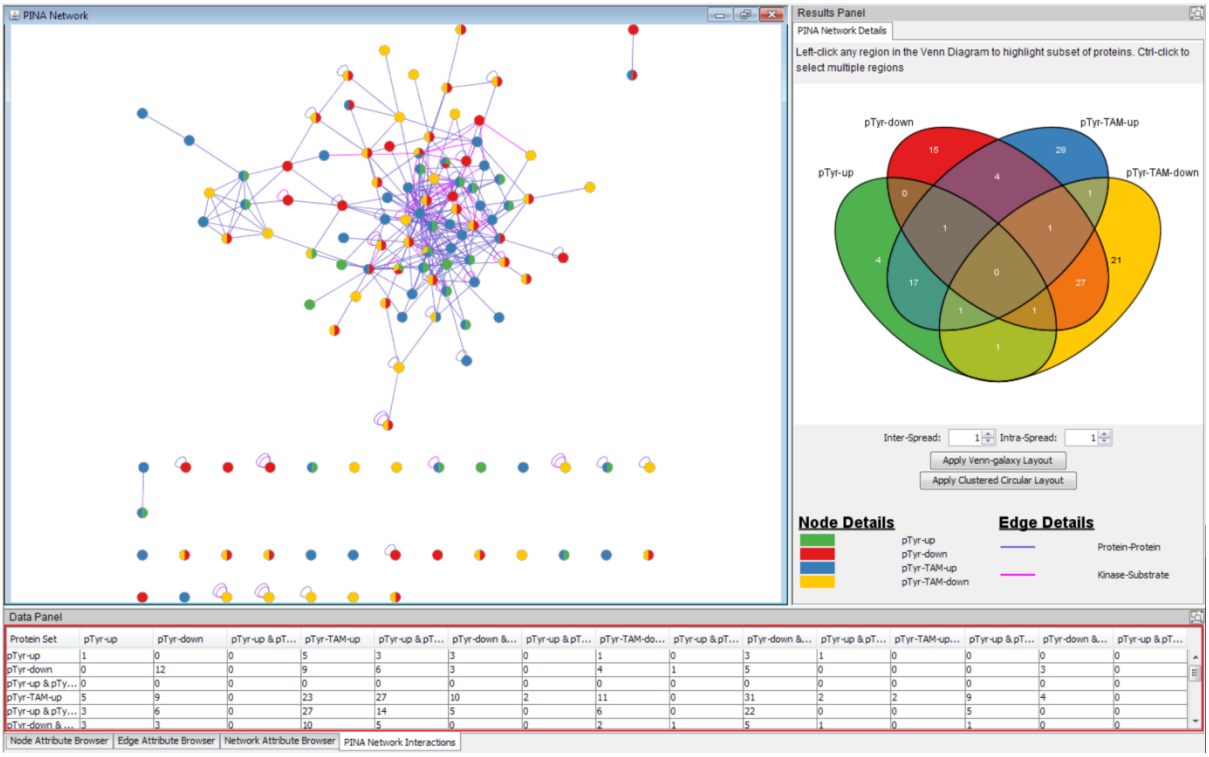


3. By default, all exclusive areas, nodes and edges are highlighted after the network has been generated:

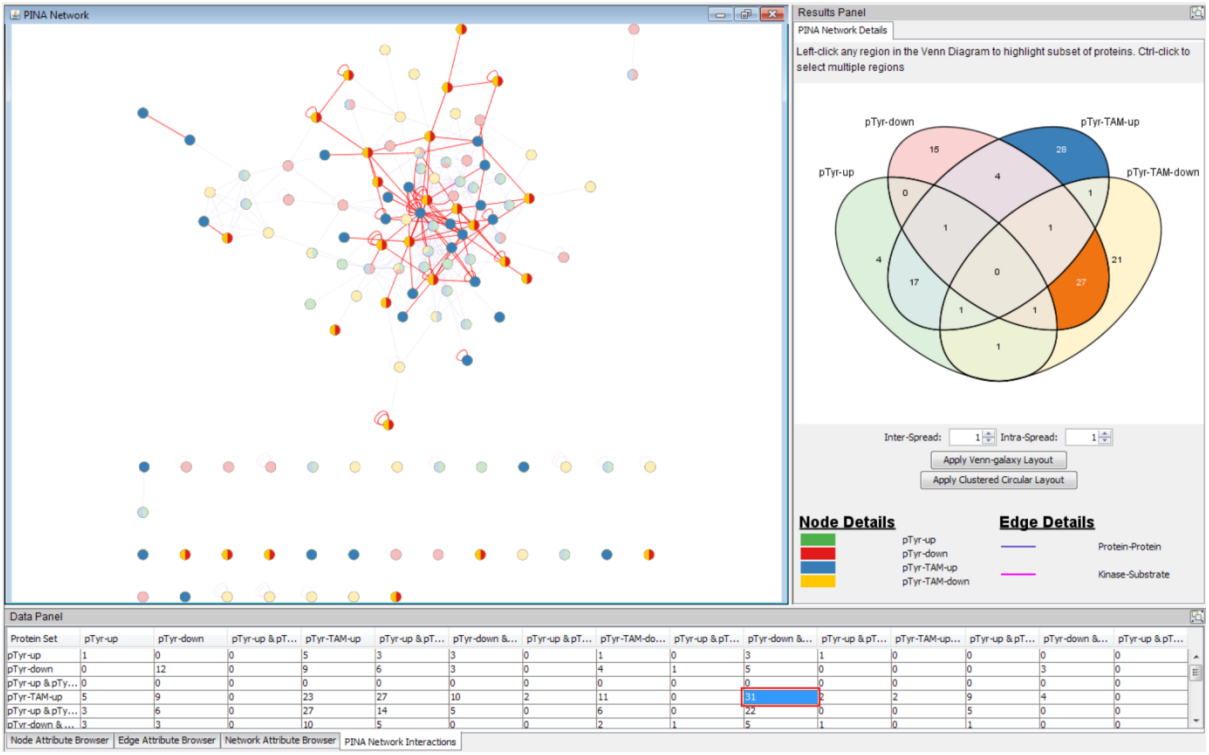


Interaction Table

In addition to the Venn diagram, an interaction table is provided on the bottom data panel to display the number of interactions between any two exclusive Venn sets:

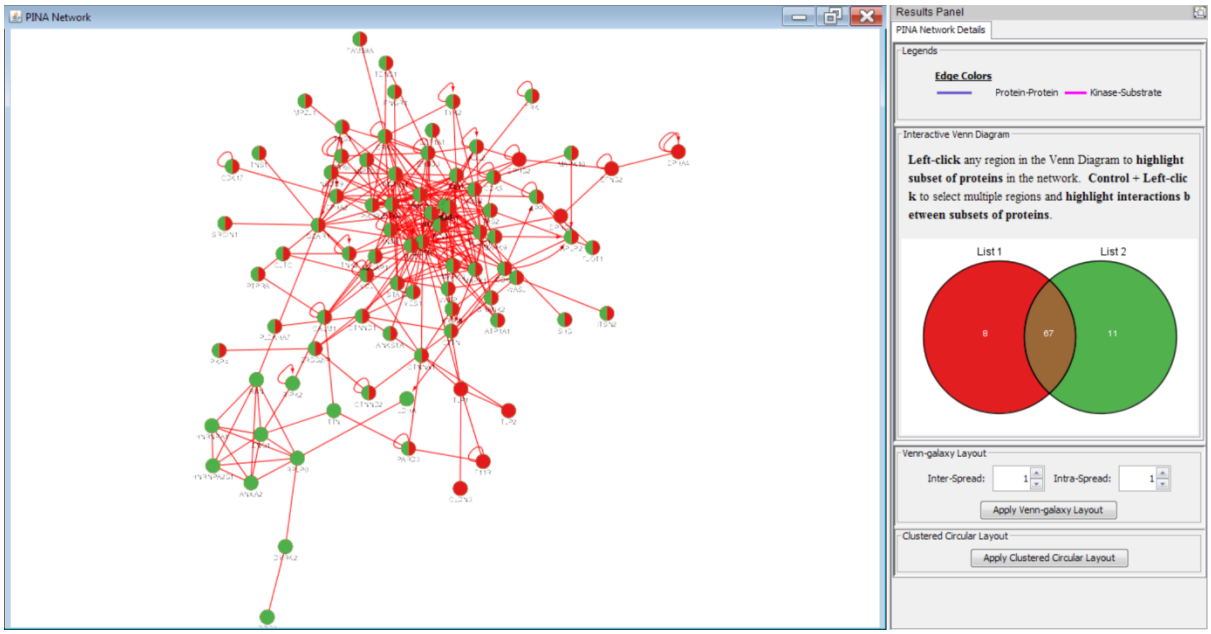


Similar to the Venn diagram, each cell in the table is clickable and can be used to quickly select the two exclusive Venn sets involved. For example, the following example shows a cell being selected and the corresponding Venn sets are highlighted in the network as well as the Venn diagram:

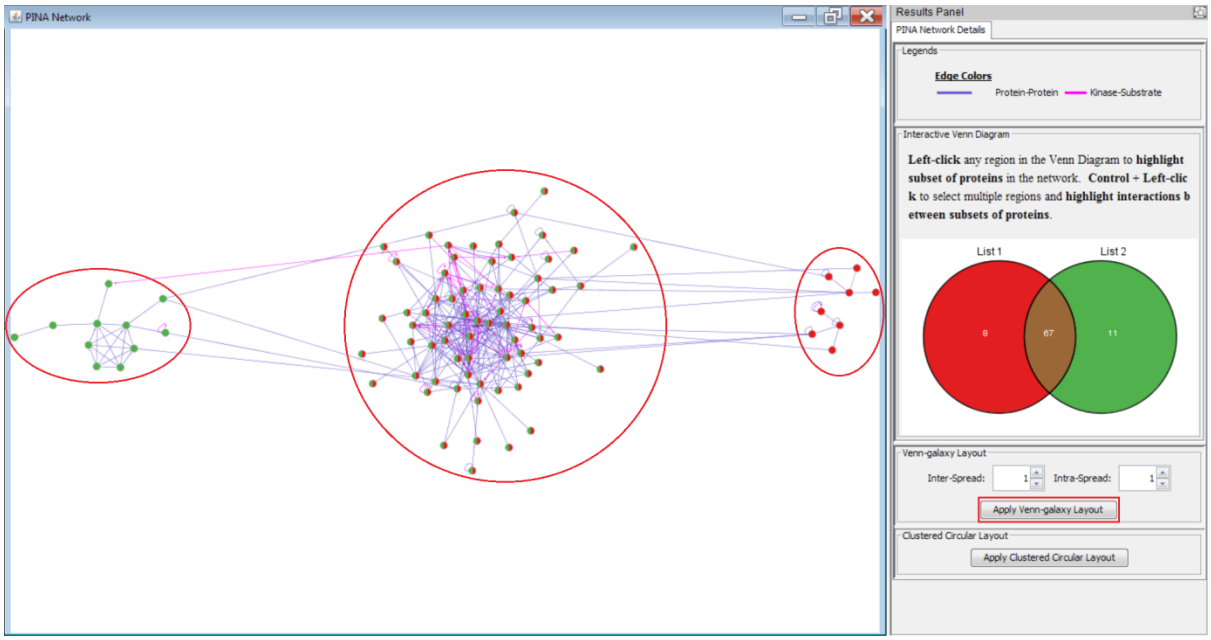


Venn-Galaxy Layout

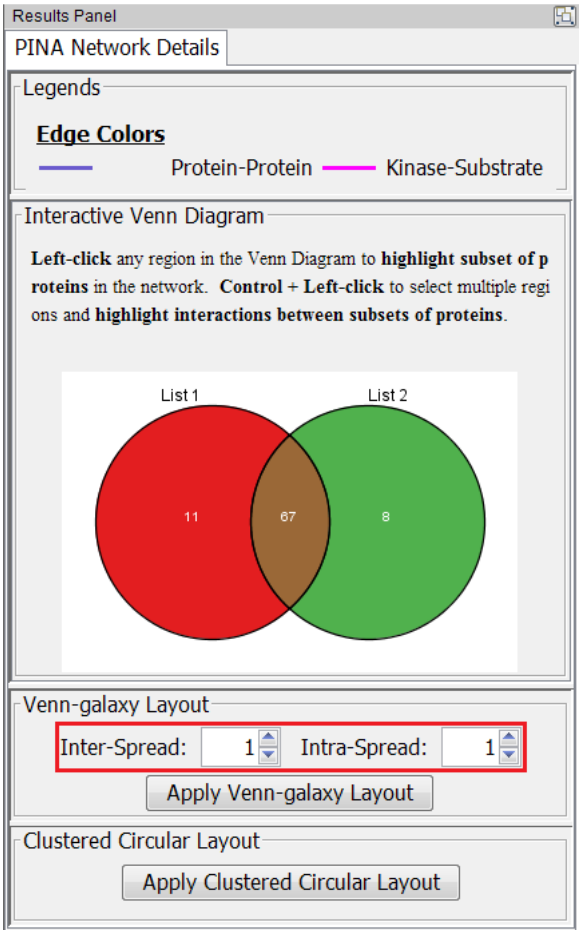
1. By default, PINA networks are displayed with force-directed layout (unweighted):



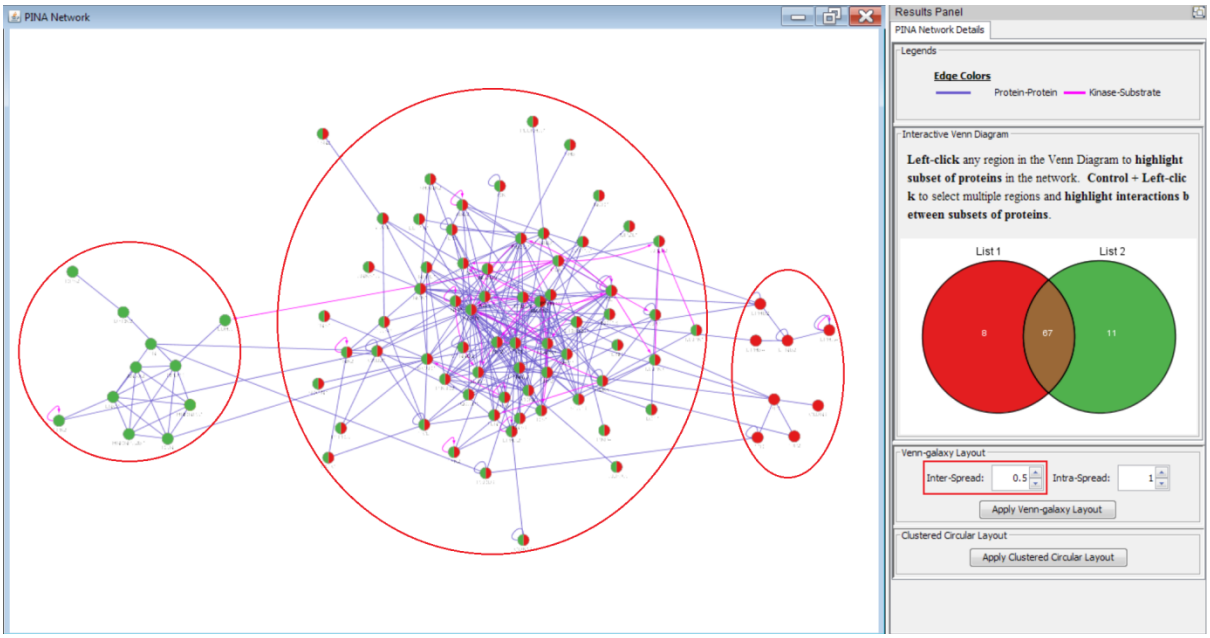
2. To rearrange the network and place protein nodes from the same exclusive Venn set together, click the **Apply Venn-Galaxy Layout** button on the side panel:



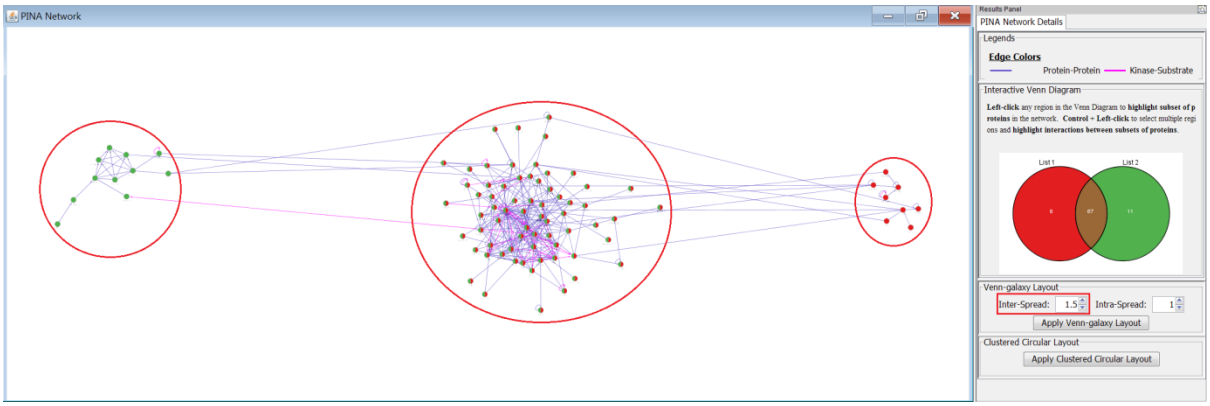
3. To increase or decrease the spread of the exclusive Venn sets in the layout, change the **Spread** attributes. The range is 0.01 – 2, with values < 1 compacting the layout and values > 1 spreading the nodes further apart. There are two **Spread** attributes: **Inter-** and **Intra-Spread**. Changing **Inter-Spread** controls the distance *between* the exclusive Venn sets, while **Intra-Spread** controls the spread of the nodes *within* each exclusive Venn set:

4. 

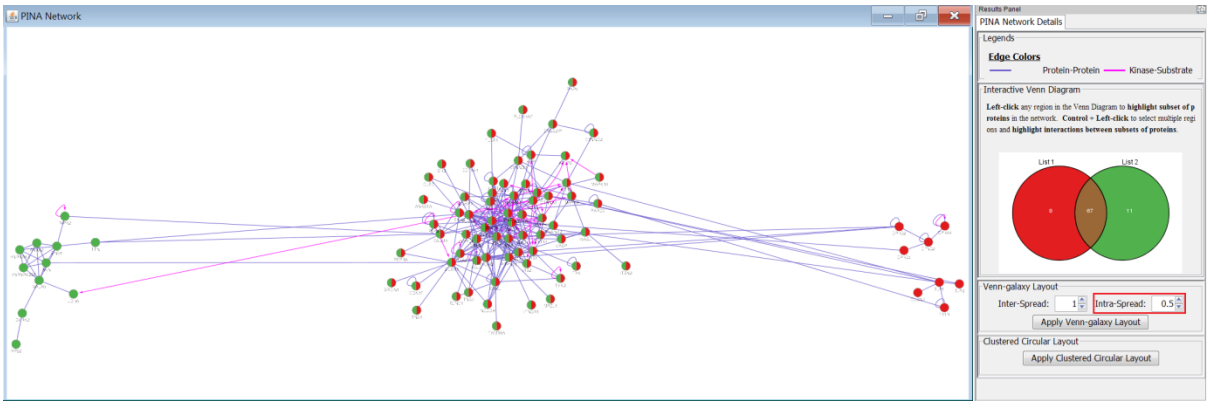
5. Set **Inter-Spread** to 0.5 and click the **Apply Venn-Galaxy Layout** button. The exclusive Venn sets will be placed closer with each other:



6. Set **Inter-Spread** to 1.5 and click the **Apply Venn-Galaxy Layout** button. The exclusive Venn sets will be placed further apart:

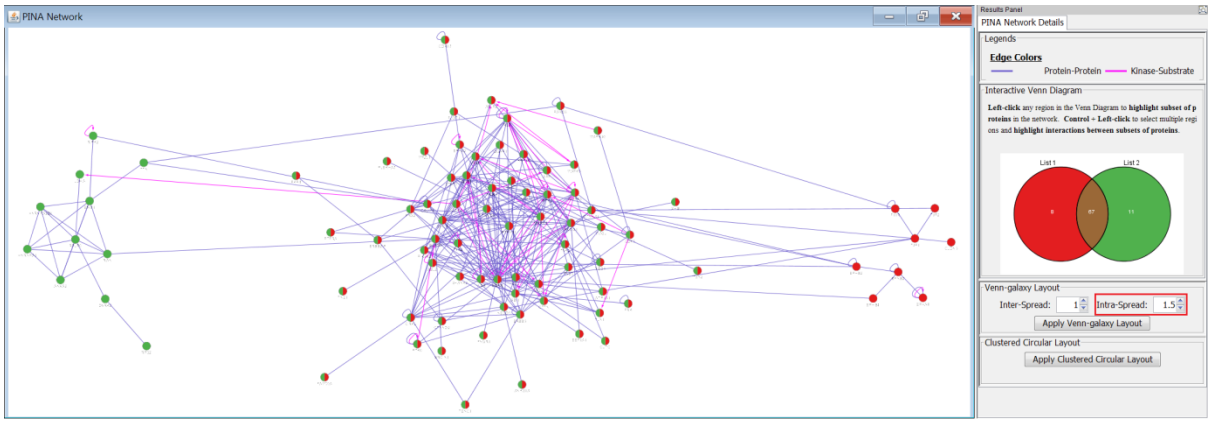


7. Keep **Inter-Spread** at 1, set **Intra-Spread** to 0.5 and click the **Apply Venn-Galaxy Layout** button:

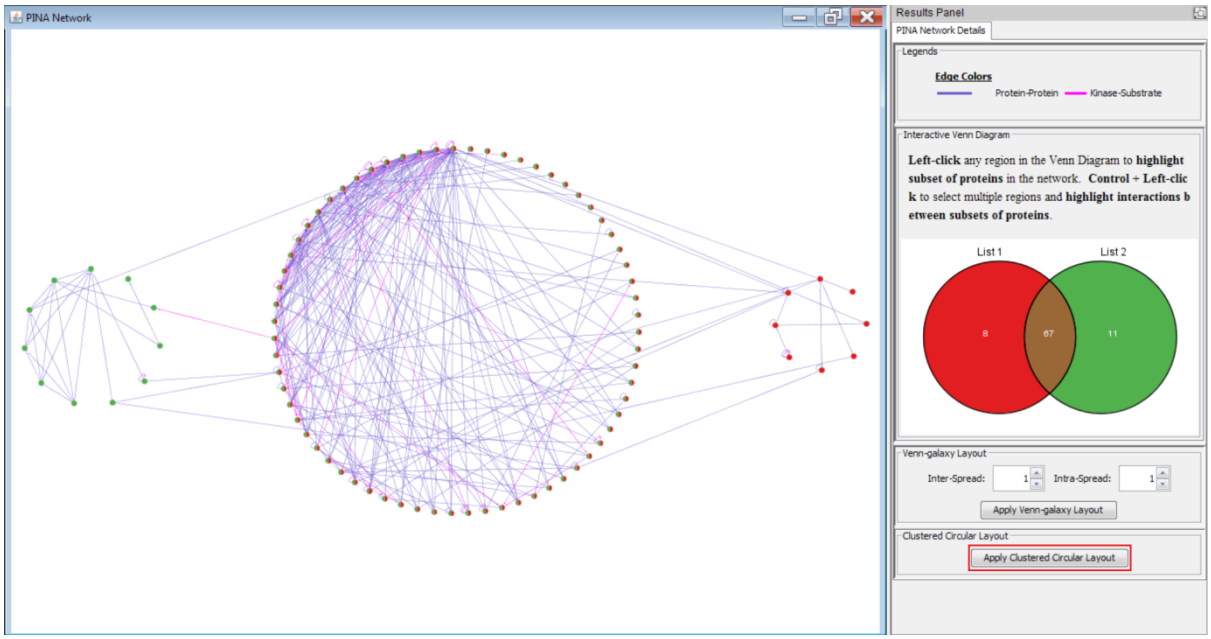


8. Set **Intra-Spread** to 1.5 and click the **Apply Venn-Galaxy Layout** button. The distance between nodes will increase without changing the distance between the centres of the Venn

sets:



9. Alternatively, use Clustered Circular Layout to separate the exclusive Venn sets. Click the **Apply Clustered Circular Layout** button:

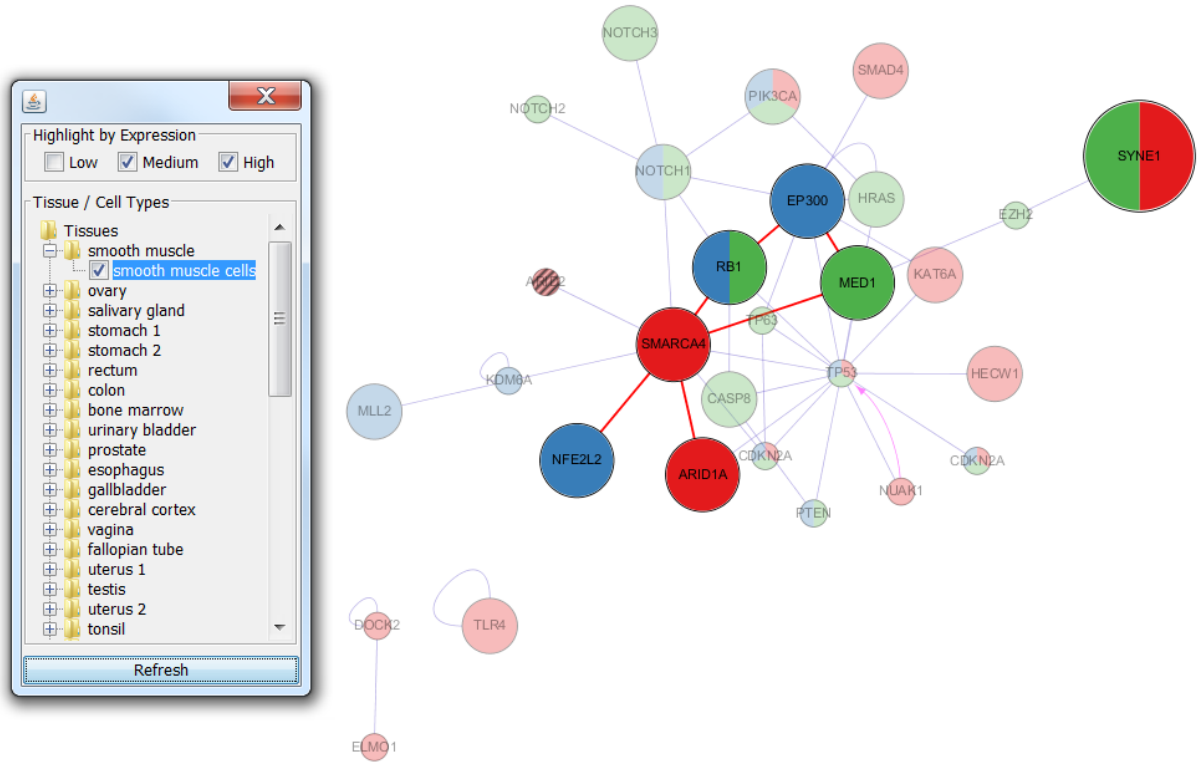


THE HUMAN PROTEIN ATLAS

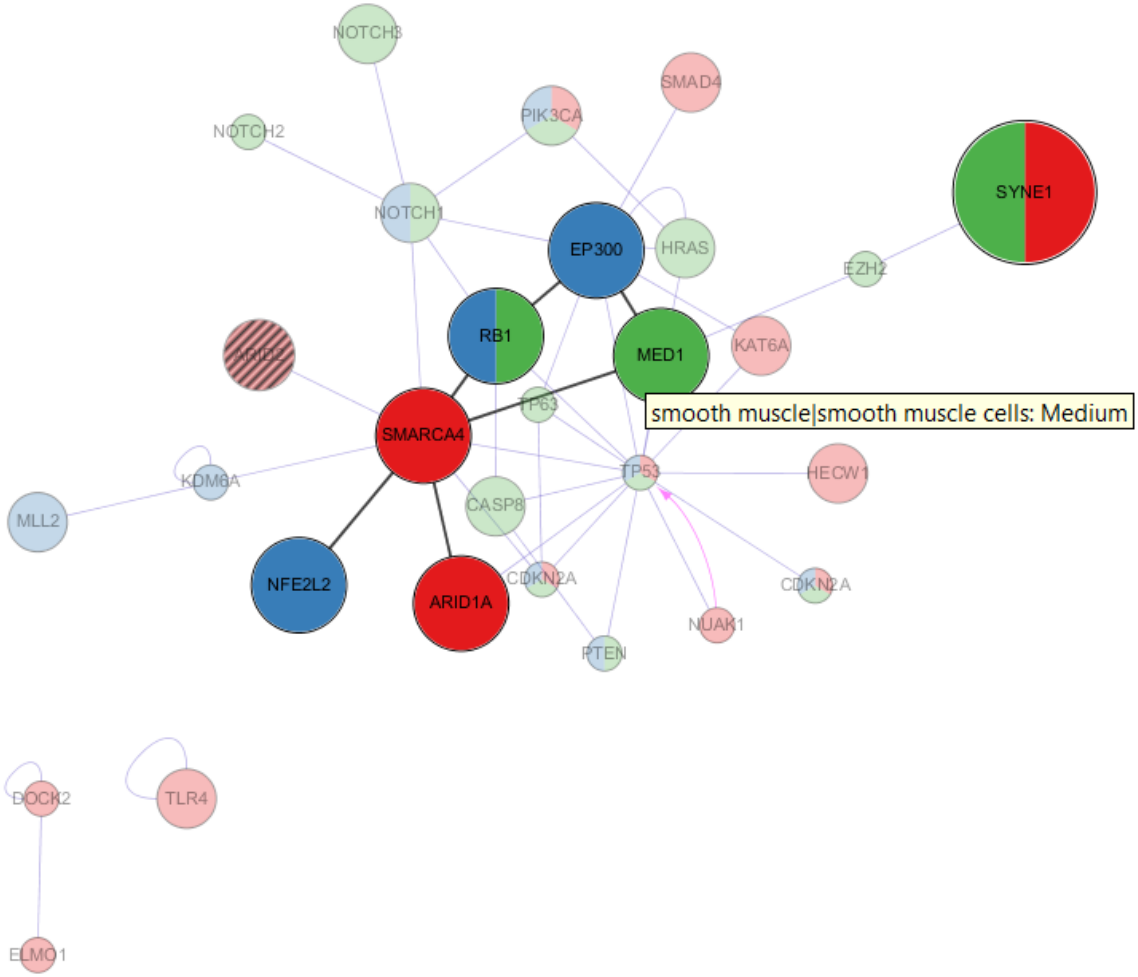
The curated protein profiling data for 80+ normal tissue / cell types, 20 cancer tumor types and 20 subcellular locations from the Human Protein Atlas (HPA, http://www.proteinatlas.org/) have been incorporated into PINA4MS and users can select proteins based on their expression data in the three different contexts. Note that the selection of proteins based on HPA data is complementary to previously mentioned functionalities (e.g. Venn diagram, Interaction Table), in that users can select and highlight nodes and edges belonging to a cancer type for example, and additionally use the functionalities above to further select for those cancer proteins in an exclusive Venn set or sets. However, selection using tissue, cancer, subcellular location data are mutually exclusive. That is, selecting tissue will remove any previous selection on cancer / subcellular location, and vice versa.

Tissue Protein Expression Panel

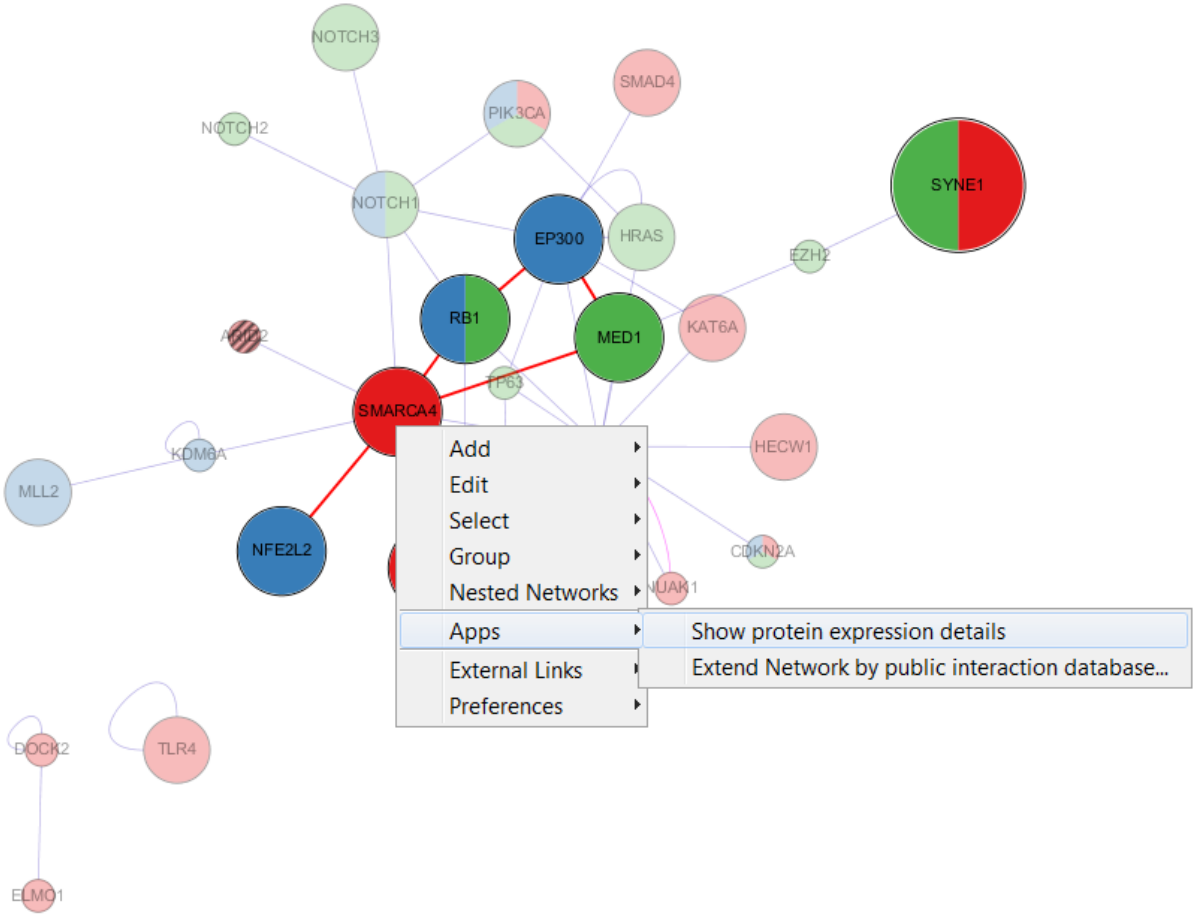
The pop-up Tissue Panel (accessible via **Apps → PINA4MS → Identify sub-network by protein expression → in Human Normal Tissue …**) contains a clickable checkbox tree to support quick selection and highlight of proteins belonging to each particular tissue / cell type. Check any tissue / cell type and click **'Highlight sub-Network'**. The proteins that are expressed (with HPA expression level of either "Low", "Medium" or "High") are highlighted, the rest are dimmed. In addition, proteins with no HPA normal tissue information are dimmed and overlaid with a Moiré pattern. The size of the node reflects the expression level of the protein in the tissue. Users can additionally choose to highlight only proteins with certain expression level(s), by using the check boxes at the top of the tissue panel.



For extra expression information for each protein, users can hover the mouse over a node, and tooltips will provide a brief summary of the node information.



Right-click a node → **Apps** → **Show protein expression details** will display a more detailed summary of the protein information on the control (West) panel (see below):



Network | Style | Select | PINA Network Protein Expression

Protein

Gene Name	SYNE1
UniProt Accession	Q8NF91

Cancer

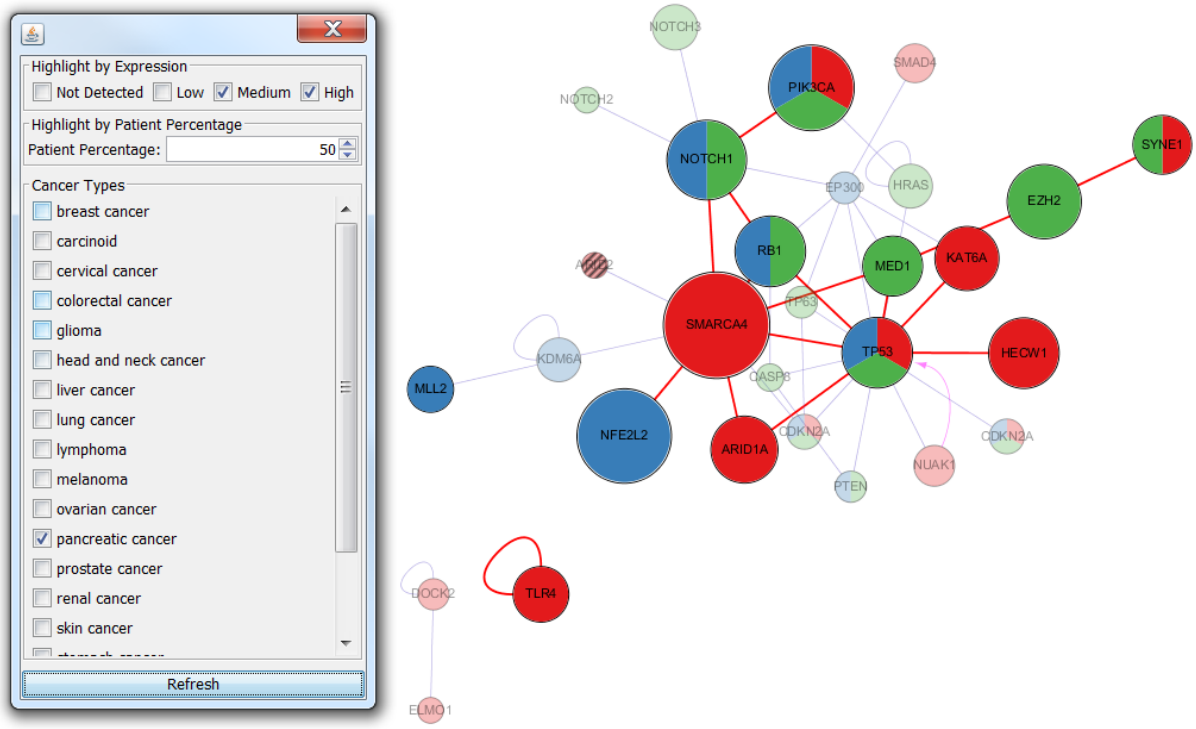
Selected Cancer Type	-
High Expression Patients	0
Medium Expression Patients	0
Low Expression Patients	0
No Detected Expression Patie...	0

Normal Tissue

Selected Tissue Type	smooth muscle
Selected Cell Type	smooth muscle cells
Tissue Expression Level	High

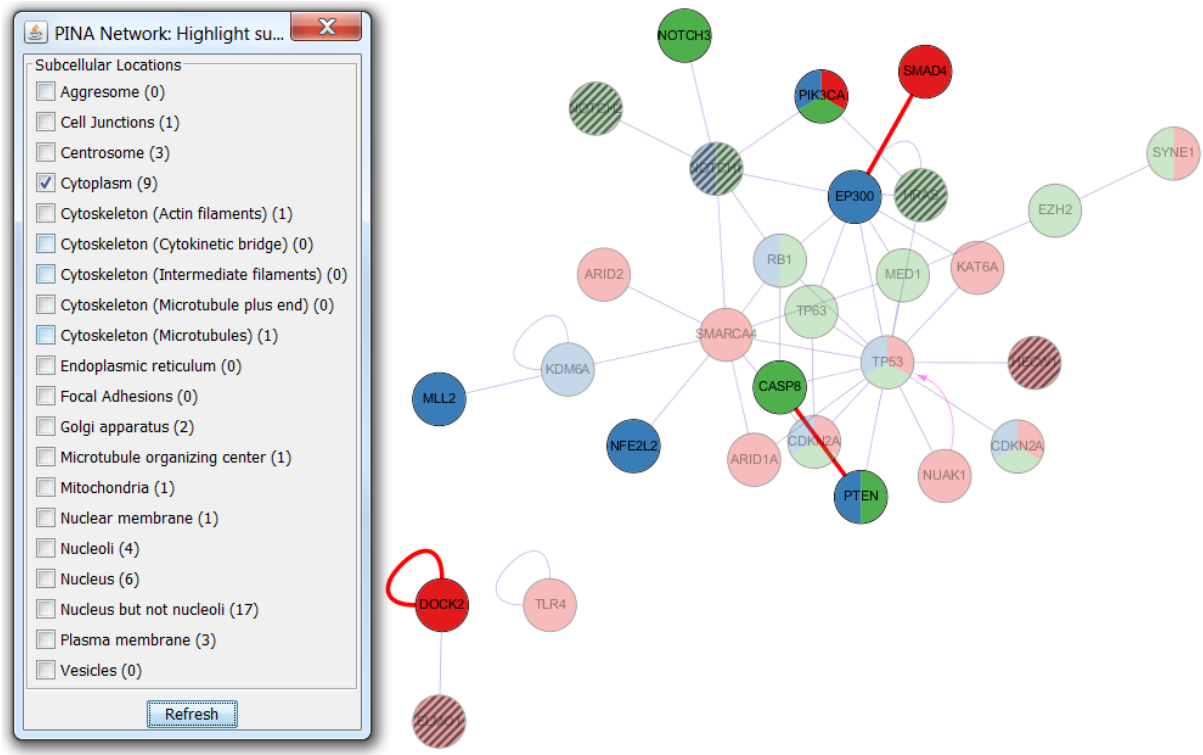
Cancer Protein Expression Panel

The pop-up Cancer Panel (accessible via **Apps → PINA4MS → Identify sub-network by protein expression → in Human Tumor Tissue …**) contains a series of check boxes to support quick selection and highlight of proteins belonging to each cancer type. The usage is similar to the Tissue Panel and the nodes are similarly displayed, with the node size proportional to the average expression level of the patients. An additional feature at the second panel from the top allows users to set a percentage cutoff, which will highlight only those proteins in the selected cancer type if the percentage of patients with the selected expression level(s) exceeds the cutoff.



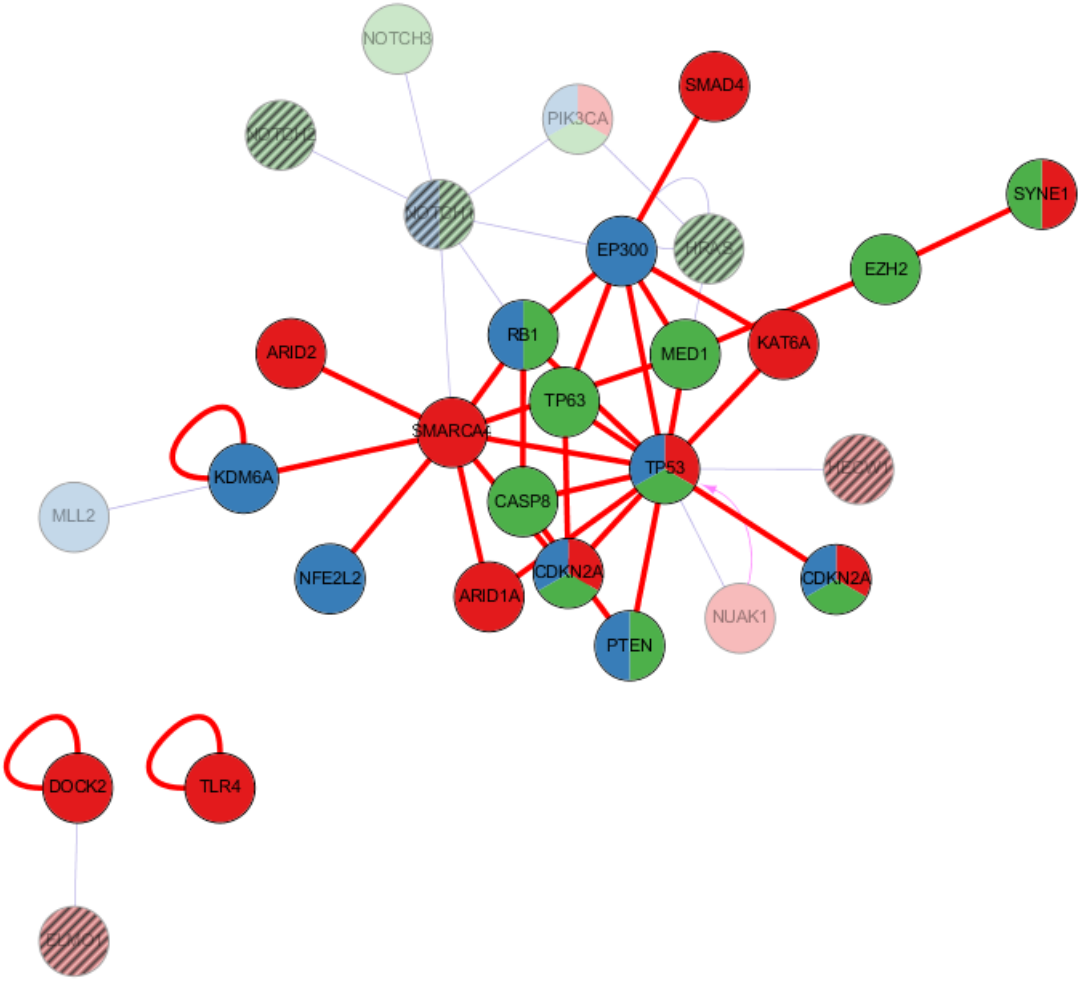
Subcellular Location Menu

Selection of proteins that are expressed in one or more subcellular location(s) can be achieved by accessing the pop-up menu (**Apps → PINA4MS → Identify sub-network by Subcellular Location → highlight proteins with specified location(s) …**) and selecting one or more locations. This will highlight proteins which are labelled to be localized in the subcellular location(s) in either their HPA "Main location" column or "Other location" column. In addition, the edges connecting co-localized proteins are highlighted and thickened.



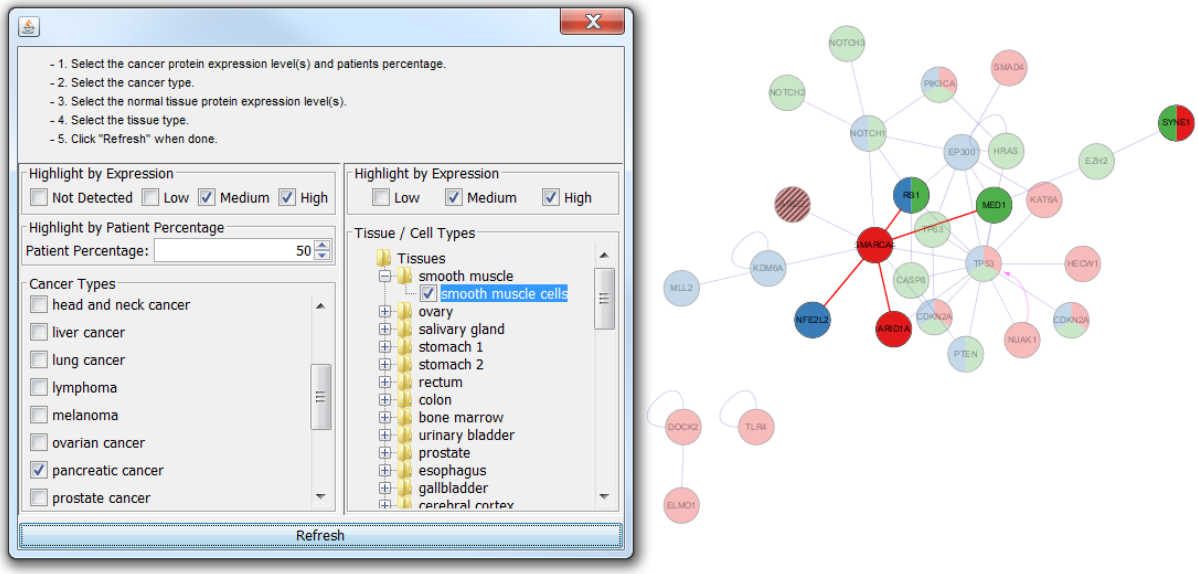
Co-Subcellular Location

The co-subcellular location menu (**App → PINA4MS → Identify sub-network by Subcellular Location → highlight interactions which interacting proteins share the same locations**) highlights *interacting* pairs of proteins that share subcellular location(s), based on their HPA "Main Location" or "Other Location" columns. Note that only proteins that interact are highlighted and the interacting edges are highlighted and thickened.



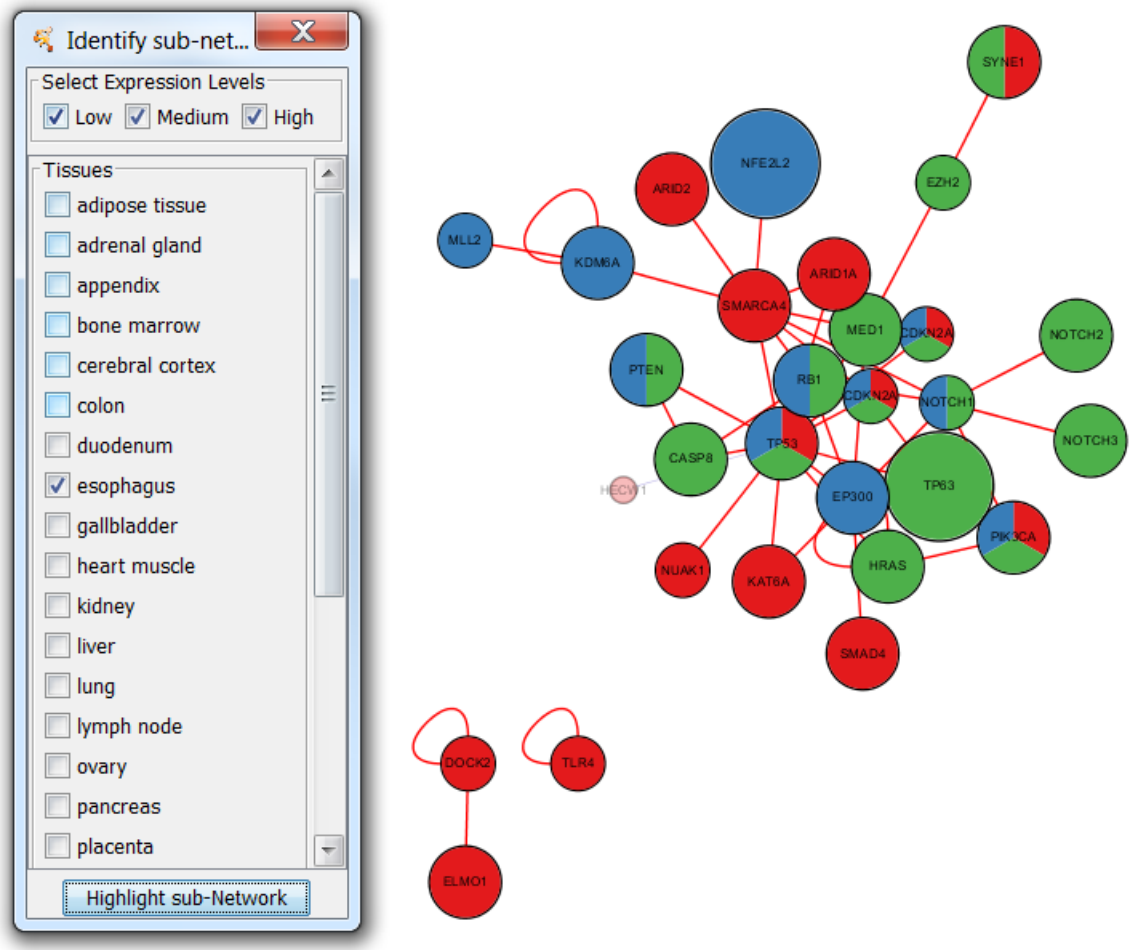
Cancer versus Normal Tissue

Apps → PINA4MS → Identify sub-network by Protein Expression → difference between Human Tumor and Normal Tissue … brings up a pop-up dialog with a side-by-side view of the Cancer Panel and the Tissue Panel. Users can select proteins that are expressed in one cancer type and one normal tissue type, and highlight the proteins with specific expression level(s) in the cancer and tissue.



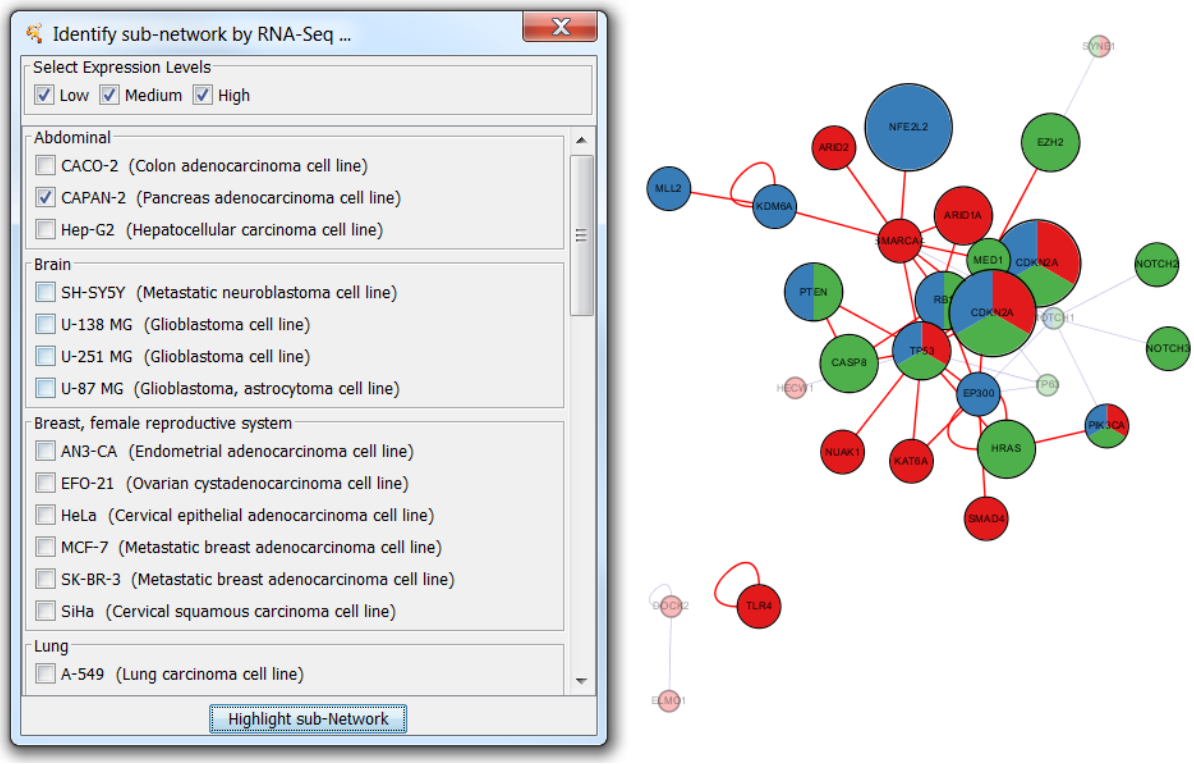
Normal Tissue RNA-Seq Panel

App → PINA4MS → Identify sub-network by mRNA Expression (RNA-Seq) → in Human Normal Tissue … brings up a pop-up panel with a series of check boxes to support quick selection and highlight of proteins based on RNA-sequencing estimate of transcript abundance.



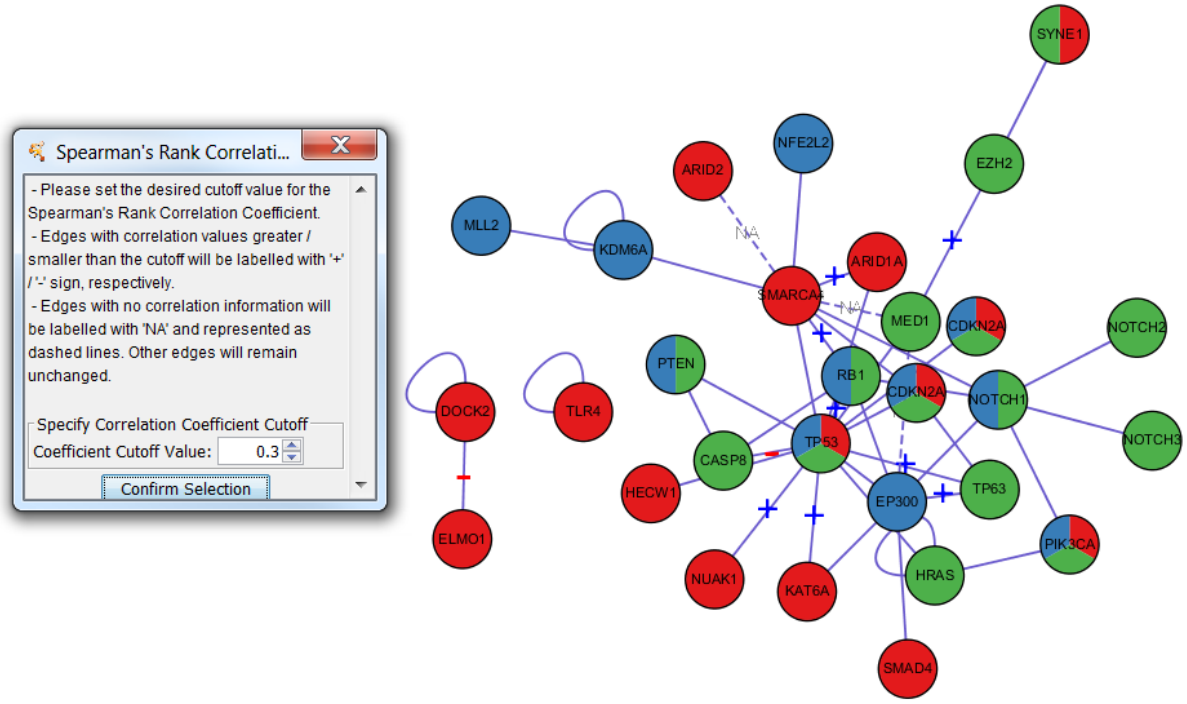
Cancer RNA-Seq Panel

App → PINA4MS → Identify sub-network by mRNA Expression (RNA-Seq) → in Human Cell Line … brings up the Cancer RNA-Seq Panel which contains a series of check boxes for quick selection of proteins based on mRNA transcript abundance in cancer cell lines.



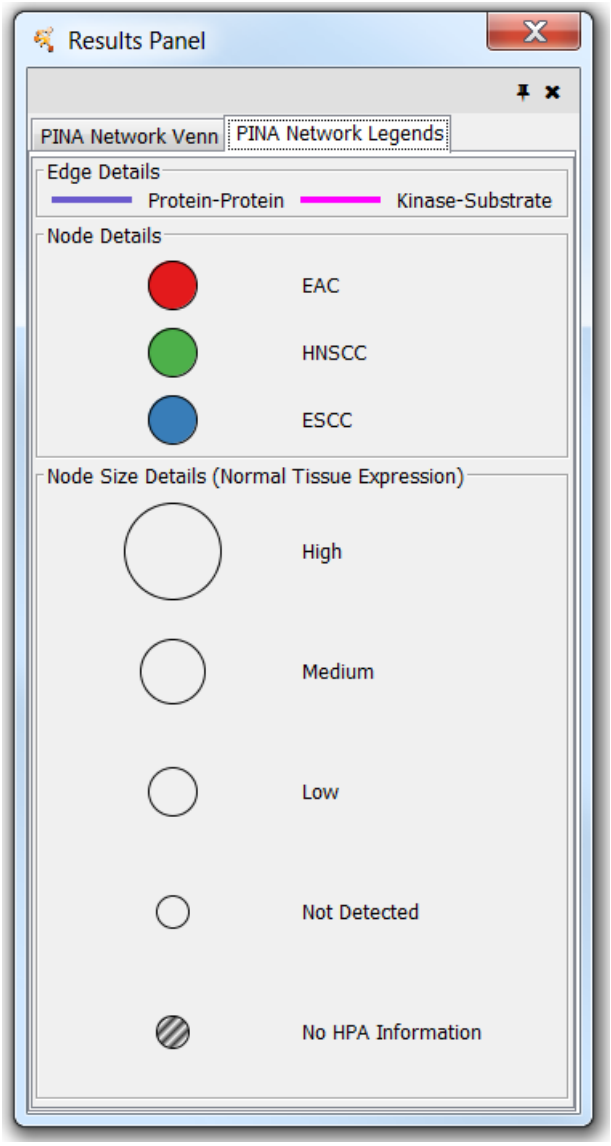
Expression Correlation Analysis

The correlation between the expression profiles of proteins can be access via **App → PINA4MS → Identify sub-network by Expression Correlation**. There are four types of correlation accessible, based on protein expression normal tissue profile; protein expression cancer profile; RNA expression normal tissue profile; and RNA expression cancer cell line profile. Users can set a correlation coefficient cutoff value to limit the selection to significant correlations. For each type of correlation, proteins that are positively / negatively correlated will be labelled with a positive / negative ('**+**' / '**-**') sign, respectively, on their interacting edges. Edges with no correlation information are displayed as dashed lines and labelled as '**NA**', while all remaining edges are unchanged if they do not meet the cutoff value.



Legends Panel

The legends panel is automatically generated on the Results (East) panel. It contains information about the meaning of the edge colours; node colours and input list names; the node sizes and the corresponding protein expression levels.



FAQs

Network creation takes very long (>10 minutes).

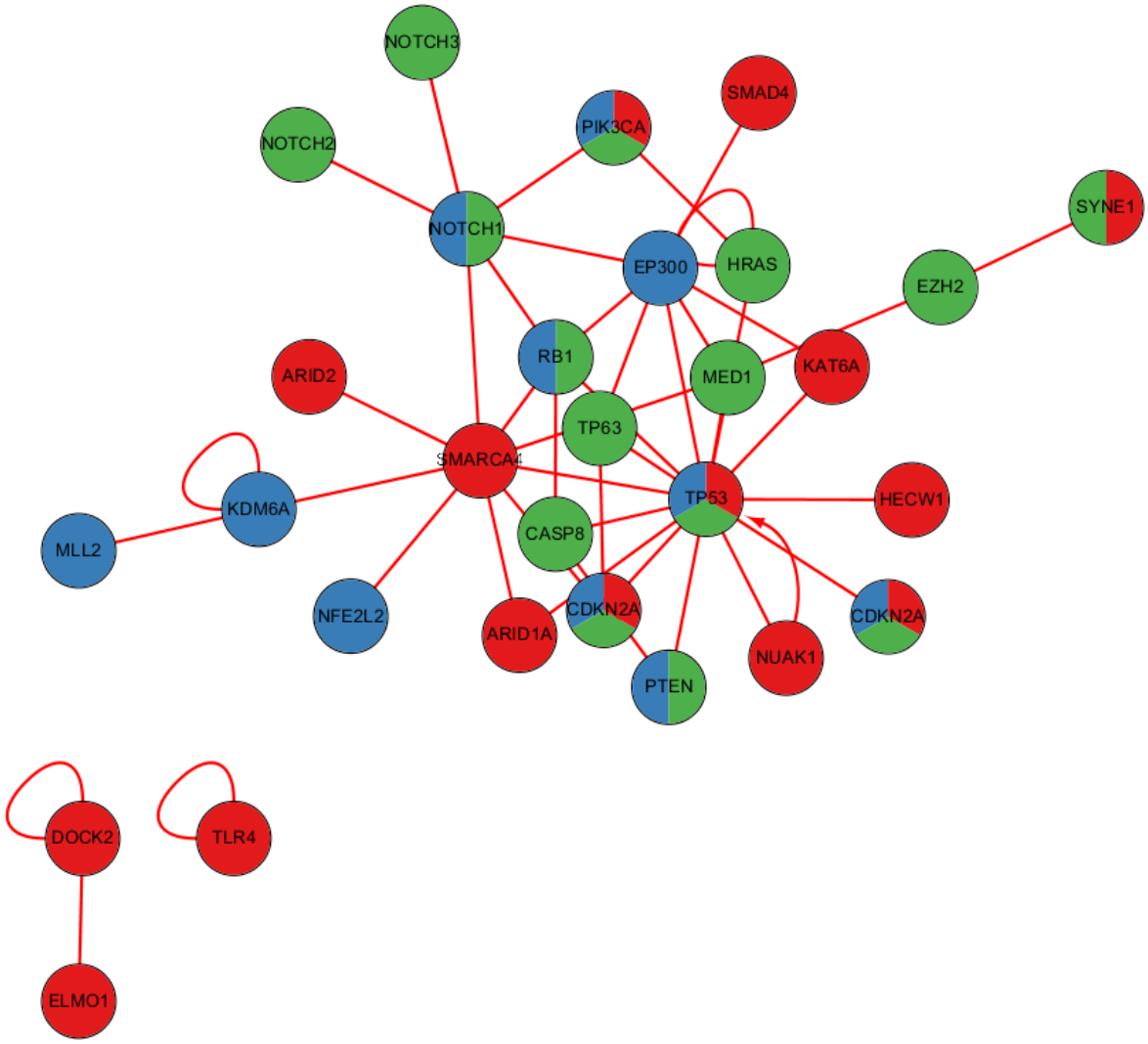
Network creation may take several minutes to fetch information from the PINA server for large networks, for example >1,000 input UniProt ACs, or if the interacting partner search feature is used. For this reason, a network in Cytoscape may not contain a display view yet. In general, most networks we have tested (up to 3,000 input UniProt ACs) took <10 minutes to load. If this is not the case, please restart Cytoscape and try the PINA4MS built-in example (**App → PINA4MS → Create Network … → Example → Search**), which should take <5 minutes to load. Please contact the PINA team (http://cbg.garvan.unsw.edu.au/pina/contact.do) if it takes longer than that to ascertain the web server is functioning properly.

No pie charts are shown after network creation.

This is most likely due to Cytoscape rendering engine's Level of Detail (LOD) settings (http://wiki.cytoscape.org/Cytoscape_3/UserManual/Rendering_Engine). If you want to display every detail of the network, please try full details mode by **View → Show Graphics Details (or CTRL + SHIFT + F on Windows/Linux, Command + SHIFT + F for Mac)**.

Only nodes are shown after network creation, but not edges.

There are two possible scenarios here. One is that the input proteins have no known interactions in the current versions of PINA and PhosphoSite. Alternatively, the PINA web server has an issue and is not returning the proper information. Please try the PINA4MS built-in example (**App → PINA4MS → Create Network … → Example → Search**), which should create a network with nodes and edges similar to the one below:



If this is not the case, please contact the PINA team (http://cbg.garvan.unsw.edu.au/pina/contact.do) to ascertain the web server is functioning properly.