

HOSPITAL FOR SICK CHILDREN & UNIVERSITY OF TORONTO

OrthoNets v1.0

For Cytoscape

**Yanqi Hao, Anna Merkoulovitch, James Vlasblom, Shuye Pu, Andrei L. Turinsky, Denitza Roudeva, Brian Turner,
Jack Greenblatt, Shoshana J. Wodak**

29/11/2010

Table of Contents

Introduction.....	2
1 – Getting and Installing OrthoNets	2
2 – Loading Interaction Networks.....	3
3 – Working with Networks.....	5
3.1 – The Networks	6
3.2 - View Panel	7
3.3 - Visual Options / Legend Panel	8
3.4 – The Information Panel	9
3.5 – The Data Panel	12
4 – Customization	12
4.1 – Basic Customization.....	12
4.1.1 - Custom interaction networks	12
4.1.2 - Custom genes of interest	14
4.1.3 - Overlaying Experimental Data	14
5 – Advanced Options	16
5.1 – Advanced Customization	16
5.2 – Advanced View Options.....	18
6 – Glossary of Terms.....	19
7 – References	20

Introduction

The availability of biological information varies widely for different model organisms. For example, the protein-protein interactome of the yeast *S. cerevisiae* is comparatively well characterized, whereas only a small fraction of the human interactome is known. Conversely, there are thousands of human genes that have been linked to disease, but there is little equivalent information in yeast. The OrthoNets plug-in for Cytoscape enables the simultaneous visualization and navigation of interaction networks of multiple organisms, supporting knowledge transfer from one organism to another. A major difficulty is that protein-protein interaction data are often scattered across multiple databases, necessitating a potentially long and tedious phase of gathering data prior to any analysis. OrthoNets avoids this difficulty by taking advantage of the iRefWeb resource (<http://wodaklab.org/iRefWeb/>), which provides access to a repository of non-redundant interaction data retrieved from ten public interaction databases.

OrthoNets allows interaction networks from two organisms to be viewed simultaneously, and synchronizes the views based on pre-computed orthology relationships between proteins (genes). Each protein is annotated with all aliases available in iRefWeb, OMIM disease annotations, and domain information from PFAM - and every edge is annotated with the number and types of experimental evidence supporting that interaction. Up to four simultaneous views are supported for any two organisms. These views are derived from the protein interaction network for each organism, where the nodes are proteins or genes, respectively. OrthoNets also integrates with iRefWeb to provide filtered views of the interaction networks, and allows the user to quickly center the views on specified genes of interest (for example, genes involved in a particular cellular processes or cellular localization), and infer information between orthologous genes/proteins, their respective environments in the interaction networks and the disease annotations, whenever applicable.

1 – Getting and Installing OrthoNets

Before installing OrthoNets, please verify that your computer has Cytoscape 2.7 or newer installed, available from <http://www.cytoscape.org>. Cytoscape and OrthoNets require Java 5 or newer, available from: <http://www.oracle.com/technetwork/java/javase/downloads/index.html> To install java, click the ‘Download JRE’ button at the above webpage, and follow the installation instructions located at: <http://java.sun.com/javase/6/webnotes/install/index.html>

OrthoNets can require a lot of RAM memory, depending on the size of the interaction networks being visualized and queried. For the included iRef networks, at least 2GB of RAM memory is required, and approximately 100MB of disk space. If OrthoNets still runs very slowly, you may need to increase the amount of memory allocated to java, as detailed at http://cytoscape.wodaklab.org/wiki/How_to_increase_memory_for_Cytoscape

OrthoNets is available via the Cytoscape plugin manager (recommended), and from <http://wodaklab.org/orthonets>, both as a stand-alone package and a “webstart” version. Both packages provide the same features and functionality, although the webstart version requires an internet connection to operate, and is pre-packaged with Cytoscape 2.7. The webstart version is best suited for those who want to

explore OrthoNets without the additional complexity of installing the plugin, whereas the standalone version is most appropriate for those who want to work with OrthoNets offline, or for those who wish to use OrthoNets in tandem with other Cytoscape plugins.

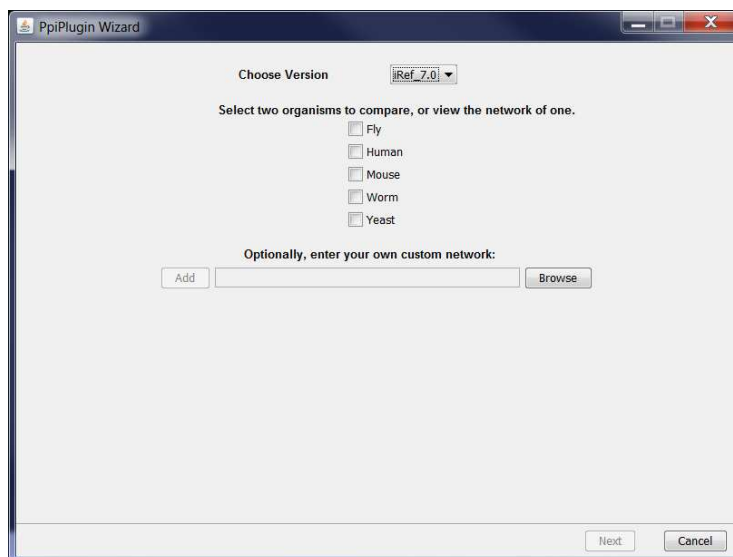
Installing the standalone version via Cytoscape plugin manager (recommended): OrthoNets is available via the Cytoscape plugin manager. To access the plugin manager, launch Cytoscape and click on Plugins -> Manage Plugins. Under the analysis category, select OrthoNets and then click install.

Launching webstart: Navigate to <http://wodaklab.org/orthonets> and click the “Webstart” tab. Depending on the configuration of your browser and installed operating system, you may receive security warnings. These can be safely disregarded. Once Cytoscape starts, please proceed to the Section 2 - “Loading Interaction Networks”.

Alternative standalone installation method: Click the OrthoNets standalone version download link from the OrthoNets webpage, as shown above, and save the .jar file to your Cytoscape plugins directory. By default, the plugins directory is inside the Cytoscape 2.7 installation directory.

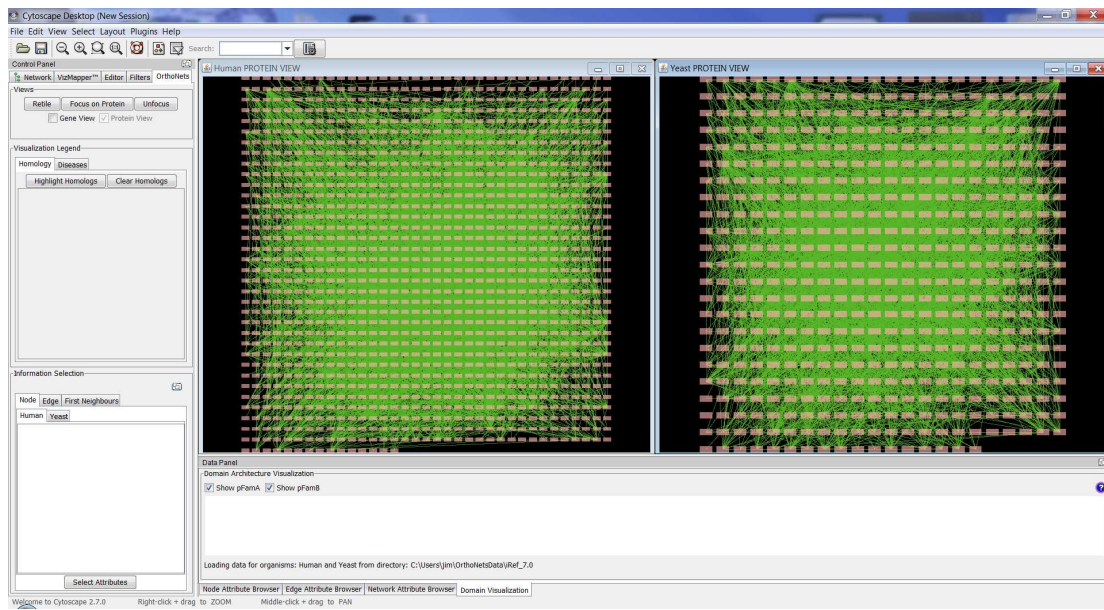
2 – Loading Interaction Networks

From the Plugins menu in Cytoscape, select OrthoNets->Networks->Load Networks. If this is the first time running OrthoNets, a notification will pop-up explaining that additional files need to be downloaded before the plug-in can run. After these files are downloaded, a dialog will prompt you to select the interaction networks you want to work with:

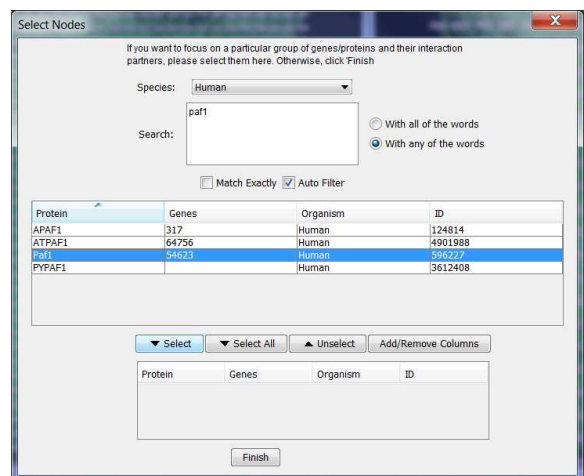
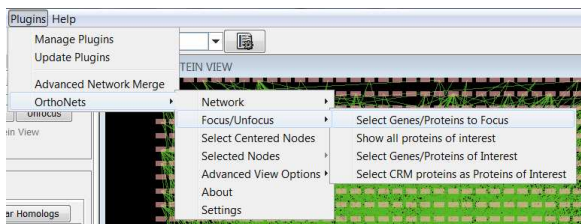


The “Version” drop-down menu lists the iRef database version that will be loaded. Immediately following this drop down menu are a series of 5 checkboxes, corresponding to the five model organisms supported by OrthoNets. Below this are any custom networks available, and add/browse buttons to add additional custom networks. Customization is covered in further detail in Section 3.

After selecting one or two organisms and clicking next, the plugin will begin loading gene/protein annotations and interaction data into memory. Depending on the computer's speed and available RAM, this process can take some time – but typically finishes within 20-30s on modern hardware. After this process completes, the networks corresponding to the selections made in the previous step will become visible. By default, all proteins corresponding to **genes of interest** (see Section 3.2) are shown as nodes, and the interactions between them are shown as edges. If more than one interaction network was selected, both networks will be shown side-by-side in a tiled arrangement:



By default, the **genes of interest** are all those genes known or predicted to be involved in the chromatin modification biological process[1, 2]. For some organisms (e.g. human and yeast, as shown above), this results in large and unwieldy networks, and so the next step is to **focus**, or **center**, on a smaller subset of genes or proteins, via the protein search/selection dialog. This search dialog is automatically activated immediately after loading networks, and can be activated later via the “Select Genes/Proteins to Focus” menu, or by clicking the “Focus on Protein” button in the top left corner of the view panel (see also Section 3.2).

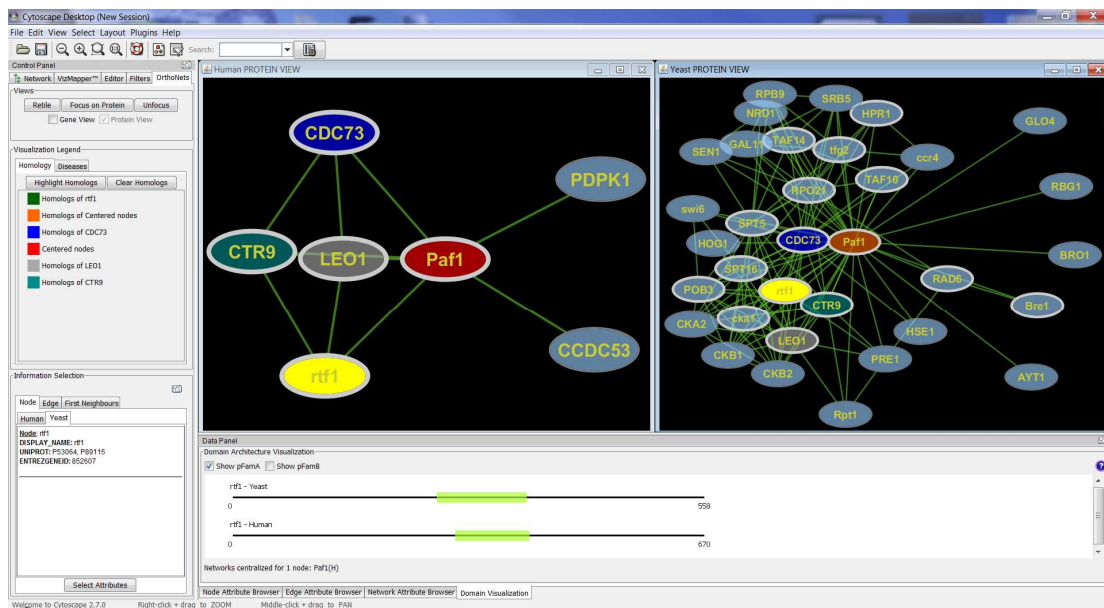


The search dialog “Species” drop-down enables the search to be constrained by species, or to be applied to all species. By default the dialog will return all proteins whose protein name, Entrez Gene ID, or iRefWeb identifier partially match any of the entered search terms. Clicking Add/Remove columns enables searching based on other criteria. For example, in the above dialog the column “**MORBID**” was added, allowing the user to search for genes linked to specific diseases. Search hits are selected by highlighting the row corresponding to the desired protein, and clicking ‘Select’. Clicking ‘Finish’ will then **center** the network views on those selected genes. (To undo this operation, select “Focus/Unfocus” -> “Show all proteins of interest”)

When the ‘Auto Filter’ is toggled off, the search hits are not updated as you type. This is a useful feature if pasting in long lists of numeric ids (e.g. Entrez Gene IDs), and is often used in tandem with the ‘Match Exactly’ toggle to quickly locate a specific list of genes. If the ‘Use AND Search’ toggle is activated, all the search hits must match all the entered terms (although the match can be from any of columns).

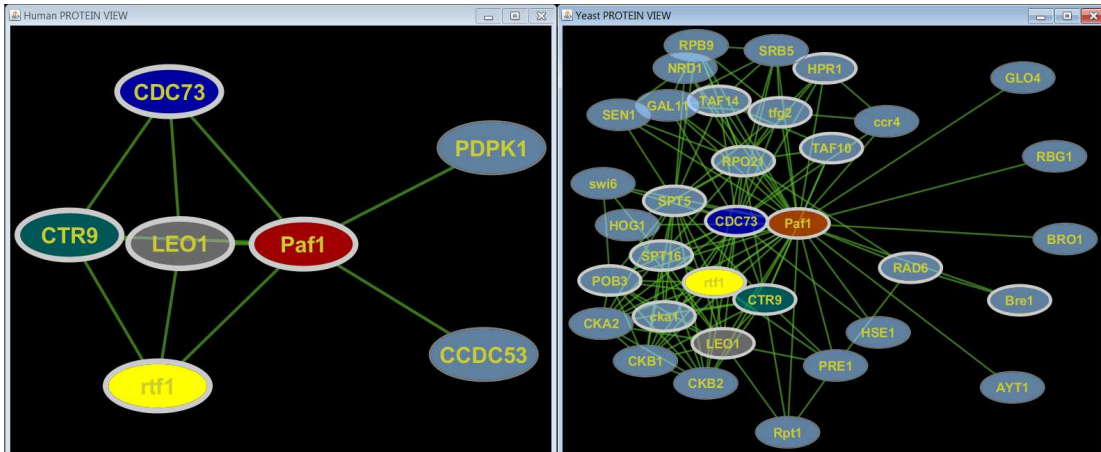
3 – Working with Networks

A typical OrthoNets session is shown below:



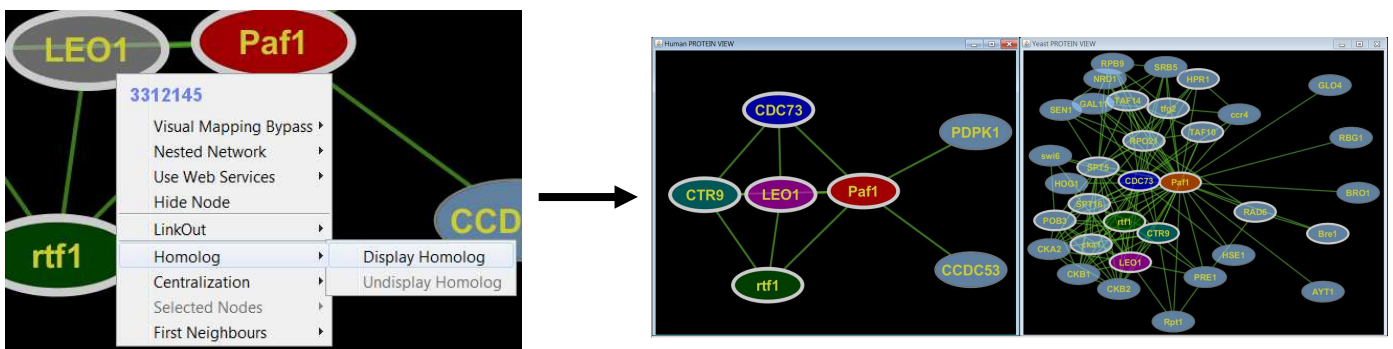
The display can be divided into five main areas - the networks themselves, and the four main panels: the view panel, the Visualization Legend panel, the Information Selection panel, and the customized Cytoscape Data panel.

3.1 – The Networks

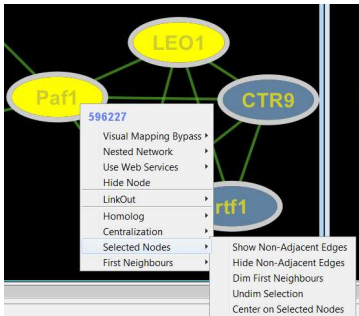


Here the networks for two organisms – Human (left) and yeast (right) have been loaded, and the view **centered** on the human protein Paf1. On the right, this centering operation loaded all interactions involving the centered protein, as well as all interactions involving interactors of Paf1. The **centered** proteins are always represented by red nodes, and **genes of interest** are highlighted with a white border.

On the left, the view is **centered** on all proteins orthologous to yeast Paf1. These nodes are always highlighted in orange. In general, the centering option applies to all specified genes, and all orthologs to these genes. Thus it is possible to enter a list of genes from both organisms, which would result in a mix of red and orange nodes in both of the networks. The left and right networks are synchronized, so that selecting a protein in one network automatically selects the orthologous proteins in the other network. Hovering the mouse over a node displays the node identifier. For the included networks, this identifier is the iRef redundant object identifier (ROG)[3]. Right clicking a node presents several options, in addition to the default options provided by Cytoscape. Selecting ‘Display Homolog’ shows all proteins orthologous to the node, using a color coding noted in the legend panel. Below the node/protein LEO1 in Human has been clicked, and assigned the color purple. The orthologous nodes in Yeast have simultaneously been color coded purple. In some cases, additional proteins can become unhidden if the ‘Display Homolog’ option is selected – if those proteins were not adjacent to the centered protein, and thus not displayed prior to selecting the ‘Display Homolog’ option.



The centralization option provides a convenient shortcut for **centralizing** on a protein of interest, bypassing the usual search dialog covered in the previous section. If one or more nodes are selected (highlighted yellow by clicking them – an operation built into Cytoscape), the “Selected Nodes” menu entry will become active – exposing several options. Note that the same ‘Selected Nodes’ section is accessible via the Plugins->OrthoNets menu.



Hide Non-Adjacent Edges: Shows/Hides all interactions that do not involve one of the selected proteins.

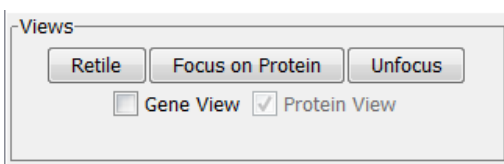
Dim First Neighbours: Reduces the transparency of all nodes connected to one of the selected nodes.

Undim Selection: Restores the default opacity to the selected nodes.

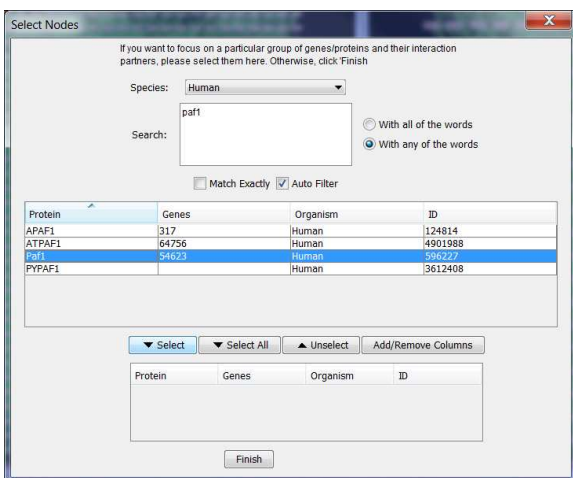
Center on Selected Nodes: Applies the centering operation to the selected nodes, bypassing the search dialog.

The last “First Neighbours” option is a shortcut to the graphing functionality detailed in the information panel Subsection 3.4.

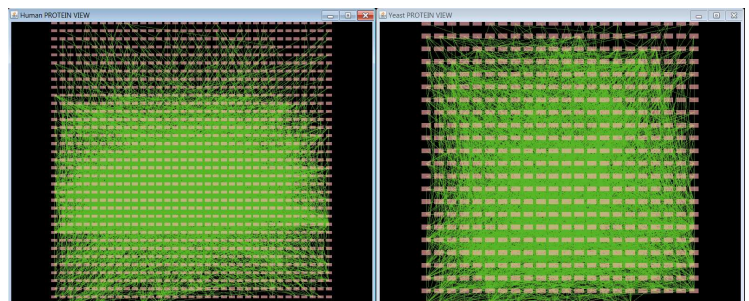
3.2 - View Panel



The view panel contains a ‘Retile’ button, which resets the network windows to their usual side-by-side arrangement, without affecting the displayed layout. This is a useful option if the windows have been manually resized, and it is desired to restore them back to their default layout. The ‘Focus on Protein’ button is a shortcut to the gene/protein search dialog (below left), and the ‘Unfocus’ option shows all the proteins corresponding to the **genes of interest** (below right).

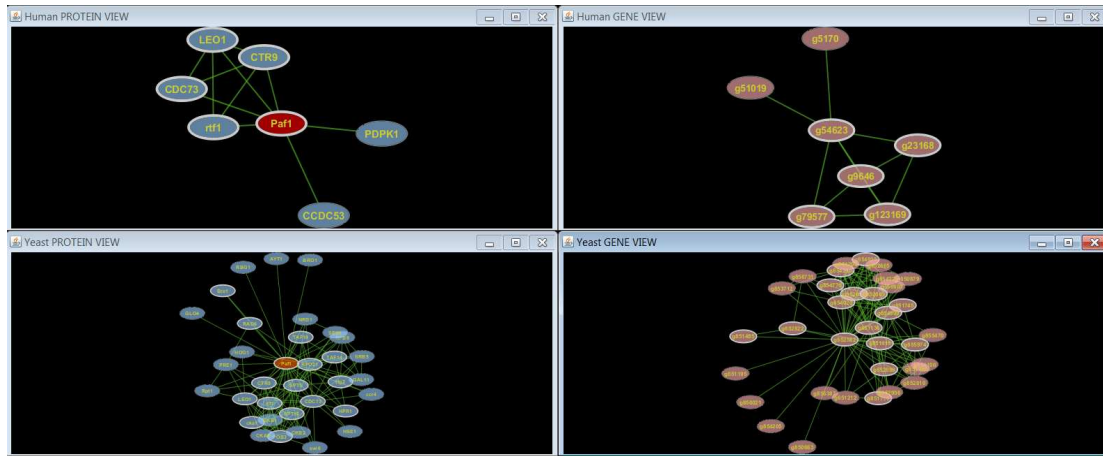


Effect of clicking ‘Focus’



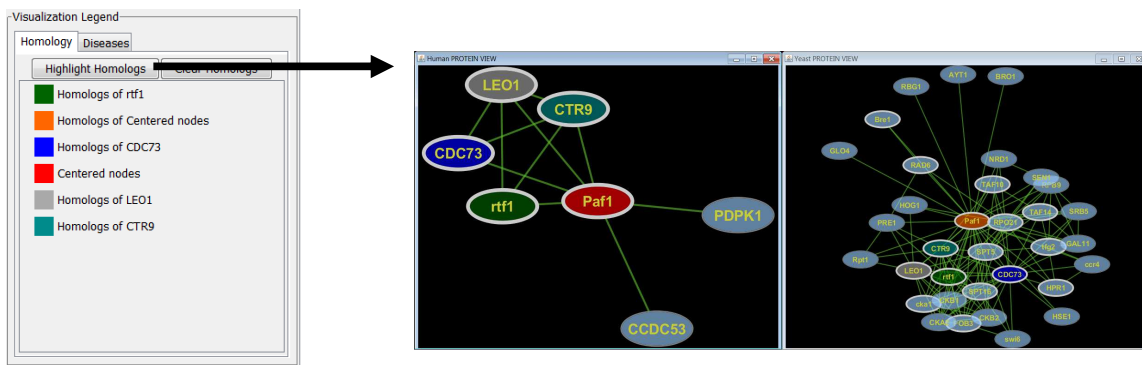
Effect of clicking ‘Unfocus’

The view panel also contains a checkbox toggle to enable the gene view. This view is always shown simultaneously with the default protein view, and hence doubles the number of networks currently visible:



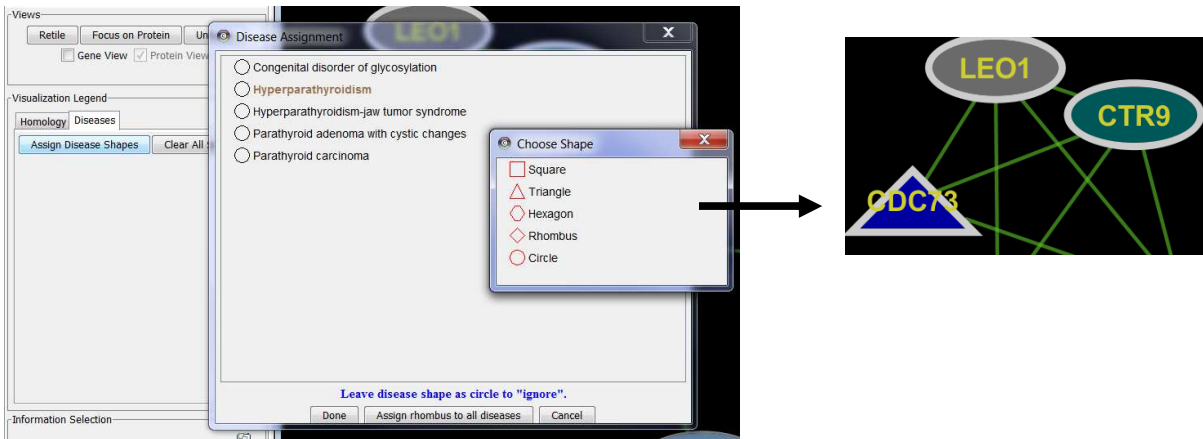
toggling the gene view arranges the protein networks on the left, and the gene networks on the right. If two organisms were selected for visualization, the top two networks will correspond to the first organism, and the bottom two networks to the second organism. In the gene networks each node corresponds to a gene, and a link is drawn between two genes if any of their corresponding protein products interact. In cases where genes code for multiple isoforms, the gene view can be especially convenient for organizing the display. All views remain synchronized by orthology, and selecting a node in any network also selects all the corresponding genes, proteins, and orthologs in the other networks.

3.3 - Visual Options / Legend Panel



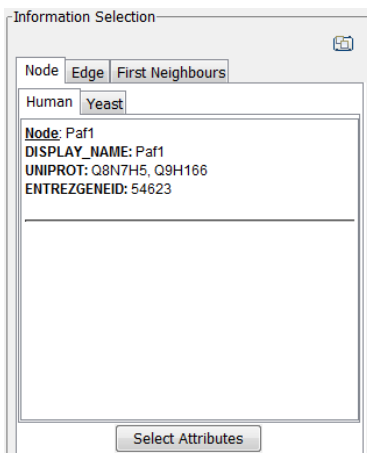
This panel is used to affect the display of network nodes, based on homology or disease information. Under the “Homology” tab, the colours used to identify the centered protein and the orthologous genes/proteins are listed. Clicking a colored square will automatically select all orthologs assigned to that color (or the centered node, if the red square is clicked). When new networks views are created by **centering** on proteins, the homology legend is cleared and no colours are assigned except red – for the centered node, and orange – for the orthologs of the centered node. Colours can be individually chosen for each set of orthologs by right clicking a node of interest and selecting the “Display Homolog” – as discussed in the previous subsection. Alternatively, clicking the “Highlight Homologs” in this panel will assign separate colours to each group of orthologous proteins.

The disease legend is accessible by clicking the “Diseases” tab in the legend panel. Here you can use node shapes to indicate which genes are linked to disease according to **OMIM**. First, click ‘Assign Disease Shapes’ to bring up the disease assignment dialog. This lists all diseases that are linked to any of the proteins currently displayed. Clicking ‘Assign Diamond to all Diseases’ will change the nodes linked to any disease to a diamond shape. Those nodes linked to several diseases will change to octagons, and those without any disease association will remain elliptical. Alternatively, you can select individual diseases from the list and assign them a custom node shape. For example, below, nodes associated with Hyperparathyroidism have been changed to triangles:



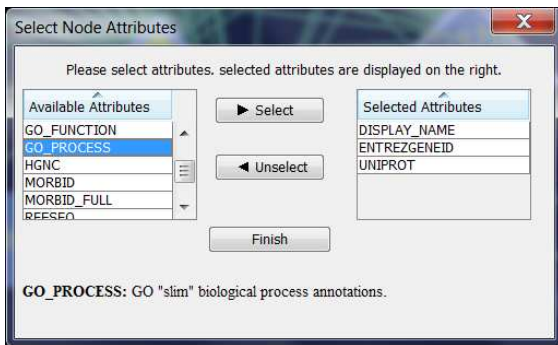
Finally, just like the homology legend, clicking any of the legend entries will automatically select all of the corresponding nodes.

3.4 – The Information Panel

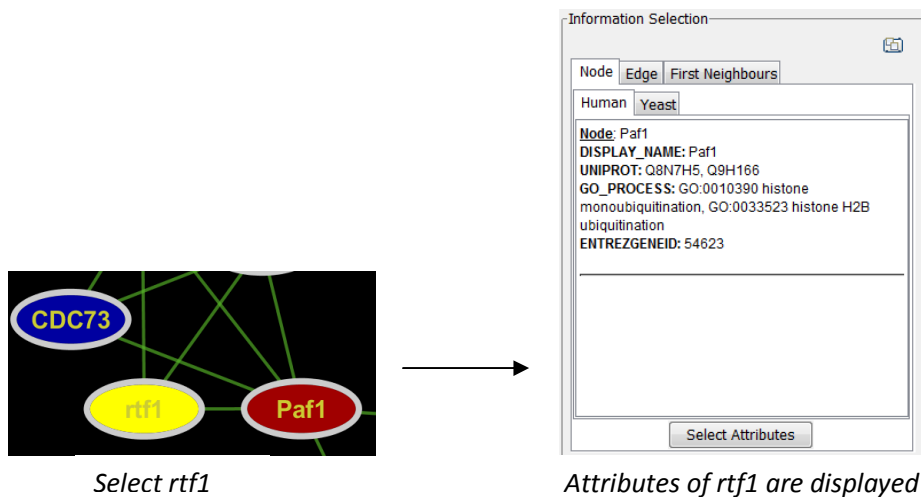


The information panel is primarily used to display node and edge attributes, and largely replaces the Cytoscape node and edge attribute browser. This allows for the simultaneous display of node and edge attributes in the information panel, and protein domain architectures in the data panel. In OrthoNets, node attributes include various gene/protein aliases such as UniProt, Entrez Gene IDs, and NCBI RefSeq; GO annotations; and disease information from **OMIM/MORBID**. Edge attributes record information related to each interaction, such as the pubmed identifiers of supporting publications and the types of experiments used to detect the interaction.

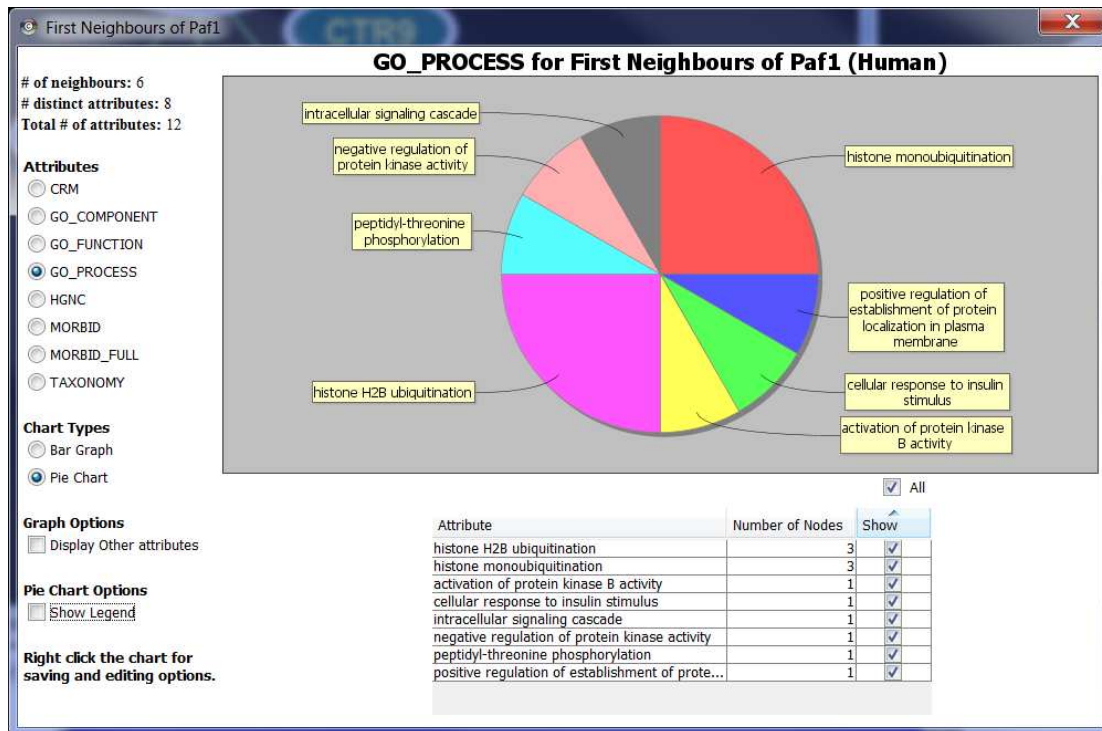
To select the node or edge attributes to display for an organism, click the Node (Edge) tab, the organism's tab, and then 'Select Attributes'. This invokes the Select Node (Edge) Attributes dialog:



Here, the available attributes are listed on the left, and those selected for display are listed on the right. The select and unselect buttons are used to select/unselect attributes for display. After clicking finish, all of the selected attributes will be shown for all selected nodes:



Finally, the 'First Neighbours' tab displays the number of interactors containing the selected protein (i.e. 'first neighbours'), and indicates how many of these have diseases, and how many are **genes of interest**. This tab also has a "Create first neighbour" graph button, that launches the OrthoNets graphing tool:

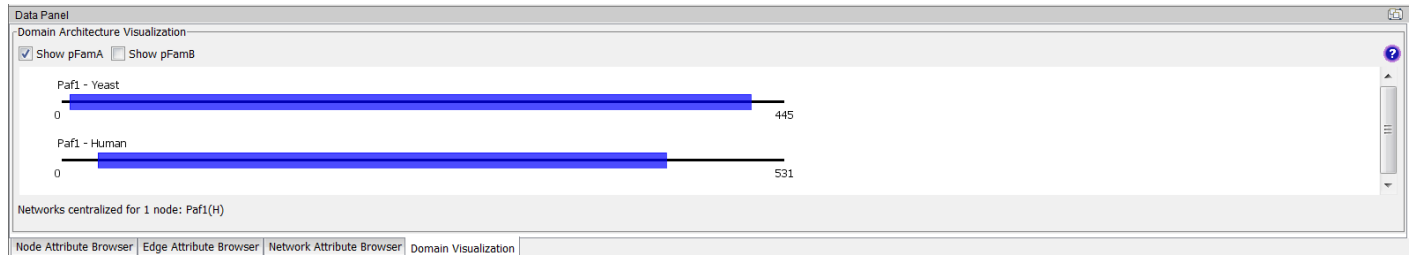


This tool provides a visual summary of the interaction neighbourhood of a selected protein. Above, the gene Paf1 was selected prior to clicking 'Create First Neighbour Graph'. As shown in the upper left of this window, Paf1 participates in 6 distinct interactions (i.e. has 6 neighbours). In the 'Attributes' section of this dialog, the user can select which attribute to be graphed – in this case, the GO Biological Process annotations. The total # of distinct biological process annotations among the first neighbours of Paf1 is given as 8, in the upper left of the window. To reduce display clutter, the OrthoNets graphing tool shows only the top 10 most common attribute values – as also indicated by “# of distinct attributes” in the upper left. Also, those attributes that are not set among any of the neighbouring proteins will not be shown in the Attributes section of this dialog (e.g. if none of the neighbouring proteins are linked to a disease in OMIM, the MORBID and MORBID_FULL options would not be shown). Using the scrollable table at the bottom middle of the dialog, the user can override this behaviour by selecting different attribute values, or by selecting all attribute values.

If some, but not all, of the first neighbours are not assigned an attribute value, then by default these nodes are not considered by the graphing tool. To group these nodes together into an 'Other' category, the user can check the “Display Other Attributes” option.

Various options for saving/exporting the graph, changing the range, and zooming are available by right-clicking the graph.

3.5 – The Data Panel



The data panel is comprised of four tabs: the node attribute browser, edge attribute browser, and network attribute browser are Cytoscape built-in features, and are documented in the Cytoscape manual. These tabs are seldom used in OrthoNets, being largely supplanted by the information panel documented in Subsection 3.3. The fourth tab, and the one visible by default, is the protein domain visualization panel – which shows the **PFam domain architectures** for all currently selected nodes. These diagrams represent a protein as a line, and the domains as colored rectangles sized proportional to their sequence length. Rolling the mouse over these colored rectangles provides additional information about that domain - including the name, PFam identifier, and the amino acid start and end positions of the domain on the peptide. Left clicking a domain automatically launches a web browser with the PFam summary for that domain.

4 – Customization

All of the data used by OrthoNets is customizable. Section 4.1 covers the basic customizations that are built into the OrthoNets graphical user interface (GUI). Some file formatting may be required, but OrthoNets was designed to make the import of this custom data as simple as possible. Section 4.2 covers the more advanced customizations that involve replacing the data files that OrthoNets uses. No special GUI facility is provided for this type of customization, and formatting these files may require more effort than the basic customizations outlined in 4.1. In addition, the advanced customizations will require additional knowledge about how OrthoNets operates internally.

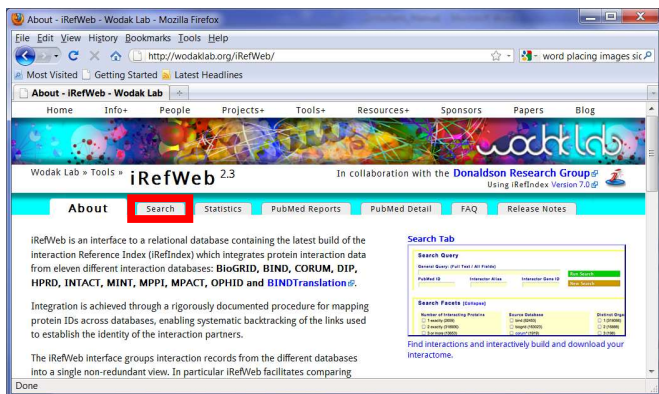
4.1 – Basic Customization

There are three main aspects of OrthoNets that are customizable: Custom interaction networks, **genes of interest**, and the overlay of experimental data.

4.1.1 - Custom interaction networks

To load a custom interaction network, navigate to Plugins->OrthoNets->Networks->Load Network. This will launch the first step of a wizard, as covered in Section 2 of this document. To add a custom network, click 'Browse' and select a custom interaction network file (format covered below), and then click add. This will add the network in the "Custom" section of the dialog.

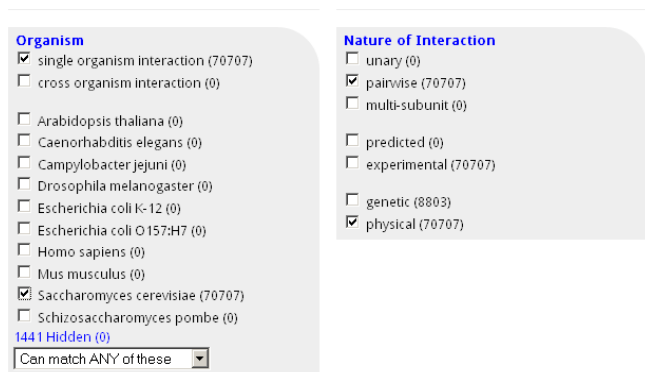
The simplest way to provide custom networks is to use the iRefWeb filtering interface, available at <http://wodaklab.org/iRefWeb/>. This interface allows you to query the iRef database with customized 'filters'. These filters act as restrictions on the returned result set. For example, the user can elect to only consider interactions supported by 2 or more publications, from small-scale experimental methods:



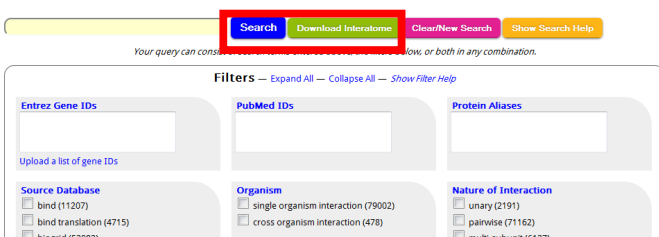
1. Navigate to <http://wodaklab.org/iRefWeb> and select the 'Search' Tab.



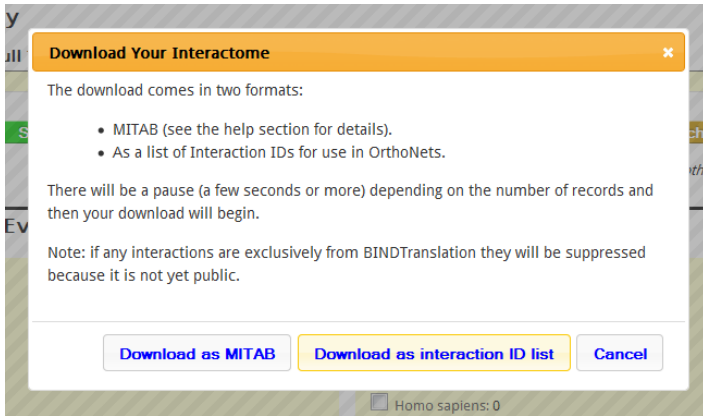
2. Select the 'Expand All' option in the filters section to view the list of available filters.



3. Check off the desired organisms. If more than one organism is checked, then OrthoNets will automatically create separate networks for each organism. At this time OrthoNets only supports pairwise physical interactions within the same organism, so also check the 'physical' and 'pairwise' options under the 'Nature of Interaction' section, and the 'single organism interaction' option under the 'Organism' section. Additional filtering options can also be selected.



4. After checking the desired filters, click 'Search'
5. Once the search is complete, click 'Download Interactome'.



6. Click 'Download as interaction ID List', and save the file for import into OrthoNets.

Alternatively, you can specify your own files using one of the two file formats that are supported by OrthoNets for custom networks. The first is a newline separated list of iRef interaction identifiers (RIGs), prefaced by the header '# iref_interaction_id' (without quotes). For example:

```
# iref_interaction_id
14
25
...
```

The second file format is a newline separated list of tab-delimited pairs of Entrez Gene identifiers, prefaced by the header '# entrezgeneid' (without quotes). For example:

```
# entrezgeneid
855710<tab>854063
854663<tab>854063
856885<tab>854063
...
```

For the second file format, OrthoNets will assume interactions between all isoforms corresponding to these identifiers (i.e. for a given pair of entrez gene identifiers, the assumed interactions are the Cartesian product of the two sets of isoforms).

4.1.2 - Custom genes of interest

By default, OrthoNets identifies those genes known or predicted to be involved in chromatin modification as **genes of interest**. These genes and proteins corresponding to these genes will always be highlighted with a white border within the OrthoNets network views. In addition, several of the graphing options and "Advanced View Options" documented in Section 3 provide customized views based on these genes of interest, and their interaction neighbourhood. Genes/proteins of interest can be specified using the Plugins -> OrthoNets -> Select Genes/Proteins of Interest option. This raises a dialog very similar to the one used for **centering** – see Section 2 for details.

4.1.3 - Overlaying Experimental Data

In OrthoNets, a user-specified network can be overlaid on an existing network view. This can be used to visually query which interactions in the user network have been previously reported in the literature - as

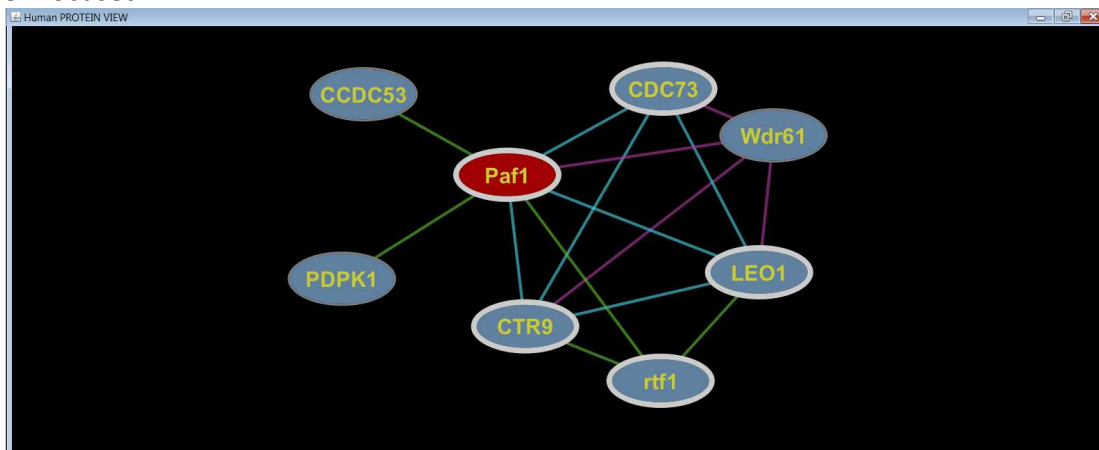
recorded in the iRef database – and which are novel. An experiment file is formatted as a tab delimited list of entrez gene identifiers, gene names, and several optional data columns prefaced by a header. For example, the following example is from an affinity purification experiment:

```
Bait<tab>Protein<tab>BaitName<tab>PreyName<tab>SpectralCount
10428<tab>10856<tab>CFDP1<tab>RUVBL2<tab>12.51
8930<tab>3853<tab>MBD4<tab>KRT6A<tab>50.434
...
```

Here the 'SpectralCount' is a score associated with each interaction. This score will be loaded as an edge attribute, and can be selected in the information panel (see Subsection 3.4) for display. Note that only the first two columns are required – but providing gene names is recommended. To load an experimental file such as the above, select OrthoNets->Plugins->Network->Overlay Edges. This will overlay the interaction file, but will not change which proteins are **centered**. Therefore, overlaying data is usually a two step process. First, choose the file to overlay, and second, choose proteins to center on via the "Plugins->OrthoNets->Focus/Unfocus ->Select genes/proteins to focus" dialog.

The overlaid data is color coded based on whether the user-specified interactions exist in the base network or not. Here the base network is either one of the included iRef networks from Mouse, Human, Worm, Fly, or Yeast – or a user-specified custom network (see "Custom Interaction Networks" subsection above). Green indicates those edges recorded in the base network that are not present in the overlaid data, whereas purple indicates those edges in the overlaid data that are not in the base network. Cyan edges are those present in both. In addition, OrthoNets highlights in gold those edges present in the overlaid data that are not in the base network, but whose interologs are present in any of the other four organisms.

In the below example, experimental TAP/MS data is overlaid on the human network, and the human protein Paf1 has been focused:



5 – Advanced Options

5.1 – Advanced Customization

Internally, the primary node entity is a protein, which is uniquely identified by a numeric iRef identifier (ROG). Whenever the plugin accepts alternative identifiers – such as Entrez Gene identifiers, it first automatically maps them to ROGs. All node and edge attributes are assigned to ROGs or pairs of ROGs. The first time the plugin is run, it downloads all data to a ppiPluginData subdirectory under the users home directory. This creates the following files and subdirectories:

```
iRef_x.x/attributes/*.na  
iRef_x.x/attributes/*.ea  
iRef_x.x/organisms/Fly|Yeast|Human|Worm|Mouse/*.sif  
iRef_x.x/orthology/orthologs.omf  
iRef_x.x/pda/domain_arch.pda
```

These files are covered below.

iRef_x.x/attributes/*.na

These are the node attribute files, containing gene/protein aliases and other information extracted from iRef. By default the plugin loads all node attribute files found in this directory. Thus to add customized node attributes, create a formatted node attribute file and copy it to this location. Consider the example node attribute file REFSEQ.na:

```
REFSEQ (class = java.lang.String)  
1657917=(31560057::75788::NP_001033716::NP_001033716.1::NP_083714::NP_0837  
14.3)  
5256241=(6679355::NP_032886::NP_032886.2)  
3068764=(NP_035707)
```

This is the standard file format for node attributes as covered in the Cytoscape manual (<http://www.cytoscape.org/documentation/users.html>). The first line is the name of the attribute, which should match the name of the file minus the .na extension. This is followed in parentheses by the class of the attribute, which can be any type supported by Cytoscape – including java.lang.String, java.lang.Integer, or java.lang.Double. In OrthoNets, most attributes are list attributes, meaning that each protein may be assigned multiple values for that attribute. In this case, the protein with ROG 1657917 is assigned the NCBI RefSeq aliases 31560057, 75788, NP_001033716, etc. A list is specified as a :: delimited set of values, enclosed in parentheses. In general, the format for a node attribute is:

```
ATTRIBUTE_NAME (class)  
ROG=(value::value::value...)  
...
```

iRef_x.x/attributes/*.ea

These are the edge attribute files, which store information about the publications and experiment types supporting protein-protein interactions. The .ea file format is the standard edge attribute format documented

in the Cytoscape manual. OrthoNets loads all edge attributes contained in this directory automatically. Consider the example PUBMED.ea edge attribute file:

```
PUBMED (class=java.lang.String)
2255804 (pp) 4772972=(14592989)
2761514 (pp) 4705328=(11805837)
3083028 (pp) 4629183=(18467557)
139374 (pp) 5124145=(11283351::10655498)
```

The file begins with the name of the attribute followed by the class name in parentheses (see *.na description above). An edge is identified by a pair of ROGs, separated by (pp). Most edge attributes are also lists, specified as a :: delimited set of values enclosed in paranthesis. For example, for the above, the edge connecting the protein identified by ROG 139374 to the protein identified by ROG 5124145 is supported by the publications with pubmed identifiers 11283351 and 10655498. The format for an edge attribute is:

```
ATTRIBUTE_NAME (class)
ROG (pp) ROG = (value::value::value...)
...
```

iRef_x.x/organisms/Fly|Yeast|Human|Worm|Mouse/*.sif

These are the interaction files themselves for each of the five organisms, identified in sif format. Each line consists of a pair ROGs, delimited by (pp):

```
ROG (pp) ROG
ROG (pp) ROG
...
```

Usually these files should not be modified, as custom interaction networks can be added via the GUI facility documented in Subsection 4.1.

iRef_x.x/orthology/orthologs.omf

This file lists all pairs of orthologs between all organisms. Each line is consists of a pair of ROGS, delimited by tab:

```
ROG<tab>ROG
ROG<tab>ROG
...
```

iRef_x.x/pda/domain_arch.pda

The domain architectures for each protein are contained in this file, one domain per line. For example:

```
1000673 2140 DUF3398 PFamA PF11878 Domain of unknown function (DUF3398) protein 65 159
1000673 2140 Ded_cyto PFamA PF06920 Dedicator of cytokinesis protein 1930 2107
1001137 752 KA1 PFamA PF02149 Kinase associated domain 1 protein 706 752
```

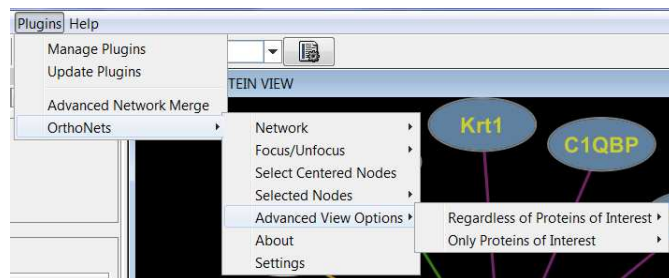
The file is contains eight columns, delimited by tabs. In order, these are:

1. iRef ROG identifier of the protein.
2. Length of the protein.

3. Short name for the domain.
4. One of either 'PFamA' or 'PFamB'.
5. PFam Identifier
6. Short Description of Domain
7. Amino acid start position of the domain
8. Amino acid end position of the domain.

5.2 – Advanced View Options

The advanced view options apply to the entire interaction network – not just the currently centered view. Advanced viewing options can be accessed from Plugins->OrthoNets->Advanced View Options:



Regardless of Proteins of Interest:

Only Diseases	All nodes with at least one disease (MORBID) annotation
Only Diseases and First Neighbours	In addition to above, also show all proteins that interact with a disease-associated protein.
Only Diseases and First Neighbours and edges incident with Proteins of Interest	As above, but only interactions involving a protein of interest are shown.

9.

Only Proteins of Interest:

Only Proteins of Interest	Only show Proteins of Interest
Proteins of Interest with Diseases	Only show those Proteins of Interest with disease (MORBID) annotations
Proteins of Interest with Diseases and First Neighbours	In addition to the above, also show all proteins that interact with a disease-associated Protein of Interest
Proteins of Interest with Diseases and First Neighbours, and edges incident with Proteins of Interest	As above, but only interactions involving a Protein of Interest are shown.
ALL Proteins of Interest, and First Neighbours with diseases, and edges incident with Proteins of Interest	Shows all proteins of interest, regardless of disease association, and all disease-associated proteins interacting with these proteins of interest. Only those

Proteins of Interest with diseases, and first neighbours with diseases, and edges incident with Proteins of Interest	edges incident with a protein of interest are shown. Shows all disease-annotated proteins of interest, and all disease-associated proteins interacting with these proteins of interest. Once again, only those interactions involving a protein of interest are shown.
---	--

6 – Glossary of Terms

Term	Description
Center (Centralize)	Centering is an operation applied to a group of proteins, which hides all proteins not interacting with a centered protein, or an ortholog of a centered protein (if two networks are loaded and visible). Centered proteins are colored as red nodes, and if two networks are loaded, orthologs of centered proteins are colored orange.
DAnCER	The Disease Annotated Chromatin Epigenetics Resource[1].
Focus	See ‘Center’
Genes of Interest	A user specified list of genes whose corresponding proteins (see <i>Proteins of Interest</i>) are always highlighted.
Homolog	For OrthoNets, the term homolog is used synonymously with ortholog. See also ‘Ortholog’.
Interolog	An interaction between two proteins is said to be an interolog if orthologs of those proteins interact.
MORBID	The Online Mendelian Inheritance in Man “MORBID” map is the source of disease-gene associations used by OrthoNets. See also http://www.ncbi.nlm.nih.gov/omim
OMIM	In OrthoNets, this is synonymous with MORBID.
Ortholog	Two genes are orthologs if they derive from a common ancestor.
Overlay Network	The operation of superimposing a user-specified network on a network derived from iRef, or on another user-specified network (see Subsection 4.1.1). Edges that are in common or unique to each network are highlighted in different colors (Subsection 4.1.3).
PFam	The source of domain architecture information used by OrthoNets. See also http://pfam.sanger.ac.uk/ for more information.
Proteins of Interest	Proteins of interest are always highlighted with a white border. By default, the proteins of interest are those known or predicted to be involved in chromatin modification[4], but this can be overridden using the interface documented in Subsection 4.1.2. Proteins of interest are also treated as a separate category in the information panel (see Section 3), and affect certain advanced display options.
RIG	The numeric identifier assigned by IREF to every unique protein interaction stored in the database. Seldom used in OrthoNets, except to specify custom networks (see Subsection 4.1)

ROG

The numeric identifier assigned by IREF to every unique protein stored in the database. ROGs are the primary identifier used by OrthoNets, but are only used internally. Except for certain advanced customizations, the user can elect to specify their data using the more common Entrez Gene identifier.

7 – References

1. Turinsky, A.L., et al., *DAnCER: Disease-Annotated Chromatin Epigenetics Resource*. Nucleic Acids Res.
2. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
3. Razick, S., G. Magklaras, and I.M. Donaldson, *iRefIndex: a consolidated protein interaction database with provenance*. BMC Bioinformatics, 2008. **9**: p. 405.
4. Pu, S., et al., *Expanding the landscape of chromatin modification (CM)-related functional domains and genes in human*. PLoS One. **5**(11): p. e14122.