

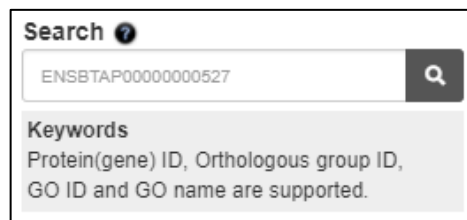
INTERESPIA tutorial for using the visualization interface

1. Controlling options

(1) Searching a protein or an orthologous protein group

User can search for the specific protein (node) using the genes or protein ID, orthologous protein group ID, GO ID or GO name.

This search supports not only the “Ensembl protein” but also “UniprotKB”, “EntrezGene”, “Gene symbol” and “Ensembl gene”.



Search ?

ENSBTAP00000000527

Keywords
Protein(gene) ID, Orthologous group ID,
GO ID and GO name are supported.

(2) Selecting layouts

(2-1) Grouping by PPIs

User can set the network layout based on protein-protein interactions.

This format is the default setting and interacting nodes are grouped together.

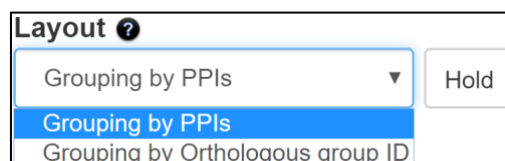
(2-2) Grouping by Orthologous group IDs

User can set the network layout based on orthologous protein relationships.

This format allows user to better identify interaction patterns among different orthologous protein groups.

(2-3) Hold button

When user clicks the “Hold” button, all nodes are fixed in their current position. If user changes the layout or interaction level, the “Hold” button is automatically turned off.



Layout ?

Grouping by PPIs ▼

Grouping by PPIs

Grouping by Orthologous group ID

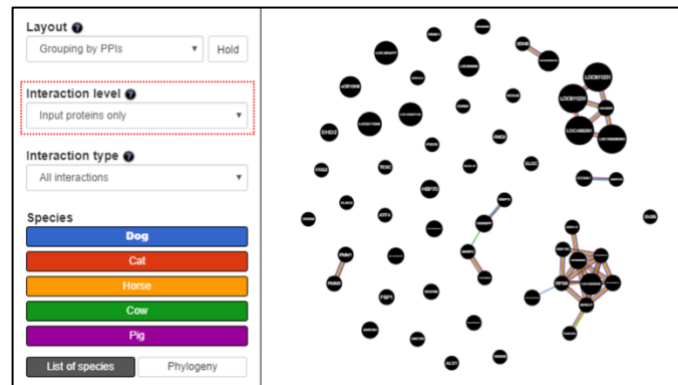
Hold

(3) Handling interaction level

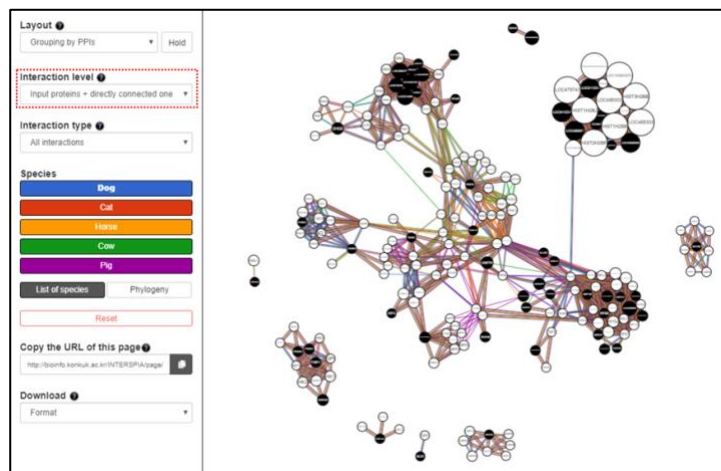
The interaction level can be adjusted according to the type of nodes based on the result of the protein extension process.

If the protein extension process is not used, there is only “Input proteins only” in the dropdown box.

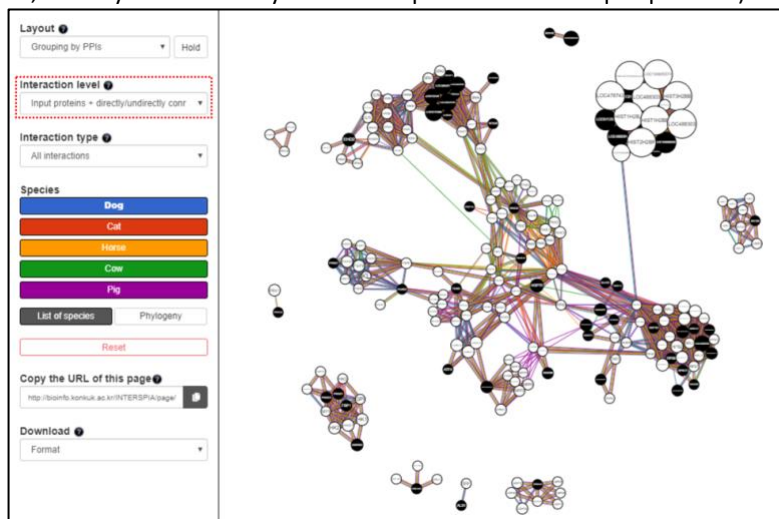
- Input proteins only: Only showing input proteins provided by user



- Input proteins + directly connected ones: Showing both input proteins provided by user and only directly connected proteins to the input proteins. The directly connected proteins are obtained by the protein extension process.



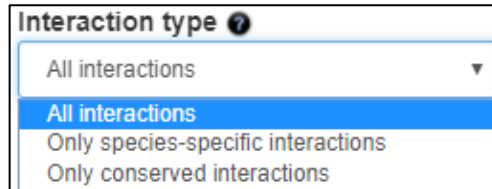
- Input proteins + directly/indirectly connected ones: Showing all proteins (input proteins provided by user, directly and indirectly connected proteins to the input proteins). The directly/indirectly



connected proteins are obtained by the protein extension process.

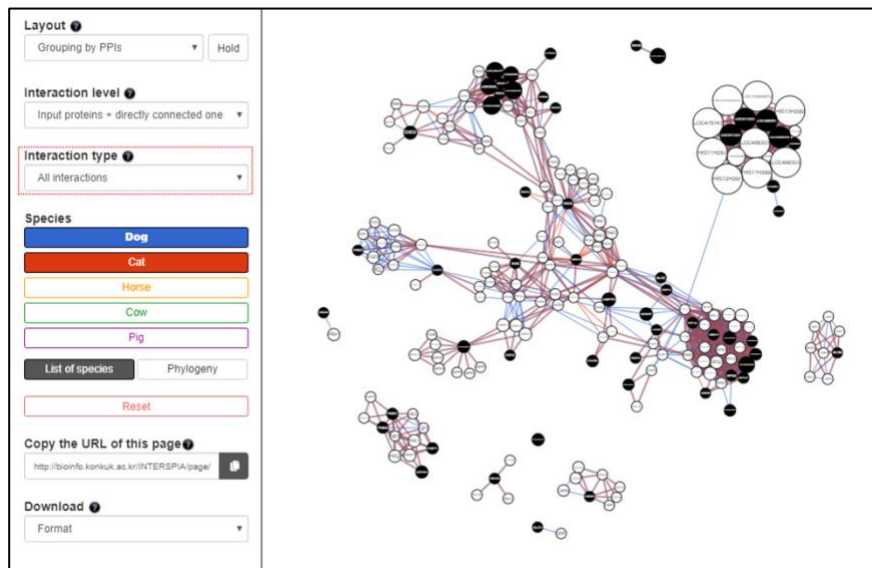
When user changes the “Interaction level”, the numbers in the “Interactions” panel (orthologous protein groups, proteins and interactions) are updated on the fly.

(4) Visualization setting for protein-protein interaction (PPI) type



- All interactions

Showing PPIs observed in at least one of the selected species.



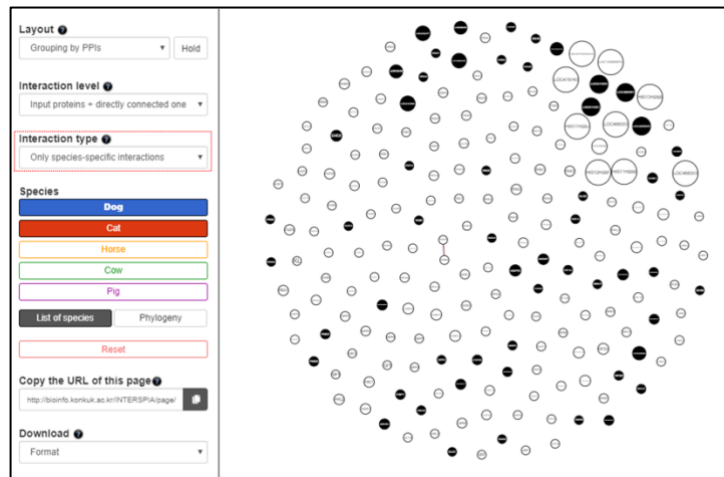
- Only species-specific interactions

Showing PPIs common in the selected species but not observed in the unselected species.

For example, in the figure below, dog and cat are selected among five species (dog, cat, horse, cow and pig). And the “Only species-specific interactions” is selected (in red dotted box).

In this case, the PPIs visualized in the network are the ones existing in both dog and cat, but not existing in the other species horse, cow and pig.

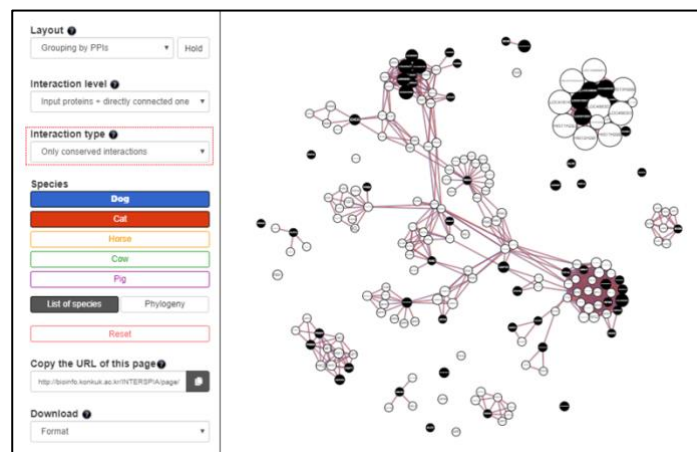
When the “Only species-specific interactions” is selected, it is not allowed to select all species.



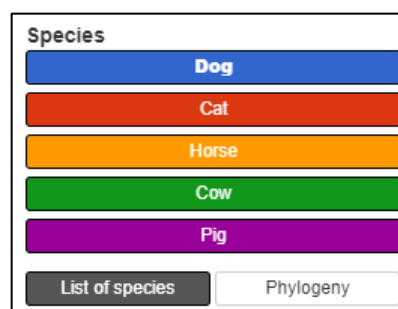
- Only conserved interactions

Showing PPIs common in the selected species. In this case, the absence of PPIs in the unselected species is not checked.

For example, in the figure below, the PPIs shown in the network are the ones existing in both dog and cat, regardless of the existence of the PPIs in the unselected species (horse, cow, and pig).



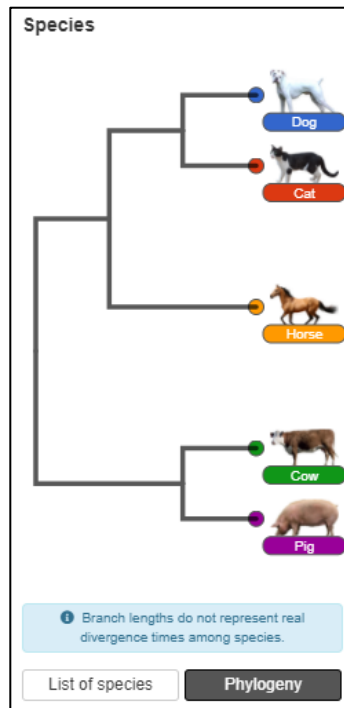
- List of species: Showing the list of the species.



The top-most species in the interactions panel is the target species.

The color of each species matches with the colors of PPIs (edges).

- Phylogeny: Showing the phylogenetic tree of the species. The branches are highlighted (thickened) when species are selected.



When user changes the “Interaction type” and “Species”, the number of interactions in the “Interactions” panel is updated on the fly.

(5) Going back to the initial state

When clicking the “Reset” button, the network is refreshed to the initial state.

(6) Copying the URL of an analysis page

When clicking the “Copy the URL of this page” button, the URL of the analysis page is copied.

This URL is only valid for 14 days.

If user revisits INTERSPIA using this URL, they can restart their analysis from the final session they copied.



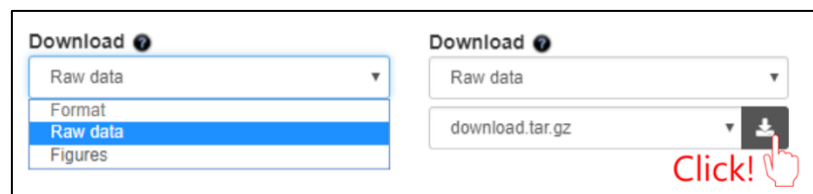
(7) Downloading figures and data

(9-1) Raw data

User can download raw data used in our web interface. The filename is “download.tar.gz” which

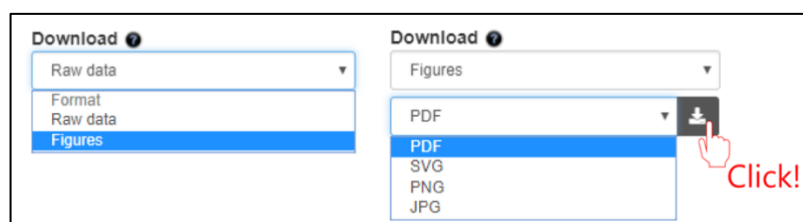
containing the following files.

- Analysis_results_url.txt: The URL of the analysis page
- [Target species].INTERSPIA.PPI.network.txt: This file contains the results of the analysis. The first and second column indicate the orthologous protein group ID, the third and fourth column show the protein ID of the orthologous groups. From the fifth to the last column indicate the mean value of the STRING confidence scores of PPIs across all orthologous protein pairs in species.
- [Target species].unused.input.protein.list.txt: The first column indicates the unused protein IDs among the user's input genes or proteins, and the second column indicates their gene symbols.
- [Target species].used.input.protein.list.txt: The used protein IDs among the user's input genes or proteins
- significant_go.txt: This file contains the results of the GO enrichment test. The first column indicates the GO ID, the second column shows GO domain (BP: Biological Process, MF: Molecular Function, CC: Cellular Component), the third column shows the GO name, and the fourth column represents the significant score represented by Q value.
- id_map.txt: This file contains the ID mapping results based on the information of UniProt (<http://www.uniprot.org/>). Columns represent gene symbol, Uniprot accession ID, EntrezGene ID, Ensembl gene ID, Ensembl protein ID of a protein in order.
- README: This file includes the parameter information related to analysis.



(9-2) Figures

User can download a figure by following formats: PDF, SVG, PNG and JPG.



2. Handling network

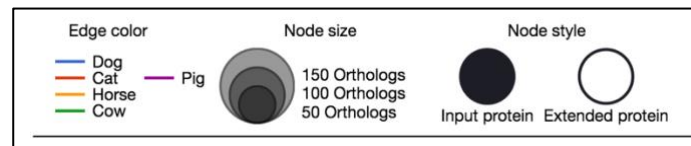
(1) Handling nodes

The black nodes indicate the user's input proteins, and the white nodes represent the extended proteins by the protein extension process.

The size of nodes shows the total number of orthologous proteins in other selected species.

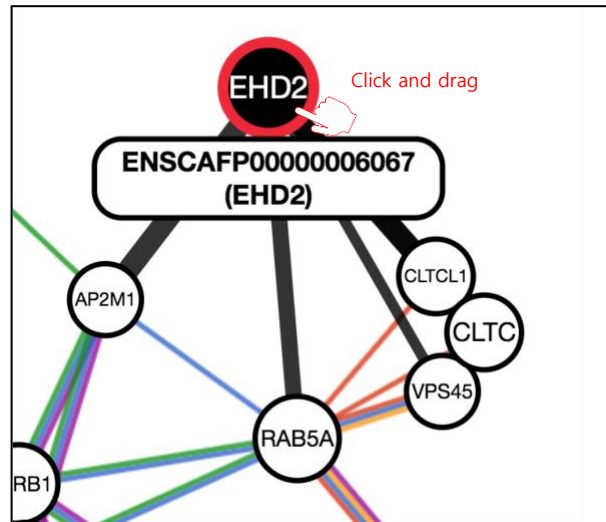
A gene symbol is shown inside a node. If a gene symbol is not available, the Ensembl protein ID is used

instead.



(1-1) Handling one of group of nodes

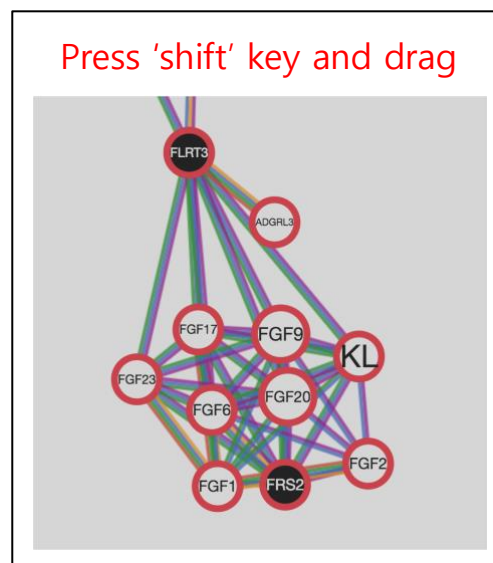
If a node is dragged by a mouse, it is sticking to a moved position.



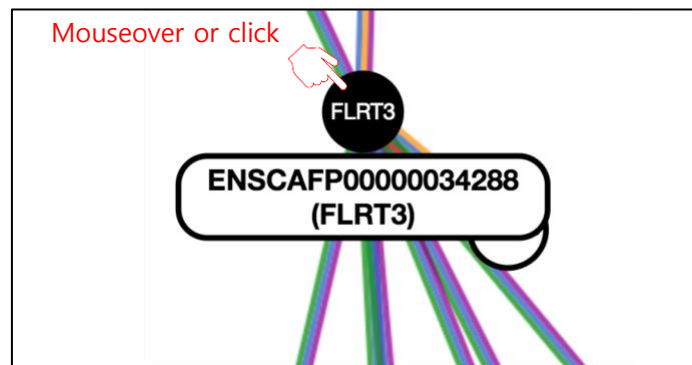
User can select multiple nodes by holding down the shift key and dragging. The selected nodes can be moved together by a mouse click and drag.

When a mouse pointer is on a node, the Ensembl protein ID and a gene symbol (if available) are displayed in a small popup window.

The border of the selected nodes becomes red.



If user clicks or hovers a mouse over a node, the node name is magnified.



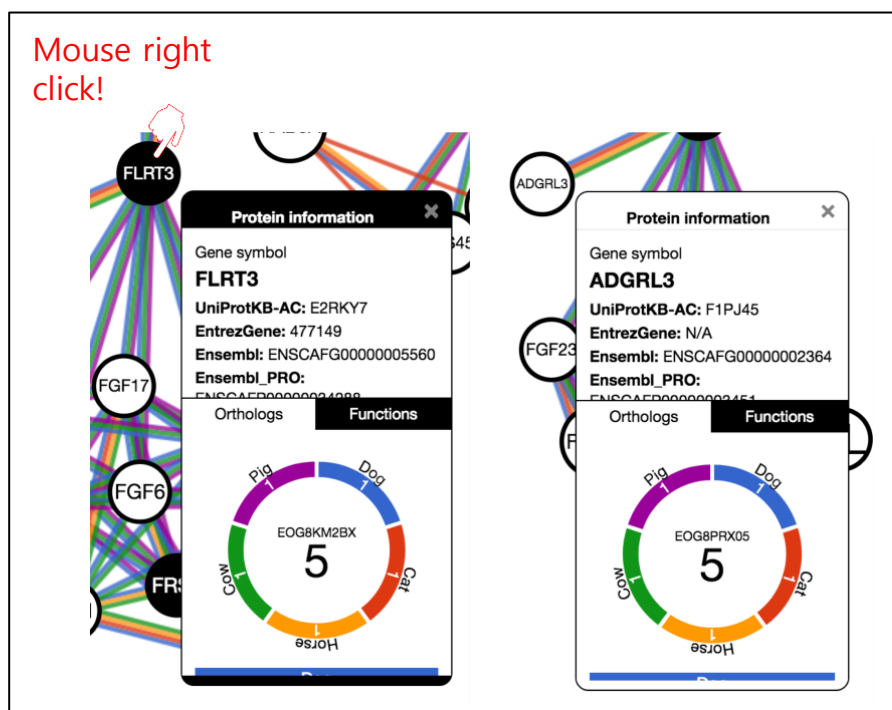
(1-2) Looking for information related to nodes

When right-clicking on a node, the user can see the information related to this node as shown below.

“Ensembl protein”, “UniprotKB”, “EntrezGene”, “Gene symbol” and “Ensembl gene” are display at the top of the window.

In the “Orthologs” section, the number of the orthologous protein in each of the used species and the protein ID are displayed. The protein IDs are linked to the information page in the ensembl web server (<https://ensembl.org/>).

GO ID and GO name of the protein are listed in the “Functions” section. The GO IDs and names are linked to the information page of this GO term in the AmiGO web server (<http://amigo.geneontology.org/amigo/>).



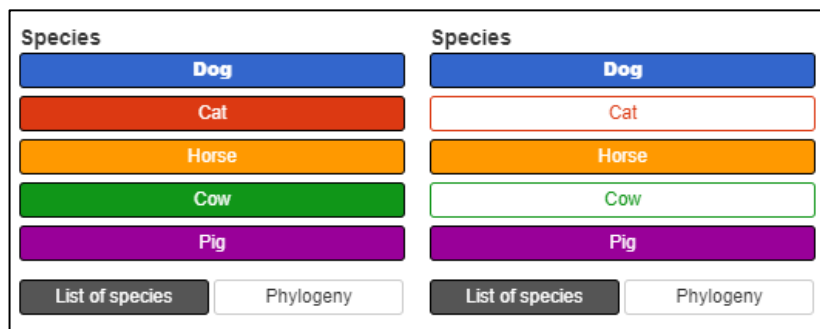
When right-clicking on a node among the input proteins, the top of the information box is labeled to black. Whereas, when right-clicking on a node among the extended proteins, the top is labeled to

white.

(2) Handling edges

(2-1) On/off the interactions of species

When user checks/unchecks the species button in the left panel, on/off the PPIs of each checked/unchecked species. Colors of each buttons match with the colors of PPIs (edges). Refer to the above section 1. (4) for the details of the functions of these buttons.



If an edge exists in multiple species, multiple colors corresponding to those species are used together to draw the edge.

3. Handling additional information

(1) Information of orthologous protein groups, proteins and interactions.

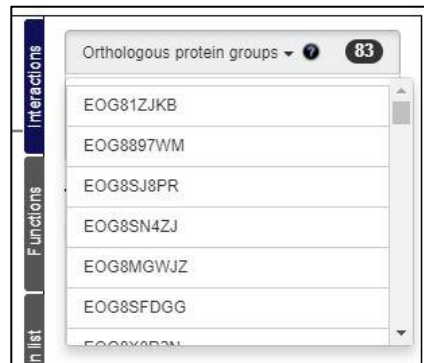
The right panel on the main web interface shows the following additional information related with the shown network. The following numbers are updated on the fly as proteins and interactions in a current network are changed by users.

(1-1) Orthologous protein groups

It shows the list of orthologous protein group IDs with their total number in the black circle as shown below.

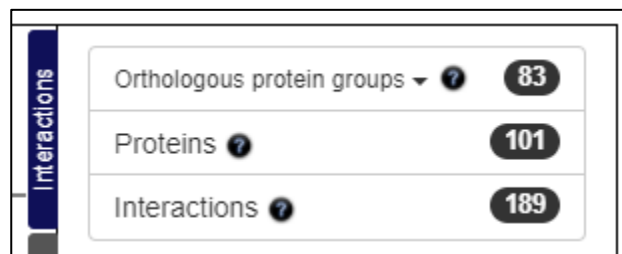
When user clicks on the “Orthologous protein groups”, the list of orthologous protein group IDs is expanded.

When user click one in the list of orthologous protein group IDs, the corresponding proteins are highlighted in the network.

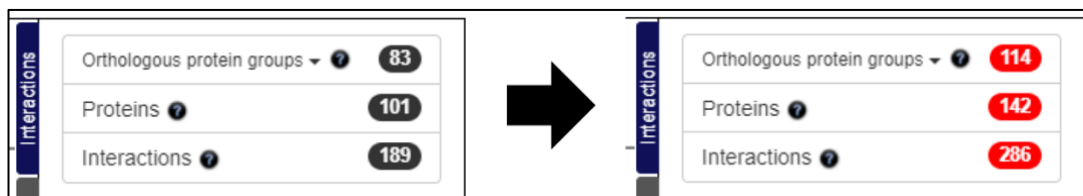


(1-2) Proteins and interactions

The number of proteins and interactions used in the visualized network are shown as below.



If user changes the current network, these numbers are updated on the fly as shown below.



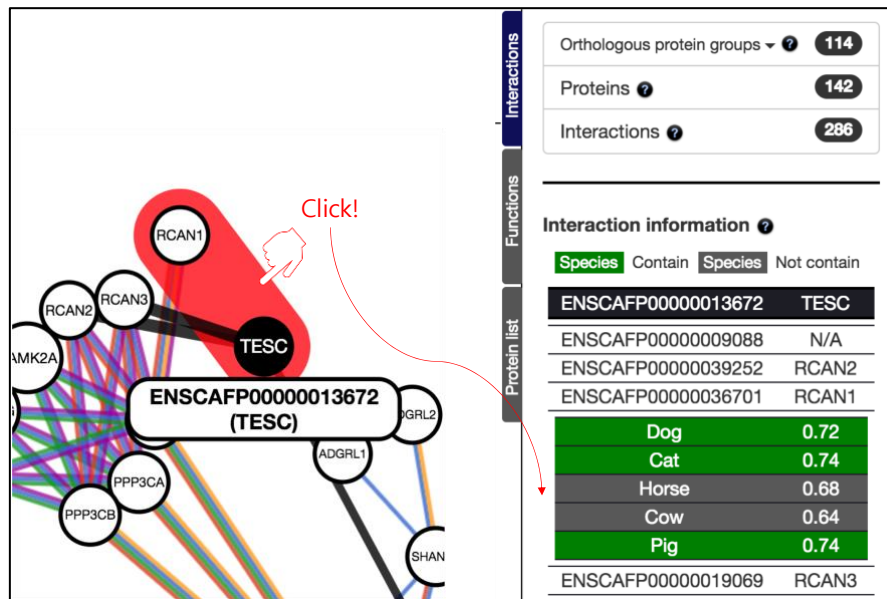
(2) Handling selected edges

When user click a node, the name of the node is listed together with the names of interacting nodes on the “Interactions” panel on the right of the main web page.

On the “Interactions” panel, the selected node is shown with a black background, and the interacting nodes are shown with a white background.

User can see which species have a specific edge by clicking a specific interacting node on the “Interactions” panel. The edge corresponding to this click is highlighted using a red color as shown below.

The above information can be also seen by first clicking a specific node, and then clicking the connected edge after it is highlighted (thickened) as below.



The selected nodes are shown as a tabular form with the Ensembl protein ID and gene symbol. If the node does not have a gene symbol, it is indicated as 'N/A'.

User can highlight a specific edge in the visualized network by a mouseover on a specific interacting node on the “Interactions” panel.

The number next to the species name means the STRING confidence score for the target species, and the mean value of the confidence scores of all interactions between two proteins in different orthologous protein groups for the comparative species.

(3) Zoom-in and moving the network

User can zoom in/out the network using the mouse wheel.

When clicking the blank space in the network, user can move the network to any position.

(4) Handling the results of Gene Ontology enrichment analysis

The “Functions” panel on the right side of the main web page shows the results of GO enrichment test using input proteins and extended proteins (if available).

The panel allows sorting of the GO enrichment test results by Q value, GO IDs and GO names. and filtering those below the specified score set by the user for the Q value.

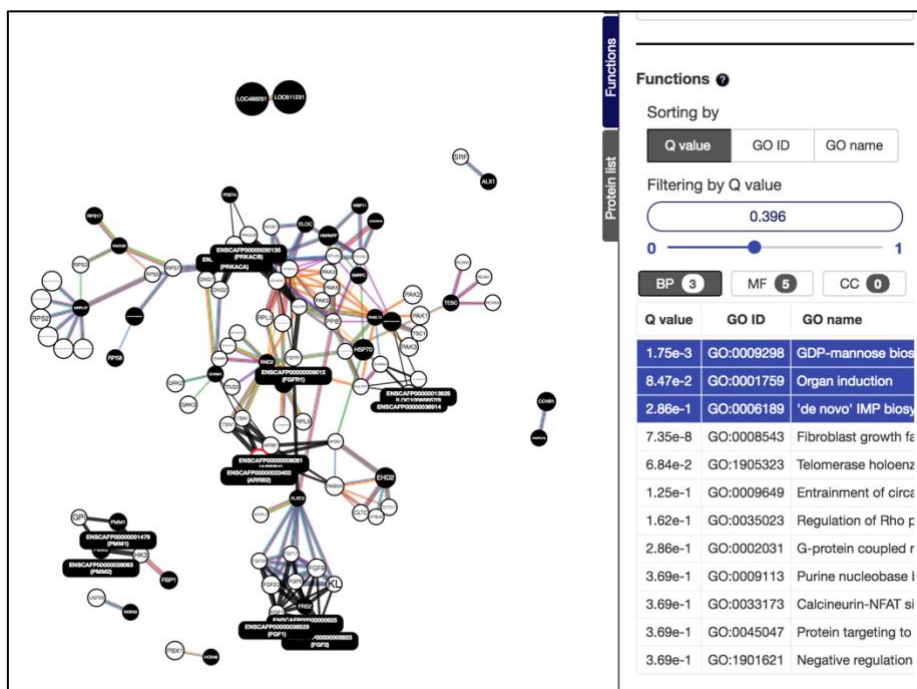
User can also use Q value cutoff score.

When user click a GO term, nodes associated with this function are highlighted using labels with a black color in the network.

In the “Functions” panel, the selected GO terms are highlighted by a blue background color and listed at the top of the list regardless of their original sorting order.

The selection of GO terms can be turned off by clicking them again.

The selected number of GO terms are shown in parentheses in the GO domain buttons (BP, MF, and CC).



In this figure, the number of the selected GO terms is 3 (if the type is BP (Biological Process)), 5 (if the type is MF (Molecular Function)) and 0 (if the type is CC (Cellular Component)).

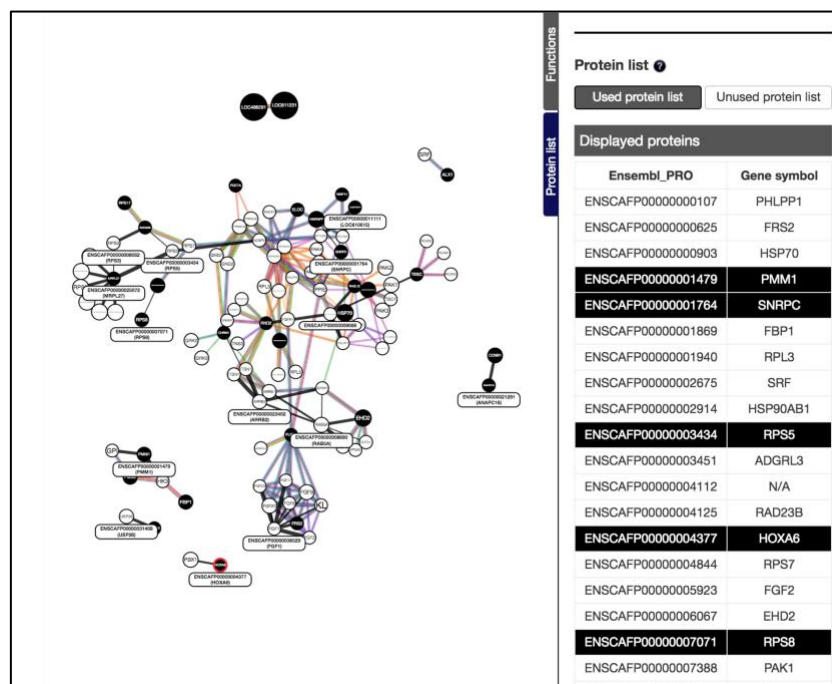
(5) Handling the list of proteins

The “Protein list” panel on the right side of the main web page shows used and unused proteins in the current analysis.

“Used protein list” contains user input proteins and proteins that were obtained from the protein extension process that are used in the current analysis. “Displayed proteins section” contains proteins currently displayed in the network. “Hidden proteins section” shows proteins that are not shown as the nodes in the current network because of a chosen interaction level.

Protein information is shown as a tabular form with the Ensembl protein ID and gene symbol. If there is no gene symbol, it is indicated as 'N/A'.

When user click one of proteins in “Used protein list”, it is highlighted with a node name in the network.



“Unused protein list” contains user’s input proteins not used in the current analysis. Some proteins can be filtered out due to missing information in the STRING, and orthoDB databases.

Protein list	
Functions	
Protein list	
Protein list	
Used protein list Unused protein list	
Ensembl_PRO	Gene symbol
ENSCAFP00000038557	N/A
ENSCAFP00000034923	N/A
ENSCAFP00000028722	MOB3A
ENSCAFP00000025048	KCTD13
ENSCAFP00000031488	N/A
ENSCAFP00000035542	LOC483480
ENSCAFP00000008560	N/A
ENSCAFP00000016624	SV2B
ENSCAFP00000023285	EEF1G
ENSCAFP00000017326	N/A
ENSCAFP00000035746	N/A
ENSCAFP00000031382	N/A
ENSCAFP00000035473	LOC611988
ENSCAFP00000032964	N/A
ENSCAFP00000032911	LOC612247
ENSCAFP00000022624	BRI3
ENSCAFP00000007315	C1QL2
ENSCAFP00000010358	SOX12
ENSCAFP00000007020	ZNF404
ENSCAFP00000001692	MNF1