

FunRich Tool Documentation

Version 2.1.2



Shivakumar Keerthikumar, Mohashin Pathan, Johnson Agbinya and Suresh Mathivanan

Mathivanan Lab
<http://www.mathivananlab.org>
La Trobe University
LIMS1
Department of Biochemistry
Bundoora, Vic-3086

Contents

System requirements and Installation	3
Highlights of FunRich tool	4
Overview of FunRich tool features	5-9
Types of query/input list	6
Enrichment analysis	7
Interaction network analysis	8
Characteristic feature of Interaction network analysis	9
Background database	10-15
FunRich database	11-12
UniProt database	13-14
Custom database	15
Walk through examples	16-28
How to generate Venn diagram	16-18
How to perform Enrichment analysis	19-25
Using 'FunRich' database as background	19-21
Using 'UniProt' database as background	22
Using 'Custom' database as background	23-24
How to generate Heatmap	25
How to generate interaction networks	26-28
Using 'FunRich' database as background	26-27
Using 'Custom' database as background	28

System requirements and Installation

System requirements

- Windows XP SP3, Windows Vista, Windows7 and Windows 8 operating systems
- RAM: 2GB recommended
- Hard disk: 850 MB free space required
- Microsoft.NET version 4 or later required
- Unzip/Unrar/7-Zip file manager required.

Installation

- Please download the tool available at <http://www.funrich.org/download/>
- After downloading unzip/Unrar using any unzip file manager
- After unzipping, navigate into the folder 'FunRich' and click on the application 'FunRich' to open the FunRich tool.

Highlights of FunRich

The results of the enrichment analysis can be visualized in a variety of graphical formats like column and bar graphs, pie chart, Venn diagram, heatmap and doughnut charts.

The user can customize the graphs by changing/editing the legends, axis labels, colors, styles, fonts and title of the graphs in FunRich.

The number of enriched terms displayed on the graphs can be selected, controlled and sorted.

Users can perform enrichment analysis against custom background database irrespective of species.

Interaction network analysis can be performed by using FunRich and custom database options.

Overview of features in FunRich

FunRich can be used for the enrichment and interaction network analysis of genes and proteins (Fig. 1).

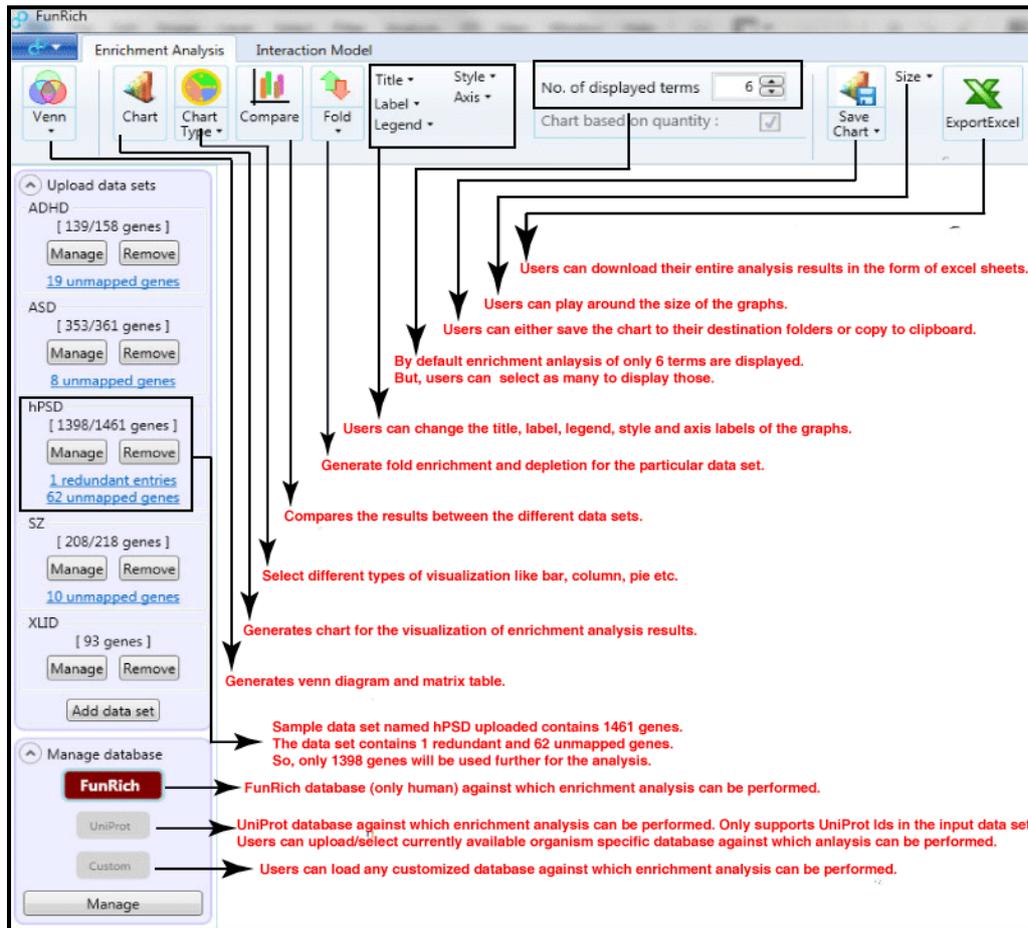


Figure 1. Overview of features in FunRich

Type of query/input list

FunRich accepts following type of genes and protein ids separated by new lines for analysis.

1. Official gene symbols (for example: BTK, BRCA1, EGFR)
2. Entrez Ids (for example: 695, 672, 1956)
3. UniProt Ids (for example: Q06187, P00533)
4. UniProt accessions (for example: BTK_HUMAN, BRCA1_HUMAN)
5. RefSeq Ids (for example: NP_958441.1, NP_958440.1)

Enrichment analysis

Enrichment analysis can be performed by selecting different background database options implemented in FunRich.

For example, against 'FunRich' database background, the following enrichment analysis can be performed.

1. Cellular process (CC)
2. Biological process (BP)
3. Molecular function (MF)
4. Protein domains
5. Site of expression
6. Biological pathways
7. Transcription factors
8. Clinical phenotypes

Secondly, against 'UniProt' database background, only enrichment of CC, BP and MF terms can be performed. But, by using 'Custom database' as background database the users can perform enrichment analysis, not just limited to above terms alone.

Interaction network analysis

Interaction network analysis can be performed using both FunRich and Custom database options implemented in FunRich.

Currently, FunRich has many interaction network layouts like planetary, circular, packed and square.

Planetary: By default, based on the degree distribution of the nodes, the one with higher degree distribution is always placed in the center and rest of the nodes in the periphery.

Circular: By default, irrespective of degree distribution, all the nodes and its interacting partners are circularized randomly.

Packed: By default, the nodes with higher degree distribution is always represented with higher radius and always emphasized in a circular fashion as well as placed closely with the other groups having similar degree distribution.

Square: This layout is very similar to 'Packed' network layout. However, unlike 'Packed' the nodes with higher degree of connectivity and nodes with less degree of connectivity contains equal radius.

Characteristic features of interaction network analysis

1. By default, biological pathways enriched in the interaction networks will be displayed while using 'FunRich database as background.'
2. If users are interested in the interaction network analysis but with custom pathway datasets, the same can be uploaded using 'Custom database' options.
3. List of genes enriched in particular pathways can be highlighted within the interactions network and separate sub-network can be created for further analysis.
4. Users can further analyze that particular sub-network by using below features.
 - Add direct neighbors (interacting partners) for those nodes in the sub-network and visualize.
 - Users can further add interacting partners not part of the input/query gene list for those nodes in the network and visualize.
 - Users can further add all the genes which are involved in that particular highlighted pathway and known to interact with those nodes in the sub-network.
5. Users can focus on the particular node and plot the interacting partners of the focused nodes.
6. Users can right click on the node and get the detailed description of that particular node.
7. Users can further change the node size, color, label and its font size.

FunRich database: FunRich database option can only be used to search human datasets (Fig. 3). In order to use this option, following steps needs to be carried out.

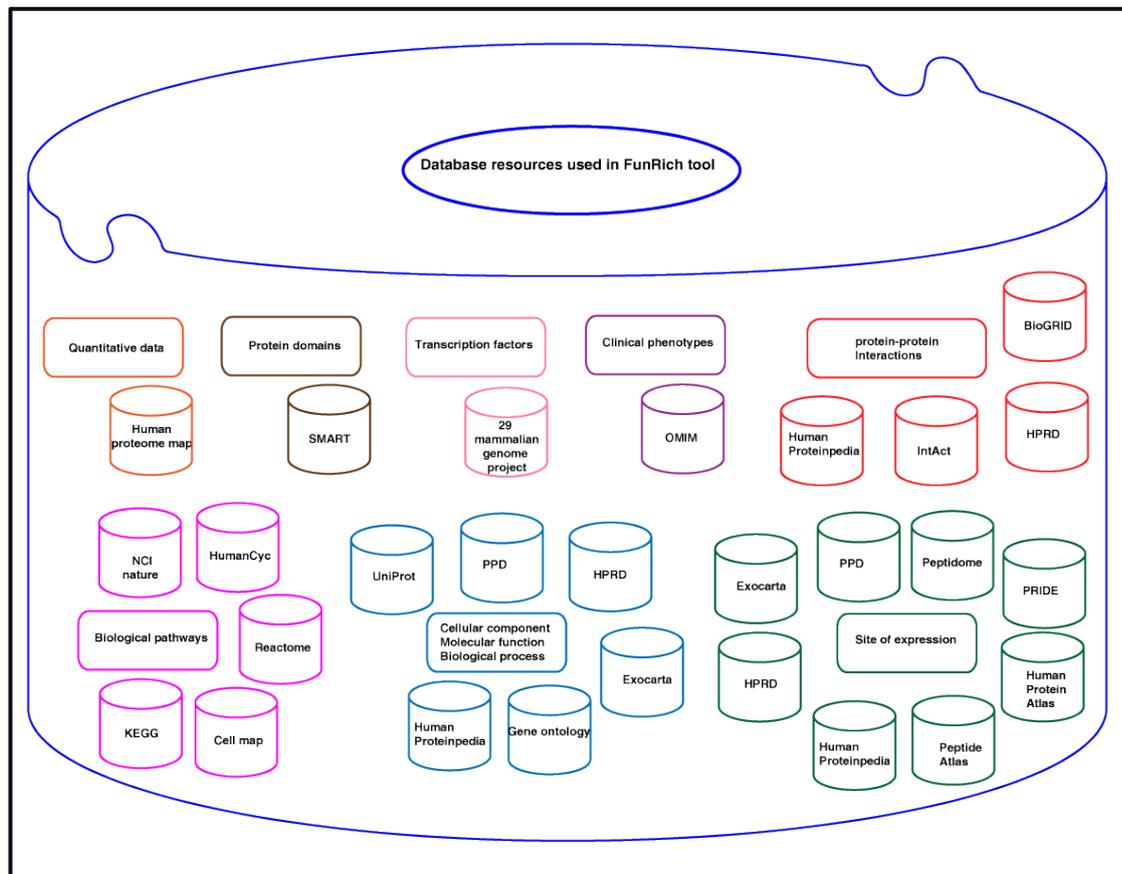


Figure 3. Types of human background data sets used in FunRich database

1. Click on the 'add data set' under the category 'upload data sets'.
2. Click on 'Browse' button to upload the query gene/protein lists. Users can also simply 'copy' the genes/protein list and click on the 'apply' the button. Users can submit all types of genes/protein list as explained in the 'type of query/input list' under the section 'Overview of FunRich tool features'.
3. Enter the name of the data set. A short name is advisable.
4. In order to select the background FunRich database, click on the 'manage' button under the category 'Manage database'. Select 'change' button to select the respective database.
5. After loading your query list and selecting the appropriate background database, the FunRich tool automatically maps your list of genes/proteins to the background datasets and displays the mapped and unmapped number of genes/proteins list.

6. The 'Chart' icon on the Enrichment analysis tab gets enabled. Click the icon to depict the results of the enrichment analysis against different background human datasets (by default, column graph for the cellular component is displayed).
7. Click on the interested tabs (named as Cellular component, Molecular function, Biological process, Biological pathway, Protein domain, Site of expression, Transcription factor, Clinical phenotype) displayed below the graphs.
8. Users can visualize the results in different chart types by clicking on the icon named 'Chart type'.
9. By default, only 6 best (based on the p-values) enriched terms are displayed. Users can further increase these numbers of displayed terms by clicking on the 'No. of displayed terms'.
10. Users can further customize the graph by selecting the enriched terms.
11. Users can also edit the title, axis, labels, legends and style of the graphs in various formats.
12. Users can display separate number of fold enriched/depleted against these background datasets by clicking on the icon named 'Fold'.
13. Besides downloading the individual graphs, users can download the entire enrichment analysis results against all these data sets in the single Excel sheet for further validations by clicking on the icons 'Save Chart' and 'Export Excel' respectively.

UniProt database:

The UniProt database option can be used to search against different organisms currently supported by UniProt. Currently, the FunRich supports all the list of query/input genes/protein list as specified in this documentation under the section ‘types of input/query list’. In order to use this option, following steps needs to be carried out.

1. Repeat all the above steps except step 4.
2. To select UniProt database, click on the ‘manage’ button under the category ‘Manage database’. Further, click on the ‘change’ button and select ‘UniProt database’ options and confirm the same.
3. Changing the database to UniProt from the other database will result in deleting the entire user submitted query list. The same message will be displayed as popup message. Please select ‘Yes’ to proceed further.
4. Next click on the ‘Configure UniProt’ button under the category ‘Configure database’, a new window will appear asking the user either to ‘Add new UniProt’ database or ‘Select’ one from the already existing ‘Saved UniProt database’ list.
5. If the user decides to add new UniProt database, then click on the ‘Add new UniProt’ button. A new window appears asking the user, either to select one from the list of organism to download the corresponding organism specific UniProt database. Currently, the list contains ~11 different taxonomical level (Fig. 4) against which enrichment analysis can be performed. Within the same window, users can either upload the UniProt database stored in their local system or can directly click on the URL link (Fig. 5) provided to download the most updated database from UniProt directly and can further upload the same.

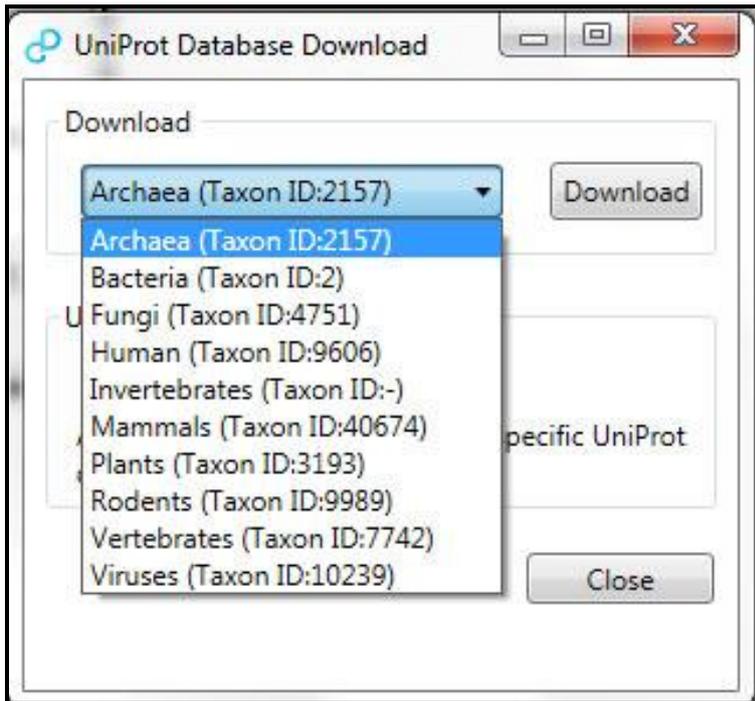


Figure 4. UniProt database download features

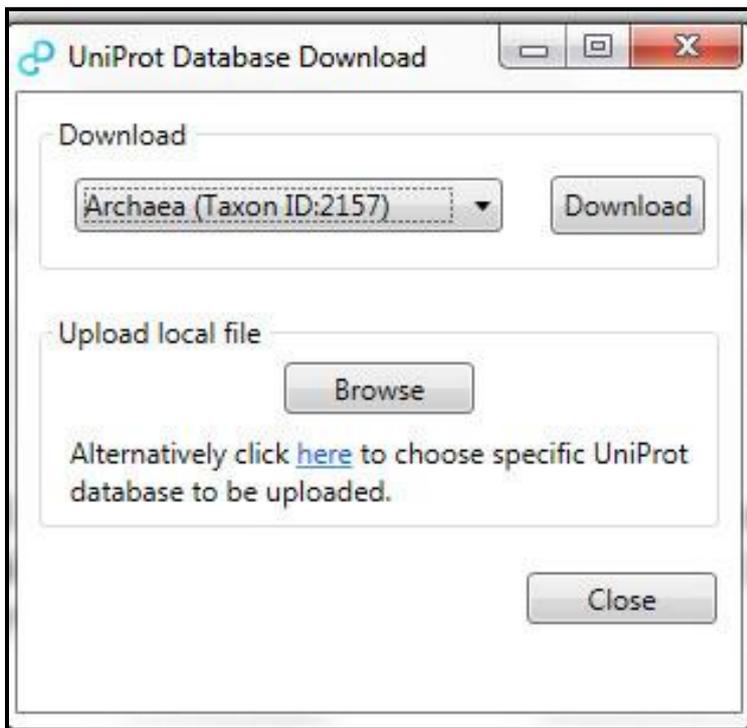
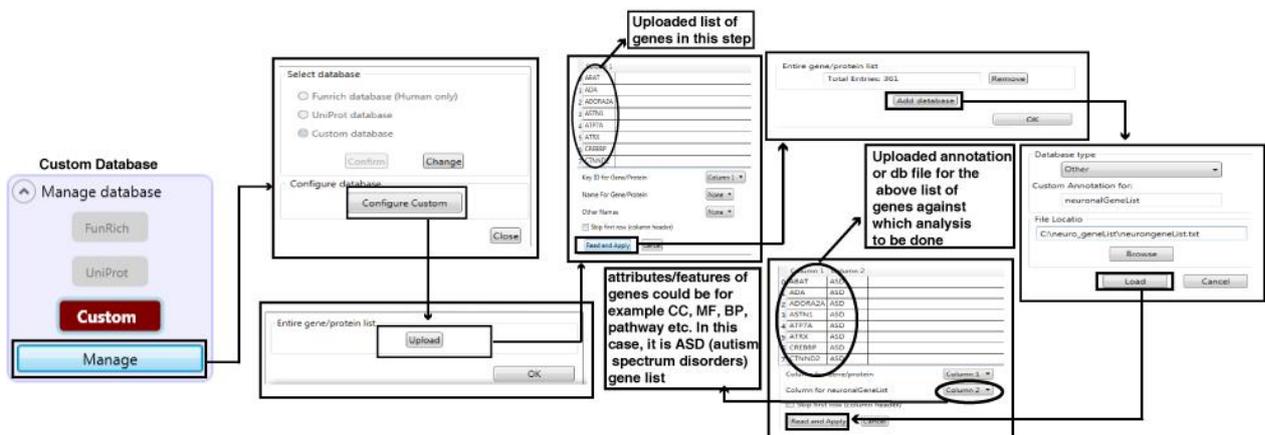


Figure 5. Uploading UniProt database directly from UniProt site

6. Enrichment analysis and further graphical visualizations can be performed as explained in the above steps under the section 'FunRich database'.

Custom database: The power of comparing any query/input list against any user specified database irrespective of any organisms exists in the custom database options. Please follow these steps given below for further analysis.

1. Repeat the above steps 2 and 3 as given under the ‘UniProt database’ section.
2. Click on the ‘Configure Custom’ button under the category ‘Configure database’. A new window will appear displaying pre-existing or earlier saved gene/protein list and its corresponding annotation/database file. Users can choose either to remove the existing one or upload the new one.



3. If user decides to upload the new one, just click on the ‘remove’ buttons, then the users will be prompted to upload the new entire list of genes/proteins.
4. This entire list of genes/proteins actually represents the list of genes/proteins as key identifier for which annotation/database file would be uploaded against which users are interested in enrichment analysis.
5. After uploading the entire list of genes/proteins, the users are asked to select the particular column (also known as key identifier) for which further database/annotation file will be uploaded in the next step (Fig. 2).
6. Once uploaded, click on the ‘Add database’ button. Here, user can upload the database file and select the type of database either from the existing list of names or give their own name.
7. In the next step, users are again asked to select the particular column containing the list of genes/proteins. Please make user you have selected the same identifier exist in this database file as selected in the above step 5. Besides, the users must select another additional column which contains some features/category/phenotype for those genes/protein lists.
8. Click on ‘Read and Apply’ button and click on the ‘OK’ button for the ‘Custom database’ to be highlighted/selected. Once this custom database is selected, users are required to follow steps mentioned in the ‘FunRich database’ section on how to upload their query gene/protein lists and further carry the enrichment analysis of that list.

Walk through examples

This section details the steps required to generate few graphs using sample datasets like neurological disorders and lipidomics gene lists.

How to generate Venn diagram

1. Download the sample data at:

Input/query gene lists: http://www.funrich.org/sample_data/sample_data.7z

Background custom annotation/database file: http://www.funrich.org/customdb/custom_db.7z

2. Upload the downloaded input/query gene lists named 'ASD.txt', 'ADHD.txt' and 'SZ.txt' into the FunRich tool using the 'Upload datasets as explained above.

3. Click on the icon named 'Venn' to generate Venn diagram between these three datasets. By default, scalable Venn diagram for these datasets is generated.

4. Users can also right click on those numbers to further select those genes for further analysis (Fig. 6).

5. Load 4th input/query gene list named 'XLID.txt' and automatically, Venn diagram between these four datasets will be generated (Fig. 6).

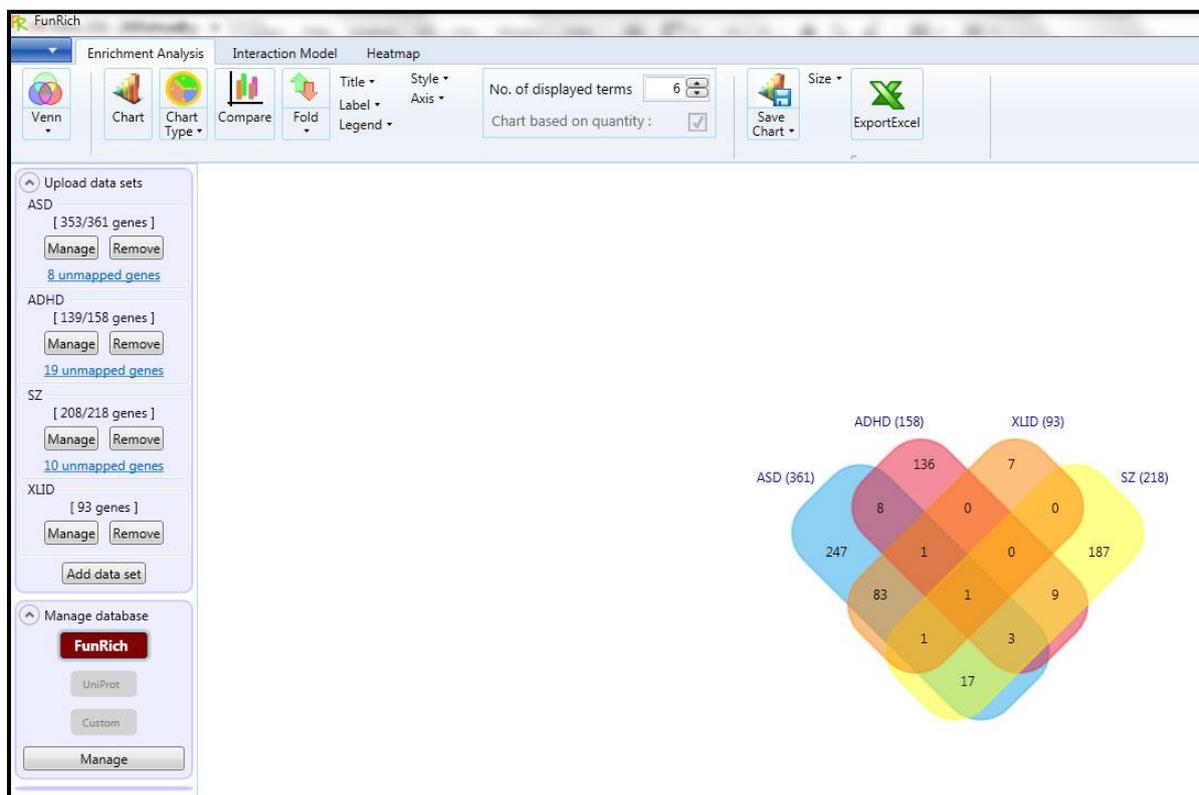


Figure 6. Venn diagram depicting four dataset generated by FunRich tool

6. Click on the 'Venn' icon options and then select gene/protein list. Here, uncheck the fourth dataset named 'XLID' to revert back to the above Scalable-Venn diagram (Fig.7).

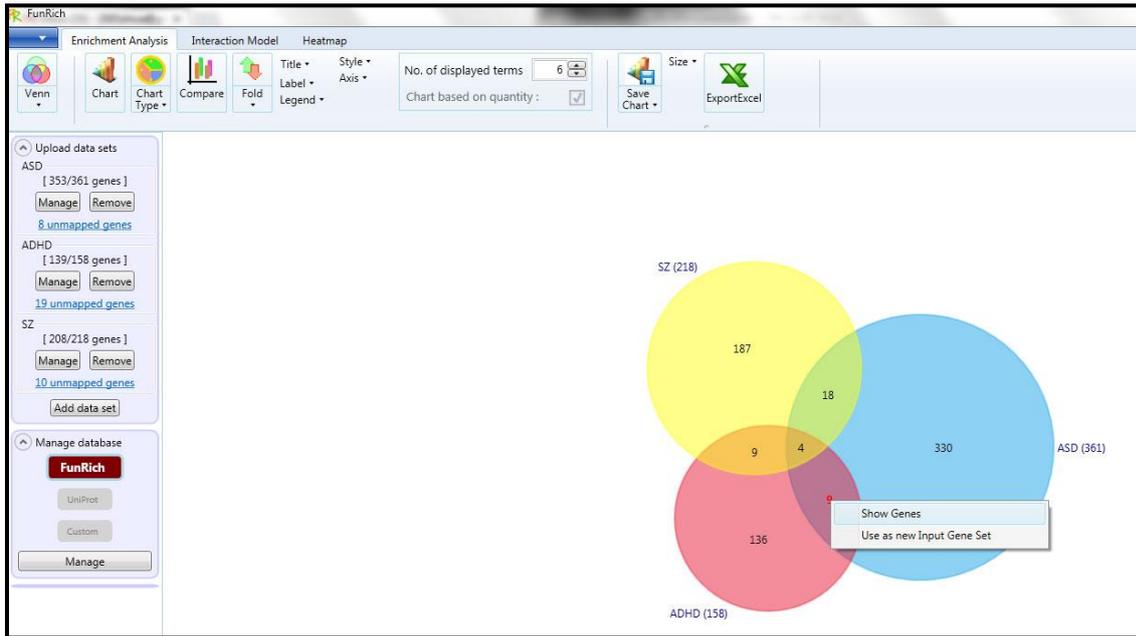


Figure 7. Scalable-Venn diagram of the sample datasets generated using FunRich tool.

7. Click on this  icon just located above the 'Venn' icon. Select 'Option' and uncheck the option named 'Area to be proportional to the number of genes' and then click on the 'Apply' button. This will generate unscalable-Venn diagram (Fig.8).

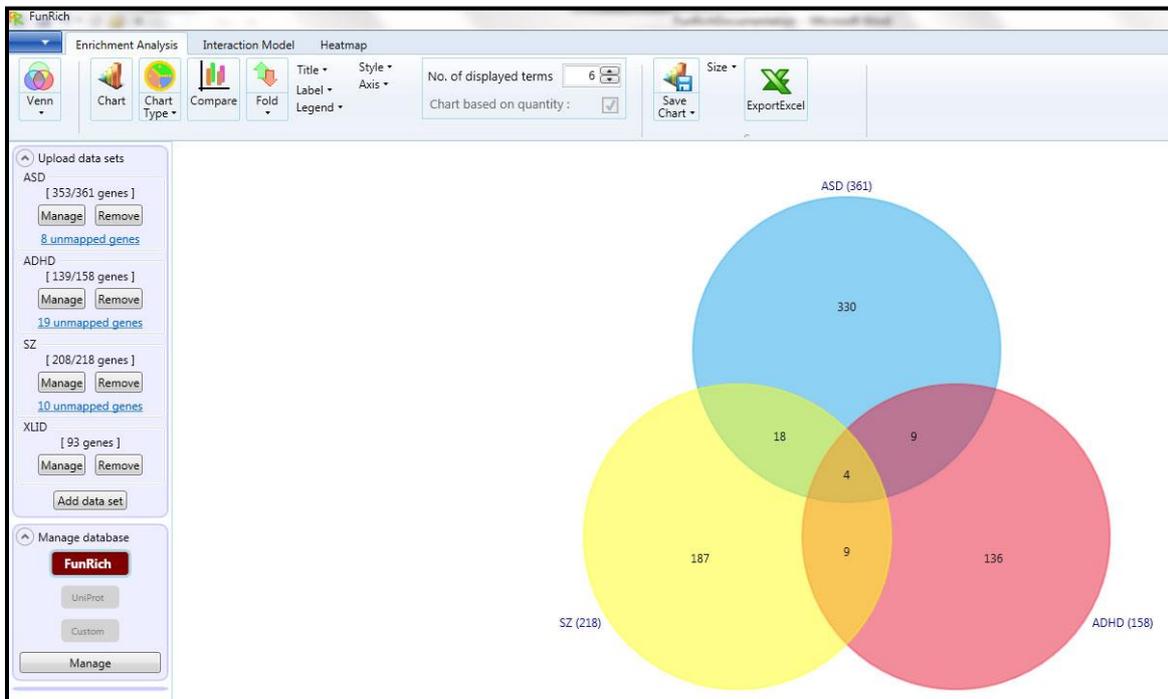


Figure 8. Unscalable-Venn diagram generated by FunRich tool

8. To generate matrix table also known as pairwise comparison, again repeat the above step7 but check the option named ‘Show pair-comparison only’ and then click on the ‘Apply’ button (Fig.9).

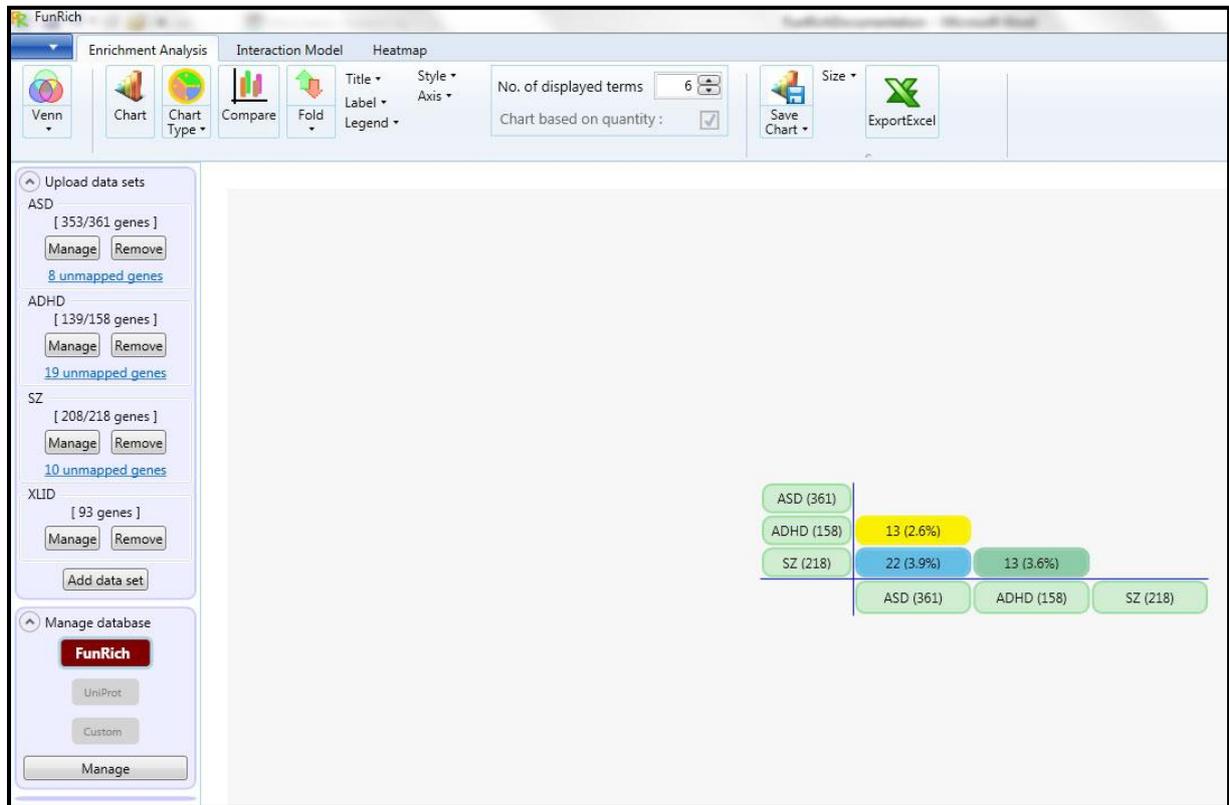


Figure 9. Matrix table (pair-wise comparison) generated by FunRich tool.

How to perform enrichment analysis

This section describes steps to perform enrichment analysis using the sample dataset against different background databases.

Against FunRich background database

1. Upload the downloaded sample datasets named 'ASD.txt' and 'XLID.txt' into the FunRich tool as explained in the above steps.

2. By default, FunRich database is selected as background database against which enrichment analysis can be performed.

3. Click on the  icon named 'Chart' and select 'ASD' as input dataset and click 'OK' to display the enrichment results against FunRich background database.

4. By default, 'Column graph' depicting the enrichment analysis of 'ASD' gene list for first six 'Cellular Component (CC)' terms is displayed based on the p-value.

5. Click on this icon  and increase the number to 8 to show eight different CC terms. The size of the graph displayed can be anytime increased or decreased by clicking on this icon .

6. Click on the 'Axis' button and select the 'X-axis label rotation' option to specific angle for clear visualization of the terms on the 'X-axis' (Fig. 10).

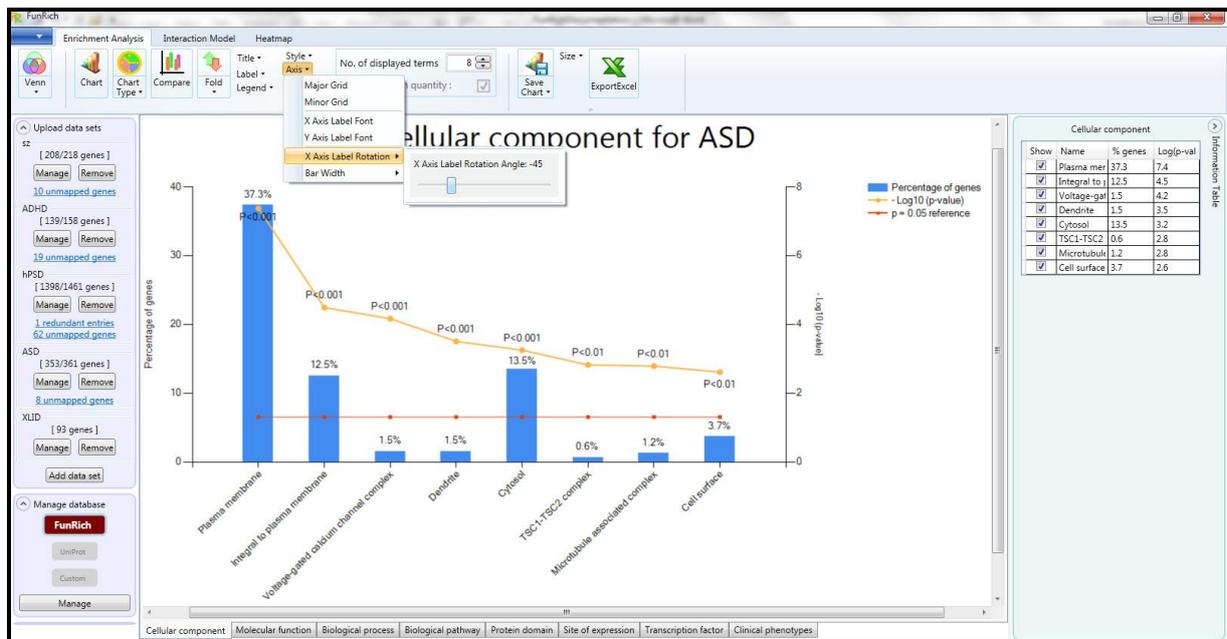


Figure 10. Enrichment analysis of sample dataset against FunRich database

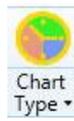
7. On the right side of the graph, a table containing all those eight CC terms is displayed. Users can uncheck any of those terms for not to be shown on the graph.

8. Users can also play around the other graph options like title, style, label and legends to make it more customized.

9. To further analyze same dataset 'ASD' for the enrichment of Molecular Function (MF), Biological Process (BP), Biological Pathway (BP), Protein Domain, Site of Expression, Transcription factor and Clinical Phenotype click on the tabs displayed on the bottom of the graph respectively.



10. Click on the icon 'Compare' and select dataset named 'ASD' from the set A and select dataset named 'XLID' from the set B and click 'OK' button.



11. Click on the icon 'Chart Type' and select 'Doughnut chart' and select on the tab named 'Clinical phenotype' on the bottom of the graph. Click on the 'Style' icon and select 'SoftEdge' as one of the Style option.



12. Click on the 'Legend' icon and reduce the height of the legends by selecting the option 'Reduce Height' under the 'Adjust Legend Size' option.



13. To save the chart click on the icon 'Save Chart'. The FunRich tool generated 'Doughnut chart' with 'SoftEdge' style depicting the 'Clinical Phenotype' enrichment results of the two datasets 'ASD' on the outer circle and 'XLID' on the inner circle will be saved in your local system (Fig. 11).

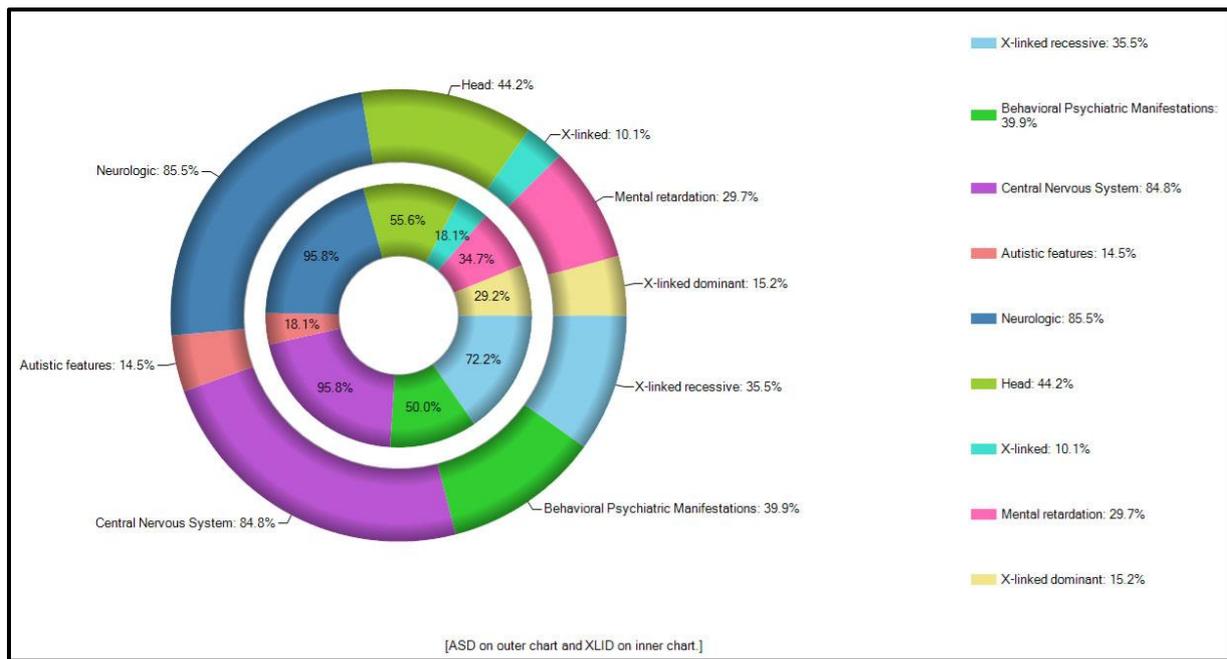


Figure 11. Enriched Clinical Phenotypes compared between the two sample datasets generated by FunRich tool.

Against UniProt background database

Please follow the below steps to perform enrichment analysis against UniProt background database.

1. Click on the 'Manage' button under the category 'Manage database'. Select 'Change' and check 'UniProt database' and 'Confirm' the same. You will be prompted saying that 'Changing the database will remove all the datasets'. Please click on the 'Yes' button to proceed further.
2. Click on the 'Configure UniProt' to add the UniProt database. Click on the 'Add UniProt database' and from the download list select 'human (Taxon ID: 9606)' and click on the 'Download' button to download the human UniProt background database. Within few minutes the database will be automatically downloaded, uncompressed and uploaded. After loading, click on the 'close' button and UniProt database will be selected as background database against which further enrichment analysis can be performed.
3. Upload the sample input list as explained in the previous steps to perform the enrichment analysis.

Against Custom background database

Please follow the below steps to perform enrichment analysis against Custom background database

1. Please repeat the above step1 under the section ‘Against UniProt database’. But, instead of selecting ‘UniProt’ check ‘Custom database’ to perform the enrichment analysis against custom database.

2. Click on the ‘Configure Custom’ button and click on the ‘Remove’ button to remove all the existing or pre-saved gene/protein list and annotation/database files.

3. Now, click on the ‘Upload’ button to upload the entire gene/protein list and upload the file name ‘custom_gene_list.txt’ downloaded from the sample input datasets. After uploading, select the ‘column1’ button as the name of the ‘Key ID for gene/protein’ and click on the ‘Read and Apply’ button to proceed further. A window showing ‘Total Entries: 2064’

4. In the next step, we will add custom annotation/database file for the above uploaded 2064 gene list. It could be any annotation like CC, MF, BP, Pathways, disease phenotype etc. In this example, we will load custom neuronal and neurological disorder annotation file for the above gene list.

5. Click on the ‘Add database’ button and from the list of database type select ‘Other’. Enter the name of this custom annotation/database file as ‘Neuronal_Neurological disorders’ and click on the ‘Browse’ button to upload the custom annotation/database file named ‘custom_neuronal_neurological_db.txt’ from the downloaded sample datasets.

6. After loading the custom annotation/database file, select the ‘column1’ button as the name for the gene/protein column. Please make sure the data types for this column should be same as selected in the step3. Next, select the ‘Column2’ as the name for the Neuronal_Neurological disorders column’ which represent the actual feature or custom annotation for the 2064 gene list. Click on the ‘Read and Apply’ button to proceed further to upload custom annotation/database file containing 2281 entries. Further ‘Click’ on the button and select ‘Close’ button.

7. Now, upload the input/query gene list named ‘SZ.txt’ downloaded from the sample datasets

as explained in the previous steps and click on the  icon , select the char type ‘Bar chart’

from the icon  to display the enrichment of neuronal/neurological disorder phenotype in the ‘Bar graph’ format (Fig.12).

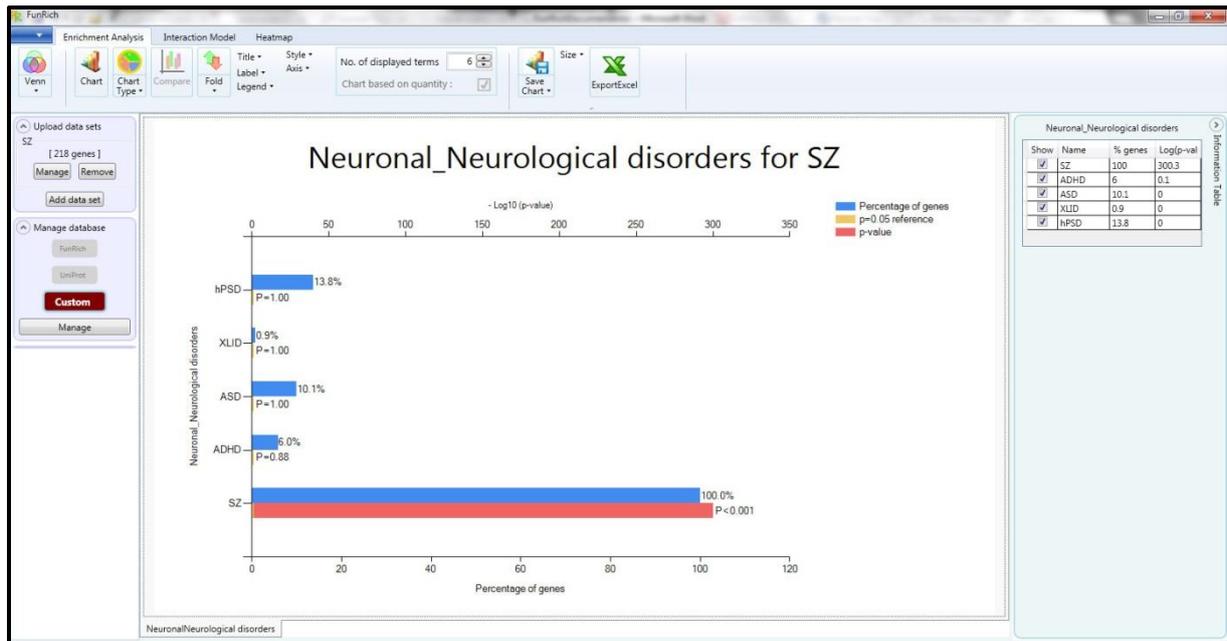


Figure 12. Snapshot of enrichment analysis against custom database using FunRich tool

How to generate interaction networks

Interaction networks of query genes list can be depicted by using both 'FunRich' and 'Custom' database as background.

Using 'FunRich' as background database.

1. Upload the sample gene list named 'custom_gene_list.txt' file downloaded from the sample datasets.

2. Using 'Manage Database' options, select the 'FunRich' database (if get stuck in this step, always follow the steps explained above to successfully select the 'FunRich' database as background).

3. Click on the icon  named 'Interaction' which is under the tab named 'Interaction Model' to generate protein-protein interaction (PPI) networks of sample input list. By default, 'Planetary' layout is displayed and users can right click on the interaction network layout space and select the 'packed' layout (Fig. 14).

4. Adjacent to PPI networks, the sample genes enriched for particular biological pathways are also shown (Fig. 14) along with their p-values.

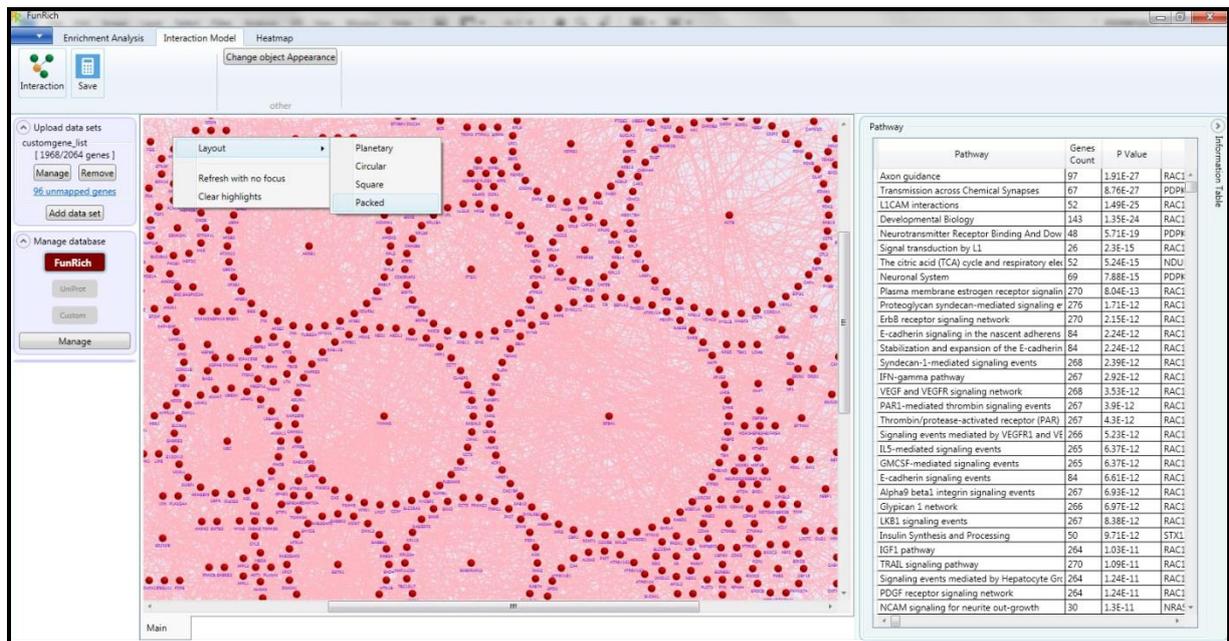


Figure 14. Snapshot of protein-protein interaction networks generated using FunRich tool

5. Click on the pathway term named 'Axon guidance' and all those 97 nodes known to be significantly enriched in that particular pathway will be highlighted.

6. For further analysis on this subset of nodes, right click on that particular pathway and select 'Plot to new network'. A new window pops up displaying three check box options using which further analysis can be carried out.

First check box option named 'Add direct neighbors from dataset' represent adding interacting partners which directly interacts with those 97 nodes.

Second check box option named 'Add direct neighbors outside dataset' represent adding interacting partners which are known to directly interact with those 97 nodes but not the part of interactors interacting with sample input list.

Third check box option named 'Include other genes involved in the pathway' represent adding all those genes involved in that particular pathway (here Axon guidance) as interacting partners which are known to interact with those 97 nodes.

7. Select the first check box and click 'OK' button to visualize all the direct interactors of 97 nodes separately and change the layout to 'Planetary'.

8. In order to further analyze those 97 nodes, again right click on the pathway term 'Axon guidance' select 'Show nodes' and save the list of genes for further enrichment analysis using FunRich tool.



9. Click on the icon  named 'Save' to save the interaction network graph in your local system folder.

10. Users can further change the color, size, thickness and font of the nodes etc. to make it more customized.

Using 'Custom' as background database

Irrespective of species, custom database can be used as background database even for the interaction network analysis by uploading custom PPI network database as given in the steps below.

1. Under 'Manage database' category click on the 'Manage' button and select the 'Custom database' as background database.
2. Click on the 'Configure Custom' and remove the entire previously loaded and existing gene/protein list.
3. Click on the 'Upload' button and upload the file named 'Custom_gene_list.txt'. Select the 'Column1' button as Key ID for those list of genes and click 'Read and Apply' button. All those 2064 entries from that gene list will be loaded.
4. Now, click on the 'Add database' button and from the database type choose 'Interaction' and click on the 'Browse' button to upload the interaction database file. Upload the sample databases file named 'PPI_human_dataset.xlsx' and click on the 'Load' button.
5. Select 'Column1' button as the first gene/protein column and select 'column2' as second gene/protein column. Further select 'Read and Apply' button to load the interaction database file. After loading click on the 'OK' and 'Close' button.
6. Using 'Upload dataset' option upload file name 'hPSD.txt' from the downloaded sample input gene list. Click on the 'Interaction Model' tab and select the icon  named 'Interaction' to generate the protein-protein interaction networks using the custom interaction database file uploaded.
7. Click on this  icon just located above the 'Venn' icon. Select the 'Option' button and click on the 'Interaction' menu. Now, check the option button named 'Include nearest neighbors' to increase the minimum number of interactions that neighbor must have to 2 and click on the 'Apply' button. Again, click on the  icon to further generate the interaction networks with above updated parameters. This will further add minimum of 2 interacting partners to all those nodes generated using the file name 'hPSD.txt'.