

Multi-Omics Data Integration and Interpretation Using the Network Explorer Module

By: Jasmine Chong, Jeff Xia

Date: 14/02/2018

The aim of this tutorial is to demonstrate how the Network Explorer module of MetaboAnalyst can be used to integrate multi-omics data using a knowledge-based network approach, thereby permitting users to gain novel insights and develop new hypotheses. The example data used in this tutorial is a toy dataset that was built to highlight the functionality of the module.

| | |
|--|----|
| Introduction to Network Explorer | 3 |
| Network Explorer Step-by-Step | 4 |
| Step 1. Start | 4 |
| Step 2. Network Explorer | 4 |
| Step 3. Network Explorer Data Upload..... | 5 |
| Step 4. Feature Mapping..... | 6 |
| Step 5. Network Selection | 7 |
| Step 6. KEGG Global Metabolic Network..... | 8 |
| Step 7. Network Selection | 12 |
| Step 8. Mapping Overview..... | 13 |
| Step 9. Network Explorer View | 14 |
| Step 10. Node Explorer Menu | 15 |
| Step 11. Function Explorer Menu | 17 |
| Step 12. Path Explorer | 19 |
| Step 14. Network Customization..... | 21 |
| 1. What is the network size limit? | 21 |
| 2. How do I delete nodes from the network?..... | 21 |
| 3. Can I create a 300 dpi high-resolution network image for publication purposes?..... | 21 |
| 4. Can I change the color and size of my nodes? | 21 |
| 5. How can I get node labels to appear? | 22 |
| 6. Can I move a node cluster? | 22 |
| 7. Can I extract nodes from the network?..... | 22 |

Introduction to Network Explorer

Metabolomics is increasingly used with other omics platforms such as transcriptomics, proteomics, and metagenomics to characterize and gain functional insight into complex diseases/conditions. However, multi-omics data integration and interpretation at a systems level remains a significant challenge. A common strategy is to analyze each set of omics data individually, and then piece together the “big picture” using individual lists of significant features (metabolites, genes, proteins, *etc.*). In particular, biological networks are a useful and flexible means to convey our biological knowledge at a systems level. By harnessing the power of networks and *a priori* biological knowledge, these lists of significant features can be co-projected onto such knowledge-based networks to reveal important links between them, as well as their associations with diseases or other interesting phenotypes. Therefore to address concerns with multi-omics data integration and interpretation, we introduce the Network Explorer module.

The aim of this module is to provide users with an easy-to-use tool that permits the mapping of their metabolites and/or genes (including KEGG orthologs or KOs) onto different types of molecular interaction networks. In particular, we aim to support the integration of transcriptomics and metabolomics data, as well as metagenomics and metabolomics data. The network visualization can then be used to gain novel insights or assist users with the development of new hypotheses. The main steps for Network Explorer are as following and will be described in further detail below:

- i. Upload a list of metabolites and/or genes.
- ii. User’s uploaded list/s of compounds are mapped to MetaboAnalyst’s internal database.
- iii. Select a network to visually explore your data.

Network Explorer Step-by-Step

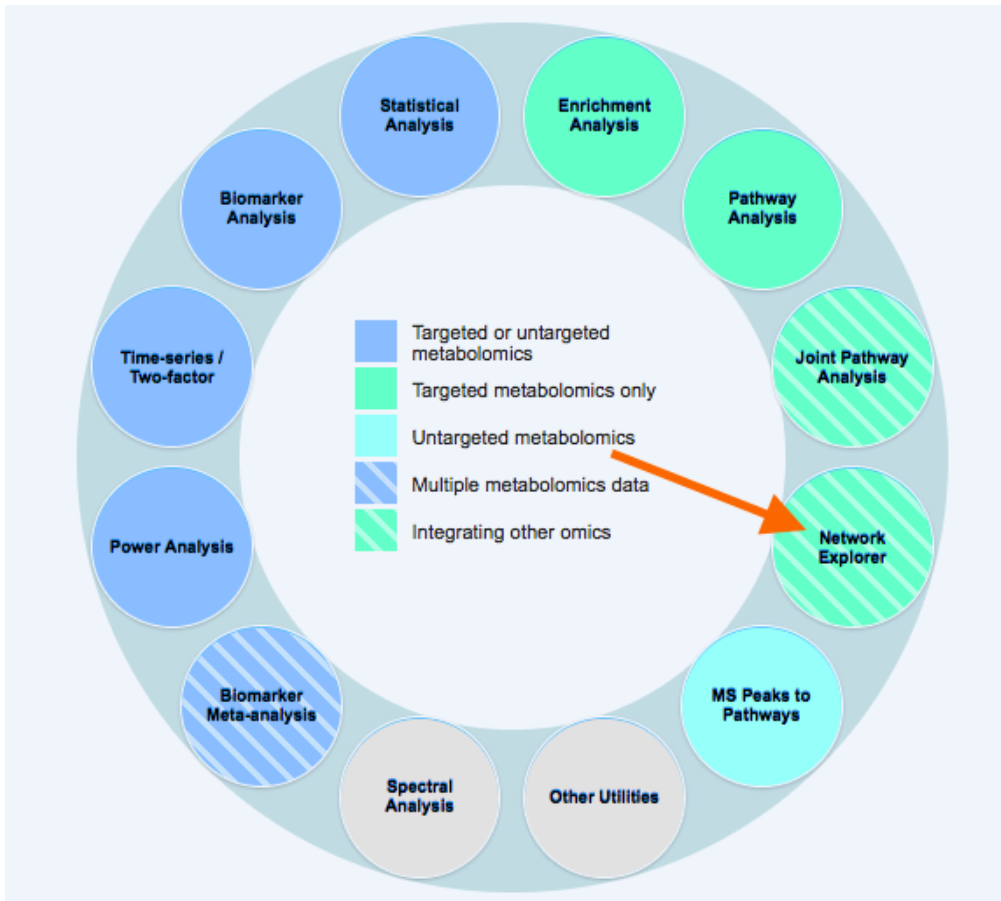
Step 1. Start

On the MetaboAnalyst home page, press “**click here to start**” to enter the module overview.



Step 2. Network Explorer

On the Module View page, click the “**Network Explorer**” circle to begin.



Step 3. Network Explorer Data Upload

In the Network Explorer upload page, copy-and-paste your list of genes and/or metabolites into the respective box, “Gene List” or “Metabolite List”. Click the drop-down menu next to “ID Type” to specify the type of data you are uploading. For the metabolite list, MetaboAnalyst currently accepts compound names, HMDB IDs, or KEGG compound IDs as metabolite identifiers. For the gene/protein list, Entrez IDs, ENSEMBL IDs, official gene symbols, or KEGG orthologs (KOs) are currently supported. Click “[Submit](#)” to upload your data.

Upload a list of human genes or KEGG orthologs, (optional) with a list of metabolites to explore their relationships. You can first [try our example data](#).

Gene List

Metabolite List

Gene list with optional fold changes

Compound list with optional fold changes

ID Type: --- Not Specified ---

ID Type: -- Please specify

Submit

Use Case: Click “[try out example data](#)” on the top of your screen to use the example toy data. A dialogue box will appear, giving you the choice between uploading 1) a list of metabolites (KEGG) and a list of genes (Entrez), or 2) a list of metabolites (KEGG) and a list of KOs (KEGG). Both datasets contain human data.

Please select one example data below

| | Details |
|--|--|
| <input checked="" type="radio"/> Metabolites - genes | Organism: Human ID Type: KEGG Compound ID and Entrez Gene ID |
| <input type="radio"/> Metabolites - KOs | Organism: Human ID Type: KEGG Compound ID and KEGG Orthology ID |

Yes Cancel

For this tutorial, we will select the “Metabolites-genes” data to be uploaded. Click “Yes” to continue. As per the screenshot below, the data is automatically pasted into their respective boxes.

Upload a list of human genes or KEGG orthologs, (optional) with a list of metabolites to explore their relationships. You can first [try our example data](#).

Gene List

Gene list with optional fold changes

```
#Entrez logFC
1737 -1.277784317
83440 -1.034136439
3939 -2.231729728
10911 -1.045657875
10690 -0.968308832
10010 -0.861541301
11224 -1.187399591
63826 -1.405238611
11031 0.785011172
4190 -1.778774832
10782 -2.140715987
10993 -0.925083829
10455 1.732172706
10963 1.177511121
10282 -1.20754269
```

ID Type: (Human) Entrez ID

Metabolite List

Compound list with optional fold changes

```
#KEGG logFC
C00116 1.010972619
C00565 -0.714283001
C00033 -0.822193121
C00583 -1.005192252
C00022 -0.623838569
C00719 -0.406052491
C05984 -0.390152174
C00207 -0.932835099
C00065 0.903658797
C00031 0.548035915
C00079 0.416744818
C02632 -0.515041676
C00064 -0.497216411
C00114 1.102078837
C00073 0.516193785
```


ID Type: KEGG ID

Submit

Click “Submit” to continue.

Step 4. Feature Mapping

Following uploading of the data, the results of the compound/gene name mapping is shown in table/s as per the screenshot below. There are two tabs on this page, the first is the “Compound Name Mapping”, and the second is the “Gene Name Mapping”.



MetaboAnalyst-- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- ID mapping
- Set parameter
- Network viewer
- Download
- Exit

The system requires all the IDs (except common compound names) to be matched exactly. The table below shows the matched genes and compounds from the underlying databases. For common compound names, users can further perform [approximate match](#) by clicking the View link in the Details column. To remove a gene or compound from further analysis, use the Delete link in the last column. Items in the Gene Name Mapping table with lightgray indicates cases where some information is missing while lightred highlights ones that has no matches with in databases.

Compound Name Mapping
Gene Name Mapping

| Query | Hit | HMDB | KEGG | Details | |
|--------|-----------------------|-----------------------------|------------------------|---------|------------------------|
| C00116 | Glycerol | HMDB0000131 | C00116 | | Delete |
| C00565 | Trimethylamine | HMDB0000906 | C00565 | | Delete |
| C00033 | Acetic acid | HMDB0000042 | C00033 | | Delete |
| C00583 | Propylene glycol | HMDB0001881 | C00583 | | Delete |
| C00022 | Pyruvic acid | HMDB0000243 | C00022 | | Delete |
| C00719 | Betaine | HMDB0000943 | C00719 | | Delete |
| C05984 | 2-Hydroxybutyric acid | HMDB0000008 | C05984 | | Delete |
| C00207 | Acetone | HMDB0001659 | C00207 | | Delete |
| C00065 | L-Serine | HMDB0000187 | C00065 | | Delete |
| C00031 | D-Glucose | HMDB0000122 | C00031 | | Delete |
| C00079 | L-Phenylalanine | HMDB0000159 | C00079 | | Delete |
| C02632 | Isobutyric acid | HMDB0001873 | C02632 | | Delete |
| C00064 | L-Glutamine | HMDB0000841 | C00064 | | Delete |

R Command History

Keep collapsed Save

```
1. initDataObjects("comp", "network", FALSE)
2. mSet<-SetOrganism(mset, "hsa")
3. geneListFile<-readChar("replace_with_your_file_name")
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mset, geneList, "hsa", "entrez");
6. compListFile<-readChar("replace_with_your_file_name")
7. compList<-readChar(compListFile, file.info(compListFile)$size)
8. mSet<-PerformIntegCompMapping(mset, compList, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mset)
```

The screenshot below shows an example of the mapping of uploaded-gene names to MetaboAnalyst's internal database. Note that queries highlighted in grey indicate genes that that have missing information. For instance, 1737 did not have a match to our KEGG Orthology KO database. Queries that are highlighted in red (not in the example data) represent compounds/genes with no matches at all.

The screenshot displays the MetaboAnalyst 4.0 interface. The main title is "MetaboAnalyst-- a comprehensive tool for metabolomics analysis and interpretation". Below the title, there is a navigation menu with options: "Upload ID mapping", "Set parameter", "Network viewer", "Download", and "Exit".

The central part of the interface shows a "Gene Name Mapping" table. Above the table, a note states: "The system requires all the IDs (except common compound names) to be matched exactly. The table below shows the matched genes and compounds from the underlying databases. For common compound names, users can further perform [approximate match](#) by clicking the View link in the Details column. To remove a gene or compound from further analysis, use the Delete link in the last column. Items in the Gene Name Mapping table with **lightgrey** indicates cases where some information is missing while **red** highlights ones that has no matches with in databases."

| Query | Hit | Symbol | KEGG Orthology IDs (KO) | Name | |
|-------|-------|--------|-------------------------|---|--------|
| 1737 | 1737 | DLAT | | dihydropyrimidine S-acetyltransferase | Delete |
| 83440 | 83440 | ADPGK | K08074 | ADP-dependent glucokinase | Delete |
| 3939 | 3939 | LDHA | | lactate dehydrogenase A | Delete |
| 10911 | 10911 | UTS2 | K05248 | ustensin 2 | Delete |
| 10690 | 10690 | FUT9 | K03063 | fucosyltransferase 9 (alpha (1,3) fucosyltransferase) | Delete |
| 10010 | 10010 | TANK | K12660 | TRAF family member-associated NPKB activator | Delete |

On the right side, there is an "R Command History" panel with a "Keep collapsed" button and a "Save" button. It contains a list of R commands used for the analysis, such as:

```

1. InitDataObjects("comp", "network", FALSE)
2. mdet<-setOrganism(mdet, "hse")
3. geneListFile<-"replace_with_your_file_name"
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. mdet<-PerformIntegGeneMapping(mdet, geneList, "hse", "entrez")
6. compListFile<-"replace_with_your_file_name"
7. compList<-readChar(compListFile, file.info(compListFile)$size)
8. mdet<-PerformIntegCompMapping(mdet, compList, "hse", "kegg")
9. mdet<-createMappingResultTable(mdet)
  
```

Users can download the name mapping at the bottom of the tables by scrolling down the page and clicking on the "You can download the result here" link. Click "Submit" to continue.

This screenshot shows the bottom of the mapping table interface. It features a navigation bar with buttons for "Previous", "Next", "Home", "Search", "Help", and "Logout". Below the navigation bar, there is a message: "You can download the result [here](#)". At the bottom center, there is a prominent "Submit" button.

Step 5. Network Selection

Following compound/gene name mapping, we can choose from 1 of 5 networks to explore the data. Currently, the Network Explorer module supports five types of biological networks including a KEGG global metabolic network, a gene-metabolite interaction network, a metabolite-disease interaction network, a metabolite-metabolite interaction network, and a metabolite-gene-disease interaction network. Please note that the last four networks are created based on information gathered from HMDB and STITCH databases, and are applicable to human studies only. Note that there are detailed descriptions beneath each network option. In this tutorial, we will go through using the

“KEGG Global Metabolic Network” and the “Metabolite-Gene-Disease Interaction Network”. To begin, click “[KEGG Global Metabolic Network](#)”.

Network Analysis Options :

Users can choose one of five different modes of networks analysis:

[KEGG Global Metabolic Network](#)
Users can map metabolites and enzymes/KOs (KEGG Orthologs), and then visually explore the results in the KEGG global KEGG metabolic network (ko01100). This feature is specially suitable to integrate results from joint metabolomics and metagenomics studies.

[Gene-Metabolite Interaction Network](#)
The gene-metabolite interaction network enables exploration and visualization of interactions between functionally related metabolites and genes. The chemical and human gene associations were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.

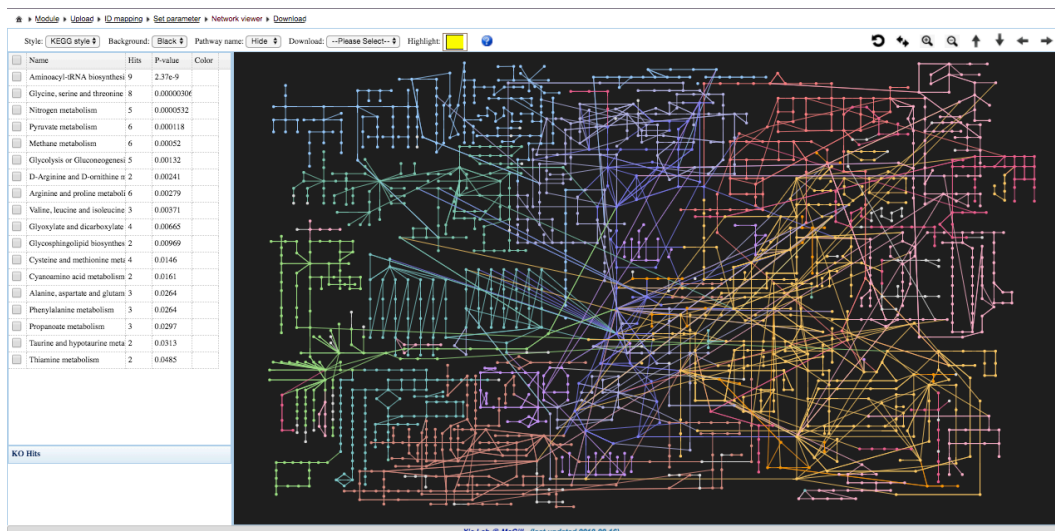
[Metabolite-Disease Interaction Network](#)
The metabolite-disease interaction network enables exploration of disease-related metabolites. The associations were obtained from HMDB. Disease association have been added to HMDB via the Human Metabolome Project's literature curation team.

[Metabolite-Metabolite Interaction Network](#)
The metabolite-metabolite interaction network helps to highlight potential functional relationships between a wide set of annotated metabolites. The chemical-chemical associations for the metabolites network were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.

[Metabolite-Gene-Disease Interaction Network](#)
The metabolite-gene-disease interaction network provides a global view of potential functional relationships

Step 6. KEGG Global Metabolic Network

The screenshot below exemplifies the default view of your data that can be mapped to the KEGG global metabolic network. Here, metabolites will be represented as nodes (circles), and enzymes as edges (lines).

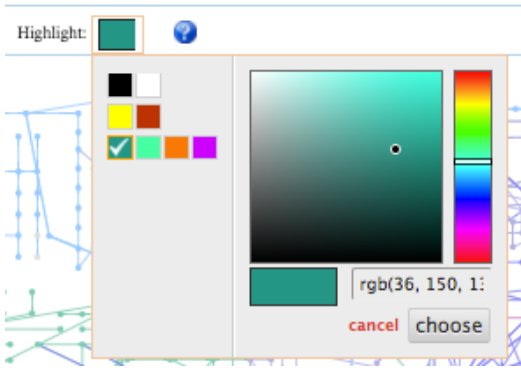


From the screenshot above you can see the global metabolic network page consists of three sections; the **top** section contains a toolbar for user-enabled editing, the **left** section contains the results of pathway analysis performed on your uploaded compounds and/or genes, and the **central** section contains the metabolic network.

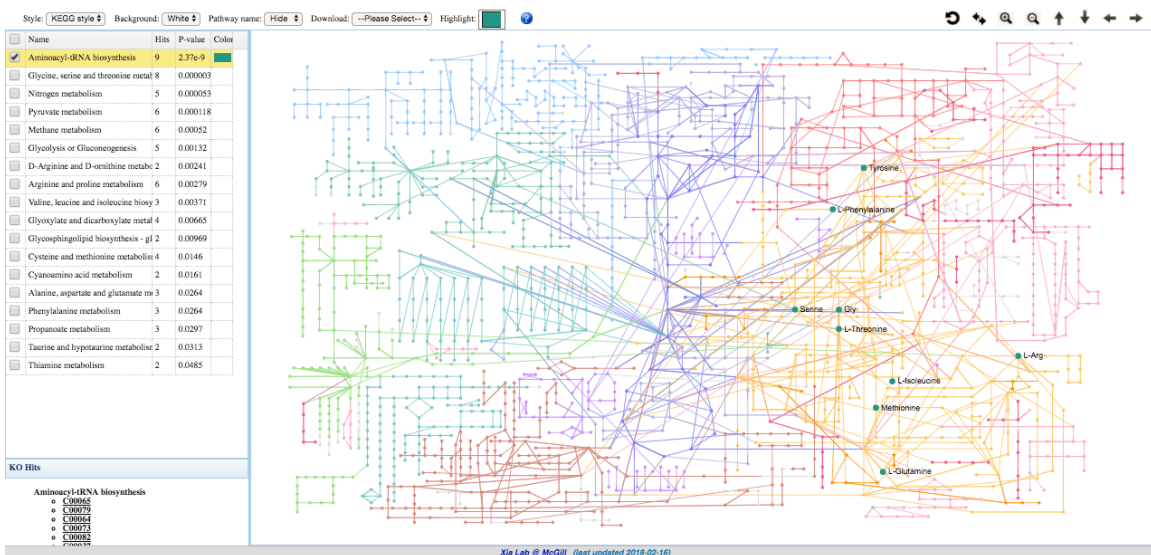
To demonstrate how we can visualize the data, we will map the top 3 enriched pathways onto the KEGG global metabolic network. To start, we will adjust the background color of the network from black to white. Using the toolbar (screenshot below) on the top section of the page, click the drop-down menu next to “Background:” and select “White”. The background color of the network should now be changed to white.



Next, we will change the highlight color by clicking on the yellow colored box next to “Highlight”. From the color palette, as per the screenshot below, you can create any color. In this case, we will create a deep green-blue. Click “chosed” to use this color.

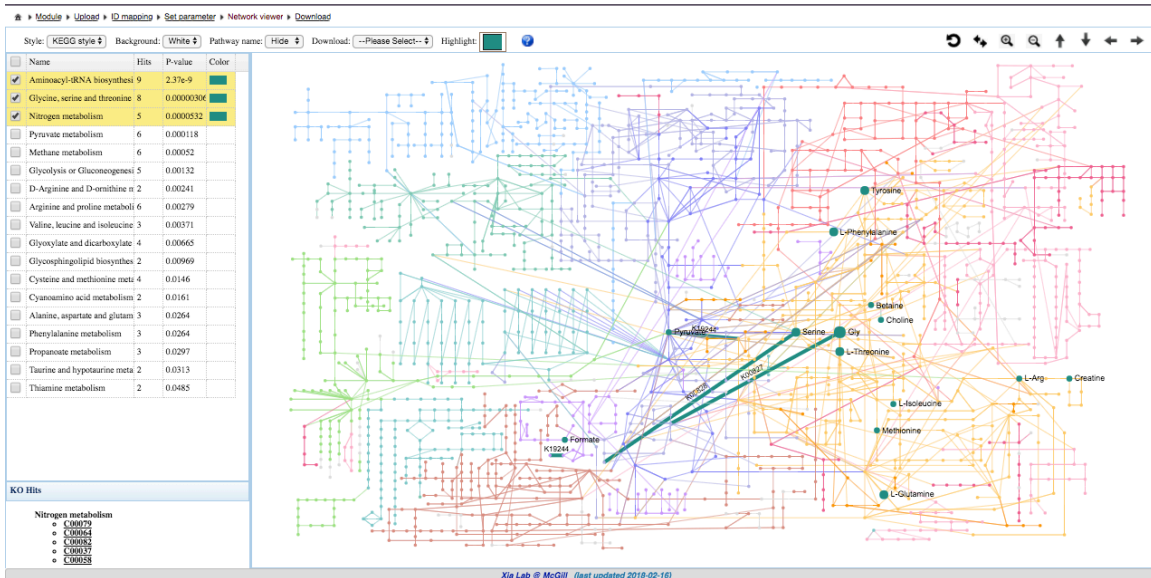


Next, on the left-side panel of the network, click the empty box next to “Aminoacyl-tRNA biosynthesis” to highlight the metabolites/genes that are significantly enriched (screenshot below).



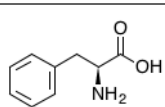
Now we will click on the empty boxes next to “Glycine, serine and threonine” and “Nitrogen Metabolism”. All the hits from these 3 pathways are now mapped onto the network (screenshot below). Note that as you select more pathways, common metabolite hits between the pathways are reflected in the node size. In this case, the Glycine node is bigger than all of the other nodes, informing you that this is a common node amongst the

selected pathways and could be of importance. You can use the scroll-button on your mouse to zoom in/out of the network.



From the screenshot above, the “Hits” box in bottom left corner of the page contains the details of the hits per each selected pathway. For Nitrogen metabolism, if we select “C00079”, a new tab appears that links us to the KEGG page for that compound. Here, we can get further details about this compound that can help to interpret our results.

KEGG **COMPOUND: C00079** [Help](#)

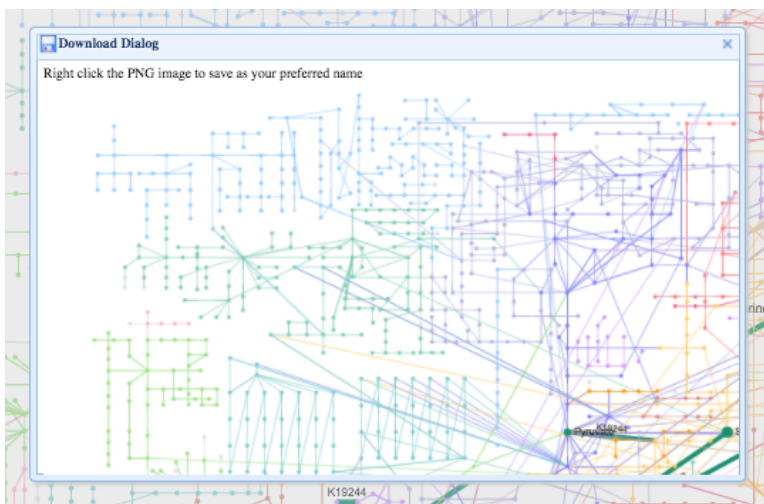
| | | |
|-------------------|---|----------|
| Entry | C00079 | Compound |
| Name | L-Phenylalanine; (S)-alpha-Amino-beta-phenylpropionic acid | |
| Formula | C9H11NO2 | |
| Exact mass | 165.079 | |
| Mol weight | 165.1891 | |
| Structure |  <p>C00079</p> <p>Mol file KCF file DB search Jmol KegDraw</p> | |
| Remark | Same as: D00021 | |
| Reaction | R00686 R00688 R00689 R00690 R00691 R00692 R00693 R00694 R00695 R00697 R00698 R00699 R01375 R01376 R01795 R03660 R06744 R07211 R08463 R08652 R08690 R09535 R09578 R09579 R10286 R10287 R10495 R10499 R10729 R11068 R11918 | |
| Pathway | map00360 Phenylalanine metabolism map00400 Phenylalanine, tyrosine and tryptophan biosynthesis map00460 Cyanoamino acid metabolism map00940 Phenylpropanoid biosynthesis | |

All links

- Ontology (1)
- KEGG BRITE (1)
- Pathway (25)
- KEGG PATHWAY (21)
- KEGG MODULE (4)
- Drug (2)
- KEGG DRUG (1)
- CHEMBL (1)
- Chemical substance (42)
- PubChem (1)
- ChEBI (1)
- 3DMET (1)
- HMDB (1)
- HSDB (1)
- KNAPSACK (1)
- MASSBANK (34)
- NIKKAJI (1)
- PDB-CCD (1)
- Chemical reaction (60)
- KEGG ENZYME (29)
- KEGG REACTION (31)
- Gene (36719)
- KEGG GENES (36719)
- All databases (36849)

[Download RDF](#)

Returning to the network, we can save the created network as a PNG image. Click the drop-down menu next to “Download:” on the toolbar at the top of the page. A Download Dialog box will then appear. Right click the image to save the network under whatever name you prefer.



Step 7. Network Selection

Now that we have covered the KEGG global metabolic network, we will return to the Network Analysis Options page by clicking the “Set Parameter” link at the very top of the page.

Module > Upload > ID mapping > Set parameter > Network viewer > Download

Style: KEGG style Background: White Pathway name: Hide Download: PNG image Highlight:

| <input type="checkbox"/> | Name | Hits | P-value | Color |
|-------------------------------------|-------------------------------|------|------------|--|
| <input checked="" type="checkbox"/> | Aminoacyl-tRNA biosynthesis | 9 | 2.37e-9 | |
| <input checked="" type="checkbox"/> | Glycine, serine and threonine | 8 | 0.00000306 | |
| <input checked="" type="checkbox"/> | Nitrogen metabolism | 5 | 0.0000532 | |

Only highly confident interactions were extracted from STITCH to create the Gene-Metabolite and Metabolite-Metabolite networks. Most of these associations are based on co-mentions highlighted in PubMed abstracts, including reactions from similar chemical structures and molecular activities. The associations for the Metabolite-Disease network were obtained from HMDB. The Metabolite-Gene-Disease network is an integration of

Gene-Metabolite, Metabolite-Disease and Gene-Disease interaction networks. Click on the “[Metabolite-Gene-Disease Interaction Network](#)” to explore relationships between your data that go beyond metabolic pathways.

Network Analysis Options :

Users can choose one of five different modes of networks analysis:

[KEGG Global Metabolic Network](#)
Users can map metabolites and enzymes/KOs (KEGG Orthologs), and then visually explore the results in the KEGG global KEGG metabolic network (ko01100). This feature is specially suitable to integrate results from joint **metabolomics and metagenomics** studies.

[Gene-Metabolite Interaction Network](#)
The gene-metabolite interaction network enables exploration and visualization of interactions between functionally related metabolites and genes. The chemical and human gene associations were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.

[Metabolite-Disease Interaction Network](#)
The metabolite-disease interaction network enables exploration of disease-related metabolites. The associations were obtained from HMDB. Disease association have been added to HMDB via the Human Metabolome Project's literature curation team.

[Metabolite-Metabolite Interaction Network](#)
The metabolite-metabolite interaction network helps to highlight potential functional relationships between a wide set of annotated metabolites. The chemical-chemical associations for the metabolites network were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.

[Metabolite-Gene-Disease Interaction Network](#)
The metabolite-gene-disease interaction network provides a global view of potential functional relationships

Step 8. Mapping Overview

You will now be taken to the “Mapping Overview” page, which provides an overview of the mapping of the example data to the “Metabolite-Gene-Disease Interaction Network”. The mapped metabolites and/or genes (called seeds) are mapped onto the selected interaction network to create subnetworks containing these seeds and their direct neighbors (i.e. first-order subnetworks). This often produces one big subnetwork (“continent”) with several smaller ones (“islands”). Subnetworks with at least 3 nodes will be listed in the table. You will be able to visually explore all of these subnetworks in the next step. As well, these subnetworks can be downloaded as SIF (simple interaction format) files to be explored in other tools (i.e. Cytoscape). Click “[Proceed](#)” to continue to the next step.

Upload
ID mapping
Set parameter
Network viewer
Download
Exit

Mapping Overview

The input metabolites and genes (seeds) are mapped to the selected molecular interaction network to create subnetworks containing these seeds and their direct neighbours (i.e. first-order subnetworks). The procedure often produces one big subnetwork ("continent") with several smaller ones ("islands"). Subnetworks with **at least 3 nodes** are listed below. You can visually explore them in the next step. These subnetworks can be downloaded as SIF (simple interaction format) files to be explored in other tools (i.e. Cytoscape).

| Networks | Nodes | Edges | Seeds | Interactions (SIF) |
|-------------|-------|-------|-------|--------------------------|
| subnetwork1 | 60 | 77 | 28 | Download |
| subnetwork2 | 5 | 4 | 2 | Download |
| subnetwork3 | 5 | 4 | 1 | Download |
| subnetwork4 | 3 | 2 | 1 | Download |
| subnetwork5 | 3 | 2 | 1 | Download |
| subnetwork6 | 3 | 2 | 1 | Download |
| subnetwork7 | 3 | 2 | 1 | Download |
| subnetwork8 | 3 | 2 | 1 | Download |

R Command History

Keep collapsed [Save](#)

```

1. initDataObjects("cont", "network", FA
LSE)
2. mset<-setOrganism(mset, "hse")
3. geneListFile<-replace_with_your_file
_name"
4. geneList<-readChar(geneListFile, fil
e.info(geneListFile)$size)
5. mset<-performIntegGeneMapping(mset, g
eneList, "has", "entrez");
6. compdListFile<-replace_with_your_file
_name"
7. compdList<-readChar(compdListFile, fil
e.info(compdListFile)$size)
8. mset<-performIntegCompdMapping(mset, c
ompdList, "has", "kegg");
9. mset<-createMappingResultTable(mset);
10. mset<-prepareNetworkData(mset);
11. mset<-prepareQuerySeeds(mset);
12. mset<-performNetworkAnalysis_3001100
(mset, "pathway", "network_enrichment
_pathway_3");
13. mset<-searchNetwork(mset, "pheno", "glo
bal", FALSE, 0.5)
14. mset<-createGraph(mset)

```

[Previous](#)
[Proceed](#)

Xia Lab @ McGill (last updated 2018-02-16)

Step 9. Network Explorer View

The screenshot below shows the default view of your data mapped onto the interaction network. Here, the page consists of four sections; the **top** section contains a toolbar for user-enabled editing, the **left** section contains the Node Explorer menu, the **right** section contains the Function Explorer menu, and the **central** section contains the interaction network. In this network, metabolites are represented as diamonds, genes are represented as circles, and diseases are represented as squares. Further, the size of the features corresponds to its node degree, and their color corresponds to its betweenness centrality values (further details below).

Module > Upload > ID mapping > Set parameter > Network viewer > Download

Network: subnetwork1
Background: Black
View: Topology
Layout: Specify
Scope: Specify
Download: Specify
More Options

Node Explorer

| ID | Name | Degree | Betweenness | Expr. |
|--------|------------|--------|-------------|-------|
| 104300 | ALZHEIM | 10 | 908.6 | 0 |
| C00037 | Glycine | 9 | 522.67 | 0.692 |
| C00031 | D-Glucose | 9 | 469.58 | 0.548 |
| 211980 | LUNG CA7 | 7 | 284.71 | 0 |
| 181500 | SCHIZOP7 | 7 | 362.28 | 0 |
| C00077 | Quinine | 6 | 165 | -0.29 |
| C00062 | L-Arginine | 6 | 165 | 0.291 |
| C00065 | L-Serine | 5 | 212.28 | 0.905 |
| C00114 | Choline | 5 | 195.71 | 1.102 |
| C00073 | L-Methion | 5 | 226 | 0.516 |
| C00116 | Glycerol | 5 | 203.28 | 1.011 |
| C00079 | L-Phenylal | 5 | 162.9 | 0.416 |
| 189 | AGXT | 5 | 311.83 | -1.04 |
| C00082 | L-Tyrosine | 4 | 105.9 | -0.30 |
| 125853 | DIABETE4 | 4 | 183.83 | 0 |
| C00188 | L-Threonin | 3 | 64.159 | 0.419 |
| C00064 | L-Glutam | 3 | 38.558 | -0.49 |
| C00207 | Acetone | 3 | 98.967 | -0.93 |
| 63825 | SRB | 3 | 60 | -1.40 |
| C00022 | Pyruvic ac | 2 | 38 | -0.62 |

Current Selections

Function Explorer

Query: All nodes

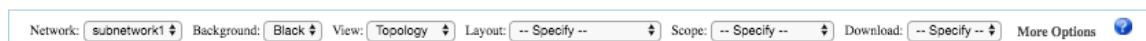
Database: KEGG

Name Hit P-value Col

Path Explorer
Batch Selection

Xia Lab @ McGill (last updated 2018-02-16)

Tips: Within the network, you can use your mouse/scroll-pad to directly drag and drop nodes. Double-click on nodes within the network to highlight them, or select nodes in the Node Explorer on the left-side panel.



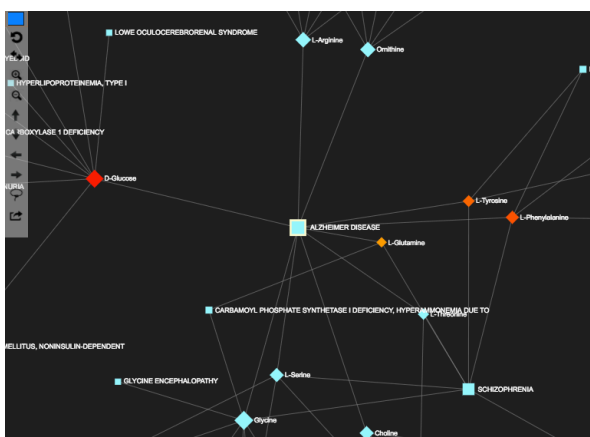
In the toolbar at the top of the page, we can see several options to customize the network. For instance, from the drop-down menu next to “Network”, we can explore all of the different subnetworks that were created in step 8. From the drop-down menu next to “View”, we can change the coloring of the network from “Topology” (default) to “Expression”; nodes will then be colored by their expression levels (if provided). Next, we can change the composition of the network by clicking the drop-down menu next to “Layout”, which will reveal 6 options available for fast alignment of network nodes. Selecting one of these options will automatically re-organize the features. Next, there are a few options to set the “Scope” from the drop-down menu when highlighting and moving the nodes: 1) Single node, to highlight/move only the node being clicked, 2) Node-neighbors, to select a node and its direct neighbors, 3) All-highlights, to select all highlighted nodes and their direct neighbors, and 4) Current function, to select all nodes from a pathway in the Function Explorer.

Step 10. Node Explorer Menu

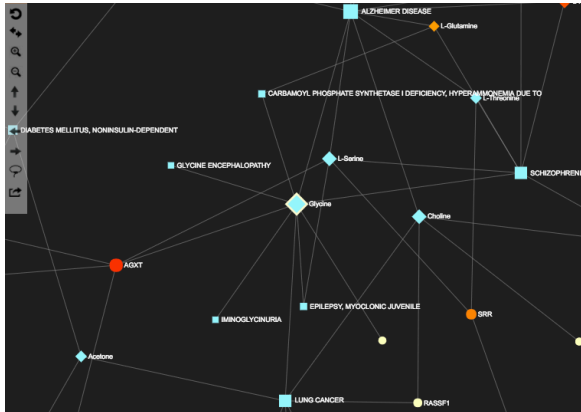
The module provides two popular topological measures found on the left-panel to provide users greater insight into their networks, node degree and betweenness centrality. Node degree refers to the number of links a node has to other nodes, and betweenness centrality represents the degree of centrality a node has in a network by measuring the number of shortest paths that pass through that node. Nodes with high scores in both measures are more likely to be important hubs. Note, you can sort the node table based on either degree or betweenness values by double clicking the corresponding column header.

| ID | Name | Degree | Betweenness | Expr. |
|--------|---------------|--------|-------------|---------|
| 104300 | ALZHEIMER | 10 | 908.6 | 0 |
| C00037 | Glycine | 9 | 522.67 | 0.69285 |
| C00031 | D-Glucose | 9 | 469.58 | 0.54804 |
| 211980 | LUNG CANCER | 7 | 284.71 | 0 |
| 181500 | SCHIZOPHRENIA | 7 | 362.28 | 0 |

In this case, “Alzheimer Disease” seems to be an important node as it has the highest node degree and highest betweenness centrality. If we click on the empty box next to “Alzheimer Disease”, the network automatically zooms into the selected node. From the screenshot below, we can see that all the nodes linked to “Alzheimer Disease” are metabolic compounds. Remember that these links are metabolite-disease associations from HMDB.



The second most important node in this example data is Glycine (screenshot below), according to its node degree and betweenness. When we select it in the Node Explorer, we can see that it is linked to several diseases and two genes (AGXT and GRIN3B). If we google “AGXT and Glycine”, we see that AGXT is responsible for converting glyoxylate to glycine in the peroxisome. Their linking in the network therefore makes sense. Relationships can be further explored by examining the links between the data.



In the bottom left corner of the page (under the Node Explorer box) is the “Current Selections” box, which provides further details of a selected feature. It will provide a link to the corresponding feature’s database such as KEGG, GenBank, or OMIM, as well as the symbol used in the network and its corresponding full name.

Current Selections

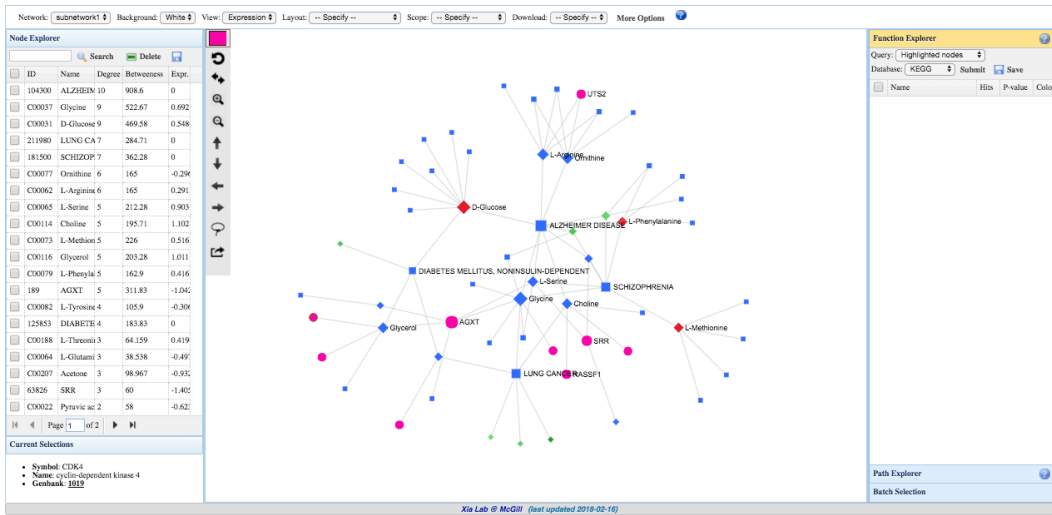
- **Symbol:** AGXT
- **Name:** alanine-glyoxylate aminotransferase
- **Genbank:** [189](#)

Step 11. Function Explorer Menu

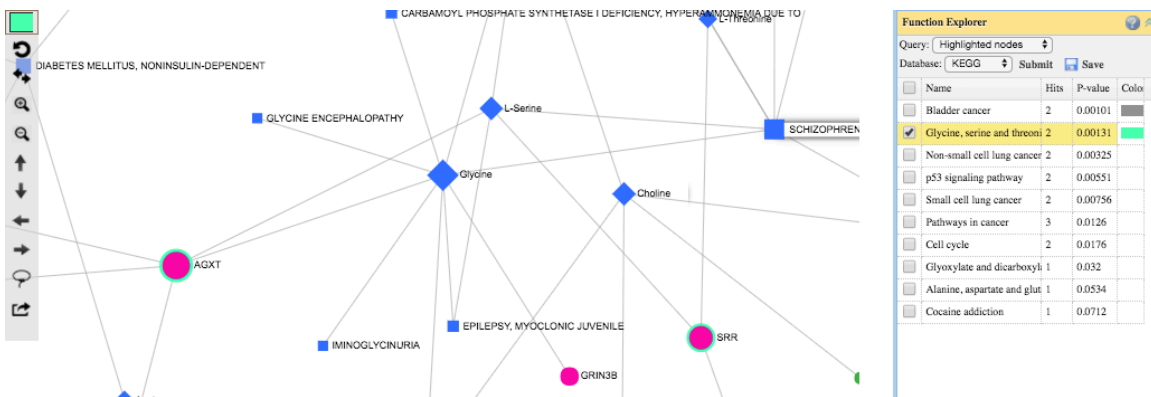
For further functional insights into your data, pathway and enrichment analysis can be performed on the data. The Function Explorer can be found in the right-hand section of the page (screenshot below). For metabolites, you can test for enriched KEGG pathways, and for genes you can test for enriched gene ontologies or pathways (KEGG/Reactome).

To perform functional enrichment analysis (screenshot below), four query options are available under the “Query” drop-down menu: 1) All nodes, 2) Up-regulated nodes, 3) Down-regulated nodes, and 4) Highlighted nodes. Up and down-regulated nodes are based on their expression levels if provided and highlighted nodes are those selected in the Node Explorer menu or highlighted in the network. Next, under “Database”, select a database for the functional enrichment analysis. Remember, only queried genes can be used for the Reactome database. Otherwise, the KEGG pathways can be used for all metabolites and genes. The aim of this is to test whether any functional pathways from the selected database are significantly enriched amongst the selected queries within the

network. Hypergeometric tests are used to compute the enrichment p-values. For this tutorial, we will change the background color of the network to white for better visibility. Then, highlight all the gene nodes using your mouse/scroll-pad to double click the circular nodes, which will now be pink. Then in the Function Explorer, set the query to “Highlighted Nodes” and set the database to KEGG. Click the **Submit** button to perform the analysis.

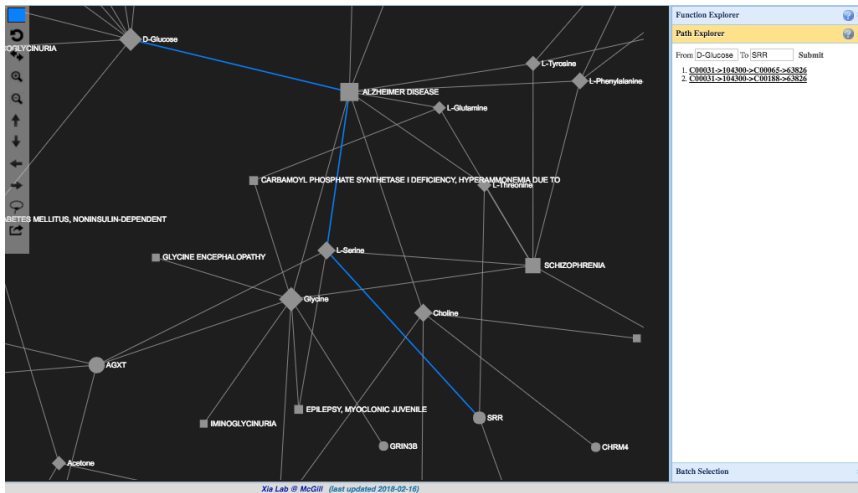


Select the empty box next to the pathways you wish to highlight in the network. For instance, if we select Glycine, serine, and threonine metabolism, the two-hits are highlighted in the network. In this case, the highlighting color was changed to a bright green-blue prior to checking the box, and AGXT and SSR both have two rings surrounding the node (screenshot below).



Step 12. Path Explorer

The path explorer can be used to find the shortest path between any 2 nodes in the network. In this case, let us find the shortest path between D-Glucose and SSR. In the screenshot below, we type “D-Glucose” in the From box, and “SSR” in the To box, and then click “Submit”. There are 2 shortest paths between these two nodes, both crossing 4 nodes. The first path will be highlighted in the network.



Step 13. Batch Selection

First open the “Batch Selection” menu from the bottom right side of the page. Enter a list of node IDs or names, one per row. In this case, enter:

D-Glucose

Choline

LUNG CANCER

Ornithine

L-Arginine

SCHIZOPHRENIA

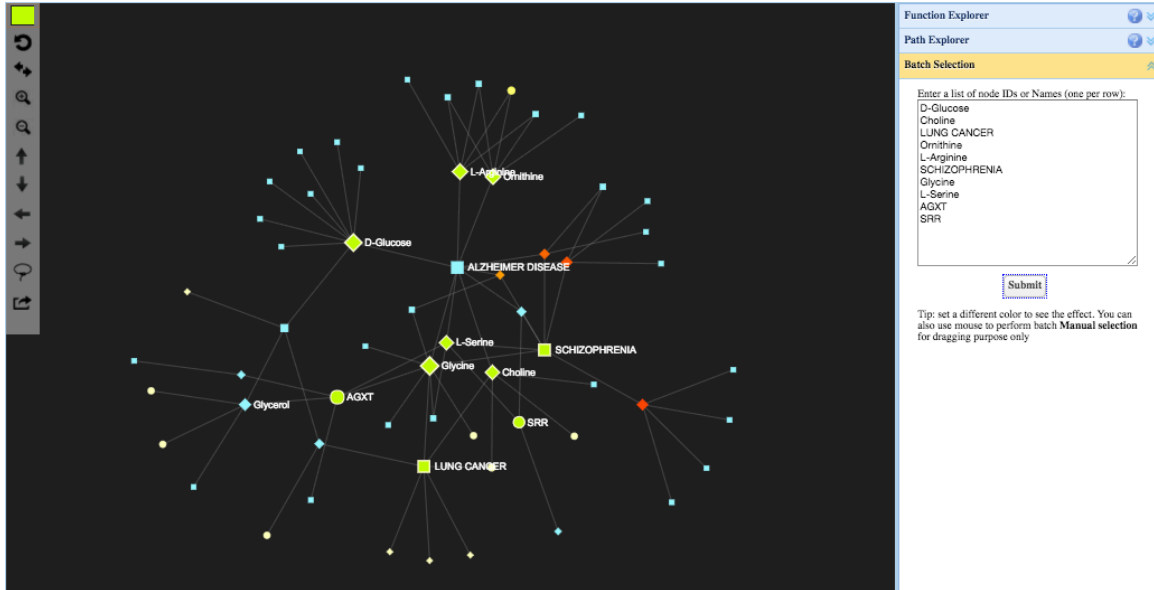
Glycine

L-Serine

AGXT

SRR

Click “Submit” to highlight these selected nodes. As per the screenshot below, the selected nodes are now highlighted in green.



All together, the Network Explorer permits users to emphasize biologically meaningful interactions/associations between nodes, as well as identify important hubs. This suite of editing tools also enables users to create customizable and easily comprehensible networks with publication quality. The integration of network topological analysis, interactive network exploration, and functional enrichment analysis provides users with different views on their data. Interpreting metabolomic data and/or gene expression data in such a context is far more insightful, and will lead to the generation of testable experimental hypotheses.

Step 14. Network Customization

For the remaining of the tutorial, the options for network customization will be demonstrated in the form of questions and answers.

1. What is the network size limit?

The network visualization is limited by the performance of a user's computer and screen resolution. Too many nodes will make the network too dense to visualize and the computer too slow to respond. We therefore recommend limiting the total number of nodes to between 200 ~ 2000 for the best experience. For very large networks, please make sure you have a decent computer equipped with a modern browser (we recommend the latest Google Chrome).

2. How do I delete nodes from the network?

To delete nodes (with their associated edges) from the current network, first select the nodes from the Node Explorer in the left-hand section. Then click the "Delete" button at the top of the Node Explorer table. A confirmation dialog will appear asking if you really want to delete these nodes. Click "Ok" to delete the nodes. Deleting nodes will trigger network re-arrangement, especially if hub nodes are removed. In addition, "orphan" nodes, nodes that are no longer attached to any nodes, may be produced due to removal. These nodes will also be excluded during re-arrangement.

3. Can I create a 300 dpi high-resolution network image for publication purposes?

Use the "Download" option from the top toolbar to download and save your image. There are three options available, PNG, SVG, and GraphML. For publications, select the PNG format. A dialog will appear asking you to right-click the image created link of your image and select "Save link as" to save your image.

4. Can I change the color and size of my nodes?

To change the color of the nodes, select the colored box on the left side of the network image. A palette will appear where you can create the color you want, then click “choose” to use this color. Your next highlighting selections will be in the new color. To change the size of the nodes, double-click the selected node to increase or decrease their size.


5. How can I get node labels to appear?

Node labels appear once the node reaches a certain size. Therefore double-click a node to make it bigger until its labels appear. Remember you can use the “Scope” to select more than just a single node, such as “Node-neighbors” to select a node and its direct neighbors.

6. Can I move a node cluster?

To move a node cluster, change the “Scope” at the top toolbar to “Node-neighbors”, and then drag the central node to move the cluster.

7. Can I extract nodes from the network?

To extract nodes from the network, first highlight the nodes (double-click in the network) or select them in the Node Explorer menu. Next, click the “Extract ”  icon on the left tool bar of the network view window. This will prompt an Extract Confirmation dialogue to appear, asking you to confirm the creation of a new module. Click “Ok” to extract the nodes. The network view will automatically change to the new module, which will be named “moduleX”, and will now be available in the Network drop-down menu.