



MetaboAnalyst 5.0

A Web-based Tool for streamlined
Metabolomics Data Analysis

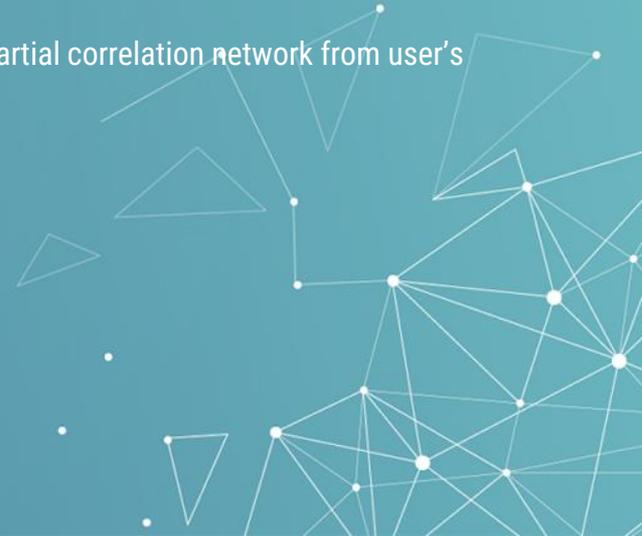
2022.07.12

6. Network Explorer

Network analysis of multi-Omics data could enhance the interpretation on the biological sense with a intuitive way from the system level. The knowledge-based network exploration on the multiple omics data has been implemented since version 4. **Network Explorer** module of MetaboAnalyst has added support for data-driven network analysis from Version 5.

Highlights:

- Added the Debiased Sparse Partial Correlation (DSPC) algorithm to calculate a partial correlation network from user's uploaded data (Basu et al. 2017).



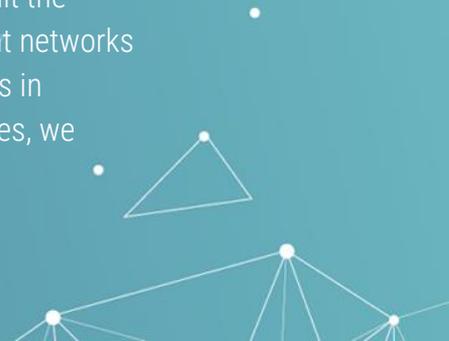
6.0 Knowledge & Background

Knowledge-driven network

- The general concept of knowledge-driven network is to analyze each set of omics data individually and then map the significant features (e.g., metabolites, genes) into the context of our knowledge framework in the forms of networks in order to uncover meaningful links among them, as well as their associations with disease phenotypes ([Zhou et al., 2020](#)).

Data-driven network

- However, knowledge-driven approaches are limited due to an insufficient coverage of the metabolome and lack of knowledge of metabolite-metabolite interactions. Meanwhile, data-driven approaches that permit the inclusion of unknown compounds can overcome these limitations to construct biologically relevant networks and even aid in identifying unknown compounds ([Basu et al., 2017](#)). Therefore, to address concerns in incomplete knowledge of metabolic networks and infer the putative identity of unknown metabolites, we introduce a data-driven network feature in **MetaboAnalyst 5.0**.



6.1 Start Network Explorer

MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down for more details)					
Raw Spectra (mzML, mzXML or mzData)			LC-MS Spectral Processing			
MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
Annotated Features (compound list or table)		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
Generic Format (.csv or .txt table files)	Statistical Analysis	Biomarker Analysis	Time-series/Two-factor Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities

Show R command history

Click here to start

The background is a solid teal color. On the left side, there is a complex network diagram consisting of numerous white nodes (dots) connected by thin white lines. Some nodes are larger than others. To the right of the network, there are several faint, light blue geometric shapes, including triangles and quadrilaterals, scattered across the space. In the upper right corner, there are small, faint white circles of varying sizes, resembling a starry sky or a data visualization.

6.2 Starting from a list

Knowledge-driven network analysis

6.2.1 Data Upload Page – list(s)

TIP: The Fold change is optional. The titles of the 2 columns need to start with '#'.

MetaboAnalyst 5.0

MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

Choose one of the following options to proceed

Lists of genes/compounds | A concentration table

Upload a list of genes (**human only**) or KEGG orthologs, (optional) with a list of metabolites.

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

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

[Try our example](#)

Select the specific tag to match your data ("Lists of genes/compounds" in this case).

Click "**Submit**" to continue.

6.2.2 Name Mapping

MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

ID Mapping
The tables below show ID mapping results based on our databases. To remove

[Compound Name Mapping](#) [Gene Name Mapping](#)

For common compound names, users can further perform approximate match

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	HMDB0000043	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	HMDB0000641	C00064	Delete
C00114	Choline	HMDB0000097	C00114	Delete
C00073	L-Methionine	HMDB0000686	C00073	Delete
C00082	L-Tyrosine	HMDB0000158	C00082	Delete
C00186	(S)-Lactate	-	C00186	Delete
C00037	Glycine	HMDB0000123	C00037	Delete
C00543	Dimethylamine	HMDB0000087	C00543	Delete
C00077	Ornithine	HMDB0000214	C00077	Delete
C00058	Formic acid	HMDB0000142	C00058	Delete
C00188	L-Threonine	HMDB0000167	C00188	Delete
C00407	L-Isoleucine	HMDB0000172	C00407	Delete
C00791	Creatinine	HMDB0000562	C00791	Delete
C00062	L-Arginine	HMDB0000517	C00062	Delete

View compound name mapping or gene name mapping by clicking the corresponding tabs.

Name mapping results from user's data. Scroll down and click "Submit" to continue.

Users can also 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.

```
OK
A total of 163 unique genes were uploaded.

OK
Name matching OK, please inspect (and manual correct) the results then proceed.
1. mSet<-readChar(readLines("your_file_name", FALSE))
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-replace_with_your_file_name
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet, geneList, "hsa", "entrez");
6. cpmlistFile<-replace_with_your_file_name
7. cpmlist<-readChar(cpmlistFile, file.info(cpmlistFile)$size)
8. mSet<-PerformIntegCmpoMapping(mSet, cpmlist, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mSet)
10. mSet<-GetNetworkGeneMappingResultTable(mSet)
```

6.2.3 Network Selection



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis



Upload

GO mapping

Set parameters

Network viewer

Download

Exit

Networks Analysis Options

KEGG Global Metabolic Network

KEGG map (V5) KEGG map (V4)

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

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.

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-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 between metabolites, connected genes, and target diseases. The network is an integration of gene-metabolite, metabolite-disease and gene-disease interaction networks.

Debiased Sparse Partial Correlation (DSPC) Network

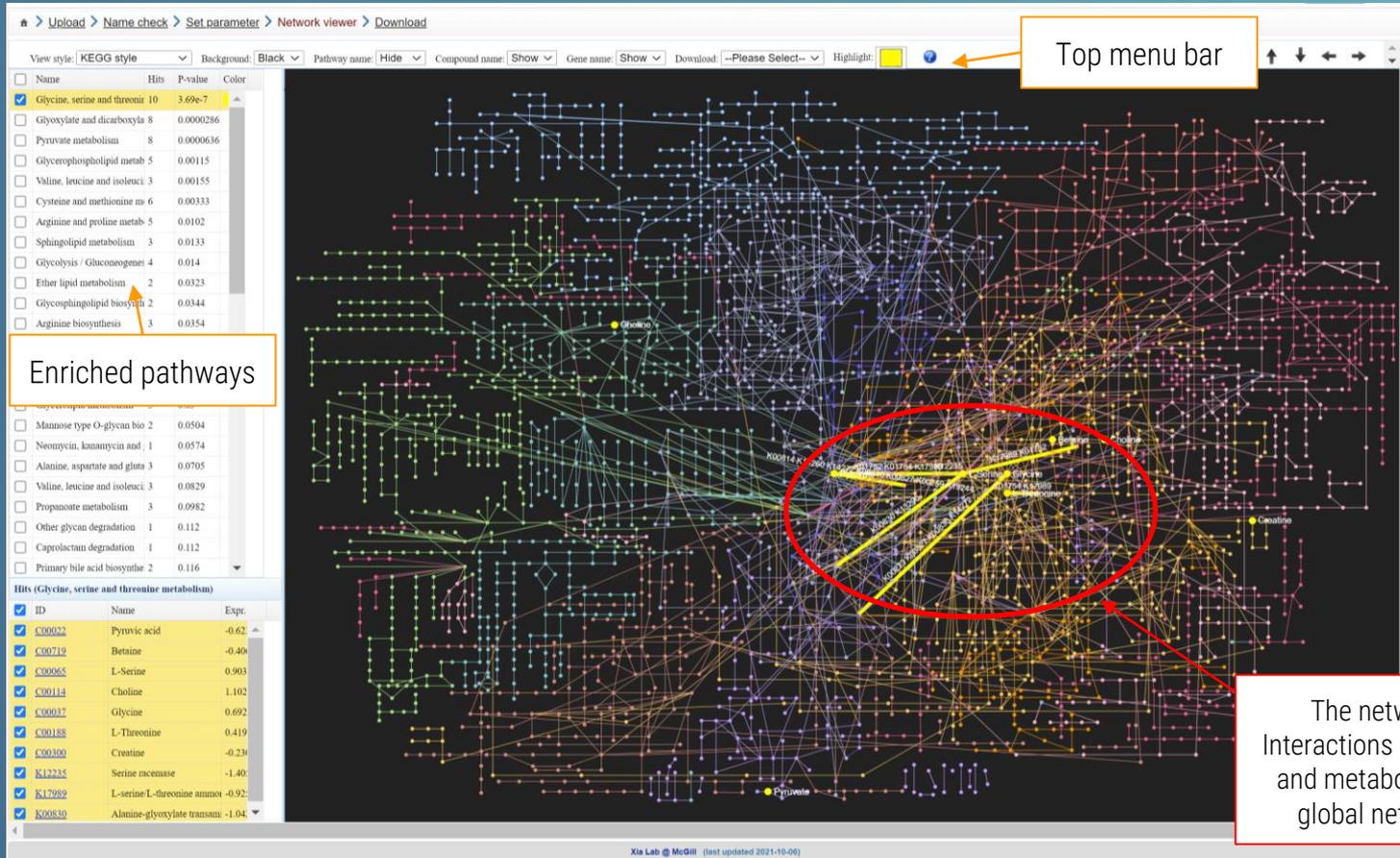
Debiased Sparse Partial Correlation (DSPC) algorithm is based on the de-sparsified graphical lasso modeling procedure (Jankova, 2015). A key assumption is that the number of true connections among the metabolites is much smaller than the available sample size. DSPC reconstructs a graphical model and provides partial correlation coefficients and P-values for every pair of metabolic features in the dataset. Thus, DSPC allows discovering connectivity among large numbers of metabolites using fewer samples (Basu et al., 2017).

Users can choose a network option to explorer the knowledge-based network.

TIP: The KEGG global metabolic network has been updated to the latest version in MetaboAnalyst, but the old version is still being provided for reproducibility with the previous version. This option will be phased out in the future.

In this tutorial, we will mainly demonstrate using the “**KEGG Global Metabolic Network**” and the “**Gene-Metabolite Interaction Network**”. Other parts is working with the same mechanism, and will be introduced briefly.

6.2.4 KEGG Global Metabolic Network



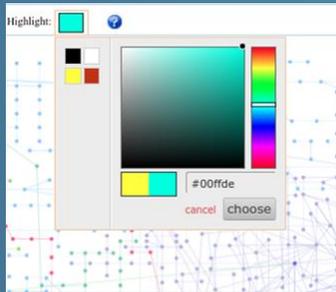
6.2.5 Highlight Enriched Pathways

1) Switch the background color to white



TIP 1: White background is better for publication or reports purposes. Please try to zoom in to find our more details of the interaction.

2) Choose a highlight color



3) Click an enriched pathway to highlight

You can click the link to get further details about the compound.

You can double-click the edges to view the reaction info.

Reaction: 2-Oxoglutarate -> Glycine

ID	Name	Expr.
<input checked="" type="checkbox"/>	C00521	Pyruvic acid -0.62
<input checked="" type="checkbox"/>	C00719	Betaine -0.49
<input checked="" type="checkbox"/>	C00051	L-Serine 0.90
<input checked="" type="checkbox"/>	C00114	Choline 1.10
<input checked="" type="checkbox"/>	C00017	Glycerol 0.69
<input checked="" type="checkbox"/>	C00108	L-Threonine 0.41
<input checked="" type="checkbox"/>	C00030	Creatine -0.28
<input checked="" type="checkbox"/>	K02262	Serine biosynthesis -0.40
<input checked="" type="checkbox"/>	K02262	L-serine L-threonine biosynthesis -0.92
<input checked="" type="checkbox"/>	K00010	Alanine glyoxylate transaminase -1.04

6.2.7 Gene-Metabolite Interaction Network



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

- Upload
 - to metabolomics
 - with parameters
- Network viewer
- Download
- Exit

Networks Analysis Options

KEGG Global Metabolic Network

KEGG map (V5) KEGG map (V4)

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

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.

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-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 between metabolites, connected genes, and target diseases. The network is an integration of gene-metabolite, metabolite-disease and gene-disease interaction networks.

Debiased Sparse Partial Correlation (DSPC) Network

Debiased Sparse Partial Correlation (DSPC) algorithm is based on the de-sparsified graphical lasso modeling procedure (Jankova, 2015). A key assumption is that the number of true connections among the metabolites is much smaller than the available sample size. DSPC reconstructs a graphical model and provides partial correlation coefficients and P-values for every pair of metabolic features in the dataset. Thus, DSPC allows discovering connectivity among large numbers of metabolites using fewer samples (Basu et al., 2017).

Click the "Gene-Metabolite Interaction Network" link.

```
R Command History
Keep collapsed Save
1. mSet<-InitDataObjects("conc", "network
k", FALSE)
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-replace_with_your_file
name"
4. geneList<-readChar(geneListFile, file
info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet, ge
neList, "hsa", "entrez");
6. cmpdListFile<-replace_with_your_file
name"
7. cmpdList<-readChar(cmpdListFile, file
info(cmpdListFile)$size)
8. mSet<-PerformIntegCmpdMapping(mSet, ca
mpdList, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mSet)
10. mSet<-GetNetworkGeneMappingResultTa
ble(mSet)
11. mSet<-PrepareNetworkData(mSet);
```

6.2.8 Network Overview



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis



- Upload
- Go Home
- Get Organisms
- Network viewer
- Download
- Exit

Network Overview

To generate knowledge-based networks, the input metabolites and genes (seeds) are mapped to the selected 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). When the networks are too big or complex for visualization, you can use the **Network Tools** at the bottom to reduce the network size.

Networks	Nodes	Edges	Seeds	Interactions (.SIF)
subnetwork1	35	44	35	Download
subnetwork2	4	3	4	Download

Proceed

Network Tools: ?



Click **"Proceed"** to view the network.

You can also refine the networks by using the network tools.

R Command History

```
1. sSet<-InitDataObjects("conc", "network", FALSE)
2. sSet<-SetOrganism(sSet, "hsa")
3. geneListFile<-"replace_with_your_file_name"
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. sSet<-PerformIntegGeneMapping(sSet, geneList, "hsa", "entrez");
6. cpdListFile<-"replace_with_your_file_name"
7. cpdList<-readChar(cpdListFile, file.info(cpdListFile)$size)
8. sSet<-PerformIntegCpdMapping(sSet, cpdList, "hsa", "kegg");
9. sSet<-CreateMappingResultTable(sSet)
10. sSet<-GetNetworkGeneMappingResultTable(sSet)
11. sSet<-PrepareNetworkData(sSet);
12. sSet<-SearchMetDB(sSet, "pheno", "gene_metabolites", FALSE, 0.5)
13. sSet<-CreateGraph(sSet)
```

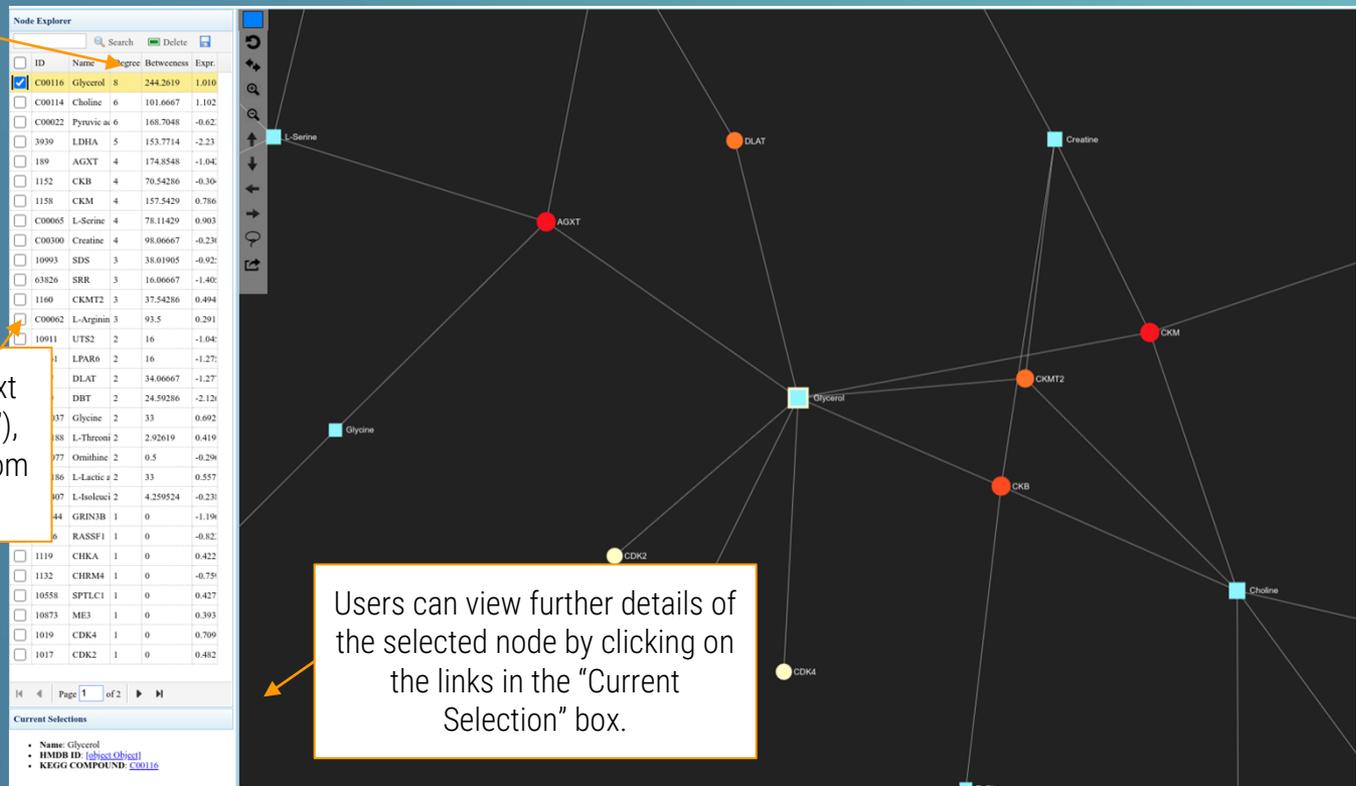
6.2.9 Network Viewer

The screenshot displays the Network Viewer application interface. At the top, the 'Top menu bar' contains options like 'Network viewer', 'Download', and 'Set search'. Below this is the 'Node Explorer' panel, which features a search bar and a table of nodes. The central area is a network graph with nodes and connecting edges. To the right, the 'Function Explorer' panel shows a query dropdown set to 'All nodes' and a table with columns for 'Name', 'Hits', 'P-value', and 'Color'. At the bottom right, the 'Path Explorer' and 'Batch Selection' panels are visible. The 'Batch Selection' panel at the very bottom shows 'Page 1 of 2' and 'Current Selections'.

ID	Name	Degree	Betweenness	Exp.
C00116	Glycerol	8	244.2619	1.010
C00114	Choline	6	101.6667	1.102
C00022	Pyruvic acid	6	168.7048	-0.62
3939	LDHA	5	153.7714	-2.23
189	AGXT	4	174.8548	-1.04
1152	CKB	4	70.54286	-0.36
1158	CKM	4	157.5429	0.786
C00065	L-Serine	4	78.11429	0.903
C00300	Creatine	4	98.06667	-0.23
10993	SDS	3	38.01905	-0.92
63826	SRR	3	16.06667	-1.40
1160	CKMT2	3	37.54286	0.494
C00062	L-Arginin	3	93.5	0.291
10911	UTS2	2	16	-1.04
10161	LPAR6	2	16	-1.27
1737	DLAT	2	34.06667	-1.27
1629	DBT	2	24.59286	-2.12
C00037	Glycine	2	33	0.692
C00188	L-Threoni	2	2.92619	0.419
C00077	Oxalidic acid	2	0.5	-0.20
C00186	L-Lactic acid	2	33	0.557
C00407	L-Isoleuci	2	4.259524	-0.23
116444	GRIN3B	1	0	-1.19
11186	RASSF1	1	0	-0.82
1119	CHKA	1	0	0.422
1132	CHRM4	1	0	-0.79
10558	SPTLC1	1	0	0.427
10873	ME3	1	0	0.393
1019	CDK4	1	0	0.709
1017	CDK2	1	0	0.482

6.2.10 Node Explorer

You can sort the node table by clicking the column header based on either degree or betweenness values.



The Node Explorer interface consists of a table on the left and a network graph on the right. The table lists nodes with their IDs, names, degrees, betweenness values, and expected values. The 'Degree' column is currently sorted in descending order. The network graph shows a central node 'Glycerol' (ID: C00116) connected to several other nodes, including 'L-Serine', 'DLAT', 'Creatine', 'CKM', 'CKMT2', 'CKB', 'Choline', 'Glycine', 'L-Threonine', 'Ornithine', 'L-Lactic acid', 'L-Isoleucine', 'GRINB', 'RASSF1', 'CDK2', 'CDK4', 'CHK2', 'CHKM4', 'SPTLC1', 'ME3', and 'CDK4'. The 'Current Selection' box at the bottom left provides details for the selected node, Glycerol.

ID	Name	degree	Betweenness	Expr.	
<input checked="" type="checkbox"/>	C00116	Glycerol	8	244.2619	1.010
<input type="checkbox"/>	C00114	Choline	6	101.6667	1.102
<input type="checkbox"/>	C00022	Pyruvic ac	6	168.7048	-0.62
<input type="checkbox"/>	3939	LDHA	5	153.7714	-2.23
<input type="checkbox"/>	189	AGXT	4	174.8548	-1.04
<input type="checkbox"/>	1152	CKB	4	70.54286	-0.30
<input type="checkbox"/>	1158	CKM	4	157.5429	0.786
<input type="checkbox"/>	C00065	L-Serine	4	78.11429	0.903
<input type="checkbox"/>	C00300	Creatine	4	98.06667	-0.23
<input type="checkbox"/>	10993	SDS	3	38.01905	-0.92
<input type="checkbox"/>	63826	SRR	3	16.06667	-1.40
<input type="checkbox"/>	1160	CKMT2	3	37.54286	0.494
<input type="checkbox"/>	C00062	L-Arginin	3	93.5	0.291
<input type="checkbox"/>	10911	UTS2	2	16	-1.04
<input type="checkbox"/>	1	LPAR6	2	16	-1.27
<input type="checkbox"/>	DLAT	2	34.06667	-1.27	
<input type="checkbox"/>	DBT	2	24.59286	-2.12	
<input type="checkbox"/>	37	Glycine	2	33	0.692
<input type="checkbox"/>	88	L-Threon	2	2.92619	0.419
<input type="checkbox"/>	77	Ornithine	2	0.5	-0.29
<input type="checkbox"/>	86	L-Lactic a	2	33	0.557
<input type="checkbox"/>	307	L-Isoleuc	2	4.25924	-0.23
<input type="checkbox"/>	44	GRINB	1	0	-1.19
<input type="checkbox"/>	6	RASSF1	1	0	-0.82
<input type="checkbox"/>	1119	CHKA	1	0	0.422
<input type="checkbox"/>	1132	CHRM4	1	0	-0.75
<input type="checkbox"/>	10558	SPTLC1	1	0	0.427
<input type="checkbox"/>	10873	ME3	1	0	0.393
<input type="checkbox"/>	1019	CDK4	1	0	0.709
<input type="checkbox"/>	1017	CDK2	1	0	0.482

Page 1 of 2

Current Selections

- Name: Glycerol
- HMDB ID: [\[Subject Object\]](#)
- KEGG COMPOUND: [C00116](#)

If you click on the empty box next to the ID (e.g., "C00116, Glycerol"), the network will automatically zoom into the selected node.

Users can view further details of the selected node by clicking on the links in the "Current Selection" box.

6.2.11 Function Explorer

Indicate the "Query" and "Database" type and then click the "Submit" to perform functional enrichment analysis

The screenshot shows the Function Explorer interface. On the left, a network diagram displays various nodes and their connections. Nodes are color-coded: yellow (e.g., L-Isoleucine, L-Threonine, Pyruvic acid, L-Serine, AGXT, Glycine, GRIN3B), pink (e.g., DBT, ME3, SRR, SDS, DLAT, CKM, CKB, AKR1A1, CDK4), blue (e.g., Creatinine, L-Lactic acid, Creatine, L-Arginine, Ornithine, D-Glucose, CHRM4, CHKA, RASSF1), and red (e.g., LDHA, CKM2, CKM). A color selection tool is visible in the top-left corner with a callout box that says "Select a color to highlight".

On the right, the "Function Explorer" panel shows the following details:

- Query: All nodes
- Database: KEGG (G)
- Submit and Save buttons
- Table with columns: Name, Hits, P-value, Colo

Name	Hits	P-value	Colo
<input checked="" type="checkbox"/> Glycine, serine and thr...	9	1.27e-8	
<input type="checkbox"/> Arginine and proline m...	7	0.00001	
<input type="checkbox"/> Pyruvate metabolism	5	0	
<input type="checkbox"/> Valine, leucine and iso...	3	0	
<input type="checkbox"/> Glycolysis or Gluconeos...	5	0	
<input type="checkbox"/> Aminoacyl-tRNA biosy...	5	0	
<input type="checkbox"/> Glyoxylate and dicarbo...	4	0	
<input type="checkbox"/> Cysteine and methionin...	4	0	
<input type="checkbox"/> Propanoate metabolism	3	0	
<input type="checkbox"/> Arginine biosynthesis	2	0	
<input type="checkbox"/> Neomycin, kanamycin	1	0	

This close-up shows the "Function Explorer" panel with the following settings:

- Query: All nodes
- Database: All nodes
- Submit and Save buttons
- Table with columns: Name, Hits, P-value, Colo

Name	Hits	P-value	Colo
<input type="checkbox"/> N...			
<input type="checkbox"/> Upregulated nodes			
<input type="checkbox"/> Downregulated nodes			
<input checked="" type="checkbox"/> Highlighted nodes		27e-8	

Below this, the "Database" dropdown is open, showing the following options:

- KEGG (G)
- Reactome
- GO:BP
- GO:MF
- GO:CC
- Motif

The "KEGG (G)" option is selected, and the table below shows the following results:

Name	Hits	P-value	Colo
<input checked="" type="checkbox"/> Gly...	9	1.27e-8	
<input type="checkbox"/> Arg...	7	0.00001	
<input type="checkbox"/> Pyr...	5	0.00008	

6.2.12 Path Explorer

The screenshot displays the Path Explorer interface. On the left, a vertical toolbar contains icons for home, search, zoom, and navigation. The main area shows a metabolic network with nodes and edges. A path is highlighted in blue, starting from CLTC and ending at CDK4. The path consists of the following nodes: CLTC, L-Lactic acid, LDHA, Pyruvic acid, AGXT, Glycerol, and CDK4. On the right, the 'Function Explorer' panel shows the search criteria: 'From: CLTC' and 'To: CDK4'. Below this, a list of five paths is displayed, each with a unique ID and a 'Submit' button. An orange arrow points from the text box to the 'Submit' button of the first path.

Function Explorer

Path Explorer

From: CLTC To: CDK4 Submit

1. [\[213-C00186-3939-C00072-189-C00116-21019\]](#) Submit
2. [\[213-C00186-3939-C00072-1737-C0016-21019\]](#) Submit
3. [\[213-C00186-3939-C00300-1152-C0016-21019\]](#) Submit
4. [\[213-C00186-3939-C00300-1158-C0016-21019\]](#) Submit
5. [\[213-C00186-3939-C00300-1160-C0016-21019\]](#) Submit

Batch Selection

Users can use the “Path Explorer” to find the shortest path between any 2 nodes in the network.

6.2.13 Batch Selection

View: **Topology** Layout: **Specify --** Scope: **All highlights** Download: **Specify --** Save View for Report More Options

Function Explorer
Path Explorer
Batch Selection

Enter a list of node IDs or Names (one per row):

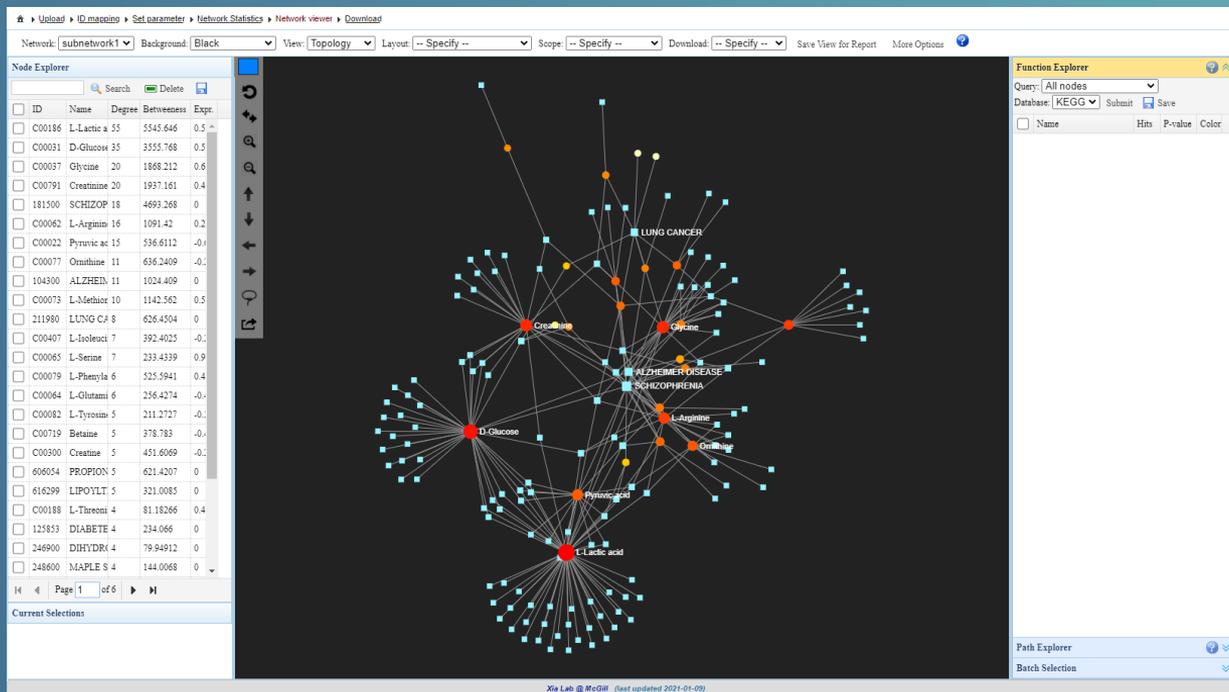
```
C00116  
C00114  
C00022  
3939  
189
```

Submit

Tip: set a different color to see the effect. You can also use mouse to perform batch **Manual selection** for dragging purpose only

Users can use the “Batch Selection” to highlight and drag a list of nodes for further analysis.

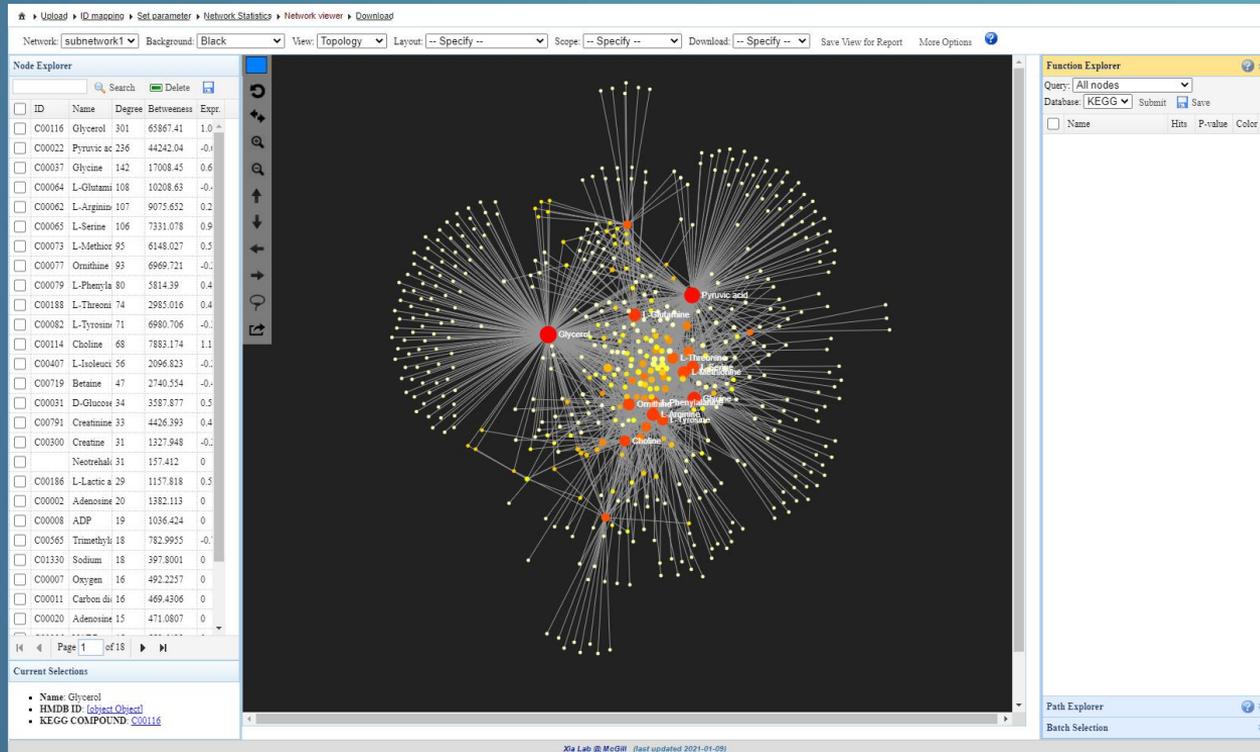
6.2.14 Metabolite-gene-disease interaction



TIP 1: This module is used to show the interactions among the metabolites and disease within a network. Most buttons of this module is working as the modules introduced above.

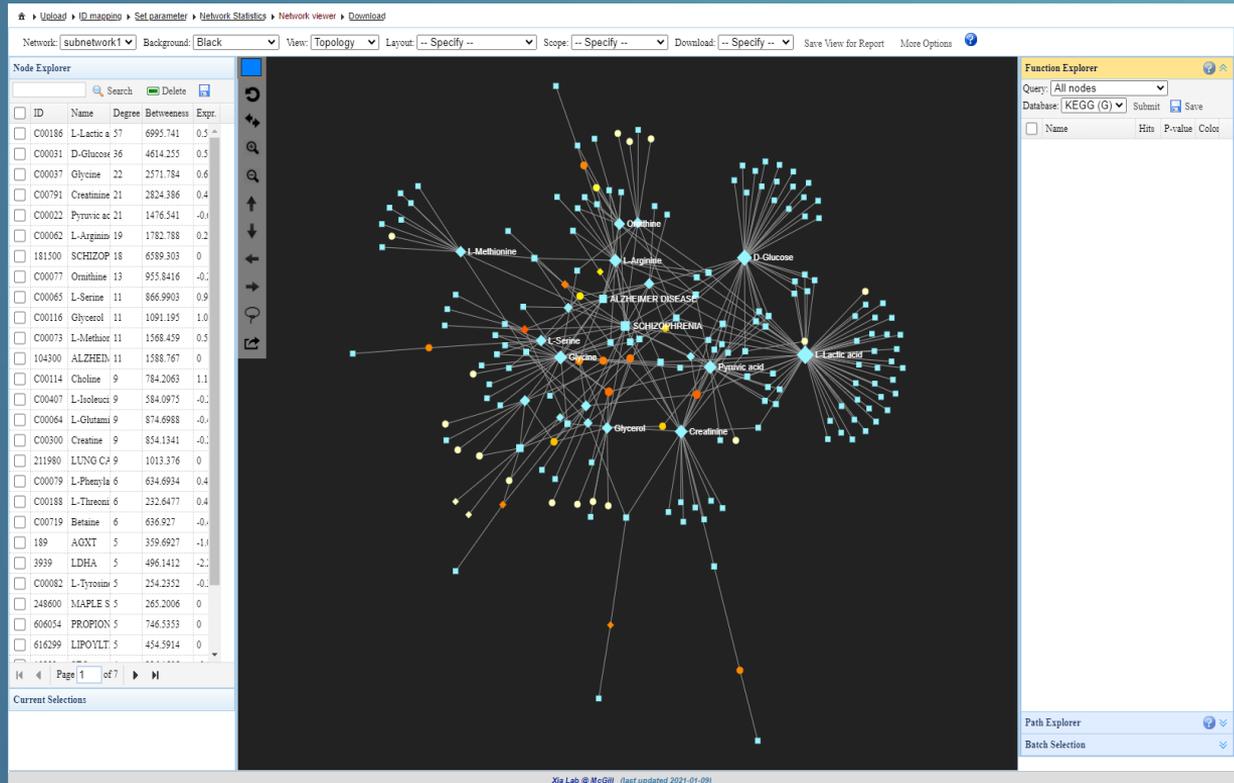
TIP 2: The topological characteristics of different nodes can be ranked by click the header in the 'Node Explorer'. Go and read the [FAQs](#) part to find out more about the introduction of topology.

6.2.15 Metabolite-metabolite interaction



TIP 1: This module is used to show the interactions among the metabolites within a network. Most buttons of this module is working as the modules introduced above.

6.2.16 Metabolite-gene-disease interaction



TIP 1: This module is used to show the interactions among the metabolites, genes and diseases within a network. Most buttons of this module is working as the modules introduced above.



The background is a solid teal color. On the left side, there is a complex network diagram consisting of numerous white nodes (dots) connected by thin white lines (edges). Some nodes are larger than others. Scattered across the background are several faint, light-colored geometric shapes, including triangles and quadrilaterals, some of which are partially filled or outlined. The overall aesthetic is clean and modern, typical of a data science or network analysis presentation.

6.3 Starting from a table

Data-driven network analysis

6.3.1 Data Upload Page – table(s)



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

Choose one of the following options to proceed

[Lists of genes/compounds](#) [A concentration table](#)

Upload your concentration data (.csv or .txt)

ID Type:

Data Format:

Data File: No file chosen

Download from Metabolomics Workbench

Study ID:

Try our test data:

Data	Description
<input checked="" type="radio"/> Concentration table	23 plasma amino acids concentrations from 240 samples measured by GC-MS (Basu S et al.).
<input type="radio"/> Peak intensity table	200 peak lists from 12 mice spinal cord samples measured by LC-MS (Saghatelian et al.).



Upload

Processing

Normalization

1. Upload your data table and specify the format etc.

1. Input the Metabolomic Workbench Study ID to import the data.

2. Click "Submit" to continue.

TIP1: To create a data-driven network, users must upload a concentration table.

TIP2: MetaboAnalyst now allows users to use the study results from Metabolomics Workbench directly by simply providing the STUDY ID.

TIP3: The data pre-processing steps, including data integrity check, ID standardization, and normalization need to be performed step by step.

6.3.2 Name Mapping Page



MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights

- Upload
- Processing
 - Data check
 - Submit check**
 - Missing value
 - Data filter
 - Data editor
- Normalization
- Network
- Download
- Exit

Name mapping results from user's data. Scroll down and click "**Submit**" to continue.

Compound Name/ID Standardization:

- For enrichment analysis, only well-annotated HMDB compounds (i.e. those in our pathway libraries & metabolite sets) will be mapped. For general-purpose name mapping, use **Compound ID Conversion** tool in **Other Utilities** module;
- Greek alphabets are not recognized, they should be replaced by English names (i.e. alpha, beta);
- Query names in normal white indicate **exact match** - marked by "1" in the download file;
- Query names highlighted indicate **no exact or unique match** - marked by "0" in the downloaded file;
- For **compound name**, you should click the **View** link to perform **approximate search** and manually select the correct match if found;
- For **KEGG ID**, it is possible to have multiple hits, you should click the **View** link to manually select the correct match if found;

Query	Hit	HMDB	PubChem	KEGG	Details
Alanine	L-Alanine	HMDB0000161	5950	C00041	
Sarcosine	Sarcosine	HMDB0000271	1088	C00213	
Glycine	Glycine	HMDB0000123	750	C00037	
Alpha-aminoisobutyric acid	2-Aminoisobutyric acid	HMDB0001906	6119	C03665	
Valine	L-Valine	HMDB0000883	6287	C00183	
Leucine	L-Leucine	HMDB0000587	6106	C00123	
Isoleucine	L-Isoleucine	HMDB0000172	6306	C00407	
Threonine	L-Threonine	HMDB0000167	6288	C00188	
Serine	L-Serine	HMDB0000187	5951	C00065	
Proline	L-Proline	HMDB0000162	145742	C00148	
Asparagine	L-Asparagine	HMDB0000168	6267	C00152	
Aspartic acid	L-Aspartic acid	HMDB0000191	5960	C00049	
Methionine	L-Methionine	HMDB0000696	6137	C00073	
4-Hydroxyproline	4-Hydroxyproline	HMDB0000725	5810	C01157	
Glutamic acid	L-Glutamic acid	HMDB0000148	33032	C00025	
Phenylalanine	L-Phenylalanine	HMDB0000159	6140	C00079	
Glutamine	L-Glutamine	HMDB0000641	5961	C00064	
Ornithine	Ornithine	HMDB0000214	6262	C00077	
Lysine	L-Lysine	HMDB0000182	5962	C00047	
Histidine	L-Histidine	HMDB0000177	6274	C00135	
Tyrosine	L-Tyrosine	HMDB0000158	6057	C00082	
Tryptophan	L-Tryptophan	HMDB0000929	6305	C00078	
Cystine	L-Cystine	HMDB0000192	67678	C00491	

```
R Command History
Keep collapsed Save
1. mSet<-InitDataObjects("conc", "networ
k", FALSE)
2. mSet<-SetOrganism(mSet, "hsa")
3. mSet<-Read.TextData(mSet, "Replacing_wi
th your file path", "row", "disc");
4. mSet<-CrossReferencing(mSet, "name");
5. mSet<-CreateMappingResultTable(mSet)
```

6.3.3 Network Parameters



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis



Upload

Processing

Data check

Name check

Missing value

Data filter

Data editor

Normalization

Network

Download

Exit

Networks Analysis Options

KEGG Global Metabolic Network

KEGG map (V5) KEGG map (V4)

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

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.

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-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 between metabolites, connected genes, and target diseases. The network is an integration of gene-metabolite, metabolite-disease and gene-disease interaction networks.

[Debiased Sparse Partial Correlation \(DSPC\) Network](#)

Debiased Sparse Partial Correlation (DSPC) algorithm is based on the de-sparsified graphical lasso modeling procedure ([Jankova, 2015](#)). A key assumption is that the number of true connections among the metabolites is much smaller than the available sample size. DSPC reconstructs a graphical model and provides partial correlation coefficients and P-values for every pair of metabolic features in the dataset. Thus, DSPC allows discovering connectivity among large numbers of metabolites using fewer samples ([Basu et al., 2017](#)).

Click "**Debiased Sparse Partial Correlation Network**" to create the network.

TIP: If you are using table uploading option, DSPC will be enabled. Otherwise, the other options will be enabled.

6.3.4 Network Overview



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis



- Upload
- Processing
 - Data check
 - Name check
 - Missing value
 - Data filter
 - Data editor
 - Normalization
- Network
 - Download
 - Exit

Network Overview

In the Debiased Sparse Partial Correlation (DSPC) network ([Basu et al. 2017](#)), the nodes are input metabolites, while the edges represent the association measures. For better visualization, the default DSPC network only shows the top correlations (edges) based on their p-value rankings (top 20% when the total number of edges is less than 1000 or the top 100 edges when the total number of edges greater than 1000).

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). When the networks are too big or complex for visualization, you can use the **Network Tools** at the bottom to reduce the network size.

Networks	Nodes	Edges	Seeds	Interactions (.SIF)
subnetwork1	22	52	22	Download

Proceed

Click **“Proceed”** to view the DSPC network.

Network Tools: ?



Optional filters to customize the network can be found here.

Correlation Filter

Specify significance cutoff for correlation

P-value cutoff:

Based on: Raw p-val Adj. p-val

Submit

Specify ranges for correlation coefficients

Negative [-1, 0]:

Between -1.0 and 0.0

Positive [0, 1]:

Between 0.0 and 1.0

Submit

6.3.5 Network View

Perform enrichment analysis on selected nodes here.

The screenshot displays a network visualization software interface. On the left is a 'Node Explorer' table with columns for ID, Name, Degree, Betweenness, and Expr. The first three rows are selected. The main area shows a network graph with nodes labeled with amino acids and metabolites, connected by edges of varying thickness and color (red for positive, blue for negative). On the right is a 'Function Explorer' table with columns for Name, Hits, P-value, and Color. A 'More Options' menu is highlighted at the top of the interface.

ID	Name	Degree	Betweenness	Expr.	
<input checked="" type="checkbox"/>	C00123	Leucine	7	31.36014	0
<input checked="" type="checkbox"/>	C00407	Isoleucine	7	22.61046	0
<input checked="" type="checkbox"/>	C00065	Serine	7	21.53698	0
<input type="checkbox"/>	C00152	Asparagin	7	38.16319	0
<input type="checkbox"/>	C00064	Glutamine	7	34.8342	0
<input type="checkbox"/>	C00135	Histidine	7	26.68189	0
<input type="checkbox"/>	C00041	Alanine	6	20.68099	0
<input type="checkbox"/>	C00183	Valine	6	10.46944	0
<input type="checkbox"/>	C00047	Lysine	6	6.282684	0
<input type="checkbox"/>	C03665	Alpha-am	5	15.22143	0
<input type="checkbox"/>	C01157	4-Hydroxy	5	4.683009	0
<input type="checkbox"/>	C00077	Ornithine	5	13.58784	0
<input type="checkbox"/>	C00037	Glycine	4	19.1044	0
<input type="checkbox"/>	C00188	Threonine	4	2.533333	0
<input type="checkbox"/>	C00082	Tyrosine	4	1.033333	0
<input type="checkbox"/>	C00025	Glutamic	4	13.74524	0
<input type="checkbox"/>	C00078	Tryptoph	4	3.472727	0
<input type="checkbox"/>	C00148	Proline	3	0	0
<input type="checkbox"/>	C00213	Sarcosine	2	0	0
<input type="checkbox"/>	C00049	Aspartic	2	0	0
<input type="checkbox"/>	C00079	Phenylal	1	0	0
<input type="checkbox"/>	C00073	Methionin	1	0	0

Name	Hits	P-value	Color
Aminoacyl-tRNA biosynt	18	3.31e-26	
Valine, leucine and isoleuc	4	0.000001	
Alanine, aspartate and glut	5	0.000024	
Arginine biosynthesis	4	0.000024	
Glyoxylate and dicarboxyl	4	0.000747	
Glycine, serine and threon	4	0.000842	
Phenylalanine, tyrosine an	2	0.00107	
Histidine metabolism	3	0.00113	
Arginine and proline meta	4	0.00145	
D-Glutamine and D-glutan	2	0.00263	
Nitrogen metabolism	2	0.00263	
Glutathione metabolism	3	0.00592	
Phenylalanine metabolism	2	0.00762	
Valine, leucine and isoleuc	3	0.014	
Pantothenate and CoA bio	2	0.0269	
beta-Alanine metabolism	2	0.0324	

Network: subnetwork1 Background: Black View: Topology Layout: -- Specify -- Scope: Node-neighbours Download: -- Specify -- Save View for Report More Options

Select network style

View edge colors through the "More Options" menu. Blue edges represent negative correlations while red edges represent positive correlations.

Node information from user's uploaded data table.

Double-click on the edge to view the edge info.

Weighted network visualization of the first subnetwork. The thicker the edge, the stronger the correlation between the features.

Edge: Leucine, Isoleucine
P-value: 8.15e-36
Partial correlation coefficient: 1

Xia Lab @ McGill (last updated 2021-01-08)

6.3.6 DSPC in Statistical Analysis module

TIP: You can do DSPC analysis from network analysis module or statistics module.

The screenshot shows the MetaboAnalyst 5.0 web interface. On the left is a navigation menu with options: Upload, Processing (Data check, Pre-process, Missing value, Data filter, Data editor, Normalization), Statistics (highlighted), Download, and Exit. The main content area is titled 'MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis' and contains the instruction 'Select an analysis path to explore :'. Below this, several analysis categories are listed: Univariate Analysis (with 'Correlation Networks (DSPC)' highlighted in a red box), Chemometrics Analysis, Feature Identification, Cluster Analysis, and Classification & Feature Selection. An orange callout box with an arrow points to the 'Correlation Networks (DSPC)' link, containing the text: 'Users can also perform DSPC network analysis in the **Statistical Analysis** module.'

6.4 Result Downloading



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis



- Upload
- Processing
 - Data check
 - Name check
 - Missing value
 - Data filter
 - Data editor
 - Normalization

Download Results & Start New Journey

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** contains all the files in your home directory. You can also generate a PDF analysis report using the button. Finally, you can continue to explore other compatible modules using the **Start New Journey** tab.

Results Download Start New Journey

Generate Report

Download.zip	networkanalyst_0.json
Rhistory.R	data_normalized.csv
data_processed.csv	data_original.csv
orig_node_list.csv	orig_edge_list.csv
node_table.csv	norm_0_dpi72.png
name_map.csv	snorm_0_dpi72.png

Logout

Click the **“Generate Report”** to download a pdf report summarizing your analysis.

Thanks

*If you have any questions please read through the FAQs or contact us at
[Zhiqiang.pang\[at\]xialab.ca](mailto:Zhiqiang.pang@xialab.ca) or [Jeff.xia\[at\]xialab.ca](mailto:Jeff.xia@xialab.ca)*

