



ConceptMetab User Manual

UNIVERSITY OF MICHIGAN

December 2015

Table of Contents

Terms of Use	2
General Use Policy.....	2
For academic and non-profit institutions.....	2
For commercial and for-profit institutions.....	2
Citation Policy.....	2
Overview	3
What is ConceptMetab?.....	3
What Can ConceptMetab Do?.....	3
Methods Used to Construct ConceptMetab	4
Concept Construction.....	4
Compound Dictionary	4
Enrichment Testing	5
Network Graphs	5
Heatmaps	5
Accessing ConceptMetab	6
Searching on a Concept.....	6
Performing a Concept Search.....	6
Exploring Enrichment Results.....	7
Filtering Enrichment Results	8
By p-value, q-value, or odds ratio	8
By selecting individual concepts.....	8
Creating a Heatmap of Significantly Enriched Concepts	8
Creating Concept Networks	10
Building a Complete Network of Concepts	10
Exploring Relationships within a Network of Concepts.....	11
Creating a Star Network of Concepts	12
Searching on a Compound	13
Performing a Compound Search	13
Creating a Concept Network	14
Building a Complete Network of Concepts	14
Exploring Relationships within a Complete Network of Concepts	15

Terms of Use

General Use Policy

For academic and non-profit institutions

- Permission is granted to access, use and/or download the Tool for internal use only.
- If the user desires to create derivative works of the Tool, source code or access to databases may be made available through request by sending an e-mail to conceptmetab-help@umich.edu.
- Use of the Tool must be acknowledged in resulting publications (see citation policy below).

For commercial and for-profit institutions

- Permission is granted to access, use, and/or download the Tool for internal use only.
- To create derivative works of the Tool for commercial purposes, source code or access to databases may be permitted through negotiation for a commercial license. Please send request to conceptmetab-help@umich.edu

Citation Policy

Please cite: Raymond Cavalcante, Snehal Patil, Terry Weymouth, Alla Karnovsky, and Maureen A Sartor. ConceptMetab: Exploring relationships among metabolite sets to identify links among biological concepts. (In revision) 2015.

Overview

What is ConceptMetab?

ConceptMetab is an interactive, web-based tool for mapping and exploring relationships among thousands of biologically-defined metabolite sets (concepts). It allows you to identify and explain relationships among biological concepts based on metabolites and other compounds. ConceptMetab uses many data sources to define the concepts.

Data Type	Source
Functional annotations	Gene Ontology (GO) Biological Process ¹ GO Cellular Component GO Molecular Function Enzymes Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways ²
Literature-derived annotations	Medical Subject Headings (MeSH) ³ - Anatomy MeSH – Diseases MeSH – Organisms MeSH – Phenomena and Processes MeSH – Psychiatry and Psychology MeSH – Technology, Industry, and Agriculture

ConceptMetab finds metabolite sets that are significantly over-represented among the concept types (the above data sources). Over-representation is tested for using a modified Fisher's Exact Test (FET), and significance is assessed by p-value, q-value, and odds ratios. The q-values are the False Discovery Rate (FDR) transformed p-values to account for the millions of tests performed.

What Can ConceptMetab Do?

With ConceptMetab, you may browse the biological concepts as organized by their data source above. You may also search for a particular concept of interest (e.g. "apoptotic process"), or for a particular metabolite/compound of interest by name or KEGG ID (e.g. "lysine").

Concept Search: After selecting the concept result that best matches your concept search, you can find other concepts that contain a significant number of metabolites in common according to the FET. You may filter by concept type, p-value, q-value, or odds ratio to limit your results. You may then examine the relationship of the concept with its significant partners via networks, a table, or a heatmap.

¹ Harris MA, et al. 2004. The Gene ontology (GO) database and informatics resource. *Nucleic Acids Research* 32(Database issue): D258-D261.

² Kanehisa M, et al. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* 40(Database issue): D109-D114.

³ Coletti MH, Bleich HL. 2001. Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association* 8(4): 317-323.

Compound Search: After selecting the metabolite result that best matches your metabolite search, you are shown a list of concepts containing the metabolite. Once you select a concept of interest, you may do everything described in the previous paragraph.

Methods Used to Construct ConceptMetab

Concept Construction

ConceptMetab defines concepts (metabolite/compound sets) based on many types of biological knowledge (concept types, see data sources above) in order to identify relationships among the concepts in the context of their common metabolites. Data used to construct the metabolite sets are downloaded from a variety of sources, and processed into metabolite sets which are stored in a MySQL database. Concept sizes are limited to between 5 and 4,000 to avoid uninformative concepts. Concept types were built as follows:

KEGG Pathways: KEGG Pathways are based on the summer 2011 freeze (before KEGG FTP became a paid subscription resource), and take the form of XML files containing the Entrez Gene IDs, reactions, and compounds associated with each pathway. Metabolite sets are formed from KEGG Pathways by extracting all compounds from each pathway XML file.

Gene Ontology (GO): Gene Ontology terms (GOBP, GOCC, and GOMF) are associated with Entrez Gene IDs using the *org.Hs.eg.db* R Bioconductor package. We use the KEGG Pathway XML to map compounds to associated genes based on reactions. GO terms are then associated to compounds via genes.

Enzymes: Enzymes are associated with genes according to the *org.Hs.eg.db* R Bioconductor package, and genes are associated with compounds using KEGG Pathway XML (see [Gene Ontology concept construction](#) for more detail).

Medical Subject Headings (MeSH): Compounds are annotated to MeSH terms by their co-occurrences in biomedical literature (PubMed database, version 14, May 19, 2014), using the [Metab2MeSH](#)⁴ database. Therefore, a MeSH concept in ConceptMetab consists of all compounds whose annotations to PubMed articles have significant co-occurrences with that MeSH term.

Compound Dictionary

KEGG Pathways, GO terms, and Enzymes are based on KEGG compound IDs, whereas MeSH terms are based on PubChem IDs. In order to make the concept types comparable, we created a compound dictionary based on Metab2MeSH. KEGG IDs are linked to PubChem substance IDs (SIDs) via the KEGG REST API (<http://www.kegg.jp/kegg/rest/>). The SIDs are linked to PubChem compound IDs (CIDs) via PubChem.

⁴ Sartor MA, et al. 2012. Metab2MeSH: annotating compounds with medical subject headings. *Bioinformatics* 28(10): 1408-1410.

Enrichment Testing

All pairs of concepts from all concept types were tested for overlap using a modified Fisher's Exact Test.⁵ P-values are adjusted for multiple testing by calculating q-values with the FDR method. By default, concepts with FDR < 0.05 are displayed to the user, but options for different p-value, q-value, and odds ratio thresholds are available. Background sets are determined for each concept type as the union of all compounds in all the concepts contained in the concept type. For the FET, the background set used is the intersection of the concept type background sets for the respective concepts being tested.

Network Graphs

Networks displaying the relationship between a query concept and its associated biological concepts (FET with FDR < 0.05, default) are displayed using Cytoscape web. There are two types of network graphs available.

Star Network: This network displays the concept of interest at the center and all the associated concepts (FET with FDR < 0.05), based on the selected concept types, at the periphery. The user may select a concept (node) and view the concept ID, concept type, and compound in the corresponding concept. The user may select an edge in the graph to view the results from the FET on the concepts (nodes) connected by the edge, and the overlapping compounds.

Complete Network: This contains the star network as a subnetwork, but also displays an edge between all concept pairs that meet the p-value / q-value and odds ratio thresholds given by the user (default FDR < 0.05, odds ratio > 0).

Heatmaps

Heatmaps allow the user to graphically view which compounds are responsible for the enrichment of each concept and the relationship among them. Viewing the results in this way can help pinpoint a compound or group of compounds that disproportionately cause concept associations.

⁵ Sartor MA, et al. 2010. ConceptGen: a gene set enrichment and gene set relation mapping tool. *Bioinformatics* 26(4): 456-463.

Accessing ConceptMetab

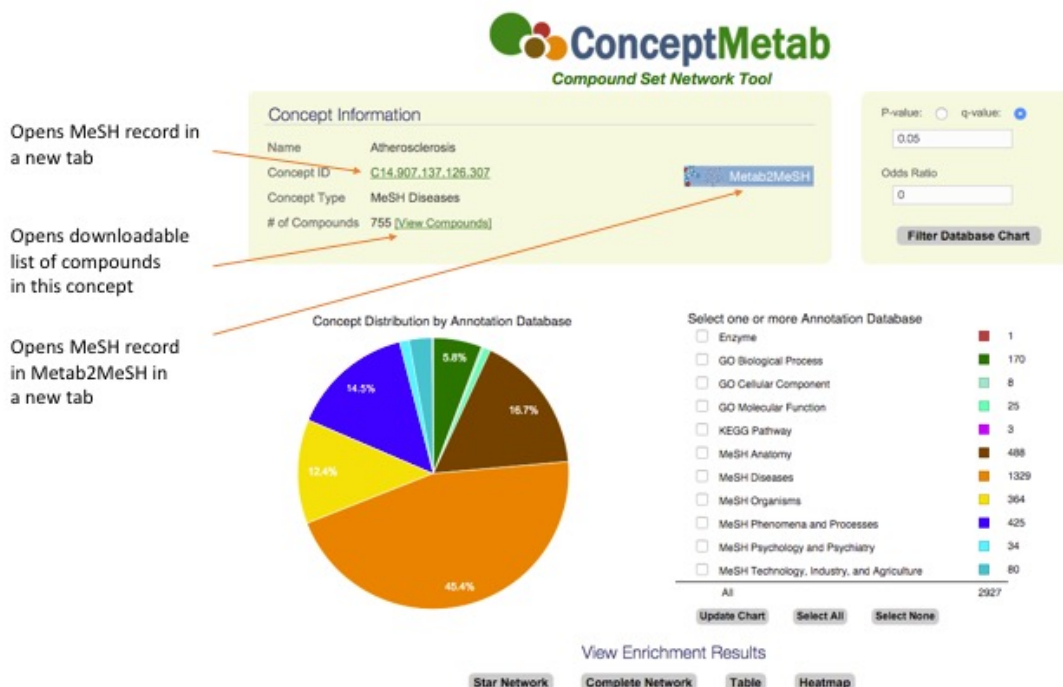

ConceptMetab can be accessed using any web browser at <http://conceptmetab.med.umich.edu/>.

Searching for a Concept

A search can be performed for a concept or for a compound.

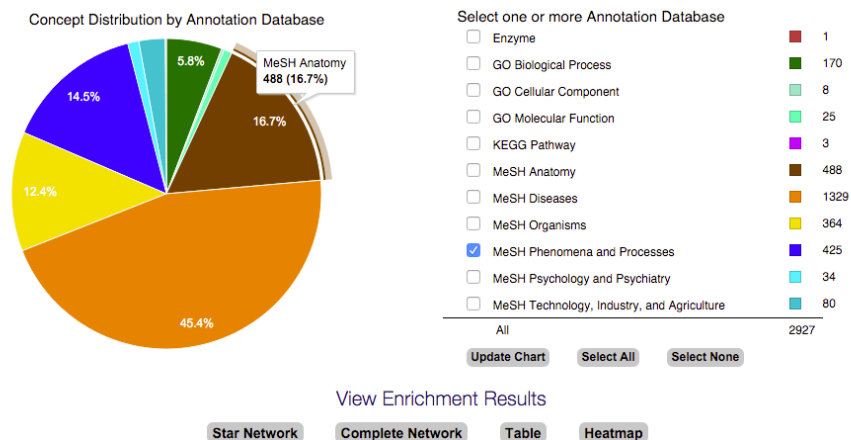
Performing a Concept Search

1. Above the search box, make sure the radio button next to Concept is selected.
2. Checking the box next to Exact Match will perform an exact match search; not checking this box will provide results that do not necessarily match the keywords exactly.
3. In the search box, enter a keyword (for example: atherosclerosis)
4. A dropdown list of predefined concepts will appear as you type. If appropriate, select a matching term.
5. Click the Search button.
6. A list of predefined concepts matching the search criteria will appear.
7. Click on the desired concept.
8. A summary window will appear, showing information about the concept, along with the enrichment results.



Exploring Enrichment Results

1. The pie chart shows the distribution of annotation databases (concept sources) amongst the enrichment results. Mousing over a portion of the pie will show the label for that piece of the pie.



2. Check the boxes next to the annotation database(s) of interest (for example: MeSH Phenomena and Processes).
3. Click on the Table button.
4. A new page opens with a downloadable table of the enriched concepts that belong to the specified annotation database.
5. Click on a column heading to sort by that column.

Concept Information

Concept Name: Atherosclerosis **Concept ID:** C14.907.137.126.307
Concept Type: MeSH Diseases **Number of Compounds:** 755

Input Parameters

q-Value < 0.05 **Odds Ratio** > 0
Selected Annotation Databases:

[CSV](#) [Excel](#)

[Go Back](#) [Star Network](#) [Complete Network](#) [Heatmap](#)

Index	Concept Name	Concept ID	Concept Type	P-value	q-Value	Overlap	Odds Ratio
1	<input type="checkbox"/> Absorption	G02.149.023	MeSH Phenomena and Processes	2.823E-9	6.336E-8	77	2.236
2	<input type="checkbox"/> Acetylation	G02.111.087.019.052	MeSH Phenomena and Processes	1.629E-8	3.466E-7	56	2.465
3	<input type="checkbox"/> Acrosome Reaction	G08.686.785.760.277.800.100	MeSH Phenomena and Processes	4.308E-28	2.435E-26	53	8.635
4	<input type="checkbox"/> Active Transport, Cell Nucleus	G03.495.166.310.100	MeSH Phenomena and Processes	1.38E-20	5.724E-19	59	4.998
5	<input type="checkbox"/> Acylation	G02.111.087.019	MeSH Phenomena and Processes	1.124E-19	4.476E-18	85	3.506

6. This table can be downloaded as a CSV file or an Excel file by clicking on the respective link at the top of the table.
7. Click the Go Back button above the table to return to the summary page containing the pie chart and list of annotation databases.

Filtering Enrichment Results

By p-value, q-value, or odds ratio

1. Several parameters for determining enriched concepts, located at the top right of the concept summary page, can be changed: q-value, p-value, and odds ratio.
2. For example, enter 8 in the text box under Odds Ratio to change it from 0 to 8.

The diagram illustrates the process of changing the Odds Ratio parameter. On the left, a form titled 'Filter Database Chart' has a 'P-value' of 0.05 and an 'Odds Ratio' of 0. A red arrow points to the right, where the same form is shown with the 'Odds Ratio' changed to 8. The 'q-value' radio button is selected in both forms.

3. Click Filter Database Chart.
4. Notice that the number of enriched concepts decreases as a result of the new parameter.

By selecting individual concepts

1. When viewing enrichment results as a table, it is possible to select individual concepts by checking the boxes next to a concept name (see table figure above).
2. Once the desired concepts are selected, users can choose to view a star network, complete network, or heatmap based on the relationships among the selected concepts.

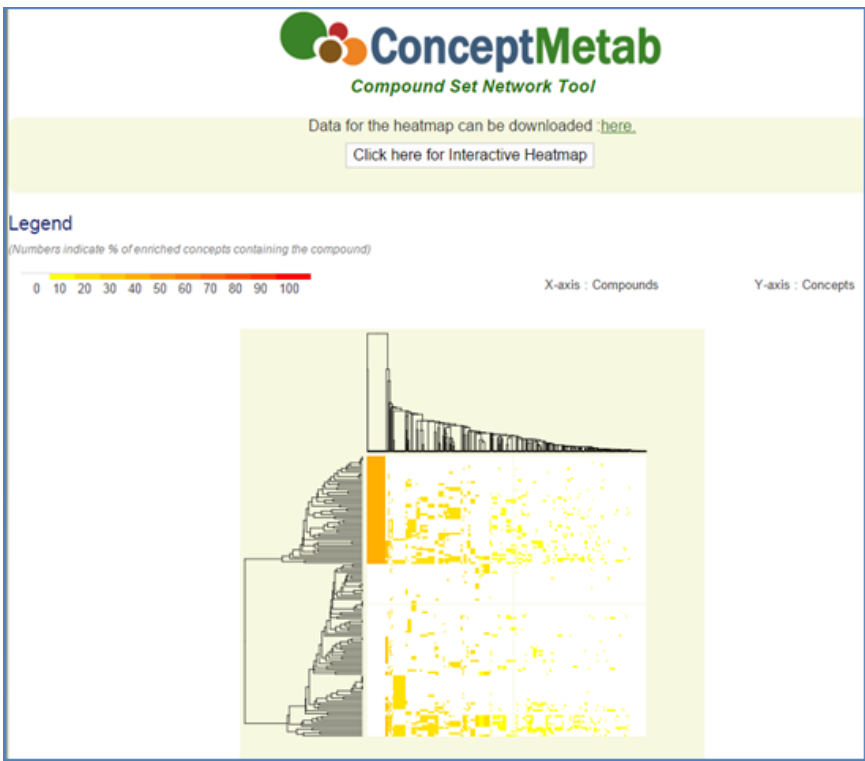
Creating a Heatmap of Significantly Enriched Concepts

1. Search on a concept of interest by following the above steps outlined in the [Performing a Concept Search](#), [Exploring Enrichment Results](#), and [Filtering Enrichment Results](#).
 - a. For example, search on atherosclerosis, and then filter the results with an odds ratio > 8.
2. Check the boxes next to the annotation database(s) of interest (for example: MeSH Phenomena and Processes).
3. Click the Heatmap button at the bottom of the window.

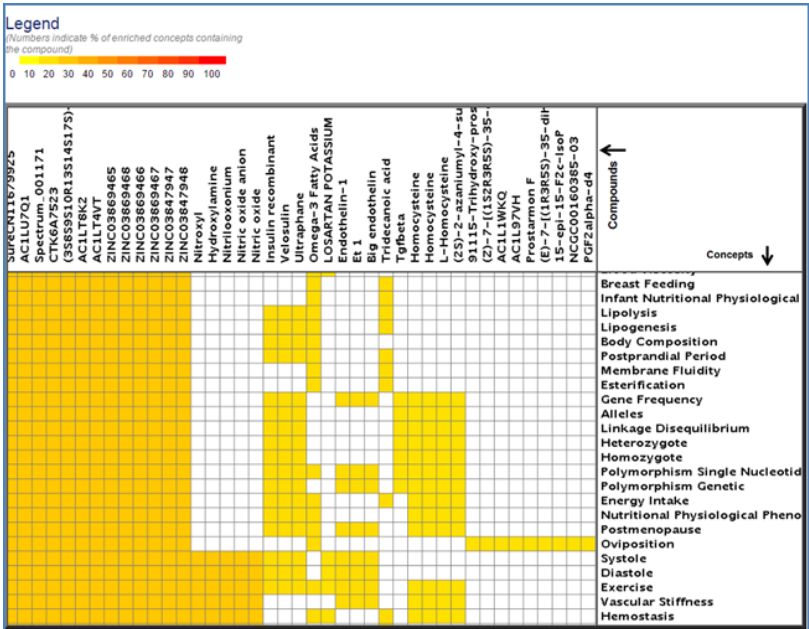
NOTE: When the concept of interest has more than 2000 compounds, up to 200 concepts can be selected for the heatmap. When the concept of interest has between 1000 and 2000 compounds, up to 500 concepts can be selected.

The screenshot shows the 'View Enrichment Results' interface with four buttons: 'Star Network', 'Complete Network', 'Table', and 'Heatmap'. The 'Heatmap' button is highlighted with a red box.

4. A heatmap of the significantly associated MeSH concepts will appear. Compounds are along the x-axis, while concepts are along the y-axis.



5. The heatmap data can be downloaded as a .txt file by clicking the “here” link at the top of the page.
6. To go to the interactive version of the heatmap, click the “Click here for Interactive Heatmap” button at the top of the window. Colored squares indicate which compounds are associated with each concept.



Creating Concept Networks

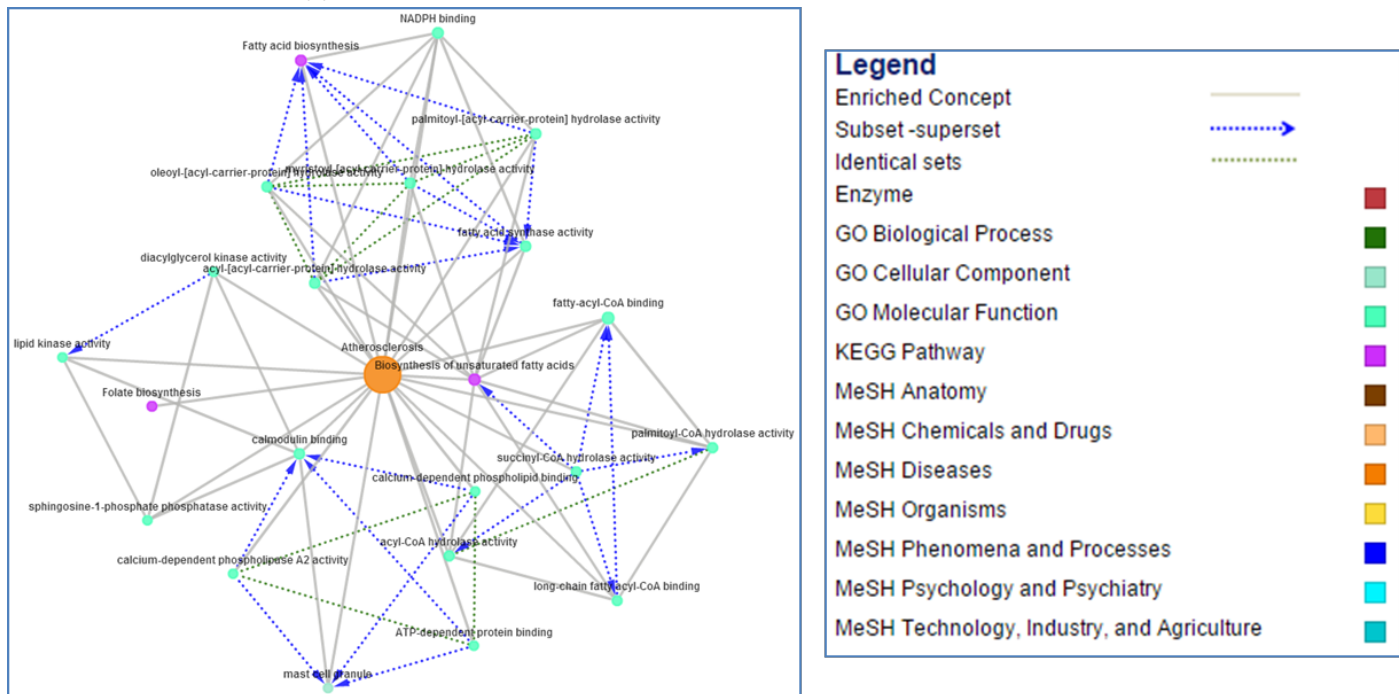
Building a Complete Network of Concepts

A complete network displays all edges, including those connected to the selected concept and those between two enriched concepts. We limit complete networks to 200 concepts for clarity of visualization.

1. Search on a concept of interest by following the above steps outlined in the [Performing a Concept Search](#), [Exploring Enrichment Results](#), and [Filtering Enrichment Results](#) sections.
 - a. For example, search on atherosclerosis, and then filter the results with an odds ratio > 8.
2. Check the box next to the annotation database(s) of interest (for example: GO Cellular Component, GO Molecular Function, and KEGG Pathway).
3. Click on the Complete Network button, located at the bottom of the page.

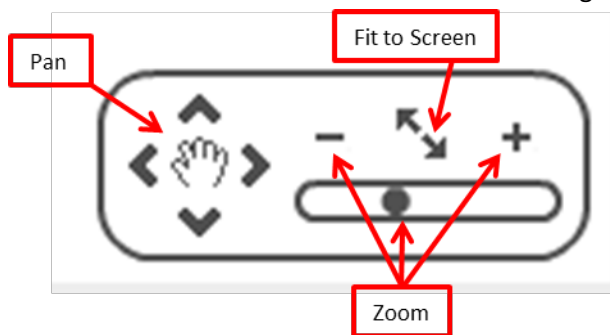


4. A network now appears.



5. A legend is provided for the network
 - a. Nodes represent concepts.
 - b. Edges represent compound overlap between two concepts.
 - c. Node color represents concept type.
 - d. Node size represents concept size.
 - e. Solid grey edges represent enriched concepts.
 - f. Dotted green edges represent identical sets.
 - g. Dotted blue arrow edges represent subset-superset.
6. Nodes can be repositioned by clicking and dragging them to a desired location.

- Pan and zoom tools are located at the bottom right of the screen.



Exploring Relationships within a Network of Concepts

- Click on an edge to explore the overlap between two concepts (for example, click on the edge between Atherosclerosis and NADPH binding).
- A popup window containing information about the edge and a list of compounds shared between the two concepts will appear.
 - Clicking on a compound name will run a search in ConceptMetab on that compound.
 - Clicking on a PubChem ID opens a new browser tab with the PubChem record for that compound.
 - Clicking on a KEGG ID opens a new browser tab with the KEGG Compound record for that compound.

Edge Information

Concept 1:	NADPH binding
Concept 2:	Atherosclerosis
P-value	2.762E-3
q-Value	3.535E-2
Odds Ratio	9.589E0
Intersection	5

Overlapping Compounds

[CSV Excel](#)

Index	Compound Name	Pubchem ID	Kegg Id	# of Concepts
1	Dihydrobiopterin	133246:	C00268:	259
2	Methylglyoxal	880:	C00546:	413
3	Prostaglandin F2alpha	5280363:	C00639:	380
4	Tetradecanoic acid	11005:	C06424:	438
5	Tetrahydrobiopterin	44257:	C00272:	274

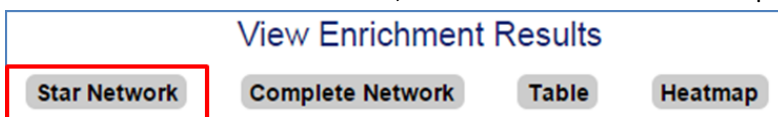
- Click on a concept node to get more information about the concept.
- A popup window containing information about the concept will appear.

Concept Information	
Name	Atherosclerosis
Concept-Id	C14.907.137.126.307
# of Compounds	755
Concept Type	MeSH Diseases

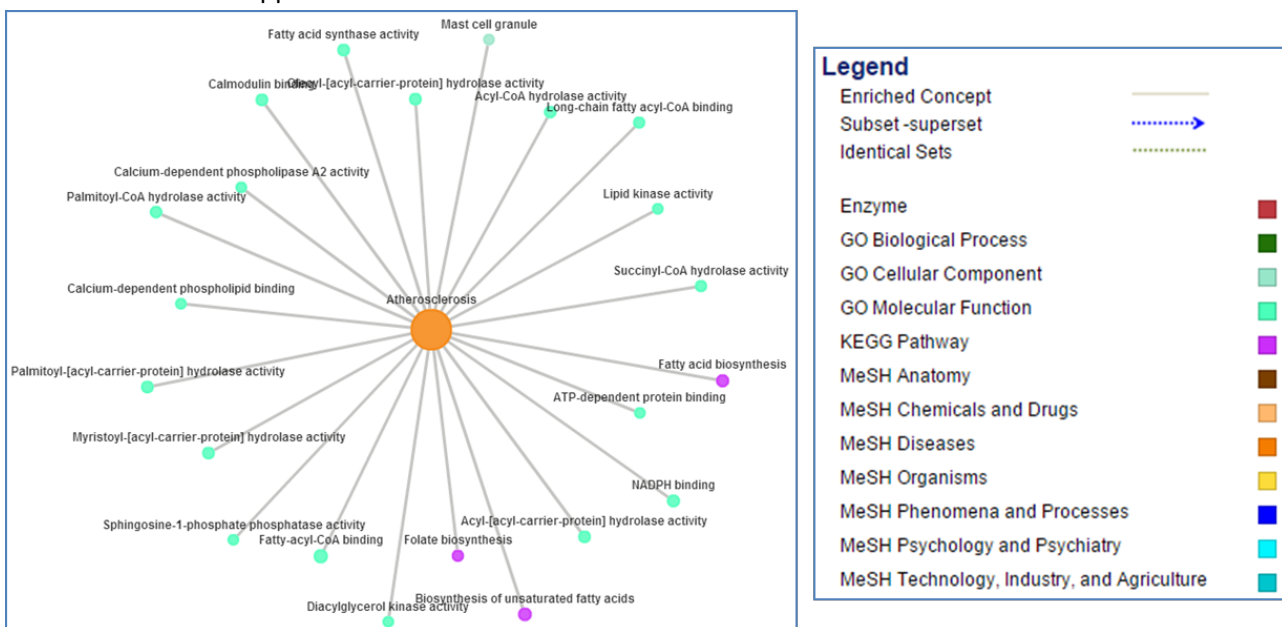
Creating a Star Network of Concepts

A star network displays only edges connected to the selected concept.

1. Search on a concept of interest by following the above steps outlined in the [Performing a Concept Search](#), [Exploring Enrichment Results](#), and [Filtering Enrichment Results](#) sections.
 - a. For example, search on atherosclerosis, and then filter the results with an odds ratio > 8.
2. Check the box next to the annotation database(s) of interest (for example: GO Cellular Component, GO Molecular Function, and KEGG Pathway).
3. Click on the Star Network button, located at the bottom of the page.

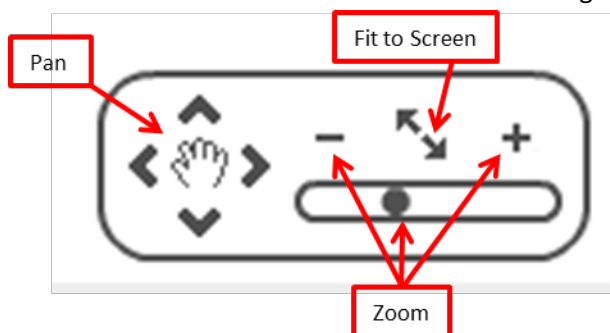


4. A star network now appears.



5. A legend is provided for the network
 - a. Nodes represent concepts.
 - b. Edges represent compound overlap between two concepts.
 - c. Node color represents concept type.
 - d. Node size represents concept size.
 - e. Solid grey edges represent enriched concepts.
 - f. Dotted green edges represent identical sets.

- g. Dotted blue arrow edges represent subset-superset.
- 6. Nodes can be repositioned by clicking and dragging them to a desired location.
- 7. Pan and zoom tools are located at the bottom right of the screen.



- 8. To explore a star network, see the above section titled [Exploring Relationships within a Network of Concepts](#) as the information applies to both complete networks and star networks.

Searching on a Compound

A search can be performed on a concept or on a compound.

Performing a Compound Search

1. Above the search box, make sure the radio button next to Compound is selected.
2. Checking the box next to Exact Match will perform an exact match search; not checking this box will provide results that do not necessarily match the keywords exactly.
3. In the search box, enter a keyword (for example: gamma-hydroxybutyrate, where the hyphen is necessary), KEGG ID, or PubChem ID.
4. A dropdown list of matching terms will appear as you type. If appropriate, select a matching term.
5. Click the Search button.

6. A list of compounds matching the search criteria will appear.
7. Click on the desired compound.
8. A summary window will appear, showing information about the compound, along with the concepts that include the compound.

Compound Information

Name: 4-Hydroxybutanoate, gamma-hydroxybutyrate, GHB
 Kegg ID: --
 Pubchem ID: [3037032](#)
 # of Concepts: 108

Pubchem Structure

CSV Excel

Select All Select None Complete Network

Search:

Index	Concept Name	Concept ID	Concept Type	Concept Size	# of Enrichments
1	<input type="checkbox"/> Brain	A08_186_211	MeSH Anatomy	3581	3603
2	<input type="checkbox"/> Central Nervous System	A08_186	MeSH Anatomy	914	3092
3	<input type="checkbox"/> Cerebral Cortex	A08_186_211_730_885_287_500	MeSH Anatomy	2548	3982
4	<input type="checkbox"/> Corpus Striatum	A08_186_211_730_885_287_249_487	MeSH Anatomy	1782	2728
5	<input type="checkbox"/> Hair	A17_360	MeSH Anatomy	673	3023
6	<input type="checkbox"/> Hippocampus	A08_186_211_464_405	MeSH Anatomy	2381	3296

- The concepts table can be downloaded as a CSV file or an Excel file by clicking on the respective CSV or Excel links at the top of the table.

Creating a Concept Network

Building a Complete Network of Concepts

A complete network displays all edges between a node and the selected concept and between all pairs of enriched concepts.

- Search on a compound of interest by following the above steps outlined in the [Performing a Compound Search](#) section.
 - For example, search on gamma-hydroxybutyrate.
- Click on a column heading to sort alphabetically or numerically by that column.
- Check the boxes next to the concepts of interest (for example: Central Nervous System, Alcoholism, and Brain Ischemia).
- Click on the Complete Network button, located above the concepts table.

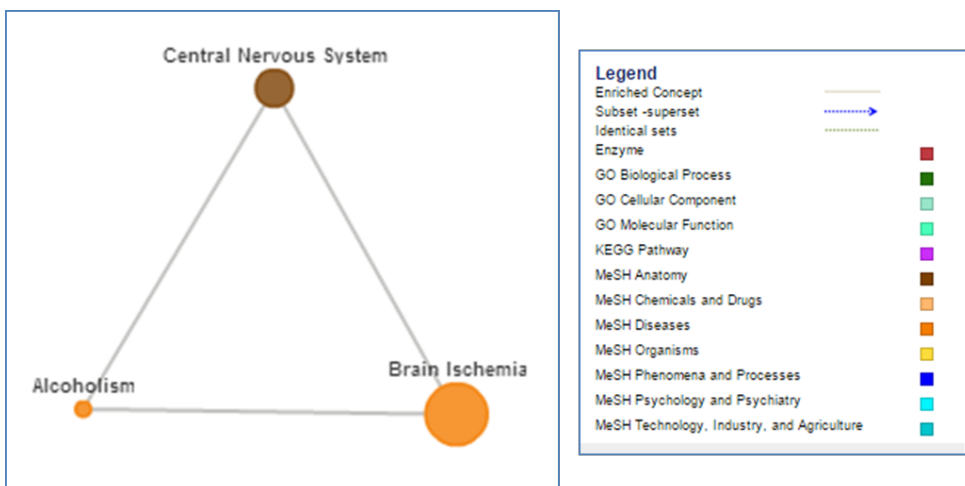
CSV Excel

Select All Select None Complete Network

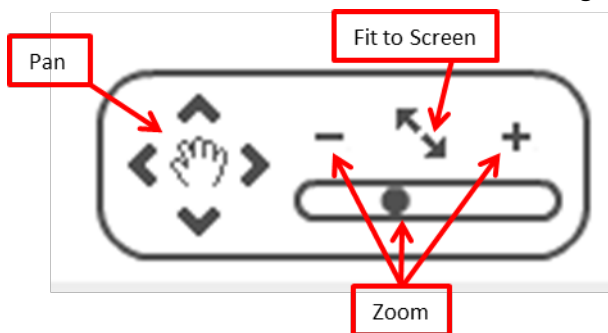
Search:

Index	Concept Name	Concept ID	Concept Type	Concept Size	# of Enrichments
9	<input type="checkbox"/> Akathisia, Drug-Induced	C25_100_249	MeSH Diseases	333	1221
85	<input type="checkbox"/> Alcohol Drinking	F01_145_317_269	MeSH Psychology and Psychiatry	552	2179

- A network now appears.



6. A legend is provided for the network
 - a. Nodes represent concepts.
 - b. Edges represent compound overlap between two concepts.
 - c. Node color represents concept type.
 - d. Node size represents concept size.
 - e. Solid grey edges represent enriched concepts.
 - f. Dotted green edges represent identical sets.
 - g. Dotted blue arrow edges represent subset-superset relationships, where the arrow points to the superset.
7. Nodes can be repositioned by clicking and dragging them to a desired location.
8. Pan and zoom tools are located at the bottom right of the screen.



Exploring Relationships within a Complete Network of Concepts

5. Click on an edge to explore the overlap between two concepts (for example, click on the edge between Brain Ischemia and Alcoholism).
6. A popup window containing information about the edge and a list of compounds shared between the two concepts will appear.
 - a. Clicking on a compound name will run a search in ConceptMetab on that compound.
 - b. Clicking on a PubChem ID opens a new browser tab with the PubChem record for that compound.

- c. Clicking on a KEGG ID opens a new browser tab with the KEGG Compound record for that compound.

Edge Information			
Concept 1:	Brain Ischemia		
Concept 2:	Alcoholism		
P-value	2.427E-2		
q-Value	2.647E-1		
Odds Ratio	1.468E0		
Intersection	35		
Overlapping Compounds			
Index	Compound Name	Pubchem ID	Kegg Id
1	dipyridamole	3108	-
2	piracetam	4843	-
3	Clomethiazole	10783	-
4	Dopamine hydrochloride	65340	-
5	MEMANTINE HYDROCHLORIDE	181458	-
6	4-aminobutanoic acid hydrochloride(1:1)	210329	-
7	AC1L9P8R	450881	-
8	Tocris-0773	1263681	-
9	N-Acetyl-L-aspartate	2042020	-
10	4-Hydroxybutanoate.gamma-hydroxybutyrate.GHB	3037032	-
11	intropin	3713809	-
12	SBB061255	4068592	-
13	2-acetamidobutanedioate	4217221	-
14	4-aminobutanoate	5460232	-
15	AC1O7GG9	6604403	-
16	AC1OCL95	6941837	-

7. Click on a concept node to get more information about the concept.
8. A popup window containing information about the concept will appear.

Concept Information	
Name	Brain Ischemia
Concept-Id	C10.228.140.300.150, C14.907.253.092
# of Compounds	1,195
Concept Type	MeSH Diseases

Additional Network Visualization: MetScape

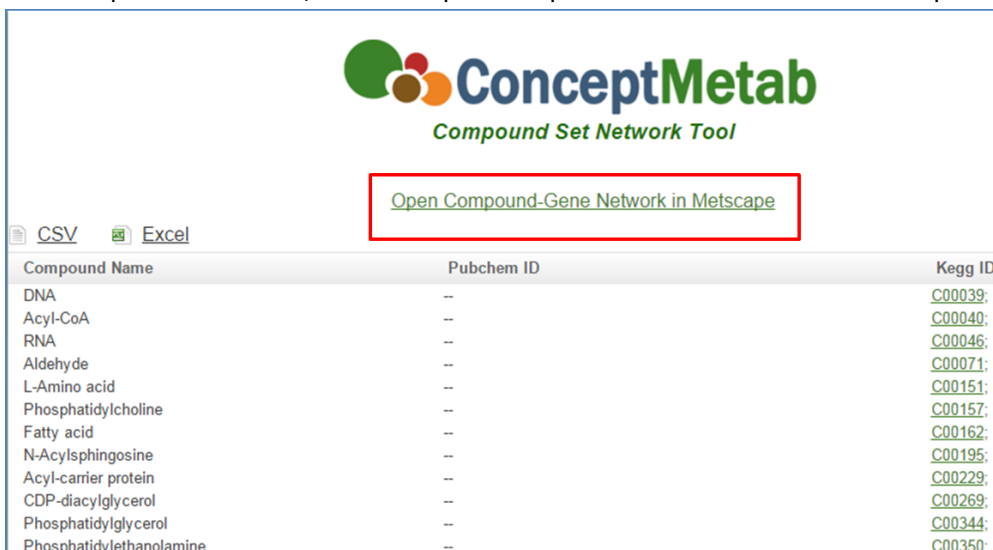
A list of metabolites that make up a concept can be visualized in [MetScape](#)⁶, a [Cytoscape](#)⁷ app for exploring relationships between compounds, reactions, enzymes, and genes within the context of human metabolism. This option is available for concepts where at least three compounds have KEGG IDs.

1. Search on a concept of interest by following the above steps outlined in the [Performing a Concept Search](#) section.
 - a. For example, do an exact match search on membrane

⁶ Karnovsky A, et al. 2012. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* 28(3): 373-380.

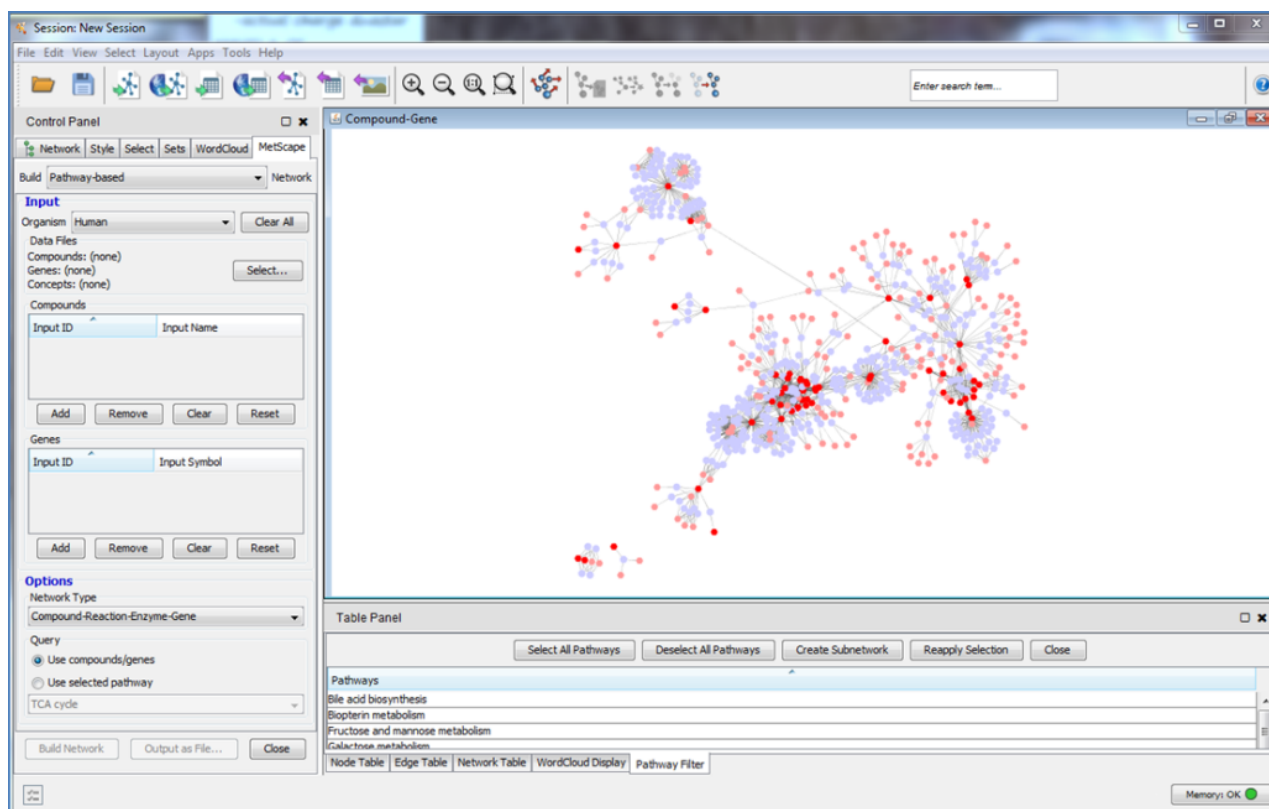
⁷ Shannon P, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11): 2498-2504.

2. Click on the View Compounds link, located next to the # of Compounds in the Concept Information section.
3. A list of compounds that make up the concept will now appear.
4. At the top of the window, click the Open Compound-Gene Network in MetScope link



Compound Name	Pubchem ID	Kegg ID
DNA	--	C00039
Acyl-CoA	--	C00040
RNA	--	C00046
Aldehyde	--	C00071
L-Amino acid	--	C00151
Phosphatidylcholine	--	C00157
Fatty acid	--	C00162
N-Acylsphingosine	--	C00195
Acyl-carrier protein	--	C00229
CDP-diacylglycerol	--	C00269
Phosphatidylglycerol	--	C00344
Phosphatidylethanolamine	--	C00350

5. A cytoscape.jnlp file will download.
6. Opening the cytoscape.jnlp file will launch Cytoscape with a MetScope-built network based on the specified list of compounds from ConceptMetab. See the [MetScope Help & Support](#) webpage for more information about using MetScope.
7. The below image is of the MetScope-built network.



Session: New Session

File Edit View Select Layout Apps Tools Help

Control Panel

Network Style Select Sets WordCloud MetScope

Build Pathway-based Network

Input

Organism: Human [Clear All]

Data Files

Compounds: (none) [Select...]

Genes: (none)

Concepts: (none)

Compounds

Input ID Input Name

[Add] [Remove] [Clear] [Reset]

Genes

Input ID Input Symbol

[Add] [Remove] [Clear] [Reset]

Options

Network Type: Compound-Reaction-Enzyme-Gene

Query

Use compounds/genes

Use selected pathway

TCA cycle

[Build Network] [Output as File...] [Close]

Compound-Gene

Table Panel

[Select All Pathways] [Deselect All Pathways] [Create Subnetwork] [Reapply Selection] [Close]

Pathways

Bile acid biosynthesis

Biopterin metabolism

Fructose and mannose metabolism

Galactose metabolism

[Node Table] [Edge Table] [Network Table] [WordCloud Display] [Pathway Filter]

Memory: OK

