

NCIBI Workshop

Module 4

Exploring Metabolomic Data Analysis Tools



National Center for Integrative Biomedical Informatics (NCIBI)

The University of Michigan Medical School

Palmer Commons Room 2017 Palmer Commons

100 Washtenaw Avenue Ann Arbor, MI 48109-2218

Phone: 734-615-5510 Fax: 734-615-6553

web site: <http://www.ncibi.org/>

email: ncibiinfo@umich.edu

Getting Started with Michigan Metabolomics Tools

MiMI Web with Metabolome and MetScape

Introduction

MiMI Web with Metabolome and Metscape are the tools built by the National Center of Integrative Biomedical Informatics (NCIBI) to help you analyze and interpret metabolomic data. Using these tools, you can decipher molecular networks underlying various biological conditions.

MiMI Web (Michigan Molecular Interactions) with Metabolome is a query and analysis interface to the MiMI database (Tarcea et al., 2009). Metscape is a plug-in for Cytoscape that lets you query the MiMI database, bring in your own experimental data and compounds, and graphically display their metabolic reactions and associated pathways (<http://www.cytoscape.org/>). Both tools use [KEGG](#) (Kanehisa et al., 2008) and [EHMN](#) (Ma et al., 2007) human metabolic networks data, stored locally in a Microsoft SQL Server database, to infer the relationships between compounds, chemical reactions, pathways, enzymes and genes.

With these tools you can upload a list of metabolites or genes to MiMI-Metabolome or to Metscape, identify reactions, genes and pathways that are associated with those, and explore their relationships. You can also import normalized experimental metabolite data and display them within the network of metabolic reactions.

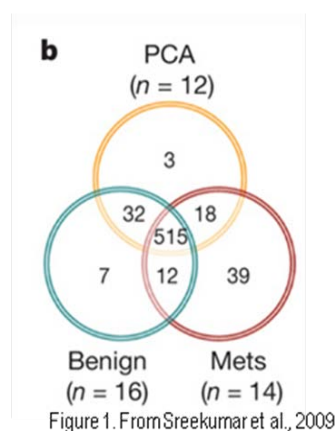
Biological Problem and Exploratory Purpose

Deciphering the molecular networks that distinguish localized prostate cancer from metastatic disease represents a novel systems biology approach to identification of critical biomarkers for cancer invasion and disease aggressiveness. At the University of Michigan, a study performed in Arul Chinnayian's lab focused on deriving the metabolomic profiles of benign prostate, clinically

localized prostate cancer and metastatic disease (Sreekumar et al., 2009). The exploratory case described in this tutorial is based on that study. We will demonstrate how you can use the NCIBI tools to quickly annotate the data described in the paper.

Sreekumar and co-authors found that benign samples, localized cancer and metastatic tissues shared 515 metabolites (Fig. 1).

In benign tissue vs. localized tumors, 87 metabolites were changed significantly ($p < 0.05$). 36 out of 87 were known compounds. 12 compounds were down-regulated and 24 were up-regulated. We will use this set of 36 compounds in this tutorial.



In metastatic samples vs. localized tumors, 226 metabolites were differential ($p < 0.05$). Out of the 226 metabolites, 124 were increased and 102 were decreased in metastatic samples. 81 metabolites out of 226 were identified.

Notably, the researchers found that six metabolites (sarcosine, kynurenine, glycerol-3-phosphate, leucine, proline and uracil) were significantly increased on disease progression from benign to localized to metastatic prostate cancer. These compounds are potential biomarkers for progressive disease. In fact, the authors showed that sarcosine is a key metabolite increased robustly in metastatic prostate cancer and is detectable in urine of men with organ-confined disease.

Objectives

In this Tutorial You Will

Part 1: Focus on sarcosine

- ▶ Learn how to query on a list of compounds, using the list of 36 compounds that are differential between benign samples and localized prostate cancer
- ▶ Focus on and examine one compound – sarcosine from the query results
- ▶ Explore the reactions and pathways that involve the query compounds
- ▶ Identify the enzymes/genes that are related to the query compounds

Part 2: Explore visualizations of compound and pathway networks

- ▶ Visualize networks of compounds, using the MetScape Cytoscape plugin/plugin.
- ▶ Upload experimentally measured values for of the set of 36 compounds
- ▶ Visualize the compounds in the context of metabolic pathways

Data

Access the two files dataset at <http://portal.ncibi.org/gateway/virtual-workshop.html>. It contains:

A text list of 36 compounds – known metabolites, differential between benign prostate and PCA samples (Sreekumar et al., 2009).

A text file of experimental measurements for the 36 compounds in the required format for importing data into Metscape.

Getting Started: Query on Your List of Compounds

1. Download the text file of 36 compounds to your hard drive using the link above.
2. Open MiMI Web with Metabolome: <http://mimi.ncibi.org/MimiWeb/main-page.jsp> Go the List Search tab. The Search page appears for entering your query.

Figure 2. MiMI search page

3. Load your compound list by clicking Browse under "Upload Gene or Compound List." Find the text file and select it.
4. Click Copy to Text. MiMI automatically enters the list under "List of genes to search for."
5. Under Select Type (next to the "List of genes to search for") click the Cids (compound IDS) radio button.
6. Leave the "Limit Search by Organism" box on Homo sapiens. Click the MiMI Search button.

The Summary Results Screen lists the compounds you entered.

Search Results

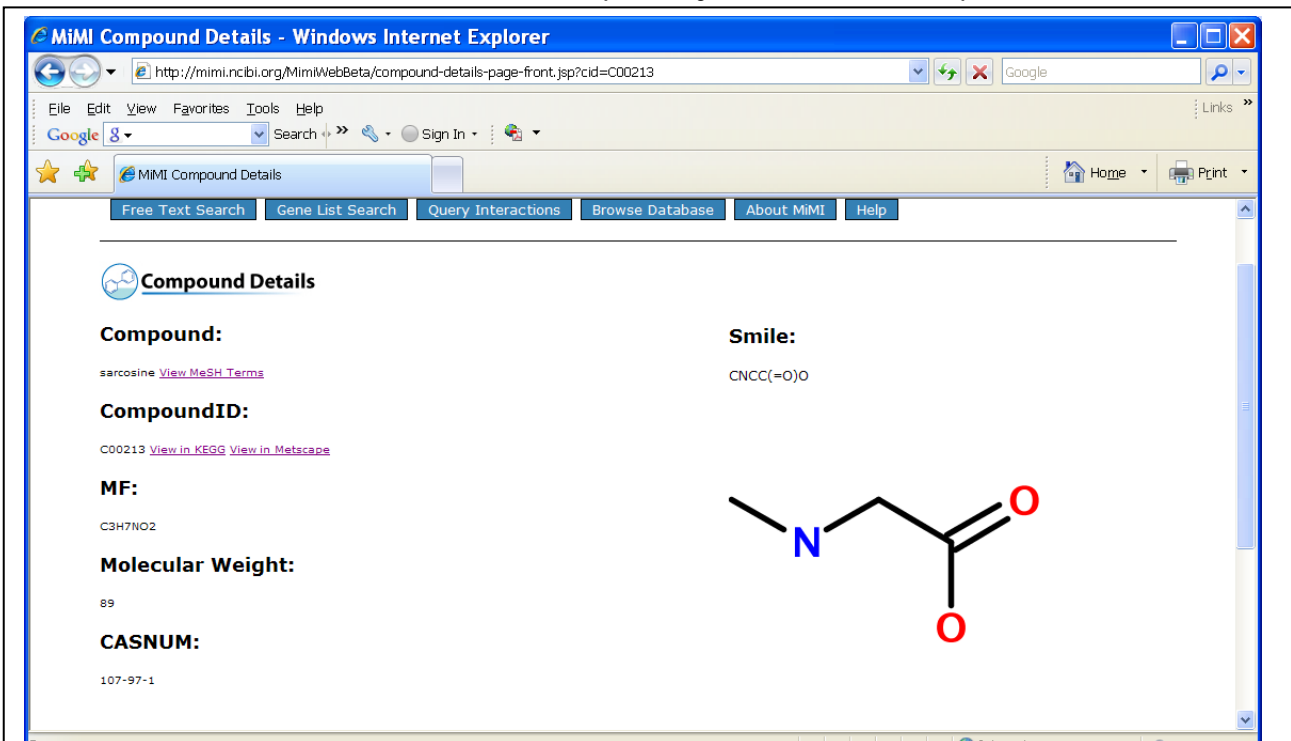
36 compounds found, displaying page 1 of 2.
[\[First/Prev\]](#) [1](#), [2](#) [\[Next/Last\]](#)

Id	Name	MF	Mol. Weight	CASNUM	Smile
C00025	glutamic acid	C5H9NO4	147	56-86-0	<chem>C(CC(=O)O)C(C(=O)O)N</chem>
C00037	glycine	C2H5NO2	75	56-40-6	<chem>C(C(=O)O)N</chem>
C00051	glutathione	C10H17N3O6S	307	70-18-8	<chem>C(CC(=O)NC(CS)C(=O)NCC(=O)O)C(C(=O)O)N</chem>
C00093	glycerol-3-phosphate	C3H9O6P	172	57-03-4	<chem>C(C(COP(=O)(O)O)O)O</chem>
C00099	beta-alanine	C3H7NO2	89	107-95-9	<chem>C(CN)C(=O)O</chem>
C00106	uracil	C4H4N2O2	112	66-22-8	<chem>C1=CN(C(=O)NC1=O</chem>
C00123	leucine	C6H13NO2	131	61-90-5	<chem>CC(C)CC(C(=O)O)N</chem>
C00137	myo-inositol	C6H12O6	180	87-89-8	<chem>C1(C(C(C(C(C1O)O)O)O)O)O</chem>
C00140	n-acetyl-d-glucosamine	C8H15NO6	221	7512-17-6	<chem>CC(=O)NC1C(C(C(OC1O)CO)O)O</chem>
C00148	proline	C5H9NO2	115	147-85-3	<chem>C1CC(NC1)C(=O)O</chem>
C00152	asparagine	C4H8N2O3	132	70-47-3	<chem>C(C(C(=O)O)N)C(=O)N</chem>
C00153	nicotinamide	C6H6N2O	122	98-92-0	<chem>C1=CC(=CN=C1)C(=O)N</chem>

Figure 3. Compound search results

Drill Down to Details on Compounds: Reactions, Enzymes and Genes

1. Find and click on C00213 - sarcosine. The Compound Details page for sarcosine appears. Since you are going to be exploring sarcosine networks in Part 2 of this tutorial, you want to get more background information about it here.
2. Scroll to the bottom of the page to see the molecular formula and the structure of sarcosine and the list of reactions and pathways in which this compound is involved.



The screenshot shows a web browser window titled "MiMI Compound Details - Windows Internet Explorer". The address bar shows the URL: <http://mimi.ncibi.org/MimiWebBeta/compound-details-page-front.jsp?cid=C00213>. The page content includes:

- Compound:** sarcosine [View MeSH Terms](#)
- CompoundID:** C00213 [View in KEGG](#) [View in Metscape](#)
- MF:** C3H7NO2
- Molecular Weight:** 89
- CASNUM:** 107-97-1
- Smile:** CNCC(=O)O

A chemical structure of sarcosine is displayed on the right side of the page. The structure shows a methyl group (CH3) attached to a nitrogen atom (N), which is further attached to a methylene group (CH2), which is attached to a carboxyl group (COOH). The oxygen atoms in the carboxyl group are highlighted in red.

Figure 4. Compound Details page

3. Click on Show/hide to display the list of reactions. You will see the [list of all reactions that involve sarcosine](#).
4. Click on Reaction ID R00367. This will bring up the Reaction Details page. In this reaction sarcosine is produced from glycine. Note the EC number of the enzyme that catalyzed this reaction - 2.1.1.20.
5. Click on a gene symbol below the EC number of the enzyme. The name of the enzyme that generates sarcosine from glycine is glycine-N-methyl transferase (GNMT). Clicking on the symbol brings up the Gene Details page for GNMT. This is a standard MiMI gene page where you can find out more about protein interactions and literature citations for GNMT.
6. Use your browser "Back" button to return to the Reaction Details page. Note the subcellular location graph on the top half of the page. In this case it shows that the enzyme glycine-N-methyl transferase is found in the cytoplasm.

7. Use your browser "Back" button to return to the Compound Details page. You can look at other reactions involving sarcosine on your own. When you do, you will see that there are two reactions in which glycine is produced from sarcosine and one reaction where glycine is converted into sarcosine. From the Reaction Details pages for R00610 and R00611, you find that the level of sarcosine is regulated by three enzymes: glycine-N-methyl transferase, sarcosine dehydrogenase and dimethylglycine dehydrogenase.
8. Still on the Compound Details page, click Show/Hide to display the list of pathways in which the compound is involved. Note that sarcosine is involved in the glycine, serine, alanine and threonine metabolism pathway. You may follow the link to the KEGG pathway on the right side of the table to see the pathway chart in KEGG.

You now have gained insights about reactions and the pathway that involve sarcosine and identified the enzymes that regulate the level of sarcosine.

If you explore the rest of the compounds on the list you will find that most of them are involved in amino acid metabolism and nitrogen breakdown pathways.

Move from MiMI Web to Metscape Networks

1. Still in MIMI on the sarcosine Compound Details page, click on the "View in MetScape" button under Compound ID heading. This operation prompts Cytoscape with the Metscape plugin to load via Java Web Start. The first time you open Cytoscape, the Java Web Start software caches (stores) the entire application locally on your computer. From then on, every time you start the application, the Java WebStart software component checks the application's website to see if a new version is available, and if so, automatically downloads and launches it.
2. Click "Run" in the Java web start dialogue box when prompted. Cytoscape opens and displays the Metscape network associated with sarcosine. (Fig. 5). Note the legend in the right side of the window.

NOTE: Further in this tutorial we will see a different network view, where the hexagonal nodes represent compounds and the edges that connect them represent the reactions.

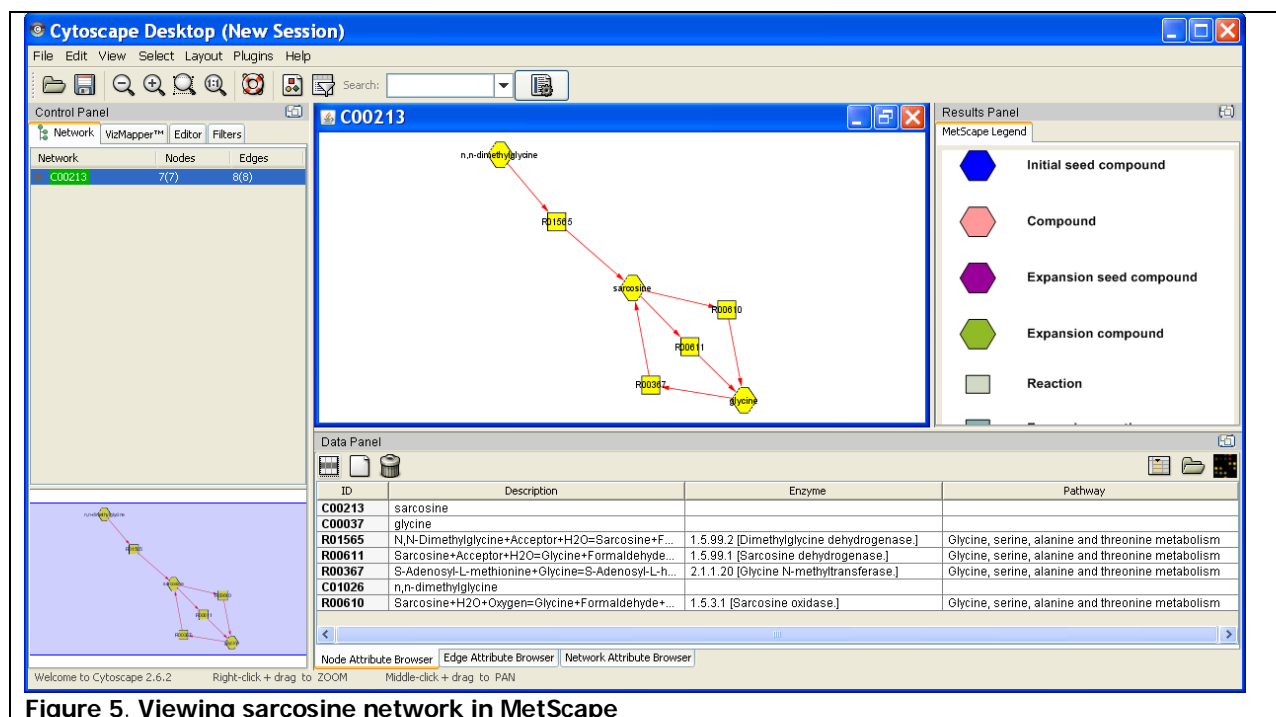


Figure 5. Viewing sarcosine network in MetScape

3. Select a group of nodes by making a box around them with your mouse. Alternatively you may use Control+Alt+A. The nodes will become highlighted (yellow) to indicate the selection.
4. Click on the Table icon in the Data Panel below to select data fields to display.
5. Choose these fields: Description, Enzyme, Pathway. Click anywhere outside the list to close it. Information now fills the data panel about the enzymes that regulate the sarcosine level.

Expanding the Network and Building Subnetworks

Try on your own

1. Identify the node that represents glycine and do the following steps: **Right click on the node → pop-up menu → MetScape → expand in a new window**. A new window with the glycine network will appear. Note that glycine is in the center of the network. Now you can identify the reactions in which glycine is produced and converted into other compounds.
2. If you are interested in a particular subset of compounds and reactions you can create a subnetwork. To do this select a compound (e.g. serine) and adjacent reaction node (R00945), then right click on selected nodes. Select **MetScape → create subnetwork** from a drop down menu. A new subnetwork window will appear. You can copy and paste the subnetwork that you just created into a presentation or Word document and continue to explore it further. Subnetwork analysis is particularly useful in for large complex networks.

After you have familiarized yourself with MetScape features and gained the information about sarcosine and glycine networks, we are ready to visualize the experimental data for the 36 compounds that we looked at in MiMI.

Load and View the Experimental Data

1. On the top menu, go to **Plugins** → **Metscape plugin** → **Build Compound Reaction Network**. A query dialogue box appears with three tabs on top. Each tab lets you conduct a different type of query.
2. Click the second tab - "Load Experiment File."

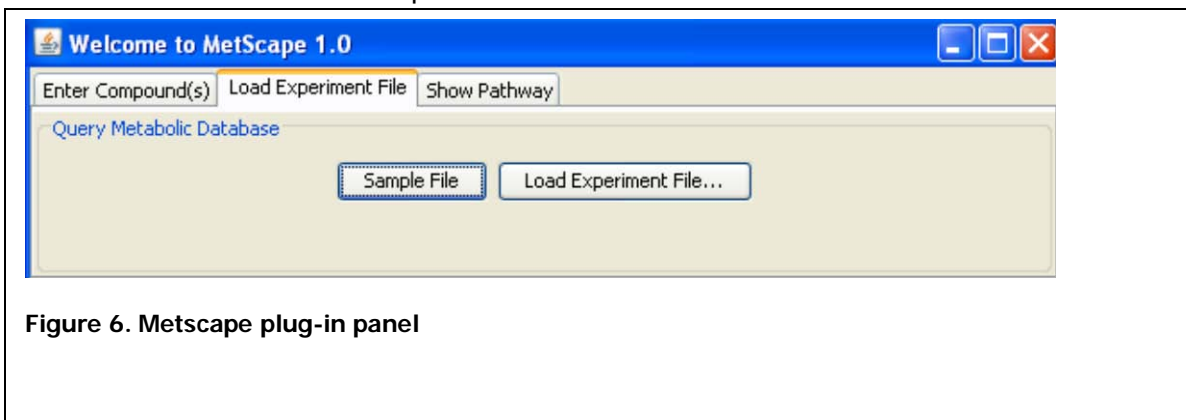


Figure 6. Metscape plug-in panel

When you load experimental data, Metscape automatically color codes and sizes the nodes to represent experimental values. The experimental file can have up to two types of values (e.g. concentration and fold change). The color intensity of the node represents the first value type. If you have a second value type, it is represented by node size. The digit in the left-most column indicates to MetScape whether your experimental data have one or two types of values.

One parameter file												
1	t1	t2	t3	t4	t5	t6						
Two parameters file												
2	t1	t5	t15	t20	t45	t0	t1	t5	t15	t20	t45	

The data file that we will use in this tutorial has two types of values. We will use the node color to represent fold change and the node size to represent normalized experimental measurements.

3. Click on the Load Experiment File button. A dialog box will appear. Select the pre-formatted data file that you downloaded at the beginning of this tutorial. Click open.
4. You will see the network that consists of the experimental compounds. Examine the network. You will notice that the compounds form several separate subnetworks. You may explore these later on your own, using techniques covered in the tutorial.

5. At this point you are also able to view experimentally measured compound in the context of known biochemical reactions and pathways.
6. From the top menu select **Plugins → MetScape → Build Compound Reactions Network**. When the dialog box appears select the Show Pathway tab. This gives you a different perspective from the earlier network you analyzed. Now the compounds are shown by nodes, and reactions (and enzymes are represented by edges (links between nodes)
7. Go to the top of the drop-down list of pathways and select All. Click the Draw Pathway button. You will see a large network of all metabolic pathways. Zoom in on any part of the network. Remember that the edges represent reactions.
8. Select any compound node and right click on it. Select **MetScape → Dynamic Expression**.
Now only the compounds that have experimental values are colored. The Dynamic expression window has three points that correspond to three samples – benign, localized and metastatic. You can navigate between these and change the node color and size.
9. Use search box to find sarcosine. Identify glycine node nearby. Note the difference in size of the nodes. Use the dynamic expression dialog box to see how the levels of these two compounds change between different samples. Note how sarcosine level increases from benign to localized and from localized to metastatic tumor samples. Also note the difference in node sizes, which indicates the compound concentration.

Explore Individual Pathways

1. You already know that sarcosine is involved in the glycine, serine, alanine and threonine metabolism pathway. You may want to look at the experimental data in the context of this pathway.
To do this, go through the following steps:
2. From the top menu select **Plugins → MetScape → Build Compound Reactions Network**.
From MetScape dialog box select Show Pathway tab. From the dropdown menu select Glycine, serine, alanine and threonine metabolism pathway. Click Draw Pathway button. You will now see the reaction network for this specific pathway. Note that two of the experimentally measured differential compounds are found in this pathway in addition to sarcosine and glycine. These are threonine and homocysteine. Use the Dynamic Expression dialog to view their experimental values. Note that sarcosine is the one with the most robust increase.

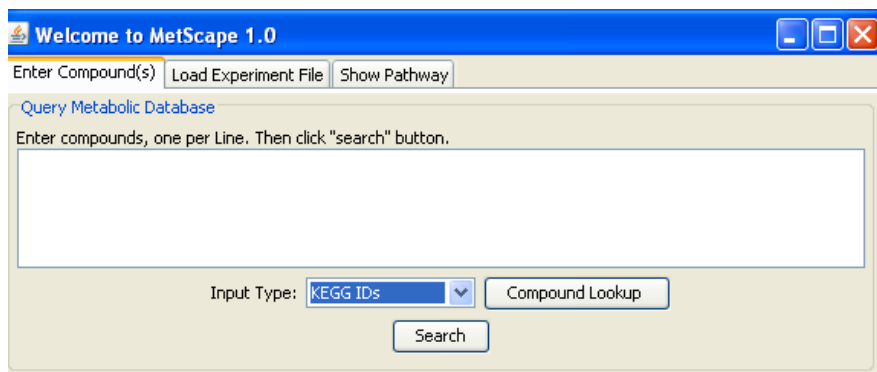
Sreekumar and coauthors experimentally confirmed the relevance of sarcosine level for tumor progression. They also performed a number of RNAi experiments to investigate the roles of the enzymes that regulate sarcosine level in tumor progression and demonstrated that attenuation of GNMT and GMGDH (enzymes that produce sarcosine) reduced the level of sarcosine and resulted in decreased cell invasiveness. In contrast, knock-down of SARDH resulted in increased sarcosine level and increased cell invasiveness.

As mentioned in the introductory part of this tutorial, six metabolites were consistently increased on disease progression from benign to localized to metastatic prostate cancer.

We explored the experimental data in the context of reaction and pathway information for one of them. You may want to do the same for the remaining 5 compounds (kynurenine, glycerol-3-phosphate, leucine, proline and uracil)

Load Compounds Using the Enter Compound(s) Tab

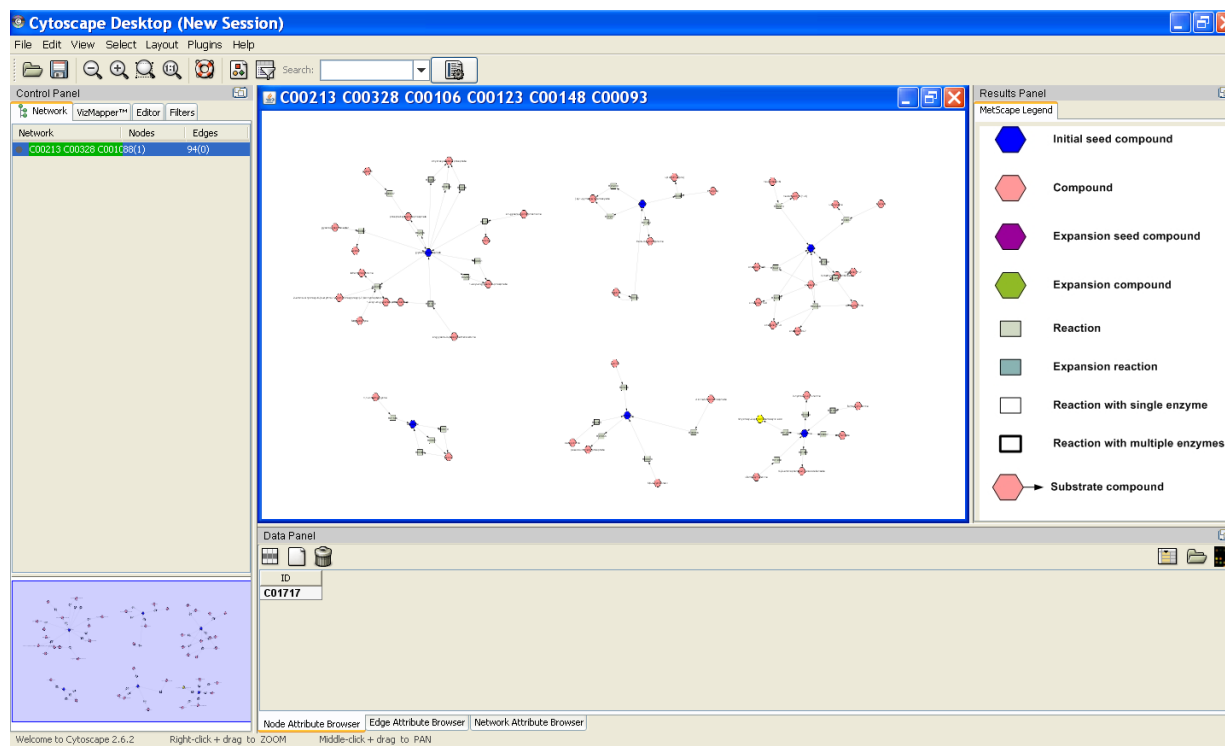
1. On the top menu, go to Plugins and in the drop down menu click **Metscape plugin** → **Build Compound Reaction Network**. A query dialogue box appears.



2. Choose input type: KEGG IDs
3. Select the first tab. Enter the following 5 Compound IDs as a string. You do not need to put in commas.

C00328 C00106 C00123 C00148 C00093

4. Click Search. Compound-reaction networks for each compound are displayed. Each compound has its own sub-network



Use dynamic expression dialog box to display experimental data as described above. Determine the enzymes and pathway for each of the 5 compounds.

References

- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008 Jan;36(Database issue):D480-4.
- Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol.* 2007;3:135.
- Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature.* 2009 Feb 12;457(7231):910-4.
- Tarcea VG, Weymouth T, Ade A, Bookvich A, Gao J, Mahavisno V, Wright Z, Chapman A, Jayapandian M, Ozgür A, Tian Y, Cavalcoli J, Mirel B, Patel J, Radev D, Athey B, States D, Jagadish HV. Michigan molecular interactions r2: from interacting proteins to pathways. *Nucleic Acids Res.* 2009 Jan;37 (Database issue):D642-6.