

MetScape 3.1 User Manual

An App for Cytoscape

National Center for Integrative Biomedical Informatics

March 2015

2014 University of Michigan This work is supported by the National Center for Integrative Biomedical Informatics through NIH Grant# 1U54DA021519-01A1 and the Michigan Regional Comprehensive Metabolomics Research Core through NIH Grant# U24DK097153.

Contents

| OVERVIEW | 4 |
|---|-----------|
| About data sources | 4 |
| Workflow overview | 4 |
| Installing Cytoscape and the MetScape app on your local computer | 5 |
| PART 1: PATHWAY-BASED NETWORK | 6 |
| Entering Data | 6 |
| Option 1: Enter a list of compounds | 6 |
| Option 2: Load an experiment file | 9 |
| Pathway-Based Visualization | .3 |
| MetScape Tab1 | .4 |
| Rules Used to Build Different Network Types1 | .6 |
| Table Panel Display | .6 |
| Choosing attributes provided through MetScape1 | .6 |
| Rearranging attributes in the table panel1 | .9 |
| Pathway filter in the table panel2 | 20 |
| Concept filter in the table panel2 | 21 |
| MetScape Results Panel (Additional Node Information)2 | 24 |
| Animation of Data2 | 24 |
| Building an Animation2 | 24 |
| Manipulating the animation: zooming2 | 27 |
| Manipulating the animation: bar chart2 | 28 |
| Playing the animation2 | 28 |
| Manipulating Network2 | 29 |
| Expanding a network2 | 29 |
| Option 1: Expanding a network in the current network window (Example using Compound- Reaction Network) | <u>29</u> |
| Option 2: Expanding a network in a new window (Example using Compound Network from Selected Genes) | 30 |
| Collapsing a network | 31 |
| Creating a subnetwork | 31 |
| Destroying a network3 | 32 |
| PART 2: CORRELATION-BASED NETWORK | 3 |

| Entering Data | 33 |
|-----------------------------------|----|
| File Formats | 33 |
| Building a Network | 34 |
| Group Definition | 36 |
| Correlation Visualization | 36 |
| Additional Information | 37 |
| Correlation Network Results Panel | 37 |
| SAVING AND REOPENING A SESSION | 38 |
| Saving a session | 38 |
| Reopening a session | 38 |
| SUPPLEMENT: METDISEASE APP | 38 |
| How to use MetDisease | 38 |

Please note that due to continuous software upgrades, the images in this handout may not exactly mimic what you see on the screen.

OVERVIEW

About data sources

MetScape is an app for <u>Cytoscape</u>, the bioinformatics network visualization tool. The app can be used to visualize and interpret metabolomics and gene expression data in the context of human metabolic networks.

MetScape uses a metabolite database developed by extracting and integrating information from the following sources:

1. Edinburgh Human Metabolic Network (EHMN) -

http://www.ehmn.bioinformatics.ed.ac.uk/

2. KEGG COMPOUND Database — <u>http://www.genome.jp/kegg/compound/</u>

MetScape allows users to load a list of metabolites with experimentally determined concentrations, a list of genes with experimentally determined expression values, and a list of concepts or pathways, and display them in the context of relevant metabolic networks.

Workflow overview

With MetScape, you can:

- Trace the connections between metabolites and genes.
- Integrate multidimensional data
- Visualize compound, reaction, enzyme and gene networks and display compound structures, as well as information for reactions, enzymes, genes, and pathways.
- Visually animate changes in compound concentrations over time and across experimental conditions.

Pathway-Based Workflow

The basic steps in the **pathway** workflow include:

- 1. <u>Enter data</u>. You can type or paste a list of compounds and/or genes, load a file containing experimental data, or start from a biological pathway.
- 2. <u>Select compound and reaction attributes</u>. Choose which attributes to display in a table as you work with your visual network graph.
- 3. Explore the visual network and table of attributes:
 - <u>Expand</u> and <u>collapse</u> a network.
 - <u>Create a subnetwork</u>
 - Visualize your data in a wide variety of network layouts provided by Cytoscape.
 - Use color, size, and other effects to visually reflect a set of attribute values.
- 4. Save your session and reopen it later.

Correlation-Based Workflow

The basic steps in the **correlation** workflow include:

- 1. <u>Enter data</u>. You can load a file containing experimental data.
- 2. <u>Select the desired Edge Mapping and Significance Range</u>. Use the slider to designate the desired significance range.
- 3. Explore the visual network and table attributes.
- 4. Save your session and reopen it later.

Installing Cytoscape and the MetScape app on your local computer

- 1. Install Cytoscape on your computer. For more information, go to the Cytoscape website at http://cytoscape.org/.
- 2. After Cytoscape is installed, start the application.
- 3. To install the MetScape app, select **Apps -> App Manager** from the Cytoscape menu.
- 4. Under the **Install Apps** tab, scroll down until you find MetScape in the second column.

| 🎯 App Manager | | × |
|-------------------|--|---|
| | GASOLINE GeneMANIA GenomeSpace GFD-Net GIANT HyperModules jActiveModules JEPETTO KEGGscape KeyPathwayMiner | Manage Sites PletScape 3.0.2 (installed: 3.0.1) Maps human, mouse and rat metabolomics and gene expression data to human metabolic networks and |
| Install from File | MCODE (Installed) MetDisease MetScape, (Installed) PathExplore PathExplore Pathexplore PEPPER PINA4WS ReactomeFIPlugIn SemScape setSApp Venn and Euler Diagrams WikiPathways | Find MetScape app in second column |

- 5. Click on the **MetScape** app.
- 6. Click Install.
- 7. When installation is complete, click **Close.**

A MetScape option is added to the Cytoscape **Apps** menu.

Registration: MetScape is a free program. We ask you to register because it helps us to keep track of the number of downloads. Your information will be stored in a secure database and we will not share it with anyone. We may send you infrequent e-mails about future MetScape releases.

Note: This app requires Cytoscape 3.0+ to run correctly.

PART 1: PATHWAY-BASED NETWORK

Entering Data

- 1. To begin a Cytoscape session with the MetScape app, first start Cytoscape.
- 2. Choose one of the following methods to get started.

Option 1: Enter a list of compounds

- Select Apps -> MetScape-> Build Network -> Pathway-based from the Cytoscape menu.
- 2. A MetScape tab now appears on the left of the Cytoscape screen.

| Control Panel | A Network |
|--|--|
| Retwork Style Select MetScape | |
| Build Pathway-based Network | |
| Input Choose species | |
| Organism Human Choose species | Add Companyeds |
| Compounds: (none) | K Add Compounds |
| Genes: (none) Select | Enter a list of compound names or KEGG IDs. KEGG IDs may be entered all on one line (separated by commas or |
| Compounds | spaces), or one per line. Compound names should be |
| Input ID Input Name | entered one per line. |
| 7 | DL-Lipoyl-L-lysine trimethylalamine |
| Add Remove Clear Reset | glucose |
| Genes | N |
| Input ID Input Symbol | |
| a por symbol | |
| | - |
| Enter compounds Add Remove Clear Reset | OK Canci |
| and/or genes | |
| Network Type | Table Panel |
| Compound-Reaction-Enzyme-Gene | |
| Query Use compounds/genes | |
| Use selected pathway | shared name Enter compounds into |
| TCA cyde "ADD" buttons | "Add Compounds" |
| Abb battons | popup window |
| Build Network Output as File Close | Node Table Edge Table Network Table |
| | Node Table Luge Table Metholic Table |

- 3. Select a species. Choices are: human, rat, mouse
- Manually enter (or copy and paste) compound ID(s) or name(s) and/or Entrez gene ID(s) or symbol(s). Click Add and enter the appropriate ID type in the popup box. Lists of IDs should be separated by spaces; lists of compound names should be entered one per line.
- 5. Click OK
- 6. If you enter a compound name, it will map to its KEGG ID a popup window will appear. If there is more than one potential match, use the dropdown arrow to choose the best match. If the compound is not found in the system, it will say "Not Found." The mapping selection will be saved so that your selection will appear as the default option in the future.

| Select Compound Mappings | | | | | |
|--------------------------|---|---------------|--|--|--|
| | Select names for compounds | | | | |
| Please select the match | One or more of the compounds in your query had multiple matches in the database. Please select the match that is best - or '(none)' if you do not wish the compound to appear in the results. | | | | |
| Input Name | Potential Matches | KEGG ID | | | |
| DL-Lipoyl-L-lysine | DL-Lipoyl-L-lysine | CE2102 | | | |
| glucose | Glucose 🗸 🗸 | <u>C00293</u> | | | |
| trimethylamine | Not Found | | | | |
| | | | | | |
| | OK Cancel Save | | | | |

- 7. Click OK
- 8. After the experimental data has been loaded, if any of the genes/compounds that you submitted were not mapped to the database objects, a **Missing Data** window will appear. After viewing the **Missing Data** information, click **SAVE** to save the data for later viewing or click **OK** to close the window without making it available for future viewing. (See the <u>Missing Data Window</u> section below, under Option 2: Load an experiment file, for more information about missing data).

| Missing Data |
|--|
| Why are these elements missing 🕜 |
| The following input data could not be found in the MetScape database: |
| Compounds: Trimethylamine |
| |
| |
| |
| OK Save |

- 9. Select a **Network Type** by selecting one of the following from the dropdown menu:
 - Compound-Reaction-Enzyme-Gene (1)
 - Compound-Reaction (2)
 - Compound-Gene (3)
 - Compound (4)



Note: When selecting Compound as the Network Type, a dropdown menu appears under Query, providing the option of choosing between compounds or genes. When any other Network Type is selected, the only option is to use compounds/genes unless using a selected pathway.

10. Click **Build Network** to query the database and create the network.

Option 2: Load an experiment file

Use to load experimental data to visualize and explore compound networks over time or in varying experimental conditions. The input can be an Excel, comma, or tab delimited file. MetScape allows users to load three types of files – compound file, gene file, and concept file. Each type is optional, e.g. you can load only compounds, only genes, only concepts, or any combination of the above.

- Select Apps -> MetScape-> Build Network -> Pathway-based from the Cytoscape menu.
- 2. A MetScape tab now appears on the left of the Cytoscape screen.
- 3. Select a species. Choices are: human, rat, mouse
- 4. Load experimental data by clicking the **Select...** button on the MetScape tab.

Compound File

The compound data file must meet the following requirements:

- The first row must be a heading row, in which:
 - The columns in the first row are column headings to label the data.
- All other rows contain experimental data, in which:
 - The first column contains KEGG Compound IDs or names.
 - The remaining columns contain experimental data. Multiple experimental values are permissible in the same spreadsheet.

Below is a portion of an example compound file with Compound IDs, significance values, and fold change values:

| | А | В | С |
|----|--------|-------------|-------------|
| 1 | | Fold Change | P-value |
| 2 | C00002 | 0.51 | 0.006174265 |
| 3 | C00004 | 1.14 | 0.646801757 |
| 4 | C00008 | 0.99 | 0.970303452 |
| 5 | C00020 | 2.12 | 0.068052875 |
| 6 | C00025 | 0.91 | 0.115376267 |
| 7 | C00037 | 0.99 | 0.904658012 |
| 8 | C00041 | 0.95 | 0.574835153 |
| 9 | C00042 | 1.09 | 0.809172269 |
| 10 | C00044 | 0.56 | 0.100724407 |
| 11 | C00047 | 1.36 | 0.189989302 |
| 12 | C00049 | 0.92 | 0.755789133 |
| 13 | C00064 | 0.48 | 0.186707182 |
| 14 | C00065 | 1.22 | 0.805151078 |
| 15 | C00073 | 0.80 | 0.356405722 |

Gene File

The gene file must meet the following requirements:

- The first row must be a heading row that includes column headings to label the data.
- The first column contains Entrez Gene IDs or Official Gene Symbols.
- The remaining columns contain experimental data. Multiple experimental values are permissible in the same spreadsheet.

Below is a portion of an example gene file with Gene IDs, significance values, and fold change values:

| | А | В | С | |
|----|----------------|-------------|-----------------|--|
| 1 | ENTREZ_GENE_ID | Pvalue | Log_fold_change | |
| 2 | 1 | 0.002335904 | -0.02 | |
| 3 | 2 | 2.07E-08 | -0.06 | |
| 4 | 2 | 2.03E-08 | 0.04 | |
| 5 | 9 | 0.003919472 | 0.04 | |
| 6 | 10 | 0.00026716 | -0.08 | |
| 7 | 12 | 0.000237925 | 0.05 | |
| 8 | 13 | 0.002489364 | -0.05 | |
| 9 | 14 | 0.506396205 | 0.00 | |
| 10 | 15 | 8.25E-06 | -0.04 | |
| 11 | 16 | 0.022222108 | -0.01 | |
| 12 | 18 | 1.47E-08 | -0.07 | |
| 13 | 18 | 5.24E-06 | -0.09 | |
| 14 | 18 | 2.12E-05 | -0.07 | |
| 15 | 19 | 0.820594937 | 0.00 | |
| 16 | 19 | 1.76E-08 | 0.06 | |
| 17 | 19 | 9.46E-09 | 0.04 | |
| 18 | 19 | 0.660631201 | 0.00 | |
| 19 | 20 | 1.16E-07 | -0.05 | |

Concept File

The concept file can be generated by a gene set enrichment analysis tool such as LRpath or GSEA from gene expression data.

Note: Gene set enrichment testing is an approach used to test for predefined biologicallyrelevant gene sets that contain more significant genes from an experimental dataset than expected by chance.

- GSEA (Subramanian at al., Proc. Natl. Acad. Sci. USA, 2005, 102:15545-15550).
- LRpath (Sartor et al., Bioinformatics, 2009, 25(2):211-7).

Below is a portion of an example concept file:

| | А | В | С | D | E | F | G | Н | I |
|----|--------------------|--------------|---------|------------|------------|------------|------------|-----------|--------------------------------------|
| 1 | Concept.name | ConceptType | n.genes | coeff | odds.ratio | p.value | FDR | Direction | sig.genes |
| 2 | Citrate cycle (TC/ | KEGG Pathway | 32 | 0.44858373 | 16.2447876 | 1.23E-08 | 1.67E-06 | up | 24368, 24399, 24401, 25179, 25721, |
| 3 | Fatty acid metab | KEGG Pathway | 28 | 0.40269739 | 12.2142885 | 1.10E-06 | 7.48E-05 | up | 24158, 25363, 25618, 25757, 140547, |
| 4 | Alanine and aspa | KEGG Pathway | 18 | 0.47651019 | 19.323541 | 1.93E-06 | 8.75E-05 | up | 24379, 24401, 25721, 81670, 81829, |
| 5 | Reductive carbox | KEGG Pathway | 11 | 0.55839769 | 32.1439027 | 6.68E-06 | 2.27E-04 | up | 24368, 24399, 24401, 25721, 79250, |
| 6 | Oxidative phospl | KEGG Pathway | 44 | 0.27758123 | 5.61284966 | 3.64E-05 | 9.89E-04 | up | 116550, 291103, 295923, 301011, 31 |
| 7 | Urea cycle and m | KEGG Pathway | 25 | 0.32301096 | 7.44383786 | 2.00E-04 | 0.00441518 | up | 24368, 24379, 24399, 24401, 24600, 3 |
| 8 | PPAR signaling p | KEGG Pathway | 32 | 0.28585824 | 5.90912002 | 2.33E-04 | 0.00441518 | up | 24158, 24450, 25045, 25757, 29171, 3 |
| 9 | Carbon fixation | KEGG Pathway | 11 | 0.44333373 | 15.7233268 | 3.59E-04 | 0.00542234 | up | 24401, 25721, 81670, 81829, 114508, |
| 10 | Arginine and pro | KEGG Pathway | 20 | 0.33385131 | 7.96259603 | 5.14E-04 | 0.00699086 | up | 24368, 24379, 24399, 24401, 24600, 3 |
| 11 | Butanoate metak | KEGG Pathway | 30 | 0.26301141 | 5.12695967 | 0.00101503 | 0.01254945 | up | 24379, 24399, 24401, 24450, 25721, |
| 12 | Valine, leucine a | KEGG Pathway | 24 | 0.27898384 | 5.66198894 | 0.00167291 | 0.01625109 | up | 24158, 24450, 29711, 140547, 17046 |
| 13 | Glutathione met | KEGG Pathway | 24 | 0.27372392 | 5.4799003 | 0.00205204 | 0.01860519 | up | 24379, 24399, 24401, 24422, 25721, |
| 14 | Propanoate meta | KEGG Pathway | 16 | 0.32105291 | 7.35380592 | 0.00265316 | 0.02122524 | up | 24158, 140547, 170465, 171155, 298 |
| 15 | Nitrogen metabo | KEGG Pathway | 15 | 0.30373897 | 6.60362465 | 0.00596513 | 0.04056287 | up | 24379, 24399, 24401, 25721, 29242, 3 |
| 16 | C5-Branched dib | KEGG Pathway | 19 | 0.26819205 | 5.29471153 | 0.00686636 | 0.04226683 | up | 24158, 24379, 24399, 24401, 25721, |
| 17 | Pyruvate metabo | KEGG Pathway | 19 | 0.26689922 | 5.25234187 | 0.00714807 | 0.04226683 | up | 81829, 298942, 306198, 307858, 361 |
| 18 | Phenylalanine m | KEGG Pathway | 21 | 0.25034932 | 4.73898475 | 0.00822755 | 0.0447139 | up | 24368, 24401, 25721, 81683 |
| 19 | Antigen processi | KEGG Pathway | 21 | 0.2496329 | 4.71793226 | 0.00841626 | 0.0447139 | up | 24223, 24812, 25217, 25599 |

To generate a concept file, you can click the LRpath button from MetScape that will take you to the LRpath website.

Note: LRpath performs gene set enrichment testing, an approach used to test for predefined biologically-relevant gene sets that contain more significant genes from an experimental dataset than expected by chance (<u>Sartor et al., 2009</u>). To run LRpath, you need a Gene Expression file with fold change (or log fold change) values and p-values. The Gene Expression file needs to contain all gene records, not just those for significant genes; LRpath will determine the significant genes from the input.

Example of selecting experimental data, including importing files and designating fold change, p-values, and thresholds:

| Select Experimental Data | pose species |
|---|---|
| Organism Human | P-value, Fold Change, and Threshold |
| Compounds | and Threshold |
| Select experimental data (none) | Import File |
| P-value (none) | Threshold |
| Fold Change (none) | Threshold |
| Genes Select experimental data | Import File button |
| (none) | Import File |
| P-value (none) | ✓ Threshold |
| Fold Change (none) | ✓ Threshold |
| Concepts | |
| Select experimental data | |
| (none) | ▼ Import File |
| Generate a concept file from genes using LA | RPath: LR Path ? |
| ОК | Cancel Button for accessing LRpath website. |

Page **11** of **39**

Below is an example of the ID mapping window from a loaded compound experiment file. The mapping selection will be saved so that your selection will appear as the default option in the future.

| | Select names for comp | oounds | |
|-------------------|---|----------------------------|---|
| Please se | ore of the compounds in your query had n lect the match that is best - or '(none)' if y I to appear in the results. | | |
| Input Name | Potential Matches | KEGG ID | |
| Glutamine | D-Glutamine | - C00819 | |
| GSH+Lysine | D-Glutamine | | |
| Ile | L-Glutamine Protein glutamine | CE2890 | |
| Inositol | Protein L-glutamine | C00137 | |
| Lactate | Protein N5-alkylglutamine | C05823 | |
| Leu | (none) (zs)-aipna-ceucine | C00123 | |
| Lysine+Arginine | Not Found | | |
| OH-Butyrate | Not Found | | |
| PtdCholine | Not Found | | |
| PtdEthanolamine | Not Found | | |
| PtdSerine | Not Found | | |
| Sphingomyelin | Sphingomyelin | C00550 | |
| Succinate | Succinate | ✓ C00042 | = |
| TAG | (+)-Prostaglandin F2a | · C00639 | |
| TMAO+Betaine | Not Found | | |
| Total cholines | Not Found | | |
| Total Glucose | Not Found | | |
| Total Glutathione | Not Found | | |
| Tyrosine | Tyrosine | · <u>C00082</u> | |
| Val | (-)-5-(3'.4'.5'-trihydroxyphen | CE2124 | * |

Note: If you select **Save** on the **Select Compound Mappings** window, you will get a .csv file containing mapping information for all compounds that MetScape successfully mapped.

Note: Multiple networks can be built within the same user session. After a network is built, pull up the **Select Experimental Data** window by clicking on the **Select...** button and change the data files. Then build the new network. Both networks will remain accessible. If you change species, however, all existing data will be lost. The below warning window will pop up in this situation.

| Warning | x |
|---------|---|
| ? | All existing data will be lost. Continue? |
| | OK Cancel |

Missing Data Window

After the experimental data has been loaded, if any of the genes/compounds that you submitted were not mapped to the database objects, a **Missing Data** window will appear.

Genes, compounds, and concepts may appear on this missing elements list because:

- Genes and compounds that you supply may not be in the database. If they are not found in the database then they are reported as missing.
- If your input genes are not human (Rat, for example), then they are mapped to human genes using homologs from NCBI's HomoloGene. If this mapping fails, then those genes are reported as missing.
- MetScape will display only the genes that encode metabolic enzymes. If an input gene does not encode metabolic enzymes, it will appear on the missing elements list.
- A concept (pathway) will appear on the missing list if all of its significant genes are missing. The list of significant genes for a concept comes from the input file or from LRpath.

After viewing the **Missing Data** information, click **SAVE** to save the data for later viewing or click **OK** to close the window without making it available for future viewing.

| ſ | Missing Data | |
|---|---|---|
| | Why are these elements missing 🛛 🕜 | |
| | The following input data could not be found in the MetScape database: | 1 |
| | Compounds: Creatine +Crn GSH +Lysine Lysine +Arginine OH-Butyrate PtdCholine PtdCholine PtdEthanolamine PtdSerine | Ш |
| | TMAO +Betaine Total Glucose Total Glutathione | - |
| | OK Save | |

Pathway-Based Visualization

MetScape includes a legend explaining its various shapes and colors. The legend will be specific to the current network type (pathway or correlation).

Access the MetScape Legend from Apps -> MetScape -> Show Legend.



Additional information about the network is expressed through visualization:

- A compound is red in the original network and subsequent subnetworks if it was in the original data loaded into MetScape.
- A green border surrounding a node represents a significant gene/compound.
- Node size represents the direction of the change. Larger nodes represent an increase and smaller nodes represent a decrease in gene/metabolite. The actual amount of the change is not represented visually.
- When a node is expanded, the edges between the original node and the expanded nodes become blue.

Note: In a pathway-based network, selected edges are highlighted yellow instead of red.

MetScape Tab

This tab has options for choosing Network Types, the data you entered, and more.

From this tab you can:

- Select between building a pathway-based network or a correlation-based network
- Select the type of network that you want to build

Select **Network Type** by selecting one of the following from the dropdown menu:

- o Compound-Reaction-Enzyme-Gene
- Compound-Reaction
- Compound-Gene
- Compound

Note: When selecting Compound as the Network Type, a dropdown menu appears under Query, providing the option of choosing between compounds or genes. When any other

Network Type is selected, the only option is to use compounds/genes unless using a selected pathway.

- Add or Remove data.
- Select a pathway
 - Use to view all the compounds and reactions associated with a metabolic pathway. Select a specific pathway from the drop-down list.



- Save the **Output as a File**.
- Build Network graph based on the data.

| | Control Panel | × | |
|------------------|---|-------------------------|--------------------------|
| | Select | MetScape | |
| | Build Pathway-based | ▼ Ne | twork |
| | Input | | |
| | Organism Human | All | |
| | Data Files Compounds: CompoundFile Genes: GeneFile Concepts: ConceptFile | Select pathway build | |
| | Compounds | | |
| | Input ID | Input Name | |
| | C00047 | L-Lysine | |
| | C00064 | L-Glutamine | - |
| | Add Remo | Clear Reset | |
| | Genes | | |
| | Input ID | Input Symbol | Add or remove |
| | 38 | ACAT1 | |
| | 47 | ACLY | - |
| | Add Remo | ve Clear Reset | |
| Select a pathway | Options Network Type | | Choose network type |
| patriway | Compound | | • |
| <u>۱</u> | Query | | |
| | Use compounds | | - |
| | Use selected pathway | | |
| | TCA cycle | | Save output as a file |
| | | asa me | |
| | Build Network | Output as File Close | |

Note: When the organism is human, the Genes section will show Input ID and Input Symbol. However, when the organism is not human (ex. rat), the Genes section will show Input ID, Input Symbol, and Human Symbol.

Rules Used to Build Different Network Types

Compound-Reaction-Enzyme-Gene (C-R-E-G), Compound-Reaction (C-R), and Compound–Gene (C-G) networks

The C-R-E-G, C-R, and C-G networks are all built from the same underlying data. That data is derived, in each case, by finding compounds that participate in reactions that are catalyzed by enzymes that are encoded by genes. If only genes are input, then all the enzymes, reactions, and compounds that match those genes are used. If only compounds are input, then all the reactions, enzymes, and genes that match those compounds are used. If both genes and compounds are input, then only those C-R-E-G couplings that match both a compound from the compound input and a gene from the gene input are used.

If a concept file is provided, genes from that file will be used as input instead of the genes from a provided gene file.

If only a gene file is provided, all genes from that file are used as input. In this case, we recommend that you load a smaller set of genes (e.g. most significant differentially expressed genes).

• Compound (C) networks built from compound input

If a list of compounds is provided, the resulting network will include the query compounds (shown in red) plus any compounds that participate in the same reactions as query compounds. The edges will be drawn between "seed" compounds and their neighboring compounds

• Compound (C) networks built from gene input

If a list of genes is provided, the resulting network will include the compounds that are related to the query genes via the reactions in which they participate, and the enzymes that catalyze these reactions. The edges will be drawn between all compounds.

• Pathway networks

Pathway-specific C-R-E-G, C-R, and C-G networks are built from a set of genes, enzymes, reactions and compounds defined in the EHMN database. Pathway-specific Compound networks are built similarly to the Compound networks built from a set of input genes.

Table Panel Display

Attributes of compounds, reactions, pathways, and concepts that you have selected in the network appear in the Cytoscape Table Panel.

Choosing attributes provided through MetScape

To choose which attributes to view:

- To choose attributes for compounds, click the Node Table tab in the Table Panel. You can choose to view node, edge, or network attributes by clicking on the respective tabs at the bottom of the Table Panel. Pathway data can be viewed under the Pathway Filter tab and concept data can be viewed under the Concept Filter tab in the Table Panel.
- 2. Click the **Hide All Columns** icon in the Table Panel toolbar. This clears the Table Panel of all columns.
- 3. Click the **Show Column** icon in the Table Panel toolbar. A list of attributes with check boxes will appear.

| Table Panel | × □ |
|---|------------|
| $\textcircled{\begin{tabular}{cccccccccccccccccccccccccccccccccccc$ | |
| Show Column icon Hide All Columns icon | |
| Node Table Edge Table Network Table Pathway Filter | |

4. Select the attributes you want displayed in the Table Panel. Attributes are listed with item type before the attribute (ex. Compound.name; Enzyme.name). The available attributes will depend on the type of network created.

| Table Panel | | | | | | | | | |
|---------------|------------------|--|--|--|--|--|--|--|--|
| ٠ | | | | | | | | | |
| Compound.name | | | | | | | | | |
| | Compound.smiles | | | | | | | | |
| Node T | direction | | | | | | | | |
| Enzyme.ecnum | | | | | | | | | |
| 00 | Gene.description | | | | | | | | |
| | < III > | | | | | | | | |

5. When you are done, click anywhere outside of the list and the table will be populated with the appropriate data.

Note: In Cytoscape 3.0, clicking on the two horizontal squares with check marks will select all attributes.

Below is a table of available attributes for each node type:

| A | В | | | | | |
|------------------------|--|--|--|--|--|--|
| Attribute 👻 | Description | | | | | |
| ID | Unique identifier, required by Cytoscape | | | | | |
| Compound.approximateMW | Compound Molecular weight (integer, data source - EHMN database) | | | | | |
| Compound.casnum | Chemical Abtracts Service compound identifier | | | | | |
| Compound.cid | KEGG compound unique identifier | | | | | |
| Compound.file_name | Rows with data specific to your experiments | | | | | |
| Compound.formula | Molecular formula | | | | | |
| Compound.mw | Molecular weight | | | | | |
| Compound.name | Compound name | | | | | |
| Compound.pubchemcid | PubChem compound identifier | | | | | |
| Compound.smiles | SMILES string for compound | | | | | |
| Compound.synonyms | Other names for compound | | | | | |
| Enzyme.ecnum | Enzyme Comission number | | | | | |
| Enzyme.name | Enzyme name | | | | | |
| Gene.file_name | Relative gene data values | | | | | |
| Gene.column_name | Gene data p-value | | | | | |
| Gene.description | Gene description | | | | | |
| Gene.human.geneid | Human gene identifier | | | | | |
| Gene.locations | Subcellular location | | | | | |
| Gene.symbol | Official gene symbol | | | | | |
| Genegeneid | Ortholog gene ID if the input organism is not human | | | | | |
| Reaction.equation | Equation for a reaction | | | | | |
| Reaction.locations | Reaction subcellular location (EHMN database) | | | | | |
| Reaction.pathway | Pathway involving reaction | | | | | |
| Reaction.reversible | T= reaction is reversible; F=reaction is non reversible | | | | | |
| Reaction.rid | Reaction identifier | | | | | |
| Туре | Node type (gene, enzyme, reaction or compound) | | | | | |
| canonicalName | Node name that is shown on node label | | | | | |

*Please note that the reactions shown in the Reaction.equation attribute display all compounds, while the networks built from the list of input genes/compounds show only "main" compounds (see <u>Ma et al., Mol. Syst. Biol., 2007, 3:135</u> for details). Pathway-specific networks show all compounds.

Below is a table of available attributes for each edge:

| A | | В | | | | | |
|---------------------|---|---|--|--|--|--|--|
| Attribute | - | Description 💌 | | | | | |
| Enzyme.names | | Enzyme name | | | | | |
| Enzyme.ecnums | | Enzyme Comission number | | | | | |
| Reaction.equation | | Equation for reaction | | | | | |
| Reaction.locations | | Reaction subcellular location (EHMN database) | | | | | |
| Reaction.pathway | | Pathway involving reaction | | | | | |
| Reaction.reversible | | T= reaction is reversible; F=reaction is non reversible | | | | | |
| Reaction.rid | | Unique identifier for reaction | | | | | |
| canonicalName | | Common name | | | | | |
| direction | | Direct or undirected | | | | | |
| interaction | | Type of interaction | | | | | |

6. Select nodes and/or edges in the graph to view their attributes in the Table Panel. If nothing is selected, attributes for all nodes and edges are shown.



Notes:

• Some attributes (such as Formula, Mass, and Smile) apply only to compound nodes, while others (such as Enzyme, Pathway, and Reversibility) apply only to reaction nodes.

Rearranging attributes in the table panel

To reorder attributes in the Table Panel, click on an attribute column heading and drag it to a new location:

| Compound.pubchemcid | Compound.cid | ∆ Co | mpound.name | |
|-------------------------------|----------------|-----------------------------|---------------|---------------------------|
| 5280749 | C05952 | Leukotrien | e E4 | Click on heading and drag |
| 6435286 | C05951 | Leukotrien | e D4 | |
| | CE5560 | glycolithocholate | | with mouse (notice: |
| 9903 | C03990 | 3alpha-Hydroxy-5beta-cholan | | Compound.name is |
| • | 054004 | III | 400 budeen is | currently to the right of |
| Node Table Edge Table Network | Table Pathwa | | | Compound.cid) |
| | | | • | |
| Compound.pubchemcid | ∆ Compound | name | Compound.ci | d |
| 5280749 Leukotriene E4 | | | C05952 | |
| 6435286 | Leukotriene D4 | | C05951 | |

 glycolithocholate
 CE5560

 3alpha-Hydroxy-5beta-cholan...
 C0390

 Image: Compound com

To sort by an attribute, click the attribute column heading; to sort in the reverse direction, click again.

| Compound.pubchemcid | | mpound.name | | Compound.cid | | |
|-----------------------------|---------------------|-------------|--------|--------------|--|--|
| 5280749 | 5280749 Leukotriene | | | C05952 | | |
| 6435286 | Leukotrien | e D4 | | C05951 | | |
| | glycolithoch | nolate | | CE5560 | | |
| 0000 | roxy-5beta-cholan | | C03990 | | | |
| Click on column title to so | | | | | | |
| Notice the arrow appears | ay Filter | | | | | |
| column is sorted A-Z (an a | arrow in | | | | | |
| the reverse direction app | | | | | | |
| when sorted Z-A). | | | | | | |

Pathway filter in the table panel

9903

Node Table

•

MetScape provides two ways to access pathway information. The **Pathway Filter** tab lists all pathways represented in the network. Selecting one or more pathways will highlight in the network all nodes in the pathway(s). Pathways are also displayed as attributes in the Node Table tab.

| Table Panel | | | | | | | | |
|---|-------------------|-------------------------|---|--|--|--|--|--|
| Select All Pathways Deselect All Pathways | Create Subnetwork | Reapply Selection Close | | | | | | |
| Pathways | | | | | | | | |
| 3-oxo-10R-octadecatrienoate beta-oxidation | | | | | | | | |
| Aminosugars metabolism | | | E | | | | | |
| Androgen and estrogen biosynthesis and metabolism | | | | | | | | |
| Arachidonic acid metabolism | | | | | | | | |
| Ble acid biosynthesis | | | | | | | | |
| Biopterin metabolism | | | | | | | | |
| Butanoate metabolism | | | | | | | | |
| C21-steroid hormone biosynthesis and metabolism | | | | | | | | |
| De novo fatty acid biosynthesis | | | | | | | | |
| Dimethyl-branched-chain fatty acid mitochondrial beta-oxidation | | | | | | | | |
| Di-unsaturated fatty acid beta-oxidation | | | Ŧ | | | | | |
| Node Table Edge Table Network Table Pathway Filter Concept Filter | | | | | | | | |

A subgraph can be created for a given pathway. The subgraph consists of a subset of nodes from an active network that belong to the selected pathway. To create a subgraph:

- 1. Select a pathway in the **Pathway Filter** tab.
- 2. Click the **Create Subnetwork** button at the top of the Table Panel (make sure you are on the **Pathway Filter** tab).
- 3. A new graph appears in the graph window. This is the subnetwork for the selected pathway.

Concept filter in the table panel

The content of the created or loaded concept file is put into the **Concept Filter** tab.

| | | | Di | rection of | f the change | |
|--------|---|----------------------------------|----------------------------|--------------|------------------------|------|
| Concep | t names | | | | | |
| | Table Panel | | | | P-value | 5 🗙 |
| | Select All Concepts Deselect | t All Concepts Create Subnetwork | k Save Concepts | Reappy Selec | tion Close | |
| 4 | Concept Name | Number of Enriched-Driving Genes | Number of Genes in Network | ↑↓ P-va | alue False Discovery F | Rate |
| | 1,4-Dichlorobenzene degradation | 21 | 9 | ↓ 0.35 | 9 0.583 | |
| | 1- and 2-Methylnaphthalene degradation | 25 | 7 | ↓ 0.29 | 3 0.546 | |
| | 3-Chloroacrylic acid degradation | 15 | 8 | ↓ 0.17 | 2 0.392 | |
| | Acute myeloid leukemia | 53 | 8 | ↑ 0.46 | 7 0.675 | |
| | Adherens junction | 74 | 12 | ↑ 0.04 | 2 0.198 | DR |
| | Adipocytokine signaling pathway | 71 | 12 | ↑ 0.90 | 1 0.95 | |
| | Node Table Edge Table Network Table Pathway F | lter Concept Filter | | | | |

1. Concept Name - official name of the concept.

2. Number of Enriched-Driving Genes – the number of enriched-driving genes in the concept.

3. Number of Genes in Network - the number of input genes that belong to the concept.

4. Direction – direction of the change.

5. P-value - the p-value for enrichment/depletion.

6. False Discovery Rate - the FDR, estimated using the method of Benjamini and Hochberg (When results are sorted by p-value, a FDR < 0.05 is interpreted to mean that approximately 5% of the concepts with lesser or equal p-value are false positives).

To map a concept on the graph, click on the concept name in the Table Panel. The concept will then be highlighted in the graph (associated nodes and edges turn yellow.).



A subgraph can be created for a given concept. The subgraph consists of genes considered to be significant to the concept (those genes that drive the concept), not all genes in the concept. To create a subgraph:

- 1. Select a concept in the **Concept Filter** tab.
- 2. Click the **Create Subnetwork** button at the top of the Table Panel (make sure you are on the **Concept Filter** tab).
- 3. A new graph appears in the graph window. This is the subnetwork for the selected concept.



The **Reapply Selection** button will re-select the last selected concept in the graph. If you select a concept and then select off of it, clicking the **Reapply Selection** button will re-select that concept.

The concept information can be saved as a file. To save this information:

- 1. Under Concept Filter tab, click the Save Concepts... button.
- 2. A new window pops up, asking where to save the file. Choose the desired location.
- 3. Click **Save**. The file saves as a CSV file that can be opened with most text editing or spreadsheet programs. The file format is identical to the input concept file.

| 🛃 Subnetw | ork1 | | | | | |
|----------------|--------------------------------------|------------------|-------------------------|-------------------------------|-----------|-----------------------|
| | Save Concept | ts | | | | |
| | Save in: Local Disk (Q:) | | | -] 💋 | | |
| | Recent Items | | | | | |
| | Choose loca Desktop for saving fi | | | n | | |
| | My Documents | | | | | A |
| | Computer | | | Click Sav | /e | A |
| | | File name: | | | | Save Xotprovein |
| Data Pane | Network | Files of type: | All Files | | • | Cancel |
| | | | | | | Save Concepts |
| Concept Nam | | ^ | | Number of Enriched-Driving Ge | nes Numbe | er of Mapped Genes |
| Citrate cyde | | | | 5 | 8 | 4.29771487419 |
| | kine receptor intera | ction | | 32 | 6 | 1 5.79931328275 |
| DNA replicatio | m | | 3 | | 3 | 2 50964646324 |
| Focal adhesio | | | | 53 | 4 | 970956 |
| | e and threonine met | | 2 | | 2 | Click on Save 1576989 |
| | ipid biosynthesis - gl | obo series | 1 | | 2 | 175247 |
| Hematopoietic | c cell lineage | | 6 | 4 | 1 | Concepts |
| Node Attribu | te Browser Edge A | ttribute Browser | Network Attribute Brows | er Concept Selection | | concepts |

Page 23 of 39

MetScape Results Panel (Additional Node Information)

Additional information can be obtained for each node by double clicking on it.

To get additional information about any compound, gene, reaction, or enzyme:

- 1. Double click on the node of interest.
- 2. On the right side of the screen, the MetScape Results Panel will show up.
- 3. Use the scroll bar or click on the undock icon 🗖 to see all the data.



Animation of Data

Create an animation of the data to see how it changes over time and across treatments.

Building an Animation

1. Create a subnetwork (for example, create a subnetwork for the TCA Cycle. See the <u>Concept Filter</u> section in the Table Panel section for how to do this).

Note: It is not necessary to create a subnetwork before doing data animation; the animation can be done on any network. For example, the Compounds file can have additional data columns (see image below) that are loaded when initially importing the Compounds file.

| A | В | С | D | E | F | G | Н | 1 |
|--------|----------|-------------|----------|----------|----------|----------|----------|----------|
| CID | p value | Fold-change | t1_0 | t1_1 | t1_2 | t2_0 | t2_1 | t2_2 |
| C00219 | 0.040982 | 1.49 | 0.477255 | 0.811411 | 0.907349 | 0.677623 | 0.349178 | 0.686385 |
| C00503 | 0.008525 | 1.53 | 0.902665 | 0.698903 | 0.309585 | 0.33507 | 0.761891 | 0.042653 |
| C00187 | 0.030047 | 1.85 | 0.489314 | 0.029276 | 0.314592 | 0.705353 | 0.084872 | 0.340032 |
| C02721 | 0.024311 | 2.81 | 0.86923 | 0.125058 | 0.843886 | 0.667656 | 0.008175 | 0.131301 |
| C00047 | 0.017356 | 1.03 | 0.312654 | 0.651588 | 0.386625 | 0.413499 | 0.650567 | 0.926665 |
| C16868 | 0.004029 | 2.01 | 0.19994 | 0.379436 | 0.374019 | 0.064547 | 0.325439 | 0.605136 |
| C05122 | 0.000312 | 1.75 | 0.002804 | 0.477482 | 0.486432 | 0.683021 | 0.562158 | 0.268456 |
| C00064 | 0.000021 | -1.2 | 0.113069 | 0.635121 | 0.403885 | 0.769097 | 0.885804 | 0.783052 |
| C05629 | 0.000252 | -1.38 | 0.775584 | 0.198474 | 0.490499 | 0.713089 | 0.756346 | 0.96571 |
| C00079 | 0.036583 | -1.15 | 0.139549 | 0.464563 | 0.037484 | 0.041068 | 0.592104 | 0.816224 |
| C00398 | 0.016353 | -1.07 | 0.115421 | 0.541825 | 0.8363 | 0.829824 | 0.001483 | 0.302741 |
| C00294 | 0.000014 | -1.4 | 0.125432 | 0.275928 | 0.741933 | 0.513364 | 0.90882 | 0.575744 |

2. From the **Apps** menu, choose **MetScape -> Animate...**



- 3. A new window pops up, titled Initialize Coupled Animation of Multiple Data Columns.
- 4. Column headings from your data are listed on the right side of the new window. Determine how many animations (number of rows) and data items per animation (number of columns) you will need. Enter this information into the Rows and Cols boxes. If you change these numbers from the default, click on **Reconfigure** Layout.

| | number of rows) and the | the right into the table of data is number of data items per animal | | | the number of he text fields and |
|--------|-------------------------|--|-------------|--------|-------------------------------------|
| Rows 2 | Cols 3 | Reconfigure Layout | Restart | Cancel | Build Animatio |
| | Column 0 | Column 1 | Column 2 | | p value |
| t | (Undefined) | (Undefined) | (Undefined) | | Fold-change |
| t | (Undefined) | (Undefined) | (Undefined) | | t1_0 t1_1 |
| | | | | | t1_2 t2_0 t2_1 t2_2 |

5. Click and drag the column names into the grid to build the animation.

| Network title: Co | abels from the list on the right abels from the list on the right | f Multiple Data Column syme-Gene spht into the table of data lab ther of data items per animab | els for each an | | | Network title: Drag and dro | Compound-Reaction plabels from the list on pumber of rows) and the | on of Multiple Data Colu n-Enzyme-Gene the right into the table of data las e number of data items per animal | bels for each a | | |
|-------------------|--|---|-----------------|--------|-----------------|--------------------------------|--|--|-----------------|--------|---------------------|
| Rows 2 | Cols 3 | Reconfigure Layout | Restart | Cancel | Build Animation | Rows 2 | Cols 3 | Reconfigure Layout | Restart | Cancel | Build Animation |
| | Column 0 | Column 1 | Column 2 | | p value | | _0 | Column 1 | Column 2 | | p value |
| DataSeries_0 | | t1_1 | t1_2 | | 12.0 | DataSeries | | t1_1 | t1_2 | | Fold-change t2_1 |
| DataSeries_1 | (Undefined) | (Undefined) | (Undefined) | | t2 1 | DataSeries | _t t2_0 | (Undefined) | (Undefined) |) | t2_1 |
| | Click headir and drop in | | | | 12_2 | | | Heading dropp appropriate loo in grid | | | |

6. After grid is completely filled in, click on **Build Animation**.

Note: If you want to undo what you filled out in the grid, click on **Restart** and the grid will be cleared.

7. Two new windows appear in the graph panel, each representing a different treatment. In addition, an **Animation Controls** window appears.



Manipulating the animation: zooming

- 1. **Zoom** in on one graph to the appropriate view size (zoom the same way you do with any Cytoscape graph).
- 2. On the **Animation Controls** window, click on **Realign All**. This brings all the treatment graphs to the same view.



Page 27 of 39

Manipulating the animation: bar chart

1. The bar chart in the **Animation Controls** window shows the range of the data and its frequency. The range of measurements are on the x-axis, while the frequency with which they occur in the experimental data set are on the y-axis.



- 2. Click on a bar in the bar chart and a vertical line (slider) will show up on the bar chart. The slider can be moved by clicking on the bar chart and dragging the line. When moved, the values shown in black are also shown in black in the graph and the set of the color scale is reset.
- 3. The selected portion of the graph determines what the animation will show. This allows outliers to be removed if desired.
- 4. The color range can be changed by clicking the dropdown arrow next to **Blue-Red**.



Playing the animation

To play the animation, click **Play** on the **Animation Controls** window. Colors and node sizes change within the three treatments, showing changes over time. To stop the animation, click **Stop** on the **Animation Controls** window.



Page **28** of **39**

Manipulating Network

Expanding a network

From a compound node, you can expand the current network to include additional reactions and related compounds, genes, and enzymes.

Option 1: Expanding a network in the current network window (Example using Compound-Reaction Network)

1. Right click on the compound node you want to expand (this node is known as the expansion seed node). A menu of options will pop up.



2. Go to MetScape -> Expand -> Expand in Existing Network

- 3. Additional compounds and reactions are added to the network. As a result, the network is often redrawn.
- 4. The edges between the original node and the expanded nodes are now blue.



Option 2: Expanding a network in a new window (Example using Compound Network from Selected Genes)

1. Right click on the compound node you want to expand (this node is known as the expansion seed node). A menu of options will pop up.



2. Select MetScape -> Expand -> Expand in Subnetwork

3. A new network is created in a new window. This network includes only the expansion seed compound and its related compounds, reactions, enzymes, and genes (depending on the Network Type). The original color designations are used when the graph is created in a new window.



Collapsing a network

To collapse a network that is expanded in a current network window:

- 1. Right click on the expansion seed compound node.
- 2. Select **MetScape -> Collapse** and then:
 - To collapse only the branch expanded from that expansion seed compound, select **Collapse**.
 - To collapse all expanded branches, select **Restore Original Network**.

Creating a subnetwork

A subnetwork of a current network can be created and will appear in a separate window. A subnetwork will include all highlighted nodes and edges (selected nodes and edges should be yellow).

For example:

- 1. Select a compound in the built network. For this example, I will select Sarcosine.
- 2. Go to the **Select** menu in Cytoscape.
- 3. Choose Nodes -> First Neighbors of Selected Nodes -> Undirected. Now all the first neighbors of Sarcosine should be yellow.
- 4. Go to the **Select** menu in Cytoscape.
- Choose Edges -> Select Adjacent Edges. Now all the adjacent edges should be yellow.
- 6. Right click on Sarcosine. A menu of options will pop up.



Page 31 of 39

7. Choose MetScape -> Create Subnetwork.

Note: When choosing options from the menu, be careful not to move the cursor outside the menu panels. Doing so will remove the highlighting of the nodes and edges, resulting in an empty subnetwork.



8. A new subnetwork is created in the graph window.

Destroying a network

To destroy a network no longer needed:

- 1. Make sure you really do want to destroy the network.
 - Cytoscape will ask you to confirm the deletion.
 - Destroying the network is irreversible.
 - **Note:** If you want to destroy a network view without destroying the network itself, use **Destroy View** instead of **Destroy Network**.
- 2. On the **Network** tab in the Cytoscape Control Panel, right click on the network you want to destroy and select **Destroy Network**.

Note: If you destroy a network that has subnetworks, the subnetworks are **NOT** destroyed. They are promoted up one network level.

PART 2: CORRELATION-BASED NETWORK

Entering Data

File Formats

Two correlation file formats are accepted (column-based is recommended):

- 1. Column-based
 - The first row must contain column headings of the user's choosing
 - The first two columns contain metabolite names
 - The next column(s) contains values, such as p-values.

Below is a portion of an example column-based correlation file:

| | А | В | С | D | E |
|----|---------|------------|----------|----------|----------|
| 1 | metab1 | metab2 | pcor | pval | qval |
| 2 | Alanine | Alanine | 1 | 0 | 0 |
| 3 | Alanine | Sarcosine | 0.110571 | 0.101242 | 0.287801 |
| 4 | Alanine | Glycine | 0.255278 | 0.000101 | 0.001017 |
| 5 | Alanine | Alpha-am | 0.091832 | 0.1743 | 0.428134 |
| 6 | Alanine | Valine | 0.040047 | 0.554916 | 0.74299 |
| 7 | Alanine | Leucine | 0.006732 | 0.921006 | 0.960585 |
| 8 | Alanine | Isoleucine | 0.071432 | 0.291444 | 0.548247 |
| 9 | Alanine | Threonine | 0.100668 | 0.136092 | 0.358658 |
| 10 | Alanine | Serine | -0.13329 | 0.047578 | 0.164894 |
| 11 | Alanine | Proline | 0.468989 | 5.19E-15 | 6.56E-13 |
| 12 | Alanine | Asparagin | -0.05741 | 0.396934 | 0.616099 |
| 13 | Alanine | Aspartic a | 0.059375 | 0.380926 | 0.609963 |
| 14 | Alanine | Methionir | 0.102612 | 0.128614 | 0.346165 |
| 15 | Alanine | 4-Hydroxy | -0.11061 | 0.101123 | 0.287801 |

2. Matrix format

- The first row and column contain metabolite names
- The rest of the rows/columns contain correlation values

Below is a portion of an example matrix format correlation file:

| | А | В | С | D | E | F | G | Н | 1 |
|----|-------------|----------|-----------|----------|-----------|----------|----------|------------|-----------|
| 1 | | Alanine | Sarcosine | Glycine | Alpha-ami | Valine | Leucine | Isoleucine | Threonine |
| 2 | Alanine | 1 | 0.215426 | 0.171592 | 0.167991 | 0.450516 | 0.287503 | 0.352786 | 0.213065 |
| 3 | Sarcosine | 0.215426 | 1 | 0.060618 | 0.158684 | 0.21722 | 0.106856 | 0.175509 | 0.074988 |
| 4 | Glycine | 0.171592 | 0.060618 | 1 | 0.130999 | -0.08656 | -0.12006 | -0.09012 | 0.164396 |
| 5 | Alpha-amino | 0.167991 | 0.158684 | 0.130999 | 1 | 0.417381 | 0.322295 | 0.266963 | 0.064821 |
| 6 | Valine | 0.450516 | 0.21722 | -0.08656 | 0.417381 | 1 | 0.773163 | 0.775301 | 0.131847 |
| 7 | Leucine | 0.287503 | 0.106856 | -0.12006 | 0.322295 | 0.773163 | 1 | 0.884062 | 0.165156 |
| 8 | Isoleucine | 0.352786 | 0.175509 | -0.09012 | 0.266963 | 0.775301 | 0.884062 | 1 | 0.258682 |
| 9 | Threonine | 0.213065 | 0.074988 | 0.164396 | 0.064821 | 0.131847 | 0.165156 | 0.258682 | 1 |
| 10 | Serine | 0.070274 | -0.03839 | 0.441288 | 0.104243 | 0.046118 | 0.023038 | 0.09493 | 0.553733 |
| 11 | Proline | 0.593237 | 0.129416 | 0.071866 | 0.02707 | 0.517814 | 0.420597 | 0.452117 | 0.306851 |
| 12 | Asparagine | 0.110042 | 0.039533 | 0.282938 | 0.156343 | -0.01262 | 0.106846 | 0.066391 | 0.602028 |

Building a Network

- 1. To begin a Cytoscape session with the MetScape app, first start Cytoscape.
- 2. Select **Apps -> MetScape-> Build Network -> Correlation-based** from the Cytoscape menu.
- 3. A MetScape tab now appears on the left of the Cytoscape screen.
- 4. Use the **Select** button to upload the appropriate file.
- 5. Compound names will map to their KEGG id; a popup window will appear. If there is more than one potential match, use the dropdown arrow to choose the best match. If a compound is not found in the system, it will say "Not Found." The mapping selection will be saved so that your selection will appear as the default option in the future.

| | Select names for compo | ounds | |
|----------------------------|--|--------------------------|---|
| in the datab | of the compounds in your quer ase. Please select the match the wish the compound to appear i | at is best - or '(none)' | |
| Input Name | Potential Matches | KEGG ID | |
| 4-Hydroxyproline | Not Found | | - |
| Alanine | (5-L-glutamyl)-L-alanine | ▼ CE0469 | |
| Alpha-aminoisobutyric acid | Not Found | | |
| Asparagine | L-Asparagine | • <u>C00152</u> | |
| Aspartic acid | D-Aspartic acid | ▼ <u>C00402</u> | E |
| Cystine | Homocystine | | |
| Glutamic acid | D-Glutamic acid | ▼ <u>C00217</u> | |
| Glutamine | D-Glutamine | ▼ C00819 | |
| Glycine | Glycine | • <u>C00037</u> | |
| Histidine | 2-(3-Carboxy-3-aminopr | ▼ <u>C04441</u> | |
| Isoleucine | L-Isoleucine | ▼ <u>C00407</u> | |
| Leucine | (2S)-alpha-Leucine | - <u>C00123</u> | |
| Lysine | 3-Hydroxy-N6,N6,N6-tri | ▼ <u>C01259</u> | |
| Methionine | Methionine | - <u>C00073</u> | |
| Ornithine | D-Ornithine | · C00515 | - |

- 5. Click OK.
- 6. Under Edge Mapping, use the dropdown menu next to **Base Edges on** and select the appropriate column from your data file.

Note: The **Range for Edges** slider changes to match the data type chosen. In the example files used in this manual, the slider is 0 to 1 for pval but -1 to 1 for pcor.

| Range for Edges | | Edge Mapping based on Pval from Example File | Range for Ed | ges | | Edge Mapping based on pcor from Example File | <u> </u> |
|-----------------|-------------|--|--------------|------|------------|--|----------|
| | ignificance | | Vegative | | Correlatio | on 💽 | Positive |
| 0.00 | 1.00 | | -1.00 | 0.00 | 0.0 | 0 1.00 | |
| 0 | | 1 | -1 | -0.5 | | 0.5 | 1 |

- 7. Under Edge Mapping, use the dropdown menu next to **Tooltip Labels** to select additional values that will be viewable when mousing over an edge in the built network.
- 8. **Optional:** Load **group definition** file. For more information, go to the <u>Group</u> <u>Definition</u> section of this document.
- 9. Under Range for Edges, drag the arrows left and right OR enter numbers in the text boxes to select the desired **significance range**. A log scale is used to allow for very small p-values.
- 10. The number of edges that will appear in the network built with the current parameters will appear below the slider.



11. Click Build Network.



Group Definition

MetScape provides the option of creating groups using a **group definition file**. A group definition file is a simple 2-column file with metabolite names in the first column and group names in the second column. Group names can be anything that you choose. A group definition file can be loaded using the **Select...** button under Group Definition on the MetScape tab. If a group definition file is used, a **Group Filter tab** will appear in the Table Panel after the network is drawn. Clicking on a group name in the Group Filter tab will select all nodes in the network that are part of that group.

Correlation Visualization

MetScape provides a legend explaining its various shapes and colors. The legend will be specific to the current network type (pathway or correlation).

While on a correlation network, access the MetScape Legend from Apps -> MetScape -> Show Legend.

| Legend |
|----------------------|
| Legend |
| |
| Known Compound |
| Unknown Compound |
| Correlation Strength |
| Positive Correlation |
| Negative Correlation |
| Significance-Based |
| Close |
| |

Additional information about the network is expressed through visualization:

- Edges
 - o If the dataset only has values 0 to 1, all edges will be black.
 - If a dataset has positive and negative correlation values:
 - A pink edge represents a positive correlation.
 - A blue edge represents a negative correlation.
 - The thicker the edge, the stronger the correlation.

Note: In a correlation-based network, selected edges are highlighted yellow instead of red.

- Nodes:
 - o Purple nodes represent mapped compounds.
 - White nodes represent compounds that did not map to a known compound in the MetScape database.

Additional Information

Node and edge attributes are displayed in the Table Panel. See the <u>Table Panel Display</u> section above for details on this panel. A Pathway Filter tab is also available in the Table Panel. See the <u>Pathway</u> <u>Filter in the Table Panel</u> section above for details about this panel.

When a correlation-based pathway is built, its name in the Network tab reflects significance information for that network. This naming convention is to help distinguish between networks when creating multiple networks in the same Cytoscape session.

Correlation Network Results Panel

Details about nodes can be viewed by double clicking on a node of interest. To learn more about the details window for **known compounds**, see the <u>MetScape Results Panel</u> section above. Double clicking on an **unknown compound** will bring up a Results Panel that contains links to <u>HMDB</u>, <u>ChemSpider</u>, <u>MassBank</u>, and <u>METLIN</u>.

| Results Panel | × □ |
|-------------------------|------------|
| Node/Edge Details | |
| Close | |
| Compound Node | |
| Links for m/z searches: | |
| HMDB | |
| ChemSpider | |
| <u>MassBank</u> | |
| METLIN | |
| | |

Each of these links will allow you to search based on **mass-to-charge values**, helping to identify the unknown compound.

In addition, a **node attribute file** can be imported into Cytoscape with one column for nodes and a second column for mass-to-charge. The mass-to-charge column needs to have one of the following as the heading: m/z, M/Z, m\z, M\Z, mz, or MZ. If the mass-to-charge values are entered for unknown compounds, clicking on the HMDB link in the Results Panel will go to the HMDB results for that mass-to-charge value. Please note that the other three links will still go to their search pages.

SAVING AND REOPENING A SESSION

Saving a session

To save a Cytoscape session containing one or more MetScape app networks:

- 1. Select **File -> Save** (or **Save As...**) from the Cytoscape menu.
- 2. Browse to a location for saving the file.
- 3. Name the file.
- 4. Click Save.

Reopening a session

To reopen a saved session containing MetScape app data:

- 1. Select **File -> Open** from the Cytoscape menu.
- 2. Navigate to the saved file location.
- 3. Select the file.
- 4. Click Open.
- 5. After Cytoscape reports that the session file was successfully loaded, click Close.

SUPPLEMENT: METDISEASE APP

MetDisease is designed to annotate metabolites with Medical Subject Headings (MeSH) disease terms. The underlying data comes from the Metab2MeSH data set (Sartor et al., Bioinformatics. 2012 May 15;28(10):1408-10). To use MetDisease, you need a network of compounds with KEGG or PubChem Ids or compound name. If using MetScape 2.3.2 or higher, you can use MetDisease on the MetScape networks.

How to use MetDisease

Once you have built a MetScape network:

- 1. Select Apps -> MetDisease -> Find MeSH Terms...
- 2. Choose the appropriate identifier type and attribute. Click **OK**.
- 3. A hierarchical tree of MeSH Disease terms should be displayed in the Table Panel.
- 4. Selecting a tree node will select respective metabolites in the network. The numbers shown after the descriptors represent the number of matching metabolites in the active network.

Note: Descriptors that have no nodes in the active network can be collapsed and hidden.

5. To view relevant PubMed citations, right click on a selected compound node and select **MetDisease -> PubMed Citations**.

For more information about MetDisease, go to <u>http://metdisease.ncibi.org/</u>. The MetDisease user manual can be downloaded at <u>http://metdisease.ncibi.org/pdf/MetDisease-User-Manual.pdf</u>.