

# Pathway Analysis with MetaboAnalyst

From a talk by  
David Wishart

[https://www.youtube.com/watch?v=1EEl9Cze\\_0I&t=319s](https://www.youtube.com/watch?v=1EEl9Cze_0I&t=319s)

# Metabolite Set Enrichment Analysis (MSEA)



<http://www.msea.ca>  
Now part of MetaboAnalyst

- Designed to handle lists of metabolites (with or without concentration data)
- Modeled after Gene Set Enrichment Analysis (GSEA)
- Supports over-representation analysis (ORA), single sample profiling (SSP) and quantitative enrichment analysis (QEA)
- Contains a library of 6300 pre-defined metabolite sets including 85 pathway sets & 850 disease sets

# Enrichment Analysis

**Purpose:** To test if there are **biologically meaningful** groups of metabolites that are significantly enriched in your data

**Biological meaningful in terms of:**

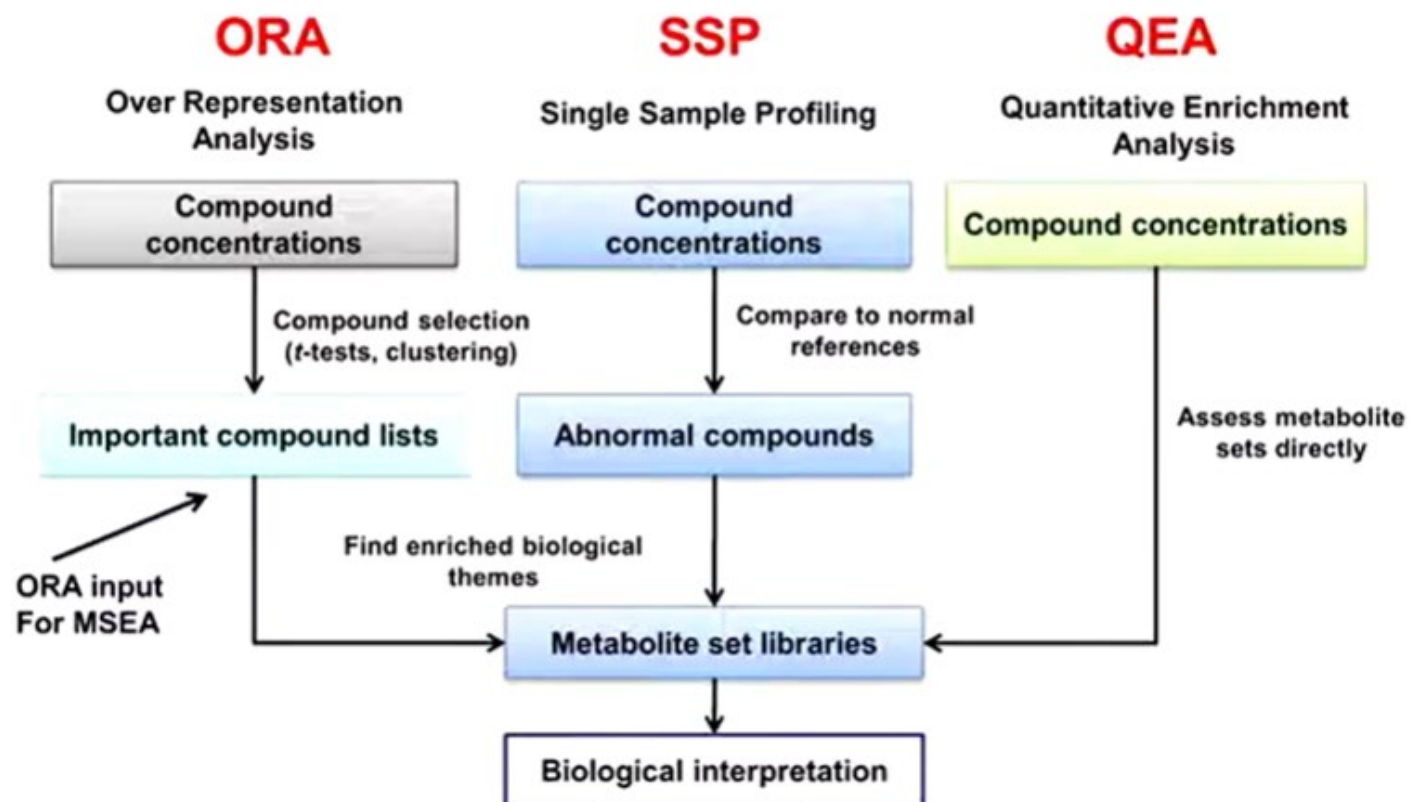
- Pathways
- Disease
- Localization

**Currently, MSEA only supports human metabolomic data**

# MSEA

- **Accepts 3 kinds of input files**
  - list of metabolite names only (ORA – over representation analysis)
  - list of metabolite names + concentration data from a single sample (SSP – single sample profiling)
  - a concentration table with a list of metabolite names + concentrations for multiple samples/patients (QEA – quantitative enrichment analysis)

# The MSEA Approach



# Data Set Selected

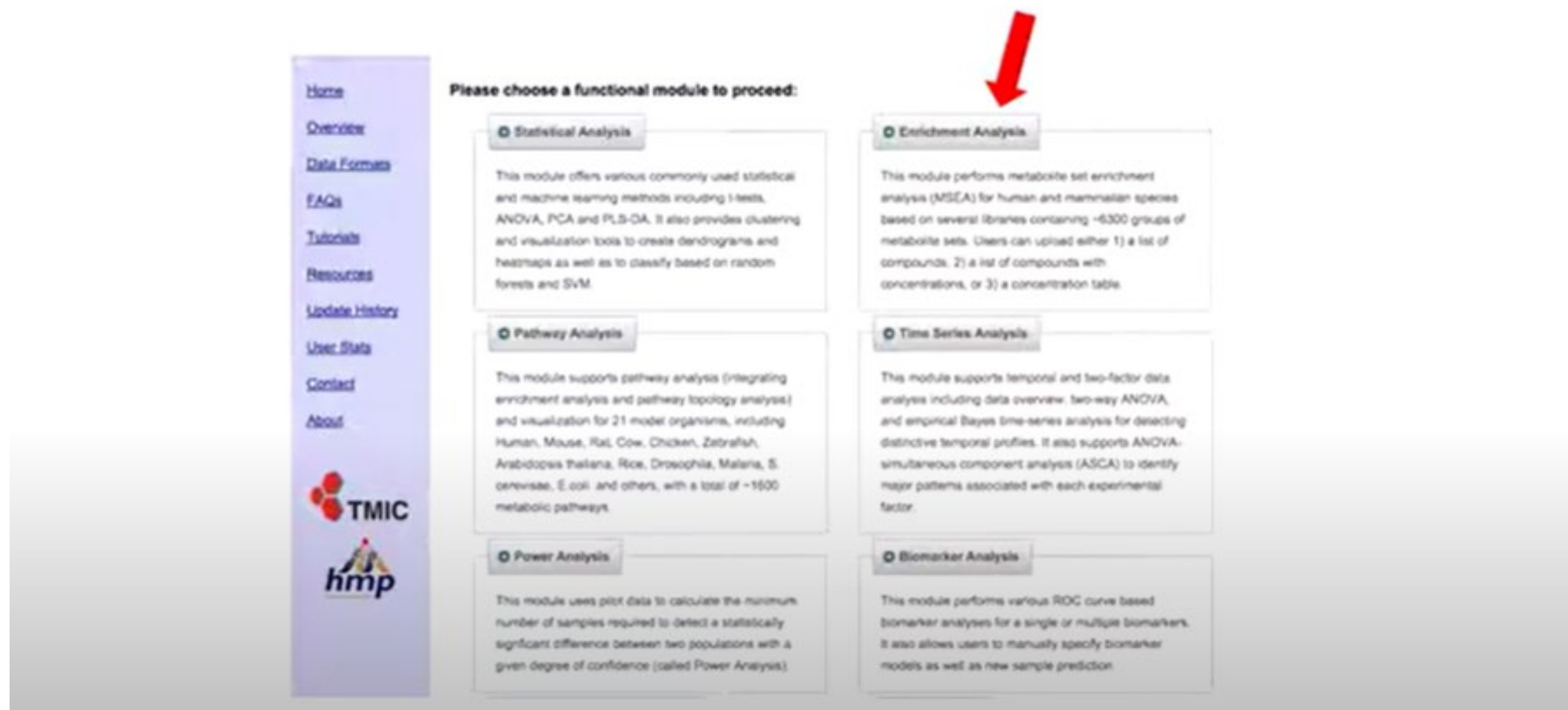
- Here we are using a collection of metabolites identified by NMR (compound list + concentrations) from the urine from 77 lung and colon cancer patients, some of whom were suffering from cachexia (muscle wasting)

2) Try our test data : ( You can download these data [here](#) )

Data Type	Description
<input checked="" type="radio"/> Concentrations <a href="#">Tutorial Report</a>	Metabolite concentrations of 77 urine samples from cancer patients measured by 1H NMR ( <a href="#">Eisner R, et al.</a> ). Group 1- cachexic; group 2 - control
<input type="radio"/> Concentrations	Metabolite concentrations of 39 rumen samples measured by proton NMR from dairy cows fed with different proportions of barley grain ( <a href="#">Amatai BN, et al.</a> ). Group label - 0, 15, 30, or 45 - indicating the percentage of grain in diet.
<input type="radio"/> NMR spectral bins <a href="#">Tutorial Report</a>	Binned 1H NMR spectra of 50 urine samples using 0.04 ppm constant width ( <a href="#">Psilogios NG, et al.</a> ) Group 1- control; group 2 - severe kidney disease.



# Start with a Compound List for ORA



The screenshot displays the TMIC hmp website interface. On the left is a vertical navigation menu with links: Home, Overview, Data Formats, FAQs, Tutorials, Resources, Update History, User Stats, Contact, and About. Below these links are the TMIC and hmp logos. The main content area is titled "Please choose a functional module to proceed:" and contains six modules arranged in a 3x2 grid. A large red arrow points to the "Enrichment Analysis" module.

**Please choose a functional module to proceed:**

- Statistical Analysis**  
This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA and PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify based on random forests and SVM.
- Enrichment Analysis**  
This module performs metabolite set enrichment analysis (MSEA) for human and mammalian species based on several libraries containing ~6300 groups of metabolite sets. Users can upload either 1) a list of compounds, 2) a list of compounds with concentrations, or 3) a concentration table.
- Pathway Analysis**  
This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 21 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, S. cerevisiae, E.coli and others, with a total of ~1600 metabolic pathways.
- Time Series Analysis**  
This module supports temporal and two-factor data analysis including data overview, two-way ANOVA, and empirical Bayes time-series analysis for detecting distinctive temporal profiles. It also supports ANOVA-simultaneous component analysis (ASCA) to identify major patterns associated with each experimental factor.
- Power Analysis**  
This module uses pilot data to calculate the minimum number of samples required to detect a statistically significant difference between two populations with a given degree of confidence (called Power Analysis).
- Biomarker Analysis**  
This module performs various ROC curve based biomarker analyses for a single or multiple biomarkers. It also allows users to manually specify biomarker models as well as new sample prediction.

# Upload Compound List

Choose one of the following options to proceed

☒ A list of compound names (over representation analysis)

Please enter a one-column compound list:

Acetoacetic acid  
Beta-Alanine  
Creatine  
Dimethylglycine  
Fumaric acid  
Glycine  
Homocysteine  
L-Cysteine  
L-Isoleucine  
L-Phenylalanine  
L-Serine  
L-Threonine  
L-Tyrosine  
L-Valine  
Phenylpyruvic acid  
Propionic acid  
Pyruvic acid  
Sarcosine

Input Type:

☒ Use example data (input type: compound names)

☐ A list of compounds with concentration values (single sample profiling)

☐ A concentration table (quantitative enrichment analysis)

Normally GSEA would require a list of all known genes for the given platform. Here we just use the list of metabolites found in KEGG. ORA is a “weak” analysis in MSEA



# Perform Compound Name Standardization



## Compound Name/ID Standardization:

Please note:

- Query names in normal white indicate exact match - marked by "1" in the downloaded file;
- Query names highlighted in red indicate no match - marked by "0" in the downloaded file;
- For compound name mapping, the no match query names will be highlighted in yellow indicate no exact match found. You should click the **View** link to perform **approximate search** and manually select the correct match if found;
- Greek alphabets are not recognized, they should be replaced by English names (i.e. alpha, beta)

Query	Hit	HMDB	PubChem	KEGG	Details
Acetic acid	Acetic acid	HMDB000060	96	C00156	
Beta-Alanine	Beta-Alanine	HMDB000096	229	C00099	
Creatine	Creatine	HMDB000064	585	C00200	
Dimethylglycine	Dimethylglycine	HMDB000092	673	C00126	
Fumaric acid	Fumaric acid	HMDB001134	664972	C00122	
Glycine	Glycine	HMDB001123	750	C00031	
Homocysteine	Homocysteine	HMDB001742	778	C06330	
L-Cysteine	L-Cysteine	HMDB000674	9862	C00097	
L-Histidine		-	-	-	<a href="#">View</a>
L-Phenylalanine	L-Phenylalanine	HMDB001159	8140	C00079	
L-Serine	L-Serine	HMDB001157	5951	C00065	
L-Threonine	L-Threonine	HMDB001167	6288	C00188	
L-Tyrosine	L-Tyrosine	HMDB001158	5057	C00082	
L-Valine	L-Valine	HMDB000683	6287	C00183	
Phenylpyruvic acid	Phenylpyruvic acid	HMDB002205	397	C00158	
Propionic acid	Propionic acid	HMDB000217	1032	C00163	
Pyruvic acid	Pyruvic acid	HMDB000243	1080	C00092	
Sarcosine	Sarcosine	HMDB000271	1086	C00213	

You can download the result [here](#)

Submit

## Name Standardization (cont.)

Query name: **L-Isoleucine**

	Matched Name	HMDB	PubChem	KEGG
<input checked="" type="checkbox"/>	L-Isoleucine	<a href="#">HMDB00172</a>	<a href="#">6306</a>	<a href="#">C00407</a>
<input type="checkbox"/>	L-Alloisoleucine	<a href="#">HMDB01056</a>	<a href="#">98792</a>	<a href="#">C00415</a>
<input type="checkbox"/>	(+/-)-erythro-Isoleucine	<a href="#">HMDB33923</a>	<a href="#">791</a>	<a href="#">C16434</a>
<input type="checkbox"/>	Nebularine	<a href="#">HMDB39001</a>	<a href="#">245869</a>	
<input type="checkbox"/>	L-gamma-glutamyl-L-isoleucine	<a href="#">HMDB29884</a>	<a href="#">11469098</a>	
<input type="checkbox"/>	Angiotensin IV	<a href="#">HMDB01564</a>	<a href="#">123727</a>	<a href="#">C00570</a>
<input type="checkbox"/>	Casomorphin	<a href="#">HMDB34173</a>	<a href="#">4063834</a>	<a href="#">C10514</a>
<input type="checkbox"/>	None of the above			

OK Cancel

# Select a Metabolite Set Library

Home

Processing

Normalization

Enrichment

**Set parameters**

View result

Download

Exit

Set parameters for enrichment analysis:

Please select a metabolite set library

☒ Pathway-associated metabolite sets

This library contains 88 metabolite sets based on normal metabolic pathways.

☐ Disease-associated metabolite sets (Blood)

This library contains 416 metabolite sets reported in human blood.

☐ Disease-associated metabolite sets (Urine)

This library contains 346 metabolite sets reported in human urine.

☐ Disease-associated metabolite sets (CSF)

This library contains 124 metabolite sets reported in human cerebral spinal fluid (CSF).

☐ SNP-associated metabolite sets

This library contains 4,500 metabolite sets based on their associations with the detected single nucleotide polymorphisms (SNPs) loci.

☐ Predicted metabolite sets

This library contains 912 metabolic sets that are predicted to be changed in the case of dysfunctional enzymes using genome-scale network model of human metabolism.

☐ Location-based metabolite sets

This library contains 57 metabolite sets based on organ, tissue, and subcellular localizations.

☐ Self-defined metabolite sets

[Click here to upload your own customized metabolite set library](#)

☐ Only use metabolite sets containing at least  compounds

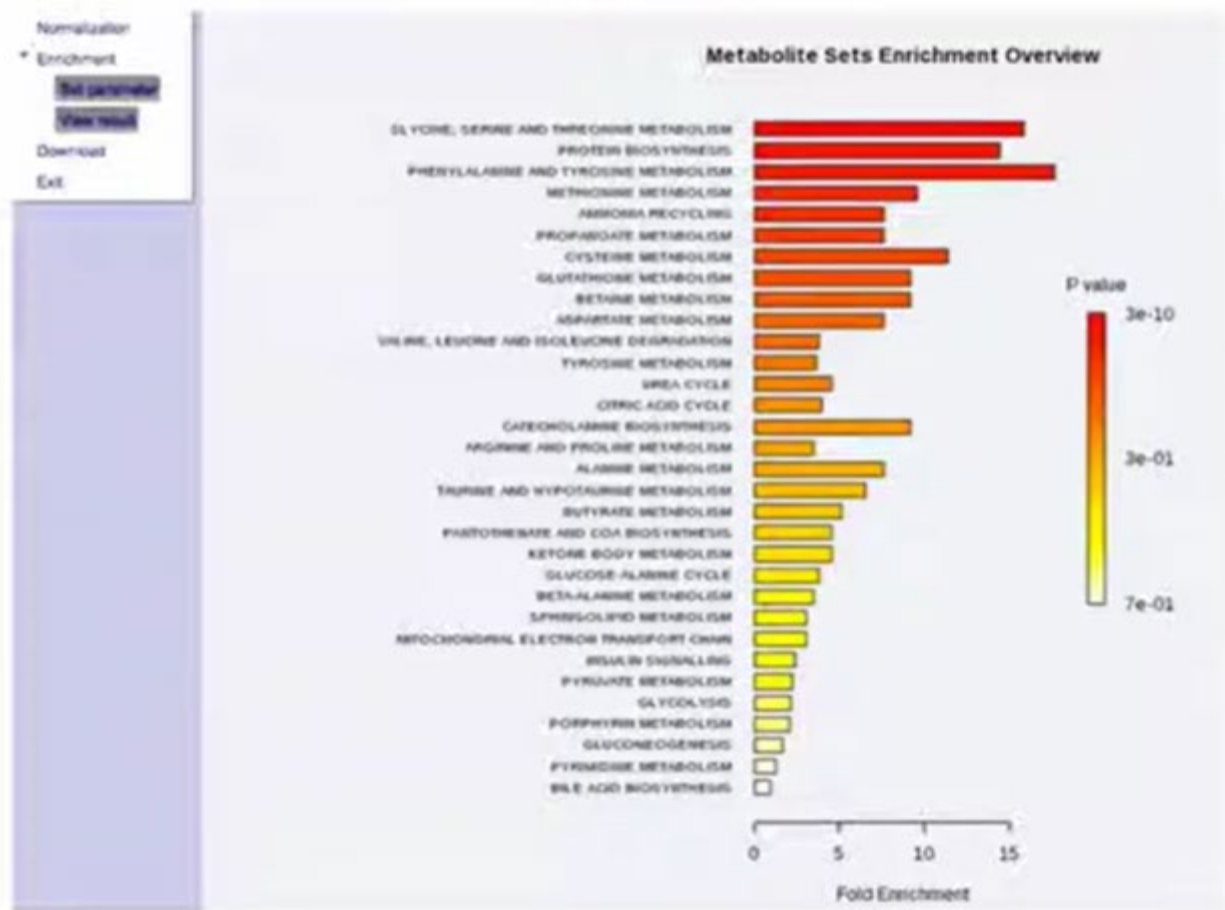
Please specify a reference metabolome

☒ Use all the compounds in the selected metabolite set library

☐ Upload a reference metabolome based on your analytical platform

Submit

# Result



# Result (cont.)

Click on details  
to see more

Metabolic Set	Total	Hits	Expect	P value	Mean P	FDR	Details
GLYCINE, SERINE AND THREONINE METABOLISM	25	9	0.567	2.74E-10	2.19E-8	2.19E-8	<a href="#">View</a>
PROTEIN BIOSYNTHESIS	19	6	0.415	9.93E-7	7.85E-5	3.97E-5	<a href="#">View</a>
PHENYLALANINE AND TYROSINE METABOLISM	13	5	0.294	3.15E-6	2.46E-4	6.4E-5	<a href="#">View</a>
METHIONINE METABOLISM	24	5	0.524	8.98E-5	0.00991	0.0018	<a href="#">View</a>
AMMONIA RECYCLING	18	3	0.393	0.00581	0.441	0.0774	<a href="#">View</a>
PROPIONATE METABOLISM	18	3	0.393	0.00581	0.441	0.0774	<a href="#">View</a>
CYSTEINE METABOLISM	8	2	0.175	0.0117	0.883	0.133	<a href="#">View</a>
GLUTATHIONE METABOLISM	10	2	0.218	0.0183	1.0	0.162	<a href="#">View</a>
BETAIN METABOLISM	10	2	0.218	0.0183	1.0	0.162	<a href="#">View</a>
ASPARTATE METABOLISM	13	2	0.262	0.0261	1.0	0.209	<a href="#">View</a>
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	36	3	0.795	0.0397	1.0	0.298	<a href="#">View</a>
TYROSINE METABOLISM	30	3	0.829	0.0456	1.0	0.304	<a href="#">View</a>
UREA CYCLE	20	2	0.436	0.0677	1.0	0.417	<a href="#">View</a>
CITRIC ACID CYCLE	23	2	0.502	0.0968	1.0	0.496	<a href="#">View</a>
CATECHOLAMINE BIOSYNTHESIS	5	1	0.109	0.105	1.0	0.536	<a href="#">View</a>
ARGININE AND PROLINE METABOLISM	26	2	0.567	0.107	1.0	0.536	<a href="#">View</a>
ALANINE METABOLISM	6	1	0.131	0.124	1.0	0.585	<a href="#">View</a>
TAURINE AND HYPOTAURINE METABOLISM	7	1	0.153	0.144	1.0	0.638	<a href="#">View</a>
BUTYRATE METABOLISM	9	1	0.196	0.181	1.0	0.758	<a href="#">View</a>
PANTOTHENATE AND COA BIOSYNTHESIS	10	1	0.218	0.199	1.0	0.768	<a href="#">View</a>
KETONE BODY METABOLISM	10	1	0.218	0.199	1.0	0.758	<a href="#">View</a>
GLUCOSE-ALANINE CYCLE	10	1	0.262	0.234	1.0	0.851	<a href="#">View</a>
BETA-ALANINE METABOLISM	13	1	0.284	0.251	1.0	0.873	<a href="#">View</a>
SPHINGOLIPID METABOLISM	15	1	0.327	0.284	1.0	0.908	<a href="#">View</a>
MITOCHONDRIAL ELECTRON TRANSPORT CHAIN	15	1	0.327	0.284	1.0	0.908	<a href="#">View</a>
INSULIN SIGNALLING	19	1	0.415	0.345	1.0	1.0	<a href="#">View</a>
PYRUVATE METABOLISM	20	1	0.436	0.36	1.0	1.0	<a href="#">View</a>
GLYCOLYSIS	21	1	0.458	0.374	1.0	1.0	<a href="#">View</a>
PORPHYRIN METABOLISM	22	1	0.48	0.388	1.0	1.0	<a href="#">View</a>
GLUCONEOGENESIS	27	1	0.589	0.454	1.0	1.0	<a href="#">View</a>
PYRIMIDINE METABOLISM	36	1	0.795	0.556	1.0	1.0	<a href="#">View</a>
BILE ACID BIOSYNTHESIS	49	1	1.07	0.872	1.0	1.0	<a href="#">View</a>

Submit

# The Matched Metabolite Set

Metabolite Set	Total	Hits	Expect	P-value	holm P	FDR	Details
GLYCINE, SERINE AND THREONINE METABOLISM	38	9	0.567	2.74E-10	2.19E-8	2.19E-8	<a href="#">View</a>
PROTEIN BIOSYNTHESIS	38	6	0.415	8.93E-7	7.85E-5	3.57E-5	<a href="#">View</a>
PHENYLALANINE AND TYROSINE METABOLISM	13	5	0.294	3.15E-6	2.49E-4	8.4E-5	<a href="#">View</a>
METHIONINE METABOLISM							<a href="#">View</a>
AMMONIA RECYCLING							<a href="#">View</a>
PROPIONATE METABOLISM							<a href="#">View</a>
CYSTEINE METABOLISM							<a href="#">View</a>
GLUTATHIONE METABOLISM							<a href="#">View</a>
BETAINE METABOLISM							<a href="#">View</a>
ASPARTATE METABOLISM							<a href="#">View</a>
VALINE, LEUCINE AND ISOLEUCINE METABOLISM							<a href="#">View</a>
TYROSINE METABOLISM							<a href="#">View</a>
UREA CYCLE							<a href="#">View</a>
CITRIC ACID CYCLE							<a href="#">View</a>
CATECHOLAMINE METABOLISM							<a href="#">View</a>
ARGININE AND PROLINE METABOLISM							<a href="#">View</a>
ALANINE METABOLISM	5	1	0.131	0.124	1.0	0.555	<a href="#">View</a>
TAURINE AND HYPOTAURINE METABOLISM	7	1	0.153	0.144	1.0	0.638	<a href="#">View</a>
BUTYRATE METABOLISM	2	1	0.196	0.181	1.0	0.716	<a href="#">View</a>

Current metabolite set:

Set Name	Metabolites	References
PHENYLALANINE AND TYROSINE METABOLISM	Ammonia, <b>Acetoacetic acid</b> , Homogentisic acid, <b>Fumaric acid</b> , L-Tyrosine, L- <b>Phenylalanine</b> , <b>Phenylpyruvic acid</b> , 4-Hydroxyphenylpyruvic acid, 4-Fumarylacetoacetic acid, Oxygen, Malylacetoacetic acid, Water, Hydrogen peroxide	<a href="#">SMPDB</a>

OK

Click on SMPDB  
to see more  
information

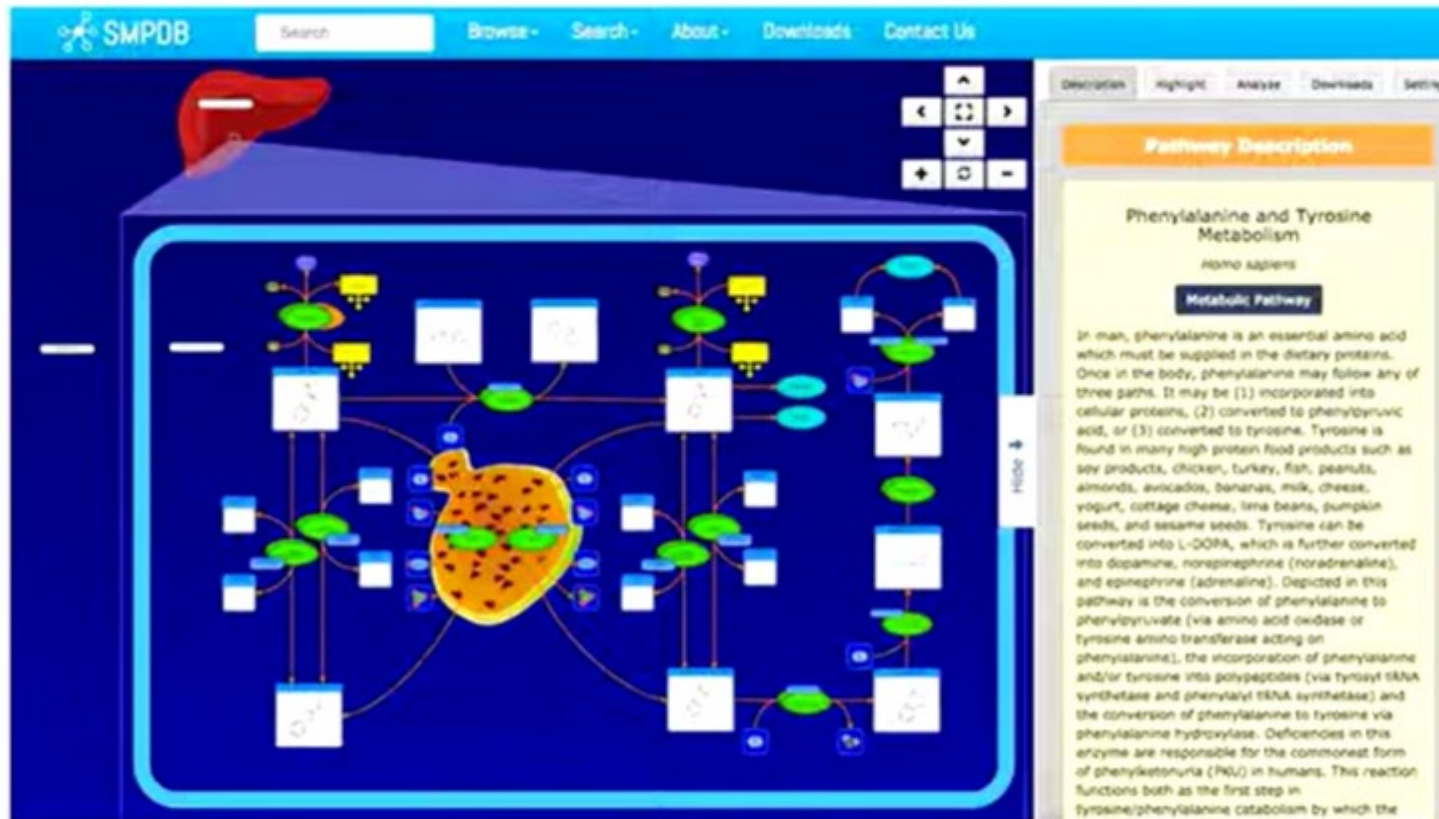


# Phenylalanine and Tyrosine Metabolism in SMPDB

so over-representation analysis is actually a veryweak form of doing pathway analysis

it's just giving you as an exampleit's arguably the easiest one to do because

*all you need to do isprovide a simple list of differentially changed metabolites*



# Single Sample Profiling (SSP)

(Basically used by a physician to analyze a patient)

Choose one of the following options to proceed

- A list of compound names (over representation analysis)
- A list of compounds with concentration values (single sample profiling)

Enter your data below (two-column data):

L-Isoleucine	0.34
Fumaric acid	0.47
Acetone	0.56
Succinic acid	9.4
1-Methylhistidine	9.6
L-Asparagine	19.62
3-Methylhistidine	9.7
L-Threonine	93.19
Creatine	720
DL-Aspartic acid	14.39
L-Tryptophan	35.78
L-Carnitine	16.01
L-Serine	17.32
L-Tyrosine	67.51
L-Alanine	219.02
D-Fucose	20.37
D-Glucose	23.92
Pyroglutamic acid	26.38

Input Type:

Biofluid (unit):

☒ Use the example data  
- urine sample (umol/mol\_creatinine)

• A concentration table (quantitative enrichment analysis)

Single sample profiling is basically what you would do if you were analyzing a single individual and **you're trying to determine whether they're sick or healthy**

You might be taking out a profile of an individual and here instead of just providing a list of a town length to actually providing a list of metabolites

You have to indicate whether it's blood or serum or urine or wherever but

As long as it knows what type of bio fluid is working with and as long as it knows that it's human because this is what it's going to be referencing

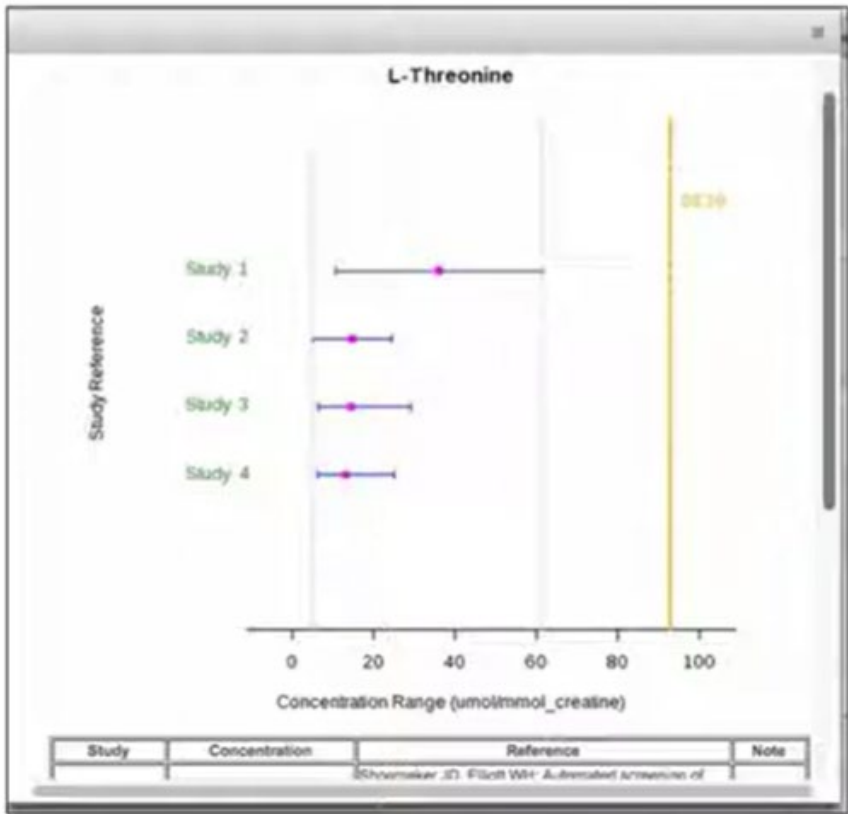
## Concentration Comparison

**Comparison with Reference Concentration**

Note: reference concentrations are in the form of **mean(min - max)** format. In cases where the ranges were not reported in the original literature, the min and max were calculated using the 95% confidence intervals. In the Comparison column, **H**, **M**, **L** means **higher**, **medium** (within range), **lower** compared to the reference concentrations. Click the **Image** icon link to see a graphical summary for the comparisons.

Compound	Concentration	Reference concentrations	Comparison	Detail	Image
L-Ascorbic acid	0.34	3.75 (1.6-5.5); 3.55 (1.7-5.4); 0.02125 (0.0088 - 0.0338); 1.3 (0.5-2.7); 1.3 (0.6-2.6)	M	<a href="#">View</a>	
Exfoliant acid	0.47	0.85 (0.52 - 1.88); 0.4 (0.2 - 0.8); 10.4 (2.8 - 83.7); 0.9 (0.1 - 1.7); 10.7 (0.1 - 28.2); 0.1 (0.1 - 1.7); 0.25 (0.1 - 0.4); 0.7 (0.2 - 1.7)	M	<a href="#">View</a>	
Ascorbate	0.58	2.24 (0 - 6.37); 3.9 (0.9 - 17.8)	M	<a href="#">View</a>	
Succinic acid	9.4	12.6 (0.47 - 24.72); 7.5 (0.5 - 18); 7.7 (1.9 - 20); 157.2 (29.4 - 486.2); 185.4 (0 - 342.6); 11.8 (4 - 27.3); 14.85 (11.28 - 17.88); 8.25 (0.5 - 18); 5.8 (1.2 - 8.4); 9.8 (4.9 - 14.8); 14.4 (8.9 - 19.3); 9.2 (2.5 - 13.5); 4.7 (1.1 - 14.5); 6 (0.3 - 33.2)	M	<a href="#">View</a>	
L-Methylniacid	8.6	5.8 (1.9 - 7.3); 2.3 (0 - 7.4); 46.1 (0 - 89.6); 15.9 (0 - 35.4); 28.1 (0 - 58.9); 1.3 (0 - 4.04); 45.5 (3.9 - 87.1); 30.8 (0 - 70); 15.9 (0 - 35.4); 30 (0 - 73); 0.00288 (0.0019 - 0.0038); 8.3 (2.4 - 28.4)	M	<a href="#">View</a>	
L-Ascorbate	18.62	0.98 (0.31 - 1.87); 10.52 (6.87 - 14.37); 10 (4.8 - 16.32); 10.585 (4.85 - 16.55); 8.8 (4.8 - 12.7); 9.3 (0 - 26); 10.1 (4.8 - 17.8)	M	<a href="#">View</a>	
3-Methylniacid	9.7	42.78 (19.82 - 85.5); 12.5 (8.3 - 18.7); 0.0148 (0.0012 - 0.0286); 16.5 (2.8 - 58.8)	M	<a href="#">View</a>	
L-Threonine	93.18	36.2 (10.82 - 61.58); 14.88 (5.17 - 24.59); 14.6 (8.6 - 20.3); 13.3 (6.4 - 25.2)	H	<a href="#">View</a>	
Ornithine	720	113 (0 - 854); 113 (0 - 854); 46 (0 - 448)	H	<a href="#">View</a>	
DL-Ascorbic acid	14.39	19 (2.7 - 44); 67.9 (14.3 - 100.7); 73.8 (84 - 130.3); 37.9 (17.3 - 83.3); 29.8 (14.7 - 85.1); 54.5 (32.4 - 76.4); 15.3 (5.2 - 16.3); 20.9 (3.8 - 95.3)	M	<a href="#">View</a>	
L-Tyrosine	35.78	13.52 (5.15 - 20.89); 5.6 (0.3 - 2.1); 6.3 (0.4 - 11.1)	H	<a href="#">View</a>	
L-Cysteine	18.01	4.5 (0.82 - 16.2); 9 (0.7 - 16.4)	M	<a href="#">View</a>	

# Concentration Comparison (cont.)



it will do essentially a concentration comparison so for each of the compounds that are in there it will provide a list of from the known or relative concentration ranges that are expected to be normal and how far whether things are significantly off and abnormal or or normal you can then further view these things and

where we're looking at Concentration Comparison (cont.) levels for threonine

in this particular individual they're read out in urine with something like 90 micro molar and the information that's been collected on humans in urine over last number of years there's four studies that were published all the values are typically reported or somewhere between 20 and 40 micro molar

so this individual clearly has abnormally high values

# Quantitative Enrichment Analysis (QEA)

now you're looking at a population with a whole bunch of metabolite concentrations and in this case we'll be indicating names and concentrations this case got a sample data step you can upload

Choose one of the following options to proceed

- A list of compound names (over representation analysis)
- A list of compounds with concentration values (single sample profiling)
- A concentration table (quantitative enrichment analysis)

Upload your concentration data (.csv or .txt)

Group Label: ☒ Discrete (Classification) ☐ Continuous (Regression)

ID Type:

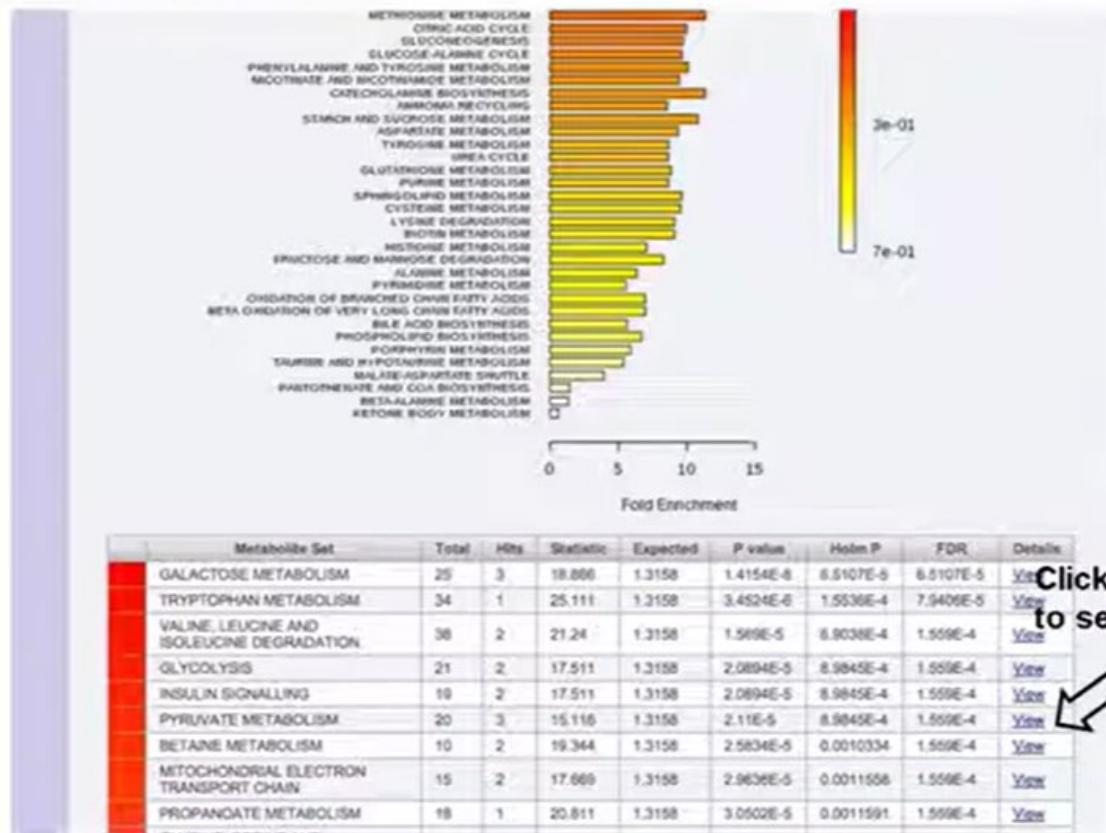
Data File:  No file chosen

---

Try our test data:

Data	ID Type	Group Label	Description
<a href="#">Data 1</a>	Common name	Discrete	Urinary metabolite concentrations from 77 cancer patients measured by 1H NMR. Phenotype: N - cachectic; Y - control
<a href="#">Data 2</a>	PubChem CID	Continuous	Urinary metabolite concentrations from 97 cancer patients measured by 1H NMR. Phenotype: muscle gain (percentage within 100 days, negative values indicate muscle loss)

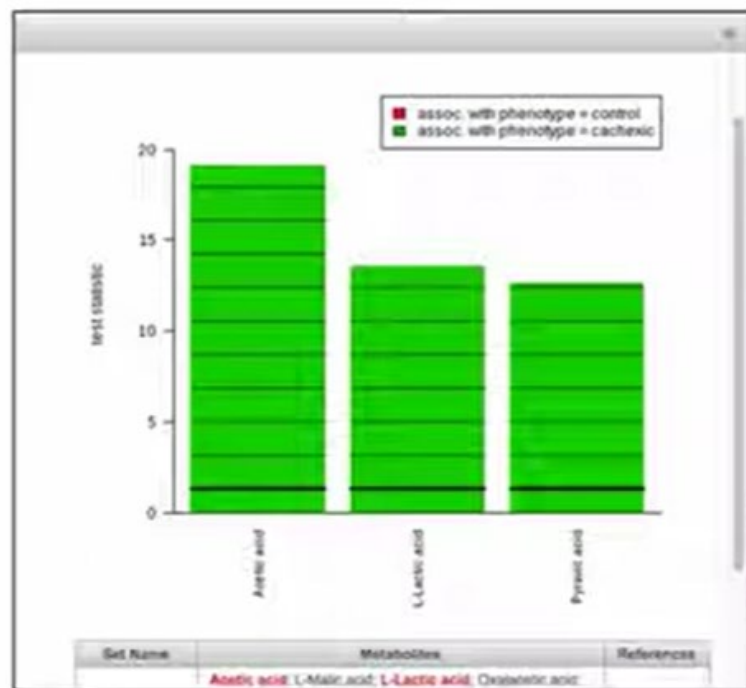
# Result



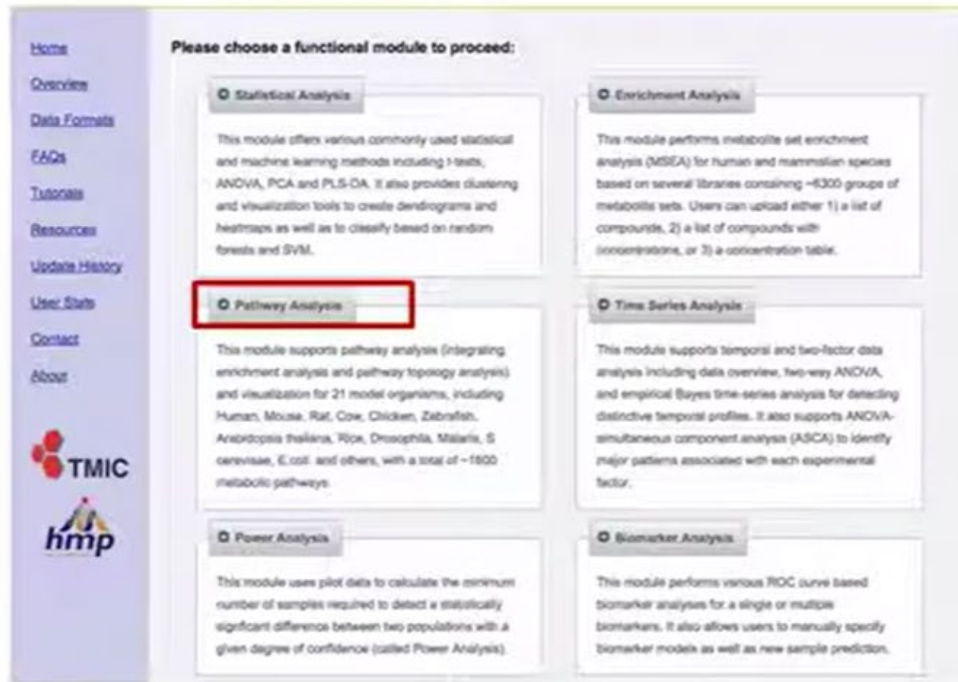
Click on details  
to see more



# The Matched Metabolite Set



# Select a Module (Pathway Analysis)



While enrichment analysis tells you which diseases pathways other things are modified

this one is largely more focused on pathways but

it looks not just at the Association of metabolites in pathways but also

It considers the pathway structures whether metabolites that are being changed or representative of hubs or spokes in the pathway whether they play a central role and

rather than just being restricted only to humans as the over-representation or MSC is this allows you to look at things from 21 model organisms

# Pathway Analysis

- **Purpose: to extend and enhance metabolite set enrichment analysis for pathways by**
  - Considering pathway structures
  - Supporting pathway visualization
- **Currently supports analysis for 21 diverse (model) organisms such as humans, mouse, Drosophila, Arabidopsis, E. coli, yeast, etc. (KEGG pathways only)**

While enrichment analysis tells you which diseases pathways other things are are modified

this one is largely more focused on pathways but

it looks not just at the Association of metabolites in pathways but also

It considers the pathway structures whether metabolites that are being changed or representative of hubs or spokes in the pathway whether they play a central role and

rather than just being restricted only to humans as the over-representation or MSC is this allows you to look at things from 21 model organisms

## Data Set Selected

- Here we are using a collection of metabolites identified by NMR (compound list + concentrations) from the urine from 77 lung and colon cancer patients, some of whom were suffering from cachexia (muscle wasting)

1

Home

Processing

Normalization

Pathway

Download

Exit

Please enter a one-column compound list:

Input Type:

Or upload a concentration table (.csv or .txt):

Group Label: ☐ Discrete (Classification) ☐ Continuous (Regression)

ID Type:

Data File:

☒ Use the example data

Data	Description
<a href="#">Download</a>	Urinary metabolite concentrations from 77 cancer patients measured by <sup>1</sup> H NMR. Phenotype: N = cachectic; Y = control

# Perform Data Normalization

**Data Normalization:**

The normalization procedures are grouped into three categories. The sample normalization allows general purpose adjustment for differences among samples; data transformation and scaling are two different approaches to make features more comparable. You can use one or combine them to achieve better results.

**Sample normalization**

☒ None

☐ Sample specific normalization (i.e. dry weight, volume) [Click here to specify](#)

☐ Normalization by sum

☐ Normalization by median

☐ Normalization by reference sample

☒ Specify a reference sample

☐ Create a pooled average sample from group

☐ Normalization by reference feature

**Data transformation**

☒ None

☐ Log transformation (generalized logarithm transformation or glog)

☐ Cube root transformation (take cube root of data values)

**Data scaling**

☐ None

☒ Auto scaling (mean-centered and divided by the standard deviation of each variable)

☐ Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)

☐ Range scaling (mean-centered and divided by the range of each variable)



# Select Pathway Libraries

Please select a pathway library:

☒ Homo sapiens (human) [80]  
☐ Mus musculus (mouse) [82]  
☐ Rattus norvegicus (rat) [81]  
☐ Bos taurus (cow) [81]  
  
☐ Gallus gallus (chicken) [78]  
  
☐ Danio rerio (zebrafish) [81]  
  
☐ Drosophila melanogaster (fruit fly) [79]  
  
☐ Caenorhabditis elegans (nematode) [78]  
  
☐ Saccharomyces cerevisiae (yeast) [80]  
  
☐ Oryza sativa japonica (Japanese rice) [83]  
☐ Arabidopsis thaliana (thale cress) [87]  
  
☐ Schistosoma mansoni [69]  
☐ Plasmodium falciparum 3D7 (Malaria) [47]  
☐ Trypanosoma brucei [54]  
  
☐ Escherichia coli K-12 MG1655 [87]  
☐ Bacillus subtilis [80]  
☐ Pseudomonas putida KT2440 [89]  
☐ Staphylococcus aureus N315 (MRSA/VISA) [73]  
☐ Thermotoga maritima [57]  
☐ Synechococcus elongatus PCC7942 [75]  
☐ Mesorhizobium loti [86]

# Perform Network Topology Analysis

Please specify a reference metabolome:

- ☒ Use all compounds in the selected pathways
- ☐ [Upload a reference metabolome based on your technical platform](#)

Please specify pathway analysis algorithms:

Pathway Enrichment Analysis

- ☒ Global Test
- ☐ Global Ancova

Pathway Topology Analysis

- ☒ Relative-betweenness Centrality
- ☐ Out-degree Centrality

Submit

# Pathway Position Matters

Which positions are important?

- ▶ Hubs

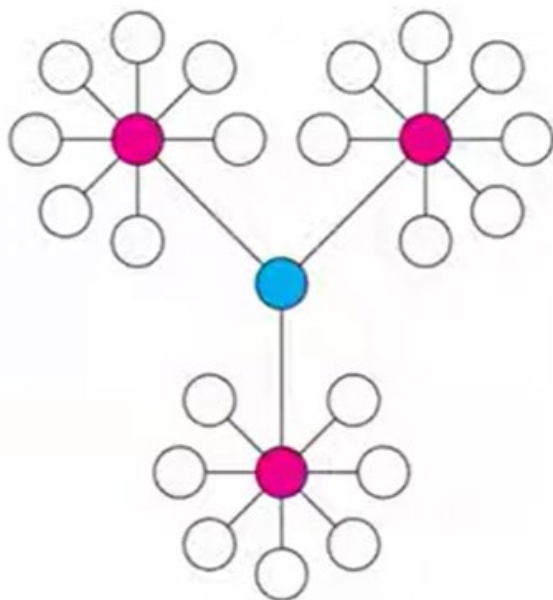
- ▶ Nodes that are highly connected (red ones)

- ▶ Bottlenecks

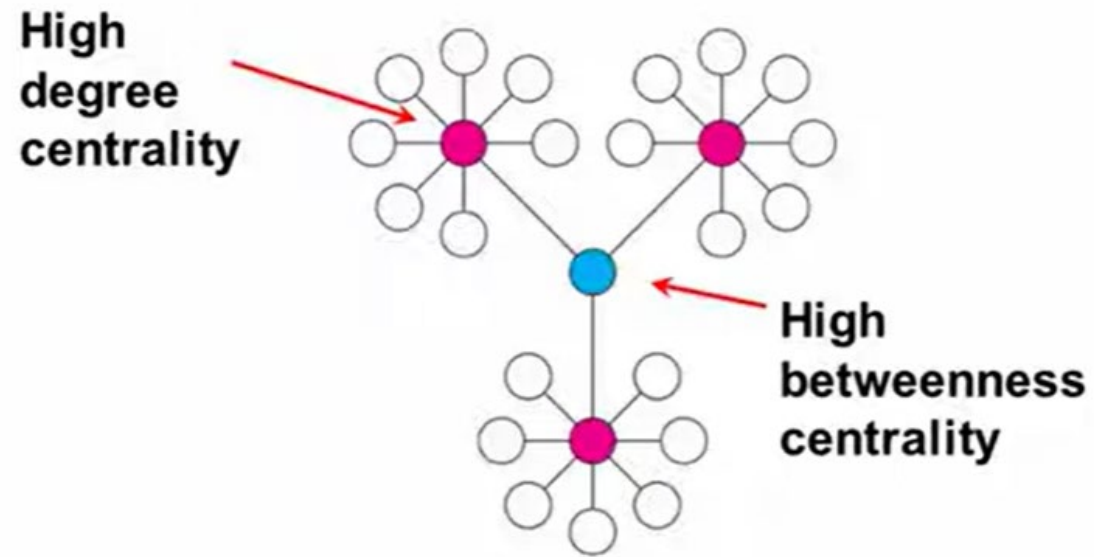
- ▶ Nodes on many shortest paths between other nodes (blue ones)

Graph theory

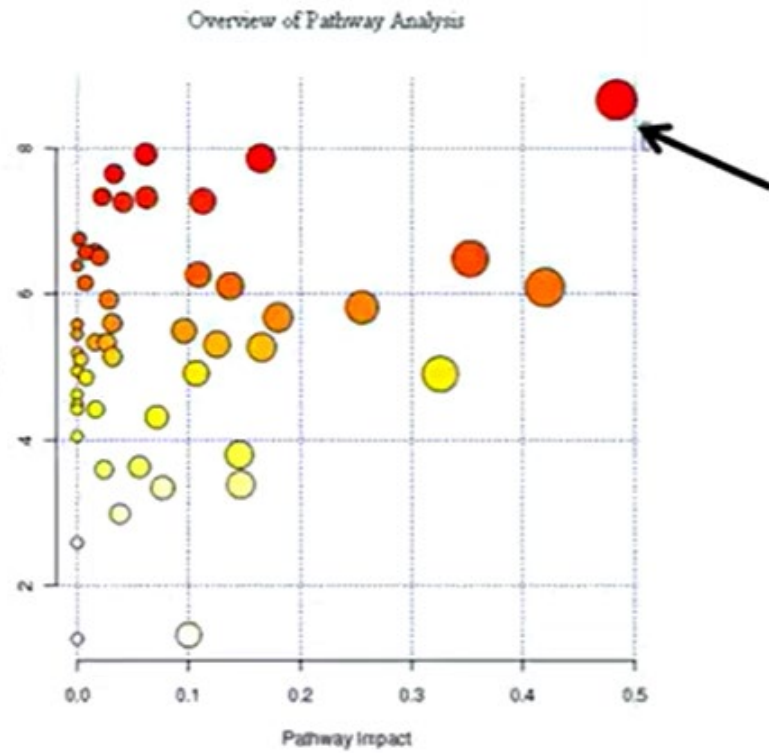
- ▶ Degree centrality
- ▶ Betweenness centrality



# Which Node is More Important?

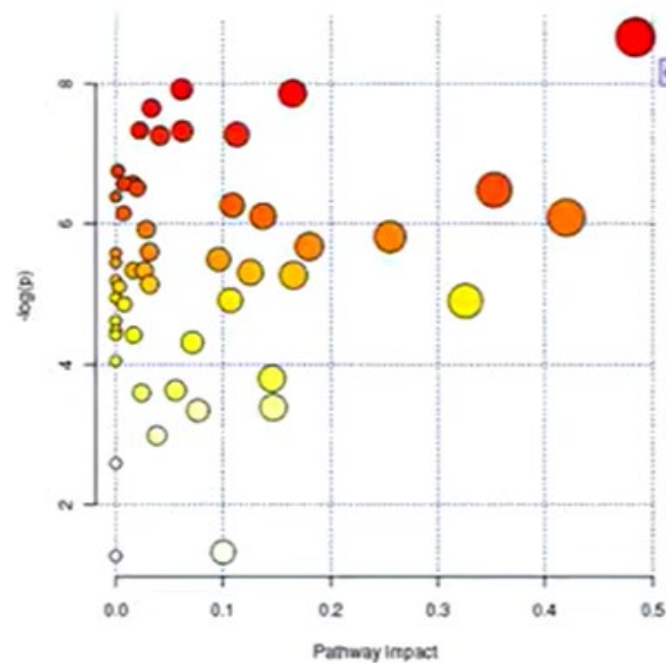


# Pathway Visualization

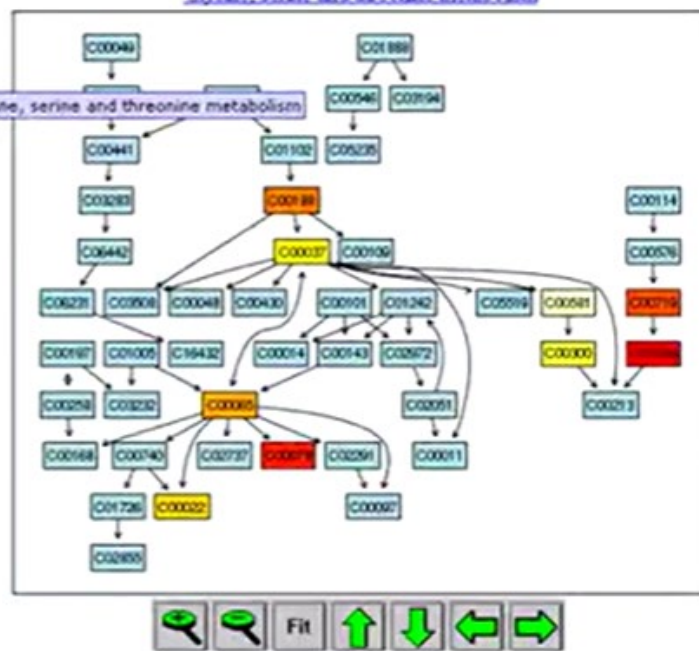


# Pathway Visualization

Overview of Pathway Analysis



Glycine, serine and threonine metabolism





# Pathway Impact

- Incorporates parameters such as the log fold-change of the DE metabolites, the statistical significance of the set of pathway genes and the topology of the signaling pathway
- Combines the pathway topology with the over-representation evidence

# Result

Pathway Name	Total	Hits	p	-log(p)	Holm p	FDR	Impact	Details
<a href="#">Valine, leucine and isoleucine degradation</a>	40	2	1.1954E-4	9.0319	0.0059769	0.0031356	0.02232	<a href="#">KEGG SMP</a>
<a href="#">Valine, leucine and isoleucine biosynthesis</a>	27	4	1.2542E-4	8.9838	0.0061458	0.0031356	0.04823	<a href="#">KEGG SMP</a>
<a href="#">Glycine, serine and threonine metabolism</a>	48	8	2.4586E-4	8.3107	0.011801	0.0040977	0.48394	<a href="#">KEGG SMP</a>
<a href="#">Methane metabolism</a>	34	6	3.8485E-4	7.6626	0.018088	0.0043833	0.16466	<a href="#">KEGG</a>
<a href="#">Sulfur metabolism</a>	18	2	4.755E-4	7.6512	0.021873	0.0043833	0.03307	<a href="#">KEGG SMP</a>
<a href="#">Arginine and proline metabolism</a>	77	6	6.578E-4	7.3266	0.029601	0.0043833	0.06203	<a href="#">KEGG SMP</a>
<a href="#">Aminoacyl-tRNA biosynthesis</a>	75	10	6.6275E-4	7.3191	0.029601	0.0043833	0.11268	<a href="#">KEGG</a>
<a href="#">Nicotinate and nicotinamide metabolism</a>	44	5	7.0133E-4	7.2625	0.030157	0.0043833	0.04113	<a href="#">KEGG SMP</a>
<a href="#">Glutathione metabolism</a>	38	2	0.0011587	6.7605	0.048664	0.0063514	0.0019	<a href="#">KEGG SMP</a>
<a href="#">Propanoate metabolism</a>	35	4	0.0013834	6.576	0.057129	0.0063514	0.01603	<a href="#">KEGG SMP</a>
<a href="#">Galactose metabolism</a>	41	3	0.001486	6.5118	0.059441	0.0063514	0.01982	<a href="#">KEGG SMP</a>
<a href="#">Taurine and hypotaurine metabolism</a>	20	3	0.0015243	6.4862	0.059449	0.0063514	0.35252	<a href="#">KEGG SMP</a>
<a href="#">Cyanoamino acid metabolism</a>	16	4	0.0016826	6.3874	0.06394	0.0064716	0.0	<a href="#">KEGG</a>
<a href="#">Nitrogen metabolism</a>	39	7	0.0021434	6.1454	0.079305	0.0070701	0.00763	<a href="#">KEGG SMP</a>
<a href="#">Inositol phosphate metabolism</a>	39	1	0.002215	6.1125	0.079741	0.0070701	0.13703	<a href="#">KEGG SMP</a>
<a href="#">Pyruvate metabolism</a>	32	4	0.0022624	6.0913	0.079741	0.0070701	0.41957	<a href="#">KEGG SMP</a>
<a href="#">Cysteine and methionine metabolism</a>	56	2	0.0026796	5.9221	0.091106	0.0078811	0.02846	<a href="#">KEGG SMP SMP</a>
<a href="#">Alanine, aspartate and glutamate metabolism</a>	24	6	0.0029727	5.8183	0.0981	0.0082576	0.25548	<a href="#">KEGG SMP SMP SMP</a>
<a href="#">Pantothenate and CoA biosynthesis</a>	27	4	0.0034143	5.6798	0.10926	0.0089486	0.18014	<a href="#">KEGG SMP</a>
<a href="#">Phenylalanine metabolism</a>	45	6	0.0036864	5.6026	0.11434	0.0089486	0.0315	<a href="#">KEGG SMP</a>

Submit

