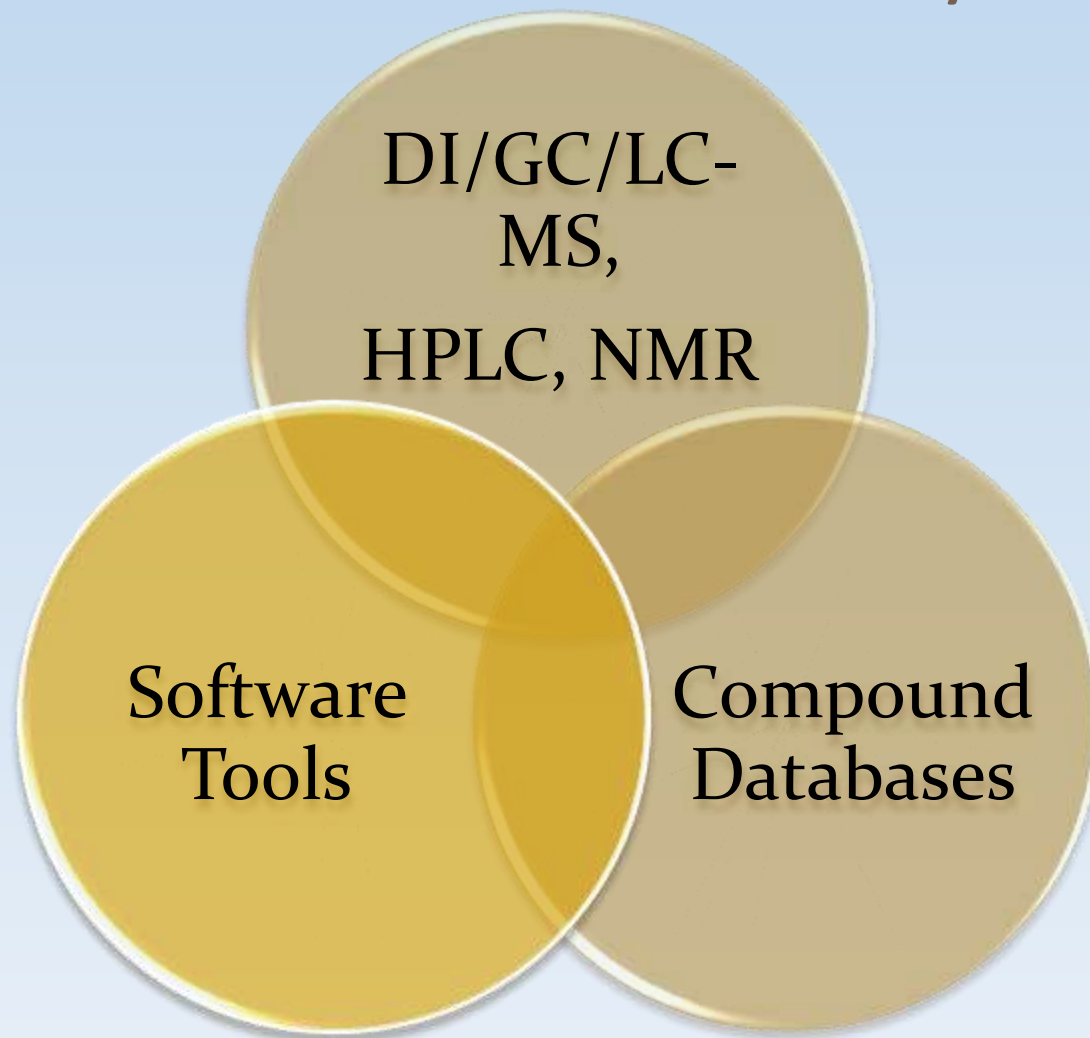


# Use of web-based tools for quantitative metabolomics

Jianguo (Jeff) Xia  
Wishart Research Group  
University of Alberta, Canada

# Metabolomics in the University of Alberta



# Outline

- Introduction
  - Omics data overview
  - Lessons from other omics research
  - Project goals
- Web-based metabolomics tools
  - I. General data processing & statistical analysis
    - I. MetaboAnalyst (<http://www.metaboanalyst.ca>)
  - II. Identify functionally interesting patterns
    - I. MSEA (<http://www.msea.ca>)
  - III. Metabolic Pathway Analysis
    - I. MetPA (<http://metpa.metabolomics.ca>)
- Public databases
- Summary

# The '-omics' data overview

Genomics	DNA sequence	100,000 - 1,000,000
Transcriptomics	Gene expression	10,000 - 100,000
Proteomics	Protein expression/ interaction	1,000 - 10,000
Metabolomics	Compound concentration	100 - 1,000

# Common questions

1. Are there some **interesting patterns** present in my data?
2. What are the most **important features** associated with different phenotypes?
3. Is there a **real difference** between the groups?
4. Can I use this data to **predict** a phenotype?
5. How to **interpret** these features / patterns?
6. How does my result **compared with published data**?

# Common approaches

1 <sup>st</sup>	Classical statistics	T-tests, ANOVA	Since 1950s
2 <sup>nd</sup>	High-dimensional feature selection; Machine learning	SAM, Limma; SVM, Neural networks	Since 1990s
3 <sup>rd</sup>	Group-based enrichment analysis	GSEA, GSA, Globaltest	Since 2003
4 <sup>th</sup>	Pathway Analysis	SPIA, TopoGSA	Since 2007

# Project Goals

- Provide well-established methods proven highly successful in other 'omics' studies;
  - Do not re-invent the wheel!
- Support traditional approaches
  - Cheminformatics approaches
  - Data processing & normalization procedures
- Easy-to-use
  - Not command-line
  - Target users – bench biologists

# Identify influential algorithms

[\[HTML\] Cluster \*\*analysis\*\* and display of genome-wide expression patterns](#)

MB Eisen, PT Spellman, PO Brown, ... - Proceedings of the ... , 1998 - National Acad Sciences

... **Microarray**-based genomic surveys and other high-throughput approaches (ranging from genomics to ... with the addition of uncharacterized genes (the results of this **analysis** will be ... Finally, the functional concordance of coexpressed genes imparts biological **significance** to the ...

[Cited by 9813](#) - [Related articles](#) - [BL Direct](#) - [All 184 versions](#)

## [Significance analysis of microarrays](#)

VG Tusher, R Tibshirani, G Chu - US Patent 7,363,165, 2008 - Google Patents

US007363165B2 (12) United States Patent Tusher et al. (io) Patent No.: US 7,363,165 B2

(45) Date of Patent: Apr. 22, 2008 (54) **SIGNIFICANCE ANALYSIS OF MICROARRAYS** (75)

Inventors: Virginia Goss Tusher, Palo Alto, CA (US); Robert Tibshirani, Palo Alto, CA (US); ...

[Cited by 5558](#) - [Related articles](#) - [BL Direct](#) - [All 81 versions](#)

## [Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles](#)

... , P Tamayo, VK Mootha, S Mukherjee, ... - Proceedings of the ... , 2005 - National Acad Sciences

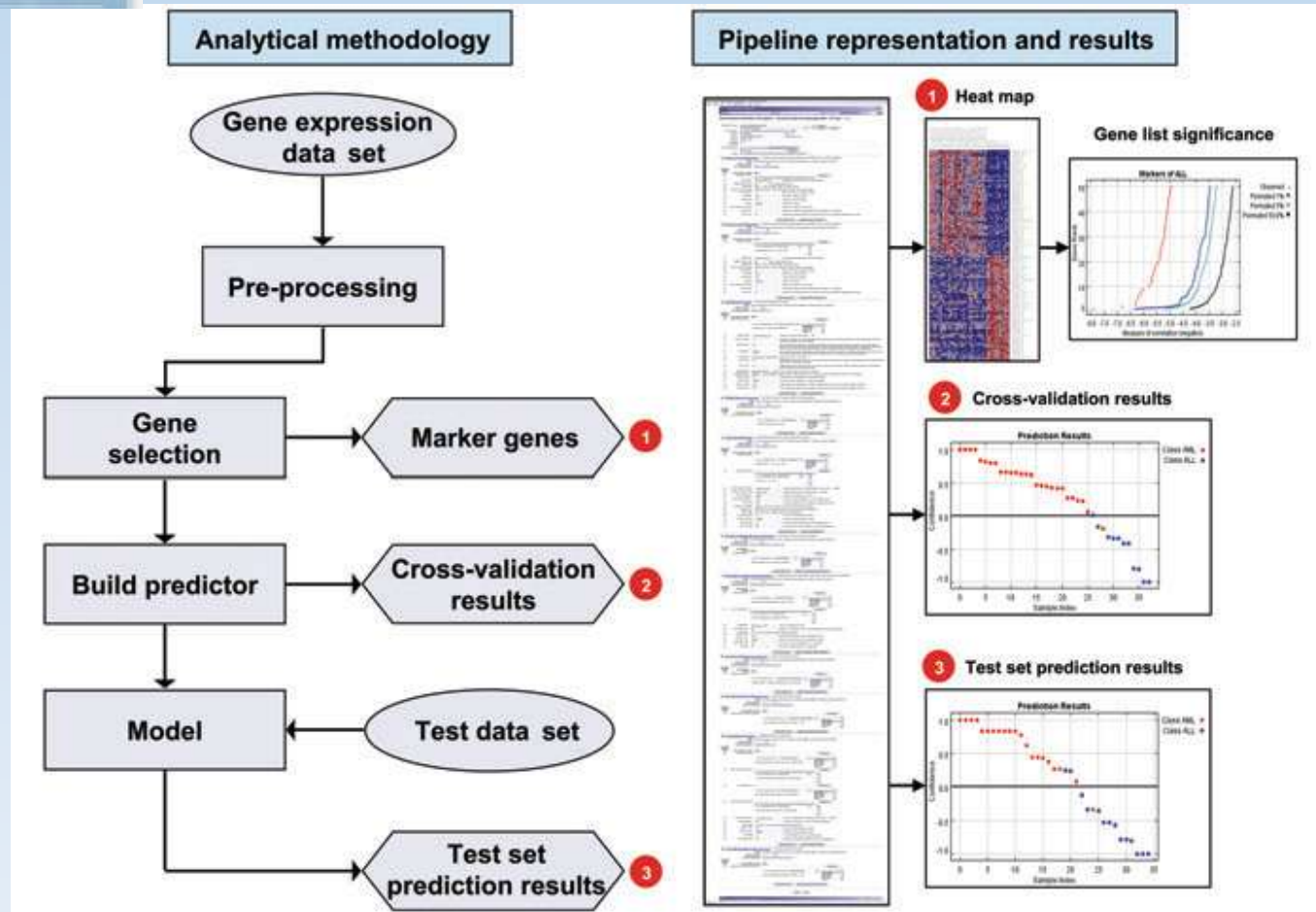
Although genomewide RNA expression **analysis** has become a routine tool in biomedical research, extracting biological insight from such information remains a major challenge.

Here, we describe a powerful analytical method called **Gene Set Enrichment Analysis** ( ...

[Cited by 1738](#) - [Related articles](#) - [BL Direct](#) - [All 38 versions](#)



# Identify the best practices



*Nature Genetics* - 38, 500 - 501 (2006)

# Metabolomics web applications

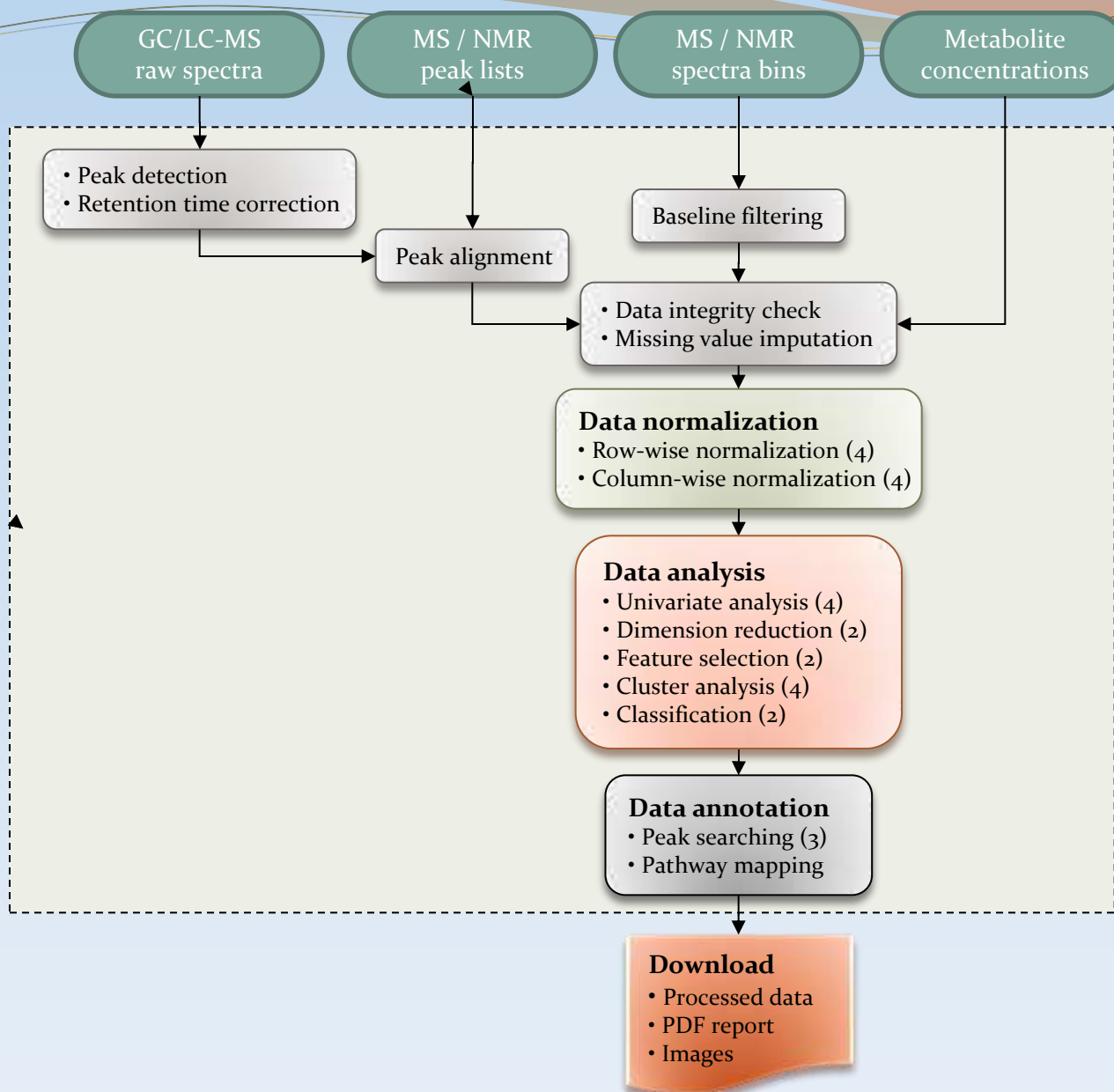
- General data processing & analysis
  - MetaboAnalyst
  - <http://www.metaboanalyst.ca>
- Metabolite Set Enrichment Analysis
  - MSEA
  - <http://www.msea.ca>
- Metabolomic Pathway Analysis
  - MetPA
  - <http://metpa.metabolomics.ca>

# MetaboAnalyst

- <http://www.metaboanalyst.ca>
- General metabolomics data processing, normalization, and statistical analysis
  - Support two-group and multi-group analysis
  - 20+ well-established methods
  - Dynamic graphical presentation
  - Automatic report generation

# What MetaboAnalyst can:

- Basic data processing:
  - Peak picking, Peak alignment, Baseline filtering, etc.
- Data normalization
  - probabilistic quotient normalization, scaling, etc.
- Data overview
  - PCA, Heatmaps, etc.
- Identify important features
  - t-tests, ANOVA, SAM, etc.
- Classification
  - PLS-DA, random Forest, SVM, etc.



# MetaboAnalyst



**Welcome** ([click here to start](#))

[Home](#)

[Overview](#)

[Data Formats](#)

[Tutorials &  
FAQs](#)

[Resources](#)

## Release Notes and Updates:

- Please upgrade your browser if this page does not display properly.
- Bug fix: Color inconsistencies b/w the confidence ellipses and sample class labels (used in PCA and PLS-DA 2D plot)(06/17/10) ;
- Bug fix: Updated the interface for zip file upload to support multiple-group analysis of peak lists and spectra data (06/15/10) ;
- Introducing **Data Editor** to enable samples/features exclusion (i.e. outliers) during analysis (06/14/10); **NEW**
- Added **ANOVA** and associated post-hoc analyses for multi-group data (06/10/10); **NEW**
- MetaboAnalyst now supports data analysis for **more than two groups**. (06/01/10); **NEW**
- For data collected from human or other mammalian species, you may also want to visit our new web application **MSEA** (<http://www.msea.ca>) for more advanced data analysis. (05/11/10);

[Read more ...](#)

**Please Cite:**

# Data Upload

[Home](#)

**Steps**

- [1. Upload](#)
- [2. Process](#)
- [\\*\\*Data Editor](#)
- [3. Normalize](#)
- [4. Analyze](#)
  - [\\_Univariate](#)
  - [\\_ANOVA](#)
  - [\\_PCA](#)
  - [\\_PLSDA](#)
  - [\\_SAM](#)
  - [\\_EBAM](#)
  - [\\_Tree & heatmap](#)
  - [\\_Kmean & SOM](#)
  - [\\_RandomForest](#)
  - [\\_R-SVM](#)
- [5. Peak Search](#)
- [6. Pathway Mapping](#)
- [7. Download](#)
- [8. Log Out](#)

Your home directory is now set up. You can choose either 1) [Upload your data](#) or 2) [Try our test data](#) in order to proceed. Please note, the uploaded data and analysis result will remain in the server for 72 hours before being deleted automatically.

## 1) Upload your data ([Data Format](#))

**Comma Separated Values (.csv) :**

**Data type :**  Concentrations  Spectral bins  Peak intensity table

**Format :**

**Data file :**

**Zipped**  

**Data type :**

**Data :**

**Pairs :**   (required for paired comparison)

**Tip:** You should create a **separate folder** for each group, and upload them as a **single .zip file**.

# Data processing and integrity check

[Home](#)

**Steps**

- 1. Upload
- 2. Process**
- \*\*Data Editor**
- 3. Normalize
- 4. Analyze
  - Univariate
  - ANOVA
  - PCA
  - PLSDA
  - SAM
  - EBAM
  - Tree & heatmap
  - Kmean & SOM
  - RandomForest
  - R-SVM
- 5. Peak Search
- 6. Pathway Mapping
- 7. Download
- 8. Log Out

## Data Integrity Check

**Details:**

1. The class labels must contain only two groups.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. Compound concentration or peak intensity values cannot be negative.

**Data processing information**

Checking data content ...passed

Two groups were detected based on the sample labels.

Samples are not paired.

All data values are numeric.

All data values are non-negative.

A total of 0 , ( 0 %) zero values were detected

A total of 0 , ( 0 %) missing values were detected

By default, these values will be replaced by a small value

Click **Skip** button if you accept the default practice

Or click **Missing value imputation** to use other methods



# Data normalization (1)

**Row-wise normalization**

None

Normalization by sum

Normalization by a reference sample

Exclude the reference sample after normalization

Normalization by a reference feature

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

**Column-wise normalization**

None

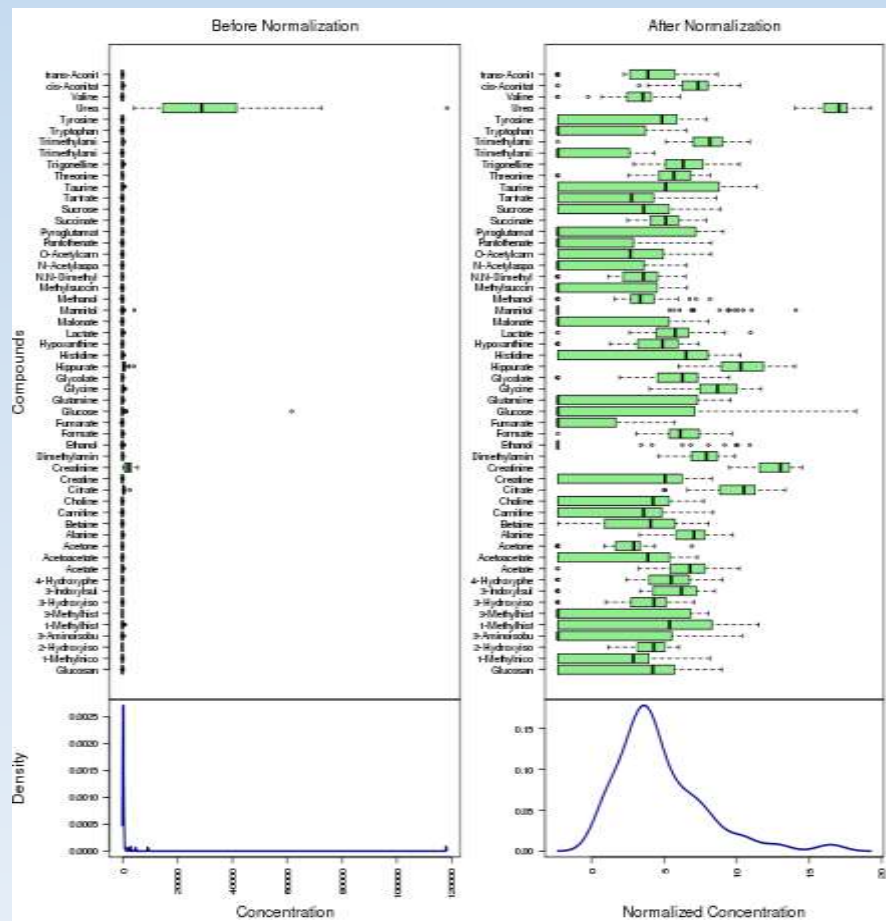
Log (log<sub>2</sub> transformation)

Autoscaling (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)

# Data normalization (2)



# Data Analysis

[Home](#)

**Steps**

- [1. Upload](#)
- [2. Process](#)
- [\\*\\*Data Editor](#)
- [3. Normalize](#)
- [4. Analyze](#)
  - [\\_Univariate](#)
  - [\\_ANOVA](#)
  - [\\_PCA](#)
  - [\\_PLSDA](#)
  - [\\_SAM](#)
  - [\\_EBAM](#)
  - [\\_Tree & heatmap](#)
  - [\\_Kmean & SOM](#)
  - [\\_RandomForest](#)
  - [\\_R-SVM](#)
- [5. Peak Search](#)
- [6. Pathway Mapping](#)
- [7. Download](#)
- [8. Log Out](#)

## **Select an analysis path to explore :**

### **Univariate Analysis**

[Fold Change Analysis, t-Tests, and Volcano plot](#) **(two-group only)**

[One-way Analysis of Variance \(ANOVA\)](#)

### **Chemometrics**

[Principal Component Analysis \(PCA\)](#)

[Partial-Least Square - Discriminant Analysis \(PLS-DA\)](#)

(permutation is only available for two-group data)

### **Significant Feature Identification**

[Significance Analysis of Microarray \(and Metabolites\) \(SAM\)](#)

[Empirical Bayesian Analysis of Microarray \(and Metabolites\) \(EBAM\)](#) **(two-group only)**

### **Cluster Analysis**

[Hierarchical Clustering - Dendrogram and Heatmap](#)

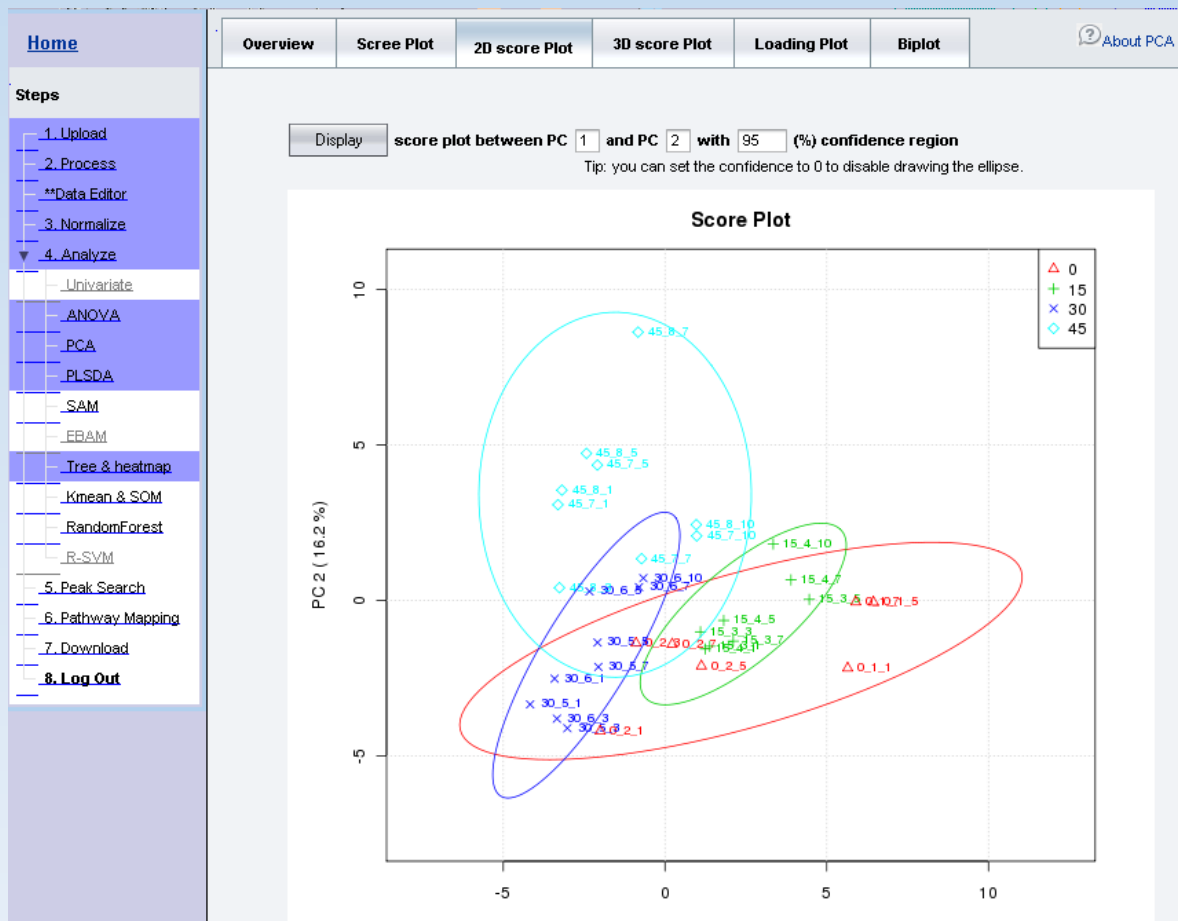
[Partitional Clustering - K-Means and Self Organizing Map \(SOM\)](#)

### **Classification & Feature Selection**

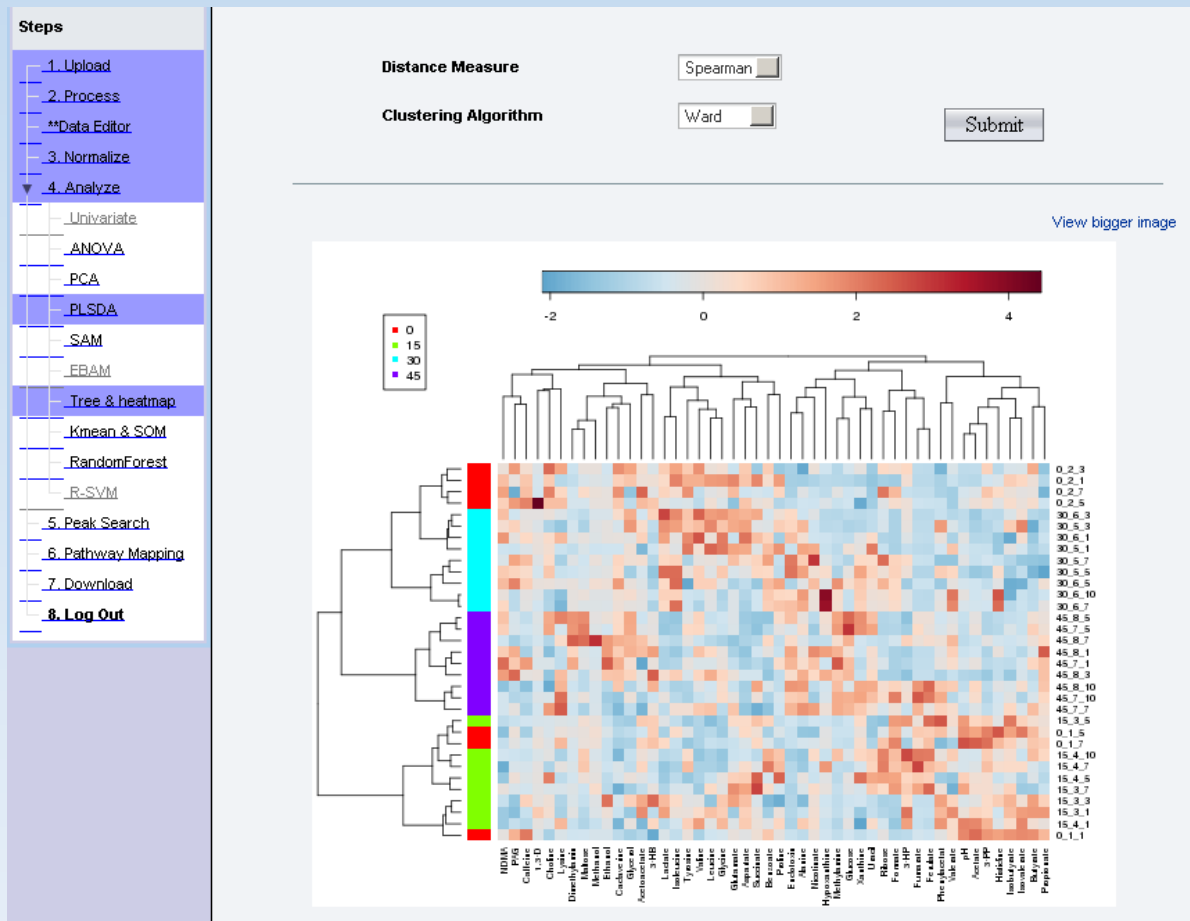
[Random Forest](#)

[Support Vector Machine \(SVM\)](#) **(two-group only)**

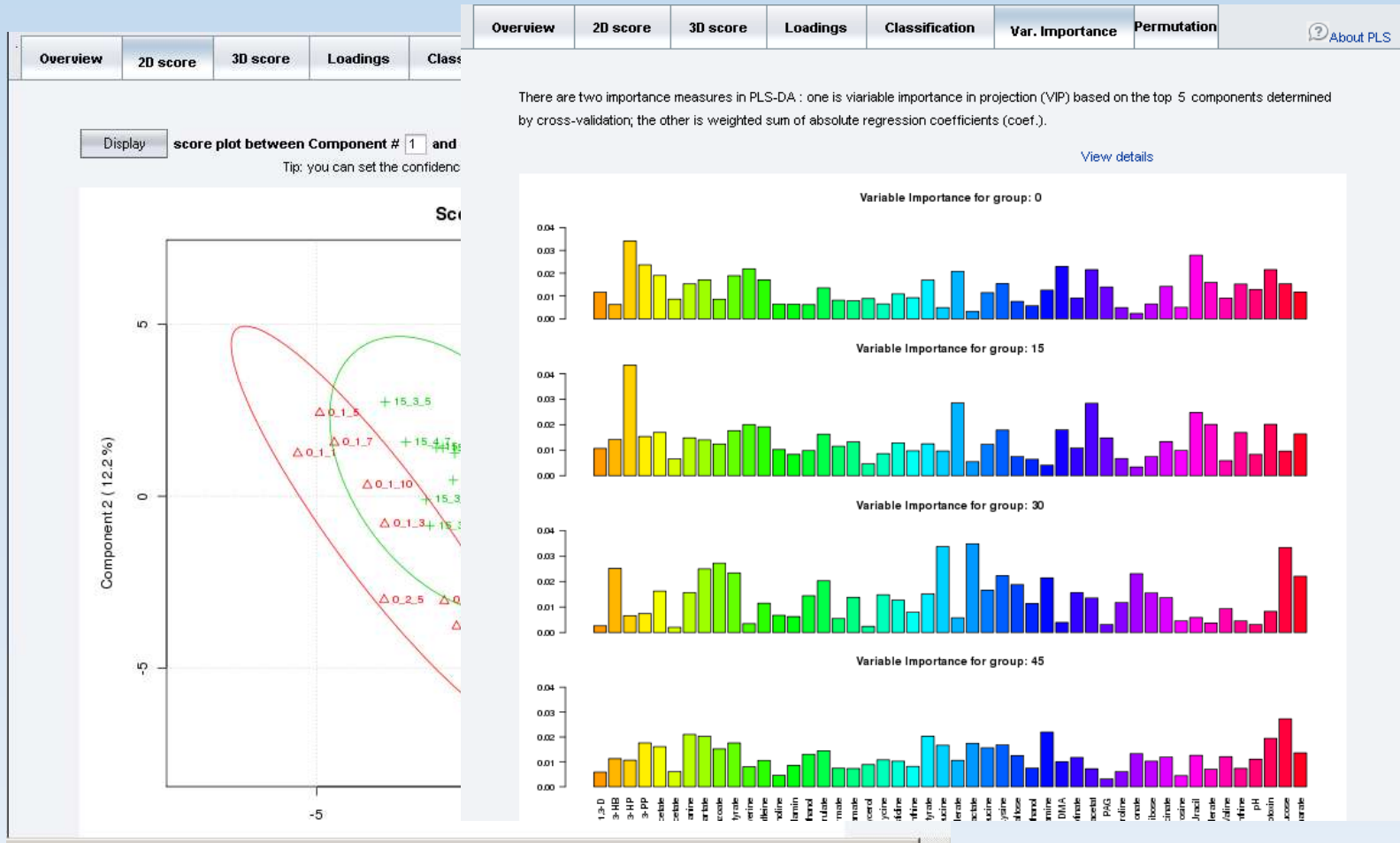
# Clustering with PCA



# Hierarchical clustering



# Supervised approach – PLS-DA



# Feature selection - ANOVA

**Home**

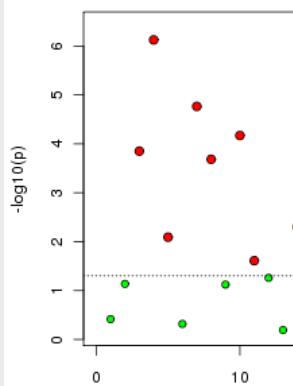
**Steps**

1. Upload
2. Process
- \*\*Data Editor
3. Normalize
- ▼ 4. Analyze
  - \_Univariate
  - **\_ANOVA**
  - \_PCA
  - \_PLSDA
  - \_SAM
  - \_FBAM
  - \_Tree & heatmap
  - \_Kmean & SOM
  - \_RandomForest
  - \_R-SVM
5. Peak Search
6. Pathway Mapping
7. Download
8. Log Out

## One-way Analysis of Variance (ANOVA)

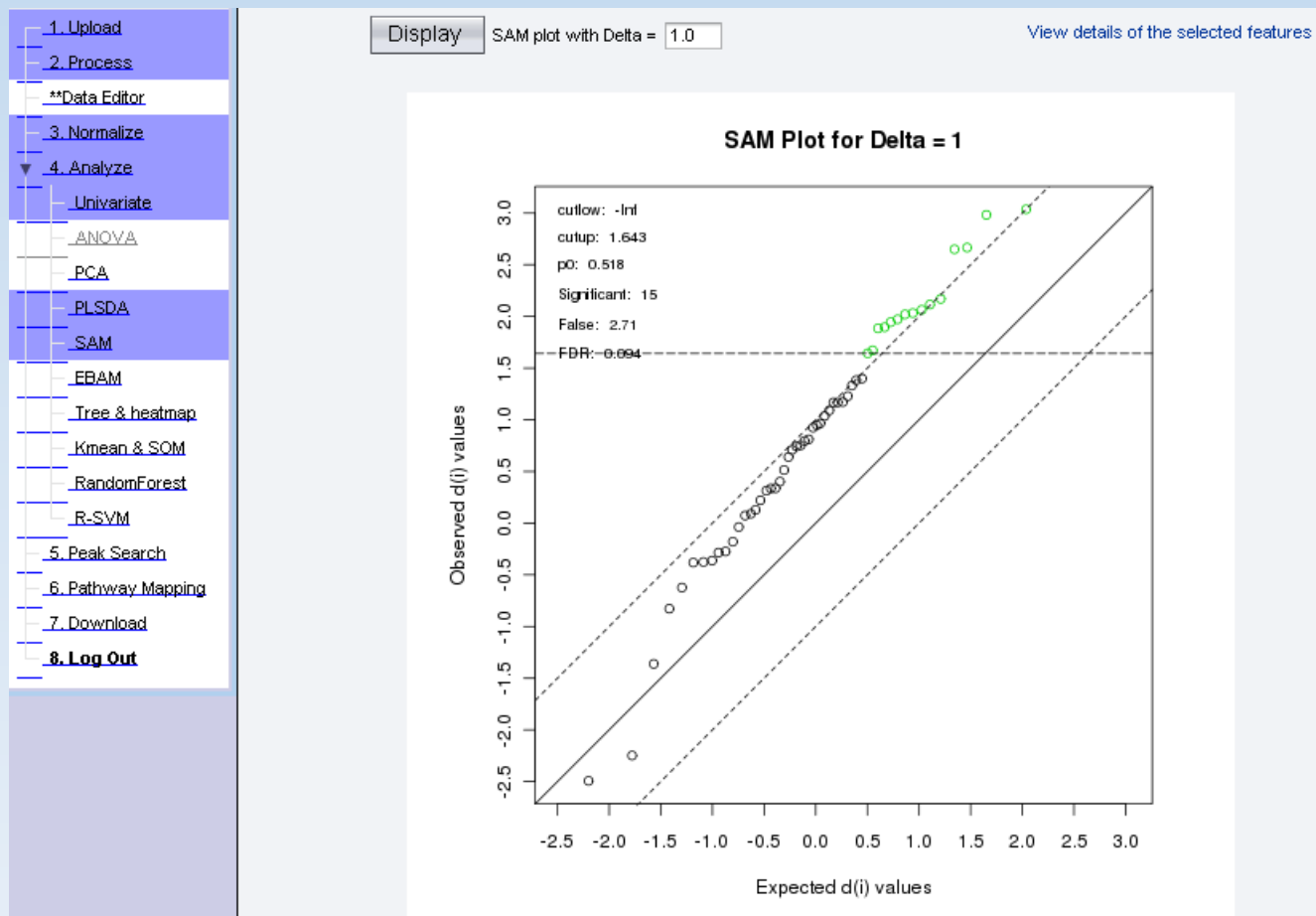
Significance Level (alpha) :

Post-hoc Analysis



Compounds	p.value	$-\log_{10}(p)$	Post-hoc (Fisher's LSD)
Methylamine	0.0	6.44521	45 - 0; 45 - 15; 45 - 30
Endotoxin	0.0	6.17744	30 - 0; 45 - 0; 30 - 15; 45 - 15
3-PP	0.0	6.12539	0 - 30; 0 - 45; 15 - 30; 15 - 45
Glucose	0.0	5.96225	45 - 0; 45 - 15; 45 - 30
Alanine	2.0E-5	4.76571	30 - 0; 45 - 0; 30 - 15; 45 - 15
Butyrate	7.0E-5	4.17052	0 - 30; 15 - 30; 45 - 30
Isoleucine	7.0E-5	4.13585	30 - 0; 0 - 45; 30 - 15; 30 - 45
3-HP	1.4E-4	3.85052	15 - 0; 15 - 30; 15 - 45
Lactate	2.0E-4	3.6996	30 - 0; 30 - 15; 30 - 45
Aspartate	2.1E-4	3.68363	0 - 45; 15 - 45; 30 - 45
Isobutyrate	4.9E-4	3.30596	0 - 30; 0 - 45; 15 - 30; 15 - 45
Uracil	9.0E-4	3.04602	15 - 0; 30 - 0; 45 - 0
Dimethylamine	0.00502	2.29908	45 - 0; 45 - 15; 45 - 30
Propionate	0.00536	2.27083	45 - 0; 45 - 30
Lysine	0.00559	2.2524	0 - 30; 45 - 15; 45 - 30
Acetate	0.0081	2.0913	0 - 30; 0 - 45; 15 - 30
pH	0.01032	1.98623	0 - 30; 0 - 45; 15 - 45
NDMA	0.01665	1.77856	0 - 15; 30 - 15; 45 - 15
Ferulate	0.0173	1.76206	15 - 30; 45 - 30

# Feature selection - SAM





# Data Download

[Home](#)

**Steps**

- 1. Upload
- 2. Process
- 3. Data Editor
- 4. Analyze
  - Univariate
  - ANOVA
  - PCA
  - PLSDA
  - SAM
  - EBAM
  - Tree & heatmap
  - Kmean & SOM
  - RandomForest
  - R-SVM
- 5. Peak Search
- 6. Pathway Mapping
- 7. Download
- 8. Log Out

## Download

The "Download.zip" can be downloaded via email link. The data is available via email address

The email service is available below.

- [Download.zip](#)
- [compounds.csv](#)
- [data\\_normalized.csv](#)
- [data\\_processed.csv](#)
- [pls\\_loading.png](#)
- [pls\\_permut.png](#)
- [pls\\_score3d.png](#)
- [rf\\_cls.png](#)
- [Rhistry.R](#)
- [sam\\_fdr.png](#)
- [univar\\_t.png](#)

## 2.2 Principal Component Analysis (PCA)

PCA is an unsupervised method aiming to find the directions that best explain the variance in a data set ( $X$ ) without referring to class labels ( $Y$ ). The data are summarized into much fewer variables called *scores* which are weighted average of the original variables. The weighting profiles are called *loadings*. The PCA analysis is performed using the `prcomp` package. The calculation is based on singular value decomposition.

The Rscript `chemometrics.R` is required. Figure 6 is pairwise score plots providing an overview of the various separation patterns among the most significant PCs; Figure 7 is the scree plot showing the variances explained by the selected PCs; Figure 8 shows the 2-D score plot between selected PCs; Figure 9 shows the 3-D score plot between selected PCs; Figure 10 shows the loading plot between the selected PCs; Figure 11 shows the biplot between the selected PCs.

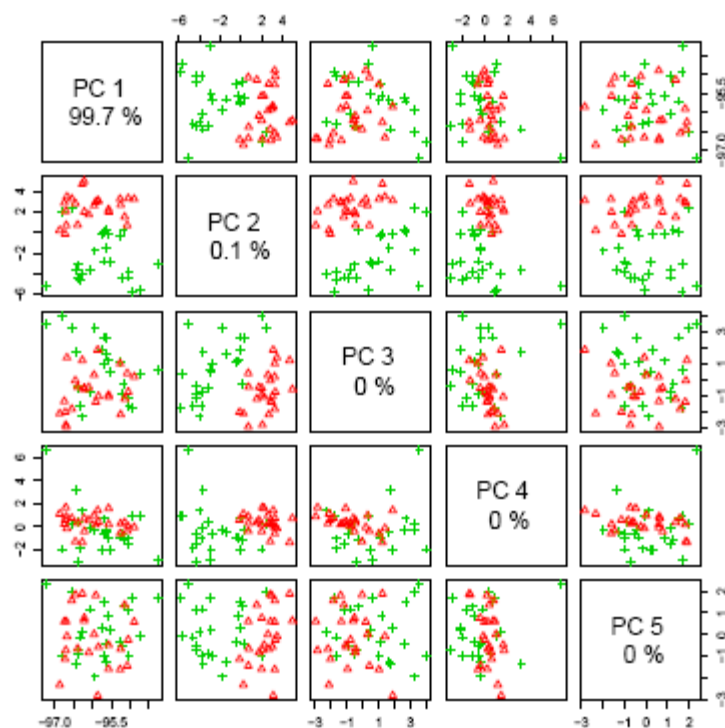


Figure 6: Pairwise score plots between the selected PCs. The explained variance of each PC is shown in the corresponding diagonal cell.



# Updates & Forecast

- Recently upgraded
  - Support for multiple group analysis
  - One-way ANOVA & post-hoc analysis
- To be added
  - To add some advanced methods for
    - Association analysis / ROC / OPLS
  - To enhance web interfacing with XCMS
  - Allow local installation
    - To be released by the end of this summer

# Data Interpretation

- Manual approach
  - Background knowledge plus literature search
  - Basic & Intuitive
  - Can be very accurate
  - Issues
    - Time-consuming
    - Subjective
    - Lack of statistical strength

# Introducing MSEA

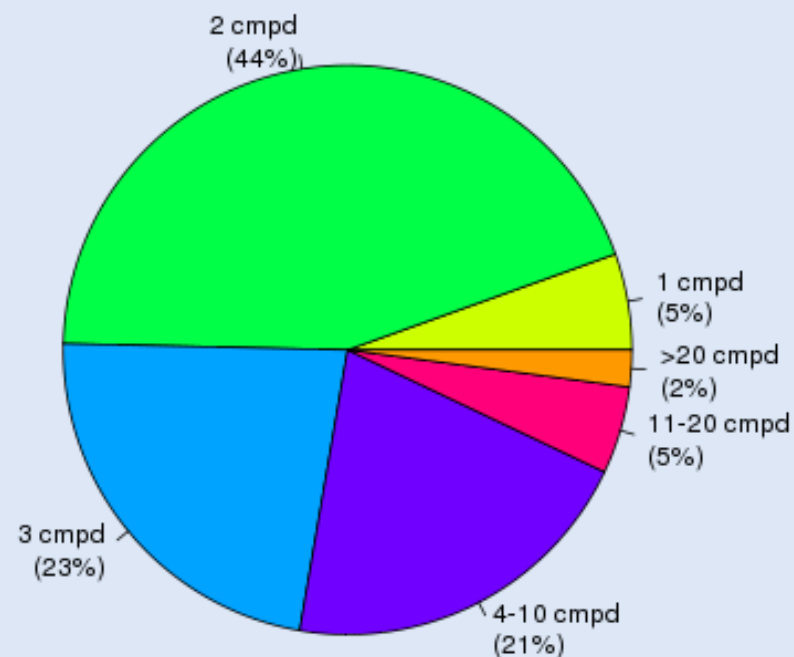
- <http://www.msea.ca>
- Metabolite Set Enrichment Analysis
- Identify **biological meaningful patterns** from quantitative metabolomics data

# Biologically meaningful

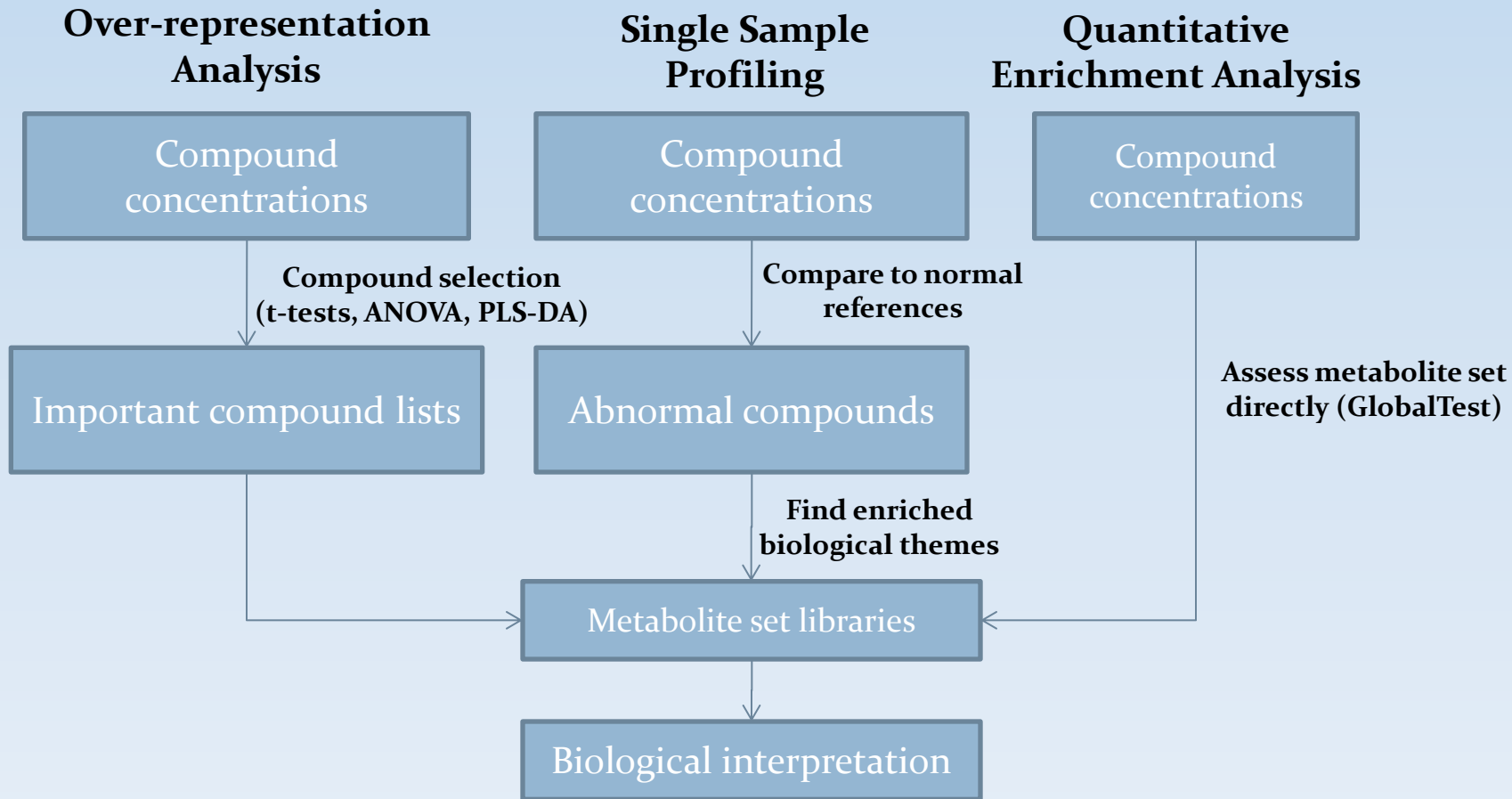
Summary of Human Metabolite Set Libraries

Category	Metabolite set #
All	5,380
Disease associated (blood)	344
Disease associated (urine)	290
Disease associated (CSF)	108
Single Nucleotide Polymorphism (SNP) associated	4,501
Biochemical pathways	80
Tissue or sub-cellular location	57

Distribution of cmpd # in metabolite sets



# The MSEA approach



# MSEA @ www.msea.ca



## Metabolite Set Enrichment Analysis (MSEA)

*- discover biologically meaningful patterns in quantitative metabolomic data*

### News and Updates:

- Added 4,500 **SNP-associated metabolite sets** (06/05/10); **NEW**
- Added support for [Biocrates](#) metabolite IDs (06/01/10);

### Overview

MSEA is a web-based tool to help identify and interpret patterns of metabolite concentration changes in a biologically meaningful context for **human** and **mammalian** metabolomic studies.

MSEA provides three types of enrichment analyses:

- ORA performs over representation analysis for a [list of metabolites](#);
- SSP performs single sample profiling on a [biofluid sample](#) by first comparing the measured compound concentrations to their normal ranges reported in literature and then testing for potentially interesting patterns;
- QEA performs quantitative enrichment analysis directly on a [compound concentration table](#) with either discrete (binary, multi-class) or continuous phenotype labels.

The analyses are based on five built-in metabolite set libraries containing over 1,000 biologically meaningful groups of metabolites. In addition, users can upload their self-defined metabolite sets (i.e. defined for other species) for enrichment analysis.

MSEA enables simultaneous biomarker discovery and functional

### Enrichment Analysis

[Over Representation Analysis \(ORA\)](#)

[Single Sample Profiling \(SSP\)](#)

[Quantitative Enrichment Analysis \(QEA\)](#)

### Other Tasks

[Compound ID Conversion](#)

[Browse Metabolite Set Libraries](#)

### Documentation

[MSEA Workflow](#)

[Library Descriptions](#)

[Screenshot Tutorials](#)



# Over-representation analysis



Over Rep

## Single Sample Profiling (SPP)

Enter your data below (two-column data):

- compound labels and concentration values separated by tab

L-Isolecine	0.34
Fumaric acid	0.47
Acetone	0.58
Succinic acid	9.4
1-Methylhistidine	9.6
L-Asparagine	19.62
3-Methylhistidine	9.7
L-Threonine	93.19
Creatine	720
cis-Aconitic acid	14.39
L-Tryptophan	35.78
L-Carnitine	16.01
L-Serine	17.32
L-Tyrosine	67.51
L-Alanine	219.02
L-Fucose	20.37
D-Glucose	23.92
Pyroglutamic acid	26.38
Formic acid	26.72
Indoxyl sulfate	34.21
Dimethylamine	38.28
Ethanolamine	39.29
Glycolic acid	41.39
L-Glutamine	52.99
L-Histidine	55.95
Trigonelline	57.4
3-Aminoisobutanoic acid	89.76
Taurine	116

Compound label:

Compound names

Biofluid (unit)

Urine (umol/mmol\_creatinine)

# Compound label standardization



## Metabolite Set Enrichment Analysis (MSEA)

- discover biologically meaningful patterns in quantitative metabolomic data

### Compound Label Standardization:

[<< Back](#)

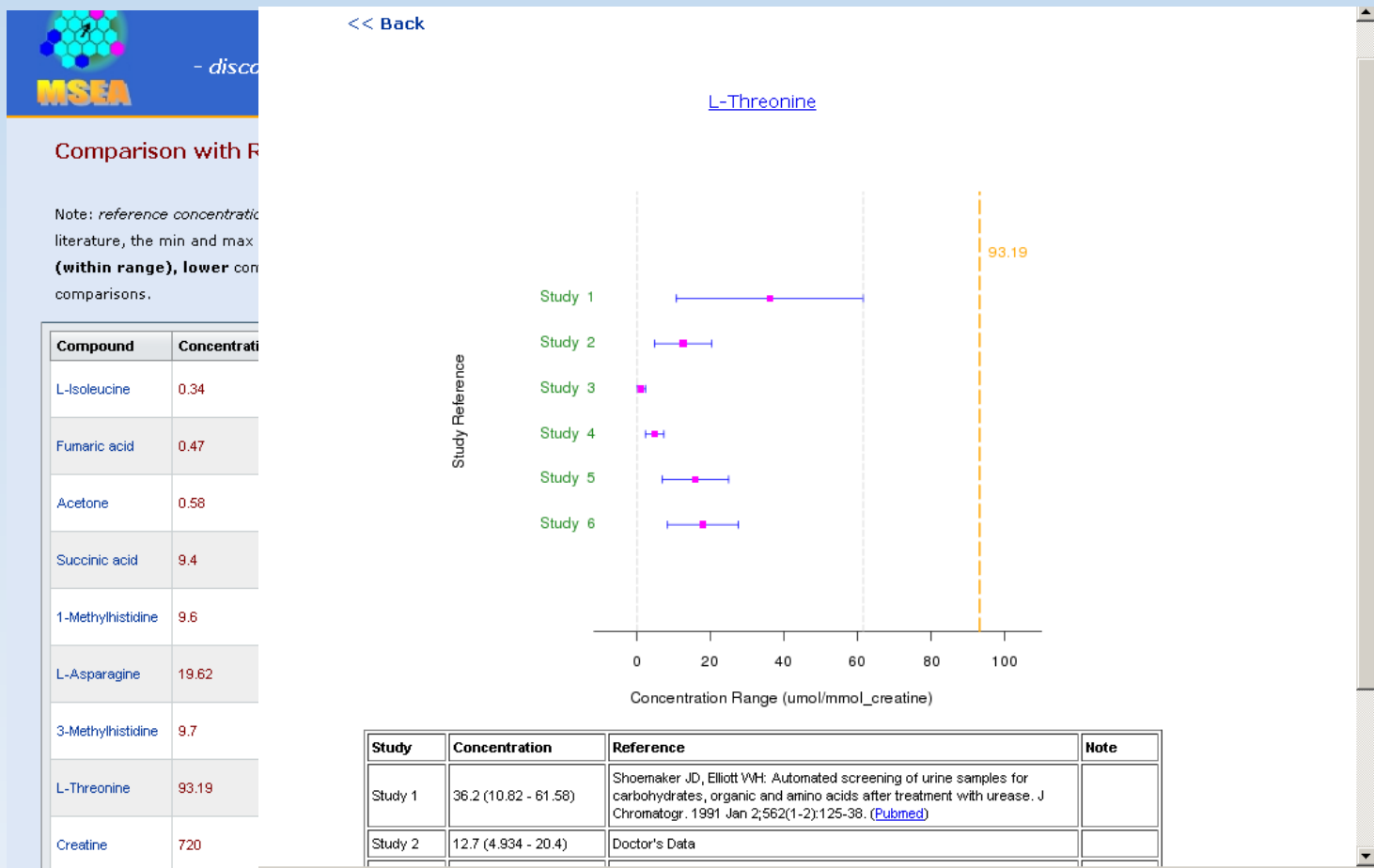
[Home](#)

Please note:

- Query names in normal white indicate exact match - marked by "1" in the download file;
- Query names highlighted in **yellow** indicate **approximate matches** (for compound name matches) - marked by "2" in the downloaded file. Users should manually select the correct match by clicking the [View](#) link of the corresponding compounds. Otherwise, the first match will be used;
- Query names highlighted in **red** indicate **no match** - marked by "0" in the downloaded file;
- Greek alphabets are not recognized, they should be replaced by English names (i.e. alpha, beta)

Query	Best Match	HMDB	Details
<a href="#">1,6-Anhydro-beta-D-glucose</a>	Glucosan	HMDB00640	
<a href="#">1-Methylnicotinamide</a>	1-Methylnicotinamide	HMDB00699	
<a href="#">2-Aminobutyrate</a>	L-Alpha-aminobutyric acid	HMDB00452	
<a href="#">2-Hydroxyisobutyrate</a>	Alpha-Hydroxyisobutyric acid	HMDB00729	
<b>2-Oxoglutarate</b>	Oxoglutaric acid	HMDB00208	<a href="#">View</a>
<a href="#">3-Aminoisobutyrate</a>	3-Aminoisobutanoic acid	HMDB03911	
<a href="#">3-Hydroxybutyrate</a>	3-Hydroxybutyric acid	HMDB00357	
<a href="#">3-Hydroxyisovalerate</a>	3-Hydroxyisovaleric acid	HMDB00754	
<a href="#">3-Indoxylsulfate</a>	Indoxyl sulfate	HMDB00682	
<a href="#">4-Hydroxyphenylacetate</a>	p-Hydroxyphenylacetic acid	HMDB00020	
<a href="#">Acetate</a>	Acetic acid	HMDB00042	
<a href="#">Acetone</a>	Acetone	HMDB01659	
<a href="#">Adipate</a>	Adipic acid	HMDB00448	
<a href="#">Alanine</a>	L-Alanine	HMDB00161	
<a href="#">Asparagine</a>	L-Asparagine	HMDB00168	
<a href="#">Betaine</a>	Betaine	HMDB00043	
<a href="#">Carnitine</a>	L-Carnitine	HMDB00062	
<a href="#">Citrate</a>	Citric acid	HMDB00094	
<a href="#">Creatine</a>	Creatine	HMDB00064	

# Identify abnormal concentration



# Library selection

## Choose a Metabolite Set Library

[Home](#)

**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 strong association ( $p$  value  $< 1e-3$ ) with detected single nucleotide polymorphisms (SNPs) loci.

**Location-based metabolite sets**

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

[Self-defined metabolite sets](#)


Click the link above to upload your own customized metabolite set library

**Only use metabolite sets containing at least**

Submit



# Download



## Metabolite Set Enrichment Analysis (MSEA)




*- discover biologically meaningful patterns in quantitative metabolomic data*

**Result Download** [Home](#)

The "Download.zip" contains all the files in your home directory. These data will remain in the server for 72 hours before being deleted automatically.

Download.zip
MSEA_Report.pdf
msea-gea.png
name_map.csv
pca-load.png
pca-score.png
pls-load.png
pls-score.png
Rhistory.R

[Log Out](#)

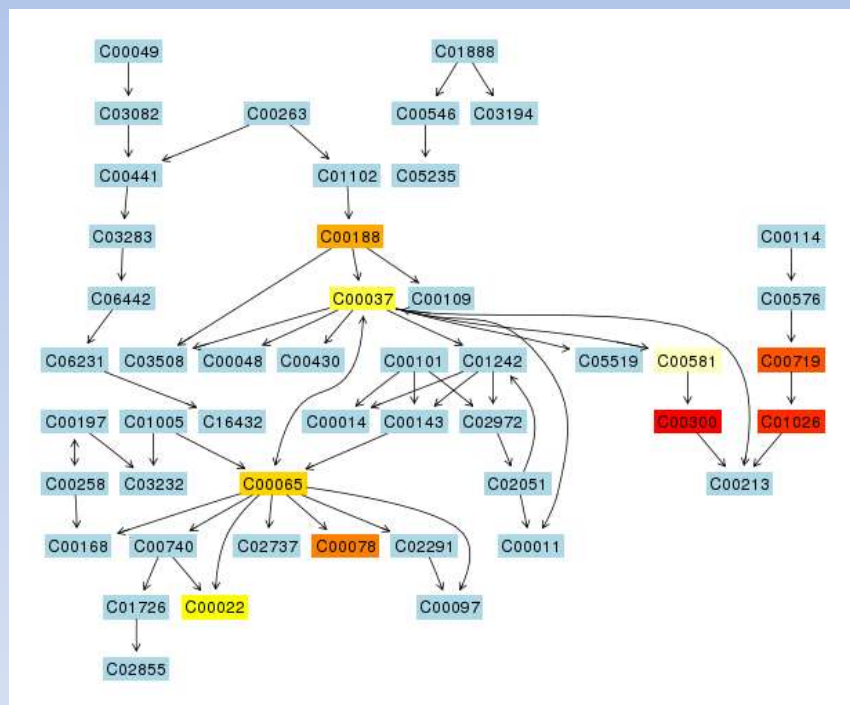
**Contact Information:**  
If you have comments or questions, please contact:  
[jianguox@ualberta.ca](mailto:jianguox@ualberta.ca)

# MSEA summary

- More biologically-motivated
- Simultaneously biomarker identification and interpretation
- Automatic comparison with published data
  - Important patterns
  - Reference concentrations
- Potential issues
  - Limited by the size and quality of the knowledge database
  - For pathway-based metabolite sets
    - Does not consider the **pathway topology**

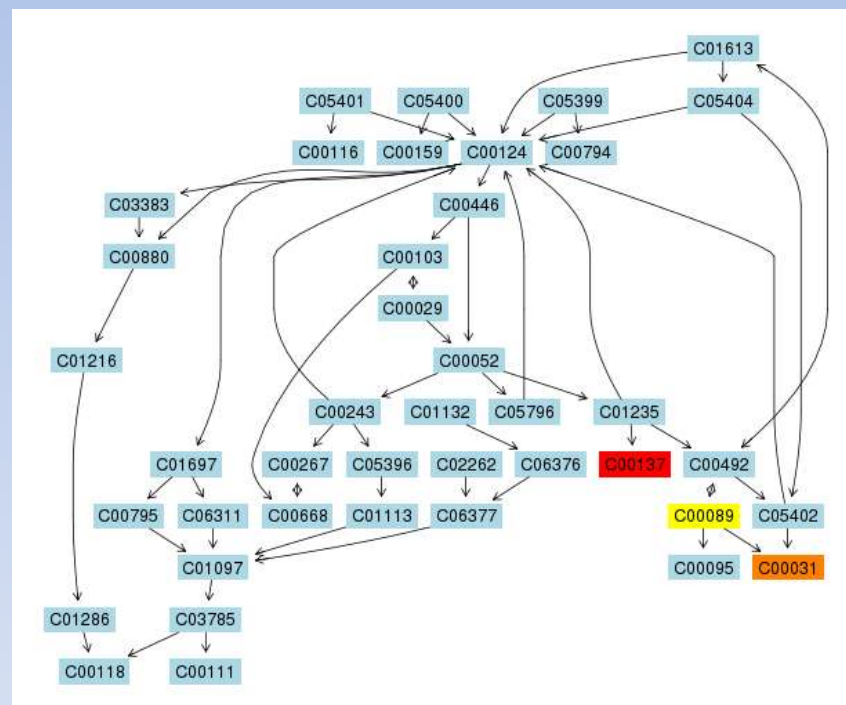
# Topology matters

Glycine, serine and threonine metabolism



$p = 1e-5$

Galactose metabolism



$p = 1e-7$



# Introducing MetPA

- <http://metpa.metabolomics.ca>
- Pathway Analysis Tool
  - 884 pathways covering 11 model organisms
  - Enrichment Analysis
    - Global Test
    - Global ANCOVA
  - Topology Analysis
    - Degree Centrality
    - Betweenness Centrality
  - Google-map style visualization

# MetPA



MetPA  
MetPA

A web-based metabolomics tool for pathway analysis & visualization

[Home](#) [Help](#) [Library](#)

## Welcome ( [Click here to start analysis](#) )

MetPA (Metabolomics Pathway Analysis) is a free and easy-to-use web application designed to perform pathway analysis and visualization of quantitative metabolomic data.

- MetPA accepts either a list of important compounds identified from your studies, or a metabolite concentration table with [binary](#), [multi-group](#), or [continuous](#) phenotype labels.
- MetPA combines three complementary analyses - [pathway enrichment analysis](#) ( including [hypergeometric test](#), [Fishers' exact test](#), [Globaltest](#), and [GlobalAncova](#) ), [pathway topology analysis](#) ( based on [degree centrality](#) or [betweenness centrality](#) measures ), and [univariate analysis](#) ( including [t-test](#), [ANOVA](#), and [linear regression](#) ), to help identify the most relevant metabolic pathways involved in the conditions under study;
- MetPA implements a [Google-Map style interactive network visualization system](#) which provides a comprehensive three-level view - [metabolome view](#), [pathway view](#), and [compound view](#). Users can intuitively explore the analysis results through point and click. The system also supports lossless zooming, dragging, linking, and highlighting;
- MetPA currently supports pathway analysis for 11 model organisms, including Human, Mouse, Rat, Cow, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Budding yeast, and E.coli., with a total of 884 pathways.



### Contact Information:

If you have comments or questions, please contact: [jianguox@ualberta.ca](mailto:jianguox@ualberta.ca).

# 11 pathway libraries (KEGG)

Please select a pathway library :

## Mammals

- Homo sapiens (human) [80]
- Mus musculus (mouse) [82]
- Rattus norvegicus (rat) [81]
- Bos taurus (cow) [81]

## Fishes

- Danio rerio (zebrafish) [81]

## Insects

- Drosophila melanogaster (fruit fly) [79]

## Nematodes

- Caenorhabditis elegans (nematode) [78]

## Fungi

- Saccharomyces cerevisiae (yeast) [65]

## Plants

- Oryza sativa japonica (Japanese rice) [83]
- Arabidopsis thaliana (thale cress) [87]

## Prokaryotes

- Escherichia coli K-12 MG1655 [87]

# Combining Enrichment analysis & Topology analysis

Please specify pathway analysis algorithms :

<b>Over Representation Analysis</b>	<input checked="" type="radio"/> Hypergeometric Test
	<input type="radio"/> Fisher's Exact Test
<hr/>	
<b>Pathway Topology Analysis</b>	<input checked="" type="radio"/> Relative-betweeness Centrality <sup>?</sup>
	<input type="radio"/> Out-degree Centrality <sup>?</sup>

Submit

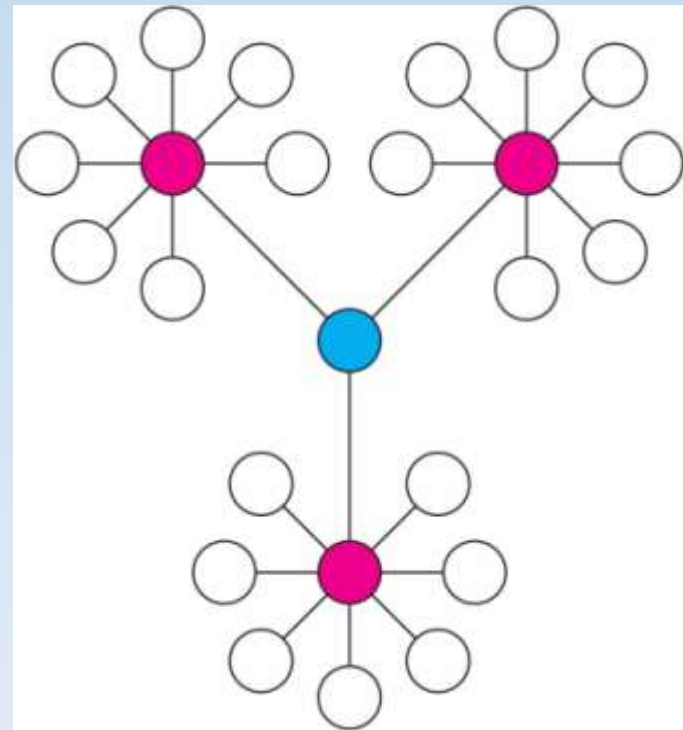
Please specify pathway analysis algorithms :

<b>Pathway Enrichment Analysis</b>	<input checked="" type="radio"/> Global Test
	<input type="radio"/> Global Ancova
<hr/>	
<b>Pathway Topology Analysis</b>	<input checked="" type="radio"/> Relative-betweeness Centrality <sup>?</sup>
	<input type="radio"/> Out-degree Centrality <sup>?</sup>

Submit

# Node importance measure: centrality

- Degree Centrality
  - Local structure;
  - Highly connected (hub)
  - **The Red nodes**
- Betweenness Centrality
  - Global structures;
  - Sits on many shortest paths between other nodes
  - **The Blue node**

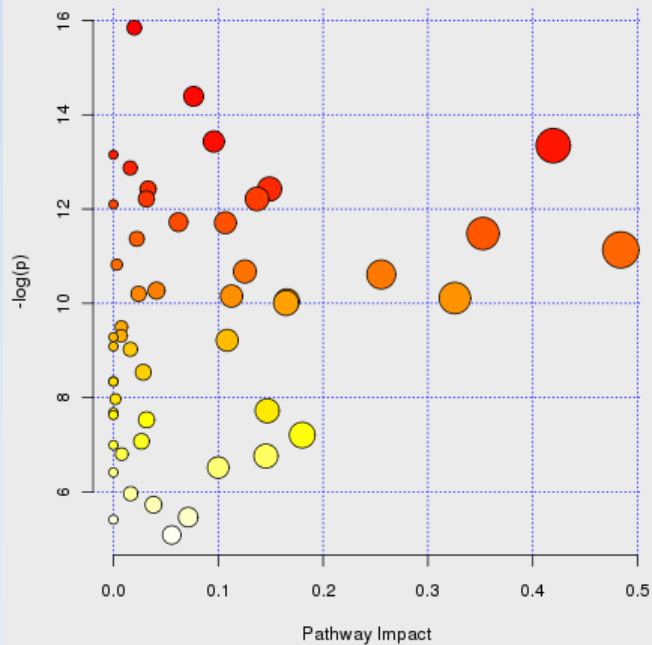


Junker *et al.* *BMC Bioinformatics* 2006

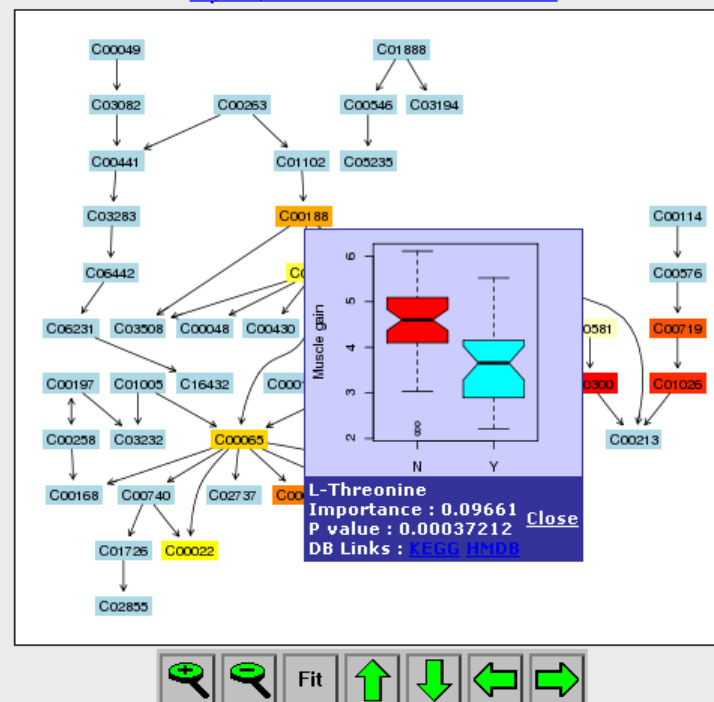
# Point and Click

The pathway can be launched either by clicking the corresponding node on the left image or by clicking the pathway name from the table below. Please note, each node (compound) is clickable. You can zoom in and out using the control buttons below, and then drag the image to the locations of your interest. Place mouse over each metabolite node will reveal its common name. Click the node will trigger **compound view** of the selected compound.

Overview of Pathway Analysis



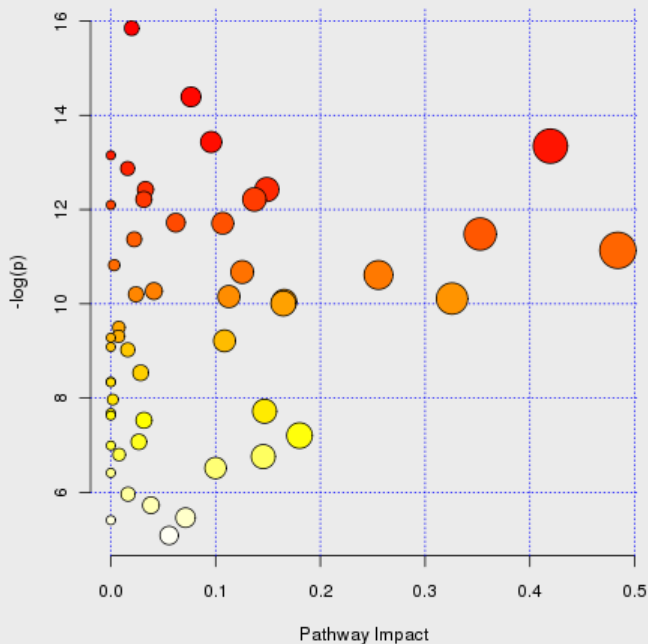
Glycine, serine and threonine metabolism



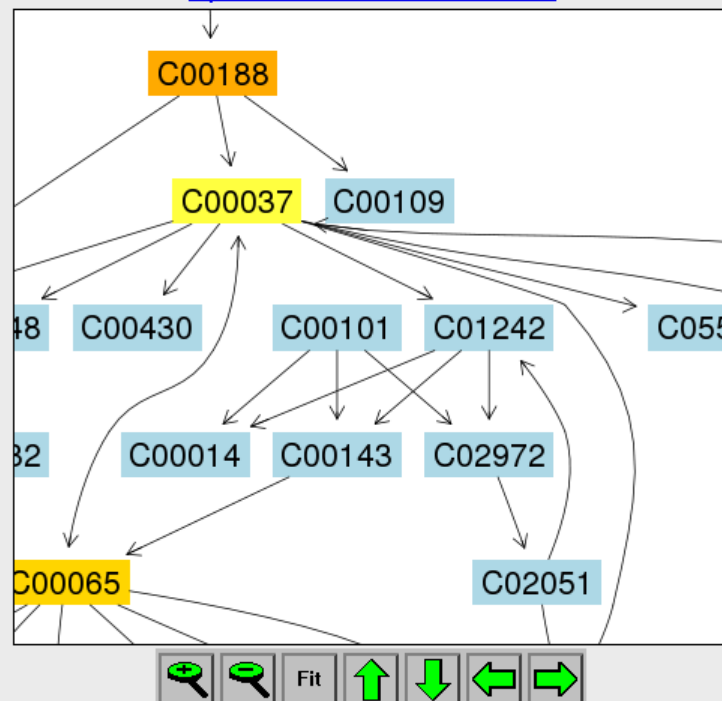
# Lossless zooming

The pathway can be launched either by clicking the corresponding node on the left image or by clicking the pathway name from the table below. Please note, each node (compound) is clickable. You can zoom in and out using the control buttons below, and then drag the image to the locations of your interest. Place mouse over each metabolite node will reveal its common name. Click the node will trigger **compound view** of the selected compound.

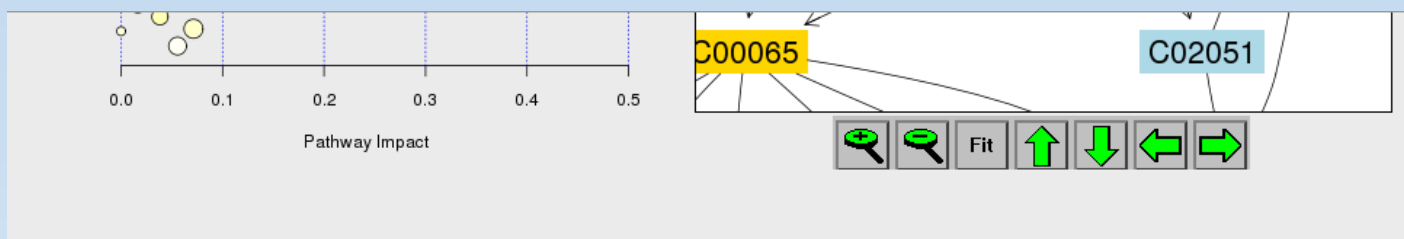
Overview of Pathway Analysis



[Glycine, serine and threonine metabolism](#)



# Result table



Pathway Name	Total Cmpd	Hits	Raw p	$-\log(p)$	Holm p	FDR	Impact	Details
Galactose metabolism	41	3	1.3105E-7	6.8826	6.6833E-6	6.6833E-6	0.01992	<a href="#">KEGG SMP</a>
Starch and sucrose metabolism	50	3	5.6226E-7	6.2501	2.8113E-5	1.4338E-5	0.0765	<a href="#">KEGG SMP</a>
Glycolysis or Gluconeogenesis	31	4	1.4659E-6	5.8339	7.183E-5	1.9816E-5	0.09576	<a href="#">KEGG SMP SMP</a>
Pyruvate metabolism	32	4	1.5995E-6	5.796	7.6774E-5	1.9816E-5	0.41957	<a href="#">KEGG SMP</a>
Amino sugar and nucleotide sugar metabolism	88	3	1.9428E-6	5.7116	9.1311E-5	1.9816E-5	0.0	<a href="#">KEGG SMP SMP</a>
Propanoate metabolism	35	4	2.5699E-6	5.5901	1.1822E-4	2.1844E-5	0.01603	<a href="#">KEGG SMP</a>
Valine, leucine and isoleucine biosynthesis	27	6	4.0178E-6	5.396	1.808E-4	2.5278E-5	0.14892	<a href="#">KEGG SMP</a>
Sulfur metabolism	18	2	4.019E-6	5.3959	1.808E-4	2.5278E-5	0.03307	<a href="#">KEGG SMP</a>
Phenylalanine metabolism	45	6	4.9533E-6	5.3051	2.1299E-4	2.5278E-5	0.0315	<a href="#">KEGG SMP</a>
Inositol phosphate metabolism	39	1	4.9565E-6	5.3048	2.1299E-4	2.5278E-5	0.13703	<a href="#">KEGG SMP</a>
Pentose phosphate pathway	32	2	5.5736E-6	5.2539	2.2852E-4	2.5841E-5	0.0	<a href="#">KEGG SMP</a>
Arginine and proline metabolism	77	6	8.1109E-6	5.0909	3.2443E-4	3.2248E-5	0.06203	<a href="#">KEGG SMP</a>
Tyrosine metabolism	76	5	8.22E-6	5.0851	3.2443E-4	3.2248E-5	0.10681	<a href="#">KEGG SMP SMP</a>
Taurine and hypotaurine metabolism	20	3	1.0339E-5	4.9855	3.9288E-4	3.7664E-5	0.35252	<a href="#">KEGG SMP</a>
Valine, leucine and isoleucine degradation	40	3	1.1562E-5	4.937	4.278E-4	3.9311E-5	0.02232	<a href="#">KEGG SMP</a>
Glycine, serine and threonine metabolism	48	9	1.4605E-5	4.8355	5.2577E-4	4.6552E-5	0.48394	<a href="#">KEGG SMP</a>
Selenoamino acid metabolism	22	1	2.0003E-5	4.6989	7.001E-4	6.0008E-5	0.00321	<a href="#">KEGG SMP</a>
Butanoate metabolism	40	5	2.3138E-5	4.6357	7.867E-4	6.5559E-5	0.12541	<a href="#">KEGG SMP</a>
Alanine, aspartate and glutamate metabolism	24	6	2.4638E-5	4.6084	8.1306E-4	6.6134E-5	0.25546	<a href="#">KEGG SMP SMP SMP</a>
Nicotinate and nicotinamide metabolism	44	5	3.4668E-5	4.4601	0.0011094	8.8404E-5	0.04113	<a href="#">KEGG SMP</a>
Pentose and glucuronate interconversions	53	2	3.7128E-5	4.4303	0.001151	9.0104E-5	0.02401	<a href="#">KEGG</a>
Aminoacyl-tRNA biosynthesis	75	12	3.8942E-5	4.4096	0.0011683	9.0104E-5	0.11268	<a href="#">KEGG</a>
Citrate cycle (TCA cycle)	20	6	4.0635E-5	4.3911	0.0011784	9.0104E-5	0.32569	<a href="#">KEGG SMP</a>



# Downloads

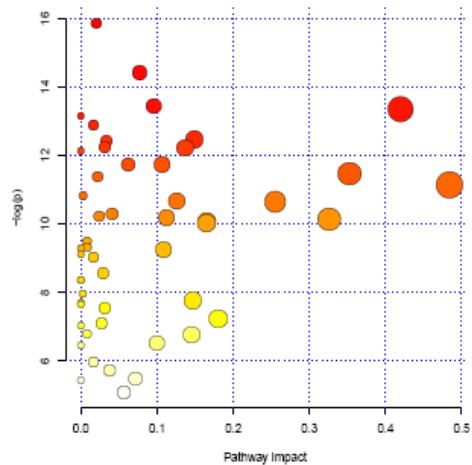


Figure 1: Summary of Pathway Analysis

The table below shows the detailed results from the pathway analysis. Since we are testing many pathways at the same time, the statistical p values from enrichment analysis are further adjusted for multiple testings. In particular, the Total is the total number of compounds in the pathway; the Hits is the actually matched number from the user uploaded data; the Raw p is the original p value calculated from the enrichment analysis; the Holm p is the p value adjusted by Holm-Bonferroni method; the FDR p is the p value adjusted using False Discovery Rate; the Impact is the pathway impact value calculated from pathway topology analysis.

Table 2: Result from Pathway Analysis with MetPA

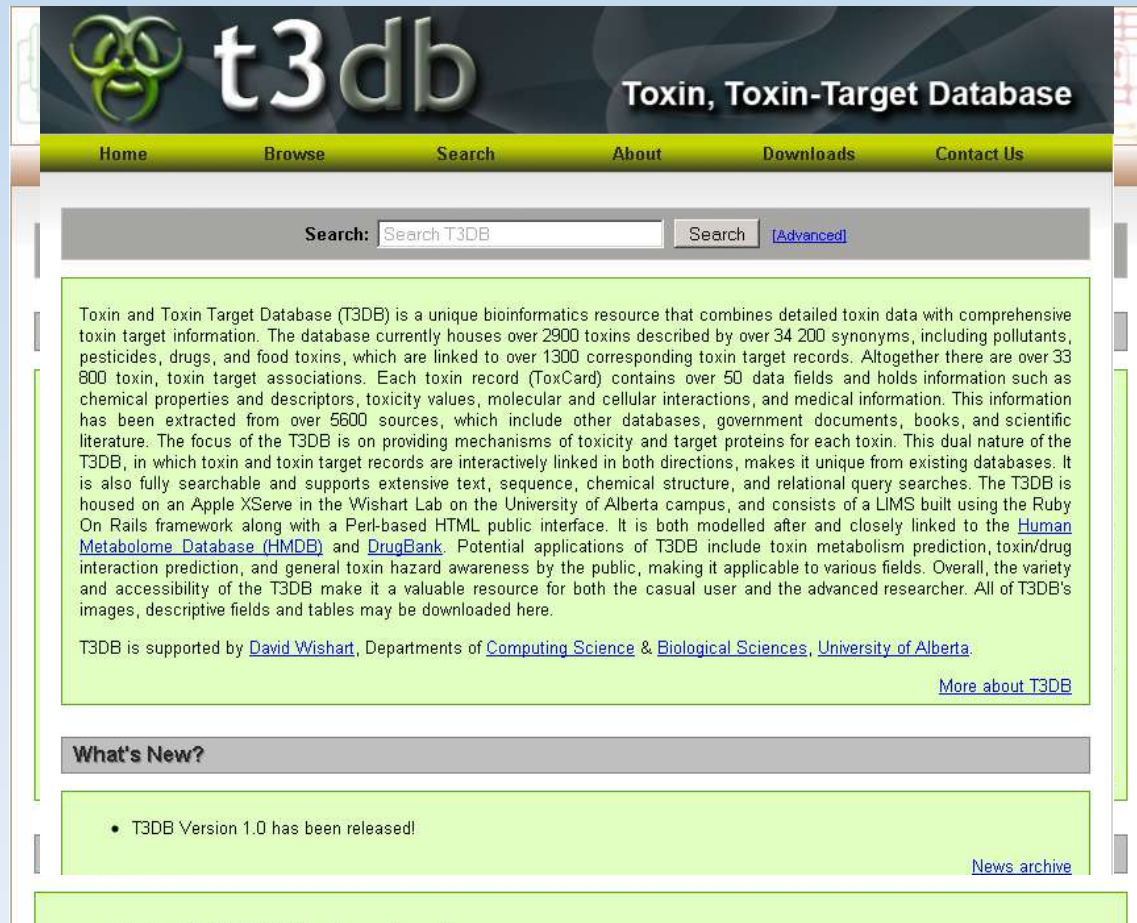
Metabolic Pathway	Total	Hits	Raw p	Holm p	FDR p	Impact
Alanine metabolism	41	3	1.21E-07	8.39E-05	4.28E-05	1.88E-04
Glact and inosine metabolism	50	3	5.62E-07	8.25E-08	2.81E-05	1.43E-02
Glycolysis or Gluconeogenesis	31	4	1.47E-06	5.89E-05	7.18E-05	1.98E-02
Pyruvate metabolism	32	4	1.60E-06	5.81E-05	7.08E-05	1.98E-02
Amino sugar and nucleoside sugar metabolism	88	3	1.84E-06	5.71E-05	8.19E-05	1.98E-02
Propanoate metabolism	35	4	2.57E-06	5.59E-05	1.18E-04	1.18E-02
Valine, leucine and isoleucine biosynthesis	27	4	4.02E-06	5.02E-05	1.81E-04	2.52E-02
Sulfur metabolism	18	2	4.02E-06	5.02E-05	1.81E-04	2.52E-02
Phenylalanine metabolism	45	4	4.02E-06	5.02E-05	1.81E-04	2.52E-02
Resolyl phosphate metabolism	39	3	4.98E-06	5.39E-05	2.13E-04	2.52E-02
Protein catabolism pathway	22	2	5.57E-06	1.25E-05	2.29E-04	2.52E-02
Arginine and proline metabolism	77	4	8.11E-06	5.09E-05	3.24E-04	3.22E-02
Tyrosine metabolism	78	5	8.28E-06	5.09E-05	3.24E-04	3.22E-02
Taurine and hypoxanthine metabolism	20	3	1.02E-05	4.99E-05	3.02E-04	3.17E-02
Valine, leucine and isoleucine degradation	40	3	1.16E-05	4.94E-05	4.28E-04	3.92E-02
Oxyma acids and thiamine metabolism	48	3	1.46E-05	4.84E-05	1.28E-04	4.89E-02
Selenoamino acid metabolism	22	1	2.00E-05	4.70E-05	7.00E-04	6.00E-02
Nucleoside metabolism	40	2	2.31E-05	4.64E-05	7.17E-04	8.89E-02
Alanine, aspartate and glutamate metabolism	34	4	3.49E-05	4.41E-05	8.13E-04	8.81E-02
Nucleoside and nucleotamide metabolism	44	5	3.47E-05	4.40E-05	1.11E-03	9.04E-02
Protein and glutathione interconversions	53	2	2.71E-05	4.40E-05	1.15E-03	9.04E-02
Anticancer-ATM biosynthesis	15	12	2.99E-05	4.41E-05	1.17E-03	9.04E-02
Citrate cycle (TCA cycle)	20	4	4.06E-05	4.39E-05	1.16E-03	9.04E-02
Oxycarboxylate and dicarboxylate metabolism	50	8	4.39E-05	4.38E-05	1.22E-03	9.04E-02
Methane metabolism	34	4	4.51E-05	4.37E-05	1.22E-03	9.04E-02
Nitrogen metabolism	39	4	7.81E-05	4.14E-05	1.01E-03	1.47E-04
Phenylalanine, tyrosine and tryptophan biosynthesis	27	2	9.02E-05	4.04E-05	2.29E-03	1.70E-04
Hexose metabolism	11	1	9.92E-05	4.03E-05	2.29E-03	1.70E-04
Tryptophan metabolism	79	3	9.98E-05	4.01E-05	2.29E-03	1.70E-04
Synthetic and degradation of ketone bodies	4	2	1.13E-04	3.92E-05	2.50E-03	1.92E-04
Ascorbate and aldarate metabolism	45	3	1.20E-04	3.92E-05	2.50E-03	1.92E-04
Cysteine and methionine metabolism	58	2	1.98E-04	3.71E-05	3.02E-03	2.13E-04
Threonine and other serine-threonine biosynthesis	28	1	2.29E-04	3.62E-05	4.02E-03	3.02E-04
Cysteine and methionine metabolism	18	2	2.42E-04	3.52E-05	4.02E-03	3.02E-04
Glutathione metabolism	38	2	2.47E-04	3.49E-05	5.90E-03	5.05E-04
Lysine degradation	47	3	4.44E-04	3.25E-05	7.10E-03	6.29E-04
Glutamine metabolism	25	1	4.81E-04	3.14E-05	7.10E-03	6.29E-04
Tryptophan biosynthesis	27	1	4.86E-04	3.11E-05	7.10E-03	6.29E-04
Protein and nucleic acid metabolism	48	1	5.98E-04	3.07E-05	7.10E-03	6.29E-04
Pantoic acid and CoA biosynthesis	22	4	7.44E-04	1.13E-05	8.92E-03	8.89E-04
D-Glucosamine and D-galactosamine metabolism	11	2	8.51E-04	1.07E-05	8.92E-03	1.08E-02
Thiamine metabolism	31	1	9.10E-04	1.02E-05	8.92E-03	1.12E-02
Urea metabolism	62	1	1.11E-03	9.82E-05	1.02E-02	1.22E-02
Nucleic acid metabolism	104	3	1.16E-03	2.00E-05	1.00E-02	1.22E-02
Lysine biosynthesis	22	2	1.48E-03	2.48E-05	1.48E-02	1.82E-02
Porphyry and chlorophyll biosynthesis	44	2	1.64E-03	2.79E-05	1.64E-02	1.82E-02
Primary bile acid biosynthesis	47	1	1.82E-03	2.99E-05	1.82E-02	2.02E-02
Vitamin B6 metabolism	22	2	2.26E-03	2.49E-05	1.81E-02	3.47E-02
Pyrimidine metabolism	61	2	2.26E-03	2.49E-05	1.81E-02	3.47E-02
Hex-amine metabolism	38	1	4.48E-03	1.25E-05	1.21E-02	4.21E-02
Oxycarboxylate metabolism	22	1	4.48E-03	1.25E-05	1.21E-02	4.21E-02
Oxycarboxylate metabolism	29	1	8.10E-03	2.21E-05	1.21E-02	8.10E-02

# MetPA summary

- Combine statistical analysis and topological analysis
  - Results are more close to manual identification
- Highly interactive visualization system
  - allows easy hierarchical navigation within a large amount of information

# Public Databases

- HMDB
- DrugBank
- SMPDB
- T<sub>3</sub>DB



The screenshot shows the homepage of the Toxin, Toxin-Target Database (T3DB). The header features the T3DB logo, which consists of three interlocking green rings, and the text "t3db" in a large, white, sans-serif font. To the right of the logo, the text "Toxin, Toxin-Target Database" is displayed in a smaller, white font. Below the header is a navigation bar with the following links: Home, Browse, Search, About, Downloads, and Contact Us. The main content area is divided into several sections. The first section is a search bar with the text "Search: Search T3DB" and a "Search" button, along with a link to "Advanced" search. Below the search bar is a large green box containing a detailed description of the T3DB. The description states that T3DB is a unique bioinformatics resource that combines detailed toxin data with comprehensive toxin target information. It currently houses over 2900 toxins described by over 34,200 synonyms, including pollutants, pesticides, drugs, and food toxins, which are linked to over 1300 corresponding toxin target records. Altogether there are over 33,800 toxin, toxin target associations. Each toxin record (ToxCARD) contains over 50 data fields and holds information such as chemical properties and descriptors, toxicity values, molecular and cellular interactions, and medical information. This information has been extracted from over 5600 sources, which include other databases, government documents, books, and scientific literature. The focus of the T3DB is on providing mechanisms of toxicity and target proteins for each toxin. This dual nature of the T3DB, in which toxin and toxin target records are interactively linked in both directions, makes it unique from existing databases. It is also fully searchable and supports extensive text, sequence, chemical structure, and relational query searches. The T3DB is housed on an Apple Xserve in the Wishart Lab on the University of Alberta campus, and consists of a LIMS built using the Ruby On Rails framework along with a Perl-based HTML public interface. It is both modelled after and closely linked to the [Human Metabolome Database \(HMDB\)](#) and [DrugBank](#). Potential applications of T3DB include toxin metabolism prediction, toxin/drug interaction prediction, and general toxin hazard awareness by the public, making it applicable to various fields. Overall, the variety and accessibility of the T3DB make it a valuable resource for both the casual user and the advanced researcher. All of T3DB's images, descriptive fields and tables may be downloaded here. T3DB is supported by [David Wishart](#), Departments of [Computing Science](#) & [Biological Sciences](#), [University of Alberta](#). A link to "More about T3DB" is provided at the bottom right of the green box. Below the green box is a "What's New?" section with a grey header. The content of this section is a single bullet point: "T3DB Version 1.0 has been released!". A link to "News archive" is provided at the bottom right of this section.

# Summary

1<sup>st</sup> & 2<sup>nd</sup>  
generation

- MetaboAnalyst: general data processing & analysis



3<sup>rd</sup> generation

- MSEA: Metabolite set enrichment analysis



4<sup>th</sup>  
generation

- MetPA: Metabolomics Pathway Analysis



5<sup>th</sup>  
generation

- Integrate with other omics data



## Acknowledgement

- Dr. David Wishart



- ❖ Alberta Ingenuity Fund (AIF)
- ❖ The Human Metabolome Project (HMP)
- ❖ University of Alberta, Canada