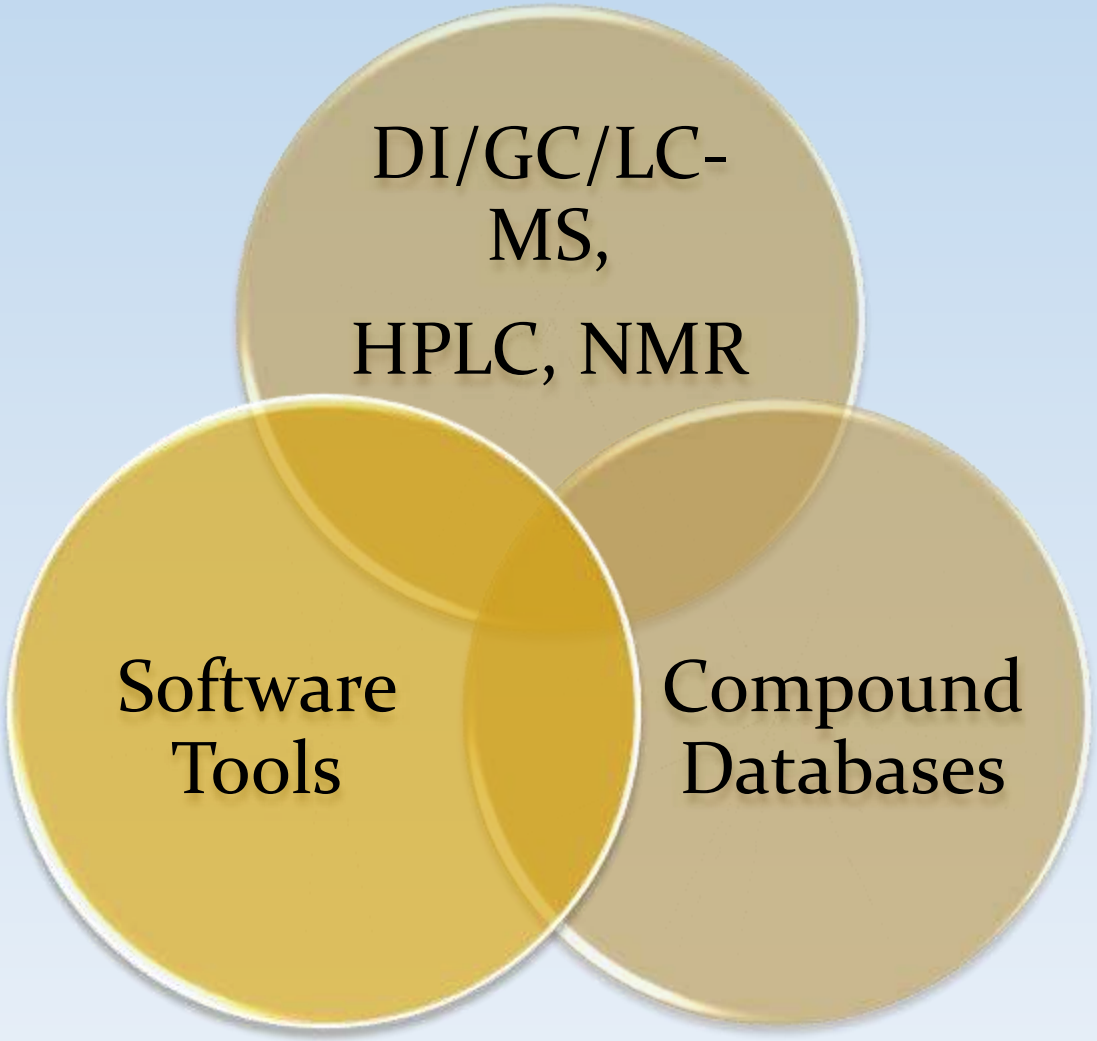
Use of web-based tools for quantitative metabolomics

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

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

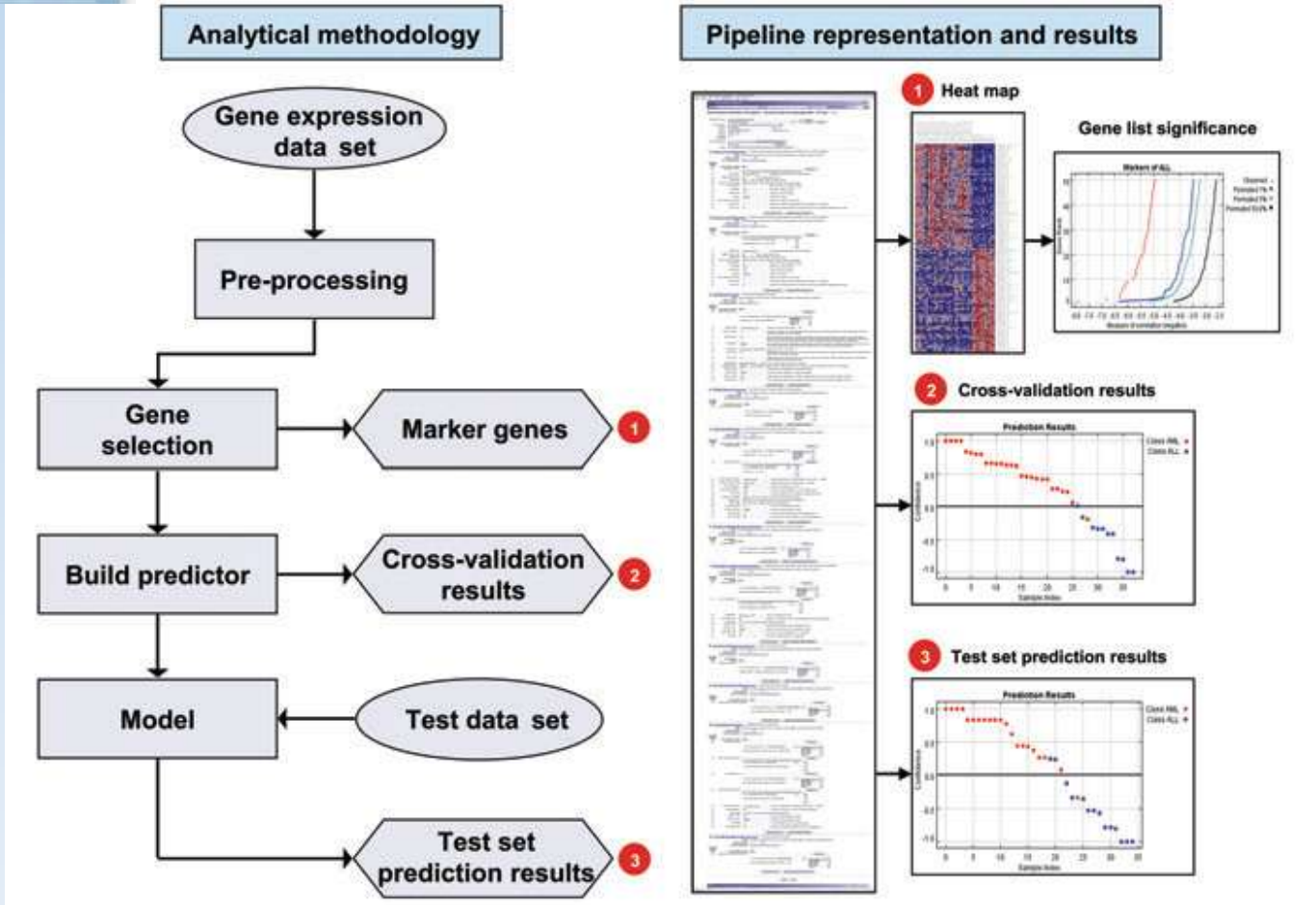
Indentify influential algorithms

[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 lusher 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

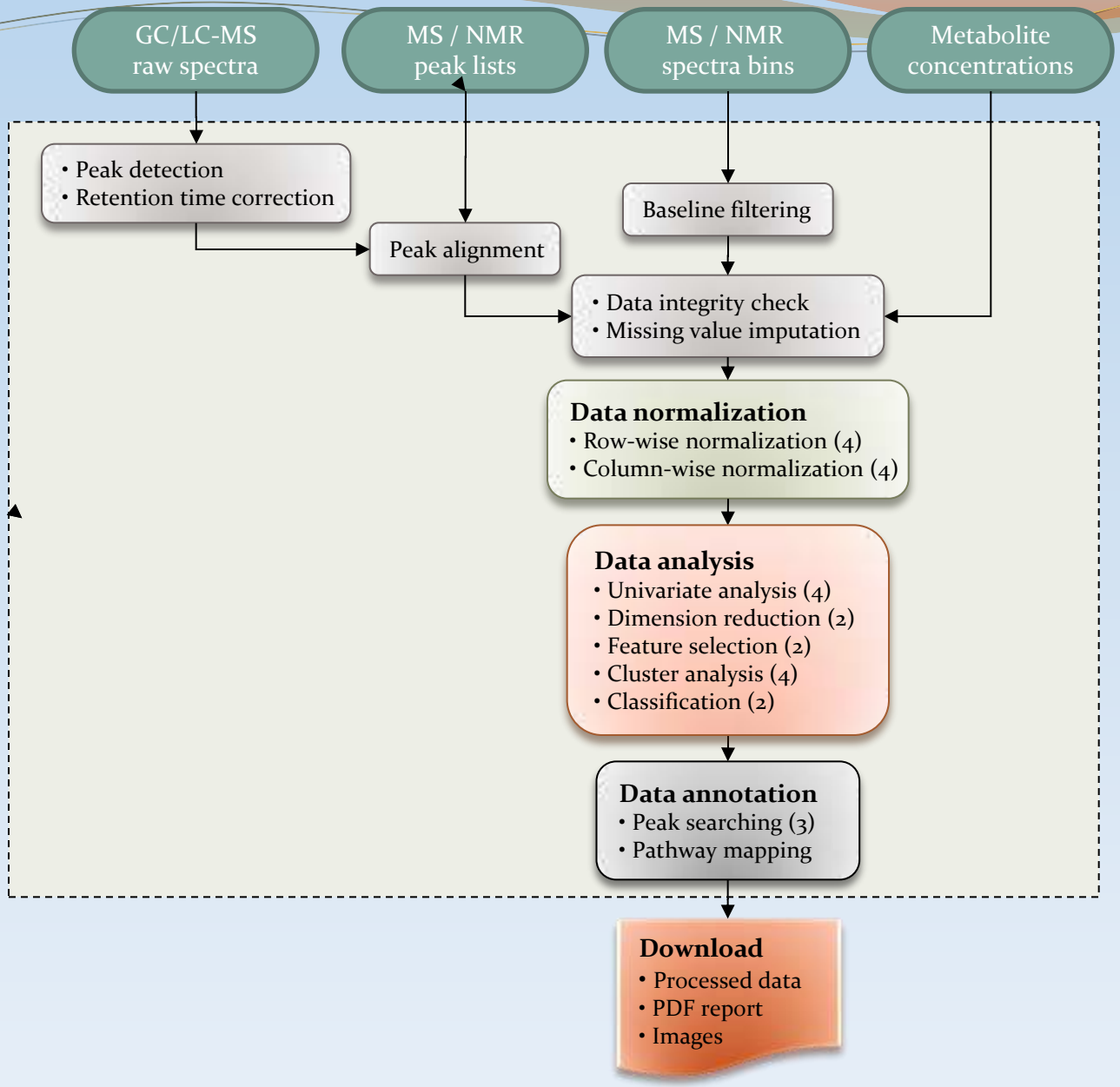
- General data processing & analysis
  - MetaboAnalyst
  - http://www.metaboanalyst.ca
- Metabolite Set Enrichment Analysis
  - MSEA
  - http://www.msea.ca
- Metabolomic Pathway Analysis
  - MetPA
  - http://metpa.metabolomics.ca

MetaboAnalyst

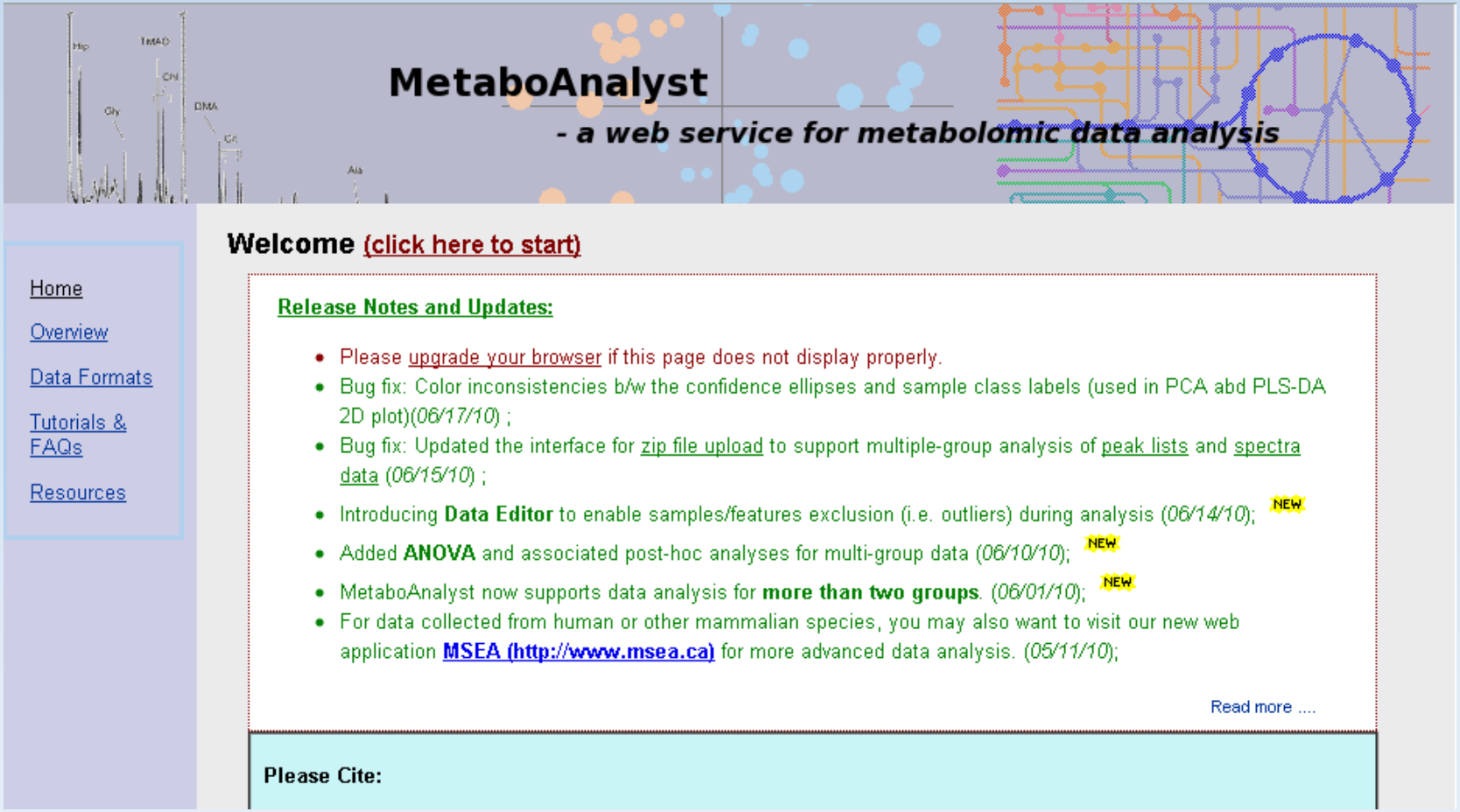
- http://www.metaboanalyst.ca
- General metabolomics data processing, normalization, and statistical analysis
  - Support two-group and multi-group analysis
  - 20+ well-established methods
  - Dynamic graphical presentation
  - Automatic report generation

What MetaboAnalyst can:

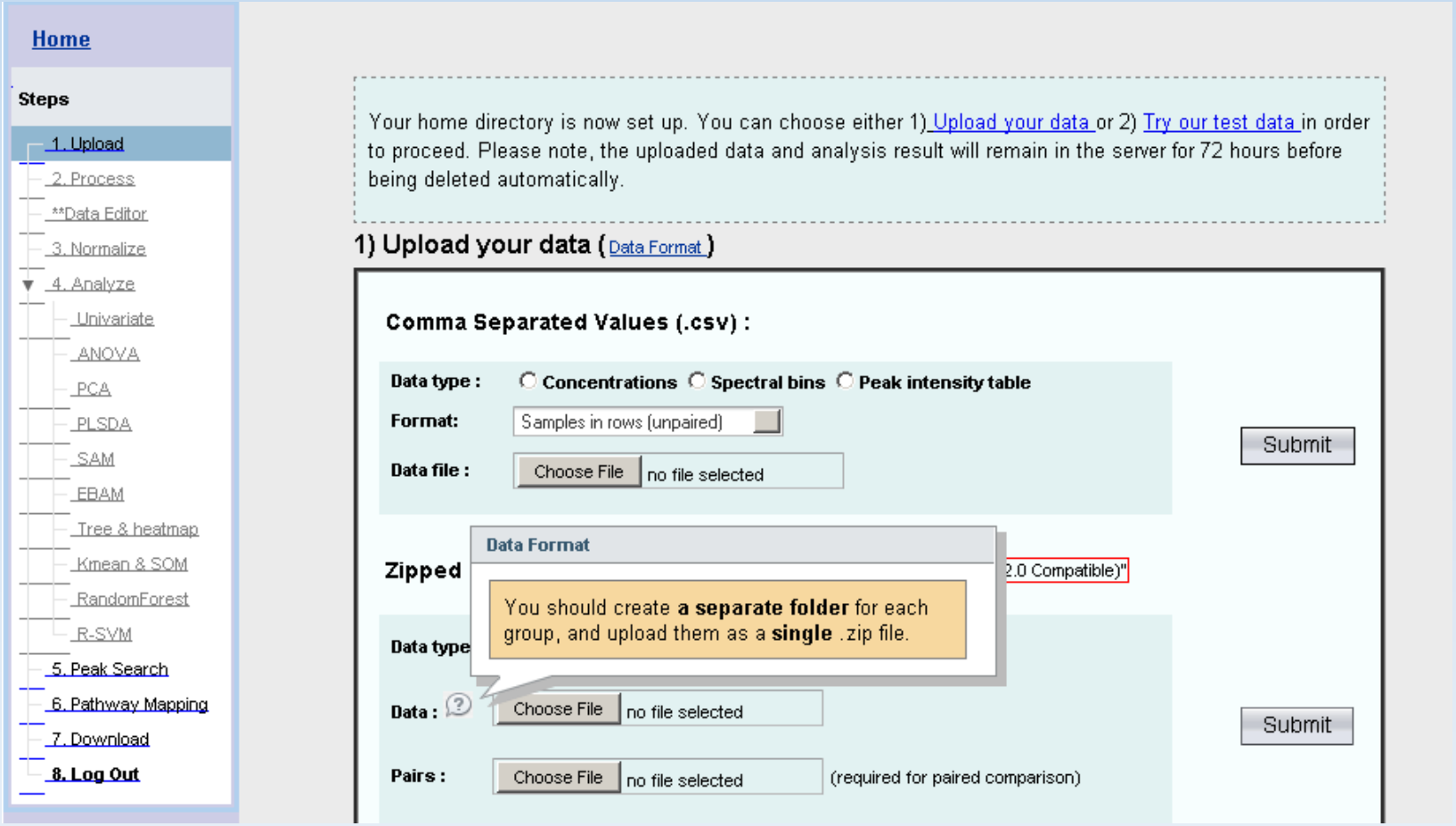
- Basic data processing:
  - Peak picking, Peak alignment, Baseline filtering, etc.
- Data normalization
  - probabilistic quotient normalization, scaling, etc.
- Data overview
  - PCA, Heatmaps, etc.
- Identify important features
  - t-tests, ANOVA, SAM, etc.
- Classification
  - PLS-DA, random Forest,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


GC/LC-MS raw spectra
MS / NMR peak lists
MS / NMR spectra bins
Metabolite concentrations
• Peak detection
• Retention time correction
Peak alignment
Baseline filtering
• Data integrity check
• Missing value imputation
Data normalization
• Row-wise normalization (4)
• Column-wise normalization (4)
Data analysis
• Univariate analysis (4)
• Dimension reduction (2)
• Feature selection (2)
• Cluster analysis (4)
• Classification (2)
Data annotation
• Peak searching (3)
• Pathway mapping
Download
• Processed data
• PDF report
• Images

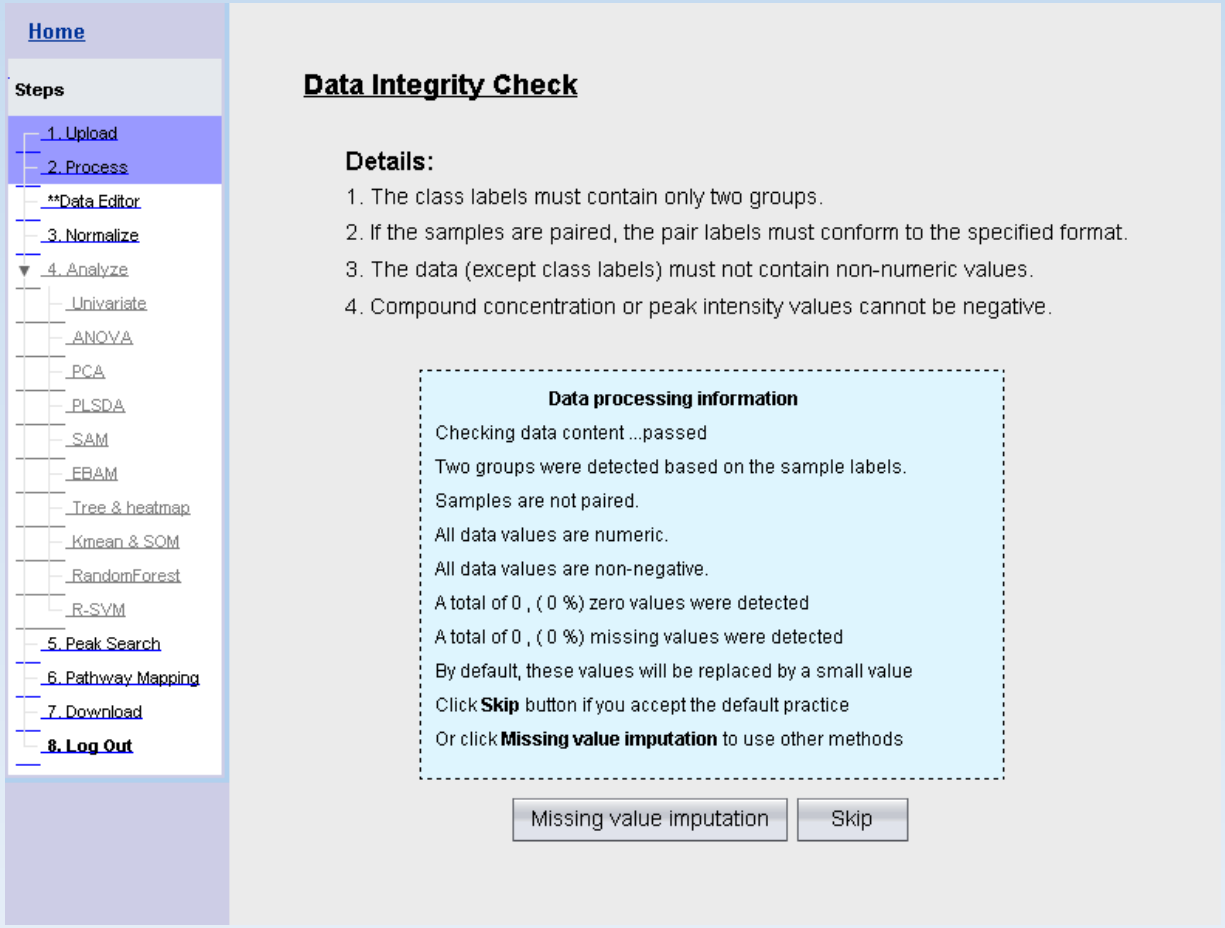
MetaboAnalyst



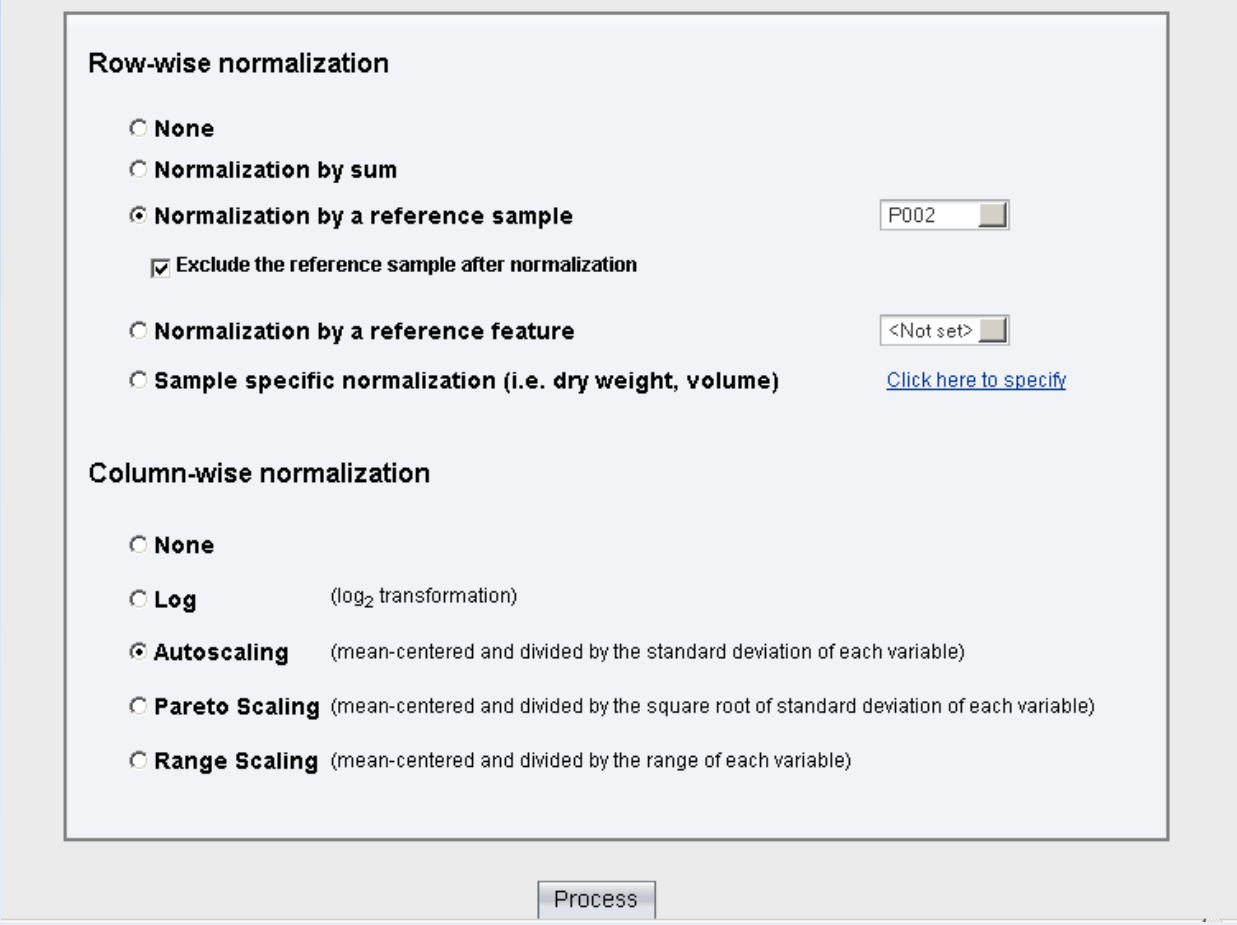
Data Upload


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: Samples in rows (unpaired)
Data file : Choose File no file selected
Submit
Zipped
Data Format
You should create a separate folder for each group, and upload them as a single .zip file.
2.0 Compatible)"
Data type
Data : Choose File no file selected
Submit
Pairs : Choose File no file selected (required for paired comparison)

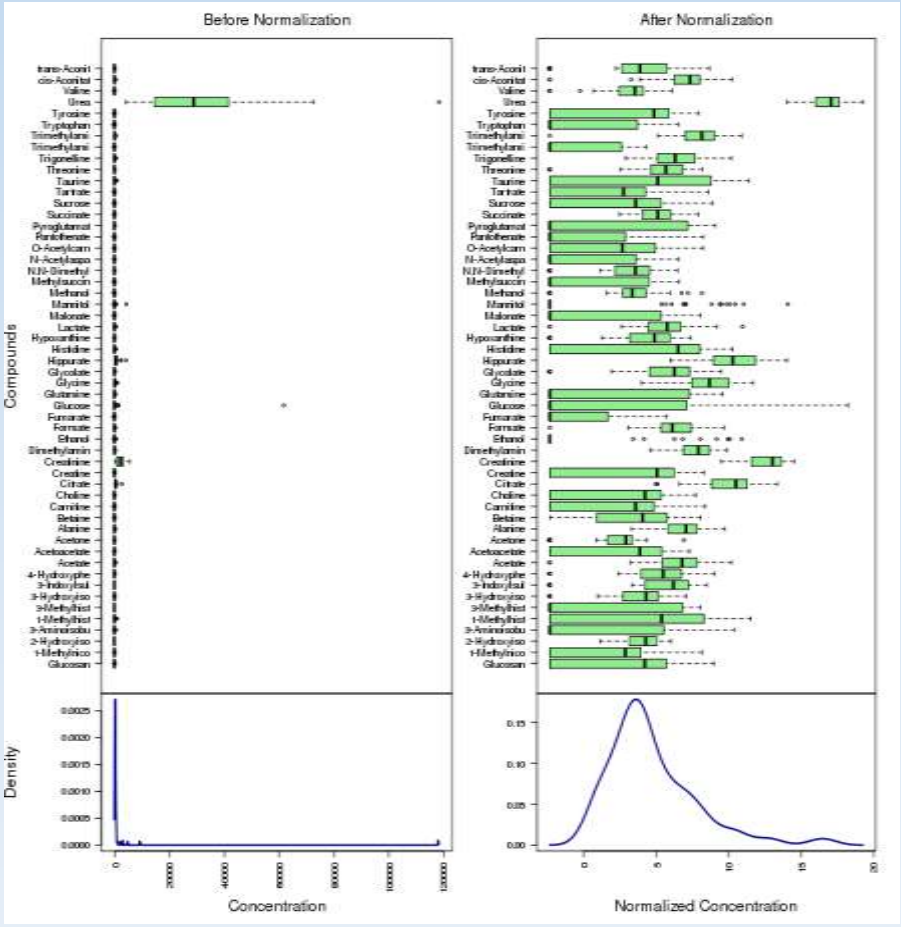
Data processing and integrity check



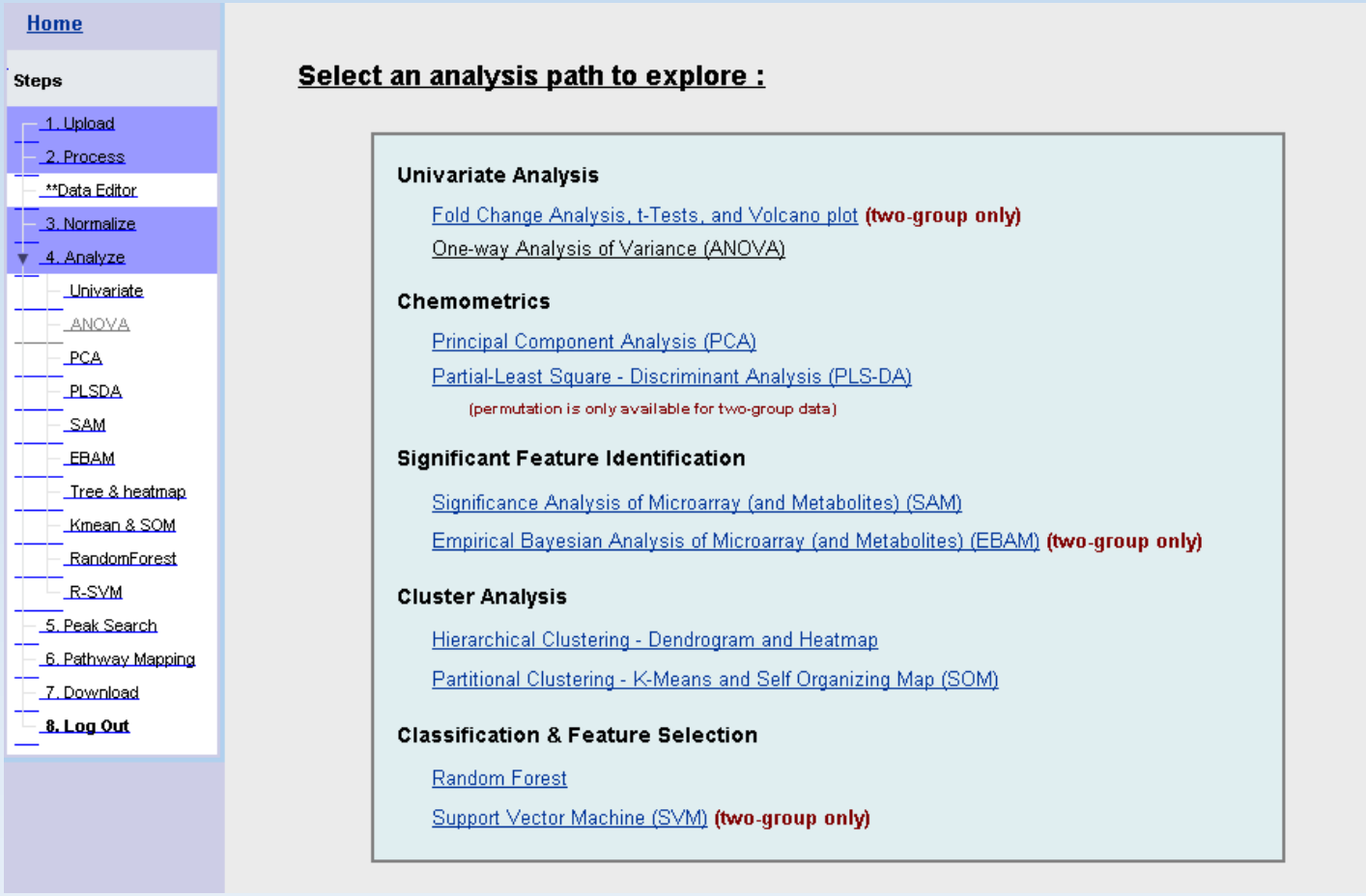
Data normalization (1)



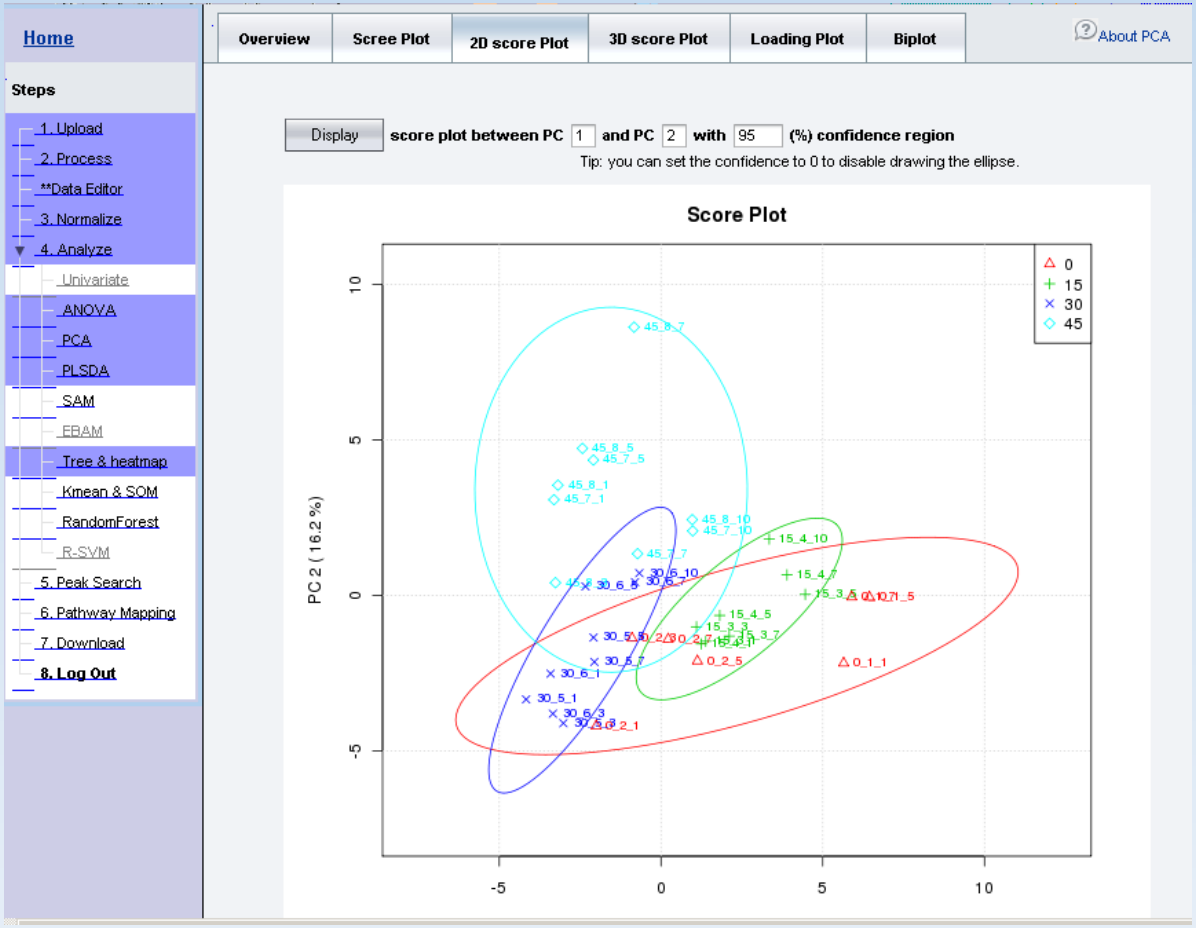
Data normalization (2)



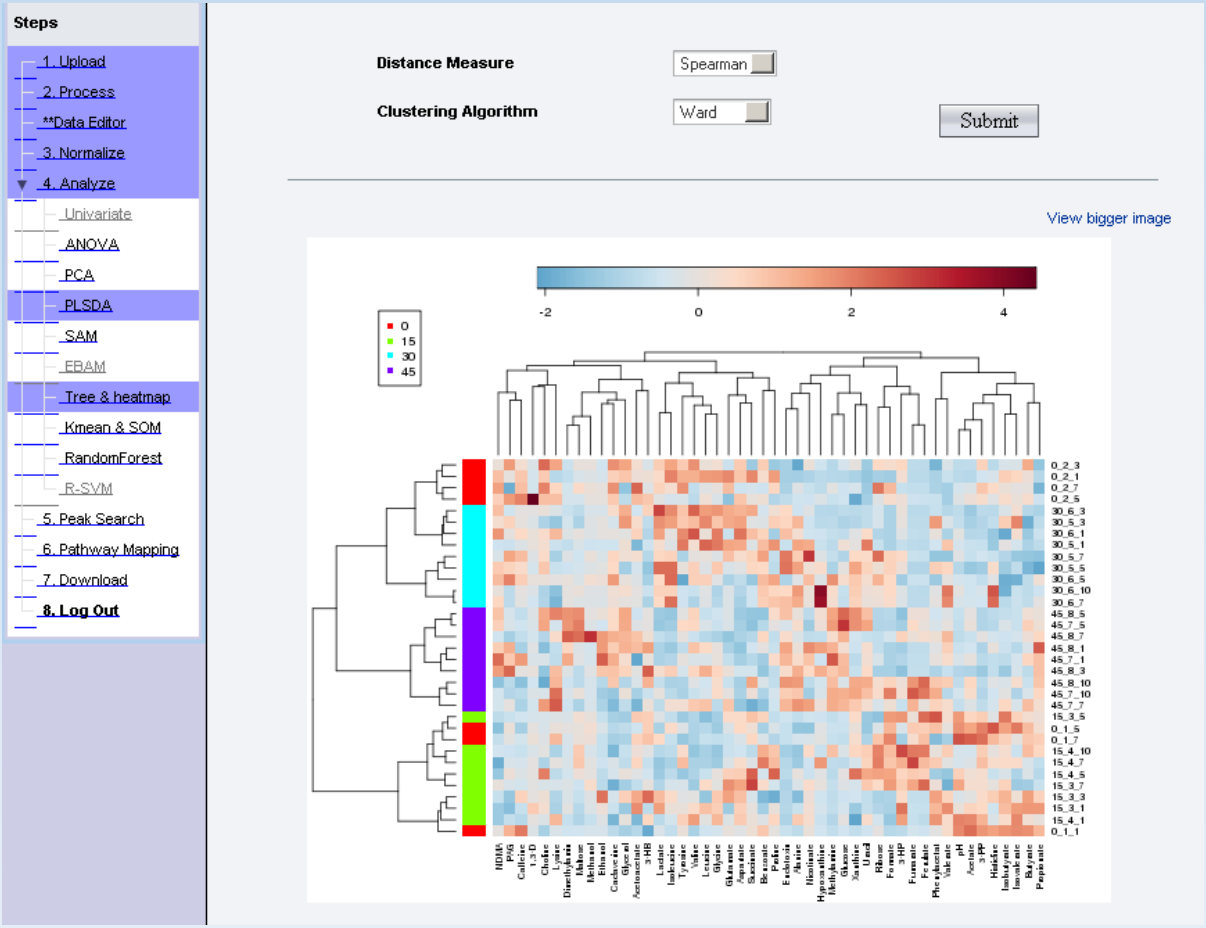
Data Analysis



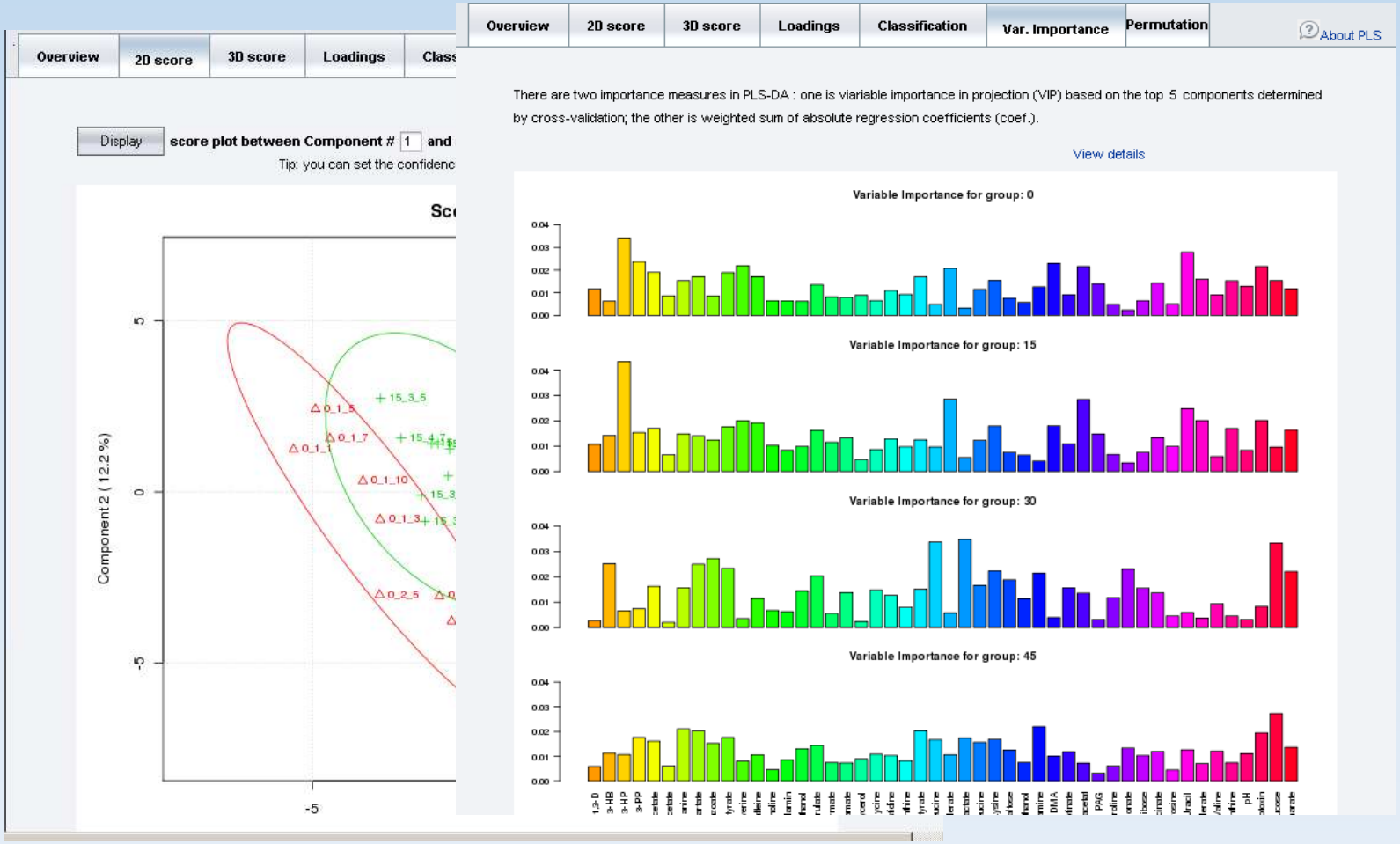
Clustering with PCA



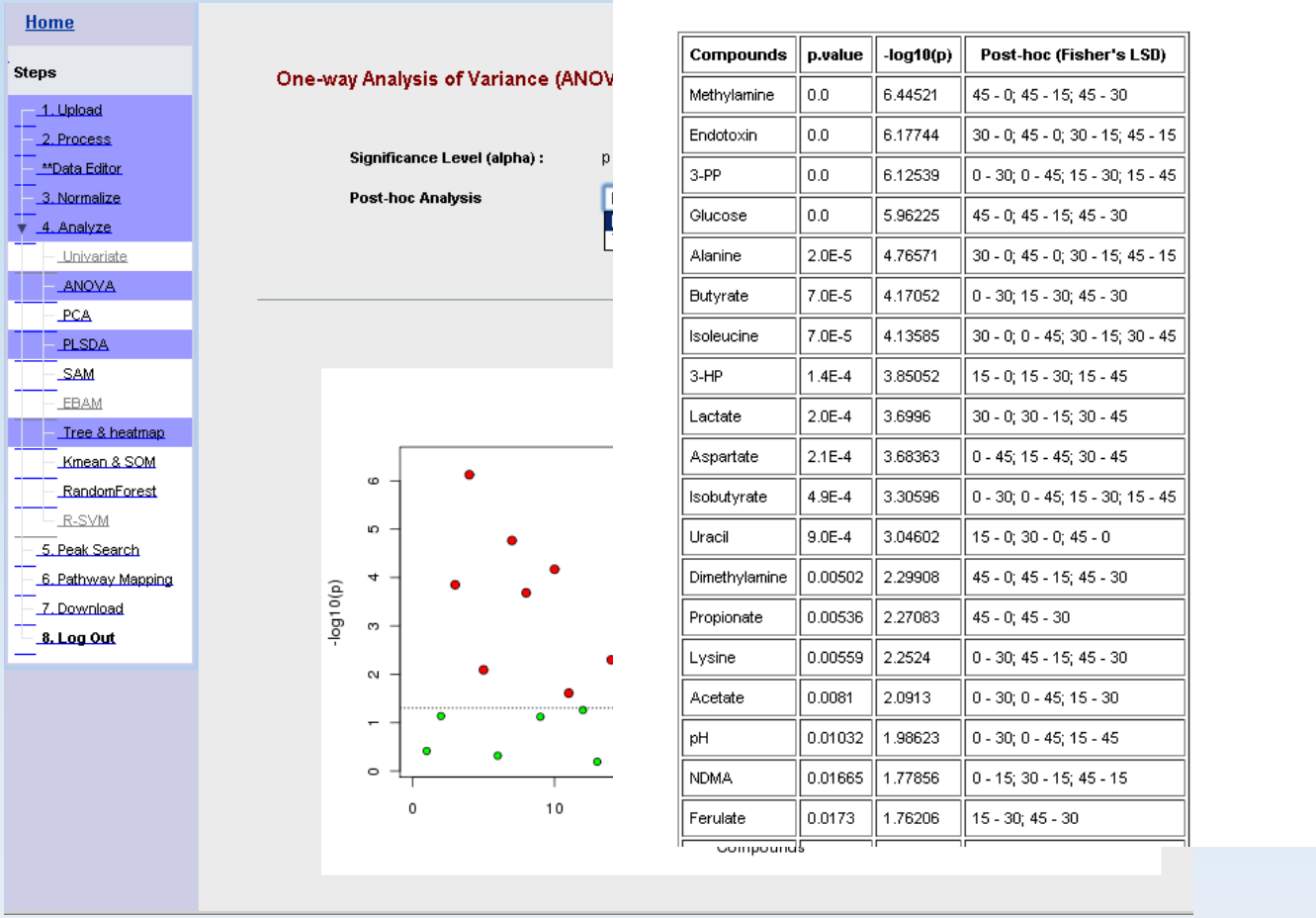
Hierarchical clustering



Supervised approach – PLS-DA

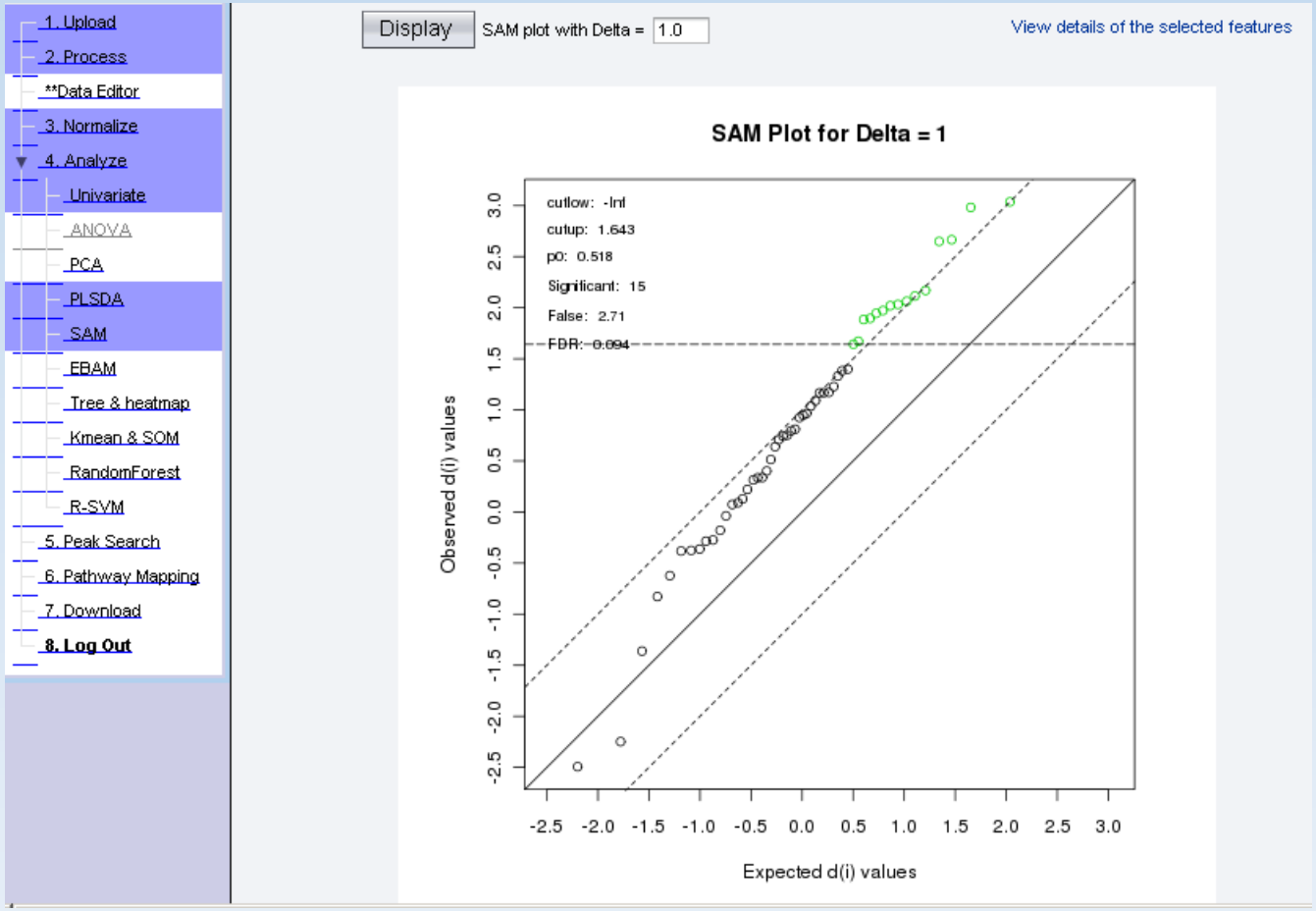


Feature selection - ANOVA



Compounds	p.value	-log10(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 Downloa

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 seperation 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.

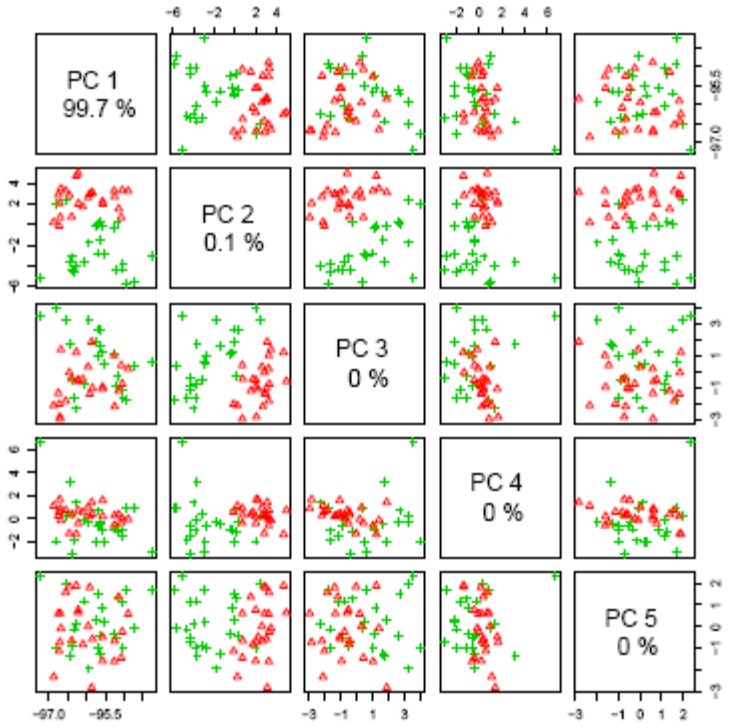


Figure 6: Pairwise score plots between the selected PCs. The explained variance of each PC is shown in the corresponding diagonal cell.


Visits
1
328

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

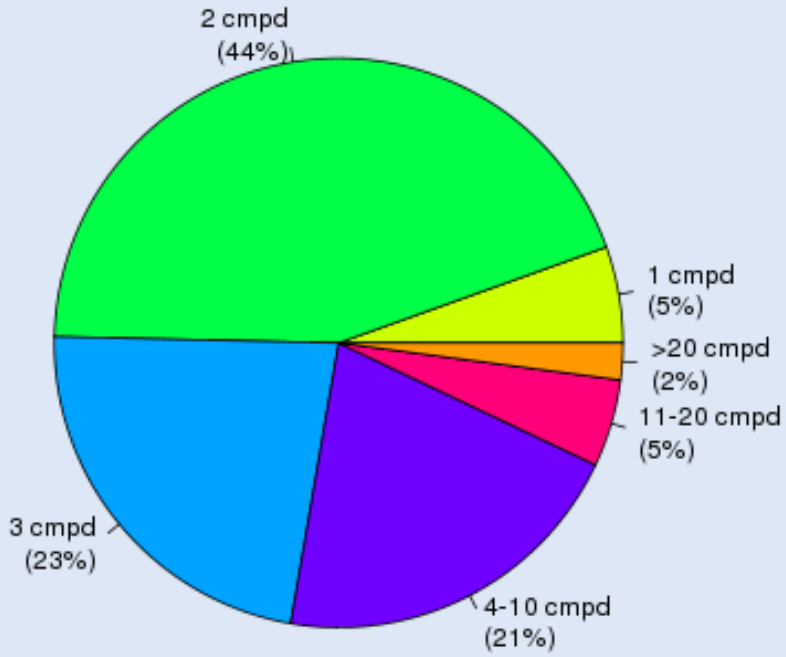
- **http://www.msea.ca**
- Metabolite Set Enrichment Analysis
- Identify **biological meaningful patterns** from quantitative metabolomics data

Biologically meaningful

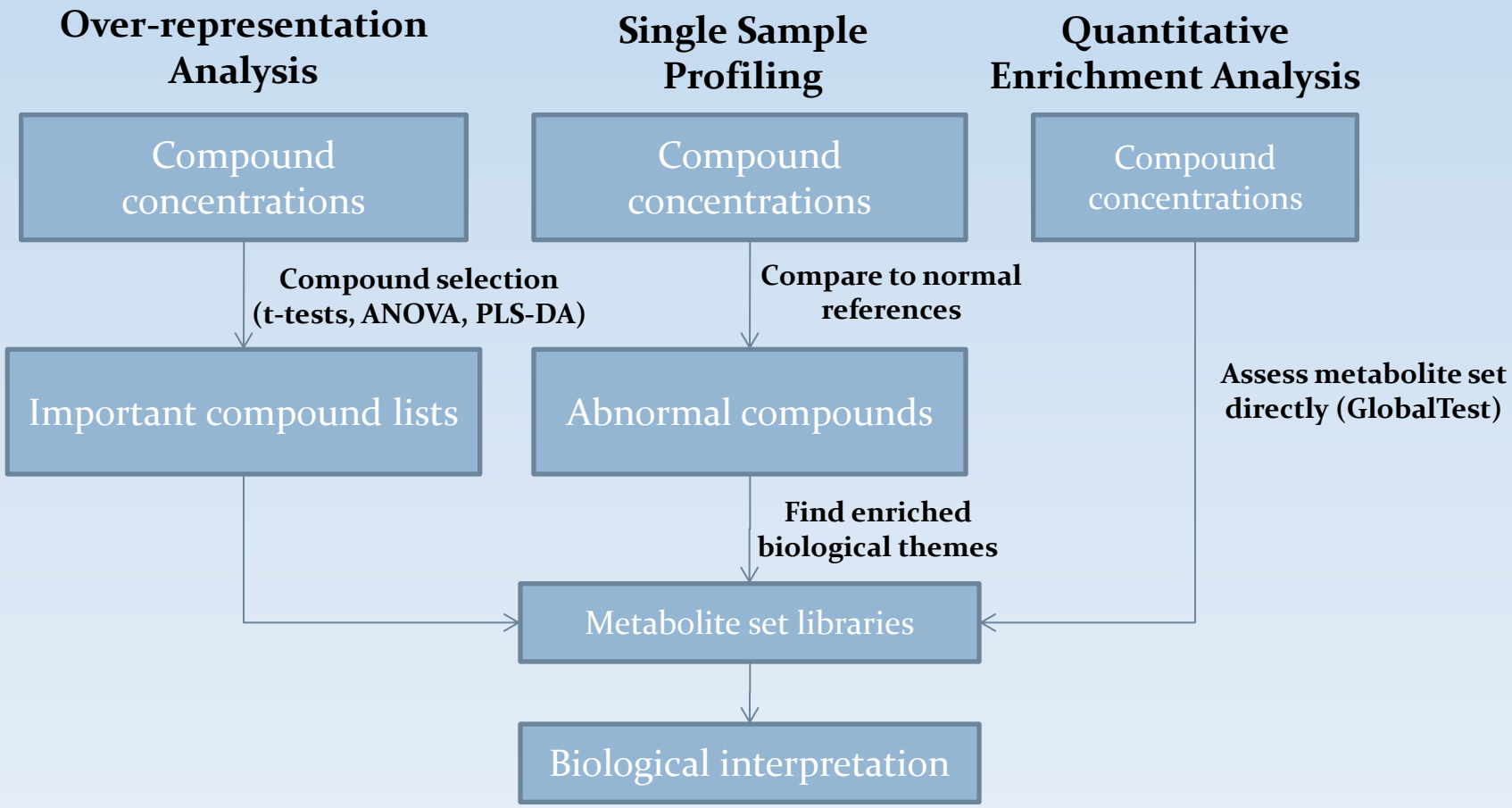
Summary of Human Metabolite Set Libraries

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

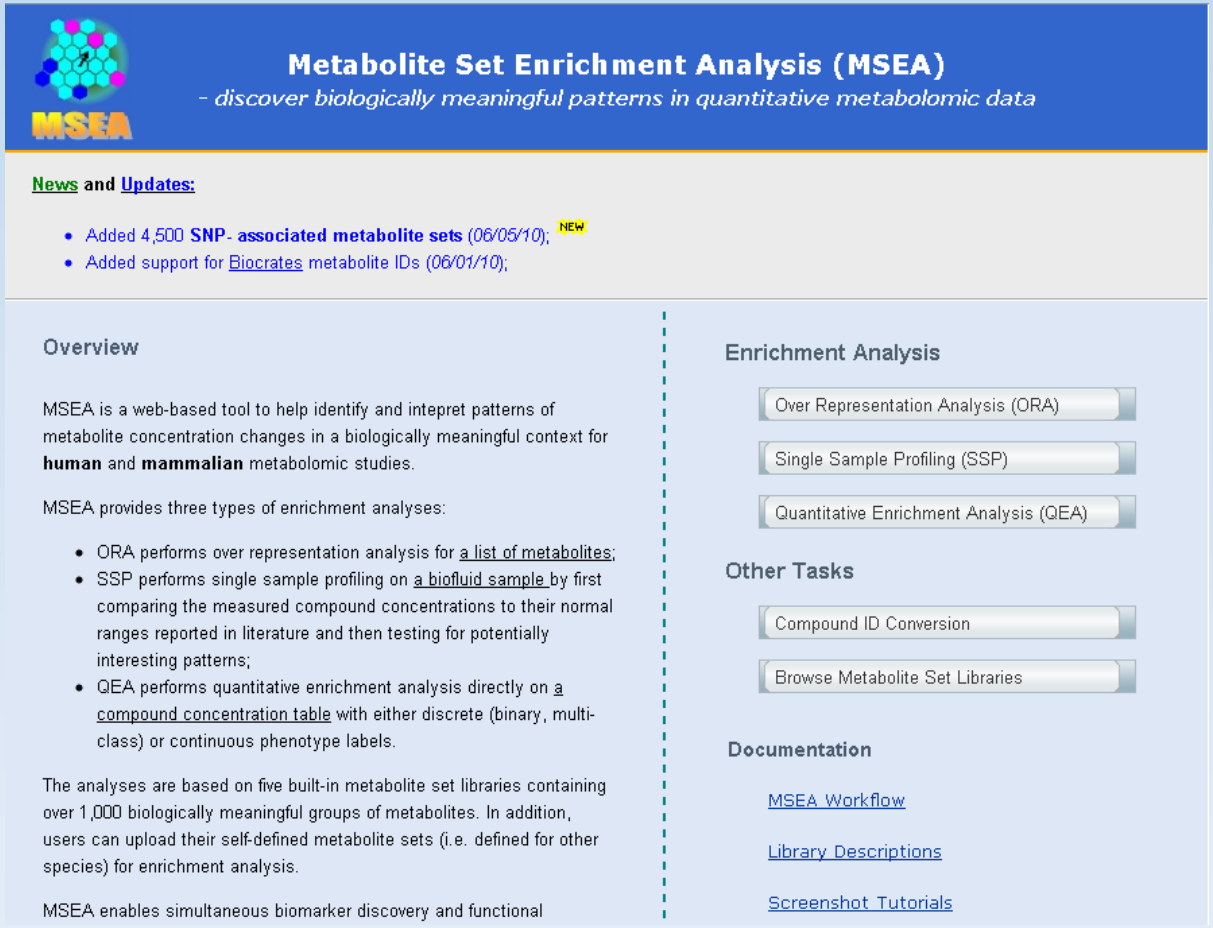
Distribution of cmpd # in metabolite sets



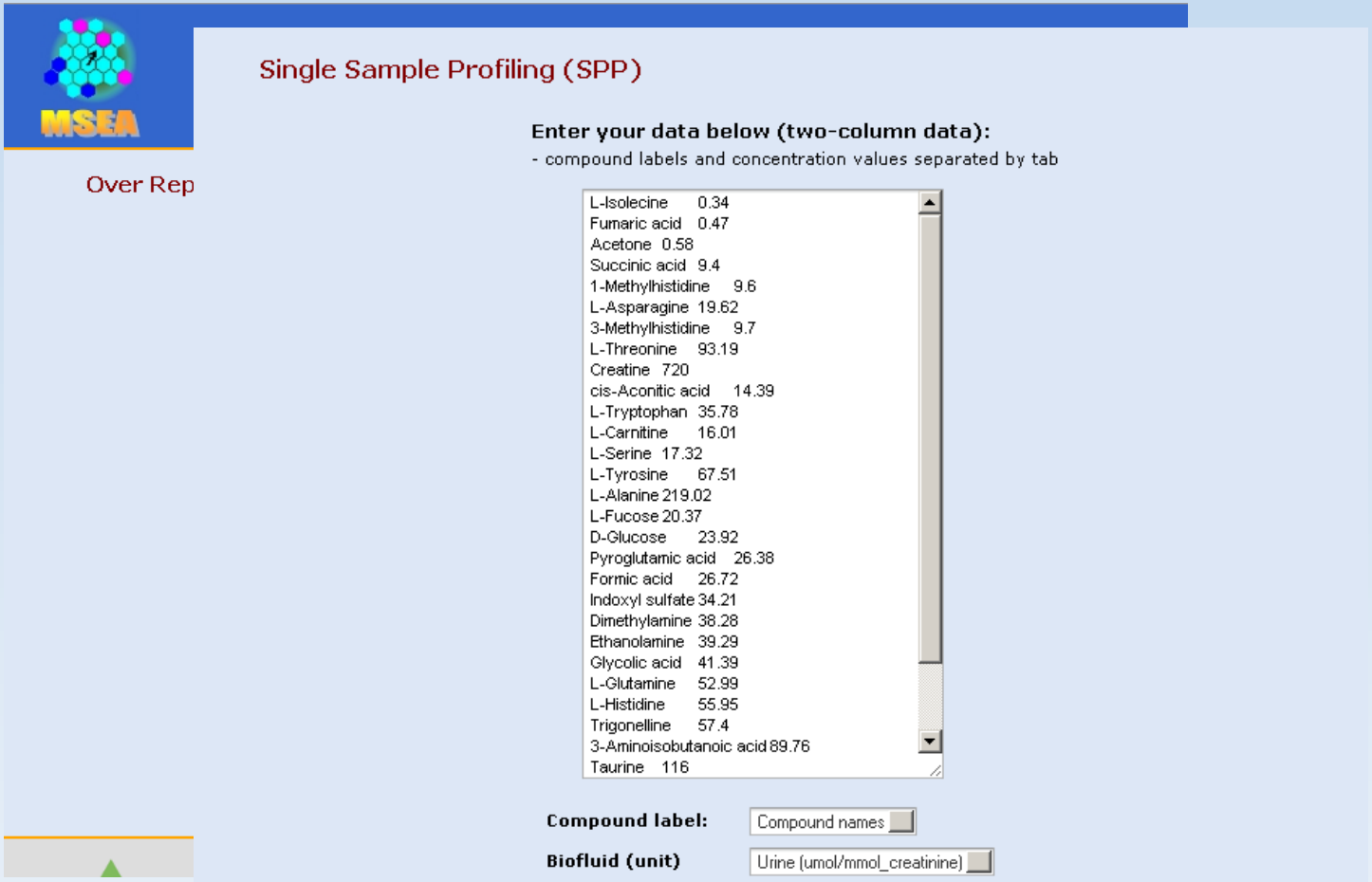
The MSEA approach



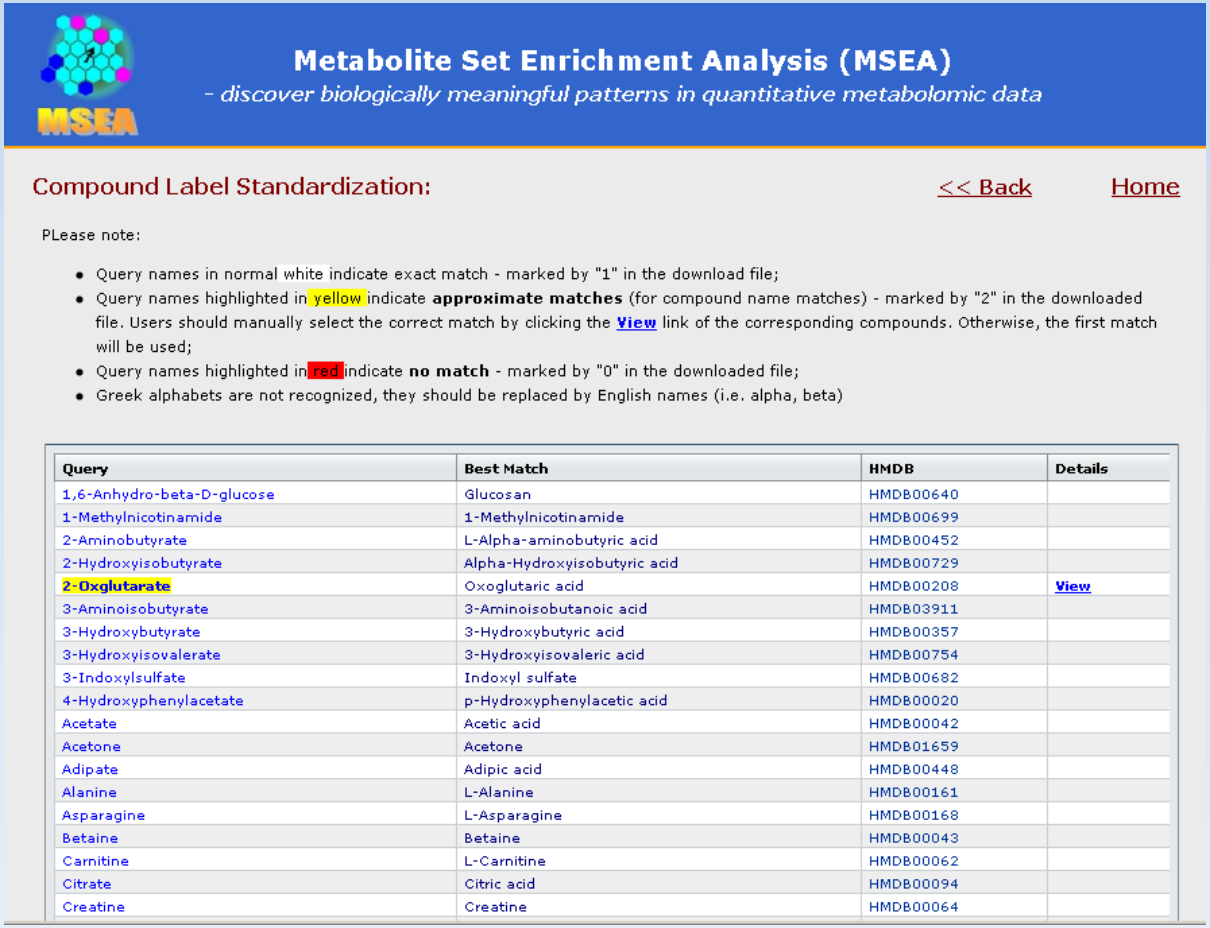
MSEA @ www.msea.ca



Over-representation analysis



Compound label standardization



Metabolite Set Enrichment Analysis (MSEA)

- discover biologically meaningful patterns in quantitative metabolomic data

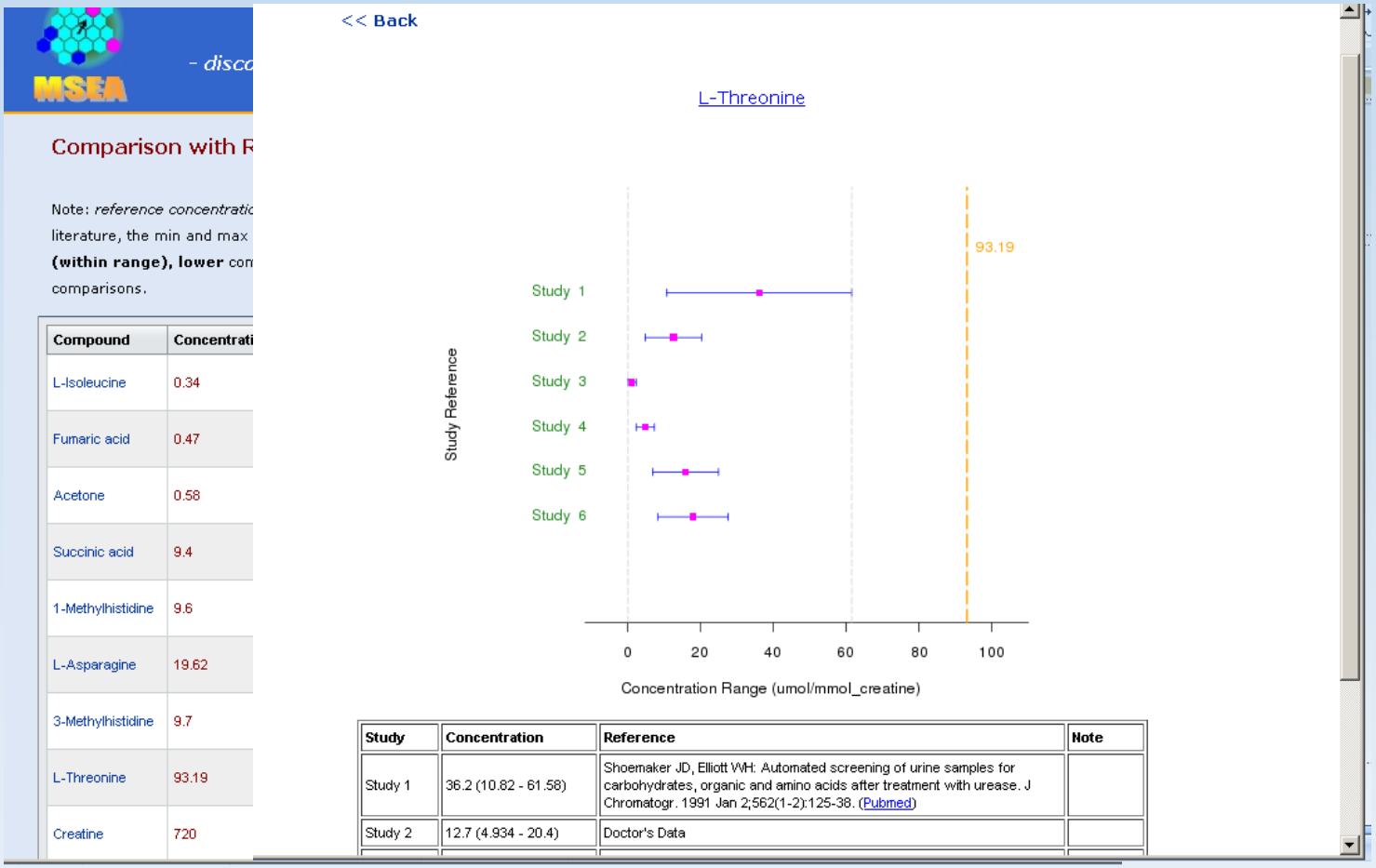
Compound Label Standardization: << Back Home

PLease note:

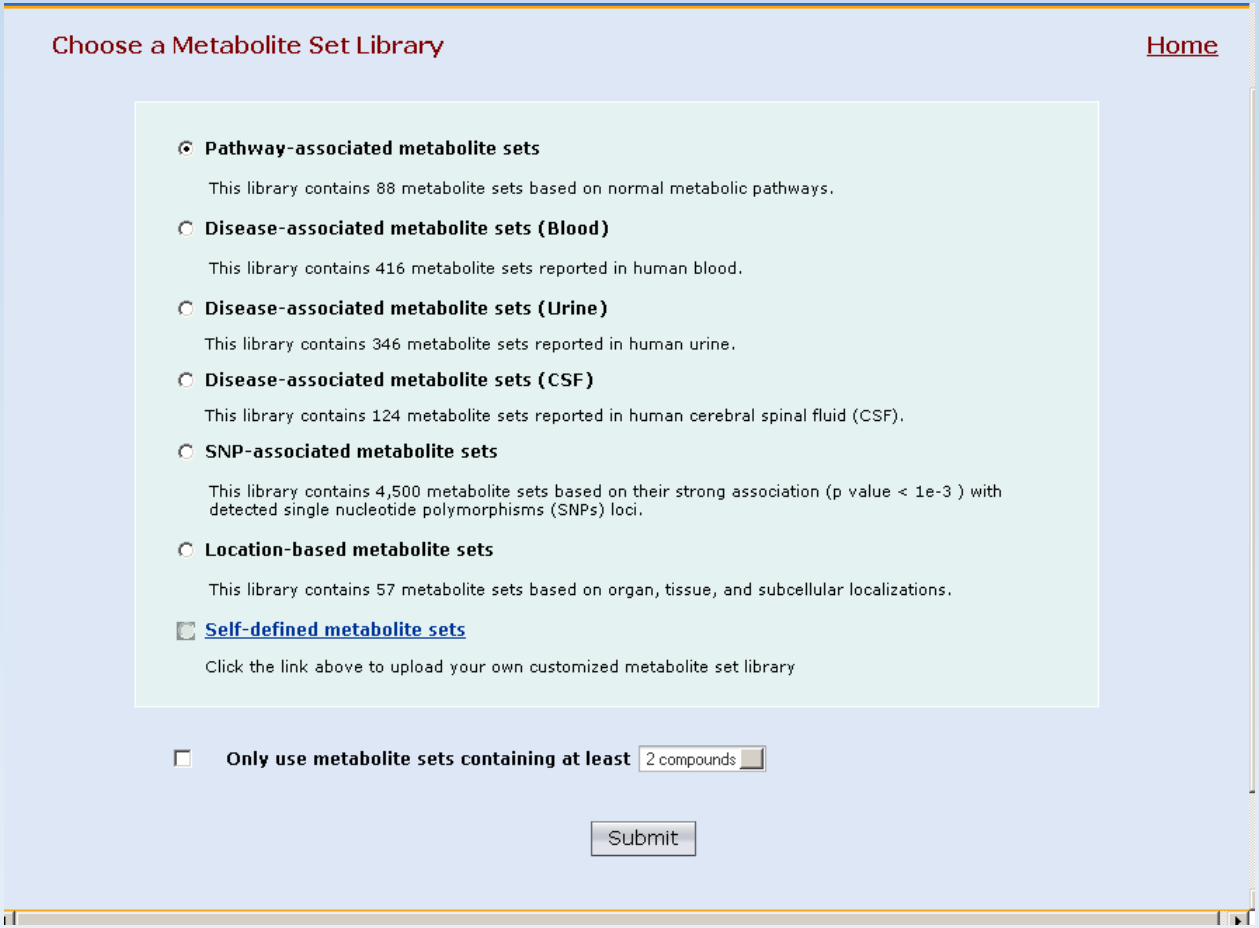
- Query names in normal white indicate exact match - marked by "1" in the download file;
- Query names highlighted in yellow indicate **approximate matches** (for compound name matches) - marked by "2" in the downloaded file. Users should manually select the correct match by clicking the **View** link 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
1,6-Anhydro-beta-D-glucose	Glucosan	HMDB00640	
1-Methylnicotinamide	1-Methylnicotinamide	HMDB00699	
2-Aminobutyrate	L-Alpha-aminobutyric acid	HMDB00452	
2-Hydroxyisobutyrate	Alpha-Hydroxyisobutyric acid	HMDB00729	
2-Oxglutarate	Oxoglutaric acid	HMDB00208	View
3-Aminoisobutyrate	3-Aminoisobutanoic acid	HMDB03911	
3-Hydroxybutyrate	3-Hydroxybutyric acid	HMDB00357	
3-Hydroxyisovalerate	3-Hydroxyisovaleric acid	HMDB00754	
3-Indoxylsulfate	Indoxyl sulfate	HMDB00682	
4-Hydroxyphenylacetate	p-Hydroxyphenylacetic acid	HMDB00020	
Acetate	Acetic acid	HMDB00042	
Acetone	Acetone	HMDB01659	
Adipate	Adipic acid	HMDB00448	
Alanine	L-Alanine	HMDB00161	
Asparagine	L-Asparagine	HMDB00168	
Betaine	Betaine	HMDB00043	
Carnitine	L-Carnitine	HMDB00062	
Citrate	Citric acid	HMDB00094	
Creatine	Creatine	HMDB00064	

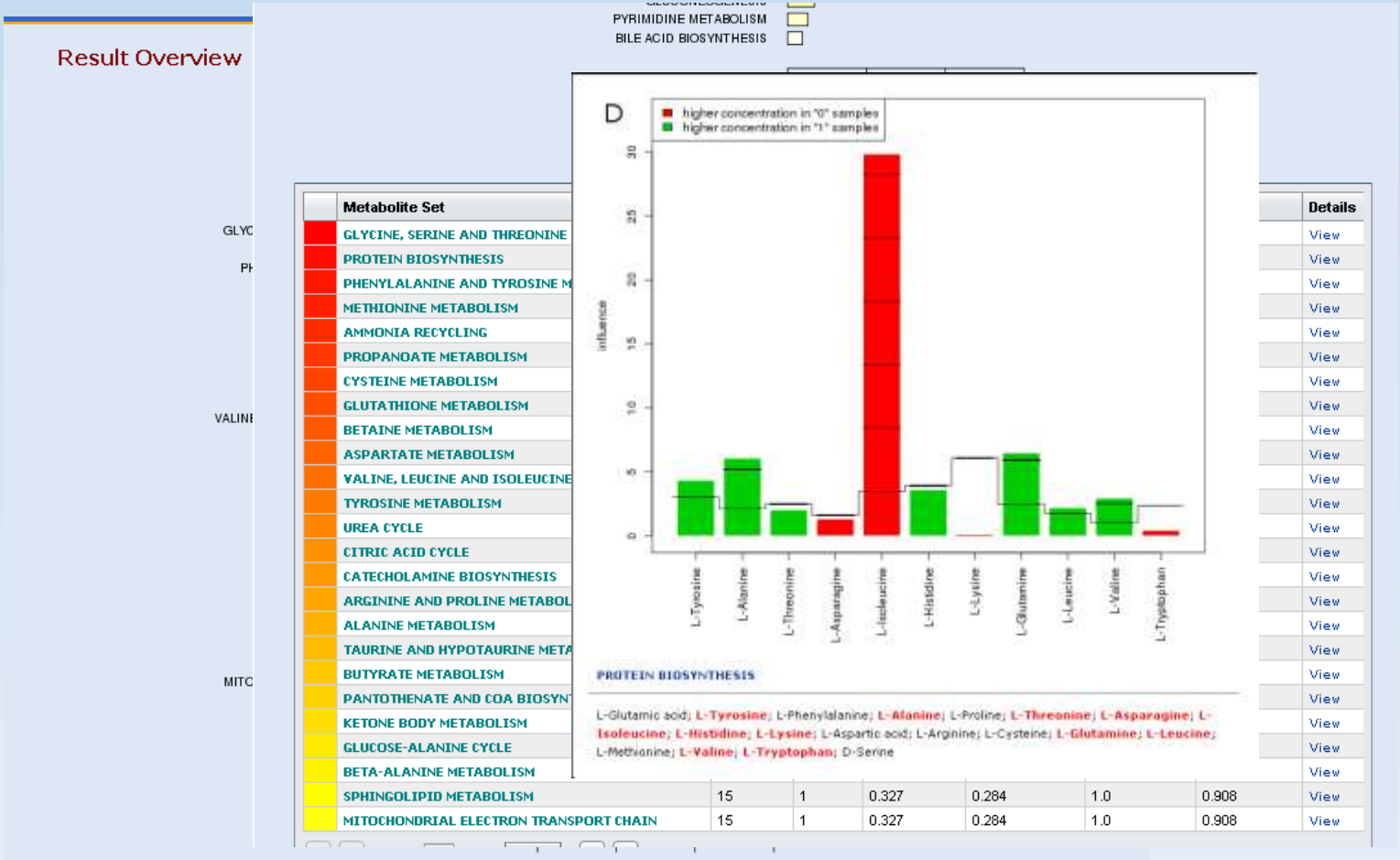
Identify abnormal concentration



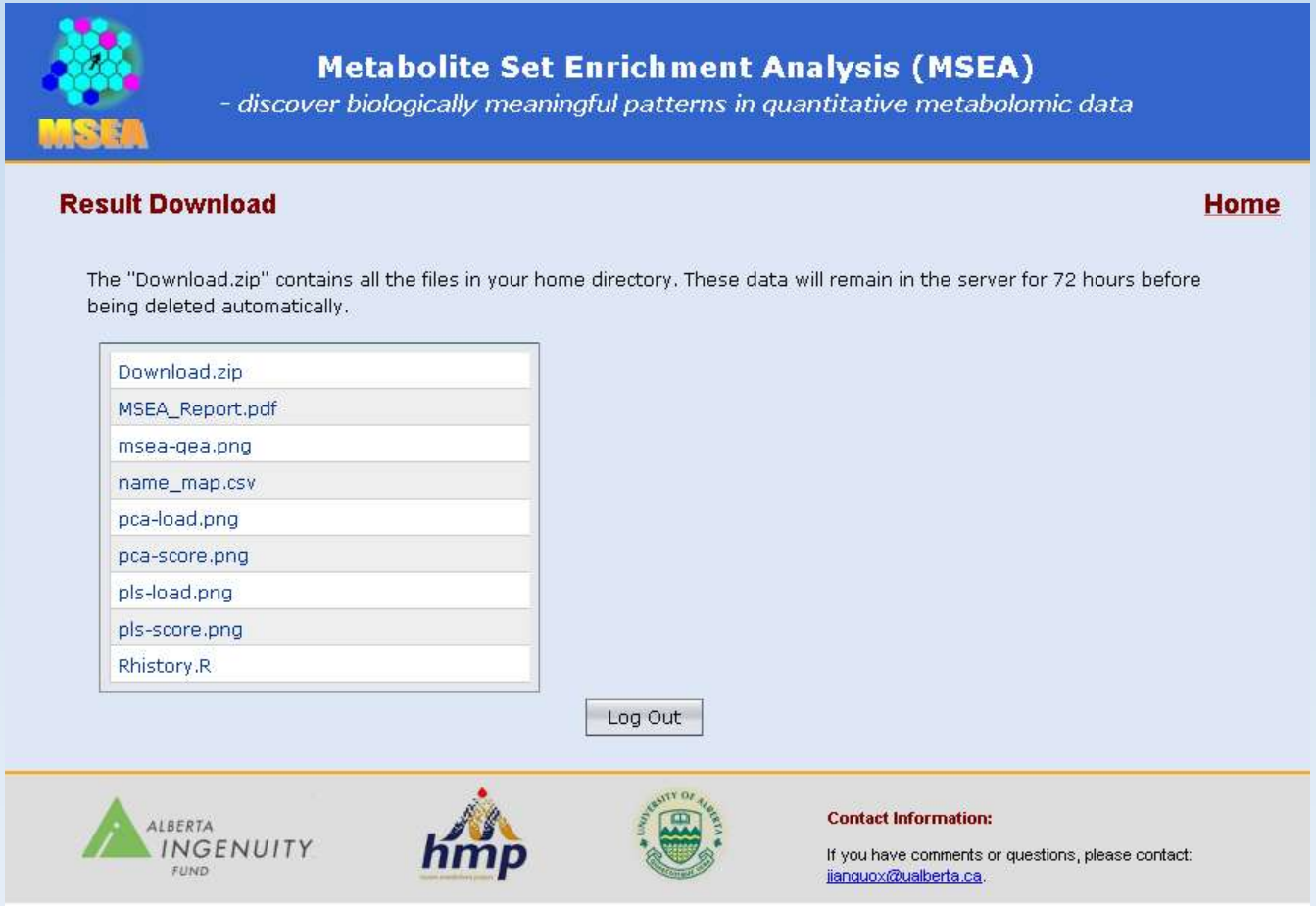
Library selection



Result


Result Overview
PYRIMIDINE METABOLISM
BILE ACID BIOSYNTHESIS
D
higher concentration in "0" samples
higher concentration in "1" samples
influence
L-Tyrosine
L-Alanine
L-Threonine
L-Asparagine
L-Isoleucine
L-Histidine
L-Lysine
L-Glutamine
L-Leucine
L-Valine
L-Tryptophan
PROTEIN BIOSYNTHESIS
L-Glutamic acid; L-Tyrosine; L-Phenylalanine; L-Alanine; L-Proline; L-Threonine; L-Asparagine; L-Isoleucine; L-Histidine; L-Lysine; L-Aspartic acid; L-Arginine; L-Cysteine; L-Glutamine; L-Leucine; L-Methionine; L-Valine; L-Tryptophan; D-Serine
Metabolite Set
Details
GLYCINE, SERINE AND THREONINE
PROTEIN BIOSYNTHESIS
PHENYLALANINE AND TYROSINE M
METHIONINE METABOLISM
AMMONIA RECYCLING
PROPANOATE METABOLISM
CYSTEINE METABOLISM
GLUTATHIONE METABOLISM
BETAINE METABOLISM
ASPARTATE METABOLISM
VALINE, LEUCINE AND ISOLEUCINE
TYROSINE METABOLISM
UREA CYCLE
CITRIC ACID CYCLE
CATECHOLAMINE BIOSYNTHESIS
ARGININE AND PROLINE METABOL
ALANINE METABOLISM
TAURINE AND HYPOTAURINE META
BUTYRATE METABOLISM
PANTOTHENATE AND COA BIOSYNT
KETONE BODY METABOLISM
GLUCOSE-ALANINE CYCLE
BETA-ALANINE METABOLISM
SPHINGOLIPID METABOLISM 15 1 0.327 0.284 1.0 0.908 View
MITOCHONDRIAL ELECTRON TRANSPORT CHAIN 15 1 0.327 0.284 1.0 0.908 View

Download

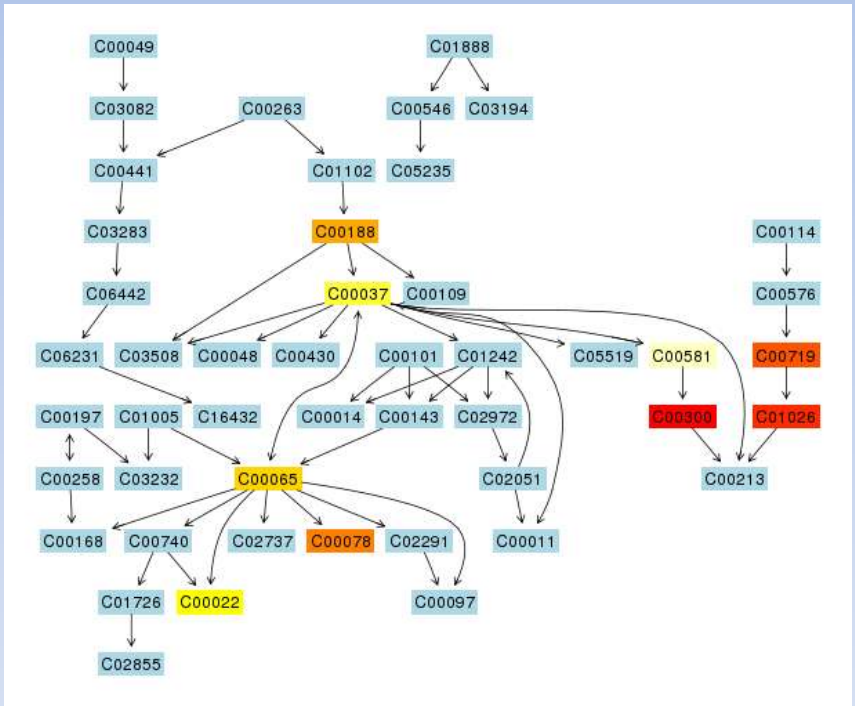

Metabolite Set Enrichment Analysis (MSEA)
- discover biologically meaningful patterns in quantitative metabolomic data
Result Download
Home
The "Download.zip" contains all the files in your home directory. These data will remain in the server for 72 hours before being deleted automatically.
Download.zip
MSEA_Report.pdf
msea-qea.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.

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

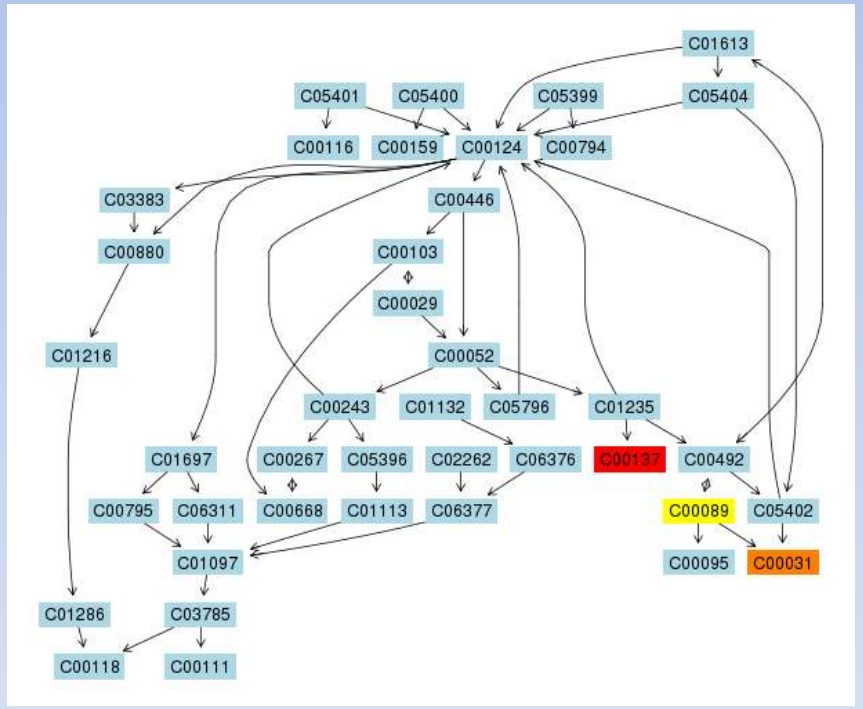
Topology matters

Glycine, serine and theonine metabolism



p = 1e-5

Galactose metabolism

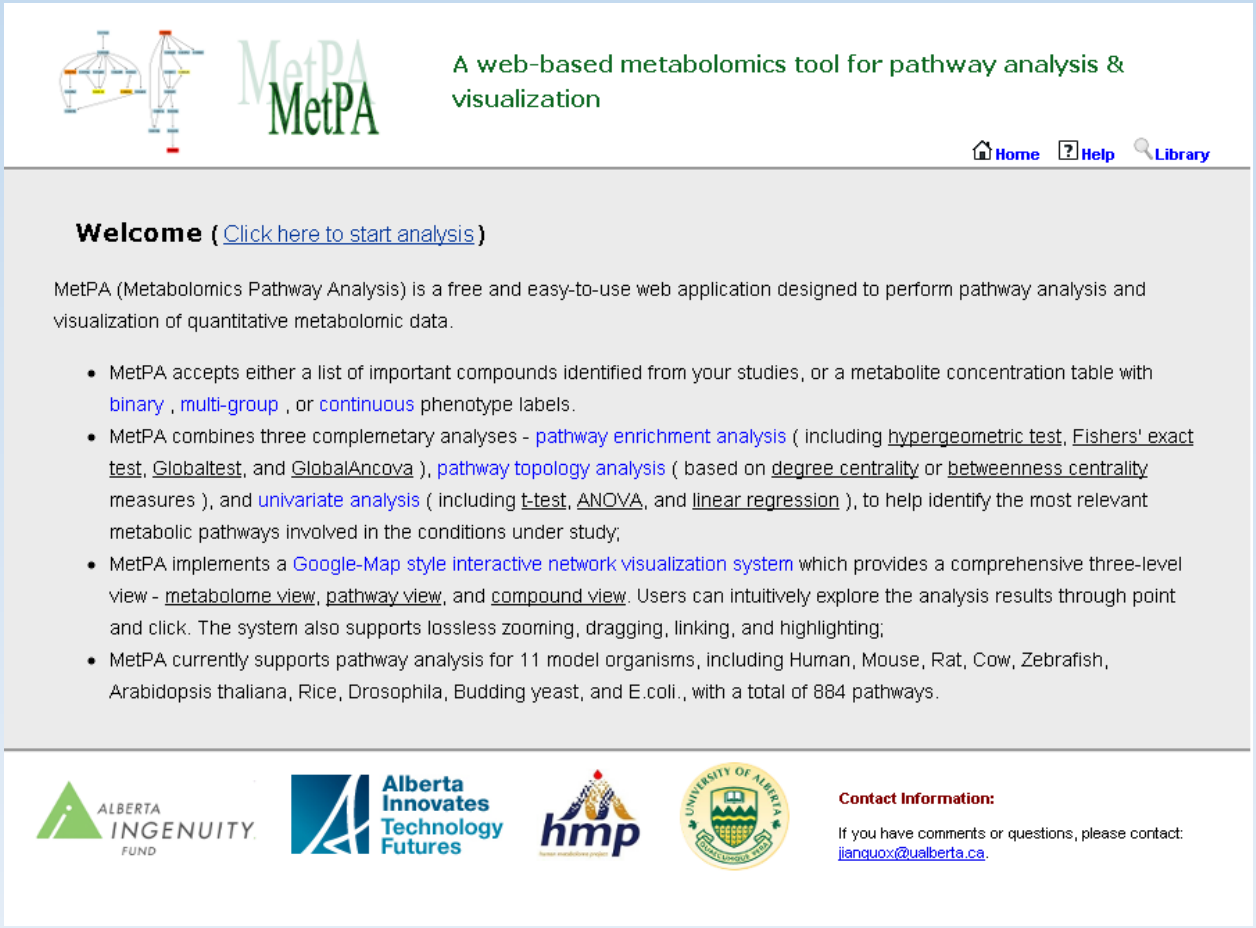


p = 1e-7

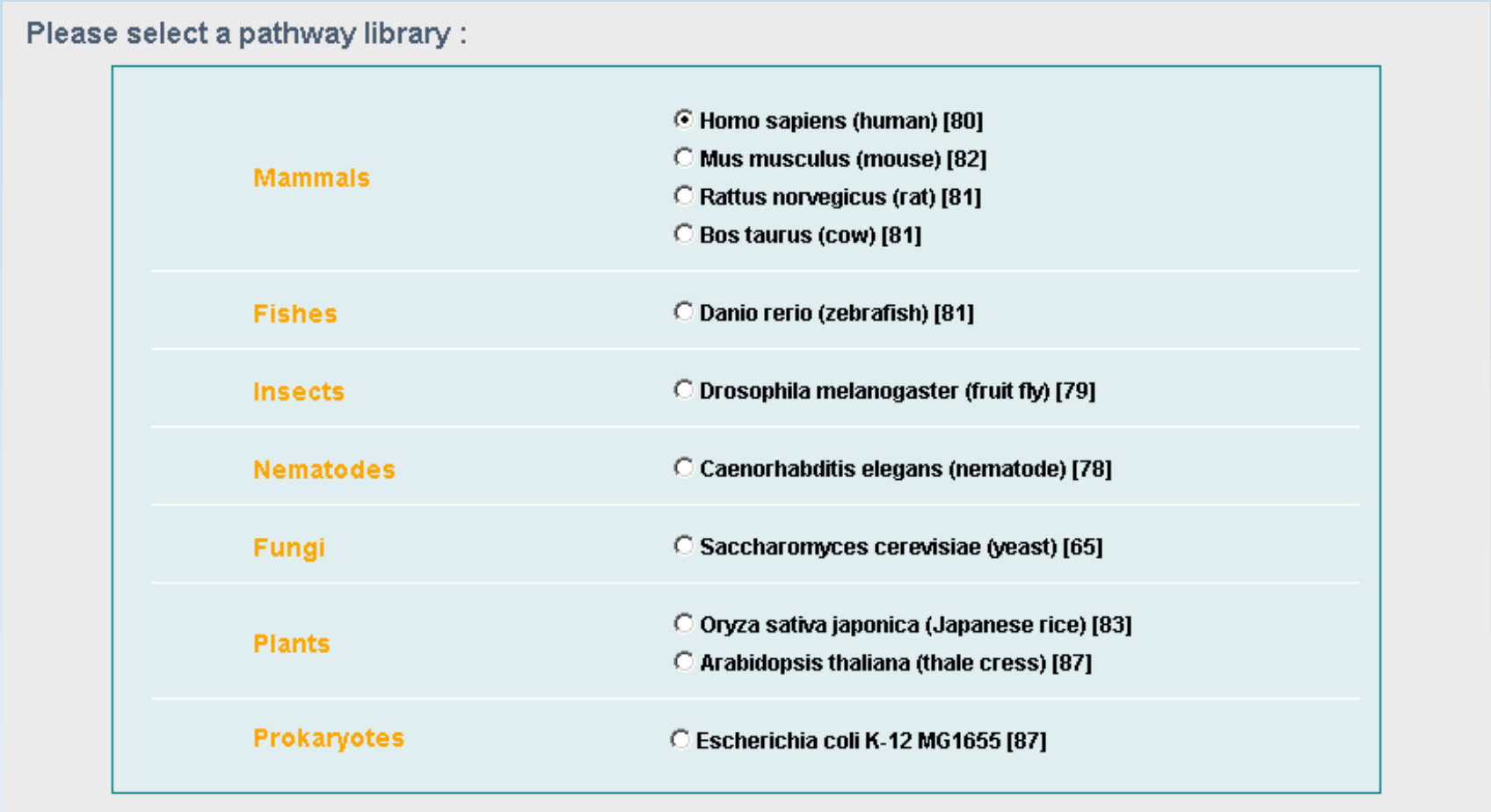
Introducing MetPA

- **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

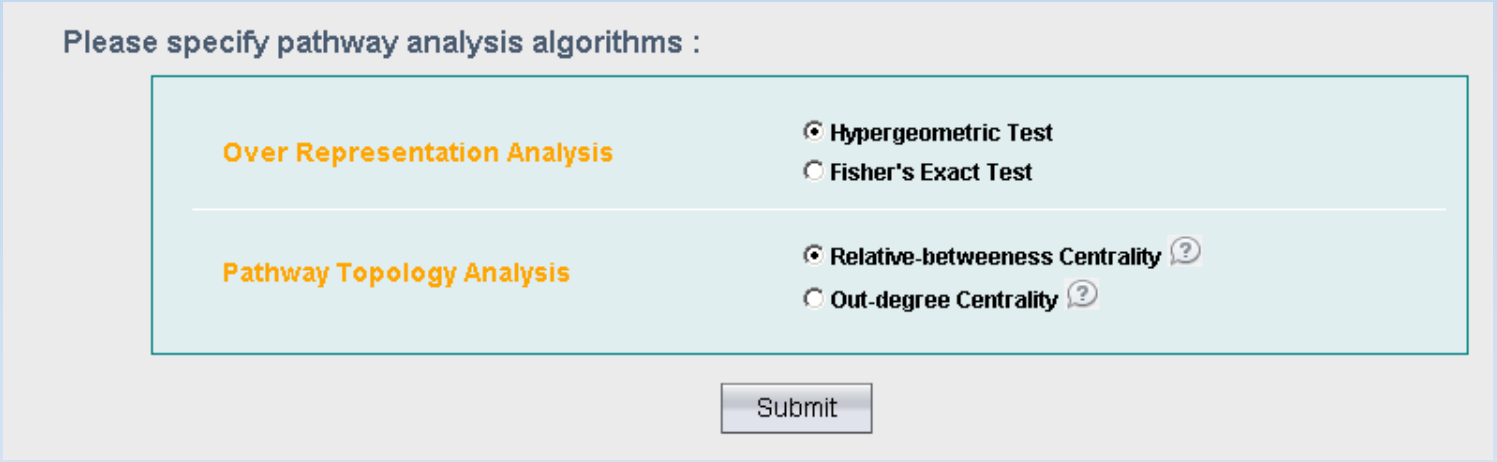
MetPA

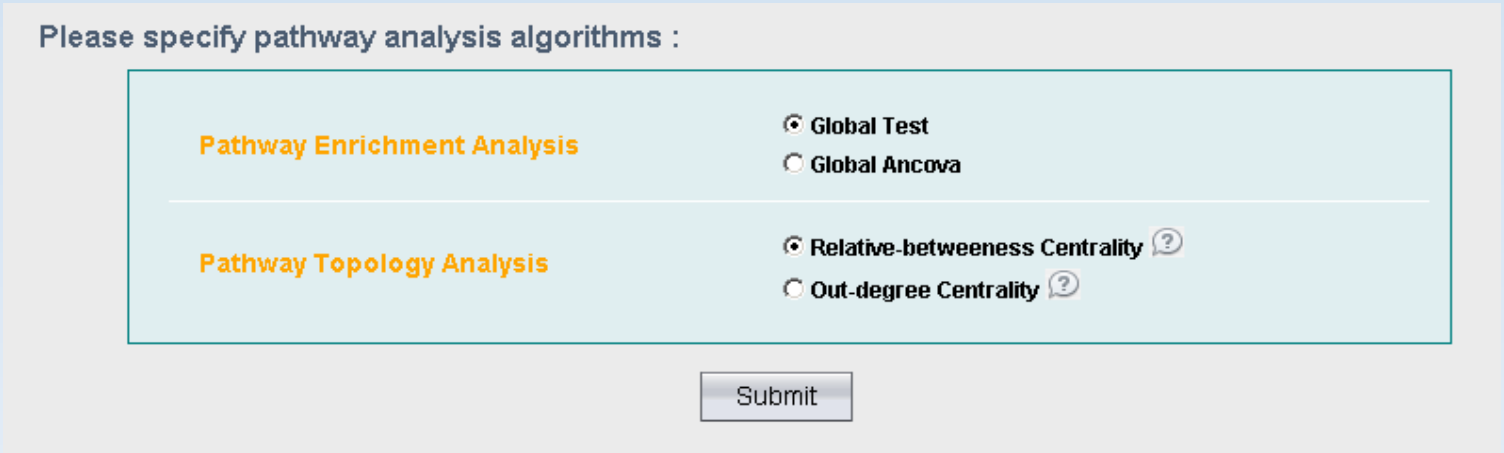

MetPA
MetPA
A web-based metabolomics tool for pathway analysis & visualization
Home
Help
Library
Welcome (Click here to start analysis)
MetPA (Metabolomics Pathway Analysis) is a free and easy-to-use web application designed to perform pathway analysis and visualization of quantitative metabolomic data.
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 complemetary 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.
ALBERTA INGENUITY FUND
Alberta Innovates Technology Futures
hmp
Contact Information:
If you have comments or questions, please contact:
jianguox@ualberta.ca.

11 pathway libraries (KEGG)



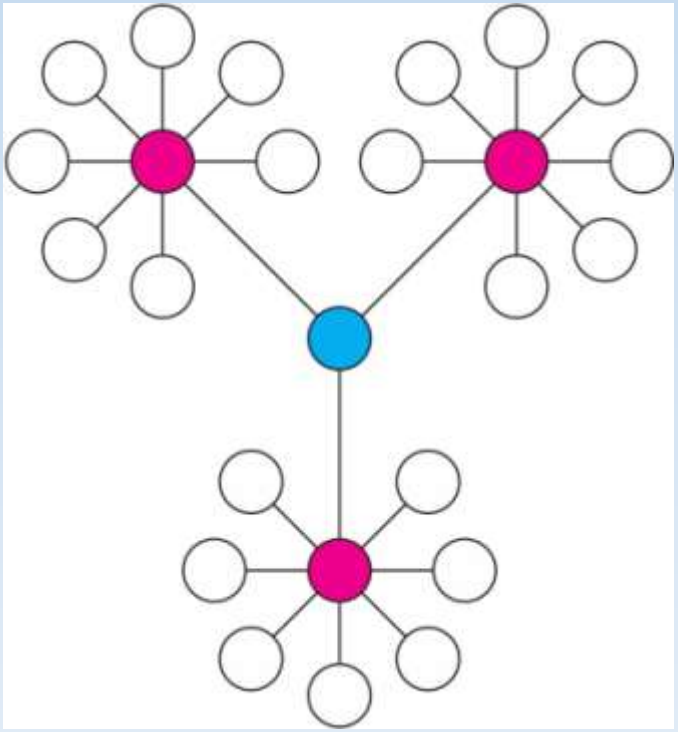
Combining Enrichment analysis & Topology analysis





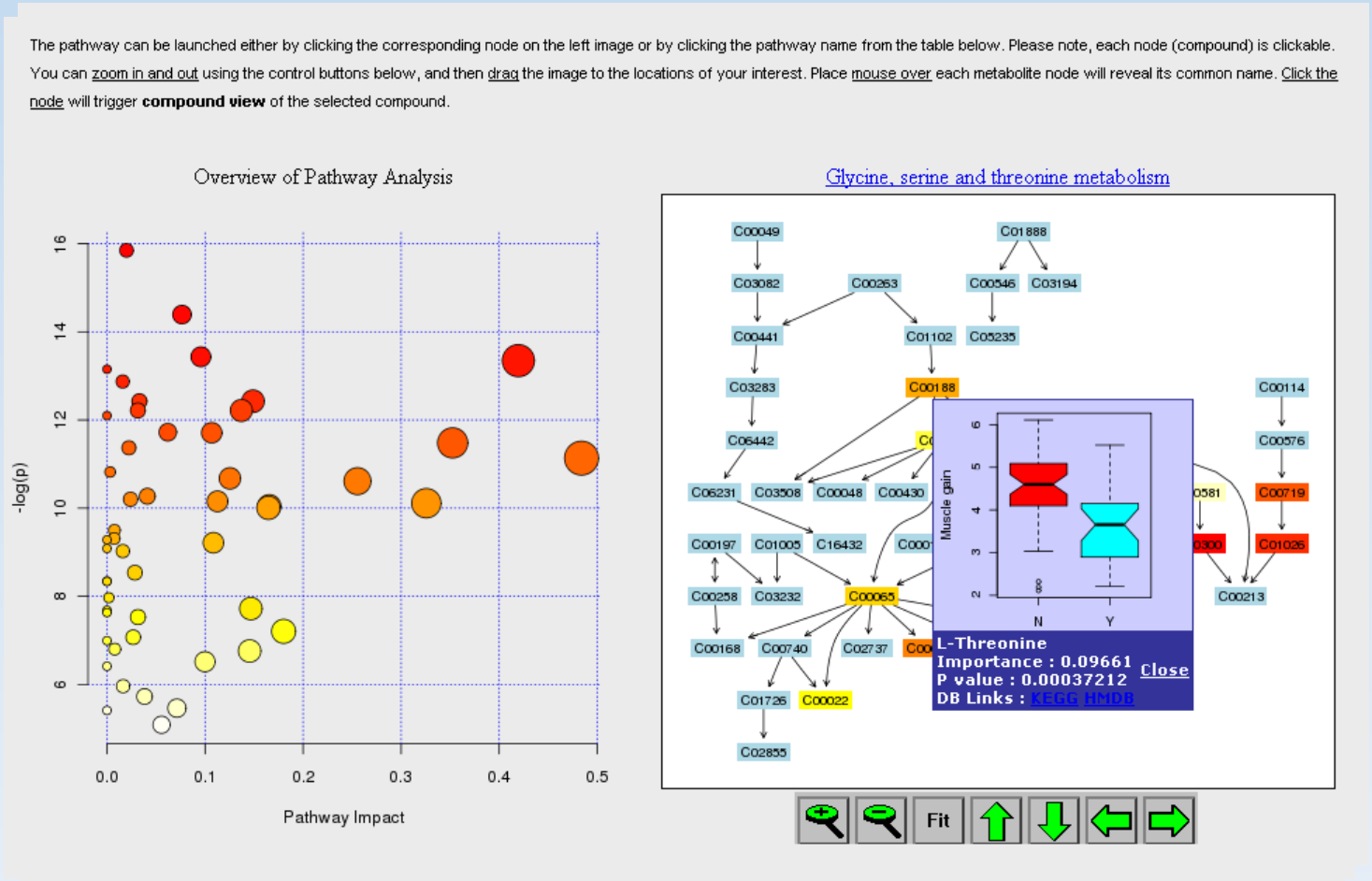
Node importance measure: centrality

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

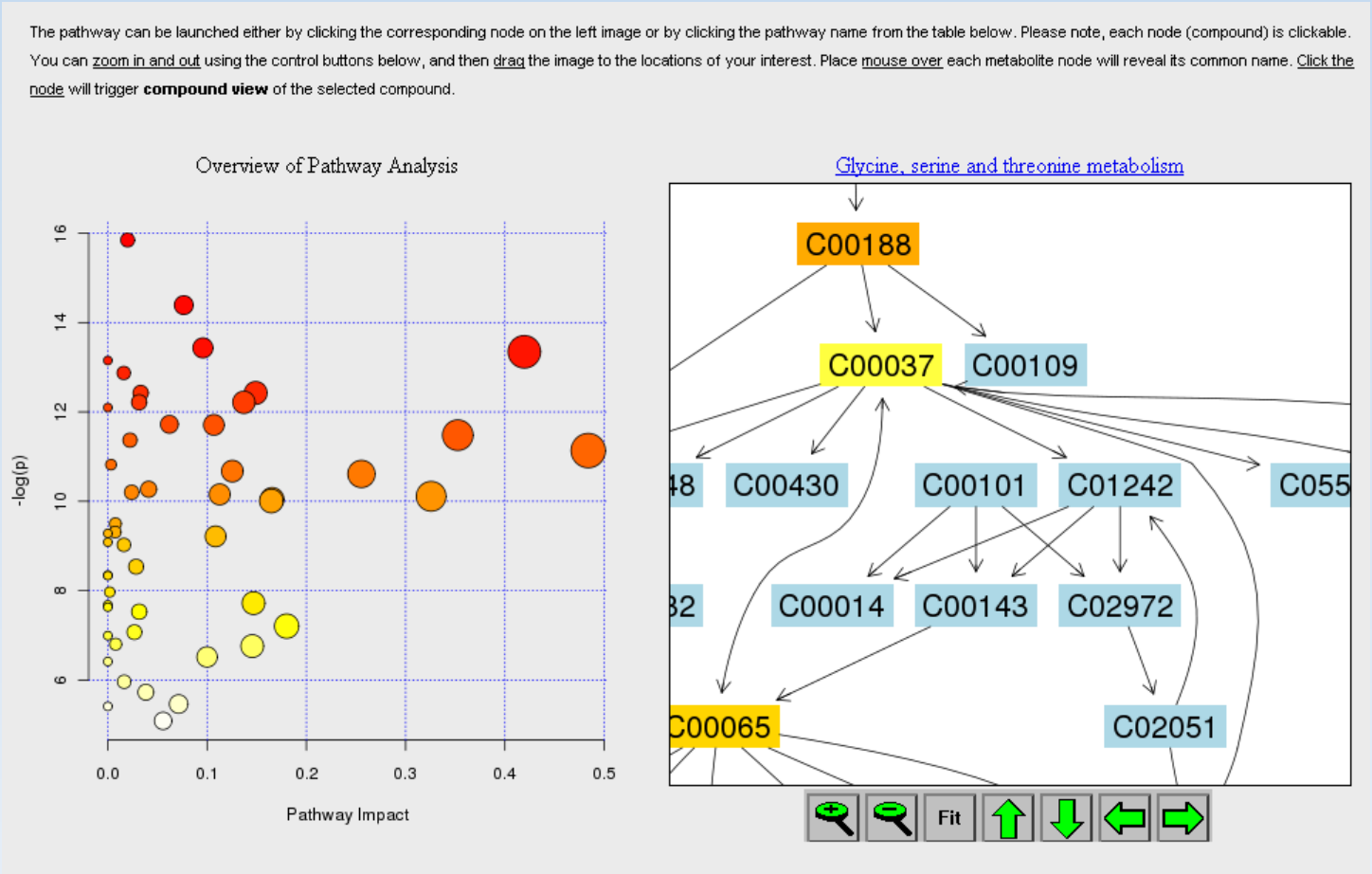


Junker *et al. BMC Bioinformatics* 2006

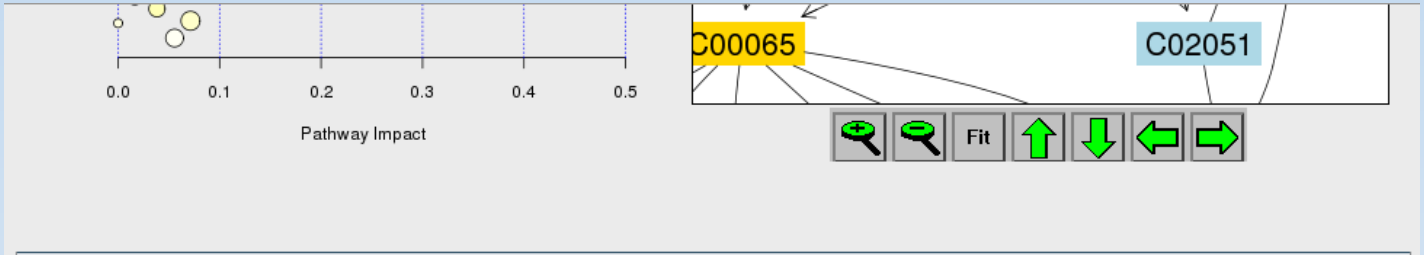
Point and Click



Lossless zooming



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

Downloads

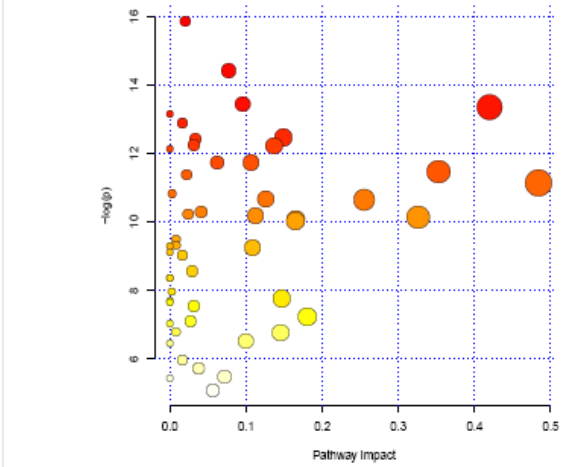


Figure 1: Summary of Pathway Analysis

5

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

	Total Cmpd	Hits	Raw p	-log(p)	Holm adjust	FDR	Impact
Galactose metabolism	41	3	1.21E-07	8.89E+00	6.28E-06	6.28E-06	0.02
Starch and sucrose metabolism	10	3	5.62E-07	8.25E+00	2.91E-05	1.45E-05	0.08
Glycolysis or Gluconeogenesis	31	4	1.47E-06	7.83E+00	7.18E-05	1.96E-05	0.18
Pyruvate metabolism	32	4	1.60E-06	7.80E+00	7.98E-05	1.96E-05	0.42
Amino sugar and nucleotide sugar metabolism	88	3	1.94E-06	7.71E+00	9.13E-05	1.96E-05	0.00
Propanoate metabolism	35	4	2.57E-06	7.59E+00	1.16E-04	2.16E-05	0.02
Valine, leucine and isoleucine biosynthesis	27	6	4.02E-06	7.40E+00	1.81E-04	2.53E-05	0.15
Sulfur metabolism	16	2	4.02E-06	7.40E+00	1.81E-04	2.53E-05	0.03
Phenylalanine metabolism	45	6	4.95E-06	7.31E+00	2.13E-04	2.53E-05	0.03
Inositol phosphate metabolism	39	1	4.96E-06	7.30E+00	2.13E-04	2.53E-05	0.16
Pentose phosphate pathway	22	2	5.57E-06	7.25E+00	2.29E-04	2.58E-05	0.00
Arginine and proline metabolism	77	6	6.11E-06	7.09E+00	3.24E-04	3.22E-05	0.06
Tyrosine metabolism	76	5	6.22E-06	7.09E+00	3.24E-04	3.22E-05	0.11
Taurine and hypotaurine metabolism	20	3	1.03E-05	6.99E+00	3.90E-04	3.77E-05	0.35
Valine, leucine and isoleucine degradation	40	3	1.16E-05	6.96E+00	4.28E-04	3.80E-05	0.02
Glycine, serine and threonine metabolism	48	8	1.46E-05	6.85E+00	5.24E-04	4.66E-05	0.48
Selenoamino acid metabolism	22	1	2.00E-05	6.70E+00	7.00E-04	6.00E-05	0.00
Butanoate metabolism	40	5	2.31E-05	6.64E+00	7.97E-04	6.56E-05	0.13
Alanine, aspartate and glutamate metabolism	24	6	2.45E-05	6.61E+00	8.13E-04	6.81E-05	0.26
Nicotinate and nicotinamide metabolism	44	5	3.47E-05	6.46E+00	1.11E-03	8.84E-05	0.04
Pentose and glucuronate interconversions	53	2	3.71E-05	6.42E+00	1.15E-03	9.01E-05	0.02
Aminoacyl-tRNA biosynthesis	75	12	3.86E-05	6.41E+00	1.17E-03	9.01E-05	0.11
Citrate cycle (TCA cycle)	20	6	4.06E-05	6.39E+00	1.18E-03	9.01E-05	0.33
Glyoxylate and dicarboxylate metabolism	50	8	4.35E-05	6.36E+00	1.22E-03	9.21E-05	0.17
Methane metabolism	34	6	4.51E-05	6.35E+00	1.22E-03	9.21E-05	0.16
Nitrogen metabolism	39	6	5.51E-05	6.12E+00	1.36E-03	1.47E-04	0.01
Phenylalanine, tyrosine and tryptophan biosynthesis	27	2	8.05E-05	6.04E+00	2.26E-03	3.70E-04	0.01
Biotin metabolism	11	1	8.32E-05	6.03E+00	2.26E-03	3.70E-04	0.00
Tryptophan metabolism	79	1	8.96E-05	6.00E+00	2.29E-03	3.75E-04	0.11
Synthesis and degradation of ketone bodies	6	2	1.13E-04	5.95E+00	2.50E-03	3.93E-04	0.00
Ascorbate and aldarate metabolism	45	3	1.20E-04	5.92E+00	2.53E-03	3.96E-04	0.02
Cysteine and methionine metabolism	58	2	1.96E-04	5.71E+00	3.90E-03	5.11E-04	0.03
Ubiquinone and other terpenoid-quinone biosynthesis	36	1	2.26E-04	5.62E+00	4.52E-03	5.59E-04	0.00
Cyanoamino acid metabolism	18	2	2.40E-04	5.62E+00	4.52E-03	5.59E-04	0.00
Glutathione metabolism	26	2	2.47E-04	5.60E+00	4.89E-03	5.61E-04	0.00
Lysine degradation	47	3	4.44E-04	5.35E+00	7.10E-03	8.29E-04	0.15
Sphingolipid metabolism	25	1	4.61E-04	5.34E+00	7.10E-03	8.29E-04	0.00
Terpenoid backbone biosynthesis	33	1	4.66E-04	5.31E+00	7.10E-03	8.52E-04	0.00
Proteasome and mannose metabolism	48	1	5.26E-04	5.27E+00	7.10E-03	7.03E-04	0.03
Pantothenate and CoA biosynthesis	27	2	7.44E-04	5.17E+00	9.92E-03	9.48E-04	0.18
D-Glutamine and D-glutamate metabolism	11	2	8.51E-04	5.07E+00	9.36E-03	1.06E-03	0.03
Thiamine metabolism	24	2	9.16E-04	5.04E+00	9.36E-03	1.12E-03	0.00
Purine metabolism	92	3	1.11E-03	5.00E+00	1.00E-02	1.32E-03	0.01
Histidine metabolism	44	3	1.16E-03	4.96E+00	1.00E-02	1.51E-03	0.15
Lysine biosynthesis	32	2	1.48E-03	4.91E+00	1.04E-02	1.86E-03	0.18
Porphyrin and chlorophyll metabolism	104	2	1.64E-03	4.79E+00	1.04E-02	1.82E-03	0.00
Primary bile acid biosynthesis	47	2	2.58E-03	4.59E+00	1.29E-02	2.80E-03	0.02
Vitamin B6 metabolism	32	2	2.26E-03	4.49E+00	1.31E-02	3.47E-03	0.04
Pyrimidine metabolism	80	3	4.25E-03	4.37E+00	1.31E-02	4.42E-03	0.07
beta-Alanine metabolism	26	3	4.48E-03	4.30E+00	1.31E-02	4.57E-03	0.00
Glycerophospholipid metabolism	39	1	6.16E-03	4.21E+00	1.31E-02	6.16E-03	0.06

6

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
- T3DB

Summary

1st & 2nd generation	• MetaboAnalyst: general data processing & analysis ✓
3rd generation	• MSEA: Metabolite set enrichment analysis ✓
4th generation	• MetPA: Metabolomics Pathway Analysis ✓
5th generation	• Integrate with other omics data 

Acknowledgement

- Dr. David Wishart

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