

Tutorial to run MetENPWeb

Sonal Choudhary
Metabolomics Workbench

Steps

Example 1. Datasets already on Metabolomics Workbench

The screenshot shows the Metabolomics Workbench homepage. At the top left is the logo, and at the top right are 'Log in / Register' and a search bar. A navigation menu includes 'Home', 'Data Repository', 'Databases', 'Protocols', 'Tools', 'Training / Events', 'About', and 'Search'. Below the menu is a welcome message. The main content area is titled 'National Metabolomics Data Repository' and features three buttons: 'Upload and Manage Studies', 'Browse and Search Studies', and 'Analyze Studies'. A red arrow points to the 'Analyze Studies' button. Below the buttons, text states that as of 11/17/20, the NMDR contains 1263 publicly available studies. A section titled 'Recently released studies on NMDR' lists three studies with their IDs and descriptions. On the right side, there is a 'Quick Links - Key Resources' dropdown menu, an 'EVENTS CALENDAR' link, and a list of NIH Common Fund Stage 2 Metabolomics Consortium Centers, including the Metabolomics Consortium Coordinating Center (M3C) and Compound Identification Cores (CIDCs).

Metabolomics Workbench

METABOLOMICS WORKBENCH

Log in / Register

Search the Metabolomics Workbench

Home | Data Repository | Databases | Protocols | Tools | Training / Events | About | Search

Welcome to the UCSD Metabolomics Workbench, a resource sponsored by the Common Fund of the National Institutes of Health.

National Metabolomics Data Repository

[Upload and Manage Studies](#) | [Browse and Search Studies](#) | [Analyze Studies](#)

As of 11/17/20 the National Metabolomics Data Repository (NMDR) contains **1263 publicly available studies**. A total of **1471 studies** have been processed by the NMDR and the remainder (**208**) will be made available subject to their embargo dates.

Recently released studies on NMDR

ST001378 - Global metabolomics of COPD2020; *Homo sapiens*; [Seoul National University Hospital](#)

ST001512 - Diel investments in phytoplankton metabolite production influenced by associated heterotrophic bacteria; *Thalassiosira pseudonana*; [University of Georgia](#)

ST001498 - Interday validation of of developed quantatation method; *Fecal Bacteria*; [Helmholtz Centre for Environmental Research - UFZ](#)

Quick Links - Key Resources

EVENTS CALENDAR

NIH Common Fund Stage 2 Metabolomics Consortium Centers

[Metabolomics Consortium Coordinating Center \(M3C\)](#)
Richard Yost, U. of Florida

[Metabolomics Workbench/NMDR](#)
Shankar Subramaniam, UC San Diego (this website)

[Compound Identification Cores \(CIDCs\)](#)
Arthur Edison, U. of Georgia
Alexey Nesvizhskii, U. of Michigan
Oliver Fiehn, UC Davis
Dean Paul Jones, Emory University
Thomas Metz, Pacific Northwest Nat. Lab.

Click on Analyze Studies


Study selection

Analyze Studies

Analyze studies using [Jupyter Notebooks](#) or the following online tools.


MS/NMR studies identifying named metabolites

Select a study for analysis:

ST001140: Changes in the Canine Plasma Lipidome after Short- and Long-Term Exces... (Life Sciences Institute, National University of Singapore) 

Submit

Analysis tools may also be accessed from within each [study page](#) using the 'Perform statistical analysis' link

 Choose study

Comparative analysis across studies

- [Perform meta-analysis on selected studies \(compare ratios of 2 selected metabolites\)](#)
- [Compare list of metabolites in 2 selected studies \(all analyses\)](#)
- [Compare list of metabolites in 2 selected studies \(individual analyses\)](#)

MS untargeted experiments containing unidentified ions

- [Search Untargeted MS data by m/z, retention time, instrumentation](#)
- [Superimpose unknown m/z on RefMet mass defect plot](#)

Perform data analysis on user-uploaded data

- [Load and analyze your own dataset](#)

Select MetENP

Metabolite classes (all analyses combined)

- Pie chart of metabolite super classes
- Pie chart of metabolite main classes
- Pie chart of metabolite sub classes

Normalization and averaging

- Show Metabolite averages per experimental factor
- Perform normalization on data
- Create Relative log abundance plots

Univariate analysis

- Perform multi-condition dot plot analysis **New!**
- Perform Volcano plot analysis
- Perform ANOVA analysis



MetENP: Metabolite enrichment and species-specific pathway annotation **New!**

- MetENPWeb analysis
- MetENP R package [🔗](#)

Clustering and correlation

- Perform hierarchial or heatmap cluster analysis
- Perform Clustered correlation analysis
- Perform Network analysis on correlated metabolites

Multivariate analysis

- Perform Principal component analysis
- Perform Linear discriminant analysis
- Perform Partial least-squares discriminant analysis (PLS-DA)

Classification and feature analysis

- Perform OPLS-DA and VIP projection
- Random Forest and VIP projection

Mapping metabolites to human biochemical pathways

- Map study metabolites to HMDB and KEGG pathways
- Map study metabolites to pathways with ratio/t-test data

Select analysis and factor column

[Overview](#) | [Upload / Manage Studies](#) | [Browse / Search Studies](#) | [Analyze Studies](#) | [Data Sharing Policy](#) | [Tutorials](#) | [FAQ](#)

Choose a analysis:

Hint: You can either select all, or use ctrl + select to choose multiple

Select all
Phospholipids, Chol. esters and Diacylglycerols
Sphingolipids
Derivatized Spingosine-1-phosphates

Check the experimental factors of this study in the table below. The first column is grouped (combined) factors and subsequent columns are individual factors

	combined_factors	TreatmentGroup	TreatmentDuration	SamplingTimePoint
1	TreatmentGroup:Prednisolone TreatmentDuration:0d SamplingTimePoint:before	Prednisolone	0d	before
2	TreatmentGroup:Prednisolone TreatmentDuration:4d SamplingTimePoint:after	Prednisolone	4d	after
3	TreatmentGroup:Tetracosactide TreatmentDuration:00w SamplingTimePoint:before	Tetracosactide	00w	before
4	TreatmentGroup:Tetracosactide TreatmentDuration:25w SamplingTimePoint:after	Tetracosactide	25w	after

Choose factor column:

Take a hint from the table above

combined_factors ▼
combined_factors
TreatmentGroup
TreatmentDuration
SamplingTimePoint

Workbench, a resource sponsored by the Common Fund of the National Institutes of Health

[Terms of use](#) | [Site map](#) | [Contact](#) | [Personnel](#)

Parameters

[Home](#) | [Data Repository](#) | [Databases](#) | [Protocols](#) | [Tools](#) | [Training / Events](#) | [About](#) | [Search](#)

[Overview](#) | [Upload / Manage Studies](#) | [Browse / Search Studies](#) | [Analyze Studies](#) | [Data Sharing Policy](#) | [Tutorials](#) | [FAQ](#)

Choose among these factors as Group 1 and Group 2. Multiple factors can be chosen for one group

Group1:
Prednisolone
Tetracosactide

← Choose factor 1

Group2:
Prednisolone
Tetracosactide

← Choose factor 2

Padjust method:

Handle missing data:

[Hint:](#)

- **half_of_min:** where the NAs are replaced by half of the min values in the data.
- **remove_NAs:** where metabolites with NAs values are removed and
- **50percent:** where metabolites with more than 50% NAs values are removed.

P-value cutoff:

Log2 fold change cutoff:

Choose a Metabolite class:

Minimum # of (most significant) metabolites per class to use in group calculation:

Kegg id of species is automatically filled. For bacterial species, please check our notes below

Kegg id of species: Species -Dog

Check this box only if you want your pathway enrichment analysis based on all the KEGG compounds associated to pathways.

Not applicable
 All pathways

Do you want to associate genes to pathways?
If you select 'Perform gene analysis' option, please be aware it may take longer than 5 mins to analyze the study

Don't perform gene analysis Perform gene analysis

The analysis may take 2-5 minutes, please wait and do not close the window. Longer datasets may take even longer. You may want to run MetENP package on the R commandline for quicker results

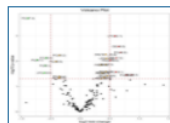
Please note that some bacterial species have different strain name in kegg and hence different kegg codes, e.g. : Salmonella enterica subsp. enterica serovar Typhimurium has 'seo' for Salmonella enterica subsp. enterica serovar Typhimurium 14028S but 'sev' for Salmonella enterica subsp. enterica serovar Typhimurium D23580, in such cases, if study in metabolomics workbench has the name Salmonella enterica subsp. enterica serovar Typhimurium, it may not be able to pick up the kegg code. Please check the correct kegg code for your analysis.

Results

[Home](#) | [Data Repository](#) | [Databases](#) | [Protocols](#) | [Tools](#) | [Training / Events](#) | [About](#) | [Search](#)

[Overview](#) | [Upload / Manage Studies](#) | [Browse / Search Studies](#) | [Analyze Studies](#) | [Data Sharing Policy](#) | [Tutorials](#) | [FAQ](#)

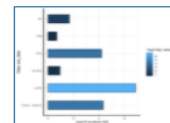
Click on individual image to get a full size downloadable figure. Picture displayed below just represents the type of analysis and are not the actual results.



Volcano plot of all significant metabolites



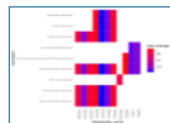
Metabolite count grouped by metabolite class



Metabolite enrichment grouped by metabolite class



Pathway network of significant metabolites



Heatmap of significant metabolites and significant pathways



Dotplot of significant metabolites and significant pathways

[Change axis label fonts in heatmap](#)

[Change axis label fonts in dot plot](#)

[Modify heatmap](#)

[Modify Dotplot](#)

Zippered Result file

[Allresults.zip](#)

Individual Result file

[Kegg_pathway_enrichment.csv](#)

[met_enrichmentscore.csv](#)

[met_path.csv](#)

[sig_metabolites_kegg_id.csv](#)

[significant_met.csv](#)

[stats_metabolites.csv](#)

Parameters Chosen:

Study ID: ST001140

Analysis: Phospholipids, Chol. esters and Diacylglycerols--Sphingolipids--Derivatized Spingosine-1-phosphates--Triacylglycerols

Factor column: TreatmentGroup

Group1: Prednisolone

Group2: Tetracosactide

p-adj: fdr; **NA filtering criteria:** 50percent; **p-value cutoff:** 0.05; **fold change cutoff:** 0.5

Metabolite class: sub_class

Number of metabolites: 3

Species: cfa

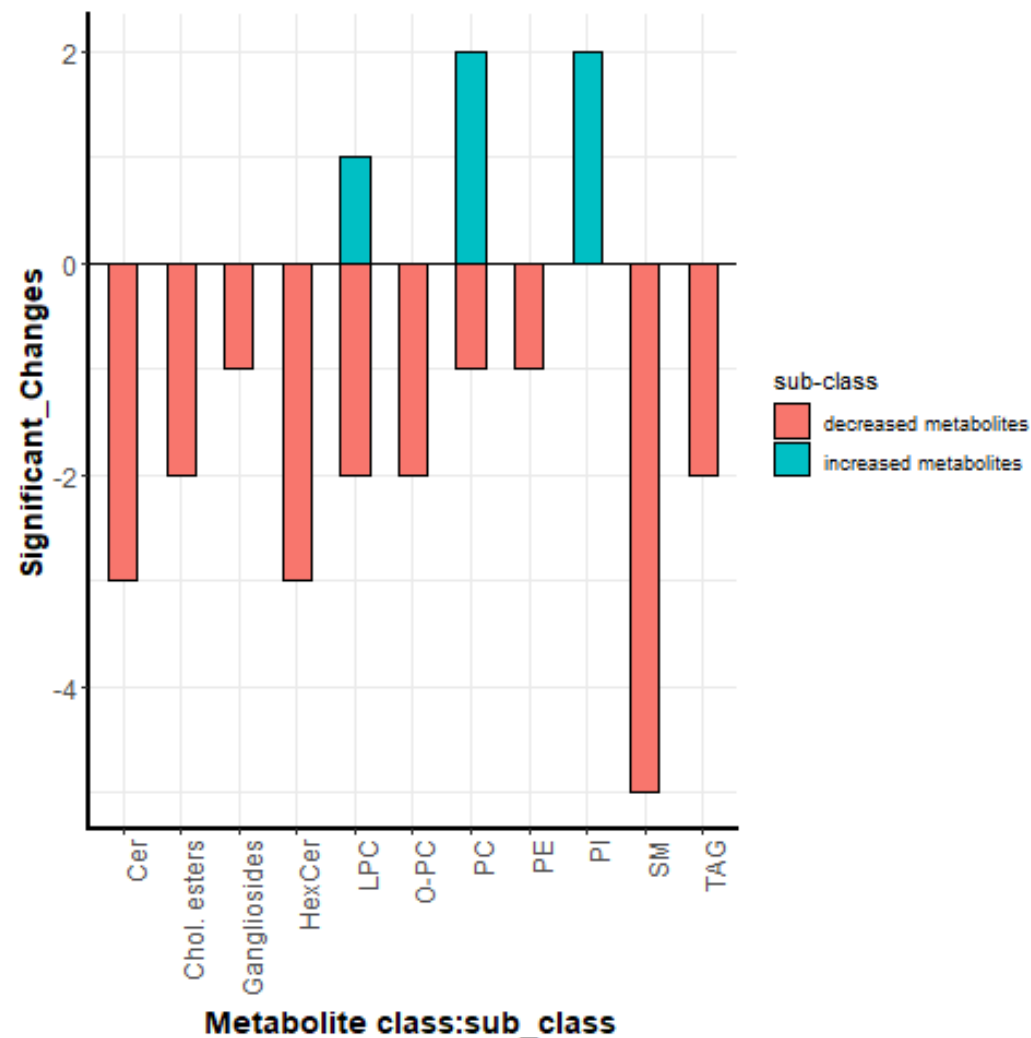
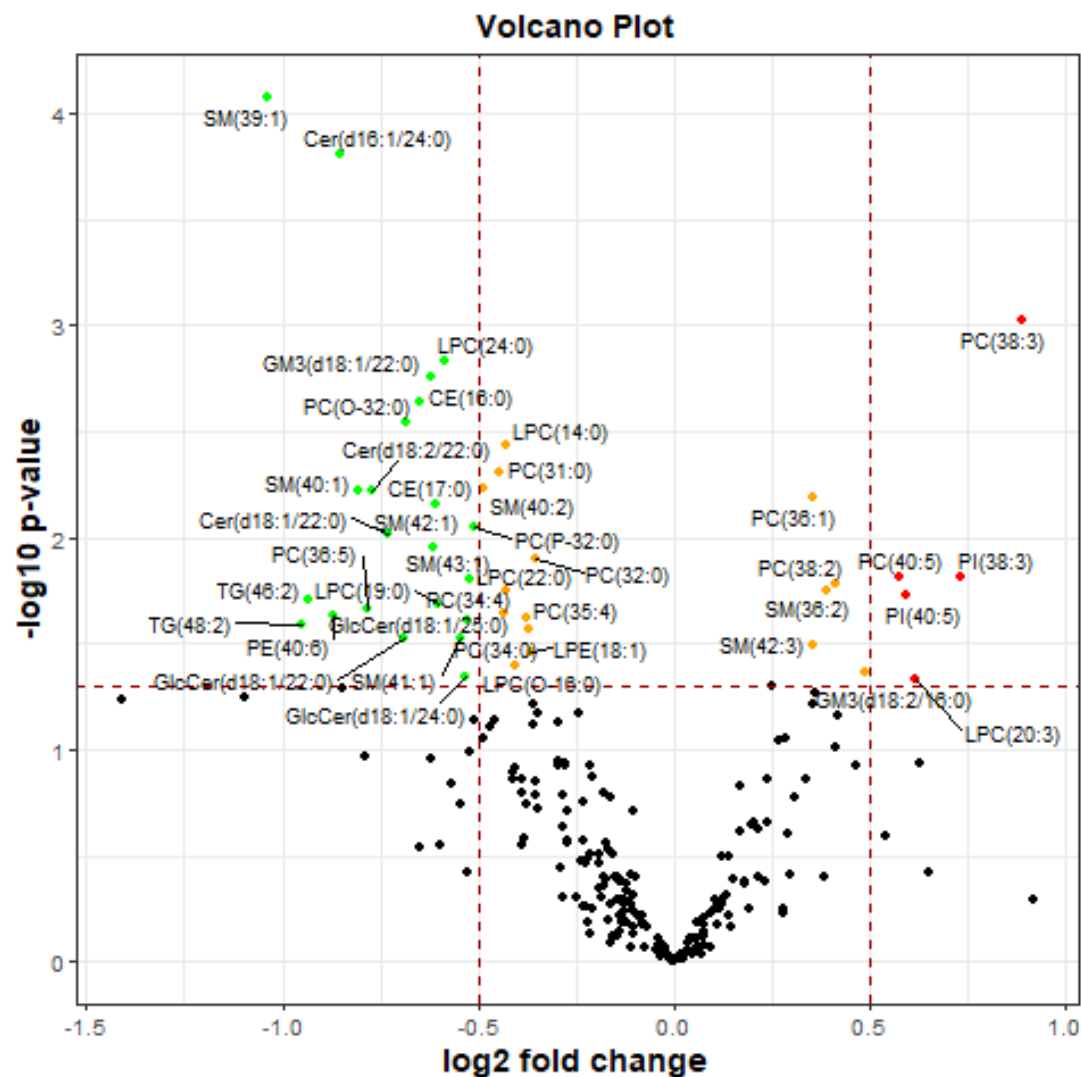
Pathway enrichment (Pathways from all KEGG pathways): FALSE

Gene analysis: FALSE

ERROR:

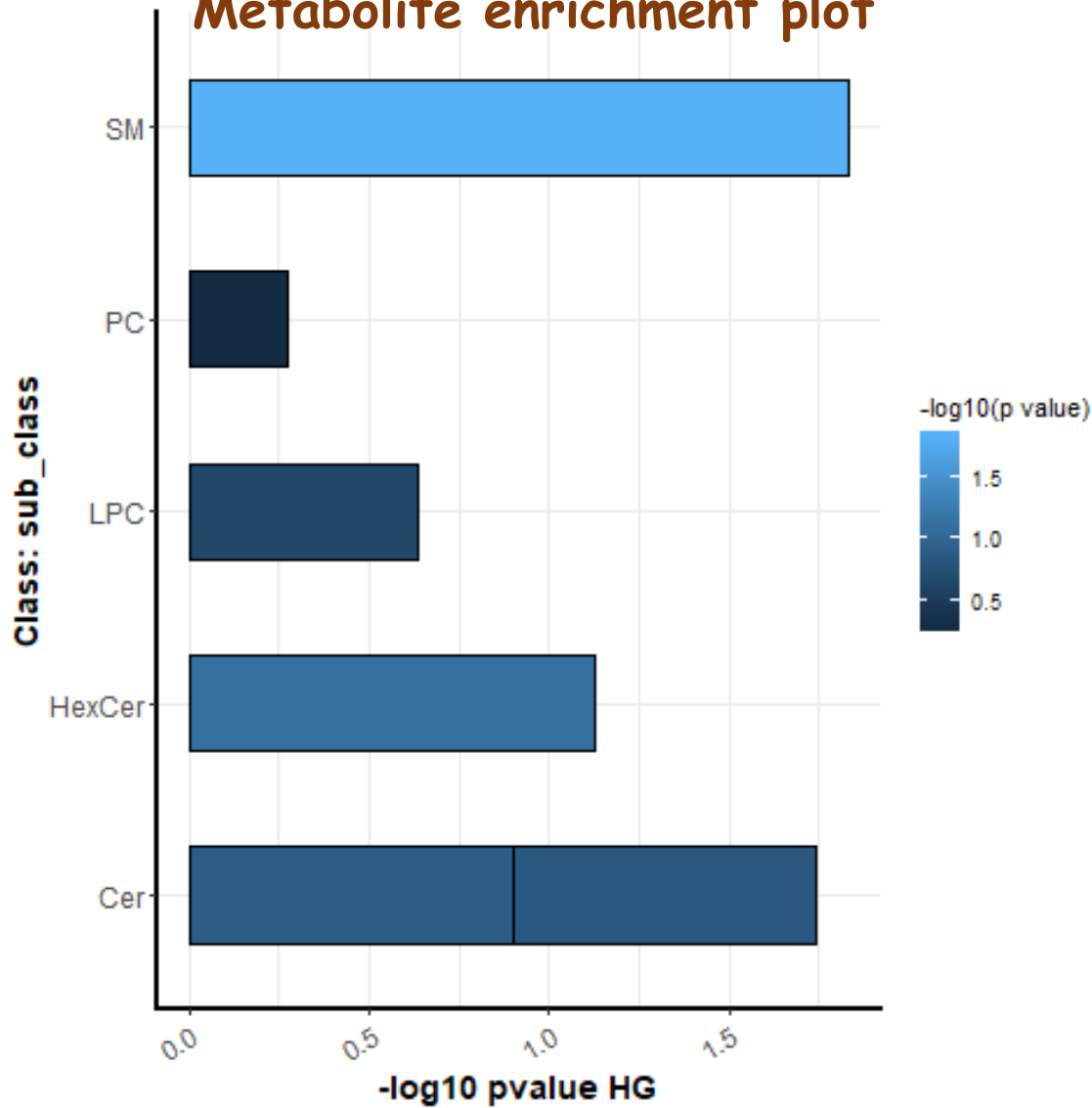
For any difficulty in running the application, contact Sonal Choudhary: kschoudhary@eng.ucsd.edu or biozonal@gmail.com

Visualization Plots

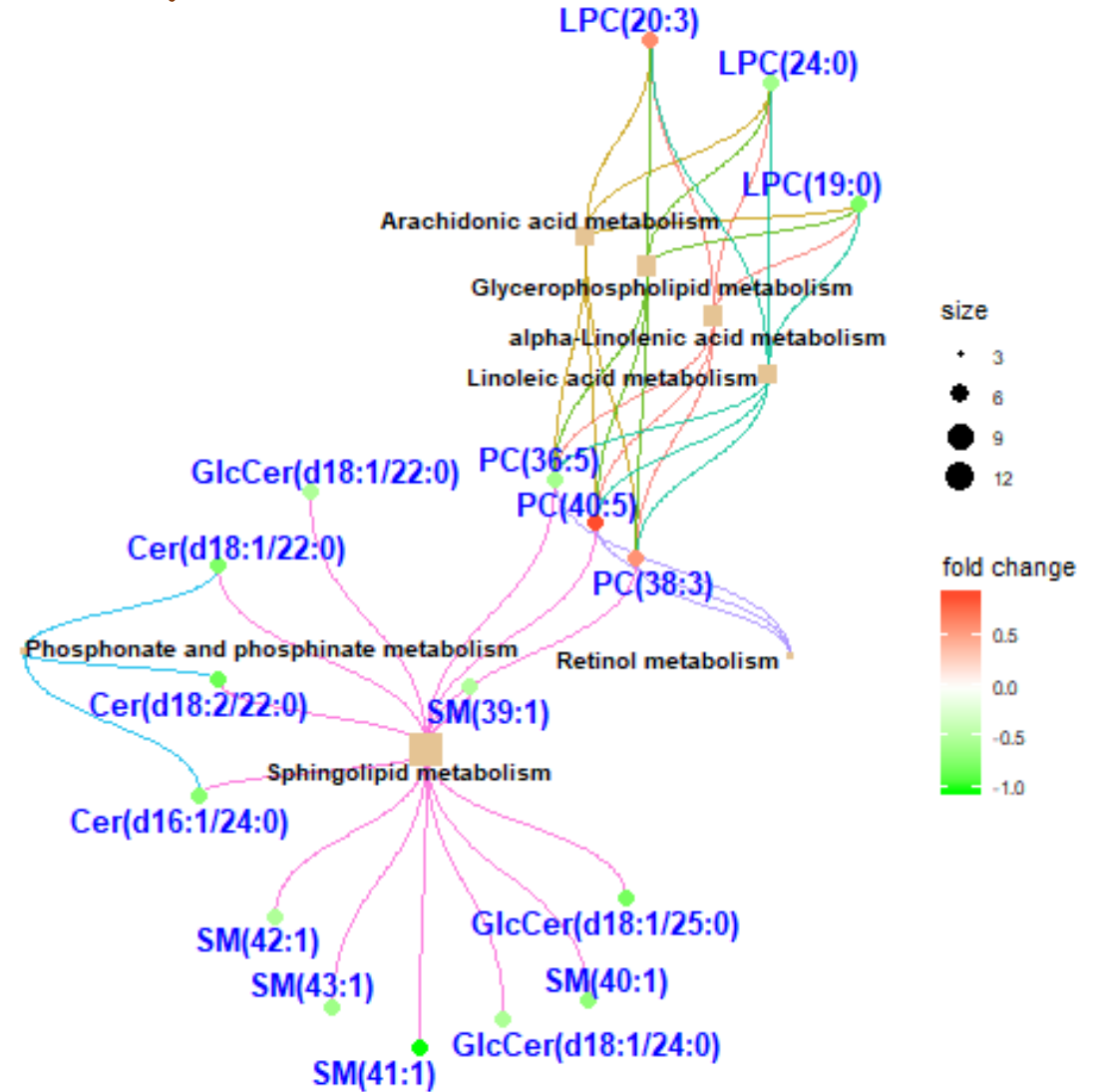


Visualization Plots

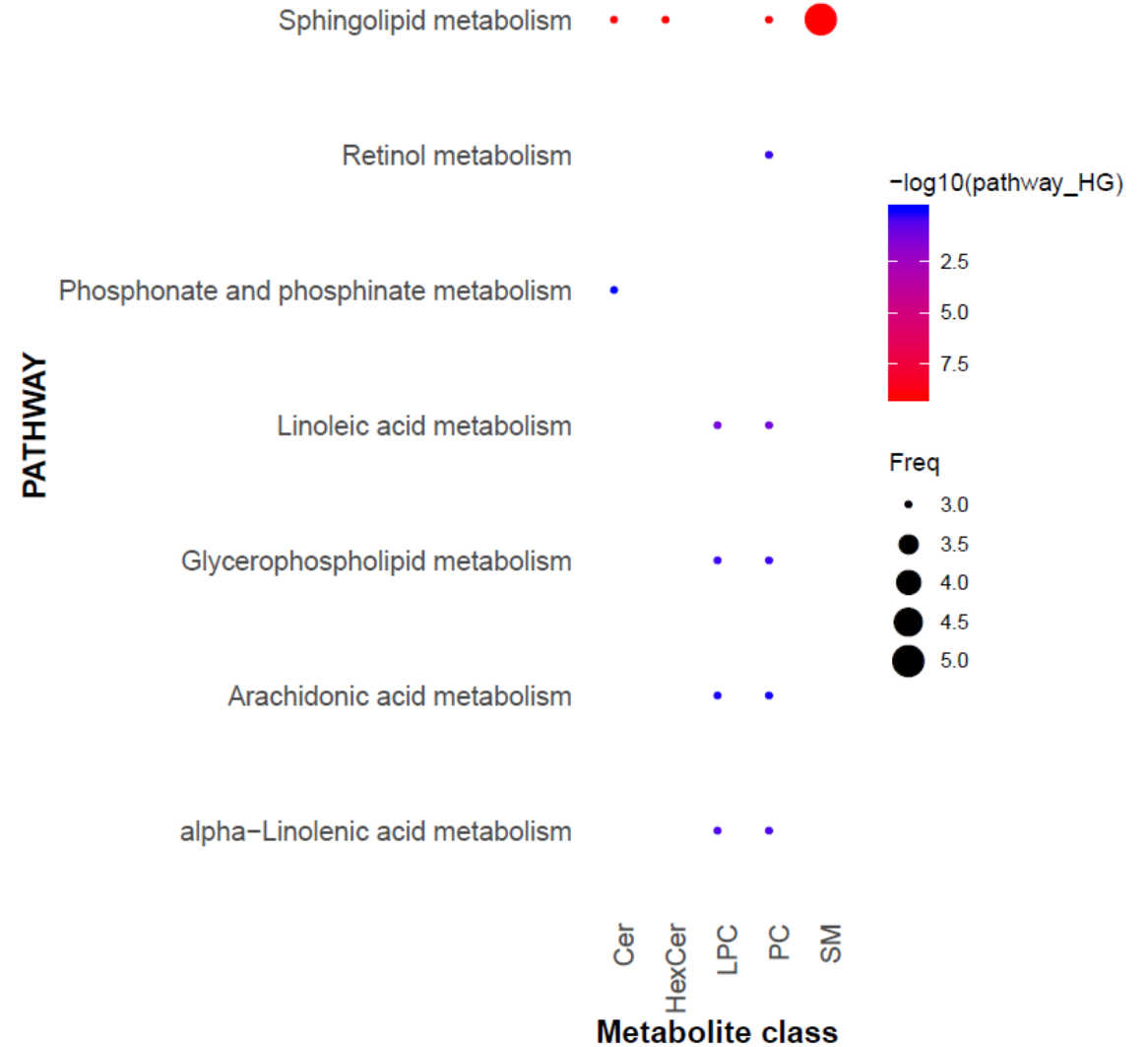
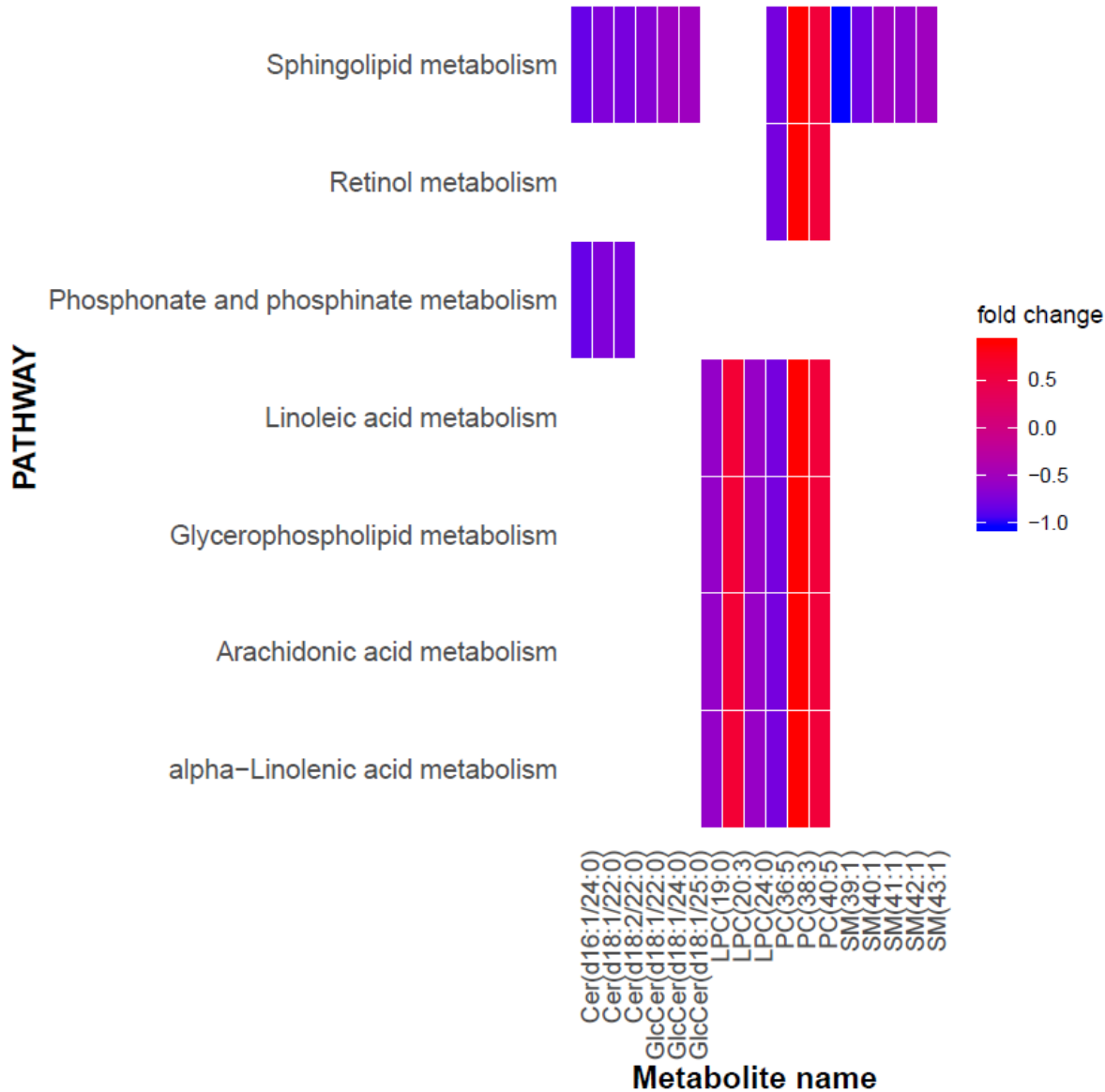
Metabolite enrichment plot



Pathway- Metabolite network



Visualization Plots



User dataset

Example 2. Custom dataset

The screenshot shows the Metabolomics Workbench website. At the top left is the logo, and in the center is the title 'METABOLOMICS WORKBENCH'. On the right, there is a 'Log in / Register' link and a search bar with the placeholder text 'Search the Metabolomics Workbench'. Below the header is a navigation menu with links for Home, Data Repository, Databases, Protocols, Tools, Training / Events, About, and Search. A welcome message follows: 'Welcome to the UCSD Metabolomics Workbench, a resource sponsored by the Common Fund of the National Institutes of Health.' The main content area is divided into three columns. The left column is titled 'National Metabolomics Data Repository' and contains three buttons: 'Upload and Manage Studies', 'Browse and Search Studies', and 'Analyze Studies'. Below these buttons is a paragraph stating that as of 11/17/20, the NMDR contains 1263 publicly available studies, with a total of 1471 studies processed, and 208 more to be made available. Below this is a section for 'Recently released studies on NMDR' with three entries: ST001378, ST001512, and ST001498. The middle column has a 'Quick Links - Key Resources' dropdown menu. The right column is titled 'EVENTS CALENDAR' and lists 'NIH Common Fund Stage 2 Metabolomics Consortium Centers' with three entries: Metabolomics Consortium Coordinating Center (M3C), Metabolomics Workbench/NMDR, and Compound Identification Cores (CIDCs).

National Metabolomics Data Repository

[Upload and Manage Studies](#) | [Browse and Search Studies](#) | [Analyze Studies](#)

As of 11/17/20 the National Metabolomics Data Repository (NMDR) contains **1263 publicly available studies**. A total of **1471 studies** have been processed by the NMDR and the remainder (**208**) will be made available subject to their embargo dates.

Recently released studies on NMDR

ST001378 - Global metabolomics of COPD2020; *Homo sapiens*; [Seoul National University Hospital](#)

ST001512 - Diel investments in phytoplankton metabolite production influenced by associated heterotrophic bacteria; *Thalassiosira pseudonana*; [University of Georgia](#)

ST001498 - Interday validation of of developed quantatation method; *Fecal Bacteria*; [Helmholtz Centre for Environmental Research - UFZ](#)

Quick Links - Key Resources

EVENTS CALENDAR

NIH Common Fund Stage 2 Metabolomics Consortium Centers

[Metabolomics Consortium Coordinating Center \(M3C\)](#)
Richard Yost, U. of Florida

[Metabolomics Workbench/NMDR](#)
Shankar Subramaniam, UC San Diego (this website)

[Compound Identification Cores \(CIDCs\)](#)
Arthur Edison, U. of Georgia
Alexey Nesvizhskii, U. of Michigan
Oliver Fiehn, UC Davis
Dean Paul Jones, Emory University
Thomas Metz, Pacific Northwest Nat. Lab.

Click on Analyze Studies

Perform data analysis on user-uploaded data

- Load and analyze your own dataset



- [Load and analyze your own dataset by MetENP](#) ^{New!}

Click here

Data Upload

[Home](#) | [Data Repository](#) | [Databases](#) | [Protocols](#) | [Tools](#) | [Training / Events](#) | [About](#) | [Search](#)

[Overview](#) | [Upload / Manage Studies](#) | [Browse / Search Studies](#) | [Analyze Studies](#) | [Data Sharing Policy](#) | [Tutorials](#) | [FAQ](#)

Upload a metabolomics data file (.csv or .txt): No file chosen

Upload the metabolomics data.

Please check the data format of your file.

- Sample names in 1st column
- Metabolites names in 1st column


Click the correct format

Input file structure of metabolomics data file

Metabolomics data file. Please check here for formats of the input file:

- **Metabolites names in 1st column**
- **Sample names in 1st column**

Check example files

The second example file is taken from [Metaboanalyst](#) 

Example upload



Upload a metabolomics data file (.csv or .txt): example.txt

Please check the data format of your file.

- Sample names in 1st column
- Metabolites names in 1st column

Input file structure of metabolomics data file

Metabolomics data file. Please check here for formats of the input file:

- **Metabolites names in 1st column**
- **Sample names in 1st column**

The second example file is taken from [Metaboanalyst](#)

After you hit 'Start Upload' button,
you can run the whole analysis like in
Example 1