MetENP: MetENPWeb & Jupyter Notebook

An R package and web application for metabolomics enrichment and pathway analysis in Metabolomics Workbench

PI: Shankar Subramaniam

Kumari Sonal Choudhary, Eoin Fahy, Kevin Coakley, Sumana Srinivasan, Mano R Maurya, Manish Sud, Shankar Subramaniam

NIH CFDE DCC Tools
August 4, 2022

Funding: NIH Common Fund, NIH grant U2C-DK119886, and CFDE program

https://www.biorxiv.org/content/10.1101/2020.11.20.391912v1

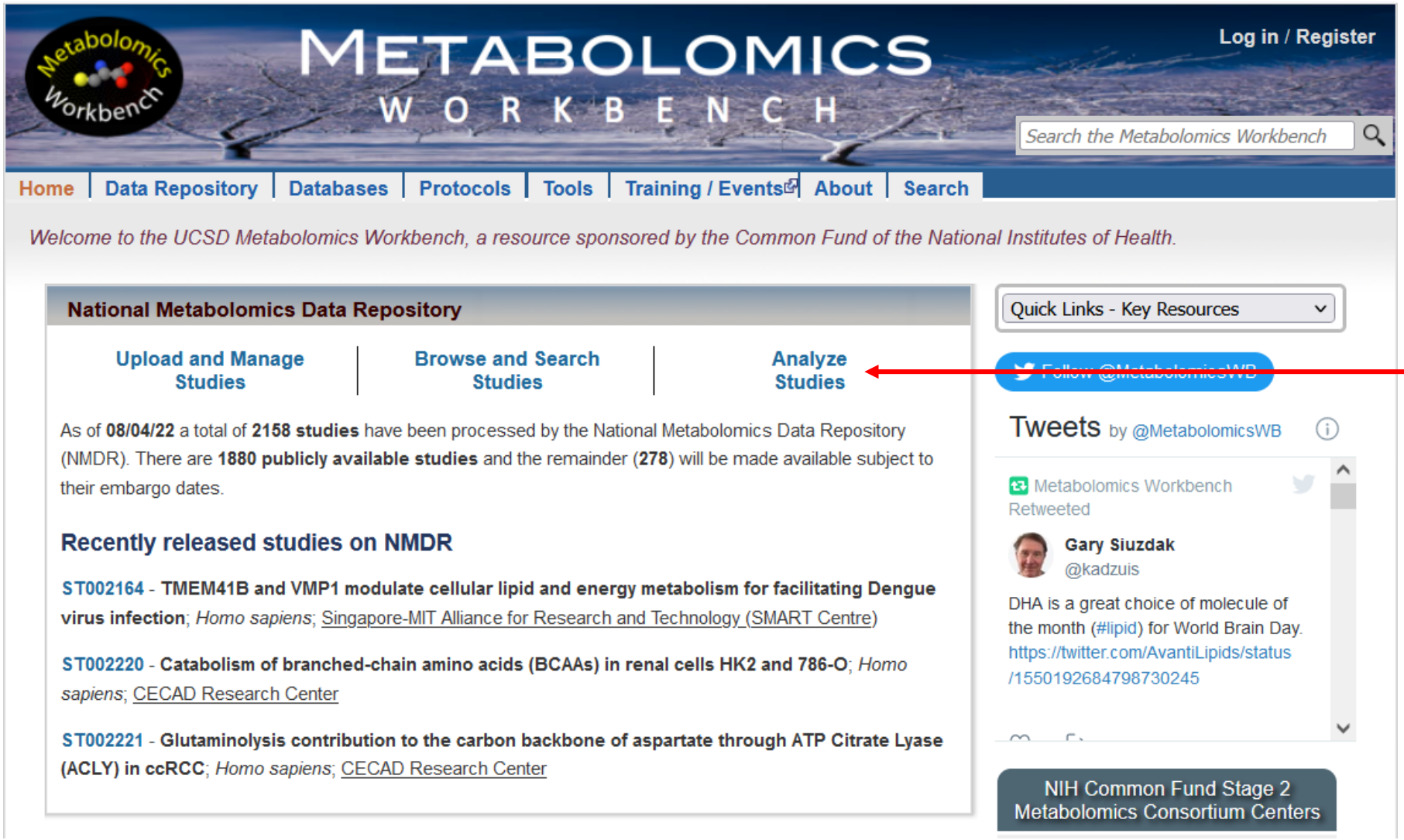
MetENP is a R package that enables detection of significant metabolites from metabolite information (names or names and concentration along with metadata information). MetENP provides enrichment score of metabolite classes, maps to pathway of the species of choice, calculates enrichment score of pathways, plots the pathways and shows the metabolite increase or decrease, and extracts and presents reaction, gene and enzyme information.

MetENP is available as a R package (https://github.com/metabolomicsworkbench/MetENP), Jupyter notebook (https://github.com/metabolomicsworkbench/MetENP/blob/main/vignettes/MetENP_vignette_Jupyter_notebook.ipynb) and web resource (https://www.metabolomicsworkbench.org/data/analyze.php).

A tutorial for the web-version is available at https://www.metabolomicsworkbench.org/data/MW-MetENP-demo.pdf. A vignette is available at https://github.com/metabolomicsworkbench/MetENP/tree/main/vignettes.

How to use MetENPWeb: Steps

Example 1. Datasets already on Metabolomics Workbench



Click on Analyze Studies

Study selection

https://www.metabolomicsworkbench.org/data/analyze.php

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

Clustering and correlation

- Perform hierarchial or heatmap cluster analysis
- Perform Clustered correlation analysis
- Perform Network analysis on correlated metabolites (mapped to classification)
- Perform Network analysis on correlated metabolites (mapped to fold-change)

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

MetaBatch Omic Browser (MD Anderson Cancer Center)

(Clustered Heat Maps, PCA+, UMAP, box plot, violin plot, and other visualizations)

- Load this study
- Load this analysis (AN001870)

MetENP: Metabolite enrichment and species-specific pathway annotation

- MetENPWeb analysis
- MetENP R package
- MetENP tutorial

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

Direct URL: https://www.metabolomicsworkbench.org/data/analysis_factor.php?STUDY_ID=ST001140

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

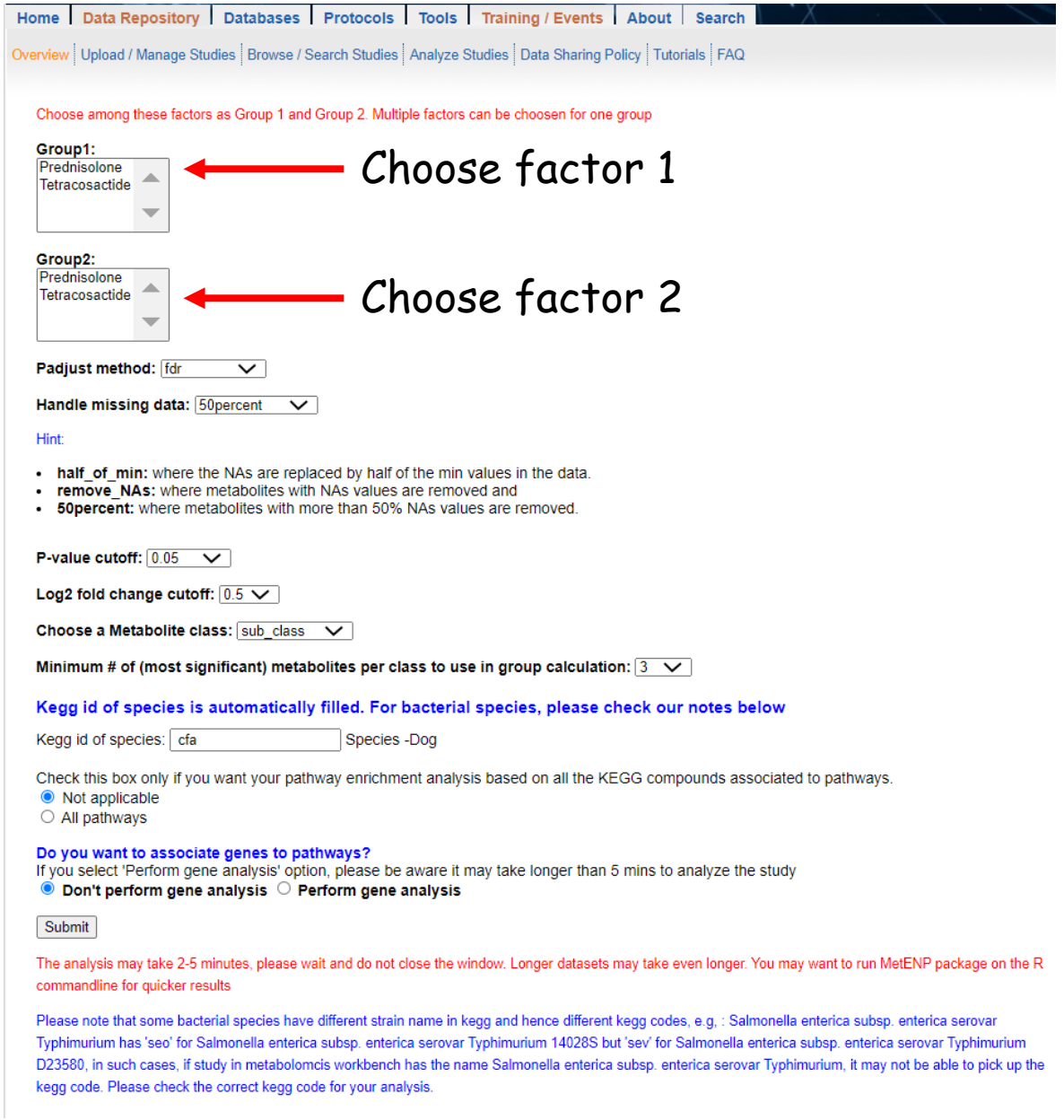
Choose factor column:

Take a hint from the table above

combined_factors

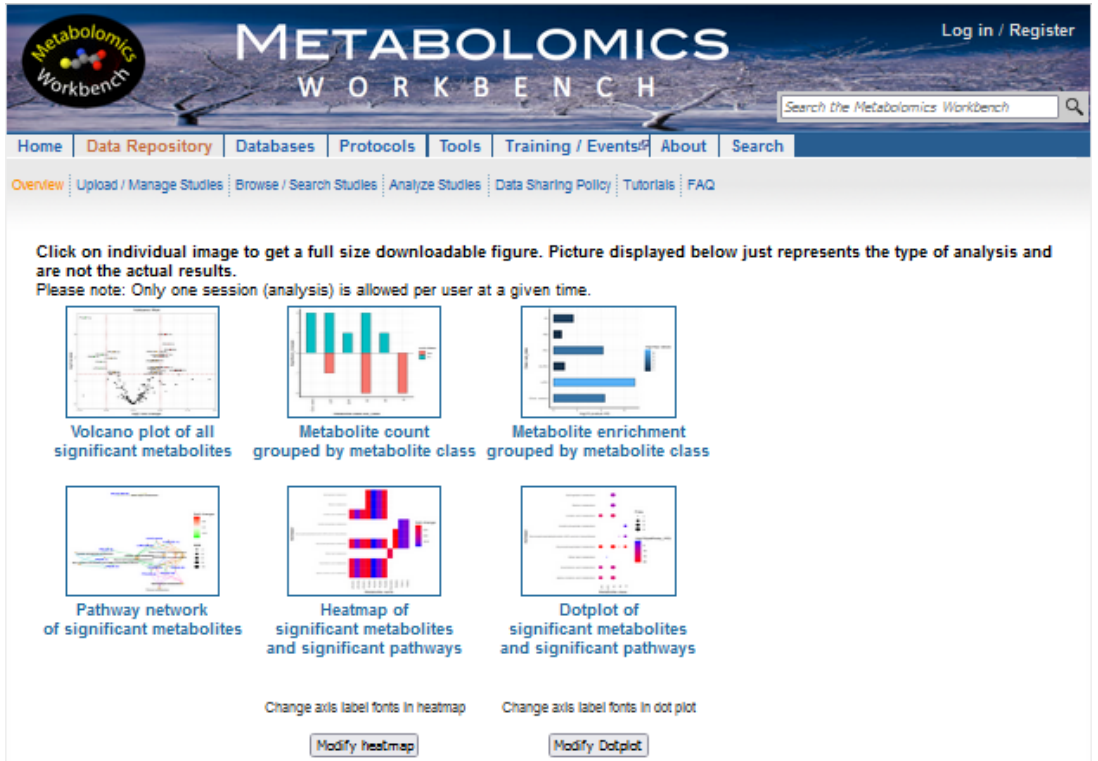
- combined_factors
- TreatmentGroup
- TreatmentDuration
- SamplingTimePoint

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 choosen for one group
Group1:
Prednisolone
Tetracosactide
Choose factor 1
Group2:
Prednisolone
Tetracosactide
Choose factor 2
Padjust method: fdr
Handle missing data: 50percent
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: 0.05
Log2 fold change cutoff: 0.5
Choose a Metabolite class: sub_class
Minimum # of (most significant) metabolites per class to use in group calculation: 3
Kegg id of species is automatically filled. For bacterial species, please check our notes below
Kegg id of species: cfa
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
Submit
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 metabolomcis 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

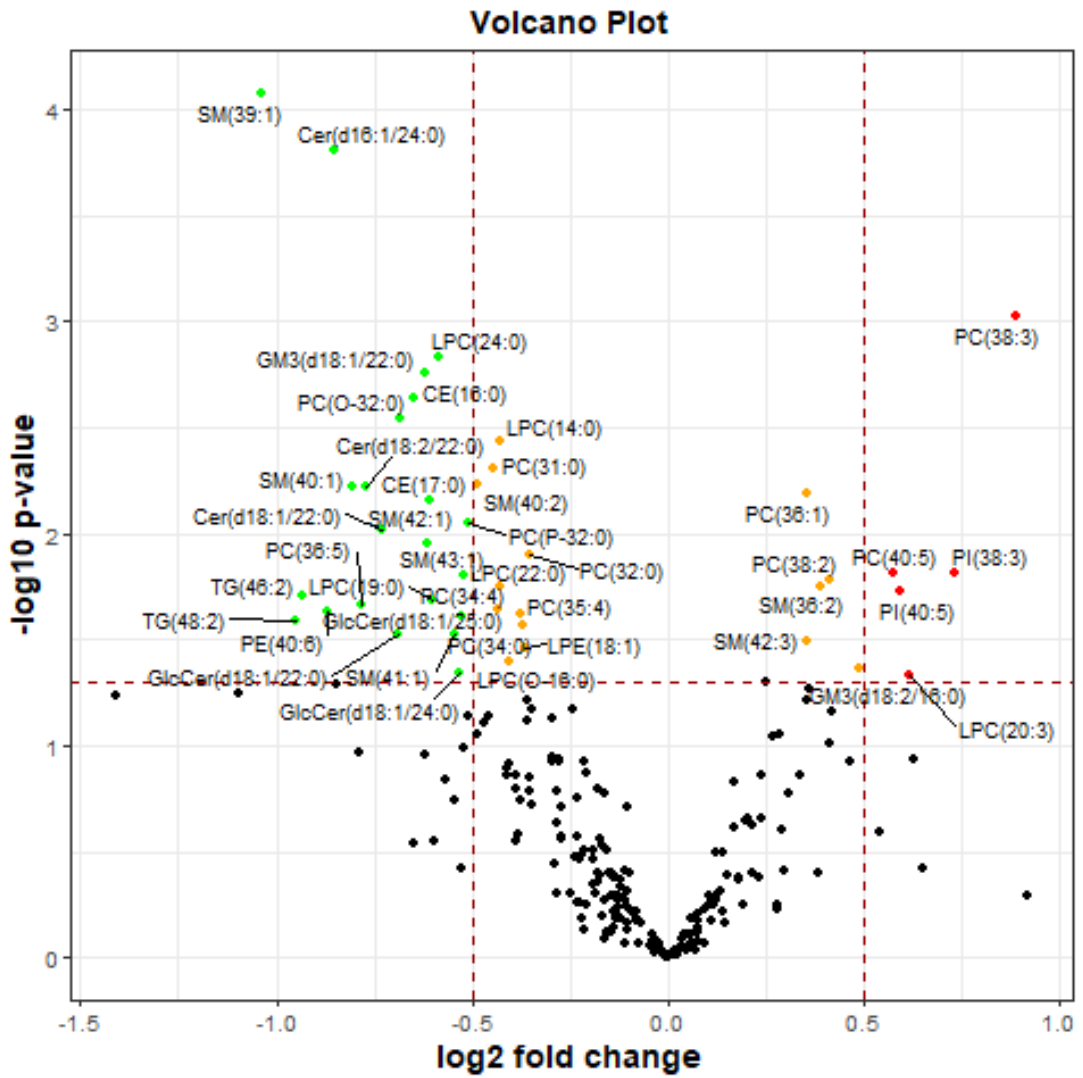
For any difficulty in running the application, contact Mano Maurya (mano@sdsc.edu) or Sumana Srinivasan (susrinivasan@eng.ucsd.edu).

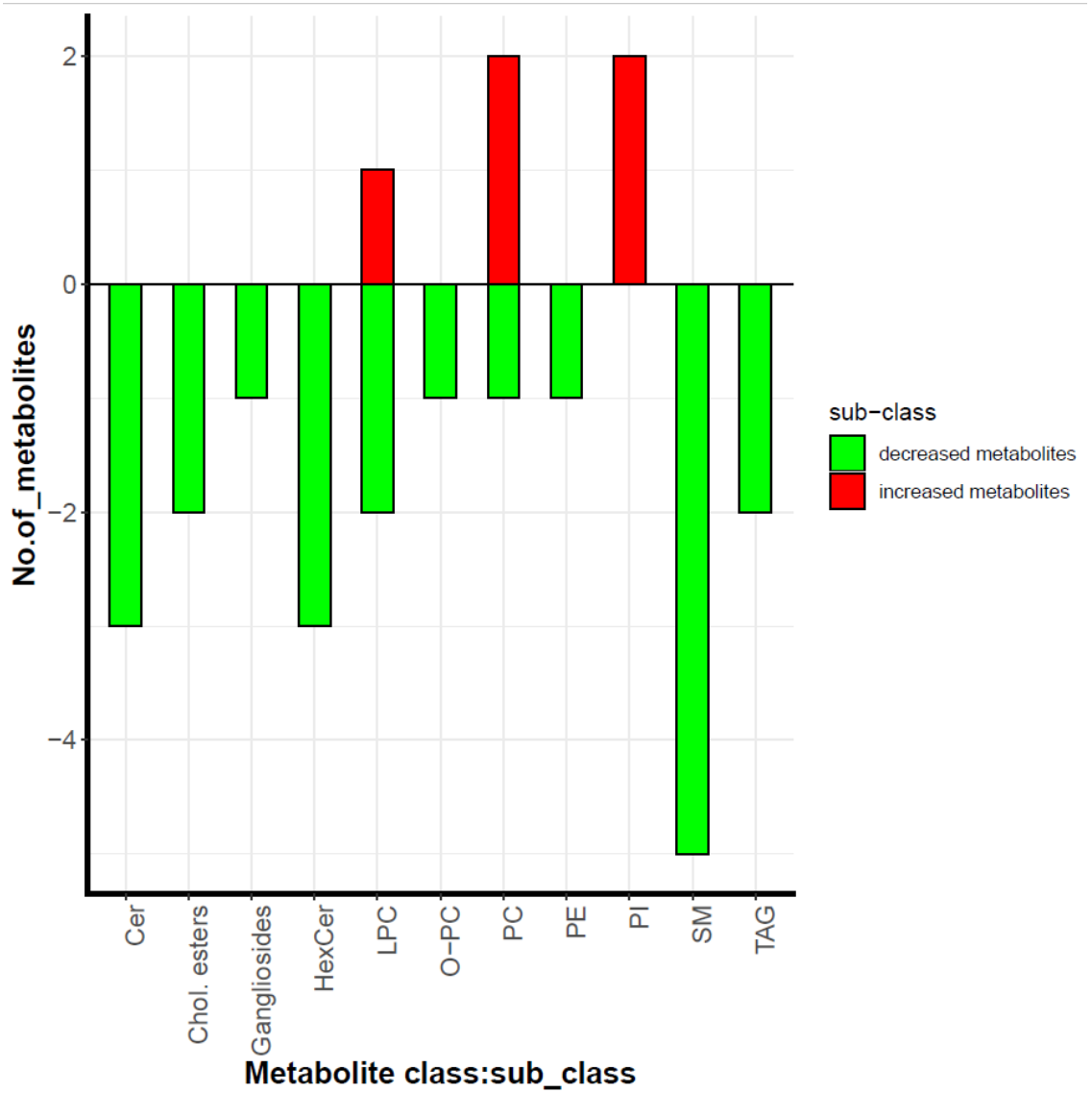


Download results as a combined zip file or individual files

Parameter values used are saved in MetENP_parameters.json inside the zip file

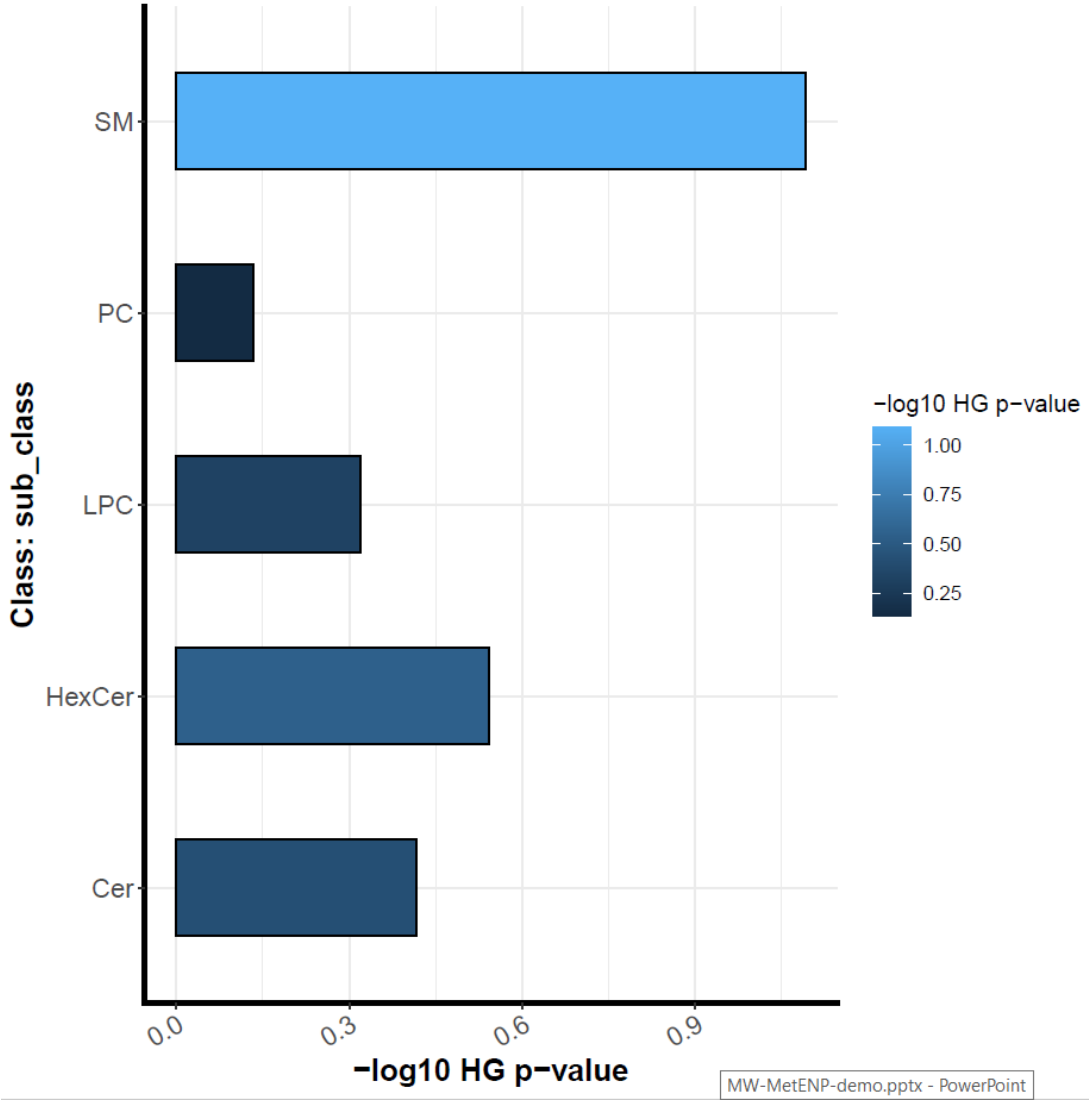
Visualization Plots





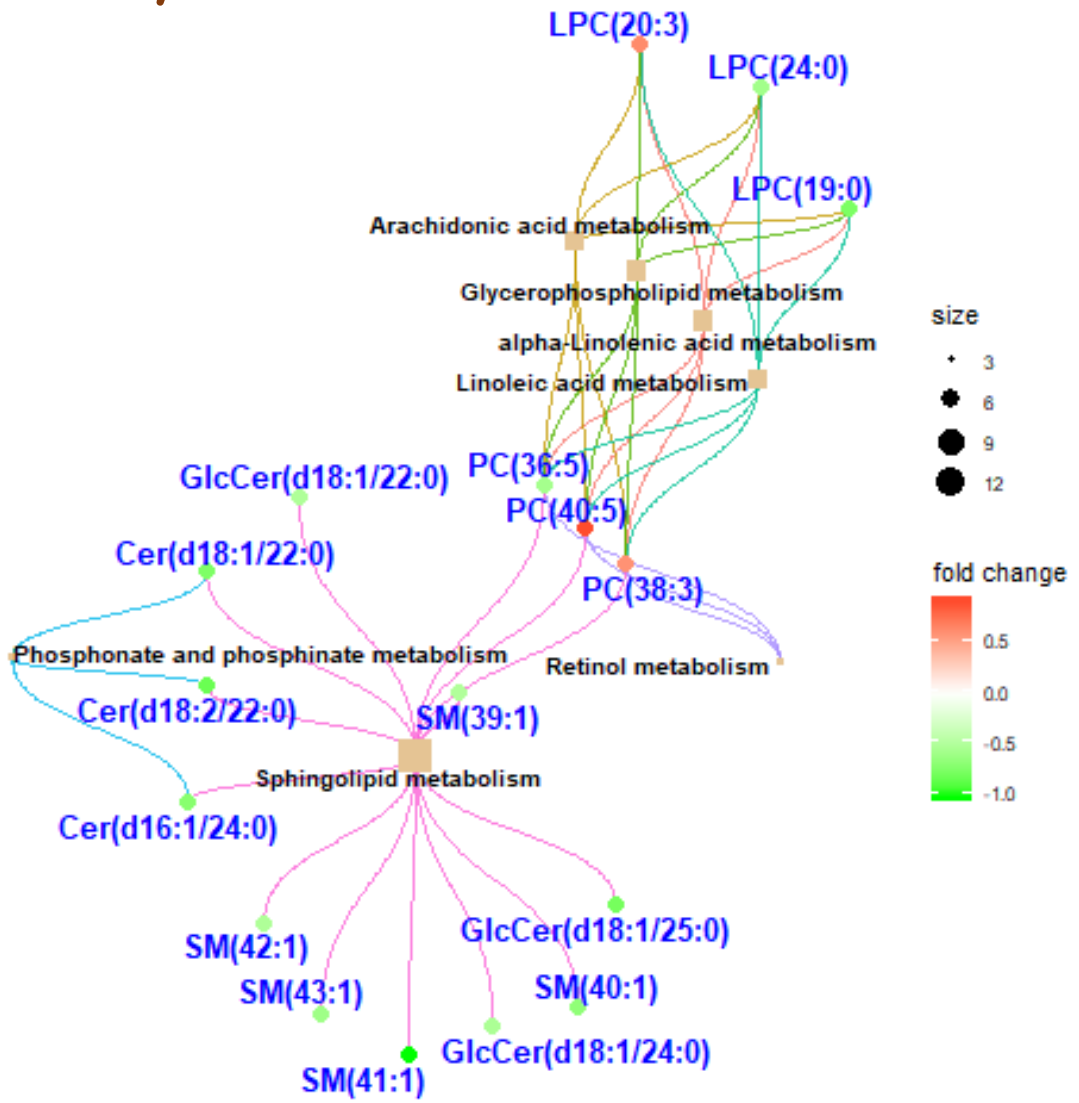
Visualization Plots

Metabolite enrichment plot

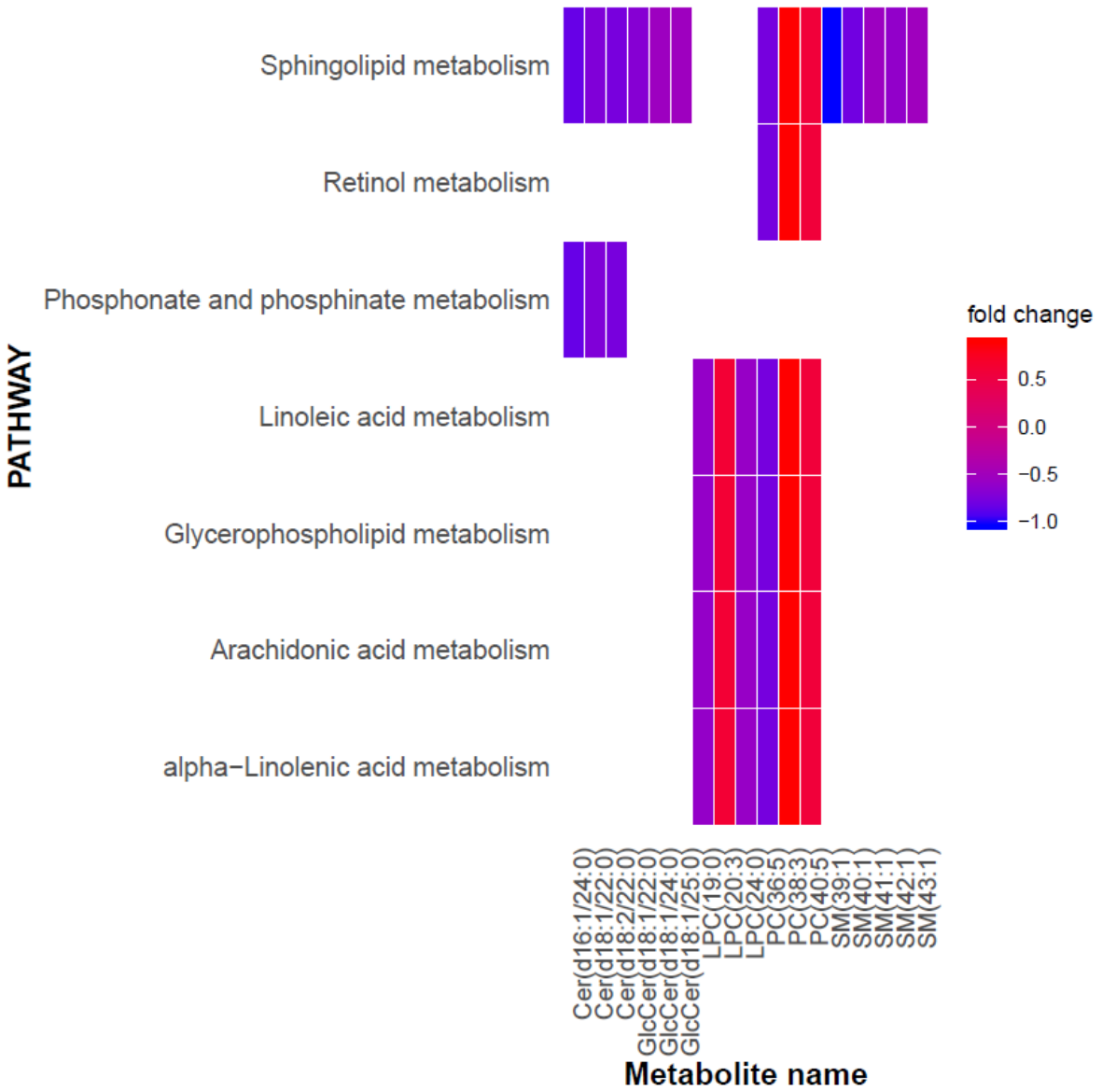


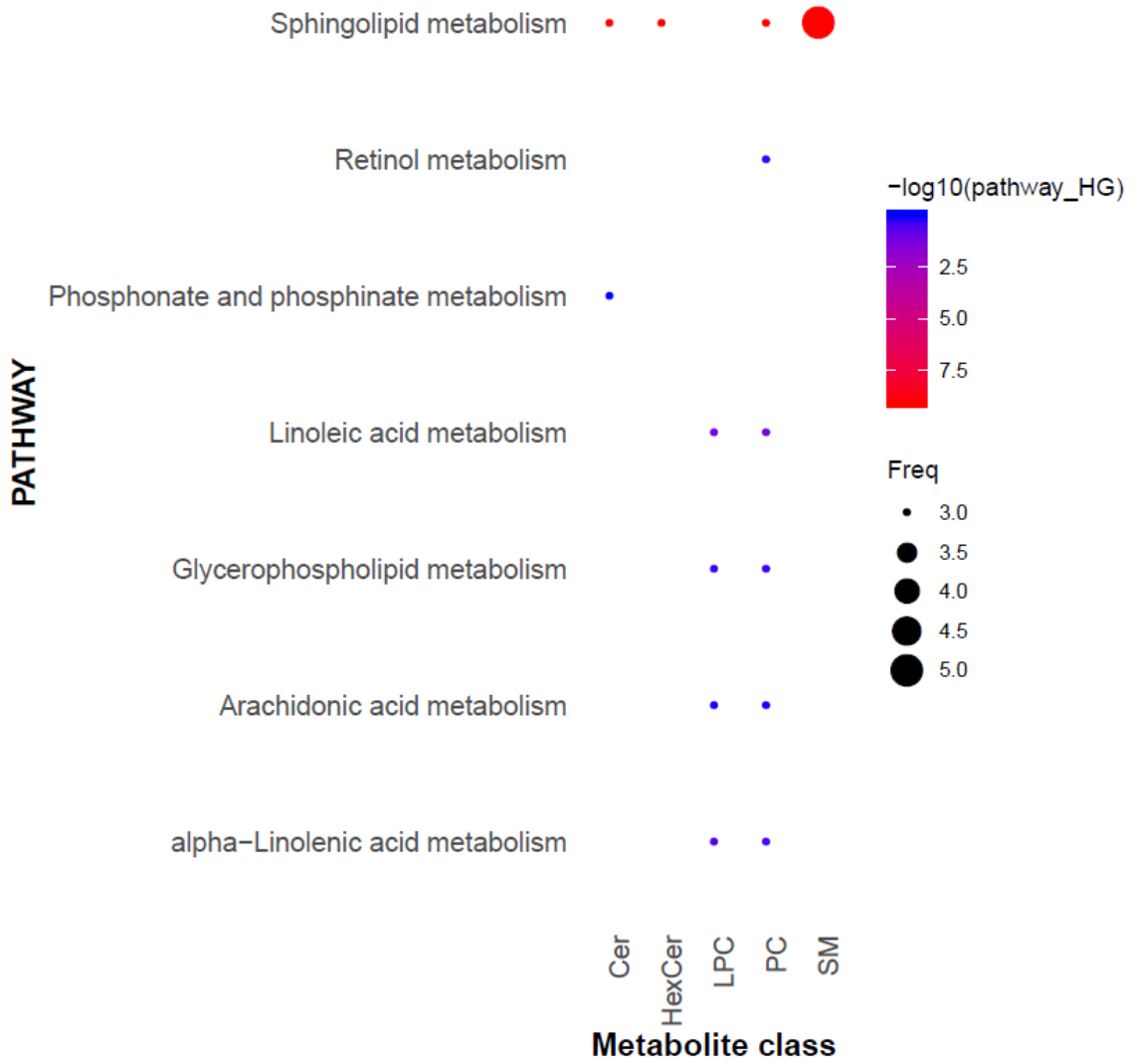
MW-MetENP-demo.pptx - PowerPoint

Pathway- Metabolite network



Visualization Plots





Other Tools

Overview | Upload / Manage Data | Browse / Search Studies | Analyze Studies | Tutorials | FAQ

Statistics Toolbox for Study: ST000915

Title: Biomarkers of NAFLD progression: a lipidomics approach to an epi

Select a dataset:
Core G Fatty acids/Eicosanoids

Run analyses on data in Study ST000915 Dataset: Core G Fatty acids/Eicosanoids

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

- Perform sample normalization / Show metabolite averages / Run cluster analysis
- Perform analyte scaling on data
- Create Relative log abundance plots

Univariate analysis

- Perform multi-condition dot plot analysis
- Perform Volcano plot analysis
- Perform ANOVA analysis

Clustering and correlation

- Perform hierarchial or heatmap cluster analysis
- Perform Clustered correlation analysis
- Perform Network analysis on correlated metabolites (mapped to classification)
- Perform Network analysis on correlated metabolites (mapped to fold-change)

Multivariate analysis

- Perform Principal component analysis
- Perform Linear discriminant analysis

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

Overview | Upload / Manage Data | Browse / Search Studies | Analyze Studies | Tutorials | FAQ

Return to Statistics Toolbox

Volcano Plot analysis for Study ST000915

(Analysis All analyses used)

Select one or more experimental factors for Groups 1 and 2. The members of each group should be DIFFERENT.

Group1	Experimental factor	Group2
☐	Diagnosis:Normal (31)	☑
☐	Diagnosis:Steatosis (17)	☐
☐	Diagnosis:NASH (20)	☐
☑	Diagnosis:Cirrhosis (20)	☐

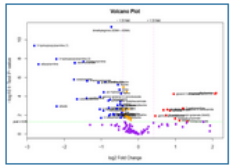
P-value cutoff: 0.1 **Fold-change cutoff:** 1.1 **Sample normalization:** Mean

Group by metabolite classification: Sub class **Use:** Submitted metabolite names

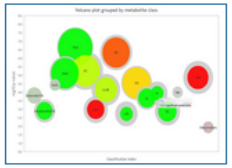
Maximum # of (most significant) metabolites per class to use in pvalue group calculation: All

Analysis: Core G Fatty acids/Eicosanoids **Combine data for all analyses?:** ☑ Run Volcano Plot

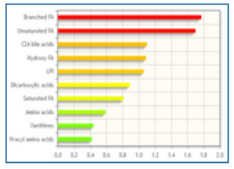
Experimental factors selected in each group			
Group1:	Diagnosis:Cirrhosis	**Group2:**	Diagnosis:Normal



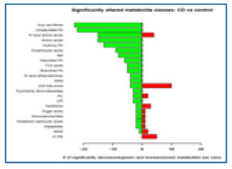
Volcano plot of all significant metabolites



Bubble plot of results grouped by metabolite class



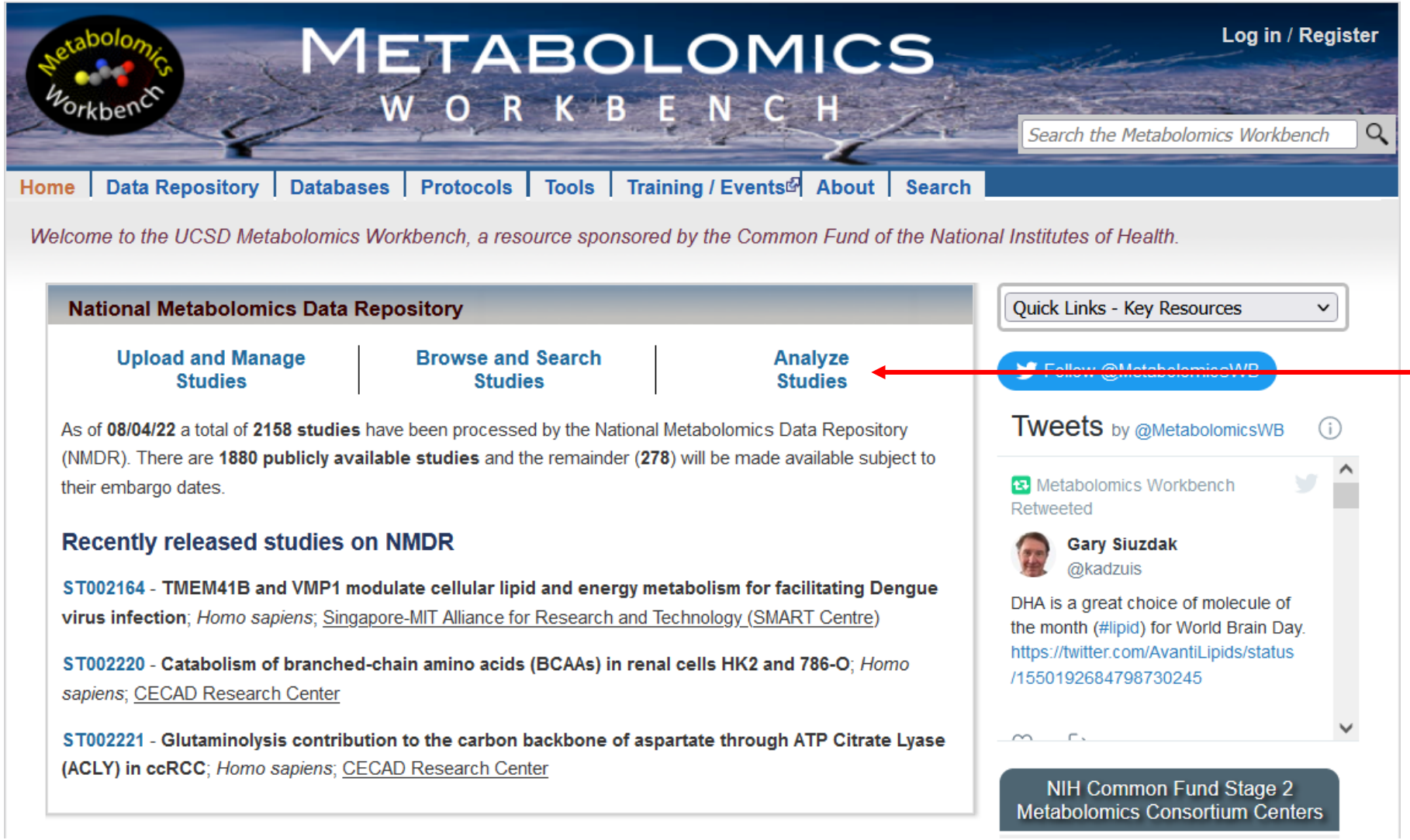
Enrichment by metabolite class



Barplot of significant metabolites by metabolite class

User dataset

Example 2. Custom dataset



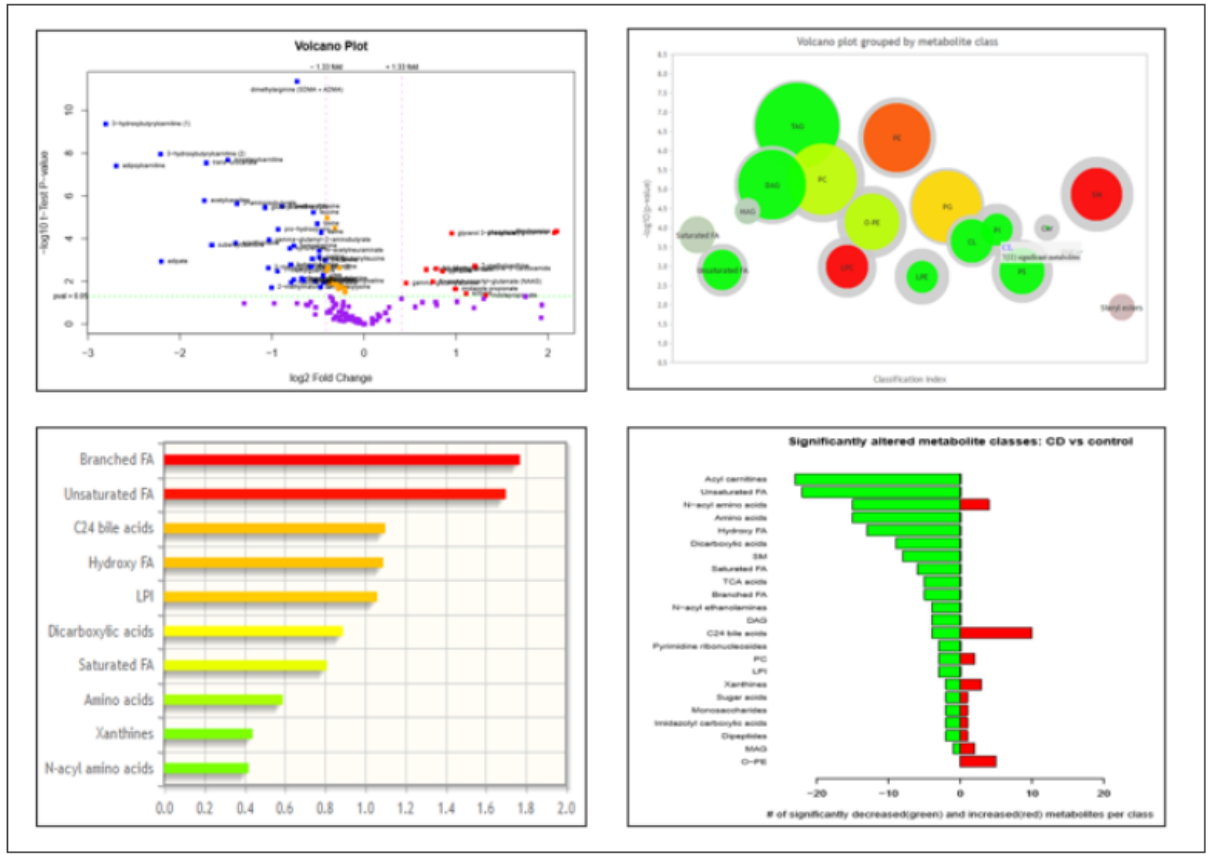
Click on Analyze Studies

https://www.metabolomicsworkbench.org/data/analyze.php

https://www.metabolomicsworkbench.org/data/analyze.php

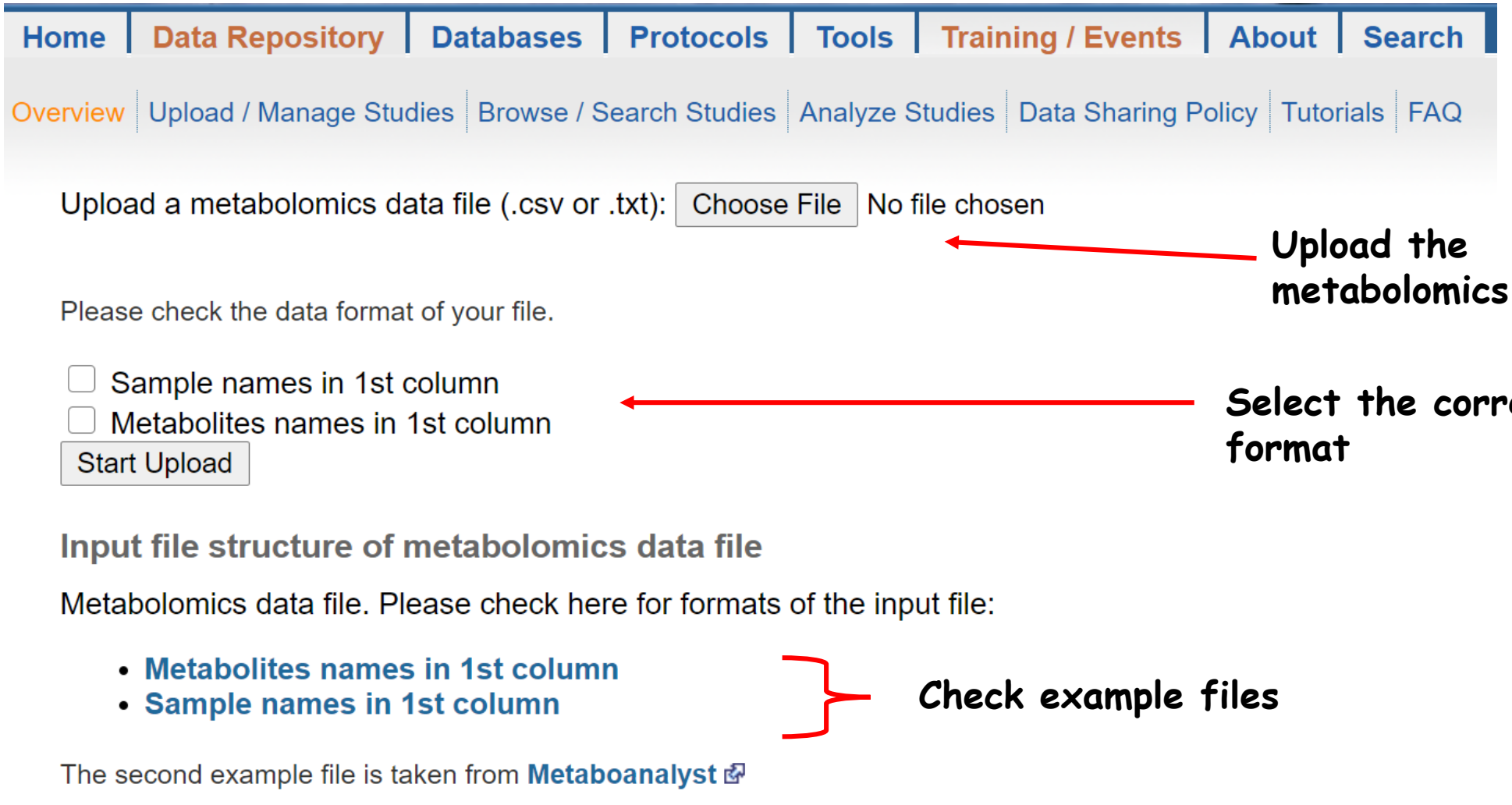
Perform data analysis on user-uploaded data

- **Load and analyze your own dataset**



- **Load and analyze your own dataset by MetENP** New! ← Click here

Data Upload

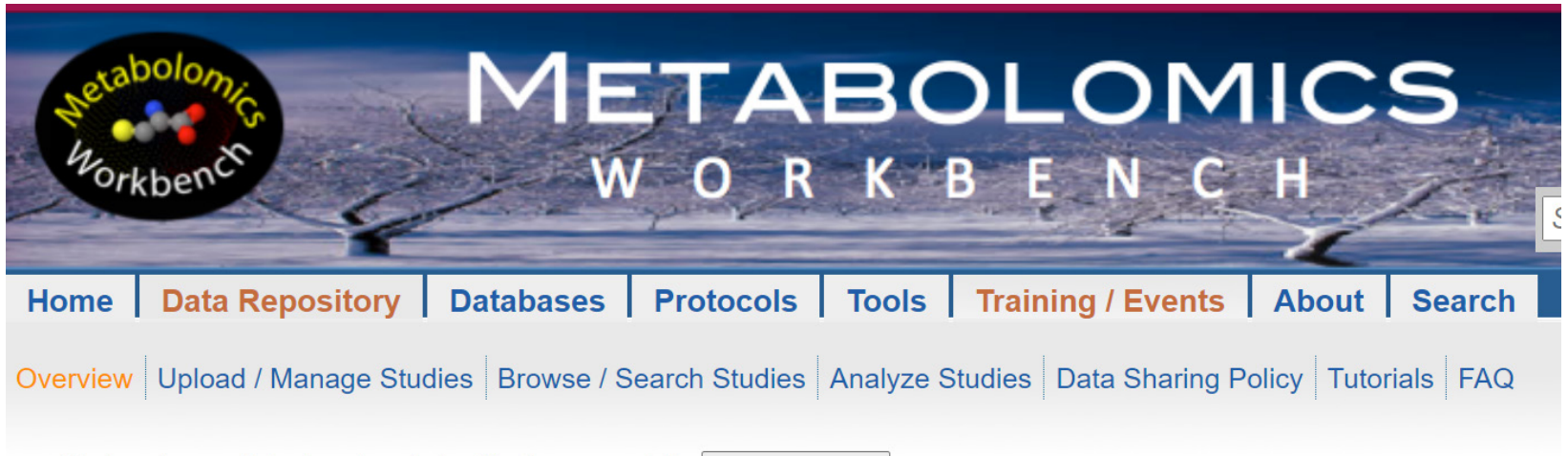


Upload the metabolomics data.

Select the correct format

Check example files

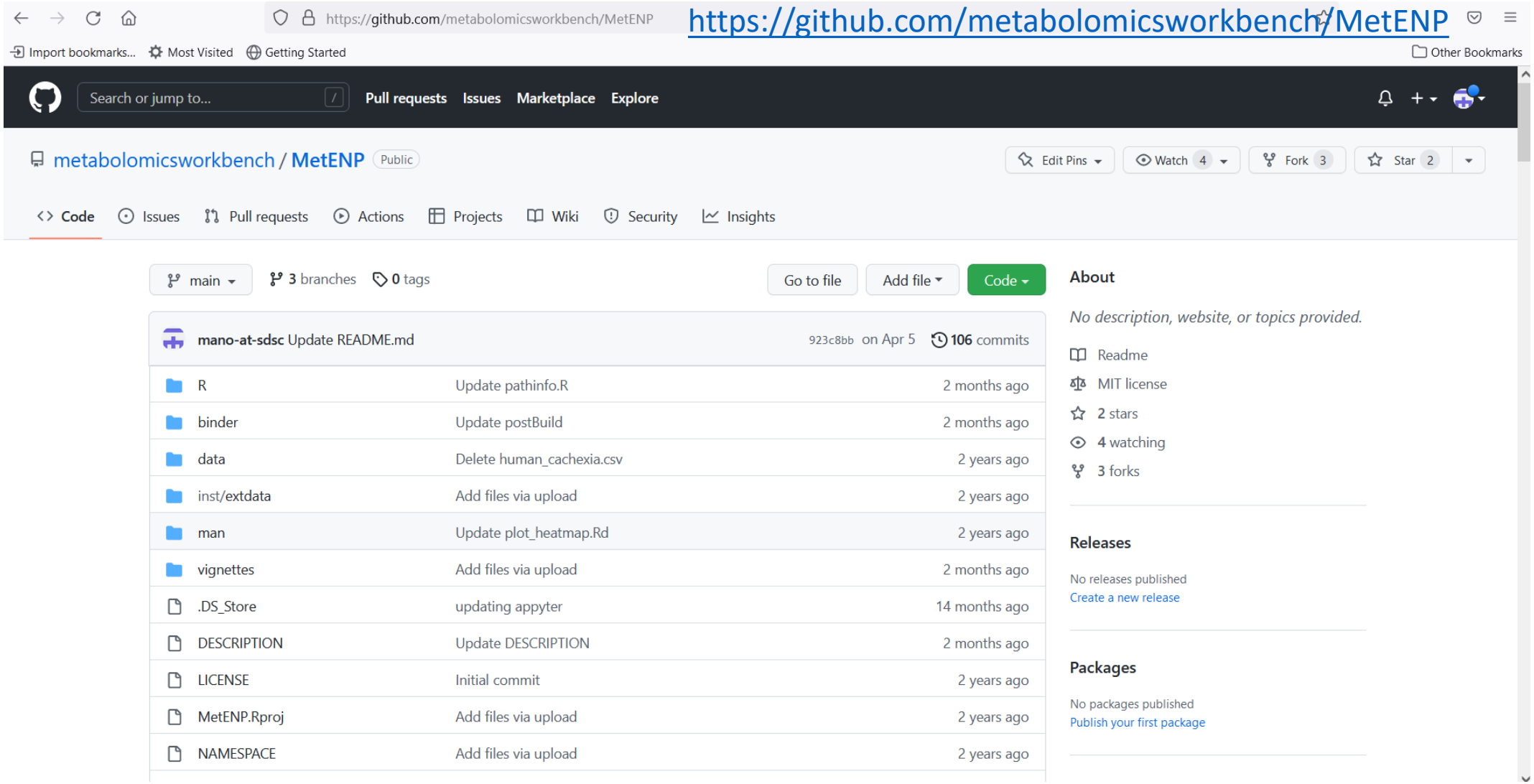
Example upload



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

MetENP Jupyter notebook available on GitHub

https://github.com/metabolomicsworkbench/MetENP

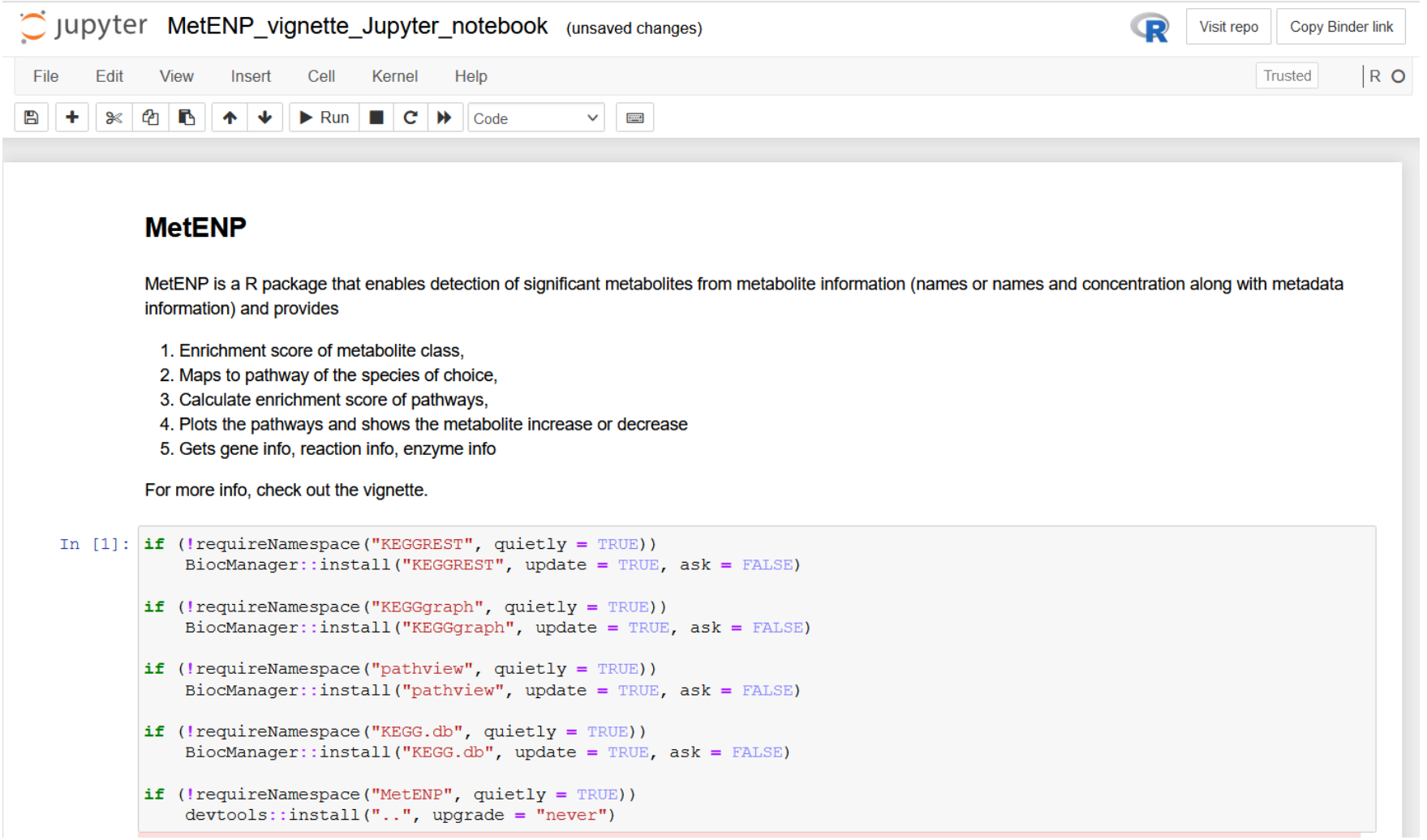


Unix command to clone MetENP: git clone https://github.com/metabolomicsworkbench/MetENP.git MetENP

MetENP on Mybinder

Search/find the studies of interest on the MW website, then use Jupyter notebook to do custom analysis related to that study.
Example study: https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST000915

https://mybinder.org/v2/gh/metabolomicsworkbench/MetENP/test01?filepath=vignettes%2FMetENP_vignette_Jupyter_notebook.ipynb



Questions

For any difficulty in running the application or other questions, contact
Mano Maurya (mano@sdsc.edu),
Sumana Srinivasan (susrinivasan@eng.ucsd.edu) or
Shankar Subramaniam (shankar@ucsd.edu).