Metabolomics Workbench and the National Metabolomics Data Repository

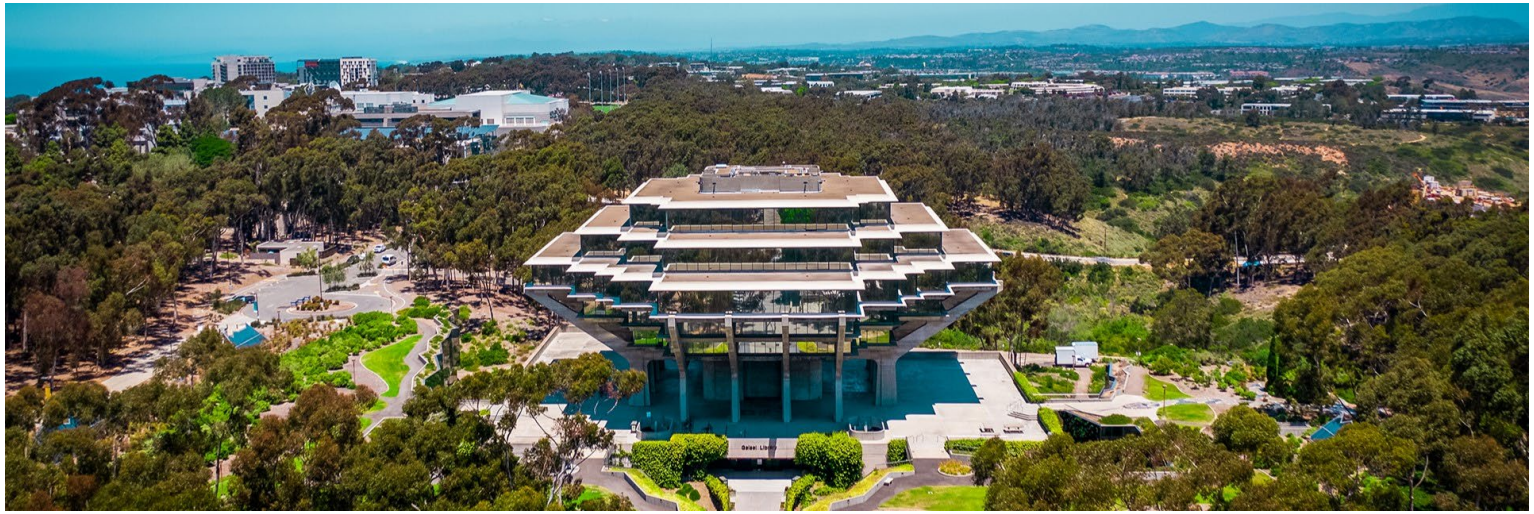
University of California San Diego
and
San Diego Supercomputer Center

Overview and Infrastructure

NIH Common Fund's National Metabolomics Data Repository
(supported by NIH grant, U2C-DK119886)

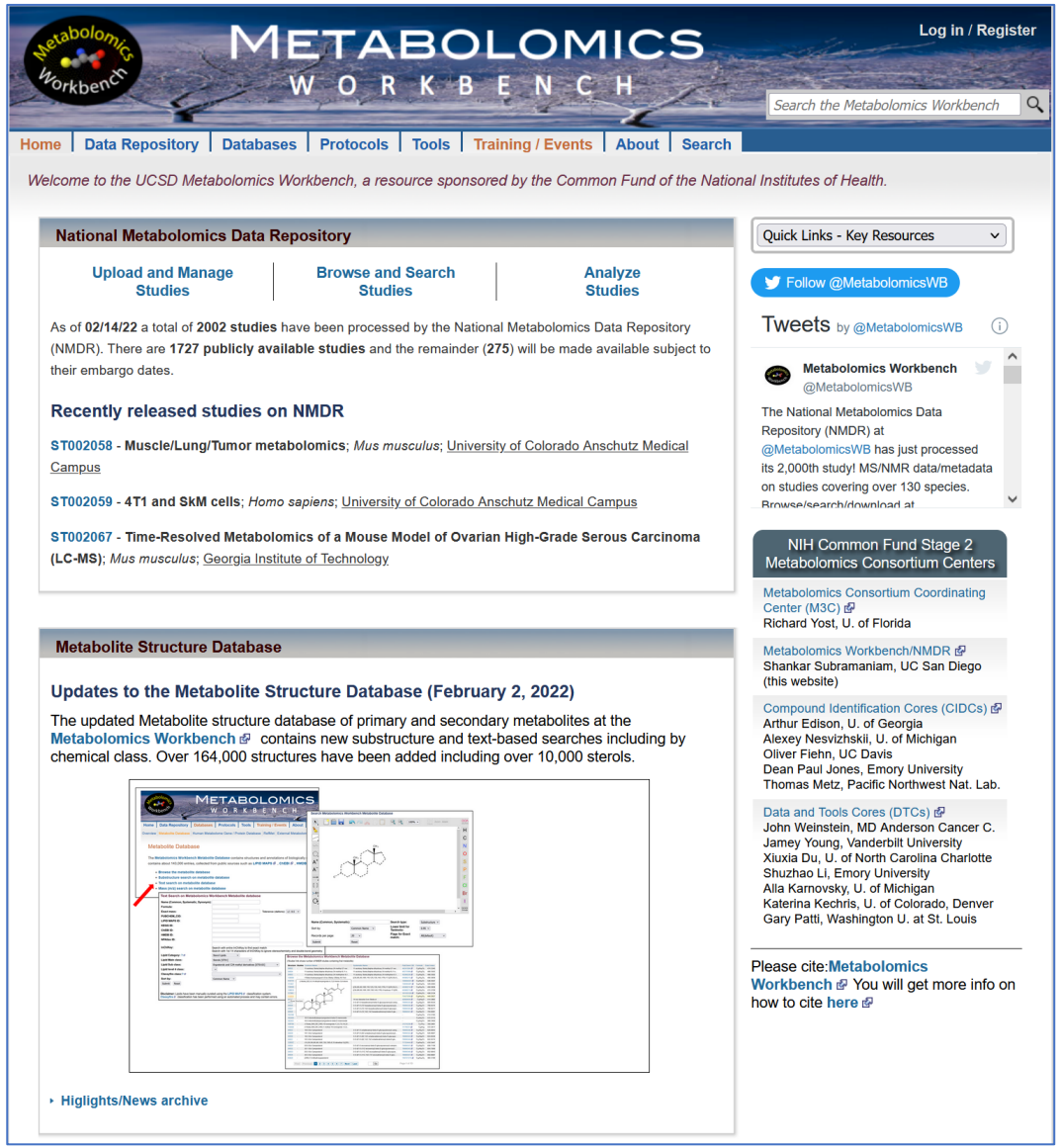
Overview of the Metabolomics Workbench

The National Institutes of Health (NIH) Common Fund Metabolomics Program was developed with the goal of increasing national capacity in metabolomics by supporting the development of next generation technologies, promoting data/metadata sharing and collaboration and providing training and mentoring opportunities.
In support of this effort, the Metabolomics Workbench website was created at the University of California, San Diego in 2013. The Metabolomics Workbench houses the National Metabolomics Data Repository (NMDR) which serves as a national and international center for metabolomics data and metadata and provides analysis tools and access to metabolite standards, protocols and other resources to the global community.



Metabolomics Workbench: https://www.metabolomicsworkbench.org

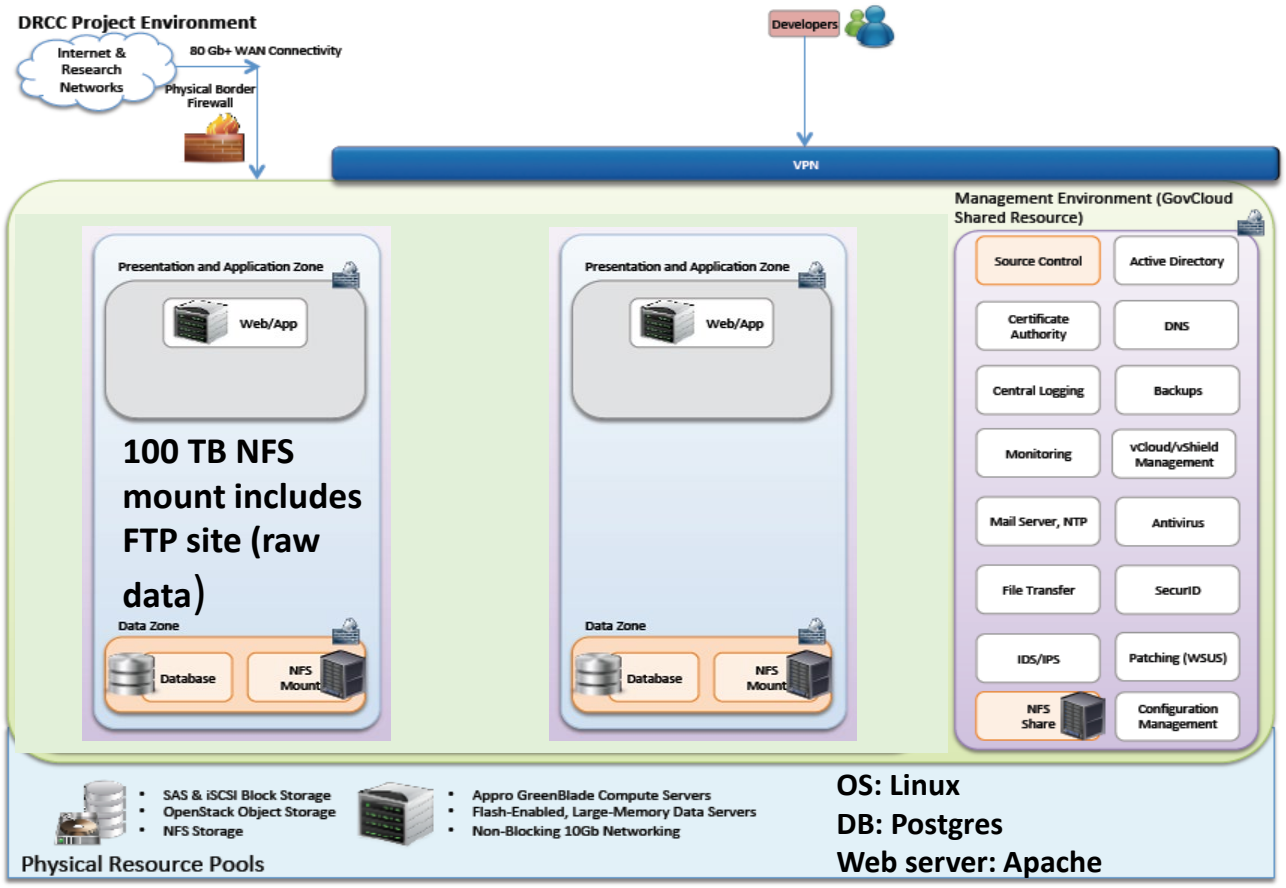
Contains the National Metabolomics Data Repository (NMDR)



Metabolomics Workbench website: what does it contain?

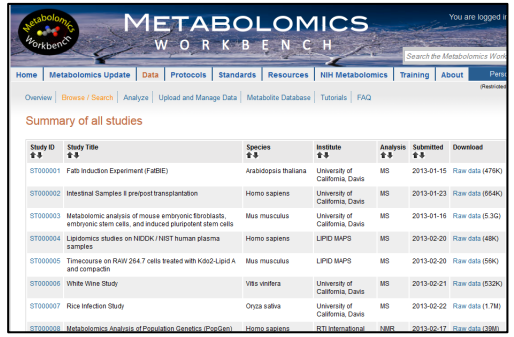
- National Metabolomics Data Repository (NMDR)
  - MS and NMR metabolomics studies
  - Metadata, targeted/untargeted measurements, raw data
- Metabolomics Workbench Metabolite database
- RefMet standardized metabolite nomenclature resource
- Online suite of statistical analysis tools
  - For NMDR studies and ad-hoc user–supplied datasets
- MetStat summary reporting tool
- Human gene/protein database of metabolism-related genes
- Protocols for metabolomics experiments
- REST service
- MS search tools
- Other metabolomics software (MW group and collaborators)

Overview of NMDR cloud computing infrastructure
(located at the San Diego Supercomputer Center)

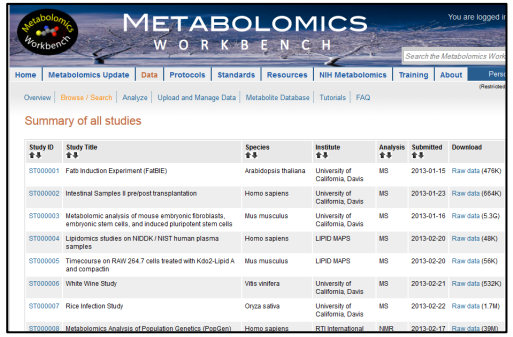


NMDR online portals

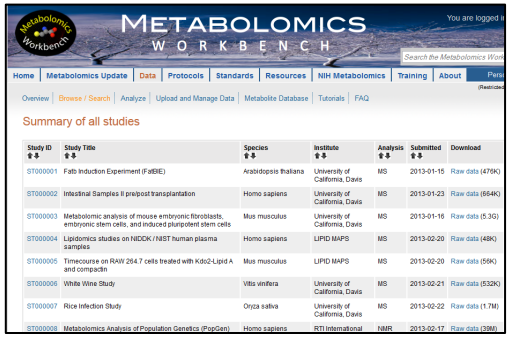
Public website

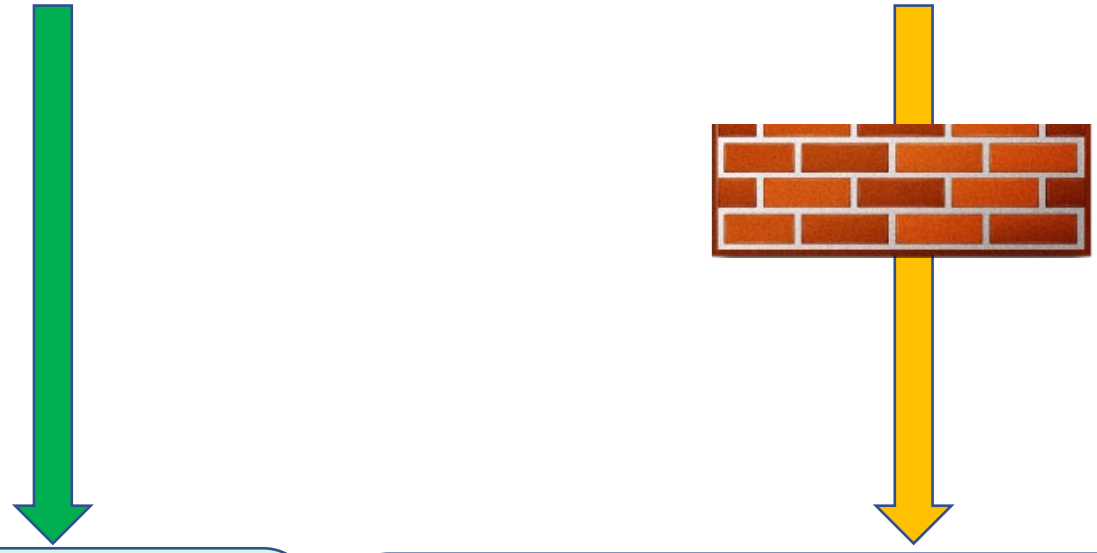


Data submission/review portal



Development websites





Open access

(non-embargoed studies)

Access controlled

Submit/View data/metadata submitted by your institution (or other institutions for which you have permission)

NMDR only

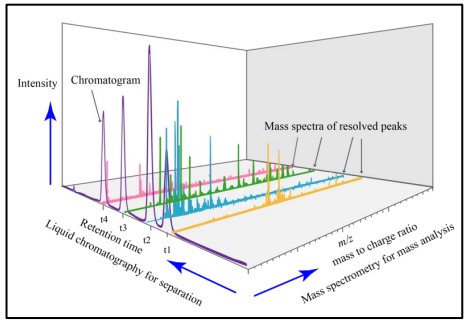
(testing/development)

Metabolomics study workflow

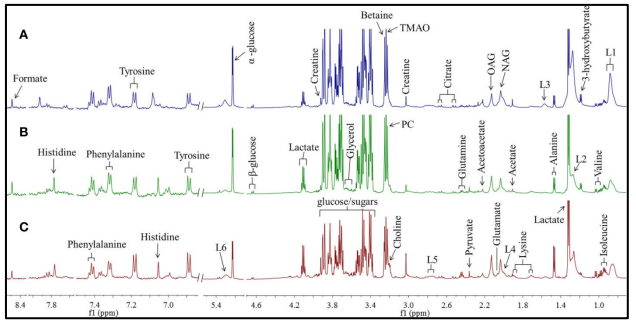
Study design

Sample prep

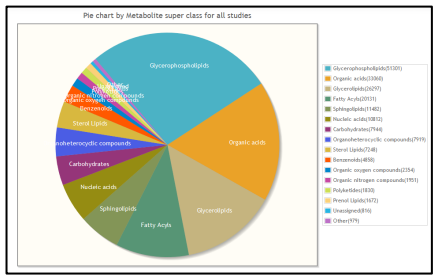
Data processing/databasing

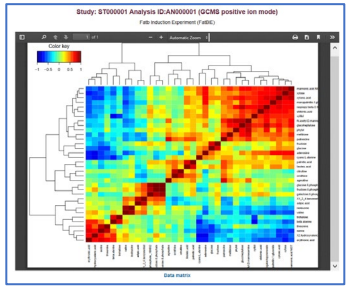


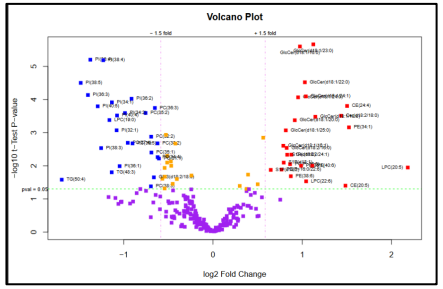
MS

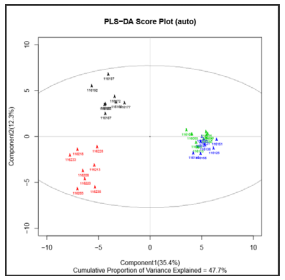


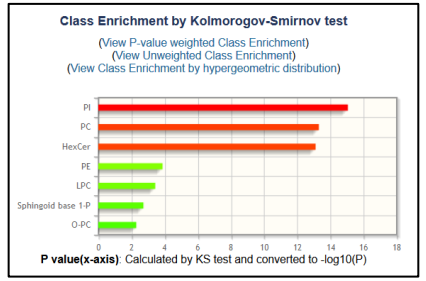
NMR

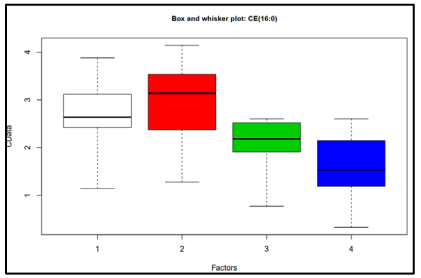








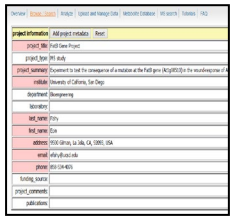




Data analysis

Overview of NMDR Inputs and Outputs

Online data submission



Metadata
Targeted data measurements
Protocols/Methods files
Untargeted data measurement files

Raw data (MS/NMR binary files) + other pertinent files

FTP

NMDR Data Repository

Browsing/searching/statistical analysis via web browser

- Targeted data (named metabolites)
- Untargeted data (m/z-ret. time features, NMR binned increments)

Download targeted measurements
Download untargeted measurements

Download Raw Data (via HTTP or FTP)

Download mwTab file (Metadata/data)

Formats:
Plain text
JSON

REST service

NMDR: Overall Strategy

Experimental data

Online study submission system

Named metabolite data/ smaller datasets (paste into web forms)

Relational Database (Data/Metadata)

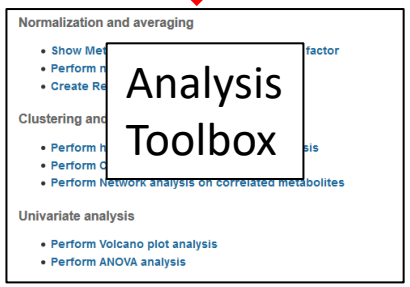
Untargeted data/ larger datasets (upload as files)

Metadata

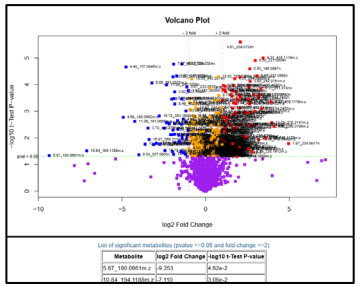
"Big data" file system: Mwtab files, Results files

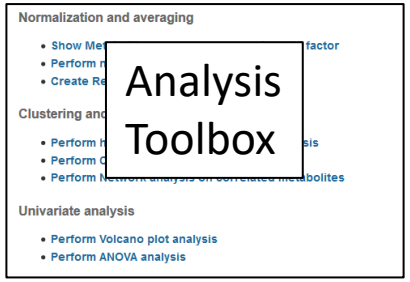
Transpose
Fix zero/null data
Average technical reps.
Apply filters

Processed text files(smaller) Save to file system for data analysis



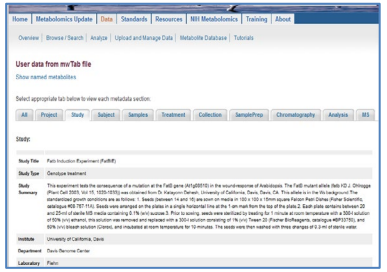
Data analysis



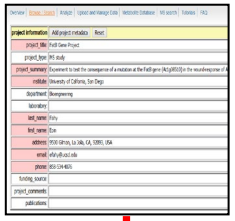


NMDR Online submission overview

Preview metadata



Online data submission forms



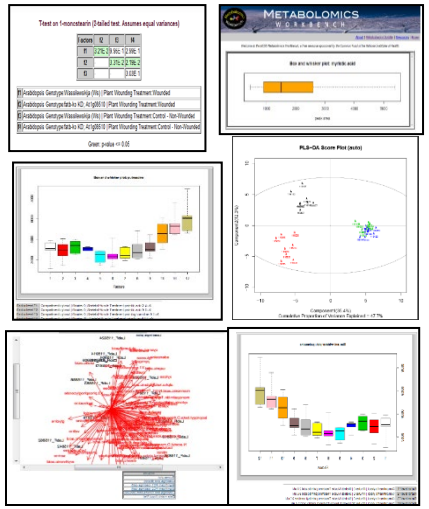
Metadata

Results

Protocol files

mwTab file

Preview data prior to submission



NMDR Curation

himal_experiment.docx	Download		
	View mwTab	Edit mwTab	Create SQL
DongHo_20160115_035137_mwtab_analysis_1.txt	View mwTab	View online	
	View mwTab	Edit mwTab	Create SQL
txt	View mwTab	View online	
	View mwTab	Edit mwTab	Create SQL
efahy_20151218_15[illegible]318_mwtab_analysis_1.txt	View mwTab	View online	

Automated SQL generation

Database

The mwTab format: A “common currency” for metadata/data sharing and storage

SECTIONS:

Metadata

Project
Study
Experimental variables (factors)
Subject
Collection
Treatment
Sample preparation
Chromatography
Analysis
MS
NMR

Data

Named metabolite measurements (table)
Named metabolites and annotations
File names for untargeted datasets

Additional

Comments preceded by a #

```
#METABOLOMICS WORKBENCH Nagireddy_Putluri_20211115_090410 DATATRACK_ID:2929 STUDY_ID:ST002005 ANALYSIS_ID:AN003268 PROJECT_ID:PR001271
VERSION                 1
CREATED_ON              November 29, 2021, 7:26 pm
#PROJECT
PR:PROJECT_TITLE                        Alterations of lipids in tumor tissues from African American and European
PR:PROJECT_TITLE                        American patient with bladder cancer
PR:PROJECT_SUMMARY                      Cancer affects all individuals in the United States, unfortunately due to
PR:PROJECT_SUMMARY                      socioeconomics, and environmental disadvantages, certain group of populations
PR:PROJECT_SUMMARY                      especially African American (AA) community bear a high burden of cancer than the
PR:PROJECT_SUMMARY                      other communities. Based on different social epidemiological study reported that
PR:PROJECT_SUMMARY                      higher incidence and mortality rate of bladder cancer in AA community. To
PR:PROJECT_SUMMARY                      understand and reveal the biological mechanism in terms of lipidomics,
PR:PROJECT_SUMMARY                      lipidomics profile were performed in 98 bladder cancer (African American and
PR:PROJECT_SUMMARY                      European America) tissues including benign.
PR:INSTITUTE                            Baylor College of Medicine
PR:LAST_NAME                            Putluri
PR:FIRST_NAME                           Nagireddy
PR:ADDRESS                              One Baylor Plaza, Houston, Texas 77030
PR:EMAIL                                putluri@bcm.edu
PR:PHONE                                (713) 798-3139
#STUDY
ST:STUDY_TITLE                          Alterations of lipids in tumor tissues from African American and European
ST:STUDY_TITLE                          American patient with bladder cancer
ST:STUDY_SUMMARY                        Cancer affects all individuals in the United States, unfortunately due to
ST:STUDY_SUMMARY                        socioeconomics, and environmental disadvantages, certain group of populations
ST:STUDY_SUMMARY                        especially African American (AA) community bear a high burden of cancer than the
ST:STUDY_SUMMARY                        other communities. Based on different social epidemiological study reported that
ST:STUDY_SUMMARY                        higher incidence and mortality rate of bladder cancer in AA community. To
ST:STUDY_SUMMARY                        understand and reveal the biological mechanism in terms of lipidomics,
ST:STUDY_SUMMARY                        lipidomics profile were performed in 98 bladder cancer (African American and
ST:STUDY_SUMMARY                        European America) tissues including benign.
ST:INSTITUTE                            Baylor College of Medicine
ST:DEPARTMENT                           Molecular and Cellular Biology
ST:LAST_NAME                            Putluri
ST:FIRST_NAME                           Nagireddy
ST:ADDRESS                              One Baylor Plaza, Houston, Texas 77030
ST:EMAIL                                putluri@bcm.edu
ST:PHONE                                (713) 798-3139
#SUBJECT
SU:SUBJECT_TYPE                         Human
SU:SUBJECT_SPECIES                      Homo sapiens
SU:TAXONOMY_ID                          9606
#FACTORS
#SUBJECT_SAMPLE_FACTORS:                SUBJECT(optional)[tab]SAMPLE[tab]FACTORS(NAME:VALUE pairs separated by |)[tab]Raw file names an
SUBJECT_SAMPLE_FACTORS                  -       03-22-2021-Pos-34734 BENIGN-30  Group:Benign
SUBJECT_SAMPLE_FACTORS                  -       03-22-2021-Pos-34880 BENIGN-32  Group:Benign
SUBJECT_SAMPLE_FACTORS                  -       03-22-2021-Pos-35088 BENIGN-36  Group:Benign
SUBJECT_SAMPLE_FACTORS                  -       03-22-2021-Pos-35603 BENIGN-38  Group:Benign
SUBJECT_SAMPLE_FACTORS                  -       03-22-2021-Pos-35599 BENIGN-40  Group:Benign
SUBJECT_SAMPLE_FACTORS                  -       03-22-2021-Pos-35774 BENIGN-42  Group:Benign
```

MwTab specification: https://www.metabolomicsworkbench.org/data/mwTab_specification.pdf