

# **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)

The screenshot shows the Metabolomics Workbench website. At the top, there is a navigation bar with links for Home, Data Repository, Databases, Protocols, Tools, Training / Events, About, and Search. A search bar is located on the right side of the header. Below the navigation bar, a welcome message states: "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 several sections. On the left, there is a section for the National Metabolomics Data Repository (NMDR) with three sub-sections: Upload and Manage Studies, Browse and Search Studies, and Analyze Studies. Below these, a paragraph states: "As of 02/14/22 a total of 2002 studies have been processed by the National Metabolomics Data Repository (NMDR). There are 1727 publicly available studies and the remainder (275) will be made available subject to their embargo dates." This is followed by a section titled "Recently released studies on NMDR" with three entries: ST002058 - Muscle/Lung/Tumor metabolomics; Mus musculus; University of Colorado Anschutz Medical Campus; ST002059 - 4T1 and SKM cells; Homo sapiens; University of Colorado Anschutz Medical Campus; and ST002067 - Time-Resolved Metabolomics of a Mouse Model of Ovarian High-Grade Serous Carcinoma (LC-MS); Mus musculus; Georgia Institute of Technology. On the right side, there is a "Quick Links - Key Resources" dropdown menu, a "Follow @MetabolomicsWB" button, and a "Tweets by @MetabolomicsWB" section. Below the tweets, there is a section for "NIH Common Fund Stage 2 Metabolomics Consortium Centers" with a list of centers and their coordinators. At the bottom of the page, there is a section for "Metabolite Structure Database" with a sub-section titled "Updates to the Metabolite Structure Database (February 2, 2022)". This section contains a paragraph: "The updated Metabolite structure database of primary and secondary metabolites at the Metabolomics Workbench contains new substructure and text-based searches including by chemical class. Over 164,000 structures have been added including over 10,000 sterols." Below this text is a screenshot of the Metabolite Structure Database interface, showing a search results page with a table of metabolites and their structures. At the bottom of the page, there is a link to "Highlights/News archive".

**National Metabolomics Data Repository**

Upload and Manage Studies | Browse and Search Studies | Analyze Studies

As of 02/14/22 a total of 2002 studies have been processed by the National Metabolomics Data Repository (NMDR). There are 1727 publicly available studies and the remainder (275) will be made available subject to their embargo dates.

**Recently released studies on NMDR**

**ST002058** - Muscle/Lung/Tumor metabolomics; *Mus musculus*; University of Colorado Anschutz Medical Campus

**ST002059** - 4T1 and SKM cells; *Homo sapiens*; University of Colorado Anschutz Medical Campus

**ST002067** - Time-Resolved Metabolomics of a Mouse Model of Ovarian High-Grade Serous Carcinoma (LC-MS); *Mus musculus*; Georgia Institute of Technology

**Metabolite Structure Database**

**Updates to the Metabolite Structure Database (February 2, 2022)**

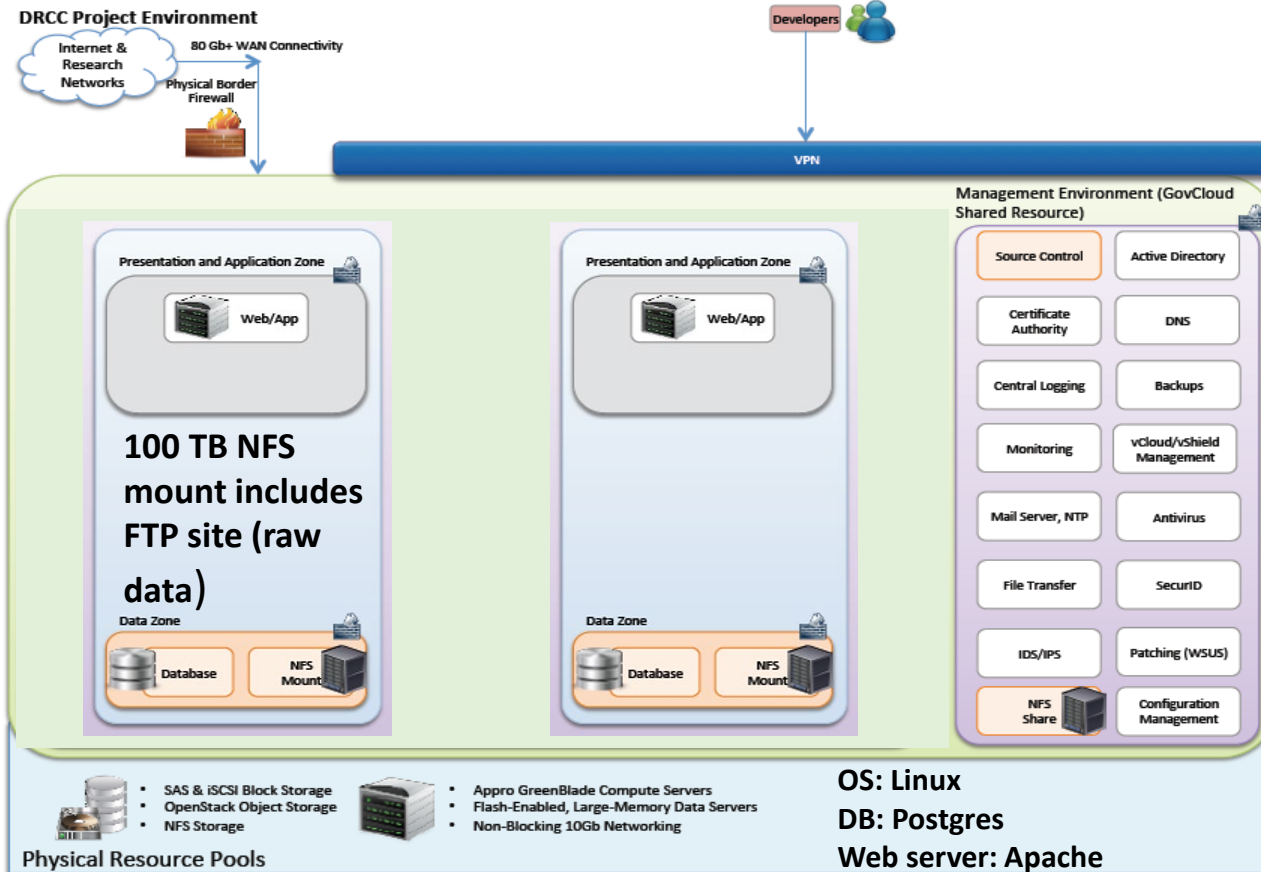
The updated Metabolite structure database of primary and secondary metabolites at the Metabolomics Workbench contains new substructure and text-based searches including by chemical class. Over 164,000 structures have been added including over 10,000 sterols.

Highlights/News archive

# 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

Study ID	Study Title	Species	Institute	Analysis	Submitted	Download
ST000001	Fab Induction Experiment (FABE)	Arabidopsis thaliana	University of California, Davis	MS	2013-01-15	Raw data (479K)
ST000002	Intestinal Samples II prepost transplantation	Homo sapiens	University of California, Davis	MS	2013-01-23	Raw data (654K)
ST000003	Metabonomic analysis of mouse embryonic fibroblasts, embryonic stem cells, and induced pluripotent stem cells	Mus musculus	University of California, Davis	MS	2013-01-16	Raw data (5.3G)
ST000004	Lipidomics studies on NIDDK/NIST human plasma samples	Homo sapiens	LIPID MAPS	MS	2013-02-20	Raw data (48K)
ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compound	Mus musculus	LIPID MAPS	MS	2013-02-20	Raw data (59K)
ST000006	White Wine Study	Vitis vinifera	University of California, Davis	MS	2013-02-21	Raw data (532K)
ST000007	Rice Infection Study	Oryza sativa	University of California, Davis	MS	2013-02-22	Raw data (1.7M)
ST000008	Metabonomics Analysis of Population Genetics (MapGen)	Homo sapiens	RTI International	FMR	2013-02-17	Raw data (120M)

Study ID	Study Title	Species	Institute	Analysis	Submitted	Download
ST000001	Fab Induction Experiment (FABE)	Arabidopsis thaliana	University of California, Davis	MS	2013-01-15	Raw data (479K)
ST000002	Intestinal Samples II prepost transplantation	Homo sapiens	University of California, Davis	MS	2013-01-23	Raw data (654K)
ST000003	Metabonomic analysis of mouse embryonic fibroblasts, embryonic stem cells, and induced pluripotent stem cells	Mus musculus	University of California, Davis	MS	2013-01-16	Raw data (5.3G)
ST000004	Lipidomics studies on NIDDK/NIST human plasma samples	Homo sapiens	LIPID MAPS	MS	2013-02-20	Raw data (48K)
ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compound	Mus musculus	LIPID MAPS	MS	2013-02-20	Raw data (59K)
ST000006	White Wine Study	Vitis vinifera	University of California, Davis	MS	2013-02-21	Raw data (532K)
ST000007	Rice Infection Study	Oryza sativa	University of California, Davis	MS	2013-02-22	Raw data (1.7M)
ST000008	Metabonomics Analysis of Population Genetics (MapGen)	Homo sapiens	RTI International	FMR	2013-02-17	Raw data (120M)

Study ID	Study Title	Species	Institute	Analysis	Submitted	Download
ST000001	Fab Induction Experiment (FABE)	Arabidopsis thaliana	University of California, Davis	MS	2013-01-15	Raw data (479K)
ST000002	Intestinal Samples II prepost transplantation	Homo sapiens	University of California, Davis	MS	2013-01-23	Raw data (654K)
ST000003	Metabonomic analysis of mouse embryonic fibroblasts, embryonic stem cells, and induced pluripotent stem cells	Mus musculus	University of California, Davis	MS	2013-01-16	Raw data (5.3G)
ST000004	Lipidomics studies on NIDDK/NIST human plasma samples	Homo sapiens	LIPID MAPS	MS	2013-02-20	Raw data (48K)
ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compound	Mus musculus	LIPID MAPS	MS	2013-02-20	Raw data (59K)
ST000006	White Wine Study	Vitis vinifera	University of California, Davis	MS	2013-02-21	Raw data (532K)
ST000007	Rice Infection Study	Oryza sativa	University of California, Davis	MS	2013-02-22	Raw data (1.7M)
ST000008	Metabonomics Analysis of Population Genetics (MapGen)	Homo sapiens	RTI International	FMR	2013-02-17	Raw data (120M)



**Open access**

**(non-embargoed studies)**

**Access controlled**

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

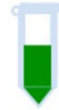
**NMDR only**

**(testing/development)**

# Metabolomics study workflow



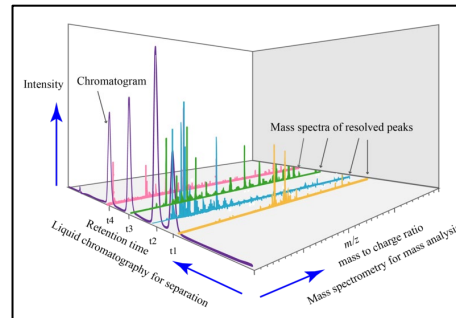
Study design



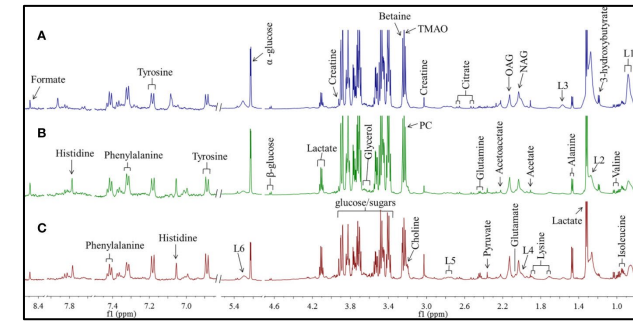
Sample prep



Data processing/databasing



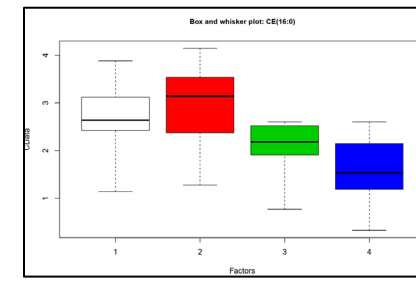
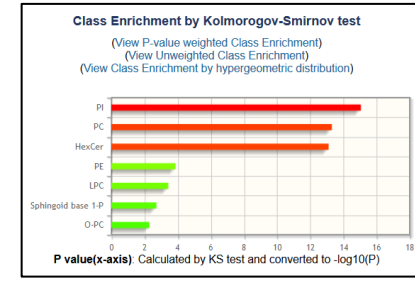
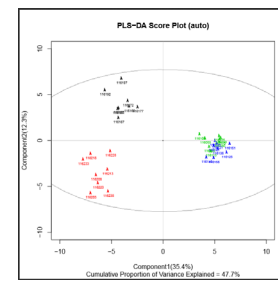
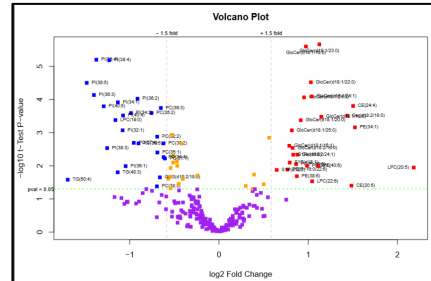
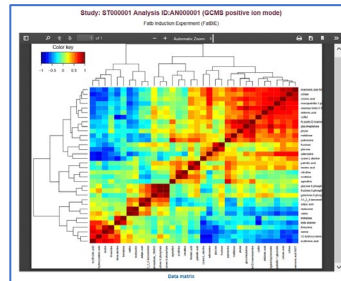
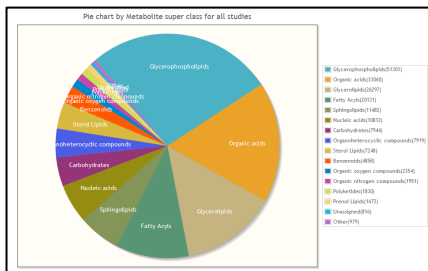
MS



NMR



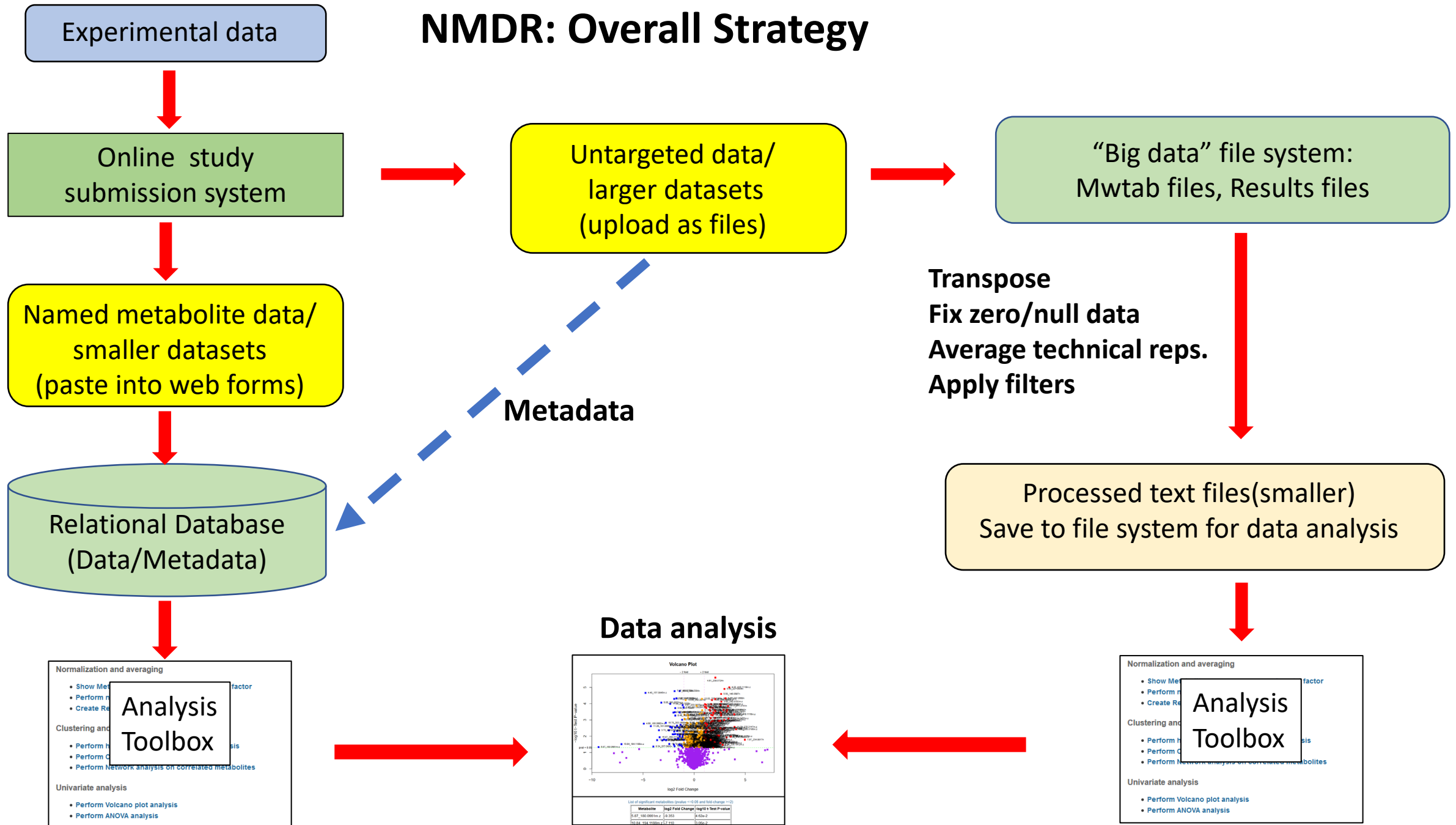
Data analysis





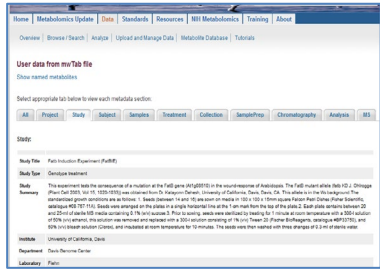


# NMDR: Overall Strategy



# NMDR Online submission overview

## Preview metadata



## Online data submission forms

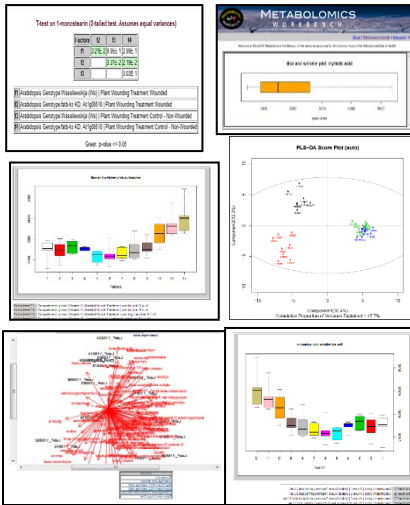
Project ID	1000000001
Project Name	1000000001
Project Description	1000000001
Project PI	1000000001
Project Co-PI	1000000001
Project Address	1000000001
Project City	1000000001
Project State	1000000001
Project Country	1000000001
Project Zip	1000000001
Project Institution	1000000001
Project Department	1000000001
Project Title	1000000001
Project Position	1000000001
Project Role	1000000001
Project Organization	1000000001
Project Department	1000000001
Project City	1000000001
Project State	1000000001
Project Country	1000000001
Project Zip	1000000001
Project Institution	1000000001
Project Department	1000000001
Project Title	1000000001
Project Position	1000000001
Project Role	1000000001
Project Organization	1000000001

- ← Metadata
- ← Results
- ← Protocol files

**mwTab file**



## Preview data prior to submission



**NMDR Curation**

**Automated SQL generation**



animal_experiment.docx	Download		
DongHo_20160115_035137_mwtab_analysis_1.txt	View mwTab	Edit mwTab	Create SQL
efahy_20151218_150318_mwtab_analysis_1.txt	View mwTab	View online	
	View mwTab	Edit mwTab	Create SQL
	View mwTab	View online	
	View mwTab	Edit mwTab	Create SQL
	View mwTab	View online	

# 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 socioeconomic, 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 socioeconomic, 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 and
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
```