


RTI International



# NMR Data Analysis Exercise

## UAB Metabolomics Training Course

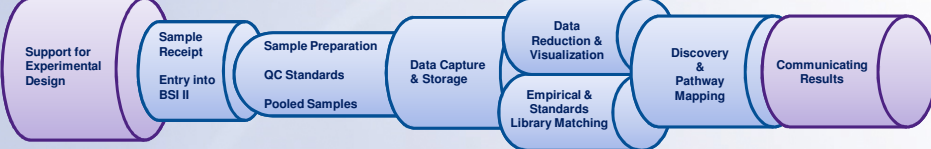
### July 17-21, 2016

Wimal Pathmasiri, Rodney Snyder  
NIH Eastern Regional Comprehensive Metabolomics Resource Core  
(RTI RCMRC)

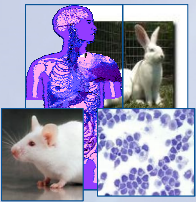
RTI International is a trade name of Research Triangle Institute.

[www.rti.org](http://www.rti.org)

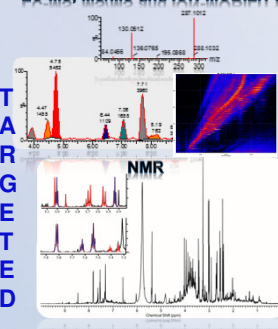
### NIH Eastern Regional Comprehensive Metabolomics Resource Core at RTI



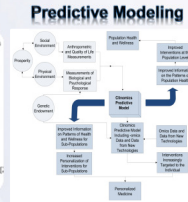
**TARGETED**



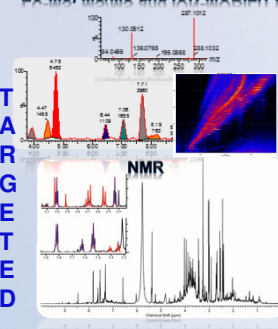
**BROAD SPECTRUM**



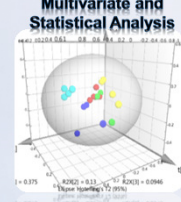
**Predictive Modeling**




**LC-MS, MS/MS and ION-MOBILITY**



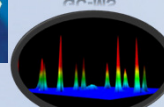
**Multivariate and Statistical Analysis**




**Pathway Mapping  
Metabolites-Proteins-Genes**




**GC-MS**

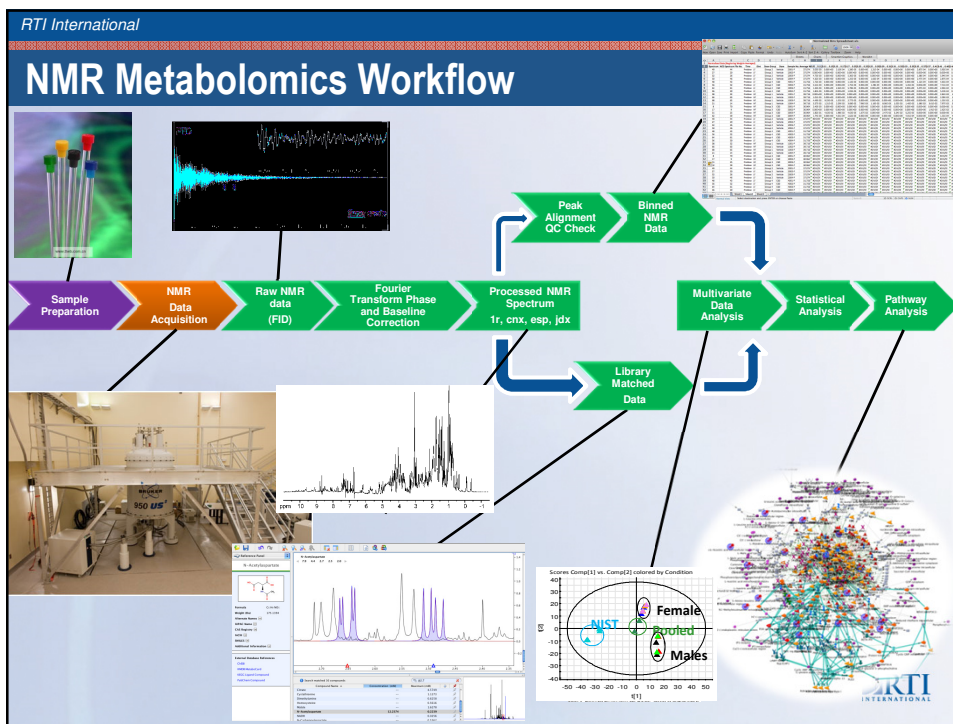


**ICP-MS**



WOMEN'S HEALTH • DRUG DISCOVERY • ENVIRONMENT • URIDLOGY • IMMUNOLOGY • OBESITY





RTI International

## NMR Metabolomics

- Broad Spectrum
  - High throughput
  - NMR Binning
  - Multivariate analysis and other statistics
  - Identifying bins important for separating study groups
  - Library matching of bins to metabolites
- Targeted Metabolomics
  - Identifying a set of metabolites
  - Quantifying metabolites
  - Multivariate analysis and other statistics
- Pathway analysis
  - Use identified metabolites
  - Use other omics data for integrated analysis

RTI INTERNATIONAL

## Some Software available for NMR Based Metabolomics

### FREE

- NMR Data Processing
  - ACD Software for Academics (ACD Labs, Toronto, Canada)
- Multivariate data analysis
  - MetaboAnalyst 3.0 (<http://www.metaboanalyst.ca>)
  - MetATT (<http://metatt.metabolomics.ca/MetATT/>)
  - MUMA (<http://www.biomolnmr.org/software.html>)
  - Other R-packages
- Library matching and Identification
  - BATMAN
  - Use of databases
    - Birmingham Metabolite library, HMDB, BMRB
- Pathway analysis
  - Metaboanalyst, metaP Server, Met-PA, Cytoscape, KEGG, IMPALA

Also available through [www.metabolomicsworkbench.org](http://www.metabolomicsworkbench.org)



## Some Software Available for NMR Based Metabolomics

### COMMERCIAL

- NMR Data-preprocessing
  - ACD Software (ACD Labs, Toronto, Canada)
  - Chenomx NMR Suite 8.1 Professional
- Multivariate data analysis
  - SIMCA 14
- Other statistical analysis
  - SAS, SPSS
- Library matching and quantification
  - Chenomx NMR Suite 8.1 Professional
- Pathway analysis
  - GeneGo (MetaCore Module)
  - Ingenuity Pathway Analysis (IPA)



## NMR Hands On Exercise

- Drug Induced Liver Injury (DILI) Study using Rat Model
- 3 Study groups and 2 time points
  - Vehicle Control (time matched)
  - Low Dose (“no effect” level, Day 01 and Day 14)
  - High Dose (Day 01 and Day 14)
- 24h Urine collected
- Samples prepared by mixing an aliquot of urine with Phosphate buffer + Chenomx ISTD (DSS, D<sub>2</sub>O, and Imidazole)
  - DSS (Chemical shift and line shape reference)
  - Imidazole (pH reference)



## Binned Data

- Three (3) Spreadsheets provided
  1. UAB\_RFA\_Metaboanalyst.csv
  2. UAB\_RFA\_Metaboanalyst\_D14\_NoPools.csv
  3. UAB\_RFA\_Metaboanalyst\_D14\_Vehicle\_vs\_HighDose.csv
- Spreadsheets 2-3 were derived from the initial spreadsheet no. 1 (for easy upload into Metaboanalyst in the subsequent analyses)



RTI International

Please go to the webpage:  
<http://www.metaboanalyst.ca/MetaboAnalyst/>

# MetaboAnalyst 3.0

**MetaboAnalyst 3.0**  
 – a comprehensive tool suite for metabolomic data analysis

Home  
 Overview  
 Data Formats  
 FAQs  
 Tutorials  
 Resources  
 Update History  
 User Stats  
 Contact  
 About

TMIC  
 hmp

Welcome [click here to start](#) >> [access old version](#)

**News & Updates**

- Updated the **confidence interval** graphics for both chemometrics and ROC curves. (01/06/2015) **NEW**
- Updated the **Heatmaps** function for better visualization of large data. (12/22/2014)
- Added a new module for **Integrated Pathway Analysis** on genes and metabolites that have both changed significantly under the same experimental conditions. (12/17/2014)
- Added a new module for **Biomarker Analysis**. (12/12/2014)
- Added sorting and filtering support in the feature details table. (11/12/2014)
- Added new functions to support **Interactive 3D PCA and PLS-DA** visualization. (10/31/2014)
- Added a new module on **Power Analysis** to support sample size and power analysis for pilot metabolomic studies. (10/30/2014)

[Read more](#)

**Please Cite:**

Xia, J., Mandal, R., Simeonkov, I., Broadhurst, D., and Wishart, D.S. (2012) [MetaboAnalyst 2.0 – a comprehensive server for metabolomic data analysis](#). *Nucl. Acids Res.* 40, W127-W133.

Xia, J., Psychogiou, N., Young, N., and Wishart, D.S. (2009) [MetaboAnalyst: a web server for metabolomic data analysis and interpretation](#). *Nucl. Acids Res.* 37, W652-660.

**Project objective:** To provide a user-friendly, web-based analytical pipeline for high-throughput metabolomics studies. In particular, MetaboAnalyst aims to offer a variety of commonly used procedures for metabolomic data processing, normalization, multivariate statistical analysis, as well as data annotation. The current implementation focuses on exploratory statistical analysis, functional interpretation, and advanced statistics for translational metabolomics studies.

**Data formats:** Diverse data types from current metabolomic studies are supported ([details](#)) including compound concentrations, NMRMS spectral bins, NMRMS peak intensity table, NMRMS peak lists, and LC/MS spectra.

**Data processing:** Depending on the type of the uploaded data, different data processing options are available ([details](#)). This is followed by data normalization steps including normalization by constant sum, normalization by a reference sample/feature, sample specific normalization, auto-Parabola scaling, etc.

**Statistical analysis:** A wide array of commonly used statistical and machine learning methods are available: [univariate](#) - fold change analysis, t-tests, volcano plot, and one-way ANOVA, correlation analysis; [multivariate](#) - principal component analysis (PCA) and partial least squares - discriminant analysis (PLS-DA); [high-dimensional feature selection](#) - significance analysis of microarrays (and [more](#)...); [RTI](#) and [hmp](#) are supported by the University of Toronto, Institute of Metabolism and Diabetes (IM&D), University of Guelph, and the University of Alberta.

RTI INTERNATIONAL

RTI International

# MetaboAnalyst: Functional Modules

Please choose a functional module to proceed:

**Statistical Analysis**

This module offers various commonly used statistical and machine learning methods from t-tests, ANOVA to PCA and PLS-DA. It also provides clustering and visualization such as dendrogram, heatmap, K-means, as well as classification based on random forests and SVM.

**Enrichment Analysis**

This module performs metabolite set enrichment analysis (MSEA) for human and mammalian species based on several libraries containing ~6300 groups of biologically meaningful metabolite sets. Users can upload a list of compounds, a list of compounds with concentrations, or a concentration table.

**Pathway Analysis**

This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 21 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, Budding yeast, E. coli, etc., with a total of ~1600 metabolic pathways.

**Time Series Analysis**

This module supports data overview (PCA and heatmaps), two-way ANOVA, multivariate empirical Bayes time-series analysis for detecting distinctive temporal profiles across different experimental conditions, and ANOVA-simultaneous component analysis (ASCA) for identification of major patterns associated with each experimental factor.

**Power Analysis**

This module allows you to upload a pilot data set to calculate the minimum number of samples required to detect the existence of a difference between two populations with a given degree of confidence.

**Biomarker Analysis**

To perform various ROC curve based biomarker analysis. It supports classical single biomarker analysis, multivariate biomarker analysis, and manual biomarker selection and evaluation.

**Integrated Pathway Analysis**

To perform joint metabolic pathway analysis on results obtained from metabolomics and gene expression studies under the same experimental or biological


**Other Utilities**

This module contains some utility functions commonly used for metabolomics data manipulation and analysis. At this moment, compound ID conversion is

RTI INTERNATIONAL

RTI International

## Data Upload



**MetaboAnalyst 3.0**  
– a comprehensive tool suite for metabolomic data analysis

Upload

Processing

Normalization

Statistics

Download

Log out

### 1) Upload your data

**Comma Separated Values (.csv) :**

Data Type:  Concentration  Spectral bins  Peak intensity table

Format:


Data File:  No file chosen

**Zippped Files (.zip) :**

Data Type:  NMR peak list  MS peak list  MS spectra


Data File:  No file chosen

Pair File:  No file chosen



RTI International

## Data Integrity Check



**MetaboAnalyst 3.0**  
– a comprehensive tool suite for metabolomic data analysis

Upload

Processing

Pre-process

**Data check**

Missing value

Data filter

Data editor

Color picker

Normalization

Statistics

Download

Log out

### Data Integrity Check:

1. Checking the class labels - at least three replicates are required in each class.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. The presence of missing values or features with constant values (i.e. all zeros)

**Data processing information:**

Checking data content...passed

Samples are in rows and features in columns

The uploaded file is in comma separated values (.csv) format

The uploaded data file contains 38 (samples) by 231 (spectra bins) data matrix.

7 groups were detected in samples.

Samples are not paired.

All data values are numeric.


A total of 0 (0%) missing values were detected.

By default, these values will be replaced by a small value.

Click **Skip** button if you accept the default practice

Or click **Missing value imputation** to use other methods

Last modified 2015-02-08



## Data Filtering

**MetaboAnalyst 3.0**  
— a comprehensive tool suite for metabolomic data analysis

**Data Filtering:**

The purpose of the data filtering is to identify and remove variables that are unlikely to be of use when modeling the data. No phenotype information are used in the filtering process, so the result can be used with any downstream analysis. This step is strongly recommended for untargeted metabolomics datasets (i.e. spectral binning data, peak lists) with large number of variables, many of them are from baseline noises. Filtering can usually improve the results. For details, please refer to the paper by [Blackstone et al.](#)

Non-informative variables can be characterized in two groups: variables of very small values - these variables can be detected using mean or median; variables that are near-constant throughout the experiment conditions - these variables can be detected using standard deviation (SD), or the robust estimate such as interquartile range (IQR). The relative standard deviation (RSD = SD/mean) is another useful variance measure independent of the mean. The following empirical rules are applied during data filtering:

- **Less than 250 variables:** 5% will be filtered;
- **Between 250 - 500 variables:** 10% will be filtered;
- **Between 500 - 1000 variables:** 25% will be filtered;
- **Over 1000 variables:** 40% will be filtered;

Please note, in order to reduce the computational burden to the server, the **None** option is only for less than 2000 features. Over that, if you choose **None**, the IQR filter will still be applied. In addition, the maximum allowed number of variables is 5000; if over 5000 variables were left after filtering, only the top 5000 will be used in the subsequent analysis.

Interquartile range (IQR)  
 Standard deviation (SD)  
 Median absolute deviation (MAD)  
 Relative standard deviation (RSD = SD/mean)  
 Non-parametric relative standard deviation (MAD/median)  
 Mean intensity value  
 Median intensity value  
 None (less than 2000 features)

Process

RTI  
INTEGRAL.DMG

## Data Normalization

**Data Normalization:**

The normalization procedures are grouped into three categories. The sample normalization allows general-purpose adjustment for differences among samples; data transformation and scaling are two different approaches to make features more comparable. You can use one or combine them to achieve better results.

**Sample normalization**

None  
 Sample specific normalization (i.e. dry weight, volume) [Click here to specify](#)  
 Normalization by sum  
 Normalization by median  
 Normalization by reference sample  
 Specify a reference sample   
 Create a pooled average sample from group   
 Normalization by reference feature

**Data transformation**

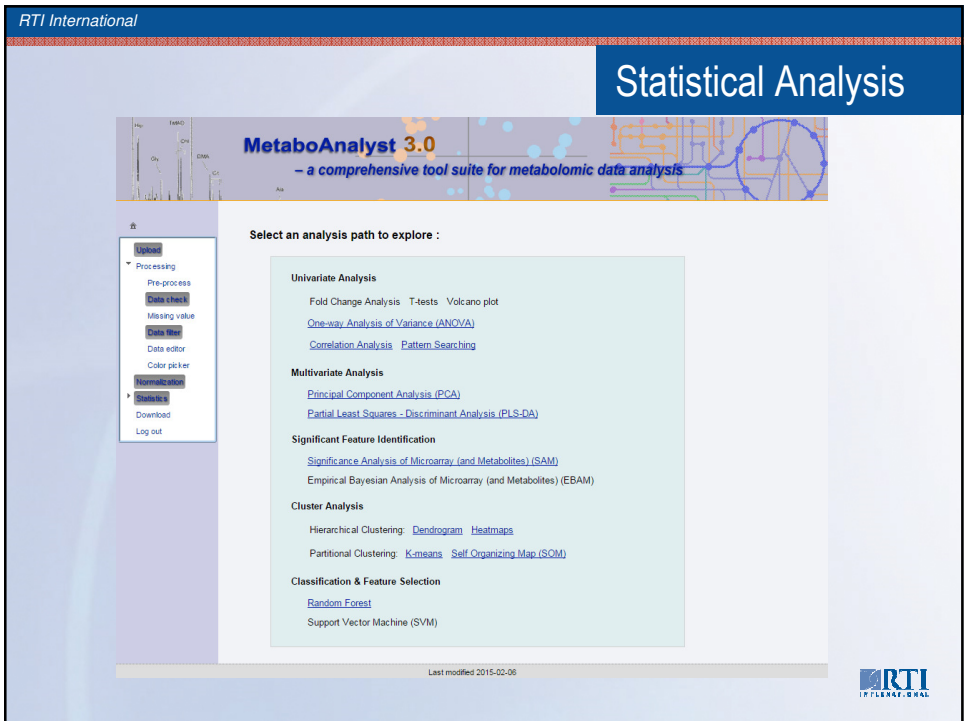
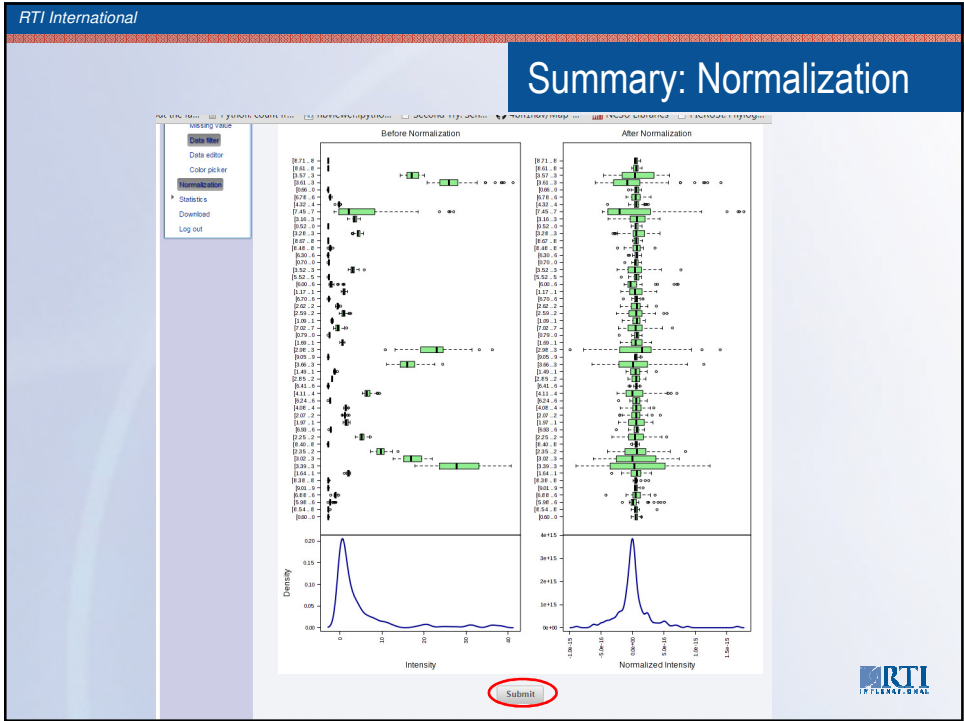
None  
 Log transformation (generalized logarithm transformation or loglog)  
 Cube root transformation (take cube root of data values)

**Data scaling**

None  
 Auto scaling (mean-centered and divided by the standard deviation of each variable)  
 Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)  
 Range scaling (mean-centered and divided by the range of each variable)

Submit

RTI  
INTEGRAL.DMG





RTI International

## Pooled QC Samples

**MetaboAnalyst 3.0**  
— a comprehensive tool suite for metabolomic data analysis

Overview | Score Plot | 2D Scores Plot | 3D Scores Plot | Loadings Plot | Biplot

Dragging to rotate the view around the axis. Clicking on any data point to view a summary of the corresponding sample. Scrolling to zoom in and out; use the Export Image button below to export the current view as a PNG image.

Update | Export

Legend:

- HD\_D01
- HD\_D14
- LD\_D01
- LD\_D14
- Pooled\_All
- V\_D01
- V\_D14

RTI International

## PCA Day 01 and Day 14

**Scores Plot**

Legend:

- △ HD\_D01
- + HD\_D14
- × LD\_D01
- × LD\_D14
- ▽ Pooled\_All
- V\_D01
- V\_D14

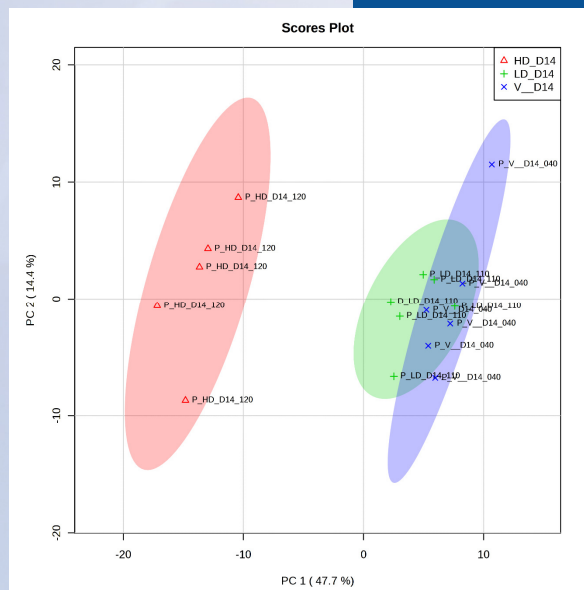
## Day 14: Vehicle, Low Dose, and High Dose Groups

Please go back to the start page and upload the data

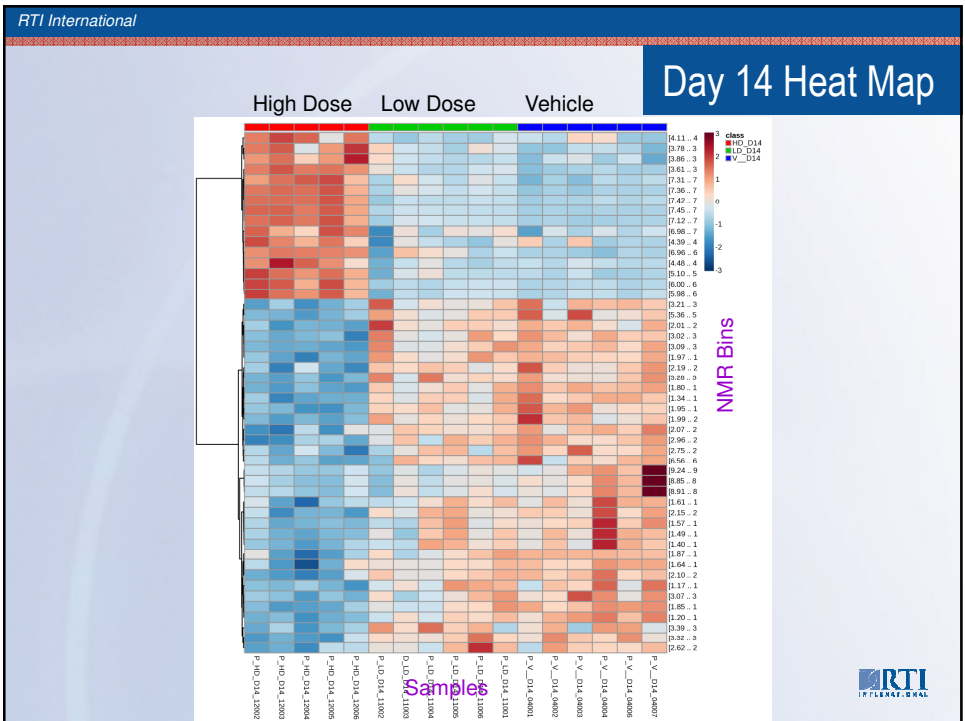
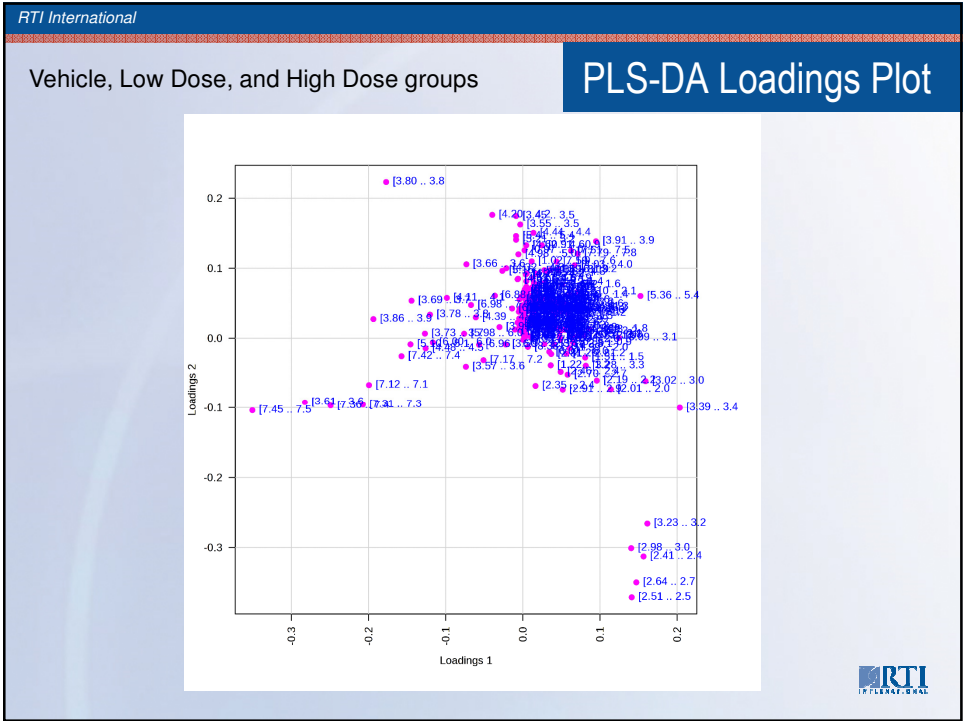
- We will compare high dose vs vehicle
  - 2. UAB\_RFA\_Metaboanalyst\_D14\_NoPools.csv
- Perform PCA
- Perform PLS-DA
- Heat map

Vehicle, Low Dose, and High Dose groups

## Day 14 PCA Scores Plot





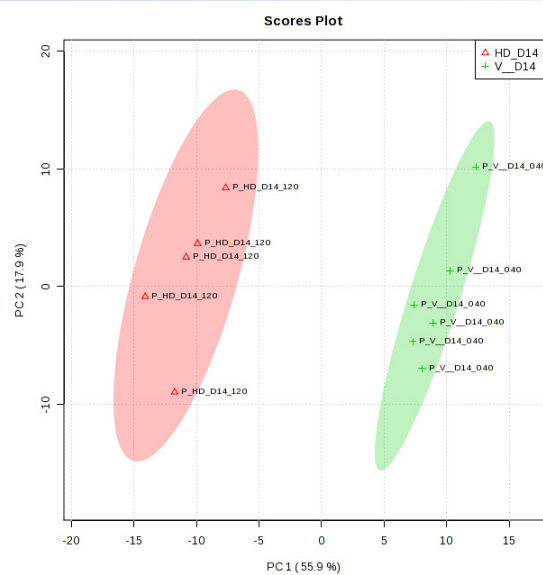


## Comparison of Day 14 High Dose and Vehicle

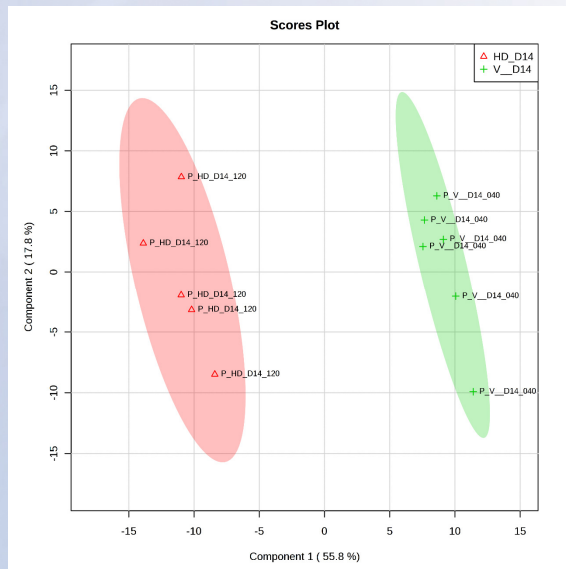
Please start from the start page and upload the data

- We will compare high dose vs vehicle
  - 3. UAB\_RFA\_Metaboanalyst\_D14\_Vehicle\_vs\_HighDose.csv
- Perform PCA
- Perform PLS-DA
- VIP Plot
- Heat map

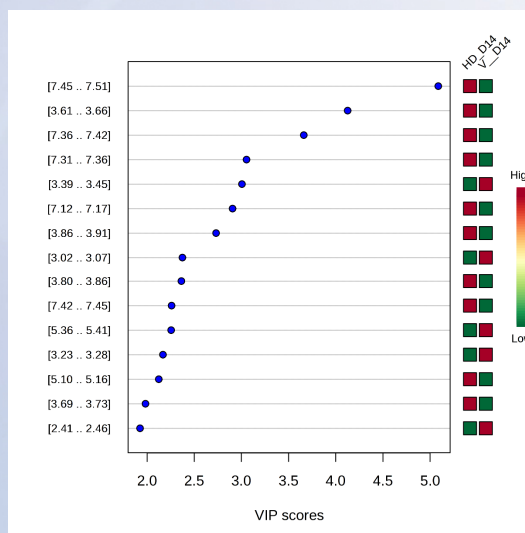
## Day 14 PCA Scores Plot: High Dose vs Vehicle

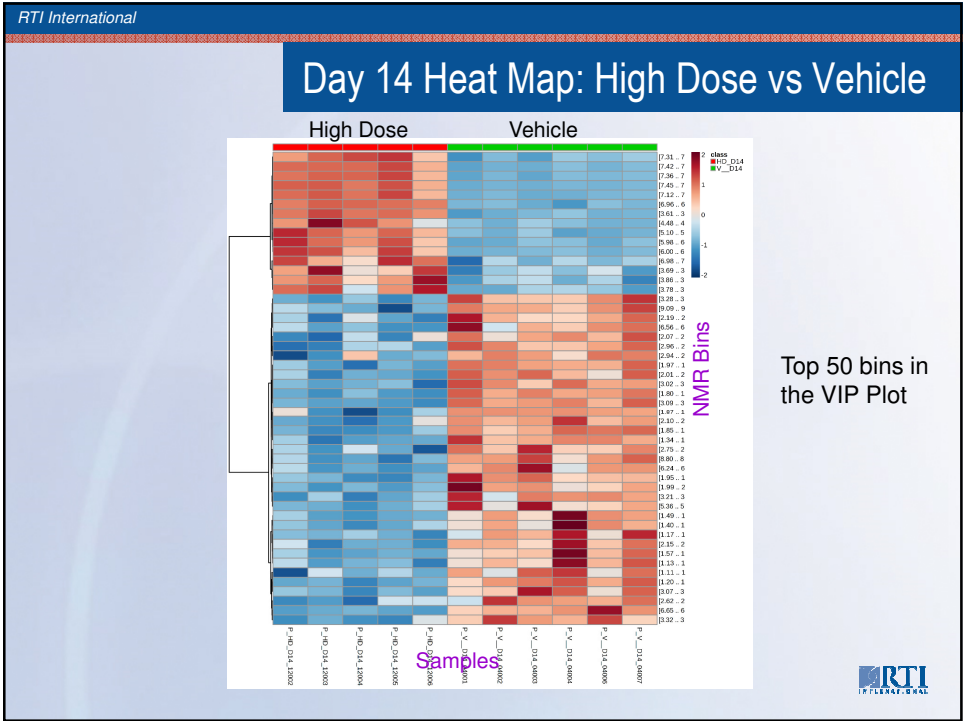


## Day 14 PLS-DA Scores Plot: High Dose vs Vehicle



## Day 14 PLS-DA VIP Plot: High Dose vs Vehicle





RTI International

## STS Center

**Wimal Pathmasiri**  
NMR & GC-MS

**Jim Carlson**  
LC- and GC-MS

**Jessica Gooding**  
LC-MS

**Kelly Mercier**  
NMR

**Susan Sumner**

**Susan McRitchie**  
Data Analysis

**Bob Clark**  
Genetics

**Zach Acuff**  
Biostatistics

**Jason Burgess**  
Program Coordinator

**Andrew Novokhatny**  
NMR and QC

**Aurora Cabrera**  
LC-MS/MS

**Jocelin Spruill**  
GC-MS

**Tammy Cavallo**  
Biology and QC

**Delisha Stewart**  
Cell Biology

**Ninell Mortensen**  
Microbiology

**Maria Moreno**  
Biology

**Rose Ewald**  
intern

**Keith Levine**  
Metallomics

**Yuanyuan Li**  
LC-MS

**Rod Snyder**  
LC-MS

**Sherry Black**  
In vivo and in vitro Metabolism

**Scott Watson**  
Neurotransmitter LC/MS

**Skip Gaudette**  
Systems

**Puvi Patel**  
In vitro metabolism

**Yan Lan Yueh**  
LC-MS

**Tim Fennell**  
Metabolism

**Hieu Vu**  
LC-MS

**Melody Markley**  
Model Systems

**Sue Clark**  
Administrative Support

**Courtney Whitaker**  
LC/MS

Thank You!

If you have any questions, please e-mail me

[wpathmasiri@rti.org](mailto:wpathmasiri@rti.org)

Useful link:

Metabolomics Workbench

<http://www.metabolomicsworkbench.org/>