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)
NMR Metabolomics Workflow

- 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
Some Software available for NMR Based Metabolomics

FREE

- NMR Data Processing
  - ACD Software for Academics (ACD Labs, Toronto, Canada)
- Multivariate data analysis
  - Metabo Analyst 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

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$_2$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)
Please go to the webpage:
http://www.metaboanalyst.ca/MetaboAnalyst/

MetaboAnalyst 3.0

MetaboAnalyst: Functional Modules

Choose a functional module to proceed:

1. **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 dendrograms, heat maps, and classification based on random forests and SVM.

2. **Pathway Analysis**
   This module supports pathway analysis including enrichment analysis and pathway topology analysis and visualization for 37 model organisms, including Homo sapiens, Rattus norvegicus, Danio rerio, Gallus gallus, Drosophila melanogaster, Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces cerevisiae, Escherichia coli, and Streptomyces coelicoflavus.

3. **Disease Analysis**
   This module allows you to upload a 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.

4. **Integrate Pathway Analysis**
   To perform joint metabolite analysis on results obtained from metabolomics and gene expression studies under the same experimental conditions.
**Data Upload**

1) Upload your data

- **Comma Separated Values (.csv):**
  - **Data Type:** Concentration / Spectral Data
  - **Format:** Samples in rows (recommended)
  - **Data Files:** Choose File

- **Zipped Files (.zip):**
  - **Data Type:** MS spectra, MS peak list
  - **Data Files:** Choose File

**Submit**

---

**Data Integrity Check**

1. Checking the data content: please check the data content is accurate and complete.
2. Checking the data columns: please check the data columns are consistent with the expected format.
3. Checking the data rows: please check the data rows are consistent with the expected format.
4. Checking the data values: please check the data values are within the expected range.

- **Data processing information:**
  - Samples are in rows and features in columns.
  - The expected file format is a tab-separated values file format.
  - The file contains 200 samples and 30 features.
  - All data values are numeric.
  - A list of missing or invalid values will be displayed.

**Missing value options:**

- **None**: All missing values will be ignored.
- **Mean**: Missing values will be replaced with the mean value.
- **Median**: Missing values will be replaced with the median value.
- **Zero**: Missing values will be replaced with zero.

**Submit**
Data Filtering:
The purpose of data filtering is to identify and remove variables that are unlikely to be of use when analyzing the data. Any pre-established information can be used in the filtering process, so the results can be used with any subsequent analysis. The following methodology is recommended to identify unwanted variables:

1. **Identify variables**:
   - Less than 1% of the data is within one standard deviation of the mean.
   - Between 50% and 100% of the data is within one standard deviation of the mean.
   - Between 90% and 100% of the data is within one standard deviation of the mean.
   - Median intensity > mean intensity.
   - None (less than 200 features).

For example, to reduce the computational burden on the server, the filter option is only for less than 2000 features. Once the thresholds are met, the selected features are applied to data filtering.

Data Normalization:
The normalization criteria are grouped into two categories. The sample normalization allows you to adjust for differences among samples, while the variance and scaling are two different approaches to make data more comparable. You can use one or a combination of both to achieve better results.

**Sample normalization**:
- None
- Specific normalization (e.g., heavy or light)
- Normalization by sample
- Normalization by experiment
- Normalization by reference sample
- Specific reference sample
- Normalization by reference sample

**Data transformation**:
- None
- Log transformation
- Square root transformation
- Variance scaling
- Mean scaling
- Median scaling
Summary: Normalization

Statistical Analysis

MetaboAnalyst 3.0
— a comprehensive tool suite for metabolomic data analysis

Select an analysis path to explore:

- Univariate Analysis
  - Fold Change Analysis
  - Volcano Plot
- Principal Component Analysis (PCA)
- Partial Least Square - Discriminant Analysis (PLS-DA)
- Significance Analysis of Metabolome (SAM)
- Empirical Bayesian Analysis of Mass Spectra (EMMA)
- Cluster Analysis
  - Hierarchical Clustering
  - k-means
  - Self-Organizing Map (SOM)
- Classification & Feature Selection
  - Random Forest
  - Support Vector Machine (SVM)
Pooled QC Samples

PCA Day 01 and Day 14
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

**PCA Loadings Plot**

**PLS-DA Scores Plot**
Vehicle, Low Dose, and High Dose groups

PLS-DA Loadings Plot

Day 14 Heat Map

Samples NMR Bins

High Dose  Low Dose  Vehicle
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 PLS-DA Scores Plot: High Dose vs Vehicle

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

Top 50 bins in the VIP Plot

RTI International

STS Center

Susan Sumner

Wimal Pathmasiri
NMR & GC-MS

Jim Carlson
LC and GC-MS

Jessica Gooding
LC-MS

Kelly Mercier
NMR

Susan McRitchie
Data Analysis

Tammy Cavalo
Biology and QC

Deisha Stewart
Cell Biology

Ninell Mortensen
Microbiology

Rose Ewald
Intern

Yuanyuan Li
LC-MS

Red Snyder
LC-MS

Sherry Black
In vivo and in vitro Metabolism

Scott Watson
Neurotransmitter LC/MS

Puvi Patal
In vitro metabolism

Nan Liu, Yueh
In vitro metabolism

Melody Markley
Model Systems

Courtney Whitaker
LC/MS

Hieu Vu
LC-MS

Jason Burgess
Program Coordinator

Keith Levine
Metallomics

Andrew Novokhathy
NMR and QC

Aurora Cabrera
LC-MS/MS

Joel Spruill
GC-MS Neurotransmitter

Tammy Cavalo
Biology and QC

Deisha Stewart
Cell Biology

Ninell Mortensen
Microbiology

Rose Ewald
Intern

Yuanyuan Li
LC-MS

Red Snyder
LC-MS

Sherry Black
In vivo and in vitro Metabolism

Scott Watson
Neurotransmitter LC/MS

Puvi Patal
In vitro metabolism

Nan Liu, Yueh
In vitro metabolism

Melody Markley
Model Systems

Courtney Whitaker
LC/MS

Hieu Vu
LC-MS
If you have any questions, please e-mail me

wpathmasiri@rti.org

Useful link:
Metabolomics Workbench
http://www.metabolomicsworkbench.org/