NMR Metabolomics Analysis

UAB Metabolomics Training Course
February 13, 2015

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

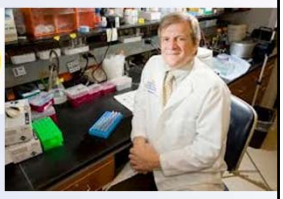
RTI International is a trade name of Research Triangle Institute.

www.rti.org

RTI International

NIH Common Fund Metabolomics Cores

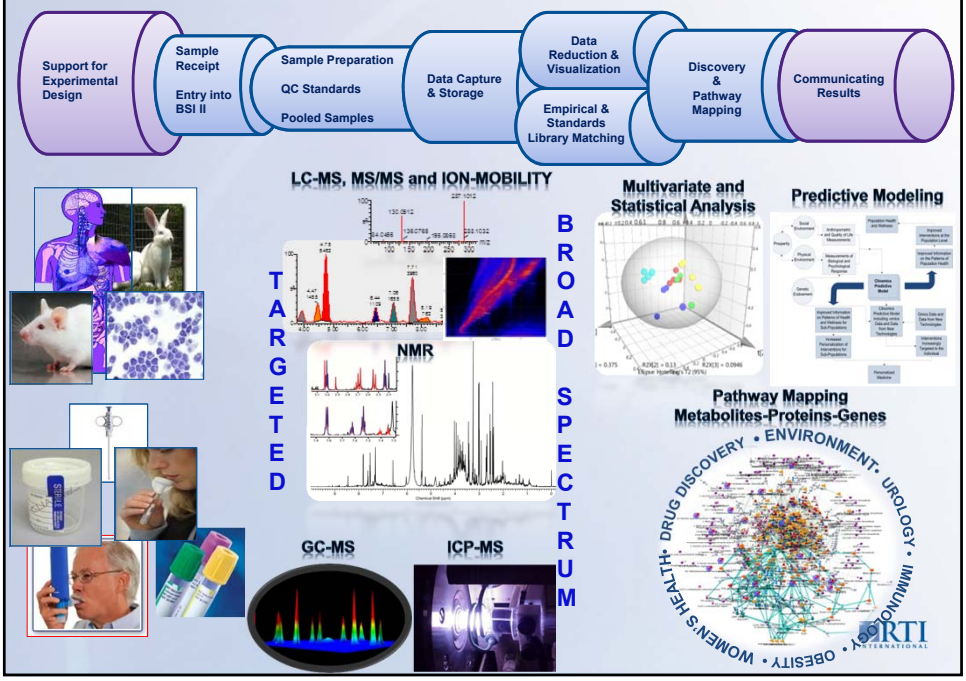
NIH Metabolomics Centers Ramp Up | November 4, 2013 Issue - Vol. 91 Issue 44 | Chemical & Engineering News. by Jyllian Kemsley



Outline of Today's Training

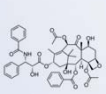
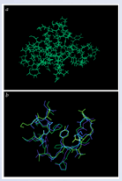
- Introduction: Wimal Pathmasiri
- NMR Metabolomics: Rodney Snyder, Susan McRitchie, Kelly Mercier
  - Study Design
  - Sample Preparation
  - Data Acquisition
  - Data Pre-processing
  - Statistical Analysis
  - Library Matching
  - Pathway Analysis
- Hands On Exercise: Wimal Pathmasiri

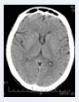
NIH Eastern Regional Comprehensive Metabolomics Resource Core at RTI



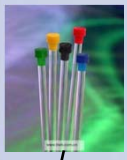
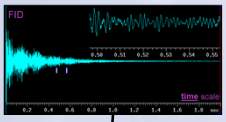
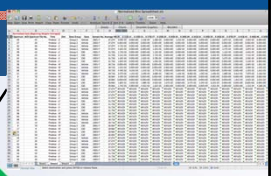
Nuclear Magnetic Resonance (NMR) Spectroscopy

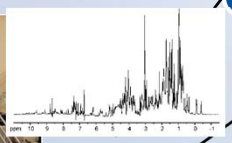
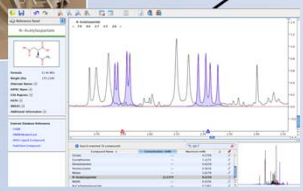
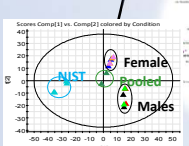
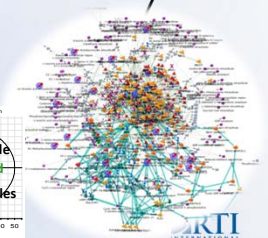
- Detects NMR active nuclei
- Robust and highly reproducible
- Non-destructive
- Quantitative
- Used in
  - Structure elucidation
    - Small molecules
    - Macromolecules (DNA, RNA, Proteins)
  - A number of techniques
    - 1D , 2D, 3D
  - Molecular motion and dynamics
- Similar method used in Imaging (MRI, fMRI)



NMR Metabolomics Workflow

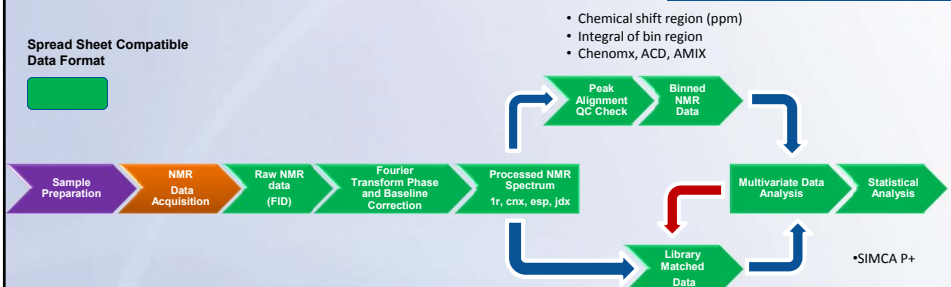




Peak Alignment QC Check
Binned NMR Data
Sample Preparation
NMR Data Acquisition
Raw NMR data (FID)
Fourier Transform Phase and Baseline Correction
Processed NMR Spectrum 1r, cnx, esp, jdx
Multivariate Data Analysis
Statistical Analysis
Pathway Analysis
Library Matched Data

Sample Preparation, Data Acquisition, and Pre-processing

NMR Work Flow



- Optimized methods (biofluids, tissues, urine)
- QC samples prepared in parallel

- Bruker Avance III NMR Spectrometer
- TopSpin Software
- Varian Unity Innova NMR Spectrometer
- VNMR Software

- Bruker Avance III NMR Spectrometer
- TopSpin Software
- Varian Unity Innova NMR Spectrometer
- VNMR Software

- metabolite
- concentration
- Chenomx
- AMIX

- SIMCA P+
- MATLAB 2011b including Simulink and Toolboxes (MathWorks, MA)
- R Statistical Programming

- SIMCA P+
- MATLAB 2011b including Simulink and Toolboxes (MathWorks, MA)
- R Statistical Programming
- Spotfire

Mapping
- KEGG
- GeneGo

NMR data acquisition is performed by using methods cited in Beckonert et al. (2007), *Nature Protocols*, 2 (11), 2692-2703.

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

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

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)

Important Steps in Metabolomics Analysis

- Study design
  - Match for factors such as gender, ethnicity, age, BMI (human studies)
  - Use of same strains in animal studies
- Sample collection
  - Collection vials, anticoagulant use (heparin, citrate, EDTA)
- Sample storage
  - -20 °C, -80 °C, minimize freeze-thaw cycles
- Sample preparation
  - Optimize the methods and use them consistently throughout study
  - Daily balance and pipette checks
- Use of Quality Check (QC) samples
  - Pooled QC samples (Phenotypic and combined pooled samples)
  - Use matching external pooled QC samples where pool samples cannot be prepared from study samples
- **Consistency and reproducibility are the keys for a successful metabolomics study**

RTI International

Sample Preparation for Metabolomics Analysis

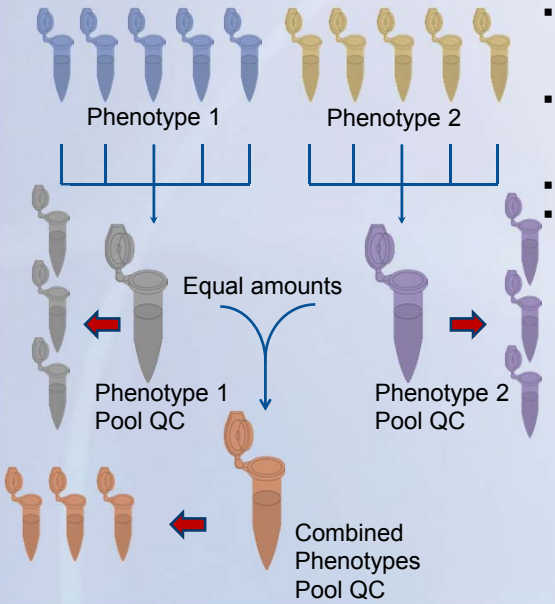
Current sample preparation practices (in brief)

- **Biofluids**
  - Dilute with D_2O/ buffer/ 0.9% Saline
  - Add internal standard (ISTD, eg. Chenomx) solution or formate (for serum).
  - Centrifuge and transfer an aliquot into NMR tube
- **Tissue and Cells**
  - Homogenization performed in ice cold 50/50 acetonitrile/water
  - Supernatant dried down (lyophilized)
  - Reconstituted in D_2O and ISTD (eg. Chenomx) solution
- **Pooled QC Samples (Sample Unlimited)**
  - Mix equal volume of study samples to get pooled QC samples
  - 10% QC samples
- **Pooled QC Samples (Sample Limited)**
  - Use independent pool of similar samples
  - 10% QC samples
- **Daily balance and pipette check**

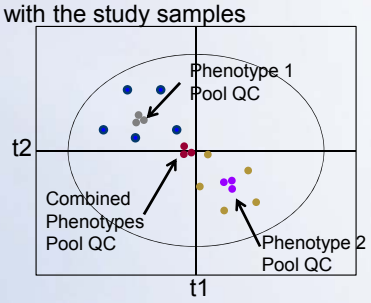
Samples are randomized for preparation and data acquisition

RTI International

Preparing Pooled QC Samples



- Aliquots from each sample in the study phenotype are pooled (phenotypic pool)
- Equal amount of each phenotypic pools are pooled (Combined phenotypic pool)
- Replicates of pools are prepared
- Pool samples are prepared along with the study samples

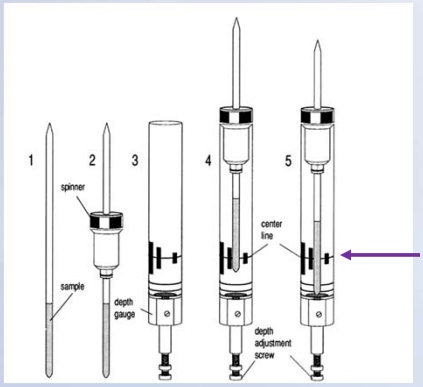


Pooled samples should cluster tightly

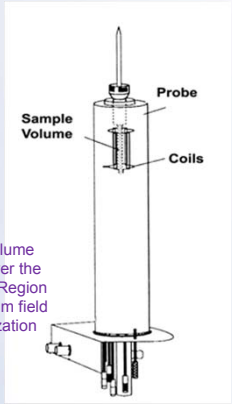
NMR Data Acquisition

- 1D NMR
  - 1st increment of NOESY
    - noesyprid (Bruker)
  - CPMG (serum or plasma)
    - cpmgpr1d (Bruker)
    - To remove broadening of signals due to macromolecules (eg. Proteins and lipids)
- 2D NMR (for structure elucidation)
  - 2D J-Resolved
  - COSY
  - TOCSY
  - HSQC
  - HMBC

Sample Amount in NMR tube



Sample volume should cover the NMR Coil Region For optimum field homogenization



- At least 10% D_2O in the sample
- Optimum volume
  - 550 – 600 uL (5mm tube)
  - 200 uL (3 mm tube)
- Sample gauge is used

For very small sample amounts, a NMR with a microcoil probe is an option.

AVANCE Beginners User Guide 004 (Bruker, Germany)

NMR Data

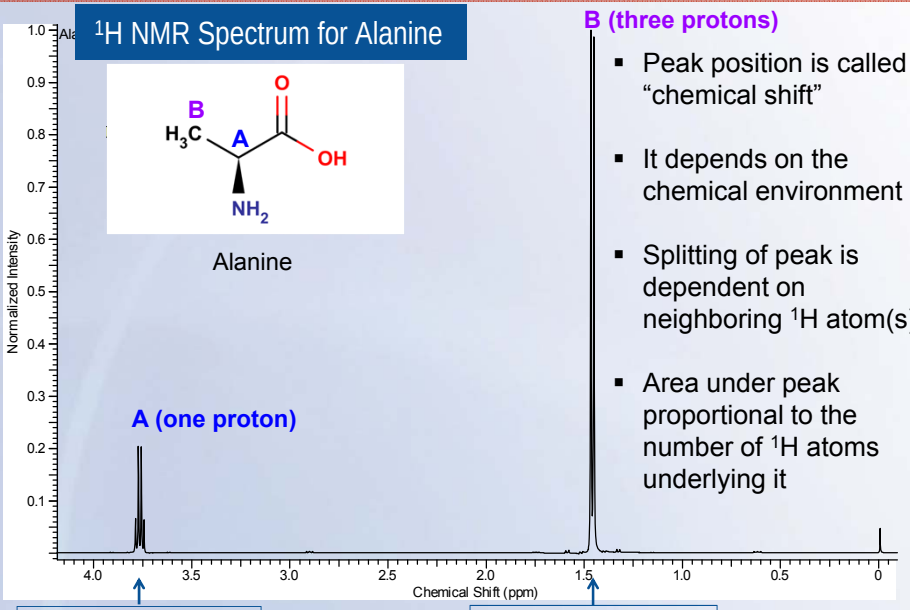
- A typical ^{1}H NMR Spectrum consists of thousands of sharp lines or signals.
- The intensity of the peak is directly related to the number of protons underlying the peak.
- The position of a particular peak in the X-axis of the NMR spectrum is called the “Chemical Shift” and it is measured in ppm scale
- The NMR spectrum obtained for the biological sample is referenced using a reference compound such as DSS, TSP, or Formate added to the sample in sample preparation step.
- pH indicator may also be used (for example, Imidazole)

DSS=4,4-dimethyl-4-silapentane-1-sulfonic acid, TSP=Trimethylsilyl propionate

RTI International

^{1}H NMR Spectrum for Alanine

Alanine



A (one proton)

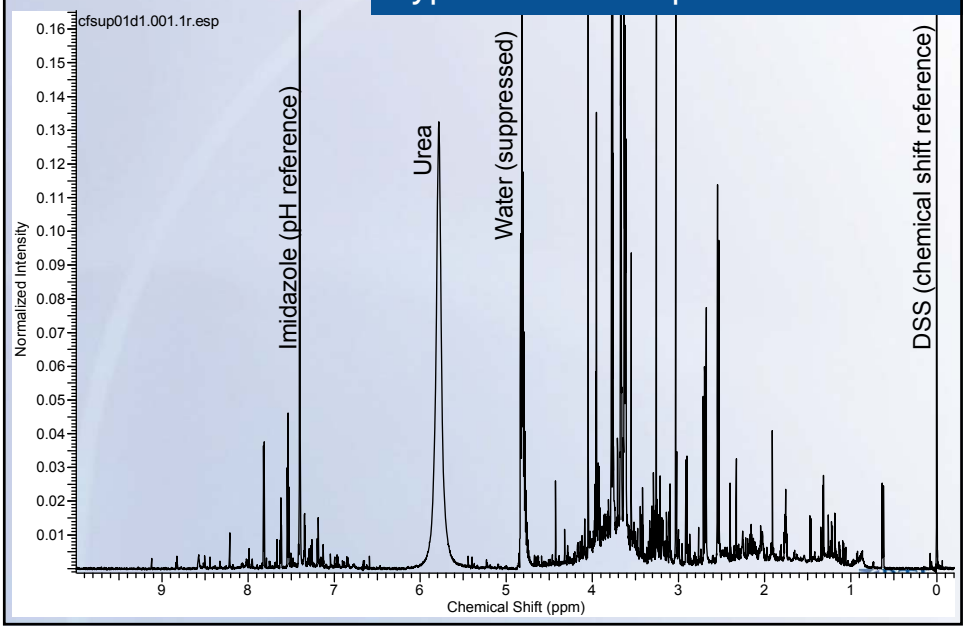
B (three protons)

- Peak position is called “chemical shift”
- It depends on the chemical environment
- Splitting of peak is dependent on neighboring ^{1}H atom(s)
- Area under peak proportional to the number of ^{1}H atoms underlying it

Chemical shift of C**H**
4 lines (quadrat)

Chemical shift of C**H_3**
2 lines (doublet)

Typical ^{1}H NMR Spectrum of Urine



Data Pre-processing

- After NMR data acquisition, the result is a set of spectra for all samples.
- For each spectrum, quality of the spectra should be assessed.
  - Line shape, Phase, Baseline
- Spectra should be referenced
  - Compounds commonly used: DSS, TSP, Formate
- Variations of pH, ionic strength of samples has effects on chemical shift
  - Peak alignment
  - Binning or Bucket integration
- Remove unwanted regions
- Normalize data (remove variation in concentration of samples)

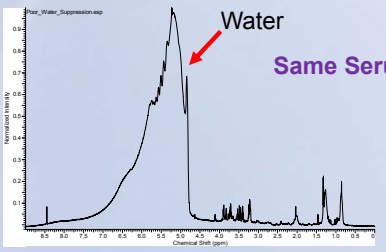
High quality data are needed

Quality Control Steps

- Quality of metabolomics analysis depends on data quality
- Typical problems
  - Water peak (suppression issues)
  - Baseline (not set at zero and not a flat line)
  - Alignment of peaks (chemical shift, due to pH variation)
  - Variation in concentration (eg. Urine)
- High quality of data is needed for best results

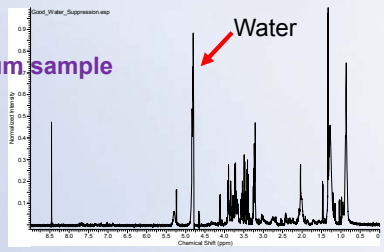
Water Suppression Effects and Other Artifacts

- If water is not correctly suppressed or removed there will be effects on normalization
- Need to remove other artifacts
- Remove drug or drug metabolites



Poor water suppression

Same Serum sample

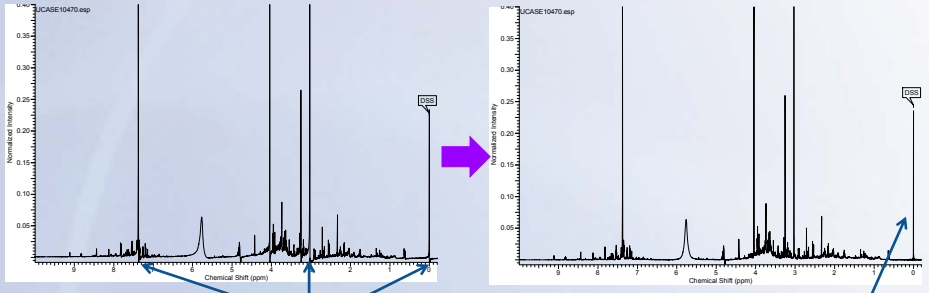


Good water suppression

NMR Pre-processing

Before

After



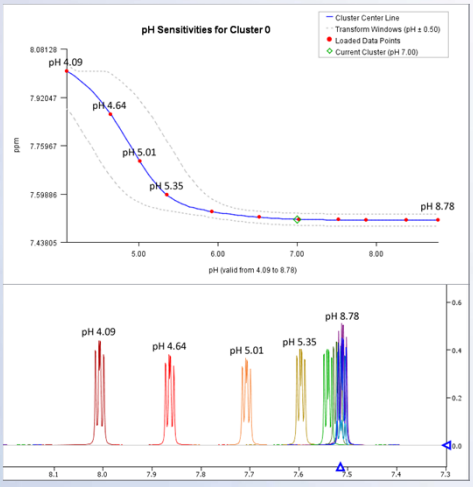
Phase
Baseline to be corrected

Reference,
Line shape

pH Dependence of Chemical Shift

Chemical shift variability
- pH
- ionic strength
- metal concentration

Methods to overcome this problem
- Use a buffer when preparing samples
- Binning (Bucketing)
  - Fixed binning
  - Intelligent binning
  - Optimized binning
- Available data alignment tools
  - Recursive Segment-wise Peak Alignment (RSPA)
  - Icoshift
  - speaq



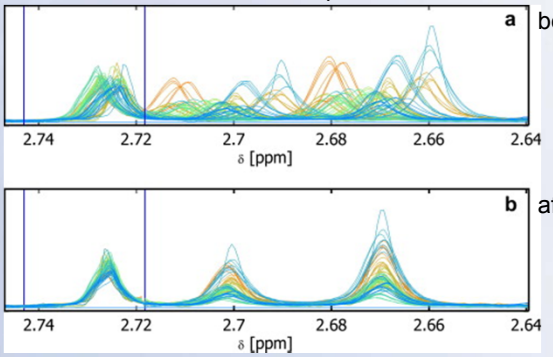
http://www.chenomx.com/software/software.php
Savorani , F. et al, Journal of Magnetic Resonance, Volume 202, Issue 2, 2010, 190 – 202
Vu, T. N. et al., *BMC Bioinformatics* 2011, **12**:405

Peak Alignment

Example

icoshift

One of the Citrate peaks



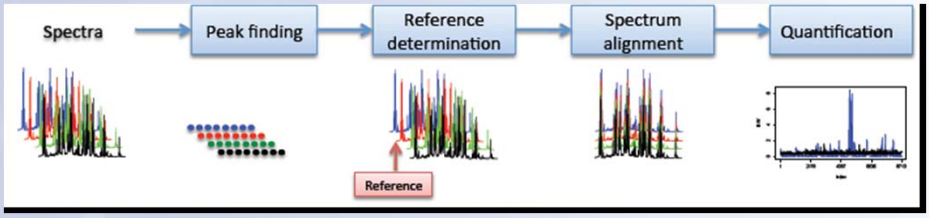
Savorani , F. et al, Journal of Magnetic Resonance, Volume 202, Issue 2, 2010, 190 - 202

RTI International

Peak Alignment

Example

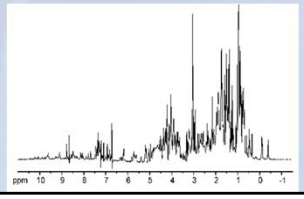
speaq

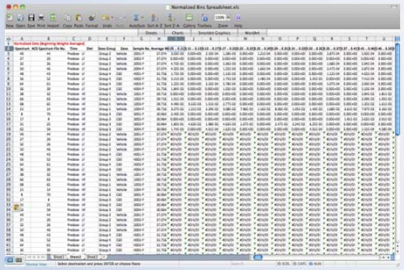


Vu, T. N. et al., *BMC Bioinformatics* 2011, **12**:405

NMR Binning

- A form of quantification that consists of segmenting a spectrum into small areas (bins/buckets) and attaining an integral value for that segment
- Binning attempts to minimize effects from variations in peak positions caused by pH, ionic strength, and other factors.
- Two main types of binning
  - Fixed binning
  - Flexible binning



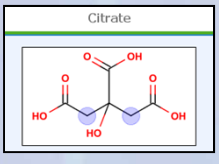


RTI International

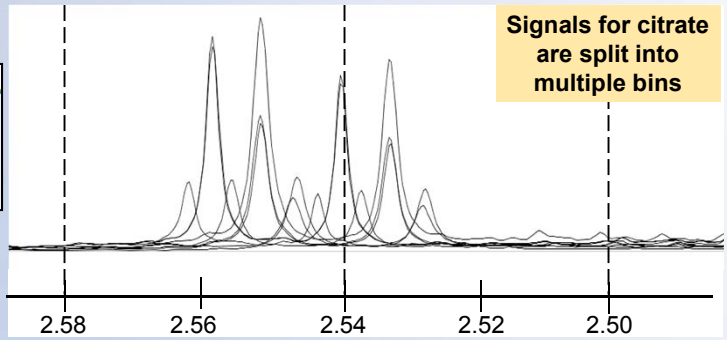
NMR Binning

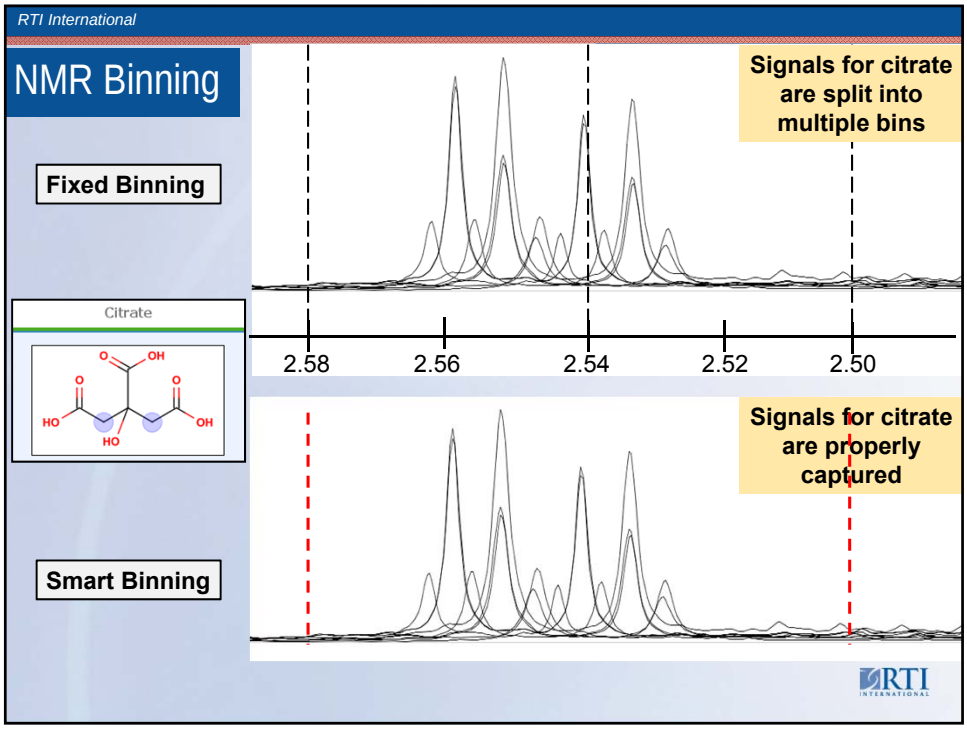
Peak shift can cause the same peak across multiple samples to fall into different bins

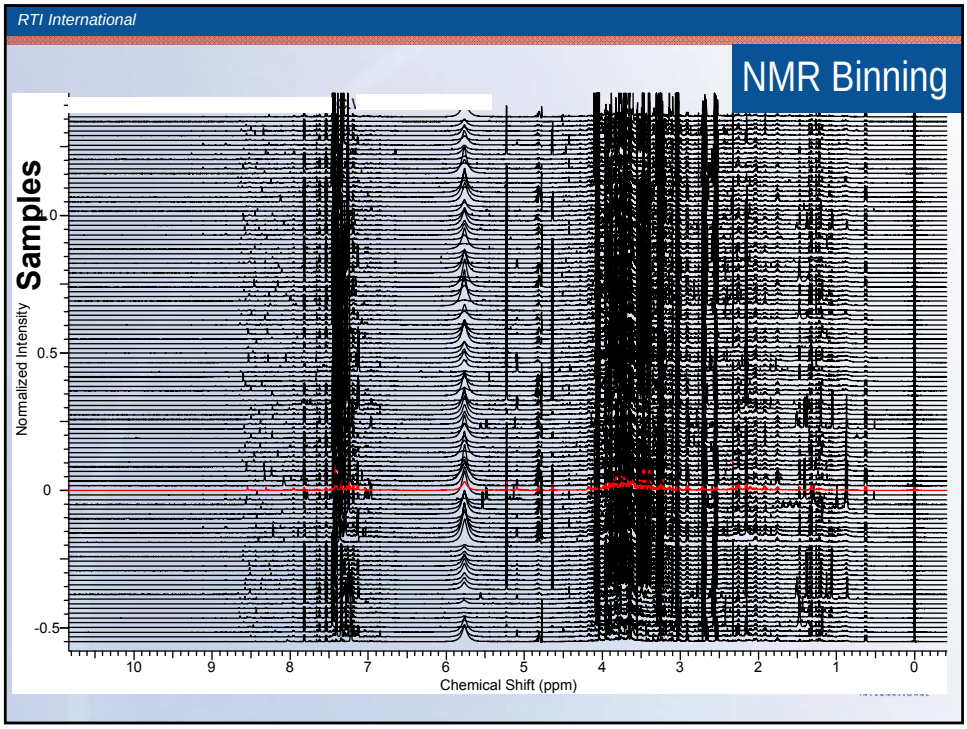
- The entire NMR spectrum is split into evenly spaced integral regions with a spectral window of typically 0.04 ppm.
- The major drawback of fixed binning is the non-flexibility of the boundaries.
- If a peak crosses the border between two bins it can significantly influence your data analysis

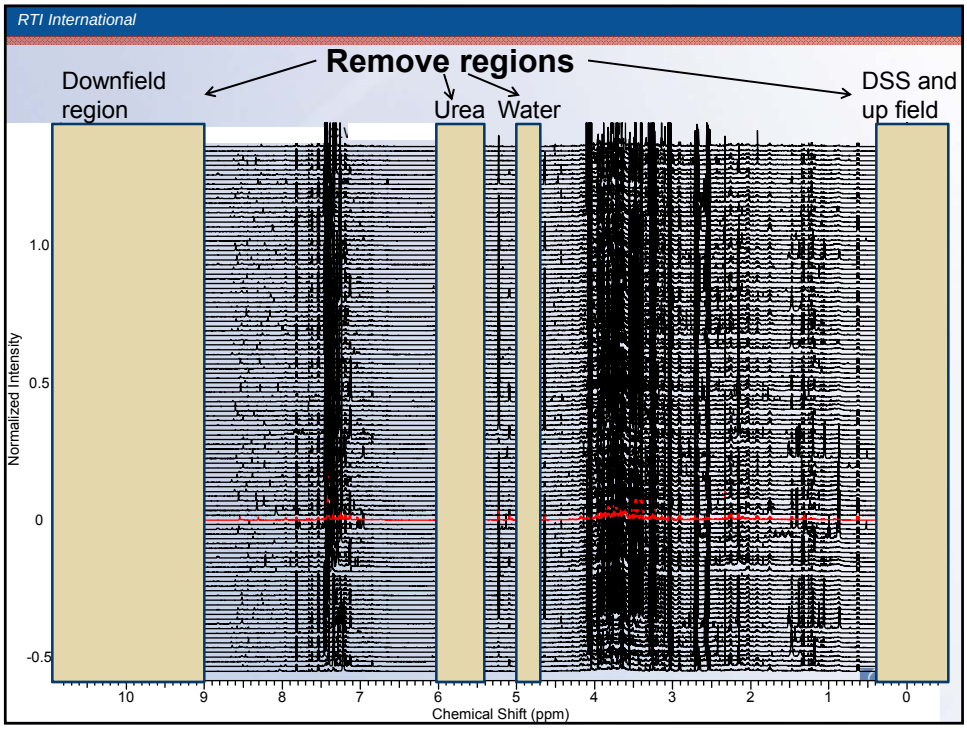


Fixed Binning




RTI International
NMR Binning
Fixed Binning
Citrate
Signals for citrate are split into multiple bins
2.58
2.56
2.54
2.52
2.50
Signals for citrate are properly captured
Smart Binning


RTI International
NMR Binning
Samples
Normalized Intensity
1.0
0.5
0
-0.5
10
9
8
7
6
5
4
3
2
1
0
Chemical Shift (ppm)



Binning

- Integrate bins (0.04 ppm bin size)
- Normalize integral of each bin to the total integral of each spectrum
- Merge metadata
- Result is a spreadsheet ready for further multivariate data analysis and other statistical analysis

Sample ID	Disease Group	[0.40 .. 0.46]	[0.46 .. 0.52]	[0.52 .. 0.54]	[0.54 .. 0.57]	[0.57 .. 0.60]	[0.60 .. 0.66]	[0.66 .. 0.68]	[0.68 .. 0.71]	[0.71 .. 0.75]
C0559	Cases	7.60E-05	0.00E+00	7.32E-02	8.48E-02	3.20E-02	1.84E+00	1.31E-01	3.60E-01	3.67E-01
C0629	Cases	0.00E+00	1.78E-02	0.00E+00	2.18E-02	0.00E+00	1.08E+01	0.00E+00	0.00E+00	3.02E-02
C0640	Cases	3.44E-04	0.00E+00	1.83E-03	1.86E-04	0.00E+00	4.51E+00	0.00E+00	0.00E+00	0.00E+00
C0835	Cases	6.41E-04	0.00E+00	6.44E-03	0.00E+00	3.96E-03	3.28E+00	0.00E+00	5.12E-03	1.75E-02
D0613	Cases	6.63E-03	0.00E+00	0.00E+00	1.06E-02	0.00E+00	5.79E+00	0.00E+00	6.36E-02	3.02E-01
D0762	Cases	0.00E+00	0.00E+00	1.79E-02	1.98E-02	0.00E+00	9.37E+00	0.00E+00	0.00E+00	1.74E-02
D1113	Cases	3.14E-03	2.42E-03	8.02E-02	1.04E-01	5.32E-03	3.74E+00	0.00E+00	2.02E-02	1.84E-01
D1158	Cases	0.00E+00	3.71E-03	2.35E-02	4.83E-02	0.00E+00	5.02E+00	0.00E+00	1.91E-02	0.00E+00
D2090	Cases	0.00E+00	0.00E+00	2.45E-03	9.98E-04	0.00E+00	5.76E+00	0.00E+00	1.24E-02	1.04E-02
E0004	Cases	1.72E-03	0.00E+00	6.85E-02	3.05E-02	0.00E+00	1.47E+00	6.90E-02	3.61E-01	4.08E-01
E0195	Cases	0.00E+00	1.69E-03	5.57E-02	6.29E-02	0.00E+00	2.77E+00	1.34E-01	2.04E-01	4.56E-01
E0325	Cases	1.25E-03	0.00E+00	4.40E-03	1.69E-02	0.00E+00	9.17E+00	0.00E+00	1.08E-02	2.30E-02
E0509	Cases	4.11E-03	0.00E+00	2.23E-02	7.54E-03	3.08E-03	3.54E+00	0.00E+00	3.28E-02	9.09E-01
E0487	Cases	1.72E-03	0.00E+00	0.00E+00	1.00E-02	0.00E+00	4.00E+00	0.00E+00	1.36E-02	0.00E+00
F0036	Cases	1.66E-02	0.00E+00	0.00E+00	2.06E-02	0.00E+00	1.22E+01	1.04E-02	0.00E+00	5.97E-01
F0108	Cases	0.00E+00	2.31E-03	6.30E-03	1.11E-02	0.00E+00	7.17E+00	0.00E+00	1.65E-02	2.21E-01
A0233	Control	0.00E+00	1.86E-02	0.00E+00	1.82E-02	0.00E+00	1.61E+01	0.00E+00	2.91E-03	0.00E+00
A0490	Control	0.00E+00	0.00E+00	2.99E-03	3.60E-02	0.00E+00	2.97E+00	0.00E+00	4.00E-02	5.46E-01
A2003	Control	0.00E+00	0.00E+00	3.45E-02	2.20E-02	0.00E+00	1.80E+00	0.00E+00	0.00E+00	0.00E+00
C0586	Control	0.00E+00	1.69E-02	0.00E+00	6.64E-03	0.00E+00	1.92E+01	0.00E+00	6.51E-02	0.00E+00
C2177	Control	0.00E+00	0.00E+00	3.02E-02	3.59E-02	0.00E+00	2.35E+00	0.00E+00	3.19E-02	1.49E-01
D0177	Control	9.21E-03	0.00E+00	1.69E-02	1.47E-02	0.00E+00	2.43E+00	0.00E+00	4.46E-02	0.00E+00
D0729	Control	0.00E+00	1.88E-03	5.58E-02	7.87E-02	2.92E-02	3.16E+00	6.59E-02	2.80E-01	4.30E-01
D0909	Control	0.00E+00	1.08E-03	0.00E+00	5.69E-03	0.00E+00	2.49E+00	0.00E+00	1.01E-02	1.87E-01
D0945	Control	0.00E+00	4.79E-04	7.00E-03	0.00E+00	4.19E-03	3.99E+00	0.00E+00	1.11E-03	3.96E-02
D1174	Control	0.00E+00	9.33E-04	0.00E+00	3.43E-03	1.30E-02	7.21E+00	6.53E-03	0.00E+00	1.66E-02
D2054	Control	1.55E-03	0.00E+00	0.00E+00	1.22E-02	0.00E+00	2.07E+00	0.00E+00	1.28E-02	3.90E-01
D2062	Control	2.39E-05	0.00E+00	6.04E-02	2.99E-02	0.00E+00	4.94E+00	0.00E+00	9.95E-03	0.00E+00
D2079	Control	2.73E-02	0.00E+00	1.81E-03	1.17E-02	0.00E+00	3.38E+01	7.87E-02	0.00E+00	5.91E+00

Metadata

Normalized binned data

Data Normalization, Transformation, and Scaling

Data Normalization

- Normalization reduces the sample to sample variability due to differences in sample concentrations—particularly important when the matrix is urine
  - Normalization to total intensity is the most common method
    - For each sample, divide the individual bin integral by the total integrated intensity
  - Other Methods
    - Normalize to a peak that is always present in the same concentration, for example normalizing to creatinine
    - Probabilistic quotient normalization
    - Quantile and cubic spline normalization

Centering, Scaling, and Transformations

I Centering

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_i$$

II Autoscaling

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{s_i}$$

Range scaling

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\left(x_{i_{max}} - x_{i_{min}}\right)}$$

Pareto scaling

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}}$$

Vast scaling

$$\tilde{x}_{ij} = \frac{\left(x_{ij} - \bar{x}_i\right)}{s_i} \cdot \frac{\bar{x}_i}{s_i}$$

Level scaling

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\bar{x}_i}$$

III Log transformation

$$\tilde{x}_{ij} = {}^{10}\log\left(x_{ij}\right)$$

$$\tilde{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_i$$

Power transformation

$$\tilde{x}_{ij} = \sqrt{\left(x_{ij}\right)}$$

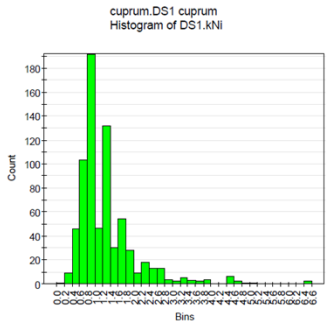
$$\tilde{x}_{ij} = \tilde{x}_{ij} - \bar{\tilde{x}}_i$$

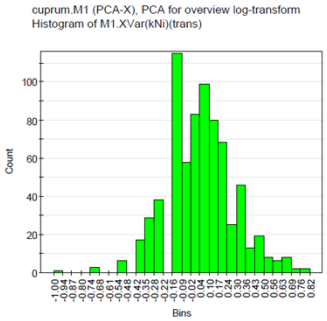
Analysis results vary depending on the scaling/ transformation methods used.

Van den Berg et al 1006, BMC Genomics, 7, 142

Data Transformation

- Before transformation
  - skew distribution
- After log-transformation
  - More close to normal distribution





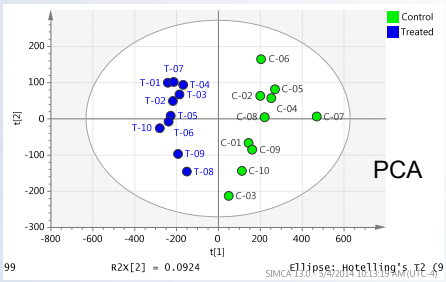
Susan Wicklund, Multivariate data analysis for omics, Sept 2-3 2008, Umetrics training

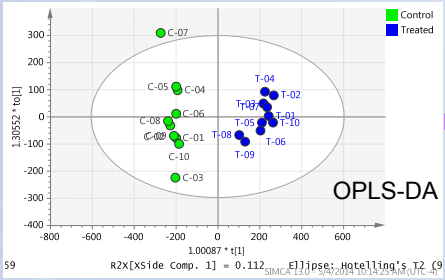
Scaling

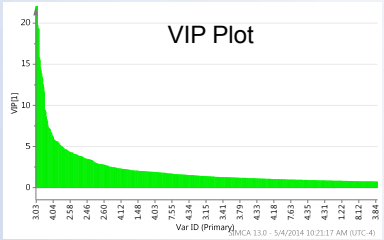
- Unit variance (autoscaling) divides the bin intensity by the standard deviation
  - May increase your baseline noise
  - Dimensionless value after scaling
- Pareto scaling divides the bin intensity by the square root of the standard deviation
  - Not dimensionless after scaling
- For NMR data, centering with pareto scaling is commonly used

Multivariate Data Analysis and Other Statistical Analyses

- Mean centered and scaled data
- Non-supervised analysis
  - Principal component analysis (PCA)
- Supervised analysis
  - PLS-DA and OPLS-DA
- Loadings plots and VIP Plots to identify discriminatory bins
- p-Value, fold change







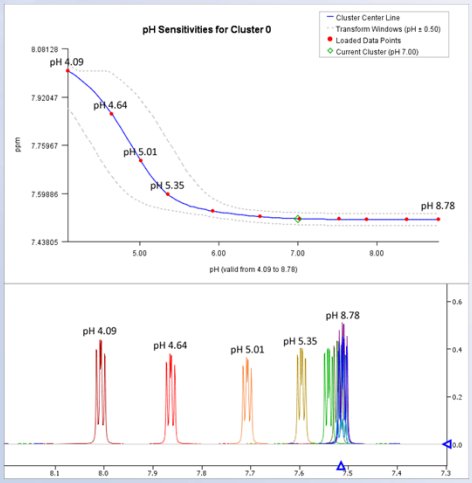
Library Matching
Pathway Analysis

Chenomx Library

- Over 320 metabolites
- pH sensitive library of ^{1}H NMR Spectra
- Customizable

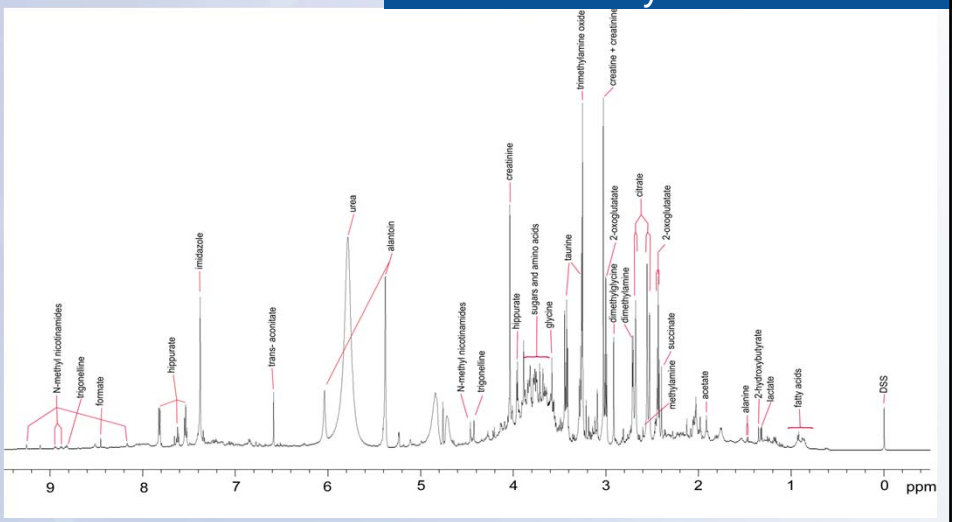
1,3-Dihydroxyacetone, 1,3-Dimethylurate, 1,6-Anhydro-β-D-glucose, 1,7-Dimethylxanthine, 1-Methylnicotinamide, 2'-Deoxyadenosine, 2'-Deoxyguanosine, 2'-Deoxyinosine, 2-Aminoadipate, 2-Aminobutyrate, 2-Ethylacrylate, 2-Furoate, 2-Hydroxy-3-methylvalerate, 2-Hydroxybutyrate, 2-Hydroxyglutarate, 2-Hydroxyisobutyrate, 2-Hydroxyisocaproate, 2-Hydroxyisovalerate, 2-Hydroxyphenylacetate, 2-Hydroxyvalerate, 2-Methylglutarate, 2-Octenoate, 2-Oxobutyrate, 2-Oxocaproate, 2-Oxoglutarate, 2-Oxoisovalerate, 2-Phosphoglycerate, 3,4-Dihydroxymandelate, 3,5-Dibromotyrosine, 3-Aminoisobutyrate, 3-Chlorotyrosine, 3-Hydroxy-3-methylglutarate, 3-Hydroxybutyrate, 3-Hydroxyisovalerate, 3-Hydroxymandelate, 3-Hydroxyphenylacetate, 3-Indoxylsulfate, 3-Methyl-2-oxovalerate, 3-Methyladipate, 3-Methylxanthine, 3-Phenyllactate, 3-Phenylpropionate, 4-Aminobutyrate, 4-Aminohippurate, 4-Hydroxy-3-methoxymandelate, 4-Hydroxybenzoate, 4-Hydroxyhippurate, 4-Hydroxyphenethylamine, 4-Hydroxyphenyllactate, 4-Pyridoxate, 5,6-Dihydrothymine, 5,6-Dihydrouracil, 5-Aminolevulinate, 5-Hydroxyindole-3-acetate, 5-Hydroxylysine, 5-Methoxysalicylate, Acetaldehyde, Acetamide, Acetaminophen, Acetate, Acetoacetate, Acetone, Acetylsalicylate, Adenine, Adenosine, Adipate, Alanine, Allantoin, Alloisoleucine, Anserine, Arginine, Argininosuccinate, Asparagine, Aspartate, Benzoate, Betaine, Bilirubin, Butyrate, Caffeine, Caprate, Caprylate, Carnitine, Carnosine, Choline, Cinnamate, Citrate, Citrulline, Creatine, Creatinine, Cysteine, Cystine, Cytidine, Cytosine, DSS (Chemical Shift Indicator), Dimethylamine, Epicatechin, Ethanol, Ethanolamine, Ethylene glycol, Ethylmalonate, Ferulate, Formate, Fructose, Fucose, Fumarate, Galactarate, Galactitol, Galactonate, Galactose, Gentisate, Glucarate, Glucose, Glutamate, Glutamine, Glutarate, Glutaric acid monomethyl ester, Glutathione, Glycerate, Glycerol, Glycine, Glycolate, Glycylproline, Guanidoacetate, Guanine, Hippurate, Histidine, Homocitrulline, Homocystine, Homogentisate, Homoserine, Homovanillate, Hypoxanthine, Ibuprofen, Imidazole, Indole-3-acetate, Inosine, Isobutyrate, Isocaproate, Isocitrate, Isoleucine, Isopropanol, Isovalerate, Kynurenate, Kynurenine, Lactate, Lactose, Leucine, Levulinate, Lysine, Malate, Maleate, Malonate, Mannitol, Mannose, Methanol, Methionine, Methylamine, Methylguanidine, Methylmalonate, Methylsuccinate, N,N-Dimethylformamide, N,N-Dimethylglycine, N-Acetylaspartate, N-Acetylglutamate, N-Acetylglycine, N-Carbamoyl-β-alanine, N-Carbamoylaspartate, N-Isovaleroylglycine, NAD+, Niacinamide, Nicotinate, O-Acetylcarnitine, O-Phosphocholine, O-Phosphoethanolamine, O-Phosphoserine, Ornithine, Oxalacetate, Oxypurinol, Pantothenate, Phenol, Phenylacetate, Phenylacetylglycine, Phenylalanine, Pimelate, Proline, Propionate, Propylene glycol, Protocatechuate, Pyridoxine, Pyroglutamate, Pyruvate, Quinolinate, Riboflavin, Ribose, S-Adenosylhomocysteine, S-Sulfocysteine, Salicylate, Salicylurate, Sarcosine, Serine, Suberate, Succinate, Succinylacetone, Sucrose, Tartrate, Taurine, Theophylline, Threonate, Threonine, Thymine, Thymol, Tiglylglycine, Trigonelline, Trimethylamine, Trimethylamine N-oxide, Tryptophan, Tyramine, Tyrosine, Uracil, Urea, Uridine, Urocanate, Valerate, Valine, Valproate, Vanillate, Xanthine, Xanthosine, Xylose, cis-Aconitate, myo-Inositol, o-Cresol, p-Cresol, trans-4-Hydroxy-L-proline, trans-Aconitate, β-Alanine, π-Methylhistidine, τ-Methylhistidine

Chemical Shift and pH Dependence



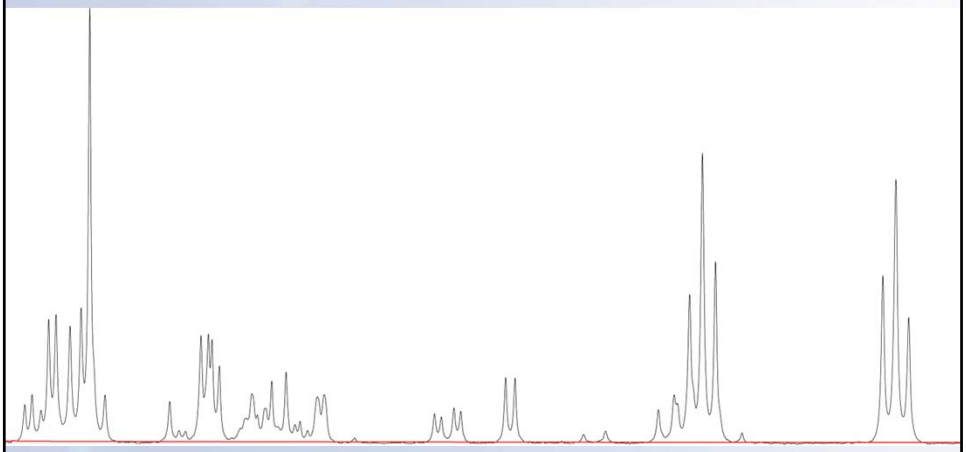
Source: http://www.chenomx.com/software/

NMR Spectrum of Urine with Chenomx Library Fit of Metabolites



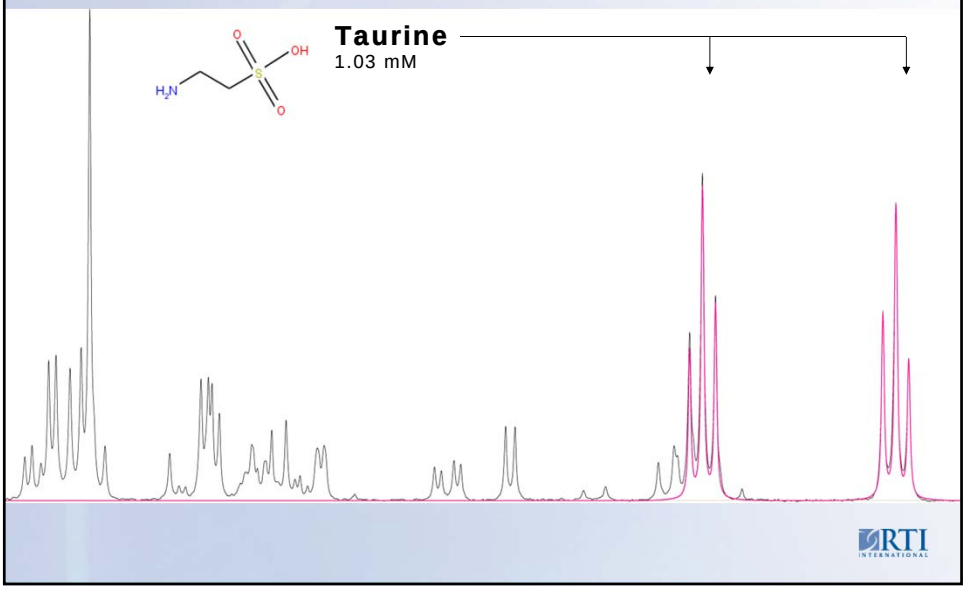
RTI International

Fitting of Metabolites



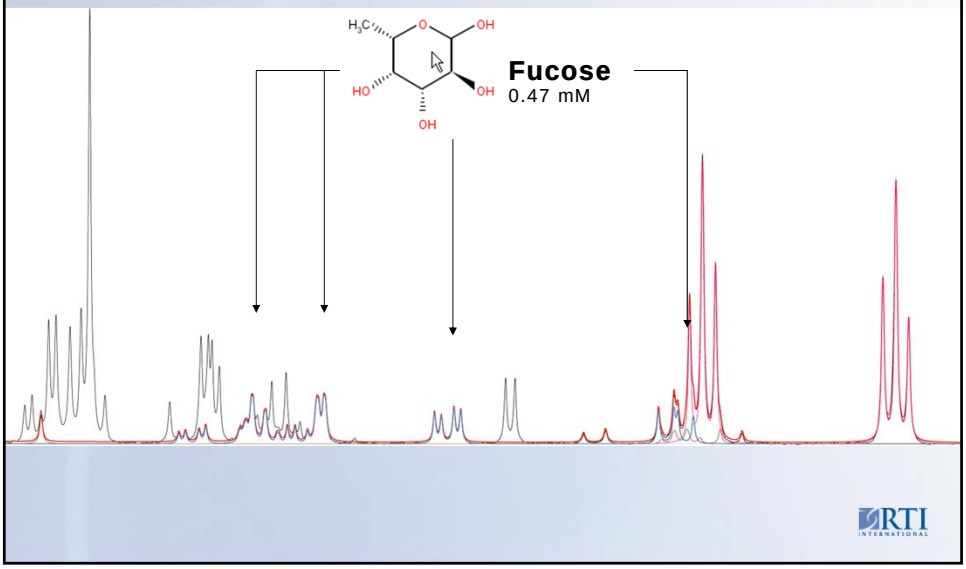
RTI International

Fitting Taurine



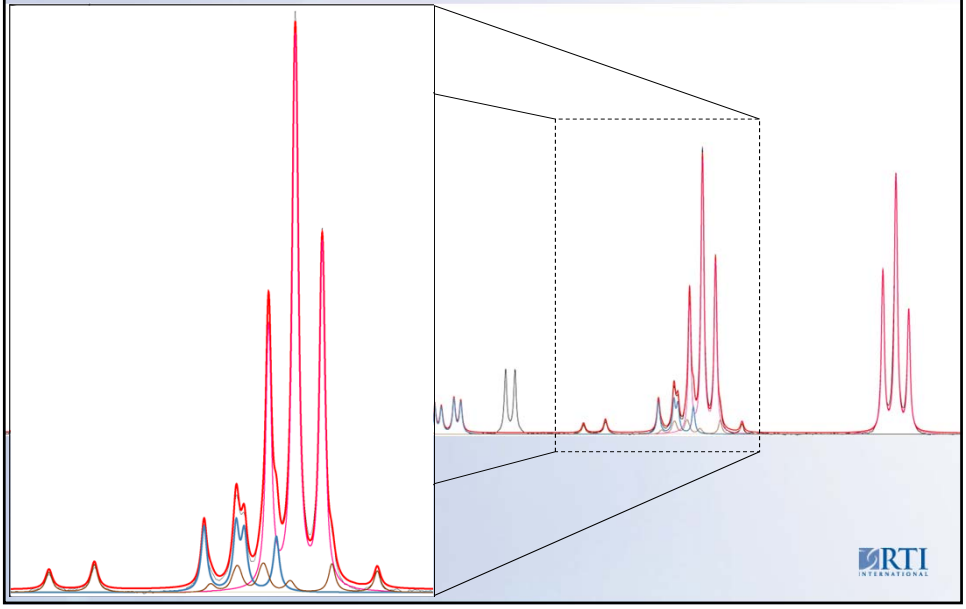
RTI International

Fitting Fucose

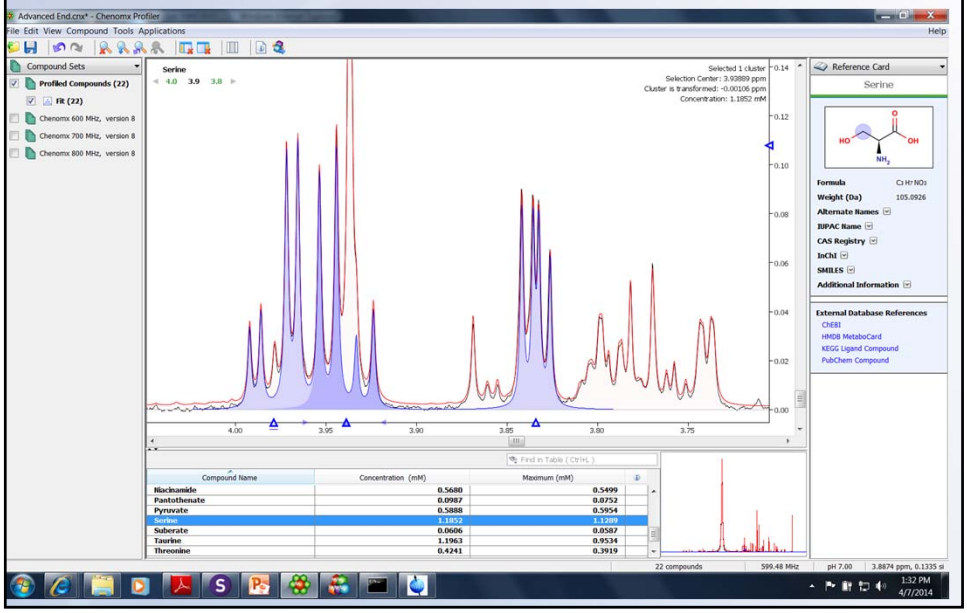


RTI International

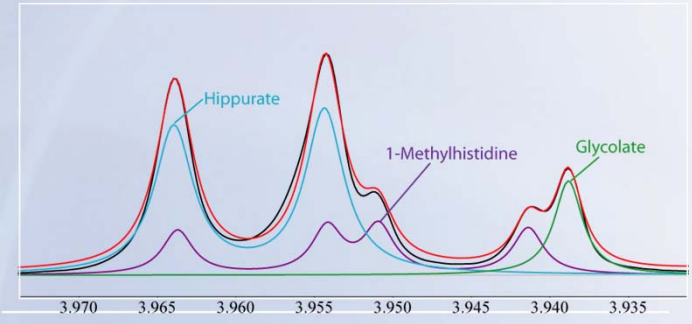
Additive Fit



Additive Fit



Chenomx Helps Resolving Ambiguity in Highly Overlapped Regions



Interpreting Results and Pathway Analysis

Once we have performed a metabolomics analysis:

- We find some important metabolites that are responsible for the separation of study groups.
- The next question is “What it means?
- How do you correlate these finding to your study questions?
- Does it explain any findings that are meaningful for your study hypotheses?
- Does it generate a new hypothesis?
- How do you answer these questions?

Next step is to interpret results and metabolic pathway analysis

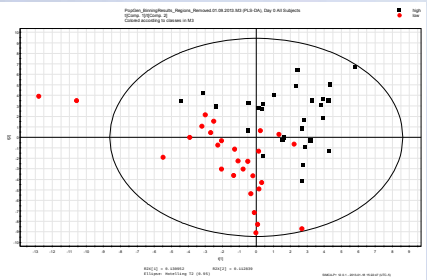
RTI International

Interpreting Results and Pathway Analysis

- There are a number of freely available software
  - meta-P Server, Metaboanalyst, Met-PA, web based KEGG Pathways, Cytoscape.
  - GeneGo, Ingenuity Pathway Analysis (Commercial)
- Another way of interpreting metabolomics results is to use traditional biochemistry text books.
- The input for pathway analysis is typically a list of metabolites (with any fold change or p-value information)
- Genomics, transcriptomics, and/or proteomics data can be integrated
- Once these pathways are identified, you may perform a targeted metabolomics analysis to validate the findings from global analysis.

Day 0 serum- Predicting Day 28 Response to Vaccine

PLS-DA
Day 0 – High Responders (Black) vs Low Responders (Red)



Subset of Metabolites that Influence the Separation of Subjects at Day 0 (VIP ≥ 1 or p-value ≤ 0.1)	
Isoleucine**	Creatinine**
Leucine**	Cysteine**
Valine	Histidine
3-Methyl-2-oxo-isovalerate	Choline
3-Hydroxybutyrate	Glucose
Lactate	Betaine
Alanine	TMAO
Acetate**	Glycine
Proline*	Glycerol
Glutamate**	Serine
Glutamine**	Creatine
Pyruvate	Tyrosine*
2-Oxoisocaproate	Histidine
Methylguanidine**	Tryptophan
Formate	Phenylalanine

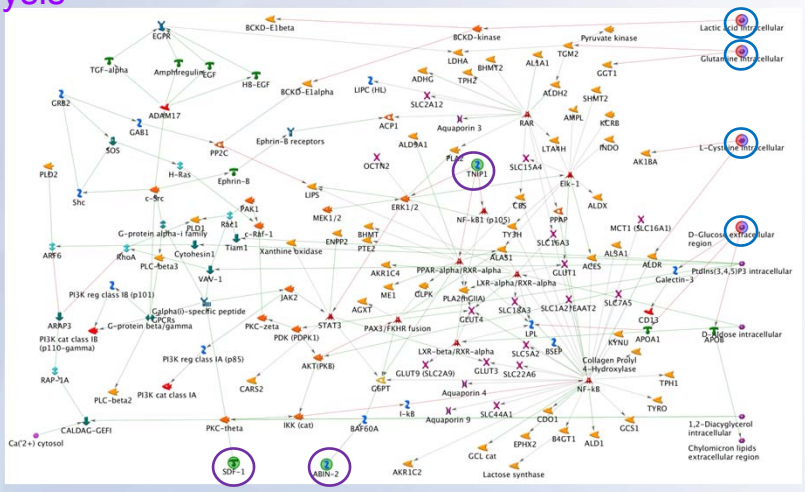
Preliminary results

*p-value < 0.05, **p-value ≤ 0.1

RTI International

GeneGo Network Analysis

Day 0 High vs Low Responders



- Receptor ligands/binding proteins related to gene markers from genetics analysis. Majumder et al. 2012, Eur. J. Human Genetics, 1-7
- Metabolites that linked in the pathways

Preliminary results

Literature

NMR data acquisition is performed by using methods cited in Beckonert et al. (2007), Nature Protocols, 2 (11), 2692-2703.

Xia, J. et al (2011) Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst, Nature Protocols 6, 743–760 (2011) doi:10.1038/nprot.2011.319

Hao, J. et al. (2014) Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN, Nature Protocols 9, 1416–1427 (2014) doi:10.1038/nprot.2014.090

Savorani , F. et al, Journal of Magnetic Resonance, Volume 202, Issue 2, 2010, 190 – 202

Vu, T. N. et al., BMC Bioinformatics 2011, 12:405

RTI International

NMR Metabolomics Hands-On Exercise

NMR Hands On Exercise

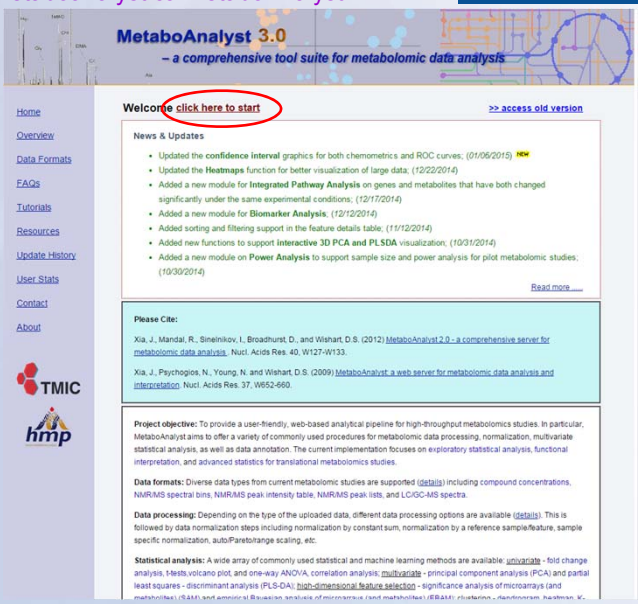
- Drug Induced Liver Injury (DILI) Study using Rat Model
- 3 Study groups and 2 time points
  - Vehicle Control (time matched)
  - Low Dose ("no effect" level, Day 01 and Day 14)
  - High Dose (Day 01 and Day 14)
- 24h Urine collected
- Samples prepared by mixing an aliquot of urine with Phosphate buffer + Chenomx ISTD (DSS, D_2O, 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)

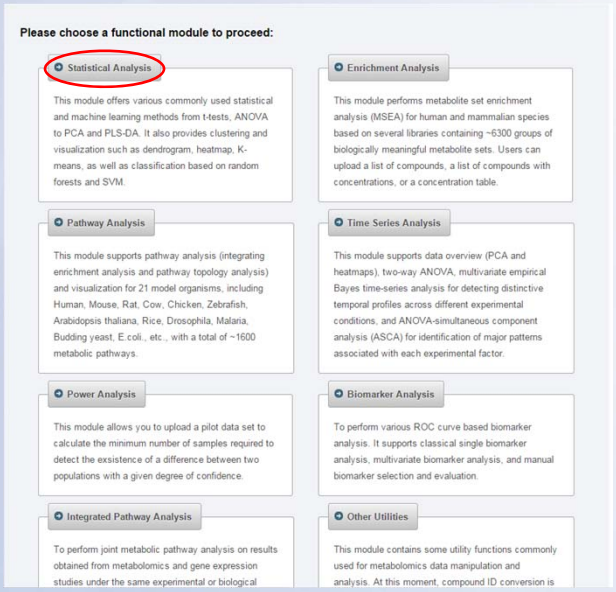
MetaboAnalyst 3.0

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

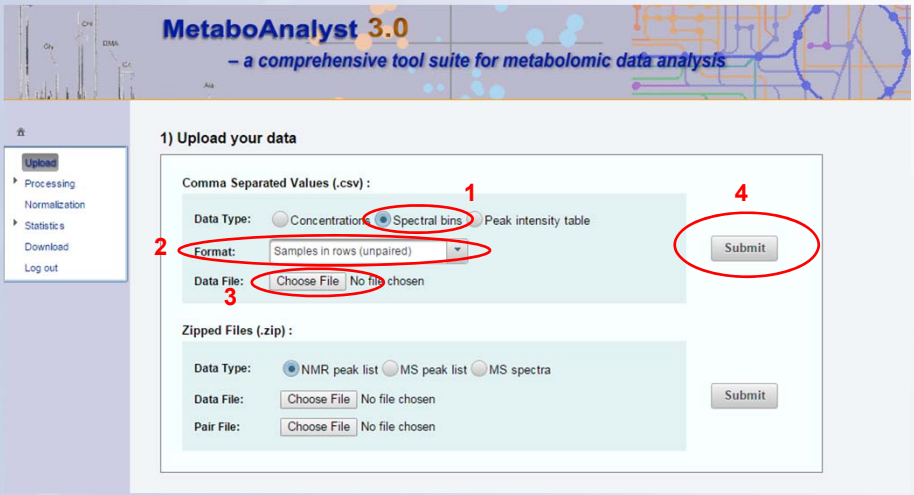


RTI International

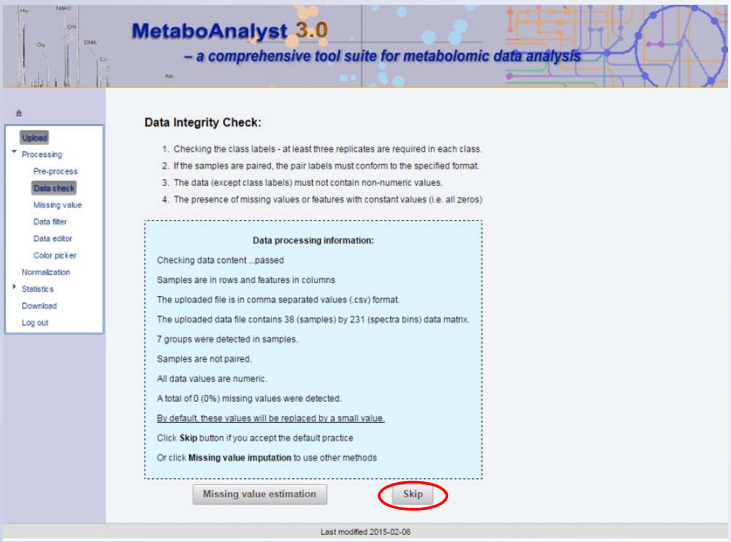
MetaboAnalyst: Functional Modules



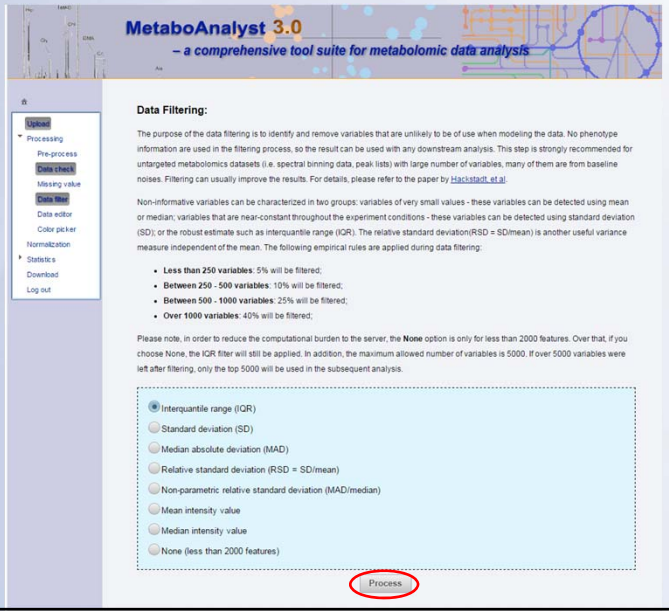
Data Upload



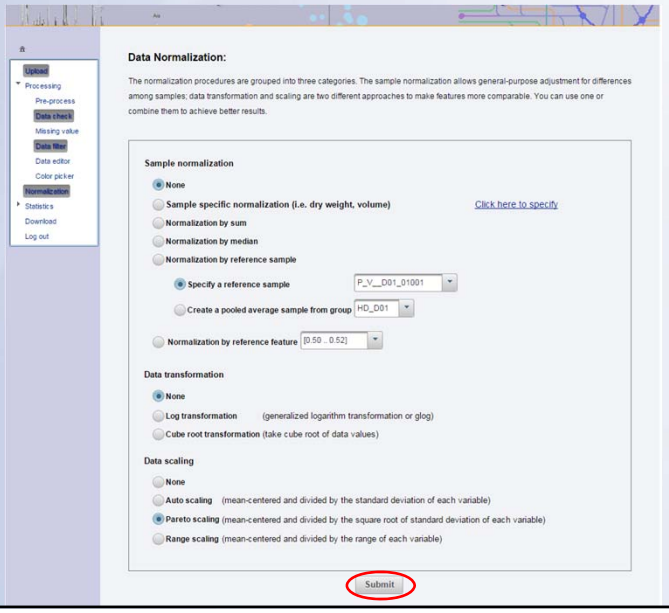
Data Integrity Check



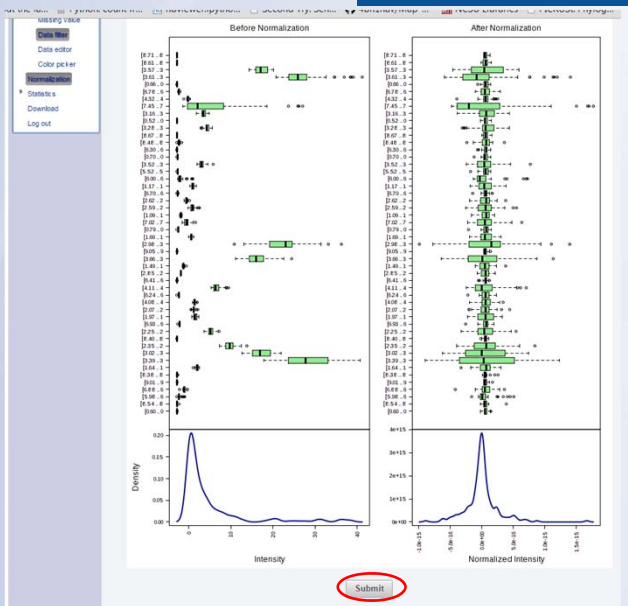
Data Filtering



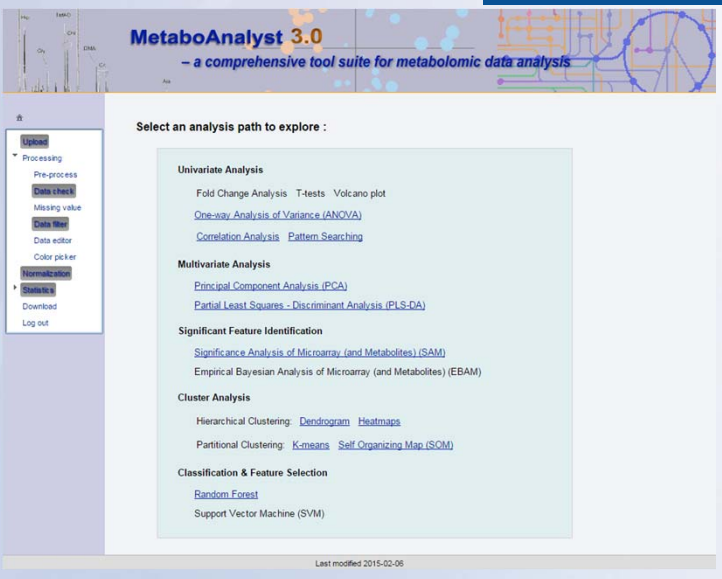
Data Normalization



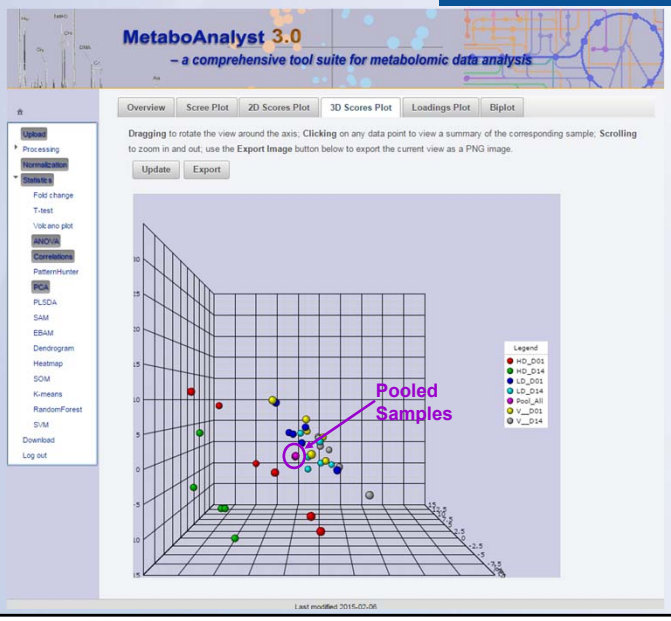
Summary: Normalization



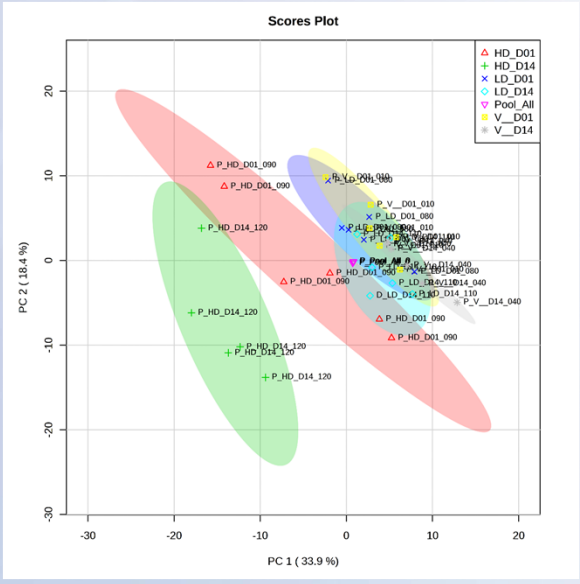
Statistical Analysis



Pooled QC Samples



PCA Day 01 and Day 14



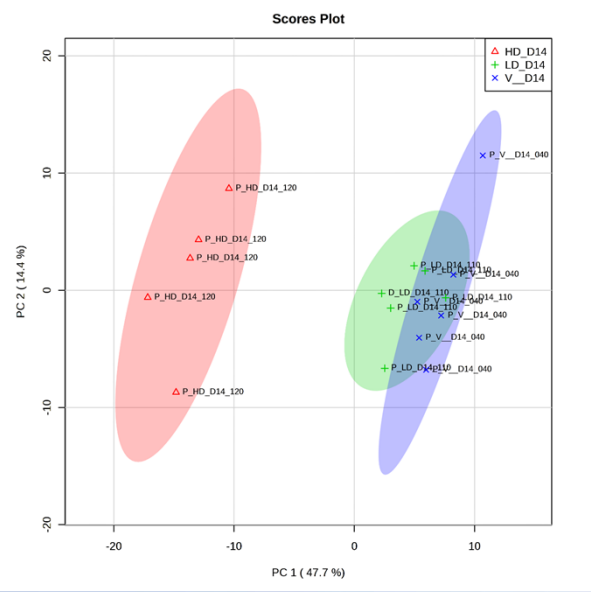
Day 14: Vehicle, Low Dose, and High Dose Groups

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

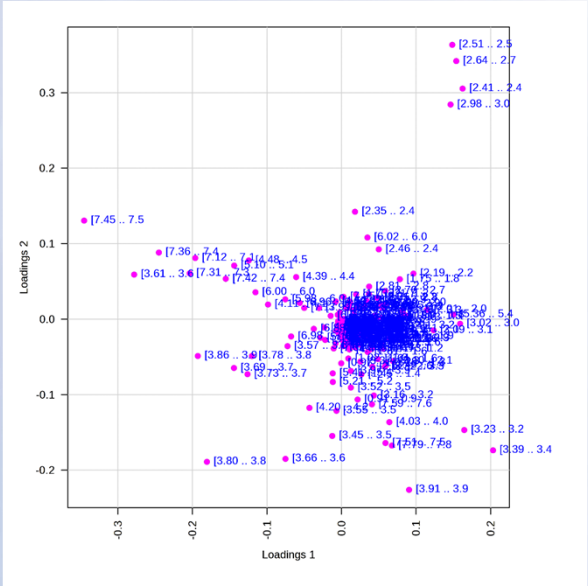
Day 14 PCA Scores Plot



RTI International

Vehicle, Low Dose, and High Dose groups

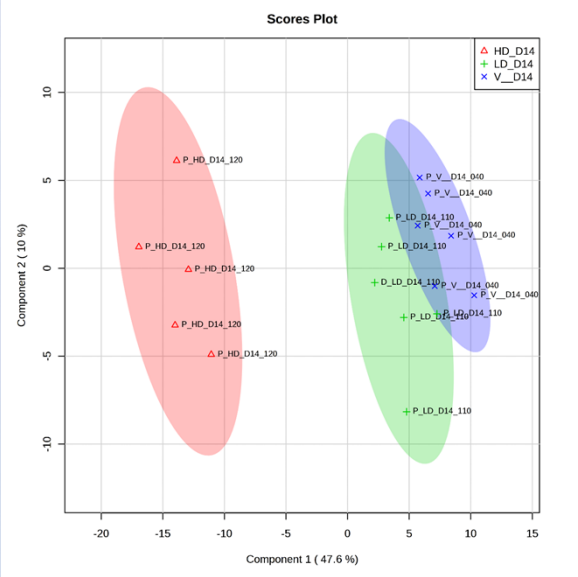
PCA Loadings Plot



RTI International

Vehicle, Low Dose, and High Dose groups

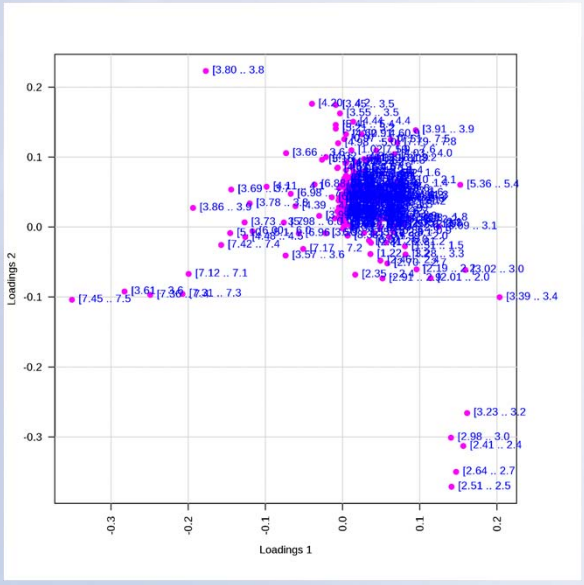
PLS-DA Scores Plot



RTI International

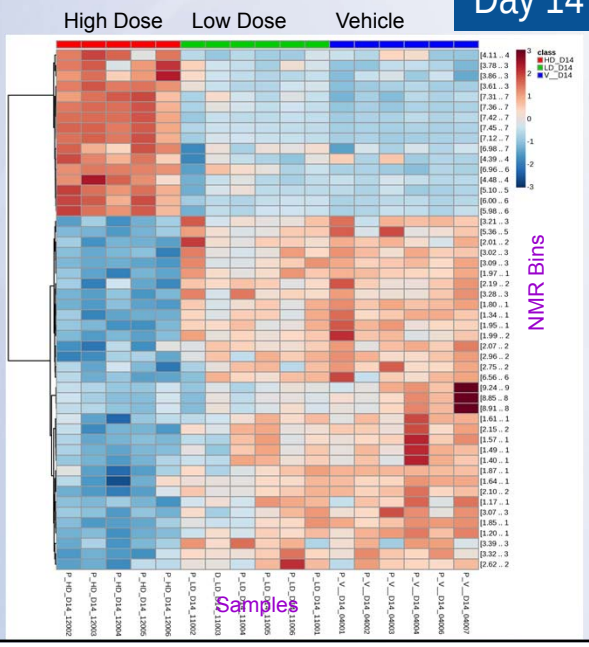
Vehicle, Low Dose, and High Dose groups

PLS-DA Loadings Plot



RTI International

Day 14 Heat Map

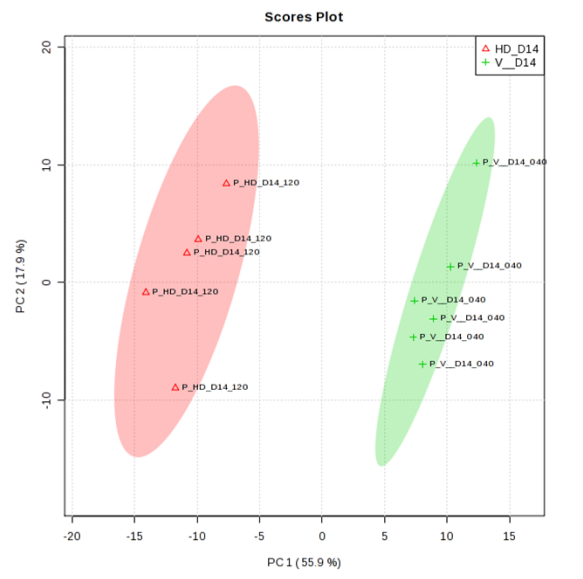


Comparison of Day 14 High Dose and Vehicle

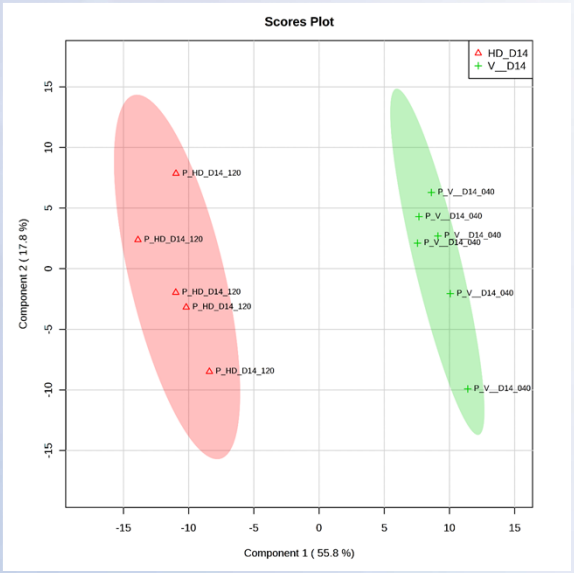
Please start from the start page and upload the data

- 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

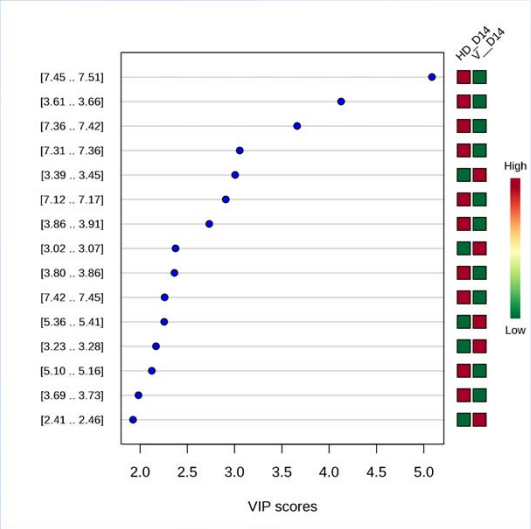
Day 14 PCA Scores Plot: High Dose vs Vehicle



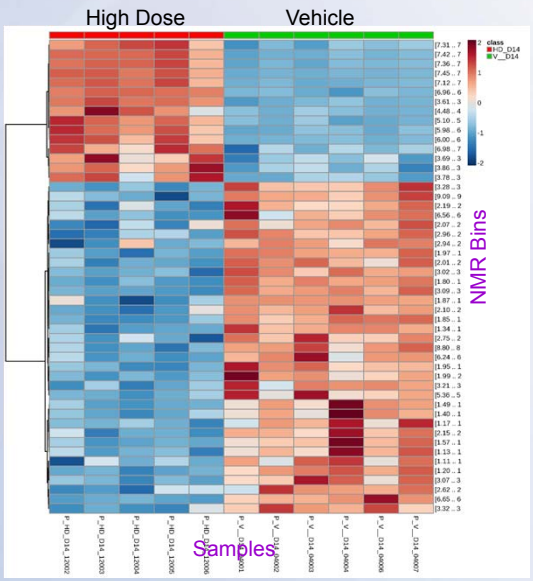
Day 14 PLS-DA Scores Plot: High Dose vs Vehicle



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



Day 14 Heat Map: High Dose vs Vehicle



Top 50 bins in the VIP Plot

Thank You!

If you have any questions, please e-mail me

wpathmasiri@rti.org

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

ACKNOWLEDGEMENTS

Director RTI RCMRC
Susan Sumner, PhD

Program Coordinator
Jason Burgess, PhD

NIH Scientific Officer
David Balshaw, PhD, NIH/NIEHS

Internship Program
Stella Lam, BS

Feasibility Studies
Susan Sumner, PhD
Susan McRitchie, MS
Executive Committee

Website
Roger Austin, MS

Biochemistry and Molecular Biology
Timothy Fennell, PhD
Ninell Mortensen, PhD
Delisha Stewart, PhD

Biorepository
Brian Thomas, PhD
Mike McCleary, BS

Interns
Tammy Cavallo
Aastha Ghimire
Zachery Acuff

LC-MS Metabolomics
Suraj Dhungana, PhD
Brian Thomas, PhD
James Carlson, MS
Alex Kovach, BS
Rodney Snyder, MS
Moses Darko, BS

GC-MS Metabolomics
Wimal Pathmasiri, PhD
Jocelin Deese-Spruill, BS
Keith Levine, PhD
James Harrington, PhD
William Studabaker, PhD

NMR Metabolomics
Wimal Pathmasiri, PhD
Kelly Mercier, PhD
Rodney Snyder, MS
Tammy Cavallo, BS
Kevin Knagge, PhD, DHMRI
Jason Winnike, PhD, DHMRI

Statistics, Bioinformatics, and Computing
Susan McRitchie, MS
Robert Clark, PhD
Andrew Novokhatny, BS

Advisors: Imperial College, UK
Jeremy Nicholson, PhD Elaine Holmes, PhD
Ian Wilson, PhD