Metabolomics 101

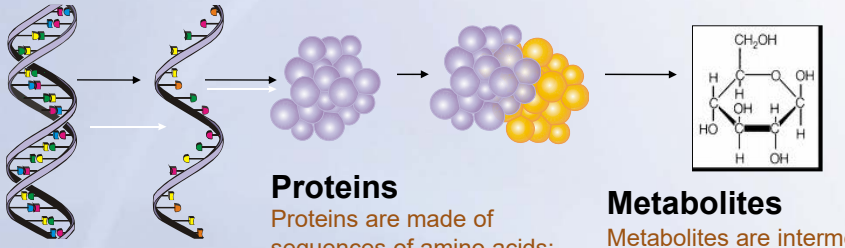
UAB Metabolomics Training Course
July 17-21, 2017

Wimal Pathmasiri and Delisha Stewart
NIH Eastern Regional Comprehensive Metabolomics Resource Core (ERCMRC)
Department of Nutrition – Nutrition Research Institute
University of North Carolina at Chapel Hill

Metabolomics

- The metabolome is the low molecular weight complement of cells, tissues, or biological fluids.
- Metabolomics investigations generally employ NMR or one of a number of types of chromatography coupled MS methods
- Metabolomics makes it feasible to uniquely profile the biochemistry of an individual, or model, apart from, or in addition to, the genome.
- Metabolomics is being used to reveal biomarkers for the early detection and diagnosis of disease, to predict outcomes, monitor therapeutic treatments and interventions, and to provide insights into biological mechanisms.

The relation of proteins and metabolites to the genome



DNA

DNA contains genetic instructions to
- make components of cells
- regulate the use of these components

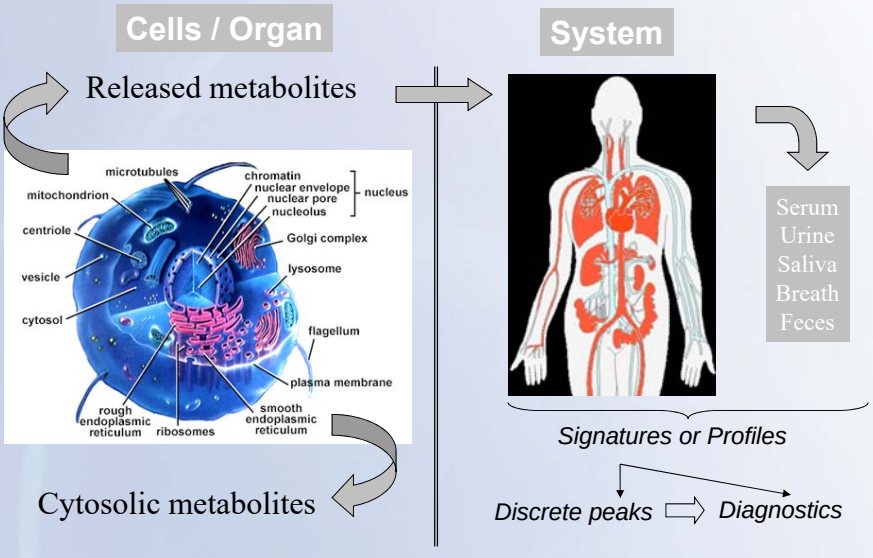
Proteins

Proteins are made of sequences of amino acids; the sequence defined by the gene.
Proteins are the enzymes that catalyze or accelerate chemical reactions in metabolism

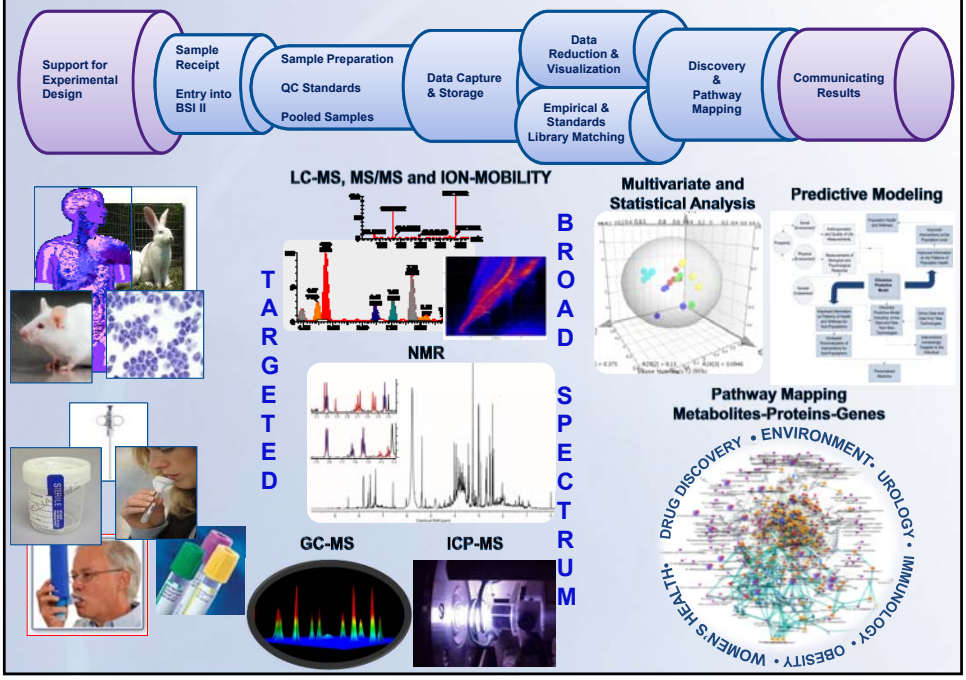
Metabolites

Metabolites are intermediates and products of metabolism.
Catabolism: the processes to breaks down large molecules.
Anabolism: the process to use catabolism energy to synthesize molecules

Cells, Tissues, and Noninvasive Fluids

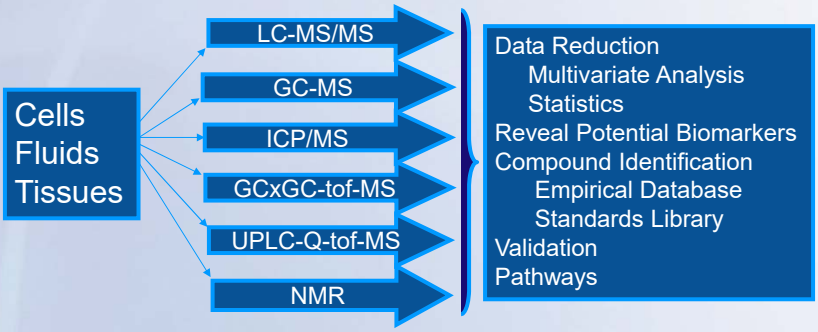


NIH Eastern Regional Comprehensive Metabolomics Resource Core at NRI



ERCMRC Metabolomics Technologies

The analysis of the small molecule diversity present in a biological system and the pattern of changes arising from disease, dysfunction, disorder, or from the therapeutic or adverse effects of drugs.

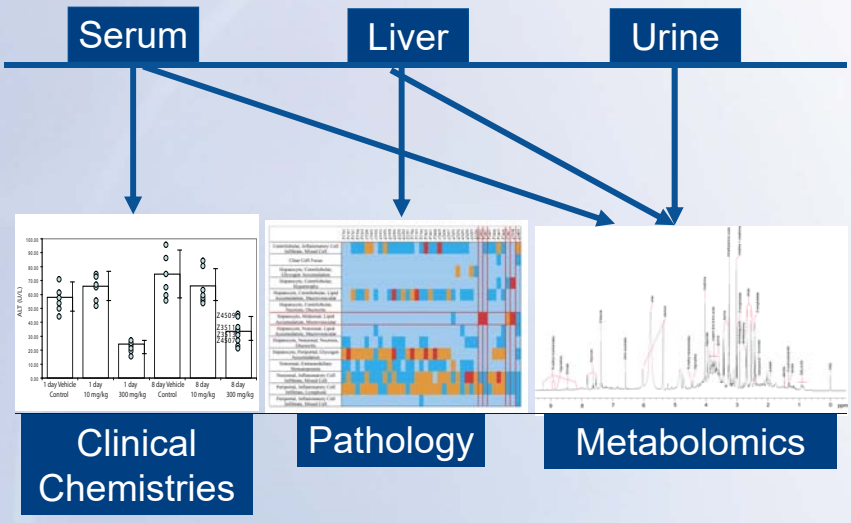


~6,500 discrete small molecule metabolites, ~ 25,000 genes, ~100,000 transcripts, and 1,000,000 proteins

Optimal and Minimal Sample Volumes

	Minimum sample for MS Based Detection	Minimum Sample for NMR-Based Detection	Optimal Sample
Serum	50 ul	100 ul	1 ml
Urine	50 ul	200 ul	1 ml
Feces	20 mg	20 mg	500 mg
Tissue	50 mg	100 mg	500 mg
Cells	$1x10^6$	$1x10^7$	$1x10^7$

Cells, Tissues, and Noninvasive Fluids

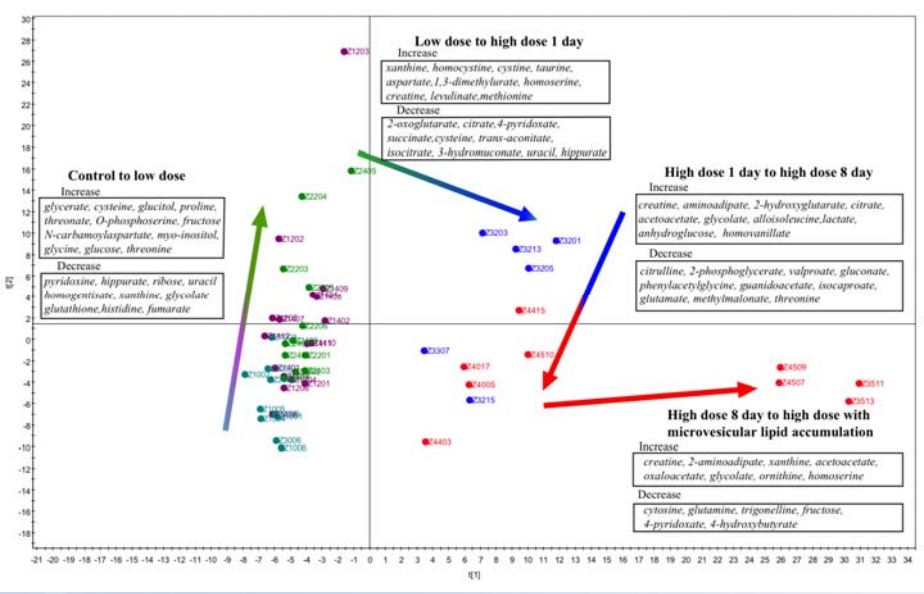


Preclinical: Monitoring for Adverse Side Effects: DILI

- Drug-induced liver injury (DILI) accounts for 80% of the drug failure rate: pre-clinical through post market.
- Non-invasive markers are needed to determine the potential for DILI during treatment.
- Patients taking the anti-TB drug, isoniazid (INH), are at risk for developing liver injury. INH is one of the five top drugs with causal relation to liver injury and transplant in the US.
- Rats were dosed with INH for 1 or 8 days at low dose 'no affect' levels and at concentrations that resulted in microvesicular lipid accumulation (MVLA) of the liver- a reversible pathology currently diagnosed by biopsy and pathology.
- Metabolomics was used to determine urinary markers to correlate with MVLA diagnosis and its onset.

Sumner et al., 2009 NIH Grant GM75903

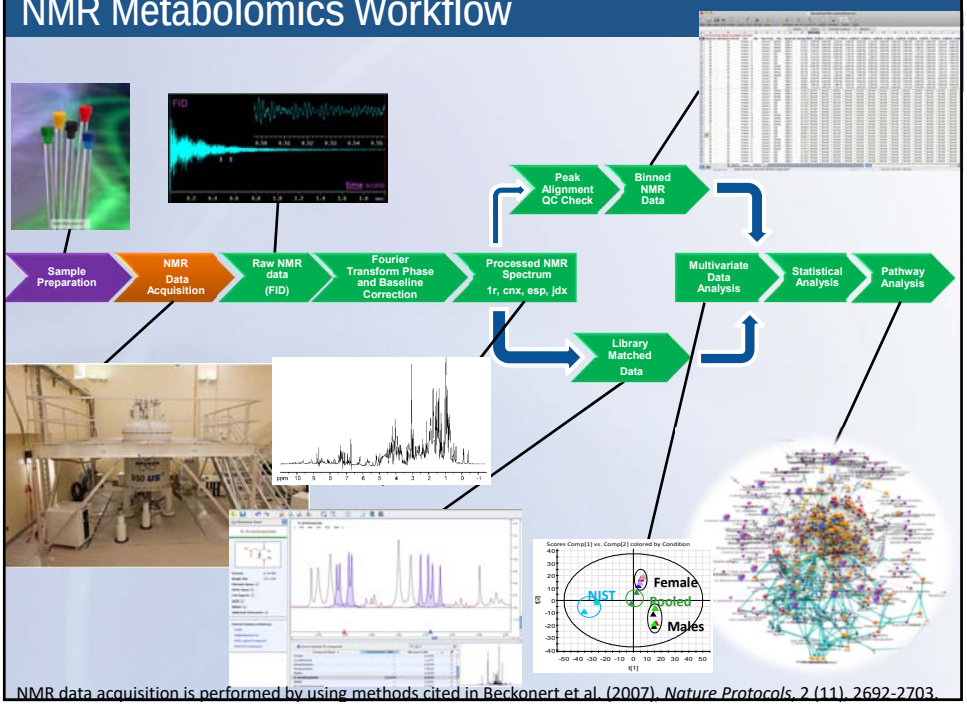
Cells, Tissues, and Noninvasive Fluids



NMR Based Metabolomics Analysis

- NMR Spectroscopy
  - A robust, reliable, and highly reproducible technique in metabolomics analysis
  - Quantitative and non-destructive method
  - Most labs use 600 – 950 MHz Spectrometers
  - The higher the field strength, the higher the sensitivity and resolution
- Broad-spectrum metabolomics
  - NMR binning (high throughput)
- Targeted metabolomics
  - Metabolite profiling and quantification of selected metabolites or a panel of metabolites

NMR Metabolomics Workflow



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

Important Steps

- 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**

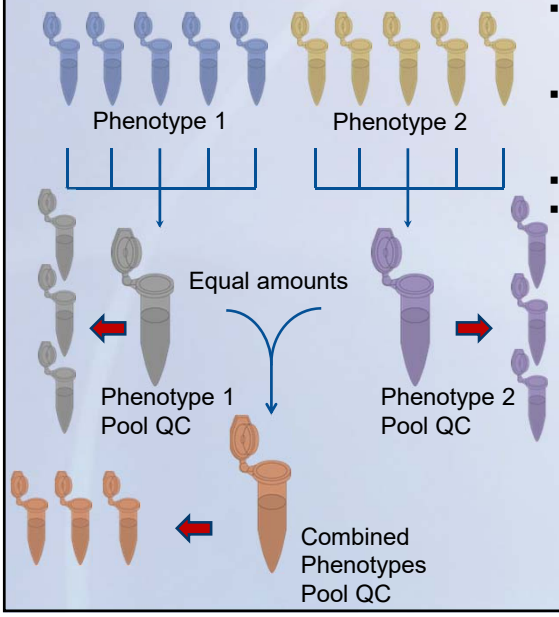
Sample Preparation for Metabolomics Analysis

Current sample preparation practices (in brief)

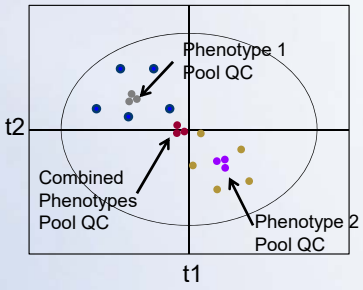
- **Biofluids**
  - Dilute with D_2O/ buffer/ 0.9% Saline
  - Add internal standard (ISTD, eg. Chenomx) solution or formate (for serum).
  - Centrifuge and transfer an aliquot into NMR tube
- **Tissue and Cells**
  - Homogenization performed in ice cold 50/50 acetonitrile/water
  - Supernatant dried 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

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

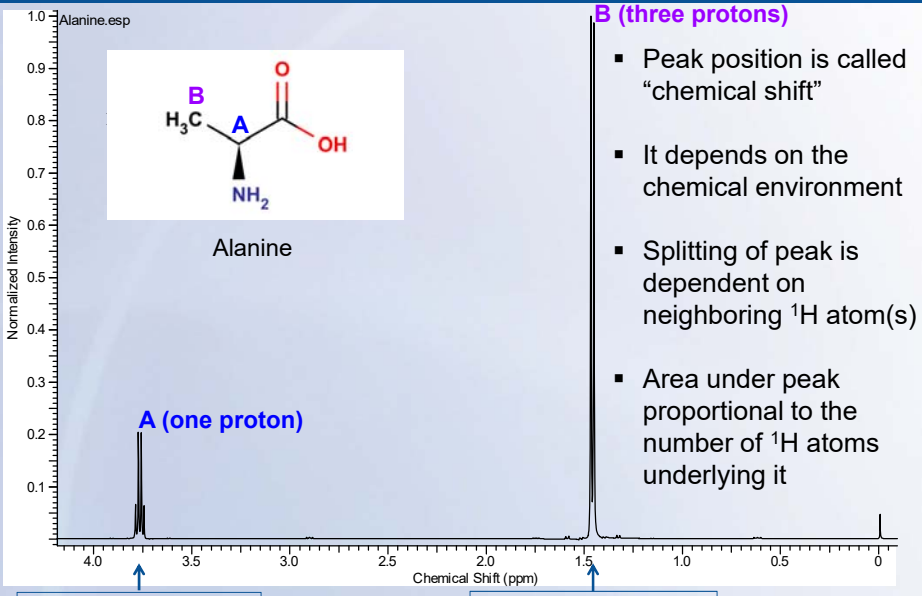


Pooled samples should cluster tightly

NMR Data

- 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)

^{1}H NMR Spectrum for alanine



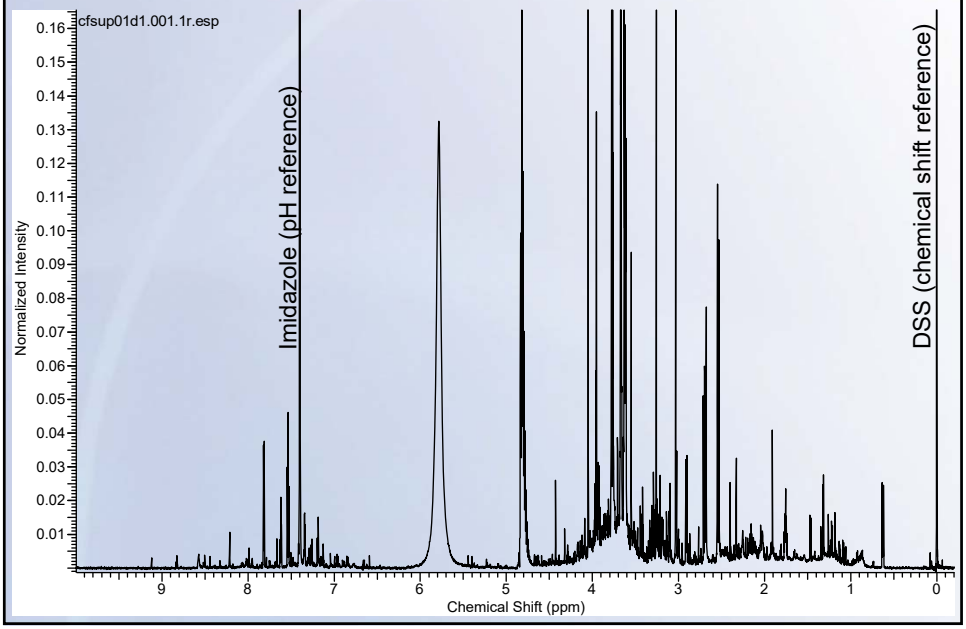
Alanine

- 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

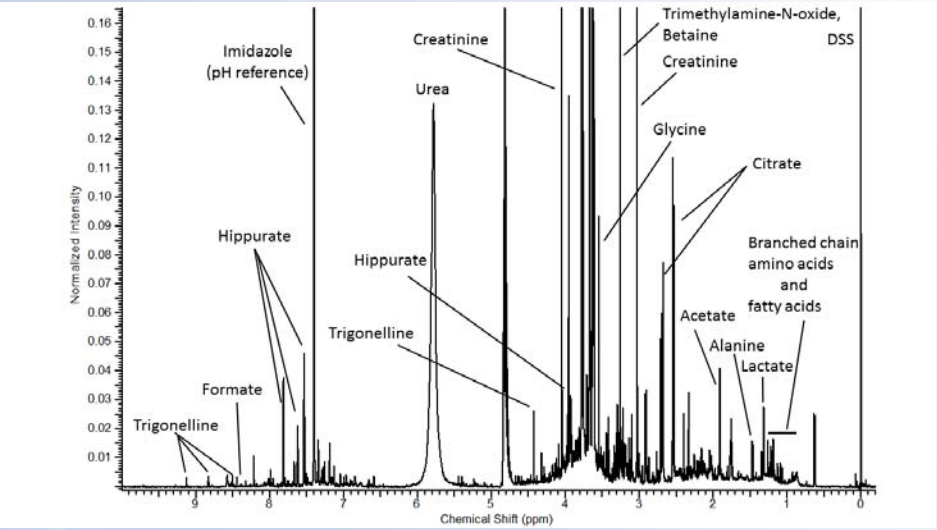
Chemical shift of C**H**
4 lines (quadrat)

Chemical shift of C**H$_3$**
2 lines (doublet)

Typical ^{1}H NMR Spectrum of urine

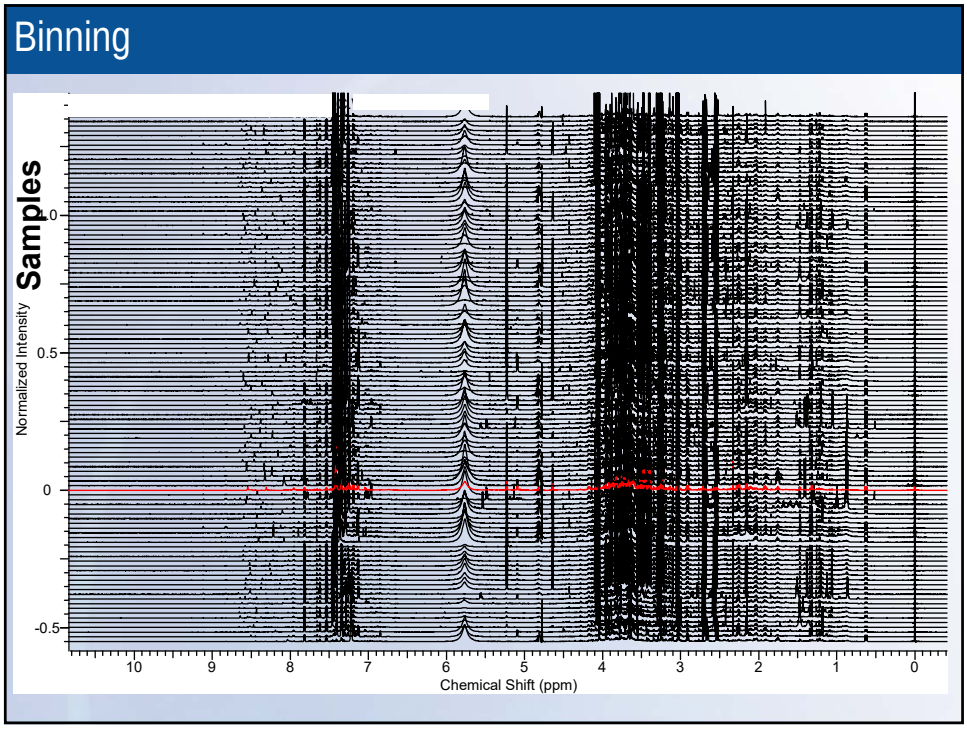


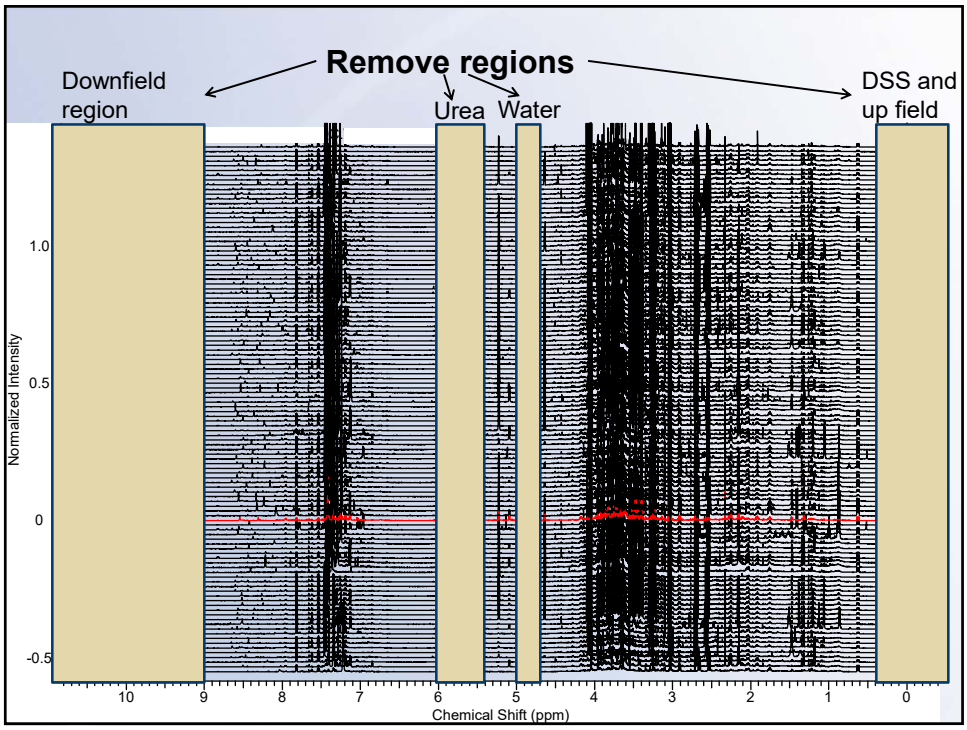
Typical ^{1}H NMR Spectrum of Urine (annotated)



Broad Spectrum Metabolomics
NMR Binning

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)


Remove regions
Downfield region
Urea
Water
DSS and up field
Normalized Intensity
1.0
0.5
0
-0.5
10
9
8
7
6
5
4
3
2
1
0
Chemical Shift (ppm)

Binning

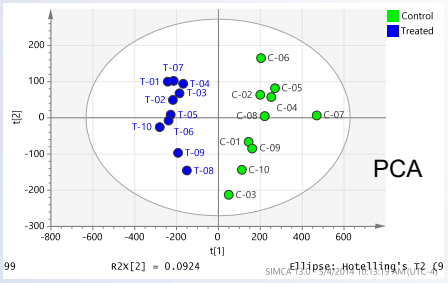
- Integrate bins (0.04 ppm bin size)
- Normalize bins to the total integral of each spectrum
- Merge metadata
- Result is a spreadsheet ready for further multivariate data analysis and other statistical analysis

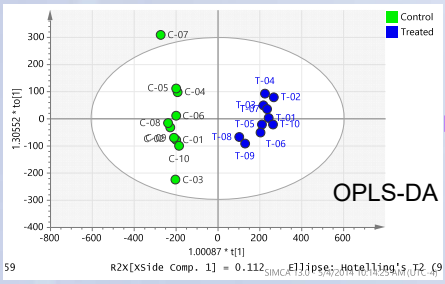
Sample ID	Disease Group	[0.40 .. 0.46]	[0.46 .. 0.52]	[0.52 .. 0.54]	[0.54 .. 0.57]	[0.57 .. 0.60]	[0.60 .. 0.66]	[0.66 .. 0.68]	[0.68 .. 0.71]	[0.71 .. 0.75]
C0559	Cases	7.60E-05	0.00E+00	7.32E-02	8.48E-02	3.20E-02	1.84E+00	1.31E-01	3.60E-01	3.67E-01
C0629	Cases	0.00E+00	1.78E-02	0.00E+00	2.18E-02	0.00E+00	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
E0225	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
E0309	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

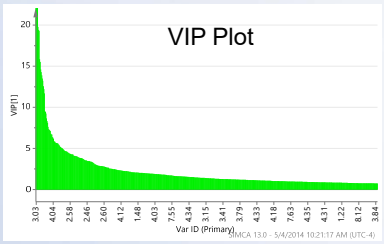
Multivariate Data Analysis & Other Statistical Analysis

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







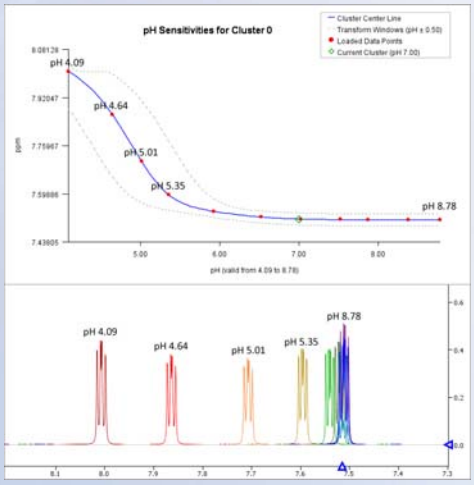
Library Matching (and quantifying) Using Chenomx

Chenomx Library

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

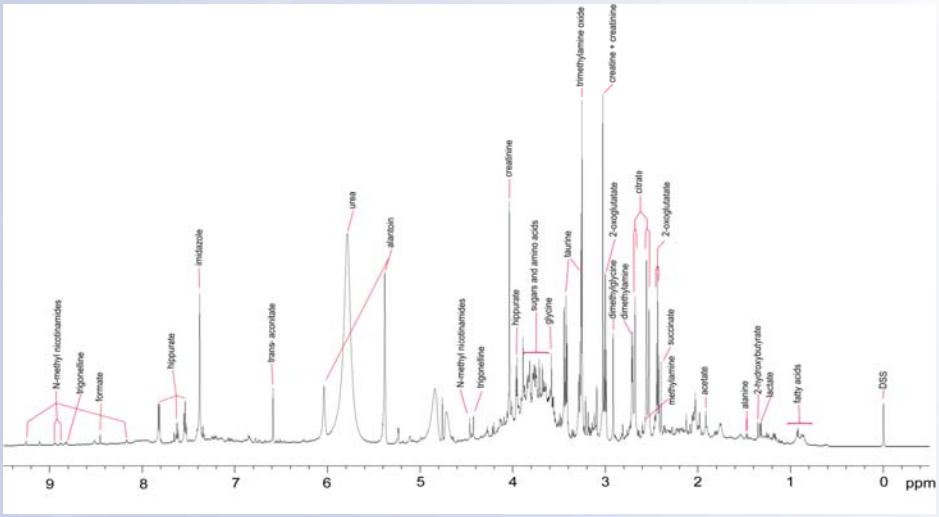
1,3-Dihydroxyacetone, 1,3-Dimethylurate, 1,6-Anhydro-β-D-glucose, 1,7-Dimethylxanthine, 1-Methylnicotinamide, 2'-Deoxyadenosine, 2'-Deoxyguanosine, 2'-Deoxyinosine, 2-Aminoadipate, 2-Aminobutyrate, 2-Ethylacrylate, 2-Furoate, 2-Hydroxy-3-methylvalerate, 2-Hydroxybutyrate, 2-Hydroxyglutarate, 2-Hydroxyisobutyrate, 2-Hydroxyisocaproate, 2-Hydroxyisovalerate, 2-Hydroxyphenylacetate, 2-Hydroxyvalerate, 2-Methylglutarate, 2-Octenoate, 2-Oxobutyrate, 2-Oxocaproate, 2-Oxoglutarate, 2-Oxoisocaproate, 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-Hydroxyphenylacetate, 4-Hydroxyphenyllactate, 4-Pyridoxate, 5,6-Dihydrouridine, 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, Butyrate, Caffeate, 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

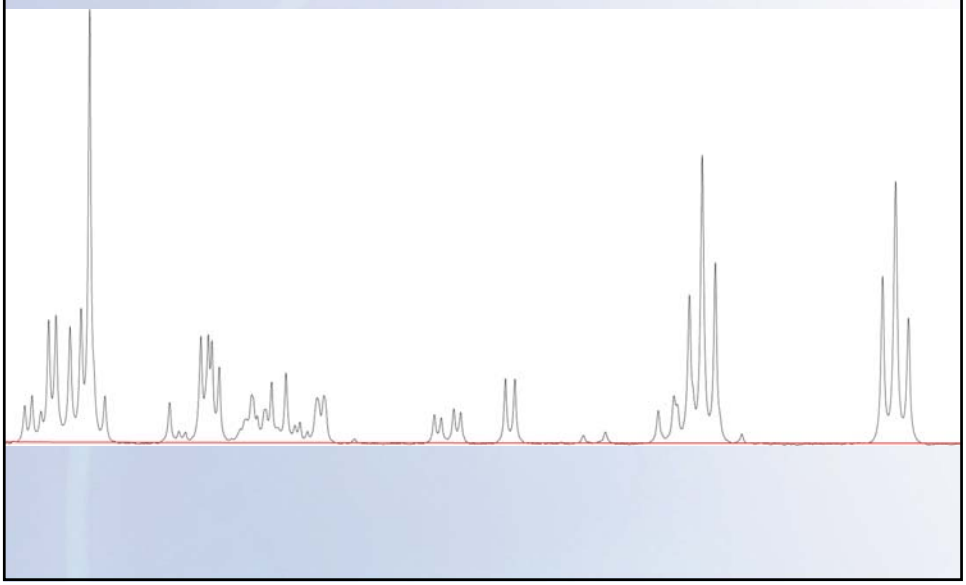


Source: http://www.chenomx.com/software/

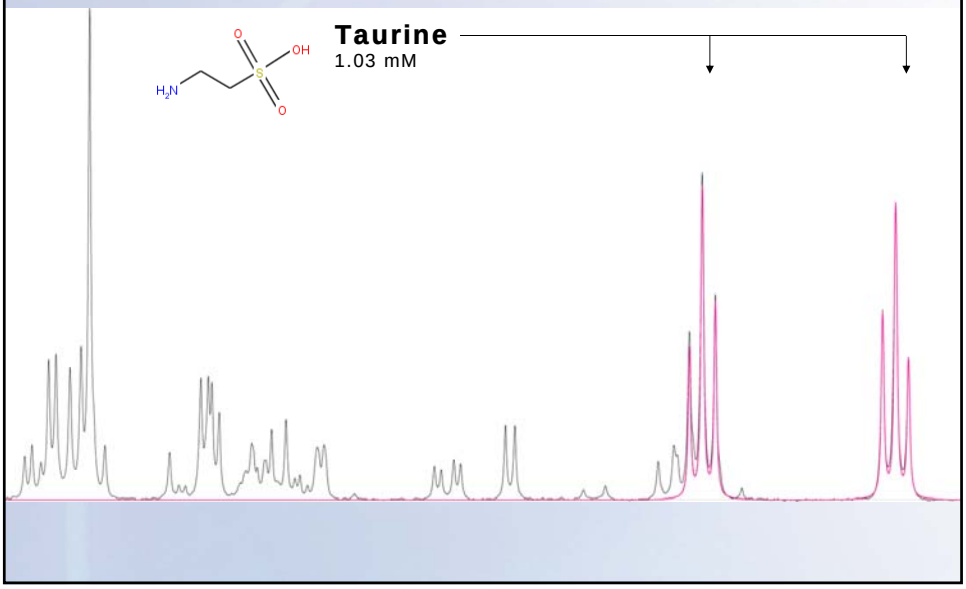
NMR Spectrum of Urine with Chenomx Library Fit of Metabolites



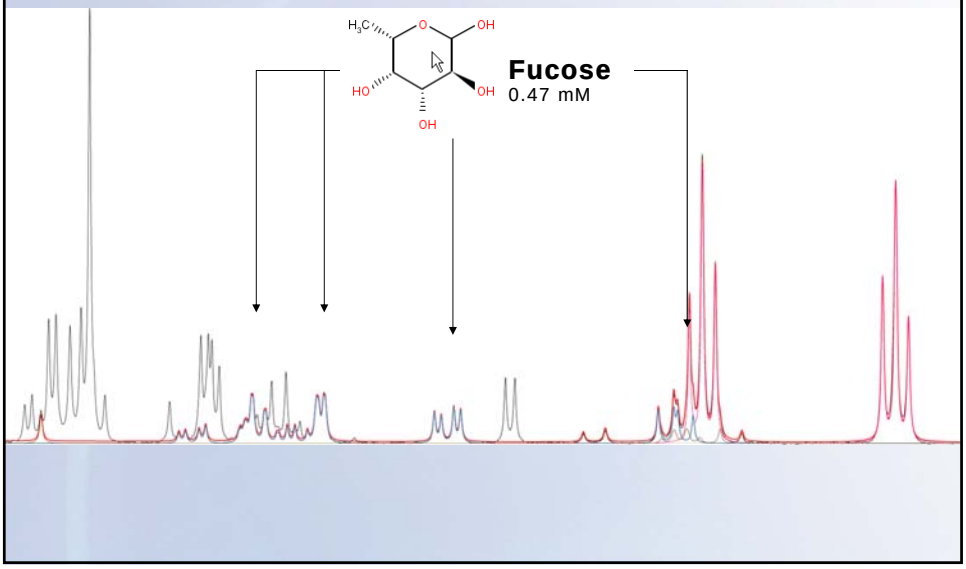
Fitting of metabolites



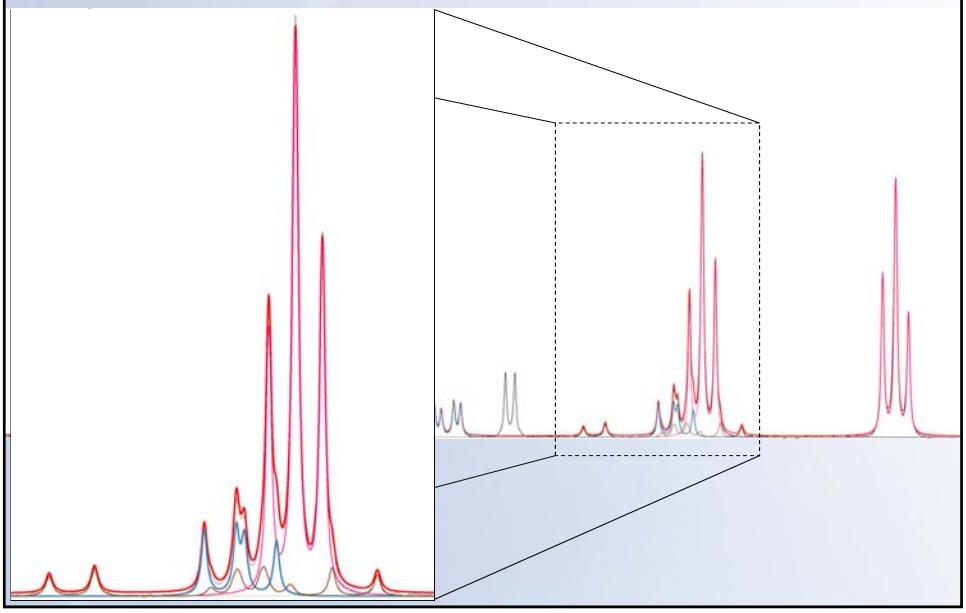
Fitting taurine



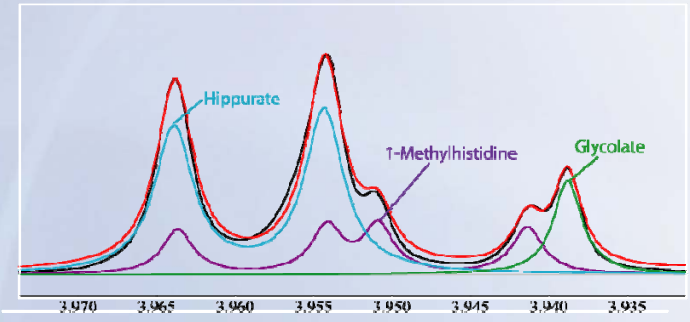
Fitting fucose



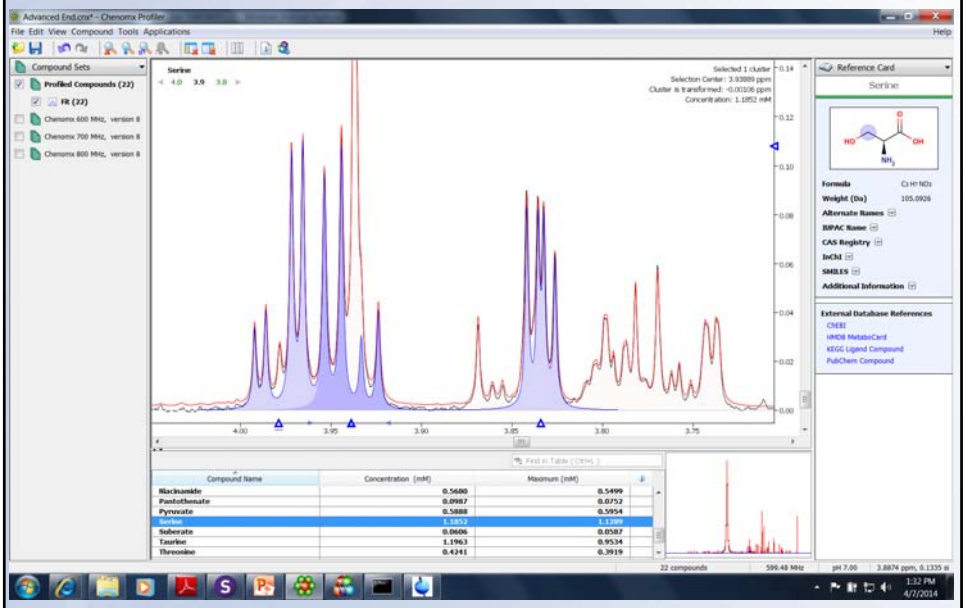
Additive Fit



Chenomx Helps Resolving Ambiguity in Highly Overlapped Regions



Additive fit



Interpretation &
Metabolic Pathway Analysis

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?

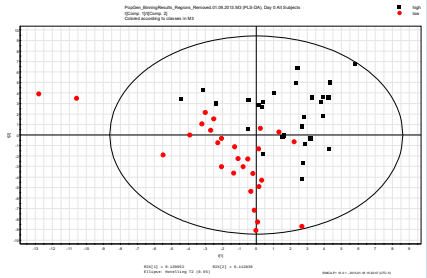
Next step is to interpret results and metabolic pathway analysis

Interpreting results and Pathway Analysis

- There is a number of freely available software
  - Metaboanalyst, MetScape 3 for Cytoscape, metaP-Server, web based KEGG Pathways.
- 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)



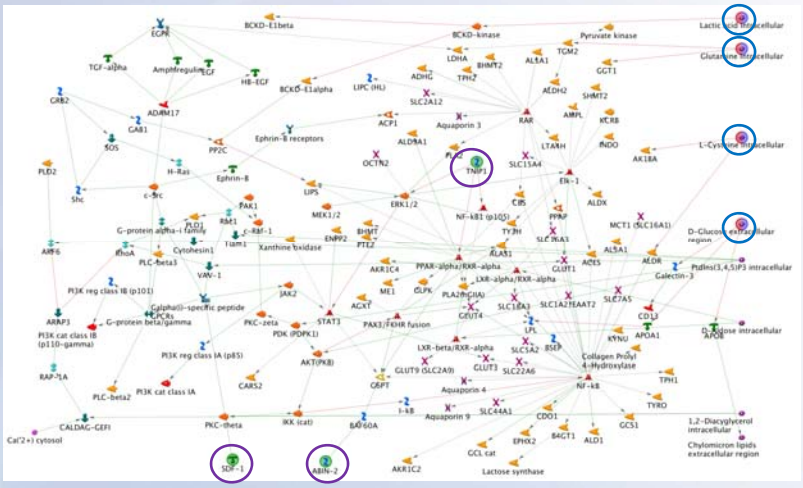
Subset of Metabolites that Influence the Separation of Subjects at Day 0 (VIP ≥ 1 or p-value ≤ 0.1)	
Isoleucine**	Creatinine**
Leucine**	Cysteine**
Valine	Histidine
3-Methyl-2-oxo-isovalerate	Choline
3-Hydroxybutyrate	Glucose
Lactate	Betaine
Alanine	TMAO
Acetate**	Glycine
Proline*	Glycerol
Glutamate**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

*p-value < 0.05, **p-value ≤ 0.1

Preliminary results

Day 0 High vs Low Responders

GeneGo Network Analysis



- 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

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
  - Bayesil (http://bayesil.ca/, includes quantification)
  - BATMAN
  - Use of databases
    - Birmingham Metabolite library, HMDB, BMRB
- Pathway analysis
  - Metaboanalyst, MetScape 3 for Cytoscape, metaP-Server, KEGG

Also available through www.metabolomicsworkbench.org

Some Software available for NMR Based Metabolomics

COMMERCIAL

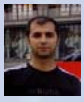
- NMR Data-preprocessing
  - ACD Software (ACD Labs, Toronto, Canada)
  - Chenomx
- Multivariate data analysis
  - SIMCA 13
- Other statistical analysis
  - SAS, SPSS
- Library matching and quantification
  - Chenomx
- Pathway analysis
  - GeneGo (MetaCore Module)
  - Ingenuity Pathway Analysis (IPA)

ERCMRC

Susan Sumner (PI/ Director)

Susan McRitchie
Program Coordinator
Data Analysis

Delisha Stewart
Cell Biology & NMR



Reza Ghanbari
Visiting NIDA Scholar

NUTRITION RESEARCH INSTITUTE

Wimal Pathmasiri
NMR & GC-MS

Yuanyuan Li
LC-MS

Rose Ewald
Intern

Kevin Knagge

Jason Winnike

David Kirchner
LC-MS, Biocrates

DHM | RI

Tim Fennell
Metabolism
LC-MS & GC-MS)

Jessica Gooding
LC-MS

Rod Snyder
LC-MS

Courtney Whitaker
LC/MS

RTI INTERNATIONAL

Maria Moreno
Biology

Keith Levine
Metallomics

Scott Watson
Neurotransmitters