Metabolomics and Metabolic Flux

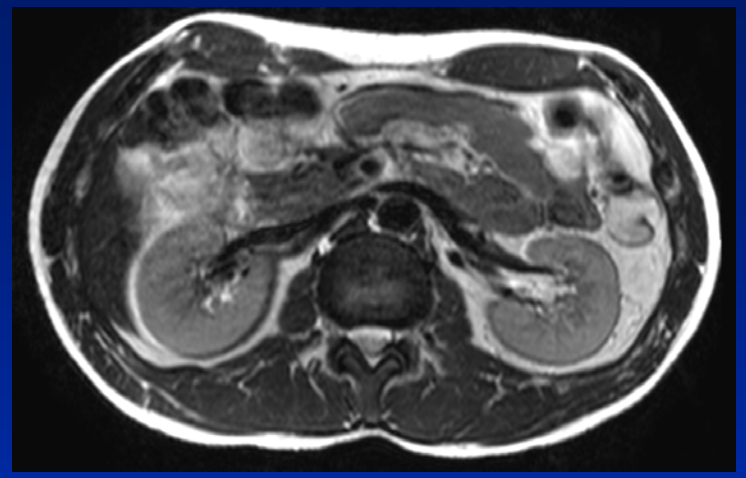
Matthew Merritt

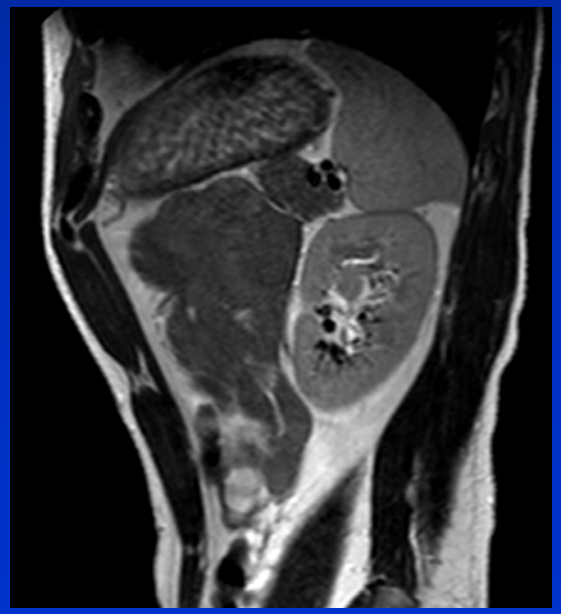
Department of Biochemistry and Molecular Biology

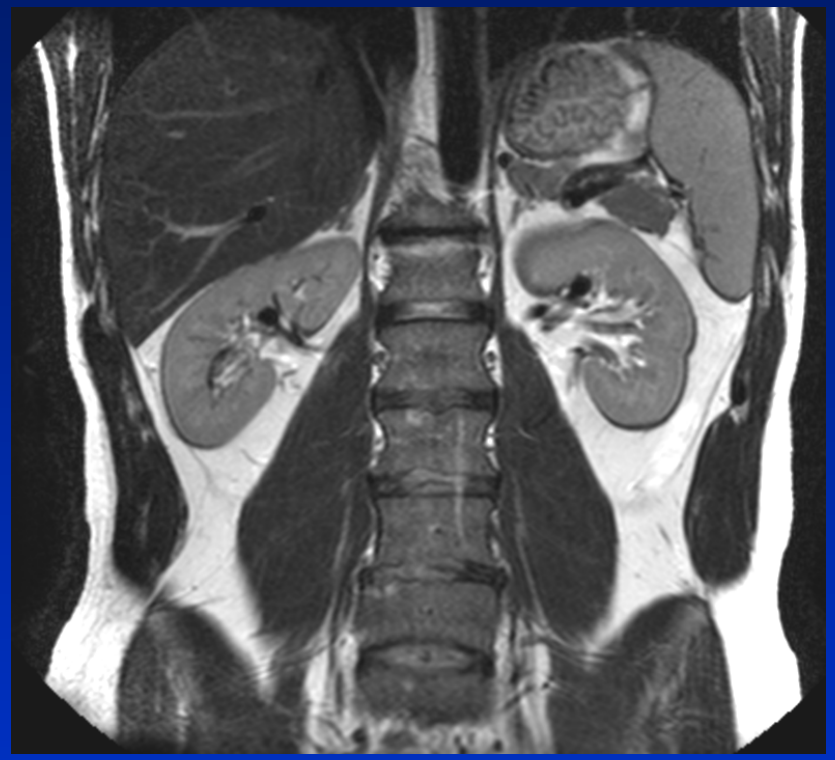
Room R3-226B, Academic Research Building

matthewmerritt@ufl.edu

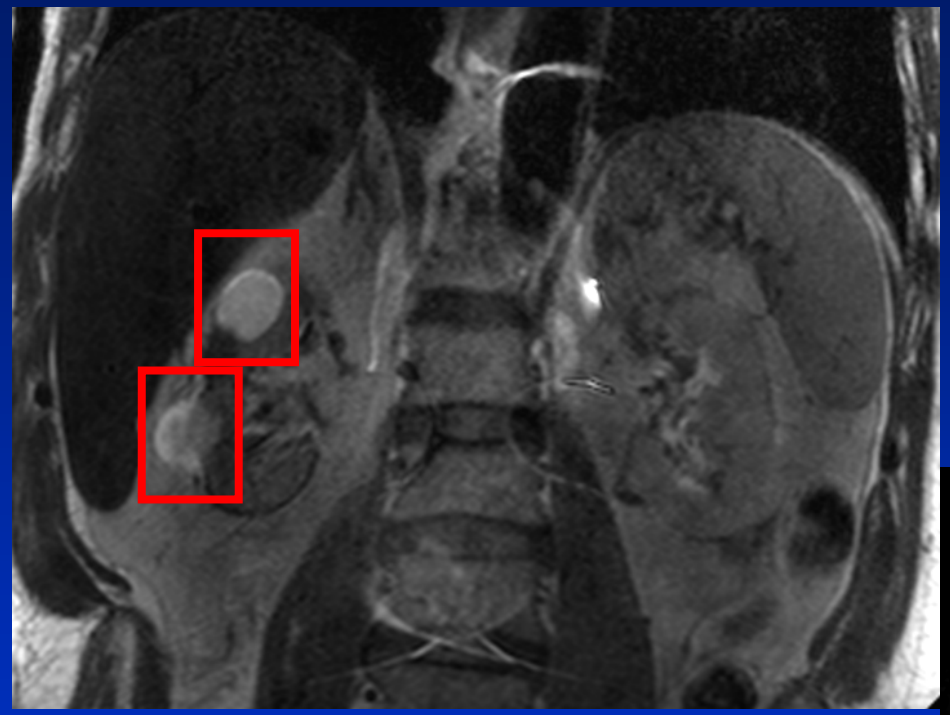
Structure

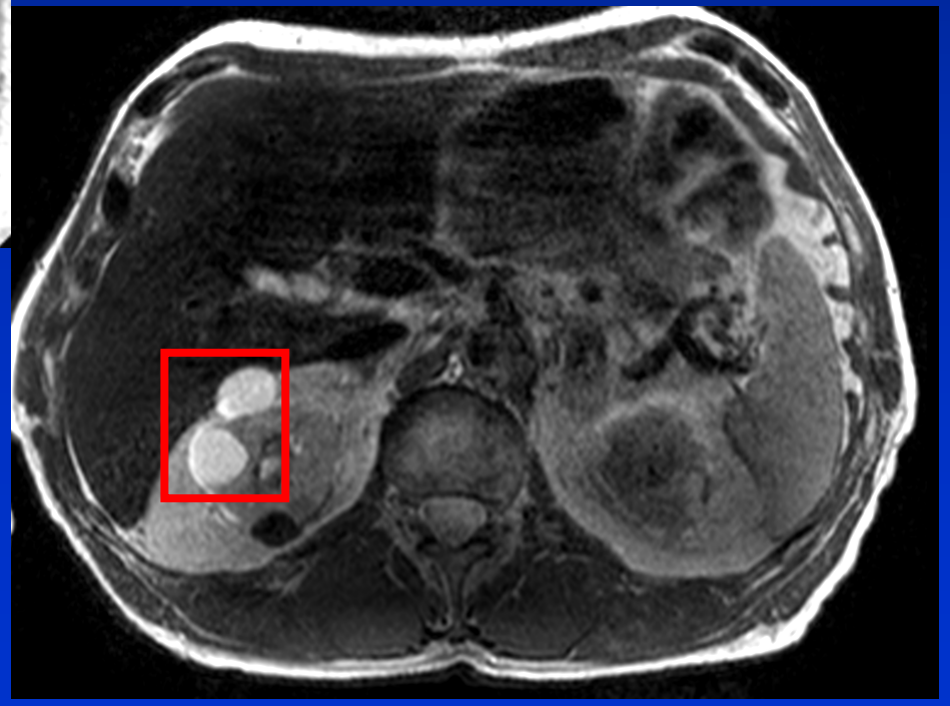






Structure/FUNCTION



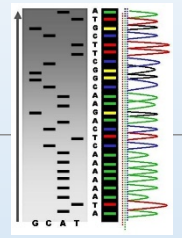


Metabolomics vs Fluxomics

Pool Sizes

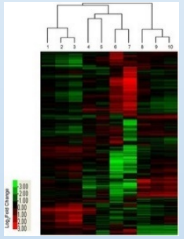
Flux

The "Omics" Revolution



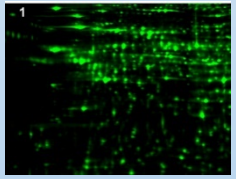
Genomics
~40,000 Genes

DNA



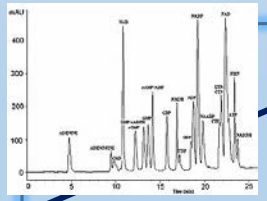
Transcriptomics
~150,000 transcripts

RNA



Proteomics
~1,000,000 proteins

Proteins



Metabolomics
>20,000 metabolites

Metabolites

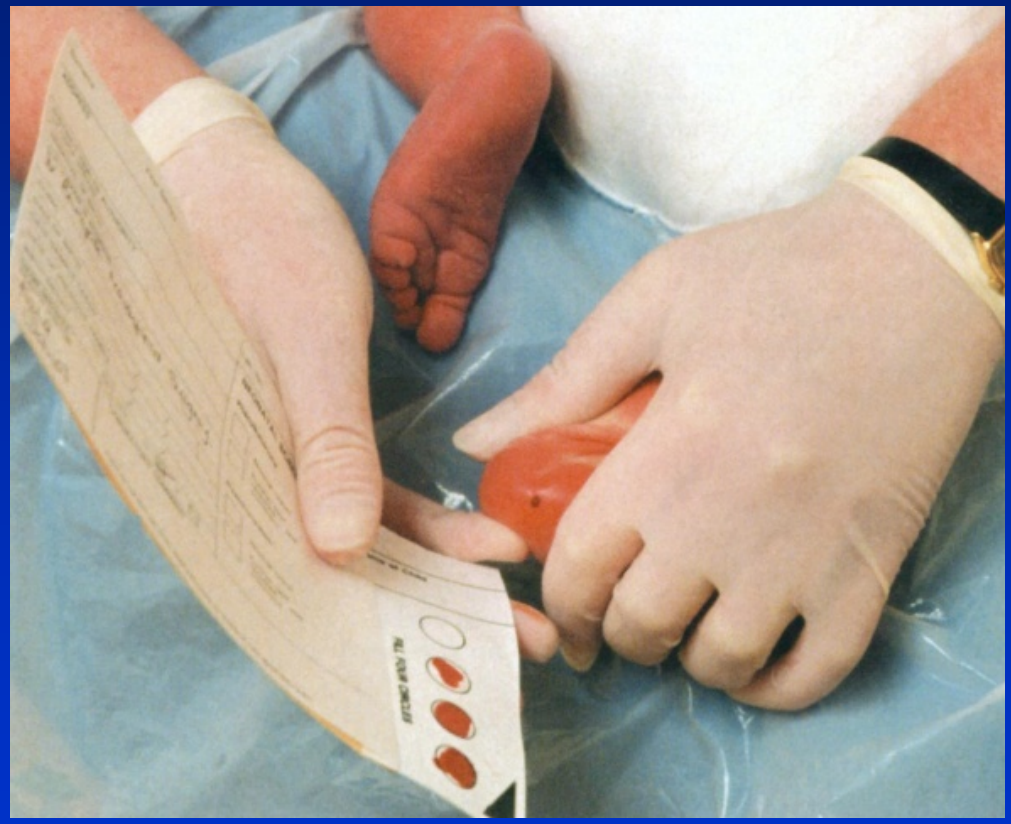
Amino Acids | Sugars | Nucleotides | Lipids

INFORMATION

Davis, V. W.; Bathe, O. F.; et.al. Metabolomics and surgical oncology: Potential role for small molecule biomarkers. *J. Surg. Oncol.* **2011**, *103*, 451-459.
Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Curr. Metabolomics* **2013**, *1*, 92-107.

Sample Collection

- Trivial amount of blood is drawn
- Blood spots on paper are extremely stable!
- This is the PKU test (phenylketonuria) that has been common in US since 1964
- PKU arises when the enzyme catalyzing conversion of phenylalanine to tyrosine is not expressed properly



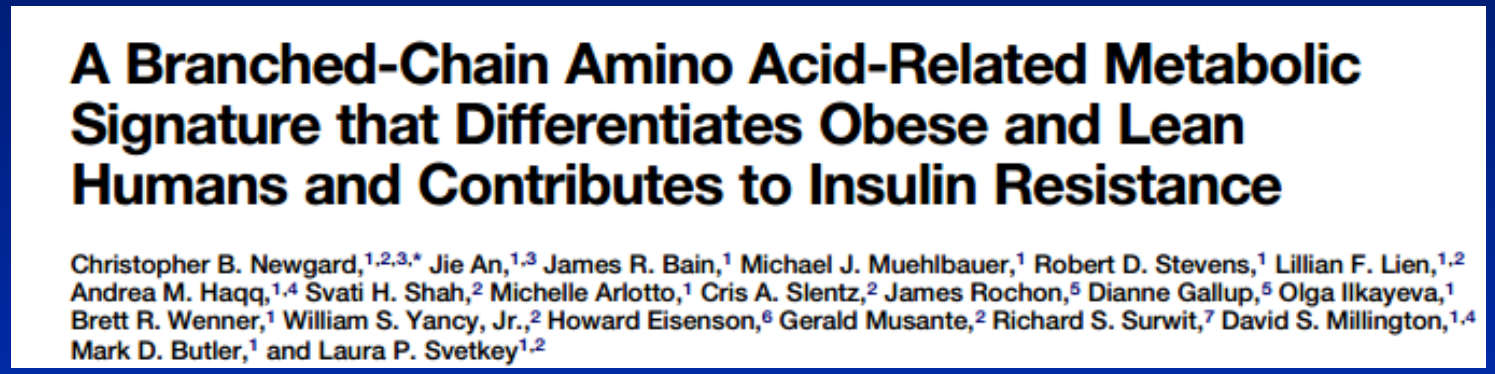
PKU treated versus untreated

- The first success story
  - Both children have PKU
  - One was born before test
  - The girl immediately went to phenylalanine free diet

Metabolomics Definition

- Study of the metabolites present within a living system
- Implication that it is some sort of high throughput screening
  - Want to be able to study a population
  - What is normal? What is pathophysiological?
- The “metabolome” changes every time a new technique is used or the sensitivity increases
  - Started out with ~200 compounds
  - Now the number is ~20000, but really could be significantly larger if lipids are considered

A Metabolic Survey that Changes How We Think About Diabetes



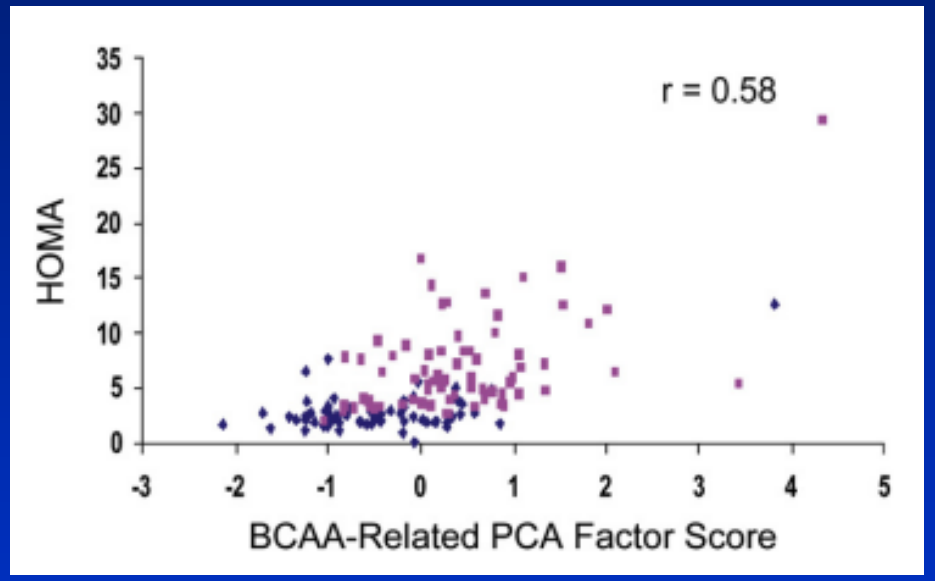
A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance

Christopher B. Newgard,[1,2,3,*] Jie An,[1,3] James R. Bain,[1] Michael J. Muehlbauer,[1] Robert D. Stevens,[1] Lillian F. Lien,[1,2] Andrea M. Haqq,[1,4] Svati H. Shah,[2] Michelle Arlotto,[1] Cris A. Slentz,[2] James Rochon,[5] Dianne Gallup,[5] Olga Ilkayeva,[1] Brett R. Wenner,[1] William S. Yancy, Jr.,[2] Howard Eisenson,[6] Gerald Musante,[2] Richard S. Surwit,[7] David S. Millington,[1,4] Mark D. Butler,[1] and Laura P. Svetkey[1,2]

- Comprehensive metabolic profile on 74 obese and 67 lean subjects
- This is a survey: what can we see?

A Metabolic Survey that Changes How We Think About Diabetes

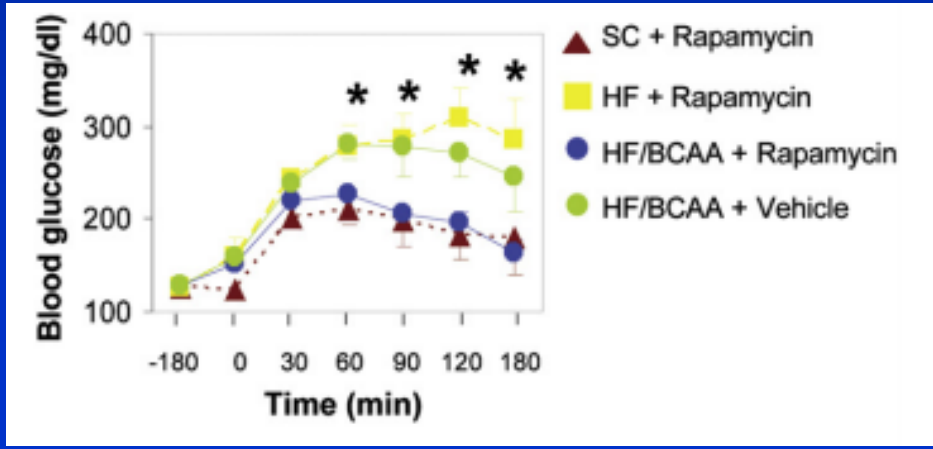
- HOMA is a measure of insulin sensitivity
  - A lower score means you are insulin sensitive (lowers blood glucose)
- Principal components analysis combines multiple variables into a single quantity that captures the maximum variance

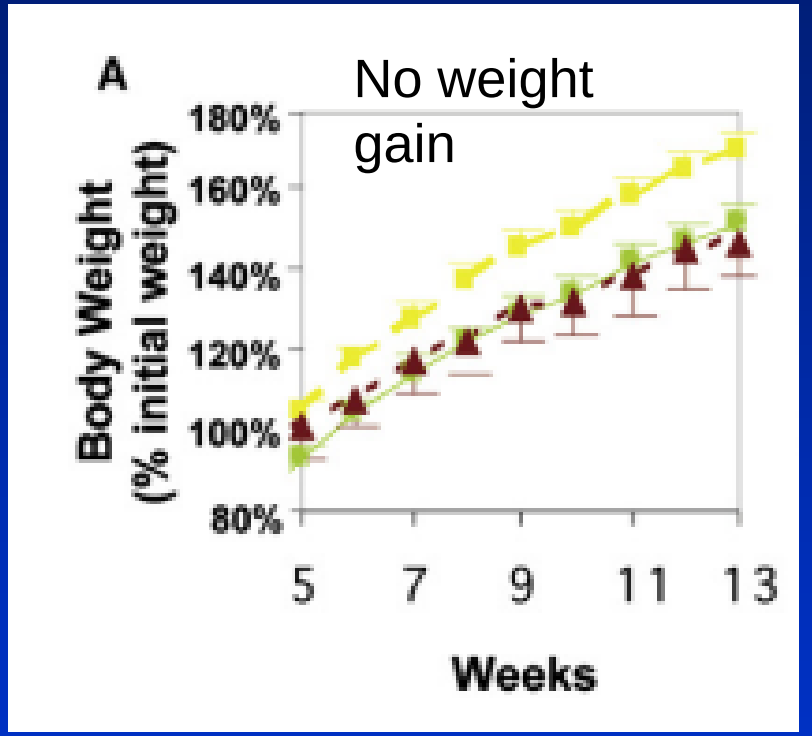


Note the scatter in the obese subjects; BCAAs (valine, leucine, isoleucine) were primary members of the principal component

Human Observations Guided a Series of Rat Based Studies

- Control, high fat, and BCAA supplemented diets
- BCAAs cause insulin

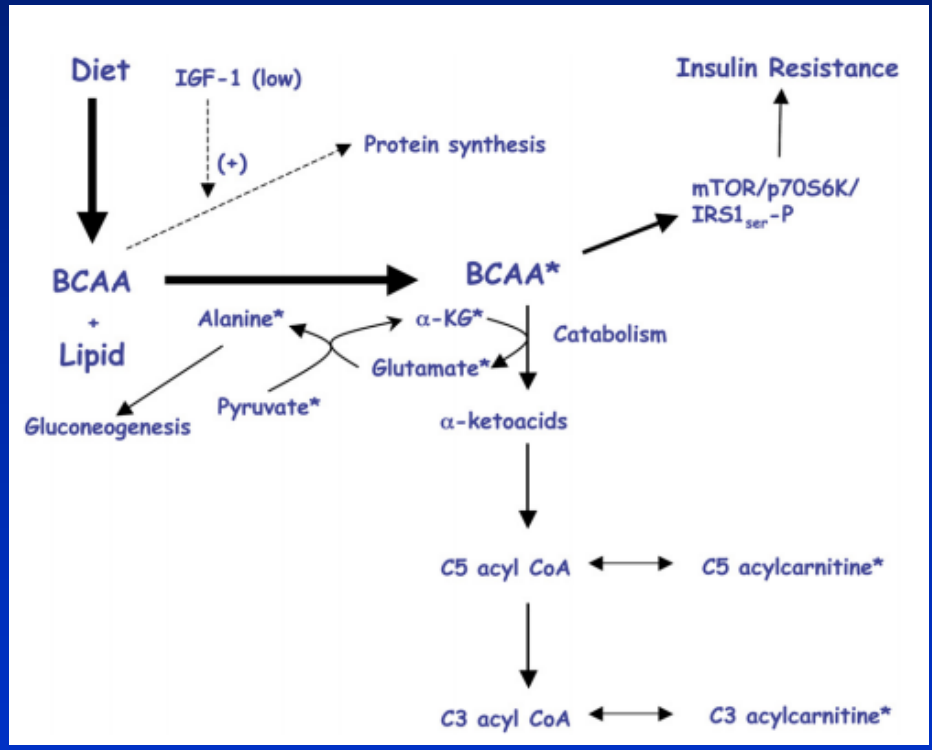




Rapamycin inhibits the putative target of the BCAA activation

A Metabolic Survey that Changes How We Think About Diabetes

- Global profiling of humans led to an observation that had not even been suggested before
- Led to experiments with testable hypothesis
- LC/MS can be used in a large cohort of people



Targets for Study

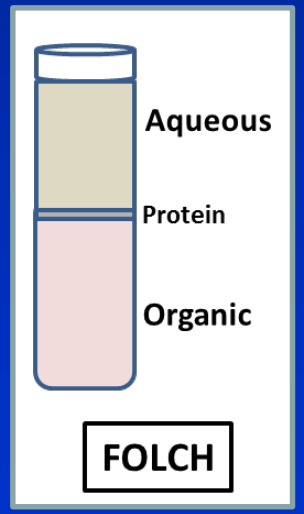
- Collect samples correctly
  - Urine
  - Blood
  - Cerebrospinal Fluid
  - Tissue

Handling Urine Samples

- Remove particulates with centrifugation
  - protein can be eliminated in urine
- pH can vary- for Nuclear Magnetic Resonance this can change peak locations and cause unnecessary scatter in the data
- Salts can vary- big issue for LC/MS (ionization efficiency)
- If stored for any significant period it must be in a -80 ^{o}C fridge
  - stop metabolism

Handling Blood Samples

- Remove red blood cells by centrifugation
  - RBCs are highly metabolic
- Decision- which metabolites am I interested in?
  - Water soluble
  - Organic soluble
- Folch extraction – Chloroform/Methanol
  - Lipids in organic phase
  - Water soluble in methanol
- Perchloric acid – Water soluble only



Blood derived samples

- Plasma- blood without the RBCs
- Serum- plasma that has removed clotting factors (the protein fibrinogen primarily)

Extraction Artifacts

- Chloroform/methanol extraction does not capture water soluble compounds as effectively as perchloric acid
- Even across amino acids, perchloric acid extracts capture various amounts

Prior to any instrumental issues, sample handling plays a large, essential, crucial part in the metabolic profile measured

Tissue Samples

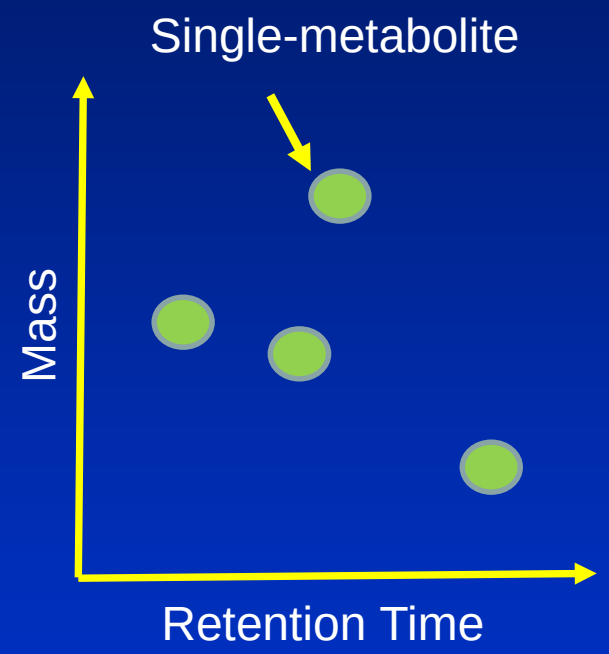
- By definition there will be fewer samples
- Immediately arrest metabolism
  - Drop in liquid N_2
  - Preferably Freeze Clamp
  - Store at -80 ^{o}C at least, preferably at LN_2 temperature
- Your choice of extraction methods

Choose a Research Platform

- Liquid chromatography –Mass spectrometry
- NMR

Increased Power of Hyphenated Methods

- LC separates on polarity (solubility between phases) of the analytes
- MS measures the mass/charge
- The combination provides the tremendous resolution needed to sort out ~20000 compounds



LC/MS vs NMR

- NMR depends upon sensitivity and chemical shift resolution (limits to ~100 compounds)
  - Higher field
  - Cryoprobes
  - Quantitative in most circumstances
- LC/MS
  - Very sensitive
  - Matrix effects are strong
  - Must add an isotopic standard to quantitate

Measuring Metabolism with Nuclear Magnetic Resonance (NMR)

Magnetic Resonance

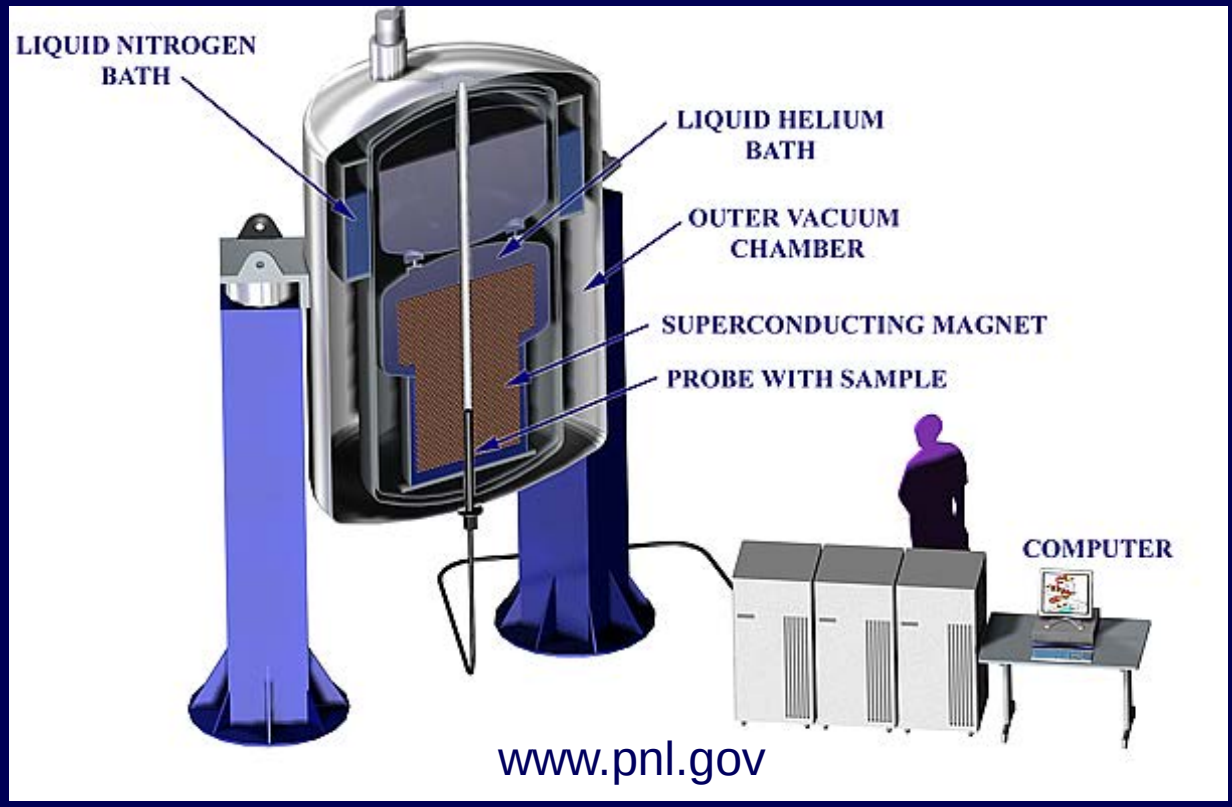
Human Imager (3T)

High Resolution NMR (14T)

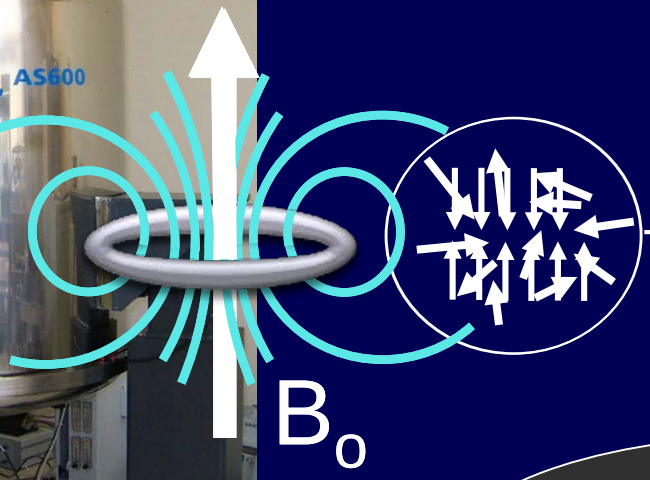
Why Magnetic Resonance?

1) Nuclei detection is frequency dependent
2) Chemical shift
3) Spin-spin coupling
4) Potential for *in vivo* studies

The NMR Spectrometer



Nuclear Magnetic Resonance
The Basics



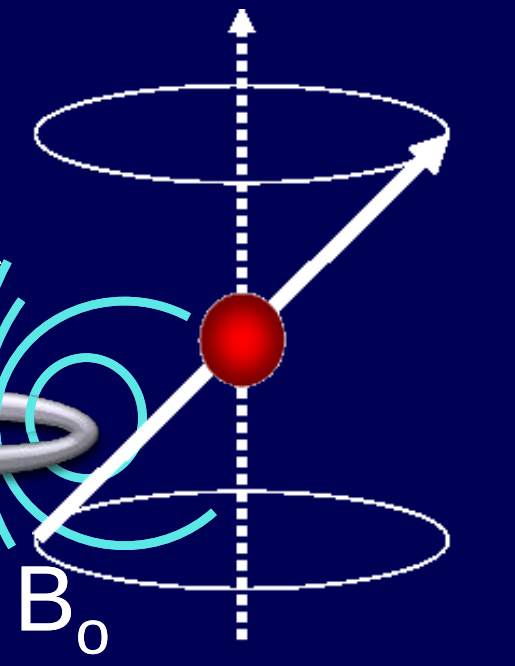
Nucleus	Spin	Natural abundance	Rel . NMR sensitivity
^{1}H	1/2	99.9%	100
^{31}P	1/2	100%	16
^{13}C	1/2	1.1%	2
^{2}H	1	0.015%	10
^{15}N	1/2	0.34%	0.2

$$\frac{\alpha}{\beta} = e^{-\Delta E/kT} = 1 - \gamma h \mathbf{B}_o / kT$$

Only 1 Carbon-13 in 100,000 is visible

^{13}C 1×10^{-5}

Nuclear Magnetic Resonance
The Basics

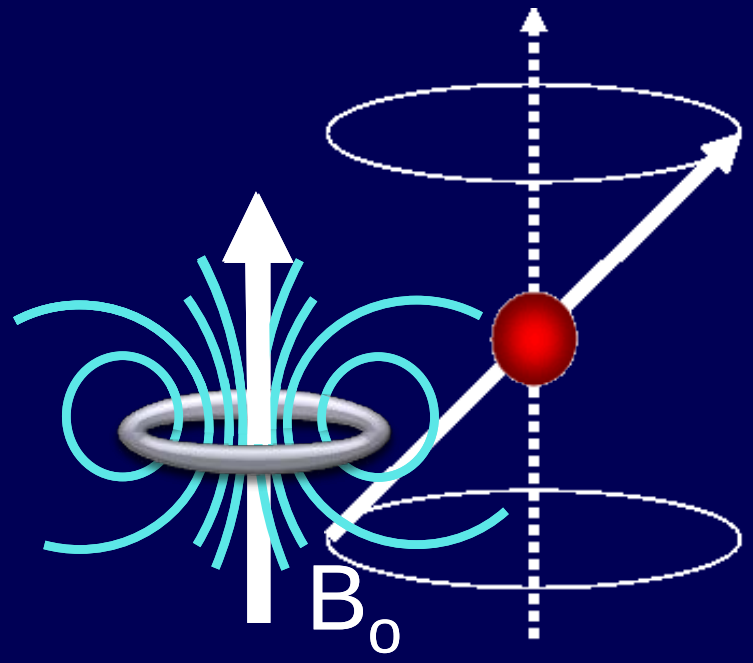


Resonance Frequency

Larmor Frequency of Common Nuclei at 14T

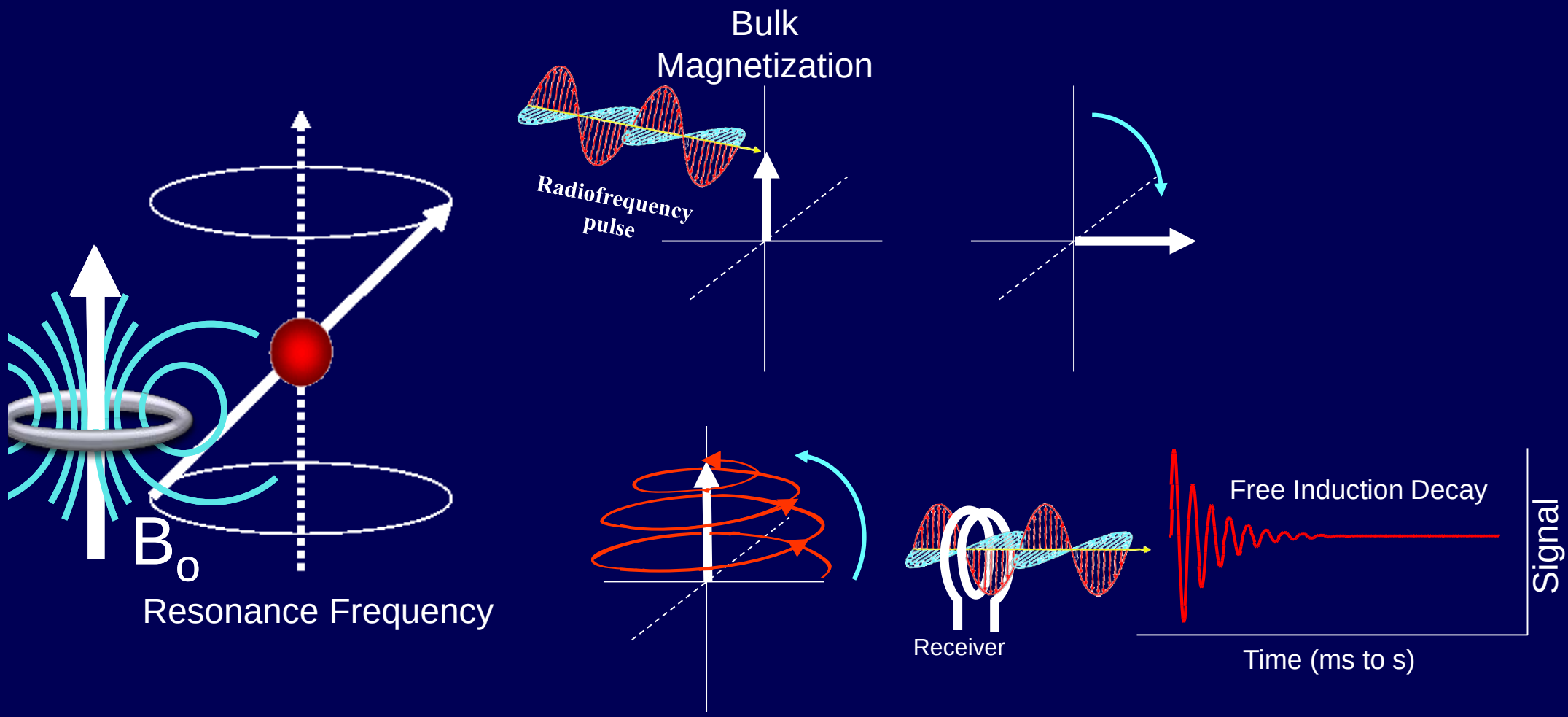
Nucleus	Frequency
^{1}H	600 MHz
^{31}P	243 MHz
^{13}C	151 MHz
^{2}H	92 MHz
^{15}N	61 MHz

Nuclear Magnetic Resonance
The Basics

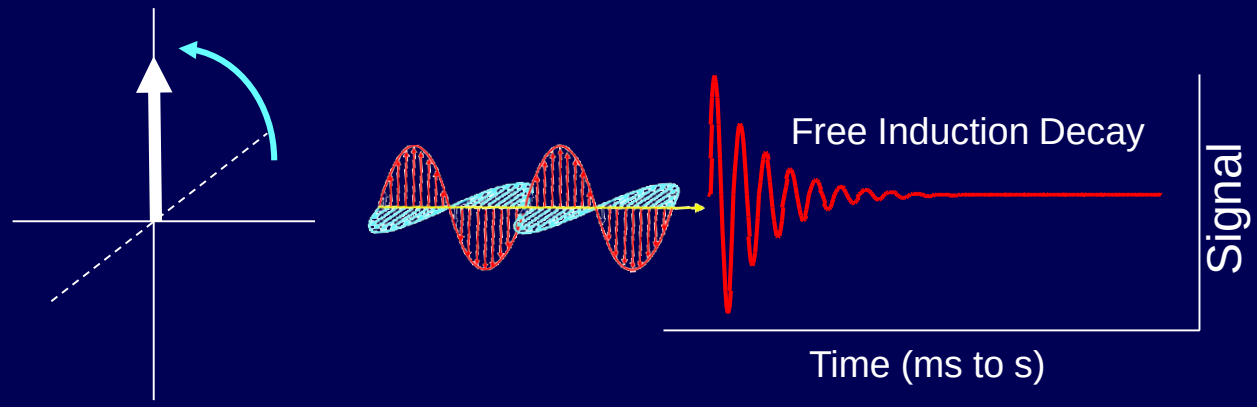


Resonance Frequency

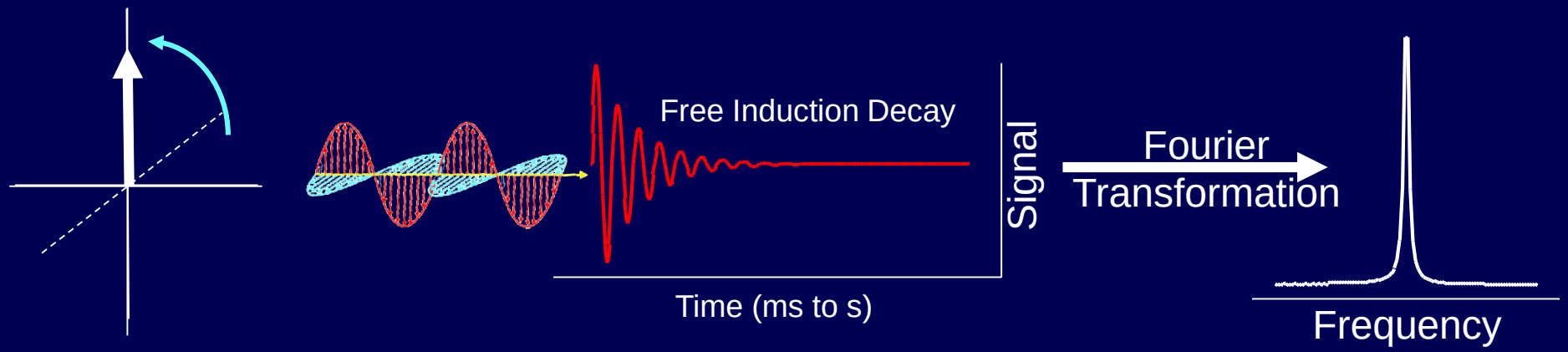
Nuclear Magnetic Resonance
The Basics

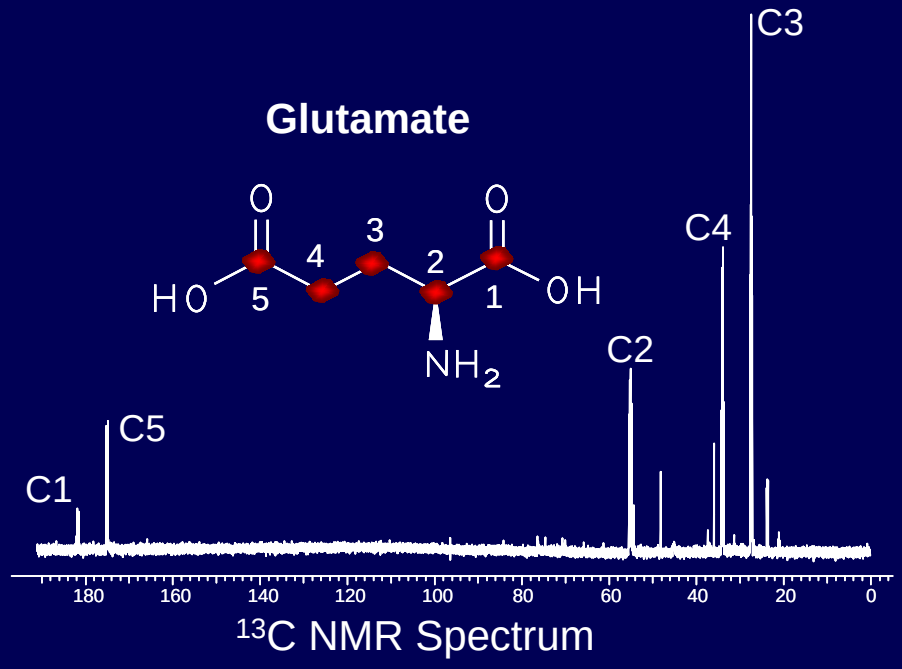


Nuclear Magnetic Resonance
The Basics



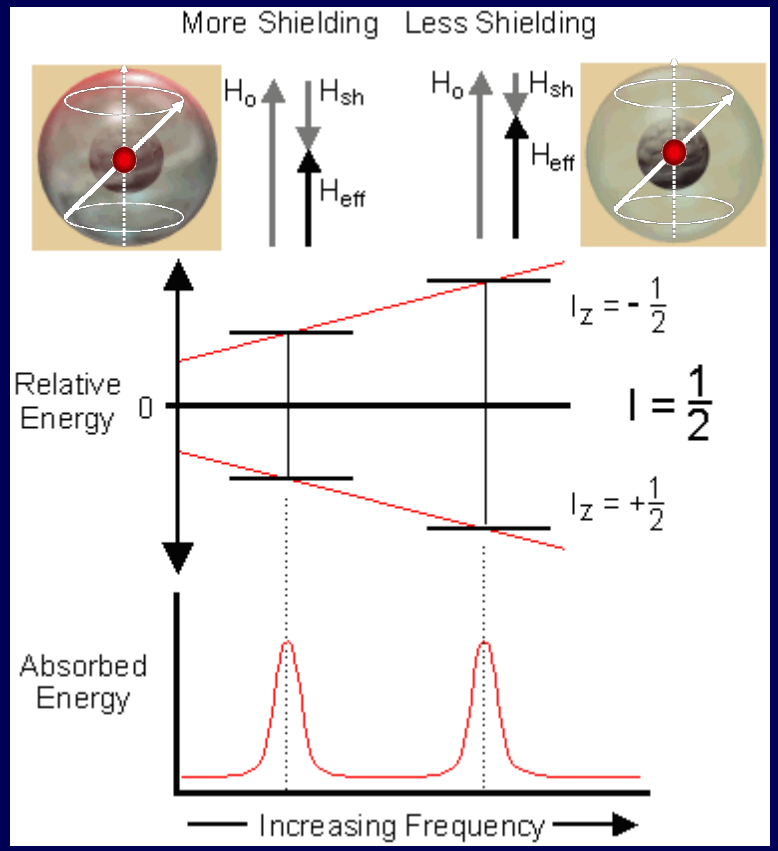
Nuclear Magnetic Resonance
The Basics



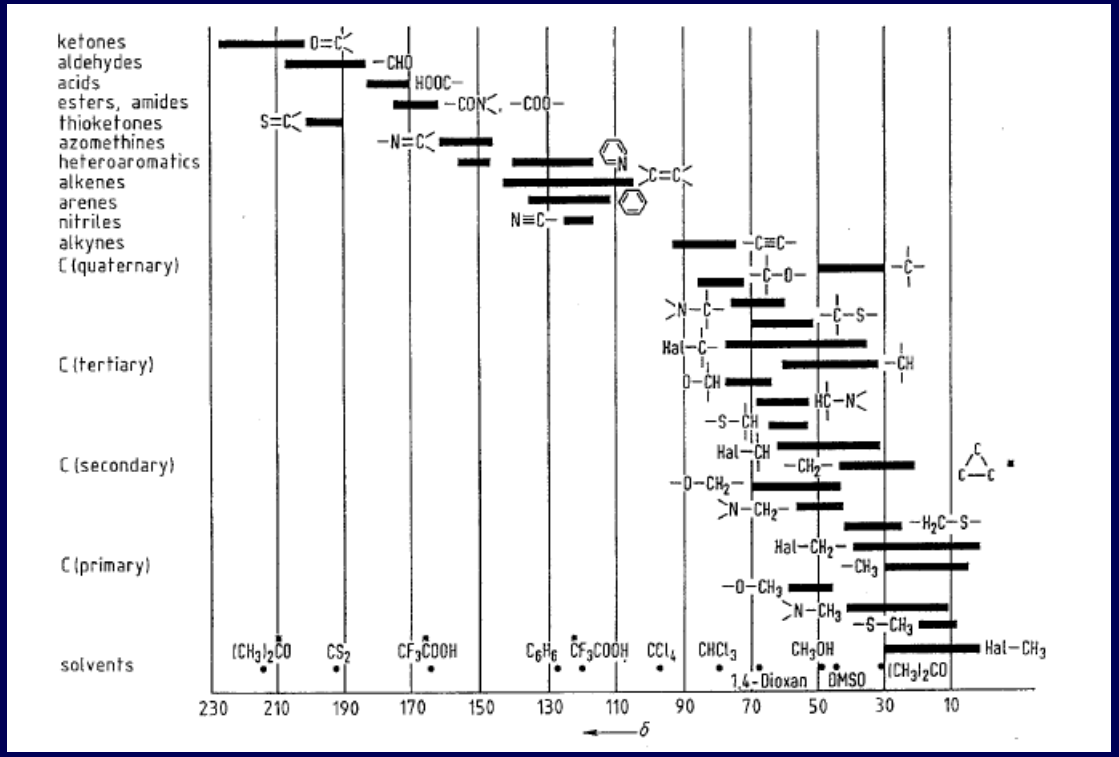


How is NMR able to resolve the individual hydrogens and carbons of metabolites?

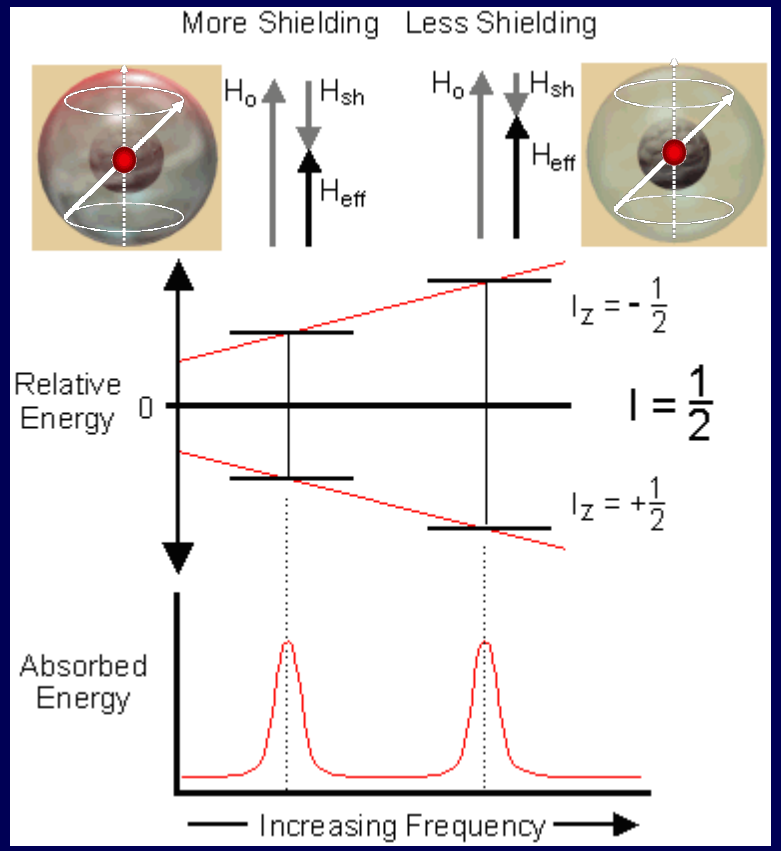
The Resonance of an NMR Signal Depends on The Chemical Environment of the Nuclei



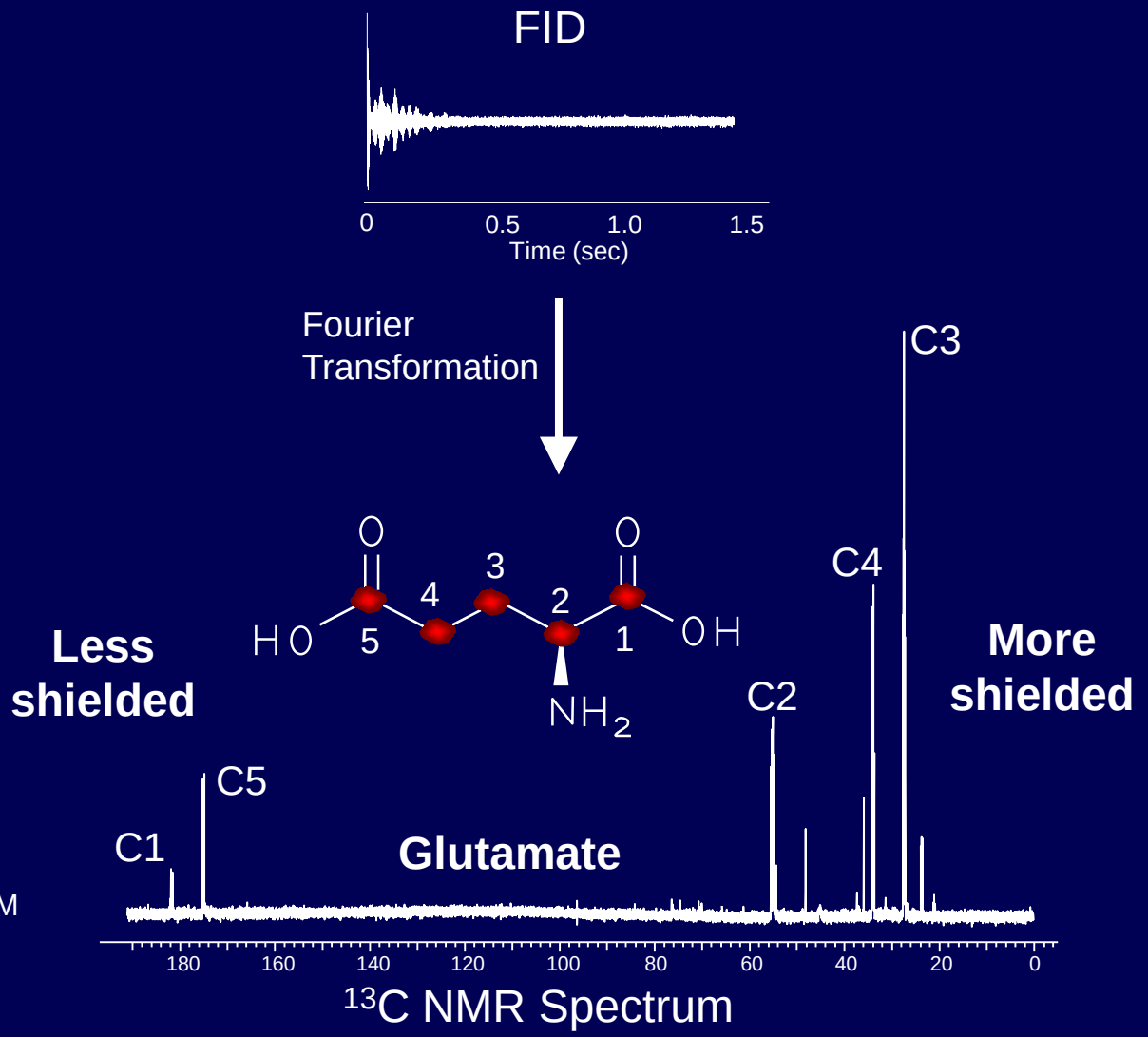
http://www.chembio.uoguelph.ca/driguana/NMR/SHIELD1.HTM



The Resonance of an NMR Signal Depends on The Chemical Environment of the Nuclei



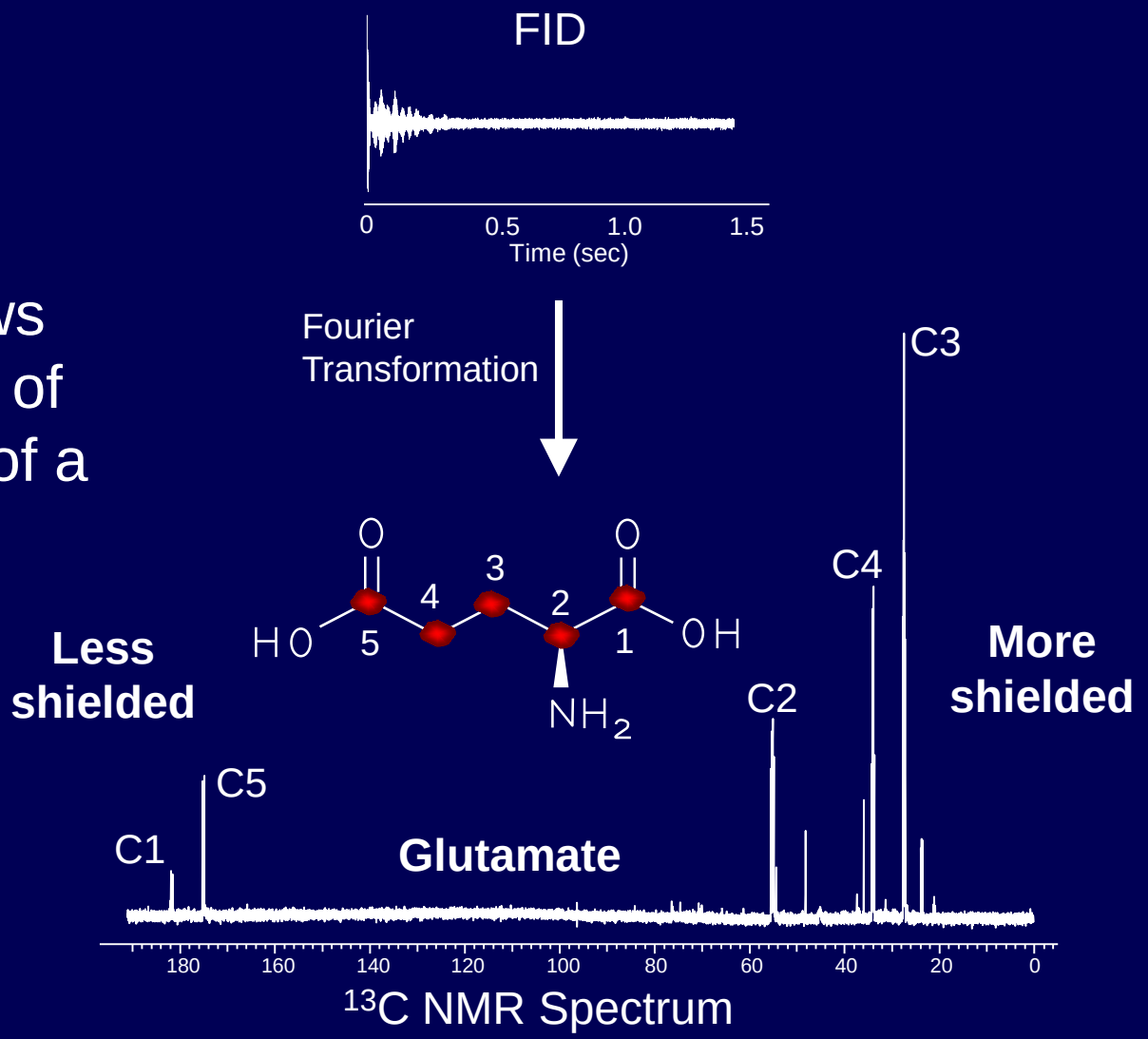
http://www.chembio.uoguelph.ca/driguana/NMR/SHIELD1.HTM



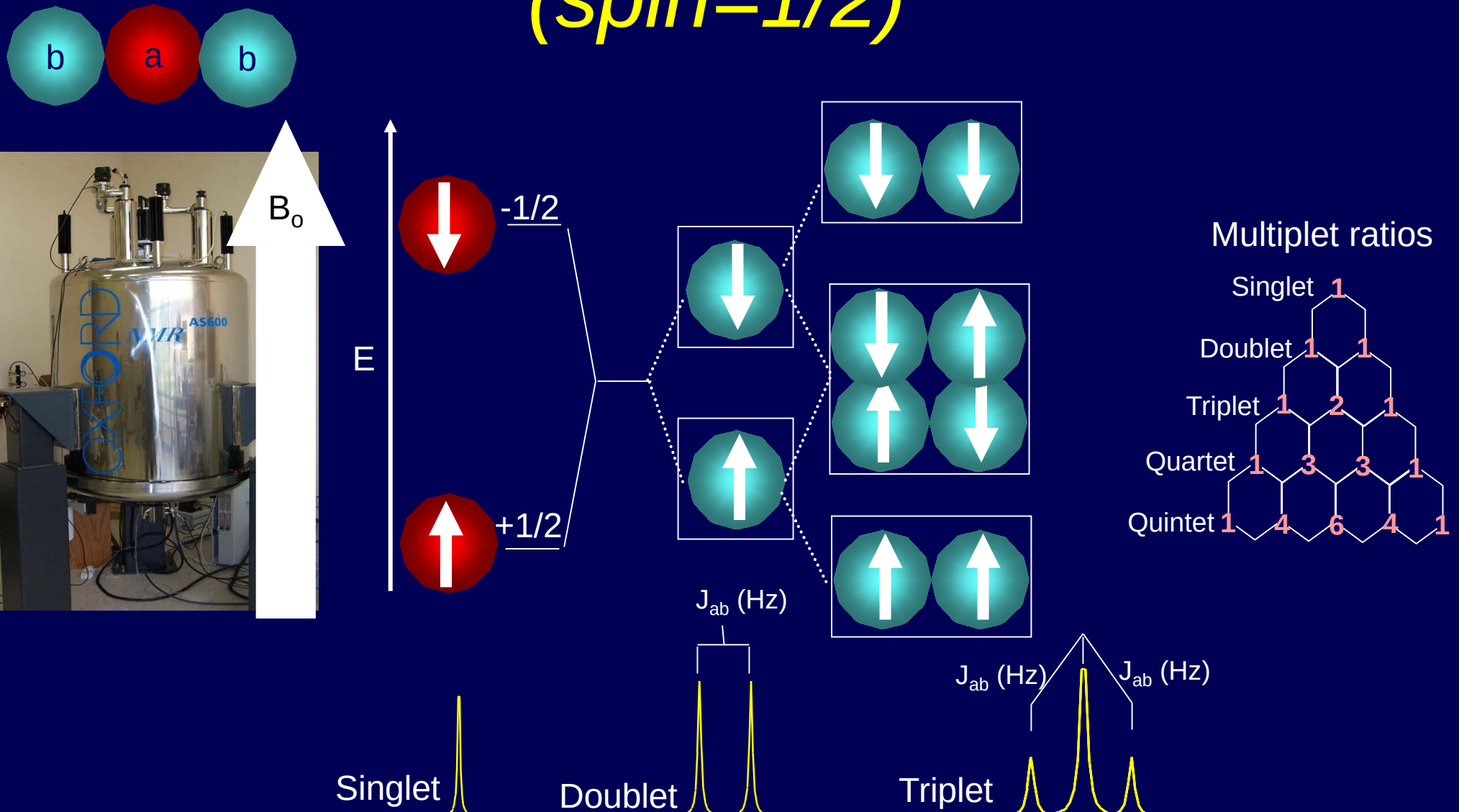
The Resonance of an NMR Signal Depends on The Chemical Environment of the Nuclei

Key Point:

NMR Chemical Shift allows simultaneous assessment of enrichment in all positions of a metabolite



Spin Coupling (j-coupling)
(spin=1/2)

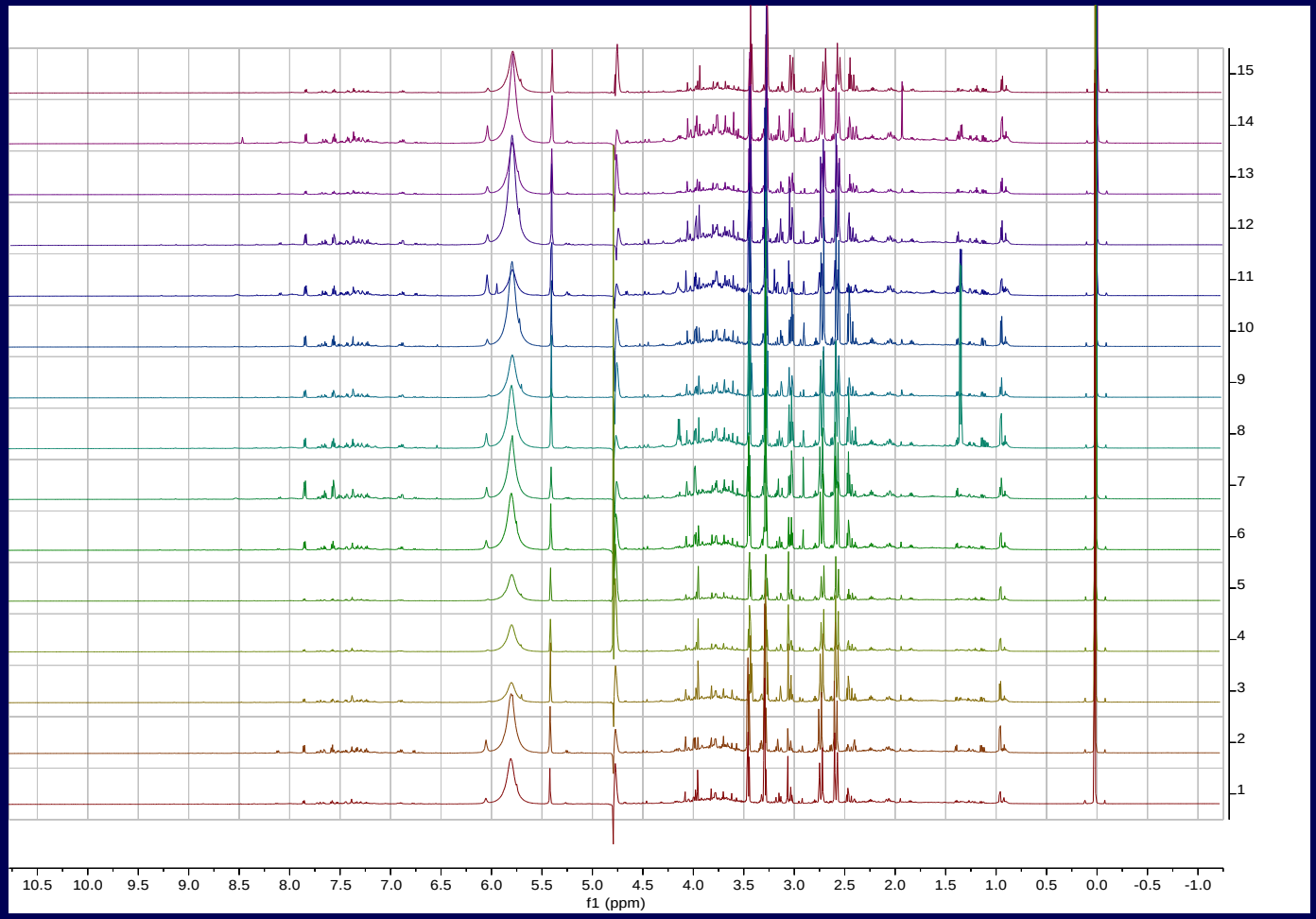


Multiplet ratios

Singlet	1				
Doublet	1	1			
Triplet	1	2	1		
Quartet	1	3	3	1	
Quintet	1	4	6	4	1

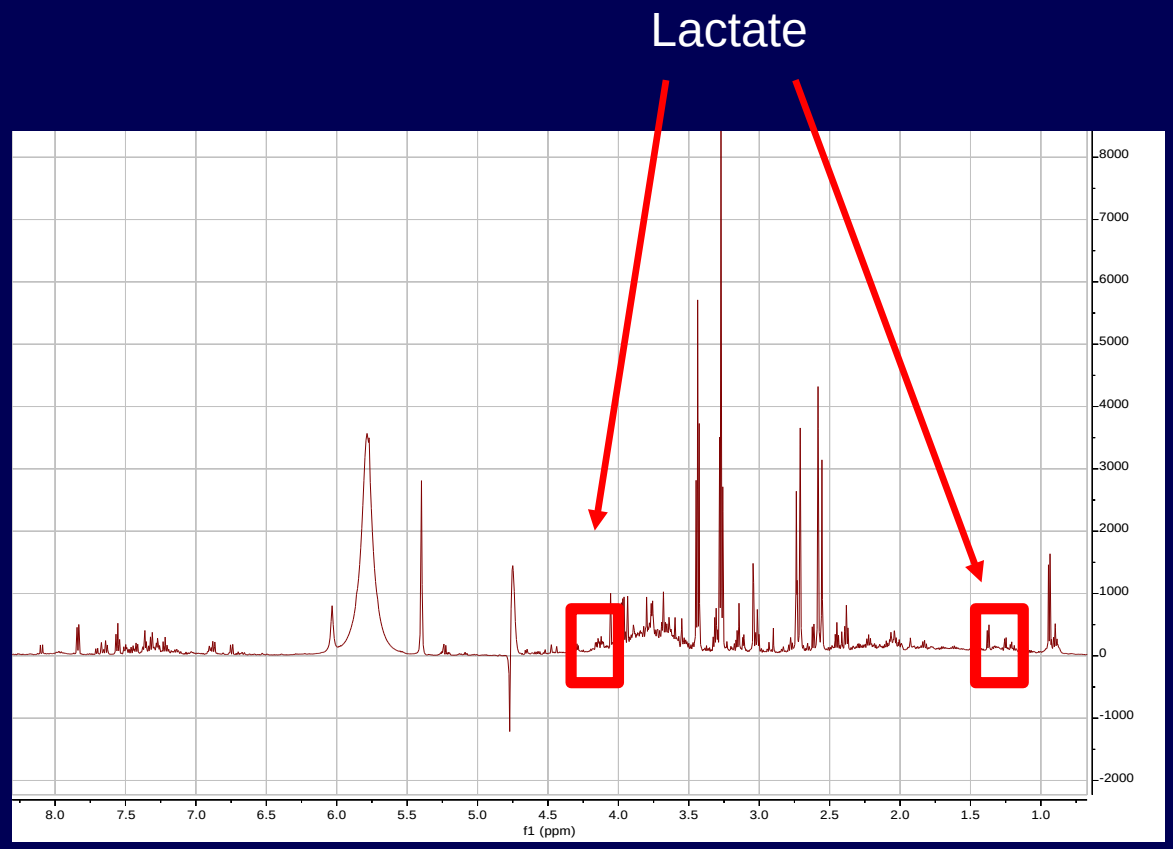
Hydrogens have the largest magnetic moment of possible nuclei and a high natural abundance

- ^{1}H NMR meets requisites for metabolomics technique
- Sensitivity is less than LC/MS methods



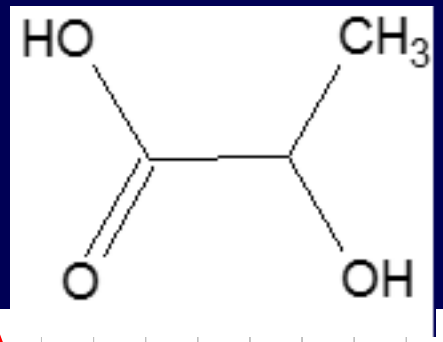
^{1}H NMR Spectra

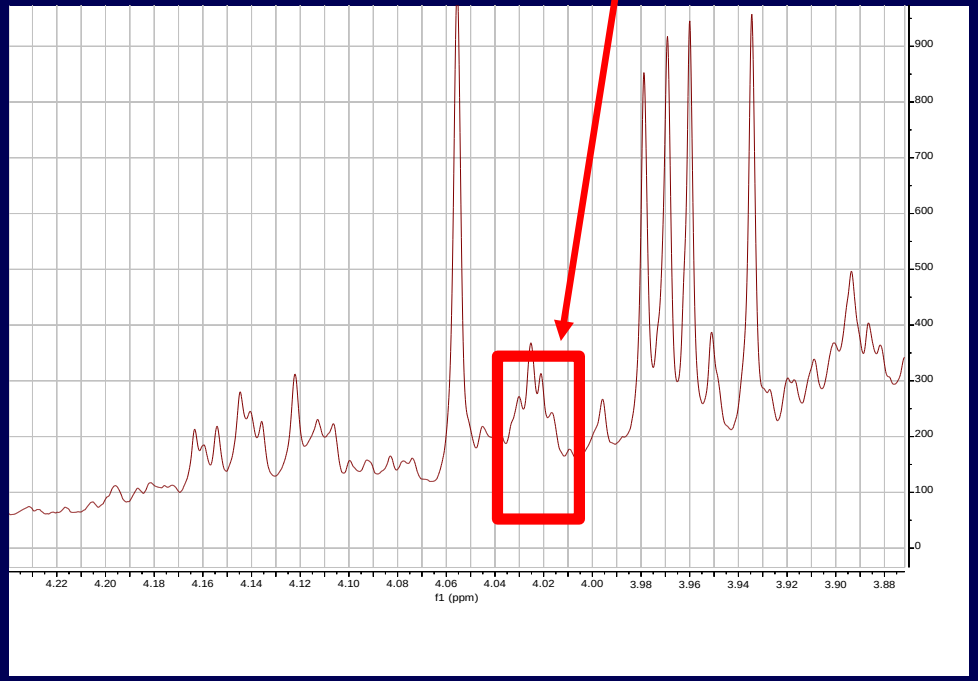
- Chemical shifts and j-couplings lead to complicated spectra
- Many resonances are correlated

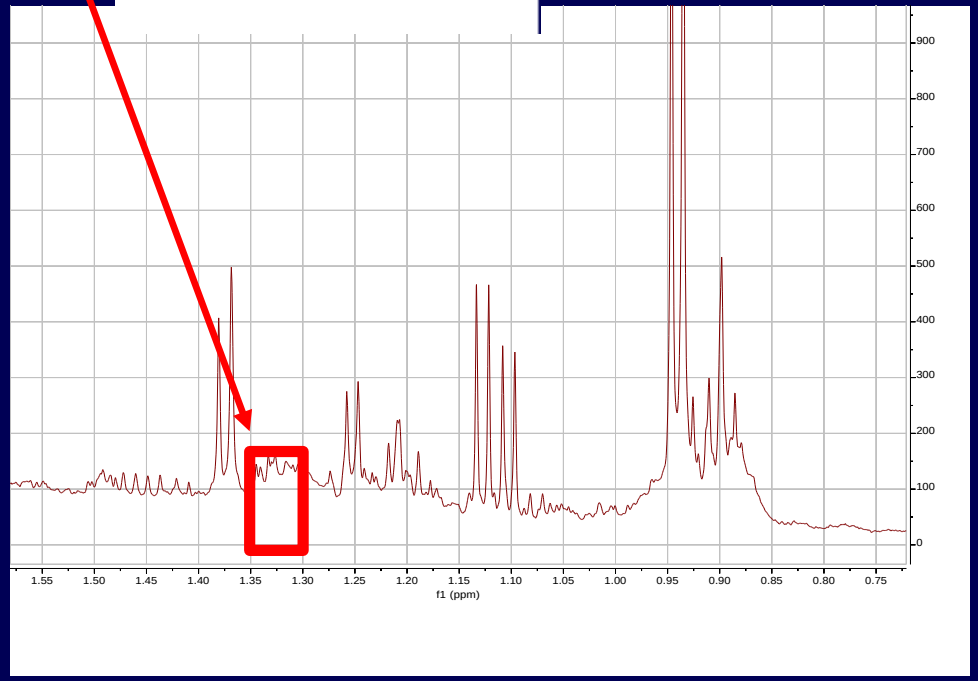


Single Molecule has Multiple Resonances

Lactate

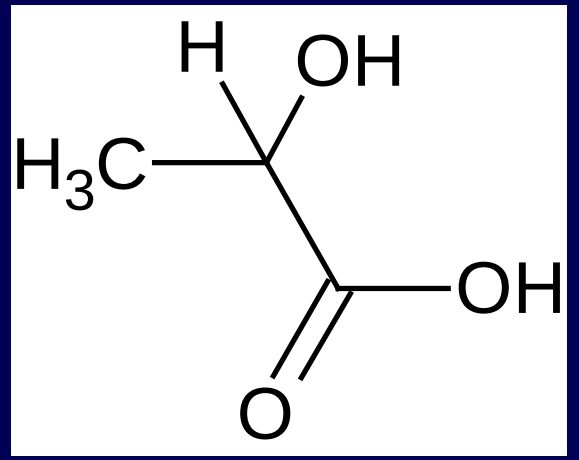




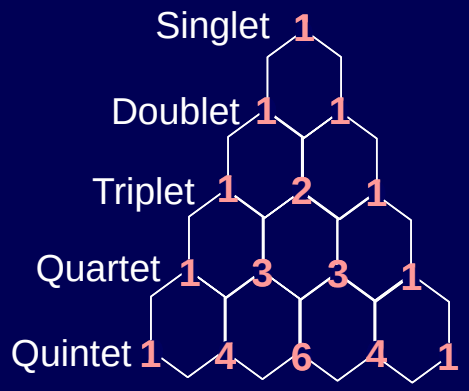


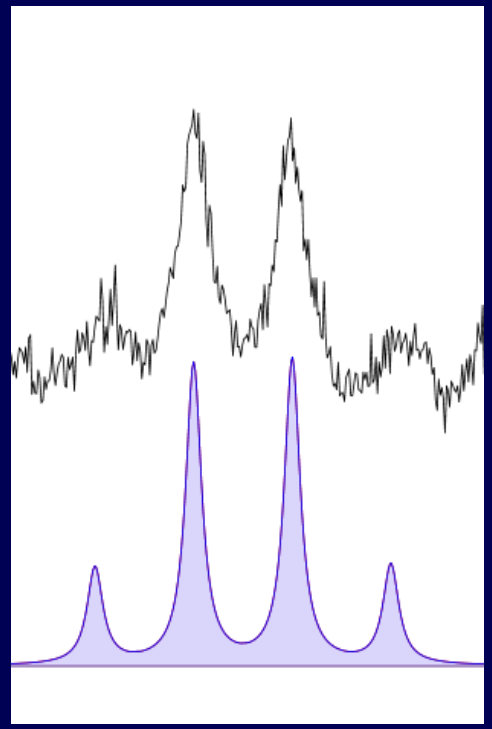
Origin of Multiple Signals/Resonance Frequency

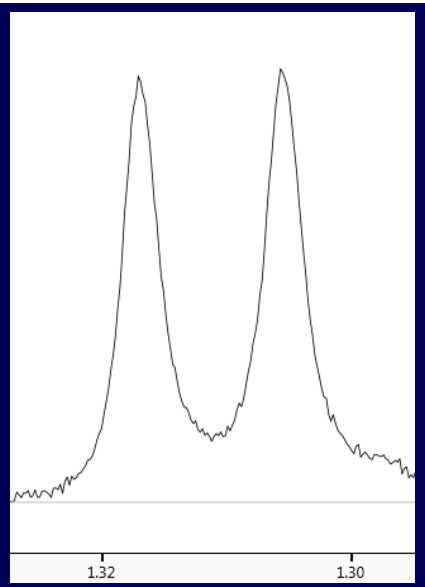
- Note the methyl group has a single hydrogen neighbor (n+1)
- The single hydrogen has three neighbors (n+1)



Multiplet ratios





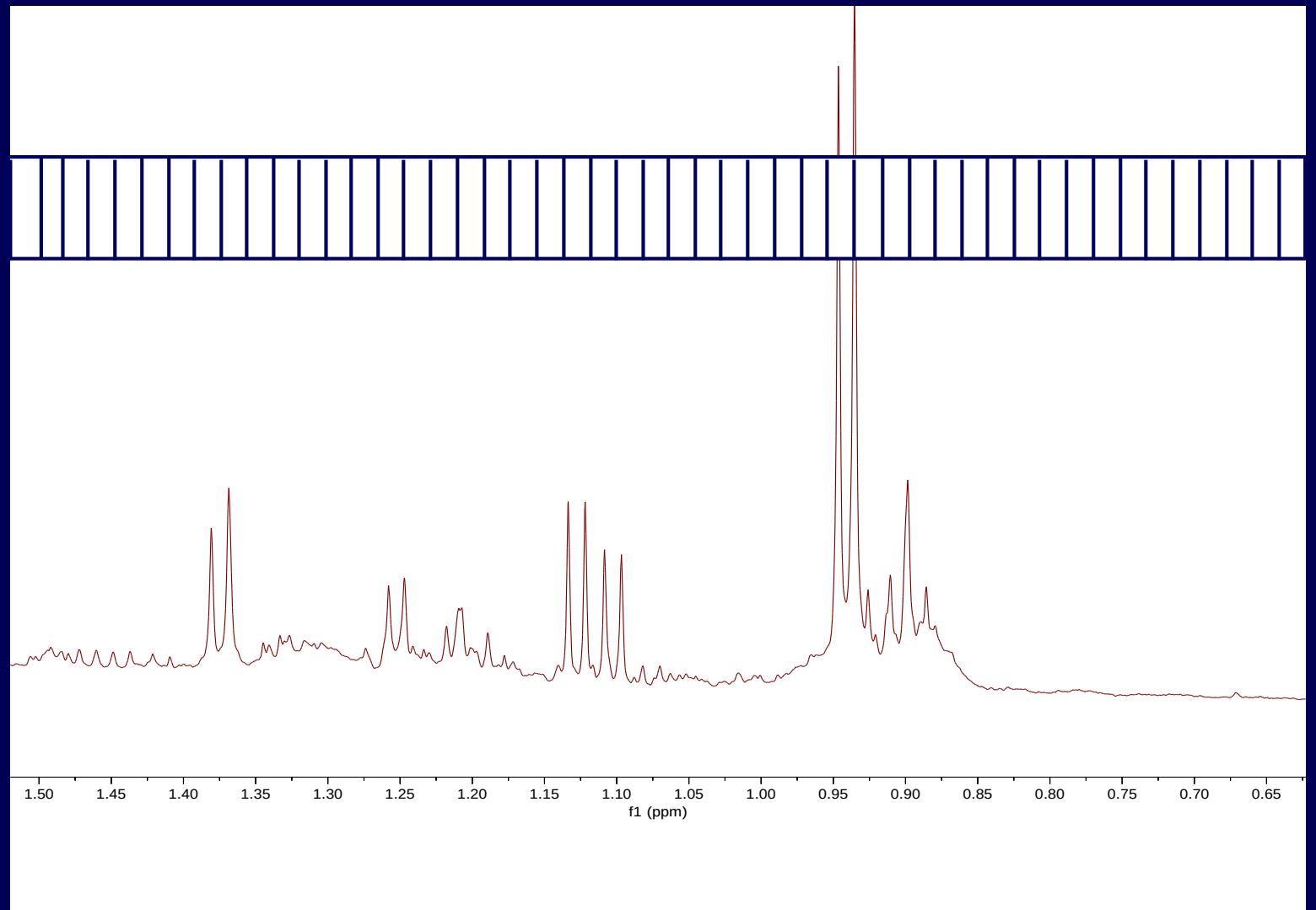


Assignment Issues

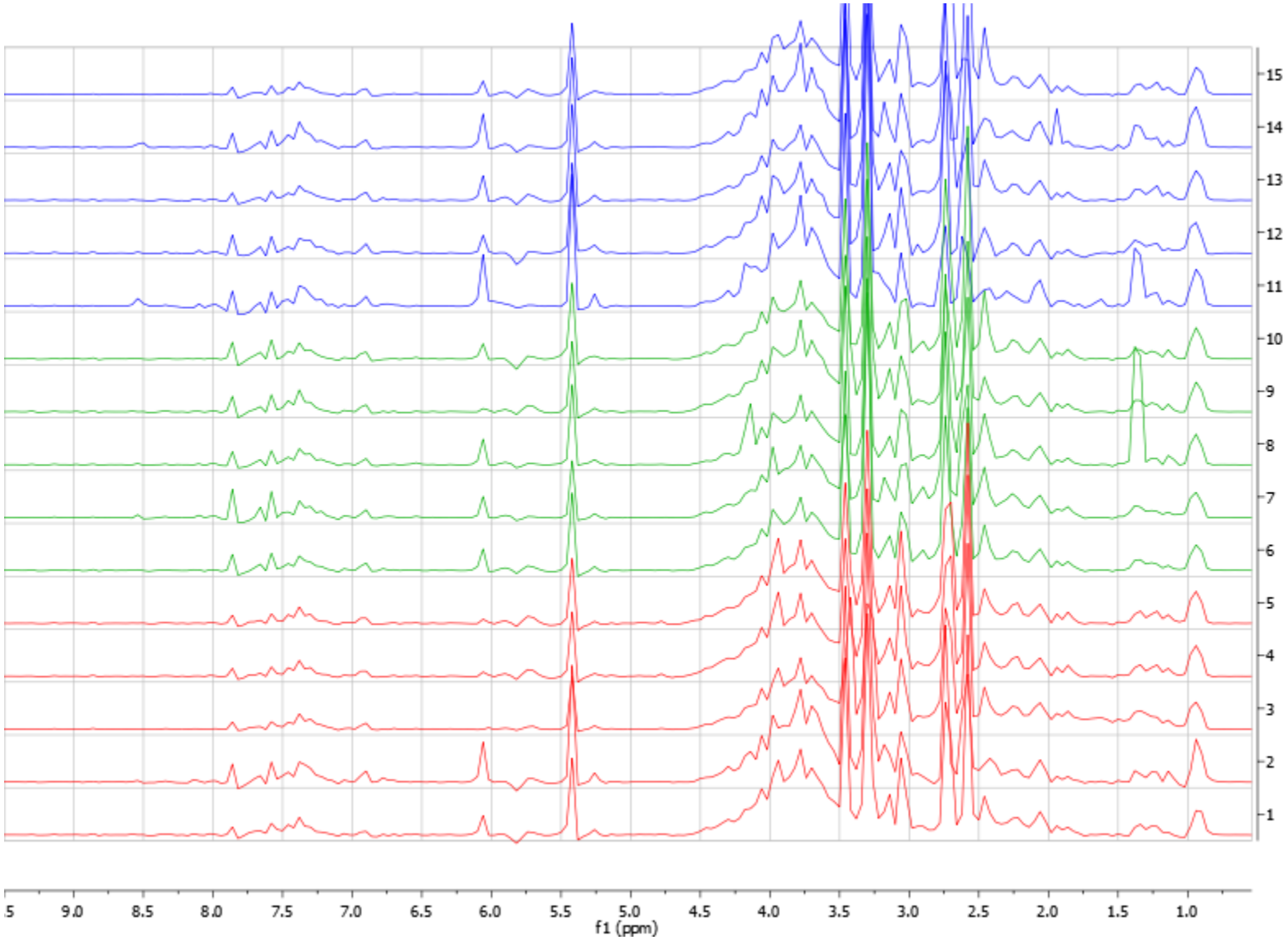
- Multiple overlapping resonances leads to difficulties in compound assignment
- Most straightforward approach is to quantify the individual metabolites
- For screening, the practice of binning was introduced

Binning

- Add up amplitude across the binned areas to produce reduced spectra



Series of Binned ^{1}H NMR Spectra



Data Reduction

Not your grandmother’s statistics

Many More Variables Than Samples

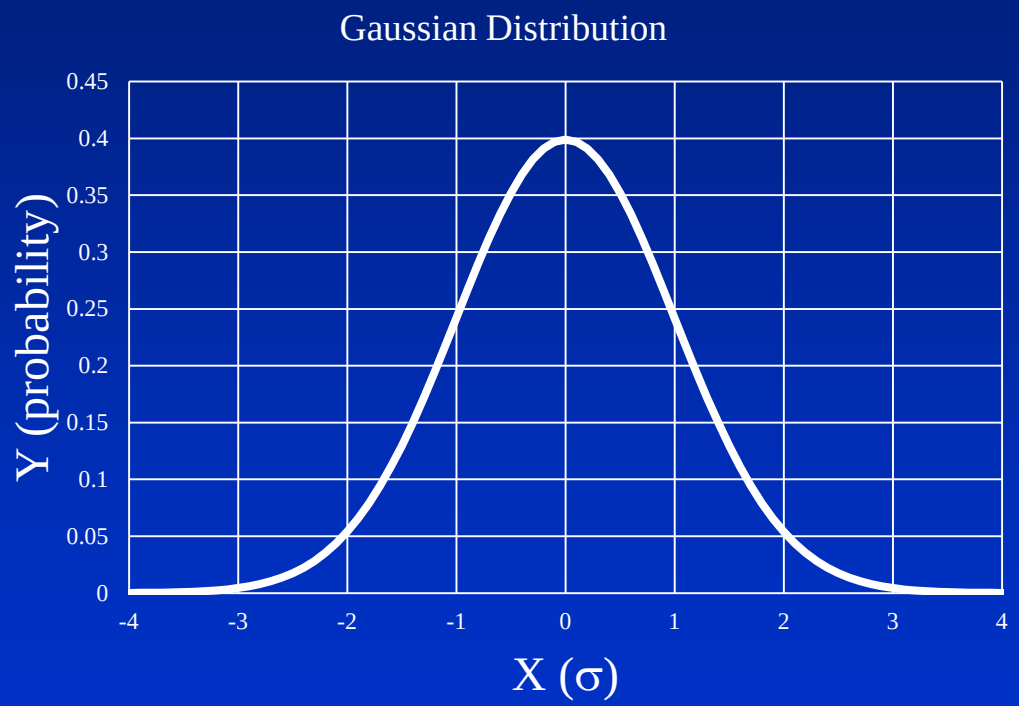
- Typical biologically based experiments involve a small number of tests (blood glucose, triglycerides, cholesterol) across many subjects
- LC/MS and NMR can generate thousands of variables, and it is often not practical to run on very many subjects

Estimating Error Using a Gaussian Probability

- If we assume a Gaussian probability for a series of measurements, it has the form:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

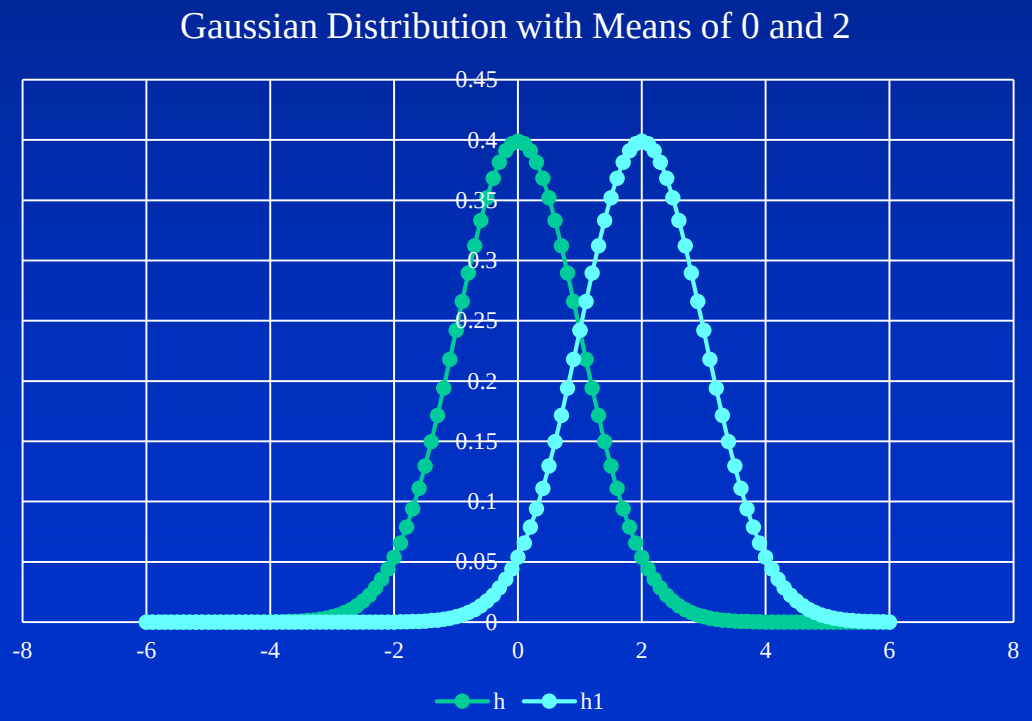
where σ is the standard deviation and μ is the average value



The Null Hypothesis

- The null hypothesis, H_0, is that there is not a significant difference between two sample populations

- When we sample the Gaussian we get a sample mean, $\bar{x}$, and a sample standard deviation, s, that is calculated from data



Confidence Intervals

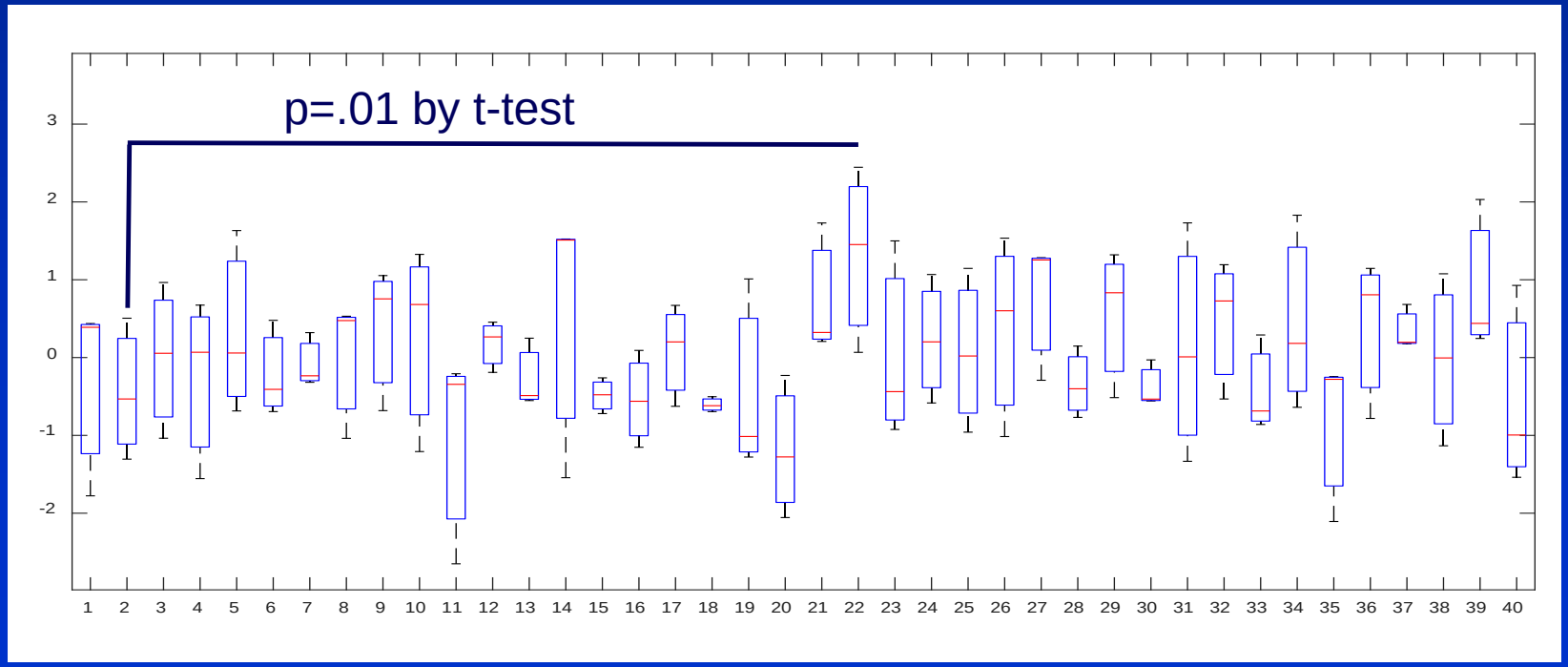
- A confidence interval uses Student's t-test to assess the likelihood that an average, $\bar{x}_1$, is not the same as the second average, $\bar{x}_2$.
- Usually a p-value is reported that gives the probability that two sets of samples are not drawn from distributions that have the same averages
- A 95% confidence interval means that you will randomly get a positive result 1 in every 20 measurements

Statistics (Type I and Type II Error)

- Type I error is detecting an effect that is not present
- Type II error is failing to detect an effect that is present

Using Students t-test 1000 times can generate a lot of type I errors

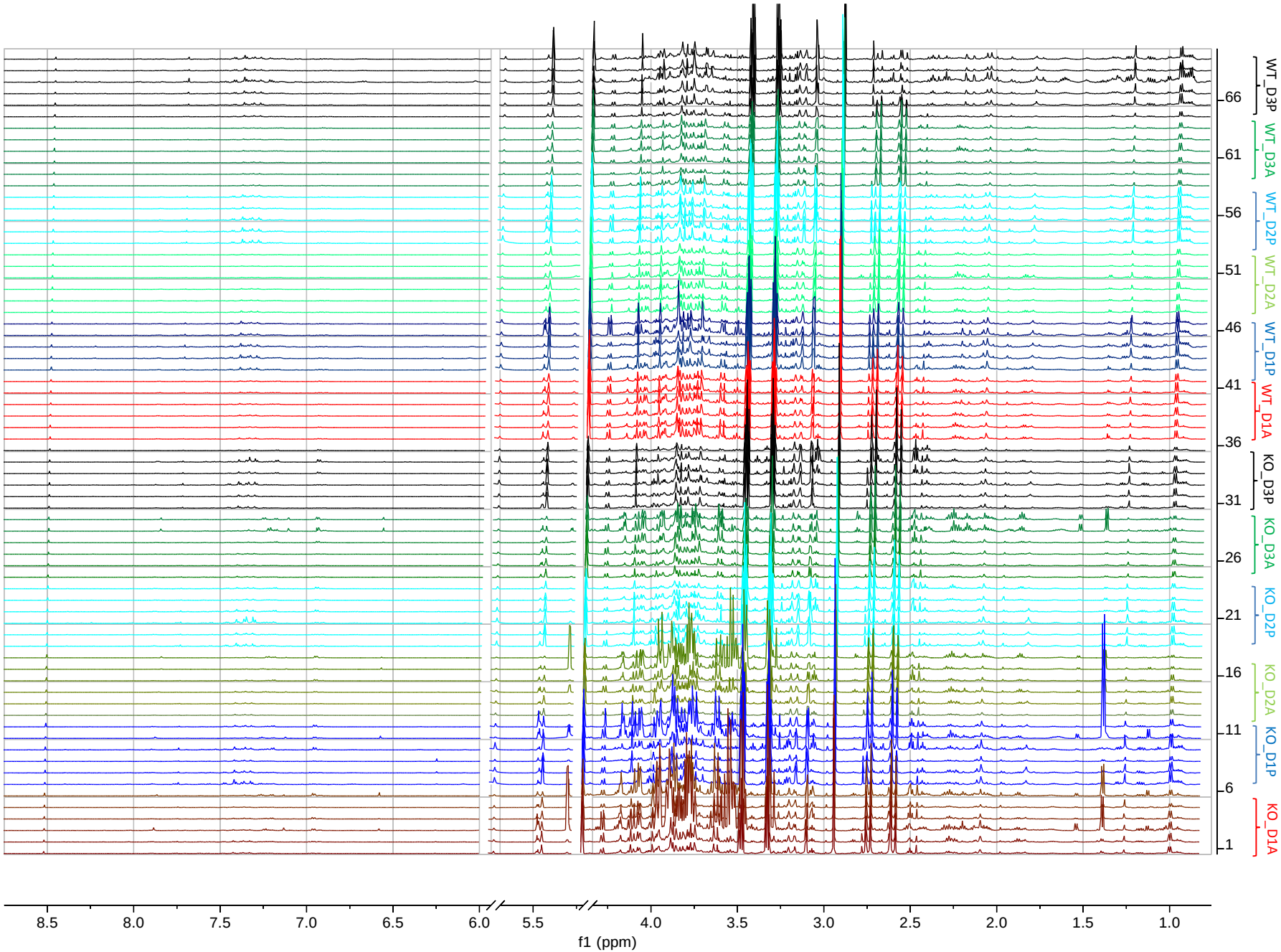
- 20 variables
- 3 “measures”
- Generated a set of Gaussian distributed numbers with average =0 and sigma=1



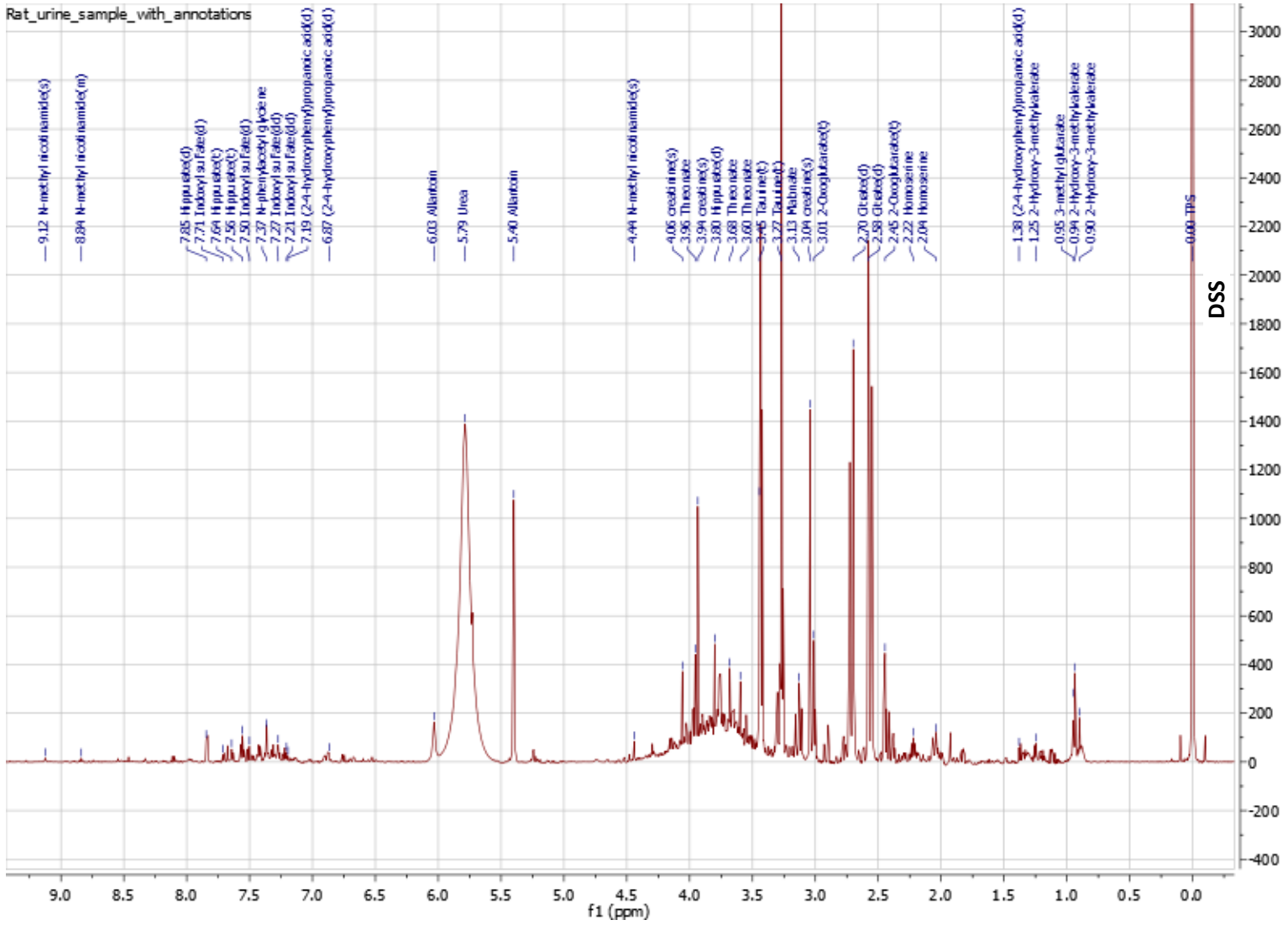
Example of Variable Bin Size

- The same set of proton NMR data was binned at .001 ppm vs .05 ppm
- Data was analyzed with the metaboanalyst pipeline

Overlay of ^{1}H NMR spectra for 70 samples



Assignments for Urine sample

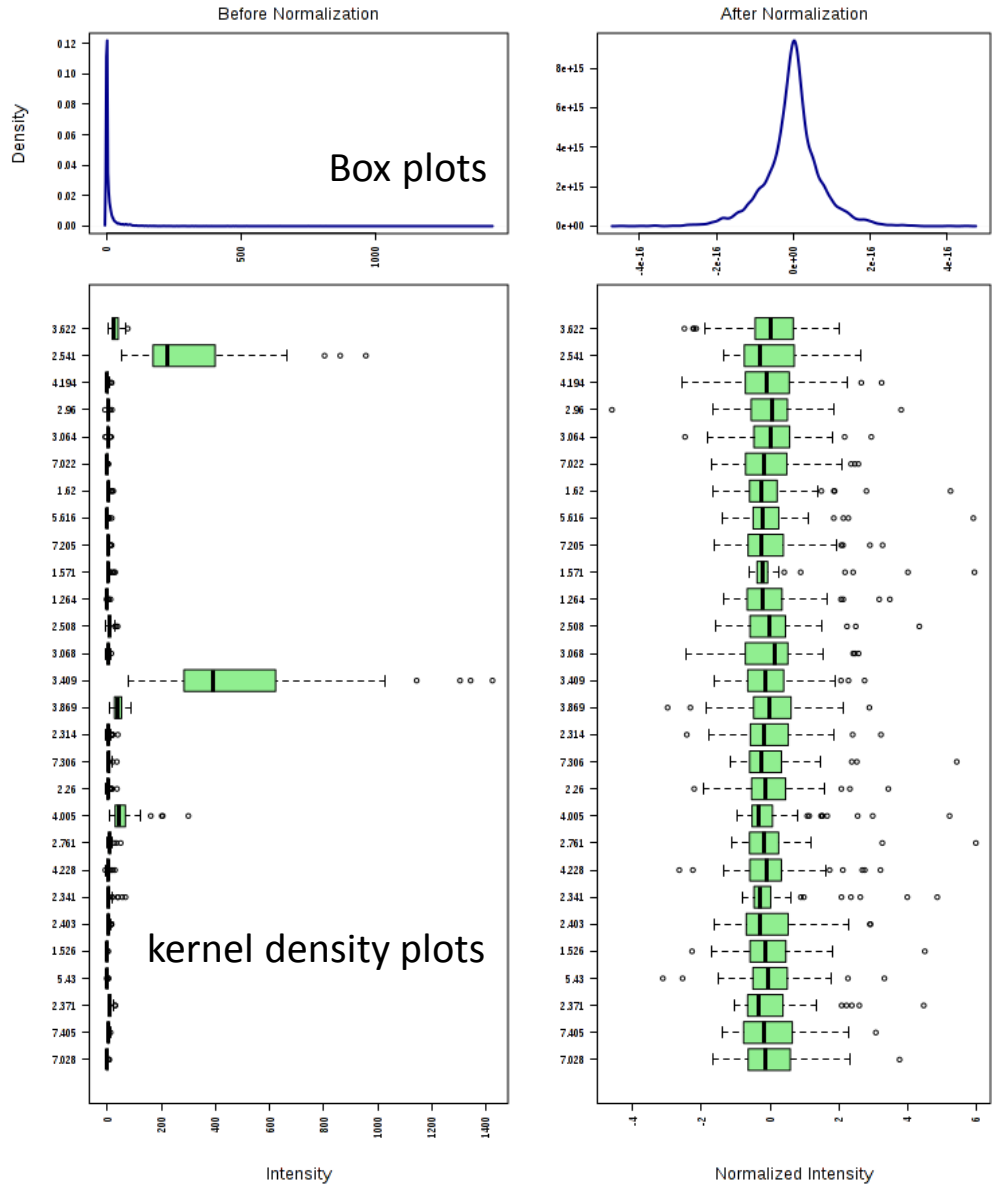


Parameters applied in Mestrenova and Metaboanalyst3.0 for the spectra of 70 samples

- All spectra were phase and baseline corrected
- Downfield region (>9.5ppm), upfield region (<0.5 ppm), Urea and water area were removed
- Normalization: probabilistic quotient normalization (PQN)
- Binned: 0.001 ppm
- Locally aligned: 162 regions
- Autoscaling (mean-centered and divided by s.d.
- CSV file was extracted and used in Mataboanalyst analysis

Normalization

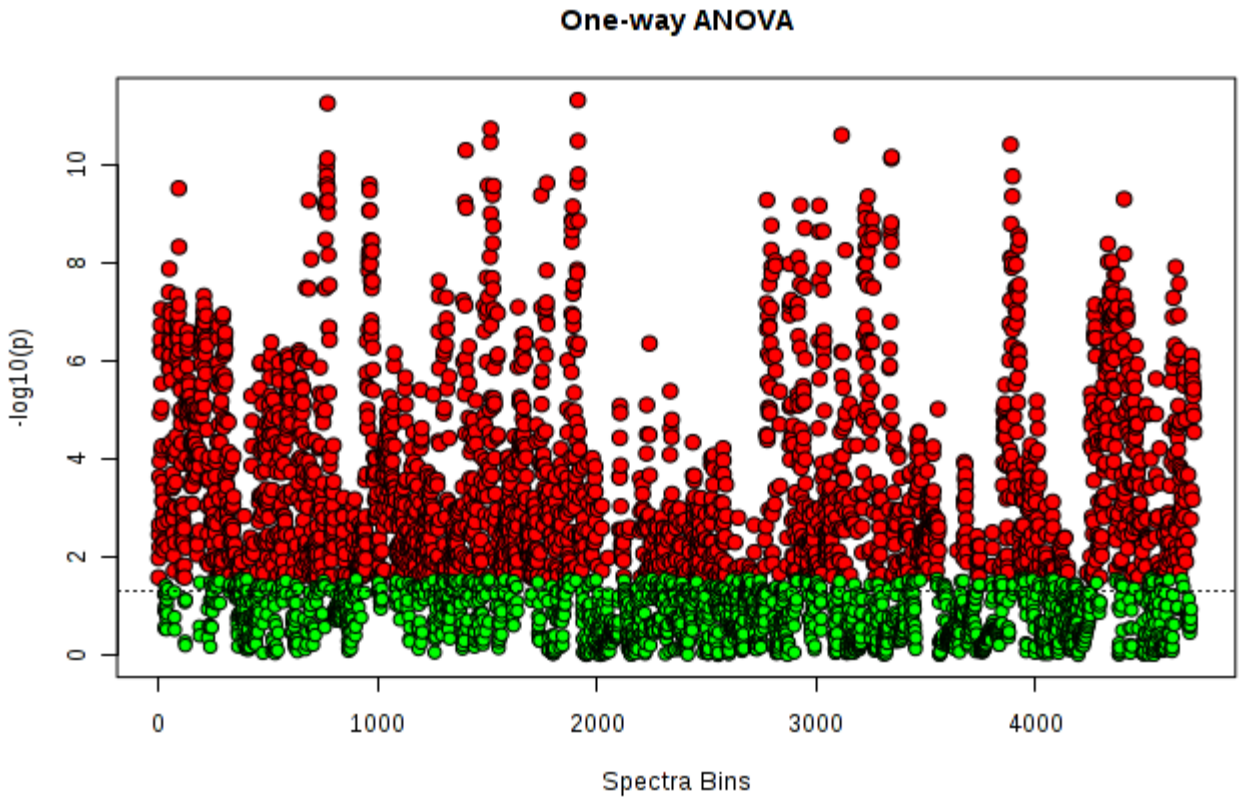
- Sample normalization allows general-purpose adjustment for differences among samples
- Performed by probabilistic quotient normalization (PQN)



One Way Analysis of Variance (ANOVA)

- Comparison column shows the comparisons between different levels that are significant given the p value threshold (dashed line with p=0.05).

- Red points indicates the most significant levels and subject of interests
- 2619 chemical shifts differed significantly among the different groups



Top 50 features identified by One-way ANOVA

- Metabolites = Spectra Bins (ppm)
- Top 50 features identified by One-way ANOVA
- Can be considered significant

Spectra Bins (ppm)	f.value	p.value
2.683	28.853	4.90E-12
1.493	28.636	5.63E-12
2.285	26.803	1.87E-11
3.887	26.363	2.52E-11
2.684	25.952	3.33E-11
2.284	25.874	3.51E-11
6.277	25.702	3.94E-11
2.172	25.316	5.14E-11
4.115	24.868	7.01E-11
1.492	24.758	7.56E-11
4.114	24.756	7.57E-11
1.489	24.202	1.12E-10
2.685	23.696	1.60E-10
1.49	23.604	1.71E-10
6.291	23.577	1.74E-10
2.682	23.158	2.35E-10
2.542	23.13	2.40E-10
1.488	23.118	2.42E-10
1.728	23.05	2.54E-10
1.491	23.017	2.60E-10
2.273	22.977	2.68E-10
2.298	22.95	2.73E-10
0.764	22.789	3.07E-10
1.494	22.757	3.14E-10
1.729	22.66	3.37E-10

Spectra Bins (ppm)	f.value	p.value
2.294	22.361	4.20E-10
2.516	22.354	4.22E-10
6.29	22.277	4.46E-10
4.007	22.264	4.51E-10
7.404	22.103	5.08E-10
3.546	22.037	5.33E-10
1.382	22.006	5.45E-10
1.496	21.991	5.51E-10
1.487	21.974	5.58E-10
2.168	21.914	5.83E-10
3.699	21.709	6.79E-10
3.785	21.688	6.90E-10
1.485	21.671	6.99E-10
2.66	21.623	7.24E-10
1.486	21.602	7.36E-10
2.173	21.548	7.66E-10
3.995	21.443	8.29E-10
1.727	21.392	8.61E-10
1.73	21.372	8.74E-10
1.495	21.221	9.79E-10
2.286	21.185	1.01E-09
3.996	20.899	1.25E-09
4.028	20.83	1.31E-09
2.657	20.791	1.35E-09

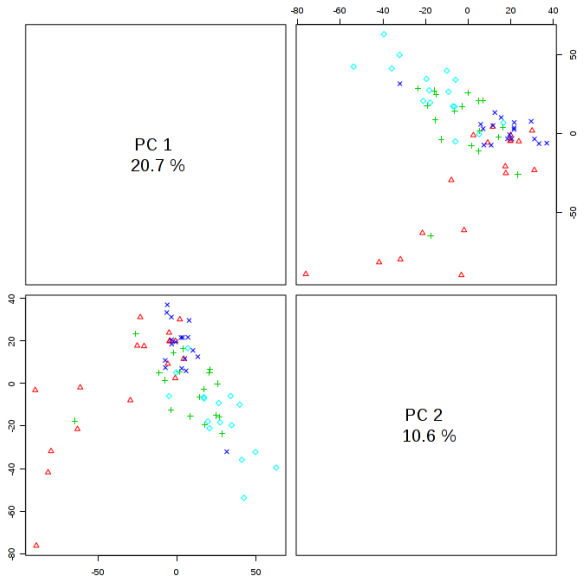
Corresponding metabolites for top 50 chemical shifts obtained from ANOVA

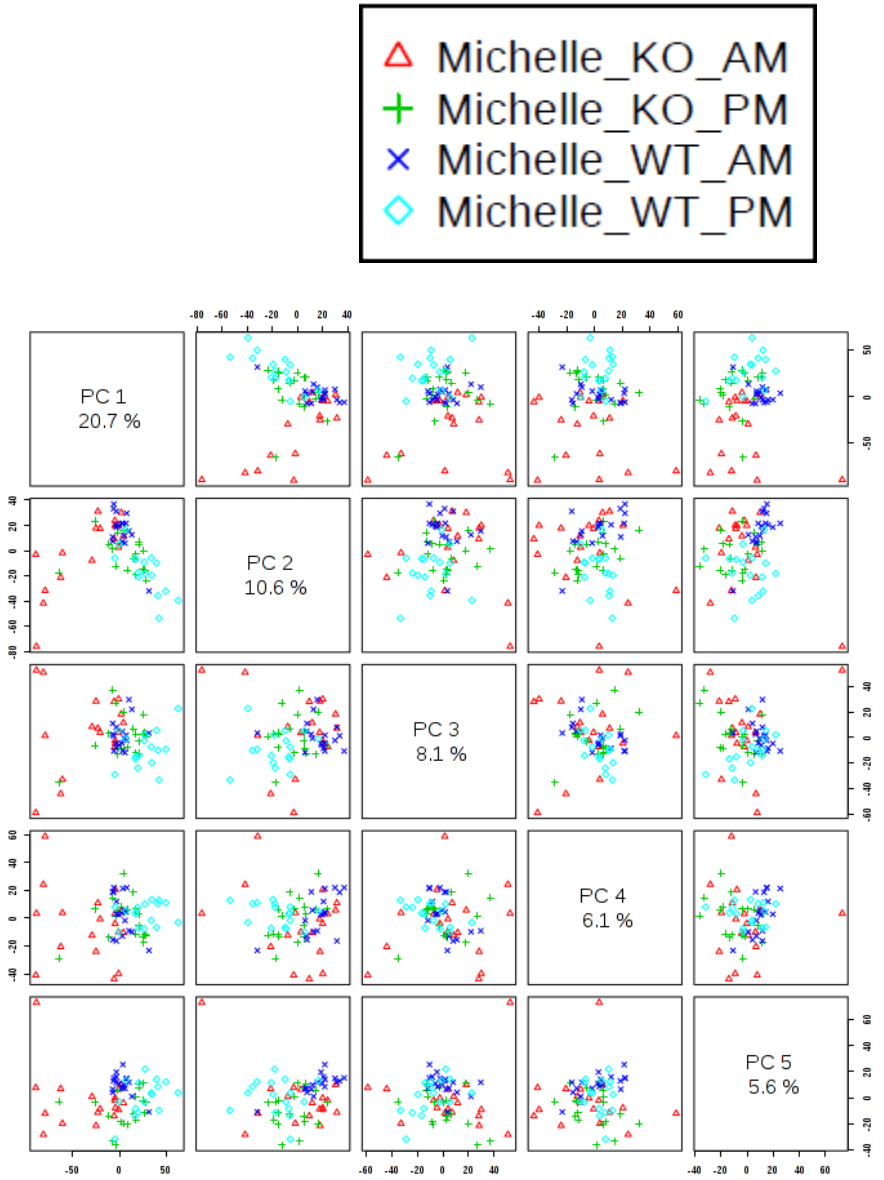
Spectra Bins(ppm)	Corresponding compounds
2.683	Citrate
1.493	Pimelate
2.285	2-Aminoadipate
3.887	Glucose
2.684	Citrate
2.284	2-Aminoadipate
6.277	2-Deoxyuridine/Thymidine?
2.172	Glutarate
4.115	Lactate
1.492	Pimelate
4.114	Lactate
1.489	Pimelate
2.685	Citrate
1.49	Pimelate
6.291	2-Deoxyuridine/Thymidine?
2.682	Citrate
2.542	Citrate
1.488	Pimelate
1.728	Glutarate
1.491	Pimelate
2.273	2-Aminoadipate
2.298	2-Aminoadipate
0.764	Unknown
1.494	Pimelate
1.729	Glutarate

Spectra Bins(ppm)	Corresponding compounds
2.294	2-Aminoadipate
2.516	Citrate
6.29	dCTP
4.007	Phenylalanine
7.404	Phenylalanine
3.546	Glucose
1.382	2-Octenoate??
1.496	Pimelate
1.487	Pimelate
2.168	Glutarate
3.699	Glucose
3.785	Glucose
1.485	Pimelate
2.66	Citrate
1.486	Pimelate
2.173	Glutarate
3.995	Phenylalanine
1.727	Glutarate
1.73	Glutarate
1.495	Pimelate
2.286	2-Aminoadipate
3.996	Phenylalanine
4.028	Phenylalanine
2.657	Citrate

PCA analysis

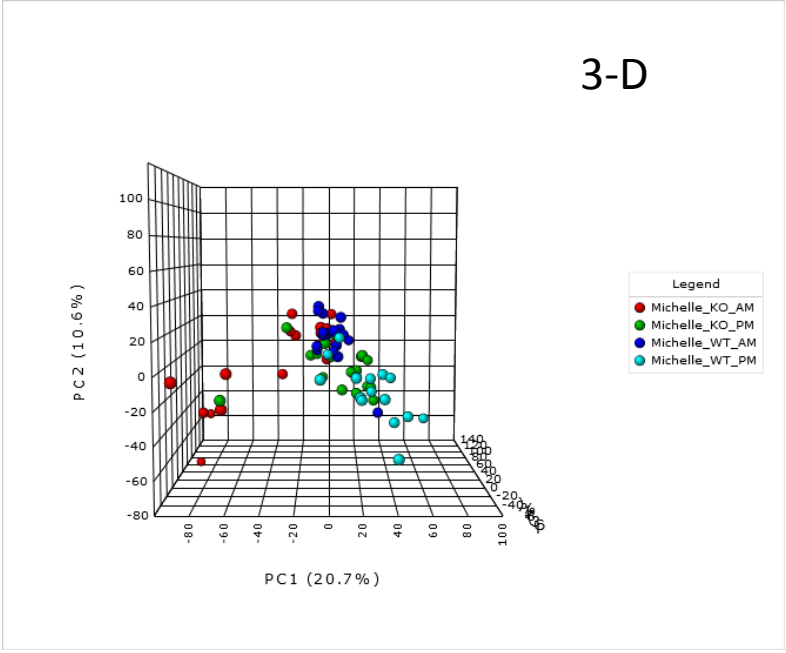
- Statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called **principal components**
- Unsupervised method
- The explained variance of each PC is shown in the corresponding diagonal cell
- Pairwise score plots providing an overview of the various separation patterns among the most significant PCs
- Components 1 and 2 explained 31.3% of the total variance among the groups

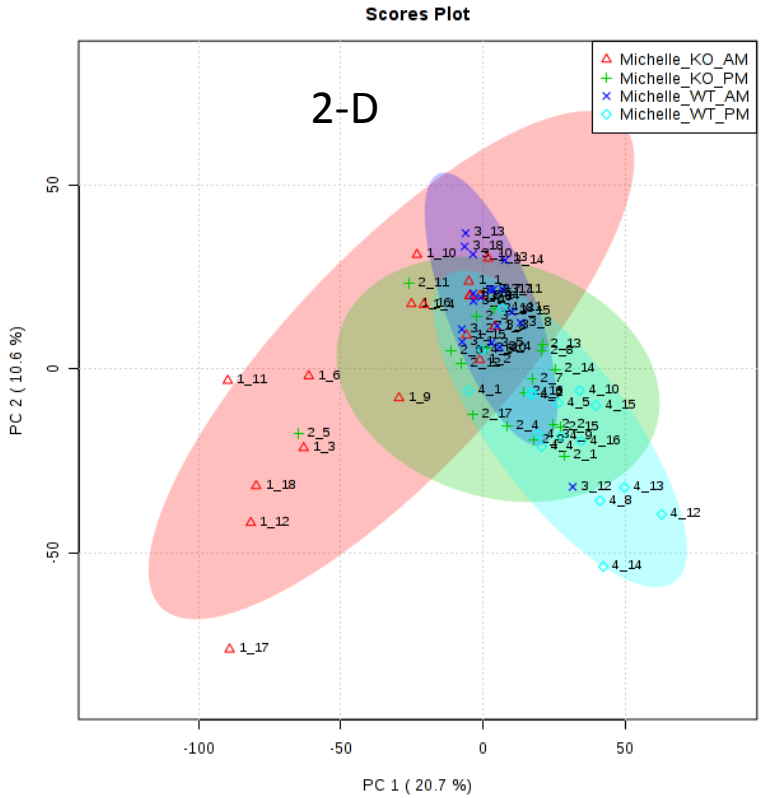




Scores plot (two and three-dimensional)

- Scores plot between PC1 vs PC2
- Explained variances are shown in brackets
- Not much separation among the groups

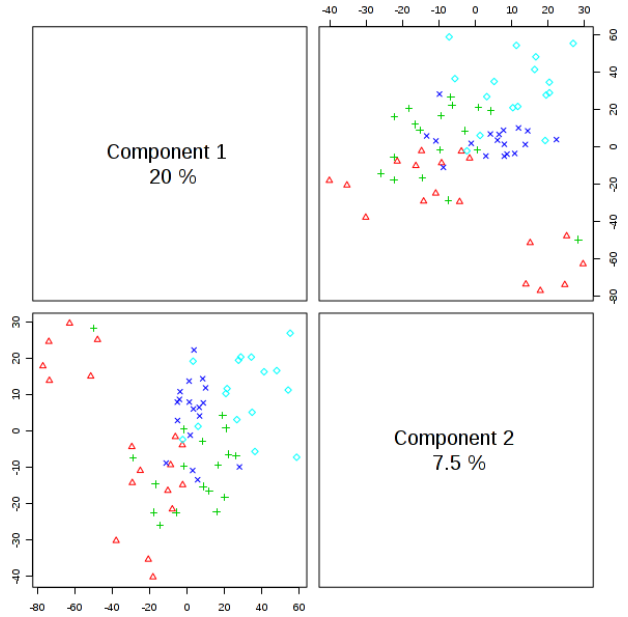


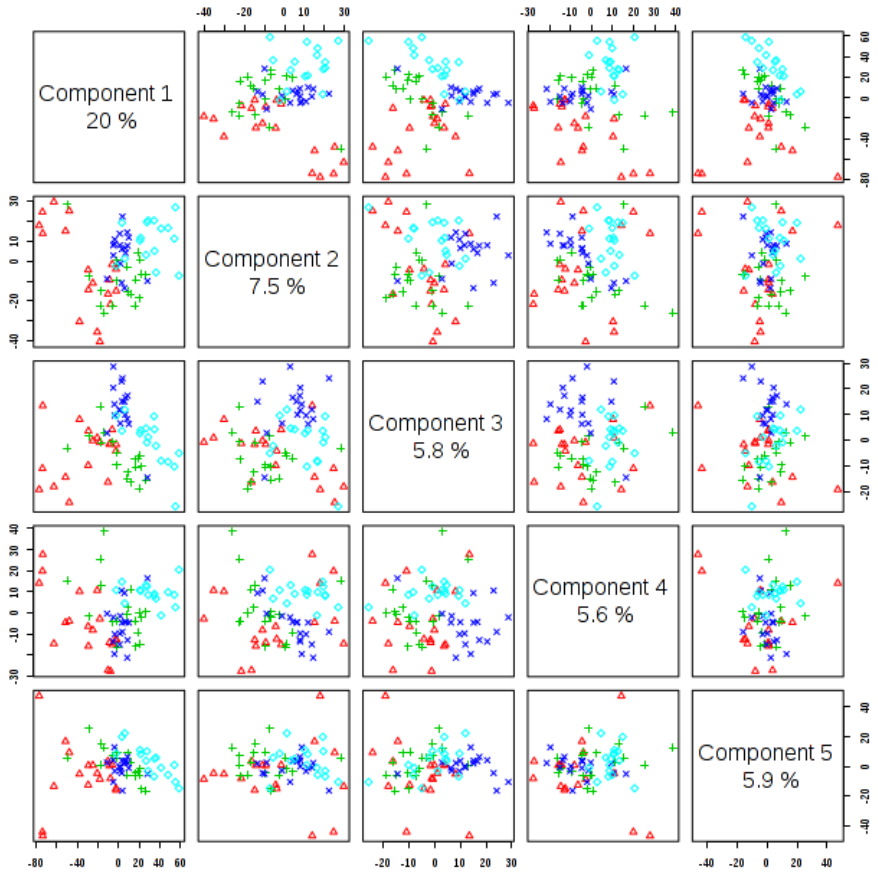


Partial Least Squares - Discriminant Analysis (PLS-DA)

- Supervised method
- Uses multivariate regression techniques to extract via linear combination of original variables (X) the information that can predict the class membership (Y)
- PLS-DA “rotates” the PCA axes to maximize separation

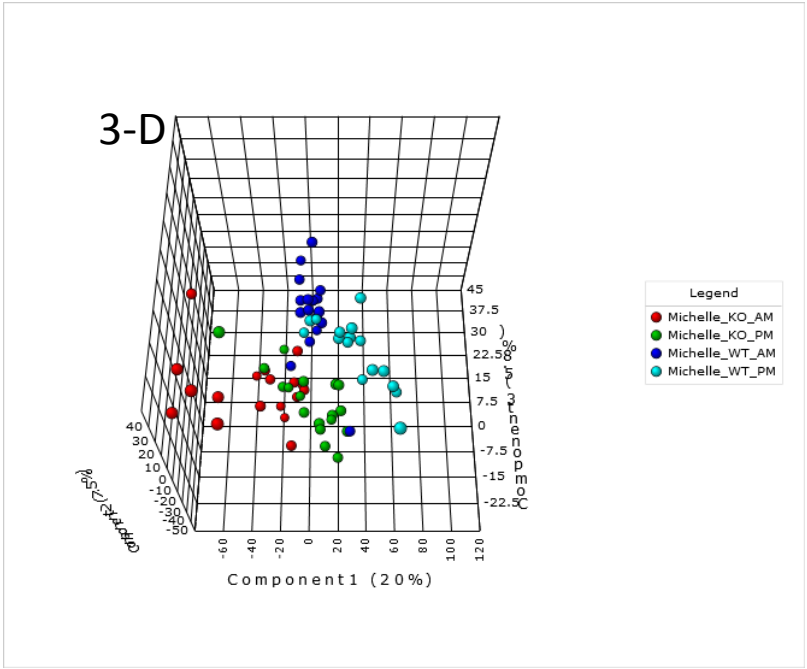
- Figure shows pairwise scores plots between the selected components
- The explained variance of each component is shown in the corresponding diagonal cell.
- Component 1 and 2 explained 27.5% of the total variance among the groups

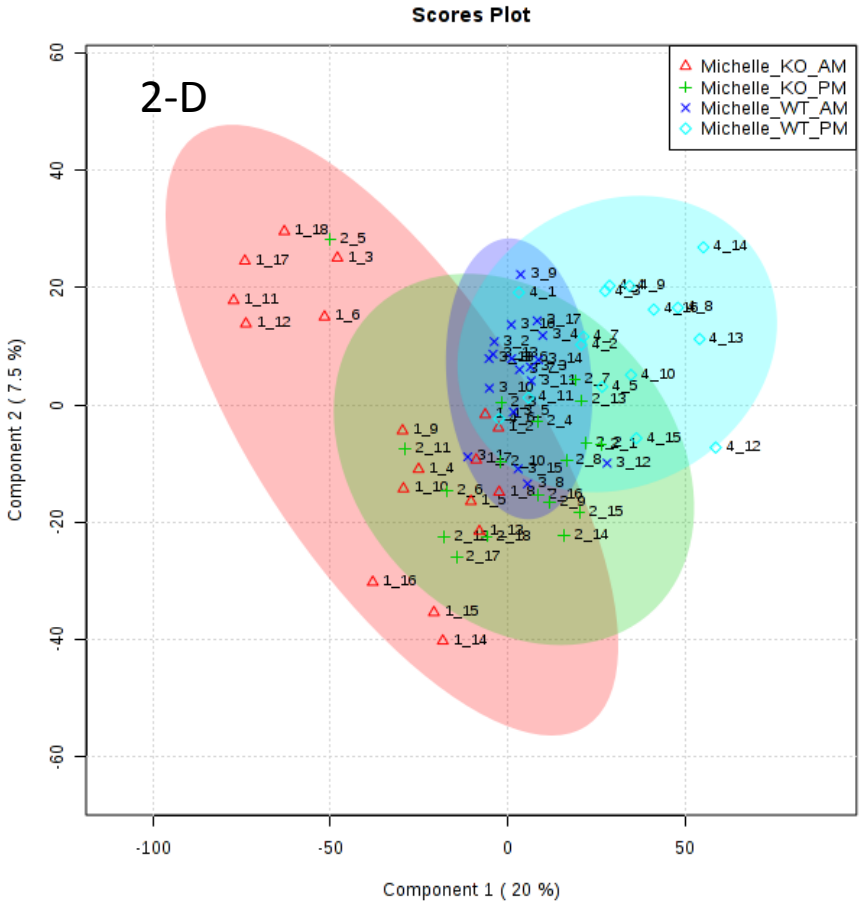




Scores plot obtained from PLS-DA (two and three-dimensional)

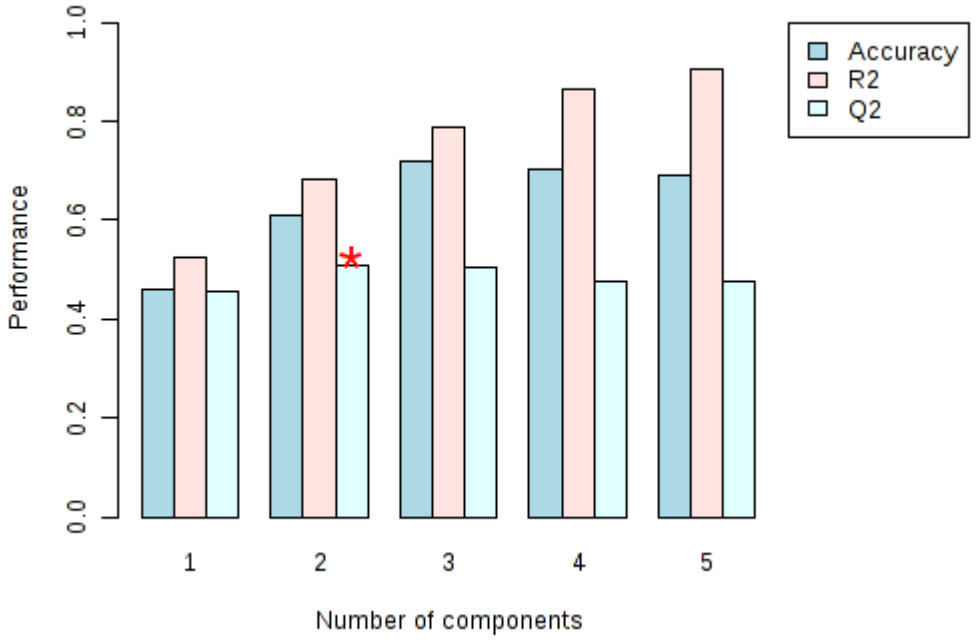
- Scores plot between the selected PCs
- The explained variances are shown in brackets
- Four clusters show some separation





Evaluation of PLS-DA Model for urine samples

- PLS-DA Model evaluated by cross validation of Q^2 and R^2
  - **R^2 – provides a measure of model fit to the original data**
  - **Q^2 – provides an internal measure of consistency between the original and cross-validation predicted data**



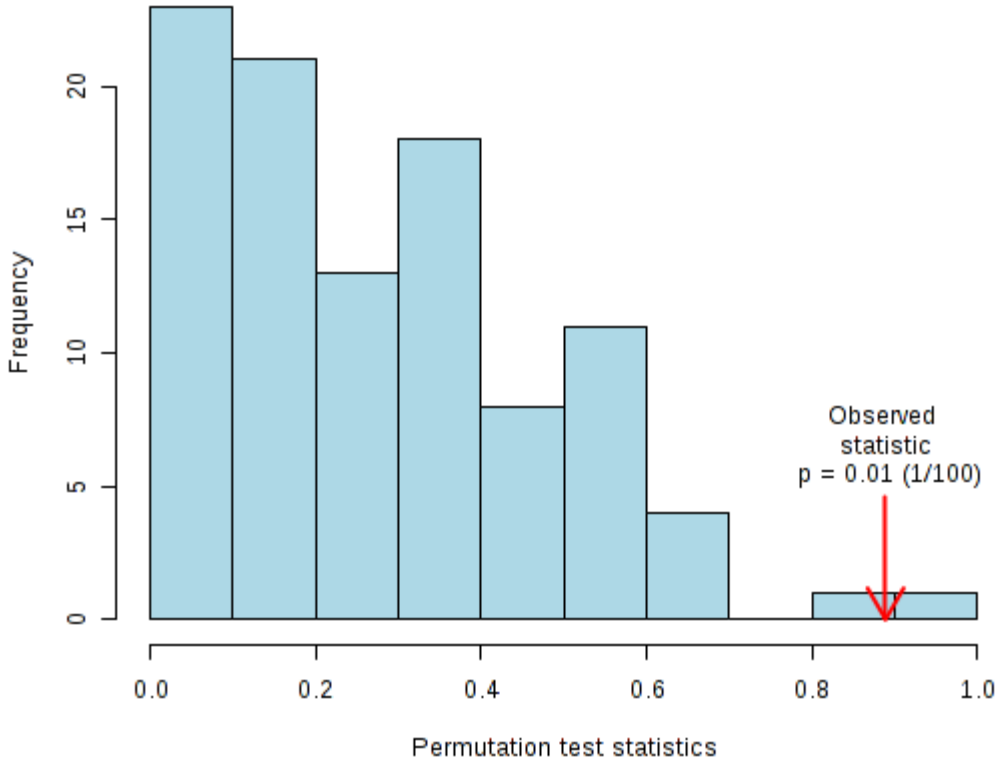
- Using too many components can overfit
- $Q^2 > 0.4$ for an acceptable model
- $R^2 >> Q^2$ possible indicator of model over-fitting
- $R^2/Q^2 > 0.5$: Acceptable model
- Since R^2 is slightly larger than Q^2 and not >> Q^2, there is a less chance of model over-fitting

PLS-DA cross validation details:

Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.45833	0.61111	0.71875	0.70486	0.69097
R2	0.52341	0.6813	0.78632	0.86473	0.90455
Q2	0.45661	0.50643	0.50266	0.47418	0.47477

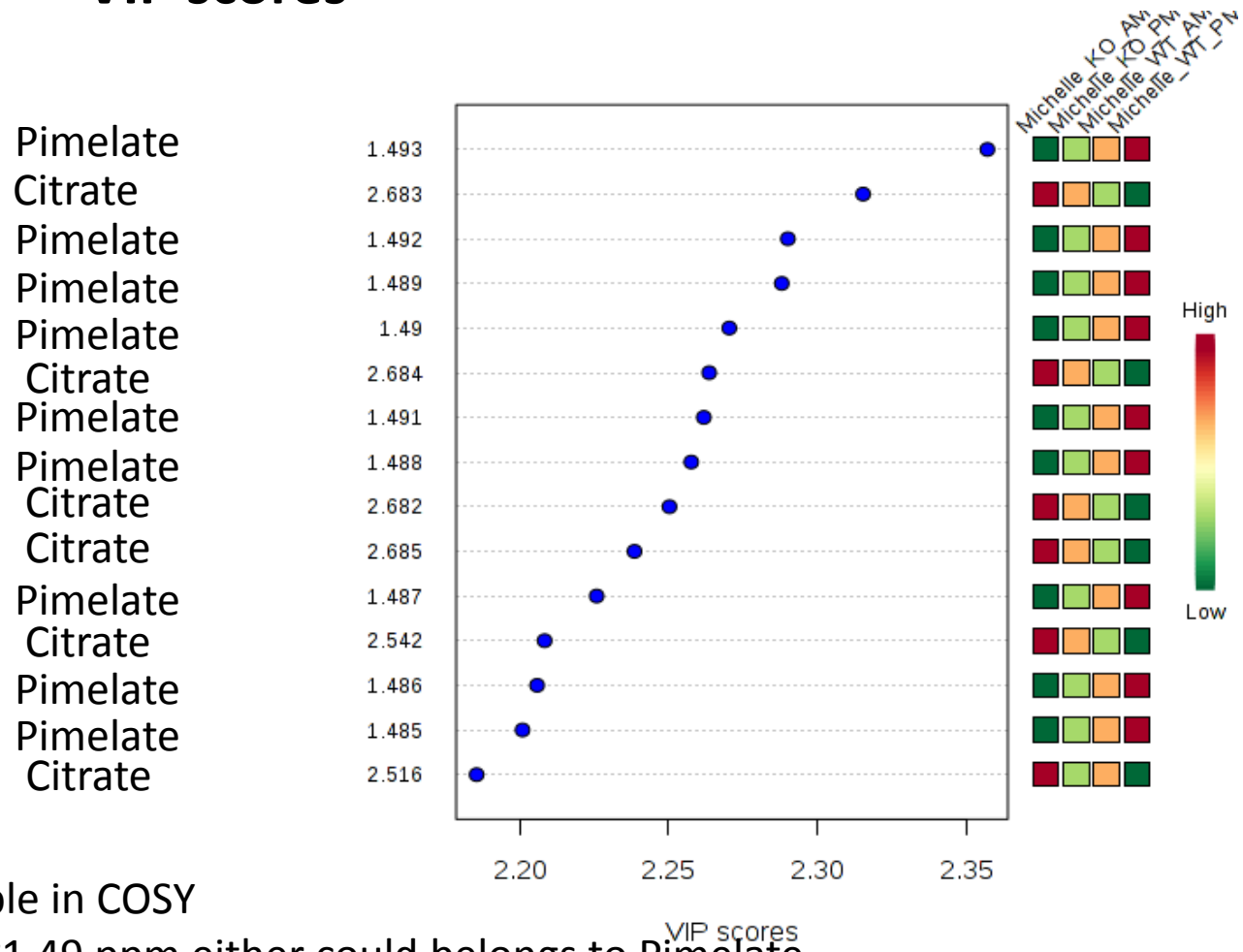
Model validation for urine samples

- PLS-DA model validation by permutation tests based on separation distance
- Performed with 10-fold CV and 100 permutation points
- The p value based on permutation is p = 0.01, the classifier is powerful



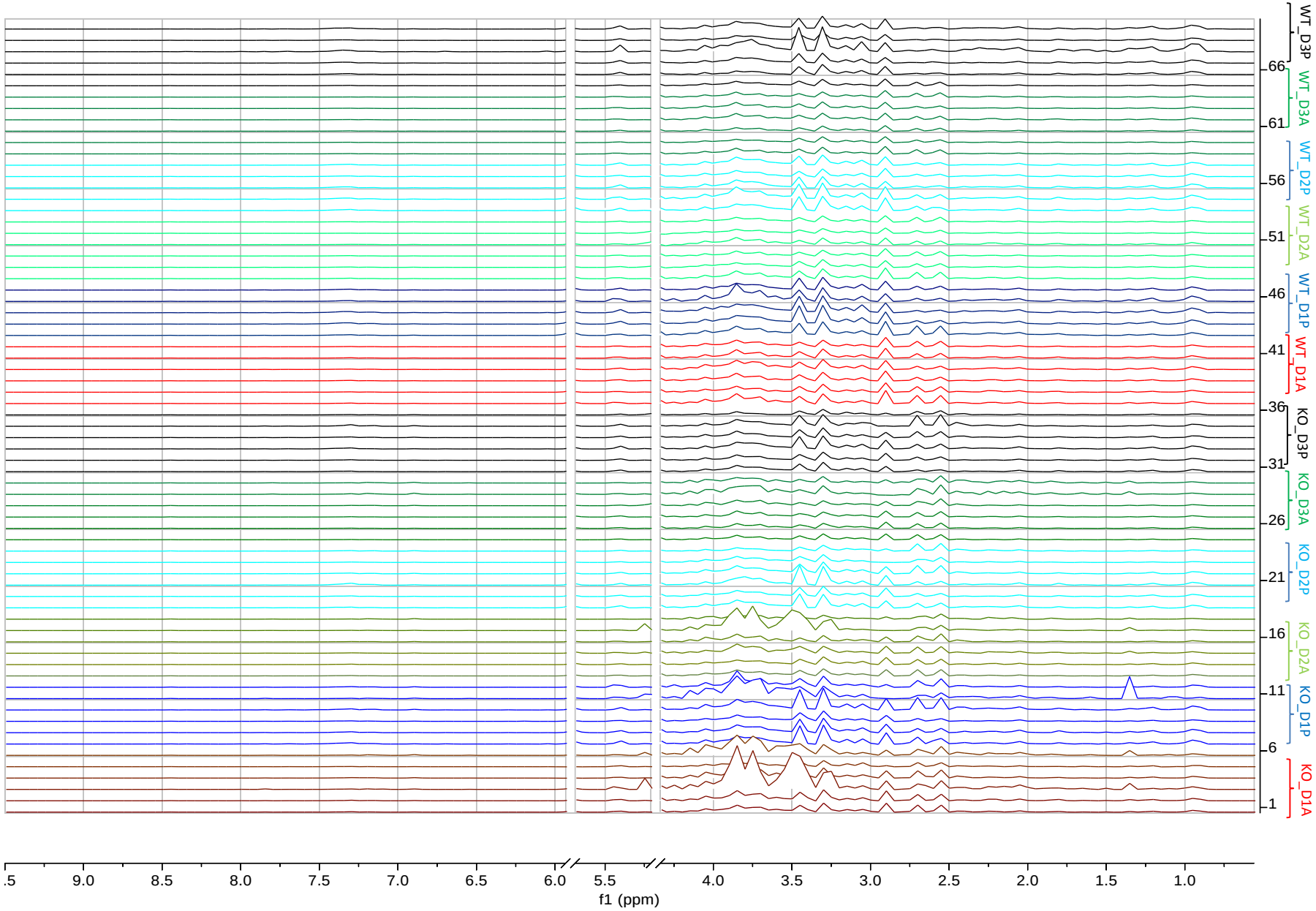
VIP scores

- VIP score is a measure of a variable's importance in the PLS-DA model
- Important features identified by PLS-DA
- The numbers on y-axis represents the chemical shifts in ppm

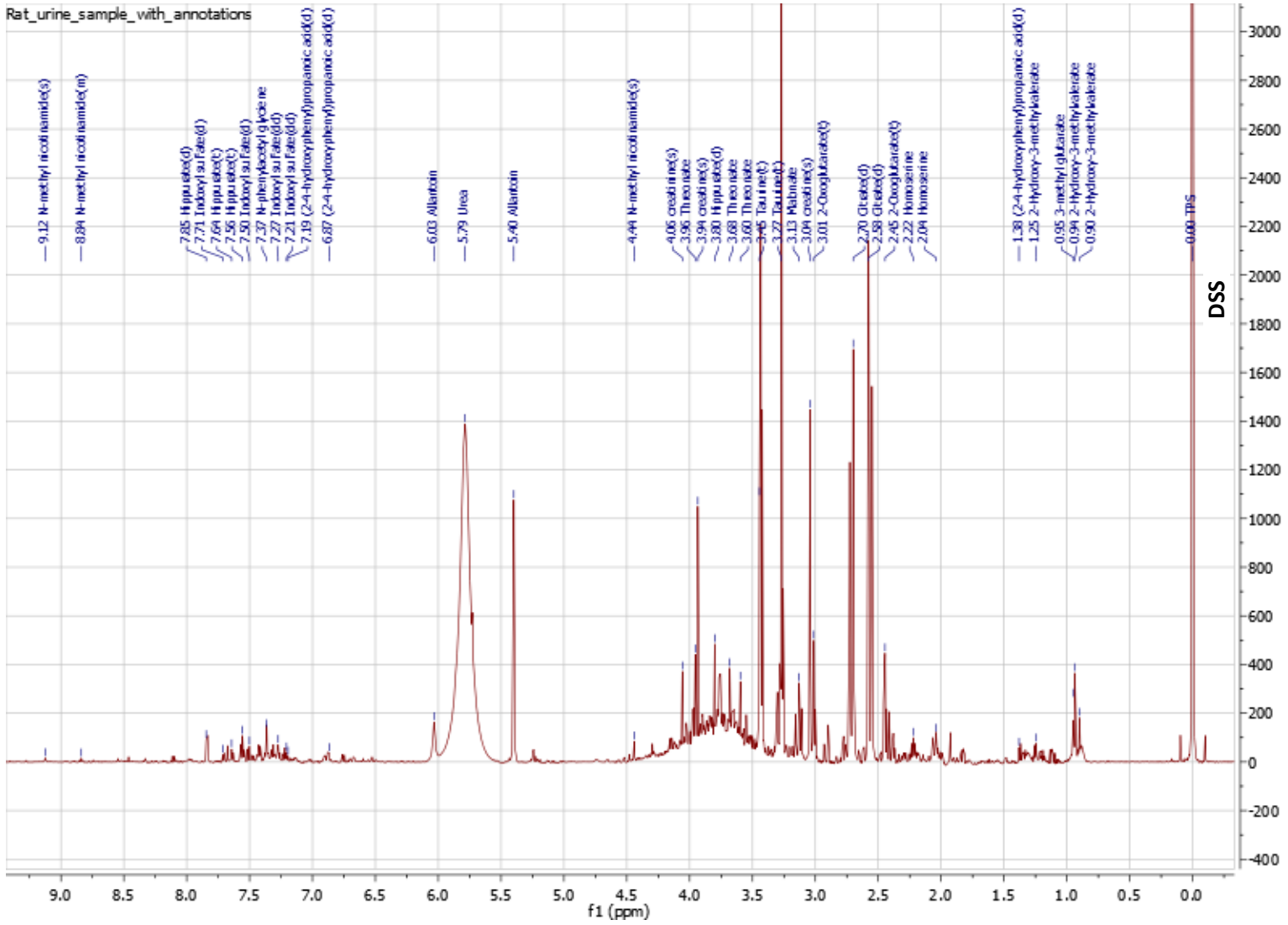


- No peaks for Pimelate visible in COSY
- Chenomx suggest peaks @~1.49 ppm either could belongs to Pimelate
- Correlation table suggest peaks @~1.49 ppm could belongs to Pimelate with correlation factor ≥ 0.54 for the peaks 1.49 ppm, 0.9 ppm and 1.33 ppm.
- Most probable compound is pimelate since it involved in the biosynthesis of lysine, while valerate (another possible compound) is found naturally in the perennial flowering plant valerian.

Overlay of ^{1}H NMR spectra for 70 samples



Assignments for Urine sample

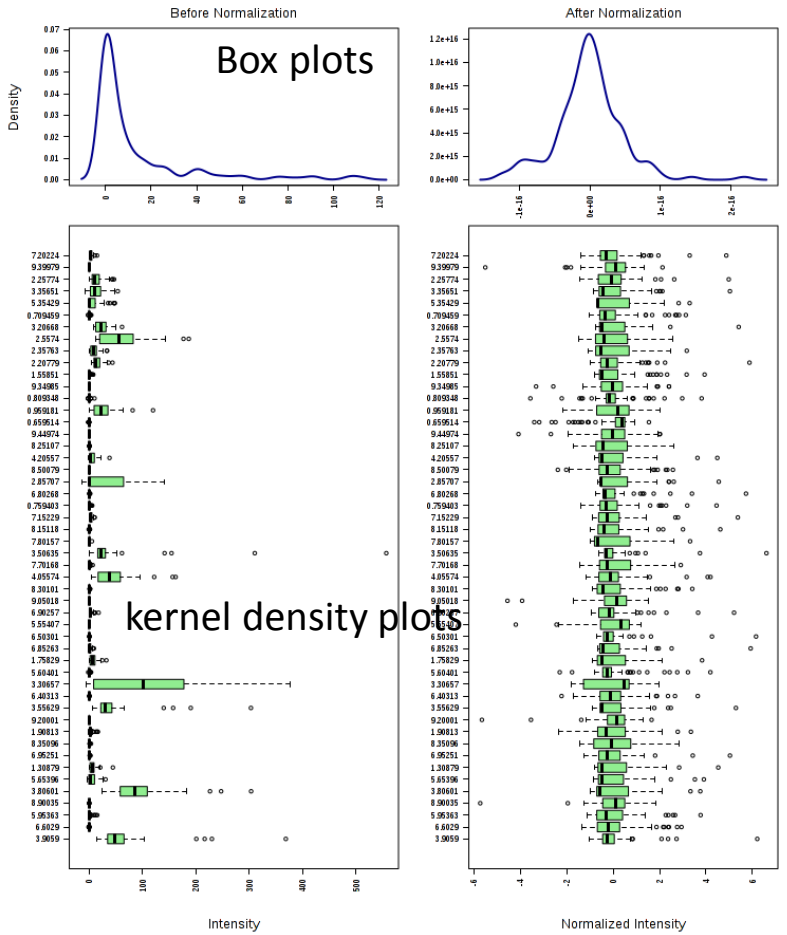


Parameters applied in Mestrenova and Metaboanalyst3.0 for the spectra of 70 samples

- All spectra were phase and baseline corrected
- Downfield region (>9.5ppm), upfield region (<0.5 ppm), Urea and water area were removed
- Normalization: probabilistic quotient normalization (PQN)
- Binned: 0.05 ppm
- Locally aligned: 162 regions
- Autoscaling
- CSV file was extracted and used in Mataboanalyst analysis

Normalization

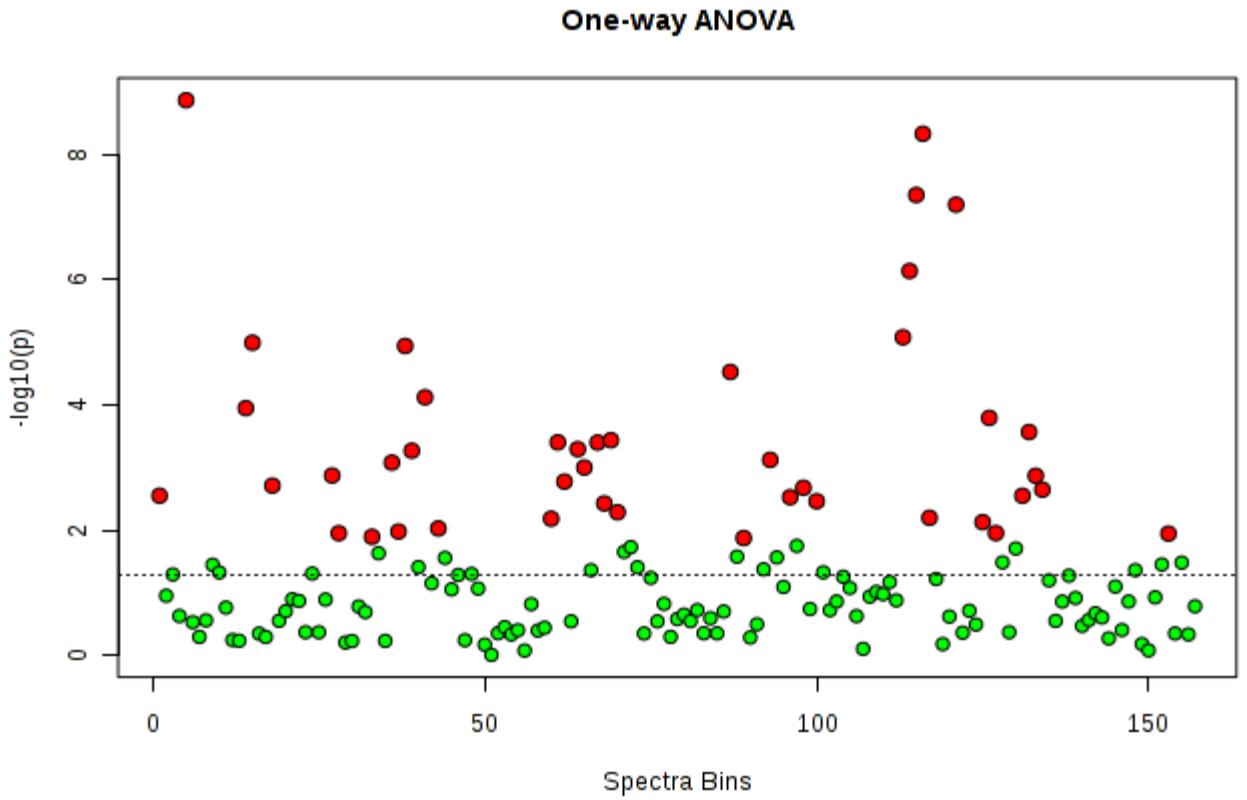
- Sample normalization allows general-purpose adjustment for differences among samples
- Performed by probabilistic quotient normalization (PQN)



One Way Analysis of Variance (ANOVA)

- Comparison column shows the comparisons between different levels that are significant given the p value threshold (dashed line with p=0.05).

- Red points indicates the most significant levels and subject of interests
- 43 chemical shifts differed significantly among the different groups



All 43 features identified by One-way ANOVA

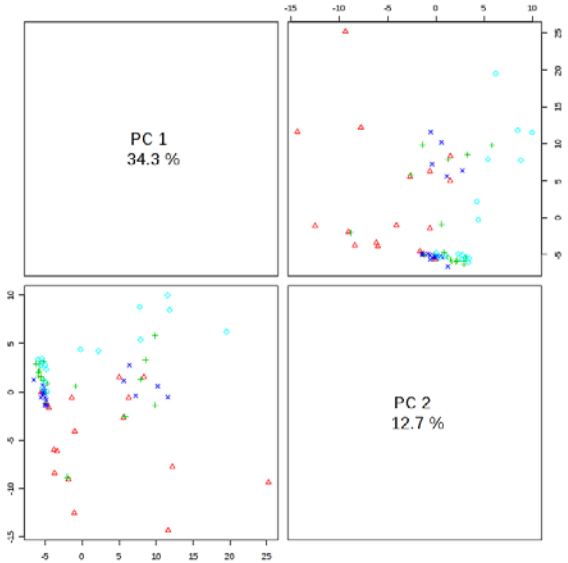
- Metabolites = Spectra Bins (ppm)
- Top 50 features identified by One-way ANOVA
- Can be considered significant

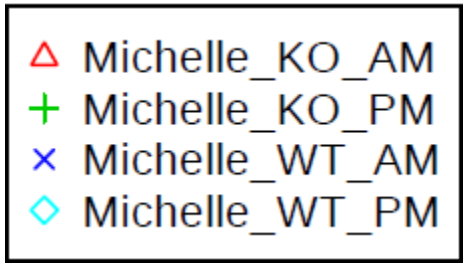
Spectra Bins (ppm)	f.value	p.value
0.759403	20.767	1.38E-09
7.45196	19.175	4.73E-09
7.40201	16.417	4.48E-08
7.70168	15.998	6.38E-08
7.35207	13.239	7.29E-07
7.30213	10.668	8.32E-06
1.25885	10.465	1.02E-05
2.40757	10.344	1.14E-05
6.00357	9.3941	2.97E-05
2.5574	8.4905	7.54E-05
1.2089	8.1115	0.000112
7.9514	7.7731	0.000161
8.25107	7.29	0.00027
3.95585	7.0134	0.000364
3.55629	6.9419	0.000394
3.85596	6.934	0.000397
3.70613	6.7122	0.000506
2.45751	6.6603	0.000536
6.30324	6.3564	0.00075
2.30768	6.2662	0.000829
3.75607	6.1041	0.000993
1.85818	5.8423	0.001332
8.30101	5.8345	0.001344
3.60624	5.644	0.001666
1.40868	5.5138	0.001932

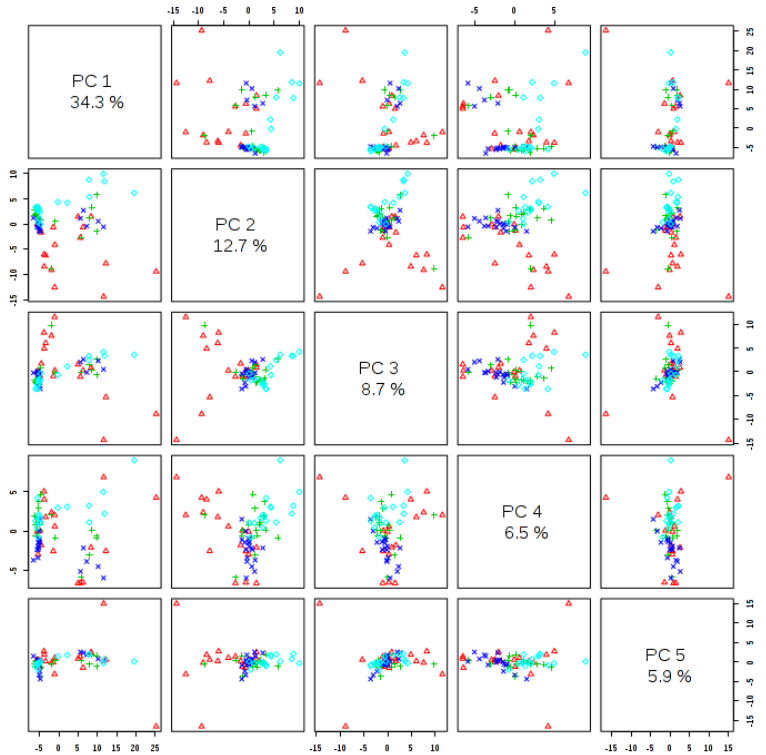
Spectra Bins (ppm)	f.value	p.value
6.55296	5.4457	0.002087
8.35096	5.3843	0.002238
0.559625	5.1886	0.002799
8.20113	5.188	0.002801
6.45307	5.1405	0.002958
6.65285	5.0119	0.003429
3.9059	4.943	0.003713
4.00579	4.6632	0.005133
7.5019	4.4879	0.006295
3.50635	4.4592	0.006509
7.90146	4.3508	0.007389
2.65729	4.1539	0.009309
2.35763	4.0578	0.010424
8.00135	4.0097	0.011032
1.90813	4.0064	0.011075
9.2999	3.9932	0.011249
2.15785	3.8934	0.012656
6.10346	3.8561	0.013228
6.55296	5.4457	0.002087
8.35096	5.3843	0.002238
0.559625	5.1886	0.002799
8.20113	5.188	0.002801
6.45307	5.1405	0.002958
6.65285	5.0119	0.003429

PCA analysis

- Statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called **principal components**
- Unsupervised method
- The explained variance of each PC is shown in the corresponding diagonal cell
- Pairwise score plots providing an overview of the various separation patterns among the most significant PCs
- Components 1 and 2 explained 47.1% of the total variance among the groups



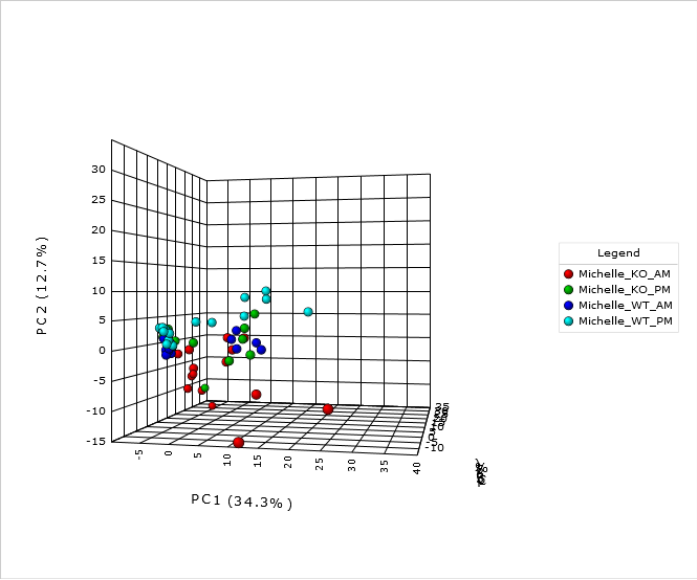




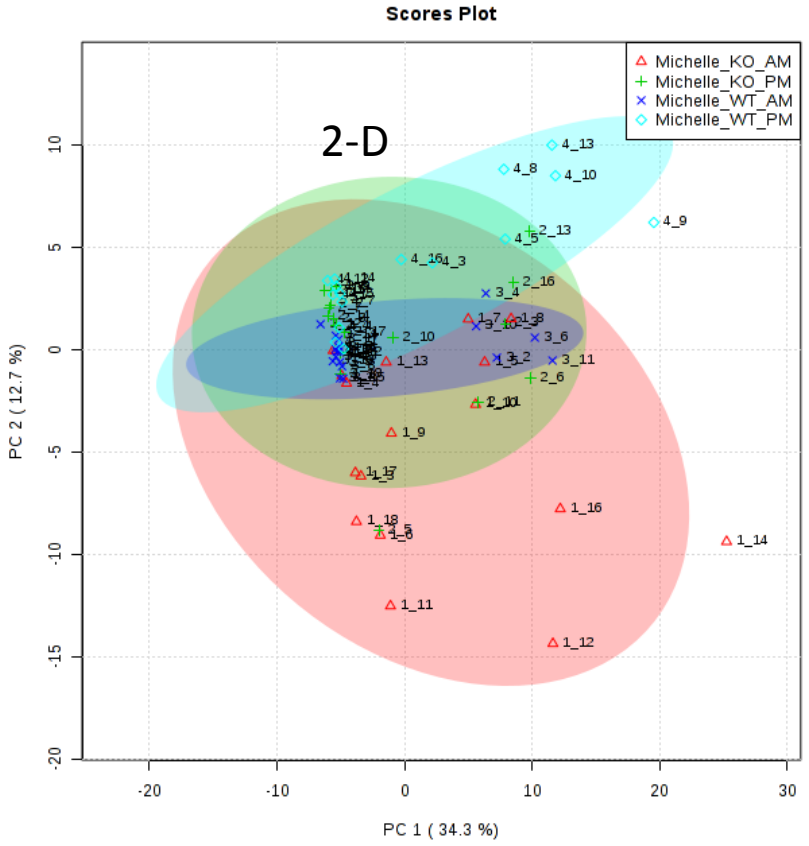
Scores plot (two and three-dimensional)

- Scores plot between PC1 vs PC2
- Explained variances are shown in brackets
- Not much separation among the groups

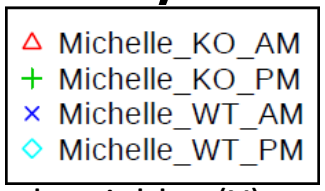
3-D



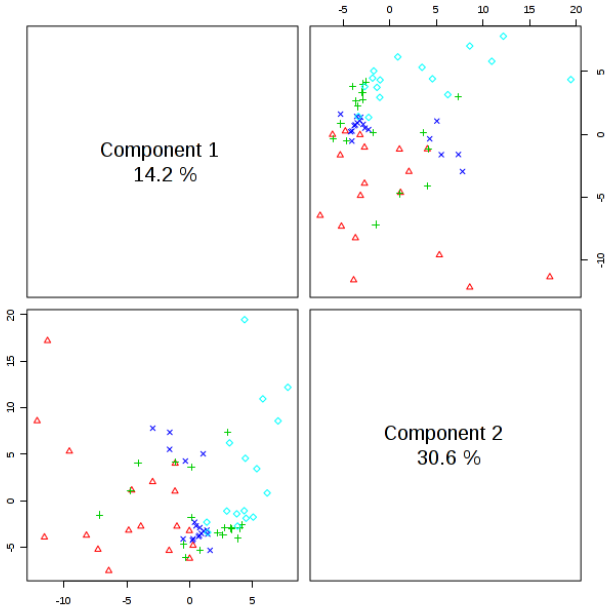
2-D

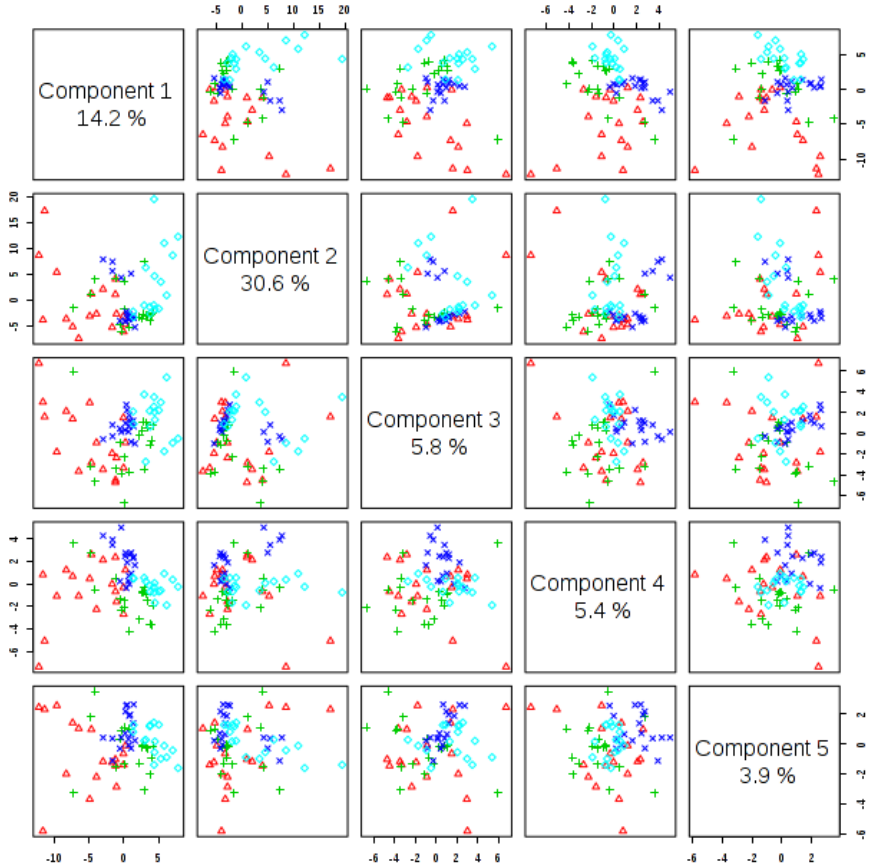


Partial Least Squares - Discriminant Analysis (PLS-DA)



- Supervised method
- Uses multivariate regression techniques to extract via linear combination of original variables (X) the information that can predict the class membership (Y)
- PLS-DA “rotates” the PCA axes to maximize separation
- Figure shows pairwise scores plots between the selected components
- The explained variance of each component is shown in the corresponding diagonal cell.
- Component 1 and 2 explained 44.8% of the total variance among the groups

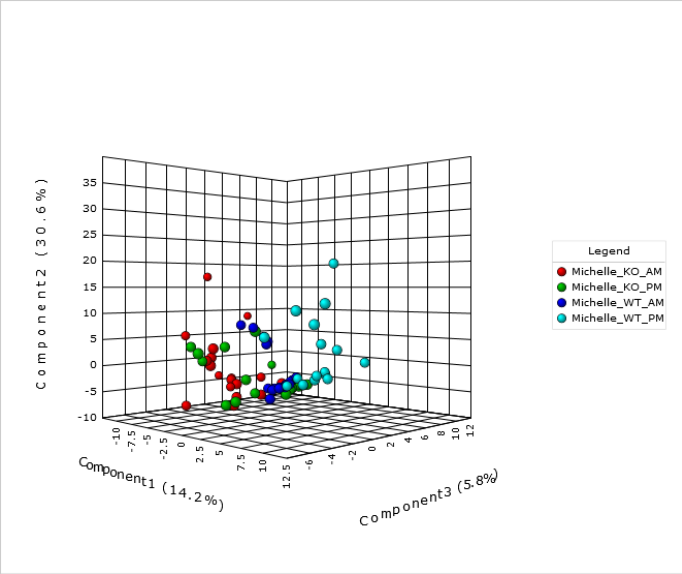




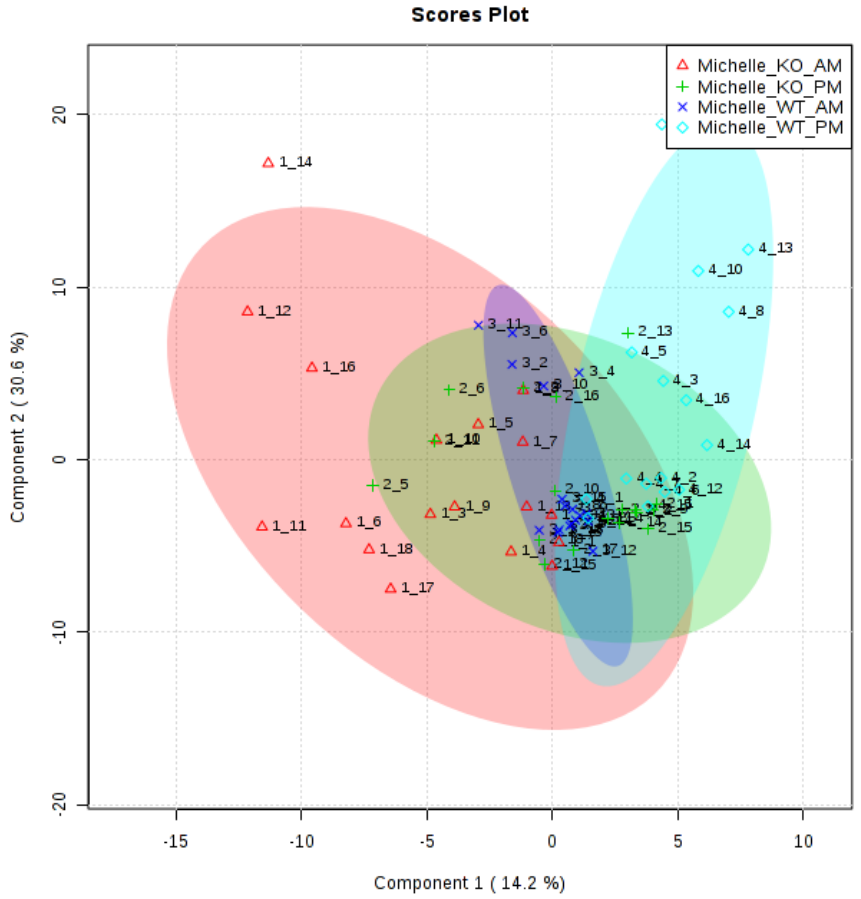
Scores plot obtained from PLS-DA (two and three-dimensional)

- Scores plot between the selected PCs
- The explained variances are shown in brackets
- Four clusters show not great separation

3-D

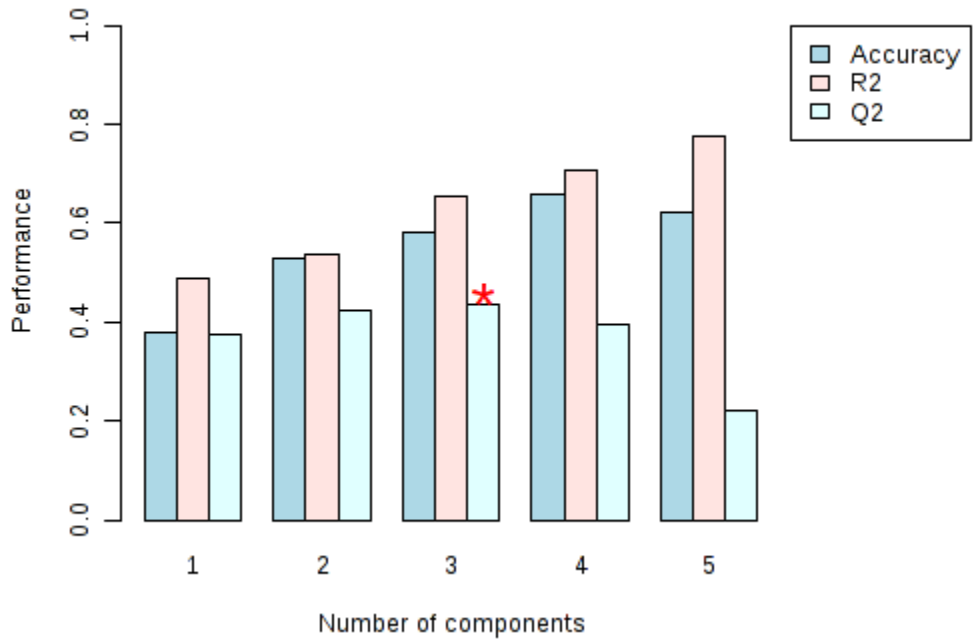


2-D



Evaluation of PLS-DA Model for urine samples

- PLS-DA Model evaluated by cross validation of Q^2 and R^2
  - **R^2 – provides a measure of model fit to the original data**
  - **Q^2 – provides an internal measure of consistency between the original and cross-validation predicted data**
- Using too many components can overfit
- $Q^2 > 0.4$ for an acceptable model
- $R^2 >> Q^2$ possible indicator of model over-fitting
- $R^2/Q^2 > 0.5$: Acceptable model
- Since R^2 is slightly larger than Q^2 and not >> Q^2, there is a less chance of model over-fitting

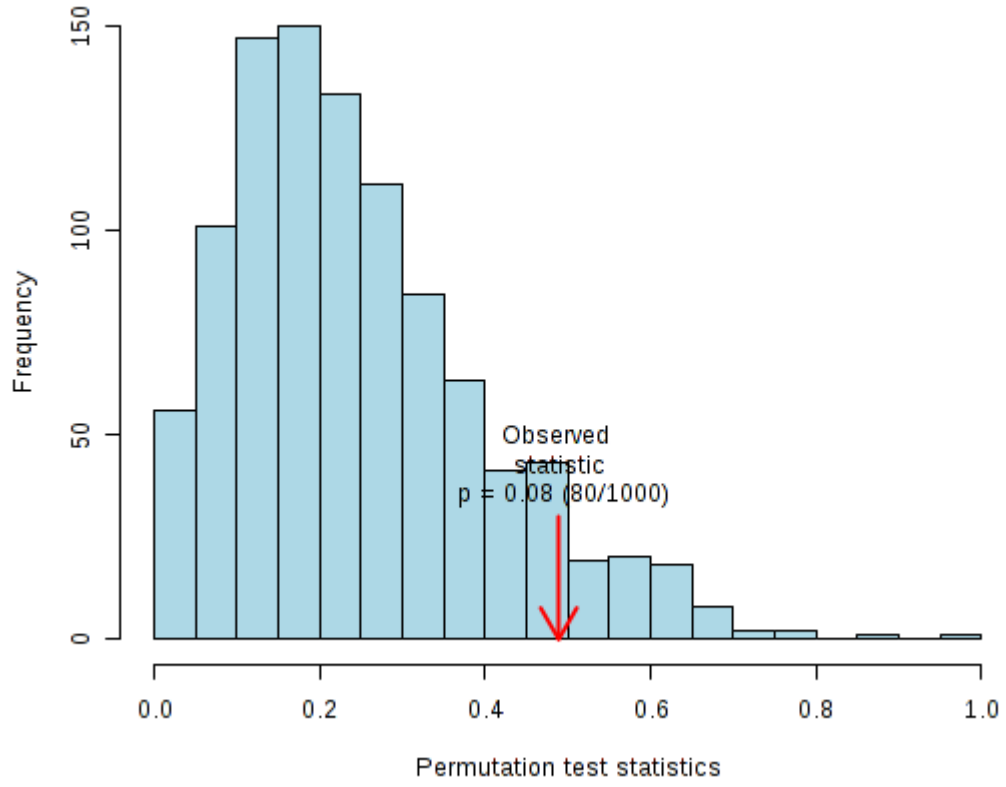


PLS-DA cross validation details:

Measure	1 comps	2 comps	3 comps	4 comps	5 comps
Accuracy	0.38	0.53	0.58333	0.65667	0.62
R2	0.48639	0.53492	0.65527	0.70858	0.77417
Q2	0.37411	0.425	0.43503	0.39493	0.21958

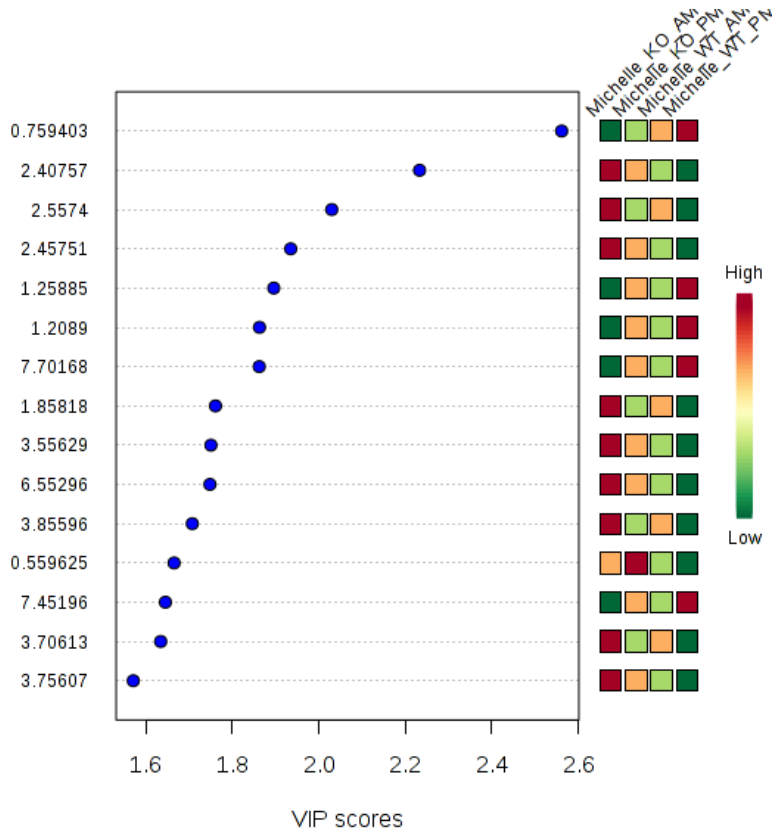
Model validation for urine samples

- PLS-DA model validation by permutation tests based on separation distance
- Performed with 10-fold CV and 100 permutation points
- The p value based on permutation is p = 0.08, the classifier is not powerful



VIP scores

- VIP score is a measure of a variable's importance in the PLS-DA model
- Important features identified by PLS-DA
- The numbers on y-axis represents the chemical shifts in ppm



Think carefully about statistics:
Then consult a statistician