Analysis of NMR Data
Thursday, July 25, 2013 (10:20-12:00 noon)

NM Data Format and Processing:
- Software for NMR Processing
- Fourier Transformation
- Line Broadening
- Base Line Correction

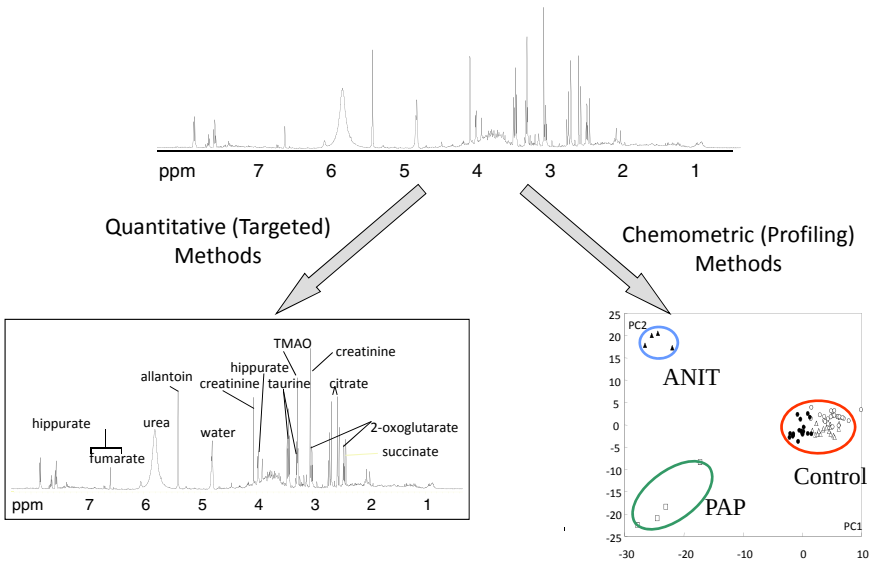
Chemical Shift Libraries:
- Metabolite Identification
- Peak Integration
- Metabolite Quantification
- Bucketing Approach

Multivariate Analysis:
- Group Separation
- Outliners
- PCA
- PLS-DA

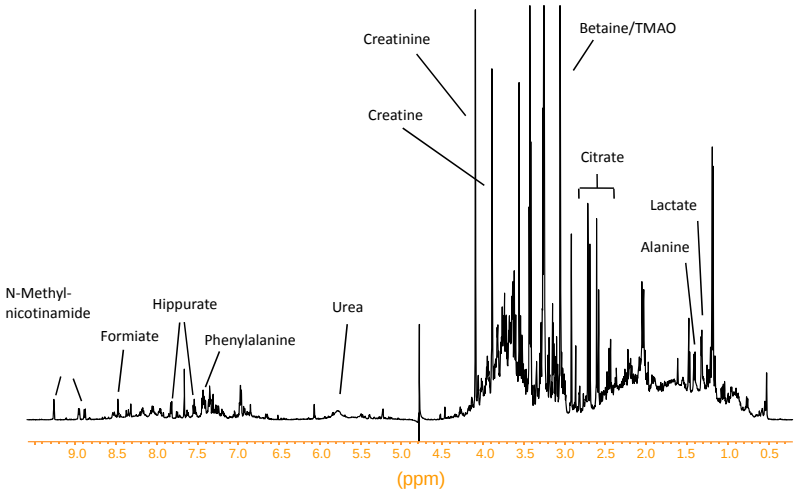
Biomarker Validation:
- Group Comparison (t-Test, ANOVA)
- Pathway Analysis
- System Biology

Clinical Metabolomics:
- Clinical Laboratory Tests
- Translational Imaging (MRSI, PET)
- Phase I-III Biomarker Trials

2 Routes to Metabolomics

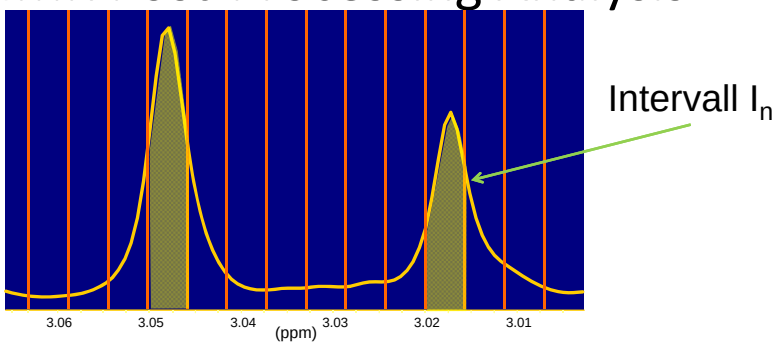


NMR Post-Processing Analysis



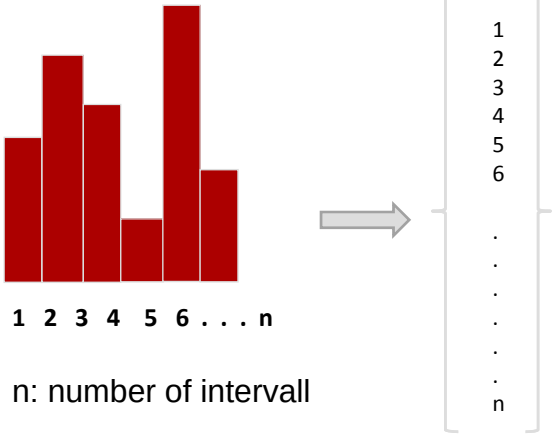
600 MHz spectrum of human neonate urine

NMR Post-Processing Analysis



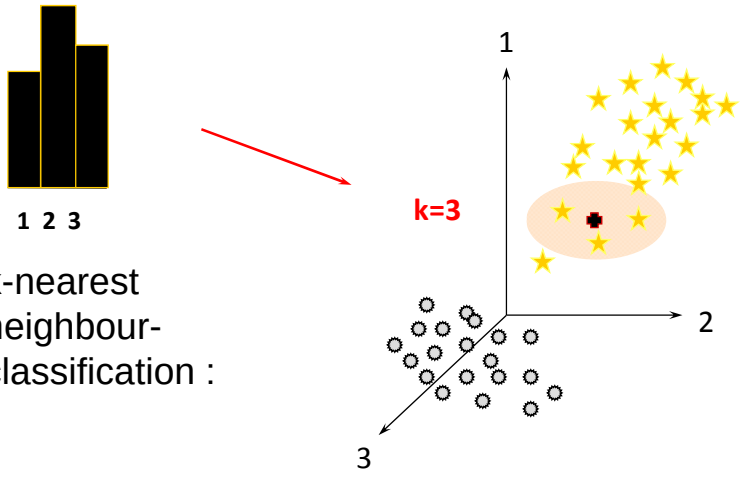
normation : $$\frac{\text{integral } I_n}{\text{sum of integrals}}$$

NMR Post-Processing Analysis



n: number of intervall

NMR Post-Processing Analysis



k-nearest neighbour-classification :

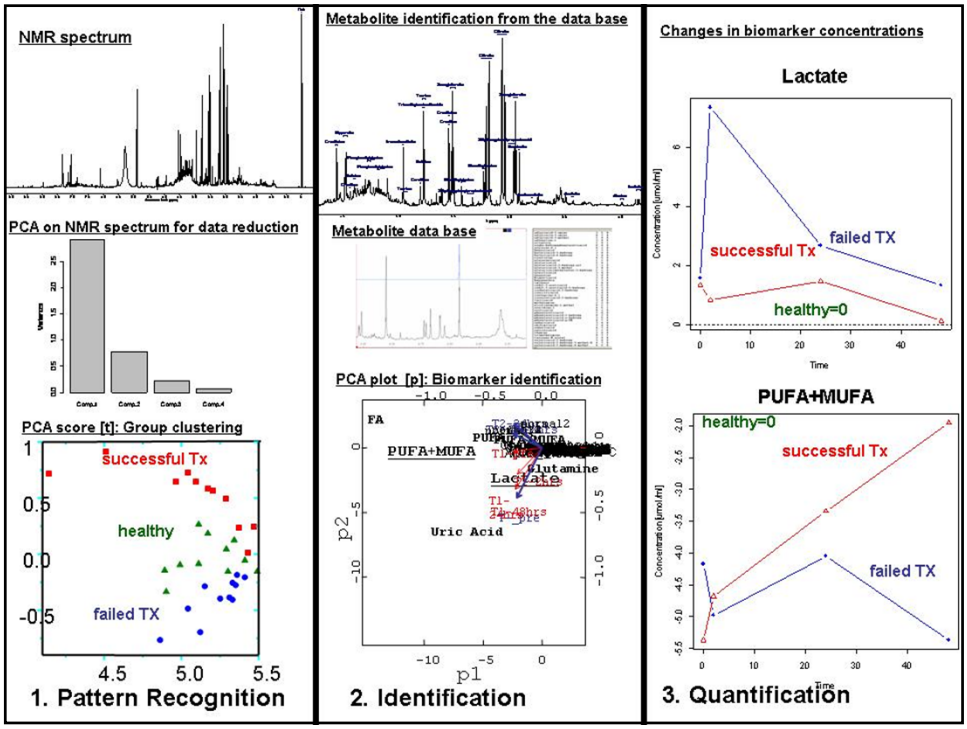
Multivariate Statistics

- Multivariate means multiple variables
- If you measure a population using multiple measures at the same time such as height, weight, hair colour, clothing colour, eye colour, etc. you are performing multivariate statistics
- Multivariate statistics requires more complex, multidimensional analyses or dimensional reduction methods

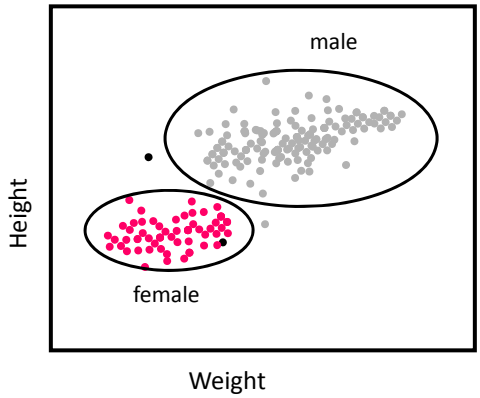
Multivariate Statistics – The Trick

Metabolomics requires multivariate analysis

- The key trick in multivariate statistics is to find a way that effectively reduces the multivariate data into univariate data
- Once done, then you can apply the same univariate concepts such as p-values, t-Tests and ANOVA tests to the data
- The trick is dimensional reduction



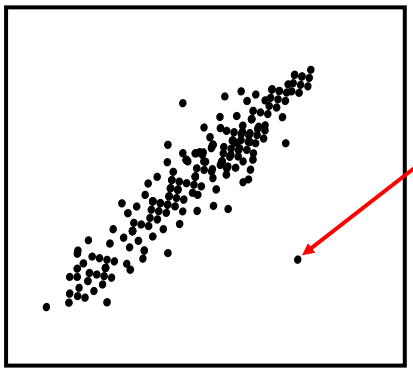
Multivariate Analysis: Group Separation



Multivariate Analysis: Outliers

- Can be both "good" and "bad"
- When modeling data -- you don't like to see outliers (suggests the model is bad)
- Often a good indicator of experimental or measurement errors -- only you can know!
- When plotting metabolite concentration data you do like to see outliers
- A good indicator of something significant

Multivariate Analysis: Outliers

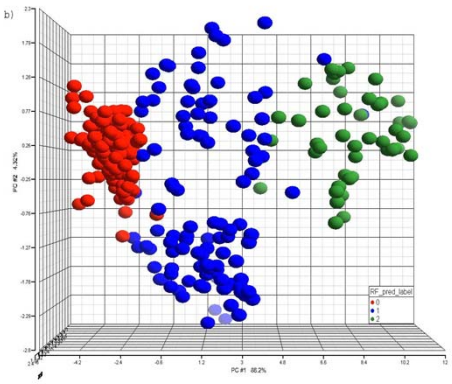


A single "bad" point can destroy a good correlation

Multivariate Analysis: Outliers

- Dealing with outliers
  - Detected mainly by visual inspection
  - May be corrected by normalization
  - May be excluded
- Noise reduction
  - More of a concern for spectral bins/ peak lists
  - Usually improves downstream results

Dimension Reduction & PCA

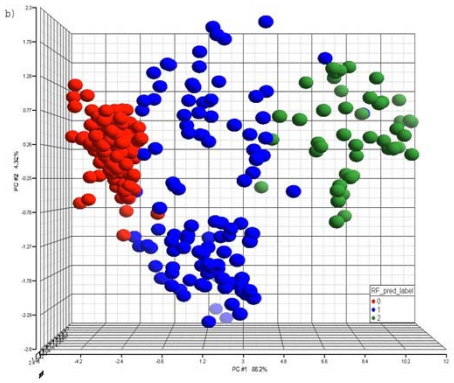


Scores plot

- PCA – Principal Componenent Analysis
- Process that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components
- Reduces 1000's of variables to 2-3 key features

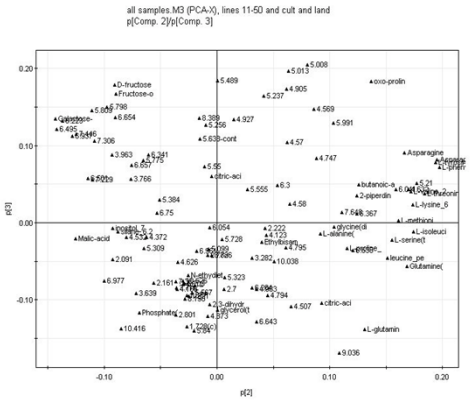
PCA Plot Nomenclature

- PCA Generate 2 kinds of plots, the scores plot and the loadings plot
- Scores plot (on right) plots the data using the main principal components



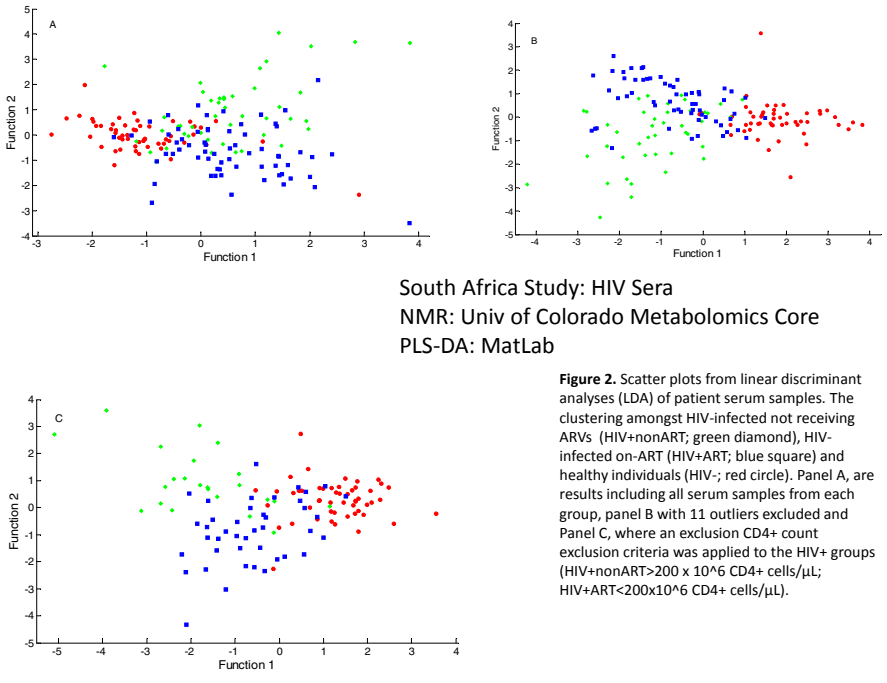
PCA Loadings Plot

- Loadings plot shows how much each of the variables (metabolites) contributed to the different principal components
- Variables at the extreme corners contribute most to the scores plot separation



PCA vs. PLS-DA

- Partial Least Squares Discriminant Analysis
- PLS-DA is a supervised classification technique while PCA is an unsupervised clustering technique
- PLS-DA uses "labeled" data while PCA uses no prior knowledge
- PLS-DA enhances the separation between groups of observations by rotating PCA components such that a maximum separation among classes is obtained



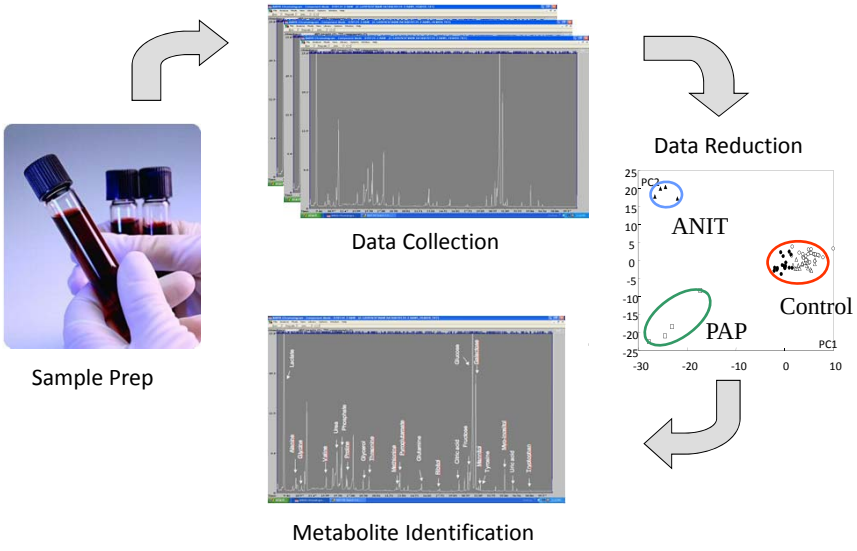
South Africa Study: HIV Sera
NMR: Univ of Colorado Metabolomics Core
PLS-DA: MatLab

Figure 2. Scatter plots from linear discriminant analyses (LDA) of patient serum samples. The clustering amongst HIV-infected not receiving ARVs (HIV+nonART; green diamond), HIV-infected on-ART (HIV+ART; blue square) and healthy individuals (HIV-; red circle). Panel A, are results including all serum samples from each group, panel B with 11 outliers excluded and Panel C, where an exclusion CD4+ count exclusion criteria was applied to the HIV+ groups (HIV+nonART>200 x 10^6 CD4+ cells/µL; HIV+ART<200x10^6 CD4+ cells/µL).

Other Supervised Classification Methods

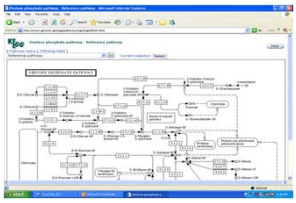
- SIMCA – Soft Independent Modeling of Class Analogy
- OPLS – Orthoganol Projection of Latent Structures
- Support Vector Machines
- Random Forest
- Naïve Bayes Classifiers
- Neural Networks

Profiling (Untargeted)

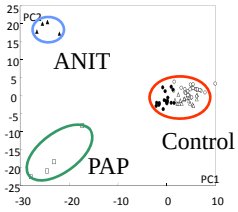


Quantitative (Targeted)

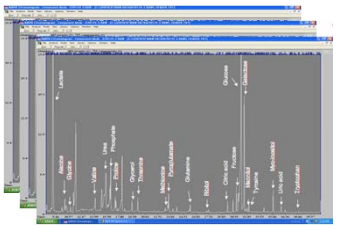
Sample Prep



Biological Interpretation

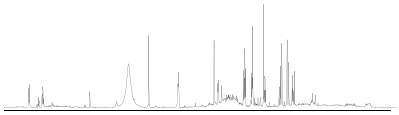


Data Reduction



Metabolite Identification & Quantification

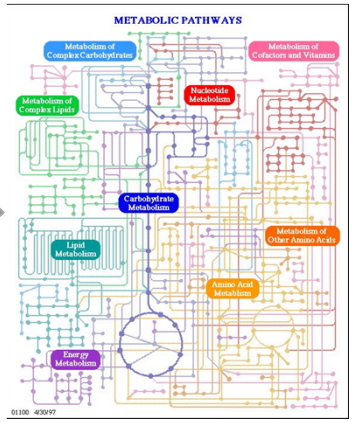
From Spectra to Lists



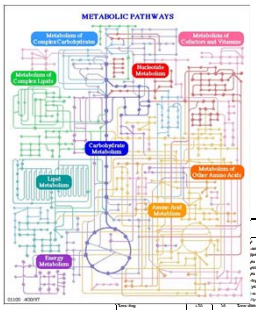
Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns -o-phospho -L-serine	0.92	<D.L. *	**Dns-Ile**	6.35	25
Dns -o-phospho -L-tyrosine	0.95	<D.L.	**Dns-3-aminosalicylic acid**	6.44	0.5
Dns -adnosine monophosphate	0.99	<D.L.	**Dns-pipecolic acid**	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	**Dns-Leu**	6.54	54
Dns-glucosamine	1.06	22	**Dns-cystathionine**	6.54	0.3
Dns -o-phospho -L-threonine	1.09	<D.L.	**Dns-Leu-Pro**	6.60	0.4
Dns -6-dimet hylamine purine	1.20	<D.L.	**Dns-5-hydroxylysine**	6.65	1.6
Dns-3-methyl -histidine	1.22	80	**Dns-Cystine**	6.73	160
Dns-taurine	1.25	834	**Dns-N-norleucine**	6.81	0.1
Dns-carnosine	1.34	28	Dns -5-hydroxydopamine	7.17	<D.L.
Dns-Arg	1.53	36	**Dns-dimethylamine**	7.33	293
Dns-Asn	1.55	133	**Dns-5-HIAA**	7.46	18
Dns-hypotaurine	1.58	10	**Dns-umbelliferone**	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns -2,3 -diaminopropionic acid	7.63	<D.L.
Dns -guanidine	1.62	<D.L.	**Dns-L-ornithine**	7.70	15
Dns-Gln	1.72	633	**Dns-4-acetyaminiophenol**	7.73	51
Dns-allantoin	1.83	3.8	**Dns-procaine**	7.73	8.9
Dns-L-citrulline	1.87	2.9	**Dns-homocystine**	7.76	3.3
Dns-1 (or 3 -)-methylhistamine	1.94	1.9	**Dns-acetaminophen**	7.97	82
Dns-adenosine	2.06	2.6	**Dns-Phe-Phe**	8.03	0.4
Dns -methylguanidine	2.20	<D.L.	**Dns-5-methyo xysalicylic acid**	8.04	2.1
Dns-Ser	2.24	511	**Dns-Lys**	8.16	184
Dns-aspartic acid amide	2.44	26	Dns -urdine	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	**Dns-leu-Phe**	8.22	0.3
Dns-Glu	2.57	21	**Dns-His**	8.35	1550
Dns-Asp	2.60	90	Dns -4-thialysine	8.37	<D.L.
Dns-Thr	3.03	157	Dns -benzylamine	8.38	<D.L.
Dns -epinephrine	3.05	<D.L.	**Dns-1-ephedrine**	8.50	0.6
Dns-ethanolamine	3.11	471	**Dns-tryptamine**	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns -pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns -2 -methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.88	593	**Dns-5-hydroxyrptpphan**	9.25	0.12
Dns-aminolevulinic acid	3.97	30	**Dns-1,3 -diaminopropane**	9.44	0.23
Dns-r-amino -butyric acid	3.98	4.6	**Dns-putrescine**	9.60	0.5
Dns-p- amino -hippuric acid	3.98	2.9	**Dns-1,2 -diaminopropane**	9.66	0.1
Dns-5-hydro xymethyluricil	4.58	1.9	**Dns-tyrosinamide**	9.79	29
Dns-tryptophanamide	4.70	5.5	**Dns-dopamine**	10.08	140
Dns -isoguanine	4.75	<D.L.	**Dns-cadaverine**	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	**Dns-histamine**	10.19	0.4
Dns-sarcosine	4.81	7.2	**Dns-3-methoxy -tyramine**	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	**Dns-Tyr**	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns -cysteamine	10.44	<D.L.

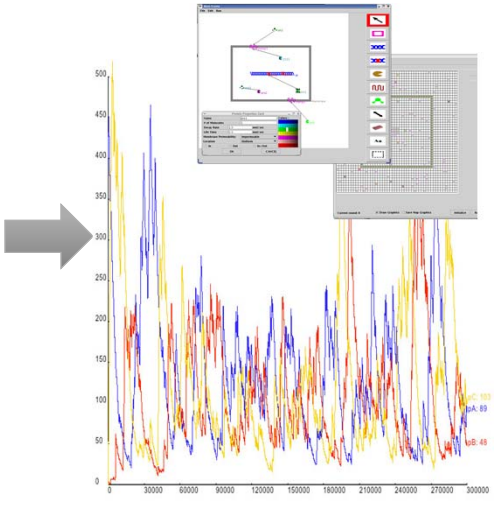
From Lists to Pathways

Compound	Retention Time (min)	Conc. in Urine (µM)	Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho -L-serine	0.92	<D.L. *	**Dns-Ile**	6.35	25
Dns-o-phospho -L-tyrosine	0.95	<D.L.	**Dns-3-aminosalicylic acid**	6.44	0.5
Dns -adnosine monophosphate	0.99	<D.L.	**Dns-pipecolic acid**	6.50	0.5
Dns-o-phosphoethanolamine	1.06	16	**Dns-Leu**	6.54	54
Dns-glucosamine	1.06	22	**Dns-cystathionine**	6.54	0.3
Dns-o-phospho -L-threonine	1.09	<D.L.	**Dns-Leu-Pro**	6.60	0.4
Dns-6-dimet hylamine purine	1.20	<D.L.	**Dns-5-hydroxylysine**	6.65	1.6
Dns-3-methyl -histidine	1.22	80	**Dns-Cystine**	6.73	160
Dns-taurine	1.25	834	**Dns-N-norleucine**	6.81	0.1
Dns-carnosine	1.34	28	Dns-5-hydroxydopamine	7.17	<D.L.
Dns-Arg	1.53	36	**Dns-dimethylamine**	7.33	293
Dns-Asn	1.55	133	**Dns-5-HIAA**	7.46	18
Dns-hypotaurine	1.58	10	**Dns-umbelliferone**	7.47	1.9
Dns-homocarnosine	1.61	3.9	Dns-2,3 -diaminopropionic acid	7.63	<D.L.
Dns -guanidine	1.62	<D.L.	**Dns-L-ornithine**	7.70	15
Dns-Gln	1.72	633	**Dns-4-acetyamidophenol**	7.73	51
Dns-allantoin	1.83	3.8	**Dns-procaine**	7.73	8.9
Dns-L -citrulline	1.87	2.9	**Dns-homocystine**	7.76	3.3
Dns-1 (or 3 -)-methylhistamine	1.94	1.9	**Dns-acetaminophen**	7.97	82
Dns-adenosine	2.06	2.6	**Dns-Phe-Phe**	8.03	0.4
Dns -methylguanidine	2.20	<D.L.	**Dns-5-methyo xysalicylic acid**	8.04	2.1
Dns-Ser	2.24	511	**Dns-Lys**	8.16	184
Dns-aspartic acid amide	2.44	26	Dns -aniline	8.17	<D.L.
Dns-4-hydroxy -proline	2.56	2.3	**Dns-leu-Phe**	8.22	0.3
Dns-Glu	2.57	21	**Dns-His**	8.35	1550
Dns-Asp	2.60	90	Dns -4-thialysine	8.37	<D.L.
Dns-Thr	3.03	157	Dns -benzylamine	8.38	<D.L.
Dns -epinephrine	3.05	<D.L.	**Dns-1-ephedrine**	8.50	0.6
Dns-ethanolamine	3.11	471	**Dns-tryptamine**	8.63	0.4
Dns-aminoadipic acid	3.17	70	Dns -pyridoxamine	8.94	<D.L.
Dns-Gly	3.43	2510	Dns -2 -methyl -benzylamine	9.24	<D.L.
Dns-Ala	3.88	593	**Dns-5-hydroxytryptophan**	9.25	0.12
Dns-aminolevulinic acid	3.97	30	**Dns-1,3 -diaminopropane**	9.44	0.23
Dns-r -amino -butyric acid	3.98	4.6	**Dns-putrescine**	9.60	0.5
Dns-p -amino -hippuric acid	3.98	2.9	**Dns-1,2 -diaminopropane**	9.66	0.1
Dns-5-hydro xymethyluricil	4.58	1.9	**Dns-tyrosinamide**	9.79	29
Dns-tryptophanamide	4.70	5.5	**Dns-dopamine**	10.08	140
Dns -isoguanine	4.75	<D.L.	**Dns-cadaverine**	10.08	0.08
Dns-5-aminopentanoic acid	4.79	1.6	**Dns-histamine**	10.19	0.4
Dns-sarcosine	4.81	7.2	**Dns-3-methoxy -tyramine**	10.19	9.2
Dns-3-amino -isobutyrate	4.81	85	**Dns-Tyr**	10.28	321
Dns-2-aminobutyric acid	4.91	17	Dns -cysteamine	10.44	<D.L.

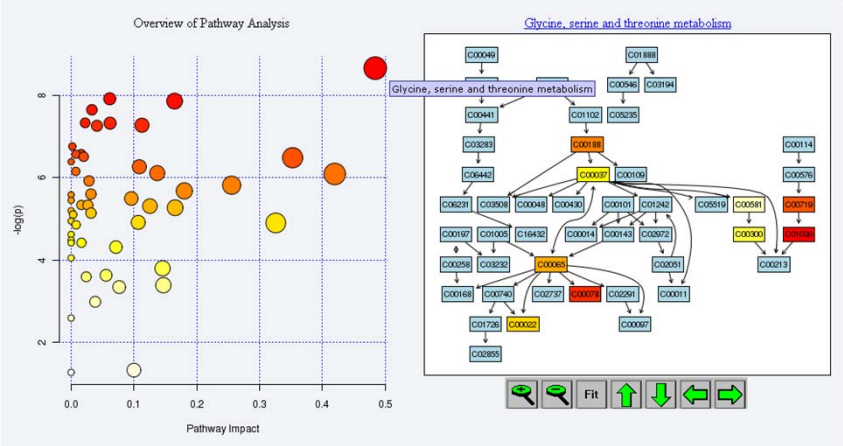


From Pathways & Lists to Models & Biomarkers

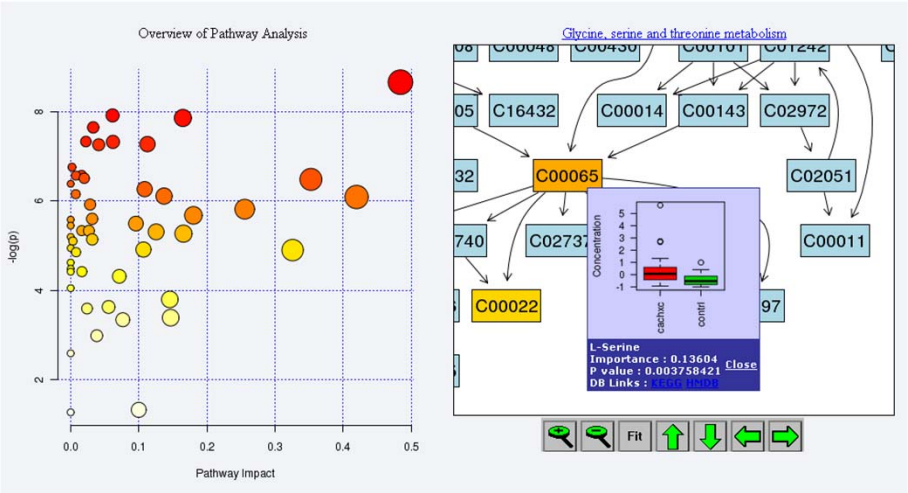




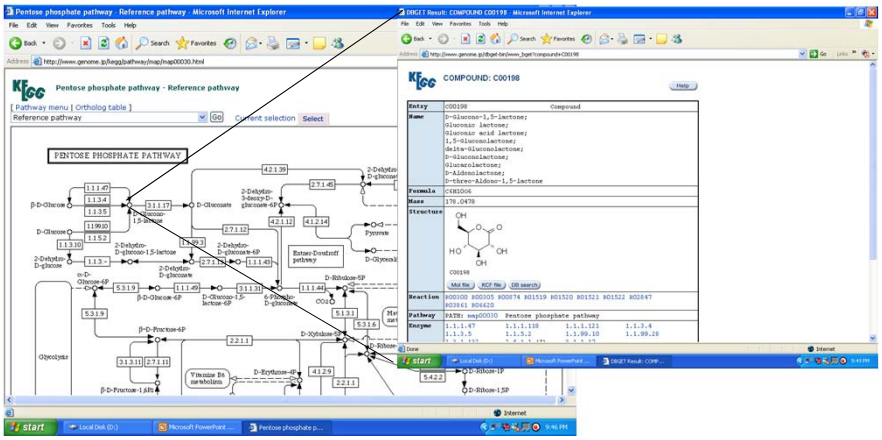
Pathway Visualization



Pathway Visualization



KEGG – Kyoto Encyclopedia of Genes and Genomes



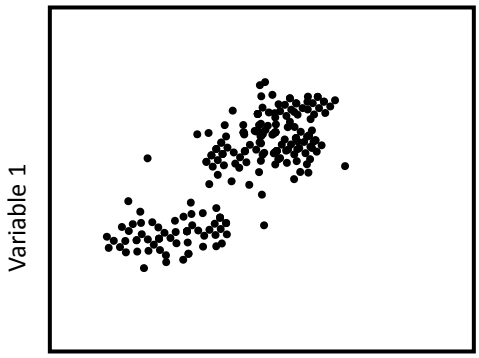
http://www.genome.jp/kegg/

Final Validation: Student's t-Test

- Also called the t-Test
- Used to determine if 2 populations are different
- Formally allows you to calculate the probability that 2 sample means are the same
- If the t-Test statistic gives you a p=0.4, and the α is 0.05, then the 2 populations are the same
- If the t-Test statistic gives you a p=0.04, and the α is 0.05, then the 2 populations are different
- Paired and unpaired t-Tests are available, paired if used for "before & after" expts. while unpaired is for 2 randomly chosen samples

Student's t-Test

- A t-Test can also be used to determine whether 2 clusters are different if the clusters follow a normal distribution

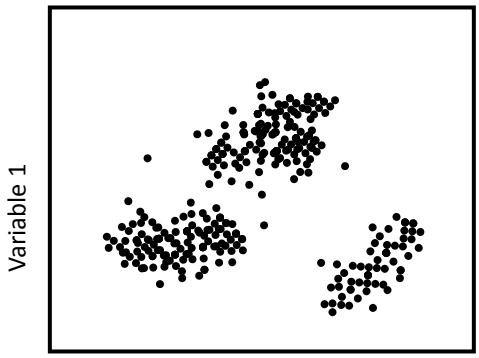


ANOVA

- Also called Analysis of Variance
- Used to determine if 3 or more populations are different, it is a generalization of the t-Test
- Formally ANOVA provides a statistical test (by looking at group variance) of whether or not the means of several groups are all equal
- Uses an F-measure to test for significance
- 1-way, 2-way, 3-way and n-way ANOVAs, most common is 1-way which just is concerned about whether any of the 3+ populations are different, not which pair is different

ANOVA

- ANOVA can also be used to determine whether 3+ clusters are different if the clusters follow a normal distribution



Biomarker Definition (NIH):

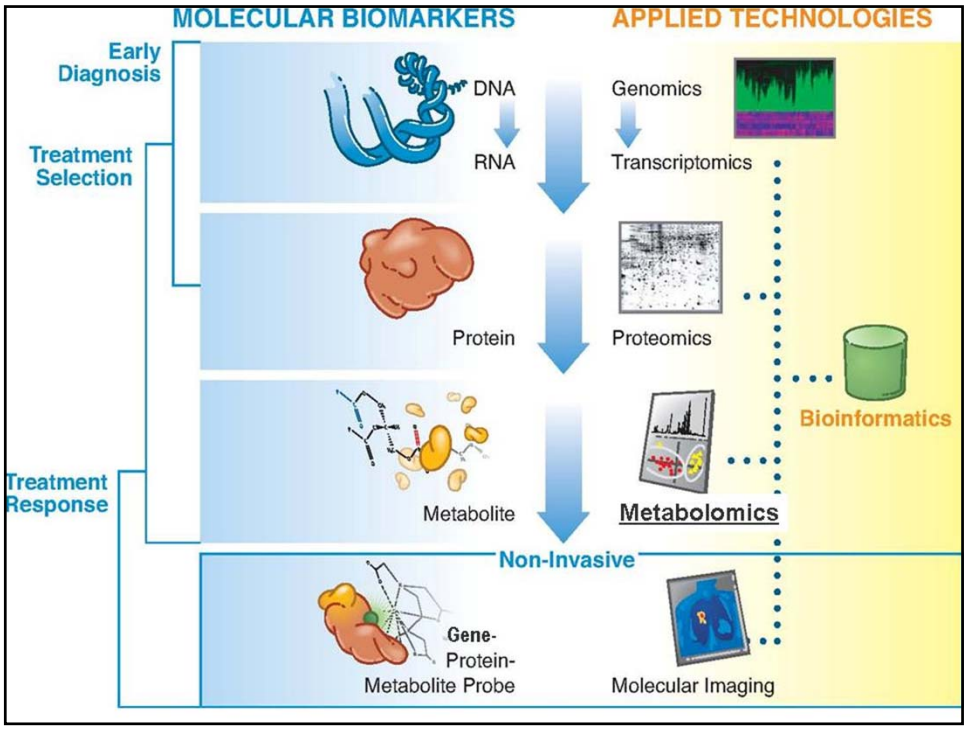
A characteristic that is objectively measured and evaluated as an:

- Indicator of normal biological process;
- Pathologic process;
- Responses to therapeutic intervention (serum/ blood, body fluids, tissue)

Sensitivity and Specificity!

Types of Biomarkers

- Diagnostic
  - Do you have a given disease/condition
- Prognostic
  - How well will you do with this disease/condition
- Predictive
  - Odds of getting a given disease/condition
- Marker of Response or Toxicity
  - Response to drug/food/toxin intake
- Marker of Exposure
  - Indication of drug/toxin/food consumption



Biomarker Statistics

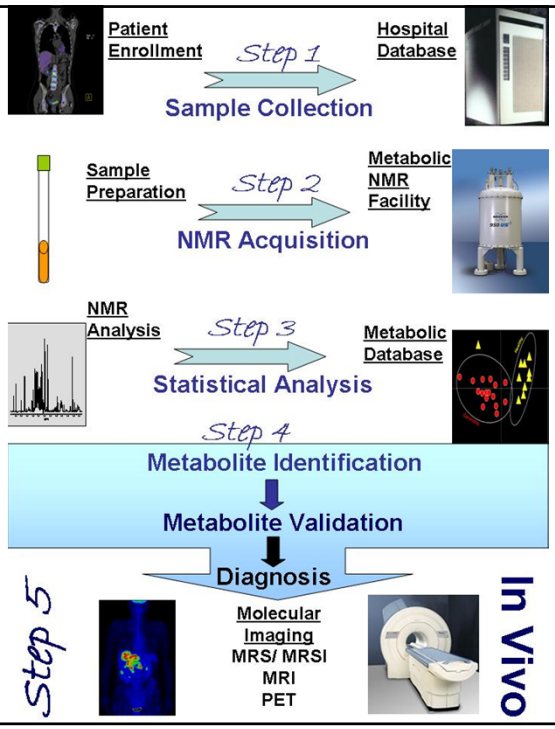
- Number of "approved" tests arising from or using Genomics – 2027 (5% of diseases)
- Number of tests arising from or using Transcriptomics – 5 (AmpliChip CYP450, MammaPrint, TargetPrint, BluePrint, ColoPrint)
- Number of "approved" tests arising from or using Proteomics - 0
- Number of "approved" tests arising from or using Metabolomics/Clinical chemistry – 150 (80% of diseases)

Examples

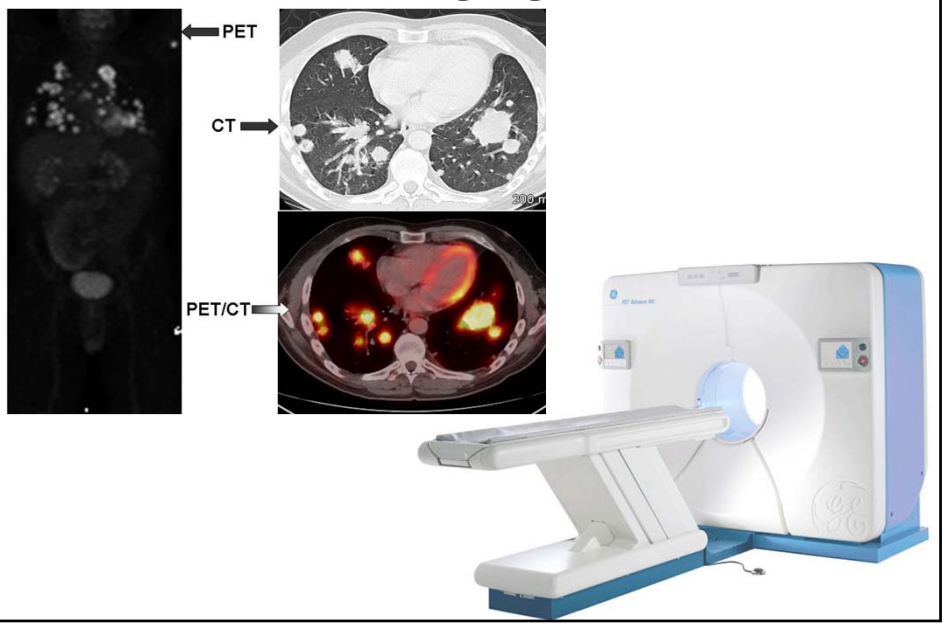
Disease	Metabolite	Cure or Treatment
Goiter/Cretinism	Low Iodine	Iodized salt
PKU	High Phenylalanine	Low Phe diet
Epilepsy	Low Ketone Bodies	Ketogenic diet
Glycogen Storage Disease I	Low Glucose	Corn starch
Lactic acidosis (Genetic)	High Lactate	Thiamine supp.

Examples

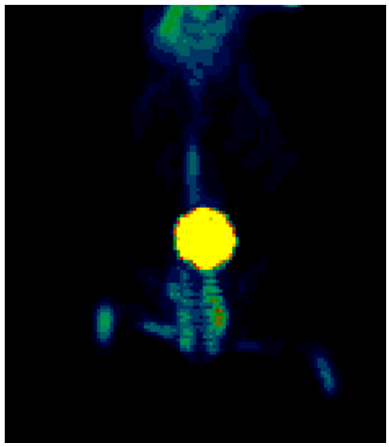
Disease	Metabolite	Cure or Treatment
Hypertension	High Salt, Low Vitamin D	Low salt diet, Vitamin D supplements
Hypercholesterolemia Atherosclerosis	High Cholesterol	Low cholesterol diet, Niacin
Diabetes	High Glucose	Insulin, Low carbohydrate diet
Gout	High Uric Acid	Vit. C, Xanthine analogues
Fatty Liver Disease	Low Choline	Choline supp.



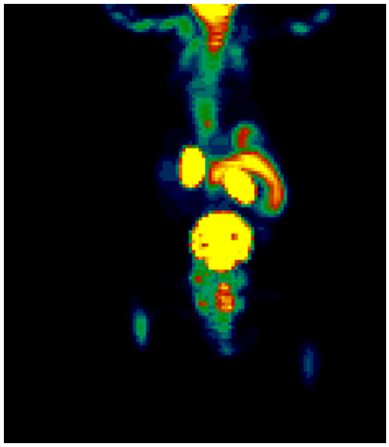
Metabolic Imaging: MRSI, PET



FDG-PET:
Increased Glucose Uptake/ Oncology



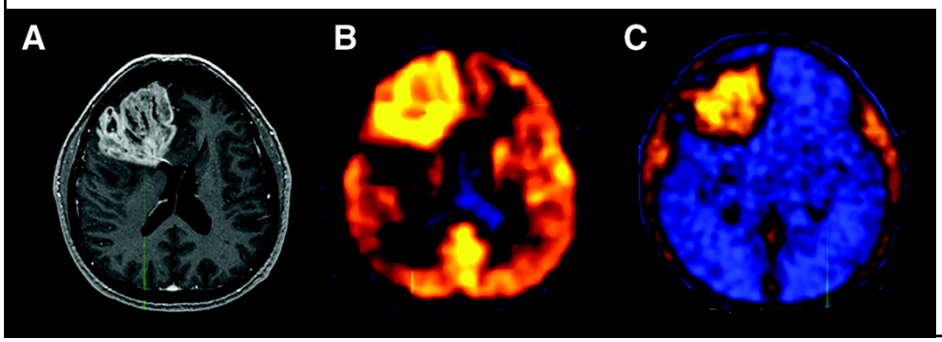
Control WT

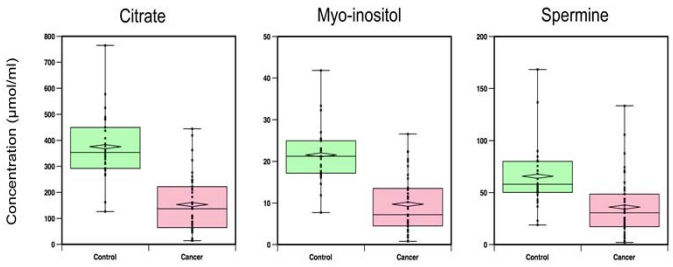


Leukemic Tg

FLT (18F-Thymidine) PET: Increased Proliferation/ DNA Synthesis

MRI FDG-PET FLT-PET

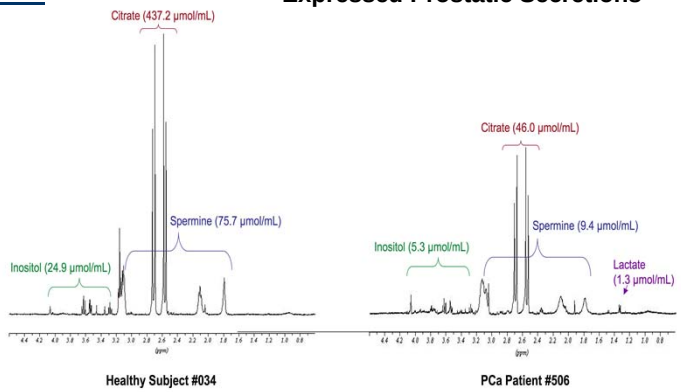




SPECIFIC MARKERS

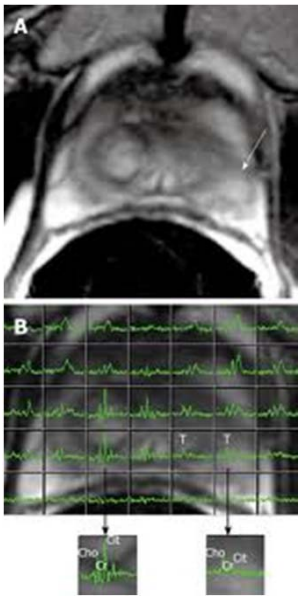
Expressed Prostatic Secretions

Citrate for Prostate Cancer

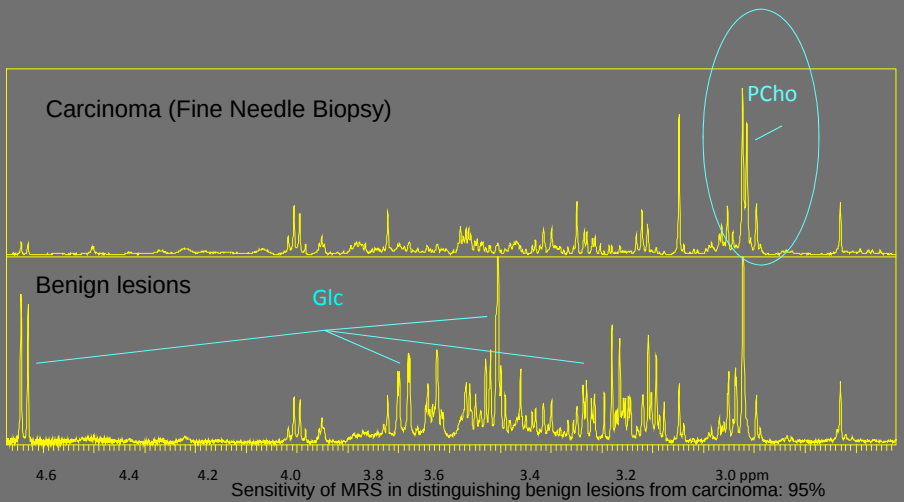


Healthy Subject #034 PCa Patient #506

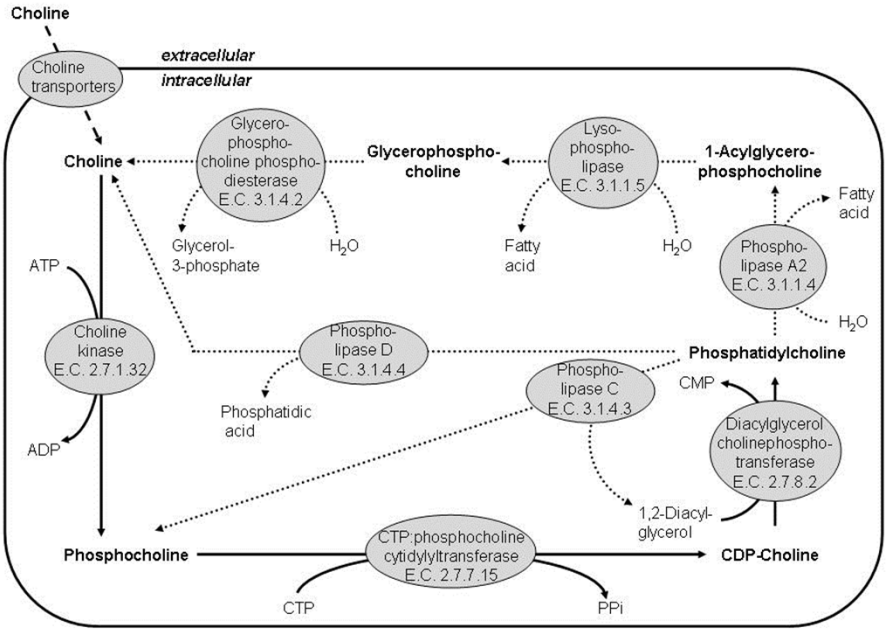
Citrate MRSI: Prostate Cancer



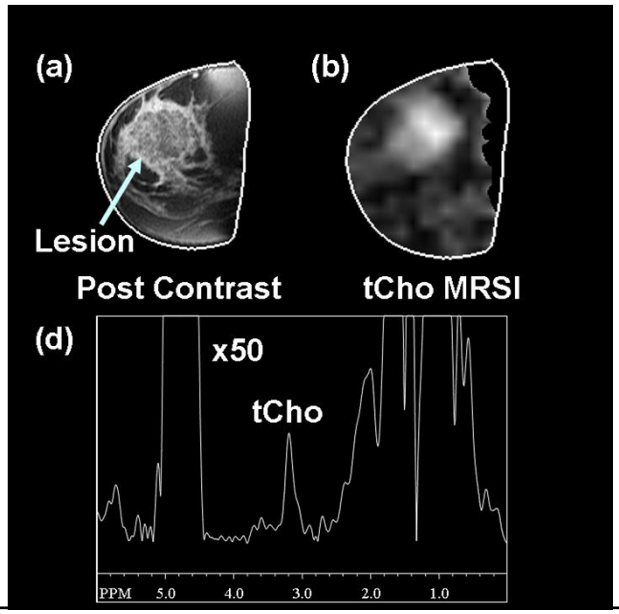
Choline NMR: Breast Cancer



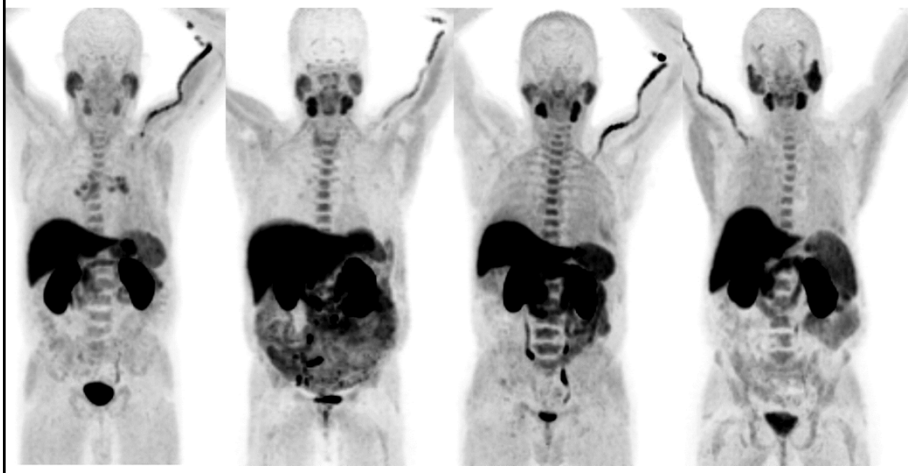
Sensitivity of MRS in distinguishing benign lesions from carcinoma: 95%
Specificity of MRS in distinguishing benign lesions from carcinoma: 96%



Choline MRSI: Breast Cancer



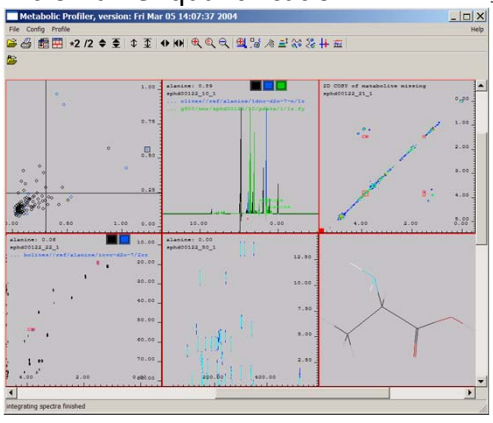
^{18}F-Choline PET



Summary

^{1}H-MRS of biofluids and tissues is useful for identification of metabolic pathways and biomarkers:

- metabolic pattern recognition,
- biomarker identification,
- biomarker quantification.



$$Cx = \frac{Ix : Nx \times C}{I : 9} \times V : M$$