Overview on Metabolomics

Josephine   **Linke**   **Yibeltal**

"Science is built up with facts, as a house is with stones. But a collection of f acts is no more a science than a heap of stones is a house." - Jules Henri Poincaré

Definitions

- ## Metabolomics

  - Newly emerging field of 'omics' research

  - Comprehensive and simultaneous systematic determination of metabolite levels in the metabolome and their changes over time as a consequence of stimuli

- ## Metabolome

  - Refers to the complete set of small-molecule metabolites

  - Dynamic

- ## Metabolites

  - Intermediates and products of metabolism

  - Examples include antibiotics, pigments, carbohydrates, fatty acids and amino acids

  - Primary and secondary metabolites

History

- 2000-1500 BC

- The first paper was titled, "Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography", by Robinson and Pauling in 1971.

- The name metabolomics was coined in the late 1990s (the first paper using the word metabolome is Oliver, S. G., Winson, M. K., Kell, D. B. & Baganz, F. (1998). Systematic functional analysis of the yeast genome.

- Many of the bioanalytical methods used for metabolomics have been adapted (or in some cases simply adopted) from existing biochemical techniques.

- Human Metabolome project – first draft of human metabolome in 2007

Data gathering

- Four main points in Analysis of metabolomics data :

  - Efficient and unbiased

  - Separation of analytes

  - Detection

  - Identification and quantification

Data gathering

- Separation Techniques

  - Gas Chromatography (GC)

  - Capillary Electrophoresis (CE)

  - High Performance Liquid Chromatography (HPLC)

  - Ultra Performance Liquid Chromatography (UPLC)

- Combination of Techniques

  - GC-MS

  - HPLC-MS

- Detection Techniques

  - Nuclear Magnetic Resonance Spectroscopy (NMR)
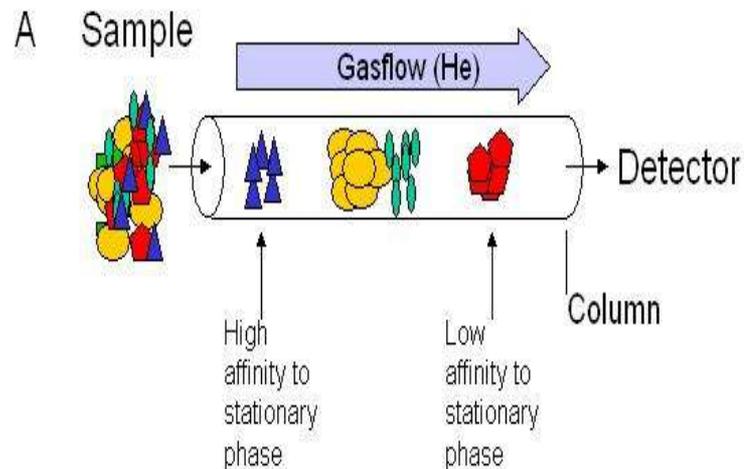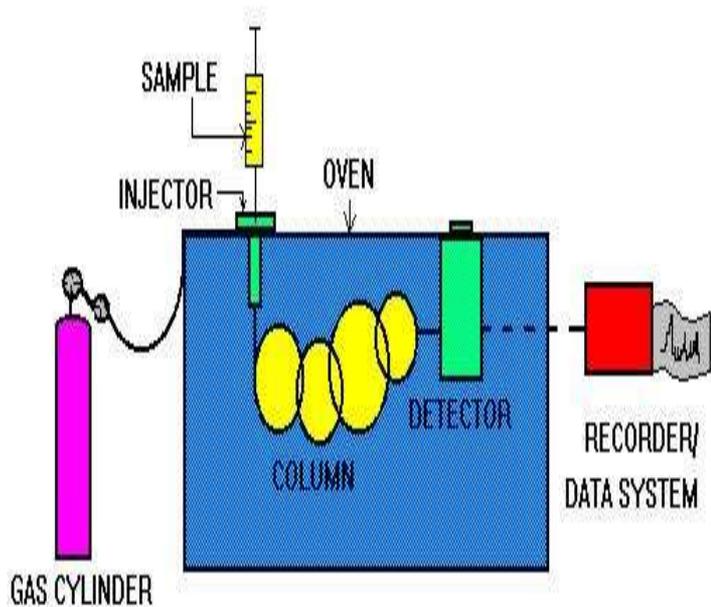
  - Mass Spectrometry (MS)

Seperation Technique - GC

- Mostly in Organic Chemistry

- High Chromatographic resolution

- Require chemical derivatization

- Mobile and stationary phase

- Alternative names
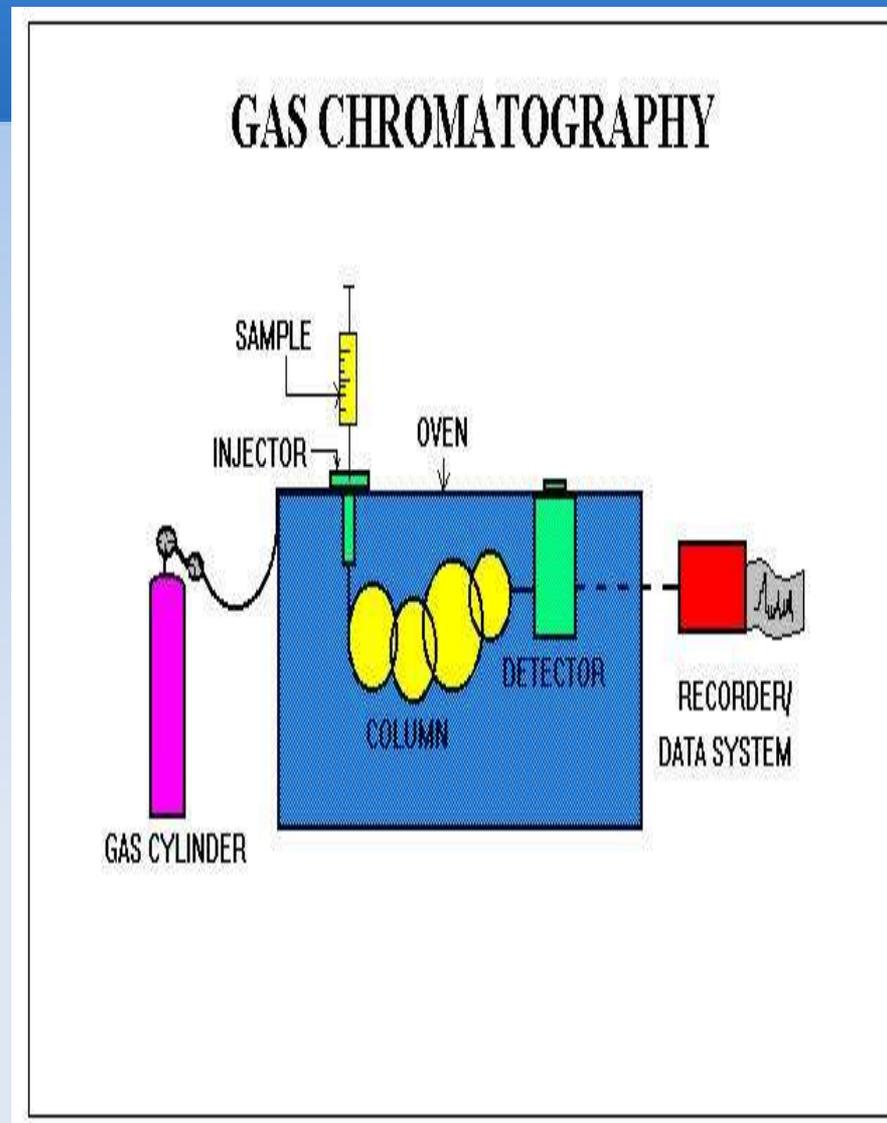
Seperation Technique - GC
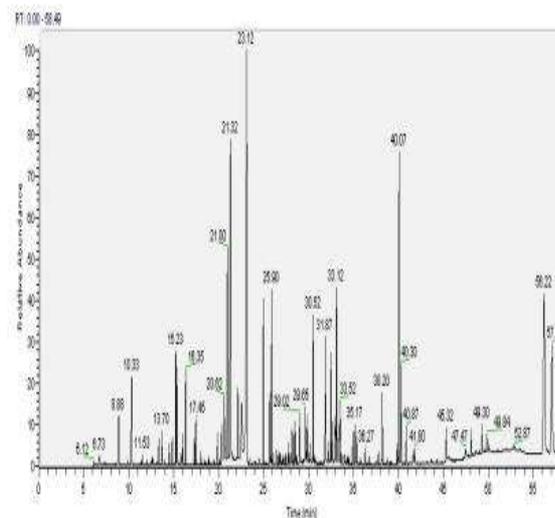
Seperation Technique - HPLC

- Biochemistry and analytical chemistry
- Lower chromatographic resolution
- Wide range analytes
- Mobile and stationary phase
- Retention time

solvent reservoir

pump to produce high pressure

processing unit and display

signal to processor

sample injection

HPLC tube

detector

waste
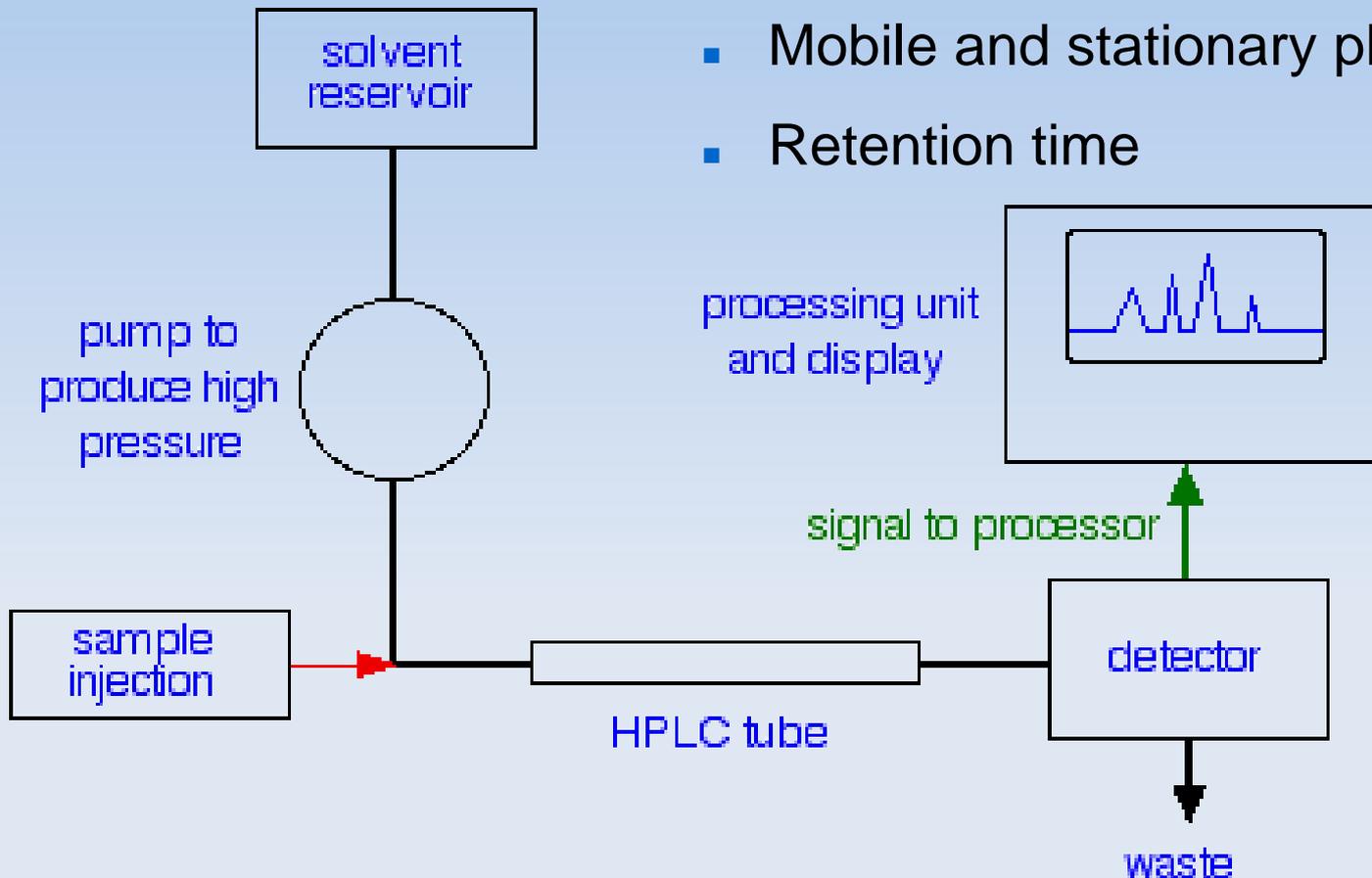
HPLC compared to UPLC
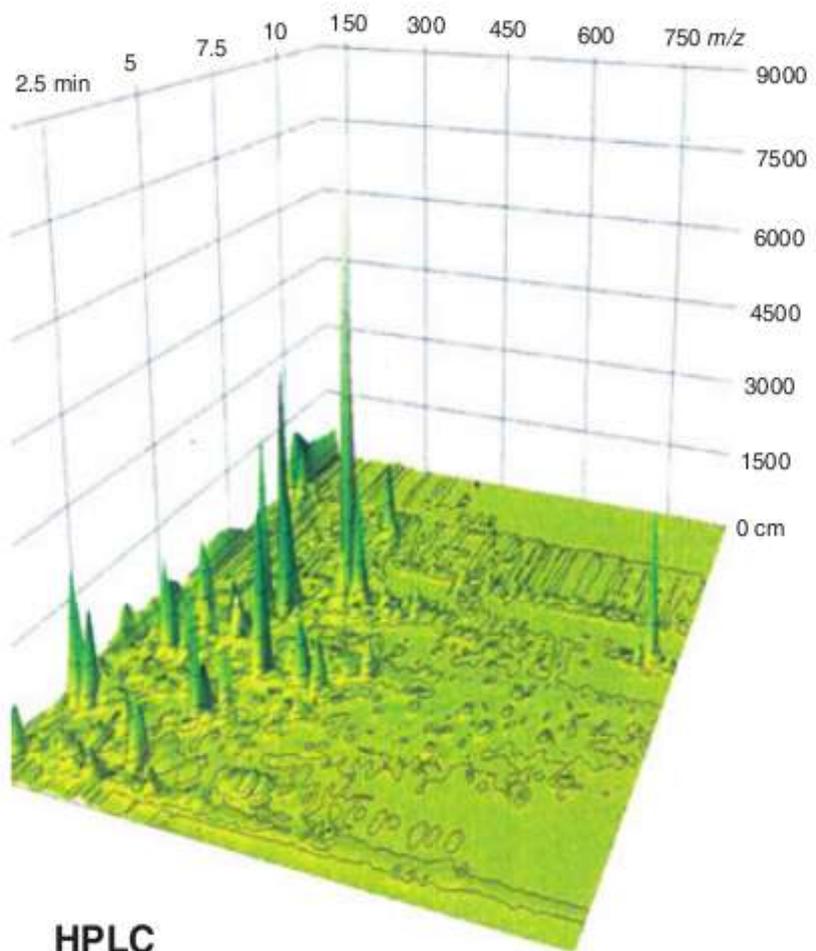
HPLC
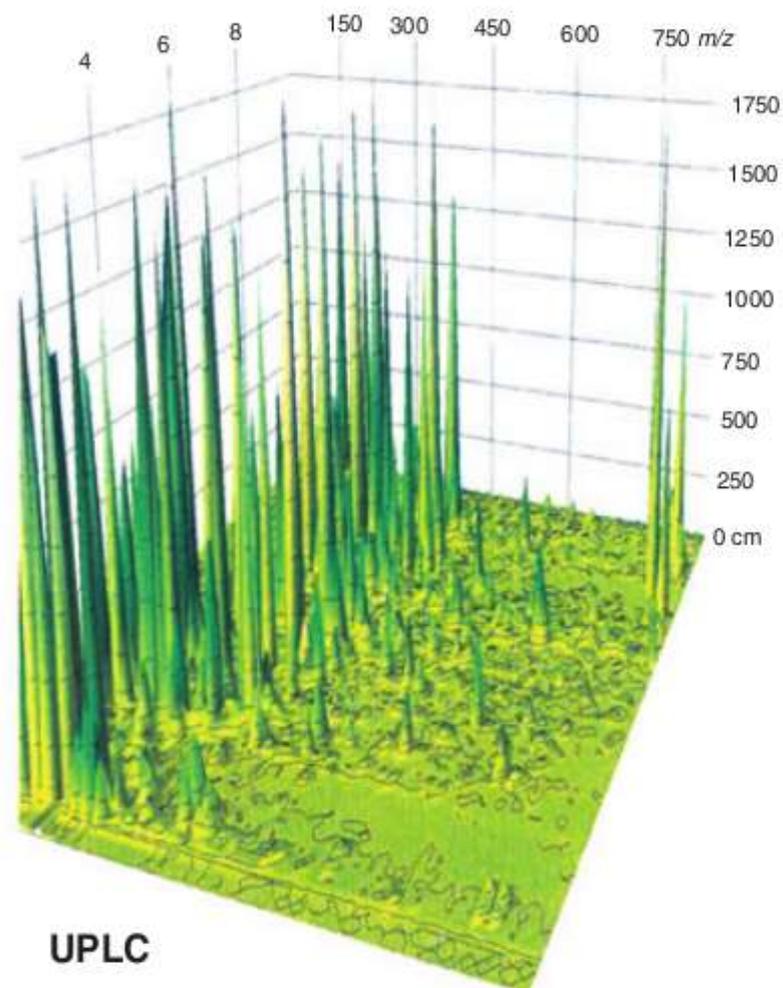
UPLC

Seperation Technique - CE

- Introduced in 1960s
- Higher separation efficiency than HPLC
- Wide range of metabolites than GC
- Charged analytes

Detection Technique - NMRS

- Doesn't depend on separation

- Relatively insensitive

- NMR spectra difficult for interpretation

- Applicable in MRI

NMR Experiment

- A current through (green)

- generates a strong magnetic field

- polarizes the nuclei in the sample material (red).

- It is surrounded by the r.f. coil (black)

- delivers the computer generated r.f. tunes that initiate the nuclear quantum dance.

- At some point in time, the switch is turned and now the dance is recorded through the voltage it induces.

- the NMR signal, in the r.f. coil.

- The signals Fourier transform (FT) shows "lines" for different nuclei in different electronic environments.
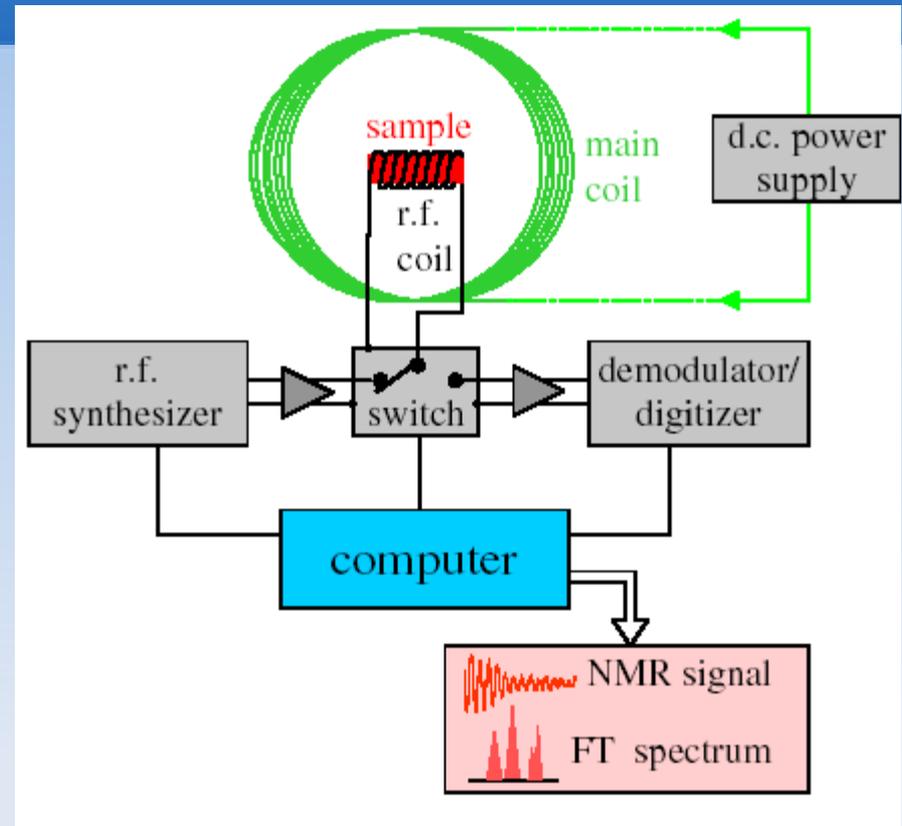
Detection Technique - NMR

- A typical 950-MHz H NMR spectrum of urine showing the degree of spectral complexity
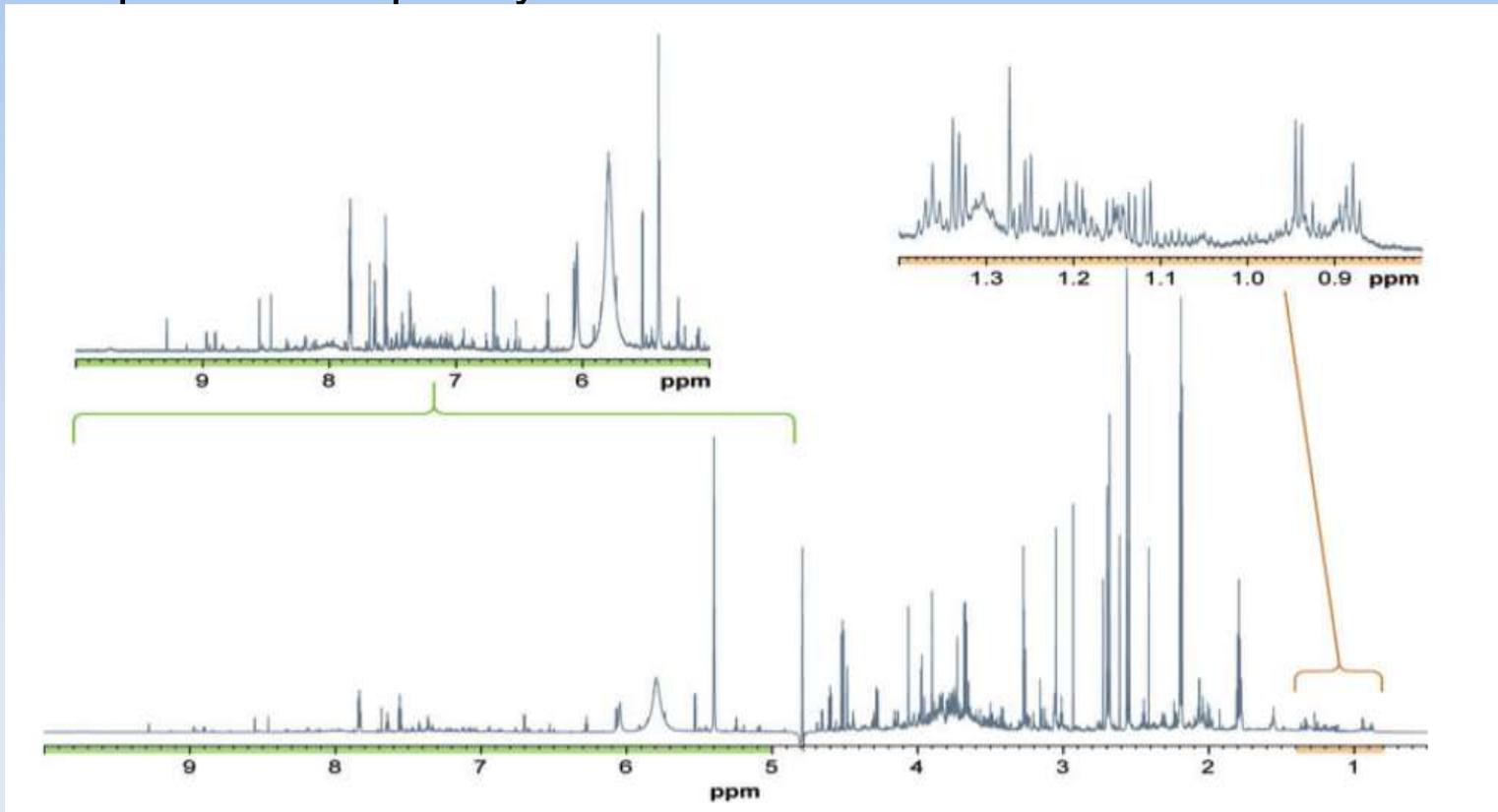
Detection Technique - MS

- To identify and to quantify metabolites

- Serves to both separate and to detect

- Mass to charge ratios

- Using electron beam

- Ion source, mass analyzer and detector

The relative strengths and weaknesses of nuclear magnetic resonance and mass spectrometry for metabolic profiling[a]

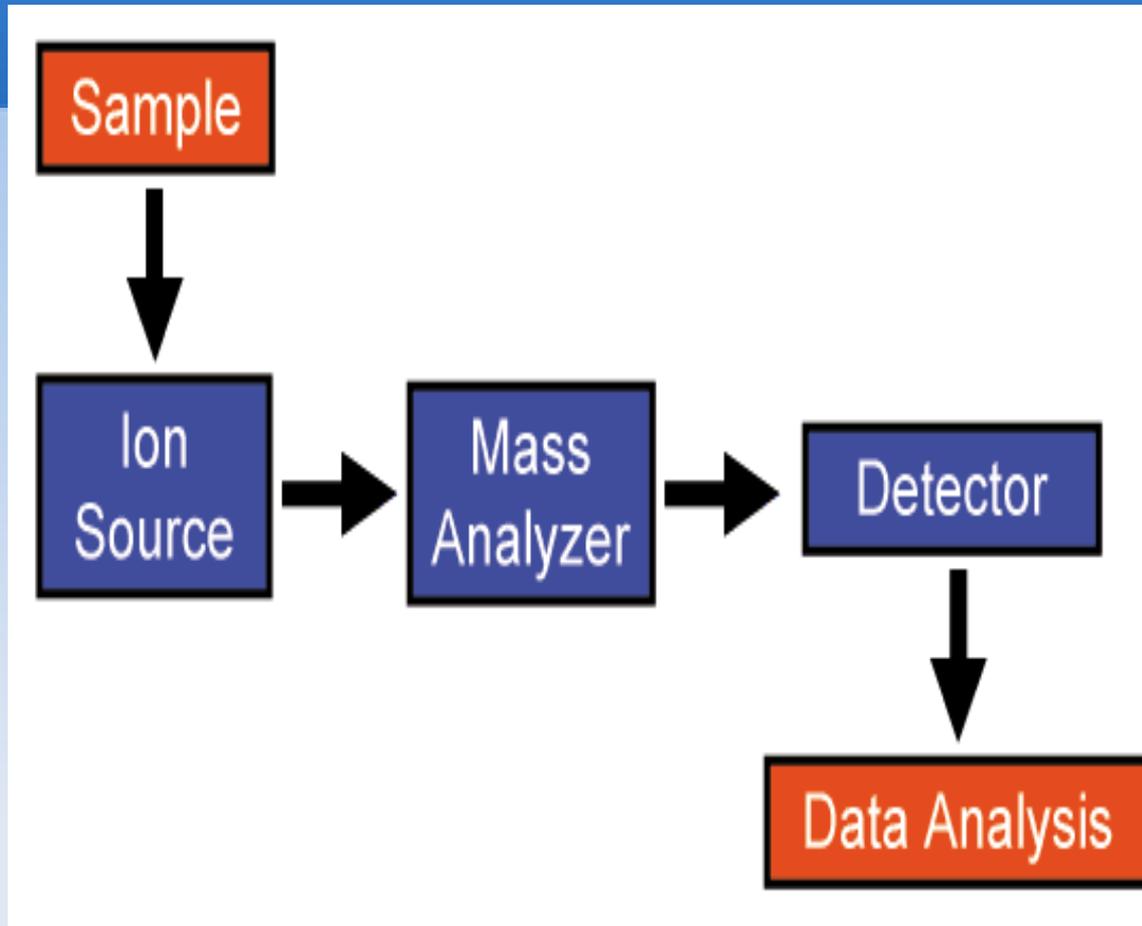|	NMR	MS
Detection limits	Low-micromolar at typical observation frequencies (600 MHz), but nanomolar using cryoprobes	Picomolar with standard techniques, but can be much lower with special techniques
Universality of metabolite detection	If metabolite contains hydrogens it will be detected, assuming the concentration is sufficient or protein binding does not cause marked line broadening	Usually needs a more targeted approach. There can be problems with poor chromatographic separation; with the loss of metabolites in void volumes; with ion suppression (but this is reduced when using UPLC); lack of ionization; ability to run both +ve and −ve ion detection gives extra information
Sample handling	Whole sample analyzed in one measurement	Different LC packings and conditions for different classes of metabolite; usually samples have to be extracted into a suitable solvent; samples have to be aliquoted but some recent studies have avoided the need for chromatography
Amount of sample used	Typically 200–400 μL, but much less for microcoil probes, down to 5–10 μL	Low μL range
Sample recovery	Technique is nondestructive	Technique is destructive but only small amounts used
Analytical reproducibility	Very high	Fair
Sample prepreparation	Minimal: addition of buffer, D_2O and chemical shift reference (not always required)	Can be substantial; often needs different LC columns and protein precipitation
Ease of molecular identification	High, both from databases of authentic material and by self-consistent analysis of 1D and 2D spectra	Difficult, often only the molecular ion is available; this needs xtra experiments, such as routine tandem MS; GC-MS is generally better with accurate retention times and comprehensive databases of spectra
Time to collect basic data	5 min for 1D ^1H NMR	10 min for UPLC-MS run
Quantitation	1–5%	5% intraday and interday is now common with or without prior chromatography
Robustness of instruments	High	Low
Molecular dynamics information	Yes, from T1, T2 relaxation time and diffusion coefficient measurements	No
Analysis of tissue samples	Yes, using MAS NMR	No
Availability of databases	Not yet comprehensive but increasing; several are available freely on the web; some commercial products also exist	Comprehensive databases for electron impact MS allow spectral comparisons; For electrospray ionization, as is usual in LC-MS, only mass values can be compared

Data analysis and interpretation

- Data collected represented in a matrix

Variables going across in different columns					
Objects going down in different rows	X-var 1	X-var 2	X-var 3	Y-var 1	Y-var 2
Sample 1					
Sample 2…					

A propositional approach to describing and using metabolomics data (the x-data) for analyzing complex systems. These may have other specific properties (the y-data) which one may also wish to 'explain' in terms of the x-data.

- Chemometric Approach

  - Principle Component Analysis (PCA)

  - Soft Independent Modeling of Class Analogy (SIMCA)

  - Partial Least-Squares (PLS) Method by Projections to Latent Structures

  - Orthogonal PLS (OPLS)

- Targeted Profiling

PCA

- Unsupervised

- Multivariate analysis based on projection methods

- Main tool used in chemometrics

- Extract and display the systematic variation in the data

- Each Principle Component (PC) is a linear combination of the original data parameters

- Each successive PC explains the maximum amount of variance possible, not accounted for by the previous PCs

- PCs Orthogonal to each other

- Conversion of original data leads to two matrices, known as scores and loadings

- The scores(T) represent a low-dimensional plane that closely approximates X. Linear combinations of the original variables. Each point represents a single sample spectrum.

- A loading plot/scatter plot(P) shows the influence (weight) of the individual X-variables in the model. Each point represents a different spectral intensity.

- The part of X that is not explained by the model forms the residuals(E)

A principal component analysis (PCA) model approximates the variation in a data table by a low dimensional model plane. This model plane provides a score plot, where the relation among the observations or samples in the model plane is visualized, for example, if there are any groupings, trends, or outliers. The loading plot describes the influence of the variables in the model plane, and the relation among them. An important feature is that directions in the score plot correspond to directions in the loading plot, and vice versa.

S

Illustration of SIMCA classification. In the left panel, the one-class classifier is shown, referred to as the asymmetric case. In the right panel, the SIMCA classification is shown with two classes, separately modeled by PCA.

- Supervised learning method based on PCA

- Construct a seperate PCA model for each known class of observations

- PCA models used to assign the class belonging to observations of unknown class origin

- Boundaries defined by 95% class interval

- Recommended for use in one class case or for classification if no interpretation is needed
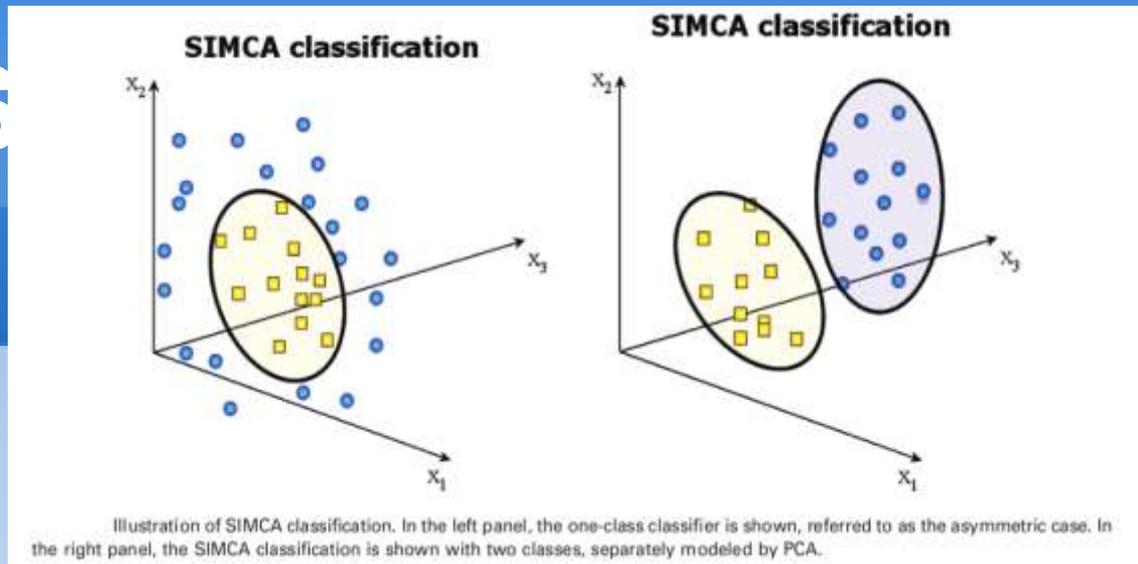
CLASS SPECIFIC STUDIES

- One-class problem: Only disease observations define a class; control samples are too heterogeneous, for example, due to other variations caused by diseases, gender, age, diet, lifestyle, etc.

- Two-class problem: Disease and control observations define two seperate classes

PLS

- Supervised learning method.

- Recommended for two-class cases instead of using SIMCA.

- Principles that of PCA. But in PLS, a second piece of information is used, namely, the labeled set of class identities.

- Two data tables considered namely X (input data from samples) and Y (containing qualitative values, such as class belonging, treatment of samples)

- The quantitive relationship between the two tables is sought.

- $X = TP^T + E$

- $Y = TC^T + E$

- The PLS algorithm maximizes the covariance between the X variables and the Y variables

- PLS models negatively affected by systematic variation in the X matrix not related to the Y matrix (not part of the joint correlation structure between X-Y.

(a) Input data   (b) Mathematical transformation(s)   (c) Output classes

The class assignment problem. The inputs can be considered, and are referred to, as the "explanatory variables" or "x-data" whereas the functional or the other classes of interest, which are still variables associated with the samples, are referred to as "dependent variables" or "y-data" and are to be obtained as the outputs.

Class information can also be used to construct an additional matrix, hereinafter called the Y matrix, consisting of a discrete 'dummy' variable where [1]/[0] indicate the class belonging.

A geometrical illustration of the difference between the PLS-DA and OPLS-DA models. In the left panel, the PLS components cannot separate the between-class variation from the within-class variation, and the resulting PLS component loadings mixes both types of variations. In the right panel, the OPLS components are able to separate these two different variations. Component 1 ($t1_p$) is the predictive component and displays the between-class ([blue circles], [yellow squares]) variation of the samples. The corresponding loading profile can be used for identifying variables important for the class separation. Component 2 ($t2_o$) is the **Y**-orthogonal component and models the within group (within-class) variation.

- OPLS method is a recent modification of the PLS method to help overcome pitfalls

- Main idea to seperate systematic variation in X into two parts, one linearly related to Y and one unrelated (orthogonal).

- Comprises two modeled variations, the Y-predictive ($T_p P_p^T$) and the Y-orthogonal ($T_o P_o^T$) compononents.

- Only Y-predictive variation used for modeling of Y.

- $X = T_p P_p^T + T_o P_o^T + E$

- $Y = T_p C_p^T + F$

- E and F are the residual matrices of X and Y

- OPLS-DA compared to PLS-DA

Remarks on pattern classification

- Intent in using these classification techniques not to identify specific compound

- Classify in specific categories, conditions or disease status

- Traditional clinical chemistry depended on identifying and quantifying specific compounds

- Chemometric profiling interested in looking at all metabolites at once and making a phenotypic classification of diagnosis

Targeted profiling

- Targeted metabolomic profiling is fundamentally different than most chemometric approaches.

- In targeted metabolomic profiling the compounds in a given biofluid or tissue extract identified and quantified by comparing the spectrum of interest to a library of reference spectra of pure compounds.

- Key advantage: Does not require collection of identical sets = More amenable to human studies or studies that require less day-to-day monitoring.

- Disadvantage: Relatively limited size of most current spectral libraries = bias metabolite identification and interpretation.

- A growing trend towards combining the best features of both chemometric and targeted methods.

Databases

- Large amount of data

- Need for databases that can be easily searched

- Better databases will help in combining chemometric and targeted profiling methods

- Newly emerging databases

- HMDB good model for other databases

- Challenge of standardisation

Summary of metabolomic databases

| Database name	URL or web address	Comments
Human metabolome database	http://www.hmdb.ca	Largest and most complete of its kind. Specific to humans only
BioMagResBank (BMRB – metabolimics)	http://www.bmrb.wisc.edu/metabolomics/	Emphasis on NMR data, no biological or biochemical data
		Specific to plants (Arabadopsis)
BiGG (database of biochemical, genetic and genomic metabolic network reconstructions)	http://bigg.ucsd.edu/home.pl	Database of human, yeast and bacterial metabolites, pathways and reactions as well as SBML reconstructions for metabolic modeling
Fiehn metabolome database	http://fiehnlab.ucdavis.edu/compounds/	Tabular list of ID'd metabolites with images, synonyms and KEGG links
Golm metabolome database	http://csbdb.mpimp-golm.mpg.de/csbdb/ gmd/gmd.html	Emphasis on MS or GC–MS data only
		No biological data
		Few data fields
		Specific to plants
METLIN metabolite database	http://metlin.scripps.edu/	Human specific
		Mixes drugs, drug metabolites together
		Name, structure, ID only
NIST spectral database	http://webbook.nist.gov/chemistry/	Spectral database only (NMR, MS, IR)
		No biological data, little chemical data
		Not limited to metabolites
Spectral database for organic compounds (SDBS)	http://www.aist.go.jp/RIODB/SDBS/cgi-bin/ direct.frame.top.cgi?lang=eng	Spectral database only (NMR, MS, IR)
		No biological data, little chemical data
		Not limited to metabolites

Text/Synonym Search

Quick Search

Keywords :

Location (optional) :

*Go

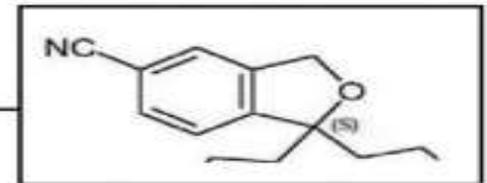Location Hint: Enter a city, a state, city comma state, or zipcode.

Structure Search

NC

(S)

O

NMe₂

F
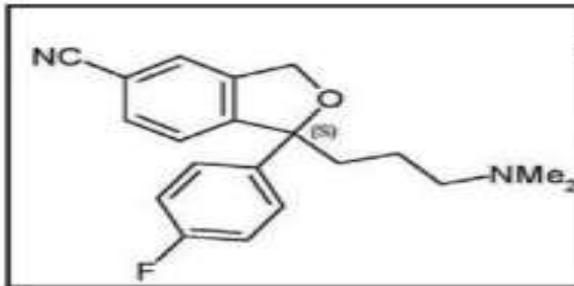
Mol Weight Search

Smiles & Formula Search

CN1C=C(N=C1)

Human Metabolome Database (HMDB)

Substructure Search

NC

O

(S)

MS Spectral Search

CH₃CHCH₂CH₃
|
CH₃
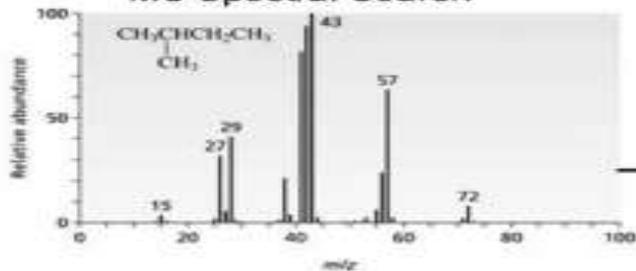
NMR Spectral Search
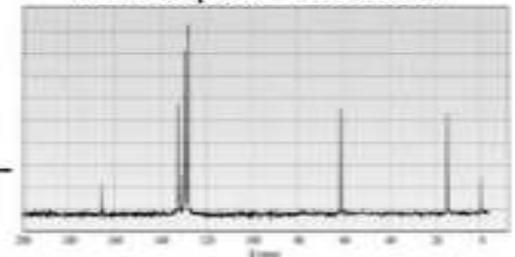
Sequence Search

METQPRDSKNACDEG
LIGGEDCANKSDRPQ
VTREYQPASFHGFDL
PIYTREWQQNMLKAF

Integration of metabolomics with other 'omics' fields

- Integrating genomics and metabolomics for engineering plant metabolic pathways - Kirsi-Marja Oksman-Caldentey and Kazuki Saito (2005)

- Proteomic and metabolomic analysis of cardioprotection: Interplay between protein kinase C epsilon and delta in regulating glucose metabolism of murine hearts

- Recent studies (2005) to integrate transcriptomics, proteomics and metabolomics in an effort to enhance production efficiency under stressful conditions of grapes.

- Nutrigenomics is a generalised term which links genomics, transcriptomics, proteomics and metabolomics to human nutrition.

Main Applications

- Drug assessment
- Clinical toxicology
- Nutrigenomics
- Functional genomics

Examples of interesting research projects

- Metabolomics and its Application for non-invasive embryo assessment in IVF

- Nonivasive metabolomic profiling of embryo culture media using proton nuclear magnetic resonance correlates with reproductive potential of embryos in women undergoing in vitro fertilization

- Nonivasive metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a prospective blinded pilot study

- Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression

- A Multivariate Screening Strategy for Investigating Metabolic Effects of Strenuous Physical Exercise in Human Serum

IVF

- Statistics

- Grading system based on embryo morphology and cleavage rates the mainstay of embryo assessment worldwide

- Not sufficiently precise

- Investigations to demonstrate underlying metabolic difference between embryos resulting in pregnancy and those that do not.

IVF

- Aim of the method:

  - To increase pregnancy rates and reduce number of embryos implanted

  - To enhance treatment outcomes and a reduction in multiple birth rate

  - To reduce time and cost of achieving a successful pregnancy

  - To expand the IVF market

IVF

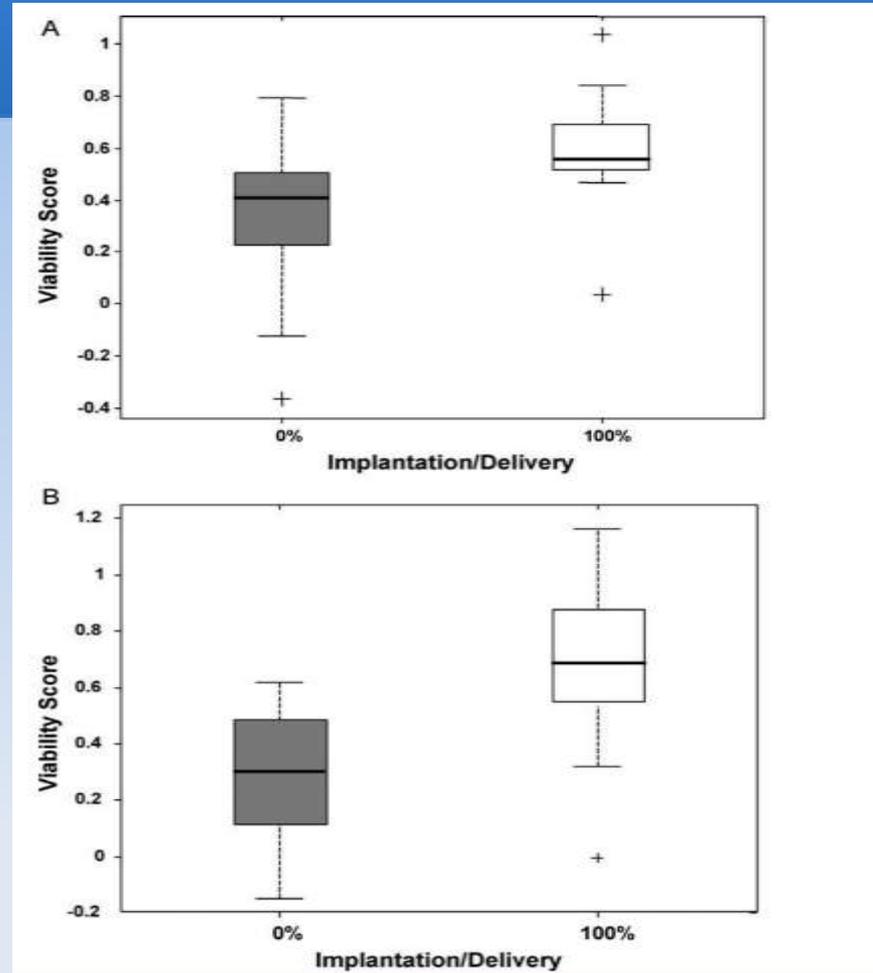Studies of non-invasive metabolomic profiling of embryo culture media to assess embryo viability in IVF.

Study	Study design	n	Day of transfer	Number of embryos transferred	Analytical technique	Center	Findings
Seli *et al.* (2007)	Algorithm development	36	Day 3	MET	Raman	YFC	A
Scott *et al.* (2008)	Blinded analysis	41	Day 3 and 5	MET	Raman	RMANJ	B
Seli *et al.* (2007)	Algorithm development	33	Day 3	MET	NIR	RMANJ	A
Seli *et al.* (2007)	Blinded analysis	16	Day 3	MET	NIR	YFC	B
Seli *et al.* (2008b)	Algorithm development	121	Day 2	SET	NIR	KLC	A, C
Seli *et al.* (2008b)	Blinded analysis	60	Day 2	SET	NIR	KLC	B, D
Vergouw *et al.* (2008)	Algorithm development	29	Day 2	SET	NIR	VUMC	A, C
Vergouw *et al.* (2008) Seli *et al.* (2008b)	Algorithm development	304	Day 3	SET	NIR	VUMC	A, C
Hardarson *et al.* (2008)	Algorithm development	137	Day 5	SET	NIR	FCG, SG	A, C, D

SET, single embryo transfer; MET, multiple embryo transfer; YFC, Yale Fertility Center, New Haven, CT, USA; RMANJ, Reproductive Medicine Associates, Morristown, New Jersey, USA; VUMC, Vrije Universiteit Medical Center, Amsterdam, The Netherlands; KLC, Kato Ladies Clinic, Tokyo, Japan; FCG, Fertilitets centrum, Göteborg, Sweden; SG, Shady Grove Fertility Reproductive Science Center, Rockville, Maryland, USA.

A, the mean viability score of embryos that implanted and resulted in fetal cardiac activity or live birth was significantly higher compared with the mean viability score of embryos that failed to implant; B, spectroscopic analysis by an observer blinded to pregnancy outcome, using a previously established regression algorithm demonstrated that the mean viability score of embryos that resulted in a pregnancy was higher compared with embryos that failed to implant; C, study showed the metabolomic profile of embryo culture media to be independent of morphology; D, a positive correlation was detected between increasing viability scores and the potential of individual embryos to result in a pregnancy.

IVF

Viability score calculated using (A) NIR and (B) Raman spectra of culture media are shown for embryos that implanted and lead to delivery (empty) and those that did not implant (shaded).

IVF

- Result:
  - Glutanate concentrations
  - Viability indices
- Conclusion
- Correlation of metabolic profile of spent embryo culture media with reproductive potential of embryos

¹H NMR Spectroscopy
glutamate and alanine/lactate ratio model

Using a model based on glutamate and alanine/lactate ratio quantities, Seli and co-workers demonstrated that embryos with proven reproductive potential have significantly higher vialibility indices compared to embryos that did not implant. Proton NMR spectrocopy identified implantation/pregnancy with a sensitivity of 88.25% and a specificity of 88.2%

Future challenges and development

- Database

- Standardisation

- Diversity/variation of metabolomic data

  - More efficient ways of identification

- Better models for interpretation of data

- Integration with other 'OMICS'

Bibliography

- Current Progress in computational metabolomics – David S.Wishart, 2007

- Metabonomics in pharmaceutical R & D – John C.Lindon, Elaine Holmes and Jeremy K.Nicholson

- Wikipedia – search on "Metabolomics"

- Metabolomics Basics : What is Metabolomics? - Thermo Scientific, www.thermo.com

- A metabolome pipeline : from concept to data to knowledge – Marie Brown, Warwick B.Dunn, David I.Ellis, Roystone Goodacre, Julia Handl, Joshua D.Knowles, Steve O'Hagan, Irena Spasic and Douglas B.Kell, 2004

- Integrative Genomics and Functional Explanation – Jo Davies, Thorunn Rafner, Garrett Hellenthal and Jotun Hein, 2009

- Chemometrics in Metabonomics – Johan Trygg, Elaine Holmes and Torbjorn Lundstedt

- Wikipedia – search on "Chromatography"

- Spectroscopic and Statistical Techniques for Information Recovery in Metabonomics and Metabolomics – John C.Lindon and Jeremy K.Nicholson

- Wikipedia – search on "NMR"

- Wikipedia – search on "Spectroscopy"

- SetupX: A public study design database for metabolomic projects – Scholz M, Fiehn O, 2007

- New bioinformatics resources for metabolomics – Markley JL, Anderson ME, Cui Q

- Non-Invasive Metabolic profiling of embryo culture media using proton nuclear magnetic resonance correlates with reproductive potential of embryos in women undergoing in vitro fertilization. - E. Seli, L.Botros, D.Sakkas, D. Burns

- Non-Invasive Metabolomic profiling of human embryo culture media using Raman spectroscopy predicts embryonic reproductive potential: a proospective blinded pilot study.- Scott R., et al.2008

The End