MRC-HPA Centre for Environment and Health

Imperial College London

MRC Medical Research Council

Health Protection Agency

KING'S College LONDON

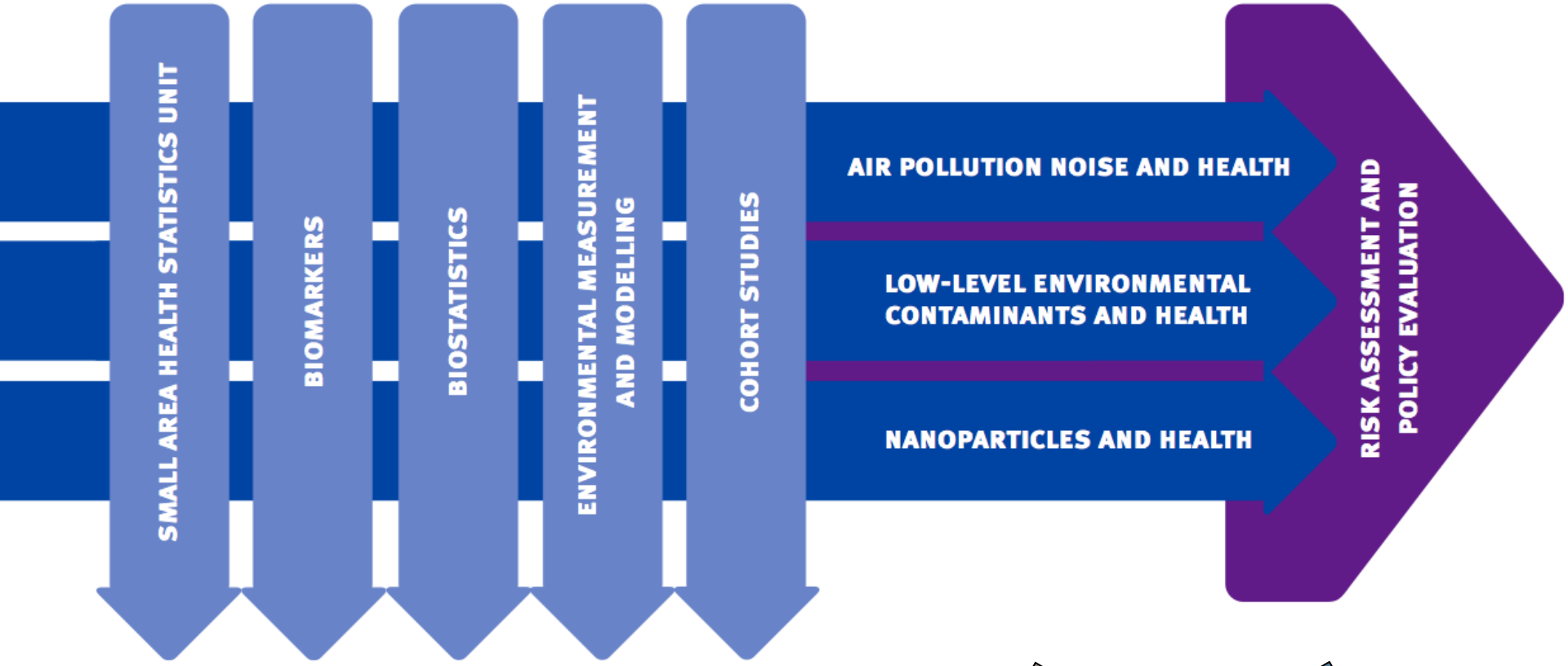
Designed Biofluid Mixtures Allow Feature-Wise Evaluation of Metabolic Profiling Analytical Platforms

Toby Athersuch

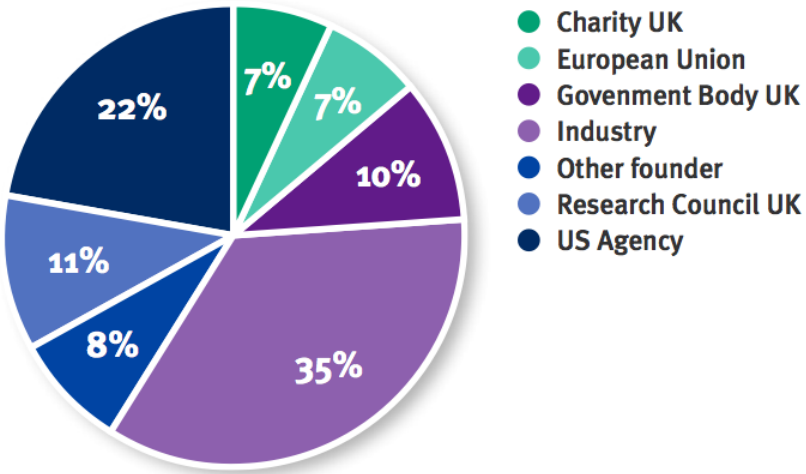
Isaac Newton Institute for Mathematical Sciences

30/08/2011

toby.athersuch@imperial.ac.uk



CENTRE FUNDING 2010–11



Biomarkers

- NIH definition:
  - "*a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention* "

    Biomarkers Definitions Working Group (2001).

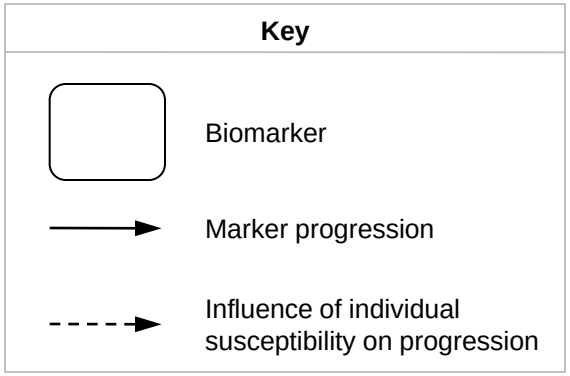
- Where do we find them?
  - Many sampling source (urine / plasma / tissue / breath / etc.)
  - Combinations of measurements

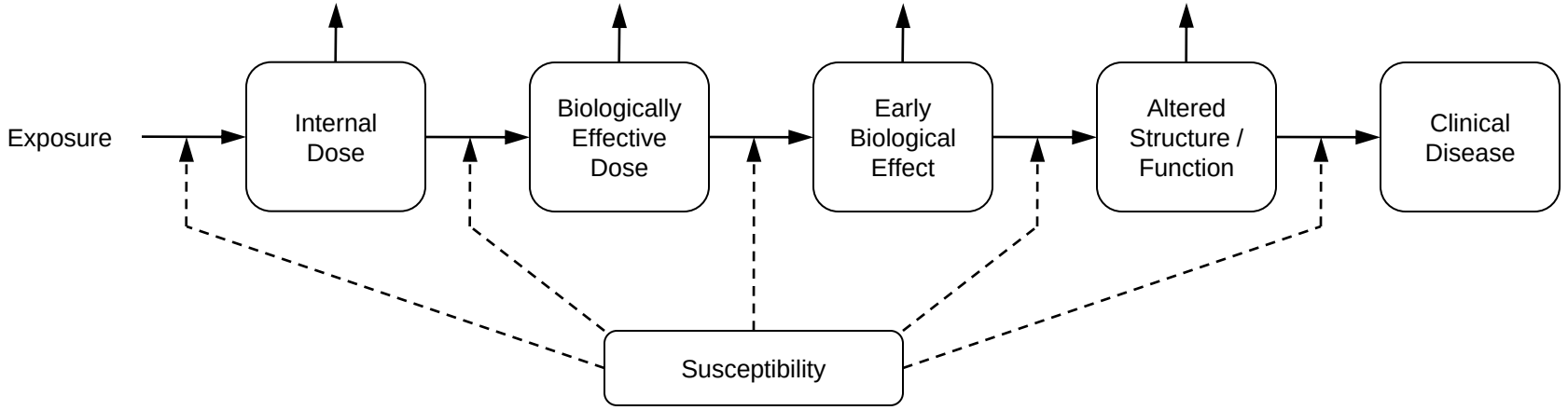
- Considerations for sampling sources:
  - Urine — Time-average systemic excretion
  - Blood — 'Snapshot' of system
  - Tissue — Likely site of injury/concern/action/effect

- Biomarkers can potentially:
  - Report *more specifically* on type of exposure
  - Give a better idea on the exposure that has *actually occurred*
  - Reflect a variety of stages in the initiation and progression of disease, and efficacy of treatment

Biomarkers

- Finding biomarkers
  - EXPOSURE
    - External exposure
    - Internal dose
  - BIOLOGICALLY EFFECTIVE DOSE
  - EFFECT
    - Biological response / early biological effect
    - Altered biological structure or function
    - Disease onset
  - SUSCEPTIBILITY
    - Underlying characteristics that facilitate/modulate exposure and response





- Adapted from: Committee on Biological Markers of the National Research Council. Biological markers in environmental health research. *Environ. Health Perspect*. 74: 3-9, 1987.

Omics and Global Profiling

- Measuring many (10's – 1000's) of biomolecules at once
  - Largely untargeted approach
  - (Hopefully) good coverage of molecular classes of interest (e.g. metabolites, gene transcripts, etc.)
  - Minimise assumptions
  - Useful for hypothesis generation – getting a list of potential candidates for further validation
- The 'profile' may be considered as a whole
  - Combinations of molecules may respond in concert
  - Potentially more sensitivity and specificity
- Approach across 'omics' sciences have numerous common elements
  - Differ according to level/type of biological information being investigated and associated methodologies
- Generate a multiparametric dataset
- Apply multivariate statistics to derive putative biomarkers

Some Definitions

Metabonomics

"*…the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification…"* *- Metabonome*

Nicholson JK, Lindon JC, Holmes E. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*. 29:1181-1189.

Metabolomics

"...the complete set of metabolites / low-molecular-weight intermediates, which are context dependent, varying according to the physiology, developmental or pathological state of the cell, tissue, organ or organism…" *- Metabolome*

Fiehn O. 2002. Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Biol.* 48(1-2) 155-171.

Metabolic Profiling

General term for 'omics' metabolite measurements

Metabolic Profiling

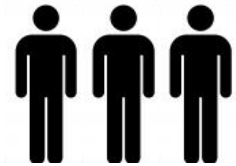
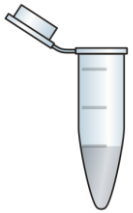
- Profiles can be any description of the small molecule composition of a ‘biological sample’
  - For each sample measure **as many different molecules** as possible
  - Measure in as **unbiased** a manner as possible
  - Usually derived using **spectroscopy**

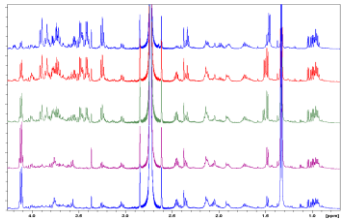
- Can simultaneously **generate (multiple) hypotheses** and test them
  - No *a priori* assumptions needed about what is important (untargeted analysis)
  - Can be used to focus on specific classes or set of metabolites (targeted)
  - **Multivariate analysis** makes this process **efficient for biomarker discovery**
  - **Combinations** of biomarkers can be more sensitive and specific than one

- Profiles **must** be:
  - Amenable to quantitative interpretation
  - Reproducible (relative to ‘biological’ variation)

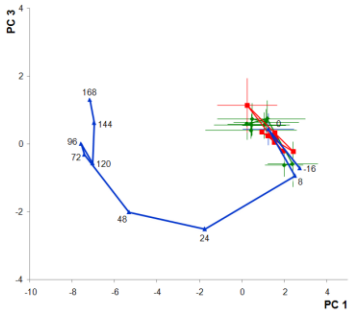
- Profiles **need not** be:
  - Comprehensive (limited by instrumental considerations)
  - Fully resolved and annotated

Outline of Analysis Strategy



Collection ➡ Preparation ➡ Analysis ➡ Data Analysis

---

1. **Generate profiles** that reflect/report on metabolic state of biological system under study
2. Use multivariate analysis / pattern recognition to indicate **profile features related to intervention**/disease…
3. Identify **metabolites responsible** for these profile features – putative biomarkers
4. Generate **testable hypotheses** to validate putative biomarkers
5. Use profiles or features **to derive classification models**

MRC-HPA Centre for Environment & Health

Imperial College London · MRC Medical Research Council · Health Protection Agency · King's College London

Analytical Platforms for Metabolic Profiling

NMR Spectroscopy

1D NMR
Multidimensional NMR
Magic angle spinning NMR
Flow injection NMR
Capillary NMR

Chromatography

Gas chromatography
Capillary electrophoresis
UPLC
HPLC
SPE

Hyphenation

Mass Spectrometry

Single quad MS
Triple quad MS
Time-of-flight MS
Ion mobility MS
Ion trap MS
FTMS

High Throughput

Preparation Robotics
Automation
LIMS

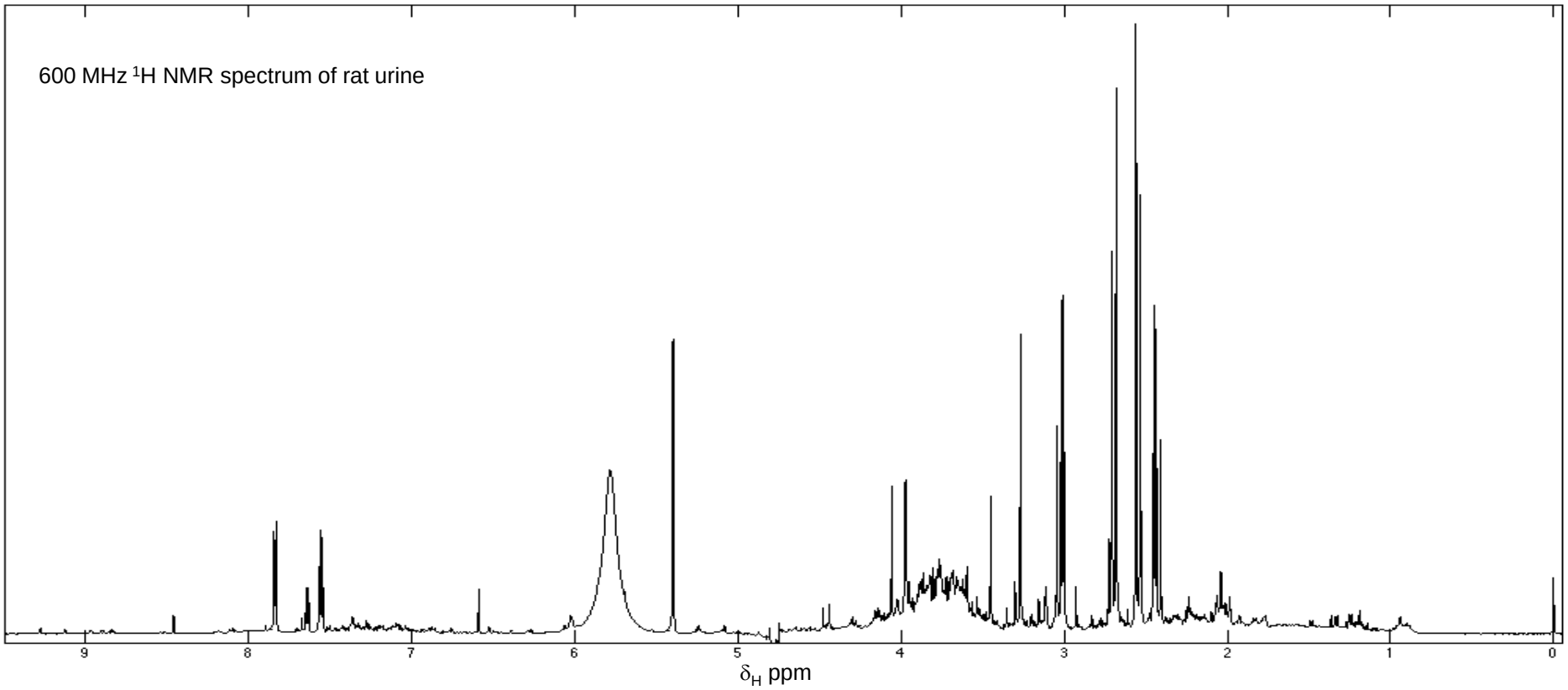
MRC-HPA Centre for Environment & Health

Imperial College London
MRC Medical Research Council
Health Protection Agency
King's College London

Biofluid NMR for Metabolic Profiling

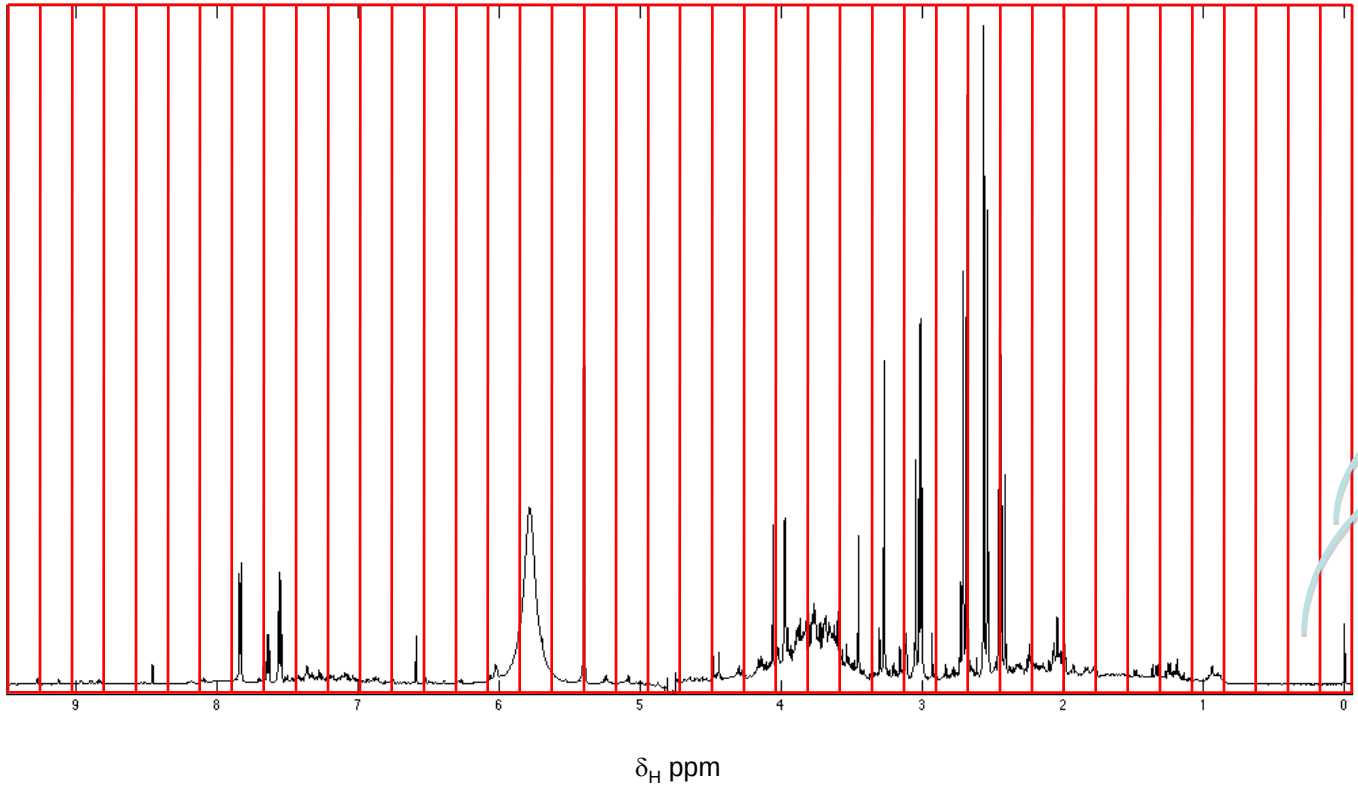
- Spectra of biofluids (urine, plasma, bile, CSF) are typically very complex, with many overlapped peaks

- Spectral assignment is complex, but necessary in drug metabolism work and metabonomics



Spectral Data for Multivariate Analysis

- 1D ^{1}H NMR data amenable to analysis
  - Full resolution
    - e.g. 32 K points / spectrum
    - high computational load
    - Potentially affected by minor chemical shift variation
  - Integrated regions
    - Arbitrary / targeted
    - Ability to incorporate moderate chemical shift variation
    - Less specific output from data analysis

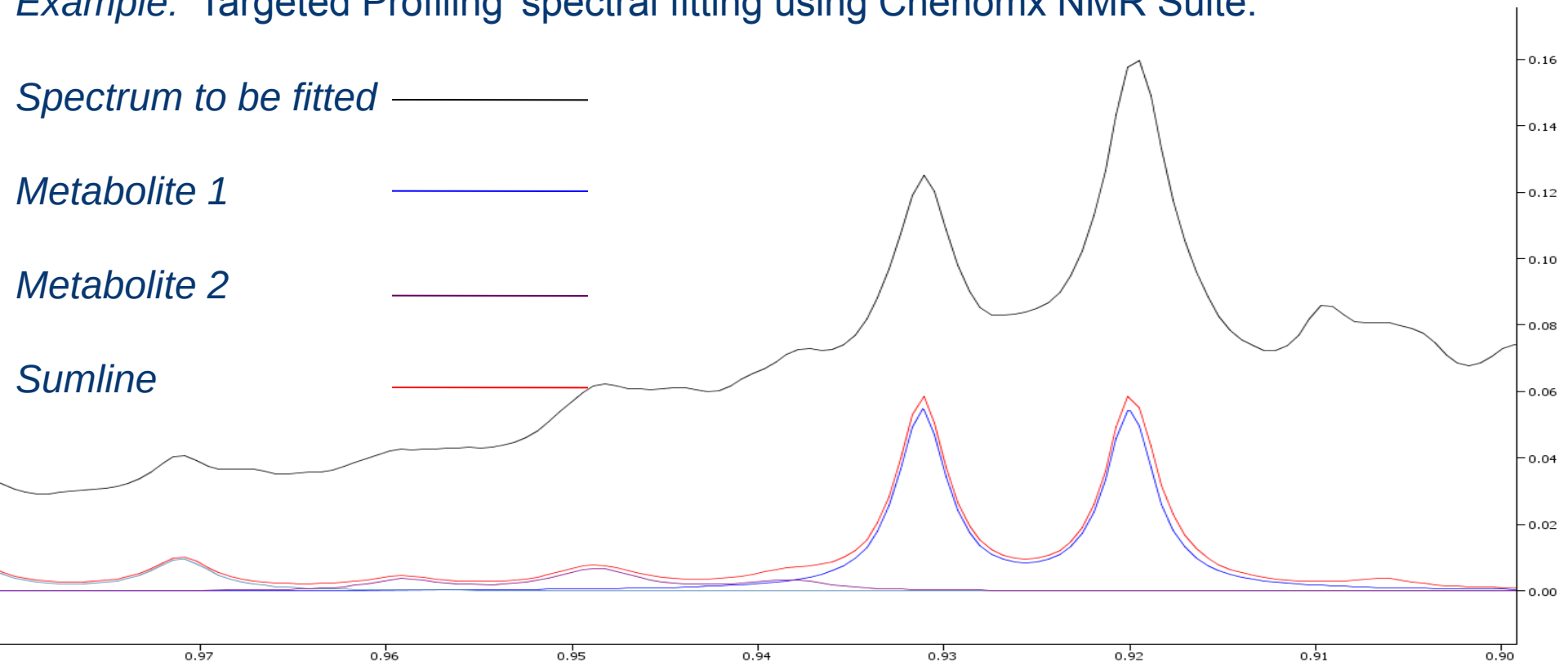


	Region			
	1	2	3	4
Sample 1				
Sample 2				
Sample 3				
…				

Spectral Data for Multivariate Analysis

- 1D ^{1}H NMR data amenable to analysis
  - Targeted analysis
    - Fit model spectra of well-characterised metabolites to target spectrum
    - Quantification possible through reference to internal standard

Example: 'Targeted Profiling' spectral fitting using Chenomx NMR Suite:



Designed Mixtures in Biofluid Analysis

Background:
Traditional ‘spike-in’ experiments used for the assessment of precision and accuracy of analyte measurements, are adversely affected by variation in the **background matrix composition**. Such approaches my underestimate the risk of matrix effects confounding accurate measurement of major metabolites

Aim:
To evaluate and illustrate how mixing intact biofluids according to a know, suitable experimental design can help:

- evaluate individual measurements
- compare performance
- enable optimisation

… of quantitative spectral analysis tools under conditions that better approximate a real metabonomics experiment.

Pilot Study – Human Urine

STUDY DESIGN - Simplex Lattice

A {*q, m*} simplex-lattice design for *q* components consists of points defined by the following coordinate settings:

The proportions assumed by each component take the *m*+1 equally spaced values from 0 -1,

x_i = 0, 1/*m*, 2/*m*, ... , 1 for *i* = 1, 2, ... , *q*

All of the possible combinations of the proportions described are sampled in the experimental.

STRATEGY

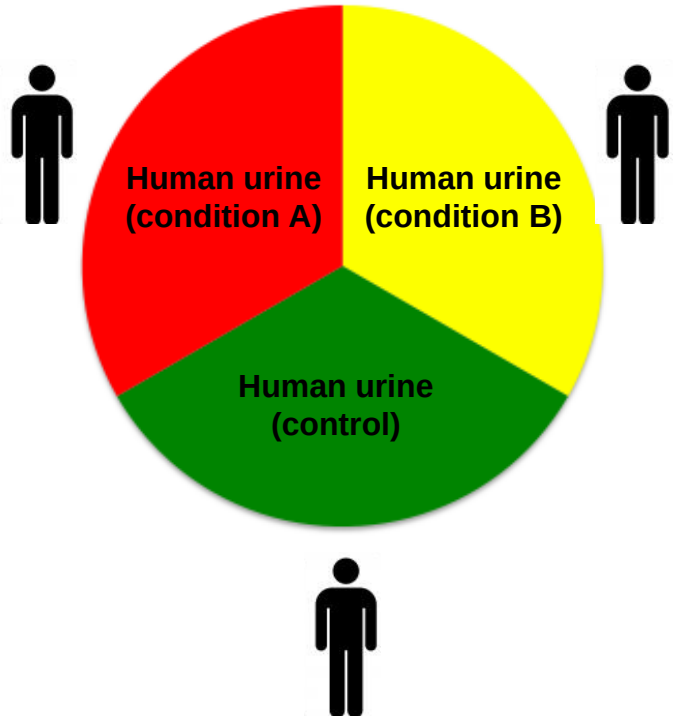
If a metabolite is accurately measured by targeted analysis from this data set, then the variation in the measurements will fit a linear regression to this experimental design. In this study we have compared a **'targeted profiling'** approach using Chenomx NMR Suite with a **'targeted bucketing'** approach.

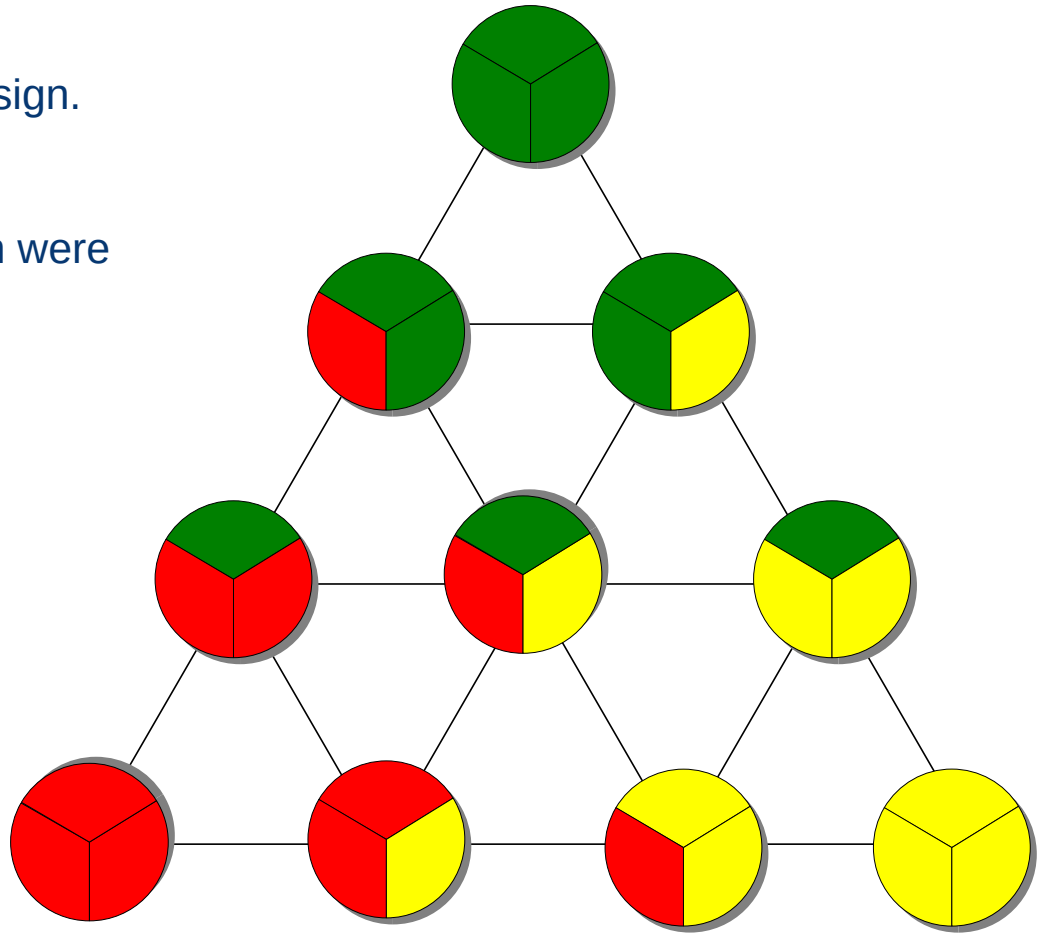
From: *NIST/SEMATECH e-Handbook of Statistical Methods,* http://www.itl.nist.gov/div898/handbook/, Updated 6/2010

Pilot Study – Human Urine

Schematic showing the experimental design.

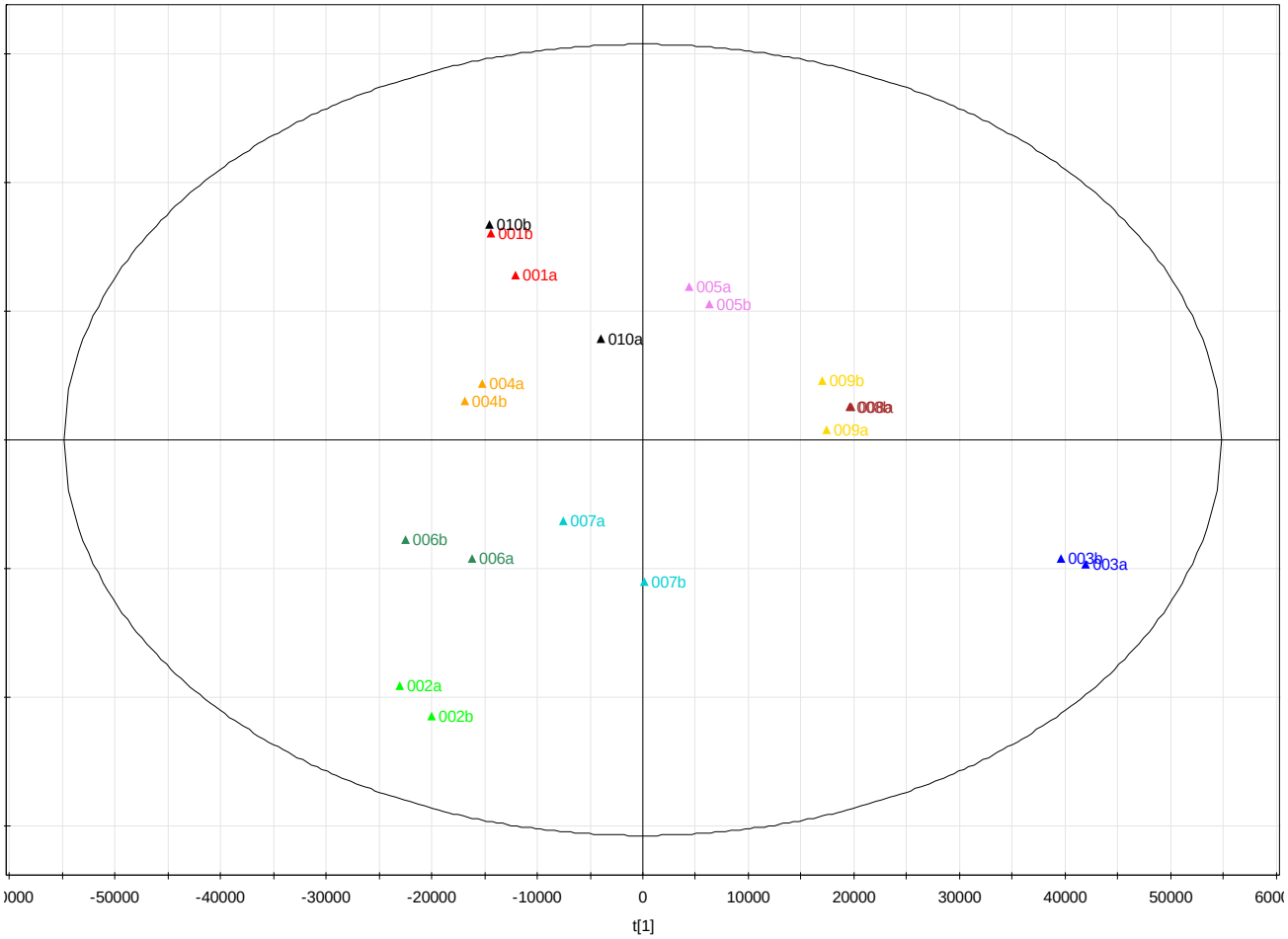
Pooled urines taken from a previously completed human nutritional intervention were mixed in known proportions





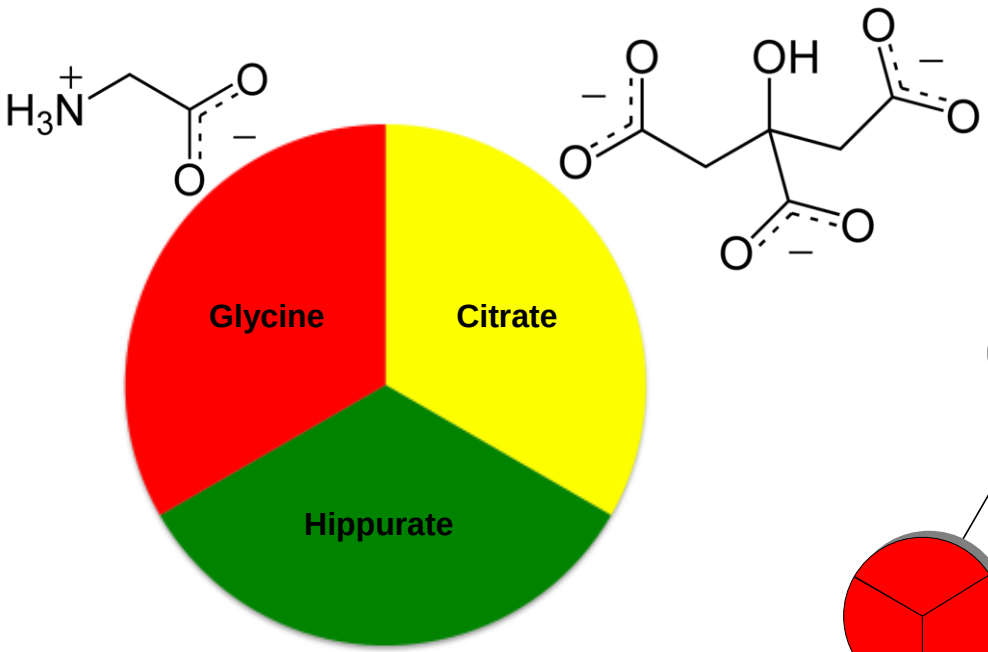
Pilot Study – Human Urine

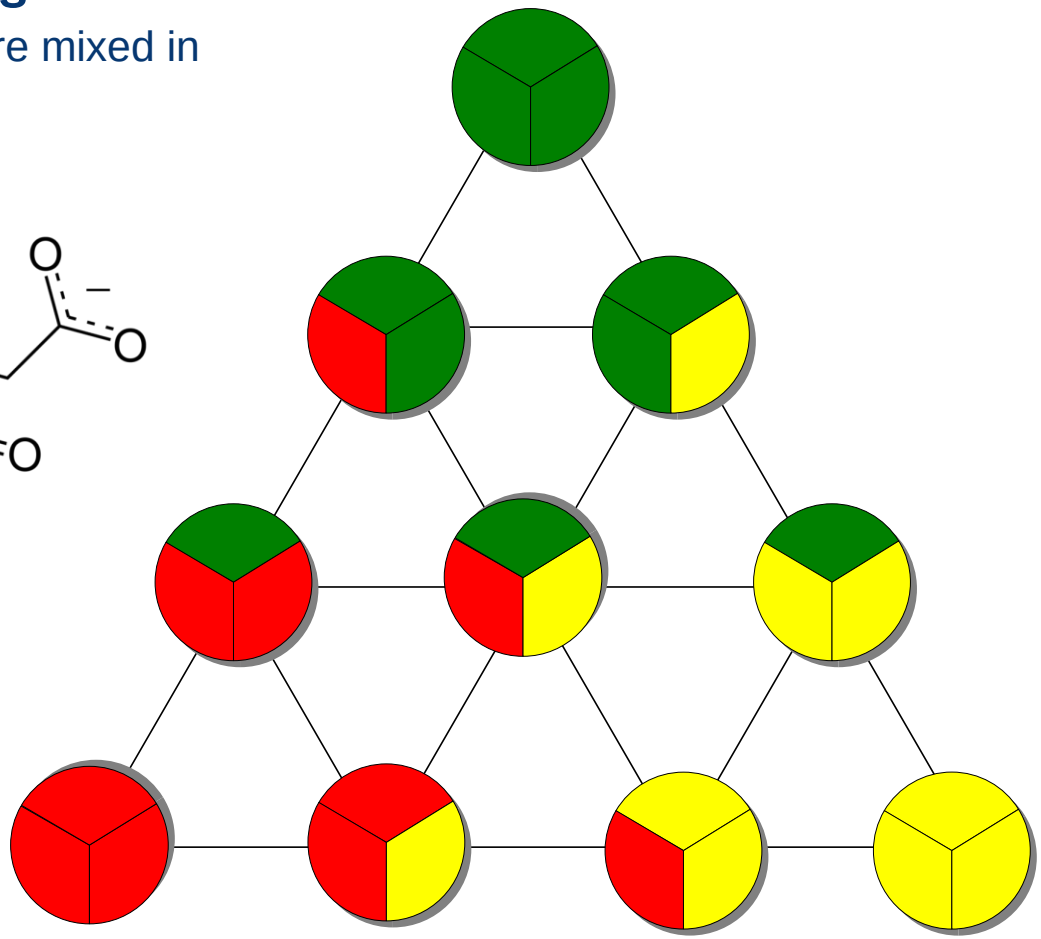
Student (1st) attempt at preparing ‘real’ biofluid samples
PCA revealed relatively weak concordance with experimental design



Pilot Study - Standard Mixtures

Three different standard compounds were mixed in known proportions

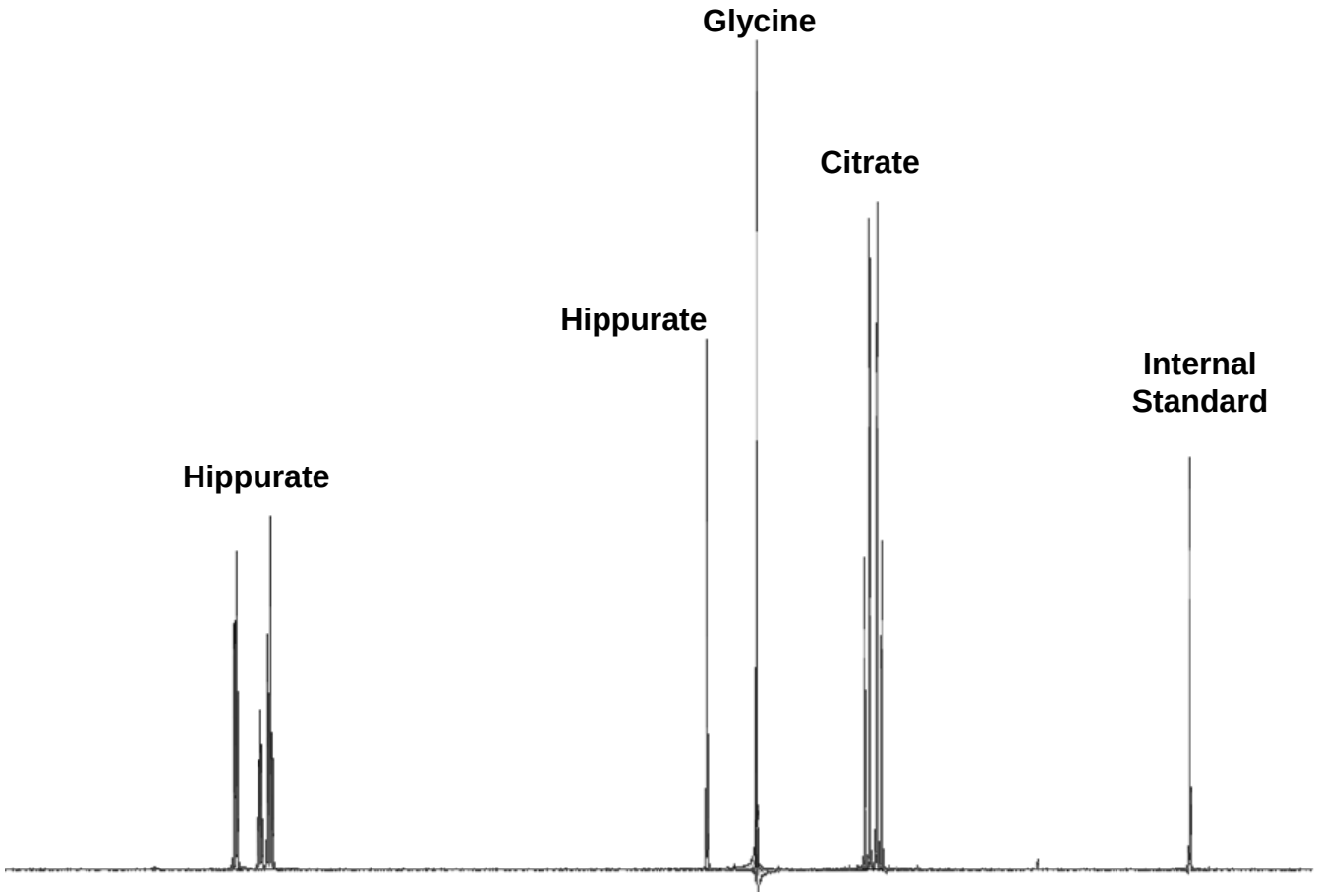




Pilot Study - Standard Mixtures

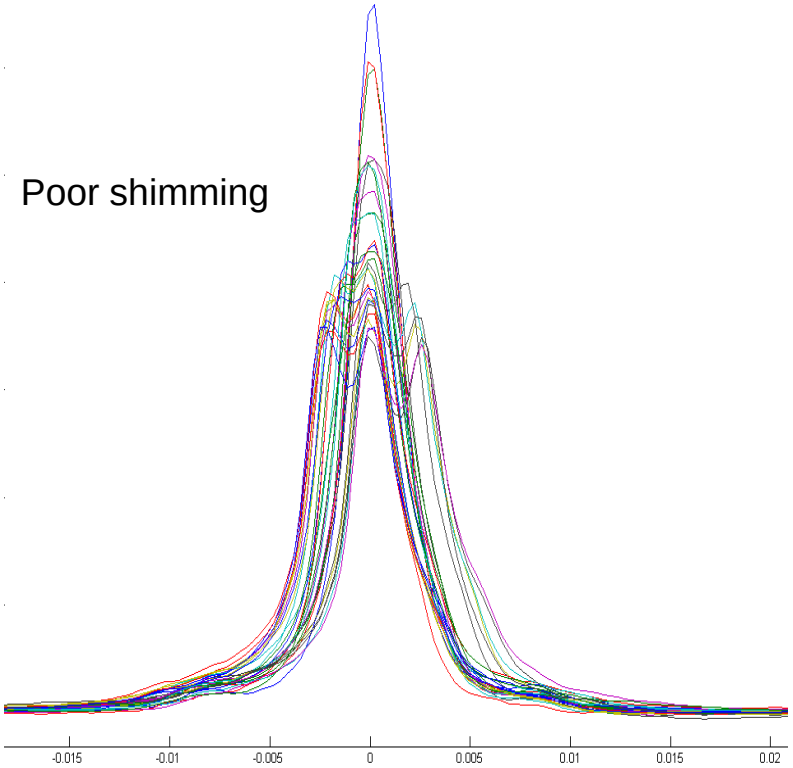
Spectra are sparse

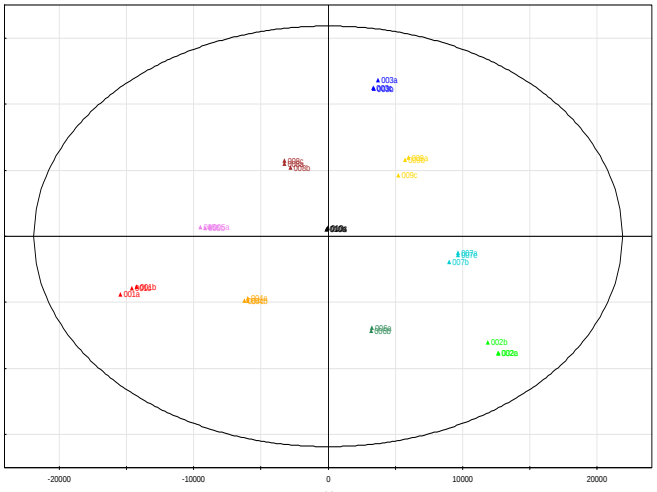
No spectral overlap between compounds



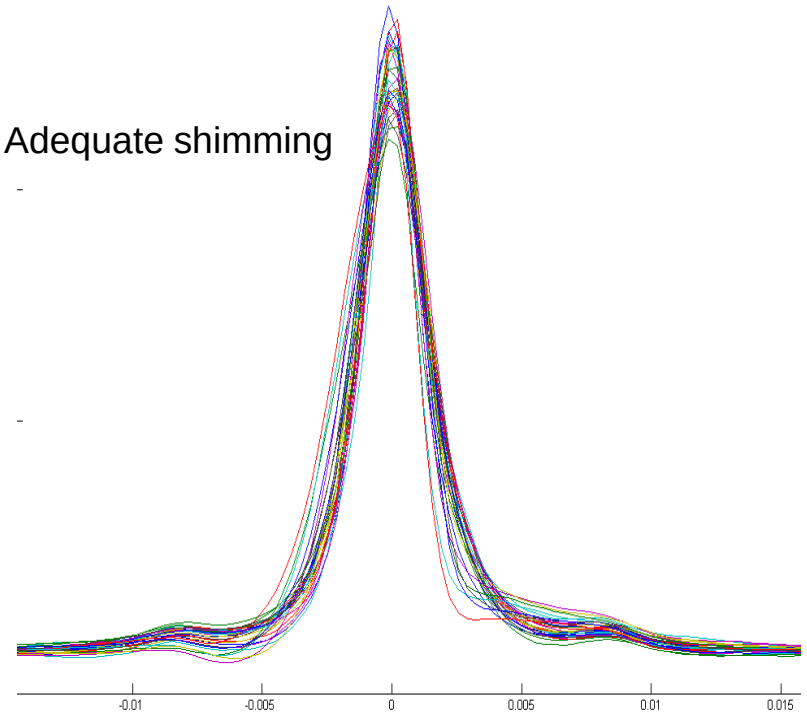
Results - 2

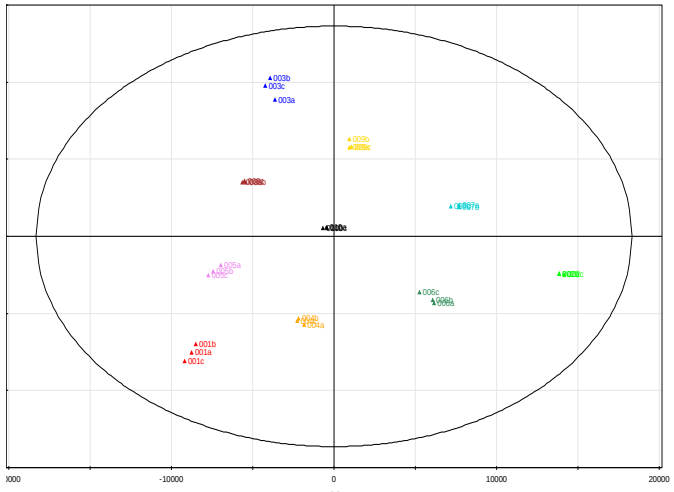
Poor shimming





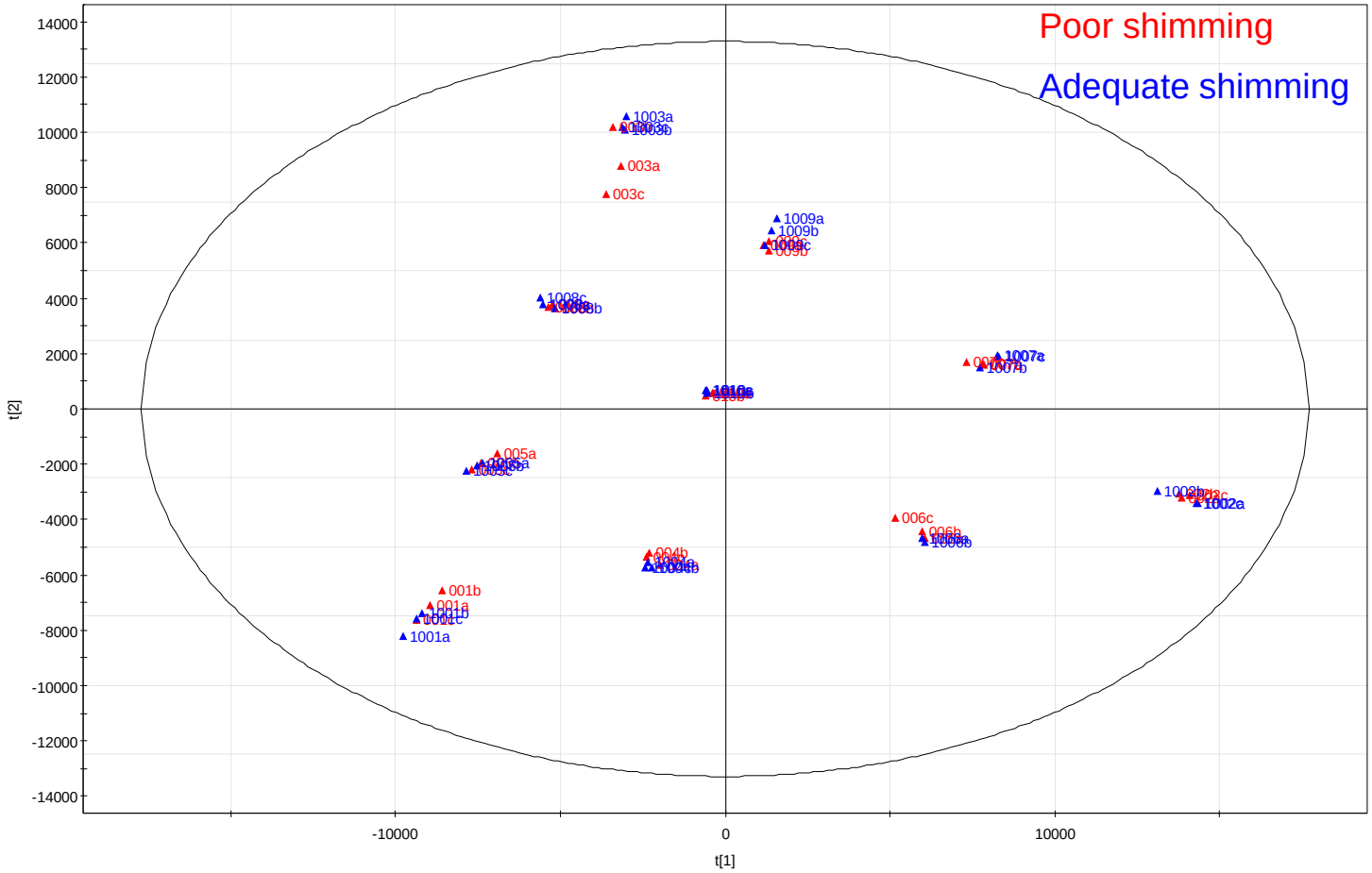
Adequate shimming





Pilot Study – Standard Mixtures

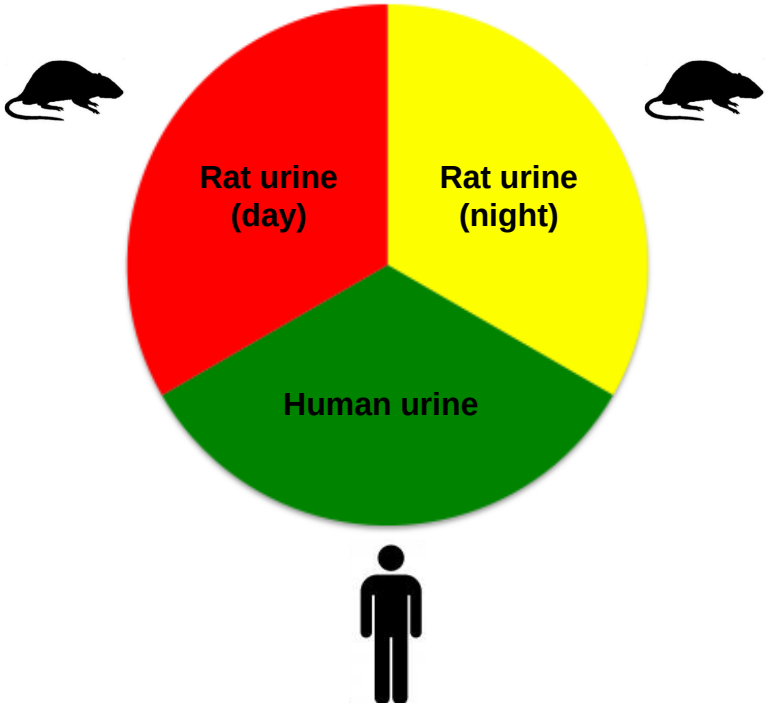
Comparison of all samples by PCA revealed minimal contribution from shimming artifact
Explained by this being the result of a similar change to all measured signals.

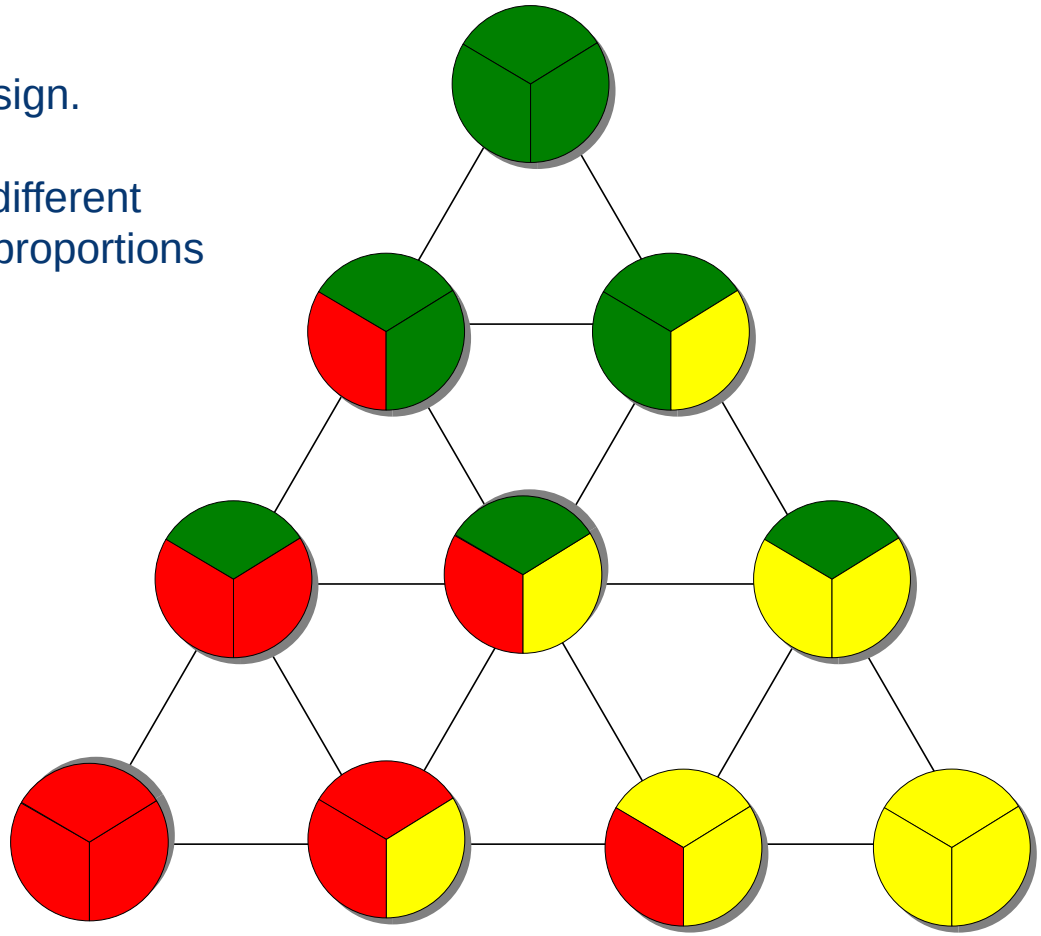


Study - Different Urines

Schematic showing the experimental design.

Three different urines with substantially different metabolic profiles were mixed in known proportions





Experimental - 3

EXPERIMENTAL METHODS

NMR Acquisition

Quantitative 1D ^{1}H NMR spectra were obtained
600 MHz
NOESYPR1D pulse sequence
Sum of 128 free
Spectral width of 20 ppm.
Analytical triplicates were used for each biofluid sample

NMR Data Processing

Line broadening of 1 Hz prior to Fourier transformation.
Phase and baseline distortions
Internally to the TSP-d_4 resonance relative to DSS (δ_{1H} 0.00).

Targeted Profiling

Quantification of individual urinary metabolites
Chenomx NMR Suite 4.6 (Chenomx Inc., Edmonton, Canada).
Reference spectra from the Chenomx 600 MHz library
Relative concentrations of each metabolite present determined.

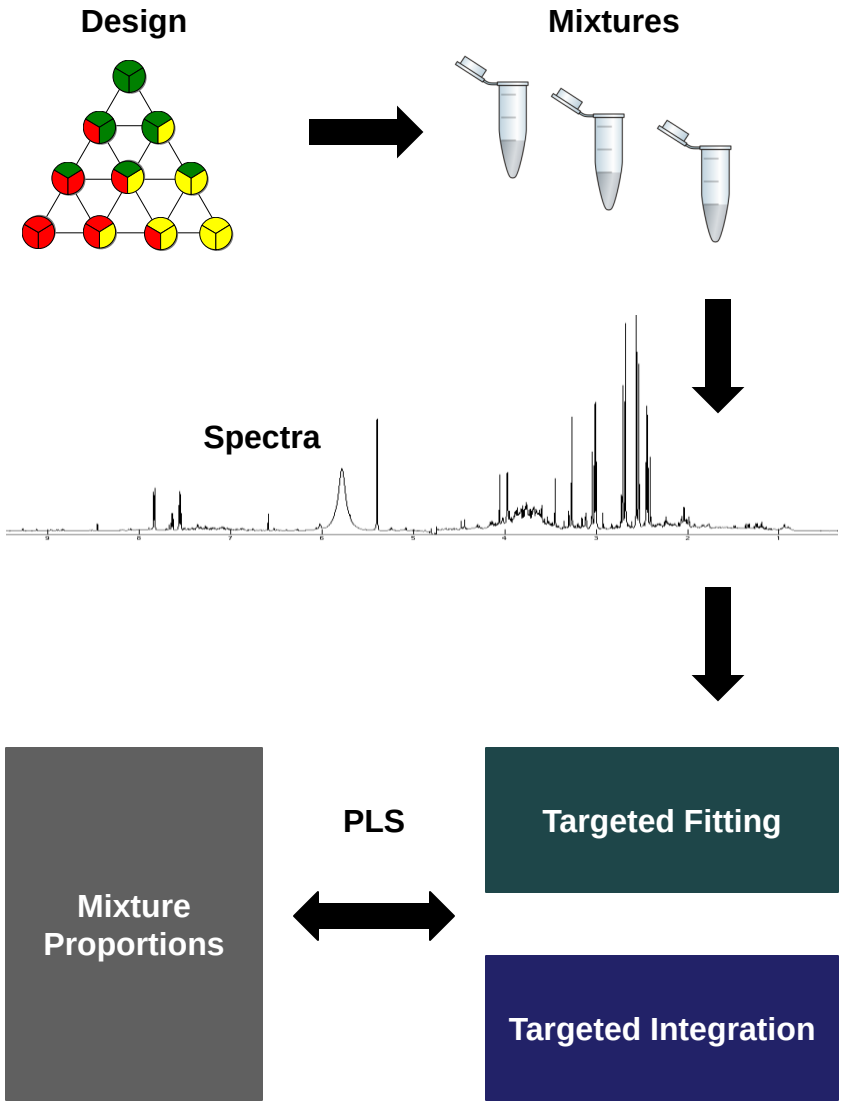
Targeted Bucketing

Spectral regions were defined for each metabolite of interest
The integral area of these regions was calculated

Pattern Recognition

Partial least squares regression
- Metabolite concentration data (X)
- Experimental design matrix (Y)

Goodness-of-fit (R^2) and goodness-of-prediction (Q^2) estimates



Results - 3

Study – Different Urine

PLS scores for “Targeted Profiling” method.

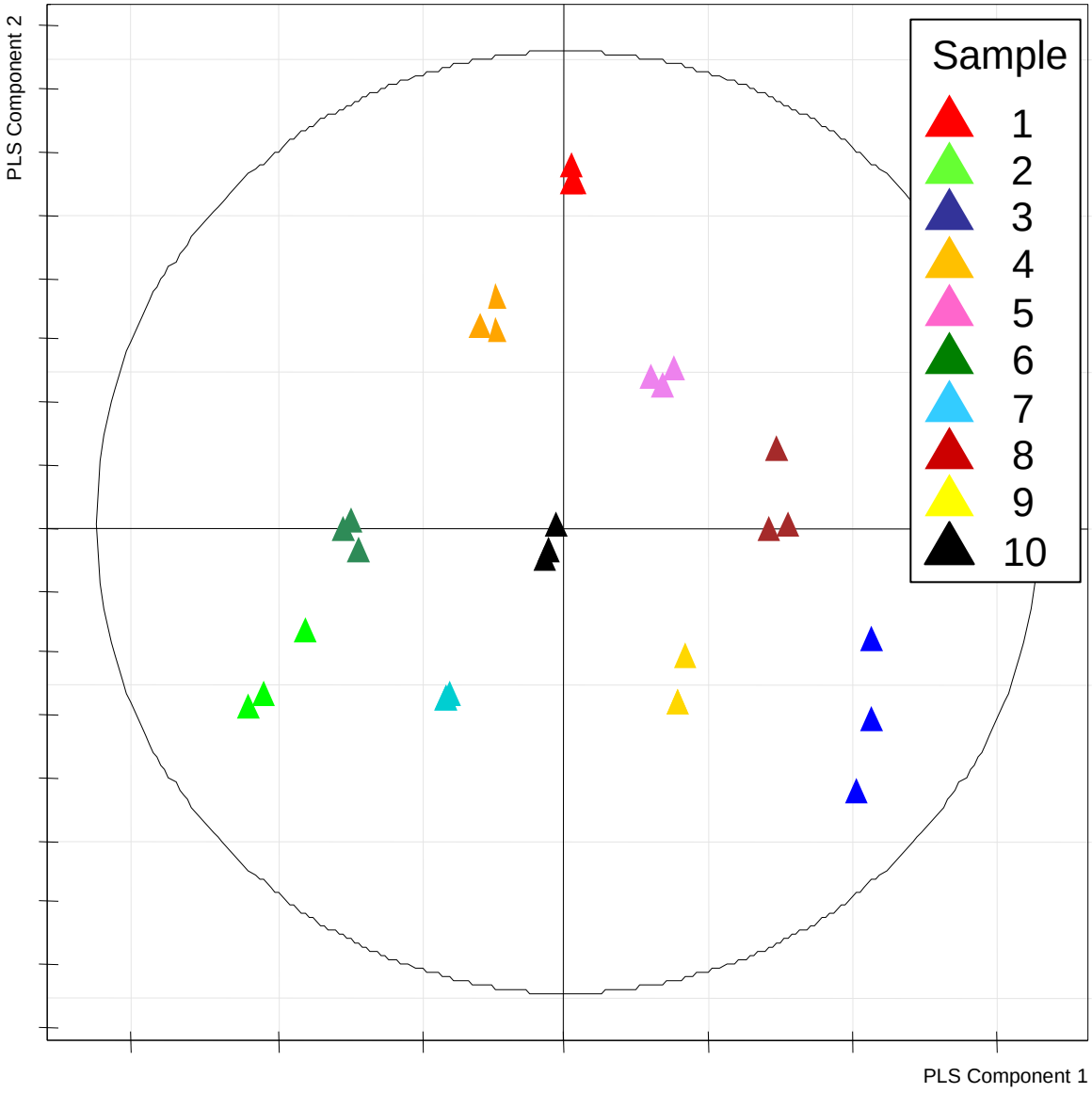
Show a relatively tight fit to the experimental design

Samples clustered in replicates

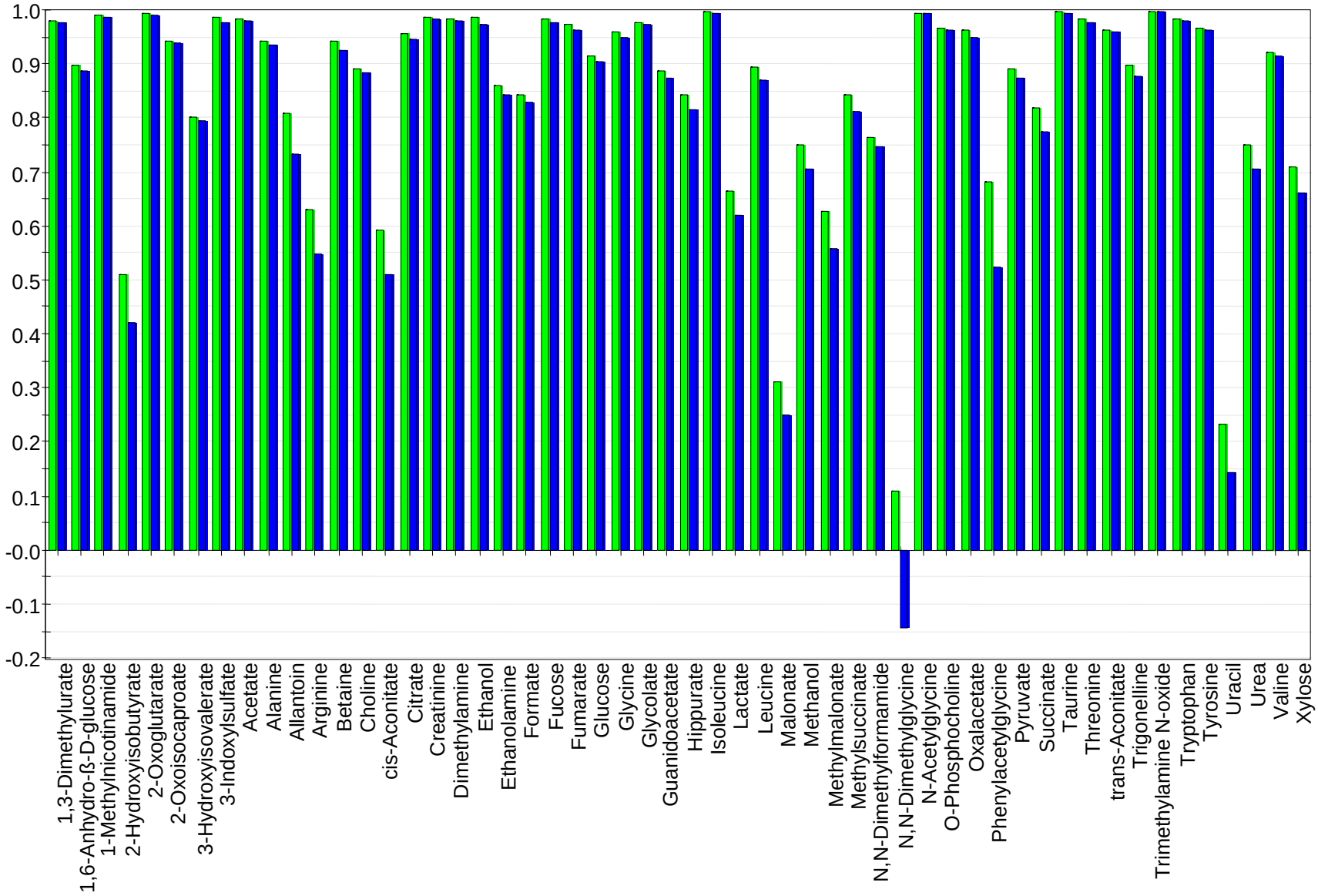
Several metabolites that were affected by overlap or could not be accurately quantified identified

$R^2 < 0.6$ = 5 of 54

$R^2 < 0.8$ = 14 of 54



Results - 3



PLS model goodness-of-fit (R^2) and goodness-of-prediction (Q^2) metrics for individual metabolite concentrations.

Study – Different Urines

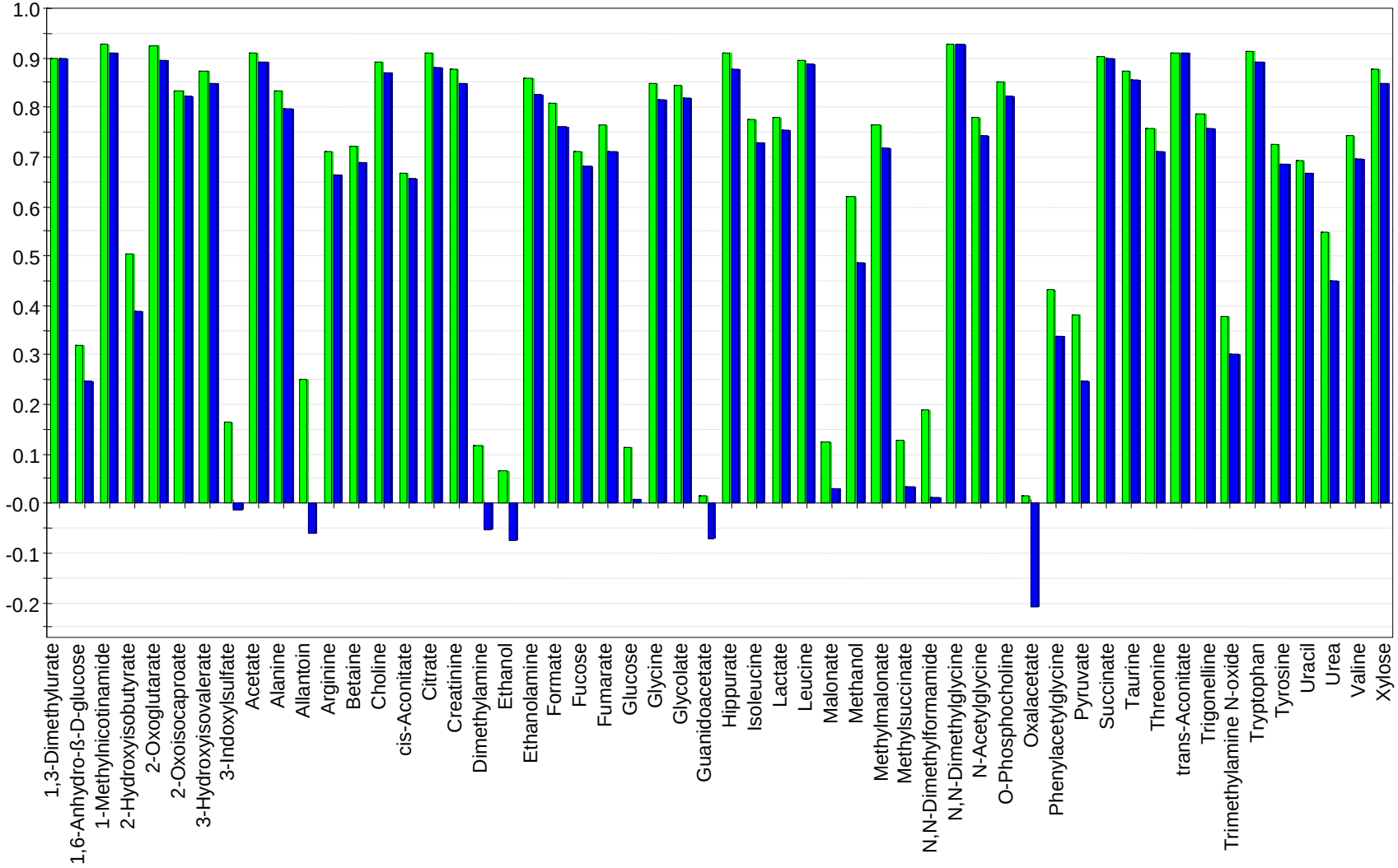
Gross urine concentration varies considerable between samples as a consequence of highly variable dilution factor.

To remove/reduce this effect from the spectra being models, probabilistic quotient normalisation (PQN, Dieterle *et al.* 2006) was applied.

The normalised spectra were regressed against the experimental design.

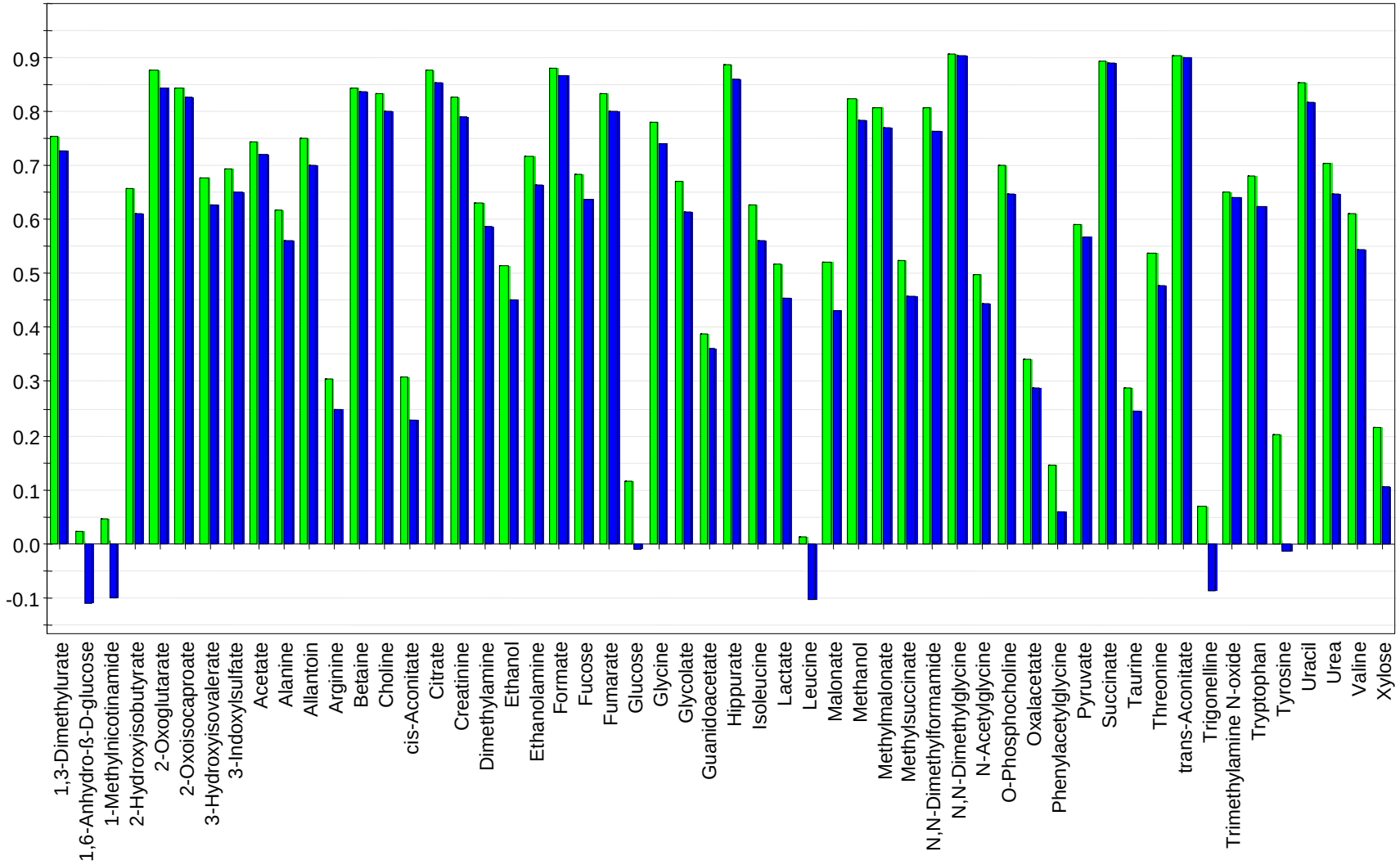
Remaining R^2 values should therefore better reflect how well the metabolite is measured across all samples, without the influence of overall urine dilution.

Targeted Profiling



PLS model goodness-of-fit (R^2) and goodness-of-prediction (Q^2) metrics for individual metabolite concentrations.

Targeted Integration



PLS model goodness-of-fit (R^2) and goodness-of-prediction (Q^2) metrics for individual metabolite concentrations.

Study – Different Urines

Consequences of Normalisation
For the analysis of the ‘Targeted Profiling’ approach, variation in most metabolites still strongly associated with the differences between the original samples,

A greater number of metabolites with lower performance were identified:

$R^2 < 0.6$ = 16 of 54
$R^2 < 0.8$ = 31 of 54

Comparison of Approaches
An equivalent analysis using a ‘Targeted Integration’ approach to represent the same metabolites indicated that such a lower distribution of

$R^2 < 0.6$ = 19 of 54
$R^2 < 0.8$ = 38 of 54

Study – Different Urines

By being able to predict the relationship between measurements on different samples, each with a different background, we were able to **estimate the reliability of quantification** – based on explained variance (R^2) - on **a metabolite-by-metabolite basis**, and to refine the set of metabolites that we could safely interpret from the urinary data.

There was broad agreement between ‘Targeted Profiling’ and ‘Targeted Integration’ approaches.

Targeted Profiling has an advantage in quantifying metabolites with **peaks in overlapped or baseline-dominated** spectral regions.

Targeted Integration has the **advantage of being very rapid**, and requiring minimal repetitive, time-consuming spectral fits to be made (a process that has not been entirely automated yet).

Using designed biofluid mixtures provides a rational basis for exploiting information from several samples whereas spectral deconvolution is typically applied to one spectrum at a time. Thus, the strategy may have general benefits for quantitative analysis in biofluid metabolic profiling.

References

1. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky C M: Targeted profiling: quantitative analysis of ^{1}H NMR metabolomics data. *Anal. Chem.* 2006. 78, 4430-4442.

2. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. *Multi- and megavariate data analysis. Principles and applications.* Umetrics AB, Umea, Sweden. 2001.

3. Dieterle F, Ross A, Schlotterbeck G, Senn H: Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Anal. Chem.* 2006. 78, 4281-4290.

Future Work

Evaluate Simple Mixtures

- No matrix
- Spiked into matrix
- Variety of Matrices

Implementation in Future Studies

Start routinely adding in mixtures of biofluids incorporating known experimental groups (e.g. control *vs* treatment A *vs* treatment B)

Investigate Higher-Dimensionality Mixtures

Investigate which experimental designs fit best within a routine profiling context

Suggestions welcome!

Summary

- Metabolic profiles are information-rich phenotypes that can be characterised by spectroscopic platforms such as NMR
- Such platforms provide information on an array of metabolites simultaneously.
- Multivariate analysis of the spectral sets can help derive biomarkers in a variety of contexts (e.g. toxicology, efficacy, disease progression, diet, etc).
- Measurements are often made on complex biofluids with varying background interference from the underlying matrix. Characterising how this can influence the measurements across a sample set can help inform the analyst of their reliability.
- Using a simplex lattice design, we evaluated the performance of NMR spectroscopy to characterise >50 common metabolites using two alternative methods for metabolite quantification.
- We conclude that inclusion of designed mixtures in routine metabolic profiling work may improve our understanding of which putative markers are reliable.

Acknowledgements and Collaborators

Biomolecular Medicine

- Prof Jeremy Nicholson
- Prof Elaine Holmes
- Prof John Lindon
- **Dr Hector Keun**
- Dr Tim Ebbels
- Dr Muireann Coen
- Dr Jake Bundy
- Dr James Ellis
- Dr Rachel Cavill
- Dr Olaf Beckonert
- Dr Jake Pearce
- Dr Anas Kamleh

Epidemiology & Biostatistics

- Prof Paul Elliott
- Prof Paolo Vineis
- Dr Anna Hansell
- Dr Claus Svendsen
- Dr David Spurgeon
- Dr Marta Blangiardo
- Dr John Gulliver

IARC

- Prof Chris Wild
- Dr Mazda Jenab
- Prof Pierre Hainaut

NHRF

- Soterios Kyrtopoulos

U. Maastricht

- Dr Theo de Kok

U. Utrecht

- Dr Roel Vermeulen

SeRMN

- Dr Teodor Parella
- Dr Miriam Perez-Trujillo
- Mr Sergi Gil

Chenomx

- **Aalim Weljie**
- Jack Newton
- Shahid Malik

Funding and Support

THE ROYAL SOCIETY

CHENOMX

carcinoGENOMICS
a Project of the European Union

envirogenomarkers

NATURAL
ENVIRONMENT
RESEARCH COUNCIL

http://www.environment-health.ac.uk