Validated Biomarkers of Caloric Restriction in Rats: Markers of Disease Risk in Humans?

Bruce S. Kristal, PhD

Dept of Neurosurgery, Brigham and Women's Hospital
Dept of Surgery, Harvard Medical School

Industry Affiliations (Disclosure):
ESA (collaborative, inactive); Mead Johnson (Consultant)
Metabolon: Consultant/Equity, (IP agreements via/Cornell Med)

“Caloric Intake” and Disease

Humans:

Increased BMI associated with increased risk of neoplasia, type II diabetes, cardio- and cerebro-vascular disease …

Laboratory Rodents:

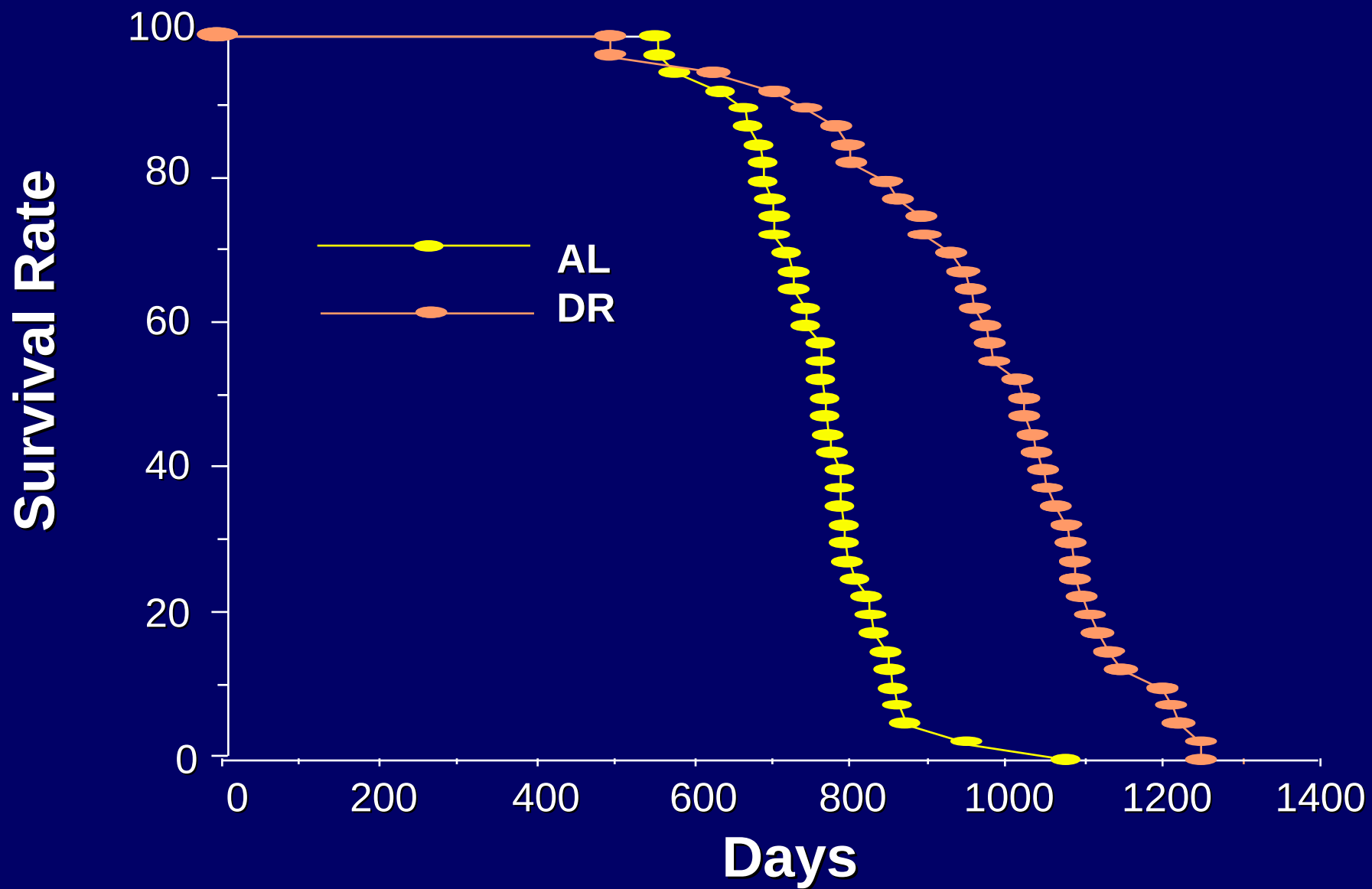
Low calorie diets increase longevity and delay morbidity

Caloric Restriction (CR)

CR is an experimental paradigm in which the dietary/caloric intake of a group of animals is reduced relative to that eaten by *ad libitum* fed controls

Caloric restriction is the most potent, most robust, and most reproducible known means of reducing morbidity and mortality in mammals

Survival Data, 1987 Cohort, Casein Diet



How do we study complex biological/clinical problems?

How do we address such questions in humans, where our ability to manipulate and analyze the system is limited?

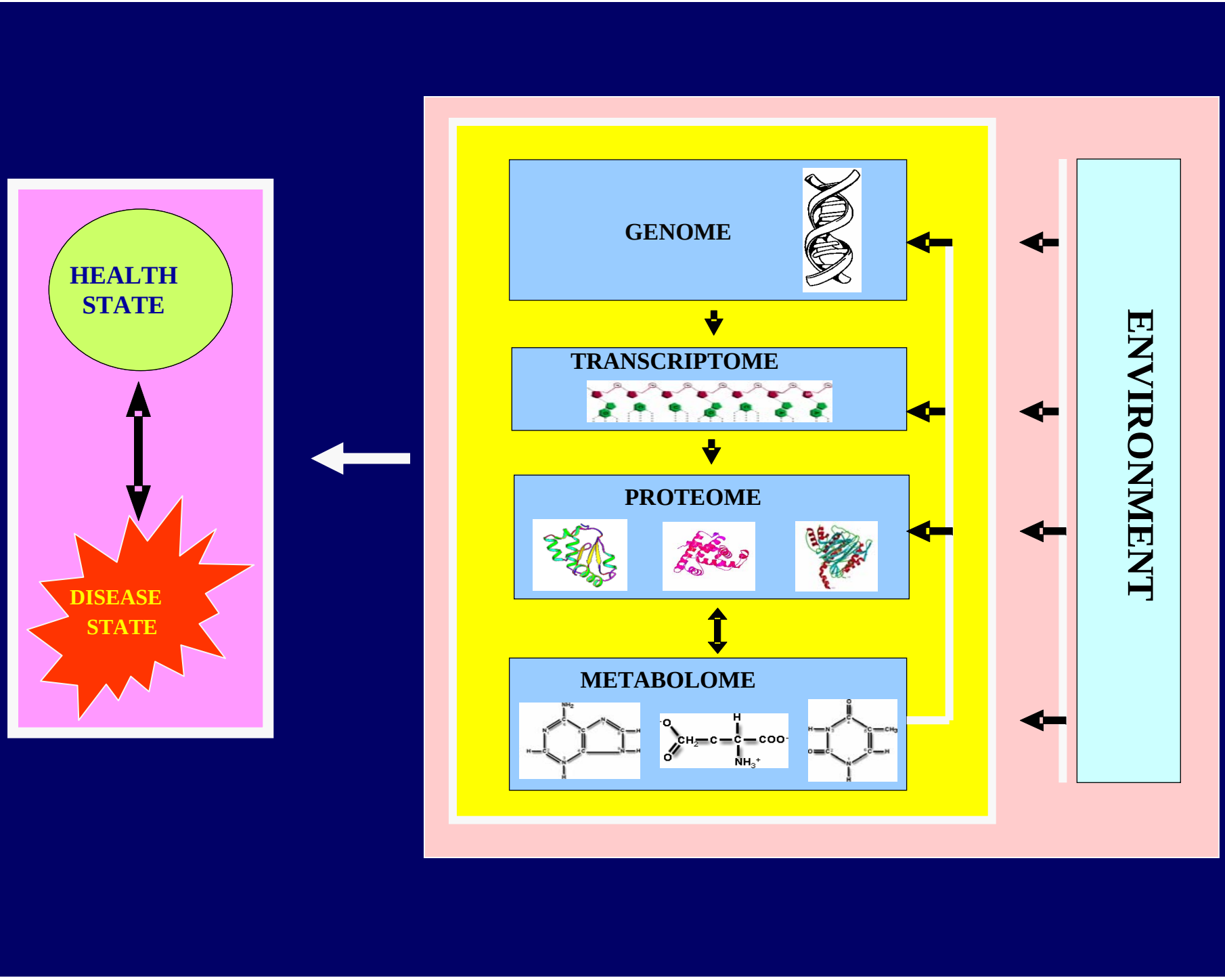
High Throughput and/or Data Density Studies

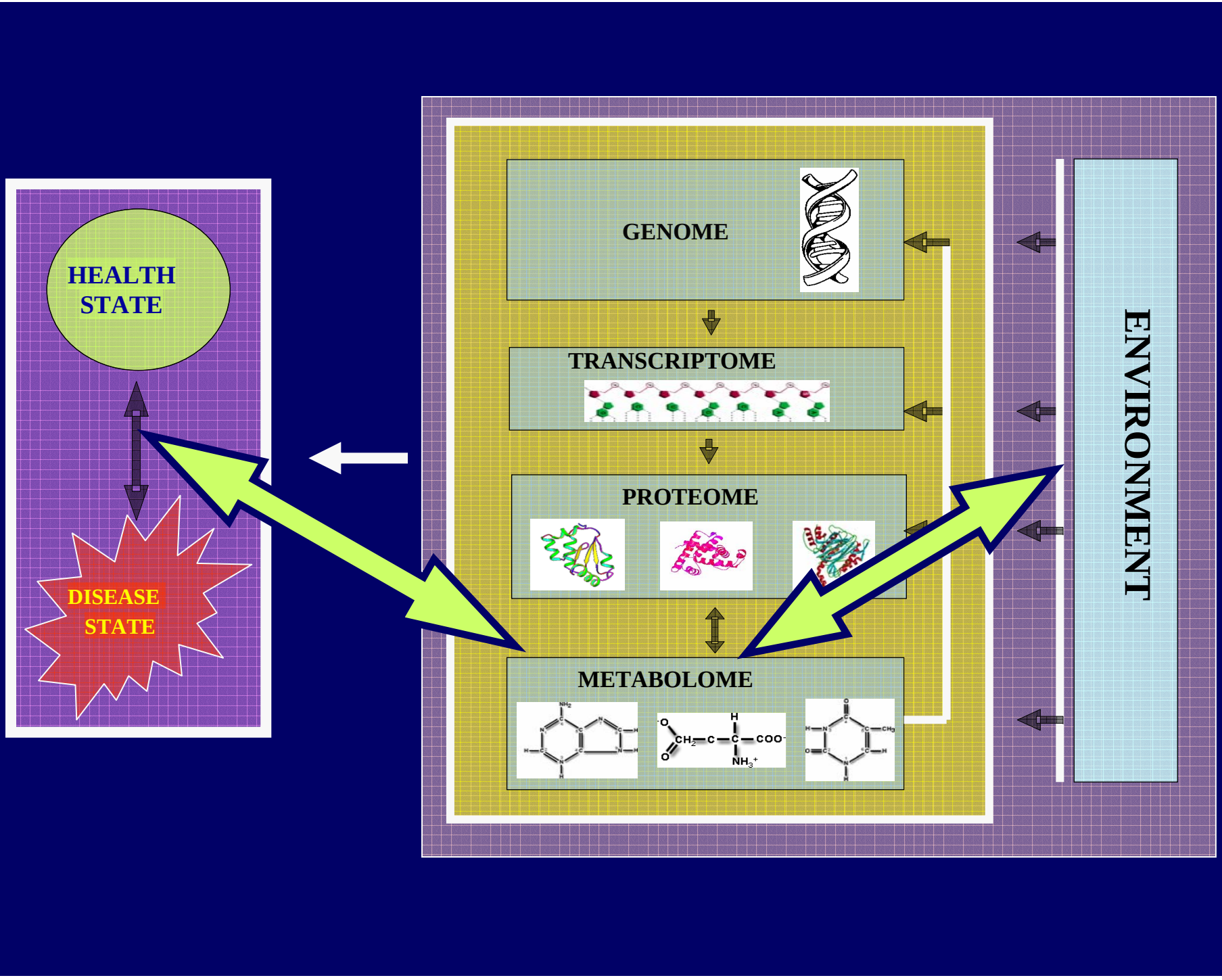
- Genomics/SNPs
- mRNA expression arrays
- Proteomics
- Small metabolites

Metabolomics:
The –omics face of biochemistry

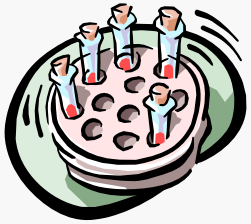
Measurement of changes in populations of low molecular weight metabolites under a given set of conditions

Fiehn

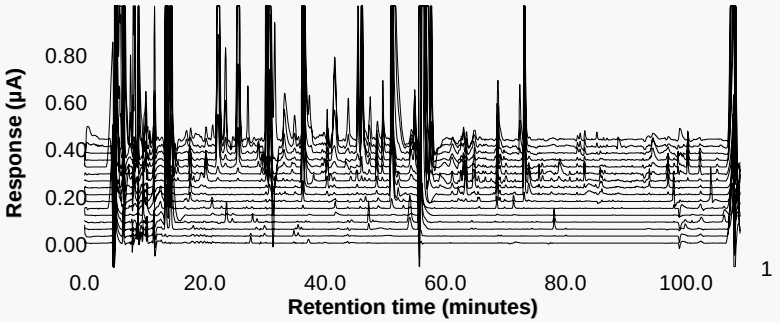

HEALTH STATE
DISEASE STATE
GENOME
TRANSCRIPTOME
PROTEOME
METABOLOME
ENVIRONMENT


HEALTH STATE
DISEASE STATE
GENOME
TRANSCRIPTOME
PROTEOME
METABOLOME
ENVIRONMENT

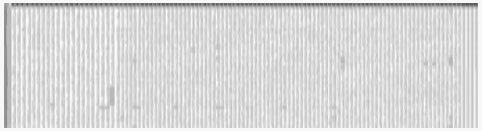
Sample Collection



Sample Analysis

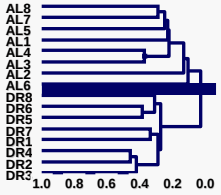


Database Curation

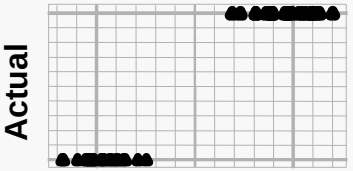


Mechanistic Insight
Drug Development
Toxicology
Classification
Prediction
Functional genomics
Sub-threshold studies
Others

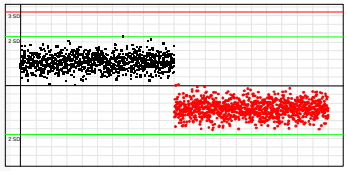
Objectively Defining Class Identity



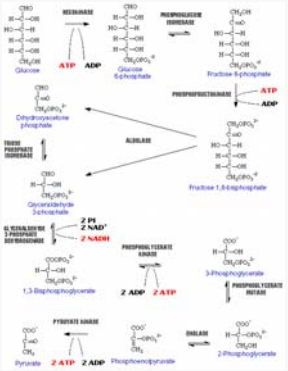
Observed Values vs. Predicted Values



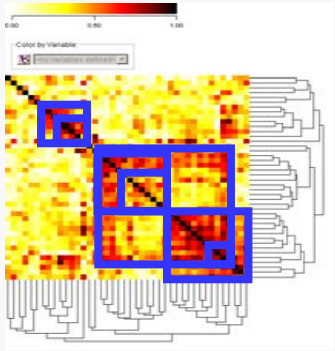
Computational Modeling of Metabolic Serotypes



Following Biochemical Pathways



Modeling Metabolic Interactions



Bioinformatics

What we measure -- biochemically

Metabolites – small molecules

Pathways (eg, purine catabolites)

Interactive pathways (eg, amino acid metabolism)

Compound classes (eg, lipids)

Conceptually linked systems

eg antioxidants, redox damage products

What we measure -- conceptually

Biochemical constituents

Excretion products

Precursor – product

Balances (eg, redox systems)

“collection depots”

Flux

Snapshot view of biochemistry

Integrated signal from genome and environment

Short and long term status

Temporal image

Sub-threshold changes (eg (toxicology, nutrition)

Metabolomics – Some Advantages

Sensitivity

- **"silent phenotypes"/sub-threshold effects**
- **Discovery**

Knowledge base (ie, metabolic pathways)

Limited repertoire – simplifies possibilities

- **(2500 non-lipid endogenous metabolites??)**

Metabolome integrates signal

- **Nature and Nurture -- genome and environment**
- **Measurement of system status/defects**

Metabolome has the fastest response time

Metabolomics – Some Disadvantages

Too Sensitive?

- cohort effects, site effects, time effects
- sample handling
- individual metabolites responsive to multiple factors
  - genes, environment, health status, location
- experiment design must account for all factors
  - controlled or fuzzy, multiple sources

Practical

- Set-up costs
- Possible need for multiple platforms (NMR, MS, HPLC)
- early industry dominance – lots of propriety data
- incompatible data standards

Informatics

Data Validation, Data Normalization,
Missing Data Decisions, Inclusion/Exclusion Criteria

↓

Subgroups, Class-specific models

↓

Outlier removal ←→ scaling ←→ transformations

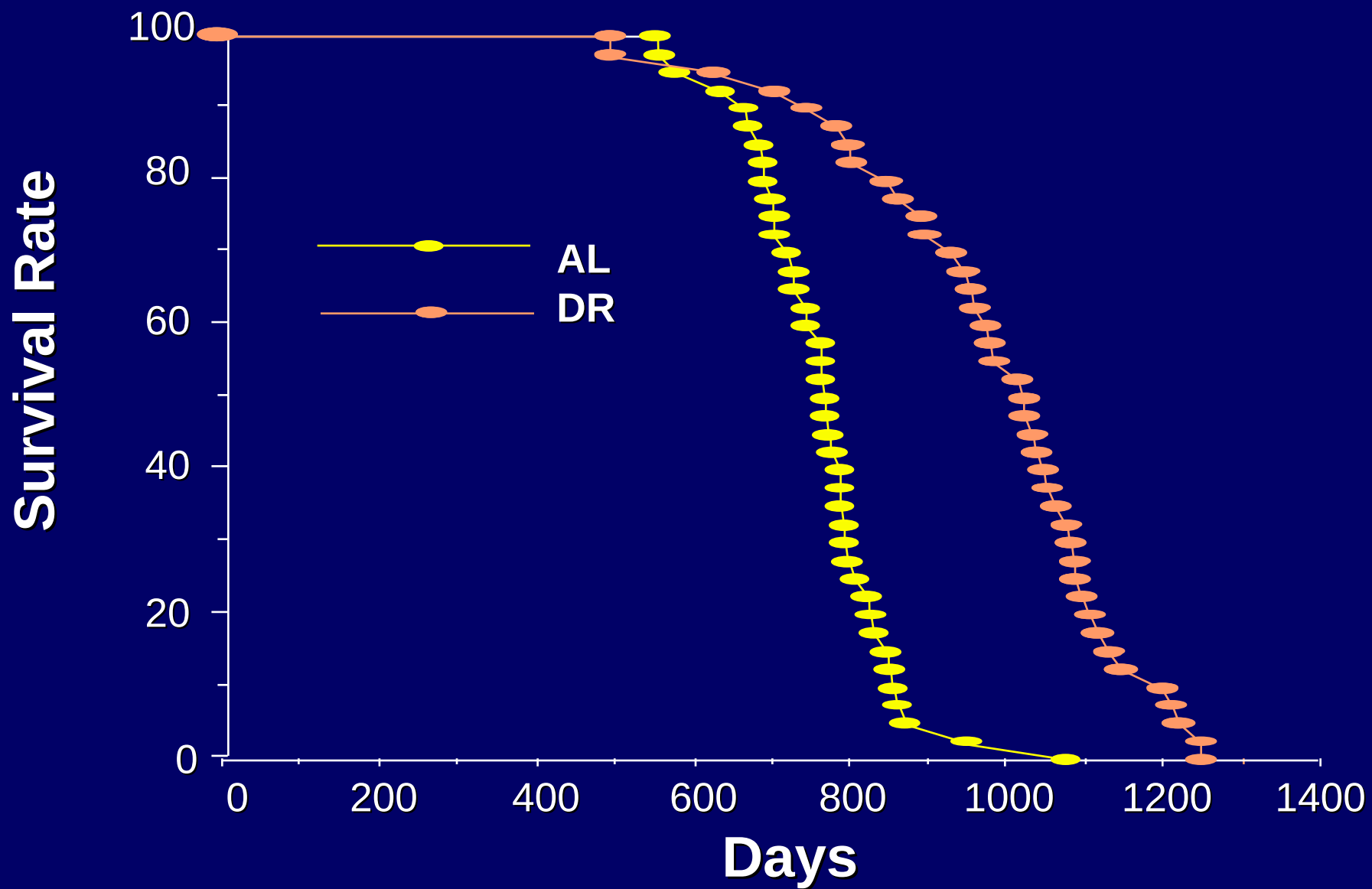
↓

Unsupervised: Clustering SOMs PCA
Supervised: kNN SIMCA PLS PLS-DA Random Forest
Machine learning: Neural Nets GAs GPs

↓

Overfit tests, Internal validation, optimization,
External validation, optimization, 2^{o} validation

Survival Data, 1987 Cohort, Casein Diet



Hypothesis:

Long-term, low-calorie diets induce changes in metabolism that persist throughout the lifespan

Predictions

- CR alters the sera “metabolome”
- There exists a “CR Serotype”
- …Part of “CR serotype” reflects beneficial physiological status --- ie, serotype defines health without reference to disease…

Goals

1) Insights into the mechanism of CR
2) Recognize CR in other organisms

   (e.g., non-human primates)

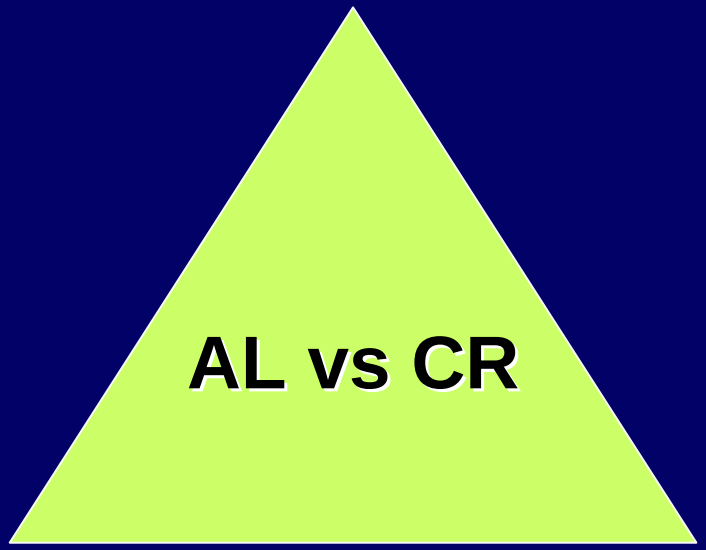
3) Biochemically determine the effective, long-term caloric intake of an individual

   (e.g., for epidemiological studies)

4) Identify predictive markers of disease

   (e.g., to intervene/prevent/focus resources;
   focus on diseases where intervention is possible)

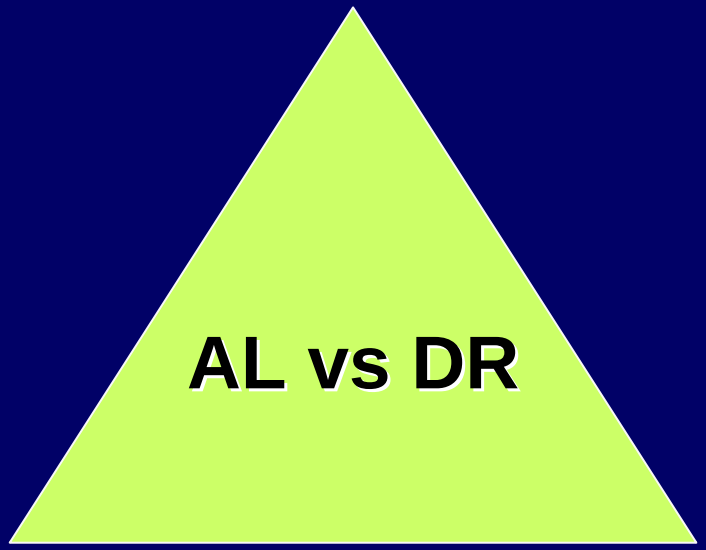
Experimental Design



Analytical Issues

Biological Issues

Experimental Design



Analytical Issues

Biological Issues

Experimental Design

Model: F344 x BN F_1 Rat

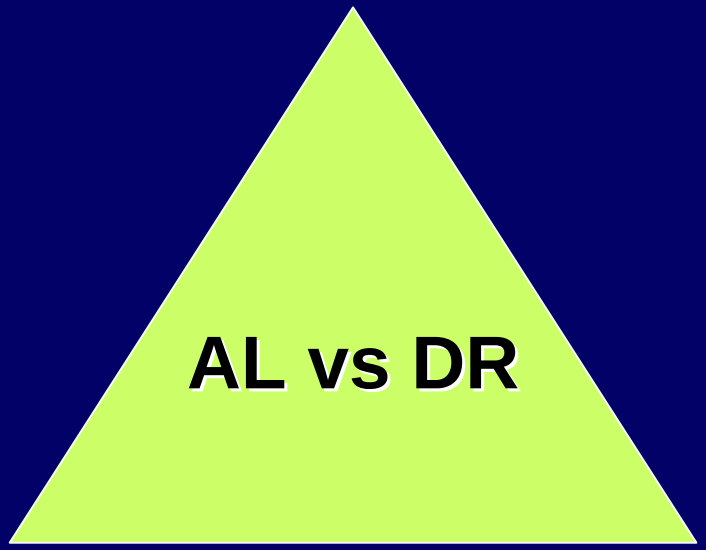
Overall Design:

AL/CR, male/female, 5 different ages
Different extents and duration of diets
Total experiment ~36 groups, 82 cohorts.

Approach:

HPLC separations with coulometric array detection
(LC/LC-MS for plasma proteomics)
Multilayer statistical and data analysis

Experimental Design



Analytical Issues

Biological Issues

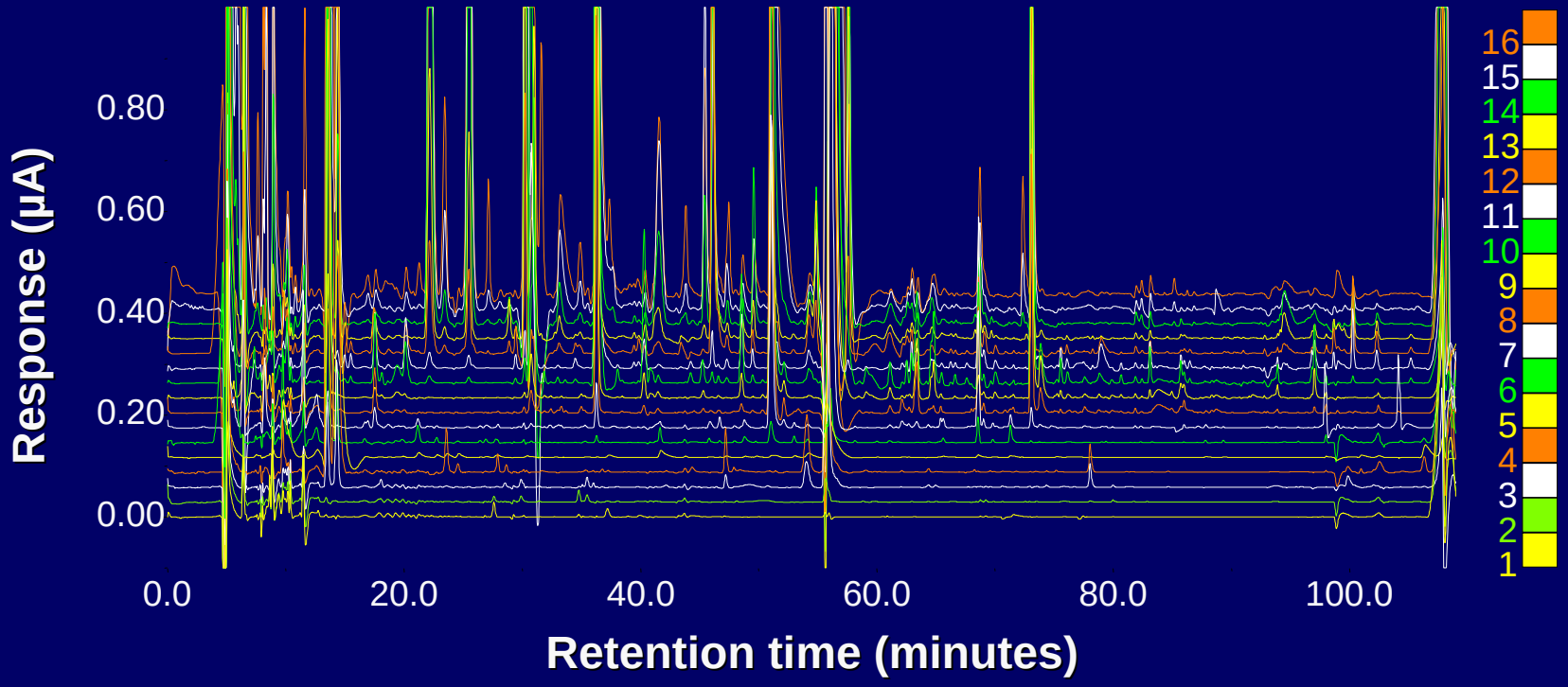
Coularray

- HPLC separations coupled with coulometric array detectors

  - Sensitivity to femtomole levels of analyte
  - Resolution of co-eluting peaks
  - Qualitative characterization of peaks
    - Biochemical identity
    - Purity

HPLC-EC on Pooled Rat Sera

1075 analytically detectable peaks
sensitivity ~300 pA = ~10 fmole/125 μl sera

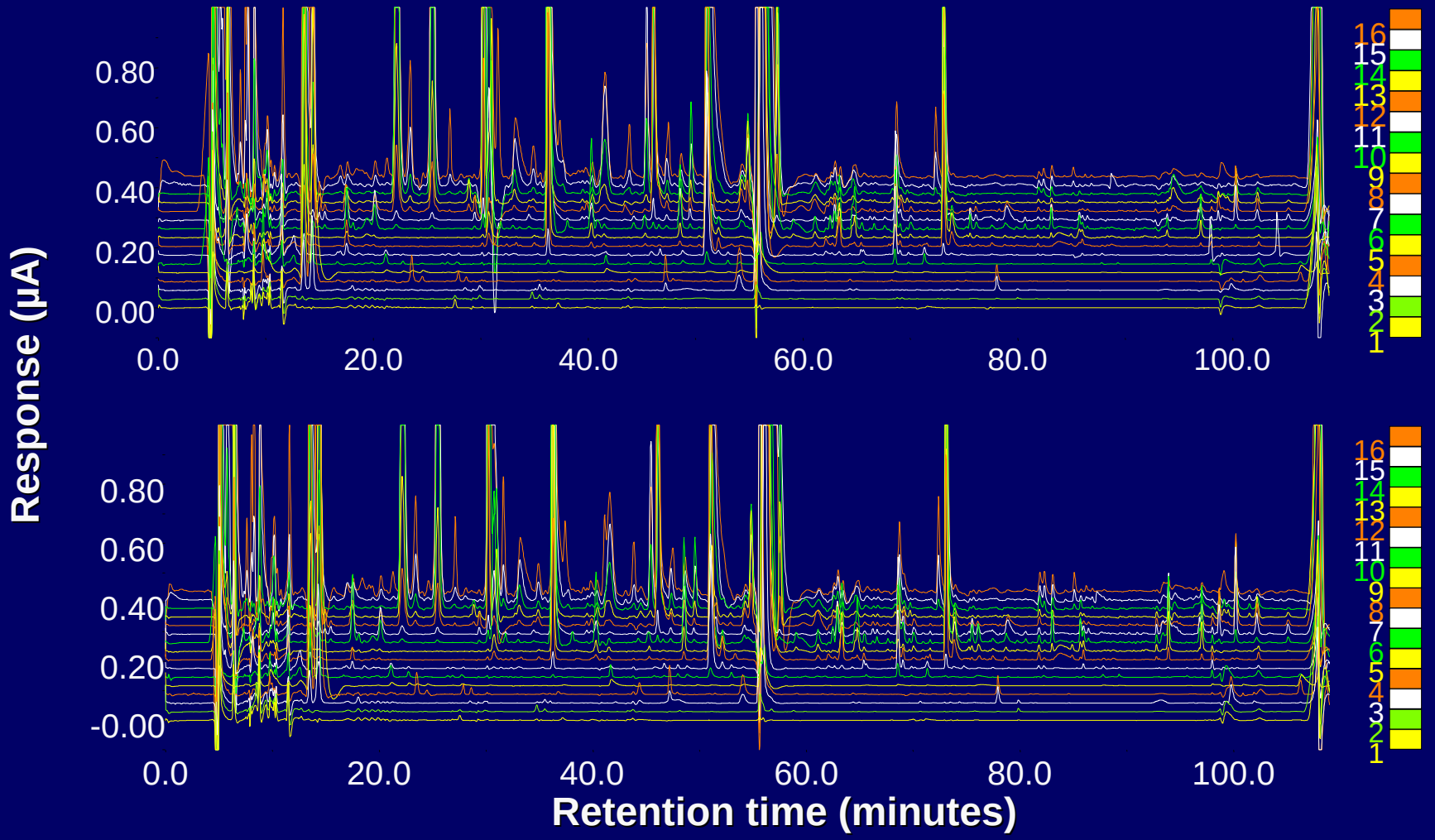


Analytical Stability

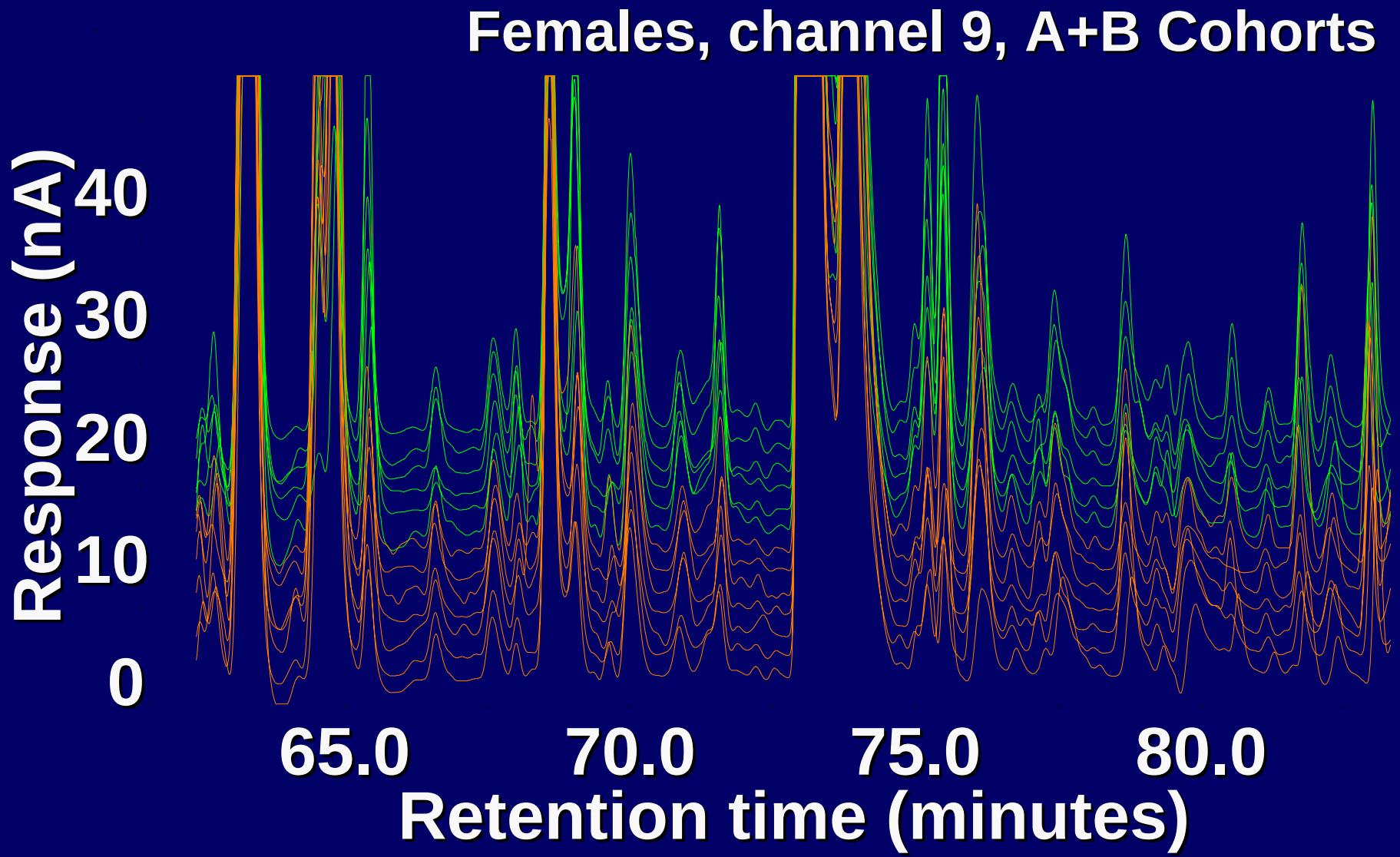
Biologic Variability

Analytical Stability/Biological Variability

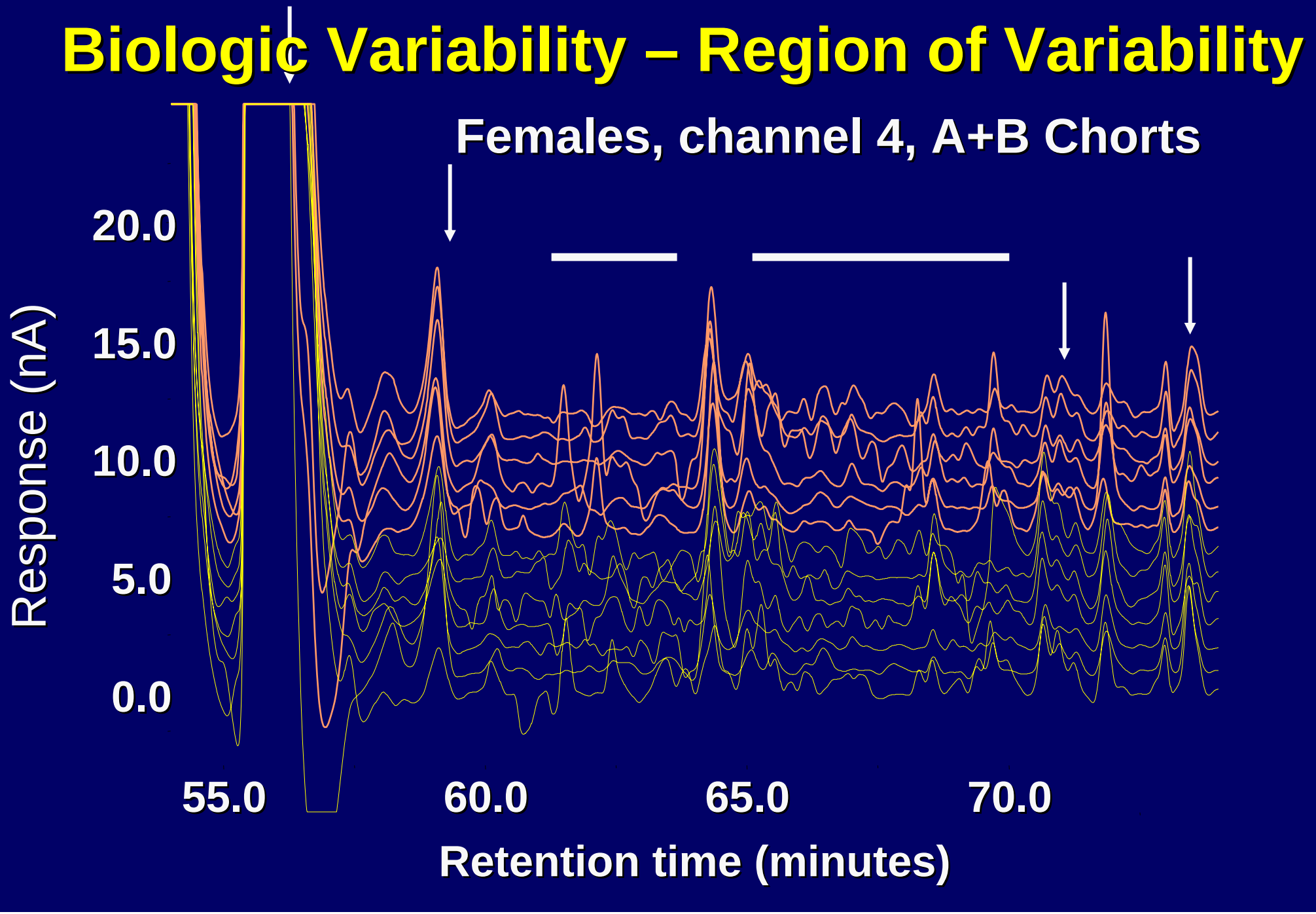
Females--Cohort A + B – Rat sera looks like rat sera



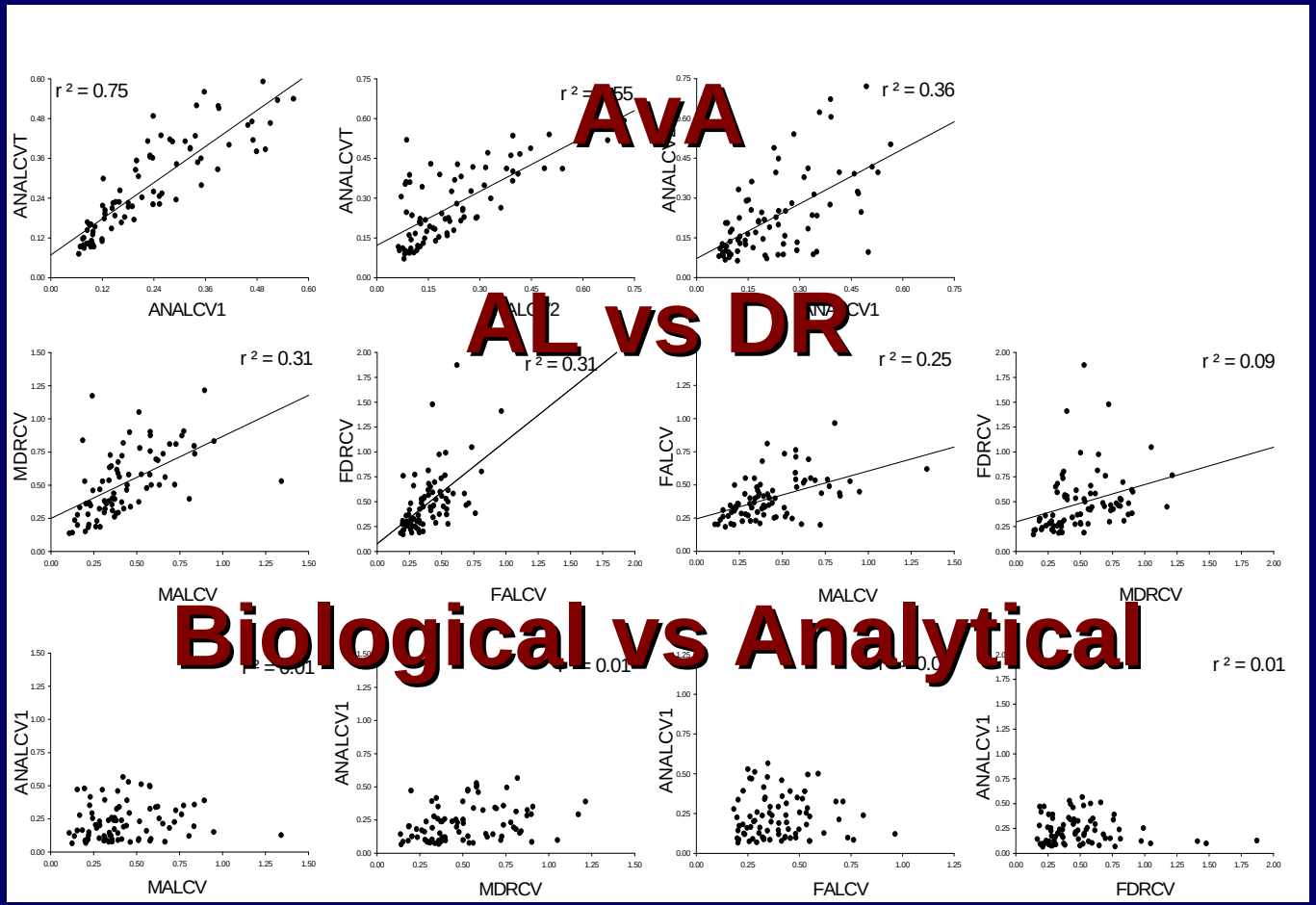
Biologic Variability – Region of Stability



Biologic Variability – Region of Variability



Analytical vs Biological Variation



In Rats:

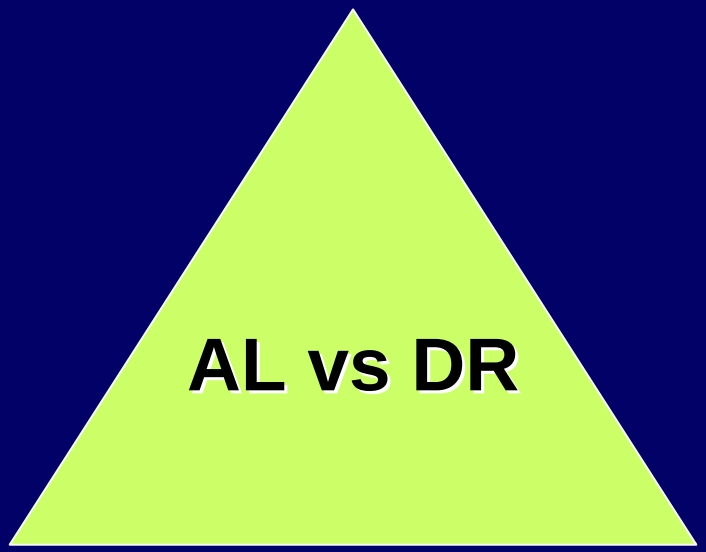
Biological variability 5 fold greater than analytical variability

Analytical variability does not influence biological variability

Primary Data Analysis

- Multivariate analyses are relatively noise-resistant
- Minimize loss of informative metabolites
  - Reduce false negatives (Type II errors)
  - Increase false positives (Type I errors)

Experimental Design



Analytical Issues

Biological Issues

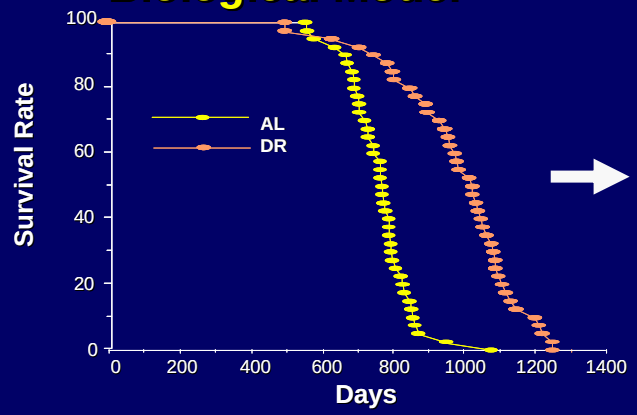
Does Serotype Encode Sufficient Information to Identify Diet Group?

Data Exploration and Classification Analysis

- **Hierarchical Cluster Analysis (HCA)**
  - **Identifies natural groups in data**
- **Principal Component Analysis (PCA)**
  - **Finds linear combinations of original variables that account for maximal variation**

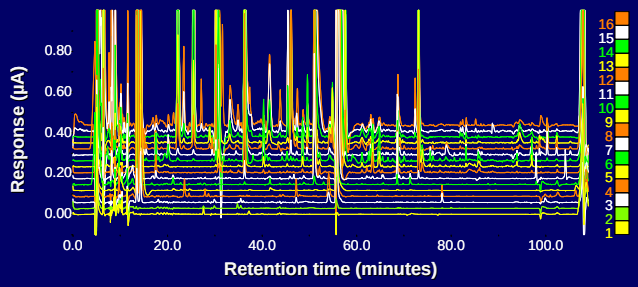
Model

Biological Model

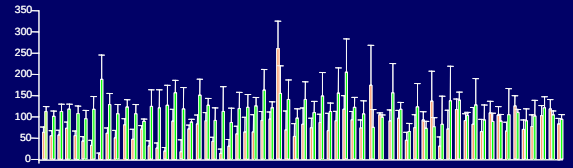


Feature Selection

1075 analytically detectable peaks
sensitivity ~300 pA = ~10 fmole/125 µl sera



T-tests, p<0.2 ?!

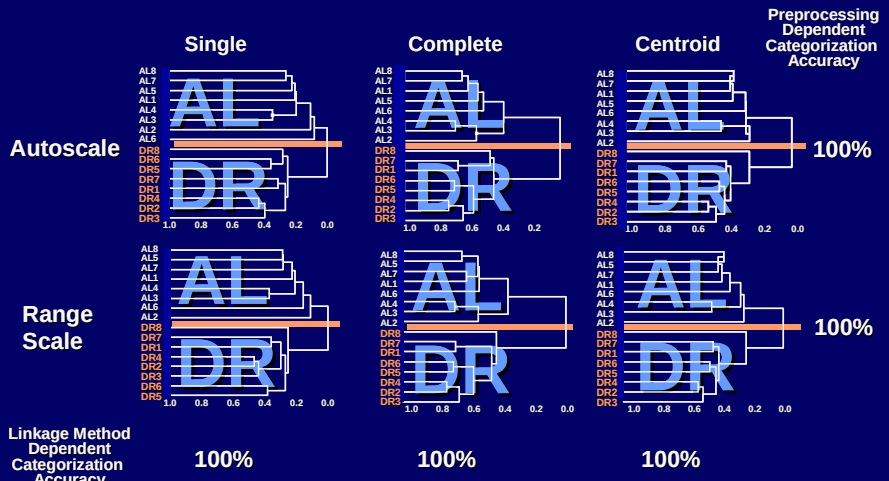


HCA

Proof of Principle

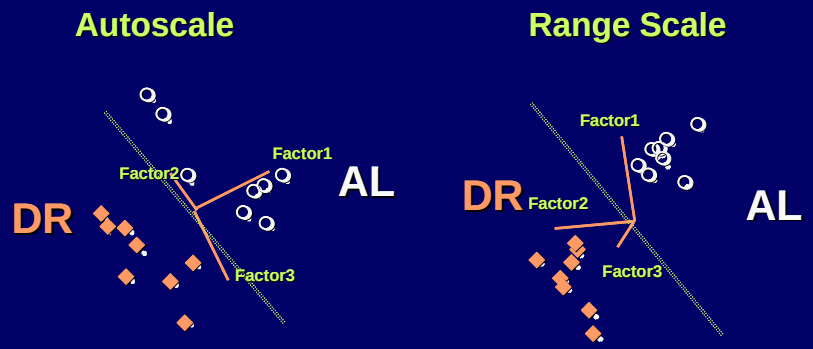
PCA

HCA Distinguishes Female AL and DR Rats



63 variables (confirmed), Non-independent samples

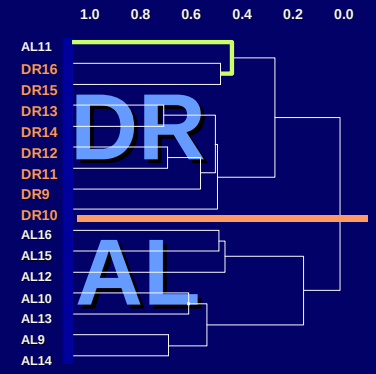
PCA Distinguishes AL and DR Female Rats



63 variables (confirmed), Non-independent samples

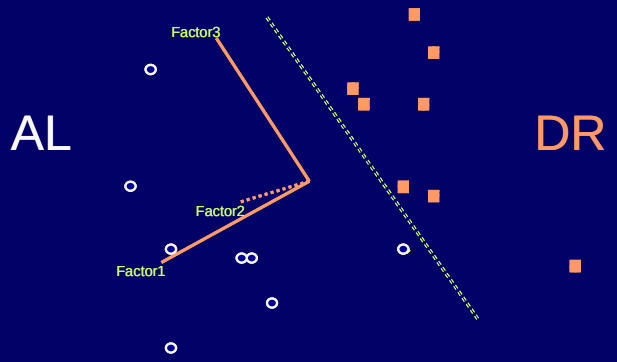
HCA Validation PCA

Complete



94% Accuracy

PCA Distinguishes AL and DR Female Rats



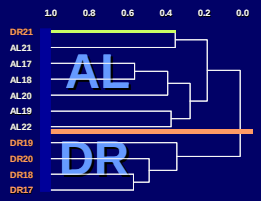
63 variables (confirmed), independent female Cohort #2

HCA Simplify Model PCA

Removed variables can contribute to AL-DR separations

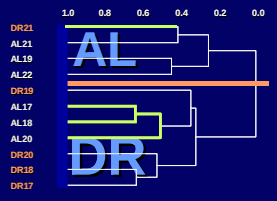
Autoscale, complete

63 Variables



91% Accuracy

36 Variables

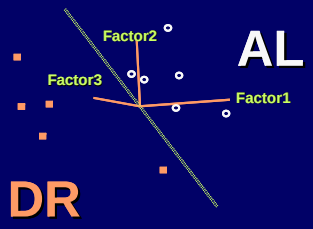


63% Accuracy

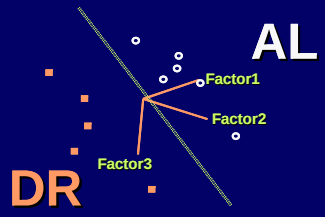
63/(37/36) variables (confirmed), Independent female Cohort #3

PCA Distinguishes AL/DR Using Either Dataset

63 Variables



36 Variables



37 variables (confirmed), Independent female Cohort #3

Status

Proof of principle accuracy:

- HCA (100%)
- PCA (100%)

Validation Accuracy:

- HCA (94%)
- PCA (100%) - subjective rotation

Simplification –

- HCA (Fails)
- PCA (100% Accuracy)

Use larger models?

Test components vs distance

“Expert Systems/Supervised Analysis”

KNN

- k-nearest neighbor analysis
- Supervised HCA (HCA is KNN with K=1)
- Distance-based metric
- Strength is with small (training) datasets

SIMCA

- Soft Independent Modeling of Class Analogy
- Supervised PCA
- Component-based metric
- Strength is modeling flexibility (eg, group-specific interactions)

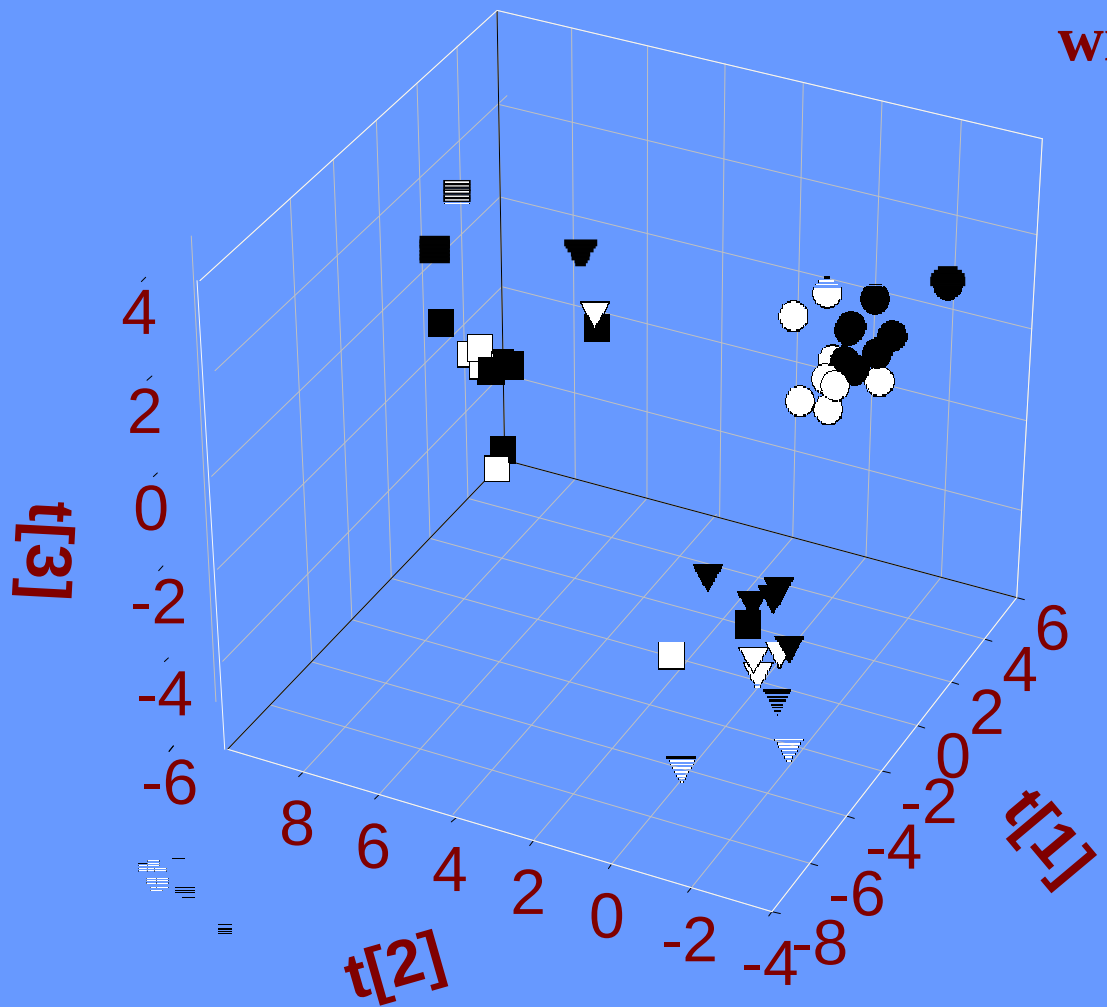
In our DR sera metabolomics data – components greatly outperform distance-based algorithms

In **OUR DR SERA METABOLOMICS** data – components greatly outperform distance-based algorithms

Profiles are cohort specific

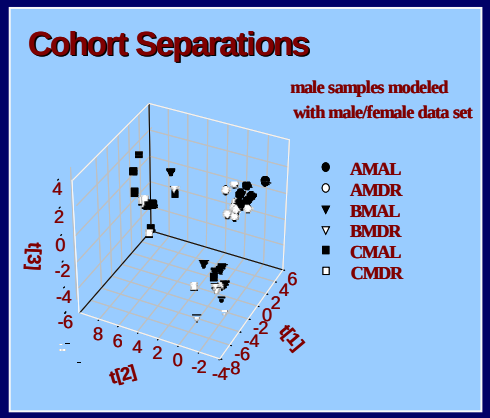
Cohort Separations

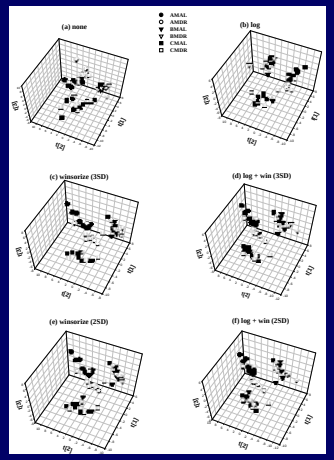
male samples modeled with male/female data set

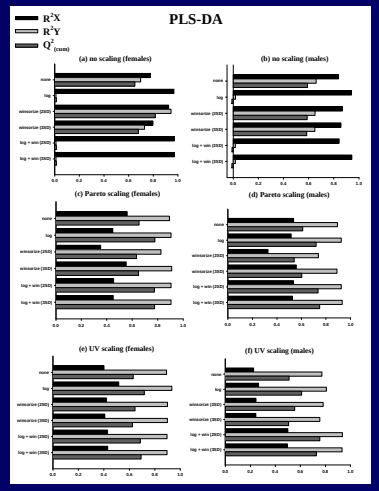


- AMAL
- AMDR
- BMAL
- BMDR
- CMAL
- CMDR

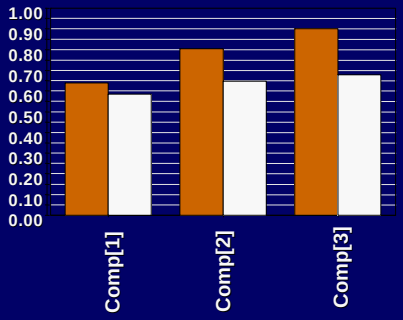
Cohort Effects



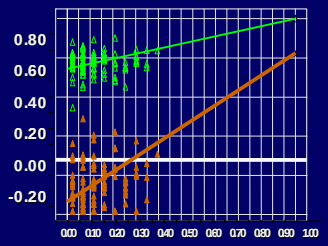


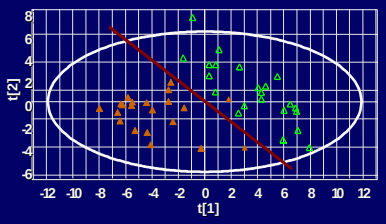


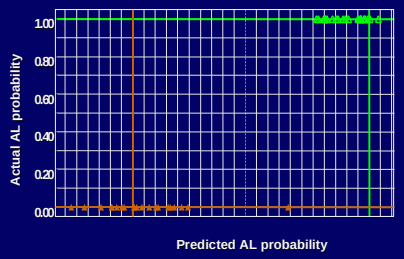
PLS-DA

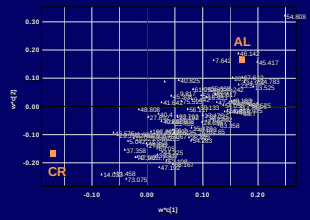


p<0.001

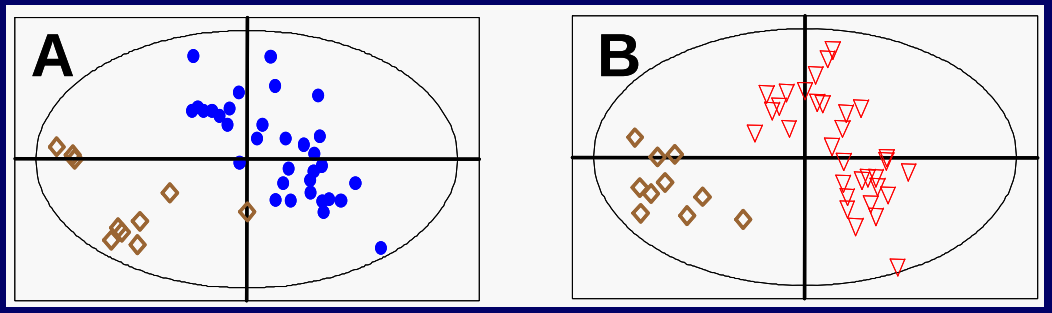




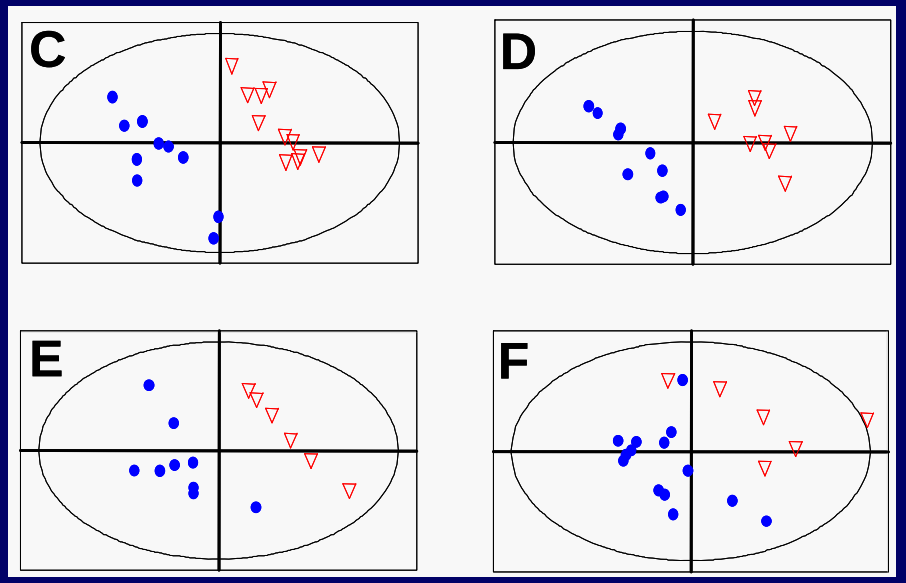




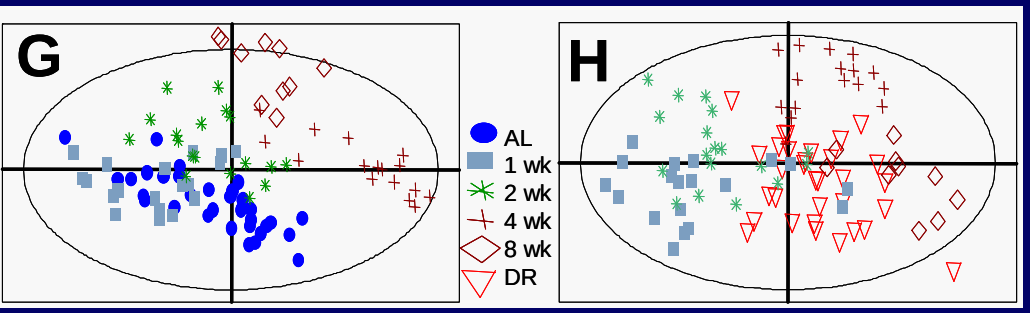
Unpublished Rat Work – Proof of Concept



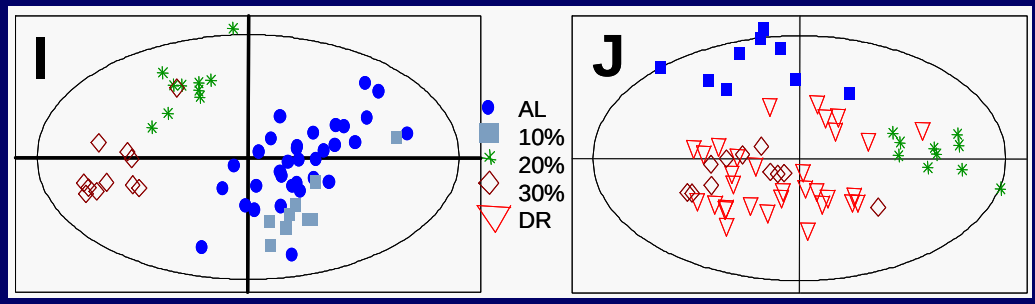
Fasting



Aging

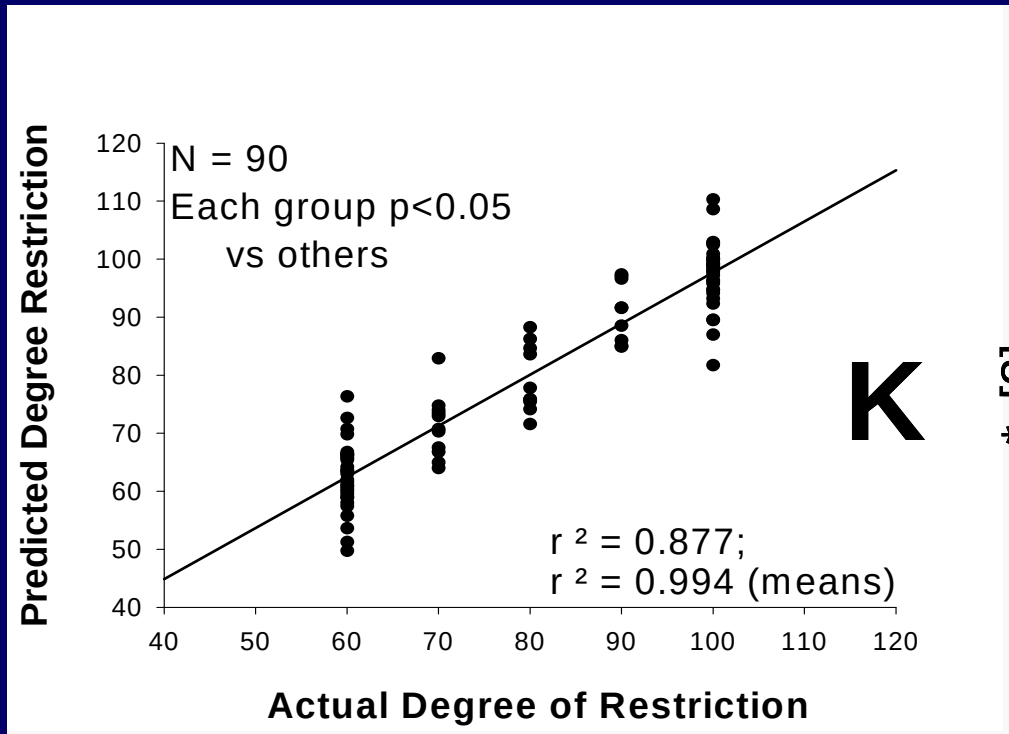


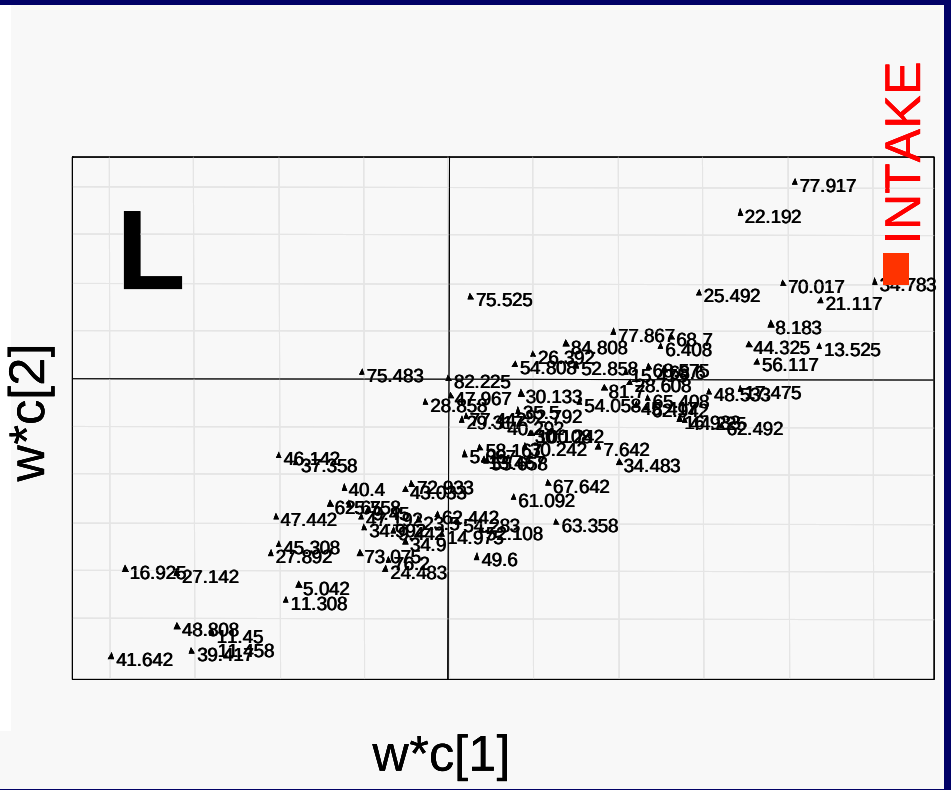
Duration



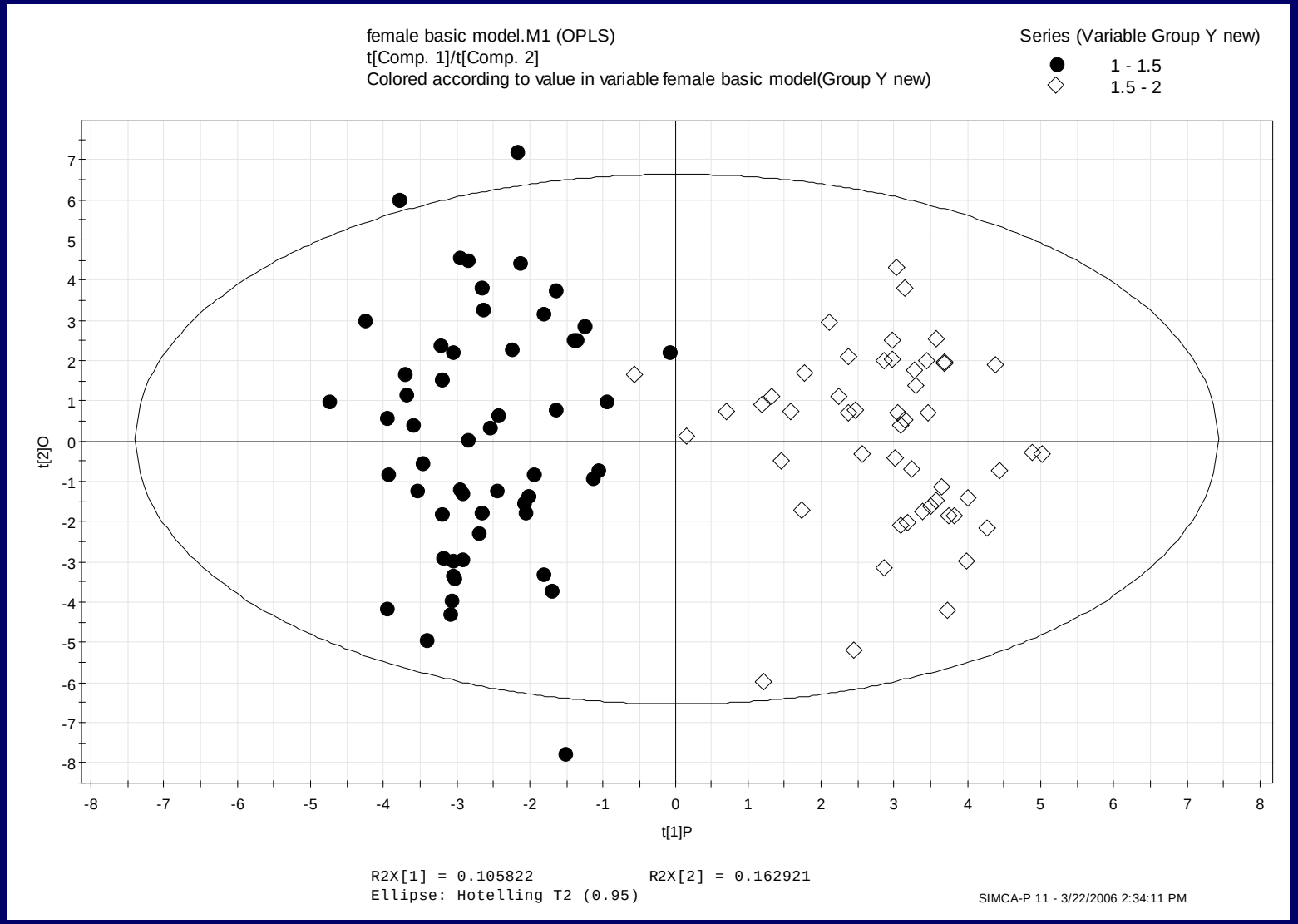
Extent

Markers "Predict" Caloric Intake with High Quantitative Accuracy -- Proof of Concept --

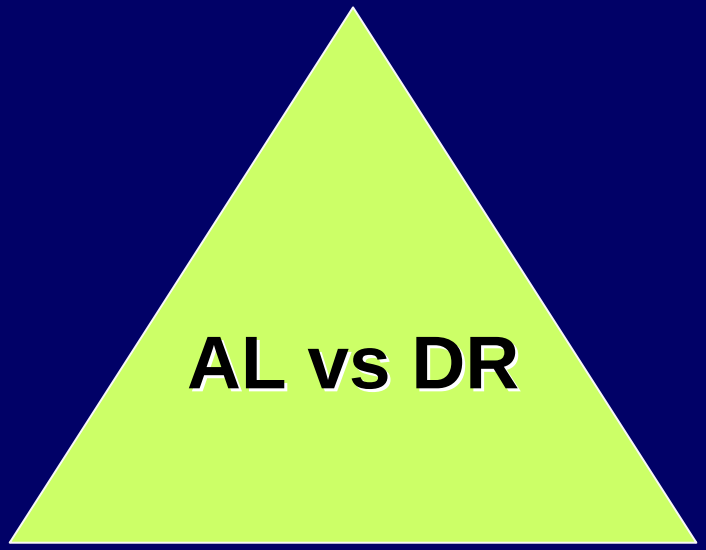




O-PLS models built for better testing



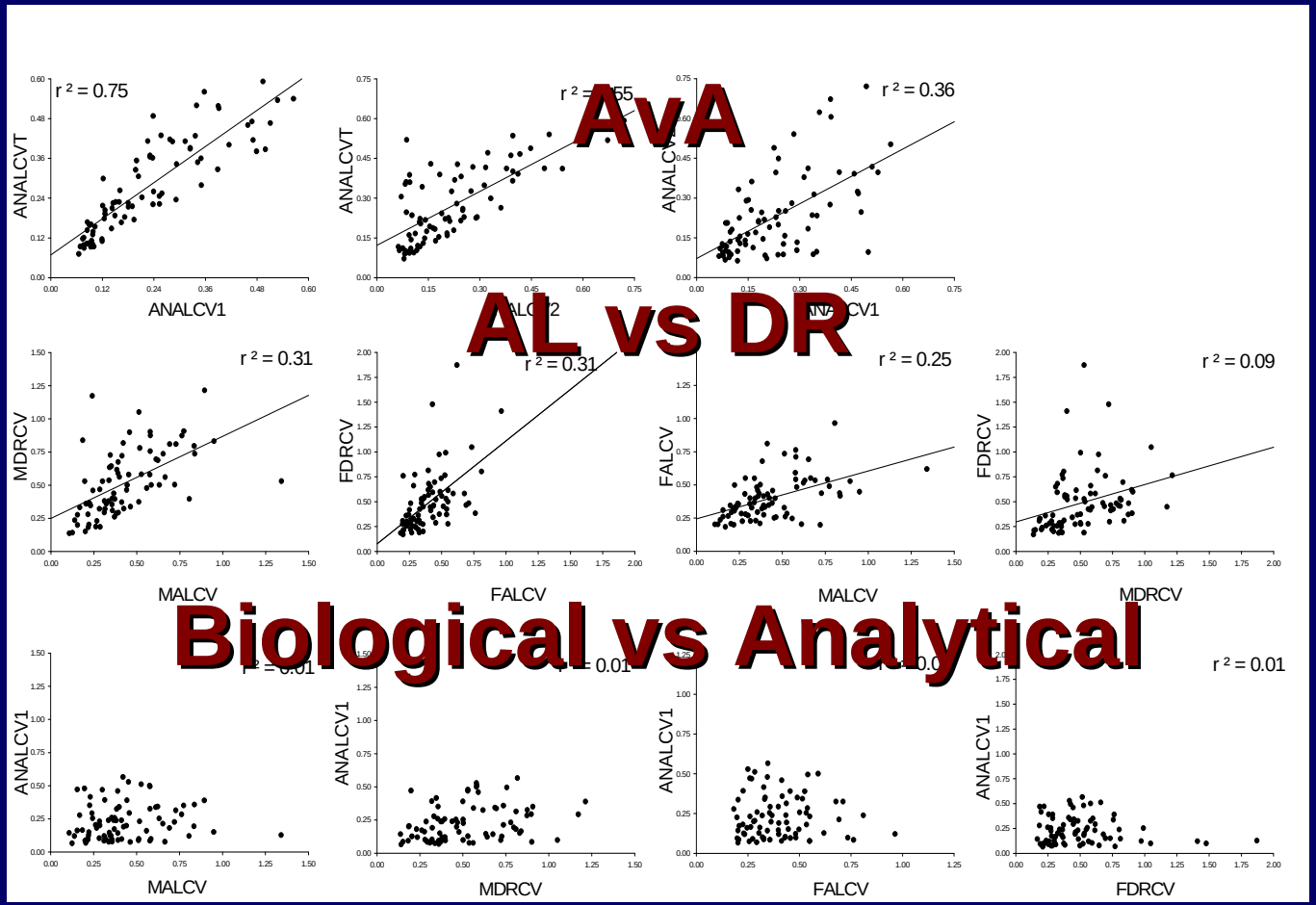
Experimental Design



Analytical Issues

Biological Issues

Analytical vs Biological Variation



In Rats:

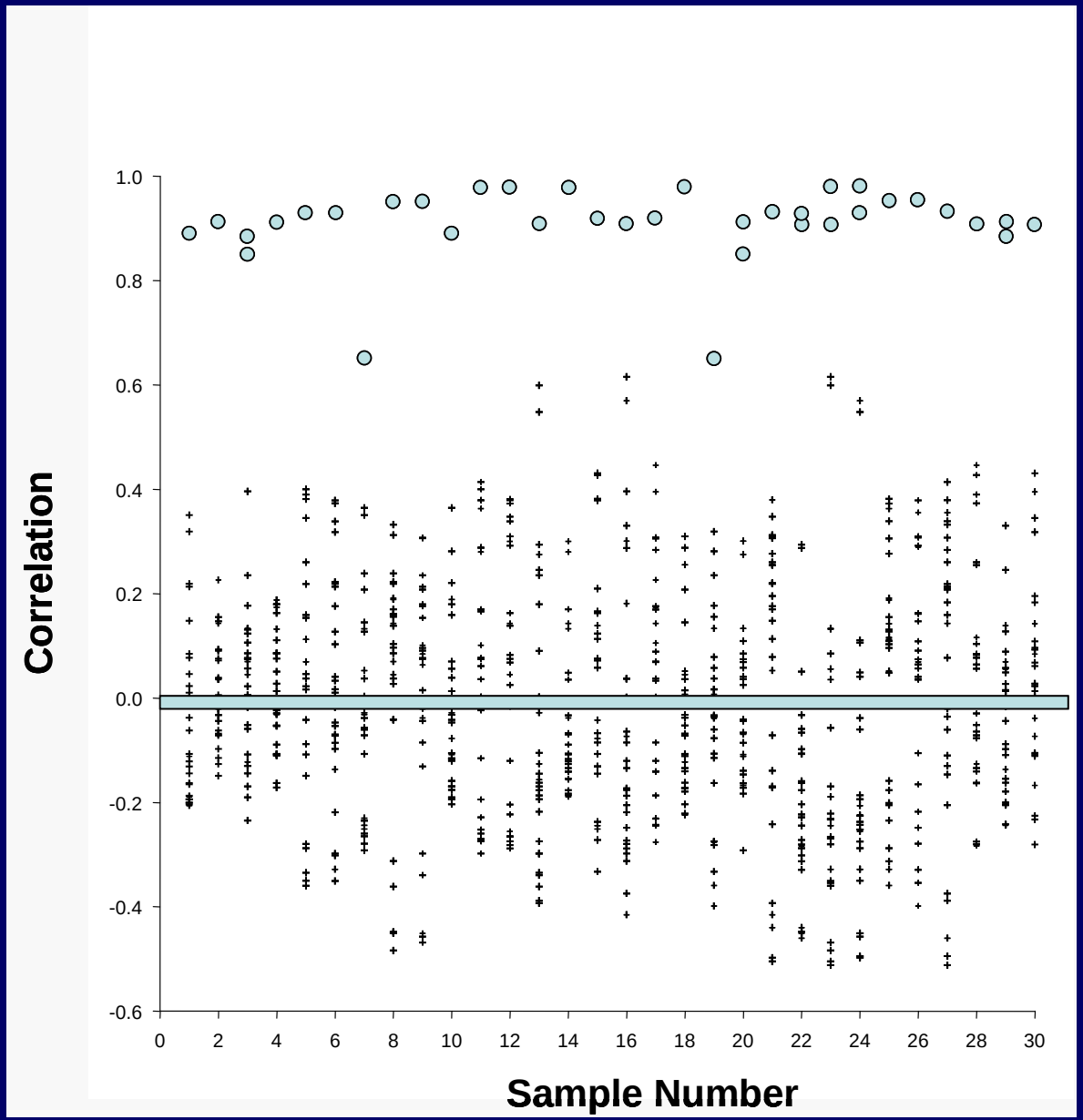
Biological variability 5 fold greater than analytical variability

Analytical variability does not influence biological variability

Human Studies:

Analytical Controls

Duplicates (triplicates)



Analytical Parameters

Means and medians (Key points highlighted)

	pre-polishing data	post-polishing data
Range, median CVs for 66 variables	2-54%	2-32%
Range, mean CVs for 66 variables	2-58%	3-38%
Overall mean CV	19%	16%
Overall median CV	**11%**	**12%**
By 13 pairs, overall mean CV	19%	16%
By 13 Pairs, overall median CV	**9%**	**9%**

3 variables dropped from further study

Analytical Parameters

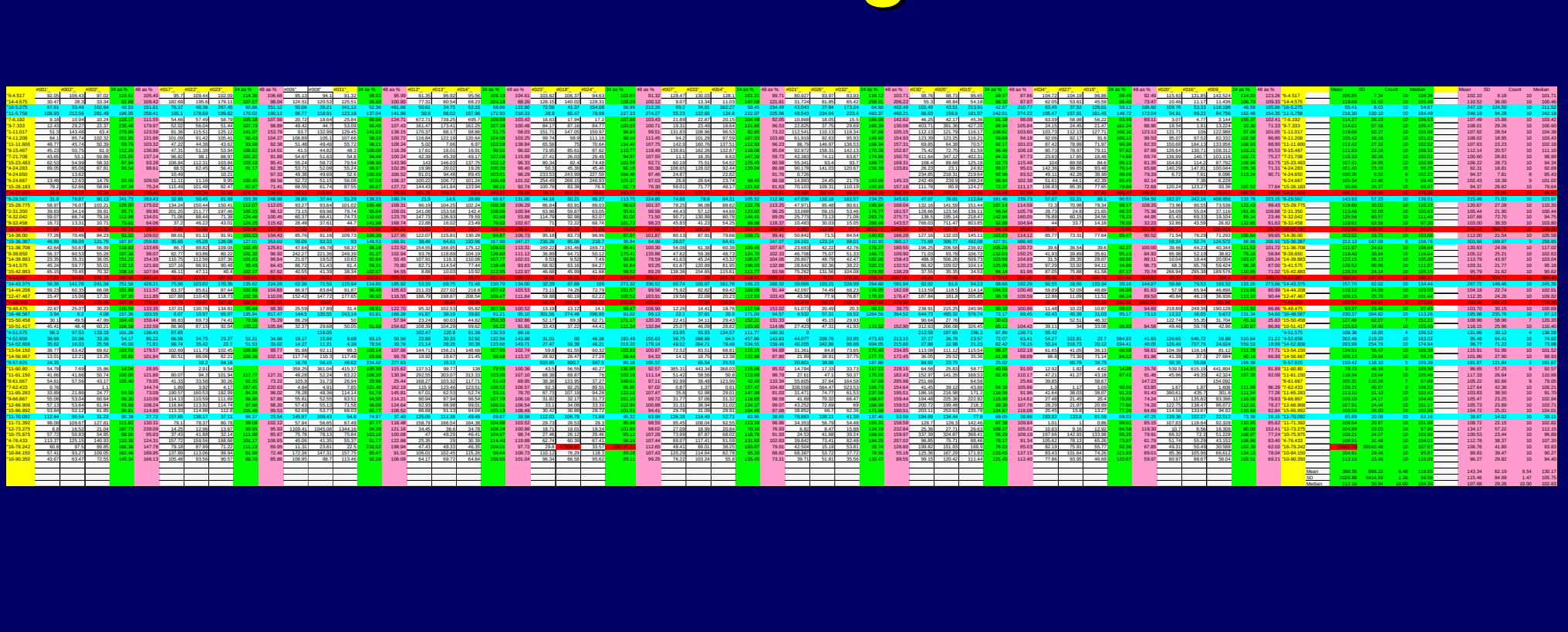
Means and medians (Key points highlighted)

	pre-polishing data	post-polishing data
Range, median CVs for 66 variables	2-54%	2-32%
Range, mean CVs for 66 variables	2-58%	3-38%
Overall mean CV	19%	16%
Overall median CV	**11%**	**12%**
By 13 pairs, overall mean CV	19%	16%
By 13 Pairs, overall median CV	**9%**	**9%**

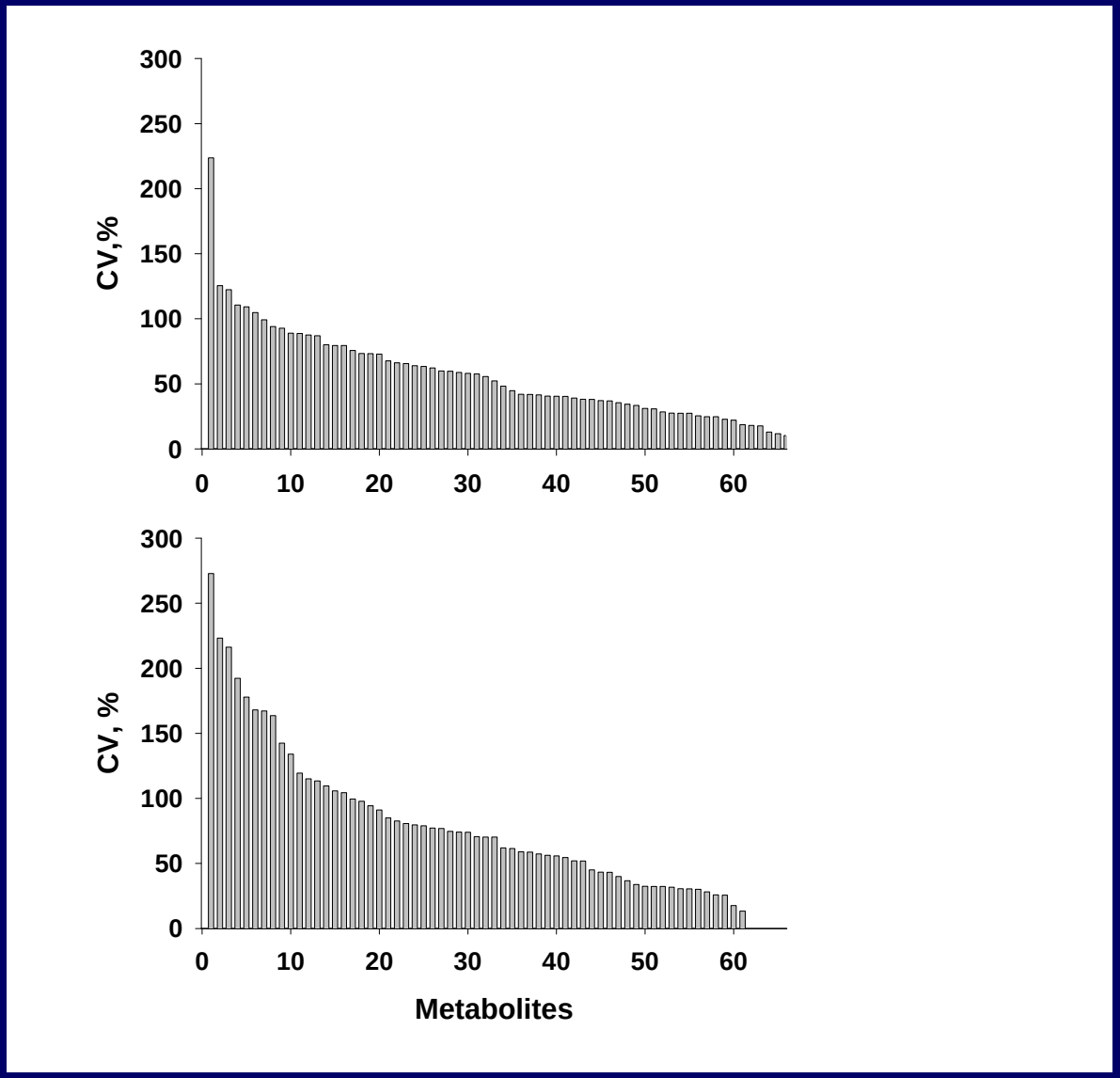
3 variables dropped from further study

Additional study:
Biological CV approximately 82-86%,
Analytical CV 13-16%, (N=69)
Signal:Noise is approximately 5:1

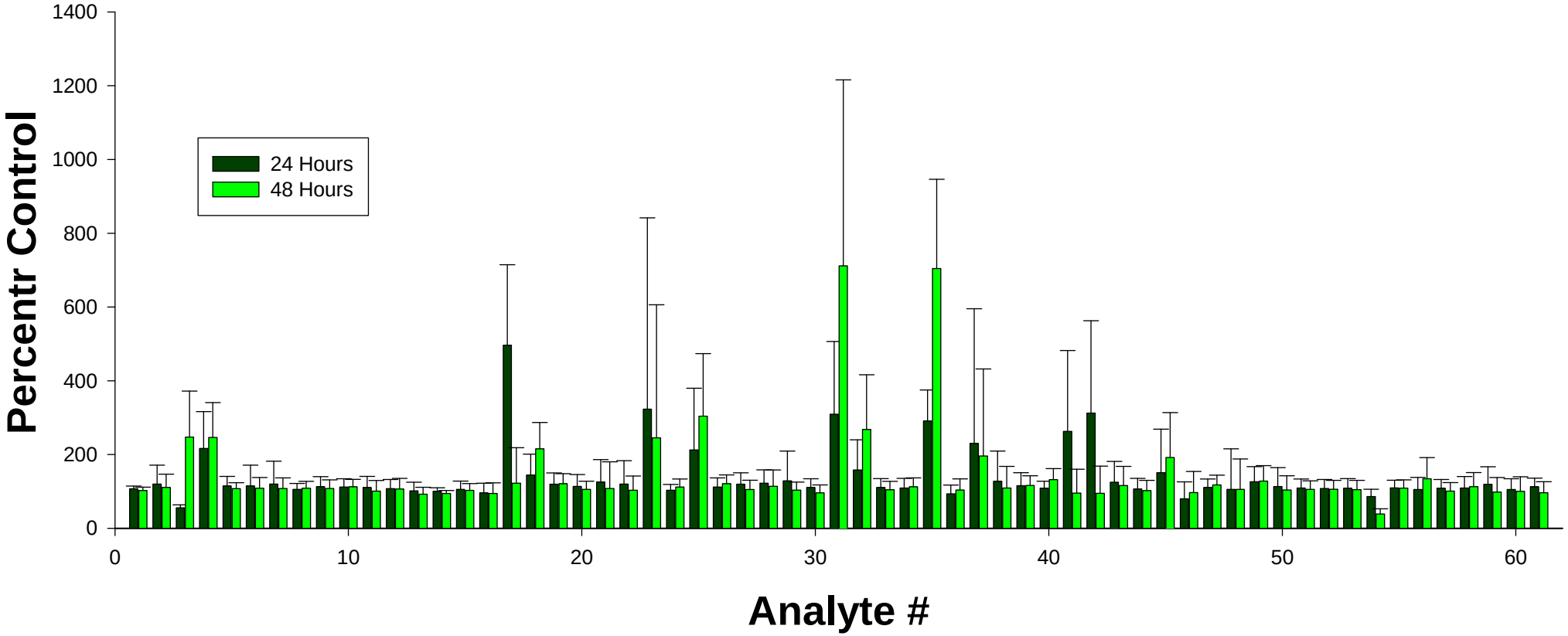
Processing Test



Variability at 24 and 48 hours

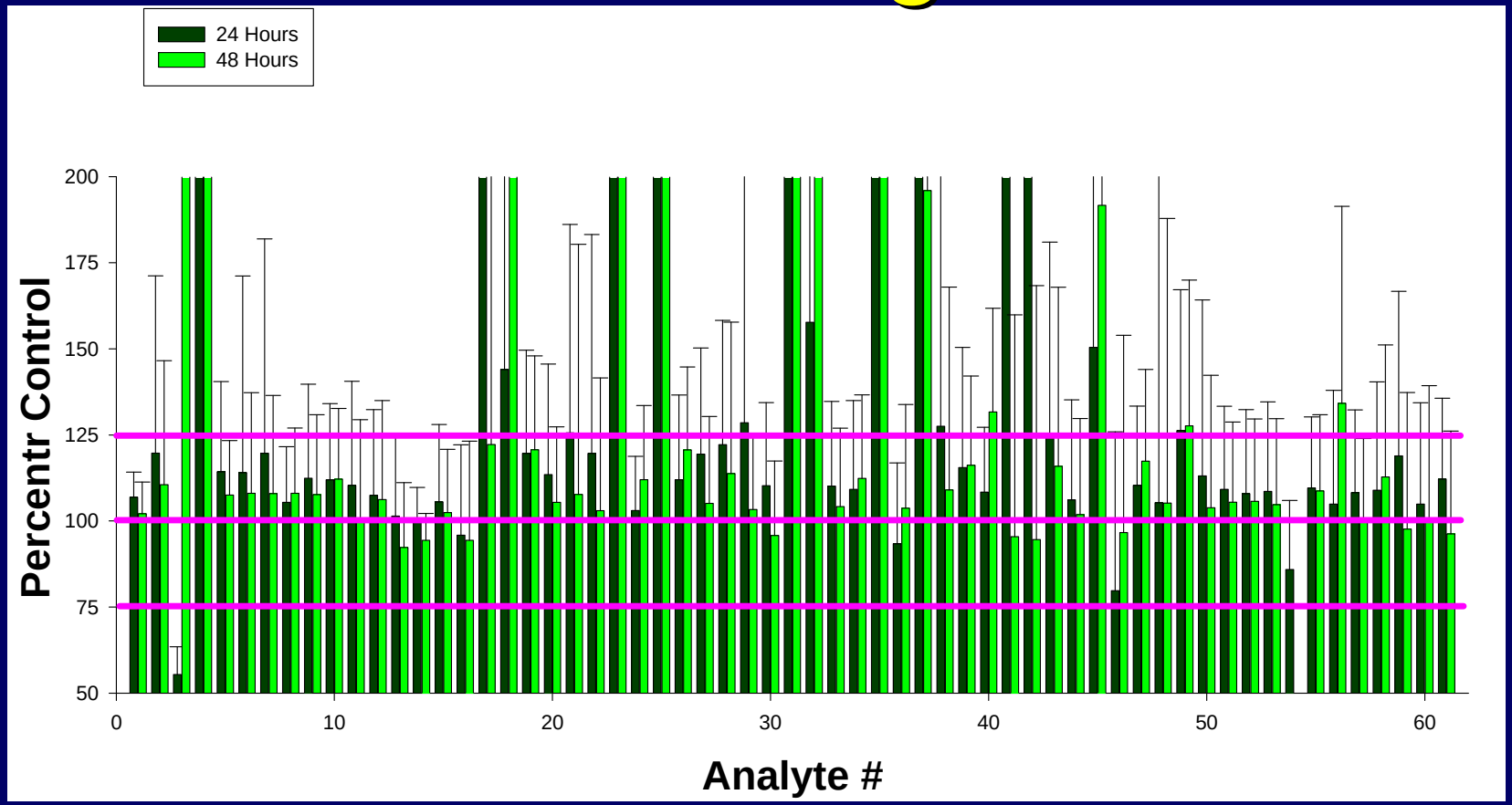


Processing Test



47 of 61 are stable to 48 hours
2 dropped in this study

Processing Test



Processing Test

	match 2	match 3	1 and 2	1 and 3	2 and 3	
1						
2*	3	19	0.606	0.615	0.614	0.611667
3	2	19	0.606	0.614	0.615	0.611667
4	7	33	0.843	0.712	0.773	0.776
5	15	30	0.942	0.929	0.887	0.919333
6	8	31	0.988	0.96	0.974	0.974
7	4	33	0.843	0.773	0.712	0.776
8	6	31	0.988	0.974	0.96	0.974
9	20	16	0.821	0.5	0.722	0.681
10	11		0.965			0.965
11	10		0.965			0.965
12	13	14	0.925	0.921	0.927	0.924333
13	12	14	0.925	0.927	0.921	0.924333
14	12	13	0.921	0.927	0.925	0.924333
15	5	30	0.942	0.887	0.929	0.919333
16	20	9	0.722	0.5	0.821	0.681
17	22	23	0.836	0.635	0.679	0.716667
18						
19	2	3	0.615	0.614	0.606	0.611667
20	9	16	0.821	0.722	0.5	0.681
21	26	32	0.751	0.73	0.797	0.759333
22	17	23	0.836	0.679	0.635	0.716667
23	17	22	0.635	0.679	0.836	0.716667
24	25		0.916			0.916
25	24		0.916			0.916
26	21	32	0.751	0.797	0.73	0.759333
27*	maybe 34?					
28	29		0.892			0.892
29	28		0.892			0.892
30	5	15	0.929	0.887	0.942	0.919333
31	6	8	0.96	0.974	0.988	0.974
32	21	26	0.73	0.797	0.751	0.759333
33	4	7	0.712	0.773	0.843	0.776
34	maybe 27?					
						0.8211

* Other decent matches -- are there more than triplets

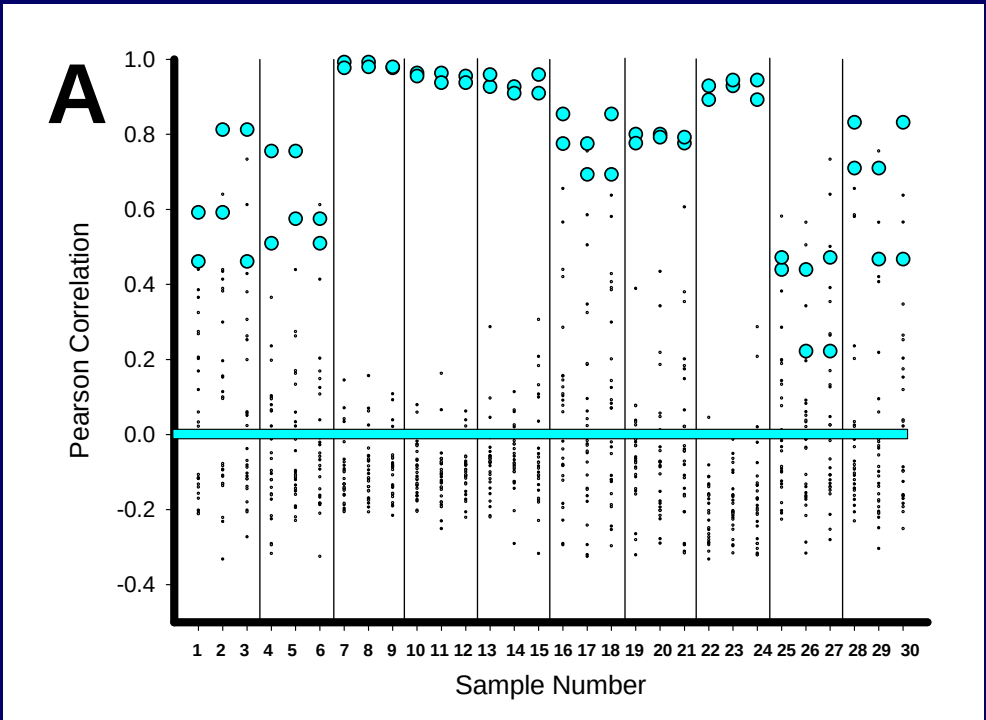
DONOR #	ID processed at 0hr	ID processed at 24hr	ID processed at 48hr
14	001	002	003
21	017	022	023
28	006	008	031
35	012	013	014
42	025	018	024
49	007	033	004
56	021	026	032
63	030	005	015
70	034	027	019
77	020	016	009

NHS QC's	ID
Postmenopausal	010
Postmenopausal	011
Premenopausal	028
Premenopausal	029

All Three Correct (two for Pools)
Two Correct
Not Assigned
Incorrectly assigned

34 samples Total
29 correctly grouped
4 not grouped* two marked as possible
1 incorrectly assigned (r2=>0.6 to group assigned)

Markers have high utility – even in worst case scenarios

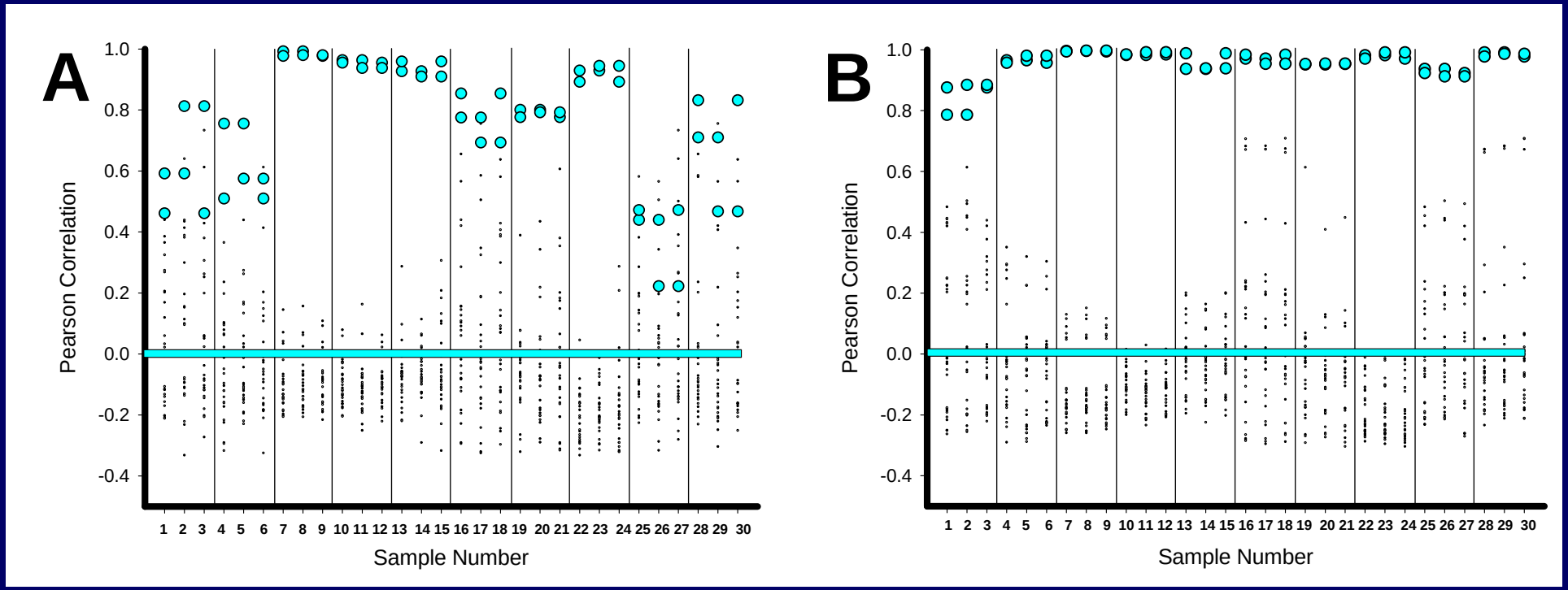


Key point: Under worst case

- **all unstable metabolites included**
- **48 hour delay**
- **No model optimization**

85% Origin ID'd correctly; only 3 % absolute error

Markers have high utility – even in worst case scenarios



Key point: Under worst case,

all unstable metabolites included

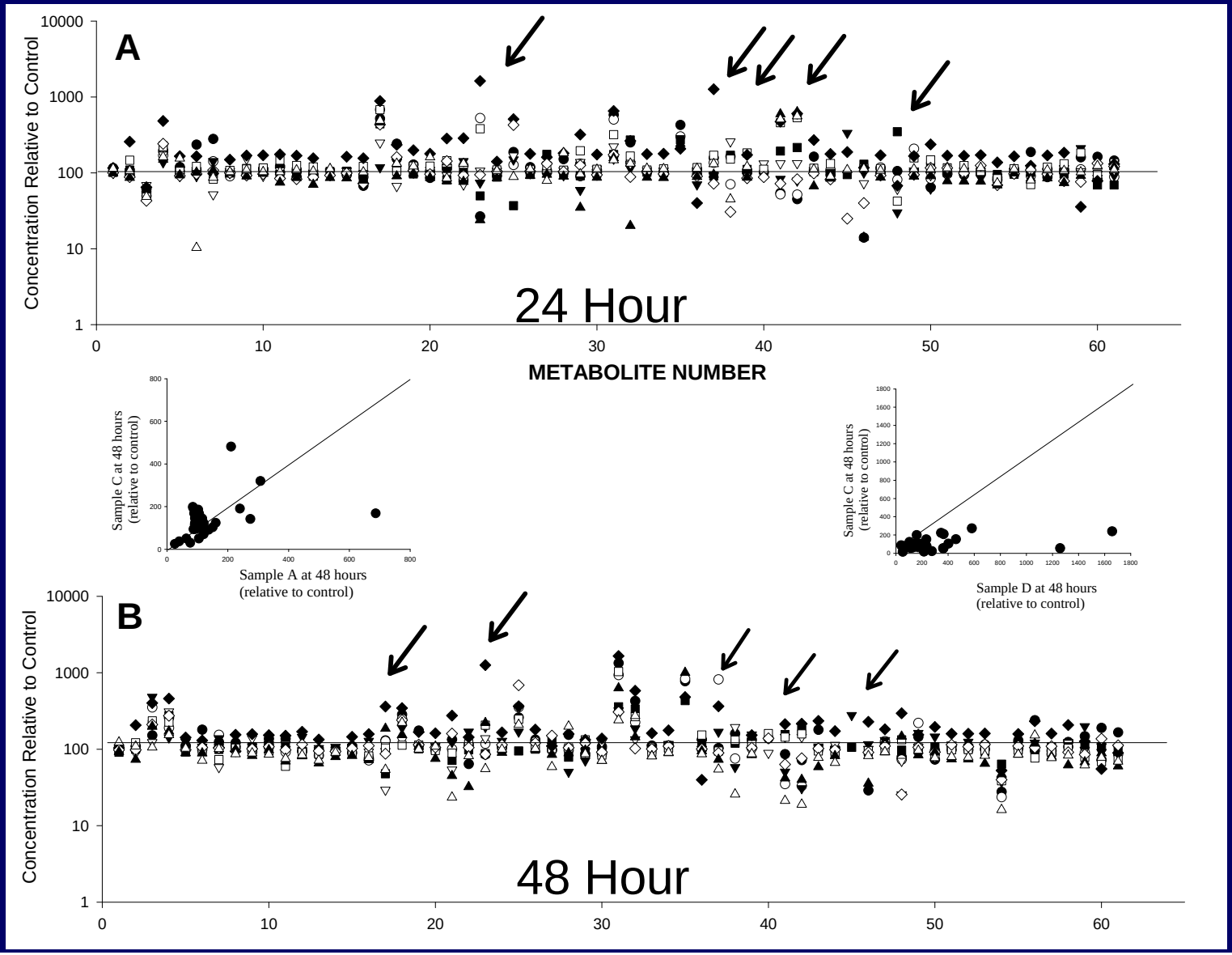
48 hour delay

No model optimization

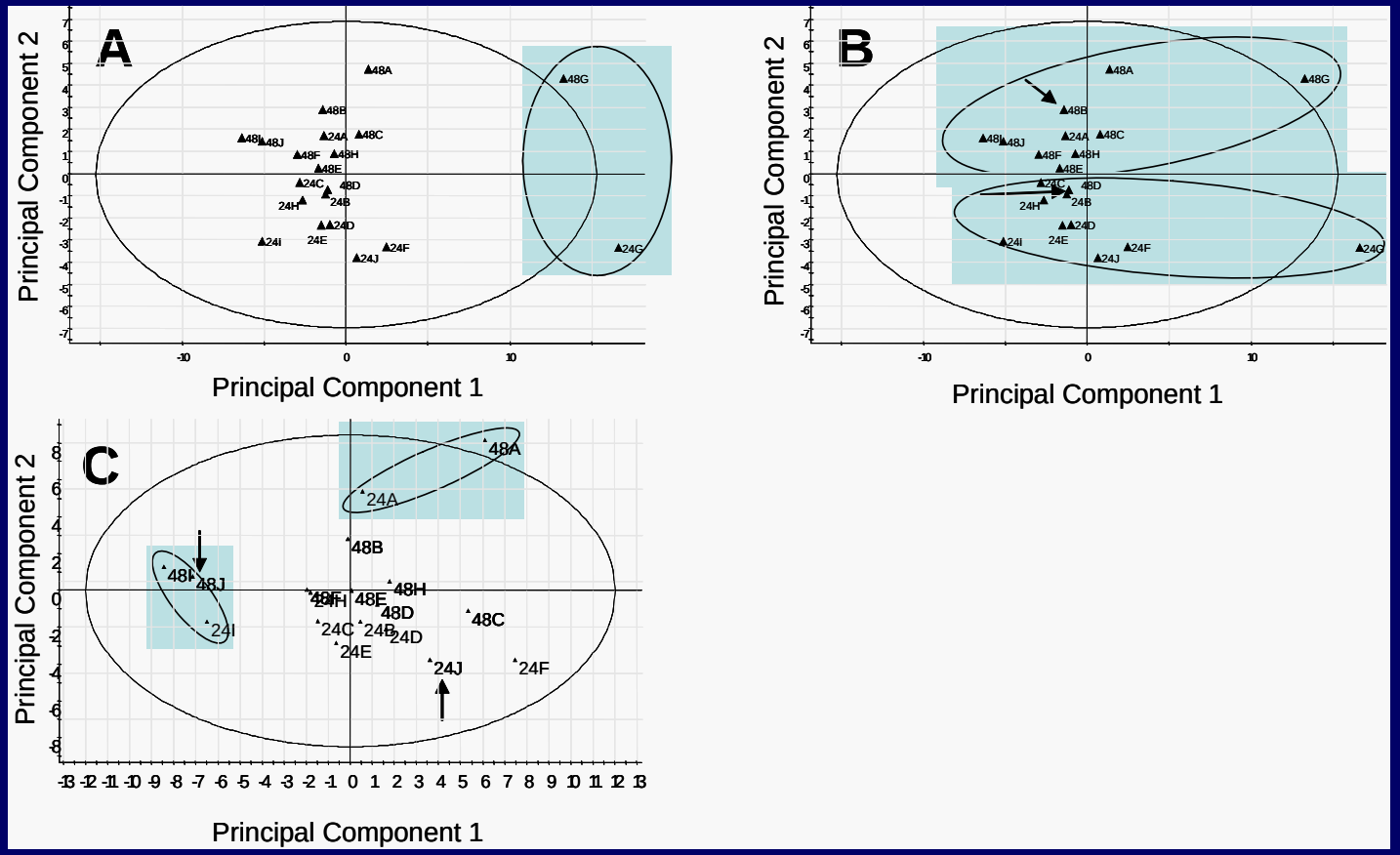
85% Origin ID'd correctly; only 3 % absolute error

Fix this, 100% accuracy

Biological Variability Affects Stability



Biological Variability impacts Analytical Parameters



Human Studies: Inter- vs Intra-Person Variability

- Test only best rat profile
  - 30 variables
    - Knowns (uric acid, tryptophan, tryptophol, 5-hydroxytrphophan, norepinephrine, methionine, 4-hydroxyphenyllactic acid)
  - 69 Triplets scored
- Single Profile Tested
  - (ie, true test; not re-fitted or optimized)
- ICC = 0.76

Human Studies:
Profile = ...

- Recent Food (ie, Fast)
- BMI
- Food intake
- Physiology

Human Studies:
Profile = ...

- Recent Food (ie, Fast)
- BMI
- Food intake
- Physiology

What is coming (funded)

Equivalent proteomics models (NIA)

Studies in CALERIE (NIA)

(0-3-6-12-18-24 mos pre/post diet)

Some validation against double-labeled water

750 case/control pairs for breast cancer (NCI) –

NHS, samples taken 2-10 years before onset

1000 case/control pairs for Type II diabetes (NIA) –

NHS, samples taken 2-10 years before onset

What is coming (funded)

- NIH Genes/Environment Initiative (NIEHS)
  - Exposure Biology Program
  - Biological Response indicators
    - Non-fat based models for primary fat in diet
    - Models for glycemic index
    - models for 24 combinations of fat/GI
    - Cross of all models into breast cancer
    - Cross of all models into Type II diabetes
    - Ties of all models to liver mitochondrial function

Summary

- Created and validated a working model of the CR serotype in both male and female rats
  - Can identify group of origin with high accuracy
  - Wide capture of metabolome
  - Reduced dimensionality of the problem
  - Identify critical markers
- Profiles distinguish diet
  - Fasted vs AL and CR
  - Across lifespan
  - across duration and extent of diet (r^2 =0.88)

Summary

- Markers are passing analytical tests in human plasma
  - Can identify split duplicates (100% accuracy)
  - Good analytical parameters (5:1 signal/noise)
  - Handles worst case shipping conditions
  - Good inter/intra-individual variability

The metabolomic markers and profiles identified appear *analytically* and *biologically* suitable for studies in defined human populations such as national clinical trials and epidemiological cohorts

Acknowledgements

- Brigham and Women's and Burke/Cornell
  - Yevgeniya Shurubor, Ugo Paolucci, Honglian Shi, Ruiwen Zhou, Vasant Marur, Diane Sheldon, Sophie Guo, Neil Russell

- Outside Collaborators
  - Walter Willett, Sue Hankinson, Frank Hu, Paul Vouros
  - Wayne Matson, Karen Vigneau-Callahan, Paul Milbury
  - Tom Vogl

- Funding
  - NIH (NIA; NCI; NIEHS)
  - Winifred Masterson Burke Relief Foundation
  - Brigham and Women's Hospital