The Metabolic Serotype of Dietary Restriction: Markers for Disease Risk in Humans?

Bruce S. Kristal, Ph.D.

Department of Neuroscience
Weill Medical College of Cornell University

Dementia Research Service
Burke Medical Research Institute

“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

American Cancer Society Study

900,000 adults

Calle et al., NEJM, 348:1625-1638, 2003

Overall Relative Cancer Risk (BMI >40 vs 18.5 to 24.9):
M: 1.52 (1.13 – 2.05): F: 1.62 (1.40 – 1.87)

ncreased risk of colorectal, pancreatic, liver, esophagus, kidney, multiple myeloma, non-Hodgkin’s lymphoma, gallbladder, prostate, breast, cervical, ovarian, uterine

“Current patterns of overweight and obesity in the United States could account for 14% of all deaths from cancer in men and 20% of those in women”

Guidelines for Healthy Weight

Nurses' Health Study: Willett et al., NEJM, 341:427-434, 1999

BMI of 26 vs 21

Coronary Heart Disease:	2x increase
Hypertension:	2-3x increase
Cholelithiasis:	2-3x increase
Type II Diabetes:	8x increase

Weight Change of 15 kg

Coronary Heart Disease:	2x increase
Hypertension:	2-3x increase
Cholelithiasis:	2-3x increase
Type II Diabetes:	6x increase

Dietary Restriction (DR)

DR 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

A Brief History of DR -- Phenomenology

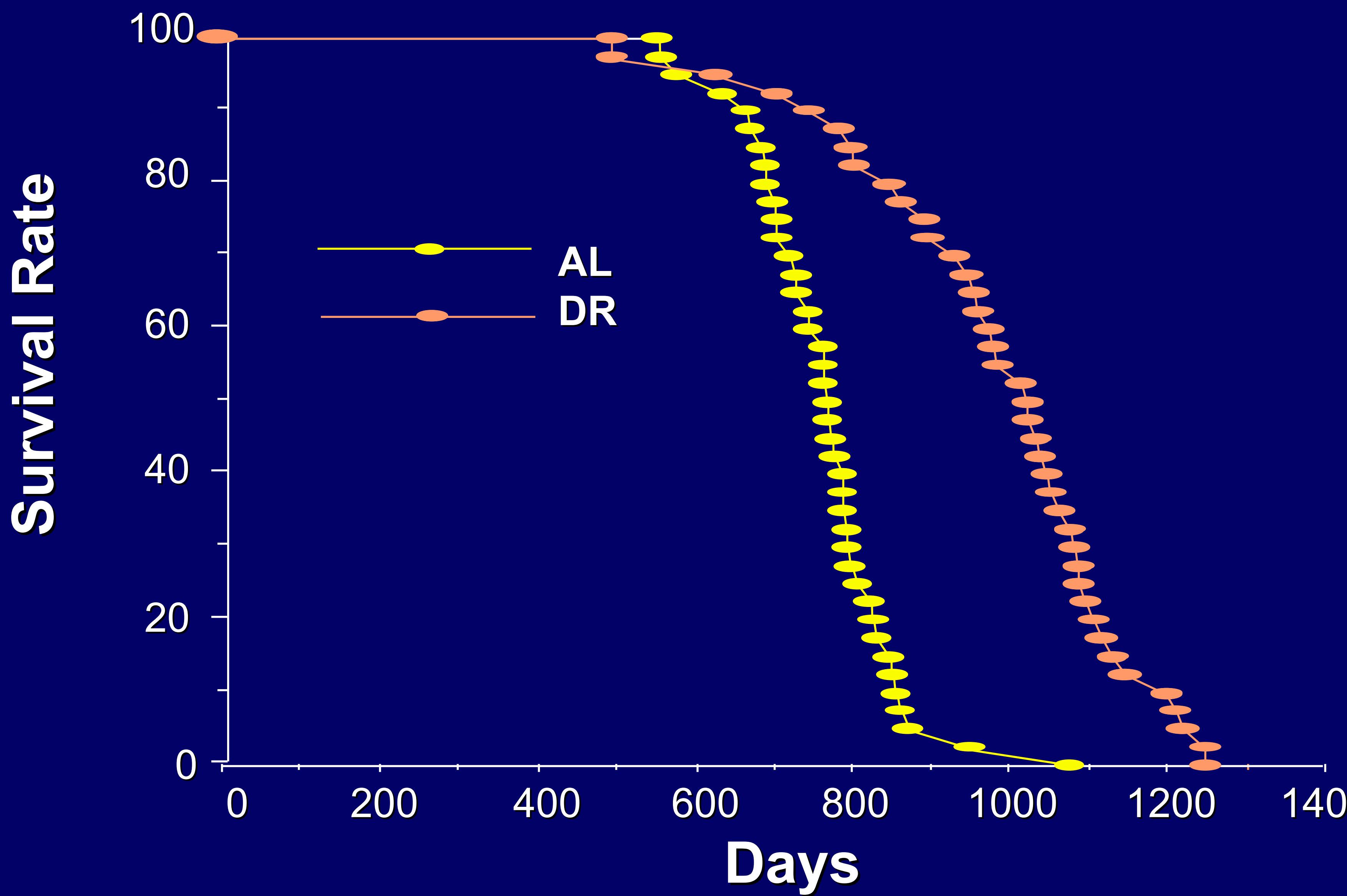
- Cellulose in diet extends longevity (McKay, 1934)
- Stair-step restriction extends longevity (McKay, 1935, 1939
- Refed animals retain benefits DR (McKay, 1941)
- DR modulates pathology (Saxton, 1941)
- DR modulates leukemia (Saxton, 1944)
- ntermittent fasting is protective (Carlson, 1946)
- DR slows the mortality rate doubling time (Ross, 1959)
- DR is robust across diets (Ross, 1965)
- 20%/40% differ in longevity, physiology (Nolen, 1972)

A Brief History of DR -- Phenomenology

- DR works in barrier colonies (Weisbroth, 1972)
- DR works both early and late in lifespan
- Calories are a key determinant (Ross, 1972)
- DR works better than protein restriction or source
- DR works better than fat/mineral restriction (Iwasaki, 1988)
- DR works in germ free animals (Snyder, 1989)
- Methionine restriction is very effective (Orentreich, 1993)
- DR works in Dwarf Mice (Bartke, 2001)

Dietary 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



High Throughput and/or Data Density Studies

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

Multivariate Approaches

- High density data collection
- Bioinformatics/Data mining

(General) Hypothesis

↓

Data Driven Analysis

(Discovery-Based Analysis)

Hypothesis:

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

Predictions

- DR alters the sera “metabolome”
- There exists a “DR Serotype”

Goals

1) Insights into the mechanism of DR
2) Recognize DR 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)

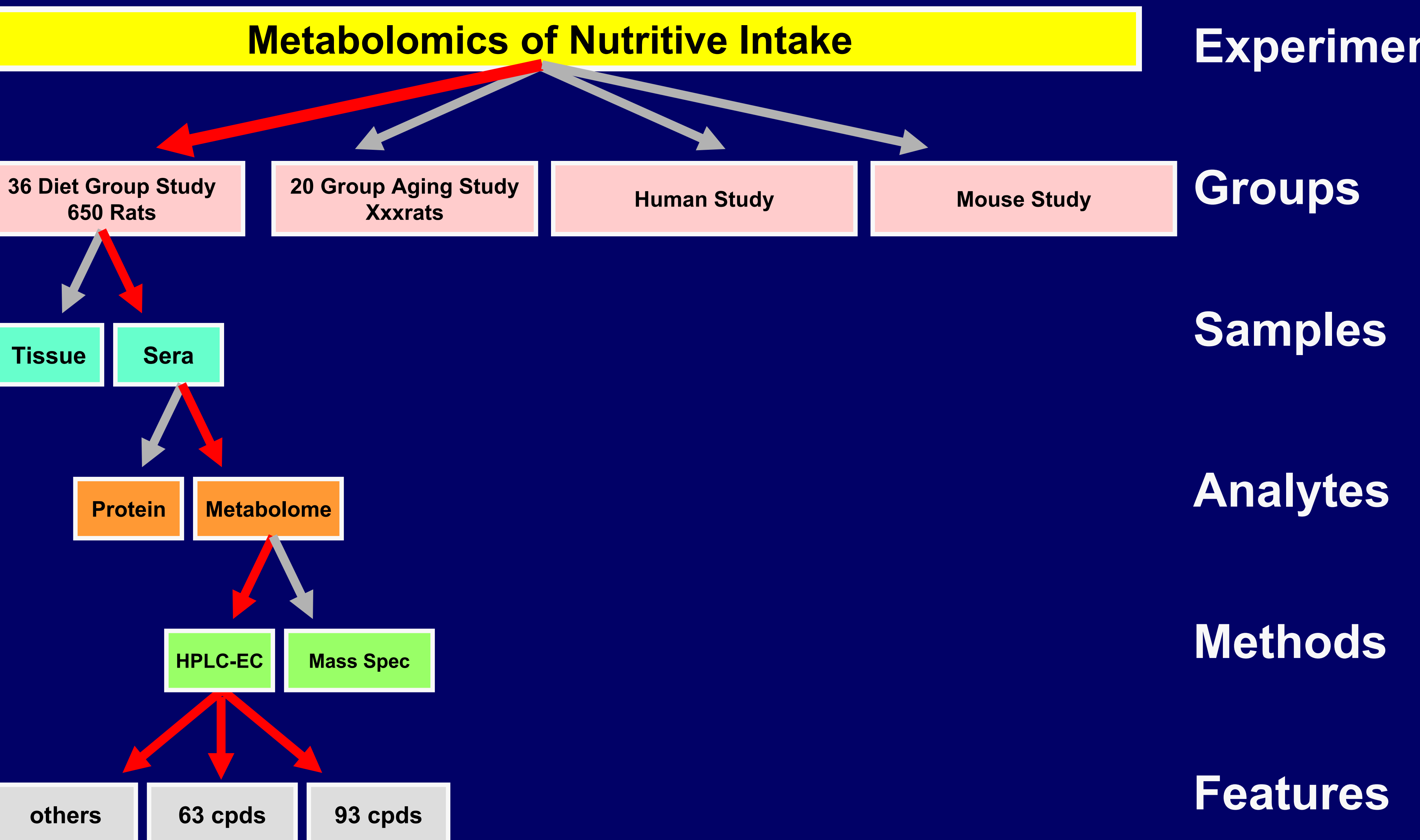
4) Identify predictive markers of disease

Experimental Design

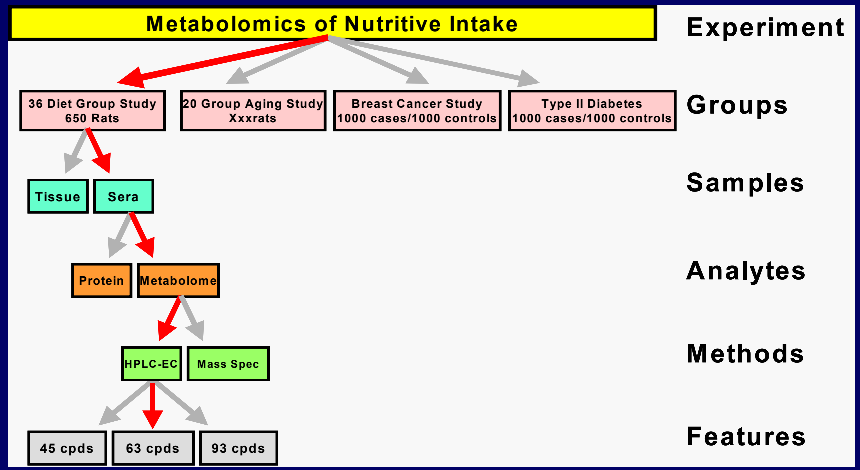
Model: F344 x BN F_1 Rat

Overall Design:

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



Metabolomics of Nutritive Intake
Experimen
Groups
36 Diet Group Study
650 Rats
20 Group Aging Study
Xxxrats
Human Study
Mouse Study
Samples
Tissue
Sera
Analytes
Protein
Metabolome
Methods
HPLC-EC
Mass Spec
Features
others
63 cpds
93 cpds



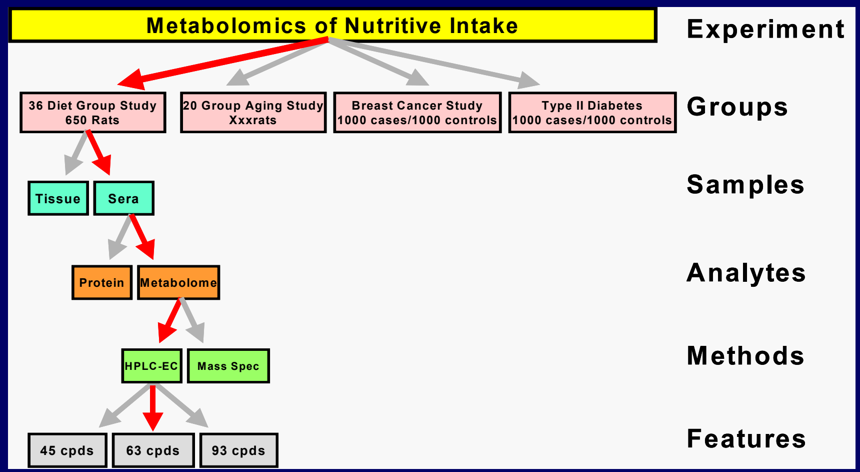
Data Validation, Data Normalization, Missing Data Decisions, Inclusion/Exclusion Criteria

Subgroups, Class-specific models

Outlier removal ←→ scaling ←→ transformations

lustering SOMs PCA kNN SIMCA PLS PLS-DA Random Forest Neural Nets GAs GP

Overfit tests, Internal validation, optimization, External validation, optimization, 2º validation



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

↓

Subgroups, Class-specific models

↓

Outlier removal ←→ scaling ←→ transformations

↓

lustering ~~SOMs~~ PCA kNN SIMCA PLS PLS-DA ~~Random Forest~~ ~~Neural Nets~~ ~~GAs~~ ~~G~~

↓

Overfit tests, Internal validation, optimization, External validation, optimization, 2º validation

Analytical Approach

HPLC separations with coulometric array detection

Multilayer statistical and data analysis

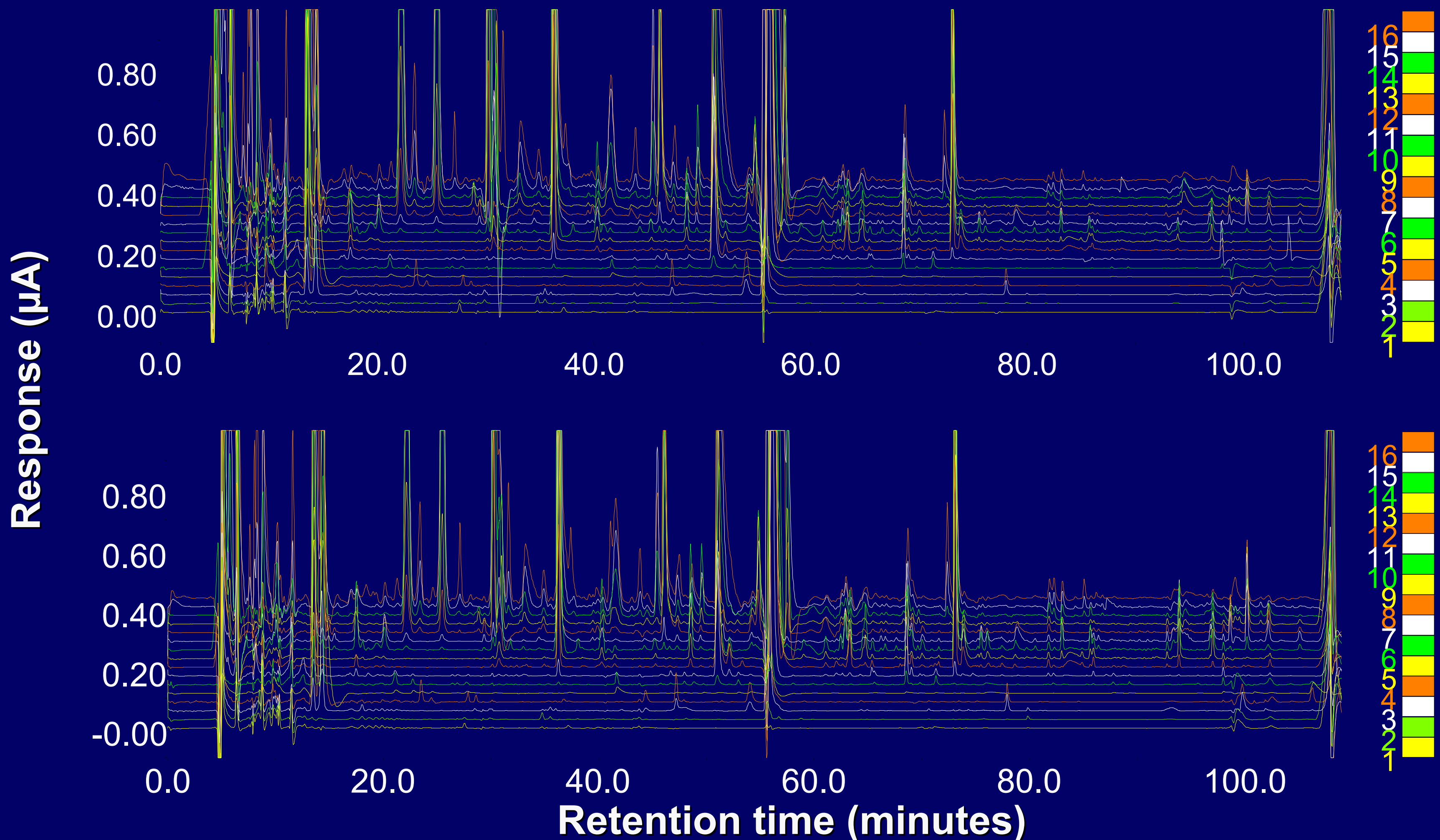
DATA QUALITY
NOT QUANTITY

Analytical Stability

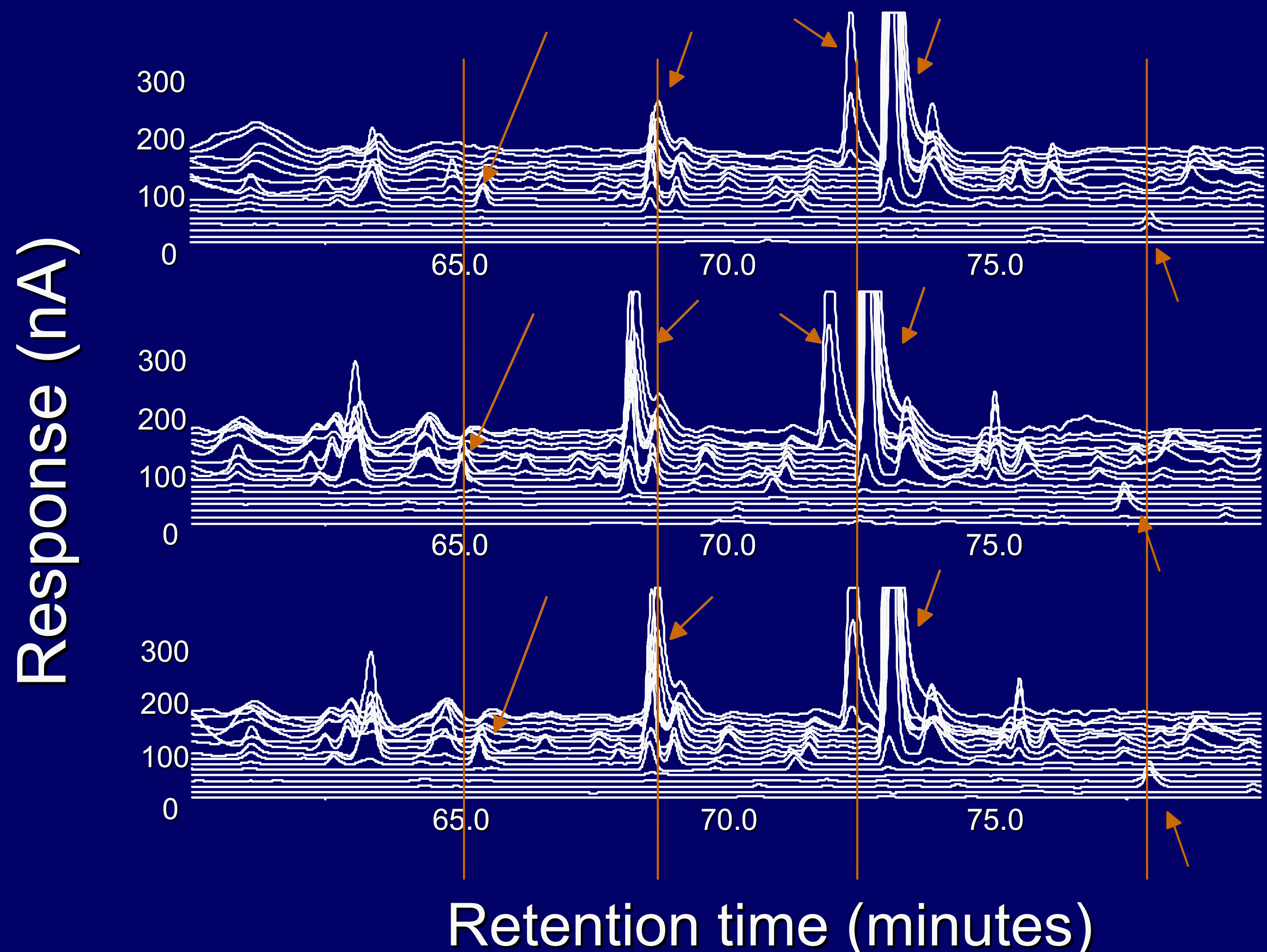
Biologic Variability

Analytical Stability/Biological Variability

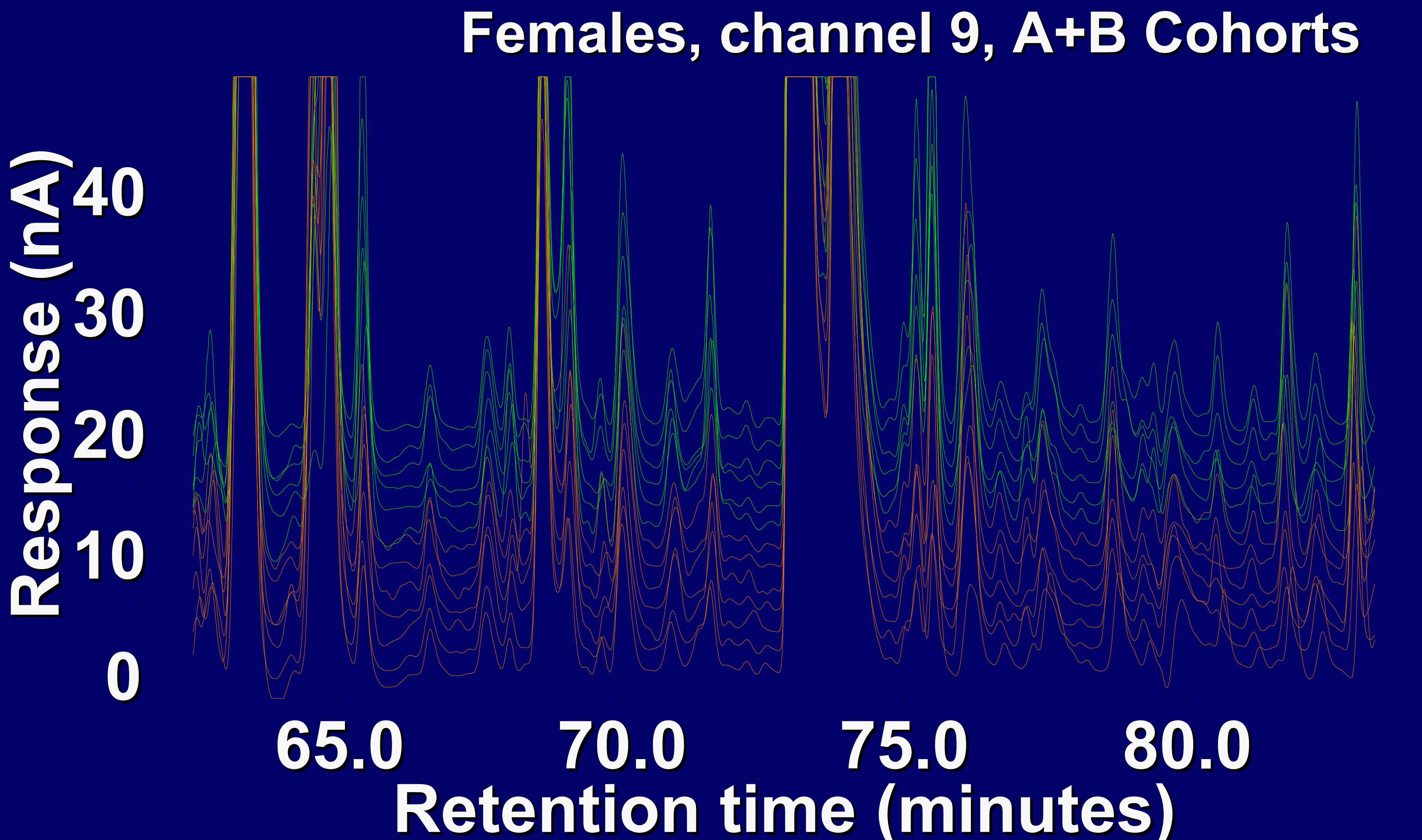
Females--Cohort A + B



Chromatogram Normalization

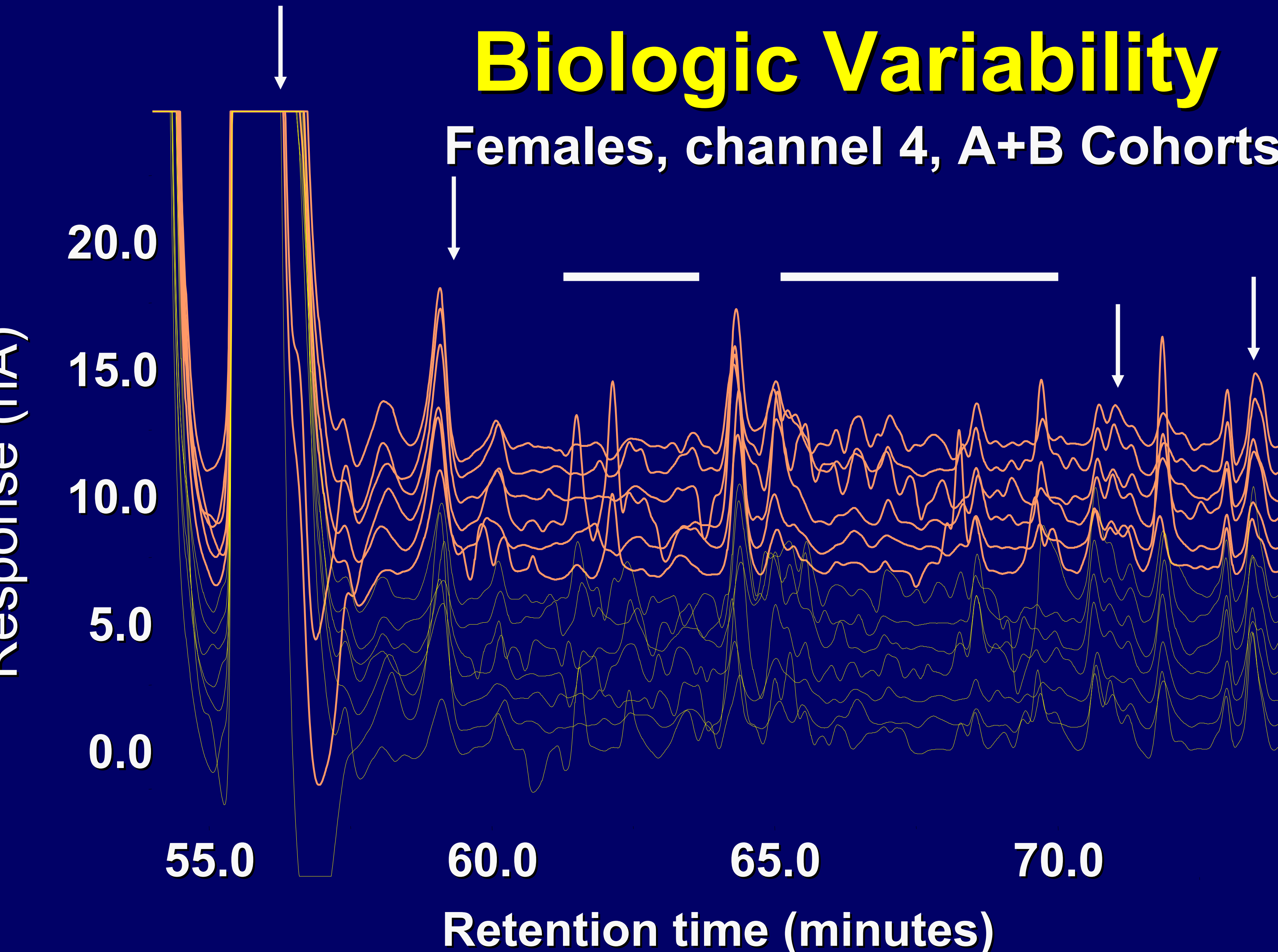


Biologic Variability



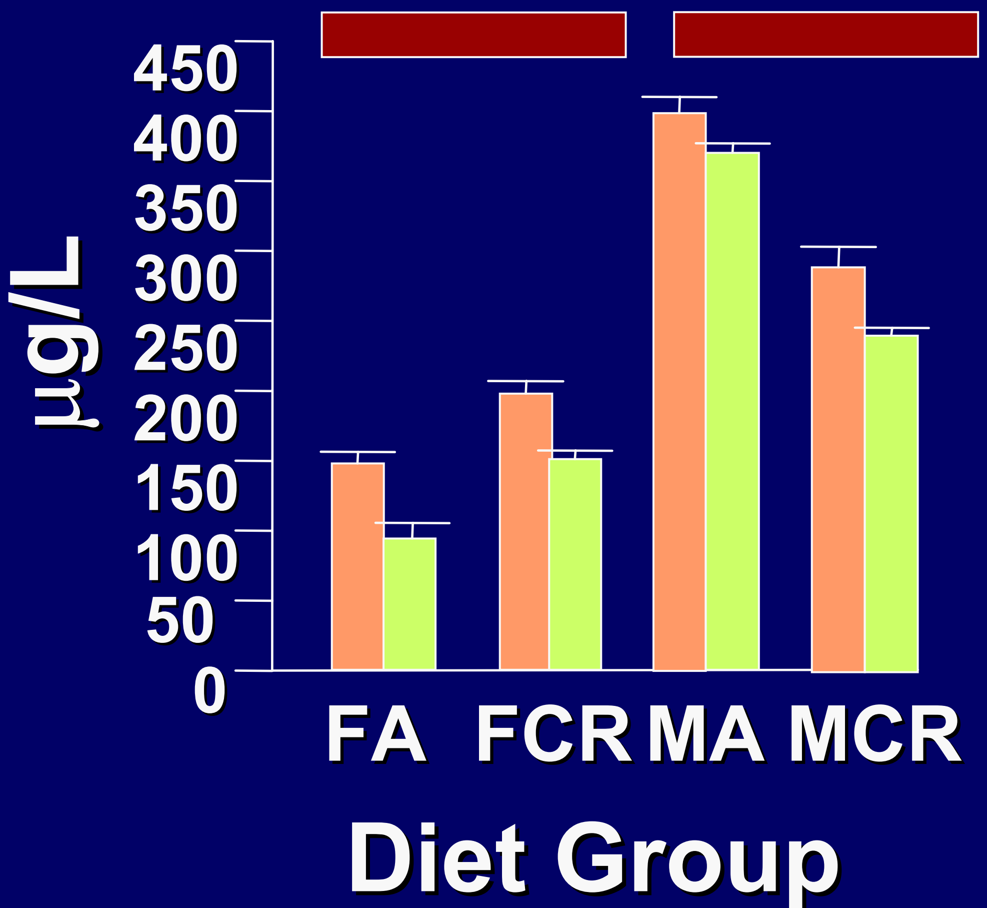
Biologic Variability

Females, channel 4, A+B Cohorts

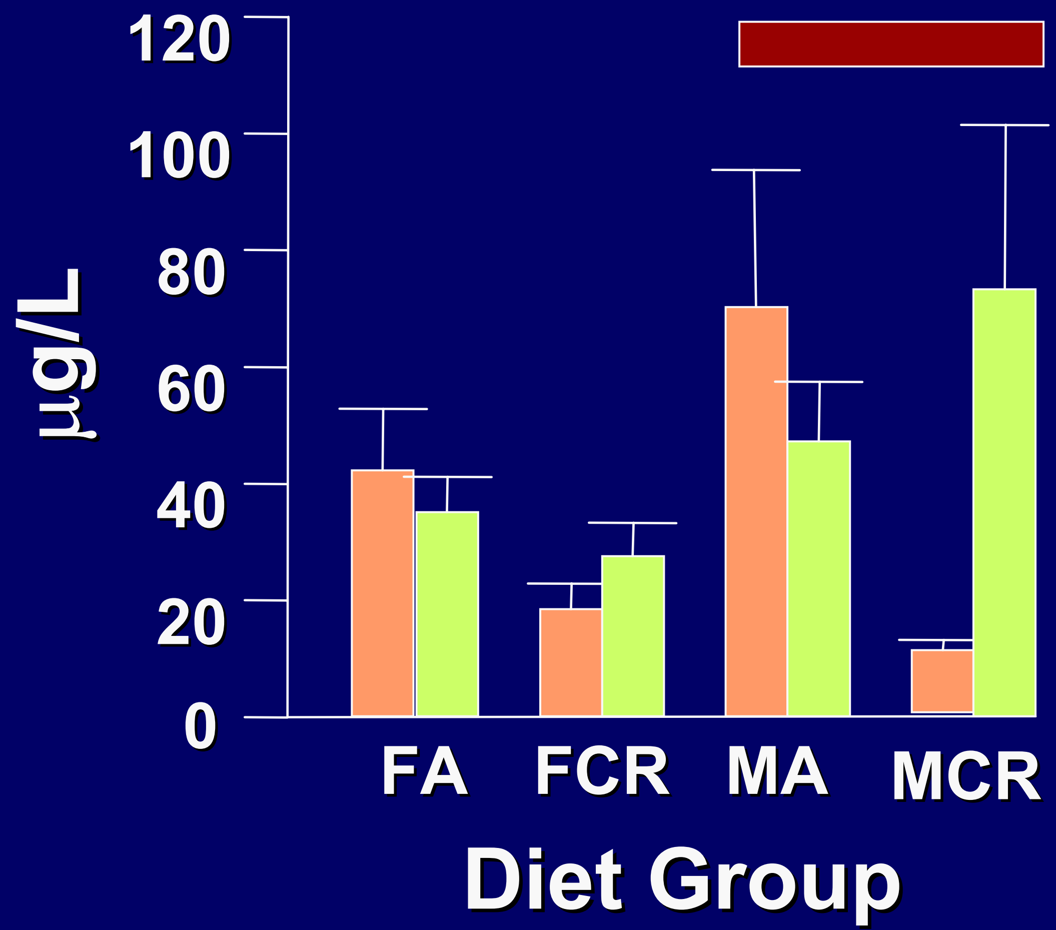


Gender and Cohort Specificity

Retinol



Retinol-Palmitate



Analytical Issues

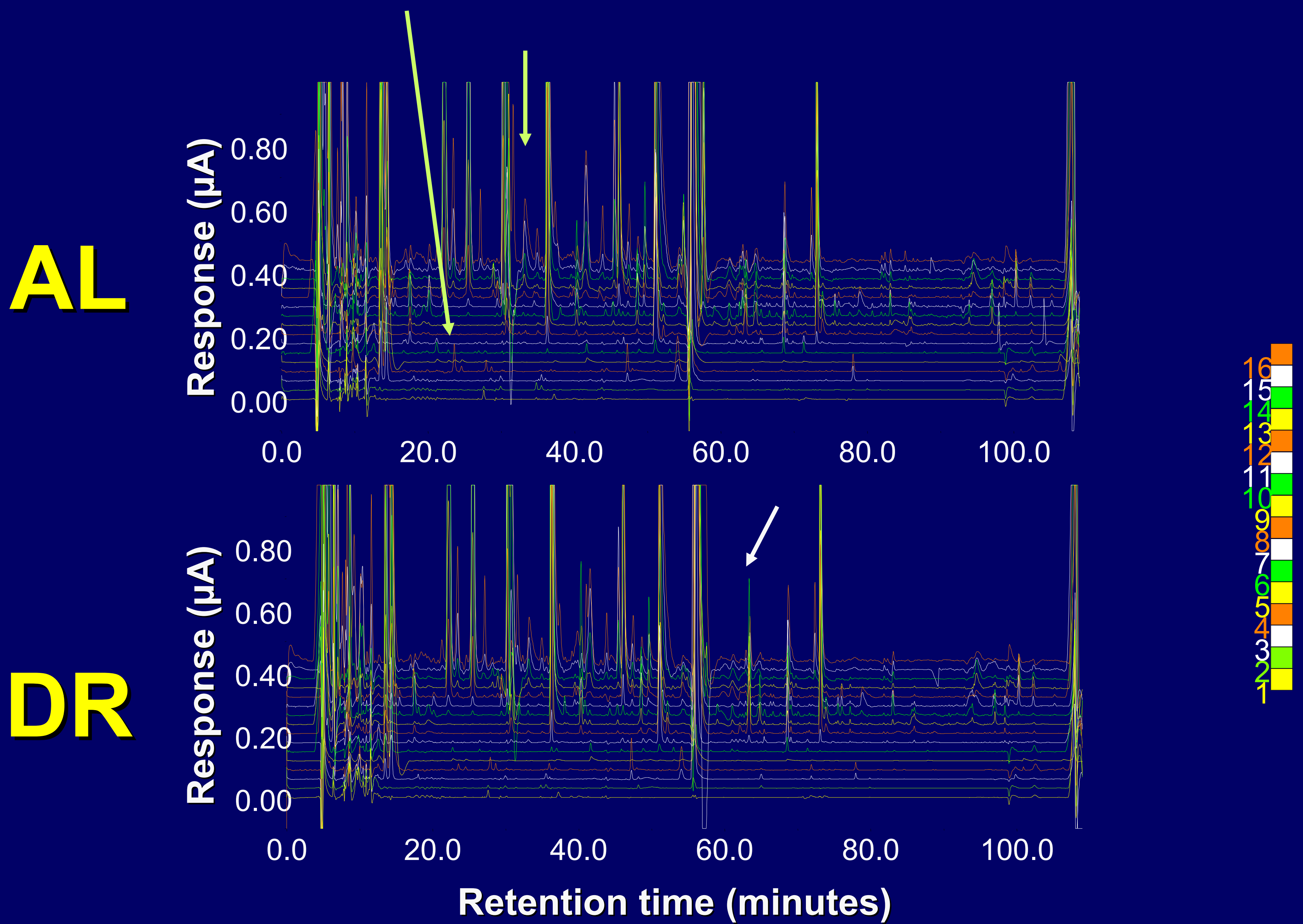
Stretching algorithm helps keep chromatography
"stable" over time

290/325 peaks are analytically valid
(2 different automated methods)

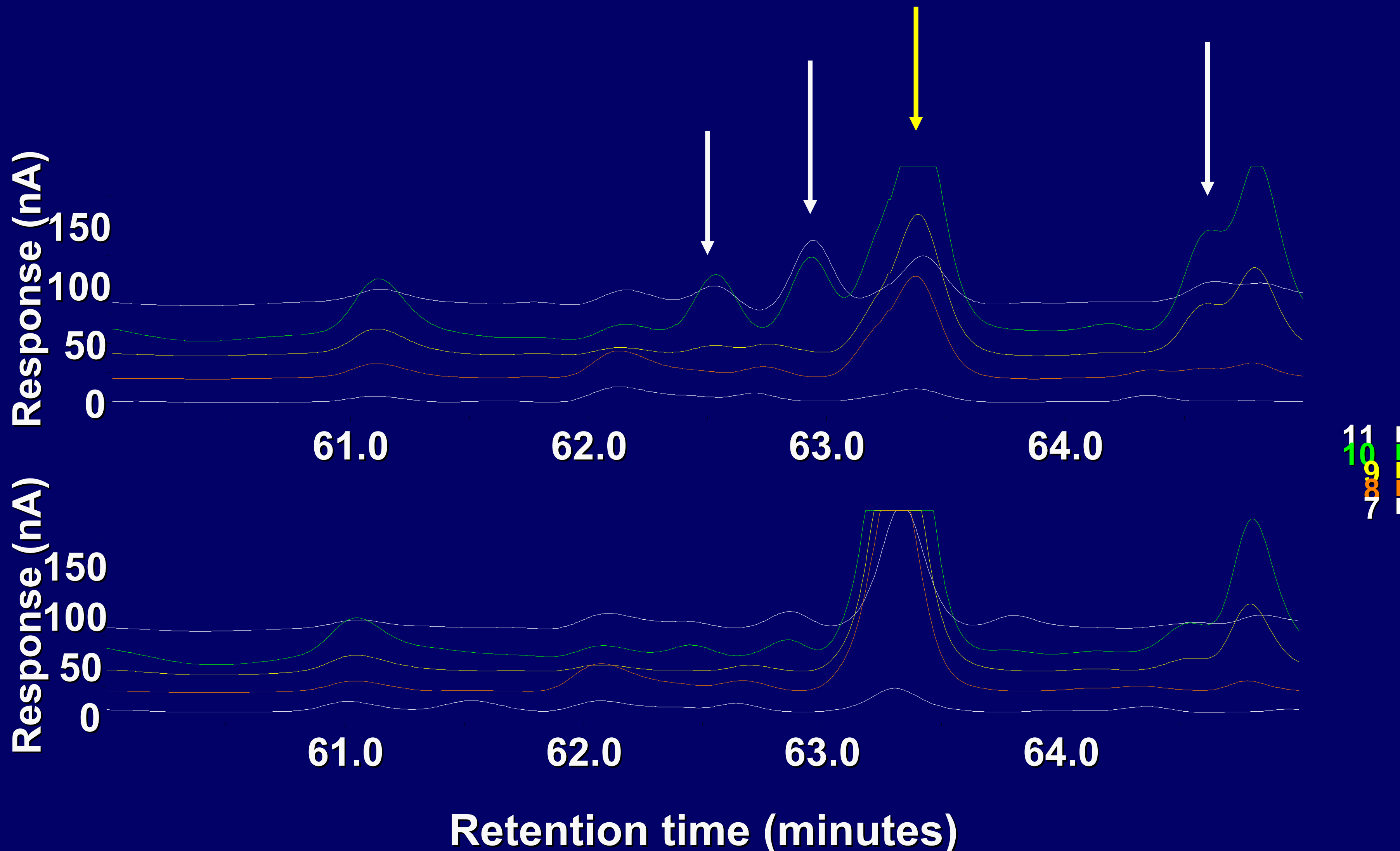
~240/290 peaks are biologically valid (F/M)

>400 peaks biologically/analytically valid overall

Do AL and DR Sera Differ?

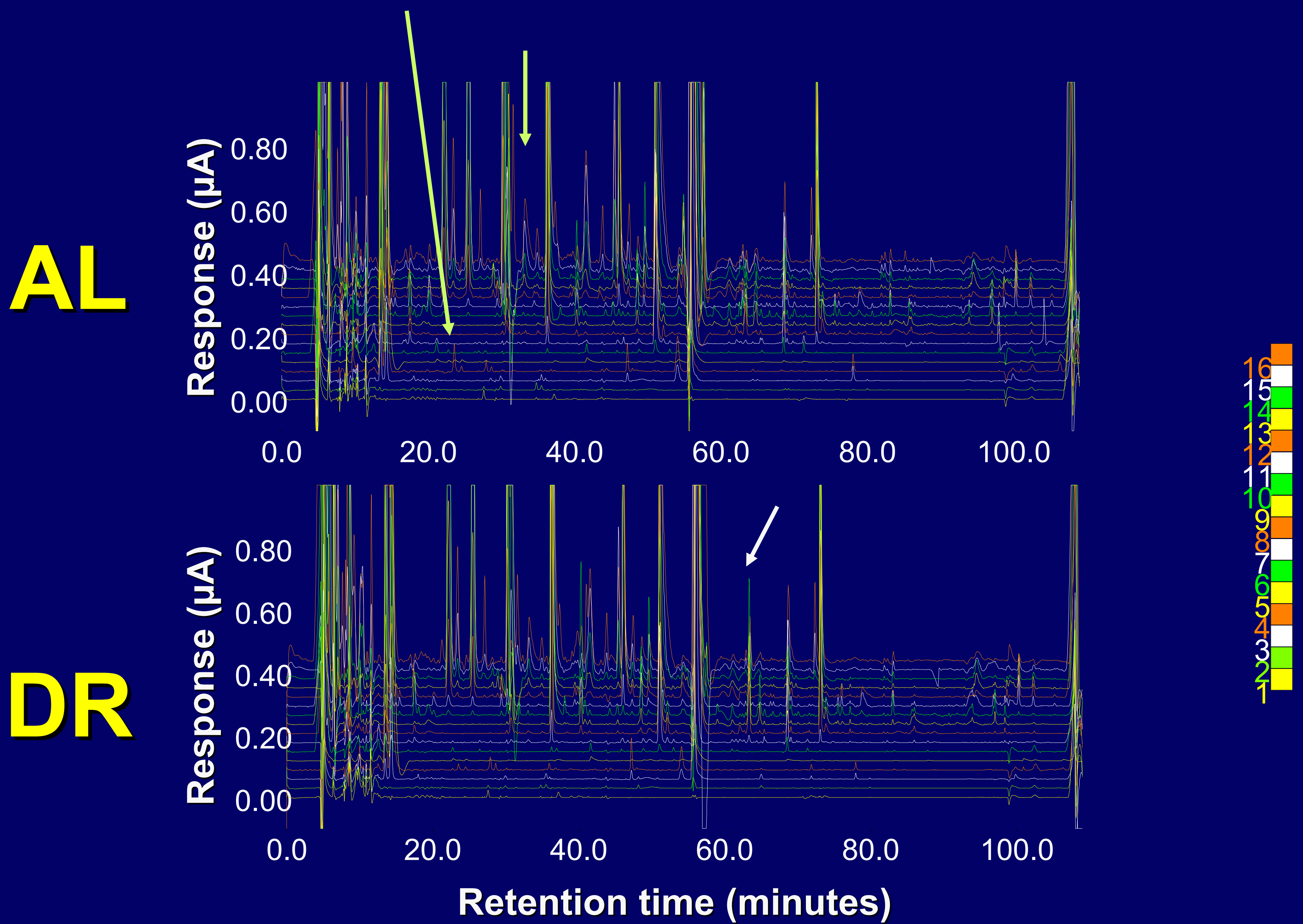

AL
DR
Response (µA)
0.80
0.60
0.40
0.20
0.00
0.0
20.0
40.0
60.0
80.0
100.0
Retention time (minutes)
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

AL vs DR



Response (nA)
150
100
50
0
61.0
62.0
63.0
64.0
11
10
9
8
7
Response (nA)
150
100
50
0
61.0
62.0
63.0
64.0
Retention time (minutes)

Do AL and DR Sera Differ?


AL
DR
Response (µA)
0.80
0.60
0.40
0.20
0.00
0.0
20.0
40.0
60.0
80.0
100.0
Retention time (minutes)
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

Data Analysis

- t-Test
- Multi-/Mega-variate analysis
  - HCA
  - PCA
- Pattern recognition

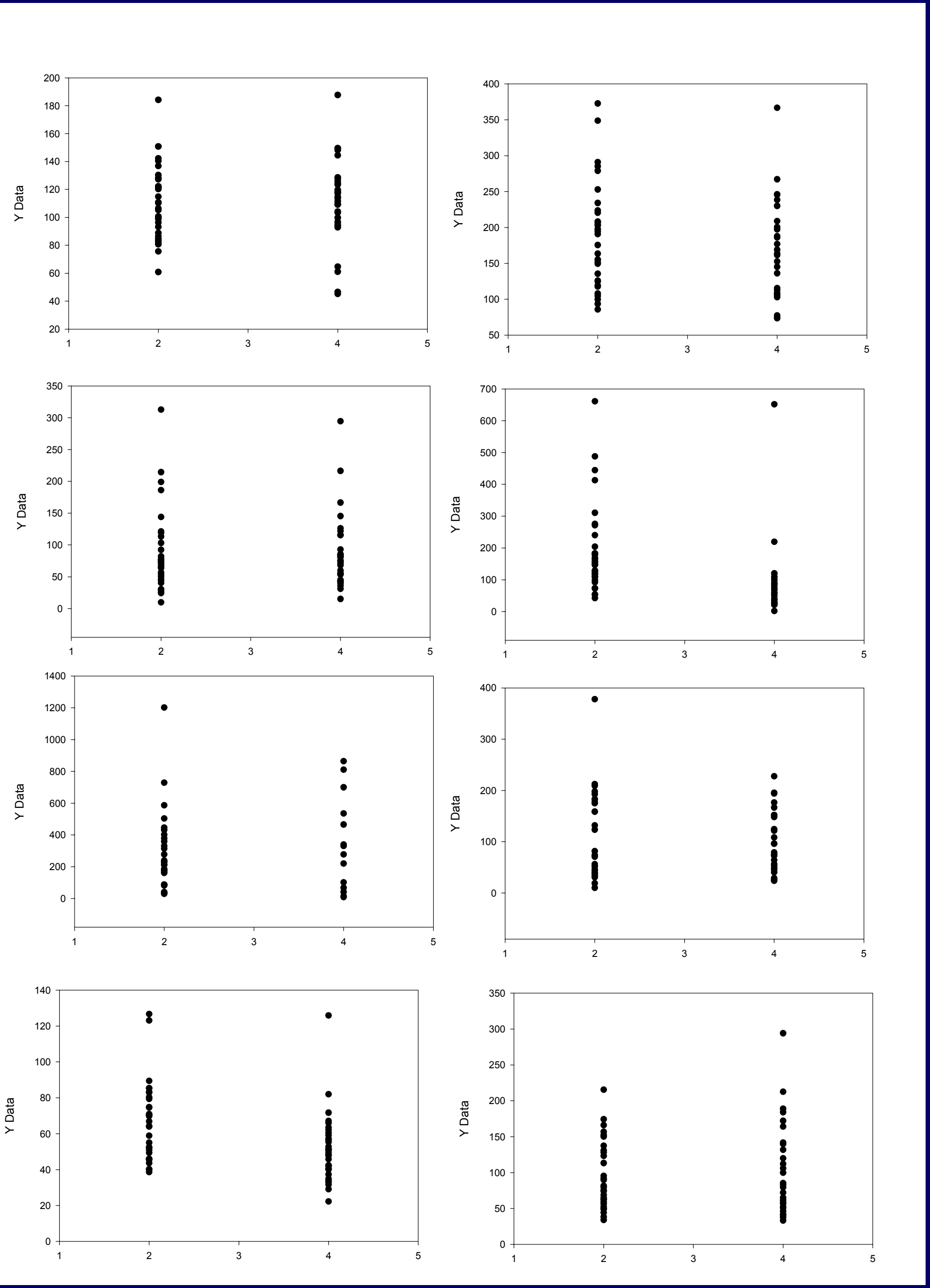
Primary Data Analysis

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

Primary Data Analysis
“Cut” by t-test at $p<0.2$

- Weak Criteria
  - Type II errors now mostly analytical errors
- ~100/~300 peaks statistically differ
- ~60 may be Type I errors (Bonferroni)

Single Markers Don’t help



Profiles out-perform single markers

we hope…

How do we build robust profiles?

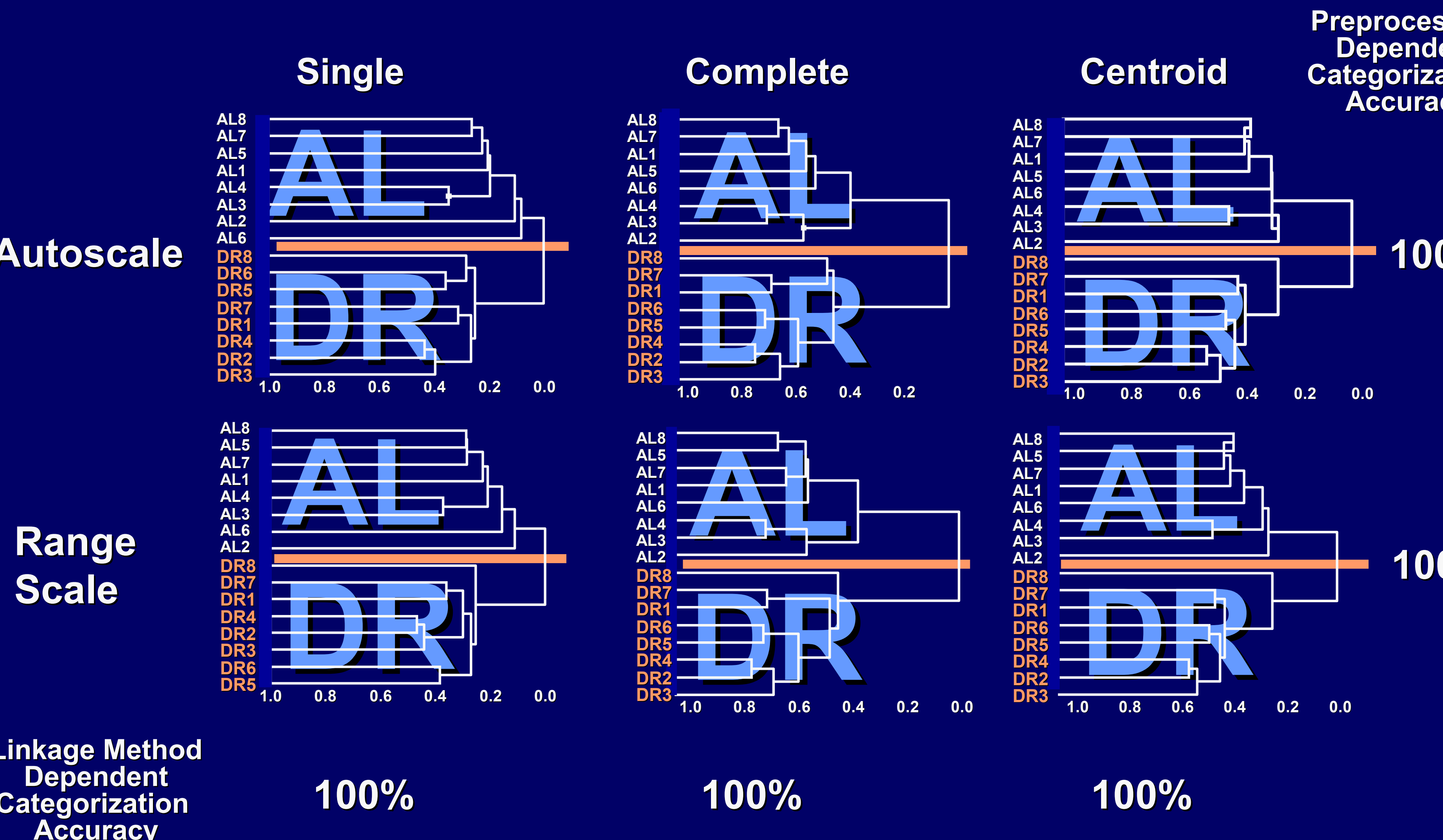
Does Serotype Encode Sufficient Information to Identify Diet Group?

Proof of Principle

Classification Analysis

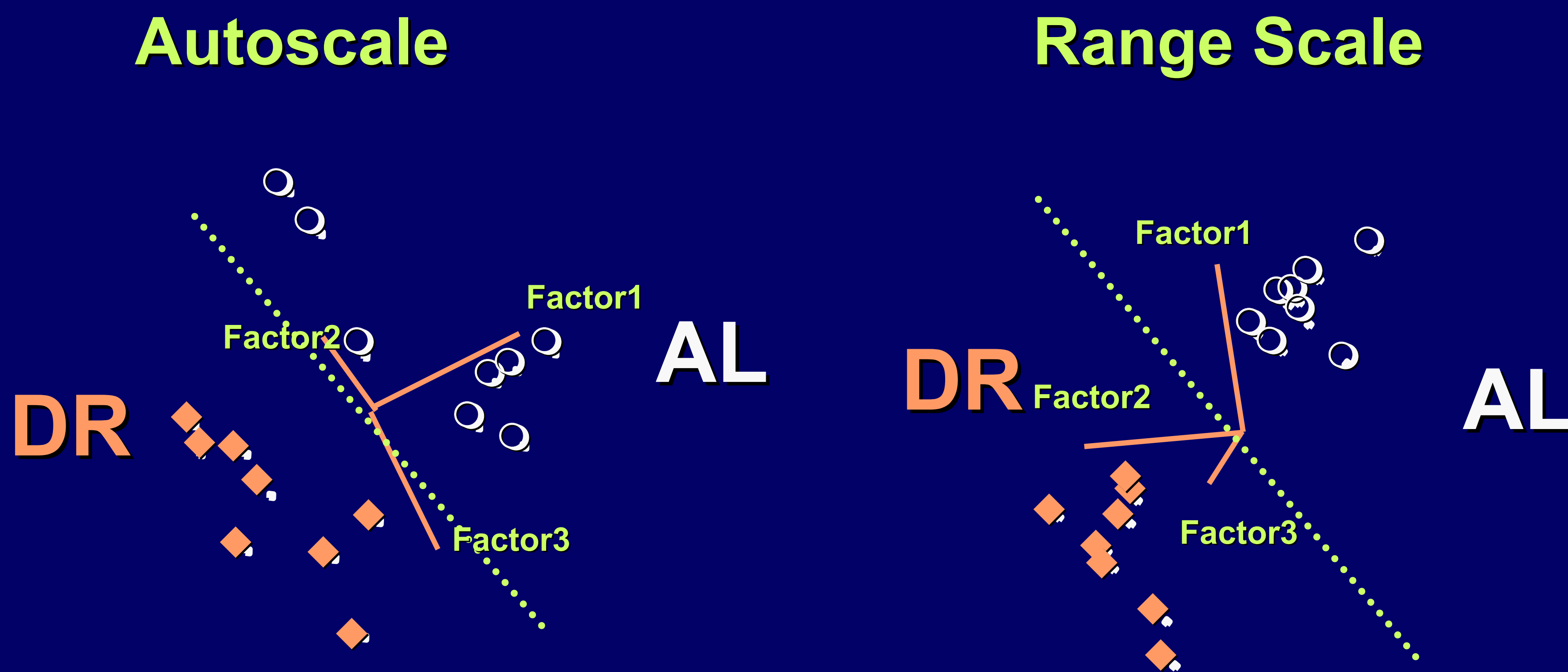
- Hierarchical Cluster Analysis (HCA)
  - Identifies natural groups in data
- Principal Component Analysis (PCA)
  - Finds linear combinations of original variables that account for maximal variation
- Pattern Recognition
  - Develop algorithms that do classification without intervention

HCA Distinguishes Female AL and DR Rats



63 variables (confirmed), Non-independent samples

PCA Distinguishes AL and DR Female Rats

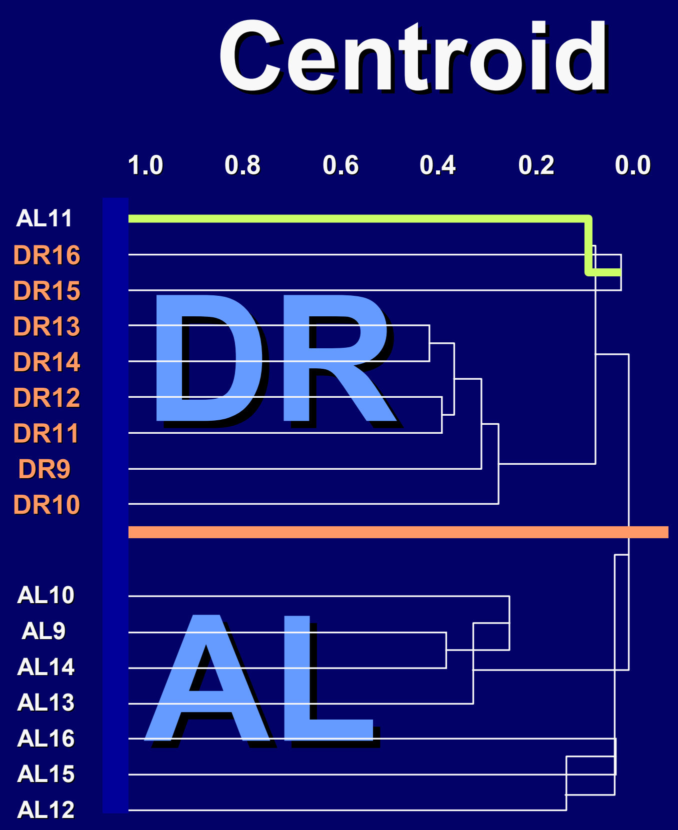


63 variables (confirmed), Non-independent samples

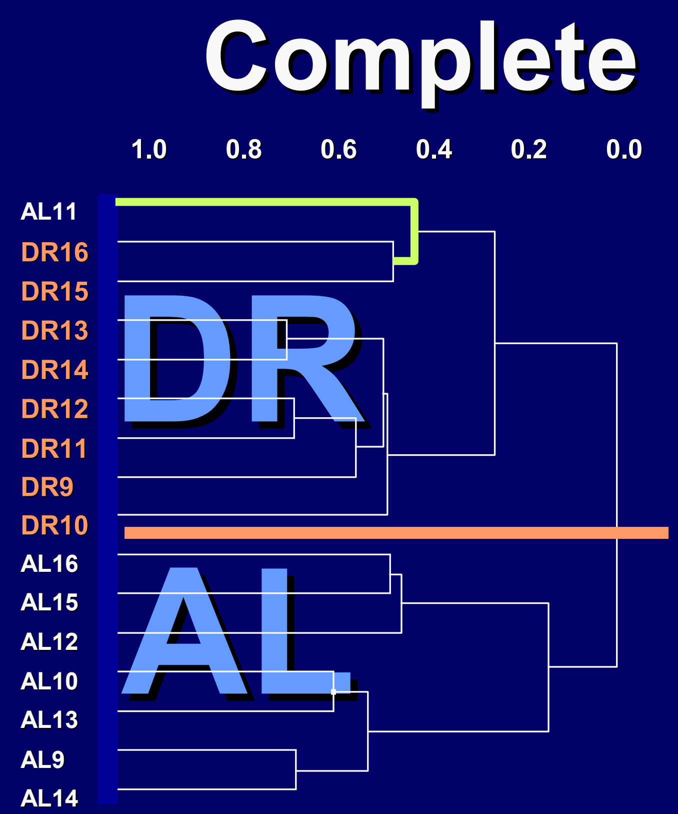
Does Serotype Encode Sufficient Information to Identify Diet Group?

Primary Validation Studies

HCA Distinguishes Female AL and DR Rats



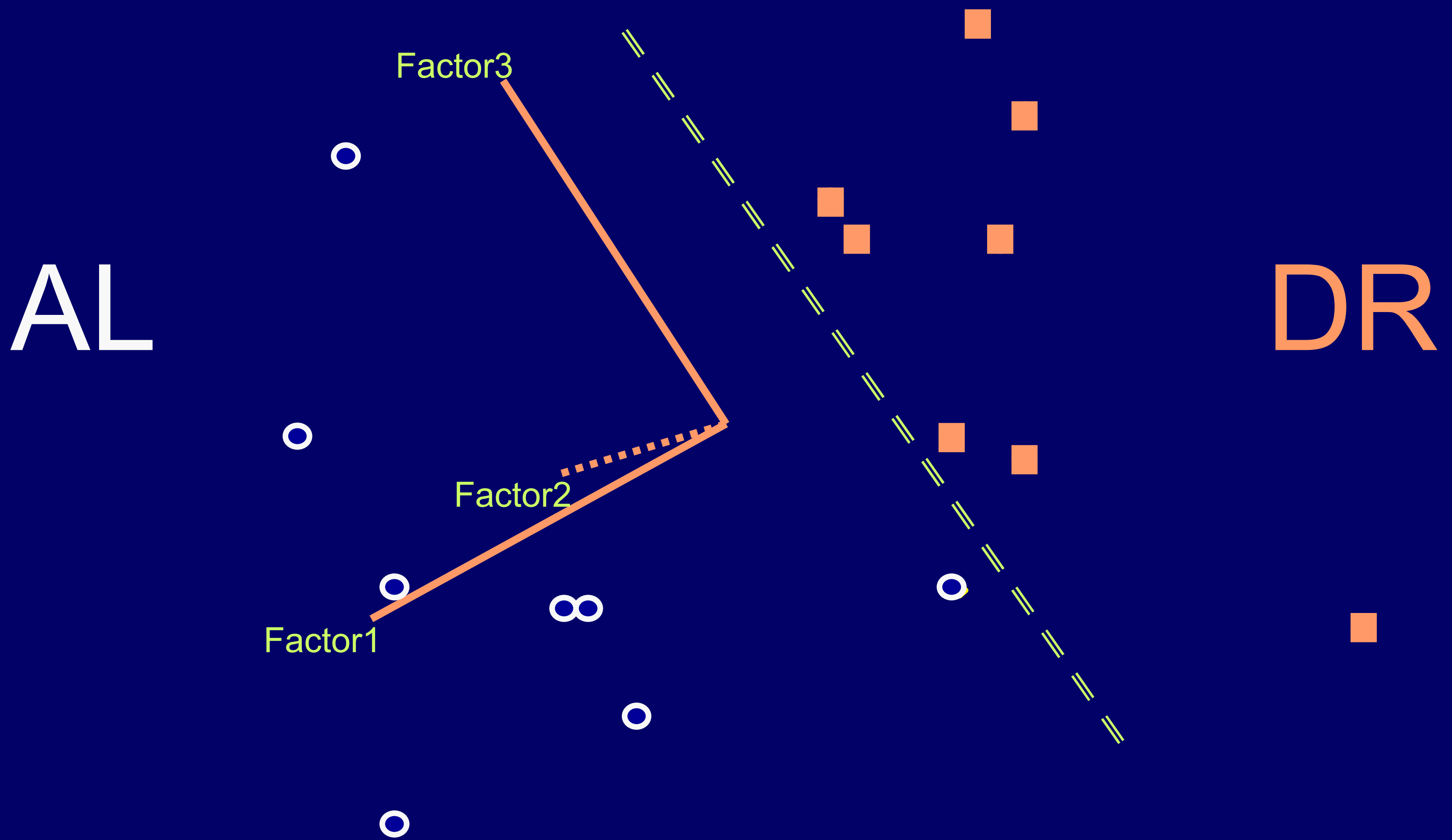
94% Accuracy



94% Accuracy

63 variables (confirmed), independent Cohort #2

PCA Distinguishes AL and DR Female Rats



63 variables (confirmed), independent female Cohort #2

Dataset Reduction:

"Recut" by t-test at $p<0.2$

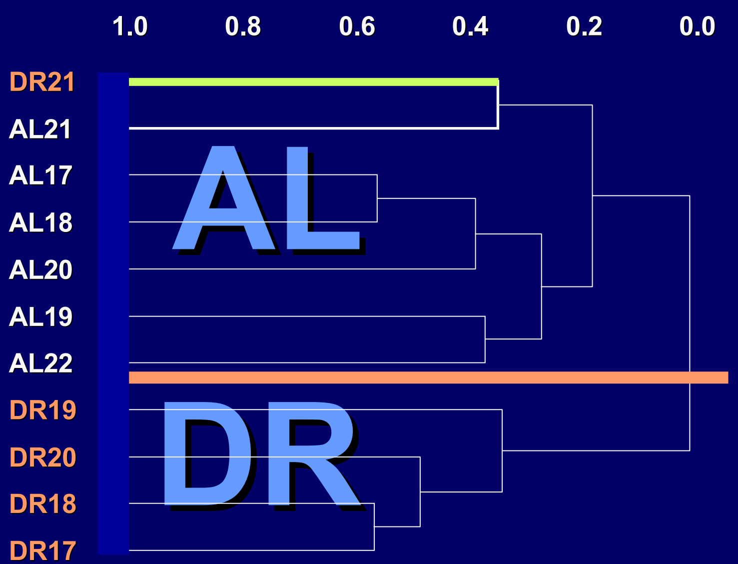
Confirm HCA and PCA separations

Validate in Cohort 3

Removed variables can contribute to AL-DR separations

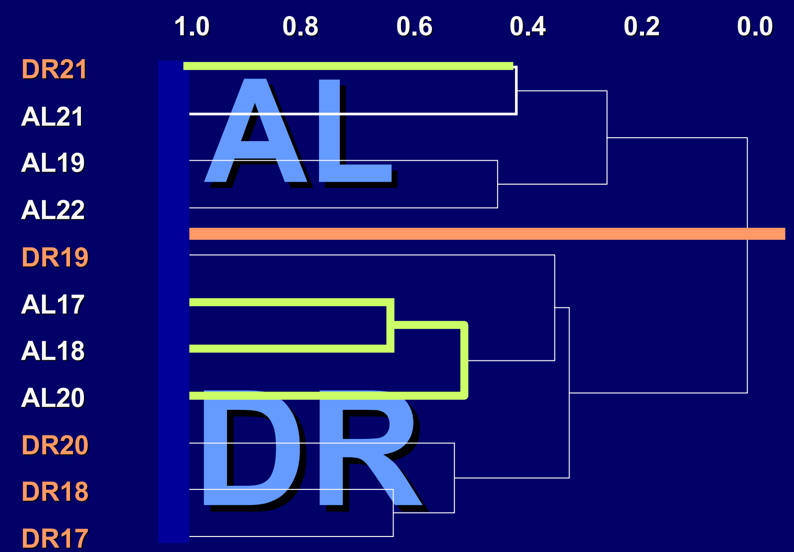
Autoscale, complete

63 Variables



91% Accuracy

36 Variables



63% Accuracy

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

Larger profiles are more robust

Larger profiles are more robust.

Or

Larger profiles are more robust!

Or

Larger profiles are more robust?

“Expert Systems/Supervised Analysis”

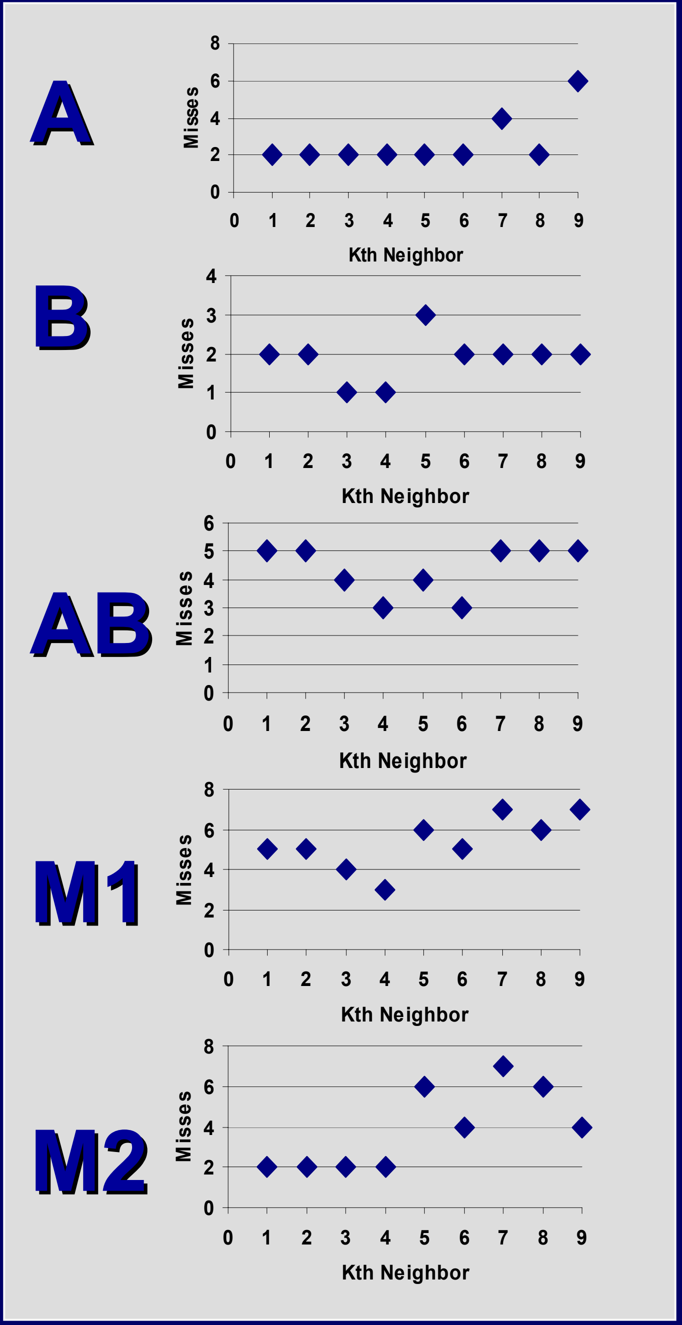
KNN

- 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)

KNN, Males Training Sets



Training Set	Test Set	Accuracy
Cohort A		85%
Cohort B		92%
Cohorts A+B		92%
Mix 1		75%
Mix 2		85%
Cohort A	Cohort B	58%
Cohort B	Cohort A	62%
Mix 1	Mix 2	83%
Mix 2	Mix 1	92%

KNN, MALES

Training Set	Test Set	Accuracy
Cohort A	Cohort A	100%
Cohort A	Cohort A	100%
Cohort B	Cohort B	100%
Cohort B	Cohort B	100%
Cohort C	Cohort C	100%
Cohort C	Cohort C	100%
Mix 1	Mix 1	100%
Mix 2	Mix 2	100%

SIMCA, Females

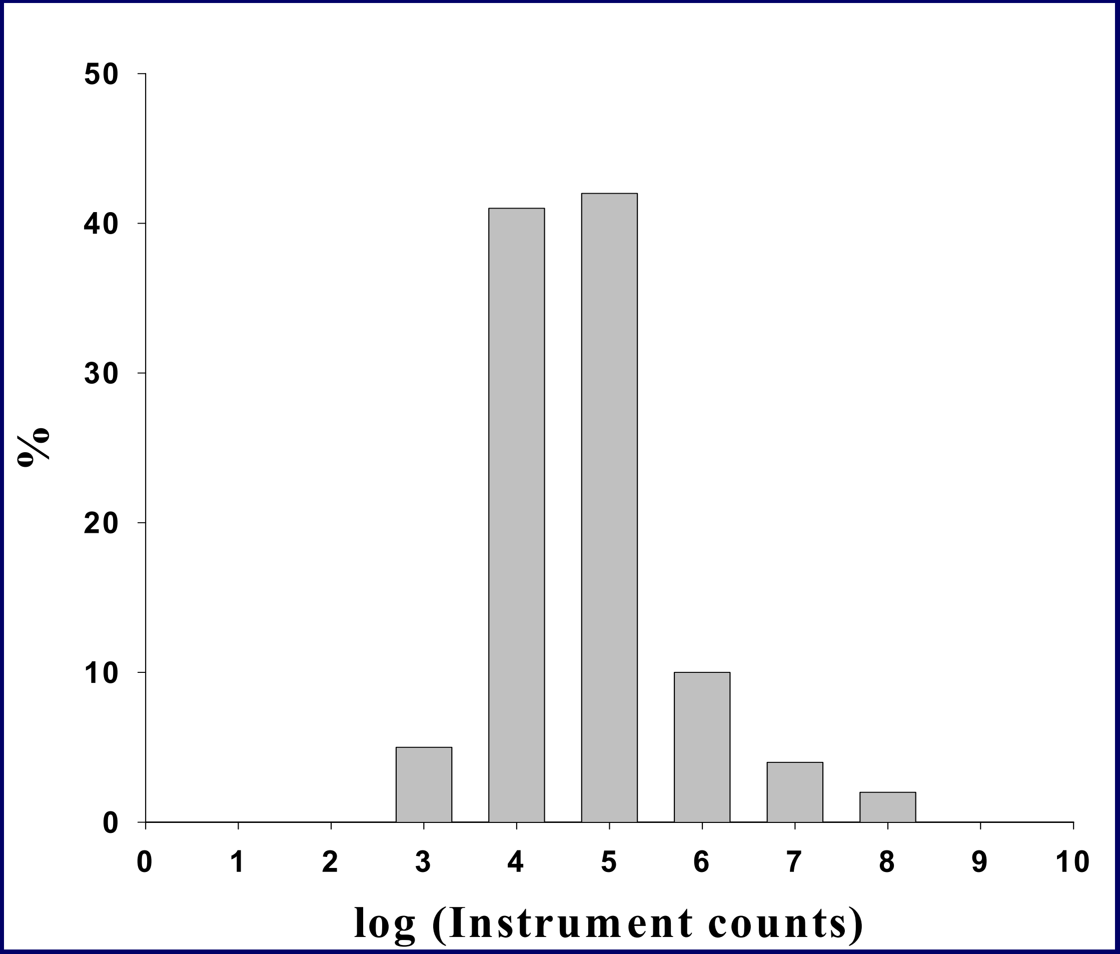
Training Set	Test Set	Accuracy
Cohort A	Cohort B	94%
Cohort A	Cohort C	94%
Cohort B	Cohort A	45%
Cohort B	Cohort C	0%
Cohort C	Cohort A	0%
Cohort C	Cohort B	13%
Mix 1	Mix 2	86%
Mix 2	Mix 1	100%

SIMCA, Females

Cross-Gender Study

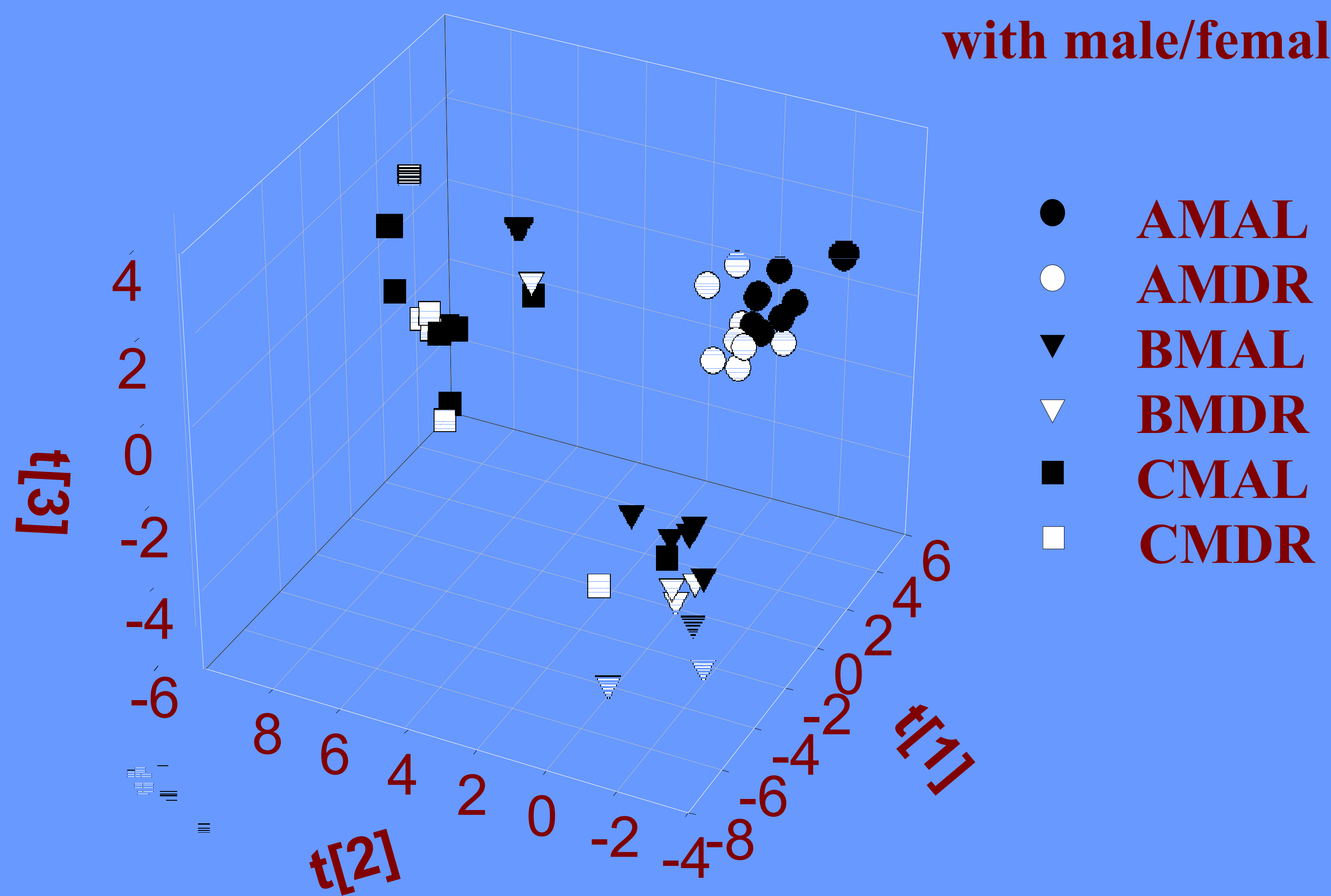
- Combine Datasets
  - Do we have good coverage of the metabolome?
    - Solubility
    - Quantitation (Dynamic Range)
  - Are the metabolites identical?
  - Are the profiles identical
  - Are our informatics appropriate?

Metabolite Distribution

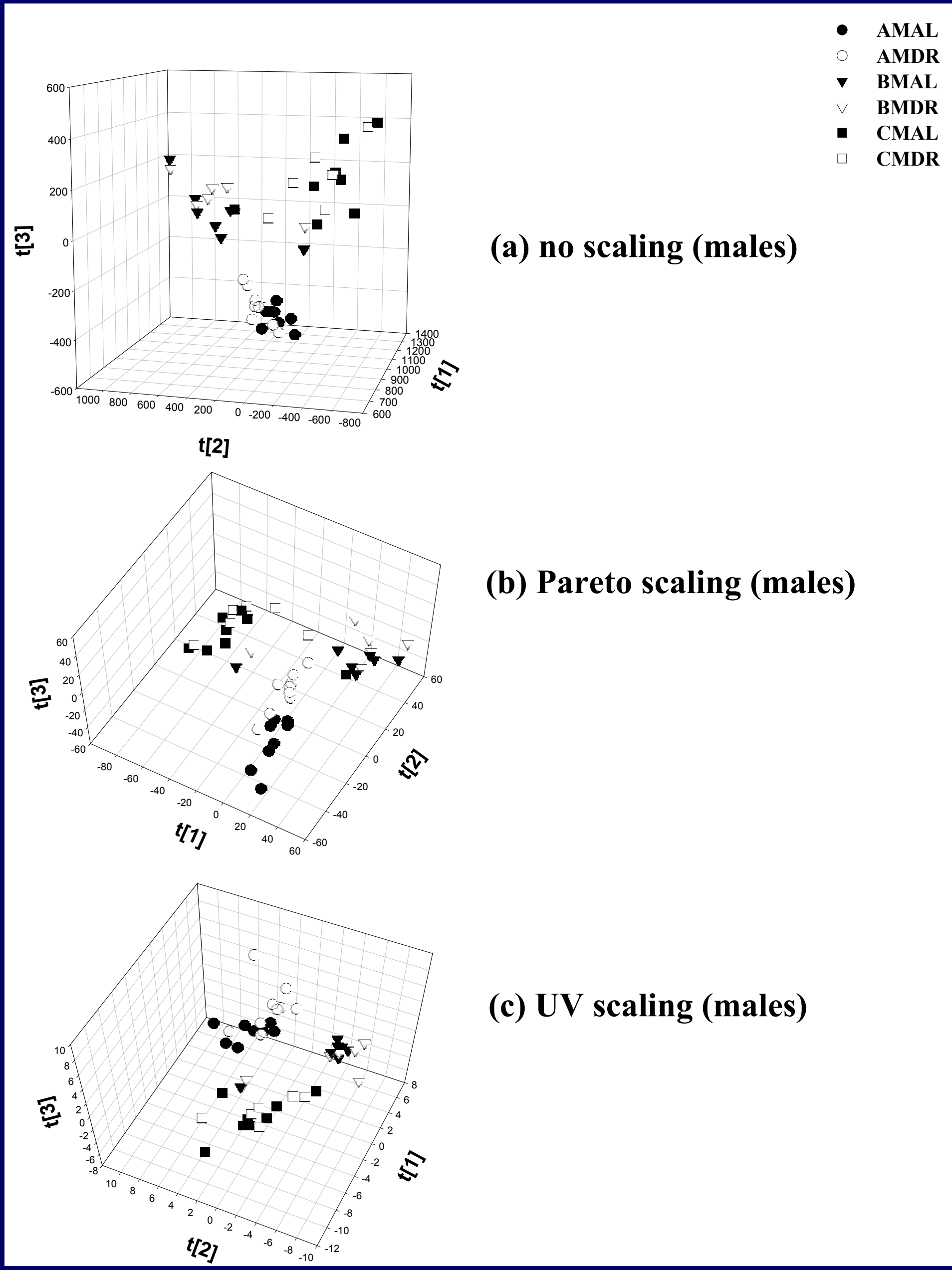


Cohort Separations

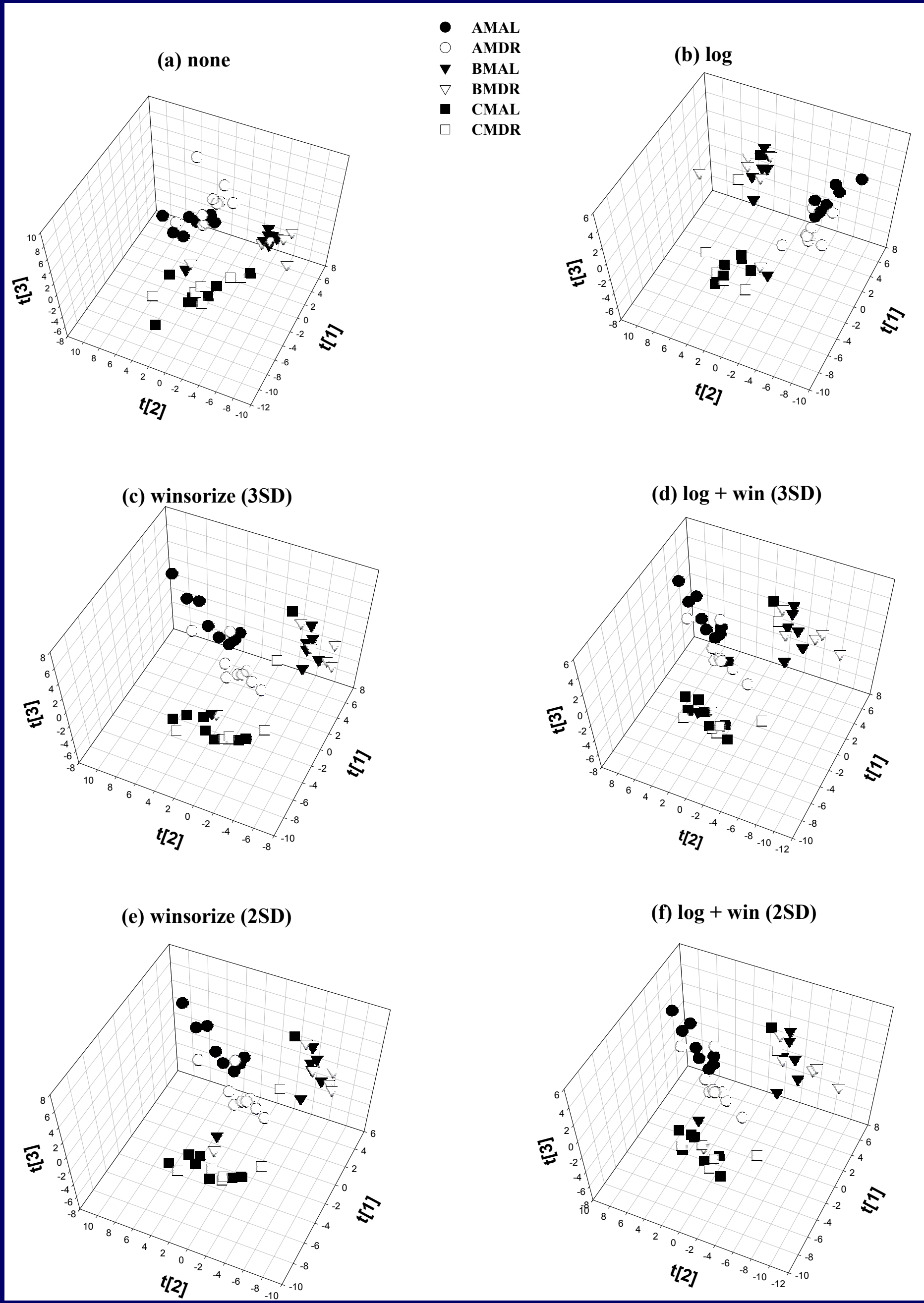
male samples modeled with male/female data set



Scaling Fails To solve the problem



Transforming fails to solve the problem



Data handling does not solve cohort separation problems:

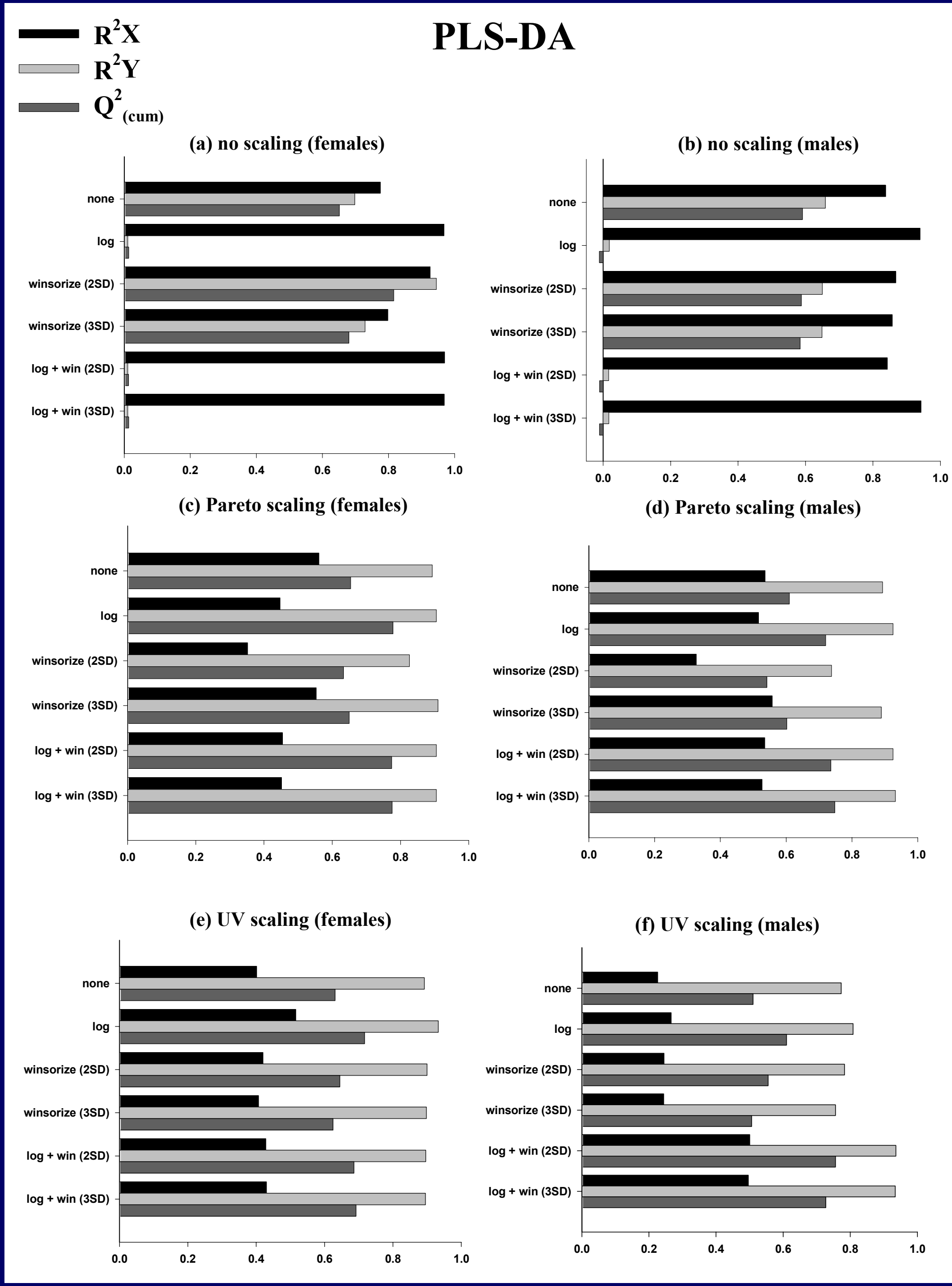
Switch from descriptive to discriminant techniques:

Move to PLS-DA

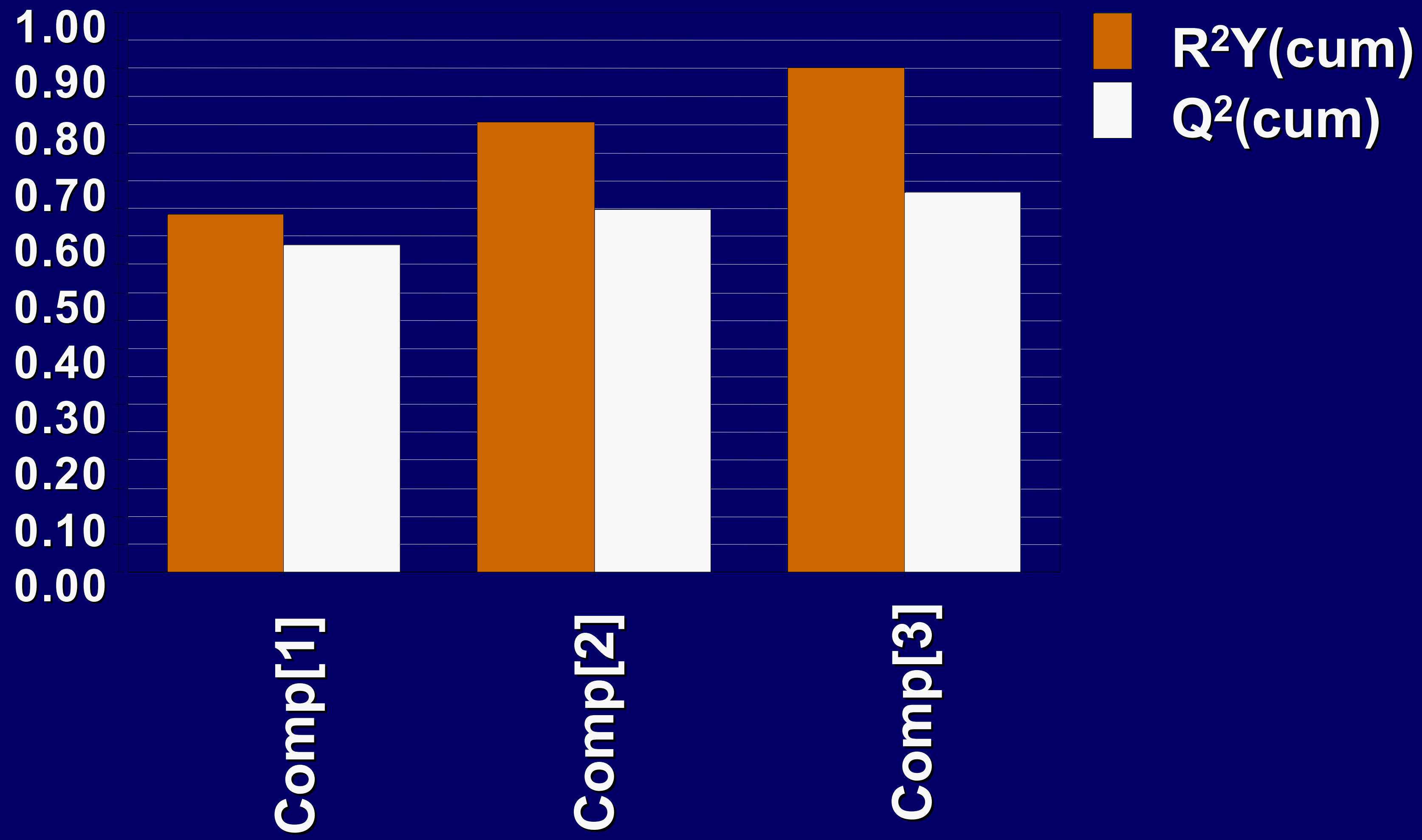
Cross-Gender Study

- Combine Datasets
  - Do we have good coverage of the metabolome? Yes
    - Solubility
    - Quantitation (Dynamic Range)
  - Are the metabolites identical? – 27%
  - Are the profiles identical – No
  - Are our informatics appropriate? Close

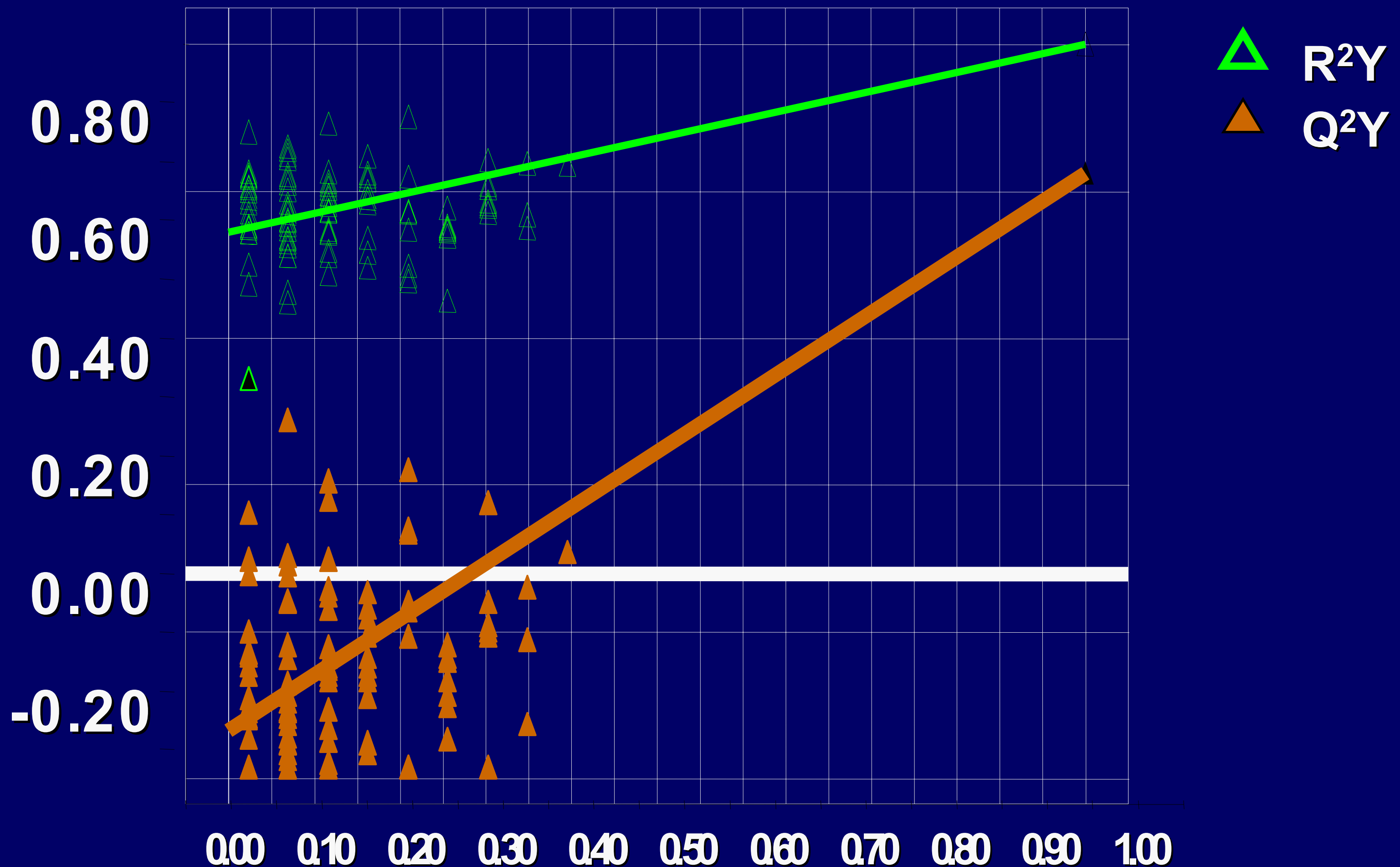
Empirically determine valid parameters



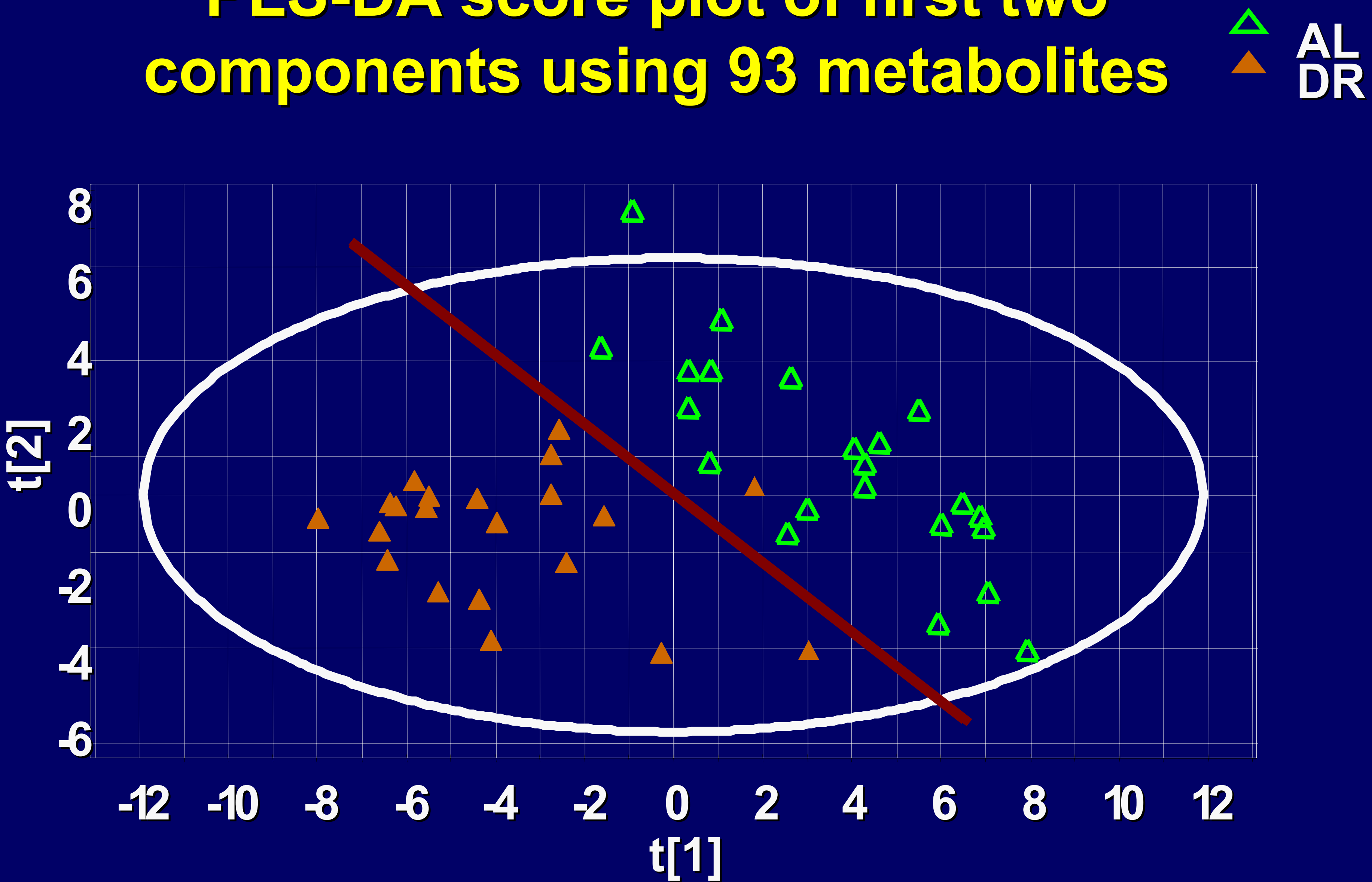
PLS-DA of Female samples (93 metabolites) (3 of 4 components)



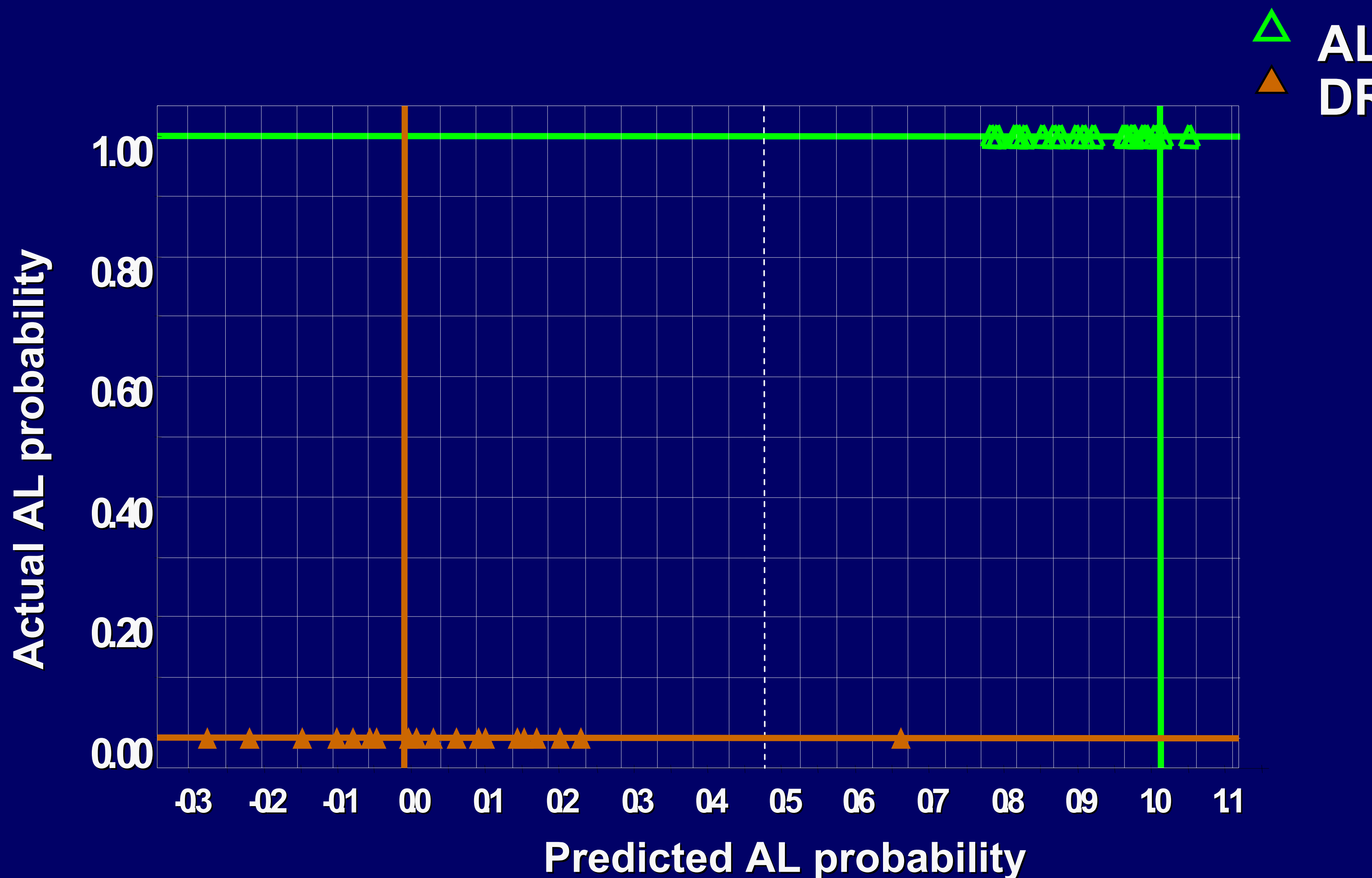
PLS-DA validate plot (after 100 permutations)
Intercepts: R^2=(0.0, 0.577), Q^2=(0.0, -0.268)



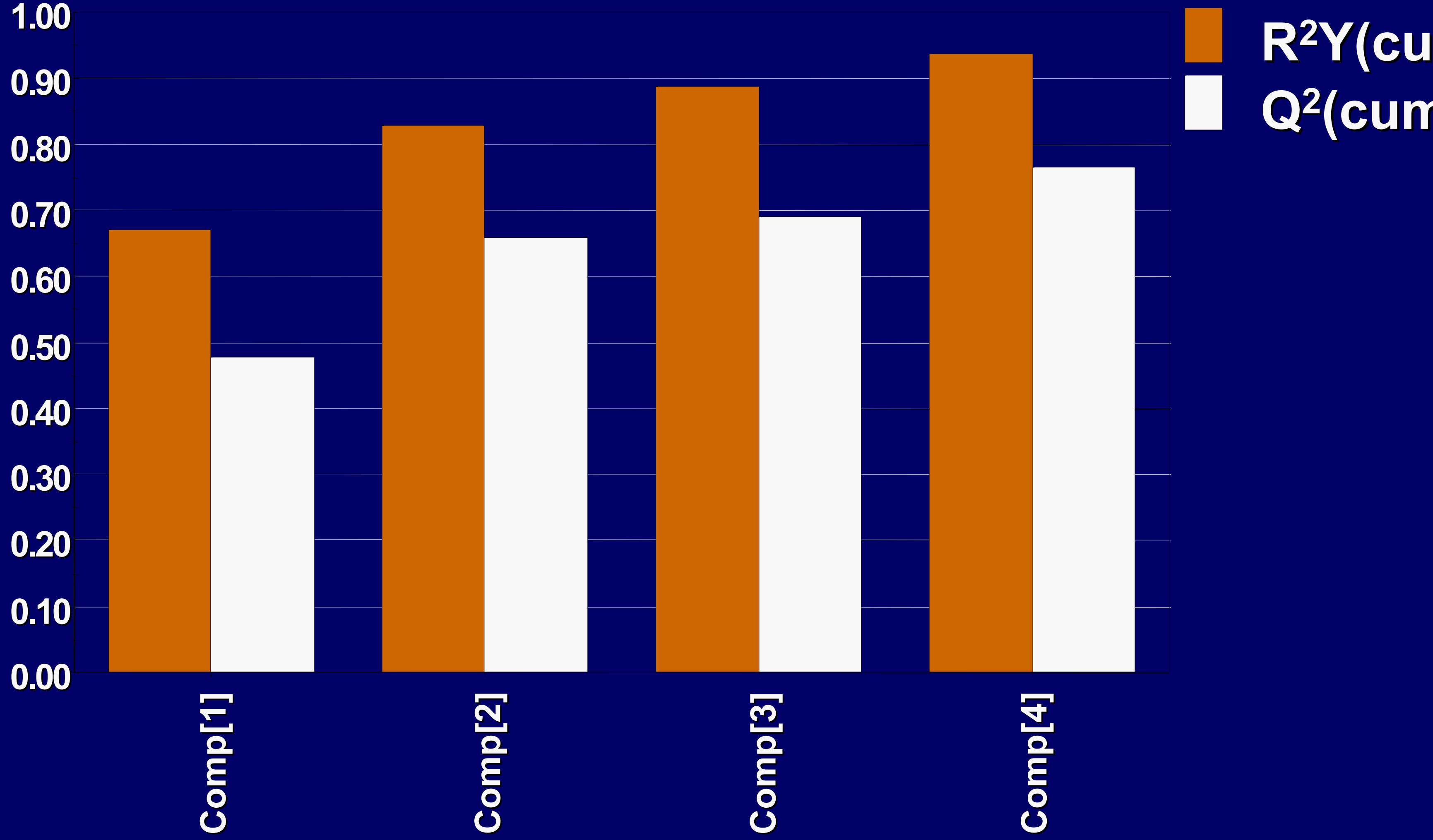
PLS-DA score plot of first two components using 93 metabolites



Observed values vs. Predicted values

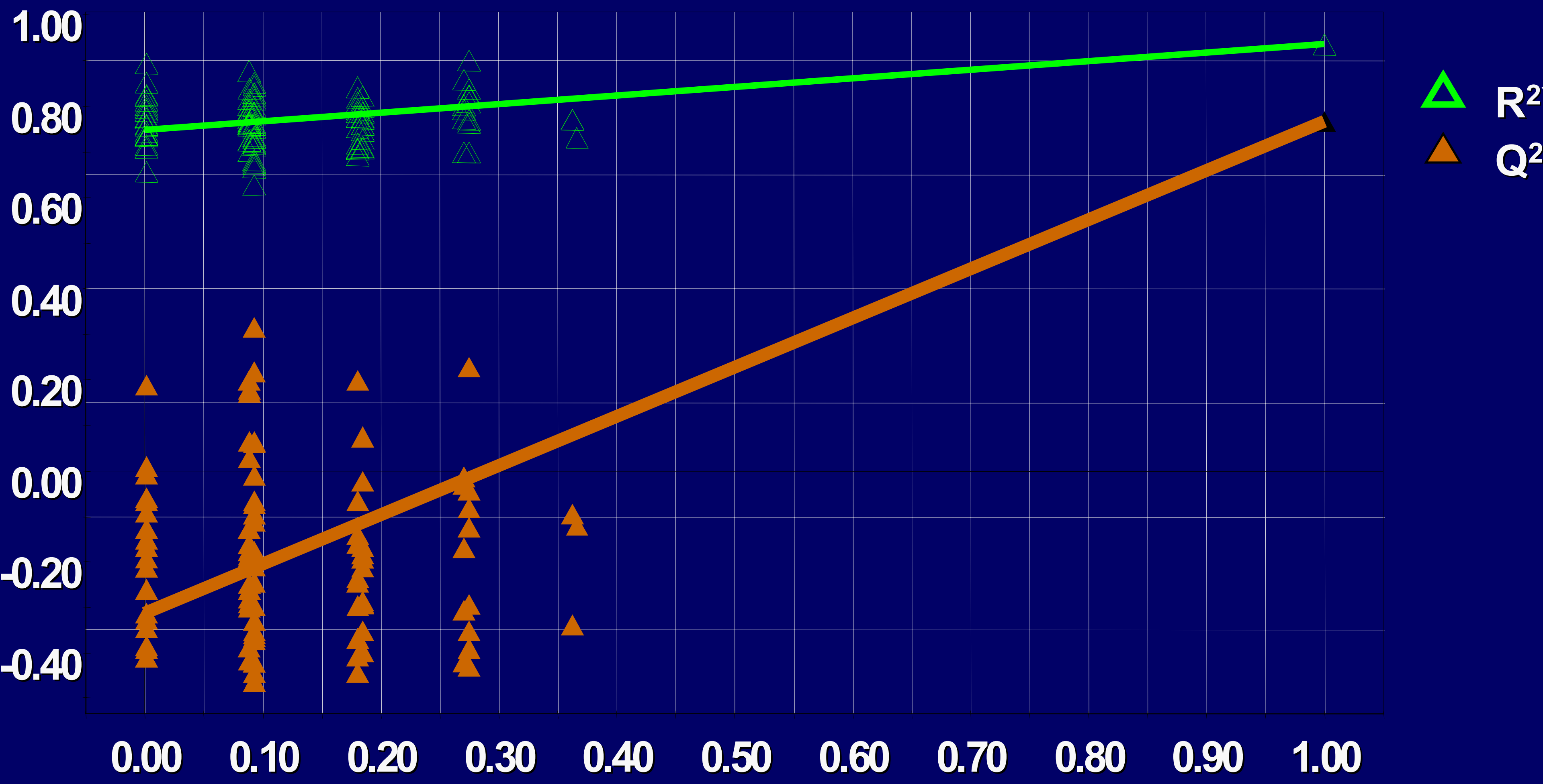


PLS-DA of Male samples using 93 metabolites

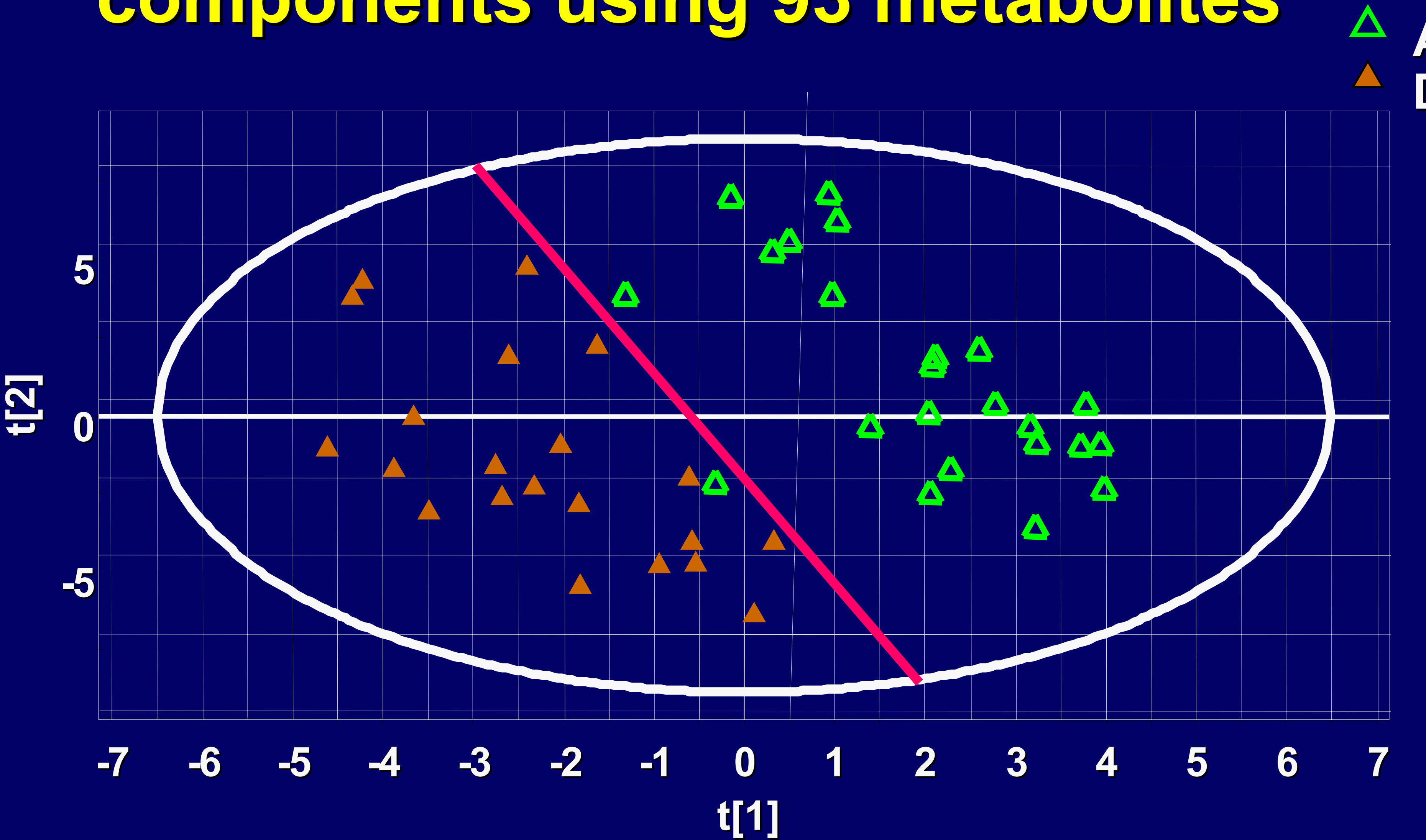


PLS-DA validate plot (after 100 permutations)

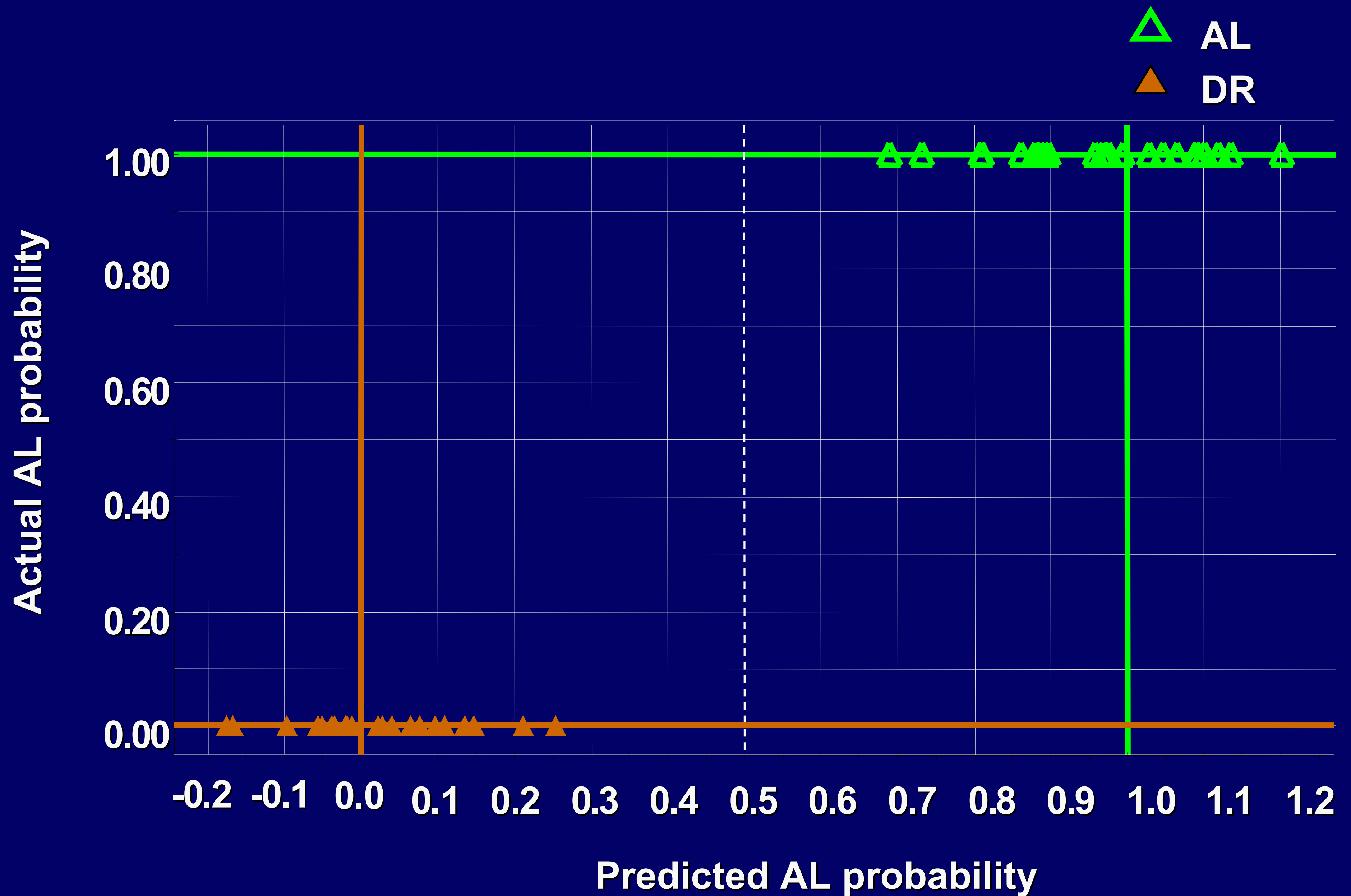
Intercepts: $R^2=(0.0, 0.747)$, $Q^2=(0.0, -0.315)$



PLS-DA score plot of first two components using 93 metabolites



Observed Values vs. Predicted Values



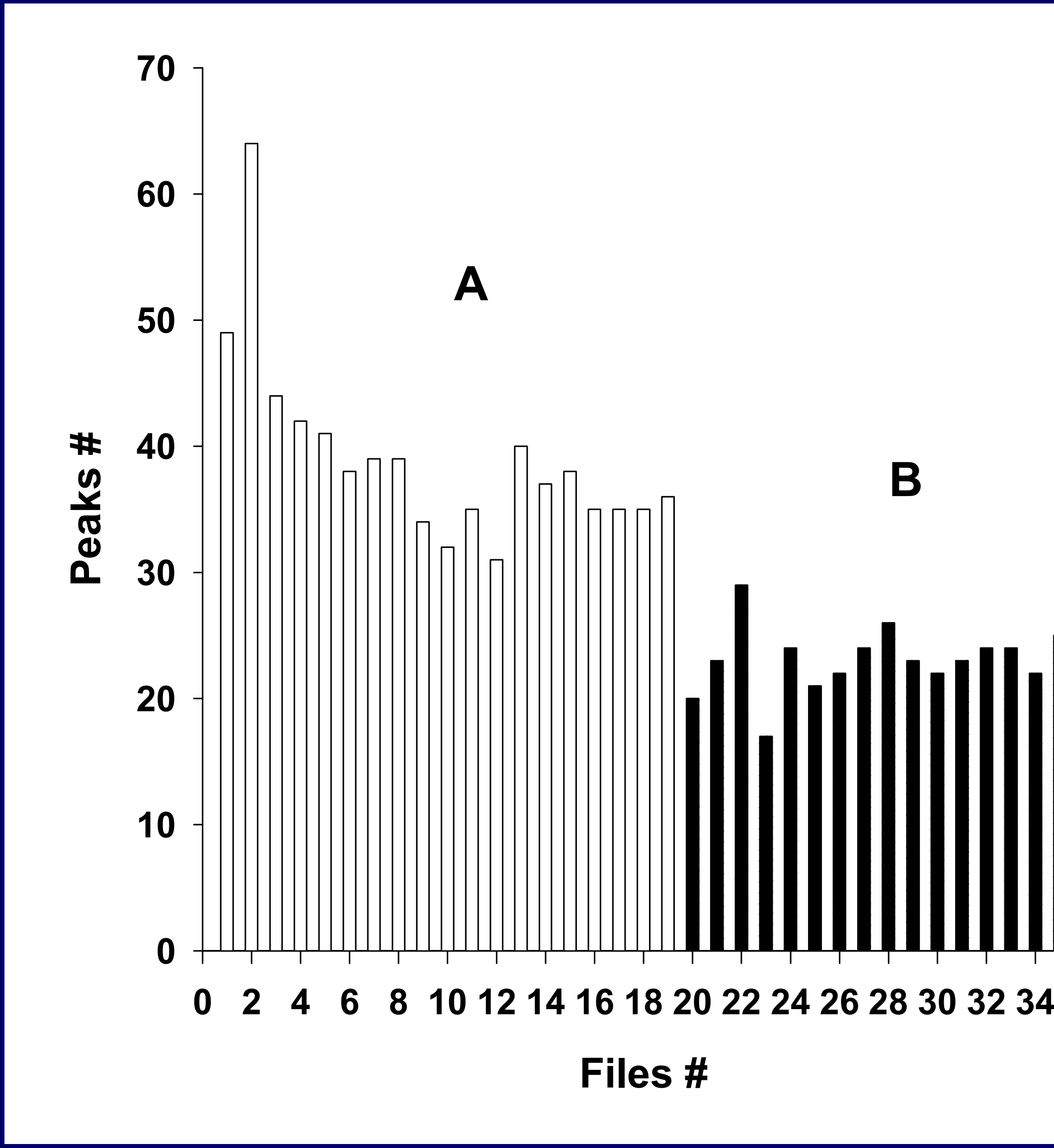
Moving Forward

- Increase samples
- Cross-species analyses
- Increase # of variables measured
- Marker identification/mechanism studies
- Modeling

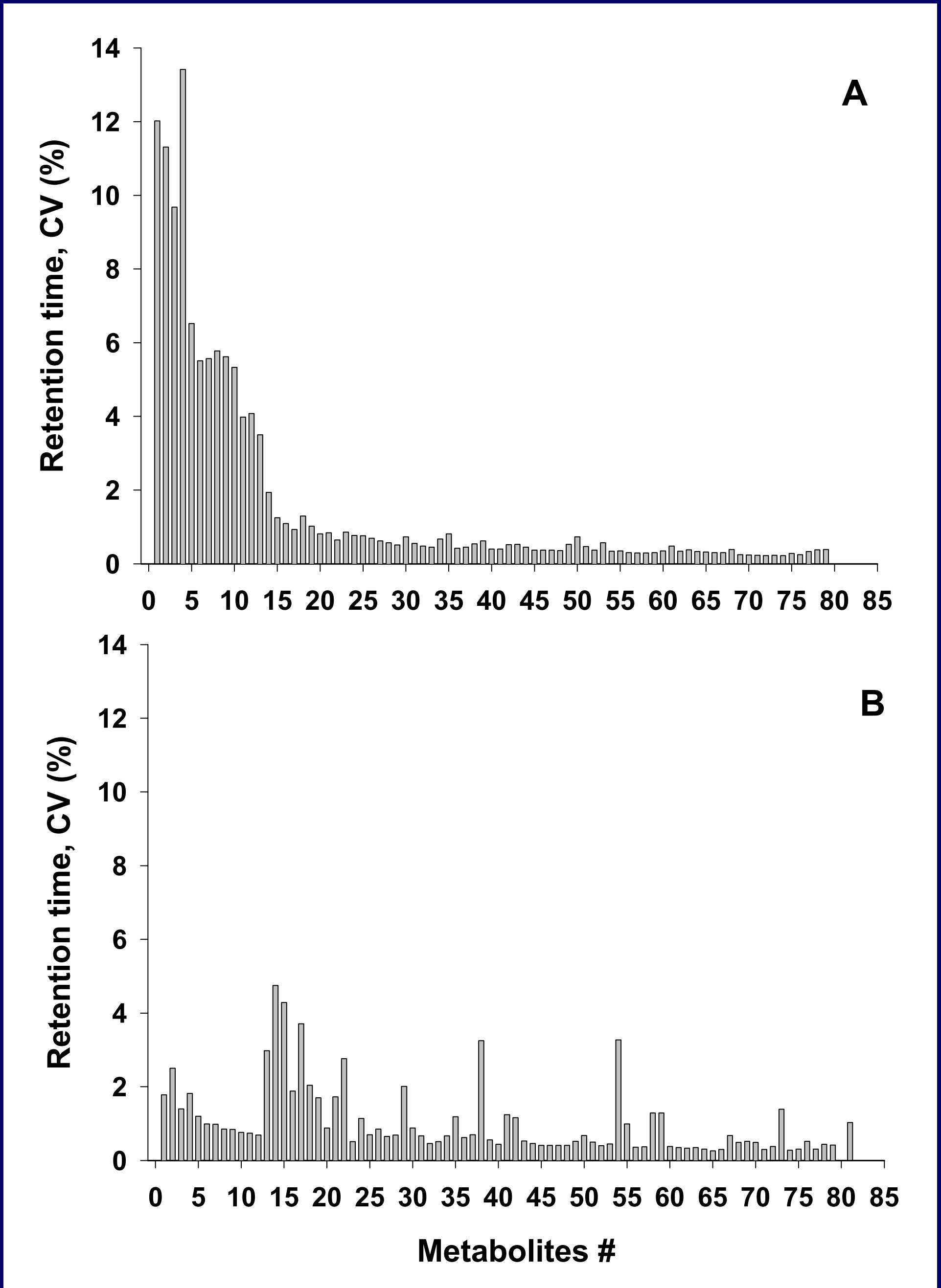
These slides removed to avoid publication problems, please contact me at bkristal@burke.org if you have questions or wish to discuss this data

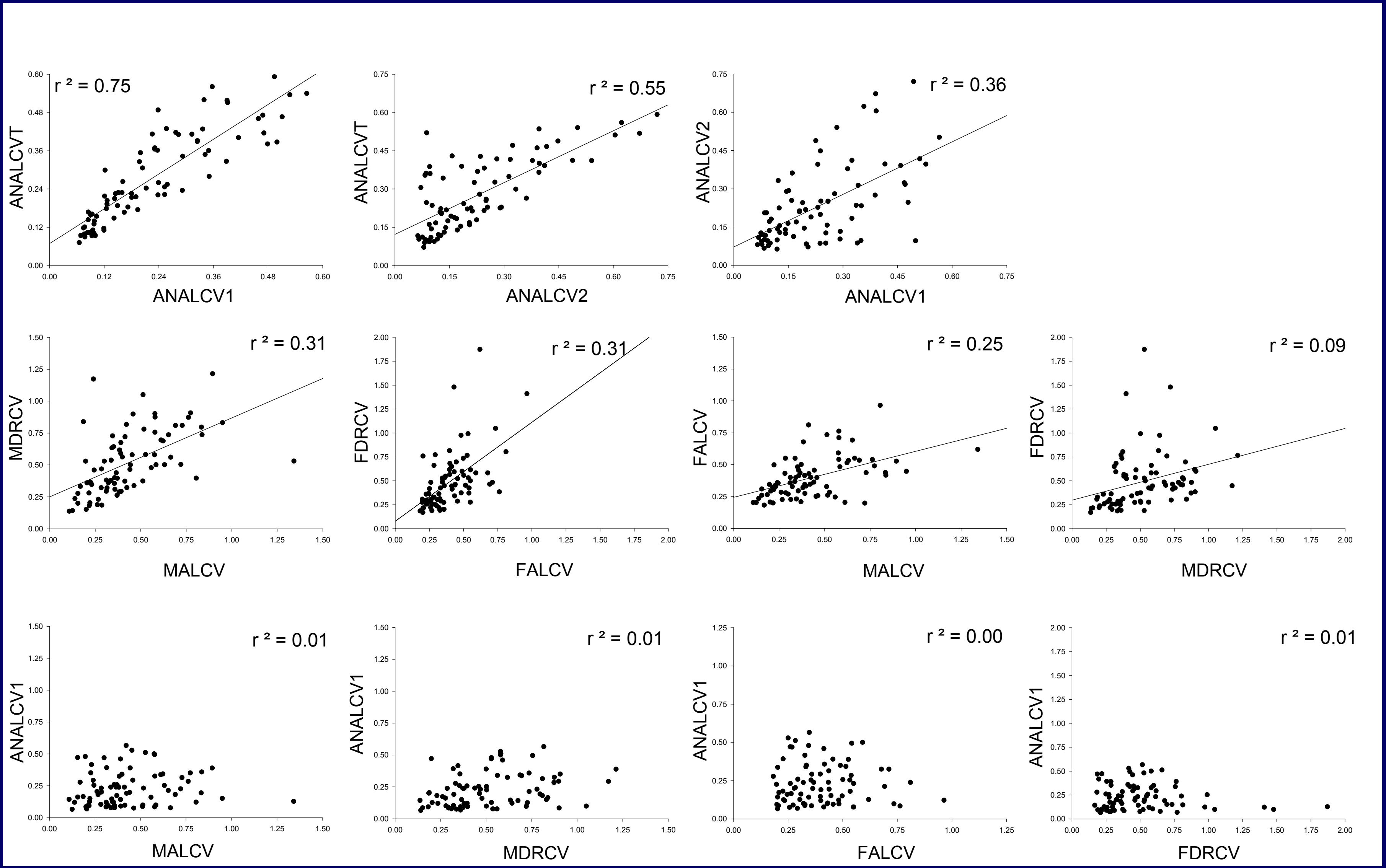
Analytical
Shifts

Auto peak ID before and after column change



Retention time CV before and after column change



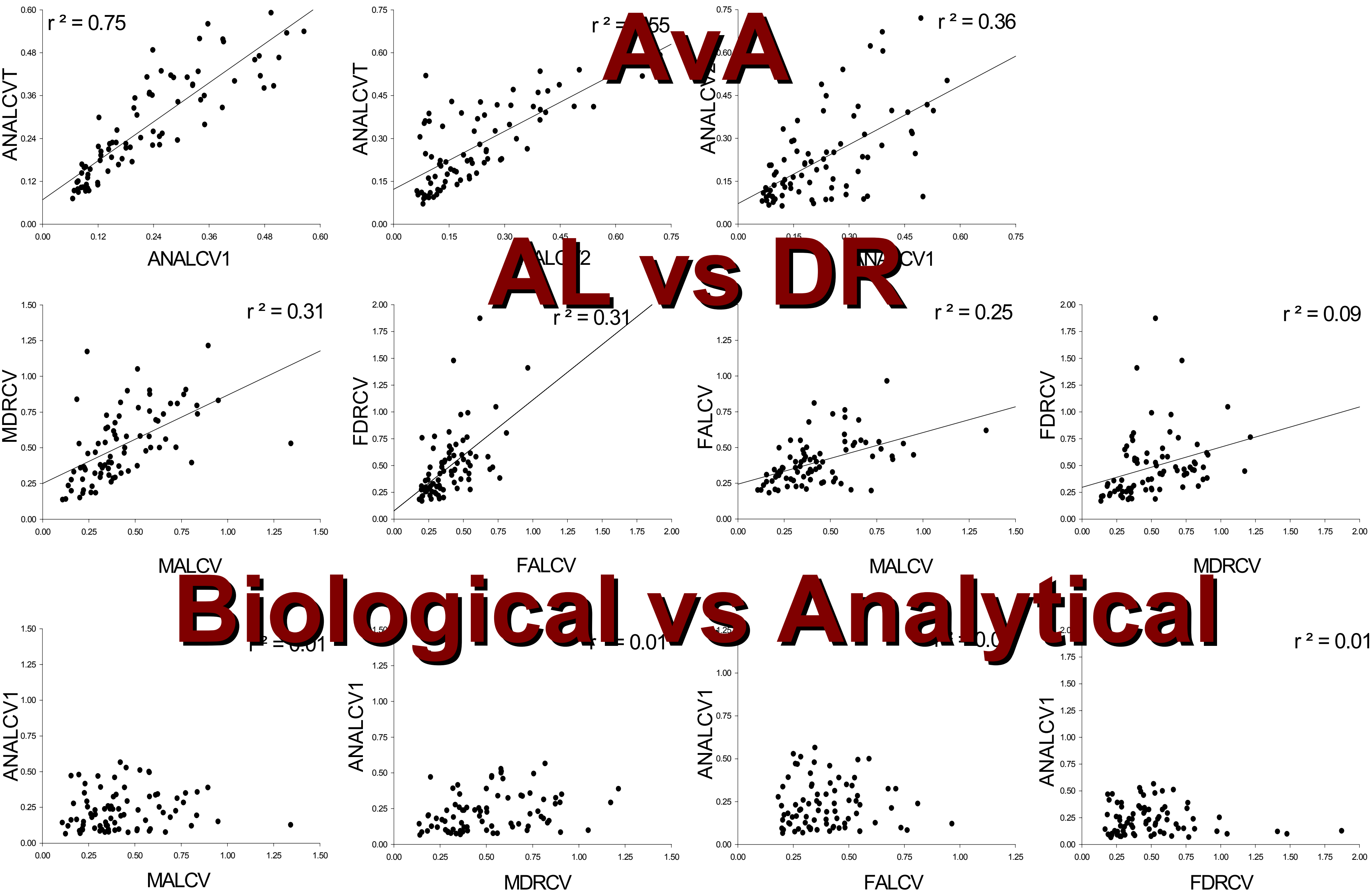


Analytical vs Biological Variation

AvA

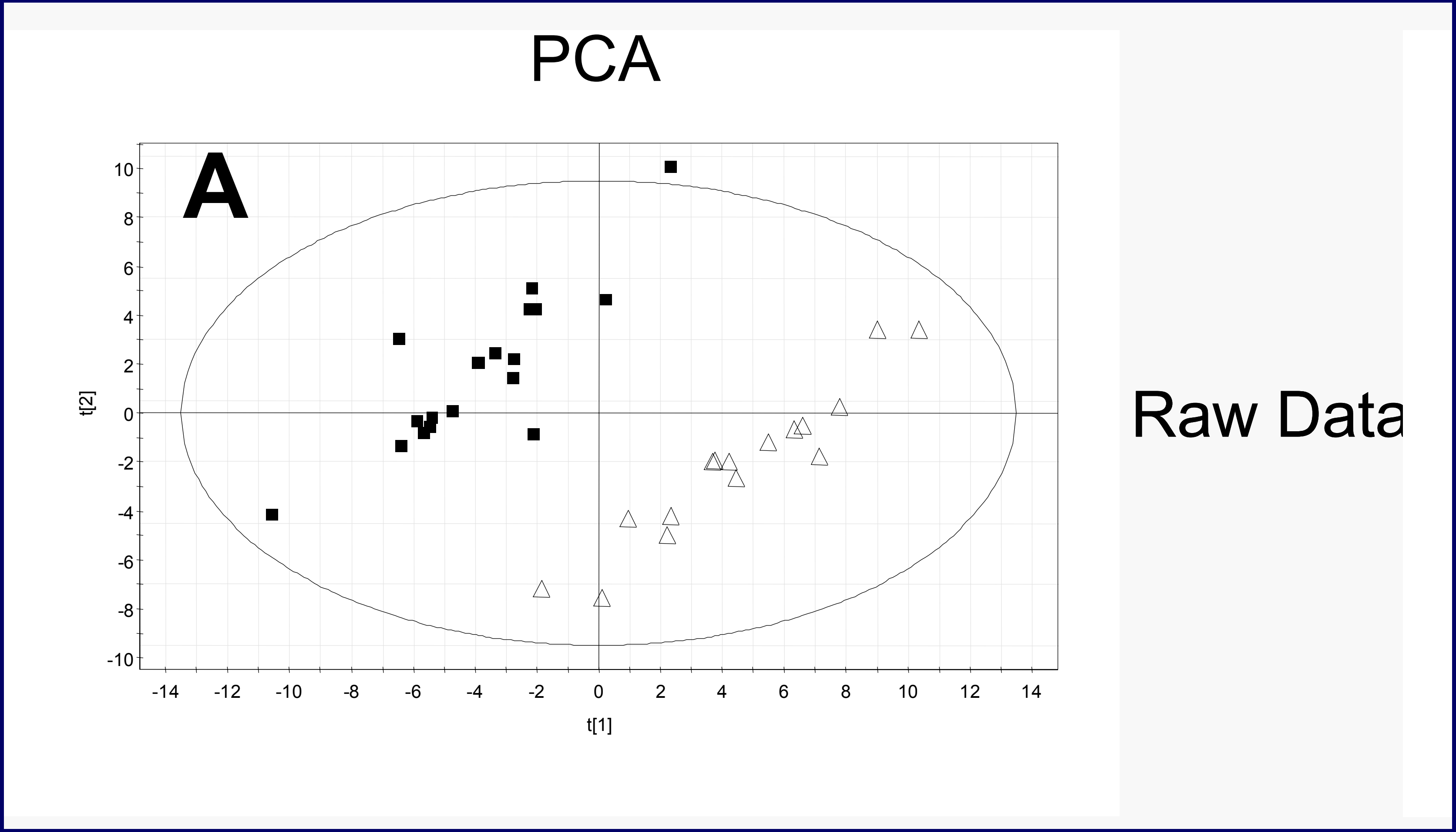
AL vs DR

Biological vs Analytical



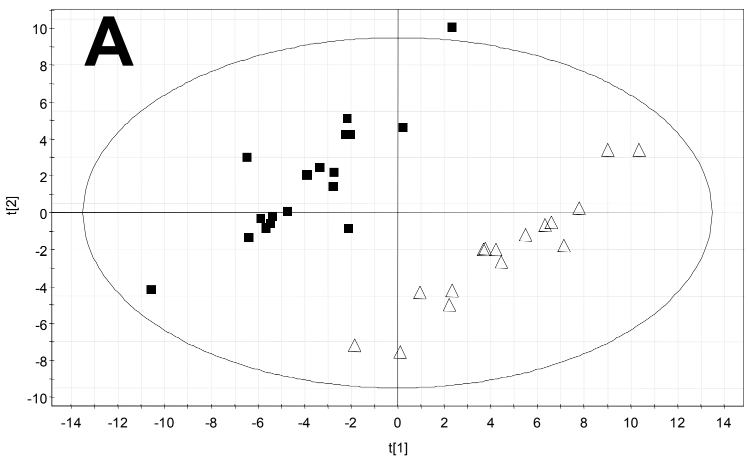
Analytical Problem: Column Crash

PCA Distinguishes before and after pools

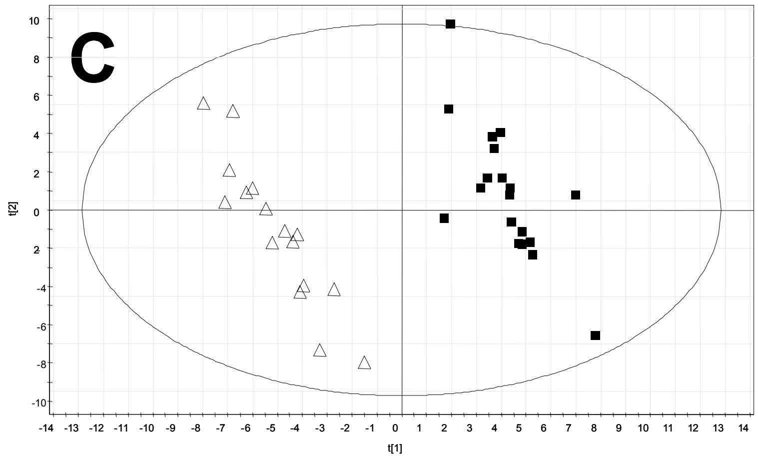


PCA

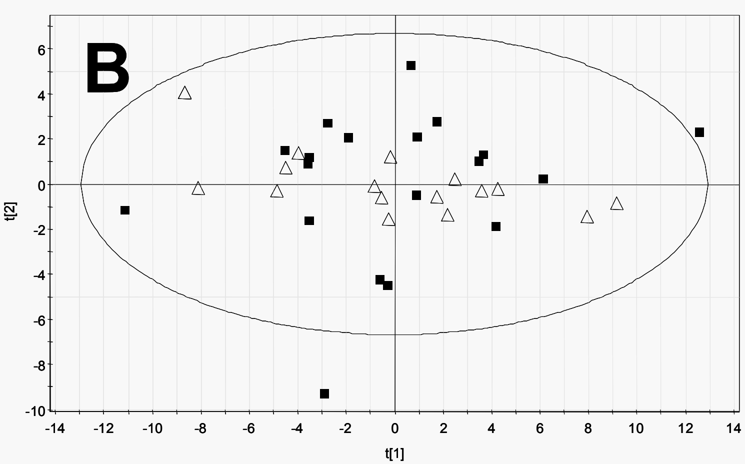
PLS-DA



Raw Data



■ “Set 1” “Set 2” △



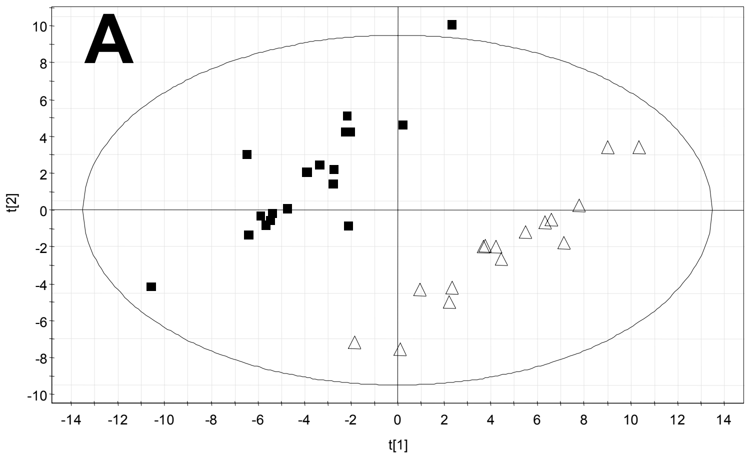
Scaled in k dimensions

D

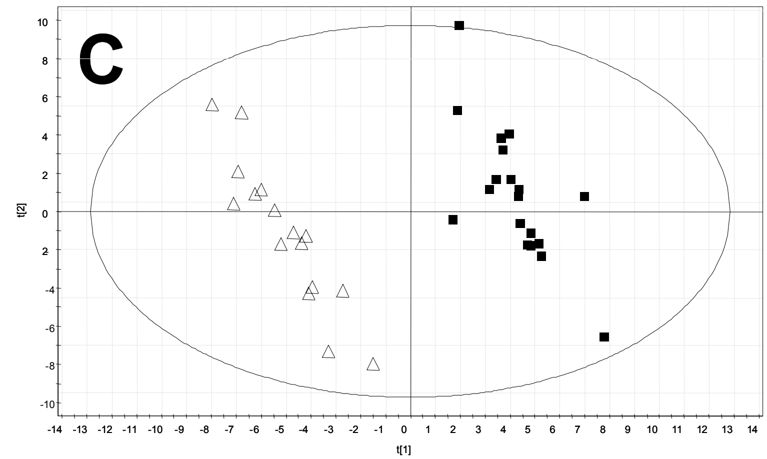
No Valid Separation Remains

PCA

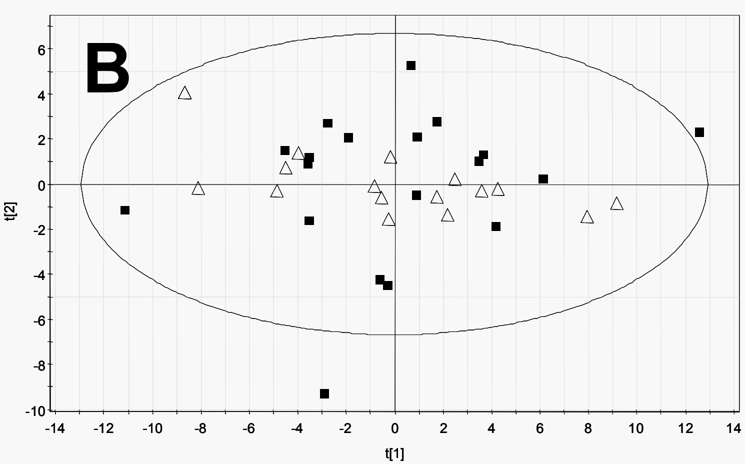
PLS-DA



Raw Data



"Set 1" "Set 2"



Scaled in k dimensions

D

No Valid Separation Remains

Missing Data

Danger:
Preliminary Analysis Only!

Summary

Created and validated a working model of the DR serotype in both male and female rats

- Can identify group of origin
  - 100% training; 85-95% test; 89-94% with expert systems;
  - >97% with optimized projection techniques (PLS-DA)
- Explains the majority of variation in the metabolites
  - Markers range in concentration by 33,000-fold
  - Wide hydrophilicity range
- Reduces dimensionality of the problem
- Identify critical markers

Clarified computational approach

Acknowledgements

Cornell/Burke

Alexander Shestopalov
Honglian Shi
Ugo Paolucci
Yevgeniya Shurubor

ESA, Inc

Wayne Matson
Karen Vigneau-Callahan

Paul Milbury
Tom Vogl
Walter Willett
Sue Hankinson
Rolf Martin

Funding

NIH (NIA); ESA

Statistics/ Informatics Positions Available