

# Statistical Validation of Metabolic Differences and Variable Importance

---

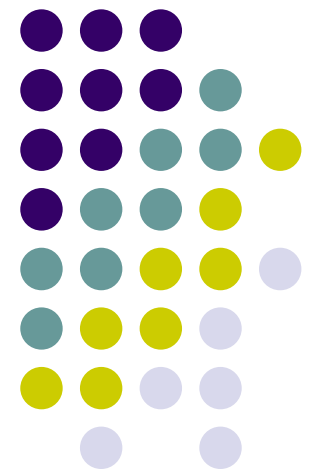
**Johan Westerhuis**

*Biosystems Data Analysis*

*Swammerdam Institute for Life Sciences*

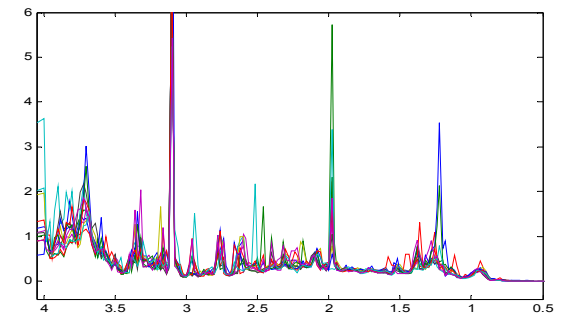
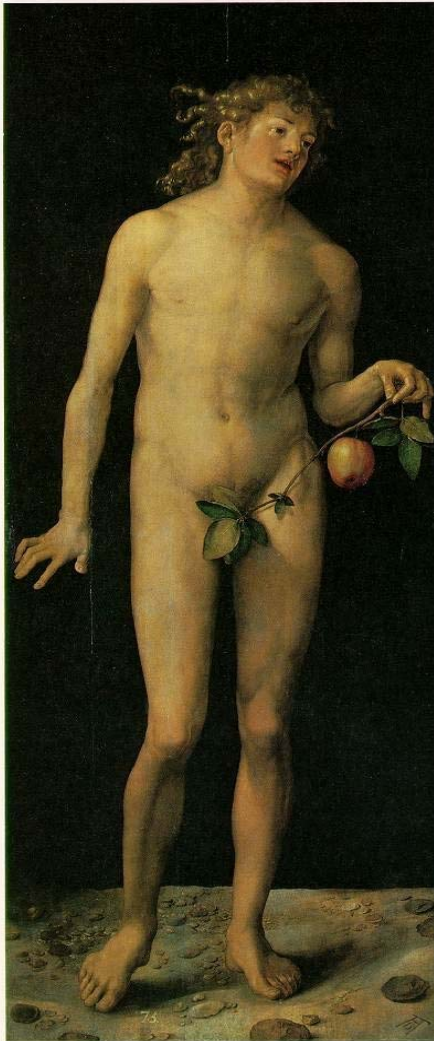
*Universiteit van Amsterdam*

***[www.bdagroup.nl](http://www.bdagroup.nl)***

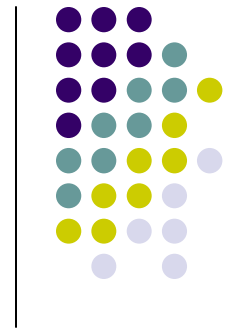


# Do men and women differ ?

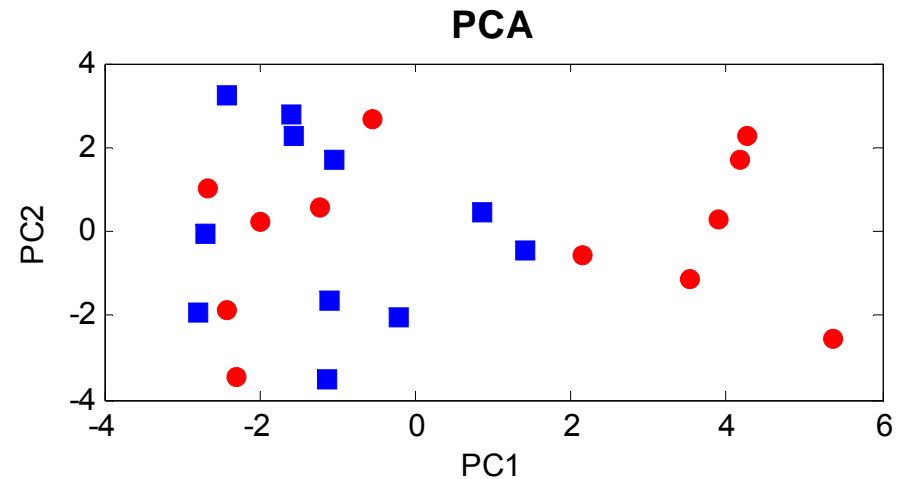
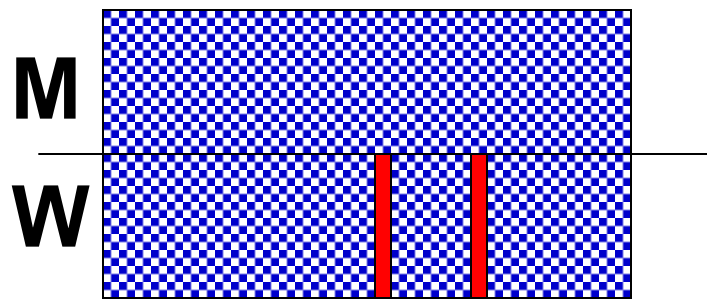
# Which are the important biomarkers ?



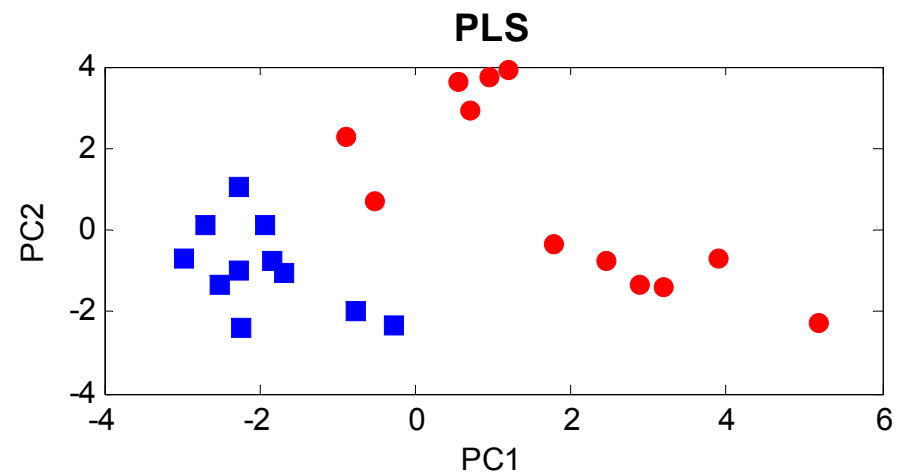
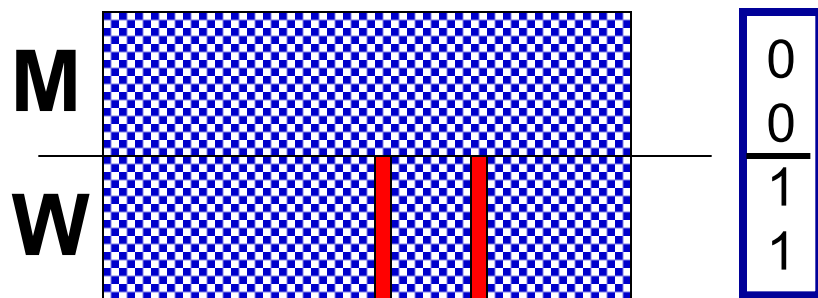
# Classification / Discrimination



- Nonsupervised PCA



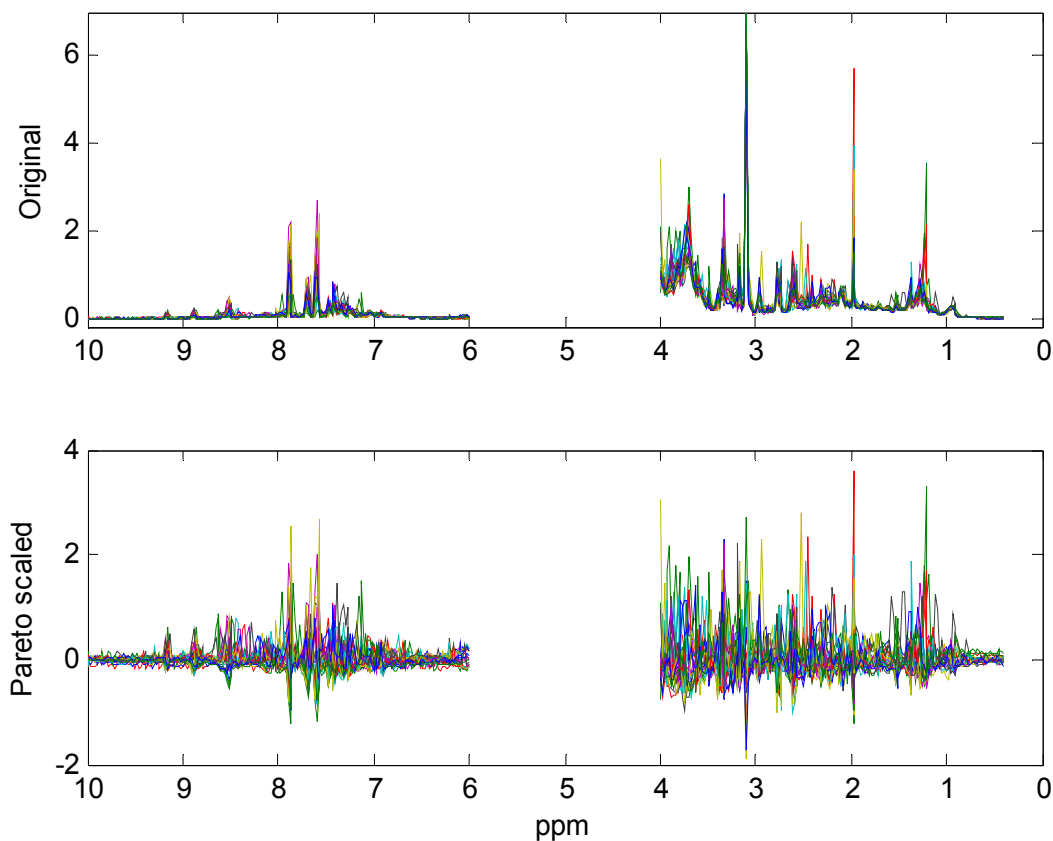
- Supervised PLSDA



# Discrimination by gender

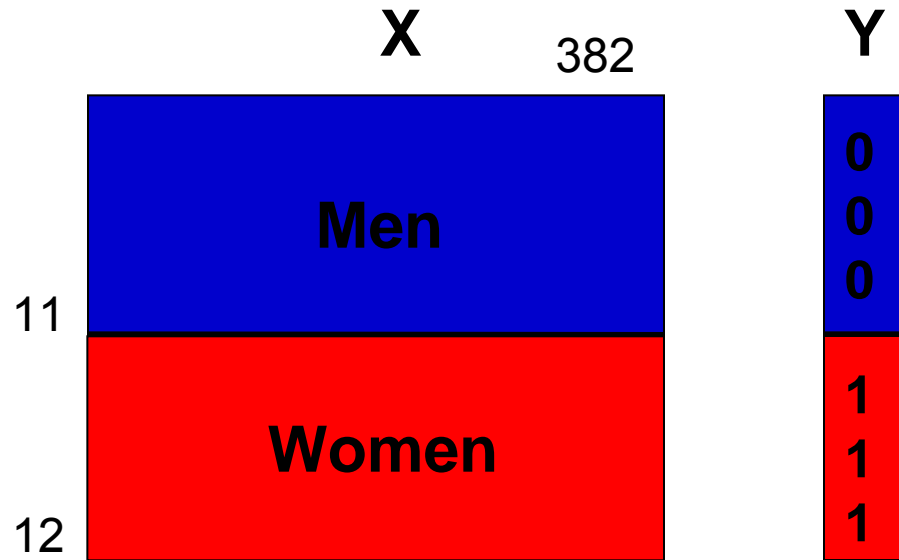


## NMR spectra

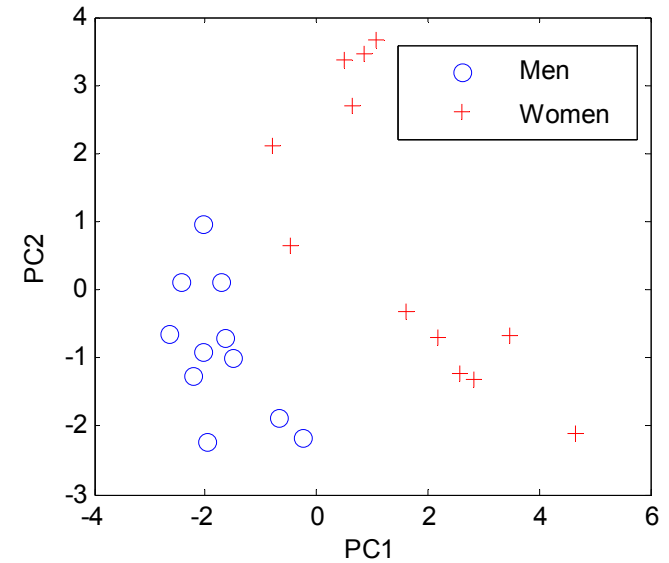


- Urine samples of healthy volunteers: women (12) and men (11).
- Spectra (0.4-4, 6-10 ppm) were binned (0.02 ppm).
- Each bin normalized to account for volume differences
- Paretoscaling (meancenter + scaling by  $s^{1/2}$ ) before analysis

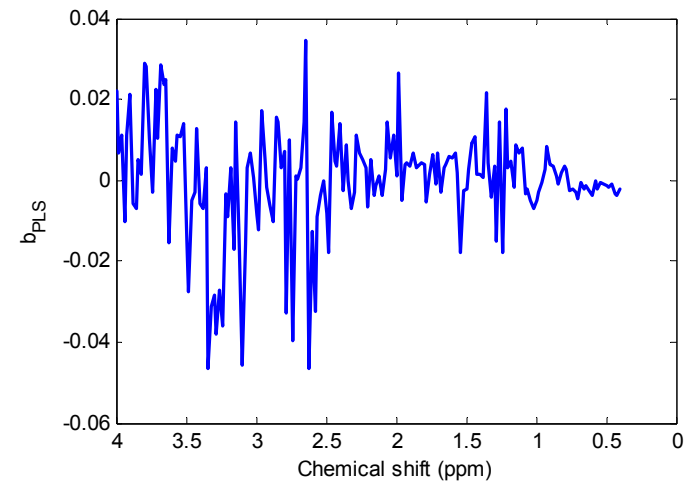
# Typical PLSDA application

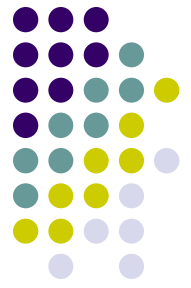


$$Q^2 = 0.27$$



- Is there really a difference between the groups ?
- What is the meaning of  $Q^2$  ?
- Which are the most important peaks for discrimination ?



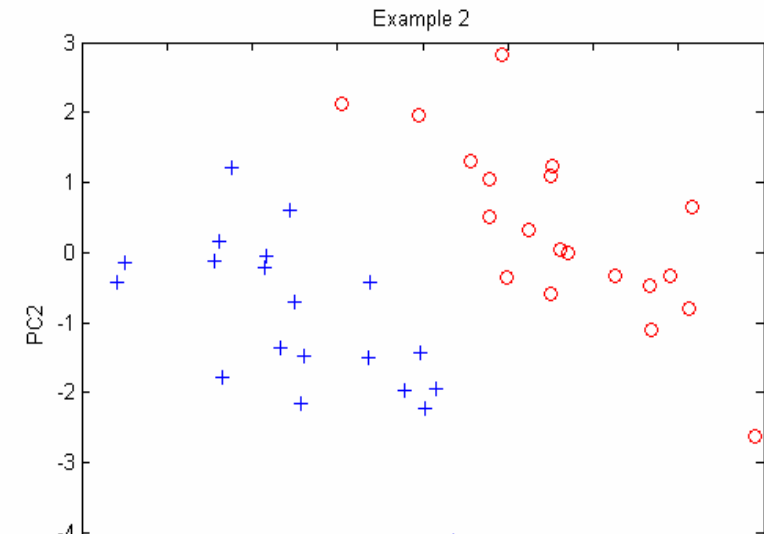
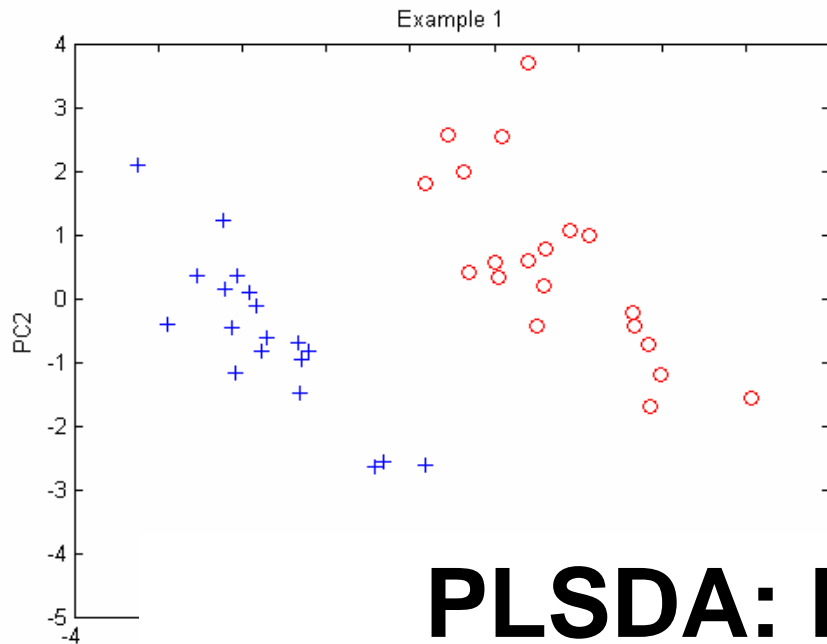


# First test using random data

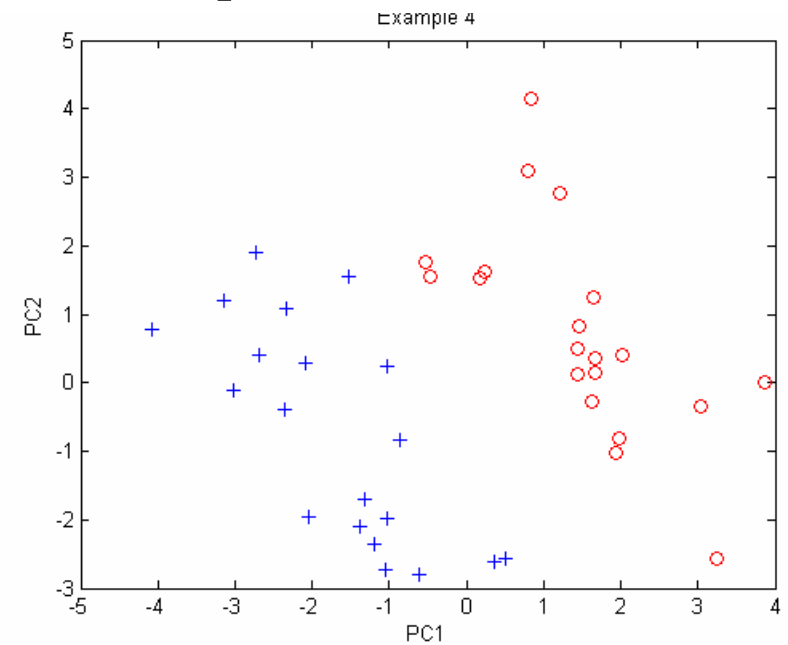
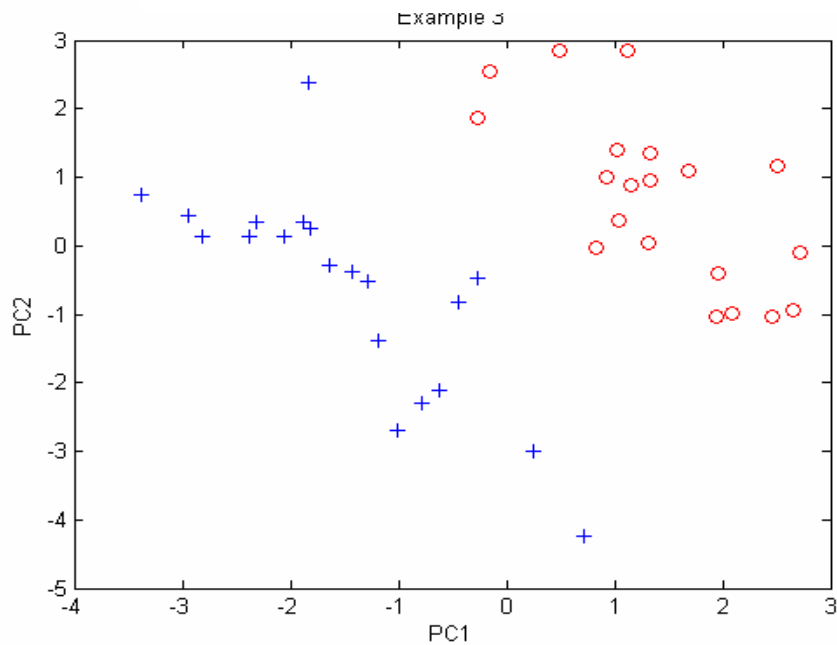


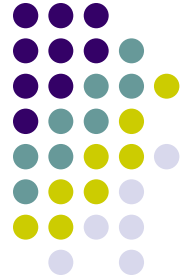
- Make PLS-DA model between X and Y.
- No difference between groups expected !

# Simulations results



## PLSDA: Eager to please





Summary recommendations for  
standardization and reporting of  
metabolic analyses

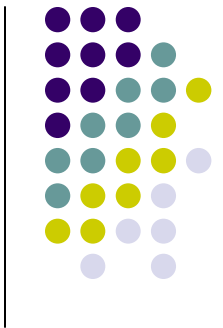
om/naturebiotechnology

## Cross validation:

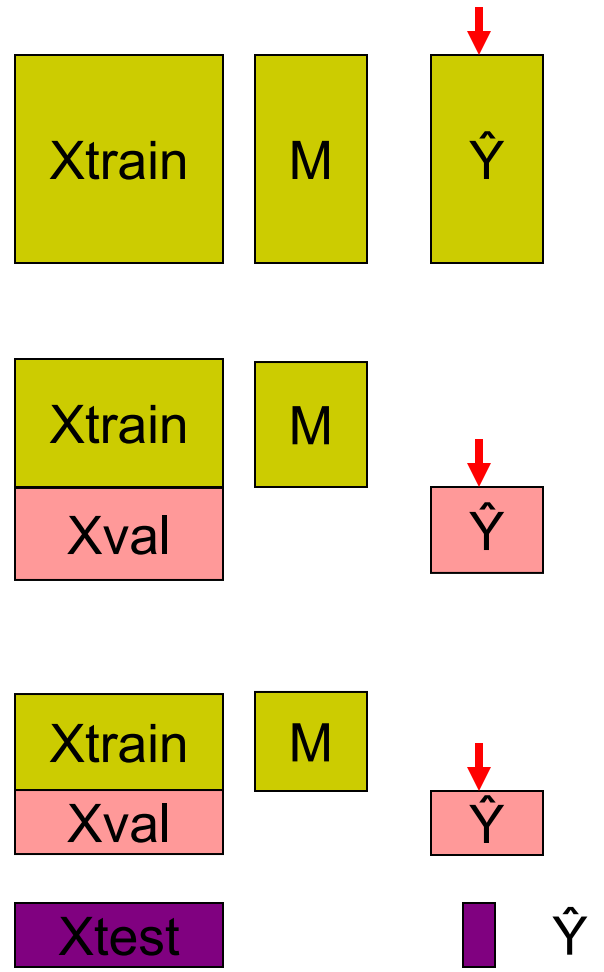
Each observation must be left out in turn and the prediction error calculated in each case. These may then be averaged to give an overall model predictability.



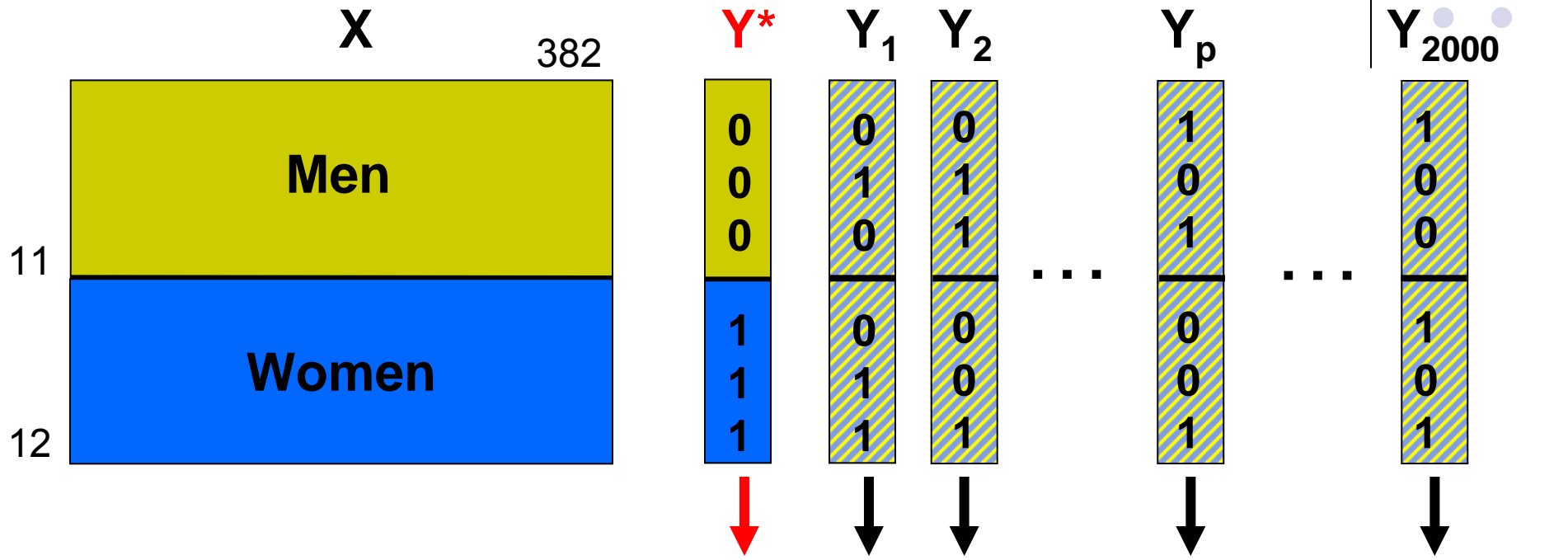
# Cross validation



- Cross validation to make better use of small number of samples
- FIT:
  - Predict class using model based on same samples
  - **BAD**
- 1CV (single cross validation)
  - Validation set is used to select number of components
  - **Much better**
- 2CV (double cross validation)
  - Predict class of separate test set not used for modelling
  - **Perfect**



# Cross validation assessment using permutation testing



**H0: Class 0 = Class 1**

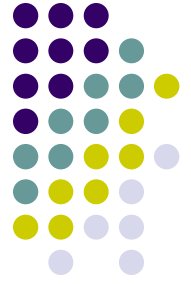
Calculate distribution of any statistics for H0:

Make model:  
Predict class

} FIT  
1CV  
2CV

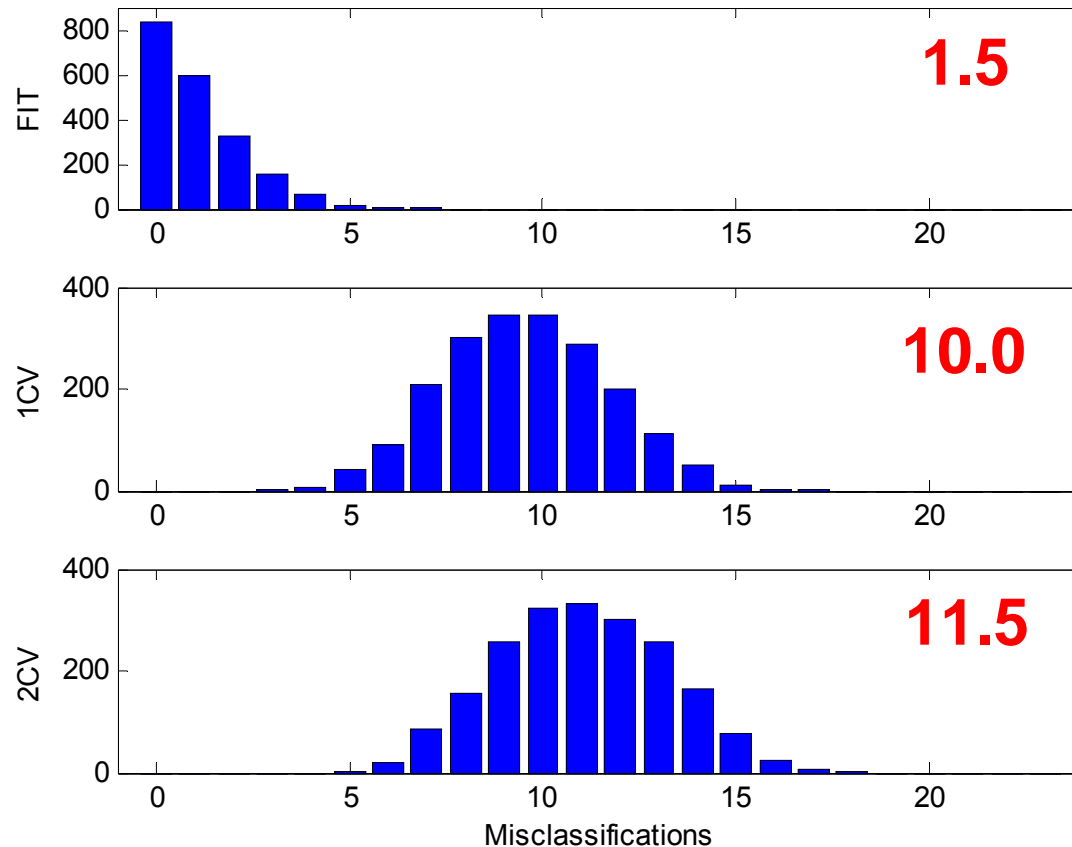
Misclassifications  
 $Q^2$   
Area under ROC curve

# Evaluation of CV procedure



- **2000 permutations**

- 20 times cross validation using FIT, 1CV, 2CV
- Calculate average number of misclassifications

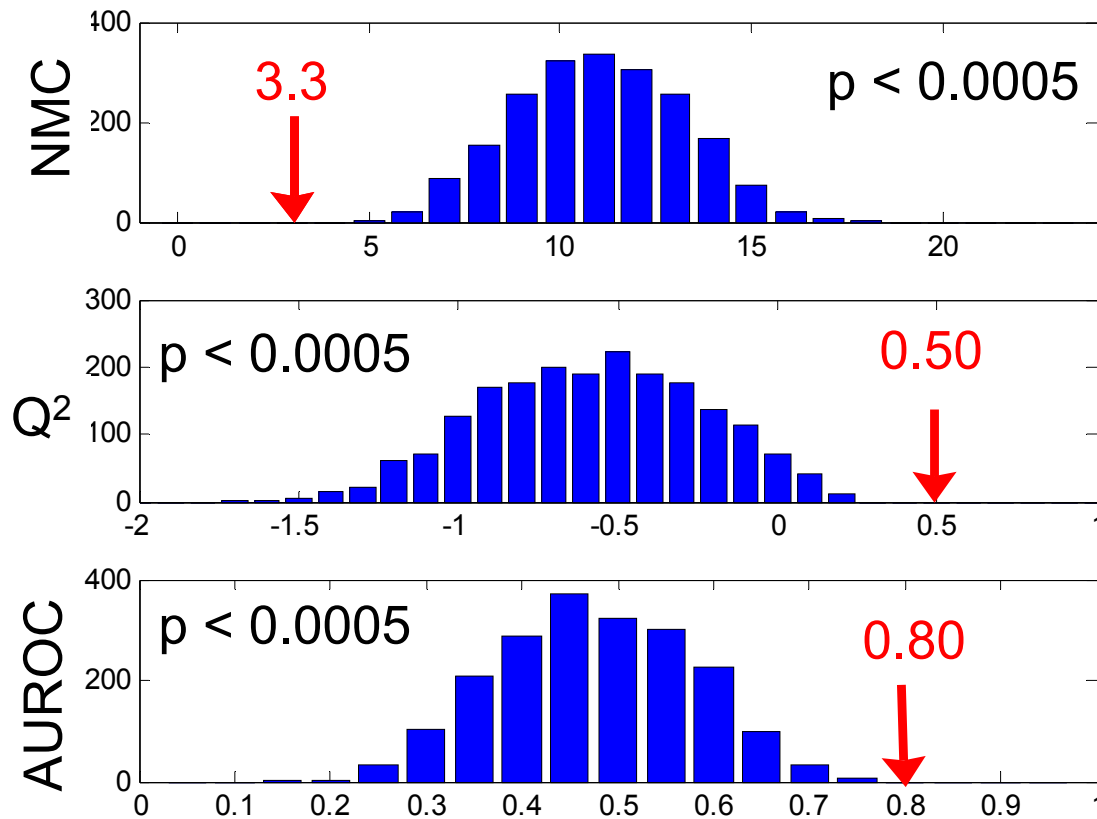


- 23 samples
- No group difference (due to permutation), so 11.5 misclassifications expected.
- FIT: **BAD**
- 1CV: **Much better** but too optimistic
- 2CV: **Perfect**

# Evaluation of gender difference



- **Original classification / 2000 permutations**
  - 20 times 2CV cross validation, NMC,  $Q^2$ , AUROC



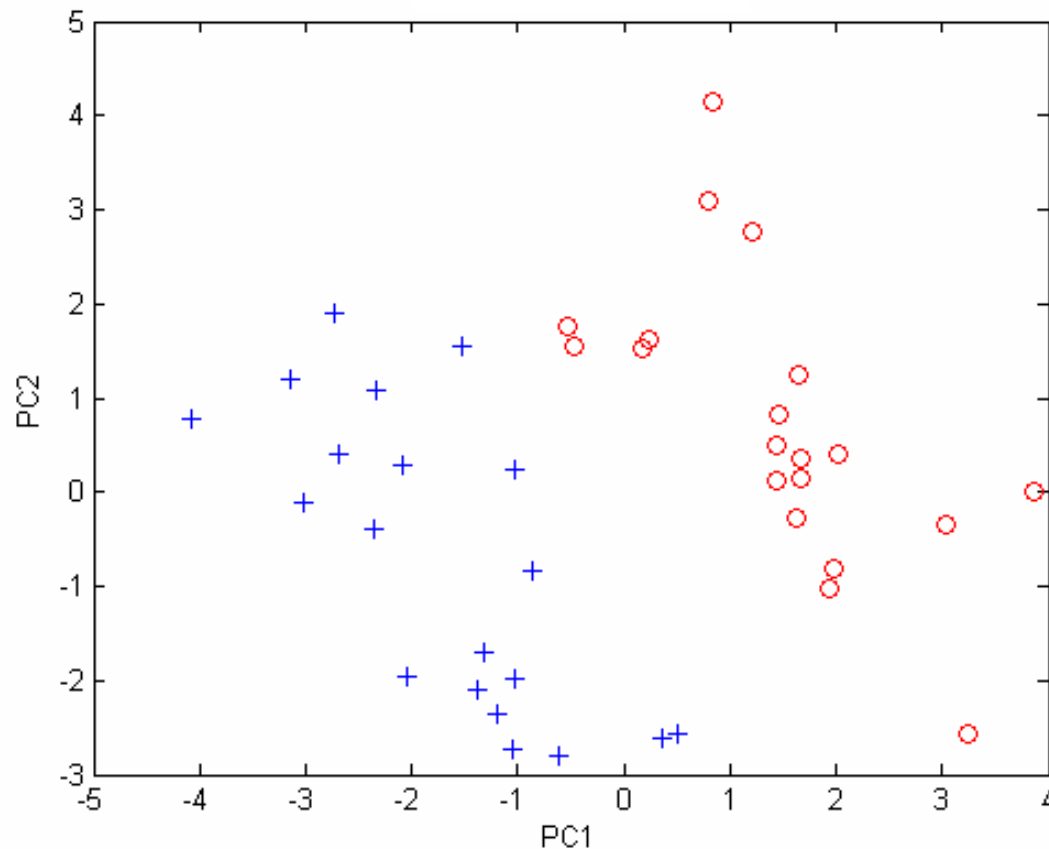
- **Permutations**
  - $Q^2$  on average -0.5
  - 6% of permutations have  $Q^2 > 0$
  - AUROC on average 0.5
- **Original**
  - 3.3 misclassifications
  - $Q^2 = 0.5$
  - AUROC = 0.80
  - All statistics  $p < 0.0005$

# Visualisation

## Score plots / Class predictions



- Do not conclude class separation from score plots.



Such results can also be obtained from random data without class difference

For studying within class variation use PCA score plots instead

# Visualisation

## Score plots / Class predictions



- Show class predictions of multiple (2CV) instead.

### Show Class predictions

Use multiple model (2CV)  
to predict class.

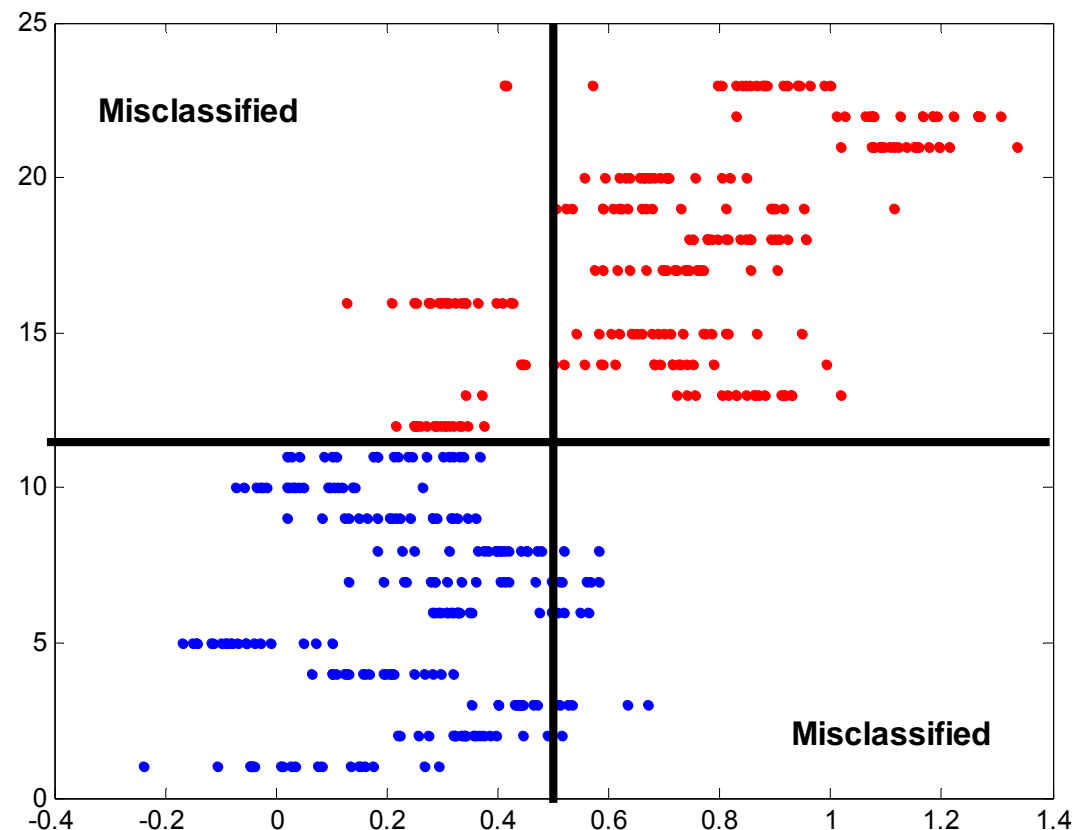
### Disadvantage:

All models based on less  
samples.

### Advantage:

Majority voting (bagging).

Idea of spread in class  
prediction.



# Validation of biomarker selection using Rank Products<sup>1</sup> (RP)



- RP introduced to detect differentially expressed genes
- RP: Nonparametric method **does not assume** normal or symmetric distribution.
- Robust due to ranking approach
- Idea: Biomarker must consistently be important in multiple models of same system.

**Variable importance**

M1	1.2	0.9	0.6	1.4	4.2	3.3	-2.1	-0.7	-2.2
M2	1.0	1.1	0.9	1.5	3.6	3.7	-3.3	0.1	-2.4
M3	0.8	1.5	1.0	1.3	1.1	3.3	-2.1	-0.3	-1.9

**Ranked importance**

6	7	9	5	1	2	4	8	3
7	6	8	5	2	1	3	9	4
8	4	7	5	6	1	2	9	3

**RP**

336	168	504	125	12	2	24	648	36
-----	-----	-----	-----	----	---	----	-----	----

Geometric mean **RP<sup>1/3</sup>**

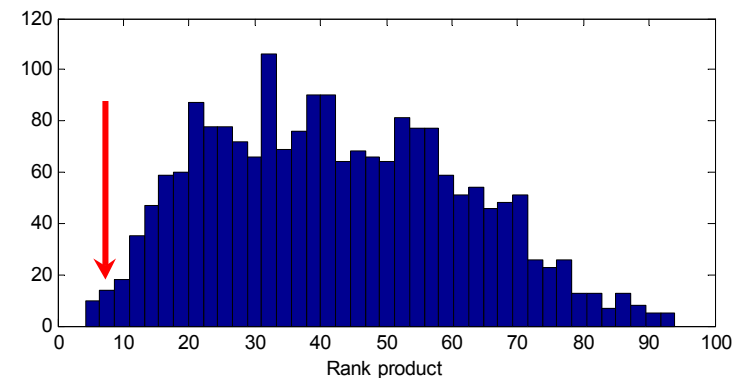
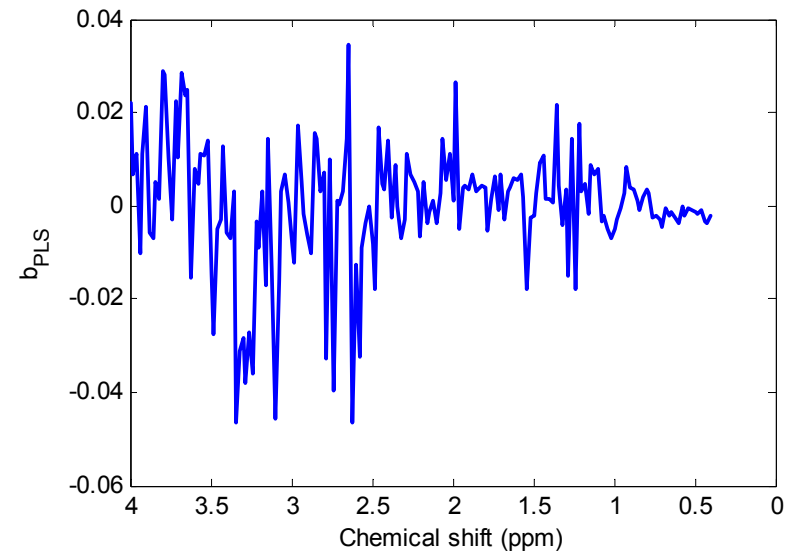
6.9	5.5	8.0	5	2.3	1.3	2.9	8.7	3.3
-----	-----	-----	---	-----	-----	-----	-----	-----

<sup>1</sup>Breitling, FEBS lett, 2004, 537, 83-92

# Rank products<sup>1</sup>, cross validation and permutation to select biomarkers



- During 2CV approach multiple models are developed
- Calculate RP of each variable.
- Calculate RP in same way for 2CV models of permuted data.
- Compare RP distributions of permuted data and original RP.



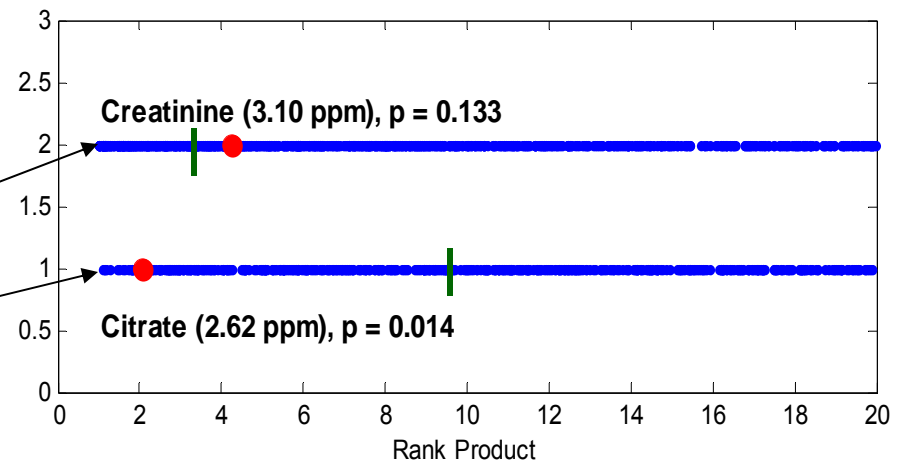
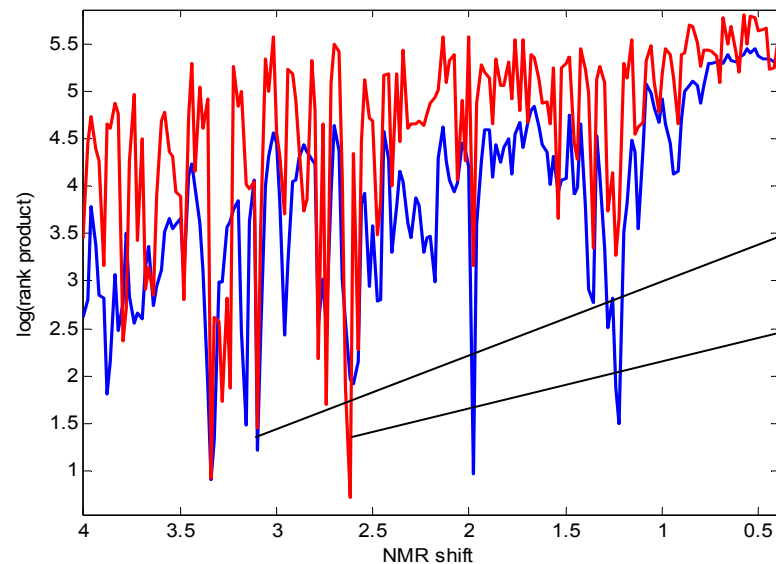
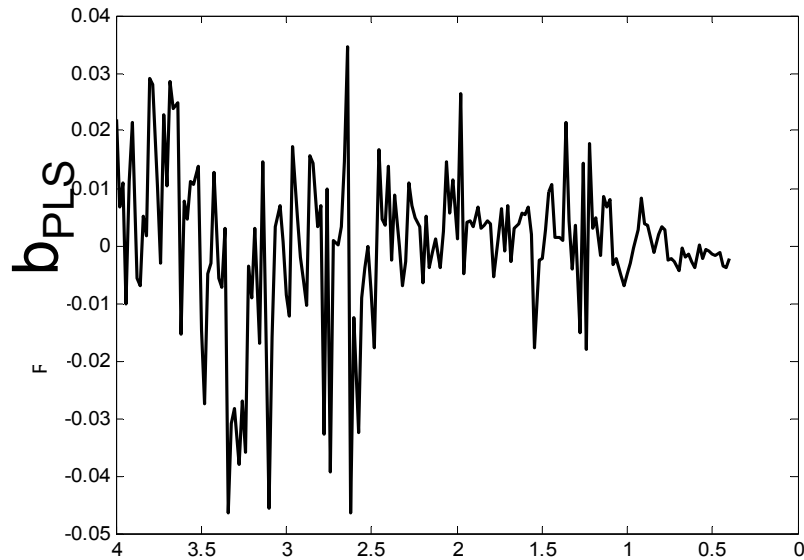
<sup>1</sup>Breitling, FEBS lett, 2004, 537, 83-92



# Biomarkers by rank products and permutations



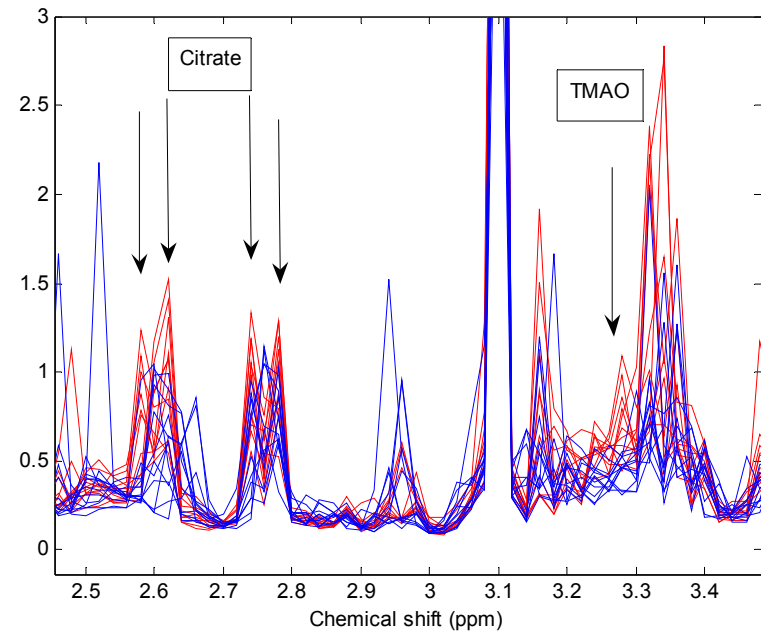
- 2000 permutations, 20 2CV models.
- Calculate average RP
- **Mean rank crossvalidation** vs **10% quantile of permutations**
- Large peaks are not always important for classification.



# Models with / without biomarkers



- Based on 10% quantiles of rank product distribution, 30 peaks were found significant.
- Among these peaks were citrate(4), TMAO, which are also found in other gender studies.
- New models

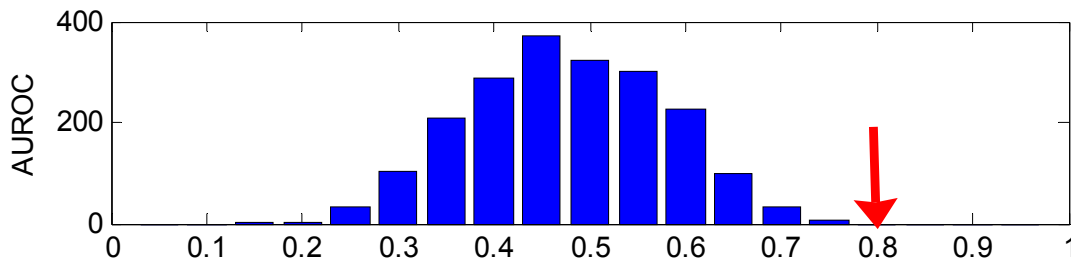
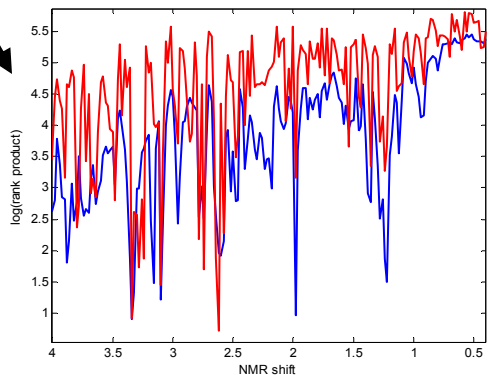
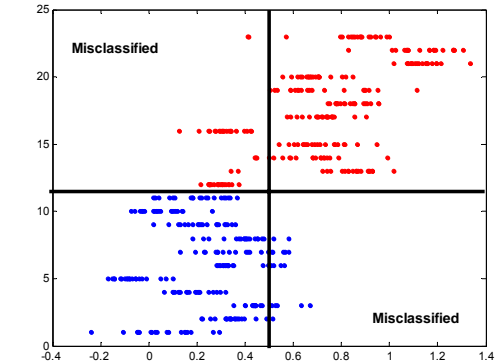
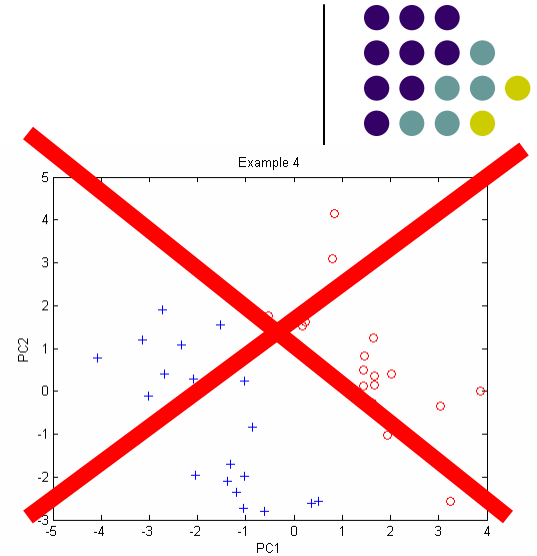


**Women / Men**

	Original model	Model with biomarkers	Model without biomarkers
Misclass.	3.35	1.05	6.80
Q <sup>2</sup>	0.50	0.77	0.18
AUROC	0.80	0.95	0.67

# Recommendations

- Use random data as first step to test new data analysis methods.
- Do not conclude class separation based on score plots.
- Show class predictions based on multiple 2CV models.
- Combine 2CV, rank products and permutation testing to find potential biomarkers.
- Use permutation testing to validate the whole discrimination procedure.



# Multilevel analysis of a nutritional randomized placebo-controlled cross-over study of grape polyphenols

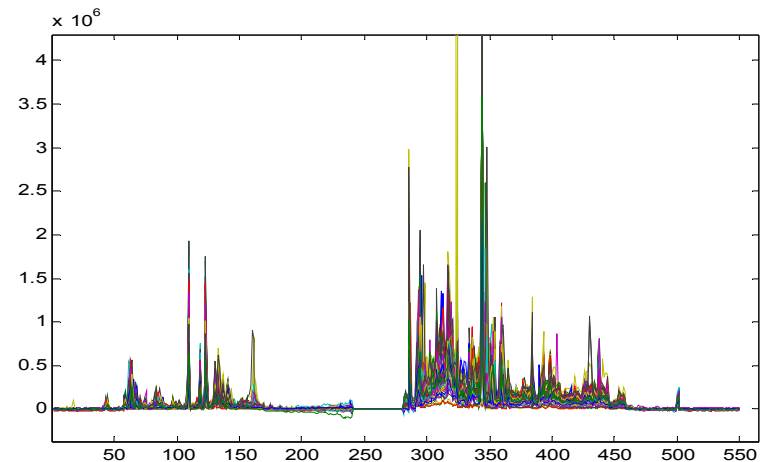


Unilever

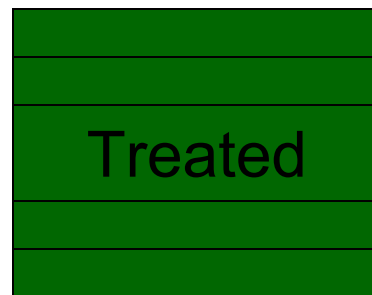
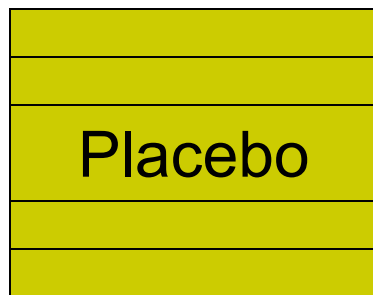
## Study effect of grape polyphenols on blood pressure

29 individuals,

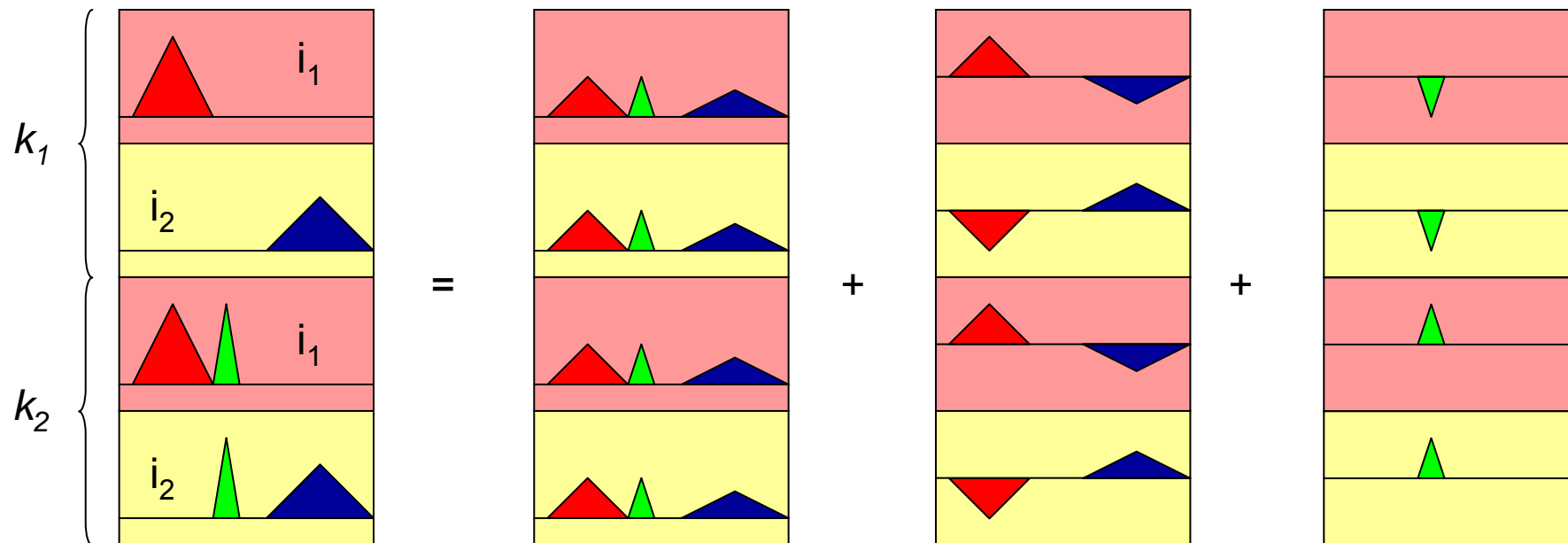
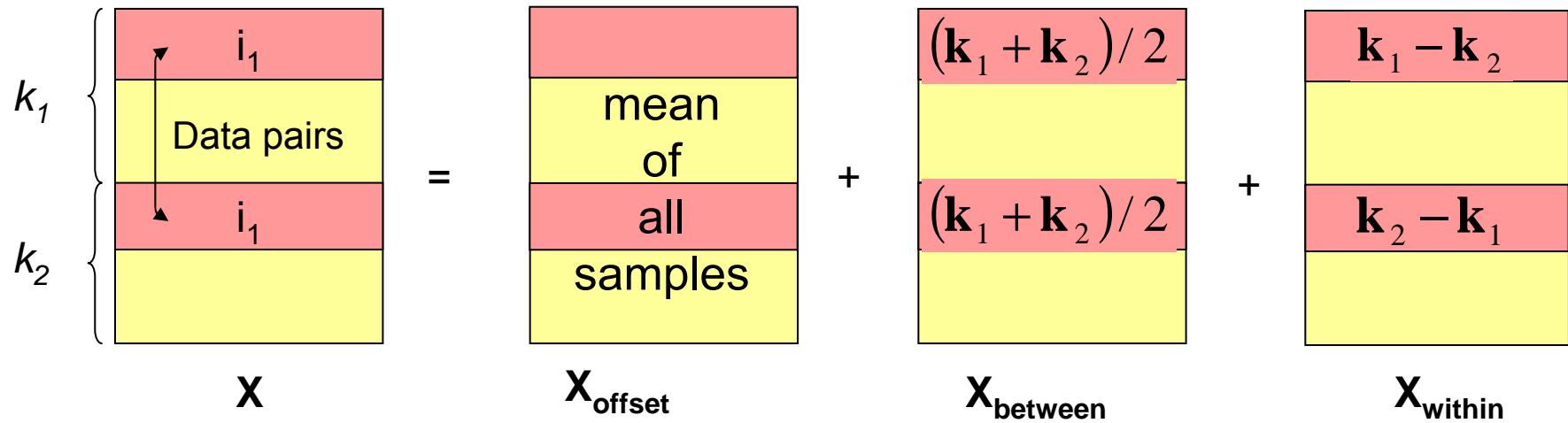
- Placebo and grape polyphenol intervention (randomized)
- The urinary  $^1\text{H}$  NMR profiles
- Autoscaling was used
- **NB!! Data is paired, USE IT !!**



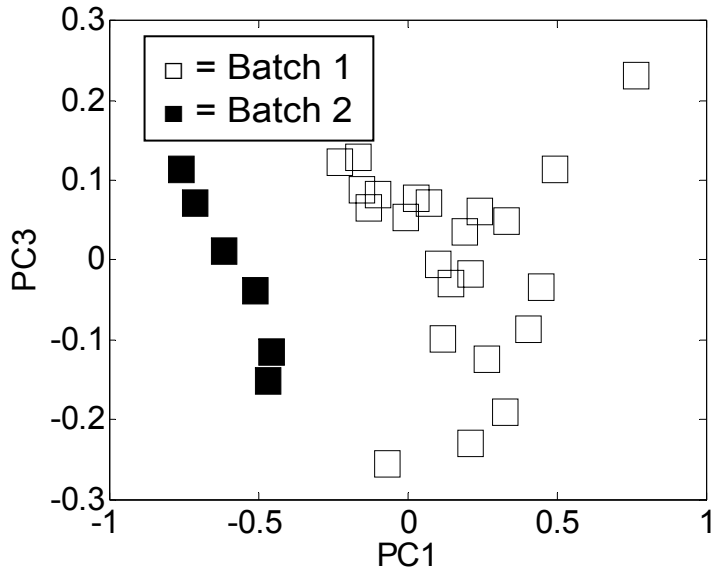
29



# Multilevel analysis: ANOVA based partitioning of variation

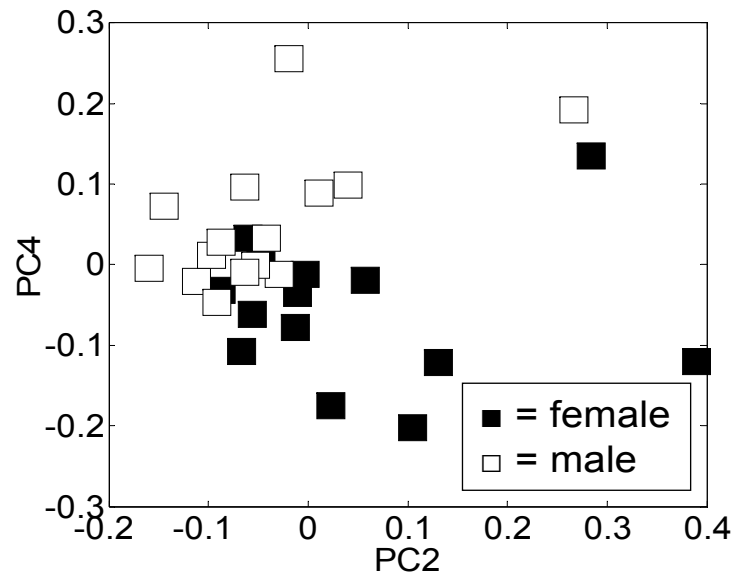
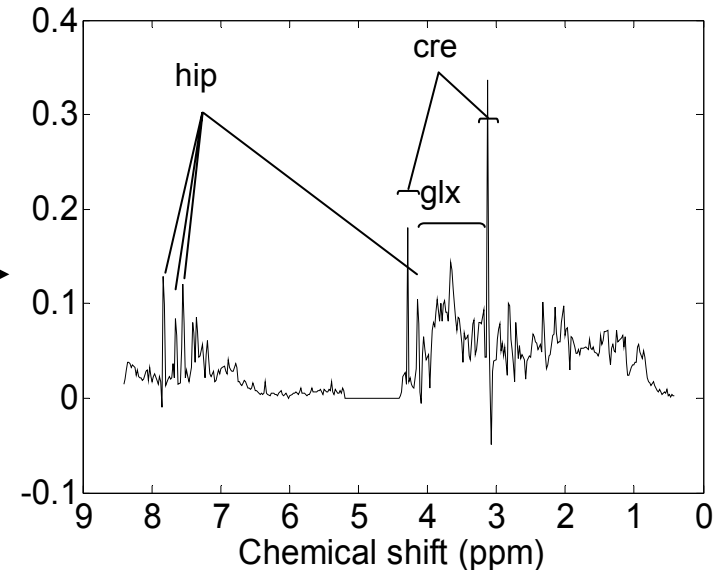


# PCA of between subject variation (79%).



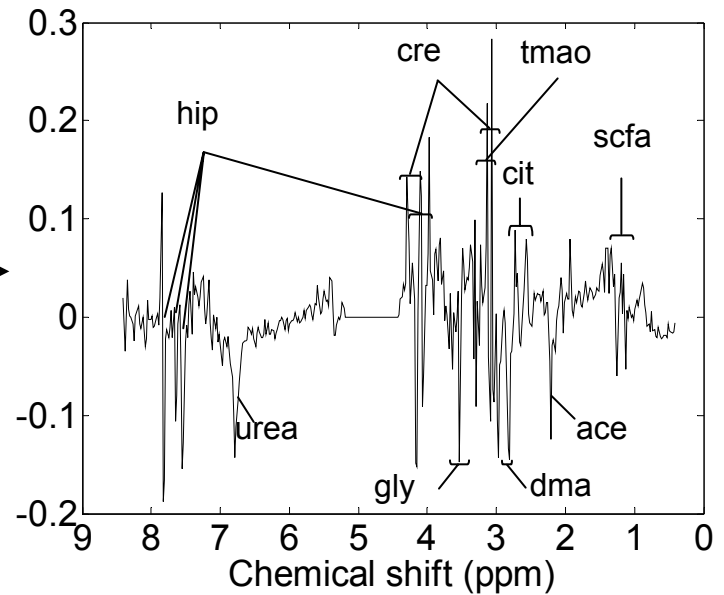
← **Experimental Variation** →

$R^2_{(PC1)}: 58\%$



← **Biological Variation** →

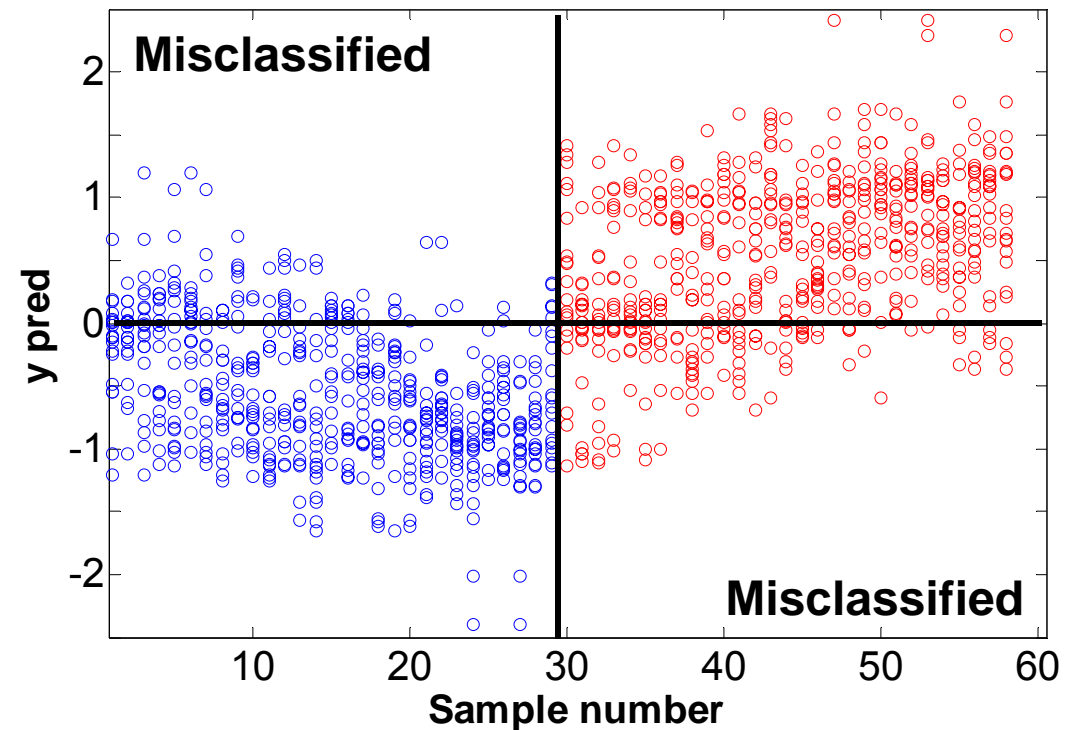
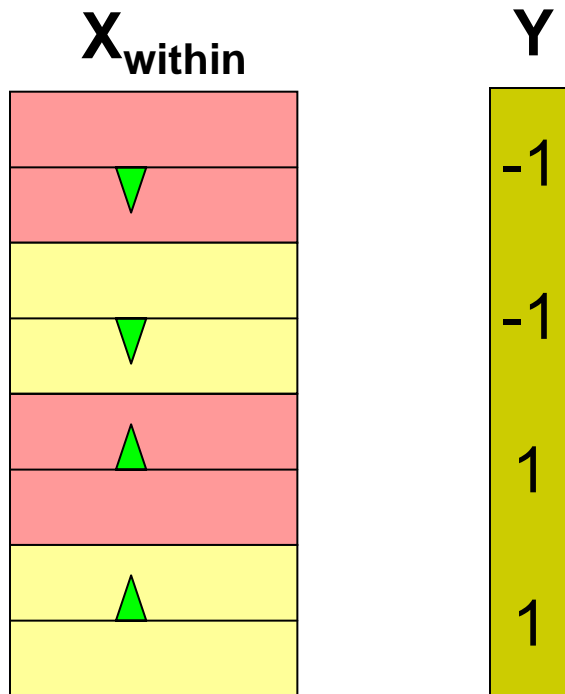
$R^2_{(PC4)}: 4\%$

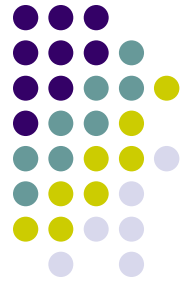


# Multilevel PLSDA of within subject variation (21%)



## 2CV class predictions





# Validation of polyphenol effect by permutation testing

Validate the PLSDA obtained treatment effect by comparing with PLSDA obtained effects when class labels were **randomized**.

Normal paired data

P1 control	P1 intervention
P2 control	P2 intervention
P3 control	P3 intervention
P4 control	P4 intervention
P5 control	P5 intervention
P6 control	P6 intervention
P7 control	P7 intervention
P8 control	P8 intervention
P9 control	P9 intervention
P10 control	P10 intervention

**Correct  
treatment order**

Permuted data

P1 control	P1 intervention
P2 control	P2 intervention
P3 intervention	P3 control
P4 intervention	P4 control
P5 control	P5 intervention
P6 intervention	P6 control
P7 control	P7 intervention
P8 intervention	P8 control
P9 control	P9 intervention
P10 control	P10 intervention

**Randomized  
treatment order**

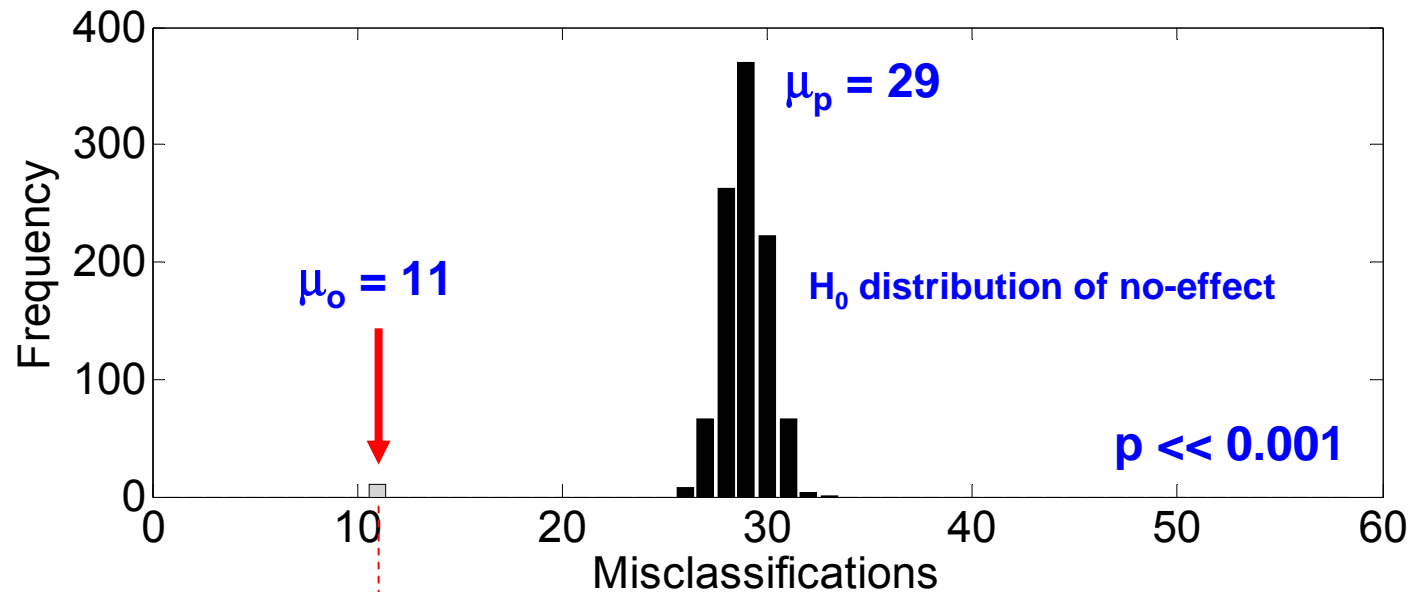
**Class predictions  
are average  
based on 20  
double cross  
validation rounds**



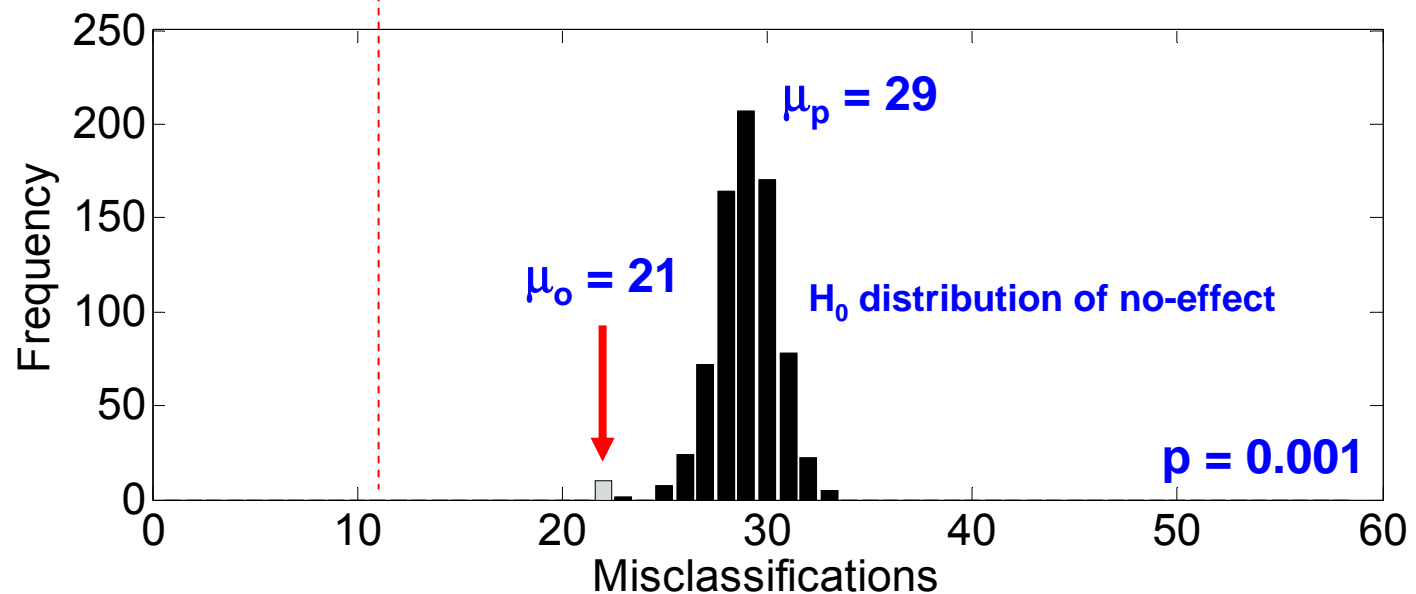
# Validation results (1000 permutations)



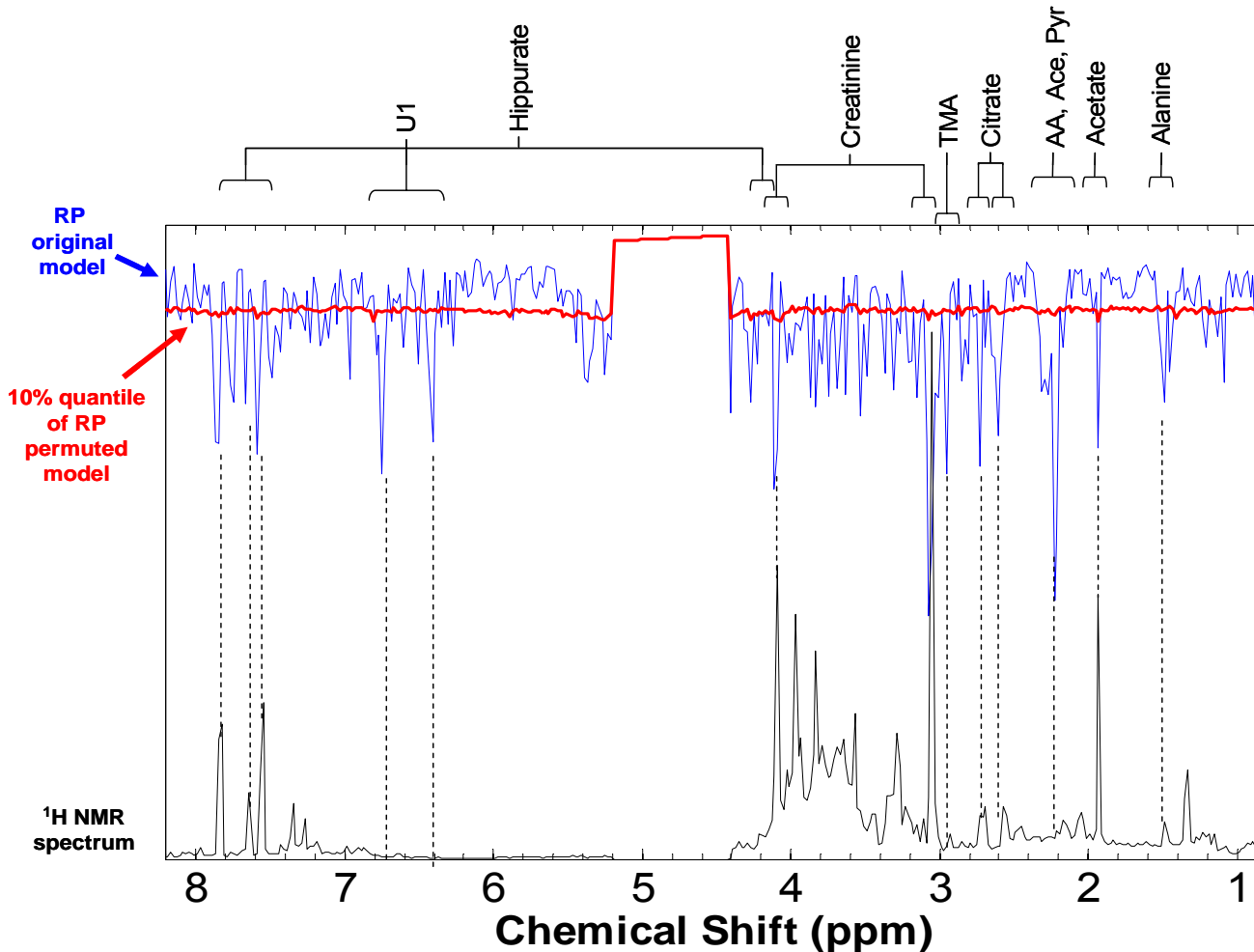
**Multilevel  
PLS-DA**



**PLS-DA  
Without  
multilevel  
variance  
splitting**



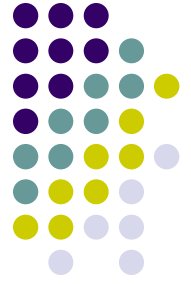
# Biomarker selection (Rank Products)



Many peaks of within individual variation were found significant.

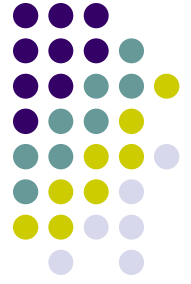
Hippurate:  
degradation product  
of polyphenols

# Conclusions



- Metabolomics data is often structured due to experimental design underlying the study.
- Take structure of data into account to improve power.
- Remove between individual variation to focus on treatment effect.
- Use 2CV, permutation testing and Rank Products to validate metabolic differences and important potential biomarkers.

# Acknowledgements



- Universiteit van Amsterdam:
  - Age Smilde, Huub Hoefsloot, Suzanne Smit, Daniel Vis



- Unilever Research Vlaardingen (metabolomics):
  - Ewoud van Velzen, John van Duynhoven, Ferdi van Dorsten