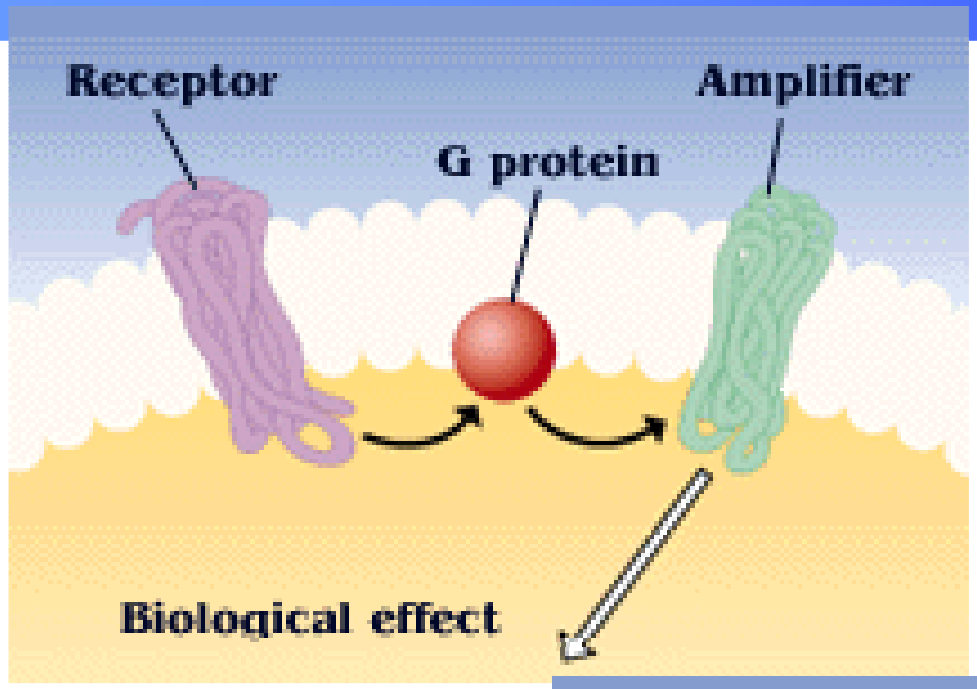
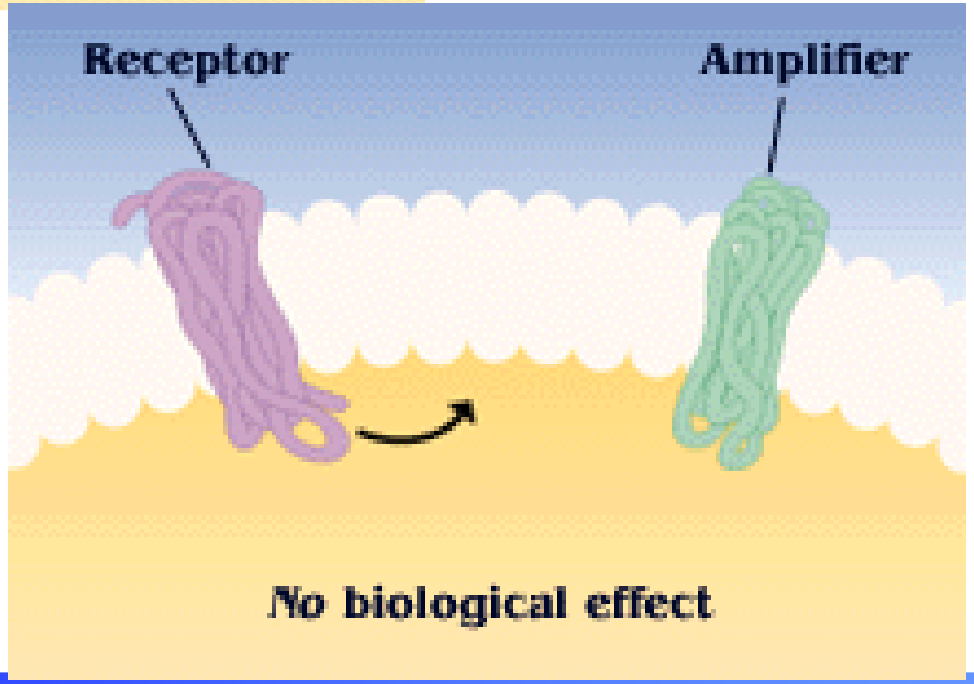
Bioinformatics, Metabolomics and personalized medicine

Torbjörn Lundstedt

KI 080609



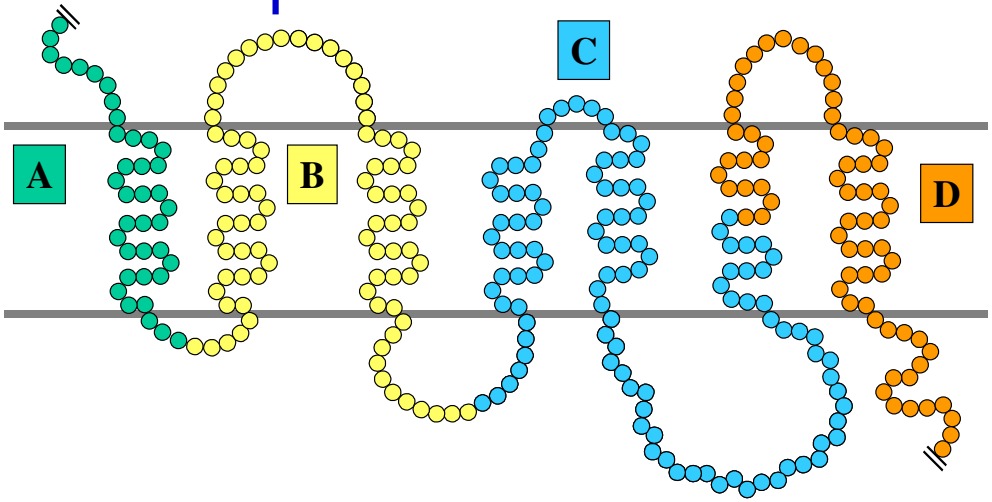
Multivariate analysis of G-protein coupled receptors



Figures from http://www.nobel.se

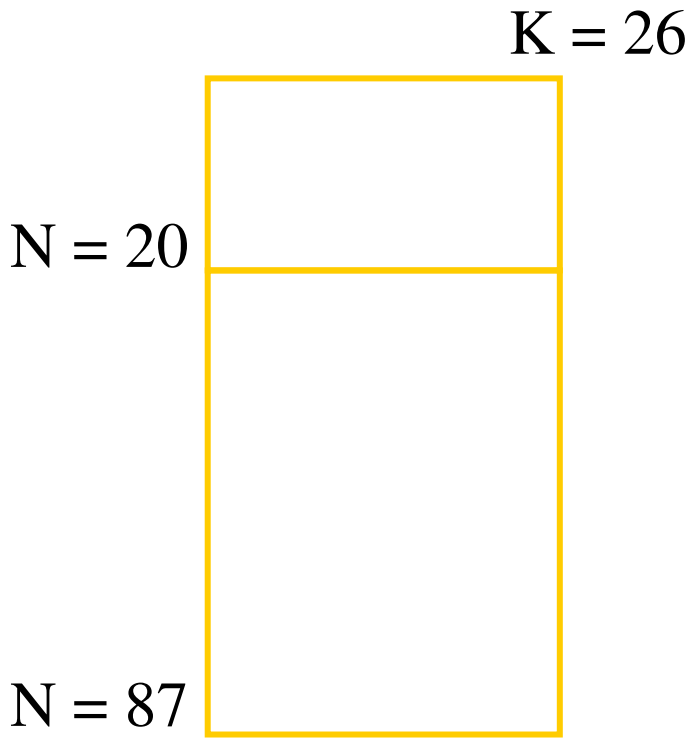
Bioinformatic examples

- G-proteincoupled receptors 2D structure GPCR
- 7 trans-membrane regions 7TM
- Amino acid sequences

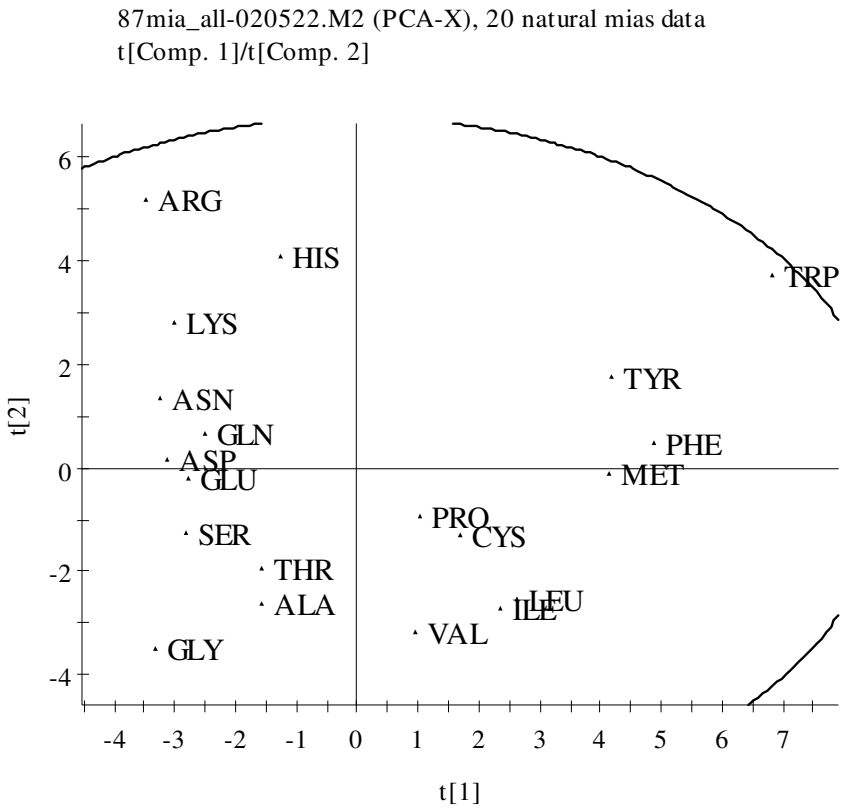


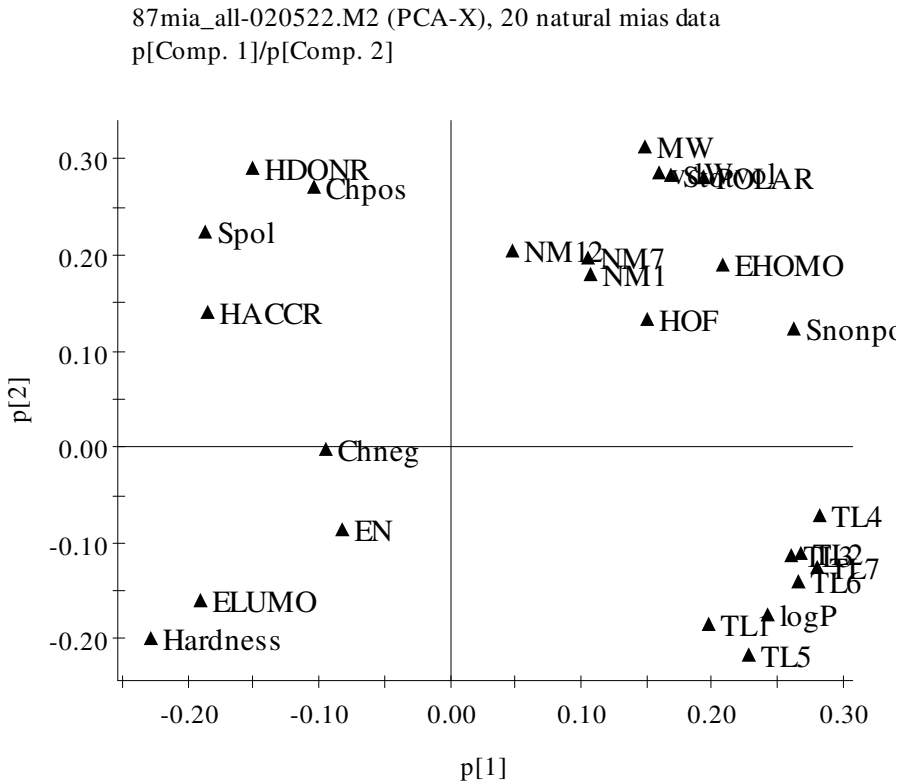
20 natural amino acids

- 20 natural amino acids
- Characterised with both experimental and calculated variables
- Observations = 20
- Variables = 26 (in this example)
- Have been expanded to 87 amino acids (Maria Sandberg et al.)



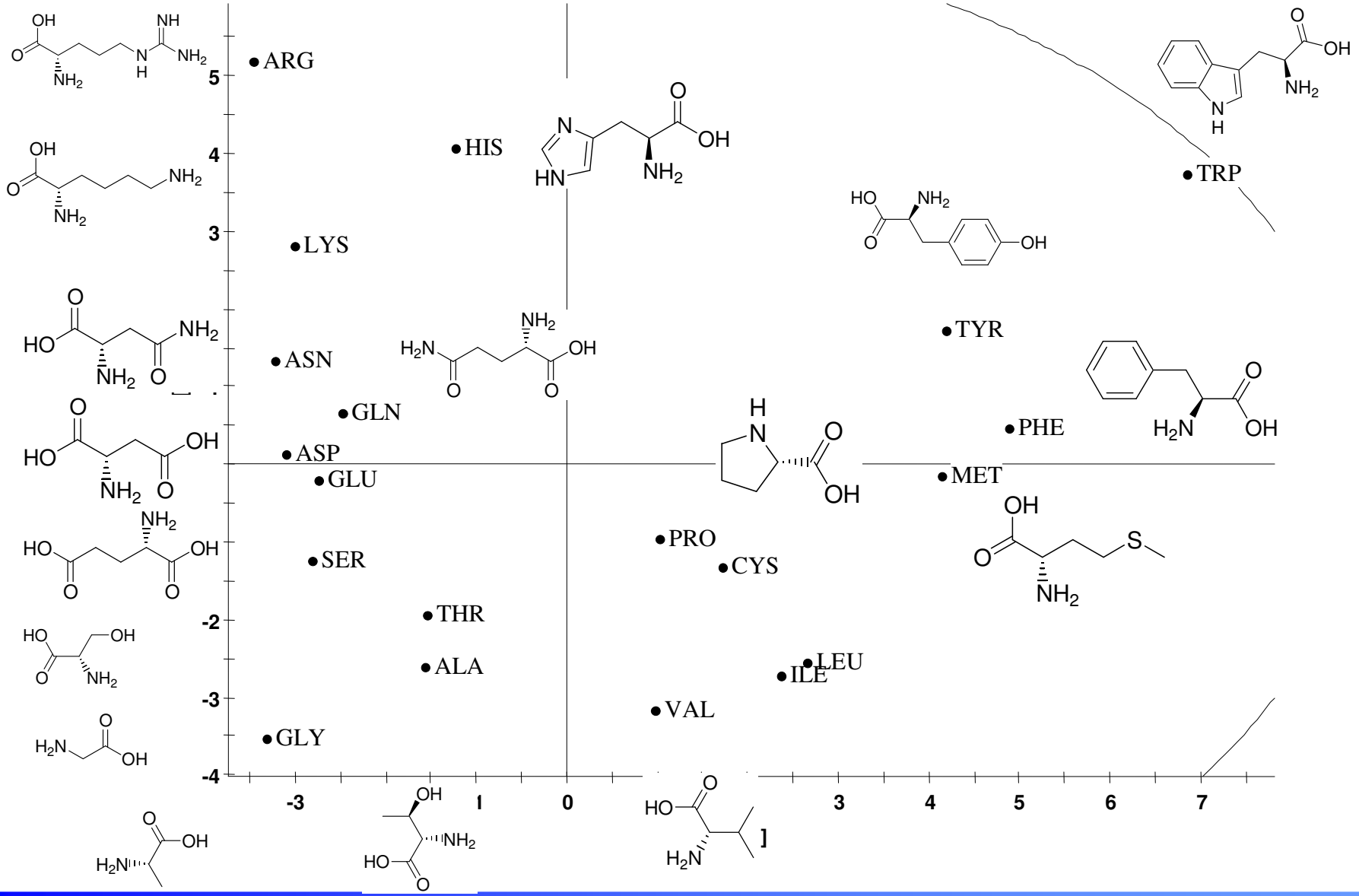
20 natural amino acids





- Two first principal components t_1 vs. t_2 ➔ compare with p_1 and p_2 for comparision

20 natural amino acids



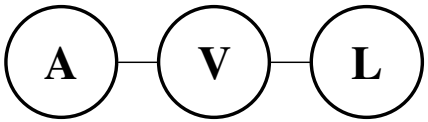
Principal properties
— natural amino acids —

No.	Name	Name	One letter code	z_1	z_2	z_3
1	Alanine	ALA	A	0.07	-1.73	0.09
2	Valine	VAL	V	-2.69	-2.53	-1.29
3	Leucine	LEU	L	-4.19	-1.03	-0.98
4	Isoleucine	ILE	I	-4.44	-1.68	-1.03
5	Proline	PRO	P	-1.22	0.88	2.23
6	Phenyalanine	PHE	F	-4.92	1.30	0.45
7	Tryptophan	TRP	W	-4.75	3.65	0.85
8	Methionine	MET	M	-2.49	-0.27	-0.41
9	Lysine	LYS	K	2.84	1.41	-3.14
10	Arginine	ARG	R	2.88	2.52	-3.44
11	Histidine	HIS	H	2.41	1.74	1.11
12	Glycine	GLY	G	2.23	-5.36	0.30
13	Serine	SER	S	1.96	-1.63	0.57
14	Threonine	THR	T	0.92	-2.09	-1.40
15	Cysteine	CYS	C	0.71	-0.97	4.13
16	Tyrosine	TYR	Y	-1.39	2.32	0.01
17	Aspargine	ASN	N	3.22	1.45	0.84
18	Glutamine	GLN	Q	2.18	0.53	-1.14
19	Aspartic acid	ASP	D	3.64	1.13	2.36
20	Glutamic acid	GLU	E	3.08	0.39	-0.07

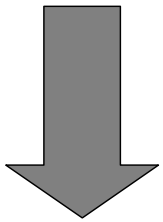
Hellberg *et al.*, 1987 (20 aa´s)
Jonsson *et al.* 1989 (55 aa´s)
Sandberg *et al.* 1998 (87 aa´s)

- Amino acids characterised with experimental and calculated variables
- PCA generates three descriptors for each amino acid
  z_1 high lipofilic, low hydrofobic
  z_2 high large, low small
  z_3 elektronic properties

Principal properties
— natural amino acids —



Principal Property translation



(z-scale)

A			V			L		
z_1	z_2	z_3	z_1	z_2	z_3	z_1	z_2	z_3
0.07	-1.73	0.09	-2.69	-2.53	-1.29	-4.19	-1.03	-0.98

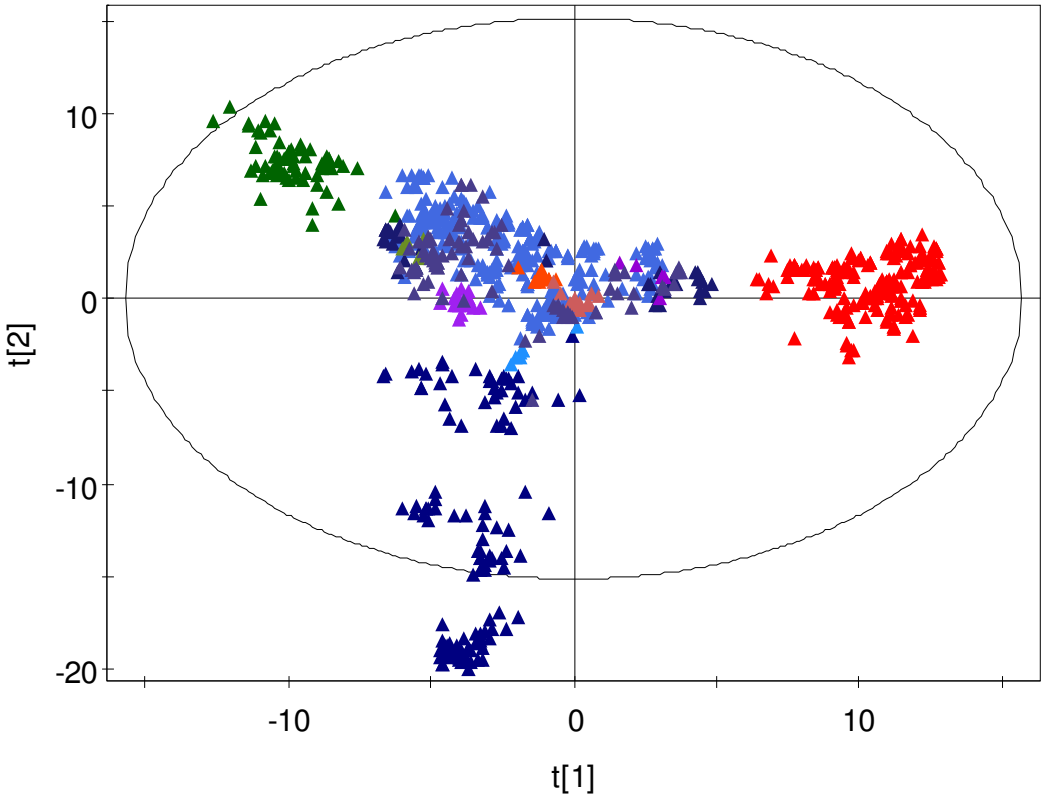
Used to characterise amino acid sequences!

7TM GPRCs analysis, t1/t2

675 variables

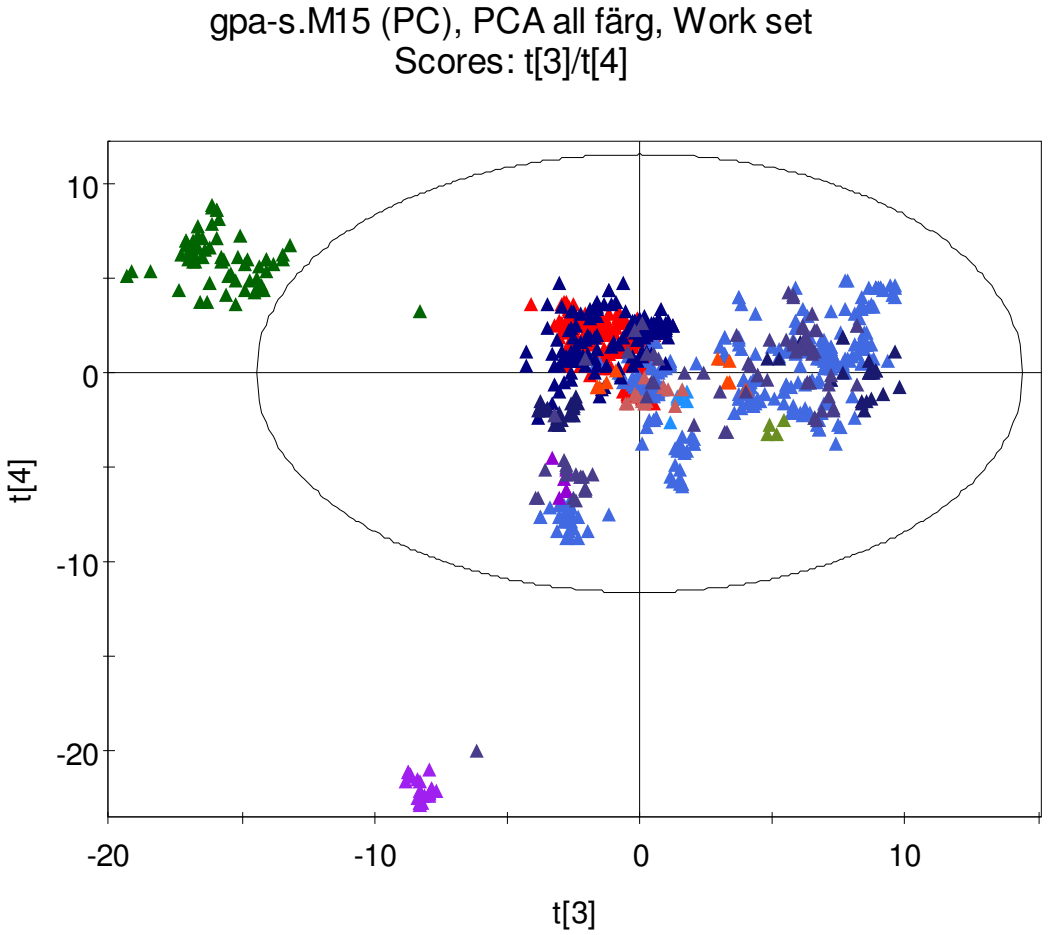
(135×5)

X



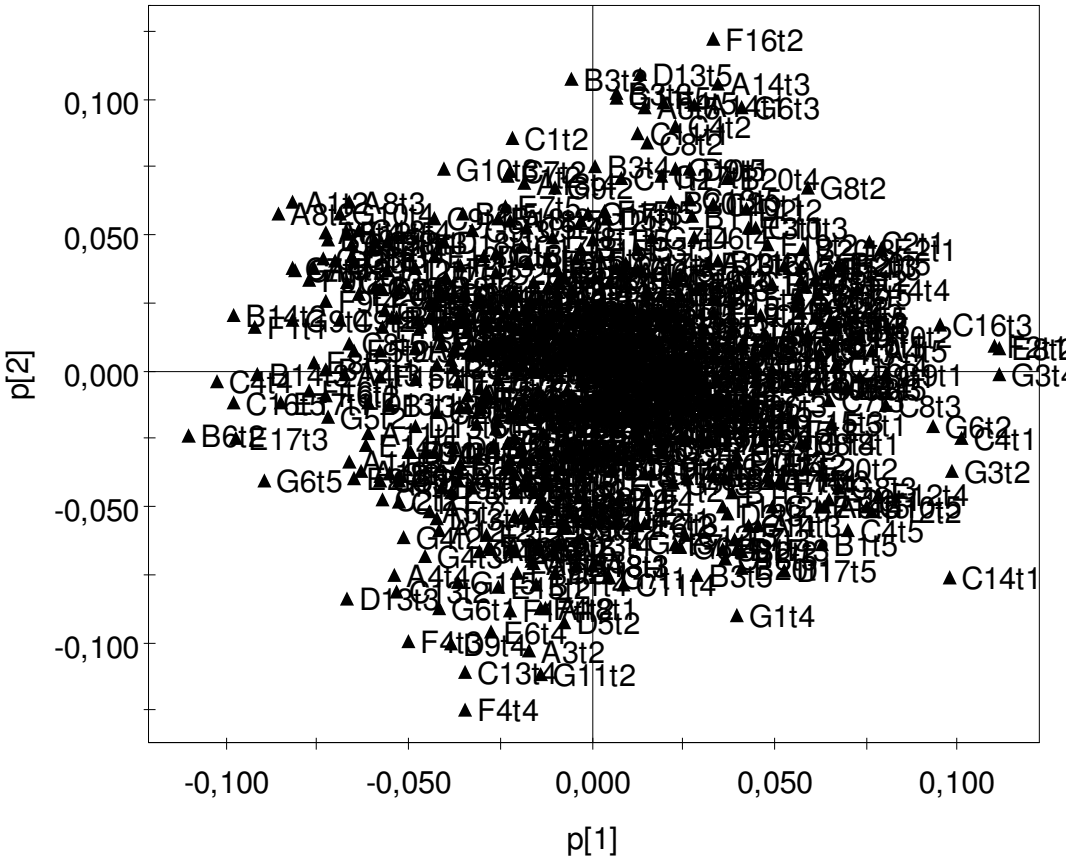
897 sequences

7TM GPRCs analysis, t3/t4



Interpretation – difficulties

gpa-s.M15 (PC), PCA all färg, Work set
Loadings: p[1]/p[2]



Simca-P 8.0 by Umetrics AB 2001-09-26 08:39

Hierarchical PCA

- Dividing the data in different levels depending on origin
- More than one PCA model to examine one problem
- Simplify interpretation

Characterisation

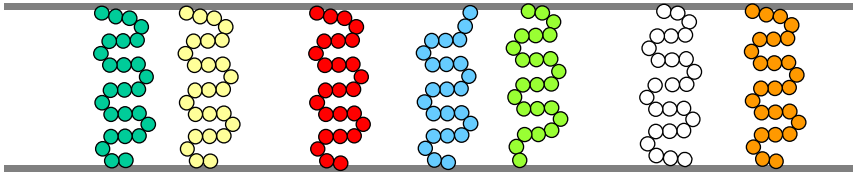
Level 1 – Interpretation

Receptor characterisation

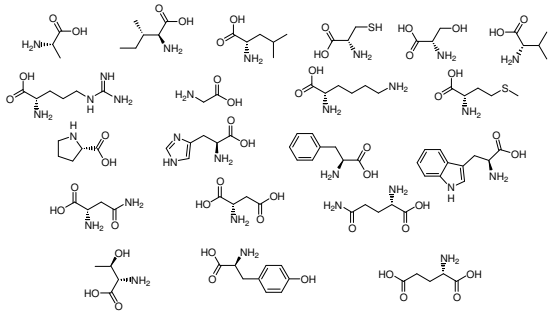
A B C D E F G

Trans-membrane characterisation

$A(T_{aa})$ $B(T_{aa})$ $C(T_{aa})$ $D(T_{aa})$ $E(T_{aa})$ $F(T_{aa})$ $G(T_{aa})$



Amino acid characterisation

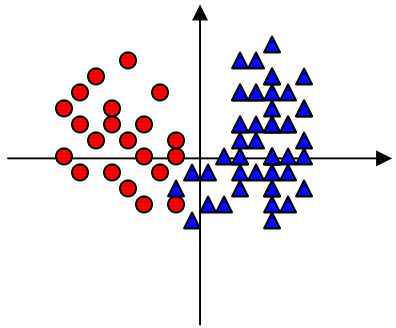


PCA

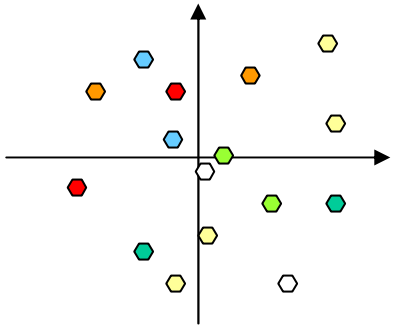
X

One receptor level

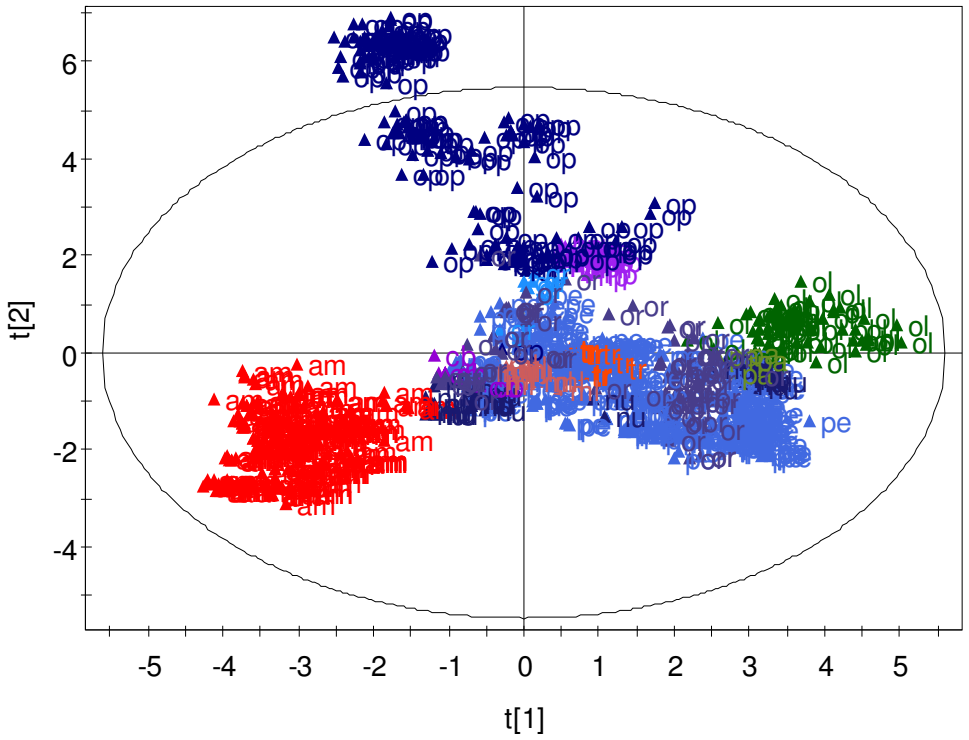
Principal scores

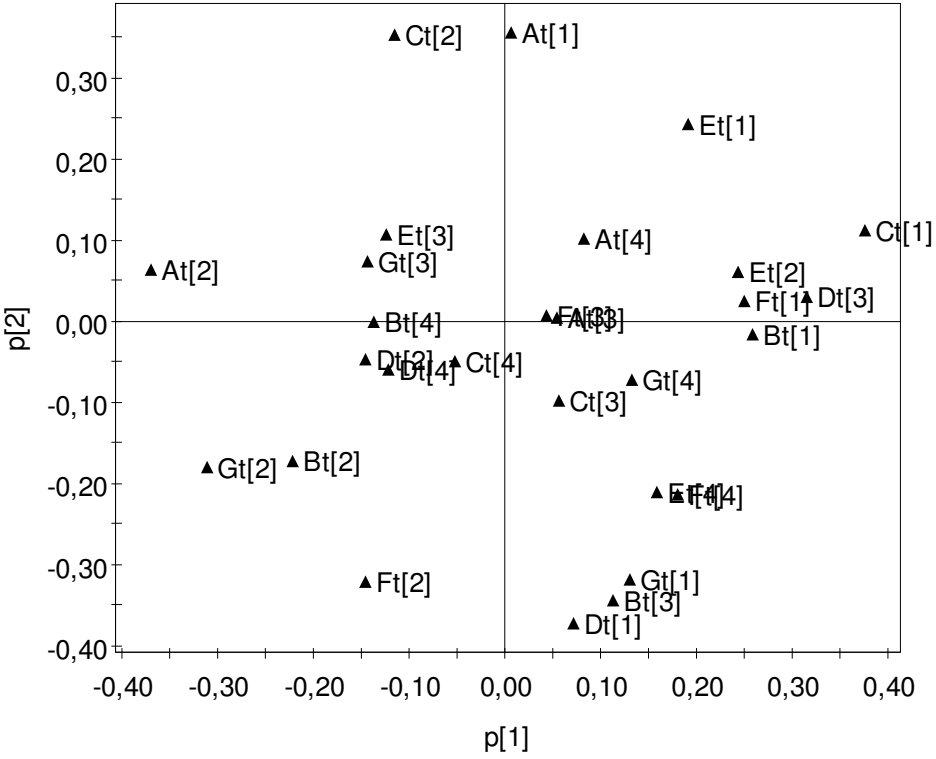


Principal loadings



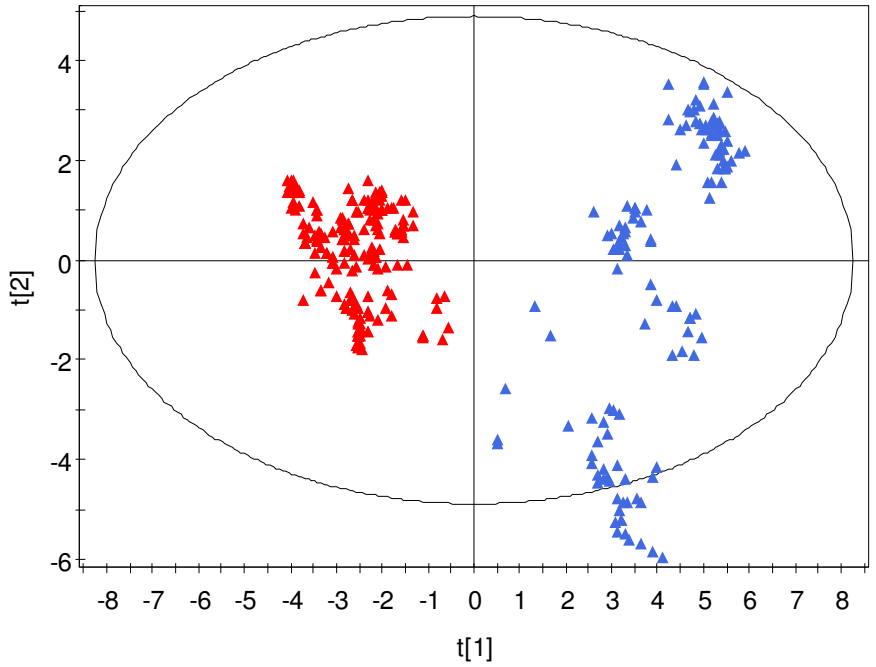
Hierarchical PCA





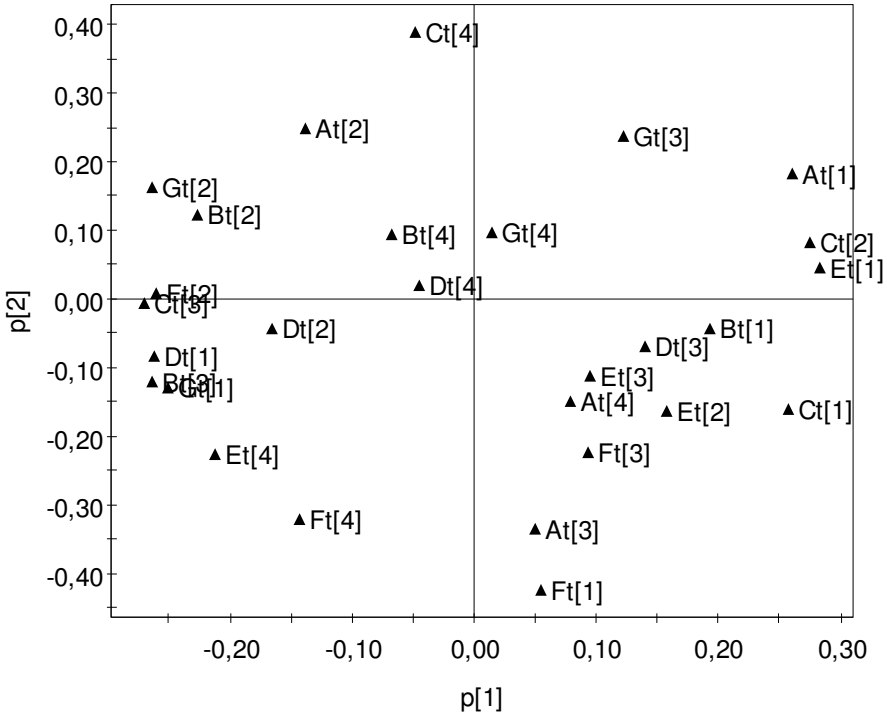
Analysis of sub-groups

s_hierar.M5 (PC), fyra komp., Work set
Scores: t[1]/t[2]



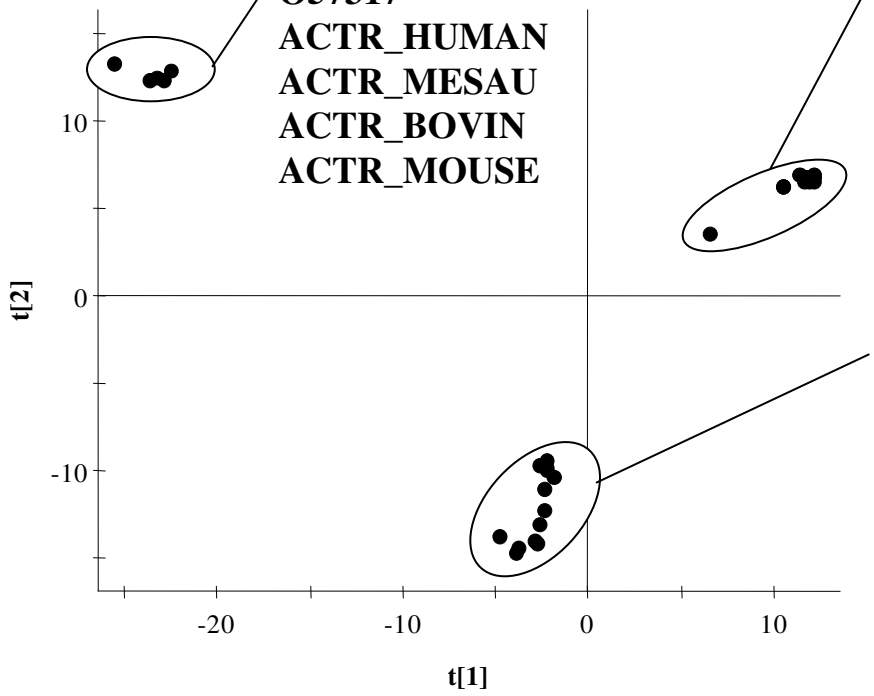
Ellipse: Hotelling T2 (0,05)
Simca-P 8.0 by Umetrics AB 2001-09-10 09:24

s_hierar.M5 (PC), fyra komp., Work set
Loadings: p[1]/p[2]



Classification – GPCR

Analysis of the melanocortin receptors; MC1, MC2, MC3, MC4 and MC5. MC1 and MC2 are quite different compared to MC3-MC5.



MC2

O57317
ACTR_HUMAN
ACTR_MESAU
ACTR_BOVIN
ACTR_MOUSE

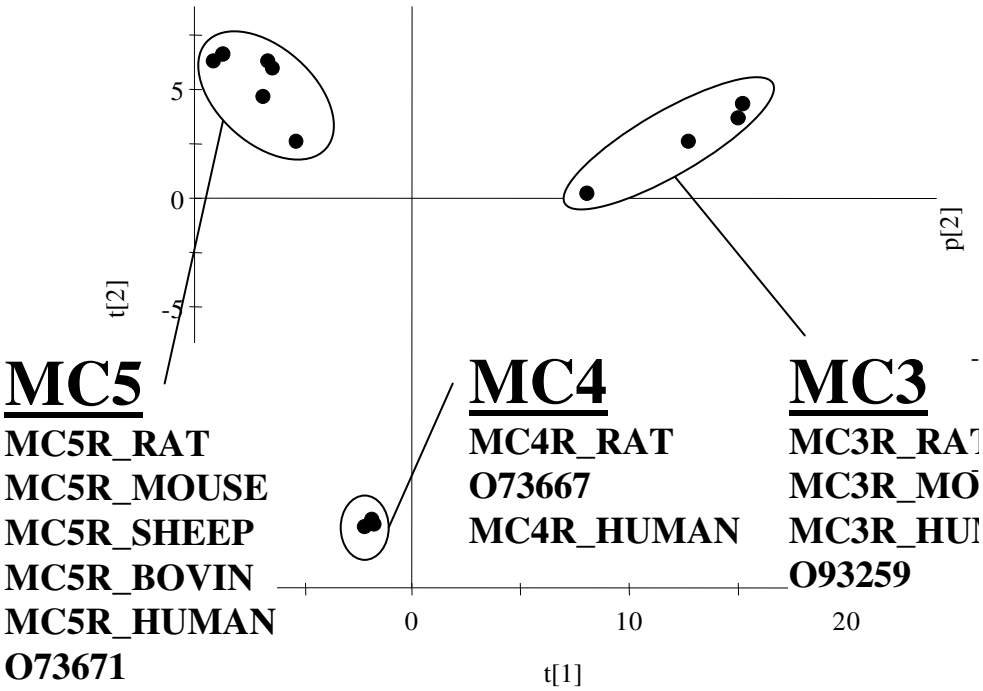
MC1

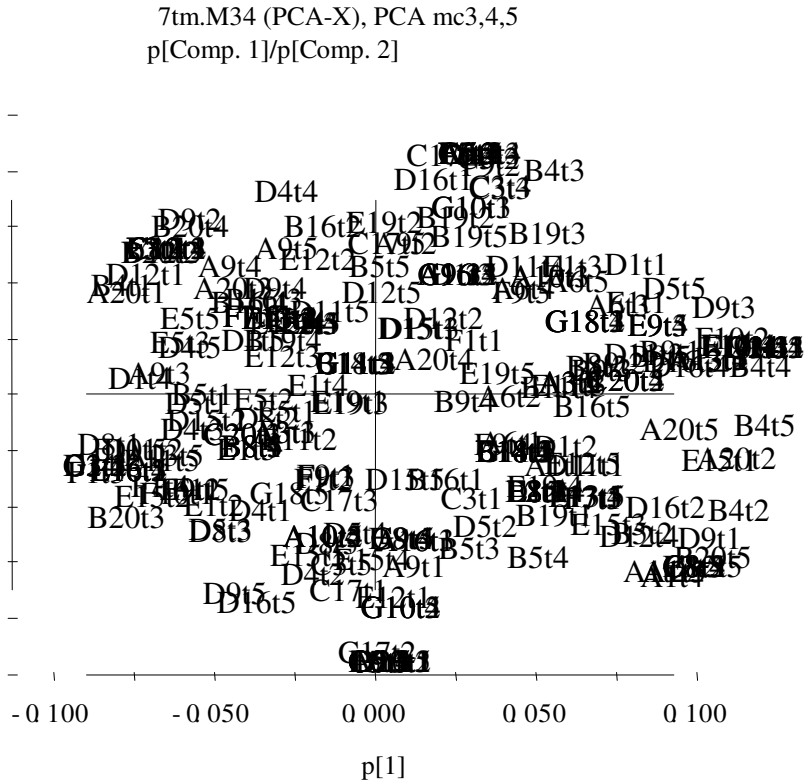
O77616	MSHR_SHEEP
MSHR_CAPCA	MSHR_RANTA
MSHR_CEREL	MSHR_CAPHI
MSHR_ALCAA	MSHR_BOVIN
MSHR_DAMDA	MSHR_OVIMO
MSHR_HUMAN	MSHR_MOUSE
MSHR_VULVU	MSHR_CHICK1

MC3,4,5

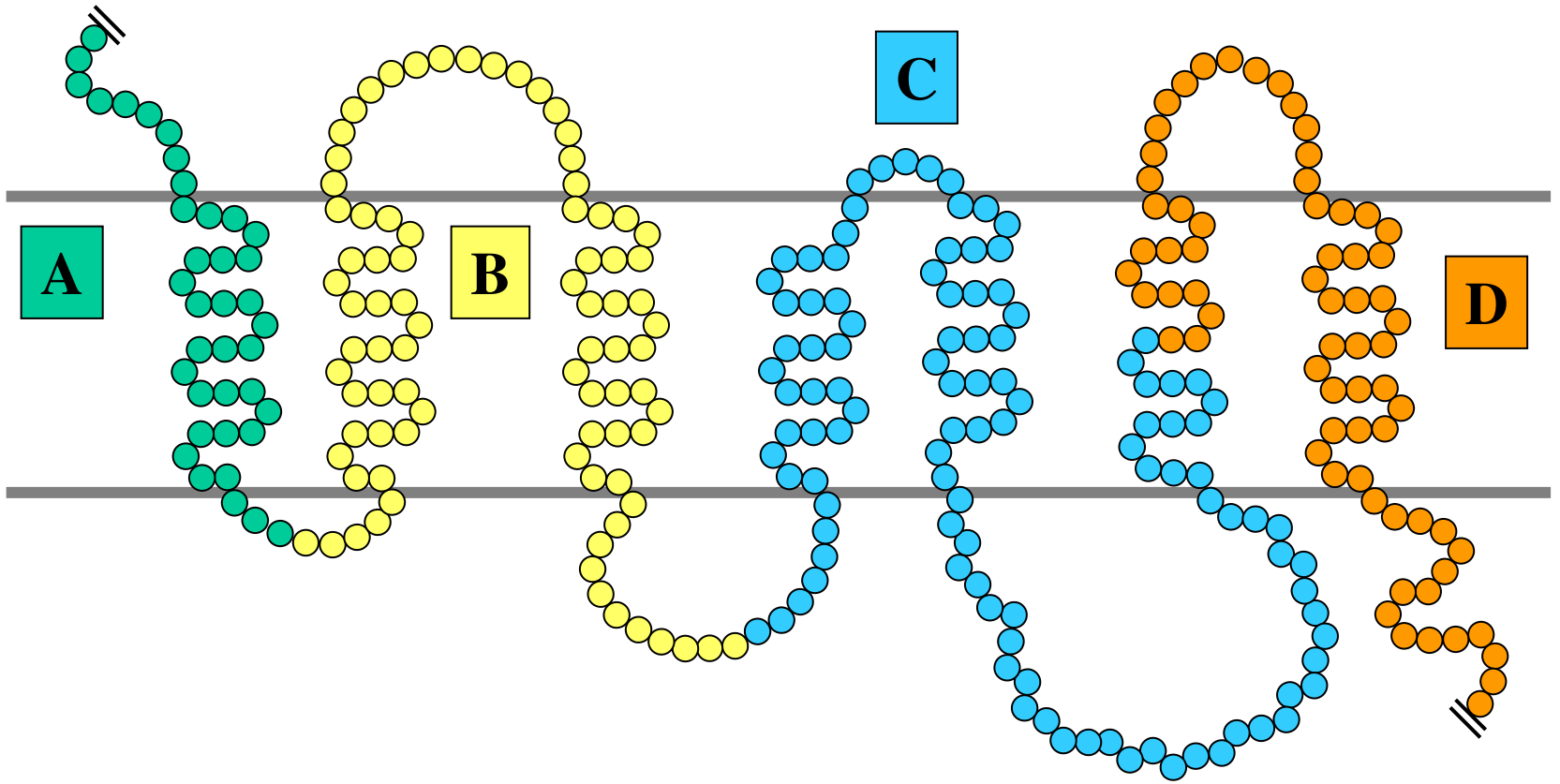
O93259	O73667
MC3R_HUMAN	MC4R_HUMAN
MC4R_RAT	MC3R_MOUSE
MC3R_RAT	O73671
MC5R_BOVIN	MC5R_MOUSE
MC5R_HUMAN	MC5R_SHEEP
MC5R_RAT	

Classification – GPCR





Schematic Overview of MC1 and MC3 Receptor Chimeras



Divided in four parts – characterised with binary coding

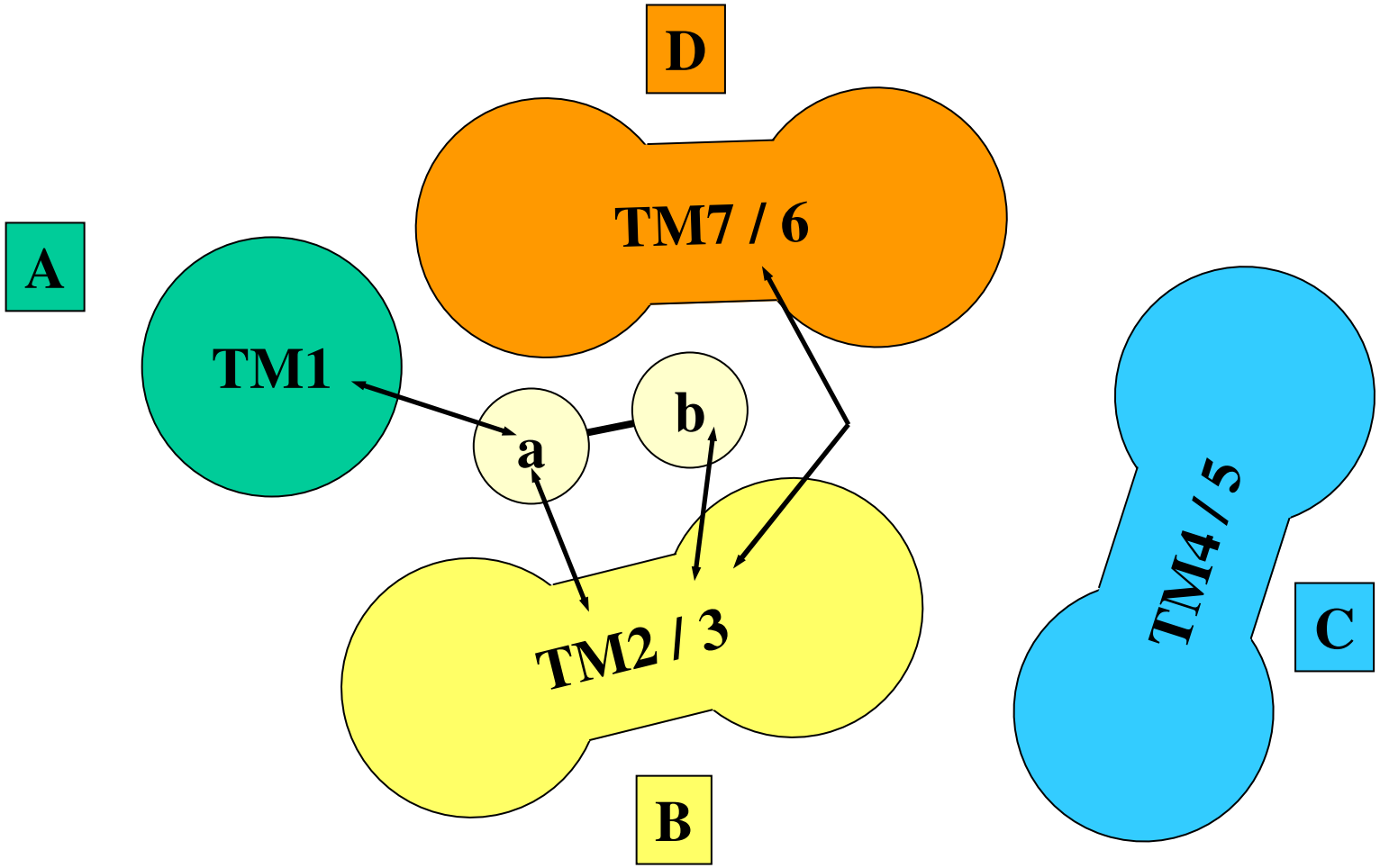
Sequences of Melanocortin Receptor Active Peptides

Position	1	2	3	4	5	6	7	8	9	10	11	12	13
Seq. Part	**a**										**b**		
Peptide													
MSH	**Ser**	**Tyr**	**Ser**	**Met**	**Glu**	His	Phe	Arg	Trp	Gly	**Lys**	**Pro**	**Val**
MS04	**Ser**	**Ser**	**Ile**	**Ile**	**Ser**	His	Phe	Arg	Trp	Gly	**Leu**	**Cys**	**Asp**
MS05	**Ser**	**Ser**	**Ile**	**Ile**	**Ser**	His	Phe	Arg	Trp	Gly	**Lys**	**Pro**	**Val**
MS06	**Ser**	**Tyr**	**Ser**	**Met**	**Glu**	His	Phe	Arg	Trp	Gly	**Leu**	**Cys**	**Asp**

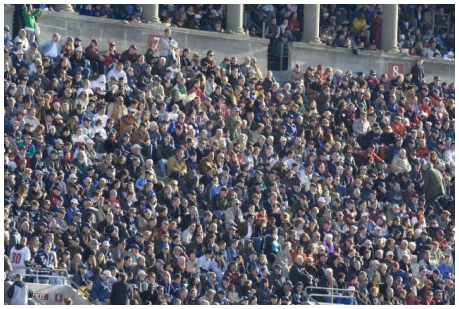
Receptor-Ligand-Response Matrix

Molecules		Descriptors X Y						Response
Z	X	Z				X		Y
Receptor	Peptide	A	B	C	D	a	b	Log(K_i)
$A_1 B_1 C_1 D_1$	$a_1 b_1$	0	0	0	0	0	0	0.21
$A_1 B_1 C_1 D_1$	$a_2 b_2$	0	0	0	0	1	1	0.82
$A_1 B_1 C_1 D_1$	$a_2 b_1$	0	0	0	0	1	0	-0.07
$A_1 B_1 C_1 D_1$	$a_1 b_2$	0	0	0	0	0	1	1.65
$A_1 B_2 C_2 D_2$	$a_1 b_1$	0	1	1	1	0	0	1.37
$A_1 B_2 C_2 D_2$	$a_2 b_2$	0	1	1	1	1	1	3.82
$A_1 B_2 C_2 D_2$	$a_2 b_1$	0	1	1	1	1	0	2.03
$A_1 B_2 C_2 D_2$	$a_1 b_2$	0	1	1	1	0	1	2.72
$A_1 B_1 C_1 D_2$	$a_1 b_1$	0	0	0	1	0	0	1.37
$A_1 B_1 C_1 D_2$	$a_2 b_2$	0	0	0	1	1	1	3.36
$A_1 B_1 C_1 D_2$	$a_2 b_1$	0	0	0	1	1	0	2.3
$A_1 B_1 C_1 D_2$	$a_1 b_2$	0	0	0	1	0	1	2.18
$A_1 B_2 C_2 D_1$	$a_1 b_1$	0	1	1	0	0	0	1.56
$A_1 B_2 C_2 D_1$	$a_2 b_2$	0	1	1	0	1	1	4.14
$A_1 B_2 C_2 D_1$	$a_2 b_1$	0	1	1	0	1	0	2.31
$A_1 B_2 C_2 D_1$	$a_1 b_2$	0	1	1	0	0	1	2.83
$A_1 B_2 C_1 D_1$	$a_1 b_1$	0	1	0	0	0	0	1.85
$A_1 B_2 C_1 D_1$	$a_2 b_2$	0	1	0	0	1	1	4.19
$A_1 B_2 C_1 D_1$	$a_2 b_1$	0	1	0	0	1	0	2.57
$A_1 B_2 C_1 D_1$	$a_1 b_2$	0	1	0	0	0	1	3.7
$A_1 B_1 C_2 D_2$	$a_1 b_1$	0	0	1	1	0	0	1.55
$A_1 B_1 C_2 D_2$	$a_2 b_2$	0	0	1	1	1	1	3.11
$A_1 B_1 C_2 D_2$	$a_2 b_1$	0	0	1	1	1	0	2.41
$A_1 B_1 C_2 D_2$	$a_1 b_2$	0	0	1	1	0	1	2.4
$A_1 B_1 C_2 D_1$	$a_1 b_1$	0	0	1	0	0	0	0.54
$A_1 B_1 C_2 D_1$	$a_2 b_2$	0	0	1	0	1	1	1.84
$A_1 B_1 C_2 D_1$	$a_2 b_1$	0	0	1	0	1	0	0.86
$A_1 B_1 C_2 D_1$	$a_1 b_2$	0	0	1	0	0	1	1.57
$A_2 B_2 C_1 D_1$	$a_1 b_1$	1	1	0	0	0	0	1.77
$A_2 B_2 C_1 D_1$	$a_2 b_2$	1	1	0	0	1	1	3.08
$A_2 B_2 C_1 D_1$	$a_2 b_1$	1	1	0	0	1	0	2.76
$A_2 B_2 C_1 D_1$	$a_1 b_2$	1	1	0	0	0	1	2.88
$A_2 B_2 C_2 D_1$	$a_1 b_1$	1	1	1	0	0	0	1.84
$A_2 B_2 C_2 D_1$	$a_2 b_2$	1	1	1	0	1	1	3.67
$A_2 B_2 C_2 D_1$	$a_2 b_1$	1	1	1	0	1	0	2.63
$A_2 B_2 C_2 D_1$	$a_1 b_2$	1	1	1	0	0	1	2.86
$A_2 B_2 C_2 D_2$	$a_1 b_1$	1	1	1	1	0	0	1.97
$A_2 B_2 C_2 D_2$	$a_2 b_2$	1	1	1	1	1	1	4.49
$A_2 B_2 C_2 D_2$	$a_2 b_1$	1	1	1	1	1	0	3.02
$A_2 B_2 C_2 D_2$	$a_1 b_2$	1	1	1	1	0	1	3.17

Interaction Between Peptide and MCR

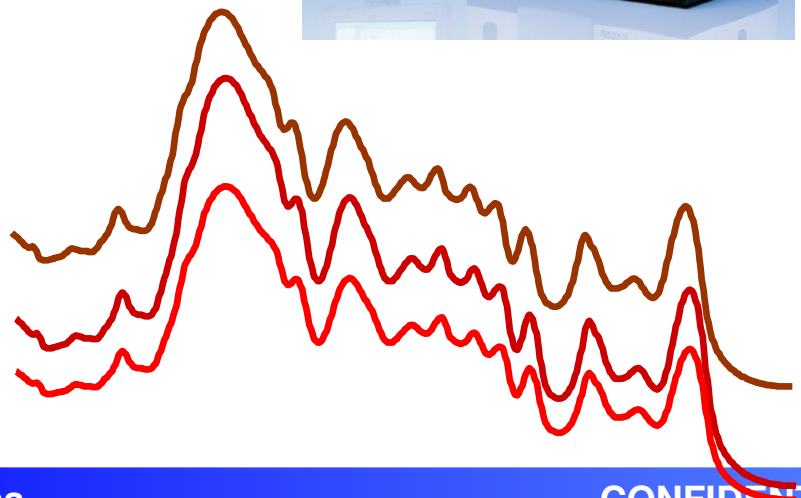


Metabonomics – Simplified view



Biological samples

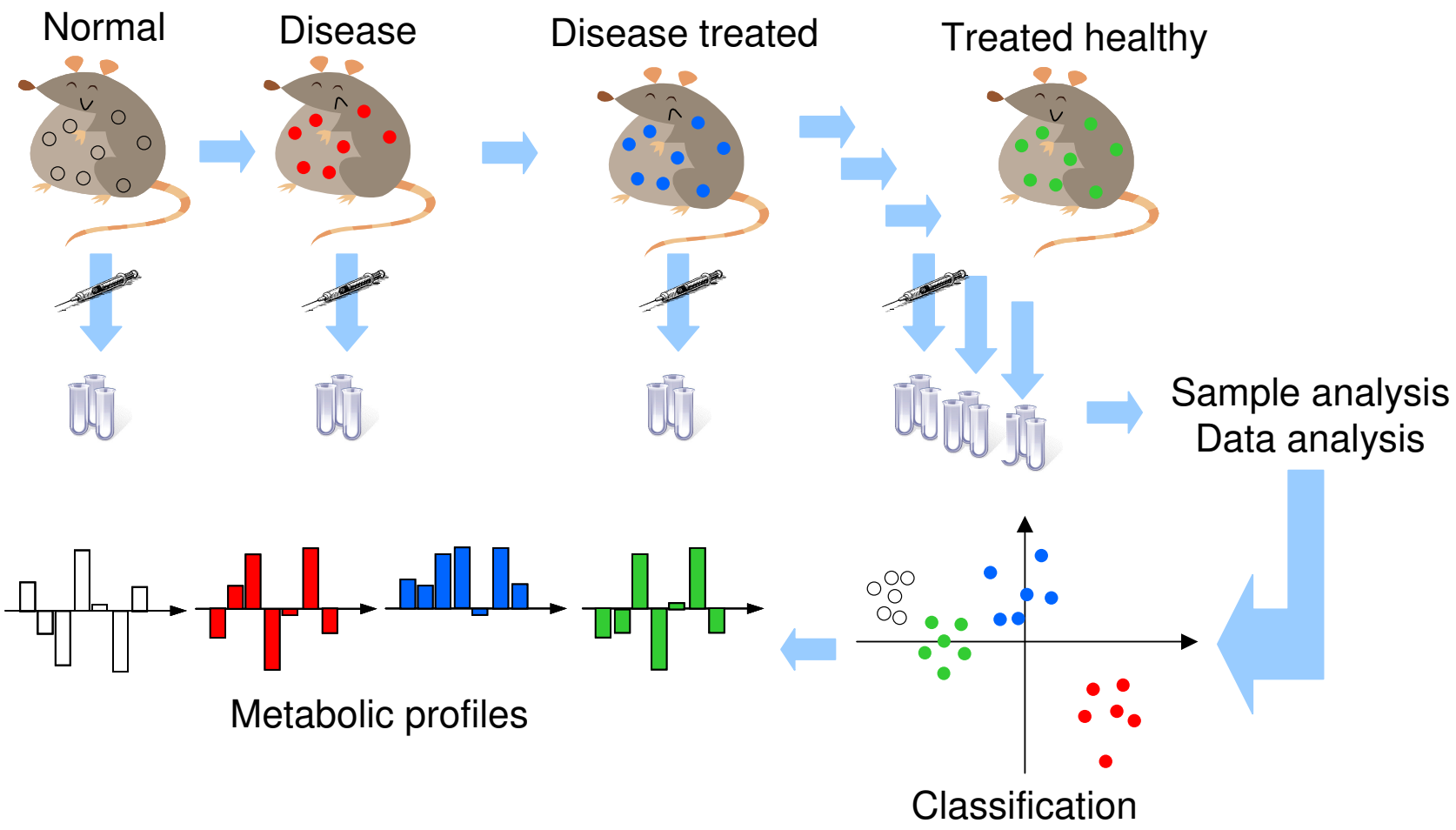
Biochemical analysis of endogenous metabolites



Information?

The challenge in modern biology is not in data collection but rather in maximizing information in data - **How to do it!**

Metabonomics at a glance



Strategy

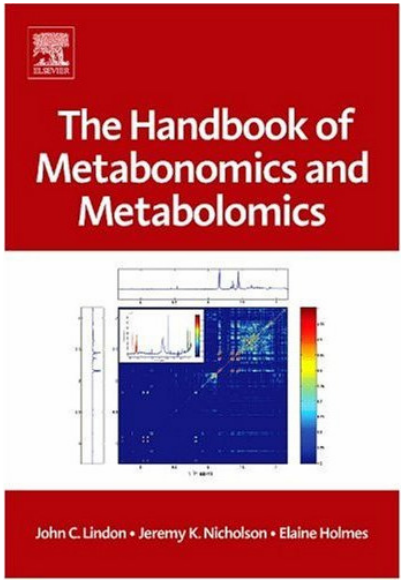
1. Formalize the aim
   - What do we want?

2. Selection of objects
   - Design of Experiments (DOE)
   - Multivariate design (MVD)

3. Sample preparation and profiling of human and animal samples
   - I*n vivo, in vitro* samples
   - Blood, Urine, Cerebral Spinal Fluid (CSF)
   - Synovial fluid (joint), Bowel fluids, Feces, Tissues

4. Integration and evaluation of collected data
   - Exploratory analysis, Interpretation & Visualization
   - Prediction models
   - Patterns
   - Target identification

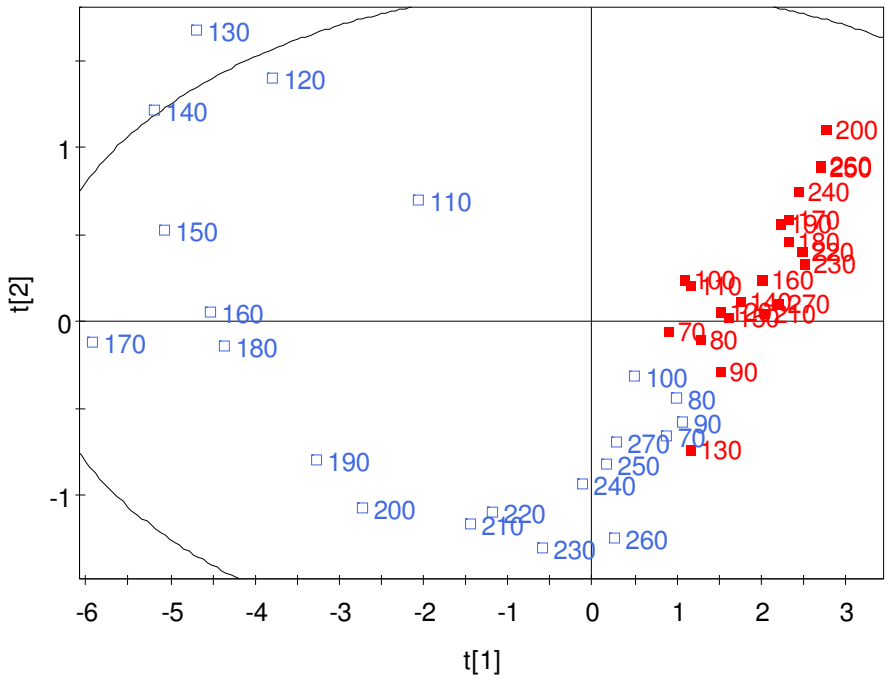
5. Identify and define IPR opportunities and strategies



Pigs

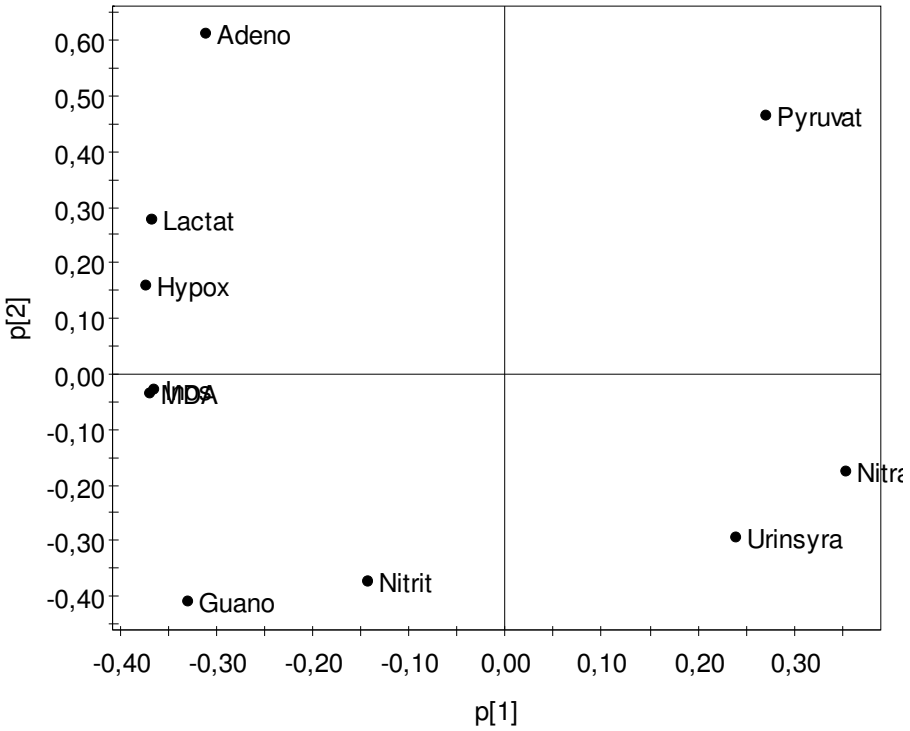
- Heart Ich/rep
- Microdialysis

SpacePig.M10 (PLS), Untitled, Work set
Scores: t[1]/t[2]



Ellipse: Hotelling T2 (0,05)
Simca-P 7.01 by Umetri AB 2001-06-06 11:01

SpacePig.M10 (PLS), Untitled, Work set
Loadings: p[1]/p[2]



Simca-P 7.01 by Umetri AB 2001-06-06 10:51

”Data filtering”

- Large data – difficult to handle and overview
- Starting point – ask specific question and filter the data, use prior information
- Depending on the question the analysis can be focused at finding different patterns
- Information identified in data subsets can thereafter be applied to whole dataset
- Combine metabonomics data with questionnaire ➔ connection genetics vs. lifestyle and environment

Parameters ?

- Example of filtering questions
- Compound classification – prior information
  - Mechanism of action
  - Classes of compound
  - Indication
  - Side effects
  - Etc
- Use structure characterisation and PCA to generate principal properties ➔ make representative/diverse selection to base first model on (e.g. sub-set)

Multivariate design

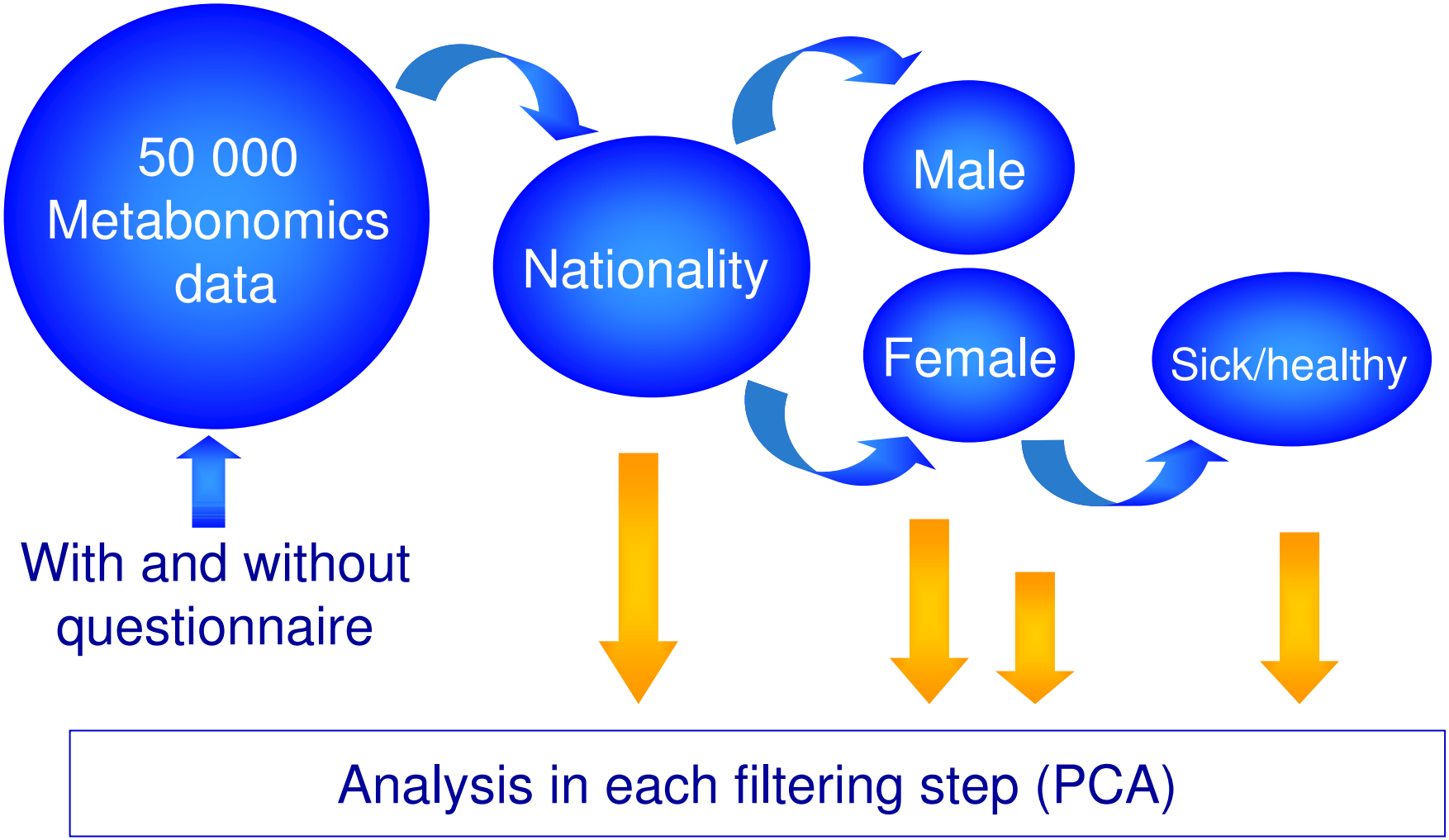
- Make PCA on reduced data set
- Make multivariate design(s) in the generated scores (principal properties) ➔ cluster based design
- Select training set and test set according to a multivariate design – if possible
  - Validate the model
  - Obtain a more robust model
  - Classify
  - Identify biomarkers

Example of questions for design

- Gender
  - male/female
    Looking for genetic patterns and lifestyle patterns/markers separating the sexes
- Genetic profile
  - Asian, European, American
    Looking for genetic patterns/markers
- Nationality
  - China, UK or Asian, European
    Looking for lifestyle and environment patterns/markers
- Anamnesis
  - ”healthy” vs. un-healthy
    Looking for disease pattern/markers

And combinations there of

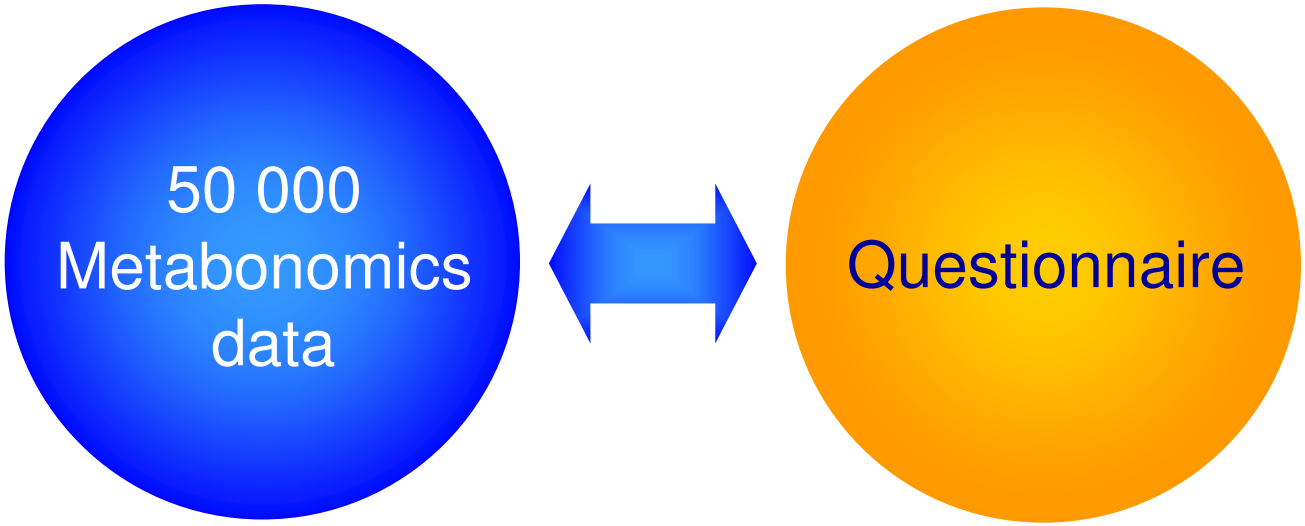
Data layout


50 000 Metabonomics data
Nationality
Male
Female
Sick/healthy
With and without questionnaire
Analysis in each filtering step (PCA)

Data layout

- Pregnant (folic acid, hormones)
- Diabetes (insulin)
- Gastric/stomach ulcer (Losec/Nexium)
- Head ace (paracetamole)
- Smoker/non-smoker
- Age < 50 <
- BMI

Lifestyle vs. metabonomics



- Identifying patterns in lifestyle by relating metabonomics to questionnaire (PLS)

Parameters ?

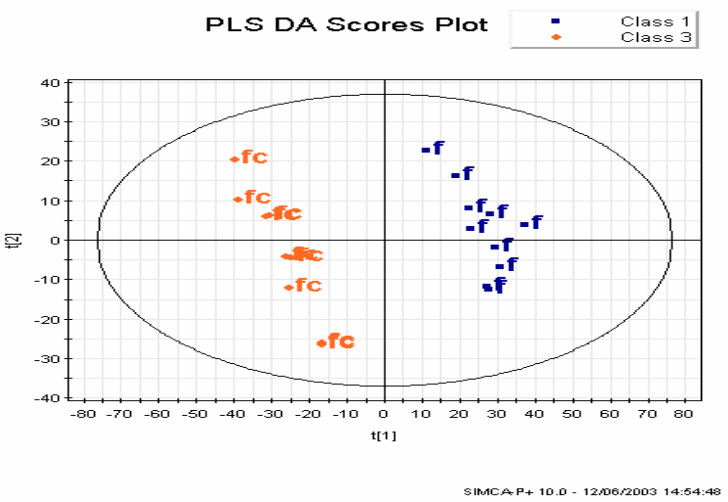
- Example of filtering questions
- Compound classification – prior information
  - Mechanism of action
  - Classes of compound
  - Indication
  - Side effects
  - Etc
- Use structure characterisation and PCA to generate principal properties ➔ make representative/diverse selection to base first model on (e.g. sub-set)

Possible problems

- Selection of control group *in vivo*?
- Selection of reference structures?
- Selection of colon's (LC/GC)?
- Sample handling/work up?
- Training set, test set – model validation, classification, biomarker
- Fast/Slow responders

Class specific models 2 - PLS-DA

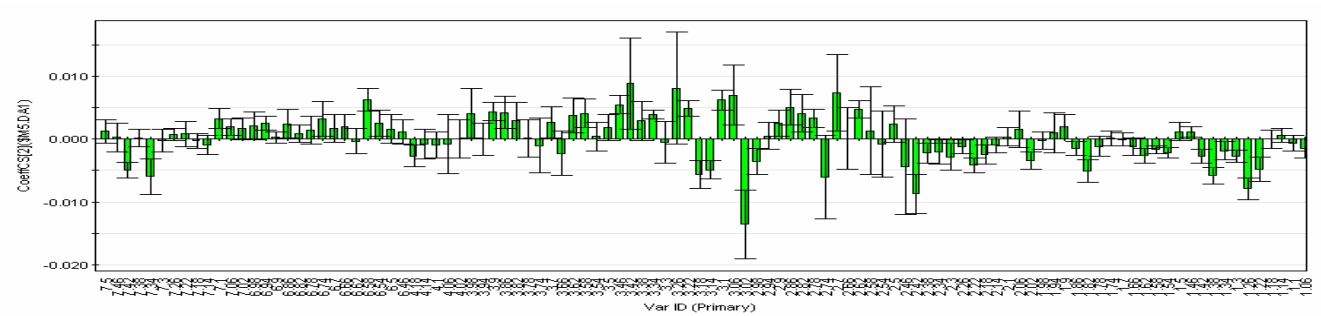
- PLS-DA Maximum separation projection
- PLS model built on group membership (1 or 0)
- Coefficients show as a whole how do the variables change when going from one group to another
- Here controls (f) and treated (fc) are separated



PLS-DA between “f” and “fc”

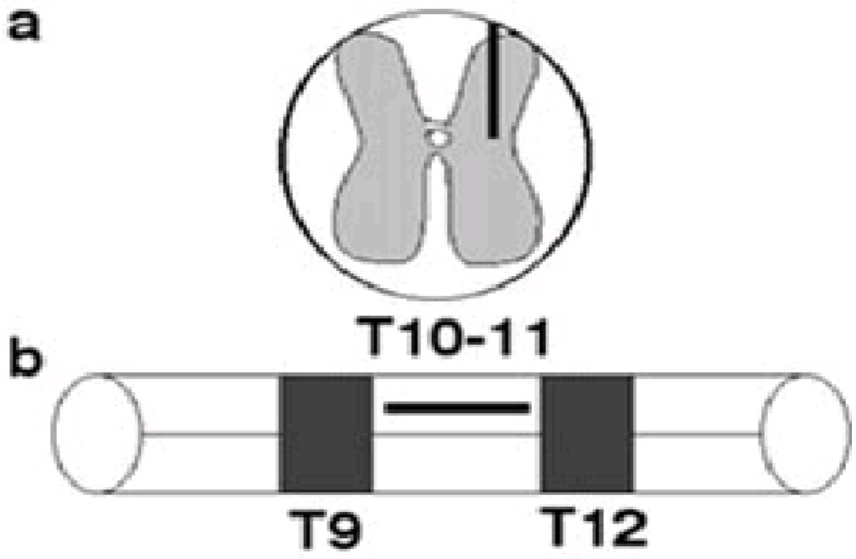
- Coefficients show as a whole how do the variables change when going from one group to another

- Chemical shift regions or masses influential for the separation of the two classes



Incision model short description

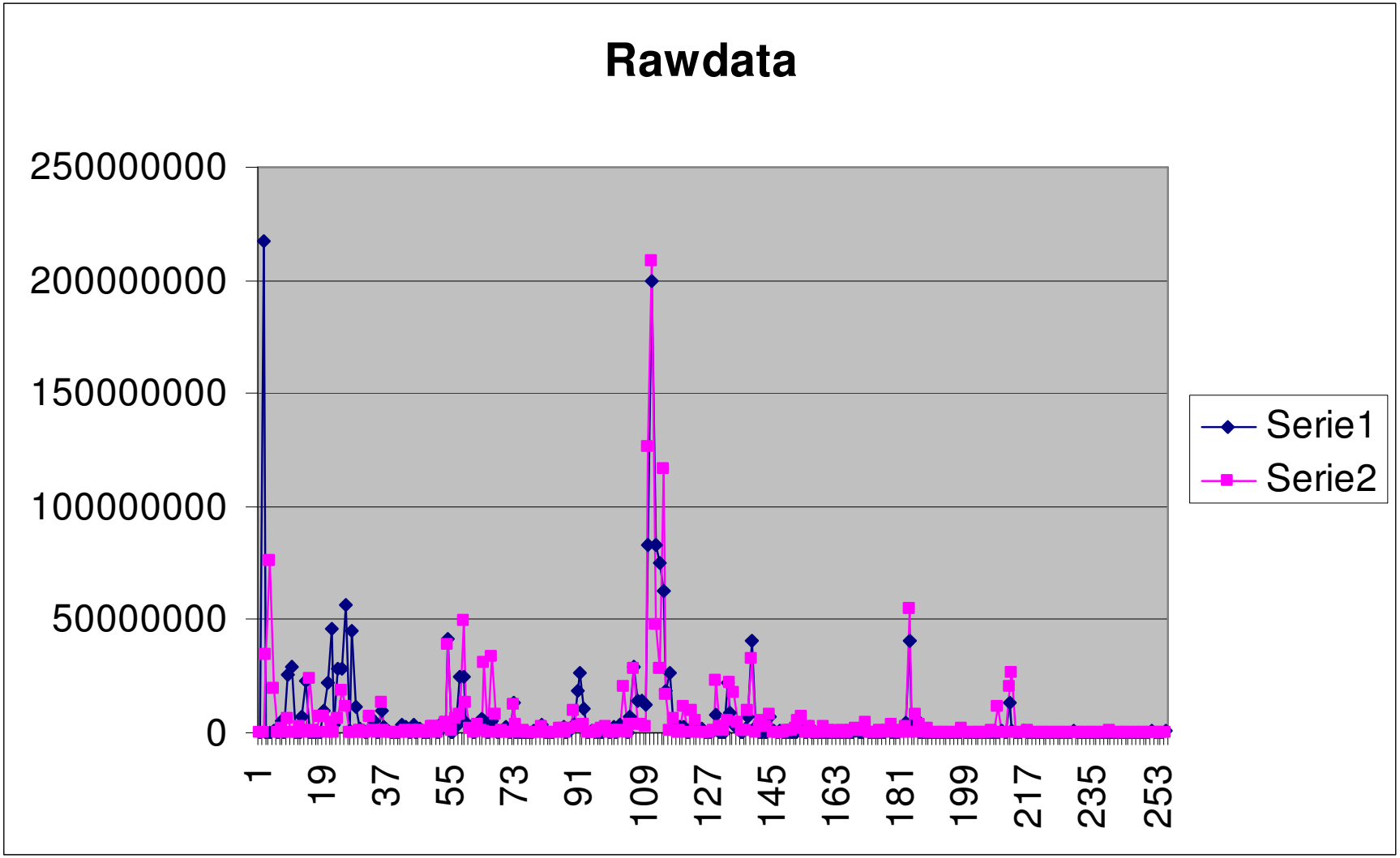
- Surgical intervention in rats. The SCI was induced by making a longitudinal incision into the right dorsal horn of the T10-11 segment
- AP173 and AP713 was given post-injury 5 minutes after injury with a topical dose of 5 µg in saline solution
- Tarlov scale was tested during 1 to 7 days
- The other responses was estimated day 7

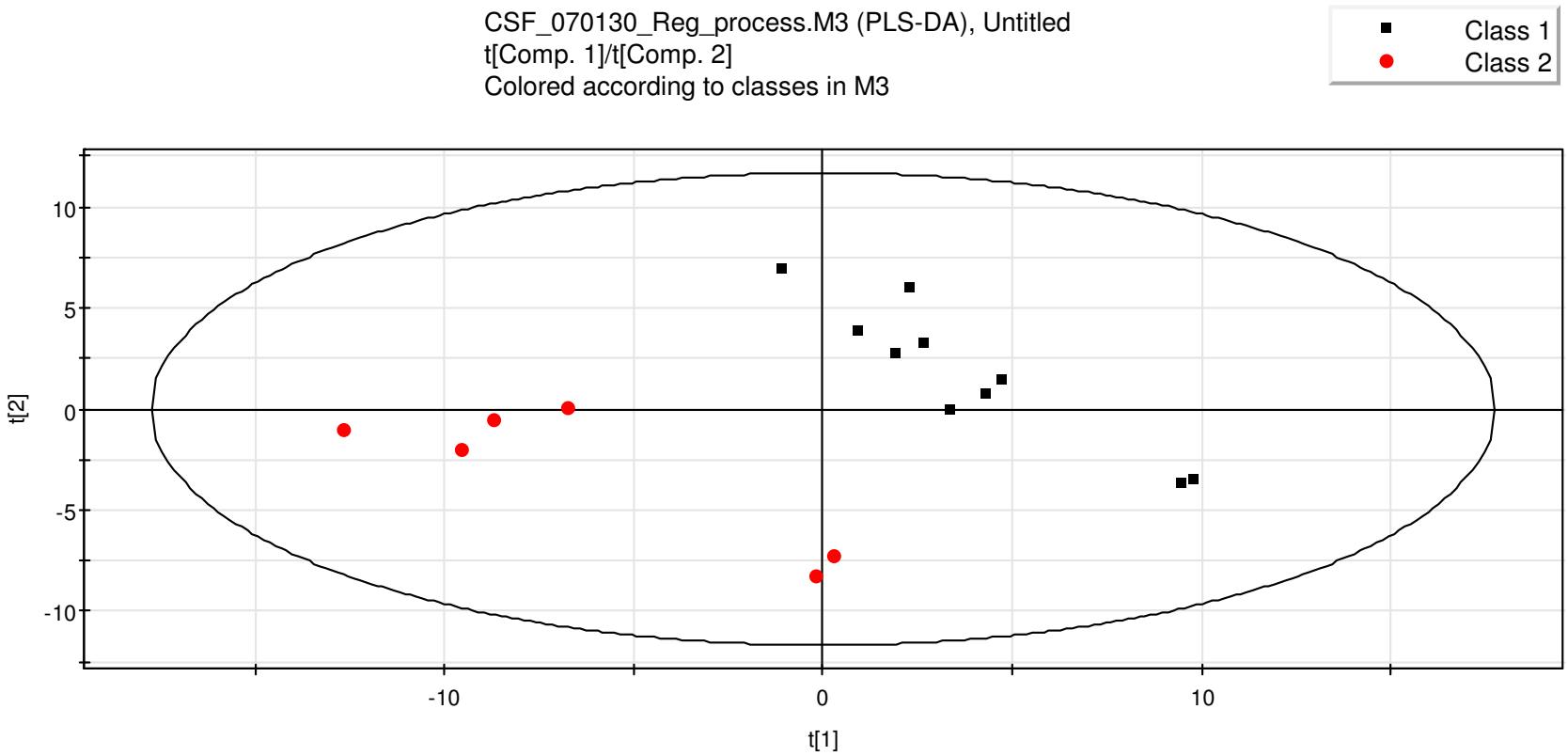


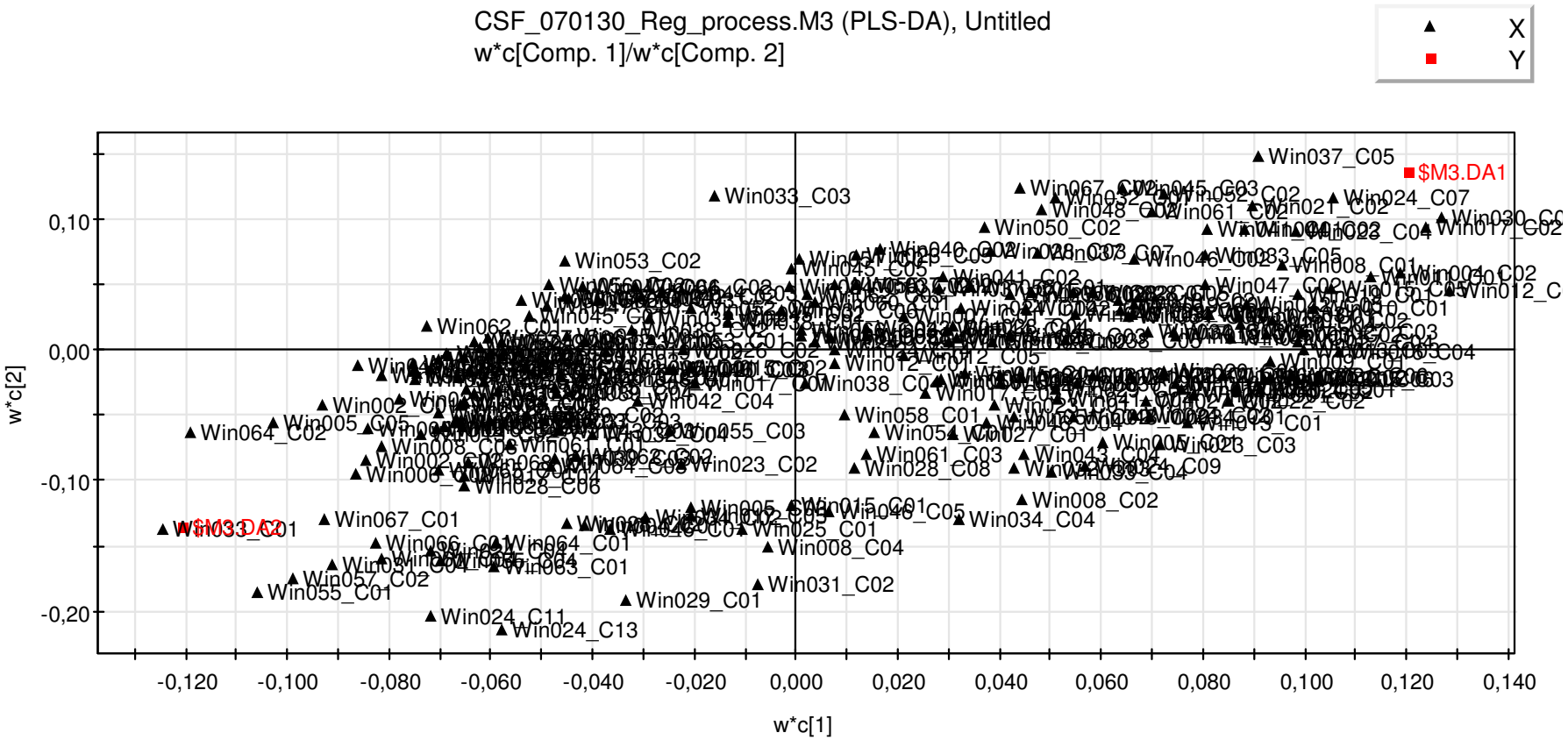
Status

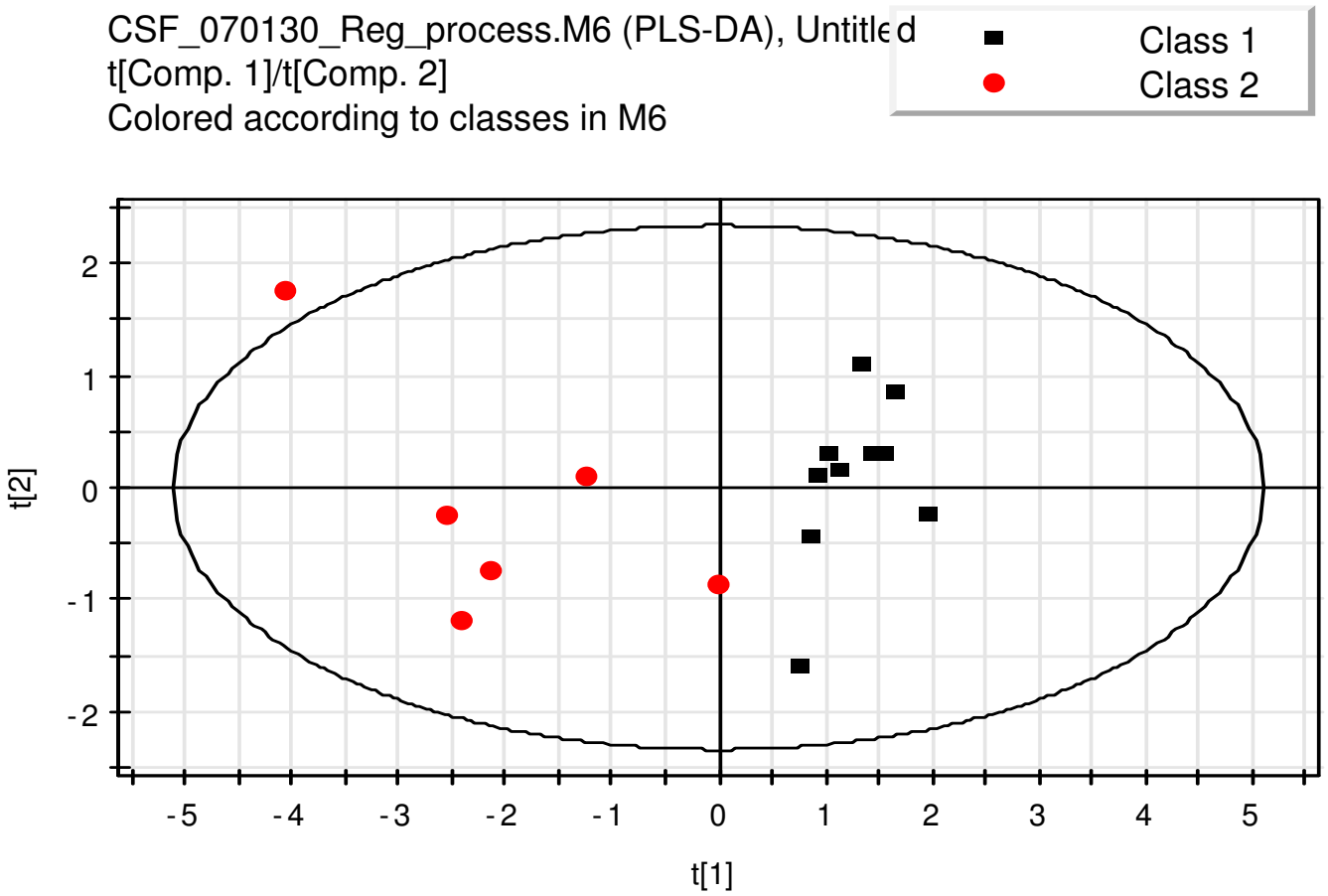
- CSF, Plasma and three segments of spinal cord from normal rats
- CSF, plasma and three segments of spinal cord 5h after injury
- GC-MS and solid phase NMR

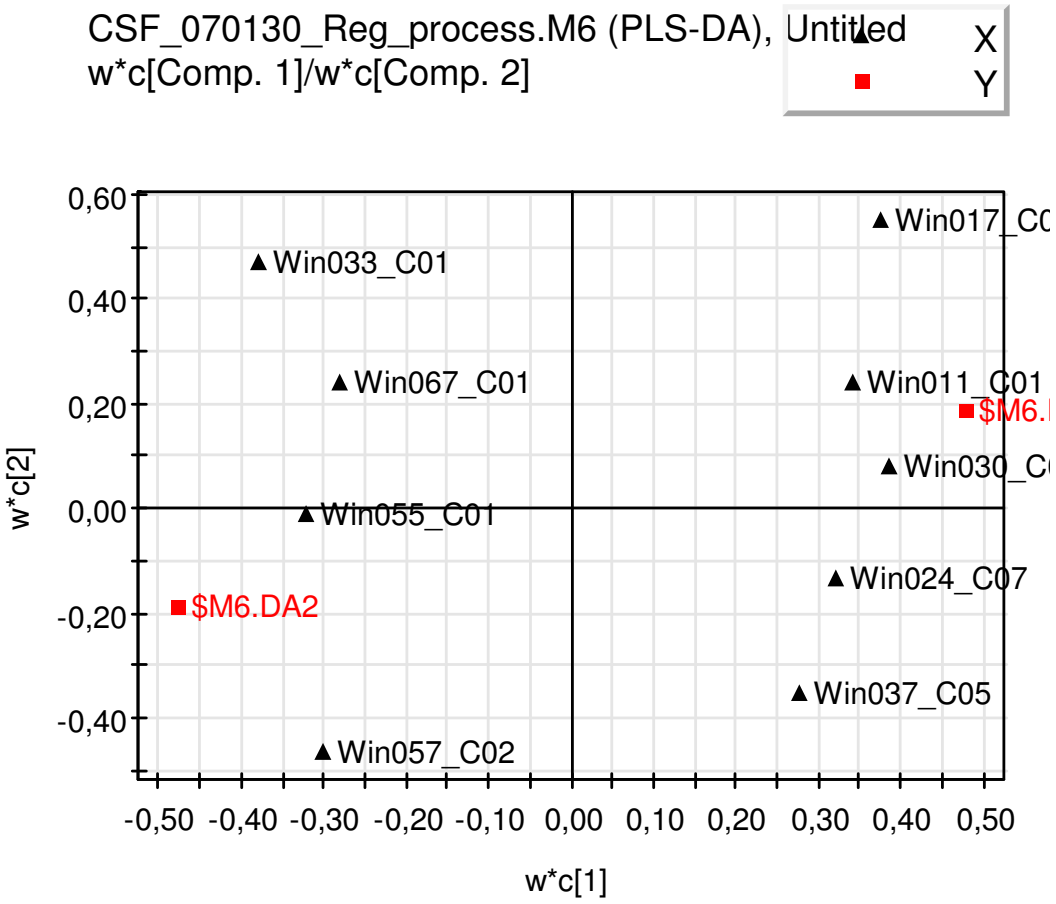
GC/MS PROFILE




CSF_070130_Reg_process.M3 (PLS-DA), Untitled
t[Comp. 1]/t[Comp. 2]
Colored according to classes in M3
Class 1
Class 2
10
5
0
-5
-10
t[2]
-10
0
10
t[1]


CSF_070130_Reg_process.M3 (PLS-DA), Untitled
w*c[Comp. 1]/w*c[Comp. 2]
X
Y
$M3.DA1
$M3.DA2
w*c[1]
w*c[2]
0,140
0,120
0,100
0,080
0,060
0,040
0,020
0,000
-0,020
-0,040
-0,060
-0,080
-0,100
-0,120
0,10
0,00
-0,10
-0,20


CSF_070130_Reg_process.M6 (PLS-DA), Untitled
t[Comp. 1]/t[Comp. 2]
Colored according to classes in M6
Class 1
Class 2
t[2]
t[1]
2
1
0
-1
-2
-5
-4
-3
-2
-1
0
1
2
3
4
5


CSF_070130_Reg_process.M6 (PLS-DA), Untitled
w*c[Comp. 1]/w*c[Comp. 2]
X
Y
0,60
0,40
0,20
0,00
-0,20
-0,40
w*c[2]
Win033_C01
Win017_C0
Win067_C01
Win011_C01
$M6.
Win030_C
Win055_C01
Win024_C07
$M6.DA2
Win037_C05
Win057_C02
-0,50 -0,40 -0,30 -0,20 -0,10 0,00 0,10 0,20 0,30 0,40 0,50
w*c[1]

Arthritis – Diagnosis

- Early diagnosis critical
  - Reduce symptoms of wear
  - More successful treatment with early medication*
- Rheumatoid arthritis
  - Physical examination, antibodies (today not specific for RA), X-ray
- Osteoarthritis
  - Physical examination, X-ray
- New diagnostic tools are needed…

* http://www.reumatikerforbundet.org/start.asp?sida=3955

Case study 1: Sample information

Analysed by gas chromatography/time of flight mass spectrometry (GC/TOFMS)

Deconvolution (MCR v.1.13)

- 59 Samples
- 181 Spectral Profiles

Outliers (possible)

- Not primary detected

Included Isotope labeled mixture

- 11 Internal standards + Methyl Stearate

Main Groups

- 3 Health Groups
  - **Control**
  - **Ra**
  - **Oa**

Rheumatoid Arthritis – brief background

- Worldwide prevalence of approximately 1%*
- Autoimmune disease, the body attacks itself, aetiology largely unknown**
- Treatment; irreversible disease, no known cure, medication to maintain mobility and ease pain
- Early diagnosis critical
  - Reduce symptoms of wear
  - More successful treatment with early medication*

- Diagnosis for rheumatoid arthritis
  - Physical examination, antibodies (today not specific for RA), X-ray, MRI

- New diagnostic tools are needed…

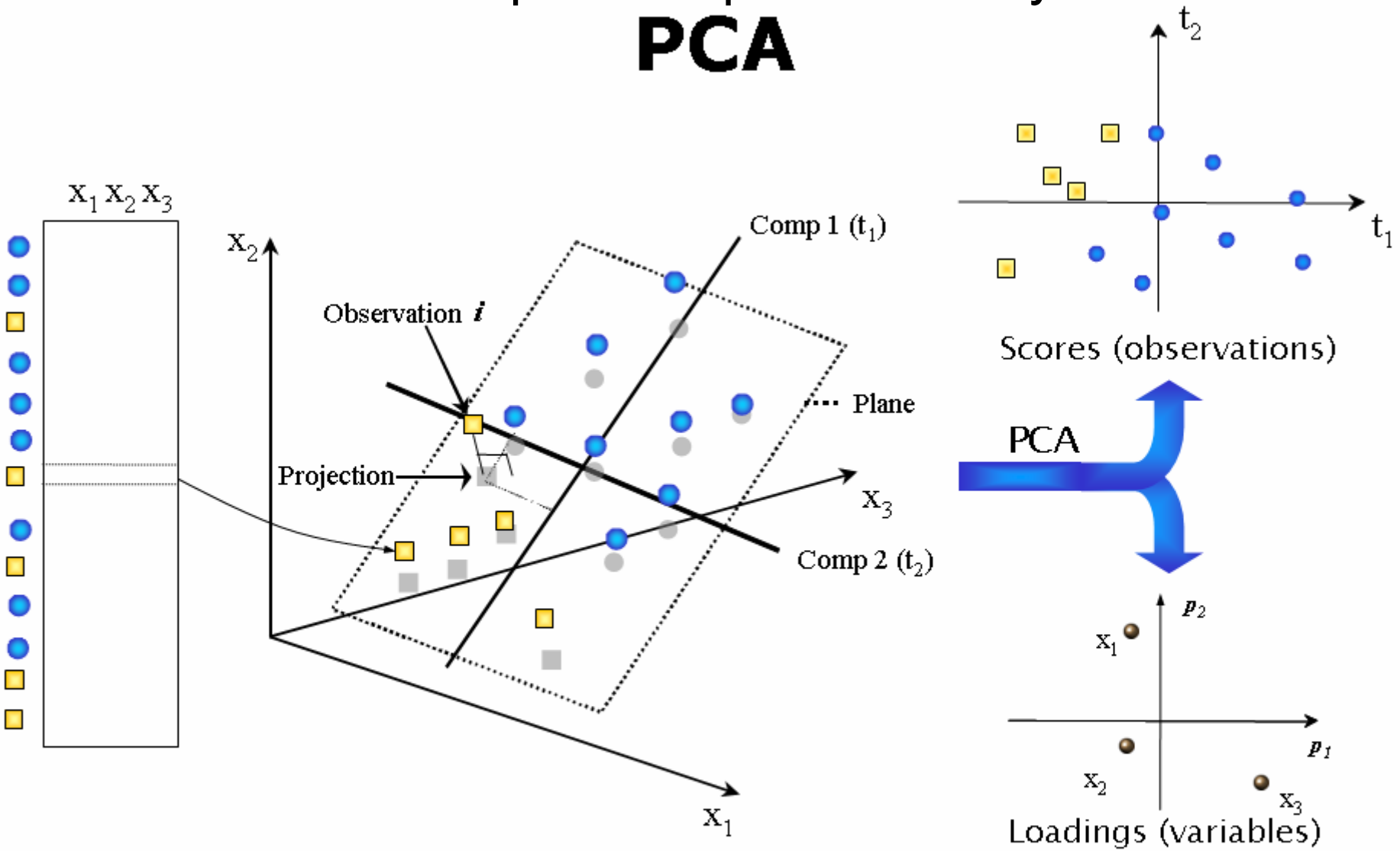
* Feldmann, M. et al., Cell. 85: 307-310 (1996)
** Krishnan, E., Joint Bone Spine. 70: 496-502 (2003)

CONFIDENTIAL

Isotope Labeled Internal Standards

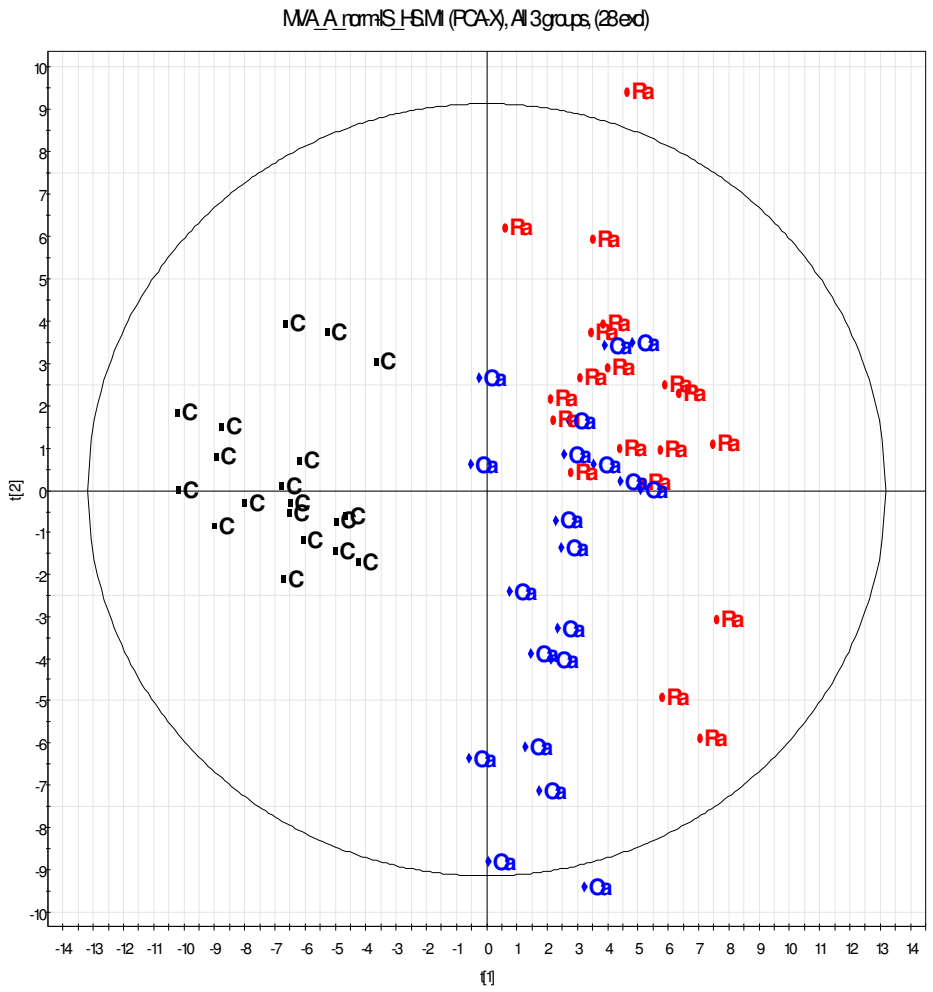
L-Proline-13C5-TMS (221 sec)	Win04_C02
Succinic Acid-D4-2TMS (225 sec)	Win04_C07
Salicylic Acid-D6-2TMS (265 sec)	Win08_C01
Di-Na-alfa-Ketoglutarat-13C4-2TMS (277 sec)	Win09_C03
L-Glumatic Acid-13C5-15N (285 sec)	Win10_C02
Putrescine-D4-4TMS (305 sec)	Win11_C07
Myristic Acid-13C3-TMS (324 sec)	Win13_C03
D-Glucose-13C6-5TMS (331 sec)	Win14_C01
Hexadecanoic Acid-13C4-TMS (354 sec)	Win16_C02
Methyl stearate (365 sec)	Win17_C03
Sucrose-13C12-8TMS (430 sec)	Win24_C01
Cholesterol-D7-TMS (483 sec)	Win29_C02

Principal Component Analysis

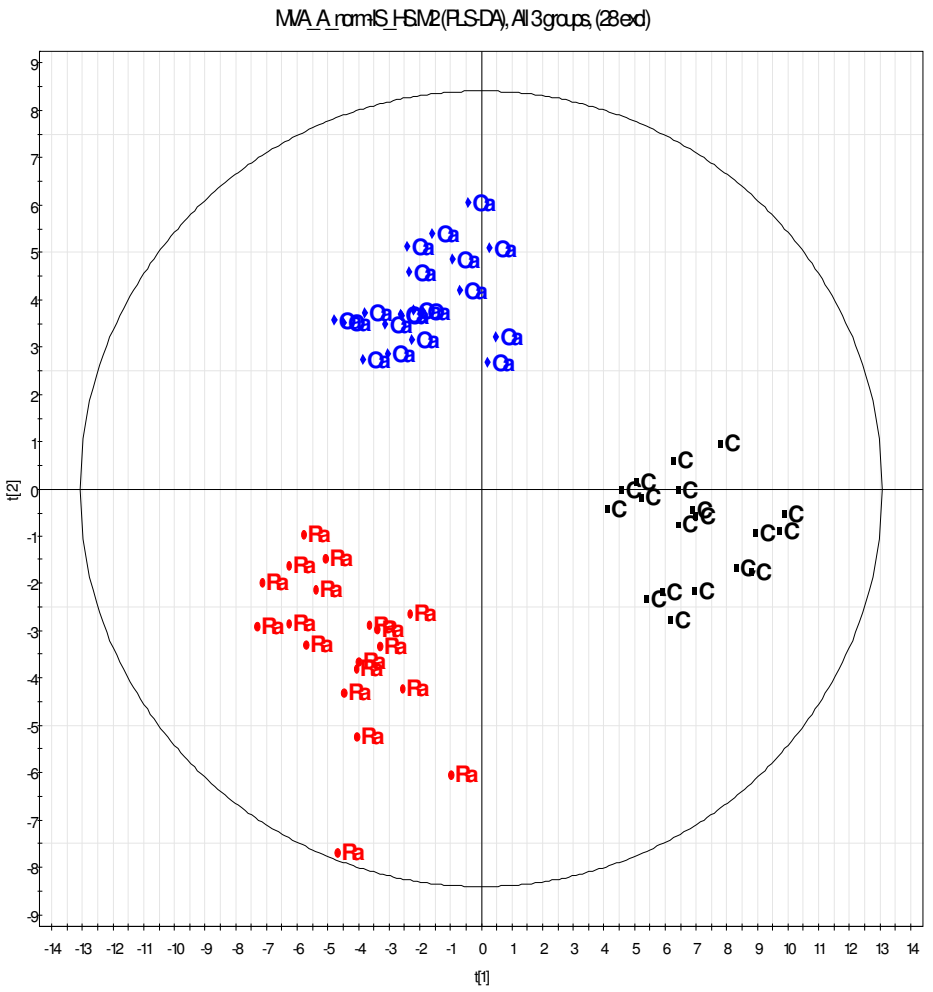


Data Overview

PCA – 181 variables

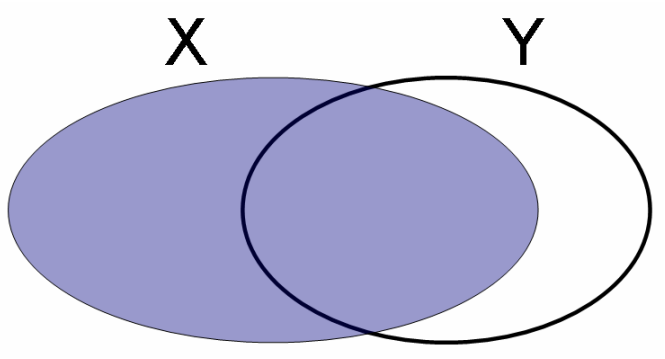


PLS-DA – 181 variables



The OPLS method

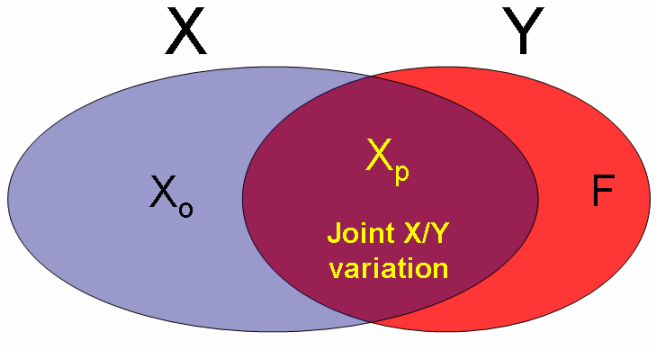
PLS, MLR, PCR, RR etc...



Mixes Y-orthogonal and Y-predictive variation
Uni-directional, Models Y from X

OPLS

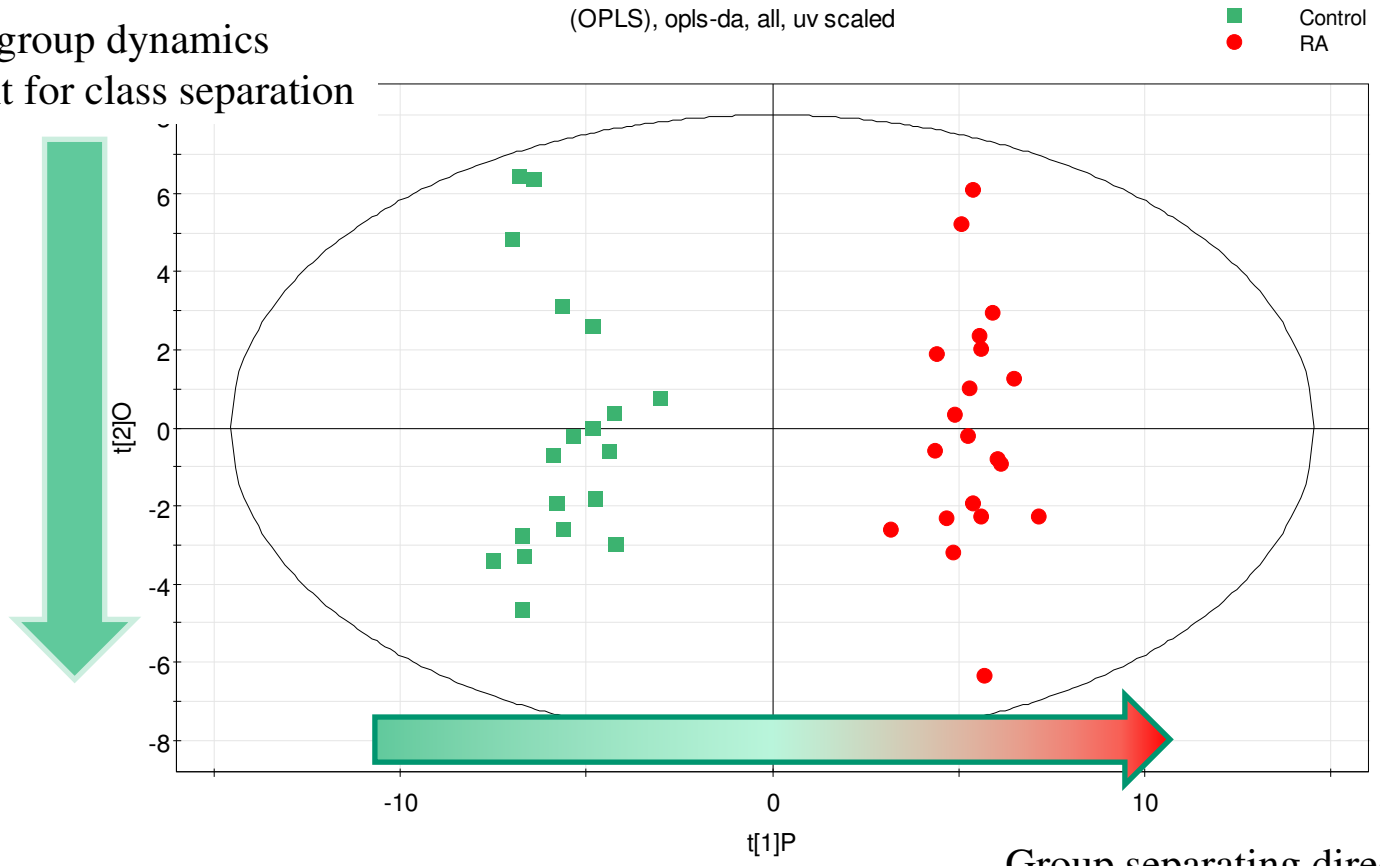
Orthogonal Projections of Latent Structures



Models X AND Y
Separates Orthogonal and Predictive variation
-e.g. 'between block' from 'within block'
Bi-directional
*O*nly uses predictive variation for modeling Y

Rheumatoid arthritis: Control vs. RA
OPLS-DA*- 204 putative biomarkers

Within group dynamics
not important for class separation



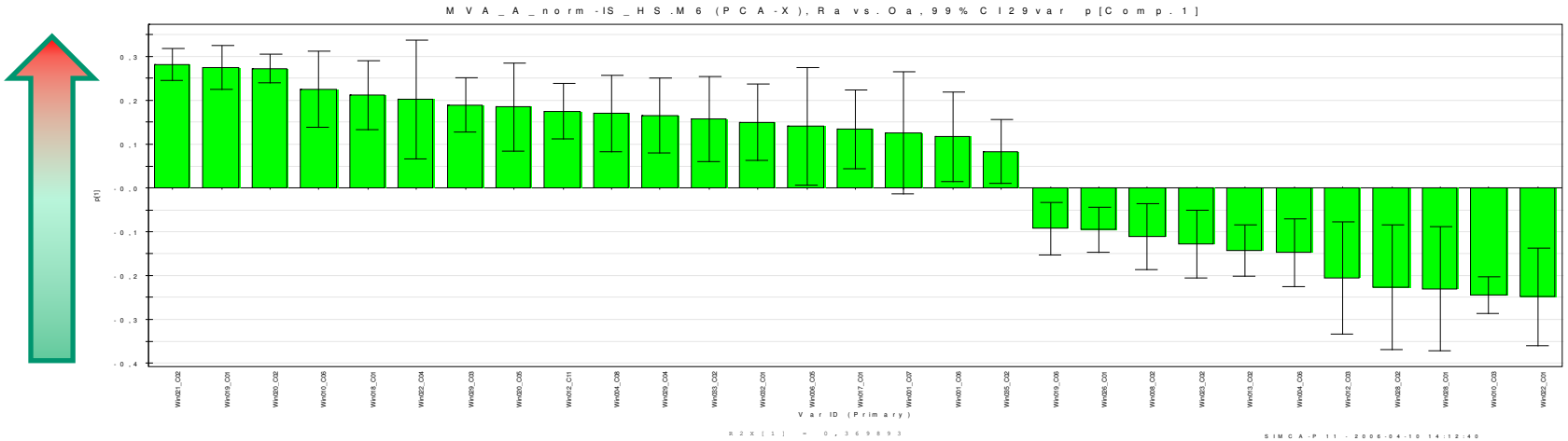
Group separating direction
Specific metabolites for healthy and diseased

*Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J.,
OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification.
J. Chemometrics, 20, 341-351 (**2006**).

Rheumatoid arthritis: Control vs. RA

- Significant (subset) metabolites for separation of RA samples from healthy controls.
  - Variables represents endogenous metabolites
  - Data set includes identified and yet to be identified potential biomarkers for diagnosis of RA

Positive correlation to Ra



Positive correlation to Control

RA: Comparison of the human case and animal models

- Great overlap of metabolites between humans and animals
  - Different metabolites show overlap in different animal models
  - Allows for identification of relevant animal models
  - Selection of model system for treatment studies

BM	Human Rheumatoid Arthritis	Mouse Collagen Induced Arthritis	Rat Adjuvant Induced Arthritis
EC001	↑	na	Na
EC002	↑	?	?
EC003	↑	↓	↓
EC004	↑	0/↓	↓
EC005	↓	na	na
EC006	↓	↓	↓
EC007	↓	↓	↓
EC008	↓	↓	↑
EC009	↓	↓	↓
EC010	↓	↑	↑
EC011	↓	0/↓	↓
EC012	↓	na	na
EC013	↓	↓	↓
EC014	↓	↓	?
EC015	↓	↓	↓
EC016	↓	?	↓
EC017	0	↓	↓
EC018	↑	↑	↓ / ?
EC019	↓	↓	↓
EC020	↓	↓ / ?	↓
EC021	?	↑ / ?	↑
EC022	↓	↓	↓
EC023	0	↓	↓
EC024	↑	↓	0/↑
EC025	↑	↓	↓
EC026	0/↑	↓	↑

CONFIDENTIAL

RA: Comparison of therapies in animal model

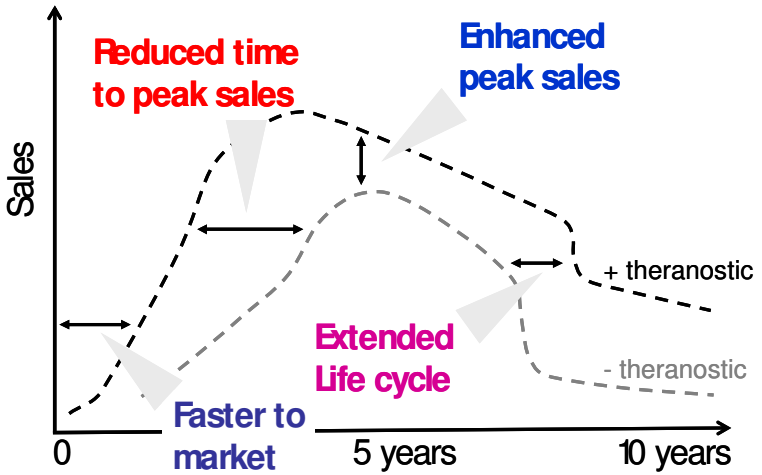
- Metabolites levels are affected by administered therapeutics
  - New drug (X) restore levels in more metabolites compared to MTX*
  - Useful in development of novel drugs
  - Tool in clinical studies to verify therapeutic effect in clinical studies
  - Concomitant development of novel drug and diagnostic test, theranostics?

	Vehicle	MTX	X 1mg	X 3mg	X 10mg
EC004	0/↑	↓	↓	0/↓	↓
EC006	0/↑ /?	0/?	0	↑	↑
EC007	↓	0/↑	0/↑	0/↓	↑
EC009	0	↑	↓	↑	↑
EC010	↑	↑	↑	↑	↑
EC011	0	0/↓	↓	0/↓	↑
EC012	0/↓	↑	0/↓	↑	↑
EC013	↑	0/↑	0/↓	↑	0/↑
EC014	↑	0/?	↑	↑	↑
EC015	0/↑	↑	0/↓	↑	↑
EC016	0	↓	↑	↑	↓
EC017	↓	↓	↓	↓	↓
EC018	↓	↓	↓	0/↑	0/↑
EC019	↓	↓	↓	↓	↓
EC022	↑	↑	0/↑	↑	↑
EC023	↓	0/↓	↓	↓	↓
EC024	↓	↓	↓	↓	↓
EC025	↓	↓	↓	↓	↓
EC026	↑	↑	↑	↑	↑

*MTX, methotrexate

Theranostics

- Theranostic approaches aim to target **thera**pies to those patients most likely to benefit by combining with diag**nostic** or prog**nostic** biomarkers
- Greatest benefit is realised with therapies that are effective, but in unknown patient sub-populations – HA therapy in OA is an excellent example
- An effective theranostic can drive sales of therapies as illustrated
- Regulation can be achieved within current pivotal clinical trial protocols
- Potential for the approach to track therapeutic efficacy



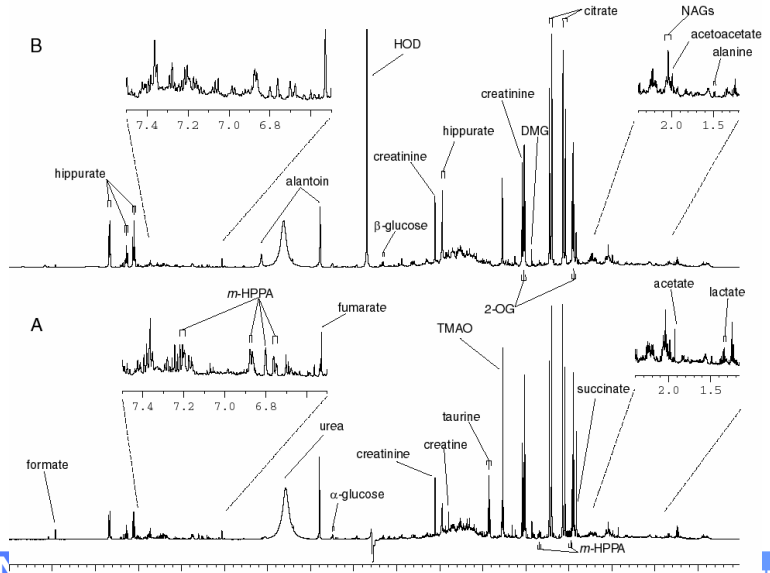
Adapted from Gilham (2002) Theranostics: an emerging tool in drug discovery and commercialisation. *Drug Discovery World*

Benefits

- Identification of responder/non responders populations
- Targeted patient recruitment to pivotal clinical trials, resulting in shorter, lower risk studies
- Stronger claims through mechanism of action knowledge and potential for data on therapeutic efficacy
- Improved HE evidence and addressing payer "entitlement" concern will increase payer uptake
- Much higher adoption in smaller market size – 50% penetration of 50% market size is better than 1-5% of the full market

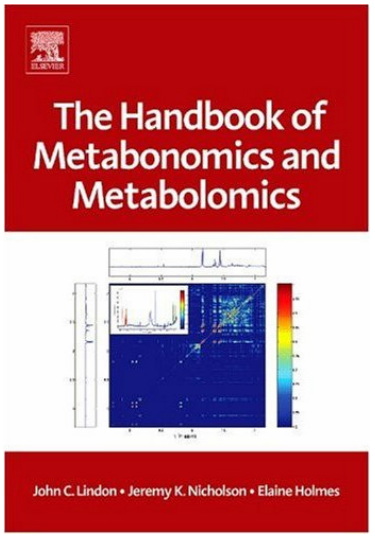
Case study 2: Functional foods study

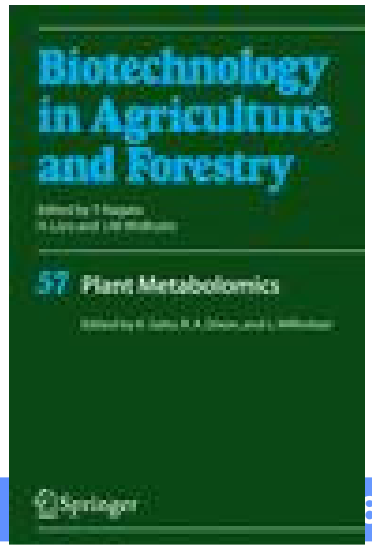
- NMR spectroscopy to detect metabolism changes due to food supplement
- N = 47 biofluid samples
  - 9 healthy individuals
  - Given prepared foodstuff for consumption
  - Multiple visits – document effect over time
- K = 32 768 variables
  - NMR shifts
- Goal: Effect of food supplement?



Chemometrics
- the information aspect for studying complex systems

1. **Define the aim**
   - ❑ What do we want?
2. **Selection of objects** (e.g. samples, time points, internode, experiments)
   - Design of Experiments (DOE)
   - Multivariate design (MVD)
3. **Sample preparation and characterisation**
   - Experimental protocol (e.g. FTIR, GCMS, Microarray)
   - Data processing (e.g. normalisation of FTIR, microarrays)
4. **Evaluation/Validation of collected data**
   - Exploratory analysis, Interpretation & Visualization
   - Dynamic study

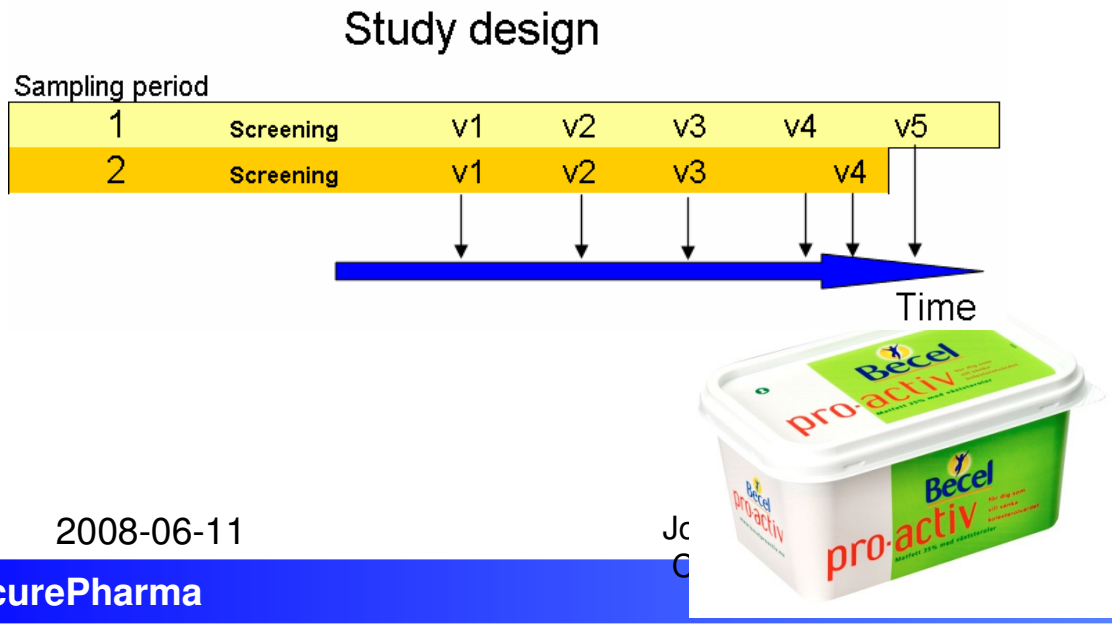




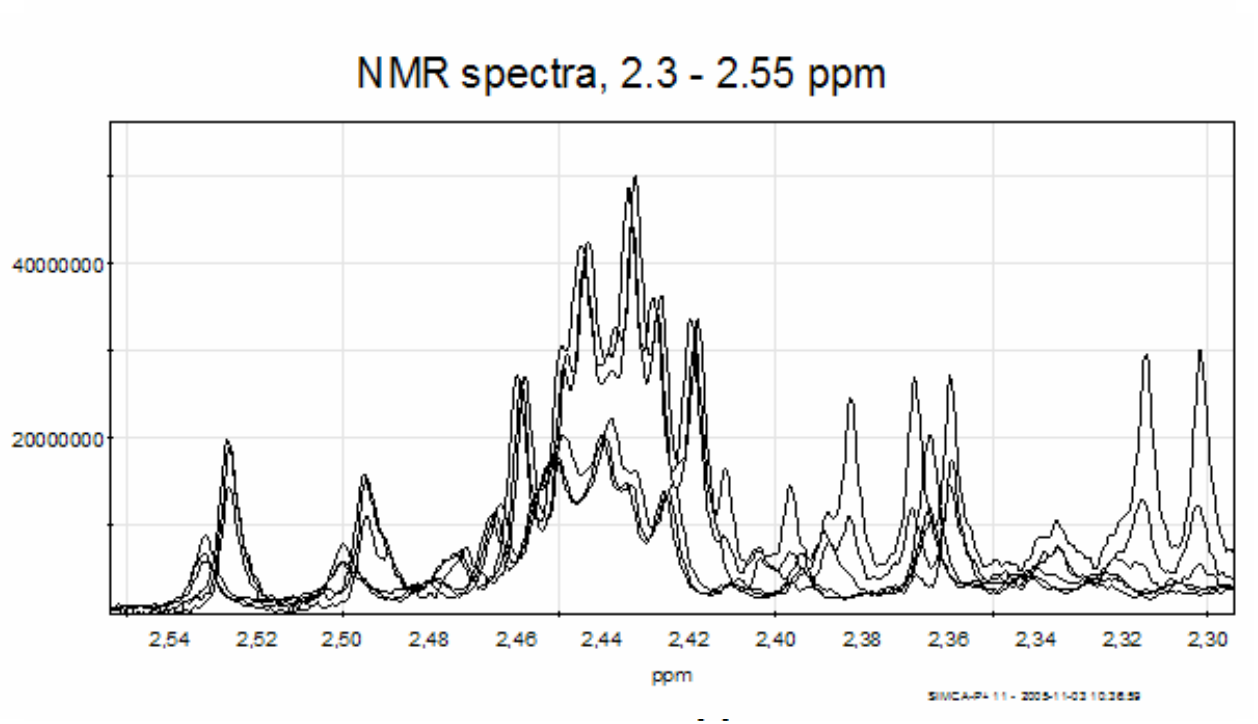
Example:
Functional foods study

- Functional foods: Foodstuffs with a documented health-promoting effect – besides energy addition

- Centre for Human Studies of Foodstuffs, Sweden
  - ❑ Inclusion/exclusion criteria
  - ❑ 9 individuals given prepared fo
  - ❑ Multiple visits – document effec



Case study 2: Overview of all individuals/samples

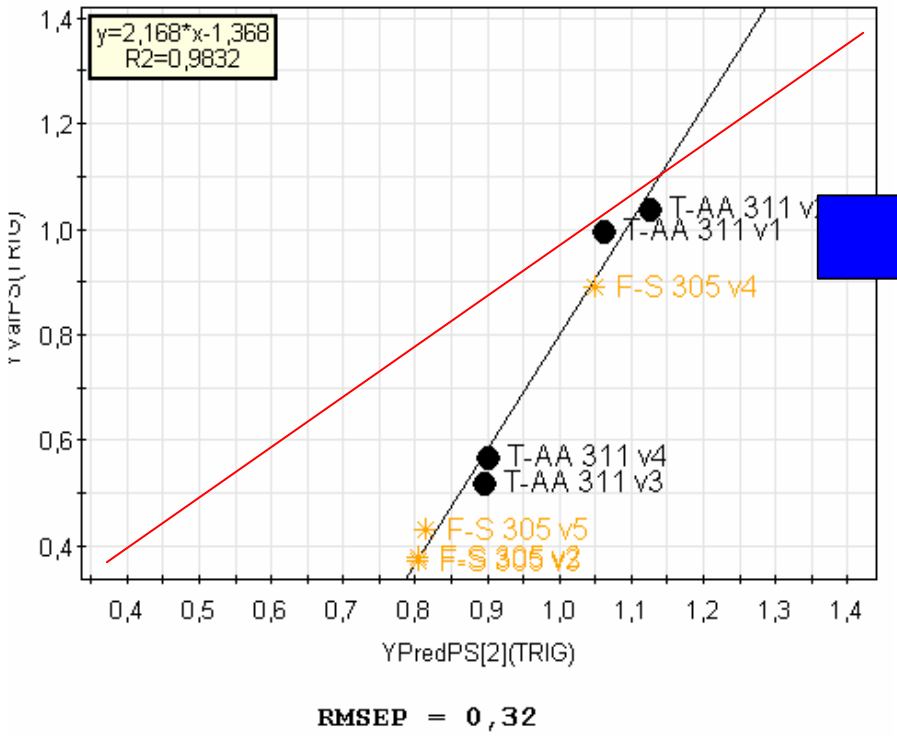


Clear separation.

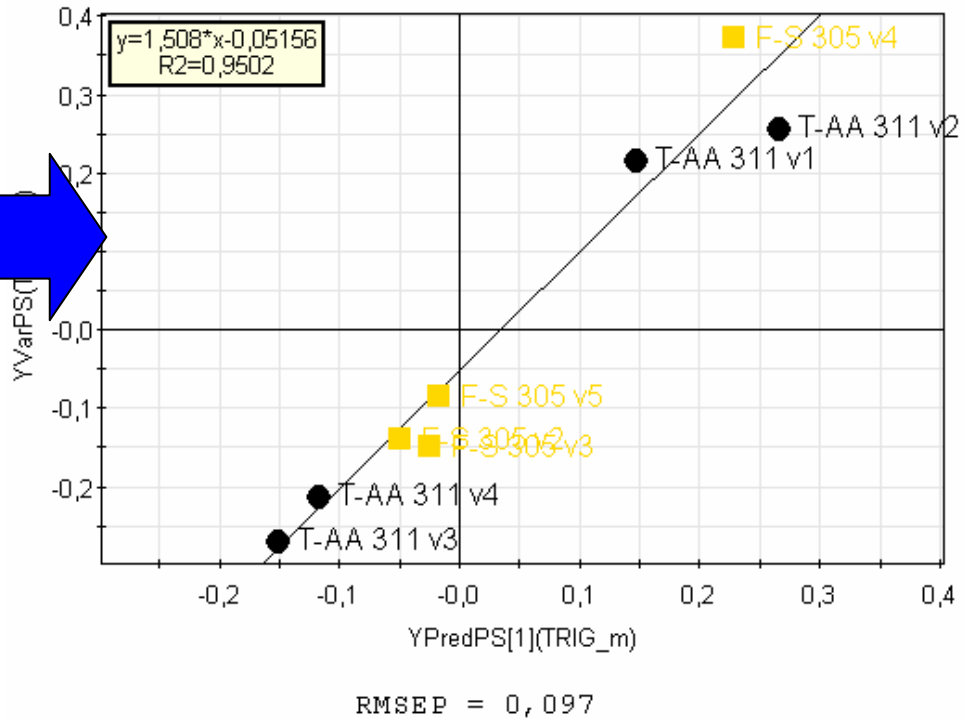
But ☹ … due to different sampling periods

Prediction of triglycerides
Raw NMR data vs Metabolic effect modelling

Prediction of triglycerides
Raw NMR data (major bias)

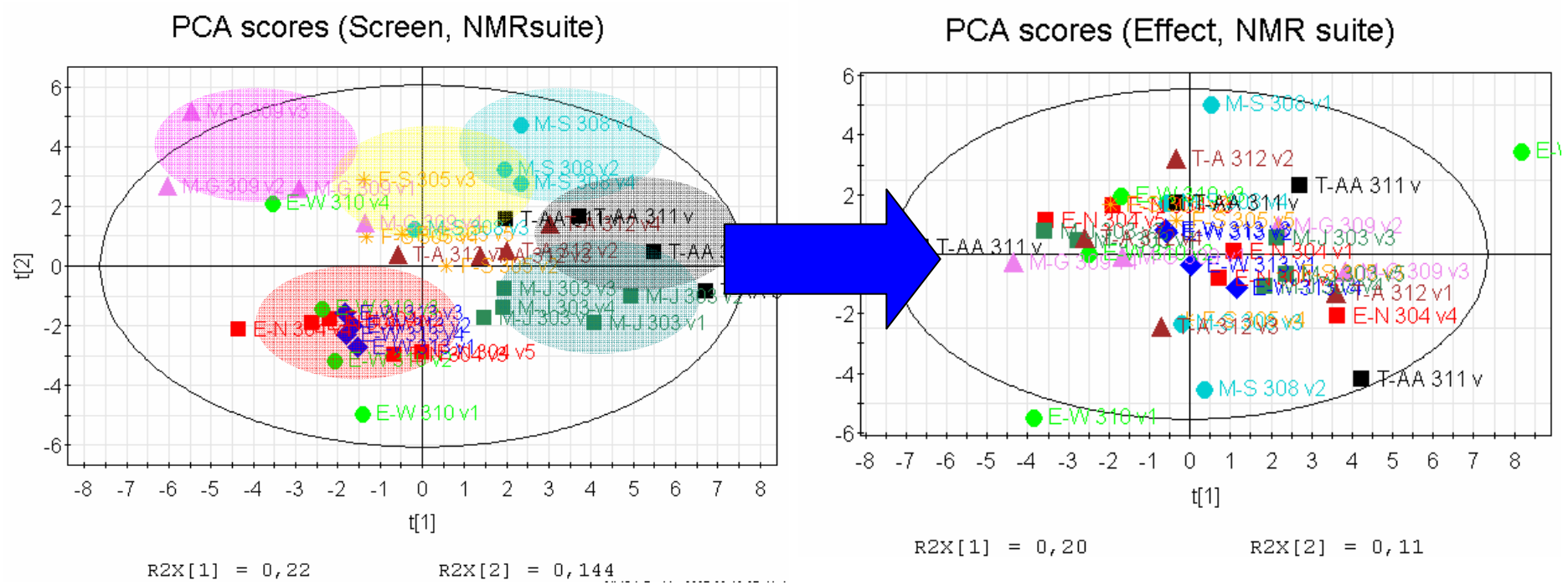


Prediction of triglycerides
Metabolic effect modelling

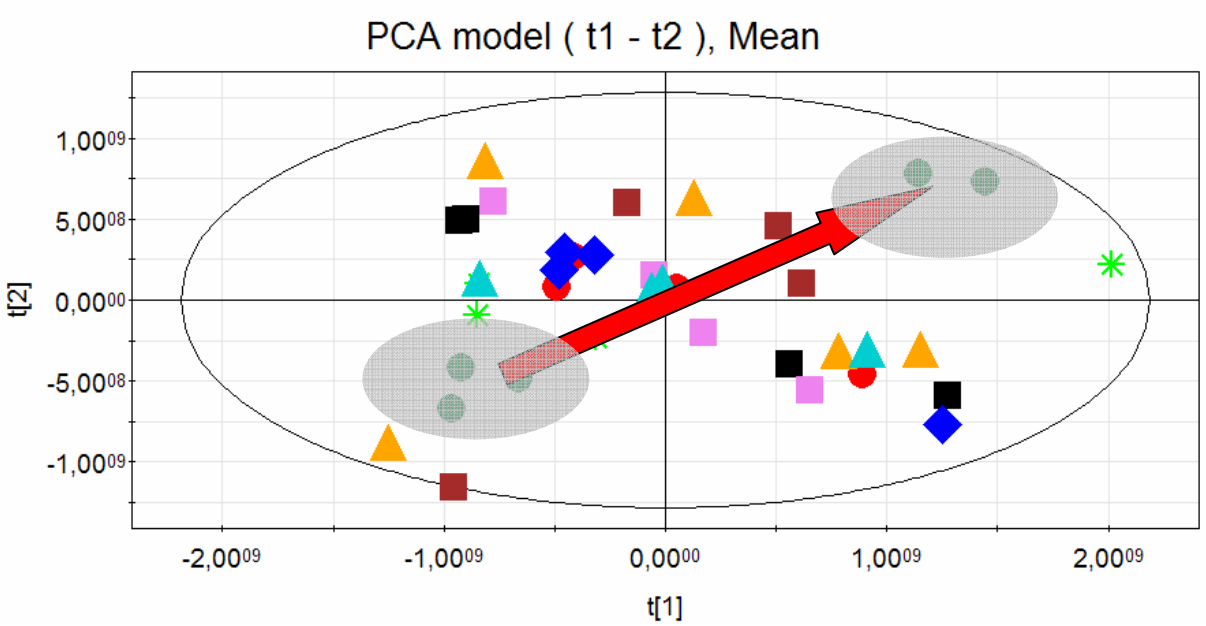


Case study 2: Modelling dynamic metabolomic time series data

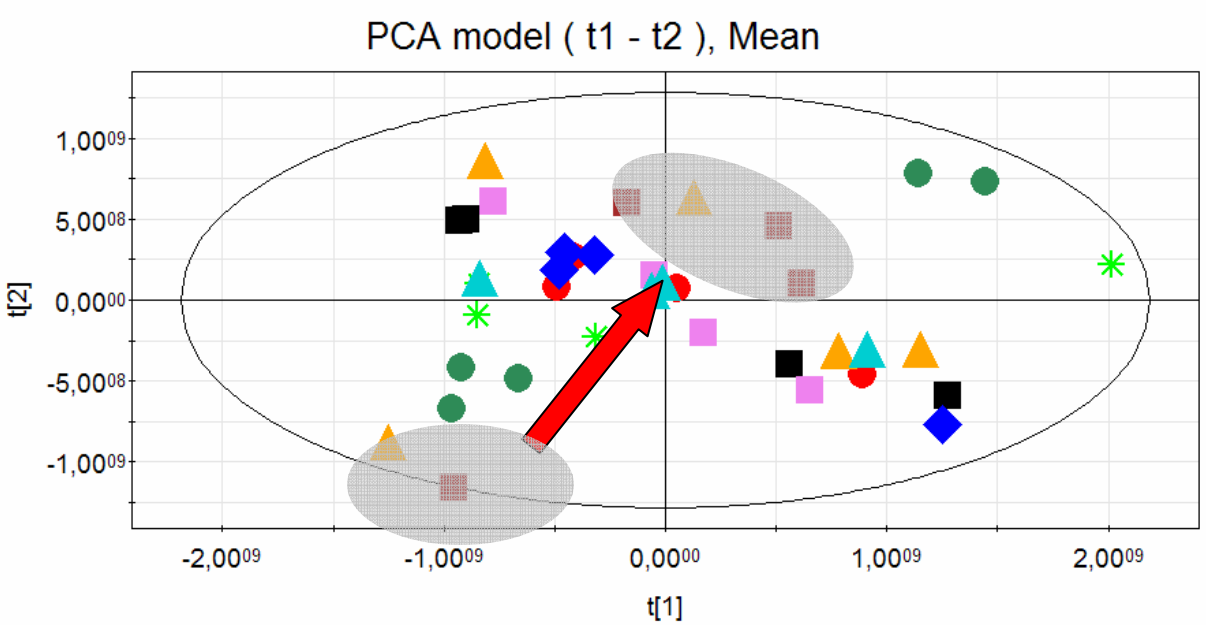
- Assumption 1: The metabolic baseline can be different over all individuals
- Assumption 2: The metabolic effect of the treatment is similar over all individuals



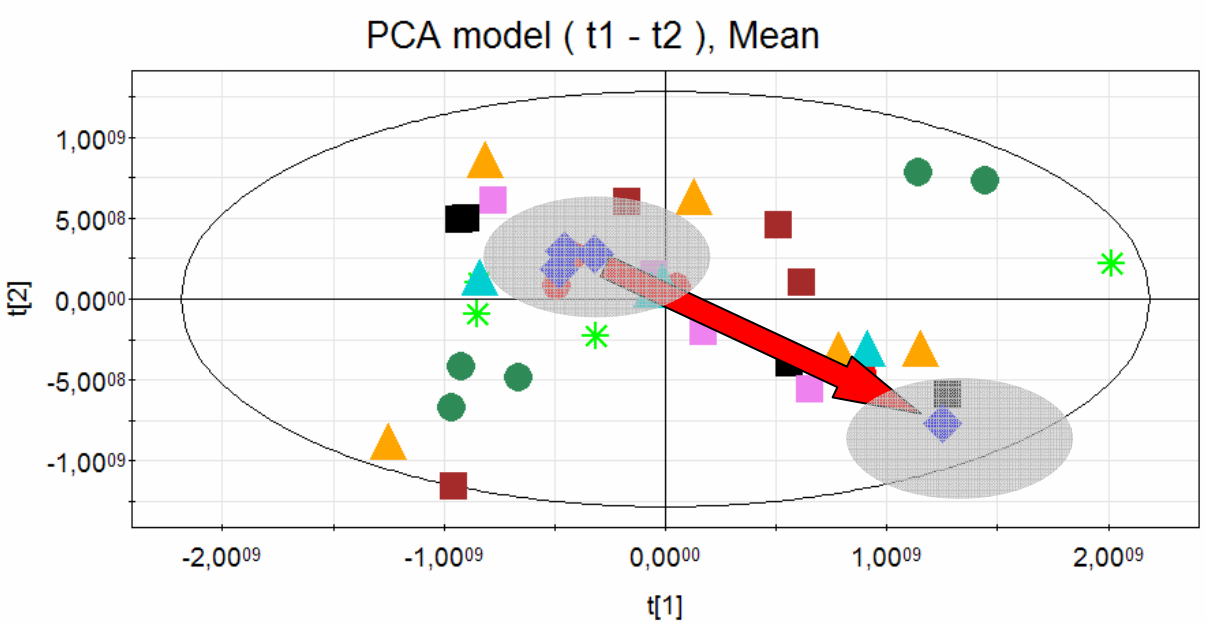
Case study 2: Overview of all individuals/samples … after pre-processing/filtering



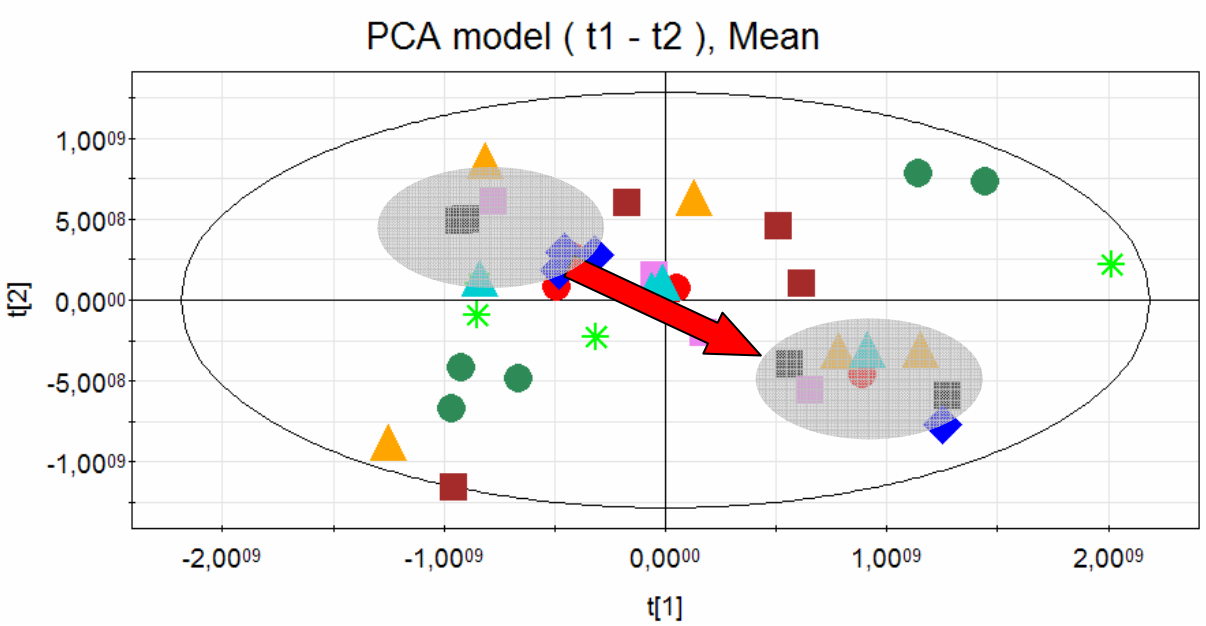
Case study 2: Overview of all individuals/samples … after pre-processing/filtering



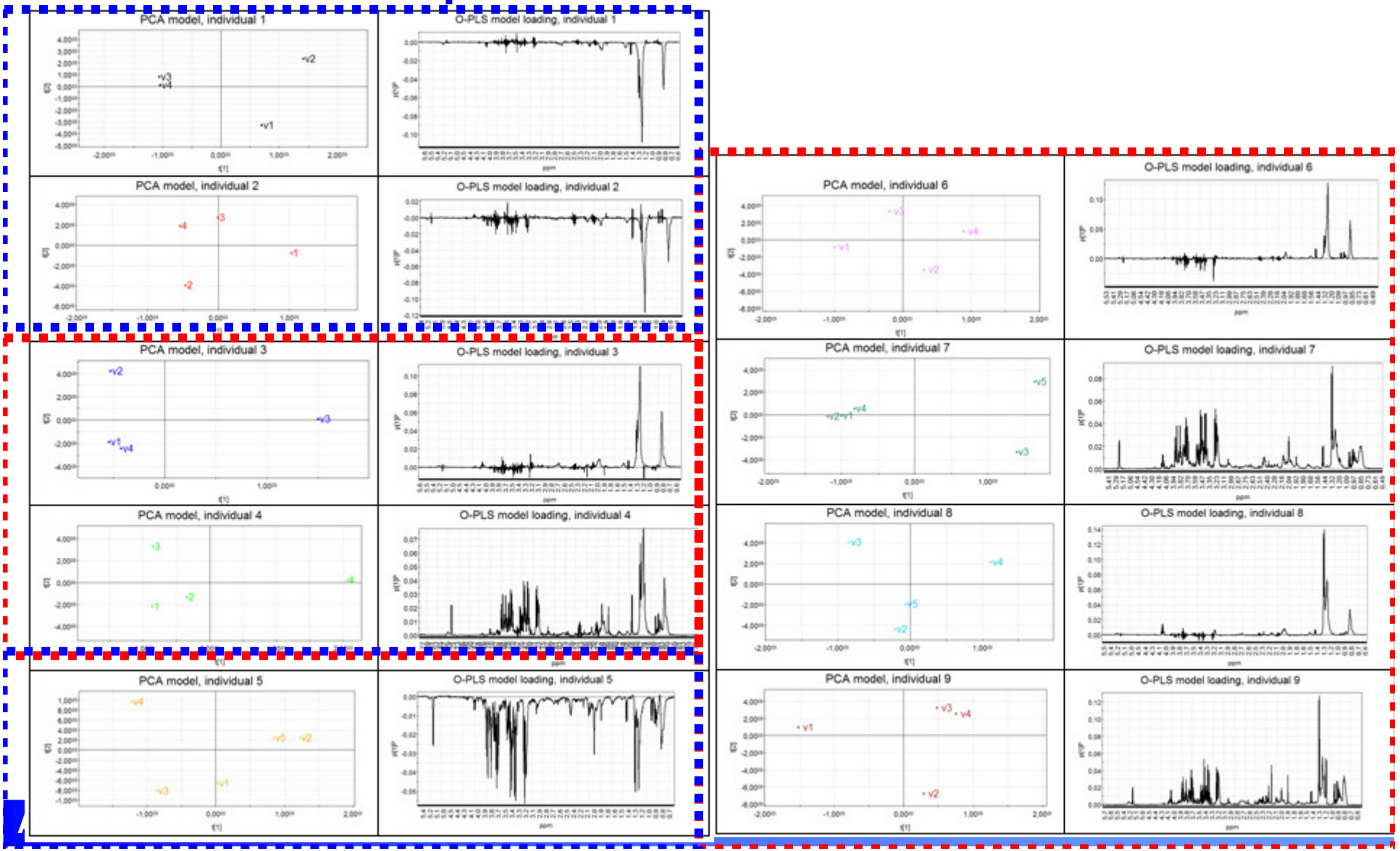
Case study 2: Overview of all individuals/samples … after pre-processing/filtering



Case study 2: Overview of all individuals/samples … after pre-processing/filtering

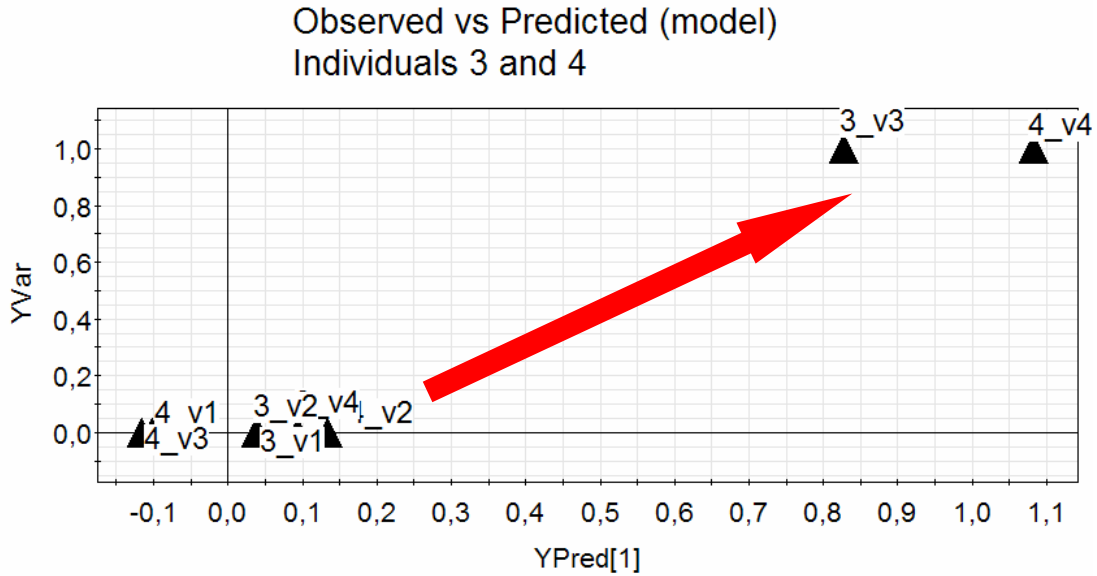


Case study 2: Metabolic profiles for each individual
Assumption: the effect of foodstuff treatment

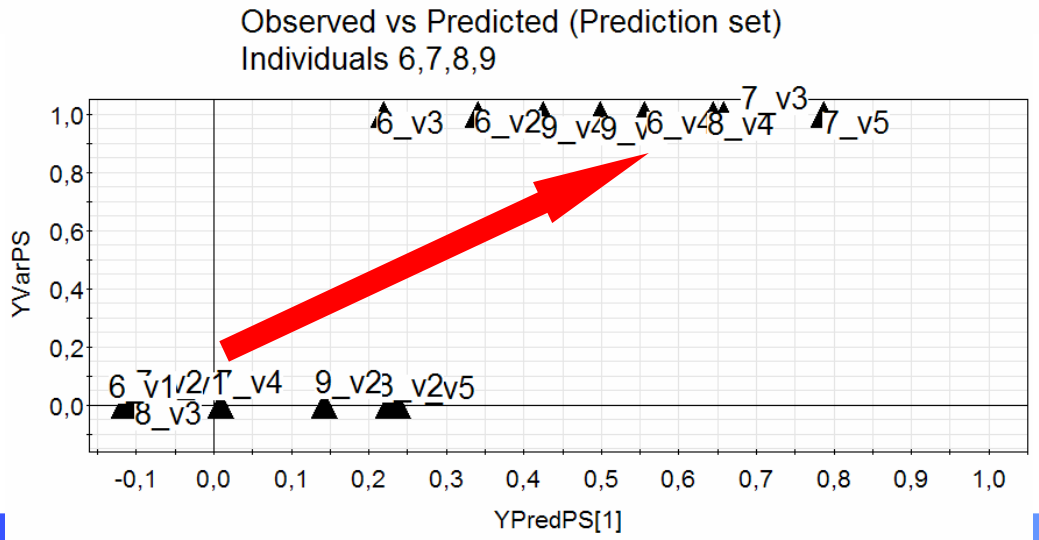


Case study 2: Prediction of effect in NMR profile

Model

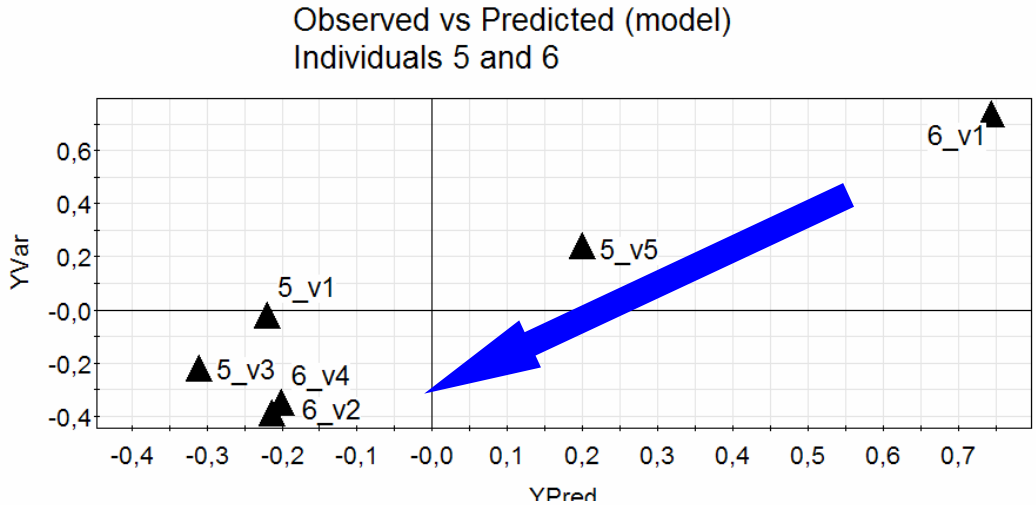


Prediction

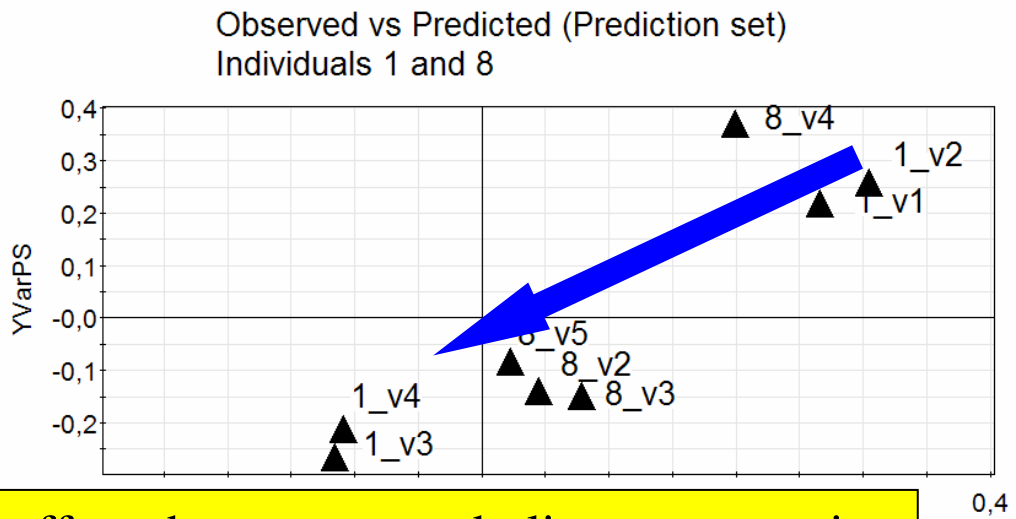


Case study 2: Prediction of health effect in endogenous metabolites (”biomarkers”)

Model

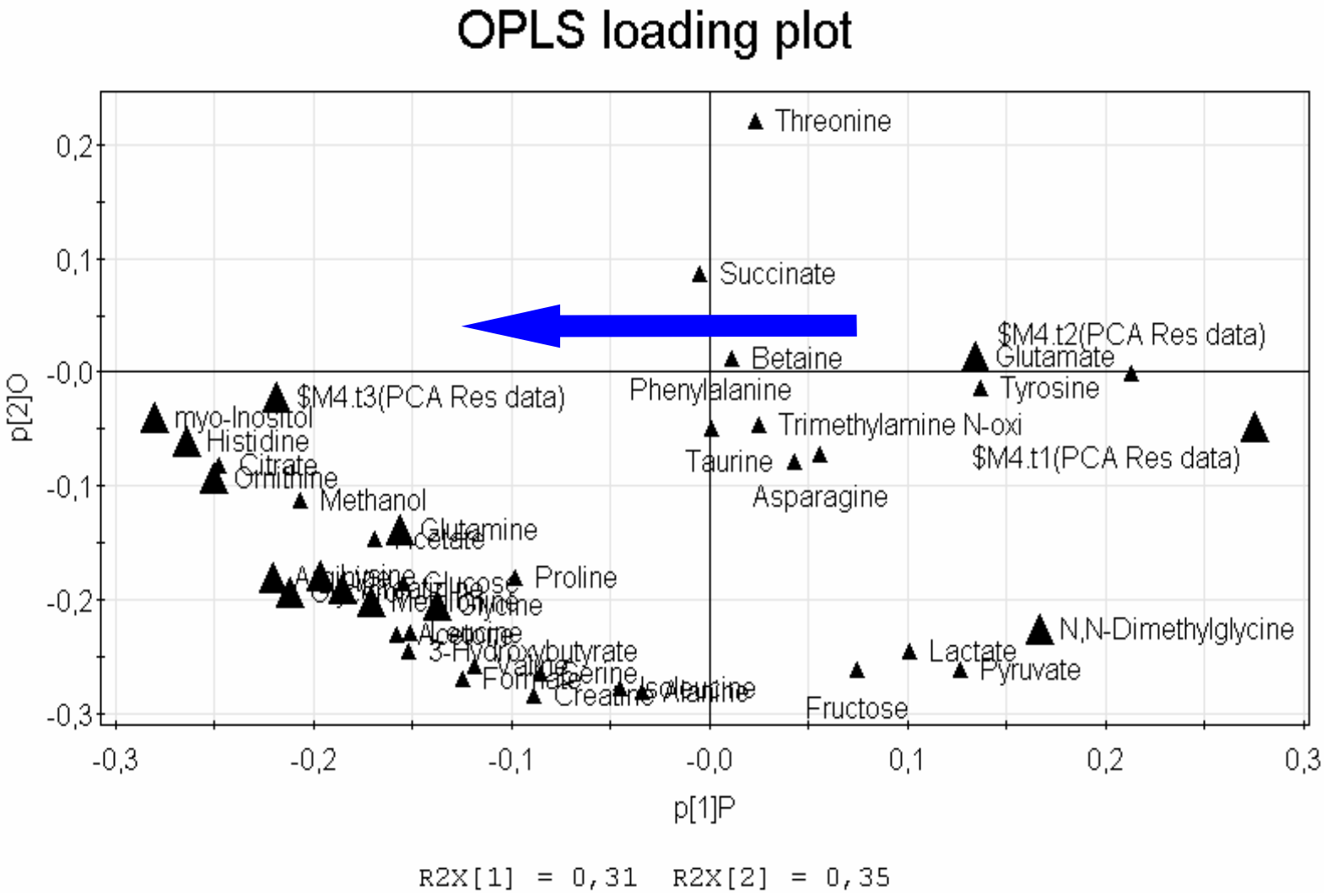


Prediction



Health effect shown, as metabolite concentration *decreases* after consumption of foodstuff

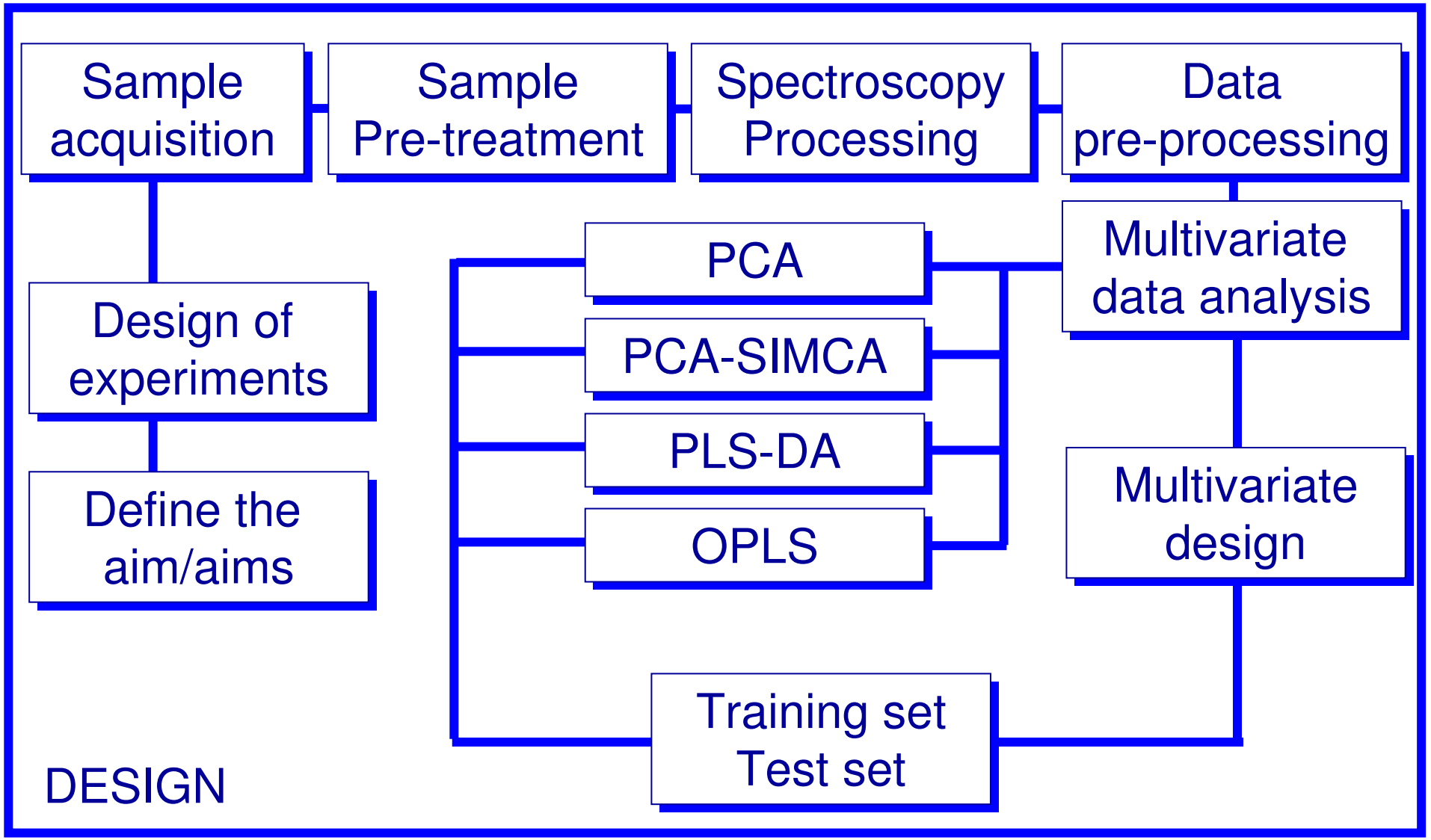
Case study 2: Analysis of health effect in endogenous metabolites ”biomarkers”)



Case study 2: Conclusions

- Effect of food supplement established
  - Increase in Myo-Inositol
  - Decrease in triglycerides
- Opens new possibilities for development of functional foods
  - Proof of physiological health effect
  - Target identification

General analysis strategy



Strategy

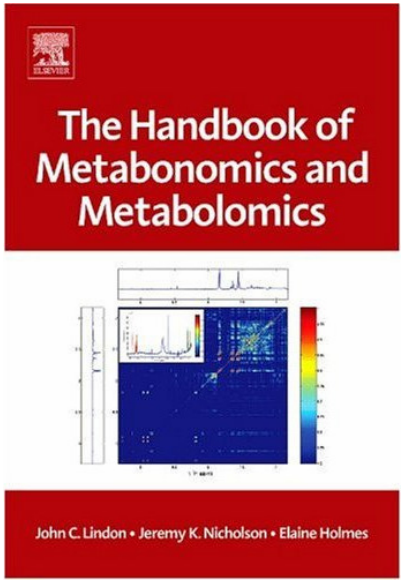
1. Formalize the aim
   - What do we want?

2. Selection of objects
   - Design of Experiments (DOE)
   - Multivariate design (MVD)

3. Sample preparation and profiling of human and animal samples
   - I*n vivo, in vitro* samples
   - Blood, Urine, Cerebral Spinal Fluid (CSF)
   - Synovial fluid (joint), Bowel fluids, Feces, Tissues

4. Integration and evaluation of collected data
   - Exploratory analysis, Interpretation & Visualization
   - Prediction models
   - Patterns
   - Target identification

5. Identify and define IPR opportunities and strategies



Acure
omics

Projects in disease diagnostics	Services CRO
Osteoarthritis (patent application filed)	Biofluid profiling
Rheumatoid arthritis (patent application filed)	Biomarker identification
Fibromyalgia	Toxicity monitoring
Chronic Fatigue Syndrome	
Type 1 Diabetes (patent application filed)	

Summary

- Historical data
  - Always a good starting point for analysis (PCA)
  - Determine data structure, preferred format/output
  - Get acquainted with the process and the data
  - Find “hidden” information
  - Good starting point for discussions
- Determine
  - The aim – is there a defined stop criteria (yield, purity, accepted batches, ID important/”sensitive” variables etc.)
  - Important to define prior to investigation ➔ know when to stop
  - The experimental domain (variables, settings, responses)

Summary

- Design of Experiments (DoE)
  - Simplify analysis
  - Ensure a systematic variation in the investigated experimental domain
  - Small design within the defined limits
  - (Design in historical data)
- Analysis
  - PCA (SIMCA etc.)
  - PLS, PLS-DA
- Next step
  - Aim(s) reached
  - Important variables
  - New optimal experiments

Acknowledgements

- Per Lek, AcurePharma
- Thomas Moritz, UPSC
- Johan Trygg, Umeå University
  - Rasmus Madsen, Umeå University
- Elisabeth Seifert, AcurePharma
- Jon Gabrielsson, AcureOmics
- Johan Olsson, Uppsala University
- Mattias Hedenström, Umeå University
- Katrin Lundstedt-Enkel, Uppsala University