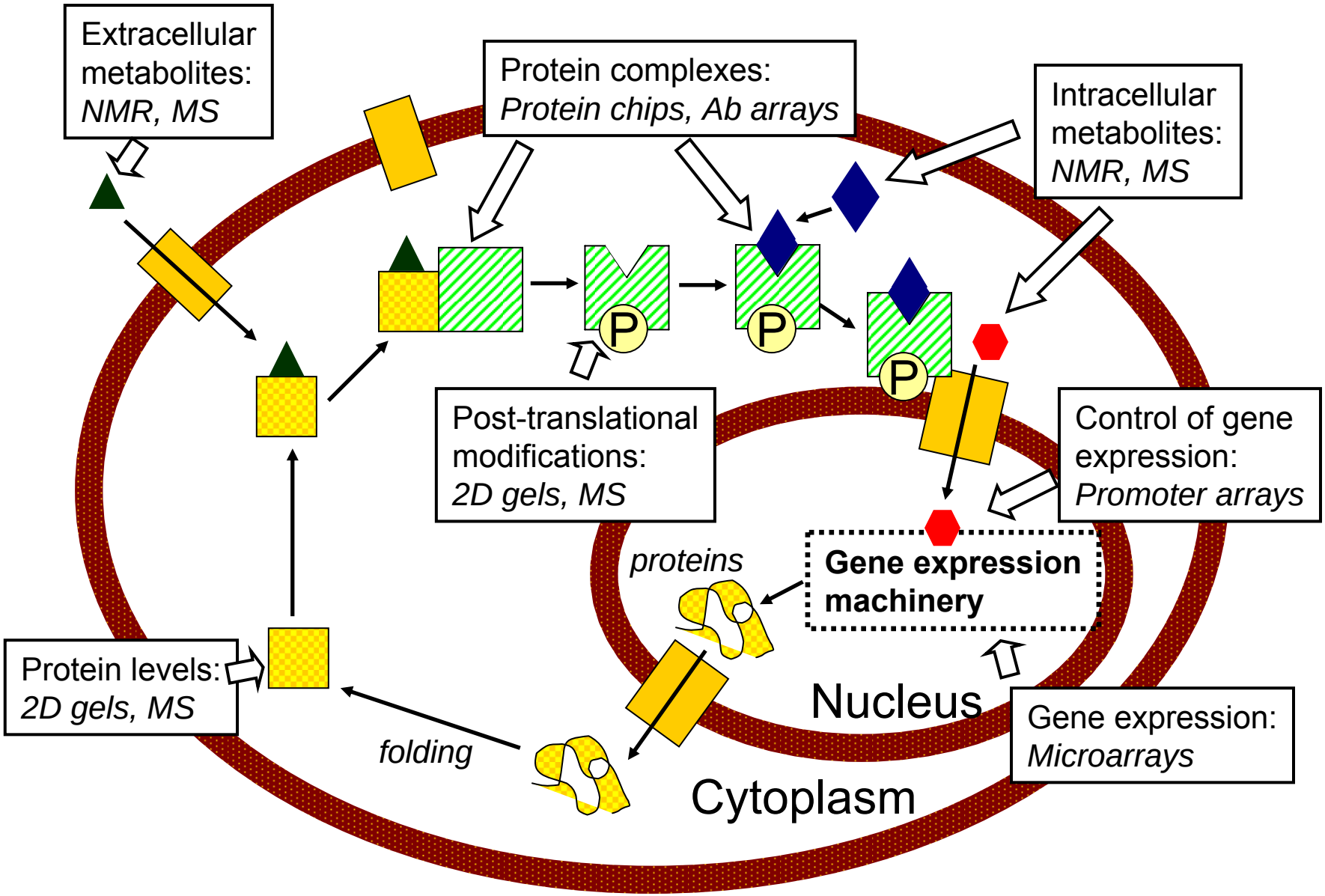
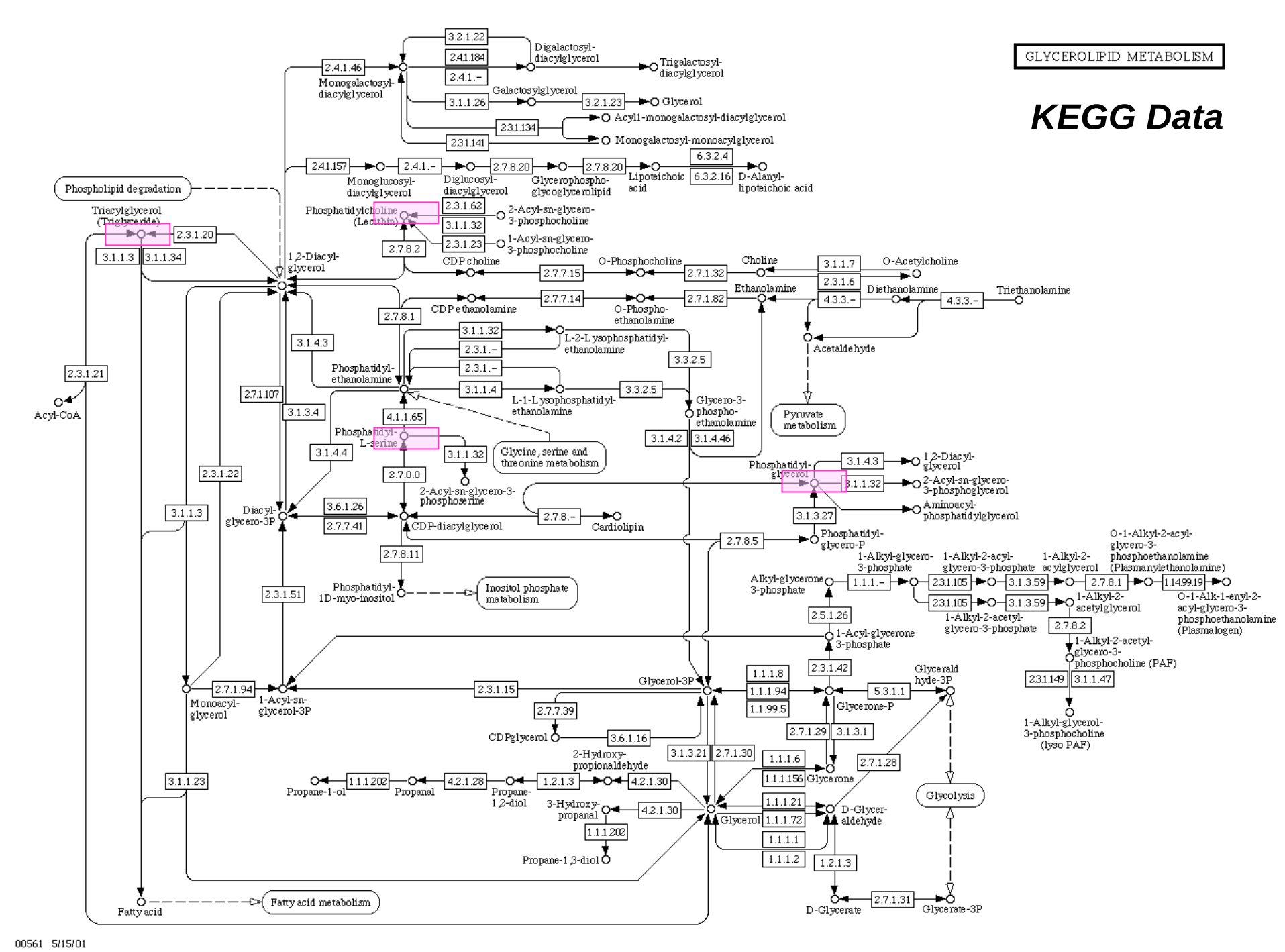
Analysis and interpretation of metabolomics and proteomics data

Matej Orešič

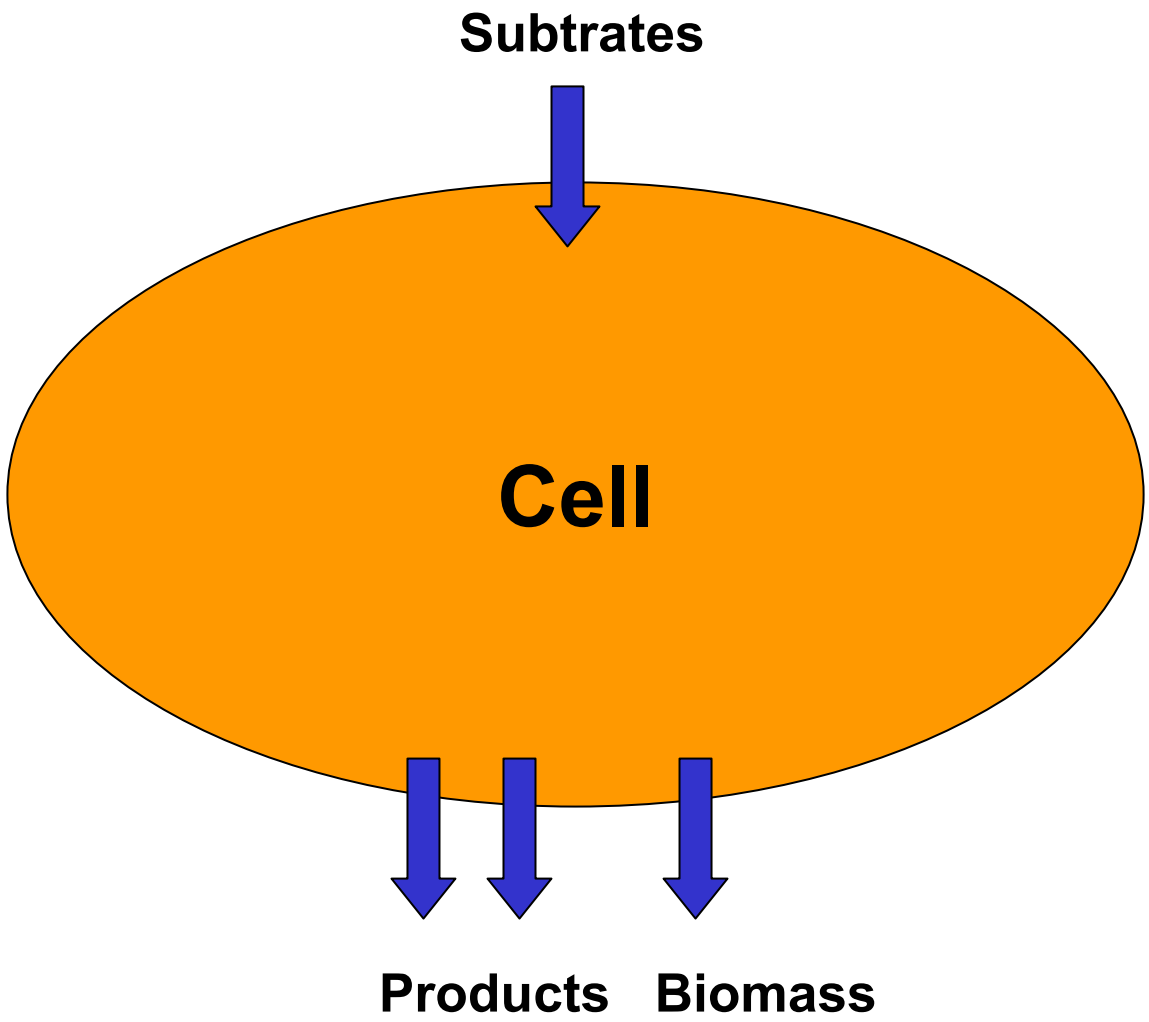
27.9.2005

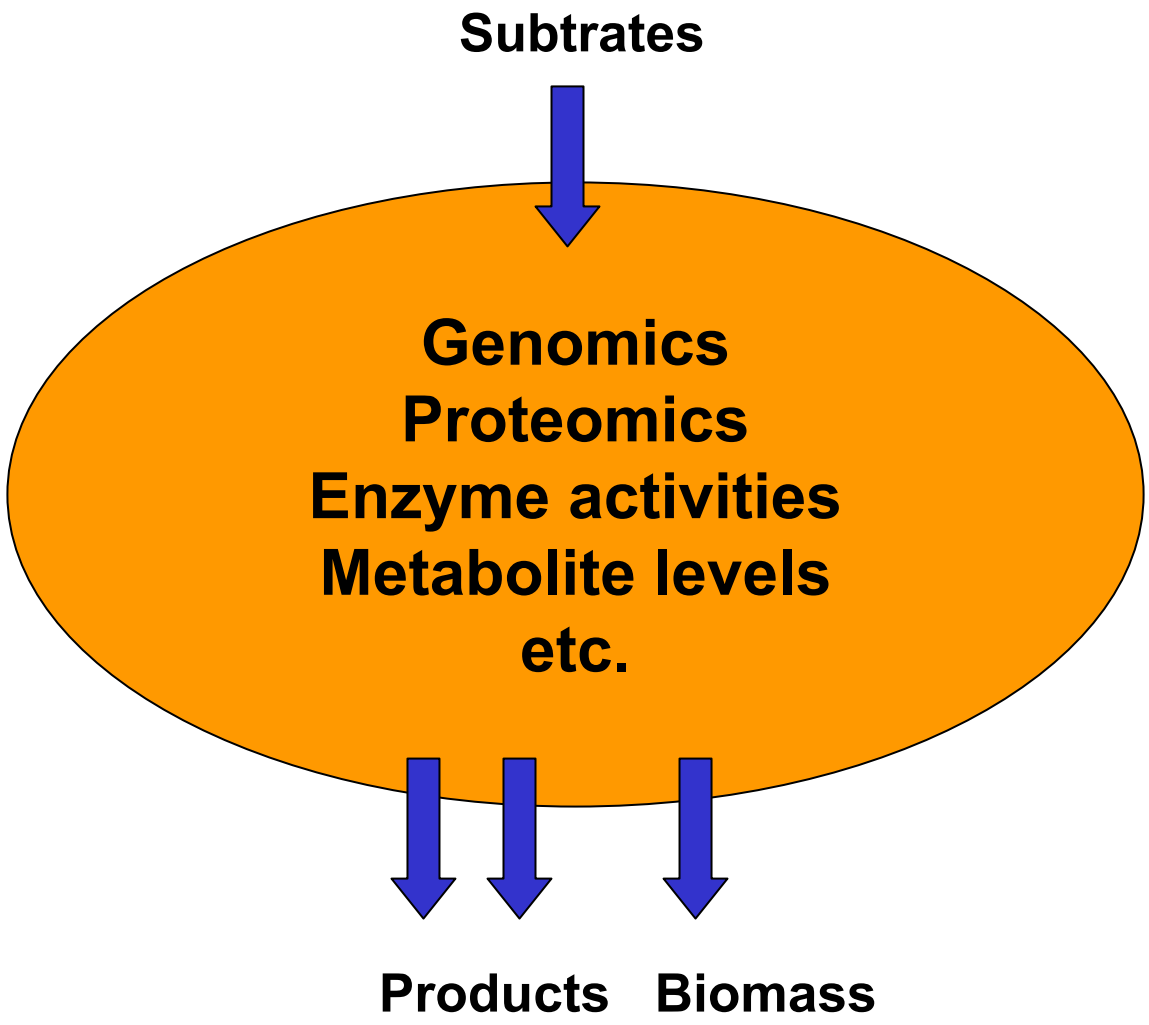
Technologies for systems biology studies at the cellular level

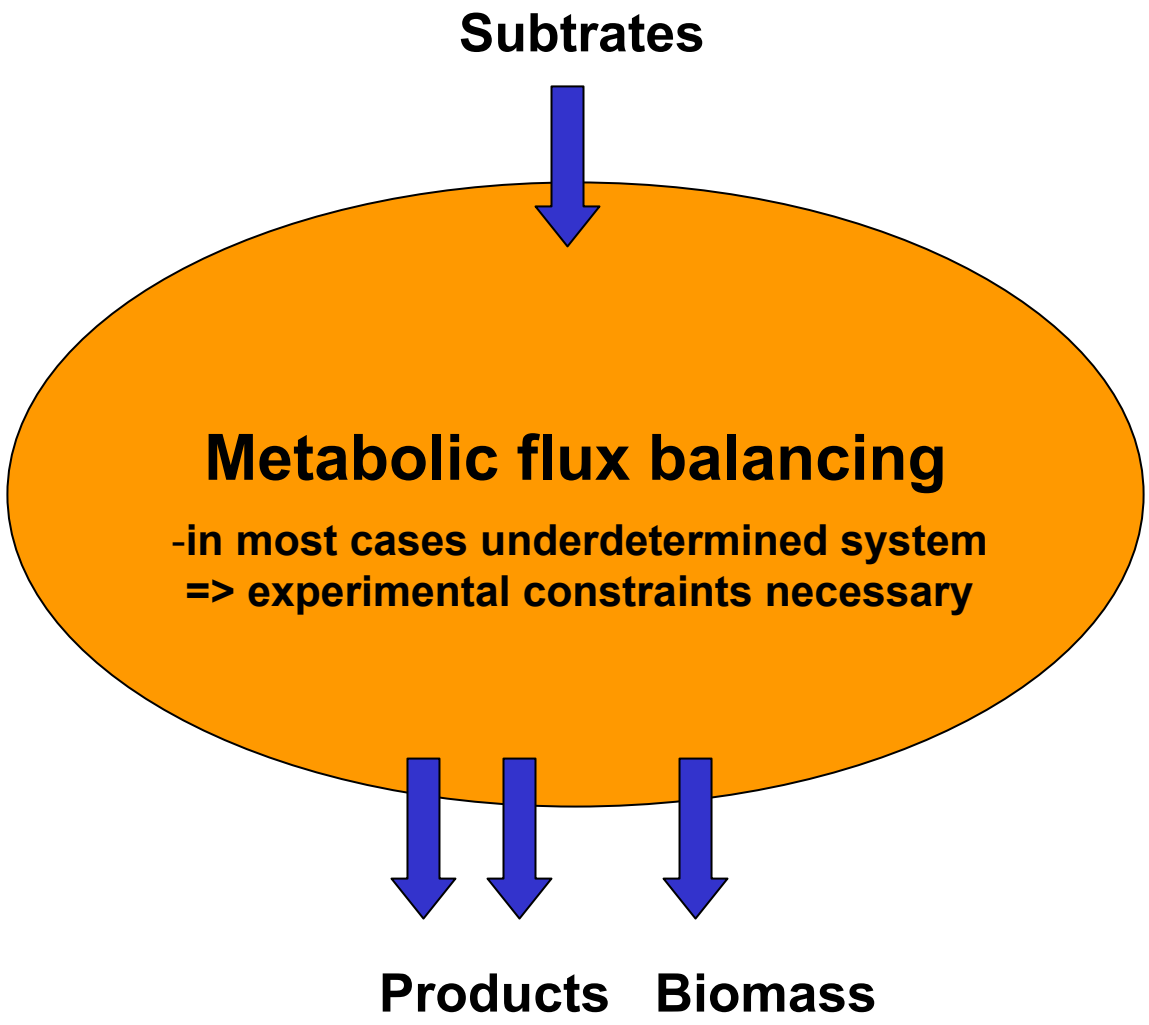



GLYCEROLIPID METABOLISM
KEGG Data
Phospholipid degradation
Triacylglycerol (Triglyceride)
Acyl-CoA
1,2-Diacyl-glycerol
Monoacyl-glycerol
1-Acyl-sn-glycerol-3P
Diacyl-glycero-3P
Fatty acid
Fatty acid metabolism
Monogalactosyl-diacylglycerol
Digalactosyl-diacylglycerol
Trigalactosyl-diacylglycerol
Galactosylglycerol
Glycerol
Acyl1-monogalactosyl-diacylglycerol
Monogalactosyl-monoacylglycerol
Monoglucosyl-diacylglycerol
Diglucosyl-diacylglycerol
Glycerophospho-glycoglycerolipid
Lipoteichoic acid
D-Alanyl-lipoteichoic acid
Phosphatidylcholine (Lecithin)
2-Acyl-sn-glycero-3-phosphocholine
1-Acyl-sn-glycero-3-phosphocholine
CDP-choline
O-Phosphocholine
Choline
O-Acetylcholine
CDP-ethanolamine
O-Phospho-ethanolamine
Ethanolamine
Diethanolamine
Triethanolamine
Acetaldehyde
Pyruvate metabolism
Phosphatidyl-ethanolamine
L-2-Lysophosphatidyl-ethanolamine
L-1-Lysophosphatidyl-ethanolamine
Glycero-3-phospho-ethanolamine
Phosphatidyl-L-serine
Glycine, serine and threonine metabolism
2-Acyl-sn-glycero-3-phosphoserine
CDP-diacylglycerol
Cardiolipin
Phosphatidyl-1D-myo-inositol
Inositol phosphate metabolism
Phosphatidyl-glycerol
1,2-Diacyl-glycerol
2-Acyl-sn-glycero-3-phosphoglycerol
Aminoacyl-phosphatidylglycerol
Phosphatidyl-glycero-P
Alkyl-glycerone 3-phosphate
1-Alkyl-glycero-3-phosphate
1-Alkyl-2-acyl-glycero-3-phosphate
1-Alkyl-2-acylglycerol
O-1-Alkyl-2-acyl-glycero-3-phosphoethanolamine (Plasmanylethanolamine)
O-1-Alk-1-enyl-2-acyl-glycero-3-phosphoethanolamine (Plasmalogen)
1-Alkyl-2-acetylglycerol
1-Alkyl-2-acetyl-glycero-3-phosphate
1-Alkyl-2-acetyl-glycero-3-phosphocholine (PAF)
1-Alkyl-glycerol-3-phosphocholine (lyso PAF)
1-Acyl-glycerone 3-phosphate
Glycerol-3P
Glycerone-P
Glyceraldehyde-3P
CDPglycerol
2-Hydroxy-propionaldehyde
Propane-1-ol
Propanal
Propane-1,2-diol
3-Hydroxy-propanal
Propane-1,3-diol
Glycerone
Glycerol
D-Glycer-aldehyde
Glycolysis
D-Glycerate
Glycerate-3P
00561 5/15/01

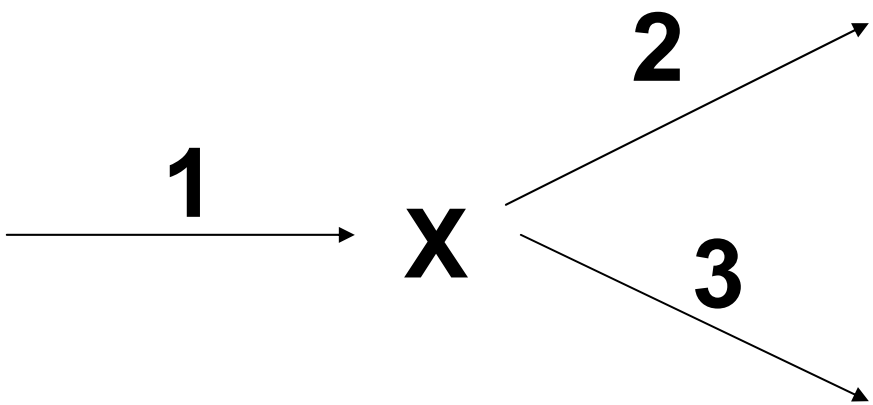
Metabolic Fluxes


Subtrates
Cell
Products
Biomass


Subtrates
Genomics
Proteomics
Enzyme activities
Metabolite levels
etc.
Products
Biomass


Subtrates
Metabolic flux balancing
-in most cases underdetermined system
=> experimental constraints necessary
Products
Biomass

Flux balancing

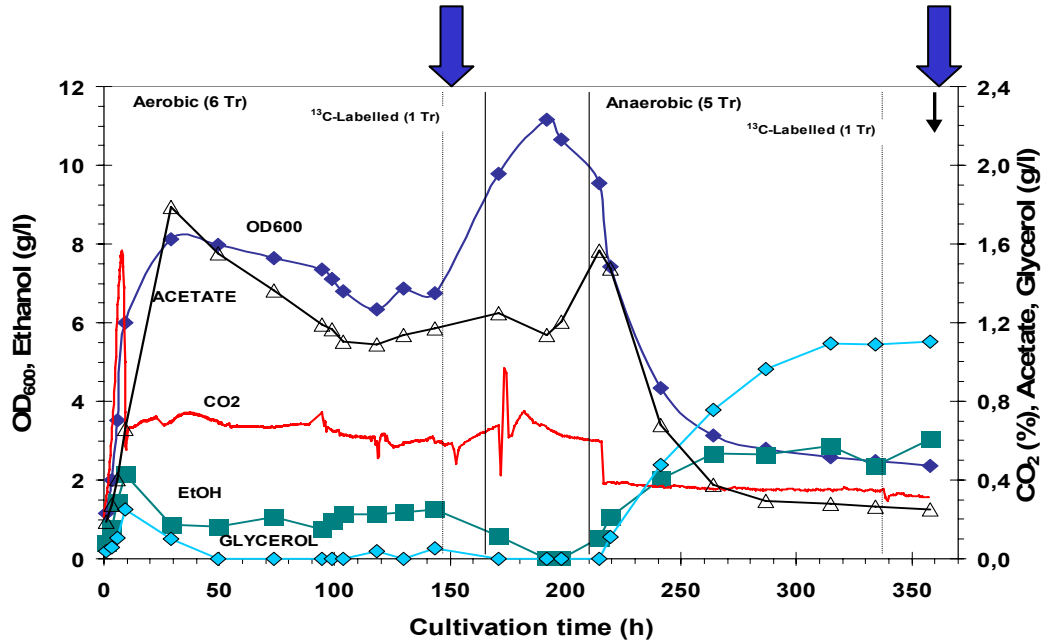


$$v_3+v_2-v_1=0$$

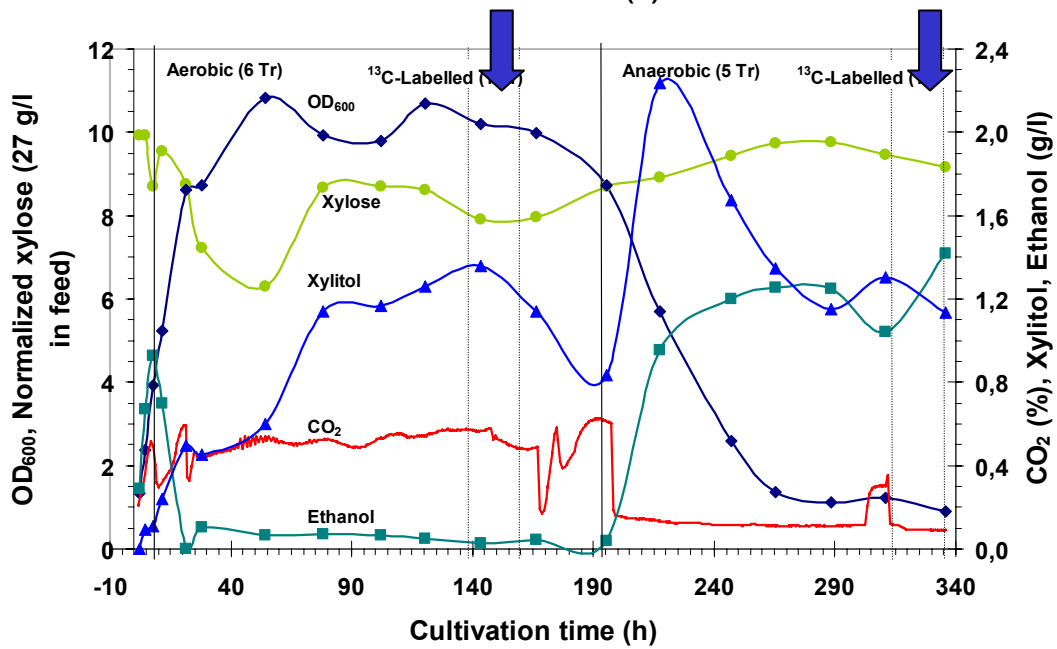
Chemostat cultures

H2490: XR/XDH+XKmc

10g/L Glucose

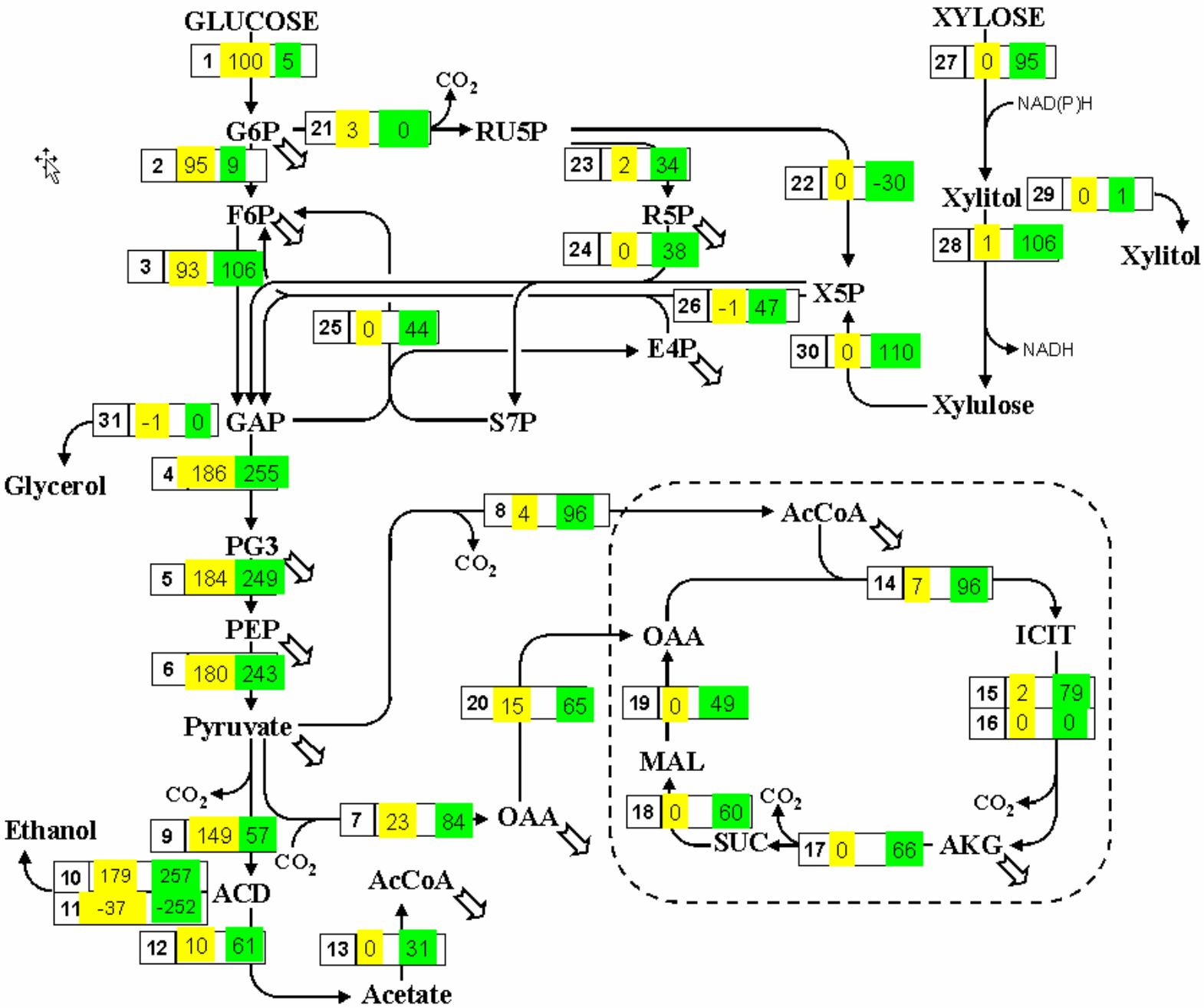


3 g/L glucose
+ 27g/L xylose



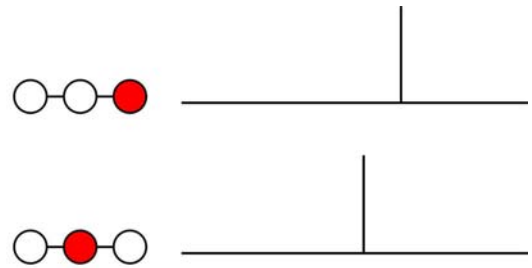
Samples:

- **aerobic culture**
- **anaerobic culture**
- **5, 30, and 60 minutes after the switch off oxygen supply**

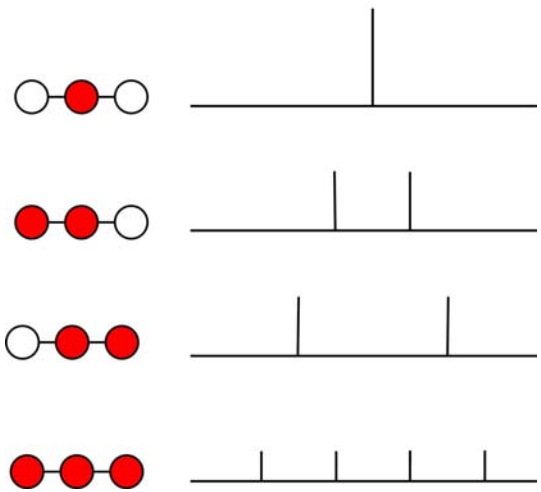

GLUCOSE
1 100 5
G6P
21 3 0
CO2
RU5P
2 95 9
F6P
3 93 106
23 2 34
R5P
24 0 38
22 0 -30
XYLOSE
27 0 95
NAD(P)H
Xylitol
29 0 1
Xylitol
28 1 106
NADH
Xylulose
X5P
26 -1 47
30 0 110
25 0 44
E4P
S7P
31 -1 0
GAP
Glycerol
4 186 255
PG3
5 184 249
PEP
6 180 243
Pyruvate
8 4 96
CO2
AcCoA
14 7 96
OAA
ICIT
20 15 65
19 0 49
15 2 79
16 0 0
MAL
CO2
CO2
18 0 60
SUC
17 0 66
AKG
CO2
Ethanol
9 149 57
7 23 84
OAA
10 179 257
11 -37 -252
ACD
AcCoA
12 10 61
13 0 31
Acetate

Benefits of ^{13}C labeling & NMR

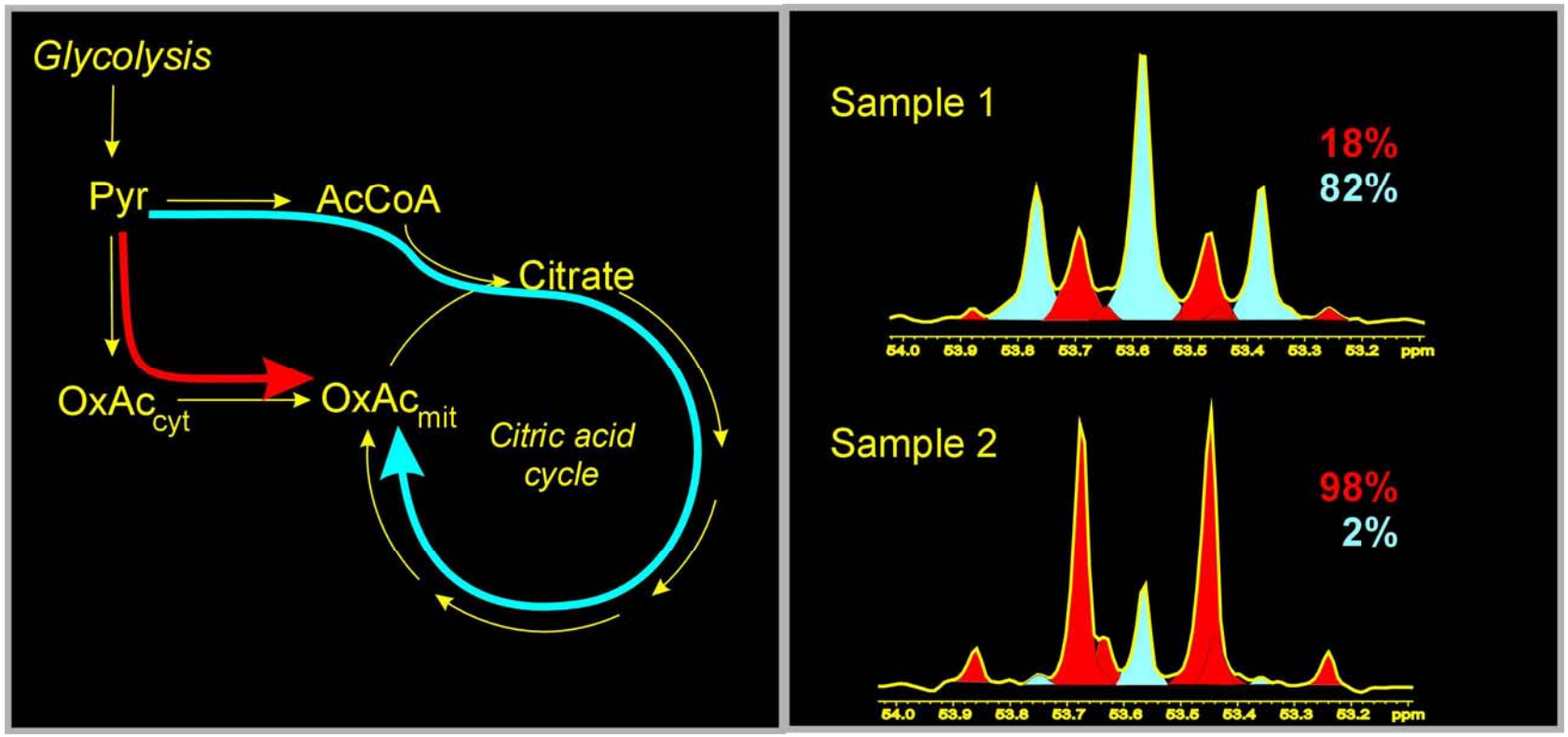
- position sensitive

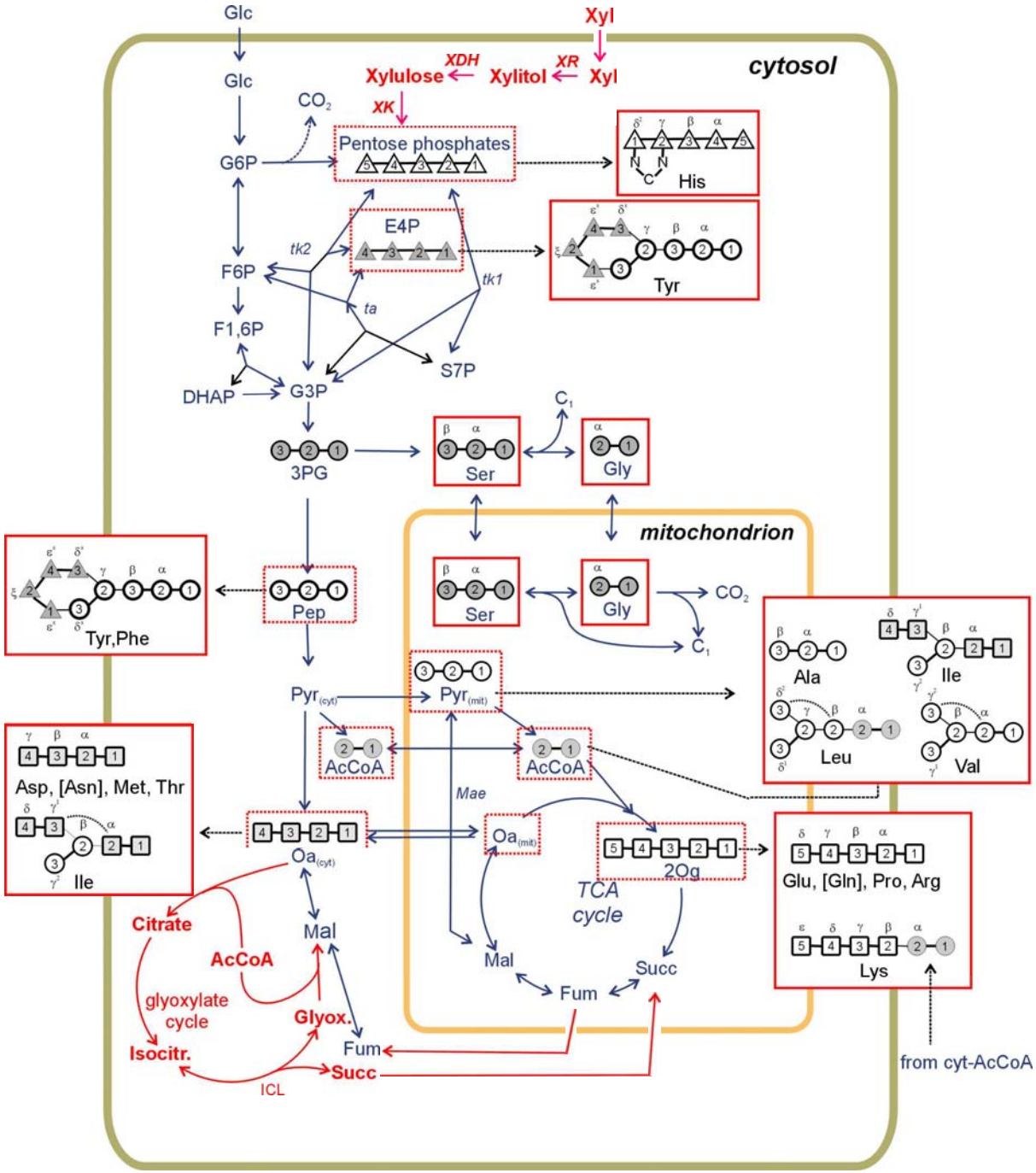


- isotopomer sensitive



METAFoR
(metabolic flux analysis)




Glc
Xyl
cytosol
Glc
XDH
Xylulose
Xylitol
XR
Xyl
CO2
XK
Pentose phosphates
G6P
His
E4P
tk2
F6P
tk1
Tyr
ta
F1,6P
S7P
DHAP
G3P
C1
3PG
Ser
Gly
mitochondrion
Tyr,Phe
Pep
Ser
Gly
CO2
C1
Ala
Ile
Pyr(cyt)
Pyr(mit)
Leu
Val
AcCoA
AcCoA
Mae
Asp, [Asn], Met, Thr
Oa(cyt)
Oa(mit)
2Og
Glu, [Gln], Pro, Arg
Ile
TCA cycle
Citrate
Mal
Mal
AcCoA
Succ
glyoxylate cycle
Glyox.
Fum
Fum
Isocitr.
Succ
Lys
ICL
from cyt-AcCoA

Mass spectrometry

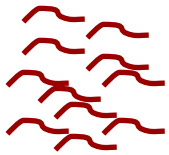
LC/MS proteomics platform and data processing

Sample

Proteins, purified and separated

digest

labelling



Peptides

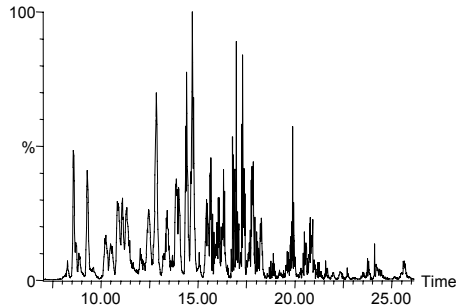
Separation

RPLC/CapLC

+ESI

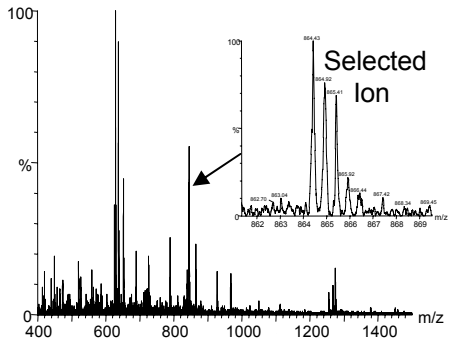
Mass spectrometer (Qtof)

MS-only ion chromatogram



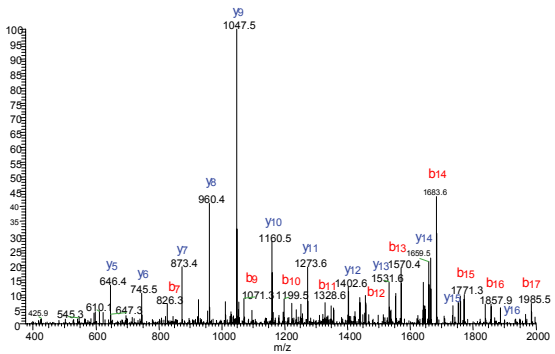
Retention time

MS survey scan



Mass / Charge

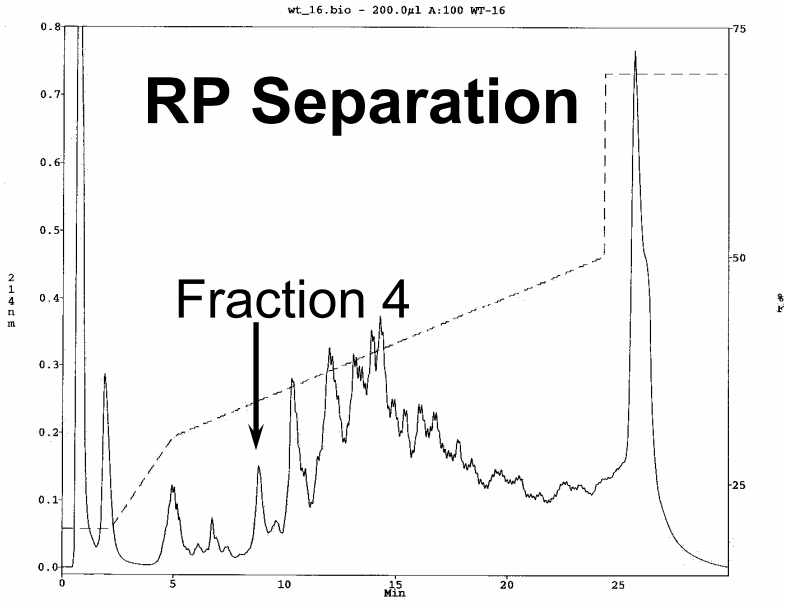
MS/MS fragmentation pattern

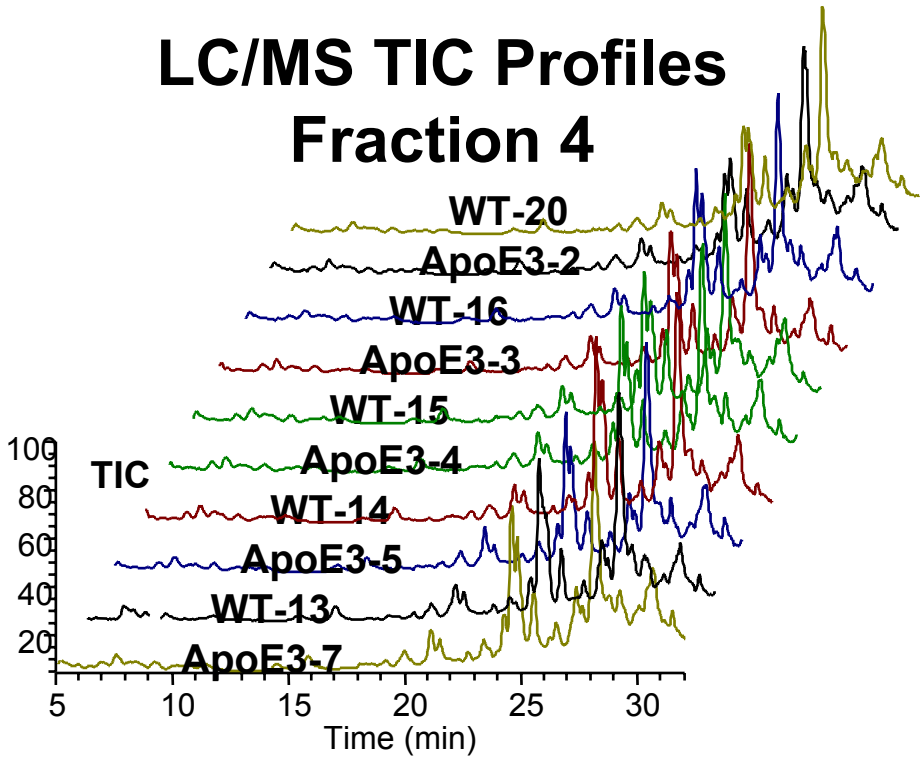


Mass / Charge

Liver Protein Profiling

Fractionation using Reversed Phase Chromatography

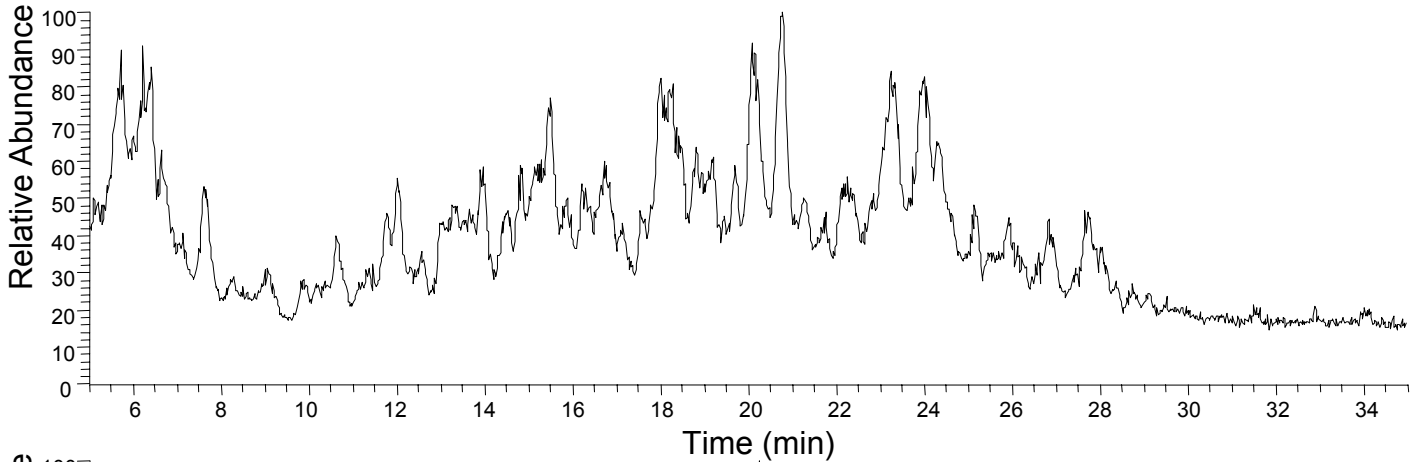




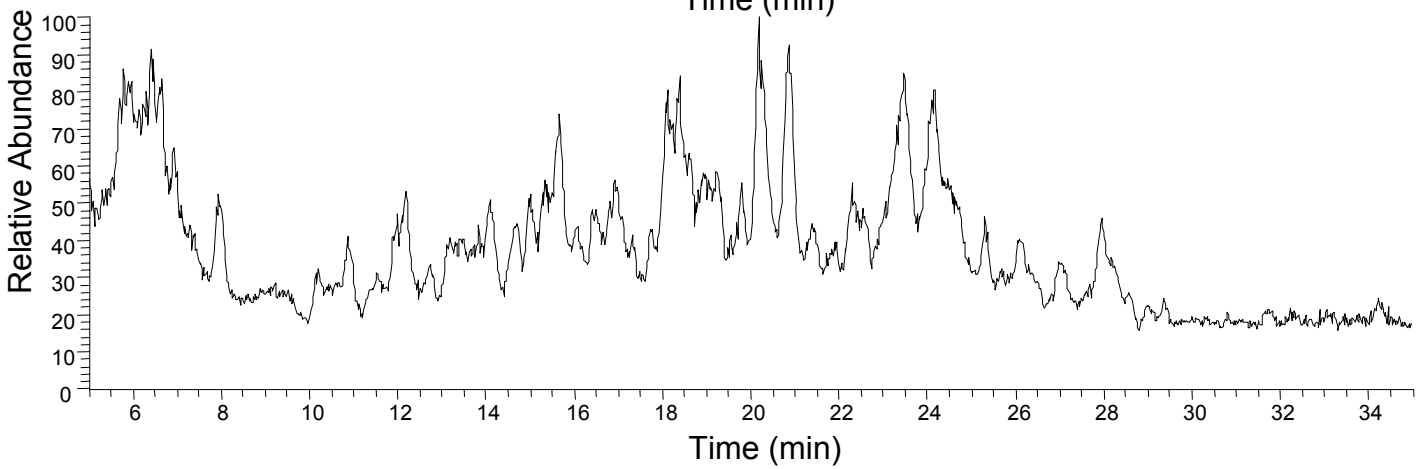
Plasma Protein Profiling

LC/MS of digested SEC fraction

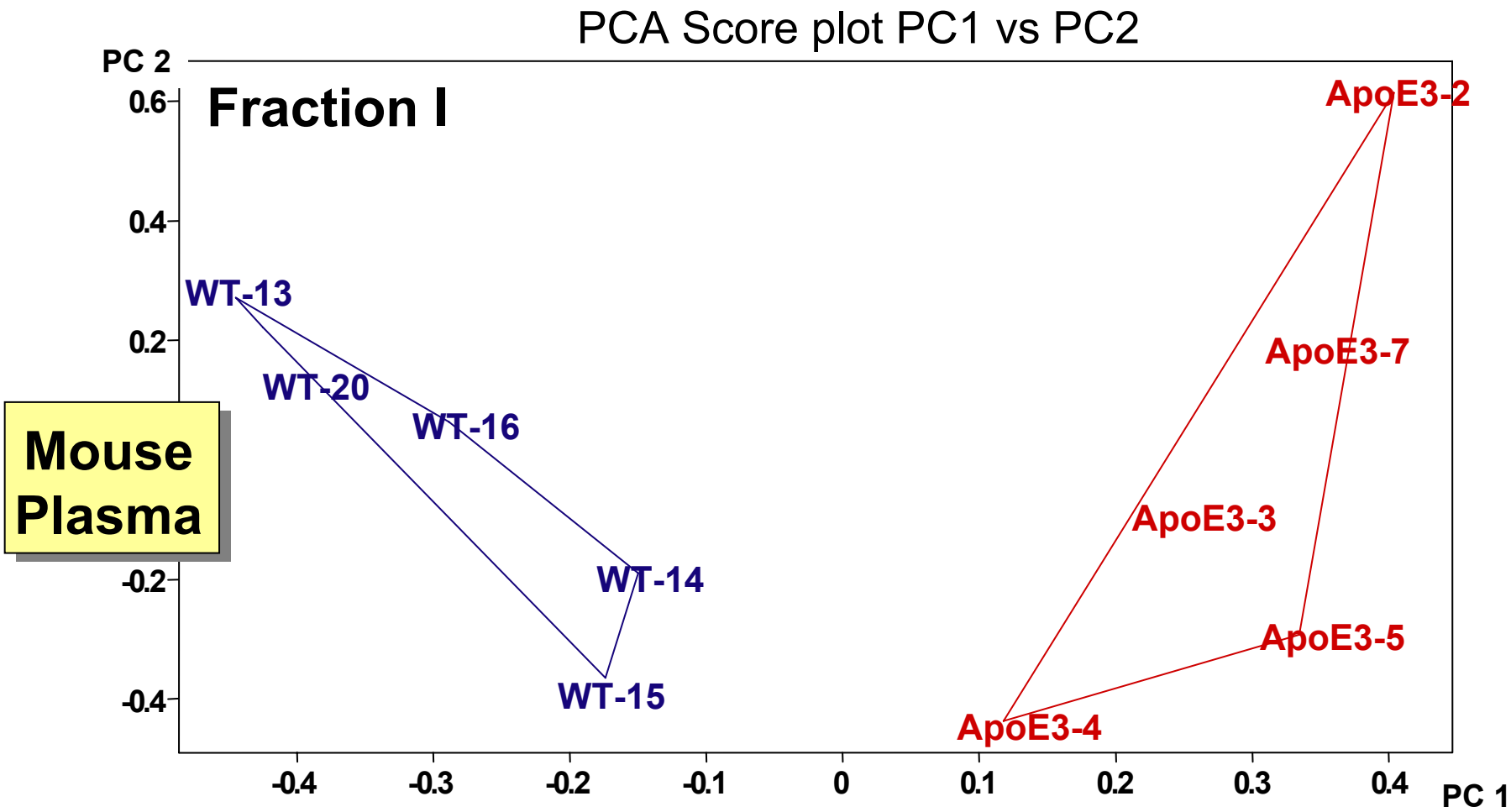
ApoE3



Wildtype



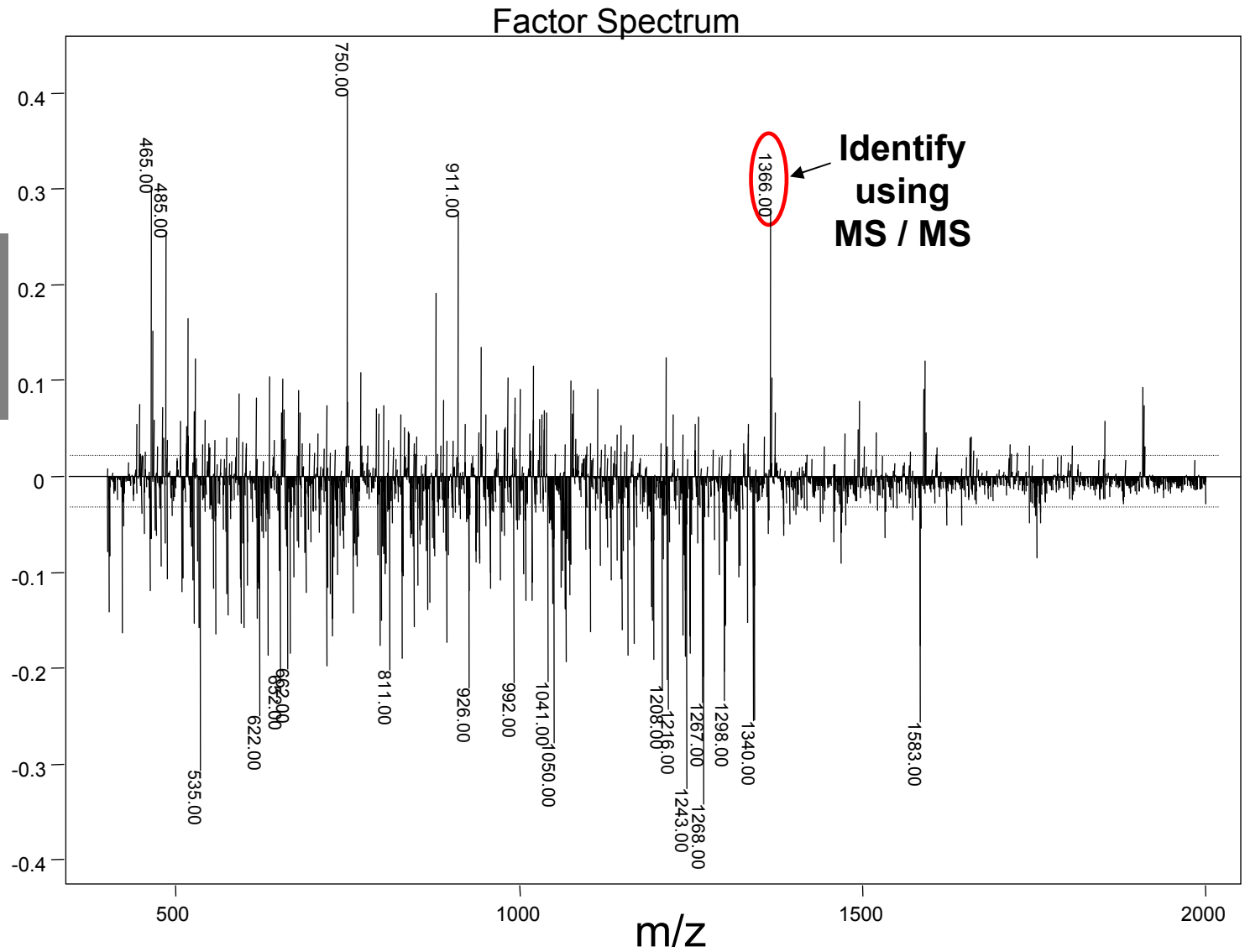
Plasma Protein Profiling
Principal Component Analysis: Fraction I



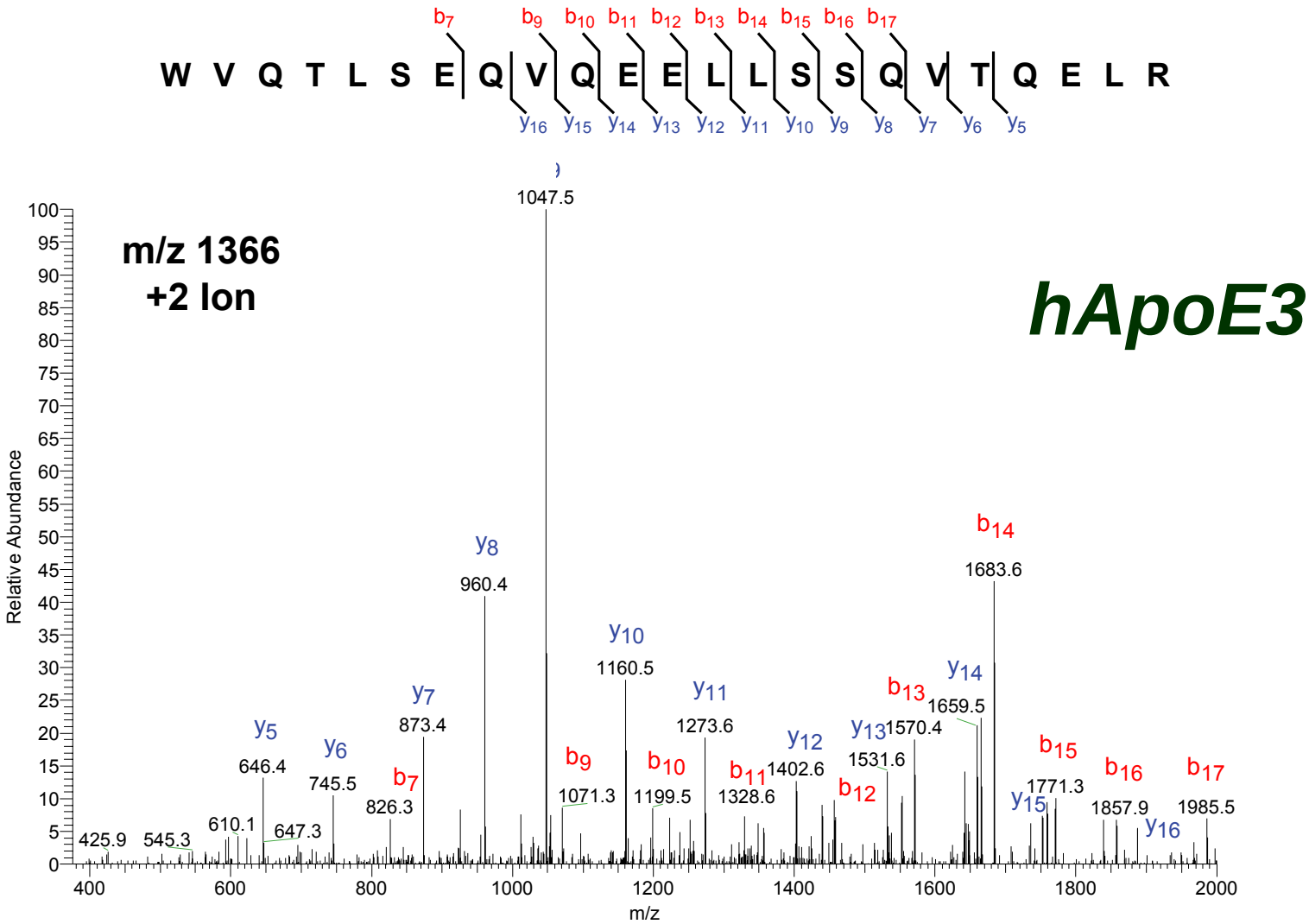
Unsupervised clustering reveals differences at 9 week age

Plasma Protein Profiling
Factor Spectrum: Peptides Exhibiting Differences

Mouse Plasma Fraction II



Peptide Sequencing using MS/MS



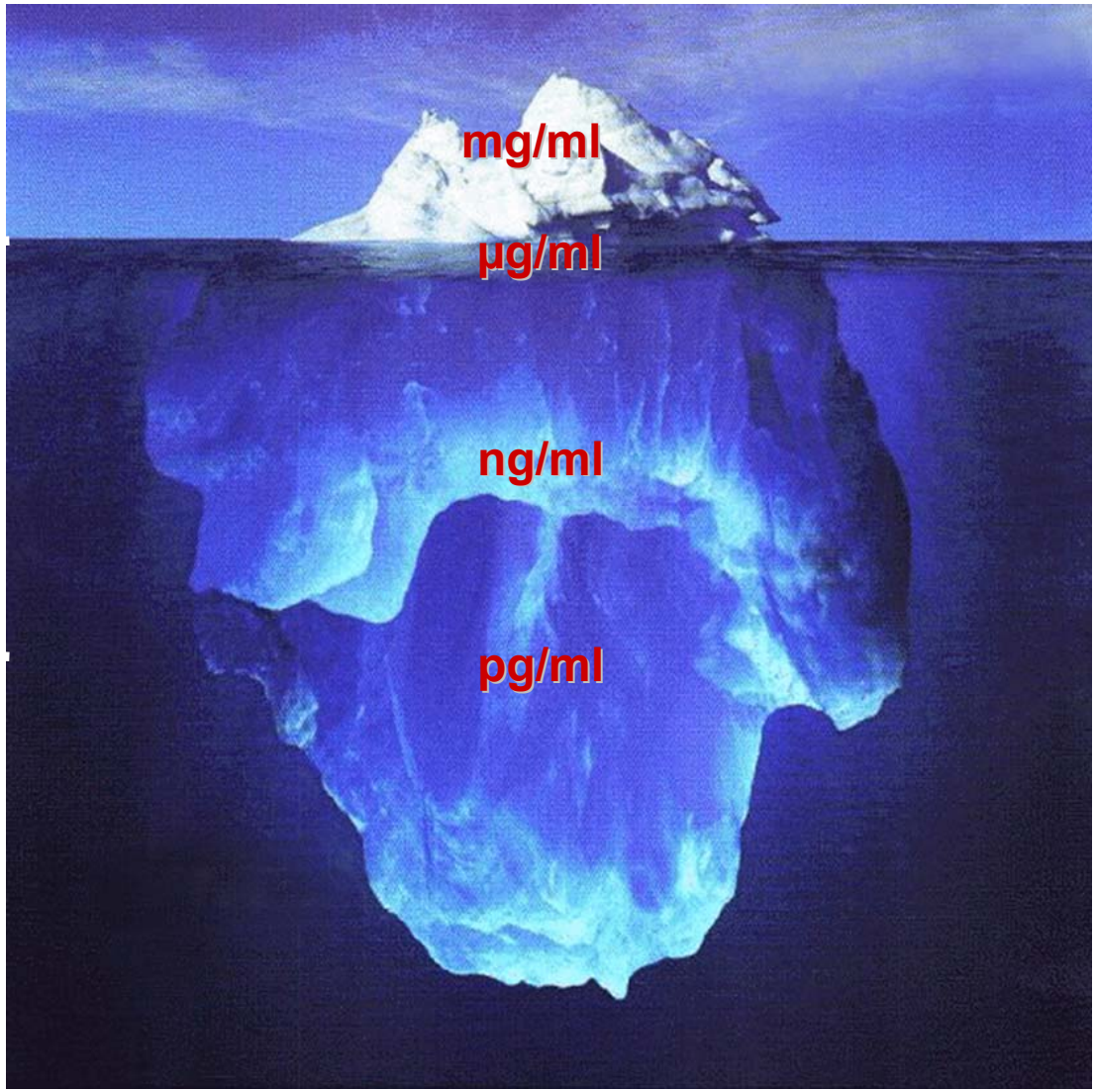
Metabolomics

Study of small molecules , *or metabolites, contained in a cell, tissue or organ (including fluids) and involved in primary and intermediary metabolism.*

NMR

Mass Spectrometry

CUSTOM



Homeostasis 'Housekeeping'

Organic Acids

Lipids

Amino Acids

Nucleotides

Steroids

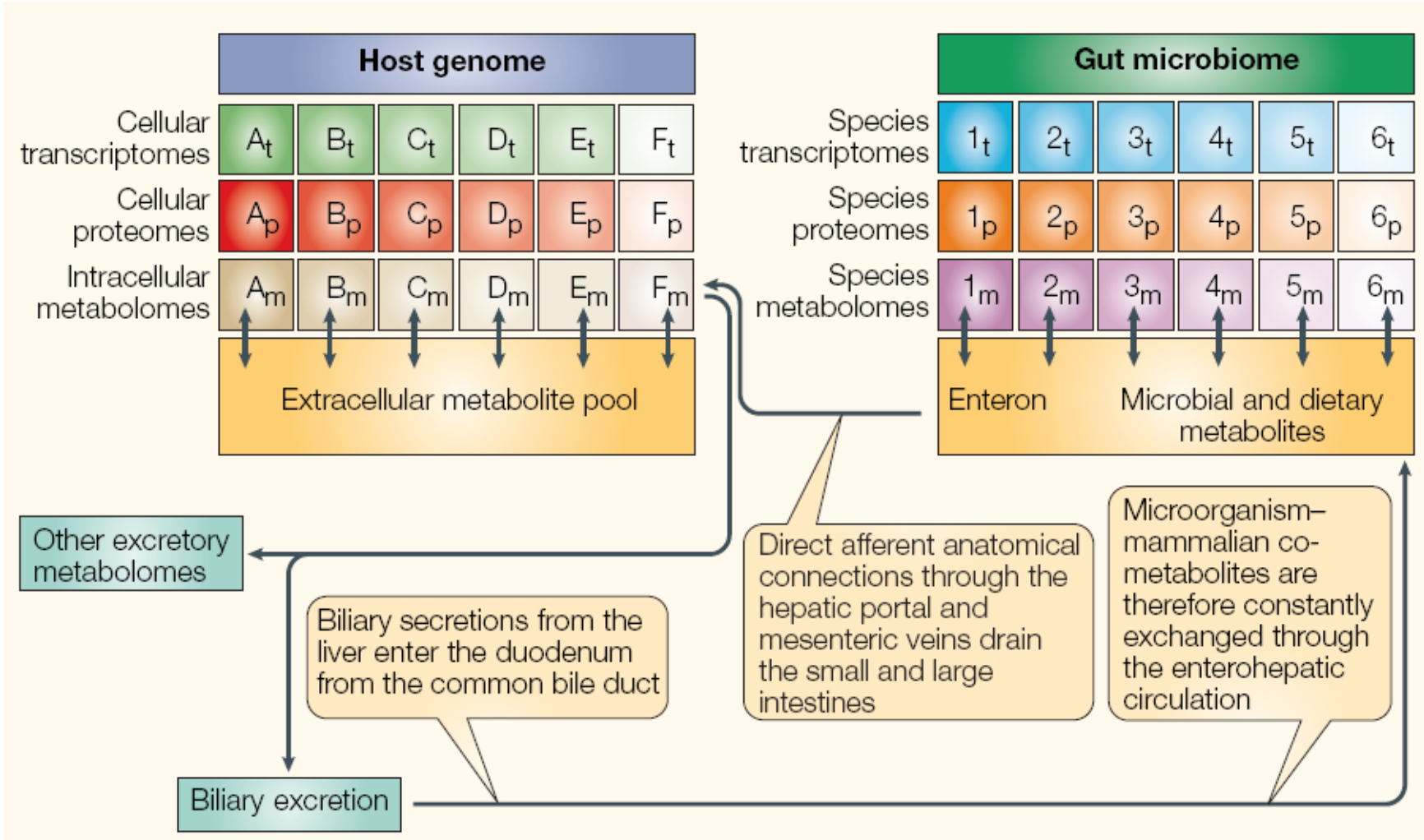
Eicosanoids

Neurotransmitters

Peptides

Trace elements

We are not alone genomewise …



From Nicholson et al., Nature Reviews Microbiology (2005)

Historical note

1500-2000BC
China
•Ants used to detect patients with diabetes

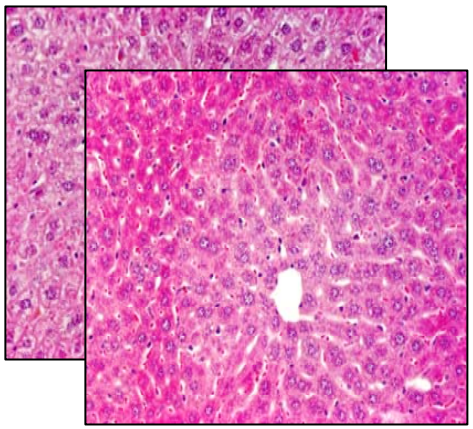
1940s-1970s
•Advances in analytics
•Pattern recognition
→ Metabolic profiling

21st century
•Advances in analytics
•Biostatistics & Bioinformatics
→ Modern era of metabolomics and systems biology

Modern metabolomics platform

Experiment design + Analytical chemistry + Chemometrics + Bioinformatics

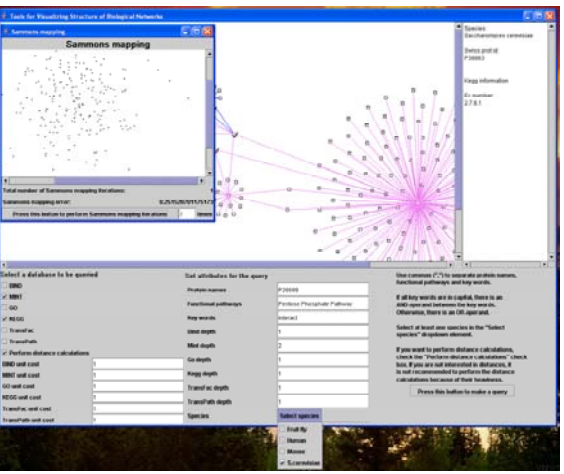
Samples



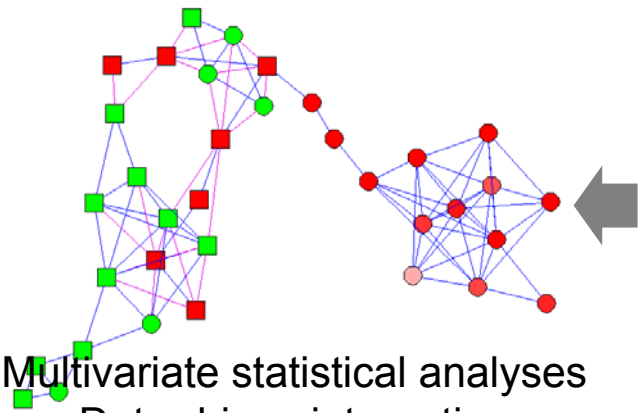
Metabolite extraction methods and analytical platforms

- Global screening (H_2O soluble)
- Global screening (Lipids)
- Primary metabolites
  Eicosanoids and fatty acids
- Other

Profiling experiments (LC/MS, GC/MS)

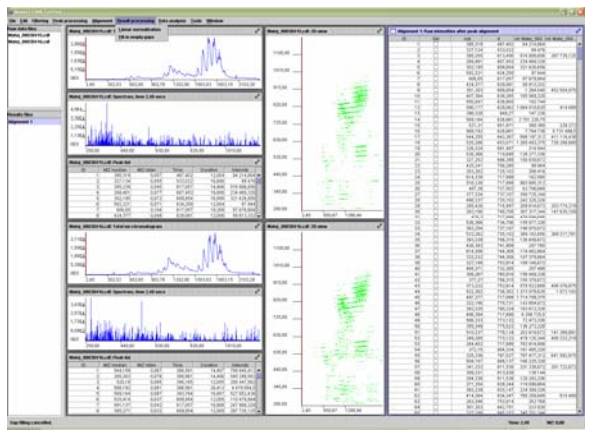


Bio-/chemo-informatics knowledge mining



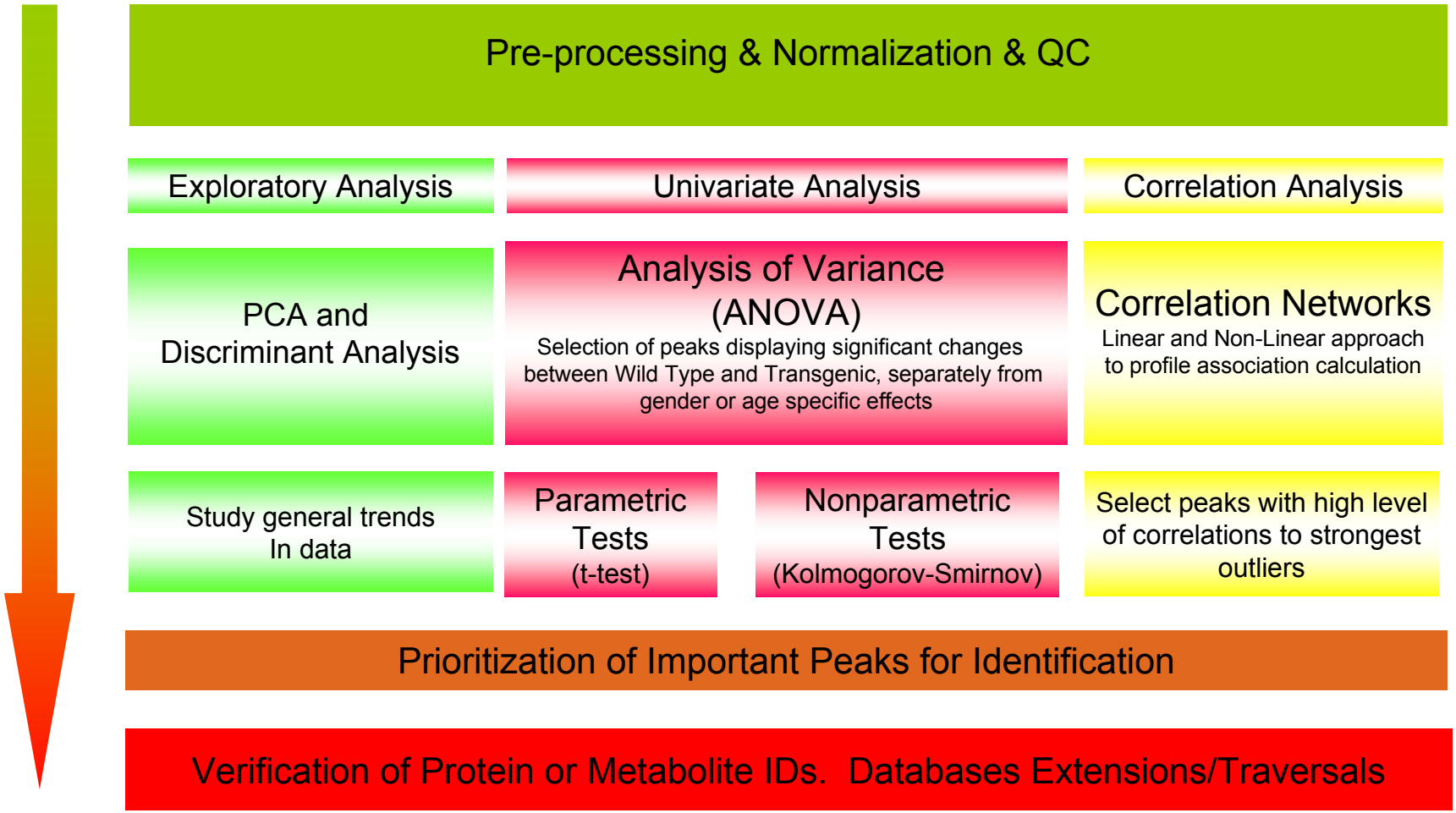
Multivariate statistical analyses
Data-driven integration

Biological insight

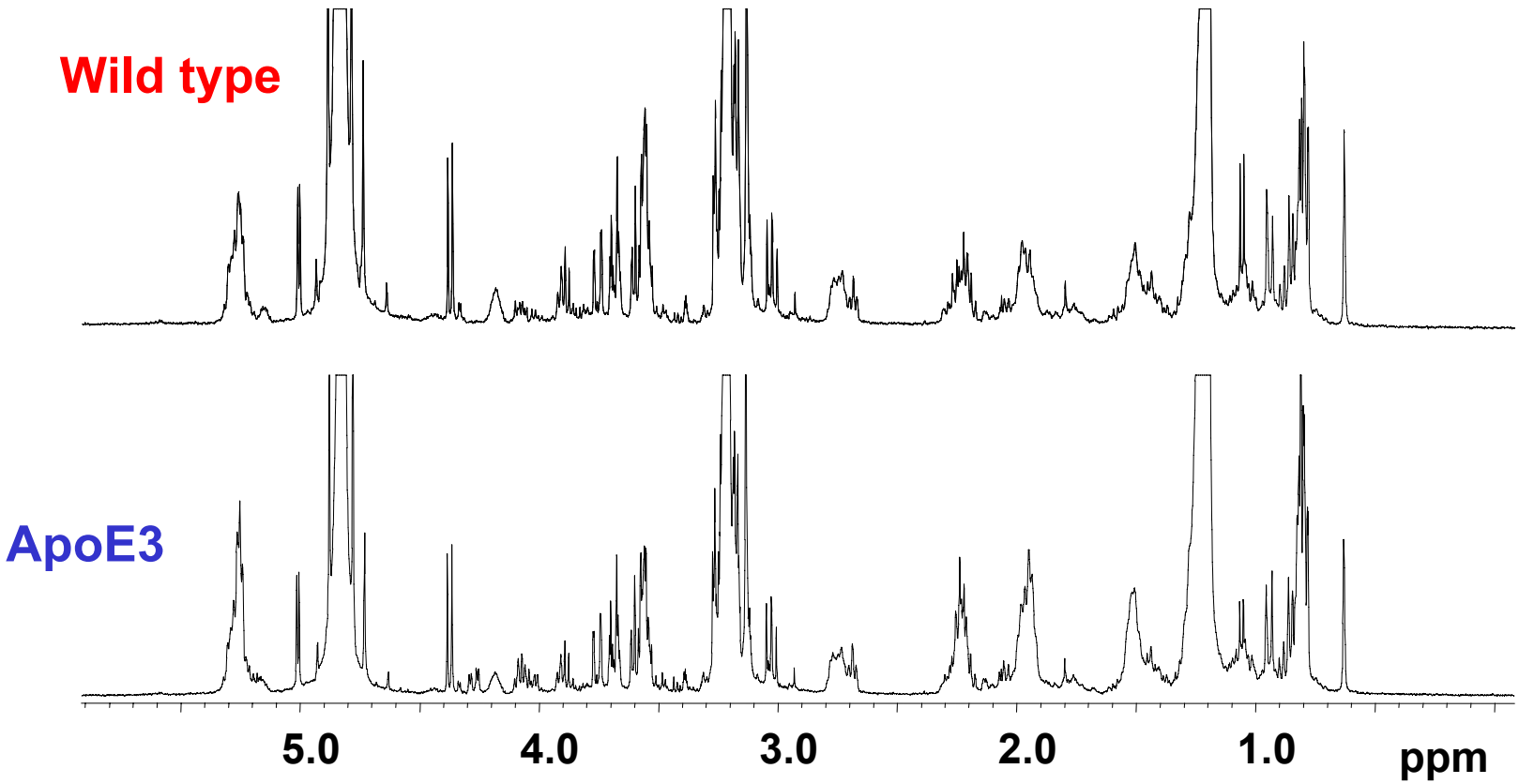


Data processing
Identification

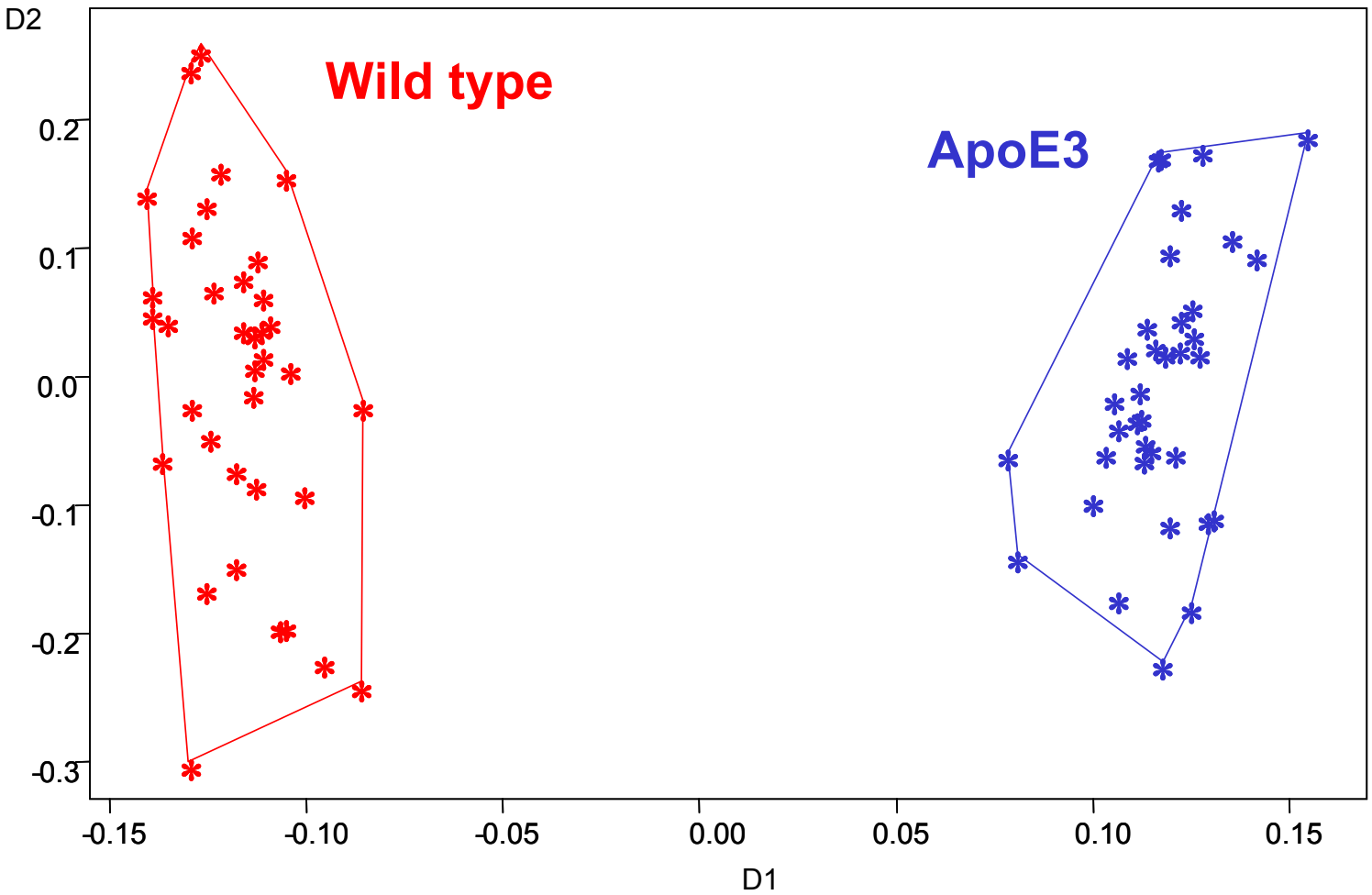
Data processing



Global Metabolite Analysis
NMR Spectra of Plasma



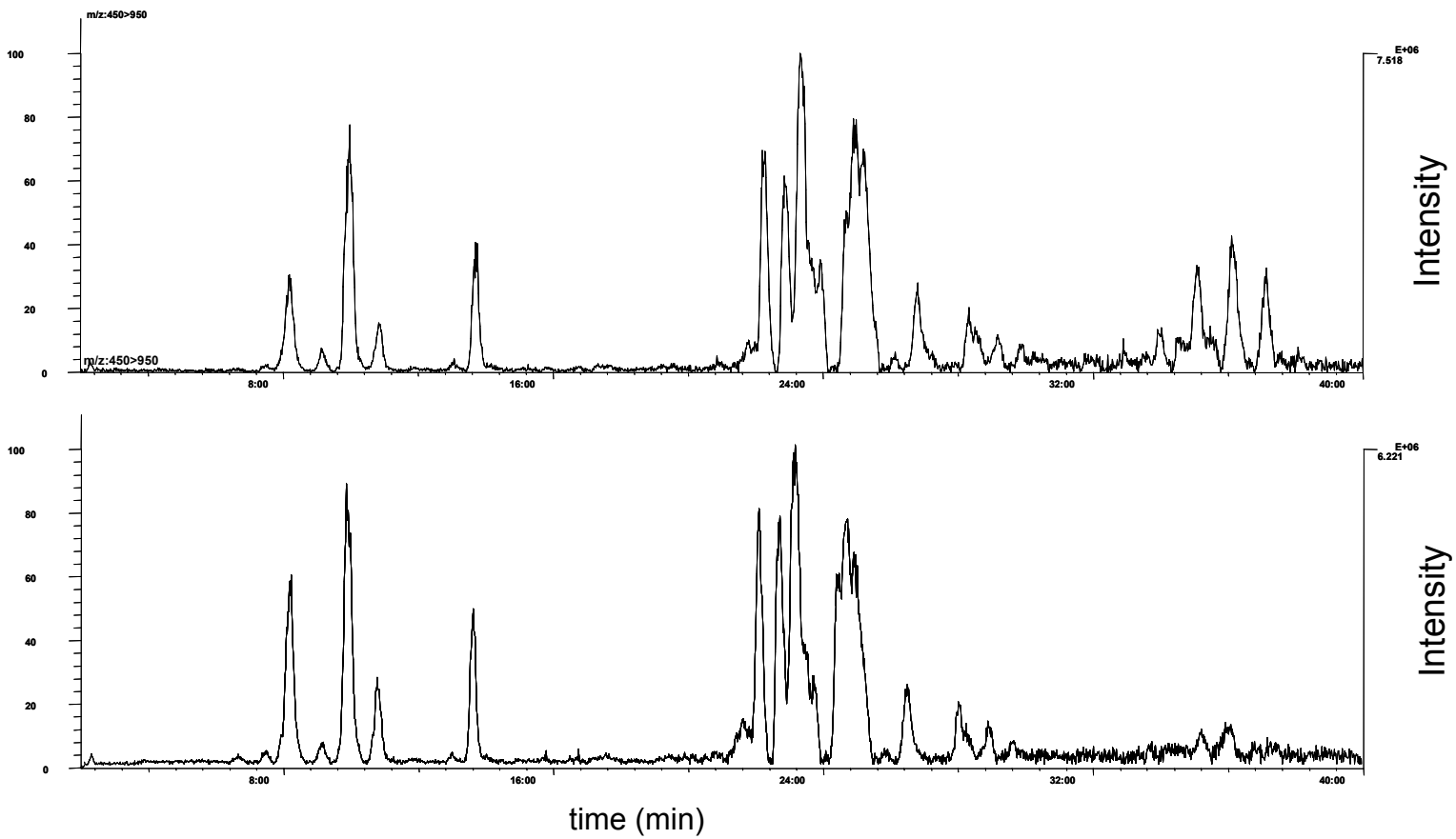
Metabolite Analysis
Plasma NMR Principal Component & Discriminant Analysis



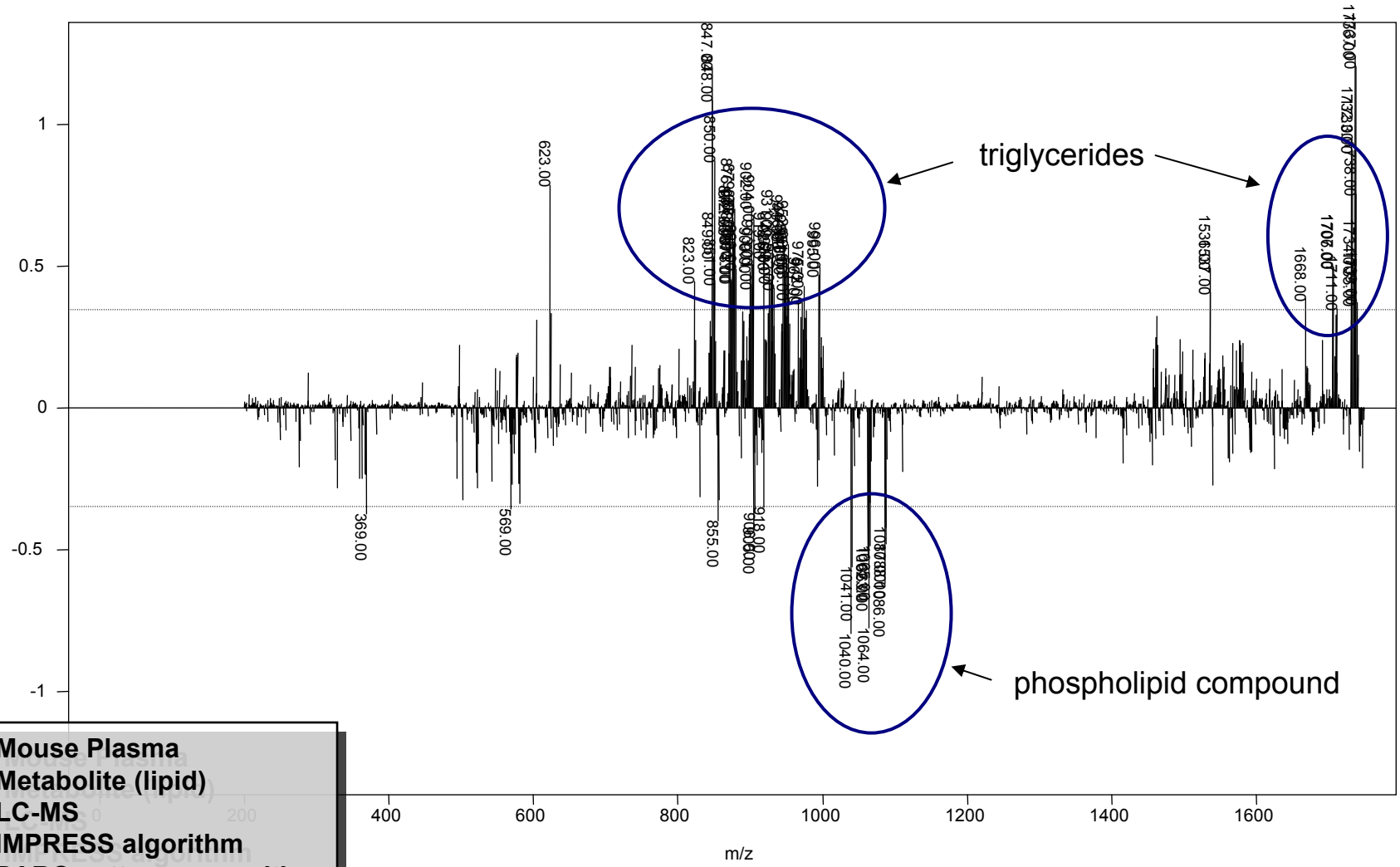
Metabolite Analysis
- LC/MS of Plasma Lipids

ApoE3 vs. WT: LC-MS Plasma Lipid Profiles

Plasma Lipids



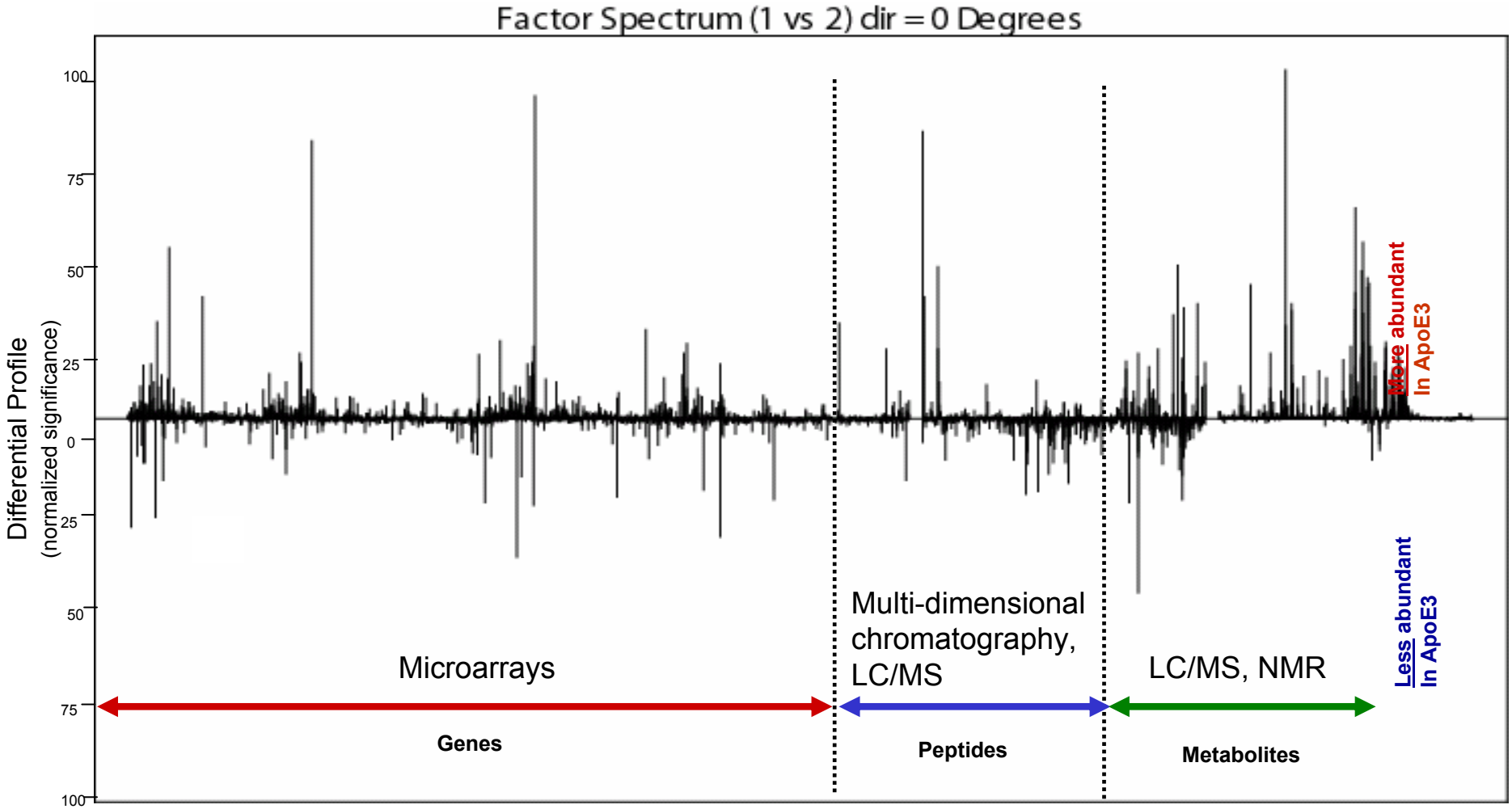
Metabolite Analysis
ApoE3 vs. WT: Plasma Lipid Difference Factor Spectrum



- **Mouse Plasma**
- **Metabolite (lipid)**
- **LC-MS**
- **IMPRESS algorithm**
- **PARC pattern recognition**

Novel Metabolite Information

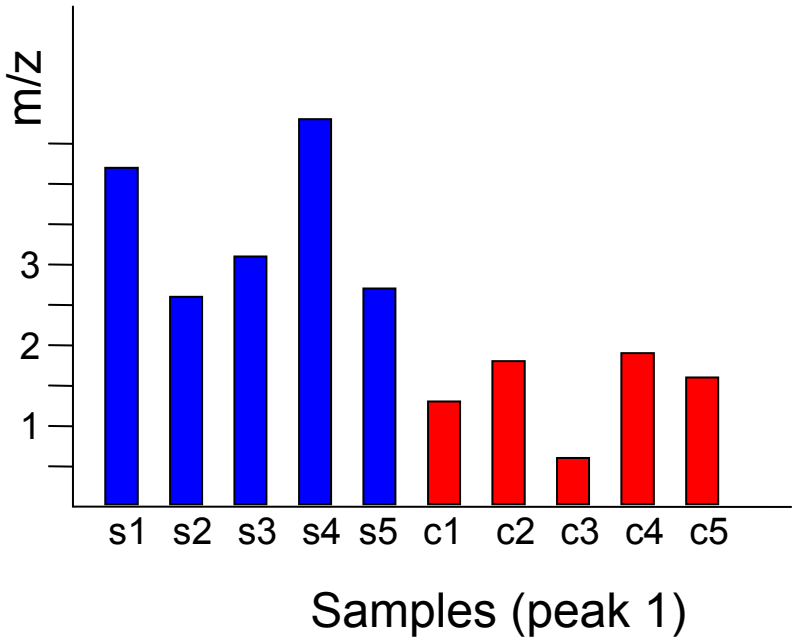
Normalized Integrated Differential Profile

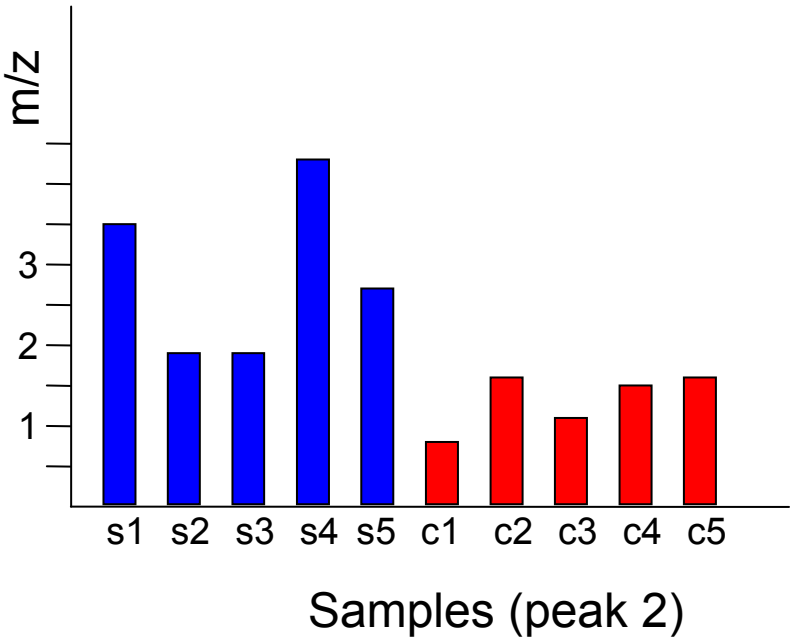


- Mouse Liver
- mRNA +Protein + Metabolite
- Normalization
- Pattern recognition

Similarity

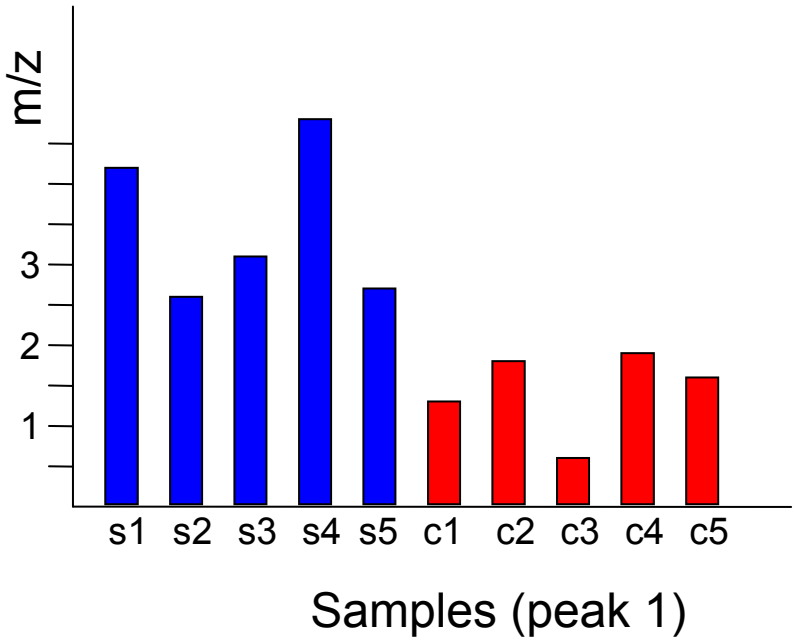
Example: highly correlated peaks



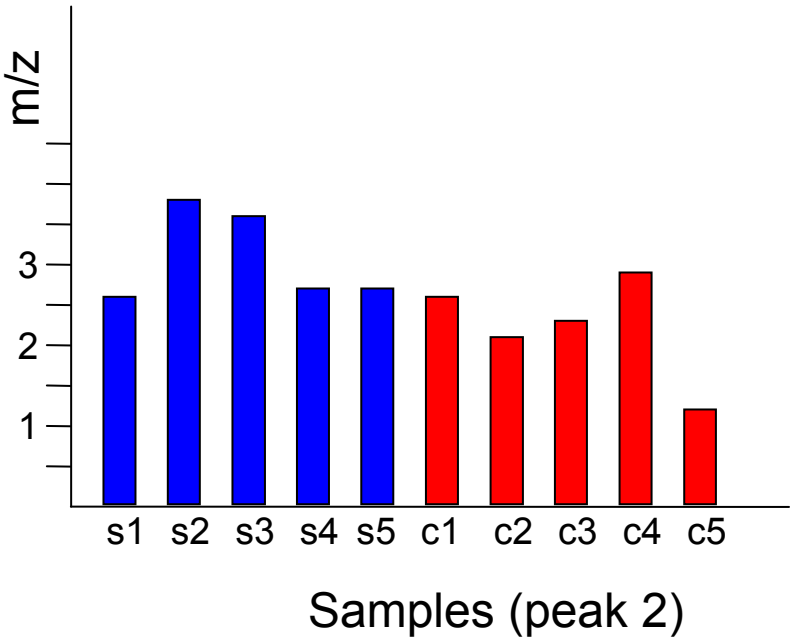


Similarity

Example: uncorrelated peaks



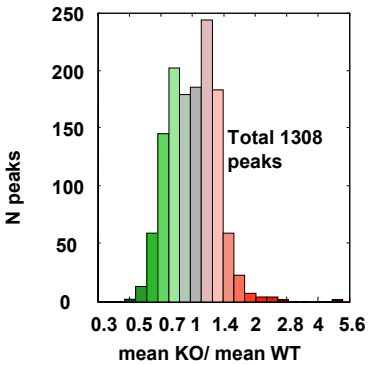
Samples (peak 1)



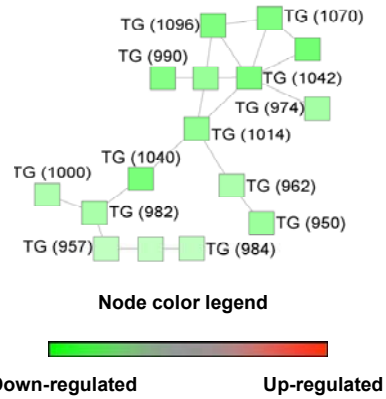
Samples (peak 2)

Correlation networkscan reveal patterns of changes relevant to the physiological response

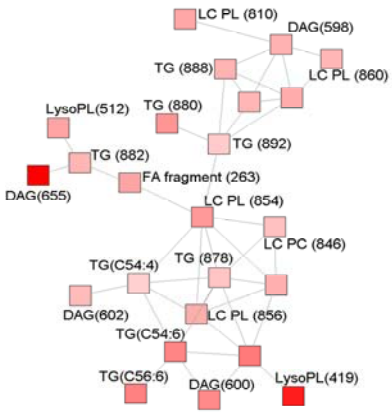
A. Histogram of the distribution of peaks (lipid compounds) according to up-/down-regulation.



B. Down-regulated long-chain triacylglycerol cluster

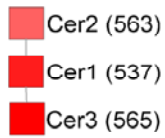


C. Up-regulated lipids (mainly long chain phospholipids, short-chain triacylglycerols, and diacylglycerols)



D. Downregulated cluster containing three C32 phsophatidylcholine lipids

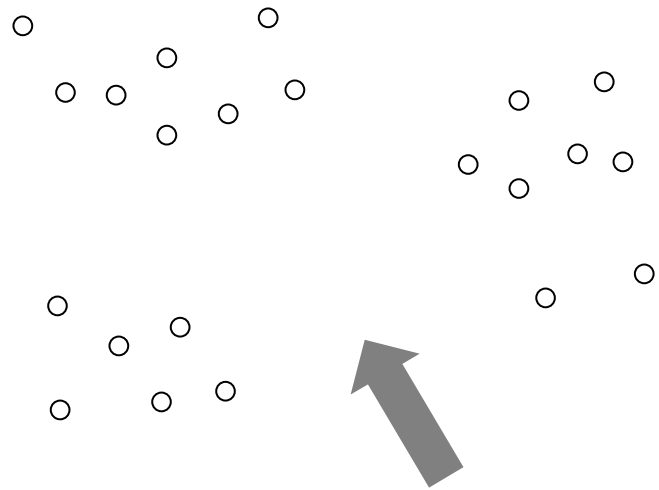
E. Upregulated ceramide cluster



G. Medina Gomez et al., Diabetes (2005)

Subspace clustering methods

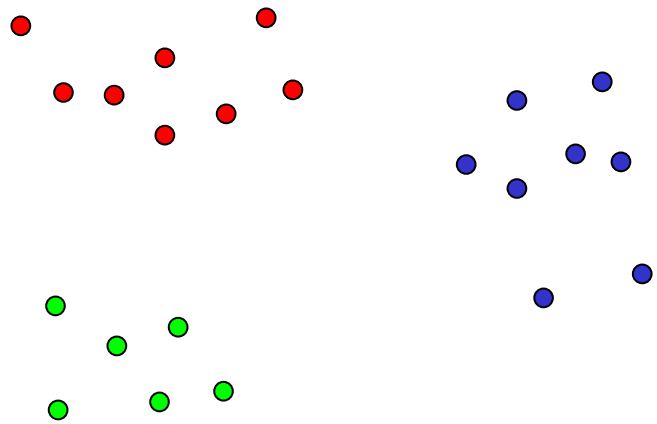
Unsupervised clustering
No prior information used



Set of “objects” (e.g. samples), each described by several variables (e.g. gene expression, metabolite profiles)

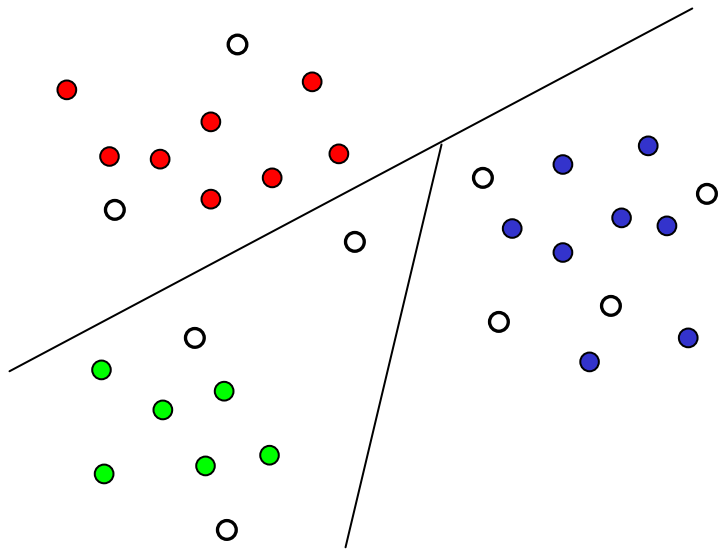
Unsupervised clustering
No prior information used

- Find groups of objects with small within-group distances and large between-group distances
- Several choices of distance metrics
- Examples: K-Means, Hierarchical, Subspace clustering methods

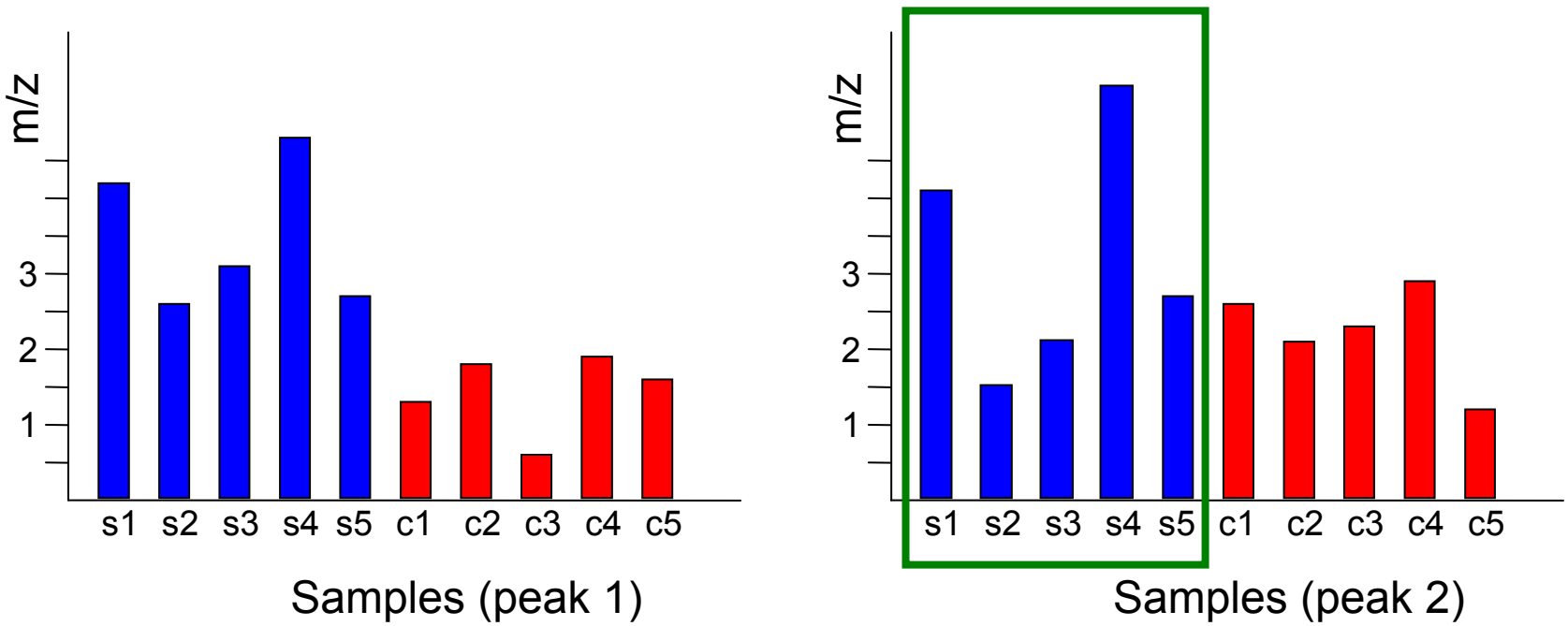


Supervised clustering
Prior grouping information available → Classification

- Find a model for each group, in order to be able to classify previously ungrouped objects
- Examples: Neural networks, Genetic algorithms, Support vector machines, Linear discriminant analysis
- Main problem in clinical applications (biomarkers, diagnostics): Lack of proper validation and overfitting.



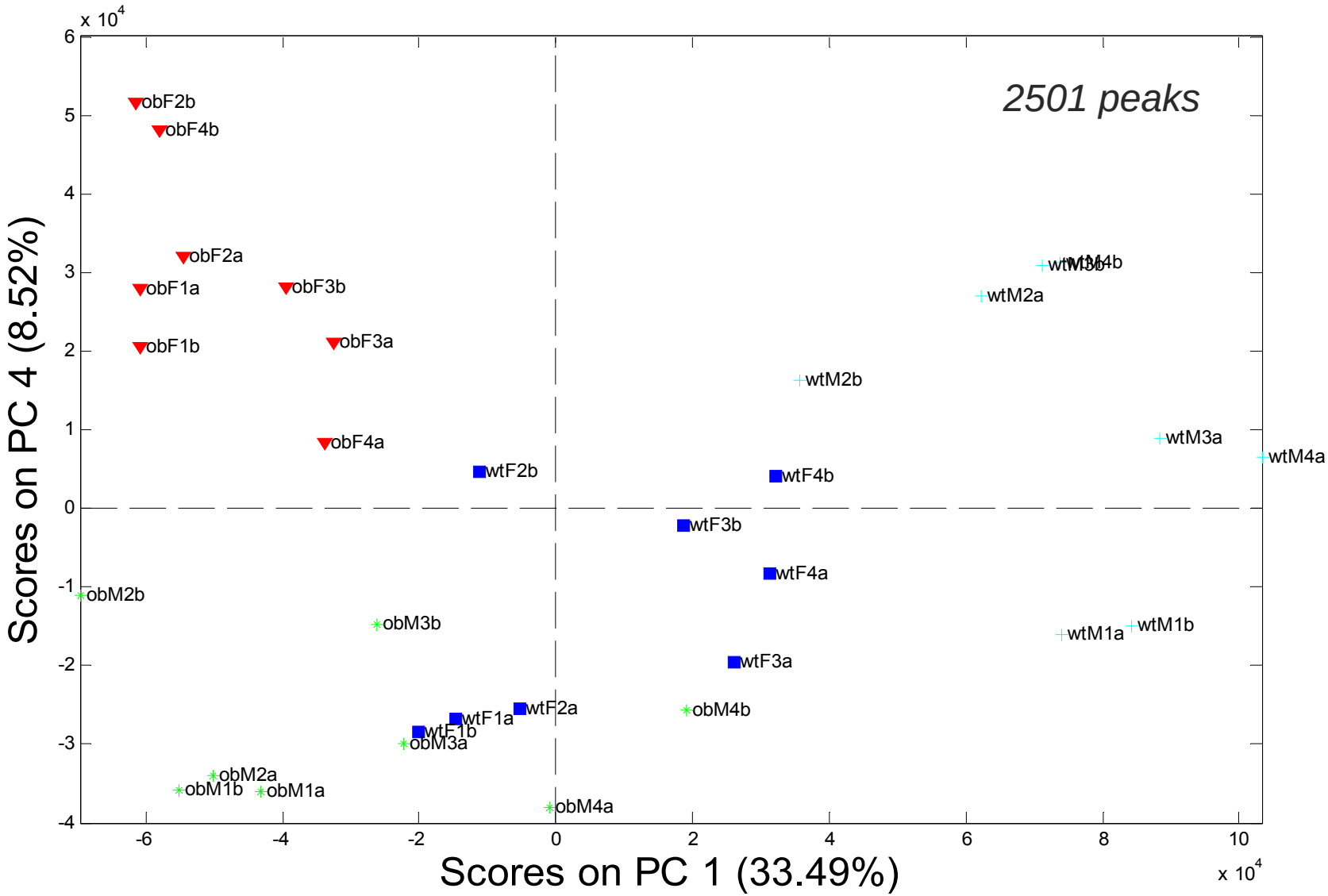
Subspace similarity
Metabolites may be dynamically (de)coupled under specific conditions



Example 2: Functional genomics
ob/ob mouse model

- Spontaneous mutation in *ob* gene resulting in lack of leptin (product of *ob* gene)
- Leptin hormone is a satiety signal
  - hormone secreted from adipose tissue
  - modulates energy intake and utilization
- Model for ea e obesity

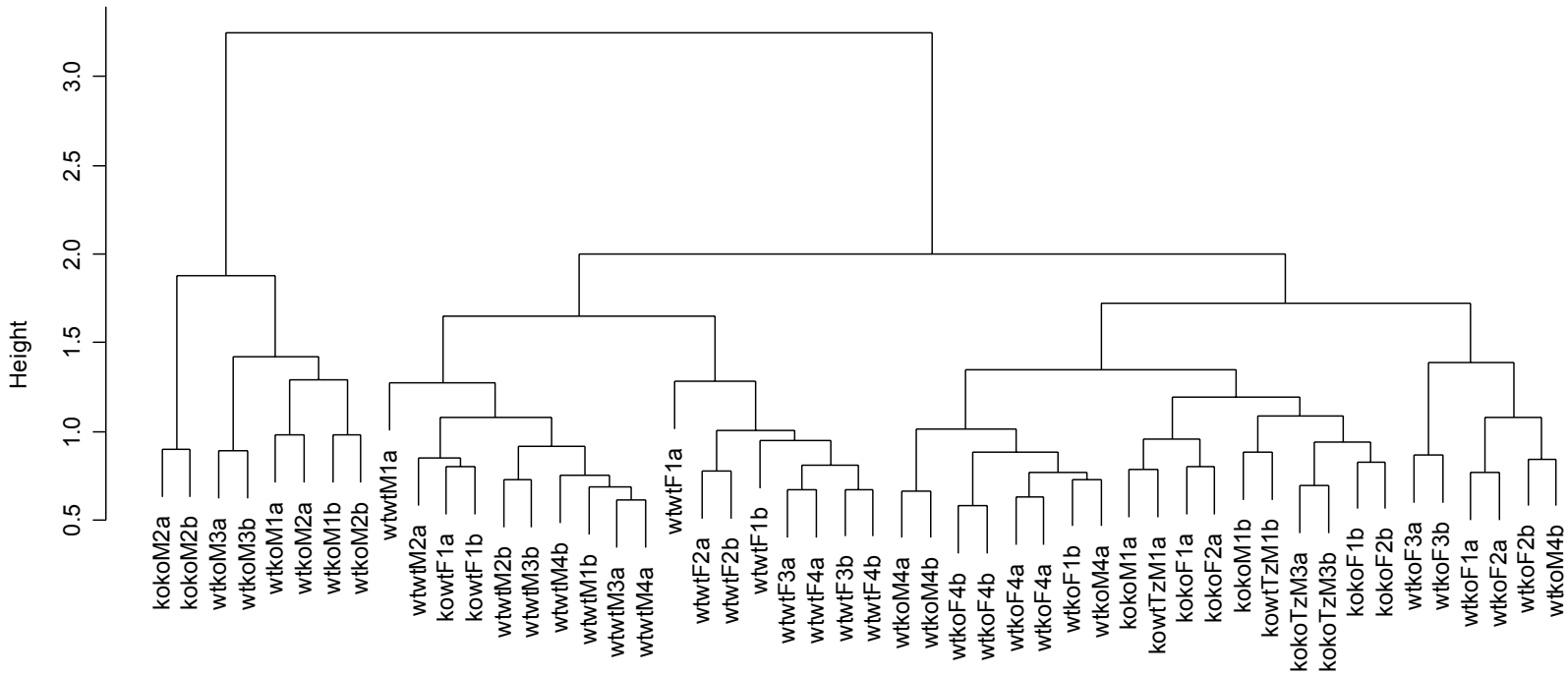
ob/ob and WT mouse white adipose tissue Lipidomic profiles reveal gender-specific differences



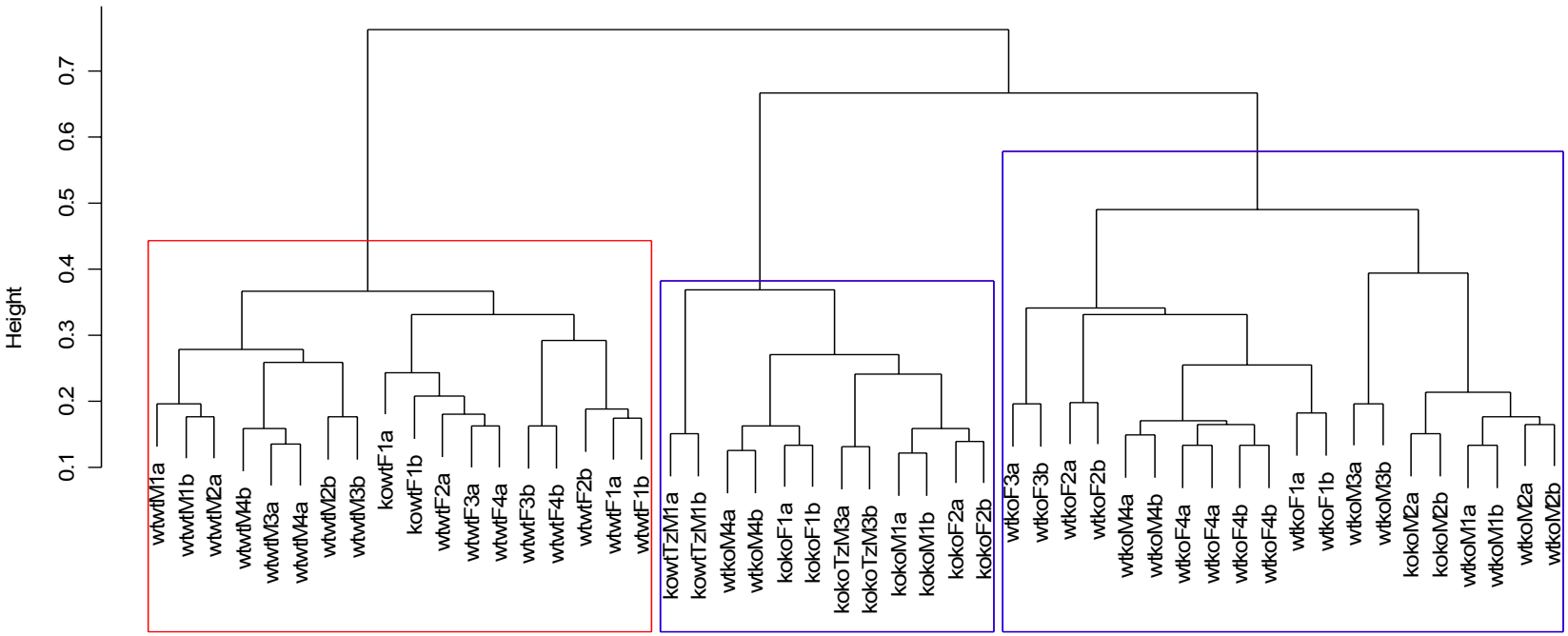
Oresic and Vidal-Puig

Double KO models (ob/ob and PPARγ2) WT/WT, WT/KO, KO/WT, and KO/KO

Clustering with Euclidian distance metric



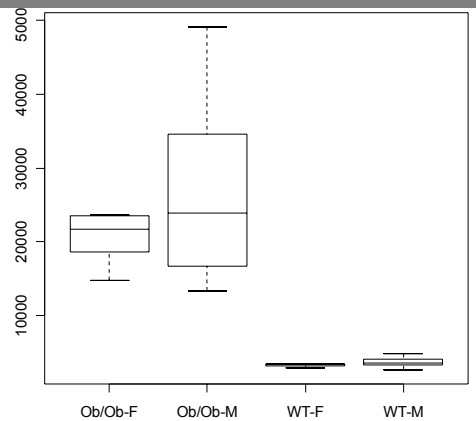
Subspace clustering (no a priori grouping assumed) COSA method



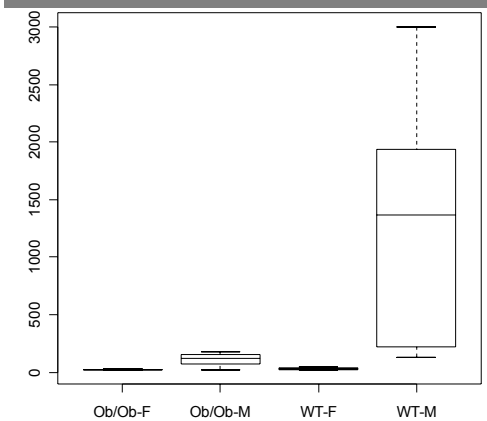
Three major groups identified from lipidomic profiles:
mainly WT/WT, mainly KO/KO, mainly WT/KO

ob/ob and WT mouse white adipose tissue Lipidomic profiles reveal gender-specific differences

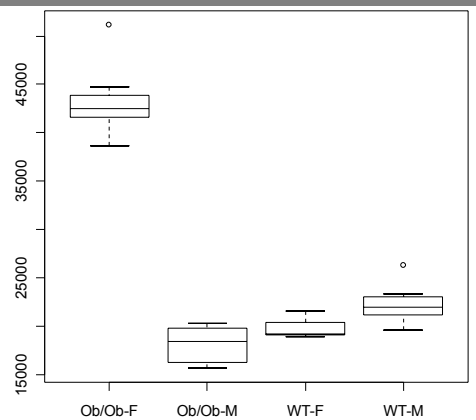
Monoacylglycerol



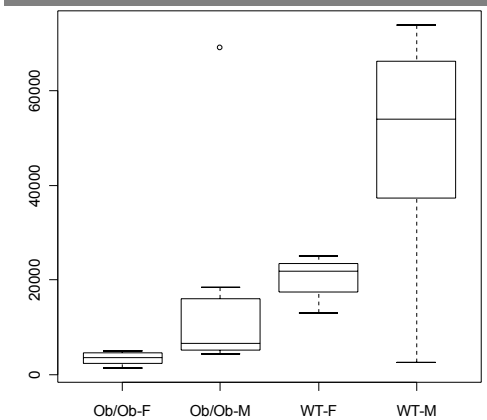
Sphingomyelin



Triacylglycerol



Triacylglycerol



References

- *Metabolic profiling: Its role in biomarker discovery and gene function analysis.* Harrigan and Goodacre, Eds. (Kluwer, 2003)
- J.H. Friedman and J.J. Meulman. Clustering objects on subsets of attributes. *J. R. Statist. Soc. B,* **66,** 1-25 (2004).
- M. Oresic, C.B. Clish, E.J. Davidov, E. Verheij, J.T.W.E. Vogels, L.M. Havekes, E. Neumann, A. Adourian, S. Naylor, J.v.D. Greef, and T. Plasterer. Phenotype characterization using integrated gene transcript, protein and metabolite profiling. *Appl. Bioinformatics,* **3,** 205-217 (2004).
- Katajamaa and Oresic, Processing methods for differential analysis of LC/MS profile data, *BMC Bioinformatics* **6**, 179 (2005).