Metabolomics

Introduction to statistical analysis and visualisation

Dipl.Inf. Henry Wirth

Interdisciplinary Center for Bioinformatics,
University of Leipzig

Department of Proteomics,
UFZ - Helmholtz-Centre for Environmental Research

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Introduction

Metabolites:

- small molecules

- intermediates and products of metabolisms

- form large networks of interaction

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Introduction

Metabolites:

- small molecules

- intermediates and p

- form large network
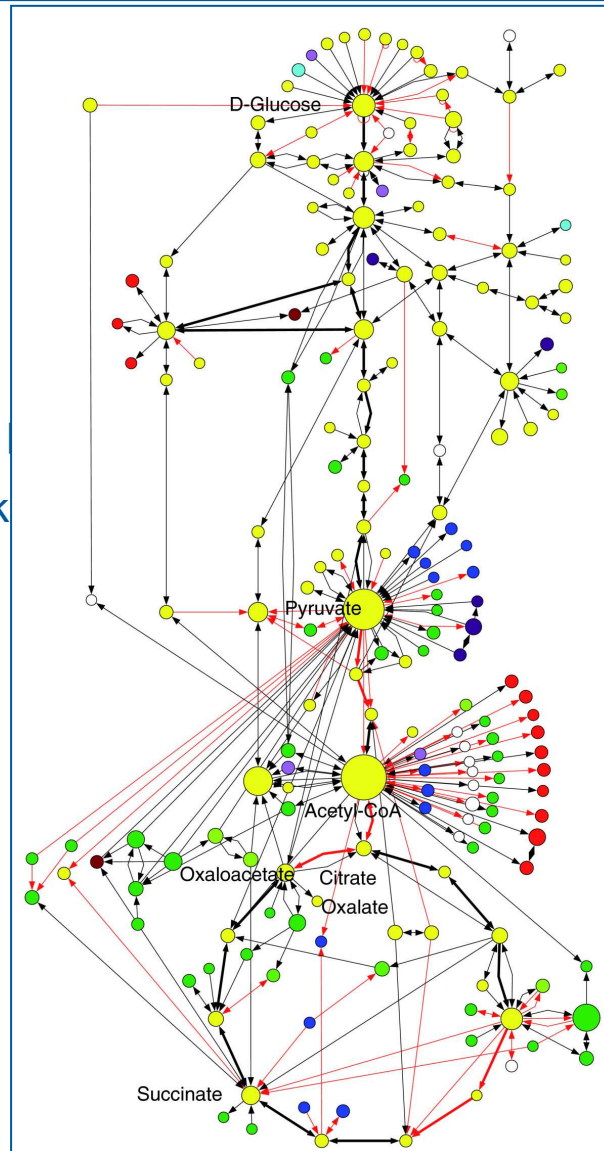
glycolysis

citric acid cycle

http://genomebiology.com/2007
/8/9/R182/figure/F3

Introduction

Metabolomics:

- study of global metabolite profiles in a system

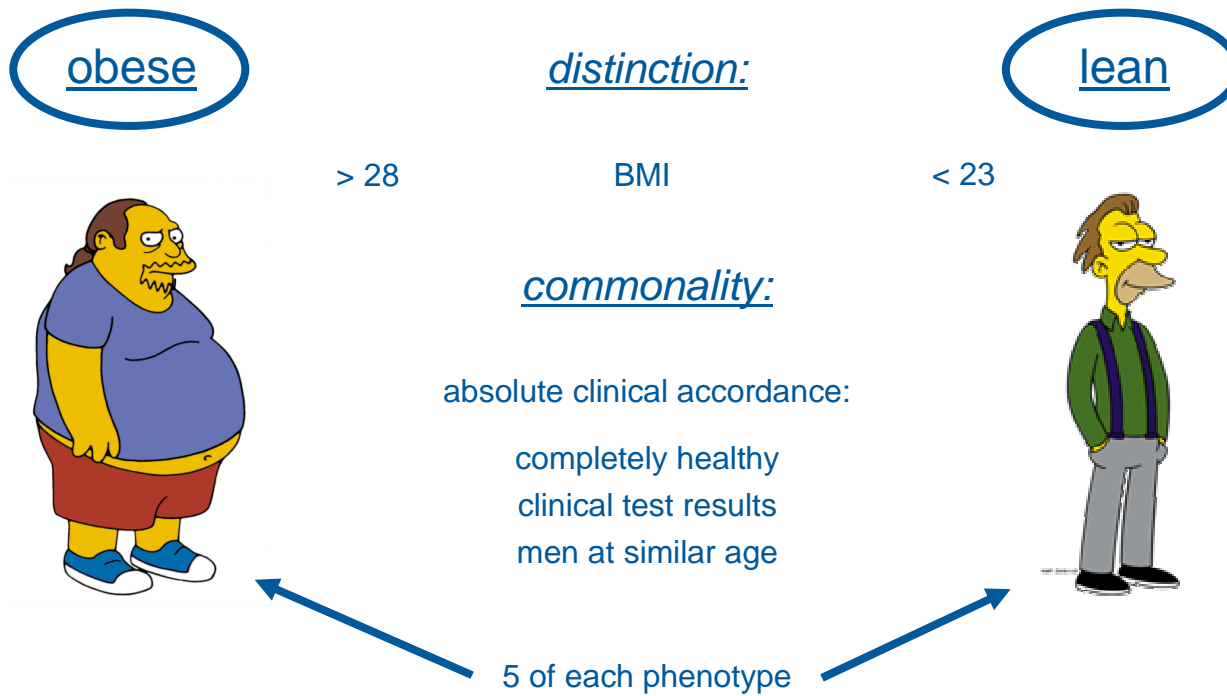- „Show me your trashcan, so I´ll tell you how you live!" (by Andreas Oberbach)

Study: Oberbach, spring 2007

Introduction

Study:

2 types of people:

obese

distinction:

lean

> 28

BMI

< 23

commonality:

absolute clinical accordance:

completely healthy
clinical test results
men at similar age

5 of each phenotype

IZBI UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Introduction

Study:

3 points of time:

before exercise

after exercise

after 24h regeneration

IZBI UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Introduction

Dataset : Kit : „Biocrates AbsoluteIDQ"

quantify concentrations of 163 metabolite in 5 groups:

carnitines: transport of fatty acids into mitochondria

aminoacids: building blocks and breakdown products of

proteins / other molecules

phospholipids: major components of cell membranes

sphingolipids: components of membranes;important mediators

in signaling cascades

sugar: energy provider

Introduction

Analysis Agenda:

time

	before exercise	after exercise	after 24h regeneration
	„pre"	„post"	„24h"

groups

„obese"

are there differences in metabolism despirte of clinical accordance ??
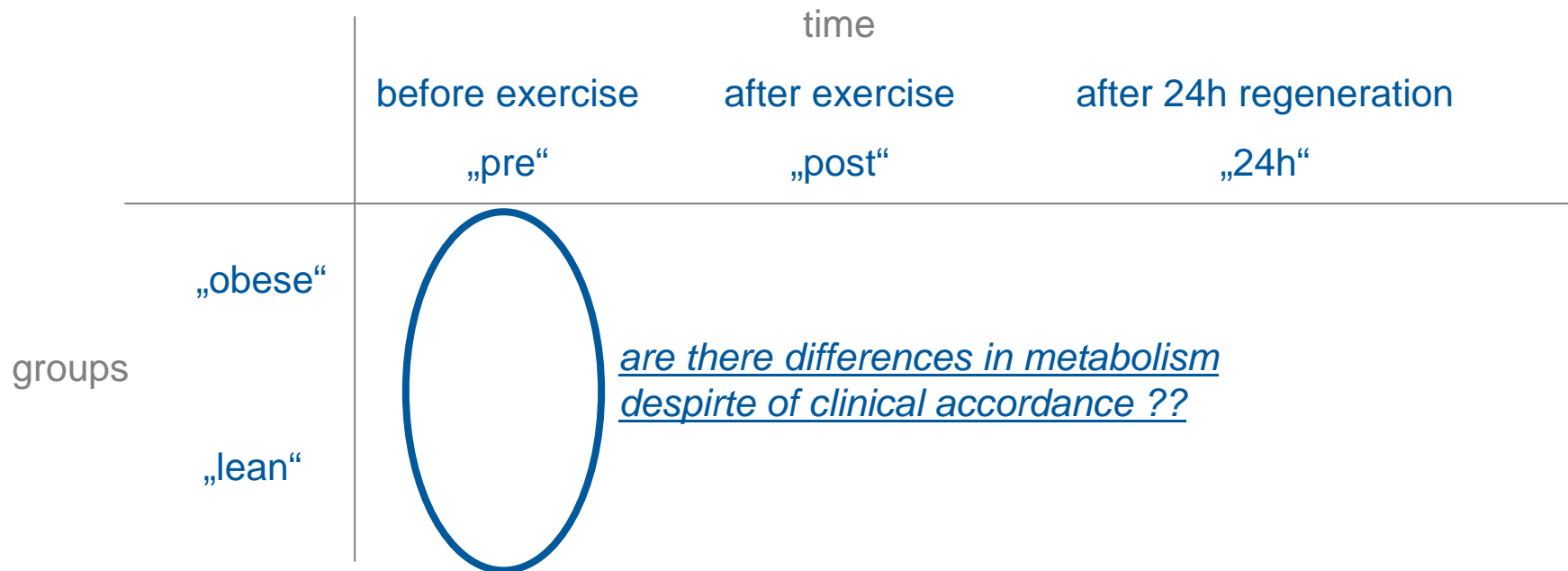
„lean"

group comparison: obese pre – lean pre

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Introduction

Analysis Agenda:

	time		
	before exercise	after exercise	after 24h regeneration
	„pre"	„post"	„24h"
groups „obese"	*how are metabolites changing over time ??*		
„lean"	*how are metabolites changing over time ??*		

behaviour under stress: obese pre-post-24h

lean pre-post-24h

Introduction

Analysis Agenda:

	time		
	before exercise	after exercise	after 24h regeneration
	„pre"	„post"	„24h"
„obese"	T-Test : *which metabolites change among groups ?*		
groups	PCA : *can groups be recovered with a certain set of metabolites ?*		
„lean"	Clustering : *are there structures in a specific data(sub)set ?*		

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Analysis – „T-Test"

Idea:

- we have: 2 groups and a set of features (concentrations, spotvolumes, attributes…)

- objective: feature extraction

→ decide, which of the features are significant different

Analysis – „T-Test"

Example:
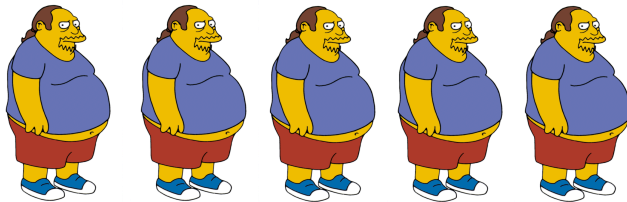
Groups: 5x obese   5x lean

Analysis – „T-Test"

Example:

Features: Size   Weight

sO   -   sL   wO   -   wL
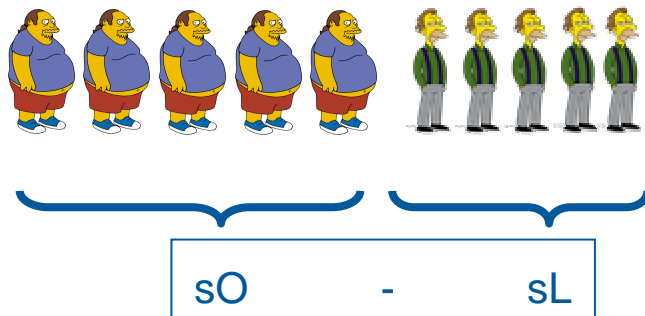
groupwise-mean difference   groupwise-mean difference

Analysis – „T-Test"

Example:
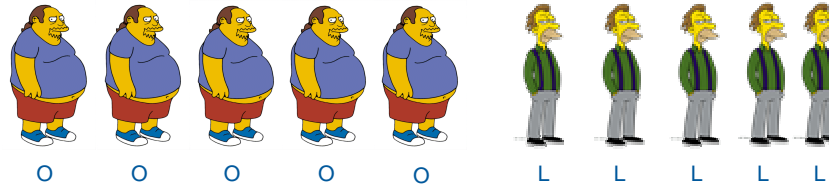
O - obese, L - lean

Features: Size   Weight

groupwise-means difference   groupwise-means difference

high probablity   very small probablity

-100   0   100   -100   0   100
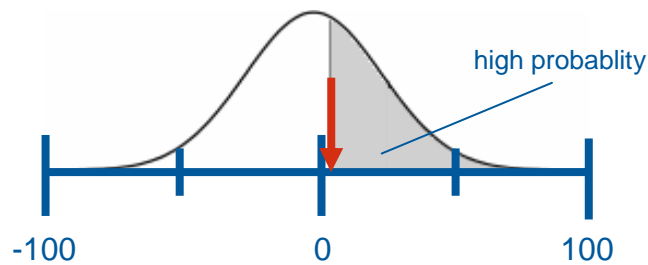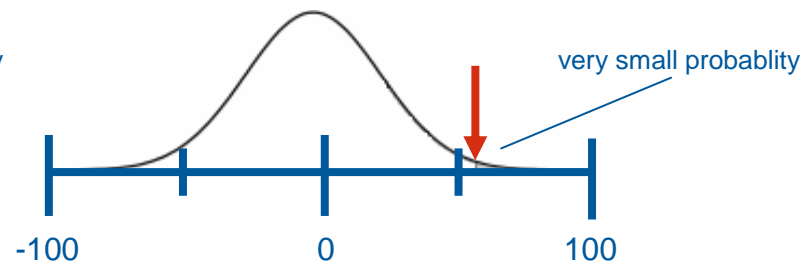
ordinary difference   rare difference

(significance)

Analysis – „T-Test"

Visualisation: Volcano-Plots (2 group comparison)
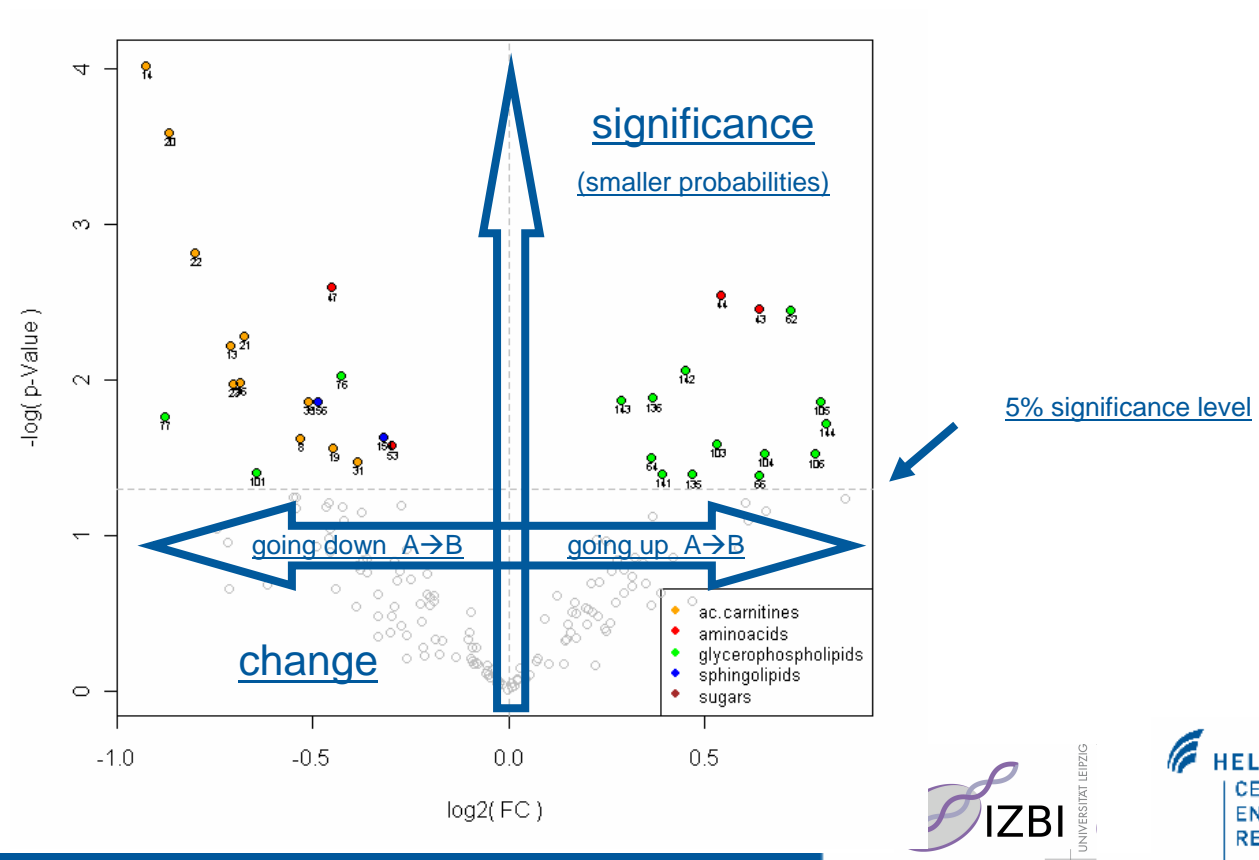
Analysis – „T-Test"

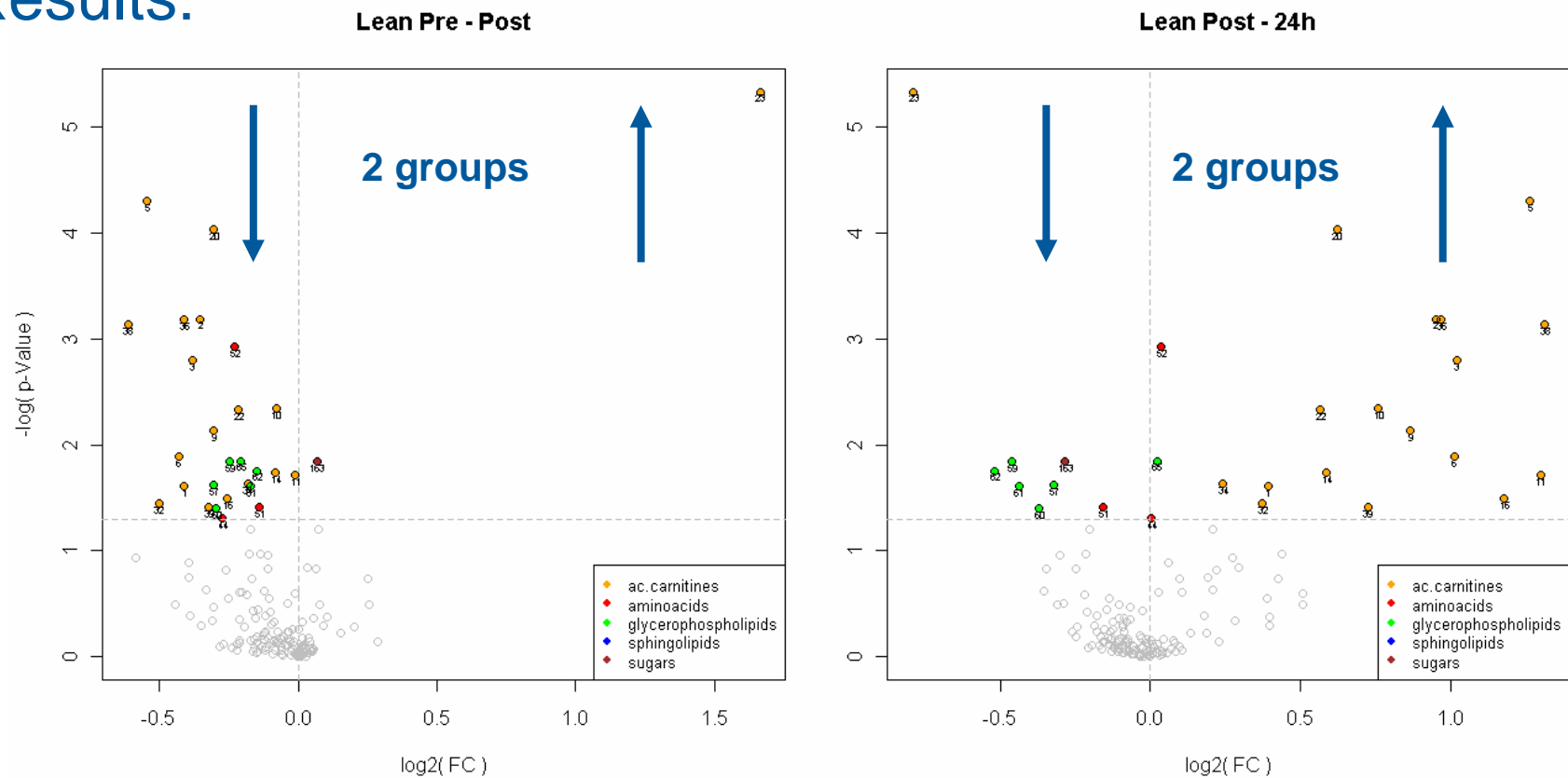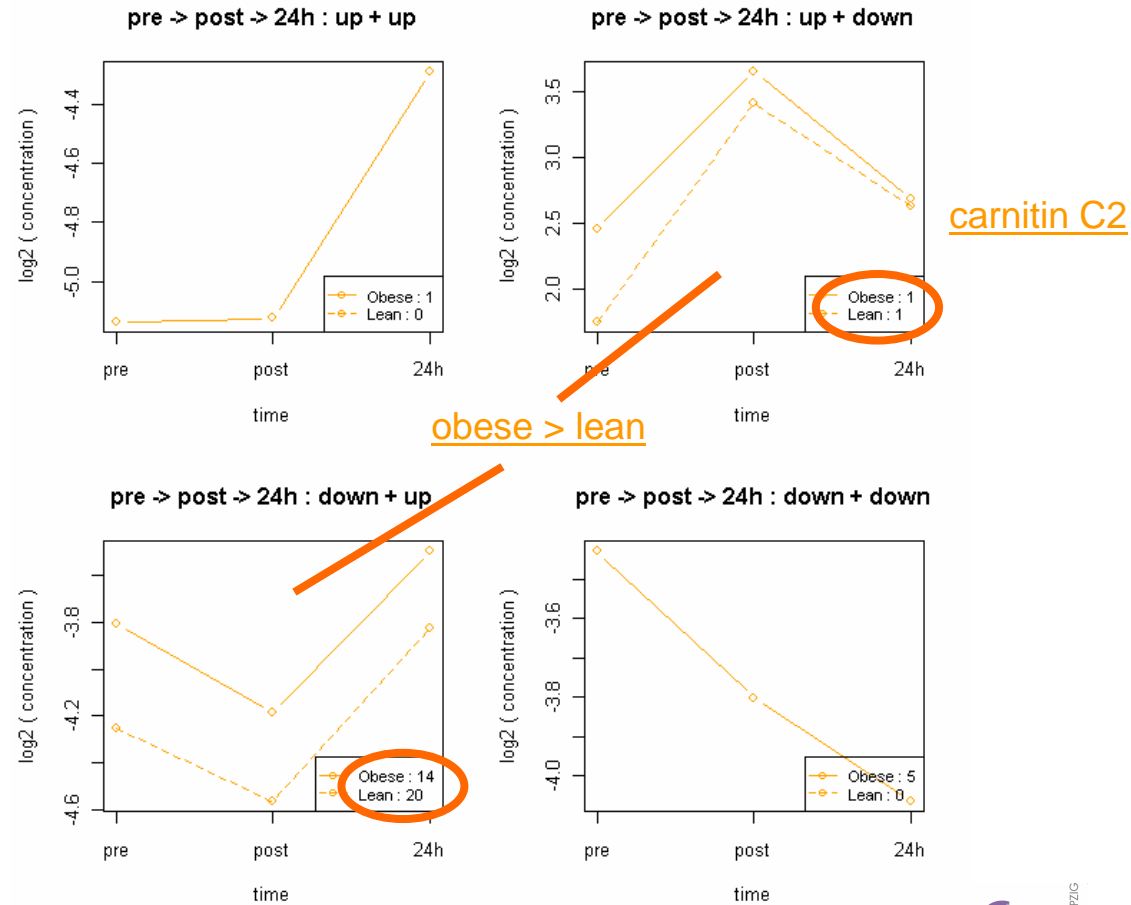Results:
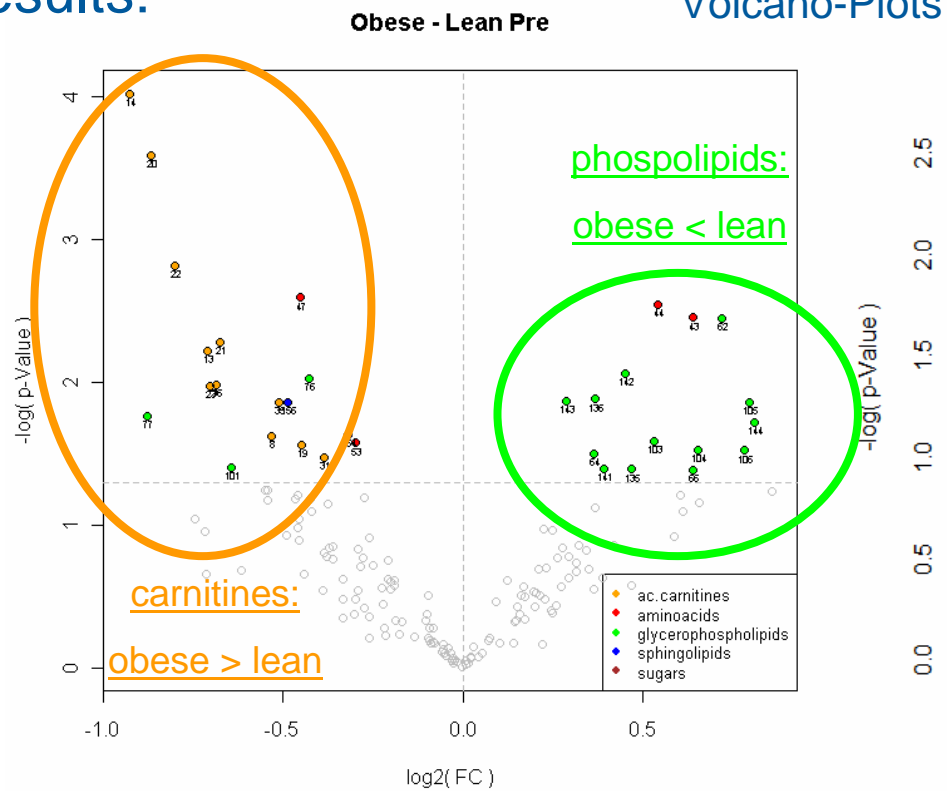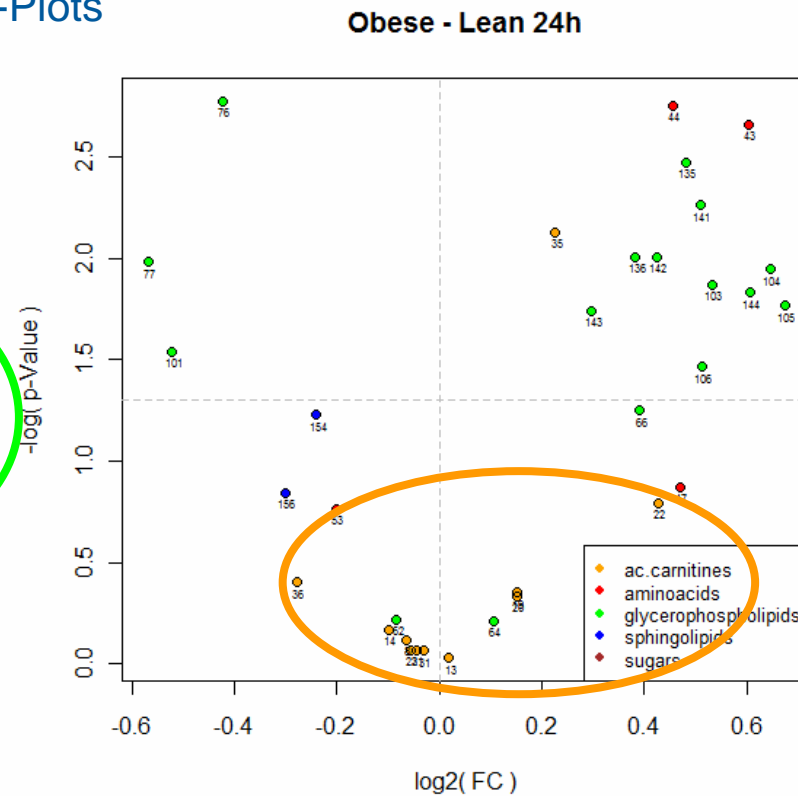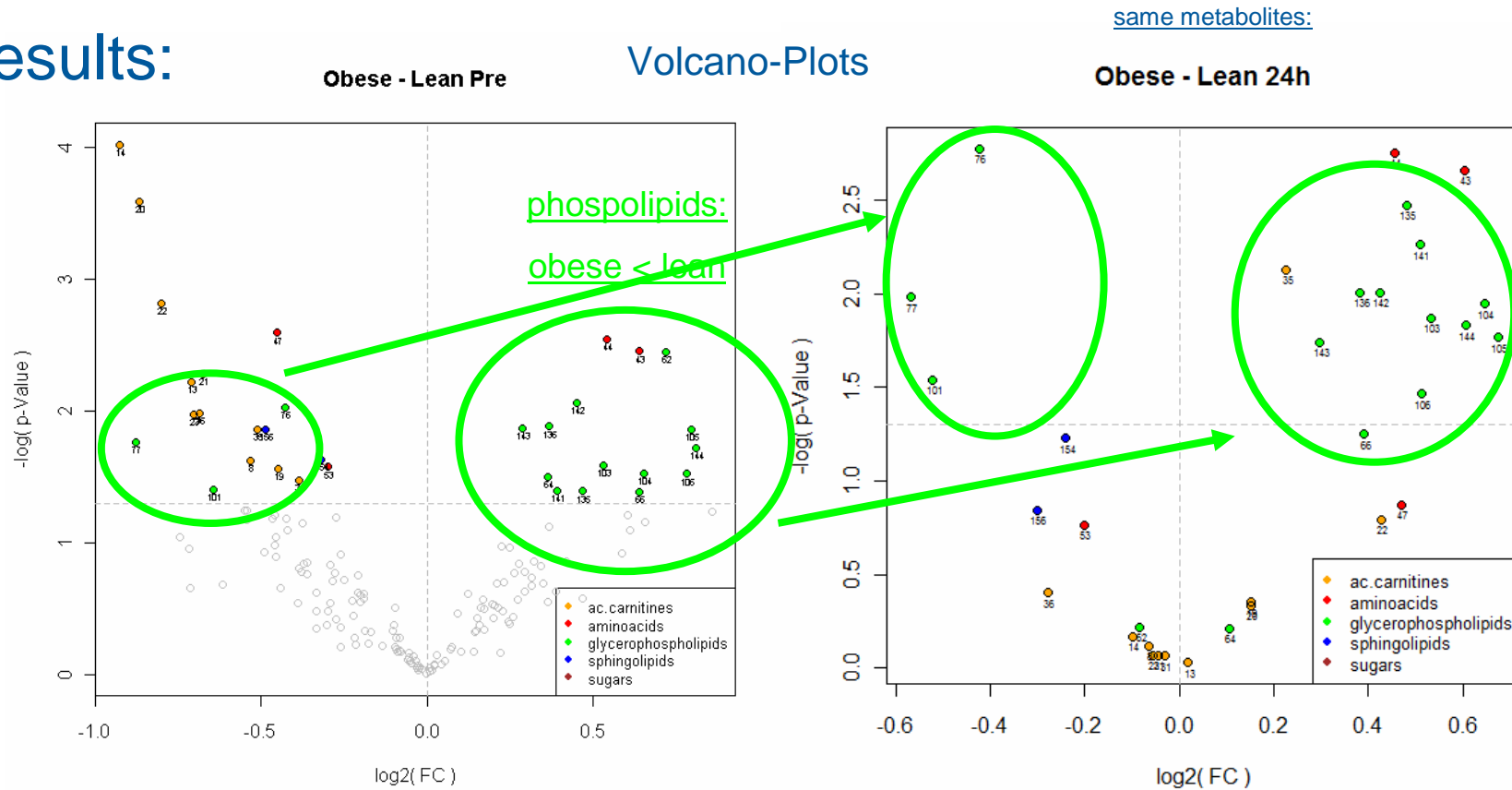
Analysis – „T-Test"

Results: Volcano-Plots

Analysis – „T-Test"

Results: Volcano-Plots

Analysis – „T-Test"

Results:

Volcano-Plots

carnitin C2

Analysis – „T-Test"

Results:

Analysis – „T-Test"

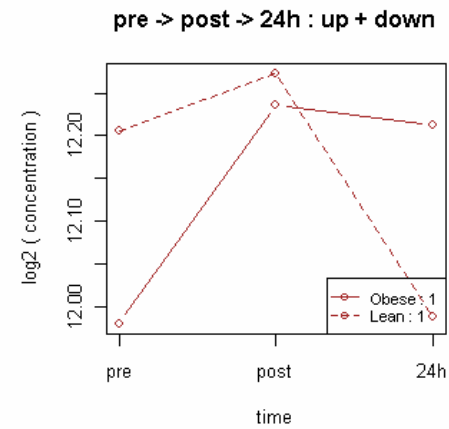Results:

Timeseries: pre $\rightarrow$ post $\rightarrow$ 24h

4 groups:				
1)		up		up
2)		up		down
3)		down		up
4)		down		down

IZBI

UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
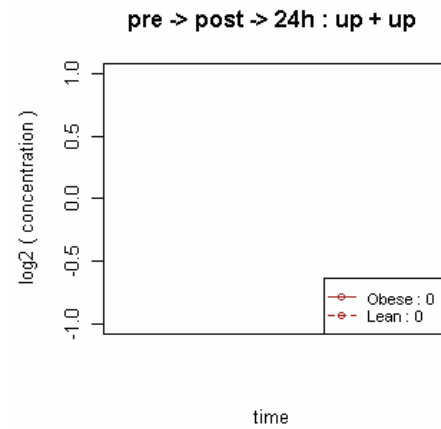RESEARCH – UFZ
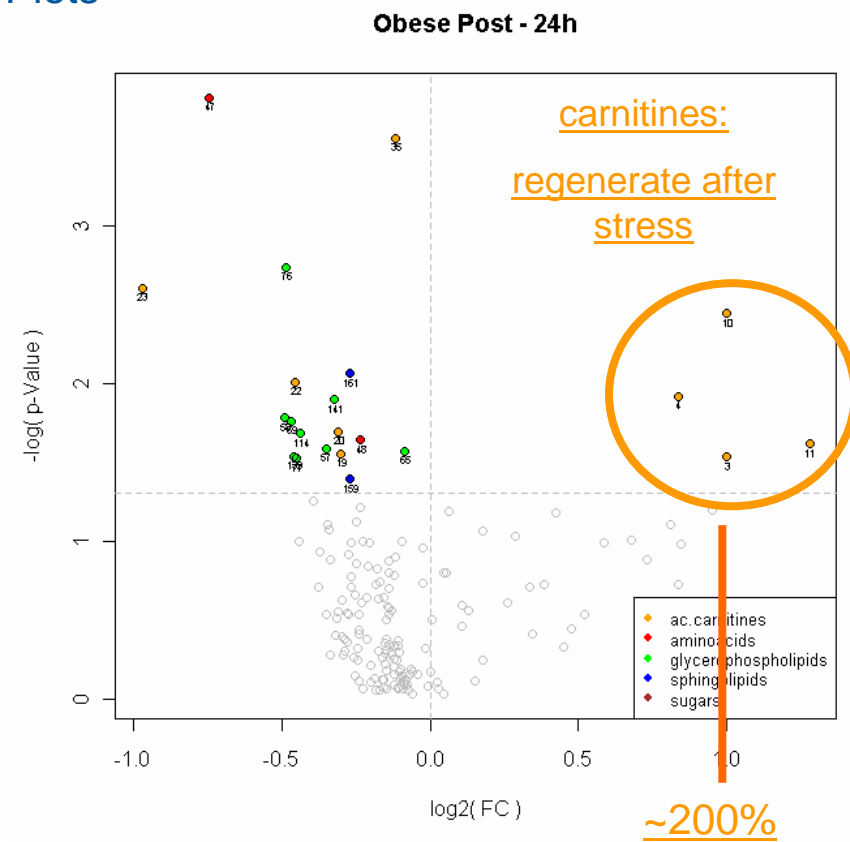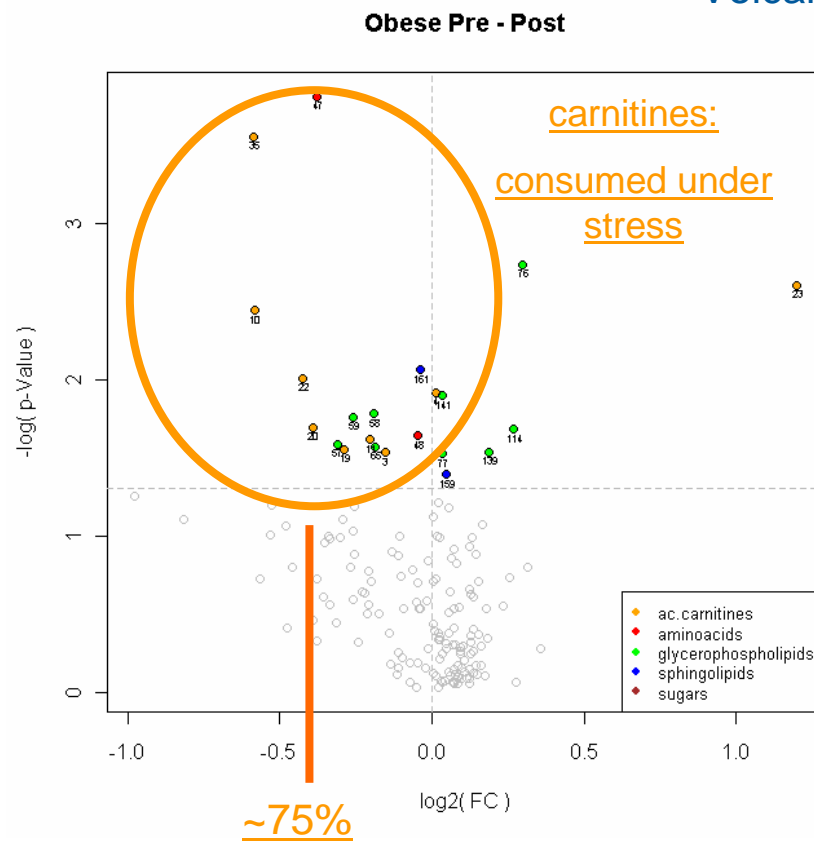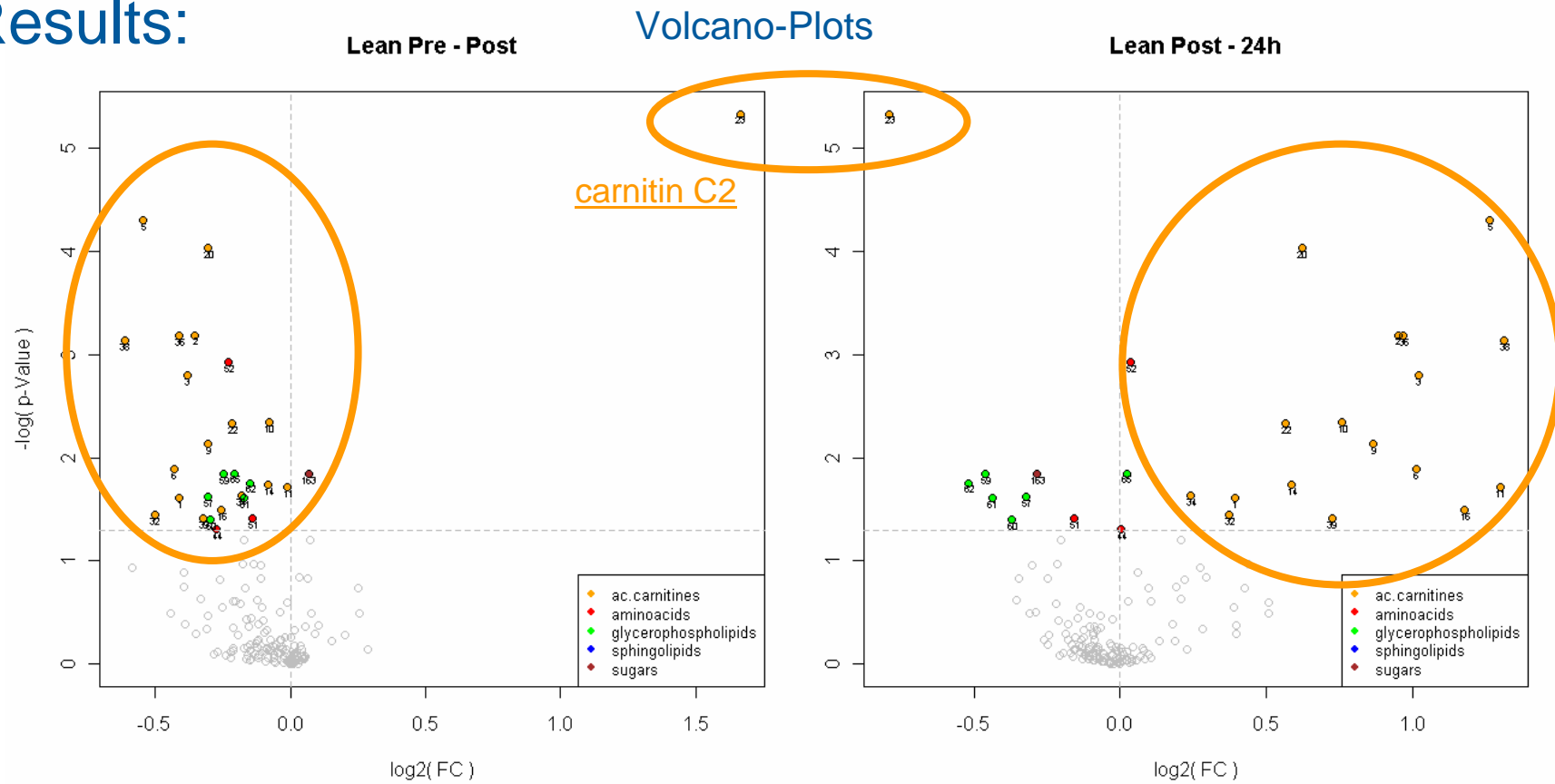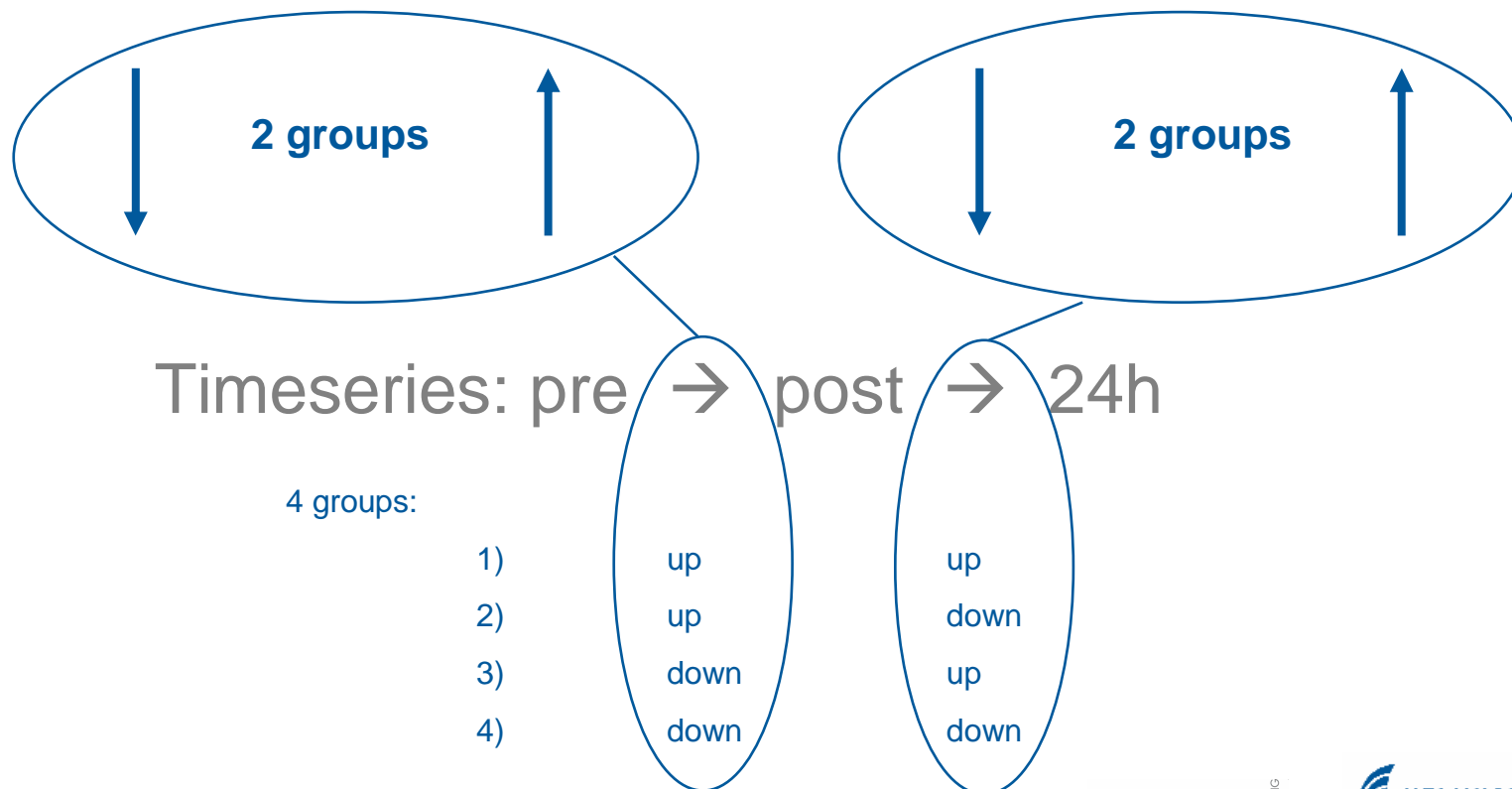
Analysis – „T-Test"

Results:

Analysis – „T-Test"

sugar

Results:
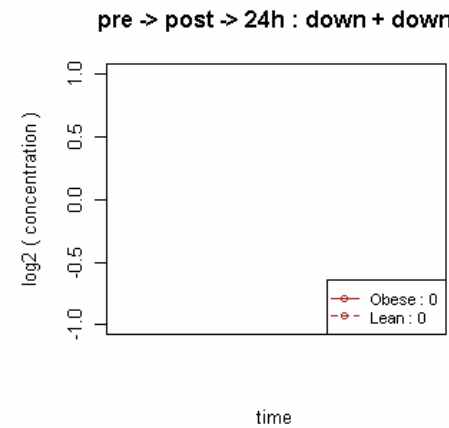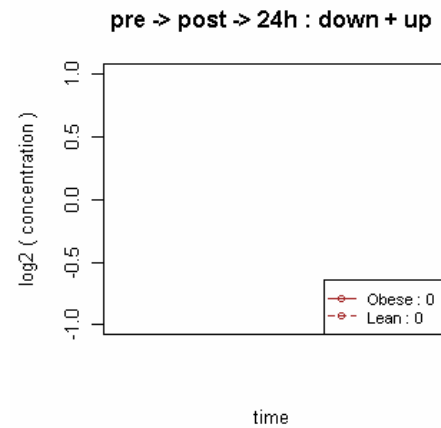
obese:

cells import more sugar, but can't get rid of it

?!

[diabetes mellitus type 2]

chronic high blood sugar level

Analysis – „T-Test"

Insight:

carnitines: underlie changes while akute physical load

(consumption and regeneration)

not capable to distinguish groups

more carnitines changing in lean people

C2 acts different than most carnitins

phospolipids: underlie changes caused by phenotype

sugar: hint of correlation obesity <> diabetes type 2

are there differences in metabolism despirte of clinical accordance ??

YES, in normal state as under stress

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ
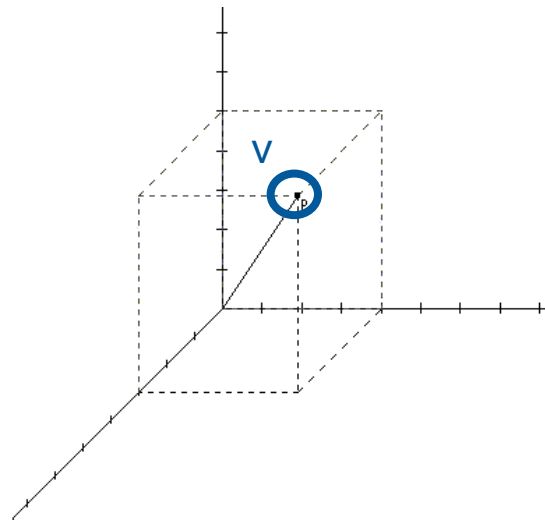
Analysis – „PCA"

Idea:

- PCA = „principle component analysis"

- unsupervised → „blind" processing – no information about grouping

  afterwards: recover group separation   ?

- concept: dimension reduction to visualizeable „principle components"

Analysis – „PCA"

Example dimensions and spaces:

$$v = (3, 4, 5)$$

3-dimensional vector

3-dimensional space

Analysis – „PCA"

Example dimensions and spaces:

$v = (3, 4, 5, 1)$ 4-dimensional vector

?   4-dimensional space

not visualizable

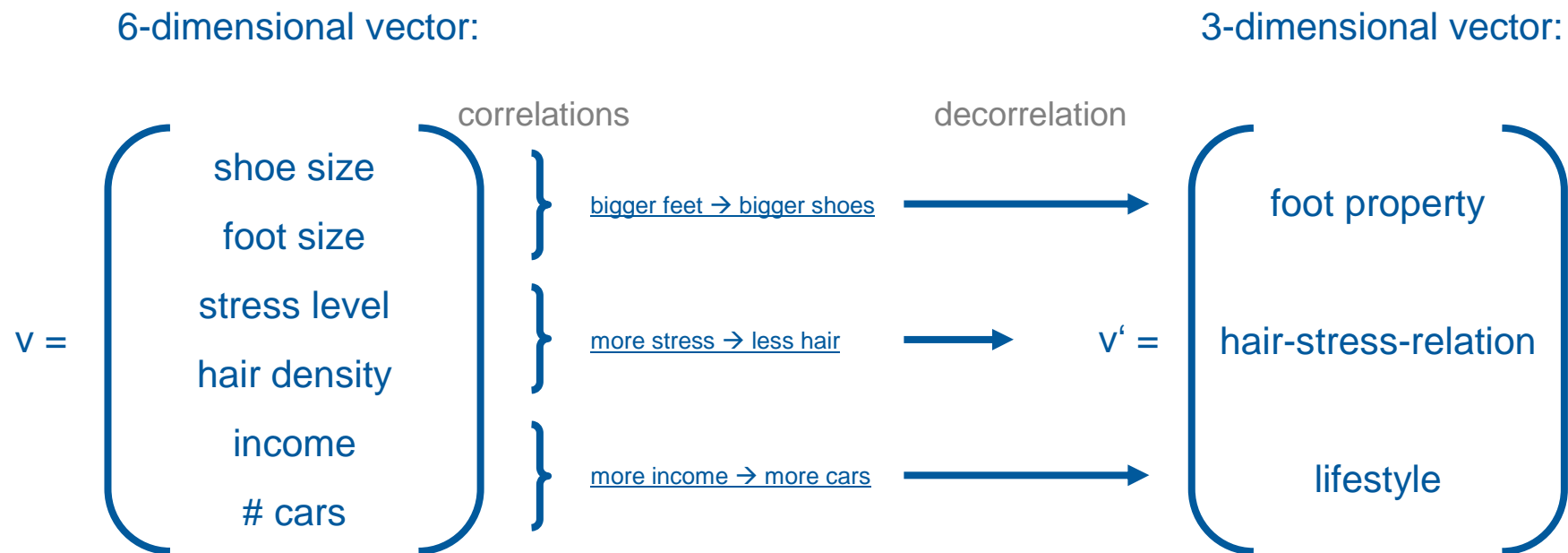# Analysis – „PCA"

Example dimensions and spaces:

$v = ($ 163 metabolite concentrations $)$ 163-dimensional vector

$v = ($ 56000 gene expresions $)$ 56000-dimensional vector
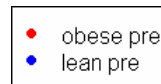
dimension reduction to 2 or 3 dimensions

IZBI UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Analysis – „PCA"

Example dimension reduction:

6-dimensional vector: 3-dimensional vector:

$$
v = \begin{pmatrix} \text{shoe size} \\ \text{foot size} \\ \text{stress level} \\ \text{hair density} \\ \text{income} \\ \text{\# cars} \end{pmatrix}
$$

correlations

bigger feet → bigger shoes

more stress → less hair

more income → more cars

decorrelation

$$
v' = \begin{pmatrix} \text{foot property} \\ \text{hair-stress-relation} \\ \text{lifestyle} \end{pmatrix}
$$

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Analysis – „PCA"

Results:

all metabolites

Obese - Lean pre

Idea:

using significant metabolites from the volcano-plots

Analysis – „PCA"

Results:

Obese - Lean pre
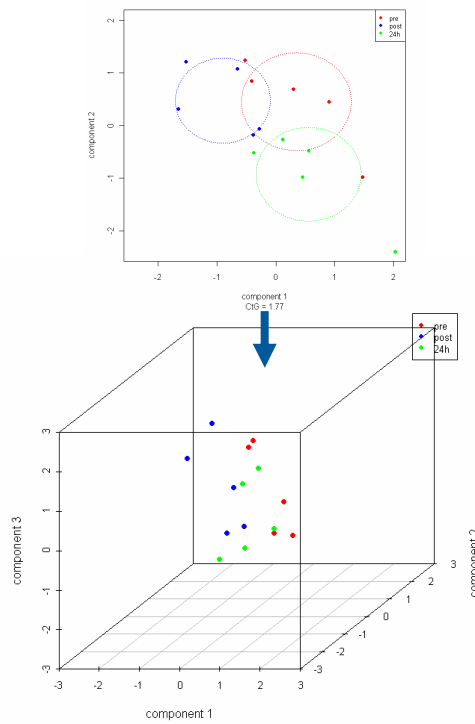

all metabolites

significant metabolites

IZBI UNIVERSITÄT

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ
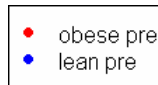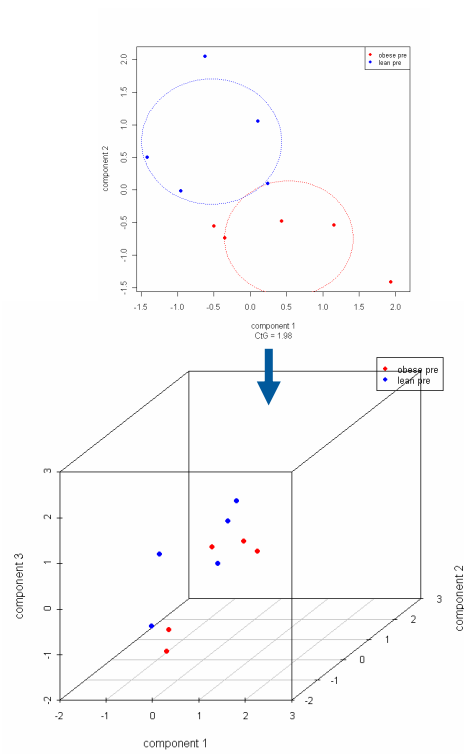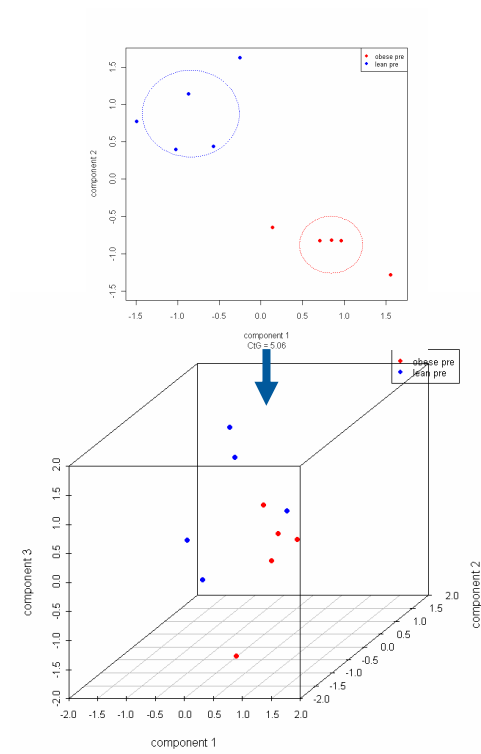
Analysis – „PCA"
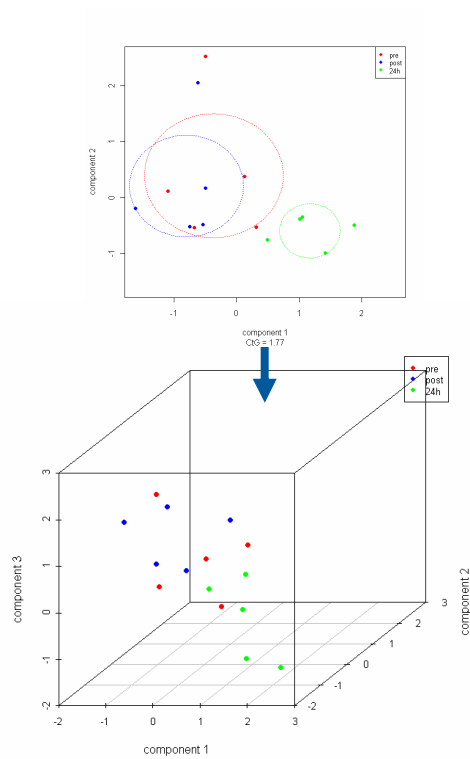
Results:

Obese pre-post-24h

Analysis – „PCA"

Results:

Lean pre-post-24h
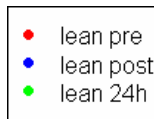
all metabolites

significant metabolites
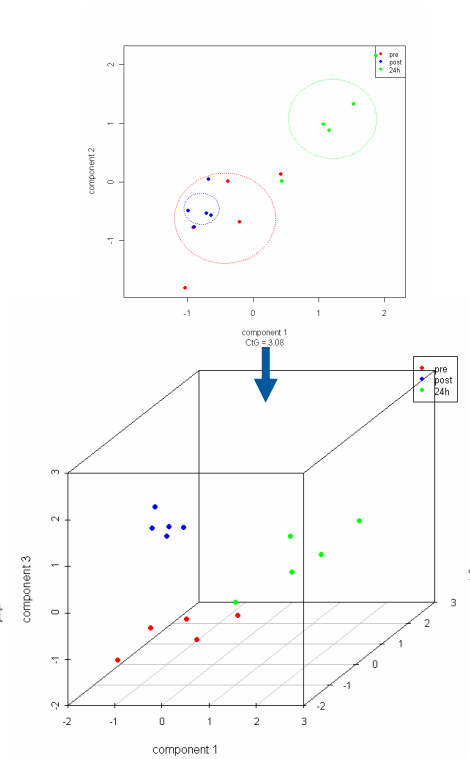
group-separation

in 3rd component

IZBI

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Analysis – „PCA"

Insight:

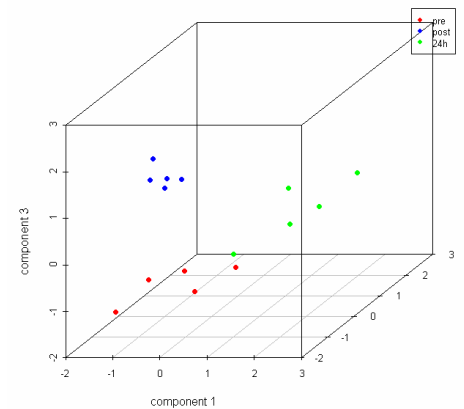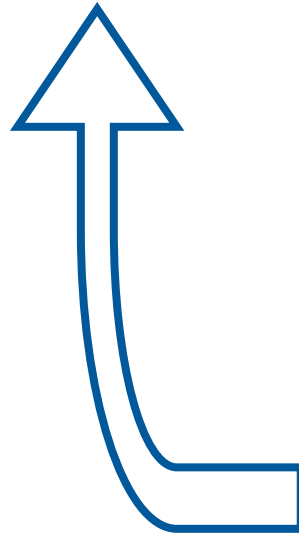- PCA cannot guarantee a separation of your groups within the first 2 or 3

   principle components

IZBI

Analysis – „PCA"

Insight:

- PCA cannot guarantee a separation of your groups within the first 2 or 3 principle components

- PCA can help validating the ability of a set of features (genes, gel-spots, metabolites) to separate the given groups

- remarkably better result while working with filtered metabolites

PCA : *can groups be recovered with a certain set of metabolites ?*

YES, with p-filter in first 3 principal components

Analysis – „Clustering"

Idea of Hierarchical agglomerative clustering:

- objective: unsupervised grouping within a similarity-structure,

  afterwards recovering groups in features and samples

- concept: beginning with n clusters and step-by-step putting together the 2

  most similar clusters

Analysis – „Clustering"

Example: Clustering of 5 people

- beginning with n clusters and step-by-step putting together the 2 most similar clusters
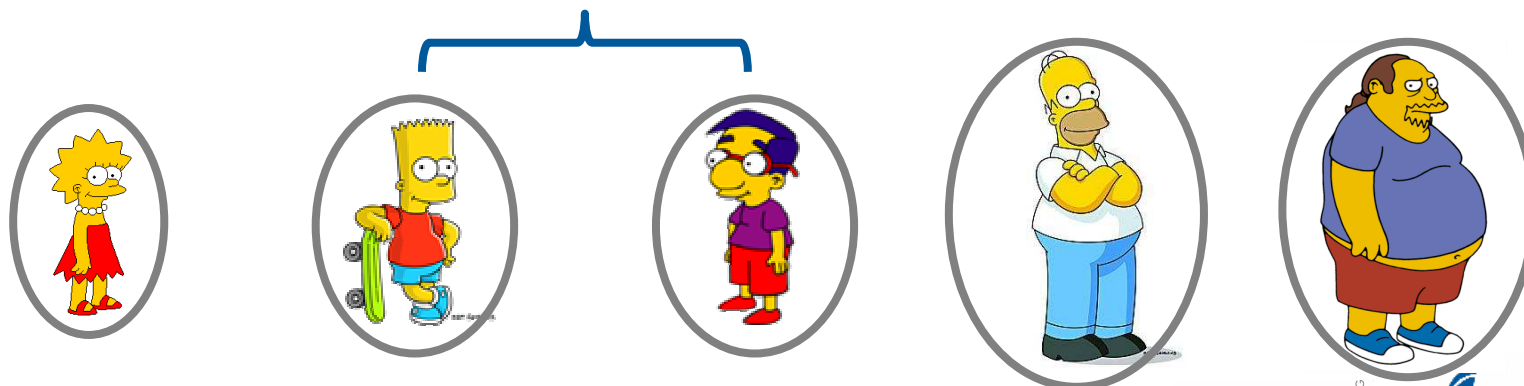
Analysis – „Clustering"

Example: Clustering of 5 people

- beginning with n clusters and step-by-step putting together the 2 most similar clusters
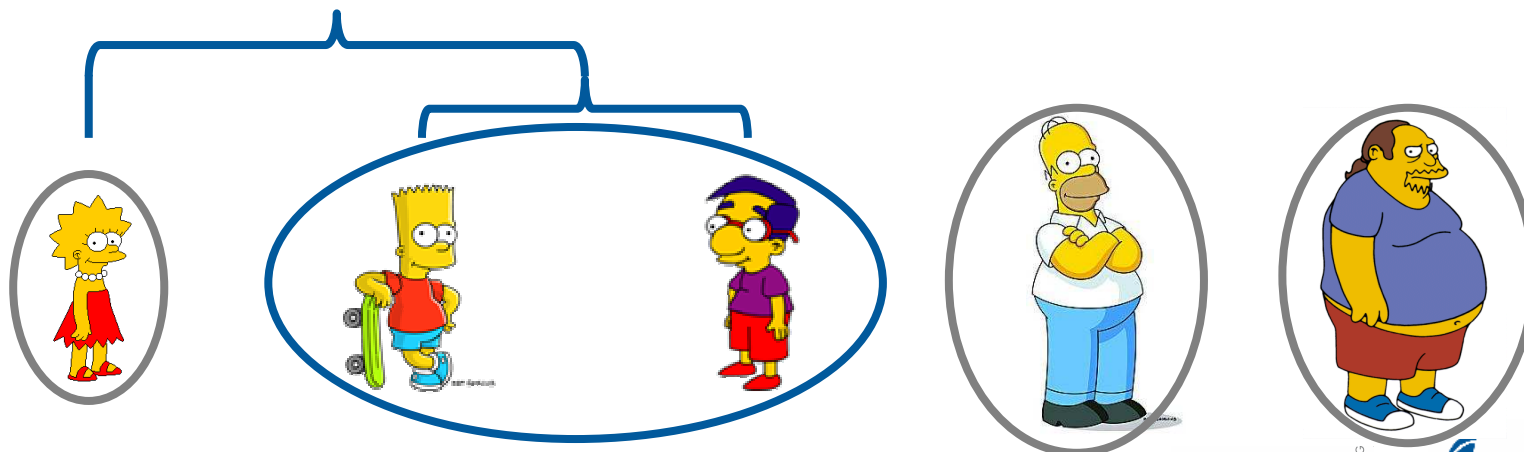
similarity-measurement:
weight

Analysis – „Clustering"

Example: Clustering of 5 people

- beginning with n clusters and step-by-step putting together the 2 most similar clusters
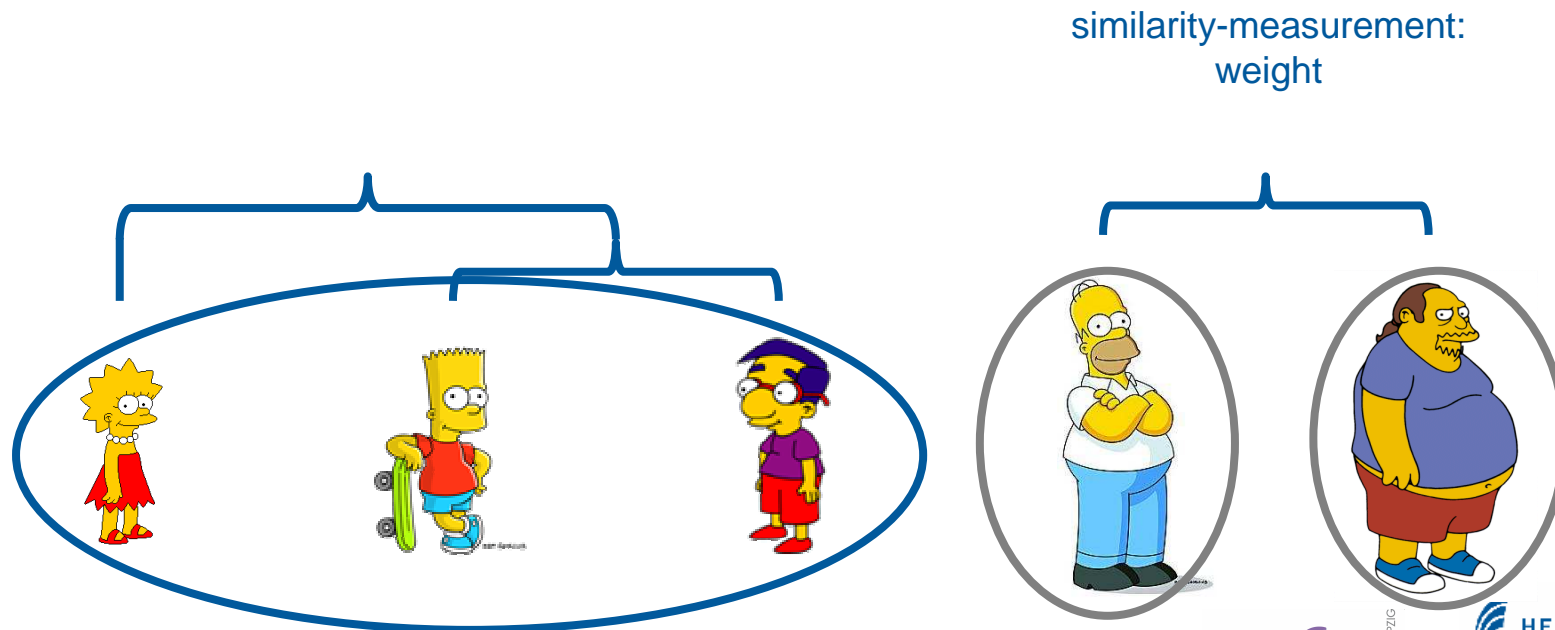
similarity-measurement:
weight

Analysis – „Clustering"

Example:

Clustering of 5 people

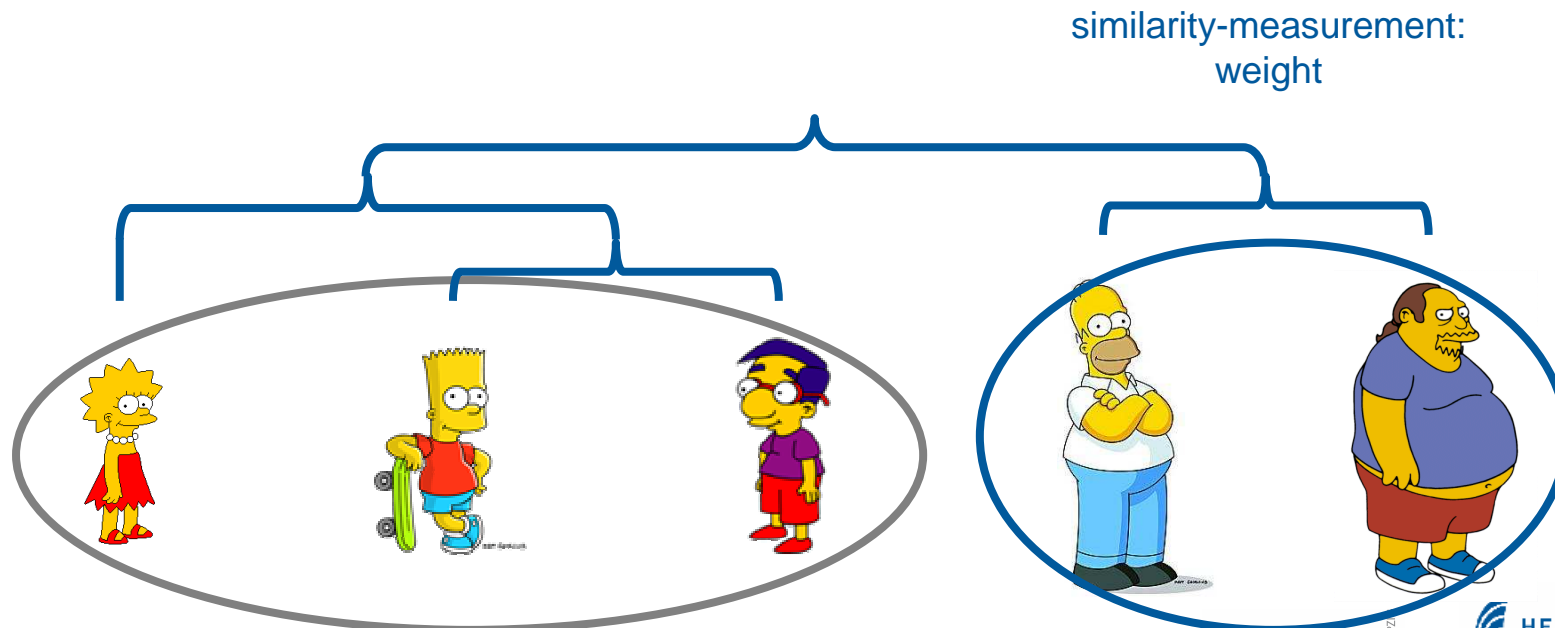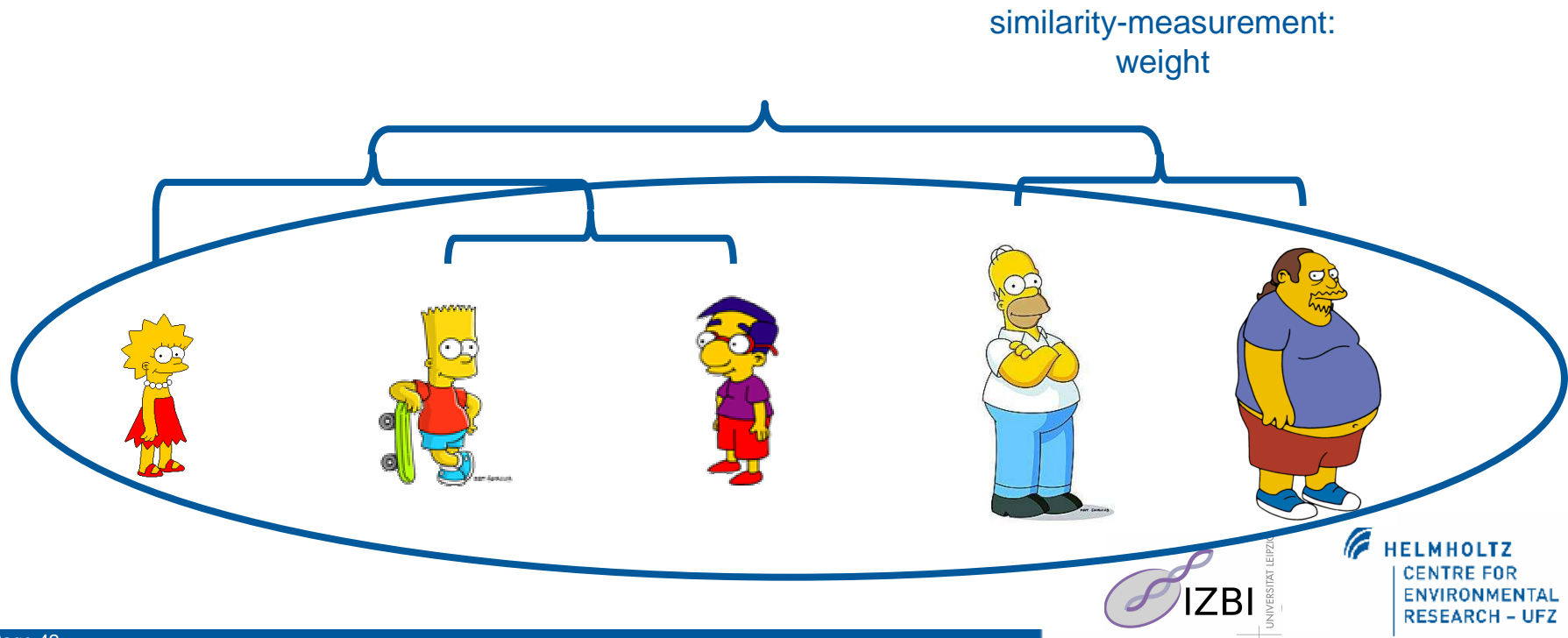- beginning with n clusters and step-by-step putting together the 2 most similar clusters

similarity-measurement:
weight

Analysis – „Clustering"

Example:

Clustering of 5 people

- beginning with n clusters and step-by-step putting together the 2 most similar clusters

similarity-measurement:
weight

Analysis – „Clustering"

Example: Clustering of 5 people

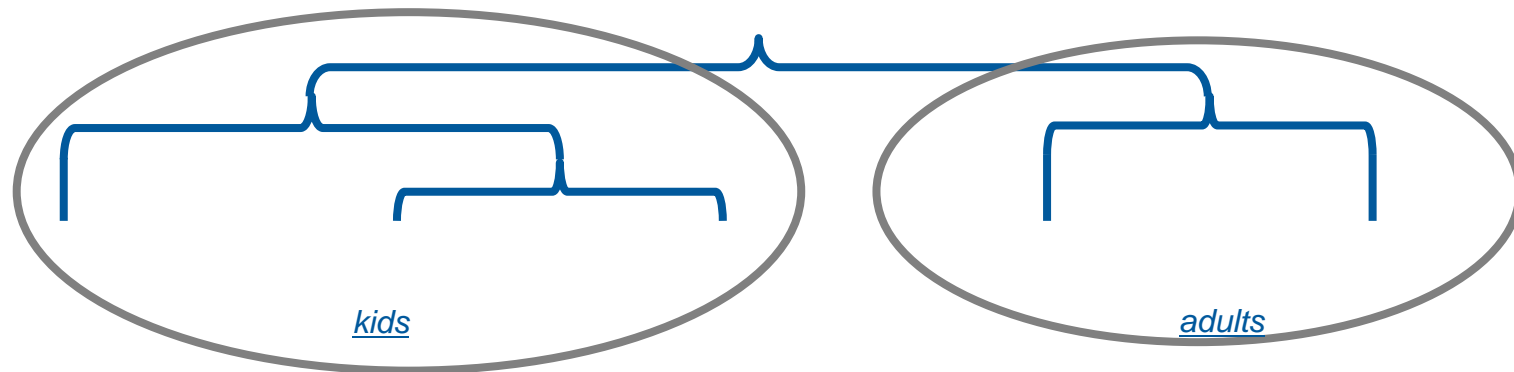- beginning with n clusters and step-by-step putting together the 2 most similar clusters

similarity-measurement:
weight

IZBI UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Analysis – „Clustering"

Example: Clustering of 5 people

- dendrogram shows a trace of the clustering process:

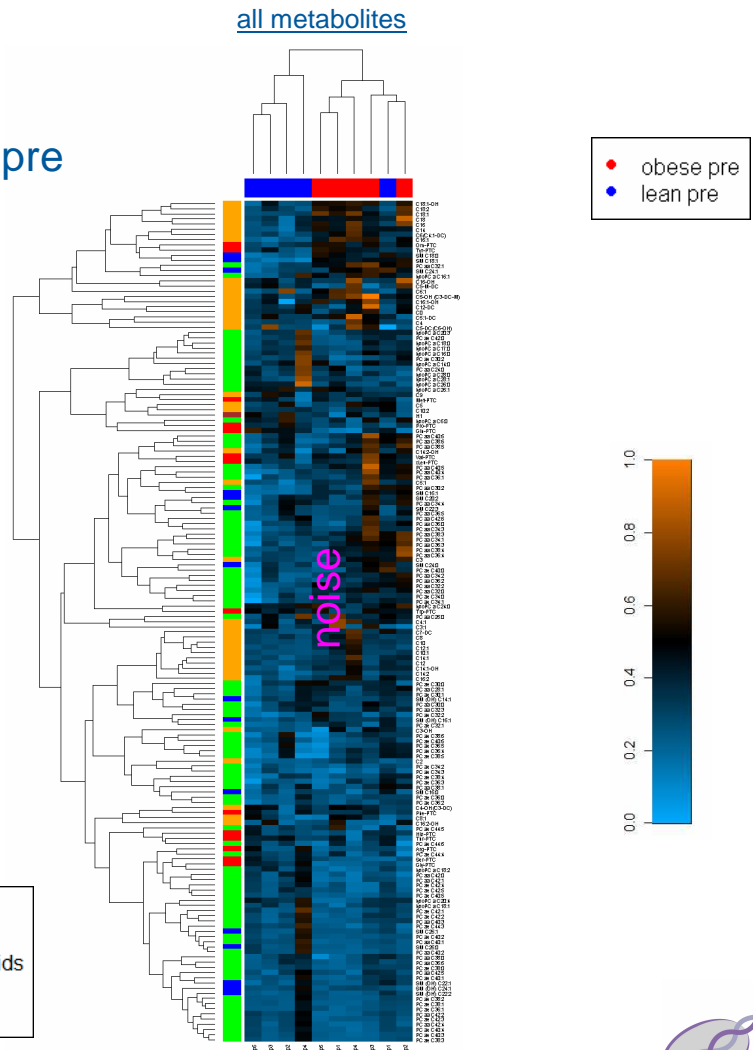_kids_   _adults_

here: the 5 people form 2 groups with regard to the similarity-measurement weigth

Analysis – „Clustering"

Results: Obese - Lean pre

Idea:
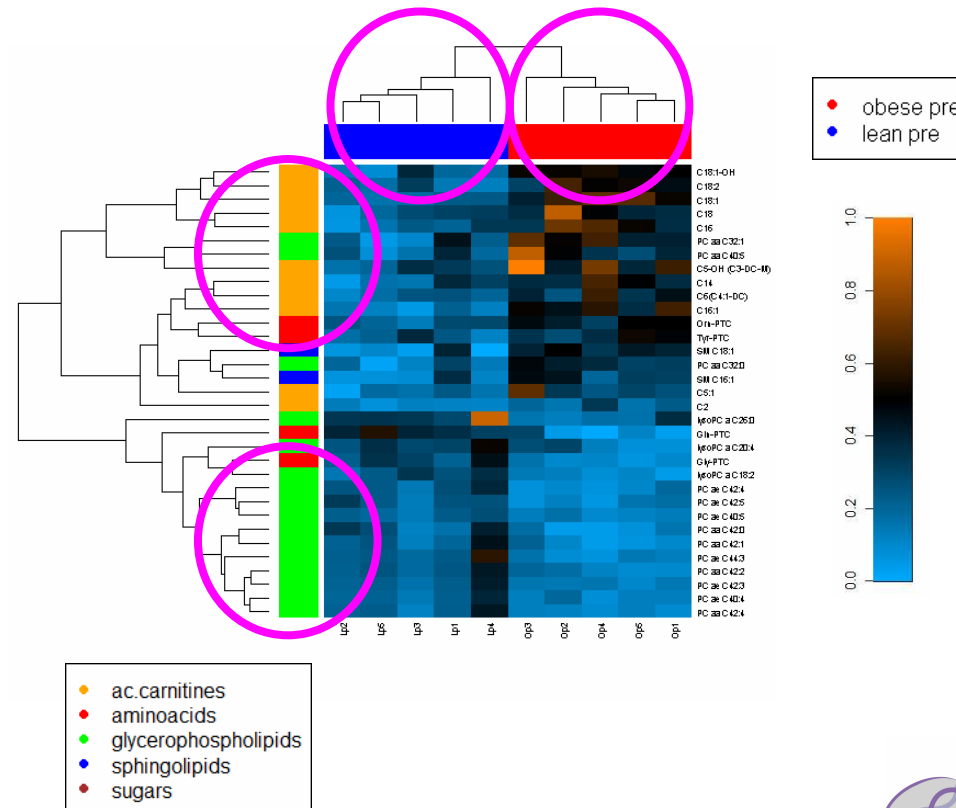
again only using
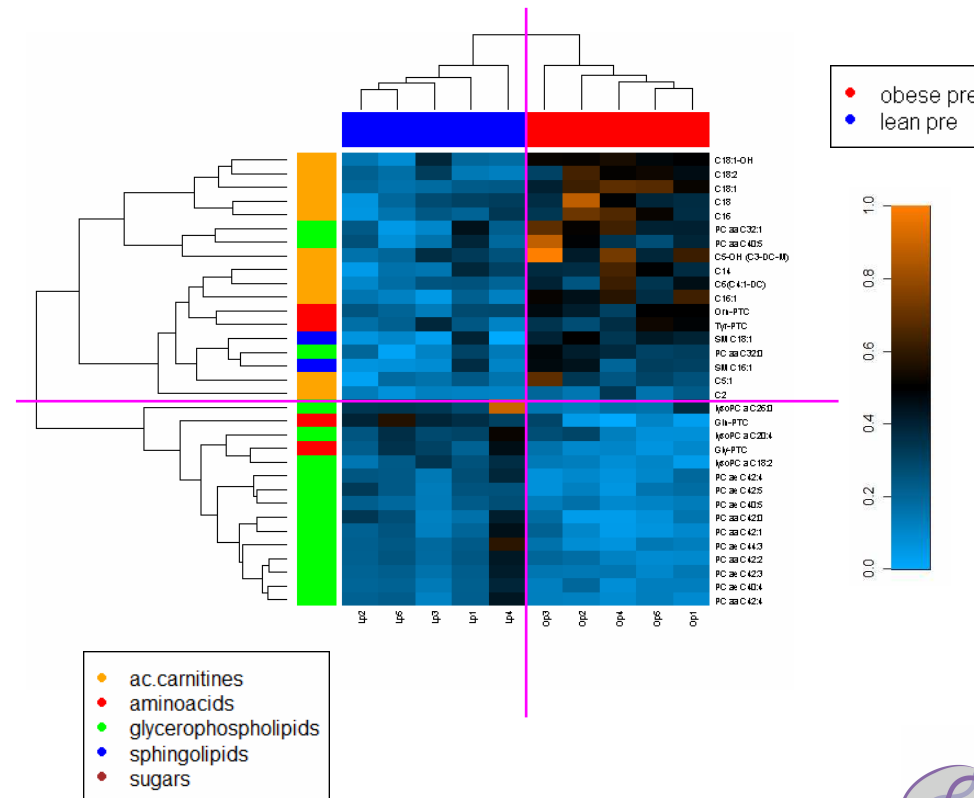significant
metabolites from
the volcano-plots

Analysis – „Clustering"

Results: Obese - Lean pre (significant metabolites)
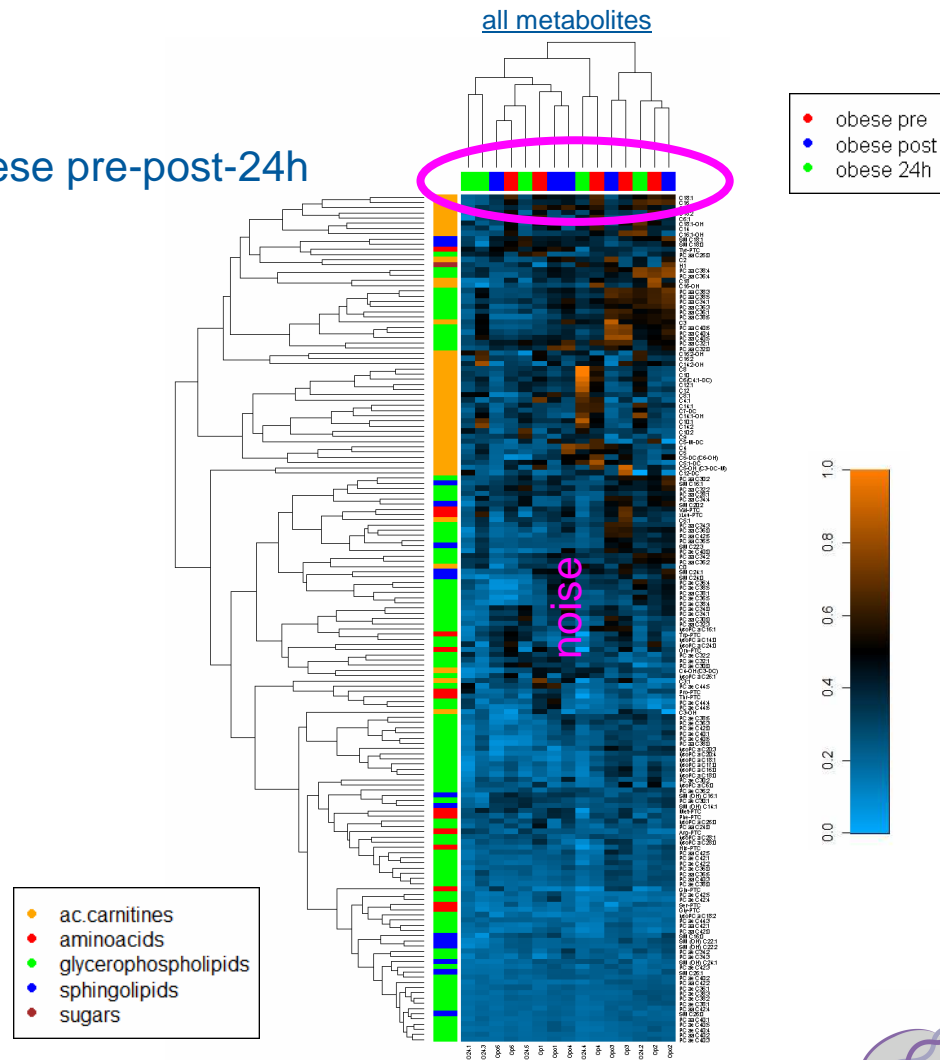
Analysis – „Clustering"

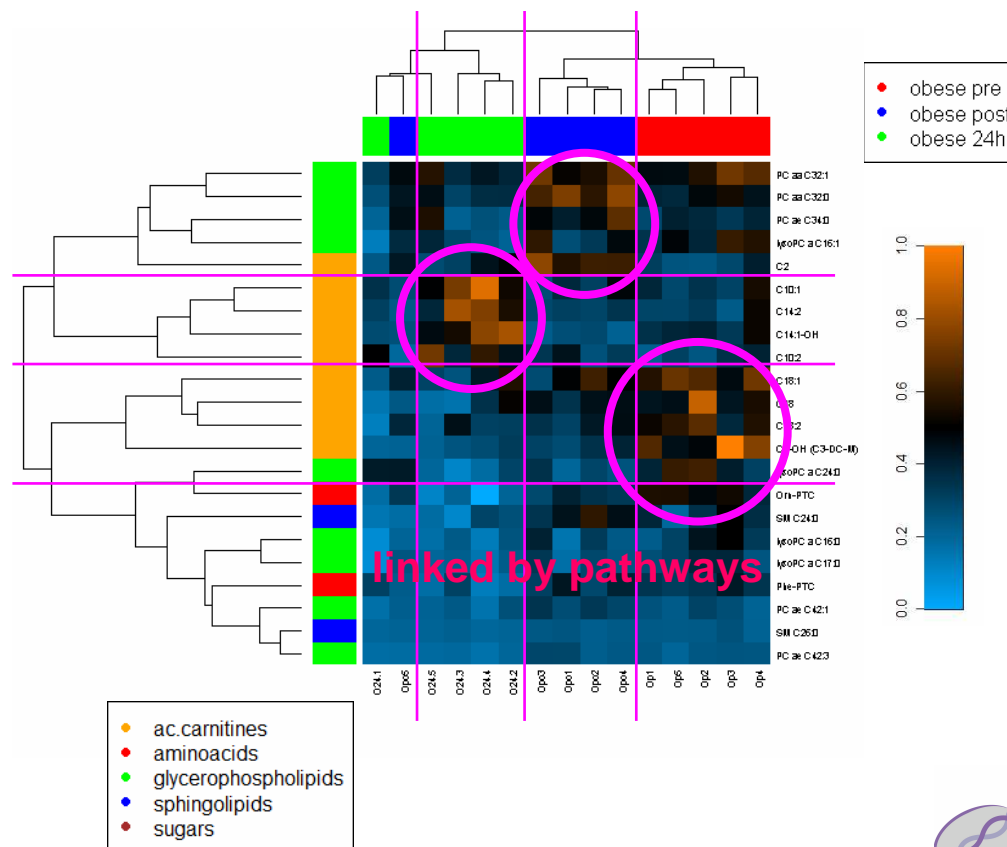Results: Obese - Lean pre (significant metabolites)

Analysis – „Clustering"
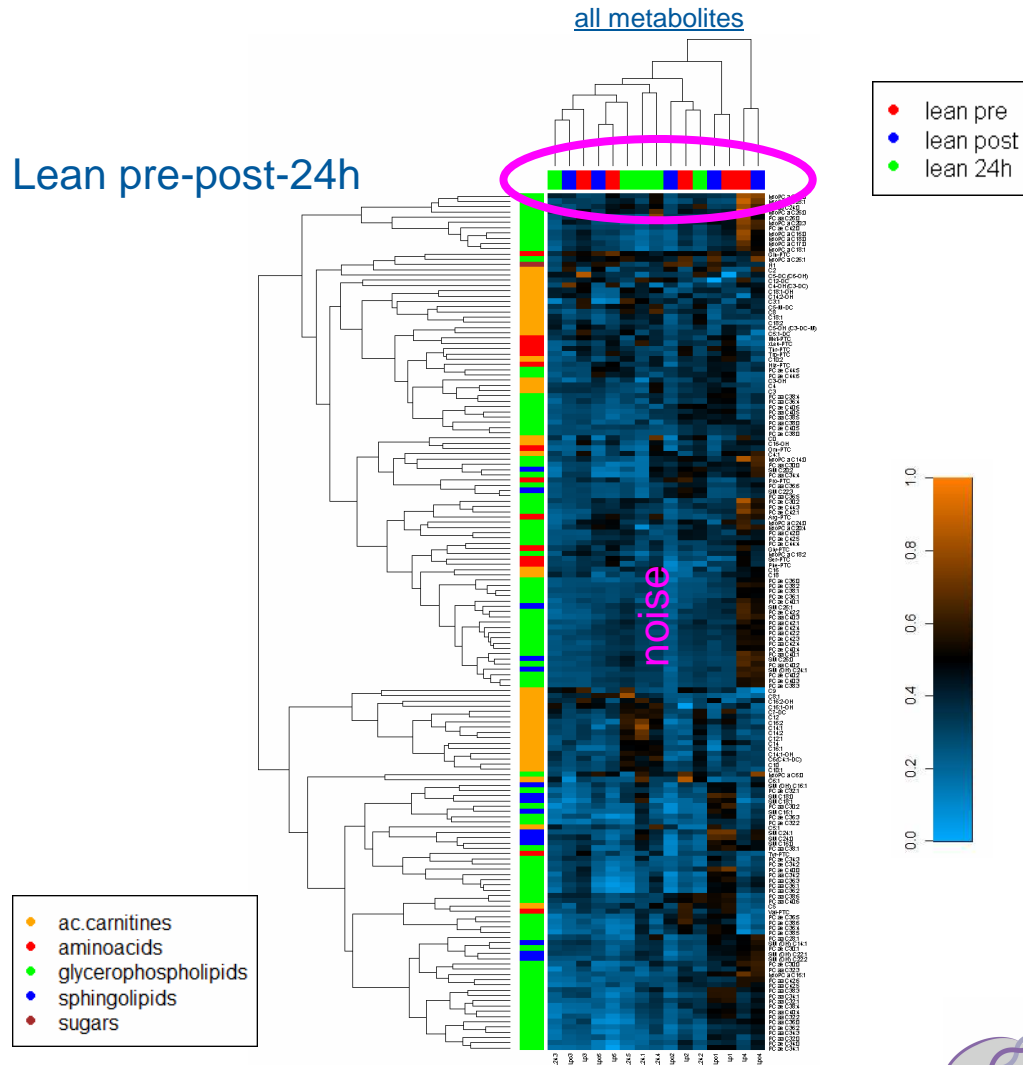
Results: Obese pre-post-24h

Analysis – „Clustering"

Results: Obese pre-post-24h (significant metabolites)
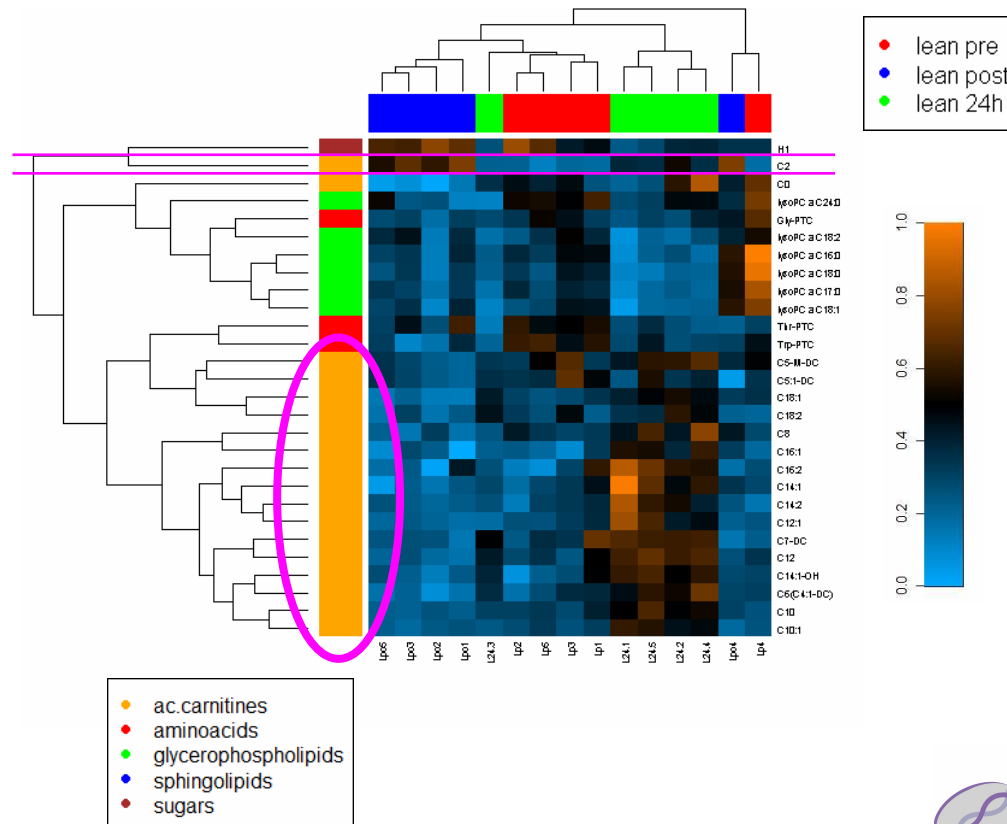
Analysis – „Clustering"

Results: Lean pre-post-24h

Analysis – „Clustering"

Results: Lean pre-post-24h (significant metabolites)

huge distance in dendrogram

Analysis – „Clustering"

Insight:

- many uninteresting metabolites (no change with regard to groups)

  → noise / bad group separation

- using only significant metabolites (p-value-filter) leads to clear structured

  heatmaps and good group separations

Clustering : *are there structures in a specific data(sub)set ?*

YES, [pathway linked] structures

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

Other Analysis Techniques

quality control: density analysis, scatterplots, M-A-Plots

significance analysis: ANOVA, Shrinkage-T-Statistics, SAM

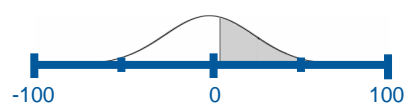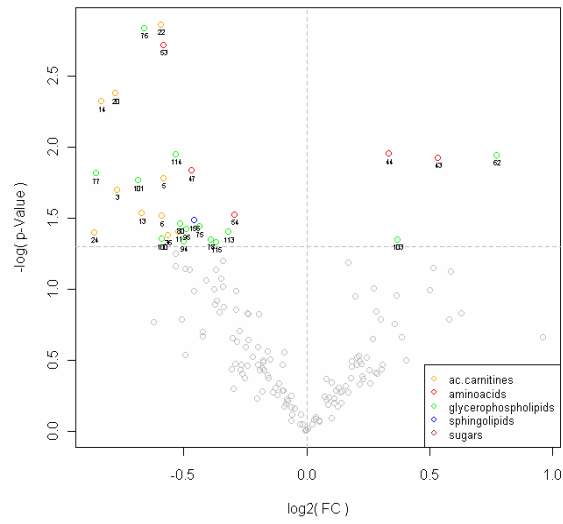discriminant analysis: ICA, SOM, PAM, RDA

cluster analysis: DCA

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ

IZBI
UNIVERSITÄT LEIPZIG

Outlook

„Trans-Omic": combining metabolomic data with proteomic data

Analysis: reproducing pathways and protein-metabolite-connections with an artificial neural net
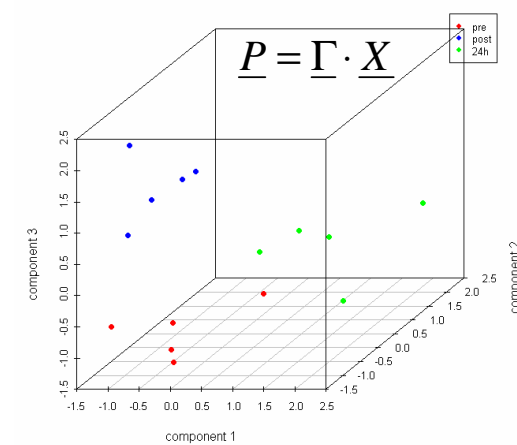
development of advanced t-tests adaptive to the datas nature
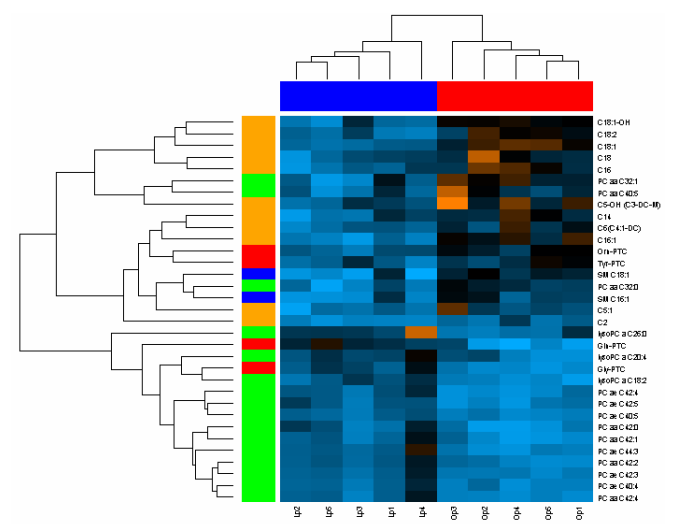
Don´t eat too much junk food

$$\underline{\Gamma} = \begin{pmatrix} \underline{\gamma}_A & \underline{\gamma}_B \cdots \underline{\gamma}_P \end{pmatrix} = \begin{pmatrix} \gamma_{1A} & \gamma_{1B} \cdots \gamma_{1P} \\ \gamma_{2A} & \gamma_{2B} \cdots \gamma_{2P} \\ \vdots & \vdots \\ \gamma_{PA} & \gamma_{PB} \cdots \gamma_{PP} \end{pmatrix}$$

$$\underline{P} = \underline{\Gamma} \cdot \underline{X}$$

$$d = \sum_{i=1}^{n} (x_i - y_i)^2$$

$$t = \sqrt{n}\,\frac{\overline{x} - \overline{y}}{\sqrt{(s_x^2 + s_y^2)}}$$

Thanks for your attention!

IZBI
UNIVERSITÄT LEIPZIG

HELMHOLTZ
CENTRE FOR
ENVIRONMENTAL
RESEARCH – UFZ