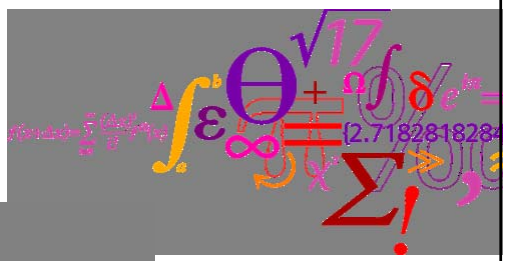
Biomarkers as a tool in general research

Lars I. Hellgren, Associate Prof.

Center for Biological Sequence Analysis

Dept. Systems Biology

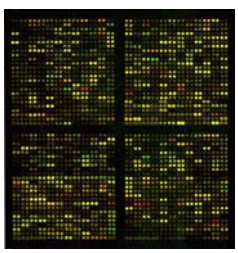
lih@bio.dtu.dk

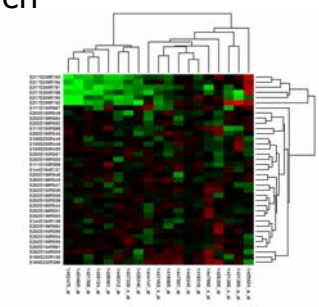


DTU Biosys
Department of Systems Biology

Content

- Biomarker - Definitions
- Classical biomarkers
- -omics techniques for establishing new "biomarker pattern"
- Some examples on this approach





Fatty acids and Metabolic Syndrome

Biomarkers - Definition

- A **biomarker** is a substance used as an **indicator** of a **biologic state**. It is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or **exposure** to a particular **environmental substance** (food component, environmental contaminant etc).
- A biomarker is normally assayed in blood, urine, faeces or other easily accesible material.

Biomarker types

- **Prognostic markers** –markers for increased risk of developing a pathological state
- **Disease marker –** markers for the presence of a pathological state
- **Exposure marker –** markers for exposure to a certain substance (food component etc)

Classical biomarkers - Examples

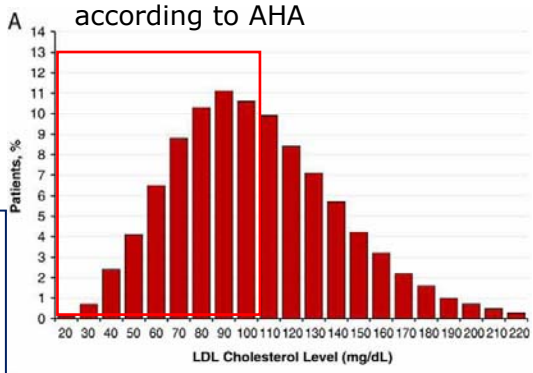
- **LDL-cholesterol** – prognostic marker for the risk of developing cardiovascular diseases
- **Fasting blood glucose** – marker for decreased insulin sensitivity/diabetes melittus
- **N-3 fatty acids in erythrocytes / adipose tissue** – marker of n-3 PUFA intake

Biomarker – limitations of single biomarkers

LDL-cholesterol in 136.000 patients hospitalized with coronary heart disease.

~ 50% of the hospitalized patients have LDL-cholesterol defined as optimal according to the American Heart Association

Optimal LDL concentration according to AHA



Sachdeva et al. 2009, American Heart Journal 157: 11-17

Add more markers … increase prognostic value …

C-reactive protein and LDL-cholesterol is minimally correlated, but both are excellent prognostic markers for the risk of cardiovascular events.

Ridker *et al*. 2002 N. Engl. J. Med. 347:1557.

Cohort of 27.000 persons, 8 years follow up.

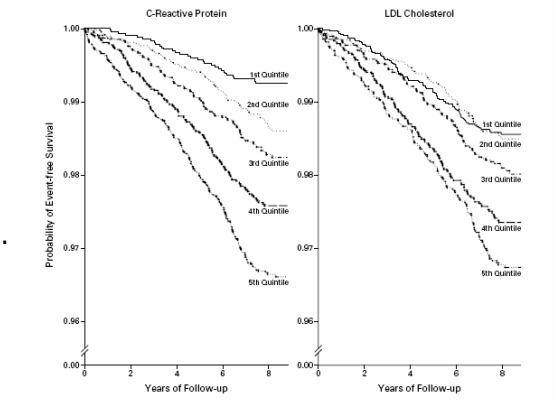


Figure 1. Event-free Survival According to Base-Line Quintiles of C-Reactive Protein and LDL Cholesterol.
The range of values for C-reactive protein was as follows: first quintile, <0.49 mg per liter; second quintile, >0.49 to 1.08 mg per liter; third quintile, >1.08 to 2.09 mg per liter; fourth quintile, >2.09 to 4.19 mg per liter; fifth quintile, >4.19 mg per liter. For LDL cholesterol, the values were as follows: first quintile, <97.6 mg per deciliter; second quintile, >97.6 to 115.4 mg per deciliter; third quintile, >115.4 to 132.2 mg per deciliter; fourth quintile, >132.2 to 153.9 mg per deciliter; fifth quintile, >153.9 mg per deciliter. To convert values for LDL cholesterol to millimoles per liter, multiply by 0.02586. Note the expanded scale on the ordinate.

Add more markers … increase prognostic value …

Ridker *et al*. 2002 N. Engl. J. Med. 347:1557.

Cohort of 27.000 persons, 8 years follow up.

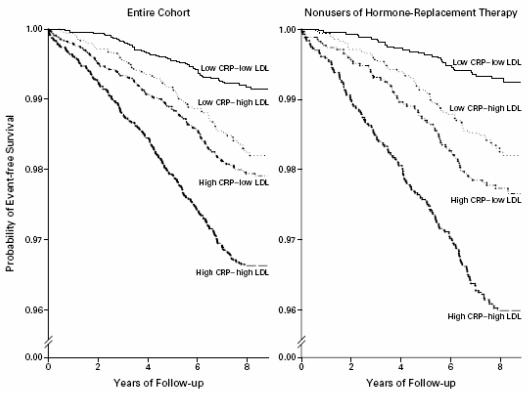


Figure 3. Event-free Survival among Women with C-Reactive Protein (CRP) and LDL Cholesterol Levels above or below the Median for the Study Population.
Data are shown for the entire cohort (27,939 women) and for women who were not taking hormone-replacement therapy at base line (15,745 women). The median values were as follows: for C-reactive protein, 1.52 mg per liter; for LDL cholesterol, 123.7 mg per deciliter (3.20 mmol per liter). Note the expanded scale on the ordinate.

Combining baseline LDL-cholesterol and C-reactive protein substantially increase prognostic power.

Life is multivariate….

- ~ 22,000 genes
- ~150,000 different RNA species (??)
- ~90,000 proteins
- ~50,000 (??) metabolites

Health or disease are the result of the integration of all these factors in each individual at each timepoint …

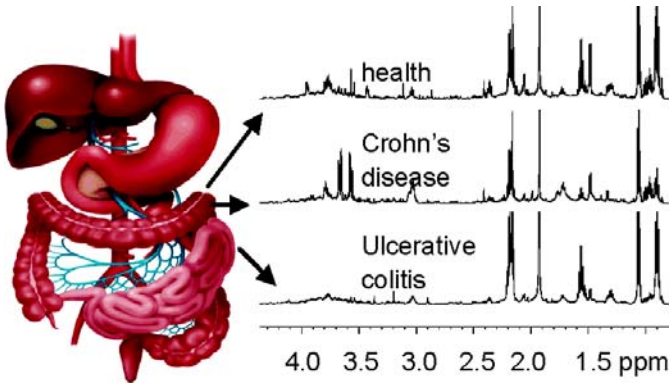
Life is multivariate….

- ~ 22,000 genes
- ~150,000 different RNA species (??)
  - **NB! The importance of non-coding RNA!**
- ~90,000 proteins
- ~50,000 (??) metabolites

Health or disease are the result of the integration of all these factors in each individual at each timepoint …

DTU Biosys, Technical University of Denmark

Multivariate biomarker – Pattern of changes…



Picture from Marchesi *et al.* 2006

"Multivariate" biomarker – Pattern of changes…

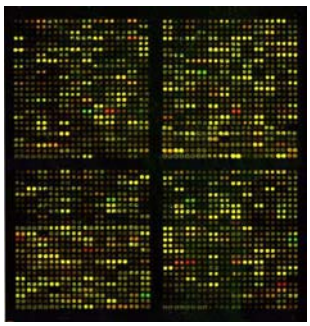
- **The "–omics revolution"**
  - **Genomics**
    - Epigenetic modifications of the genome → Altered disease susceptibility?
  - **Transcriptomics**
    - The patterns of expressed genes and non-coding DNA
  - **Proteomics**
    - The total pattern of proteins in a certain cell type under certain circumstances (sub-field; phospho-proteomics, glyco-proteomics etc)
  - **Metabolomics / Metabonomics**
    - The pattern of metabolites present in a certain cell type under certain circumstances (sub-fields; glycomics, lipidomics etc…)

Transcriptomics

- Map the non-rRNA expressed under specific conditions.

Methods: Chip-array

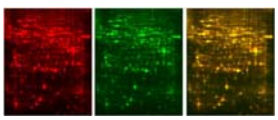
Quantify expression of both coding and non-coding parts of the genome



But note: 't Hoen et al, 2008 "Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms" Nucleic Acid Research 2008:36 e141

Proteomics

Techniques: 2-D gel electrophoresis combined with matrix-assisted laser desorption/ionization mass-spectrometry. (MALDI)

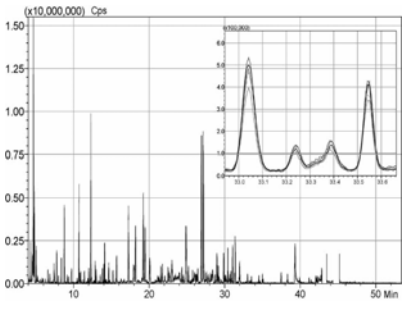


Typical experimental flow:

1. Isolate proteins from different samples
2. Label different samples with fluorescent probes emitting at different wavelength
3. Cut out proteins that are present in different amounts
4. Identify the proteins through MALDI-MS

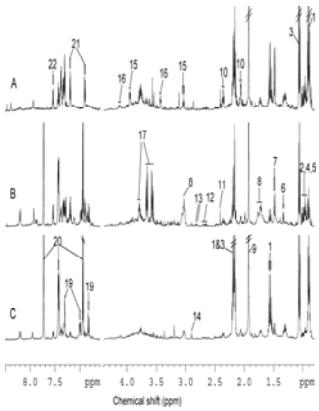
Metabonomics / Metabolomics

- **Techniques:** GC-MS, LS-MS/MS, NMR



GC/MS chromatogram of rat urine (MS in EI-mode, full scan m/z 50 – 650.

Pasikanti *et al* (2008) Rapid Com. Mass. Spect. 22:2984

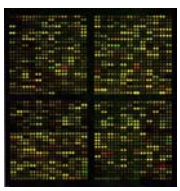


1H-NMR spectra of faeces samples.

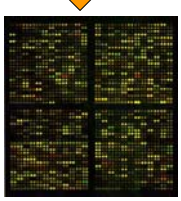
Marchesi *et al. (2007)*

The challenge – How to see unique patterns in the forest of data?

- Normalization, filtering out noise
- Mine the huge dataset for patterns that separate the studied groups from each other.
- Identify the factors in the dataset, causing the group -separation
- Unbiased construction of correlation networks that can be used as "biomarking-patterns"

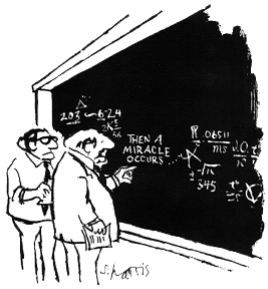


Is there a unique pattern separating the two treatments?



The challenge – How to see unique patterns in the forest of data?

- Normalization, filtering out noise
- Mine the huge dataset for patterns that separate the studied groups from each other.
- Identify the factors in the dataset, causing the group -separation
- Unbiased construction of correlation networks that can be used as "biomarking-patterns"



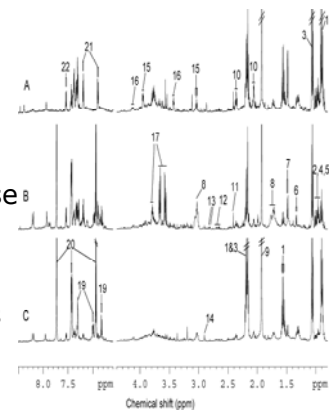
"I think you should be more explicit here in step two."

Example 1 – Metabonomic markers of IBD

Healthy

Crohn´s disease patients

Ulcerative colitis patients

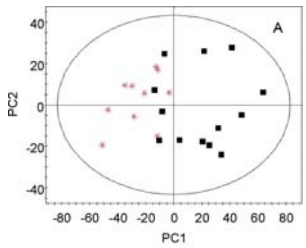


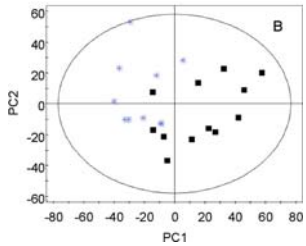
1H-NMR spectra of aqueous extract of fecal samples

Marchesi *et al*. *J. Proteome Res.* **2007,** 6, 546-551.

Example 1 – Metabonomic markers of IBD

- Principal component analysis (PCA) was performed on the normalized spectra and a score plot was generated





Healthy controls (black boxes) Crohn's disease (red stars) (A) control volunteers (black boxes) and those with ulcerative colitis (blue stars) (B). Each dot is one sample.

Marchesi *et al*. *J. Proteome Res.* **2007,** 6, 546-551.

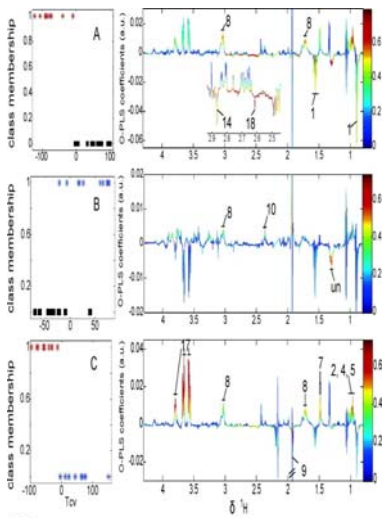
Example 1 – Metabonomic markers of IBD

- Orthogonal-Partial Least Square-Discriminant Analysis (O-PLS-DA) to identify metabolites contributing to the separation between the groups.

metabolite	CD vs Control	UC vs Control	CD vs UC
	R^2 = 0.92	R^2 = 0.77	R^2 = 0.93
	Q^2 = 0.80	Q^2 = 0.52	Q^2 = 0.70
acetate	−0.67		−0.50
alanine	+0.64		+0.72
butyrate	−0.64	−0.36[b]	
glutamate		+0.68	
glycerol			+0.82
isoleucine	+0.69		+0.66
leucine	+0.74		+0.74
lysine	+0.72	+0.62	+0.67
methylamine	−0.73	−0.39	
trimethylamine	−0.52	−0.35	
unknown (1.33)	−0.82	−0.69	
valine	+0.66		+0.71

R^2 = total explained variation by the model

Q^2 = predictability of the model

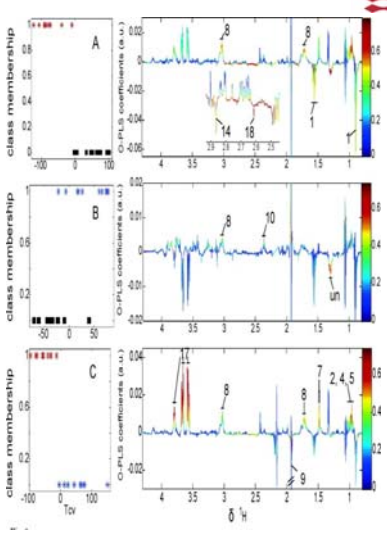


Example 1 – A "biomarker pattern" of IBD

metabolite	CD vs Control	UC vs Control	CD vs UC
	R^2 = 0.92	R^2 = 0.77	R^2 = 0.93
	Q^2 = 0.80	Q^2 = 0.52	Q^2 = 0.70
acetate	−0.67		−0.50
alanine	+0.64		+0.72
butyrate	−0.64	−0.36[b]	
glutamate		+0.68	
glycerol			+0.82
isoleucine	+0.69		+0.66
leucine	+0.74		+0.74
lysine	+0.72	+0.62	+0.67
methylamine	−0.73	−0.39	
trimethylamine	−0.52	−0.35	
unknown (1.33)	−0.82	−0.69	
valine	+0.66		+0.71

R^2 = total explained variation by the model

Q^2 = predictability of the model



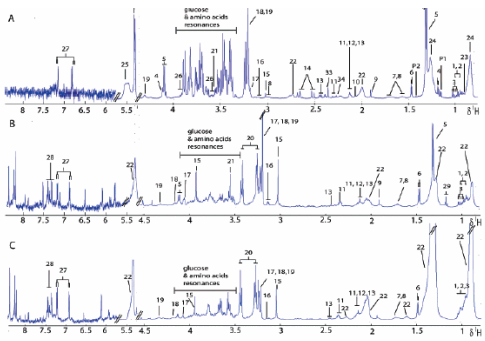
Example 2- Effects of L. paracasei on a model of IBD in mice

Interactions of *Trichinella spiralis* Infection with Dietary *Lactobacillus paracasei* Supplementation. *Martin et al. (2006) Journal of Proteome Research 5:2185*

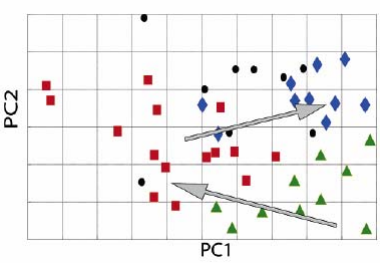
^{1}H NMR CPMG spectrum of control plasma

^{1}H MAS NMR CPMG spectra of jejunal wall

^{1}H MAS NMR CPMG spectra of jejunal longitudinal myenteric muscle (LMM)



Example 2- Effects of L. paracasei on a model of IBD in mice



PCA scoreplot of ^{1}HNMR CPMG spin echo spectra of plasma in infected mice and infected mice treated with *L. paracasei.*
Green = healthy
red = infected
blue=infected treated with *L. paracasei*
Black=treated with spent culture medium

Martin et al. (2006) Journal of Proteome Research 5:2185

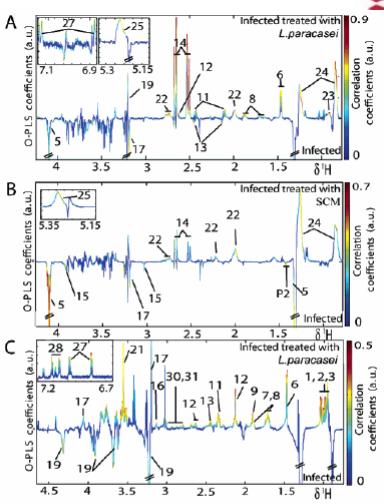
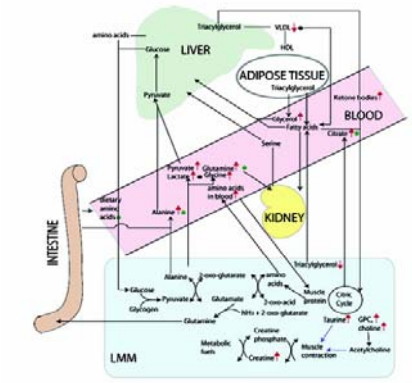


Figure 4. O-PLS-DA coefficient plots derived from ^{1}H NMR CPMG spectra of plasma (**A**, **B**), and ^{1}H MAS NMR CPMG spectra of jejunal wall (**C**), illustrating the discrimination between untreated infected mice (negative) and infected mice (positive) treated either with *L. paracasei* in culture medium (**A**, **C**) or with SCM (**B**). The color code corresponds to the correlation coefficients of the variables. The key is given in Table 1.

Example 2- Create a metabolic model explaining the results



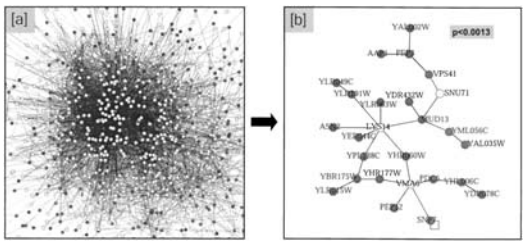
Effects of *T. spiralis infection on* major fuel and amino acid flows between organs and amino acid interconversions in LMM.

Effects of L. *paracasei* is indicated with dots.

Metabolic patterns that might be developed into a biomarker.

Creating networks of correlation between different biomolecules

The open source software platform "Cytoscape" can be used to construct networks of correlating biomolecules.



Shannon P. et.al. Genome Res. 2003;13:2498-2504

Data from different "-omics" analyses can be pooled and used to visualize correlations

www.cytoscape.org