

Linking medicine to metabolomics and proteomics

Erkki Karjalainen
Dept. Clin. Chem, University of Helsinki
11.6.2008 HDR08

Linking...

- Introduction
- The machinery of human body
- Medicine and molecules
- The promise of metabolomics
- Conclusions

Introduction

- Computers 100 x cheaper/decade
- Instruments 10 x more data/decade
- Diagnoses ? x better/decade

The *omics boom

- Genomics
- Proteomics
- Metabolomics
- Diagno*#!mics?



- ### The machinery...
- Robust function by redundancy
 - Critical parts are duplicated
 - Alternative metabolic pathways
 - Self-repairing system

- ### Medicine and molecules
- Clinical chemistry confirms a diagnosis
 - Wasteful testing practices
 - Redundancy is not understood
 - Multivariate interpretation needed

- ### The role of metabolomics
- An overabundance of raw data
 - The bottlenecks in data analysis
 - The importance of calibration
 - Diagnosis by metabolomics?

An overabundance of raw data

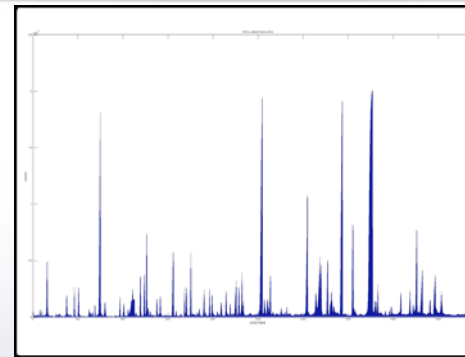
- Hard disks fill up rapidly
- Uncompressed data cannot be kept
- Compression destroys essential info

The bottlenecks in data analysis

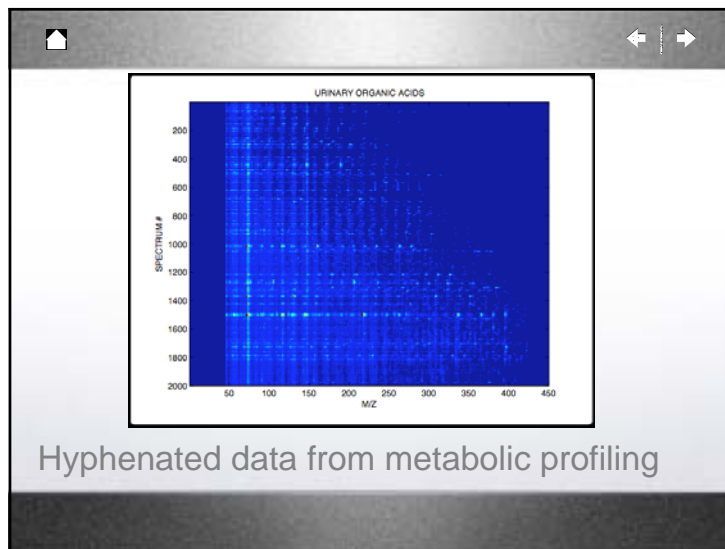
- One sample requires minutes of CPU
- The problem of scaling up
- Possible shortcuts
- A speedup of 100 is 10 years



GC-MS (Agilent MSD)



Metabolic profiling



SEARCH INSIDE™

Data Analysis for Hyphenated Techniques

by E.J. Vandecasteele (Author), M.P. Koelmans (Author)

No customer reviews yet. [See the first.](#)

Available from these sellers.

4 used & new available from \$157.86

Share your own customer images
Search inside this book

Are You an Author or Publisher?
Find out how to publish your own Kindle books

Open source attempt...

- ### The deconvolution process
- Current procedures very local
 - AMDIS traps each peak separately
 - General inversion possible - very slow
 - Topdown OSCAR faster but still too slow

- ### The scaling problem
- Bottom-up spectra are from one sample
 - Multi-sample statistics not in libraries
 - Statistical power of multiple samples?

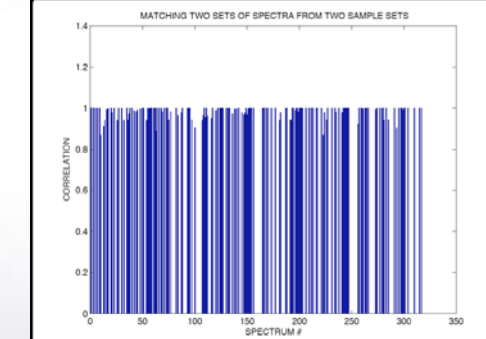
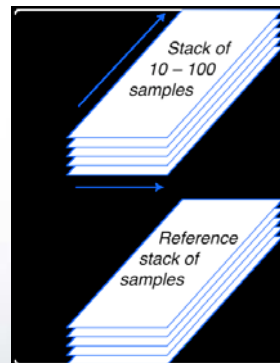
A new top-down approach

- Several samples processed as an entity
- CPU time is seconds, not minutes
- Spectra must be repeatably found in several samples
- The rest are stored in “once” database

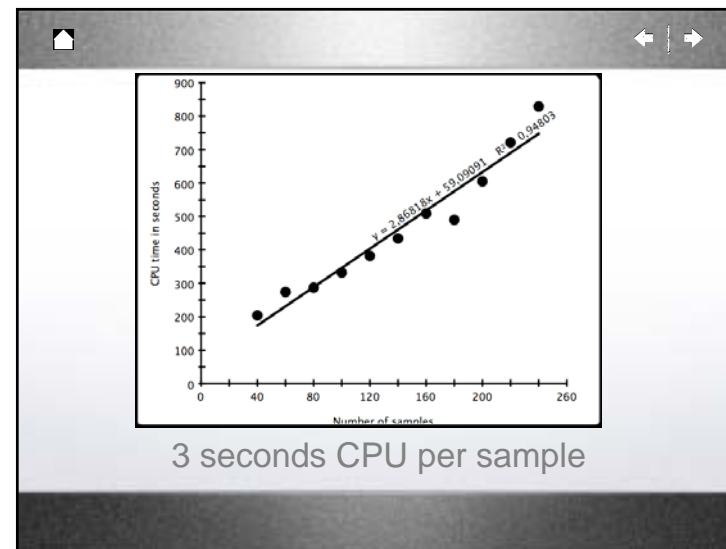
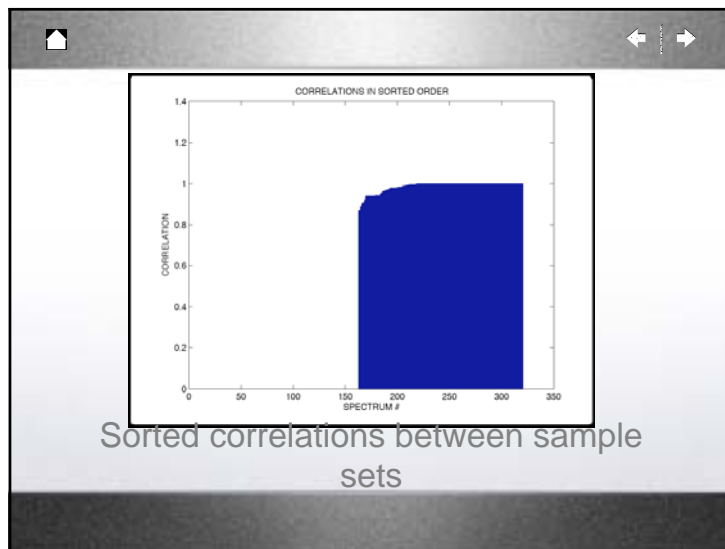
GC-MS runs are time-corrected

The new process

- Spectra in stacks of several GC-MS runs are isolated
- Spectra compared between both stacks
- Time-matched identical spectra are accepted into database of repeatable spectra



Matching spectra from two sample sets



- ### The good and the bad...
- + The system works in a hands-off fashion
 - + Spectra pool data from several samples
 - - Many samples needed for libraries!
 - - Linking back to instruments





Mac mini and Mac "maxi"

The role of calibration

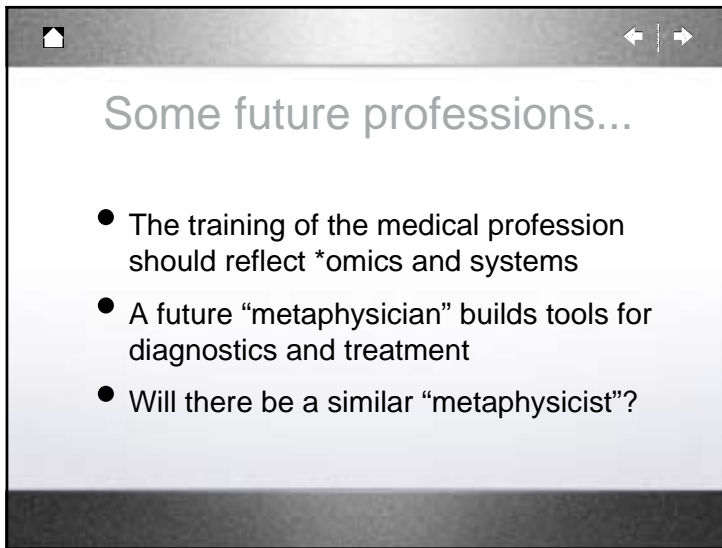
- Diagnosis by finding similar cases
- Global raw data storage
- Comparison needs a stable calibration
- Mechanisms can be inferred later

Cost-effective diagnosis

- Metabolic state must be reflected in data
- Total cost of analysis must be low
- Quality assurance by trained personnel
- Reducing missed diagnoses by profiling

Conclusions

- Metabolomics is not ready yet
- Total cost per sample must be lowered
- Raw data and calibrations must be pooled
- "Googling" raw data can suggest diagnoses



Some future professions...

- The training of the medical profession should reflect *omics and systems
- A future “metaphysician” builds tools for diagnostics and treatment
- Will there be a similar “metaphysicist”?