

Pathway search engine for proteomics - Dream, fact or fiction?

Mikhail M. Savitski, Y.M. Eva Fung,
and Roman A. Zubarev

Uppsala University
www.brimis.uu.se

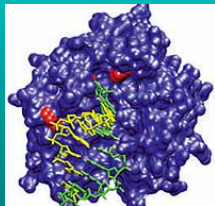
Leigh Anderson:
Critical Role of Biomarker Verification

- March 2008: 0 (zero) biomarkers discovered by proteomics
- Rate of FDA approval of (non-proteomics discovered) protein biomarkers dropped below 1/year

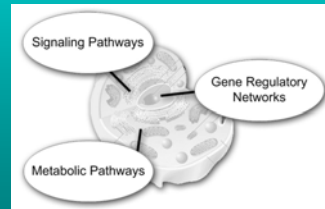
Reductionism is a dead end?...

Two paradigms in Biology

Reductionist
Molecular Biology:



Pathway Biology:



All systems go...

... *reductionism* - the idea that things can best be understood by reducing them to their smallest components, [molecules]. This turned out to be immensely useful in chemistry and physics...

In biology, however, the idea has its limits. The Human Genome project ... was a triumph of reductionism. **But merely listing genes does not explain how they collaborate to build and run an organism.** Nor do ... biological molecules give full insight into the causes and development of diseases.

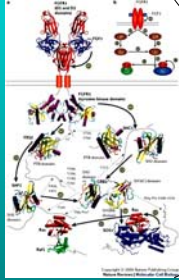
A complete understanding of biological processes means putting the bits back together again - and that is what systems biologists are trying to do, by using the results of a zillion analytical experiments to build software models of living organisms.

The pharmaceutical industry stands to gain much from this approach...

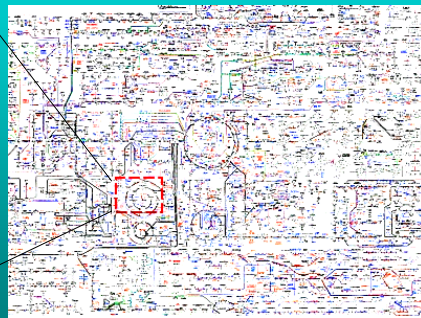
The Economist, Oct 27-Nov 2, 2007

Paradigm: Cell is a collection of pathways

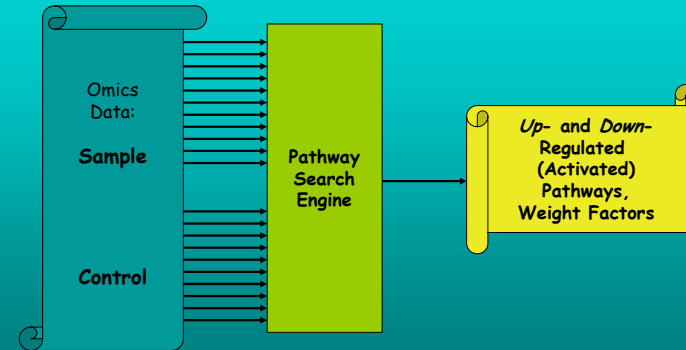
Fibroblast-growth-factor signaling pathway



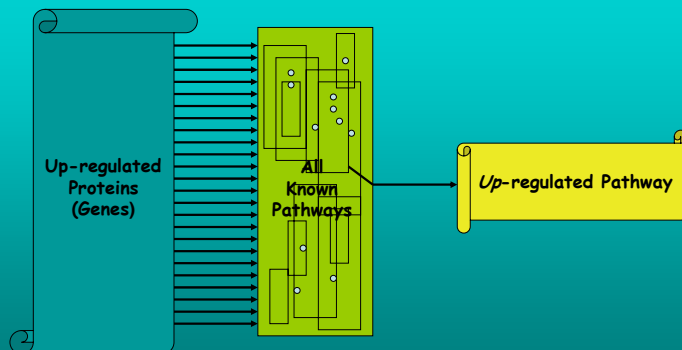
Roche Applied Science "Biochemical Pathways"



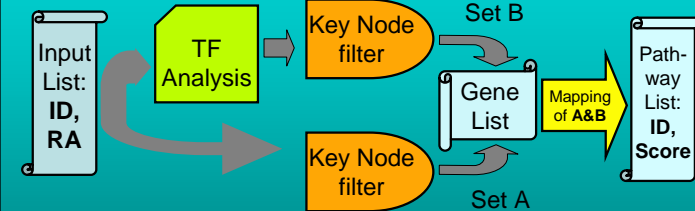
Analytical Pathway Biology



Direct Mapping (Enrichment Analysis)



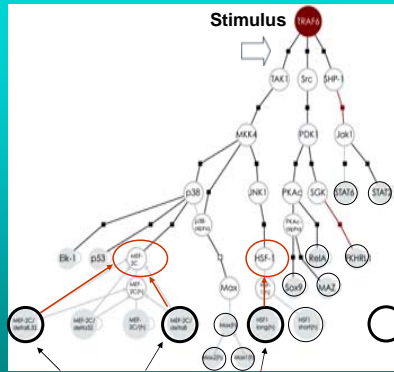
Pathway Analysis Workflow



Zubarev et al., J. Proteomics, 2008

Database: BioBase

KeyNode-Mediated Analysis: Upstream



Up-regulated
Proteins
Observed

Score

KeyNode₁ 3050

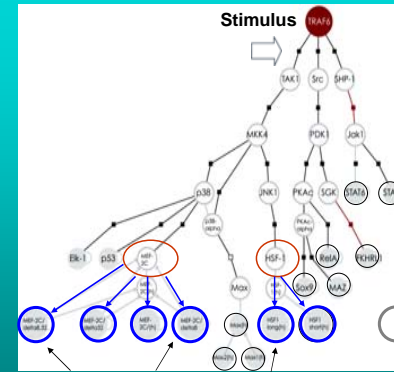
KeyNode₂ 2987

KeyNode₃ 2073

...

KeyNode_N 25

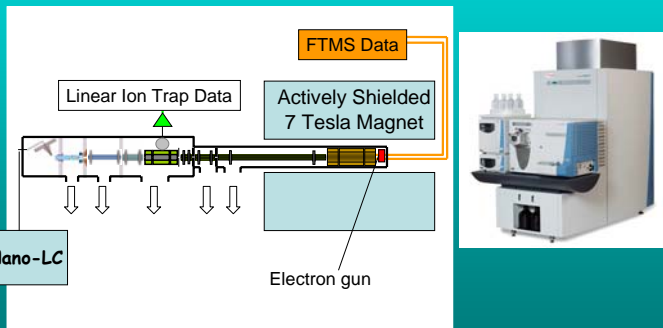
KeyNode-Mediated Analysis: Downstream



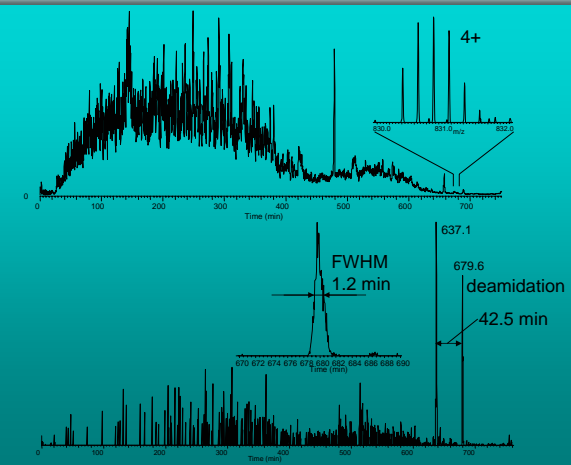
Proteins
Deduced

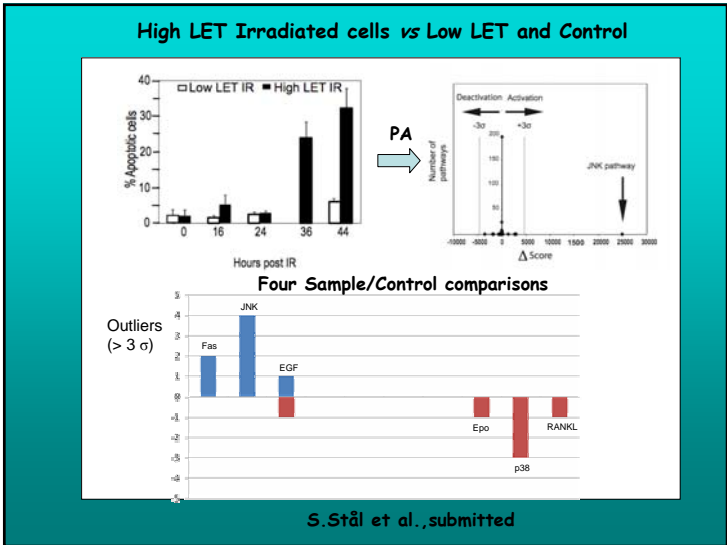
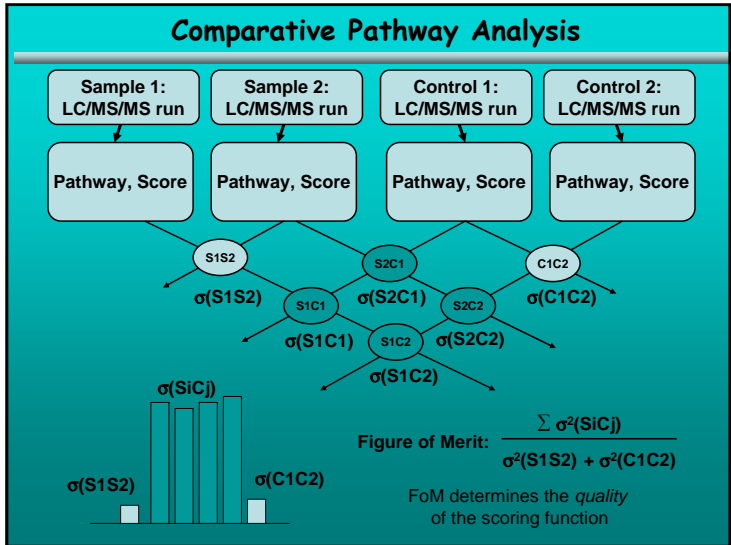
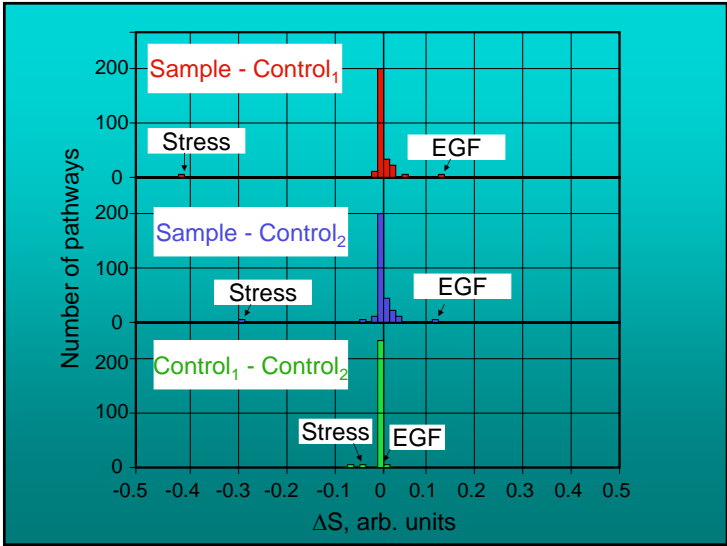
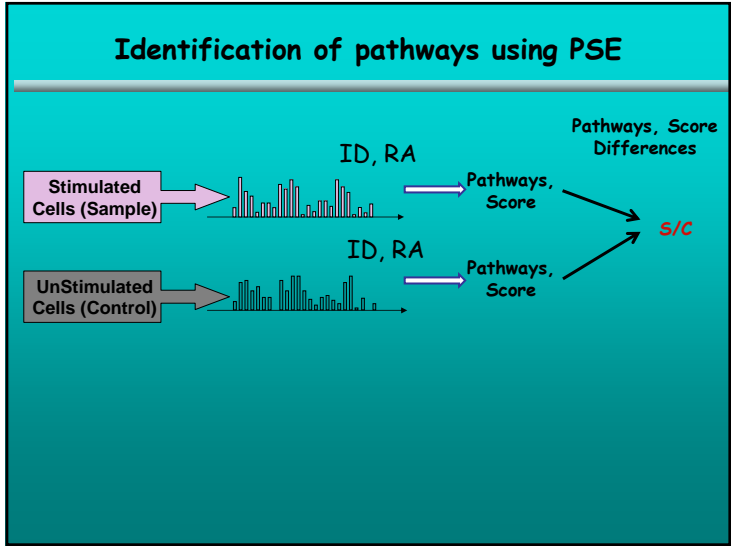
This works!

LTQ FT Fourier Transform Ion Cyclotron Resonance (FTICR) Mass Spectrometer



Full proteome of K562 cell line in a single LC/MS/MS experiment





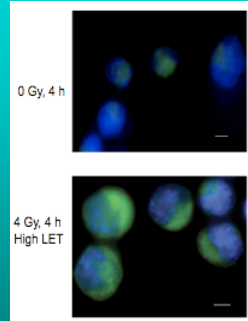
Pathway Analysis: Validation

Validation of XYZ pathway



Pathway marker

With Angela Bachi



S. Stål et al., submitted

Pathway Analysis: Criticism

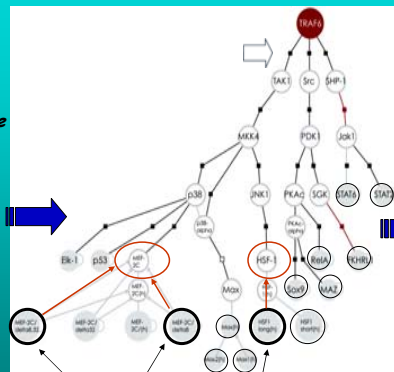
- In some cases, activated pathways are known from previous studies
- PSE can only identify pathways that *are* in the database

KeyNode analysis:

What key molecules are responsible for the observed proteome change?

Key Node Analysis: Proteins to KeyNodes

Dataset
Protein Abundance
IPI... 10752
IPI... 252
IPI... 171353
...
IPI... 672



KeyNode Score
KN₁ S₁
KN₂ S₂
KN₃ S₃
...
KN_N S_N

Proteins Observed

Key Node Analysis: Pairwise comparison

Control	Sample	Rank Shift	Sample vs Control	New Rank
KeyNode A	KeyNode C	+2	KeyNode C	1
KeyNode B	KeyNode A	-1	KeyNode D	2
KeyNode C	KeyNode D	+1	KeyNode A	3
KeyNode D	KeyNode B	-2	KeyNode B	4

Sorting →

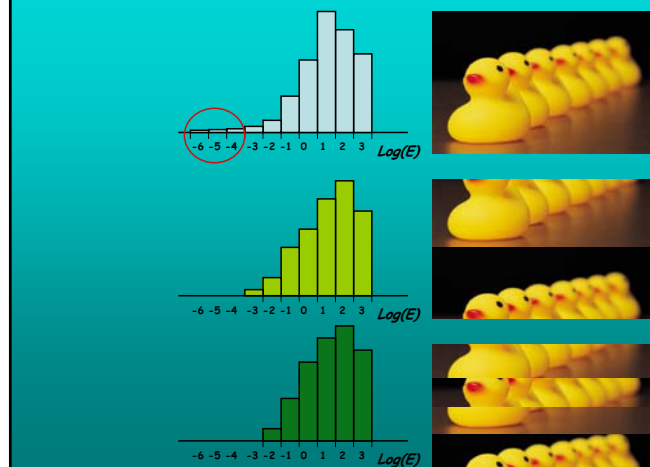
Key Node Analysis: E values

Sample 1 vs Control 1	Sample 1 vs Control 2	Sample 2 vs Control 1	Sample 2 vs Control 2	E-value	Sorting
KeyNode A	KeyNode B	KeyNode A	KeyNode D	$4.8 \cdot 10^{-8}$	KeyNode A
KeyNode B	KeyNode A	KeyNode B	KeyNode B	$8 \cdot 10^{-9}$	KeyNode B
KeyNode C	KeyNode C	KeyNode D	KeyNode A	$6 \cdot 10^{-9}$	KeyNode D
KeyNode D	KeyNode D	KeyNode C	KeyNode C	$1.4 \cdot 10^{-7}$	KeyNode C
...

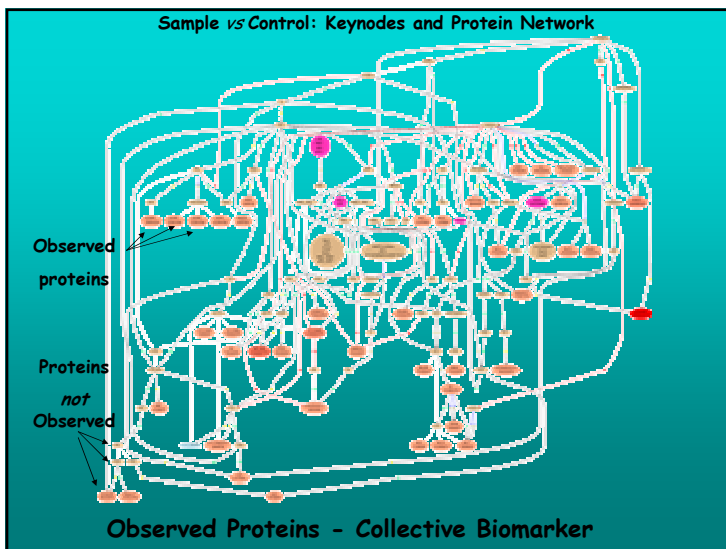
$$E = N \cdot (\text{Rank}_1/N) \cdot (\text{Rank}_2/N) \cdot (\text{Rank}_3/N) \cdot (\text{Rank}_4/N)$$

$$= \text{Rank}_1 \cdot \text{Rank}_2 \cdot \text{Rank}_3 \cdot \text{Rank}_4 / N^3$$

Key Node Analysis: Statistical Evaluation

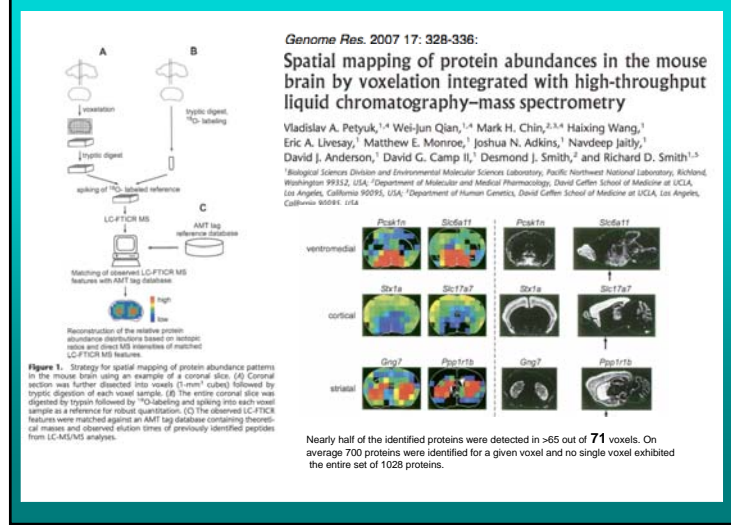


Sample vs Control: KeyNodes and Protein Network



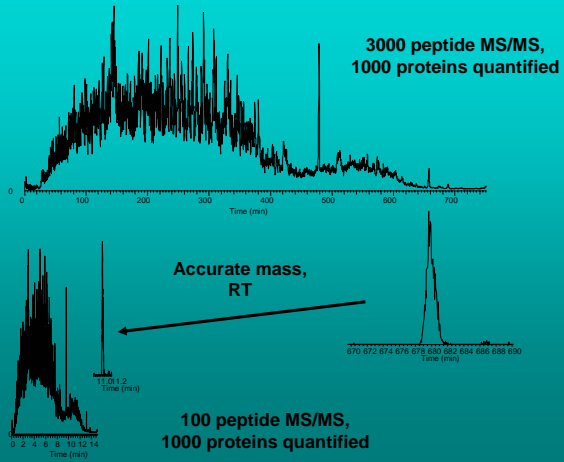
Genome Res. 2007 17: 328-336: Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry

Vladislav A. Petyuk,^{1,4} Wei-Jun Qian,^{1,4} Mark H. Chin,^{2,3,4} Haixing Wang,¹ Eric A. Livesay,¹ Matthew E. Monroe,¹ Joshua N. Adkins,¹ Navdeep Jaitly,¹ David J. Anderson,¹ David G. Camp II,¹ Desmond J. Smith,² and Richard D. Smith^{1,5}
¹Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352, USA; ²Department of Molecular and Medical Pharmacology, David Geffen School of Medicine at UCLA, Los Angeles, California 90095, USA; ³Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, California 90095, USA; ⁴UT-ORNL

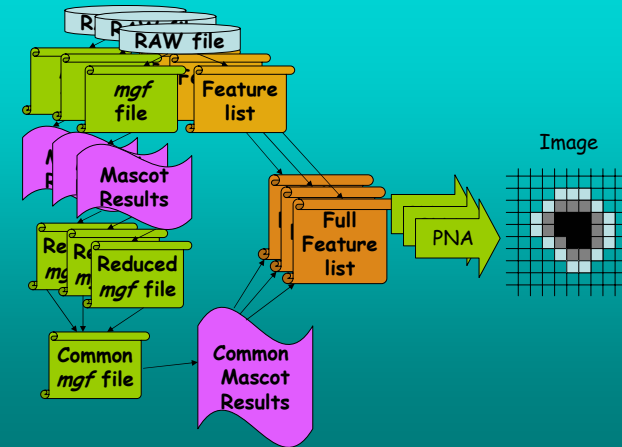


Nearly half of the identified proteins were detected in >65 out of 71 voxels. On average 700 proteins were identified for a given voxel and no single voxel exhibited the entire set of 1028 proteins.

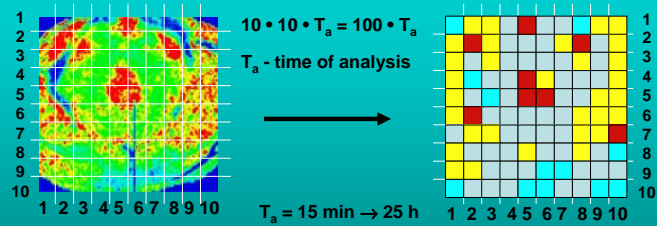
Pathway Analysis in Tissue Profiling?



Data reduction in PNA tissue imaging: Workflow



Pathway Analysis in Tissue Profiling?



Dr. Corina Mayrhofer
Clinical Institute of Pathology
Medical University of Vienna



International Conference on Pathway Analysis in Proteomics

November 8-11, 2008

Lisbon, Portugal

Requests:

Roman.Zubarev@icm.uu.se

Acknowledgements

Corina Meyrhofer
Konstantin Artemenko
Per Andren
Åke Engström
Ulf Pettersson

Alan Marshall
Matthias Mann
Peter Roepstorff
Mikhail Gorshkov
Per Håkansson
Peter O'Connor
Youri Tsybin

Kim Haselmann
Bogdan Budnik
Alexandra Patriksson
David van der Spoel
Serhiy Souchelnitskiy

Thermo
Wolfgang Metelmann-Strupat
Stevan Horning
Robert Malek
Martin Zeller
Michiko Kanai

Funding:
Swedish Research Council
K&A Wallenberg foundation