What's in a proteome? Lessons from a visual account of protein investment in cellular functions

Ron Milo

Department of Plant Sciences
Weizmann Institute of Science

Ask us about……

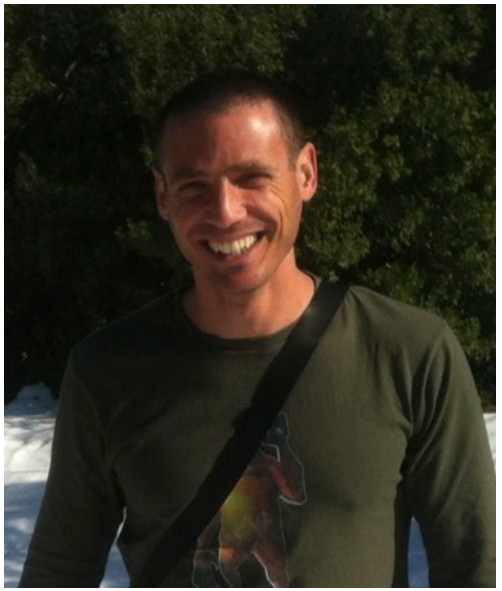
- Glycolytic strategy as a tradeoff between energy yield and protein cost (Avi Flamholz)

Ask us about……

- Glycolytic strategy as a tradeoff between energy yield and protein cost (Avi Flamholz)
- Enzyme cost of metabolic pathway fluxes (Wolf Liebermeister)

Ask us about……

- Glycolytic strategy as a tradeoff between energy yield and protein cost (Avi Flamholz)
- Enzyme cost of metabolic pathway fluxes (Wolf Liebermeister)
- Increased concentration of proteins with growth rate can result from passive resource redistribution (Uri Barenholz)



In search of a quantitative view of the proteome

- When you sum over recent reports:
  60 million in *S. pombe* (volume ≈100 um^3) ;
  1-2 billion in Hela/U2OS cell lines (volume ≈3000 um^3)

- How can we make a sanity check?
- What will a back of the envelope calculation suggest?

estimating the number of proteins per cell volume

Many more examples in:
"Cell Biology by the Numbers"
→ freely available electronically

protein mass per volume (≈ 0.2 g/ml)

number of proteins per cell volume

$$\frac{N}{V} = \frac{c_p}{l_{aa} \times m_{aa}}$$

m_{aa}: mass aa (≈ 110 Da)

l_{aa}: aa per protein ($\approx 350 \frac{aa}{protein}$)

Avogadro's number

$$\frac{N}{V} = \frac{0.2\,[g/ml] \times 6 \times 10^{23}\left[\frac{Da}{g}\right] \times 10^{-12}\left[\frac{ml}{\mu m^3}\right]}{350\left[\frac{aa}{protein}\right] \times 110\left[\frac{Da}{aa}\right]} \approx 3 \times 10^6 \frac{proteins}{\mu m^3}$$

organism	characteristic volume	number of proteins
E. coli	$\approx 1\ \mu m^3$	$\approx 3 \times 10^6$
budding yeast	$\approx 30\ \mu m^3$	$\approx 100 \times 10^6$
HeLa cell line	$\approx 3{,}000\ \mu m^3$	$\approx 10 \times 10^9$

In search of a quantitative view of the proteome

- 2-4 million proteins per micron cubed
  - i.e. ≈3 million in *E. coli* medium growth rate; ~100 million in budding yeast; ~10 billion in a human cell line
  - Several publications probably require normalization factors
  - Sanity checks are useful…
  - "What is the total number of protein molecules per cell volume? A call to rethink some published values" (Bioessays 2013)

- **What are the functions that require the most protein investment in a cell?**
- **What are the "top 10" most highly expressed proteins?**

Novel methodologies measure proteome-wide abundances in a variety of model organisms

Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise

John R. S. Newman[1,2], Sina Ghaemmaghami[1,2]†, Jan Ihmels[1,2], David K. Breslow[1,2], Matthew Noble[1], Joseph L. DeRisi[1,3] & Jonathan S. Weissman[1,2]

Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation

Peng Lu[1–3], Christine Vogel[1,2], Rong Wang[1,2], Xin Yao[1] & Edward M Marcotte[1]

On what does the cell spend its budget?
How can we convey all this information visually?

The quantitative proteome of a human cell line

Martin Beck[1,9], Alexander Schmidt[2,9], Johan Malmstroem[3,4], Manfred Claassen[5], Alessandro Ori[1], Anna Szymborska[1], Franz Herzog[6], Oliver Rinner[4], Jan Ellenberg[1] and Ruedi Aebersold[6,7,8,*]

Proteome Organization in a Genome-Reduced Bacterium

Sebastian Kühner,[1*] Vera van Noort,[1*] Matthew J. Betts,[1] Alejandra Leo-Macias,[1] Claire Batisse,[1] Michaela Rode,[1] Takuji Yamada,[1] Tobias Maier,[2] Samuel Bader,[1] Pedro Beltran-Alvarez,[1] Daniel Castaño-Diez,[1] Wei-Hua Chen,[1] Damien Devos,[1] Marc Güell,[2] Tomas Norambuena,[3] Ines Racke,[1] Vladimir Rybin,[1] Alexander Schmidt,[4] Eva Yus,[2] Ruedi Aebersold,[4] Richard Herrmann,[5] Bettina Böttcher,[1]† Achilleas S. Frangakis,[1] Robert B. Russell,[1] Luis Serrano,[2,6] Peer Bork,[1]‡ Anne-Claude Gavin[1]‡

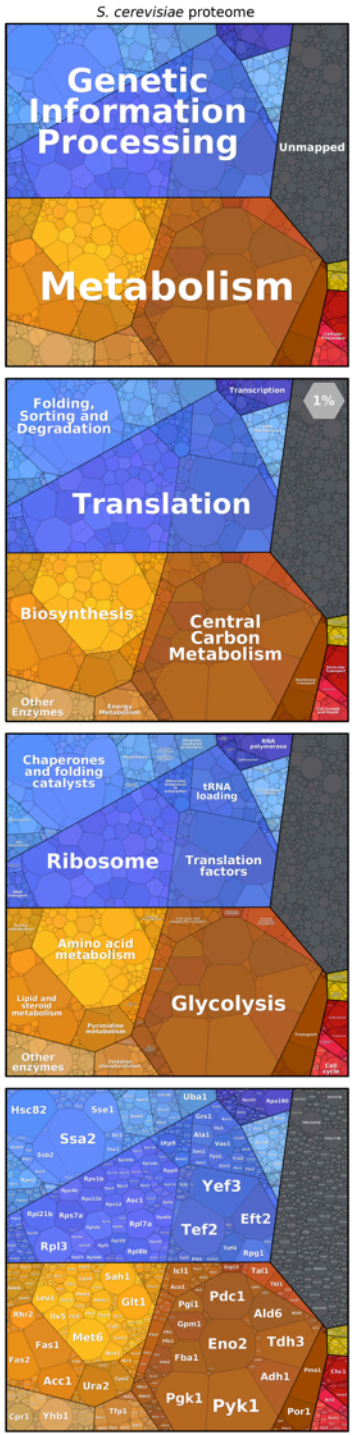
The power of global cross fertilization

- Wolfram Liebermeister – Charité University Berlin

- Jörg Bernhardt - Greifswald University

Proteomaps are Voronoi tree maps

Proteomaps give a quantitative view of cellular investment in the proteome from single protein level through specific functions to broad functional categories



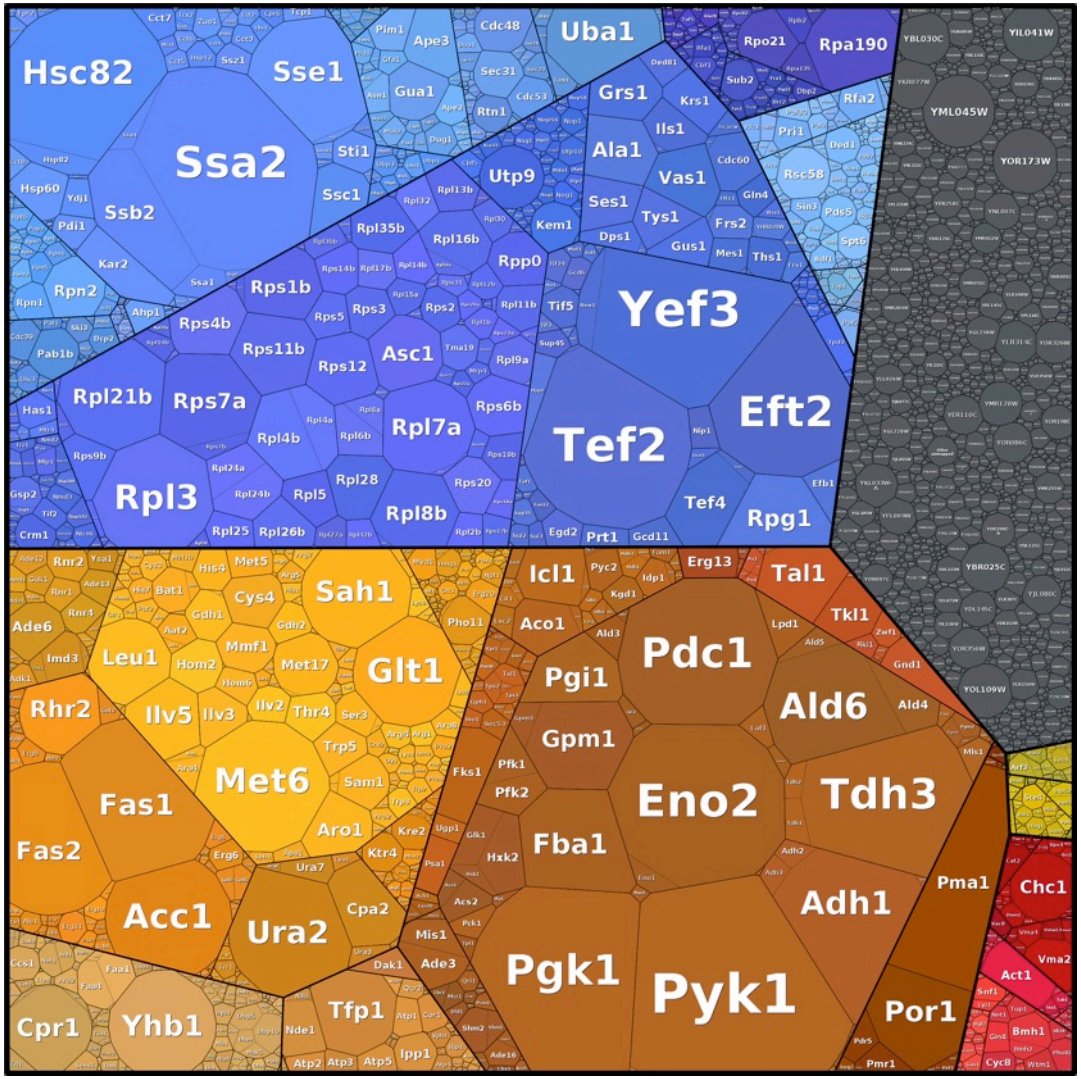
S. cerevisiae proteome in rich medium

MS data from: Nagaraj 2012

Proteomaps enable quantitative visualization of the proteome

- Polygon size is proportional to abundance

- We usually calculate length weighted (i.e. amino acid investment) but also present on website by copy numbers

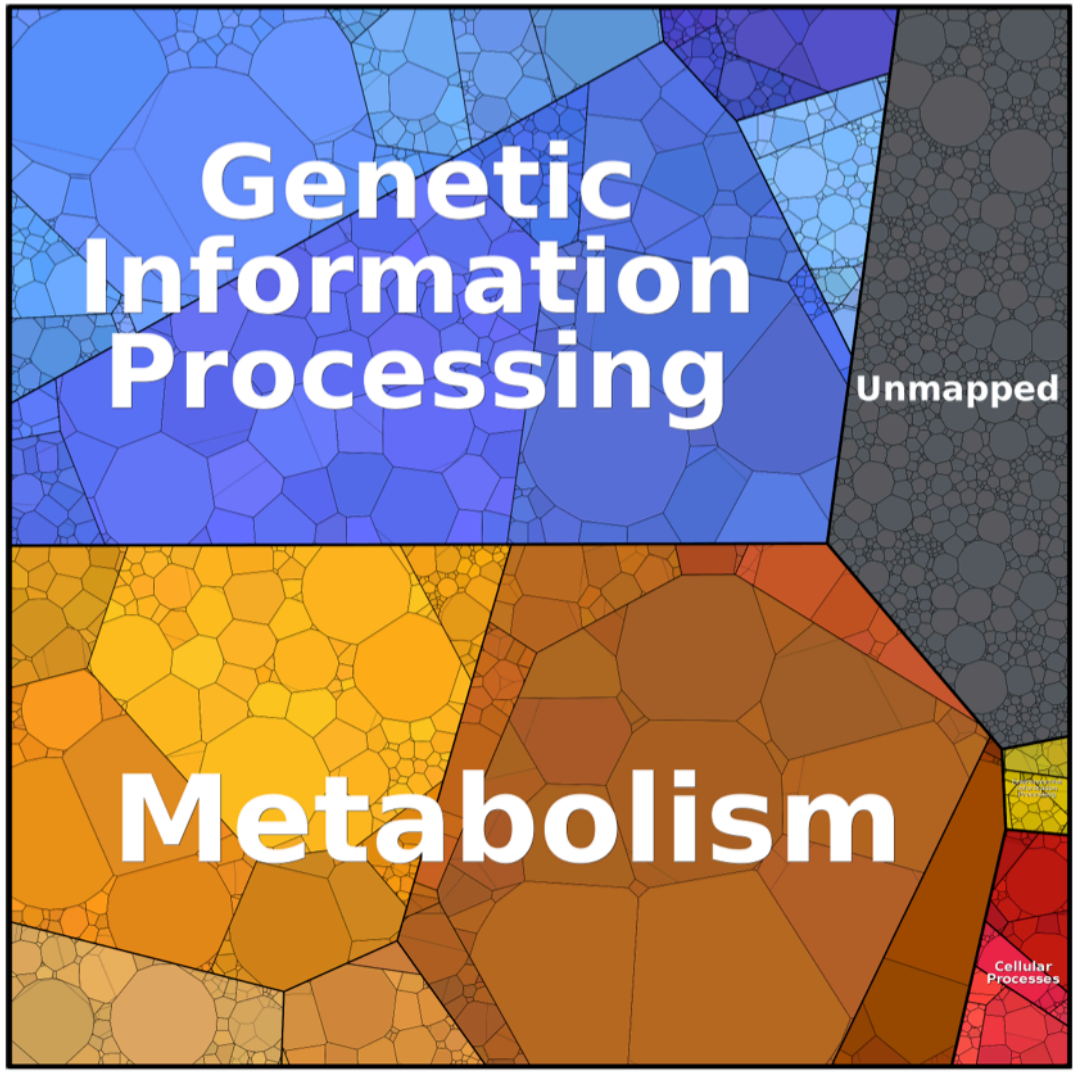


S. cerevisiae proteome in rich medium

MS data from: Nagaraj 2012

Proteomaps enable quantitative visualization at the functional level

- Metabolism and the Central dogma dominate the investment in single-cell organisms
- Signaling (green) is a minor fraction

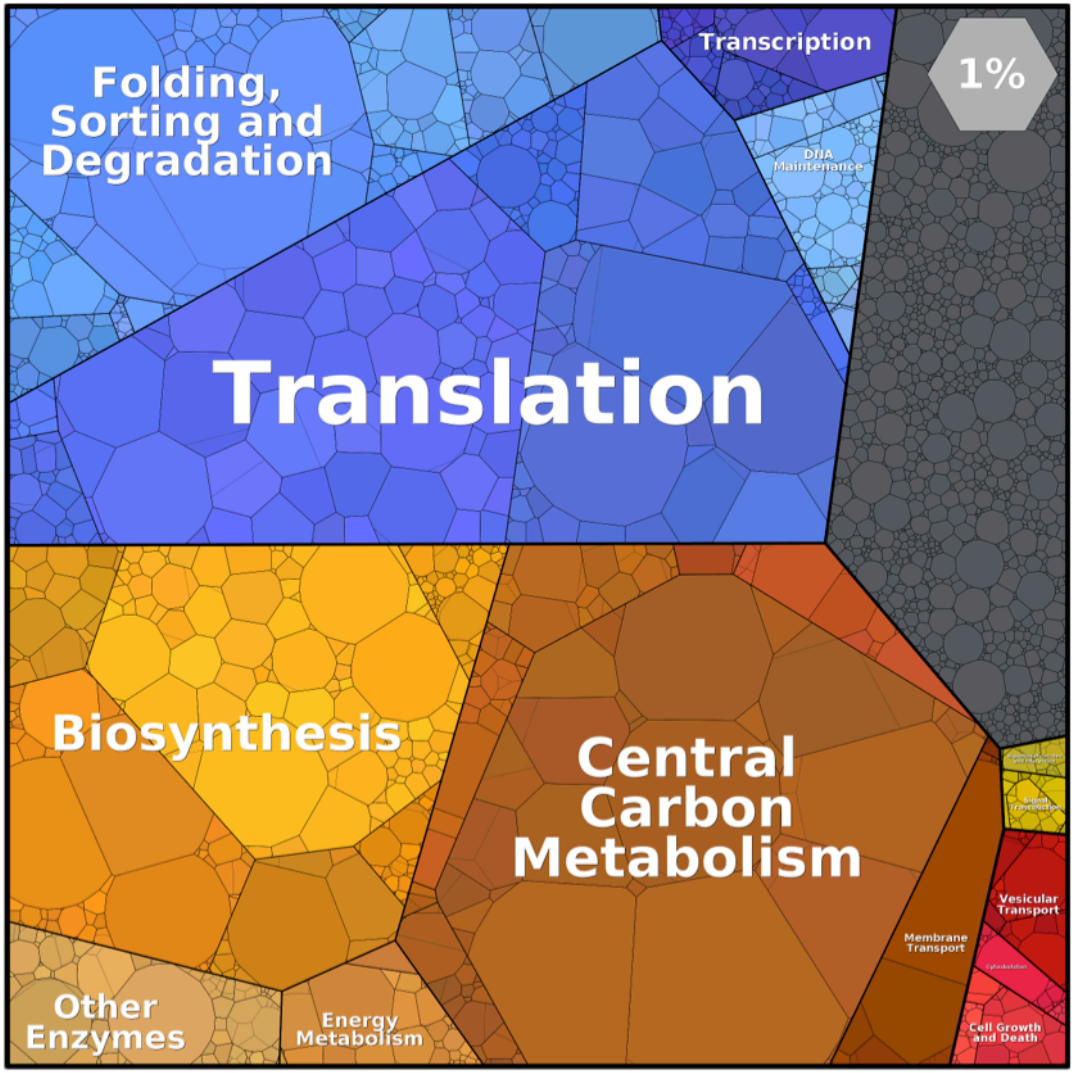


S. cerevisiae proteome in rich medium

MS data from: Nagaraj 2012

Proteomaps enable quantitative visualization at the functional level

- Functional clustering and color code
- Investment in central dogma is not homogenous:
  Translation>>Transcription
  Translation>> Replication

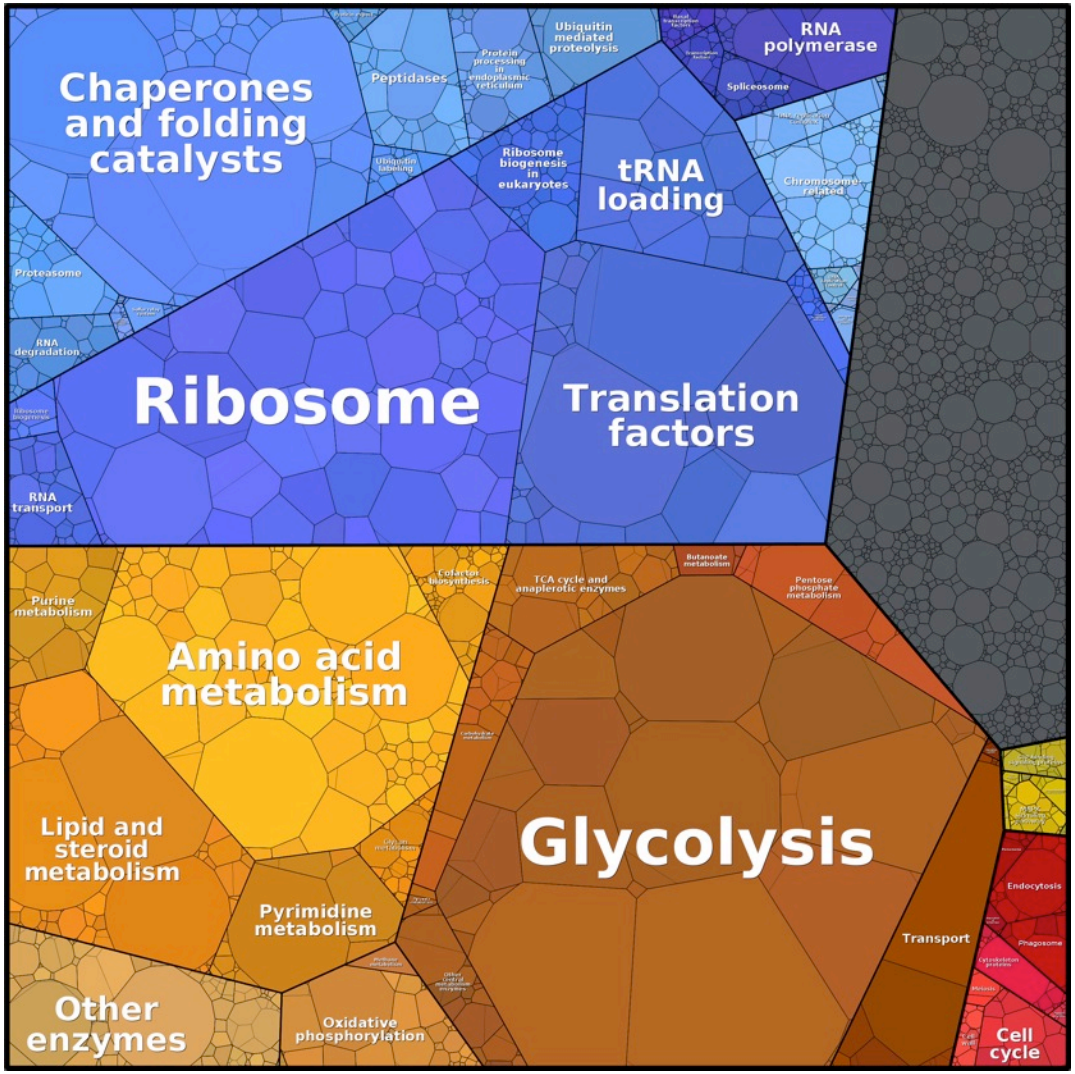


S. cerevisiae proteome in rich medium

MS data from: Nagaraj 2012

Proteomaps enable quantitative visualization at the functional level

- Functional clustering and color code
- Glycolysis >10% of proteome though <1% of genome

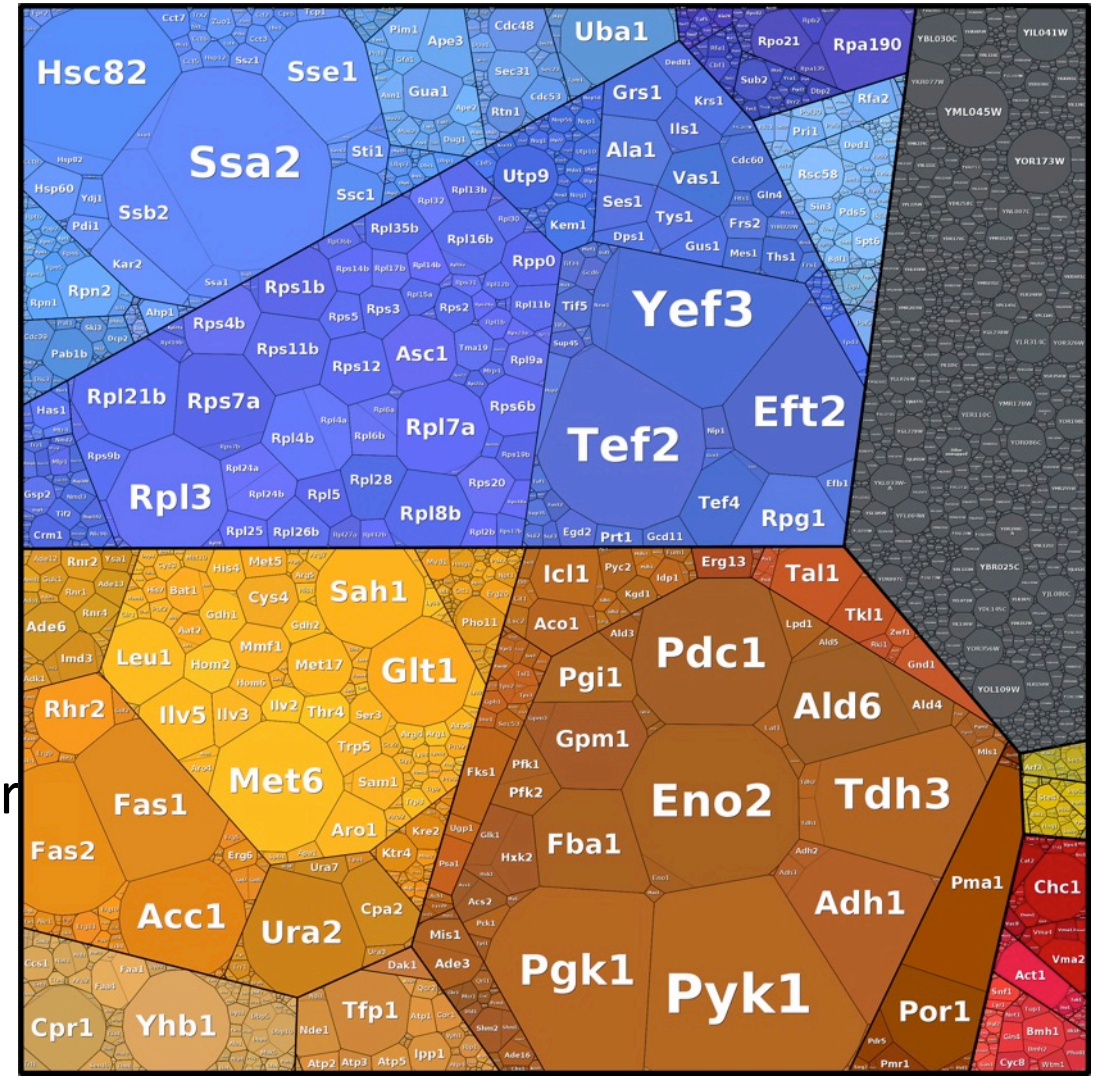


S. cerevisiae proteome in rich medium

MS data from: Nagaraj 2012

Proteomaps enable quantitative visualization of the proteome

- Most abundant proteins are easily recognized
- Our brains are good at extracting information from images
- Our intuition & memory gains from images (rather than only tables etc.)



S. cerevisiae proteome in rich medium

MS data from: Nagaraj 2012

Functional clustering of proteome based on KEGG Pathway Maps ontology

Genetic Information Processing

- Transcription
- Translation
- Folding, Sorting and Degradation
- DNA maintenance
- RNA family

Environmental Information Processing

- Signal Transduction
- Signaling Molecules and Interaction

Cellular Processes

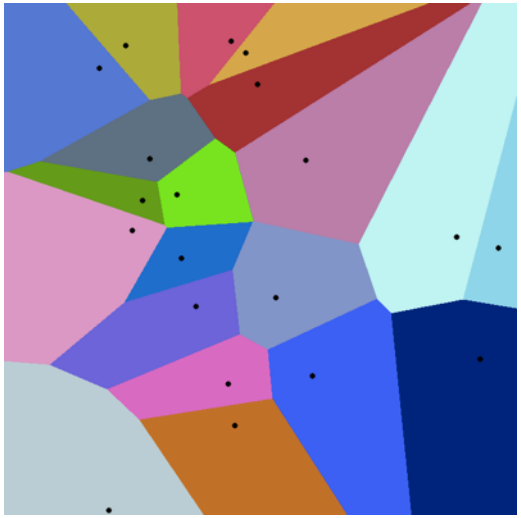
- Cytoskeleton
- Cell cycle
- Transport
- Cell Motility
- Cell Growth and Death
- Cell Communication

Metabolism

- Carbohydrate Metabolism
- Energy Metabolism
- Lipid Metabolism
- Nucleotide Metabolism
- Amino Acid Metabolism
- Metabolism of Other Amino Acids
- Glycan Biosynthesis and Metabolism
- Metabolism of Cofactors and Vitamins
- Metabolism of Terpenoids and Polyketides
- Biosynthesis of Other Secondary Metabolites
- Xenobiotics Biodegradation and Metabolism
- Further enzymes

Not mapped

Proteomaps are created as consecutively refined Voronoi diagrams

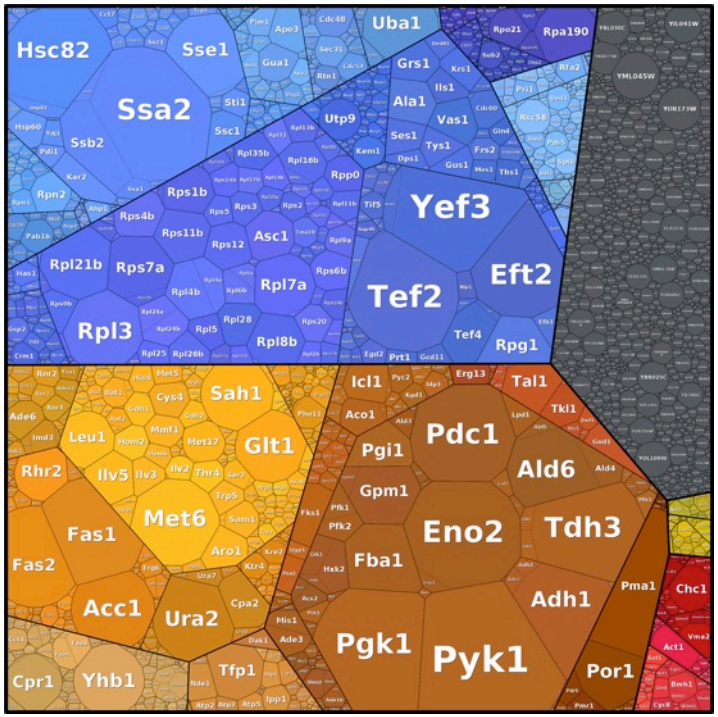


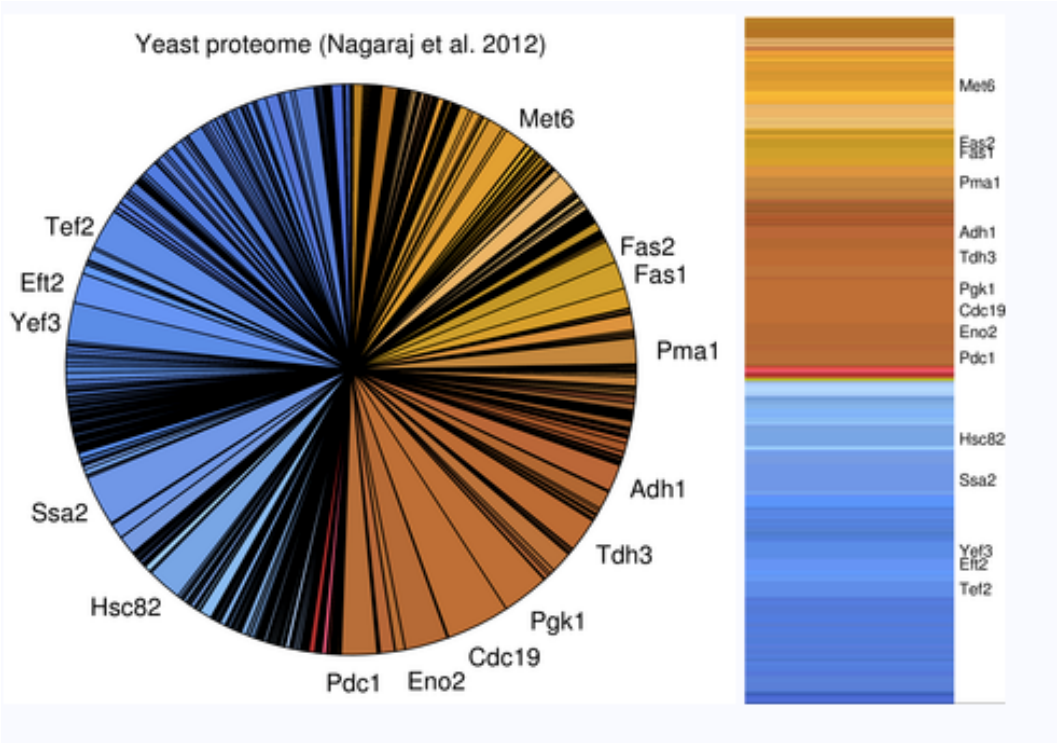
Bernhardt J, Funke S, Hecker M, Siebourg J (2009) in Sixth International Symposium on Voronoi Diagrams, pp. 233-241

Proteome wide quantification and mapping has various caveats

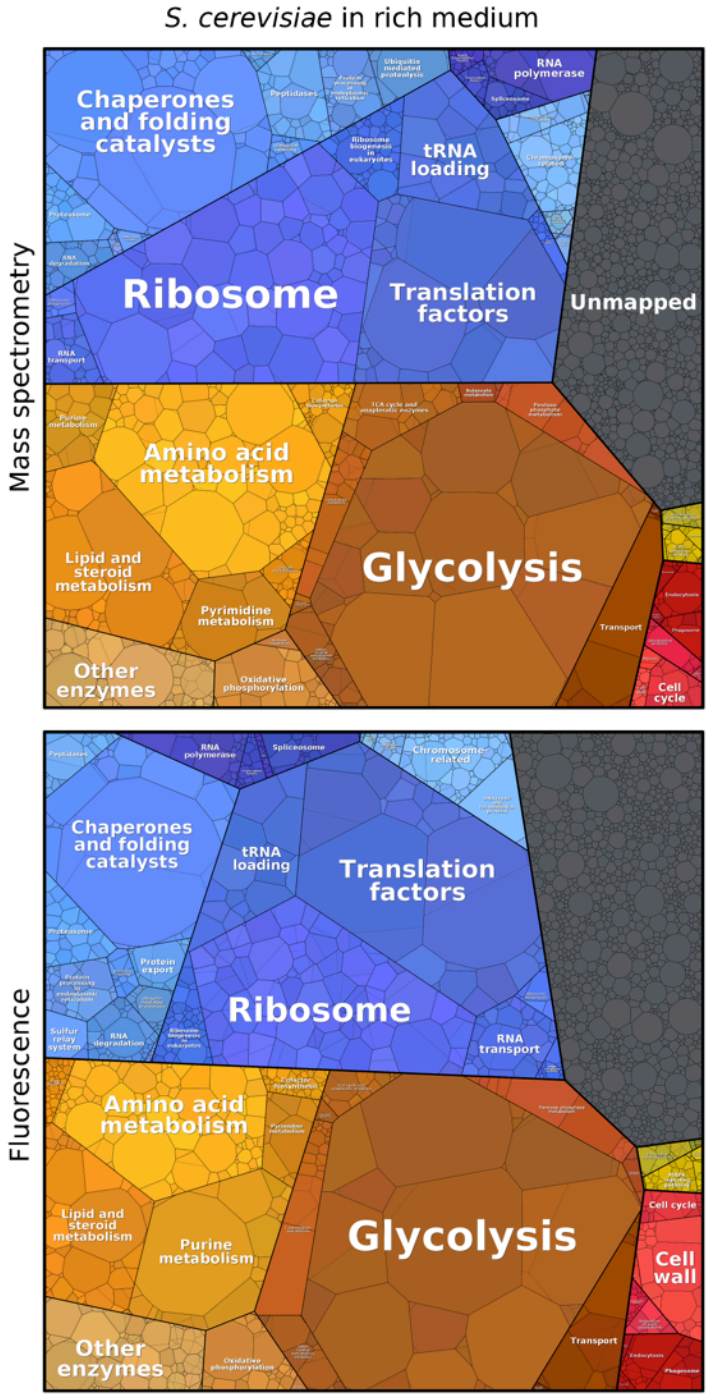
- Quantification issues:
  - Biases in extraction (compartments etc.)
  - Biases due to tagging/spectrometry
- Mapping issues:
  - A protein is forced to belong to only one functional class
  - Mapping is deterministic but not unique (local minimum)

A 2D tiling (e.g. proteomap) has advantages over a pie or bar diagram





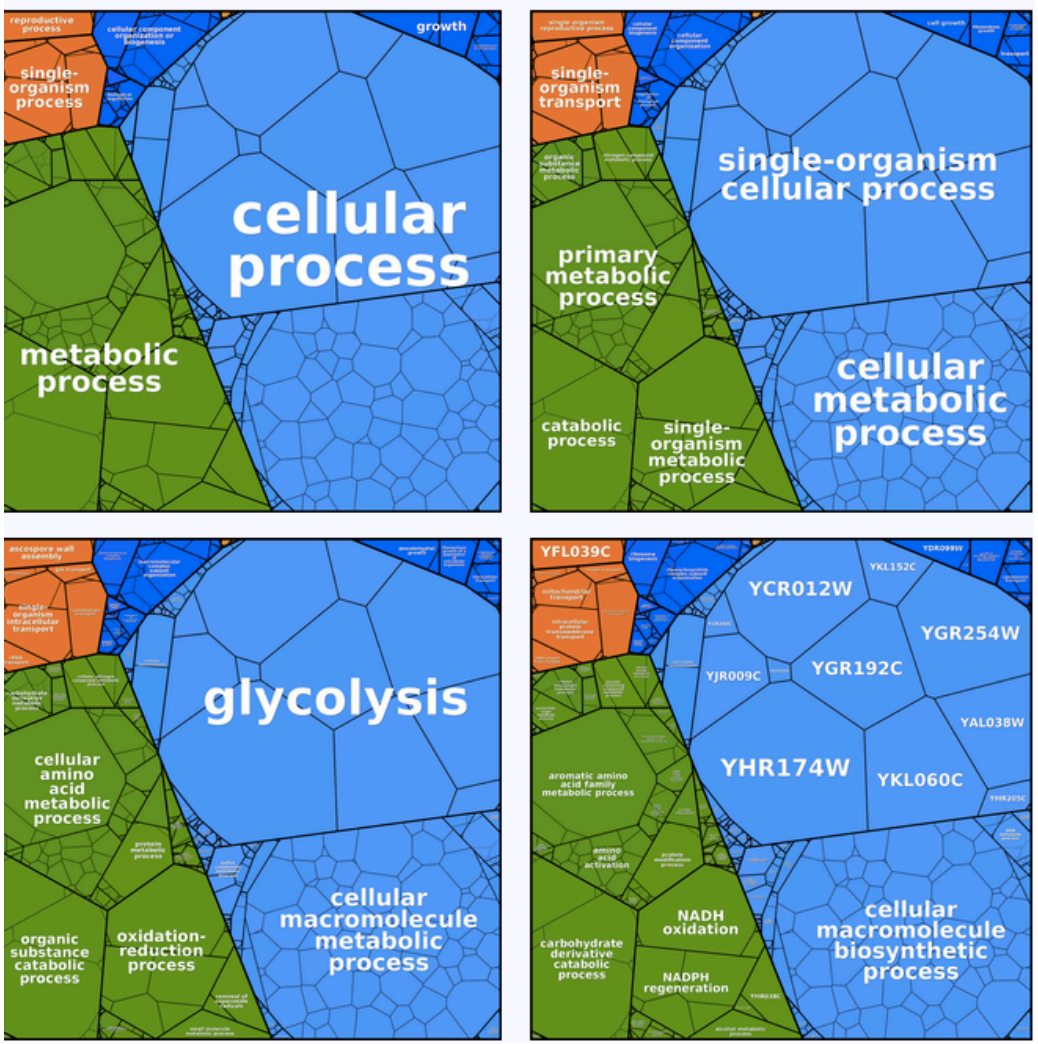
Different measurement methods are quite consistent



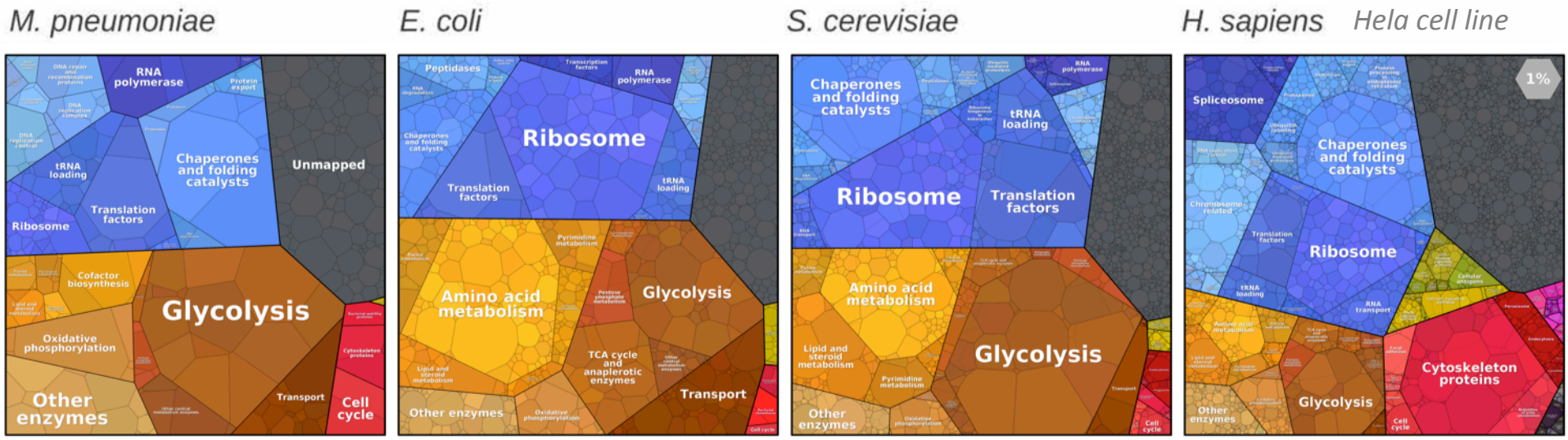
Data from: Nagaraj, 2012; Newman, 2006

GO annotation is less suitable due to a different hierarchy tree structure

(we use the well balanced KEGG hierarchy)



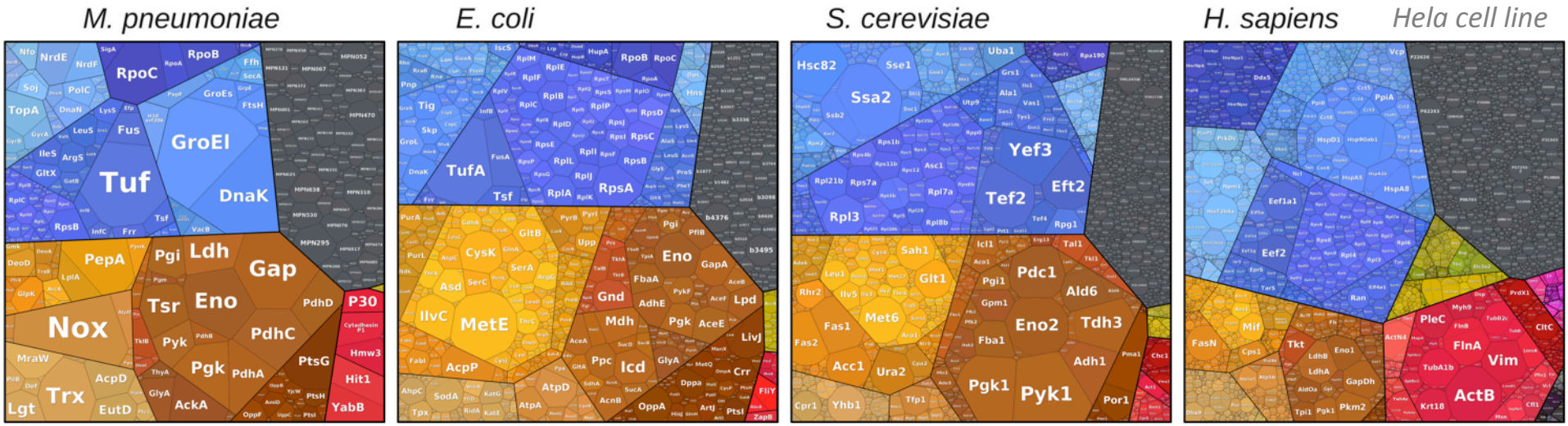
Comparative view of the proteome: commonalities and species-specific trends



Data from: Kühner, 2009, Lu, 2007

Top proteins abundance and function

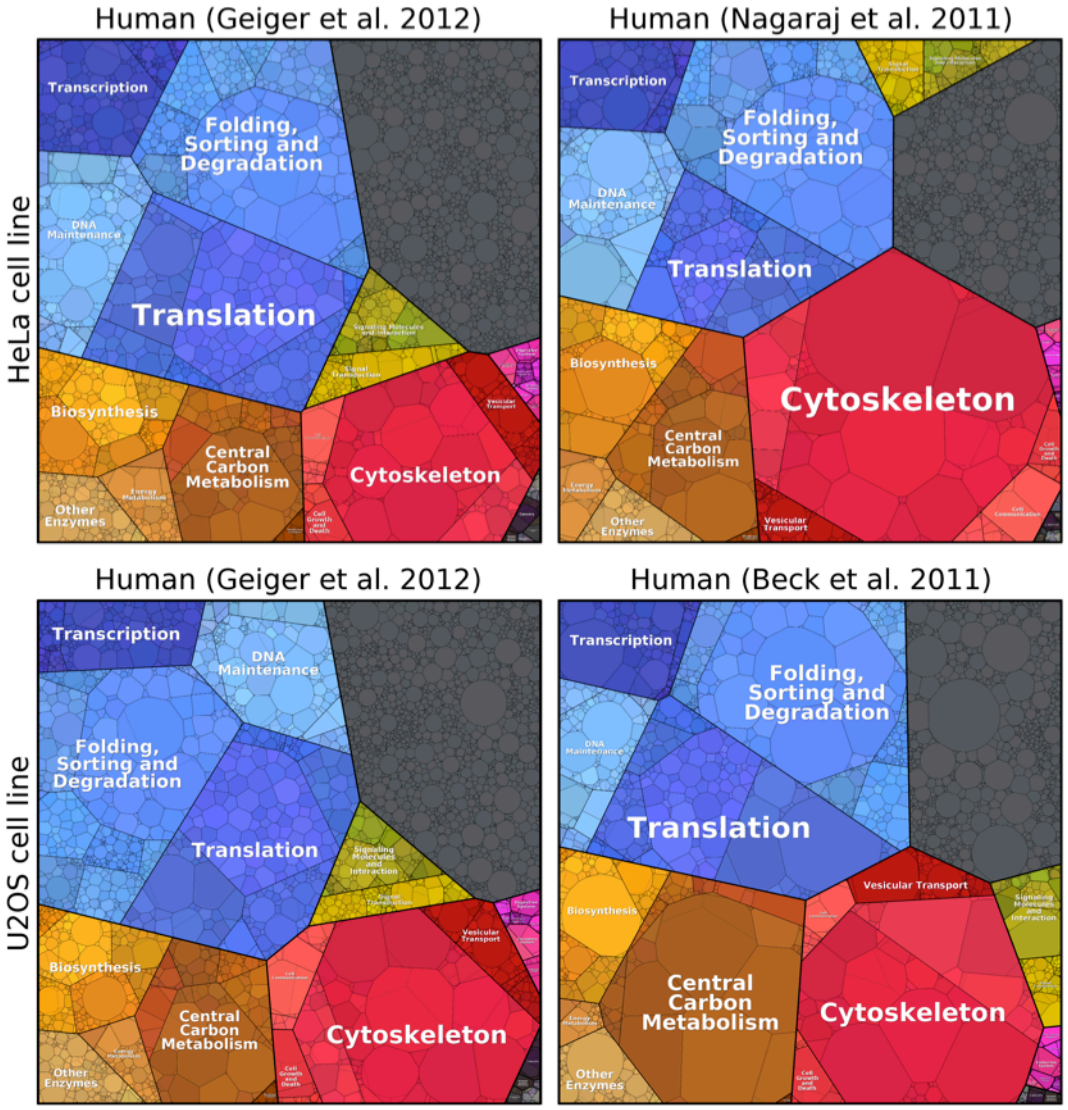
- Highest expressing protein usually 3-5% of proteome
- Usually 10-20 proteins >1% of proteome each
- Common top 10 stars: Glycolytic (Eno, Gap, Pgk, Pyk), Elongation factors (Tuf/Yef), Chaperones (GroEL, DnaK)



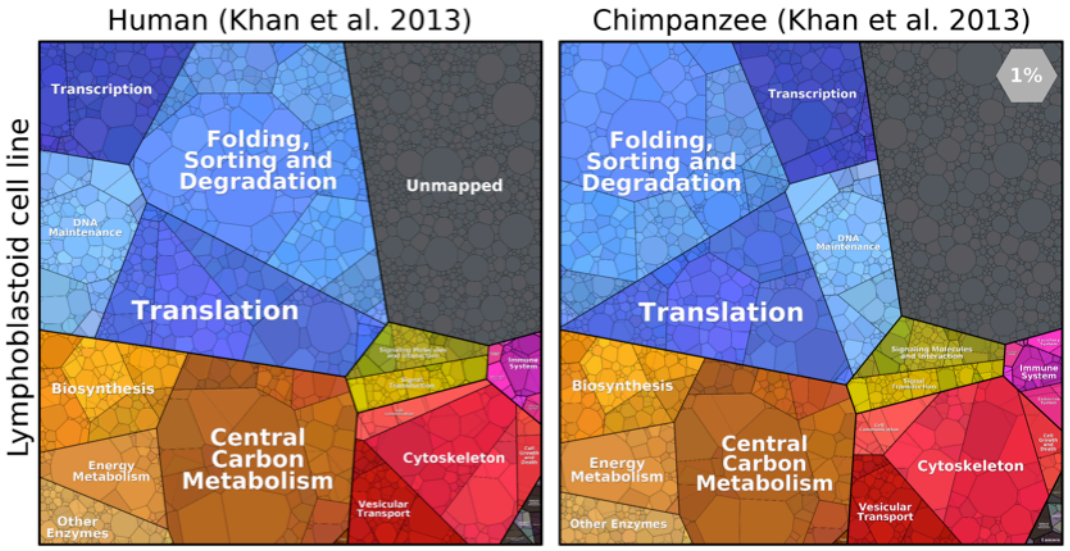
Data from: Kühner, 2009, Lu, 2007

How different is a human cell line from other human cell lines and from the same cell line in chimp?

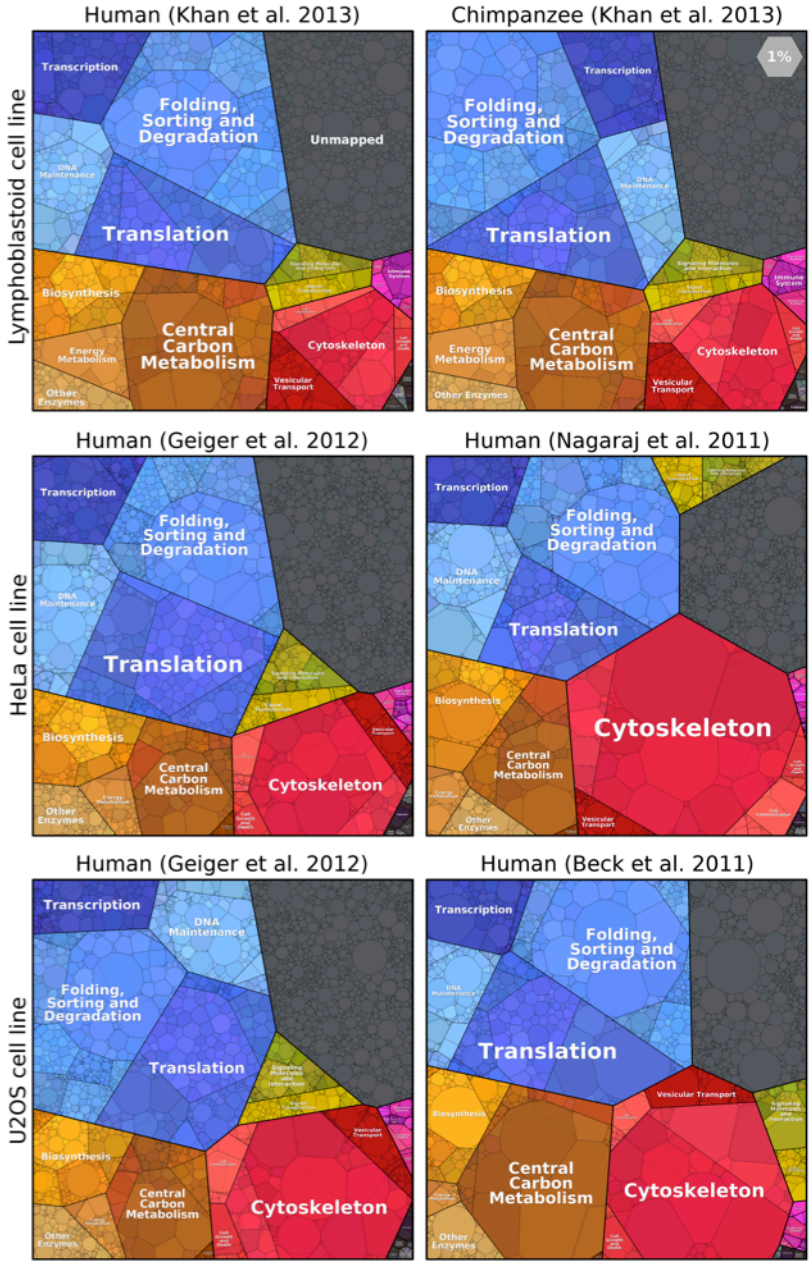
How similar are different cell lines?



How similar are we to a chimp?



How different is a human cell line from other human cell lines and from the same cell line in chimp?



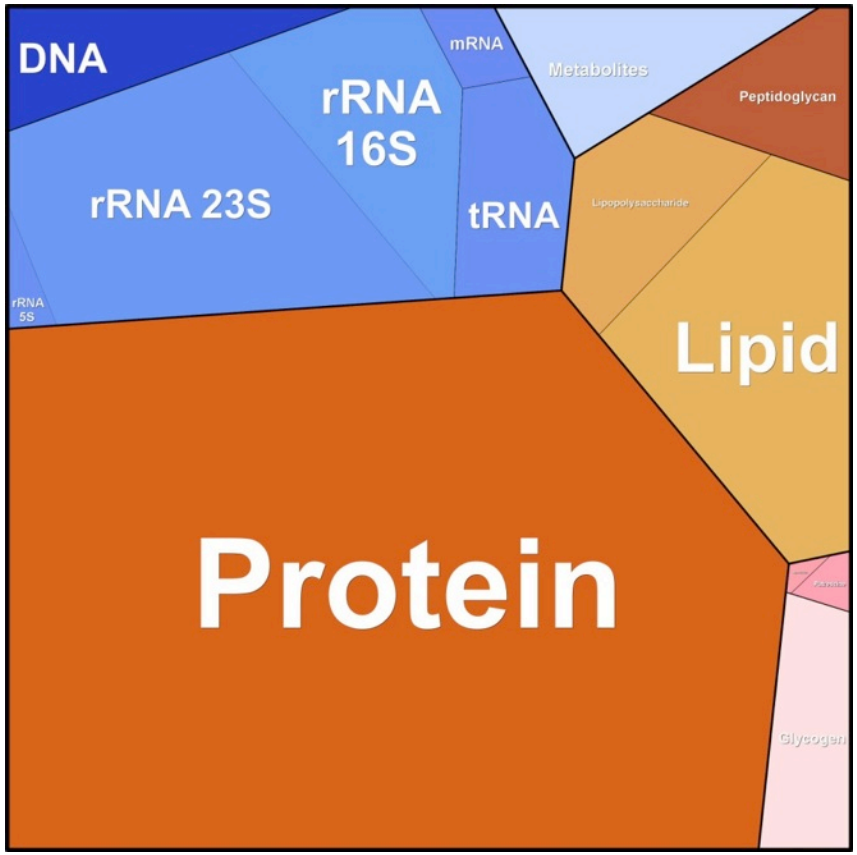
Proteomaps can serve as a learning and research tool

- Way to interactively explore abundant proteins and functional categories
- Our intuition & memory gains from images (rather than only tables etc.)

Check out in website how you can hover over a functional group or protein to find fraction and number of copies and also functional annotation in KEG

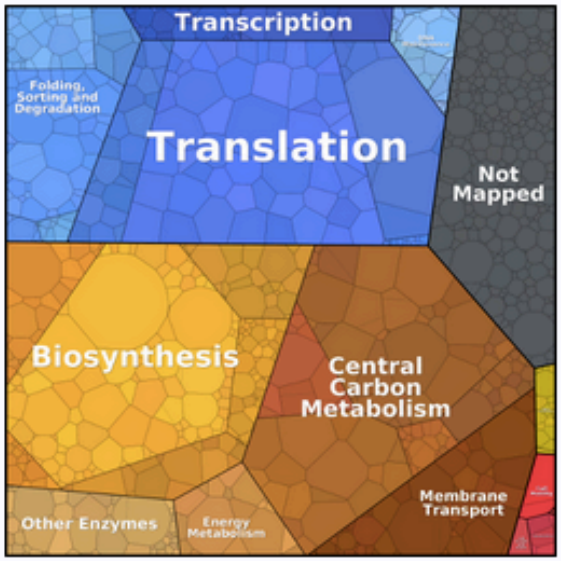
Towards a quantitative view of all cell constituents

- A flood of proteome wide data is on the way
- Various organisms, conditions, perturbations
- Also useful for transcriptome, metabolome, dry mass etc.
- Where cell resources are invested (biomass, nitrogen, ribosome machinery, cell volume, folding and solubility)



Neidhardt et al, 1990

www.proteomaps.net

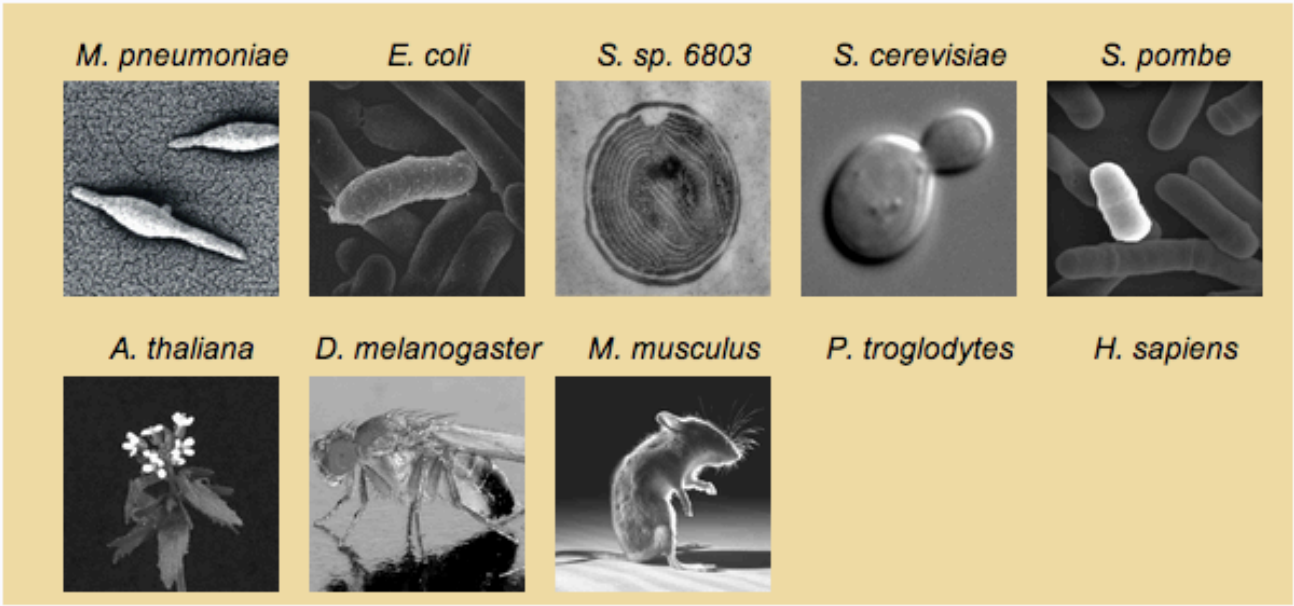


Depicting the protein composition of cells

Proteomaps show the quantitative composition of proteomes with a focus on protein function. They are built automatically from proteome data and based on the KEGG Pathways gene classification. Each protein is shown by a polygon and functionally related proteins are arranged in common regions. To emphasize highly expressed proteins, polygon areas represent protein abundances, weighted by protein size.

To open a proteomap, choose an organism and click on a proteomap symbol in the list. The line above the proteomap allows you to navigate between hierarchy levels. Hover the mouse pointer over a map to see the abundance values.

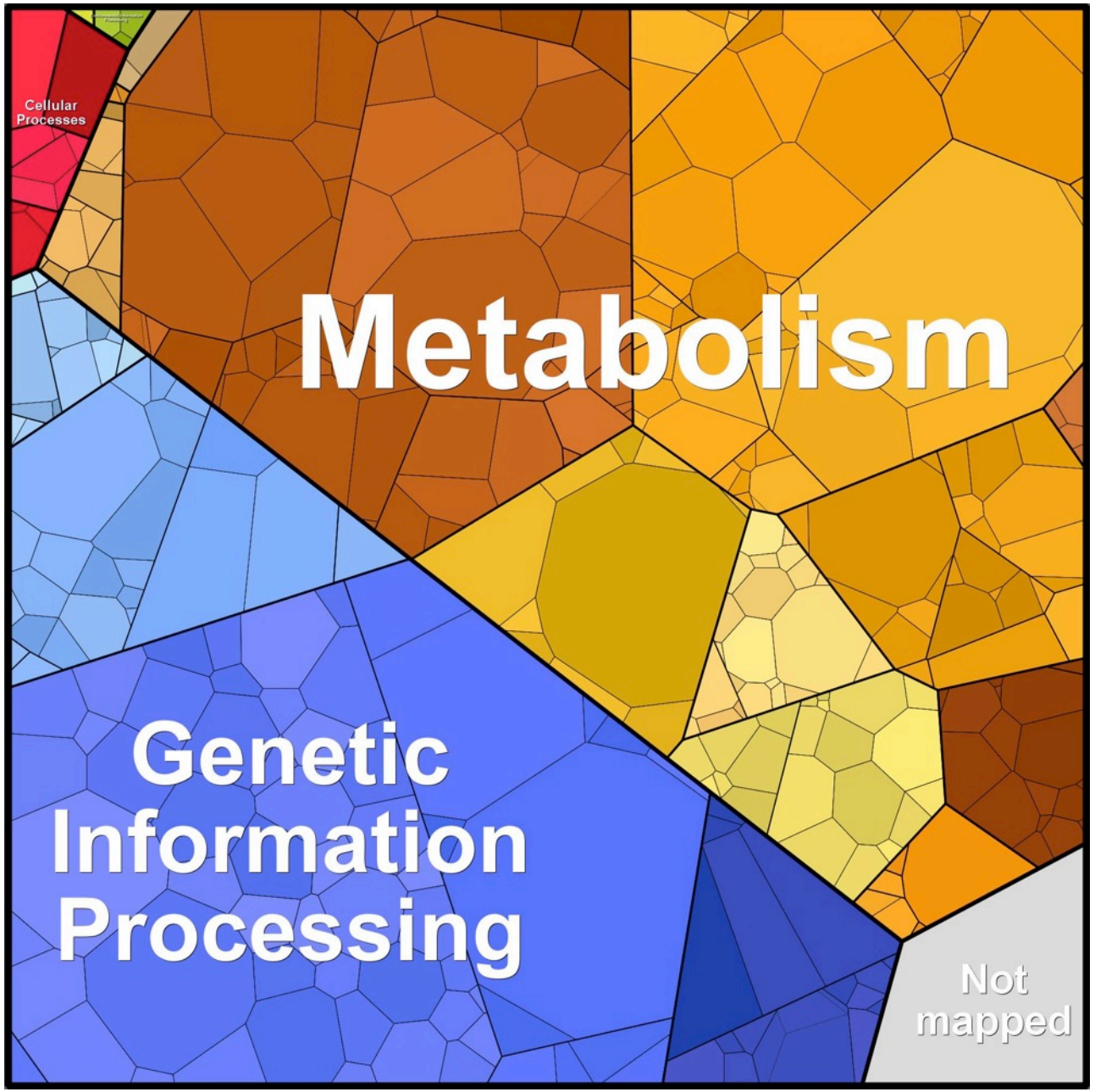
See methods description and data tables



The Proteomaps team

- Elad Noor & Avi Flamholz & Dan Davidi - Weizmann


- Wolfram Liebermeister – Charité University Berlin
- Jörg Bernhardt - Greifswald University
- More at: http://www.proteomaps.net/
  Liebermeister et al, PNAS 2014



Data from:
Lu, 2007

E. coli, Exponential growth on minimal media