Human metabolic network reconstruction

Hedi Peterson

BIIT

jclub
12.02.2007

Motivation

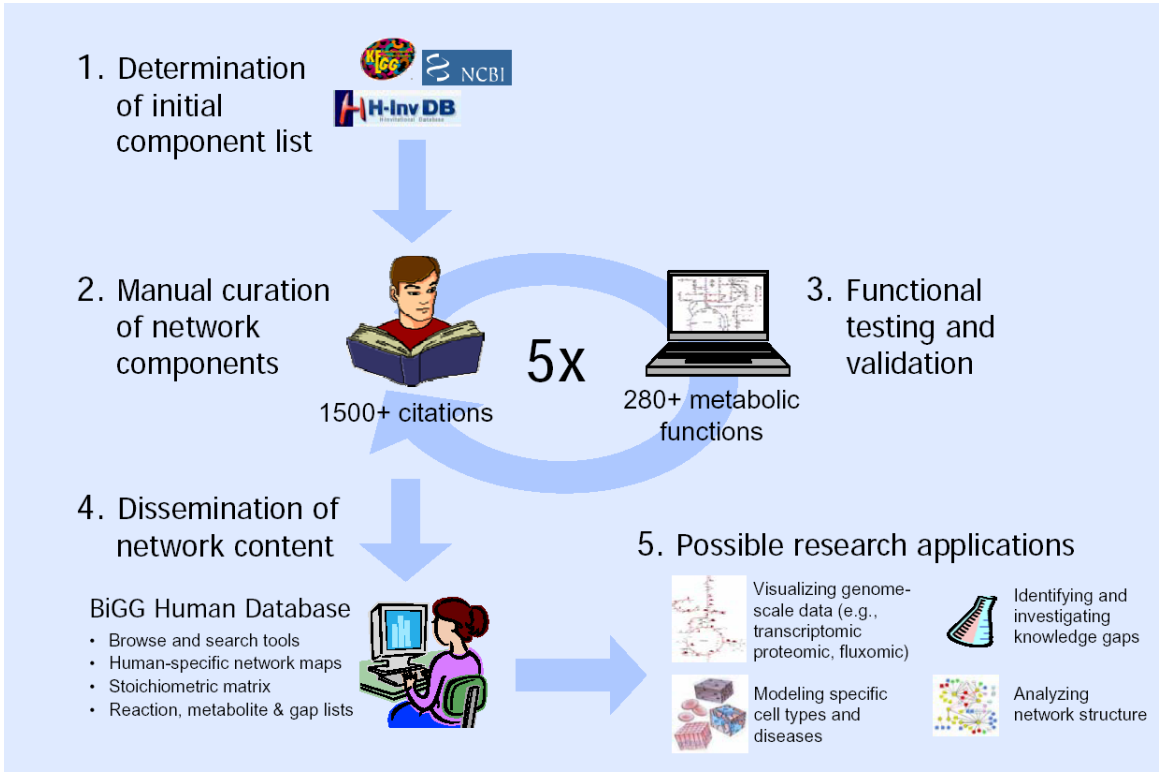
- discovery of missing information
- formulation of *in silico* model
- structured context for analyzing high-throughput biological data sets

Used data

- Human genome Build 35 - 1865 genes from KEGG
- Genes mapped to 3623 metabolic enzymes & 3672 reactions
- Over 1500 articles, reviews, biochemical books (from more than 50 years)

BiGG

- Input: bottom-up reconstruction of **Bi**ochemically, **G**enetically, **G**enomically structured reconstruction
- Outcome: *in silico* model for computing allowable network states under governing chemical and genetic constraints
- Validation: simulating 288 known metabolic functions *in silico* (5 iterative rounds)


1. Determination of initial component list
KEGG
NCBI
H-Inv DB
2. Manual curation of network components
1500+ citations
5x
280+ metabolic functions
3. Functional testing and validation
4. Dissemination of network content
BiGG Human Database
• Browse and search tools
• Human-specific network maps
• Stoichiometric matrix
• Reaction, metabolite & gap lists
5. Possible research applications
Visualizing genome-scale data (e.g., transcriptomic proteomic, fluxomic)
Identifying and investigating knowledge gaps
Modeling specific cell types and diseases
Analyzing network structure

Detail levels

- Formulated metabolites and reactions
  - known reaction stoichometry
  - substrate/cofactor specificity
  - substrate/cofactor directionality
  - overall conservation of mass and charge-based metabolite ioniation states at pH 7.2
- Compartmentalization of metabolites
  - intracellular: cytoplasm, mitochondria, nucleus, ER, Golgi complex, lysososome, peroxisome
  - extracellular environment
- Boolean descriptions of gene-protein relationships
  - alternative spliced variants
  - protein complexes
  - isozymes
- Confidence scores and literature references

Detail levels

- Formulated metabolites and reactions
  - known reaction stoichometry
  - substrate/cofactor specificity
  - substrate/cofactor directionality
  - overall conservation of mass and charge-based metabolite ioniation states at pH 7.2
- Compartmentalization of metabolites
  - intracellular: cytoplasm, mitochondria, nucleus, ER, Golgi complex, lysososome, peroxisome
  - extracellular environment
- Boolean descriptions of gene-protein relationships
  - alternative spliced variants
  - protein complexes
  - isozymes
- Confidence scores and literature references

Detail levels

- Formulated metabolites and reactions
  - known reaction stoichometry
  - substrate/cofactor specificity
  - substrate/cofactor directionality
  - overall conservation of mass and charge-based metabolite ioniation states at pH 7.2
- Compartmentalization of metabolites
  - intracellular: cytoplasm, mitochondria, nucleus, ER, Golgi complex, lysososome, peroxisome
  - extracellular environment
- Boolean descriptions of gene-protein relationships
  - alternative spliced variants
  - protein complexes
  - isozymes
- Confidence scores and literature references

Detail levels

- Formulated metabolites and reactions
  - known reaction stoichometry
  - substrate/cofactor specificity
  - substrate/cofactor directionality
  - overall conservation of mass and charge-based metabolite ioniation states at pH 7.2
- Compartmentalization of metabolites
  - intracellular: cytoplasm, mitochondria, nucleus, ER, Golgi complex, lysososome, peroxisome
  - extracellular environment
- Boolean descriptions of gene-protein relationships
  - alternative spliced variants
  - protein complexes
  - isozymes
- Confidence scores and literature references

Component	Number
Genes	1,496
Transcripts*	1,905
Proteins	2,004
Complex-associated reactions*	248
Isozyme-associated reactions*	946
Intrasystem reactions	3,311
Metabolic	2,233
Transport†	1,078
Exchange reactions†	432
Compartment-specific metabolites	2,712
Cytoplasm	995
Extracellular space	388
Mitochondrion	383
Golgi apparatus	279
Endoplasmic reticulum	231
Lysosome	207
Peroxisome	139
Nucleus	90
Citations	1,587
Primary literature	1,378
Review articles	188
Textbooks	21
Validated metabolic functions	288
Knowledge gaps‡	356

Building the network

- Initial component list
- Eight metabolic subsets
  - Amino acids
  - Carbohydrates
  - Energy
  - Glycans
  - Lipids
  - Nucleotides
  - Secondary metabolites/xenobiotics
  - Vitamins & cofactors

Building the network

- Reaction directionality from thermodynamic data or inferred from legacy data and textbooks
- Compartmentalization from protein localization data, sequence targeting signals and indirect physiological evidence (else cytoplasmic)
- GTPR relationships were manually identified from literature
- Relationships were formulated as Boolean logic statements
  - **OR** Isozymes - distinct proteins catalyzing same substrate- and compartment specific reaction (1G - 2T or 2G)
  - **AND** Protein complexes when reactions depend on more than one protein

Pathway categories

- I - extensive primary literature
- II - 1/2 highly characterized enzymes & 1/2 with moderate biological evidence
- III - wide range of confidence scores and gene coverage

Confidence scores

- 3 - classical biochemical or genetical evidence
- 2 - physiological data or biochem/genetical data from non-humans
- 1 - *in silico* modeling
- 0 - unevaluated

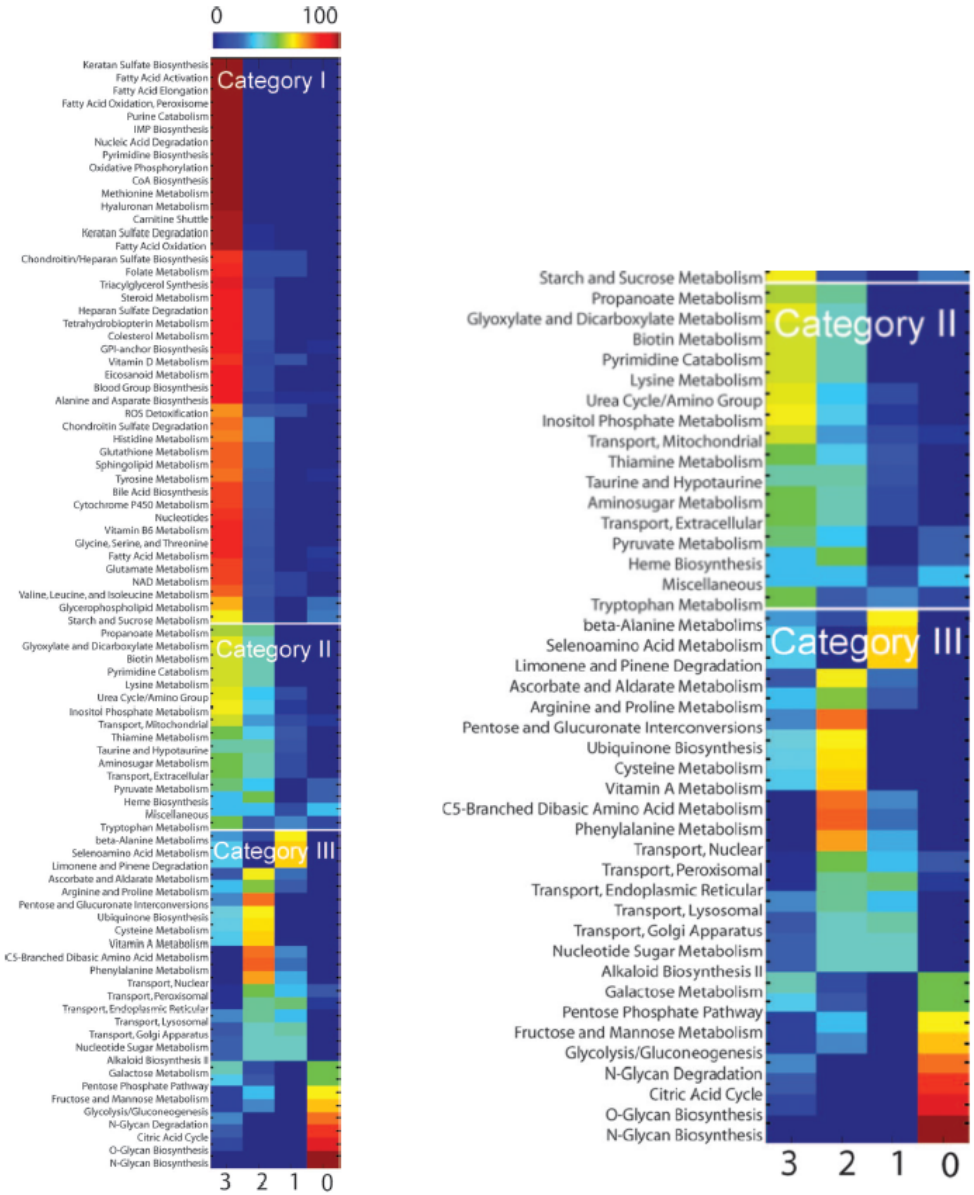
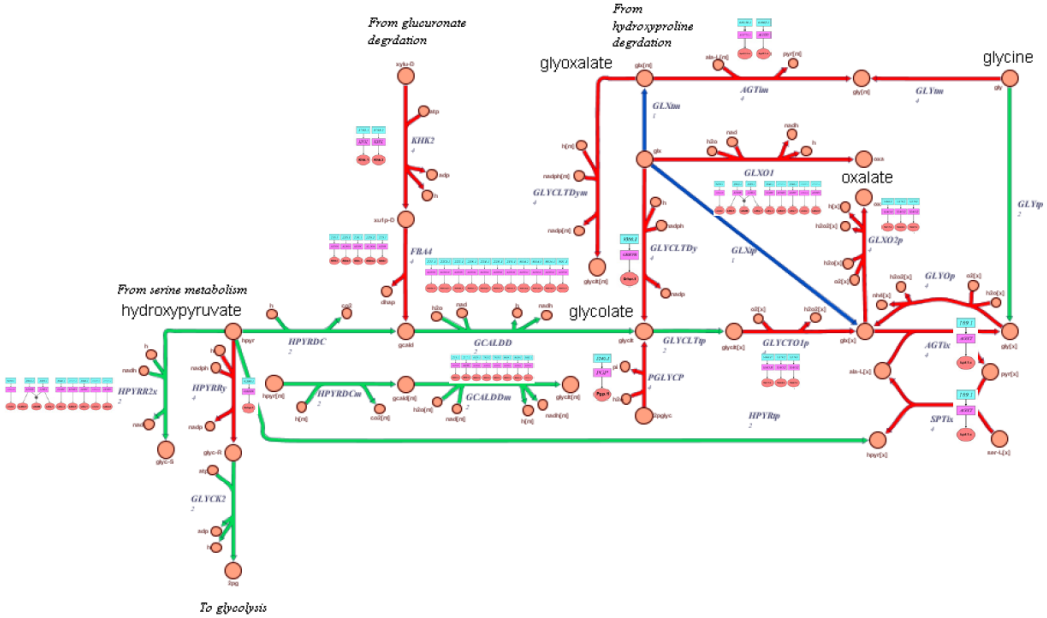


Fig. 1. Human metabolic knowledge landscape. Colors represent the percentage of reactions within a pathway that have a confidence score of 3 (biochemical or genetic evidence), 2 (physiological data or evidence from a nonhuman mammalian cell), 1 (modeling evidence), or 0 (unevaluated). Metabolic pathways (primarily defined by the Kyoto Encyclopedia of Genes and Genomes LIGAND database) were classified into three categories based on their level of characterization as detailed in the text.

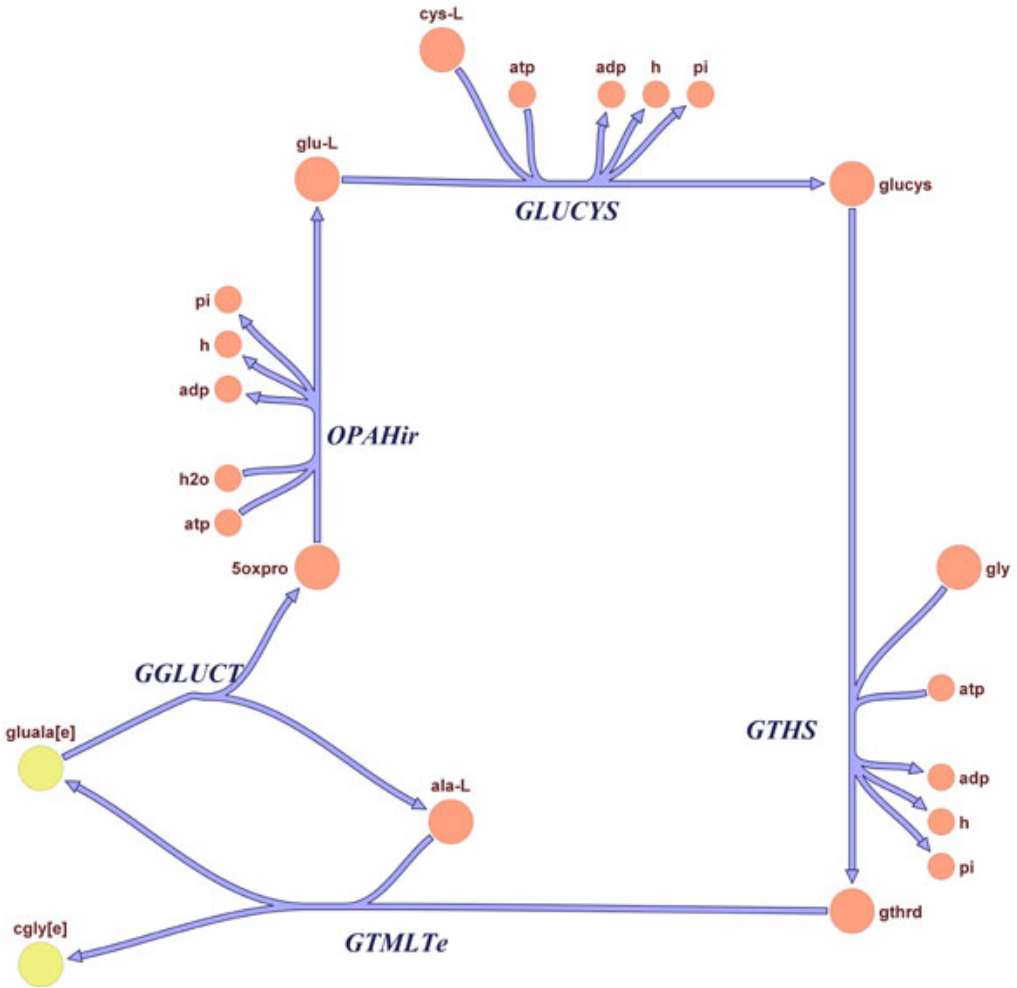


Category II pathway - glyoxylate to glycine

Coupled reaction sets - subnetworks

- Coupled reaction sets consist of reactions that are active together in functional states of a network
- Flux coupling analysis was used to identify coupled reaction sets under aerobic glucose metabolic conditions
- Example: glutathione reaction set

Glutathione reaction set



Both glutathione synthetase and glutamate-cysteine ligase deficiencies lead to hemolytic anemia.

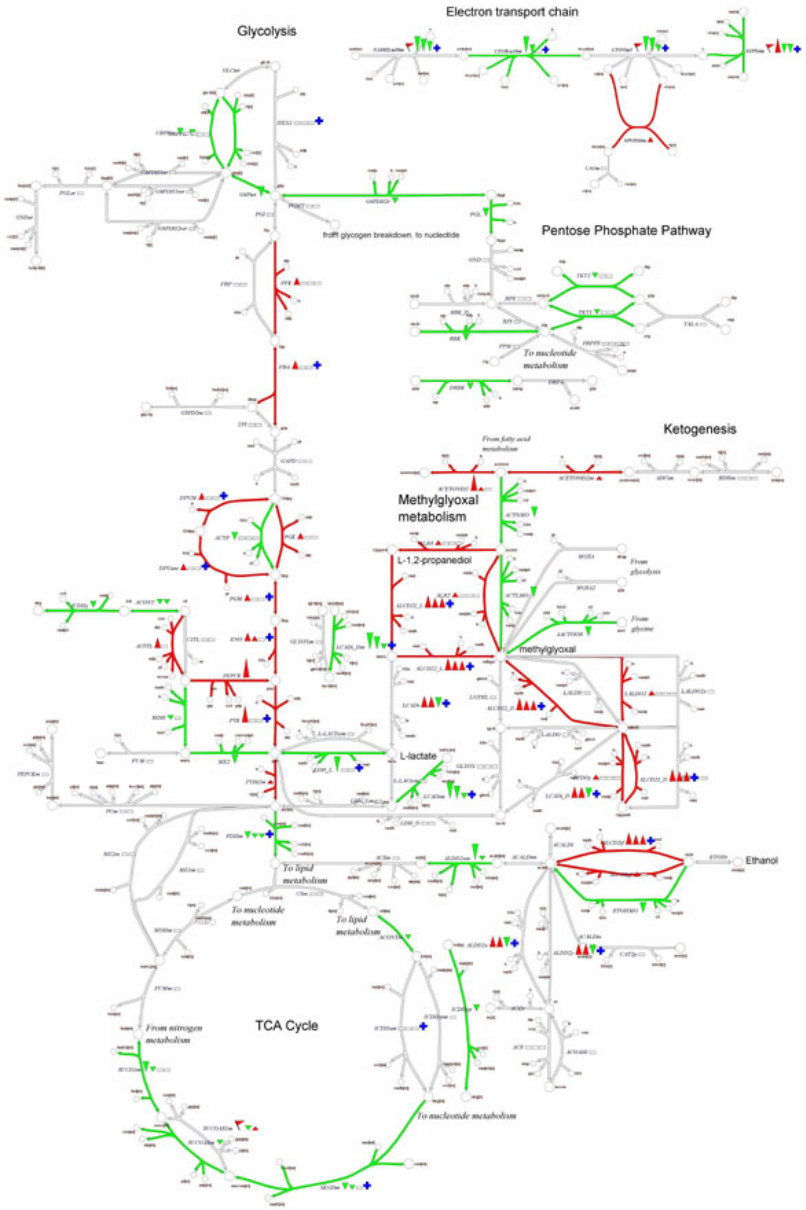
Systems biology -OMICS

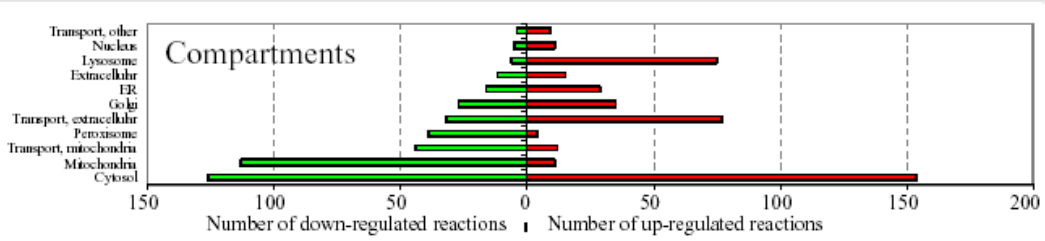
- transcriptomic
- proteomic
- fluxomic
- metabolomic

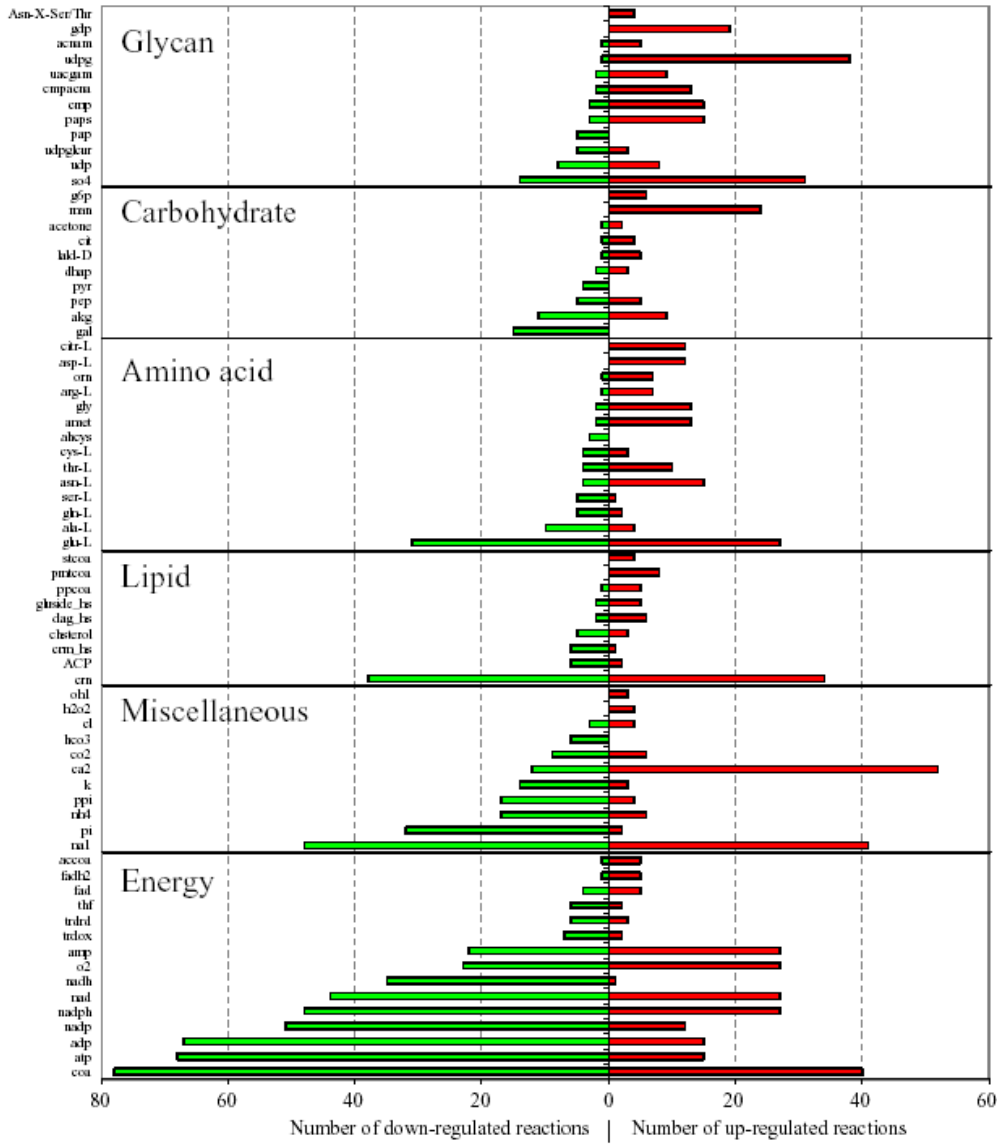
Combining expression data with metabolic networks - an example

- Gastric bypass surgery effects on skeletal muscle metabolism
- Gene expression data before and after surgery
- Up-regulated anaerobic metabolism & down-regulated oxidative phosphorylation after surgery
- Genes in glycolysis, pentose phosphate pathway, methylglyocal metabolism, oxidative phosphorylation expression change

Expression change related to gastric bypass surgery




Compartments
Transport, other
Nucleus
Lysosome
Extracellular
ER
Golgi
Transport, extracellular
Peroxisome
Transport, mitochondria
Mitochondria
Cytosol
150
100
50
0
50
100
150
200
Number of down-regulated reactions
Number of up-regulated reactions


Glycan
Carbohydrate
Amino acid
Lipid
Miscellaneous
Energy
80
60
40
20
0
20
40
60
Number of down-regulated reactions
Number of up-regulated reactions

Used article

- Duarte, Becker, Jamshidi, Thiele, Mo, Vo, Srivas & Paulsson
  Global reconstruction of the human metabolic network based on genomic and bibliomic data
  PNAS Feb 6th, vol 104 1777-1782
- Accessible at http://bigg.ucsd.edu

Single Value Decomposition (SVD)

Example [edit]

Consider the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix}$$

Its singular value decomposition is

$$U = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 2.236 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, V^* = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0.447 & 0 & 0 & 0 & 0.894 \\ 0 & 0 & 0 & 1 & 0 \\ -0.894 & 0 & 0 & 0 & 0.447 \end{bmatrix}$$

that is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 2.236 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0.447 & 0 & 0 & 0 & 0.894 \\ 0 & 0 & 0 & 1 & 0 \\ -0.894 & 0 & 0 & 0 & 0.447 \end{bmatrix}$$

Notice above that Σ only has values in its diagonal. Furthermore, as you can see below, multiplying the matrices U and V^* by their transpose yield an identity matrix.

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & 0 & 0.447 & 0 & -0.894 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.894 & 0 & 0.447 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0.447 & 0 & 0 & 0 & 0.894 \\ 0 & 0 & 0 & 1 & 0 \\ -0.894 & 0 & 0 & 0 & 0.447 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$