

# Human Metabolic Network Reconstruction



Bernhard Ø. Palsson, Ph.D.  
Departments of Bioengineering and Medicine  
University of California, San Diego



## Why reconstruct a human metabolic network?

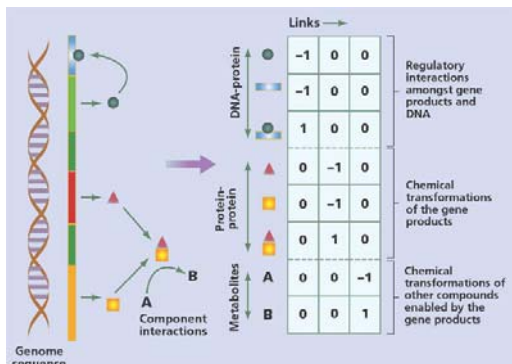
Metabolism is a key contributor to major human diseases, including diabetes, obesity, cardiovascular disease and cancer. The recent sequencing and annotation of the human genome sequence [1] along with the diverse high-throughput data sets available for human disease states have spawned a new era in biomedicine, holding the promise of an understanding of human biology to a degree heretofore unachieved. Integrating these high throughput and legacy biochemical data has inherent bioinformatic challenges which are best met by a systems biology approach. Our goal is to develop a genome-scale reconstruction of human cellular metabolism based on heterogeneous data types and accounting for cell and tissue specificity that can be used for predictive simulations of physiological and pathophysiological metabolic states.

### A human metabolic reconstruction will:

- Enable progression from a 'component annotation' of the genome to a 'systemic annotation' [2].
- Provide a bridge between high-throughput data and *in silico* analysis.
- Provide a valuable resource for other research groups as demonstrated by the growing number of publications using our reconstructed networks for *E. coli* and yeast [3,4].
- Provide a framework for studying human disease.

## A systemic annotation of the human genome

Information on cellular components and their interactions can be extracted from the human genome annotation and used to form a genome-scale interaction network, or systemic genome annotation. An interaction network provides a common denominator for many disease studies, making this undertaking particularly relevant to genomically-enabled research and medicine.



Genome-scale interaction networks mathematically represent the stoichiometric relationship between molecular components encoded in the genome (rows) and their interactions (columns).

## How will the network be reconstructed?

The human cellular metabolic network will be reconstructed according to the following strategy:

**Step 1:** A list of **candidate genes** will be compiled from genome annotation databases.

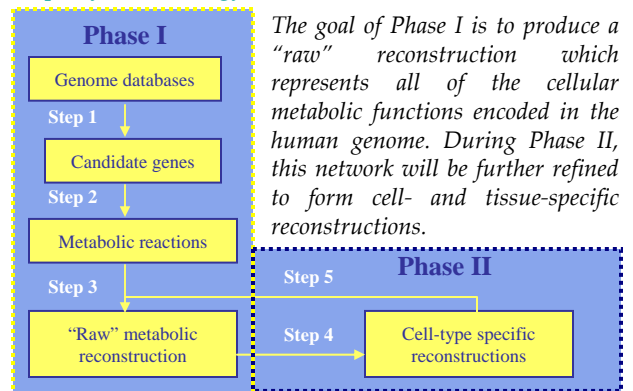
**Step 2:** The list of candidate metabolic genes will be mapped to **metabolic reactions** in the Kyoto Encyclopedia of Genes and Genomes [5] and the current yeast metabolic reconstruction [6].

**Step 3:** The pathway-associated reaction list will be **manually curated** using the resources listed below and assembled into a "raw" network reconstruction.

**Step 4:** The raw reconstruction will be further refined to form **cell-type specific** reconstructions.

**Step 5:** Literature and experimental data will be used to **continually update and expand** the metabolic network.

Network curation is exceedingly labor-intensive due to the manual curation process, however phase I (steps 1-3) is nearing completion. The contents can be found at <http://systemsbiology.ucsd.edu>.



## Resources for manual network curation

### Genome annotation databases

- NCBI Entrez Gene <http://www.ncbi.nlm.nih.gov/entrez/>
- Ensembl <http://www.ensembl.org/>
- H-Invitational Database <http://www.h-invitational.jp/>
- UCSC Genome Bioinformatics <http://genome.ucsc.edu/>
- GeneCards <http://bioinformatics.weizmann.ac.il/cards/>
- Human Genome Database <http://gdbwww.gdb.org/>

### Enzyme, reaction, and pathway databases

- Human Protein Reference Database <http://www.hprd.org/>
- BRENDA <http://www.brenda.uni-koeln.de/>
- KEGG LIGAND <http://www.genome.jp/ligand/>
- ExPASy Pathways <http://www.expasy.org/>
- Reactome <http://www.reactome.org/>

## What data is included in a reconstruction?

Genome-scale network reconstructions include a variety of biological data types (genomic, transcriptomic, proteomic, metabolomic, phenomic) that are manually collected from the literature and biological databases. Examples of how these data are used in a reconstruction are described below.

**Functional annotations:** Each gene's functional annotation is verified based on evidence from sequence analysis, cDNA cloning and expression, and biochemical characterization.

**Gene-protein-reaction associations:** Evidence from sequence analysis, gene knockout studies, and protein purifications are used to form gene-protein-reaction associations that represent alternative splicing, isozymes, multimeric proteins, and protein complexes (see panel below).

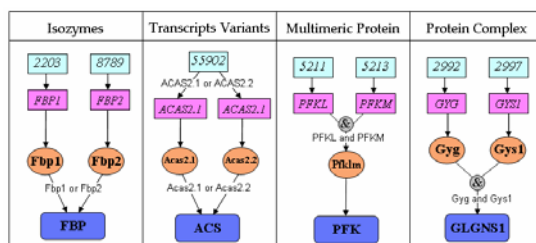
**Compartmental localization:** Protein localization data is used to compartmentalize proteins in the cytosol, mitochondrion, nucleus, endoplasmic reticulum, Golgi apparatus, lysosome, peroxisome, and/or extracellular space.

**Enzymatic functions:** Biochemical and thermodynamic data are used to determine reaction stoichiometry and reversibility as well as preferred substrates and cofactors.

**Tissue distributions:** Northern blots and microarray data provide information on each gene's relative expression level in various tissue categories: neural, digestive, connective, reproductive, blood/lymphatic, muscle, liver, lung, and kidney/bladder.

## Gene-protein-reaction associations

Gene-protein-reaction associations are Boolean logic statements that describe the functional relationship between open reading frames, transcripts, proteins, and reactions. Four types are included in the human metabolic reconstruction.



**Isozymes:** *FBP1* and *FBP2* encode fructose-1,6-bisphosphatases. *Fbp1* and/or *Fbp2* must be present for the corresponding reaction to occur.

**Transcript Variants:** Locus 55902 encodes two alternatively spliced transcripts which both produce acetyl-Coenzyme A synthetases. At least one transcript must be present for the protein to exist.

**Multimeric Proteins:** The phosphofructokinase *Pfklm* is a heterotetramer of L and M subunits. Both transcripts must be present to produce this protein.

**Protein Complex:** The glycogen synthase complex is composed of *Gyg* (glycogenin) and *Gys1* (glycogen synthase). Both proteins must be present for the reaction to occur.

## What will the reconstruction enable?

This project is a long term effort which is anticipated to become a useful resource for the biomedical research community at large, enabling:

1. The **dissemination of content** for other investigators to use for research purposes [3,4].
2. The identification of **physical gaps** in networks and **conceptual gaps** in our understanding of human metabolism.
3. The **visualization of biological datasets** on a comprehensive collection of human-specific maps.
4. Physiological and pathophysiological **simulations** in cell/tissue-specific, compartmentalized model.
5. The development of **translational technologies** such as personalized simulations of drug metabolism.
6. A **framework for future network expansion**, including further iterations with physiological data and additional cellular processes such as regulation and signaling.

## Current status

Due to a focused, parallelized effort, five man years of work have been completed for the *in silico Homo sapiens Build 1* which includes carbohydrate, lipid, amino acid, nucleotide, vitamin, cofactor, energy, glycan, and secondary metabolite metabolism.

The researchers involved have firsthand experience developing and analyzing metabolic networks for a variety of organisms, ranging from prokaryotes to eukaryotes, which have been widely used by the research community [3,4].

## Contact Information

Bernhard Ø. Palsson  
Departments of Bioengineering and Medicine  
University of California, San Diego  
9500 Gilman Drive, La Jolla, CA 92093-0412  
858.534.5668 palsson@ucsd.edu  
<http://systemsbiology.ucsd.edu>

## References

1. I.H.G.S Consortium, *Finishing the euchromatic sequence of the human genome*. Nature, 2004. 431(7011): p. 931-45.
2. Palsson, B., *Two-dimensional annotation of genomes*. Nat Biotechnol, 2004. 22(10): p. 1218-9.
3. Systems Biology Research Group, *Analysis of the E. coli network by other research groups*. [http://systemsbiology.ucsd.edu/organisms/ecoli/ecoli\\_others.html](http://systemsbiology.ucsd.edu/organisms/ecoli/ecoli_others.html)
4. Systems Biology Research Group, *Analysis of yeast model by other research groups*. [http://systemsbiology.ucsd.edu/organsims/yeast/yeast\\_others.html](http://systemsbiology.ucsd.edu/organsims/yeast/yeast_others.html)
5. Kanehisa, M., et al., *The KEGG databases at GenomeNet*. Nucleic Acids Res, 2002. 30(1): p. 42-6.
6. Duarte, N.C., M.J. Herrgard, and B. Palsson, *Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model*. Genome Res, 2004. 14(7): p. 1298-309.