UNIVERSITY OF CALIFORNIA, SAN DIEGO

GENOME-SCALE RECONSTRUCTION AND ANALYSIS OF EUKARYOTIC METABOLIC NETWORKS

A Dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Bioengineering

by

Natalie Christine Hurlen

Committee in charge:

Professor Bernhard Ø. Palsson, Chair
Professor Edward A. Dennis
Professor Jeffrey D. Esko
Professor Andrew D. McCulloch
Professor Lanping Amy Sung

2006

The Dissertation of Natalie Christine Hurlen is approved, and it is acceptable in quality and form for publication on microfilm:

____

____

____

____

____
Chair

University of California, San Diego

2006

This Dissertation is dedicated to

Mom, Pops, Mike, and “the boys”
My dearest Erik and the Hurlen family

and is in loving memory of

Bohdan and Olga Chapla
Henry and Eleanor Duarte

We've called the human genome the book of life, but it's really three books:
a history book, a shop manual and parts list, and a textbook of medicine more
profoundly detailed than ever.

Francis Collins, Director of the National Human Genome Research Institute

TABLE OF CONTENTS

Signature Page iii

Dedication.. iv

Epigraph.. v

Table of Contents.. vi

List of Figures xi

List of Tables.. xiv

Preface………………………………………………………………………........ xiv

Acknowledgments... xvi

Vita………….. xix

Abstract……………………………………………………………………….....xxi

CHAPTER 1: INTRODUCTION 1

1.1 What is systems biology? 1

1.2 Network reconstruction 1

1.3 Dissertation overview... 3

CHAPTER 2: TOOLS FOR GENOME-SCALE RECONSTRUCTION AND ANALYSIS 6

2.1 Reconstruction resources 6

2.1.1 Genome annotation databases 6

2.1.2 Enzyme, reaction, and pathway databases 7

2.1.3 Textbooks and review articles 8

2.2 Constraint-based modeling 8

2.2.1 Flux balance analysis 9

2.2.2 Phase planes and shadow price analysis 10

2.2.3 *In silico* gene deletions 11

CHAPTER 3: GENOME-SCALE RECONSTRUCTION OF THE *SACCHAROMYCES CEREVISIAE* METABOLIC NETWORK 15

3.1 Previous models of *S. cerevisiae* metabolism 15

3.2 Reconstruction of *S. cerevisiae* iND750 17

3.2.1 Reconstruction procedure 17

3.2.2 Summary of *S. cerevisiae* iND750 20

3.3 Validation of *S. cerevisiae* iND750 with a large-scale gene deletion study 24

3.3.1 Comparison of *in silico* and *in vitro* growth phenotypes 24

3.3.2 Analysis of false predictions by pathway and compartment 25

3.3.3 Analysis of false predictions by source 27

3.4 Conclusions 36

CHAPTER 4: INTEGRATED ANALYSIS OF *SACCHAROMYCES CEREVISIAE* METABOLIC PHENOTYPES 50

4.1 *In silico* characterization of metabolic phenotypes 50

4.1.1 *S. cerevisiae* phenotypic phase plane 51

4.1.2 Simulation of optimal metabolic phenotypes 52

4.1.3 Further characterization of oxidative-fermentative phases 54

4.2 Integration of experimental data and *in silico* predictions 56

4.3 Conclusions 57

CHAPTER 5: GENOME-SCALE RECONSTRUCTION OF THE GLOBAL HUMAN METABOLIC NETWORK 63

5.1 Previous models of mammalian metabolism 63

5.1.1 Small-scale, cell-specific metabolic models... 64

5.1.2 Mammalian genome sequencing and the first genome-scale model 65

5.1.3 Top-down reconstructions of human metabolism 66

5.2 Manual reconstruction of the global human metabolic network 67

5.2.1 Overall reconstruction strategy 67

5.2.2 Initial component list 68

5.2.3 Curation of network content 68

5.2.4 Functional validation and gap analysis 72

5.2.5 The result, *H. sapiens* Recon 1 72

5.3 Knowledge landscapes 74

5.4 Conclusions 76

CHAPTER 6: GENE EXPRESSION ANALYSIS IN THE CONTEXT OF A GENOME-SCALE NETWORK 97

6.1 Integrated analysis of skeletal muscle metabolism 97

6.1.1 The obesity epidemic 98

6.1.2 Metabolic gene expression in obese skeletal muscle 99

6.2 Tailoring the global metabolic network with pathway analysis 103

6.3 Conclusions 106

CHAPTER 7: CONCLUSIONS 122

7.1 Contributions to the field 123

7.1.1 Advancements in reconstructing eukaryotic metabolic networks 123

7.1.2 Identification of metabolic knowledge gaps.. 124

7.1.3 Integrated analysis of metabolic phenotypes....................................... 124

7.2 What's next for *S. cerevisiae* iND750 and *H. sapiens* Recon 1? 125

7.3 Improving genome-scale reconstruction and analysis of eukaryotes.................. 127

7.3.1 Enhancing component representations... 127

7.3.2 Improving communication with the research community.................... 130

7.3.3 Fine-tuning integration of gene-expression data.................................. 130

7.4 Concluding remarks.. 132

BIBLIOGRAPHY.. 136

LIST OF FIGURES

Figure 1.1: The hierarchical relationship of network components. 5

Figure 2.1: Overview of constraint-based modeling. 14

Figure 3.1: Overall reconstruction strategy for *S. cerevisiae* iND750. 38

Figure 3.2: Gene-protein-reaction associations in *S. cerevisiae* iND750. 39

Figure 3.3: Compartmental distribution of *S. cerevisiae* iND750's metabolites. 40

Figure 3.4: The complete *S. cerevisiae* iND750 metabolic map. 41

Figure 3.5: False predictions by compartment (*A*) and metabolic pathway (*B*). 42

Figure 3.6: Mispredicted gene deletion phenotypes in central metabolism. 43

Figure 3.7: Mispredicted gene deletion phenotypes in amino acid metabolism. 44

Figure 3.8: Breakdown of the false predictions by source.. 45

Figure 4.1: The yeast glucose-oxygen phenotypic phase plane. 59

Figure 4.2: Yeast’s optimal growth behaviors as a function of oxygen availability.... 60

Figure 4.3: Growth experiments shown on the phenotypic phase plane. 61

Figure 5.1: Timeline of the Human Genome Project. 78

Figure 5.2: The development, dissemination, and possible applications of *H. sapiens* Recon 1. 79

Figure 5.3: Snapshot of the initial human metabolic component list. 80

Figure 5.4: Snapshot of the human gene index. 81

Figure 5.5: Assembling reconstruction data in SimPheny. 82

Figure 5.6: Timeline of the human metabolic reconstruction. 83

Figure 5.7: Glycan representations in *Homo sapiens* Recon 1.................................... 84

Figure 5.8: Snapshot of the BiGG human database. 85

Figure 5.9: Global human metabolic knowledge landscape.. 86

Figure 5.10: Knowledge landscape for heparan sulfate degradation. 87

Figure 5.11: Knowledge landscape for glyoxylate and dicarboxylate metabolism. 88

Figure 5.12: Knowledge landscape for vitamin C metabolism. 89

Figure 6.1: Schematic of Roux-en-Y gastric bypass surgery. 108

Figure 6.2 Expression changes in central metabolism post-gastric bypass............... 109

Figure 6.3 Expression changes in oxidative phosphorylation and reactive oxygen species (ROS) detoxification post-gastric bypass. 110

Figure 6.4: Expression changes in collagen *(A)* and glycosaminoglycan *(B)* metabolism post-gastric bypass. 111

Figure 6.5 Compartmentalization and metabolite connectivity of up-regulated and down-regulated reaction networks post-gastric bypass. 112

Figure 6.6: Expression differences in central metabolism in lean versus morbidly obese subjects. 113

Figure 6.7: Expression differences in oxidative phosphorylation in lean versus morbidly obese subjects. 114

Figure 6.8: Hierarchical clustering of pathway analysis for gene expression in 79 healthy human tissues. 115

Figure 7.1: Compartmental distribution of *H. sapiens* Recon 1's metabolites. 133

Figure 7.2: Number of transcripts per open reading frame (ORF) in the human gene index. 134

LIST OF TABLES

Table 3.1: Comparison of *S. cerevisiae* iFF708 and iND750......................................46

Table 3.2: Comparison of transport reactions in *S. cerevisiae* iND750......................47

Table 3.3: Model changes suggested by the gene deletion study................................48

Table 4.1: Comparison of *in silico* and *in vitro* flux measurements. 62

Table 5.1: Comparison of the initial and finished human genome sequences. 90

Table 5.2: Divisions of the initial component list for parallel manual curation...........91

Table 5.3: Examples of metabolic objectives used for functional validation. 92

Table 5.4: Classification of network gaps...93

Table 5.5: *H. sapiens* Recon 1 network statistics..94

Table 5.6: Examples of queries available at the BiGG Human Database...................95

Table 5.7: The BiGG Human Database links to several public databases..................96

Table 6.1: Weight classification by body mass index (BMI)...................................116

Table 6.2: Expression changes in oxidative phosphorylation for the gastric bypass study. 117

Table 6.3: Expression changes in glycolysis/gluconeogenesis for the gastric bypass study. 119

Table 6.4: Expression changes in pentose phosphate pathway for the gastric bypass study. 120

Table 6.5: Protocol for mapping Affymetrix present/absent calls to on/off calls. 121

Table 7.1: Candidates for cell-specific human metabolic models.............................135

PREFACE

On assembling this Dissertation, I realized that the Human Genome Project (HGP) has been closely intertwined with many aspects of my academic career. While this is in part due to good timing, I have also been fortunate to work with many pioneering scientists who have played pivotal roles in various stages of the project. As an undergraduate at Boston University (BU), I conducted my senior research thesis under the supervision of Dr. Charles DeLisi, who led the effort to launch the HGP while he was a scientist at the Department of Energy in the mid-1980s. Dr. DeLisi has since received the Presidential Citizens Medal from President Bill Clinton for his HGP efforts, and has made significant strides towards training a new generation of integrative, multidisciplinary scientists by establishing one of the nation's first and largest bioinformatics graduate programs at BU. The inaugural year of the bioinformatics program coincided with my senior year of college, and as a result I had the chance to take newly offered graduate courses on biological database analysis as well as play an active role in establishing research collaborations between the College of Engineering and BU's School of Medicine.

The skills I learned at BU proved valuable to a small, San Diego start-up company named Egea Biosciences, at which I worked as a Bioinformatics Research Specialist from 2001-2002. At the time, the company consisted of less than 15 employees, all of whom directly reported to President and Chief Executive Officer Dr. Glen Evans. Prior to founding Egea, Dr. Evans was a principal investigator in the HGP, leading the task force to sequence human chromosome 11 at The University of Texas Southwestern Medical Center at Dallas. I can vividly remember Dr. Evans' excitement on the day that the initial human genome sequence was released, and the February 15, 2001 edition of *Nature* that he gave me retains a prominent spot on my bookshelf in honor of his great work and that of hundreds of other scientists worldwide.

It seems only fitting that when I left Egea to return to my graduate studies full-time, I was immediately drawn to the work of Dr. Bernhard Palsson's Genetic Circuits Research Group (now known as the Systems Biology Research Group or SBRG). Their research required my bioinformatic expertise, which included an array of data analysis and mining tools, but also included components of mathematical modeling and experimental studies, two topics that I was eager to learn more about. While my first projects in the SBRG were focused on reconstructing and modeling the yeast *Saccharomyces cerevisiae*, the release of a finished human genome sequence in 2003 opened up possibilities for new avenues of research, and in November 2004 I was asked to organize a team of students to construct the first genome-scale model of human metabolism. Thus, this Dissertation reports the first of many milestones in human systems biology, which was enabled by an extraordinarily ambitious research project prompted by Dr. Charles DeLisi in 1985.

La Jolla

November 2006

ACKNOWLEDGMENTS

I am deeply indebted to the kindness, support, and encouragement of many people who have helped me get to where I am today. I would like to acknowledge the National Science Foundation and the National Institutes of Health, whose research and training grants that have funded my graduate studies at UCSD. I am also grateful to my advisor Dr. Bernhard Palsson for his generous support. Dr. Palsson, thanks for recognizing my potential and for giving me ample opportunities to develop and showcase my communication, leadership, and organizational skills. I would also like to acknowledge Drs. Dennis, Esko, McCulloch, and Sung, my doctoral committee members. Thank you for giving me a broader perspective of how this work fits into the biomedical community at large and for your guidance in preparing this Dissertation.

I would like to extend special thanks to Dr. Shu Chien, who played an integral role in my decision to come to UCSD, and Dr. Shankar Subramaniam and Ms. Irene Jacobo, who helped me through many challenges during the first years of my graduate career. Dr. Chien, having lunch with you during my recruitment trip reaffirmed that UCSD is where I wanted to be. Dr. Subramaniam, you have truly bent over backwards to help me make the most of my graduate studies, and I am sincerely grateful for your advice. Irene, you are without a doubt one of the department's greatest assets, and I am very thankful for your academic and personal support.

For the past four years, I have been fortunate to work an extraordinary group of bright, cooperative, and entertaining colleagues in the Systems Biology Research Group. I would first like to recognize the human reconstruction team (a.k.a., the

"BiGG HuRT") – Scott Becker, Neema Jamshidi, Ines Thiele, Monica Mo, Thuy Vo, and Rohith Srivas – thank you for your dedication and hard work. I would also like to acknowledge Jan Schellenberger and Tom Conrad for putting together the BiGG Human Database. I am grateful to Iman Famili, who was my mentor when I first joined the SBRG, and Jennie Reed, who's been a great friend and co-worker for many years. Iman, we make a terrific team, both in the lab and on the dance floor! Jennie, your patience and resourcefulness never cease to amaze me; thank you for all of your advice throughout the years (and the frosting too!) JR, IT, MM, TV, IF: thank you for all of the great girls' nights. SAB & NJ: consider yourselves warned that I WILL get you back someday! CB: thanks for helping me stay slim by mooching off my food. MH: I promise to take good care of Hissy in Arizona! AR: Your enthusiasm and positive energy are sorely missed; hopefully we can reunite in Vegas! AJ, VP, EK: thanks for making lab retreats memorable. AF: I swear we never took your Padres cozy. TEA: Thanks for the great North County restaurant tips. NP: Thanks for reminding me why it's not a good idea to bring leftovers for lunch. JP: You will go down in history as the only person I know who's brave enough to eat cow brains. KA & IT, my fellow night owls: I know I can always count on you to be awake at 3am!

Finally, I would like to thank my family for teaching me to work hard, make things "pretty," and have fun. Pops, thanks for instilling in me the importance of discipline, organization, and integrity. Mom, thank you for nurturing my aesthetic eye and for being my greatest ally. Mike, thanks for including me in all of your fun adventures. And to my husband, Erik, thank you for giving me the courage to try many new things (such as vegetables that don't begin with the letter 'c') and for entertaining my questions on everything from bike gear mechanics to French pronunciations!

The text of Chapters 2 and 4, in part or in full, is a reprint of the material as it appears in N.C. Duarte, B.O. Palsson, and P. Fu. 2004. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. *BMC Genomics*, **5**:63. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

The text of Chapter 3, in part or in full, is a reprint of the material as it appears in N.C. Duarte, M.J. Herrgard, and B.O. Palsson. 2004. Characterization and validation of *Saccharomyces cerevisiae* iND750: a fully compartmentalized genome-scale metabolic model. *Genome Res* **14**:1298-309. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

The text of Chapters 5 & 6, in part or in full, is a reprint of the material as it appears in N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.O. Palsson. 2006. Global reconstruction of human metabolic network based on genomic and bibliomic data. Submitted to *Proc Natl Acad Sci U.S.A*. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

VITA

2000	B.S., Biomedical Engineering, Summa cum laude Boston University
2000-2006	Graduate Student Researcher Systems Biology Research Group University of California, San Diego
2001-2002	Bioinformatics Research Specialist Egea Biosciences, San Diego, California
2003	M.S., Bioengineering University of California, San Diego
2006	Ph.D., Bioengineering University of California, San Diego

PUBLICATIONS

Duarte NC, Herrgard MJ, Palsson BO: Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Biology* 14(7):1298-309, 2004.

Duarte NC, Palsson BO, Fu P: Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. *BMC Genomics* 5(1):63, 2004.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* (submitted), 2006.

ORAL PRESENTATIONS

Duarte NC, Palsson BO: Overview of the systems biology research group. *17th Annual Bioengineering Research Symposium*, San Diego, CA, 2003.

Duarte NC, Herrgard MJ, Palsson BO: Genome-scale reconstruction of *Saccharomyces cerevisiae*. *UC Systemwide Bioengineering Symposium*, Irvine, CA, 2004.

Duarte NC, Herrgard MJ, Palsson BO: Genome-scale reconstruction of

Saccharomyces cerevisiae. *Yeast Genetics and Molecular Biology Meeting*, Seattle, WA, 2004.

POSTER PRESENTATIONS

Duarte NC, Herrgard MJ, Palsson BO: *Saccharomyces cerevisiae* iND750: A fully compartmentalized genome-scale metabolic model. *4th International Conference on Systems Biology*, St. Louis, MO, 2003.

Duarte NC, Herrgard MJ, Palsson BO: *Saccharomyces cerevisiae* iND750: A fully compartmentalized genome-scale metabolic model. *18th Annual Bioengineering Research Symposium*, San Diego, CA, 2004.

Duarte NC, Herrgard MJ, Palsson BO: Reconstruction and validation of *Saccharomyces cerevisiae* iND750. *Yeast Genetics and Molecular Biology Meeting*, Seattle, WA, 2004.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: *Homo sapiens* Recon 1: A genome-scale reconstruction of human cellular metabolism. *Systems to Synthesis Symposium, San Diego Consortium for Systems Biology*, San Diego, CA, 2006.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: *Homo sapiens* Recon 1: A genome-scale reconstruction of human cellular metabolism. *Genetics Society of America Annual Meeting*. San Diego, CA, 2006.

PATENTS

Palsson BO, Duarte NC, Becker SA, Jamshidi N: Genome-scale reconstruction of human metabolism. US Patent Application No. 20040029149, February 12, 2004.

Note that some records are in my previous name of Natalie Christine Duarte.

ABSTRACT OF THE DISSERTATION

GENOME-SCALE RECONSTRUCTION AND ANALYSIS OF EUKARYOTIC METABOLIC NETWORKS

by

Natalie Christine Hurlen

Doctor of Philosophy in Bioengineering

University of California, San Diego, 2006

Professor Bernhard Ø. Palsson, Chair

Cells are comprised of complex, highly integrated networks of genes, proteins, and chemical compounds that interact with one another to achieve biological functions. A goal of systems biology is to develop comprehensive reconstructions of these networks in order to study their emergent properties. With the growing availability of whole genome sequences, cellular ‘part lists’ can now be defined for many organisms. The procedure for assembling genome-scale microbial networks is well established. However, such efforts have been limited for eukaryotes, especially in multicellular species. Thus, the overall goal of this Dissertation was to advance the reconstruction and analysis of eukaryotic systems by developing genome-scale metabolic models of *Saccharomyces cerevisiae* and a generic human cell.

We first describe the reconstruction of *S. cerevisiae* iND750, a fully compartmentalized metabolic network that includes systemic gene-protein relationships, pH-specific metabolite formula and charge, and elementally and charge-balanced reactions. iND750 was manually assembled with component-by-component (*i.e.*, bottom-up) approach and then functionally validated by comparing its predictions of 4,200 gene deletion phenotypes to *in vitro* data.

Next we discuss the human reconstruction project, which required a combination of top-down and bottom-up approaches to construct a comprehensive, high quality network within a reasonable time frame. This entailed automated extraction of a candidate component list from the genome annotation and parallelized, manual curation by a team of researchers. The resultant network, named *Homo sapiens* Recon 1, collectively represents 1,497 genes, 2,005 proteins, and 3,311 reactions found in a variety of human cell types, and is the largest genome-scale reconstruction to date.

Finally, we demonstrate the applications of these networks as mathematical models and as a context for high-throughput data analysis. *In silico* and *in vitro* growth experiments revealed that yeast exhibits few optimal phenotypes over a range of glucose and oxygen uptake rates, and that there are distinct combinations of these rates that yield maximal biomass and ethanol production. Qualitative assessment of gene expression levels in obese skeletal muscle highlighted consistencies between metabolic states post-gastric bypass and under caloric restriction. Pathway analysis of gene expression data also provided to initial steps towards generating tissue-specific metabolic reconstructions.

CHAPTER 1: INTRODUCTION

1.1 What is systems biology?

The goal of systems science is to integrate information about individual components to study how a system behaves as a whole. While this contextual, holistic approach is an established practice in many fields, including engineering, sociology, political science, and economics [1-3], cell biology has historically favored reductionistic methods in which genes and enzymes are independently characterized in great detail. However, the advent of new high-throughput technologies has created an overwhelming supply of comprehensive, data-rich information that necessitates an integrative, multi-disciplinary approach. This has led to the burgeoning field of systems biology, in which the mathematical tools of systems science have been used to understand biological systems from a network perspective.

1.2 Network reconstruction

Networks are comprehensive reconstructions of a system's components and their interactions [4]. There are two general strategies for assembling a reconstruction: top-down and bottom-up. Top-down approaches rely on inference methods to identify and formulate relationships between network components. They are typically implemented in a computer, enabling rapid assembly of large, comprehensive networks. In contrast, bottom-up networks are manually assembled in a component-by-component manner based on direct physical evidence from multiple data sources.

Manual reconstruction can be a time-consuming and laborious process, but is oftentimes favored for modeling applications because it produces self-consistent networks. Thus, top-down and bottom-up methods each have distinct advantages and disadvantages, and the best choice for reconstructing a network of interest can depend on many factors, including time, data availability, and the number of components.

In systems biology, networks are commonly reconstructed at the cellular level and are considered ‘genome-scale’ if they include all of the components encoded by an organism’s DNA. The interactions between genes, proteins, and chemical compounds constitute the two-dimensional annotation of a genome [5, 6], and there are many parallels between their hierarchy and the one-dimensional, sequence-based annotation (Figure 1.1). Tools are available for automated, top-down assembly of cellular networks [7], and the bottom-up procedure for reconstructing microbial networks is well established [6]. However, there has been limited progress in extending these methods to eukaryotes, especially mammals. Therefore, the primary goals of this Dissertation were:

1. To develop more advanced representations of eukaryotic metabolic networks, first in the unicellular budding yeast *S. cerevisiae* and then in a generic human cell; and
2. To use these reconstructions for integrated analysis of metabolic behaviors under various genetic and environmental constraints, including single gene deletions and pathophysiological states.

1.3 Dissertation overview

This Dissertation begins with a primer on resources available for genome-scale reconstruction and analysis (Chapter 2). The following four chapters provide in-depth descriptions of our studies in yeast (Chapters 3-4) and humans (Chapters 5-6); detailed overviews of these chapters are provided below. Finally, we conclude with a summary of my contributions to systems biology and future extensions of this work (Chapter 7).

S. cerevisiae reconstruction and analysis

Chapter 3 describes the reconstruction and validation of *S. cerevisiae* iND750, a genome-scale model of yeast metabolism with 750 genes, 1149 reactions, and 646 metabolites. Unlike its predecessor, iND750 is fully compartmentalized, accounting for seven intracellular localizations (cytosol, mitochondrion, nucleus, endoplasmic reticulum, Golgi apparatus, peroxisome, and vacuole), directly incorporates logical relationships between genes and proteins, and contains carefully formulated compounds and reactions that satisfy elemental and charge balance constraints at pH 7.2. The model was functionally validated by comparing its *in silico* predictions of more than 4,200 gene deletion phenotypes to *in vitro* measurements.

In Chapter 4, we present an integrated computational and experimental study that analyzes yeast's growth behavior over a range of glucose and oxygen uptake conditions. We found that, as compared to *Escherichia coli*, yeast exhibits relatively few distinct metabolic phenotypes under these conditions. This chapter also introduces a method for computing *in silico* secretion profiles, which are convenient, graphical descriptions of allowable uptake and secretion rates during optimal growth states.

Human reconstruction and analysis

Chapter 5 describes the genome-scale reconstruction of *H. sapiens* Recon 1, a global metabolic network with 1,496 genes, 3,311 reactions, and 2,712 metabolites that are collectively found in a variety of human cell types. The network was reconstructed by a combination of top-down and bottom-up methods, including an exhaustive, manual survey of the scientific literature, and resulted in a detailed, quantitative assessment of the human metabolic knowledge landscape.

In Chapter 6, we provide illustrative examples of how high-throughput data (specifically, gene expression profiles) can be integrated with *H. sapiens* Recon 1 to interrogate metabolic states in obese skeletal muscle and generate automated reconstructions of cell-specific networks.

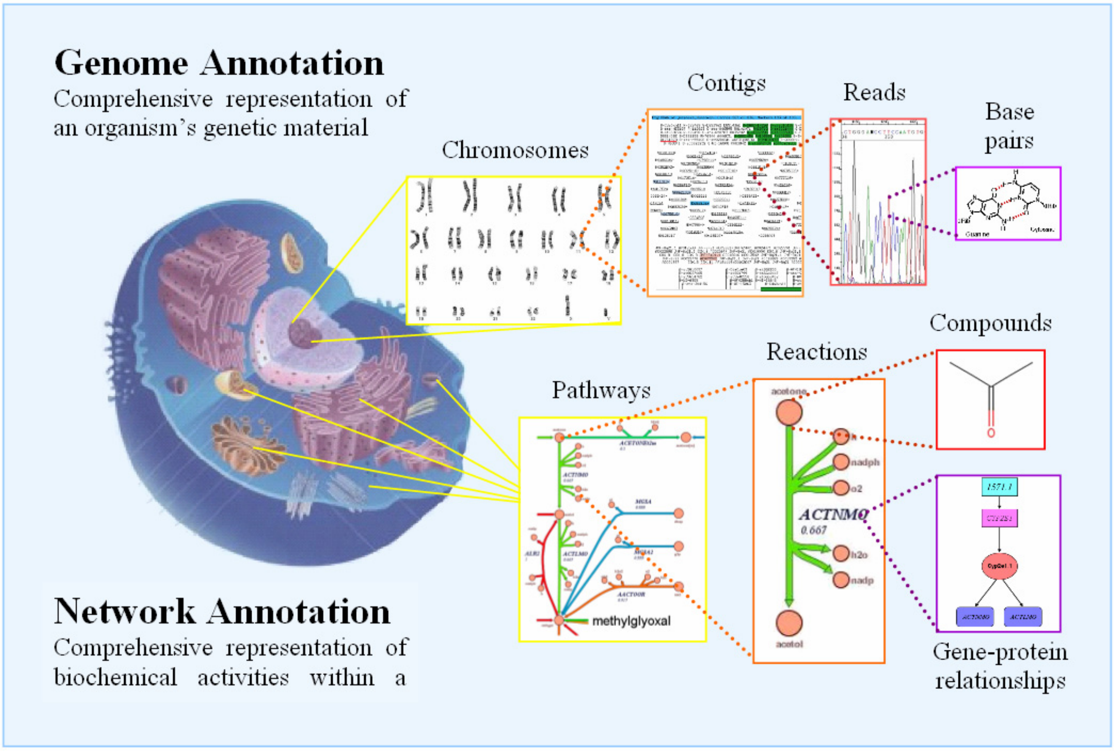

Genome Annotation
Comprehensive representation of an organism's genetic material
Chromosomes
Contigs
Reads
Base pairs
Compounds
Reactions
Pathways
acetone
nadph
o2
ACTNMO
0.667
h2o
nadp
acetol
methylglyoxal
Network Annotation
Comprehensive representation of biochemical activities within a
Gene-protein relationships

Figure 1.1: The hierarchical relationship of network components. There are many parallels between levels of detail in genome and network annotations. Chromosomes are analogous to cellular biochemical networks, which are described in terms of reactions. Contigs are delineated by sequence reads, which describe individual base pairs, the primary components of a sequence annotation. Similarly, reactions are catalyzed by enzymes, which are derived from genes and their transcripts, and act on compounds, the primary components in a network annotation. While the scope of genome annotation is clearly defined and has been significantly characterized, biochemical networks vary across cell types and states and, as a whole, are largely uncharacterized.

CHAPTER 2: TOOLS FOR GENOME-SCALE RECONSTRUCTION AND ANALYSIS

The wealth of data on metabolic components and their interactions enables construction of high confidence, biologically accurate networks that can be used to computationally interrogate metabolic states. This chapter introduces key resources for collecting these data (2.1) and provides a brief overview of the constraint-based modeling tools used in this Dissertation (2.2).

2.1 Reconstruction resources

Reconstructions rely on many types of biological evidence, including genetic, biochemical, and physiological data. While much of this information can be readily obtained from online databases (2.1.1 & 2.1.2), detailed, organism-specific data is must usually be extracted from the scientific literature (2.1.3). The main references for our yeast and human metabolic reconstructions are described here; other relevant databases for microbial reconstructions have been discussed elsewhere [6].

2.1.1 Genome annotation databases

Genome annotation databases are comprehensive, gene-centric resources that provide an abundance of information on gene identifiers (*e.g.*, abbreviations, names, and synonyms), gene-protein relationships (*e.g.*, alternative transcripts, isozymes, and protein complexes), and protein localization. They also typically contain links to primary research articles in PubMed [8], making them an ideal starting point for manual network curation. One of the largest and most widely used annotation databases is Entrez Gene [9], which consists of over 3,600 taxa to date. Most model organisms have species-specific genome annotation databases that are actively updated by their research community, such as the *Saccharomyces* Genome Database [10] and Comprehensive Yeast Genome Database [11]. Several human genome annotation databases have been generated computationally using data mining algorithms (GeneCards [12]) and high-throughput data (H-Invitational Database [13], Kyoto Encyclopedia of Genes and Genomes (KEGG) [14], and HumanCyc [15]).

2.1.2 Enzyme, reaction, and pathway databases

Metabolism has been extensively studied in a variety of organisms, resulting in a collective knowledge base that includes many mechanistic reactions and well-characterized interactions. Brenda [16] is an enzyme database with extensive records of cofactor preferences, kinetic measurements, and reaction stoichiometry for a variety of organisms. Entries are organized by Enzyme Commission (E.C.) numbers [17], which are universal (non-organism-specific) four digit codes based on enzymatic reaction mechanisms. E.C. numbers are also linked to metabolic maps in KEGG LIGAND, a suite of databases with detailed information on genes, reactions, and compounds involved in a variety of cellular processes [14]. Another universal vocabulary for describing gene and protein functions is Gene Ontology (GO) [18]. GO is hierarchically organized based on three primary classifications: molecular functions, biological processes, and cellular components. A convenient web-based application for searching and viewing these hierarchies is the AmiGO browser [18].

2.1.3 Textbooks and review articles

While the internet has facilitated the exchange and archive of tremendous amounts of genomic and biochemical data, textbooks and review articles are the most useful resources for collecting organism-specific physiological information. Biochemistry textbooks by Lubert Stryer [19] and Donald and Judith Voet [20] are excellent general resources, providing a basic overview of common metabolic functions. We also recommend Biochemical Pathways by Gerhard Michal [21], which includes detailed descriptions of Boehringer Mannheim wall charts, and Metabolism at a Glance by Jack Salway [22], which provides human-specific genetic and physiological data for select metabolic pathways. Additional texts specific to yeast and human biochemistry are also available [23-30]

2.2 Constraint-based modeling

Network reconstructions are the basis of mathematical models used in a variety of applications [4]. Constraint-based approaches have been successfully used in modeling microbial metabolism [31-33], and here they are used to simulate *S. cerevisiae* growth phenotypes (Chapters 3-4) and nearly 300 metabolic functions known to exist in human cells (Chapter 5). The basic premise of constraint-based modeling is to generate a solution space that consists of a collection of all possible cellular behaviors [34-38]. This space is defined by many factors, including genetics, network stoichiometry, thermodynamics, and environmental conditions, and can be further refined as additional constraints are introduced (Figure 2.1).

This section describes three constraint-based approaches used in this Dissertation: flux balance analysis (2.2.1), phenotypic phase plane analysis (2.2.2), and *in silico* gene deletion analysis (2.2.3). Additional information on the growing toolbox for constraint-based analysis can be found in a recent review [31].

2.2.1 Flux balance analysis

In flux balance analysis (FBA), stoichiometry of the metabolic reaction network and linear programming are used to compute optimal metabolic phenotypes [38, 39]. The relationship between metabolite concentrations, ***x***, and reaction activities, ***v***, is described by the dynamic mass balance equation [40, 41]:

$$\frac{dx}{dt} = S \bullet v$$

where ***S*** is an $m \times n$ matrix of stoichiometric coefficients, ***x*** is an $m \times 1$ vector of metabolite concentrations, and ***v*** is and $n \times 1$ vector of reaction activities. Thus, the rows of **S** correspond to the internal metabolites and the columns represent the reactions in the network. Under steady-state conditions, the dynamic mass balance equation simplifies to:

$$S \bullet v = 0$$

Since the number of reactions is often greater than the number of metabolites, the dynamic mass balance equation is underdetermined and contains multiple solutions. By defining an objective function (such as cellular biomass composition) along with a set of inputs and outputs that correspond to growth conditions, one can use standard linear programming techniques to determine a flux distribution that maximizes cell growth [36]. While the optimal value of the objective function will be unique, this value may be produced by more than one flux distribution. Additional methods have been developed to characterize these multiple alternative optima [42-44].

2.2.2 Phase planes and shadow price analysis

Phenotypic phase plane (PhPP) analysis explores how changes in two environmental variables, such as oxygen and a carbon uptake rates, affect optimal growth rates. Some applications include comparisons of growth phenotypes on alternate carbon sources [45], evaluation of microbial adaptability [46, 47], assessment of network functions and capacities [48], and investigation of gene regulation effects [49]. Thus, PhPP analysis provides a way to guide experiments and analyze phenotypic functions based on genome-scale metabolic networks.

In PhPP analysis, FBA (2.2.1) and linear programming are used to map all of the cellular growth conditions represented by two environmental variables onto a two-dimensional plane and identify phases with distinct metabolic pathway utilization patterns. Phases are determined by the calculation of shadow prices [38], which describe the sensitivity of the objective function (Z) to changes in the availability of each metabolite:

$$\gamma_i = \frac{-dZ}{db_i}$$

where b_i is the i^{th} metabolite and γ_i is the i^{th} shadow price. By definition, phases are regions of the PhPP in which all of the points have the same shadow prices. Phase boundaries therefore describe transitions between metabolic states, and can oftentimes highlight interesting, organism-specific relationships such as the optimal ratios of environmental conditions to produce maximal biomass yield (*i.e.*, lines of optimality) [38].

Analyzing the shadow prices of key metabolites across the PhPP can provide physiological interpretations of its phases. According to the convention defined in [38], a negative shadow price indicates that a metabolite is limiting, *e.g*, the value of the objective function increases if the metabolite's net production increases or its net consumption decreases, and a positive shadow price indicates that a metabolite is available in excess. A shadow price equal to zero indicates that a change in the availability of the metabolite does not affect the objective value. Secretion profiles can be generated for extracellular metabolites with null shadow prices by maximizing and minimizing their corresponding exchange fluxes.

2.2.3 *In silico* gene deletions

Quantitative data on growth rates of individual gene deletion strains can be directly compared to *in silico* predictions of metabolic phenotypes. Previous studies in *H. pylori* [50], *H. influenzae* [51], *E. coli* [52], and *S. cerevisiae* [53] have typically had accuracy rates of 60% to 90% under a variety of experimental conditions. Such comparisons enable identification of potential problem areas in the network, allow verification of hypothesized metabolic reactions, and suggest specific experiments that can be used to verify components of the network, such as the enzymatic function of particular genes [6].

The effects of a single gene deletion are simulated by constraining the flux through its corresponding reaction to zero. FBA is then performed to find the predicted growth rate of the *in silico* deletion strain. The deletion is considered deleterious if its optimal growth flux is lower than that of the wild type *in silico* model.

The text of this chapter, in part or in full, is a reprint of the material as it appears in N.C. Duarte, B.O. Palsson, and P. Fu. 2004. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. *BMC Genomics*, **5**:63. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

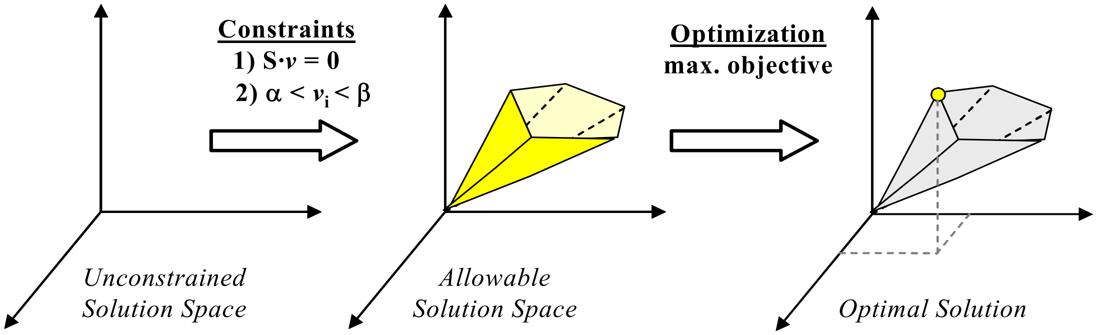


Figure 2.1: Overview of constraint-based modeling. The unconstrained solution space is defined by the flux through reactions in the stoichiometric network. Imposing steady-state conditions as well as upper and lower flux boundaries eliminates some potential cellular behaviors, resulting in an allowable solution space that contains many possible solutions. Linear optimization can be used to calculate optimal solutions for a defined objective function, which for flux balance analysis (2.2.1) is typically cellular biomass.

CHAPTER 3: GENOME-SCALE RECONSTRUCTION OF THE *SACCHAROMYCES CERVISIAE* METABOLIC NETWORK

S. cerevisiae, commonly known as baker's or brewer's yeast, is an important model organism. Many of its cellular processes, including metabolism, cell cycling, mRNA processing, and protein sorting, are generally conserved with higher eukaryotes. There are several technical advantages that make yeast an ideal experimental system, such as its non-pathogenicity, rapid growth, and malleable genetics, and this has resulted in a wealth of genetic, biochemical, and physiological data [54-56].

This chapter reports the reconstruction and validation of an expanded genome-scale model of *S. cerevisiae* metabolism [57]. We begin with a discussion of previous modeling efforts (3.1), which includes the original yeast reconstruction [58] that formed the basis of this work.

3.1 Previous models of *S. cerevisiae* metabolism

Several modeling approaches have been used to study yeast metabolism. Most flux-balance models (see 2.2.1) use small-scale network reconstructions for specific growth conditions, such as anaerobic, glucose-limited metabolism [59], aerobic growth on galactose [60] or growth on mixtures of glucose and ethanol [61]. Dynamic models of simplified central metabolic networks [62, 63] and full-scale kinetic models of glycolysis [64, 65] and the pentose phosphate pathway [66] have also appeared. These small-scale reconstructions have been useful for studying detailed metabolic events such as changes in individual metabolite concentrations and key flux splits. However, the applications of small-scale reconstructions are limited since many cellular processes are dependent on the interaction of components at the whole-cell level.

The sequencing and annotation of the *S. cerevisiae* genome [67] provided a parts list of cellular components and their interactions. This data, together with known physiological and biochemical data were used to reconstruct *S. cerevisiae* iFF708, the first genome-scale model of *S. cerevisiae* metabolism [58]. iFF708 contained a total of 708 open reading frames, 1175 metabolic reactions, and 733 metabolites compartmentalized between the cytosol and mitochondrion. The model was validated through *in silico* gene deletion studies [53] and the calculation of key physiological parameters [68].

Reconstruction of the *S. cerevisiae* metabolic network demonstrated that the constraint-based approach can be applied to networks of higher complexity, such as those with multiple compartments. A goal of this Dissertation was to expand the scope of this metabolic network by including a more biologically-accurate description of yeast's cellular components, namely: 1) the logical relationship between genes, transcripts, proteins, and reactions, 2) the cell-wide conservation of mass and charge through elementally and charge-balanced reactions, and 3) the full compartmentalization of yeast's metabolites and proteins. The expanded yeast metabolic network then served as a prototype of a fully compartmentalized, genome-scale model of human metabolism (Chapter 5).

3.2 Reconstruction of *S. cerevisiae* iND750

The previous genome-scale metabolic reconstruction of *S. cerevisiae* (iFF708; 3.1) was the starting point for reconstructing iND750, a fully compartmentalized yeast model that requires a cell-wide proton balance and includes associations between its genes, proteins, and reactions (Figure 3.1). This section summarizes the changes made to iFF708 as well as key properties of iND750.

3.2.1 Reconstruction procedure

Starting with the list of ORFs included in iFF708, the corresponding gene names, E.C. numbers (2.1.2) [17], and reactions were all re-evaluated to check their consistency with genome annotation databases and recently published reports. Special attention was given to compartmentalization, elemental and charge balancing of reactions, and the relationships between genes, proteins, and reactions, which are discussed below.

Compartmentalization

Since reactions in iFF708 were restricted to only the cytosol, mitochondria, and extracellular space, the localization of each gene product was revised to take into consideration the five additional compartments included in iND750 (peroxisome, endoplasmic reticulum, Golgi apparatus, nucleus, and vacuole). Information on the localization of the gene products was primarily taken from the *Saccharomyces* Genome Database [69] and Comprehensive Yeast Genome Database [70]. If there was little or no evidence that a gene product was found in a particular compartment, then it was assumed to be located in the cytosol. An additional assumption was also needed for membrane proteins, since oftentimes there was no evidence regarding the location of their catalytic domains. Unless there was evidence to the contrary, it was assumed that reactions catalyzed by membrane proteins occurred in the cytosol. Finally, all of the compartments were modeled as if there were only one boundary between the cytosol and its lumen. For example, since the mitochondria's intercompartmental space is considered to be equivalent to the cytosol in its metabolite and ion concentrations [20], proteins that are localized to these regions are considered cytosolic. Similarly, the cell wall and periplasmic space are both treated as part of extracellular compartment.

Intercompartmental transport

Additional transport reactions were needed to describe the exchange of compounds between iND750's eight cellular compartments. The transport processes across the plasma membrane have been well studied; many genes have been identified that encode transport proteins (see [23] and [26] for comprehensive list). These genes and their documented transport mechanisms have been included in iND750. In addition, many metabolites are known to diffuse across the yeast cell wall [23, 26]. For those compartments in which there was little information about transport processes, most of the exchange reactions had to be inferred. A primary assumption was that a particular compound was transported across various membranes by a similar process. For example, since tyrosine is known to cross the plasma membrane via proton symport, it was also assumed to be transported across the peroxisomal membrane by the same mechanism. Transport reactions were also inferred based on the known characteristics of some membranes, such as the nuclear membrane, which contains pores that allow substrates less than 9 nm or 60 kDa to pass freely [71].

Consequently, most of the compounds transported into and out of the nucleus are exchanged by simple diffusion.

Elemental and charge balancing

The reactions in iND740 are elementally and charge balanced. The formula and charge of the metabolites were determined based on their ionization state at a pH of 7.2. For simplicity, all of the compartments were assumed to have the same pH. By introducing ionized compounds, water molecules and protons that participate in the reactions are explicitly accounted for so that the reactions had no net charge change and obeyed elemental balances. Water molecules were allowed to freely diffuse into all of the compartments. However, the protons could only enter and leave the various compartments by participating in active transport reactions. Thus, the production and consumption of protons had to be balanced within each compartment.

Gene-protein-reaction associations

Unlike the iFF708, which only considered one-to-one associations between genes and reactions, the logical relationship between genes, proteins, and reactions are all modeled in iND750. To do this, the entry of each gene was examined to see if there was any evidence that its gene product was multifunctional, an isozyme, a protein subunit, or a participant in a protein complex. Multifunctional proteins were defined as those that can catalyze more than one reaction (Figure 3.2A). Distinct proteins that could catalyze the same reaction were labeled as isozymes (Figure 3.2B). Proteins were classified as multimeric if more than one transcript was required to catalyze an enzymatic function (Figure 3.2C). Key words used to identify multimeric proteins were “chains” or “subunits” of proteins. Proteins could also form complexes; this is defined as a functional entity in which proteins from different transcripts must act together to catalyze a reaction (Figure 3.2D). There were also more complex cases in which a protein belonging to a complex was made up of subunits, such as in the fatty acid synthase complex. Boolean logic statements [6] representing these relationships were formulated for all of iND750's genes and reactions.

Network assembly and testing

The updated metabolic network was constructed using the SimPheny™ software package (Genomatica, San Diego, CA). It was verified separately that iND750 is capable of predicting whole cell functions such as P/O ratios and byproduct secretion rates under a variety of conditions with similar or improved accuracy compared to iFF708 [68] (data not shown).

3.2.2 Summary of *S. cerevisiae* iND750

S. cerevisiae iND750 describes our current knowledge of yeast metabolism and is the first fully compartmentalized genome-scale reconstruction. It includes direct representations of gene-protein relationships and accounts for elemental and charge balancing. A summary of these new features is provided here.

Updated gene, reaction, and metabolite lists

The extent of the changes that were made to iFF708 to form iND750 is reflected in Table 3.1, which compares the number of genes, reactions, and metabolites in the two models. Nearly all of the genes in iFF708 are accounted for in iND750. The additional genes primarily encode tRNA synthetases (26 genes) and ATPases found in the vacuole and Golgi apparatus (13 genes). Both models also share a large number of metabolites, although the compartmental location of the metabolites has not been considered in this comparison. Most of the metabolites added to iND750 are found in reactions that have been expanded, *i.e.*, reactions that were lumped in iFF708 and are now included as individual steps or with distinct metabolites in the new model. For example, the replacement of a generic ceramide metabolite with two specific moieties led to the introduction of approximately 20 additional metabolites in subsequent reactions. The most notable difference between the models is in their reaction sets. Of iND750's 1149 reactions[1], only 56% are the same as those in iFF708 even after accounting for changes required for elemental and charge balancing. Most of these changes are the result of iND750's five additional compartments; many of the reactions that were previously listed as cytosolic were reassigned to a new compartment, and more than 80 reactions were added to represent the metabolite exchange for these five compartments. Also, as mentioned earlier, many types of metabolic reactions were expanded, especially in fatty acid degradation, where four individual steps in iND750 replaced the one lumped reaction for each fatty acid included in iFF708. Other changes that are not noted in Table 3.1 include: the introduction of a systemic definition of the associations between genes, proteins, and reactions, the removal of redundant compound abbreviations and duplicated reactions, and updates to gene names and E.C. numbers.

Full network compartmentalization

S. cerevisiae iND750 accounts for eight cellular compartments, three of which were included in iFF708 (extracellular space, cytosol, and mitochondria) and five

[1] Counting different isozymes as separate reactions, iND750 includes a total of 1489 reactions.

additional compartments (peroxisome, nucleus, Golgi apparatus, endoplasmic reticulum, and vacuole). To evaluate the connectivity of these compartments, iND750's 646 distinct metabolites were analyzed according to their compartmental location (Figure 3.3). Most notably, almost 90% of the metabolites appear in cytosolic reactions. Half of these metabolites are unique to the cytosol; this large percentage is not surprising since reactions were assigned to the cytosol by default. The seven other compartments vary significantly in their number of metabolites and connectivity. For example, more than 75 metabolites can be found in the mitochondria, extracellular space, and peroxisome. All of the metabolites in the extracellular space are shared with other compartments, whereas the mitochondria and peroxisome have a defined set of unique metabolites that do not appear in other compartments. The nucleus, Golgi apparatus, endoplasmic reticulum, and vacuole have less than 35 metabolites, almost all of which can be found in multiple compartments.

Developing a fully compartmentalized *S. cerevisiae* network required the addition of many intercompartmental transport reactions. Table 3.2 shows the 297 transport reactions included in iND750. The majority of these reactions represent transport across the plasma and mitochondrial membranes. The primary transport mechanisms across the plasma membrane and the intracellular membranes are noticeably different. Nearly two-thirds of the metabolites exchanged between the cytosol and the extracellular space occur by symport, typically a primary metabolite and proton transported in the same direction, whereas most of the metabolites exchanged between the intracellular compartments are transported by diffusion. The membranes also vary in their number of gene-associated reactions. The plasma membrane has the largest proportion of gene-associated reactions (almost 50%), while the nuclear, endoplasmic reticular, Golgi apparatus, and vacuolar membranes do not have any. As a result, many of the transport reactions across the intracellular membranes had to be inferred based on reactions known to take place in these compartments.

All of the compartments in iND750 were assumed to have a pH of 7.2. Consequently, the charge and formulae of all metabolites were determined by their ionization form at this pH. By including water molecules and protons in iND750's reactions, more than 99% could be written so that they were both elementally and charge-balanced. The few imbalanced reactions are typically those catalyzed by enzymes whose mechanism is not fully understood, such as biotin synthase (E.C. 2.8.1.6). Structuring the reactions in this manner forces the proton production and consumption to be balanced within each compartment and thus in the entire cell. This global proton balancing has implications for cellular growth, as has been demonstrated for *E. coli* grown on various carbon sources [72].

Addition of gene-protein-reaction associations

Unlike iFF708, which does not systematically represent the relationship between its genes and reactions, iND750's gene-protein-reaction associations can be viewed as graphical representations of the logical relationships between its ORFs, transcripts, proteins, and reactions. For example, proteins classified as multifunctional can catalyze more than one reaction (Figure 3.2A). Distinct proteins that can individually catalyze a reaction are defined as isozymes (Figure 3.2B). Multimeric proteins are defined as those formed by more than one transcript (Figure 3.2C). Finally, a protein complex is a set of proteins that are required to catalyze a reaction (Figure 3.2D).

Reaction and metabolite lists, gene-protein-reaction associations, and metabolic maps (Figure 3.4) for *S. cerevisiae* iND750 are available in Supplement A and at http://systemsbiology.ucsd.edu.

3.3 Validation of *S. cerevisiae* iND750 with a large-scale gene deletion study

Genome-scale networks can be used to predict metabolic phenotypes that can be verified experimentally. Comparing these *in silico* predictions to experimental data allows for the identification of possible problem areas in the network and can suggest specific experiments to probe network components in detail. Since iND750 represents the current understanding of yeast metabolism as completely as possible within a stoichiometric framework, analysis of its failure modes is important because they can be used to highlight inconsistencies in the body of information used in the reconstruction.

3.3.1 Comparison of *in silico* and *in vitro* growth phenotypes

S. cerevisiae iND750 was validated and interrogated in detail by comparing model predictions for deletion strain phenotypes with published results from two large-scale growth phenotyping studies [73, 74] in seven different media conditions. The media conditions included in this study were aerobic growth on glucose minimal media (MMD) and on rich media with six different carbon sources: glucose (YPD), galactose (YPGal), glucose-ethanol-glycerol mixed media (YPDGE), glycerol (YPG), ethanol (YPE), and lactate (YPL). *In silico* gene deletions were performed using established flux balance analysis procedures (2.2.1) [32, 34, 35]. In order to make the *in vivo* data and *in silico* predictions comparable, both were converted from continuous-value relative fitness scores to a discrete viable/retarded growth assessment for each gene deletion strain and condition (see [57] for detailed methods). The *in silico* phenotype predictions were classified into one of four categories: true positive (TP; experimentally and *in silico* viable), true negative (TN; experimentally and *in silico* growth retarded), false positive (FP; experimentally growth retarded, *in silico* viable), and false negative (FN, experimentally viable, *in silico* growth retarded). Deletion phenotype predictions for at least one condition were done for 682 of the total of 750 genes in the model and the predictions were classified as described above. No experimental deletion data was available for the remaining genes.

A total of 4,154 comparisons between *in silico* and *in vivo* deletions were analyzed. The overall correct prediction rate was 82.6%, which is similar to that obtained in more limited studies with other organisms as well as yeast [48, 50, 51, 53]. The true positive rate (true positive predictions/total number of *in vivo* normal growth phenotypes) was 96.6%, indicating that the model correctly captures the built-in redundancy in metabolism in that most gene deletions have no phenotypic effect under most conditions. However, the false positive rate (false positive predictions/total number of *in vivo* deleterious phenotypes) was 77.0% showing that less than one quarter of slow growth phenotypes were predicted correctly.

3.3.2 Analysis of false predictions by pathway and compartment

To further investigate the sources of the false predictions, their distribution was analyzed with respect to cellular compartments and metabolic subsystems. The overall false prediction rates as well as false negative and false positive rates for genes in particular cellular compartments are shown in Figure 3.5A. There was surprisingly large variability in the false prediction rates between genes in different compartments. The mitochondrial compartment had the highest overall error rate and most of these errors were false positive predictions. Since the mitochondria were shown to have a distinct set of metabolites (Figure 3.3), it seems surprising that iND750 may not have fully captured its unique role in cellular growth. Further analysis of the failure modes in terms of pathways (Figure 3.5B) revealed a high percentage of false prediction rates for genes in quinone biosynthesis (45.3%) and phospholipid biosynthesis (39.3%), suggesting that the model may not accurately represent mitochondrial maintenance. Also, in this model, we have assumed that the outer mitochondrial membrane is like a sieve allowing free diffusion of metabolites; however, there is evidence to suggest that the permeability of outer mitochondrial membrane may be regulated [75]. This variation in permeability may have important implications for controlling energy metabolism [76]. Other areas which were found to have high false prediction rates include oxidative phosphorylation in central metabolism (Figure 3.6) and amino acid biosynthesis (Figure 3.7).

Peroxisomal reactions were one of the most significant additions to iFF708 as the peroxisome has its own defined set of metabolites and plays a crucial role in the degradation of fatty acids. The high correct prediction rate obtained for peroxisomal genes (96.9%) indicates that the model fairly accurately accounts for the metabolic function of this important cellular compartment. Low false prediction rates were also obtained for genes involved in extracellular transport (3.2%), histidine metabolism (5.0%), and glutamate metabolism (5.1%). Overall, the distribution of the false predictions can be seen to be quite uneven with a few metabolic subsystems accounting for the majority of the problems.

3.3.3 Analysis of false predictions by source

A total of 246 gene knockouts had false predictions under one or more conditions. Causes of the false predictions were evaluated by studying the relevant literature on previously determined mutant phenotypes for each gene and by interrogating its role in the metabolic model. The results of this evaluation for each media condition as well as for all false positives and false negatives, for false predictions under a unique condition, and for all false predictions together are shown in Figure 3.8. The primary sources of false predictions were organized into 10 different categories (detailed in the caption for Figure 3.8). Overall, more than half of the false predictions can be accounted for by the involvement of the genes in other cellular processes in addition to metabolism (33.7%) and problems in the biomass composition assumed in the *in silico* deletion study (17.5%). Interestingly, the reasons for false positive and false negative predictions were quite different with majority of the false positives arising for the above mentioned reasons whereas the majority of false negatives could be traced to uncertainty in the *in silico* media composition (50.0%) and issues related to the gene-protein-reaction relationships in the model (18.4%). The different media conditions had similar distributions of the sources of false predictions, but the majority of the false predictions that arose because of missing *in silico* biomass components were related to essential genes. The sources of false predictions for genes with a unique false prediction under one experimental condition were also quite different from the overall pattern with a particularly high fraction of false predictions with no clear reason for the false assessment (25.5%).

Specific examples of false predictions in these categories are provided below (for a detailed description of all erroneous predictions, please refer to Supplement A). In many cases, the false predictions led to direct suggestions of how to potentially improve the model (Table 3.3) or of specific experiments that could be performed to further improve our understanding of yeast metabolism.

Model structure

Analysis of the sources of false predictions revealed five genes for which the false predictions are probably due to missing or extraneous functionalities in the model. POS5 (coding for a mitochondrial NADH/NADPH kinase) deletion resulted in a false positive prediction, because the model can produce NADPH in mitochondria using other mechanisms, whereas it has been recently shown experimentally that Pos5p is the major source of mitochondrial NADPH [77]. The false positive predictions for ERG2, ERG3, and ERG6 are due to a bypass in the model in ergosterol metabolism that allows direct synthesis of ergosterol from zymosterol. Although this bypass has been suggested to exist in yeast [78], based on the current study it appears that this alternate route in yeast does not bypass Erg2p, Erg3p, and Erg6p. The mitochondrial pyrophosphatase PPA2 deletion is a false positive, because the model can utilize a cytoplasmic pyrophosphatase instead and transport phosphate and pyrophosphate between the two compartments. If this transport capacity was limited as it is likely to be in vivo the PPA2 deletion would result in a lower growth rate due to limitation in mitochondrial metabolism. All these false predictions suggest directly potential changes into the actual structure of the model and also possibly re-evaluating our understanding of the specific parts of yeast metabolism as in the case of the ergosterol biosynthetic pathway.

Gene-protein-reaction associations

The false predictions relating to gene-protein-reaction associations are primarily due to either potentially missing isozymes (false negatives) or the existence of a dominant isozyme whose activity can not be fully compensated for by the other isozymes (false positives). The Gal2p galactose transporter is an example of the latter class as it is known that other hexose transporters can also transport galactose [79], but based on the comparison between simulation results and experimental data it appears that these transporters are insufficient to maintain maximal galactose uptake. A typical example of the latter class is the false negative prediction for Bat2p transaminase, which was found to be due to the lack of valine transamination functionality of the Bat1p isozyme in the model. This function has not been experimentally proven [80], but based on the results presented here it appears likely that BAT1 gene product can catalyze valine transamination in addition to other transamination reactions. The false predictions that were due to gene-protein-reaction associations suggest modifications to the model that relate to how the gene-to-enzymatic function mapping occurs in vivo.

Regulatory mechanisms

The lack of incorporation of regulatory mechanisms in the model could only clearly explain false model predictions for two of the genes – CDC19 (pyruvate kinase) and ADH1 (alcohol dehydrogenase). Both of these genes have isozymes that are capable of catalyzing the same reaction, but are known to be down-regulated under the particular condition in which the false positive prediction was done. The lack of regulatory restraints in the current model could also partially explain the observed general pattern of higher false prediction rates for conditions with glucose as the main carbon source as one would expect that the model would otherwise be more accurate for glucose than for less well characterized carbon sources. Due to the extensive metabolic reprogramming in glucose-grown cells utilizing different glucose repression mechanisms regulation plays a more significant role on glucose containing media than on other media conditions. In future generations of constraint-based metabolic models transcriptional regulation will be at least qualitatively incorporated in the models [81] so that regulatory effects will be more accurately accounted for.

Dead ends in the model

For eight genes with false positive predictions the reaction catalyzed by the gene product leads to a dead end in the model whereas in vivo the product of the reaction clearly is necessary for cellular function. This result indicates that either the model is missing some metabolic functions or there are gaps in the literature in understanding specific metabolic subsystems. Many of the dead ends are in phospholipids metabolism where the corresponding genes participate in the biosynthesis of complex phospholipids that are not utilized within the model, but that are probably converted to essential membrane phospholipids. Not all the dead ends in the model result in false predictions so that the eight-gene subset provides direct suggestions for further experimental work necessary for understanding the role of the currently unutilized metabolites in yeast cellular function.

Accumulation of toxic intermediates

In few cases the primary reason for a false positive prediction by the model appears to be the accumulation of a toxic intermediate in vivo when a particular enzyme further down the pathway is removed. For example, although folate biosynthesis is not required in rich media, genes involved in the biosynthetic pathway

(FOL1/FOL2/FOL3/DFR1) are essential, which is most likely due to toxicity of dihydropteorate (DHP), a precursor in the pathway [82]. Similarly in vivo MET22 null mutant accumulates phosphoadenylyl sulfate (PAPS), which is cytotoxic [83]. The in silico model does not account for non-specific chemical toxicity effects since these are usually not directly related to the metabolic function, but it is also possible that the model allows balancing of an toxic intermediate even if this would not happen in vivo and hence fails to predict the deleterious phenotype.

Media composition

Uncertainties in the in silico media compositions used to mimic the experimental conditions were the primary source of 32 false predictions most of which were false negatives. There are two separate sources of errors that can be identified: 1) wrong media composition, and 2) incorrect numerical values of maximum uptake rates of key nutrients. The former category includes examples such as TPS1/2 (trehalose 6-phosphate synthase/phosphatase), which both are false positive predictions on rich media due to the fact that the in silico YP medium contains trehalose and hence these genes that are essential for trehalose biosynthesis are not needed. However, it has been shown that trehalose is indeed a major component of the yeast extract medium [84] so that the false positive prediction is probably due to the inability of the yeast to utilize the trehalose in the media in the experimental deletion studies. The latter category of errors is typically manifested as a function of either the major carbon source or oxygen uptake rate or both. For many genes involved in mitochondrial respiration either lowering or raising the oxygen uptake rate would result in better predictive power. However, the maximum oxygen uptake rates in a batch culture are hard to estimate as they depend both on the degree of aeration provided and on the growth-rate dependent limitations due to the Crabtree effect. Most of the false predictions unique to a specific experimental condition could be traced back to uncertainties in the in silico media composition or maximal uptake rates indicating that a more careful evaluation of these failure modes would require performing the in vivo deletion studies in well defined media conditions that can be reproduced more accurately in silico.

Biomass composition

As noted already in our earlier deletion study utilizing iFF708 [53] the biomass composition utilized in the model is a major source of false predictions as it determines which metabolites are considered to be essential for cellular function and in what relative quantities these metabolites have to be produced. The biomass composition is derived primarily from experimental data on the composition of yeast cells growing in the exponential phase and it only includes the major biomass components as measuring trace components is difficult [58]. Typical examples of false positive predictions by the model are all genes involved in heme and quinone biosynthesis as these cofactors are obviously essential for cellular function. However, while the model utilizes these and other cofactors, they are recycled in the reactions and unless there is a drain of cofactors to the biomass they do not need to be synthesized de novo. An example of false negative predictions that relate to the in silico biomass composition are certain genes in membrane lipid and steroid biosynthesis. While some of the lipids are essential they can often be utilized interchangeably by the cell so that any particular type of lipid or sterol may not be essential as long as sufficient overall amount of e.g. phospholipids is produced. Since the model biomass requires fixed amounts of certain types of phospholipids and steroids this leads to false negative predictions. The false predictions due to the biomass composition could be easily corrected by including trace amounts of essential cofactors in the biomass and allowing more flexible usage of phospholipids and steroids, but it would be difficult to estimate exactly the relative amounts of the metabolites required without further experimentation.

Other cellular processes

The single most common source of false predictions in this study was the involvement of metabolic genes in other cellular processes that are not accounted for in the current model. As mentioned earlier the model does not currently include mRNA and protein synthesis and thus all pathways resulting in the biosynthesis of various RNA species such as transfer RNAs are dead ends although these functions are clearly essential for cellular function. Since methods for incorporating protein synthesis into the constraint-based modeling framework have been developed [85, 86] in future versions of the model these currently missing functionalities can be accounted for. Another type of false positive prediction that arises from the involvement of metabolic genes in other cellular processes is the role of these genes in overall cellular maintenance. For example, false positive predictions were made for all vacuolar ATPase components as their disruption in vivo results in major problems in pH balancing and the current model does not yet implement full pH balancing between compartments. Similarly, although the model does correctly predict the phenotypes for deletions of ATP synthase subunits on non-fermentable carbon sources, on fermentable carbon sources the model does not require the mitochondrial ATP synthase although in vivo this functionality is required for general mitochondrial maintenance. As the constraint-based framework is extended to include other types of cellular processes besides metabolism and regulation it can be expected that many of the false predictions will be corrected and that the comparison between in silico and in vivo gene deletions will provide valuable assistance for the expanded model building.

Discrepancies in experimental data

There were 16 genes with false predictions, for which apparent discrepancies in experimental data were found. These included cases such as PRO3, which is listed as an essential gene in one study [74], but appears to be non-essential in the other study [73]. In addition to discrepancies between the two genome-wide deletion studies there were also genes whose phenotype in the large-scale studies disagreed with the reported phenotype in the literature (e.g. THR1 null mutant should only be a threonine auxotroph and should grow on rich media). False predictions for cases where apparent discrepancies in experimental data were found were not further analyzed as it was not clear which data set would be the most trustworthy.

Unknown sources of false predictions

There were 31 genes whose predicted false phenotypes could not be explained by any of the reasons listed above even after careful evaluation of both the model and experimental data. Many of the genes in this list are related to a few separate metabolic subsystems with false phenotypic predictions under specific media conditions indicating that there may be important unidentified biochemical mechanisms present in these systems. An especially interesting example is the high number of false predictions related to methionine and homocysteine biosynthesis, which have been extensively studied both in yeast and in higher eukaryotes because of the role of homocysteine in cardiovascular and neurodegenerative diseases [87, 88].

The key gene in this system is MET6, which codes for the methionine synthetase responsible for converting homocysteine into methionine. This deletion has no phenotype on rich media in vivo, but the model predicted the deletion to be lethal due to inability to balance homocysteine in absence of the methionine synthetase reaction. However, the model currently accounts for all the biochemical transformations with homocysteine either as a reactant or product that are known to be present in yeast indicating that there is may still some unknown mechanisms by which the homocysteine balancing is accomplished in vivo. The genes with false predictions with no clearly identifiable reason for the false result provide clues to areas where further experimental work is clearly needed in order to improve our understanding of eukaryotic metabolism.

Model Changes Suggested by Gene Deletion Study

Detailed analysis of model failures resulted in 27 direct suggestions for improving the current model either by changing its reaction structure or the gene-protein-reaction associations (Table 3.3). Some of these suggestions are straightforward, such as making a component of a complex non-essential for the enzymatic function, whereas others, such as restricting phosphate transport across the mitochondrial membrane, would be somewhat more challenging to implement. For all of the 27 cases, the model represents the current knowledge on metabolic biochemistry, genetics, and physiology as well as possible, and the changes primarily relate to the interpretation of the available information. These suggestions demonstrate how model-driven evaluation of experimental gene deletion phenotypes can be used to systematically fine tune a model and improve our understanding of the particular biological system.

3.4 Conclusions

We have shown that multi-compartmental *in silico* metabolic models of eukaryotic cells with elementally and charge-balanced reactions can be successfully built. In addition, these models can be used to compute growth phenotypes of organisms with altered genotypes in various media conditions. The growth phenotypes computed with the compartmentalized eukaryotic model were found to be consistent with 83% of the *in vivo* results. Detailed case-by-case analysis of the false predictions led to the identification of gaps or inconsistencies in our knowledge base that require either changes in the model structure or further experimental investigation. This high correct prediction rate demonstrates the growing predictive power of constraint-based metabolic models even under variable environmental conditions and the overall importance of network topology in determining phenotypic consequences of genotypic changes.

The text of this chapter, in part or in full, is a reprint of the material as it appears in N.C. Duarte, M.J. Herrgard, and B.O. Palsson. 2004. Characterization and validation of *Saccharomyces cerevisiae* iND750: a fully compartmentalized genome-scale metabolic model. *Genome Res* **14**:1298-309. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

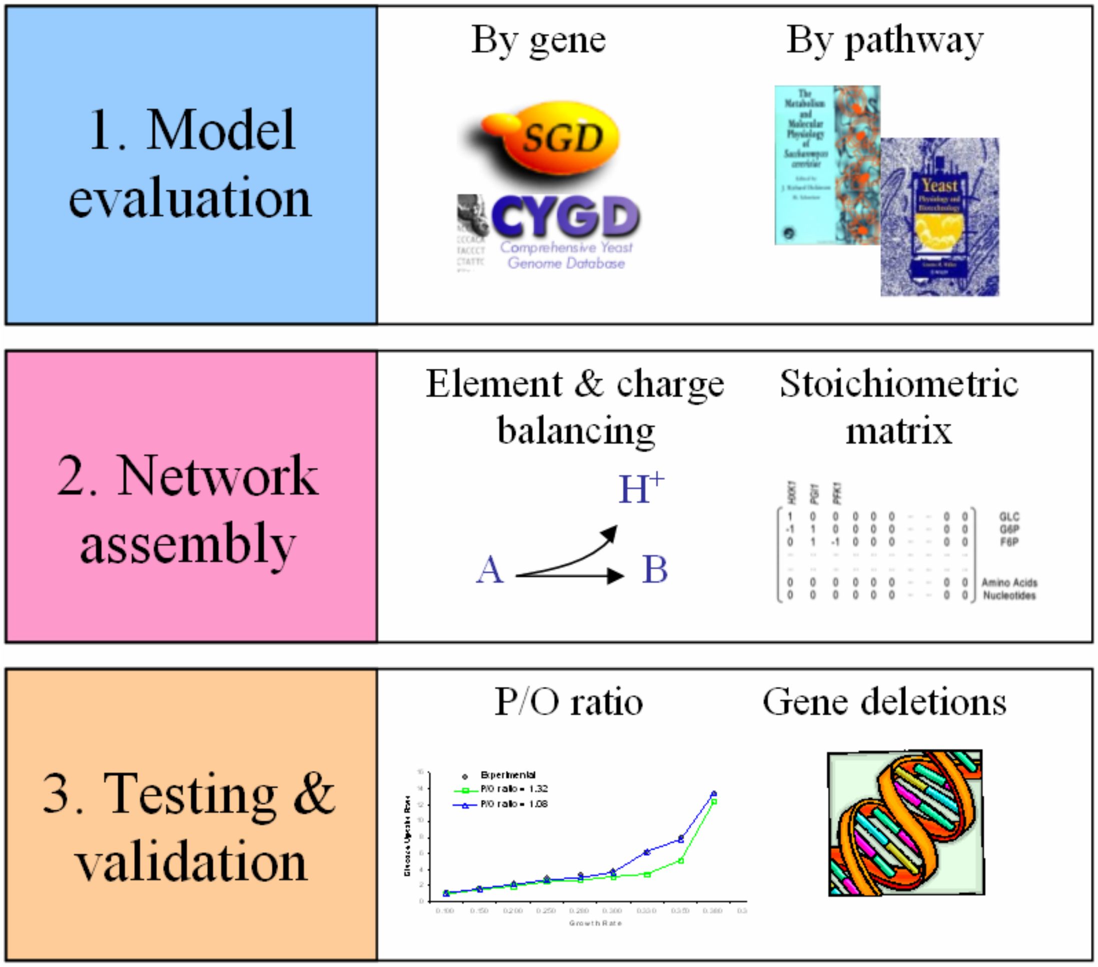


Figure 3.1: Overall reconstruction strategy for *S. cerevisiae* iND750. *S. cerevisiae* iFF708 [58] was used a starting point for building an updated and expanded yeast metabolic network. Step 1) Information on localization, database identifiers, and gene-protein relationships were collected from genome annotation databases. New genes and reactions were also identified from the literature. Step 2) Elementally and charge-balanced reactions were then formulated based on metabolite structures at pH 7.2 and used to form a stoichiometric matrix. Step 3) The model's predictions of physiological parameters, such as the P/O ratio, and gene deletion phenotypes were used to verify network content. The result is *S. cerevisiae* iND750, the first fully compartmentalized eukaryotic network. The contents of iND750 can be found in Supplement A and online at http://systemsbiology.ucsd.edu.

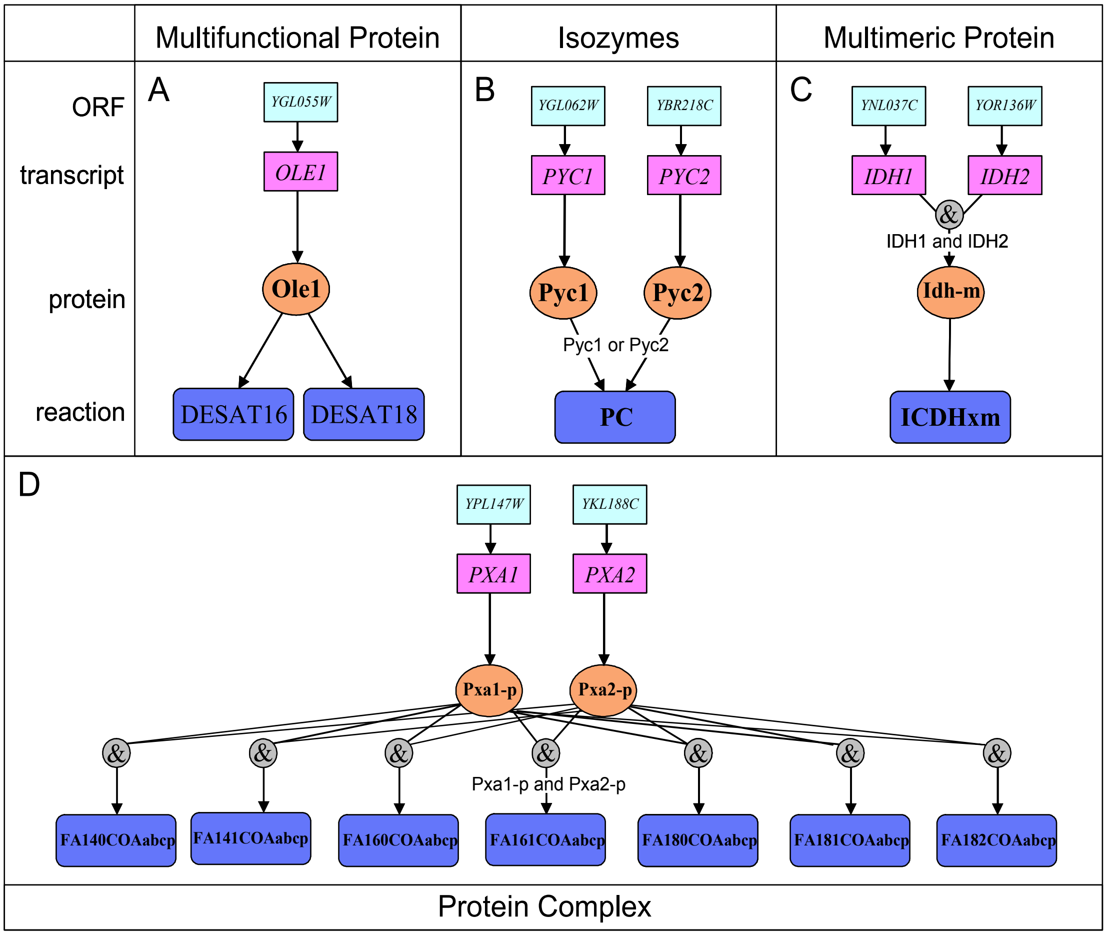


Figure 3.2: Gene-protein-reaction associations in *S. cerevisiae* iND750. Gene-protein-reaction associations represent the detailed logical relationships between open reading frames (ORFs), transcripts, proteins, and reactions in the model. *(A)* A multifunctional protein, such as Ole1, can catalyze more than one reaction. *(B)* Pyc1 and Pyc2 are examples of isozymes, or proteins that can catalyze the same reaction independently. *(C)* Idh-m is an example of a multimeric protein; it is formed by the association of two transcripts. *(D)* Proteins Pxa1-p and Pxa2-p form a protein complex. Both proteins are required to catalyze the reactions. All the gene-protein-reaction associations in *S. cerevisiae* iND750 are available in Supplement A and online at http://systemsbiology.ucsd.edu.

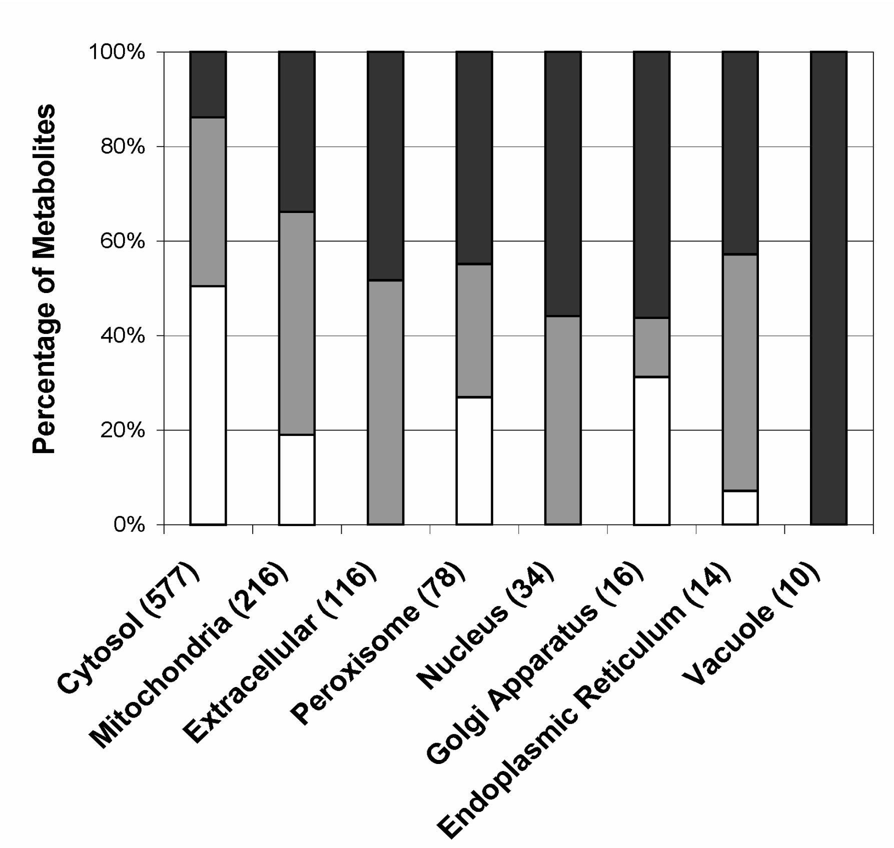


Figure 3.3: Compartmental distribution of *S. cerevisiae* iND750's metabolites. The number of metabolites found in each compartment is shaded based on its connectivity. Metabolites that are unique to a particular compartment are shown in white; metabolites found in two compartments are shaded in grey; and metabolites found in three or more compartments are shaded in black.

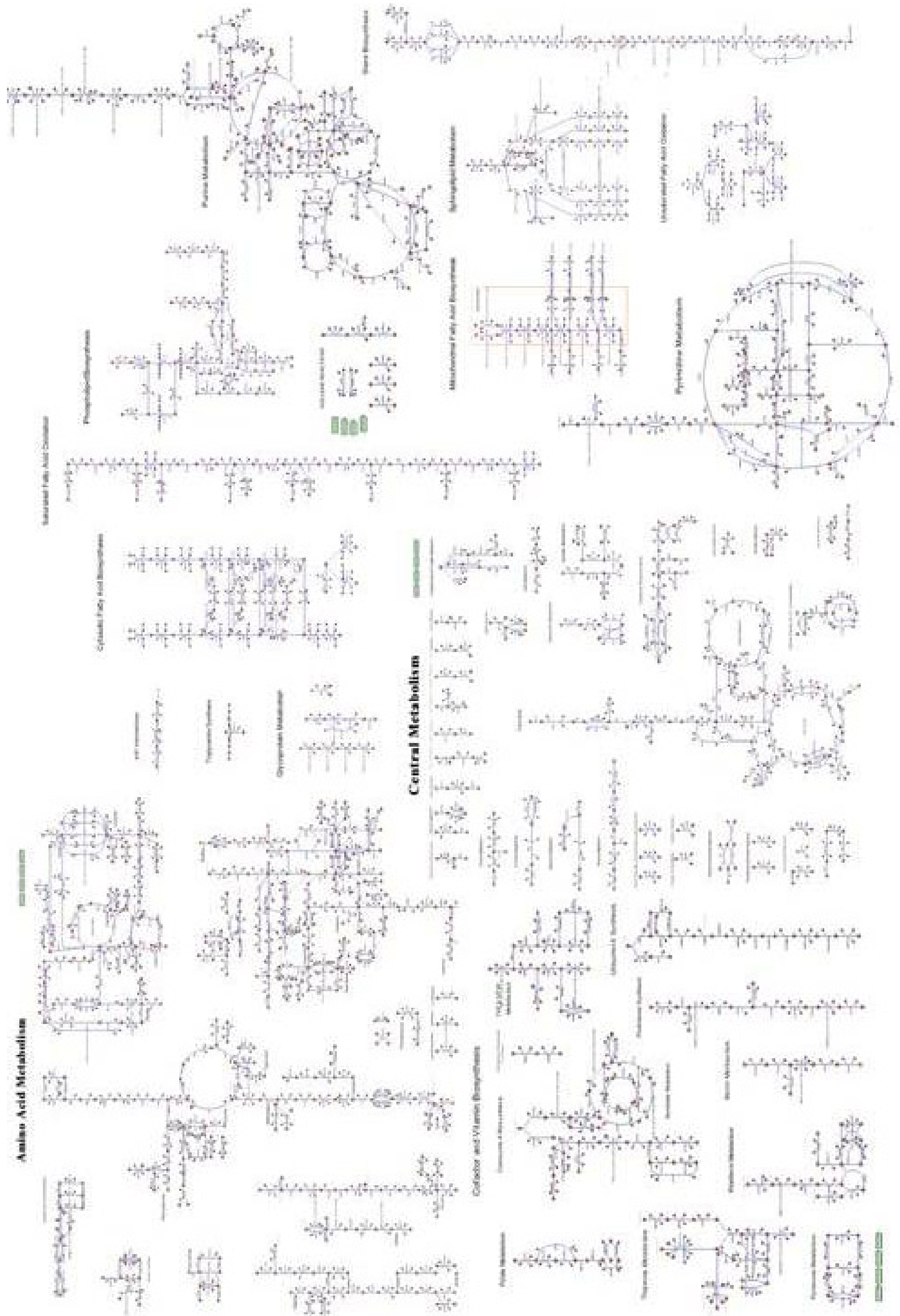

Amino Acid Metabolism
Central Metabolism
Purine Metabolism
Sterol Biosynthesis
Sphingolipid Metabolism
Unsaturated Fatty Acid Oxidation
Mitochondrial Fatty Acid Biosynthesis
Pyrimidine Metabolism
Phospholipid Biosynthesis
Saturated Fatty Acid Oxidation
Cytosolic Fatty Acid Biosynthesis
Cofactor and Vitamin Biosynthesis

Figure 3.4: The complete *S. cerevisiae* iND750 metabolic map. All of iND750's 646 metabolites (nodes) and 1179 reactions (connections) can be visualized on a collection of yeast-specific metabolic maps available in Supplement A and at http://systemsbiology.ucsd.edu.

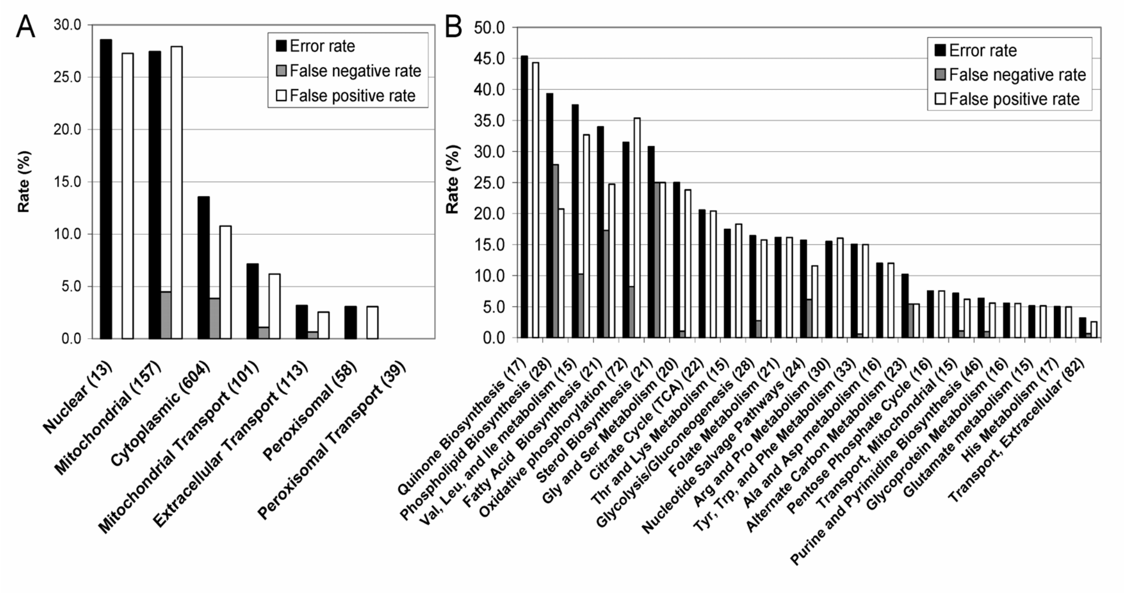


Figure 3.5: False predictions by compartment (A) and metabolic pathway (B). The overall error rate is the percentage of false predictions out of all of the predictions. The false-negative (FN) rate is the percentage of FN predictions out of all predictions in which the experimental data show normal growth. The false-positive (FP) rate is the percentage of FP predictions out of all predictions in which the experimental data show retarded growth. Genes that participate in transport functions between compartments are classified according to Table 3.2. Compartments with at least 10 genes and metabolic subsystems with at least 15 genes are included.

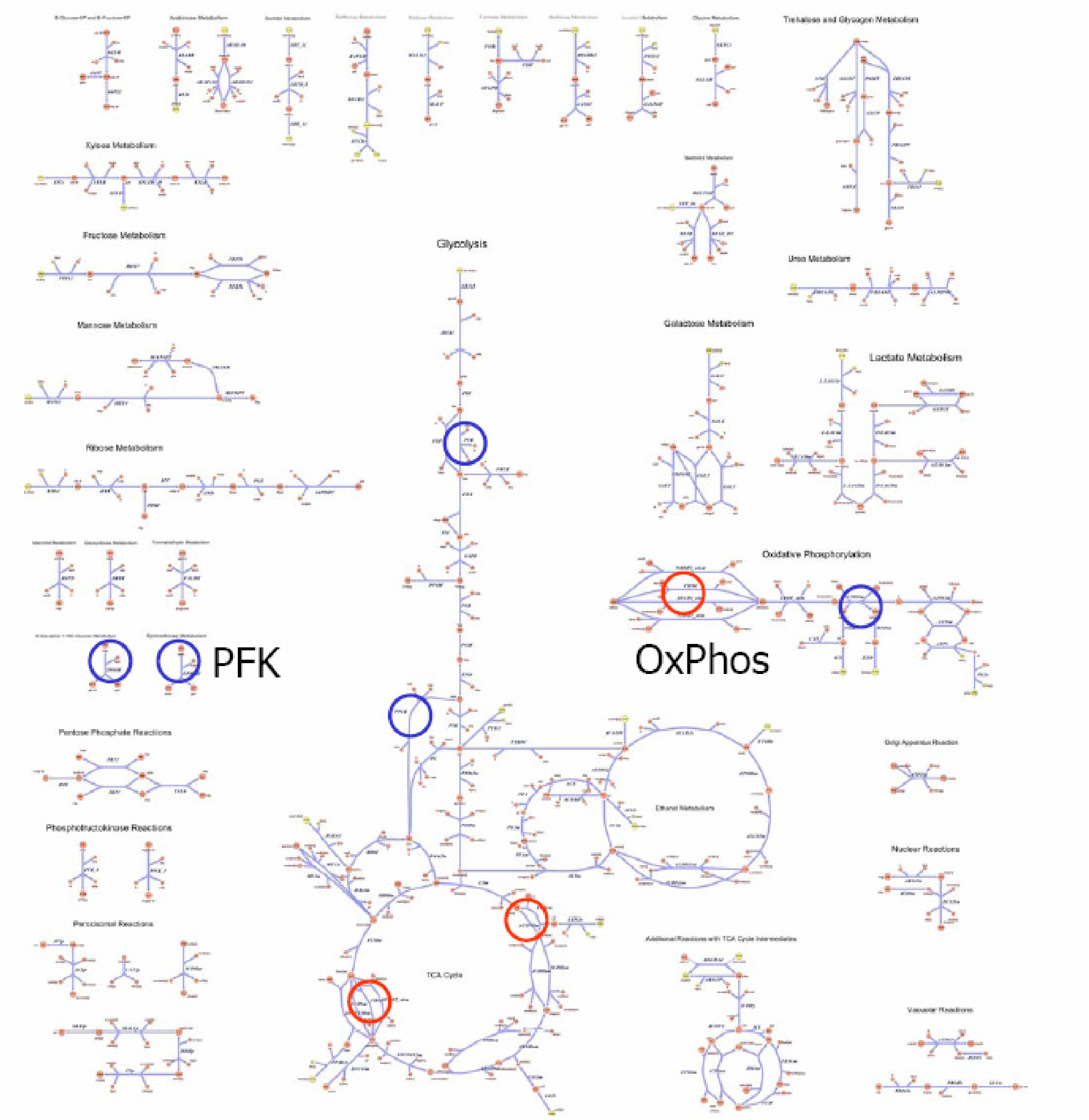


Figure 3.6: Mispredicted gene deletion phenotypes in central metabolism. A high false prediction rate was found in genes related to oxidative phosphorylation (OxPhos, 31.4%). Most of the erroneous predictions in glycolysis were false negatives (blue circles) whereas those in the tricarboxylic acid (TCA) cycle were false positives (red circles).

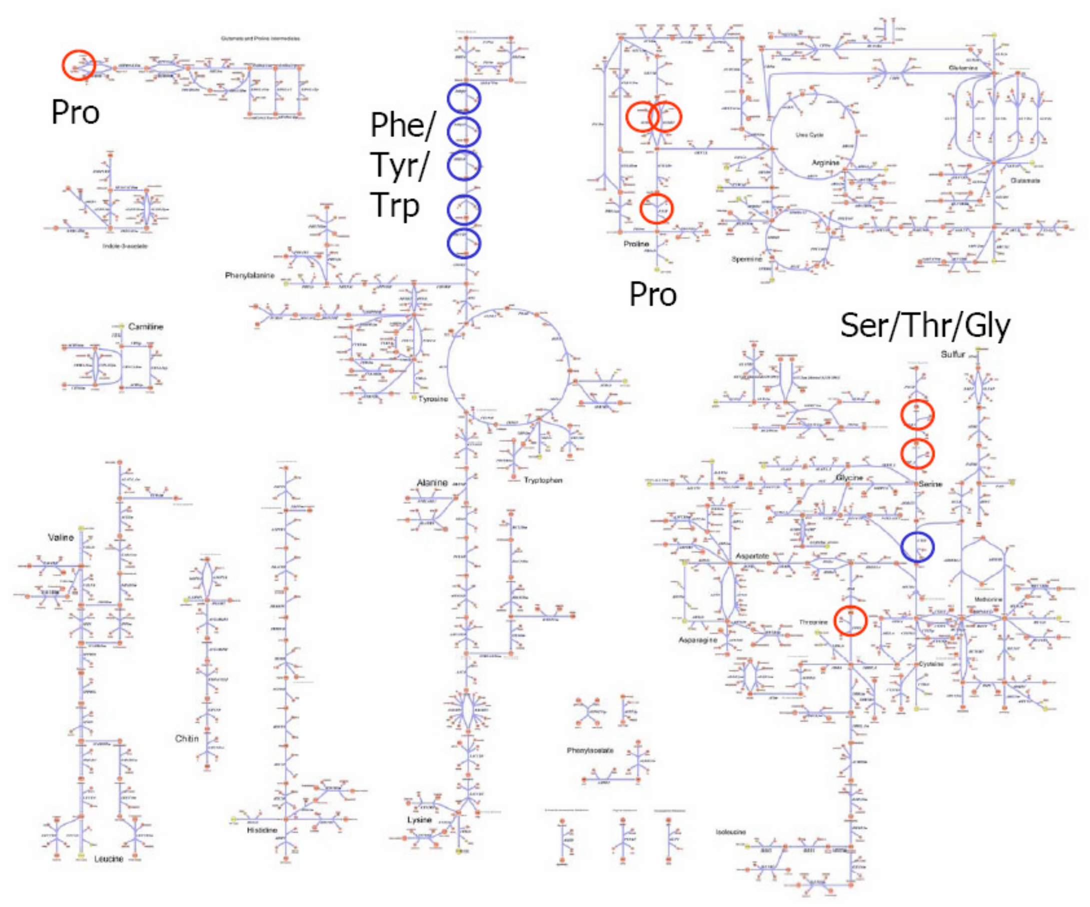


Figure 3.7: Mispredicted gene deletion phenotypes in amino acid metabolism. High false prediction rates were obtained for genes in branched chain amino acid biosynthesis (37.5%). Several false negative (blue circles) were associated with the initial steps in aromatic amino acid biosynthesis whereas genes related to proline metabolism were generally false positive predictions (red circles).

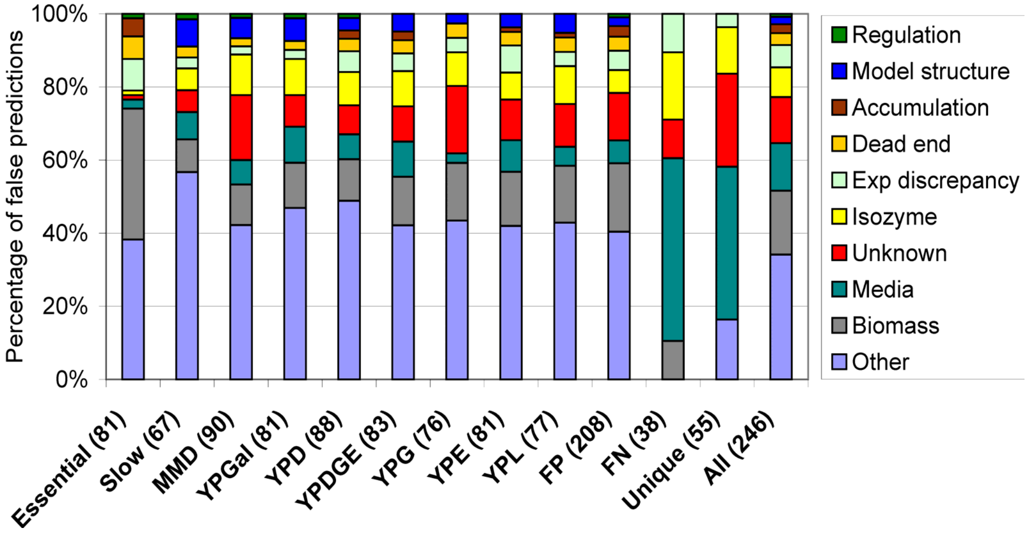


Figure 3.8: Breakdown of the false predictions by source. The reasons for false predictions are: transcriptional regulation (regulation), model structure, accumulation of toxic intermediate in vivo (accumulation), dead ends in the model (dead end), discrepancy in the experimental data (exp discrepancy), gene–protein-reaction associations (isozyme), unknown, *in silico* media composition (media), in silico biomass composition (biomass), and other cellular processes not included in the model (other). Results are shown for each experimental condition, including essential genes (essential) and slow growth genes (slow) on rich media. In addition, the distributions of the sources of false predictions are shown for false-positive (FP), false-negative (FN), and unique false predictions (unique) separately.

Table 3.1: Comparison of *S. cerevisiae* iFF708 and iND750. [a] The total number of metabolites irrespective of their compartmental locations. [b] The number of unique reactions (isozymes are not counted as separate reactions).[c] Reactions that differ in protons and water molecules are considered to be conserved.

Network component	iFF708	iND750	% Conserved
Genes	708	750	94
Metabolites[a]	584	646	90
Unique Reactions[b]	842	1149	56[c]

Table 3.2: Comparison of transport reactions in *S. cerevisiae* iND750. The transport mechanisms have been classified as diffusion (exchange of only a primary metabolite), symport (a primary and secondary metabolite transported in the same direction), antiport (a primary and secondary metabolite transported in opposite directions), or other (ABC transporters and ADP/ATP exchange reactions). For each membrane/mechanism combination, the number of gene-associated reactions is shown in parentheses next to the total number of reactions in that category.

Transport category	# of reactions	Transport mechanism (# gene associated)			
		Diffusion	Symport	Antiport	Other
Extracellular	113	36 (9)	74 (46)	3 (1)	0
Mitochondrial	101	65 (0)	21 (2)	14 (13)	1 (1)
Peroxisomal	39	19 (2)	6 (0)	5 (0)	9 (9)
Nuclear	23	18 (0)	5 (0)	0	0
Endoplasmic Reticular	10	9 (0)	1 (0)	0	0
Vacuolar	7	5 (0)	2 (0)	0	0
Golgi Apparatus	4	3 (0)	0	1 (0)	0

Table 3.3: Model changes suggested by the gene deletion study. Abbreviations: ORF – open reading frame, Gene – gene abbreviation.

ORF	Gene	Reason for false prediction	Suggested changes and comments
YPL188W	*POS5*	Model structure	Change the model so that only Pos5p can provide NADPH in mitochondria.
YMR267W	*PPA2*	Model structure	Force the model to utilize Ppa2p instead of the cytoplasmic isoforms by restricting phosphate transport out of the mitochondria.
YMR202W	*ERG2*	Model structure	Modify the interconversion between zymosterol and ergosterol biosynthesis to require *ERG2*.
YLR056W	*ERG3*	Model structure	See *ERG2*.
YML008C	*ERG6*	Model structure	See *ERG2*.
YDR178W	*SDH4*	Isozyme	Make Sdh4p a non-essential part of the succinate dehydrogenase complex.
YML123C	*PHO84*	Isozyme	There are multiple alternative isozymes for the phosphate transporters, but Pho84p should be the dominant one.
YBR069C	*TAT1*	Isozyme	There are multiple alternative isozymes for amino acid transporters in the model, but they need to be made less efficient than Tat1p.
YLR081W	*GAL2*	Isozyme	Model includes other isozymes (HXT genes) that are not nearly as efficient for gal transport so disabling their gal transport ability should result correct prediction.
YMR105C	*PGM2*	Isozyme	Pgm2p is major isoform of phosphoglucomutase. Do not allow the minor isoform (Pgm1p) to fully compensate for loss of Pgm2p.
YHR137W	*ARO9*	Isozyme	Aro8p should be able to compensate for *ARO9* deletion on minimal media – modify the gene-protein-reaction association to reflect this.
YGL125W	*MET13*	Isozyme	Met13p is the dominant isozyme. Do not allow isozyme (Met12p) to compensate fully for the loss of Met13p.
YHR046C	*INM1*	Isozyme	Add the gene product of *YDR287W* as an isozyme for Inm1p.
YHR001WA	*QCR10*	Isozyme	This subunit should be made a non-essential part of the cytochrome bc1 complex since it only plays structural role.
YFR033C	*QCR6*	Isozyme	Deletion of *QCR*6 does not have significant effect on the formation or stability of cytochrome bc complex so that it should not play an essential role in complex formation.
YKL067W	*YNK1*	Isozyme	Null mutant retains 10% of nucleoside diphosphate kinase activity. Sources of remaining enzyme activity are unknown. Reaction without gene associations should be added to the model to represent these unidentified enzymes.
YLR304C	*ACO1*	Isozyme	The isozyme coded by *YJL200C* should not be able to fully compensate for *ACO1* deletion.
YNL052W	*COX5A*	Isozyme	Cox5Ap is the dominant isoform – Cox5Bp should not be able to fully compensate.

Table 3.3, continued.

ORF	Gene	Reason for false prediction	Suggested changes and comments
YKL148C	*SDH1*	Isozyme	Sdh1p should not be considered to be an essential part of the succinate dehydrogenase complex.
YGL008C	*PMA1*	Isozyme	This is the major isoform of the cytosolic ATPase, but in the model a minor isoform (which contains Pma2p instead of Pma1p) can compensate for the function. Do not allow the minor isoform to fully compensate for the loss of the major isoform.
YLR342W	*FKS1*	Isozyme	There are three alternate isozymes in the model, but Fks1p should be made the dominant isozyme.
YHR183W	*GND1*	Isozyme	This is the major isozyme (80% of activity) – other isozymes should be made less efficient.
YLR044C	*PDC1*	Isozyme	There are three alternate isozymes in the model, but *PDC1* deletion alone is sufficient to reduce pyruvate decarboxylase activity significantly enough to result in a slow growth phenotype. Should have Pdc1p as the major isozyme.
YJR148W	*BAT2*	Isozyme	*BAT2* single deletion should not be lethal as there is a mitochondrial isozyme (Bat1p) - double deletion should be lethal. Bat1p currently does not catalyze valine transamination so this functionality should be added.
YCL009C	*ILV6*	Isozyme	Ilv6p is the regulatory subunit of phenylalanine transaminase. This subunit should be made non-essential for the enzymatic function.
YAL038W	*CDC19*	Gene Regulation	Pyk2p isozyme should only be expressed under conditions of very low glycolytic flux.
YOL086C	*ADH1*	Gene Regulation	This isozyme (out of five) should be the only one active under severely glucose repressed conditions.

CHAPTER 4: INTEGRATED ANALYSIS OF *SACCHAROMYCES CEREVISIAE* METABOLIC PHENOTYPES

Genome-scale metabolic networks of microorganisms, namely *Escherichia coli*, *Haemophilus influenzae*, and *Helicobacter pylori*, have led to useful insights into substrate preferences, the effects of gene deletions, optimal growth patterns, outcomes of adaptive evolution, and shifts in expression profiles [32]. With the recent reconstruction of *S. cerevisiae*'s genome-scale metabolic network [57, 58], these analytical techniques can now be applied to the first genome-scale model of a eukaryotic cell. In this study, we examine the function and capacity of yeast's metabolic machinery and show that its phenotypic phase plane (2.2.2) can be used to accurately predict metabolic phenotypes and to interpret experimental data in the context of a genome-scale model.

4.1 *In silico* characterization of metabolic phenotypes

In the following sections, we formulate a glucose-oxygen phenotypic phase plane for yeast (4.1.1) based on its recent genome-scale metabolic reconstruction [58] and calculate respiratory quotients and secretion profiles for a range of oxygenation conditions (4.1.2). The growth states predicted by the PhPP are then characterized using shadow price analysis (4.1.3), *in silico* gene deletion simulations (4.1.3), and *in vivo* growth experiments (4.2).

4.1.1 *S. cerevisiae* phenotypic phase plane

The *S. cerevisiae* genome-scale metabolic network constructed by Forster *et al.* [58] was used to generate a phenotypic phase plane (PhPP; 2.2.2) [89] that describes yeast's metabolic states at various levels of glucose and oxygen availability (Figure 4.1). Points on the surface of the three-dimensional PhPP correspond to maximal growth rates allowable for each pair of glucose and oxygen uptake rates in the *x*-*y* plane (Figure 4.1A). All of the points on or below this three-dimensional surface represent feasible metabolic growth behaviors.

The two-dimensional projection of the PhPP (Figure 4.1B) has been divided into seven regions, or "phases," to allow for qualitative comparisons (P_1 – P_7). Each phase represents a metabolic phenotype with specific pathway utilization. These pathway utilization patterns are defined by shadow price analysis (section 1.4.2), which uses parameters generated by the linear programming solution (shadow prices) to identify how changes in metabolite levels affect biomass formation [89]. Shadow prices are constant within a phase and change continuously at the boundary from one phase to the next.

Two regions of the PhPP have infeasible steady-state flux distributions: the area along the *y*-axis and the small square near the origin. Growth is infeasible in the region between the ordinate and P1 since yeast cannot use more than six oxygen molecules per glucose molecule. The two red lines in Figure 4.1B are lines of optimality (LO). LO_{growth} represents optimal aerobic glucose-limited growth of *S. cerevisiae* in which substrates are completely oxidized to produce biomass and is comparable to the sole line of optimality that has been identified in *E. coli* [72]. $LO_{ethanol}$ corresponds to maximum ethanol production under microaerobic conditions while growth is maximized. In lieu of glycerol production, NADH is reoxidized via maximal ethanol formation.

The phenomenon described by $LO_{ethanol}$ is a distinguishing feature of the yeast PhPP, and is supported by many research reports in the literature [90-92]. For example, Cysewski and Wilke [90] found a sharp stimulation of the specific ethanol productivity at a very low but non-zero level of dissolved oxygen. Later studies showed that a value of 10 ppb of dissolved oxygen maximized ethanol production in yeast chemostat cultures [92]. Thus, $LO_{ethanol}$, the second line of optimality predicted by the genome-scale model, is consistent with the experimental observations.

4.1.2 Simulation of optimal metabolic phenotypes

Flux balance analysis (2.2.1) was used to illustrate how the optimal metabolic phenotypes change across the seven phases of the yeast phase plane (Figure 4.1B). For the simulations, the glucose uptake rate was arbitrarily set to 5 mmol/gDCW/hr and the oxygen uptake rate (OUR) was varied from 0 to 20 mmol/gDCW/hr. This allowed us to study the influence of a single environmental variable on cellular metabolism. Small amounts of NH_3, sulfate and phosphate were introduced for the biomass synthesis. During anaerobic conditions (OUR = 0, on the *x*-axis), the growth rate was low and the respiratory quotient (CO_2 evolution rate / OUR) was infinite by definition (Figure 4.2A). As the oxygen uptake rate increased to 13 mmol/gDCW/hr to reach LO_{growth}, the growth rate increased to its maximum value and the respiratory quotient approached 1.06. Further increasing the oxygen uptake rate caused both the growth rate and respiratory quotient to decrease due to futile cycles in which a combination of two or more biochemical reactions resulted only in the hydrolysis of ATP or other high-energy compounds [89].

Metabolic by-product secretion profiles (2.2.2) were also calculated with increasing oxygen uptake rates. Since alternative optimal solutions exist in the genome-scale metabolic flux models [42], a range of secretion rates can be found amongst all of the equivalent optimal solutions for a fixed point in the PhPP. Remarkably, there was less than 1% difference between the maximum and minimum allowable secretion rates for a fixed maximal growth rate; thus, only the maximum predicted secretion fluxes for ethanol, succinate, glycerol, and acetate are shown (Figure 4.2B). During anaerobic fermentation, ethanol, glycerol, and succinate were produced. Maximum ethanol production occurred at an oxygen uptake rate of 0.5 mmol/gDCW/hr, a condition defining $LO_{ethanol}$. Glycerol production ceased at this point. With a slight increase in oxygen uptake rate above $LO_{ethanol}$, acetate began to be secreted but succinate secretion decreased to zero. Ethanol and acetate were no longer secreted once the oxygen uptake rate was equal to or greater than 13 mmol/gDCW/hr, a point on LO_{growth} where the metabolic pathway utilization enables complete aerobic growth.

The results of this analysis suggest that yeast has only a few primary phenotypes, designated by the various phases. In P_1, the oxygen supply is sufficient for growth by aerobic respiration, resulting in carbon dioxide as the sole by-product. Phases P_2-P_6 correspond to states of oxidative-fermentative growth, which is characterized by secretion of oxidative and fermentative metabolic by-products, *i.e.*, acetate and ethanol, respectively. Finally, P_7 represents microaerobic conditions. In this environment, yeast grows primarily by fermentation and secretes ethanol, glycerol, and succinate. This limited range of metabolic states is strikingly different from that found for *E. coli*, whose glucose-oxygen PhPP has five distinct optimal *in silico* phenotypes [72].

4.1.3 Further characterization of oxidative-fermentative phases

The secretion profile (Figure 4.2B) does not show any phenotypic differences between phases $P_2 – P_6$. These states are highly similar since the phases are essentially co-planar in the 3-dimensional PhPP (Figure 4.1A). However, through the use of shadow price analysis and *in silico* gene deletions, distinct pathway utilization patterns could be found for each phase.

Shadow price analysis (2.2.2) evaluates how small changes in metabolite production affect optimal growth rates [89]. A positive shadow price indicated that a metabolite was available in excess, meaning that a decrease in its availability would increase biomass synthesis, and a negative shadow price indicated that a metabolite was limiting such that increasing its availability would increase the biomass synthesis.

In silico gene deletions (2.2.3) were also performed in order to determine which reactions were essential in each phase. This approach was especially useful for interpreting the physiological differences between growth states in phases 2 – 6 since their phenotypes were indistinguishable in terms of their secretion profiles.

Phase 2

In phase 2, the ratio of oxygen uptake rate and glucose uptake rate is lower than that on the line of optimality. As a result, the cell is oxygen limited and begins to ferment. Mitochondrial NAD+ is available in excess, meaning that the biomass synthesis would improve if its availability decreased. In order to maintain the cell's redox balance, the excess mitochondrial NAD+ must be reduced. This is done through the production of acetate and ethanol, which begin to be secreted in this phase. Thus it is the production of acetate and ethanol that makes the optimal growth rate less than that defined on the line of optimality.

Phase 3

As the ratio of oxygen and glucose uptake rates is further decreased, three lower glycolysis reactions (fructose bis-phosphate aldolase, triose phosphate dehydrogenase, and phosphoglycerate kinase) become essential for growth in phase 3. These reactions are also essential in subsequent phases as oxygen uptake rate is further decreased. Due to the limited oxygen, more carbons “overflows” into the fermentation pathway while at the same time oxidative metabolism becomes less effective.

Phase 4

Shifting from phase 3 to phase 4, the pentose phosphate pathway is utilized to generate NADPH because not enough NADPH is produced through respiration at the lower oxygen uptake rate. The NADPH is then converted to NADH which is subsequently used for ATP production.

Phase 5

Further lowering the ratio of oxygen and glucose uptake rates restricts the cell’s ability to produce pyruvate in phase 5. Yeast can no longer utilize the oxidative pathways because an insufficient amount of cytosolic NAD+ is produced. When comparing phases 4 and 5, all of the metabolites with shadow price sign changes were folate intermediates. These are important energy carriers that are directly linked to the availability of both cytosolic and mitochondrial NAD+ and NADP+.

Phase 6

As you enter phase 6, acetate production completely ceases. Ethanol is secreted as the only metabolic by-product to balance the redox potential of the cell.

4.2 Integration of experimental data and *in silico* predictions

A useful application of the *S. cerevisiae* PhPP is to qualitatively classify yeast's metabolic state based on phenotypic observations made *in vivo*. Three groups of experiments were conducted under different growth conditions in the PhPP: aerobic, glucose-limited growth (Figure 4.3B), oxidative-fermentative batch growth (Figure 4.3C), and microaerobic batch growth (Figure 4.3D) (see [93] for methods). These data were then projected on the *S. cerevisiae* PhPP (Figure 4.3A) using the experimentally measured OUR and glucose uptake rates. The metabolite concentration profiles obtained from these experiments were found to qualitatively agree with the corresponding metabolic states predicted by the PhPP. For example, in growth conditions near $LO_{ethanol}$, cells are expected to grow almost entirely by fermentation, with significant production of ethanol and lesser amounts of glycerol, acetate and succinate secretion. This phenotype is qualitatively similar to experimental observation, in which more ethanol is produced than acetate as shown in Figure 4.3D.

Points representative of each growth state were then used as constraints in computer simulations to quantitatively predict yeast's metabolic phenotype (Table 4.1). Overall, the experimental observations and the *in silico* predictions are in good agreement. However, the predicted growth rates are slightly higher than the measured values. This difference may result from the model's prediction of optimal performance while growth *in vivo* is actually suboptimal.

4.3 Conclusions

In this study, the *S. cerevisiae* genome-scale metabolic network was used to formulate a phenotypic phase plane that displays the maximum allowable growth rate and distinct patterns of metabolic pathway utilization for all combinations of glucose and oxygen uptake rates. *In silico* predictions of growth rate and secretion rates and *in vivo* data for three separate growth conditions (aerobic glucose-limited, oxidative-fermentative, and microaerobic) were concordant. Thus, constraint-based methods such as phase plane analysis can be used to explore *in silico* the metabolic capabilities of microorganisms, generate new hypotheses as to how these organisms operate, and highlight the impact of individual cellular components on the organism as a whole.

The text of this chapter, in part or in full, is a reprint of the material as it appears in N.C. Duarte, B.O. Palsson, and P. Fu. 2004. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. *BMC Genomics*, **5**:63. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

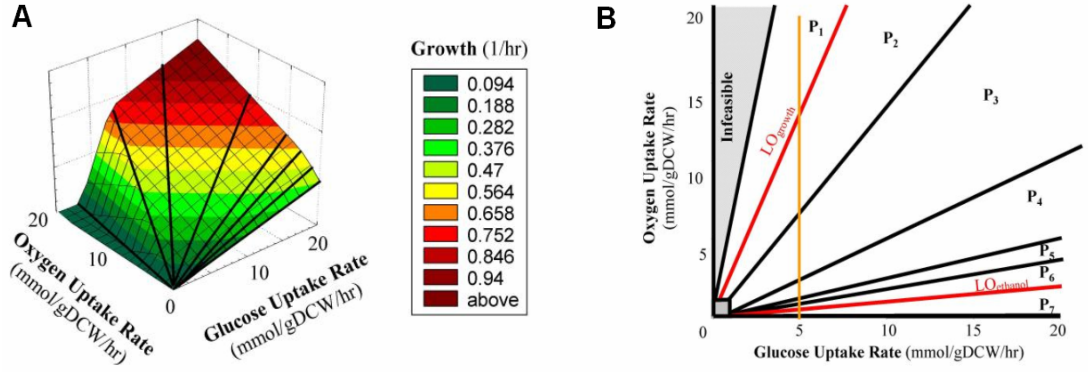


Figure 4.1: The yeast glucose-oxygen phenotypic phase plane (PhPP). *(A)* The three-dimensional *S. cerevisiae* PhPP drawn with Statistica™ (Statsoft, Tulsa, OK). The *x* and *y* axes represent the glucose uptake rate and oxygen uptake rate, respectively. The third dimension is the cellular growth rate. *(B)* A two-dimensional projection of the 3-D polytope in panel A. The two lines of optimality are shown in red. LO$_{growth}$ represents optimal aerobic glucose-limited growth and LO$_{ethanol}$ corresponds to maximum ethanol production under microaerobic conditions. P$_1$ - P$_7$ represent phases with various metabolic phenotypes. The shaded regions correspond to infeasible growth conditions. The orange line (glucose uptake flux = 5 mmol/gDCW/hr) represents the conditions which were used for the simulations in Figure 4.2.

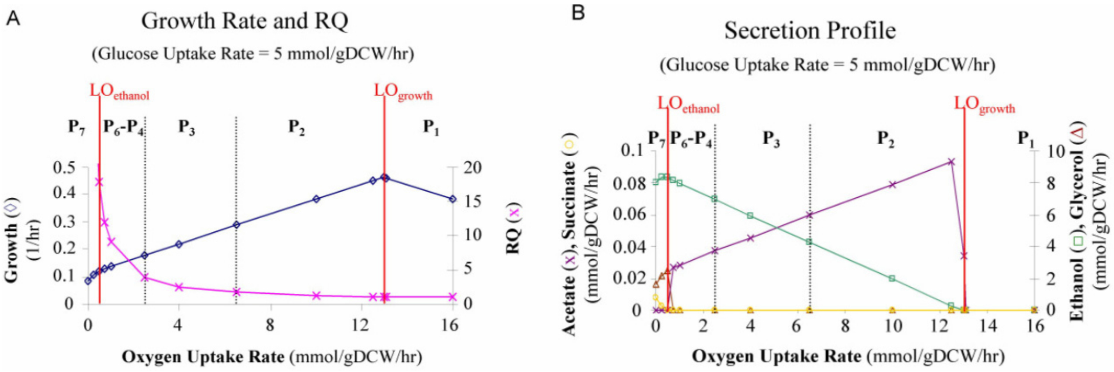


Figure 4.2: Yeast's optimal growth behaviors as a function of oxygen availability. Simulations were performed in conditions ranging from completely anaerobic fermentation to completely aerobic growth. The range of oxygen uptake rates used in the simulations (orange line, Figure 4.1B) allows for the characterization of the PhPP's seven phases (P_1 - P_7) and two lines of optimality (LO_{growth}, $LO_{ethanol}$). *(A)* Growth rate and respiratory quotient (RQ). *(B)* Secretion profile for acetate, succinate, ethanol, and glycerol.

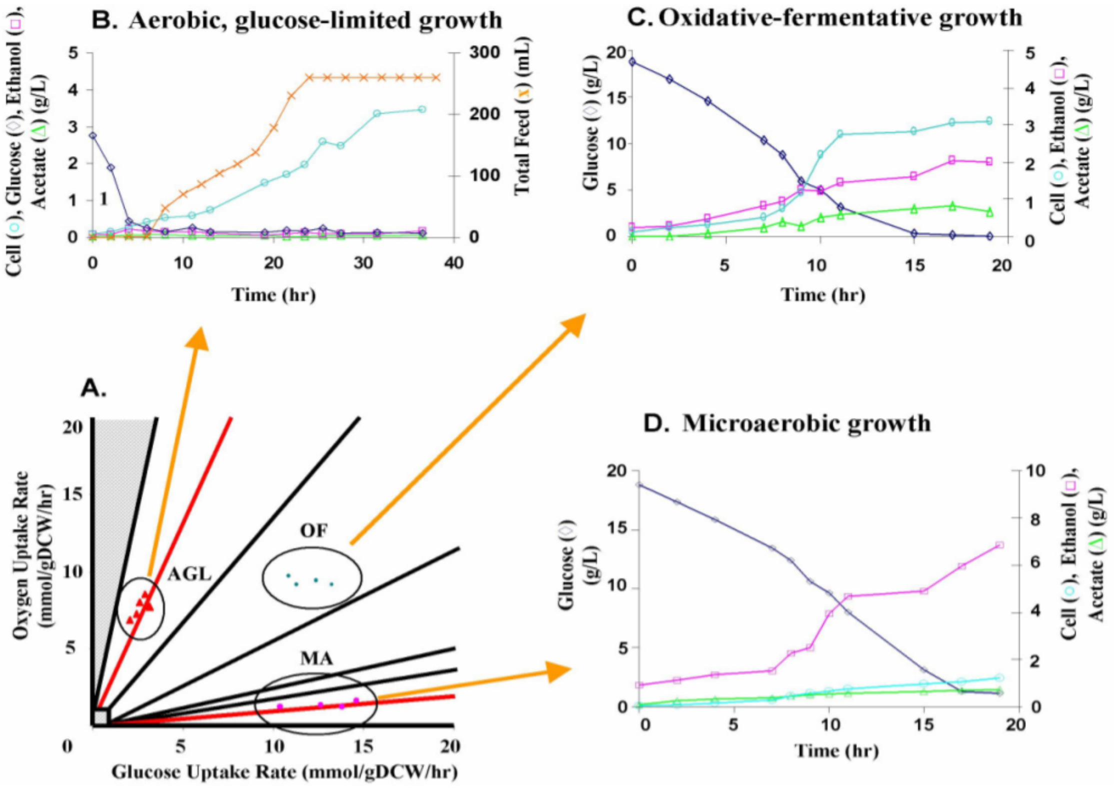


Figure 4.3: Growth experiments shown on the phenotypic phase plane. *(A)* The three groups of experimental data displayed on the *S. cerevisiae* PhPP were used as an index for the time course profiles in panels B, C and D. *(B)* Aerobic glucose-limited growth (AGL) controlled by fed-batch operation. *(C)* Oxidative-fermentative growth (OF) with unlimited glucose and oxygen availability. *(D)* Microaerobic growth (MA) with unlimited glucose and very low oxygen availability. The AGL and MA data sets are located on lines of optimality and as a result are stable metabolic states with only one degree of freedom (glucose for AGL and oxygen for MA). OF is an unstable metabolic state with two degrees of freedom (glucose and oxygen), making it more difficult to control this type of growth condition. By perturbing the environmental conditions, cells in OF can be shifted to either AGL or MA.

Table 4.1: Comparison of *in silico* and *in vitro* flux measurements. Abbreviations: OUR – oxygen uptake rate, GUR – glucose uptake rate. Units: growth rate (1/hr), substrate uptake rates and metabolite production rates (mmol/gDCW/hr).

	Microaerobic fermentation OUR = 1, GUR = 14		**Oxidative fermentation OUR = 9, GUR = 12**		**Aerobic growth OUR = 8, GUR = 2.5**	
	In silico	**Experimental**	***In silico***	**Experimental**	***In silico***	**Experimental**
Growth rate	0.33	0.31	0.53	0.51	0.22	0.20
Ethanol	21.29	20.08	11.98	11.07	0	0.16
Acetate	0.26	0.22	2.62	2.57	0	0.31

CHAPTER 5: GENOME-SCALE RECONSTRUCTION OF THE GLOBAL HUMAN METABOLIC NETWORK

An individual's metabolism is determined by one's genetics, environment, and nutrition. With the recent sequencing and annotation of the human genome [95-97], we can now identify the human body's complement of metabolic enzymes. An extensive knowledge base of human genetic, biochemical, and physiological data also exists in the literature from decades worth of scientific studies. The stage has thus been set for constructing the first genome-scale reconstructions of human cellular processes.

This chapter begins with a brief history of the Human Genome Project and previous models of mammalian metabolism (5.1). This is followed by an in-depth description of the reconstruction and validation procedures used to assemble the global human metabolic network (5.2), which we have termed *H. sapiens* Recon 1. Finally, we conclude with a discussion of how *H. sapiens* Recon 1 can be used a strategic tool for discovery research (5.3).

5.1 Previous models of mammalian metabolism

The course of this dissertation has witnessed tremendous growth in the analysis and modeling of mammalian systems. For example, the first mammalian genome was released in 2002 [94], and since that time more than ten others have been completed, including our own [95]. While several small-scale models of mammalian metabolism have appeared, the first genome-scale metabolic networks did not emerge until 2005. These efforts are described in greater detail here.

5.1.1 Small-scale, cell-specific metabolic models

The majority of small-scale, cell-specific models are of hepatocytes and myocytes; these cells have clearly defined roles in metabolism, and thus there is great interest in understanding their function in health, injury, and disease. Mass-balance models of hepatocyte metabolism have been successful in studying the effects of severe burn injuries [96], induced liver failure [97], and hypermetabolism [98] in rats as well as to simulate fibrosis [99] and von Gierke's and Hers' diseases [100] in humans. Martin Yarmush and his colleagues have also used their rat model to investigate why cultured hepatocytes grow so poorly in plasma [101, 102], which has been a major hurdle in the development of extracorporeal blood filtering devices and other biotechnological applications. The regulation of ATP-to-ADP concentrations [103] and consequences of moderate burn injury [104] have been examined with mass-balance models of human skeletal muscle cells. Dynamic [105] and full-kinetic models [106] of skeletal myocytes were also used to interrogate the relationship between fluxes through the glycogenolytic pathway and oxidative phosphorylation. Thermokinetic [107] and flux-balance [108] models of human cardiac mitochondria have also appeared. Modeling approaches that have been applied to other cell types include metabolic flux analysis of human embryonic kidney cells [109] and a full-kinetic model of human red blood cell metabolism [110]. These small, cell-specific models are useful for studying cellular behaviors under a defined set of conditions. However, to fully explore metabolic genotype-phenotype relationships, comprehensive reconstructions of the inherently complex metabolic networks that exist *in vivo* are needed. As described in 1.2, the availability of whole genome sequences is central to developing genome-scale metabolic models.

5.1.2 Mammalian genome sequencing and the first genome-scale model

The Human Genome Project was launched in 1990 with the goal of obtaining a high-quality euchromatic sequence of the human genome (Figure 5.1). The widely publicized race between the publicly-funded International Human Genome Sequencing Consortium (IHGSC) and privately-held Celera Genomics to complete an initial draft of the human genome sequence ended in February 2001 when the groups simultaneously published their efforts on the covers of *Nature* [96] and *Science* [111], respectively (Table 5.1). While the IHGSC pushed forward to complete a finished human genome sequence, the first high-quality draft of a mammalian genome sequence was announced for the laboratory mouse *Mus musculus* in December 2002 [94]. The mouse genome sequence was widely recognized as an important tool for comparative analysis of human genes, and has subsequently enabled many insights into gene assignments, regulation, and evolution [112].

A finished draft of the mouse genome was foundational to constructing the first genome-scale model of mammalian metabolism [113]. While the authors faced a conceptual leap in modeling a multicellular eukaryotic organism, much of their basic reconstruction approach followed directly from their previous work with *S. cerevisiae* ([58]; 3.1). Rather than reconstructing a particular cell type, the authors assembled a generalized network of mouse metabolism that accounted for 872 metabolites and 1,220 biochemical reactions globally found across all murine cells. The model's consistency and predictive ability were validated by demonstrating that its *in silico* predictions of growth, single gene deletion phenotypes, substrate requirements, and antibody production were concordant with experimentally measured values from a murine hybridoma cell line. This work represents a significant achievement for genome-scale, constraint-based modeling, as the authors were able to successfully apply reconstruction methods and modeling tools reserved for single-celled microorganisms to a multicellular, eukaryotic system.

5.1.3 Top-down reconstructions of human metabolism

The final draft of the euchromatic human genome sequence was completed in October 2004 (Build 35; [114]). Since its release, a handful of top-down reconstructions (1.2) of human metabolism have appeared. For example, the HumanCyc metabolic database (http://HumanCyc.org) was generated by mapping genes from Build 31 of the genome annotation (November 2002) to a universal, multi-organism pathway database using the PathoLogic algorithm [15]. Similarly, the Kyoto Encyclopedia of Genes and Genomes pathway database (KEGG; 2.1.2) contains a comprehensive collection of metabolic maps in which enzymes are selectively highlighted based on the genome annotation of an organism of interest [14]. As described in the next section, these top-down reconstructions can be a valuable starting point for building genome-scale models. However, additional data must be manually compiled from the scientific literature to ensure that the network components and their interactions are based on direct physical evidence and reflect the current knowledge of human metabolism.

5.2 Manual reconstruction of the global human metabolic network

The wealth of detailed biochemical information available for humans, combined with the recent sequencing and annotation of the human genome [95], enabled the first genome-scale, bottom-up reconstruction of the global human metabolic network. Its reconstruction and validation is described here; Chapter 6 discusses how the network can be used a context for interpreting large-scale biological data sets and as a basis for developing cell-specific reconstructions.

5.2.1 Overall reconstruction strategy

The goal of this project was to develop a genome-scale metabolic model that comprehensively represents all of the biochemical activities found in human cells. The reconstruction procedure and some of potential applications of the global metabolic network are outlined in Figure 5.2. Briefly, an initial component list was generated using a top-down approach (5.2.2). The network was then manually curated to validate the automated content, expand the network scope, fill in pathway gaps, and include additional dimensionality to the annotation (5.2.3). Several rounds of reconstruction and testing under strict quality control were required to obtain a BiGG, high quality network (5.2.4). The resultant network, *H. sapiens* Recon 1, is the first manually-curated, bottom-up reconstruction of human metabolism and the largest eukaryotic reconstruction to date (5.2.5). The entire contents of Recon 1, including reaction lists, metabolic maps, and the stoichiometric matrix, are freely available in a searchable database at http://bigg.ucsd.edu. Reconstruction of the global human metabolic network has enabled the identification of knowledge gaps (5.3) and provides a context for the analysis of genome-scale data, such as gene expression measurements (Chapter 6).

5.2.2 Initial component list

A well-annotated genome sequence is critical to bottom-up reconstruction since it enables the rapid identification of candidate network components [6] and the assembly of a preliminary network [15] that can be used as a starting point for manual curation (Figure 5.2, Supplement B). We used E.C. numbers [17] and GO terms [115] (2.1.2) to identify an initial set of 1,865 human metabolic genes from the November 2004 (Build 35) annotations of three publicly available genome annotation databases: KEGG [14], NCBI's LocusLink [116] (now Entrez Gene [9]), and the H-Invitational Database [13]. These genes were mapped to a rudimentary network of 3,623 metabolic enzymes from KEGG Orthology [117], and 3,673 reactions from KEGG's LIGAND database [14] and *S. cerevisiae* iND750, the compartmentalized yeast metabolic reconstruction (see Chapter 3) [57].

5.2.3 Curation of network content

In addition to establishing initial network scope, LIGAND's pathway-based organizational structure also facilitated parallel network assembly. The initial component list was divided into eight metabolic subsets that were simultaneously curated by a team of researchers (Table 5.1). Curation entailed the verification of computationally predicted gene assignments, formulation of enzymatic reactions, and identification of protein compartmentalization and gene/protein interactions based on biological evidence found in literature articles, textbooks, and internet resources (see 2.1).

Verification of gene assignments

Putative gene assignments were verified based on evidence collected from genome annotation databases, namely Entrez Gene [9] and Gene Cards [12], and the scientific literature. Genes were not included in the network unless there was compelling evidence for their function. At a minimum, this evidence included sequence based annotations, such as computationally predicted functional domains or high sequence similarity to a well characterized mammalian homolog.

Formulation of metabolites and reactions

The KEGG and iND750 reactions mapped in the automated procedure were thoroughly revised during the manual curation process. Substrate and cofactor preferences were identified using BRENDA [16] and published literature. Metabolite formulae and charge were calculated based on their ionization state at pH 7.2, which for simplicity was presumed to be constant across all compartments. Extensive literature surveys were required to identify and formulate compounds with variable, organism-specific compositions (*e.g.,* dolichol, phospholipids, quinones). Metabolic reactions were formulated based on known stoichiometry and were subjected to mass- and charge-balance constraints. Reaction directionality was determined from thermodynamic data or inferred from legacy data and textbooks.

Protein compartmentalization

Protein compartmentalization was entirely determined by manual curation, as this information was not captured by the KEGG Ontology. The reactions in Recon 1 were compartmentalized in the cytosol, mitochondrion, nucleus, endoplasmic reticulum, Golgi apparatus, lysosome, peroxisome, and/or extracellular space based on protein localization data, sequence targeting signals, and indirect physiological evidence. If this data was unavailable, reactions were modeled as cytosolic. For double-membrane organelles such as the nucleus and mitochondrion, the composition of the intermembrane space was considered to be equivalent to the cytosol. Consequently, proteins identified in the outer membrane and intermembrane space were modeled as cytosolic.

Gene-transcript-protein-reaction relationships

Gene-transcript-protein-reaction relationships [6, 113] were manually identified from the literature and formulated as Boolean logic statements (see Figure 3.2). Isozymes (an "or" relationship) were defined as distinct proteins that catalyze the same substrate- and compartment-specific reaction and could arise from one gene due to alternative splicing or be encoded by independent genes. All isozymes were modeled as having the same reaction rate regardless of known substrate preferences. Cases in which a reaction was dependent on the presence of more than one gene/protein (an "and" relationship, *e.g.,* proteins with multiple subunits/chains or complexes composed of multiple enzymes) were classified as protein complexes. These associations are typically composed of both enzymatic and structural or regulatory components. More complicated gene and protein interactions also exist; for example, the mitochondrial ATP synthase includes an F1 catalytic core with 5 subunits, an F0 protein channel with 8 subunits, and several assembly proteins. Most of the genes are alternatively spliced and some subunits can be encoded by a number of genes [118].

Confidence scores

Confidence scores were assigned based on biological evidence associated with each reaction. Evidence from classical biochemical or genetic experiments, such as gene cloning and protein characterization, were given the highest confidence score (3). Mid-level scores (2) were assigned to reactions based on physiological data or biochemical/ genetic evidence from non-human mammalian cell (typically mouse, rat, or rabbit). Reactions with the lowest confidence score (1) were included solely based on *in silico* modeling because, during the process of model validation, they were deemed mandatory for a particular metabolic function.

Network assembly in SimPheny

The contents of the reconstruction were assembled in the SimPheny™ software package (Genomatica, San Diego). A new gene index with more than 16,000 loci was compiled based on Build 35 of the human genome annotation (Figure 5.3). Entrez Gene IDs [9] were used as unique identifiers for genes. Alternative transcripts were defined based on known RefSeq mRNA IDs [116] associated with each locus and were designated with a numeric suffix after each locus id. For example, the two transcripts associated with locus 55902 are recorded as "55902.1" and "55902.2" in the gene index. Each gene was automatically associated with at least one transcript, even if its RefSeq ID was unknown. Most of the loci (82%) were associated with only one transcript and approximately 13% with two transcripts. Only 244 loci had five or more transcripts, and these were mainly genes involved in tumorgenesis, cell cycle, and signaling.

An example of how data was extracted from online databases and entered into the SimPheny™ software package is provided in Figure 5.4. The contents of the network were fixed after each round of reconstruction and the stoichiometric matrix was formulated as previously described [119].

5.2.4 Functional validation and gap analysis

Bottom-up reconstructions can be represented mathematically, enabling the computational interrogation of their properties [6]. Since the global network accounts for all known metabolic enzymes encoded in the human genome, it collectively represents metabolic functions found in a variety of cell types. We validated the basic functionality of the human metabolic network by using flux balance analysis (2.2.1) [120] to simulate over 280 metabolic functions, ranging from simple pathway-level objectives to parameter calculations (Table 5.3).

Comprehensive gap analysis of the stoichiometric matrix was performed after each round of functional validation. Each "dead-end" metabolite that could not be produced or consumed was manually re-examined by returning to the literature to identify possible reactions describing its degradation, production, or transport. A description of the 474 unresolved gaps from the final round of validation is provided in Supplement B. Gaps were classified as either knowledge base (*i.e.*, 'dead end' metabolites that require further experimental investigation to be resolved) or model-scope gaps (*i.e.,* compounds whose metabolism is outside the scope of our reconstruction). Examples of each are provided in Table 5.4.

5.2.5 The result, *H. sapiens* Recon 1

Like genome sequencing, network reconstruction is an iterative process, requiring several rounds of iterative gap analysis and comprehensive revalidation under strict quality control to achieve an accurate, high quality network [4]. It took a team of seven researchers nearly 18 months to assemble the final global human metabolic network (Figure 5.6), which we have named *H. sapiens* Recon 1. Recon 1 is the largest functionally-validated reconstruction of a multi-cellular eukaryote to date (Table 5.5) and is the first human cellular process to be comprehensively modeled at this level of detail. It was almost entirely constructed from human-specific data, containing over 50 years of biochemical evidence collected from more than 1500 primary literature articles, reviews, and biochemistry textbooks. Many of its reactions were directly extracted from the literature and are not described in any chart or database. A complete list of genes, metabolites, reactions, citations, and curator comments are available in Supplement B.

Metabolic maps

Recon 1 also includes a comprehensive collection of high-quality, human specific maps (Supplement B) that are useful for navigating its and interpreting cell-scale data, such as gene expression measurements (Chapter 6) and functionally coupled reaction sets [121]. Glycan structures have been included on the maps to improve visualization of the glycosylation pathways (Figure 5.7).

BiGG human website

The Biochemically, Genetically and Genomically integrated (BiGG) Human Database (http://bigg.ucsd.edu) was developed as a resource for disseminating the contents of *H. sapiens* Recon 1. The website allows users to perform customizable searches on a static, internal database using convenient drop-down menus (Figure 5.8, Table 5.6) and contains links to public database entries associated with Recon 1's genes, proteins, reactions, and metabolites (Table 5.7). Future plans for the expanding the website are discussed in Chapter 7.

Recon 1 summarizes our current knowledge of the human metabolic network in a structured, mathematical format that enables systematic studies of human metabolism and its properties. Two of its potential research applications include global assessment of network confidence (5.3) and interpretation of large-scale data sets in the context of a biochemically accurate, genetically and genomically integrated database (Chapter 6).

5.3 Knowledge landscapes

Bottom-up reconstruction of Recon 1 required extensive, manual surveys of the primary literature to evaluate biological evidence associated with each gene, protein, and reaction. Viewing our confidence (5.2.3) in these individual components at the system level reveals a global knowledge landscape with specific "peaks" and "valleys" in our understanding of human metabolism (Figure 5.9). Three categories of metabolic pathways were identified based on the degree of characterization of their corresponding reactions.

Category I Pathways

Category I pathways represent the "peaks" of our knowledge landscape, with roughly 80% or more of the reactions having the highest confidence score. For example, genes have been identified for nearly all of the steps in chondroitin sulfate catabolism (Figure 5.10) except for its initial proteolysis and final degradation step. This pathway also contains a "dead-end" metabolite (5.2.3), as degradation of the glycosaminoglycan attachment site that is produced in the initial reaction (CBPASEly) is outside the scope of the current reconstruction.

Category II Pathways

Category II pathways, such as glyoxylate metabolism (Figure 5.11), have a roughly equal proportion of highly characterized enzymes and those with moderate biological evidence. For instance, while the peroxisomal and mitochondrial degradation of glyoxylate to L-glycine (AGTix, AGTim, respectively) has been extensively studied, the presence of glycerate kinase (GLYCK2) was inferred based on the observation that individuals with D-glycericaciduria (who lack this enzyme) cannot further metabolize D-glycerate and excrete gram amounts of it in their urine [122].

Category III Pathways

Category III pathways exhibit a wide range of confidence scores and gene coverage. The fact that some of these pathways have not been completely elucidated is surprising, and arguably these knowledge deficits may not have been identified without a systems approach. For example, the mechanism which cycles the end products of vitamin C degradation back to the glycolytic pathway appear to be poorly understood (Figure 5.12) despite evidence in human erythrocytes that it may be used as an energy source [123]. A large number of intracellular transport reactions are also included in this category, indicating that as a whole they require considerably more investigation to elucidate precise mechanistic reactions. Thus, the reconstruction of *H. sapiens* Recon 1 has resulted in a comprehensive review of our knowledge of human metabolism and has lead to direct suggestions where further experimental studies are needed (see Supplement B).

5.4 Conclusions

H. sapiens Recon 1 is the largest eukaryotic reconstruction to date and the first manually assembled, functionally validated model of human metabolism. Recon 1 was constructed based on the recent human genome annotation (Build 35) and over 50 years of genetic, biochemical, and physiological data that was extracted from the scientific literature. This work represents a significant milestone in human systems biology, and we foresee three primary applications of Recon 1 that will clearly be of interest to biomedical community. First, when subjected to genetic and chemical constraints, Recon 1 can be used as a mathematical model for computational interrogation of healthy and pathophysiological states. Second, Recon 1 is a discovery tool, consistently describing known aspects of human metabolism and defining knowledge gaps which require future experimental investigation. Third, Recon 1 is a comprehensive, high-confidence network that provides a context for mapping biological content such as genomic, transcriptomic, and proteomic data. This final application is the subject of Chapter 6.

The text of this chapter, in part or in full, is a reprint of the material as it appears in N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.O. Palsson. 2006. Global reconstruction of human metabolic network based on genomic and bibliomic data. Submitted to *Proc Natl Acad Sci U.S.A*. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

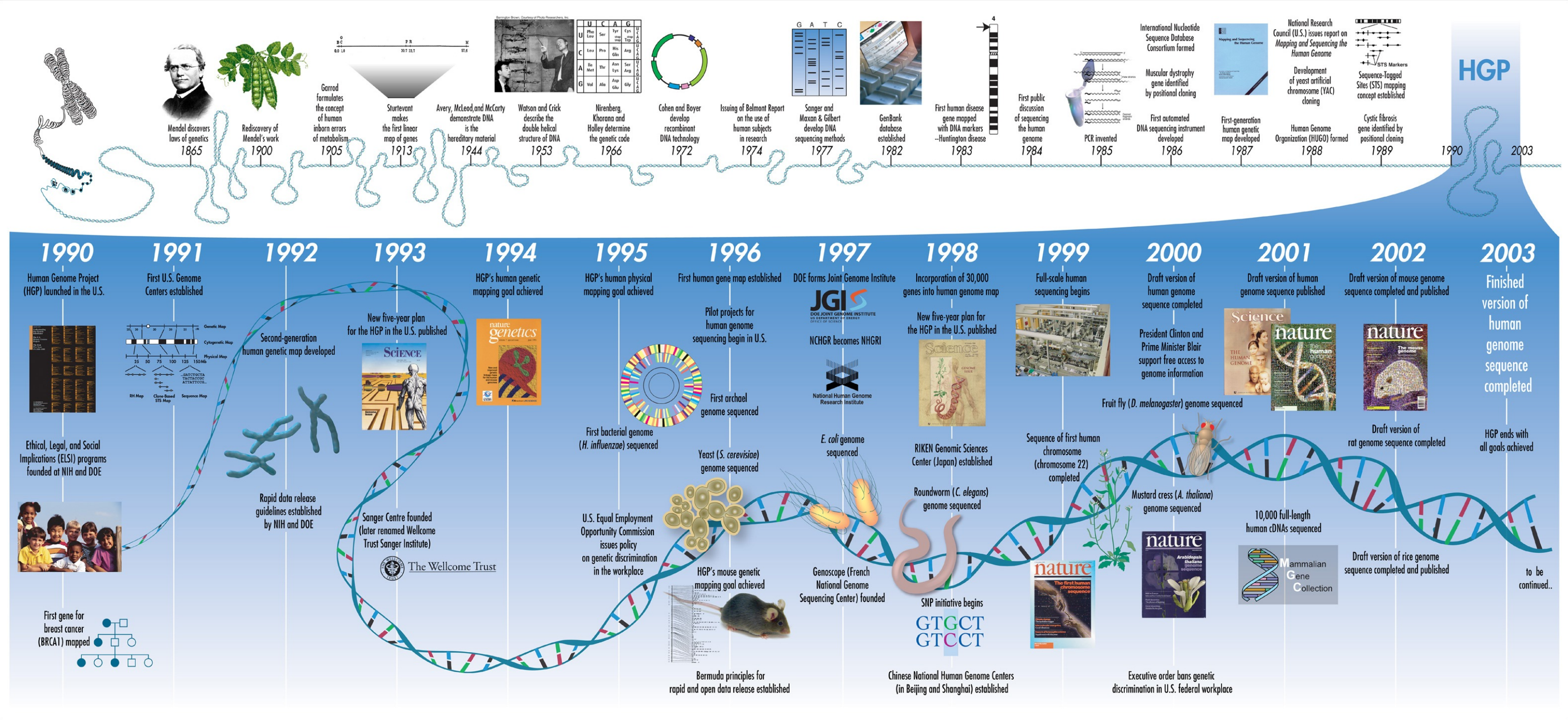

Mendel discovers laws of genetics 1865
Rediscovery of Mendel's work 1900
Garrod formulates the concept of human inborn errors of metabolism 1905
Sturtevant makes the first linear map of genes 1913
Avery, McLeod,and McCarty demonstrate DNA is the hereditary material 1944
Watson and Crick describe the double helical structure of DNA 1953
Nirenberg, Khorana and Holley determine the genetic code 1966
Cohen and Boyer develop recombinant DNA technology 1972
Issuing of Belmont Report on the use of human subjects in research 1974
Sanger and Maxan & Gilbert develop DNA sequencing methods 1977
GenBank database established 1982
First human disease gene mapped with DNA markers --Huntington disease 1983
First public discussion of sequencing the human genome 1984
PCR invented 1985
International Nucleotide Sequence Database Consortium formed
Muscular dystrophy gene identified by positional cloning
First automated DNA sequencing instrument developed 1986
First-generation human genetic map developed 1987
National Research Council (U.S.) issues report on Mapping and Sequencing the Human Genome
Development of yeast artificial chromosome (YAC) cloning
Human Genome Organization (HUGO) formed 1988
STS Markers
Sequence-Tagged Sites (STS) mapping concept established
Cystic fibrosis gene identified by positional cloning 1989
HGP
1990
2003
1990
Human Genome Project (HGP) launched in the U.S.
Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE
First gene for breast cancer (BRCA1) mapped
1991
First U.S. Genome Centers established
1992
Second-generation human genetic map developed
Rapid data release guidelines established by NIH and DOE
1993
New five-year plan for the HGP in the U.S. published
Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)
The Wellcome Trust
1994
HGP's human genetic mapping goal achieved
1995
HGP's human physical mapping goal achieved
First bacterial genome (H. influenzae) sequenced
U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace
1996
First human gene map established
Pilot projects for human genome sequencing begin in U.S.
First archael genome sequenced
Yeast (S. cerevisiae) genome sequenced
HGP's mouse genetic mapping goal achieved
Bermuda principles for rapid and open data release established
1997
DOE forms Joint Genome Institute
NCHGR becomes NHGRI
E. coli genome sequenced
Genoscope (French National Genome Sequencing Center) founded
1998
Incorporation of 30,000 genes into human genome map
New five-year plan for the HGP in the U.S. published
RIKEN Genomic Sciences Center (Japan) established
Roundworm (C. elegans) genome sequenced
SNP initiative begins
GTGCT
GTCCT
Chinese National Human Genome Centers (in Beijing and Shanghai) established
1999
Full-scale human sequencing begins
Sequence of first human chromosome (chromosome 22) completed
2000
Draft version of human genome sequence completed
President Clinton and Prime Minister Blair support free access to genome information
Fruit fly (D. melanogaster) genome sequenced
Mustard cress (A. thaliana) genome sequenced
Executive order bans genetic discrimination in U.S. federal workplace
2001
Draft version of human genome sequence published
10,000 full-length human cDNAs sequenced
Mammalian Gene Collection
2002
Draft version of mouse genome sequence completed and published
Draft version of rat genome sequence completed
Draft version of rice genome sequence completed and published
2003
Finished version of human genome sequence completed
HGP ends with all goals achieved
to be continued...

Figure 5.1: Timeline of the human genome project. Reproduced with permission from [124].

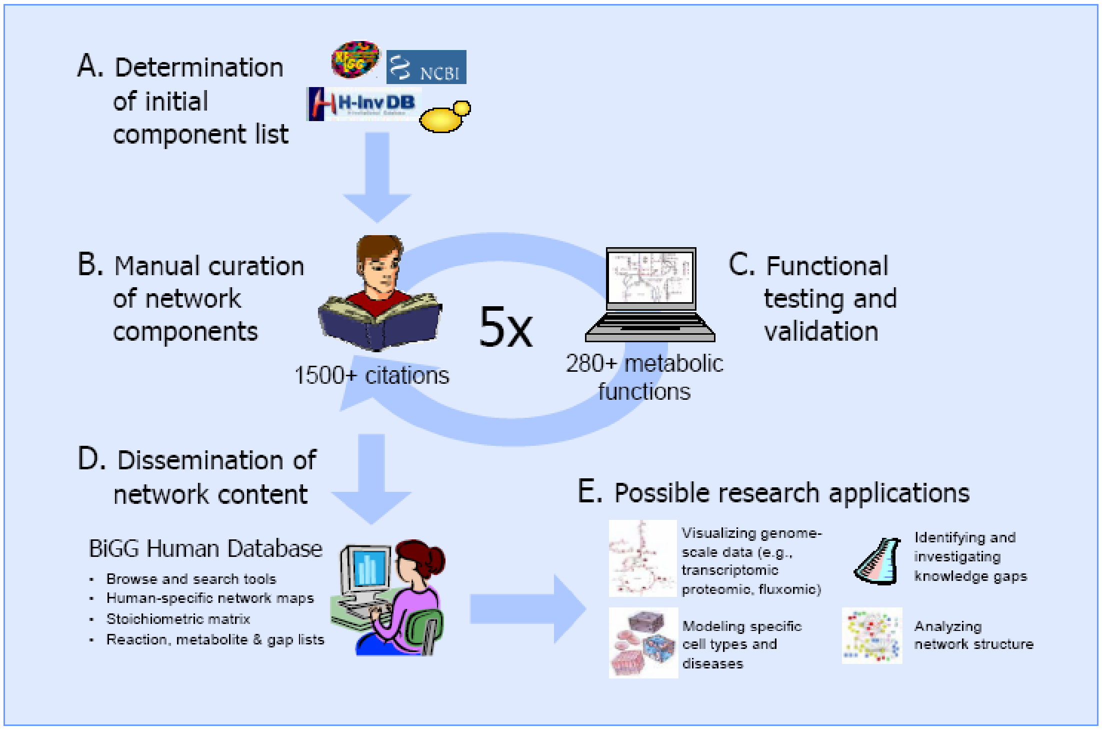
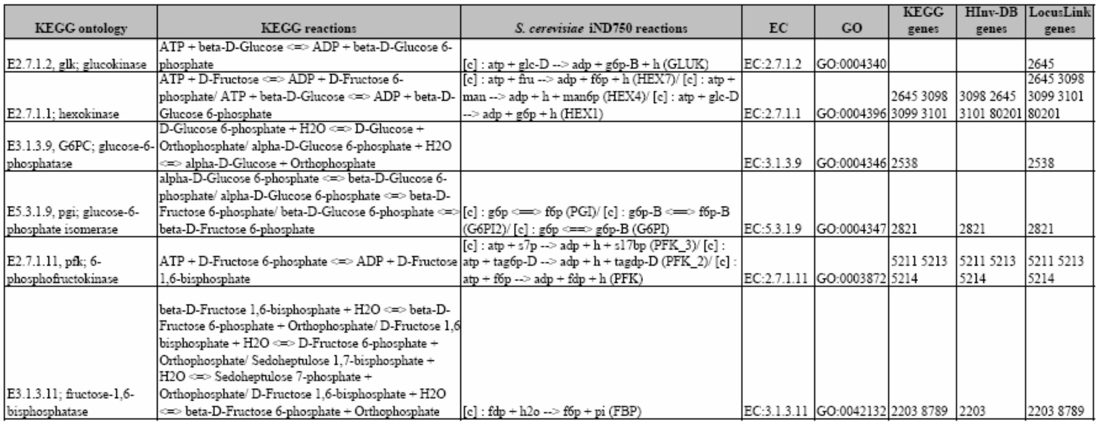
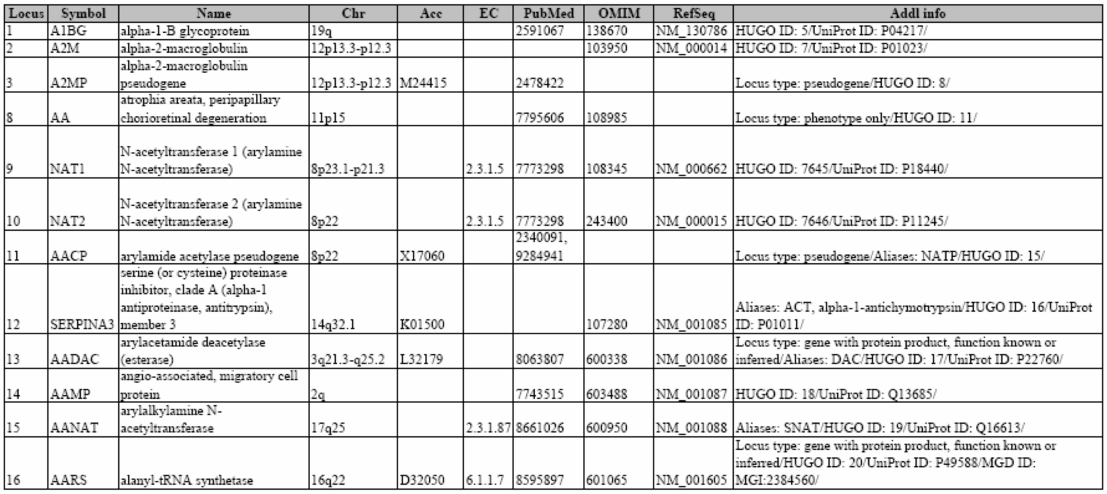

A. Determination of initial component list
NCBI
H-Inv DB
B. Manual curation of network components
1500+ citations
5x
280+ metabolic functions
C. Functional testing and validation
D. Dissemination of network content
BiGG Human Database
Browse and search tools
Human-specific network maps
Stoichiometric matrix
Reaction, metabolite & gap lists
E. 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

Figure 5.2: The development, dissemination, and possible applications of *H. sapiens* Recon 1. Initial component lists were derived from the genome annotation and pathway databases. Iterative rounds of manual reconstruction and gap analysis were required to form a functional, predictive model. The result of this procedure is *H. sapiens* Recon 1, a biochemically, genetically, and genomically integrated (BiGG) database whose contents are available at http://bigg.ucsd.edu. Recon 1 has many potential applications, many of which may aid in the elucidation and treatment of human disease.



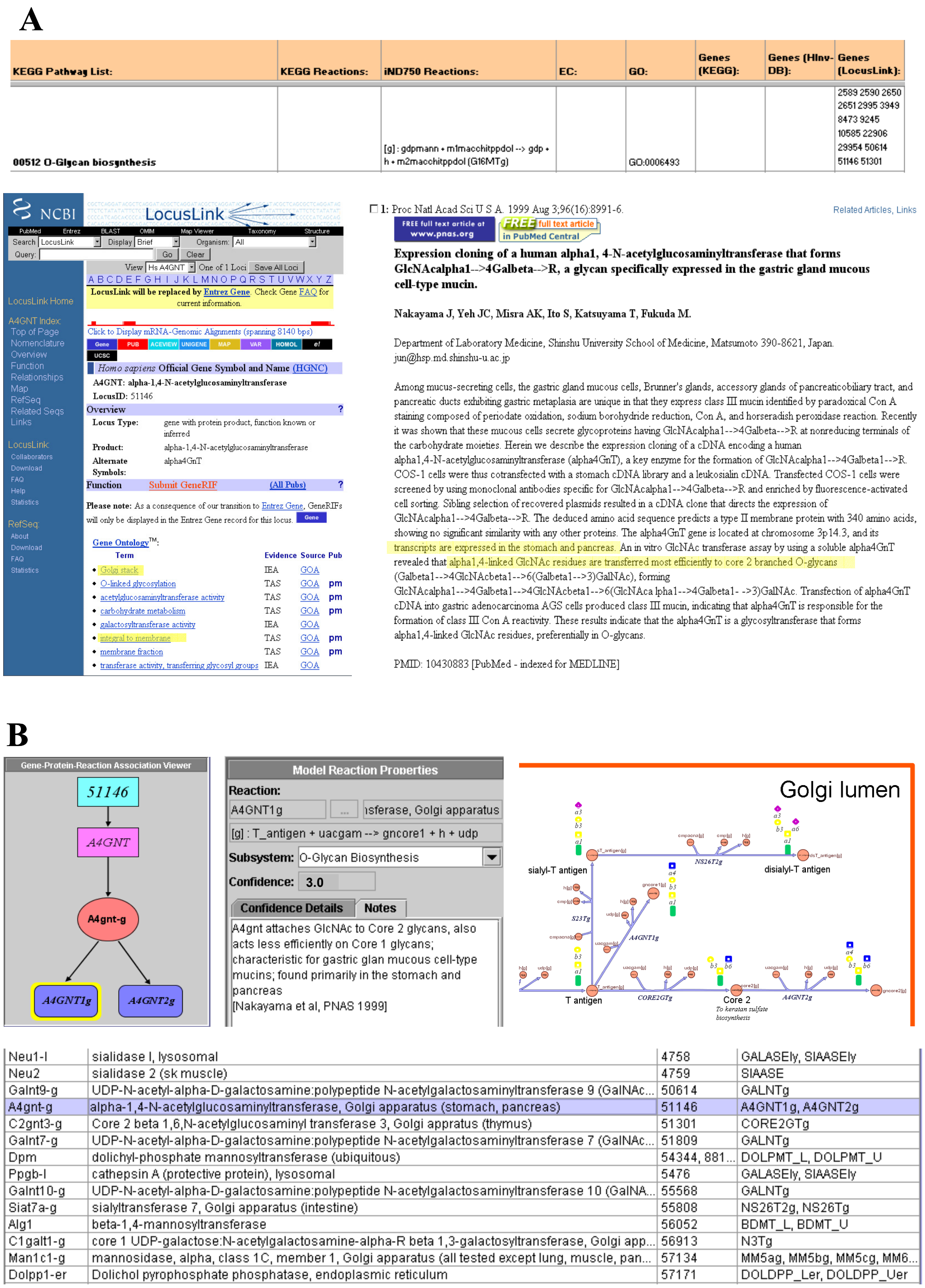
KEGG ontology	KEGG reactions	*S. cerevisiae* iND750 reactions	EC	GO	KEGG genes	HInv-DB genes	LocusLink genes
E2.7.1.2, glk; glucokinase	ATP + beta-D-Glucose <=> ADP + beta-D-Glucose 6-phosphate	[c] : atp + glc-D --> adp + g6p-B + h (GLUK)	EC:2.7.1.2	GO:0004340			2645
E2.7.1.1; hexokinase	ATP + D-Fructose <=> ADP + D-Fructose 6-phosphate/ ATP + beta-D-Glucose <=> ADP + beta-D-Glucose 6-phosphate	[c] : atp + fru --> adp + f6p + h (HEX7)/ [c] : atp + man --> adp + h + man6p (HEX4)/ [c] : atp + glc-D --> adp + g6p + h (HEX1)	EC:2.7.1.1	GO:0004396	2645 3098 3099 3101	3098 2645 3101 80201	2645 3098 3099 3101 80201
E3.1.3.9, G6PC; glucose-6-phosphatase	D-Glucose 6-phosphate + H2O <=> D-Glucose + Orthophosphate/ alpha-D-Glucose 6-phosphate + H2O <=> alpha-D-Glucose + Orthophosphate		EC:3.1.3.9	GO:0004346	2538		2538
E5.3.1.9, pgi; glucose-6-phosphate isomerase	alpha-D-Glucose 6-phosphate <=> beta-D-Glucose 6-phosphate/ alpha-D-Glucose 6-phosphate <=> beta-D-Fructose 6-phosphate/ beta-D-Glucose 6-phosphate <=> beta-D-Fructose 6-phosphate	[c] : g6p <==> f6p (PGI)/ [c] : g6p-B <==> f6p-B (G6PI2)/ [c] : g6p <==> g6p-B (G6PI)	EC:5.3.1.9	GO:0004347	2821	2821	2821
E2.7.1.11, pfk; 6-phosphofructokinase	ATP + D-Fructose 6-phosphate <=> ADP + D-Fructose 1,6-bisphosphate	[c] : atp + s7p --> adp + h + s17bp (PFK_3)/ [c] : atp + tag6p-D --> adp + h + tagdp-D (PFK_2)/ [c] : atp + f6p --> adp + fdp + h (PFK)	EC:2.7.1.11	GO:0003872	5211 5213 5214	5211 5213 5214	5211 5213 5214
E3.1.3.11; fructose-1,6-bisphosphatase	beta-D-Fructose 1,6-bisphosphate + H2O <=> beta-D-Fructose 6-phosphate + Orthophosphate/ D-Fructose 1,6 bisphosphate + H2O <=> D-Fructose 6-phosphate + Orthophosphate/ Sedoheptulose 1,7-bisphosphate + H2O <=> Sedoheptulose 7-phosphate + Orthophosphate/ D-Fructose 1,6-bisphosphate + H2O <=> beta-D-Fructose 6-phosphate + Orthophosphate	[c] : fdp + h2o --> f6p + pi (FBP)	EC:3.1.3.11	GO:0042132	2203 8789	2203	2203 8789

Figure 5.3: Snapshot of the initial human metabolic component list. Metabolic enzymes in KEGG Orthology were used as a scaffold for mapping an initial set of 1,865 candidate metabolic genes and 3,673 reactions (Supplement B). Metabolite and reaction definitions for *S. cerevisiae* iND750 can be found in Supplement A. The complete list can be found in Supplement B. Genes are listed by their unique locus identifiers. Abbreviations: KEGG – Kyoto Encyclopedia of Genes and Genomes, EC – Enzyme Commission numbers, GO – Gene Ontology terms, HInv-DB – H-Invitational Database.



Locus	Symbol	Name	Chr	Acc	EC	PubMed	OMIM	RefSeq	Addl info
1	A1BG	alpha-1-B glycoprotein	19q			2591067	138670	NM_130786	HUGO ID: 5/UniProt ID: P04217/
2	A2M	alpha-2-macroglobulin	12p13.3-p12.3				103950	NM_000014	HUGO ID: 7/UniProt ID: P01023/
3	A2MP	alpha-2-macroglobulin pseudogene	12p13.3-p12.3	M24415		2478422			Locus type: pseudogene/HUGO ID: 8/
8	AA	atrophia areata, peripapillary chorioretinal degeneration	11p15			7795606	108985		Locus type: phenotype only/HUGO ID: 11/
9	NAT1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	8p23.1-p21.3		2.3.1.5	7773298	108345	NM_000662	HUGO ID: 7645/UniProt ID: P18440/
10	NAT2	N-acetyltransferase 2 (arylamine N-acetyltransferase)	8p22		2.3.1.5	7773298	243400	NM_000015	HUGO ID: 7646/UniProt ID: P11245/
11	AACP	arylamide acetylase pseudogene	8p22	X17060		2340091, 9284941			Locus type: pseudogene/Aliases: NATP/HUGO ID: 15/
12	SERPINA3	serine (or cysteine) proteinase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	14q32.1	K01500			107280	NM_001085	Aliases: ACT, alpha-1-antichymotrypsin/HUGO ID: 16/UniProt ID: P01011/
13	AADAC	arylacetamide deacetylase (esterase)	3q21.3-q25.2	L32179		8063807	600338	NM_001086	Locus type: gene with protein product, function known or inferred/Aliases: DAC/HUGO ID: 17/UniProt ID: P22760/
14	AAMP	angio-associated, migratory cell protein	2q			7743515	603488	NM_001087	HUGO ID: 18/UniProt ID: Q13685/
15	AANAT	arylalkylamine N-acetyltransferase	17q25		2.3.1.87	8661026	600950	NM_001088	Aliases: SNAT/HUGO ID: 19/UniProt ID: Q16613/
16	AARS	alanyl-tRNA synthetase	16q22	D32050	6.1.1.7	8595897	601065	NM_001605	Locus type: gene with protein product, function known or inferred/HUGO ID: 20/UniProt ID: P49588/MGD ID: MGI:2384560/

Figure 5.4: Snapshot of the human gene index. The human gene index contains more than 16,000 genes and includes references to several biological databases. Abbreviations: Chr – chromosomal location, Acc – Swiss-Prot Accession number, EC – Enzyme Commission number, PubMed – PubMed identifier, OMIM – Online Mendelian Inheritance in Man identifier, RefSeq – RefSeq mRNA identifier, Addl info – additional information, which include: aliases, locus type, UniProt identifiers, Human Genome Organization identifiers (HUGO), Mouse Genome Database (MGD) and Mouse Genome Index (MGI) identifiers.



KEGG Pathway List:	KEGG Reactions:	iND750 Reactions:	EC:	GO:	Genes (KEGG):	Genes (HInv-DB):	Genes (LocusLink):
00512 O-Glycan biosynthesis		[g] : gdpmann + m1macchitppdol --> gdp + h + m2macchitppdol (G16MTg)		GO:0006493			2589 2590 2650 2651 2995 3949 8473 9245 10585 22906 29954 50614 51146 51301

Search: LocusLink Display: Brief Organism: All
Query:
View: Hs A4GNT One of 1 Loci

LocusLink will be replaced by Entrez Gene. Check Gene FAQ for current information.

Click to Display mRNA-Genomic Alignments (spanning 8140 bps)

Homo sapiens **Official Gene Symbol and Name (HGNC)**

A4GNT: alpha-1,4-N-acetylglucosaminyltransferase

LocusID: 51146

Overview ?

Locus Type: gene with protein product, function known or inferred

Product: alpha-1,4-N-acetylglucosaminyltransferase

Alternate Symbols: alpha4GnT

Function Submit GeneRIF (All Pubs) ?

Please note: As a consequence of our transition to Entrez Gene, GeneRIFs will only be displayed in the Entrez Gene record for this locus.

Gene Ontology™:

Term	Evidence	Source	Pub
Golgi stack	IEA	GOA	
O-linked glycosylation	TAS	GOA	pm
acetylglucosaminyltransferase activity	TAS	GOA	pm
carbohydrate metabolism	TAS	GOA	pm
galactosyltransferase activity	IEA	GOA	
integral to membrane	TAS	GOA	pm
membrane fraction	TAS	GOA	pm
transferase activity, transferring glycosyl groups	IEA	GOA	

□ 1: Proc Natl Acad Sci U S A. 1999 Aug 3;96(16):8991-6.

Expression cloning of a human alpha1, 4-N-acetylglucosaminyltransferase that forms GlcNAcalpha1-->4Galbeta-->R, a glycan specifically expressed in the gastric gland mucous cell-type mucin.

Nakayama J, Yeh JC, Misra AK, Ito S, Katsuyama T, Fukuda M.

Department of Laboratory Medicine, Shinshu University School of Medicine, Matsumoto 390-8621, Japan.
jun@hsp.md.shinshu-u.ac.jp

Among mucus-secreting cells, the gastric gland mucous cells, Brunner's glands, accessory glands of pancreaticobiliary tract, and pancreatic ducts exhibiting gastric metaplasia are unique in that they express class III mucin identified by paradoxical Con A staining composed of periodate oxidation, sodium borohydride reduction, Con A, and horseradish peroxidase reaction. Recently it was shown that these mucous cells secrete glycoproteins having GlcNAcalpha1-->4Galbeta-->R at nonreducing terminals of the carbohydrate moieties. Herein we describe the expression cloning of a cDNA encoding a human alpha1,4-N-acetylglucosaminyltransferase (alpha4GnT), a key enzyme for the formation of GlcNAcalpha1-->4Galbeta1-->R. COS-1 cells were thus cotransfected with a stomach cDNA library and a leukosialin cDNA. Transfected COS-1 cells were screened by using monoclonal antibodies specific for GlcNAcalpha1-->4Galbeta-->R and enriched by fluorescence-activated cell sorting. Sibling selection of recovered plasmids resulted in a cDNA clone that directs the expression of GlcNAcalpha1-->4Galbeta-->R. The deduced amino acid sequence predicts a type II membrane protein with 340 amino acids, showing no significant similarity with any other proteins. The alpha4GnT gene is located at chromosome 3p14.3, and its transcripts are expressed in the stomach and pancreas. An in vitro GlcNAc transferase assay by using a soluble alpha4GnT revealed that alpha1,4-linked GlcNAc residues are transferred most efficiently to core 2 branched O-glycans (Galbeta1-->4GlcNAcbeta1-->6(Galbeta1-->3)GalNAc), forming GlcNAcalpha1-->4Galbeta1-->4GlcNAcbeta1-->6(GlcNAca lpha1-->4Galbeta1- ->3)GalNAc. Transfection of alpha4GnT cDNA into gastric adenocarcinoma AGS cells produced class III mucin, indicating that alpha4GnT is responsible for the formation of class III Con A reactivity. These results indicate that the alpha4GnT is a glycosyltransferase that forms alpha1,4-linked GlcNAc residues, preferentially in O-glycans.

PMID: 10430883 [PubMed - indexed for MEDLINE]

Model Reaction Properties

Reaction: A4GNT1g ... ısferase, Golgi apparatus

[g] : T_antigen + uacgam --> gncore1 + h + udp

Subsystem: O-Glycan Biosynthesis

Confidence: 3.0

A4gnt attaches GlcNAc to Core 2 glycans, also acts less efficiently on Core 1 glycans; characteristic for gastric glan mucous cell-type mucins; found primarily in the stomach and pancreas
[Nakayama et al, PNAS 1999]

Neu1-l	sialidase I, lysosomal	4758	GALASEly, SIAASEly
Neu2	sialidase 2 (sk muscle)	4759	SIAASE
Galnt9-g	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 9 (GalNAc...	50614	GALNTg
A4gnt-g	alpha-1,4-N-acetylglucosaminyltransferase, Golgi apparatus (stomach, pancreas)	51146	A4GNT1g, A4GNT2g
C2gnt3-g	Core 2 beta 1,6,N-acetylglucosaminyl transferase 3, Golgi appratus (thymus)	51301	CORE2GTg
Galnt7-g	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 7 (GalNAc...	51809	GALNTg
Dpm	dolichyl-phosphate mannosyltransferase (ubiquitous)	54344, 881...	DOLPMT_L, DOLPMT_U
Ppgb-l	cathepsin A (protective protein), lysosomal	5476	GALASEly, SIAASEly
Galnt10-g	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (GalNA...	55568	GALNTg
Siat7a-g	sialyltransferase 7, Golgi apparatus (intestine)	55808	NS26T2g, NS26Tg
Alg1	beta-1,4-mannosyltransferase	56052	BDMT_L, BDMT_U
C1galt1-g	core 1 UDP-galactose:N-acetylgalactosamine-alpha-R beta 1,3-galactosyltransferase, Golgi app...	56913	N3Tg
Man1c1-g	mannosidase, alpha, class 1C, member 1, Golgi apparatus (all tested except lung, muscle, pan...	57134	MM5ag, MM5bg, MM5cg, MM6...
Dolpp1-er	Dolichol pyrophosphate phosphatase, endoplasmic reticulum	57171	DOLDPP_Ler, DOLDPP_Uer

Figure 5.5: Assembling reconstruction data in SimPheny. *(A)* Genome annotation databases (2.1.1) are used to review published reports on genes identified in the initial component list. *(B)* Recording this information in SimPheny as gene-protein-reaction associations, protein and reaction entries, and on metabolic maps.

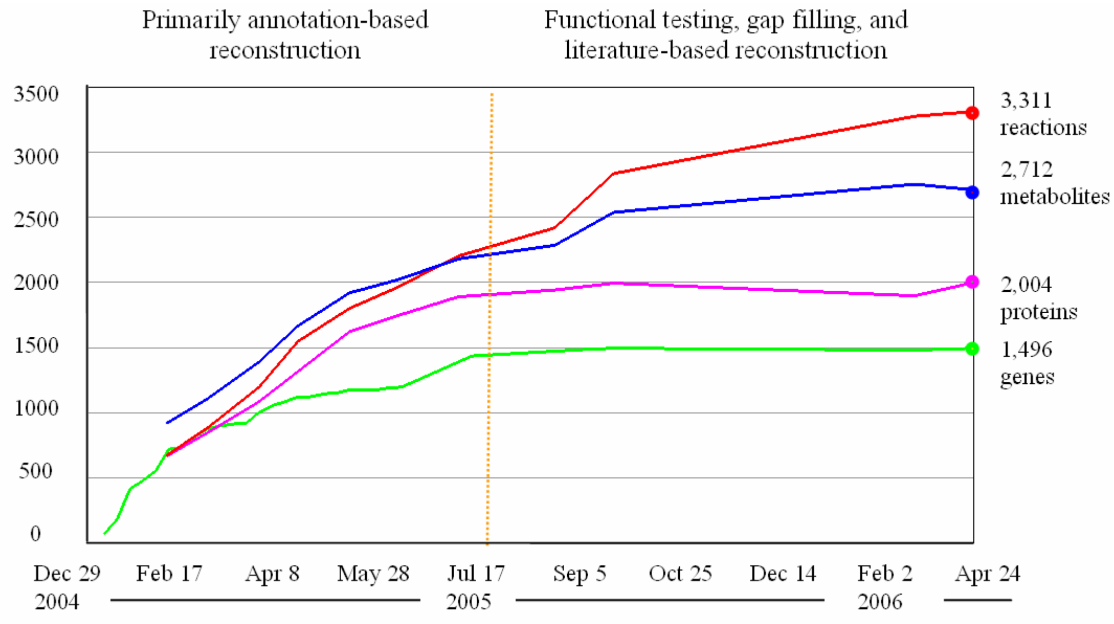
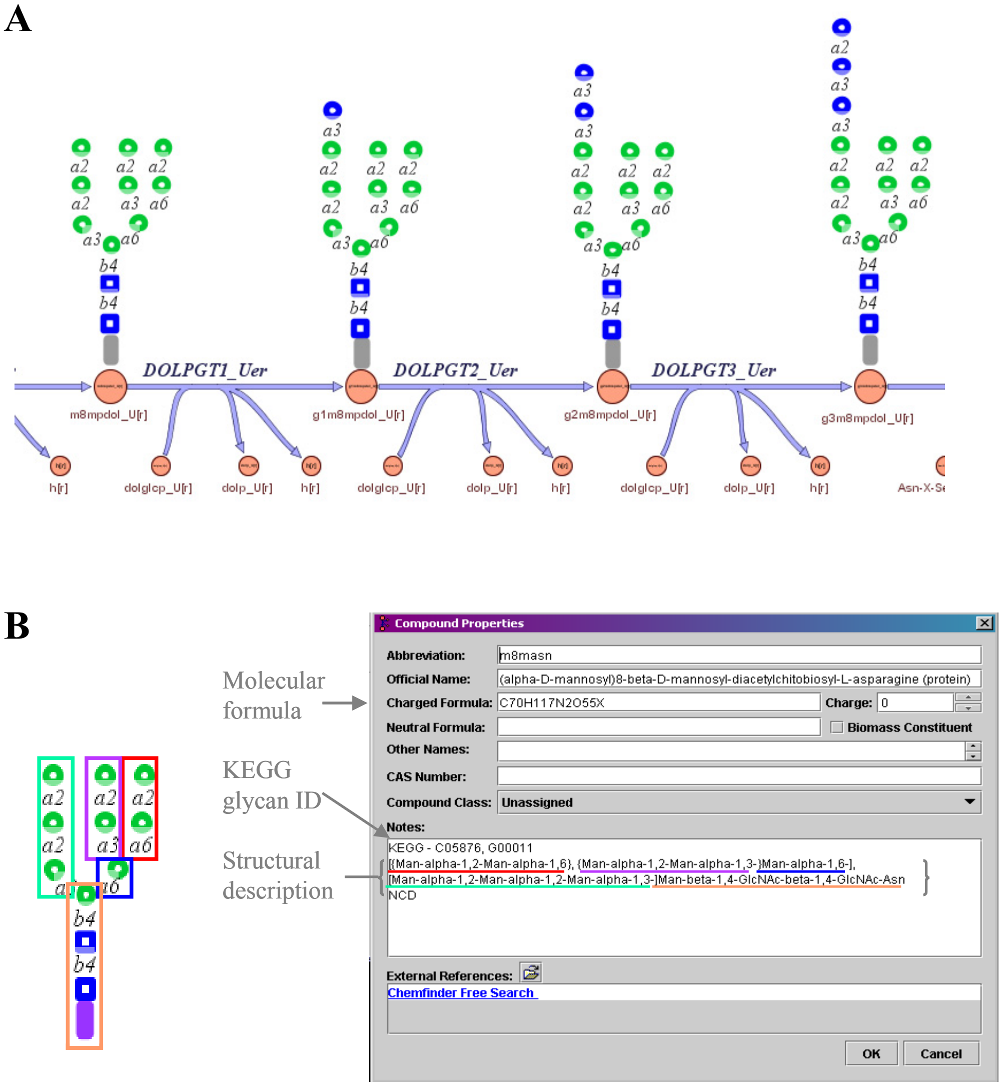


Figure 5.6: Timeline of the human metabolic reconstruction. The reconstruction can be roughly divided into phases of annotation- and literature-based reconstruction. Note that there is a rapid increase in the number of components during the annotation based phase, in which network content was determined by direct curation of the automated component list. Most of the reactions added during the literature-based reconstruction were not gene associated, and were typically included to resolve network gaps. The overall reconstruction and debugging required 18 months.



Compound Properties

Abbreviation: m8masn

Official Name: (alpha-D-mannosyl)8-beta-D-mannosyl-diacetylchitobiosyl-L-asparagine (protein)

Charged Formula: C70H117N2O55X Charge: 0

Neutral Formula:

□ Biomass Constituent

Other Names:

CAS Number:

Compound Class: Unassigned

Notes:

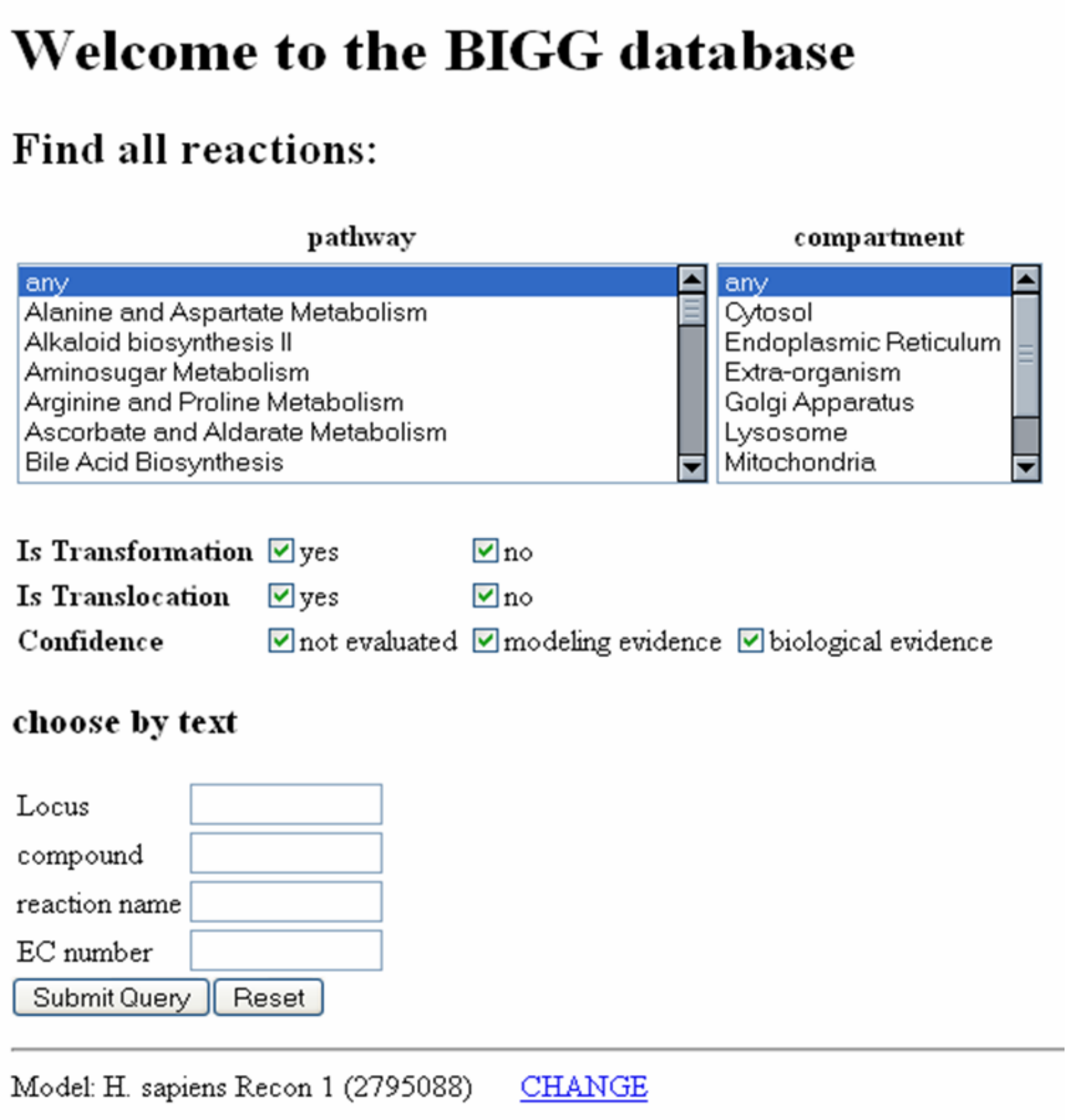
KEGG - C05876, G00011
[{Man-alpha-1,2-Man-alpha-1,6}, {Man-alpha-1,2-Man-alpha-1,3-}Man-alpha-1,6-],
[Man-alpha-1,2-Man-alpha-1,2-Man-alpha-1,3-]Man-beta-1,4-GlcNAc-beta-1,4-GlcNAc-Asn
NCD

External References:

Chemfinder Free Search

OK Cancel

Figure 5.7: Glycan representations in *H. sapiens* Recon 1. *(A)* Glycan structures are drawn directly on Recon 1's metabolic maps according to the convention described in Essentials of Glycobiolgy [125]. *(B)* Compound entries also include molecular formula and structural descriptions of glycans.



Welcome to the BIGG database

Find all reactions:

pathway	compartment
any	any
Alanine and Aspartate Metabolism	Cytosol
Alkaloid biosynthesis II	Endoplasmic Reticulum
Aminosugar Metabolism	Extra-organism
Arginine and Proline Metabolism	Golgi Apparatus
Ascorbate and Aldarate Metabolism	Lysosome
Bile Acid Biosynthesis	Mitochondria

Is Transformation ☑ yes ☑ no

Is Translocation ☑ yes ☑ no

Confidence ☑ not evaluated ☑ modeling evidence ☑ biological evidence

choose by text

Locus

compound

reaction name

EC number

Model: H. sapiens Recon 1 (2795088) CHANGE

Figure 5.8: Snapshot of the BiGG human database. The BiGG database (http://bigg.ucsd.edu) is a website for browsing and searching the contents of *H. sapiens* Recon 1. It also well integrated with a variety of other databases (Table 5.6).

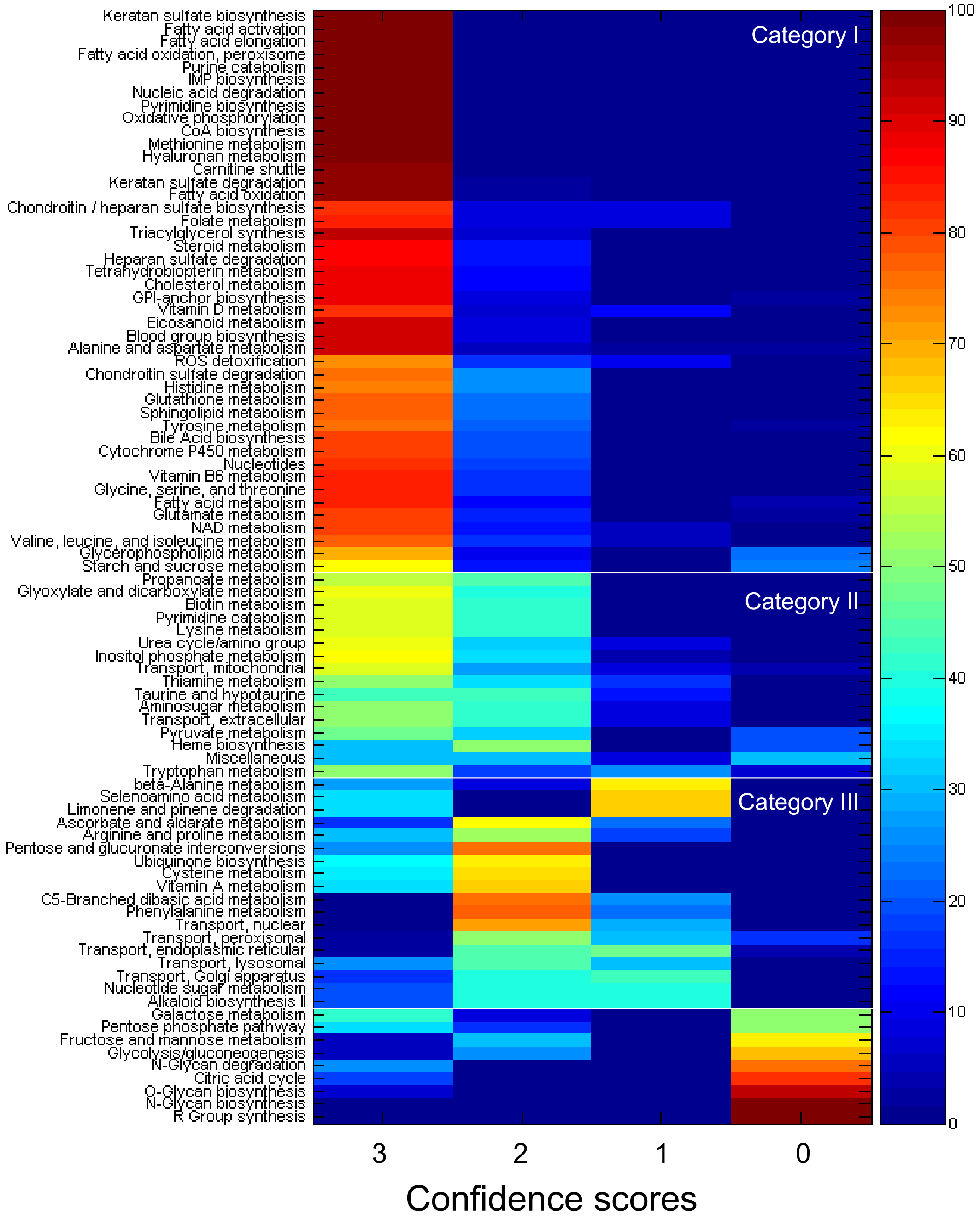


Figure 5.9: Global human metabolic knowledge landscape. Colors represent the percent of reactions within a pathway which have a confidence score of 3 (biochemical or genetic evidence), 2 (physiological data or evidence from a non-human mammalian cell), 1 (modeling evidence), or 0 (unevaluated). Pathways were classified into three categories based on their level of characterization (Figures 5.10-5.12).

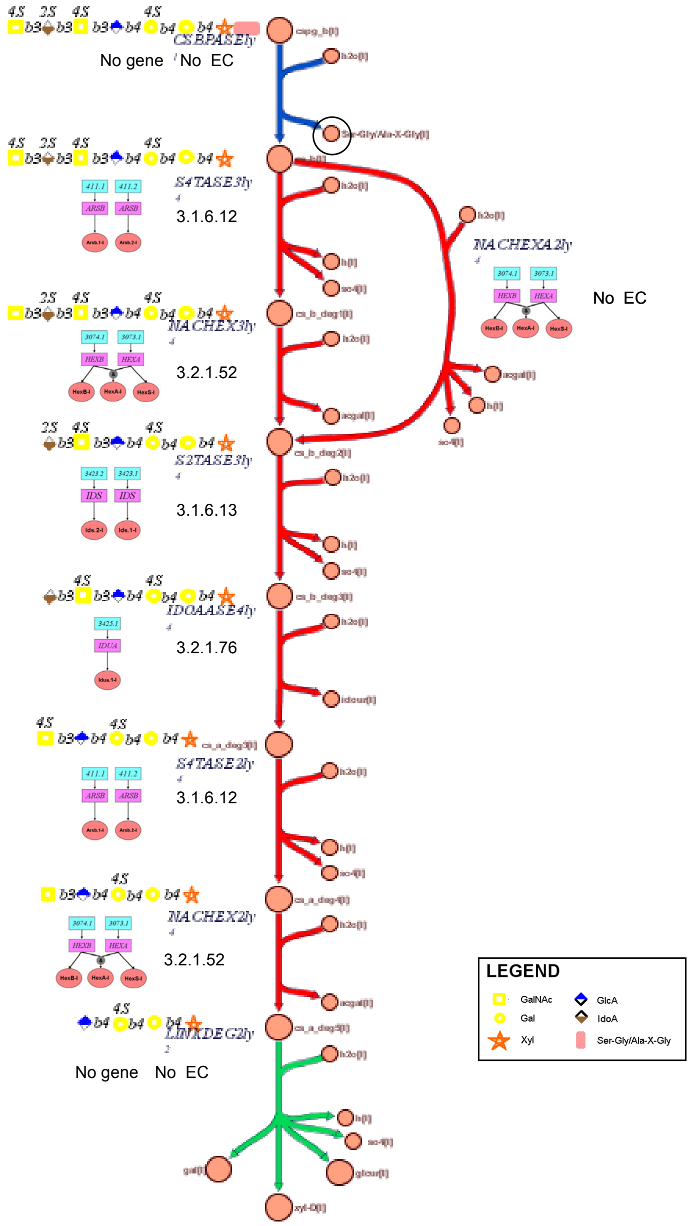


Figure 5.10: Knowledge landscape for heparan sulfate degradation. Proteoglycans presumably undergo initial cleavage extracellularly, producing short peptides with single chondroitin sulfate chains (cspg_b). These chains are then endocytosed and further degraded by endosomal endoglycosidases to produce free chondroitin sulfate chains (cs_b). No biological evidence supporting this mechanism has been identified yet [126]. Note that the peptide by-product (Ser-Gly/Ala-X-Gly) is a “dead-end” metabolite that is produced but not consumed. Final degradation of the core tetrasaccharide linkage (LINKDEG2ly) was inferred based on enzymes identified in rabbit [126]. Reactions are color-coded by confidence scores: 3 – red, 2 – green, 1 – blue. Gene, metabolite, and reaction abbreviations are defined in Supplement B.

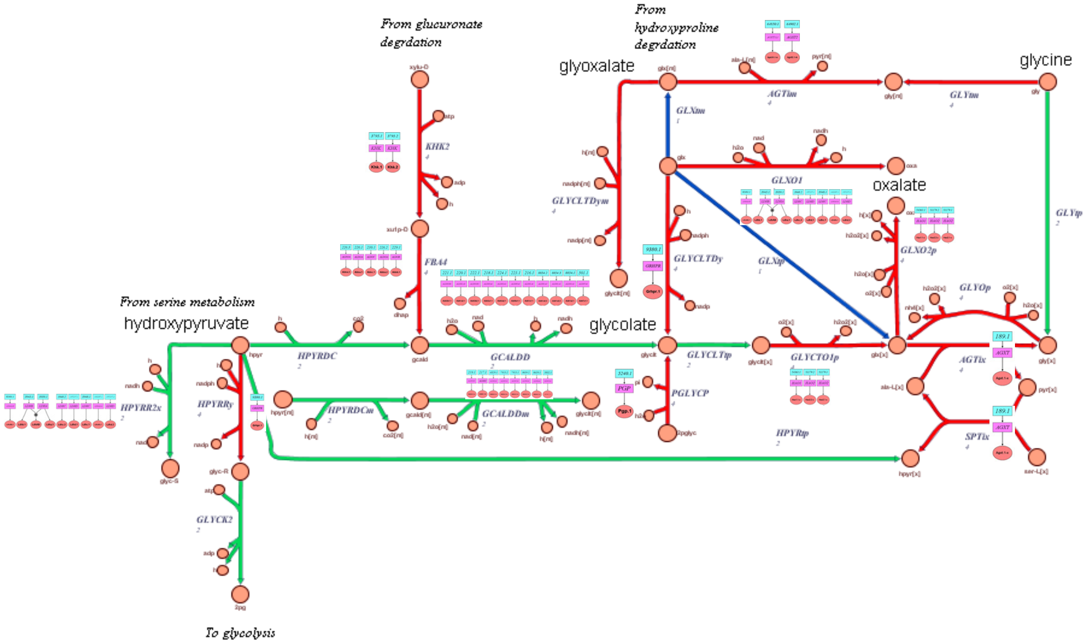


Figure 5.11: Knowledge landscape for glyoxylate and dicarboxylate metabolism. Category II pathways typically have a combination of well-known functions, such as the degradation of glyoxylate (glx) to glycine (gly) in normal physiological conditions and overproduction of oxalate (oxa) in oxalosis, and those that are poorly understood, such as the feedback of glycolate intermediate hydroxypyruvate (hpyr) to glycolysis**.** Reactions are color-coded by confidence scores: 3 – red, 2 – green, 1 – blue. Metabolite and reaction abbreviations are defined in Supplement B.

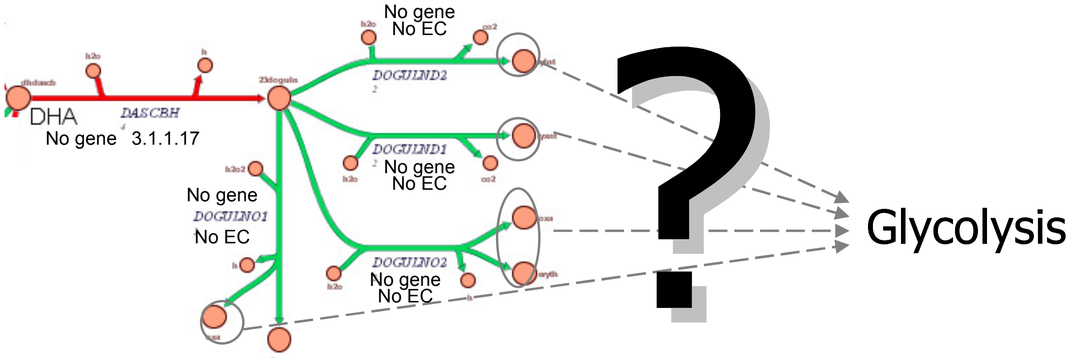


Figure 5.12: Knowledge landscape for vitamin C metabolism. The degradation of 2,3-dioxo-L-gulonate (23doguln) to L-xylonate (xylnt), L-lyxonate (lyxnt), and L-threonate (thrnt) are supported by physiological evidence [123, 127], but the exact reaction mechanisms in which these four- and five-carbon sugar acids are converted to glycolytic intermediates could not be identified in the literature. Reactions are color-coded by confidence scores: 3 – red, 2 – green, 1 – blue. Metabolite and reaction abbreviations are defined in Supplement B.

Table 5.1: Comparison of the initial and finished human genome sequences.

	Draft sequences [95, 96]	**Finished sequence** [114]
Date released	February 2001	October 2004
Euchromatin coverage	90%	99%
Length (bp)	3.08 billion	2.85 billion
Gaps	150,000	341
Error rate (events per bp)	1 in 1,000	1 in 100,000

Table 5.2: Division of the initial component list for parallel manual curation.

Subset	Primary researcher(s)	# of enzymes	% with reaction	% with gene
Carbohydrates	Natalie Hurlen	725	96	50
Glycans	Natalie Hurlen	280	90	71
Nucleotides	Ines Thiele	182	94	73
Vitamins & cofactors	Ines Thiele	330	94	42
Lipids	Neema Jamshidi	305	95	71
Amino acids	Scott Becker & Monica Mo	807	97	58
Secondary metabolites	Scott Becker	123	78	34
Energy	Rohith Srivas & Thuy Vo	680	78	55

Table 5.3: Examples of metabolic objectives used for functional validation. A complete list is available in Supplement B.

Simple pathway objectives
• Synthesize thyroid hormones from L-tyrosine • Synthesize cholesterol from HMG-CoA • Degrade heparan sulfate • Catabolize histidine to alpha-ketoglutarate • Synthesize UMP from L-glutamine
Comprehensive physiological functions
• Simplified cell biomass (as defined in [113]) • Synthesize glucose from L-alanine • Synthesize ketone bodies from L-leucine and L-isoleucine • Degrade glycogen into free glucose
Parameter calculations
• P/O ratio • Maximal ATP production

Table 5.4: Classification of network gaps. The source of each metabolite gap was thoroughly investigated and led to their classification as either knowledge- or model scope-limited. A complete list of gaps and their classifications is available in Supplement B.

Metabolite	Compartment	Classification
Glucaric acid	Mitochondria	Knowledge-base gap. Seems to be a dead-end metabolite in humans [128].
Hydroxypyruvate	Mitochondria	Knowledge-base gap. Very few reports describing the properties of the corresponding enzyme (hydroxypyruvate decarboxylase), none in human tissues.
Isocitrate	Peroxisome	Knowledge-base gap. The source of peroxisomal isocitrate has not been determined [129].
Maltotriose	Lysosome	Model-scope gap. Maltotriose usually arises from glycogen degradation, but it is not a degradation product of the representative glycogen structure included in our network.
Phosphatidylinositol	Endoplasmic reticulum	Model-scope gap. This gap is due to alternative localization of an enzyme that is part of a complete (functional) cytosolic pathway.
Ser-Gly/Ala-X-Gly	Endoplasmic reticulum	Model scope gap. Represents peptide binding site for glycosaminoglycan chains, which is not synthesized in our model.

Table 5.5: *H. sapiens* Recon 1 network statistics. * Transcripts, complexes, and isozymes are defined in 5.2.3. † Transport reactions describe metabolite transport across organellar and plasma membranes, whereas exchange reactions describe metabolite transport into and out of the extra-cellular space from the surrounding medium. ‡ Number of dead-end metabolites that are only produced or consumed. A complete list of Recon 1's content can be found in Supplement B.

Component	Count
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**
Cytosol	995
Extra-cellular 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**

Table 5.6: Examples of queries available at the BiGG Human Database.

Sample queries
• What are all of the peroxisomal transport reactions? • What are all of the reactions which are not gene associated OR are only supported by modeling data? • What are all of the reactions which are located in the cytosol AND are involved in L-alanine metabolism?

Table 5.7: The BiGG database links to several public databases. Abbreviations: HUGO – Human Genome Organization, KEGG – Kyoto Encyclopedia of Genes and Genomes, GO – Gene Ontology, MGD – Mouse Genome Database, OMIM – Online Mendelian Inheritance in Man.

Genes	Functionality
Entrez Gene [9] HUGO [130] Ensembl [131]	GO [18] Brenda [16] KEGG [14]
Transcripts	**Homology**
RefSeq [9]	MGD [132]
Proteins	**Disease**
UniProt [133] Swiss-Prot [135]	OMIM [134]
Metabolites	**Literature**
KEGG [14] NIST Chemistry Web [136]	PubMed [8]

CHAPTER 6: GENE EXPRESSION ANALYSIS IN THE CONTEXT OF A GENOME-SCALE NETWORK

Recent advances in high-throughput experimentation have led to an accumulation of large data sets that simultaneously measure the state of thousands of cellular components. For example, since their development in the early 1990s, more than 7,000 scientific publications using Affymetrix's GeneChip technology have appeared [137]. As a result, there has been a growing interest in using these data to characterize cellular processes at the system level. In this chapter, we present examples of how gene expression data can be used in conjunction with *H. sapiens* Recon 1 (Chapter 5) to interrogate metabolic states (6.1) and to develop cell and tissue specific models (6.2).

6.1 Integrated analysis of skeletal muscle metabolism

Skeletal muscle metabolism is directly linked to its composition, containing a mixture of oxidative (Type I) and glycolytic (Type II) fibers (for a review, see [138]). The distribution of these fiber types is the result of both genetics and a variety of external stimuli such as nutrition, environment, activity, and loading [139-142], and varies considerably from person-to-person [143]. There are known differences between the relative distribution of these fibers in lean and obese populations [144-148]. Here we present a case study that explores how dramatic weight loss and nutritional restriction effects metabolism in obese skeletal muscle.

6.1.1 The obesity epidemic

Obesity has been named the biggest health problem of the century. Over 1.5 billion adults and 10% of children worldwide are overweight or obese [149]. The prevalence of obesity is especially alarming in industrialized countries, where the incidence has doubled over the past decade [150]. In the U.S. alone, 65% of adults are overweight [151], and obesity-related conditions result in 300,000 annual deaths [152]. The total direct and indirect cost of obesity-related medical treatments in the U.S. was $117 billion for the year 2000 [153]. Clearly, the social and economic effects of obesity are staggering, and the effects are being felt worldwide.

<u>Clinical classifications of obesity</u>

Obesity is typically defined using a using a height-weight metric known as the body mass index (BMI):

$$BMI = 703 * \frac{weight(lb)}{height(in)^2} = \frac{weight(kg)}{height(m)^2}$$

The range of BMI scores and corresponding weight classifications are shown in Table 6.1. There has been much discussion about the validity of the BMI scale because factors such as body frame and excessive muscularity can skew scores [153]. Consequently, modifications have been suggested based on race, sex, and age [154-159]. Despite these discrepancies, BMI remains to be the industry standard for clinical diagnosis of obesity.

<u>Bariatric surgery as a treatment for morbid obesity</u>

According to the National Heart, Blood, and Lung Institute Guidelines [160], surgical intervention is good option for some patients with clinically severe obesity because the benefits of surgical intervention outweigh obesity-associated health risks, which include type 2 diabetes, hypertension, coronary artery disease, gallbladder disease, osteoarthritis, cancer, and early death [161]. Gastric bypass is quickly becoming an established treatment for obesity, with a rise from 16,000 performed annually in the early 1990s to more than 103,000 in 2003 [162]. Patients who have undergone bypass surgery typically experienced short-term loss of 40-80% of their excess body weight [163, 164] and significant improvement in co-morbidities [164].

The most common form of bariatric surgery performed in the U.S. is Roux-en-Y gastric bypass [165]. This procedure bypasses the majority of the stomach and duodenum, leaving only a small 10 to 30 mL stomach pouch and shortened small intestine (Figure 1). The surgery is designed to both reduce stomach capacity and restrict absorption of fats. However, it also commonly results in several known nutritional and metabolic complications, namely vitamin B12 deficiency, iron deficiency, thiamine deficiency, metabolic bone disease, and cholelithiasis [166-168].

6.1.2 Metabolic gene expression in obese skeletal muscle

H. sapiens Recon 1 was used to investigate the metabolic effects of gastric bypass in human skeletal muscle. Published gene expression data [150] from the vastus lateralis muscle of three morbidly obese patients was examined before and one year after Roux-en-Y gastric bypass (once weight had stabilized). The patients experienced significant reductions in weight (45%) and BMI (46%), with an average post-surgery weight of 200 lbs (90.6 kg) and BMI of 32.9. While their weight loss was dramatic, it is important to note that the patients were still considered obese one year post-surgery. However, the authors noticed a significant reduction in post-treatment insulin levels, which is indicative of improved insulin sensitivity and a decreased risk for co-morbidities such as type 2 diabetes.

Mapping gene expression data to *H. sapiens* Recon 1

The procedure used to map gene expression measurements to the *H. sapiens* Recon 1 reaction network is outlined here. The gastric bypass study included six gene expression data sets (one for each patient pre- and post-surgery) collected with Affymetrix U133A Plus 2.0 chips. The data sets were downloaded from Gene Expression Omnibus (GEO) (GEO ID: GSE5109; [169]) and normalized by dividing signal measurements by the average of all measurements on the chip. The log10 ratio of post/pre-surgery expression signals were then calculated for each patient. Probes were mapped to Entrez Gene and RefSeq mRNA IDs [9] in *H. sapiens* Recon 1 based on database identifiers in the Affymetrix U133A Plus 2.0 annotation file [170], resulting in a set of 2,071 candidate metabolic probes. The probe list was then further refined to remove probes whose expression ratio was qualitatively inconsistent across all three patients (*i.e.*, not all up or all down). The remaining 516 probes were matched with their corresponding genes and an additional 24 were removed due to qualitative conflicts at the gene level. The average expression ratio was calculated for each gene and then mapped to the reaction network using Recon 1's gene-transcript-protein-reaction associations (Supplement B; 5.2.3).

Metabolic gene expression patterns observed pre- and post-gastric bypass surgery

H. sapiens Recon 1's comprehensive collection of integrated, genome-scale metabolic maps (Supplement B) was used to obtain a pathway-level view of gene expression trends from the gastric bypass data set (Figures 6.2-6.4). We observed a general trend of up-regulated anaerobic metabolism and down-regulated oxidative phosphorylation post-surgery, with many genes in glycolysis, pentose phosphate pathway, methylglyoxal metabolism, and oxidative phosphorylation showing subtle but consistent overall patterns of expression change (Tables 6.2-6.4). This pattern appears to be consistent with the down-regulation of genes involved in mitochondrial bioenergetics that has been observed in the skeletal muscle of rhesus monkeys subjected to long-term caloric restriction [171]. Distinct changes in gene expression associated with collagen and glycosaminoglycan metabolism were also observed (Figure 6.4), suggesting that gastric bypass may lead to extensive remodeling of the extracellular matrix and cell surface proteoglycans. These changes may be attributed to alterations in hormone levels, which are known to strongly modulate the composition of the extracellular matrix [172, 173].

Gene expression changes were also examined in the context of reaction compartmentalization and metabolite connectivity (Figure 6.5). This was done by defining distinct reaction networks based on the up- and down-regulated genes. Genes that were more highly expressed pre-surgery (*i.e.*, in the morbidly obese state) generally involved a larger number of mitochondrial and peroxisomal functions, whereas post-surgery there is a shift towards higher metabolic activity in the cytosol and lysosome with a concomitant increase in extracellular metabolite exchange. Inspection of these networks at the metabolite level revealed that most compartmentalization differences could be attributed to increased oxidative energy metabolism and reactive oxygen species production pre-surgery and increased glycosaminoglycan degradation and amino acid-sodium co-transport post-surgery. The shifts between the types of glycosaminoglycans predominantly metabolized in each state were also apparent, with a decrease in mannose and glucose-6-phosphate utilization and an increase galactose and sulfur carriers post-surgery.

Metabolic gene expression patterns observed in lean and morbidly obese populations

To further investigate which effects were due to weight loss, gene expression measurements from the gastric bypass study were compared to published gene expression data from a cross-sectional group of eight lean and eight morbidly obese subjects (GEO ID: GSE474, [174]). Each of the 16 data sets (one per patient, taken from the rectus abdominus muscle and profiled with Affymetrix U133A microarrays) was normalized by its signal average, and then individual probe measurements were pooled by calculating average values for lean and morbidly obese populations. The log10 of the lean-to-obese expression ratio was calculated and reaction mappings for central metabolism and oxidative phosphorylation (Figures 6.6-6.7) were compared to results from the gastric bypass patients (Figures 6.2-6.3).

Visual inspection of gene expression patterns in the cross-sectional study revealed distinct differences from the gastric bypass patients. For example, genes associated with oxidative phosphorylation were more highly expressed in lean subjects than morbidly obese, and those involved in ketogenesis and lactate metabolism were down-regulated. While these results are consistent with previous observations that obese individuals have a smaller percentage of Type I oxidative fibers and higher percentage of Type II glycolytic fibers than their lean counterparts [144-148], they seem to conflict with the changes we observed in the leaner, post-gastric bypass data set. Thus, this may suggest that nutrition, not weight loss, is the overriding factor in skeletal muscle metabolism one year post-gastric bypass surgery, as the observed metabolic gene expression patterns are generally more consistent with caloric restriction than the those in lean versus morbidly obese populations.

While the work presented here represents only a limited case study, it effectively illustrates how *H. sapiens* Recon 1 can be used as a context for visualizing and interpreting genome-scale content. The basic integration method and analysis tools developed here can be easily extended to accommodate future studies with larger patient populations and more stringent expression thresholds.

6.2 Tailoring the global metabolic network with pathway analysis

The global human metabolic network is a comprehensive representation of the metabolic capabilities encoded in the DNA of all human cells. However, gene expression is a highly regulated process, and consequently tissues express only a subset of their total complement of enzymes at any given time. Thus, refinement of network content is needed to perform accurate, quantitative simulations of metabolic functions in tissue- and condition-specific states.

In this study, we employ a top-down strategy (1.2) to rapidly generate rough drafts of tissue-specific metabolic networks based on their pathway-level gene expression patterns. This strategy employs established pathway analysis statistics [175], but is based on present/absent calls rather than raw expression measurements.

Determining whether a gene is “on” or “off” in a particular tissue

We would ideally like to incorporate all of the genes and reactions known to be expressed in a tissue in order to comprehensively model its full range of metabolic capabilities. Thus, our general strategy for tailoring the global human metabolic network is to be inclusive, only eliminating components if they are never present in the tissue of interest. However, a major challenge with microarray data is that it can only conclusively prove that a gene is expressed in, not absent from, a particular sample since it there are many reasons why it might be undetected (*e.g.*, poor probe design, insufficient sample size, or low expression levels). The simplest approach for determining whether a gene is "on" (present) or "off" (absent) in a particular system is to therefore to assume that genes are absent if they are undetected in all replicates [176]. Our implementation of this approach using Affymetrix's present, marginal, and absent calls is summarized in Table 6.5.

Statistical assessment of expression patterns in metabolic pathways

With a metric in place for calling genes on or off, the next step in our analysis was to implement a statistical method for assessing the likelihood of finding a certain number of genes turned on in a particular pathway (which can be loosely regarded as the probability that the pathway is expressed). A combined chi-squared/Fisher Exact test was implemented as described in [177]. Briefly, the chi-squared test describes how the observed number of hits deviates from what is expected. In terms of pathway analysis, the chi-squared test is counting the number of times genes from a specified gene list (*e.g.*, those that are up-regulated) occur in a particular pathway. This test gives only an approximate p-value and can only be used for cases in which there are five or more observations of each type [175]. The Fisher exact test is performed in cases with less than five observations and calculates the probability of seeing an observed number of hits or more. The null hypothesis in pathway analysis is that the relative expression changes of genes in a particular pathway (or pathway enrichment) are a random subset of those observed in the experiment as a whole [175]. Therefore the probability represents the chance that a pathway would contain as many or more affected genes as actually observed [178].

Pathway analysis of 79 healthy human tissues

Our pathway-based approach was used to generate global snapshot of metabolic pathway utilization in 79 healthy human tissues [179]. Expression data was downloaded from the Genomics Institute of the Novartis Research Foundation [180] and mapped to *H. sapiens* Recon 1 as described in section 6.1.2. A combined chi-square and Fisher's exact test was used to calculate p-values for each of *H. sapiens* Recon 1's 93 metabolic pathways based on its distribution of on/off calls (Figure 6.8).

A few pathways appear to be "on" in nearly all tissues. These pathways, which include glycolysis, oxidative phosphorylation, and propanoate metabolism, have a highly significant number of present calls for their associated genes. Several pathways were found to be differentially expressed, with some tissues expressing a significantly larger percentage of genes than others. Notably, glutamine metabolism was highly expressed in the brain, which is consistent with its known role as a prominent neurotransmitter. Glycosylation pathways were strongly expressed in blood cell lines, which may be indicative of the importance of signaling and communication in these tissues. While most samples expressed roughly the same number of metabolic pathways, two subgroups appeared to have higher- and lower-than average range of metabolic activity. The former category includes the liver, kidney and adipocytes, which clearly have distinct roles in general metabolism, while the latter includes the appendix, several nervous tissues and surprisingly, skeletal muscle and skin.

The results presented here demonstrate the utility of comprehensive, pathway-level analysis of gene expression data in determining high-level, functional differences between tissues. Our pathway analysis strategy was a simple extension of established pathway analysis methods [175], using present/absent calls and a highly curated reaction network rather than arbitrarily set expression cutoffs and genome annotations. Continual development of this method holds the promise of using gene expression data to refine the global human metabolic network into cell specific networks (Chapter 7).

6.3 Conclusions

Gene expression profiling is an effective tool for collecting comprehensive information on a system of interest under controlled conditions. Genome-scale networks such as *H. sapiens* Recon 1 can provide a context for the interpretation of such data sets and lead to new hypotheses on cellular function. Expression measurements are also useful for top-down tailoring of global networks, providing first-pass approximations of specific cell types, conditions, and disease states.

The text of this chapter, in part or in full, is a reprint of the material as it appears in N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.O. Palsson. 2006. Global reconstruction of human metabolic network based on genomic and bibliomic data. Submitted to *Proc Natl Acad Sci U.S.A*. I was the primary author of the publication and the co-authors participated and supervised the research which forms the basis for this chapter.

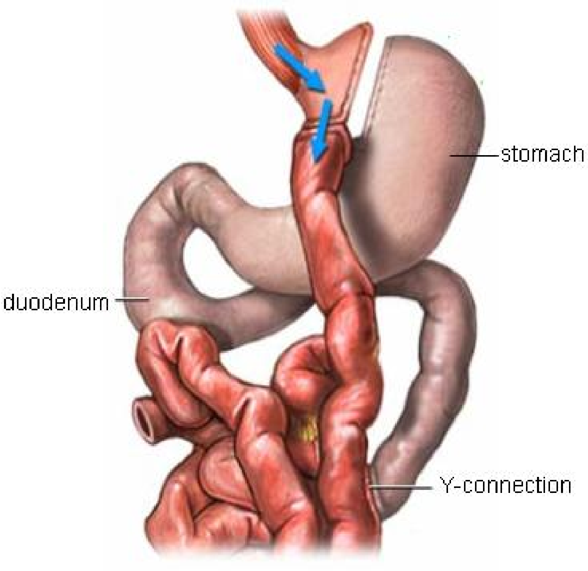


Figure 6.1: Schematic of Roux-en-Y gastric bypass surgery. A small pouch is surgically created from the stomach and connected to the jejunum, or mid-region of the small intestine. The bypassed stomach and duodenum are reattached in a Y-connection to the jejunum to allow the flow of digestive juices into the small intestine. Modified from [167].

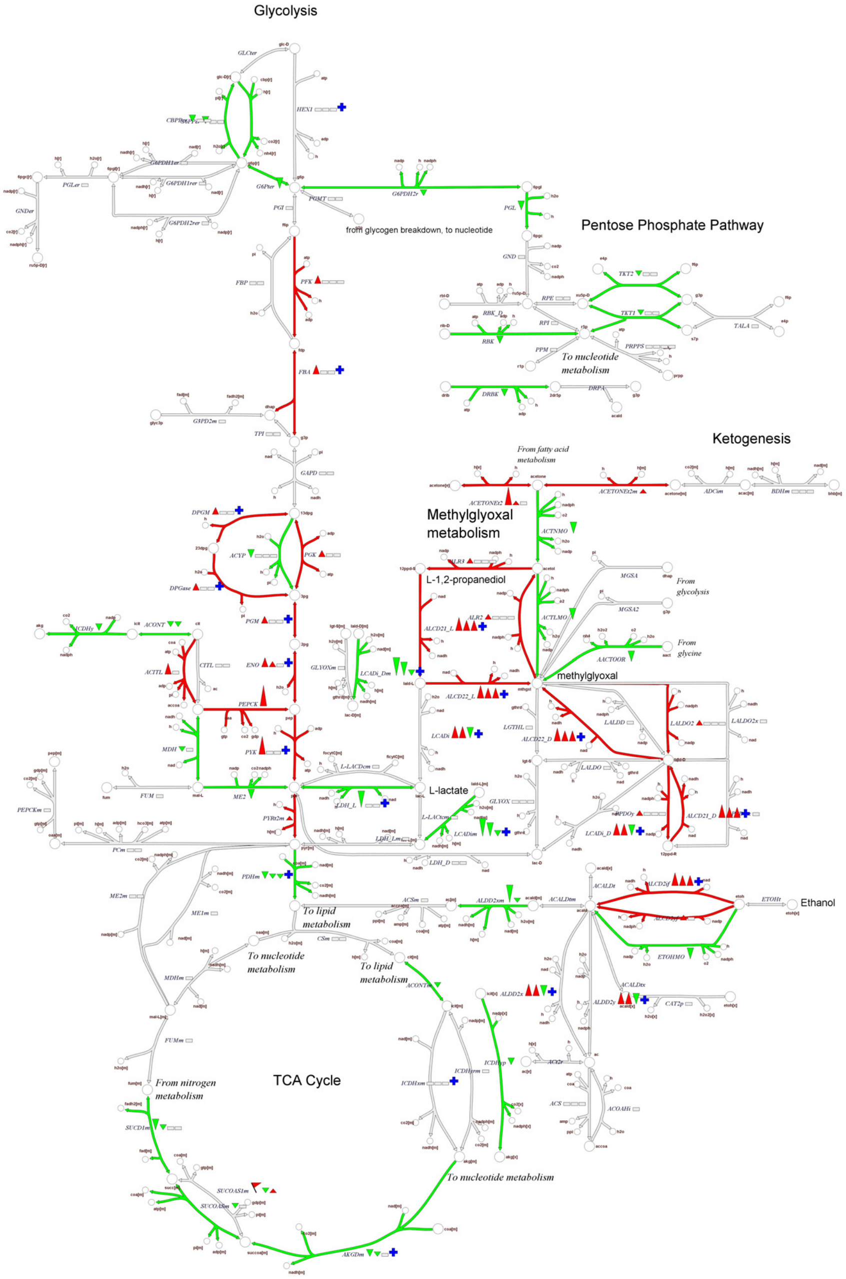


Figure 6.2: Expression changes in central metabolism post-gastric bypass. Relative gene expression levels (green – down, red – up, white – no data available or reaction level conflict) have been mapped to reactions in the global human metabolic network. Arrows next to reaction abbreviations indicate the magnitude of expression changes on a log10 scale (grey boxes indicate no data available). Genes in glycolysis, methylglyoxal metabolism, and ketogenesis were generally down-regulated, whereas those in the pentose phosphate pathway and tricarboxylic acid (TCA) cycle were generally up-regulated. Reaction and metabolite abbreviations can be found in Supplement B.

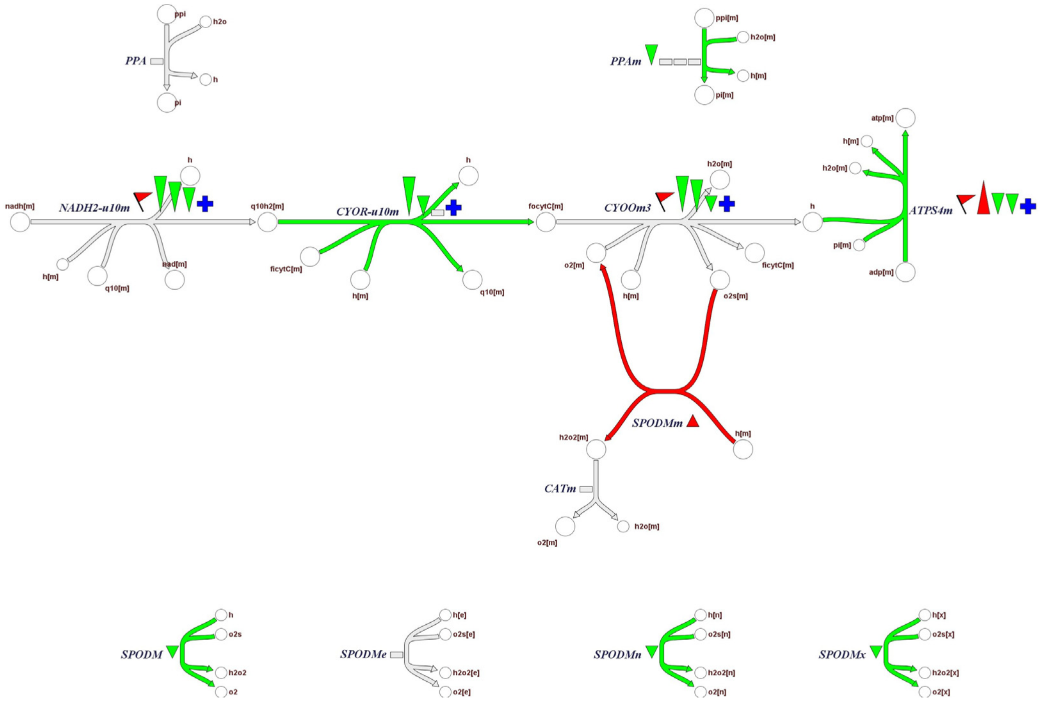


Figure 6.3: Expression changes in oxidative phosphorylation and reactive oxygen species (ROS) detoxification post-gastric bypass. The majority of the genes encoding the electron transport chain were observed to be strongly down-regulated post-surgery. Please refer to Figure 6.2 for color and symbol definitions and Supplement B for reaction and metabolite abbreviations.

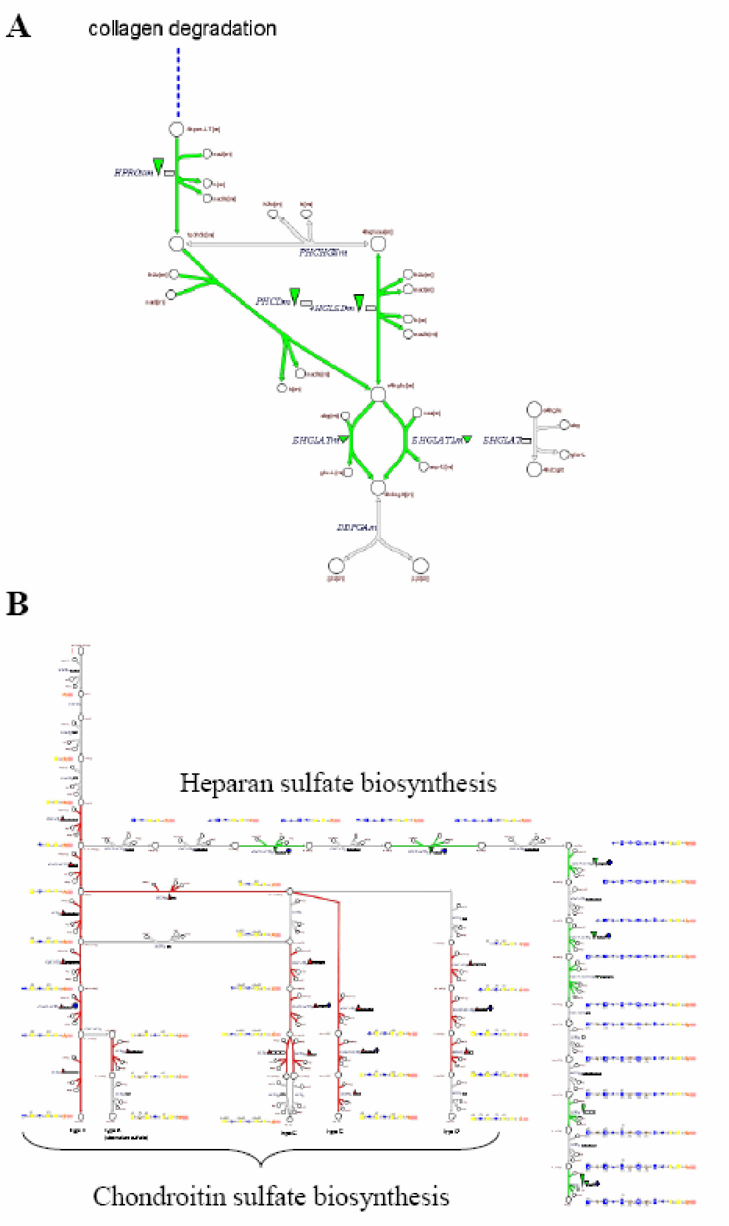


Figure 6.4: Expression changes in collagen (A) and glycosaminoglycan (B) metabolism post-gastric bypass. Expression of genes involved in collagen degradation and heparan sulfate biosynthesis decreased post-surgery, whereas those associated with chondroitin/dermatan sulfate biosynthesis increased. Other notable changes in glycan expression include down-regulation of genes involved in N-glycan biosynthesis and up-regulation of those in keratan sulfate degradation and hyaluronan biosynthesis (data not shown). Please refer to Figure 6.2 for color and symbol definitions and Supplement B for reaction and metabolite abbreviations.

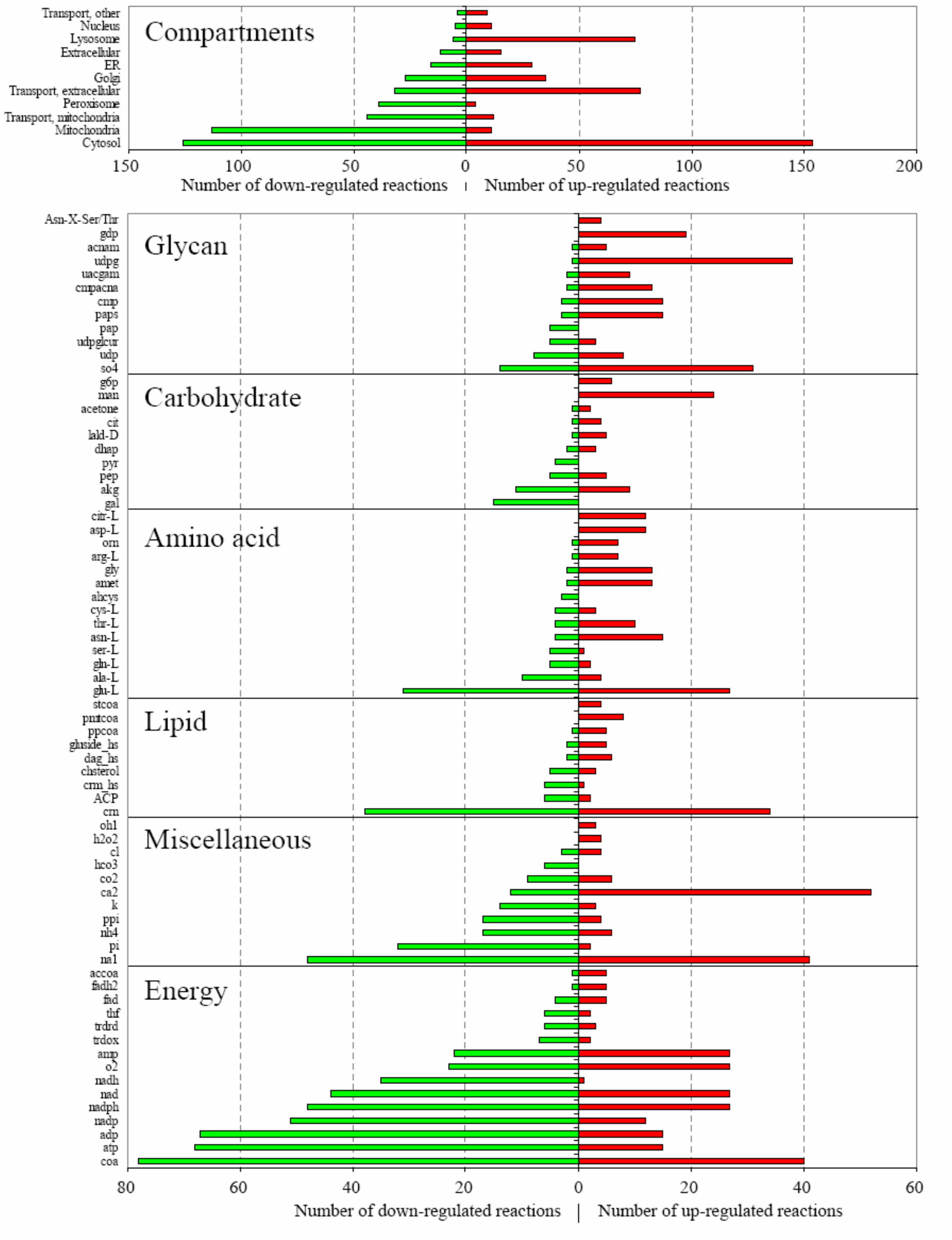


Figure 6.5: Compartmentalization and metabolite connectivity of up-regulated and down-regulated reaction networks post-gastric bypass. Most down-regulated reactions (424 total) relate to mitochondrial bioenergetics and peroxisomal oxidation whereas up-regulated reactions (432 total) reflect a shift towards amino acid-sodium co-transport and lysosomal degradation. See Supplement B for abbreviations.

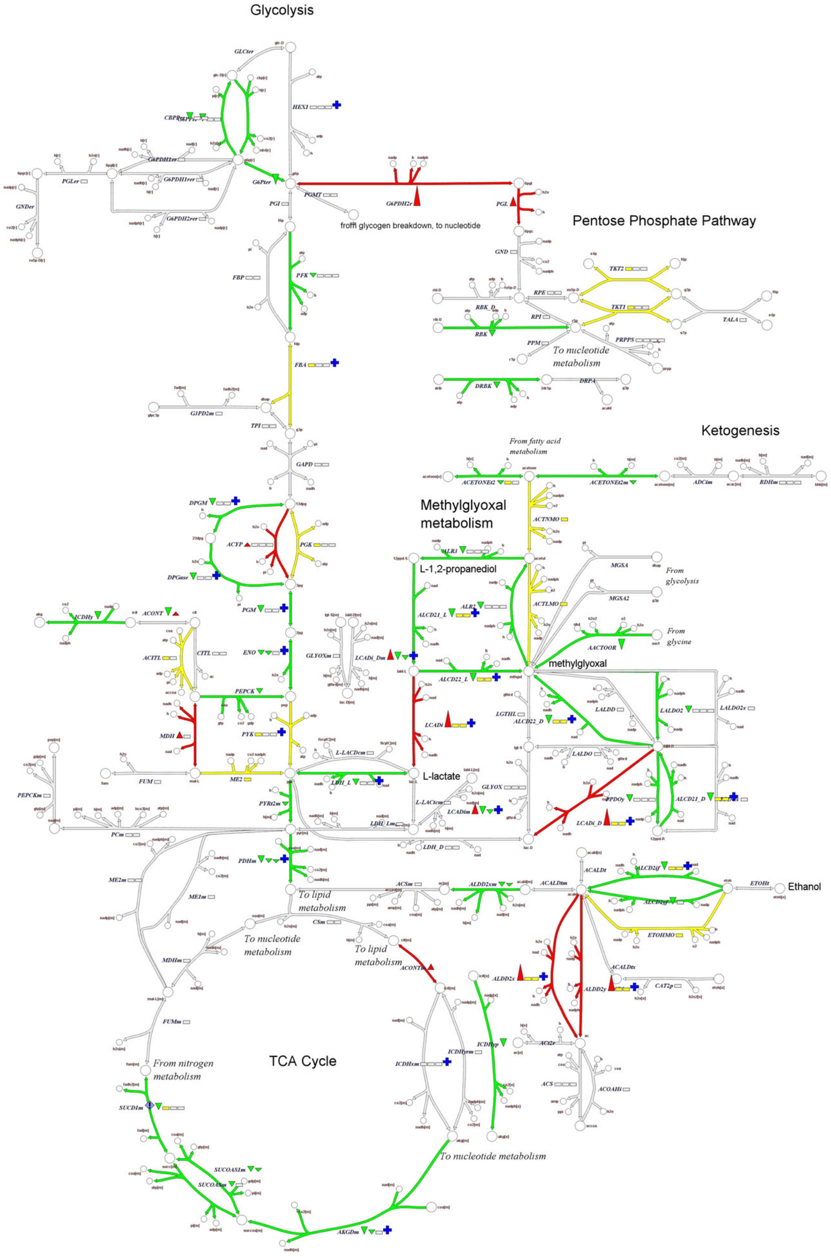


Figure 6.6: Expression differences in central metabolism in lean versus morbidly obese subjects. Relative gene expression levels (green – down, red – up, white – no data available or reaction level conflict, yellow – probe level conflict) have been mapped to reactions in the global human metabolic network. Genes involved in anaerobic metabolism were generally observed to be down-regulated in lean subjects whereas the majority of those encoding the electron transport chain were up-regulated. Please refer to Figure 6.2 for a symbol definitions and Supplement B for reaction and metabolite abbreviations.

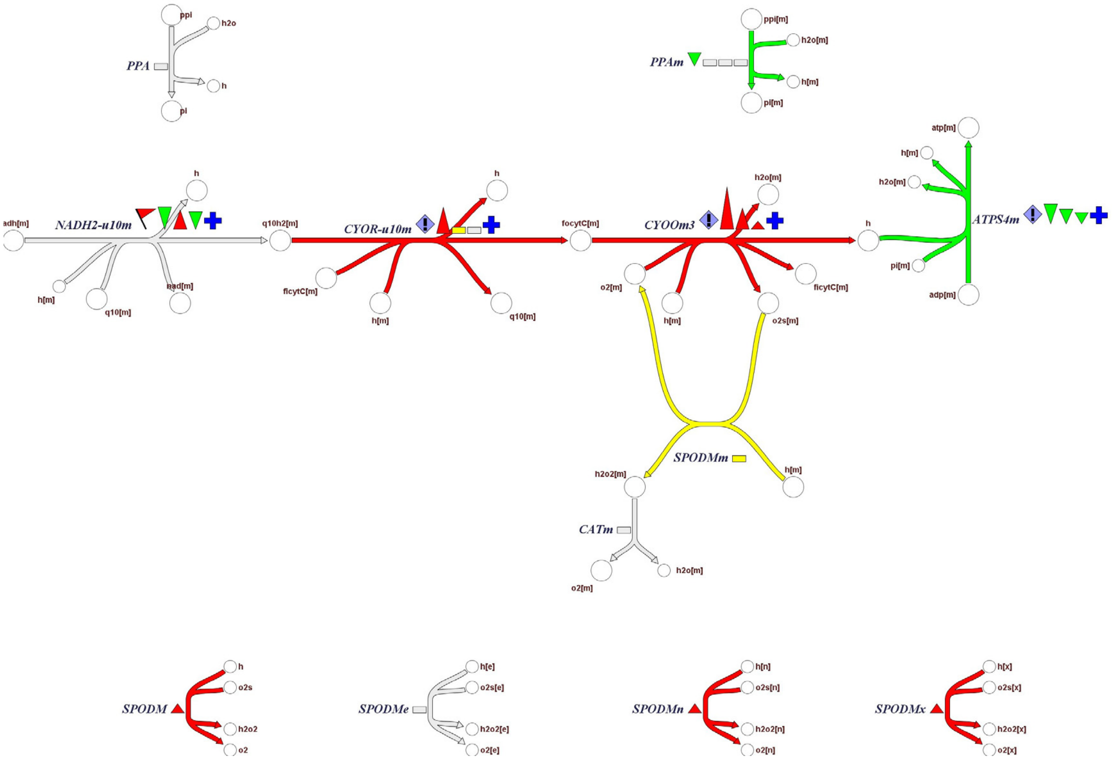


Figure 6.7: Expression differences in oxidative phosphorylation in lean versus morbidly obese subjects. Relative gene expression levels (green – down, red – up, white – no data available or reaction level conflict, yellow – probe level conflict) have been mapped to reactions in the global human metabolic network. Genes involved in anaerobic metabolism were generally observed to be down-regulated in lean subjects whereas the majority of those encoding the electron transport chain were up-regulated. Please refer to Figure 6.2 for a symbol definitions and Supplement B for reaction and metabolite abbreviations.

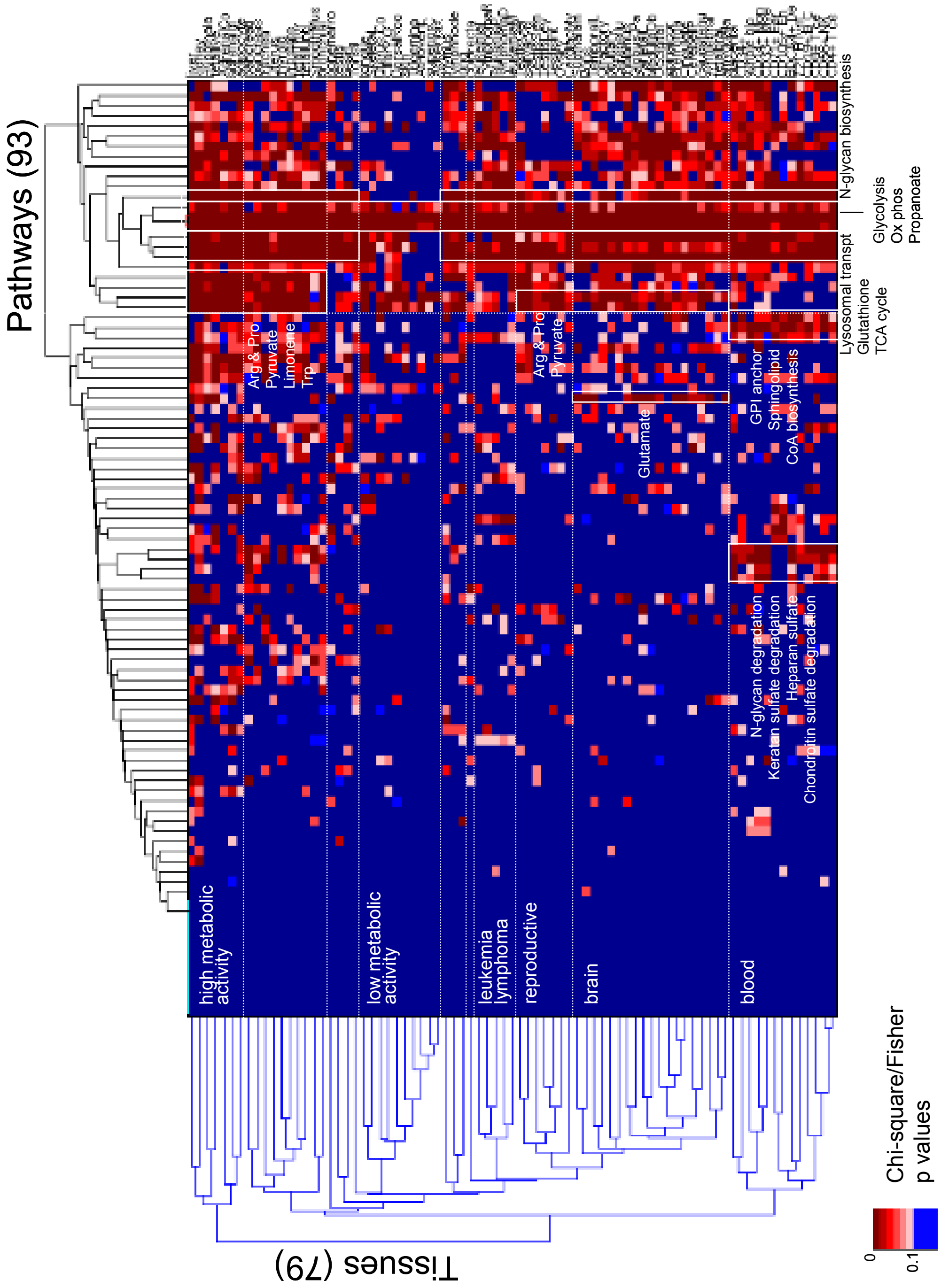


Figure 6.8: Hierarchical clustering of pathway analysis for gene expression in 79 healthy human tissues. Genes have been classified into metabolic pathways (columns) based on their annotation in *H. sapiens* Recon 1. P-values describe the likelihood that a pathway is present in a particular tissue (row) based on its distribution of present/absent calls and were calculated using a combined Chi-square/Fisher's exact test. Only p-values <0.1 are displayed.

Table 6.1: Weight classification by body mass index (BMI) [181].

BMI range	Classification
<20	Underweight
18.5-24.9	Normal
25-29.9	Overweight
30-39.9	Obese
>40	Morbidly obese

Table 6.2: Expression changes in oxidative phosphorylation for the gastric bypass study. Negative ratios indicate a decrease in expression post-surgery. Expr ratio – average log10 expression ratio (n=3), Reaction – reaction abbreviation in *H. sapiens* Recon 1 (see Supplement B).

Locus	Expr Ratio	Fold change	Gene	Putative function	Reaction
4694	-0.10133	-1.26278	NDUFA1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 1, 7.5kDa	NADH2-u10m
4696	-0.10096	-1.26171	NDUFA3	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9kDa	NADH2-u10m
4702	-0.06429	-1.15955	NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	NADH2-u10m
4708	-0.07274	-1.18234	NDUFB2	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	NADH2-u10m
4709	-0.0755	-1.18988	NDUFB3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa	NADH2-u10m
4711	-0.05262	-1.12881	NDUFB5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5, 16kDa	NADH2-u10m
4714	-0.05517	-1.13544	NDUFB8	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8, 19kDa	NADH2-u10m
4717	-0.09094	-1.23293	NDUFC1	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1, 6kDa	NADH2-u10m
4718	-0.08633	-1.21991	NDUFC2	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 2, 14.5kDa	NADH2-u10m
4719	-0.08673	-1.22105	NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	NADH2-u10m
4728	-0.21907	-1.65604	NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8, 23kDa (NADH-coenzyme Q reductase)	NADH2-u10m
4729	-0.12911	-1.34621	NDUFV2	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	NADH2-u10m
7991	-0.13311	-1.35865	TUSC3	tumor suppressor candidate 3	NADH2-u10m
51079	-0.17639	-1.50105	GRIM19	cell death-regulatory protein GRIM19	NADH2-u10m
4713	0.117673	1.311213	NDUFB7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa	NADH2-u10m

Table 6.2, continued.

Locus	Expr Ratio	Fold change	Gene	Putative function	Reaction
7381	-0.24155	-1.74401	UQCRB	ubiquinol-cytochrome c reductase binding protein	CYOR-u10m
29796	-0.10188	-1.2644	HSPC051	ubiquinol-cytochrome c reductase complex (7.2 kD)	CYOR-u10m
1327	-0.02967	-1.07071	COX4I1	cytochrome c oxidase subunit IV isoform 1	CYOOm3
1329	-0.2183	-1.65312	COX5B	cytochrome c oxidase subunit Vb	CYOOm3
1339	-0.19251	-1.55779	COX6A2	cytochrome c oxidase subunit VIa polypeptide 2	CYOOm3
1349	-0.07335	-1.18399	COX7B	cytochrome c oxidase subunit VIIb	CYOOm3
9167	0.04488	1.108869	COX7A2L	cytochrome c oxidase subunit VIIa polypeptide 2 like	CYOOm3
509	-0.03965	-1.09558	ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	ATPS4m
514	-0.10436	-1.27163	ATP5E	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit	ATPS4m
516	-0.05736	-1.1412	ATP5G1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 1	ATPS4m
517	-0.1111	-1.29152	ATP5G2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9), isoform 2	ATPS4m
518	-0.10221	-1.26534	ATP5G3	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit c (subunit 9) isoform 3	ATPS4m
521	-0.04207	-1.10171	ATP5I	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit e	ATPS4m
522	-0.09347	-1.24013	ATP5J	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit F6	ATPS4m
4905	-0.07962	-1.2012	NSF	N-ethylmaleimide-sensitive factor	ATPS4m
10632	0.201196	1.589264	ATP5L	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g	ATPS4m

Table 6.3: Expression changes in glycolysis/gluconeogenesis for the gastric bypass study. Negative ratios indicate a decrease in expression post-surgery. Expr ratio – average log10 expression ratio (n=3), Reaction – reaction abbreviation in *H. sapiens* Recon 1 (see Supplement B).

Locus	Expr Ratio	Fold change	Gene	Putative function	Reaction
5211	0.163217	1.456187	PFKL	phosphofructokinase, liver	PFK
229	0.188382	1.543057	ALDOB	aldolase B, fructose-bisphosphate	FBA2, FBA4, FBA, FBA5
5230	0.099133	1.256414	PGK1	phosphoglycerate kinase 1	PGK
5223	0.125026	1.333602	PGAM1	phosphoglycerate mutase 1 (brain)	DPGase, DPGM, PGM
2023	0.075391	1.189573	ENO1	enolase 1, (alpha)	ENO
26237	0.18623	1.535429	ENO1B	enolase alpha, lung-specific	ENO
5315	0.481154	3.027987	PKM2	pyruvate kinase, muscle	PYK
5105	0.567078	3.690443	PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)	PEPCK
98	-0.16043	-1.44689	ACYP2	acylphosphatase 2, muscle type	ACYP

Table 6.4: Expression changes in pentose phosphate pathway for the gastric bypass study. Negative ratios indicate a decrease in expression post-surgery. Expr ratio – average log10 expression ratio (n=3), Reaction – reaction abbreviation in *H. sapiens* Recon 1 (see Supplement B).

Locus	Expr Ratio	Fold change	Gene	Putative function	Reaction
2539	-0.09416	-1.24211	G6PD	glucose-6-phosphate dehydrogenase	G6PDH2r
8277	-0.06275	-1.15544	TKTL1	transketolase-like 1	TKT2, TKT1
25796	-0.19013	-1.54927	PGLS	6-phosphogluconolactonase	PGL
64080	-0.13661	-1.36964	RBKS	ribokinase	DRBK, RBK

Table 6.5: Protocol for mapping Affymetrix present/absent calls to on/off calls. Genes are only called off if they are absent in all replicates. Since replicates 1 and 2 are interchangeable, the order of the present/absent calls does not affect the on/off call.

Replicate #1	Replicate #2	Call
Absent	Absent	Off
Marginal	Absent	On
Present	Absent	On
Marginal	Marginal	On
Present	Marginal	On
Present	Present	On

CHAPTER 7: CONCLUSIONS

The time and cost of data generation was once a hindrance to biological research, limiting studies to the independent characterization of individual genes and proteins. However, with the recent explosion of high-throughput experimental methods, such as genome sequencing, gene expression profiling, and massively parallel phenotyping, we have now entered a "data rich" era that provides rapid, simultaneous analysis of thousands of cellular components. The inherent challenge of integrating and analyzing these large, comprehensive data sets is arguably best met with a systems approach [182-185].

This Dissertation encompasses three fundamental aims of systems biology as they relate to eukaryotic cellular metabolism:

1. Reconstructing comprehensive, self-consistent networks based on a variety of data types (Chapters 3 & 5).
2. Using network reconstructions as a context for large-scale data analysis (Chapter 6).
3. Formulating mathematical models for network validation (Chapter 3) and hypothesis-driven experimentation (Chapter 4).

The following sections summarize my contributions to systems biology (7.1), describe key applications of the *S. cerevisiae* iND750 and *H. sapiens* Recon 1 metabolic networks, and highlight areas of eukaryotic reconstruction that need further improvement (7.3).

7.1 Contributions to the field

While the reconstruction and modeling of prokaryotic networks is well established, especially in *Escherichia coli* [186], previous attempts at genome-scale reconstruction and analysis in eukaryotes are limited (3.1, 5.1). In this Dissertation, we have shown that fully compartmentalized, biochemically, genetically, and genomically integrated (BiGG) eukaryotic networks can be reconstructed, and demonstrate their applications in studying optimal growth behaviors (Chapter 3), predicting gene deletion phenotypes (Chapter 4), and investigating metabolic states in specific human tissues (Chapter 6). A summary of key results and conclusions is provided here.

7.1.1 Advancements in reconstructing eukaryotic metabolic networks

Reconstruction principles first introduced in Forster and Famili's original yeast network (3.1, [58]) were further developed in the assembly of *S. cerevisiae* iND750 (3.2, [57]) to include full compartmentalization, direct incorporation of gene-transcript-protein associations, and elementally and charge balanced reactions. This work subsequently led to a new generation of metabolic models that also incorporate these features [119, 187, 188], including the global human reconstruction (Chapter 5, [121]).

A novel combination of top-down and bottom-up methods (1.2) were employed in the human reconstruction project to achieve careful, quality controlled curation within a manageable time frame. This approach (5.2), in which a candidate component list generated from the genome annotation was simultaneously curated by a team of researchers, could be used as a prototype for future mammalian reconstructions in rat, cow, dog, rabbit, and chicken, whose genomes have been fully sequenced or are nearing completion [189-191].

7.1.2 Identification of metabolic knowledge gaps

Our large-scale gene deletion study (3.3, [57]) and comprehensive survey of the human metabolic knowledge base (5.3, [121]) demonstrate the utility of genome-scale reconstructions as strategic tools for discovery-driven research. For example, detailed examination of the 246 failure modes in the yeast deletion study led to 27 direct suggestions of how to potentially improve the model (Table 3.3) and specific experiments that could be performed to further improve our understanding of yeast metabolism.

Systemic assignment of reaction confidence scores in *H. sapiens* Recon 1 revealed an apparent bias in the characterization of metabolic pathways (Figure 5.9), with nearly 20% of those in Recon 1 only moderately supported by biochemical and/or genetic evidence, and another 20% largely based on physiological data or modeling assumptions alone. In addition to highlighting these “thin” areas at the reaction and pathway levels, a list of specific metabolites that require additional study was also assembled based on careful evaluation of network gaps (Supplement B).

7.1.3 Integrated analysis of metabolic phenotypes

The integration of genome-scale networks and experimental data is essential to fully characterizing an organism’s genotype-phenotype relationships. For instance, our *in silico* and *in vitro* analysis of yeast’s optimal growth patterns (Chapter 4, [93]) revealed that *S. cerevisiae* exhibits only a few dominant phenotypes over a range of glucose uptake and oxygenation conditions, which is strikingly different than behaviors observed in *E. coli* [72]. We also generated new hypotheses on the function and capacity of yeast's metabolic machinery in these states, which includes the presence of an optimal glucose-oxygen uptake ratio for maximal ethanol production during optimal growth.

The global human reconstruction and its collection of human-specific metabolic maps provided a comprehensive, quality-controlled context for visualizing and interpreting gene expression in obese skeletal muscle (6.1; [121]) and across a panel of 79 human tissues (6.2). Specifically, pathway analysis of these data suggested that gastric bypass surgery may induce transcriptional changes similar to long-term caloric restriction, and that tissues may exhibit some high-level transcriptional differences that can be used to differentiate their metabolic networks.

7.2 What's next for *S. cerevisiae* iND750 and *H. sapiens* Recon 1?

For *S. cerevisiae* iND750, additional iterations of testing and validation with experimental data will be required to continually improve its interpretive and predictive capabilities [192]. In fact, since its release in 2004, two updated metabolic models have appeared: *S. cerevisiae* iLL672 [193], which has improved predictability of gene deletion phenotypes, and *S. cerevisiae* iMM910 [194], which includes 160 additional genes and extensively validated protein and reaction localizations. A transcriptional regulatory network has also been assembled to describe the regulation of metabolic genes in *S. cerevisiae* iND750 [195].

The stoichiometric framework of *S. cerevisiae* iND750 is the basis of a variety of mathematical tools [4], and as a result, it has numerous modeling applications in the research community. For example, *S. cerevisiae* iND750 has been used to explore structure-function relationships related to gene essentiality [196-198], environmental conditions [199, 200], and optimal bioreactor design [201]. *S. cerevisiae* iND750 is also part of a growing collection of genome-scale reconstructions that have been used for comparative studies of network structure [202, 203], analysis of evolutionary network conservation [204], and as a basis for top-down assembly of new networks [205]. We anticipate that metabolic models of yeast will continue to be actively developed and expanded, and that the mathematical and experimental tools developed for yeast will lay the groundwork for future studies in higher eukaryotes, including human cells.

The next big step for *H. sapiens* Recon 1 will be to develop cell-specific models for quantitative simulations of physiological and pathophysiological metabolic states. While the possibilities seem endless, the best candidate cell types for mathematical modeling will have defined metabolic objectives and good data availability (Table 7.1). For example, while skeletal muscle, cardiac muscle, and liver all have clear metabolic functions, obtaining healthy, homogeneous samples of these tissues can be difficult. On the other hand, blood cells and cell lines can be easily sampled and effectively grown *in vitro*. However, the metabolic objectives of these cells may be limited to biomass production, and it remains to be seen whether immortalized cell lines are suitable representations of *in vivo* cells. Thus, careful selection of reconstruction targets is critical to advancing constraint-based modeling of human cells.

7.3 Improving genome-scale reconstruction and analysis of eukaryotes

Lessons learned from the reconstruction and analysis of *S. cerevisiae* iND750 and *H. sapiens* Recon 1 have led to many suggestions for improving component representations (7.3.1), community interaction (7.3.2), and the integration of gene expression data (7.3.3). A summary of these insights and their implications for future extensions of this work are described here.

7.3.1 Enhancing component representations

Comparison of metabolite compartmentalization in the yeast and human metabolic networks revealed that we are getting better at defining the unique functionalities found within these compartments (Figure 7.1). For example, in addition to the mitochondria and peroxisome, which have a significant number of unique metabolites in both the yeast and human networks, the distinct metabolic roles of the Golgi apparatus, endoplasmic reticulum, and lysosome also appear to have been captured by the human metabolic network. Furthermore, singular value decomposition [206] of the yeast and human stoichiometric matrices [121] confirmed that the introduction of compartmentalization added new, non-redundant functionalities to these networks.

So what can be done to improve our representation of intracellular compartments? A first step might be to remove the simplifying assumption that there is a constant pH of 7.2 across all compartments and reformulate metabolites so that they have compartment-specific formulae and charge. Although this could lead to potential bookkeeping and computational challenges (as the number of metabolites and reactions could each increase eight-fold), such a study might provide new insights on compartmental relationships by introducing transformations that have varying energetic costs depending on their localization.

Better biochemical characterization of intracellular compartments is also needed to improve the overall accuracy of our reconstructions. While green fluorescent protein (GFP)-tagging has been suggested for high-throughput localization assessment, our observations from a case study with three yeast mitochondrial data sets suggests that it might be premature to assign localizations using this data alone (data not shown). For membrane-bound enzymes, it is especially important to identify active site localizations to ensure proper compartmentalization of their corresponding reactions. High quality localization data may also reduce the need to add metabolite exchange reactions, which as a whole are poorly understood.

In addition to compartmentalization, another feature introduced in the yeast and human reconstructions is the systemic representation of isozymes and complexes as Boolean logic statements. For yeast, these relationships were defined based on multiple, independent genes (Figure 3.2). However, mammals, unlike yeast, are known to have high levels of alternative splicing, with recent estimates of alternative splicing in 40-60% of human genes [199-204]. Therefore, we extended definitions of isozymes and complexes in the human reconstruction to include multiple splice variants of the same gene. These alternative transcripts were identified in terms of RefSeq mRNA identifiers [9], a non-redundant set of transcript sequences provided by the National Center for Biotechnology Information.

While we found RefSeq's coverage of alternative transcripts in human genes to be limited (Figure 6.2), there were two primary reasons for choosing this identifier to define alternative transcripts in our reconstruction. First, RefSeq identifiers are highly curated and well integrated other types of data, including Affymetrix's microarrays. Second, cataloging splice variants from the literature would have been challenging (if not impossible), as their descriptions are typically not reported consistently across research groups, and sequence information is rarely available. The ambiguities present in defining alternative transcripts are reminiscent of those that once existed with enzymes [207] and eventually led to the development of Enzyme Commission numbers [17]. Unfortunately, a major obstacle in establishing a similar universal nomenclature system for transcript variants is that the extent of their cross-species conservation is unclear [208]. Thus, while comprehensive identification and classification of human alternative splicing events is greatly needed, there appears to be many logistical challenges which must be overcome before this can be accomplished.

Our representation of enzyme kinetics could also be improved in future versions of the yeast and human metabolic networks. For example, since we do not currently account for enzyme substrate and cofactor affinities, all isozymes are considered to be equivalent and can catalyze reactions at the same maximal rate *in silico*. This presents a challenge in gene expression analysis, as oftentimes isozymes are differentially regulated, leading to conflicting results when measurements are mapped to fluxes in the reaction network. Furthermore, we could account for sequence variations with known enzymatic defects by reducing maximum flux values for their corresponding reactions [209]. The inclusion of detailed kinetic data may therefore provide a more accurate link between changes in component properties and measured physiological functions.

7.3.2 Improving communication with the research community

As demonstrated by *S. cerevisiae* iND750, genome-scale reconstructions have many applications as integrated databases, structural networks, and mathematical models. Consequently, their contents need to be disseminated in a variety of formats to best serve the needs of the research community. The current BiGG Human Database (http://bigg.ucsd.edu) provides basic search/browse capabilities, cross-references gene, protein, reaction, and metabolite entries, and links to several biological databases (Table 5.7). In the future, we hope to expand the database to include other genome-scale reconstructions as well as add many more features, including:

- Community forums for network curation and feedback;
- Data repositories, advanced query tools, and technical support to facilitate information exchange;
- Searchable, integrated maps for data visualization; and
- Tools such as FBA (2.2.1) and pathway analysis (6.2) for *in silico* analysis.

Making these improvements would facilitate community-based curation of metabolic networks and provide greater access to genome-scale visualization and analysis tools.

7.3.3 Fine-tuning integration of gene expression data

As described in 7.2, a long-term goal of the human metabolic reconstruction project is to tailor the global network to specific cells, tissues, and disease states. This

Dissertation describes an approach that uses presence or absence of whole metabolic pathways to potentially refine network content (6.2). While traditional pathway definitions are subjective, overlapping, and may have poor gene coverage, a key strength of this method is that it offers tremendous flexibility in the classification of functionally related genes. Thus, future implementations could use *H. sapiens* Recon 1's network structure (*e.g.,* extreme pathways [210], correlated reaction sets [211], flux coupled reaction sets [120]) or biological modules (*e.g.,* compartments or chromosomal locations) as "pathways." In addition, gene presence or absence could alternatively be assessed by determining an appropriate cutoff expression value through statistical analysis or, instead of binary on/off calls, quantitative expression measurements such as standard deviations from the mean, or percentile rank could be used to determine genes with significant expression.

An alternative to the pathway-based approach is integrating gene expression data at the individual reaction level. While this strategy is arguably less crude than pathway analysis, it also has some inherent challenges. First, results can be confusing when mapping expression data to reactions associated with more than one gene. For example, is an enzyme present if only one of its subunits is called on? Another potential difficulty is that this approach may lead to "incomplete" networks that can no longer achieve some of their desired functionalities. Methods for handling these challenges have been proposed [212] and are currently underway in the Systems Biology Research Group at UCSD.

7.4 Concluding remarks

To paraphrase the introductory quote by Dr. Francis Collins, the human genome sequence is not simply the book of life, but a context for evolutionary analysis, a catalog of cellular components, and a basis for improved understanding and treatment of disease. Much of the work described in this thesis relates to the second application, using the human genome (as well as that of yeast) to systemically reconstruct metabolic networks. We have also made additional strides in understanding yeast physiology during optimal growth under a variety of genetic and environmental constraints. However, for human studies, the most exciting implications of this work are yet to come. Efforts are already underway to use the global network as a stepping stone for comprehensive, cell type-specific reconstructions that will enable detailed, quantitative analysis of human physiology and pathophysiology. Careful integration of genomic, transcriptomic, and metabolomic data may also yield translational technologies such as personalized simulations of drug metabolism. Finally, with the growing number of 1-D and 2-D annotations, new insights may be generated on the evolution and adaptation of metabolic networks.

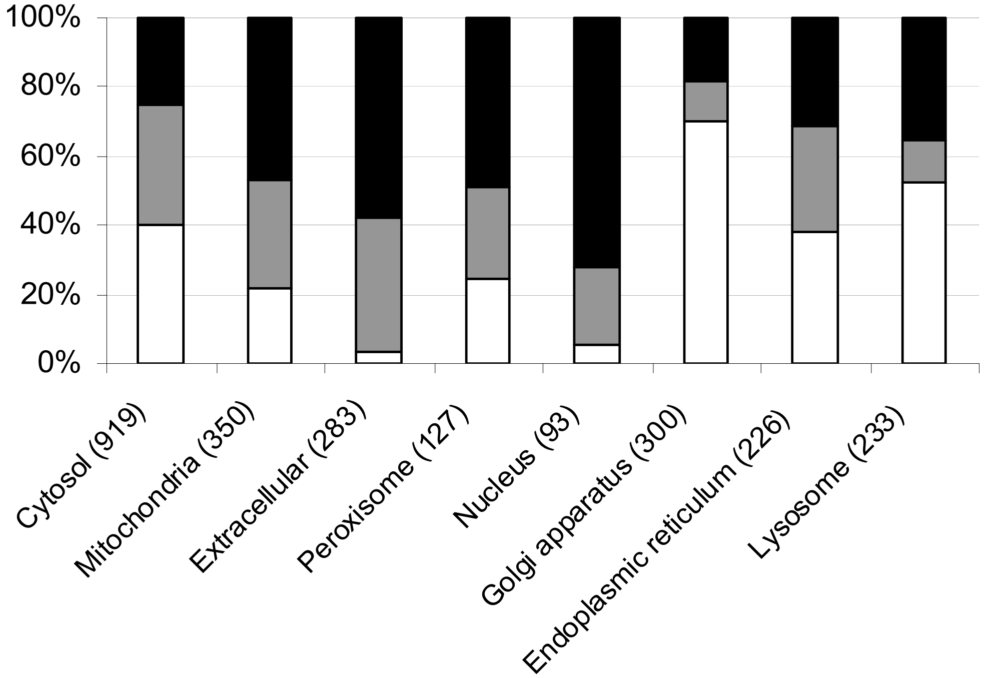


Figure 7.1: Compartmental distribution of *H. sapiens* Recon 1's metabolites. The number of metabolites found in each compartment is shaded based on its connectivity. Metabolites that are unique to a particular compartment are shown in white; metabolites found in two compartments are shaded in grey; and metabolites found in three or more compartments are shaded in black. Compare to Figure 3.3 for yeast.

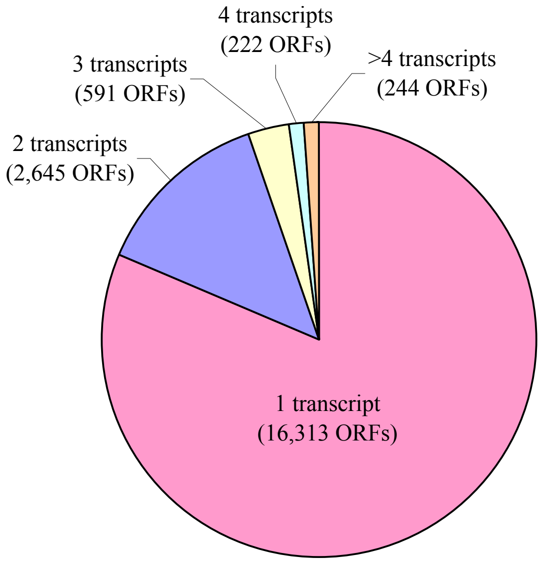


Figure 7.2: Number of transcripts per open reading frame (ORF) in the human gene index. An initial set of 20,015 human ORFs and 25,883 transcripts were identified based on the November 2004 genome annotation in LocusLink [116] and RefSeq mRNA identifiers [9] (see 5.2.2).

Table 7.1: Candidates for cell-specific human metabolic models. Legend – High quality, homogeneous data can be readily obtained (♦♦♦), data can be readily obtained, but is usually heterogeneous (♦♦), data is difficult to obtain or only available under limited conditions (♦).

Cell type	Potential objective functions	Data availability
Cardiac myocytes	ATP production	♦
Skeletal myocytes	ATP production	♦♦
Hepatocytes	Bile, cholesterol, glycogen, and urea production	♦
Cancer	Biomass production	♦♦
Cell lines	Biomass production	♦♦♦
B cell	Biomass production	♦♦♦

BIBLIOGRAPHY

1. Skyttner, L., *General systems theory : ideas & applications*. 2001, Singapore ; River Edge, N.J.: World Scientific. xii, 459 p.

2. Weinberg, G.M., *An introduction to general systems thinking / Gerald M. Weinberg*. Silver anniversary ed. 2001, New York: Dorset House. xxi, 279 p.

3. Laszlo, E., *The systems view of the world : a holistic vision for our time*. Advances in systems theory, complexity, and the human sciences. 1996, Cresskill, NJ: Hampton Press. viii, 103 p.

4. Palsson, B.O., *Systems biology: properties of reconstructed networks*. 2006, New York: Cambridge University Press.

5. Palsson, B., *Two-dimensional annotation of genomes.* Nat Biotechnol, 2004. **22**(10): p. 1218-9.

6. Reed, J.L., I. Famili, I. Thiele, and B.O. Palsson, *Towards multidimensional genome annotation.* Nat Rev Genet, 2006. **7**(2): p. 130-41.

7. Karp, P.D., S. Paley, and P. Romero, *The Pathway Tools software.* Bioinformatics, 2002. **18 Suppl 1**: p. S225-32.

8. Wheeler, D.L., D.M. Church, R. Edgar, S. Federhen, W. Helmberg, T.L. Madden, J.U. Pontius, G.D. Schuler, L.M. Schriml, E. Sequeira, T.O. Suzek, T.A. Tatusova, and L. Wagner, *Database resources of the National Center for Biotechnology Information: update.* Nucleic Acids Res, 2004. **32 Database issue**: p. D35-40.

9. Maglott, D., J. Ostell, K.D. Pruitt, and T. Tatusova, *Entrez Gene: gene-centered information at NCBI.* Nucleic Acids Res, 2005. **33**(Database issue): p. D54-8.

10. Christie, K.R., S. Weng, R. Balakrishnan, M.C. Costanzo, K. Dolinski, S.S. Dwight, S.R. Engel, B. Feierbach, D.G. Fisk, J.E. Hirschman, E.L. Hong, L. Issel-Tarver, R. Nash, A. Sethuraman, B. Starr, C.L. Theesfeld, R. Andrada, G. Binkley, Q. Dong, C. Lane, M. Schroeder, D. Botstein, and J.M. Cherry, *Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms.* Nucleic Acids Res, 2004. **32 Database issue**: p. D311-4.

11. Mewes, H.W., C. Amid, R. Arnold, D. Frishman, U. Guldener, G. Mannhaupt, M. Munsterkotter, P. Pagel, N. Strack, V. Stumpflen, J. Warfsmann, and A. Ruepp, *MIPS: analysis and annotation of proteins from whole genomes.* Nucleic Acids Res, 2004. **32 Database issue**: p. D41-4.

12. Safran, M., I. Solomon, O. Shmueli, M. Lapidot, S. Shen-Orr, A. Adato, U. Ben-Dor, N. Esterman, N. Rosen, I. Peter, T. Olender, V. Chalifa-Caspi, and D. Lancet, *GeneCards 2002: towards a complete, object-oriented, human gene compendium.* Bioinformatics, 2002. **18**(11): p. 1542-3.

13. Imanishi, T., T. Itoh, Y. Suzuki, C. O'Donovan, S. Fukuchi, K.O. Koyanagi, R.A. Barrero, T. Tamura, Y. Yamaguchi-Kabata, M. Tanino, K. Yura, S. Miyazaki, K. Ikeo, K. Homma, A. Kasprzyk, T. Nishikawa, M. Hirakawa, J. Thierry-Mieg, D. Thierry-Mieg, J. Ashurst, L. Jia, M. Nakao, M.A. Thomas, N. Mulder, Y. Karavidopoulou, L. Jin, S. Kim, T. Yasuda, B. Lenhard, E. Eveno, C. Yamasaki, J. Takeda, C. Gough, P. Hilton, Y. Fujii, H. Sakai, S. Tanaka, C. Amid, M. Bellgard, F. Bonaldo Mde, H. Bono, S.K. Bromberg, A.J. Brookes, E. Bruford, P. Carninci, C. Chelala, C. Couillault, S.J. de Souza, M.A. Debily, M.D. Devignes, I. Dubchak, T. Endo, A. Estreicher, E. Eyras, K. Fukami-Kobayashi, G.R. Gopinath, E. Graudens, Y. Hahn, M. Han, Z.G. Han, K. Hanada, H. Hanaoka, E. Harada, K. Hashimoto, U. Hinz, M. Hirai, T. Hishiki, I. Hopkinson, S. Imbeaud, H. Inoko, A. Kanapin, Y. Kaneko, T. Kasukawa, J. Kelso, P. Kersey, R. Kikuno, K. Kimura, B. Korn, V. Kuryshev, I. Makalowska, T. Makino, S. Mano, R. Mariage-Samson, J. Mashima, H. Matsuda, H.W. Mewes, S. Minoshima, K. Nagai, H. Nagasaki, N. Nagata, R. Nigam, O. Ogasawara, O. Ohara, M. Ohtsubo, N. Okada, T. Okido, S. Oota, M. Ota, T. Ota, T. Otsuki, D. Piatier-Tonneau, A. Poustka, S.X. Ren, N. Saitou, K. Sakai, S. Sakamoto, R. Sakate, I. Schupp, F. Servant, S. Sherry, R. Shiba, N. Shimizu, M. Shimoyama, A.J. Simpson, B. Soares, C. Steward, M. Suwa, M. Suzuki, A. Takahashi, G. Tamiya, H. Tanaka, T. Taylor, J.D. Terwilliger, P. Unneberg, V. Veeramachaneni, S. Watanabe, L. Wilming, N. Yasuda, H.S. Yoo, M. Stodolsky, W. Makalowski, M. Go, K. Nakai, T. Takagi, M. Kanehisa, Y. Sakaki, J. Quackenbush, Y. Okazaki, Y. Hayashizaki, W. Hide, R. Chakraborty, K. Nishikawa, H. Sugawara, Y. Tateno, Z. Chen, M. Oishi, P. Tonellato, R. Apweiler, K. Okubo, L. Wagner, S. Wiemann, R.L. Strausberg, T. Isogai, C. Auffray, N. Nomura, T. Gojobori and S. Sugano, *Integrative annotation of 21,037 human genes validated by full-length cDNA clones.* PLoS Biol, 2004. **2**(6): p. 856-75.

14. Kanehisa, M., S. Goto, M. Hattori, K.F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa, *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Res, 2006. **34**(Database issue): p. D354-7.

15. Romero, P., J. Wagg, M.L. Green, D. Kaiser, M. Krummenacker, and P.D. Karp, *Computational prediction of human metabolic pathways from the complete human genome.* Genome Biol, 2005. **6**(1): p. R2.

16. Schomburg, I., A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg, *BRENDA, the enzyme database: updates and major new developments.* Nucleic Acids Res, 2004. **32**(Database issue): p. D431-3.

17. *Enzyme nomenclature 1992 : IUB- recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*, ed. E.C. Webb. 1992, San Diego: Academic Press. 862.

18. *The Gene Ontology (GO) project in 2006.* Nucleic Acids Res, 2006. **34**(Database issue): p. D322-6.

19. Stryer, L., *Biochemistry*. 3rd ed. 1988, New York: W.H. Freeman. xxxii, 1089.

20. Voet, D., Voet, J. G., Pratt, C. W., *Fundamentals of Biochemistry*. 1999, New York: Wiley.

21. Michal, G., *Biochemical pathways : an atlas of biochemistry and molecular biology*. English language ed. 1999, New York, Heidelberg: Wiley ; Spektrum. xi, 277.

22. Salway, J.G., *Metabolism at a glance*. 2nd ed. 1999, Oxford ; Malden, MA: Blackwell Science. 111.

23. Dickinson, J.R. and M. Schweizer, *The metabolism and molecular physiology of Saccharomyces cerevisiae*. 1999, London ; Philadelphia: Taylor & Francis. xii, 343.

24. Rose, A.H. and J.S. Harrison, *Metabolism and physiology of yeasts*. 2nd ed. The Yeasts ; v. 3. 1989, London ; San Diego: Academic Press. xxiv, 635.

25. Strathern, J.N., E.W. Jones, and J.R. Broach, *The Molecular biology of the yeast Saccharomyces : metabolism and gene expression*. Cold Spring Harbor monograph series ; [11B]. 1982, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory. x, 680.

26. Walker, G.M., *Yeast physiology and biotechnology*. 1998, Chichester ; New York: J. Wiley & Sons. ix, 350.

27. Orten, J.M., *Human Biochemistry*. 1975, C. V. Mosby.

28. Devlin, T.M., *Textbook of biochemistry : with clinical correlations*. 5th ed. 2002, New York: Wiley-Liss. xxiv, 1216.

29. Frisell, W.R., *Human biochemistry*. 1982, New York: Macmillan. ix, 845 p.

30. Garrett, R. and C.M. Grisham, *Principles of biochemistry : with a human focus*. 2002, Fort Worth: Harcourt College Publishers. 1 v. (various pagings).

31. Price, N.D., J.L. Reed, and B.O. Palsson, *Genome-scale models of microbial cells: evaluating the consequences of constraints.* Nat Rev Microbiol, 2004. **2**(11): p. 886-897.

32. Price, N.D., J.A. Papin, C.H. Schilling, and B. Palsson, *Genome-scale microbial in silico models: the constraints-based approach.* Trends in Biotechnology, 2003. **21**(4): p. 162-169.

33. Covert, M.W., I. Famili, and B.O. Palsson, *Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology?* Biotechnol Bioeng, 2003. **84**(7): p. 763-72.

34. Varma, A. and B.O. Palsson, *Metabolic Flux Balancing: Basic concepts, Scientific and Practical Use.* Bio/Technology, 1994. **12**: p. 994-998.

35. Bonarius, H.P.J., G. Schmid, and J. Tramper, *Flux analysis of underdetermined metabolic networks: The quest for the missing constraints.* Trends in Biotechnology, 1997. **15**(8): p. 308-314.

36. Edwards, J.S., R. Ramakrishna, C.H. Schilling, and B.O. Palsson, *Metabolic Flux Balance Analysis*, in *Metabolic Engineering*, S.Y. Lee and E.T. Papoutsakis, Editors. 1999, Marcel Deker.

37. Schuster, S., T. Dandekar, and D.A. Fell, *Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering.* Trends Biotechnol, 1999. **17**(2): p. 53-60.

38. Edwards, J.S., M. Covert, and B. Palsson, *Metabolic modeling of microbes: the flux-balance approach.* Environmental Microbiology, 2002. **4**(3): p. 133-40.

39. Covert, M.W., C.H. Schilling, I. Famili, J.S. Edwards, I.I. Goryanin, E. Selkov, and B.O. Palsson, *Metabolic modeling of microbial strains in silico.* Trends Biochem. Sci., 2001. **26**: p. 179-186.

40. Horn, F. and R. Jackson, *General mass action kinetics.* Arch. Rational Mech. Anal., 1972. **47**: p. 81-116.

41. Reich, J.G. and E.E. Sel'kov, *Energy metabolism of the cell : a theoretical treatise*. 1981, London ; New York: Academic Press. viii, 345.

42. Lee, S., C. Phalakornkule, M.M. Domach, and I.E. Grossmann, *Recursive MILP model for finding all the alternate optima in LP models for metabolic networks.* Comp Chem Eng, 2000. **24**: p. 711-16.

43. Mahadevan, R. and C.H. Schilling, *The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.* Metab Eng, 2003. **5**(4): p. 264-76.

44. Reed, J.L. and B.O. Palsson, *Genome-Scale In Silico Models of E. coli Have Multiple Equivalent Phenotypic States: Assessment of Correlated Reaction Subsets That Comprise Network States.* Genome Res, 2004. **14**(9): p. 1797-805.

45. Edwards, J.S., R.U. Ibarra, and B.O. Palsson, *In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.* Nature Biotechnology, 2001. **19**: p. 125-130.

46. Fong, S.S., J.Y. Marciniak, and B.Ø. Palsson, *Description and Interpretation of Adaptive Evolution of Escherichia coli K-12 MG1655 Using a Genome-*

scale in silico Metabolic Model. Journal of Bacteriology, 2003. **185**(21): p. 6400-8.

47. Ibarra, R.U., J.S. Edwards, and B.O. Palsson, *Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth.* Nature, 2002. **420**(6912): p. 186-9.

48. Edwards, J.S. and B.O. Palsson, *The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities.* Proceedings of the National Academy of Sciences, 2000. **97**(10): p. 5528-5533.

49. Price, N.D., J.L. Reed, J.A. Papin, S.J. Wiback, and B.O. Palsson, *Network-based Analysis of Metabolic Regulation in the Human Red Blood Cell.* Journal of Theoretical Biology, 2003. **225**(2): p. 185-194.

50. Schilling, C.H., M.W. Covert, I. Famili, G.M. Church, J.S. Edwards, and B.O. Palsson, *Genome-scale metabolic model of Helicobacter pylori 26695.* Journal of Bacteriology, 2002. **184**(16): p. 4582-4593.

51. Edwards, J.S. and B.O. Palsson, *Systems properties of the Haemophilus influenzae Rd metabolic genotype.* Journal of Biological Chemistry, 1999. **274**(25): p. 17410-6.

52. Edwards, J.S., and Palsson, B.O., *Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. BMC Bioinformatics*, 2000. **1/1**.

53. Forster, J., I. Famili, B.O. Palsson, and J. Nielsen, *Large-scale evaluation of in silico gene knockouts in Saccharomyces cerevisiae.* Omics, 2003. **7**(2): p. 193-202.

54. Raamsdonk, L.M., B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M.C. Walsh, J.A. Berden, K.M. Brindle, D.B. Kell, J.J. Rowland, H.V. Westerhoff, K. van Dam, and S.G. Oliver, *A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations.* Nat Biotechnol, 2001. **19**(1): p. 45-50.

55. Lee, T.I., N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, and R.A. Young, *Transcriptional regulatory networks in Saccharomyces cerevisiae.* Science, 2002. **298**(5594): p. 799-804.

56. Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E.S. Lander, *Sequencing and comparison of yeast species to identify genes and regulatory elements.* Nature, 2003. **423**(6937): p. 241-54.

57. 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.

58. Forster, J., I. Famili, P.C. Fu, B.O. Palsson, and J. Nielsen, *Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network.* Genome Research, 2003. **13**(2): p. 244-53.

59. Nissen, T.L., U. Schulze, J. Nielsen, and J. Villadsen, *Flux distributions in anaerobic, glucose-limited continuous cultures of Saccharomyces cerevisiae.* Microbiology, 1997. **143**(Pt 1): p. 203-18.

60. Ostergaard, S., L. Olsson, and J. Nielsen, *In vivo dynamics of galactose metabolism in Saccharomyces cerevisiae: metabolic fluxes and metabolite levels.* Biotechnol Bioeng, 2001. **73**(5): p. 412-25.

61. Vanrolleghem, P.A., P. De Jong-Gubbels, W.H. Van Gulik, J.T. Pronk, J.P. Van Dijken, and S. Heijnen, *Validation of a metabolic network for Saccharomyces cerevisiae using mixed substrate studies.* Biotechnology Progress, 1996. **12**(4): p. 434-448.

62. Visser, D., R. van der Heijden, K. Mauch, M. Reuss, and S. Heijnen, *Tendency modeling: a new approach to obtain simplified kinetic models of metabolism applied to Saccharomyces cerevisiae.* Metab Eng, 2000. **2**(3): p. 252-75.

63. Lei, F., M. Rotboll, and S.B. Jorgensen, *A biochemically structured model for Saccharomyces cerevisiae.* J Biotechnol, 2001. **88**(3): p. 205-21.

64. Hynne, F., S. Dano, and P.G. Sorensen, *Full-scale model of glycolysis in Saccharomyces cerevisiae.* Biophys Chem, 2001. **94**(1-2): p. 121-63.

65. Rizzi, M., M. Baltes, U. Theobald, and M. Reuss, *In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae .2. Mathematical model.* Biotechnology and Bioengineering, 1997. **55**(4): p. 592-608.

66. Vaseghi, S., A. Baumeister, M. Rizzi, and M. Reuss, *In vivo Dynamics of the pentose phosphate pathway in Saccharomyces cerevisiae.* Metabolic Engineering, 1999. **1**: p. 128-140.

67. Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S.G. Oliver, *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.

68. Famili, I., J. Forster, J. Nielsen, and B.O. Palsson, *Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13134-9.

69. Weng, S., Q. Dong, R. Balakrishnan, K. Christie, M. Costanzo, K. Dolinski, S.S. Dwight, S. Engel, D.G. Fisk, E. Hong, L. Issel-Tarver, A. Sethuraman, C.

Theesfeld, R. Andrada, G. Binkley, C. Lane, M. Schroeder, D. Botstein, and J. Michael Cherry, *Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins.* Nucleic Acids Res, 2003. **31**(1): p. 216-8.

70. Mewes, H.W., D. Frishman, U. Guldener, G. Mannhaupt, K. Mayer, M. Mokrejs, B. Morgenstern, M. Munsterkotter, S. Rudd, and B. Weil, *MIPS: a database for genomes and protein sequences.* Nucleic Acids Res., 2002. **30**(1): p. 31-4.

71. Allen, T.D., J.M. Cronshaw, S. Bagley, E. Kiseleva, and M.W. Goldberg, *The nuclear pore complex: mediator of translocation between nucleus and cytoplasm.* J Cell Sci, 2000. **113 (Pt 10)**: p. 1651-9.

72. Reed, J.L., T.D. Vo, C.H. Schilling, and B.O. Palsson, *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).* Genome Biology, 2003. **4**(9): p. R54.1-R54.12.

73. Steinmetz, L.M., C. Scharfe, A.M. Deutschbauer, D. Mokranjac, Z.S. Herman, T. Jones, A.M. Chu, G. Giaever, H. Prokisch, P.J. Oefner, and R.W. Davis, *Systematic screen for human disease genes in yeast.* Nat Genet, 2002. **22**: p. 22.

74. Giaever, G., A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A.P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.D. Entian, P. Flaherty, F. Foury, D.J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J.H. Hegemann, S. Hempel, Z. Herman, D.F. Jaramillo, D.E. Kelly, S.L. Kelly, P. Kotter, D. LaBonte, D.C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S.L. Ooi, J.L. Revuelta, C.J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D.D. Shoemaker, S. Sookhai-Mahadeo, R.K. Storms, J.N. Strathern, G. Valle, M. Voet, G. Volckaert, C.Y. Wang, T.R. Ward, J. Wilhelmy, E.A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J.D. Boeke, M. Snyder, P. Philippsen, R.W. Davis, and M. Johnston, *Functional profiling of the Saccharomyces cerevisiae genome.* Nature, 2002. **418**(6896): p. 387-91.

75. Mannella, C.A., *The 'ins' and 'outs' of mitochondrial membrane channels.* Trends Biochem Sci, 1992. **17**(8): p. 315-20.

76. Vander Heiden, M.G., N.S. Chandel, X.X. Li, P.T. Schumacker, M. Colombini, and C.B. Thompson, *Outer mitochondrial membrane permeability can regulate coupled respiration and cell survival.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4666-71.

77. Outten, C.E. and V.C. Culotta, *A novel NADH kinase is the mitochondrial source of NADPH in Saccharomyces cerevisiae.* Embo J, 2003. **22**(9): p. 2015-24.

78. Parks, L.W., *Metabolism of sterols in yeast.* CRC Crit Rev Microbiol, 1978. **6**(4): p. 301-41.

79. Wieczorke, R., S. Krampe, T. Weierstall, K. Freidel, C.P. Hollenberg, and E. Boles, *Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in Saccharomyces cerevisiae.* FEBS Lett, 1999. **464**(3): p. 123-8.

80. Kispal, G., H. Steiner, D.A. Court, B. Rolinski, and R. Lill, *Mitochondrial and cytosolic branched-chain amino acid transaminases from yeast, homologs of the myc oncogene-regulated Eca39 protein.* J Biol Chem, 1996. **271**(40): p. 24458-64.

81. Covert, M.W. and B.O. Palsson, *Transcriptional Regulation in Constraints-based Metabolic Models of Escherichia coli.* J. Biol. Chem., 2002. **277**(31): p. 28058-64.

82. Bayly, A.M. and I.G. Macreadie, *Cytotoxicity of dihydropteroate in Saccharomyces cerevisiae.* FEMS Microbiol Lett, 2002. **213**(2): p. 189-92.

83. Thomas, D., R. Barbey, and Y. Surdin-Kerjan, *Gene-enzyme relationship in the sulfate assimilation pathway of Saccharomyces cerevisiae. Study of the 3'-phosphoadenylylsulfate reductase structural gene.* J Biol Chem, 1990. **265**(26): p. 15518-24.

84. Zhang, J., J. Reddy, B. Buckland, and R. Greasham, *Toward consistent and productive complex media for industrial fermentations: Studies on yeast extract for a recombinant yeast fermentation process.* Biotechnol Bioeng, 2003. **82**(6): p. 640-52.

85. Allen, T.E., M.J. Herrgard, M. Liu, Y. Qiu, J.D. Glasner, F.R. Blattner, and B.O. Palsson, *Genome-scale analysis of the uses of the Escherichia coli genome: model-driven analysis of heterogeneous data sets.* J Bacteriol, 2003. **185**(21): p. 6392-9.

86. Allen, T.E. and B.O. Palsson, *Sequenced-Based Analysis of Metabolic Demands for Protein Synthesis in Prokaryotes.* Journal of Theoretical Biology, 2003. **220**(1): p. 1-18.

87. Lievers, K.J., L.A. Kluijtmans, and H.J. Blom, *Genetics of hyperhomocysteinaemia in cardiovascular disease.* Ann Clin Biochem, 2003. **40**(Pt 1): p. 46-59.

88. Mattson, M.P. and F. Haberman, *Folate and homocysteine metabolism: therapeutic targets in cardiovascular and neurodegenerative disorders.* Curr Med Chem, 2003. **10**(19): p. 1923-9.

89. Edwards, J.S., R. Ramakrishna, and B.O. Palsson, *Characterizing the metabolic phenotype: a phenotype phase plane analysis.* Biotechnol Bioeng, 2002. **77**(1): p. 27-36.

90. Cysewski, G.R. and C.R. Wilke, *Rapid ethanol fermentations using vacuum and cell recycle.* Biotechnology and Bioengineering, 1977. **19**(8): p. 1125-1144.

91. Grosz, R. and G. Stephanopoulos, *Physiological, Biochemical, and Mathematical Studies of Micro-Aerobic Continuous Ethanol Fermentation By Saccharomyces-Cerevisiae .1. Hysteresis, Oscillations, and Maximum Specific Ethanol Productivities in Chemostat Culture.* Biotechnology and Bioengineering, 1990. **36**(10): p. 1006-1019.

92. Nishizawa, Y., I.J. Dunn, and J.R. Bourne. *The influence of oxygen and glucose on anaerobic ethanol production in continuous cultivation of microorganisms*. in *Proc of the 7th Symp.* 1980. Prague.

93. Duarte, N.C., B.O. Palsson, and P. Fu, *Integrated analysis of metabolic phenotypes in Saccharomyces cerevisiae.* BMC Genomics, 2004. **5**(1): p. 63.

94. Waterston, R.H., K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S.E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M.R. Brent, D.G. Brown, S.D. Brown, C. Bult, J. Burton, J. Butler, R.D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A.T. Chinwalla, D.M. Church, M. Clamp, C. Clee, F.S. Collins, L.L. Cook, R.R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K.D. Delehaunty, J. Deri, E.T. Dermitzakis, C. Dewey, N.J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D.M. Dunn, S.R. Eddy, L. Elnitski, R.D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G.A. Fewell, P. Flicek, K. Foley, W.N. Frankel, L.A. Fulton, R.S. Fulton, T.S. Furey, D. Gage, R.A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T.A. Graves, E.D. Green, S. Gregory, R. Guigo, M. Guyer, R.C. Hardison, D. Haussler, Y. Hayashizaki, L.W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D.B. Jaffe, L.S. Johnson, M. Jones, T.A. Jones, A. Joy, M. Kamal, E.K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W.J. Kent, A. Kirby, D.L. Kolbe, I. Korf, R.S. Kucherlapati, E.J. Kulbokas, D. Kulp, T. Landers, J.P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D.R. Maglott, E.R. Mardis, L. Matthews, E. Mauceli, J.H. Mayer, M. McCarthy, W.R. McCombie, S. McLaren, K. McLay, J.D. McPherson, J. Meldrim, B. Meredith, J.P. Mesirov, W. Miller, T.L. Miner, E. Mongin, K.T. Montgomery, M. Morgan, R. Mott, J.C. Mullikin, D.M. Muzny, W.E. Nash, J.O. Nelson, M.N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M.J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K.H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C.S. Pohl, A. Poliakov, T.C. Ponce, C.P. Ponting, S. Potter, M. Quail, A. Reymond, B.A. Roe, K.M. Roskin, E.M. Rubin, A.G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M.S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J.B. Singer, G. Slater, A. Smit, D.R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J.P. Vinson, A.C. Von Niederhausern, C.M. Wade, M. Wall, R.J. Weber, R.B. Weiss, M.C. Wendl, A.P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J.

Wierzbowski, D. Willey, S. Williams, R.K. Wilson, E. Winter, K.C. Worley, D. Wyman, S. Yang, S.P. Yang, E.M. Zdobnov, M.C. Zody and E.S. Lander, *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.

95. *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

96. Lee, K., F. Berthiaume, G.N. Stephanopoulos, D.M. Yarmush, and M.L. Yarmush, *Metabolic flux analysis of postburn hepatic hypermetabolism.* Metab Eng, 2000. **2**(4): p. 312-27.

97. Arai, K., K. Lee, F. Berthiaume, R.G. Tompkins, and M.L. Yarmush, *Intrahepatic amino acid and glucose metabolism in a D-galactosamine-induced rat liver failure model.* Hepatology, 2001. **34**(2): p. 360-71.

98. Lee, K., F. Berthiaume, G.N. Stephanopoulos, and M.L. Yarmush, *Profiling of dynamic changes in hypermetabolic livers.* Biotechnol Bioeng, 2003. **83**(4): p. 400-15.

99. Calik, P. and A. Akbay, *Mass flux balance-based model and metabolic flux analysis for collagen synthesis in the fibrogenesis process of human liver.* Med Hypotheses, 2000. **55**(1): p. 5-14.

100. Beard, D.A. and H. Qian, *Thermodynamic-Based Computational Profiling of Cellular Regulatory Control in Hepatocyte Metabolism.* Am J Physiol Endocrinol Metab, 2004.

101. Chan, C., F. Berthiaume, K. Lee, and M.L. Yarmush, *Metabolic flux analysis of cultured hepatocytes exposed to plasma.* Biotechnol Bioeng, 2003. **81**(1): p. 33-49.

102. Chan, C., F. Berthiaume, K. Lee, and M.L. Yarmush, *Metabolic flux analysis of hepatocyte function in hormone- and amino acid-supplemented plasma.* Metab Eng, 2003. **5**(1): p. 1-15.

103. Jeneson, J.A., H.V. Westerhoff, and M.J. Kushmerick, *A metabolic control analysis of kinetic controls in ATP free energy metabolism in contracting skeletal muscle.* Am J Physiol Cell Physiol, 2000. **279**(3): p. C813-32.

104. Banta, S., T. Yokoyama, F. Berthiaume, and M.L. Yarmush, *Quantitative effects of thermal injury and insulin on the metabolism of the skeletal muscle using the perfused rat hindquarter preparation.* Biotechnol Bioeng, 2004.

105. Lambeth, M.J. and M.J. Kushmerick, *A computational model for glycogenolysis in skeletal muscle.* Ann Biomed Eng, 2002. **30**(6): p. 808-27.

106. Vicini, P. and M.J. Kushmerick, *Cellular energetics analysis by a mathematical model of energy balance: estimation of parameters in human skeletal muscle.* Am J Physiol Cell Physiol, 2000. **279**(1): p. C213-24.

107. Cortassa, S., M.A. Aon, E. Marban, R.L. Winslow, and B. O'Rourke, *An integrated model of cardiac mitochondrial energy metabolism and calcium dynamics.* Biophys J, 2003. **84**(4): p. 2734-55.

108. Vo, T.D., H.J. Greenberg, and B.O. Palsson, *Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data.* J Biol Chem, 2004. **279**(38): p. 39532-40.

109. Nadeau, I., J. Sabatie, M. Koehl, M. Perrier, and A. Kamen, *Human 293 cell metabolism in low glutamine-supplied culture: interpretation of metabolic changes through metabolic flux analysis.* Metab Eng, 2000. **2**(4): p. 277-92.

110. Joshi, A. and B.O. Palsson, *Metabolic dynamics in the human red cell. Part I--A comprehensive kinetic model.* Journal of Theoretical Biology, 1989. **141**(4): p. 515-28.

111. Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N.N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K.V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M.

Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu, *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

112. Miller, W., K.D. Makova, A. Nekrutenko, and R.C. Hardison, *Comparative genomics.* Annu Rev Genomics Hum Genet, 2004. **5**: p. 15-56.

113. Sheikh, K., J. Forster, and L.K. Nielsen, *Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus.* Biotechnol Prog, 2005. **21**(1): p. 112-21.

114. Consortium, I.H.G.S., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

115. Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock, *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

116. Pruitt, K.D. and D.R. Maglott, *RefSeq and LocusLink: NCBI gene-centered resources.* Nucleic Acids Res, 2001. **29**(1): p. 137-40.

117. Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori, *The KEGG resource for deciphering the genome.* Nucleic Acids Res, 2004. **32 Database issue**: p. D277-80.

118. Leyva, J.A., M.A. Bianchet, and L.M. Amzel, *Understanding ATP synthesis: structure and mechanism of the F1-ATPase (Review).* Mol Membr Biol, 2003. **20**(1): p. 27-33.

119. Feist, A.M., J.C.M. Scholten, B.O. Palsson, F.J. Brockman, and T. Ideker, *Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri.* 2006. **2**(1): p. msb4100046-E1-msb4100046-E14.

120. Burgard, A.P., E.V. Nikolaev, C.H. Schilling, and C.D. Maranas, *Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions.* Genome Res, 2004. **14**(2): p. 301-12.

121. Duarte, N.C., Becker, S.A., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.O., *Global reconstruction of the human metabolic network based on genomic and bibliomic data.* Proc Natl Acad Sci U S A, 2006. **Submitted**.

122. Poore, R.E., C.H. Hurst, D.G. Assimos, and R.P. Holmes, *Pathways of hepatic oxalate synthesis and their regulation.* Am J Physiol, 1997. **272**(1 Pt 1): p. C289-94.

123. Banhegyi, G., L. Braun, M. Csala, F. Puskas, and J. Mandl, *Ascorbate metabolism and its regulation in animals.* Free Radic Biol Med, 1997. **23**(5): p. 793-803.

124. Leja, D., *Human Genome Project Timeline*, 38-72.jpg, Editor. 2002, National Human Genome Research Institute. p. Illustration of Human Genome Project Timeline.

125. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., & Marth, J. (Eds.), *Essentials of Glycobiology*. 1999, Plainview, N.Y.: Cold Spring Harbor Laboratory Press.

126. Winchester, B.G., *Lysosomal metabolism of glycoconjugates.* Subcell Biochem, 1996. **27**: p. 191-238.

127. Simpson, G.L. and B.J. Ortwerth, *The non-oxidative degradation of ascorbic acid at physiological conditions.* Biochim Biophys Acta, 2000. **1501**(1): p. 12-24.

128. Marsh, C.A., *Biosynthesis of D-glucaric acid in mammals: a free-radical mechanism?* Carbohydr Res, 1986. **153**(1): p. 119-31.

129. Geisbrecht, B.V. and S.J. Gould, *The human PICD gene encodes a cytoplasmic and peroxisomal NADP(+)-dependent isocitrate dehydrogenase.* J Biol Chem, 1999. **274**(43): p. 30527-33.

130. Eyre, T.A., F. Ducluzeau, T.P. Sneddon, S. Povey, E.A. Bruford, and M.J. Lush, *The HUGO Gene Nomenclature Database, 2006 updates.* Nucleic Acids Res, 2006. **34**(Database issue): p. D319-21.

131. Curwen, V., E. Eyras, T.D. Andrews, L. Clarke, E. Mongin, S.M. Searle, and M. Clamp, *The Ensembl automatic gene annotation system.* Genome Res, 2004. **14**(5): p. 942-50.

132. Blake, J.A., J.T. Eppig, C.J. Bult, J.A. Kadin, and J.E. Richardson, *The Mouse Genome Database (MGD): updates and enhancements.* Nucleic Acids Res, 2006. **34**(Database issue): p. D562-7.

133. Wu, C.H., R. Apweiler, A. Bairoch, D.A. Natale, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, R. Mazumder, C. O'Donovan, N. Redaschi, and B. Suzek, *The Universal Protein Resource (UniProt): an expanding universe of protein information.* Nucleic Acids Res, 2006. **34**(Database issue): p. D187-91.

134. Hamosh, A., A.F. Scott, J.S. Amberger, C.A. Bocchini, and V.A. McKusick, *Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human*

genes and genetic disorders. Nucleic Acids Res, 2005. **33**(Database issue): p. D514-7.

135. Gasteiger, E., E. Jung, and A. Bairoch, *SWISS-PROT: connecting biomolecular knowledge via a protein database.* Curr Issues Mol Biol, 2001. **3**(3): p. 47-55.

136. Linstrom, P.J., Mallard, W.G., *NIST Chemistry WebBook.* NIST Standard Reference Database Number 69. 2005, Gaithersburg, MD: National Institute of Standards and Technology.

137. Affymetrix, *Scientific publications*. 2006.

138. Zierath, J.R. and J.A. Hawley, *Skeletal muscle fiber type: influence on contractile and metabolic properties.* PLoS Biol, 2004. **2**(10): p. e348.

139. Booth, F.W. and D.B. Thomason, *Molecular and cellular adaptation of muscle in response to exercise: perspectives of various models.* Physiol Rev, 1991. **71**(2): p. 541-85.

140. Chibalin, A.V., M. Yu, J.W. Ryder, X.M. Song, D. Galuska, A. Krook, H. Wallberg-Henriksson, and J.R. Zierath, *Exercise-induced changes in expression and activity of proteins involved in insulin signal transduction in skeletal muscle: differential effects on insulin-receptor substrates 1 and 2.* Proc Natl Acad Sci U S A, 2000. **97**(1): p. 38-43.

141. Fluck, M. and H. Hoppeler, *Molecular basis of skeletal muscle plasticity--from gene to form and function.* Rev Physiol Biochem Pharmacol, 2003. **146**: p. 159-216.

142. Hawley, J.A., *Adaptations of skeletal muscle to prolonged, intense endurance training.* Clin Exp Pharmacol Physiol, 2002. **29**(3): p. 218-22.

143. Simoneau, J.A. and C. Bouchard, *Genetic determinism of fiber type proportion in human skeletal muscle.* Faseb J, 1995. **9**(11): p. 1091-5.

144. Lillioja, S., A.A. Young, C.L. Culter, J.L. Ivy, W.G. Abbott, J.K. Zawadzki, H. Yki-Jarvinen, L. Christin, T.W. Secomb, and C. Bogardus, *Skeletal muscle capillary density and fiber type are possible determinants of in vivo insulin resistance in man.* J Clin Invest, 1987. **80**(2): p. 415-24.

145. Wade, A.J., M.M. Marbut, and J.M. Round, *Muscle fibre type and aetiology of obesity.* Lancet, 1990. **335**(8693): p. 805-8.

146. Tanner, C.J., H.A. Barakat, G.L. Dohm, W.J. Pories, K.G. MacDonald, P.R. Cunningham, M.S. Swanson, and J.A. Houmard, *Muscle fiber type is associated with obesity and weight loss.* Am J Physiol Endocrinol Metab, 2002. **282**(6): p. E1191-6.

147. Hickey, M.S., J.O. Carey, J.L. Azevedo, J.A. Houmard, W.J. Pories, R.G. Israel, and G.L. Dohm, *Skeletal muscle fiber composition is related to*

adiposity and in vitro glucose transport rate in humans. Am J Physiol, 1995. **268**(3 Pt 1): p. E453-7.

148. Helge, J.W., A.M. Fraser, A.D. Kriketos, A.B. Jenkins, G.D. Calvert, K.J. Ayre, and L.H. Storlien, *Interrelationships between muscle fibre type, substrate oxidation and body fat.* Int J Obes Relat Metab Disord, 1999. **23**(9): p. 986-91.

149. *Obesity world's biggest health hurdle, conference told.* 2006, Canadian Broadcasting Corporation News.

150. Park, J.J., J.R. Berggren, M.W. Hulver, J.A. Houmard, and E.P. Hoffman, *GRB14, GPD1, and GDF8 as potential network collaborators in weight loss-induced improvements in insulin action in human skeletal muscle.* Physiol Genomics, 2006.

151. Flegal, K.M., M.D. Carroll, C.L. Ogden, and C.L. Johnson, *Prevalence and trends in obesity among US adults, 1999-2000.* Jama, 2002. **288**(14): p. 1723-7.

152. Allison, D.B., K.R. Fontaine, J.E. Manson, J. Stevens, and T.B. VanItallie, *Annual deaths attributable to obesity in the United States.* Jama, 1999. **282**(16): p. 1530-8.

153. Kushner, R., *Roadmaps for Clinical Practice: Case Studies in Disease Prevention and Health Promotion—Assessment and Management of Adult Obesity: A Primer for Physicians.* 2003, American Medical Association: Chicago, IL.

154. *Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies.* Lancet, 2004. **363**(9403): p. 157-63.

155. Deurenberg, P., M. Deurenberg Yap, J. Wang, F.P. Lin, and G. Schmidt, *The impact of body build on the relationship between body mass index and percent body fat.* Int J Obes Relat Metab Disord, 1999. **23**(5): p. 537-42.

156. Stevens, J., J. Cai, E.R. Pamuk, D.F. Williamson, M.J. Thun, and J.L. Wood, *The effect of age on the association between body-mass index and mortality.* N Engl J Med, 1998. **338**(1): p. 1-7.

157. Deurenberg, P., J.A. Weststrate, and J.C. Seidell, *Body mass index as a measure of body fatness: age- and sex-specific prediction formulas.* Br J Nutr, 1991. **65**(2): p. 105-14.

158. Deurenberg, P., M. Yap, and W.A. van Staveren, *Body mass index and percent body fat: a meta analysis among different ethnic groups.* Int J Obes Relat Metab Disord, 1998. **22**(12): p. 1164-71.

159. Gallagher, D., M. Visser, D. Sepulveda, R.N. Pierson, T. Harris, and S.B. Heymsfield, *How useful is body mass index for comparison of body fatness across age, sex, and ethnic groups?* Am J Epidemiol, 1996. **143**(3): p. 228-39.

160. National Institutes of Health Heart, L., and Blood Institute, North American Association for the Study of Obesity, *Practical Guide to the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults*. 2002, National Institutes of Health: Bethesda, MD.

161. Kushner, R.F. and R.L. Weinsier, *Evaluation of the obese patient. Practical considerations.* Med Clin North Am, 2000. **84**(2): p. 387-99, vi.

162. Ullman, K., *Primary Care Takes on Bariatric Care*. 2006. p. 14-15.

163. Fisher, B.L. and P. Schauer, *Medical and surgical options in the treatment of severe obesity.* Am J Surg, 2002. **184**(6B): p. 9S-16S.

164. Richardson, D.W. and A.I. Vinik, *Metabolic implications of obesity: before and after gastric bypass.* Gastroenterol Clin North Am, 2005. **34**(1): p. 9-24.

165. Kim, J.J., M.E. Tarnoff, and S.A. Shikora, *Surgical treatment for extreme obesity: evolution of a rapidly growing field.* Nutr Clin Pract, 2003. **18**(2): p. 109-23.

166. Malinowski, S.S., *Nutritional and metabolic complications of bariatric surgery.* Am J Med Sci, 2006. **331**(4): p. 219-25.

167. *Roux-en-Y stomach surgery for weight loss (A.D.A.M. Medical Encyclopedia)*. 2004, A.D.A.M., Inc.

168. Sugerman, H.J., Nguyen, N.T., *Management of Morbid Obesity*. 2005, New York, NY: Taylor & Francis Group. 264 pages.

169. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository*. 2002. p. 207-210.

170. Liu, G., A.E. Loraine, R. Shigeta, M. Cline, J. Cheng, V. Valmeekam, S. Sun, D. Kulp, and M.A. Siani-Rose, *NetAffx: Affymetrix probesets and annotations*. 2003. p. 82-86.

171. Kayo, T., D.B. Allison, R. Weindruch, and T.A. Prolla, *Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5093-8.

172. Wilson, V.J., M. Rattray, C.R. Thomas, B.H. Moreland, and D. Schulster, *Growth hormone increases IGF-I, collagen I and collagen III gene expression in dwarf rat skeletal muscle.* Mol Cell Endocrinol, 1995. **115**(2): p. 187-97.

173. Wilson, V.J., M. Rattray, C.R. Thomas, B.H. Moreland, and D. Schulster, *Effects of hypophysectomy and growth hormone administration on the mRNA levels of collagen I, III and insulin-like growth factor-I in rat skeletal muscle.* Growth Horm IGF Res, 1998. **8**(6): p. 431-8.

174. Hulver, M.W., J.R. Berggren, R.N. Cortright, R.W. Dudek, R.P. Thompson, W.J. Pories, K.G. MacDonald, G.W. Cline, G.I. Shulman, G.L. Dohm, and J.A. Houmard, *Skeletal muscle lipid metabolism with obesity.* Am J Physiol Endocrinol Metab, 2003. **284**(4): p. E741-7.

175. Curtis, R.K., M. Oresic, and A. Vidal-Puig, *Pathways to the analysis of microarray data.* Trends Biotechnol, 2005. **23**(8): p. 429-35.

176. Akesson, M., J. Forster, and J. Nielsen, *Integration of gene expression data into genome-scale metabolic models.* Metab Eng, 2004. **6**(4): p. 285-93.

177. Draghici, S., P. Khatri, R.P. Martins, G.C. Ostermeier, and S.A. Krawetz, *Global functional profiling of gene expression.* Genomics, 2003. **81**(2): p. 98-104.

178. Grosu, P., J.P. Townsend, D.L. Hartl, and D. Cavalieri, *Pathway Processor: a tool for integrating whole-genome expression results into metabolic networks.* Genome Res, 2002. **12**(7): p. 1121-6.

179. Su, A.I., T. Wiltshire, S. Batalov, H. Lapp, K.A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M.P. Cooke, J.R. Walker, and J.B. Hogenesch, *A gene atlas of the mouse and human protein-encoding transcriptomes.* Proc Natl Acad Sci U S A, 2004. **101**(16): p. 6062-7.

180. *Genomics Institute of the Novartis Research Foundation Gene Expression Database*. 2006.

181. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. National Institutes of Health.* Obes Res, 1998. **6 Suppl 2**: p. 51S-209S.

182. Joyce, A.R. and B.O. Palsson, *The model organism as a system: integrating 'omics' data sets.* Nat Rev Mol Cell Biol, 2006. **7**(3): p. 198-210.

183. Ge, H., A.J. Walhout, and M. Vidal, *Integrating 'omic' information: a bridge between genomics and systems biology.* Trends Genet, 2003. **19**(10): p. 551-60.

184. Palsson, B.O., *In silico biology through "omics".* Nat. Biotechnol., 2002. **20**(7): p. 649-650.

185. Kitano, H., *Systems biology: a brief overview.* Science, 2002. **295**(5560): p. 1662-4.

186. Reed, J.L. and B.O. Palsson, *Thirteen Years of Building Constraint-Based In Silico Models of Escherichia coli.* J Bacteriol, 2003. **185**(9): p. 2692-9.

187. Becker, S.A. and B.O. Palsson, *Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation.* BMC Microbiol, 2005. **5**(1): p. 8.

188. Thiele, I., T.D. Vo, N.D. Price, and B. Palsson, *An Expanded Metabolic Reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single and double deletion mutants.* J Bacteriol., 2005. **187**(16): p. 5818-5830.

189. Gibbs, R.A., G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, S. Scherer, G. Scott, D. Steffen, K.C. Worley, P.E. Burch, G. Okwuonu, S. Hines, L. Lewis, C. DeRamo, O. Delgado, S. Dugan-Rocha, G. Miner, M. Morgan, A. Hawes, R. Gill, Celera, R.A. Holt, M.D. Adams, P.G. Amanatides, H. Baden-Tillson, M. Barnstead, S. Chin, C.A. Evans, S. Ferriera, C. Fosler, A. Glodek, Z. Gu, D. Jennings, C.L. Kraft, T. Nguyen, C.M. Pfannkoch, C. Sitter, G.G. Sutton, J.C. Venter, T. Woodage, D. Smith, H.M. Lee, E. Gustafson, P. Cahill, A. Kana, L. Doucette-Stamm, K. Weinstock, K. Fechtel, R.B. Weiss, D.M. Dunn, E.D. Green, R.W. Blakesley, G.G. Bouffard, P.J. De Jong, K. Osoegawa, B. Zhu, M. Marra, J. Schein, I. Bosdet, C. Fjell, S. Jones, M. Krzywinski, C. Mathewson, A. Siddiqui, N. Wye, J. McPherson, S. Zhao, C.M. Fraser, J. Shetty, S. Shatsman, K. Geer, Y. Chen, S. Abramzon, W.C. Nierman, P.H. Havlak, R. Chen, K.J. Durbin, A. Egan, Y. Ren, X.Z. Song, B. Li, Y. Liu, X. Qin, S. Cawley, K.C. Worley, A.J. Cooney, L.M. D'Souza, K. Martin, J.Q. Wu, M.L. Gonzalez-Garay, A.R. Jackson, K.J. Kalafus, M.P. McLeod, A. Milosavljevic, D. Virk, A. Volkov, D.A. Wheeler, Z. Zhang, J.A. Bailey, E.E. Eichler, E. Tuzun, E. Birney, E. Mongin, A. Ureta-Vidal, C. Woodwark, E. Zdobnov, P. Bork, M. Suyama, D. Torrents, M. Alexandersson, B.J. Trask, J.M. Young, H. Huang, H. Wang, H. Xing, S. Daniels, D. Gietzen, J. Schmidt, K. Stevens, U. Vitt, J. Wingrove, F. Camara, M. Mar Alba, J.F. Abril, R. Guigo, A. Smit, I. Dubchak, E.M. Rubin, O. Couronne, A. Poliakov, N. Hubner, D. Ganten, C. Goesele, O. Hummel, T. Kreitler, Y.A. Lee, J. Monti, H. Schulz, H. Zimdahl, H. Himmelbauer, H. Lehrach, H.J. Jacob, S. Bromberg, J. Gullings-Handley, M.I. Jensen-Seaman, A.E. Kwitek, J. Lazar, D. Pasko, P.J. Tonellato, S. Twigger, C.P. Ponting, J.M. Duarte, S. Rice, L. Goodstadt, S.A. Beatson, R.D. Emes, E.E. Winter, C. Webber, P. Brandt, G. Nyakatura, M. Adetobi, F. Chiaromonte, L. Elnitski, P. Eswara, R.C. Hardison, M. Hou, D. Kolbe, K. Makova, W. Miller, A. Nekrutenko, C. Riemer, S. Schwartz, J. Taylor, S. Yang, Y. Zhang, K. Lindpaintner, T.D. Andrews, M. Caccamo, M. Clamp, L. Clarke, V. Curwen, R. Durbin, E. Eyras, S.M. Searle, G.M. Cooper, S. Batzoglou, M. Brudno, A. Sidow, E.A. Stone, J.C. Venter, B.A. Payseur, G. Bourque, C. Lopez-Otin, X.S. Puente, K. Chakrabarti, S. Chatterji, C. Dewey, L. Pachter, N. Bray, V.B. Yap, A. Caspi, G. Tesler, P.A. Pevzner, D. Haussler, K.M. Roskin, R. Baertsch, H. Clawson, T.S. Furey, A.S. Hinrichs, D. Karolchik, W.J. Kent, K.R. Rosenbloom, H. Trumbower, M. Weirauch, D.N. Cooper, P.D. Stenson, B. Ma, M. Brent, M. Arumugam, D. Shteynberg, R.R. Copley, M.S. Taylor, H. Riethman, U. Mudunuri, J. Peterson, M. Guyer, A. Felsenfeld, S. Old, S. Mockrin and F. Collins, *Genome sequence of the Brown Norway rat yields insights into mammalian evolution.* Nature, 2004. **428**(6982): p. 493-521.

190. Lindblad-Toh, K., C.M. Wade, T.S. Mikkelsen, E.K. Karlsson, D.B. Jaffe, M. Kamal, M. Clamp, J.L. Chang, E.J. Kulbokas, 3rd, M.C. Zody, E. Mauceli, X. Xie, M. Breen, R.K. Wayne, E.A. Ostrander, C.P. Ponting, F. Galibert, D.R. Smith, P.J. DeJong, E. Kirkness, P. Alvarez, T. Biagi, W. Brockman, J. Butler, C.W. Chin, A. Cook, J. Cuff, M.J. Daly, D. DeCaprio, S. Gnerre, M. Grabherr,

M. Kellis, M. Kleber, C. Bardeleben, L. Goodstadt, A. Heger, C. Hitte, L. Kim, K.P. Koepfli, H.G. Parker, J.P. Pollinger, S.M. Searle, N.B. Sutter, R. Thomas, C. Webber, J. Baldwin, A. Abebe, A. Abouelleil, L. Aftuck, M. Ait-Zahra, T. Aldredge, N. Allen, P. An, S. Anderson, C. Antoine, H. Arachchi, A. Aslam, L. Ayotte, P. Bachantsang, A. Barry, T. Bayul, M. Benamara, A. Berlin, D. Bessette, B. Blitshteyn, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, A. Brown, P. Cahill, N. Calixte, J. Camarata, Y. Cheshatsang, J. Chu, M. Citroen, A. Collymore, P. Cooke, T. Dawoe, R. Daza, K. Decktor, S. DeGray, N. Dhargay, K. Dooley, K. Dooley, P. Dorje, K. Dorjee, L. Dorris, N. Duffey, A. Dupes, O. Egbiremolen, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, P. Ferreira, S. Fisher, M. FitzGerald, K. Foley, C. Foley, A. Franke, D. Friedrich, D. Gage, M. Garber, G. Gearin, G. Giannoukos, T. Goode, A. Goyette, J. Graham, E. Grandbois, K. Gyaltsen, N. Hafez, D. Hagopian, B. Hagos, J. Hall, C. Healy, R. Hegarty, T. Honan, A. Horn, N. Houde, L. Hughes, L. Hunnicutt, M. Husby, B. Jester, C. Jones, A. Kamat, B. Kanga, C. Kells, D. Khazanovich, A.C. Kieu, P. Kisner, M. Kumar, K. Lance, T. Landers, M. Lara, W. Lee, J.P. Leger, N. Lennon, L. Leuper, S. LeVine, J. Liu, X. Liu, Y. Lokyitsang, T. Lokyitsang, A. Lui, J. Macdonald, J. Major, R. Marabella, K. Maru, C. Matthews, S. McDonough, T. Mehta, J. Meldrim, A. Melnikov, L. Meneus, A. Mihalev, T. Mihova, K. Miller, R. Mittelman, V. Mlenga, L. Mulrain, G. Munson, A. Navidi, J. Naylor, T. Nguyen, N. Nguyen, C. Nguyen, T. Nguyen, R. Nicol, N. Norbu, C. Norbu, N. Novod, T. Nyima, P. Olandt, B. O'Neill, K. O'Neill, S. Osman, L. Oyono, C. Patti, D. Perrin, P. Phunkhang, F. Pierre, M. Priest, A. Rachupka, S. Raghuraman, R. Rameau, V. Ray, C. Raymond, F. Rege, C. Rise, J. Rogers, P. Rogov, J. Sahalie, S. Settipalli, T. Sharpe, T. Shea, M. Sheehan, N. Sherpa, J. Shi, D. Shih, J. Sloan, C. Smith, T. Sparrow, J. Stalker, N. Stange-Thomann, S. Stavropoulos, C. Stone, S. Stone, S. Sykes, P. Tchuinga, P. Tenzing, S. Tesfaye, D. Thoulutsang, Y. Thoulutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, V. Venkataraman, A. Vo, T. Wangchuk, T. Wangdi, M. Weiand, J. Wilkinson, A. Wilson, S. Yadav, S. Yang, X. Yang, G. Young, Q. Yu, J. Zainoun, L. Zembek, A. Zimmer and E.S. Lander, *Genome sequence, comparative analysis and haplotype structure of the domestic dog.* Nature, 2005. **438**(7069): p. 803-19.

191. Hillier, L.W., W. Miller, E. Birney, W. Warren, R.C. Hardison, C.P. Ponting, P. Bork, D.W. Burt, M.A. Groenen, M.E. Delany, J.B. Dodgson, A.T. Chinwalla, P.F. Cliften, S.W. Clifton, K.D. Delehaunty, C. Fronick, R.S. Fulton, T.A. Graves, C. Kremitzki, D. Layman, V. Magrini, J.D. McPherson, T.L. Miner, P. Minx, W.E. Nash, M.N. Nhan, J.O. Nelson, L.G. Oddy, C.S. Pohl, J. Randall-Maher, S.M. Smith, J.W. Wallis, S.P. Yang, M.N. Romanov, C.M. Rondelli, B. Paton, J. Smith, D. Morrice, L. Daniels, H.G. Tempest, L. Robertson, J.S. Masabanda, D.K. Griffin, A. Vignal, V. Fillon, L. Jacobbson, S. Kerje, L. Andersson, R.P. Crooijmans, J. Aerts, J.J. van der Poel, H. Ellegren, R.B. Caldwell, S.J. Hubbard, D.V. Grafham, A.M. Kierzek, S.R. McLaren, I.M. Overton, H. Arakawa, K.J. Beattie, Y. Bezzubov, P.E. Boardman, J.K. Bonfield, M.D. Croning, R.M. Davies, M.D. Francis, S.J. Humphray, C.E. Scott, R.G. Taylor, C. Tickle, W.R. Brown, J. Rogers, J.M. Buerstedde, S.A. Wilson, L. Stubbs, I. Ovcharenko, L. Gordon, S. Lucas, M.M. Miller, H. Inoko, T. Shiina, J. Kaufman, J. Salomonsen, K. Skjoedt, G.K. Wong, J. Wang, B. Liu, J. Wang, J. Yu, H. Yang, M. Nefedov, M.

Koriabine, P.J. Dejong, L. Goodstadt, C. Webber, N.J. Dickens, I. Letunic, M. Suyama, D. Torrents, C. von Mering, E.M. Zdobnov, K. Makova, A. Nekrutenko, L. Elnitski, P. Eswara, D.C. King, S. Yang, S. Tyekucheva, A. Radakrishnan, R.S. Harris, F. Chiaromonte, J. Taylor, J. He, M. Rijnkels, S. Griffiths-Jones, A. Ureta-Vidal, M.M. Hoffman, J. Severin, S.M. Searle, A.S. Law, D. Speed, D. Waddington, Z. Cheng, E. Tuzun, E. Eichler, Z. Bao, P. Flicek, D.D. Shteynberg, M.R. Brent, J.M. Bye, E.J. Huckle, S. Chatterji, C. Dewey, L. Pachter, A. Kouranov, Z. Mourelatos, A.G. Hatzigeorgiou, A.H. Paterson, R. Ivarie, M. Brandstrom, E. Axelsson, N. Backstrom, S. Berlin, M.T. Webster, O. Pourquie, A. Reymond, C. Ucla, S.E. Antonarakis, M. Long, J.J. Emerson, E. Betran, I. Dupanloup, H. Kaessmann, A.S. Hinrichs, G. Bejerano, T.S. Furey, R.A. Harte, B. Raney, A. Siepel, W.J. Kent, D. Haussler, E. Eyras, R. Castelo, J.F. Abril, S. Castellano, F. Camara, G. Parra, R. Guigo, G. Bourque, G. Tesler, P.A. Pevzner, A. Smit, L.A. Fulton, E.R. Mardis and R.K. Wilson, *Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.* Nature, 2004. **432**(7018): p. 695-716.

192. Palsson, B.O., *The challenges of in silico biology.* Nat Biotechnol, 2000. **18**(11): p. 1147-50.

193. Kuepfer, L., U. Sauer, and L.M. Blank, *Metabolic functions of duplicate genes in Saccharomyces cerevisiae.* Genome Res, 2005. **15**(10): p. 1421-30.

194. Mo, M.L., Palsson, B.O. *Saccharomyces cereivisiae iMM910, an improved and expanded yeast metabolic reconstruction*. in preparation [cited.

195. Herrgard, M.J., B.S. Lee, V. Portnoy, and B.O. Palsson, *Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae.* Genome Res, 2006. **16**(5): p. 627-35.

196. Wunderlich, Z. and L.A. Mimy, *Using the topology of metabolic networks to predict viability of mutant strains.* Biophysical Journal, 2006. **91**(6): p. 2304-2311.

197. Mahadevan, R. and B.O. Palsson, *Properties of metabolic networks: structure versus function.* Biophys J, 2005. **88**(1): p. L07-9.

198. Samal, A., S. Singh, V. Giri, S. Krishna, N. Raghuram, and S. Jain, *Low degree metabolites explain essential reactions and enhance modularity in biological networks.* BMC Bioinformatics, 2006. **7**(1): p. 118.

199. Almaas, E., Z.N. Oltvai, and A.L. Barabasi, *The Activity Reaction Core and Plasticity of Metabolic Networks.* PLoS Comput Biol, 2005. **1**(7): p. e68.

200. Urbanczik, R. and C. Wagner, *Functional stoichiometric analysis of metabolic networks.* Bioinformatics, 2005. **21**(22): p. 4176-80.

201. Wang, L. and V. Hatzimanikatis, *Metabolic engineering under uncertainty--II: analysis of yeast metabolism.* Metab Eng, 2006. **8**(2): p. 142-59.

202. Becker, S.A., N.D. Price, and B.O. Palsson, *Metabolite coupling in genome-scale metabolic networks.* BMC Bioinformatics, 2006. **7**: p. 111.

203. Zhao, J., H. Yu, J.H. Luo, Z.W. Cao, and Y.X. Li, *Hierarchical modularity of nested bow-ties in metabolic networks.* Bmc Bioinformatics, 2006. **7**.

204. Bilu, Y., T. Shlomi, N. Barkai, and E. Ruppin, *Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states.* Plos Computational Biology, 2006. **2**(8): p. 932-938.

205. Chen, L.F. and D. Vitkup, *Predicting genes for orphan metabolic activities using phylogenetic profiles.* Genome Biology, 2006. **7**(2).

206. Famili, I. and B.O. Palsson, *Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices.* J Theor Biol, 2003. **224**(1): p. 87-96.

207. Green, M.L. and P.D. Karp, *Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers.* Nucleic Acids Res, 2005. **33**(13): p. 4035-9.

208. Lareau, L.F., R.E. Green, R.S. Bhatnagar, and S.E. Brenner, *The evolving roles of alternative splicing.* Curr Opin Struct Biol, 2004. **14**(3): p. 273-82.

209. Jamshidi, N. and B.O. Palsson, *Systems biology of SNPs.* Mol Syst Biol, 2006. **2**: p. 38.

210. Papin, J.A., N.D. Price, S.J. Wiback, D.A. Fell, and B.O. Palsson, *Metabolic pathways in the post-genome era.* Trends Biochem Sci, 2003. **28**(5): p. 250-8.

211. Papin, J.A., J.L. Reed, and B.O. Palsson, *Hierarchical thinking in network biology: the unbiased modularization of biochemical networks.* Trends Biochem Sci, 2004. **29**(12): p. 641-7.

212. Becker, S.A., *Senate exam*. 2006, Department of Bioengineering, University of California, San Diego.