UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Characterizing complex phenotypes in metabolism : an "omics"-driven systems approach

Permalink

https://escholarship.org/uc/item/5ds062cp

Author

Mo, Monica L.

Publication Date

2009

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Characterizing complex phenotypes in metabolism: An "omics"-driven systems approach

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Monica L. Mo

Committee in charge:

Professor Bernhard Ø. Palsson, Chair
Professor Steven Briggs
Professor Jeff Hasty
Professor Andrew McCulloch
Professor Shyni Varghese

2009

Copyright
Monica L. Mo, 2009
All rights reserved.

The dissertation of Monica L. Mo is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2009

DEDICATION

To Don...I look forward to all that's to come.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

These past five years in graduate school has been quite a journey, one that I could not have gone through without the help and support of some important people. First, I'd like to thank the National Institutes of Health and National Science Foundation for the research grants that have funded my graduate studies during my time at UCSD. I'd like to also thank all of my thesis committee members for their time and commitment. Dr. Palsson, I am extremely thankful for your constant encouragement and the career (and travel!) opportunities you've pushed my way through these past several years.

I'd like to acknowledge the members of the Systems Biology Research Group, past and present. It has been a pleasure to have been a part of such a uniquely talented group of researchers, and I look forward to continuing to collaborate on new projects with you all. To my office mates Karsten and Vasiliy, thank you for always keeping lab life light and humorous. Jan, thanks for always having the answers to all of my statistics and computer questions. Neema, thanks for all of the edible treats during our talks about research and everything else random. Markus, thank you for being in the lab at a time when I needed guidance the most in those first few years of graduate school. I'll always be thankful for your constructive criticism and having you as a sounding board for my random ideas during our long conversations. To my fellow Recon 1 reconstructors, thank you for the dedication and time you put into the work that preceded much of my thesis studies that followed. You made my life much easier.

Thank you, Mom and Dad for getting me to this point in my life. I may not always admit that I follow your advice, but much of what I've done up to this point has been an extension of what you've taught me. Laura, as cliche as it sounds, you really are the best little sister one can ask for. To Jen H, Esther, and Jen A, great friends are hard to come by, especially ones who support you unconditionally through all aspects of life. I will always treasure our conversations over caffeinated drinks, training runs, poker nights, and mini-vacations. You kept me grounded and sane through this entire graduate school process, and most importantly, helped me maintain perspective on what life should be about. To this little being growing inside of me, thank you for providing the "urgency" to complete this chapter of my life. I can't wait to finally meet you. Don, thank you for your unwavering love and support through these past 8 years. You're right, patience and persistence really does pay off in the end.

The text of Chapter One, in full, is a reprint of the material as it appears in M.L.

Mo and B.Ø. Palsson. 2009. Understanding human metabolic physiology: A genome-to-systems approach. Trends in Biotechnology, 27(1):37-44. I was the primary author of this publication and the co-author participated and supervised the research, which forms the basis for this chapter.

The text of Chapter Two, in part, is a reprint of the material as it appears in M.L. Mo, B.Ø. Palsson, and M.J. Herrgard. 2009. Connecting extracellular metabolomic profiles to intracellular metabolic states in yeast. BMC Systems Biology. 3:37. I was the primary author of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

The text of Chapter Three, in full, is a reprint of the material as it appears in M.L. Mo, N. Jamshidi, and B.Ø. Palsson. 2007. A Genome-scale, Constraint-based Approach to Systems Biology of Human Metabolism., Molecular Biosystems. 3:9 and in M.L. Mo and B.Ø. Palsson. 2009. Understanding human metabolic physiology: A genome-to-systems approach. Trends in Biotechnology, 27(1):37-44. I was the primary author of these publications and the co-authors participated and/or supervised the research which forms the basis for this chapter.

The text of Chapter Four, in full, is a reprint of the material as it will likely appear in M.L. Mo, M.J. Herrgard, and B.Ø. Palsson. 2009. Characterizing global drug response phenotypes in human metabolism. (In preparation). I was the primary author of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

The text of Chapter Five, in full, is a reprint of the material as it appears in M.L. Mo, B.Ø. Palsson, and M.J. Herrgard. 2009. Connecting extracellular metabolomic profiles to intracellular metabolic states in yeast. BMC Systems Biology. 3:37. I was the primary author of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

The text of Chapter Six, in part, is a reprint of the material as it will likely appear in Yanes, O., Clark, J., Mo, M.L., Wong, D.M., Sanchez-Ruiz, A., Benton, P., Trauger, S.A., Desponts, C., Patti, G.J., Palsson, B.Ø., Ding, S., Siuzdak, G., Highly reactive endogenous metabolites characterize embryonic stem cells. Nature Chemical Biology (In review). I was the primary author of the text in that portion of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

The text of Chapter Seven, in part, is a reprint of the material as it appears in M.L. Mo and B.Ø. Palsson. 2009. Understanding human metabolic physiology: A genome-

to-systems approach. Trends in Biotechnology, 27(1):37-44. I was the primary author of this publication and the co-author participated and supervised the research, which forms the basis for this chapter.

VITA

2004	B.S., Bioengineering: Biotechnology, University of California, San Diego
2007	M.S., Bioengineering, University of California, San Diego
2009	Ph.D., Bioengineering, University of California, San Diego

PUBLICATIONS

Duarte, N.D., Becker, S.A., Jamshidi, N., Thiele, I., **Mo, M.L.**, Vo, T.D., Srivas, R., Palsson, B.Ø., Global reconstruction of the human metabolic network based on genomic and bibliomic data , Proc Natl Acad. Sci U S A, 104(6):1777-82 (2007)

Becker, S.A., Feist, A.M., **Mo, M. L.**, Hannum, G., Palsson, B.O., Herrgard, M.J. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox., Nat. Protocols, 2, 727-738 (2007)

Mo, M.L., Jamshidi, N., Palsson, B.Ø., A Genome-scale, Constraint-based Approach to Systems Biology of Human Metabolism., Molecular Biosystems. 3:9 (2007)

Herrgrd, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., Blthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novre, N., Li, P., Liebermeister, W., **Mo, M. L.**, Oliveira, A. P., Petranovic, D., Pettifer, S., Simeonidis, E., Smallbone, K., Spasi, I., Weichart, D., Brent, R., Broomhead, D. S., Westerhoff, H. V., Kardar, B., Penttil, M., Klipp, E., Palsson, B.Ø., Sauer, U., Oliver, S. G., Mendes, P., Nielsen, J. and Kell, D. B., A consensus yeast metabolic network obtained from a community approach to systems biology., Nature Biotechnology, 26: 1155-1160 (2008).

Mo, M.L. and Palsson, B.Ø. Understanding human metabolic physiology: A genome-to-systems approach, Trends in Biotechnology. 27(1):37-44 (2009).

Mo, M.L., Palsson, B.Ø., Herrgard, M.J., Connecting extracellular metabolomic profiles to intracellular metabolic states in yeast. BMC Systems Biology. 3:37 (2009).

Yanes, O., Clark, J., **Mo, M.L.**, Wong, D.M., Sanchez-Ruiz, A., Benton, P., Trauger, S.A., Desponts, C., Patti, G.J., Palsson, B.Ø., Ding, S., Siuzdak, G., Highly reactive endogenous metabolites characterize embryonic stem cells. Nature Chemical Biology (In review)

Mo, M.L., Herrgard, M.J., Palsson, B.Ø. Characterizing global drug response phenotypes in human metabolism (In preparation).

ABSTRACT OF THE DISSERTATION

Characterizing complex phenotypes in metabolism: An "omics"-driven systems approach

by

Monica L. Mo
Doctor of Philosophy in Bioengineering
University of California San Diego, 2009
Professor Bernhard Ø. Palsson, Chair

The advent of high-throughput technologies has resulted in an explosion of molecular data. A major challenge is found in interpreting and understanding these different types of data sets at a phenotypic level. Systems biology has capitalized on these technologies by consolidating various types of biological information into structured networks for their analysis and computation. The bottom-up systems biology approach, in particular, has been crucial in providing mechanistic foundations for systems-level modeling in microorganisms, and its extension to eukaryotic metabolism has made it possible to elucidate complex phenotypes in a systematic manner. The work presented in this dissertation describes the integrative use of high-throughput data and genome-scale network reconstructions to characterize complex phenotypes of eukaryotic metabolism. First, the genome-scale reconstructions of yeast and human metabolism are discussed, which provide the contextual basis in which "omics" data is analyzed. Previously developed constraint-based modeling approaches were refined to analyze "omics" data sets, in particular for transcriptomic and metabolomic data. Finally, example applications are presented in the evaluation of physiological and perturbed metabolic states of yeast and human cellular systems. The studies discussed herein are: (1) analyzing drug response phenotypes of human metabolism; (2) evaluating genetic and environmentally perturbed processes in yeast ammonium assimilation; and (3) characterizing the pluripotent phenotype of embryonic stem cell metabolism. The work described in this dissertation represents advancement towards integrating bottom-up and data-driven approaches to understanding broader "omics"-to-phenotype relationships.

Chapter 1

Modeling complex phenotypes of metabolism

1.1 The human genome: A system defined

The human genome sequence is generally touted as a blueprint to physiological functions of the human body. However, the translation from the annotated sequence to cellular physiology remains elusive. The annotated sequence contains a wealth of information about the gene products (proteins) and the biochemical processes that they mediate. This information essentially provides a parts list of biological components that exist in human cells, but it is thfe interplay between these components which governs physiological behavior.

Human metabolic physiology arises collectively from different levels of biological organization, as illustrated in Figure 1.1, and thus requires a systems perspective to understand it as a whole. A system is a collection of individual parts that work synergistically as a single, functional unit. Much of human physiology can be described in a synonymous manner; for example, tissue-level functions are the result of interactions between multiple cell types, and the human body is comprised of interconnected organ and tissue systems that enable various whole-body functions. This overall systems view on biology is the conceptual foundation for the construction and use of networks to understand human physiologic functions.

The complex nature of biological interactions has led to the need for and development of systems biology, an emerging field of research that combines high-throughput ex-

perimentation with computational tools to systematically analyze biological systems. Two approaches are commonly used in systems biology, top-down and bottom-up; both aim at understanding the interactions arising under various physiological conditions, but from different perspectives. The top-down strategy analyzes from a global point-of-view by dissecting the overall system into its smaller interacting parts and generally requires statistical measures. Alternatively, the bottom-up approach first specifies in detail the individual components and interactions of a system. The components are then pieced together into a larger system, thus providing a mechanistic basis to studying its underlying biology based on the known, specified parts. The data type largely determines the approach which is most amenable to answering a particular biological question and is dependent on the level of interaction detail available.

In the post-genomic era, the genome-to-life concept has ignited (molecular) systems biology and the implementation of its central paradigm: molecular components -> networks -> computational models -> physiological studies [1]. Since much of the component data is genome-derived, this paradigm is tantamount to developing a mechanistic genotype-phenotype relationship. The prospects of such a development would indeed be transformative in biology and, if successfully realized, can have a broad impact on the life sciences and life science-based industries. The shift towards human systems biology has been initiated with the emergence of bottom-up, genome-scale human metabolic networks [2, 3, 4], which will play an important role in studying human health and disease.

1.2 Genome-scale reconstructions in metabolic systems biology

Metabolism is widely known to play an important role in human physiology, and its function is important to understanding disease states and progression [5, 6], aging and nutrition [7, 8], and sports, astronaut and soldier performance [9, 10]. In particular, metabolism is a key factor in human disease, including diabetes and obesity [11, 12], cancer [13, 14], neurodegenerative and psychiatric diseases [15, 16], alcoholism [17], and ischemia [18]. Successful implementation of molecular systems biology of human metabolism is thus likely to have broad consequences.

The reconstruction of genome-scale metabolic networks in microorganisms is now well developed [19, 20, 21, 22, 23, 24] and has been successfully implemented in computing microbial metabolic phenotypes for a variety of biological applications [25, 26]. While the

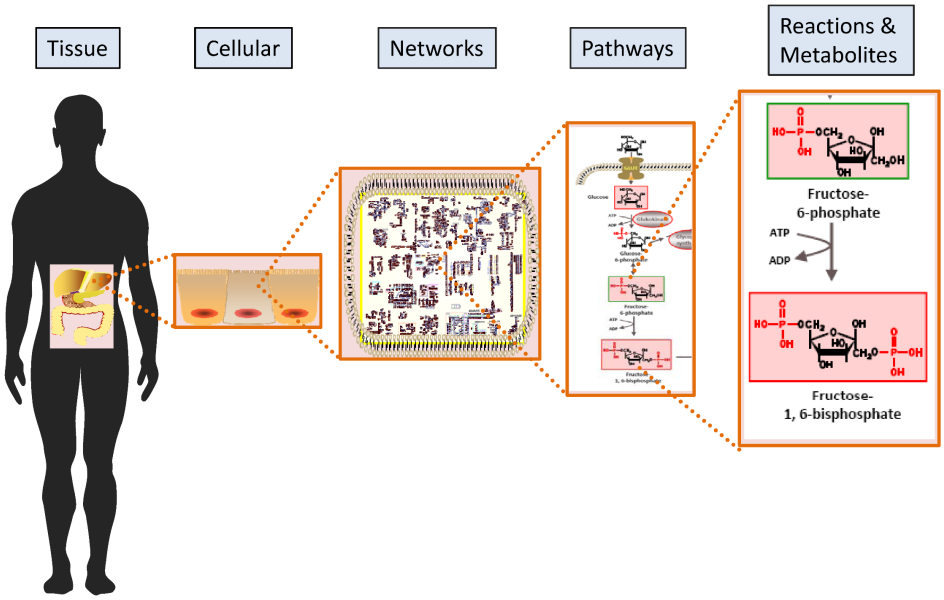


Figure 1.1: Dissecting the different biological levels of interaction in human metabolic physiology.

Human metabolism is the consequence of interacting components organized at the tissue, cellular, pathway, and reaction levels. Metabolic functions of organs result from interacting tissues. Tissue functions are mediated by cell-to-cell interactions, which are controlled by networks of pathway activity within each cell. Pathways are comprised of connecting reactions that carry out specific metabolic functions, linked together by the metabolites in which they process and form.

Table 1.1: List of available genome-scale human metabolic networks.

Metabolic Network	**Publication**	**Description**
HumanCyc	Genome Biology (2005)	An automated draft reconstruction based on mapping from Ensembl and Entrez Gene databases. PathoLogic alrogithm was used to fill pathway holes with candidate reactions based on sequence comparison.
Recon 1	PNAS (2007)	A manually curated network reconstruction and functional *in silico* model. Manual curation, reaction gap assessment and 288 flux balance analysis (FBA)-based functional validation tests were performed.
Edinburgh Human Metabolic Network	Molecular Systems Biology (2007)	Network created from automated genome annotation mapping and inclusion of biochemical legacy data. EMP database was used to obtain and integrate biochemical literature-based information.

biochemical makeup of enzymes and metabolites vary, the general framework of metabolism is consistent across all organisms and metabolic studies have been enabled for a wide range of species using similar methodologies. Thus, a logical next step is to extend the successful development and analysis of microbial metabolic networks to an analogous effort to study human metabolism. Reconstructing metabolism provides a starting basis to building large-scale, mechanistically-accurate networks for human physiology as the biochemical transformations of metabolism are well studied and documented at both pathway and mechanistic levels. Metabolic systems may thus become the first process in human cells and tissues where the application of molecular systems biology will bear fruit.

Two approaches have emerged in systems biology, top-down and bottom-up; both aim at understanding the interactions arising under various physiological conditions, but from different perspectives. The top-down strategy analyzes from a global point-of-view by dissecting the overall system into its smaller interacting parts with statistical approaches. Alternatively, the bottom-up approach first specifies in detail the individual components and interactions of a system. Components are then pieced together into a larger system, thus providing a mechanistic basis to studying its underlying biology based on known, specified parts. Since much of the components are genome-derived, a bottom-up network reconstruction approach is crucial to developing a mechanistic genotype-phenotype relationship.

An annotated genome along with literature, or bibliome, data defines the known components present in a biological system. Because this information exists in many dif-

ferent domains, there is a need to compile the data in a structured format which catalogs genes, their associated protein products, and related biological functions. A component-by-component, or bottom-up, approach to network reconstruction results in a biochemically, genetically, and genomically (BiGG) structured format that serves as a knowledge base of genome-derived biological components and a framework for computational modeling. This part of the reconstruction process has been described extensively for microbial metabolic networks [26, 27] and can thus be readily applied towards human metabolism.

Figure 1.2 illustrates the general information that is derived and considered from annotation databases and the bibliome as it is incorporated into a BiGG reconstruction. The information is structured such that the whole genome is dissected into its gene parts (genomically), genes are related to its encoded proteins (genetically), and protein enzymes are linked to their catalytic reaction functions and the metabolite species in which they interconvert (biochemically). Such information is structured to describe the connections between genes, proteins, and their respective metabolic functions. Complex relationships, such as reactions catalyzed by more than one enzyme (i.e. isozymes) and multi-functional proteins, can be textually and graphically described as Boolean logic relationships as an additional layer of reaction information [2, 3]. This multi-level structure distinguishes a knowledge base network from a standard database by providing an integrative view of disparate data types and placing them in a relevant biological context. Automation and manual curation approaches can be used to reconstruct a network. Automated methods have been used to enable quicker mining and cross-comparison of data from different resources and automatically assign potential functions to annotated genes. However, this essentially generates a rough draft of the knowledge content that requires additional refinement to enhance its content quality. Further investigation and consideration of experimental data based on literature reading provides evidence for a biochemical reactions addition to the network and has been emphasized as an integral part of the reconstruction process [3, 4]. While manual verification of literature evidence is a timely and laborious process, it is a crucial QC/QA procedure that a reconstruction process must undergo to ensure content quality. Such procedures include charge- and mass-balancing reactions, determining localization of reaction activity, identifying substrate and cofactor specificity, and, more specific to human metabolic reconstructions, the incorporation of alternatively spliced variants. These data workflows have previously been defined and have been implemented in metabolic networks for microbial and human systems [26, 27, 28].

While the genome annotation can be used to derive a majority of human metabolic

reactions, missing reactions exist where the genome has not been fully elucidated and require additional evaluation to fill in the gaps. Gap-filling can be done algorithmically; for example, one approach involves prediction of candidate reactions required to fill human metabolic pathway gaps by projecting known pathways from other organisms [2]. Alternatively, reactions can be inferred directly from network topology [29]. A more arduous approach involves manual assessment of published literature and adding those reactions that were not automatically identified from the annotated genome. Literature mining and review not only increases the confidence of adding a reaction, but also places the physiological functions that each reaction fulfills into context (e.g. at the metabolic pathway level).

Further consideration can be made to ensure reaction gaps are filled in the context of metabolic functions. Cellular biomass growth or energy demands are metabolic functions that are primarily assessed for networks of single-celled, microbial organism. However, current genome-scale human networks represent metabolic reactions that exist in any human cell and are therefore a global depiction of all human metabolic functions. Thus, in reconstructing a network for global human metabolism, a variety of basic metabolic functions that can exist in any human cell must be considered. Functional validation testing was implemented, as described in [3], during the reconstruction procedure and ensured that 288 basic metabolic processes compiled from literature were indeed computationally functional. Tested metabolic processes represent a defined list of known physiological functions and include ketogenesis, ATP production, and biosynthesis of non-essential amino acids from their respective precursors. Algorithms have now been developed which can be useful during reconstruction to computationally identify gaps and determine candidate reactions required to fulfill a particular metabolic function [30, 31].

Once formed, the BiGG knowledge base becomes the basis for a mathematical representation of the network that is reconstructed from it. A network is comprised of nodes and links, where the nodes are biological components and the links are chemical transactions between components. Thus, a metabolic network is comprised of metabolites and biochemical reactions that catalyze transformations between them as the respective nodes and links. Its mathematical description culminates in a matrix format of stoichiometric coefficients where the rows are the metabolite components and the columns the reaction links, effectively representing a two-dimensional annotation of the genome [32]. This representation thus provides the mathematical context to quantitatively study human metabolism as a whole as well as by compartment (e.g. cell- or tissue-specific).

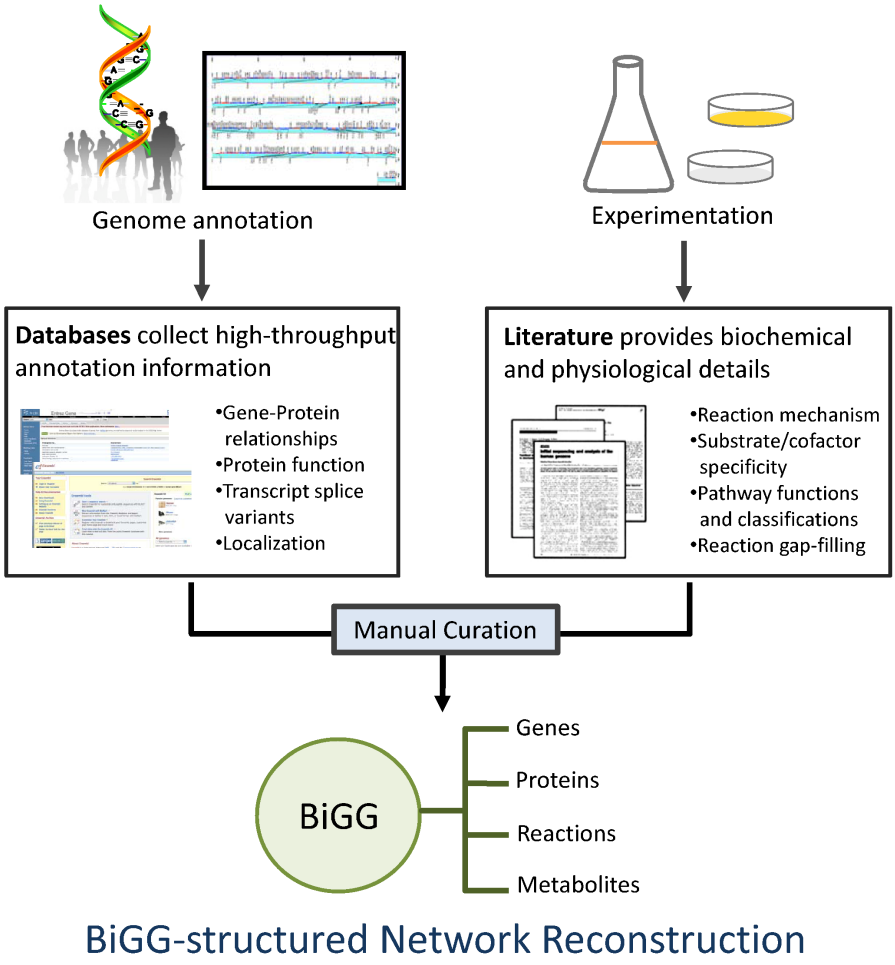


Figure 1.2: Incorporation of genomic and biochemical knowledge derived from the genome annotation and experimental literature into a BiGG-structured knowledge base network. High-throughput annotation data provides information on gene products, transcript variants, and their associated functions as well as localization (i.e. cellular compartment and tissue). Literature documents specific biochemical details from experiments on the gene product functions, such as reaction mechanism and substrate specificity.

1.3 Integrating high-throughput information and bottom-up systems biology

High-throughput technologies can now simultaneously measure high quality measurements of biomolecules in biological systems. The molecules of interest are primarily of the four main 'omes': the genome, transcriptome, proteome, and metabolome. Genomic analysis is performed by genotyping the sequence to accurately identify single nucleotide polymorphisms (SNPs) in the genome that are related to a particular phenotypic trait. Transcriptomic analysis is considered the most established of the high-throughput methodologies, in which the expression state of all gene transcripts are measured simultaneously on individual DNA probes on microarrays [33]. Proteomics analogously detects and identifies all proteins present in a cell using tandem mass spectroscopy. More recently, quantitative proteomics has been made possible through the use of isotope labeling techniques in tandem with mass spectroscopy to quantify relative protein abundances [34]. More recently, the development of new technologies has advanced towards acquiring metabolomic data [35]. The metabolome is the set of intracellular and extracellular metabolites present under a given physiological condition and is essentially the metabolic phenotype culminating from various upstream control points at the signaling, transcriptional, and proteomic levels. Taken together, these datasets represent the biological components at varying mechanistic levels that are simultaneously present to give rise to its resulting phenotype.

The vast amount of information generated by high-throughput technologies requires computational tools to facilitate their biological interpretation. Bottom-up network reconstructions are organized such that the incorporation and analysis of different data types can be done in the context of a genomically- and biochemically-structured framework (i.e. "context for content"). Since human metabolic networks represent global human metabolism, the networks can be further refined by datasets for specified conditions. The gene-protein-reaction structured format allows the integration of gene and protein expression data to determine metabolic reaction activity present for a given condition [28]. For instance, a method exploiting this format was recently used to predict tissue-specific metabolism based on gene expression and proteomic information mined from various public databases [36]. By integrating tissue-specific high-throughput data with a global metabolic network, specific metabolic activities for 10 different tissues were accurately predicted and determined a significant role for post-transcriptional regulation in tissue-specific metabolic phenotypes.

In addition to gene and protein expression data, metabolomic information can also be directly incorporated as they are explicitly represented in metabolic networks. High-throughput metabolomics has become increasingly available and provide quantitative information on various metabolic states. For example, the HMDB database contains detailed physiological and disease concentration metabolite levels from several biofluids and tissues [37]. Since metabolic networks describe the mechanistic relationships between metabolites, they can be used to systematically identify underlying pathways associated with the measured metabolites. Such applications are an alternative to the standard, top-down statistical approaches used to analyze high-throughput "omics" data and enable the interpretation of high-throughput data in the context of a mechanistically-detailed modeling framework.

1.4 Dissertation outline

"Omic" data sets represent simultaneous measurements for thousands of biochemical molecules, and understanding the relationships between these different data types will require a systems approach. Genome-scale network reconstructions have been built for metabolism in a number of target microorganisms [38], and constraint-based computational methods applied to these networks have proven successful in evaluating many aspects of metabolic phenotypes. Recent studies have highlighted bottom-up network analysis of high-throughput data as an potential alternative to standard statistical approaches as it provides a context in which the relationships between the components have already been defined. Therefore, these networks provide a "context for content" to systematically characterize phenotypic behavior and offer an exciting prospect for using such an approach to better understand complex phenotypes of higher-order organisms since a large amount of information already exists and continues to be generated.

This dissertation focuses on the integrative analysis of gene expression and metabolomic data using constraint-based network approaches in eukaryotic metabolism. The work presented herein can be categorized into three general areas: (1) the bottom-up network reconstructions for yeast and human metabolism as contexts for integration and analysis of high-throughput data; (2) the development of method approaches to analyze quantitative transcriptomic and metabolomic data in the context of network reconstructions; and (3) the application of these methods on genome-scale network reconstruction to characterize distinguishing metabolic features under physiological and perturbation conditions.

Chapter two introduces genome-scale network reconstructions of yeast metabolism, which serves as a model organism for reconstructing and analyzing eukaryotic networks. Examples of studies that were enabled by a previous genome-scale yeast reconstruction, iND750, are discussed, followed by the reconstruction process and validation of the most current yeast metabolic network, iMM904.

Chapter three discusses the first global reconstruction of human metabolism, Recon 1, and some of the initial applications that been used to study the network. It begins by introducing studies on an organelle-level human metabolic network that illustrates the promising uses of a smaller human metabolic network. The reconstruction process and initial applications of the global human metabolic network which followed are discussed.

Chapter four presents a study on integrating pharmacogenomic response data with the global network of human metabolism to study drug-specific metabolic response signatures. It discusses the use of a constraint-based approach to assess the overall network effects of metabolic intervention of candidate drug molecules and highlighting metabolic pathways as proposed mechanisms of action.

In chapter five, extracellular metabolome data measured in response to environmental and genetic perturbations of ammonium assimilation pathways was integrated with the updated iMM904 yeast network to characterize intracellular metabolic pathways that are broadly perturbed. The chapter presents a proof-of-concept study that shows how extracellular metabolomic data can be incorporated with a genome-scale network to evaluate the metabolic behavior of the intracellular system based on metabolite-level constraints.

Chapter six demonstrates the integration of metabolomic data with the genome-scale human metabolic network to characterize the stemness phenotype of embryonic stem cells. A constraint-based approach following similar principles from the previous chapter is used in evaluating the network behavior of stem cell metabolism, and experimental studies further supported the results from the analysis.

1.5 Terminology and Definitions

Bibliome: the comprehensive compilation of published literature text and associated information.

Genome-scale model: a computational model converted from a genome-scale network reconstruction by defining and implementing mathematical parameters to calculate phenotypic behavior of the reconstructed organism.

High-throughput technology: technological methods that enable rapid large-scale, high-quality measurements of biomolecules (i.e. genome, transcriptome, proteome, metabolome) for scientific experimentation.

Genome-scale network reconstruction: a genome annotation-based, manually curated assembly of biochemical transformations which describe specific biological processes (e.g. metabolism, transcriptional regulation) that exists in a specific organism or cell-type.

Knowledge base: the collection of knowledge information (i.e. from databases, literature, etc.) stored as a logically organized structure that represents relationships between the stored information.

Systems biology: an integrative field of research that combines experimental and computational approaches to systematically study relationships and interactions between various components of a biological system.

The text of this chapter, in full, is a reprint of the material as it appears in M.L. Mo and B.Ø. Palsson. 2009. Understanding human metabolic physiology: A genome-to-systems approach. Trends in Biotechnology, 27(1):37-44. I was the primary author of this publication and the co-author participated and supervised the research, which forms the basis for this chapter.

Chapter 2

Yeast as a model system for genome-scale modeling in eukaryotes

Saccharomyces cerevisiae has been one of the most important model organisms to study eukaryotic cellular processes. It has been an ideal experimental system to study primarily because of its ability to be genetically manipulated as well as its ability to grow under various culture conditions. As a result, a large availability of high-throughput and physiological experimental data for *S. cerevisiae* can be integrated and analyzed in a modeling framework [39, 40]. With the completion of its annotated genome sequence [41], the reconstruction of the first genome-scale eukaryotic metabolic network, iFF708. was possible. The network consisted of 708 genes, 1175 metabolic reactions, and 733 metabolites compartmentalized in the cytosol and mitochondria [42], and was used to calculate key physiological parameters and behavior that were found to accurately recapitulate physiological experimental data [42, 43].

This chapter presents a brief overview of the genome-scale studies resulting from an expanded yeast reconstruction, followed by the reconstruction of the most recent *S. cerevisiae* metabolic network, iMM904 and its validation as a modeling framework for yeast metabolism.

2.1 Examples uses of the expanded iND750 yeast reconstruction

Subsequent expansions of the iFF708 networks have since progressed over the years. An expanded network, iND750, was reconstructed based on the iFF708 network to include more detailed descriptions of additional cellular compartments, defined logical relationships between genes, proteins, and reactions, as well as charge- and elementally balanced metabolic reactions [44]. The iND750 yeast reconstruction represented the first computable genome-scale eukaryotic metabolic network to capture detailed relationships between genetic and biochemical components as well as comprehensive cellular localization of charge- and elementally-balanced metabolic reactions. A more recent version of the yeast metabolic network, iMM904, was expanded to account for 904 genes, 1,223 metabolites, and 1,412 reactions with detailed expansion of lipid and carbohydrate metabolism as well as pathways involved in the metabolism and transport of observed secreted metabolites.

The continual process of improving and updating network reconstructions has provided the essential basis for enabling the analysis of high-throughput data. One such example of an integrative use of the iND750 network was described in a study that utilized computational and experimental validation approaches to uncover novel regulatory and metabolic interactions. The iND750 network provided a central framework for integrative analysis of known regulatory mechanisms to construct the first transcriptional regulatory network for eukaryotic metabolism. The resulting integrated metabolic/regulatory network, iMH805/775, accounted for 805 genes (55 transcription factors and 750 metabolic genes) and 775 regulatory interactions, with 82 intra- and extracellular metabolites act as input signals to the regulatory network [45]. In addition to representing regulatory interactions, iMH805/775 also included rules describing the mode of combinatorial control by different transcription factors (TF) at each promoter. The regulatory network represented an integration of genome-wide datasets on protein-DNA interactions (ChIP-chip) and TF binding motifs by combining ChIP-chip data, binding site motifs on promoters, and expression change in TF knock-out/overexpression strains. By combining experimental and computational approaches,novel regulatory mechanisms for specific genes were identified to be involved in utilization of particular carbon sources such as PGM2 (galactose) and HXT1 (mannose) [45]. Regulatory interactions identified through this process were included in iMH805/837 and demonstrated that a systematic approach can be used to analyze different types of high-throughput datasets and fill in missing eukaryotic regulatory mechanisms.

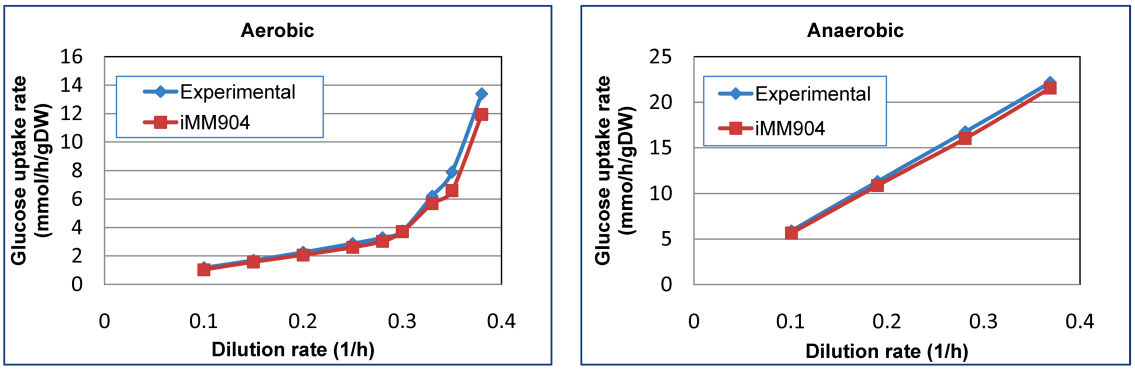


Figure 2.1: Comparison of experimental and predicted aerobic and anaerobic glucose uptake rates at different dilution rates.
Measured experimental and computed iMM904 values were plotted and shown to be comparably similar under both conditions.

Another study using the iND750 network was used to derive a systematic condition-dependent annotation (CDA) of metabolic genes that reflects the context-dependent nature of their functionality [46]. Based on the CDA predictions, novel functionally annotated pathways were derived and experimentally verified using substrate auxotrophy measurements of single and double gene deletion strains. For example, the CAR2 gene which that was annotated to be involved in arginine biosynthesis was found to also interact in proline biosynthesis, a finding that had not been previously annotated [46]. Thus, metabolic network reconstructions can serve as a framework to systematically evaluate metabolic gene annotations as a function of its environmental condition.

A more recent version of the yeast metabolic network, iMM904, was expanded to account for 904 genes, 1,223 metabolites, and 1,412 reactions with detailed expansion of lipid and carbohydrate metabolism as well as pathways involved in the metabolism and transport of observed secreted metabolites. As shown in previous studies, the continual process of improving and updating network reconstructions has provided the essential basis for enabling the analysis of high-throughput data.

2.2 Method approaches to constructing and validating the iMM904 network model

2.2.1 Reconstruction methods

The previously reconstructed iND750, a fully compartmentalized and elementally-balanced S. cerevisiae metabolic network, was used as the basis for reconstructing the iMM904 network [44]. The network was further expanded to include additional genes and reactions based on genomic, biochemical, and physiological information [see Additional file S1]. The details of existing reactions (substrate and cofactor specificity, reaction reversibility, and compartmentalization) in the iND750 network were also re-evaluated to update the model based on existing literature. The iMM904 network was reconstructed using the SimPheny modeling software (Genomatica Inc, San Diego, CA). Existing gene-protein-reaction (GPR) associations from iND750 were also reviewed and several were modified to include additional genes and proteins. GPR associations are Boolean representations of the logical relationship between ORFs and their corresponding transcripts, proteins, and reactions to enable mapping of genes to their respective functions.

2.2.2 Methods of converting network to model

The network reconstruction was converted to a constraint-based model using established procedures [47]. Network reactions and metabolites were assembled into a stoichiometric matrix S containing the stoichiometric coefficients of the reactions in the network. The steady-state solution space containing possible flux distributions is determined by calculating the null space of S: $S \cdot v = 0$, where v is the reaction flux vector. Minimal media conditions were set through constraints on exchange fluxes corresponding to the experimental measured substrate uptake rates. All the model-based calculations were done using the Matlab COBRA Toolbox [47] utilizing the glpk or Tomlab/CPLEX (Tomopt, Inc.) optimization solvers.

2.3 Method approaches to validating the iMM904 network model

2.3.1 Chemostat data validation

The iMM904 model was initially validated by simulating wild type yeast growth in aerobic and anaerobic carbon-limited chemostat conditions and comparing the simulation results to published experimental data on substrate uptake and byproduct secretion in these conditions [48]. The study was performed following the approach taken to validate the iFF708 model in a previous study [43]. The predicted glucose uptake rates were determined by setting the in silico growth rate to the measured dilution rate, which are equivalent under continuous culture growth, and minimizing the glucose uptake rate. The accuracy of in silico predictions of substrate uptake and byproduct secretion by the iMM904 model was similar to the accuracy obtained using the iFF708 model and results are shown in Figure 2.1.

2.3.2 Genome-scale gene deletion validation

The iMM904 network was further validated by performing genome-scale gene lethality computations following established procedures to determine growth phenotypes under minimal medium conditions and compared to published data. A modified version of the biomass function used in previous iND750 studies was set as the objective to be maximized and gene deletions were simulated by setting the flux through the corresponding reaction(s) to zero. The biomass function was based on the experimentally measured composition of major cellular constituents during exponential growth of yeast cells and was reformulated to include trace amounts of additional cofactors and metabolites with the assumed fractional contribution of 10^{-6}. These additional biomass compounds were included according to the biomass formulation used in the iLL672 study to improve lethality predictions through the inclusion of additional essential biomass components [49]. The model was constrained by limiting the carbon source uptake to 10 mmol/h/gDW and oxygen uptake to 2 mmol/h/gDW. Ammonia, phosphate, and sulfate were assumed to be non-limiting. The experimental phenotyping data was obtained using strains that were auxotrophic for methionine, leucine, histidine, and uracil. These auxotrophies were simulated by deleting the appropriate genes from the model and supplementing the in silico strain with the appropriate supplements at non-limiting, but low levels. Furthermore, trace amounts of essential

nutrients that are present in the experimental minimal media formulation (4-aminobenzoate, biotin, inositol, nicotinate, panthothenate, thiamin) were supplied in the simulations [49].

Three distinct methods to simulate the outcome of gene deletions were utilized: Flux-balance analysis (FBA) [1, 50, 51], Minimization of Metabolic Adjustment (MoMA) [52], and a linear version of MoMA (linearMoMA). In the linearMoMA method, minimization of the quadratic objective function of the original MoMA algorithm was replaced by minimization of the corresponding 1-norm objective function (i.e. sum of the absolute values of the differences of wild type FBA solution and the knockout strain flux solution). The computed results were then compared to growth phenotype data (viable/lethal) from a previously published experimental gene deletion study [49].

The comparison between experimental and in silico deletion phenotypes involved choosing a threshold for the predicted relative growth rate of a deletion strain that is considered to be viable. We used standard ROC curve analysis to assess the accuracy of different prediction methods and models across the full range of the viability threshold parameter, with results shown in Figure 2.2. The ROC curve plots the true viable rate against the false viable rate thus allowing comparison of different models and methods without requiring arbitrarily choosing this parameter *a priori* [53]. The optimal prediction performance corresponds to the point closest to the top left corner of the ROC plot (i.e. 100% true viable rate, 0% false viable rate). The values reported in Table 2.1 correspond to selecting the optimal viability threshold based on this criterion. We summarized the overall prediction accuracy of a model and method using the Matthews Correlation Coefficient (MCC) [53]. The MCC ranges from -1 (all predictions incorrect) to +1 (all predictions correct) and is suitable for summarizing overall prediction performance in our case where there are substantially more viable than lethal gene deletions. ROC plots were produced in Matlab (Mathworks, Inc.).

2.4 Reconstruction content of the iMM904 network

A previously reconstructed *S. cerevisiae* network, iND750, was used as the basis for the construction of the expanded iMM904 network. Prior to its presentation here, the iMM904 network content was the basis for a consensus jamboree network that was recently published but has not yet been adapted for FBA calculations [54]. The majority of iND750 content was carried over and further expanded on to construct iMM904, which accounts for 904 genes, 1,228 individual metabolites, and 1,412 reactions of which 395 are transport

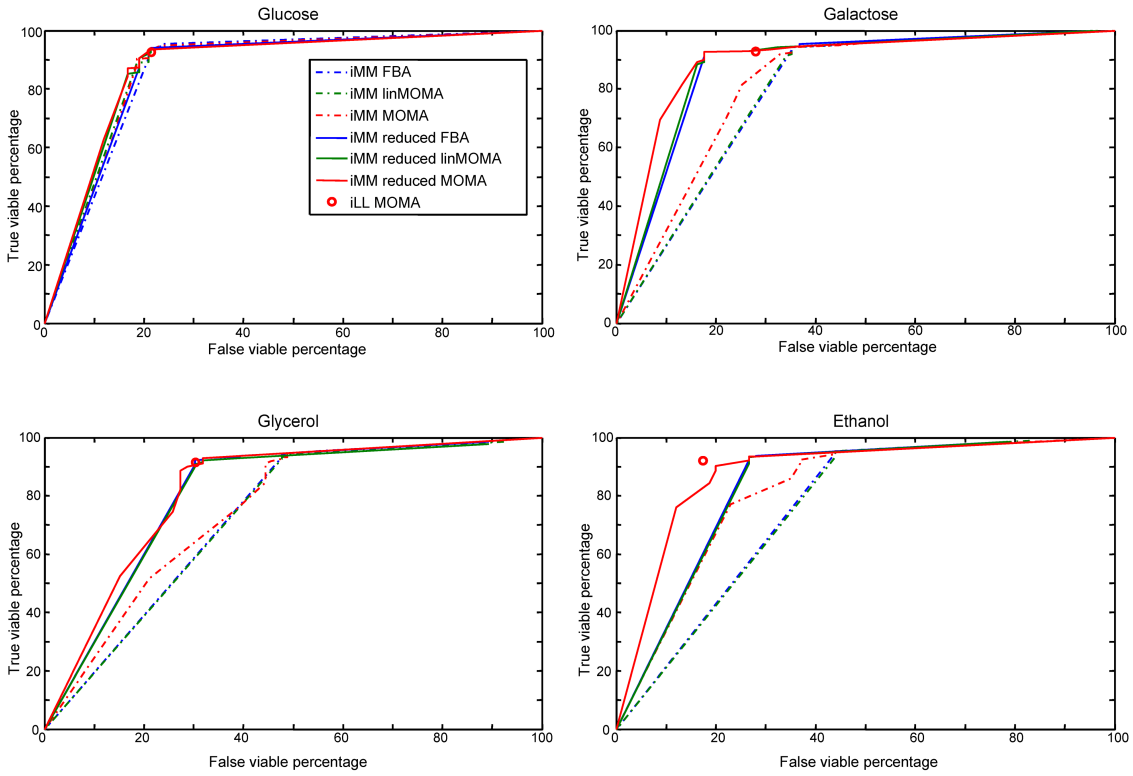


Figure 2.2: ROC curve plots of iMM904 and iLL672 growth predictions using different optimization analysis methods.

reactions. Both the number of gene-associated reactions and the number of metabolites increased in iMM904 compared with the iND750 network. Additional genes and reactions included in the network primarily expanded the lipid, transport, and carbohydrate subsystems. The lipid subsystem includes new genes and reactions involving the degradation of sphingolipids and glycerolipids. Sterol metabolism was also expanded to include the formation and degradation of steryl esters, the storage form of sterols. The majority of the new transport reactions were added to connect network gaps between intracellular compartments to enable the completion of known physiological functions. We also added a number of new secretion pathways based on experimentally observed secreted metabolites [55].

A number of gene-protein-reaction (GPR) relationships were modified to include additional gene products that are required to catalyze a reaction. For example, the protein compounds thioredoxin and ferricytochrome C were explicitly represented as compounds in iND750 reactions, but the genes encoding these proteins were not associated with their corresponding GPRs. Other examples include glycogenin and NADPH cytochrome p450 reductases (CPRs), which are required in the assembly of glycogen and to sustain catalytic activity in cytochromes p450, respectively. These additional proteins were included in iMM904 as part of protein complexes to provide a more complete representation of the genes and their corresponding products necessary for a catalytic activity to occur.

Major modifications to existing reactions were in cofactor biosynthesis, namely in quinone, beta-alanine, and riboflavin biosynthetic pathways. Reactions from previous *S. cerevisiae* networks associated with quinone, beta-alanine, and riboflavin biosynthetic pathways were essentially inferred from known reaction mechanisms based on reactions in previous network reconstructions of E. coli [42, 44]. These pathways were manually reviewed based on current literature and subsequently replaced by reactions and metabolites specific to yeast. Additional changes in other subsystems were also made, such as changes to the compartmental location of a gene and its corresponding reaction(s), changes in reaction reversibility and cofactor specificity, and the elucidation of particular transport mechanisms.

2.5 Predicting deletion growth phenotypes for genome-scale validation

The updated genome-scale iMM904 metabolic network was validated by comparing in silico single-gene deletion predictions to in vivo results from a previous study used to analyze another S. cerevisiae metabolic model, iLL672 [49]. This network was constructed

based on the iFF708 network [42], which was also the starting point for reconstructing the iND750 network [44]. The experimental data used to validate the iLL672 model consisted of 3,360 single-gene knockout strain phenotypes evaluated under minimal media growth conditions with glucose, galactose, glycerol, and ethanol as sole carbon sources. Growth phenotypes for the iMM904 network were predicted using FBA [1, 50, 51], MoMA [52], and linear MoMA methods as described in Methods and subsequently compared to the experimental data (Table 2.1). Each deleted gene growth prediction comparison was classified as true lethal, true viable, false lethal, or false viable. The growth rate threshold for considering a prediction viable was chosen for each condition and method separately to optimize the tradeoff between true viable and false viable predictions (maximum Matthews correlation coefficient, see Methods).

Since iMM904 has 212 more genes than iLL672 with experimental data, we also present results for the subset of iMM904 predictions with genes included in iLL672 (reduced iMM904 set). When the same gene sets are compared, iMM904 improves gene lethality predictions under glucose, galactose, and glycerol conditions over iLL672 somewhat, but is less accurate at predicting growth phenotypes under the ethanol condition. It should be noted that the iLL672 predictions were obtained directly from [49] and thus the growth rate threshold was not optimized similarly to iMM904 predictions. Overall, when viability cutoff is chosen as indicated above for each method separately, the three prediction methods (FBA, MOMA, and linear MOMA) perform similarly.

While the full gene complement in iMM904 greatly increased the number of true viable predictions, the full model also made significantly more false viable predictions compared with reduced iMM904 and iLL672 predictions. However, it is important to note that 143 reactions involved in dead-end biosynthetic pathways were actually removed from iFF708 to build the iLL672 reconstruction [49]. These dead-ends are considered knowledge gaps in pathways that have not been fully characterized and, as a result, lead to false viable predictions when determining gene essentiality if the pathway is in fact required for growth under a certain condition [44, 56]. As more of these pathways are elucidated and included in the model to fill in existing network gaps, we can expect false viable prediction rates to consequently decrease. Thus, while a larger network has a temporarily reduced capacity to accurately predict gene deletion phenotypes, it captures a more complete picture of currently known metabolic functions and provides a framework for network expansion as new pathways are elucidated [30].

Table 2.1: Comparison of iMM904 (full and reduced) and iLL672 gene deletion predictions and experimental data under minimal media conditions.

Media	**Model**	**Method**	**True viable**	**False viable**	**False lethal**	**True lethal**	**MCC**
Glucose	iMM904 full	FBA	647	10	32	33	0.60
	iMM904 full	linMOMA	644	10	35	33	0.58
	iMM904 full	MOMA	644	10	35	33	0.58
	iMM904 reduced	FBA	440	9	28	33	0.61
	iMM904 reduced	linMOMA	437	9	31	33	0.60
	iMM904 reduced	MOMA	437	9	31	33	0.60
	iLL672 full	MOMA	433	9	35	33	0.57
Galactose	iMM904 full	FBA	595	32	36	59	0.58
	iMM904 full	linMOMA	595	32	36	59	0.58
	iMM904 full	MOMA	595	32	36	59	0.58
	iMM904 reduced	FBA	409	12	33	56	0.67
	iMM904 reduced	linMOMA	409	12	33	56	0.67
	iMM904 reduced	MOMA	409	12	33	56	0.67
	iLL672 full	MOMA	411	19	31	49	0.61
Glycerol	iMM904 full	FBA	596	43	36	47	0.48
	iMM904 full	linMOMA	595	44	37	46	0.47
	iMM904 full	MOMA	598	44	34	46	0.48
	iMM904 reduced	FBA	410	20	34	46	0.57
	iMM904 reduced	linMOMA	409	21	35	45	0.56
	iMM904 reduced	MOMA	412	21	32	45	0.57
	iLL672 full	MOMA	406	20	38	46	0.55
Ethanol	iMM904 full	FBA	593	45	29	55	0.54
	iMM904 full	linMOMA	592	45	30	55	0.54
	iMM904 full	MOMA	592	44	30	56	0.55
	iMM904 reduced	FBA	408	21	27	54	0.64
	iMM904 reduced	linMOMA	407	21	28	54	0.63
	iMM904 reduced	MOMA	407	20	28	55	0.64
	iLL672 full	MOMA	401	13	34	62	0.68

2.6 Recapitulation

The continual update and expansion of the *S. cerevisiae* metabolic reconstruction has resulted in high-quality, well-curated networks that provide a mathematical structure for modeling experimental data. The development of these network reconstructions have driven the advancement of methods used to interrogate network properties and have shown to improve the scope and reliability of biological studies. The successful development of genome-scale yeast reconstructions have shown that such networks can enable a systems-level understanding of eukaryotic metabolism. The lessons learned in developing the yeast networks provide the foundation for building similar computational frameworks for human metabolism that have the potential to advance research and development in disease and health.

The text of this chapter, in part, is a reprint of the material as it appears in M.L. Mo, M.J. Herrgard, and B.Ø. Palsson. 2009. Connecting extracellular metabolomic profiles to intracellular metabolic states in yeast. BMC Systems Biology. 3:37. I was the primary author of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

Chapter 3

Towards constraint-based modeling of human metabolism

The procedures for reconstructing genome-scale networks and extending its capabilities to a non-kinetic, constraint-based model are now well-established, enabling hypothesis-driven biology and the analysis of metabolic capabilities in microorganisms [25]. With a more complex system in humans than in single-celled organisms, there is a greater need for the application of constraint-based models to understand the basis of normal and pathophysiological metabolism. As shown in the last chapter, the development of the genome-scale yeast metabolic reconstruction has served as a foundational platform for systems modeling of eukaryotic metabolism. The approaches developed for integrative analysis of genome-scale yeast metabolic reconstructions led to an analogous effort to reconstructing and studying human metabolism.

The completion of the first genome-scale human metabolic network, Recon 1, is a key step to applying computational systems biology in a biomedical context [3]. As a comprehensive, literature-based reconstruction of human metabolism, Recon 1 accounts for 1,496 genes encoding for proteins, 2,004 proteins, 2,766 metabolites, and 3,311 metabolic and transport reactions. This network reconstruction was converted into an *in silico* model of human metabolism and validated through the simulation of 288 known metabolic functions found in a variety of cell and tissue types. This chapter presents a brief overview of limited human metabolic networks, the reconstruction process and completion of the *H. sapiens* Recon 1 network, and early network applications adapted to understand human metabolism at a systems-level.

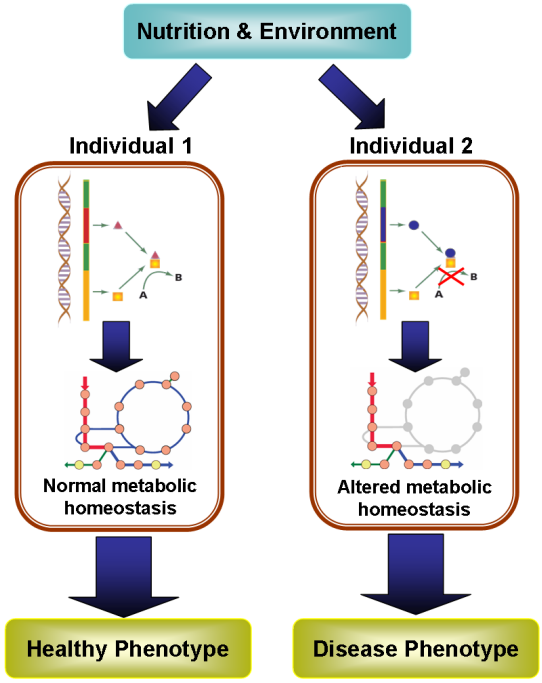


Figure 3.1: Metabolic phenotype as a consequence of the interactions between external (environmental and nutritional) and internal (e.g. genetic and proteomic) factors.
While both individual 1 and 2 are subjected to the same external factors, different metabolic phenotypes arise from the unique interactions between external factors and an individual's internal factors. These differences can result in a healthy phenotype for individual 1 and a disease phenotype for individual 2. Due to the large size and complexity of the interactions and resulting networks, computational models will be critical for mechanistic analysis and interpretation of results.

3.1 Metabolism as a complex system

Metabolism plays a central role in determining the functional state within a cell, underscored by its involvement in integral cellular processes. It is an interconnected system of chemical transformations necessary to sustain various cellular functions and includes the degradation of molecules for energy production and the assembly of essential cellular constituents. The components in metabolism consist of thousands of metabolites, molecules that are either synthesized through anabolic processes or degraded in catabolic pathways, and the enzymes that catalyze these biochemical conversions. Together, they form intricate interaction networks that govern cellular behavior which ultimately determines how the body functions.

The complex nature of human metabolism is further evinced when considering diseases with complex genotype-phenotype relationships, such as obesity and Type 2 diabetes, which have emerged as prevalent epidemics in the U.S. and increasingly on a worldwide basis [57, 58]. Steps have been made towards the understanding of genotype-phenotype relationships in metabolic syndrome through the identification of genes resulting in metabolic dysfunction [57, 59, 60, 61]. While a genetic predisposition exists in acquiring these disorders, environmental factors are often a triggering influence on their development, which has led to more recent studies on how nutrient-gene interactions alter metabolic homeostasis [62, 63]. The development of metabolic disorders is essentially the consequence of an intricate interplay between genetic, environmental, and nutritional factors (Figure 3.1). It has thus become clear that this is a multifaceted problem that requires a systemic approach to explore a system of causes.

3.2 Analyzing human metabolism in systems biology

Reconstruction of genome-scale human metabolic networks has initiated development towards studying human physiology *in silico* at a systems-level. There are four crucial steps to this process which have been described [1], with the biological representation becoming more focused and detailed at each level and thus culminating in a systems framework to analyze and model human metabolic phenotypes (Figure 3.2).

Two general approaches used to study metabolic network systems are (1) topology-based analysis and (2) *in silico* modeling. Metabolic networks are data structures describing known biochemical interactions between metabolites and, as such, can be used to define structural, or topological, features of metabolism. A key motivation for understanding net-

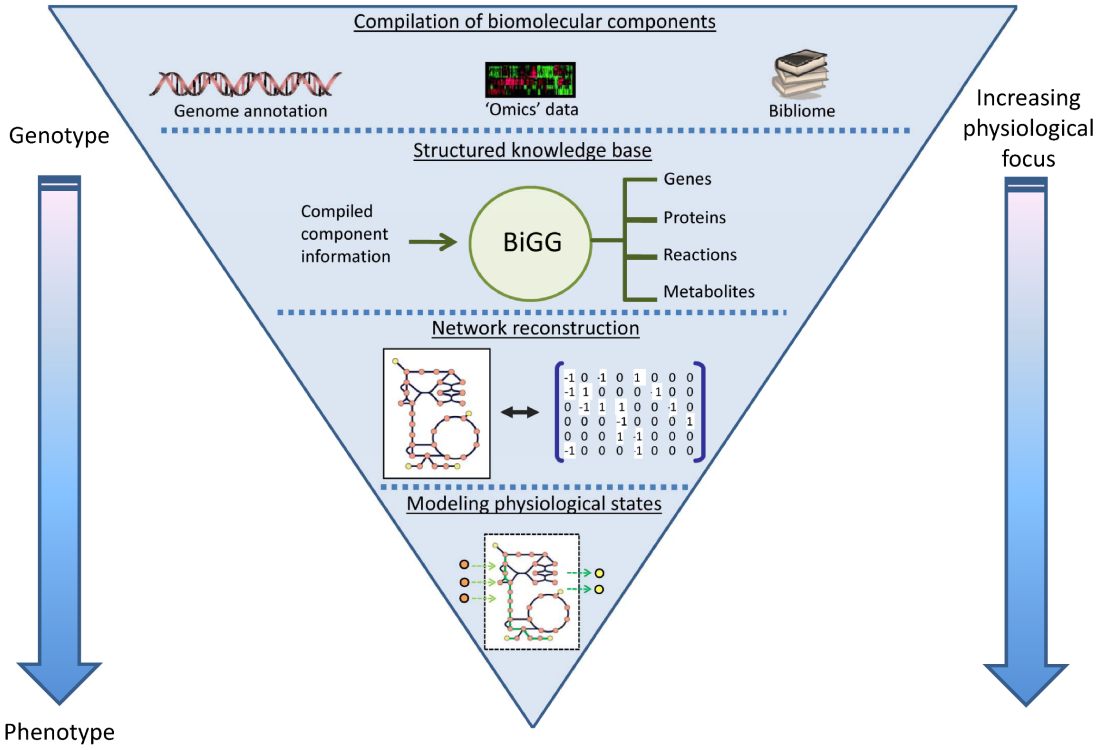


Figure 3.2: The four general steps of bottom-up systems biology that enables studying a physiological system *in silico*.

Components from various biological sources is first compiled and incorporated into a BiGG structured format. The assembled information becomes the basis for a network reconstruction specific to the system of interest. The network can be converted into a computational model by imposing mathematical parameters relevant to specific biological conditions (e.g. genetic, environmental). Each step in the process formulates a more detailed *in silico* representation of the studied system, thus increasing the level of physiological focus and refining the genotype-phenotype relationship.

work topology is that metabolic functions, and its corresponding malfunctions, are thought to be determined in part by how their interactions are structured. This topology-function relationship has already been extensively studied in microbial networks [64, 65, 66, 67, 68, 69] as well as in smaller-scale human metabolic and signaling networks [70, 71, 72, 73], and a similar approach can be readily applied towards understanding human physiological and disease functions at a larger scale.

The imposition of modeling parameters adds another dimension to analyzing genome-scale networks. By formulating mathematical constraints corresponding to different physiological environments, a network can be adapted into a computational, or *in silico*, model of metabolism for a specified condition. While a network itself can be used to look at connectivity properties, its converted model form can be used to study its functional metabolic states. Various mathematical procedures have already been developed that explore systems properties, in particular around the steady-state, of genome-scale models [25, 47]. Once in its *in silico* form, the genome-scale model can be used to perform a number of biological studies as demonstrated by the historical use of the E. coli metabolic reconstructions [26, 74]. Progress has already been made to adapt computational approaches to study smaller human metabolic networks, including calculating candidate metabolic states under diabetic, ischemic, and dietetic conditions in the cardiomyocyte mitochondria [75] and simulating effects of plasma environment on hepatocyte metabolism [76].

3.3 Preliminary work at the organelle-scale: the human cardiomyocyte mitochondria

Prior to the establishment of Recon 1, a metabolic model of the cardiomyocyte mitochondria was reconstructed and used successfully in the systemic analyses of various biological properties [77]. The network was constructed predominantly using proteomic and biochemical data and consisted of 189 elementally and charge-balanced reactions. Three functional capabilities were assessed (ATP production, heme synthesis, and phospholipid synthesis), wherein flux variability analysis was used to characterize inherent network flexibility when each metabolic function was carried out [77]. The study indicated that a large portion of mitochondrial network reactions was accessible for optimal heme and phospholipid biosyntheses while it was severely restrictive for ATP production.

A study on diabetic, ischemic, and dietetic metabolic states was performed in which random sampling of the steady-state solution space was used to calculate candidate

flux distributions based on imposed constraints derived from experimental data [75]. Subsequent simulations demonstrated how reaction flux distributions, reaction correlations, and network flexibility under these conditions deviated from the normal physiologic condition. Interestingly, the flux through pyruvate dehydrogenase (PDH) was found to be severely limited under a diabetic state in comparison to the normal condition. This was in agreement with published observations that PDH activity is often restricted in diabetics [78], thus suggesting that PDH inhibition *in vivo* may be due in part to stoichiometric restrictions rather than entirely to regulatory actions.

One approach in constraint-based models to simplify metabolic networks includes the calculation of correlated reaction sets (co-sets), sets of reactions that are always active or inactive together due to mass conservation constraints. This means that a flux through one reaction implies a flux through other reactions in the co-set (conversely if there is a reaction without any net flux in one reaction, all of the other reactions in the co-set do not carry any flux). These co-sets can be calculated by running a series of optimizations as in flux coupling [79] or by sampling the flux solution space and calculating the perfectly correlated sets of reactions [75]. Mapping of calculated mitochondrial co-sets to diseases described in the Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/) enabled the classification and correlation of causal SNPs in a systemic and mechanistic fashion [80]. Most of the co-sets involved contiguous sets of reactions, but those involving non-contiguous groups of reactions could potentially provide novel insights. One such example involved a reaction co-set found in the urea cycle, in which deficiencies due to different reactions presented similar phenotypes for three out of the four reactions in the co-set. The results from this study indicated that network-wide analysis of reaction co-sets can be used to elucidate the genotype-phenotype relationship.

3.4 *H. sapiens* Recon 1: A Genome-scale Network Reconstruction of Global Human Metabolism

3.4.1 Building the Recon 1 network

The overall procedure for the reconstruction has previously been described in detail [3] and will be discussed in brief while highlighting the major differences from previous metabolic network reconstructions. The initial component list based on the genome annotation of Build 35 provided the basic framework for the whole reconstruction and was

followed by a manual curation phase which involved perusing through over 1,500 citations in the "bibliome" in addition to the use of available internet resources and databases. These manual curation steps (outlined in Figure 3.3) are unchanged in comparison to the standardized approach and accounts for the most time-consuming process.

The reconstruction effort was carried out in a parallelized manner divided among six researchers based on metabolic pathways. Like the other metabolic reconstructions, the reconstruction and model building process involved iterative steps; however, these debugging steps were more involved in building Recon 1 due to the fact that multiple researchers were reconstructing different parts of the network at the same time and a global cell was being reconstructed, not a single cell. The latter point was addressed by determining 288 bibliome-defined, metabolic objectives that exist in the human body, which include basic physiological functions such as ATP production and ketogenesis [3]. In addition, an internal QC/QA process was maintained in which changes and additions were tracked and each metabolic function was re-simulated at every iteration, thus ensuring that newer model iterations did not lose functionality. These efforts resulted in a BiGG-structured reconstruction of global human metabolism.

3.4.2 Characterizing the knowledge landscape

The manual curation procedure combined with the testing of functional capabilities resulted in the definition of the knowledge landscape for human metabolism. Almost all of the reactions were assigned confidence scores which provided an indication of the strength of evidence for making the gene annotation and the associated reactions. Experimental evidence, such as demonstrated enzymatic activity, had the highest confidence scores, whereas reactions that were added due to modeling evidence (i.e. a simulation could not be carried out without the inclusion of a particular reaction) received the lowest. Once the reconstruction was complete, it was possible to look over the whole set of reactions and classify the pathways in terms of confidence scores. These classifications are obviously subjective due to the semi-standardized pathway classifications. Pathway definitions, for example, are neither one-to-one nor completely standardized; however, it can highlight areas in which little experimental evidence is available. The dominant pathways in the areas with low confidence were intracellular transport reactions. This relatively neglected area of metabolism can have significant impact on the types of predictions made by the model. For example, the difference between active and passive transport can have quantitative and qualitative implications in energetic predictions. An interesting observation across many of

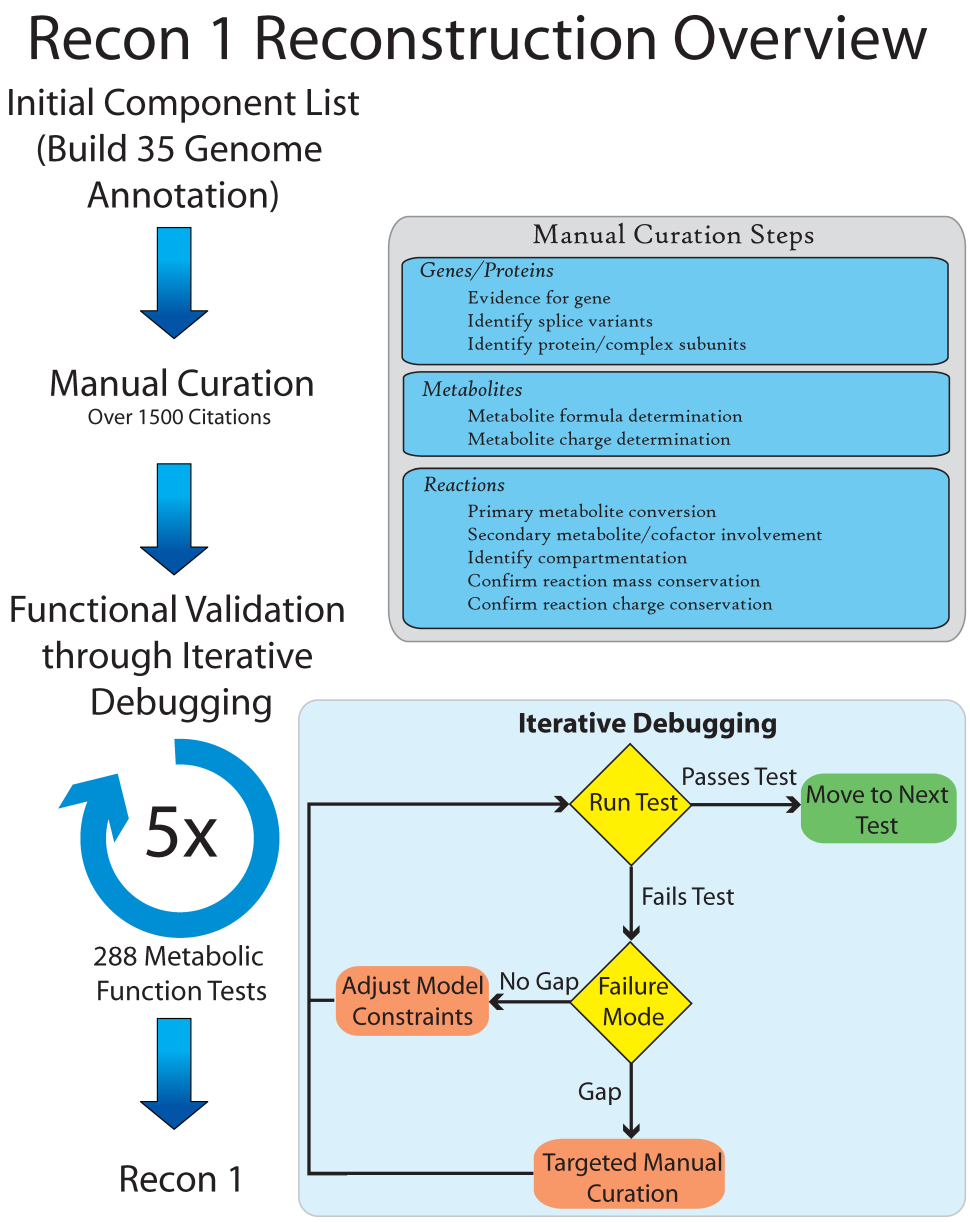


Figure 3.3: An overview of the Recon 1 reconstruction process from the Build 35 genome annotation.

An initial list of components was compiled from Build 35 and manually curated. The manual curation steps are among the most time consuming aspects of the reconstruction, however they also contribute significantly towards making a biologically accurate, functional model. The iterative debugging process was carried out for the 288 functional tests at each iteration to ensure proper physiological functioning of the Recon 1 metabolic network.

the pathways was that the knowledge landscape was uneven, that is, there were significant differences among all the pathways and within pathways as well. The dips, or so-called valleys, in the knowledge landscape direct areas in which further experiments can help make new contributions to the scientific infrastructure.

3.4.3 Identifying potential alternative drug targets

Recon 1 was analyzed under aerobic glucose conditions using flux coupling [79] to identify over 250 coupled reaction sets. One of the largest involved the cholesterol biosynthesis pathway, which passes multiple intracellular compartments (cytosol, endoplasmic reticulum, and peroxisome). 3-hydroxy-3-methylglutaryl-Coenzyme A (HMG-CoA) reductase, the metabolic target of the anti-lipidemic class of statin drugs, was in this coupled reaction set. As a variation on the study with the cardiac mitochondria which suggested that causal SNP associated diseases in one co-set have similar phenotypes [80], one can make the case for proposing alternative drug targets for reactions in the same co-set. With the HMG-CoA co-set, other members of the set are thus identified as potential alternative drug targets for treating hyperlipidemia. As demonstrated for the human mitochondria, deficiencies in enzymes belonging to the same functionally coupled reaction set may have similar phenotypes [80]. Another example with potential disease applications involved a coupled reaction set associated with glutathione metabolism in which multiple enzymes were causally associated with hemolytic anemia (when deficiencies of the enzymes were observed in patients). These examples demonstrate that *in silico* simulations can provide an analytical approach to study the causes and consequences of disease, leading to potential insight into new drug treatment targets.

3.4.4 Mapping and analyzing expression data

H. sapiens Recon 1 includes information about biological relationships between gene transcripts, their corresponding protein products, and the reactions these proteins catalyze. As such, gene expression profiles from the skeletal muscle of patients before and after bariatric surgery were mapped to the network [60]. This particular data set was valuable because gene expression profiling was carried out on the same patient before and after surgery, so each patient could serve as their own control. Rather than identifying a handful of genes that were statistically determined as up-regulated or down-regulated, all metabolic genes from each expression profile were mapped onto Recon 1 to observe overall expression

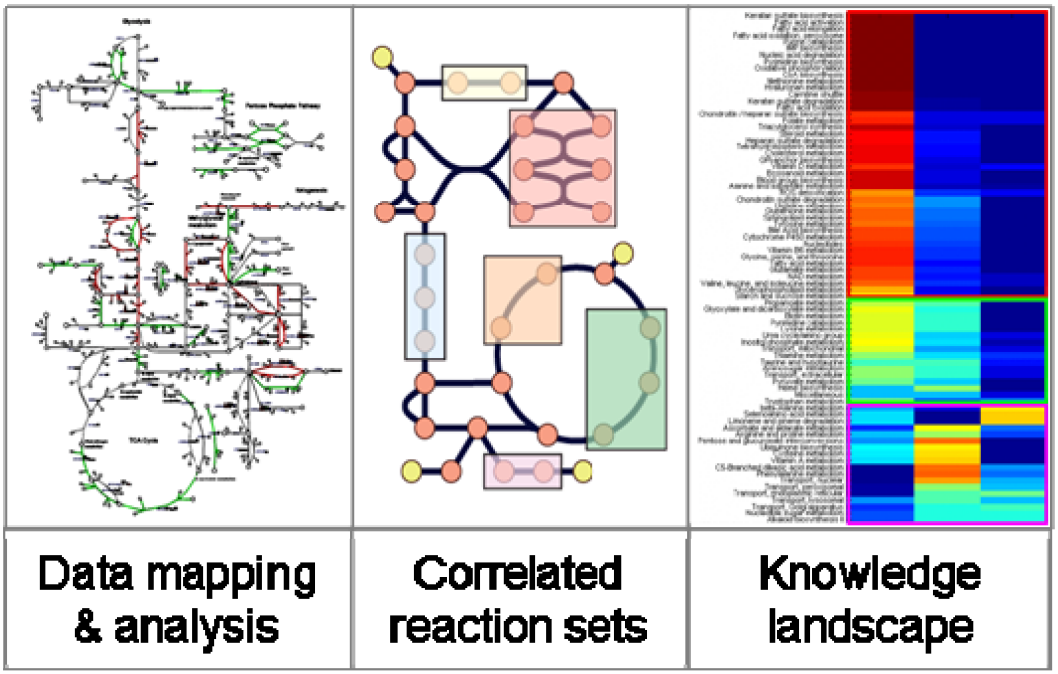


Figure 3.4: The three initial applications of the Recon 1 network to demonstrate its use. The mapping of high-throughput data can be mapped to the Recon 1 network through GPR associations to visualize metabolic trends. Correlated reaction sets, or co-sets, can be calculated to indicate reactions that are used together under a particular condition. A knowledge landscape describing the confidence of literature evidence used to reconstruct Recon 1 systematically characterizes areas of metabolism which are abundant or lacking in direct experimentation.

changes. Changes among the three patients were observed to be consistent with one another, wherein the overall patterns observed indicated an increased reliance on anaerobic pathways and decreased expression in enzymes involved in oxidative phosphorylation. These patterns were consistent with expression profiles in primates subjected to long-term caloric restriction [81] thus suggesting the muscle tissue in patients experience caloric restriction one year following surgery.

3.4.5 Network-based disease phenotype characterization

More recently, an expanded study was published on a metabolic disease network that was constructed to describe gene-disease phenotype relationships using the Online Mendelian Inheritance in Man (OMIM) database and Recon 1 [82]. A large-scale comparison with patient records indeed showed that adjacent reaction connectivity is correlated with co-expression of disease phenotypes and is thus a possible predictor of disease co-morbidity. Detailed analysis of disease associations provided additional insight into underlying reaction mechanisms contributing to a shared pathophysiology. An interesting example discussed the high association between diabetes and hemolytic anemia, which, upon further analysis, was explained by a NADPH deficiency via glucose-6-phosphate dehydrogenase leading to glutathione deficiency, resulting in hemolytic anemia. Such studies indicate that human metabolic network-based studies can expand into mechanistic analysis of disease relevant to diagnostic and therapeutic applications.

3.5 Recapitulation

The analysis of fundamental genotype-phenotype relationships in human biology has been initiated with the reconstruction of a global human metabolic network. Recon 1 represents a milestone in metabolic systems biology as a mathematically-structured knowledge base that enables systematic studies of human metabolism and its properties. Genome-scale metabolic reconstruction have shown to be useful, systematic frameworks to studying yeast metabolism, as discussed in the previous chapter. Similar approaches have be broadly extended to studying human metabolism, and preliminary work with both a smaller cardiomyocyte mitochondrial network and the global metabolic network have been promising. Hence, continued development of applications for the human network reconstruction will enhance its use in systems analysis and discovery.

The text of this chapter, in full, is a reprint of the material as it appears in M.L.

Mo, N. Jamshidi, and B.Ø. Palsson. 2007. A Genome-scale, Constraint-based Approach to Systems Biology of Human Metabolism., Molecular Biosystems. 3:9 and in M.L. Mo and B.Ø. Palsson. 2009. Understanding human metabolic physiology: A genome-to-systems approach. Trends in Biotechnology, 27(1):37-44. I was the primary author of these publications and the co-authors participated and/or supervised the research which forms the basis for this chapter.

Chapter 4

Characterizing drug response phenotypes in human metabolism

The pharmaceutical industry aims to develop drugs that achieve maximum therapeutic benefits for the condition treated while minimizing adverse side effects. Research efforts have historically focused on identifying "druggable" molecular components for therapeutic intervention based on binding and accessibility characteristics of the identified target [83, 84]. Physiological understanding of drug intervention, however, requires not only evaluating static effects on the target itself but also the dynamic consequence on overall cellular responses resulting from both direct and indirect drug interactions. The common occurrence of adverse reactions and the emerging application of drug repositioning indicate the existence of off-target binding effects that are often unaccounted for in the primary analysis of the target [85]. Drug intervention based on one target may lead to broader, downstream effects than another, given the complex interconnected nature of biological interactions and pathways. This increases the relevance of identifying an overall therapeutic effect to understand mechanisms of action and presses the need for integrative approaches to elucidate systems-level drug phenotypes.

A number of network-based studies have emerged as a result of the wide availability of pharmacological data such as, for example, on drug-gene interactions and side effect similarities [86, 87, 88]. These approaches, however, are limited to information that is already anticipative of a drug's action. Alternatively, response data records profiles of molecular measurements (e.g. at gene and protein expression levels) that represent actual biological responses to drug perturbations [89, 90, 91, 92]. Network-based prediction of tissue-specific

metabolic states from transcriptomic data have recently highlighted the constraint-based approach as an useful alternative to standard methods that rely primarily on statistical inference [93, 36]. Given the interest in metabolic enzymes and regulators as therapeutic targets [84], incorporating drug response data with genome-scale metabolic networks can be a first step towards evaluating systemic drug response phenotypes.

This chapter focuses on a network-based approach to characterize systemic phenotypes of human metabolism based on pharmacogenomic response data. An integrative network approach was applied to evaluate drug-specific metabolic reaction activities using expression profiles from the Connectivity Map database [94] in the context of a global human metabolic reconstruction. The resulting reaction response signatures were further evaluated to identify profiles of metabolite markers relevant to potential mechanisms of action. The results of this study show that the metabolic network-based approach can be a valuable tool in pharmacogenomic analysis for characterizing systemic drug response phenotypes.

4.1 Method approaches to analyzing pharmacogenomic data

4.1.1 Data processing and mapping

Gene expression CEL files for the AffyMetrix U133 chips were downloaded from the Connectivity Map database and MAS5-normalized with the BioConductor package [95] to obtain probe intensities and detection p-values for each array. Human Entrez Gene identifiers associated with the probes were used to map detection p-values to their corresponding reactions based on the Boolean gene-protein-reaction (GPR) associations as was previously described [93]. Reactions which were not associated with a gene were assigned a p-value according to the 90^{th}-percentile for that expression profile. The Recon 1 metabolic network was constrained to allow uptake at an arbitrary exchange value of -1 of metabolites according to a supplemented DMEM media composition, described as the culture medium environment from the original dataset [94].

4.1.2 Gene expression analysis to determine reaction activity scores

The GIMME algorithm [93] was used to analyze quantitative gene expression profiles in the context of the genome-scale human metabolic network. Briefly, the GIMME algorithm uses quantitative gene expression data and flux balance analysis (FBA) [1, 50, 51] as a cue to activate reactions in a network model to construct condition-specific metabolic networks required to achieve pre-determined metabolic objectives. Reactions are deemed

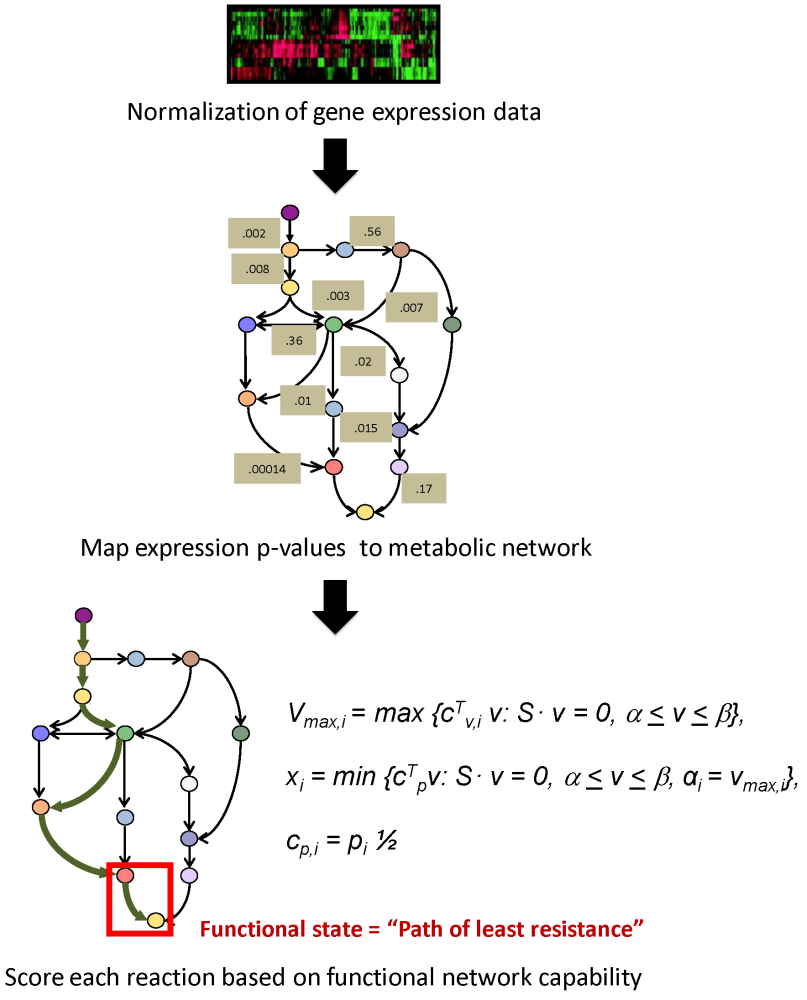


Figure 4.1: Schematic illustrating the conversion of gene expression data to reaction activity scores.

MAS5-normalization of the array data was performed to obtain Wilcoxon p-values. p-values are mapped to metabolic network reactions based on GPR associations, and each network reaction is then set as a metabolic objective in separate LP problems. The FBA-based approach calculates weighted flux distribution scores for each reaction metabolic objective.

active if above an imposed expression value cutoff, and inactive reactions are reactivated as required to fulfill the objective while minimizing the difference from the cutoff value. An "inconsistency" score is calculated to quantify the predicted networks agreement with gene expression data for a particular metabolic objective, with a lower score indicating a better network agreement with the objective. For this study, the GIMME approach was implemented as a scoring method to determine each drug-treated expression profiles agreement, or consistency, with the optimal activity of each metabolic reaction (Figure 4.1). We calculated inconsistency scores for all Recon 1 network reactions by setting each reaction as a required metabolic objective in separate GIMME problems. The method was modified to weight each reaction with a detection p-value function to determine and score the most probable flux distribution that optimizes for each network reaction (i.e. likelihood that optimal activity is achieved for each metabolic reaction). This approach allows for optimizing for the most probable functional flux distribution ("path of least resistance") when solving for a network most consistent with expression data and unbiased assessment of all reactions as objective functions. Hence, each reaction is quantified with a "functionality" score that defines its optimal activity enabled by the most probable flux distribution.

The Recon 1 network [3] was first converted into an irreversible network such that all reactions were restricted to positive flux values. The optimal flux value for each metabolic reaction objective was first calculated using FBA with the optimization problem:

(1) $\mathrm{v}_{max,i} = \max\{\mathrm{c}^{T}{}_{v,i}\mathrm{v}\text{: } S \cdot v = 0,\ \alpha < \mathrm{v} < \beta\}$,

where S is the stoichiometric matrix, v is the reaction flux vector, α and β are vectors for the lower and upper bounds of the reactions, and $\mathrm{c}_{v,i}$ corresponds to the imposed reaction objective for each i reaction. The reaction activity scores, x_i, were calculated for each i reaction by setting the lower bound to the optimal reaction activity, $\mathrm{v}_{max,i}$, and solving:

(2) $\mathrm{x}_i = \min\{\mathrm{c}^{T}{}_{p}\mathrm{v}\text{: } S \cdot v = 0,\ \alpha < \mathrm{v} < \beta,\ \alpha_i = v_{max,i}\}$,

where mapped detection p-values p were stored in a weighting vector c_p, and each vector component $\mathrm{c}_{p,i} = {p_i}^{0.5}$ mapped to each i^{th} reaction.

4.1.3 Analysis of metabolic response phenotypes (MRPs)

The drug reaction activity scores were normalized to the control activity scores to determine how each reaction activity changes relative to its respective controls:

(3) $R_i = \frac{x_{drug,i} - \mu_{control,i}}{\sigma_{control,i}/\sqrt{n_{control,i}}}$,

where R_i is the i^{th} reaction response, $\mu_{control,i}$, $\alpha_{control,i}$, and $n_{control,i}$ are the respective mean, standard deviation, and number of arrays for each i^{th} control reaction activity score. This normalization allows us to account for variability in the control activities. Since the activity scores were determined based on the weighted summation of detection probabilities, a lower drug reaction score compared to its control would indicate a higher overall probability of activity whereas a higher score would signify a lower probability. The resulting normalized scores represent a quantitative metabolic signature of relative high (activated) and low (inhibited) activity perturbations according to each array.

A Welchs t-test was used to assign statistical significance of reaction response changes in drug treatments with a minimum of 3 replicate array sets (df = n-2). Statistical significance was determined between groups of replicate drug treatment conditions and all other treatments. The Wilcoxon rank sum test was used to determine statistical significance of drug treatment pairwise correlations between similar treatments, shared gene targets, and drug metabolizing enzymes versus all possible pairwise correlations. Statistical significance of reaction frequency across all 353 drug treatments was determined with a one-variable Chi-square test (df = 1) of equivalent probability for increased and decreased activities.

4.1.4 Mapping of MCF-7 proteome data

The MCF-7 proteomic biomarker data was taken from the supplemental material of a previous published study [96]. Up-regulated and down-regulated protein biomarkers were based on the statistical significant threshold established in the original study. UniProt identifiers from the supplemental material were mapped to their respective Entrez Gene identifiers, which were then mapped to Recon 1 network reactions based on the GPR associations.

4.1.5 Analysis of metabolite markers

Perturbed metabolites were summarized based on statistical reaction significance of the MRP profiles using an approach similar to a previously published reporter metabolite analysis [97]. In brief, statistical significance of reactions are summarized around metabolites to identify significantly perturbed metabolites of the metabolic network. These reporter metabolites indicate dominant perturbation features at the metabolite level for a particular drug condition. The reporter metabolite Z-score for any metabolite j is derived from the response scores of the reactions consuming or producing j (set of reaction responses denoted as r_j) as:

(4) $\mathrm{z}_{met,j} = \frac{m_{met,j} - \mu_{met,Nj}}{\sigma_{met,Nj}}$,

where N_j is the number of reactions in r_j and $\mathrm{m}_{met,j}$ is calculated as:

(5) $\mathrm{m}_{met,j} = \frac{1}{\sqrt{Nj}} \sum_{i \epsilon r_j} \mathrm{R}_i$.

Metabolite scores were normalized by computing $\mu_{met,Nj}$ and $\sigma_{met,Nj}$ corresponding to the mean m_{met} and its standard deviation for 1,000 permutation reaction sets of size N_j. with positive scores indicating a significantly perturbed metabolic region relative to what is expected by random chance. Statistical significance of the metabolite marker Z-scores are found according to a one-tailed distribution.

4.2 Evaluating drug metabolic response phenotype (MRP) profiles

The Connectivity Map database is a collection of pharmacogenomic profiles for cancer cell lines that have been treated in culture with drug (i.e. perturbagen) molecules [94]. The likeness, or connectivity, between drugs is assessed by comparing a user-determined query set of up- or down-regulated genes to the reference drug expression profiles in the database. In this study, we evaluated the set of gene expression levels that are mapped to metabolic reactions (n = 2,114) in Recon 1 to determine the network-wide effects on optimizing each reaction activity. We used the initial published set from the database (build

01), with a focus on the most comprehensive cell line dataset available for the MCF-7 breast cancer cell line.

We use an integrative metabolic network approach to analyze gene expression profiles by asking the following question: how does the overall metabolic expression profile affect the functional capability of each biochemical reaction? Expression profiles from drug-treated and untreated cell cultures were first analyzed based on a published optimization-based (GIMME) method that computes and scores a steady-state flux distribution most consistent with an expression profile for a defined objective function [93]. Each network reaction was evaluated as an objective function with the optimization algorithm and scored to quantify the most probable flux distribution enabling its optimal activity. This generated a reaction activity vector based on each array (n_{drug}=353, $n_{control}$=59), thus defining the most likely optimal flux activities consistent with the data. Unscored reactions (no flux activity) were removed from further analysis. Pearson correlation coefficients (PCCs) between pairs of the same treatment were computed to evaluate the degree of similarity between replicate conditions. We compared the reaction score vectors and the subset of mapped normalized expression data to find that the average reaction score vector PCC = 0.955 was significantly higher than the average array data PCC = 0.776 (Wilcoxon $p < 10^{-10}$). Hence, the result indicates that the conversion to reaction activity scores is robust to data noise as it significantly improves profile similarity of the array data in which the scores were derived from.

The reaction score vectors were then evaluated to determine response signatures of metabolism, or metabolic response phenotype (MRP) profiles, related to drug action perturbations. MRPs were calculated by normalizing reaction scores of drug-treated cells to their respective control scores to identify how each reaction in the network is affected by drug intervention. The MRP profiles are represented as vectors of perturbed activity scores, in which each score component defines the relative reaction activity change between drug treatment and control. We then verified the similarity across profiles of cells treated under the same drug conditions (i.e. same drug or same drug concentration). PCCs were calculated for all possible pair-wise MRP combinations as an unbiased control (average PCC = 0.416). The MRPs between cells treated with the same drug were more significantly correlated than the control average with a mean PCC = 0.485 (Wilcoxon $p < 10^{-7}$). MRPs for cells treated with the same concentration of the same drug were only slightly higher (mean PCC = 0.501, Wilcoxon $p < 10^{-7}$). Similarity considerably improves when we only evaluate reaction activities that are significantly perturbed (mean PCC = 0.914, p <

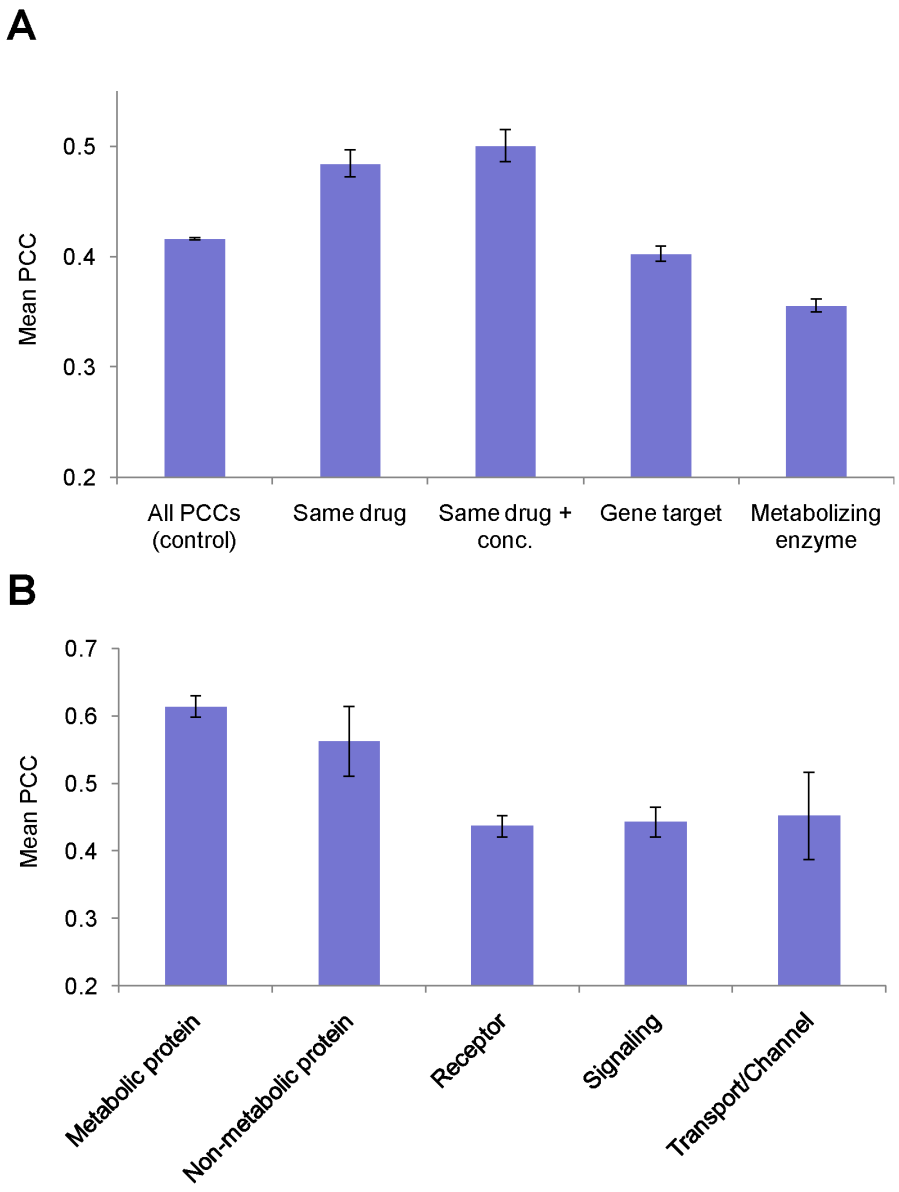


Figure 4.2: Weighted average Pearson correlation coefficients (PCCs) between MRP profiles in different categories.

(A) The similarity measure between replicate treatment conditions (i.e. same drug, same concentration of drug) were significantly higher. MRPs of the same gene target and drug metabolizing enzymes showed less similarity. (B) Drugs which targeted metabolic proteins had higher similarity between MRPs. Gene targets that were not metabolism-related were more susceptible to metabolic response variation.

0.05); overall, the results indicate MRP profiles are significantly consistent between similar treatment conditions (see Figure 4.2A).

In principle, drugs presumed to elicit a similar action, such as through its gene-targeted or drug metabolizing actions, have related response phenotypes. To assess whether drugs with the same gene targets or metabolizing enzymes indeed share similar responses, we evaluated a subset of MRPs of drugs with known shared gene targets and metabolizing enzymes based on information downloaded from the DrugBank database [98]. Surprisingly, we found that MRPs between drugs that share either the same targets or drug metabolizing enzymes have a decreased likelihood of sharing a similar metabolic response. MRPs between drug pairs with shared gene targets were slightly less correlated (mean PCC = 0.403, Wilcoxon $p < 0.02$) when compared to the average control expectation. Table 4.1 provides the subset of gene targets and respective average correlations between drugs which share that targeted action. Correlation between MRPs that shared a drug metabolizing enzyme were even lower (mean PCC = 0.356, Wilcoxon $p < 10^{-10}$), suggesting that the process in which drugs are metabolized have an insignificant impact on the systemic response phenotype. Drugs whose targets are metabolic reactions were generally more highly correlated than gene targets pertaining to non-metabolic actions (Figure 4.2B).

Among the highly correlated MRPs are those sharing targeted actions related to nucleotide metabolism-related genes such as GMPR, ADSL, AMPD1, and IMPDH1. Metabolism can be mediated by upstream controls at the gene expression and signaling levels, and a number of existing FDA-approved therapies target such controls to enhance or inhibit specific metabolic processes [84]. However, our analysis shows that non-metabolic targets generally result in a higher response variation, as indicated by a lower average PCC for non-metabolic proteins, targeted receptors, channels or transporters, and signaling proteins. The systemic metabolic responses are shown to be unconserved in drugs with similarly expected gene-targeted or metabolizing actions, particularly for those which are not directly related to a targeted metabolic activity. Furthermore, it emphasizes that the same interacting target between drugs does not guarantee a similar metabolic response as there are additional effects occurring downstream or away from the expected interaction. Thus, the development of a viable metabolic target may depend on a more direct approach to ensure a more accurate response.

Table 4.1: Subset list of gene names and mean correlation ($p < 0.02$) between drugs sharing similar gene targets.

Gene target	**Gene description**	**Mean correlation**
ABCG2	ATP-binding cassette, sub-family G, member 2	0.858
TUBB1	tubulin, beta 1	0.836
rpsD	30S ribosomal protein S4	0.819
GMPR	guanosine monophosphate reductase	0.779
ADSL	adenylosuccinate lyase	0.670
AMPD1	adenosine monophosphate deaminase 1	0.669
GMPS	guanine monophosphate synthetase	0.668
IMPDH2	IMP (inosine monophosphate) dehydrogenase 2	0.667
PPARG	peroxisome proliferator activated receptor gamma	0.665
ALOX5	arachidonate 5-lipoxygenase	0.585
HPRT1	hypoxanthine guanine phosphoribosyl transferase 1	0.580
PPAT	phosphoribosyl pyrophosphate amidotransferase	0.580
PIM1	PIM-1 kinase	0.545
CACNA2D1	calcium channel, alpha 2/delta subunit 1	0.544
HRH1	histamine receptor H1	0.538
NFKB1	nuclear factor of kappa light polypeptide gene enhancer	0.476
HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A	0.469
PIK3CG	phosphoinositide-3-kinase, gamma polypeptide	0.453
DRD1	dopamine receptor D1	0.433
ADRA2A	adrenergic receptor, alpha 2a	0.417
SCN5A	sodium channel, voltage-gated, type V, alpha subunit	0.415
ALB	albumin	0.415
PTGS1	prostaglandin-endoperoxide synthase 1	0.388
CALM1	calmodulin 1	0.385
ABCB1	ATP-binding cassette, sub-family B (MDR/TAP)	0.371
PDE4A	phosphodiesterase 4A, cAMP-specific	0.356
ESR2	estrogen receptor 2	0.355
PDPK1	3-phosphoinositide dependent protein kinase-1	0.340
CACNG1	calcium channel, voltage-dependent, gamma subunit 1	0.124

4.3 Characterizing the global drug response pattern of MCF-7 cells

A global drug response pattern was characterized based on the frequency in which metabolic reaction activities were functionally different (i.e. increased vs. decreased activities). Metabolic reactions that were significantly up- or down-regulated (Bonferonni p $< 10^{-4}$) across all 353 profiles represent commonly affected activities of the drug-treated MCF-7 cells. Among the most commonly upregulated reaction activities involve the pentose phosphate pathway, folate metabolism, and purine catabolism. An energetic dependency on mitochondrial oxidation (TCA cycle, fatty acid metabolism, oxidative phosphorylation) is emphasized as several glycolytic activities broadly decrease in response to drug-induced conditions. The oxidative activities are coupled to a redox response to mitigate reactive oxygen species (ROS) production as glutathione peroxidase and superoxide dismutase activities are among the most highly activated functions. These pathways are consistent with a cellular redox maintenance response, which has been proposed to play a role in drug toxicity and is equally critical as the target-binding event itself [99].

We compared the broader drug response to MCF-7 metabolism by assessing a subset of metabolic reactions based on a set of proteomic biomarkers previously identified in wild-type MCF-7 cell lines [96] (Figure 4.3). A published study identified a profile of protein markers that were significantly upregulated and downregulated in MCF-7 cells using an integrative proteo-transcriptomic approach, where 39 metabolic proteins mapped to 37 significantly upregulated (positive) and 12 downregulated (negative) reaction activities. Interestingly, a significant overlap (overlap = 0.83, p $< 10^{-4}$) was observed between the drug response activities and the biomarkers. Upregulated proteins involved in *de novo* nucleotide synthesis (pentose phosphate pathway, folate metabolism, and IMP biosynthesis) correspond to their respective predicted drug response activities. Nucleotide metabolism is particularly crucial for cellular proliferation, and many common chemotherapeutic agents are designed around the inhibition of purine and pyrimidine synthesis pathways [100]. Enhanced TCA cycle, ATP synthase, glutathione peroxidase, and catalase activities in conjunction with downregulated reactions of aerobic glycolysis implicate a heavier reliance on mitochondrial energy production and stress detoxification during drug intervention. While glycolysis is generally the focus in cancer metabolism, it has been shown that oxidative metabolism plays a more significant role in ATP production than glycolytic metabolism in a number of different cancer cells, including MCF-7 cells [101, 102]. These results suggest

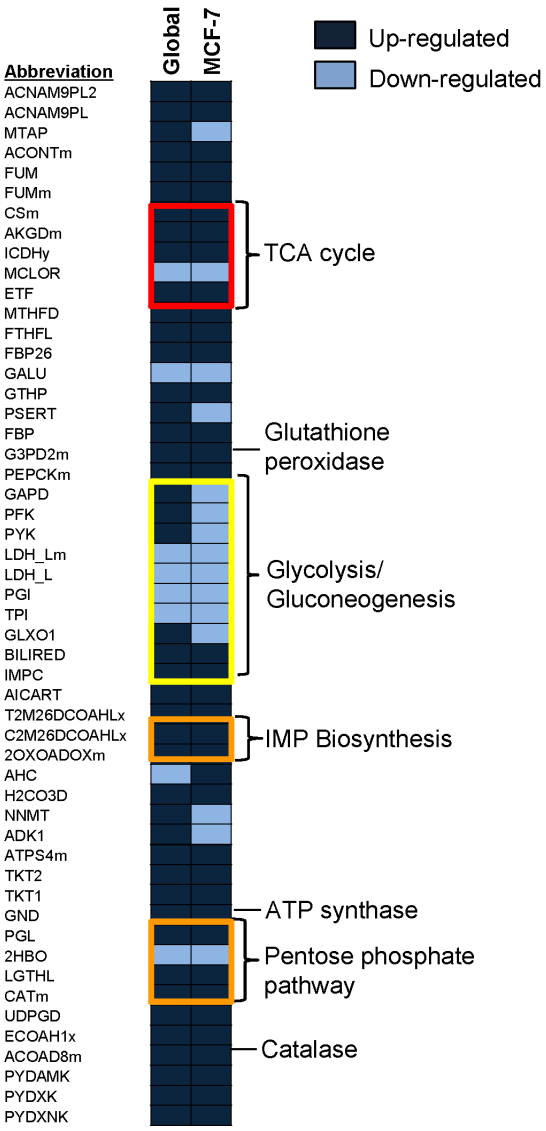


Figure 4.3: Comparison of a subset of reaction activities between the global drug response and MCF-7 proteome biomarker profile.

A significant overlap (83%) is shown for activities related to increased purine synthesis and pentose phosphate pathways. Up-regulated oxidative metabolism (TCA cycle, glutathione peroxidase, catalase, ATP synthase) and broadly down-regulated glycolytic activities is indicative of a general cellular redox maintenance response.

that redox and biosynthetic capabilities are primed in MCF-7 metabolism to moderate drug toxicity effects related to oxidative stress.

4.4 Metabolite markers are consistent with mechanisms of drug action

The effects of drug action were further summarized by reaction changes around metabolites to identify metabolic nodes where surrounding reactions were most prominently affected. These are considered central metabolite hubs where reaction perturbations converge on as a consequence of a particular drug response and can be regarded as potential markers specific to its action. Note that only a subset of 41 drugs could be analyzed as replicate drug-treated sets were required to determine the statistical significance of reaction perturbations (see Methods for further details).

A detailed analysis of the most significantly perturbed metabolites ($p < 0.05$) can be used to identify the primary response modes of drug action (Figure 4.4A). For instance, precursor metabolites involved with cortisol synthesis were highly scored for carbamazepine, a drug which primarily treats corticosteroid-induced psychotic and mood disorders [103]. Fatty acids and their acyl-CoA derivatives were the primary metabolite markers of the isoflavone genistein, indicative of its beneficial effects on treating hyperlipidemia [104]. In another example, sodium phenylbutyrate, a drug used in the treatment of hyperammonemia, has been previously reported to have a hypocholesterolemic effect *in vivo* [105, 106]. Metabolite markers related to 27-hydroxycholesterol, a biliary cholesterol derivative suggested to prevent atherosclerosis [107], proposes that the 27-hydroxylation pathway is potentially linked to this beneficial effect.

An analysis of metformin response revealed aminoimidazole carboxamide ribonucleotide (AICAR) as the key metabolite linked to its mechanism of action. Metformin, an anti-diabetic drug, acts as an energetic switch for catabolic metabolism through the activation of AMP-activated kinase (AMPK). AICAR, an intermediate of IMP biosynthesis, is commonly used as a mimetic for AMPK activation [108] and the identification of AICAR as the primary biomarker suggests regulation of folate-dependent purine synthesis plays an important role during AMPK mediation. This is further supported as metabolites involved with the reactions AICAR transformylase and IMP cyclohydrolase are also highly perturbed. (Figure 4.5).

Comparison between drugs in a similar therapeutic class can be used to iden-

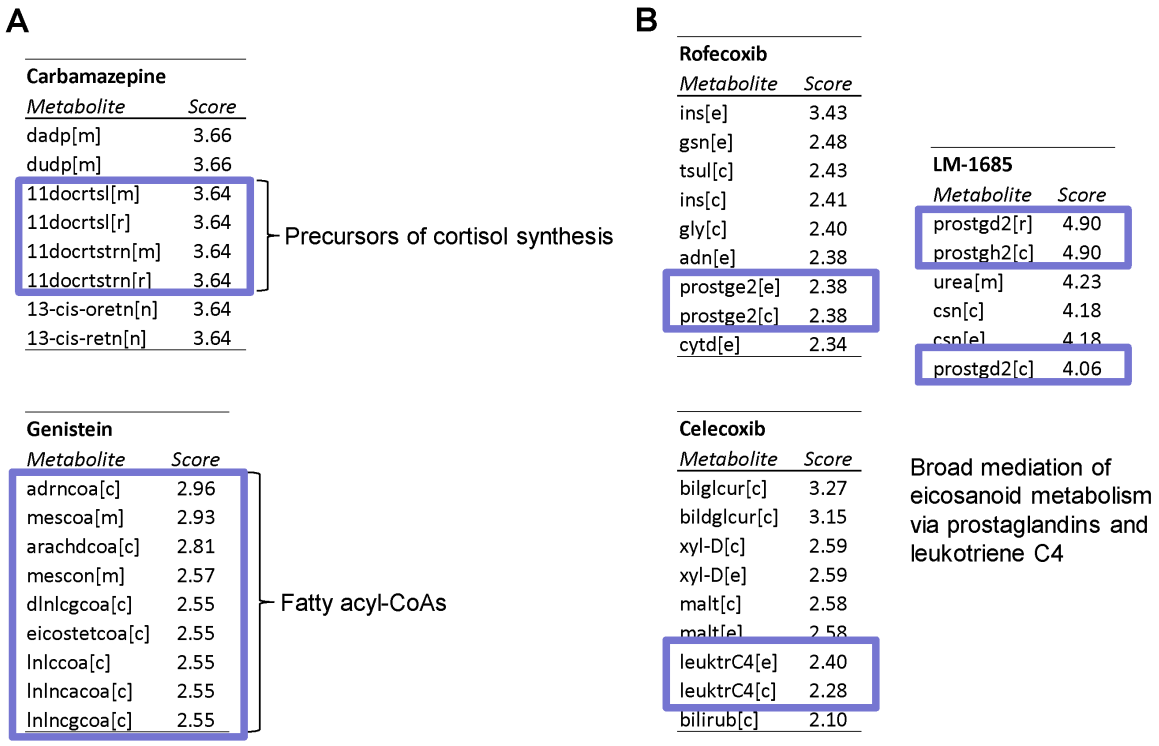


Figure 4.4: Example listings of the highest metabolite scores ($p < 0.05$) for carbamazepine, genistein, and COX-2 inhibitors.

(A) Metabolite precursors of the cortisol synthesis pathway are indicated as markers of carbamazepine, which treats corticosteroid-related psychotic disorders. The treatment of hyperlipidemia by genistein action is highly associated with the general perturbation of fatty acyl-CoAs. (B) Metabolite markers of the COX-2 inhibitors rofecoxib, celecoxib, and LM-1685 indicate a broader mediation of eicosanoid metabolic pathways related to prostaglandin and leukotriene synthesis.

tify common mechanisms of action. For instance, metabolites associated with eicosanoid metabolism (prostaglandins D2/E2/H2 and leukotriene C4) were predicted to be broadly perturbed by the COX-2 inhibitors, consistent with its targeted inhibition of eicosanoid metabolism (Figure 4.4B). Rofecoxib is also shown to be associated with metabolite markers of keratan and chondroitin sulfate degradation, although celecoxib and LM-1685, a celecoxib analogue, were more weakly associated with the signature. COX-2 inhibitors are commonly used to alleviate symptoms of osteoarthritis, and the signature is consistent with studies showing the beneficial effects of COX-2 inhibitors have on osteoarthritic cartilage through the mediation of proteoglycan catabolism [109].

Tyrosine and 3,4-dihydroxy-L-phenylalanine (L-DOPA), the respective substrate and product of tyrosine hydroxylase, are identified as common metabolite markers of the histone deacetylase (HDAC) inhibitors, valproic acid and trichostatin A. Figure 4.6 illustrates the metabolites markers that are implicated in the phenylalanine- and tyrosine-derived catecholamine and thyroid hormone synthesis pathways. HDAC inhibitors act as modulators of gene expression and have been widely used to treat psychiatric and neurological disorders. Tyrosine hydroxylase (TH) is the rate limiting enzyme in the synthesis of catecholamine neurotransmitters, and studies have shown that both drugs are involved in TH promoter regulation [110, 111]. Additionally, although reports of long-term valproate usage on endocrine functions and thyroid hormone levels have been conflicting [112], our results implicate its involvement in triiodothyronine (T3) and thyroxine (T4) synthesis.

Potential side effect mechanisms can be revealed by evaluating the metabolite markers of metabolic response. Metabolites of vitamin B6 metabolism (pyridoxal and pyridoxamine phosphate) and the B6-dependent process of oxalate synthesis (i.e. oxalate, hydroxypyruvate, glyoxylate) were strongly associated with haloperidol action. The development of tardive dyskinesia generally occurs in schizophrenic patients who are on long-term treatment of haloperidol, a dopamine antagonist. However, symptoms can be effectively reduced by treating patients with vitamin B6 [113], thus indicating deficient activation of vitamin B6 as a possible cause of the haloperidol-related side effect.

In another example, metformin response was shown to be significantly associated with metabolite markers involved in TCA cycle (acetyl-CoA, oxaloacetate, fumarate) and oxidative phosphorylation (ubiquinol-10, ubiquinone-10). Furthermore, fatty acids and their acyl-carnitine derivatives are also highly associated with metformin action. These metabolites are linked to mitochondrial oxidation, consistent with a previous study showing inhibition of mitochondrial oxidative metabolism as a common mechanism of metformin

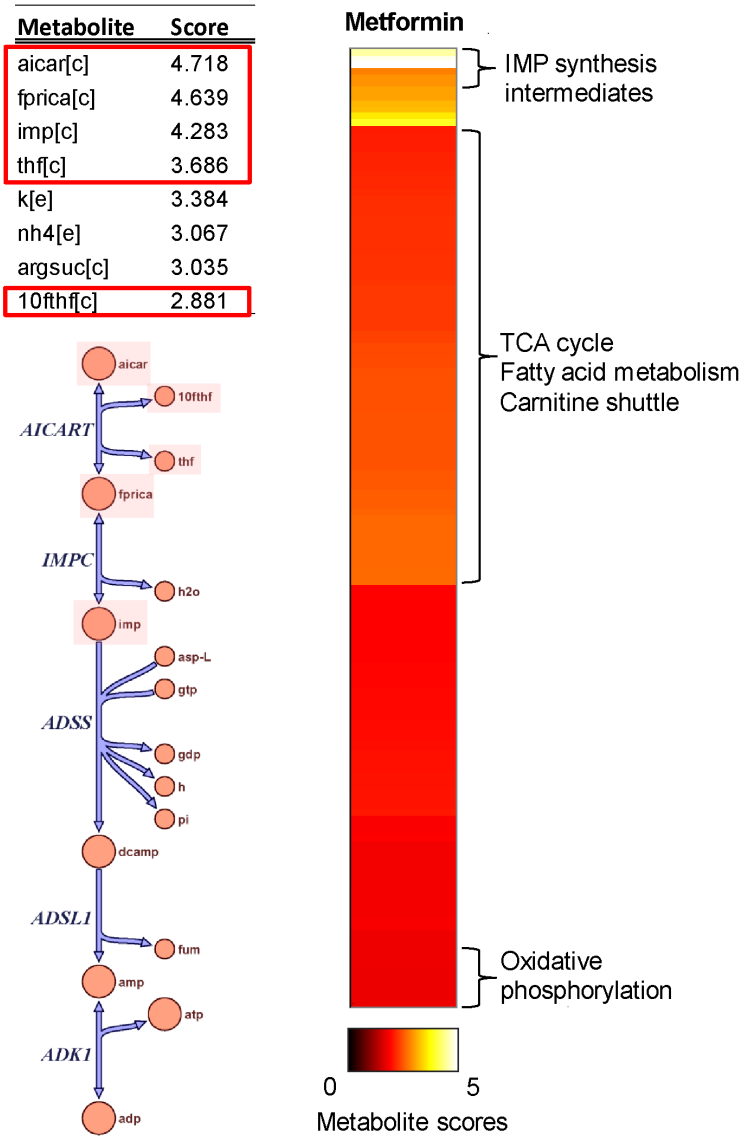
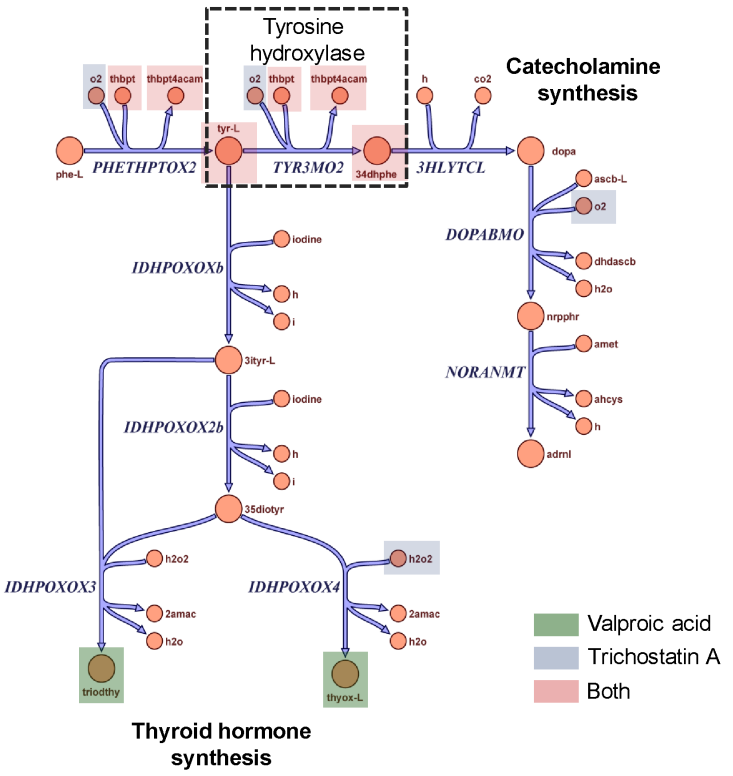


Figure 4.5: Metabolite intermediates of IMP synthesis and mitochondrial oxidation are highly associated ($p<0.05$) with metformin response.

The highest scores shown in the table implicate significantly perturbed metabolite markers involved in the reactions AICAR formyltransferase (AICART) and IMP cyclohydrolase (IMPC), as highlighted on the IMP synthesis pathway. A signature associated with mitochondrial β-oxidation, TCA cycle, and oxidative phosphorylation are also strongly mediated by metformin and acts as a potential contributor to the lactic acidosis side effect.



Metabolite	Trichostatin A	Valproic acid
tyr-L[e]	1.32	2.75
tyr-L[c]	2.74	2.69
h[g]	-0.44	2.53
trp-L[c]	1.58	2.52
k[g]	-0.43	2.45
na1[g]	-0.43	2.45
leuktrC4[r]	0.43	2.37
trp-L[e]	0.76	2.18
estradiolglc[c]	-0.26	2.16
34dhphe[e]	2.04	2.11
estradiolglc[e]	-0.12	2.09
thbpt4acam[c]	3.02	2.01
dcyt[n]	-1.42	1.95
glu-L[r]	0.11	1.92
leuktrD4[r]	0.11	1.92
thyox-L[c]	-1.20	1.91
thyox-L[e]	-1.20	1.91
triodthy[c]	-1.20	1.91
triodthy[e]	-1.20	1.91
k[c]	0.76	1.90
dmnoncoa[x]	-1.50	1.88
gthox[e]	-0.75	1.87
thbpt[c]	2.97	1.87

Figure 4.6: Highly perturbed ($p < 0.05$) metabolic intermediates of tyrosine-related pathways are associated with HDAC inhibitor response.
Catecholamine synthesis is regulated through the mediation of tyrosine hydroxylase (TH) by both valproic acid and trichostatin A responses. Thyroid hormones, T3 and T4, are also highly associated with valproic acid, indicating that endocrine function is affected by its therapeutic use.

that contributes to its well-known side effect of lactic acidosis [114] (Figure 4.5). Since diabetic patients have been shown to have dysfunctional TCA cycle activity in skeletal muscle cells [115], the amplification of the TCA cycle and oxidative phosphorylation deficiency by metformin enhances the likelihood of such an adverse effect.

4.5 Discussion

Significant effort has previously focused on identifying a drug action response that is directly associated with the anticipated target function. Genome-wide data now makes it possible to consider the systems-level effects of targeted drug intervention. Network recon-

structions provide a structured context in which gene expression data can be systematically interpreted in terms of their biochemical functions. We have demonstrated that the systematic evaluation of pharmacogenomic profiles with the human metabolic network can be used to characterize metabolic response phenotypes in drug-treated MCF-7 cells. First, we validated the MRPs by showing that the degree of similarity between reaction activities of replicate conditions is significantly correlated, whereas those sharing similarly targeted or anticipated actions are not generally conserved. Cellular redox mediation was implicated as a major downstream consequence of drug-induced actions; these pathways were also found to be highly associated with the inherent MCF-7 proteome profile, thus suggesting the cell line's endogenous metabolic signature may be less susceptible to oxidation-related drug toxicity. Furthermore, presented examples of proposed metabolite markers related to individual drug mechanisms were consistent with therapeutic and side effects that have been previously documented in literature. In the example for metformin, we were able to identify a primary metabolite marker that serves as a mimetic of the drug's principal targeted action. It is important to note that the response phenotypes were characterized with a global metabolic network rather than the cell-type in which the data was measured from. However, the global network could still be useful in inferring therapeutic effects on certain cell- or tissue-types, such as in the case of the COX-2 inhibitor rofecoxib on osteoarthritic cartilage.

Although the subset of analyzed drug targets was not entirely involved with mediating metabolism, our analysis implicates non-metabolic drug targets generally result in a higher metabolic response variation. The regulation of metabolism through various hierarchical levels (e.g. transcriptional, translational, signaling) are often complex, with many interacting factors that are still not fully elucidated. There is also the possibility of increased error when the hierarchical flow of information is coordinated among a larger cascade of components given that there are additional interactions downstream from the targeted site. Studies have already demonstrated that the influence of hierarchical regulation on metabolic fluxes can vary significantly between different enzymes [116]; thus, the optimal metabolic intervention may require targeting the biochemical enzymes themselves rather than through indirect actions at other regulatory levels. Indeed, anti-metabolites may be a more viable avenue to optimally exhibit a desired metabolic effect as they directly affect or impede the use of a metabolite.

The network-based approach presented here is currently limited to studying metabolism. Nevertheless, metabolism plays a central role in human physiology and its functional

response is essential to understanding the overall efficacy and effects to a drug. The developing discipline of pharmacometabolomics further emphasizes the emerging importance of metabolic biomarker signatures for elucidating and predicting drug response [117]. Given that genome-scale metabolic reconstructions continue to be developed for various cellular systems, a similar approach applied to pathogenic organisms could be valuable to the drug target discovery process for infectious disorders.

The text of this chapter, in full, is a reprint of the material as it will likely appear in M.L. Mo, M.J. Herrgard, and B.Ø. Palsson. 2009. Characterizing global drug response phenotypes in human metabolism (In preparation). I was the primary author of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

Chapter 5

Connecting extracellular metabolomic measurements to intracellular flux states in yeast

The measurement and identification of metabolites enables the study of unique chemical profiles which represent the downstream "end products" of gene and protein expression control. Of particular interest in metabolite biomarker discovery are the quantitative profiles of metabolites that are secreted into the extracellular environment by cells under different conditions. Recent advances in profiling the extracellular metabolome (EM) have allowed obtaining insightful biological information on cellular metabolism without disrupting the cell itself. This information can be obtained through various analytical detection, identification, and quantization techniques for a variety of systems ranging from unicellular model organisms to human biofluids [35, 118, 119, 120].

Metabolite secretion by a cell reflects its internal metabolic state, and its composition varies in response to genetic or experimental perturbations due to changes in intracellular pathway activities involved in the production and utilization of extracellular metabolites [55]. Variations in metabolic fluxes can be reflected in EM changes which can, in turn, provide insight into the intracellular pathway activities related to metabolite secretion. The extracellular metabolomic approach has already shown promise in a variety of applications, including capturing detailed metabolite biomarker variations related to disease and drug-induced states and characterizing gene functions in yeast [121, 122, 123, 124]. However, interpreting changes in the extracellular metabolome can be challenging due to

the indirect relationship between the proximal cause of the change (e.g. a mutation) and metabolite secretion.

Since metabolic networks describe mechanistic, biochemical links between metabolites, integration of such data can allow a systematic approach to identifying altered pathways linked to observed quantitative changes in secretion profiles. Measured secretion rates of major byproduct metabolites can be applied as additional exchange flux constraints that define observed metabolic behavior. For example, a recent study integrating small-scale EM data with a genome-scale yeast model correctly predicted oxygen consumption and ethanol production capacities in mutant strains with respiratory deficiencies [125]. The respiratory deficient mutant study used high accuracy measurements for a small number of major byproduct secretion rates together with an optimization-based method that are well suited for such data. Here, we expand the application range of the model-based method used in [125] to extracellular metabolome profiles, which represent a temporal snapshot of the relative abundance for a larger number of secreted metabolites. Our approach is complementary to statistical (i.e. "top-down") approaches to metabolome analysis [126] and can potentially be used in applications such as biofluid-based diagnostics or large-scale characterization of mutants strains using metabolite profiles.

This chapter will focus on the use of a constraint-based sampling approach on an updated genome-scale network of yeast metabolism to systematically determine how EM level variations are linked to global changes in intracellular metabolic flux states. By using a sampling-based network approach and statistical methods (Figure 5.1), EM changes were linked to systemic intracellular flux perturbations in an unbiased manner without relying on defining single optimal flux distributions. The inferred perturbations in intracellular reaction fluxes were further analyzed using reporter feature approaches [97, 127] in order to identify dominant metabolic features that are collectively perturbed (Figure 5.2). The sampling-based approach also has the additional benefit of being less sensitive to inaccuracies in metabolite secretion profiles than optimization-based methods and thus can more readily be used in settings such as biofluid metabolome analysis.

5.1 Methods for integration and analysis of exometabolomic data

The methods implemented in this study are shown schematically in Figures 5.1 and 5.2 and the steps are described as follows.

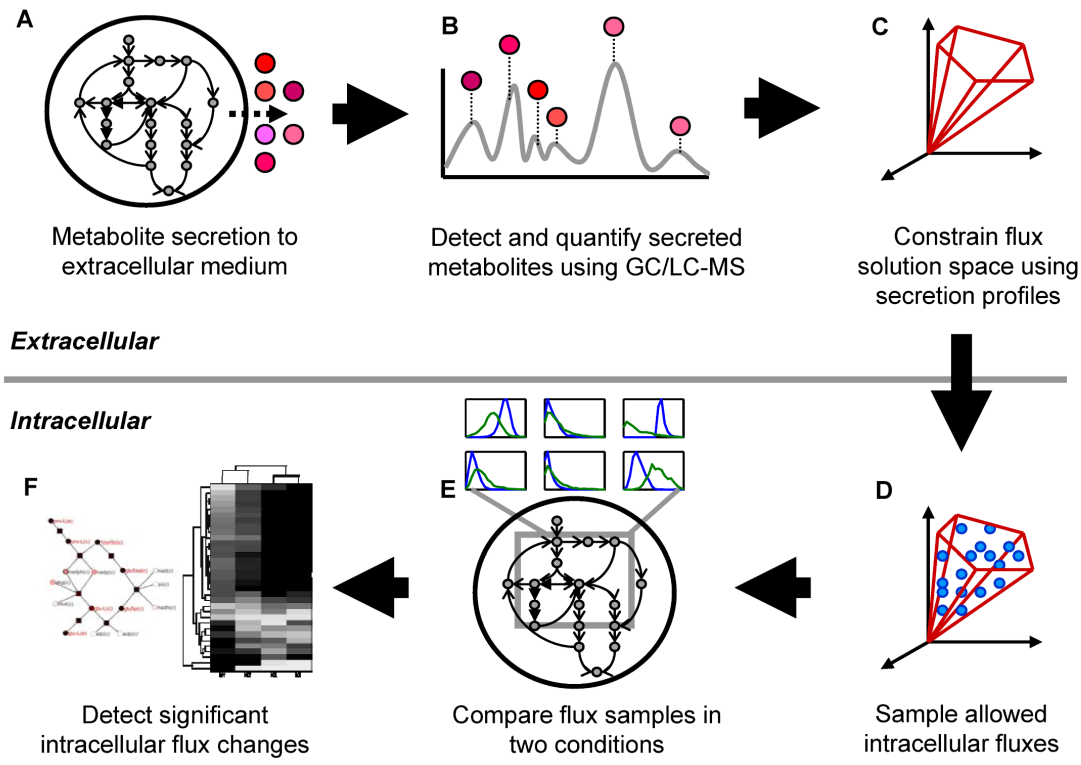


Figure 5.1: Schematic illustrating the integration of exometabolomic (EM) data with the constraint-based framework.

(A) Cells are subjected to genetic and/or environmental perturbations to secrete metabolite patterns unique to that condition. (B) EM is detected, identified, and quantified. (C) EM data is integrated as required secretion flux constraints to define allowable solution space. (D) Random sampling of solution space yields the range of feasible flux distributions for intracellular reactions. (E) Sampled fluxes were compared to sampled fluxes of another condition to determine which metabolic regions were altered between the two conditions (see Figure 5.2). (F) Significantly altered metabolic regions were identified.

5.1.1 Constraining the iMM904 network with exometabolomic data

Relative levels of quantitative EM data were incorporated into the constraint-based framework as overflow secretion exchange fluxes to simulate the required low-level production of experimentally observed excreted metabolites. The primary objective of this study is to associate relative metabolite levels that are generally measured for metabonomic or biofluid analyses to the quantitative ranges of intracellular reaction fluxes required to produce them. However, without detailed kinetic information or dynamic metabolite measurements available, we approximated EM datasets of relative quantitative metabolite levels to be proportional to the rate in which they are secreted and detected (at a steady state) into the extracellular media. This approach is analogous to approximating uptake rates based on metabolite concentrations from a previous study performing sampling analysis on a cardiomyocyte mitochondrial network to identify differential flux distribution ranges for various environmental (i.e. substrate uptake) conditions [75].

The raw data was normalized by the raw maximum value of the dataset (thus the maximum secretion flux was 1 mmol/hr/gDW) with an assumed error of 10% to set the lower and upper bounds and thus inherently accounting for sampling calculation sensitivity. The gdh1/GDH2 strains were flask cultured under minimal glucose media conditions; thus, glucose and oxygen uptake rates were set at 15 and 2 mmol/hr/gDW, respectively, for the gdh1/GDH2 strain study. In the anaerobic case the oxygen uptake rate was set to zero, and sterols and fatty acids were provided as *in silico* supplements as described in [43]. For the potassium limitation/ammonium toxicity study the growth rate was set at 0.17 1/h, and the glucose uptake rate was minimized to mimic experimental chemostat cultivation conditions. These input constraints were constant for each perturbation and comparative wild-type condition such that the calculated solution spaces between the conditions differed based only on variations in the output secretion constraints.

5.1.2 FBA optimization of EM-constrained networks

A modified FBA method with minimization of the 1-norm objective function between two optimal flux distributions was used to determine optimal intracellular fluxes based on the EM-constrained metabolic models. This method determines two optimal flux distributions simultaneously for two differently constrained models (e.g. wild type vs. mutant) - these flux distributions maximize biomass production in each case and the 1-norm distance between the distributions is as small as possible given the two sets of constraints. This

approach avoids problems with alternative optimal solutions when comparing two FBA-computed flux distributions by assuming minimal rerouting of flux distibution between a perturbed network and its reference network. Reaction flux changes from the FBA optimization results were determined by computing the relative percentage fold change for each reaction between the mutant and wild-type flux distributions.

5.1.3 Sampling of the steady-state solution space of EM-constrained network

We utilized artificial centering hit-and-run (ACHR) Monte Carlo sampling [75, 128] to uniformly sample the metabolic flux solution space defined by the constraints described above. Reactions, and their participating metabolites, found to participate in intracellular loops [129] were discarded from further analysis as these reactions can have arbitrary flux values. The following sections describe the approaches used for the analysis of the different datasets.

Sampling approach used in the gdh1/GDH2 study. Due to the overall shape of the metabolic flux solution space, most of the sampled flux distributions resided close to the minimally allowed growth rate (i.e. biomass production) and corresponded to various futile cycles that utilized substrates but did not produce significant biomass. In order to study more physiologically relevant portions of the flux space we restricted the sampling to the part of the solution space where the growth rate was at least 50% of the maximum growth rate for the condition as determined by FBA. This assumes that cellular growth remains an important overall objective by the yeast cells even in batch cultivation conditions, but that the intracellular flux distributions may not correspond to maximum biomass production [43].

To test the sensitivity of the results to the minimum growth rate threshold, separate Monte Carlo samples were created for each minimum threshold ranging from 50% to 100% at 5% increments. We also tested the sensitivity of the results to the relative magnitude of the extracellular metabolite secretion rates by performing the sampling at three different relative levels (0 corresponding to no extracellular metabolite secretion, maximum rate of 0.5 mmol/hr/gDW, and maximum rate of 1.0 mmol/hr/gDW). For each minimum growth rate threshold and extracellular metabolite secretion rate, the ACHR sampler was run for 5 million steps and a flux distribution was stored every 5000 steps. The sensitivity analysis results indicate that the reaction Z-scores are not significantly affected by either the

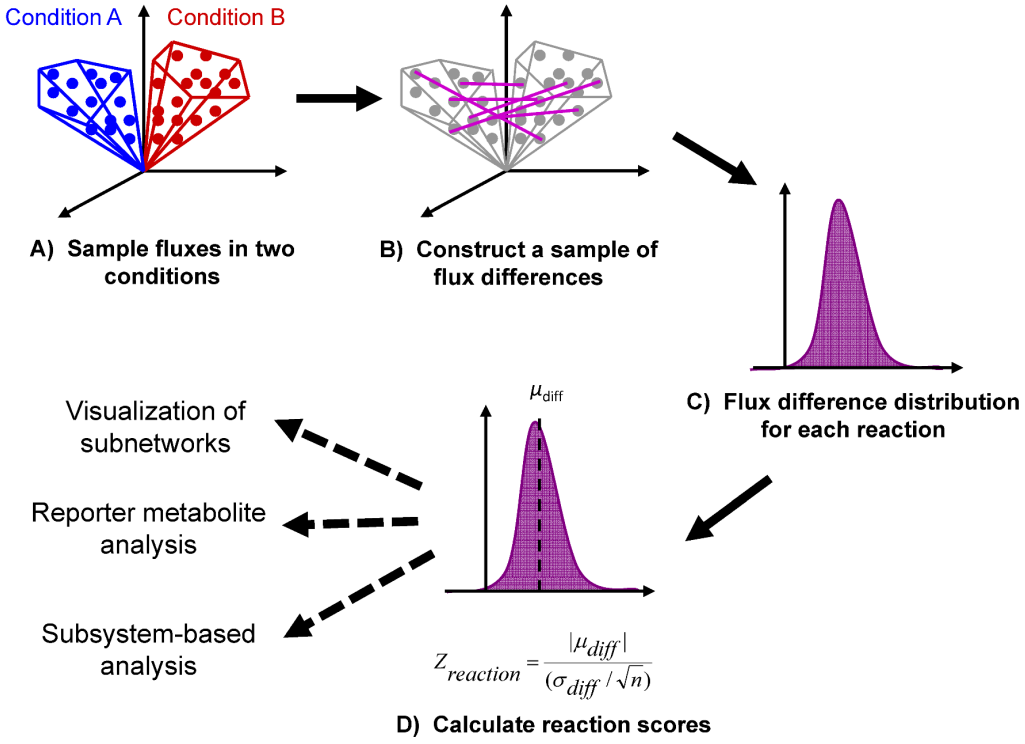


Figure 5.2: Schematic of sampling and scoring analysis to determine intracellular flux changes.

(A) Reaction fluxes are sampled for two conditions. (B & C) Sample of flux differences is calculated by selecting random flux values from each condition to obtain a distribution of flux differences for each reaction. (D) Standardized reaction Z-scores are determined, which represent how far the sampled flux differences deviates from a zero flux change. Reaction scores can be used in visualizing perturbation subnetworks and analyzing reporter metabolites and subsystems.

portion of the solution space sampled or the exact scaling of secretion rates. The final overall sample used was created by combining the samples for all minimum growth rate thresholds for the highest extracellular metabolite secretion rate (maximum 1 mmol/hr/gDW). This approach allowed biasing the sampling towards physiologically relevant parts of the solution space without imposing the requirement of strictly maximizing a predetermined objective function. The samples obtained with no EM data were used as control samples to filter reporter metabolites/subsystems for Tables 5.3 and 5.4 whose scores were significantly high due to only random differences between sampling runs.

Sampling approach used in the potassium limitation/ammonium toxicity study. Since the experimental data used in this study was generated in chemostat conditions, and previous studies have indicated that chemostat flux patterns predicted by FBA are close to the experimentally measured ones [130], we assumed that sampling of the optimal solution space was appropriate for this study. In order to sample a physiologically reasonable range of flux distributions, samples for four different oxygen uptake rates (1, 2, 3, and 4 mmol/hr/gDW with 5 million steps each) were combined in the final analysis.

5.1.4 Standardized scoring of flux differences between perturbation and control conditions

A Z-score based approach was implemented to quantify differences in flux samples between two conditions (Figure 5.2). First, two flux vectors were chosen randomly, one from each of the two samples to be compared and the difference between the flux vectors was computed. This approach was repeated to create a sample of 10,000 (n) flux difference vectors for each pair of conditions considered (e.g. mutant or perturbed environment vs. wild type). Based on this flux difference sample, the sample mean and standard deviation between the two conditions was calculated for each reaction i. The reaction Z-score was calculated as: , which describes the sampled mean difference deviation from a population mean change of zero (i.e. no flux difference between perturbation and wild type). Note that this approach allows accounting for uncertainty in the flux distributions inferred based on the extracellular metabolite secretion constraints. This is in contrast to approaches such as FBA or MoMA that would predict a single flux distribution for each condition and thus potentially overestimate differences between conditions.

The reaction Z-scores can then be further used in analysis to identify significantly perturbed regions of the metabolic network based on reporter metabolite [97] or subsys-

tem [127] Z-scores. These reporter regions indicate, or "report", dominant perturbation features at the metabolite and pathway levels for a particular condition. The reporter metabolite Z-score for any metabolite j can be derived from the reaction Z-scores of the reactions consuming or producing j (set of reactions denoted as R_j) as:

(1) $Z_{met,j} = \frac{m_{met,j} - \mu_{met,Nj}}{\sigma_{met,Nj}}$,

where N_j is the number of reactions in R_j and $m_{met,j}$ is calculated as

(2) $m_{met,j} = \frac{1}{\sqrt{Nj}} \sum_{i \epsilon R_j} Z_i$.

To account and correct for background distribution, the metabolite Z-score was normalized by computing $\mu_{met,Nj}$ and met,,N_j corresponding to the mean mmet and its standard deviation for 1,000 randomly generated reaction sets of size N_j. Z-scores for subsystems were calculated similarly by considering the set of reactions N_k that belongs to each subsystem k. Hence, positive metabolite and subsystem scores indicate a significantly perturbed metabolic region relative to other regions, whereas a negative score indicate regions that are not perturbed more significantly than what is expected by random chance. Perturbation subnetworks of reactions and connecting metabolites were visualized using Cytoscape [131].

5.2 Inferring intracellular perturbation flux states from exometabolomic profiles

5.2.1 Aerobic and anaerobic gdh1/GDH2 mutant behavior

The gdh1/GDH2 mutant strain was previously developed [132, 133] in order to lower NADPH consumption in ammonia assimilation, which would in turn favor the NADPH-dependent fermentation of xylose. In this strain, the NADPH-dependent glutamate dehydrogenase, Gdh1, was deleted and the NADH-dependent form of the enzyme, Gdh2, was overexpressed. The net effect is to allow efficient assimilation of ammonia into glutamate using NADH instead of NADPH as a cofactor. While growth characteristics remained unaffected, relative quantities of secreted metabolites differed between the wild-type and mutant strain under aerobic and anaerobic conditions.

We analyzed EM data for the gdh1/GDH2 and wild-type strains reported in [55]

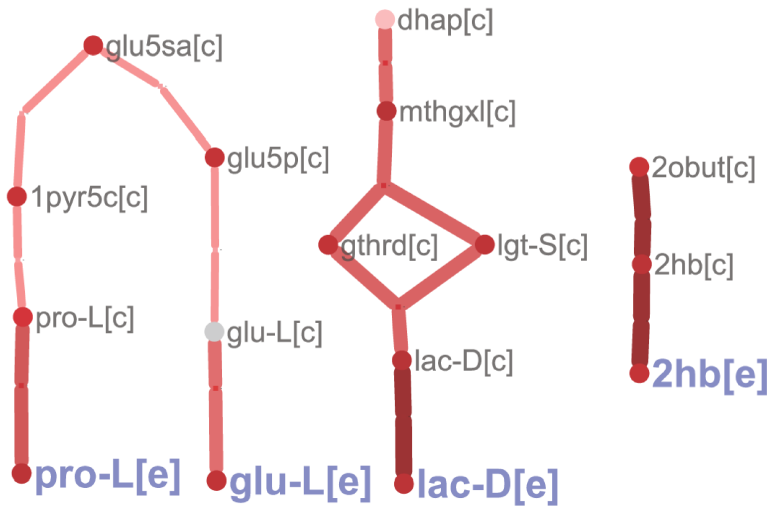


Figure 5.3: Perturbation reaction subnetwork of gdh1/GDH2 mutant under aerobic conditions.

The network illustrates a simplified subset of highly perturbed reactions connected to aerobically-secreted metabolites predicted from the sampling analysis of the gdh1/GDH2 mutant strain. The major secreted metabolites (glutamate, proline, D-lactate, and 2-hydroxybuturate) were also detected in the anaerobic condition.

under aerobic and anaerobic conditions separately using both FBA optimization and sampling-based approaches as described in Methods. 43 measured extracellular and intracellular metabolites from the original dataset [55], primarily of central carbon and amino acid metabolism, were explicitly represented in the iMM904 network. Extracellular metabolite levels were used to formulate secretion constraints and differential intracellular metabolites were used to compare and validate the intracellular flux predictions. Perturbed reactions from the FBA results were determined by calculating relative flux changes, and reaction Z-scores were calculated from the sampling analysis to quantify flux changes between the mutant and wild-type strains, with $Z_{reaction} > 1.96$ corresponding to a two-tailed p-value < 0.05 and considered to be significantly perturbed.

To validate the predicted results, reaction flux changes from both FBA and sampling methods were compared to differential intracellular metabolite level data measured from the same study. Intracellular metabolites involved in highly perturbed reactions (i.e. reactants and products) predicted from FBA and sampling analyses were identified and compared to metabolites that were experimentally identified as significantly changed ($p < 0.05$) between mutant and wild-type. Statistical measures of recall, accuracy, and precision were calculated and represent the predictive sensitivity, exactness, and reproducibility respectively. From the sampling analysis, a considerably larger number of significantly perturbed reactions are predicted in the anaerobic case (505 reactions, or 70.7% of active reactions) than in aerobic (394 reactions, or 49.8% of active reactions). The top percentile of FBA flux changes equivalent to the percentage of significantly perturbed sampling reactions were compared to the intracellular data. Results from both analyses are summarized in Table 5.1. Sampling predictions were considerably higher in recall than FBA predictions for both conditions, with respective ranges of 0.83-1 compared to 0.48-0.96. Accuracy was also higher in sampling predictions; however, precision was slightly better in the FBA predictions as expected due to the smaller number of predicted changes. Overall, the sampling predictions of perturbed intracellular metabolites are strongly consistent with the experimental data and significantly outperforms that of FBA optimization predictions in accurately predicting differential metabolites involved in perturbed intracellular fluxes.

Perturbation subnetworks can be drawn to visualize predicted significantly perturbed intracellular reactions and illustrate their connection to the observed secreted metabolites in the aerobic and anaerobic gdh1/GDH2 mutants. Figure 5.3 shows an example of a simplified aerobic perturbation subnetwork consisting primarily of proximal pathways connected directly to a subset of major secreted metabolites (glutamate, proline, D-lactate,

Table 5.1: Statistical comparison of the differential intracellular metabolite data set ($p < 0.05$) with metabolites involved in perturbed reactions predicted by FBA optimization and sampling analyses for aerobic and anaerobic gdh1/GDH2 mutant.
Overall statistics indicate combined results of both conditions (Recall = Rec; Accuracy = Acc; Precision = Pre).

	Aer FBA	**Aer Sampling**	**Ana FBA**	**Ana Sampling**	**Overall FBA**	**Overall Sampling**
Rec	0.48	0.83	0.96	1.00	0.71	0.91
Acc	0.55	0.62	0.64	0.64	0.60	0.63
Pre	0.78	0.69	0.64	0.63	0.68	0.66

and 2-hydroxybuturate). Figure 5.4 displays anaerobic reactions with Z-scores of similar magnitude to the perturbed reactions in Figure 5.3. The same subset of metabolites is also present in the larger anaerobic perturbation network and indicates that the NADPH/NADH balance perturbation induced by the gdh1/GDH2 manipulation has widespread effects beyond just altering glutamate metabolism anaerobically. Interestingly, it is clear that the majority of the secreted metabolite pathways involve connected perturbed reactions that broadly converge on glutamate. Note that Figures 5.3 and 5.4 only show the subnetworks that consisted of two or more connected reactions for a number of secreted metabolites no contiguous perturbed pathway could be identified by the sampling approach. This indicates that the secreted metabolite pattern alone is not sufficient to determine which specific production and secretion pathways are used by the cell for these metabolites.

To further highlight metabolic regions that have been systemically affected by the gdh1/GDH2 modification, reporter metabolite and subsystem methods [97, 127] were used to summarize reaction scores around specific metabolites and in specific metabolic subsystems. The top ten significant scores for metabolites/subsystems associated with more than three reactions are summarized in Tables 5.2 (aerobic) and 5.3 (anaerobic), with $Z > 1.64$ corresponding to $p < 0.05$ for a one-tailed distribution.

Perturbations under aerobic conditions largely consisted of pathways involved in mediating the NADH and NADPH balance. Among the highest scoring aerobic subsystems are TCA cycle and pentose phosphate pathway key pathways directly involved in the generation of NADH and NADPH. Reporter metabolites involved in these subsystems glyceraldehyde-3-phosphate, ribulose-5-phosphate, and alpha-ketoglutarate were also identified. These results are consistent with flux and enzyme activity measurements of the gdh1/GDH2 strain under aerobic conditions [134], which reported significant reduction in

Table 5.2: List of the top ten significant reporter metabolite and subsystem scores for the gdh1/GDH2 vs. wild type comparison in aerobic conditions.
*No. of reactions categorized in a subsystem or found to be neighboring each metabolite

Reporter metabolite	**Z-score**	**No of reactions***
L-proline [c]	2.71	4
Carbon dioxide [m]	2.51	15
Proton [m]	2.19	51
Glyceraldehyde 3-phosphate [c]	1.93	7
Ubiquinone-6 [m]	1.82	5
Ubiquinol-6 [m]	1.82	5
Ribulose-5-phosphate [c]	1.80	4
Uracil [c]	1.74	4
L-homoserine [c]	1.72	4
Alpha-ketoglutarate [m]	1.71	8
Reporter subsystem	**Z-score**	**No of reactions***
Citric Acid Cycle	4.58	7
Pentose Phosphate Pathway	3.29	12
Glycine and Serine Metabolism	2.69	17
Alanine and Aspartate Metabolism	2.65	6
Oxidative Phosphorylation	1.79	8
Thiamine Metabolism	1.54	8
Arginine and Proline Metabolism	1.44	20
Other Amino Acid Metabolism	1.28	5
Glycolysis/Gluconeogenesis	0.58	14
Anaplerotic reactions	0.19	9

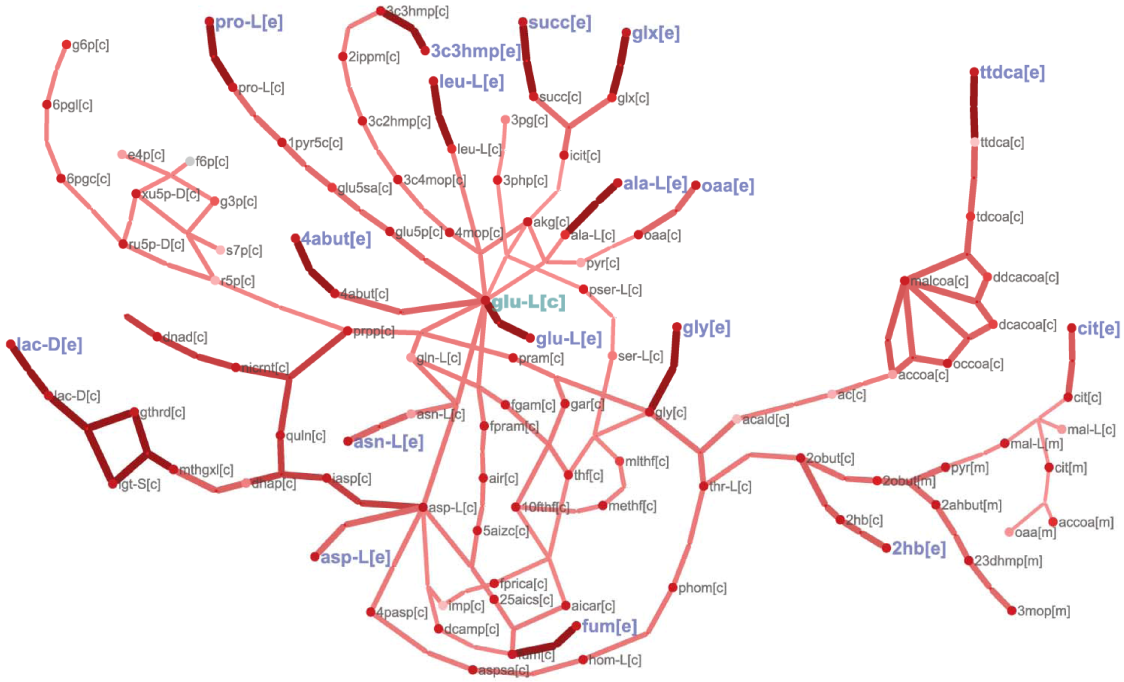


Figure 5.4: Perturbation reaction subnetwork of gdh1/GDH2 mutant under anaerobic conditions.

Subnetwork illustrates the highly perturbed anaerobic reactions of similar $Z_{reaction}$ magnitude to the reactions in Figure 5.3. A significantly larger number of reactions indicates mutant metabolic effects are more widespread in the anaerobic environment. The network shows that perturbed pathways converge on glutamate (highlighted in green), the main site in which the gdh1/GDH2 modification was introduced, which suggests that the direct genetic perturbation effects are amplified under this environment.

the pentose phosphate pathway flux with concomitant changes in other central metabolic pathways. Levels of several TCA cycle intermediates (e.g. fumarate, succinate, malate) were also elevated in the gdh1/GDH2 mutant according to the differential intracellular metabolite data. Altered energy metabolism, as indicated by reporter metabolites (i.e. ubiquinone-6, ubiquinol-6, mitochondrial proton) and subsystem (oxidative phosphorylation), is certainly feasible as NADH is a primary reducing agent for ATP production. Pentose phosphate pathway and NAD biosynthesis also appears among the most perturbed anaerobic subsystems, further suggesting perturbed cofactor balance as a common, dominant effect under both conditions.

Glutamate dehydrogenase is a critical enzyme of amino acid biosynthesis as it

acts as the entry point for ammonium assimilation via glutamate. Consequently, metabolic subsystems involved in amino acid biosynthesis were broadly perturbed as a result of the gdh1/GDH2 modification in both aerobic and anaerobic conditions. For example, the proline biosynthesis pathway that uses glutamate as a precursor was significantly perturbed in both conditions, as supported by significantly changed intracellular and extracellular levels. There were differences, however, in that more amino acid related subsystems were significantly affected in the anaerobic case (Table 5.4), further highlighting that altered ammonium assimilation in the mutant has a more widespread effect under anaerobic conditions. This effect is especially pronounced for threonine and nucleotide metabolism, which were predicted to be significantly perturbed only in anaerobic conditions. Intracellular threonine levels were amongst the most significantly reduced relative to other differential intracellular metabolites in the anaerobically grown gdh1/GDH2 strain (see [55]), and the relationship between threonine and nucleotide biosynthesis is further supported by threonine's recently discovered role as a key precursor in yeast nucleotide biosynthesis [135]. Other key anaerobic reporter metabolites are glycine and 10-formyltetrahydrofolate, both of which are involved in the cytosolic folate cycle (one-carbon metabolism). Folate is intimately linked to biosynthetic pathways of glycine (with threonine as its precursor) and purines by mediating one-carbon reaction transfers necessary in their metabolism and is a key cofactor in cellular growth [136]. Thus, the anaerobic perturbations identified in the analysis emphasize the close relationship between threonine, folate, and nucleotide metabolic pathways as well as their potential connection to perturbed ammonium assimilation processes. Interestingly, this association has been previously demonstrated at the transcriptional level as yeast ammonium assimilation (via glutamine synthesis) was found to be co-regulated with genes involved in glycine, folate, and purine synthesis [137].

In summary, the overall differences in predicted gdh1/GDH2 mutant behavior under aerobic and anaerobic conditions show that changes in flux states directly related to modified ammonium assimilation pathway are amplified anaerobically whereas the indirect effects through NADH/NADPH balance are more significant aerobically. Perturbed metabolic regions under aerobic conditions were predominantly in central metabolic pathways involved in responding to the changed NADH/NADPH demand and did not necessarily emphasize that glutamate dehydrogenase was the site of the genetic modification. The majority of affected anaerobic pathways were involved directly in modified ammonium assimilation as evidenced by 1) significantly perturbed amino acid subsystems, 2) a broad perturbation subnetwork converging on glutamate (Figure 5.4), and 3) glutamate as the

Table 5.3: List of top ten significant reporter metabolite and subsystem scores for the gdh1/GDH2 vs. wild type comparison in anaerobic conditions.
*No. of reactions categorized in a subsystem or found to be neighboring each metabolite

Reporter metabolite	**Z-score**	**No of reactions***
Glutamate [c]	4.52	35
Aspartate [c]	3.21	11
Alpha-ketoglutarate [c]	2.66	17
Glycine [c]	2.65	7
Pyruvate [m]	2.56	7
Ribulose-5-phosphate [c]	2.43	4
Threonine [c]	2.28	6
10-formyltetrahydrofolate [c]	2.27	5
Fumarate [c]	2.27	5
L-proline [c]	2.04	4
Reporter subsystem	**Z-score**	**No of reactions***
Valine, Leucine, and Isoleucine Metabolism	3.97	15
Tyrosine, Tryptophan, and Phenylalanine Metabolism	3.39	23
Pentose Phosphate Pathway	3.29	11
Purine and Pyrimidine Biosynthesis	3.08	40
Arginine and Proline Metabolism	2.96	19
Threonine and Lysine Metabolism	2.74	14
NAD Biosynthesis	2.66	7
Alanine and Aspartate Metabolism	2.65	6
Histidine Metabolism	2.24	10
Cysteine Metabolism	1.85	10

most significant reporter metabolite (Table 5.3).

5.2.2 Potassium-limited and excess ammonium environments

A recent study reported that potassium limitation resulted in significant growth retardation effect in yeast due to excess ammonium uptake when ammonium was provided as the sole nitrogen source [138]. The proposed mechanism for this effect was that ammonium could to be freely transported through potassium channels when potassium concentrations were low in the media environment, thereby resulting in excess ammonium uptake [138]. As a result, yeast incurred a significant metabolic cost in assimilating ammonia to glutamate and secreting significant amounts of glutamate and other amino acids in potassium-limited

conditions as a means to detoxify the excess ammonium. A similar effect was observed when yeast was grown with no potassium limitation, but with excess ammonia in the environment. While the observed effect of both environments (low potassium or excess ammonia) was similar, quantitatively unique amino acid secretion profiles suggested that internal metabolic states in these conditions are potentially different.

In order to elucidate the differences in internal metabolic states, we utilized the iMM904 model and the EM profile analysis method to analyze amino acid secretion profiles for a range of low potassium and high ammonia conditions reported in [138]. As before, we utilized amino acid secretion patterns as constraints to the iMM904 model, sampled the allowable solution space, computed reaction Z-scores for changes from a reference condition (normal potassium and ammonia), and finally summarized the resulting changes using reporter metabolites. Figure 5 shows a clustering of the most significant reporter metabolites ($Z > 1.96$ in any of the four conditions studied) obtained from this analysis across the four conditions studied. Interestingly, the potassium-limited environment perturbed only a subset of the significant reporter metabolites identified in the high ammonia environments. Both low potassium environments shared a consistent pattern of highly perturbed amino acids and related precursor biosynthesis metabolites (e.g. pyruvate, PRPP, alpha-ketoglutarate) with high ammonium environments. The amino acid perturbation pattern (indicated by red labels in Figure 5.5) was present in the ammonium-toxic environments, although the pattern was slightly weaker for the lower ammonium concentration. Nevertheless, the results clearly indicate that a similar ammonium detoxifying mechanism that primarily perturbs pathways directly related to amino acid metabolism exists under both types of media conditions.

In addition to perturbed amino acids, a secondary effect notably appears at high ammonia levels in which metabolic regions related to folate metabolism are significantly affected. As highlighted in green in Figure 5.5, we predicted significantly perturbed key metabolites involved in the cytosolic folate cycle. These include tetrahydrofolate derivatives and other metabolites connected to the folate pathway, namely glycine and the methionine-derived methylation cofactors S-adenosylmethionine and S-adenosylhomocysteine. Additionally, threonine was identified to be a key perturbed metabolite in excess ammonium conditions. These results further illustrate the close connection between threonine biosynthesis, folate metabolism involving glycine derived from its threonine precursor, and nucleotide biosynthesis [135] that was discussed in conjunction with the gdh1/GDH2 strain data. Taken together with the anaerobic gdh1/GDH2 data, the results consistently suggest

highly perturbed threonine and folate metabolism when amino acid-related pathways are broadly affected.

In both ammonium-toxic and potassium-limited environments, impaired cellular growth was observed, which can be attributed to high energetic costs of increased ammonium assimilation to synthesize and excrete amino acids. However, under high ammonium environments, reporter metabolites related to threonine and folate metabolism indicated that their perturbation, and thus purine supply, may be an additional factor in decreasing cellular viability as there is a direct relationship between intracellular folate levels and growth rate [139]. Based on these results, we concluded that while potassium-limited growth in yeast indeed shares physiological features with growth in ammonium excess, its effects are not as detrimental as actual ammonium excess. The effects on proximal amino acid metabolic pathways are similar in both environments as indicated by the secretion of the majority of amino acids. However, when our method was applied to analyze the physiological basis behind differences in secretion profiles between low potassium and high ammonium conditions, ammonium excess was predicted to likely disrupt physiological ammonium assimilation processes, which in turn potentially impacts folate metabolism and associated cellular growth.

5.3 Discussion

The method presented in this study presents an approach to connecting intracellular flux states to metabolites that are excreted under various physiological conditions. We showed that well-curated genome-scale metabolic networks can be used to integrate and analyze quantitative EM data by systematically identifying altered intracellular pathways related to measured changes in the extracellular metabolome. We were able to identify statistically significant metabolic regions that were altered as a result of genetic (gdh1/GD2 mutant) and environmental (excess ammonium and limited potassium) perturbations, and the predicted intracellular metabolic changes were consistent with previously published experimental data including measurements of intracellular metabolite levels and metabolic fluxes. Our reanalysis of previously published EM data on ammonium assimilation-related genetic and environmental perturbations also resulted in testable hypotheses about the role of threonine and folate pathways in mediating broad responses to changes in ammonium utilization. These studies also demonstrated that the sampling-based method can be readily applied when only partial secreted metabolite profiles (e.g. only amino acids) are available.

With the emergence of metabolite biofluid biomarkers as a diagnostic tool in human disease [140, 141] and the availability of the genome-scale human metabolic network [3], extensions of the present method would allow identifying potential pathway changes linked to these biomarkers. Employing such a method for studying yeast metabolism was possible as the metabolomic data was measured under controllable environmental conditions where the inputs and outputs of the system were defined. Measured metabolite biomarkers in a clinical setting, however, is far from a controlled environment with significant variations in genetic, nutritional, and environmental factors between different patients. While there are certainly limitations for clinical applications, the method introduced here is a progressive step towards applying genome-scale metabolic networks towards analyzing biofluid metabolome data as it 1) avoids the need to only study optimal metabolic states based on a predetermined objective function, 2) allows dealing with noisy experimental data through the sampling approach, and 3) enables analysis even with limited identification of metabolites in the data. The ability to establish potential connections between extracellular markers and intracellular pathways would be valuable in delineating the genetic and environmental factors associated with a particular disease.

The text of this chapter, in full, is a reprint of the material as it appears in M.L. Mo, M.J. Herrgard, and B.Ø. Palsson. 2009. Connecting extracellular metabolomic profiles to intracellular metabolic states in yeast. BMC Systems Biology. 3:37. I was the primary author of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

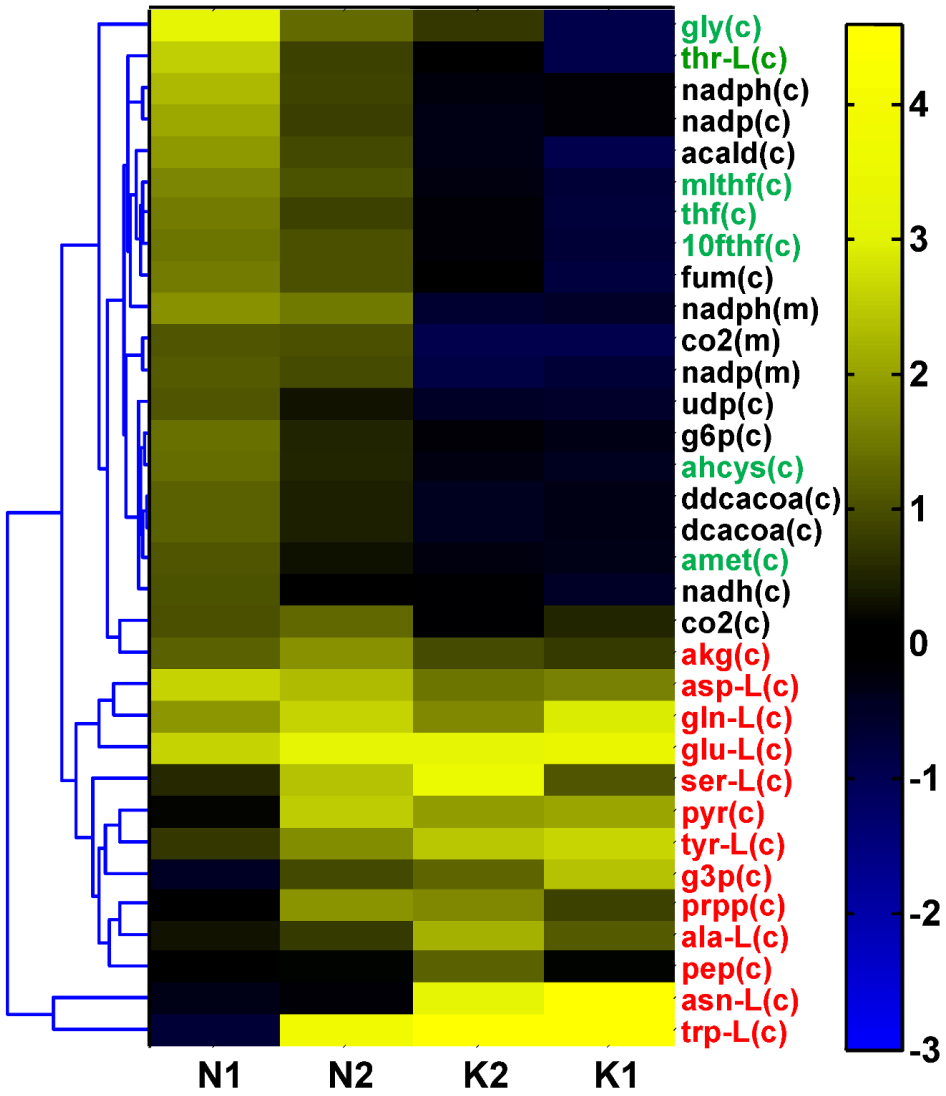


Figure 5.5: Clustergram of top reporter metabolites (i.e. in yellow) in ammonium-toxic and potassium-limited conditions.

Amino acid perturbation patterns (shown in red) were shown to be consistently scored across conditions, indicating that potassium-limited environments K1 (lowest concentration) and K2 (low concentration) elicited a similar ammonium detoxification response as ammonium-toxic environments N1 (high concentration) and N2 (highest concentration). Metabolites associated with folate metabolism (highlighted in green) are also highly perturbed in ammonium-toxic conditions.

Chapter 6

Integrative metabolomic-based analysis of embryonic stem cells

Many studies on the molecular mechanisms mediating embryonic stem cells (ESCs) have been previously focused on global gene expression levels that has led to the discovery of key genes that are involved in stem cell biology [142]. More recently, proteomic-based studies have identified the complement of functional gene products that operate in ESCs [143, 144]. Despite the advances of these studies, the molecular framework which controls the balance of ESC pluripotent and differentiation states have not yet been fully elucidated. Therefore, the study of endogenous metabolites in ESCs is an important relatively unexplored complement to understanding the stemness phenotype (i.e. pluripotency and differentiation) as it studies the resulting biochemical signatures that arise from various upstream regulatory controls.

Network reconstructions provide a structured framework to systematically integrate and analyze disparate datasets including transcriptomic, proteomic, metabolomic, and fluxomic data. Metabolomic data is one of the more relevant data types for this type of analysis as network reconstructions define the biochemical links between metabolites, and recent advancements in analytical technologies have allowed increasingly comprehensive intracellular and extracellular metabolite level measurements [35, 118]. The metabolome represents the temporal snapshot of the set of metabolites present under a given physiological condition and is the culminating phenotype resulting from various "upstream" control mechanisms. In the previous chapter, the integrative analysis of yeast exometabolomic data demonstrated that relative metabolic levels can be incorporated as required secretion constraints in a genome-scale network to infer reaction changes related to changes in

metabolomic changes. Analogously, intracellular metabolomic data can be implemented in a similar manner as an approximate measure of metabolite level that is enforced by the metabolic network, and reactions linked to these metabolite changes can be systematically inferred based on these enforced constraints.

A non-targeted, mass spectrometry-based metabolomic approach was recently used to study ESC metabolism by characterizing the metabolite properties of ESCs compared to those of differentiated neurons and cardiomyocytes. This chapter presents the application of an integrative study to analyze broader metabolic pathway usage mediated by ESCs using a constraint-based approach on the Recon 1 network based on the measured metabolite data. Metabolomic data was integrated with the metabolic network as imposed metabolite interval constraints to identify reaction activities that are linked to measured changes in metabolites characteristic of ESC metabolism. The study described here provides another example application of using metabolic networks to assess global reaction effects linked to differential metabolites that have been identified. Findings from the analysis are further supported by experimental *in vitro* studies that are briefly described.

6.1 Method approaches to analyzing ESC-mediated metabolic pathways

6.1.1 Network analysis of metabolomic data

A constraint-based modeling approach on the genome-scale human metabolic network, Recon 1 [3], was used to analyze metabolomic data measured from ESC and mature cell populations. Recon 1 was used as a general pathway framework to analyze the reaction connectivity of the metabolomic data using a constraint-based network approach [25, 47]. Metabolomic data was used to constrain the metabolic network as metabolite interval ranges, which has recently been described in published studies as an approach to defining metabolite levels in constraint-based analysis [145, 146]. Flux Variability Analysis (FVA) was used to analyze metabolic reaction flexibility as a result of increased production of metabolites observed as upregulated in ESC and mature cell populations. This method was used to analyze the network consequence on all reaction activities when there is a general increase in the production of metabolites specified in the measured data (i.e. upregulated metabolites in ESC and mature cell populations). Briefly, FVA is a variation on flux balance analysis (FBA) [1, 50, 51] which calculates both minimum and maximum reaction

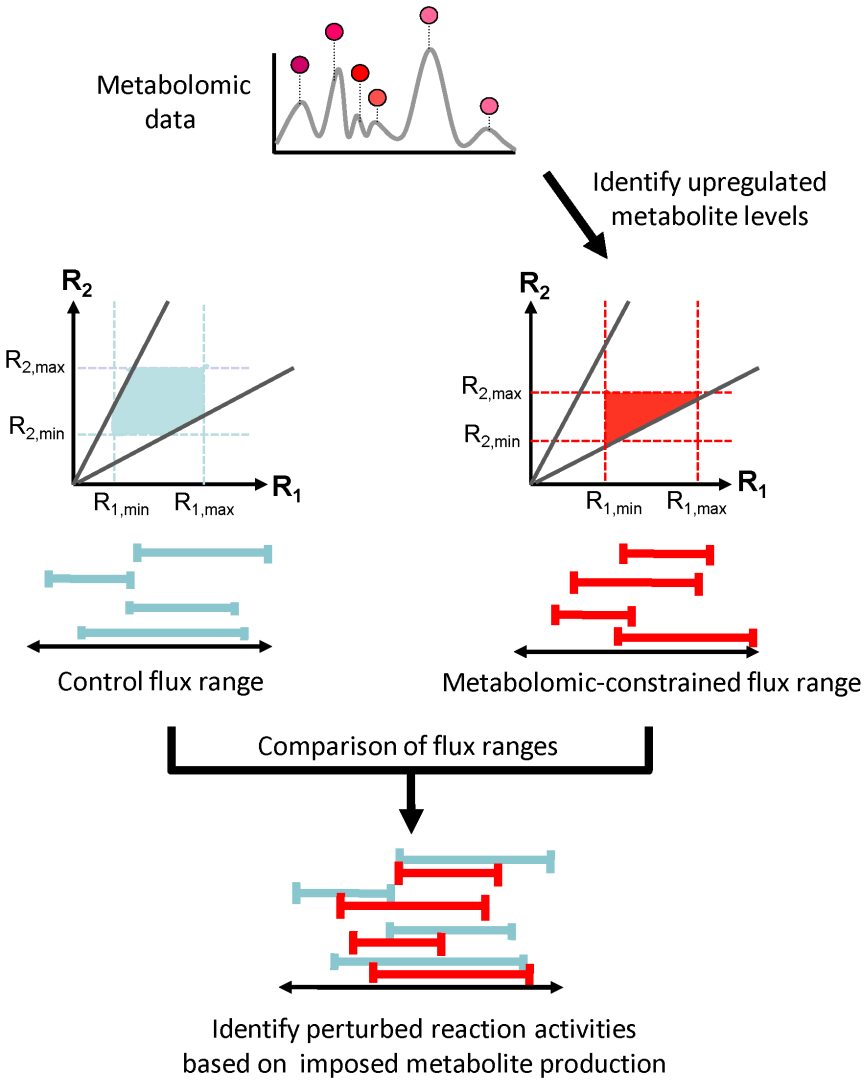


Figure 6.1: Schematic illustrating the metabolomic-based analysis using the FVA approach to identify broader reaction activites linked to upregulated metabolites.
Metabolites levels that are identified to be significantly upregulated in the metabolomic profile were applied as metabolite demand constraints. Lower and upper bounds of the interval ranges pertaining to each metabolite demand in the profile were set to its maximum value. FVA was used to determine activity ranges for all reactions as a result of maximizing production of the identified metabolites. Reactions that were affected by the metabolite demand constraints were identified by comparing the flux ranges to a control range (i.e. unconstrained network).

flux capacities through linear optimization for each network reaction under the applied constraint conditions. FVA determines the achievable range of reaction flux flexibility, or variability [147] when particular reaction constraints are imposed on the network.

For the purpose of this study, FVA was used to identify metabolic reactions which showed significant activity variation (i.e. were directly affected) when increased production of metabolites in ESC and mature cell were enforced. Any variation in reaction capacity indicates a general perturbation connected to the metabolite increased production and hence activities related to ESCs and mature cells. A comparison of reaction flexibilities between the ESC- and mature cell-constrained networks were an approximation of reaction activity changes, where decreased flexibility indicates a more restricted (or inhibited) activity, and increased flexibility indicates the reaction to be more responsive, or active. The optimal values corresponding to the maximal production for each increased metabolite were first calculated using FBA with the following linear optimization problem:

(1) $\mathrm{m}_{max,i} = \max\{\mathrm{c}^T{}_{m,i}\mathrm{v}\colon\ S \cdot v = 0,\ \alpha < \mathrm{v} < \beta\}$,

where S is the stoichiometric matrix, v is the reaction flux vector, α and β are lower and upper reaction bound vectors, with $\mathrm{c}_{m,i}$ corresponds to the imposed production objective reaction added for each ith upregulated metabolite to determine the maximum production $\mathrm{m}_{max,i}$. FVA was performed on the metabolomic-constrained networks to calculate reaction flux ranges (R_{min} and R_{max} as the upper and lower range limits) that were effectively changed relative to its control (i.e. no metabolite constraints) flux interval range. The lower and upper bounds on upregulated metabolite exchange reactions, α_m and β_m, were fixed to their respective maximum values and the flux ranges were calculated with FVA. The FVA equations are as follows:

(2) $\mathrm{R}_{min} = \min\{\mathrm{c}^T v\colon\ S \cdot v = 0,\ \alpha < v < \beta\ \ \alpha_m = \beta_m = \mathrm{m}_{max}\}$
(3) $\mathrm{R}_{max} = \max\{\mathrm{c}^T v\colon\ S \cdot v = 0,\ \alpha < v < \beta\ \ \alpha_m = \beta_m = \mathrm{m}_{max}\}$

The network was constrained to allow uptake at an arbitrary exchange value of -1 and secretion of all metabolites with an extracellular transporter reaction defined for its transport. Fixed uptake and secretion exchange constraints were used across all conditions such that calculated differences were only a function of changes in imposed metabolite demands. Reaction activity ranges were compared to their reference activity range (i.e. without

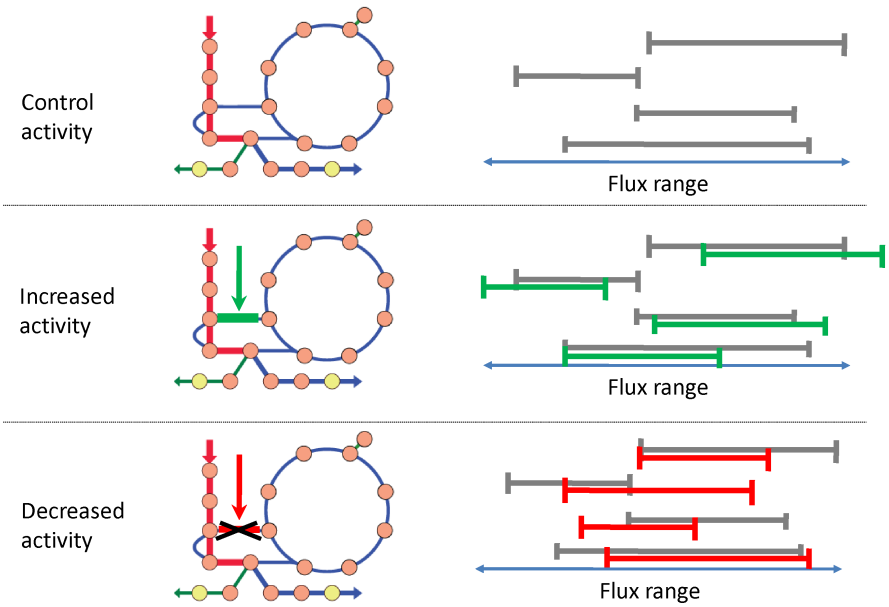


Figure 6.2: Schematic illustrating the network analysis using the FVA approach to evaluate the systemic effects of activated and inhibited reactions mediated by COX, LOX, PLA_2, and fatty acid desaturases.
The flux ranges of all reactions are compared between a control "unaltered" network and the flux ranges computed following inhibition (decreased activity) or activation (increased activity) of a reaction. This allowed for identifying broader systemic changes across all network reactions as a result of inhibiting and activating reactions associated with the synthesis and oxidative pathways of polyunsaturated fatty acids.

metabolomic constraints) to normalize reaction flexibility changes. Relative activities (i.e. increased and decreased) were determined by comparing the flux range changes between ESC and mature cell reactions. The top 80^{th}-percentile reaction changes were considered to be the highest differences between ESCs and mature cells. Non-parametric Wilcoxon rank-sum test was used to assess the statistical significance of the highly active reactions. Statistical analysis and generated plots were performed using Excel and Matlab.

6.1.2 Analyzing network effects of activated and inhibited reactions mediated by oxidative enzymes

We evaluated the network effects of activated and inhibited reactions mediated by COX, LOX, PLA_2, and fatty acid desaturase (5Δ and 6Δ desaturase) on the human

metabolic network using the previously described FVA approach (Figure 6.2). This follows a similar approach recently used to analyze systems-level effects of epigenetic perturbations [148]. FBA was first used to calculate the maximum capacity $e_{max,j}$ for each j^{th} reaction corresponding to the PLA_2, COX, LOX, and 5Δ and 6Δ desaturase enzymes, with $c_{enz,j}$ corresponding to the objective reactions set for each enzyme. FVA was used to calculate the upper and lower range limits for inhibition ($R_{min,inh}$, $R_{max,inh}$) and activation ($R_{min,act}$, $R_{max,act}$). The lower and upper bounds of the reaction were set to zero and FVA was performed to simulate reaction inhibition. FVA was then used to analyze the consequence of reaction activation by setting the lower and upper bounds to its maximum capacity, e_{max}. The equations are as follows:

(4) $e_{max,j} = \max\{c^T{}_{enz,j}\ v\text{: } S \cdot v = 0,\ \alpha < v < \beta\}$
(5) $R_{min,inh} = \min\{c^T v\text{: } S \cdot v = 0,\ \alpha < v < \beta,\ \alpha_{enz} = \beta_{enz} = 0\}$
(6) $R_{max,inh} = \max\{c^T v\text{: } S \cdot v = 0,\ \alpha < v < \beta,\ \alpha_{enz} = \beta_{enz} = 0\}$
(7) $R_{min,act} = \min\{c^T v\text{: } S \cdot v = 0,\ \alpha < v < \beta,\ \alpha_{enz} = \beta_{enz} = e_{max}\}$
(8) $R_{max,act} = \max\{c^T v\text{: } S \cdot v = 0,\ \alpha < v < \beta,\ \alpha_{enz} = \beta_{enz} = e_{max}\}$

The highest reaction activities pertaining to inhibition (ESC self-renewal) were chosen with an 80^{th} percentile cut-off. A two-tailed Students t-test was used to determine metabolic reactions that were significantly increased in the activation condition (ESC differentiation) relative to its inhibition activity. Model-based calculations were done using the Matlab COBRA Toolbox [47] utilizing the Tomlab/CPLEX (Tomopt, Inc.) optimization solver. Statistical analysis and generated tables were performed using Excel and Matlab.

6.2 Network analysis reveals altered redox status between ESC and mature populations

Endogenous metabolites using a liquid chromatography (LC)-electrospray (ESI) mass spectrometry (MS) approach were measured to find relative abundance metabolite levels in ESCs and mature (i.e. differentiated neurons and cardiomyocytes) cell populations. The identified metabolic signatures implicated significant differences in their metabolite composition 6.3, suggesting divergent pathways mediating their metabolism during differentiation. Among the most highly upregulated ESC metabolites relative to mature cell populations (fold > 2, $p < 0.01$) include lipid oxidative stress mediators and secondary lipid

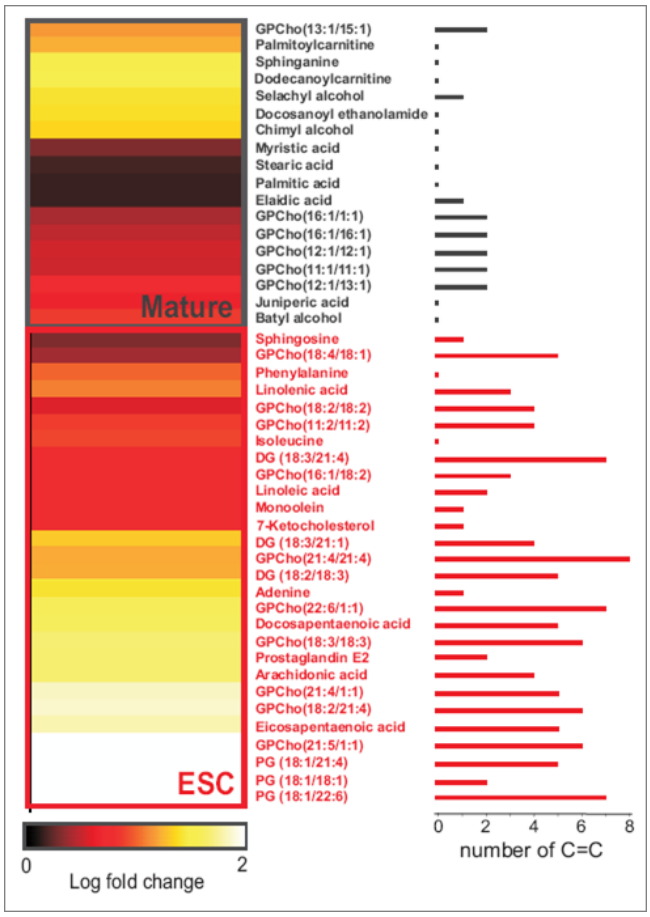


Figure 6.3: Heatmap showing 46 metabolites whose structures were identified by tandem MS.

Lighter colors (yellow and white) correspond to the largest fold changes, where fold is defined as the relative difference between the integrated peak area of each feature in ESCs relative to mature populations and vice versa. Metabolite names shown in black are up-regulated in mature populations relative to ESCs ($p<0.01$). Metabolite names shown in red are up-regulated in ESCs relative to mature populations ($p<0.01$). PCs with even-numbered carbon chains are indicated as they are the most abundant in nature. The number of C=C double bonds for each metabolite is shown by the bar graph on the right.

messengers such as arachidonic acid, prostaglandin E2, eicosapentaenoic acid, linolenic acid, diacylglycerols (DG), glycerophosphoglycerols (PG), and glycerophosphocholines (GPCho). Fatty acid constituents associated with DGs, PGs, and GPChos conferred a higher degree of unsaturation, while mature cell populations showed an abundance of saturated fatty acids and their acyl-carnitine derivatives primarily in the cardiomyocytes. Overall, elucidated structures of differentially regulated metabolites revealed there was an increased degree of unsaturation, or a larger number of carbon-carbon double bonds (C=C), in ESCs while there is a substantial decrease in unsaturated metabolites in mature cell populations.

We evaluated the differences between reactions mediating the metabolism of up-regulated ESC and mature cell metabolites in the context of a global human metabolic network1 [3] using computational methods that were previously described. The differences in reaction activities represent metabolic activities that are mediated as ESCs differentiate into mature cells. The clustergram in Figure 6.4 summarizes the metabolic reactions activities indicated to have the highest activity differences (>80th-percentile, Wilcoxon p $< 10^{-10}$) between ESCs and mature cells. Purine metabolism was highly associated with ESCs as indicated by folate metabolism, which drives de novo purine synthesis, and purine salvage pathways. Interestingly, a recent study in mouse ESCs showed that its growth is largely sustained through pathways directly associated with purine synthesis [149]. The analysis also revealed significantly increased activities in ROS detoxification pathways, glutathione metabolism, and eicosanoid metabolism in ESCs relative to mature cells, indicating a strong mediation of oxidative stress in ESCs. Responsive reaction activities involving fatty acid chain elongation/desaturation signify an important role for unsaturated fatty acids in ESC metabolism. Additionally, mitochondrial metabolism (i.e. β-oxidation) is a dominant feature in mature cells as indicated by more highly active oxidative phosphorylation and carnitine shuttle reactions relative to ESCs. These results, in conjunction with the high degree of unsaturated, or reduced, metabolite levels measured in ESCs, suggest that the regulation of redox metabolism in ESCs maintains a reduced intracellular state. Consequently, since chemical unsaturations such as C=C are highly reactive under oxidative conditions, the results suggest there is broad mediation of reduction-oxidation (redox) metabolic reaction pathways during ESC differentiation.

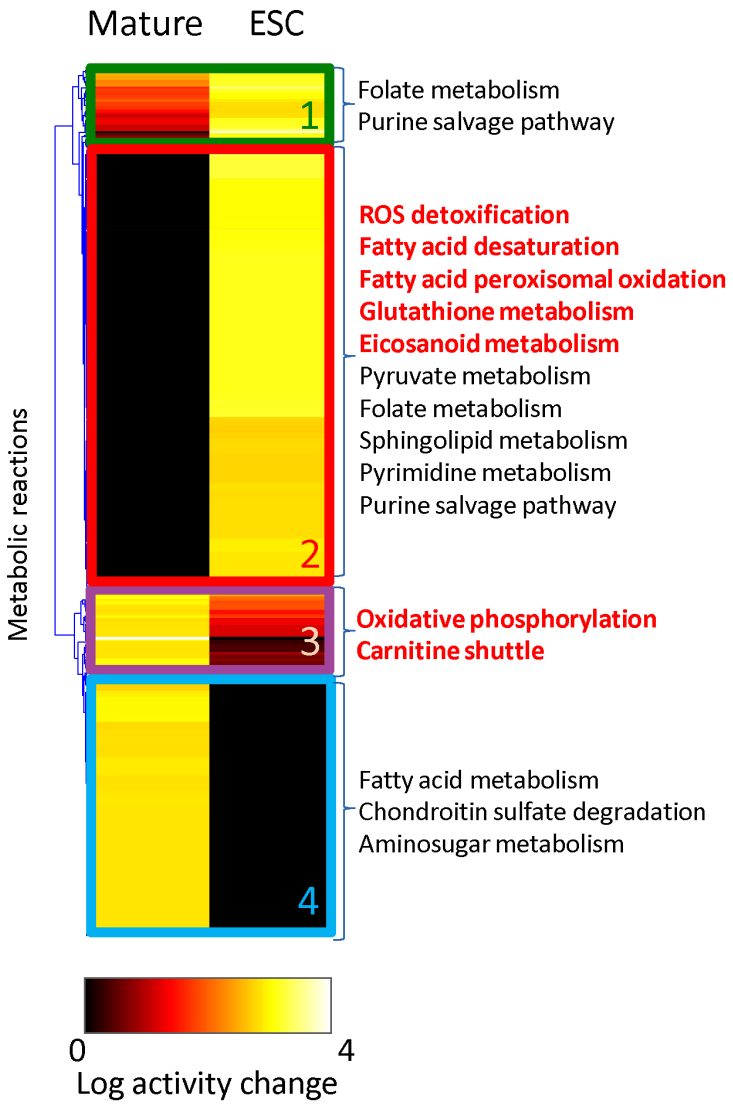


Figure 6.4: Heatmap of the highest metabolic reaction differences (> 80th-percentile) between ESCs and mature populations.

Lighter colors (yellow and white) correspond to the largest changes. Cluster 1 represents reaction activities that are up-regulated in ESCs relative to mature populations, namely such as folate and purine metabolism. Cluster 2 shows reactions that are highly active in ESCs, such as fatty acid oxidation and desaturation and oxidative stress-related pathways (ROS detoxification and glutathione metabolism). Cluster 3 and cluster 4 show up-regulated reactions in mature populations that are consistent with increased mitochondrial energy metabolism.

6.3 *In vitro* studies implicate redox state to mediate ESC pluripotency and differentiation

The mediation of ESC redox metabolism, in particular fatty acid desaturation and oxidative stress, was hypothesized to play a critical role during the differentiation process based on the metabolomic analysis. To evaluate the consequence of these activities on the ESC stemness phenotype, the enzyme activities of phospholipase A2 (PLA_2), 5Δ and 6Δ desaturases, cyclooxygenase (COX) and lipooxygenase (LOX) were inhibited *in vitro*. 5Δ and 6Δ desaturases are involved in the production of unsaturated fatty acids (e.g. arachidonic acid, eicosapentaenoic acid) from linoleic acid. PLA_2 hydrolyzes phospholipids from the cellular membrane to release unsaturated fatty acids. COX and LOX both oxidize C=C from the PLA_2-released precursors to produce important biological eicosanoid mediators of the inflammatory response, producing reactive oxygen species (ROS) in the process. The enzymatic activities were inhibited *in vitro* by culturing mESCs with their respective inhibitors and pluripotency was measured through monitoring the expression of the transcription factor markers Nanog and Oct4 that are heavily involved with self-renewal of undifferentiated ESCs [150, 151]. The results showed that there was a loss of differentiation and promotion of ESC self-renewal when reaction activities related to the oxidative stress response were inhibited. In addition, differentiation was delayed when 5Δ and 6Δ desaturases were inhibited in culture. Thus, the *in vitro* analysis demonstrated that the oxidative cascades mediating transformation of chemically reactive (unsaturated) metabolites are crucial in the mediation of stem cell differentiation as its inhibition promotes the pluripotent state of ESCs.

Glutathione levels were also measured in ESCs to evaluate the cellular redox state of ESCs (Figure 6.5). The measured metabolic signature of ESCs suggested that a general reduced cellular state is maintained during its pluripotent state and switches to a highly oxidative state upon differentiation. Results from the network-based analysis indicated glutathione metabolism and ROS detoxification to play a key role in mediating the redox status of ES cells. A primary component of the redox system is the mediation of glutathione metabolism, in which reduced reduced glutathione (GSH) is a major antioxidant that protects cells from ROS formation. GSH concentrations were determined in ESCs at day 0, 3, and 7 of differentiation, which showed that GSH levels decreased as a function of time of differentiation in both growth and chemically defined media. Overall, the *in vitro* results support the role of redox maintenance as a key determinant in promoting ESC

differentiation.

6.4 Network analysis of *in vitro* inhibited enzyme activities link stemness phenotype to broader metabolic effects

The inhibition of oxidative cascades led to the observed reduction of ESC differentiation and promotion of its pluripotent state. To further examine the broader metabolic pathways mediated during ESC self-renewal and differentiation, we used a network approach to evaluate their respective inhibited and activated effects of oxidative (COX, LOX, PLA2) and fatty acid desaturase (5Δ and 6Δ desaturases) pathways on the global human metabolic network [3]. The global effect of inhibiting and activating oxidative pathways in the human metabolic network was evaluated by using the FVA approach described in methods. Specifically, we examined pathway reactions involving PLA2, COX, LOX, and 5Δ-6Δ desaturase enzymes. When the oxidative cascades were inhibited, and thus promoting the pluripotent state, the most highly active reactions (>80th-percentile, Wilcoxon $p < 10^{-10}$) involve glycolysis and pyruvate metabolism (Figure 6.6A). The biosynthesis of sphingolipids, inositol phosphate, nucleotides, glycerophospholipids, and triacylglycerols is also highly favorable under inhibitory conditions. Formyltetrahydrofolate dehydrogenase, which participates in the one carbon pool by folate, was predicted to be active in the pluripotent state. In addition, increased activity of the oxidative phase in the pentose phosphate pathway provides NADPH required in reductive biosynthetic processes. In general, the network analysis reveals that the maintenance of the pluripotent phenotype in ESCs is associated with an active biosynthetic state (i.e. anabolic metabolism) that would be favorable under hypoxic conditions to promote its self-renewal.

The activation of oxidative cascades results in a general increase in cellular metabolic reactions relative to the inhibitory conditions. We found that the most significantly increased reactions (Bonferonni-corrected $p < 10^{-5}$) relative to the pluripotent (inhibitory oxidative) state are oxidative phosphorylation (e.g., ATP synthase), ROS detoxification (e.g., glutathione peroxidase, superoxide dismutase, catalase, peroxidase) mitochondrial and peroxisomal β-oxidation, purine catabolism (e.g., xanthine oxidase) and glutathione metabolism (e.g., glutathione peroxidase). These results show that activation of the synthesis and oxidation of polyunsaturated fatty acids in ESCs is associated with increased activities of other pro-oxidative reactions. Inhibition, in contrast, increases anabolic reaction activities that are associated with pluripotency. In the presence of oxygen, COX-

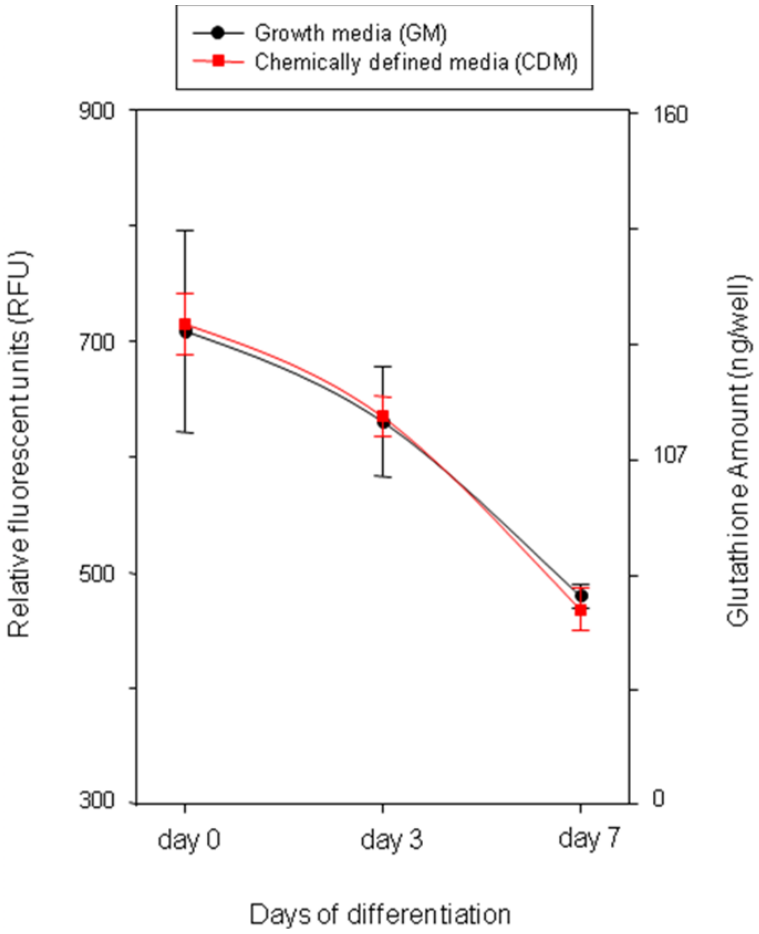


Figure 6.5: Reduced glutathione (GSH) levels as a function of days of ESC differentiation. GSH levels were measured at day 0 (mESC state) in both GM and CDM and then day 3 and 7 in N2B27 differentiated cultures from cells first cultured in GM and CDM. GSH levels were significantly lower by day 7, indicative of a highly oxidized state.

and LOX-mediated oxidation leads to the production of downstream eicosanoids such as prostaglandins [152]. The network analysis shows that activation of these oxidative pathways is associated with an overall increase in oxidative stress of the cell (e.g. superoxide dismutase, catalase, glutathione peroxidase). The result is consistent with recent experimental observations that ROS accumulation and signaling is required for differentiation [153] and that the regulation of pro-oxidative enzymes such as NADPH oxidase, and subsequent ROS generation, as well as antioxidants (e.g. glutathione peroxidase) are prerequisites for differentiation [154, 155, 156].

6.5 Discussion

The integration of ESC metabolomic data with the Recon 1 network provides a systematic approach to characterizing the stemness phenotype. A constraint-based analysis of the metabolomic data using the global metabolic network revealed that the unsaturated metabolites determined to be at higher levels in ESCs are primarily associated with oxidative mediation. The *in vitro* analysis supported that redox metabolism is a primary mediator of pluripotent and differentiation phenotypes as 1) inhibited fatty acid desaturation pathways delayed differentiation; 2) inhibited oxidative pathways of polyunsaturated fatty acids promote the pluripotent state; and 3) reduced glutathione levels dropped upon stem cell differentiation. In addition, further network analysis showed that when the ESC pluripotent state is promoted by inhibiting enzymes involved in the synthesis and oxidation of polyunsaturated fatty acids, a broader anabolic effect on metabolism is observed. Thus, the results indicate a strong association between pluripotency and intracellular oxidative regulation, supported by previous studies demonstrating that stem cells contain lower levels of ROS than their mature progeny [157, 158, 153].

Hypoxic conditions are known to maintain the pluripotent and undifferentiated phenotype of stem/precursor cells both in vitro and in vivo [159, 160, 161]. The observation is consistent with the existence of hypoxic niches housing stem cells in specific anatomic locations and in developing embryos [162]. Increased glycolysis and pyruvate metabolic pathways associated with the ESC pluripotent state is highly favorable for adapting under hypoxic conditions and is further supported by the fact that these pathways are regulated by hypoxia-inducible factors (HIFs) [163, 164]. The network analysis supports that stem cells inherently sustain highly reactive structural precursors under a reduced cellular state and, together with hypoxic conditions, makes them particularly sensitive to differentiate in

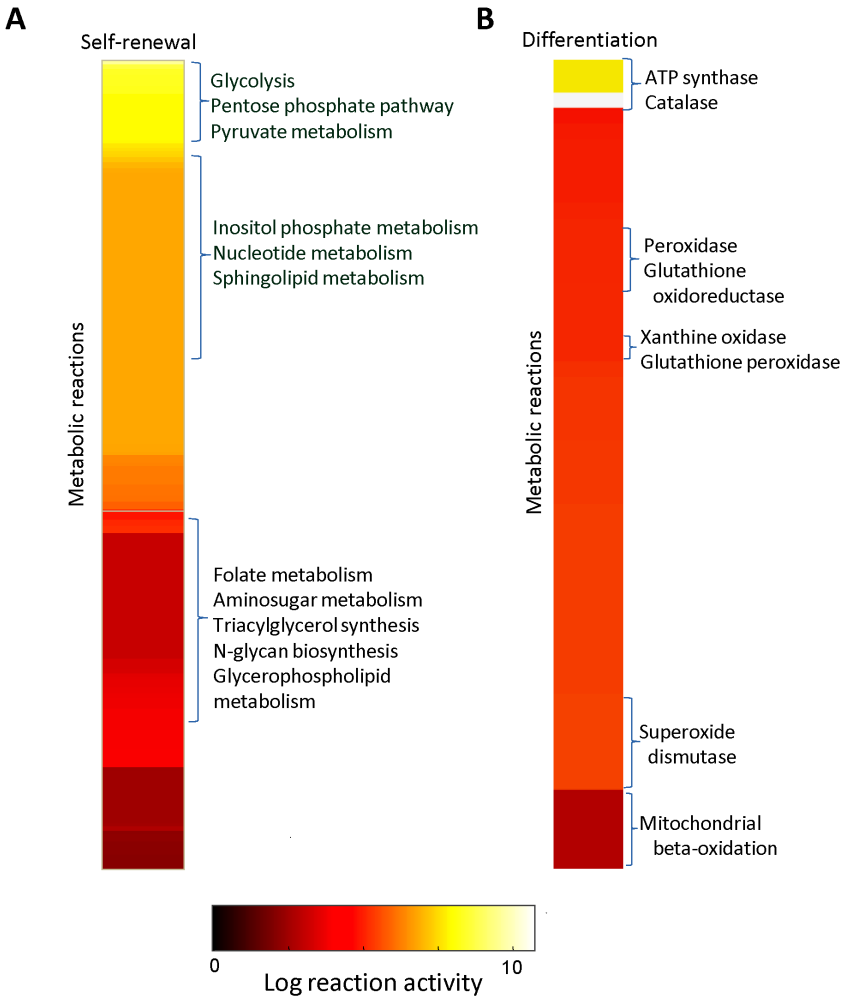


Figure 6.6: Broader reaction activities associated with inhibiting (self-renewal) and activating (differentiation) oxidative (COX, LOX, PLA2) and fatty acid desaturase (5Δ and 6Δ desaturases) pathways.

(A) Heatmap of highest reaction activities associated with inhibition (self-renewal) of oxidative cascades (>80th percentile). High glycolysis and pyruvate metabolism is supportive of hypoxic environments, and anabolic metabolism is highly upregulated. (B) Heatmap of reaction activities associated with activation (differentiation) of oxidative cascades (Bonferroni-corrected $p<10^{-5}$). Oxidative stress activities are indicated as significantly increased in ESC differentiation state.

response to oxidation and its consequent oxidative processes. Overall, this study strongly supports that the activation of oxidative metabolism is a metabolic signature of stem cell differentiation.

The text of this chapter, in part, is a reprint of the material as it will likely appear in Yanes, O., Clark, J., Mo, M.L., Wong, D.M., Sanchez-Ruiz, A., Benton, P., Trauger, S.A., Desponts, C., Patti, G.J., Palsson, B.O., Ding, S., Siuzdak, G., Highly reactive endogenous metabolites characterize embryonic stem cells. Nature Chemical Biology (In review). I was the primary author of the text in that portion of the publication and the co-authors participated and/or supervised the research which forms the basis for this chapter.

Chapter 7

In closing

We have reached a mature stage with the development of metabolic systems biology in microorganisms, and the extension of this approach to more complex organisms has been initiated with existing networks for yeast and human metabolism. A key part of this process is the bottom-up reconstruction of genome-scale metabolic networks based on our current knowledge, which requires detailed curation and incorporation of genomic and biochemical information into a mathematically-structured network. These networks facilitate systems analysis of the organism in which the network is reconstructed for by using various constraint-based methods that have been developed and established. The studies presented in this dissertation focused extensively on refining constraint-based approaches for the evaluation of "omics"-based phenotypes.

Chapters two and three described the reconstruction process and validation of genome-scale networks for yeast and human metabolism, which serve as the foundations for the "omics"-driven systems modeling in subsequent chapters. The studies described in the latter chapters of the dissertation highlighted constraint-based modeling of metabolic networks to integrate and analyze transcriptomic and metabolomic data. The use of network reconstructions as a "context for content" refines metabolic capabilities captured by experimental "omics" measurements and enables the direct computation of the "omics"-to-phenotype relationship. Chapter four presented a study that adapted the global human metabolic network to re-interpret previously published pharmacogenomic expression data. The results in the analysis showed that systemic metabolic phenotypes evaluated for cell lines could be used to interpret mechanisms of drug response based on pharmacogenomic data. Chapters five and six focused on the integration and analysis of metabolomic data to characterize network behavior at its ultimate endpoint phenotypic level. First, a study on

evaluating the ammonium assimilation process in yeast based on extracellular metabolomic profiles was presented as a proof-of-concept analysis that metabolite data can be incorporated as demand constraints in network reconstructions. Chapter six presented the integrative use of metabolomic data, genome-scale networks, and *in vitro* experimentation to study physiological aspects of stem cell metabolism that was previously not well-explored. Taken together, these applications have demonstrated that constraint-based modeling of high-throughput data in the context of genome-scale networks can broaden understanding of the "omics"-to-phenotype relationship and help drive the biological discovery process.

7.1 Lessons learned

Coming into the systems biology field, my naivete had me initially believe that I could build a model that would accurately capture physiological behavior. It seemed like a fairly simple process: build a model, compute, and output all that one needed to know about that system. Over the years, my idealism was replaced with the realization that no computational model will ever be a "perfect" one. After all, how can we reconstruct a perfect replica of a cellular system when it is not the system itself? It is important to remember the network models described in this dissertation are an engineering-applied approximation of biology, with modeling assumptions that can seem overly presumptuous. Nevertheless, just as all other engineering principles have moved towards model simulations as standard practice in their respective fields, the representation of biology in its computable form is what enables its systems-level analysis and, ultimately, the prediction of its behavior.

The bottom-up network approach to analyzing eukaryotic systems was established for yeast metabolism, laying the foundation for utilizing similar approaches for human systems. This process seems rather straightforward in principle, but as the saying goes, not all (models) are made equal. Going from a smaller yeast reaction network to the global human network, which more than doubled in size, proved to be more challenging in its analysis and interpretation. However, while there were many more components and interactions to consider, the interconnected nature of metabolic networks still makes it possible to identify both broader effects and key metabolic nodes that conditions of interest are mediated in. Ideally, we would be able to use the method approaches described in this dissertation to pinpoint an exact cause of disruption or modification to the cellular system. However, perturbations often result in systemic changes that are distal from the affected site and, in some cases, the primary changes may occur away from the perturbed region.

Hence, the best way to address this problem currently is to quantify or rank order metabolic regions by their level of participation, as was described in the method approaches used in the dissertation studies. While these approaches may not give an exact answer pertaining to a biological mechanism, the ability to identify broader levels of impact for different metabolic pathways and regions can be just as significant and relevant.

Systems biology has the potential to be applied in many different avenues of biological research as was demonstrated in the presented studies. While the modeling of human physiology is still in its infancy, its future advancements will ultimately drive biomedical discovery and development in disease and health. Top-down, statistical methods in systems biology have been useful in the computation of all measured molecular components and is more comprehensive in nature; however, inferring phenotypic relationships without context can often be inconclusive. The advantage of using the bottom-up approach that was extensively described in this dissertation is that the component-to-component relationships are generally well-defined. The connections between nodes are predefined, and generally speaking, any errors or falsely defined relationships in such types of models are traceable to the reconstruction process. Ideally, if technological advances will one day be capable of identifying all component interactions in a biological system, all models can be described in a bottom-up fashion as all known cellular constituents can be defined (i.e. a "whole-cell" model).

7.2 Future directions

While genome-scale networks have proven to be useful in elucidating biological behavior, what we can learn is still limited by the content and scope of information that is incorporated into the reconstructed network. Therefore, future advances must be made towards the integration and modeling of multiple tissue-specific networks and other biological systems to improve its physiological relevance and predictive potential. Genome-scale metabolic networks are a starting point towards modeling more complex biological processes in humans. The requirement for detailed interaction information highlights the limitations of the bottom-up network reconstructions since it is dependent on the information that is already known or confirmed. As molecular biology technologies continue to advance, coarse-grained details will become more refined and require incorporating additional levels of information. An intuitive next step is to extend the scope of human molecular systems modeling from metabolism to other biological systems as more details on their mechanisms

and interactions are elucidated. The application of similar methodologies used in reconstructing metabolism has been prototyped at a smaller scale and applied towards human signal transduction pathways [73, 166], as well as at the genome-scale for E. coli transcriptional and translational machinery [167]. The merging of these systems would indeed represent a more accurate and comprehensive description of the inherently dependent nature of cellular processes [168]. Additionally, integration of transcriptional regulation with metabolism has already been achieved for microbial networks in the form of Boolean logic networks [169, 45, 170]. While the human transcriptional regulation is markedly more complex than in microorganisms, the general framework has been formulated and can be implemented at a genome-scale as detailed data on regulatory component interactions becomes available. Such a comprehensive network would undoubtedly improve the integrative analysis of "omics" data as it would account for more components and interactions that occur under different biological conditions.

Ultimately, a multi-compartment human model can provide a better depiction of physiologically relevant metabolic states. Efforts have already been initiated towards developing context-specific metabolic networks from high-throughput data [93, 36], providing the basis for constructing specific, segmented networks from a global metabolic network such as Recon 1. Accurate representations of individual cellular and tissue compartments will be essential to such higher-level modeling and will require rigorous assessment of their unique metabolic functions and demands. Once individual networks have been formulated, they can be integrated to form multi-network models. A recent study modeling a codependent, two-organism microbial system showed to accurately predict known physiological features when considering the flux exchange and interaction between separate metabolic networks [165]. Indeed, similar approaches can also be applied in principle when modeling metabolite exchange and interaction between multiple tissue metabolic networks. With the methods to constructing and integrating context-specific metabolic networks in different cell- or tissue-types at hand, a multi-compartmentalized, "whole-body" model of human metabolism can soon be realized.

The text of chapter, in part, is a reprint of the material as it appears in M.L. Mo and B.Ø. Palsson. 2009. Understanding human metabolic physiology: A genome-to-systems approach. Trends in Biotechnology, 27(1):37-44. I was the primary author of this publication and the co-author participated and supervised the research, which forms the basis for this chapter.

Bibliography

[1] B. Ø. Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 2006.

[2] P. Romero, 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*, 6(1):R2, 2005.

[3] N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ø. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS U S A*, 104:1777–1782, 2007.

[4] H. Ma, A. Sorokin, A. Mazein, A. Selkov, E. Selkov, O. Demin, and I. Goryanin. The edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol*, 3:135, 2007.

[5] N. Matsuzawa-Nagata, T. Takamura, H. Ando, S. Nakamura, S. Kurita, H. Misu, T. Ota, M. Yokoyama, M. Honda, K.I. Miyamoto, and S. Kaneko. Increased oxidative stress precedes the onset of high-fat diet-induced insulin resistance and obesity. *Metabolism*, 57(8):1071–77, 2008.

[6] S. de Ferranti and D. Mozaffarian. The perfect storm: obesity, adipocyte dysfunction, and metabolic consequences. *Clin Chem*, 54(6):945–55, 2008.

[7] J.A. Martinez, M.D. Parra, J.L. Santos, M.J. Moreno-Aliaga, A. Marti, and M.A. Martinez-Gonzalez. Genotype-dependent response to energy-restricted diets in obese subjects: towards personalized nutrition. *Asia Pac J Clin Nutr*, 17 Suppl 1:119–22, 2008.

[8] S.K. Park and T.A. Prolla. Lessons learned from gene expression profile studies of aging and caloric restriction. *Ageing Res Rev*, 4(1):55–65, 2005.

[9] T.A. Crowder, M.D. Beekley, R.X. Sturdivant, C.A. Johnson, and A. Lumpkin. Metabolic effects of soldier performance on a simulated graded road march while wearing two functionally equivalent military ensembles. *Mil Med*, 172(6):596–602, 2007.

[10] B.C. Nindl, J.A. Alemany, M.D. Kellogg, J. Rood, S.A. Allison, A.J. Young, and SJ. Montain. Utility of circulating igf-i as a biomarker for assessing body composition changes in men during periods of high physical activity superimposed upon energy and sleep restriction. *J Appl Physiol*, 103(1):340–6, 2007.

[11] D. Feuvray and A. Darmellah. Diabetes-related metabolic perturbations in cardiac myocyte. *Diabetes Metab*, 34 Suppl 1:S3–9, 2008.

[12] D. LeRoith. Dyslipidemia and glucose dysregulation in overweight and obese patients. *Clin Cornerstone*, 8(3):38–52, 2007.

[13] N.J. Serkova, J.L. Spratlin, and S.G. Eckhardt. Nmr-based metabolomics: translational application and treatment of cancer. *Curr Opin Mol Ther*, 9(6):572–85, 2007.

[14] Robak T., E. Lech-Maranda, A. Korycka, and E. Robak. Purine nucleoside analogs as immunosuppressive and antineoplastic agents: mechanism of action and clinical activity. *Curr Med Chem*, 13(26):3165–89, 2006.

[15] S Rozen, M.E. Cudkowicz, M. Bogdanov, W.R. Matson, B.S. Kristal, C Beecher, S. Harrison, P. Vouros, J Flarakos, K. Vigneau-Callahan, T.D. Matson, K.M. Newhall, M.F. Beal, R.H. Brown, and R. Kaddurah-Daouk. Metabolomic analysis and signatures in motor neuron disease. *Metabolomics*, 1(2):101–108, 2005.

[16] R. Kaddurah-Daouk, J. McEvoy, R.A. Baillie, D. Lee, J.K. Yao, P.M. Doraiswamy, and K.R. Krishnan. Metabolomic mapping of atypical antipsychotic effects in schizophrenia. *Mol Psychiatry*, 12(10):934–45, 2007.

[17] B.P. Sokolov, L. Jiang, N.S. Trivedi, and C. Aston. Transcription profiling reveals mitochondrial, ubiquitin and signaling systems abnormalities in postmortem brains from subjects with a history of alcohol abuse or dependence. *J Neurosci Res*, 72(6):756–67, 2003.

[18] Zhou. L., H. Huang, T.A. McElfresh, D.A. Prosdocimo, and W.C. Stanley. Impact of anaerobic glycolysis and oxidative substrate selection on contractile function and mechanical efficiency during moderate severity ischemia. *Am J Physiol Heart Circ Physiol*, 293(3):H939–945, 2008.

[19] S. A. Becker and B. Ø. Palsson. Genome-scale reconstruction of the metabolic network in staphylococcus aureus n315: an initial draft to the two-dimensional annotation. *BMC Microbiology*, 5(1):8, 2005.

[20] J. S. Edwards and B. Ø. Palsson. Systems properties of the haemophilus influenzae rd metabolic genotype. *Journal of Biological Chemistry*, 274(25):17410–6, 1999.

[21] J. S. Edwards and B. Ø. Palsson. The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *PNAS U S A*, 97(10):5528–5533, 2000.

[22] J. L. Reed, T. D. Vo, C. H. Schilling, and B. Ø. Palsson. An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr)". *Genome Biology*, 4(9):R54.1–R54.12, 2003.

[23] C.H. Schilling, M.W. Covert, I. Famili, G.M. Church, J.S. Edwards, and B. Ø. Palsson. Genome-scale metabolic model of helicobacter pylori 26695. *Journal of Bacteriology*, 184(16):4582–4593, 2002.

[24] I. Thiele, T. D. Vo, N. D. Price, and B. Ø. Palsson. Expanded metabolic reconstruction of helicobacter pylori (iit341 gsm/gpr): an *in silico* genome-scale characterization of single- and double-deletion mutants. *J Bacteriol*, 187(16):5818–5830, 2005.

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

[26] A.M. Feist and B.O. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using escherichia coli. *Nat Biotech*, 26(6):659–67, 2008.

[27] J. L. Reed, I. Famili, I. Thiele, and B. Ø. Palsson. Towards multidimensional genome annotation. *Nat Rev Genet*, 7(2):130–141, 2006.

[28] M.L. Mo, N. Jamshidi, and B.O. Palsson. A genome-scale, constraint-based approach to systems biology of human metabolism. *Mol Biosyst*, 3(9):598–603, 2007.

[29] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–5, 2004. 0167-7799 Journal Article.

[30] J. L. Reed, T. R. Patel, K. H. Chen, A. R. Joyce, M. K. Applebee, C. D. Herring, O. T. Bui, E. M. Knight, S. S. Fong, and B. Ø. Palsson. Systems approach to genome annotation: Prediction and validation of metabolic functions. *Proc Natl Acad Sci U S A*, 103(46):17480–17484, 2006.

[31] V. Satish Kumar, M.S. Dasika, and C.D. Maranas. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, 8(212), 2007.

[32] B. O. Palsson. Two-dimensional annotation of genomes. *Nat Biotechnol*, 22(10):1218–9, 2004. 1087-0156 Journal Article.

[33] C.R. Farber and A.J. Lusis. Integrating global gene expression analysis and genetics. *Adv Genet*, 60:571–601, 2008.

[34] A.S. Haqqani, J.F. Kelly, and D.B. Stanimirovic. Quantitative protein profiling by mass spectrometry using isotope-coded affinity tags. *Methods Mol Biol*, 439:225–40, 2008.

[35] D.B. Kell. Metabolomics and systems biology: making sense of the soup. *Curr Opin Microbiol*, 7(3):296–307, 2004.

[36] T Shlomi, MN Cabili, MJ Herrgrd, B Palsson, and E Ruppin. Network-based prediction of human tissue-specific metabolism. *Nature Biotechnol*, 26(9):1003–10, 2008.

[37] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, M. Nikolai, L. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. Macinnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, and L. Querengesser. Hmdb: the human metabolome database. *Nucl. Acids Res.*, 35(Database issue):D521:6, 2007.

[38] A. M. Feist, M. J. Herrgard, I. Thiele, J. L. Reed, and B.O. Palsson. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 7(2):129–43, 2009.

[39] L.M. Raamsdonk, 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*, 19(1):45–50, 2001.

[40] M Kellis, N Patterson, M Endrizzi, B Birren, and ES Lander. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, 423(6937):241–54, 2003.

[41] A. Goffeau, 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*, 274(5287):546,563–7, 1996.

[42] J. Forster, I. Famili, P.C. Fu, B. Ø. Palsson, and J. Nielsen. Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network. *Genome Research*, 13(2):244–53, 2003.

[43] I. Famili, J. Forster, J. Nielsen, and B. Ø. 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.*, 100(23):13134–9, 2003.

[44] N. C. Duarte, M. J. Herrgard, and B. Ø. Palsson. Reconstruction and validation of saccharomyces cerevisiae ind750, a fully compartmentalized genome-scale metabolic model. *Genome Res*, 14(7):1298–1309, 2004.

[45] M. J. Herrgard, B. S. Lee, V. Portnoy, and B. Ø. Palsson. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in saccharomyces cerevisiae. *Genome Research*, 16(5):627–35, 2006.

[46] T. Shlomi, M.J. Herrgard, V. Portnoy, E. Naim, B.. Palsson, R. Sharan, and E. Ruppin. Systematic condition-dependent annotation of metabolic genes. *Genome Res*, 17(11):1626–33, 2007.

[47] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: The cobra toolbox. *Nat. Protocols*, 2:727–738, 2007.

[48] T. L. Nissen, U. Schulze, J. Nielsen, and J. Villadsen. Metabolic functions of duplicate genes in saccharomyces cerevisiaeflux distributions in anaerobic, glucose-limited continuous cultures of saccharomyces cerevisiae. *Microbiology*, 143(Pt 1):203–18, 2005.

[49] L. Kuepfer, U. Sauer, and L.M. Blank. Metabolic functions of duplicate genes in saccharomyces cerevisiae. *Genome Res*, 15(10):1421–1430, 2005.

[50] J. S. Edwards, M. Covert, and B. Ø. Palsson. Metabolic modeling of microbes: the flux balance approach. *Environmental Microbiology,*, 4(3):133–140, 2002.

[51] A. Varma and B. Ø. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Biotechnology*, 12:994–998, 1994.

[52] D. Segre, D. Vitkup, and G. M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*, 99(23):15112–7, 2002.

[53] P. Baldi, S. Brunak, Y. Chauvin, C.A. Andersen, and H. Nielsen. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, 16(5):412–24, 2000.

[54] M.J. Herrgard, N. Swainston, P. Dobson, W.B. Dunn, K.Y. Arga, M. Arvas, N. Bluthgen, S. Borger, R. Costenoble, M. Heinmann, E. Klipp, Palsson B.O., U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen, and D. B. Kell. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, 26:1155–60, 2008.

[55] S.G. Villas-Boas, J.F. Moxley, M. Akesson, G. Stephanopoulos, and J. Nielsen. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem J*, 388(Pt 2):669–77, 2005.

[56] S.A. Becker and B.O. Palsson. Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Syst Biol*, 2(14), 2008.

[57] D.M. Muoio and C.B. Newgard. Obesity-related derangements in metabolic regulation. *Annu Rev Biochem*, 75:367–401, 2006.

[58] C.J. Stein and G.A. Colditz. The epidemic of obesity. *J Clin Endocrinol Metab*, 89(6):2522–5, 2004.

[59] S. Matthaei, M. Stumvoll, M. Kellerer, and H.U. Haring. Pathophysiology and pharmacological treatment of insulin resistance. *Endocr Rev*, 21(6):585–618, 2000.

[60] J.J. Park, 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*, 27(2):114–21, 2006.

[61] C. Power, S.K. Miller, and P.T. Alpert. Promising new causal explanations for obesity and obesity-related diseases. *Biol Res Nurs*, 8(3):223–33, 2007.

[62] L. Afman and M. Muller. Nutrigenomics: from molecular nutrition to prevention of disease. *J Am Diet Assoc*, 106(4):569–76, 2006.

[63] M. Muller and S. Kersten. Nutrigenomics: goals and strategies. *Nat Rev Genet*, 4(4):315–22, 2003.

[64] Z. Wunderlich and L. A. Mirny. Using the topology of metabolic networks to predict viability of mutant strains. *Biophys J*, 91(6):2304–11, 2006.

[65] D. Vitkup, P. Kharchenko, and A. Wagner. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol*, 7(5):R39, 2006.

[66] M.C. Palumbo, A. Colosimo, A. Giuliani, and L. Farina. Functional essentiality from topology features in metabolic networks: a case study in yeast. *FEBS Lett*, 579(21):4642–6, 2005.

[67] A. Samal, 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*, 7(1):118, 2006.

[68] E. V. Nikolaev, A. P. Burgard, and C. D. Maranas. Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions. *Biophys J*, 88(1):37–49, 2005.

[69] S.J. Wiback, R. Mahadevan, and B.O. Palsson. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The escherichia coli alpha-spectrum. *Biotechnology and Bioengineering*, 86(3):317–31, 2004.

[70] N.D. Price, 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*, 225(2):185–194, 2003.

[71] S. J. Wiback, R. Mahadevan, and B. Ø. Palsson. Reconstructing metabolic flux vectors from extreme pathways: Defining the alpha-spectrum. *Journal of Theoretical Biology*, 24(3):313–324, 2003.

[72] I. Famili and B. Ø. Palsson. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys J*, 85(1):16–26, 2003.

[73] J. A. Papin and B. Ø. Palsson. The jak-stat signaling network in the human b-cell: an extreme signaling pathway analysis. *Biophys J*, 87(1):37–46, 2004.

[74] J. L. Reed and B. O. Palsson. Thirteen years of building constraint-based in silico models of escherichia coli. *J Bacteriol*, 185(9):2692–9, 2003. 0021-9193 Journal Article.

[75] I. Thiele, N. D. Price, T. D. Vo, and B. Ø. Palsson. Candidate metabolic network states in human mitochondria. impact of diabetes, ischemia, and diet. *J Biol Chem*, 280(12):11683–11695, 2005.

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

[77] T. D. Vo, H. J. Greenberg, and B. Ø. Palsson. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem*, 279(38):39532–39540, 2004.

[78] R.M. Mayers, B. Leighton, and E. Kilgour. Pdh kinase inhibitors: a novel therapy for type ii diabetes? *Biochem Soc Trans.*, 33(Pt 2):367–70, 2005.

[79] A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome. Res.*, 14(2):301–12, 2004.

[80] N. Jamshidi and B. Ø. Palsson. Systems biology of snps. *Mol Syst Biol*, 2:38, 2006.

[81] T Kayo, DB Allison, R Weindruch, and TA Prolla. Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. *Proc Natl Acad Sci U S A*, 98(9):5093–8, 2001.

[82] D.S. Lee, J. Park, K.A. Kay, N.A. Christakis, Z.N. Oltvai, and A.L. Barabasi. The implications of human metabolic network topology for disease comorbidity. *Proc Natl Acad Sci U S A*, 105(29):9880–5, 2008.

[83] A.L. Hopkins and C.R. Groom. The druggable genome. *Nat Rev Drug Discov*, 1(9):727–30, 2007.

[84] P. Imming, C. Sinning, and A. Meyer. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov*, 5(10):821–34, 2006.

[85] M.L. MacDonald, J. Lamerdin, S. Owens, B.H. Keon, G.K. Bilter, Z. Shang, Z. Huang, H. Yu, J. Dias, T. Minami, S.W. Michnick, and J.K. Westwick. Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol*, 2(6):329–37, 2006.

[86] MA Yildirim, KI Goh, ME Cusick, AL Barabàsi, and M. Vidal. Drug-target network. *Nat Biotechnol*, 25(10):1119–1126, 2007.

[87] M. Campillos, M. Kuhn, A.C. Gavin, L.J. Jensen, and P. Bork. Drug target identification using side-effect similarity. *Science*, 321(5886):263–6, 2008.

[88] J.M. Schwartz and J.C. Nacher. Local and global modes of drug action in biochemical networks. *BMC Chem Biol*, 9:4, 2009.

[89] L. Cabusora, E. Sutton, A. Fulmer, and C.V. Forst. Differential network expression during drug and stress response. *Bioinformatics*, 21(12):2898–905, 2005.

[90] Z. Kutalik, J.S. Beckmann, and S. Bergmann. A modular approach for integrative analysis of large-scale gene-expression and drug-response data. *Nat Biotech*, 26(5):531–9, 2008.

[91] Y. Ma, Z. Ding, Y. Qian, X. Shi, V. Castranova, E.J. Harner, and L. Guo. Predicting cancer drug response by proteomic profiling. *Clin Cancer Res*, 12(15):4583–9, 2006.

[92] A.A. Cohen, N. Geva-Zatorsky, E. Eden, M. Frenkel-Morgenstern, I. Issaeva, A. Sigal, R. Milo, C. Cohen-Saidon, Y. Liron, Z. Kam, L. Cohen, T. Danon, N. Perzov, and U. Alon. Dynamic proteomics of individual cancer cells in response to a drug. *Science*, 322(5907):1511–6, 2008.

[93] S.A. Becker and B.O. Palsson. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*, 4(5):e1000082, 2008.

[94] J Lamb, ED Crawford, D Peck, JW Modell, IC Blat, MJ Wrobelm, J Lerner, JP Brunet, A Subramanian, KN Ross, M Reich, H Hieronymus, G Wei, SA Armstrong, SJ Haggarty, PA Clemons, R Wei, SA Carr, ES Lander, and TR Golub. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–35, 2006.

[95] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y. Yang, and J. Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 5(10):R80, 2004.

[96] K. Ou, K. Yu, D. Kesuma, M. Hooi, N. Huang, W. Chen, S.Y. Lee, X.P. Goh, L.K. Tan, J. Liu, S.Y. Soon, S. Bin Abdul Rashid, T.C. Putti, H. Jikuya, T. Ichikawa, O. Nishimura, M. Salto-Tellez, and P. Tan. Novel breast cancer biomarkers identified by integrative proteomic and gene expression mapping. *J Proteome Res*, 7(4):1518–28, 2008.

[97] K. R. Patil and J. Nielsen. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A*, 102(8):2685–9, 2005.

[98] DS Wishart, C Knox, AC Guo, S Shrivastava, M Hassanali, P Stothard, Z Chang, and J. Woolsey. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*, 1(34):D668–72, 2006.

[99] D. C. Liebler and F. P. Guengerich. Elucidating mechanisms of drug-induced toxicity. *Nat Rev Drug Discov*, 4:410–420, May 2005.

[100] W.B. Parker. Enzymology of purine and pyrimidine antimetabolites used in the treatment of cancer. *Chem Rev.*, 109(7):2880–93, 2009.

[101] M. Guppy, P. Leedman, X. Zu, and V. Russell. Contribution by different fuels and metabolic pathways to the total atp turnover of proliferating mcf-7 breast cancer cells. *Biochem J*, 364(Pt 1):309–15, 2002.

[102] R. Moreno-Sanchez, S. Rodrguez-Enrquez, A. Marn-Hernndez, and E. Saavedra. Energy metabolism in tumor cells. *FEBS J*, 274(6):393–418, 2007.

[103] K. Wada, N. Yamada, T. Sato, H. Suzuki, i M. Mik, Y. Lee, K. Akiyama, and S. Kuroda. Corticosteroid-induced psychotic and mood disorders: diagnosis defined by dsm-iv and clinical pictures. *Psychosomatics*, 42(6):461–6, 2001.

[104] S. Kim, H.J. Shin, S.Y. Kim, J.H. Kim, Y.S. Lee, D.H. Kim, and M.O. Lee. Genistein enhances expression of genes involved in fatty acid catabolism through activation of pparalpha. *Mol Cell Endocrinol*, 220(1-2):51–8, 2004.

[105] K. Mokrzycki. Anti-atherosclerotic efficacy of quercetin and sodium phenylbutyrate in rabbits. *Ann Acad Med Stetin*, 46:189–200, 2000.

[106] D. Kritchevsky, A.W. Moyer, W.C. Tesar, R.F. McCandless, J.B. Logan, R.A. Brown, and M. Englert. The effect of sodium 2-phenylbutyrate in experimental atherosclerosis. *Angiology*, 7(2):156–8, 1956.

[107] W.C. Duane and N.B. Javitt. 27-hydroxycholesterol: production rates in normal human subjects. *J Lipid Res*, 40(7):1194–9, 1999.

[108] J.M. Corton, J.G. Gillespie, S.A. Hawley, and D.G. Hardie. 5-aminoimidazole-4-carboxamide ribonucleoside. a specific method for activating amp-activated protein kinase in intact cells? *Eur J Biochem*, 229(2):558–65, 1995.

[109] S.C. Mastbergen, N.W. Jansen, J.W. Bijlsma, and F.P. Lafeber. Differential direct effects of cyclo-oxygenase-1/2 inhibition on proteoglycan turnover of human osteoarthritic cartilage: an in vitro study. *Arthritis Res Ther*, 8(2):R2, 2006.

[110] H.S. Kim, J.S. Park, S.J. Hong, M.S. Woo, S.Y. Kim, and K.S. Kim. Regulation of the tyrosine hydroxylase gene promoter by histone deacetylase inhibitors. *Biochem Biophys Res Commun*, 312(4):950–7, 2003.

[111] A. D'Souza, E. Onem, P. Patel, E.F. La Gamma, and B.B. Nankova. Valproic acid regulates catecholaminergic pathways by concentration-dependent threshold effects on th mrna synthesis and degradation. *Brain Res*, 1247:1–10, 2009.

[112] L.K. Vainionp, K. Mikkonen, J. Rtty, M. Knip, A.J. Pakarinen, V.V. Myllyl, and J.I. Isojrvi. Thyroid function in girls with epilepsy with carbamazepine, oxcarbazepine, or valproate monotherapy and after withdrawal of medication. *Epilepsia*, 45(3):197–203, 2004.

[113] V. Lerner, C. Miodownik, A. Kaptsan, Y. Bersudsky, I. Libov, B.A. Sela, and E. Witztum. Vitamin b6 treatment for tardive dyskinesia: a randomized, double-blind, placebo-controlled, crossover study. *J Clin Psychiatry*, 68(11):1648–54, 2007.

[114] M.R. Owen, E. Doran, and A.P. Halestrap. Evidence that metformin exerts its anti-diabetic effects through inhibition of complex 1 of the mitochondrial respiratory chain. *Biochem J.*, 348 Pt 3:607–14, 2000.

[115] P. Schrauwen and M.K. Hesselink. Reduced tricarboxylic acid cycle flux in type 2 diabetes mellitus? *Diabetologia*, 51(9):1694–7, 2008.

[116] S. Rossell, C. C. van der Weijden, A. Lindenbergh, A. van Tuijl, C. Francke, B. M. Bakker, and H. V. Westerhoff. Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U.S.A.*, 103:2166–2171, Feb 2006.

[117] R. Kaddurah-Daouk, B. S. Kristal, and R. M. Weinshilboum. Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.*, 48:653–683, 2008.

[118] D B Kell, M Brown, H M Davey, W B Dunn, I Spasic, and S G Oliver. Metabolic footprinting and systems biology: the medium is the message. *Nat Rev Microbiol*, 3(7):557–565, Jul 2005.

[119] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, and D.B. Kell. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol*, 22(5):245–52, 2004.

[120] E.M. Lenz, J. Bright, I.D. Wilson, S.R. Morgan, and A.F. Nash. A 1h nmr-based metabonomic study of urine and plasma samples obtained from healthy human subjects. *J Pharm Biomed Anal*, 33(5):1103–15, 2003.

[121] J. Allen, H.M. Davey, D. Broadhurst, J.K. Heald, J.J. Rowland, S.G. Oliver, and D.B. Kell. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotech*, 21(6):692–6, 2003.

[122] J.K. Nicholson, J. Connelly, J.C. Lindon, and E. Holmes. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov*, 1(2):153–61, 2002.

[123] R.J. Mortishire-Smith, G.L. Skiles, J.W. Lawrence, S. Spence, A.W. Nicholls, B.A. Johnson, and J.K. Nicholson. Use of metabonomics to identify impaired fatty acid metabolism as the mechanism of a drug-induced toxicity. *Chem Res Toxicol*, 17(2):165–73, 2004.

[124] M.S. Sabatine, E. Liu, D.A. Morrow, E. Heller, R. McCarroll, R. Wiegand, G.F. Berriz, F.P. Roth, and R.E. Gerszten. Metabolomic identification of novel biomarkers of myocardial ischemia. *Circulation*, 112(25):3868–75, 2005.

[125] T. Cakir, C. Efe, D. Dikicioglu, B. Hortasu, A. Kirdar, and S.G. Oliver. Flux balance analysis of a genome-scale yeast model constrained by exometabolomic data allows metabolic system identification of genetically different strains. *Biotechnol Prog*, 23(2):320–6, 2007.

[126] J.W. Bang, D.J. Crockford, E. Holmes, F. Pazos, M.J. Sternberg, S.H. Muggleton, and J.K. Nicholson. Integrative top-down system metabolic modeling in experimental disease states via data-driven bayesian methods. *J Proteome Res*, 7(2):497–503, 2008.

[127] A.P. Oliveira, K.R. Patil, and J. Nielsen. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Syst Biol*, 2:17, 2008.

[128] N. D. Price, J. Schellenberger, and B. Ø. Palsson. Uniform sampling of steady state flux spaces: Means to design experiments and to interpret enzymopathies. *Biophysical Journal*, 87(4):2172–86, 2004.

[129] N. D. Price, I. Thiele, and Palsson B.O. Candidate states of helicobacter pylori's genome-scale metabolic network upon application of "loop law" thermodynamic constraints. *Biophysical J*, 90(11):3919–28, 2006.

[130] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. *Mol Syst Biol*, 3:119, 2007.

[131] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.

[132] T.L. Nissen, M.C. Kielland-Brandt, J. Nielsen, and J. Villadsen. Optimization of ethanol production in saccharomyces cerevisiae by metabolic engineering of the ammonium assimilation. *Metab Eng*, 2(1):69–77, 2000.

[133] C. Roca, J. Nielsen, and L. Olsson. Metabolic engineering of ammonium assimilation in xylose-fermenting saccharomyces cerevisiae improves ethanol production. *Appl Environ Microbiol*, 69(8):4732–6, 2003.

[134] M Moreira dos Santos, G Thygesen, P Ktter, L Olsson, and J Nielsen. Aerobic physiology of redox-engineered saccharomyces cerevisiae strains modified in the ammonium assimilation for increased nadph availability. *FEMS Yeast Res*, 4(1):59–68, 2003.

[135] IV Hartman, J.L. Buffering of deoxyribonucleotide pool homeostasis by threonine metabolism. *Proc Natl Acad Sci U S A*, 104(28):11700–5, 2007.

[136] C.L. Gelling, M.D. Piper, S.P. Hong, G.D. Kornfeld, and I.W. Dawes. Identification of a novel one-carbon metabolism regulon in saccharomyces cerevisiae. *J Biol Chem*, 279(8):7072–81, 2004.

[137] V. Denis and B. Daignan-Fornier. Synthesis of glutamine, glycine and 10-formyl tetrahydrofolate is coregulated with purine biosynthesis in saccharomyces cerevisiae. *Mol Gen Genet.*, 259(3):246–55, 1998.

[138] D.C. Hess, W. Lu, J.D. Rabinowitz, and D. Botstein. Ammonium toxicity and potassium limitation in yeast. *PLoS Biol*, 4(11):e351, 2006.

[139] S. Hjortmo, J. Patring, and T. Andlid. Growth rate and medium composition strongly affect folate content in saccharomyces cerevisiae. *Int J Food Microbiol*, 123(1-2):93–100, 2008.

[140] F. Kussmann, M. Raymond and M. Affolter. Omics-driven biomarker discovery in nutrition and health. *J Biotechnol*, 124(4):758–87, 2006.

[141] N.J. Serkova and C.U. Niemann. Pattern recognition and biomarker validation using quantitative 1h-nmr-based metabolomics. *Expert Rev Mol Diagn*, 6(5):717–31, 2006.

[142] B. Bhattacharya, S. Puri, and R. K. Puri. A review of gene expression profiling of human embryonic stem cell lines and their differentiated progeny. *Curr Stem Cell Res Ther*, 4:98–106, May 2009.

[143] H. Baharvand, M. Hajheidari, S. K. Ashtiani, and G. H. Salekdeh. Proteomic signature of human embryonic stem cells. *Proteomics*, 6:3544–3549, Jun 2006.

[144] H. Dihazi, G. H. Dihazi, J. Nolte, S. Meyer, O. Jahn, G. A. Mueller, and W. Engel. Multipotent adult germline stem cells and embryonic stem cells: comparative proteomic approach. *J. Proteome Res.*, Oct 2009.

[145] M.L. Mo, B.O. Palsson, and M.J. Herrgard. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol*, 3:37, 2009.

[146] T. Shlomi, M.N. Cabili, and E. Ruppin. Predicting metabolic biomarkers of human inborn errors of metabolism. *Mol Syst Biol*, 5:263, 2009.

[147] R Mahadevan and CH Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, 5:264–76, 2003.

[148] M.I. Sigurdsson, N. Jamshidi, J.J. Jonsson, and B.O. Palsson. Genome-scale network analysis of imprinted human metabolic genes. *Epigenetics*, 4(1):43–6, 2009.

[149] J. Wang, P. Alexander, L. Wu, R. Hammer, O. Cleaver, and S.L. McKnight. Dependence of mouse embryonic stem cells on threonine catabolism. *Science*, 325(5939):435–9, 2009.

[150] J. B. Kim, B. Greber, M. J. Arazo-Bravo, J. Meyer, K. I. Park, H. Zaehres, and H. R. Schler. Direct reprogramming of human neural stem cells by OCT4. *Nature*, 461:649–643, Oct 2009.

[151] D. Pei. Regulation of pluripotency and reprogramming by transcription factors. *J. Biol. Chem.*, 284:3365–3369, Feb 2009.

[152] T. E. North, W. Goessling, C. R. Walkley, C. Lengerke, K. R. Kopani, A. M. Lord, G. J. Weber, T. V. Bowman, I. H. Jang, T. Grosser, G. A. Fitzgerald, G. Q. Daley, S. H. Orkin, and L. I. Zon. Prostaglandin E2 regulates vertebrate haematopoietic stem cell homeostasis. *Nature*, 447:1007–1011, Jun 2007.

[153] M. Tsatmali, E. C. Walcott, and K. L. Crossin. Newborn neurons acquire high levels of reactive oxygen species and increased mitochondrial proteins upon differentiation from progenitors. *Brain Res.*, 1040:137–150, Apr 2005.

[154] M. Buggisch, B. Ateghang, C. Ruhe, C. Strobel, S. Lange, M. Wartenberg, and H. Sauer. Stimulation of ES-cell-derived cardiomyogenesis and neonatal cardiac cell proliferation by reactive oxygen species and NADPH oxidase. *J. Cell. Sci.*, 120:885–894, Mar 2007.

[155] J. Li, M. Stouffs, L. Serrander, B. Banfi, E. Bettiol, Y. Charnay, K. Steger, K. H. Krause, and M. E. Jaconi. The NADPH oxidase NOX4 drives cardiac differentiation: Role in regulating cardiac transcription factors and MAP kinase activation. *Mol. Biol. Cell*, 17:3978–3988, Sep 2006.

[156] Y. Yan, P. Sabharwal, M. Rao, and S. Sockanathan. The antioxidant enzyme Prdx1 controls neuronal differentiation by thiol-redox-dependent activation of GDE2. *Cell*, 138:1209–1221, Sep 2009.

[157] M. Diehn, R. W. Cho, N. A. Lobo, T. Kalisky, M. J. Dorie, A. N. Kulp, D. Qian, J. S. Lam, L. E. Ailles, M. Wong, B. Joshua, M. J. Kaplan, I. Wapnir, F. M. Dirbas, G. Somlo, C. Garberoglio, B. Paz, J. Shen, S. K. Lau, S. R. Quake, J. M. Brown, I. L. Weissman, and M. F. Clarke. Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature*, 458:780–783, Apr 2009.

[158] K. Ito, A. Hirao, F. Arai, S. Matsuoka, K. Takubo, I. Hamaguchi, K. Nomiyama, K. Hosokawa, K. Sakurada, N. Nakagata, Y. Ikeda, T. W. Mak, and T. Suda. Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature*, 431:997–1002, Oct 2004.

[159] S. J. Morrison, M. Csete, A. K. Groves, W. Melega, B. Wold, and D. J. Anderson. Culture in reduced levels of oxygen promotes clonogenic sympathoadrenal differentiation by isolated neural crest stem cells. *J. Neurosci.*, 20:7370–7376, Oct 2000.

[160] T. Ezashi, P. Das, and R. M. Roberts. Low O2 tensions and the prevention of differentiation of hES cells. *Proc. Natl. Acad. Sci. U.S.A.*, 102:4783–4788, Mar 2005.

[161] Q. Lin, Y. J. Lee, and Z. Yun. Differentiation arrest by hypoxia. *J. Biol. Chem.*, 281:30678–30683, Oct 2006.

[162] M. C. Simon and B. Keith. The role of oxygen availability in embryonic development and stem cell function. *Nat. Rev. Mol. Cell Biol.*, 9:285–296, Apr 2008.

[163] W. G. Kaelin and P. J. Ratcliffe. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol. Cell*, 30:393–402, May 2008.

[164] B. Keith and M. C. Simon. Hypoxia-inducible factors, stem cells, and cancer. *Cell*, 129:465–472, May 2007.

[165] S. Stolyar, S. Van Dien, K.L. Hillesland, N. Pinel, T.J. Lie, J.A. Leigh, and D.A. Stahl. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol*, 3:92, 2007.

[166] F. Li, I. Thiele, N. Jamshidi, and B. . Palsson. Identification of potential pathway mediation targets in Toll-like receptor signaling. *PLoS Comput. Biol.*, 5:e1000292, Feb 2009.

[167] I. Thiele, N. Jamshidi, R. M. T. Fleming, and B.O. Palsson. Genome-scale reconstruction of e. coli's transcriptional and translational machinery: A knowledge-base and its mathematical formulation. page (in review), 2008.

[168] J. Min Lee, E.P Gianchandani, J.A. Eddy, and J.A. Papin. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLos Comp Biol*, 4(5):e1000086, 2008.

[169] M. W. Covert, E. M. Knight, J. L. Reed, M.J. Herrgard, and B. Ø. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–96, 2004.

[170] E. P. Gianchandani, J. A. Papin, N. D. Price, A. R. Joyce, and B. Ø. Palsson. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol*, 2(8):e101, 2006.