Education is a progressive discovery of our own ignorance

- Will Durant -

University of Alberta

Using ^{1}H-NMR based metabolomics to investigate the pathological consequences of mitochondrial disease and human rabies infection.

by

Stacey Nichole Reinke

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

Spring 2012
Edmonton, Alberta

For Dad, Mom, and Ryan.

ABSTRACT

Mitochondrial diseases encompass a wide range of clinical phenotypes. The etiology of these disorders is extremely complex; mitochondria are central to energy metabolism and dysfunction can have a profound effect on global metabolism. The original objective of this thesis was to use ^{1}H-NMR based metabolomics to investigate the metabolic mechanisms of mitochondrial disease. Metabolomics is the systematic study of metabolites and is a powerful approach to understanding disease. In the clinical realm, metabolomics is useful for identifying biomarkers of disease and for understanding disease mechanisms. The metabolome is extremely sensitive to a number of factors, such as genetic background, age, diet, gender and stress; it is therefore difficult to meaningfully interpret results from human metabolomic data. For this reason, we chose to employ two common laboratory models of mitochondrial disease: the nematode *Caenorhabditis elegans* and the yeast *Saccharomyces cerevisiae*. These model systems provide the opportunity to study mitochondrial disease in a well-defined genetic background under controlled environmental conditions. Our findings revealed that even in a highly-controlled experiment, an appreciable amount of biological variability still exists. In the laboratory, *C. elegans* can be fed two different strains of *E. coli*: OP50 for regular maintenance and HT115 for RNAi-mediated gene suppression. We discovered that the nematode metabotype, mtDNA copy number, brood size and lifespan are significantly altered by diet. We also investigated the metabolic consequences of Complex II dysfunction in yeast, as reflected in the extracellular metabolome or exometabolome. Metabolomic studies are frequently used to identify a small number of key biomarkers that discriminate individuals of different phenotypes. Our exometabolome data

revealed that the entire metabolome contributes to discrimination between phenotypic classes. Amino acid metabolism, which is closely linked to energy metabolism, was profoundly affected by mitochondrial dysfunction in both model systems. I also investigated the pathogenic, molecular and metabolic consequences of a human rabies infection following an aggressive and controversial neuro-therapeutic treatment protocol, revealing a possible adverse and fatal consequence of this procedure.

EXAMINING COMMITTEE

Dr. Bernard Lemire, Department of Biochemistry

Dr. Brian Sykes, Department of Biochemistry

Dr. Vickie Baracos, Department of Oncology

Dr. David Broadhurst, Department of Medicine

Dr. Myrna Simpson, University of Toronto

ACKNOWLEDGMENTS

The past six years have been nothing less than an incredible journey. To everyone who I have met and to everyone who has helped me along in this journey, I sincerely thank you.

First, to my supervisors, Dr. Bernard Lemire and Dr. Brian Sykes, thank you for your guidance. Bernie: thank you for allowing me the freedom to pursue my own interests. Thank you for helping me develop my oral and written presentation skills. Brian: thank you for being a mentor and a friend. You helped me establish an expansive professional network of many great researchers; I am grateful for the opportunity to have worked with these people.

I would like to thank my additional supervisory committee member, Dr. Vickie Baracos, for providing guidance throughout my research. I would like to extend a huge thank you to Dr. David Broadhurst. Your patience while teaching me about data modelling and statistical analysis techniques is greatly appreciated. Thank you also for being an examiner on my thesis defence committee. I would also like to thank Dr. Myrna Simpson of the University of Toronto for being the external examiner for my thesis defence exam.

To the many past and present members of the Lemire and Sykes labs, thank you. Although there are too many people to name, a few names need special attention. I would like to offer a huge thank you to Dr. Sam Szeto. You offered your undying patience when I was a new graduate student. Thank you for enlisting my help on your thesis project and for working with me on our additional projects together. Most importantly, during this time you became a close friend. Above all, I thank you for your friendship. To Leanne Sayles, thank you for supervising me as an undergraduate summer student and showing me the ways of the lab. Thank you to Adrienne Baksh, Mia Hu, Dr. Natalia Volodko, and Dr. Maciej Behrendt for your friendship and help with research projects. I would also like to extend a huge thank you to all of the Lemire lab assistants who spent hours pouring plates, preparing solutions, stuffing pipette tip containers,

and autoclaving our garbage. These are necessary, but time consuming and undesirable, tasks. Having them done for us freed our time for research. From the Sykes lab, I would like to thank Dr. Olivier Julien and Dr. Ian Robertson for coming to my rescue when I had spectrometer difficulties. I would also like to thank Angela Thiessen, Dave Corson, Jeff Devries, Nic Shaw, Monica Li, and Robert Boyko for their help with technical assistance. Thank you to Dr. Ryan McKay and Deryck Webb from the National High Field Nuclear Magnetic Resonance Centre (NANUC) for their help when using NANUC instruments.

I would like to thank the researchers that I have collaborated with over the last six years: Dr. Magnus Friis and Dr. Michael Schultz, Dr. Simon Lamarre and Dr. Sean Brosnan, and Ingrid Catz and Dr. Ken Warrin. I would also like to extend a huge thank you to Dr. Chris Power, with whom I collaborated for the research presented in Chapter 5. Thank you to all of the authors that contributed to this study, in particular Will Branton and Ferdinand Maingat, of the Power lab, for preparing samples for me.

I would also like to thank the other department and university members who provided their support. Thank you to Dean Schieve for all of your help with anything computer-related. Thank you to Dr. Adrienne Wright, Dr. Jonathan Parrish, and Shannon Swan for your friendship and for giving me opportunities to expand my professional experience beyond research. Thank you to Dr. Neil Haave for giving me the opportunity to speak to your students about graduate school and about my research.

Life just wouldn't be the same without wonderful friends. Thank you to Dr. Lorissa Niebergall and Sarah Rudiger for your wonderful friendship. Regardless of whether I have needed to celebrate or just drink coffee in the presence of another person, you have both been there any time I have needed you. Thank you from the bottom of my heart.

To my parents, Daryl and Fern Reinke, thank you for your unconditional love and support. You have always believed in me. I would not have been able to achieve this level of education without you standing behind me (and partially financing me). Papa bear: you were solely responsible for lighting the flame that sparked my interest in all living things. I have such fond memories of looking at preserved creatures in jars in your lab, learning about the anatomy of the fish you caught, and looking at the “principals” (nuclei) of cells under the microscope. I apologize for asking continual questions as a child and I thank you for your patience and answering all of them. It has been a great pleasure to continually share my research with you. My little Fernie: thank you for being everything a mother should be. Thank you for telling me that I had to defend my thesis in 2011 so that all of your girlies could finish school in the same year. Thank you for caring about things like that.

To my sisters, Tammy, Melissa, and Christine Reinke, thank you for all of your sisterly love. Tammy: thank you for your support and interest in my studies. Thank you for feeding me when I was an undergrad. Melissa: thank you for all of the arguments we have had, for sharing what you have learned in chiropractic school, and for letting me share what I have learned with you. Christine: thank you for providing comic relief and for always being so liberal to highlight my short-comings. I am so happy that you finally found you’re interest in all living things. Thank you for helping me to see that I like animals and not just how they function.

Finally, to thank my husband, and best friend, Ryan Plett: thank you for everything. We have been on so many journeys together; thank you for standing beside me every step of the way on this one. I welcome whatever journeys the future may hold, because we will embark upon them together. May they all be wonderful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	one dimensional
ADP	adenosine diphosphate
AIDS	acquired immune deficiency syndrome
AIF	apoptosis initiating factor
ALS	amyotrophic lateral sclerosis
ANT	adenine nucleotide translocator
Apaf-1	apoptotic protease activating factor - 1
ATP	adenosine triphosphate
BCAA	branched chain amino acid
BG	basal ganglion
CD	cluster of differentiation
cDNA	complementary DNA
CNS	central nervous system
CSF	cerebrospinal fluid
CTX	cortex
CV	coefficient of variance
CV-ANOVA	analysis of variance testing of cross-validated predictive residuals
Cyt C	cytochrome C
DAB	3,3'-diaminobenzidine
DNA	deoxyribonucleic acid
DoE	design of experiment
dsRNA	double stranded RNA
DSS-d6	2,2-dimethyl-2-sila 3,3,4,4,5,5-hexadeutero-pentane sulphonic acid
EEG	electroencephalogram
FLAIR	fluid attenuated inversion recovery
FUdR	5-fluoro-2′-deoxyuridin
GABA	gamma-butyric acid
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
H&E	hematoxylin and eosin
H2O2	hydrogen peroxide
HIV	human immunodeficiency virus
HLA-DQA1	major histocompatability complex, class II, DQ alpha 1
IgG	immunoglobulin G
IMM	inner mitochondrial membrane
IPTG	isopropyl-ß-D-thiogalactopyranoside
IRIS	immune reconstitution syndrome
KYN	kynurenine
L1-4	larval stages 1 through 4
LDA	linear discriminant analysis
LDH	lactate dehydrogenase

LV	latent variable
mGPDH	mitochondrial glycerol phosphate dehydrogenase
MRC	mitochondrial respiratory chain
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
MS	mass spectrometry
MS	multiple sclerosis
mtDNA	mitochondrial DNA
mtPTP	mitochondrial permeability transition pore
NAA	N-acetylaspartate
NAD+/NADH	oxidized/reduced nicotinamide adenine dinucleotide
NADP+/NADPH	oxidized/reduced nicotinamide adenine dinucleotide phosphate
nDNA	nuclear DNA
neuroIRIS	neurological immune reconstitution syndrome
NIPALS	non-linear iterative partial least squares
NGM	nematode growth media
NMDA	N-methyl-D-aspartate
NMR	nuclear magnetic resonance
NR1	NMDA receptor subunit 1
OD	optical density
OXPHOS	oxidative phosphorylation
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
PDH	pyruvate dehydrogenase
PLS	projections to latent structures by means of partial least squares
PLS-DA	PLS-discriminate analysis
PPP	pentose phosphate pathway
Q	ubiquinone
Qp site	proximal quinone-binding site
RFC	relative fold change
RNA	ribonucleic acid
RNAi	RNA-intereference
ROS	reactive oxygen species
rRNA	ribosomal RNA
RSD	relative standard deviation
RT-PCR	reverse transcriptase PCR
RV	rabies virus
SDH	succinate dehydrogenase
SEM	standard error of the mean
T2	spin-spin relaxation time
TCA	trichloroacetic acid
TCA cycle	tricarboxylic acid (Kreb's) cycle

tRNA	transfer RNA
VIP	variable importance on the projection
WM	white matter

CHAPTER 1

General Introduction

Introduction

Historically, infectious disease has been the leading cause of death worldwide; epidemic infectious disease and inadequate nutrition remain major health challenges in developing nations. The advent of modern medicine (primarily the development of antibiotics), combined with improved hygiene practices, has substantially reduced death due to infectious disease in industrialized nations over the last century (Alberti 2001). Today, however, industrialized nations face a new endemic health challenge: metabolic disease (German et al. 2005). Common diseases including diabetes, atherosclerosis, and cancer have been linked to metabolic imbalance (German et al. 2005; Wallace 2005; Gross et al. 2010); our evolving understanding of the role that deranged metabolism plays in disease etiology has prompted us to view and treat modern health problems in a new light.

Metabolomics, in the clinical realm, is the systematic study of metabolism and the role it plays in disease; it is commonly applied to the identification of biomarkers of disease. Although the term metabolomics and its development as a field of study emerged around the turn of the millennium (Zhang et al. 2011), the concept of relating metabolic biomarkers to disease has been in practice for several decades. The glucose tolerance test, which measures the rate at which glucose is cleared from the blood stream following ingestion of a standard dose, has been employed as an indicator of diabetes for over sixty years (Schmidt 1950). The recent evolution of the field of metabolomics has sparked global interest in the wealth of information that is contained in the metabolome. In the past few years, specific biomarkers have been linked to a variety of diseases, including prostate

cancer, pneumonia, and early prediction of preeclampsia (Sreekumar et al. 2009; Slupsky et al. 2009; Kenny et al. 2010). This chapter introduces the concept of metabolomics and discusses various features of the metabolome. A discussion of the technological platform and statistical methods used in this thesis research is also included.

Metabolomic studies are also useful for probing the underlying molecular mechanisms of complex diseases. Mitochondria are the cellular organelles responsible for aerobic respiration; energy metabolism is central to, and highly integrated with global cellular metabolism. Mitochondrial dysfunction causes a wide variety and severity of clinical phenotypes. This chapter also describes mitochondria and the complex nature of mitochondrial disease. Finally, due to the great complexity and variability of the metabolome, in particular the human metabolome, it is often difficult to make meaningful interpretations from metabolic data. Model systems offer greater control of both genetic and environmental factors that affect the metabolome and are valuable for developing principles of metabolism that may apply to more complex systems. Two model systems of mitochondrial disease and their advantages for use in metabolomic studies are also introduced.

Metabolomics

All biological systems contain four functional components: the genome, the transcriptome, the proteome, and the metabolome (Figure 1-1) (Goodacre 2007). The transcriptome reflects the expression of the genome; the proteome reflects translation of the genome, including post-translational modifications. The metabolome consists of the entire set of metabolites found within a given

biological system; metabolites are the (bio)chemicals that are reactants, intermediates or products of enzyme-mediated reactions. As the four functional biological components reflect each other's functional status, a great deal of information can be extracted from their combined study (systems biology). However, the metabolome is considered to be the functional level most directly associated with phenotype, as the metabolome is the ultimate end product of gene expression. Any stimulation or suppression of gene expression and protein translation will respectively increase or decrease the levels of the metabolite acted upon by that protein. Similarly, gene mutations or mis-expression of proteins that act as signalling molecules can also affect the metabolic state. Many diseases are in fact the result of deranged metabolism, being caused by enzyme dysfunction, nutritional imbalance, or inadequate metabolic regulation (German et al. 2005). Thus, the metabolome offers a great deal of information that can be useful for the diagnosis, prognosis, and understanding of clinical phenotypes.

Metabolomes and metabolites

The term metabolome can describe the chemical signature of a specific tissue, organ system or of the entire organism depending on context. Both intracellular (endo-) and extracellular (exo-) metabolomes exist (Kell et al. 2005; Dunn et al. 2005); they are studied as separate but complementary entities. Whereas characterizing the endometabolome (metabolic fingerprinting) offers a snapshot of active metabolism, studying the exometabolome (metabolic footprinting) provides information about the consumption or expulsion of nutrients. Exometabolomic studies are often used to gain insight about the intracellular

metabolic state when adequate cellular sample size is difficult to obtain (such as in cell culture). When working with animal models, it is relatively easy to obtain tissue samples of sufficient quantity for metabolomic analyses; it is often extremely difficult to obtain the same from human patients. In the clinical setting, metabolic footprinting of biofluids is routinely used for diagnostic purposes; this can range from the completely non-invasive collection of urine to the highly invasive collection of cerebrospinal fluid (CSF). Both metabolomic approaches can be extremely useful depending on available resources and the hypothesis.

Metabolites are the low molecular weight molecules that are involved in biochemical reactions (as reactants, intermediates, or products) (Dunn et al. 2005). Depending on the metabolite, the tissue location and its subcellular localization, metabolite concentrations range from pM to mM (Wishart et al. 2009). Although metabolites are generally considered to be organic molecules, inorganic molecules, such as metals also play an important role in metabolism. Some metal ions are coordinated by enzymes (such as the mitochondrial respiratory chain complexes) and also some metabolites (such as ATP) (Mounicou et al. 2009). Although the definition of a metabolite seems fairly straightforward, the exact divisions between metabolite and peptide, transcript or DNA can be ambiguous (Dunn et al. 2011). For example, glutathione is a tripeptide that functions as an important metabolite for protecting against oxidative stress. Similarly, nucleotides compose both DNA (and RNA) and energy metabolites (such as NADH and ATP). To date, almost eight thousand metabolites have been identified in the human body. These include endogenous metabolites, drugs, toxins and environmental pollutants, and food components (Wishart et al. 2009).

Factors affecting metabolic state

The metabolome is extremely sensitive and highly dynamic. Whereas the transcriptome and the proteome respond within minutes or hours to a particular stress, the metabolome may respond within seconds and is an immediate indicator of stimuli. Many factors affect the status of endogenous metabolites: genetic background, diet, exercise, age, gender, stress, the diurnal cycle, and the oestrus cycle (Bollard et al. 2001; Solanky et al. 2003; Holmes et al. 2008). Endogenous metabolism is also affected by the presence of exogenous metabolites (xenobiotics) including drugs, food additives, plant secondary metabolites (such as caffeine), and pesticides (Holmes et al. 2008). Xenobiotics are metabolized by the body and can also contribute to metabolic variation (Dunn et al. 2011; Holmes et al. 2008). Lastly, the microbiome also impacts metabolism; the microbiome includes all microorganisms inhabiting the human body (Goodacre 2007). Although microbiomes exist throughout the human body, the best characterized is the gut microbiome. Gut bacteria can metabolize host food prior to nutrient absorption across the intestine; some of these products play roles in diseases such as obesity, intestinal disease, cancer, and cardiovascular disease (Holmes et al. 2008; Wang et al. 2011). The interactions between endogenous metabolites, xenobiotics and the microbiome create a highly complex metabolic network. As all of these factors are highly specific to each individual, an appreciable amount of inter-individual (and intra-individual) metabolic heterogeneity exists within healthy populations (Saude 2007; Zuppi et al. 1998).

Applications

Metabolomic studies often focus on biomarker discovery and understanding disease etiology; however, the field is applicable to a wide variety of interests. Metabolomics has begun to play a popular role in the food industry. The chemical composition of various wines has been identified in order to discriminate molecules that are associated with characteristics such as vintage, origin and quality (Cuadros-Inostroza et al. 2010; Ali et al. 2011). These studies are useful for wine makers to optimize grape varieties and yeast strains used, and storage conditions. Dutch researchers have gone one step further, studying the metabolic conversion of dietary polyphenols, first by gut microorganisms and then by the host (van Dorsten et al. 2010; van Duynhoven et al. 2011). In the horticulture industry, metabolomics has been used to understand disease resistance (Bollina et al. 2011) and stress responses (Shulaev et al. 2008) to optimize breeding (Davies et al. 2010). Environmental scientists use metabolomics to assess soil and water toxicity (Simpson and McKelvie 2009; Viant et al. 2006). These are a just a few examples of the breadth of possible applications of metabolomics.

Instrumentation

A variety of instruments and techniques can be used for metabolite quantification. Each offers unique advantages and metabolite specificity. The two most commonly used platforms are NMR spectroscopy and mass spectrometry (MS). However, various chromatographic techniques, Fourier-transform infrared and Raman spectroscopies and assay kits can also be used for the similar purposes (Fancy et al. 2006; Ellis and Goodacre 2006). For this thesis research, I chose to

use NMR spectroscopy for the advantages listed below as well as the availability of local facilities and expertise.

NMR spectroscopy as a metabolomics platform

NMR spectroscopy offers a number of unique and invaluable advantages when applied to metabolomic studies; these have been discussed in detail in a number of reviews (Lindon et al. 2003; Dunn et al. 2005; Dunn et al. 2011). First, NMR is commonly used to detect protons; the ubiquity of protons in metabolites means that many can be detected simultaneously in one sample. Additionally, other common nuclei, such as ^{13}C, ^{15}N, and ^{31}P are observable via NMR, yielding opportunities for metabolic tracer studies. Specifically labelled metabolites can be introduced and followed along biochemical pathways. Although both NMR and MS can establish the degree of labelling of nuclei in a given molecule, NMR is uniquely able to distinguish which atoms are labelled; this is highly advantageous for flux studies. Second, NMR is quantitative and highly reproducible. Each metabolite produces a highly-specific peak pattern, based on its composition and coupling constants. Targeted profiling, quantitatively characterizing NMR spectra, is easily achieved using databases that contain peak chemical shift and J-coupling information. Chenomx NMR Suite Professional (Chenomx Inc., Edmonton, Canada) is one of the most widely used databases for characterizing NMR spectra. Based on information stored in the database, Chenomx creates mathematical models of unknown metabolites; internal standards are used to quantify metabolites (Weljie et al. 2006). Contributing to the reproducibility of NMR, unlike in MS, is the fact that samples do not physically interact with the NMR spectrometer. When using MS, quality control samples must be periodically

analyzed (Sangster et al. 2006; Zelena et al. 2009); this is unnecessary to maintain reproducibility with NMR spectroscopy. Third, almost any biofluid and tissue can be analyzed with minimal preparation. Although removal of large molecular weight molecules is often an included step (discussed in next paragraph), sample preparation involves adding a chemical shift indicator (also acts as concentration reference) and sometimes a pH buffer. Fourth, as samples remain intact during acquisition, NMR spectroscopy is non-destructive. This offers the potential to repeat the analysis of any given sample using another platform.

Despite its many advantages, NMR fails to compare to MS in sensitivity. Whereas MS can detect metabolites in the nanomolar concentration range, micromolar concentrations must be present to be observed with NMR. Thus, in samples where NMR can detect tens of metabolites, MS can detect hundreds or even thousands (Dunn et al. 2011). Another more manageable disadvantage of NMR is interference of large molecular weight molecules, such as proteins and lipids. These molecules create broad peaks in NMR spectra (Figure 1-2), often overlapping and reducing resolution of peaks of interest (Dunn et al. 2011). Also, those metabolites that remain enzyme bound are not detectable. Thus, sample preparation often involves removal of these large molecules. Removal of large molecular weight molecules produces spectra with flat baselines and easily identifiable metabolites, represented by narrow peaks (Figure 1-2). Although filtration is a commonly used to preserve sample integrity, extractions produce clean samples and release enzyme-bound metabolites.

Data Analysis

Metabolomic studies generate a substantial amount of data; as with any study, when looking for changes in metabolite concentrations, statistical significance of the data must be established to avoid making false discoveries. Univariate methods (Student's t-test, ANOVA etc.) can be useful in establishing significant differences between individual metabolites of different groups. However, univariate methods treat variables independently, making it is difficult to determine correlations between variables; these correlations are important in establishing a clear understanding of the underlying biological processes of a given experiment. Often it is the combination of many metabolites that can uncover clear biological differences between two (or more) biological phenotypes. These combinations are often called metabolic signatures or metabolic profiles. For this reason, it is standard procedure to apply multivariate analytical methods to metabolomic studies.

The projection approach

There are several methods for modelling metabolomic data; however, the projection approach is one of the most widely used methods by the metabolomics community. Prior to understanding the projection approach, one must first understand the basic structure of a data matrix (Figure 1-3). Every data matrix can be represented by *n* observations (samples) and *k* variables (metabolites); this is referred to as the X-matrix (Goodacre et al. 2004). Y-variables represent responses (phenotype) and will be discussed later in this section. According to the projection approach, if each variable occupied one dimensional space, then the data matrix of *k* variables would represent a *k*-dimensional space. In order to

understand and draw meaningful conclusions from the data, the projection approach projects the *k*-dimensional space into a lower-dimensional space that can be visualized. This is done in a manner that minimal information is lost and meaningful interpretation of the data can be readily achieved. A simple example of the projection approach plots 10 observation points in a three-dimensional space, representing three variables x_1, x_2, and x_3 (Figure 1-4). The projection approach mathematically projects the data onto lower dimensional hyper-planes or latent variables (Eriksson et al. 2001). These latent variables express the data in a linear combination of the original vectors (Shlens 2009). In this example, the data are represented by two latent variables; thus, the three-dimensional data are reduced and visualized in two dimensions. The same concept can be applied to many observations in a multi-dimensional space.

Principal component analysis

Principal component analysis (PCA) is one example of how the projection approach is applied to large data sets. In the above example, the latent variables are referred to as principal components (PCs); the first principal component is the line that best fits the data points in the least squares sense (standard deviation from line to observation points is minimized) (Jewell 2000), and therefore explains the direction of maximum variance in the multidimensional space. The second and subsequent PCs represent the next mathematically orthogonal direction of maximum variance that is not modelled in the former components (Otto 1999). In this sense, PCA can be considered as a rotation of the original *k*-dimensional axes onto a new variance dependent set of *p*-dimensional axes. The number of effective PC dimensions is determined by the amount of

cumulative explained variance (Otto 1999). PCA is an unsupervised approach; that is, each metabolite variable is treated equally without *a priori* knowledge of responses (phenotypes). Thus, any correlation between biological samples is based purely on the 'natural' multivariate variance observed in the measured biochemistry. Often this variance is sufficient to see clusters of biochemically similar species; the associated model 'loadings', which reflect the importance of each metabolite in each of the projected PC dimensions, can be interpreted to give biological meaning to the correlations.

PCA approximates the data in the X-matrix using the product of two smaller matrices: the score (T) matrix and the loading (P) matrix (Figure 1-5). The T scores are the coordinate points of each observation along the PC. The loadings (P) describe the magnitude and the manner (positive or negative correlation) in which the variables contribute to the scores (Eriksson et al. 2001). Observations that differ greatly in their variables will lie in opposite directions along the principal component. Geometrically, the X-matrix consists of *n* rows (observations) and *k* columns (variables); the T-matrix consists of *n* rows and *d* columns (number of PCs); the P-matrix consists of *d* rows and *k* columns (Figure 1-5). Mathematically, PCA deconstructs the X-matrix based on the following linear relationship:

$$X = TP^T + E$$

where E represents the matrix of unexplained variance by the model (Otto 1999). Theoretically, in order to calculate the principal component scores, singular value decomposition of X^TX (the correlation matrix of X) is performed (Shlens 2009); however, in practice for large data sets, an iterative process known as *non-linear*

iterative partial least squares (NIPALS) is generally used (Jewell 2000). A full derivation of this algorithm is beyond the scope of this thesis. The result is a PCA scores model:

$$T^* = XP$$

$$PC_1: \quad \mathbf{t^*_1} = p_{1,1}\boldsymbol{x_1} + p_{1,2}\boldsymbol{x_2} + \ldots + p_{1,n}\boldsymbol{x_n}$$

$$PC_2: \quad \mathbf{t^*_2} = p_{2,1}\boldsymbol{x_1} + p_{2,2}\boldsymbol{x_2} + \ldots + p_{2,n}\boldsymbol{x_n}$$

Thus, for each PC score vector $\mathbf{t}_i^*$, there is an associated loadings vector $\mathbf{p}_i'$, which indicates the importance of each variable, $\mathbf{x}_{1\ldots n}$, on that component.

Projections to latent structures by means of partial least squares (PLS)

PCA is the most basic multivariate projection approach. As discussed above, it is optimized to describe the 'natural' variance in the data. Often this projection is not sufficient to examine less obvious differences in groups of biological samples. In these cases, a projection method optimized for group (or class) separation is more explanatory. An alternative to PCA is a type of projection method known as Linear Discriminant Analysis (LDA). There are many available LDA (Broadhurst and Kell 2006); however, they are all based on the same principle – to optimize the projection based on minimizing within-class variance whilst also maximising between-class variance (Figure 1-6). Thus, geometrically speaking, the difference between PCA and LDA is that PCA is a rotation of the original x-axis onto a new set of axes describing descending directions of maximum *variance* whereas, LDA is a rotation onto a new set of axes describing descending directions of maximum *discrimination*.

The most common and effective LDA algorithm used by the metabolomics community is projection to latent structures by means of partial least squares regression (PLS-DA) (Eriksson et al. 2001). This process uses an iterative algorithm (similar to NIPALS) to rotate and project the X-matrix into a lower dimensional set of latent variables, P, (P-dimensional space) using an additional regressor matrix, Y, containing the binary encoded class membership information. In the simplest cases of a two class model, samples of 'Class 1' will be encoded as a '1' and samples of 'Class 2' will be encoded as a '0' in the first Y-variable; the second Y-variable would encode the reverse (Figure 1-7A). Thus, mathematically the algorithm creates latent variables based on maximum variance between the X and Y variables. In other words, a PLS model will try and find the multidimensional direction in X-space that explains the maximum multidimensional variance in the Y-space (Figure 1-7B).

Similar to PCA, PLS-DA also decomposes the X-variable matrix into smaller score, loadings and error matrices (Figure 1-8):

$$X = TP^T + E.$$

However, PLS also decomposes the Y matrix thus:

$$Y = TQ^T + F$$

where T represents the latent projection scores, Q represents the Y-loadings, P represents the X-loadings and E & F represents the unexplained variance in the model in both the X and Y space. Where PCA is theoretically based on the singular value decomposition of X^TX, PLS is based on the singular value decomposition of X^TY (the cross correlation matrix). Again, full derivation of the

optimization algorithm is beyond the scope of this thesis (Wold et al. 2001). The resulting PLS-DA scores matrix for the projection of X onto latent space, T*, (similarly explained as the PCA example) is:

$$T^* = XP$$

And thus the prediction of classification, Y*, can be expressed as:

$$Y^* = XPQ^T$$

Or,

$$Y^* = XB$$

where B represents the regression coefficients mapping X directly onto Y (Otto 1999; Jewell 2000).

Although PLS-DA is commonly used for metabolomic data modelling, it does have limitations. First, PLS-DA only works well if the number of classes does not exceed six (Eriksson et al. 2001). Second, class observations must be “tight”, occupying a small portion of the X-space (Eriksson et al. 2001). However, metabolic data are often extremely heterogeneous; it is very easy to over-optimize PLS models built using heterogeneous data, where discrimination is based on erroneous variation in the data (statistical noise) but which bears little relation to ‘real’ biological deviation (Wold et al. 2001). Third, caution must be exercised to only use the number of latent variables needed to explain biological variation. By calculating too many latent variables (over-fitting), the model will fit on noise, thus reducing its predictive power (Figure 1-9) (Eriksson et al. 2001). Broadhurst and Kell (2006) have discussed this limitation and offer suggestions

on how to avoid over-fitting; a discussion of model validity in the context of this thesis will be presented in a later section. Fourth, PLS-DA (and PCA) assumes data orthogonality, that is variables are considered to be entirely independent of each other (Eriksson et al. 2001). Thus, PLS-DA cannot correlate the mathematical model to biological relevance. Despite these limitations, PLS-DA offers a means of visualizing and discriminating data and establishing quantitative relationships among observations and variables.

Data pre-processing

As metabolite concentrations can range over four orders of magnitude (Wishart et al. 2007; Wishart et al. 2009) (in this thesis, metabolite concentrations ranged over two orders of magnitude), the metric differs substantially between variable columns. In multivariate methods, such as PCA and PLS-DA, large differences in both absolute concentrations and total range can distort the model (Otto 1999). Scaling the data by multiplying it by a vector to give comparable variance to all variable columns eliminates both types of distortion (Otto 1999; Eriksson et al. 2001). The most common method of scaling is to unit variance (i.e. dividing each data point in a given variable by the total variance of that variable). In addition to scaling, data are also translated along the coordinate origin to eliminate a constant offset; this is referred to as mean centering. To achieve the mean center, the mean value of each variable is subtracted from the data. These two steps ensure that all variables are treated with *a priori* equal importance (Figure 1-10).

Model validation and significance

A valid model predicts much better than chance. Multivariate models in this thesis were generated using the Simca-P+ (v12.0.1) software package (Umetrics, Umeå, Sweden); this software assesses PLS-DA model performance in two ways (Eriksson et al. 2001). First, the model generates R^2 (goodness of fit; explained variance) and Q^2 (goodness of prediction). Normal biological systems contain an appreciable amount of variability and therefore, interpretation of these values is often subjective. Although a Q^2 value of greater than 0.9 would be considered excellent by any standards, a Q^2 value greater than 0.5 can be considered good for biological data. To ensure the model is not fitting on noise (Figure 1-9), the difference between R^2 and Q^2 must not be greater than 0.2. Second, validation plots can be generated to determine the risk of the model being spurious. A series of random permutations of the model are performed, where the Y-observations are randomly generated while the X-matrix is left intact. Simca-P+ (v12.0.1) generates validation plots as seen in Figure 1-11; validation plots must be generated for each class within the model. The horizontal axis of the plot shows the correlation between the original and permutated y-variables. The vertical axis displays the R^2 and Q^2 values for each permutation. The original y-variable is most correlated with itself, defining the highest theoretical point on the plot (top right-hand corner). All R^2 and Q^2 values from randomly-generated permutations are shown to the left of the original; R^2 and Q^2 values are represented by green triangles and blue squares, respectively. Two aspects are used to indicate model validity, although these are somewhat subject to interpretation. The first is that none of the randomly generated permutations, either R^2 or Q^2, out-perform the original (none of the points on the left are higher

than the originals on the far right). The second is that the y-intercepts are below the acceptable cut-offs of 0.4 and 0.05 for the R^2 and Q^2 regression lines, respectively. Figure 1-11A shows a validation plot derived from a model with valid data; Figure 1-11B represents a validation plot for a model that is not valid. To determine statistical significance between classes, a CV-ANOVA (Analysis of variance testing of cross-validated predictive residues) test generates a *P*-value for the model.

Data interpretation

Projection models decompose data into scores and loadings values; both can be presented in plots to interpret the data. These plots can be one-, two-, or three-dimensional and present any combination of the model's latent variables. As the first one or two latent variables explain most of the variance, they are most often presented. Figure 1-12 presents an example of a score and of a loading plot. An observation's score value represents the point at which it is located along the latent variable. Hence, observations that group close together on a score plot are most similar to each other; those that are most different will lie on opposite ends of the latent variable. Score plots provide an overview of the data, as well as possible outliers. Outliers are defined as observations that do not fit the model well; their existence may cause the model to find erroneous latent variables (Eriksson et al. 2001). Loadings plots provide information about the correlation between variables; cosines of the angles between loading vectors indicate correlation (Otto 1999). Small angles indicate variables that are highly correlated to each other in the establishment of the model. Uncorrelated variables are orthogonal to each other. The size of the loading vector (the coordinate value)

indicates its importance on the establishment of the model. Loadings near the origin of the coordinate axis are unimportant to the establishment of the model (Otto 1999). Loadings can be presented either independently, or if properly scaled can be superimposed onto the score plot. This plot is called a biplot; software programs such as Simca P+ (v12.0.1) automatically present loadings plots in this way. Additional interpretation methods will be described as they are presented throughout this thesis.

Mitochondria

Mitochondria are the cellular organelles responsible for generating energy in the form of ATP. They convert the chemical energy of nutrients into energy that can be used by the cell (Wallace 2010). The number of mitochondria in a cell can range from a few hundred to hundreds of thousands, depending on cell type and energy requirements (Scheffler 1999).

Physical characteristics

Mitochondria are very dynamic but often appear tubular in shape. Typical mitochondria, in cells such as hepatocytes, are approximately 3-4 μm in length and 1 μm in diameter. Mitochondria have four distinct compartments: the outer membrane, the intermembrane space, the inner membrane, and the matrix (Figure 1-13). Convolutions of the inner membrane (cristae) maximize its surface area for energy production (Scheffler 1999).

Mitochondrial DNA (mtDNA)

Circular mitochondrial DNA (mtDNA) resides in the mitochondrial matrix (Wallace 1982). In humans, mtDNA encodes 13 polypeptides that are subunits of the

mitochondrial respiratory chain complexes. Complex II is the only complex to be entirely encoded by nuclear DNA; Complexes I, III, IV, and V have subunits encoded by both nuclear and mitochondrial DNA. mtDNA also encodes the 12S and 16S rRNAs and 22 tRNAs required for expression of its own genes (Wallace 2010).

Mitochondrial Respiratory Chain and Oxidative Phosphorylation

The mitochondrial respiratory chain (MRC) utilizes the energy stored in metabolic reducing equivalents (NADH and succinate) to generate ATP. This is achieved by coupling the transfer of electrons and subsequent reduction of molecular oxygen to the vectorial movement of protons across the inner membrane, which re-enter the matrix through ATP synthase, thus generating ATP. This process is known as oxidative phosphorylation (OXPHOS).

The MRC consists of five complexes (Figure 1-14). Complex I, NADH-ubiquinone oxidoreductase transfers two electrons from NADH to ubiquinone while pumping four protons into the intermembrane space. Complex II, succinate dehydrogenase is both a MRC enzyme and a Kreb's cycle enzyme. It oxidizes succinate to fumarate, transferring two electrons to ubiquinone. Complex II is the only electron-transferring MRC complex that does not pump protons into the intermembrane space. Ubiquinone is a lipid soluble molecule; the reduced ubiquinol moves through the inner membrane to Complex III (ubiquinone-cytochrome *c* oxidoreductase), where the electrons are transferred to cytochrome *c*. Complex III also releases four protons into the intermembrane space. Cytochrome *c*, carries one electron at a time to Complex IV (cytochrome *c* oxidase), which reduces molecular oxygen to water. For every one-half molecule

of oxygen reduced to water, two protons are pumped to the intermembrane space. The vectorial transport of protons to the intermembrane space creates a proton motive force, establishing both a pH gradient and an electrical potential across the inner membrane. The inner membrane is impermeable to protons and under normal conditions the only way they can re-enter the matrix is through Complex V (ATP synthase), thus generating ATP (Garrett and Grisham 2010).

Mitochondrial Disease

Despite the primary role of the MRC being maintenance of cellular energy and redox levels, it also influences a vast array of metabolic and cellular processes (Wallace 1999; Scheffler 1999, 2000). The term mitochondrial disease originally described clinical phenotypes arising from OXPHOS dysfunction (Smeitink et al. 2001); presently the term includes dysfunction of any mitochondrial enzyme or RNA molecule (UMDF 2011c). Diagnosing mitochondrial dysfunction can be extremely difficult as the clinical presentations are numerous. Mitochondrial diseases are highly variable in age of onset, severity and tissue location. Several factors contribute to disease pathology and variability, making treatment and prognoses extremely difficult (Smeitink et al. 2001).

Genetic factors

The MRC is composed of about 85 proteins. With the exception of complex II, which is solely nuclear-encoded, all MRC complexes contain both nuclear-encoded and mitochondrial-encoded subunits. Disease phenotypes can arise from mutations to either genome. Mitochondrial genetics differ from Mendelian genetics, which govern the nuclear genome, in four distinct ways (Smeitink et al. 2001). First, mtDNA is exclusively maternally inherited. Second, mtDNA exists in

polyplasmy, having many copies in each cell. Each human mitochondrion contains between two and ten copies of DNA, with each cell containing between one thousand and ten thousand copies. Third, mtDNA can be heteroplasmic, a condition where both wild-type and mutated mtDNA co-exist. During mitosis, mitochondria are randomly distributed between daughter cells, resulting in varying degrees of both polyplasmy and heteroplasmy. Heteroplasmy plays a key role in the fourth difference between mitochondrial and Mendelian genetics, the threshold effect. For mtDNA mutations to exhibit a phenotypic effect, they must result in a critical biochemical effect (Rossignol et al. 2003). A pathogenic threshold can be achieved at any of the following levels: the percentage of wild type mtDNA in a heteroplasmic situation, the transcription of the mtDNA, the translation of the mRNAs, the function of the proteins or in other biochemical processes. Mutations in tRNA genes can destabilize the tertiary tRNA structure, thereby increasing their susceptibility to hydrolysis and aggravating the effects of heteroplasmy (Chomyn et al. 2000). Enzyme function can be affected either by changing the number of properly assembled enzyme complexes or the intrinsic kinetic properties of the enzymes. Ultimately, a phenotypic change only arises when any of the above disturbances affects mitochondrial energy metabolism or ATP production (Rossignol et al. 2003). Whereas nuclear-derived mitochondrial disorders are often apparent early in life, mtDNA mutations often cause delayed symptom onset and are progressive (Wallace 1999).

Reactive oxygen species (ROS) and apoptosis

Mitochondria are a key site for both ROS production and the initiation of apoptosis. Imbalances to either of these processes can lead to disease

phenotypes. During electron transfer through the MRC, some electrons escape and react with molecular oxygen, forming the superoxide anion, O_2^-. During normal respiration, up to 2% of molecular oxygen consumed is converted to O_2^- (Orrenius et al. 2011). During periods of caloric abundance and low cellular demand for ATP, the ATP synthase slows while electrons are continuously pumped across the IMM and the membrane becomes hyperpolarized. Proton pumping continues until the electrochemical gradient inhibits it and electron transfer through the MRC. Electron carriers are maximally reduced and can donate their electrons to molecular oxygen, thereby increasing ROS production (Wallace 2005). Conversely, maintaining adequate levels of physical activity, while preventing excess caloric intake prevents ROS production.

Once produced, the superoxide anion is converted via superoxide dismutase to the relatively stable hydrogen peroxide (H_2O_2) molecule. Hydrogen peroxide can be converted to water via glutathione peroxidase. Hydrogen peroxide can also diffuse out of the mitochondrion into the cytosol. H_2O_2 causes protein damage by creating carbonyl groups and oxidizing thiols; it also oxidizes polyunsaturated fatty acids, damaging phospholipids (Wallace 2005). Both the superoxide anion and hydrogen peroxide can interact with Fe^{2+} or Cu^{2+} via Haber-Weiss and Fenton reactions, thereby generating the highly reactive and toxic $^{\bullet}OH$ radical (Toninello et al. 2004). Superoxide anion-mediated oxidation of Fe-S proteins results in both enzyme inactivation and simultaneous release of Fe^{2+} and H_2O_2; this concomitant release of Fenton reaction substrates amplifies the production of the $^{\bullet}OH$ radical and cellular oxidative damage (Fridovich 1997). The $^{\bullet}OH$ radical can damage DNA, turning nuclear proto-oncogenes into functional oncogenes. Cytosolic H_2O_2 activates signalling cascades, promoting transcription and

replication which may favor tumour generation. Mitochondrial DNA is particularly vulnerable to oxidative damage for two key reasons (Orrenius et al. 2011). First, mtDNA is in close proximity to the MRC and Fe-S containing enzymes. Second, unlike nDNA, mtDNA lacks protective histones. Oxidation accounts for 10-20 times higher levels of modified bases in mtDNA than in nDNA and may contribute significantly to genomic instability that causes mitochondrial dysfunction.

In addition to causing cellular damage and promoting tumour formation, mitochondria are also integrally involved in regulating apoptosis. Mitochondria-induced apoptosis is mediated by the mitochondrial permeability transition pore (mtPTP) (Wallace 2005). The inner membrane adenine nucleotide translocator (ANT), the outer membrane porin, along with Bax, Bcl2 and cyclophilin D are thought to compose the mtPTP. Opening of the mtPTP is stimulated by increased mitochondrial calcium, oxidative stress or by decreases in the electrochemical gradient, in ADP or in ATP. Upon mtPTP opening, the electrochemical gradient collapses completely as ions equilibrate across the mitochondrial membranes. As a result, mitochondrial contents move freely into the cytosol; death-promoting proteins such as cytochrome *c*, apoptosis initiating factor (AIF), latent caspases, and endonuclease G are among the mitochondrial contents. AIF and endonuclease G degrade chromatin in the nucleus. Cytochrome *c* activates cytosolic apoptotic protease activating factor-1 (Apaf-1); Apaf-1 cleaves procaspase-9. Activated caspase-9 initiates cytosolic proteolysis. Normally, cytochrome *c* is bound to the inner mitochondrial membrane by electrostatic, hydrophobic and hydrogen-bonding interactions to cardiolipin (Orrenius et al. 2011). Cardiolipin oxidation by ROS releases cytochrome *c* during apoptosis (Ott et al. 2002; Kagan et al. 2005).

Increased ROS production stimulated by OXPHOS dysfunction can either initiate tumour promotion or apoptosis. Several cancers, including renal cell carcinoma, pheochromocytoma, paraganglioma and prostate cancer are strongly associated with ROS production (Wallace 2005). Increased ROS production and apoptotic activity are thought to play an important role in neurodegenerative diseases, such as Alzheimer's disease, amyotrophic lateral sclerosis (ALS), Parkinson's disease and Huntington's disease (Wallace 2005).

Redox balance and metabolism

The MRC drives proton transport across the IMM using electrons from NADH and succinate. In as many as 299 biochemical reactions, NAD^+ accepts electrons during fuel oxidation (Wishart et al. 2007; Wishart et al. 2009). Succinate dehydrogenase (SDH) functions in the Kreb's cycle and the MRC, converting succinate to fumarate and transferring two electrons to ubiquinone. MRC deficiency often causes an accumulation of NADH and succinate. NADH accumulation inhibits a number of enzymes, including pyruvate dehydrogenase (PDH), citrate synthase, isocitrate dehydrogenase, and α-ketoglutarate dehydrogenase of the Kreb's cycle (Garrett and Grisham 2010). An increased $NADH/NAD^+$ ratio would also dictate the inhibition of many other unregulated redox reactions by means of decreased substrate availability. One common hallmark feature of mitochondrial disease is lactic acidosis. Lactate dehydrogenase reduces pyruvate to lactate while oxidizing NADH; blood lactate/pyruvate ratios are examined in patients suspected of having mitochondrial disease (UMDF 2011a). Thus, MRC dysfunction can lead to

cellular redox imbalance, further reducing energy production while affecting many biochemical processes.

Energy metabolism is central to global metabolism. During periods of excess fuel intake, macromolecules that provide energy storage are synthesized. During fasting, these molecules are catabolised to provide the ATP that is required to maintain cellular/organismal function. Central to energy metabolism is carbohydrate metabolism. Carbohydrates are oxidized via glycolysis to provide substrates for fatty acid synthesis and the Kreb's cycle; glucose can also be oxidized via the pentose phosphate pathway (PPP), providing NADPH for reductive biosynthesis and ribose-5-phosphate for nucleic acid synthesis. Glycolysis and the Kreb's cycle provide substrates for amino acid synthesis, and conversely, amino acid carbon catabolic products can enter glycolysis and the Kreb's cycle (Garrett and Grisham 2010). As carbon and energy metabolism are so closely linked to the metabolism of all macromolecule monomers, any disturbance in the MRC can potentially affect numerous metabolic reactions.

Tissue specificity

The clinical presentations of mitochondrial disease are numerous and extremely heterogeneous. As tissues vary greatly in their energy requirements, the organ systems that are most sensitive to mitochondrial dysfunction are also those that require the most energy. These organ systems, in decreasing order of energy requirement are: the central nervous system, skeletal and cardiac muscle, the renal system, the endocrine system, and the liver (Wallace et al. 2007). Almost any organ system can be affected independently; however multiple organ systems and or systemic phenotypes often present. It is recommended to

clinicians that mitochondrial disease be suspected when three or more organ systems are affected (UMDF 2011b).

Model Systems

As the metabolome is extremely sensitive to many pressures (discussed in Chapter 2), performing meaningful metabolomic studies using human specimens can be extremely difficult. Using model systems in metabolomic experiments offers the advantage of being able to better control genetic background and environmental conditions, therefore reducing metabolic variability. Two model systems were used for this thesis research: the nematode *Caenorhabditis elegans* and the baking yeast *Saccharomyces cerevisiae*.

Caenorhabditis elegans

C. elegans are free-living soil nematodes measuring approximately 1 mm in length as adults. Being transparent, their digestive and reproductive structures are easily identifiable under a dissecting microscope (Figure 1-15). From eggs, these nematodes hatch and develop into adults through four larval stages, each separated by a cuticle moult (Figure 1-16) (Wood 1988). If undesirable growth and reproductive conditions are sensed during larval stage 1 (L1), the nematode will enter an alternative larval stage, the dauer at the second moult. Dauers are radially constricted with an internal oral orifice plug preventing food intake (Cassada and Russell 1975; Hu 2007). Dauers are able to survive up to 24 weeks without eating; upon restoration of favourable environmental conditions, the dauers enter larval stage 4 (L4) within 12 hours (Riddle 1988). Under desirable growth conditions, the *C. elegans* life cycle is about 3 days, with lifespan being about 3 weeks (Wood 1988).

C. elegans offers four distinct advantages for metabolomic studies. First, their genetic background is well-defined, being the first multicellular organism to be completely genetically sequenced (Hodgkin 2005). Although *C. elegans* exists as two sexes, hermaphrodite and male, sexual reproduction is primarily the result of self-fertilization by the hermaphrodite. Thus, *C. elegans* are isogenic, creating genetically identical populations. Second, the genetic background can be easily manipulated. Transgenic animals can be created using microinjection; DNA is injected into the distal arm of the gonad, where germ line nuclei are located (Evans 2006). RNA-interference (RNAi)-mediated gene suppression can be achieved either through micro-injection (Fire et al. 1998) or more easily through feeding (Timmons et al. 2001). Third, culture conditions can be strictly defined and controlled. *C. elegans* primarily feed on *E. coli* bacteria; however they can also be axenically cultured for optimal environmental control. Fourth, the wild-type *C. elegans* life cycle is about 3 days (Wood 1988), with adult hermaphrodites producing about 300 offspring (data presented in Chapter 3). Thus, cultures adequate for metabolomics studies are easily obtained. These four advantages combined with the structural and functional similarity of the nematode MRC to the human MRC makes *C. elegans* an ideal model system for mitochondrial disease.

Saccharomyces cerevisiae

S. cerevisiae is a yeast species commonly used in laboratory studies. Similar to *C. elegans*, *S. cerevisiae* offers the advantages of having a well-defined genetic background and being easily cultured. Yeast are also highly conducive to performing exometabolomic studies as they can be efficiently grown and

separated from small volume liquid cultures. Metabolic footprinting of yeast has been used for a number of purposes: to establish the aromatic effects of using multiple yeast strains during wine fermentation (Howell et al. 2006); to produce organic acids such as succinate, malate, pyruvate and lactate on an industrial scale (Abbott et al. 2009); and to understand growth rate control (Castrillo et al. 2007). Research for this thesis used *S. cerevisiae* as a model of succinate dehydrogenase dysfunction; the yeast enzyme is structurally and functionally similar to the mammalian enzyme (Lemire and Oyedotun 2002).

Thesis Objective

As discussed in this chapter, the metabolome is extremely sensitive to a number of pressures including gender, genetic background and lifestyle. Due to the extreme heterogeneity of the human metabolome (Zuppi et al. 1998; Saude 2007), model systems are often used in metabolomic studies. Model systems provide a well-defined genetic background and are grown under highly-controlled conditions. I predicted that, by studying the two above named model systems, greater experimental control over metabolic noise could be achieved, providing more meaningful results. Chapter 2 will discuss the level of metabolic variation in *C. elegans* fed two different diets and in complemented *S. cerevisiae* strains. Chapter 3 will explore the biological effects that dietary *E. coli* strains have on *C. elegans*. Two laboratory strains of *E. coli* are routinely used to feed *C. elegans*; however the differential effects that these two strains have on the nematode have not been previously explored. The effects that diet has on the *C. elegans* metabolome, mitochondrial DNA copy number, lifespan and brood size will be discussed.

Mitochondrial disease is extremely complex, causing a wide variety of clinical phenotypes. As mitochondria are the site of aerobic energy production and are central to global metabolism, I hypothesized that an exploration into the metabolic effects of mitochondrial dysfunction would provide insight to the molecular mechanisms underlying disease. Chapter 4 will explore the metabolic effects of mitochondrial Complex II dysfunction as reflected in the *S. cerevisiae* exometabolome.

Although the primary objective of this thesis research was to apply metabolomics to gaining an understanding of the molecular mechanisms of mitochondrial disease, I also had the opportunity to explore the effects of a rabies virus infection and neuro-therapeutic protocol. My collaborators and I hypothesized that the aggressiveness of this infection and its treatment would cause notable pathological, molecular, and metabolic neurological consequences; chapter 5 will highlight these.

Figures

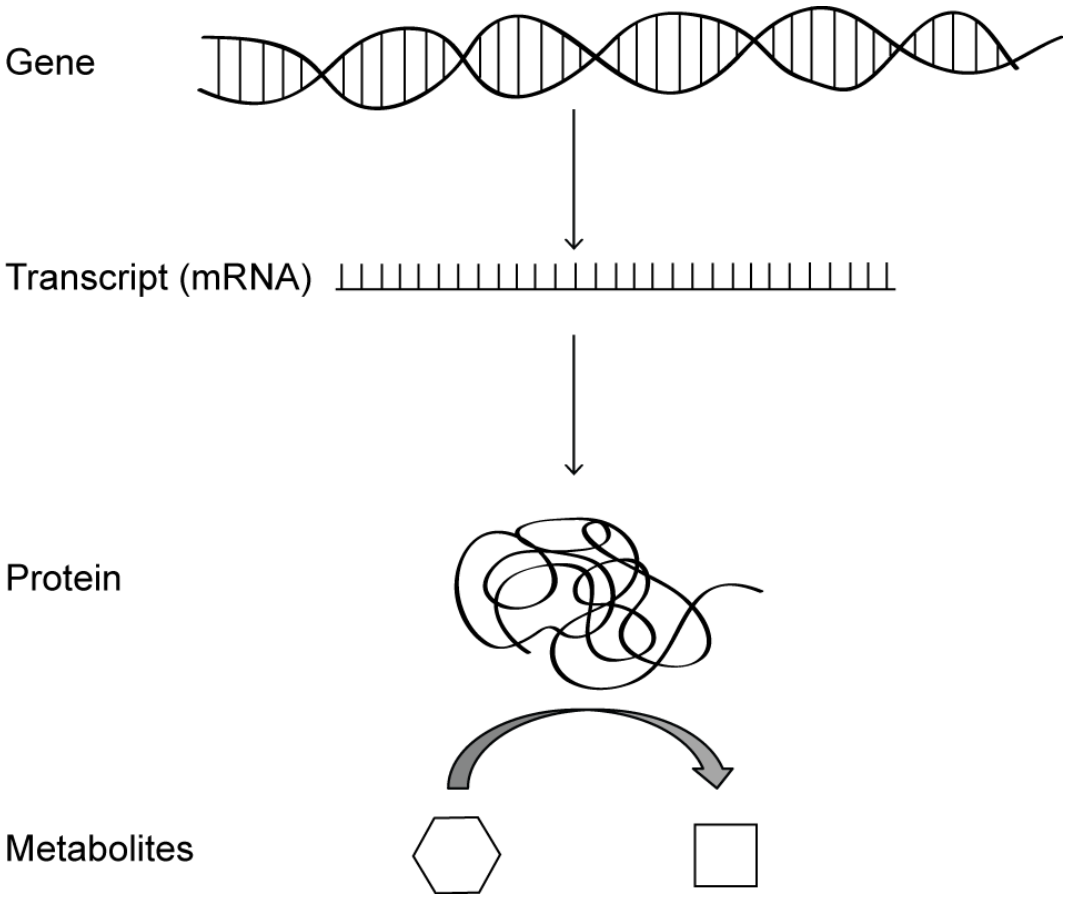


Figure 1-1. The four functional biological levels. Genes (genome) encode mRNA (transcriptome), which translate proteins (proteome). The proteome catalyzes biochemical reactions to transform metabolites (metabolome).

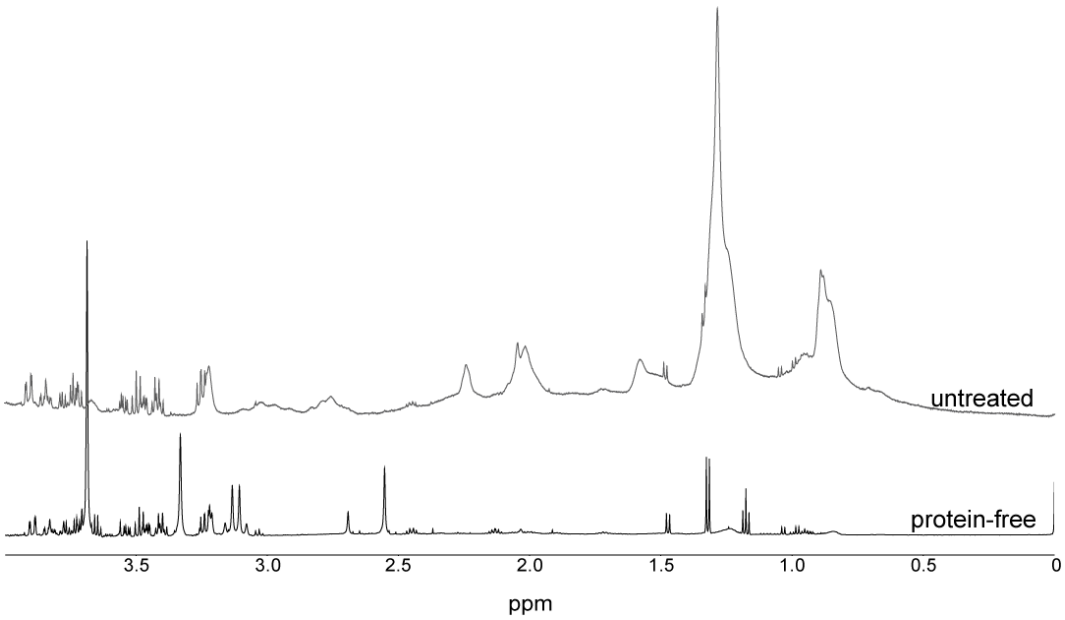


Figure 1-2. The 0-4 ppm region of representative 1H-NMR spectra. The top spectrum is from a sample that has been untreated. Broad peaks are from protein and lipids, and overlay metabolites of interest. The bottom spectrum is from a sample in which proteins have been removed. The baseline is flat and narrow metabolite peaks are easily identifiable.

	X-Variables (Predictor Variables)						Y-Variables (Responses)	
	Metabolite 1	Metabolite 2	Metabolite 3	…	Metabolite *k*-1	Metabolite *k*	Response 1	Response 2
Sample 1								
Sample 1								
Sample 1								
…								
…								
Sample *n*-2								
Sample *n*-1								
Sample *n*								

Figure 1-3. A metabolomic data set. There are *n* observations (samples) and *k* variables (metabolites, X-space; responses, Y-space) in a metabolomic data matrix. Each square represents one data point. Adapted from Goodacre et al. (2004).

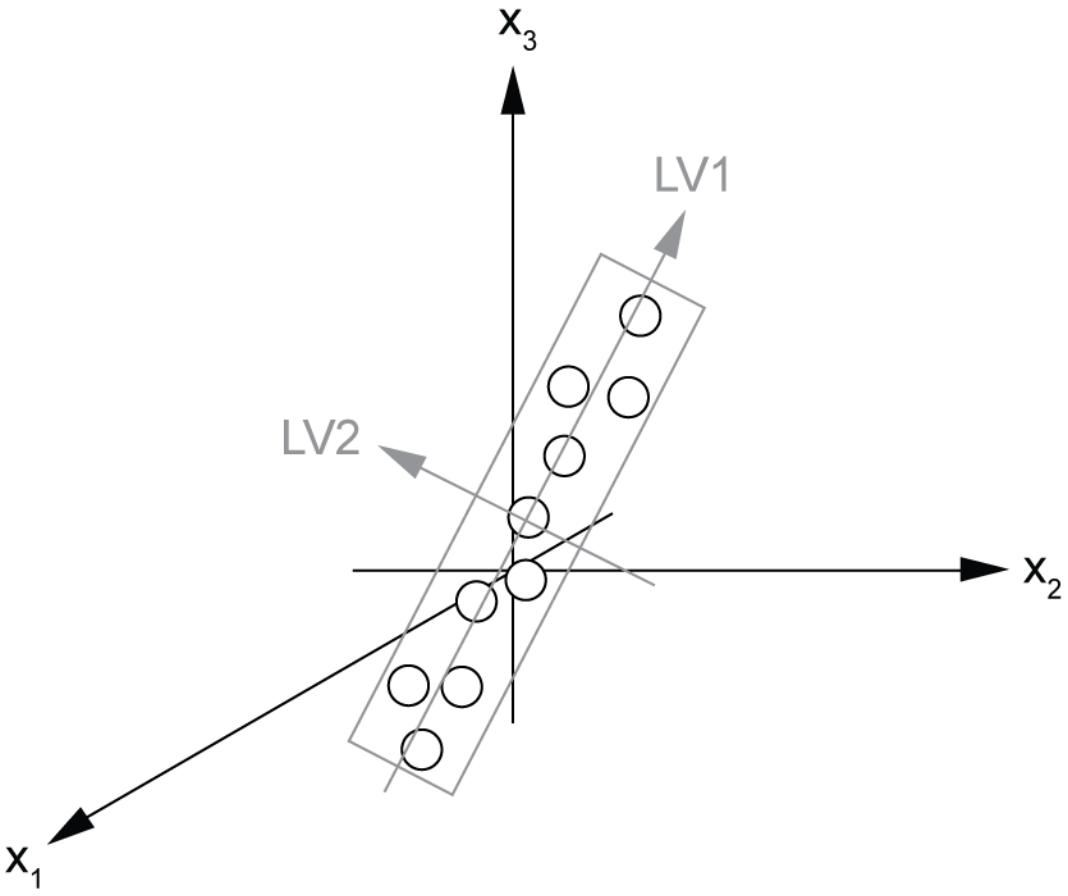


Figure 1-4. Projection model. Ten observations (circles) are plotted in three dimensions; each dimension represents one variable. The projection model finds two new latent variables (LVs). Adapted from Eriksson et al. (2001).

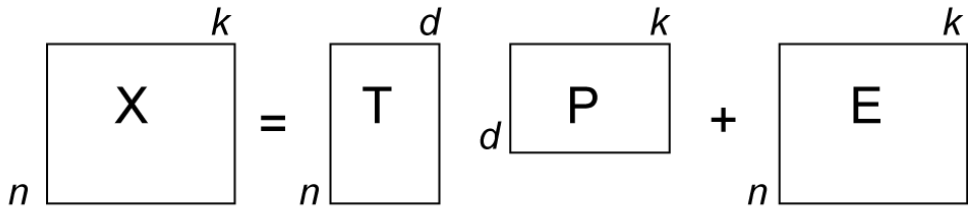


Figure 1-5. Geometric representation of PCA matrices. Data (X-matrix) are decomposed into the score (T) matrix, loadings (P) matrix and error (E) matrix. *d*, number of principal components (latent variables); *k*, number of variables (metabolites); *n*, number of observations (samples). Adapted from Otto (1999).

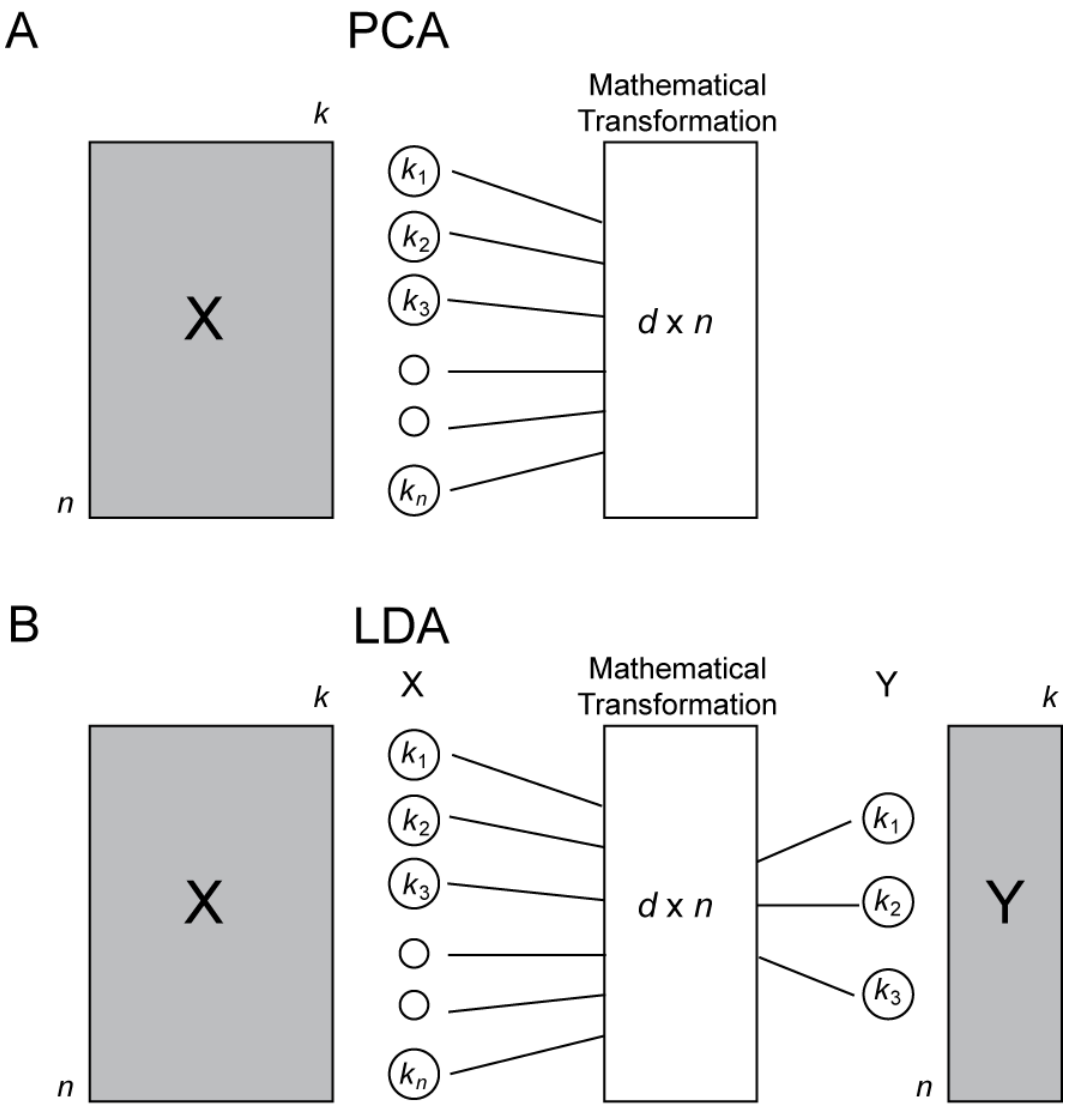


Figure 1-6. A visual representation of PCA and LDA. A, PCA mathematically transforms the X-data with *k* variables and projects it along *d* latent variables. B, LDA rotates projections from both the X- and Y-spaces to find the latent variables that explain the maximum X-variation direction in the Y-space. Adapted from Broadhurst and Kell (2006).

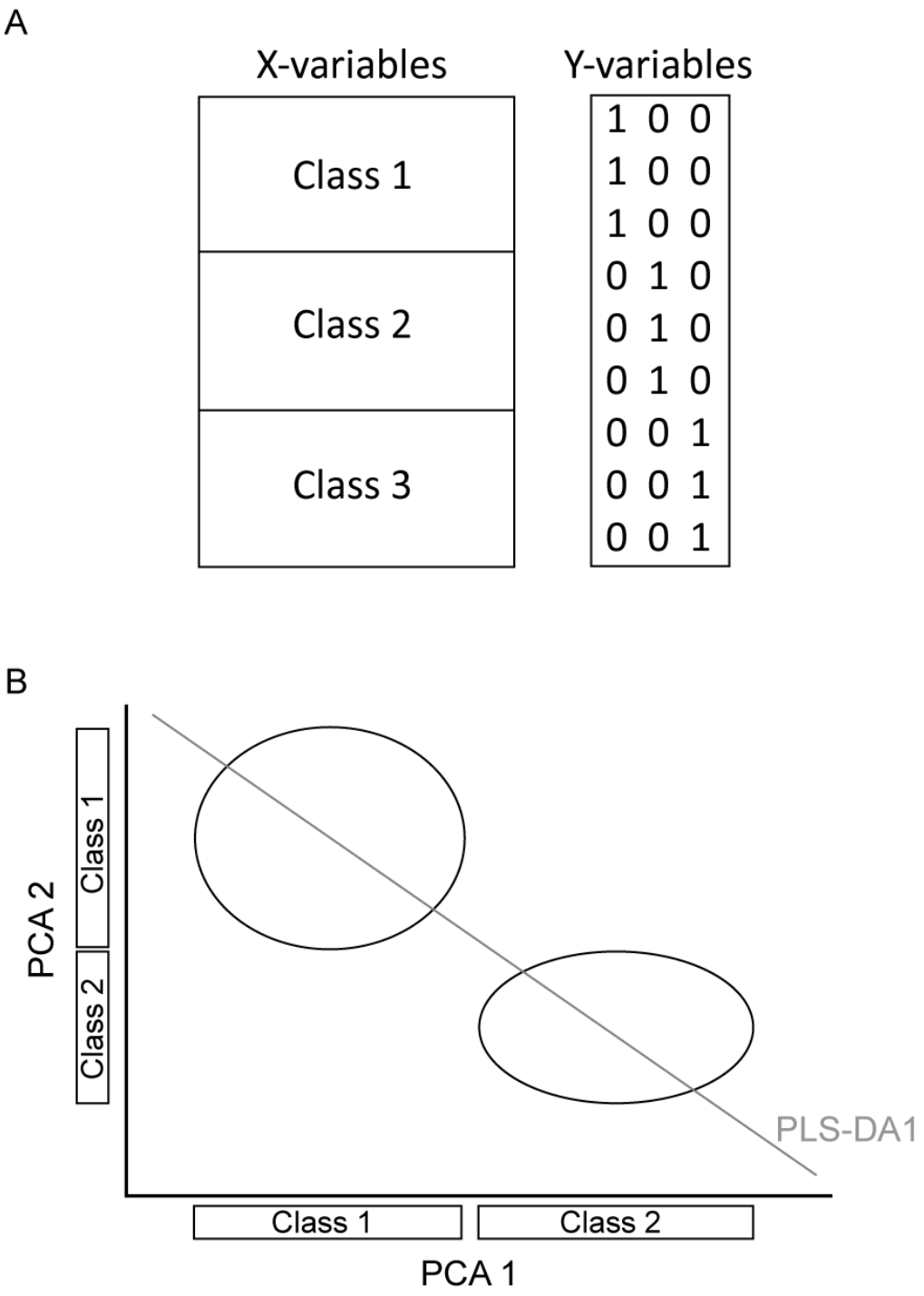


Figure 1-7. PLS-DA. A, A binary set of Y-variables is created to classify different groups of observations in the X-matrix. B, Two classes of observations are represented (ovals) along two principal components. The principal components measure the maximum variance between the two classes. PLS-DA measures the maximum discriminative variance between the two classes. Panel A adapted from Eriksson et al. (2001).

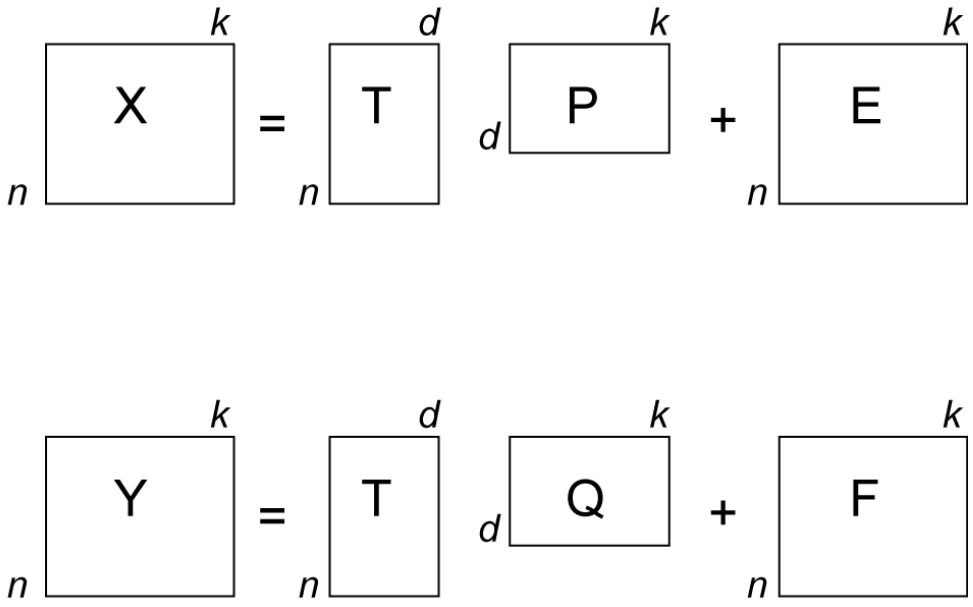


Figure 1-8. Geometric representation of the X and Y data matrices in PLS. The X and Y matrices are decomposed into score, loadings and error (unexplained) variance matrices. T, latent projection scores; P, X-loadings; Q, Y-loadings; E & F, unexplained variance in the model in both the X and Y space. *d*, number of latent variables in each matrix; *k*, number of variables in each matrix; *n*, number of observations. Adapted from Otto (1999).

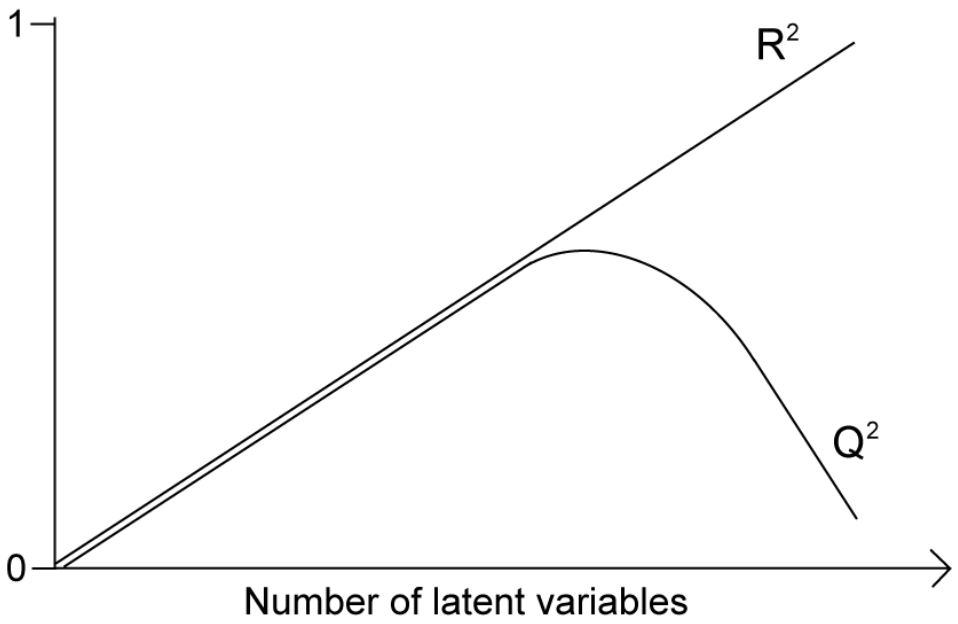


Figure 1-9. Effect of increasing the number of latent variables on model goodness of fit and goodness of prediction. As the number of latent variables increases, the explained variability approaches 100% (yielding a goodness of fit, R^2, value of 1). However, once all biological variance is explained, latent variables model noise, thus reducing the predictive power of the model (goodness of prediction Q^2 decreases).

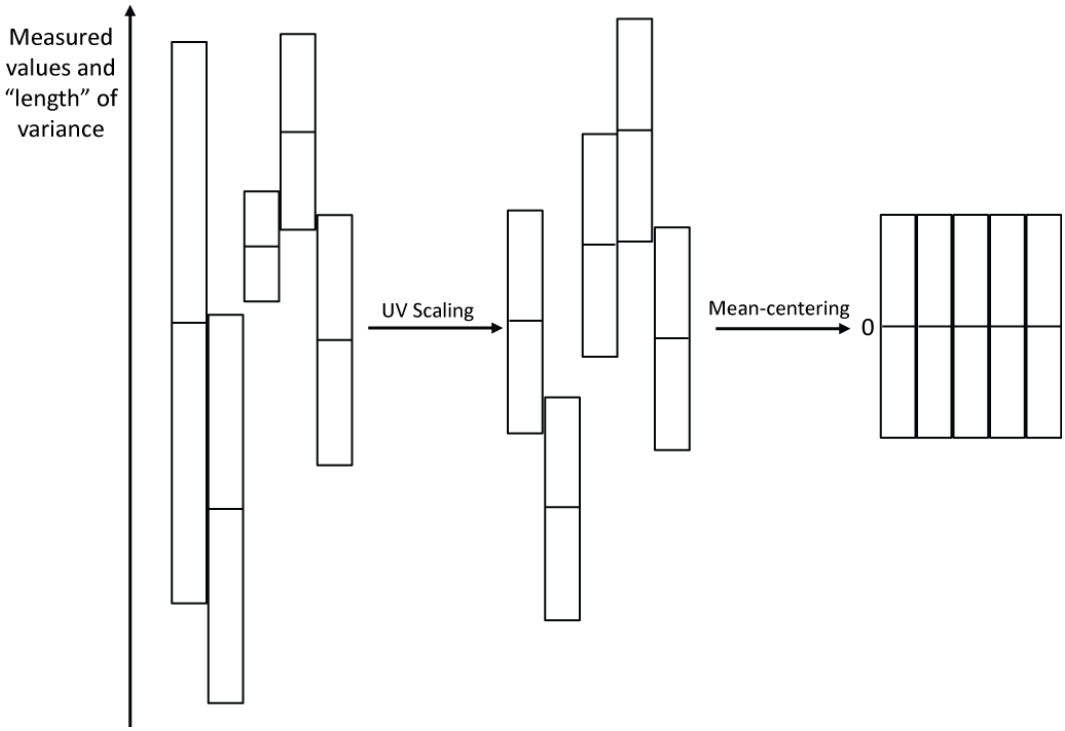


Figure 1-10. Data pre-processing. Prior to multi-variate data analysis, data are pre-processed to ensure all variables are treated with *a priori* equal importance. First, scaling reduces the effect of variables with large data ranges while amplifying the effect of variables with small data ranges. Second, the mean value of each variable is subtracted from each data point within that variable, thus eliminating constant data offset. Adapted from Eriksson et al. (2001).

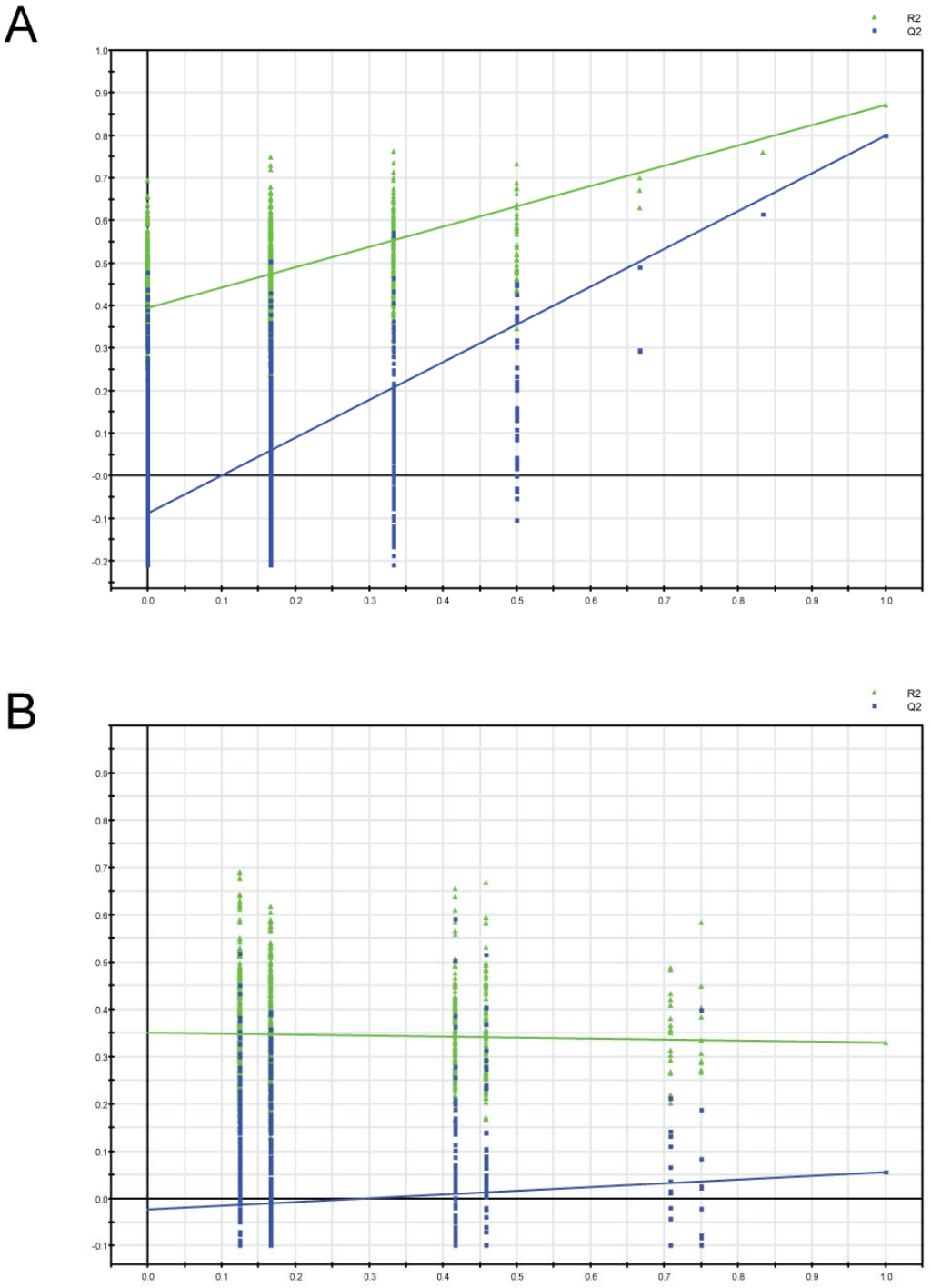


Figure 1-11. Validation Plots. A, A validation plot representing a valid data set within a PLS-DA model. Two features indicate that this plot is valid. First, the original R^2 and Q^2 values (points on far right) are higher than all permutations (all points to the left of originals). Second, the y-intercepts for the R^2 and Q^2 regression lines are not above 0.4 and 0.05, respectively. B, A validation plot representing data that are not valid; several permutations out-perform the original model.

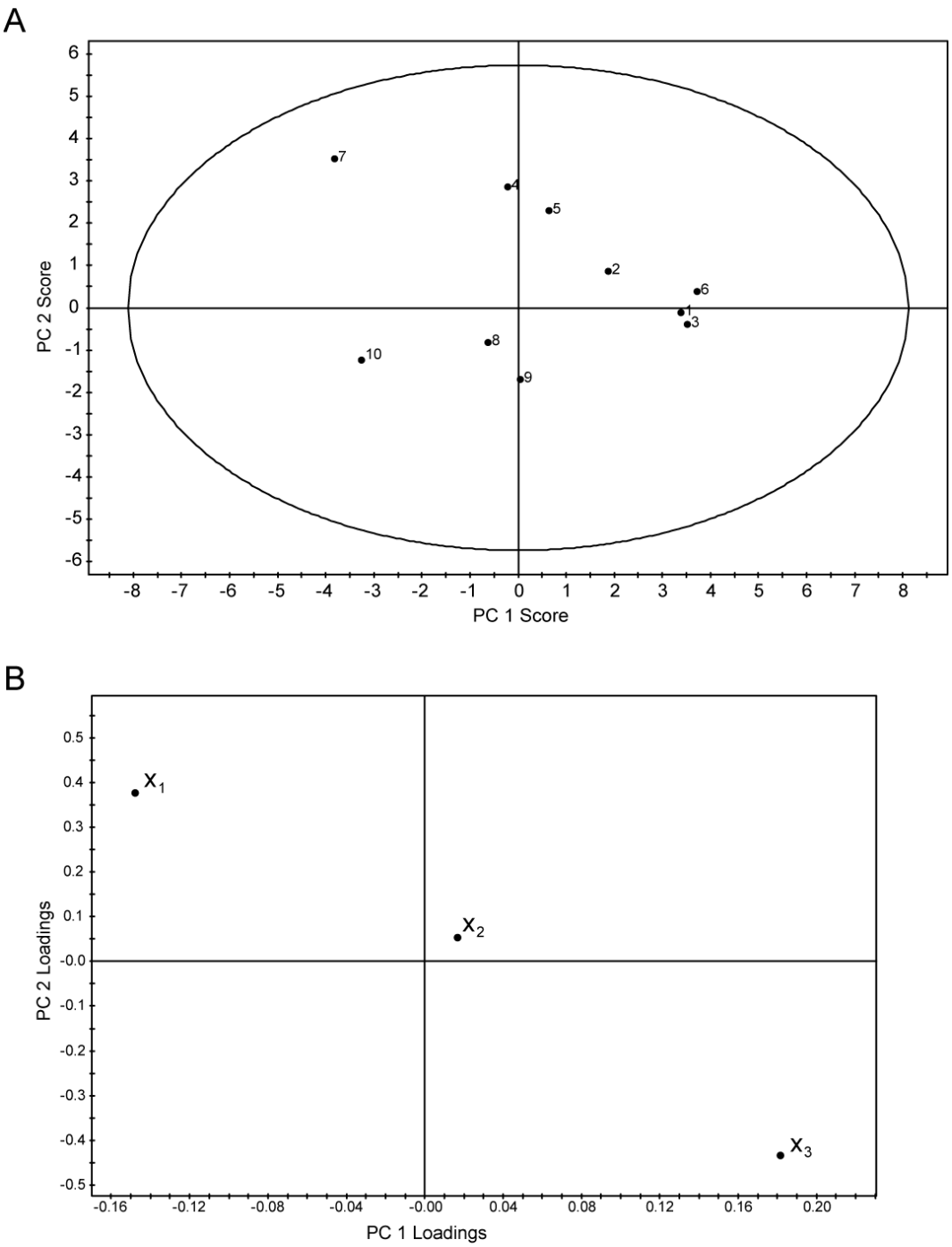


Figure 1-12. Score and loadings plots. A, Principal component (PC) scores are plotted for 10 observations (dots). B, Loadings are plotted for 3 variables: x_1, x_2, and x_3. Variables x_1 and x_3 are orthogonal to each other and thus contribute to model generation; variable x_2 is near the coordinate origin and has very little influence on the model.

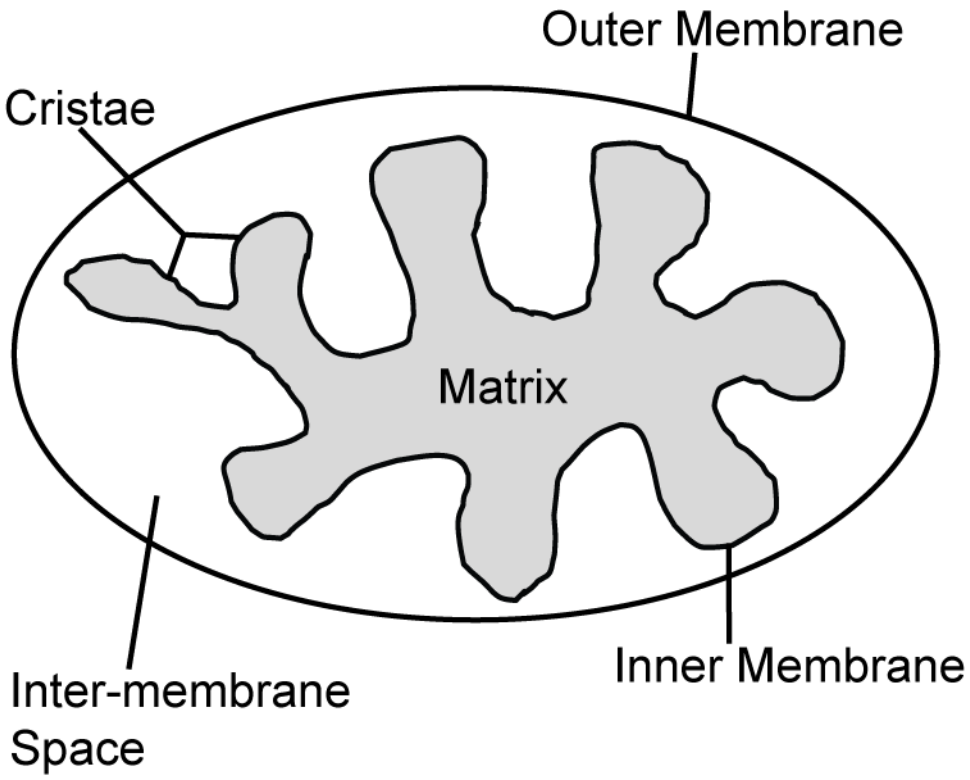


Figure 1-13. Schematic diagram of a mitochondrion. Mitochondria have two membranes separated by the inter-membrane space. The inner membrane surrounds the matrix and contains several folds, called cristae to maximize surface area for aerobic respiration.

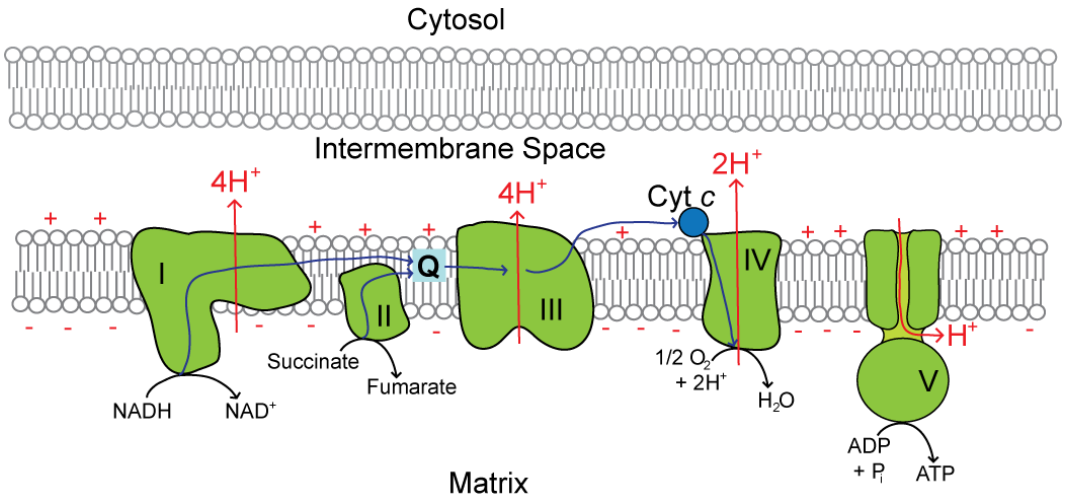


Figure 1-14. Mitochondrial respiratory chain. Complexes I – V are shown in order from left to right. Black arrows, metabolic redox and phosphorylation reactions; red arrows, proton path; blue arrows, electron path; Q, ubiquinone; Cyt *c*, cytochrome *c*. Adapted from Nelson and Cox (2005).

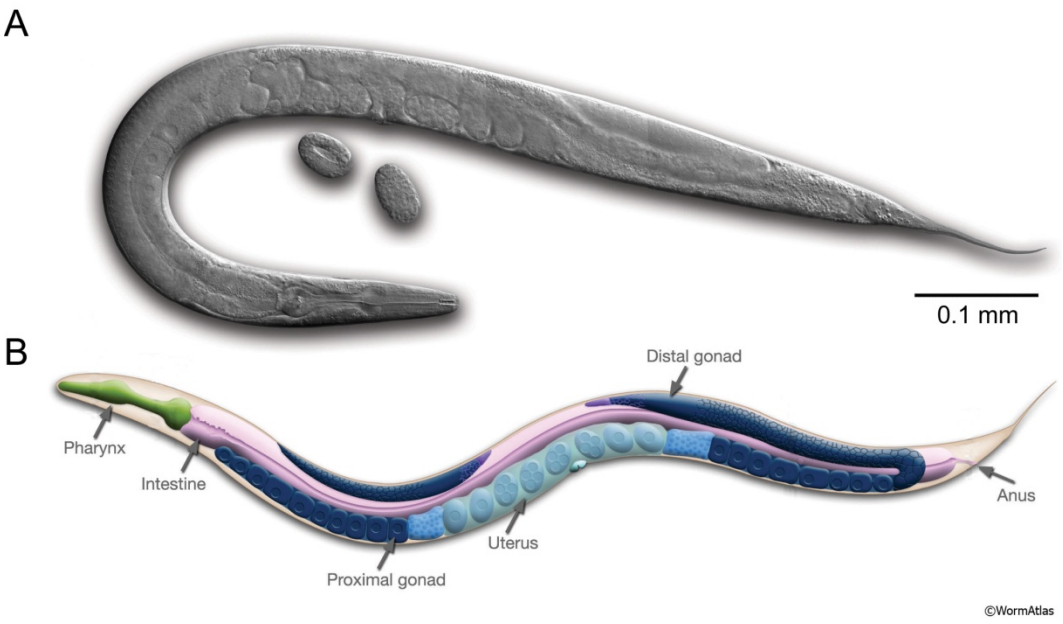


Figure 1-15. Diagram of *C. elegans* hermaphrodite. A, Photograph of adult with laid eggs. B, Illustration of adult with structures of the digestive and reproductive systems labelled. Adapted from Hall and Altun (2008).

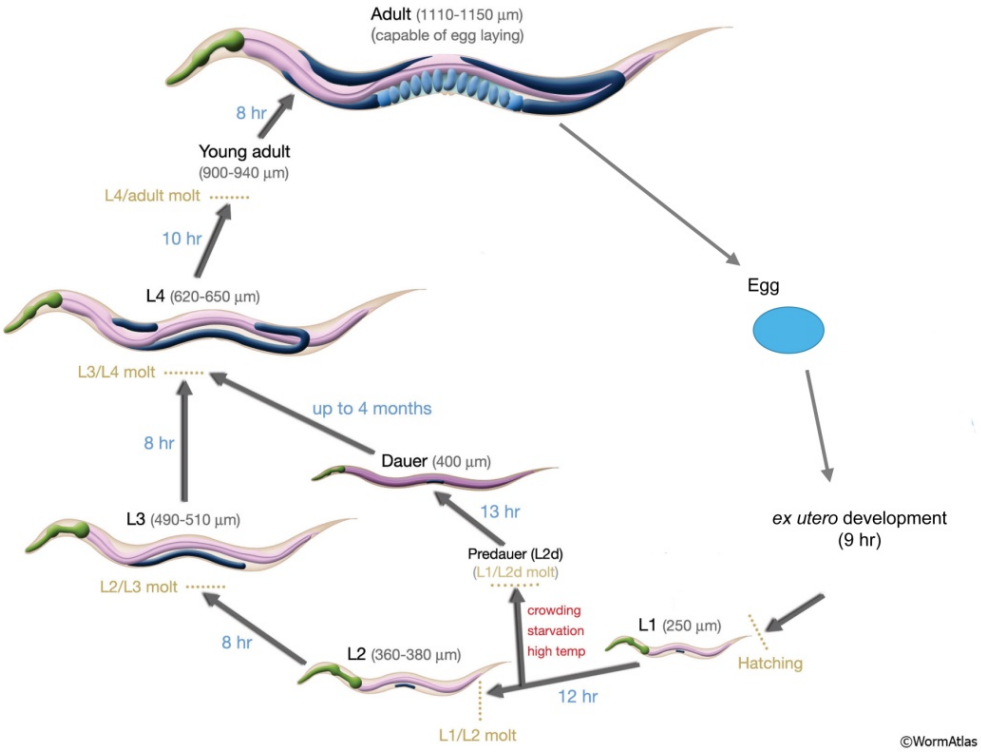


Figure 1-16. The *C. elegans* lifecycle. Adult hermaphrodites self-fertilize and lay eggs that continue to develop *ex utero* for about 9 hours before hatching. Upon hatching, worms transition through four larval stages, each separated by a cuticle moult, before entering adulthood. Under undesirable growth conditions, larvae can enter the dauer stage at the L1/L2 molt; dauers can survive several weeks and develop into L4 larvae upon sensing desirable growth conditions. Adapted from Hall and Altun (2008).

References

Abbott DA, Zelle RM, Pronk JT, van Maris AJ (2009) Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. FEMS Yeast Res 9 (8):1123-1136

Alberti G (2001) Noncommunicable diseases: tomorrow's pandemics. Bull World Health Organ 79 (10):907

Ali K, Maltese F, Toepfer R, Choi YH, Verpoorte R (2011) Metabolic characterization of Palatinate German white wines according to sensory attributes, varieties, and vintages using NMR spectroscopy and multivariate data analyses. J Biomol NMR 49 (3-4):255-266

Bollard ME, Holmes E, Lindon JC, Mitchell SC, Branstetter D, Zhang W, Nicholson JK (2001) Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution ^{1}H NMR spectroscopy of urine and pattern recognition. Anal Biochem 295 (2):194-202

Bollina V, Kushalappa AC, Choo TM, Dion Y, Rioux S (2011) Identification of metabolites related to mechanisms of resistance in barley against *Fusarium graminearum*, based on mass spectrometry. Plant Mol Biol 77 (4-5): 355-370

Broadhurst DI, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2 (4):171-196

Cassada RC, Russell RL (1975) The dauerlarva, a post-embryonic developmental variant of the nematode *Caenorhabditis elegans*. Dev Biol 46 (2):326-342

Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst D, O'Donoghue K, Hester SS, Dunkley TP, Hart SR, Swainston N, Li P, Gaskell SJ, Paton NW, Lilley KS, Kell DB, Oliver SG (2007) Growth control of the eukaryote cell: a systems biology study in yeast. J Biol 6 (2):4

Chomyn A, Enriquez JA, Micol V, Fernandez-Silva P, Attardi G (2000) The mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episode syndrome-associated human mitochondrial tRNALeu(UUR) mutation causes aminoacylation deficiency and concomitant reduced association of mRNA with ribosomes. J Biol Chem 275 (25):19198-19209

Cuadros-Inostroza A, Giavalisco P, Hummel J, Eckardt A, Willmitzer L, Pena-Cortes H (2010) Discrimination of wine attributes by metabolome analysis. Anal Chem 82 (9):3573-3580

Davies HV, Shepherd LV, Stewart D, Frank T, Rohlig RM, Engel KH (2010) Metabolome variability in crop plant species--when, where, how much and so what? Regul Toxicol Pharmacol 58 (3 Suppl):S54-61

Dunn WB, Bailey NJ, Johnson HE (2005) Measuring the metabolome: current analytical technologies. Analyst 130 (5):606-625

Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL (2011) Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chem Soc Rev 40 (1):387-426

Ellis DI, Goodacre R (2006) Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 131 (8):875-885

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S (2001) Multi- and megavariate data analysis. Part 1: Basic principles and applications. 2nd edn. Umetrics Academy, Umeå, Sweden

Evans TC, ed. Transformation and microinjection (April 6, 2006), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1985/wormbook.1.108.1, http://www.wormbook.org

Fancy SA, Beckonert O, Darbon G, Yabsley W, Walley R, Baker D, Perkins GL, Pullen FS, Rumpel K (2006) Gas chromatography/flame ionisation detection mass spectrometry for the detection of endogenous urine metabolites for metabonomic studies and its use as a complementary tool to nuclear magnetic resonance spectroscopy. Rapid Commun Mass Spectrom 20 (15):2271-2280

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391 (6669):806-811

Fridovich I (1997) Superoxide anion radical (O^{2-}.), superoxide dismutases, and related matters. J Biol Chem 272 (30):18515-18517

Garrett RH, Grisham CM (2010) Biochemistry. 4th edn. Brooks/Cole, Boston

German JB, Hammock BD, Watkins SM (2005) Metabolomics: building on a century of biochemistry to guide human health. Metabolomics 1 (1):3-9

Goodacre R (2007) Metabolomics of a superorganism. J Nutr 137 (1 Suppl):259S-266S

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol 22 (5):245-252

Gross S, Cairns RA, Minden MD, Driggers EM, Bittinger MA, Jang HG, Sasaki M, Jin S, Schenkein DP, Su SM, Dang L, Fantin VR, Mak TW (2010) Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. J Exp Med 207 (2):339-344

Hall DH, Altun ZF (2008) *C. elegans* Atlas. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Hodgkin J Introduction to genetics and genomics (September 6, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.17.1, http://www.wormbook.org

Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. Cell 134 (5):714-717

Howell KS, Cozzolino D, Bartowsky EJ, Fleet GH, Henschke PA (2006) Metabolic profiling as a tool for revealing Saccharomyces interactions during wine fermentation. FEMS Yeast Res 6 (1):91-101

Hu PJ Dauer (August 8, 2007) *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.144.1, http://www.wormbook.org

Jewell N (2000) A tutorial on PLS and regression. http://cisrg.shef.ac.uk/people/jewelln/Regression%20Tutorial/index.htm. Accessed September 20 2011

Kagan VE, Tyurin VA, Jiang J, Tyurina YY, Ritov VB, Amoscato AA, Osipov AN, Belikova NA, Kapralov AA, Kini V, Vlasova, II, Zhao Q, Zou M, Di P, Svistunenko DA, Kurnikov IV, Borisenko GG (2005) Cytochrome c acts as a cardiolipin oxygenase required for release of proapoptotic factors. Nat Chem Biol 1 (4):223-232

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG (2005) Metabolic footprinting and systems biology: the medium is the message. Nat Rev Microbiol 3 (7):557-565

Kenny LC, Broadhurst DI, Dunn W, Brown M, North RA, McCowan L, Roberts C, Cooper GJ, Kell DB, Baker PN (2010) Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. Hypertension 56 (4):741-749

Lemire BD, Oyedotun KS (2002) The *Saccharomyces cerevisiae* mitochondrial succinate:ubiquinone oxidoreductase. Biochim Biophys Acta 1553 (1-2):102-116

Lindon JC, Holmes E, Nicholson JK (2003) So what's the deal with metabonomics? Anal Chem 75 (17):384A-391A

Mounicou S, Szpunar J, Lobinski R (2009) Metallomics: the concept and methodology. Chem Soc Rev 38 (4):1119-1138

Nelson DL, Cox MM (2005) Lehninger Principles of Biochemistry. 4th edn. W.H. Freeman, New York

Orrenius S, Nicotera P, Zhivotovsky B (2011) Cell death mechanisms and their implications in toxicology. Toxicol Sci 119 (1):3-19

Ott M, Robertson JD, Gogvadze V, Zhivotovsky B, Orrenius S (2002) Cytochrome c release from mitochondria proceeds by a two-step process. Proc Natl Acad Sci U S A 99 (3):1259-1263

Otto M (1999) Chemometrics. Wiley-VCH, Weinheim, Germany

Riddle DL (1988) The Dauer Larva. In: Wood WB, reseachers Co*Ce* (eds) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp 393-414

Rossignol R, Faustin B, Rocher C, Malgat M, Mazat JP, Letellier T (2003) Mitochondrial threshold effects. Biochem J 370 (Pt 3):751-762

Sangster T, Major H, Plumb R, Wilson AJ, Wilson ID (2006) A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. Analyst 131 (10):1075-1078

Saude EJ, Adamko, D., Rowe, B.H, Marrie, T., Sykes, B.D. (2007) Variation of metabolites in normal human urine. Metabolomics 3:439-451.

Scheffler IE (1999) Mitochondria. Wiley-Liss, New York

Scheffler IE (2000) A century of mitochondrial research: achievements and perspectives. Mitochondrion 1:3-31

Schmidt V (1950) The significance of the Exton-Rose tolerance test for the diagnosis of diabetes mellitus. Acta Med Scand 136 (6):408-416

Shlens J (2009) A tutorial on principal component analysis (version 3.01). http://www.snl.salk.edu/~shlens/pca.pdf. Accessed October 6 2011

Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. Physiol Plant 132 (2):199-208

Simpson MJ, McKelvie JR (2009) Environmental metabolomics: new insights into earthworm ecotoxicity and contaminant bioavailability in soil. Anal Bioanal Chem 394 (1):137-149

Slupsky CM, Rankin KN, Fu H, Chang D, Rowe BH, Charles PG, McGeer A, Low D, Long R, Kunimoto D, Sawyer MB, Fedorak RN, Adamko DJ, Saude EJ, Shah SL, Marrie TJ (2009) Pneumococcal pneumonia: potential for

diagnosis through a urinary metabolic profile. J Proteome Res 8 (12):5550-5558

Smeitink J, van den Heuvel L, DiMauro S (2001) The genetics and pathology of oxidative phosphorylation. Nat Rev Genet 2 (5):342-352

Solanky KS, Bailey NJ, Beckwith-Hall BM, Davis A, Bingham S, Holmes E, Nicholson JK, Cassidy A (2003) Application of biofluid ^{1}H nuclear magnetic resonance-based metabonomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. Anal Biochem 323 (2):197-204

Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 457 (7231):910-914

Timmons L, Court DL, Fire A (2001) Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. Gene 263 (1-2):103-112

Toninello A, Salvi M, Pietrangeli P, Mondovi B (2004) Biogenic amines and apoptosis: minireview article. Amino Acids 26 (4):339-343

UMDF (2011a) Getting a Diagnosis. United Mitochondrial Disease Foundation. http://www.umdf.org/site/c.otJVJ7MMIqE/b.5692885/k.E71C/Getting_a_Diagnosis.htm. Accessed July 27 2011

UMDF (2011b) Possible Symptoms. United Mitochondrial Disease Foundation. http://www.umdf.org/site/c.otJVJ7MMIqE/b.5692883/k.C0C7/Possible_Symptoms.htm. Accessed July 29 2011

UMDF (2011c) What is Mitochondrial Disease. United Mitochondrial Disease Foundation. http://www.umdf.org/site/c.otJVJ7MMIqE/b.5692879/k.3851/What_is_Mitochondrial_Disease.htm. Accessed July 17 2011

van Dorsten FA, Grun CH, van Velzen EJ, Jacobs DM, Draijer R, van Duynhoven JP (2010) The metabolic fate of red wine and grape juice polyphenols in humans assessed by metabolomics. Mol Nutr Food Res 54 (7):897-908

van Duynhoven J, Vaughan EE, Jacobs DM, Kemperman RA, van Velzen EJ, Gross G, Roger LC, Possemiers S, Smilde AK, Dore J, Westerhuis JA, Van de Wiele T (2011) Metabolic fate of polyphenols in the human superorganism. Proc Natl Acad Sci U S A 108 Suppl 1:4531-4538

Viant MR, Pincetich CA, Tjeerdema RS (2006) Metabolic effects of dinoseb, diazinon and esfenvalerate in eyed eggs and alevins of Chinook salmon

(*Oncorhynchus tshawytscha*) determined by ^{1}H NMR metabolomics. Aquat Toxicol 77 (4):359-371

Wallace DC (1982) Structure and evolution of organelle genomes. Microbiol Rev 46 (2):208-240

Wallace DC (1999) Mitochondrial diseases in man and mouse. Science 283 (5407):1482-1488

Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu Rev Genet 39:359-407

Wallace DC (2010) Bioenergetics and the epigenome: interface between the environment and genes in common diseases. Dev Disabil Res Rev 16 (2):114-119

Wallace DC, Lott MT, Procaccio V (2007) Mitochondrial genes in degenerative disease, cancer, and aging. In: Rimoin DL, Connor JM, Pyeritz RE, Korf BR (eds) Emery and Rimoin's Principles and Practice of Medical Genetics, vol 1. Churchill Livingstone, Birmingham, pp 194-298

Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM, Wu Y, Schauer P, Smith JD, Allayee H, Tang WH, DiDonato JA, Lusis AJ, Hazen SL (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472 (7341):57-63

Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM (2006) Targeted profiling: quantitative analysis of ^{1}H NMR metabolomics data. Anal Chem 78 (13):4430-4442

Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, Xiong Y, Clive D, Greiner R, Nazyrova A, Shaykhutdinov R, Li L, Vogel HJ, Forsythe I (2009) HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res 37 (Database issue):D603-610

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L (2007) HMDB: the Human Metabolome Database. Nucleic Acids Res 35 (Database issue):D521-526

Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58 (2):109-130

Wood WB (1988) Introduction to *C. elegans* Biology. In: Wood WB, researchers Co*Ce* (eds) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp 1-16

Zelena E, Dunn WB, Broadhurst D, Francis-McIntyre S, Carroll KM, Begley P, O'Hagan S, Knowles JD, Halsall A, Wilson ID, Kell DB (2009) Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. Anal Chem 81 (4):1357-1364

Zhang GF, Sadhukhan S, Tochtrop GP, Brunengraber H (2011) Metabolomics, pathway regulation, and pathway discovery. J Biol Chem 286 (27):23631-23635

Zuppi C, Messana I, Forni F, Ferrari F, Rossi C, Giardina B (1998) Influence of feeding on metabolite excretion evidenced by urine ^{1}H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svaldbard). Clinica chimica acta; international journal of clinical chemistry 278 (1):75-79

CHAPTER 2

^{1}H-NMR-based metabolic profiling reveals inherent biological variation in the yeast and nematode model systems.

Samuel SW Szeto, Stacey N Reinke, and BD Lemire.

Department of Biochemistry, University of Alberta, Edmonton, Canada

A version of this chapter has been accepted for publication. Szeto SSW, Reinke SN, and BD Lemire 2011. *J Biomol NMR*, 49:245-54.

Contributions: SSWS and SNR contributed equally to this work; SNR provided all nematode data and analyzed all NMR spectra; SSWS and SNR prepared all yeast samples, acquired spectra, and generated figures; SSWS, SNR, and BDL wrote the manuscript.

Introduction

The advent of metabolomics brought substantial promise for diagnosing and understanding human disease. However, it quickly became apparent how diverse and variable the metabolome of any given tissue or biofluid could be. A number of intrinsic and extrinsic factors place pressure on the metabolome, leading to biological variability. These factors include, but are not limited to, genetic background, gender, age, lifestyle, microflora, environmental factors, and exposure to xenobiotics (Holmes et al. 2008; Walsh et al. 2008; Crews et al. 2009; Lenz et al. 2003; Solanky et al. 2003; Saude et al. 2007). The degree of metabolic variability is also dependent on sample origin. Urine is composed of waste products from the entire body and is thus highly variable. Blood is much less variable due to homeostatic control (Walsh et al. 2008). As the metabolome is highly sensitive, it reflects current stresses placed on the biological system. As these stressors fluctuate over time, so does the metabolome. Two separate studies evaluated intra-individual variability in the human urine metabolome over a period of time (Saude et al. 2007; Zuppi et al. 1998). Both studies revealed that intra-individual variability can be as pronounced as inter-individual variability; both levels of variability can lead to difficulties in data interpretation.

Technical or analytical variation can also contribute significantly to observed metabolic variation. A typical metabolomic experiment involves sample collection, extraction and/or processing, data acquisition, spectral processing, and mathematical and statistical analysis (Parsons et al. 2009). An experiment using model systems also includes media or food preparation and maintenance of

culture conditions. Thus, in addition to biological variability, experimental procedures offer several opportunities for variation to be introduced.

Relative standard deviation (RSD = standard deviation/mean *100%), or coefficient of variance (CV), provides a means for characterizing variability. RSDs can be applied to individual metabolites or to the entire data set. They are useful in assessing and improving data, and acting as frames of reference when developing new methodologies. RSDs are also particularly useful to provide target values for new metabolomic researchers to aim for (Parsons et al. 2009). When RSD values are evaluated from processing artificial samples, they can also be useful for self-evaluation to ensure analytical integrity. This can help distinguish the magnitude of actual biological variation from total observed variation.

Metabolomic studies have been applied to a number of non-human model systems in effort to gain control over factors contributing to metabolic variability. In model systems, genetic background is defined and environmental conditions can be controlled (Crews et al. 2009). Saude and colleagues analyzed urine from twenty-five guinea pigs, grown under identical experimental conditions in a sterile environment. The animals had similar genetic backgrounds. The study reported similar, or even greater, biological variability (RSD values ~50-300%) to humans (Saude et al. 2007). More recently, Parsons and colleagues determined RSD values for a number of metabolomic datasets, spanning a variety of sample types from mammals, fish, invertebrates and a cultured cell line. Inter-individual biological variation ranged from 7.2%, for rat tissue extracts, to 58.4% for fish plasma (Parsons et al. 2009). Metabolic analysis of *Arabidopsis thaliana* leaf

extracts revealed remarkable biological variation in plant model systems, with an average RSD of 40% (Fiehn et al. 2000).

In light of reports of inherent biological variation seen in model systems, we examined metabolic variability arising from two commonly used laboratory model systems: the nematode *Caenorhabditis elegans*, and the brewing yeast *Saccharomyces cerevisiae*. In this study, we calculated RSD values for quantified metabolites and report that an appreciable amount of biological variation exists. This variability is also apparent between genetically comparable yeast strains. Using multivariate statistical analysis, we are able to discriminate between the metabotypes of two succinate dehydrogenase (SDH) knockout strains and the two knockouts complemented with a plasmid carrying wild-type SDH genes; yet, these strains form two phenotypically identical pairs. Our results demonstrate that even "simple" model systems cultured under highly controller conditions exhibit significant metabolic variation.

Materials and Methods

Worm strains and culture conditions

The *C. elegans* N2 (Bristol) wild-type strain was used. Worms were cultured on either nematode growth media (NGM) plates seeded with *E. coli* OP50 or on NGM containing 1 mM IPTG and 25 μg/ml carbenicillin and seeded with *E. coli* HT115 (DE3) transformed with the L4440 vector (Lewis and Fleming 1995; Fraser et al. 2000). Cultures were inoculated with synchronized L1 stage worms and maintained at 20 ºC as described (Reinke et al. 2010). Worms were harvested, incubated in sterile water for 30 minutes to allow digestion of gut bacteria, and washed three times (Grad et al. 2007). Worm pellets were

suspended in 900 µl of sterile water and three 10 µl aliquots were removed to quantify total protein. Seven biological replicates were collected for worms fed *E. coli* OP50 and six were collected for worms fed *E. coli* HT115 L4440.

Yeast strains, media, and culture conditions

The *S. cerevisiae* knockout strains sdh3W1 (MH125, *sdh3::TRP1*) and sdh4W2 (MH125, *sdh4::TRP1*) have been described previously (Oyedotun and Lemire 1997; Oyedotun and Lemire 1999; Szeto et al. 2007). The yeast knockout strains were transformed with either empty low-copy plasmids or plasmids containing the corresponding wild-type *SDH3* or *SDH4* genes by lithium acetate-mediated transformation (Gietz et al. 1992). Strains were plated on solid SD medium containing cas amino acids (0.5% w/v) without tryptophan for plasmid retention. Cells were grown for 36 hrs at 30 ^{o}C in 2 ml of YP medium containing 0.25% glucose. This is a modified version of a medium previously used to examine acid secretion (Szeto et al. 2007; de Kok et al. 1975). To eliminate potential variations in the metabolite profiles arising from the growth medium composition, a single batch of medium was used to culture all strains in this study. Following completion of the incubation period, the optical densities (OD) of the cultures at 600 nm were determined. Cultures were centrifuged to pellet cells and the clarified media were transferred to microcentrifuge tubes as described (Szeto et al. 2010). Twelve biological replicates were collected for each strain.

Metabolite extraction

Worms were lysed on ice in trichloroacetic acid (TCA; 5% final concentration) by sonication. Clarified post growth yeast media were also treated with TCA (5%

final concentration) and incubated on ice for 30 minutes to precipitate protein. Both worm and yeast samples were centrifuged at 14 000 x *g* for 15 minutes at 4 ^{o}C. Supernatants were recovered and pH neutralized with 5 M NaOH. Protein-free lysates were re-clarified by centrifugation. All samples were flash frozen in liquid nitrogen before lyophilization.

Sample preparation for NMR spectroscopy

Dry protein-free lysates were dissolved in 570 µl D_2O (99.9%; Isotec Inc., Miamisburg, Ohio). 30 µl of 5 mM 2,2-dimethyl-2-sila 3,3,4,4,5,5-hexadeutero-pentane sulphonic acid (DSS-d_6, Chenomx Inc., Edmonton, Alberta) were added as a chemical shift indicator and concentration standard for NMR analysis. The pH was recorded for calibration purposes and samples were centrifuged at 14,000 x *g* for 3 min to remove particulate matter. 510 µl of supernatant were transferred to 5 mm diameter NMR tubes for data collection.

^{1}H NMR spectroscopy and NMR data processing

One-dimensional ^{1}H NMR spectra were acquired on a 600 MHz Varian Inova spectrometer (Varian Inc., Palo Alto, California) at 30 °C using a tnnoesy pulse sequence (circa Vnmr 6.1B software, Varian Inc.). Parameters were consistent with those previously described (Szeto et al. 2007, 2010; Reinke et al. 2010). Chenomx NMR Suite Professional software v5.1 (Chenomx Inc., Edmonton, Alberta) was used for metabolite identification and quantification. This software uses pattern recognition and lineshape deconvolution to fit spectra based on highly specific peak patterns (Weljie et al. 2006). The resonance linewidths are input from the reference standard, DSS. The spectral patterns of many

metabolites often contain more than one peak throughout the spectrum, forming complex and distinctive sets of peaks. In addition, the spectral baseline often varies across regions but fitting can be achieved by using only the top portion of any peak as long as overall fit is consistent across all regions. As such, baseline inconsistencies do not affect quantification accuracy. For additional consistency, the same sets of peaks were always utilized for quantification of any one metabolite.

Metabolite data analysis

Worm metabolic data were normalized to protein contents. Yeast metabolic data were normalized to the mass of total dried protein-free lysate and to OD of the culture. Relative standard deviation (standard deviation / mean x 100%) was determined for each metabolite in each sample set. Box plots were generated as previously described (Cann 2003; Massart et al. 2005). The Mann-Whitney test was used to analyze whether the two worm metabolite RSD distributions differed significantly (Corder and Foreman 2009). Multivariate statistical analyses were performed using Simca P+ v12.0.1 software (Umetrics, Umeå, Sweden) (Eriksson et al. 2001). Data were not transformed but were scaled to unit variance, dividing each variable by its standard deviation, and mean centered to provide equal importance to all variables. Data were visualized using both the unsupervised principal component analysis (PCA) and the supervised partial least square discriminate analysis (PLS-DA). R^2 and Q^2 values were used to assess the goodness of fit and prediction of the respective models. To examine the possibility of spurious model generation, validations of each group within each model were performed. The software randomly generated 999 permutations

of y-variables (metabolites), while keeping x-variables (strain) intact, and facilitated a comparison of the measures of fit (R^2 and Q^2) from the generated permutations to those of the original. Statistical significance of the differences between groups in each model was determined using CV-ANOVA (analysis of variance testing of cross-validated predictive residuals).

Results

Biological variation observed in the worm and yeast model systems

We determined the RSD values for each metabolite to quantify spectrum-wide variability in the metabolomes (Table 2-1, Figure 2-1) of whole worm, protein-free lysates and protein-free, exometabolome lysates of yeast. RSD values demonstrate reproducibility in large data sets but also offer a frame of reference when manipulating or developing new preparation methods for biological samples (Parsons et al. 2009). *C. elegans* are routinely cultured on NGM plates seeded with *E. coli* OP50. This strain is an uracil auxotroph and produces a thin lawn on NGM plates, allowing for easier visualization of the worms (Brenner 1974). Alternatively, another laboratory *E. coli* strain HT115 (DE3) is used for feeding-mediated gene suppression by RNA-interference (RNAi) (Kamath et al. 2001). HT115 is transformed with the empty vector L4440 as an RNAi control. Using six samples of wild-type worms fed *E. coli* HT115 L4440, variance in whole worm, protein-free lysates ranged from 17% to 46%, with an average RSD of 29%. Three metabolites glycerol, leucine and methionine had a coefficient of variance greater than 40%. Using seven samples, variance in worm lysates from wild-type worms fed *E. coli* OP50 ranged from 18% to 64%, with an average RSD of 39%. Nine metabolites had a coefficient of variance greater than 40%: choline,

fumarate, glucose, glutamate, glutamine, methionine, serine, tryptophan, and valine. The RSD distributions of these two metabolomes were compared and are significantly different (p<0.001). The average RSD of the yeast exometabolomes were 12% for the complemented strains and 11% for the knockout strains. Although the average RSD values were quite similar, there was greater variance in the replicates of the complemented *SDH4* strain, exhibiting a range of 5 to 35%. The most variable metabolites in the complemented *SDH3* exometabolome were asparagine, myo-inositol and nicotinate. Choline, asparagine, lactate and myo-inositol were the most variable exometabolites of the complemented *SDH4* metabolome. Choline was the most variable metabolite, with an RSD greater than 35%. Myo-inositol was the only metabolite in the *SDH3* knockout strain that varied by more than 20% of the mean. However, α-ketoglutarate, choline, lactate, and myo-inositol all exceeded 20% variation in the *SDH4* knockout strain. In addition, we did not observe any correlation between the RSD and the concentration of metabolites quantified. The RSD values were not significantly higher for metabolites that had either very small (<10 uM) or very large (>10mM) concentrations (data not shown). These data are also represented as box plots shown in Figure 2. The box plot yields a simple visual representation of the RSD data sets. The lower, middle, and upper lines of the boxes represent the lower quartile, median, and upper quartile of each RSD data set, respectively. The error bars show the range of the data. The yeast exometabolomes were much less variable than the whole-worm metabolome. Variance in the wild-type whole-worm metabolome was altered by food source.

Multivariate analysis of yeast exometabolomes

We employed multivariate analysis to establish metabotype differences between previously characterized yeast strains. The metabotypes of two knockout strains (*SDH3* KO and *SDH4* KO) were compared to each other. This was also done with the respective complemented strains (knockout strains with plasmid-borne copies of the wild-type *SDH3* and *SDH4* genes, respectively). Partial least squares discriminant analysis (PLS-DA) provided a supervised evaluation of the variation between the two data sets examined (Figure 3). Each colour-coded dot represents a replicate metabotype in the representative PLS-DA score plots. The PLS-DA models indicated that the complemented strains could be discriminated from each other (Figure 3A). A similar result was also obtained for the two knockout strains examined (Figure 3B). PLS-DA yields an opportunity for statistical analysis of the models. The PLS-DA model comparing the two complemented strains resolved two components with R^2X, R^2Y, and Q^2 values of 0.688, 0.872 and 0.8, respectively. The PLS-DA model comparing the two knockout stains resolved three components with R^2X, R^2Y, and Q^2 values of 0.568, 0.98 and 0.913, respectively. CV-ANOVA tests were also performed on each model. These resulted in the values $p=1.69 \times 10^{-6}$ for the complemented strain model and $p=5.67 \times 10^{-8}$ for the knockout strain model. Validation plots were also generated for each strain within each model from 999 random permutations of the data using all components (Figure 4). Two criteria indicate model validity. First, none of the permutations out-performed the original data set. The points representing the original data (the blue and green dots at the far right of the graph) are higher than the permutated data (all dots to the left of the original). Second, the Y-intercepts for the R^2 regression line (green) and the Q^2

regression line (blue) should be less than 0.4 and 0.05, respectively. Our data fit these criteria with the exception that the R^2Y-intercepts for the model comparing the knockout strains are higher than 0.4. However, when the high R^2 and Q^2 values for both models, the visual separation seen between strains in each model, the low CV-ANOVA *p* values, and the results of the validation permutations are considered, we suggest that the models are valid. To ensure that supervised evaluation did not falsely differentiate between the various yeast strains, principal component analysis (PCA) was also performed (Figure 5). PCA is unsupervised and treats each observation equally. Again, each colour-coded dot represents a replicate metabotype. In both models, visual separation between the strains in the score plots is evident. The PCA model comparing the two complemented strains resolved two components with R^2X and Q^2 values of 0.694 and 0.516, respectively. The model comparing the two knockout strains resolved three components with R^2X and Q^2 values of 0.608 and 0.314, respectively.

Discussion

The metabolome is an extremely sensitive monitor of physiological state. As such, metabolomic studies of human biofluids can show immense variability in the concentration of an individual metabolite. Concentrations of metabolites in human urine can vary by more than 200% of the mean (Saude et al. 2007). The composition of cerebrospinal fluid (CSF) is well maintained under homeostatic control, with RSD values ranging between 12 and 70% (Stoop et al. 2010). Model systems are highly conducive to metabolomic studies as genetic background and growth conditions can be well defined because genetic background and growth conditions can be controlled. The metabotypes of model systems under such

conditions should be less variable than those of their human counterparts. We show that in genetically identical animals grown under carefully controlled conditions, individual metabolite concentrations in *C. elegans* can vary by as much as 64% (Figures 1 and 2, Table 1). The less complex unicellular model system *S. cerevisiae* exhibited decreased biological variation in its exometabolomes, with a maximum RSD of 36%. However, the range of variation between the complemented strains differed. Whereas the complemented *SDH3* strain had a RSD range of 15%, the complemented *SDH4* strain had a RSD range of 31%. The yeast knockout strains were more similar in their RSD ranges: 17% for *SDH3* KO and 20% for *SDH4* KO, respectively. The observation that substantial biological variations exists in organisms, such as yeast, may not be a surprising result considering clonal populations of *E. coli* have also been shown to exhibit substantial phenotypic variation (Elowitz et al. 2002).

The greater variance in *C. elegans* than in *S. cerevisiae* may be attributed to four factors. First, *C. elegans* are social feeders. Despite adequate food and space for the whole population, local subpopulations may experience food limitation (de Bono and Bargmann 1998). Second, *C. elegans* develop through four larval stages before entering adulthood. Metabolic differences among the different stages have been documented (O'Riordan and Burnell 1989, 1990; Braeckman et al. 2009). The distribution of larval stages will vary somewhat between cultures. Third, the worms were cultured on NGM plates seeded and grown overnight with *E. coli*. Chemical composition will vary between plates and *E. coli* lawns. Additionally, several plates of worms are needed to obtain a sufficient number of worms on which to perform metabolic studies. For logistical reasons, we did not evaluate the metabolic composition of either NGM media or of the

E. coli. Fourth, intracellular metabolites were measured in worms and suboptimal worm lysis could also contribute to variation between cultures. However, we have optimized the sonication protocol to minimize incomplete lysis. The metabolic complexity of *C. elegans* and the inherently more variable culture conditions it demands likely account for most of the sample variation we detected. Our present findings also show that biological variation can also be significantly affected by diet. This observation highlights the metabolome's sensitivity to external factors. A discussion of the differential pressures that *E. coli* dietary strains exude on *C. elegans* metabolism will be presented in Chapter 3. Yeast also exhibits an appreciable amount of biological variation despite being grown under well-defined and controlled conditions. As all experiments were carried out with great care and consistency, we propose that the variation in yeast samples may be primarily due to stochastic gene expression (Raser and O'Shea 2005).

Technical or analytical variation is an important consideration in metabolomics studies and can be a significant contributing factor to the observed metabolic variation. Sources of this variation include instrumentation, sample preparation, and NMR data analysis. In our studies, we examined the contributions of technical variation to the observed variation using two approaches. First, we manually fitted and quantified the DSS concentration reference peak in all of the yeast exometabolite samples. DSS is added at a constant concentration to all samples and variations in the quantified values are a reflection of the precision of the NMR spectrometer. Since these spectra were collected at various times over a number of days, variations in DSS peak values may reflect subtle changes in the instrument that occur during operation. The average RSD value for the DSS peaks from the spectral dataset was 2.1% (data not shown). The second

approach involved quantifying the metabolite profile from a yeast exometabolite sample prepared in triplicate, with the ^{1}H-NMR spectra from these samples being collected over the course of three consecutive days. The average RSD value for the metabolites quantified from these profiles was 3.8% with a range of 0.8 to 9.1% (data not shown) and these values are significantly lower than that observed for any of the yeast exometabolite datasets. Our findings are consistent with those in several past studies that indicate that technical variability is not a significant contributing factor to the variation observed in the model systems examined (Fiehn et al. 2000; Crews et al. 2009; Parsons et al. 2009). Our results reaffirmed the conclusion by Parsons *et al.* that quantitation using 1D NMR offers excellent analytical precision. In their comparative study of a number of analytical approaches, 1D NMR generated data with the lowest RSD values (median RSD of 3.1%) out of all the analytical approaches examined (Parsons et al. 2009).

We also examined whether biological variation occurs between yeast strains with very similar genetic backgrounds. In previous studies, the complemented SDH strains were considered phenotypically wild-type and identical (Szeto et al. 2007, 2010). However, multivariate analysis using PLS-DA discriminated between these two strains (Figure 3A). A similar situation is also observed for the comparison between the two knockout strains (Figure 3B). To rule out the possibility of spurious model generation, PCA and model validations were performed (Figures 4 and 5). The results suggest the models are valid. It has been demonstrated previously that metabolic profiling can discriminate between yeast mutants that are otherwise phenotypically indistinguishable (Allen et al. 2003; Szeto et al. 2010). The results presented in this study further substantiate this notion. There may be several reasons why these pairs of yeast strains are

metabolically different. In the case of the complemented SDH strains, there may be subtle differences in the expression of the respective SDH subunits. Upstream or downstream regulatory elements may not have been cloned into the respective plasmids, affecting the plasmid-borne expression of the subunits. Another factor could be plasmid copy number. The yeast strains were each transformed with centromeric plasmids that are maintained at 1-2 copies per cell; however, precise copy number is subject to stochastic variation as cells divide. Increased heterogeneity of gene expression was observed in *E. coli* cells carrying a plasmid-borne copy of the *lacI* gene compared to expression from the chromosomal copy (Elowitz et al. 2002). Subtle changes in the activity of an enzyme can be amplified into larger changes in the metabotype, as we have shown previously (Szeto et al. 2010). The observation that the metabotypes of the knockout strains could also be discriminated is a surprising result considering that both knockouts result in the complete absence of SDH (Oyedotun and Lemire 1997; Oyedotun and Lemire 1999; Szeto et al. 2007). The different metabotypes of the two knockouts suggest that the Sdh3p and Sdh4p subunits may have additional roles beyond their presence in SDH. It was recently demonstrated that the deletion of *SDH3* results in transcriptional changes associated with fatty acid and sterol metabolism despite being in glucose-repressed conditions. This evidence suggests that nuclear gene signaling is responsive to TCA cycle dysfunction even in situations when this pathway is underutilized (Cimini et al. 2009).

In this study, we examined whether the biological variation observed in the metabolomic studies of other organisms is also observed in simpler model systems, such as yeast and nematodes. Both model systems reveal appreciable

amounts of biological variation, with *C. elegans* showing greater metabolic variance. The variation in *C. elegans* was also modulated by diet. We also examined the biological variation between phenotypically identical yeast strains and were able to discriminate between them by their metabotypes using multivariate analysis. Our results highlight the sensitivity of the metabolome, even in simpler model systems grown under controlled conditions, and the usefulness of metabolic profiling to determine these subtle differences. They also emphasize the importance of appropriate statistical analysis for interpreting data.

Limitations of Study

This study explores the amount of variance found in *C. elegans* and *S. cerevisiae* metabolomes; variance was based on RSD values. This study poses a number of limitations. First, univariate statistics were not performed to test whether RSD values for individual metabolites were statistically different between classes. Second, the RSD ranges for all classes of both model systems were presented together in Figure 2-2. We do not intend to imply that the data should be compared between model systems as they are derived from different metabolomes. Third, the biological reasons and relevance for the differential variability between compared groups is unknown and not explored. This issue could only be explored if transcriptomic or metabolic tracer studies were performed.

Tables and Figures

Table 2-1. Summary of the metabolomic data sets used in this study.

Metabolite Dataset	Metabolome type	No. of metabolites quantified	Average %RSD	Range of %RSD
C. elegans on OP50	Intracellular	26	39	18 - 64
C. elegans on HT115 L4440	Intracellular	26	29	17 - 46
S. cerevisiae SDH3 WT	Extracellular	33	12	9 - 24
S. cerevisiae SDH4 WT	Extracellular	33	12	5 - 36
S. cerevisiae SDH3 KO	Extracellular	33	8	4 – 21
S. cerevisiae SDH4 KO	Extracellular	33	11	5 - 25

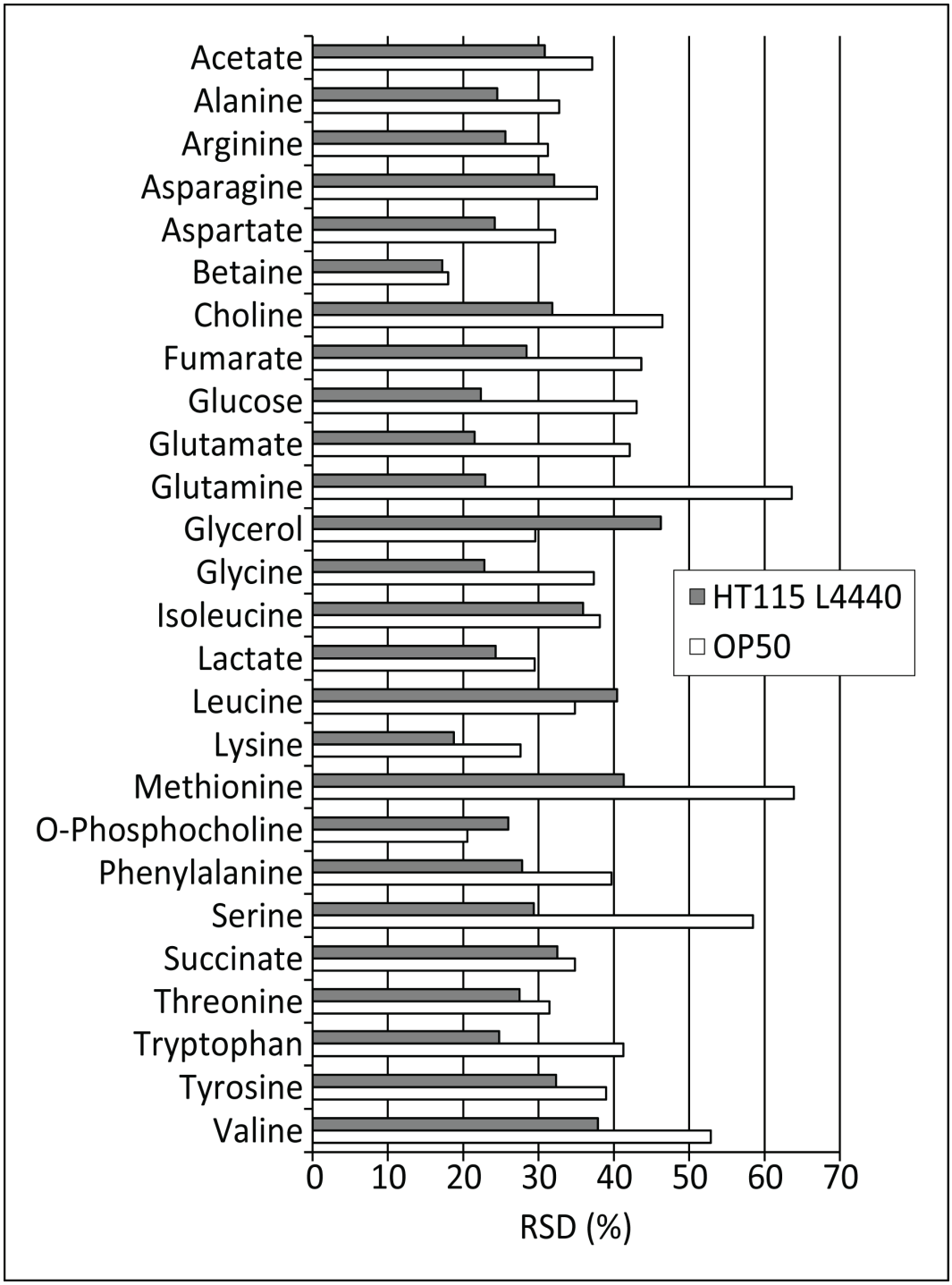


Figure 2-1A. RSD values determined for the identified metabolites. Values were determined from wild-type, whole worm, protein-free lysates from worms fed *E. coli* HT115 L4440, grey (N=6), or fed *E. coli* OP50, white (N=7).

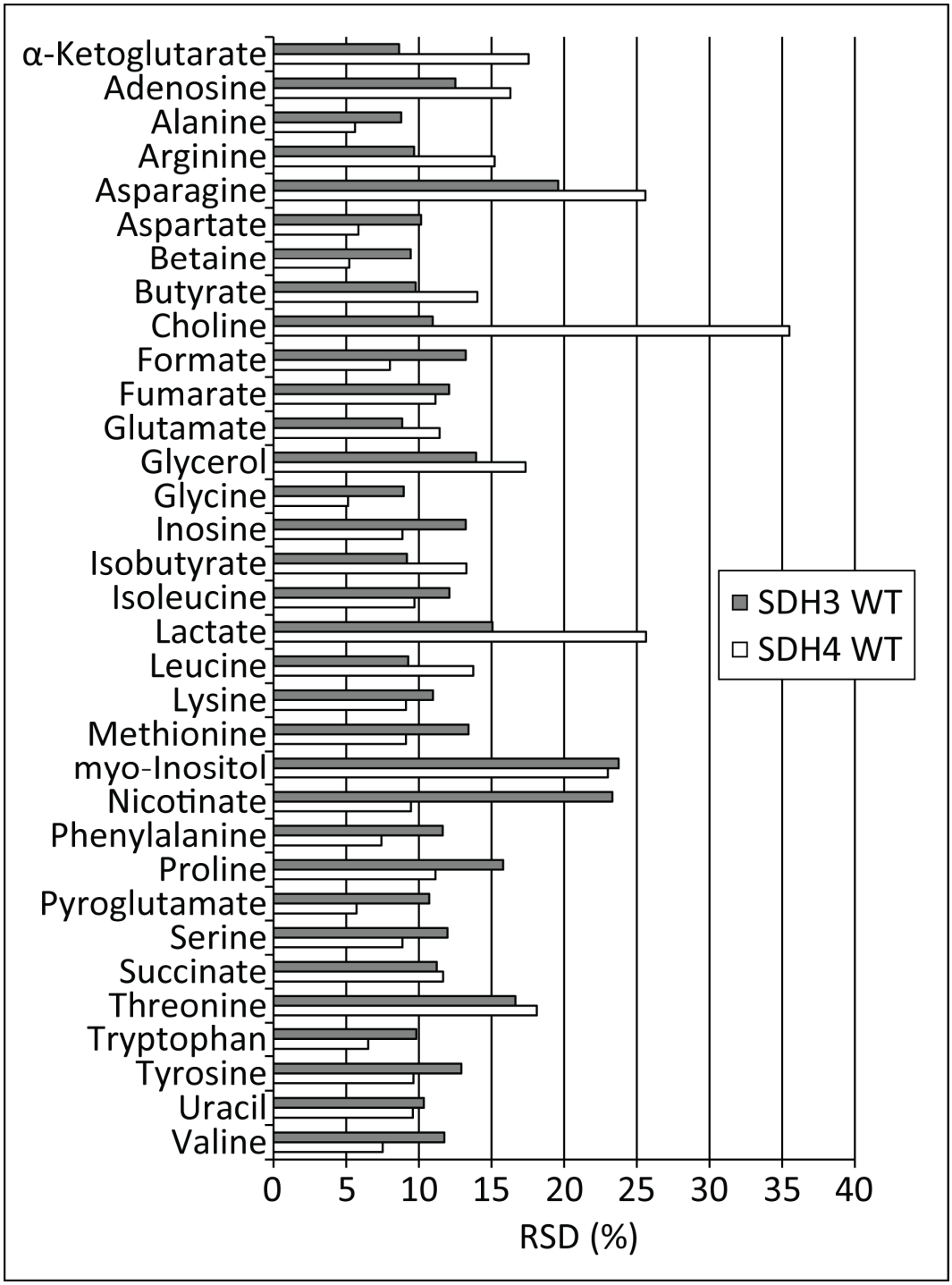


Figure 2-1B. **RSD values determined for the identified metabolites.** Values were determined from the exometabolome of the complemented SDH3, grey (N=12), and *SDH4*, white (N=12) yeast strains.

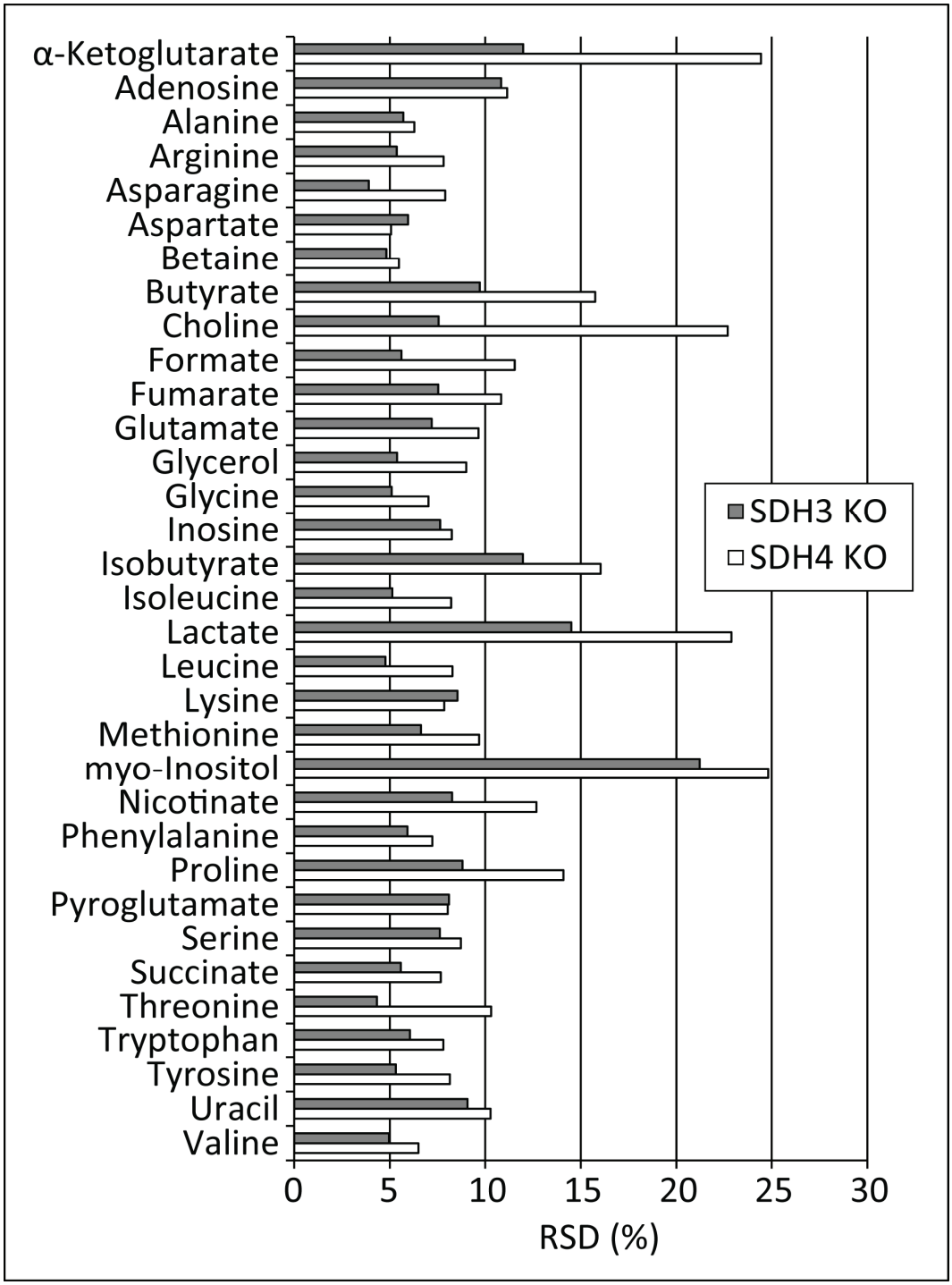


Figure 2-1C. **RSD values determined for the identified metabolites.** Values were determined from the exometabolome of the knockout SDH3, grey (N=12), and *SDH4*, white (N=12) yeast strains.

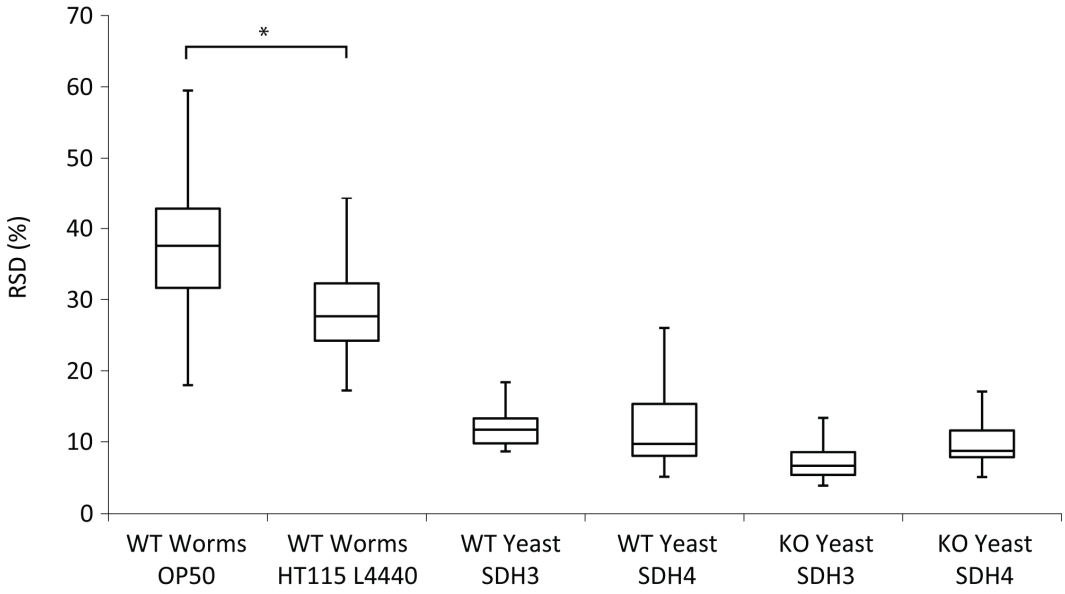


Figure 2-2. Box plot of RSD values for metabolomic datasets described in this study. Values for lower quartile, median and upper quartile are shown. Error bars show the range of data. *, $p < 0.001$.

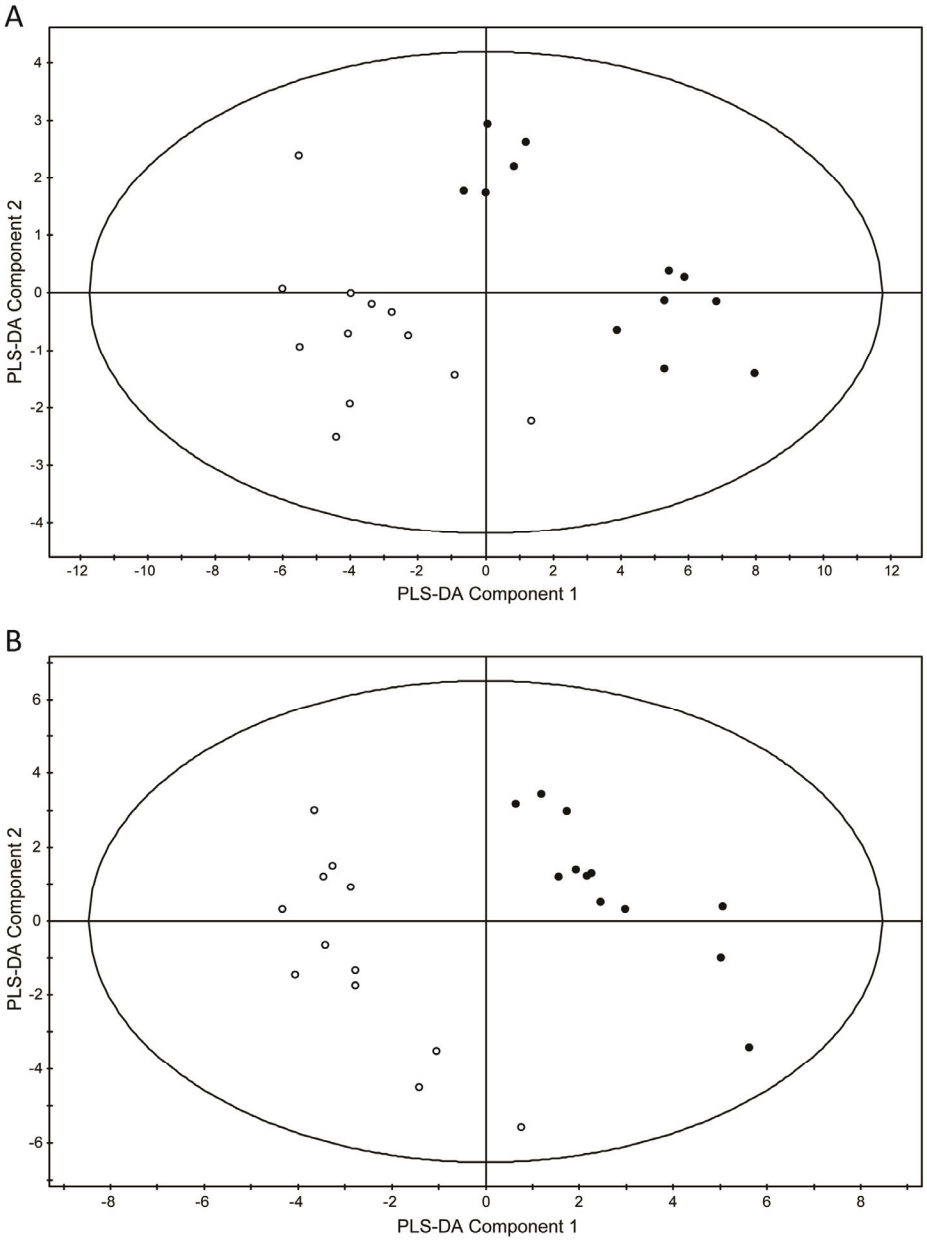


Figure 2-3. Two component PLS-DA models of ^{1}H NMR derived exometabolome profiles of *S. cerevisiae* strains. PLS-DA score scatter plots and validations were performed using SIMCA P + v12.0.1 software. A, PLS-DA score plot for the complemented yeast strains. Black dots, *SDH3*; white dots, *SDH4*. The model is characterized by the following parameters using 2 components: R^2X_1, 0.606, R^2X_2, 0.0819, R^2X_{cum}, 0.688, R^2Y_{cum}, 0.872, Q^2_{cum}, 0.8. N =12 for both strains, CV-ANOVA p=1.69 x 10^{-6}. B, PLS-DA score plot for the yeast deletion strains. Black dots, *SDH3* KO; white dots, *SDH4* KO. The model is characterized by the following parameters using 3 components: R^2X_1, 0.33, R^2X_2, 0.184, R^2X_{cum}, 0.568, R^2Y_{cum}, 0.98, Q^2_{cum}, 0.913. N = 12 for both strains, CV-ANOVA p=5.67 x 10^{-8}. Each dot represents one yeast culture.

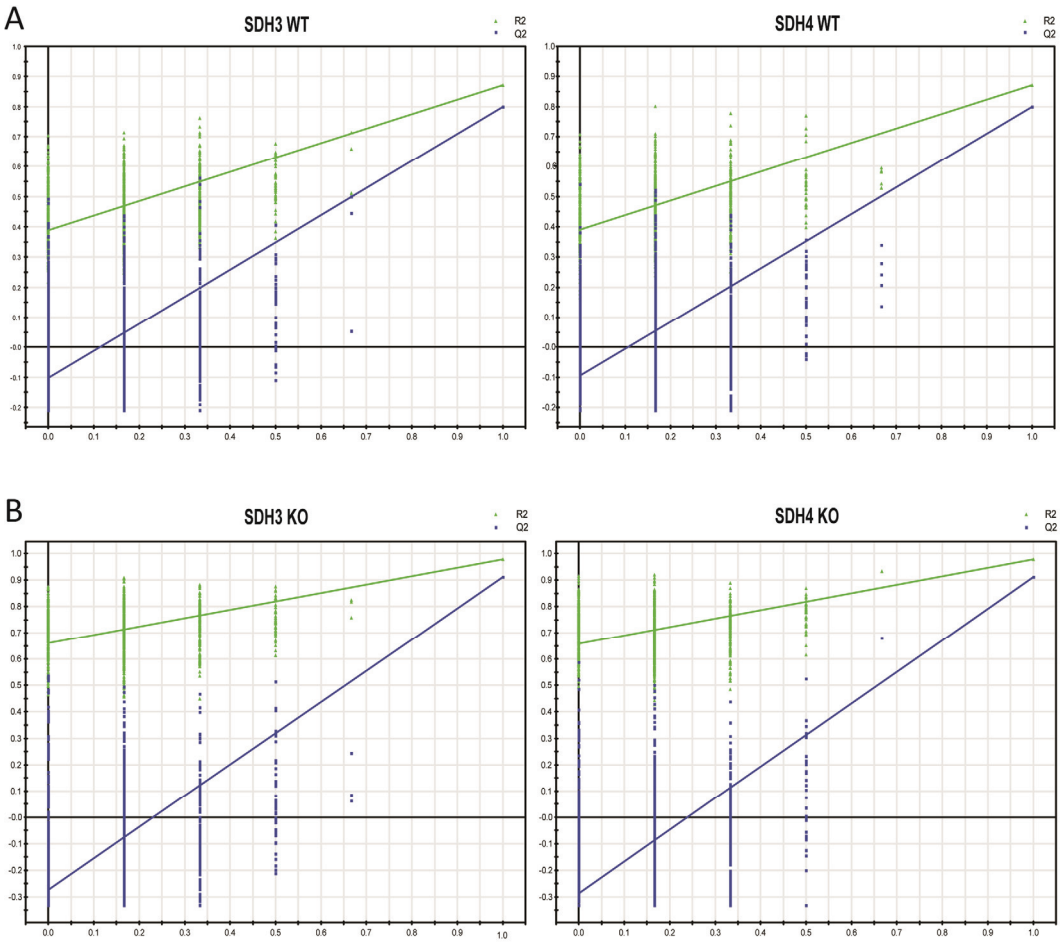


Figure 2-4. Validation plots for the complemented and knockout strain models. A, Validation model using 999 permutations across 2 components. B, Validation model using 999 permutations across 3 components. The regression line represents the correlation coefficient between the original and permuted *Y*-variables against cumulative R^2 and Q^2. Green, R2; blue, Q2.

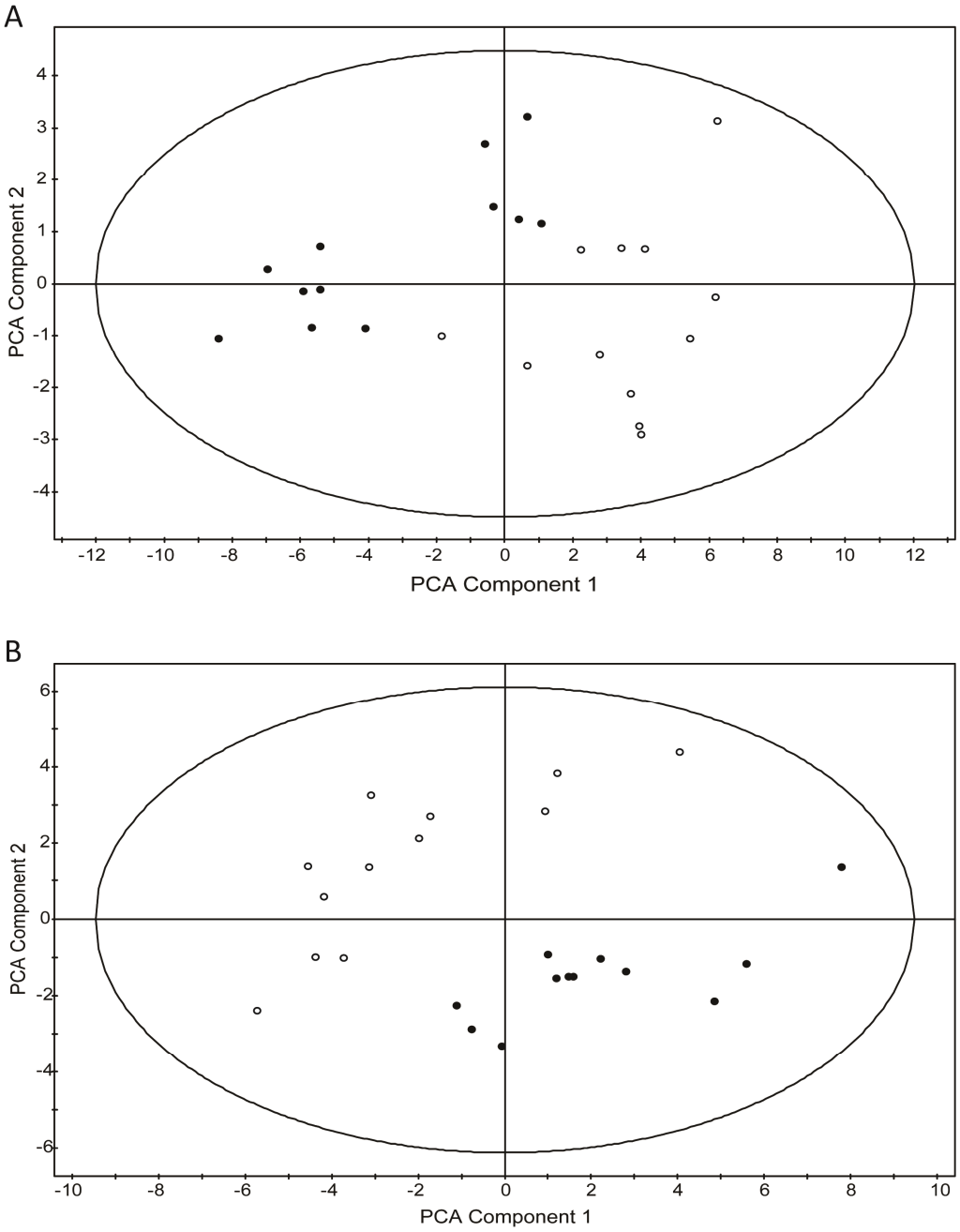


Figure 2-5. Two component PCA models of ^{1}H NMR derived exometabolome profiles of *S. cerevisiae* strains. PCA score scatter plots were performed using SIMCA P + v12.0.1 software. A, PCA score plot for the complemented yeast strains. Black dots, *SDH3*; white dots, *SDH4*. The model is characterized by the following parameters using 2 components: R^2X_1, 0.609, R^2X_2, 0.0844, R^2X_{cum}, 0.694, Q^2_{cum}, 0.516. $N = 12$ for both strains. B, PCA score plot for the yeast deletion strains. Black dots, *SDH3* KO; white dots, *SDH4* KO. The model is characterized by the following parameters using 3 components: R^2X_1, 0.366, R^2X_2, 0.153, R^2X_{cum}, 0.608, Q^2_{cum}, 0.314. $N = 12$ for both strains. Each dot represents one yeast culture.

References

Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nat Biotechnol 21 (6):692-696

Braeckman BP, Houthoofd K, Vanfleteren JR Intermediary Metabolism (February 16, 2009), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.146.1, http://www.wormbook.org

Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77 (1):71-94

Cann AJ (2003) Maths from scratch for biologists. John Wiley & Sons, Chichester, England

Cimini D, Patil K, Schiraldi C, Nielsen J (2009) Global transcriptional response of *Saccharomyces cerevisiae* to the deletion of *SDH3*. BMC Syst Biol 3 (1):17

Corder G, Foreman D (2009) Nonparametric statistics for non-statisticians: a step-by-step approach. John Wiley & Sons, Hoboken, New Jersey

Crews B, Wikoff WR, Patti GJ, Woo HK, Kalisiak E, Heideker J, Siuzdak G (2009) Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. Anal Chem 81 (20):8538-8544

de Bono M, Bargmann CI (1998) Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. Cell 94 (5):679-689

de Kok J, Muller JLM, Slater EC (1975) EPR studies on the respiratory chain of wild-type *Saccharomyces cerevisiae* and mutants with a deficiency in succinate dehydrogenase. Biochim Biophys Acta 387:441-450

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297 (5584):1183-1186

Eriksson LJE, Kettaneh-Wold N, Trygg J, Wikström C, Wold S (2001) Multi- and megavariate data analysis. Part 1: principles and applications. 2nd edn. Umetrics academy, Umeå, Sweden

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18 (11):1157-1161

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J (2000) Functional genomic analysis of *C. elegans*

chromosome I by systematic RNA interference. Nature 408 (6810):325-330

Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic acids research 20 (6):1425

Grad LI, Sayles LC, Lemire BD (2007) Isolation and functional analysis of mitochondria from the nematode *Caenorhabditis elegans*. In: Leister D, Herrmann J (eds) Mitochondria: practical protocols. Humana Press, Totowa, N.J., pp 51-66

Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. Cell 134 (5):714-717

Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J (2001) Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. Genome Biol 2 (1):research0002.0001-0010

Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF (2003) A ^{1}H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. Journal of pharmaceutical and biomedical analysis 33 (5):1103-1115

Lewis JA, Fleming JT (1995) Basic culture methods. Methods Cell Biol 48:3-29

Massart D, Smeyers-Verbeke J, Capron X, Schlesier K (2005) Visual presentation of data by means of box plots. LC•GC Eur 18 (4):215-218

O'Riordan VB, Burnell AM (1989) Intermediary metabolism in the dauer larva of the nematode *Caenorhabditis elegans* - 1. Glycolysis, gluconeogenesis, oxidative phosphorylation and the tricarboxylic acid cycle. Comp Biochem Physiol Part B: Biochem Mol Biol 92 (2):233-238

O'Riordan VB, Burnell AM (1990) Intermediary metabolism in the dauer larva of the nematode *Caenorhabditis elegans* - II. The glyoxylate cycle and fatty-acid oxidation. Comp Biochem Physiol Part B: Biochem Mol Biol 95 (1):125-130

Oyedotun KS, Lemire BD (1997) The carboxyl terminus of the *Saccharomyces cerevisiae* succinate dehydrogenase membrane subunit, SDH4p, is necessary for ubiquinone reduction and enzyme stability. J Biol Chem 272 (50):31382-31388

Oyedotun KS, Lemire BD (1999) The *Saccharomyces cerevisiae* succinate-ubiquinone oxidoreductase. Identification of Sdh3p amino acid residues involved in ubiquinone binding. J Biol Chem 274 (34):23956-23962

Parsons HM, Ekman DR, Collette TW, Viant MR (2009) Spectral relative standard deviation: a practical benchmark in metabolomics. Analyst 134 (3):478-485

Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. Science 309 (5743):2010-2013

Reinke SN, Hu X, Sykes BD, Lemire BD (2010) *Caenorhabditis elegans* diet significantly affects metabolic profile, mitochondrial DNA levels, lifespan and brood size. Mol Genet Metab 100 (3):274-282

Saude E, Adamko D, Rowe B, Marrie T, Sykes B (2007) Variation of metabolites in normal human urine. Metabolomics 3 (4):439-451

Solanky KS, Bailey NJ, Beckwith-Hall BM, Davis A, Bingham S, Holmes E, Nicholson JK, Cassidy A (2003) Application of biofluid ^{1}H nuclear magnetic resonance-based metabonomic techniques for the analysis of the biochemical effects of dietary isoflavones on human plasma profile. Anal Biochem 323 (2):197-204

Stoop MP, Coulier L, Rosenling T, Shi S, Smolinska AM, Buydens L, Ampt K, Stingl C, Dane A, Muilwijk B, Luitwieler RL, Sillevis Smitt PA, Hintzen RQ, Bischoff R, Wijmenga SS, Hankemeier T, van Gool AJ, Luider TM (2010) Quantitative proteomics and metabolomics analysis of normal human cerebrospinal fluid samples. Mol Cell Proteomics 9 (9):2063-2075

Szeto SS, Reinke SN, Sykes BD, Lemire BD (2007) Ubiquinone-binding site mutations in the Saccharomyces cerevisiae succinate dehydrogenase generate superoxide and lead to the accumulation of succinate. J Biol Chem 282 (37):27518-27526

Szeto SS, Reinke SN, Sykes BD, Lemire BD (2010) Mutations in the *Saccharomyces cerevisiae* succinate dehydrogenase result in distinct metabolic phenotypes revealed through ^{1}H NMR-based metabolic footprinting. J Proteome Res 9 (12):6729-6739

Walsh MC, Nugent A, Brennan L, Gibney MJ (2008) Understanding the metabolome – challenges for metabolomics. Nutrition Bulletin 33 (4):316-323

Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM (2006) Targeted profiling: quantitative analysis of ^{1}H-NMR metabolomics data. Anal Chem 78 (13):4430-4442

Zuppi C, Messana I, Forni F, Ferrari F, Rossi C, Giardina B (1998) Influence of feeding on metabolite excretion evidenced by urine ^{1}H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svaldbard). Clinica chimica acta; international journal of clinical chemistry 278 (1):75-79

CHAPTER 3

Caenorhabditis elegans diet significantly affects metabolic profile, mitochondrial DNA levels, lifespan, and brood size.

Stacey N Reinke, Xiaobin (Mia) Hu, Brian D Sykes, and Bernard D Lemire

Department of Biochemistry, University of Alberta, Edmonton, Canada

A version of this chapter has been accepted for publication. Reinke SN, Hu X, Sykes BD, and BD Lemire 2010. *Mol Genet Metab*. 100(3): 274-82.

Contributions: SNR performed metabolomic and brood size studies. XH evaluated lifespan and mtDNA copy number. SNR, BDS, and BDL wrote the manuscript.

Introduction

Diet is a major contributor to health and chronic dietary imbalances play a key role in global epidemics of disease (Zivkovic and German 2009). Recent advances in metabolomics offer new approaches to evaluate health and nutritional status. A cell's set of metabolites, its metabolome, is the ultimate end product of gene expression and reflects the effects of gene mutation. The metabolome is also extremely sensitive to factors such as gender, diet, exercise, drug use, and environmental conditions. Individuals vary tremendously in the types and intensities of stresses they experience, resulting in drastically variable metabolic profiles. Metabolism also responds to rapidly changing conditions, such as circadian rhythms and feeding patterns (Gu et al. 2007; Lenz et al. 2004; Lenz et al. 2003; Walsh et al. 2006; Zuppi et al. 1998). As such, the metabolome is in a constant dynamic state. The metabolome not only varies between individuals but also within an individual over time (Saude 2007). Metabolomic studies will offer novel insight into genetic and environmental impacts on health. However, the complexity of cellular metabolism, combined with human genetic and lifestyle variations have hampered the use of metabolic profiling for understanding diet and disease conditions.

At the heart of aerobic metabolism is the mitochondrial respiratory chain (MRC), which is central not only to energy metabolism but also to redox balance and the metabolism of amino acids, nucleotides, and nitrogen. Mutations or deficiencies in any of the MRC complexes can cause a wide variety of pathological conditions affecting almost any tissue or organ system. The most severe consequences are encephalomyopathies, cardiomyopathies and failure to thrive. Currently, there

are no cures for mitochondrial disease and treatment options remain limited (Smeitink et al. 2006).

As discussed in Chapter 1, the etiology of mitochondrial disease is extremely complex. As a model system, the nematode, *C. elegans* offers advantages for both mitochondrial and metabolomic studies. The MRC of the nematode closely resembles that of humans in both structure and function. Several advantages of using nematodes for metabolic studies have been highlighted in Chapter 1.

In an effort to better understand the pathology of mitochondrial disease, we investigated how the metabolome was affected in a nematode model of complex I dysfunction. Complex I, NADH:ubiquinone oxidoreductase, uses the energy of NADH oxidation to translocate protons from the mitochondrial matrix to the intermembrane space and contributes to the generation of the electrochemical proton gradient used to drive ATP synthesis (Wallace 2005). Complex I dysfunction is the most common form of mitochondrial disease (Scaglia et al. 2004). In previous work, we characterized point mutations in the *nuo-1* gene (Grad and Lemire 2004; Grad et al. 2005; Grad and Lemire 2006; DeCorby et al. 2007). The *nuo-1* gene is the worm ortholog of the human *NDUFV1* gene and encodes the 51-kDa FMN-containing active site subunit of complex I (Tsang et al. 2001). The *nuo-1* A352V and A443F mutations produce complex phenotypes with features that resemble human complex I mutations (Grad and Lemire 2004). In order to better understand their complex phenotypes, we performed metabolomic studies on these two mutants.

We used nuclear magnetic resonance (NMR) spectroscopy to simultaneously and quantitatively analyze dozens of metabolites in wild type and in complex I

mutants. We reproducibly found significant differences between mutant and wild type metabolic profiles, suggesting widespread effects of complex I dysfunction on cellular metabolism. To further explore the metabolic consequences of complex I mutations, we chose to employ RNAi against metabolic genes involved in cellular redox balance. We chose lactate dehydrogenase (LDH), which converts pyruvate to lactate in an NADH-dependent manner and mitochondrial glycerol phosphate dehydrogenase (mGPDH), which is part of the glycerol phosphate shuttle and utilizes the MRC to re-oxidize cytosolic NADH. Complex I deficiencies impair NADH oxidation and we reasoned that RNAi directed against alternative cellular pathways for NADH oxidation, such as LDH and mGPDH should aggravate the metabolic consequences of the deficiencies.

C. elegans is routinely grown with the laboratory strain *E. coli* OP50 as the primary food source (Brenner 1974). OP50 is a uracil auxotroph and only forms a thin bacterial lawn on standard culture plates, allowing for easier visualization of the animals. Another laboratory *E. coli* strain, HT115(DE3) is routinely employed for RNAi studies. HT115 is RNase III-deficient and has an isopropyl-ß-D-thiogalactopyranoside (IPTG)-inducible T7 RNA polymerase. HT115 is usually transformed with the feeding vector L4440, which contains a multiple cloning site flanked by T7 promoters (Kamath et al. 2001). Gene-specific DNA fragments are cloned into the vector and double stranded RNA (dsRNA) can be produced by IPTG induction. Upon comparison of worm metabolic profiles produced by OP50 and HT115(DE3) L4440 diets, we found the two were drastically different.

Despite the widespread use of both OP50 and HT115 diets, a critical evaluation of the metabolic and phenotypic differences they produce is lacking. We

compared the effects of the two *E. coli* diets on lifespan, fertility, mtDNA copy number, and metabolic profile. Our results show diet can significantly affect all of these. Surprisingly, we found that the metabolic effects produced by a dietary change are as profound as those produced by mutations or by RNAi treatments. The results of this study not only emphasize the caution that the *C. elegans* research community must take when designing and controlling experiments, but the global research community in general.

Materials and Methods

Strains

The following *C. elegans* worm strains were used: N2 (Bristol) wild-type; LB25, *nuo-1(ua1)* II, *unc-119(ed-3)* III, *uaEx25*[p016bA352V]; LB27, *nuo-1(ua1)* II, *unc-119(ed-3)* III, *uaEx27*[p016bA443F] (Grad and Lemire 2004). The following *E. coli* strains were used as food sources for culturing worms: OP50, *E. coli* B, uracil auxotroph; HT115(DE3), *F-, mcrA, mcrB, IN(rrnD-rrnE)1, lambda-, rnc14::Tn10*(DE3 lysogen: lacUV5 promoter-T7 polymerase).

Cultures for metabolic studies

Worms were cultured on solid nematode growth medium (NGM) seeded with *E. coli* OP50 as described (Lewis and Fleming 1995) or on NGM containing 1 mM IPTG and 25 μg ml^{-1} carbenicillin seeded with *E. coli* HT115(DE3) transformed with either the empty vector L4440 or with one of the following MRC GeneService clones: T25G3.4 (I-3B12) or F13D12.2 (II-7P20) (Fraser et al. 2000). Cultures were inoculated with synchronized L1 stage worms, grown at 20 ºC and harvested and washed with sterile water. Following a 30 minute incubation to

allow worms to digest bacteria in their gut, three additional washes were performed (Grad et al. 2007). To avoid caloric restriction, worms were monitored to ensure they did not deplete the plates of *E. coli*. Worms on plates beginning to starve were transferred to fresh plates. Cultures included worms of all stages from different generations.

Protein precipitation and assays

Protein-free worm lysates were obtained by sonication of worms in 5% (w/v) trichloroacetic acid on ice. Precipitated protein was recovered by centrifugation at 14,000 x *g* at 4 ºC for 12 minutes and quantified (Grad et al. 2007). Supernatants were adjusted to pH 7 with NaOH, clarified by centrifugation and lyophilized for NMR analysis.

^{1}H-NMR analysis

The entire, dried, protein-free lysate was dissolved in 570 µl D_2O (99.9%; Isotec Inc., Miamisburg, OH) and 30 µl of 5 mM DSS-d_6 (2,2-dimethyl-2-sila 3,3,4,4,5,5-hexadeuteropentane sulfonic acid), used as a chemical shift indicator and concentration standard (Chenomx Inc., Edmonton, AB). Particulate matter was removed by centrifugation and the pH was recorded. 500 µl of the solution were transferred to a 5 mm glass NMR tube. All spectra were acquired at 30 ºC using a 600 MHz Inova NMR spectrometer. The tnnoesy pulse sequence was used (circa Vnmr 6.1B software, Varian Inc.) with an acquisition time of 4s, preacquisition delay of 1s, mixing time of 0.1s, sweep width of 7200 Hz, and 256 transients (Saude et al. 2006). Spectra were analyzed using Chenomx NMR Suite Professional software v5.0. All metabolite concentrations were normalized to protein content.

Multivariate data analysis

Metabolic data were visualized and analyzed using SIMCA P+ v12.0.1 software (Umetrics, Umeå, Sweden). The data were not transformed but were scaled to unit variance and mean-centered. Principal component analysis (PCA) models were created to provide an overview of the data and identify possible outliers. Partial least square discriminant analysis (PLS-DA) models were then created and each model was validated using permutation testing. Differences between models were analyzed by a CV-ANOVA (analysis of variance testing of cross-validated predictive residuals) test.

mtDNA copy number

Total DNA was isolated from L1 worms digested in lysis buffer (0.2 M NaCl, 0.1 M Tris-HCl pH 8.3, 50 mM EDTA, 0.5% SDS) containing 100 µg ml^{-1} proteinase K, extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with ethanol. DNA quantification was performed using a Rotorgene 3000 Cycler (Corbett Research, UK) in 12 µl reaction volumes containing 200 nM of each primer and 3 µl of DNA template in 1X SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich, St. Louis, MO). Reactions were initiated at 94 °C for 30 seconds, followed by 40 cycles of 94 °C for 15 seconds and 60 °C for 1 minute. Primers for mtDNA were for the *ctb-1* gene, which encodes the cytochrome *b* of complex III and for nDNA, the *atp-2* gene, which encodes the ß-subunit of the ATP synthase complex were designed with Primer Express 2.0 software (Applied Biosystems, Foster City, CA). The primer sequences are:

for *ctb-1*, forward 5'-TGAAGCTGACCCTATAATGAGGC-3' and

reverse 5'-CCCTAAGACTTTATTTGGAATAGCAC-3';

for *atp-2*, forward 5'-GCAACGTTCAGAAATGCGCT-3' and

reverse 5'-TGTTTGAGCTGAGGCGGACT-3'.

Standard curves for mtDNA and nDNA were generated using 10-fold serial dilutions (10^3 to 10^7 copies) of DNA plasmids containing cloned *ctb-1* or *atp-2* genes.

Phenotypic analyses

All phenotypic analyses were performed at 20ºC. In the lifespan experiments, the L4 larval stage was considered day zero and culture plates contained 50 μM 5-fluoro-2′-deoxyuridine (FUdR) (Sigma-Aldrich, St. Louis, MO) to block the embryonic development of progeny. Survival curves were generated using GraphPad Prism 4 (GraphPad Software, La Jolla CA). The log-rank (Mantel-Cox) and the Gehan-Breslow-Wilcoxon tests were used to evaluate differences in survival between the two groups. Brood sizes were determined as previously described (Grad and Lemire 2004). Total progeny (live plus unhatched) were counted; unfertilized oocytes were not counted.

Results

Metabolic profile changes with diet

Metabolism in the wild-type worm is clearly affected by the food source. When the metabolic profile of worms cultured on HT115, under RNAi inducing conditions, is compared to the profile obtained when cultured on OP50, the concentrations of eleven of twenty nine identified metabolites changed significantly (Figure 3-1). The amino acids aspartate, glutamate, and lysine were all increased significantly in the HT115 metabolome. Other metabolites that

increased in concentration were betaine, glucose, lactate, and *o*-phosphocholine. Interestingly, formate, oxaloacetate, and propionate were only detectable in the growth on HT115 but not on OP50. The only metabolite to decrease significantly in concentration was acetate. Almost 1/3 of the metabolites analyzed change significantly depending on diet. The metabolic significance of these changes is unclear.

Multivariate data analysis of metabolic profiles

Data sets were not transformed but were mean-centered and scaled to unit variance. PCA was used to identify outliers and to perform an initial inspection of variation in the data set. PLS-DA was used to gain optimal separation between the different diet groups and to perform statistical analysis. Figure 3-2A presents the PLS-DA score scatter plot for the OP50 (stars) and HT115 L4440 (boxes) data sets. In addition, the data sets for the RNAi treatments against lactate dehydrogenase (*ldh-1*; F13D12.2; triangles) and mitochondrial glycerol phosphate dehydrogenase (mGPDH; T25G3.4; circles) are shown. The four data sets are well separated from each other and are characterized by the following parameters using the first six components: R^2X_{cum}, 0.912; R^2Y_{cum}, 0.978; Q^2_{cum}, 0.915. These R^2 and Q^2 values indicate excellent fit and predictive ability, respectively. Figure 3-2B shows model validation using 999 permutations and 6 components. The validation plots correspond with features, highlighted in Chapter 1, that indicate model validity. First, none of the permutations outperform the original (as all blue and green points on the left of the figure are lower than the respective originals at the far right). Second, the Y-intercepts of the blue lines (Q^2) are well below zero. The Y-intercepts of the green lines (R^2) are higher than

optimal (0.4); however, this is likely due to the small sample size. These validations indicate that all four classes are valid within the model. To distinguish significant differences between the groups, a CV-ANOVA test was performed (p=9.8 x 10^{-7}), indicating that the difference between models in Figure 3-2 is significant.

Diet-induced metabolic profile changes are not restricted to the wild-type strain. We also created a PLS-DA model of three different worm strains fed either OP50 or HT115: wild-type and the two *nuo-1* mutants LB25 and LB27 (Figure 3-3A). Animals are scored only by diet: OP50 (circles) or HT115 L4440 (triangles). This plot is characterized by the following parameters using the first three components: R^2X_{cum}, 0.724; R^2Y_{cum}, 0.891; Q^2_{cum}, 0.81. Validation of the model is shown in Figure 3-3B using 999 permutations across 3 components. The features described above indicate validity of the model. A CV-ANOVA test (p=5.7 x 10^{-7}) indicates the two groups are significantly different from each other. This model indicates that metabolic profiles of genetically distinct strains correlate well based solely on diet. The metabolic profiles of the wild-type, LB25, and LB27 cultured on OP50 are shown in Supplementary Figure 3-1. These profiles were analyzed and validated using a PLS-DA model; they are significantly distinct from each other (Supplementary Figure 3-2). The metabolic profiles of the strains fed HT115 are shown in Supplementary Figure 3-3. Again, PLS-DA indicates the worm strains have significantly different metabolic profiles (Supplementary Figure 3-4). In the validation plots, some of the permutations outperformed the original, although this is likely due to the small sample sizes. These results show that the metabolic profiles of wild-type, LB25 and LB27 animals are distinct regardless of diet and that the profiles also change significantly with diet.

mtDNA copy number

The mtDNA and nDNA contents of wild-type worms grown on OP50 or on HT115 were determined. The ratio of mtDNA to nDNA was calculated; the ratio in wild type worms cultured on OP50 was set to 1.0. Wild-type worms grown on HT115 have almost 1.4 times the mtDNA:nDNA ratio of those cultured on OP50 (Figure 3-4). A two-tailed unpaired Student's *t* test indicates that the mtDNA:nDNA ratios are significantly different (p=0.03).

Phenotypic analyses

Lifespan analysis and brood size are good indicators of animal fitness. Wild-type worms fed HT115 (dashed) live significantly longer than those fed OP50 (solid; Figure 3-5A). The median survival times are 12 and 13 days for OP50 and HT115, respectively (Mantel Cox $P<0.05$; Gehan-Breslow-Wilcoxon $P<0.001$). Brood size was evaluated for worms on each diet (Figure 3-5B). The average brood sizes for wild-type worms fed OP50 (black) or HT115 (white) were 304 and 322, respectively. These values are not significantly different (p=0.163). We also evaluated the brood size of a *nuo-1* mutant (LB25). The average brood sizes were 125 and 62 for *nuo-1* animals fed OP50 (black) or HT115 (white), respectively. These values are significantly different (p<0.05).

Discussion

Metabolomic studies offer considerable promise for understanding environmental and genetic factors that contribute to health or disease. One such factor is diet. How dietary components are converted into metabolites is affected by genetic

makeup, diet, exercise, and stress (Gu et al. 2007; Lenz et al. 2003; Lenz et al. 2004; Walsh et al. 2006; Zuppi et al. 1998). We chose to use a *C. elegans* model system in order to control and minimize the number of variables that might impact the metabolic profile. *C. elegans* cultures are clonal, largely eliminating genetic background as a variable. Environmental factors such as growth medium, temperature, diet, and culture density can also be controlled.

Our results emphasize the sensitivity of the metabolome to diet. In replicate experiments performed under carefully controlled conditions, we reproducibly obtained similar metabolite concentrations. We were surprised to discover, however, that the metabolic profiles of worms grown on OP50 and HT115 differed so markedly. *C. elegans* is typically grown on plates seeded with *E. coli* OP50, an uracil auxotroph as the food source (Brenner 1974). However, when performing RNAi experiments in which the dsRNA is delivered to the nematodes by feeding, it is conventional to use *E. coli* HT115 carrying a cloned fragment of a *C. elegans* gene to induce RNA interference. When using HT115, IPTG is added to the medium to induce RNA transcription and carbenicillin is added to maintain selection for the plasmid (Kamath et al. 2001; Kamath and Ahringer 2003). At least one third of all the metabolites we quantified in the wild-type worm were significantly changed in concentration between food sources. Multivariate analysis confirms that the two diets do, in fact, produce distinct metabolic profiles within the same worm strain. Furthermore, the two *E. coli* diets produce the distinct metabolic profiles in genetically different strains, and the change in metabolic profile is distinct enough to group strains by diet alone (Figure 3-3A). We also note that the variation between diet-induced metabotypes is as marked as the effects of mutations or RNAi treatments.

The diet-specific metabolic profiles are clearly linked to changes in gene expression. As our results show, diet can have a profound effect on mtDNA copy number, but it is unclear whether diet-dependent metabolic changes are in response to altered mtDNA levels or vice versa. Diet alteration is also sufficient to significantly change phenotype. Lifespan was clearly affected by diet in the wild-type worm. Diet change did not alter the brood size of the wild-type worm strain but did alter that of the *nuo-1* mutant LB25. The wild-type is healthy and may be better able to compensate for any dietary changes. However, the *nuo-1* strain, being metabolically challenged may be incapable of fully compensating for the dietary change.

There are a number of reasons why different strains of *E. coli* could so dramatically affect worm metabolism, mtDNA copy number, and phenotype. First, *E. coli* strains may have differing nutritional values because of their compositions or because of their ease of ingestion and digestion. Strains that are difficult to digest may impose caloric restriction on the worms. However, we note that the growth rates and morphology of strains on OP50 and HT115 are quite similar, suggesting a caloric restriction was not imposed. *C. elegans* can sense specific amino acids, which act as signalling molecules to modulate a starvation response; the levels of those signalling amino acids may differ between *E. coli* strains (Kang and Avery 2009). Second, the two *E. coli* strains are genetically different and will respond in a unique manner to the availability of nutrients from the growth medium. The same growth medium may be permissive for one bacterial strain but impose considerable metabolic stress on another. *E. coli* strains also have significant metabolic heterogeneity (Maharjan and Ferenci 2005). Due to logistical difficulties, the metabolic profiles of the *E. coli* were not

evaluated. Third, *E. coli* is not only a food source; it can also be pathogenic (Darby 2005). Worms are susceptible to opportunistic bacterial gut colonization, particularly as they age. Although our protocol removes bacteria from the worm gut by extensive washing before metabolites are extracted, gut colonization could produce long term effects on nutrient absorption or on stress and immune responses. Fourth, the extent of social feeding may be affected by food source. Fifth, the *E. coli* strains are grown on plates with different additives; the HT115 plates contain IPTG and carbenicillin to induce RNAi. These additions may also contribute to the altered metabolic profiles. We advise caution in comparing metabolomic information collected from nematodes grown on different strains of *E. coli*. Furthermore, it is likely that the phenotypes of well characterized nematode strains may differ considerably between seemingly similar culture conditions.

One way to more rigorously control diet is to grow the worms on an axenic growth medium that contains all essential nutrients and eliminates the need for a bacterial source of food (Szewczyk et al. 2003; Clegg et al. 2002). Axenic growth medium has the added advantage of allowing substitutions with labelled compounds for tracer studies. However, axenic growth medium is considerably more complex and expensive to prepare than traditional NGM seeded with *E. coli.* Growth with this type of medium also eliminates the possibility of utilizing feeding-induced RNAi to investigate worm metabolism.

We believe our work is the first critical evaluation of the effects of OP50 and HT115, as food sources, on nematodes. Our study demonstrates that changing diets can have a profound effect on wild-type and mutant worm strains. It should

serve as a reminder to the *C. elegans* community and to all researchers studying nutrition and metabolism of the sensitivity of the metabolome and of phenotype to diet.

Limitations of Study

To assess the metabolomic differences between worms fed different diets, both univariate and multivariate statistical analyses were performed. We made two primary assumptions in our interpretations of these analyses. First, univariate statistical power was determined using the Student's *t* test. This test assumes that data is normally distributed; a greater sample size would have more accurately indicated data distribution. Second, we assumed that the PLS-DA model presented in Figure 3-3 did not fit on noise, based on R^2 and Q^2 values and on validation testing. However, each class in this model included three different worm strains; the amount of within-class variation is presumably more substantial in this case than if each class represented a single worm strain. As PLS-DA assumes that within-class variation is small compared to between-class variation (see Chapter 1), the possibility exists that our model may have fitted on noise.

Figures

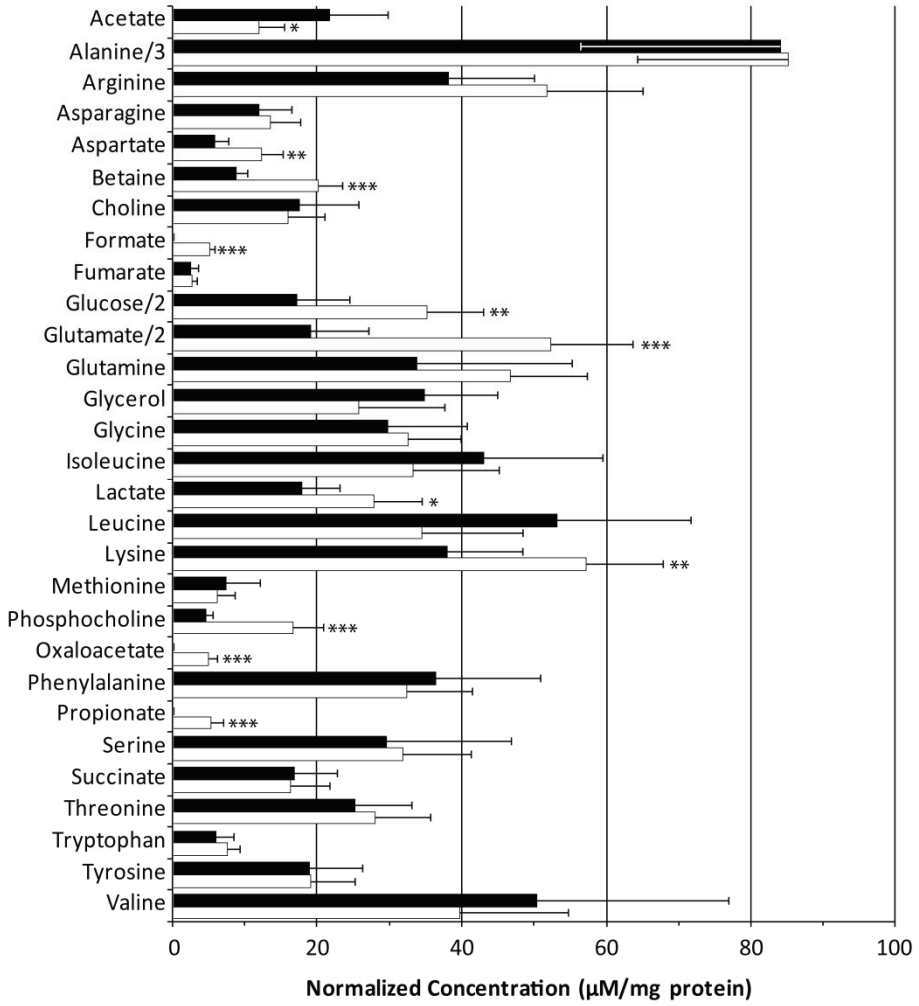


Figure 3-1. Comparison of metabolic profiles of wild-type worms fed OP50 and HT115(DE3) L4440. Metabolite concentrations were quantified using ^{1}H-NMR spectroscopy and normalized to total worm protein. For simplicity of presentation, alanine, glucose, and glutamate concentrations have been divided by 3, 2, and 2, respectively. *Black bars*, OP50 (N=7); *white bars*, HT115(DE3) L4440 (N=6). * p<0.05; ** p<0.01; *** p<0.001 using a two-tailed unpaired Student's *t* test.

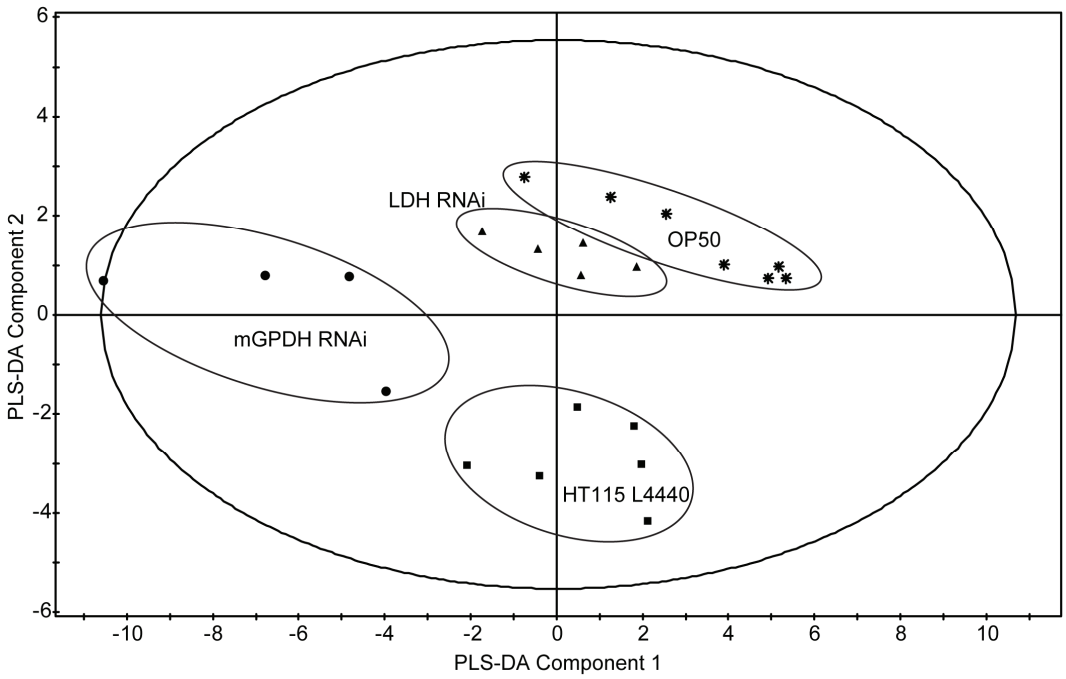


Figure 3-2A. PLS-DA score plot of ^{1}H-NMR derived metabolic profiles of wild-type *C. elegans* cultures fed with either OP50 or HT115(DE3) L4440. The model is characterized by the following parameters: R^2X_1, 0.539, R^2X_2, 0.145, R^2X_{cum}, 0.912, R^2Y_{cum}, 0.978, Q^2_{cum}, 0.915. *Stars*, worms fed OP50 (N=7); *boxes*, worms fed HT115(DE3) L4440 (N=6); *triangles*, LDH RNAi knockdown (N=5); *circles*, mGPDH RNAi knockdown (N=4). Each dot represents one replicate worm culture. Ovals are added solely to draw attention to location of each class and do not reflect any statistical relevance.

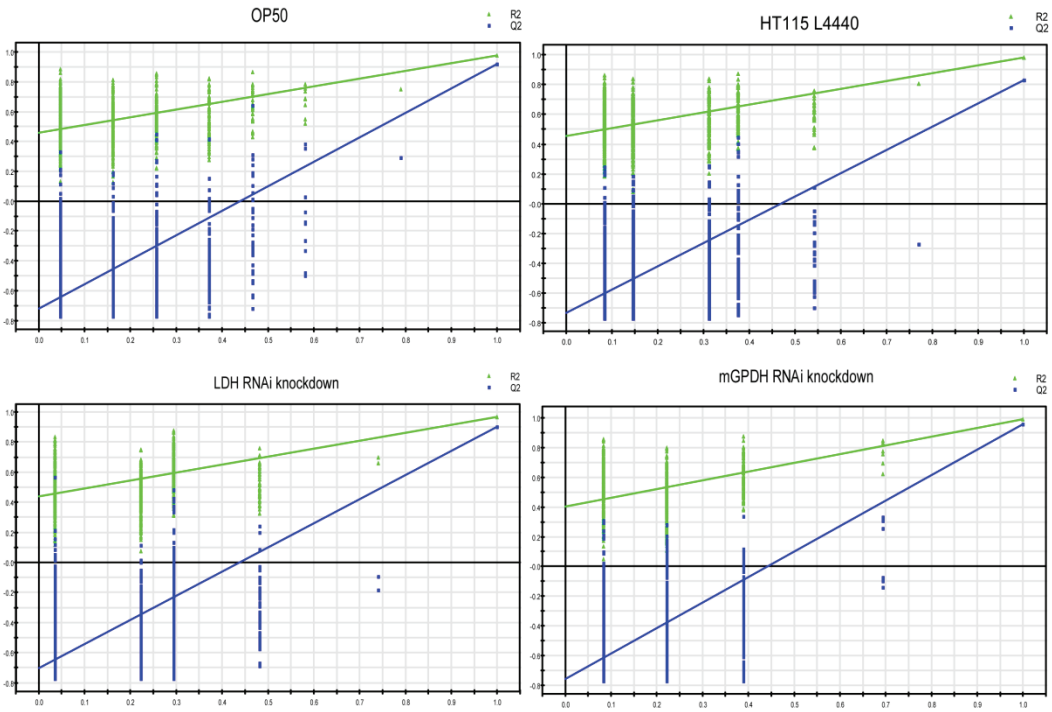


Figure 3-2B. Validation plots for PLS-DA model of ^{1}H-NMR derived metabolic profiles of wild-type *C. elegans* cultures fed with either OP50 or HT115(DE3) L4440. 999 permutations, across 6 components, were used for each of the above diets. *Green*, R^2; *blue*, Q^2.

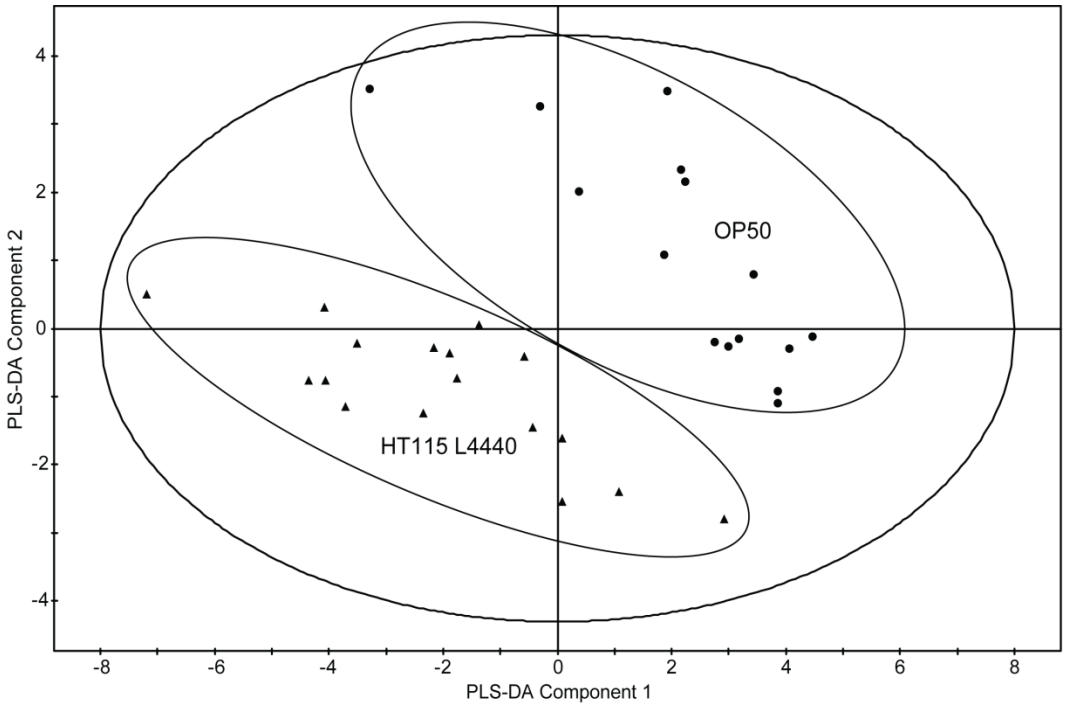


Figure 3-3A. PLS-DA score plot of ^{1}H-NMR derived metabolic profiles of wild-type and two *nuo-1* mutants fed with either OP50 or HT115(DE3) L4440. The PLS-DA model is characterized by the following parameters: R^2X_1, 0.425, R^2X_2, 0.115, R^2X_{cum}, 0.724, R^2Y_{cum}, 0.891, Q^2_{cum}, 0.81. *Circles*, all worms fed OP50 (N=15); *triangles*, all worms fed HT115 L4440 (N=17). Each dot represents one replicate worm culture. Ovals are added solely to draw attention to location of each class and do not reflect any statistical relevance.

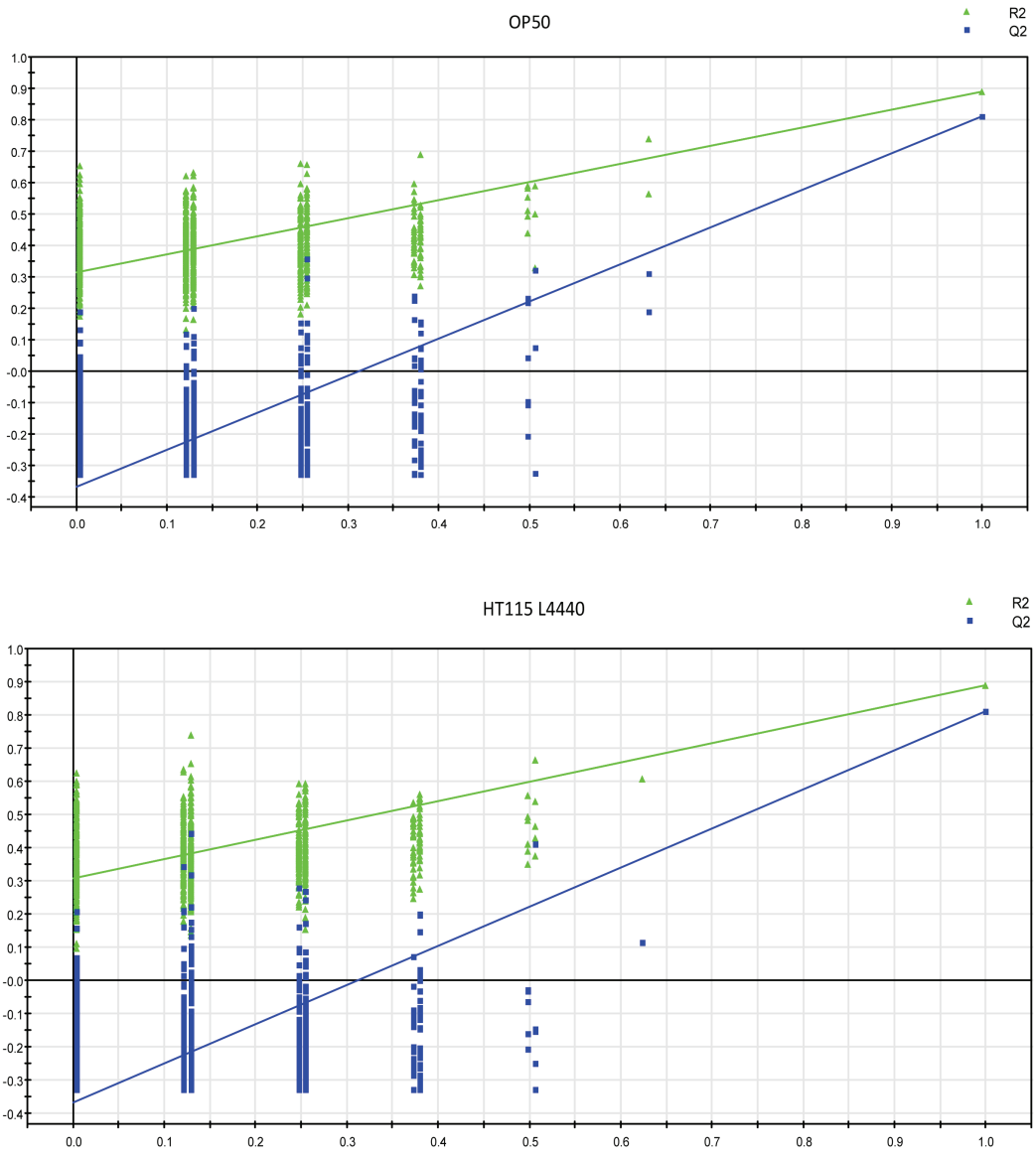


Figure 3-3B. Validation plots for PLS-DA model of ^{1}H-NMR derived metabolic profiles of wild-type and two *nuo-1* mutants fed with either OP50 or HT115(DE3) L4440. 999 permutations, across 3 components, were performed for each of the above diets. *Green*, R^2; *blue*, Q^2.

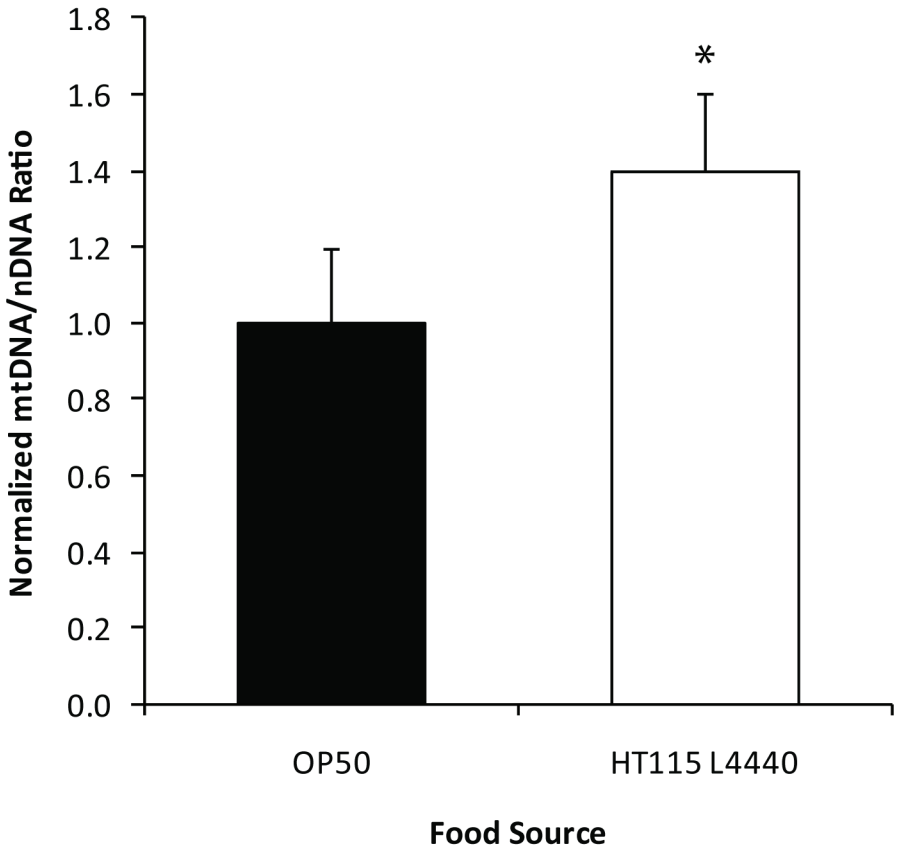


Figure 3-4. Mitochondrial DNA copy numbers. The mtDNA:nDNA ratio of wild-type worms fed OP50 was set to 1.0. The OP50 and HT115(DE3) L4440 values are averages of 6 samples plus 2 replicates, and 8 samples plus 2 replicates, respectively. * $p<0.05$.

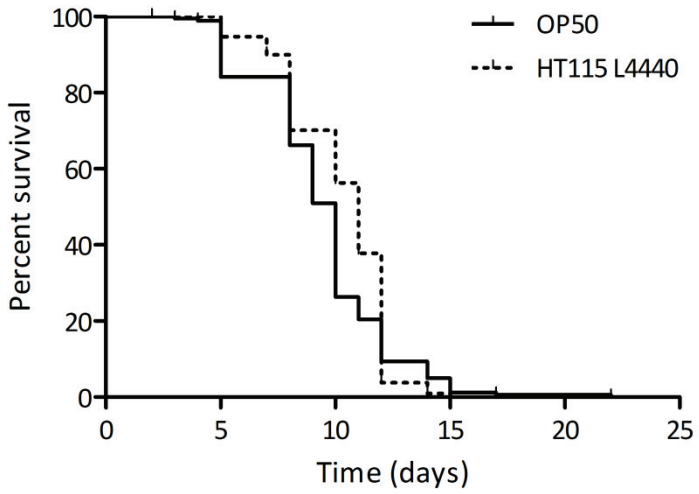


Figure 3-5A. Phenotypic analyses, lifespan. Synchronized, wild type L4-stage hermaphrodites were transferred to plates seeded with OP50 (n=184, solid) or HT115(DE3) L4440 (n=143, dashed) and monitored for survival. Log-rank (Mantel-Cox) analysis, $p<0.05$ and Gehan-Breslow-Wilcoxon analysis, $p<0.001$.

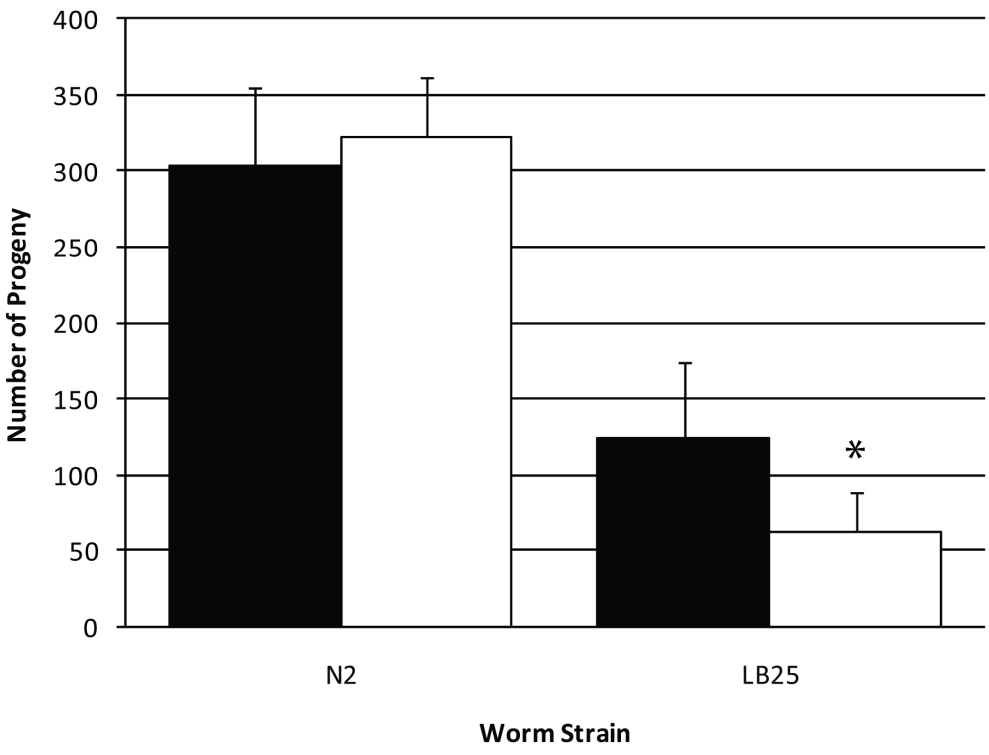


Figure 3-5B. Phenotypic analyses, brood size. Brood sizes were measured for L4-stage wild-type or LB25 hermaphrodites grown on OP50, *black*, or HT115(DE3) L4440, *white*. N=23, 23, 9, and 6 for WT(OP50), WT(L4440), LB25(OP50), and LB25(L4440), respectively. * $p<0.05$.

References

Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77 (1):71-94

Clegg ED, LaPenotiere HF, French DY, Szilagyi M Use of CeHR axenic medium for exposure and gene expression studies. In: East Coast Worm Meeting, Durham, New Hampshire, USA, June 14-16 2002.

Darby C Interactions with microbial pathogens (September 6, 2005) *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi:10.1895/wormbook.1.21.1, http://www.wormbook.org

DeCorby A, Gášková D, Sayles LC, Lemire BD (2007) Expression of Ndi1p, an alternative NADH:ubiquinone oxidoreductase, increases mitochondrial membrane potential in a *C. elegans* model of mitochondrial disease. Biochim Biophys Acta 1767 (9):1157-1163

Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. Nature 408 (6810):325-330

Grad LI, Lemire BD (2004) Mitochondrial complex I mutations in *Caenorhabditis elegans* produce cytochrome *c* oxidase deficiency, oxidative stress and vitamin-responsive lactic acidosis. Hum Mol Genet 13 (3):303-314

Grad LI, Lemire BD (2006) Riboflavin enhances the assembly of mitochondrial cytochrome *c* oxidase in *C. elegans* NADH-ubiquinone oxidoreductase mutants. Biochim Biophys Acta 1757 (2):115-122

Grad LI, Sayles LC, Lemire BD (2005) Introduction of an additional pathway for lactate oxidation in the treatment of lactic acidosis and mitochondrial dysfunction in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 102 (51):18567-18372

Grad LI, Sayles LC, Lemire BD (2007) Isolation and functional analysis of mitochondria from the nematode *Caenorhabditis elegans*. In: Leister D, Herrmann J (eds) Mitochondria: practical protocols. Humana Press, Totowa, N.J., pp 51-66

Gu H, Chen H, Pan Z, Jackson AU, Talaty N, Xi B, Kissinger C, Duda C, Mann D, Raftery D, Cooks RG (2007) Monitoring diet effects via biofluids and their implications for metabolomics studies. Anal Chem 79 (1):89-97

Kamath RS, Ahringer J (2003) Genome-wide RNAi screening in *Caenorhabditis elegans*. Methods 30 (4):313-321

Kamath RS, Martinez-Campos M, Zipperlen P, Fraser AG, Ahringer J (2001) Effectiveness of specific RNA-mediated interference through ingested

double-stranded RNA in *Caenorhabditis elegans*. Genome Biol 2 (1):research0002.0001-0010

Kang C, Avery L (2009) Systemic regulation of starvation response in Caenorhabditis elegans. Genes & development 23 (1):12-17

Lenz EM, Bright J, Wilson ID, Hughes A, Morrisson J, Lindberg H, Lockton A (2004) Metabonomics, dietary influences and cultural differences: a ^{1}H NMR-based study of urine samples obtained from healthy British and Swedish subjects. Journal of pharmaceutical and biomedical analysis 36 (4):841-849

Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF (2003) A ^{1}H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. Journal of pharmaceutical and biomedical analysis 33 (5):1103-1115

Lewis JA, Fleming JT (1995) Basic culture methods. Methods Cell Biol 48:3-29

Maharjan RP, Ferenci T (2005) Metabolomic diversity in the species *Escherichia coli* and its relationship to genetic population structure. Metabolomics 1 (3):235-242

Saude EJ, Adamko, D., Rowe, B.H, Marrie, T., Sykes, B.D. (2007) Variation of metabolites in normal human urine. Metabolomics 3:13

Saude EJ, Slupsky CM, Sykes BD (2006) Optimization of NMR analysis of biological fluids for quantitative accuracy. Metabolomics 2 (3):113-123

Scaglia F, Towbin JA, Craigen WJ, Belmont JW, Smith EO, Neish SR, Ware SM, Hunter JV, Fernbach SD, Vladutiu GD, Wong LJ, Vogel H (2004) Clinical spectrum, morbidity, and mortality in 113 pediatric patients with mitochondrial disease. Pediatrics 114 (4):925-931

Smeitink JA, Zeviani M, Turnbull DM, Jacobs HT (2006) Mitochondrial medicine: a metabolic perspective on the pathology of oxidative phosphorylation disorders. Cell metabolism 3 (1):9-13

Szewczyk NJ, Kozak E, Conley CA (2003) Chemically defined medium and *Caenorhabditis elegans*. BMC Biotechnol 3 (1):19

Tsang WY, Sayles LC, Grad LI, Pilgrim DB, Lemire BD (2001) Mitochondrial respiratory chain deficiency in *Caenorhabditis elegans* results in developmental arrest and increased lifespan. J Biol Chem 276:32240-32246

Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. Annu Rev Genet 39:359-407

Walsh MC, Brennan L, Malthouse JP, Roche HM, Gibney MJ (2006) Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. Am J Clin Nutr 84 (3):531-539

Zivkovic AM, German JB (2009) Metabolomics for assessment of nutritional status. Curr Opin Clin Nutr Metab Care 12 (5):501-507

Zuppi C, Messana I, Forni F, Ferrari F, Rossi C, Giardina B (1998) Influence of feeding on metabolite excretion evidenced by urine ^{1}H NMR spectral profiles: a comparison between subjects living in Rome and subjects living at arctic latitudes (Svaldbard). Clinica chimica acta; international journal of clinical chemistry 278 (1):75-79

CHAPTER 4

Mutations in the *Saccharomyces cerevisiae* succinate dehydrogenase result in distinct metabotypes revealed through ^{1}H-NMR based metabolic footprinting.

Samuel SW Szeto, Stacey N Reinke, Brian D Sykes, and Bernard D Lemire

Department of Biochemistry, University of Alberta, Edmonton, Canada

A version of this chapter has been accepted for publication:

Szeto SSW, Reinke SN, Sykes BD and BD Lemire 2010. *J Proteome Res*. 9:6729-39.

Contributions: SSWS and SNR contributed equally to this work. SSWS and SNR carried out sample preparation, NMR acquisition, and figure preparation. SNR quantified metabolite concentrations. SSWS, SNR, BDS, and BDL wrote the manuscript.

Preliminary results for this chapter are published in:

Szeto SSW, Reinke SN, Sykes BD, and BD Lemire. 2007. *J Biol Chem*. 282:27528-26.

Contributions: SSWS performed all experiments excluding metabolic profiling. SNR performed metabolic profiling experiments. SSWS and BDL wrote the manuscript.

Introduction

Metabolomics offers great promise in understanding how metabolite levels affect phenotype, as the metabolome is thought to be the most sensitive and functional measure of cellular state (Fernie et al. 2004; Raamsdonk et al. 2001; Holmes et al. 2008; Kell et al. 2005). Metabolic perturbations have been reported in several disease states; these metabolic changes can be useful diagnostic and prognostic biomarkers (MacIntyre et al. 2010; Brindle et al. 2002; Griffin and Shockcor 2004; Kenny et al. 2010; Lanza et al. 2010; Jansson et al. 2009). Metabolic footprinting examines the profile of extracellular metabolites, consumed from or excreted into the media by cells; this is referred to as the exometabolome (Kell et al. 2005; Allen et al. 2003; Mapelli et al. 2008). This innovative strategy exploits the notion that changes in the exometabolome are a direct reflection of intracellular metabolic state. Studying the exometabolome is efficient, non-invasive, and not subject to the technical difficulties associated with isolating intracellular metabolites (Kell et al. 2005; Allen et al. 2003). Metabolic footprinting has been applied successfully in metabolic studies of bacteria, yeast, human cell culture, and human waste products (Allen et al. 2003; Mas et al. 2007; Kaderbhai et al. 2003; Shaham et al. 2010; Slupsky et al. 2009).

Mitochondria and particularly the mitochondrial respiratory chain (MRC) play a fundamental role in aerobic metabolism. Although the primary role of the MRC is energy metabolism and the maintenance of redox balance and cellular energy levels, MRC function also influences a multitude of metabolic pathways and cellular processes not directly connected with energy metabolism (Wallace 1999; Scheffler 1999, 2000; Wallace 2005). MRC dysfunction represents the most

common group of inborn errors of metabolism (Thorburn 2004). The clinical presentations are numerous, affecting almost any tissue or organ system and being highly variable in the age of onset and severity (Smeitink et al. 2001). The factors that contribute to this variability are poorly understood, making diagnosis, prognosis, and treatment of mitochondrial disease very challenging, as discussed in Chapter 1. A number of metabolites whose levels are linked to mitochondrial disease have been identified but a consistent and reliable marker of MRC dysfunction is lacking (Mancuso et al. 2009). To facilitate a better understanding of the molecular etiology of mitochondrial diseases and define disease signatures, a more systematic and comprehensive evaluation of the metabolic alterations resulting from MRC dysfunction is needed (Shaham et al. 2010). Recently, Shaham and colleagues addressed these issues by characterizing the exometabolome of spent culture medium, using a cultured cell model of mitochondria disease. Their work demonstrated that the metabolic footprint can provide insight into the biochemical effects of mitochondrial disease; the also identified a new potential biomarker (Shaham et al. 2010).

Succinate dehydrogenase (SDH), or complex II, is a critical enzyme functioning in both the MRC and the tricarboxylic acid (TCA) cycle, forming a direct link between these two essential metabolic processes (Ackrell et al. 1992; Hägerhäll 1997; Hederstedt and Ohnishi 1992). SDH is a tetrameric, iron-sulfur flavoprotein anchored in the mitochondrial inner membrane. It catalyzes the oxidation of succinate to fumarate coupled to the reduction of the lipid soluble electron carrier ubiquinone to ubiquinol. The enzyme is comprised of two domains; the catalytic domain, formed by the *SDHA* and *SDHB*, where succinate oxidation occurs, and the membrane anchor domain, consisting of *SDHC* and *SDHD*, the site of

ubiquinone reduction (Lemire and Oyedotun 2002; Hederstedt and Ohnishi 1992).

As SDH participates in two fundamental metabolic pathways, its dysfunction can manifest in a variety of disease phenotypes. Mutations in the *SDHA* gene, encoding the flavoprotein subunit, can result in several clinical phenotypes, including optic atrophy, ataxia, muscle weakness, myopathy, and Leigh syndrome (an infantile-onset progressive neurodegenerative disorder) (Ackrell 2002; Bourgeron et al. 1995). Mutations in the four SDH genes have also been linked to a variety of cancers including paraganglioma (benign vascularised tumors of the head and neck), pheochromocytoma (tumors of the adrenal medulla), renal cell carcinoma, and colorectal cancer (Rustin et al. 2002; Burnichon et al. 2010; Sugimoto et al. 2000; Rustin and Rötig 2002; Baysal et al. 2001; Habano et al. 2003).

The yeast *Saccharomyces cerevisiae* SDH is structurally and functionally similar to its mammalian counterpart, making it a suitable model for study (Mancuso et al. 2009). The yeast SDH subunit-encoding genes are labelled *SDH1-4*, respectively. In our previous work, we characterized yeast models of point mutations in residues that are associated with tumor formation in humans (Bayley et al. 2006; Schiavi et al. 2006; Neumann et al. 2004; Milunsky et al. 2001; Szeto et al. 2007). We focussed our attention on residues lining the proximal ubiquinone-binding (Q_p) site of the membrane anchor domain. Residues from both membrane anchor subunits contribute to the formation of the Q_p site and are frequently mutated in paragangliomas and pheochromocytomas (Bayley et al. 2005; Sun et al. 2005; Oyedotun and Lemire 2004). We modeled tumorigenic

mutations in the conserved yeast residues, *SDH3* Arg-47 and *SDH4* Asp-88 (Oyedotun and Lemire 2004). *SDH3* Arg-47 was mutated to Cys, Glu, and Lys and *SDH4* Asp-88 to Asn, Glu, and Lys (Szeto et al. 2007). These mutants exhibited variable respiratory growth, which was indicative of mutation severity (Szeto et al. 2007). They also resulted in the accumulation, and excretion, of succinate. We hypothesized that mutation to SDH, a critical enzyme in energy metabolism, should result in global alterations to metabolism and that these changes should be reflected in the exometabolome. Profiling the metabolic changes, resulting from SDH dysfunction, may also reveal additional pathways that are dependent on SDH activity and facilitate a greater understanding of underlying pathogenesis.

In this study, we used a ^{1}H-NMR based metabolic footprinting approach to examine the metabolic profiles of *SDH3* and *SDH4* mutants. We identified and quantified 36 metabolites, including nucleic acid constituents, organic acids, amino acids, and sugars. Our findings show that mutations to SDH result in significant alterations in yeast metabolism, as reflected by changes in the exometabolome. Using multivariate analysis, we were able to discriminate the various yeast strains from each other on the basis of their metabolic phenotypes, or metabotypes. These metabotypes were strongly correlated to mutant respiratory growth yields, identifying a relationship between these two parameters. Our study demonstrates the usefulness of metabolic footprinting in elucidating the biochemical details underlying complex diseases, such as mitochondrial disease and provides insight into the metabolic perturbation resulting from SDH dysfunction.

Materials and Methods

Strains, media, and culture conditions

The *S. cerevisiae* strains sdh3W1 (MH125, *sdh3::TRP1*), sdh4W2 (MH125, *sdh4::TRP1*) and the various site-directed mutants used in this study have been described previously (Oyedotun and Lemire 1997; Oyedotun and Lemire 1999; Szeto et al. 2007). Plasmids were introduced into the yeast strains by lithium acetate-mediated transformation (Gietz et al. 1992). Strains were plated on solid SD medium containing casamino acids (0.5% wt/vol) for plasmid retention and to verify correct phenotypes. The cells were then grown for 36 hrs at 30 ^{o}C in 2mL of YP medium containing 0.25% glucose.

Preparation of exometabolome samples

After completion of culture growth, 100 µl was removed and used to measure the optical density at 600 nm. The remaining culture was centrifuged at 14 000 x *g* for 2 min to pellet the cells and the clarified media were transferred to new microfuge tubes. The samples were subjected to trichloroacetic acid precipitation with a final concentration of 5% and incubated on ice for 30 min. The samples were centrifuged at 14 000 x *g* for 15 min to pellet protein. The supernatants were recovered, adjusted to pH 7.0 with 5M NaOH, flash frozen with liquid N_2 and stored at -80 ^{o}C until they could be lyophilized for 2 days and stored dry at 4 ^{o}C. The lyophilized samples were dissolved in 570 µl D_2O (99.9%; Isotec Inc., Miamisburg, Ohio) along with 30 µl of 5 mM 2,2-dimethyl-2-sila 3,3,4,4,5,5-hexadeutero-pentane sulphonic acid (DSS-d_6, Chenomx Inc., Edmonton, Alberta) as a chemical shift indicator and concentration standard for NMR analysis. The pH was recorded for calibration purposes and samples were centrifuged at

14,000 x *g* for 3 min to remove particulate matter. 510 µl of the supernatants were then transferred to 5 mm diameter NMR tubes.

^{1}H-NMR spectroscopy and NMR data processing

One-dimensional ^{1}H-NMR spectra were acquired on a 600 MHz Varian Inova spectrometer (Varian Inc., Palo Alto, California) at 30 °C using a tn noesy pulse sequence (circa Vnmr 6.1B software, Varian Inc.). A pulse width of 7.95 µs was used. All spectra had an acquisition time of 4 s, a preacquisition delay of 1 s, a mixing time of 0.1 s, a sweep width of 7200 Hz and 256 transients collected (Szeto et al. 2007; Reinke et al. 2010). All spectra were Fourier transformed without line broadening applied, referenced to the DSS-d_6 singlet at 0 ppm, manually phased and baseline-corrected. Chenomx NMR Suite Professional software v4.5 (Chenomx Inc., Edmonton, Alberta) was used for identification and quantification of metabolites by computer-assisted manual fitting of selected peaks.

Metabolite data analysis

Metabolic data were subjected to multivariate statistical analysis using Simca P+ v12.0.1 software (Umetrics, Umeå, Sweden) (Eriksson et al. 2001). Data were not transformed but were mean centered and scaled to unit variance. Data was initially visualized using the unsupervised principal component analysis (PCA) to note any outliers. Data were then treated with the supervised partial least square discriminate analysis (PLS-DA) to gain optimal discrimination between groups. R^2 and Q^2 values gave indication of model goodness of fit and prediction ability. Validations of individual groups within each model were achieved by performing

999 random permutations of y-variables (strain designation) while keeping x-variables (metabolites) intact. The measures of fit (R^2 and Q^2) from the permutations were compared with those of the original model. Additionally CV-ANOVA (analysis of variance testing of cross-validated predictive residuals) tests were performed to determine significant difference between groups in each model. The Pearson product-moment correlation coefficient and level of significance were determined as described (Cann 2003).

Results

Exometabolome alterations in SDH dysfunction

Our previous work characterizing tumorgenic mutations, modeled in the yeast SDH, revealed elevated levels of succinate secreted into the media, a hallmark of SDH dysfunction associated with oncogenesis (Szeto et al. 2007). We hypothesized that mutations to SDH would cause additional perturbations to global metabolism and that these would also be reflected in changes to the exometabolome. To investigate our hypothesis, we employed a ^{1}H-NMR metabolic footprinting approach to examine the exometabolomes of yeast strains containing *SDH3* and *SDH4* mutations. These mutations, in order of decreasing respiratory capacity, are *SDH3* Arg47Lys, Arg47Cys and Arg47Glu and *SDH4* Asp88Glu, Asp88Asn and Asp88Lys (Table 4-1). The respective knock-out strains, transformed with either the corresponding wild-type genes or with empty vector, were also examined. The respiratory capacity of each strain was determined previously using two approaches: growth on solid glycerol-containing medium and growth yield in galactose-containing liquid medium. Glycerol is a non-fermentable carbon source; oxidation of glycerol requires aerobic respiration.

Galactose is a fermentable carbon source and is consumed during an initial fermentative growth phase. Upon exhaustion of galactose *S. cerevisiae* exhibit a diauxic shift to a subsequent respiratory phase (Oyedotun and Lemire 1999, 2001; Oyedotun et al. 2004). We determined the spectral profiles of the acid-soluble extracellular metabolites in the spent culture media of the various *SDH3* and *SDH4* strains using ^{1}H-NMR (Figure 4-1). A metabolite profile of the fresh culture media was also established. A total of 36 were identified using this approach ranging from amino acids, organic acids, nucleotides, and sugars, demonstrating the wealth of biochemical information that can be extracted using this approach.

We compared the metabolite concentrations of the media before and after growth to determine which metabolites were consumed or excreted by the yeast strains during the culturing time. As shown in Figure 4-2, the absolute concentrations of many of the quantified metabolites changed significantly (P<0.05), reflecting their net consumption or excretion by the yeast cells. The overall trends in the changes in metabolite concentration were consistent between both *SDH3* and *SDH4* sets of yeast strains. Two metabolites, glucose and niacinamide, were completely consumed during the incubation time as they were present in the original culture media but absent after growth of the yeast strains. Of the nutrients consumed, the majority were amino acids (alanine, arginine, asparagine, aspartate, leucine, methionine, phenylalanine, proline, serine, threonine, tryptophan, and tyrosine), but betaine, lactate, and uracil were also consumed. We detected a variety of metabolites that were excreted into the medium, these included butyrate, fumarate, succinate, nicotinate, and significant quantities of acetate. Some metabolites were altered in all five strains, while

others were only altered in a certain subset of the strains; these latter changes correlated with respiratory capacity. For example, glycerol was significantly depleted by only wild-type and respiratory-competent strains, consistent with previous results that respiratory-deficient strains are unable to grow on glycerol as a sole carbon source. In contrast, acetate was produced exclusively by respiratory-deficient strains.

To facilitate a comparison of the metabotypes between strains, we normalized the metabolite concentrations to the final optical densities achieved by the culture. These normalized concentrations were compared to those of the corresponding wild-type (Figure 4-3). All of the normalized metabolite concentrations, except for fumarate and isobutyrate, followed a consistent pattern, with their concentrations correlating with mutation severity. The respiration-competent mutants *SDH3* Arg47Lys, *SDH4* Asp88Glu and *SDH4* Asp88Asn had metabolite levels comparable to or slightly higher than wild-types. Significantly higher concentrations were measured for the respiratory-deficient strains. These results show that SDH dysfunction causes reproducible changes in metabolism. These changes are reflected in the exometabolome and are clearly dependent on mutation severity.

Multivariate data analysis

We used multivariate analysis to further analyze and interpret the metabotypes. Principal component analysis (PCA) was used to inspect data variation and identify possible outliers. PLS-DA was subsequently performed to obtain maximal discrimination between profiles and for statistical analysis (Eriksson et al. 2001). The resulting scatter plots reveal that strains can be discriminated based on their

metabolic footprints; the replicate data sets for each strain cluster into discernable groups (Figure 4-4). The groups were separated mainly along the first component within each model. Model performance was assessed using the R^2 and Q^2 values. The *SDH3* model possessed goodness of fit values, R^2X and R^2Y, of 0.96 and 0.759, while the *SDH4* model had values of 0.907 and 0.692, respectively. The Q^2 values were 0.663 and 0.593 for the *SDH3* and *SDH4* models, respectively. These values indicate that the models possess excellent fit and that the data is predictive. Model validations were performed using 999 permutations and five principle components for the *SDH3* model (Supplementary Figure 4-1) and four principle components for the *SDH4* model (Supplementary Figure 4-2). Two aspects of the validation plots indicate model validity: no permutations outperform the original data and the Y-intercepts for the R^2 and Q^2 permutation regression lines are below 0.4 and 0.05, respectively. The validation plots for all of data sets, with the exception of the *SDH4* Asp88Glu mutant, satisfied both criteria and indicate that these models are valid. The validation plot for *SDH4* Asp88Glu mutant had one permutation outperform the original data but satisfied all of the other criteria. To examine statistical significance between the groups, CV-ANOVA tests were performed and resulted in scores of $P = 4.16 \times 10^{-23}$ for the SDH3 strains and $P = 2.19 \times 10^{-20}$ for the *SDH4* strains, indicating that the differences between the groups within each model are highly significant.

Variable importance on the projection (VIP) plots reveal that a large portion of the identified metabolites contributes to discriminating the strains (Figure 4-5, A and B). A VIP score of higher than 1 indicates that the metabolite has an important influence on the separation between strains, for any given component (Eriksson et al. 2001). At least two-thirds of the metabolites contribute greatly to

strain separation along component one; this pattern is unusual but demonstrates the importance of evaluating the entire metabolic profile rather than a few key metabolites. Interestingly, the most discriminating metabolites differed between the *SDH3* and *SDH4* models. For example, acetate ranked as the most important metabolite in the *SDH4* model but its contribution to the *SDH3* model was lower. Consistent with the VIP plots, the biplots reveal that the majority of the metabolites are responsible for the clustering of the groups (Figure 4-5, C and D). Biplots display the X-variable weights superimposed onto the scores, in this case metabolites weights are superimposed onto strain classification scores. Squares are on the plots represent strain and triangles represent individual metabolites. Proximity to the center of the plot is proportional to the variable weight on discrimination. The more distal a variable is from the center, the higher its weight in model discrimination. That is, the spatial proximity of a certain metabolite to a strain reflects its increased relative concentration compared with that of other strains. These biplots show that the respiration deficient mutants and knock-out strains are enriched in the majority of the metabolites identified.

Correlation between metabotype and growth yield

PLS-DA of the data sets revealed a noticeable trend. The component 1 score appeared to correlate with mutant severity. That is, the wild-type strains are on one side of the scatter plots, the knock-out strains are on the other, and the mutants are in between the two extremes with their positions in order of respiratory capacity (Figure 4-4). For the *SDH3* strains, the Arg47Lys data set clustered near the wild-type while the Arg47Cys and Arg47Glu data sets clustered near the knock-out. The Arg47Lys mutant is respiratory-competent,

while Arg47Cys and Arg47Glu are impaired (Table 4-1). For the *SDH4* strains, the Asp88Glu and Asp88Asn data sets were clustered closer to the wild-type while the Asp88Lys data set clustered closer to the knock-out. For the *SDH4* strains, Asp88Glu and Asp88Asn are respiratory-competent and Asp88Lys is respiratory-impaired (Table 4-1). We examined the relationship between strain metabotypes and their growth phenotypes. The biplot score, for each strain, was plotted against its previously determined growth yield using a galactose growth assay (Figure 4-6). We determined the Pearson product-momentum coefficient for the *SDH3* strains (P<0.02) and *SDH4* strains (P<0.05). These results indicate a strong correlation between exometabolome and respiratory-competence.

Discussion

Metabolic footprinting is an innovative method for characterizing a cell's or organism's metabolic status, gaining intracellular metabolic information in a non-invasive manner. Such studies would be potentially very useful in the clinical setting. Exometabolome studies have been successfully performed in a number of model systems including bacteria, yeast, and cultured human cells (Kell et al. 2005; Allen et al. 2003; Kaderbhai et al. 2003; Shaham et al. 2010). In this work, we used high resolution ^{1}H-NMR spectroscopy and multivariate data analysis to examine mitochondrial dysfunction in the yeast model. We focused on the metabolic effects of single amino acid substitutions in the membrane domain of SDH, a key enzyme in the MRC and TCA cycle. We characterized the exometabolomes of these mutants, identifying a number of biochemically relevant molecules (Figure 4-1).

Our results show that single point mutations, in SDH, cause drastic alterations to the cellular metabolic state, which are reflected in the exometabolome (Figure 4-2 and 4-3). The magnitudes of the perturbations correlate with mutation severity, with larger changes exhibited by mutants with reduced respiratory-competence (Figure 4-3). More modest exometabolome alterations were observed for respiration-competent mutants. PLS-DA revealed that individual strains could be discriminated from each other (Figure 4-4). Further analysis of the PLS-DA models, using VIP plots and biplots, showed that at least two-thirds of the metabolites were important in this discrimination (Figure 4-5). Although highly unusual, this pattern highlights the importance of evaluating a more comprehensive metabolic profile, rather than a few key metabolites. We suggest that metabolic footprinting may be a robust method for discriminating between mutants that are otherwise phenotypically indistinguishable (Allen et al. 2003). The most severely impaired mutants, *SDH3* Arg47Glu and *SDH4* Asp88Lys, were easily discriminated from their respective knockout strains by their metabotypes (Figure 4-3), even though the mutants and knockout strains exhibit similar growth phenotypes on fermentable and nonfermentable carbon sources (Table 4-1). Unlike the knockout strains, these severely respiratory-impaired enzymes retain residual SDH activity when examined *in vitro* (Szeto et al. 2007). These low levels of activity may be sufficient to account for the observable metabolic differences between these yeast strains. Our results also show a strong correlation between mutant metabotypes and their respiratory competence (Figure 4-6). Metabotypes can therefore provide functional information on how phenotypes are governed or determined by metabolites (Fernie et al. 2004). Additionally, our results also suggest that characterization of metabotypes may

offer a rapid means of gaining insight into the phenotypes of a new mutation (Figure 4-7).

Metabolic analyses of these SDH mutants also provide insight into the effects of SDH dysfunction on cellular metabolism. The most obvious metabolic change is the presence of large quantities of extracellular acetate found in the most dysfunctional mutant strains (Figure 4-2). Acetogenesis is thought to be a direct consequence of TCA cycle impairment and has been observed in SDH, fumarase and malate dehydrogenase mutants (Romano and Kolter 2005). Upon glucose depletion, yeast cells undergo a diauxic shift to respiration utilizing ethanol (Krukeberg and Dickinson 2004). Acetyl-CoA is synthesized from ethanol, via the pyruvate dehydrogenase bypass pathway, in the cytosol (Boubekeur et al. 1999). Import into the mitochondria is carried out by the carnitine acetyltransferase system (Kispal et al. 1991; Kispal et al. 1993; Pronk et al. 1994). TCA cycle dysfunction will lead to a build-up of Acetyl-CoA in the mitochondrial matrix. *S. cerevisiae* is unable to transport acetyl-CoA out of the mitochondria and it is converted to acetate by acetyl-CoA hydrolase (Buu et al. 2003). Acetate, in its protonated form, can leave the matrix through passive diffusion. Alternatively, an unknown mitochondrial carrier or exchanger may participate in acetate export (Boubekeur et al. 1999; Wolfe 2005; Riviere et al. 2009). Excess acetate is likely excreted from the cell by various monocarboxylate transporters (Casal et al. 2008; Paiva et al. 2004). Interestingly, extracellular acetate levels are a better predictor of SDH dysfunction in these mutants than the succinate to fumarate ratio. In our previous work, all of the SDH mutants had elevated succinate to fumarate ratios compared to the wild-type (Table 4-1). However, these ratios do not correlate with mutation severity. That acetate excretion was only observed for

the most severe mutations suggest that the respiration-competent mutant have sufficient SDH activity to metabolize acetate via the TCA cycle; interestingly, these same mutants also excrete succinate (Szeto et al. 2007; Pollard et al. 2005).

Amino acid concentrations varied greatly between strains. Amino acids, particularly asparagine, serine, and threonine were taken up from the medium by all strains (Figure 4-2). Then normalized to cell growth, we noted that amino acids were not utilized as readily in SDH-deficient strains as in the wild-type strains. Two reasons may explain this observation. First, the mutant strains likely have decreased protein synthesis. Second, amino acids may be used as an energy source upon ethanol depletion but the TCA cycle impairment of the SDH mutants may limit these pathways. Amino acids can provide both carbon and nitrogen sources, except the branched chain and aromatic amino acids, which can only serve as a nitrogen source in yeast (Dickinson 2004; Hazelwood et al. 2008).

Metabolic profiling is emerging as a powerful tool for elucidating the molecular etiologies underlying complex human diseases. In this study, we examined metabolic consequences, of various levels of SDH dysfunction, in the exometabolome. Our findings suggest metabolic footprinting may offer unique advantages in understanding the severity of mutations. Other mitochondrial and metabolic disease models should also be amenable to a metabolomic approach, which may provide additional insight into the pathogenesis of complex diseases.

Limitations of Study

There are three limitations of this study. First, we normalized metabolic data to growth (as reflected by OD); however this may not be the best method. Future metabolomic studies involving yeast should explore other possible methods for normalizing data. Second, we did not present univariate statistical analyses of the data collected in this study. While multivariate analysis presents a holistic model of the data, univariate analysis can also provide valuable information. Third, we measured 36 of the thousands of possible metabolites that are present in yeast. Although we were able to explore the effects of SDH dysfunction on nitrogen and energy metabolism, we were unable to determine what the effects were on other areas of metabolism (such as lipid metabolism).

Tables and Figures

Table 4-1

SDH mutants used in this study and their phenotypic properties [a]

Strain	Growth on minimal glycerol medium	Growth yield on galactose medium [b]	Succinate:Fumarate [c]
***SDH3* WT**	+	100 %	60 ± 10
***SDH3* Arg47Cys**	-	35 ± 4 %	190 ± 30
***SDH3* Arg47Glu**	-	12 ± 1 %	170 ± 20
***SDH3* Arg47Lys**	+	70 ± 6 %	210 ± 30
***SDH3* KO**	-	12 ± 1 %	190 ± 20
***SDH4* WT**	+	100 %	80 ± 10
***SDH4* Asp88Asn**	+	72 ± 9 %	180 ± 20
***SDH4* Asp88Glu**	+	80 ± 7 %	150 ± 40
***SDH4* Asp88Lys**	-	10 ± 1 %	170 ± 30
***SDH4* KO**	-	10 ± 1 %	150 ± 20

[a] Reference is Szeto *et al*. (2007).

[b] Values represent the percentage growths of the yeast strains compared to their respective wild-type controls.

[c] From this study.

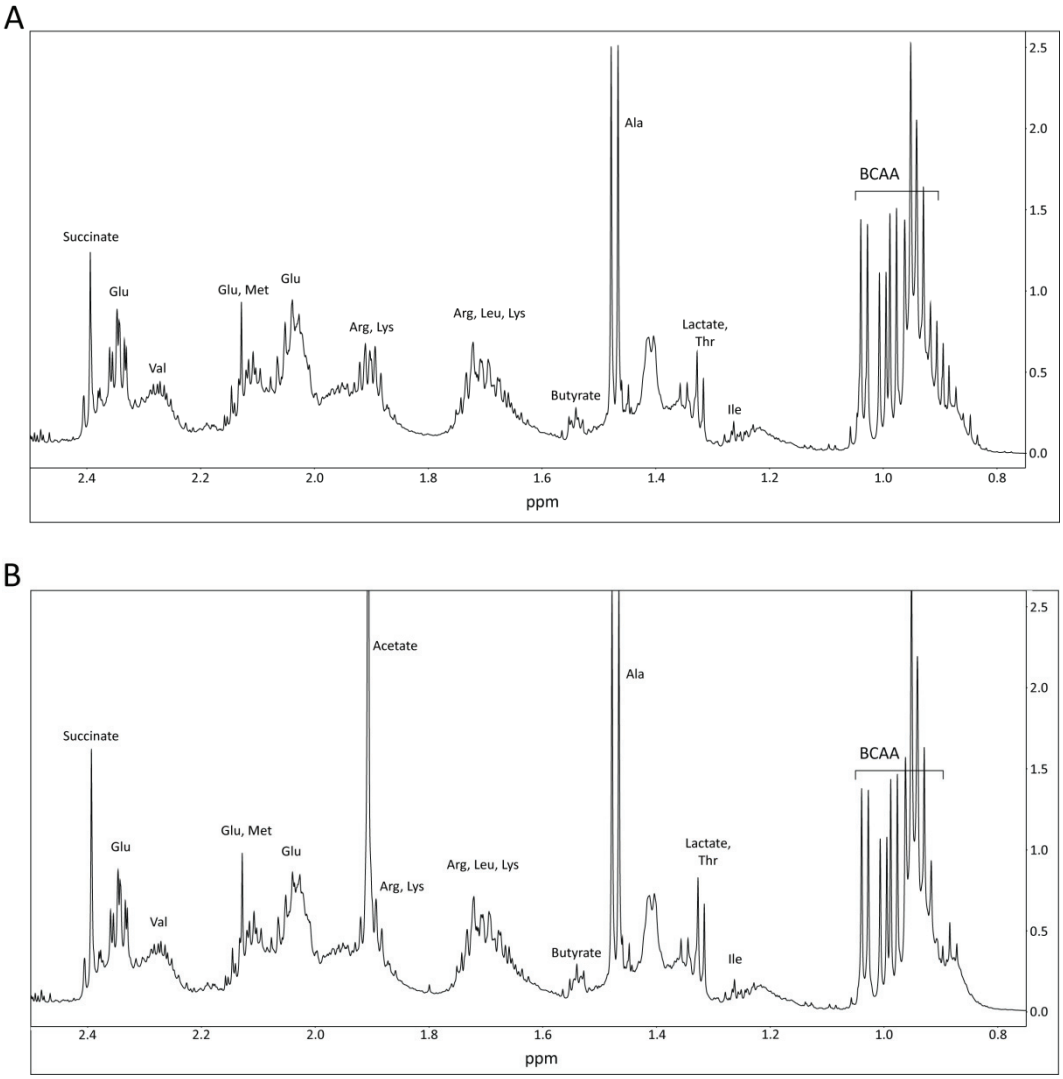


Figure 4-1. ^{1}H-NMR spectra of spent culture media. The aliphatic regions of representative spectra are shown. Key metabolites are labelled. BCAA, branched chain amino acids. A, *SDH 3* wild-type; B, Δ*SDH3*.

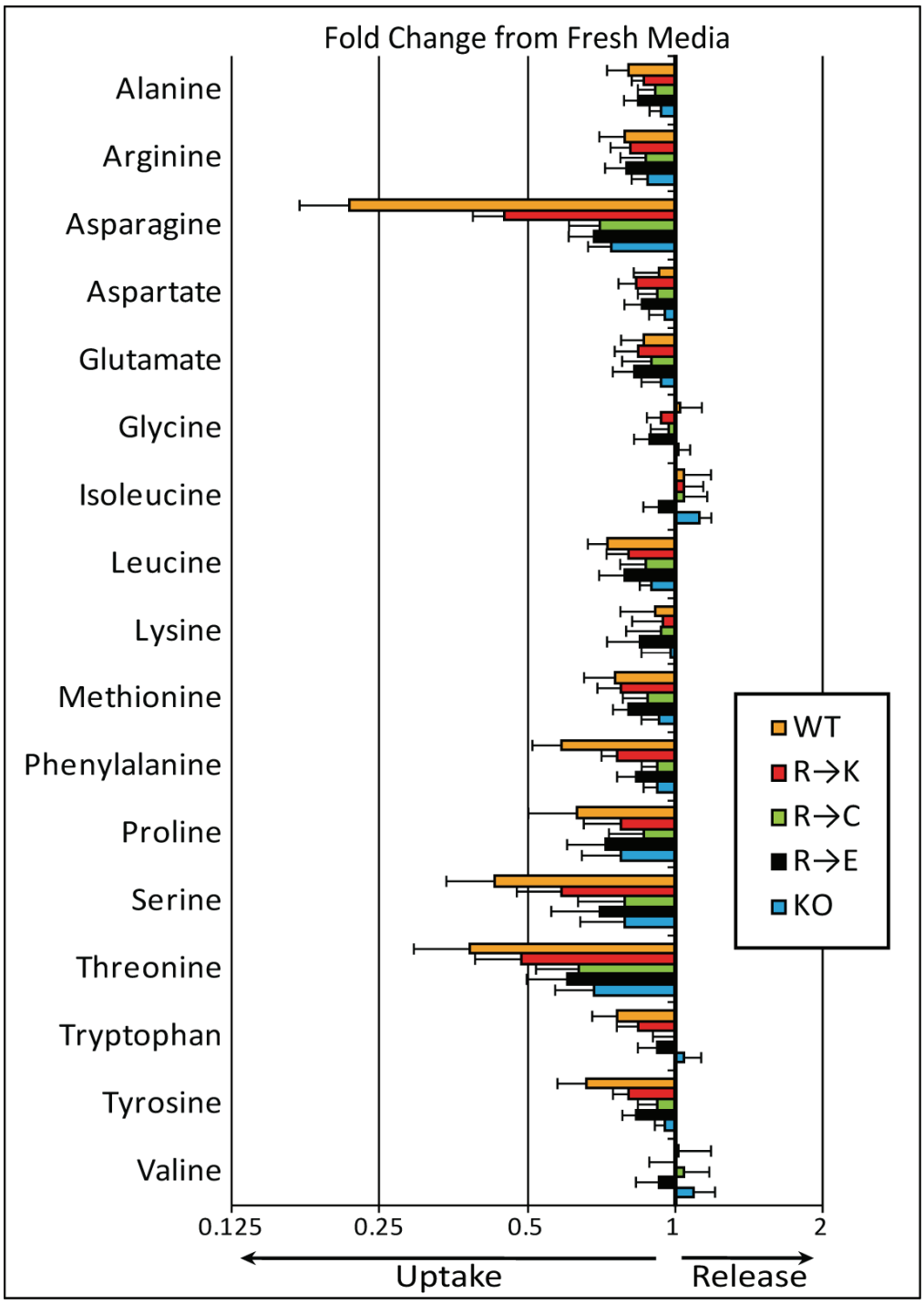


Figure 4-2A. Ratio of metabolites in post-growth to pre-growth medium. Amino acids from *SDH3* strains. Post-growth metabolite concentrations were normalized to the mass of dry protein-free lysates and compared to mass-normalized pre-growth media concentrations. Metabolite concentrations were not normalized to cell culture densities. Yellow, WT; red, Arg47Lys; green Arg47Cys; black, Arg47Glu; blue, KO.

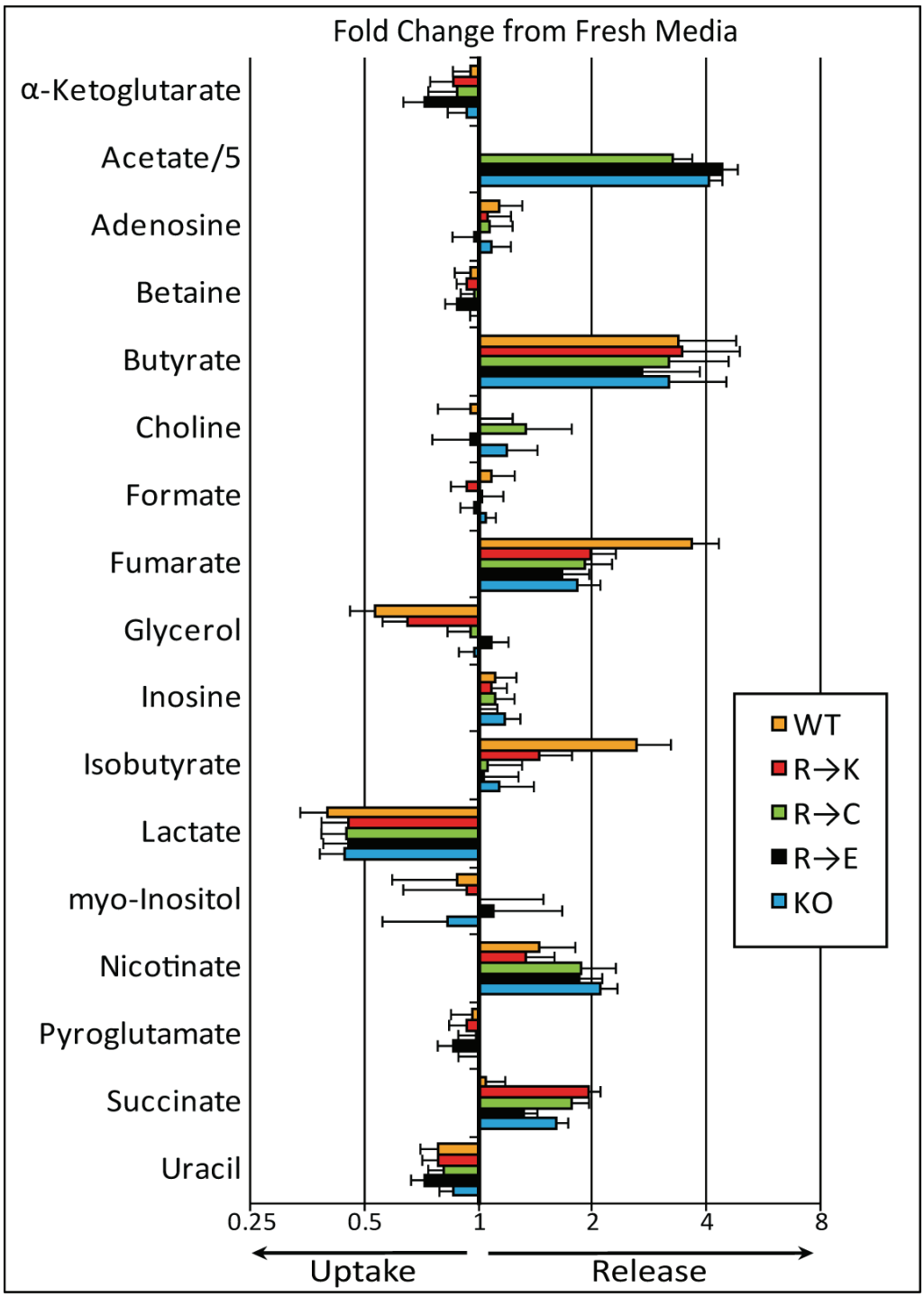


Figure 4-2B. Ratio of metabolites in post-growth to pre-growth medium. Non-amino acid metabolites from SDH3 strains. Post-growth metabolite concentrations were normalized to the mass of dry protein-free lysates and compared to mass-normalized pre-growth media concentrations. Metabolite concentrations were not normalized to cell culture densities. Yellow, WT; red, Arg47Lys; green Arg47Cys; black, Arg47Glu; blue, KO.

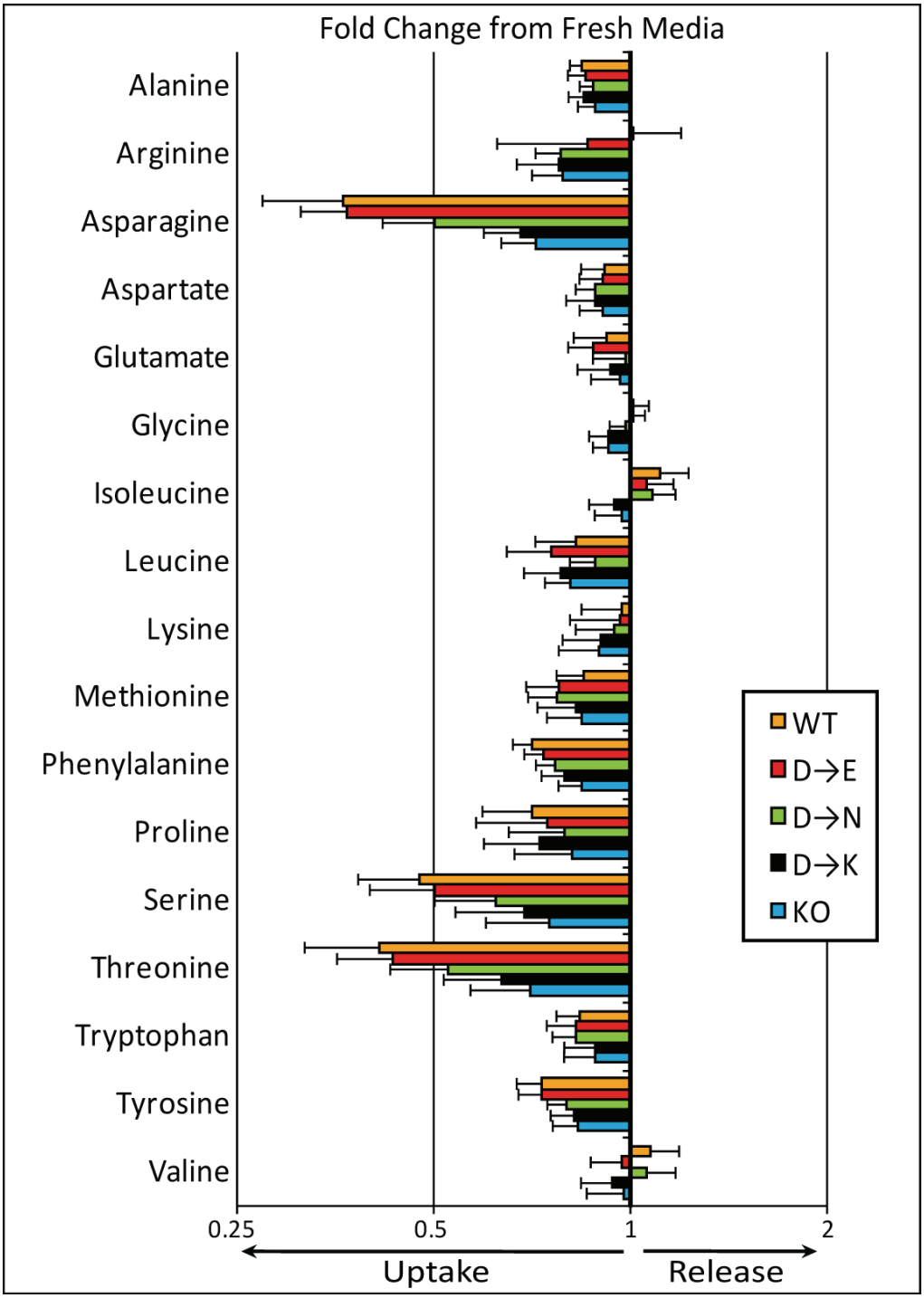


Figure 4-2C. Ratio of metabolites in post-growth to pre-growth medium. Amino acids from *SDH4* strains. Post-growth metabolite concentrations were normalized to the mass of dry protein-free lysates and compared to mass-normalized pre-growth media concentrations. Metabolite concentrations were not normalized to cell culture densities. Yellow, WT; red, Asp88Glu; green, Asp88Asn; black Asp88Lys; blue, KO.

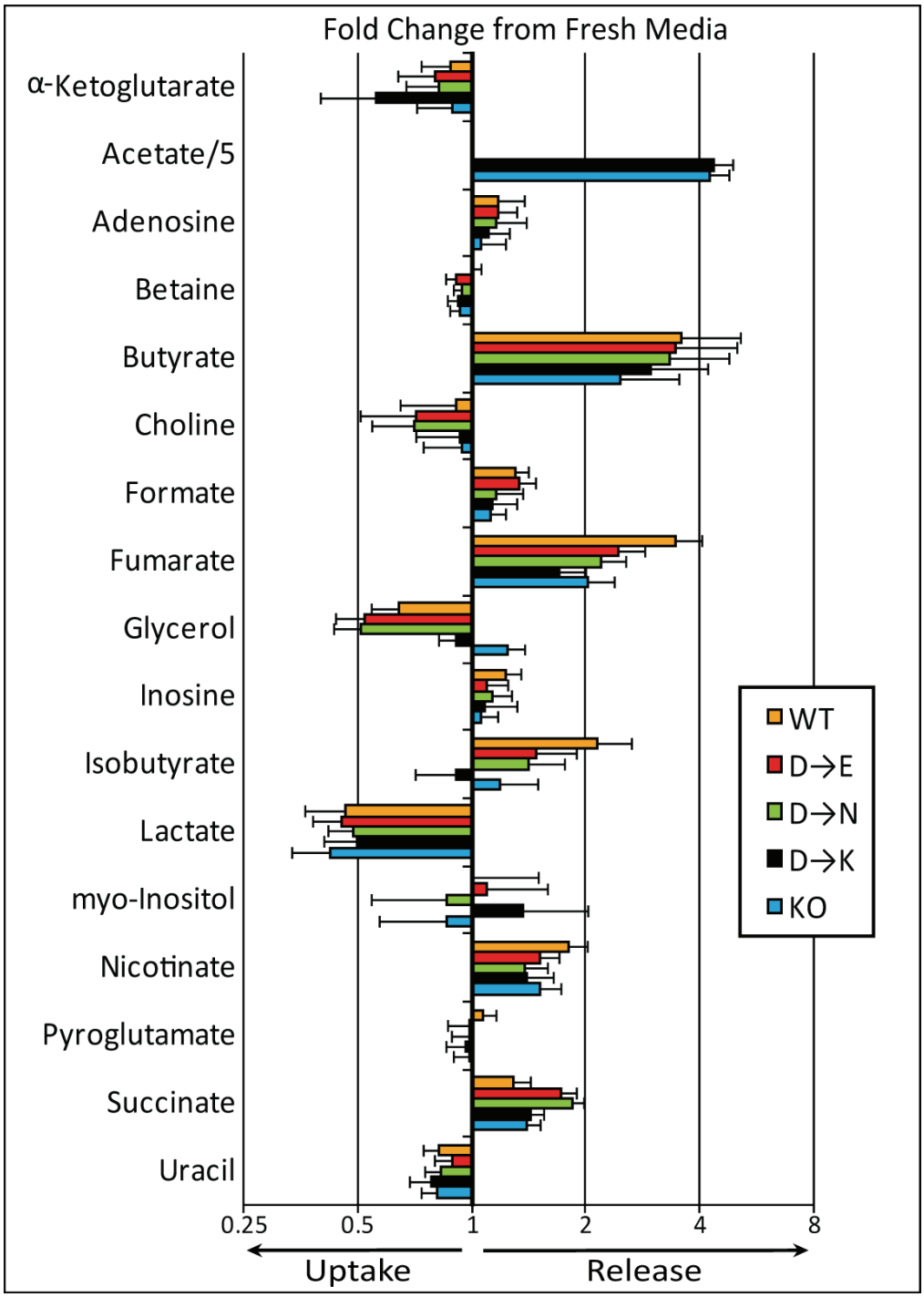


Figure 4-2D. Ratio of metabolites in post-growth to pre-growth medium. Non-amino acids from *SDH4* strains. Post-growth metabolite concentrations were normalized to the mass of dry protein-free lysates and compared to mass-normalized pre-growth media concentrations. Metabolite concentrations were not normalized to cell culture densities. Yellow, WT; red, Asp88Glu; green, Asp88Asn; black Asp88Lys; blue, KO.

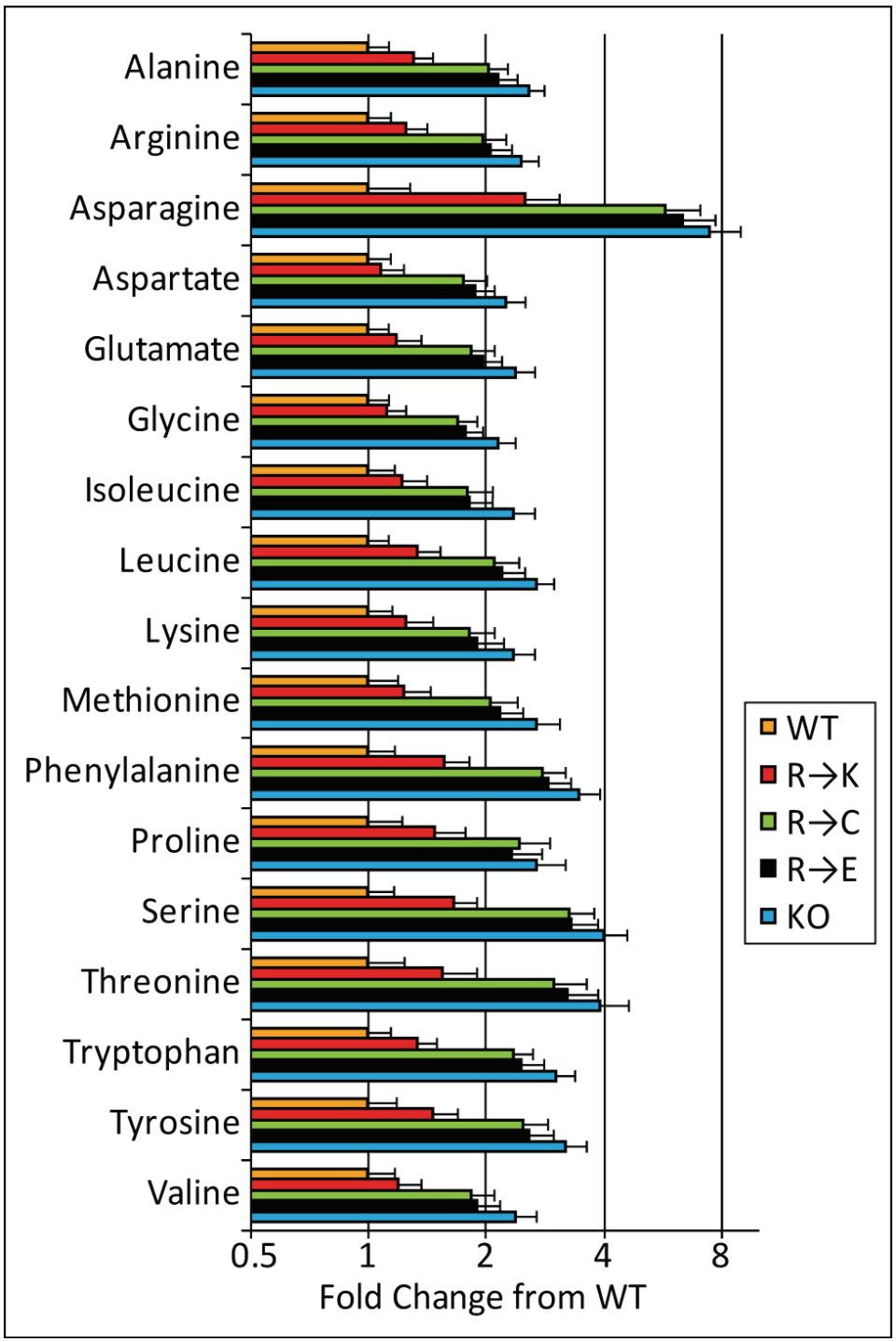


Figure 4-3A. Growth-normalized comparison of metabolites in spent medium. Amino acids in post-growth media of *SDH3* mutants. Metabolites were quantified and normalized to the mass of dry protein-free lysates and to cell culture densities. Yellow, WT; red, Arg47Lys; green, Arg47Cys; black, Arg47Glu; blue, KO.

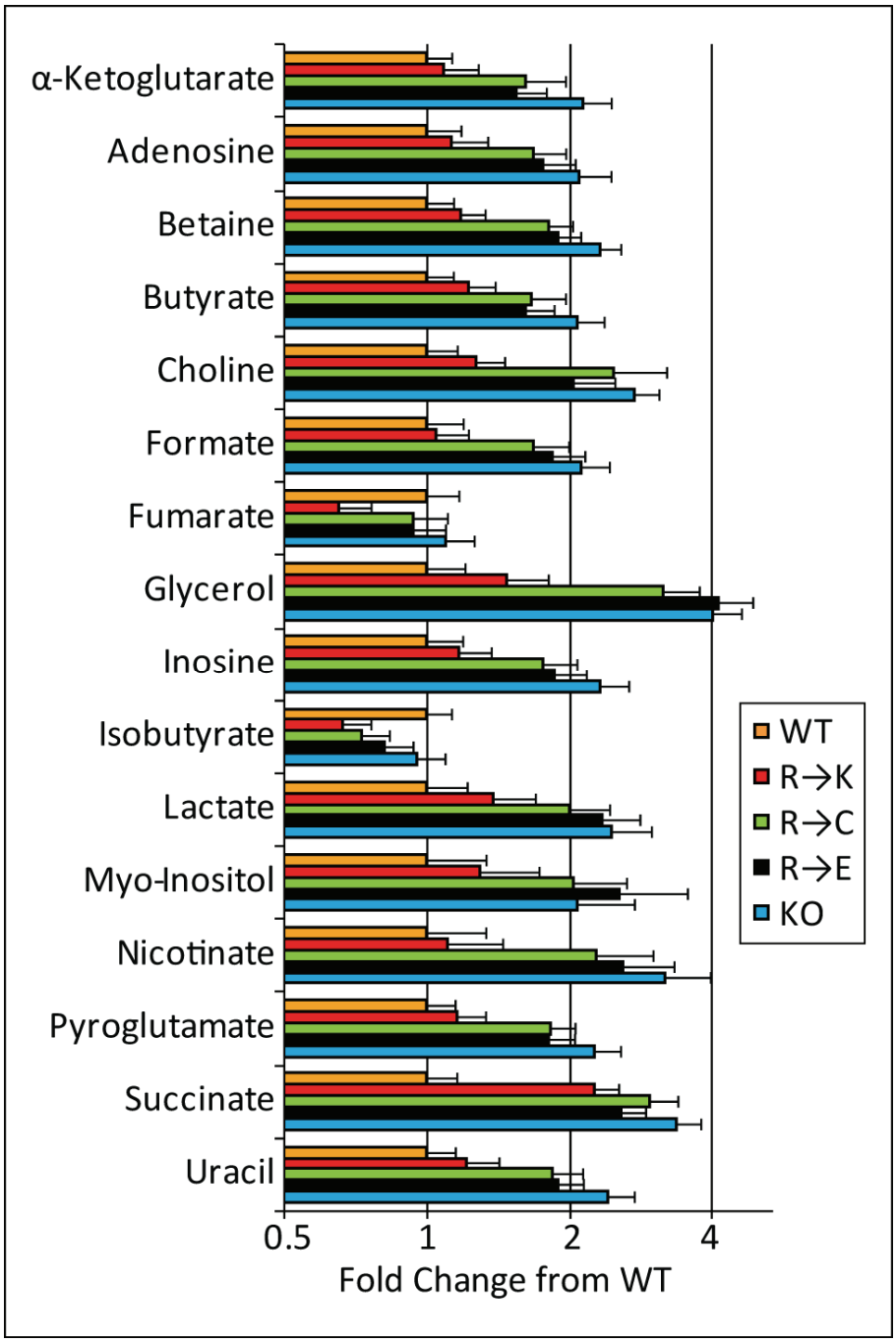


Figure 4-3B. Growth-normalized comparison of metabolites in spent medium. Non-amino acids in post-growth media of *SDH3* mutants. Metabolites were quantified and normalized to the mass of dry protein-free lysates and to cell culture densities. Yellow, WT; red, Arg47Lys; green, Arg47Cys; black, Arg47Glu; blue, KO.

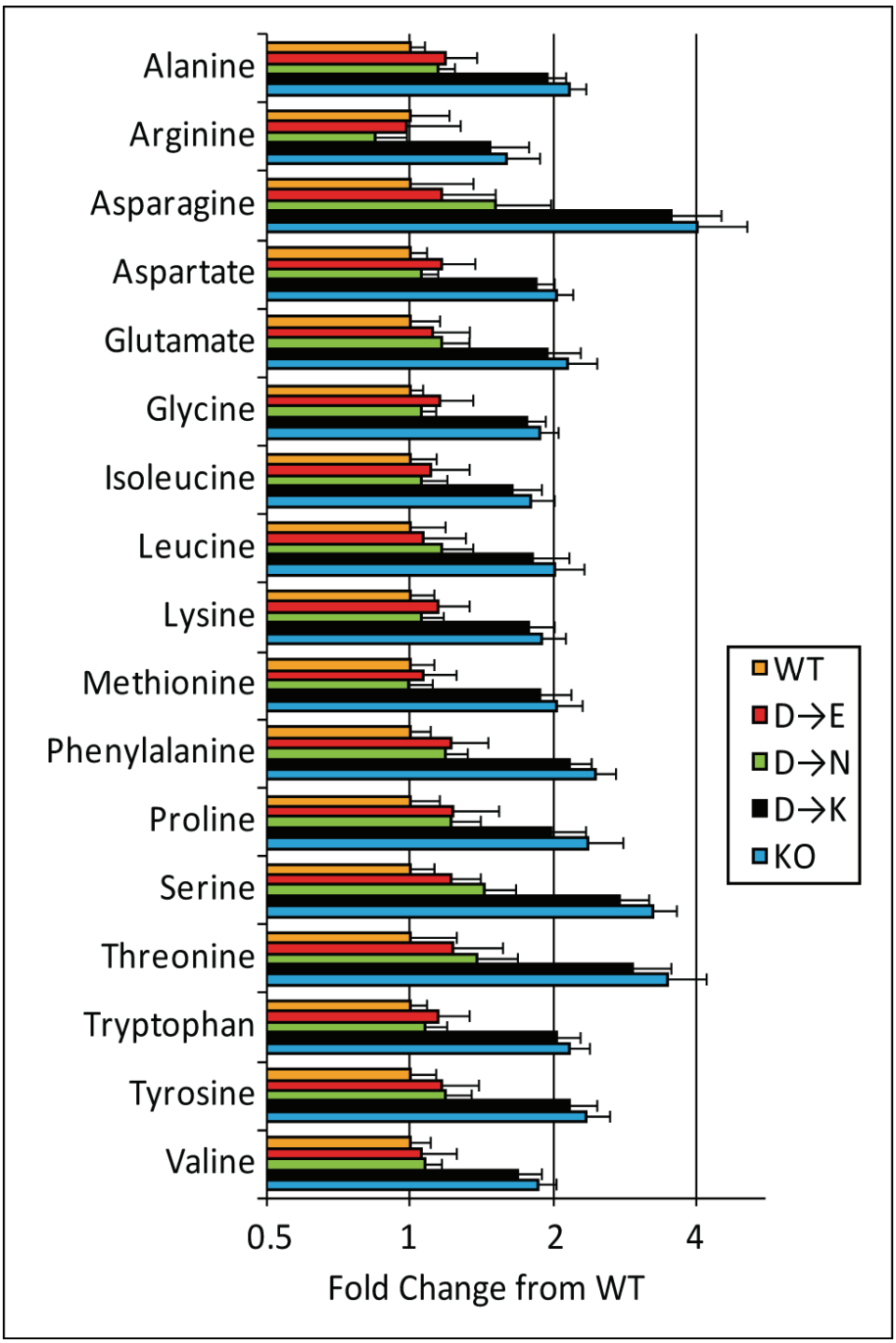


Figure 4-3C. Growth-normalized comparison of metabolites in spent medium. Amino acids in post-growth media of *SDH4* mutants. Metabolites were quantified and normalized to the mass of dry protein-free lysates and to cell culture densities. Yellow, WT; red, Asp88Glu; green, Asp88Asn; black, Asp88Lys; blue, KO.

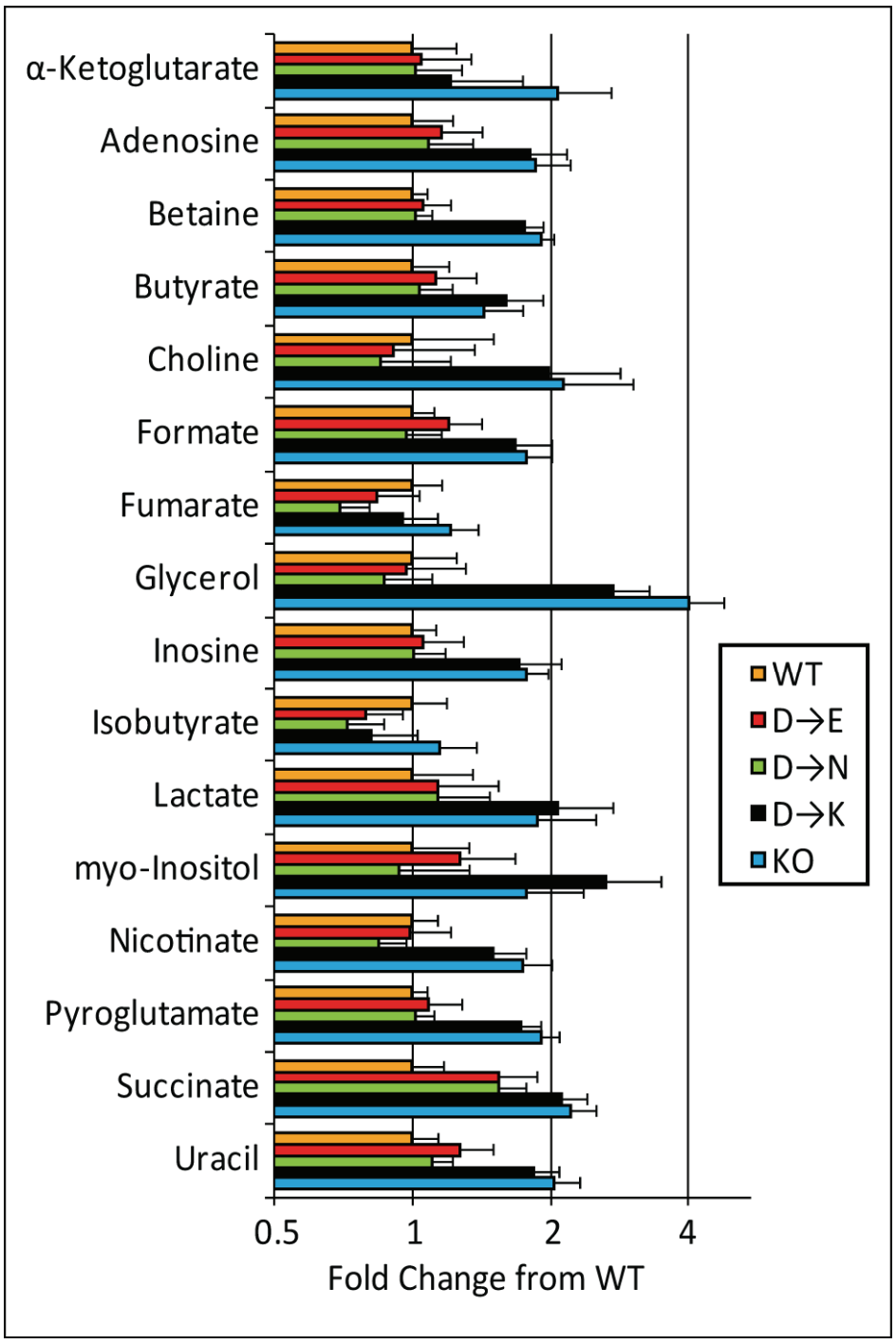


Figure 4-3D. Growth-normalized comparison of metabolites in spent medium. Non-amino acid metabolites in post-growth media of *SDH4* mutants. Metabolites were quantified and normalized to the mass of dry protein-free lysates and to cell culture densities. Yellow, WT; red, Asp88Glu; green, Asp88Asn; black, Asp88Lys; blue, KO.

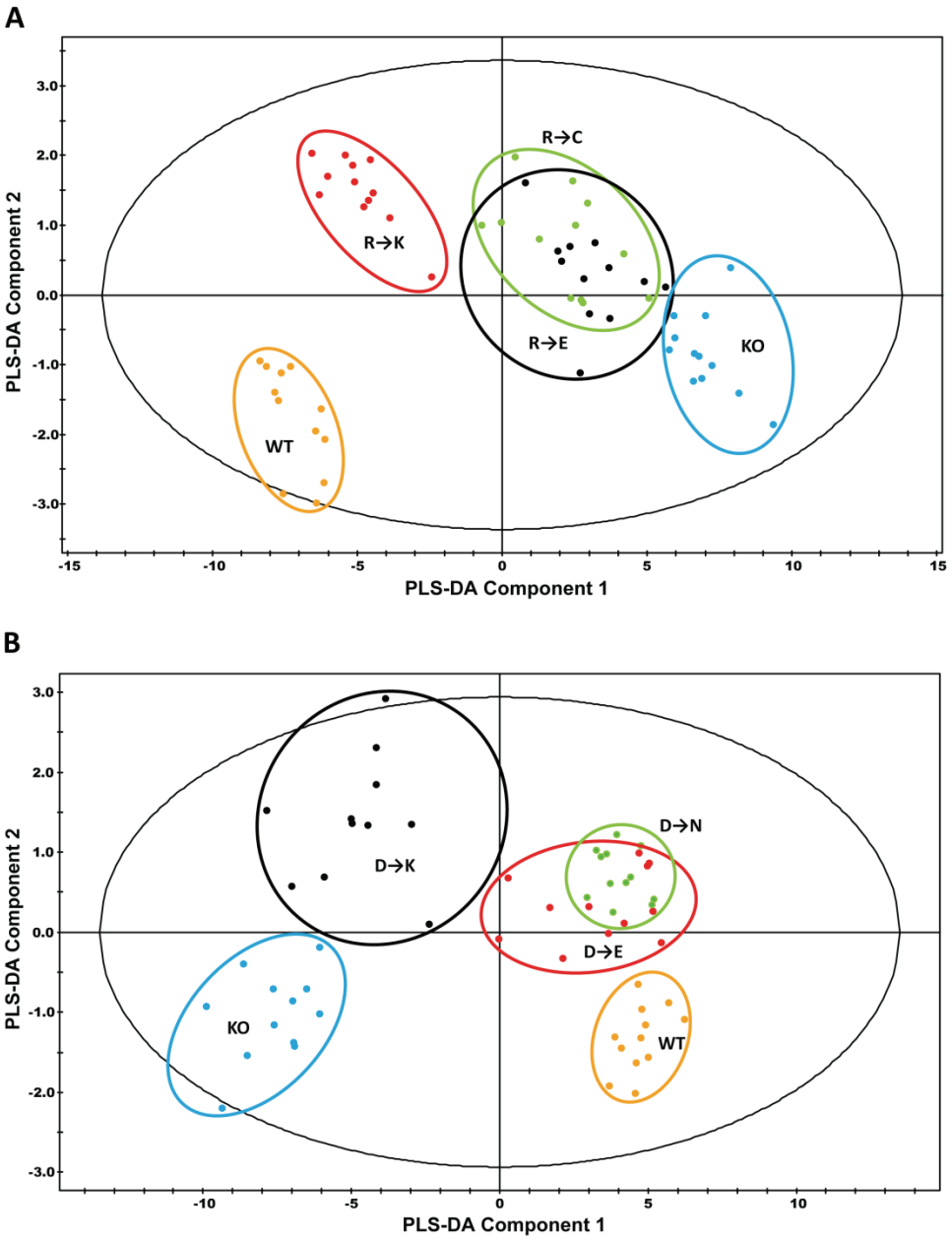


Figure 4-4. PLS-DA model plots of exometabolome profiles. A, *SDH3* mutant strains. Yellow, WT; Arg47Lys; green, Arg47Cys; black, Arg47Glu; blue, KO. The model is characterized by the following parameters using 5 components: R^2X_1, 0.873; R^2X_2, 0.0524; R^2X_{cum}, 0.96; R^2Y_{cum}, 0.759; Q^2_{cum}, 0.663. *N* = 12 for all strains, CV-ANOVA *P* = 4.16×10^{-23}. B, *SDH4* mutant strains. Yellow, WT; red, Asp88Glu; green, Asp88Asn; black, Asp88Lys; blue, KO. The model is characterized by the following parameters using 4 components: R^2X_1, 0.833; R^2X_2, 0.0401; R^2X_{cum}, 0.907; R^2Y_{cum}, 0.692; Q^2_{cum}, 0.593. *N* = 12 for all strains, CV-ANOVA *P* = 2.19×10^{-20}. Ovals are added solely to draw attention to location of each class and do not reflect any statistical relevance.

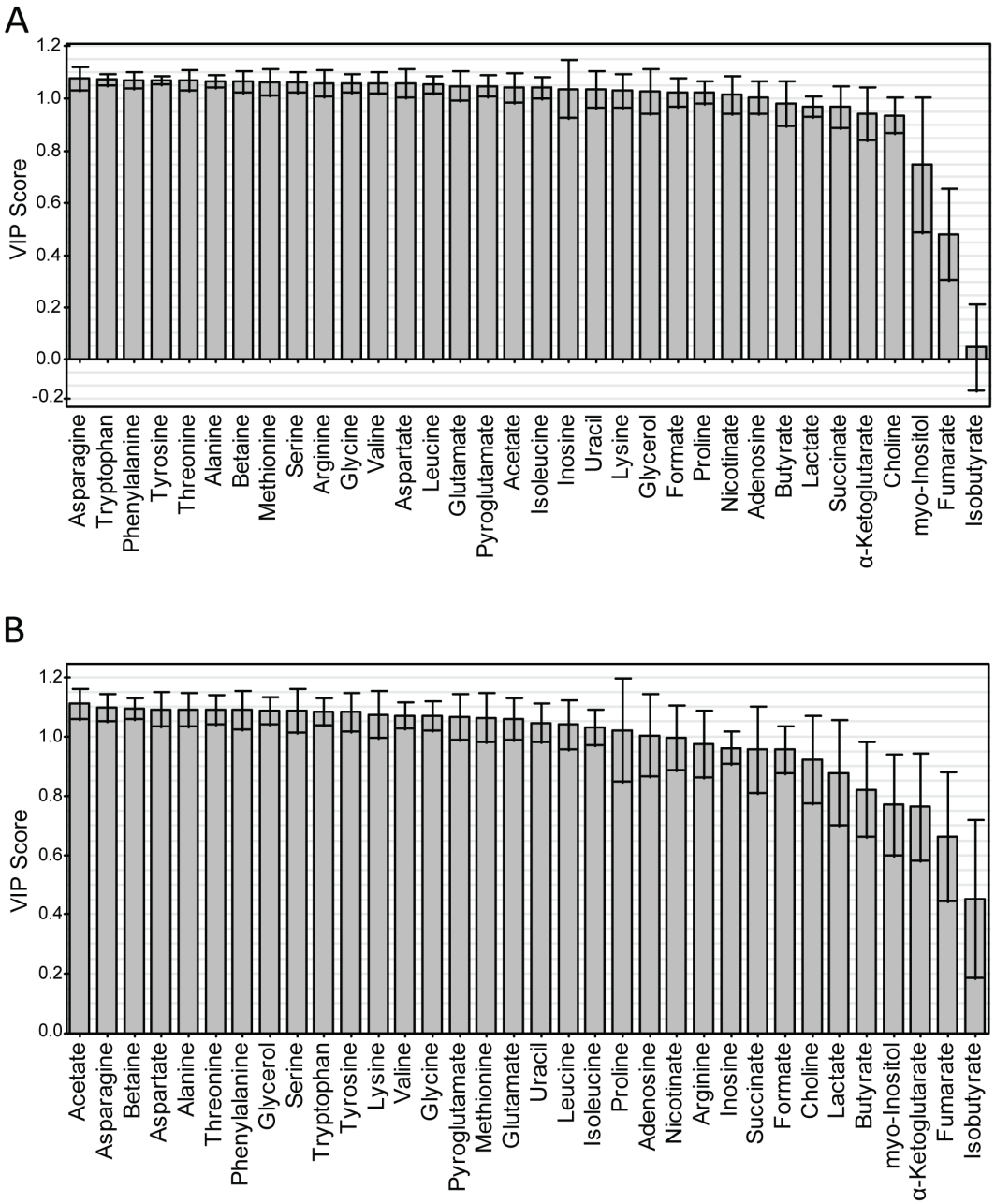


Figure 4-5 A and B. VIP Plots. Variable importance of projection plots for exometabolites along component 1. A, *SDH3* PLS-DA model. B, *SDH4* PLS-DA model.

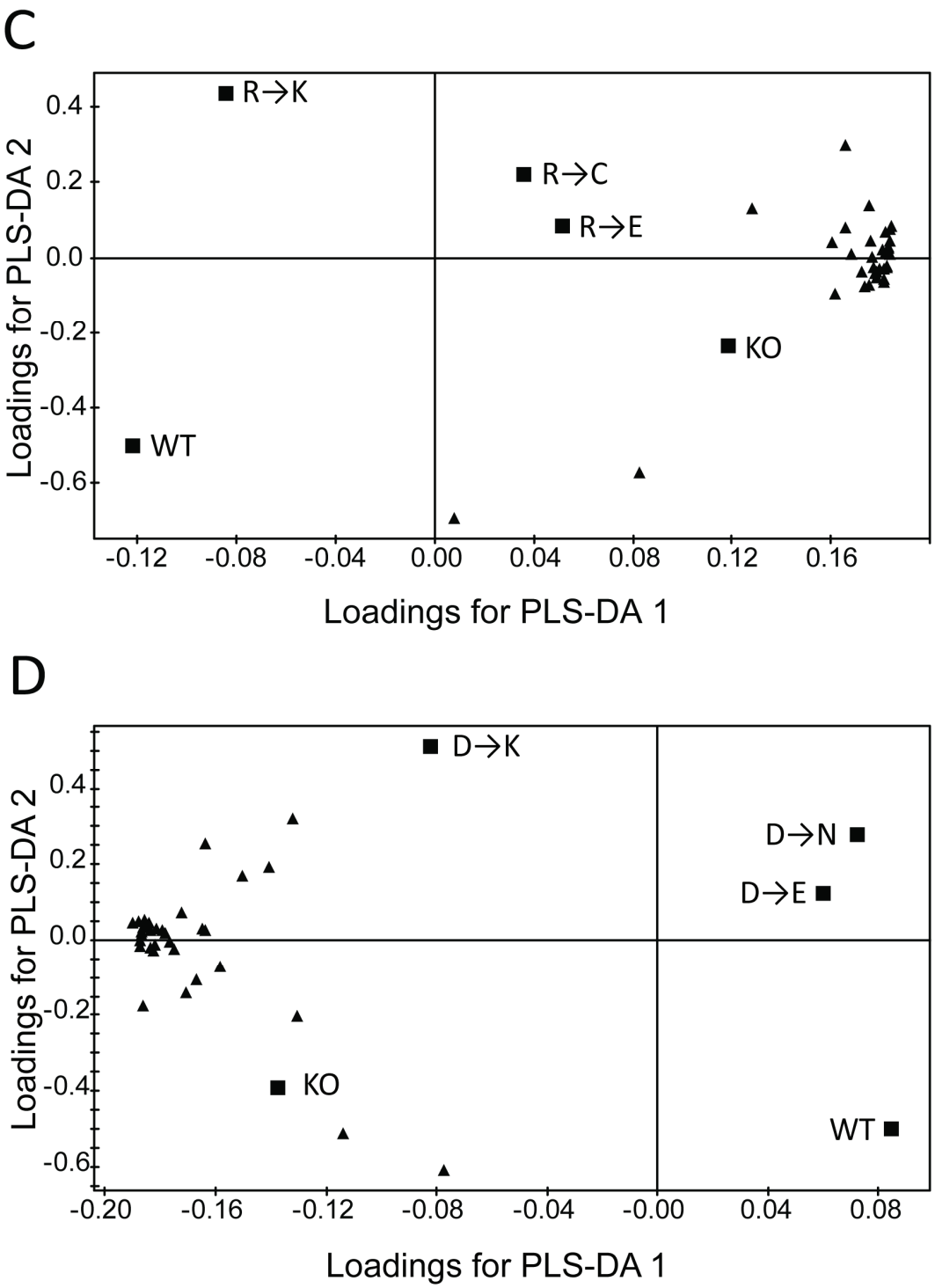


Figure 4-5 C and D. Biplots using PLS-DA components 1 and 2. C, *SDH3* model. R→K, Arg47Lys; R→C, Arg47Cys; R→E, Arg47Glu. D, *SDH4* PLS-DA model. D→E, Asp88Glu; D→N, Asp88Asn; D→K, Asp88Lys. Squares, strain groups; triangles, individual metabolites.

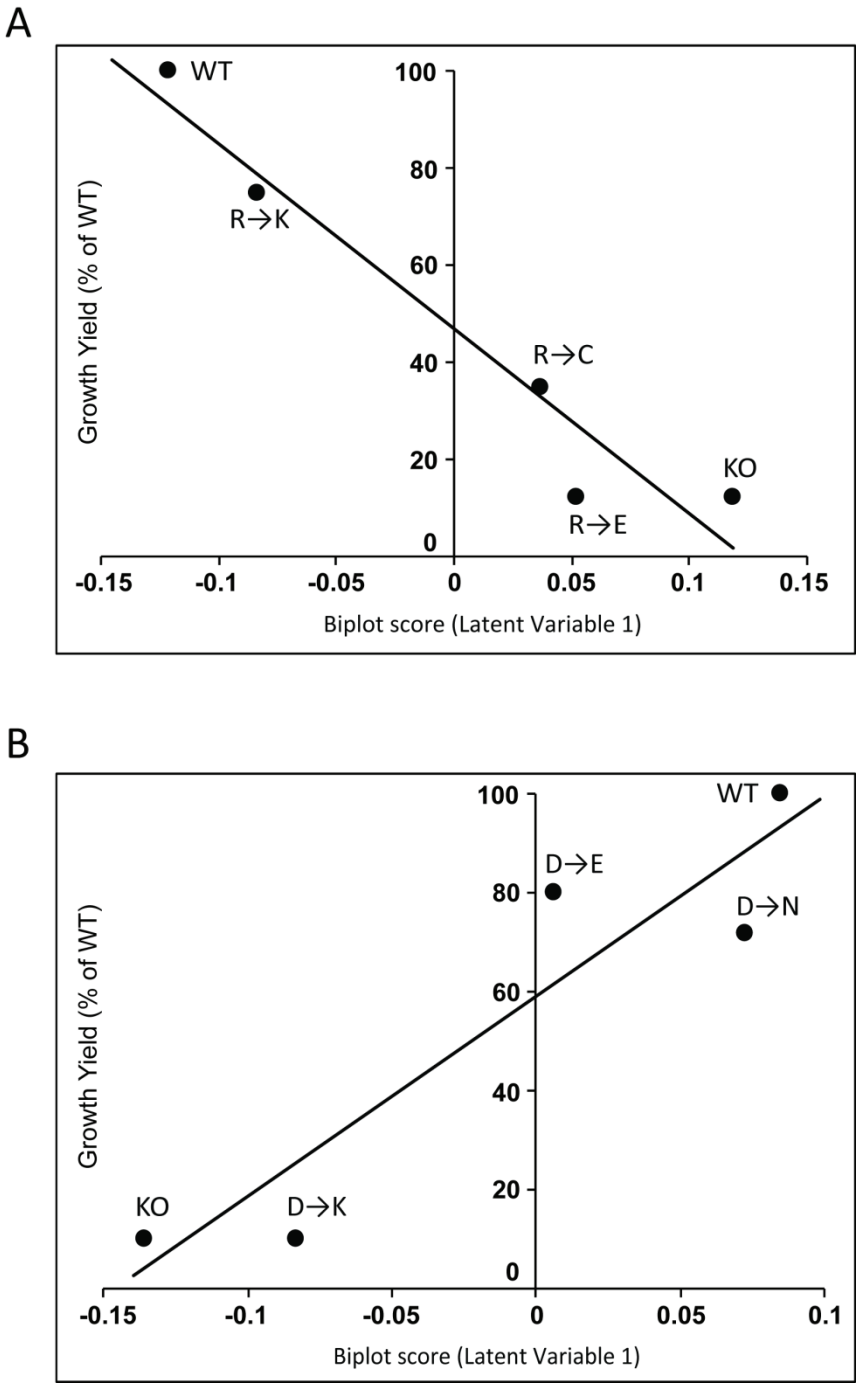


Figure 4-6. Correlation between component 1 biplot score and growth yield. Component 1 biplot scores from each PLS-DA model are plotted against previously determined growth yields on galactose medium; the wild type is set to 100%.39 A, *SDH3* mutant strains. Pearson correlation coefficient R = -0.97, degrees of freedom = 3, $p < 0.02$. B, *SDH4* mutant strains. R = 0.93, degrees of freedom = 3, $p < 0.05$.

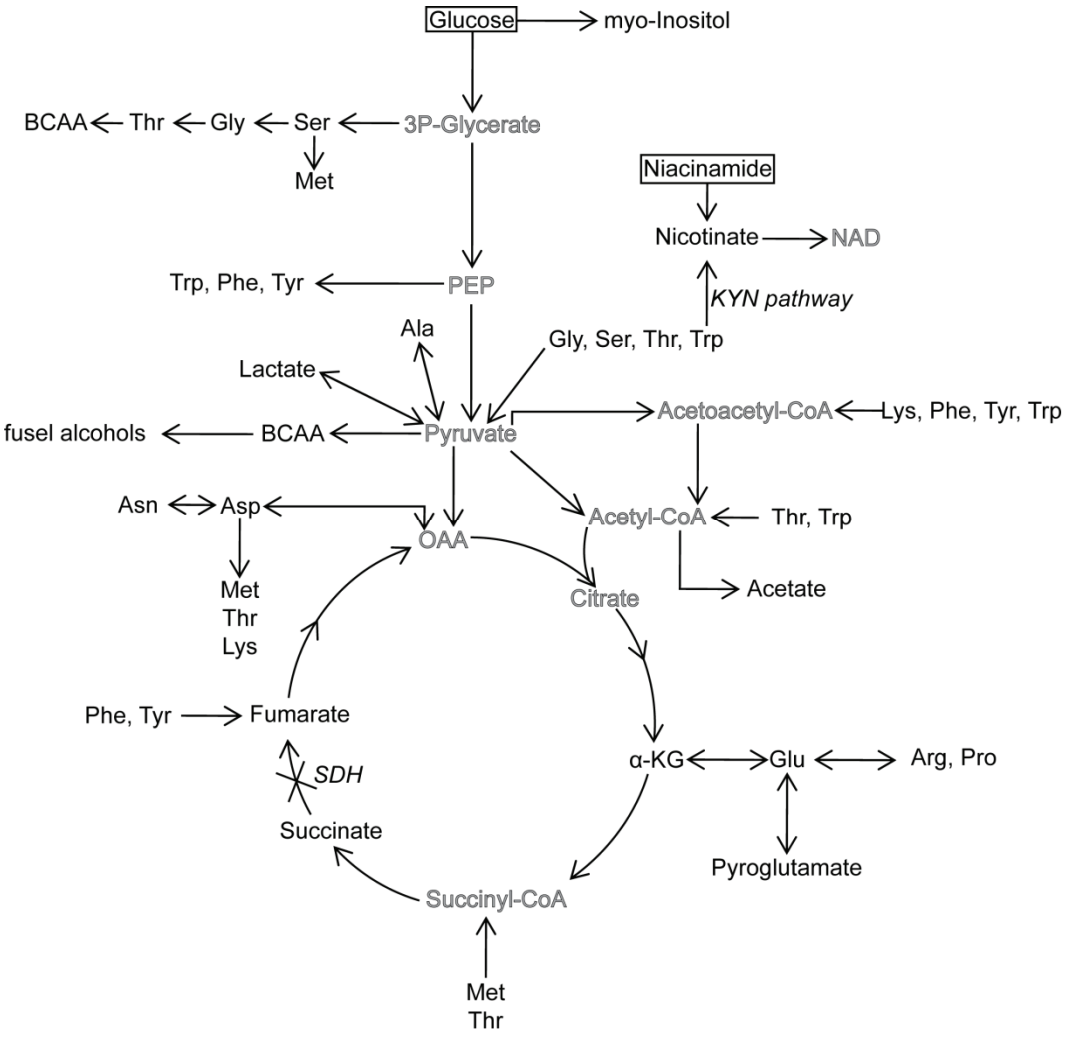


Figure 4-7. Overview of carbohydrate and amino acid metabolism in yeast. Major energy-related metabolic pathways and amino acid biosynthetic and catabolic pathways are highlighted. Black text, metabolites with increased concentrations compared to wild-type when normalized to mass and cell culture density; grey text, metabolites that were not detected; boxed, metabolites completely depleted from the medium. BCAA, branched chain amino acids. The location of SDH in the TCA cycle is indicated. KYN, kynurenine.

References

Ackrell BA (2002) Cytopathies involving mitochondrial complex II. Mol Aspects Med 23 (5):369-384

Ackrell BAC, Johnson MK, Gunsalus RP, Cecchini G (1992) Structure and function of succinate dehydrogenase and fumarate reductase. In: Müller F (ed) Chemistry and Biochemistry of Flavoenzymes. vol III. CRC Press, Boca Raton, pp 229-297

Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nat Biotechnol 21 (6):692-696

Bayley JP, Devilee P, Taschner PE (2005) The SDH mutation database: an online resource for succinate dehydrogenase sequence variants involved in pheochromocytoma, paraganglioma and mitochondrial complex II deficiency. BMC Med Genet 6 (1):39

Bayley JP, van Minderhout I, Weiss MM, Jansen JC, Oomen PH, Menko FH, Pasini B, Ferrando B, Wong N, Alpert LC, Williams R, Blair E, Devilee P, Taschner PE (2006) Mutation analysis of SDHB and SDHC: novel germline mutations in sporadic head and neck paraganglioma and familial paraganglioma and/or pheochromocytoma. BMC Med Genet 7 (1):1

Baysal BE, Rubinstein WS, Taschner PE (2001) Phenotypic dichotomy in mitochondrial complex II genetic disorders. J Mol Med 79 (9):495-503

Boubekeur S, Bunoust O, Camougrand N, Castroviejo M, Rigoulet M, Guerin B (1999) A mitochondrial pyruvate dehydrogenase bypass in the yeast *Saccharomyces cerevisiae*. J Biol Chem 274 (30):21044-21048

Bourgeron T, Rustin P, Chretien D, Birch-Machin M, Bourgeois M, Viegas-Péquignot E, Munnich A, Rötig A (1995) Mutation of a nuclear succinate dehydrogenase gene results in mitochondrial respiratory chain deficiency. Nature Genetics 11:144-149

Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, Clarke S, Schofield PM, McKilligin E, Mosedale DE, Grainger DJ (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^{1}H-NMR-based metabonomics. Nat Med 8 (12):1439-1444

Burnichon N, Briere JJ, Libe R, Vescovo L, Riviere J, Tissier F, Jouanno E, Jeunemaitre X, Benit P, Tzagoloff A, Rustin P, Bertherat J, Favier J, Gimenez-Roqueplo AP (2010) *SDHA* is a tumor suppressor gene causing paraganglioma. Hum Mol Genet 19 (15):3011-3020

Buu LM, Chen YC, Lee FJ (2003) Functional characterization and localization of acetyl-CoA hydrolase, Ach1p, in *Saccharomyces cerevisiae*. J Biol Chem 278 (19):17203-17209

Cann AJ (2003) Maths from scratch for biologists. John Wiley & Sons, Chichester, England

Casal M, Paiva S, Queirós O, Soares-Silva I (2008) Transport of carboxylic acids in yeasts. FEMS Microbiology Reviews 32 (6):974-994

Dickinson J (2004) Nitrogen Metabolism. In: Dickinson J, Schweizer M (eds) The metabolism and molecular physiology of *Saccharomyces cerevisiae*. 2nd edn. CRC Press, Philadelphia, PA, pp 104-116

Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S (2001) Multi- and megavariate data analysis. Part 1: Basic principles and applications. 2nd edn. Umetrics Academy, Umeå, Sweden

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol 5 (9):763-769

Gietz D, St Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res 20 (6):1425

Griffin JL, Shockcor JP (2004) Metabolic profiles of cancer cells. Nat Rev Cancer 4 (7):551-561

Habano W, Sugai T, Nakamura S, Uesugi N, Higuchi T, Terashima M, Horiuchi S (2003) Reduced expression and loss of heterozygosity of the *SDHD* gene in colorectal and gastric cancer. Oncol Rep 10 (5):1375-1380

Hägerhäll C (1997) Succinate:quinone oxidoreductases: variations on a conserved theme. Biochim Biophys Acta 1320 (2):107-141.

Hazelwood LA, Daran JM, van Maris AJ, Pronk JT, Dickinson JR (2008) The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. Appl Environ Microbiol 74 (8):2259-2266

Hederstedt L, Ohnishi T (1992) Progress in succinate:quinone oxidoreductase research. In: Ernster L (ed) Molecular Mechanisms in Bioenergetics. Elsevier Science Publishers, New York, pp 163-197

Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. Cell 134 (5):714-717

Jansson J, Willing B, Lucio M, Fekete A, Diksved J, Half-varson J, Tysk C, Schmitt-Kopplin P (2009) Metabolomics reveals metabolic biomarkers of Crohn's disease. PLoS One 4 (7):e6386

Kaderbhai NN, Broadhurst DI, Ellis DI, Goodacre R, Kell DB (2003) Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. Comp Funct Genomics 4 (4):376-391

Kell DB, Brown M, Davey HM, Dunn WB, Spasic I, Oliver SG (2005) Metabolic footprinting and systems biology: the medium is the message. Nat Rev Microbiol 3 (7):557-565

Kenny LC, Broadhurst DI, Dunn W, Brown M, North RA, McCowan L, Roberts C, Cooper GJ, Kell DB, Baker PN (2010) Robust early pregnancy prediction of later preeclampsia using metabolomic biomarkers. Hypertension 56 (4):741-749

Kispal G, Cseko J, Alkonyi I, Sandor A (1991) Isolation and characterization of carnitine acetyltransferase from *S. cerevisiae*. Biochim Biophys Acta 1085 (2):217-222

Kispal G, Sumegi B, Dietmeier K, Bock I, Gajdos G, Tomcsanyi T, Sandor A (1993) Cloning and sequencing of a cDNA encoding *Saccharomyces cerevisiae* carnitine acetyltransferase. Use of the cDNA in gene disruption studies. J Biol Chem 268 (3):1824-1829

Krukeberg A, Dickinson J (2004) Carbon Metabolism. In: Dickinson J, Schweizer M (eds) The metabolism and molecular physiology of *Saccharomyces cerevisiae*. 2nd edn. CRC Press, Philadelphia, PA, pp 42-103

Lanza I, Zhang S, Ward L, Karakelides H, Raftery D, Nair K (2010) Quantitative metabolomis by ^{1}H-NMR and LC-MS/MS confirms altered metabolic pathways in diabetes. PLoS One 5 (5):e10538

Lemire BD, Oyedotun KS (2002) The *Saccharomyces cerevisiae* mitochondrial succinate:ubiquinone oxidoreductase. Biochim Biophys Acta 1553 (1-2):102-116

MacIntyre DA, Jimenez B, Lewintre EJ, Martin CR, Schafer H, Ballesteros CG, Mayans JR, Spraul M, Garcia-Conde J, Pineda-Lucena A (2010) Serum metabolome analysis by ^{1}H-NMR reveals differences between chronic lymphocytic leukaemia molecular subgroups. Leukemia 24 (4):788-797

Mancuso M, Orsucci D, Coppede F, Nesti C, Choub A, Siciliano G (2009) Diagnostic approach to mitochondrial disorders: the need for a reliable biomarker. Curr Mol Med 9 (9):1095-1107

Mapelli V, Olsson L, Nielsen J (2008) Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. Trends Biotechnol 26 (9):490-497

Mas S, Villas-Boas SG, Hansen ME, Akesson M, Nielsen J (2007) A comparison of direct infusion MS and GC-MS for metabolic footprinting of yeast mutants. Biotechnol Bioeng 96 (5):1014-1022

Milunsky JM, Maher TA, Michels VV, Milunsky A (2001) Novel mutations and the emergence of a common mutation in the SDHD gene causing familial paraganglioma. Am J Med Genet 100 (4):311-314

Neumann HPH, Pawlu C, Pęczkowska M, Bausch B, McWhinney SR, Muresan M, Buchta M, Franke G, Klisch J, Bley TA, Hoegerle S, Boedeker CC, Opocher G, Schipper J, Januszewicz A, Eng C, Group ftE-APS (2004) Distinct Clinical Features of Paraganglioma Syndromes Associated With SDHB and SDHD Gene Mutations. JAMA 292 (8):943-951

Oyedotun KS, Lemire BD (1997) The carboxyl terminus of the *Saccharomyces cerevisiae* succinate dehydrogenase membrane subunit, SDH4p, is necessary for ubiquinone reduction and enzyme stability. J Biol Chem 272 (50):31382-31388

Oyedotun KS, Lemire BD (1999) The *Saccharomyces cerevisiae* succinate-ubiquinone oxidoreductase. Identification of Sdh3p amino acid residues involved in ubiquinone binding. J Biol Chem 274 (34):23956-23962

Oyedotun KS, Lemire BD (2001) The quinone-binding sites of the *Saccharomyces cerevisiae* succinate- ubiquinone oxidoreductase. J Biol Chem 276:16936-16943

Oyedotun KS, Lemire BD (2004) The quaternary structure of the *Saccharomyces cerevisiae* succinate dehydrogenase: Homology modeling, cofactor docking, and molecular dynamics simulation studies. J Biol Chem 279 (10):9424-9431

Oyedotun KS, Yau PF, Lemire BD (2004) Identification of the heme axial ligands in the cytochrome b_{562} of the *Saccharomyces cerevisiae* succinate dehydrogenase. J Biol Chem 279 (10):943-9439

Paiva S, Devaux F, Barbosa S, Jacq C, Casal M (2004) Ady2p is essential for the acetate permease activity in the yeast *Saccharomyces cerevisiae*. Yeast 21 (3):201-210

Pollard PJ, Brière JJ, Alam NA, Barwell J, Barclay E, Wortham NC, Hunt T, Mitchell M, Olpin S, Moat SJ, Hargreaves IP, Heales SJ, Chung YL, Griffiths JR, Dalgleish A, McGrath JA, Gleeson MJ, Hodgson SV, Poulsom R, Rustin P, Tomlinson IP (2005) Accumulation of Krebs cycle intermediates and over-expression of HIF1α in tumours which result from germline FH and SDH mutations. Hum Mol Genet 14 (15):2231-2239

Pronk JT, Wenzel TJ, Luttik MA, Klaassen CC, Scheffers WA, Steensma HY, van Dijken JP (1994) Energetic aspects of glucose metabolism in a pyruvate-

dehydrogenase-negative mutant of *Saccharomyces cerevisiae*. Microbiology 140 (Pt 3):601-610

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19 (1):45-50

Reinke SN, Hu X, Sykes BD, Lemire BD (2010) *Caenorhabditis elegans* diet significantly affects metabolic profile, mitochondrial DNA levels, lifespan and brood size. Mol Genet Metab 100 (3):274-282

Riviere L, Moreau P, Allmann S, Hahn M, Biran M, Plazolles N, Franconi JM, Boshart M, Bringaud F (2009) Acetate produced in the mitochondrion is the essential precursor for lipid biosynthesis in procyclic trypanosomes. Proc Natl Acad Sci U S A 106 (31):12694-12699

Romano JD, Kolter R (2005) *Pseudomonas-Saccharomyces* interactions: influence of fungal metabolism on bacterial physiology and survival. J Bacteriol 187 (3):940-948

Rustin P, Munnich A, Rötig A (2002) Succinate dehydrogenase and human diseases: new insights into a well-known enzyme. Eur J Hum Genet 10 (5):289-291

Rustin P, Rötig A (2002) Inborn errors of complex II - Unusual human mitochondrial diseases. Biochim Biophys Acta 1553 (1-2):117-122.

Scheffler IE (1999) Mitochondria. Wiley-Liss, New York

Scheffler IE (2000) A century of mitochondrial research: achievements and perspectives. Mitochondrion 1:3-31

Schiavi F, Savvoukidis T, Trabalzini F, Grego F, Piazza M, Amistà P, Demattè S, Del Piano A, Cecchini ME, Erlic Z, De Lazzari P, Mantero F, Opocher G (2006) Paraganglioma syndrome: *SDHB*, *SDHC*, and *SDHD* mutations in head and neck paragangliomas. Ann N Y Acad Sci 1073:190-197

Shaham O, Slate NG, Goldberger O, Xu Q, Ramanathan A, Souza AL, Clish CB, Sims KB, Mootha VK (2010) A plasma signature of human mitochondrial disease revealed through metabolic profiling of spent media from cultured muscle cells. Proc Natl Acad Sci U S A 107 (4):1571-1575

Slupsky CM, Rankin KN, Fu H, Chang D, Rowe BH, Charles PG, McGeer A, Low D, Long R, Kunimoto D, Sawyer MB, Fedorak RN, Adamko DJ, Saude EJ, Shah SL, Marrie TJ (2009) Pneumococcal pneumonia: potential for diagnosis through a urinary metabolic profile. Journal of proteome research 8 (12):5550-5558

Smeitink J, van den Heuvel L, DiMauro S (2001) The genetics and pathology of oxidative phosphorylation. Nat Rev Genet 2 (5):342-352

Sugimoto J, Shimohira M, Osawa Y, Matsubara M, Yamamoto H, Goto Y, Nonaka I (2000) A patient with mitochondrial myopathy associated with isolated succinate dehydrogenase deficiency. Brain Dev 22 (3):158-162

Sun F, Huo X, Zhai Y, Wang A, Xu J, Su D, Bartlam M, Rao Z (2005) Crystal structure of mitochondrial respiratory membrane protein complex II. Cell 121 (7):1043-1057

Szeto SS, Reinke SN, Sykes BD, Lemire BD (2007) Ubiquinone-binding site mutations in the Saccharomyces cerevisiae succinate dehydrogenase generate superoxide and lead to the accumulation of succinate. J Biol Chem 282 (37):27518-27526

Thorburn D (2004) Mitochondrial disorders: Prevalence, myths and advances. Jf Inherit Metab Dis 27 (3):349-362

Wallace DC (1999) Mitochondrial diseases in man and mouse. Science 283 (5407):1482-1488

Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. Annu Rev Genet 39:359-407

Wolfe AJ (2005) The acetate switch. Microbiol Mol Biol Rev 69 (1):12-50

CHAPTER 5

Metagenomic and metabolomic characterization of rabies encephalitis:

New insights into the treatment of an ancient disease.

Stacey N Reinke[1], Lothar Resch[2], Ferdinand Maingat[3], Will Branton[3], Alan C Jackson[4], R Holt[5], Caroline M Slupsky[3], Tom J Marrie[3], Brian D Sykes[1], and Chris Power[3]

Departments of [1]Biochemistry, [2]Laboratory Medicine and Pathology, and [3]Medicine, University of Alberta, Edmonton, Canada; [4]Department of Medicine (Neurology), University of Manitoba, Winnipeg, Canada; [5]Genome Science Centre, British Columbia Cancer Agency, Vancouver, Canada

A version of this chapter has been submitted for publication.

Contributions: SNR and BDS performed metabolomics experiments on brain tissue and analyzed previously acquired CSF spectra. LR performed the autopsy and described the pathological findings. FM and WB prepared tissue samples for NMR and deep sequencing, and performed all other experiments. ACJ provided intellectual property. RH performed deep sequencing studies. CMS and TJM provided NMR spectra of CSF. SNR, LR, BDS and CP wrote the manuscript.

Introduction

Rabies is one of the most historically feared diseases, with disease symptoms being described as early as the second century BC (Adamson 1977; Jackson 2007a). Rabies virus infection is a zoonotic disease spread through bite or saliva contact from an infected animal and remains common in developing countries. About 50,000 deaths are reported annually worldwide (World Health Organization 2005). While dogs are the primary reservoir for rabies in developing countries, rabies in North America continues to exist in wildlife. Between 1924 and 2007, 24 human rabies cases were reported in Canada. Seven cases were reported after 1970; 6 of these were transmitted from bats (Parker et al. 2003; Varughese 2000).

Rabies virus is a prototypic infection of the nervous system in which the virus selectively infects neurons, using retrograde axonal transport to traffic within the nervous system (Jackson 2008). Upon entry into the CNS neurons, infection becomes rapidly widespread along the neuroanatomical pathways, eventually spreading to peripheral neural networks (Jackson 2007b). The amount of time between initial infection and symptom onset, the incubation period, is typically between 3 and 13 weeks; however, incubation periods have been reported to be as short as a few days and longer than a year (Smith et al. 1991; Jackson 2007b). Early symptoms usually include localized pain at the wound site, fever, anxiety, malaise, fatigue, and irritability (Warrell 1976; Dupont and Earle 1965; Hattwick et al. 1974). These symptoms develop into one of two clinical forms of the disease: classical or paralytic rabies. Classical rabies, also called furious rabies, presents in approximately 80% of patients. It is characterized by

hyperexcitability, agitation, hallucinations, aggressiveness, and hypersalivation. Many patients also exhibit hydrophobia, the hallmark of rabies, brought on by spasms of the diaphragm and neck muscles on attempt to swallow; hydrophobia is unique to rabies infection. Paralytic, or dumb, rabies is characterized by muscle weakness and is often misdiagnosed. In either clinical presentation, death invariably results within about 14 days of symptom onset (Jackson 2007a).

Recovery from clinical rabies is rare although post-exposure prophylaxis and vaccination remain highly effective means to preventing disease onset (Hemachudha et al. 2006b). These measures include proper wound cleansing and both passive and active immunization (Manning et al. 2008). In 2005, Willoughby and colleagues reported the recovery of a 15-year old girl from clinical rabies. (Willoughby et al. 2005). The patient was treated with ketamine, a non-competitive antagonist of N-methyl-D-aspartate (NMDA) receptors, to protect against excitotoxicity. Ketamine had been previously shown to reduce rabies virus replication in rat cells (Tsiang et al. 1991). The anti-viral drugs ribavirin and amantadine were also administered; amantadine provides additional protection against excitotoxicity (Willoughby et al. 2005; Kornhuber et al. 1994). The patient was found to be clear of transmissible rabies about four weeks after symptom onset. She was released from hospital about one month later with neurological deficits that improved over time (Hu et al. 2007). This intensive treatment regimen was named the Milwaukee Protocol. Due to its original success, the Milwaukee Protocol offered optimism towards the efforts in effectively treating rabies. The protocol has been repeatedly attempted around the world. Three additional recoveries have been reported (The Milwaukee Protocol version 3.1), but not published, while several failures have been well documented

(Hemachudha et al. 2006a; Schmiedel et al. 2007; McDermid et al. 2008; Rubin et al. 2009; Hunter et al. 2010; Maier et al. 2010).

In this study, we report the outcome and pathogenic features of a patient who developed clinical rabies and was unsuccessfully treated with the Milwaukee Protocol. Molecular and histological analyses of autopsied brain tissue revealed evidence of rabies virus, neuronal loss and extensive inflammation. We used ^{1}H-NMR spectroscopy to establish metabolic profiles of excised brain tissue and serial cerebrospinal fluid (CSF) samples. Metabolomic profiling revealed discordant metabolite detection and abundance between the brain and CSF. Metabolic analysis of brain specimens supported our molecular and histological findings; analysis of CSF supported previous clinical findings from patients treated with the same therapeutic approach. As widespread inflammation is not typical of rabies infection, we suggest that the present patient developed the devastating neurological immune reconstitution inflammatory syndrome (neuro-IRIS), which was likely initiated by the use of specific therapeutic agents.

Materials and Methods

Tissue samples

Brain tissue at autopsy and serial cerebrospinal fluid (CSF) specimens (n=3, weeks 1, 4 and 12 post-admission to hospital) were collected from a patient diagnosed with rabies encephalitis (McDermid et al. 2008). For non-rabies controls, brain tissues (total, n=5; median age: 63 yr; sepsis (n=1), myocardial infarction (n=1), HIV/AIDS (n=2), cancer/leukemia (n=1)) and cerebrospinal fluid (CSF) were used from the Laboratory for Neurological Infection and

Immunity, University of Alberta (Vergote et al. 2006). These studies were approved by the University of Alberta Ethics Committee (Health Ethics Board PRO2291).

Neuropathological analyses

Formalin-fixed and paraffin-embedded brain sections were immune-labelled overnight with antibodies to a rabies virus-encoded nucleocapsid protein and human CD3, CD8, and CD68 with subsequent labelling by a biotin-conjugated anti-rabbit secondary IgG antibody (1/500 dilution; Vector Biotechnologies Inc), developed using 3,3'-diaminobenzidine (DAB) (Vector Biotechnologies Inc) and counterstained with hematoxylin (Jackson 1991).

Real-time RT PCR

Human brain, including 3 discrete samples from frontal cerebral cortex, subcortical white matter and basal ganglia from the rabies and non-rabies patients were homogenized and total RNA was purified from which cDNA was prepared. Semi-quantitative real-time PCR was performed (Power et al. 2003). All PCR primers generated amplicons, which were normalized to *GAPDH* mRNA levels and expressed as relative fold change (RFC) compared with controls (mean $\pm$SEM), as previously reported (Noorbakhsh et al. 2010).

Brain deep sequencing

cDNA from frontal subcortical white matter of the rabies and non-rabies patients was subjected to massively parallel sequencing (Freeman et al. 2009). Paired-end 75 nucleotide sequences were compared to existing databases and annotated based on similarity. Gene ontology analyses were executed and

significantly increased or diminished functional groups of genes were compared.

^{1}H-NMR based metabolomic analyses of brain and CSF

Brain tissue samples (cerebral cortex and white matter) from the rabies and a non-rabies patient (male, 70 yr, sepsis) were homogenized and metabolites were extracted with trichloroacetic acid, as described (Reinke et al. 2010). Serial rabies CSF specimens were investigated and compared with multiple non-rabies CSF specimens; absolute metabolite levels were averaged for the rabies and non-rabies patients. CSF samples were filtered prior to NMR analysis. With the exception that samples were not lyophilized, protein-free lysates were prepared for and analyzed by ^{1}H-NMR spectroscopy as described (Szeto et al. 2010).

Results

Clinical history

A 73-year old Caucasian male sustained a bat bite on his left shoulder but failed to seek medical attention (McDermid et al. 2008). Six months later, he developed severe, progressively worsening pain and weakness in his left arm. Generalized weakness, left arm spasms, dysphagia and gasping breaths subsequently followed. The patient became obtunded with autonomic instability. On examination, he showed hypersalivation, opisthotonus, arm myoclonus and generalized hyper-reflexia with preserved brainstem reflexes. Virologic and serologic testing of nuchal skin, saliva and serum confirmed rabies virus infection. Cranial MRI imaging revealed cerebral atrophy, as expected from an elderly patient (Figure 5-1A and C). Treatment commenced 15 days after

symptom onset (3 days after diagnosis). The patient was treated with rabies immune globulin and with a neuro-therapeutic regimen (the Milwaukee Protocol), including ketamine, midazolam and propofol, together with ribavirin and amantadine (The Milwaukee Protocol version 3.1). Intubation and ventilation were instituted (Willoughby et al. 2005). Despite receiving the above protocol (42 day duration), there was no neurological recovery and the protocol was discontinued. Repeat cranial MRI, using both FLAIR and T2 imaging on clinical illness day 68, revealed severe edema (Figure 5-1B and D). CSF contained elevated protein levels and lympocytosis. The patient remained unresponsive and failed to demonstrate brainstem reflexes; clinical care was subsequently withdrawn.

Rabies virus detection and neuropathology

Neuronal loss was evident in various areas of the brain. Surviving neurons contained Negri bodies (Figure 5-2A). Routine H&E staining of the leptomeninges revealed fibrosis; a mononuclear inflammatory cell infiltrate was primarily comprised of lymphocytes with occasional plasma cells and macrophages. Rabies nucleocapsid protein immunoreactivity was detectable in cells resembling neurons, particularly in the basal ganglia (Figure 5-2B). The majority of the lymphocytes were CD3 immunoreactive (Figure 5-2C), with numerous CD8 immunopositive cells (Figure 5-2C, inset). Macrophage activation was highlighted throughout the rabies brain with CD68 immunolabeling (Figure 5-2D). Hypertrophied astrocytes were evident with H&E staining (Figure 5-2E). Occasional mineralized pyramidal (ferruginated) neurons were also present (Figure 5-2F).

Molecular studies

Deep sequencing of the rabies-infected brain-derived RNA disclosed numerous sequence tags corresponding to multiple regions within the rabies genome and revealed that the viral strain infecting the present case was closely related to a bat-derived virus, SHBRV-18 (data not shown). Comparison with the non-rabies brain-derived cDNA revealed genes involved in immune response and inflammation were highly induced in the rabies-infected brain (Figure 5-3A); genes related to neuronal homeostasis and function were markedly suppressed (Figure 5-3B). Relative transcript detection in the rabies brain compared with non-rabies brain specimens using real time RT-PCR analyses revealed increased *HLA-DQA1*, a cell surface receptor on antigen-presenting cells (Figure 5-3D), *CD3ε* and *CD8*, T-cell receptor complexes (Figure 5-3E and F, respectively), and *MxA*, an interferon-activated gene (Figure 5-3G). Conversely, for the NMDA receptor subunit *NR1*, transcript levels were markedly reduced in the rabies brain specimens (Figure 5-3H). Rabies virus RNA was detectable in all examined brain regions but was highest in the basal ganglia (Figure 5-3I).

Metabolomic studies

^{1}H-NMR spectra were derived from CSF and brain (cerebral cortex and white matter) samples (Figure 5-4A). Representative spectra of cerebral cortex acid-soluble metabolites display the γ and β proton peaks of glutamine and N-acetyl-aspartate (NAA), respectively (Figure 5-4A). Increased glutamine and absent NAA were evident in the rabies-infected cerebral cortex. Urea and quinolate were selectively present in the rabies-derived CSF, based on the average metabolite concentrations of rabies-infected serial CSF samples compared to non-rabies

CSF specimens. In contrast, acetate, myo-inositol, pyroglutamate, and succinate were either lower or absent in rabies compared with non-rabies CSF samples (Figure 5-4B). Brain metabolites, expressed relative to matched creatine levels, revealed that rabies-infected cortex exhibited reduced NAA and gamma-butyric acid (GABA) levels compared to the non-rabies cortical levels (Figure 5-4C). The cortex from the rabies-infected patient also showed increased levels of lactate, taurine, phosphoethanolamine, alanine, glutamine, leucine, phenylalanine, serine, threonine, and valine, while acetate and aspartate were markedly decreased (Figure 5-4C). The white matter of the rabies-infected brain (Figure 5-4D) displayed lower acetate, aspartate, glutamate, myo-inositol, and NAA levels; conversely, glutamine, lactate, and taurine levels were higher in the rabies white matter (Figure 5-4D).

Discussion

Following its original success, the Milwaukee Protocol offered optimism towards establishing an effective treatment against rabies infection. This protocol included the use of ketamine, midazolam, propofol, ribavirin, and amantadine (The Milwaukee Protocol version 3.1). Ketamine was the primary drug of choice for three reasons: it is a dissociative anaesthetic, it has been shown to have anti-viral properties specific to rabies in rat cortical neurons, and it potentially protects against NMDA (N-methyl D-aspartate) receptor-mediated excitotoxicity (Lockhart et al. 1992). Amantadine and ribavirin were instituted for their anti-viral properties against the rabies virus (Superti et al. 1985; Jackson et al. 2003). Midazolam offered further patient sedation, lowering metabolic demand and stimulating

protective GABA receptors. Propofol has been shown to completely suppress EEG activity in rabies patients (The Milwaukee Protocol version 3.1).

In the last few years, controversy over the efficacy of this therapy has been mounted for three reasons. First, despite its original success, several failures of the Milwaukee protocol have been well documented (Hemachudha et al. 2006a; Schmiedel et al. 2007; McDermid et al. 2008; Rubin et al. 2009; Hunter et al. 2010; Maier et al. 2010). Second, the original surviving patient had anti-rabies antibodies, despite never receiving prophylactic treatment. The rabies viral strain was also not isolated in this case, and thus could have been attenuated in some way. Therefore, it is unclear what role, if any, the Milwaukee Protocol played in the patient's survival (Jackson 2005). Third, the mechanism for ketamine inhibiting rabies-virus genome transcription has never been established (Lockhart et al. 1992). More recently, Weli and colleagues failed to establish any evidence for rabies-induced excitotoxicity or protective effects of ketamine in neuronal cells and adult mice (Weli et al. 2006). In 2010, Jackson and colleagues reported that oxidative stress is responsible for neuronal injury during rabies infection (Jackson et al. 2010). Despite its many failures and ongoing controversies, the Milwaukee Protocol continues to be used to treat rabies infection. The present patient exhibited classic features of 'furious' rabies and despite extensive interventions aimed at limiting neural damage, succumbed after prolonged survival in intensive care. Herein, we report the first high-throughput metagenomic and metabolomic analyses of a rabies infection of the nervous system in the context of an intensive therapeutic regimen.

The present investigations revealed common features of a typical rabies infection, including presence of rabies virus (proteins and nucleic acids), neuronal loss, and Negri body presence. Rabies glycoprotein was immuno-detected in all regions of the brain examined, which was complemented by widespread detection of rabies virus-encoded RNA by both RT-PCR and deep sequencing. The specific transcript abundance of *MxA* reflected the activation of host response to viral infection. The extent of neuronal loss was revealed through extensive astrogliosis and diminished RNA of the NMDA receptor subunit, NR1, reflecting the predominant neuronal cellular tropism of the rabies virus. Thus, the morphological and molecular aspects of the present case were complementary. The pathological features of the present case were remarkable in that, unlike typical human rabies infection (Perl and Good 1991), neuroinflammation was a predominant feature. Evidence for neuroinflammation includes MRI imaging (Figure 5-1) and immune-detection of lymphocytes (CD3 and CD8 receptors) and macrophages (CD 68 receptor) (Figure 5-2), with supporting molecular analyses (Figure 5-3) compared to non-rabies samples. Activation of microglia/macrophages and lymphocyte infiltration were identified in all brain regions based on specific transcript (*HLA-DQA1, CD3* and *CD8*) abundance, again congruent with matching neuro- and immunopathology studies (Figure 5-3).

The availability of serial CSF and autopsied brain specimens prompted investigation of the metabolic features underlying rabies, also allowing us to verify our molecular observations (Figure 5-4). These studies revealed that CSF and brain specimens contained substantially different profiles with several metabolites exhibiting discordant detection and abundance between CSF and

brain. For example, quinolate and urea were only detectable in CSF while taurine, niacinamide, GABA, and NAA were selectively present in brain tissues. Elevated CSF quinolate levels have been reported in rabies infections (The Milwaukee Protocol version 3.1). Quinolate is an NMDA receptor agonist, inducing excitotoxicity; ketamine has been found to inhibit this activity (Keilhoff et al. 1991). This dichotomy in tissue profiles likely reflects differing origins of metabolites; CSF is largely derived from choroid plexus and the ventricular system, including the circumventricular organs, which are relatively devoid of a blood-brain barrier and in close proximity to circulating blood. In contrast, brain tissue, particularly white matter, is comparatively free of blood-derived molecules but abundant in characteristic neural molecules, such as the neurotransmitter, GABA and the neuronal metabolite NAA. Predictably, cortex and white matter diverged in terms of relative abundance of individual metabolites. However, lactate accumulation was consistently predominant in both anatomical regions, possibly reflecting mitochondrial dysfunction. Mild lactic acidosis has been previously reported during rabies infection (The Milwaukee Protocol version 3.1); this may be attributed to ribavirin exhibiting variable toxicity towards mitochondria (Fleischer et al. 2004). The cortex and white matter were also similarly more concentrated in glutamine, an indication of astrogliosis in both tissues (Hanefeld et al. 2005). Of particular interest was the robust induction of taurine in rabies brain samples, pointing towards a host defense response, as taurine is recognized as a potential cytoprotective molecule (Takahashi et al. 2003; Hagar 2004). Despite increased taurine, the present NMR studies recapitulated the above molecular and pathological analyses in that neuronal loss, indicated by diminished NAA, was a key feature of the present case. Reduced NAA levels, as

revealed through magnetic resonance spectroscopy (MRS), have been reported in other clinical rabies cases (The Milwaukee Protocol version 3.1); measuring NAA levels using MRS during rabies infection would provide information about neuronal loss during the disease and treatment course.

Despite a rigorous and prolonged neuroprotective therapeutic regimen (Milwaukee Protocol), the current patient died with high viral burden in the brain, widespread neuronal loss and remarkable lymphocyte infiltration. The persistent viral load, chiefly in the basal ganglia, might have induced a specific neuroinflammatory (CD8+ T cell) response in the same anatomical site, although this finding is atypical in human rabies (Perl and Good 1991). The robust pleocytosis at the time of the last CSF collection also suggested a concurrent robust neuroinflammatory response, which was supported by the present molecular and neuropathological studies. However, rabies is usually neuropathologically bland (except for Negri bodies) (Jogai et al. 2000) and thus, an alternative explanation for the present findings is a neurological immune reconstitution syndrome (neuroIRIS) (Legris et al. 2010; Johnson and Nath 2011). Neurological IRIS is a devastating inflammatory response that occurs during the immune system recovery following treatment with anti-viral or immune-suppressing drugs (McCarthy and Nath 2010). Even though the pathophysiological mechanisms are not well understood, IRIS is well-described in patients being treated for human immunodeficiency virus (HIV) infection and multiple sclerosis (MS), as well as transplant recipients (McCarthy and Nath 2010). Several components of the present neuro-therapeutic protocol, including ketamine (Wilson et al. 1971; Chang et al. 2005), midazolam (Freire-Garabal et al. 1992) and propofol (Galley et al. 1998) are potentially immunosuppressive.

IRIS is characterized by increased release and circulation of CD8+ cells within one to two months of immune recovery (McCarthy and Nath 2010). Immune staining and specific transcript quantitation revealed abundant presence of CD8+ T lymphocytes. Similar to previous reports of neuroIRIS cases, this protocol was ceased more than 20 days prior to the final CSF analysis before death (McCombe et al. 2009). The cessation of the immunosuppressive therapeutic protocol could permit activation of lymphocytes targeting rabies virus (RV) antigens and their ensuing entry into the brain during immune restoration. Thus, the use of intensive neuroprotective protocols should be considered carefully before initiation, as the effects of the treatment protocol might complicate or worsen the underlying disease process.

Limitations of Study

This study contains two major caveats. First, as discussed above, it may take several hours for cells to die post-mortem and autopsies are not general performed for several hours, during which anaerobic metabolism may still take place. Despite this limitation, we did see robust metabolic differences between the rabies-infected and matched control brains that correlate with pathological findings. Second, we normalized brain metabolite concentrations to creatine concentration; creatine metabolism is also likely altered post-mortem. Although normalizing metabolite concentrations to excised tissue mass or protein content would have been more suitable, these measurements were not taken due to the extremely infectious nature of the rabies virus.

Figures

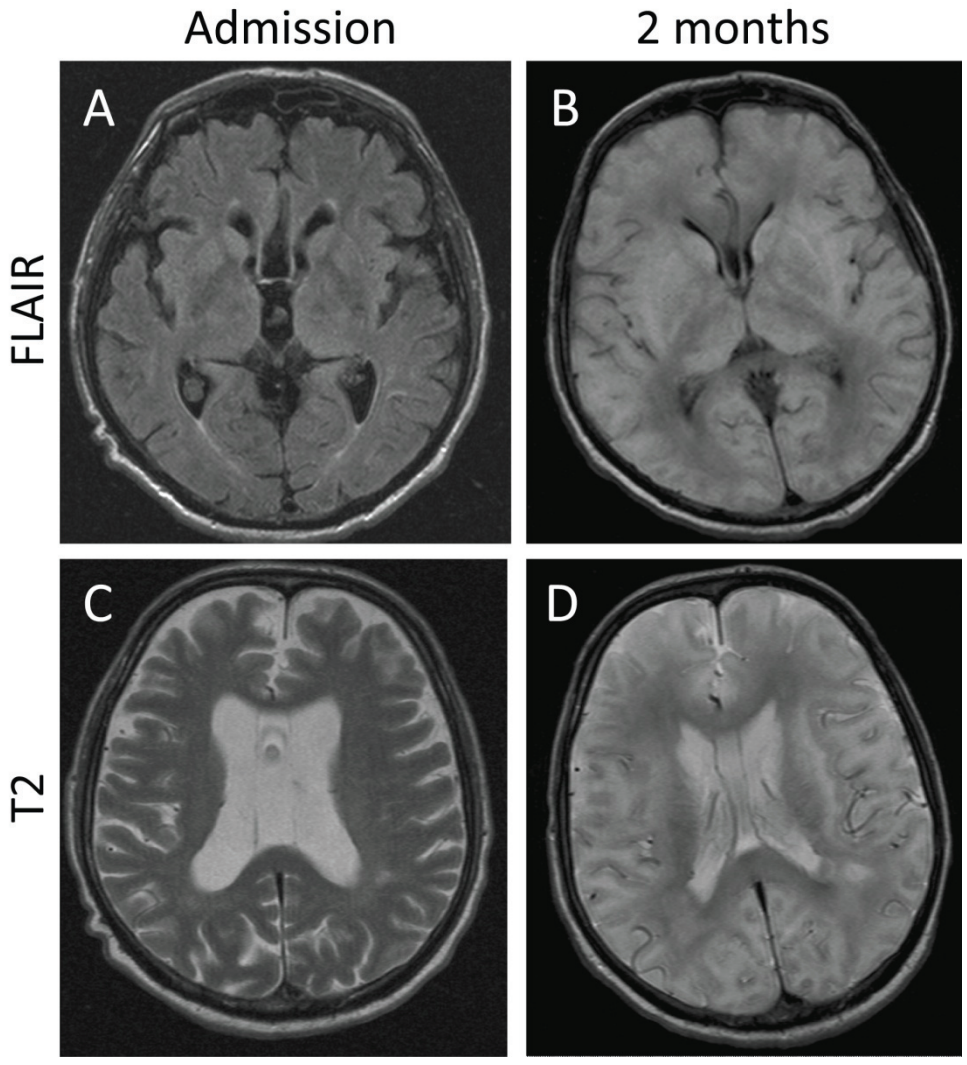


Figure 5-1. Neuroimaging showing cranial MRI images. A, Flair imaging at time of admission. B, Flair imaging at 2 months post admission. C, T2 imaging at time of admission. D, T2 imaging at 2 months post admission.

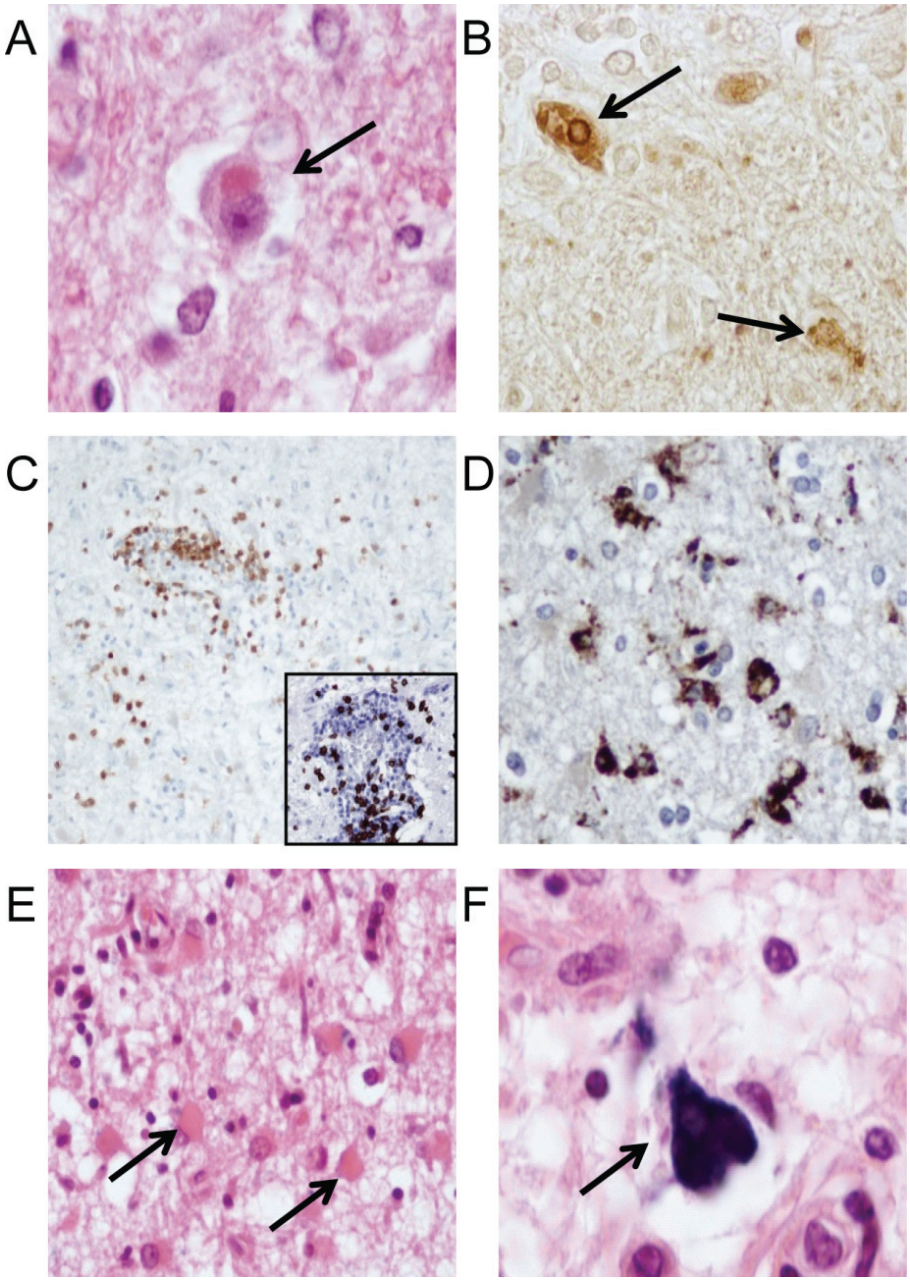


Figure 5-2. Neuropathological analyses. A, Negri body (arrow) in the pons, stained with hematoxylin and eosin (H&E). B, Rabies nucleocapsid protein immunoreactivity (arrows) in the basal ganglia; C, CD3-immunopositive lymphocytes in white matter with CD8-immunoreactive lymphocytes (inset); D, CD68-immunopositive macrophages and microglia in the basal ganglia; E, Abundant hypertrophied astrocytes (arrows) in the white matter, stained with H&E; F, Ferruginated neuron (arrow) in the cerebellum, stained with H&E.

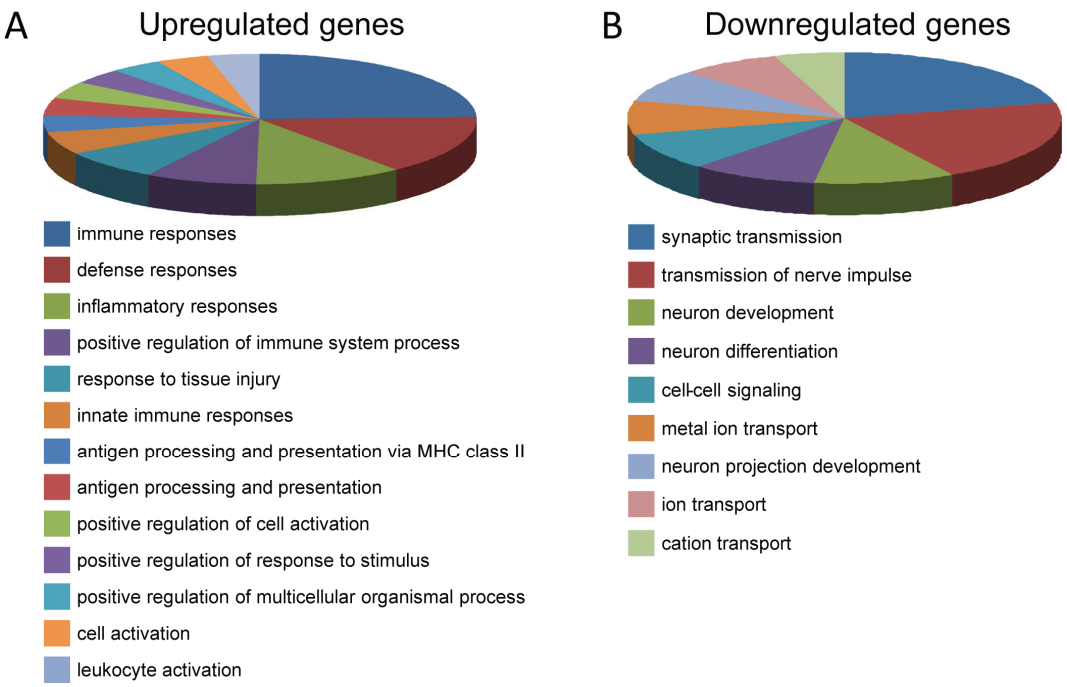


Figure 5-3 A and B. Host and viral gene expression in rabies and non-rabies brains. Deep sequencing of brain white matter-derived cDNA showed up-regulated genes (A) and down-regulated genes (B) in the rabies-infected brain compared to non-rabies brains.

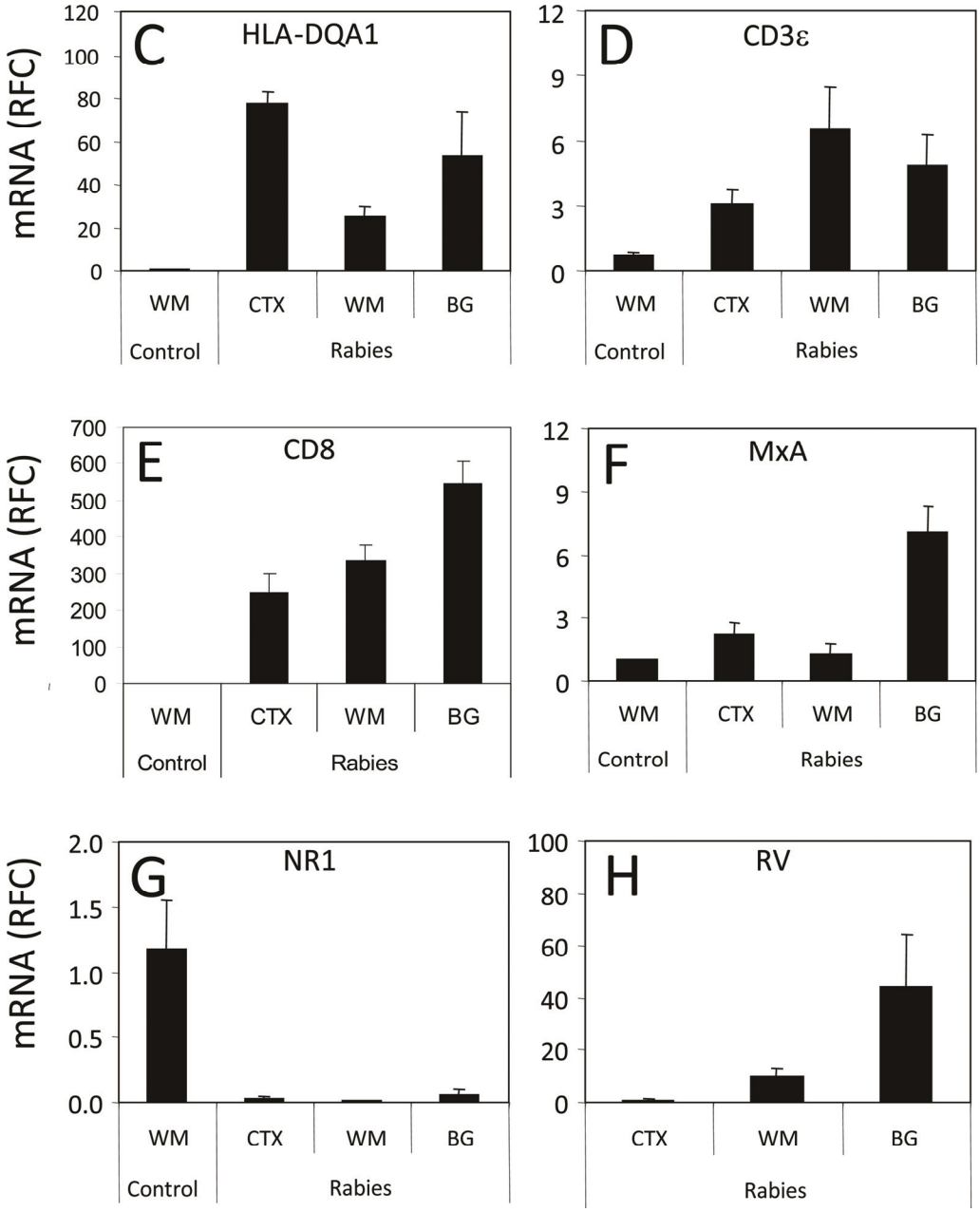


Figure 5-3C-H. Host and viral gene expression in rabies and non-rabies brains. Transcript levels were quantified, relative to GAPDH transcript abundance, in the rabies-infect brain and compared to non-rabies brains. CTX, cerebral cortex; WM, white matter; BG, basal ganglion. (C) *HLA-DQA1*. (D) *CD3ε*. (E) *CD8*. (F) *MxA*. (G) *NR1*. (H) *Rabies virus*.

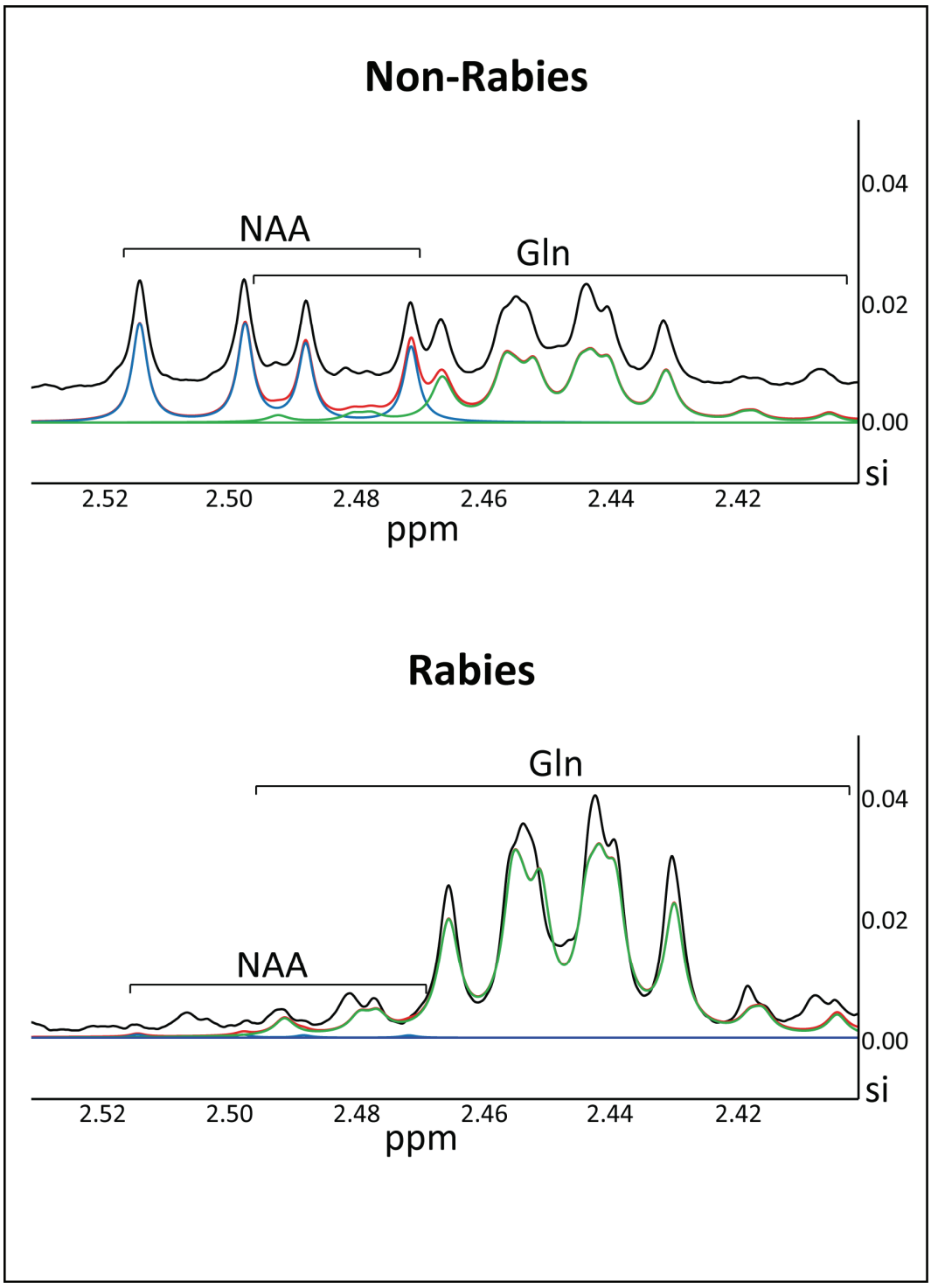


Figure 5-4A. ^{1}H-NMR derived metabolic profiles. ^{1}H-NMR spectra were obtained from cerebral cortex. Peaks from γ and β protons are shown for glutamine and NAA, respectively. Black line, spectrum; green line, glutamine; blue line, NAA; red line, sum of quantified metabolites.

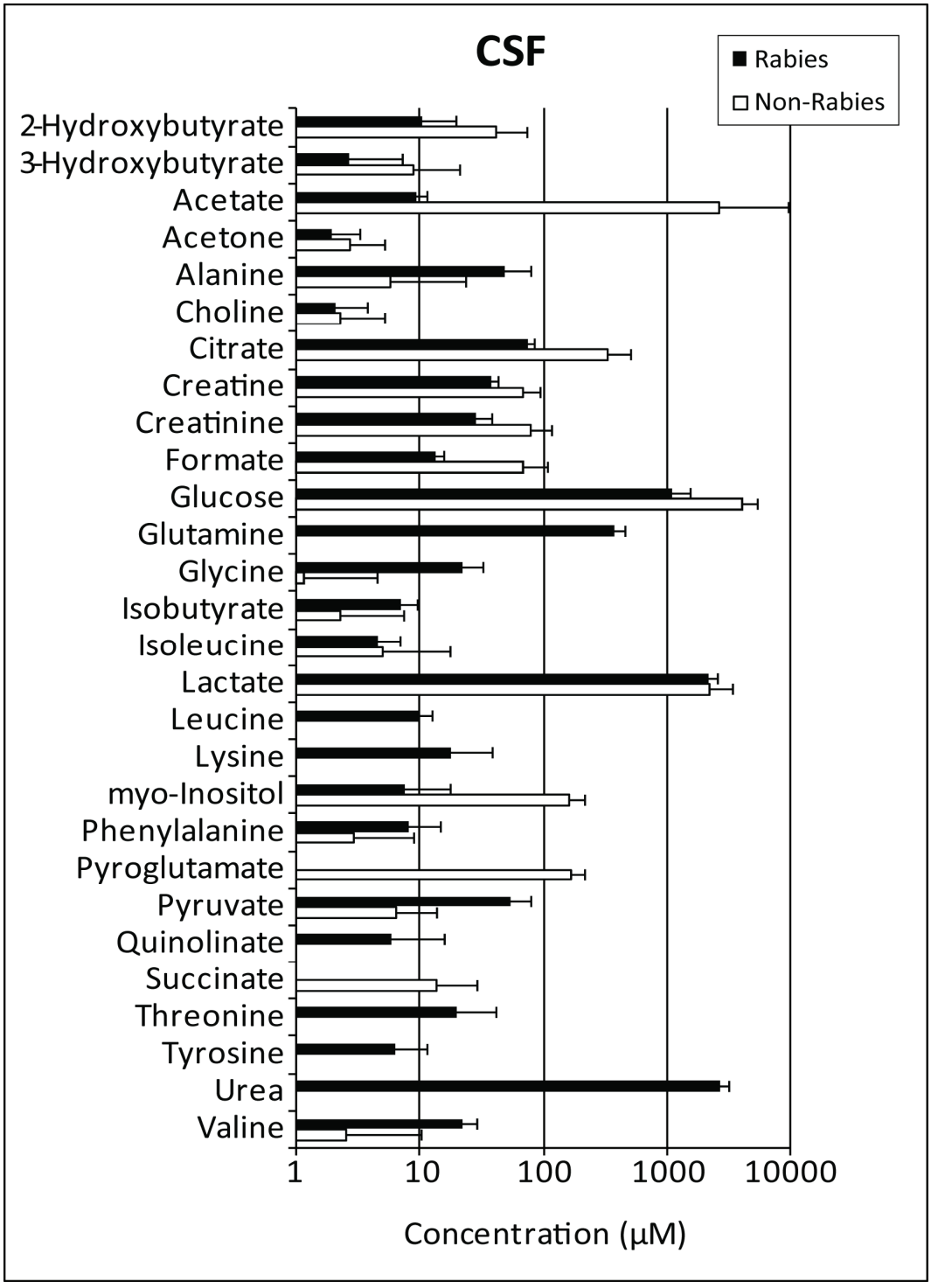


Figure 5-4B. ^{1}H-NMR derived metabolic profiles. Serial CSF specimens were collected during rabies virus infection and the averages of individual metabolite levels are shown, black, compared to an average of non-rabies (n=8) CSF samples, white.

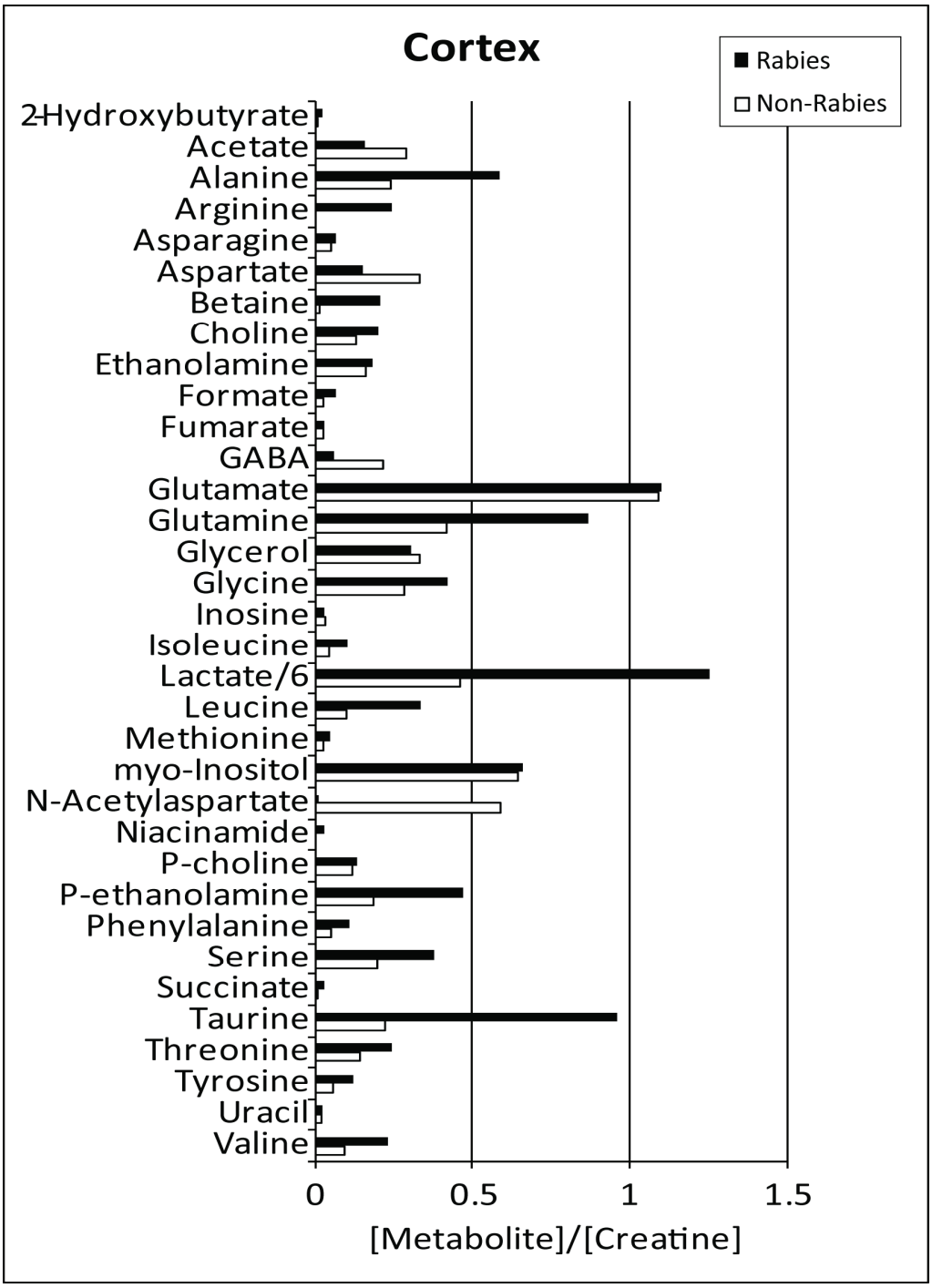


Figure 5-4C. ^{1}H-NMR derived metabolic profiles. Cerebral cortex samples (n=3) were analyzed and averaged from the rabies-infected patient, black, and compared to samples (n=3) from a non-rabies patient, white. Metabolite concentrations were normalized to matched creatine concentration. Lactate concentrations are divided by 6 for presentation purposes.

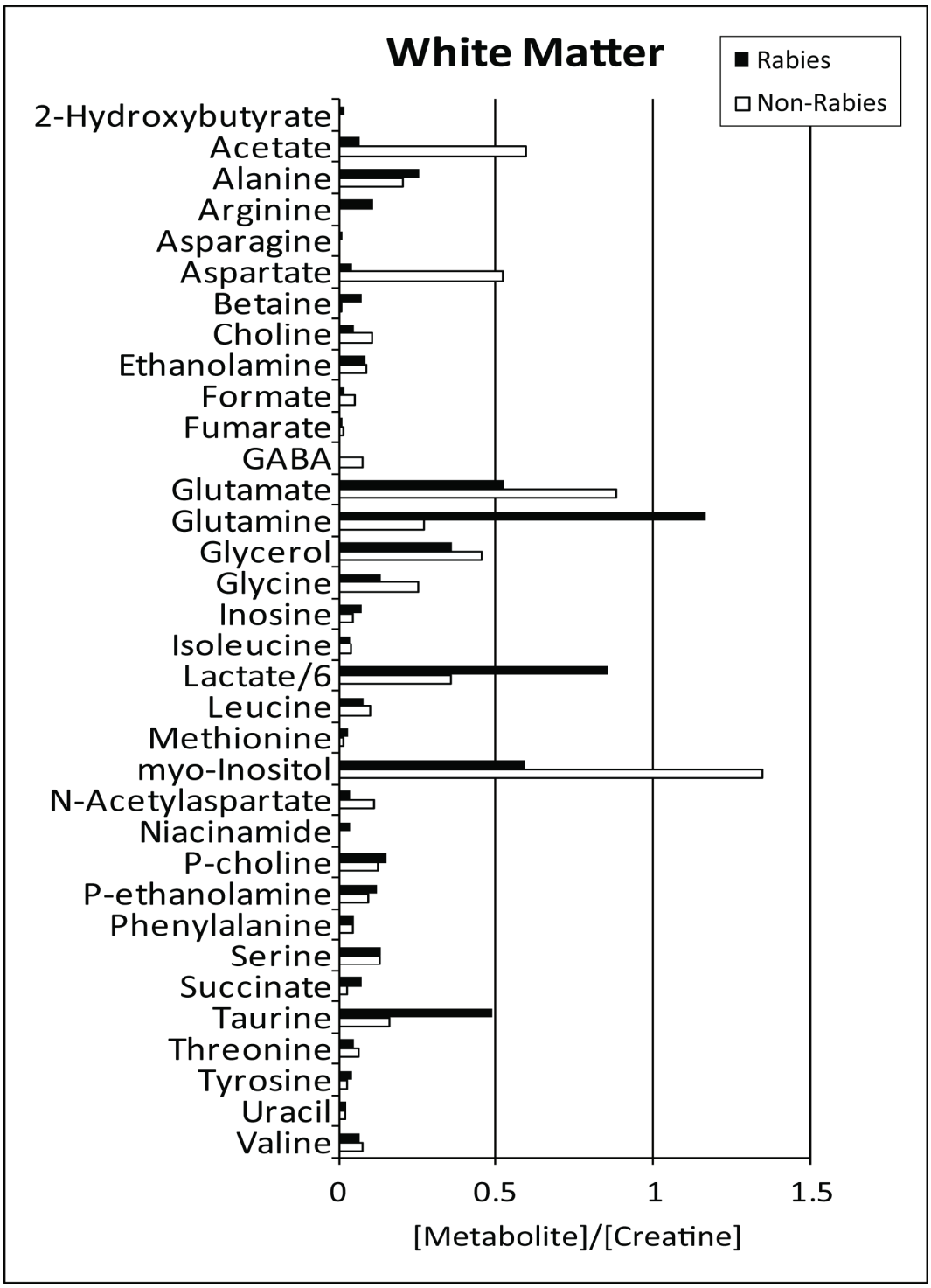


Figure 5-4D. ^{1}H-NMR derived metabolic profiles. White matter samples (n=3) were analyzed and averaged from the rabies-infected patient, black, and compared to those of the non-rabies patient, white. Metabolite concentrations were normalized to matched creatine concentration. Lactate concentrations are divided by 6 for presentation purposes.

References

Adamson PB (1977) The spread of rabies into Europe and the probable origin of this disease in antiquity. J R Asiat Soc GB Irel 2:140-144

Chang Y, Chen TL, Sheu JR, Chen RM (2005) Suppressive effects of ketamine on macrophage functions. Toxicol Appl Pharmacol 204 (1):27-35

Dupont JR, Earle KM (1965) Human rabies encephalitis. A study of forty-nine fatal cases with a review of the literature. Neurology 15 (11):1023-1034

Fleischer R, Boxwell D, Sherman KE (2004) Nucleoside analogues and mitochondrial toxicity. Clin Infect Dis 38 (8):e79-80

Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA (2009) Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. Genome Res 19 (10):1817-1824

Freire-Garabal M, Belmonte A, Balboa JL, Nunez MJ (1992) Effects of midazolam on T-cell immunosuppressive response to surgical stress in mice. Pharmacol Biochem Behav 43 (1):85-89

Galley HF, Dubbels AM, Webster NR (1998) The effect of midazolam and propofol on interleukin-8 from human polymorphonuclear leukocytes. Anesth Analg 86 (6):1289-1293

Hagar HH (2004) The protective effect of taurine against cyclosporine A-induced oxidative stress and hepatotoxicity in rats. Toxicol Lett 151 (2):335-343

Hanefeld FA, Brockmann K, Pouwels PJ, Wilken B, Frahm J, Dechent P (2005) Quantitative proton MRS of Pelizaeus-Merzbacher disease: evidence of dys- and hypomyelination. Neurology 65 (5):701-706

Hattwick MA, Rubin RH, Music S, Sikes RK, Smith JS, Gregg MB (1974) Postexposure rabies prophylaxis with human rabies immune globulin. JAMA 227 (4):407-410

Hemachudha T, Sunsaneewitayakul B, Desudchit T, Suankratay C, Sittipunt C, Wacharapluesadee S, Khawplod P, Wilde H, Jackson AC (2006a) Failure of therapeutic coma and ketamine for therapy of human rabies. J Neurovirol 12 (5):407-409

Hemachudha T, Wacharapluesadee S, Laothamatas J, Wilde H (2006b) Rabies. Curr Neurol Neurosci Rep 6 (6):460-468

Hu WT, Willoughby RE, Jr., Dhonau H, Mack KJ (2007) Long-term follow-up after treatment of rabies by induction of coma. N Engl J Med 357 (9):945-946

Hunter M, Johnson N, Hedderwick S, McCaughey C, Lowry K, McConville J, Herron B, McQuaid S, Marston D, Goddard T, Harkess G, Goharriz H, Voller K, Solomon T, Willoughby RE, Fooks AR (2010) Immunovirological correlates in human rabies treated with therapeutic coma. J Med Virol 82 (7):1255-1265

Jackson AC (1991) Biological basis of rabies virus neurovirulence in mice: comparative pathogenesis study using the immunoperoxidase technique. J Virol 65 (1):537-540

Jackson AC (2005) Recovery from rabies. N Engl J Med 352 (24):2549-2550.

Jackson AC (2007a) Human disease. In: Jackson AC, Wunner WH (eds) Rabies. 2nd edn. Elsevier Ltd., Amsterdam; London, pp 309-340

Jackson AC (2007b) Pathogenesis. In: Jackson AC, Wunner WH (eds) Rabies. 2nd edn. Elsevier Ltd., Amsterdam; London, pp 341-381

Jackson AC (2008) Rabies. Neurol Clin 26 (3):717-726

Jackson AC, Kammouni W, Zherebitskaya E, Fernyhough P (2010) Role of oxidative stress in rabies virus infection of adult mouse dorsal root ganglion neurons. J Virol 84 (9):4697-4705

Jackson AC, Warrell MJ, Rupprecht CE, Ertl HC, Dietzschold B, O'Reilly M, Leach RP, Fu ZF, Wunner WH, Bleck TP, Wilde H (2003) Management of rabies in humans. Clin Infect Dis 36 (1):60-63

Jogai S, Radotra BD, Banerjee AK (2000) Immunohistochemical study of human rabies. Neuropathology 20 (3):197-203

Johnson T, Nath A (2011) Immune reconstitution inflammatory syndrome and the central nervous system. Curr Opin Neurol 24 (3):284-290

Keilhoff G, Wolf G, Stastny F (1991) Effects of MK-801, ketamine and alaptide on quinolinate models in the maturing hippocampus. Neuroscience 42 (2):379-385

Kornhuber J, Weller M, Schoppmeyer K, Riederer P (1994) Amantadine and memantine are NMDA receptor antagonists with neuroprotective properties. J Neural Transm Suppl 43:91-104

Legris T, Massad M, Purgus R, Vacher-Coponat H, Ranque S, Girard N, Berland Y, Moal V (2010) Immune reconstitution inflammatory syndrome mimicking relapsing cryptococcal meningitis in a renal transplant recipient. Transpl Infect Dis:303-308

Lockhart BP, Tordo N, Tsiang H (1992) Inhibition of rabies virus transcription in rat cortical neurons with the dissociative anesthetic ketamine. Antimicrob Agents Chemother 36 (8):1750-1755

Maier T, Schwarting A, Mauer D, Ross RS, Martens A, Kliem V, Wahl J, Panning M, Baumgarte S, Muller T, Pfefferle S, Ebel H, Schmidt J, Tenner-Racz K, Racz P, Schmid M, Struber M, Wolters B, Gotthardt D, Bitz F, Frisch L, Pfeiffer N, Fickenscher H, Sauer P, Rupprecht CE, Roggendorf M, Haverich A, Galle P, Hoyer J, Drosten C (2010) Management and outcomes after multiple corneal and solid organ transplantations from a donor infected with rabies virus. Clin Infect Dis 50 (8):1112-1119

Manning SE, Rupprecht CE, Fishbein D, Hanlon CA, Lumlertdacha B, Guerra M, Meltzer MI, Dhankhar P, Vaidya SA, Jenkins SR, Sun B, Hull HF (2008) Human rabies prevention--United States, 2008: recommendations of the Advisory Committee on Immunization Practices. MMWR Recomm Rep 57 (RR-3):1-28

McCarthy M, Nath A (2010) Neurologic consequences of the immune reconstitution inflammatory syndrome (IRIS). Curr Neurol Neurosci Rep 10 (6):467-475

McCombe JA, Auer RN, Maingat FG, Houston S, Gill MJ, Power C (2009) Neurologic immune reconstitution inflammatory syndrome in HIV/AIDS: outcome and epidemiology. Neurology 72 (9):835-841

McDermid RC, Saxinger L, Lee B, Johnstone J, Gibney RT, Johnson M, Bagshaw SM (2008) Human rabies encephalitis following bat exposure: failure of therapeutic coma. CMAJ 178 (5):557-561

The Milwaukee Protocol version 3.1. Medical College of Wisconsin. http://www.mcw.edu/rabies. Accessed July 13 2011

Noorbakhsh F, Ramachandran R, Barsby N, Ellestad KK, LeBlanc A, Dickie P, Baker G, Hollenberg MD, Cohen EA, Power C (2010) MicroRNA profiling reveals new aspects of HIV neurodegeneration: caspase-6 regulates astrocyte survival. FASEB J 24 (6):1799-1812

Parker R, McKay D, Hawes C, Daly P, Bryce E, Doyle P, Moore W, McKenzie I, Roscoe D, Weatherill S, Skowronski DM, Petric M, Pielak K, Naus M (2003) Human rabies, British Columbia-January 2003. Can Commun Dis Rep 29 (16):137-138

Perl D, Good P (1991) The pathology of rabies in the central nervous system. In: Baer G (ed) The natural history of rabies. 2nd edn. CRC Press, Boca Raton, Florida, pp 163-190

Power C, Henry S, Del Bigio MR, Larsen PH, Corbett D, Imai Y, Yong VW, Peeling J (2003) Intracerebral hemorrhage induces macrophage activation and matrix metalloproteinases. Ann Neurol 53 (6):731-742

Reinke SN, Hu X, Sykes BD, Lemire BD (2010) *Caenorhabditis elegans* diet significantly affects metabolic profile, mitochondrial DNA levels, lifespan and brood size. Mol Genet Metab 100 (3):274-282

Rubin J, David D, Willoughby RE, Jr., Rupprecht CE, Garcia C, Guarda DC, Zohar Z, Stamler A (2009) Applying the Milwaukee protocol to treat canine rabies in Equatorial Guinea. Scand J Infect Dis 41 (5):372-375

Schmiedel S, Panning M, Lohse A, Kreymann KG, Gerloff C, Burchard G, Drosten C (2007) Case report on fatal human rabies infection in Hamburg, Germany, March 2007. Euro Surveill 12 (5):E070531-070535

Smith JS, Fishbein DB, Rupprecht CE, Clark K (1991) Unexplained rabies in three immigrants in the United States. A virologic investigation. N Engl J Med 324 (4):205-211

Superti F, Seganti L, Pana A, Orsi N (1985) Effect of amantadine on rhabdovirus infection. Drugs Exp Clin Res 11 (1):69-74

Szeto SS, Reinke SN, Sykes BD, Lemire BD (2010) Mutations in the *Saccharomyces cerevisiae* succinate dehydrogenase result in distinct metabolic phenotypes revealed through ^{1}H NMR-based metabolic footprinting. J Proteome Res 9 (12):6729-6739

Takahashi K, Ohyabu Y, Solodushko V, Takatani T, Itoh T, Schaffer SW, Azuma J (2003) Taurine renders the cell resistant to ischemia-induced injury in cultured neonatal rat cardiomyocytes. J Cardiovasc Pharmacol 41 (5):726-733

Tsiang H, Ceccaldi PE, Ermine A, Lockhart B, Guillemer S (1991) Inhibition of rabies virus infection in cultured rat cortical neurons by an N-methyl-D-aspartate noncompetitive antagonist, MK-801. Antimicrob Agents Chemother 35 (3):572-574

Varughese P (2000) Human rabies in Canada--1994-2000. Can Commun Dis Rep 26 (24):210-211

Vergote D, Butler GS, Ooms M, Cox JH, Silva C, Hollenberg MD, Jhamandas JH, Overall CM, Power C (2006) Proteolytic processing of SDF-1alpha reveals a change in receptor specificity mediating HIV-associated neurodegeneration. Proc Natl Acad Sci U S A 103 (50):19182-19187

Warrell DA (1976) The clinical picture of rabies in man. Trans R Soc Trop Med Hyg 70 (3):188-195

Weli SC, Scott CA, Ward CA, Jackson AC (2006) Rabies virus infection of primary neuronal cultures and adult mice: failure to demonstrate evidence of excitotoxicity. J Virol 80 (20):10270-10273

Willoughby RE, Jr., Tieves KS, Hoffman GM, Ghanayem NS, Amlie-Lefond CM, Schwabe MJ, Chusid MJ, Rupprecht CE (2005) Survival after treatment of rabies with induction of coma. N Engl J Med 352 (24):2508-2514

Wilson RD, Priano LL, Traber DL, Sakai H, Daniels JC, Ritzmann SE (1971) An investigation of possible immunosuppression from ketamine and 100 percent oxygen in normal children. Anesth Analg 50 (3):464-470

World Health Organization (2005) WHO expert consultation on rabies: first report. Geneva, Switzerland.

CHAPTER 6

General Discussion

Overview

The original objective of this thesis research was to explore the molecular mechanisms involved in mitochondrial disease using a metabolomics approach. We postulated that, by investigating the metabolic effects of mitochondrial dysfunction, we would gain a better understanding of disease pathogenesis. As the metabolome is extremely sensitive to both genetic and environmental factors, we chose to use two common laboratory models: the nematode *C. elegans* and the yeast *S. cerevisiae*. By performing our studies on genetically well-defined systems under highly-controlled growth conditions, we predicted that any significant metabolic changes would be the result of disease state and not of metabolic noise. Our results reveal, however, that an appreciable amount of biological variability exists even within simple model systems. The extremely sensitive nature of the metabolome to environmental changes such as diet was also revealed. We discovered that dietary changes not only significantly alter metabolism, but also mtDNA copy number and other phenotypic characteristics in *C. elegans*. Indeed, these results highlight the importance of proper metabolomic experimental design; this chapter will discuss these considerations.

As discussed in Chapter 1, mitochondrial disease etiology is extremely complex; metabolism is only one of several factors involved in pathogenesis. Our results highlight the difficulties of studying metabolism in such a complex system. Despite these difficulties, a strong understanding of how metabolism is compromised is essential for understanding and effectively treating mitochondrial disease. One of the goals of the metabolomics community is to develop personalized medicine strategies. Mitochondrial disease is a broad term that

encompasses many disorders; in Chapter 3 we show that two mutants of the same gene have both metabolic similarities and differences. These results demonstrate the need for a personalized medicine approach when investigating mitochondrial disease. In addition to studying mitochondrial disease, we also had the opportunity to examine the pathological, molecular, and metabolic effects of a rabies infection following an aggressive, and controversial therapeutic protocol; our findings reveal a potential adverse and fatal consequence of this procedure. This chapter also discusses future metabolomic and personalized medicine approaches that could be applied to the advancement of treatments for mitochondrial disorders and rabies.

Design of Experiment (DoE) Considerations

The results presented in Chapters 2 and 3 emphasize the necessity of careful consideration during metabolomic experimental design. As discussed in Chapter 2, measured metabolic variation is the product of both biological and experimental/analytical variation. Analytical variation in our experiments was assessed and found not to be a major contributor to measured variability; this finding is consistent with previous studies on model systems (Crews et al. 2009; Fiehn et al. 2000; Parsons et al. 2009). However, the possibility of introducing a considerable amount of experimental variation always exists; if authors do not report their methods, it is impossible for readers to determine data integrity. For this reason, minimum reporting standards for metabolomics experiments have been proposed by the Metabolomics Society (Griffin et al. 2007; van der Werf et al. 2007). By the standards set forth in these reports, all metabolomics studies should include a discussion of both factors specific to metabolomic experiments

and general aspects that determine experimental outcome. Metabolomics-specific considerations include details of metabolism quenching, metabolite extraction, sample storage, and data normalization. General aspects affecting experimental outcome include experimental design, biosource, growth environment, and treatments during growth (van der Werf et al. 2007). Griffin and colleagues (2007) expand their report to include considerations for mammalian experiments. As there are numerous factors that affect experimental outcome and degree of metabolic variation, design of experiment considerations should be closely evaluated and reported in all metabolomics studies.

Chapter 3 revealed the complex nature of performing metabolomic studies on the nematode *C. elegans*. Future metabolomic investigations using this system must be carefully planned. Following our discovery of the impact that diet has on the nematode metabolome, mtDNA copy number, and phenotype, we chose to employ an axenic liquid culture medium (Clegg et al. 2002). However, this medium utilized twelve different solutions, ten of which were prepared in lab. Skim milk was also added as a food source; this was prepared from skim milk powder. Media was prepared fresh prior to every use. As there were several possible steps in which experimental error could be introduced, the data was more variable than data obtained from standard NGM media seeded with *E. coli*; we were unable to draw any meaningful conclusions from its use (data not shown). The axenic media was also more difficult to work with as it was susceptible to contamination. Indeed, there is no perfect solution for investigating metabolism in *C. elegans*. Single metabolite tracer and flux studies must be performed under axenic conditions to ensure that only worm metabolism is measured. However, tracer studies that have a more general goal in mind (such

as ubiquitous ^{15}N labelling) would be exceedingly expensive if performed under axenic conditions. Worm life cycle must also be carefully considered when performing metabolomic studies. The nematode transitions through four larval stages prior to becoming a gravid adult; metabolic differences in these stages have been documented (O'Riordan and Burnell 1989, 1990; Braeckman et al. 2009). Thus, the metabolomes of mixed cultures will undoubtedly differ from synchronized cultures. Although *C. elegans* is an excellent laboratory model system, it poses a number of challenges for metabolomics studies; these challenges and their correlation to hypothesis must be carefully considered during the DoE process.

It is extremely important to design good 'wet lab' experiments; however it is equally important to design good data modelling and statistical experiments. As described in Chapter 1, PCA and PLS-DA are the most commonly used of many data analysis methods in the metabolomics community. The caveat, however, is that most of these researchers are also biologists. To quote Efron and Tibshirani (1993): "Left to our own devices, ... we are all too good at picking out non-existent patterns that happen to suit our purposes"; for those of us who are not mathematicians, it is easy to ignorantly choose the data modelling and statistical methods that best suit our needs. As discussed in Chapter 1, it is easy to generate spurious models that appear to discriminate the data well. For this reason, all authors reporting metabolomics data should familiarize themselves with the various data analysis, validation, and reporting procedures. Broadhurst and Kell (2006) and Goodacre and colleagues (2007) discuss the various data analysis methods and describe how to avoid reporting false discoveries. By becoming educated in this area and seeking advice of those familiar in this area,

authors ensure that they design and report data models that best explain the data instead of models that best explain scientific motivation.

Experimental design is the most important process in any metabolomics study. Carefully considering each step of the biological, analytical, and mathematical experiments and then reporting these considerations is essential for maintaining a high level of integrity. In a field of study where a multitude of factors can affect measured results, and several approaches can be taken to analyze and model the results, maintaining integrity is of utmost importance.

Metabolomics and Personalized Medicine Approaches

Mitochondrial Disease

As previously mentioned, the results in Chapter 3 reveal the effects that diet has on metabolism, as well as molecular and phenotypic characteristics. Although these findings complicate our original research objective, they do emphasize the magnitude of the effect that diet can have on overall health and its potentially important role in treating mitochondrial disease. Several dietary interventions, including macronutrient-specific diets and treatment with vitamins and cofactors, have been used and studied in patients affected by mitochondrial disease (Schiff et al. 2011). Although in isolated cases there have been anecdotal reports of improvement, scientific evidence supporting a universal treatment is lacking. This is not surprising, and should be expected, considering the complexity and heterogeneity of mitochondrial disease pathogenesis and the sensitivity of the metabolome. Future investigations into dietary intervention should be performed both at a more basic level to explore how dietary compounds affect cellular

metabolism, and at a more personalized level to determine how these treatment options affect the patient as a whole.

At the basic research level, metabolic footprinting studies of cultured human cell lines would provide a wealth of knowledge as to how different nutrients ameliorate the unbalanced metabolic state. Tracer and flux studies could easily be performed to delineate affected pathways. The results presented in Chapters 3 and 4 reveal that amino acid (nitrogen) metabolism is profoundly affected by energy metabolism. The specific pathways involved and the pathological consequences associated with their alteration need to be identified. As there are countless mutations involved in mitochondrial dysfunction, many cell lines will need to be established for this purpose. Although such a study would be extremely resource-intensive, it would also provide invaluable information for directing treatment of mitochondrial disease. Metabolic footprinting also provides the advantage of being an efficient method for characterizing tens (or even hundreds depending on analytical platform used) of metabolites simultaneously. In metabolomic studies, especially in those that are hypothesis-driven, it is easy to focus on a few 'important' metabolites. Our results in Chapter 4 contrarily suggest that the entire metabolome (or what can be identified) is greater than the sum of its individual components, as we were able to correlate metabotype to growth yield phenotype. Thus, cell culture metabolic footprinting would afford an excellent opportunity to explore the metabolome in both a holistic and pathway-specific manner.

Due to the heterogeneity and extremely individual phenotypes associated with mitochondrial impairment, personalized metabolomic and treatment approaches

must be taken. Because dietary intervention is highly anecdotal, metabolic state needs to be closely monitored during any treatment attempts; this can be achieved through the analysis of biofluids, such as blood and urine. Regardless of whether specific metabolic conclusions can be drawn, making qualitative interpretations such as ‘more’ or ‘less’ metabolically-compromised (as compared to the ‘normal’ metabolic state) may direct treatment options. A drawback to this approach would be patients experiencing a placebo-effect, where the treatment-associated metabolic state would not be reflective of patient improvement. In an ideal research scenario, immortal cell lines would be generated from affected tissue excised from patients; well-defined nutritional, pharmaceutical, and metabolomic footprinting studies could then be performed, taking a truly personalized approach to disease treatment. Although this approach would yield the most meaningful results for patients, it would also be enormously cost, time, and resource expensive, and thus not practically feasible. In addition, this approach would only work well for easily-excised tissues such as skeletal muscle. Dietary intervention could be integral in treating patients with mitochondrial disease; even though it is presently performed in a trial-and-error manner, metabolomic and personalized medicine studies afford the opportunity to understand how diet can be used in the treatment of mitochondrial disease.

Rabies

Following the first successful treatment of clinical rabies, the Milwaukee Protocol provided optimism towards treating this otherwise ubiquitously fatal disease. However, as discussed in Chapter 6, the efficacy of this treatment option has become highly controversial. An investigation must be performed to determine

which component, if any, of the protocol is beneficial in the treatment of rabies. If components of this protocol were found to be effective in treating clinical rabies, it would be interesting to be able to predict patient outcome. The first survivor was 15 years old (Willoughby et al. 2005); another 5-year old patient survived the rabies infection, only to pass away several days later due to complications (Rubin et al. 2009). The ages of these survivors pose an interesting question: do neuro-biology and neuro-chemistry change significantly with age? And, do these changes play a role in disease and treatment outcome in younger people? From a metabolomic point of view, it would be extremely interesting to evaluate brain neuro-chemistry (as reflected in the brain metabolome) during different life stages. However, this type of study poses significant logistical challenges. It is difficult to obtain autopsy-derived brain tissue from a large enough population to hold any statistical power. Further, autopsies are not performed immediately after death; several hours or days may pass before tissue is excised. Complete neuronal death may take several hours (Averback 1980); during this time anaerobic metabolism may occur, affecting the measured metabolic profile. Indeed, the present study poses the same caveat; however, we used a matched control with similar autopsy time to reduce possible error. The profound metabolomic differences that we observed are more indicative of the severe pathological state in the rabies infection than of autopsy time. Despite these obvious logistical difficulties, an investigation into age-related neuro-chemical differences may provide valuable information for prognostic assessment and treatment evolution.

The present study is of particular interest in that inflammation was abundant, despite the fact that neuro-inflammation is not generally noted in clinical rabies

(Perl and Good 1991). This phenomenon has also been recently reported in another unsuccessful attempt of the Milwaukee Protocol (Aramburo et al. 2011). As discussed in Chapter 5, a number of the agents used in the Milwaukee Protocol are immunosuppressive; our results suggest that a severe consequence of treatment, neurological immune reconstitution syndrome (neuroIRIS), ensued. Despite this potential and serious risk, the Milwaukee Protocol is the only available treatment for clinical rabies. Thus, patients who are administered, and then removed from, this treatment should be closely monitored for signs of developing neuroIRIS; however, there is no definitive diagnostic test for this syndrome (Johnson and Nath 2010). In this study, we were able to quantify twenty eight CSF metabolites using ^{1}H-NMR spectroscopy. Using magnetic resonance spectroscopy (MRS), the following metabolites are detectable in the *in vivo* brain: alanine, choline, creatine, glucose, glutamate, glutamine, lactate, myo-inositol, and N-acetylasparate (Soares and Law 2009). The potential for these metabolites to be used as diagnostic markers of neuroIRIS must be assessed; this can be done in two ways. First, the metabolomic signatures of CSF from neuroIRIS patients and matched controls need to be defined. MRS of the brain should also be conducted to determine if any quantifiable metabolites change significantly during an inflammatory state. Second, an exometabolomic study of cultured human neural cells with and without exposure to inflammatory stimuli should be performed. This will provide a controlled environment free from *in vivo* biological variability. Any metabolites that are excreted in high quantities or completely consumed should be cross-referenced with CSF and brain metabolomic data. In addition to MRS providing a potential diagnostic tool for neuroIRIS, it may also be used to monitor patient neurological health during

rabies infection. N-acetylasparate (NAA) is a neuronal marker and detectable by MRS (Soares and Law 2009); monitoring neuronal loss would indicate treatment efficacy.

Our investigation of the present case revealed persistent viral load in the brain despite aggressive treatment, as well as a potentially fatal side effect of treatment. As controversy over the efficacy of the Milwaukee protocol persists, treatment of clinical rabies must be attempted with great caution. However, the metabolomic studies discussed above may provide an opportunity for personalized and more effective treatment options of clinical rabies. An age-specific brain metabolome and serial monitoring of brain NAA levels could lead to greater prognostic accuracy. Establishing an inflammatory metabolic profile could offer an immediate and definitive diagnosis of neuroIRIS, thus allowing for timely pharmaceutical intervention to occur. Although rabies remains almost ubiquitously fatal, the 'young' field of metabolomics will aid in the advancement of more effective treatments of this ancient disease.

Final Remarks

I once heard metabolomics studies be compared to fishing. Both pursuits often involve blind, albeit educated, searches for something significant. The simple truth, however, is that we will not catch fish if we do not go fishing. Metabolomics is still primarily a discovery-driven science rather than a hypothesis-driven one. Although this approach is unconventional, it is extremely useful for identifying biomarkers of disease and providing a metabolic platform for future studies. Despite the fact that most metabolic pathways have been elucidated for many years, we are still relatively naïve about the consequences of these pathways

becoming unbalanced. Metabolomics is the final frontier of the 'omics' sciences. Not only does it complete the systems biology platform but it also provides an efficient and cost-effective 'red-light or green-light' approach for directing research initiatives. Together with genomic, transcriptomic, and proteomic data, metabolomics will continue to offer insight into the molecular mechanisms that underlie disease. This holistic understanding will continue to foster the development of personalized medicine strategies that will one day provide definitive and effective approaches in diagnosing, predicting, and treating human disease.

References

Aramburo A, Willoughby RE, Bollen AW, Glaser CA, Hsieh CJ, Davis SL, Martin KW, Roy-Burman A (2011) Failure of the Milwaukee Protocol in a child with rabies. Clin Infect Dis 53 (6):572-574

Averback P (1980) A study of the rate of cell depletion in solid tissue. J Pathol 130 (3):173-178

Braeckman BP, Houthoofd K, Vanfleteren JR Intermediary Metabolism (Februaru 16, 2009), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.146.1, http://www.wormbook.org

Broadhurst DI, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2 (4):171-196

Clegg ED, LaPenotiere HF, French DY, Szilagyi M Use of CeHR axenic medium for exposure and gene expression studies. In: East Coast Worm Meeting, Durham, New Hampshire, USA, June 14-16 2002.

Crews B, Wikoff WR, Patti GJ, Woo HK, Kalisiak E, Heideker J, Siuzdak G (2009) Variability analysis of human plasma and cerebral spinal fluid reveals statistical significance of changes in mass spectrometry-based metabolomics data. Anal Chem 81 (20):8538-8544

Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall, New York

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol 18 (11):1157-1161

Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, Bessant C, Connor S, Capuani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai R, Sjöström M, Trygg J, Wulfert F (2007) Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 3:231-241

Griffin JL, Nicholls AW, Daykin CA, Heald S, Keun HC, Schuppe-Koistinen I, Griffiths JR, Cheng LL, Rocca-Serra P, Rubtsov DV, Robertson D (2007) Standard reporting requirements for biological samples in metabolomics experiments: mammalian/ *in vivo* experiments. Metabolomics 3:179-188

Johnson T, Nath A (2010) Neurological complications of immune reconstitution in HIV-infected populations. Ann N Y Acad Sci 1184:106-120

O'Riordan VB, Burnell AM (1989) Intermediary metabolism in the dauer larva of the nematode *Caenorhabditis elegans* - 1. Glycolysis, gluconeogenesis, oxidative phosphorylation and the tricarboxylic acid cycle. Comp Biochem Physiol Part B: Biochem Mol Biol 92 (2):233-238

O'Riordan VB, Burnell AM (1990) Intermediary metabolism in the dauer larva of the nematode *Caenorhabditis elegans* - II. The glyoxylate cycle and fatty-acid oxidation. Comp Biochem Physiol Part B: Biochem Mol Biol 95 (1):125-130

Parsons HM, Ekman DR, Collette TW, Viant MR (2009) Spectral relative standard deviation: a practical benchmark in metabolomics. Analyst 134 (3):478-485

Perl D, Good P (1991) The pathology of rabies in the central nervous system. In: Baer G (ed) The natural history of rabies. 2nd edn. CRC Press, Boca Raton, Florida, pp 163-190

Rubin J, David D, Willoughby RE, Jr., Rupprecht CE, Garcia C, Guarda DC, Zohar Z, Stamler A (2009) Applying the Milwaukee Protocol to treat canine rabies in Equatorial Guinea. Scand J Infect Dis 41 (5):372-375

Schiff M, Benit P, Coulibaly A, Loublier S, El-Khoury R, Rustin P (2011) Mitochondrial response to controlled nutrition in health and disease. Nutr Rev 69 (2):65-75

Soares DP, Law M (2009) Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications. Clin Radiol 64 (1):12-21

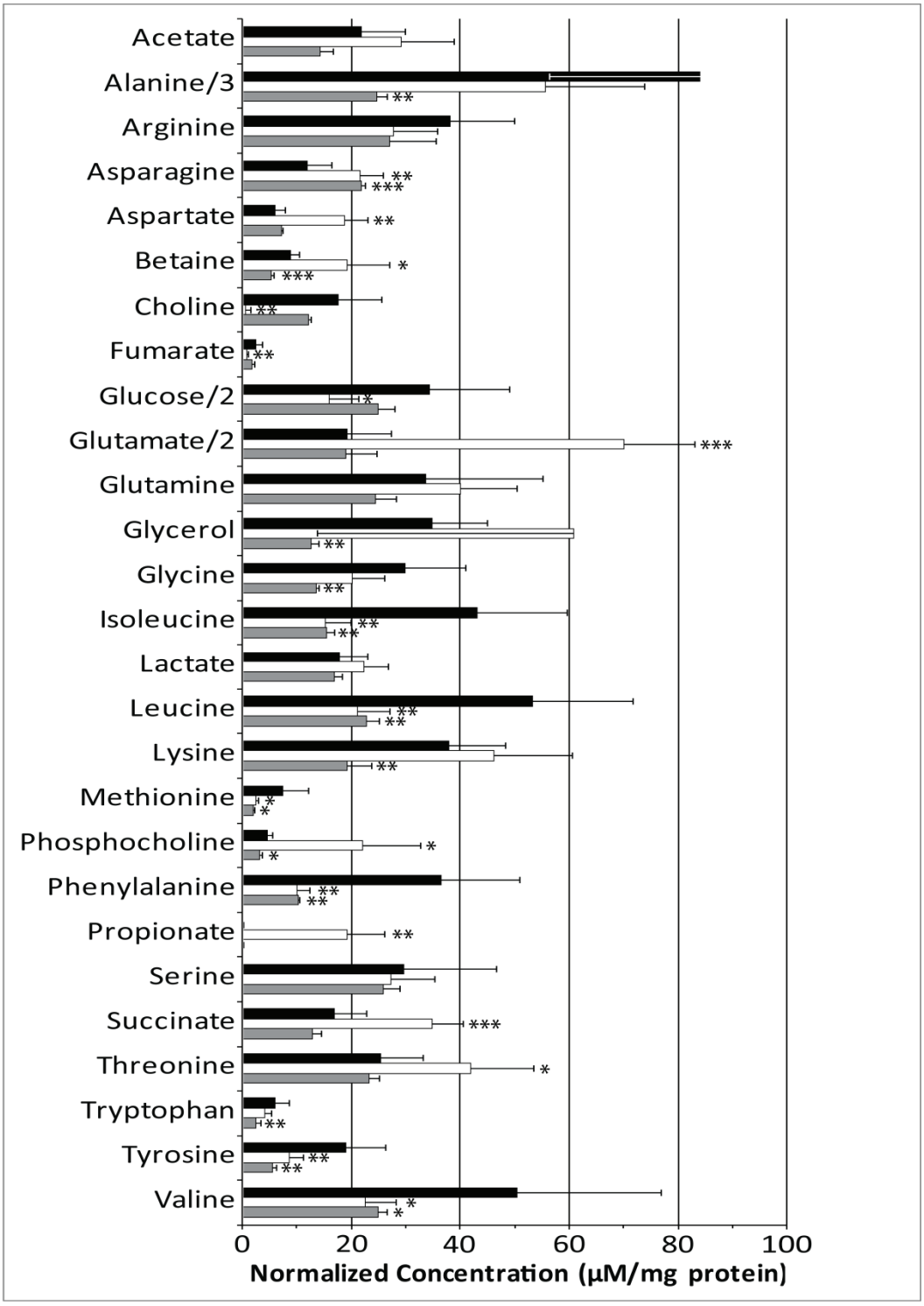
van der Werf MJ, Takors R, Smedsgaard J, Nielsen J, Ferenci T, Portais JC, Wittmann C, Hooks M, Tomassini A, Oldiges M, Fostel J, Sauer U (2007) Standard reporting requirements for biological samples in metabolomics

experiments: microbial and *in vitro* biology experiments. Metabolomics 3:189-194

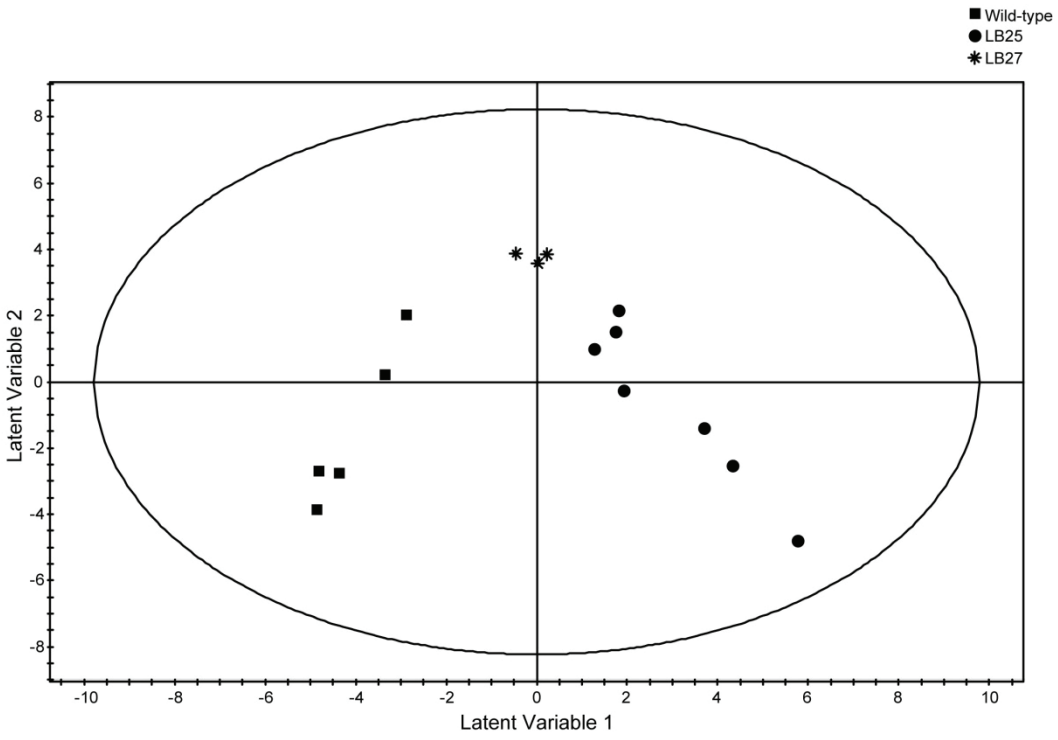
Willoughby RE, Jr., Tieves KS, Hoffman GM, Ghanayem NS, Amlie-Lefond CM, Schwabe MJ, Chusid MJ, Rupprecht CE (2005) Survival after treatment of rabies with induction of coma. N Engl J Med 352 (24):2508-2514

APPENDIX A

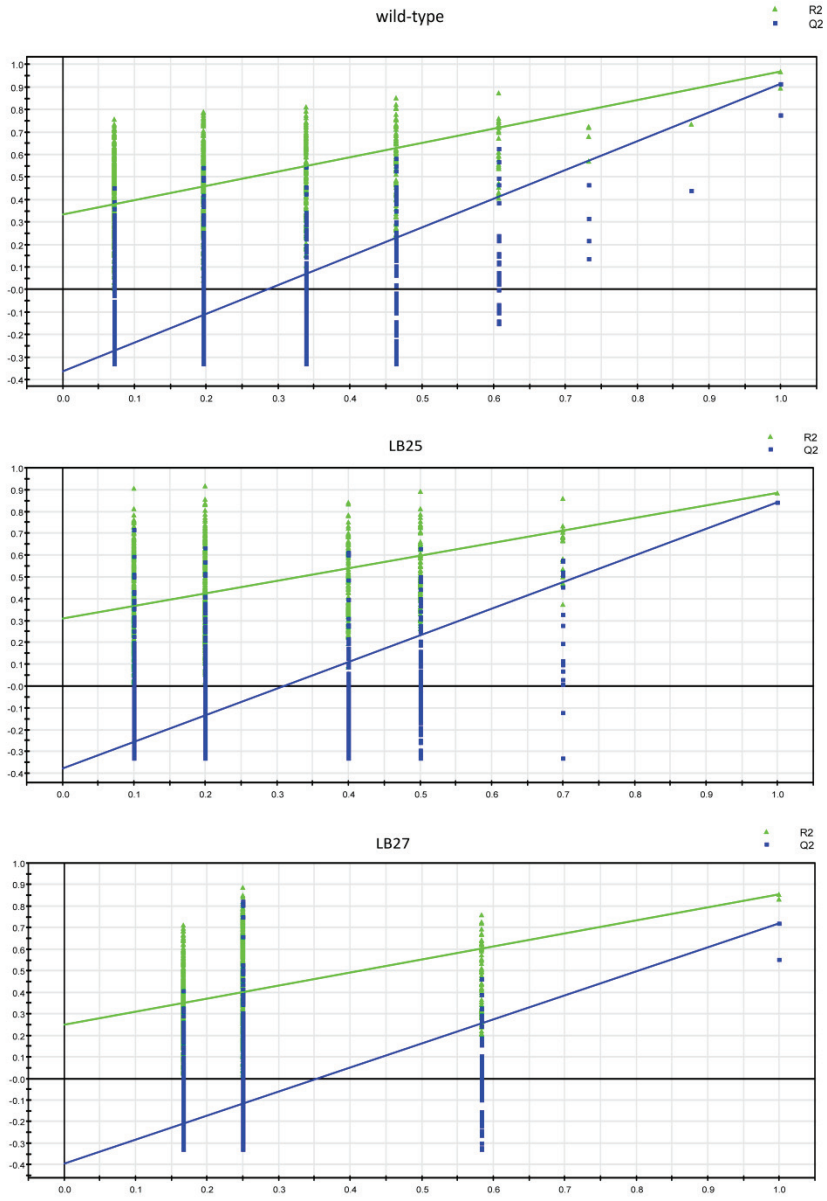
Supplementary Figures



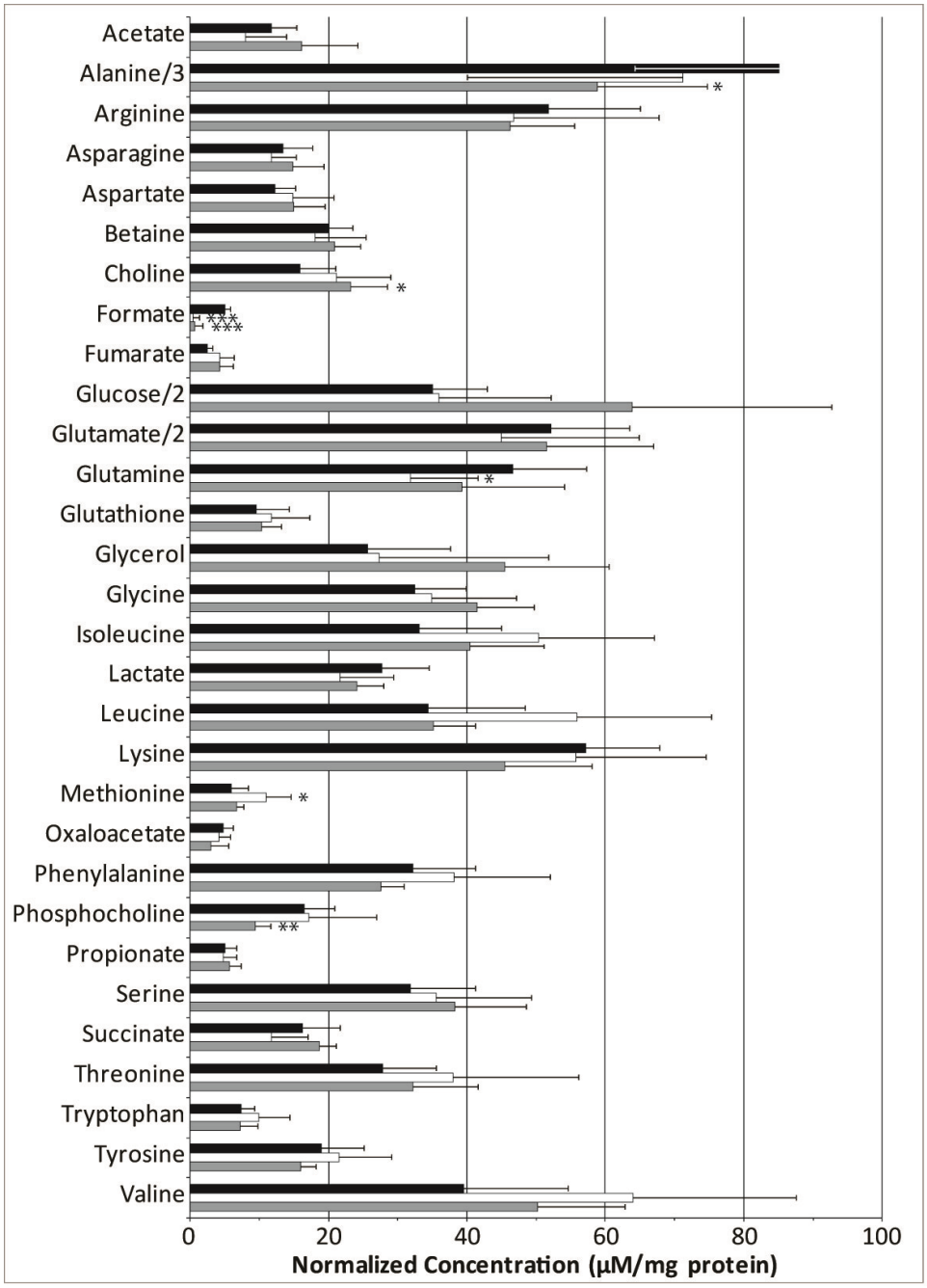
Supplementary Figure 3-1. Metabolic profiles of various *C. elegans* strains fed *E. coli* OP50. Metabolite concentrations were quantified using ^{1}H-NMR spectroscopy and normalized to total worm protein. Black, wild-type; white, LB25; grey LB27. *, $p < 0.05$; **, $p < 0.01$, ***, $p < 0.001$.



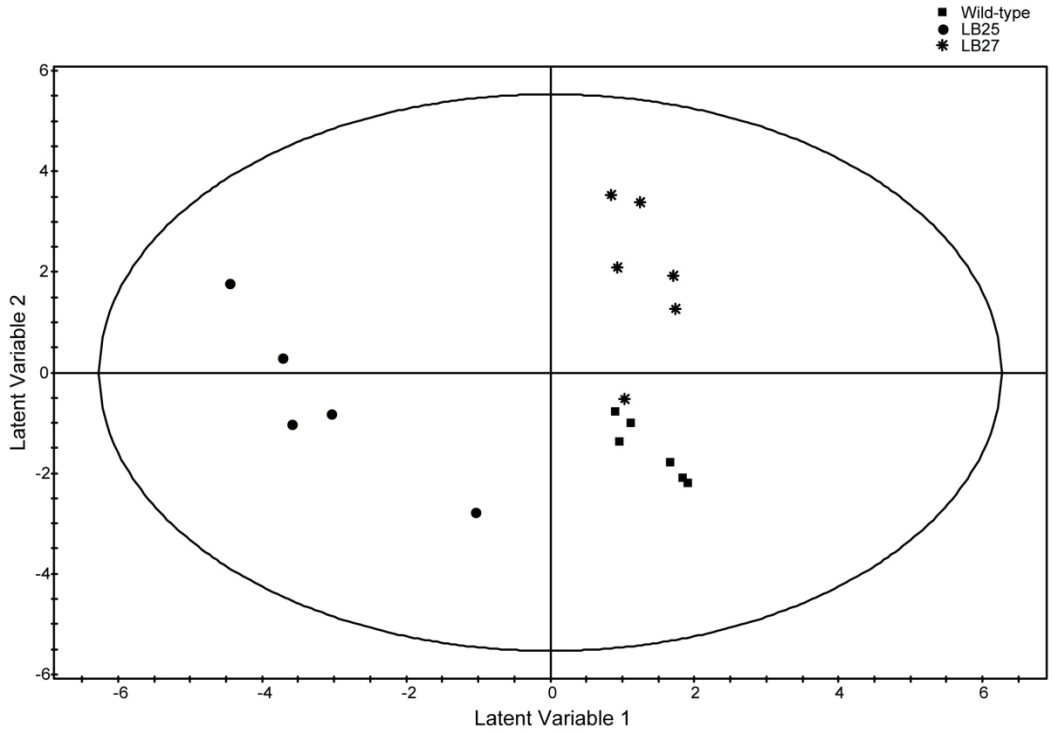
Supplementary Figure 3-2A. PLS-DA model of ^{1}H-NMR derived metabolite profiles of various *C. elegans* strains fed *E. coli* OP50. PLS-DA score plot. Boxes, wild-type (N=7); circles, LB25 (N=5); stars, LB27 (N=3).



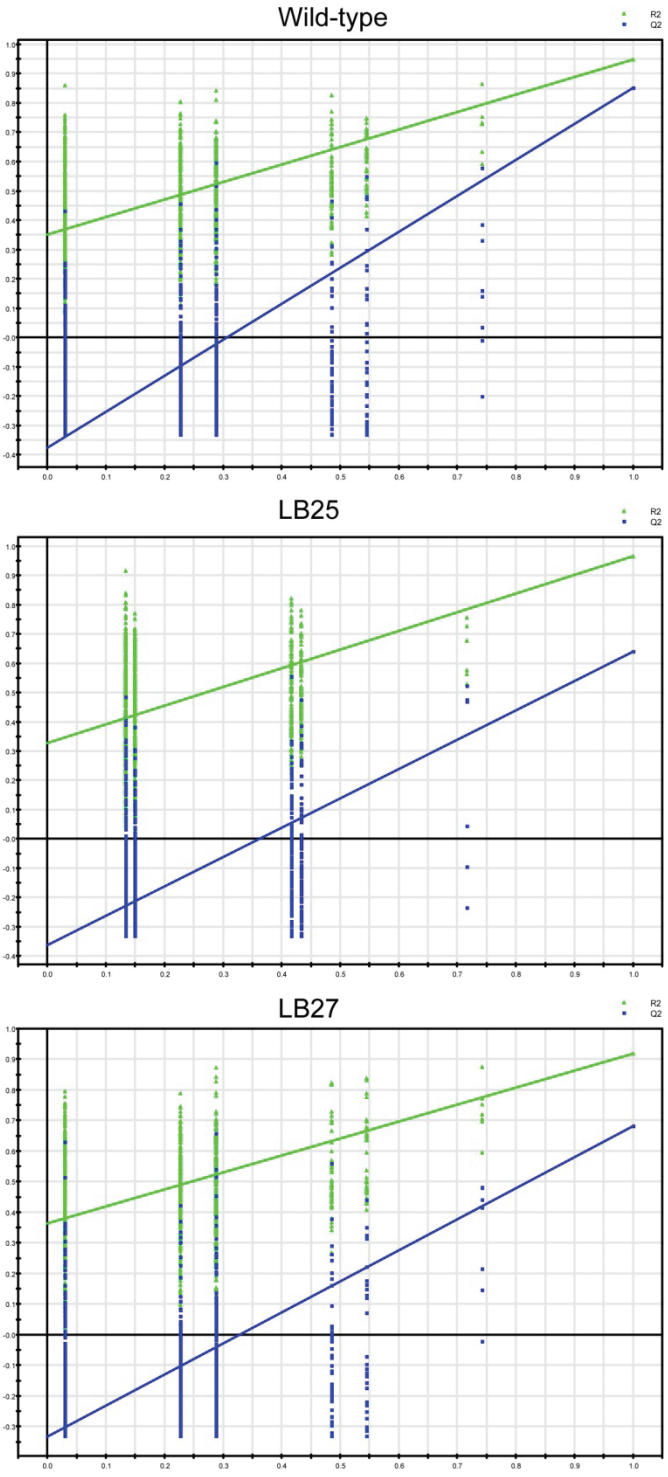
Supplementary Figure 3-2B. PLS-DA model of ^{1}H-NMR derived metabolite profiles of various *C. elegans* strains fed *E. coli* OP50. Validation plots of 999 permutations.



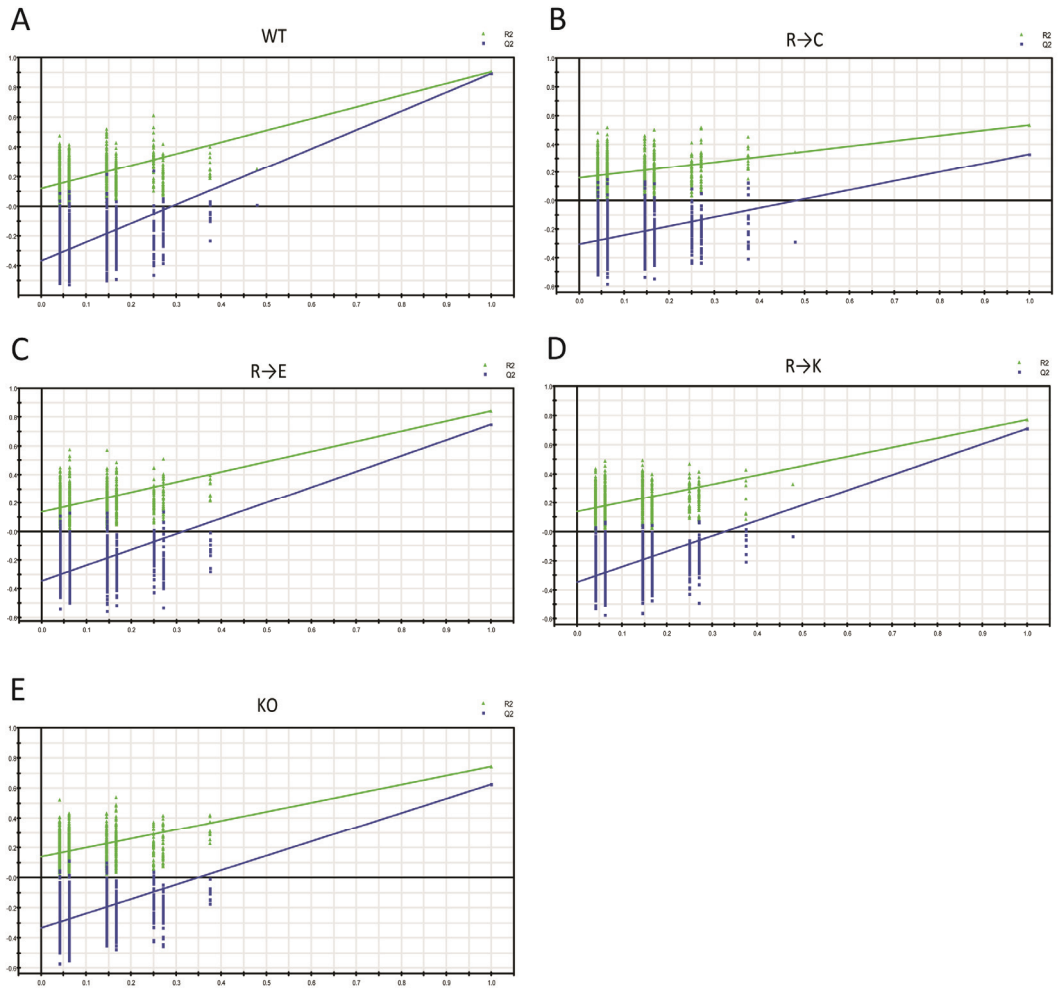
Supplementary Figure 3-3. Metabolic profiles of various *C. elegans* strains fed *E. coli* HT115 L4440. Metabolite concentrations were quantified using ^{1}H-NMR spectroscopy and normalized to total worm protein. Black, wild-type; white, LB25; grey LB27. *, $p < 0.05$; **, $p < 0.01$, ***, $p < 0.001$.



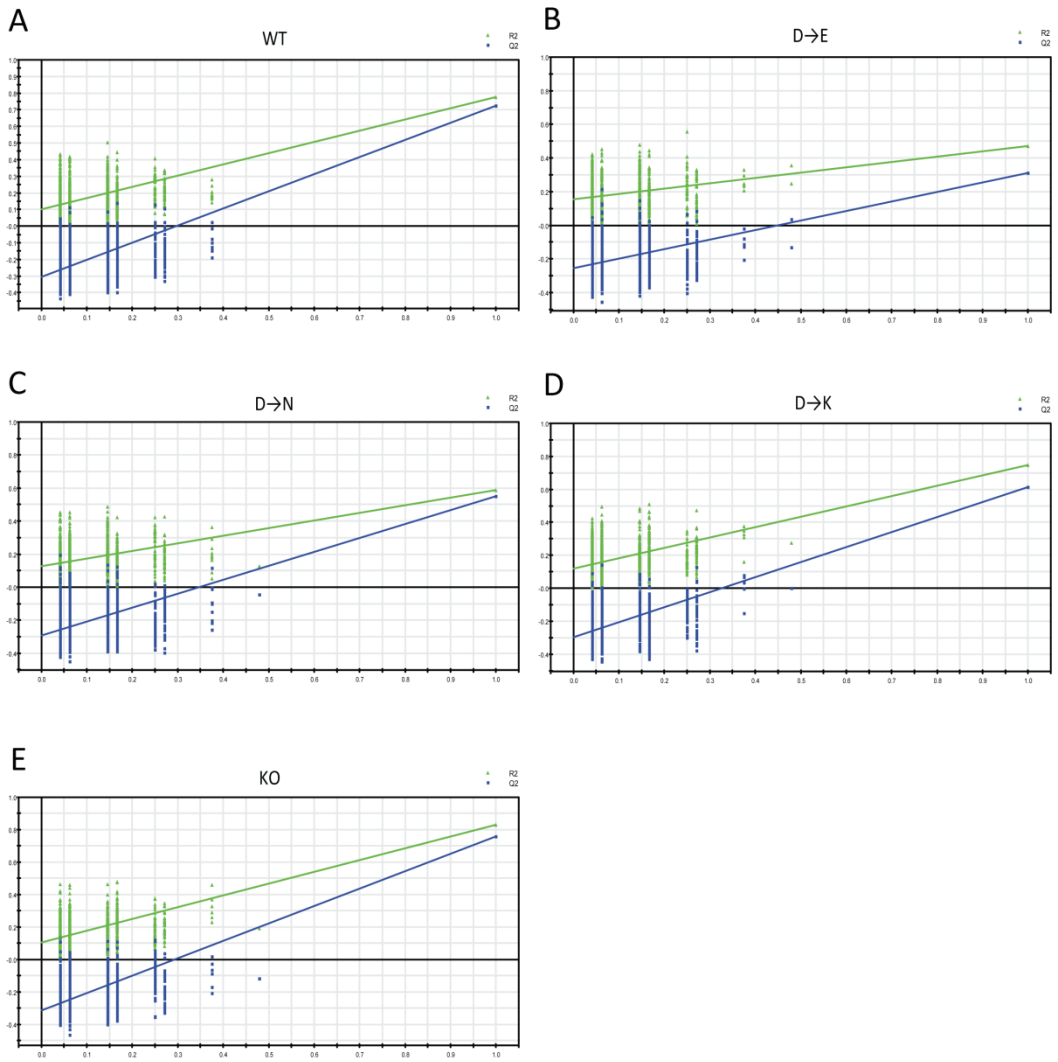
Supplementary Figure 3-4A. PLS-DA model of ^{1}H-NMR derived metabolite profiles of various *C. elegans* strains fed *E. coli* HT115 L4440. PLS-DA score plot. Boxes, wild-type (N=6); circles, LB25 (N=5); stars, LB27 (N=6).



Supplementary Figure 3-4B. PLS-DA model of ^{1}H-NMR derived metabolite profiles of various *C. elegans* strains fed *E. coli* HT115 L4440. Validation plots of 999 permutations.



Supplementary Figure 4-1. Validation plots for *SDH3* model. 999 permutations were performed for each class.



Supplementary Figure 4-2. Validation plots for *SDH4* model. 999 permutations were performed for each class.

APPENDIX B

A glycolytic burst drives glucose induction of global histone acetylation by picNuA4 and SAGA.

R Magnus N Friss, Bob P Wu, Stacey N Reinke, Darren J Hockman, Brian D Sykes, and Michael C Schultz

Department of Biochemistry, University of Alberta, Edmonton, Canada

A version of this chapter has been accepted for publication: Friis et al. 2009. *Nuclei Acids Res*. 37(12):3969-80.

SN Reinke contributed ^{1}H-NMR derived metabolomic analyses of yeast.

Summary

Little is known about what enzyme complexes or mechanisms that control global lysine acetylation in the amino-terminal tails of the histones. In yeast, it is well established that signalling pathways activated by glucose can drive transcription reprogramming (Zaman et al. 2008) and that almost 1400 genes are induced when stationary phase cells are fed glucose (Liko et al. 2007). As acetyl-CoA is an essential co-substrate for lysine acetylases, we predicted that glucose-induced histone acetylation is mediated through a metabolic mechanism. To determine if glucose-fed stationary phase cells produce precursors for acetyl-CoA synthesis, we quantified acid-soluble intracellular yeast metabolites using ^{1}H-NMR spectroscopy. Representative glucose spectra, in one feeding experiment, are shown in Figure B-1A. As expected the concentration of glucose is very low in unfed cells and rises sharply after feeding (Figure B-1B; the upper panel includes the reference signature peaks for glucose, the lower panel displays the doublet peak of the proton of the α-anomeric carbon in cells). Increases in ADP, ATP, acetate and ethanol concentrations accompany glucose uptake (Figure B-1B). We conclude that feeding stationary phase cells glucose results in new synthesis of ATP and acetate, the metabolites required for acetyl-CoA production.

Figure

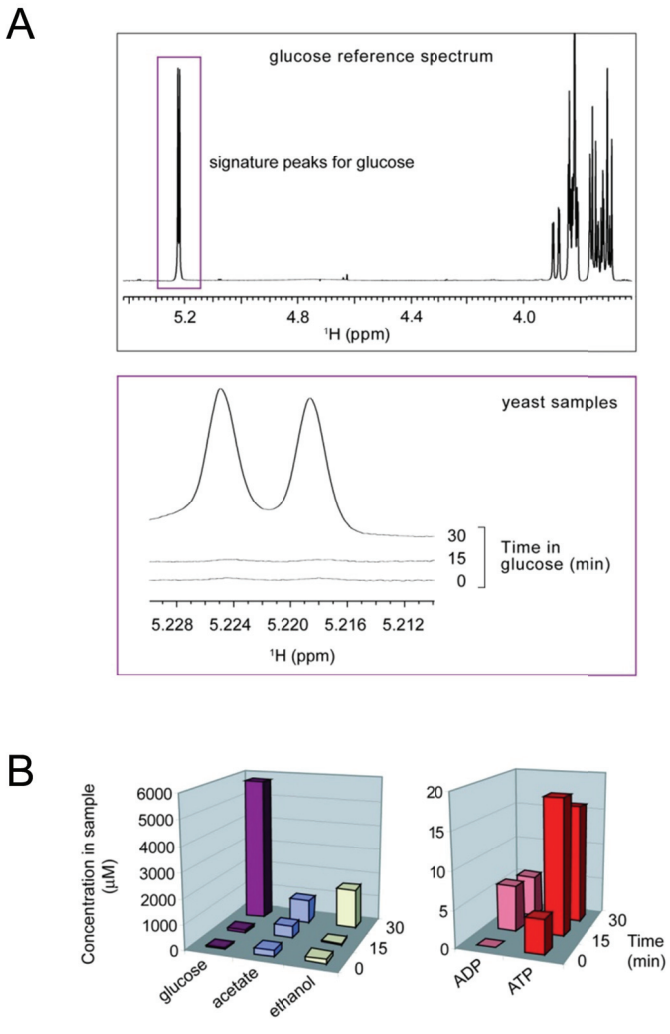


Figure B-1. Metabolite profiles in unfed and glucose-fed SP cells. A, ^{1}H-NMR spectra of a glucose reference sample (top panel) and samples of metabolites isolated from SP cells (bottom panel; only glucose peaks are shown). B, Targeted quantitative profiling of cellular metabolite levels by ^{1}H-NMR, before and after glucose feeding. Identical volumes were analyzed from samples prepared from the same number of cells and resuspended in the same final volume. The results are the average of two independent experiments.

References

Liko D, Slattery MG, Heideman W (2007) Stb3 binds to ribosomal RNA processing element motifs that control transcriptional responses to growth in *Saccharomyces cerevisiae*. J Biol Chem 282 (36):26623-26628

Zaman S, Lippman SI, Zhao X, Broach JR (2008) How *Saccharomyces* responds to nutrients. Annu Rev Genet 42:27-81

APPENDIX C

Formate can differentiate between hyperhomocysteinemia due to impaired remethylation and impaired transsulfuration.

Simon G. Lamarre[1], Anne M. Molloy[2], Stacey N. Reinke[3], Brian D. Sykes[3], Margaret E. Brosnan[1] and John T. Brosnan[1]

[1]Department of Biochemistry, Memorial University, St. John's, NL, Canada
[2]Department of Clinical Medicine, Trinity College, Dublin, Ireland
[3]Department of Biochemistry, University of Alberta, Edmonton, AB, Canada

A version of this chapter has been accepted for publication: Lamarre et al. 2011. *Am J Physiol Endocrinol Metab*. Available online September 20, 2011.

SNR contributed ^{1}H-NMR derived metabolomic analyses of rat serum.

Summary

It is generally held that vitamin B_{12} (cobalamin) provides cofactors for only two mammalian enzymes: methionine synthase, which catalyzes the remethylation of homocysteine to methionine, and methylmalonyl-CoA mutase, which catalyzes the conversion of methylmalonyl CoA to succinyl CoA. Given that the functions of only a fraction of the 26,000 mammalian genes are fully understood, we hypothesized that there might be additional, undiscovered, vitamin B_{12}-requiring enzymes. We addressed this issue by feeding rats a vitamin B_{12}-deficient diet and subjecting their sera to proton nuclear magnetic resonance (^{1}H-NMR) metabolomic analysis. We observed an approximately 6-fold increase in formate levels from 64 μM in control rats to 402 μM in vitamin B_{12}-deficient rats (Table C-1). This elevation could be attributed to impaired one-carbon metabolism, since formate is assimilated into the one-carbon pool by incorporation into 10-formyl-THF, via the enzyme 10-formyl-THF synthase. We also observed that methylmalonate was only present in B_{12}-deficient rats (Table C-1). As the concentrations of three ketone bodies were significantly lower in the B_{12}-deficient rats (Table C-1), this observation might be due to the sequestration of some of the hepatic coenzyme A pool as methylmalonyl-CoA, which may impair fatty acid oxidation or ketogenesis. Although our investigations do not identify a novel B_{12}-dependent metabolic pathway, they do reveal that formate and methylmalonate provide a novel window into cellular folate metabolism; these metabolites can be used as biomarkers of deranged one-carbon metabolism.

Table

Table C-1. Metabolites displaying more or less than twice the control concentration in the serum of B_{12} deficient rats, as measured by ^{1}H-NMR metabolomics analysis

Metabolite	Concentration (μM)		
	Control	Deficient	Fold change
Methylmalonate	ND	84.5	∞
Formate	64.3	401.9	6.25
Acetoacetate	82.9	33.8	0.41
3-Hydroxybutyrate	289.9	99.3	0.34
Acetone	29.7	7.3	0.25