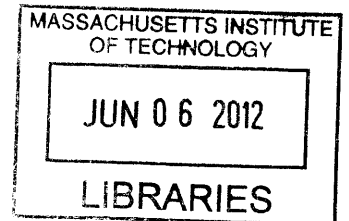


Bridging the gap between protein-tyrosine phosphorylation networks, metabolism and physiology in liver-specific PTP1b deletion mice

by

Emily R. Miraldi

Bachelor of Arts, Oberlin College (2006)



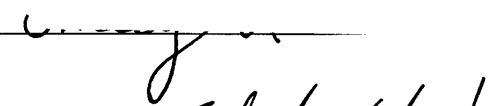
Submitted to the Computational and Systems Biology Program
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computational and Systems Biology at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY


ARCHIVES


June 2012

©2012 Emily R. Miraldi; All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Author 
Emily R. Miraldi
Computational and Systems Biology Program
April 13, 2012

Certified by 
Forest M. White
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by 
Chris B. Burge
Professor of Biology and Biological Engineering
Director, Computational and Systems Biology Graduate Program

Bridging the gap between protein-tyrosine phosphorylation networks, metabolism and physiology in liver-specific PTP1b deletion mice

by
Emily R. Miraldi

Submitted to the Computational and Systems Biology Program
on April 13, 2012, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Computational and Systems Biology

ABSTRACT

Metabolic syndrome describes a complex set of obesity-related disorders that enhance diabetes, cardiovascular, and mortality risk. Studies of liver-specific protein-tyrosine phosphatase 1b (PTP1b) deletion mice (L-PTP1b^{-/-}) suggests that hepatic PTP1b inhibition would mitigate metabolic syndrome progression through amelioration of hepatic insulin resistance, endoplasmic reticulum stress, and whole-body lipid metabolism. However, the network alterations underlying these phenotypes are poorly understood. Mass spectrometry was used to quantitatively discover protein phosphotyrosine network changes in L-PTP1b^{-/-} mice relative to control mice under both normal and high-fat diet conditions. A phosphosite set enrichment analysis was developed to identify numerous pathways exhibiting PTP1b- and diet-dependent phosphotyrosine regulation. Detection of PTP1b-dependent phosphotyrosine sites on lipid metabolic proteins initiated global lipidomics characterization of corresponding liver samples and revealed altered fatty acid and triglyceride metabolism in L-PTP1b^{-/-} mice. Multivariate modeling techniques were developed to infer molecular dependencies between phosphosites and lipid metabolic changes, resulting in quantitatively predictive phenotypic models.

Thesis Supervisor: Forest M. White
Title: Associate Professor of Biological Engineering

Acknowledgements

Foremost I am extremely grateful for the mentorship of my adviser, Forest White. Although the focus of our lab is experimental, Forest welcomed me as a student with both computational and experimental interests and provided me with an exceptional opportunity to explore both. This was a challenging undertaking for both of us, and I could not have done it without Forest's support and confidence. I admire his scientific insight, integrity, and willingness to give honest criticism and advice. Because of Forest's support and mentorship, the years I have spent at MIT have been rich both intellectually and in a broader sense.

I am indebted to Doug Lauffenburger, who took the time to discuss biological engineering with me when I was an undergraduate at Oberlin College. Doug gave me the opportunity to work in his lab during the summer between my junior and senior years of college. This experience gave me a deep-seeded enthusiasm for mathematical modeling of biological systems, which motivated me to apply for graduate school. I would also like to acknowledge John Burke, who mentored me that summer. I am also very grateful to Doug for his mentorship in graduate school. Serving as chair of my thesis committee, he inspired and challenged me to incorporate engineering principles into my thesis.

I am thankful to Alan Saghatelian, who became a collaborator on my project at the beginning of my fifth year in graduate school. He gave me the opportunity to learn an entirely new set of experimental (metabolomics) techniques in his lab, and my enthusiasm for this field stems directly from the year and a half that I spent in his lab. I have benefitted greatly from his generosity, scientific advice, and mentorship. His influence on this thesis is unmistakable.

I am grateful to Ben Neel who, along with Forest and me, designed my thesis project. As a scientist whose natural inclination is computational, I have a high threshold for experimental endeavors, but the collaborative project proposed by Ben captured my scientific curiosity. Throughout my time at MIT, Ben's input on the project was invaluable, and he challenged me to think critically about experimental data. I feel extremely fortunate to have had the opportunity to interact with him.

I was very lucky that Barbara Kahn agreed to serve on my thesis committee. She provided critical physiological insights in thesis committee meetings, many of which led to re-evaluation of conclusions from experimental data in my thesis work and suggested follow-up experiments. I also appreciate Barbara's positive outlook and enthusiasm.

I thank Jason Kim and a talented post doc in his lab, Randall Friedline, who generously contributed their time and resources at the early stage of the collaboration.

I would like to thank all members of both the White and Saghatelian labs who have mentored me, been good colleagues and friends, and contributed to experiments or computational analyses. From the White lab, I would specifically like to acknowledge Hannah Johnson, Amanda Del Rosario, Bryan Owens-Bryson, Kristen Naegle, Josh Apgar, Paul Huang, and Scott Carlson. From the Saghatelian lab, I would like to recognize Tejia Zhang and Edwin Homan.

I have benefited from conversations and interactions with innumerable others. This list is woefully incomplete, but I would like to thank Ken Lau, Hadar Sharfi, Robert Gerszten and lab members, Ernest Fraenkel and lab members (in particular, Carol Huang), Michael Hemann and

lab members (in particular, Justin Pritchard) and talented teaching assistants, Faisal Kashif and Vincent Tan. I have also greatly benefited from interaction with fellow students in the Computational and Systems Biology Ph.D. program, the Biological Engineering Department and other programs at MIT.

Finally, I would like to thank my friends and family. I appreciate constant love and support from my sister and brother, Virginia Utz and Daniel Miraldi, and my parents, David and Leslee Miraldi. In particular, I thank my parents, because much of my happiness at MIT I attribute to them. Following their examples, I have endeavored to take advantage of academic opportunities but not at the expense of family and friends. My parents helped me develop an enthusiasm for many of the activities that have enriched my life here, including art, music, cooking, and sports. They helped me develop confidence to think in terms of possibilities and to admit mistakes. All of these skills have proved invaluable at MIT.

I also appreciate funding from an NSF Graduate Fellowship.

Table of Contents

ABSTRACT	3
ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS	6
FIGURES	8
SUPPLEMENTARY FIGURES	9
TABLES	9
1. INTRODUCTION: NOVEL MECHANISMS OF CELLULAR SIGNALING AND METABOLIC CONTROL	10
Introduction	10
Regulation of metabolic enzymes by tyrosine phosphorylation	13
Regulation of metabolic enzymes by acetylation	18
Saturated fatty acid signaling molecules and ceramide derivatives	21
Monounsaturated fatty acids as signaling molecules	25
Polyunsaturated fatty acid signaling molecules and eicosanoid derivatives	28
Conclusions	32
References	34
2. BRIDGING THE GAP BETWEEN PROTEIN-TYROSINE PHOSPHORYLATION NETWORKS, METABOLISM AND PHYSIOLOGY IN LIVER-SPECIFIC PTP1B DELETION MICE	39
Introduction	40
Results	45
Study design	45
HFD and genotype drive global phosphotyrosine network changes.....	46
PTP1b-dependent phosphosites are enriched for insulin signaling, REDOX, and lipid metabolic pathways	47
PTP1b-deficient livers have an overabundance of free PUFA	50
PTP1b-deficient livers have altered triglyceride metabolism on HFD	54
Modeling the relationships between protein phosphorylation and lipid metabolites.....	56
Discussion	59
Methods	63
Animal studies.....	63
Serum measurements.....	64

Phosphotyrosine analysis	64
Principle component analysis and variance explained by diet and genotype.....	67
Unsupervised clustering analysis	69
Supervised clustering of genotype- and diet-dependent phosphorylation sites	70
Lipidomics analysis.....	71
Biochemical analysis.....	78
Stochastic multivariate regression analysis.....	78
Phosphosite set enrichment analysis	83
References.....	86
Figures.....	91
Supplementary Figures	99
Tables	104
3. LEVERAGING MS-BASED PTM DATASETS: STRATEGIES FOR MULTIVARIATE MODELING AND PATHWAY INFERENCE IN THE CONTEXT OF NOISE AND MISSING DATA.....	117
Introduction.....	117
Results	123
Study design	123
Stochastic correlation reflects measurement error	124
PTM Set Enrichment Analysis extends GSEA for PTM datasets.....	127
Modeling of measurement error using stochastic multivariate regression.....	131
Optimization of missing data methods for multivariate regression	134
Integrated SMR-PSEA analysis yields predictive models and pathway-level insight from incomplete data	137
Discussion.....	139
Methods.....	142
PTM Set Enrichment Analysis	142
Stochastic correlation	146
Stochastic multivariate regression.....	147
Missing data techniques for stochastic multivariate regression	151
Evaluation of missing data techniques	153
Final models	154
References	155
Figures.....	157
Supplementary Figure	167
4. CONCLUSIONS.....	168
Contributions.....	168
Limitations and Future Directions	175
References	180

Figures

Figure 2.1 Study Design and Phosphotyrosine Analysis	91
Figure 2.2 Diet and Genotype Dependencies in the Phosphotyrosine Network.....	93
Figure 2.3 Global Lipidomics Analysis for <i>De Novo</i> Discovery.....	94
Figure 2.4 PTP1b-dependent Changes in FFA Metabolism	95
Figure 2.5 PTP1b-dependent Changes in Triglyceride Metabolism.....	96
Figure 2.6 Multivariate Models of Lipidomics Phenotypes	97
Figure 2.7 Summary of Lipid Metabolic and Phosphosite Interactions	98
Figure 3.1 Noise and Missing Data in L-PTP1b ^{-/-} Phosphotyrosine Data.....	157
Figure 3.2 Stochastic Correlation	158
Figure 3.3 Comparison of Correlation P-values	159
Figure 3.4 Schematic of PSEA for Genotype-dependent Phosphosites	160
Figure 3.5 Results of PSEA for Genotype-dependent Phosphosites	161
Figure 3.6 Stochastic Multivariate Regression	162
Figure 3.7 Schematic for Testing Missing Data Methods	163
Figure 3.8 ROC Performance of Missing Data Methods.....	164
Figure 3.9 SMR Model Prediction.....	165
Figure 3.10 SMR Model Components and PSEA	166

Supplementary Figures

Supplementary Figure 2.1 Diet-dependent Phosphosites	99
Supplementary Figure 2.2 Co-elution to Determine PUFA Species	100
Supplementary Figure 2.3 Measurements of Hepatic Inflammation	101
Supplementary Figure 2.4 Serum Triglyceride Compositions	102
Supplementary Figure 2.5 Multivariate Models of Steatosis	103
Supplementary Figure 3.1 Complete Data Subsets.....	167

Tables

Table 1: Animal Weights and Serum Measurements	104
Table 2: Enrichment Analysis of Genotype-Dependent Phosphosites	104
Table 3: Enrichment Analysis of Diet-Dependent Phosphosites.....	107
Table 4: Quantification of Hepatic FFAs.....	108
Table 5: Quantification of Hepatic Triglycerides and Cholesterol.....	110
Table 6: Percent Compositions of Serum Triglycerides	112
Table 7: Phosphosites Sets Predictive of ω 3 / ω 6 PTP1b PUFA.....	113
Table 8: Phosphosite Sets Predictive of ω 7 / ω 9 PTP1b PUFA	114
Table 9: Phosphosite Sets Predictive of Steatosis.....	115

1. Introduction: Novel Mechanisms of Cellular Signaling and Metabolic Control

Introduction

Cellular signaling mechanisms enable living organisms to sense and respond to diverse environmental cues. Example cues include light, nutrients, and hormones; example cellular responses are movement, production of nutrient-metabolizing enzymes, and proliferation. Cues drive down-stream networks of chemically-driven molecular events, or cellular signaling mechanisms, that lead to a response.

One key cellular signaling mechanism involves classes of proteins known as kinases and phosphatases. In concert, they are responsible for the addition and removal of phosphate groups to and from particular sites on proteins. The addition or removal of the phosphate group has diverse effects on a protein's state, including enhancement or inhibition of enzymatic activity, change in cellular localization, and/or interaction with another molecule. External cues can result in robust phosphorylation changes on more than hundreds of proteins (Schmelzle et al., 2006; Zhang et al., 2005). Perturbation of these phosphorylation signaling networks by mutation of phosphosite or modulation of kinase or phosphatase protein levels can drastically alter cellular response and whole-body phenotype (Elchebly et al., 1999; Huang et al., 1997; Nakatani et al., 2004). The network of protein and molecular interactions underlying these coordinated, multicomponent phosphorylation changes are the subject of intense scientific research and of particular importance to human health. Aberrant phosphatase and kinase activities are well-documented drivers of diseases, including cancer and diabetes (Blume-Jensen and Hunter, 2001; Cohen, 2006).

Recently, the important roles of other protein post-translation modifications (PTMs) in cellular signaling have been highlighted. These PTMs include protein acetylation, methylation, glycosylation, ubiquitylation, and protein cleavage; these mechanisms have also been shown to play critical roles in the regulation of cellular responses. Individual proteins often contain multiple sites for modification by diverse PTMs, and the control of protein behavior as a function of these interacting modifications will fundamentally progress our understanding of cellular signaling transduction and control.

Significant progress in understanding protein PTM-mediated cellular signal transduction has been accomplished using traditional biochemistry and molecular biology techniques, which tend to be low throughput, qualitative and largely limited to univariate lines of inquiry. However, recent advances in biotechnology have enabled quantitative PTM measurements of increasing scale, both in number of conditions and site-modifications measured (Del Rosario and White, 2010). Of particular importance have been breakthroughs in electro-spray ionization liquid-chromatography coupled to tandem mass spectrometry (ESI-LC-MS/MS), which, in tandem with PTM enrichment techniques (especially immobilized metal affinity chromatography (IMAC) and PTM peptide immunoprecipitation (IP) reactions), have enabled the identification of thousands of PTMs (Ficarro et al., 2002; Kim et al., 2006). Development of isotopic labeling techniques, such as Stable Isotope Labeling by Amino acids in Cell culture (SILAC) and Isobaric Tag for Relative and Absolute Quantification (iTRAQ), enable simultaneous quantification and identification of phosphopeptides across multiple conditions. Quantitative MS techniques allow network-level characterization of hundreds of PTMs in response to diverse stimuli and contexts (Mann, 2006; Wolf-Yadlin et al., 2007; Zhang et al., 2005). Global MS-based methods have been adapted for the detection and quantification of protein phosphorylation, lysine acetylation

and cleavage in response to cellular signaling cues (Dix et al., 2008; Owens-Bryson; Wolf-Yadlin et al., 2007).

One particular advantage of MS-based, PTM analysis is that they enable detection and quantification of any modified site on theoretically any protein. In contrast, other high-throughput approaches typically measure a much more limited set of PTMs for which site-specific antibodies already exist (Wolf-Yadlin et al., 2009). In these experiments, measured posttranslational modification must be selected a priori, placing significant constraints on biological exploration. In contrast, MS-based techniques enable global, systematic exploration of cellular cue responses. Often, when an allegedly well-characterized PTM-based cellular response network is re-examined by MS-based analysis, novel cue-responsive PTMs and protein-protein interactions are revealed, leading to a more nuanced biological network picture (Huang et al., 2007; Schmelzle et al., 2006).

In addition to impacting the traditionally “protein-centric” field of cellular signaling, mass spectrometry advances have also revolutionized and revitalized the study of metabolism, often mistakenly considered to be a well-understood domain of biology. Similarly to genomics and proteomics, the nascent field of metabolomics is focused on global metabolic measurements across cells, tissues, and fluids. Metabolites are defined broadly as peptides, carbohydrates, lipids, nucleosides, and catabolic products of exogenous compounds (Saghatelian and Cravatt, 2005). In this field, MS-based techniques continue to reveal novel classes of metabolites and uncover significant roles for both novel and well-studied metabolites as cellular signaling molecules in human health and disease (Cao et al., 2008; Nomura et al., 2010; Spite and Serhan, 2010).

Individually, proteomics- and metabolomics-based analyses have revealed important connections between physiological response and diverse cellular signaling molecules and mechanisms. Here, I highlight two classes of PTMs, tyrosine phosphorylation and lysine acetylation, that have recently been found to control cellular metabolism via direct modification of metabolic enzymes in addition to their relatively better-characterized, indirect modes of action, through canonical signaling cascades and nuclear control of transcriptional regulation, respectively. In addition, I highlight fatty acids as a class of signaling-molecule metabolites whose mechanisms of action depart from the traditional perspective of metabolite signals as allosteric modulators of upstream and downstream metabolic enzymes only.

Regulation of metabolic enzymes by tyrosine phosphorylation

Protein tyrosine phosphorylation in growth factor and insulin signaling plays an indisputably central regulatory role in the determination of cellular metabolic response to nutrient cues. In growth factor and insulin signaling pathways, ligands activate receptor tyrosine kinases (RTKs) and ensuing tyrosine phosphorylation changes on downstream proteins lead to activation of multiple signaling pathways. The two most well-studied downstream signaling pathways are those leading to activation of the Erk-mitogen activated protein kinase (MAPK) and the phosphatidylinositol-3 kinase (PI3K)/Akt pathways (Saltiel and Pessin, 2002). Erk and Akt are serine/threonine kinases that serve as central signaling hubs, and many serine and threonine phosphorylation sites on metabolic enzymes downstream of these kinases have direct and well characterized effects on metabolic enzyme activity (Boulpaep et al., 2003). Tyrosine phosphorylation also leads to transcriptional regulation of metabolic enzymes. Growth hormone and insulin signaling, in addition to other important metabolic signaling cues, such as leptin and cytokines, promote tyrosine phosphorylation and activation of the Janus kinase (JAK) – signal

transducer and activator of transcription (STAT) pathways, leading directly to transcriptional regulation of metabolic enzymes (Richard and Stephens, 2011).

In contrast to these well-established mechanisms, experimental evidence for direct tyrosine phosphorylation of metabolic enzymes in modulating metabolic control is recent. In 1983, tyrosine phosphorylation of three glycolytic metabolic enzymes was discovered in cells transformed by the Rous sarcoma virus (Cooper et al., 1983). Although authors conceded that tyrosine phosphorylation of these enzymes could be “gratuitous”, they also postulated that these post translation changes might contribute to aerobic glycolysis, or the Warburg effect, in transformed cells. These first tyrosine phosphorylated enzymes were resolved using 2D gels and silver-staining. In the last decade, due to the technical breakthroughs described above, hundreds of tyrosine phosphorylation sites on metabolic enzymes have been reported (Hornbeck et al., 2012; Naegle et al., 2010). In addition, several quantitative MS experiments have shown that tyrosine phosphorylation sites on metabolic enzymes are dynamically regulated in response to growth hormone or insulin stimulus, as well as other metabolism-altering contexts, such as cancer (Huang et al., 2007; Schmelzle et al., 2006; Zhang et al., 2005).

Although quantitatively measured response to metabolism-altering cues suggests a functional role for these phosphotyrosine sites on metabolic enzymes, it has only been in the last several years that molecular biology experiments have convincingly established important metabolic roles for site-specific phosphotyrosine regulation of metabolic enzymes. Most of this work has been motivated by a desire to understand the role of tyrosine phosphorylation of metabolic enzymes in cancer cell metabolism.

Tyrosine phosphorylation of pyruvate kinase (PK) isoenzyme M2 (PKM2) promotes cancer cell metabolism and growth (Hitosugi et al., 2009). While PK isoenzyme M1 (PKM1) is constitutively active and expressed in non-proliferating, differentiated cells, PKM2 requires allosteric activation by fructose-bis-phosphate (FBP) and is expressed during development, in highly proliferative tissues and in tumor cells. PKM2 exists as an active tetramer or as a lower activity dimer. Because PKM2 catalyzes a rate-limiting step in glycolysis (conversion phosphoenolpyruvate (PEP) to pyruvate), it was initially unclear why cancer cells, which rely on high levels of ATP, would preferentially express the less-active isoenzyme. However, relative to PKM1, the PKM2 tetramer was found to favor glycolysis and lactate production, while the less-active PKM2 dimers favored the diversion of trioses toward biosynthesis of amino acid and lipids, processes critical to proliferation and growth. Replacement of PKM2 with PKM1 in tumor cell lines reduced aerobic glycolysis and sizes of tumors in xenographs, suggesting an important role for PKM2 in mediating the Warburg effect and proliferative advantage (Dang, 2009).

PKM2 activity is inhibited by interaction with specific tyrosine-phosphorylated peptides as well as phosphorylation of PKM2 Y105 (Hitosugi et al., 2009). Site-specific mutation of six tyrosine-phosphorylated residues to phenylalanine revealed that Y105 uniquely decreased PKM2 activity. This site was found to be phosphorylated in multiple cancer cell lines and by several oncogenic kinases, including fibroblast growth factor receptor 1 (FGFR1). Authors also demonstrated that Y105F PKM2 expression in H1299 cells specifically reduced Warburg hallmarks relative to WT PKM2 and conferred proliferative disadvantages in xenograph growth models, demonstrating the functional importance of PKM2 Y105 phosphorylation (Hitosugi et al., 2009).

However, it is often argued that the stoichiometry of metabolic enzyme tyrosine phosphorylation is too low to have cellular-level metabolic effects. Evaluation of the molecular mechanism by which PKM2 was inhibited by pY105 resulted in a model that more broadly addresses a means by which some low-stoichiometry tyrosine phosphorylation events on metabolic enzymes might potentiate physiological consequences. Authors used ^{14}C -FBP to show that incubation of WT PKM2 with the pY105-peptide destabilized the allosterically-activating binding of FBP to PKM2 (Hitosugi et al., 2009). They also demonstrated that PKM2 catalytic activity increased upon FGFR1 inhibition for WT and mutant-control Y390F PKM2, but not Y105F or the K433E PKM2 mutant, the latter of which cannot bind phosphorylated peptide; this suggested that pY105-mediated PKM2 inhibition required binding of pY105-phosphorylated PKM2. Together, these findings suggest an “intermolecular” or “transprotein” FBP-release model across PKM2 dimers and tetramers, by which one phosphorylated PKM2 monomer can destabilize FBP binding and inhibit non-phosphorylated PKM2 monomers. This model might explain how other low stoichiometric phosphorylation events could potentiate large scale changes in metabolic activity (Hitosugi et al., 2009). As discussed, PKM2 is also inhibited by binding to other tyrosine phosphopeptides, and thus tyrosine phosphorylation sites on other proteins and peptides may synergize with low-stoichiometry, phosphorylated enzymes to significantly impact metabolism.

Another study demonstrated a role for tyrosine phosphorylation of lactose dehydrogenase A (LDH-A) in the control of NADH/NAD⁺ redox homeostasis in cancer cells (Fan et al., 2011). LDH converts pyruvate to lactate, a reaction that regenerates the NAD⁺ required for glycolysis. LDH had previously been implicated in tumorigenesis, potentially contributing to aerobic glycolysis. Similarly to the PKM2 study, authors established that multiple oncogenic tyrosine

kinases could phosphorylate LDH-A, and they developed site-specific mutants to test for effect of tyrosine phosphorylation on LDH-A activity. LDH-A phosphorylation of two sites, Y10 and Y83, were necessary for increased enzymatic activity *in vitro*. LDH-A pY10 was detected in several cancer cell lines, and a gel filtration chromatography experiment following by an *in vitro* FGFR kinase assay with WT or Y10F LDH-A suggested that pY10 increased LDH-A activity by promoting tetramer formation. LDH-A Y10F site mutant expression in cancer cell lines resulted in decreased glycolytic metabolism and increased oxygen consumption, as well as higher NADH/NAD⁺ ratio under normoxic and hypoxic conditions. Without pY10 LDH-A to sustain the redox balance required for aerobic glycolysis, LDH-A Y10F mutants exhibited a growth disadvantage in xenograph models relative to control. Thus, by distinct mechanisms, PGM-2 and LDH-A tyrosine phosphorylation both modulate cellular metabolic state and tumor fitness.

Another recent study demonstrated that several oncogenic tyrosine kinases translocate to the mitochondria to phosphorylate pyruvate dehydrogenase kinase 1 (PDHK1), a serine/threonine kinase that most notoriously inactivates pyruvate dehydrogenase (PDH). Some of the oncogenic tyrosine kinases were constitutively active mutants, but authors also detected full-length FGFR as tyrosine phosphorylated in the mitochondria of H1299 cells, suggesting that FGFR1 may have a basal level of mitochondrial tyrosine kinase activity that is ligand independent. In addition, the tyrosine kinases were localized to distinct mitochondrial locations (outer membrane or matrix), and the location of the measured PDH activity corresponded to tyrosine kinase location (Hitosugi et al., 2011). Together, these data suggest that tyrosine phosphorylation of mitochondrial proteins may (1) occur outside of the context of oncogenic mutant RTKs and (2) be highly spatially regulated. Tight, spatial regulation provides another argument for a significant

functional role of tyrosine phosphorylation sites on metabolic enzymes, even in the context of inferred low-stoichiometry from whole-cell lysates.

As a final example, Src tyrosine kinase and potentially other family members translocate to the mitochondria to regulate cellular-level metabolism. Most significantly, Src phosphorylates adenine nucleotide translocator 1 (ANT1) on Y190 and Y194, and these sites modulate mitochondrial ADP/ATP exchange (Feng et al., 2010). Src may also modulate cellular redox metabolism by direct phosphorylation of several other mitochondrial enzymes. Inhibition of Src activity has been shown to decrease mitochondrial respiration via complex I, complex IV, or both complex IV and V activities in cancer cells, osteoclasts, and rat brain mitochondria, respectively. *In vitro* kinase assays coupled with *in vitro* enzyme activity assays have suggested that particular tyrosine phosphorylation sites on these respiratory-chain complex proteins modulate enzyme activity (Hebert-Chatelain et al., 2012). However, the redox environment of tissue culture is distinct from an *in vivo* environment, and it remains to be seen whether these Src-dependent mitochondrial redox phenotypes extend *in vivo*.

Together, these studies highlight the importance of site-specific tyrosine phosphorylation on metabolic enzymes in modulating cellular-level metabolism. Most of these studies have focused on metabolic enzyme tyrosine phosphorylation in the context of cancer cell metabolism and proliferation. Normal physiological and other pathophysiological conditions present exciting opportunities for the exploration of phosphotyrosine regulation of metabolic enzymes.

Regulation of metabolic enzymes by acetylation

Protein lysine acetylation was first discovered almost 50 years ago, and research in this field has defined a crucial role for lysine acetylation of transcription factors and histones in nuclear

control of gene expression (Spange et al., 2009). In 1997, the NAD-dependent protein lysine-deacetylase *Sir2p* was discovered to slow aging in yeast, and the mammalian homolog SIRT1, highly induced upon calorie restriction (CR) in mice, demonstrated similar capacity (Guarente, 2011b). Thus, a role for lysine acetylation in metabolism was hypothesized. Analogously to protein tyrosine phosphorylation, SIRT1 control of cellular metabolism under CR was first shown to be mediated by a well-established acetylation mechanism; some of the first important SIRT1 substrates, forkhead box (FOX) O1, FOXO3, and proliferator-activated receptor- γ 1 α (PGC1 α), were nuclear transcription factors (Guan and Xiong, 2011). However, the recent advent of an LC-MS/MS protocol involving IP enrichment of lysine-acetylated peptides spurred the identification of thousands of protein lysine acetylation events, and these events extend well beyond the nucleus (Kim et al., 2006).

In particular, a preponderance of acetylation on metabolic proteins in the cytosol and mitochondria has been reported, and several molecular and biochemical studies have demonstrated important roles for this modification in enzyme function. Similar to phosphorylation, acetylation has been shown to decrease or increase enzyme activity. In addition, acetylation of glyceraldehydes 3-phosphate dehydrogenase (GAPDH) was found to regulate the direction of carbon flux in *Salmonella enterica* (Guan and Xiong, 2011), suggesting another intricate mode of metabolic control.

Interestingly, in the cytosol, entire metabolic pathways are acetylated under certain metabolic perturbations, while, in the mitochondria, acetylation trends are less uniform (Guan and Xiong, 2011). A network perspective provides insight (Guarente, 2011a). Under glucose-rich conditions, glycolytic enzymes are heavily acetylated, and this modification generally appears to promote glycolysis. Glucose metabolism is often coordinated with lipogenesis, and it has been

shown that flux through ATP citrate lyase (ACLY) promotes protein acetylation, as ACLY converts citrate to acetyl-CoA for fatty acid synthesis (Wellen et al., 2009). An acetylation-driven positive feedback loop is likely to promote glycolysis under glucose-rich conditions: (1) the glycolytic product pyruvate is shuttled to the mitochondria for oxidation by the TCA cycle, (2) citrate from the TCA cycle is shuttled to the cytosol for ACLY-mediated conversion to acetyl-CoA, (3) lysine-acetyl transferases utilize the cytosolic acetyl-CoA to enhance the activity of glycolytic enzymes. In addition, these metabolic pathways increase NADH:NAD⁺ ratio, potentially limiting the activity of SIRT1 by reducing the availability of its cofactor NAD⁺. In contrast, under starvation conditions, SIRT1 becomes active, and de-acetylation of cytosolic enzyme substrates not only inhibits anabolic metabolic pathways but also releases acetate, which can feed into mitochondrial fatty-acid oxidation pathways. Consistent with this notion is the role of a second deacetylase, SIRT3, also active under starvation conditions; SIRT3 activates long-chain acyl dehydrogenase (LCAD) and acetyl-CoA synthetase (ACS), which facilitate mitochondrial metabolism of acetyl-CoA derived from both fatty acids and acetate, respectively. In the mitochondria, pathway acetylation patterns may be more complex (less uniform) than cytosolic trends due to both the flux of acetate to the mitochondria and regulation by multiple mitochondrial sirtuins (Guarente, 2011a). Hybrid feedback loops, composed of large-scale acetylation changes and metabolic fluxes, may be a paradigm for metabolism and PTM regulation in a broader context.

Thus, protein lysine acetylation not only coordinates metabolic responses via transcription factor and histone regulation but also through direct biologically functional modification of metabolic enzymes. As evidenced by recent MS experiments, the acetylation of metabolic enzymes is extensive (Guan and Xiong, 2011). In response to nutrient cues, acetyl-lysine modifications span

individual enzymes to entire pathways. The first quantitative MS-based analysis of protein-lysine acetylation was recently published (Choudhary et al., 2009). Additional quantitative characterizations of cellular responses have the potential to provide network-level insight into the restructuring of metabolism by acetyl-lysine modification.

Saturated fatty acid signaling molecules and ceramide derivatives

The traditional role of fatty acid signaling is generally limited to allosteric inhibition of upstream and downstream metabolic enzymes. It is increasingly appreciated that certain fatty acids coordinate many metabolic pathways globally, and this large-scale regulation of multiple, indirectly-related metabolic pathways places these molecules in the realm of canonical signaling proteins. In addition, several fatty acids control pathways that are not traditionally considered metabolic.

A role for fatty acids (FAs) in the regulation of inflammatory pathways emerges from the study of obesity, a state characterized by increased FA storage, expanded adipose tissue mass, and chronic activation of inflammatory pathways (Konner and Bruning, 2011). In particular, obesity (1) results in increased expression and secretion of TNF α and IL-6 from white adipose tissue, (2) leads to the dysregulation of cellular homeostatic mechanisms, such as the endoplasmic reticulum (ER) stress response, and (3) is associated with increased levels of FAs that promote inflammation and contribute to the development of insulin resistance (Konner and Bruning, 2011). These processes are accompanied by activation of downstream cellular stress kinases, such as c-Jun N-terminal kinase (JNK) and I κ B kinase (IKK), which show promise as therapeutic targets (Cai et al., 2005; Nakatani et al., 2004). Of equal importance, however, are the upstream pathways leading to activation of these kinases, and, in that regard, fatty-acid

activation of Toll-like receptors (TLRs) appears to play a critical role (Konner and Bruning, 2011).

TLRs are pattern-recognition receptors (PRRs) involved in innate immune recognition of pathogenic microbial infection. They bind to lipopolysaccharides (LPS), a component of Gram-negative bacterial cell walls, and, in a macrophage cell line, the medium-chain FA component of LPS, lauric acid (C12:0), has been shown to trigger a TLR4 response (Konner and Bruning, 2011). Further studies have shown that saturated fatty acids (SFAs) but not unsaturated fatty acids (USFAs) activate TLR4 signaling in macrophages and adipocytes, and TLR4 deficiency blocks FA-mediated inflammatory signaling in the same cells (Konner and Bruning, 2011). *In vivo*, a study of TLR4 deficient mice compared the effects of SFA-rich to USFA-rich HFDs and demonstrated that TLR4 is required for SFA-induced insulin resistance, adiposity, and adipose macrophage invasion. Interestingly, TLR4 deficiency did not protect against these phenotypes on the USFA-rich HFD (Davis et al., 2008). This highlights the diversity of pathological mechanisms in obesity as well as the specificity of TLR4-mediated inflammation for SFAs. In addition to acyl-chain saturation, induction of a TLR4 response is specific to SFA chain length. Laurate, myristate (C14:0), and palmitate (C16:0) activate TLR4-dependent inflammation in adipocytes, whereas stearate (C18:0) does not. Furthermore, the specificity for chain length is cell-type dependent. In contrast to adipocytes, macrophage TLR4-dependent inflammatory response is stimulated by both palmitate and stearate (Chait and Kim, 2010). Together, these findings suggest a stereospecific rather than a fuel-related mechanism for SFAs in TLR4-mediated inflammation.

It was originally thought that SFAs were TLR4 ligands. However, radio-labeled SFAs fail to bind TLR4, and numerous alternatives have been proposed (Chait and Kim, 2010). These

mechanisms are controversial and include SFA promotion of TLR4 dimerization, SFA induction of TLR4 surface expression, and SFA mediation of lipid-raft-dependent TLR4 activation (Chait and Kim, 2010). Recently, another possibility has been introduced (Schwartz et al., 2010). In this study, authors primed macrophage/monocyte cell lines and primary human monocytes with SFAs and then stimulated with a low dose of lipopolysaccharide (LPS). The combination of SFA and low-dose LPS synergistically amplified the inflammatory response. The amplified response was dependent on SFA metabolism to ceramide. In addition, SFA and ceramide both stimulated cellular signaling cascades not typically associated with TLR4; SFA and ceramide treatment resulted in activation of protein kinase C (PKC) ζ and the MAPKs Erk, JNK, and p38. Inhibition of these kinases, in turn, blocked the SFA-enhanced inflammatory response to LPS. In the context of obesity, these results suggest another model for SFA-induction of inflammation. On a regular basis, low-level LPS leaks into the circulation, due to absorption of gut bacteria, minor infection or food contamination. Low-level LPS in combination with HFD-induced elevation of SFAs could amplify inflammatory gene expression in macrophages and other tissues through the convergence of TLR4 with other signaling pathways (Chait and Kim, 2010). This study also highlights the role of ceramide as an important downstream signaling derivative of SFAs. Indeed, murine studies indicate that ceramide synthesis is required for TLR4-induced insulin resistance. At the same time, TLR4 is necessary for palmitate-induced insulin resistance and ceramide synthesis, and, thus, all three of these signaling molecules are necessary for induction of the insulin-resistance stress response (Bikman and Summers, 2011).

Ceramides accumulate in response to many other stress stimuli, including inflammatory cytokines, glucocorticoids and chemotherapy, and they regulate diverse metabolic phenotypes and signaling pathways (Bikman and Summers, 2011). Ceramides are formed *de novo* from

serine and palmitate in a sequence of four reactions. They can also be synthesized directly from sphingomyelin, the most abundant sphingolipid in mammals, and indirectly from catabolism of complex sphingolipids in salvage pathways. Thus, tight, hormonal regulation rather than substrate availability controls ceramide synthesis (Bikman and Summers, 2011).

Ceramide accumulation broadly inhibits cellular uptake of key nutrients, including glucose, amino acids, and fatty acids (Bikman and Summers, 2011). Short-chain ceramide mimetics prevent insulin-stimulated translocation of GLUT4 glucose transporter to the plasma membrane, while treatment of rodents with an inhibitor of *de novo* ceramide synthesis promotes insulin-dependent glucose uptake, even in the context of insulin-resistance causing factors, such as SFAs, glucocorticoids, and HFD. Although ceramides require intracellular FAs for synthesis, ceramide analogs inhibit the FA translocase CD36 from actively transporting FAs across the cell membrane. This occurs in a dose- and time-dependent manner and in the absence of CD36 mRNA expression changes, suggesting a real-time, protein-mediated response. Finally, in muscle cells, ceramide diminishes both basal and insulin-stimulated uptake of amino acids by decreasing the membrane (but not total cellular) levels of sodium-coupled neutral amino acid transporter 2 (SNAT2). Via control of central nutrient import, ceramides are important modulators of cellular survival. In addition, ceramides inhibits Akt, a major signaling protein involved in cellular survival, growth, and a variety of other responses. Ceramide inhibition of Akt is mediated by two pathways, whose relative importance varies by cell type. These pathways lead to activation of protein phosphatase 2A (PP2A) or PKC ζ . The PKC ζ -dependent pathway requires functional caveolae, small (50-100nm) lipid rafts that cause invagination of the plasma membrane. In general, lipid rafts may be an important means for cross-talk between lipid metabolic and protein signaling molecules. Interestingly, ceramide synthesis is inhibited by

expression of a constitutively active Akt, suggesting a mutually antagonistic ceramide-Akt relationship (Bikman and Summers, 2011).

The specificity of TLR4-induced inflammatory response for particular SFAs provides evidence that the role of SFAs in inflammation is stereospecific and may occur synergistically through activation of cellular signaling pathways independent of TLR4. In addition, *de novo* synthesis of ceramide from SFAs parallels a signal transduction cascade, where ceramide, as described above, shares signaling effector properties potentially as far reaching as its mutual-antagonist Akt.

Monounsaturated fatty acids as signaling molecules

Studies involving inhibition of stearoyl-CoA desaturase (SCD) have demonstrated a role for monounsaturated fatty acids (MUFAs) in the control of cellular metabolism, and, intriguingly, sometimes the origin of the MUFA, endogenous versus exogenous, is a key determinant of cellular phenotype. SCD is a $\Delta 9$ -desaturase, ER-membrane protein that catalyzes the biosynthesis of MUFA (oleate (C18:1 $\omega 9$) and palmitoleate (C16:1 $\omega 7$)) from stearate (C18:0) and palmitate (C16:0). Substantial insight into the physiological roles of MUFA comes from studies of mice with whole-body and tissue-specific SCD deficiency as well as mice treated with antisense oligonucleotides (ASO) against SCD.

Diets high in saturated fatty acids promote lipogenesis via activation of the master lipogenic transcription regulator sterol regulatory element-binding protein-1c (SREBP1c), and this effect was originally thought to be mediated by stearate. However, SCD^{-/-} mice, deficient in MUFA synthesis, are resistant to SFA-induced lipogenesis, demonstrating that MUFA are somehow involved in activation of SREBP1c. Equally surprising, in the absence of MUFA, SFAs induce a

diametrical metabolic response: activation of AMP-activated protein kinase (AMPK) and fatty-acid oxidation (Sampath et al., 2007). In addition, SCD^{-/-} mice and SCD-ASO treated-mice are resistant to lipogenesis and obesity even on a high fat diet (HFD) that consisted of 40-50% MUFA. This suggests that endogenous MUFA synthesis yields a pool of MUFA distinct from dietary, and the endogenous pool dictates the balance between energy storage and catabolism (Jiang et al., 2005; Ntambi et al., 2002).

Characterization of hepatic SCD deletion (L-SCD^{-/-}) mice refines the mechanism of endogenous MUFA in metabolic control and reveals a novel role for the MUFA oleate (Miyazaki et al., 2007). First, L-SCD^{-/-} on HFD are not resistant to obesity, and, more significantly, hepatic steatosis develops, indicating that hepatic lipogenesis continues in the absence of locally (hepatocyte-) synthesized MUFAs. Authors hypothesize that endogenous MUFAs from other tissues must compensate for loss of liver-produced MUFAs. The fundamental difference between endogenous and dietary MUFA in metabolic regulation is a mystery; perhaps endogenous MUFAs from extrahepatic tissues are secreted with cofactors that enable hepatic utilization for lipogenesis.

Additional insight into MUFA-dependent metabolic regulation is gained from L-SCD^{-/-} mice on high-carbohydrate diets (HCD). Under normal circumstances, a HCD should potentially activate hepatic SREBP1c as well as carbohydrate regulatory element-binding protein (ChREBP), leading to transcription of lipogenic enzymes, hepatic steatosis, and other metabolic syndrome symptoms. A high-sucrose, very low-fat (HSLVF) diet is an extreme form of the HCD, but L-SCD^{-/-} mice have severely limited lipogenesis and hepatic steatosis even under this condition. In addition, L-SCD^{-/-} have reduced triglyceride secretion and are hypoglycemic. Supplementing the HSLVF with 20% triolein (a triglyceride composed entirely of oleate acyl chains) but not

20% triostearin (stearate triglyceride) rescues all of the described HSLVF-induced phenotypes. Somehow supplementation with oleate, but not stearate, results in restored transcription and activation of SREBP1c and ChREBP. Activation of SREBP1c is a multi-step process that involves proteolytic cleavage and translocation to the nucleus. Authors suggest that oleate is involved in post-transcriptional processing of SREBP, because hepatic lipogenesis can be restored in global SCD^{-/-} mice by expression of a constitutively active nuclear form of SREBP1a in liver. Oleate supplementation also resolved hypoglycemia and restored the gluconeogenic metabolites in L-SCD^{-/-} mice on HSLVF, extending the known regulatory function of oleate to include carbohydrate in addition to lipid metabolism. This novel oleate-gluconeogenesis connection offers a potential explanation for oleate-mediated lipogenesis. The gluconeogenic metabolites glucose and glucose-6-phosphate (G6P) are agonists of the liver X receptor (LXR), which promotes transcription of SREBP1c and ChREBP, although the link between oleic acid and gluconeogenesis is unknown. Yet another potential mechanism connecting oleate to lipogenesis could be alteration in membrane biomechanics due to incorporation of oleate in membrane phospholipids, which, in turn, could affect SREBP1 maturation and translocation (Miyazaki et al., 2007). It remains to be seen if SCD's other MUFA product, palmitoleate, would also restore hepatic lipogenesis and gluconeogenesis on HSLVF. If the described regulatory functions are unique to oleate, then the location of the ω bond (ω 9, as opposed to ω 7 in palmitoleate) likely mediates some aspects of the regulatory mechanism.

Thus, murine models of SCD inhibition have highlighted the MUFA oleate as a signaling metabolite that regulates diverse metabolic pathways: gluconeogenesis and lipogenesis. Oleate regulation of lipogenesis may involve both transcriptional and post-transcriptional control; both mechanisms extend beyond allosteric inhibition of dedicated metabolic enzymes.

Polyunsaturated fatty acid signaling molecules and eicosanoid derivatives

The historical appreciation of polyunsaturated fatty acids (PUFAs) as more than a fuel source dates back at least 35 years (Bang et al., 1976). In an effort to understand the low incidence of cardiovascular mortality among Greenland Inuits, the lipid composition of their seafood-rich diets was compared to a typical Western diet, and ω 3 PUFA content, elevated in the Inuit population, was hypothesized to promote cardiovascular benefits (Bang et al., 1976). Subsequent clinical trials demonstrated the value of ω 3-PUFA diet supplementation in the promotion of cardiovascular health, and, in the last decade, insight into the molecular means by which PUFAs exert their physiological effects has developed (Arnold et al., 2010).

The roles of specific species and classes of PUFAs are functions of their chain length, degree of desaturation, and double-bond locations (especially the final (ω) double-bond location). The bioactivities of ω 3 and ω 6 PUFAs are the most well characterized, and, in mammals, these PUFA are products of essential fatty acids α -linolenic acid (LNA, C18:3 ω 3) and γ -linoleic acid (GLA, C18:2 ω 6), respectively. ω 7 and ω 9 PUFA are products of oleic acid (OA, C18:1 ω 9) and palmitoleic acid (POA, C16:1 ω 7), precursors supplied by diet or *de novo* lipogenesis via SCD Δ 9-desaturase activity. All PUFAs, regardless of precursor, are products of the same biosynthetic pathways, which take place in the ER and are composed of elongases and two desaturases, Fads1 (Δ 5) and Fads2 (Δ 6) (Guillou et al., 2010). However, PUFA synthesis is not entirely contained in the ER, as the synthesis of docosahexanoic acid (DHA C22:6 ω 3) and docosapentanoic acid (C22:5 ω 6) requires C24 precursor transport to the peroxisome for β -oxidation. Cellular PUFA are stored as phospholipids, and cellular concentrations are tightly controlled by binding proteins (fatty acid binding proteins (FABPs) and acyl-CoA binding proteins (ACBPs)). Well-established modes of signaling include the activation of phospholipids

(i.e., to diacyl glycerol and inositol triphosphate) and lipase-mediated release of PUFAs from phospholipids for oxidative metabolism to eicosanoids, pluripotent paracrine and autocrine signaling molecules. However, PUFA also have direct cellular signaling roles, which were first discovered in the context of nuclear-hormone-receptor control of gene expression (Sampath and Ntambi, 2005).

Inhibition of hepatic lipogenesis by ω 3 and ω 6 PUFA is potent enough to overcome prolipogenic signaling of postprandial insulin and carbohydrates (Sampath and Ntambi, 2005). These PUFA reduce cellular fat stores doubly, through simultaneous activation of fatty-acid β oxidation and inhibition of fatty acid synthesis (Sampath and Ntambi, 2005). PUFA regulation of gene expression occurs on the order of minutes, a timescale consistent with ligand-mediated regulation (Benatti et al., 2004). Indeed, ω 3 and ω 6 PUFA bind to and activate the peroxisome proliferator-activated receptor α (PPAR α), a nuclear hormone receptor that promotes expression of genes involved in mitochondrial and peroxisomal FA oxidation (Sampath and Ntambi, 2005). Numerous ω 3 and ω 6 PUFA bind PPAR α , but with varying affinities. For example, in primary hepatocytes, eicosapentanoic acid (EPA, C20:5 ω 3) is a much stronger PPAR α agonist than arachidonic acid (AA, C20:4 ω 6). Importantly, transactivation assays show that many ω 3 and ω 6 PUFA activate PPAR α at an order of magnitude lower concentration than that found in human serum, indicating a physiological role for PUFA-receptor binding (Sampath and Ntambi, 2005).

ω 3 and ω 6 PUFA also potently inhibit the master lipogenic transcription factor, SREBP1c, in liver. Evidence suggests that this regulation is mediated by enhanced decay of SREBP1c mRNA and proteasomal degradation of active, nuclear SREBP1c, although the molecular detail of these mechanisms is unknown (Jump et al., 2005). The mechanism does appear to be PUFA-species

specific. The hierarchy of PUFA-mediated SREBP1c mRNA degradation was found to be: EPA = AA > GLA > oleic acid (Benatti et al., 2004). Thus, PUFA regulation of gene expression is mediated by diverse and PUFA-species-specific mechanisms.

PUFA control more than gene expression, however, as transcriptional control does not entirely explain PUFA-mediated inhibition of hepatic triglyceride secretion (Pan et al., 2004). In order to elucidate the non-transcriptional mechanism of ω 3 PUFA inhibition of triglyceride secretion, authors treated primary hepatocytes with DHA and evaluated several hypothesized mechanisms. They found that DHA treatment resulted in lipid peroxide formation and oxidative damage to apolipoprotein B (ApoB), the critical apolipoprotein in very-low-density lipoprotein (VLDL) particles. Oxidation of ApoB lead to its degradation and diminished VLDL secretion. Interestingly, ApoE, the core component of high-density lipoprotein (HDL) particles was resistant to DHA-induced oxidative damage and degradation, suggesting that this mechanism is specific to control of VLDL. In addition, by treating the hepatocytes with other fatty acids, authors showed that ApoB secretion inversely correlated with acyl-chain desaturation, independent of the ω -bond position (Pan et al., 2004). Although authors limited their study to ω 3 and ω 6 PUFA, this mechanism could feasibly extend to ω 7 and ω 9 PUFA.

Although PUFA, especially the well-characterized ω 3 and ω 6 species, directly modulate many facets of cellular behavior, additional consideration of PUFAs as initiators of eicosanoid signaling strengthens their classification as signaling molecules. Technically, an eicosanoid is any long-chain oxygenated PUFA product; however, the “quintessential properties of ‘true’ eicosanoids are their stereochemical precision in formation and recognition, their potency in the nanomolar range *in vitro*, and their bona fide biological activities” (Funk, 2001). Eicosanoids act in an autocrine or paracrine manner and are rapidly synthesized *de novo* from lipase-mediated

membrane release of PUFAs, stimulated by mechanical stress and other cellular cues (i.e., growth factor, cytokine). There are three major branches of eicosanoid signaling: prostaglandins, leukotrienes, and other cytochrome P450 pathways, each of which contains subpathways, resulting in myriad eicosanoid products, each with diverse and specific bioactivities (Arnold et al., 2010).

In addition to the diversity of biochemical pathways, eicosanoid responses depend upon the PUFA precursor, stimulus, and cell type (Funk, 2001). AA (C20:4 ω 6) is the prototypical PUFA precursor of eicosanoids, and it is typically involved in promotion of inflammation. Other (ω 3) PUFA, such as EPA and DHA, were initially thought to mediate anti-inflammatory effects by AA-substrate competition in the eicosanoid biosynthetic pathways. However, it is now appreciated that the eicosanoid products of EPA and DHA, suggestively named resolvins and protectins, independently promote inflammation resolution and are biologically active lipid mediators in their own right (Arnold et al., 2010; Spite and Serhan, 2010). In addition, other PUFAs are processed into potent eicosanoids. For example, a recent study revealed that 5-oxo-C20:3, the 5-oxo product of the ω 9-PUFA mead acid (MA, C20:3), activated eosinophil chemotaxis via the 5-oxo-ETE receptor with potency equivalent to 5-oxo-ETE, the AA product for which the receptor was named (Patel et al., 2008). This is a rare example of a species-specific, ω 9-PUFA signaling mechanism. As the field grows, it is likely that many more roles for PUFA species in nuclear hormone activation, post-transcriptional regulation, eicosanoid signaling, and other facets of cellular regulation will be discovered.

Conclusions

Large-scale, MS-based characterizations of protein PTMs and metabolites have identified an astonishing diversity in protein and metabolic states. These MS experiments have prompted molecular-level investigations. Even though these studies seem limited in the context of so many newly-discovered and uncharacterized biomolecules, they clearly indicate that (1) metabolites must be considered as cellular signaling molecules in addition to energy sources and (2) PTM of metabolic enzymes is not gratuitous, even if that particular class of PTMs has historically been characterized in the regulation of metabolic control by other means. These conclusions present a more nuanced model of cellular control. An understanding of metabolism would be incomplete without consideration of enzyme PTMs and, likewise, prediction of cellular response relies upon incorporation of novel metabolite signaling mechanisms into known cellular signaling pathways. As potentially “complicating” as large-scale measurement of PTMs and metabolites are to prior models, they will fundamentally contribute to our understanding of cellular signaling networks and human health and disease.

The network-level acetylation changes that accompany the switch from fed to fasting metabolism present an interesting concept for PTM-metabolite coordination. Simultaneous consideration of protein acetylation and metabolic flux in the glucose-dependent context offers a glimpse of PTMs as protein metabolites and metabolites as protein modulators, and, together, suggests evidence for potent metabolite-PTM feedbacks. In the context of cancer, tyrosine phosphorylation is often dysregulated (Blume-Jensen and Hunter, 2001), and evidence suggests that tyrosine phosphorylation of metabolic enzymes plays a crucial role in the switch from oxidative phosphorylation to aerobic glycolysis (Hitosugi et al., 2009). In the context of cancer, it is tempting to speculate about the existence of large-scale, full-circle metabolite-

phosphotyrosine (as well as other PTM) feedbacks. These feedbacks may be more apparent upon metabolic perturbations (i.e., hypoxia). Conversely, perturbations of PTM-modulating proteins can result in dramatic metabolic changes (Banno et al., 2010; Bence et al., 2006), and these perturbations present a complementary approach for exploration of metabolite-PTM interactions.

Chapter 2 describes a study whose aim is to systematically develop a molecular-network understanding of a single genetic perturbation to a protein tyrosine phosphorylation network; this perturbation had previously been shown to promote metabolic changes at the physiological level by unknown molecular means (Delibegovic et al., 2009). In this study, global MS-based analyses are used to quantitatively measure site-specific protein-tyrosine phosphorylation and molecularly-resolved metabolic changes in response to this genetic perturbation *in vivo*. The system is additionally characterized in the context of a nutrient challenge, high fat diet. Surprisingly, HFD promotes network-level phosphotyrosine changes that scale with the more-direct phosphotyrosine perturbation, phosphatase deletion.

Given the breadth of the MS-based experimental analyses, several modes of statistical modeling techniques were developed to detect pathway-specificity in tyrosine phosphorylation network changes and to infer multivariate relationships between molecular measurements and metabolic phenotypes. The results from the computational analysis are described briefly in Chapter 2. In Chapter 3, these computational methods are described and motivated in full. Chapter 4 describes both contributions and limitations of Chapters 2 and 3 results. In addition, Chapter 4 proposes new research directions that develop from their consideration.

References

- Arnold, C., Konkol, A., Fischer, R., and Schunck, W.H. (2010). Cytochrome P450-dependent metabolism of omega-6 and omega-3 long-chain polyunsaturated fatty acids. *Pharmacol Rep* 62, 536-547.
- Bang, H.O., Dyerberg, J., and Hjoorne, N. (1976). The composition of food consumed by Greenland Eskimos. *Acta Med Scand* 200, 69-73.
- Banno, R., Zimmer, D., De Jonghe, B.C., Atienza, M., Rak, K., Yang, W., and Bence, K.K. (2010). PTP1B and SHP2 in POMC neurons reciprocally regulate energy balance in mice. *The Journal of clinical investigation* 120, 720.
- Benatti, P., Peluso, G., Nicolai, R., and Calvani, M. (2004). Polyunsaturated fatty acids: biochemical, nutritional and epigenetic properties. *J Am Coll Nutr* 23, 281-302.
- Bence, K.K., Delibegovic, M., Xue, B., Gorgun, C.Z., Hotamisligil, G.S., Neel, B.G., and Kahn, B.B. (2006). Neuronal PTP1B regulates body weight, adiposity and leptin action. *Nat Med* 12, 917-924.
- Bikman, B.T., and Summers, S.A. (2011). Ceramides as modulators of cellular and whole-body metabolism. *J Clin Invest* 121, 4222-4230.
- Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. *Nature* 411, 355-365.
- Boulpaep, E., Boron, W.F., Caplan, M.J., Transduction, S., Cantley, L., Igarashi, P., Aronson, P.S., Moczydlowski, E.G., Excitability, E., and Potentials, A. (2003). Medical physiology: a cellular and molecular approach. *Signal Transduction* 48, 1073-1076.
- Cai, D., Yuan, M., Frantz, D.F., Melendez, P.A., Hansen, L., Lee, J., and Shoelson, S.E. (2005). Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. *Nat Med* 11, 183-190.
- Cao, H., Gerhold, K., Mayers, J.R., Wiest, M.M., Watkins, S.M., and Hotamisligil, G.S. (2008). Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell* 134, 933-944.
- Chait, A., and Kim, F. (2010). Saturated fatty acids and inflammation: who pays the toll? *Arterioscler Thromb Vasc Biol* 30, 692-693.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834-840.
- Cohen, P. (2006). The twentieth century struggle to decipher insulin signalling. *Nat Rev Mol Cell Biol* 7, 867-873.
- Cooper, J.A., Reiss, N.A., Schwartz, R.J., and Hunter, T. (1983). Three glycolytic enzymes are phosphorylated at tyrosine in cells transformed by Rous sarcoma virus. *Nature* 302, 218-223.
- Dang, C.V. (2009). PKM2 tyrosine phosphorylation and glutamine metabolism signal a different view of the Warburg effect. *Sci Signal* 2, pe75.

- Davis, J.E., Gabler, N.K., Walker-Daniels, J., and Spurlock, M.E. (2008). Tlr-4 deficiency selectively protects against obesity induced by diets high in saturated fat. *Obesity (Silver Spring)* *16*, 1248-1255.
- Del Rosario, A.M., and White, F.M. (2010). Quantifying oncogenic phosphotyrosine signaling networks through systems biology. *Curr Opin Genet Dev* *20*, 23-30.
- Delibegovic, M., Zimmer, D., Kauffman, C., Rak, K., Hong, E.G., Cho, Y.R., Kim, J.K., Kahn, B.B., Neel, B.G., and Bence, K.K. (2009). Liver-specific deletion of protein-tyrosine phosphatase 1B (PTP1B) improves metabolic syndrome and attenuates diet-induced endoplasmic reticulum stress. *Diabetes* *58*, 590-599.
- Dix, M.M., Simon, G.M., and Cravatt, B.F. (2008). Global mapping of the topography and magnitude of proteolytic events in apoptosis. *Cell* *134*, 679-691.
- Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A.L., Normandin, D., Cheng, A., Himms-Hagen, J., and Chan, C.C. (1999). Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* *283*, 1544.
- Fan, J., Hitosugi, T., Chung, T.W., Xie, J., Ge, Q., Gu, T.L., Polakiewicz, R.D., Chen, G.Z., Boggon, T.J., Lonial, S., *et al.* (2011). Tyrosine phosphorylation of lactate dehydrogenase A is important for NADH/NAD(+) redox homeostasis in cancer cells. *Mol Cell Biol* *31*, 4938-4950.
- Feng, J., Lucchinetti, E., Enkavi, G., Wang, Y., Gehrig, P., Roschitzki, B., Schaub, M.C., Tajkhorshid, E., Zaugg, K., and Zaugg, M. (2010). Tyrosine phosphorylation by Src within the cavity of the adenine nucleotide translocase 1 regulates ADP/ATP exchange in mitochondria. *Am J Physiol Cell Physiol* *298*, C740-748.
- Ficarro, S.B., McClelland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., and White, F.M. (2002). Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature biotechnology* *20*, 301-305.
- Funk, C.D. (2001). Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* *294*, 1871-1875.
- Guan, K.L., and Xiong, Y. (2011). Regulation of intermediary metabolism by protein acetylation. *Trends Biochem Sci* *36*, 108-116.
- Guarente, L. (2011a). The logic linking protein acetylation and metabolism. *Cell Metab* *14*, 151-153.
- Guarente, L. (2011b). Sirtuins, Aging, and Metabolism. *Cold Spring Harb Symp Quant Biol*.
- Guillou, H., Zdravcevic, D., Martin, P.G., and Jacobsson, A. (2010). The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog Lipid Res* *49*, 186-199.
- Hebert-Chatelain, E., Jose, C., Cortes, N.G., Dupuy, J.W., Rocher, C., Dachary-Prigent, J., and Letellier, T. (2012). Preservation of NADH ubiquinone-oxidoreductase activity by Src kinase-mediated phosphorylation of NDUFB10. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*.
- Hitosugi, T., Fan, J., Chung, T.W., Lythgoe, K., Wang, X., Xie, J., Ge, Q., Gu, T.L., Polakiewicz, R.D., Roesel, J.L., *et al.* (2011). Tyrosine phosphorylation of mitochondrial pyruvate dehydrogenase kinase 1 is important for cancer metabolism. *Mol Cell* *44*, 864-877.

- Hitosugi, T., Kang, S., Vander Heiden, M.G., Chung, T.W., Elf, S., Lythgoe, K., Dong, S., Lonial, S., Wang, X., Chen, G.Z., *et al.* (2009). Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. *Sci Signal* 2, ra73.
- Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic acids research* 40, D261-D270.
- Huang, H.S., Nagane, M., Klingbeil, C.K., Lin, H., Nishikawa, R., Ji, X.D., Huang, C.M., Gill, G.N., Wiley, H.S., and Cavenee, W.K. (1997). The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling. *J Biol Chem* 272, 2927-2935.
- Huang, P.H., Mukasa, A., Bonavia, R., Flynn, R.A., Brewer, Z.E., Cavenee, W.K., Furnari, F.B., and White, F.M. (2007). Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci U S A* 104, 12867-12872.
- Jiang, G., Li, Z., Liu, F., Ellsworth, K., Dallas-Yang, Q., Wu, M., Ronan, J., Esau, C., Murphy, C., Szalkowski, D., *et al.* (2005). Prevention of obesity in mice by antisense oligonucleotide inhibitors of stearoyl-CoA desaturase-1. *J Clin Invest* 115, 1030-1038.
- Jump, D.B., Botolin, D., Wang, Y., Xu, J., Christian, B., and Demeure, O. (2005). Fatty acid regulation of hepatic gene transcription. *J Nutr* 135, 2503-2506.
- Kim, S.C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., *et al.* (2006). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* 23, 607-618.
- Konner, A.C., and Bruning, J.C. (2011). Toll-like receptors: linking inflammation to metabolism. *Trends Endocrinol Metab* 22, 16-23.
- Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology* 7, 952-958.
- Miyazaki, M., Flowers, M.T., Sampath, H., Chu, K., Otzelberger, C., Liu, X., and Ntambi, J.M. (2007). Hepatic stearoyl-CoA desaturase-1 deficiency protects mice from carbohydrate-induced adiposity and hepatic steatosis. *Cell Metab* 6, 484-496.
- Naegle, K.M., Gymrek, M., Joughin, B.A., Wagner, J.P., Welsch, R.E., Yaffe, M.B., Lauffenburger, D.A., and White, F.M. (2010). PTMScout, a Web resource for analysis of high throughput post-translational proteomics studies. *Mol Cell Proteomics* 9, 2558-2570.
- Nakatani, Y., Kaneto, H., Kawamori, D., Hatazaki, M., Miyatsuka, T., Matsuoka, T., Kajimoto, Y., Matsuhisa, M., Yamasaki, Y., and Hori, M. (2004). Modulation of the JNK pathway in liver affects insulin resistance status. *Journal of Biological Chemistry* 279, 45803.
- Nomura, D.K., Long, J.Z., Niessen, S., Hoover, H.S., Ng, S.W., and Cravatt, B.F. (2010). Monoacylglycerol lipase regulates a fatty acid network that promotes cancer pathogenesis. *Cell* 140, 49-61.

- Ntambi, J.M., Miyazaki, M., Stoehr, J.P., Lan, H., Kendziorski, C.M., Yandell, B.S., Song, Y., Cohen, P., Friedman, J.M., and Attie, A.D. (2002). Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. *Proc Natl Acad Sci U S A* 99, 11482-11486.
- Owens-Bryson, B.D. (unpublished work).
- Pan, M., Cederbaum, A.I., Zhang, Y.L., Ginsberg, H.N., Williams, K.J., and Fisher, E.A. (2004). Lipid peroxidation and oxidant stress regulate hepatic apolipoprotein B degradation and VLDL production. *J Clin Invest* 113, 1277-1287.
- Patel, P., Cossette, C., Anumolu, J.R., Gravel, S., Lesimple, A., Mamer, O.A., Rokach, J., and Powell, W.S. (2008). Structural requirements for activation of the 5-oxo-6E,8Z, 11Z,14Z-eicosatetraenoic acid (5-oxo-ETE) receptor: identification of a mead acid metabolite with potent agonist activity. *J Pharmacol Exp Ther* 325, 698-707.
- Richard, A.J., and Stephens, J.M. (2011). Emerging roles of JAK-STAT signaling pathways in adipocytes. *Trends Endocrinol Metab* 22, 325-332.
- Saghatelian, A., and Cravatt, B.F. (2005). Global strategies to integrate the proteome and metabolome. *Curr Opin Chem Biol* 9, 62-68.
- Saltiel, A.R., and Pessin, J.E. (2002). Insulin signaling pathways in time and space. *Trends Cell Biol* 12, 65-71.
- Sampath, H., Miyazaki, M., Dobrzyn, A., and Ntambi, J.M. (2007). Stearoyl-CoA desaturase-1 mediates the pro-lipogenic effects of dietary saturated fat. *J Biol Chem* 282, 2483-2493.
- Sampath, H., and Ntambi, J.M. (2005). Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annu Rev Nutr* 25, 317-340.
- Schmelzle, K., Kane, S., Gridley, S., Lienhard, G.E., and White, F.M. (2006). Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 55, 2171-2179.
- Schwartz, E.A., Zhang, W.Y., Karnik, S.K., Borwege, S., Anand, V.R., Laine, P.S., Su, Y., and Reaven, P.D. (2010). Nutrient modification of the innate immune response: a novel mechanism by which saturated fatty acids greatly amplify monocyte inflammation. *Arterioscler Thromb Vasc Biol* 30, 802-808.
- Spange, S., Wagner, T., Heinzl, T., and Kramer, O.H. (2009). Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int J Biochem Cell Biol* 41, 185-198.
- Spite, M., and Serhan, C.N. (2010). Novel lipid mediators promote resolution of acute inflammation: impact of aspirin and statins. *Circ Res* 107, 1170-1184.
- Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T.V., Cross, J.R., and Thompson, C.B. (2009). ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* 324, 1076-1080.
- Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A., and White, F.M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* 104, 5860.
- Wolf-Yadlin, A., Sevecka, M., and MacBeath, G. (2009). Dissecting protein function and signaling using protein microarrays. *Curr Opin Chem Biol* 13, 398-405.

Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A., and White, F.M. (2005). Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 4, 1240-1250.

2. Bridging the gap between protein-tyrosine phosphorylation networks, metabolism and physiology in liver-specific PTP1b deletion mice

Authors

Emily R. Miraldi^{1,2}, Randall H. Friedline^{3*}, Hannah Johnson^{2,4*}, Hadar Sharfi^{2*}, Tejia Zhang⁵, Ken S. Lau^{2,4,6}, Hwi Jin Ko³, Kevin M. Haigis^{4,6}, Mike Yaffe², Douglas A. Lauffenburger⁴, Barbara B. Kahn⁸, Jason K. Kim³, Benjamin G. Neel^{9,10,+}, Alan Saghatelian^{5,+}, Forest M. White^{2,4}

Affiliations

¹Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.,

²Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.,

³Program in Molecular Medicine and Department of Medicine, Division of Endocrinology, Metabolism and Diabetes, University of Massachusetts Medical School, Worcester, Massachusetts 01605, U.S.A.,

⁴Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.,

⁵Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, U.S.A.,

⁶Molecular Pathology Unit, Center for Cancer Research, and Center for Systems Biology, Massachusetts General Hospital, Charlestown, MA 02129, U.S.A.,

⁷Department of Pathology, Harvard Medical School, Boston, MA 02115, U.S.A.

⁸Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, U.S.A.

⁹Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 2M9, Canada,

¹⁰Campbell Family Cancer Research Institute, Ontario Cancer Institute and Princess Margaret Hospital, University Health Network, Toronto, ON M5G 1K7, Canada.

Introduction

Metabolic or insulin resistance syndrome describes a group of commonly associated disorders, including central obesity, hyperglycemia, hyperinsulinemia, dyslipidemia, and hypertension, that promote the development of Type 2 Diabetes Mellitus (T2DM), cardiovascular disease (CAD), cancer, polycystic ovarian disease, and nonalcoholic fatty liver disease (NAFLD) (Biddinger and Kahn, 2006; Kashyap and Defronzo, 2007). It is associated with 50% increased risk of cardiovascular mortality and 30% enhanced risk of mortality from all causes (Sidorenkov et al., 2010). In tandem with obesity, metabolic syndrome and its comorbidities have spread worldwide, affecting individuals across all socioeconomic circumstances (Popkin et al., 2012).

In the last twelve years, the liver has emerged as both a preeminent contributor to metabolic syndrome pathogenesis and a target tissue for therapeutical intervention (Rutter, 2000). These roles are consequences of hepatic insulin signaling, which critically mediates homeostatic changeover from the fasted to fed state. In response to nutrient cues, insulin is secreted from the pancreatic β cells directly into the portal circulation. The resulting high portal insulin levels prime the liver for rapid metabolic transition, stimulating fuel storage (via glycogen synthesis, lipogenesis, and lipoprotein synthesis) and suppressing pathways that supply fuel to the rest of the body (gluconeogenesis, glycogenolysis, and very low density lipoprotein (VLDL) secretion) (Michael et al., 2000).

Insulin signaling is initiated by the extracellular binding of insulin to insulin receptors (IRs), transmembrane receptor tyrosine kinases (RTKs), expressed on the surface of responsive cells. Ligand binding induces IR transphosphorylation of cytosolic tyrosine residues, which enhance IR kinase activity and promote localization and tyrosine phosphorylation of IR substrates,

including insulin receptor substrates (IRSs) and Shc. The IRSs serve as key scaffolds for the binding and activation of additional signaling proteins and cascades, which orchestrate metabolic transitions (Saltiel and Pessin, 2002).

In the context of metabolic syndrome, liver and other insulin-sensitive tissues become insulin-resistant. Hepatic insulin resistance results in unchecked gluconeogenesis, driving hyperinsulinemia and hyperglycemia. In metabolic syndrome, the liver tissue itself accumulates lipid as a result of free fatty acid (FFA) flux from insulin-resistant adipose tissue and altered hepatic lipid metabolism (Samuel et al., 2004). Accumulation of hepatic lipid (steatosis) is associated with inflammation and compromised liver function (Coleman and Lee, 2004). While the mechanisms of insulin resistance are the subject of intense debate, there is a consensus that insulin resistance develops from impaired cellular signaling downstream of the IR (Taniguchi et al., 2006).

In the context of metabolic syndrome in the liver, pathways downstream of the insulin receptor exhibit differential insulin resistance (Shimomura et al., 2000). In genetic mouse models of obesity and lipodystrophy, leptin deficiency led to hyperphagia and the hallmarks of T2DM: hyperglycemia, hyperinsulinemia, and hypertriglyceridemia. Livers were enlarged with steatosis, and there was an increase in mRNAs encoding endogenous sterol response element binding protein-1c (SREBP1c), an insulin-responsive, transcriptional regulator of lipogenesis. Insulin signaling also regulates FOXO1, a gluconeogenesis-promoting transcription factor, through inhibitory phosphorylation. Intriguingly, whereas both FoxO1 and SREBP1c are regulated by insulin, livers in both models were resistant to insulin suppression of gluconeogenesis (via FoxO1) but responsive to insulin promotion of lipogenesis (via SREBP1c) (Shimomura et al., 2000).

This phenomenon of pathway-selective hepatic insulin resistance is also relevant to metabolic syndrome in humans. A more recent study, using stable isotope feeding followed by liver biopsy, found that liver lipogenesis is not sensitive to nutritional state in obese patients with NAFLD and elevated insulin resistance (Donnelly et al., 2005). In contrast to healthy patients, whose rate of lipogenesis cycles up and down with feeding and fasting, obese, insulin resistant patients' rate of lipogenesis remained constant and did not decrease with fasting. Several rodent models of diet-induced hepatic insulin resistance also demonstrate selective hepatic sensitivity to insulin promotion of SREBP1c (Samuel et al., 2004; Shimizu et al., 2003). Orthogonal evidence for selective insulin resistance in liver comes from studies in liver-specific, insulin-receptor-knock-out (LIRKO) mice, which develop severe whole-body insulin resistance due to defective hepatic insulin signaling (Michael et al., 2000). Unlike other models of obesity and insulin resistance, SREBP1c is down-regulated in LIRKO mice relative to controls. NAFLD and hypertriglyceridemia do not occur, because LIRKO mice have absolute inhibition of hepatic insulin signaling (Biddinger et al., 2008).

Selective insulin resistance has implications for metabolic syndrome and T2DM therapy. "Brute force" treatment of T2DM with large doses of insulin may effectively control blood glucose, but high doses of insulin may also enhance hepatic lipogenesis, and thereby increase lipotoxicity and development of NAFLD. Treatments that selectively improve insulin sensitivity in the hepatic gluconeogenesis pathway are preferred. With such a treatment, insulin levels and hepatic SREBP1c levels would fall, effectively curbing NAFLD development (Brown and Goldstein, 2008).

Evidence suggests that treatment targeting inhibition of hepatic protein-tyrosine phosphatase 1B (PTP1B) would not only selectively improve insulin sensitivity in the pathway leading to

suppression of hepatic gluconeogenesis but may also selectively and independently decrease hepatic SREBP1c expression (Delibegovic et al., 2009). Liver-specific PTP1B deletion mice (L-PTP1b^{-/-}) show clinically attractive metabolic phenotypes under HFD stress. In comparison to control mice, they exhibit improved glucose homeostasis, lipid metabolism, and resistance to HFD-induced endoplasmic reticulum (ER) stress. Importantly, L-PTP1b^{-/-} mice have comparable weight gain to control on HFD, suggesting that these phenotypes are the result of altered cellular signaling and metabolism, rather than secondary effects of reduced adiposity (Delibegovic et al., 2009).

PTP1B is most notably a negative regulator of insulin and leptin signaling, as whole-body PTP1B^{-/-} mice are lean, hypersensitive to insulin, and resistant to diet-induced obesity (Elchebly et al., 1999). PTP1b is a ~50kDa protein anchored to the cytoplasmic face of the endoplasmic reticulum (ER) and is a well-established negative regulator of insulin signaling, via dephosphorylation of tyrosine residues on IR and potentially insulin receptor substrate 1 (IRS-1), and leptin signaling, via dephosphorylation of tyrosine residues on Janus kinase 2 (Jak2). In addition, PTP1b regulates specific phosphotyrosine sites on proteins in other pathways (Src, MAPK, cell adhesion and motility) (Yip et al., 2010) which may also physiologically impact PTP1b^{-/-} mice.

Hepatic PTP1b deletion potentiates insulin signaling, which leads to nutrient-dependent inhibition of hepatic glucose production and whole-body improvements in glucose homeostasis in the context of obesity. Multiple measurements indicate improved lipid metabolism in L-PTP1b^{-/-} mice relative to control mice. Gene expression of SREBP1c is decreased relative to control, as is a master cholesterol biogenesis regulator, SREBP2. Physiological measurements corroborate the gene expression data; L-PTP1b^{-/-} mice have significantly decreased serum

triglyceride and cholesterol, as well as decreased liver cholesterol content. In addition, at five weeks of HFD but not at 16 weeks, L-PTP1b^{-/-} mice have decreased hepatic steatosis (Delibegovic et al., 2009).

The resistance to HFD-induced ER stress in L-PTP1b^{-/-} mice is therapeutically attractive as well. Over the last decade, ER stress, in the context of inflammation and lipotoxicity, has emerged as a crucial component of metabolic syndrome pathogenesis (Ji and Kaplowitz, 2006; Ozcan et al., 2004). The ER plays a central role in the cellular stress response (Zhang and Kaufman, 2008). Pathological stress conditions, including excess lipid accumulation and abnormal intracellular energy fluxes, disrupt ER homeostasis and lead to accumulation of misfolded proteins in the ER lumen (Ozcan et al., 2004). ER stress sensors, embedded in the ER membrane, activate signal transduction pathways to regain ER homeostasis. Chronic obesity-induced ER stress leads to activation of stress kinase JNK and NFκB-dependent expression of inflammatory cytokines (i.e., Il-1β, Il-6, and TNFα), which cause and exacerbate insulin resistance (Gregor and Hotamisligil, 2011). In fact, inhibition of hepatic NFκB or JNK is sufficient to protect mice from HFD-induced insulin resistance (Cai et al., 2005; Nakatani et al., 2004). Hepatic PTP1b deletion disrupts activation of all three ER stress sensors (IRE1α, PERK, and ATF6)(Agouni et al., 2011; Delibegovic et al., 2009). Thus, in terms of both whole-body and liver-specific physiology, deleting or inhibiting hepatic PTP1b holds apparent therapeutic potential for treatment of metabolic syndrome.

While hepatic PTP1b inhibition or deletion has been associated with positive physiological effects, the underlying molecular mechanisms are not well understood. Reasoning that the phenotypic effects of PTP1b deletion were likely due to altered tyrosine phosphorylation of protein signaling networks, we coupled statistical modeling techniques with quantitative, mass-

spectrometry (MS)-based measurement of hepatic protein tyrosine phosphorylation to determine, in an unbiased and site-specific manner, the phosphotyrosine network drivers mediating resistance to HFD in L-PTP1b^{-/-} mice. These studies revealed that hepatic phosphotyrosine networks are sensitive to both PTP1b deletion and HFD, and that the PTP1b-dependent phosphorylation sites are significantly enriched on enzymes with diverse lipid metabolic functions. To elucidate a role for this largely uncharacterized set of PTP1b-dependent phosphosites in lipid metabolic control, we globally profiled hepatic lipids and found that hepatic PTP1b deletion alters metabolism of fatty acids and triglycerides. To determine the phosphorylation network states associated with altered lipid metabolism, multivariate statistical models were developed. These models not only highlight the mechanistic underpinnings connecting PTP1b deletion, altered tyrosine phosphorylation signaling and altered lipid metabolism, but also offer quantitative predictions for selected network perturbations.

Results

Study design

L-PTP1b^{-/-} mice were generated by crossing PTP1b-floxed (PTP1b-fl/fl) mice with mice expressing Cre recombinase under the control of the albumin promoter (Alb-Cre), resulting in Alb-Cre PTP1b-fl/fl mice that specifically lack hepatic PTP1b expression (Delibegovic et al., 2009). Twenty-one L-PTP1b^{-/-} and 28 Alb-Cre control mice were fed a normal chow (NC) diet (14% kcal from fat) until 12 weeks of age, at which point mice were placed on either HFD (55% kcal from fat) or continued on NC. These diets were maintained until 19.5 weeks of age (Figure 2.1A). To gauge the physiological consequences of HFD, relevant quantitative physiological measurements, including serum markers of insulin sensitivity and lipid metabolism, were made

at 17.5 and 18.5 weeks. As expected, both L-PTP1b^{-/-} and control mice had elevated levels of fed insulin, glucose, and leptin on HFD relative to NC, indicating that insulin and leptin resistance developed during the course of the diets (Table 2.1). While L-PTP1b^{-/-} and control mice had similar fed glucose and insulin for a given diet, L-PTP1b^{-/-} mice had significantly elevated fasted glucose relative to controls on both diets and reduced fasted insulin on NC (Table 2.1).

At 19.5 weeks, mice were sacrificed and their livers flash-frozen for biochemical analysis. To determine the molecular effects of PTP1b deletion and HFD on insulin response, liver was harvested under either basal or insulin stimulation conditions (Figure 2.1A). Each combination of experimental conditions (genotype, diet, and stimulation) included 3-8 mice, thereby enabling statistical analysis for each condition.

HFD and genotype drive global phosphotyrosine network changes

Relative quantification of protein phosphotyrosine levels across each individual liver sample was accomplished by combining results from ten 8-plexed, phosphotyrosine-immunoprecipitation (IP)-Immobilized Metal Affinity Chromatography (IMAC)-Liquid Chromatography (LC)-Tandem MS (MS/MS) experiments (Figure 2.1B) (Zhang et al., 2005). This analysis resulted in the identification and quantification of 301 tyrosine phosphosites on 266 proteins (data available upon request). To visualize the most important trends across the multidimensional phosphotyrosine datasets, we applied principle component analysis (PCA). The most dominant phosphosite trends dictate the locations of liver samples in the principle component plane, and here the plane naturally segregates liver samples according to diet and genotype conditions for both the insulin-stimulated and basal datasets (Figure 2.2A). This suggests that diet and genotype are the dominant causes of phosphotyrosine network perturbations. In fact,

individually, genotype and diet explain 21-23% and 15-17% of the phosphorylation variance. The PCA plot demonstrates that the phosphorylation data, by itself, is sufficient to distinguish genotype and diet among the various mice.

Correlation-based statistical analysis was used to detect and order PTP1b-dependent and diet-dependent phosphorylation sites in heat maps (Figure 2.2B and Supplementary Figure 2.1, respectively). Due to the intrinsic irreproducibility of data-dependent mass spectrometric analysis (Wolf-Yadlin et al., 2007), the data set is somewhat sparse, with many of the phosphorylation sites measured in only some of the mouse livers. However, using the correlation-based analysis, we identified large clusters of PTP1b-dependent and diet-dependent sites. It is important to note that alterations in the phosphorylation network reflect rewiring of the signaling network over time, leading to increased phosphorylation on some sites and decreased phosphorylation on others in response to chronic hepatic PTP1b deletion.

PTP1b-dependent phosphosites are enriched for insulin signaling, REDOX, and lipid metabolic pathways

We adapted a gene set enrichment analysis framework (Subramanian et al., 2005) for analysis of our quantitative protein phosphorylation data (Chapter 3) and identified several protein pathways that were enriched for PTP1b-dependent phosphotyrosine regulation. In total, we tested for PTP1b-dependence in 59 hand-curated phosphosite sets, composed of protein pathway annotations from Kegg, cellular locations from GO, protein structural information (PFAM and SwissPro), a curated list of PTP1b substrates (Ren et al., 2011), as well as a set of insulin-sensitive phosphorylation sites from a previous study (Schmelzle et al., 2006). In this analysis, the direction of PTP1b-dependence was taken into account, as we searched for phosphosite sets that showed a concordant increase or decrease in phosphorylation upon PTP1b deletion. Seven

phosphosite sets show significant PTP1b-dependent tyrosine phosphorylation (FDR<0.01, Table 2.2). Consistent with its role in regulating insulin sensitivity, phosphosites involved in insulin signaling have increased phosphorylation upon PTP1b deletion; these include sites on the receptor (IR), receptor substrates (IRS1, IRS2, Shc), and several other canonical insulin signaling proteins (PI3K, Erk1, Erk2, and others). Enrichment of PTP1b-dependent phosphosites on proteins localized to the ER and cytosol is in agreement with the cellular localization of PTP1b to the cytoplasmic face of the ER, providing further confirmation that this approach identifies relevant connections among the PTP1b-dependent sites.

In our analysis we detected and quantified phosphorylation of eleven of the 18 PTP1b substrate proteins and seven of the 25 PTP1b-substrate phosphosites verified in (Ren et al., 2011). These sets of phosphorylation sites were tested for PTP1b-dependence, but neither set was significantly enriched among PTP1b-dependent phosphosites. The lack of PTP1b-dependent phosphorylation on some PTP1b substrates suggests compensatory modes of regulation (e.g. through other PTPs and kinases) and that the PTP1b-dependent phenotypes are due to phosphorylation changes on additional proteins and pathways. This complexity makes it challenging to identify direct substrates of PTP1b, but the network-level phosphosite measurements provide direct insight into the particular proteins and pathways that are modulated by chronic PTP1b deletion or HFD *in vivo*.

Our unbiased, network-level approach identified several novel sets of PTP1b-dependent phosphosites, including those on proteins involved in metabolic processes, oxidation-reduction (REDOX) and lipid metabolism. Given that beneficial lipid metabolic changes had been previously reported in L-PTP1b^{-/-} mice on HFD (Delibegovic et al., 2009), we were particularly interested in the enrichment for PTP1b-dependent phosphosites on lipid metabolic proteins

(highlighted in Figure 2.2B). These proteins span a diverse set of lipid pathways, including fatty acid metabolism and transport (acyl-coA binding protein (ACBP), Fatty Acid Synthase (FAS), ATP citrate lyase (ACLY), cytochrome b5 (CYB5A), and cytochrome P450 2e1 (CYP2E1)), triglyceride synthesis, storage and transport (long-chain fatty acid-CoA ligase 5 (ACSL5), high-density-lipoprotein binding protein (vigilin)) and phospholipid metabolism and transport (peroxiredoxin 6 (PRDX6), staphylococcal nuclease and tudor domain containing 1 (SND1), S-adenosyl homocysteine hydrolase (SAHH)). Independently, phosphosites limited to proteins involved in fatty acid metabolism only are also significantly PTP1b-dependent (FDR<0.05), highlighting potential PTP1b-dependent regulation of fatty acids in particular. Some of these phosphorylation sites have been identified in previous large-scale studies, but their functions have not been characterized. Determining the role of PTP1b deletion in lipid and fatty-acid metabolism will require functional analysis of these phosphorylation sites.

The enrichment of PTP1b-dependent phosphosites on proteins involved in REDOX is intriguing as well, given that oxidative stress plays an important role in obesity-induced inflammation and ER stress (Zhang and Kaufman, 2008). These phosphorylation sites are on proteins typically thought of as metabolic enzymes rather than signaling molecules (e.g., 4-hydroxyphenylpyruvic acid dioxygenase (HPD) Y221, aldehyde dehydrogenase 1a7 (ALDH1A7) Y484), and the functions of these sites as well as the contributions of these proteins to inflammation and ER stress are largely uncharacterized. Although numerous and diverse metabolic enzymes are expressed in liver tissue, enrichment for metabolic proteins among PTP1b-dependent phosphosites was unexpected. In addition to REDOX and lipid metabolic proteins, other metabolic processes that show PTP1b-dependent phosphorylation (FDR < .05) include arginine and proline metabolism, phenylalanine metabolism, and glycolysis/gluconeogenesis. The roles

that perturbation of PTP1b plays in these pathways, as well as the functions of these phosphosites, are unexplored. However, recent work demonstrates that site-specific tyrosine phosphorylation of other metabolic enzymes critically modulates enzyme activity, affecting metabolism, REDOX environment, and growth at the cellular level (Fan et al., 2011; Hitosugi et al., 2009).

In contrast to PTP1b-dependent sites, phosphosite set enrichment analysis of diet-dependent sites yielded only enrichment for mitochondrial proteins (FDR < .01, Table 2.3, highlighted in Supplementary Figure 2.1) (note that the lack of detection reflects MS experimental design rather than a lack of pathway-specific HFD-mediated phosphotyrosine changes (discussed in Methods)). Tyrosine phosphorylation of mitochondrial proteins has been implicated in the context of cancer, as oncogenic RTKs and Src family kinases have been shown to translocate to the mitochondria and directly phosphorylate and modulate mitochondrial metabolic enzymes (Feng et al., 2010; Hitosugi et al., 2011). Interestingly, here we have determined that HFD also increases tyrosine phosphorylation on mitochondrial proteins, revealing a potential link between cancer and obesity. Further investigation of these phosphorylation sites might reveal mechanisms by which obesity promotes cancer metabolism.

PTP1b-deficient livers have an overabundance of free PUFA

Although lipid and fatty acid metabolic pathways show significant PTP1b-dependent phosphotyrosine regulation, the effects of these phosphorylation events have not been characterized. To gain insight into the potential roles of these sites and determine whether altered phosphorylation impacts hepatic lipid composition, we performed quantitative lipidomic analyses of the liver samples. Given the diversity of lipid metabolic proteins with PTP1b-dependent phosphorylation, we applied a global LC-MS-based lipidomics strategy that enables

de novo discovery of differentially abundant lipids among samples (Figure 2.3) (Homan et al., 2011). Through this approach, we were able to monitor quantitative changes in lipids between L-PTP1b^{-/-} and control mice. On HFD, the most significant PTP1b-dependent lipid metabolic changes involved sets of polyunsaturated fatty acids (PUFA) in the free-fatty-acid (free-FA) pool. To examine these changes more closely, a calibration curve, composed of isotopically labeled FA standards, was used to more accurately quantify FAs detectable by our method (28 structural isomers, ranging from C16 to C24 acyl-chain lengths and spanning four orders of magnitude in abundance (Table 2.4)). Interestingly, PTP1b deletion altered the FA pool composition rather than pool size, as total FAs were similar between L-PTP1b^{-/-} and control (Table 2.4). To visualize these results, we plotted each FA on a volcano plot as a function of compositional fold-change (L-PTP1b^{-/-} relative to control) and corresponding statistical significance (Figure 2.4A). Many of the unsaturated FAs had ion chromatogram elution profiles with multiple peaks (Figure 2.3B and Supplementary Figure 2.2); each peak corresponds to a particular isomer whose elution time is dependent on the position of the final (ω) double bond in the acyl chain. Double bond location is a critical determinant of physiological function. For example, C20:3 ω 6 is the anti-inflammatory lipid dihomo- γ -linolenic acid, while C20:3 ω 3 is a precursor of the anti-lipogenic PUFAs eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). The potential physiological roles of C20:3 ω 7 and ω 9, on the other hand, are less-well studied. As can be seen in the upper-right-hand corner of Figure 2.4A, several PUFAs were significantly increased in the L-PTP1b^{-/-} livers relative to control, including C18:3, C20:3, C22:3. Particular isomer peaks for each of these PUFAs were PTP1b-dependent, and thus determination of double bond location was necessary to understand the physiological implications of these changes.

To determine the identities of the PTP1b-dependent C18:3, C20:3 and C22:3 isomer peaks, co-injection of commercially available isomer standards was performed (Supplementary Figure 2.2 and Methods). Because standards were not available for confirmation of all assignments, uncertainty in ω -bond assignment is denoted with an asterisk. The most significantly PTP1b-dependent PUFA (C18:3, C20:3, and C22:3) species were determined to be ω 9 or a combination of ω 9 and ω 7 species and, in addition to total C20:2 (for reasons described below), are referred to as “ ω 7+ ω 9 PTP1b PUFA” (denoted in red, Figure 2.4A). The other PTP1b-dependent PUFA species (C24:6, C24:5, C24:4, C20:3, C22:3, colored in green in Figure 2.4A) were determined to be ω 3 and/or ω 6 and are subsequently referred to as “ ω 3+ ω 6 PTP1b PUFA”. Although the physiological roles of these isomers are not well-characterized, increasing levels of selected ω 3+ ω 6 or ω 7+ ω 9 PUFAs in the L-PTP1b^{-/-} mice might have pleiotropic effects, as discussed below.

To gain insight into the unknown physiological functions of the PTP1b-dependent PUFA isomers, correlation-based clustering was applied to the percent fatty acid compositions of HFD and NC livers. PTP1b-dependent PUFA clustered with several better-characterized PUFAs (Figure 2.4B). All ω 3+ ω 6 PTP1b PUFA, including the poorly characterized C24 PUFA series, coherently cluster with ω 3 and ω 6 PUFAs whose physiological effects have been characterized, including anti-lipogenic/anti-inflammatory DHA (C22:6 ω 3) and EPA (C20:5 ω 3), for which the ω 3 PTP1b-dependent PUFAs (C18:3, C20:3, and C24:6) would all be precursors. As might be expected, this cluster of PUFAs, several of which are associated with physiological benefit, is significantly depleted in the HFD mouse livers relative to NC (Table 2.4). The high level of correlation between the fatty acids in this cluster suggests a positive metabolic function for the ω 3+ ω 6 PTP1b PUFAs.

In contrast, the C20:3 ω 9/ ω 7* isomers are barely detectable under NC conditions, while C18:3 ω 7* and C22:3 ω 9*/ ω 7* isomers are not detected at all (Table 2.4), suggesting that these species are byproducts of HFD metabolism. C20:3 ω 9/ ω 7* clusters with C20:2 and monounsaturated fatty acids (MUFAs): C16:1 ω 7, C18:1 ω 9 and C20:1 (Figure 2.4B). Given that ω 7/ ω 9 MUFA are products of the Δ 9 desaturase, stearoyl-CoA desaturase (SCD), we quantified the SCD Index, which is often used as a surrogate for SCD activity (Stefan et al., 2008). The C18 SCD Index, a ratio of SCD's major product (oleic acid) to substrate (stearic acid), was quantified for each mouse under both basal and insulin-stimulation conditions. In agreement with the increased abundance of the ω 7/ ω 9 PTP1b PUFAs, the SCD Index is elevated in both NC and HFD L-PTP1b^{-/-} mice relative to control under basal conditions (Figure 2.4C). However, the trend changes upon insulin stimulation. Although insulin stimulation does not appear to affect the SCD index for control mice, the SCD index of the PTP1b-deletion livers is dramatically reduced upon insulin stimulation ($P < .05$, combined p-value for insulin-dependence in HFD and NC L-PTP1b^{-/-} livers). This result was initially surprising, as insulin should promote lipogenesis and SCD activity, and PTP1b deletion enhances insulin sensitivity. However, the dose of insulin-stimulation was hypoglycemia-inducing (10mU/g, 10min.). Given that the L-PTP1b^{-/-} livers are more sensitive to insulin, it is likely that they became hypoglycemic more quickly, leading to rapid suppression of SCD activity via hypoglycemic stress response. In support of insulin-induced hypoglycemic stress, the ratio of pT172 AMPK to total AMPK is dramatically elevated in the insulin-stimulated livers (Figure 2.4D). Thus L-PTP1b^{-/-} mice might have increased hepatic SCD activity under physiological conditions.

PTP1b-deficient livers have altered triglyceride metabolism on HFD

MS-based lipidomics analysis revealed that L-PTP1b^{-/-} mice have an approximately 2-fold increase in total hepatic triglycerides relative to control mice following HFD (Figure 2.5A). This trend was initially uncovered in a preliminary MS-lipidomics analysis of a subset of thirteen HFD liver tissue samples, verified in quantitative MS-lipidomics analyses (with triglyceride standard) and further validated with a more conventional, enzymatic assay, which correlated well with MS results ($\rho = .96$). These results were unexpected, given that we previously found a decrease in L-PTP1b^{-/-} hepatic triglycerides on HFD at an early time point (8 weeks of age) and no difference at 21 weeks, a time point more similar to this study (19.5 weeks)(Delibegovic et al., 2009). Given this discrepancy and to verify that the trends were accurate, the quantification of liver triglycerides was extended to include all HFD liver tissues in the study (12 L-PTP1b^{-/-}, 15 control) as well as a second cohort of HFD mice (4 L-PTP1b^{-/-}, 4 control). These additional samples confirmed that hepatic triglycerides were elevated in L-PTP1b^{-/-} mice (combined data shown in Figure 2.5A). These data are even more striking when liver triglyceride content is plotted as a function of mouse weight. For weight-matched pairs of mice, L-PTP1b^{-/-} mice have dramatically more hepatic triglyceride content (Figure 2.5B) with surprisingly little correlation between total hepatic triglycerides and serum triglycerides measurements (Figure 2.5C). This disparity between hepatic and serum triglycerides is consistent with the later time point in the previous study (Delibegovic et al., 2009), where L-PTP1b^{-/-} mice had significantly lower serum triglycerides but similar hepatic triglycerides relative to control. Across studies, the lack of correlation between hepatic and serum triglycerides suggests a PTP1b-dependent defect in hepatic triglyceride secretion that increases with age. Because liver steatosis is often implicated in compromised liver function and inflammation, we measured molecular markers of

inflammation: IL-6 and IL-1 β , the marker of macrophage infiltration MCP-1, and phosphorylation of the stress kinase p38 (Supplementary Figure 2.3). Despite elevated steatosis, we did not detect increased inflammation in the L-PTP1b^{-/-} mice, which is consistent with the muted ER stress and inflammation responses observed in these mice previously (Agouni et al., 2011; Delibegovic et al., 2009).

In addition to total triglycerides, MS analysis provided quantification for 28 triglyceride structural isomers (Table 2.5). Each structural isomer is characterized by the sum of carbons and double bonds in the acyl chains. Detected triglycerides ranged from 48 to 56 total acyl carbon length and contained a total of one to ten desaturations. Statistically significant PTP1b-dependent changes in triglyceride composition were visualized in a volcano plot (Figure 2.5D). As a fraction of the total pool, triglycerides with many sites of desaturation, or “PUFA-rich” triglycerides, are specifically reduced in L-PTP1b^{-/-} livers relative to control. In fact, the number of double bonds in the triglycerides correlates with the compositional depletion of that triglyceride in the L-PTP1b^{-/-} livers ($P < 10^{-5}$).

Longer chain, PUFA-rich triglycerides constitute a larger percentage of triglyceride composition in NC relative to HFD mice, regardless of PTP1b deletion. As highlighted by the heatmap of clustered triglyceride compositions (Figure 2.5E), the distinction between NC abundance and HFD depletion of these metabolites is much more pronounced than the compositional difference between HFD L-PTP1b^{-/-} and control. On HFD, esterification of diet-derived FAs into triglycerides may limit cytotoxic effects. The resulting build-up of shorter-chain triglycerides with fewer sites of desaturation would explain the lower fraction of longer-chain-PUFA-rich triglycerides in HFD versus NC mice.

While the physiological significance of the L-PTP1b^{-/-} liver triglyceride compositions on HFD is unknown, serum depletion of PUFA-rich triglyceride species is positively predictive of T2DM development in humans (Rhee et al., 2011). To determine whether the compositional depletion of PUFA-rich triglycerides in the L-PTP1b livers was associated with similar serum compositional changes, we assayed serum triglycerides by MS (Table 2.6, Supplementary Figure 2.4) and detected a parallel compositional depletion in PUFA-rich triglycerides ($P < 10^{-11}$, correlation between triglyceride double bond number and depletion in L-PTP1b^{-/-} mice).

Although total liver and serum triglyceride measurements suggest a triglyceride secretion defect, the alteration in triglyceride compositions (depletion of PUFA-rich and elevation of MUFA-rich triglycerides) suggests that elevated hepatic lipogenesis might also contribute to liver steatosis in the L-PTP1b^{-/-} mice on HFD. For this reason, we measured the ratio of pS79 acetyl-CoA carboxylates (ACC) to total ACC, a marker of lipogenesis inhibition (Supplementary Figure 2.5). Although total pS79 ACC was significantly increased in L-PTP1b^{-/-}, the ratio of pS79 ACC to ACC was not significantly elevated. This result is consistent with hepatic lipogenic mRNA expression measurements in the previous study (Delibegovic et al., 2009). Thus, the compositional triglyceride changes in L-PTP1b^{-/-} mice might reflect an increased capacity to store diet-derived FAs as neutral triglycerides. Altered acyl chain compositions in other complex lipid classes were not detected.

Modeling the relationships between protein phosphorylation and lipid metabolites

To gain insight into the phosphorylation sites that might potentiate PTP1b-dependent phenotypes, computational models of the phosphorylation and lipidomic datasets were constructed. Given the size of our study and the biologically uncharacterized nature of the measurements in the phosphotyrosine and lipidomics datasets, we built multivariate regression

models that leverage mouse-specific variation and are predictive of the novel PTP1b-dependent phenotypes. We specifically selected multivariate regression, because it is capable of predictively modeling phenotypes as the result of multiple underlying factors, and, in this way, comes closer to capturing the complex, multifactorial nature of biological systems.

As PTP1b-dependent phosphotyrosine network alterations were likely drivers of the observed lipid metabolic changes, our model-building efforts focused on predicting lipid metabolic phenotypes as functions of the 228 tyrosine phosphorylation site measurements in the HFD basal dataset. Final, independent models of steatosis, $\omega 7+\omega 9$ PTP1b PUFA, and the $\omega 3+\omega 6$ PTP1b PUFA were built on reduced sets of molecular determinants (ten or fewer). Model cross-validation indicated that these models had good predictive power ($Q^2 = .54, .66, .52$, respectively) (Figure 2.6B). In Figure 2.6A, the reduced sets of molecular determinants for models are displayed graphically for each phenotype.

The $\omega 7/\omega 9$ PUFA clusters and $\omega 3/\omega 6$ PUFA clusters have several predictive phosphorylation sites in common. These lipids have significant overlap in biosynthetic and metabolic pathways; thus, common sites are likely to contribute to processes governing both sets of lipids. CYP2E1 Y426, ACLY Y672, enolase 1 (ENO1) Y25, glutamine synthase (GLUL) Y336, and keratin 8 (K8) Y210 strongly contribute to both models. Although all of these sites have unknown function, we have attempted to infer functionality from the characterized role of the protein. For instance, CYP2E1 has stereospecific hydroxylase and epoxidase activity for PUFAs. CYP2E1 Y426 is hyperphosphorylated when PTP1b is deleted, and, were this site inhibitory, decreased CYP2E1 PUFA metabolism might contribute to the increased levels of both $\omega 3/\omega 6$ and $\omega 7/\omega 9$ PUFA. ACLY converts cytosolic citrate to acetyl-CoA and represents a potential control point in lipid metabolism. ACLY protein and activity were elevated in livers of *db/db* mice, and

shRNA depletion of ACLY dramatically improved steatosis through repression of PPAR γ and many other lipogenic genes (Wang et al., 2009). Given that the expression of lipogenic genes was decreased in two previous studies of L-PTP1b $^{-/-}$ mice (Agouni et al., 2011; Delibegovic et al., 2009) and that ACLY pY782 is hyperphosphorylated in these livers, this site could be inhibitory. However, ACLY can directly affect the levels of acetyl-CoA and protein acetylation (Wellen et al., 2009), and therefore we cannot rule out a pleiotropic effect, where altered acetylation of multiple lipid metabolic enzymes might influence both ω 3/ ω 6 and ω 7/ ω 9 PUFA.

Additional phosphorylation sites in the ω 7/ ω 9 PUFA model are implicated in lipid metabolism. For example, CYB5 is an electron-transporting heme protein of the ER that co-localizes with Δ 5, Δ 6, and Δ 9 desaturases (Koltun et al., 2009; Schenkman and Jansson, 2003). These enzymes, in coordination with elongases, are required for the biosynthesis of long-chain PUFA (Guillou et al., 2004; Guillou et al., 2010). CYB5 is necessary for SCD function and it promotes Δ 6 (FADS2) desaturase activity (Guillou et al., 2004). Because both SCD index and Cyb5 Y11 phosphorylation are elevated in L-PTP1b $^{-/-}$ livers, increased phosphorylation of Cyb5 Y11 may potentiate SCD activity. At this point, we cannot distinguish between direct effects on enzymatic activity of CYB5 versus altered protein-protein interactions regulating the activity of the desaturase complexes. Additionally, CYB5 contributes to fatty acid metabolism via interaction with cytochrome P450s, including CYP2E1, described above. It is possible that the altered phosphorylation of CYB5 Y11 modifies multiple aspects of lipid metabolism simultaneously.

As points of comparison, models of steatosis as functions of metabolites were also built. Two models were built for steatosis as a function of (1) percent fatty acid composition and (2) percent triglyceride composition ($Q^2 = .37, .82$, respectively). The predictors from the fatty acid model of steatosis are not limited to one class of fatty acids (i.e., ω 7/ ω 9 or ω 3/ ω 6), which suggests that

both classes of molecules are important for prediction. The steatosis model as a function of triglyceride composition actually has the greatest predictive power relative to the other models. The most predictive triglycerides were the most long-chain and desaturated PUFA-rich structural isomers detected: C56:9, C56:8, and C54:7. This supports a connection between PTP1b-dependent steatosis and depletion of long-chain PUFA triglycerides (Figure 2.6).

Although the multivariate models of ω 7/ ω 9 PUFA, ω 3/ ω 6 PUFA, and steatosis were built on a reduced subset of most predictive phosphorylation sites, all phosphorylation sites were ranked according to predictive power in the model-building process. To globally discover protein pathways whose tyrosine phosphorylation patterns were predictive of phenotypes, we applied phosphosite enrichment analysis to these ranked lists of phosphosites (Tables 2.7-2.9). From this analysis, REDOX proteins were significantly enriched at an FDR of .01 for all three models. The balance of oxidation and reduction is a pivotal driver of metabolic state; desaturation of fatty acids and CYP2E1 oxidation of PUFA, for example, depend on conversion of NADH to NAD⁺. Oxidation also modulates liver triglyceride secretion and hepatic steatosis (Pan et al., 2004). Therefore, net changes in oxidation and reduction mediated by tyrosine phosphorylation of REDOX enzymes could play a significant role in lipid metabolism.

Discussion

PTP1b inhibition continues to be a focus of drug development efforts to therapeutically manage obesity-related diseases, and an anti-sense oligonucleotide inhibitor of PTP1b has advanced to phase II clinical trials (Haque et al., 2011). In this study, for the first time, the effect of PTP1b inhibition on liver tissue has been characterized at the molecular network level in a broad and unbiased manner. Using MS-based technology, we profiled changes in protein tyrosine

phosphorylation *in vivo*, across 50 mouse livers under different conditions of diet and genotype. The scale of this study was crucial to developing the statistical power required to gain confidence in phosphorylation trends uncovered, detect PTP1b-dependent pathways, and to construct multivariate models of lipid metabolic phenotypes as functions of the underlying molecular network.

Our results establish new roles for PTP1b deletion in modulating hepatic lipid metabolism. Multiple phosphorylation sites on proteins involved in lipid metabolism were found to be significantly affected by PTP1b deletion, and comprehensive lipidomics analysis demonstrated that L-PTP1b^{-/-} mice have several metabolic abnormalities. For instance, although total levels of fatty acids are similar between L-PTP1b^{-/-} and control mice on HFD, L-PTP1b^{-/-} have an overabundance of longer-chain PUFAs, including ω 7/ ω 9 PUFAs and select ω 3/ ω 6 species. Unfortunately, the physiological roles of most of the PTP1b-dependent PUFAs are not well characterized. However, when HFD and NC fatty acid measurements are clustered, PTP1b-dependent ω 3/ ω 6 species cluster with DHA and EPA, suggesting anti-lipogenic and/or anti-inflammatory roles for these species.

Multivariate regression models suggest multiple mechanisms by which altered tyrosine phosphorylation levels might lead to altered PUFA profiles in the L-PTP1b^{-/-} mice (Figures 2.6A and 2.7). Foremost is the hypothesis that PTP1b deletion leads to a general increase in desaturase activity, potentially via CYB5 Y11 and interaction with the FADS1, FADS2 and SCD. As discussed, SCD is dependent on CYB5 for its activity, and, indeed, the L-PTP1b^{-/-} have elevated C18 SCD indices under NC and HFD basal conditions. Given the role of SCD in triglyceride synthesis (Man et al., 2006), an increase in SCD activity could also contribute to the compositional build-up of MUFA-rich triglycerides in L-PTP1b^{-/-} mice. Elevated SCD activity

would also limit the cytotoxicity of diet-derived saturated fatty acids (SFAs), as MUFAs are typically less toxic than SFAs and triglycerides are inert (Coleman and Lee, 2004; Liu et al., 2011) (Figure 2.7). Other PTP1b-dependent phosphosites might contribute to an excess of PUFA via defects in pathways leading to complex lipid formation. ACBP, which binds acyl-CoA esters (C14-C22) with high specificity and affinity ($K_d < .05 \text{ nM}$) (Oikari et al., 2008), is hyperphosphorylated on pY29 upon PTP1b deletion (Figure 2.1B). The crystal structure for ACBP is available, and Y29 stabilizes the 3' ribose phosphate of the acyl-CoA (Taskinen et al., 2007). Phosphorylation of Y29 would electrostatically repel the 3' ribose phosphate, potentially destabilizing the binding of acyl-CoAs to pY29-ACBP. If pY29-ACBP has compromised acyl-CoA binding, free PUFA in PTP1b-deletion livers might abound because (1) transport of acyl-CoAs to cellular sites of complex lipid synthesis is impaired and (2) untransported acyl-CoAs would cause a build-up of PUFA due to an inhibition of their activation to acyl-CoAs by long-chain acyl-CoA synthetases. This mechanism would also be consistent with the compositional depletion of long-chain PUFA-rich triglycerides in L-PTP1b^{-/-} livers. PTP1b-dependent phosphorylation of PRDX6 (Y89), a bifunctional protein with glutathione peroxidase and phospholipase A₂ activities (Manevich et al., 2007), was also detected. A pY89-dependent increase in phospholipase activity could also contribute to the PTP1b-dependent PUFA accumulation. Given the diverse roles of proteins with altered tyrosine phosphorylation, it is likely that altered PUFA composition in the L-PTP1b^{-/-} livers is due to multiple mechanisms, including an increase in desaturase activity (CYB5 Y11), decreased PUFA metabolism (CYP2E1 Y426, discussed in results), increased lipase activity (PRDX6 Y89) and modulation of pathways leading to complex lipid synthesis (ACBP Y29, ACSL5 Y69 and/or SND1 Y908) (Figure 2.7).

L-PTP1b ^{-/-} livers are ~2-fold more steatotic than controls on HFD. Elevation of liver triglycerides did not correlate with markers of lipogenesis or inflammation, suggesting that the steatosis was not associated with compromised liver function. Comparison of total liver and serum triglycerides in L-PTP1b^{-/-} mice indicates defective triglyceride secretion, which could contribute to steatosis (Figure 2.5C). In addition, MS-based analysis of serum and liver triglyceride compositions in conjunction with measurement of lipogenic markers (Figure 2.5D, Supplementary Figures 2.3, 2.4) suggests that increased formation of “MUFA-rich triglycerides” from diet-derived FAs may also promote hepatic steatosis (discussed in Results). This could potentially be mediated by increased SCD activity and funneling of MUFA products to DGAT2, as DGAT2 targets triglycerides to cytosolic storage rather than VLDL secretion (Yamazaki et al., 2005).

Elevated PUFAs could contribute to defective triglyceride secretion in L-PTP1b^{-/-} mice. ω 3 PUFA not only decrease hepatic triglyceride output via suppression of lipogenesis but also independently inhibit triglyceride secretion (Ginsberg and Fisher, 2009). DHA inhibition of triglyceride secretion was shown to be mediated by peroxidation of lipids followed by the degradation of apolipoprotein B (apoB), the central component of VLDL particles, which was specifically sensitive to the PUFA-mediated oxidative damage (Pan et al., 2004). In this same study, several other dietary PUFAs, including linoleic acid (C18:3 ω 3) and arachidonic acid (C20:4 ω 6), were shown to promote apoB degradation via peroxidation. Lipid peroxide formation was positively correlated with polyunsaturation of fatty acids, independent of the double bond position. In the current study, we suspect that PTP1b-dependent increase in PUFA might inhibit VLDL secretion by lipid peroxidation and oxidative degradation of ApoB,

potentially through altered activity of CYP2E1 and other proteins whose phosphorylation patterns were predictive in the steatosis model (Figure 2.7, Table 9).

Our study also demonstrates the utility of computational approaches to gain biological insight from complex multivariate datasets. First, we adapted gene set enrichment analysis for the discovery of PTP1b-dependent pathways from the phosphotyrosine measurements. Using this approach, we recovered known PTP1b-dependent pathways (insulin signaling) and discovered novel PTP1b-dependent pathways: redox homeostasis, amino acid metabolism, and lipid metabolism (particular fatty acid metabolism). In addition, we developed multivariate regression models that predict lipid metabolic phenotypes as functions of underlying tyrosine phosphorylation sites, providing biological insight in a quantitative and unbiased way.

Although our discussion has mainly focused on PTP1b-dependent phosphorylation of proteins involved in lipid metabolism, PTP1b-dependent phosphosites are present on a much larger set of proteins. There is much to be learned about the roles that these phosphorylation sites might play in the context of metabolic syndrome. It is our hope that this study will stimulate further research of the diverse pathways and surprising phenotypes modulated by PTP1b inhibition.

Methods

Animal studies

PTP1b-liver-specific-deletion mice were generated by crossing PTP1b-floxed (PTP1b-fl/fl) mice with mice expressing Cre recombinase under the control of the albumin promoter (Alb-Cre), resulting in Alb-Cre PTP1b-fl/fl mice that specifically lack hepatic PTP1b expression (Delibegovic et al., 2009). Alb-Cre mice were used as controls. Genotyping for the PTP1b floxed allele and the presence of Cre was performed by PCR. Hepatic PTP1b protein expression

was accessed at the end of the study, by quantitative immunoblotting (PTP1b antibody AF13661, R&D Systems). Mice were maintained on a 12-hour light/dark cycle in a temperature-controlled barrier facility with water and food freely accessible. All mice were maintained on a normal chow (NC) diet (Prolab Isopro RMH 3000, 15% calories from fat) until twelve weeks of age, at which point mice either continued on NC or were placed on high fat diet (Teklad TD.93075, 55% calories from fat). Mouse procedures were performed in accordance with the guidelines and approval of the Institutional Animal Care and Use Committee at the University of Massachusetts Medical School and the Ontario Cancer Institute.

Serum measurements

At 5.5 and 6.5 weeks of NC or HFD, fed and fasted serum samples, respectively, were collected from the tail veins of mice. Adiponectin, glucagon, cholesterol (total, HDL, and LDL) were measured in the fasting serum samples. Insulin and glucose were measured in both fasted and fed serum samples. Serum hormones (insulin, glucagon, leptin, adiponectin) were measured using reagents purchased from Millipore and run on a Bio-Plex 200 System; hormone levels were determined by a logistic 5pl standard curve fit. Lipid analysis was performed on a Roche Cobas c111 chemistry analyzer. Glucose was measured using an Analox glucometer.

Phosphotyrosine analysis

Harvesting of liver tissue. At 19.5 weeks of age, mice were fasted overnight and their tissues harvested. Based on preliminary optimization of insulin stimulation conditions, 10mU/g insulin (Novo Nordisk Novolin R) was i.p.-injected 10min. prior to tissue harvesting. All mice received i.p.-injection of 150mg/kg pentobarbital sodium (Nembutal Lundbeck, Inc) 2min. prior to tissue harvesting. While mice were unconscious, the abdominal cavity was opened and livers were

snap-frozen and extracted using liquid-nitrogen-chilled freeze clamps. Tissues were stored for subsequent biochemical analyses at <-70C.

Peptide sample processing. In a 4C room, aliquots of frozen liver tissues, at an estimated mass ~100mg, were placed in 5mL round-bottom polypropylene tubes containing 3mL ice-cold 8M urea and immediately homogenized using a Polytron homogenizer. Protein concentration was estimated by bicinchoninic acid (BCA) assay (Pierce). Sample proteins were reduced (1hr, 10mM, DTT, RT), alkylated (1hr, 55mM iodacetamide, 56C), and digested (1:50 protein to modified trypsin (Promega) ratio, in ammonium acetate buffer pH 8.9 added to original tissue lysate such that urea concentration in final solution was reduced to 800mM; reaction ran on a rotor for 20 hrs. at RT and was stopped by addition of 1mL acetic acid (99.99%). The digests were centrifuged, and supernatants were desalted and fractionated on C18 Sep-Pak Plus cartridges (Waters). Peptides that eluted with 25% acetonitrile in .1% acetic acid were lyophilized. Peptide sample aliquots (corresponding to 800ug of starting protein) were labeled with 8-plex iTRAQ (2 aliquots of label per peptide sample). Sets of eight distinctly labeled peptide samples were combined. For both insulin-stimulated and basal conditions, one liver sample was selected as a control and included in all insulin-stimulated or basal experiments to enable quantification across experiments.

Enrichment of phosphotyrosine peptides. A phosphotyrosine peptide IP was performed, as described in (Zhang et al., 2005), with slight modification to the anti-phosphotyrosine antibodies used. Here, antibodies included 12ug 4G10 (Millipore), 12ug PY100 (CST), and 12ug PT66 (Sigma), with 60uL protein G agarose beads (Calbiochem). Peptides were eluted from the antibodies with 70uL 100mM glycine, pH 2.1. To address non-specific binding from the phosphotyrosine peptide IP, an immobilized metal affinity chromatography step (IMAC) was

used for further enrichment of phosphopeptides (Zhang et al., 2005). Phosphopeptides were eluted from the IMAC column to a capillary precolumn.

LC-MS/MS. The capillary precolumn (100µm i.d., packed with 10cm of ODS-A, 12nm, S-10µm beads (YMC)) was connected to a capillary analytical column (50µm i.d., packed with 10cm of ODS-AQ, 12nm, S-5 µm beads (YMC), with a laser-pulled (Model P-200; Sutter Instrument) electrospray ionization emitter tip (<1µm diameter)) (Martin et al., 2000). Peptides were eluted (flow rate ~ 20nL/min) from the liquid chromatography column to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) with the following gradient: 0min: 0% B; 10min: 13% B; 105min: 42% B; 115min: 60% B; 122min: 100% B (solvent A = .2 mM acetic acid and solvent B = 70% acetonitrile, 2mM acetic acid). Data were collected using a data-dependent acquisition mode: a high resolution MS1 scan on the Orbitrap was followed by up to 10 pairs of data-dependent MS/MS scans of the most abundant MS1 precursor ions. Each MS/MS pair consisted of one MS/MS scan on the LTQ (isolation width 3m/z, CID fragmentation, 35% collision energy) and a second MS/MS scan on the Orbitrap (isolation width 3m/z, HCD fragmentation, 75% collision energy). The Orbitrap MS/MS provided high-resolution iTRAQ quantification, while the LTQ MS/MS scan provided better sequence information. Dynamic exclusion for the data-dependent scans was set to 120sec.

MS Data processing. Xcaliber (.raw) data files were converted to MASCOT generic format (.mgf) with DTASuperCharge (version 1.19) and searched with MASCOT v2.1 (Matrix Science) against the National Center for Biotechnology Information (NCBI) mouse proteome with peptide tolerance of 10ppm, MS/MS tolerance of .8 Dalton, 1 missed cleavage, fixed modifications (methyl-cysteine and 8-plex iTRAQ), and variable modifications (methionine oxidation, tyrosine phosphorylation, and serine/threonine phosphorylation). Mascot peptide identifications,

phosphorylation site assignments and quantification were verified manually. Phosphorylation site assignments were made using PTMScout (Naegle et al., 2010).

Relative quantification of phosphotyrosine peptides. Given the high resolution of the Orbitrap, the peak heights of iTRAQ reporter ions were used to quantify the relative amounts of phosphopeptides across an 8-plex experiment. Peak heights in a spectrum were normalized to mean iTRAQ peak intensity. If multiple, validated spectra existed for the same phosphosite, these spectra were averaged. For normalization of phosphorylation measurements, 1/500th of the supernatant from the phosphotyrosine-peptide IP was also analyzed by LC-MS/MS. This provided quantitative information for the most abundant (unphosphorylated) peptides in the sample, which were assumed, on average, to be constant across all samples (with the exception of blood proteins like hemoglobin and blood serum albumin, which were excluded from the analysis). The supernatant iTRAQ peak heights were normalized relative to the mean iTRAQ peak height for each spectrum, resulting in ~1000, 8-dimensional peptide ratios. The median of the 90% least-outlying peptide ratios (outlying by multivariate T^2 statistic) was used as a normalization factor for the phosphopeptide quantitation. After relative quantification was achieved within each individual phosphotyrosine analysis, relative quantification across analyses was accomplished by normalizing each individual analysis to the control livers. Associated error was determined by calculating the relative standard deviation for samples run multiple times but in separate phosphotyrosine peptide quantification experiments. The mean absolute and relative standard deviations were .10 +/- .10 and 10% +/- 10%, respectively.

Principle component analysis and variance explained by diet and genotype

Principle component analysis (Wold et al., 1987) was used to decompose the insulin-stimulated (insulin) and basal phosphotyrosine data matrices independently. The principle components

define a linearly independent set of axes composed of linear combinations of the phosphosite profiles and onto which the mouse liver tissue samples can be plotted. The first principle component captures the most important data variance trends, and each subsequent principle component captures the next most important variance trend that was not described by the previous principle component(s), until all data variance is explained. The contributions of all sites to each of the principle components is preserved in the loadings matrix, P , and the position of each of the liver tissue samples in principle component space is contained in the scores matrix, T . The relationship among P , T , and the corresponding data matrix, X , is:

$$X = TP^T,$$

where the dimensions of X , T , and P are $[N \times M]$, $[N \times p]$, and $[M \times p]$ and N , M , p are the number of liver tissue samples, number of phosphosite measurements, and the number of principle components used, respectively. (P^T is the transpose of P .)

This analysis requires that the data matrix be complete, and, given the missing data points in the phosphotyrosine datasets that result from the merging of multiple data-dependently acquired MS/MS experiments, missing data techniques were required for application of PCA. Analysis was limited to phosphorylation sites that had data for at least 70% of the liver samples in the insulin-stimulated or basal datasets. Multiple random imputation was used to deal with the remaining missing data points, as this approach makes minimal assumptions, namely, that the distribution of each missing phosphosite can be randomly approximated by the observations present for that phosphosite (Rubin, 1987). This procedure was repeated for multiple iterations (1000). A complete data matrix, $X_{C,i}$, was randomly imputed from the incomplete data matrix, X_I , and decomposed into corresponding scores and loadings matrices $T_{C,i}$ and $P_{C,i}$, ($i = 1, 2, \dots$

1000). Estimates of the scores and loadings, T_{est} and P_{est} , were calculated as the average of the $T_{C,i}$ and $P_{C,i}$, respectively. The columns of T_{est} and P_{est} were made orthogonal and orthonormal, respectively. The locations of the liver tissue samples plotted in the principle component plane (Figure 2.2) correspond to T_{est} .

The variance explained by factors (diet, genotype or both diet and genotype) was estimated as follows:

$$Var_{exp} = 1 - \frac{\sum_{i=1}^M \sum_{j=1}^N (phos_{i,j} - phos_{j,est(i)})^2}{\sum_{i=1}^M \sum_{j=1}^N (phos_{i,j} - phos_{i,mean})^2}$$

where M is the total number of phosphosites in the dataset, $phos_{ij}$ is the phosphorylation level for the i^{th} phosphosite in the j^{th} sample ($j = 1, 2, \dots, N$), $phos_{i,mean}$ is the mean for the i^{th} phosphosite, and $phos_{j,est(i)}$ is the estimate of $phos_{ij}$ resulting from regression of the i^{th} phosphosite phosphorylation levels onto a corresponding indicator vector of factors (genotype or diet). Note that complete data is not required for this calculation, as $phos_{j,est(i)}$ is calculated independently for each phosphosite, using only the data points for the i^{th} phosphosite that are present. This as well as all subsequent computational and statistical analyses were performed in Matlab Student Version 7.10.0.499 (R2010a).

Unsupervised clustering analysis

The affinity propagation (AFP) algorithm was selected for clustering analysis, because this algorithm does not require a complete data matrix as input but rather a matrix of similarities between measurements, where the similarity between all measurements need not be specified. In addition, despite the limited input requirements, this algorithm performs very well in comparison to other algorithms (Frey and Dueck, 2007). For clustering, measurements in datasets were

scaled using (1) the \log_2 fold-change relative to the mean or (2) a z-score for each measurement. (A z-score is calculated by subtracting the mean value of a particular measurement and then dividing by the corresponding standard deviation.) Both phosphotyrosine and lipidomics dataset measurements were clustered using correlation as the metric of similarity. For improved data visualization, measurements within AFP-derived clusters were re-ordered using hierarchical clustering with correlation distance.

Supervised clustering of genotype- and diet-dependent phosphorylation sites

To specifically and comprehensively visualize factor-dependent phosphorylation sites, phosphorylation sites were ranked according to factor dependence and the most highly ranked phosphosites were incorporated into heatmaps. Typically, calculating a correlation coefficient between a feature (phosphosite) and factor and then ranking the factors according to correlation value or corresponding p-value would be sufficient. However, the insulin and basal phosphorylation datasets are independent, and, in addition, factor-dependence may be context specific. For example, some phosphosites may be PTP1b-dependent only in the context of HFD. To enable detection and visualization of all of the factor-dependent features in both global and condition-specific contexts, a procedure, involving Fisher's method, was developed to cull factor-dependencies from independent datasets as well as context-specific subsets.

For detection of genotype-dependent phosphosites, we estimated a correlation coefficient and corresponding p-value for each genotype-phosphosite pair across the following, independent data subsets: HFD/basal, NC/basal, HFD/insulin, and NC/insulin. Then, p-values for each genotype-phosphosite pair were combined, using Fisher's method, to create p-values for each of the following situations: PTP1b dependence across all datasets, NC-only PTP1b dependence, HFD-only PTP1b dependence, PTP1b dependence under insulin-stimulation only, PTP1b dependence

under basal conditions only, and HFD/basal, NC/basal, HFD/insulin or NC/insulin conditions only. For example, to calculate a p-value for the “PTP1b-dependent under insulin-stimulation only” context, Fisher’s method was used to combine p-values from the HFD/insulin and NC/insulin correlation coefficients. Importantly, the direction of the correlation is taken into account when applying Fisher’s method. Then, after estimating p-values for the many possible context-specific correlations, the most significant correlative p-value for a given phosphosite-genotype pair was used to rank all phosphosite-genotype pairs.

In a similar manner, for detection and ranking of the most diet-dependent phosphosites, correlation coefficients and p-values were calculated for the following, independent data subsets: basal/PTP1b^{-/-}, basal/control, insulin/PTP1b^{-/-} and insulin/control, and overall p-values were estimated analogously to the genotype case. A raw p-value cutoff of .025 was selected for the heatmaps in Figure 2.2 and Supplementary Figure 2.1.

Lipidomics analysis

Lipid extraction. In a 4C room, aliquots of frozen liver tissues were weighed (~120mg per tissue sample) and placed on dry ice. Lipids were extracted in 6mL of a 2:1:1 solution of CHCl₃ : MeOH : H₂O. To quantify the absolute abundances of select lipids, reference standards were spiked into the chloroform phase (670nmol (HFD) or 330nmol (NC) glyceryl triheptadecanoate (Sigma), 1.5umol (HFD) or .75umol (NC) Cholesterol-25,26,26,26,27,27,27-d7 (C/D/N Isotopes), 5nmol UC¹³-oleic acid (Spectra Stable Isotopes), 500pmol UC¹³-palmitic acid (Spectra Stable Isotopes), 50pmol arachidonic acid-5,6,8,9,11,12,14,15-d8 (Cayman Chemical), and 5pmol eicosapentaenoic acid-19,19',20,20,20-d5 (Cayman Chemical)). Frozen tissues were placed into a 15-mL dounce tissue grinder and homogenized in the extraction solution on ice. The extract was transferred to a glass vial and centrifuged at 2,500G, 4C for 5min to separate the

organic and aqueous layers. Twice the organic layer was transferred to fresh glass vials by Pasteur pipette. The final glass vial was placed under a stream of nitrogen until the organic phase evaporated. Samples were stored at -80C and later dissolved in chloroform for LC-MS analysis (Vinayavekhin and Saghatelian, 2011).

LC-MS analysis. LC-MS analysis was performed using an Agilent 6220 LC-ESI-TOF instrument. For LC analysis in negative ion mode, a Gemini (Phenomenex) C18 column (5 μ m, 4.6mm \times 50mm) was used in combination with a precolumn (C18, 3.5 μ m, 2mm \times 20mm). Mobile phases A and B were 95:5 water:methanol and 60:35:5 isopropanol:methanol:water, respectively; both contained .1% ammonium hydroxide. For positive-mode LC analysis, a Luna (Phenomenex) C5 column (5 μ m, 4.6mm \times 50mm) was used with a precolumn (C4, 3.5 μ m, 2mm \times 20mm). Mobile phase A and B had the same solvent compositions as negative mode, except, A and B were both supplemented with .1% formic acid and 5mM ammonium formate instead of .1% ammonium hydroxide. For both analysis modes, a 60min gradient was used: 100% A, flow rate .1mL/min from 0-5min, a linear increase in solvent B from 20% to 100% at .4mL/min from 5-45min, isocratic 100% solvent B for 7min at .4mL/min, and equilibration with 100% solvent A at .5mL/min for 8min (Vinayavekhin and Saghatelian, 2011).

For preliminary, discovery-motivated experiments, 1/6th sample amounts were run in a 30 μ L injection volume for both modes. In quantitative, follow-up positive mode analyses focusing on triglycerides, 1/240th (HFD) or 1/80th (NC) of the sample extract was injected in 30 μ L. In quantitative, follow-up negative mode analyses focusing on fatty acids, two analyses were required because fatty acid abundances span more than four orders of magnitude. In the first analysis, to avoid saturation of the most abundant fatty acids (oleic, stearic and palmitic), sample amount was limited to 1/80th-1/240th of the extract, and a 15 μ L injection volume was used. In

addition, the LC lines were typically coated with a significant background of palmitic and stearic acid, so the lines of the LC were purged overnight with a 1% acetic acid DMSO solution to ensure that background from stearic and palmitic acids was < 20% of sample signal the next day. For the second analysis and to better separate isomers of low abundant fatty acids, 1/6th of the sample extract was injected using 5uL volume and an 80min gradient was developed: 0-5min 100% A at .1mL/min, linear increase in solvent B from 20% to 40% from 5-15min at .4mL/min, (slower) linear increase in solvent B from 40% to 80% from 15-55min at .4mL/min, linear increase in solvent B from 80% to 100% from 55-65 at .4mL/min, isocratic 100% B from 65-73min at .4mL/min and finally a 7min equilibration step with 100% A at .5mL/min.

MS analysis was performed with an electrospray ionization (ESI) source. The capillary voltage was set at 4kV and the fragmentor voltage to 100 V. The drying gas temperature was 350C at a flow rate of 10L/min. The nebulizer pressure was 45psi. Data were collected in both centroid and profile modes with a mass range of 100-1500Da.

Automated discovery of differential metabolites. XCMS (Smith et al., 2006) was used to match, quantify, and compare peaks across an initial, exploratory lipidomics analysis of six L-PTP1b-/- mice and seven control mouse livers under HFD, basal (HFDB) conditions. Agilent chromatogram data files (.d) for both positive and negative mode 60min-gradient analyses were obtained from Agilent MassHunter and converted to mzXML files by the software program Trapper. The mzXML files were analyzed by XCMS using the default parameters. The final output file contained ion m/z ratio, average retention time, integrated mass ion intensities (peak area), and a p-value corresponding to a t-test to determine whether the ion was differentially abundant between L-PTP1b-/- and control mice. Ions that XCMS indicated might be differentially abundant were then examined individually for verification by hand.

Absolute quantification of hepatic triglycerides and cholesterol. Detectable hepatic triglyceride species ranged from total acyl chain carbon length of 48 to 56. Manual integration of ion chromatograms (20 ppm m/z window) was performed with care, so that only ion chromatogram area corresponding to the monoisotopic peak of the triglyceride was included in the integration. In addition, triglyceride peak areas from sample analyses were compared to background from blanks, and several triglycerides were excluded from the dataset (i.e., C48:0, C50:0), as background was >20% of the signal. Integrated peak areas were then (1) normalized to the peak area of the glyceryl triheptadecanoate standard and (2) normalized to the starting tissue mass to obtain absolute quantification in nmol triglyceride / mg tissue. Total triglyceride amount was obtained by summing all triglyceride species quantified. Percent triglyceride composition was calculated by dividing each triglyceride structural isomer by total triglycerides. For liver tissue samples with replicate analyses, these replicates were combined using the principle component alignment method described below.

Absolute cholesterol quantification was obtained by normalizing to the cholesterol-d7 standard and starting tissue masses.

Absolute quantification of hepatic fatty acids. LC-MS data was inspected manually and ion chromatograms were integrated with 20ppm m/z tolerance, as above. As described above, two separate MS runs were required for quantification of high abundance and low abundance species, so that the former did not saturate the detector and the latter were detectable. The original plan for combining quantification of low and high abundance runs for single sample was to quantify each run independently: construct standard curves from the four fatty acid standards (spike into samples before processing and spanning four orders of magnitude), quantify the fatty acids using the run-specific standard curve, and average fatty acids that were quantified in both high-

abundance and low-abundance runs. However, for the analyses of highly abundant fatty acids, which were limited to small sample amounts, often only two standards were detected. As one might expect, the standard curves (in log-log space) were very sensitive to noise in the measurement of those standards. Similarly, the most abundant fatty acid standard often saturated in analysis of the low-abundance fatty acids, resulting in a calibration curve of three points, which was also very sensitive to measurement noise.

To overcome this issue, a new method for integrating the high and low abundance datasets was devised. Rather than limit calibration curve-fitting to the four or fewer standard fatty acids detected per run, our method incorporates all fatty acids common to both high- and low-abundance runs (> 10 fatty acid pairs) into each sample calibration. Principle component analysis is used to find one principle component (fit a line) from the paired peak areas from high and low abundance analyses. The fatty acids not common to both runs are then projected onto the first principle component, so that all of the fatty acids have positions on the principle component. The positions of the standards are used to determine the linear relationship between principle component position and absolute units of quantification (pmol). We found that this approach was more robust to noise in standard measurements, because it made use of more data points and resulted in a global fit. In addition, often there were more than two MS analyses per sample, and this method of principle component alignment was easily extended to enable global alignment of multiple runs.

After calibration to standards, fatty acid abundances were normalized to starting tissue mass to obtain absolute quantification (pmol fatty acid / mg tissue). Percent composition was obtained by dividing each fatty acid by the sum total of all fatty acid abundances.

Identification of selected fatty acid isomers by co-injection. C18:3 ω 3, C18:3 ω 6, 18:3 ω 9, C20:3 ω 3, C20:3 ω 6, C20:3 ω 9, C22:3 ω 3 standards were purchased from Cayman Chemicals, while C24:6 ω 3 was purchased from Laradan Fine Chemicals AB. The LC-MS protocol used for all co-injection experiments was the same as described for analysis of low abundance fatty acids above. For each fatty acid species, the following runs were done: each standard was run individually (10pmol), all isomers of that species were run together (10pmol each), an HFD liver tissue sample was run individually (1/6 sample), and finally 1/6 sample of HFD liver tissue was run co-injected with amounts of isomer standards titrated so that the co-elution of standard with particular endogenous peaks could be identified.

For C18:3, two sets of standard concentrations were used in the co-injection experiments to identify the 3rd isomer peak: (1) 1.25pmol C18:3 ω 3, C18:3 ω 6, and C18:3 ω 9 and (2) 2.5pmol C18:3 ω 3 and C18:3 ω 6 with 6.25pmol C18:3 ω 9 (Supplementary Figure 2.2A, red and purple lines, respectively). None of the standards co-eluted with the PTP1b-dependent isomer peak. Because (1) isomer standards elution times increased with ω -bond distance (i.e., eluted in the order ω 3, ω 6, ω 9), and (2) the endogenous peak of interest eluted in between the ω 6 and ω 9 peaks, we suspect that the endogenous peak corresponds to C18:3 ω 7. Because we did not have a standard to confirm this assignment, this peak is referred to as C18:3 ω 7* in the text.

For C20:3, two sets of standard concentrations were used in the co-injection experiments to identify the 2nd isomer peak: (1) 500fmol C20:3 ω 3 and C20:3 ω 6 with 3.4pmol C20:3 ω 9 and (2) 1pmol C20:3 ω 3 and C20:3 ω 6 with 3.4pmol C20:3 ω 9 (Supplementary Figure 2.2B). C20:3 ω 3 and ω 6 co-eluted with the first pick, while C20:3 ω 9 co-eluted with the second peak. In the absence of an ω 7 C20:3 standard, we concluded that the most significantly differential peak is either ω 9, ω 7, or a combination of both isomers and refer to this peak as C20:3 ω 9/ ω 7*.

For C22:3, 150fmol and 450fmol C22:3 ω 3 was run with endogenous lipid sample in two separate runs (Supplementary Figure 2.2C). C22:3 ω 3 co-eluted with the first (less significantly PTP1b-dependent) of the two C22:3 isomer peaks. Given the similarity between the profiles of C20:3 and C22:3 (two distinct isomer peaks) and that, in the reverse-phase chromatography used, isomer separation decreases with increasing carbon chain length, we suspect that C22:3 ω 6 would co-elute with C22:3 ω 3, as is the case for C20:3 ω 3 and ω 6, and that the second peak, similarly to the second peak of C20:3, corresponds to ω 7 and/or ω 9 species. Thus, we refer to the PTP1b-dependent C22:3 peak as ω 7*/ ω 9*.

For C24:6, co-injection runs with 50fmol and 300fmol C24:6 ω 3 were performed (Supplementary Figure 2.2D). The single C24:6 peak co-eluted with a C24:6 ω 3 standard, confirming the identification of a C24:6 fatty acid structural isomer. Even though isomer separation decreases with increasing carbon chain length, we can conclude that the isomer peak is not ω 7/ ω 9 because double bonds on fatty acids must be separated by at least one unsaturated carbon, and thus the only other feasible placement of the ω bond would be at the ω 6 position. Here, we denote the peak as C24:6 ω 3/ ω 6*.

Estimation of technical error. The technical error associated with the quantification of hepatic lipids was estimated via analysis of technical replicates. Here, technical replicates are separate samples from one mouse liver tissue that were processed, run, and quantified on different days. There were five such technical replicates for fatty acid analysis. The percent error associated with absolute fatty acid abundances was 17% +/- 15% (mean +/- standard deviation), while the percent error associated with percent composition estimates was 19% +/- 15%. For the triglyceride analysis, there were four such technical replicates, and the percent error associated with absolute triglyceride structural isomer abundances was 15% +/- 10%, while the percent

error associate with percent composition measurements was 22% +/- 10%. For both triglycerides and fatty acids, percent relative standard deviation was constant with respect to lipid abundance.

Enzymatic assay of hepatic triglycerides. Triglycerides were quantified as described (Delibegovic et al., 2009), using the Stanbio Triglyceride Liquicolor Test kit.

Serum triglyceride analysis. Serum triglyceride analysis was performed as described (Rhee et al., 2011), with the following changes: Glyceryl triheptadecanoate (Sigma) standard was added to the isopropanol diluent, which, when mixed with serum (9:1, isopropanol:serum), yielded a final concentration of .5 mg/dL standard. A Jupiter C4 column (4.6 × 150mm, 5µm particle size, 30nm pore size, Phenomenex) was used at a 700uL/min flow rate. Data were acquired on an Agilent 6220 LC-ESI-TOF instrument.

Biochemical analysis

Markers of hepatic inflammation. Bio-Plex Pro Mouse Cytokine Sets were used for quantification of hepatic IL-6, IL-1b, and MCP-1 content, using a Bio-Plex 200 System.

Stochastic multivariate regression analysis

Lipidomics phenotypes as functions of phosphotyrosine measurements. Individual models of steatosis, ω3/ ω6 PTP1b PUFA, and ω7/ ω9 PTP1b PUFA as functions of the phosphotyrosine measurements were constructed using stochastic multivariate regression and treatment-dependent random imputation, because these methods were determined to perform well in Chapter 3. The models were limited to the HFD basal liver tissues. For each model, the phosphotyrosine dataset, which corresponded to 228 phosphorylation sites across 13 samples, was modeled as the independent variable matrix, $X [N \times M]$ ($N = 13$, $M = 228$), while the particular lipid phenotype was modeled as a dependent variable matrix, $Y [N \times K]$, where k is the number of lipidomics

measurements being modeled. (For steatosis, $k = 1$, while, for ω_3/ω_6 PTP1b PUFA (C24:6, C24:5, C24:4, C22:3, C20:3), $k = 5$, and, for ω_7/ω_9 PTP1b PUFA (C20:3, C22:3, C20:2, C18:3), $k = 4$.) The matrices X and Y correspond to measurement means, while X_{std} and Y_{std} are matrices of the same dimensions and contain measurement standard deviation. When no technical replicate existed for a measurement, the standard deviation was estimated using the average relative standard deviation of the given dataset.

The stochastic multivariate regression algorithm, adapted to deal with missing data points by treatment-dependent random imputation, is as follows:

1. *Gaussian sampling for measurement noise.* For each data point present in X , draw a random sample, $X_{gaus}(n,m)$ from a normal distribution with mean, $X(n,m)$, and standard deviation, $X_{std}(n,m)$, $n = 1,2,\dots,N$, $m = 1,2,\dots,M$. Similarly, draw $Y_{gaus}(n,k)$, based on $Y(n,k)$ and $Y_{std}(n,k)$, $n = 1,2,\dots,N$, $k = 1,2,\dots,K$.
2. *Treatment-dependent random imputation for missing data points.* For each measurement (column) of X_{gaus} , estimate the missing data points for L-PTP1b-/- samples with a randomly selected L-PTP1b-/- sample from those present. Similarly, estimate the missing data points for control samples with a randomly selected observation from the controls. (No missing data points existed for the dependent Y matrix.)
3. *X and Y data normalization and averaging of Y matrix measurements.* For each column of X_{gaus} or Y_{gaus} , mean-center the data and divide by the standard deviation to yield matrices X_{zscore} and Y_{zscore} . If $P > 1$, average Y_{zscore} column-wise to yield $Y_{zscore,v}$ [$N \times 1$]. (Averaging Y simplifies subsequent model reduction steps.)

4. *Draw a bootstrap sample.* Randomly select paired rows of X_{zscore} and $Y_{zscore,v}$ without replacement and build X_{boot} and Y_{boot} matrices, dimensions $[N \times M]$ and $[N \times 1]$, respectively.
5. *Build regression model.* Use partial least squares regression (PLSR), with number of principle components equal to one minus the rank of X_{boot} , to estimation the linear regression coefficients B $[M \times 1]$, such that $Y_{boot} = X_{boot}B + \varepsilon$, where ε is the residual error.
6. *Multiple imputation and Gaussian sampling.* Repeat steps 1-5 to create distributions of model coefficients, B . Here, the number of iterations, $itters$, was 5000. Store each vector of coefficients B in the corresponding matrix B_{matrix} $[itters \times M]$.
7. *Estimate the significance of each independent variable measurement to prediction.* For each measurement m in X , use the corresponding coefficient distribution contained in column m of B_{matrix} , to empirically test the null hypothesis that zero belongs to this distribution.
8. *Model reduction.* Rank the independent variables according to coefficient significance determined in 7. Sequentially build and evaluate PLSR models, by varying (1) the number of principle components, pcs , in the model and (2) by incorporating the top F ranked variables in the model. For this analysis, $pcs = \{1,2,\dots,4\}$, and $F = \{1,2,\dots,20\}$.

Here, models were evaluated by estimating R^2 , goodness-of-fit, and Q^2 , goodness-of-prediction, statistics:

$$R^2 = 1 - \frac{\sum_{n=1}^N (Y^{meas,n} - Y^{pred,n})^2}{\sum_{n=1}^N (Y^{meas,n} - Y^{mean})^2}$$

$$Q^2 = 1 - \frac{\sum_{n=1}^N (Y^{meas,n} - Y^{pred,loocv(n)})^2}{\sum_{n=1}^N (Y^{meas,n} - Y^{mean,loocv(n)})^2}$$

where $Y^{\text{meas},n}$ corresponds to the experimentally determined response measurement, n , Y^{mean} is the average of the response measurements, $Y^{\text{pred},n}$ is the model prediction for response measurement, n , $Y^{\text{pred,loocv}(n)}$ is the prediction for response n , from a leave-one-out-cross-validation model, excluding measurements for sample n , and, similarly, $Y^{\text{mean,loocv}(n)}$ is the mean of the response measurements, leaving out measurement n . Importantly, estimation of any set of model predictions for the response vector Y , using the relationship $Y = XB$, requires a complete set of predictor variables X . Thus, to estimate R^2 and Q^2 , a complete matrix, X_c , was created using column-wise treatment-dependent mean imputation. Specifically, for each measurement (column) of X_c , the missing data points for L-PTP1b-/- (or control) samples were estimated as the mean of those L-PTP1b-/- (or control) samples observed for that measurement. Final model parameters, pcs and F , were determined based on optimization of Q^2 values. Important to determination of model quality was estimation of the error associated with the Q^2 values, which was accomplished by leave-two-out-cross-validation.

In addition, reduced models were built by a second method, based on the calculation of the variable importance of projection (VIP) score, commonly used to rank variables for incorporation into the reduced PLSR models (Huang et al., 2010; Kumar et al., 2007). The VIP score for a given independent variable and PLSR model with pcs principle components is:

$$VIP(m, pcs) = \sqrt{\frac{\sum_{pc=1}^{pcs} Var_{exp,Y}(pc)W(m, pc)^2}{\sum_{pc=1}^{pcs} Var_{exp,Y}(pc)}}$$

where $Var_{exp,Y}(pc)$ is the variance in Y explained by regression onto a particular principle component, pc , and $W(m,c)$ is the quantitative contribution that a particular measurement, m , makes to the pc^{th} PLSR principle component (Kumar et al., 2007). Because the PLSR principle

component plane is optimized to contain linear combinations of independent variable measurements that covary maximally with the dependent variable, this metric performs well in model reduction.

The method detailed below is an adaption of the methods described in (Kumar et al., 2007). In this variation of stochastic multivariate regression, at step 5, the X_{boot} loadings matrix P_{boot} [$M \times pcs$] was stored, P_{boot} such that $X_{boot} = T_{boot}P_{boot}^T$, $Y_{boot} = U_{boot}C_{boot}^T$, and $covariance(T_{boot}, U_{boot})$ is maximal, P_{boot} [$N \times pcs$], U_{boot} [$M \times pcs$], C_{boot} [$1 \times pcs$], $pcs = \min(4, \text{rank}(X_{boot}))$, and Z^T denotes the transpose of Z (Huang et al., 2010). The loadings were stored in a 3D-matrix P_{matrix} [$itters \times M \times pcs$] (step 6), and then used to calculate VIP scores for each of the M independent variable measurements (step 7). Specifically, the loadings matrices were averaged across the iterations' dimension of P_{matrix} to create P_{ave} [$M \times pcs$]. Thus, by averaging, variable measurements that consistently contribute to the PLSR principle component plane will have higher loading values and contribution to the model more than measurements that are randomly imputed or associated with larger amounts of error. To make this loadings matrix consistent with loadings matrix produced directly from PLSR algorithms, P_{ave} was orthonormalized row-wise. Next, the scores matrix, T [$N \times pcs$], was estimated by the equation $T = X_c P_{ave}^T$ and the weights, W , were estimated as $P_{ave} / (P_{ave}^T P_{ave})$. The rows of T were orthogonalized, and, then the Y loadings, YL [$pcs \times 1$] were calculated from the relationship $YL = (T^T T)^{-1} T^T Y$. Next, the $Var_{exp,Y}(pc)$, was calculated using the equation: $Var_{exp,Y}(pc) = 1 - (Y - T(pc)^T YL(pc))^T (Y - T(pc)^T YL(pc)) / Var_Y$, where $T(pc)$ is the pc^{th} column of T and $YL(pc)$ is the pc^{th} row of YL transposed to a column vector. Thus, after a fair amount of linear algebra, the VIP score can be calculated for each measurement and total number of pcs in the model. Step 8 is then performed based on VIP-score model reduction.

A more detailed description and discussion of these methods is described in Chapter 3.

Steatosis as functions of lipidomics measurements. Models of steatosis were also built as functions of FA composition or triglyceride composition, using the methods described above. The independent matrix, X , for the FA-composition-based steatosis model was [15 mice \times 33 FA profiles], while X was [14 mice \times 26 triglyceride species] for the triglyceride-composition-base model; corresponding dependent Y matrices were [15 \times 1] and [14 \times 1]. Final models were selected as described above for phosphotyrosine-based models.

Phosphosite set enrichment analysis

Phosphosite set enrichment analysis was accomplished by adaption of Gene Set Enrichment analysis (GSEA) as described (Subramanian et al., 2005) for set enrichment from quantitative, post translation modification (PTM) datasets; this adaption of GSEA, termed, PTM Set Enrichment Analysis (PSEA), and motivated by this study specifically, is described in Chapter 3. Briefly, the goal of PSEA is to detect whether certain PTM site features (kinase motif, protein cellular process) are significantly over-represented among PTM sites implicated in a quantitatively defined relationship (sensitivity to a perturbation or importance to phenotype prediction). PSEA inputs include (1) list of PTMs quantitatively ranked according to phenotype or condition relevance and (2) at least one set, defined by a common feature (cellular compartment, protein domain) specific to a subset of the PTMs measured.

For analysis here, PTM sets (59 in total) were hand-curated sets of phosphosites, composed of protein pathway annotations from Kegg, gene ontologies, protein structural information (PFAM and SwissPro), a curated list of PTP1b substrates (Ren et al., 2011), as well as a set of insulin-sensitive phosphorylation sites from a previous study (Schmelzle et al., 2006). These sets were

then used in five independent PSEA analyses to search for enrichment among (1) PTP1b-dependent sites, (2) HFD-dependent sites, and sites predictive of (3) steatosis, (4) ω 3/ ω 6 PTP1b PUFA, and (5) ω 7/ ω 9 PTP1b PUFA. The multivariate stochastic regression method described above was used to generate ranked lists of sites for analyses (3)-(5); specifically, the ranking of phosphosites was dependent on the significance of phosphosite regression coefficient, as calculated above. Sites were ranked according to the sign of the regression coefficient multiplied by the $-\log_{10}$ -transformed p-value associated with the coefficient. This analysis was limited to the HFD basal context only. Analyses (1) and (2) incorporated all phosphotyrosine datasets; sites were ranked according to significance of correlation with genotype or diet. In particular, the significance and direction of correlation of genotype with a phosphosite was calculated, if possible, for the four relevant data subdivisions (HFD basal, HFD insulin, NC basal, and NC insulin), then these p-values were combined using Fisher's method, accounting for the direction of the correlation. (This method is slightly different than the supervised-clustering framework described above, as it measures overall correlation only.) The same was done for diet, where the four relevant data subdivisions were basal L-PTP1b^{-/-}, insulin L-PTP1b^{-/-}, basal control, and insulin control. Phosphosites in analyses (1) and (2) were ranked according to the direction of correlation coefficient multiplied by the associated $-\log_{10}$ -transformed p-value. Thus, in all analyses described here, the sign of the relationship was taken into account, indicating that our method would detect set/pathway enrichments for groups of phosphosites that increased or decreased phosphorylation according to a particular phenotype or treatment condition. As phosphorylation has many functions, including protein inhibition or activation, a consideration of unsigned analysis for this dataset can be found in Chapter 3.

For each analysis, the lower limit on set size ensured membership of at least five phosphosites. In addition, given missing data, it was not possible to rank phosphosites according to all treatments and phenotypes. Thus, the genotype and diet analyses tested 59 and 45 sets, while 56 sets were tested for relevance to phenotype prediction. A complete list of PSEA sets tested and results are available upon request. A subset of analysis results, those sets found to be significantly enriched controlling for an FDR < .1 are found in Tables 2.2, 2.4, 2.7, 2.8, and 2.9.

Note: PSEA yielded numerous PTP1b-dependent sets but only a single, significantly diet-dependent set. This is due to experimental design. Each MS 8-plex iTRAQ experiment was optimized for comparison of L-PTP1b^{-/-} mice to control rather than for comparison of HFD to NC mice. As a result, whereas every MS experiment enabled genotype comparison, diet comparison was accomplished by pooling the overlapping sites from multiple runs. Thus, fewer samples were available for the diet comparison, and we lacked statistical power equivalent to the genotype-dependent case. If additional MS experiments were run for diet comparison, additional diet-dependent phosphotyrosine sets would likely be uncovered.

PTP1b-liver-specific-deletion mice were generated by crossing PTP1b-floxed (PTP1b-fl/fl) mice with mice expressing Cre recombinase under the control of the albumin promoter (Alb-Cre), resulting in Alb-Cre PTP1b-fl/fl mice that specifically lack hepatic PTP1b expression (Delibegovic et al., 2009). Alb-Cre mice were used as controls. Genotyping for the PTP1b floxed allele and the presence of Cre was performed by PCR. Hepatic PTP1b protein expression was accessed at the end of the study, by quantitative immunoblotting (PTP1b antibody AF13661, R&D Systems). Mice were maintained on a 12-hour light/dark cycle in a temperature-controlled

barrier facility with water and food freely accessible. All mice were maintained on a normal chow (NC) diet (Prolab Isopro RMH 3000, 15% calories from fat) until twelve weeks of age, at which point mice either continued on NC or were placed on high fat diet (Teklad TD.93075, 55% calories from fat). Mouse procedures were performed in accordance with the guidelines and approval of the Institutional Animal Care and Use Committee at the University of Massachusetts Medical School and the Ontario Cancer Institute.

References

- Agouni, A., Mody, N., Owen, C., Czopek, A., Zimmer, D., Bentires-Alj, M., Bence, K.K., and Delibegovic, M. (2011). Liver-specific deletion of protein tyrosine phosphatase (PTP) 1B improves obesity- and pharmacologically induced endoplasmic reticulum stress. *Biochem J* 438, 369-378.
- Biddinger, S.B., Hernandez-Ono, A., Rask-Madsen, C., Haas, J.T., Aleman, J.O., Suzuki, R., Scapa, E.F., Agarwal, C., Carey, M.C., Stephanopoulos, G., *et al.* (2008). Hepatic insulin resistance is sufficient to produce dyslipidemia and susceptibility to atherosclerosis. *Cell Metab* 7, 125-134.
- Biddinger, S.B., and Kahn, C.R. (2006). From mice to men: insights into the insulin resistance syndromes. *Annu Rev Physiol* 68, 123-158.
- Brown, M.S., and Goldstein, J.L. (2008). Selective versus total insulin resistance: a pathogenic paradox. *Cell Metab* 7, 95-96.
- Cai, D., Yuan, M., Frantz, D.F., Melendez, P.A., Hansen, L., Lee, J., and Shoelson, S.E. (2005). Local and systemic insulin resistance resulting from hepatic activation of IKK-beta and NF-kappaB. *Nat Med* 11, 183-190.
- Coleman, R.A., and Lee, D.P. (2004). Enzymes of triacylglycerol synthesis and their regulation. *Prog Lipid Res* 43, 134-176.
- Delibegovic, M., Zimmer, D., Kauffman, C., Rak, K., Hong, E.G., Cho, Y.R., Kim, J.K., Kahn, B.B., Neel, B.G., and Bence, K.K. (2009). Liver-specific deletion of protein-tyrosine phosphatase 1B (PTP1B) improves metabolic syndrome and attenuates diet-induced endoplasmic reticulum stress. *Diabetes* 58, 590-599.
- Donnelly, K.L., Smith, C.I., Schwarzenberg, S.J., Jessurun, J., Boldt, M.D., and Parks, E.J. (2005). Sources of fatty acids stored in liver and secreted via lipoproteins in patients with nonalcoholic fatty liver disease. *J Clin Invest* 115, 1343-1351.
- Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A.L., Normandin, D., Cheng, A., Himms-Hagen, J., and Chan, C.C. (1999). Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* 283, 1544.

- Fan, J., Hitosugi, T., Chung, T.W., Xie, J., Ge, Q., Gu, T.L., Polakiewicz, R.D., Chen, G.Z., Boggon, T.J., Lonial, S., *et al.* (2011). Tyrosine phosphorylation of lactate dehydrogenase A is important for NADH/NAD(+) redox homeostasis in cancer cells. *Mol Cell Biol* 31, 4938-4950.
- Feng, J., Lucchinetti, E., Enkavi, G., Wang, Y., Gehrig, P., Roschitzki, B., Schaub, M.C., Tajkhorshid, E., Zaugg, K., and Zaugg, M. (2010). Tyrosine phosphorylation by Src within the cavity of the adenine nucleotide translocase 1 regulates ADP/ATP exchange in mitochondria. *Am J Physiol Cell Physiol* 298, C740-748.
- Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. *Science* 315, 972.
- Ginsberg, H.N., and Fisher, E.A. (2009). The ever-expanding role of degradation in the regulation of apolipoprotein B metabolism. *J Lipid Res* 50 Suppl, S162-166.
- Gregor, M.F., and Hotamisligil, G.S. (2011). Inflammatory mechanisms in obesity. *Annu Rev Immunol* 29, 415-445.
- Guillou, H., D'Andrea, S., Rioux, V., Barnouin, R., Dalaine, S., Pedrono, F., Jan, S., and Legrand, P. (2004). Distinct roles of endoplasmic reticulum cytochrome b5 and fused cytochrome b5-like domain for rat Delta6-desaturase activity. *J Lipid Res* 45, 32-40.
- Guillou, H., Zdravec, D., Martin, P.G., and Jacobsson, A. (2010). The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog Lipid Res* 49, 186-199.
- Haque, A., Andersen, J.N., Salmeen, A., Barford, D., and Tonks, N.K. (2011). Conformation-sensing antibodies stabilize the oxidized form of PTP1B and inhibit its phosphatase activity. *Cell* 147, 185-198.
- Hitosugi, T., Fan, J., Chung, T.W., Lythgoe, K., Wang, X., Xie, J., Ge, Q., Gu, T.L., Polakiewicz, R.D., Roesel, J.L., *et al.* (2011). Tyrosine phosphorylation of mitochondrial pyruvate dehydrogenase kinase 1 is important for cancer metabolism. *Mol Cell* 44, 864-877.
- Hitosugi, T., Kang, S., Vander Heiden, M.G., Chung, T.W., Elf, S., Lythgoe, K., Dong, S., Lonial, S., Wang, X., Chen, G.Z., *et al.* (2009). Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. *Sci Signal* 2, ra73.
- Homan, E.A., Kim, Y.G., Cardia, J.P., and Saghatelian, A. (2011). Monoalkylglycerol ether lipids promote adipogenesis. *J Am Chem Soc* 133, 5178-5181.
- Huang, P.H., Miraldi, E.R., Xu, A.M., Kundukulam, V.A., Del Rosario, A.M., Flynn, R.A., Cavenee, W.K., Furnari, F.B., and White, F.M. (2010). Phosphotyrosine signaling analysis of site-specific mutations on EGFRvIII identifies determinants governing glioblastoma cell growth. *Mol Biosyst* 6, 1227-1237.
- Ji, C., and Kaplowitz, N. (2006). ER stress: can the liver cope? *J Hepatol* 45, 321-333.
- Kashyap, S.R., and DeFronzo, R.A. (2007). The insulin resistance syndrome: physiological considerations. *Diab Vasc Dis Res* 4, 13-19.
- Koltun, D.O., Zilbershtein, T.M., Migulin, V.A., Vasilevich, N.I., Parkhill, E.Q., Glushkov, A.I., McGregor, M.J., Brunn, S.A., Chu, N., Hao, J., *et al.* (2009). Potent, orally bioavailable, liver-selective stearoyl-CoA desaturase (SCD) inhibitors. *Bioorg Med Chem Lett* 19, 4070-4074.

Kumar, N., Wolf-Yadlin, A., White, F.M., and Lauffenburger, D.A. (2007). Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. *PLoS computational biology* 3, e4.

Liu, X., Strable, M.S., and Ntambi, J.M. (2011). Stearoyl CoA desaturase 1: role in cellular inflammation and stress. *Adv Nutr* 2, 15-22.

Man, W.C., Miyazaki, M., Chu, K., and Ntambi, J. (2006). Colocalization of SCD1 and DGAT2: implying preference for endogenous monounsaturated fatty acids in triglyceride synthesis. *J Lipid Res* 47, 1928-1939.

Manevich, Y., Reddy, K.S., Shuvaeva, T., Feinstein, S.I., and Fisher, A.B. (2007). Structure and phospholipase function of peroxiredoxin 6: identification of the catalytic triad and its role in phospholipid substrate binding. *J Lipid Res* 48, 2306-2318.

Martin, S.E., Shabanowitz, J., Hunt, D.F., and Marto, J.A. (2000). Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 72, 4266-4274.

Michael, M.D., Kulkarni, R.N., Postic, C., Previs, S.F., Shulman, G.I., Magnuson, M.A., and Kahn, C.R. (2000). Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol Cell* 6, 87-97.

Naegle, K.M., Gymrek, M., Joughin, B.A., Wagner, J.P., Welsch, R.E., Yaffe, M.B., Lauffenburger, D.A., and White, F.M. (2010). PTMScout, a Web resource for analysis of high throughput post-translational proteomics studies. *Mol Cell Proteomics* 9, 2558-2570.

Nakatani, Y., Kaneto, H., Kawamori, D., Hatazaki, M., Miyatsuka, T., Matsuoka, T., Kajimoto, Y., Matsuhisa, M., Yamasaki, Y., and Hori, M. (2004). Modulation of the JNK pathway in liver affects insulin resistance status. *Journal of Biological Chemistry* 279, 45803.

Oikari, S., Ahtialansaari, T., Huotari, A., Kiehne, K., Folsch, U.R., Wolfram, S., Janne, J., Alhonen, L., and Herzig, K.H. (2008). Effect of medium- and long-chain fatty acid diets on PPAR and SREBP-1 expression and glucose homeostasis in ACBP-overexpressing transgenic rats. *Acta Physiol (Oxf)* 194, 57-65.

Ozcan, U., Cao, Q., Yilmaz, E., Lee, A.H., Iwakoshi, N.N., Ozdelen, E., Tuncman, G., Gorgun, C., Glimcher, L.H., and Hotamisligil, G.S. (2004). Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. *Science* 306, 457-461.

Pan, M., Cederbaum, A.I., Zhang, Y.L., Ginsberg, H.N., Williams, K.J., and Fisher, E.A. (2004). Lipid peroxidation and oxidant stress regulate hepatic apolipoprotein B degradation and VLDL production. *J Clin Invest* 113, 1277-1287.

Popkin, B.M., Adair, L.S., and Ng, S.W. (2012). Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev* 70, 3-21.

Ren, L., Chen, X., Luechapanichkul, R., Selner, N.G., Meyer, T.M., Wavreille, A.S., Chan, R., Iorio, C., Zhou, X., Neel, B.G., *et al.* (2011). Substrate specificity of protein tyrosine phosphatases 1B, RPTPalph, SHP-1, and SHP-2. *Biochemistry* 50, 2339-2356.

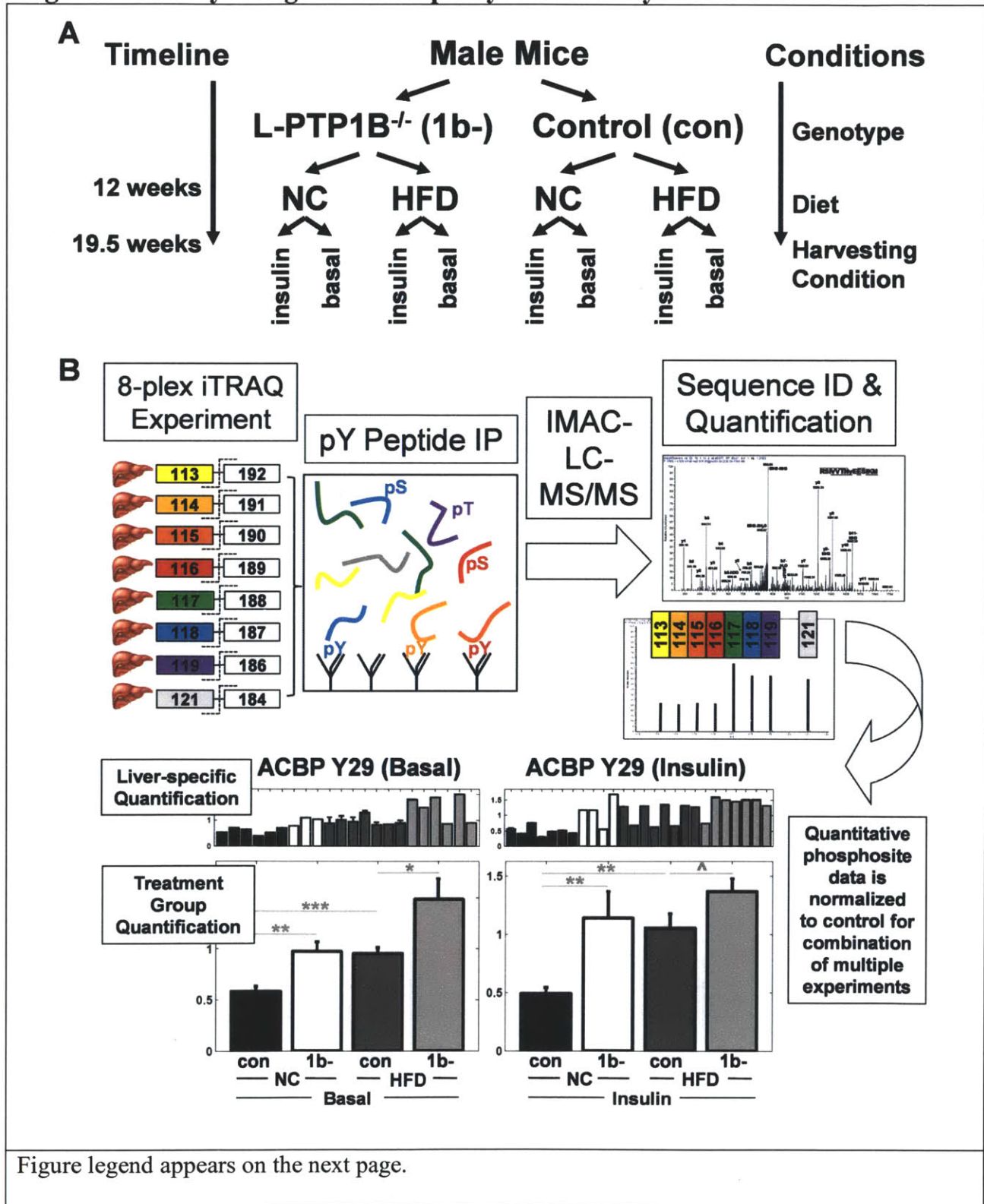
Rhee, E.P., Cheng, S., Larson, M.G., Walford, G.A., Lewis, G.D., McCabe, E., Yang, E., Farrell, L., Fox, C.S., O'Donnell, C.J., *et al.* (2011). Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J Clin Invest* 121, 1402-1411.

- Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys, Vol 519 (Wiley Online Library).
- Rutter, G.A. (2000). Diabetes: the importance of the liver. *Curr Biol* 10, R736-738.
- Saltiel, A.R., and Pessin, J.E. (2002). Insulin signaling pathways in time and space. *Trends Cell Biol* 12, 65-71.
- Samuel, V.T., Liu, Z.X., Qu, X., Elder, B.D., Bilz, S., Befroy, D., Romanelli, A.J., and Shulman, G.I. (2004). Mechanism of hepatic insulin resistance in non-alcoholic fatty liver disease. *J Biol Chem* 279, 32345-32353.
- Schenkman, J.B., and Jansson, I. (2003). The many roles of cytochrome b5. *Pharmacol Ther* 97, 139-152.
- Schmelzle, K., Kane, S., Gridley, S., Lienhard, G.E., and White, F.M. (2006). Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 55, 2171-2179.
- Shimizu, S., Ugi, S., Maegawa, H., Egawa, K., Nishio, Y., Yoshizaki, T., Shi, K., Nagai, Y., Morino, K., Nemoto, K., *et al.* (2003). Protein-tyrosine phosphatase 1B as new activator for hepatic lipogenesis via sterol regulatory element-binding protein-1 gene expression. *J Biol Chem* 278, 43095-43101.
- Shimomura, I., Matsuda, M., Hammer, R.E., Bashmakov, Y., Brown, M.S., and Goldstein, J.L. (2000). Decreased IRS-2 and increased SREBP-1c lead to mixed insulin resistance and sensitivity in livers of lipodystrophic and ob/ob mice. *Mol Cell* 6, 77-86.
- Sidorenkov, O., Nilssen, O., Brenn, T., Martiushov, S., Arkhipovsky, V.L., and Grjibovski, A.M. (2010). Prevalence of the metabolic syndrome and its components in Northwest Russia: the Arkhangelsk study. *BMC Public Health* 10, 23.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78, 779-787.
- Stefan, N., Peter, A., Cegan, A., Staiger, H., Machann, J., Schick, F., Claussen, C.D., Fritsche, A., Haring, H.U., and Schleicher, E. (2008). Low hepatic stearyl-CoA desaturase 1 activity is associated with fatty liver and insulin resistance in obese humans. *Diabetologia* 51, 648-656.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545.
- Taskinen, J.P., van Aalten, D.M., Knudsen, J., and Wierenga, R.K. (2007). High resolution crystal structures of unliganded and liganded human liver ACBP reveal a new mode of binding for the acyl-CoA ligand. *Proteins* 66, 229-238.
- Vinayavekhin, N., and Saghatelian, A. (2011). Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach. *J Am Chem Soc* 133, 17168-17171.

- Wang, Q., Jiang, L., Wang, J., Li, S., Yu, Y., You, J., Zeng, R., Gao, X., Rui, L., Li, W., *et al.* (2009). Abrogation of hepatic ATP-citrate lyase protects against fatty liver and ameliorates hyperglycemia in leptin receptor-deficient mice. *Hepatology* *49*, 1166-1175.
- Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T.V., Cross, J.R., and Thompson, C.B. (2009). ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* *324*, 1076-1080.
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems* *2*, 37-52.
- Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A., and White, F.M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* *104*, 5860.
- Yamazaki, T., Sasaki, E., Kakinuma, C., Yano, T., Miura, S., and Ezaki, O. (2005). Increased very low density lipoprotein secretion and gonadal fat mass in mice overexpressing liver DGAT1. *J Biol Chem* *280*, 21506-21514.
- Yip, S.C., Saha, S., and Chernoff, J. (2010). PTP1B: a double agent in metabolism and oncogenesis. *Trends Biochem Sci* *35*, 442-449.
- Zhang, K., and Kaufman, R.J. (2008). From endoplasmic-reticulum stress to the inflammatory response. *Nature* *454*, 455-462.
- Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A., and White, F.M. (2005). Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* *4*, 1240-1250.

Figures

Figure 2.1 Study Design and Phosphotyrosine Analysis



(A) Study design.

(B) Eight peptide samples, purified from individual livers, were chemically labeled individually with a unique 8-plex iTRAQ isobaric mass tag. Labeled peptide samples were combined and phosphotyrosine-containing peptides were enriched by phosphotyrosine peptide IP using pan-specific phosphotyrosine antibodies. Phosphorylated peptides were further enriched by IMAC prior to analysis by LC-MS/MS, resulting in hundreds of MS/MS phosphopeptide fragmentation spectra. Each MS/MS spectrum yields (1) sequence information and (2) iTRAQ reporter peaks whose intensities are proportional to the abundance of that particular phosphopeptide in the corresponding samples. To enable phosphopeptide quantification across multiple 8-plex MS experiments, each experiment was normalized to a control liver peptide sample, kept constant across experiments.

Figure 2.2 Diet and Genotype Dependencies in the Phosphotyrosine Network

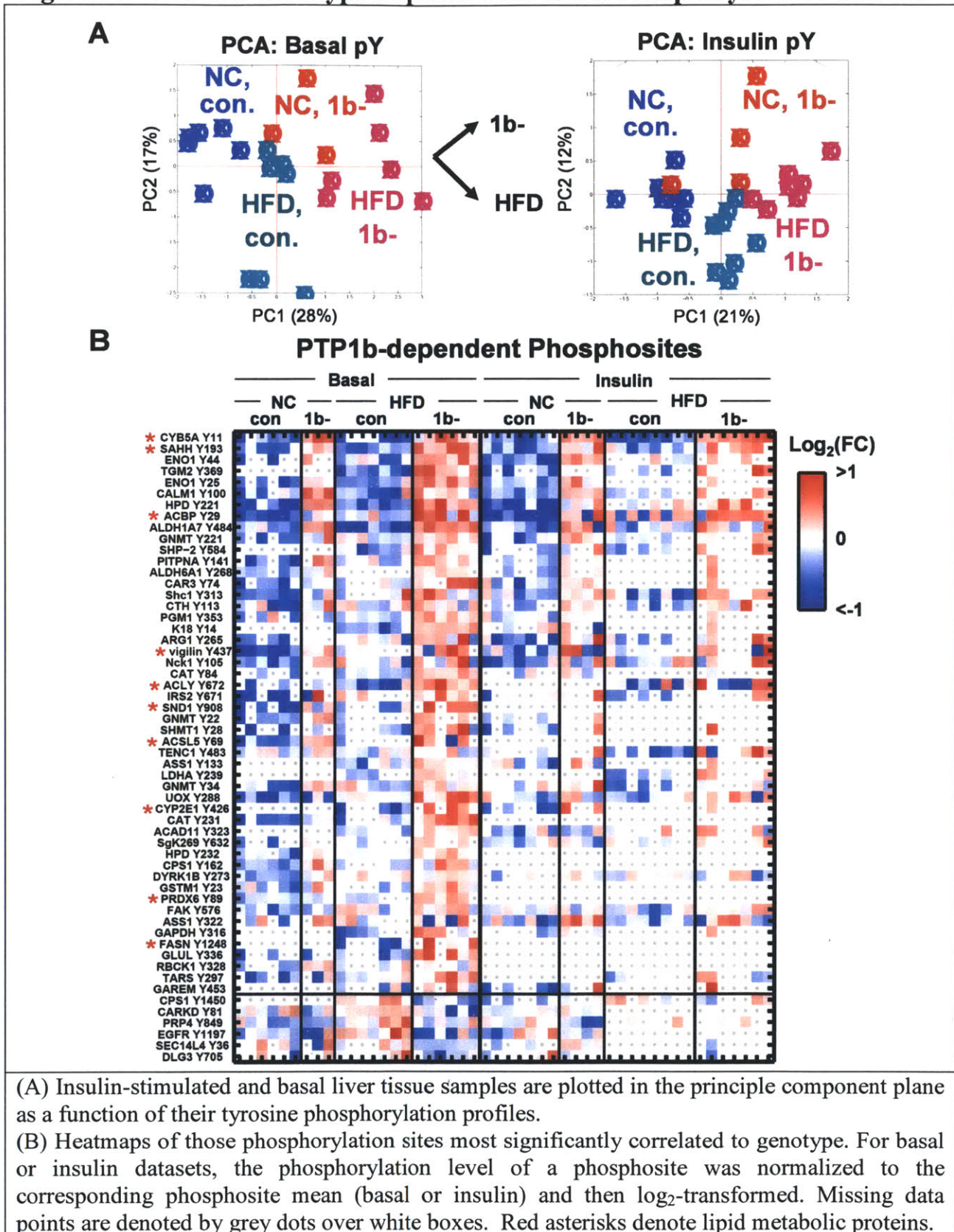


Figure 2.3 Global Lipidomics Analysis for *De Novo* Discovery

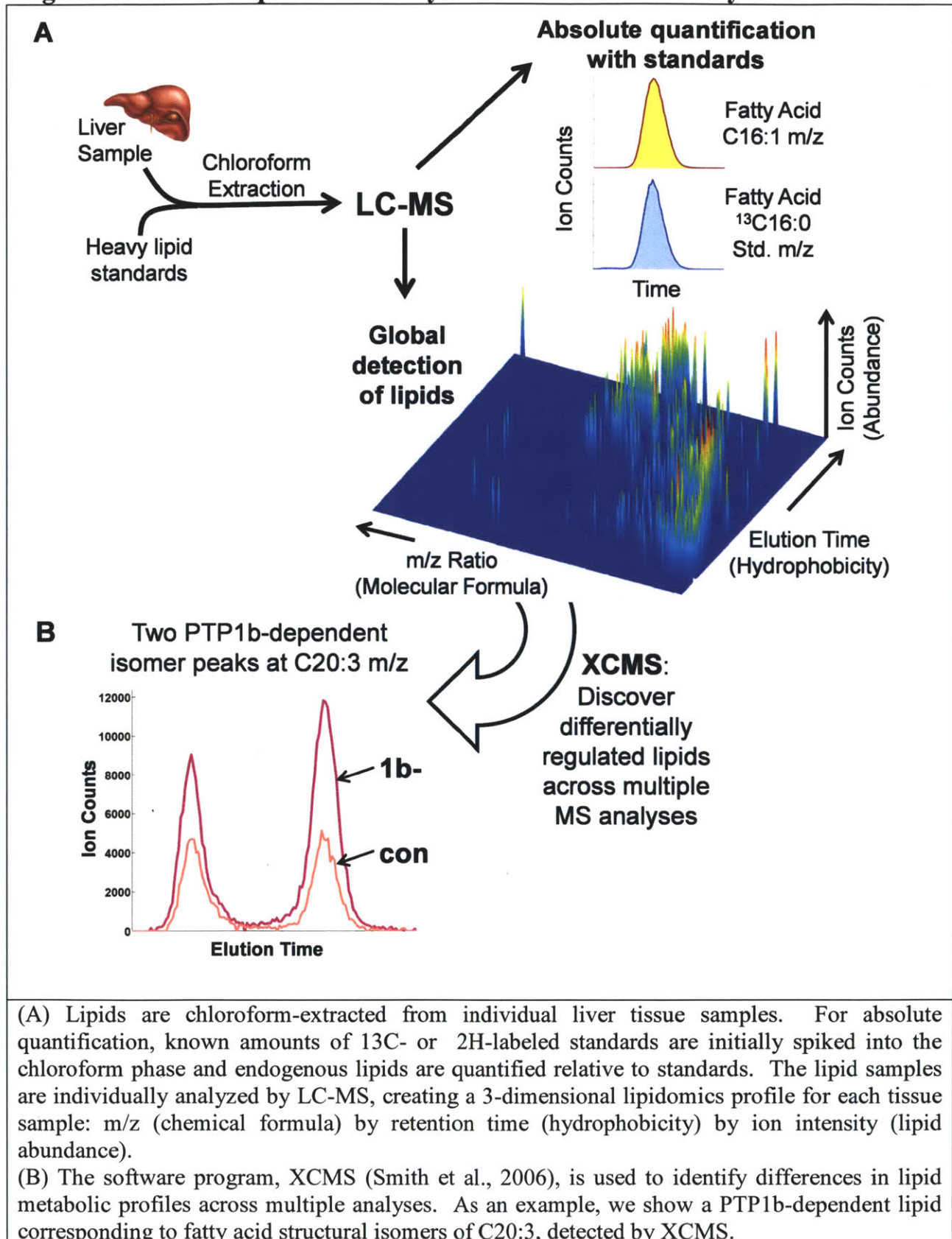
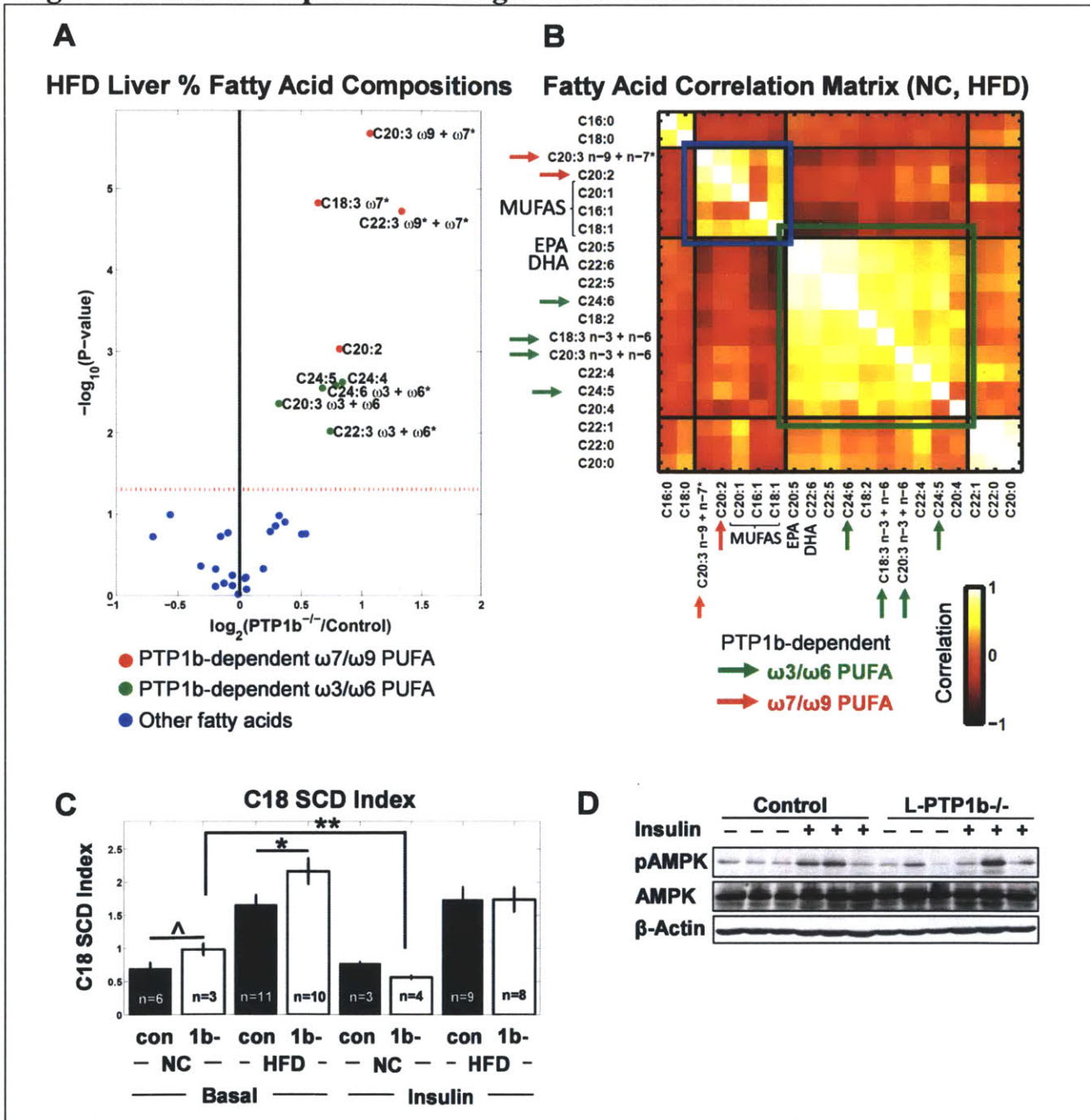


Figure 2.4 PTP1b-dependent Changes in FFA Metabolism



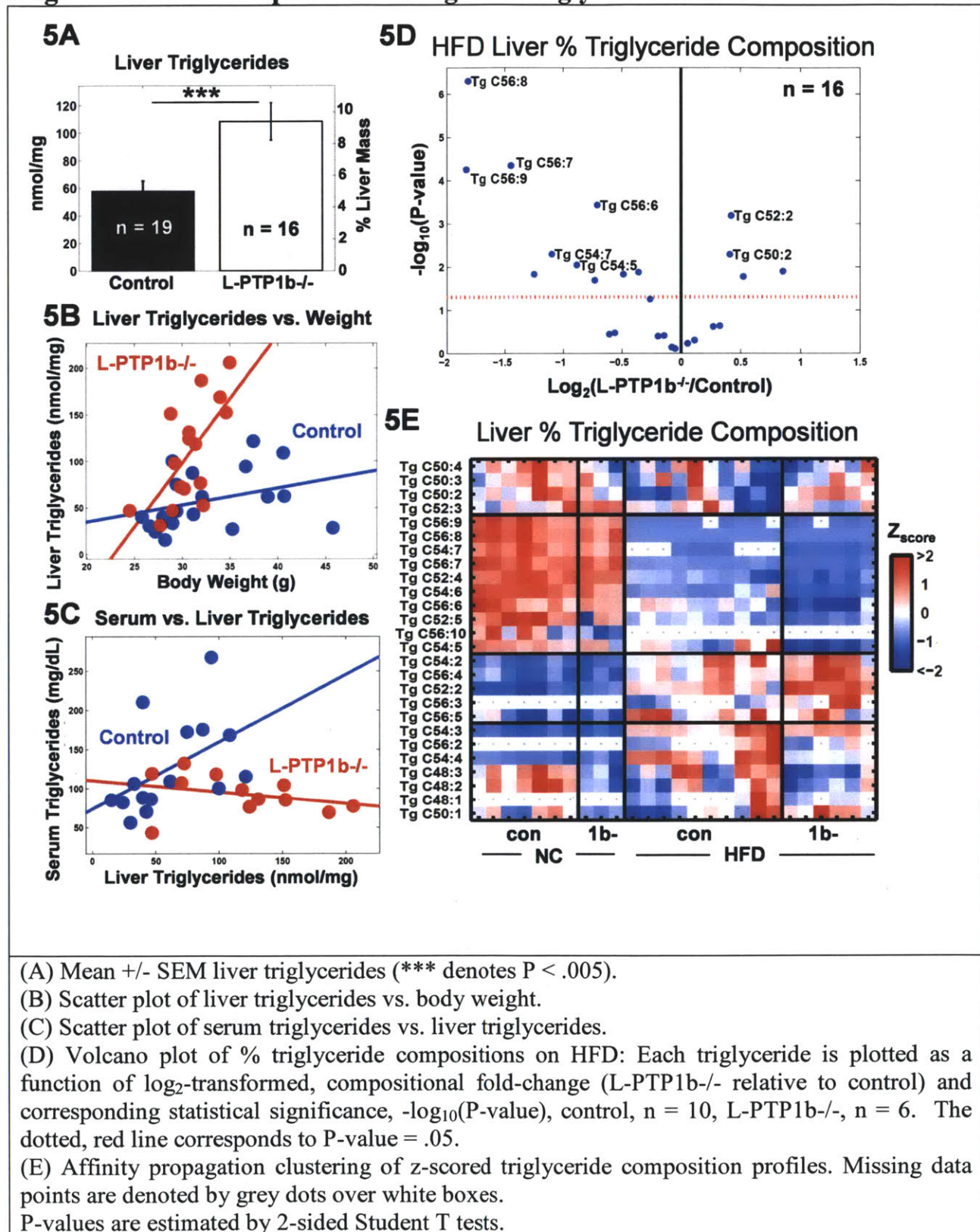
(A) Volcano plot of % FA compositions on HFD: Each FA is plotted as a function of \log_2 -transformed, compositional fold-change (L-PTP1b^{-/-} relative to control) and corresponding statistical significance ($-\log_{10}(P\text{-value})$, control, n = 19, L-PTP1b^{-/-}, n = 17). The dotted, red line corresponds to P-value = .05.

(B) The clustered FA correlation matrix; the correlation coefficients between FAs were calculated using % FA compositional measurements across NC and HFD mice.

(C) Mean \pm SEM C18 SCD indices (the ratio of C18:1 to C18:0). P-values are estimated by 2-sided Student T tests; **, *, and ^ denote P < .01, .05, and .1.

(D) Western blot of hepatic pT172-AMPK, total AMPK, and β actin in HFD mice..

Figure 2.5 PTP1b-dependent Changes in Triglyceride Metabolism



(A) Mean +/- SEM liver triglycerides (***) denotes P < .005).

(B) Scatter plot of liver triglycerides vs. body weight.

(C) Scatter plot of serum triglycerides vs. liver triglycerides.

(D) Volcano plot of % triglyceride compositions on HFD: Each triglyceride is plotted as a function of \log_2 -transformed, compositional fold-change (L-PTP1b^{-/-} relative to control) and corresponding statistical significance, $-\log_{10}(P\text{-value})$, control, n = 10, L-PTP1b^{-/-}, n = 6. The dotted, red line corresponds to P-value = .05.

(E) Affinity propagation clustering of z-scored triglyceride composition profiles. Missing data points are denoted by grey dots over white boxes.

P-values are estimated by 2-sided Student T tests.

Figure 2.6 Multivariate Models of Lipidomics Phenotypes

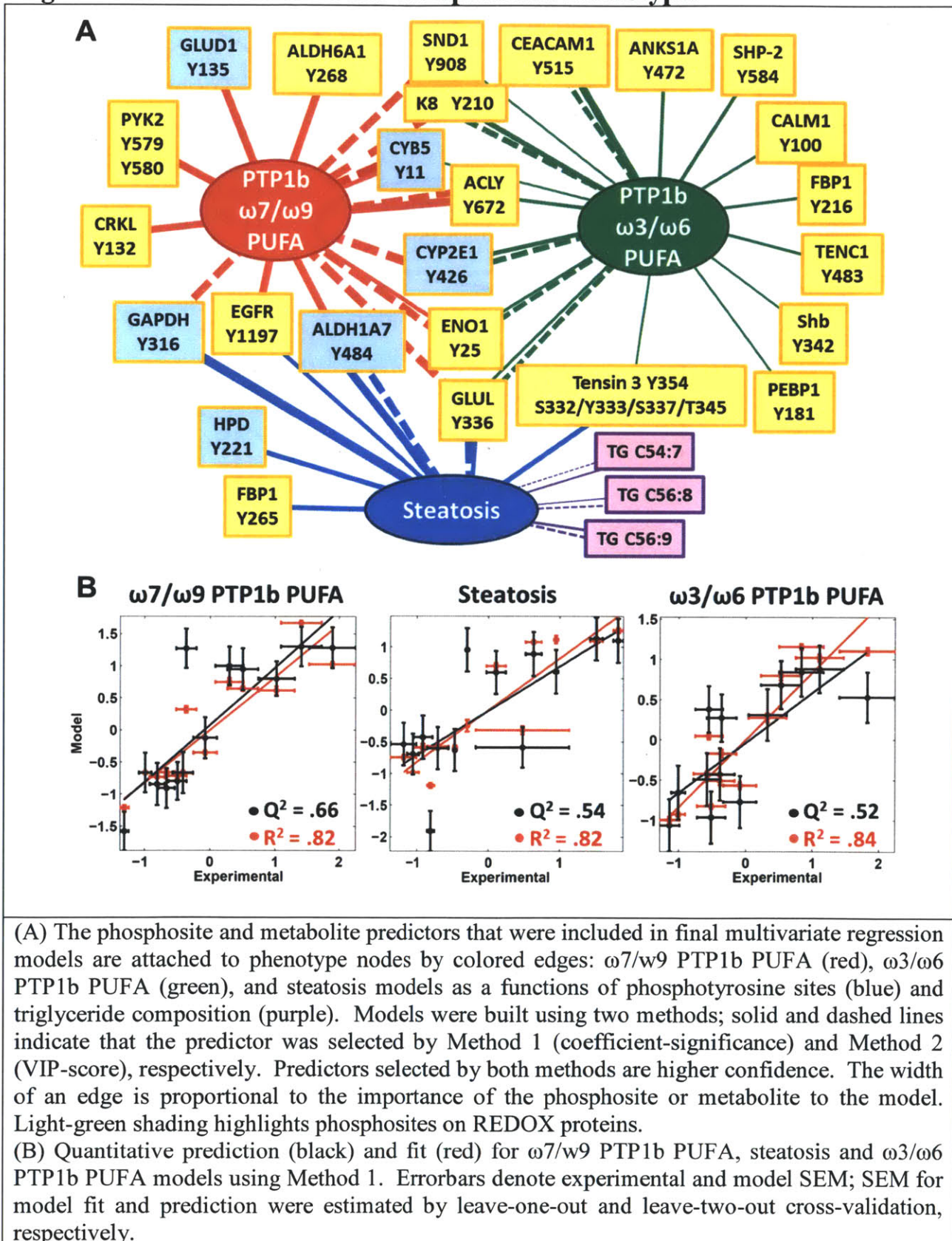
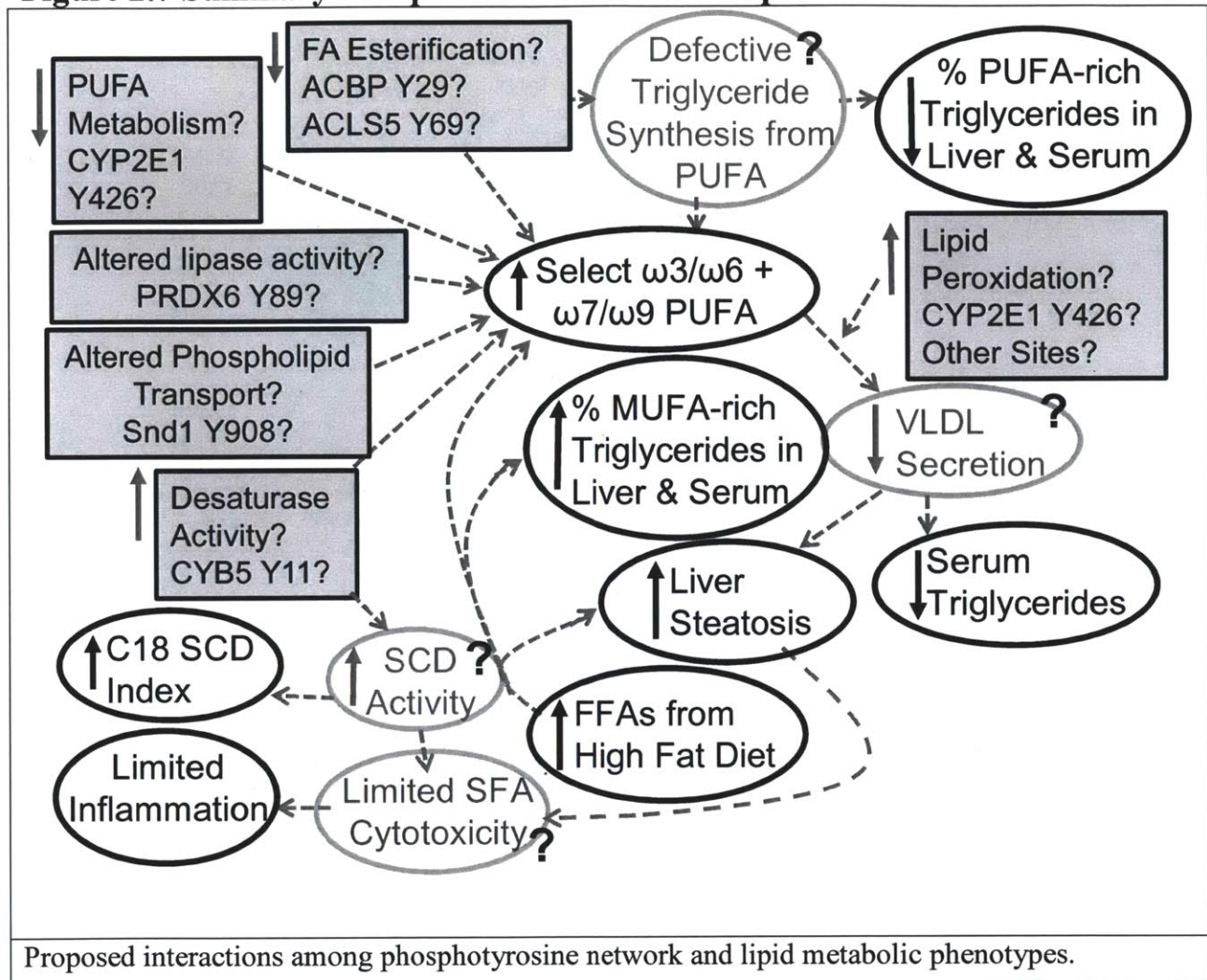
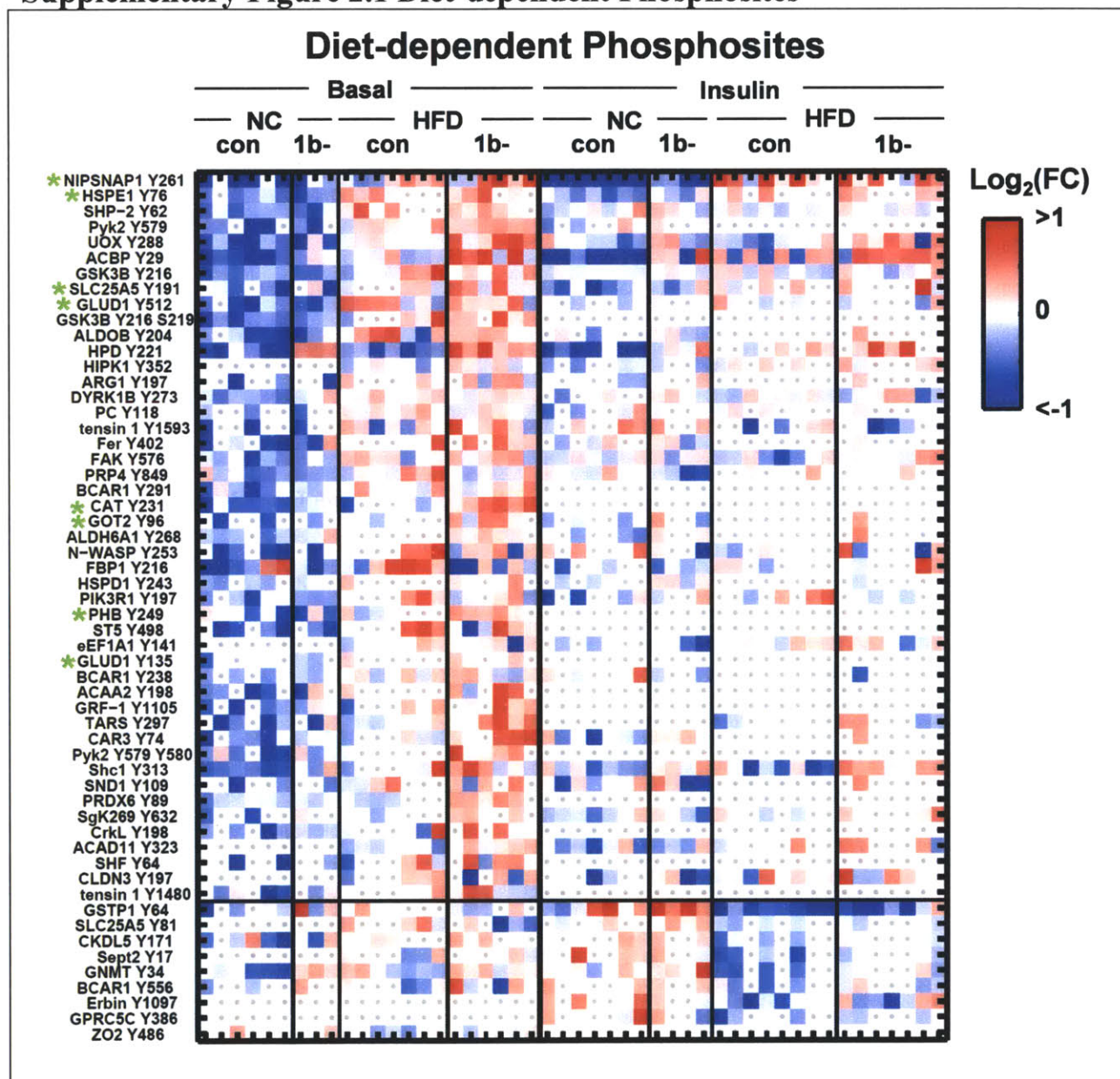


Figure 2.7 Summary of Lipid Metabolic and Phosphosite Interactions



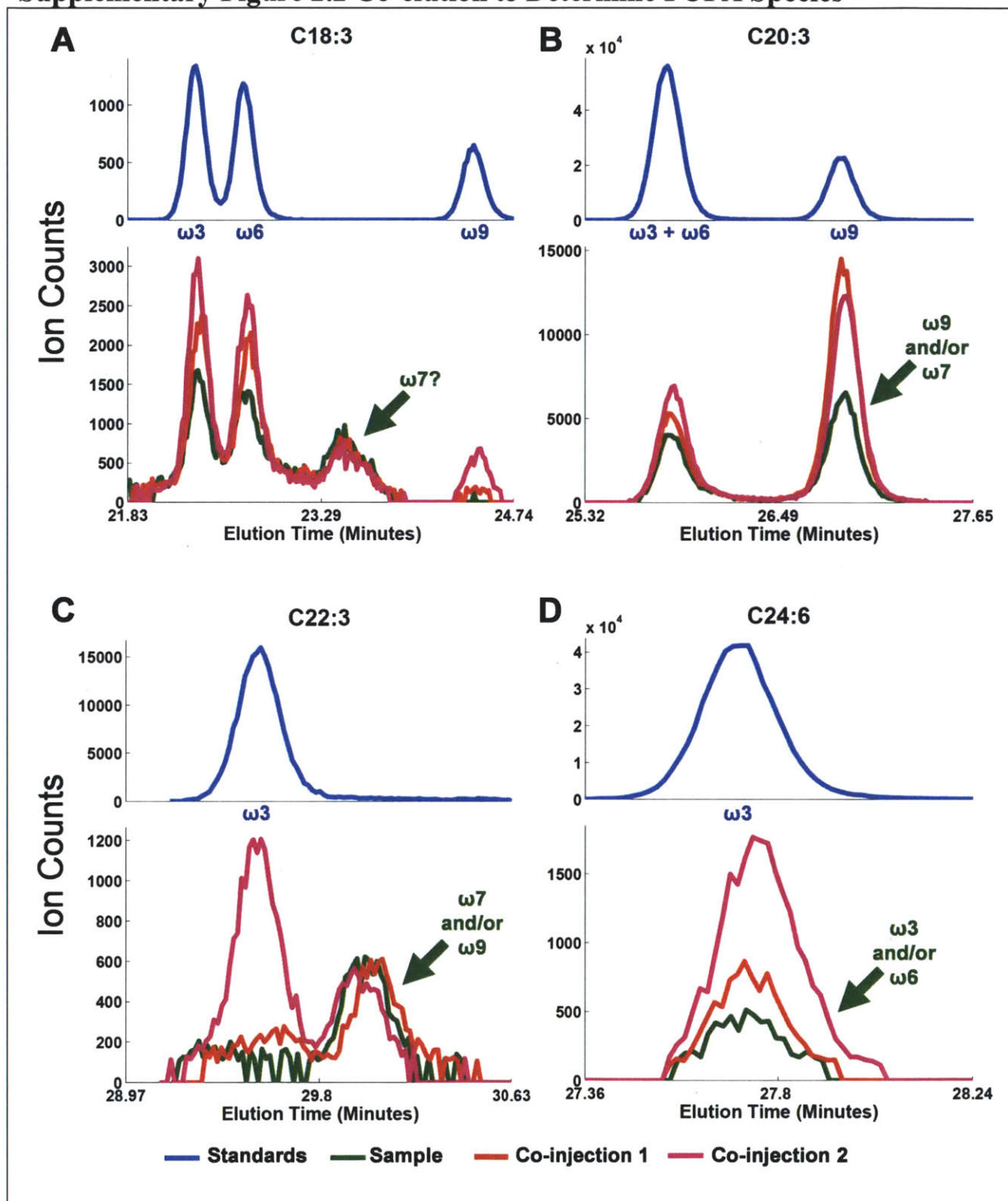
Supplementary Figures

Supplementary Figure 2.1 Diet-dependent Phosphosites



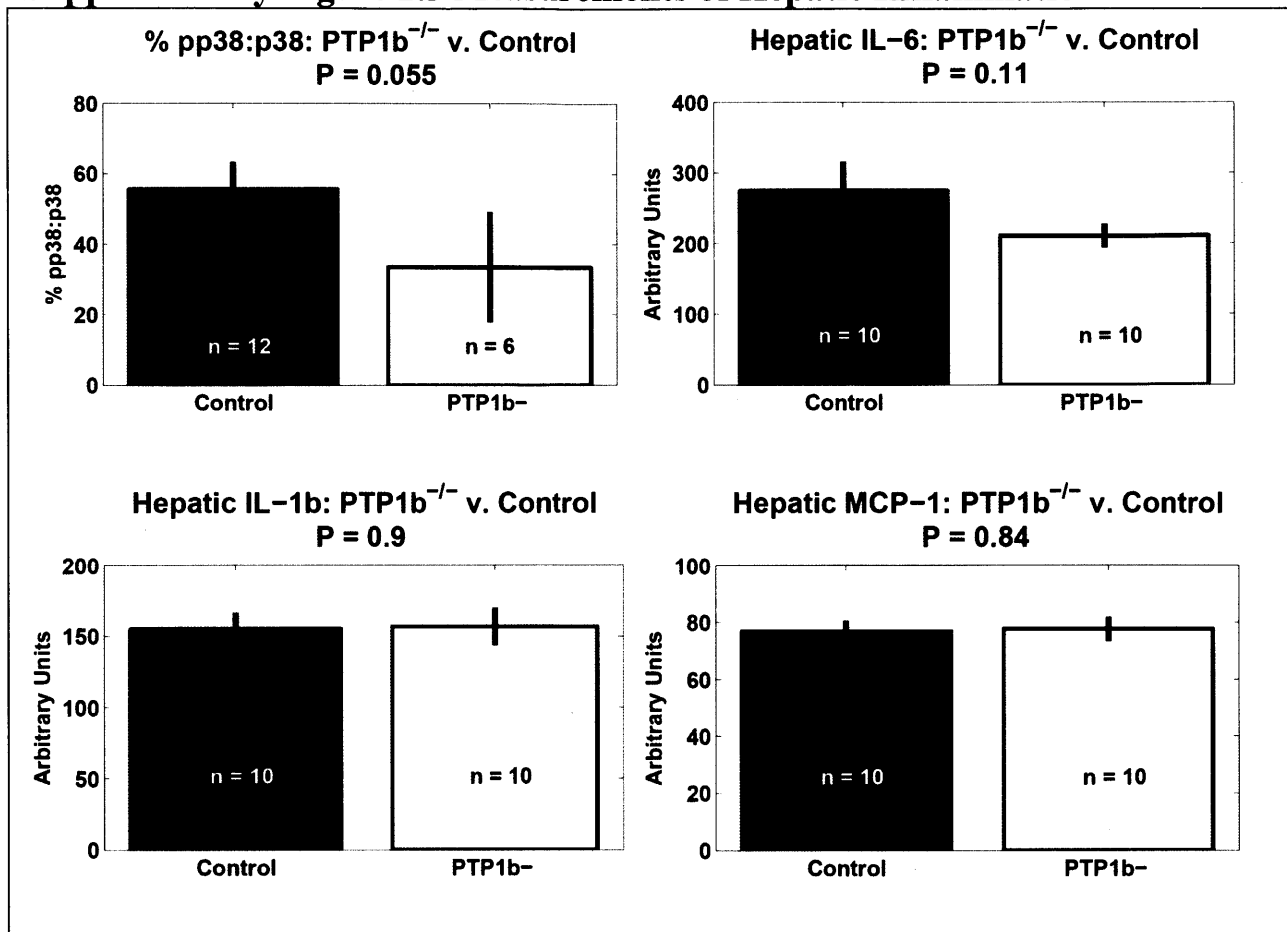
Heatmaps of those phosphorylation sites most significantly correlated to diet. For basal or insulin datasets, the phosphorylation level of a phosphosite was normalized to the corresponding phosphosite mean (basal or insulin) and then log₂-transformed. Missing data points are denoted by grey dots over white boxes. Negatively and positively correlated phosphosites are segregated. Green astericks denote mitochondrial proteins.

Supplementary Figure 2.2 Co-elution to Determine PUFA Species



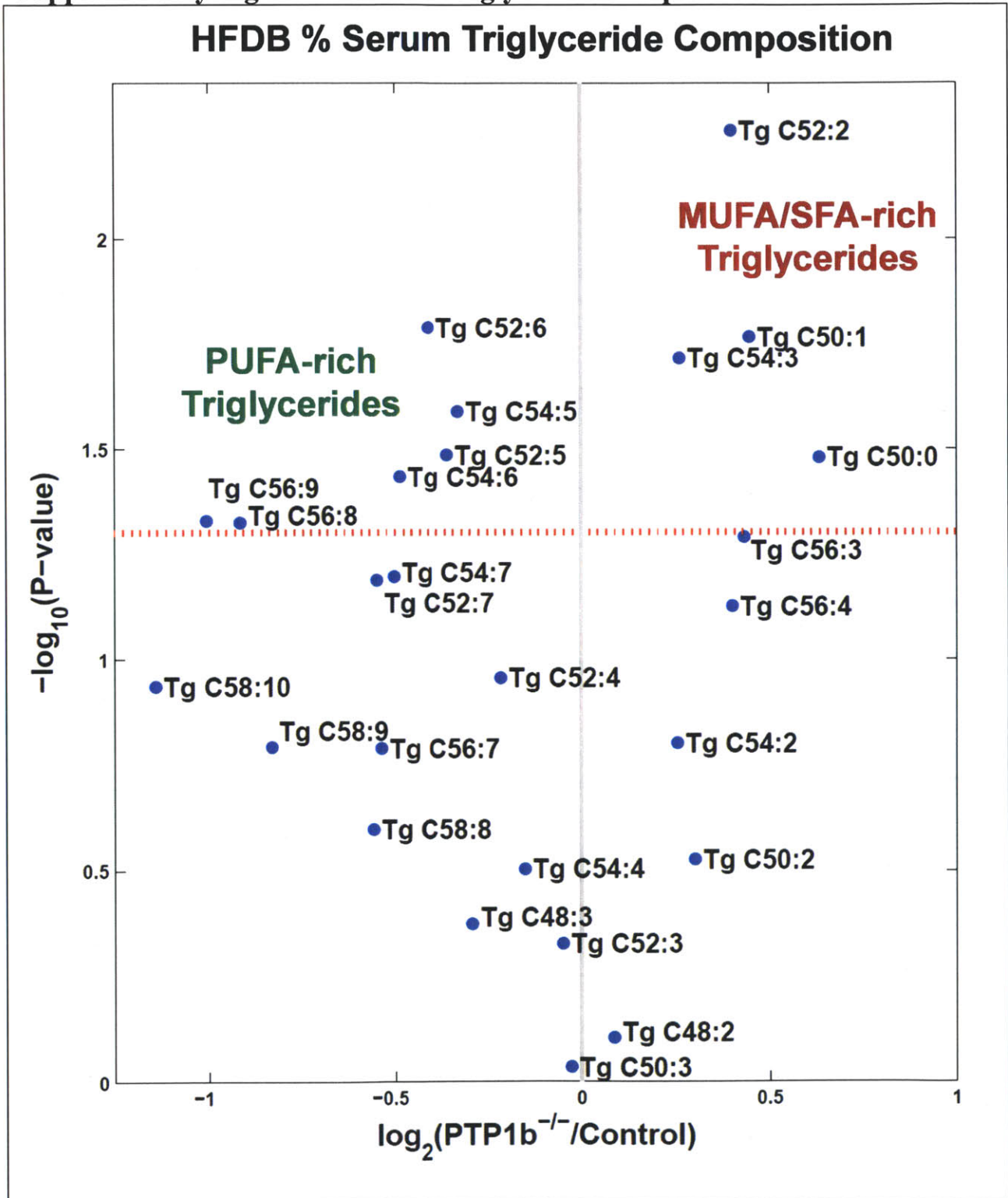
Blue and green curves correspond to ionchromatograms from standards and liver lipid samples analyzed in independent LC-MS experiments. Standards and lipid samples were co-injected (red and purple curves) in two additional experiments to determine the PTP1b-dependent isomer peak indicated by the green arrow. See Methods for detailed discuss of results.

Supplementary Figure 2.3 Measurements of Hepatic Inflammation



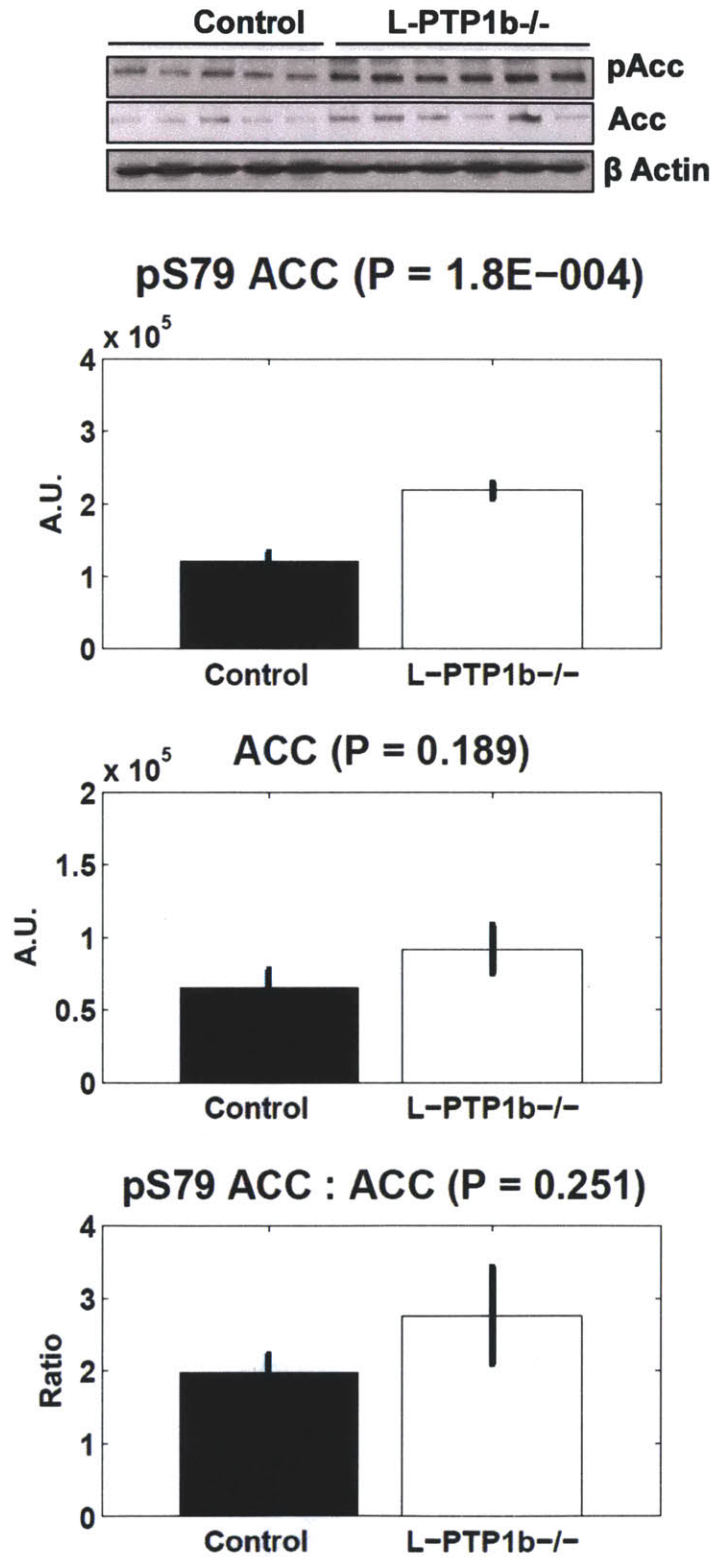
Bar graphs denote the mean and standard error associated with each of the measurements. Two-tailed Student t test was used to calculate p-values.

Supplementary Figure 2.4 Serum Triglyceride Compositions



Volcano plot of percent serum triglyceride compositions on HFD: Each triglyceride is plotted as a function of log₂-transformed, compositional fold-change (L-PTP1b^{-/-} relative to control) and corresponding statistical significance, -log₁₀(P-value), control, n = 10, L-PTP1b^{-/-}, n = 6. The dotted, red line corresponds to P-value = .05.

Supplementary Figure 2.5 Multivariate Models of Steatosis



Western blots and quantification for pS79 ACC and total ACC. Bars and error bars represent mean and standard error, respectively. Two-tailed Student t test was used to calculate p-values.

Tables

Table 1: Animal Weights and Serum Measurements

	NC		HFD	
	Control	L-PTP1b ^{-/-}	Control	L-PTP1b ^{-/-}
Weight (g)	25.4 +/- 1 (n = 13)	24.7 +/- 1.1 (n = 8)	32.8 +/- 1.4 (n = 19)**	30.3 +/- 0.8 (n = 17)
Fed Glucose (mg/dL)	192.1 +/- 6.1 (n = 13)	197.1 +/- 11.4 (n = 8)	236.8 +/- 7.2 (n = 15)**	224 +/- 11 (n = 13)
Fed Insulin (ng/mL)	0.89 +/- 0.13 (n = 7)	0.73 +/- 0.22 (n = 5)	3.47 +/- 0.91 (n = 7)**	2.92 +/- 0.94 (n = 5)
Fasted Glucose (mg/dL)	114 +/- 7.3 (n = 12)	160.1 +/- 10.5 (n = 8)**	119.9 +/- 8.3 (n = 15)	162.4 +/- 15.7 (n = 13)*
Fasted Insulin (pM)	56.4 +/- 10.2 (n = 6)	26.5 +/- 3.4 (n = 5)*	63.3 +/- 13.1 (n = 8)	60.2 +/- 19.7 (n = 6)
Serum Triglycerides (mg/dL)	91.7 +/- 10.9 (n = 12)	78.4 +/- 8.2 (n = 8)	125.9 +/- 15.9 (n = 15)	91.9 +/- 7 (n = 13)^
Serum Cholesterol (mg/dL)	98.8 +/- 6.6 (n = 6)	83.6 +/- 12.5 (n = 5)	144.6 +/- 16.2 (n = 8)	113.9 +/- 25 (n = 7)
Serum HDL (mg/dL)	81.7 +/- 6.4 (n = 6)	69.4 +/- 11.5 (n = 5)	114.6 +/- 12.1 (n = 8)	88.4 +/- 19.4 (n = 7)
Serum LDL (mg/dL)	4.3 +/- 0.8 (n = 6)	3.6 +/- 1.4 (n = 5)	10.6 +/- 1.9 (n = 8)	14.9 +/- 4.3 (n = 7)
Adiponectin (pg/mL)	12.7 +/- 1.4 (n = 13)	11.5 +/- 1.1 (n = 8)	11.9 +/- 0.9 (n = 15)	15 +/- 1.5 (n = 11)^
Glucagon (pM)	25 +/- 7 (n = 7)	18.7 +/- 4.5 (n = 5)	21.9 +/- 4.3 (n = 8)	17.6 +/- 3.4 (n = 7)
Leptin (pM)	54.2 +/- 19.6 (n = 5)	47.1 +/- 27.1 (n = 3)	169 +/- 89.5 (n = 7)**	186.6 +/- 76.4 (n = 6)

Serum measurements and weight gain are written as mean +/- standard error for each of the diet and genotype conditions. Number of mice analyzed per condition is denoted in parentheses. Statistical analysis was performed using two-tailed Student T-tests. L-PTP1b^{-/-} mice were compared to control on respective diets, and HFD control were compared to NC control. ** = P<.01, * = P < .05, ^ = P<.1.

Table 2: Enrichment Analysis of Genotype-Dependent Phosphosites

P _{raw}	NES	Category Label	N	Category Membership	Source
<2E-05**	-3.47	Metabolic process	29	HADHA Y724, CPS1 Y140, ACAT1 Y328, PCBD1 Y70, ARG1 Y218, GAPDH Y328, GLUD1 Y135, CPS1 Y1450, GLUD1 Y512, PGAM2 Y92, ARG1 Y197, CTH Y59, CAR2 Y114, CPS1 Y590, IDHI Y391, GPX1 Y147, GAPDH Y316, PAH Y166*, ARG1 Y188, CPS1 Y162, PGM1 Y353, CTH Y113, LDHA Y239, PRDX6 Y89, ARG1 Y265, TGM2 Y369, CYP2E1 Y426, ENO1 Y44, CYB5A Y11	GO
<2E-05**	-3.58	Oxidation reduction	29	HADHA Y724, UGDH Y108, GAPDH Y328, GLUD1 Y135, GLUD1 Y512, UOX Y230, HGD Y40, ACOX1 Y629, ALDH1L1 Y848, VDACL1	GO, SwissPro,

				<i>Y195, IDH1 Y391, ACAD11 Y323, UOX Y288, GPX1 Y147*, GAPDH Y316, FASN Y1248, HPD Y232, PAH Y166, ALDH6A1 Y268, ALDH1L1 Y892, CAT Y231, CAT Y84, LDHA Y239, PRDX6 Y89, CYP2E1 Y426, HPD Y221, SAHH Y193, ALDH1A7 Y484, CYB5A Y11</i>	Literature
<2E-05**	-3.10	Endoplasmic Reticulum	18	<i>CAV1 Y14 Y6, ALDOB Y204, HSPD1 Y243, PEBP1 Y181, ASS1 Y29, ASS1 Y133, HPD Y232*, ASS1 Y322, HSPD1 Y385, GLUL Y336, CAT Y231, CAT Y84, VCP Y173, ACSL5 Y69, CYP2E1 Y426, CALM1 Y100, HPD Y221, CYB5A Y11,</i>	GO
<2E-05**	-3.66	Cytosol	108	<i>Lyn Y173, Dok1 Y450, BCAR1 Y238, FBP1 Y216, Yes Y220, Fgr Y197, RPL13A Y137, p38-alpha Y182, Crk Y251, afadin Y1285, PIK3R1 Y197, Dok1 Y408, BCAR1 Y291, SHP-2 Y62, BCAR1 Y556, FAK Y577 T575, Dok1 Y361, vinculin Y822, BCAR1 Y271, SEC14L4 Y36, FBP1 Y265, PLCG1 Y771, CAV1 Y14 Y6, eEF1A1 Y177, IRS2 Y649, GRF-1 Y1087, STAT3 Y686, BCAR1 Y414, DLG1 Y783, Yes Y192, FAK Y861, InsR Y1175 Y1179, BCAR1 Y391, GOT1 Y264, vinculin Y692, UGDH Y108, CDK2 Y15, Abl Y393, GSK3B Y216 S219, afadin Y1230, CTNND1 Y904, talin 1 Y1116, BCAR1 Y253, Yes Y424, CTNND1 Y654, PGAM2 Y92, CrkL Y198, PGK1 Y196, ALDOB Y204, LHPP Y159, SCAP2 Y260, FAK Y397, FAK Y576, WASP Y293, talin 1 Y26, CrkL Y132, p38-alpha Y182 T180, Hrs Y216, Lyn Y376, OTC Y317, JAK2 Y570, CAR2 Y114, ERK2 Y185 T188/T183, ERK1 Y205 T203/T208*, SEC14L4 Y36, GSK3B Y216, FBP1 Y245, InsR Y1179/Y1175, plectin 1 Y288, IRS2 Y814, HSPD1 Y243, SHMT1 Y67, IDH1 Y391, CTNND1 Y96, TARS Y297, GRF-1 Y1105, IRS1 Y983, IRS1 Y460, IRS1 Y935, ERK2 Y185, PIK3R1 Y310, GPX1 Y147, GOT1 Y71, HSP90A Y285, IRS2 Y671, SHMT1 Y28, FASN Y1248, HPD Y232, Nck1 Y105, ERK1 Y205, ACLY Y672, HSPD1 Y385, PGM1 Y353, GNMT Y22, CAT Y231, CAT Y84, LDHA Y239, Shc1 Y313, VCP Y173, PRDX6 Y89, GNMT Y34, SHP-2 Y584, PITPNA Y141, GNMT Y221, TGM2 Y369, CALM1 Y100, HPD Y221, SAHH Y193</i>	GO
3.2E-05**	-2.41	Insulin signaling pathway	25	<i>FBP1 Y216, Crk Y251, PIK3R1 Y197, FBP1 Y265, IRS2 Y649, InsR Y1175 Y1179, GSK3B Y216 S219, CrkL Y198, CrkL Y132, ERK2 Y185 T188/T183*, ERK1 Y205 T203/T208, GSK3B Y216, FBP1 Y245, InsR Y1179/Y1175, IRS2 Y814, IRS1 Y983, IRS1 Y460,</i>	Kegg

				IRS1 Y935, ERK2 Y185, PIK3R1 Y310, IRS2 Y671, FASN Y1248, ERK1 Y205, Shc1 Y313, CALM1 Y100	
3.2E-05**	-2.40	Lipid metabolism	11	HADHA Y724, ACAT1 Y328, ACAA2 Y198, ACSM3 Y51, ACOX1 Y629, vigilin Y437*, ACLY Y672, PRDX6 Y89, ACSL5 Y69, ACBP Y29, CYB5A Y11	Kegg, SwissPro, Literature
1.4E-04**	2.45	Tight junction	20	PARD3 Y1076, CLDN2 Y194/Y195, CTNNA1 Y177, PARD3 Y1123, Yes Y220, F11R Y281, ZO1 Y1145, CLDN3 Y197, afadin Y1285*, CTNNA1 Y619, MAG11 Y361, Yes Y192, ZO1 Y1360, CGN Y229, afadin Y1230, Yes Y424, CTNNB1 Y654, ZO2 Y486, ZO1 Y1190, ZO1 Y1066	Kegg
2.3E-04*	-2.30	Fatty acid metabolism	9	HADHA Y724, ACAT1 Y328, ACAA2 Y198, ACOX1 Y629, ACLY Y672*, PRDX6 Y89, ACSL5 Y69, ACBP Y29, CYB5A Y11	Kegg, Literature
4.5E-04*	-2.26	Neurotrophin signaling pathway	25	p38-alpha Y182, Crk Y251, PIK3R1 Y197, SHP-2 Y62, PLCG1 Y771, IRS2 Y649, Abl Y393, GSK3B Y216 S219, CrkL Y198, CrkL Y132, p38-alpha Y182 T180, ERK2 Y185 T188/T183, ERK1 Y205 T203/T208, GSK3B Y216, IRS2 Y814, IRS1 Y983*, IRS1 Y460, IRS1 Y935, ERK2 Y185, PIK3R1 Y310, IRS2 Y671, ERK1 Y205, Shc1 Y313, SHP-2 Y584, CALM1 Y100	Kegg
4.5E-04*	-2.27	Arginine and proline metabolism	19	CPS1 Y140, GOT1 Y264, ARG1 Y218, GLUD1 Y135, CPS1 Y1450, GLUD1 Y512, ARG1 Y197, OTC Y317, GOT2 Y96, CPS1 Y590, ASS1 Y29, GOT1 Y71*, ASS1 Y133, ASS1 Y322, GOT2 Y284, ARG1 Y188, CPS1 Y162, GLUL Y336, ARG1 Y265	Kegg
5.2E-04*	-2.25	Mitochondria	35	Lyn Y173, PC Y118, HADHA Y724, FH Y488, CPS1 Y140, HSPE1 Y76, SLC25A5 Y191, SLC25A5 Y81, ACAT1 Y328, FH Y462, GLUD1 Y135, CPS1 Y1450, GLUD1 Y512, ACAA2 Y198, ACSM3 Y51, Lyn Y376, OTC Y317, NIPSNAP1 Y148, GOT2 Y96, NIPSNAP1 Y261, CPS1 Y590, VDAC1 Y195, HSPD1 Y243, PEBP1 Y181, ASS1 Y29, PHB Y249, ASS1 Y133, ASS1 Y322*, GOT2 Y284, CPS1 Y162, HSPD1 Y385, CAT Y231, CAT Y84, ACSL5 Y69, CYB5A Y11	GO
6.5E-04*	-2.24	Phenylalanine metabolism	8	GOT1 Y264, GOT2 Y96, GOT1 Y71*, HPD Y232, GOT2 Y284, PAH Y166, PRDX6 Y89, HPD Y221	Kegg
7.1E-04*	-2.22	Glycolysis / Gluconeogenesis	12	FBP1 Y216, FBP1 Y265, GAPDH Y328, PGAM2 Y92, PGK1 Y196, ALDOB Y204, FBP1 Y245, GAPDH Y316, PGM1 Y353*, LDHA Y239, ENO1 Y44, ENO1 Y25	Kegg
1.1E-03^	-2.16	Golgi apparatus	17	Lyn Y173, ZO1 Y1145, DYRK4 Y344, CAV1 Y14 Y6, ZO1 Y1360, Lyn	GO

				<i>Y376, ZO1 Y1190, HSPD1 Y243, PEBP1 Y181, ZO1 Y1066, FASN Y1248*</i> , HPD Y232, HSPD1 Y385, GLUL Y336, CAT Y231, CAT Y84, HPD Y221	
1.1E-03 [^]	-2.16	Insulin-sensitive pYs in adipocytes	33	<i>PROSC Y69, Lyn y173, PARD3 Y1076, Dok1 Y450, CLDN2 Y194/Y195, HADHA Y274, FBP1 Y216, FH Y488, TEC Y415, PARD3 Y1123, CPS1 Y140, EGFR Y1172, Yes Y220, RPL13A Y137, p38-alpha Y182, CLTC Y899, Crk Y251, afadin Y1285, tensin 1 Y1558, PRP4 Y849, Shb Y432, Eef1a1 y141, DYRK4 Y344, PIK3R1 Y197, Dok1 Y408, ATP1A1 Y260*</i> , HBD Y42, CAR KD Y81, BCAR1 Y556, FAK Y577 T575, tensin 1 Y1480, PECAM-1 Y702, TENC1 Y705	Schmelzle et al.
1.8E-03 [^]	-2.09	Cysteine and methionine metabolism	9	<i>GOT1 Y264, BHMT Y289, CTH Y59, GOT2 Y96, GOT1 Y71*</i> , GOT2 Y284, CTH Y113, LDHA Y239, SAHH Y193	Kegg

NES stands for normalized enrichment scores as described in Methods. ** = FDR = .01, * = FDR = .05, ^ = FDR = .1. NES > 0 indicates that phosphorylation on sites in pathway significantly decrease with PTP1b deletion, while NES < 0 indicates that phosphorylation on sites in the pathway are hyperphosphorylated upon PTP1b deletion. The phosphosites in each category membership are ordered from most PTP1b-dependent decreased phosphorylation to most PTP1b-dependent increased phosphorylation. Those phosphosites that contributed most to the enrichment (compose the “leading edge”) appear in regular font, while the other sites in that category are italicized.

Table 3: Enrichment Analysis of Diet-Dependent Phosphosites

P_{raw}	NES	Category Label	N	Category Membership	Source
2.2E-04**	2.28	Mitochondria	22	<i>HSPE1 Y76, NIPSNAP1 Y261, SLC25A5 Y191, GOT2 Y96, GLUD1 Y512, CAT Y231*</i> , <i>GLUD1 Y135, PHB Y249, ACAT1 Y328, CAT Y84, PC Y118, CPS1 Y162, Lyn Y173, ACAA2 Y198, HSPD1 Y243, ACSL5 Y69, CPS1 Y1450, CYB5A Y11, ASS1 Y322, HSPD1 Y385, Lyn Y376, SLC25A5 Y81</i>	GO

NES stands for normalized enrichment scores. ** = FDR = .01, * = FDR = .05, ^ = FDR = .1. NES > 0 indicates that phosphorylation on sites in pathway significantly increase upon HFD, while NES < 0 indicates that phosphorylation on sites in the pathway have reduced phosphorylation under HFD. The phosphosites in each category membership are ordered from most positively HFD-dependent to most negatively HFD-dependent phosphorylation. Those phosphosites that contributed most to the enrichment (compose the “leading edge”) appear in regular font, while the other sites in that category are italicized.

Table 4: Quantification of Hepatic FFAs

Absolute quantification of hepatic free fatty acids (pmol FFA / mg liver tissue)				
	NC		HFD	
	Control	L-PTP1b ^{-/-}	Control	L-PTP1b ^{-/-}
C16:3	0.054 +/- 0.0079 (n = 7)	0.063 +/- 0.017 (n = 3)	0.052 +/- 0.0099 (n = 10)	0.044 +/- 0.013 (n = 9)
C16:2	0.14 +/- 0.017 (n = 7)	0.14 +/- 0.026 (n = 3)	0.45 +/- 0.056 (n = 11)	0.64 +/- 0.22 (n = 10)
C16:1	3.4 +/- 0.35 (n = 9)	3.2 +/- 0.43 (n = 7)	8.4 +/- 1.3 (n = 19)*	8.4 +/- 1.4 (n = 17)
C16:0	44 +/- 2.9 (n = 9)	44 +/- 2.1 (n = 7)	77 +/- 16 (n = 19)*	60 +/- 4.8 (n = 17)
C18:3 n-3 + n-6	1.1 +/- 0.18 (n = 7)	1.2 +/- 0.16 (n = 3)	0.45 +/- 0.11 (n = 19)^	0.47 +/- 0.15 (n = 17)
C18:3 n-7*			0.12 +/- 0.014 (n = 19)	0.18 +/- 0.025 (n = 17)
C18:2	14 +/- 2.1 (n = 9)	12 +/- 2.3 (n = 7)	17 +/- 3.4 (n = 19)	15 +/- 1.9 (n = 17)
C18:1	17 +/- 1.5 (n = 9)	18 +/- 2.9 (n = 7)	65 +/- 13 (n = 19)**	58 +/- 6.9 (n = 17)
C18:0	24 +/- 2 (n = 9)	24 +/- 1.9 (n = 7)	37 +/- 5.7 (n = 19)	29 +/- 2.1 (n = 17)
C20:5	2.1 +/- 0.32 (n = 9)	3.2 +/- 1.5 (n = 7)	0.09 +/- 0.014 (n = 19)**	0.087 +/- 0.015 (n = 17)
C20:4	3.3 +/- 0.83 (n = 9)	2.4 +/- 0.53 (n = 7)	4.6 +/- 1.2 (n = 19)^	4.5 +/- 0.68 (n = 17)
C20:3 n-3 + n-6	0.37 +/- 0.051 (n = 7)	0.55 +/- 0.13 (n = 3)	0.43 +/- 0.058 (n = 19)	0.5 +/- 0.049 (n = 17)
C20:3 n-9 + n-7*	0.076 +/- 0.0085 (n = 7)	0.09 +/- 0.013 (n = 3)	0.33 +/- 0.058 (n = 19)**	0.64 +/- 0.077 (n = 17)**
C20:2	0.36 +/- 0.04 (n = 7)	0.48 +/- 0.079 (n = 3)	0.7 +/- 0.15 (n = 19)^	1.1 +/- 0.14 (n = 17)^
C20:1	1.1 +/- 0.16 (n = 9)	1.1 +/- 0.23 (n = 7)	3 +/- 0.64 (n = 19)**	2.9 +/- 0.26 (n = 17)
C20:0	0.77 +/- 0.087 (n = 7)	0.73 +/- 0.095 (n = 3)	1.2 +/- 0.33 (n = 19)	0.6 +/- 0.085 (n = 17)
C22:6	7.3 +/- 1.3 (n = 9)	5.5 +/- 1.1 (n = 7)	1.4 +/- 0.24 (n = 19)**	1.2 +/- 0.14 (n = 17)
C22:5	0.91 +/- 0.12 (n = 7)	1.3 +/- 0.27 (n = 3)	0.43 +/- 0.073 (n = 19)**	0.46 +/- 0.032 (n = 17)
C22:4	0.23 +/- 0.024 (n = 7)	0.31 +/- 0.082 (n = 3)	0.28 +/- 0.043 (n = 19)	0.31 +/- 0.03 (n = 17)
C22:3 n-3 + n-6*	0.011 +/- 0.0048 (n = 6)	0.023 +/- 0.0058 (n = 3)	0.018 +/- 0.003 (n = 18)	0.026 +/- 0.0029 (n = 17)^
C22:3 n-9* + n-7*			0.049 +/- 0.0095 (n = 19)**	0.089 +/- 0.01 (n = 17)**
C22:2	0.036 +/- 0.0053 (n = 7)	0.044 +/- 0.0047 (n = 3)	0.059 +/- 0.021 (n = 11)	0.065 +/- 0.0065 (n = 10)
C22:1	0.24 +/- 0.034 (n = 7)	0.27 +/- 0.02 (n = 3)	0.41 +/- 0.12 (n = 19)	0.32 +/- 0.035 (n = 17)
C22:0	0.45 +/- 0.039 (n = 7)	0.41 +/- 0.025 (n = 3)	0.71 +/- 0.26 (n = 19)	0.36 +/- 0.056 (n = 17)
C24:6	0.11 +/- 0.015 (n = 7)	0.15 +/- 0.034 (n = 3)	0.027 +/- 0.0042 (n = 19)**	0.038 +/- 0.0028 (n = 17)*
C24:5	0.083 +/- 0.013 (n = 7)	0.12 +/- 0.041 (n = 3)	0.05 +/- 0.007 (n = 19)^	0.076 +/- 0.0069 (n = 17)*
C24:4	0.015 +/- 0.0044 (n = 6)	0.026 +/- 0.011 (n = 3)	0.047 +/- 0.0076 (n = 11)*	0.074 +/- 0.011 (n = 10)*
C24:3			0.013 +/- 0.0072 (n = 3)	0.013 +/- 0.0045 (n = 4)

C24:2	0.074 +/- 0.0041 (n = 7)	0.077 +/- 0.01 (n = 3)	0.049 +/- 0.012 (n = 11)**	0.044 +/- 0.0043 (n = 10)
C24:1	0.42 +/- 0.044 (n = 7)	0.4 +/- 0.045 (n = 3)	0.34 +/- 0.099 (n = 11)**	0.23 +/- 0.025 (n = 10)
C24:0	0.47 +/- 0.049 (n = 7)	0.48 +/- 0.064 (n = 3)	0.56 +/- 0.26 (n = 11)	0.51 +/- 0.26 (n = 10)
Total	130 +/- 6 (n = 7)	137 +/- 17 (n = 3)	218 +/- 40 (n = 19)	183 +/- 17 (n = 17)

% Composition of hepatic free fatty acids pool				
	NC		HFD	
	Control	L-PTP1b^{-/-}	Control	L-PTP1b^{-/-}
C16:3	0.042 +/- 0.0044 (n = 7)	0.046 +/- 0.009 (n = 3)	0.02 +/- 0.003 (n = 10)**	0.018 +/- 0.002 (n = 9)
C16:2	0.11 +/- 0.0075 (n = 7)	0.098 +/- 0.0088 (n = 3)	0.19 +/- 0.01 (n = 11)*	0.25 +/- 0.04 (n = 10)
C16:1	2.9 +/- 0.19 (n = 7)	3.1 +/- 0.093 (n = 3)	3.9 +/- 0.19 (n = 19)*	4 +/- 0.2 (n = 17)
C16:0	41 +/- 4.1 (n = 7)	34 +/- 1.2 (n = 3)	37 +/- 1 (n = 19)	35 +/- 1 (n = 17)
C18:3 n-3 + n-6	0.84 +/- 0.077 (n = 7)	0.82 +/- 0.058 (n = 3)	0.24 +/- 0.01 (n = 19)*	0.33 +/- 0.08 (n = 17)
C18:3 n-7*			0.055 +/- 0.004 (n = 19)**	0.085 +/- 0.005 (n = 17)**
C18:2	12 +/- 1.2 (n = 7)	13 +/- 1.3 (n = 3)	7.3 +/- 0.4 (n = 19)**	7 +/- 0.3 (n = 17)
C18:1	15 +/- 0.86 (n = 7)	18 +/- 0.66 (n = 3)	28 +/- 2 (n = 19)**	29 +/- 1 (n = 17)
C18:0	23 +/- 3 (n = 7)	19 +/- 2.1 (n = 3)	21 +/- 1 (n = 19)	19 +/- 0.9 (n = 17)
C20:5	2 +/- 0.21 (n = 7)	1.9 +/- 0.22 (n = 3)	0.044 +/- 0.004 (n = 19)**	0.042 +/- 0.004 (n = 17)
C20:4	3 +/- 0.56 (n = 7)	2.6 +/- 0.51 (n = 3)	2 +/- 0.2 (n = 19)	2.3 +/- 0.3 (n = 17)
C20:3 n-3 + n-6	0.31 +/- 0.032 (n = 7)	0.4 +/- 0.077 (n = 3)	0.21 +/- 0.01 (n = 19)**	0.26 +/- 0.009 (n = 17)**
C20:3 n-9 + n-7*	0.063 +/- 0.0041 (n = 7)	0.066 +/- 0.0027 (n = 3)	0.16 +/- 0.01 (n = 19)**	0.33 +/- 0.03 (n = 17)**
C20:2	0.3 +/- 0.034 (n = 7)	0.35 +/- 0.021 (n = 3)	0.31 +/- 0.03 (n = 19)	0.54 +/- 0.06 (n = 17)**
C20:1	0.96 +/- 0.13 (n = 7)	1.2 +/- 0.078 (n = 3)	1.2 +/- 0.1 (n = 19)	1.5 +/- 0.1 (n = 17)
C20:0	0.66 +/- 0.09 (n = 7)	0.55 +/- 0.087 (n = 3)	0.46 +/- 0.08 (n = 19)**	0.31 +/- 0.03 (n = 17)
C22:6	6.6 +/- 0.75 (n = 7)	5.8 +/- 0.93 (n = 3)	0.65 +/- 0.06 (n = 19)**	0.65 +/- 0.08 (n = 17)
C22:5	0.77 +/- 0.1 (n = 7)	0.91 +/- 0.18 (n = 3)	0.21 +/- 0.02 (n = 19)**	0.25 +/- 0.02 (n = 17)
C22:4	0.19 +/- 0.023 (n = 7)	0.23 +/- 0.054 (n = 3)	0.14 +/- 0.014 (n = 19)^	0.16 +/- 0.01 (n = 17)
C22:3 n-3 + n-6*	0.014 +/- 0.0074 (n = 5)	0.017 +/- 0.003 (n = 3)	0.0083 +/- 0.001 (n = 18)	0.014 +/- 0.002 (n = 17)**
C22:3 n-9* + n-7*			0.025 +/- 0.003 (n = 19)**	0.052 +/- 0.005 (n = 17)**
C22:2	0.032 +/- 0.0066 (n = 7)	0.033 +/- 0.00092 (n = 3)	0.023 +/- 0.006 (n = 11)	0.033 +/- 0.004 (n = 10)
C22:1	0.2 +/- 0.034 (n = 7)	0.2 +/- 0.027 (n = 3)	0.18 +/- 0.04 (n = 19)	0.17 +/- 0.01 (n = 17)
C22:0	0.38 +/- 0.034 (n = 7)	0.31 +/- 0.056 (n = 3)	0.31 +/- 0.08 (n = 19)*	0.19 +/- 0.02 (n = 17)

C24:6	0.084 +/- 0.0099 (n = 7)	0.11 +/- 0.021 (n = 3)	0.013 +/- 0.002 (n = 19)**	0.021 +/- 0.002 (n = 17)**
C24:5	0.069 +/- 0.012 (n = 7)	0.085 +/- 0.025 (n = 3)	0.025 +/- 0.003 (n = 19)**	0.042 +/- 0.005 (n = 17)**
C24:4	0.013 +/- 0.0046 (n = 6)	0.019 +/- 0.007 (n = 3)	0.02 +/- 0.002 (n = 11)^	0.035 +/- 0.004 (n = 10)**
C24:3			0.0035 +/- 0.002 (n = 3)	0.0079 +/- 0.003 (n = 4)
C24:2	0.063 +/- 0.0054 (n = 7)	0.058 +/- 0.0061 (n = 3)	0.021 +/- 0.004 (n = 11)**	0.022 +/- 0.002 (n = 10)
C24:1	0.36 +/- 0.048 (n = 7)	0.3 +/- 0.027 (n = 3)	0.14 +/- 0.03 (n = 11)**	0.12 +/- 0.014 (n = 10)
C24:0	0.4 +/- 0.044 (n = 7)	0.37 +/- 0.097 (n = 3)	0.23 +/- 0.09 (n = 11)^	0.2 +/- 0.05 (n = 10)

Quantification of liver free fatty acids are written as mean +/- standard error for each of the diet and genotype conditions. Number of mice analyzed per condition is denoted in parentheses. Statistical analysis was performed using two-tailed Student T-tests. L-PTP1b^{-/-} were compared to control on respective diet, and HFD control were compared to NC control. ** = $P_{\text{raw}} < .01$, * = $P_{\text{raw}} < .05$, ^ = $P_{\text{raw}} < .1$. An asterick by fatty acid species indicates that the fatty acid measurement likely corresponds to that species, but its identification was not rigorously confirmed.

Table 5: Quantification of Hepatic Triglycerides and Cholesterol

Absolute quantification of triglycerides and cholesterol (nmol lipid / mg liver tissue)				
	NC		HFD	
	Control	L-PTP1b ^{-/-}	Control	L-PTP1b ^{-/-}
Tg C48:3	0.13 +/- 0.044 (n = 7)	0.11 +/- 0.048 (n = 3)	0.085 +/- 0.024 (n = 10)	0.12 +/- 0.022 (n = 7)
Tg C48:2	0.31 +/- 0.11 (n = 7)	0.25 +/- 0.12 (n = 3)	0.17 +/- 0.048 (n = 10)	0.36 +/- 0.077 (n = 7)*
Tg C48:1			0.095 +/- 0.024 (n = 6)	0.37 +/- 0.061 (n = 5)**
Tg C50:4	0.38 +/- 0.12 (n = 7)	0.4 +/- 0.19 (n = 3)	0.26 +/- 0.076 (n = 10)	0.56 +/- 0.12 (n = 7)*
Tg C50:3	1.8 +/- 0.6 (n = 7)	2.2 +/- 1.1 (n = 3)	1.2 +/- 0.35 (n = 10)	3 +/- 0.65 (n = 7)*
Tg C50:2	3.2 +/- 1.1 (n = 7)	3.7 +/- 1.7 (n = 3)	2 +/- 0.64 (n = 10)	6 +/- 1.3 (n = 7)**
Tg C50:1	0.96 +/- 0.29 (n = 7)	1.1 +/- 0.44 (n = 3)	0.76 +/- 0.24 (n = 10)	2.3 +/- 0.48 (n = 7)**
Tg C52:5	1.7 +/- 0.51 (n = 7)	0.95 +/- 0.58 (n = 3)	0.68 +/- 0.19 (n = 10)	1.3 +/- 0.24 (n = 7)^
Tg C52:4	11 +/- 3 (n = 7)	14 +/- 5.7 (n = 3)	4.8 +/- 1.4 (n = 10)	9.6 +/- 1.8 (n = 7)*
Tg C52:3	17 +/- 4.1 (n = 7)	25 +/- 9.6 (n = 3)	14 +/- 4 (n = 10)	34 +/- 7.3 (n = 6)*
Tg C52:2	5.5 +/- 1.2 (n = 7)	9.5 +/- 3.3 (n = 3)	11 +/- 3.8 (n = 10)^	36 +/- 8.6 (n = 6)**
Tg C54:7	1.3 +/- 0.35 (n = 7)	1.4 +/- 0.52 (n = 3)	0.12 +/- 0.043 (n = 4)**	0.086 +/- 0.016 (n = 6)
Tg C54:6	2.7 +/- 0.72 (n = 7)	3 +/- 1.5 (n = 3)	0.63 +/- 0.16 (n = 10)*	0.67 +/- 0.13 (n = 7)
Tg C54:5	3.9 +/- 1 (n = 7)	5 +/- 2.4 (n = 3)	2.6 +/- 0.59 (n = 10)	3.7 +/- 0.69 (n = 7)
Tg C54:4	3.6 +/- 0.9 (n = 7)	5.1 +/- 2.2 (n = 3)	5.8 +/- 1.4 (n = 10)	11 +/- 2.1 (n = 7)*

Tg C54:3	1.7 +/- 0.34 (n = 7)	2.4 +/- 0.9 (n = 3)	5.8 +/- 1.5 (n = 10)*	15 +/- 3.2 (n = 7)*
Tg C54:2	0.25 +/- 0.042 (n = 7)	0.31 +/- 0.087 (n = 3)	0.77 +/- 0.31 (n = 10)*	2.2 +/- 0.54 (n = 7)*
Tg C56:10	0.04 +/- 0.011 (n = 7)	0.037 +/- 0.019 (n = 3)		
Tg C56:9	0.44 +/- 0.12 (n = 7)	0.48 +/- 0.21 (n = 3)	0.018 +/- 0.0046 (n = 8)**	0.012 +/- 0.002 (n = 6)
Tg C56:8	4.5 +/- 1.1 (n = 7)	5.3 +/- 2.2 (n = 3)	0.12 +/- 0.028 (n = 10)**	0.077 +/- 0.015 (n = 7)
Tg C56:7	3.3 +/- 0.7 (n = 7)	4.5 +/- 1.9 (n = 3)	0.39 +/- 0.11 (n = 10)**	0.33 +/- 0.056 (n = 7)
Tg C56:6	0.63 +/- 0.14 (n = 7)	0.83 +/- 0.35 (n = 3)	0.41 +/- 0.13 (n = 10)	0.55 +/- 0.099 (n = 7)
Tg C56:5	0.27 +/- 0.049 (n = 7)	0.37 +/- 0.15 (n = 3)	0.38 +/- 0.11 (n = 10)	0.81 +/- 0.16 (n = 7)*
Tg C56:4	0.18 +/- 0.033 (n = 7)	0.24 +/- 0.096 (n = 3)	0.36 +/- 0.14 (n = 10)^	1.1 +/- 0.26 (n = 7)**
Tg C56:3			0.17 +/- 0.04 (n = 6)	1.4 +/- 0.38 (n = 5)**
Tg C56:2			0.031 +/- 0.012 (n = 6)	0.095 +/- 0.032 (n = 5)*
Total Triglycerides	62 +/- 16 (n = 7)	84 +/- 33 (n = 3)	58 +/- 7.5 (n = 19)	110 +/- 15 (n = 16)**
Total Cholesterol	3 +/- 0.19 (n = 7)	3 +/- 0.29 (n = 3)	5.6 +/- 0.16 (n = 10)**	5.5 +/- 0.2 (n = 7)
% Liver triglyceride composition				
	NC		HFD	
	Control	L-PTP1b^{-/-}	Control	L-PTP1b^{-/-}
Tg C48:3	0.2 +/- 0.017 (n = 7)	0.12 +/- 0.025 (n = 3)*	0.17 +/- 0.019 (n = 10)*	0.11 +/- 0.017 (n = 6)
Tg C48:2	0.47 +/- 0.042 (n = 7)	0.29 +/- 0.04 (n = 3)*	0.34 +/- 0.035 (n = 10)	0.3 +/- 0.034 (n = 6)
Tg C48:1			0.34 +/- 0.11 (n = 6)	0.23 +/- 0.023 (n = 4)
Tg C50:4	0.58 +/- 0.035 (n = 7)	0.45 +/- 0.057 (n = 3)*	0.48 +/- 0.037 (n = 10)	0.47 +/- 0.05 (n = 6)
Tg C50:3	2.6 +/- 0.21 (n = 7)	2.5 +/- 0.27 (n = 3)	2.2 +/- 0.2 (n = 10)	2.4 +/- 0.16 (n = 6)
Tg C50:2	4.9 +/- 0.37 (n = 7)	4.3 +/- 0.32 (n = 3)	3.6 +/- 0.27 (n = 10)**	4.8 +/- 0.19 (n = 6)
Tg C50:1	1.6 +/- 0.14 (n = 7)	1.3 +/- 0.055 (n = 3)	1.6 +/- 0.2 (n = 10)	1.9 +/- 0.15 (n = 6)
Tg C52:5	2.6 +/- 0.18 (n = 7)	1.5 +/- 0.73 (n = 3)*	1.4 +/- 0.063 (n = 10)*	1.1 +/- 0.093 (n = 6)
Tg C52:4	17 +/- 0.47 (n = 7)	17 +/- 0.23 (n = 3)	9.1 +/- 0.52 (n = 10)^	7.6 +/- 0.46 (n = 6)**
Tg C52:3	27 +/- 0.89 (n = 7)	30 +/- 0.31 (n = 3)^	26 +/- 1.5 (n = 10)	27 +/- 0.71 (n = 6)**
Tg C52:2	9.3 +/- 0.55 (n = 7)	12 +/- 0.84 (n = 3)*	21 +/- 1.2 (n = 10)**	28 +/- 1.1 (n = 6)**
Tg C54:7	1.9 +/- 0.13 (n = 7)	1.7 +/- 0.23 (n = 3)	0.17 +/- 0.021 (n = 4)**	0.077 +/- 0.016 (n = 6)**
Tg C54:6	4.2 +/- 0.15 (n = 7)	3.4 +/- 0.42 (n = 3)*	1.4 +/- 0.23 (n = 10)*	0.57 +/- 0.057 (n = 6)**
Tg C54:5	6.2 +/- 0.22 (n = 7)	5.6 +/- 0.8 (n = 3)	5.5 +/- 0.65 (n = 10)**	3 +/- 0.2 (n = 6)**
Tg C54:4	5.9 +/- 0.22 (n = 7)	5.9 +/- 0.72 (n = 3)	13 +/- 1 (n = 10)*	8.7 +/- 0.22 (n = 6)**
Tg C54:3	2.9 +/- 0.22 (n = 7)	2.9 +/- 0.31 (n = 3)	12 +/- 1.4 (n = 10)	12 +/- 0.55 (n = 6)**

Tg C54:2	0.47 +/- 0.078 (n = 7)	0.41 +/- 0.087 (n = 3)	1.5 +/- 0.17 (n = 10)	1.9 +/- 0.31 (n = 6)*
Tg C56:10	0.065 +/- 0.0071 (n = 7)	0.043 +/- 0.006 (n = 3)		
Tg C56:9	0.72 +/- 0.065 (n = 7)	0.56 +/- 0.052 (n = 3)	0.036 +/- 0.0031 (n = 8)**	0.011 +/- 0.0025 (n = 5)**
Tg C56:8	7.4 +/- 0.66 (n = 7)	6.6 +/- 1.1 (n = 3)	0.25 +/- 0.015 (n = 10)**	0.07 +/- 0.012 (n = 6)**
Tg C56:7	5.5 +/- 0.32 (n = 7)	5.5 +/- 0.52 (n = 3)	0.82 +/- 0.065 (n = 10)**	0.3 +/- 0.057 (n = 6)**
Tg C56:6	1.1 +/- 0.078 (n = 7)	0.98 +/- 0.031 (n = 3)	0.8 +/- 0.034 (n = 10)**	0.49 +/- 0.075 (n = 6)**
Tg C56:5	0.48 +/- 0.055 (n = 7)	0.44 +/- 0.045 (n = 3)	0.76 +/- 0.052 (n = 10)	0.69 +/- 0.069 (n = 6)*
Tg C56:4	0.34 +/- 0.045 (n = 7)	0.28 +/- 0.059 (n = 3)	0.66 +/- 0.047 (n = 10)*	0.95 +/- 0.13 (n = 6)**
Tg C56:3			0.49 +/- 0.056 (n = 6)*	0.88 +/- 0.16 (n = 4)
Tg C56:2			0.09 +/- 0.023 (n = 6)	0.061 +/- 0.015 (n = 4)

Quantification of liver triglycerides and cholesterol is written as mean +/- standard error for each of the diet and genotype conditions. Number of mice analyzed per condition is denoted in parentheses. Statistical analysis was performed using two-tailed Student's T-tests. L-PTP1b^{-/-} mice were compared to control on respective diet, and HFD control were compared to NC control. ** = $P_{\text{raw}} < .01$, * = $P_{\text{raw}} < .05$, ^ = $P_{\text{raw}} < .1$.

Table 6: Percent Compositions of Serum Triglycerides

HFD Serum Triglyceride Compositions				
	Absolute Quantification (mg/dL)		Percent Composition	
	Control	L-PTP1b^{-/-}	Control	L-PTP1b^{-/-}
Tg C48:3	0.19 +/- 0.09 (n = 7)	0.18 +/- 0.17 (n = 6)	0.39 +/- 0.13 (n = 7)	0.32 +/- 0.18 (n = 6)
Tg C48:2	0.26 +/- 0.14 (n = 7)	0.3 +/- 0.25 (n = 6)	0.53 +/- 0.19 (n = 7)	0.56 +/- 0.24 (n = 6)
Tg C50:3	0.71 +/- 0.32 (n = 7)	0.75 +/- 0.61 (n = 6)	1.47 +/- 0.46 (n = 7)	1.44 +/- 0.54 (n = 6)
Tg C50:2	1.89 +/- 1.04 (n = 7)	2.45 +/- 1.81 (n = 6)	3.89 +/- 1.54 (n = 7)	4.8 +/- 1.43 (n = 6)
Tg C50:1	1.81 +/- 0.91 (n = 7)	2.31 +/- 1.29 (n = 6)	3.68 +/- 1.14 (n = 7)	5.01 +/- 0.27 (n = 6)*
Tg C50:0	0.27 +/- 0.14 (n = 7)	0.33 +/- 0.12 (n = 6)	0.54 +/- 0.15 (n = 7)	0.83 +/- 0.28 (n = 6)*
Tg C52:7	0.09 +/- 0.03 (n = 7)	0.06 +/- 0.03 (n = 6)	0.19 +/- 0.07 (n = 7)	0.13 +/- 0.02 (n = 6)^
Tg C52:6	0.4 +/- 0.08 (n = 7)	0.31 +/- 0.18 (n = 6)	0.87 +/- 0.17 (n = 7)	0.65 +/- 0.07 (n = 6)*
Tg C52:5	2.16 +/- 0.57 (n = 7)	1.79 +/- 1.25 (n = 6)	4.55 +/- 0.53 (n = 7)	3.56 +/- 0.92 (n = 6)*
Tg C52:4	8.4 +/- 2.42 (n = 7)	7.42 +/- 4.66 (n = 6)	17.67 +/- 2.36 (n = 7)	15.23 +/- 2.72 (n = 6)
Tg C52:3	8.26 +/- 2.38 (n = 7)	7.54 +/- 3.87 (n = 6)	17.3 +/- 1.55 (n = 7)	16.72 +/- 1.17 (n = 6)
Tg C52:2	6.92 +/- 2.62 (n = 7)	8.06 +/- 3.45 (n = 6)	14.2 +/- 1.94 (n = 7)	18.73 +/- 2.79 (n = 6)**
Tg C54:7	0.42 +/- 0.13 (n = 7)	0.29 +/- 0.16 (n = 6)	0.91 +/- 0.31 (n = 7)	0.64 +/- 0.08 (n = 6)^

Tg C54:6	1.6 +/- 0.48 (n = 7)	1.12 +/- 0.62 (n = 6)	3.46 +/- 0.98 (n = 7)	2.47 +/- 0.28 (n = 6)*
Tg C54:5	3.3 +/- 0.7 (n = 7)	2.6 +/- 1.44 (n = 6)	7.08 +/- 1.3 (n = 7)	5.64 +/- 0.44 (n = 6)*
Tg C54:4	2.5 +/- 0.65 (n = 7)	2.16 +/- 1.17 (n = 6)	5.34 +/- 1.1 (n = 7)	4.81 +/- 0.59 (n = 6)
Tg C54:3	4.29 +/- 1.06 (n = 7)	4.73 +/- 2.15 (n = 6)	9.05 +/- 1.17 (n = 7)	10.86 +/- 1.2 (n = 6)*
Tg C54:2	0.9 +/- 0.47 (n = 7)	0.94 +/- 0.44 (n = 6)	1.8 +/- 0.48 (n = 7)	2.15 +/- 0.32 (n = 6)
Tg C56:9	0.14 +/- 0.07 (n = 7)	0.07 +/- 0.04 (n = 6)*	0.32 +/- 0.17 (n = 7)	0.16 +/- 0.03 (n = 6)*
Tg C56:8	0.64 +/- 0.31 (n = 7)	0.32 +/- 0.17 (n = 6)^	1.4 +/- 0.7 (n = 7)	0.74 +/- 0.16 (n = 6)*
Tg C56:7	1.12 +/- 0.55 (n = 7)	0.7 +/- 0.3 (n = 6)	2.43 +/- 1.17 (n = 7)	1.68 +/- 0.39 (n = 6)
Tg C56:4	0.38 +/- 0.11 (n = 7)	0.42 +/- 0.14 (n = 6)	0.8 +/- 0.09 (n = 7)	1.06 +/- 0.33 (n = 6)^
Tg C56:3	0.37 +/- 0.17 (n = 7)	0.42 +/- 0.16 (n = 6)	0.75 +/- 0.18 (n = 7)	1.02 +/- 0.26 (n = 6)^
Tg C58:10	0.16 +/- 0.13 (n = 7)	0.06 +/- 0.02 (n = 6)^	0.36 +/- 0.27 (n = 7)	0.16 +/- 0.05 (n = 6)
Tg C58:9	0.25 +/- 0.17 (n = 7)	0.12 +/- 0.04 (n = 6)^	0.56 +/- 0.38 (n = 7)	0.32 +/- 0.13 (n = 6)
Tg C58:8	0.21 +/- 0.14 (n = 7)	0.12 +/- 0.03 (n = 6)	0.47 +/- 0.28 (n = 7)	0.32 +/- 0.12 (n = 6)
Total	47.64 +/- 12.19 (n = 7)	45.55 +/- 24.01 (n = 6)		

Quantification of liver triglycerides and cholesterol is written as mean +/- standard error for each of the diet and genotype conditions. Number of mice analyzed per condition is denoted in parentheses. Statistical analysis was performed using two-tailed Student's T-tests. L-PTP1b^{-/-} mice were compared to control on respective diet, and HFD control were compared to NC control. ** = $P_{\text{raw}} < .01$, * = $P_{\text{raw}} < .05$, ^ = $P_{\text{raw}} < .1$.

Table 7: Phosphosites Sets Predictive of $\omega 3$ / $\omega 6$ PTP1b PUFA

Set enrichments for phosphosites predictive of the PTP1b-dependent $\omega 3$ / $\omega 6$ PUFA (C24:6, C24:5, C24:4, C22:3, C20:3)

P_{raw}	NES	Category Label	N	Category Membership	Source
1.1E-04**	2.44	Metabolic process	25	CYP2E1 Y426, CYB5A Y11, CAR2 Y114, LDHA Y239, PGM1 Y353, GAPDH Y316, PRDX6 Y89, CTH Y113, CPS1 Y162, TGM2 Y369*, ARG1 Y188, GPX1 Y147, ARG1 Y197, PCBD1 Y70, ENO1 Y44, GAPDH Y328, ACAT1 Y328, HADHA Y724, PGAM2 Y92, CPS1 Y140, GLUD1 Y135, CPS1 Y590, GLUD1 Y512, ARG1 Y265, CPS1 Y1450	GO
1.1E-04**	2.43	Oxidation reduction	24	CYP2E1 Y426, CYB5A Y11, SAHH Y193, HPD Y232, LDHA Y239, HPD Y221, HGD Y40, GAPDH Y316, PRDX6 Y89, FASN Y1248, ALDH1A7 Y484*, UOX Y288, GPX1 Y147, ALDH1L1 Y848, CAT Y231, ALDH6A1 Y268, UGDH Y108, GAPDH Y328, CAT Y84, UOX Y230, HADHA Y724, GLUD1 Y135, GLUD1 Y512, ACAD11 Y323	GO, SwissPro, Literature

1.4E-03 [^]	2.13	Endoplasmic Reticulum	16	CYP2E1 Y426, CALM1 Y100, GLUL Y336, CYB5A Y11, PEBP1 Y181, HPD Y232*, <i>HPD Y221, VCP Y173, ASS1 Y133, CAV1 Y14 Y6, CAT Y231, ALDOB Y204, ACSL5 Y69, CAT Y84, HSPD1 Y385, ASS1 Y322,</i>	GO
----------------------	------	------------------------------	----	---	----

NES stands for normalized enrichment scores. ** = FDR = .01, * = FDR = .05, ^ = FDR = .1. NES > 0 indicates that increased phosphorylation on sites in the pathway are positively predictive of proportion of $\omega 3 / \omega 6$ PUFA, while NES < 0 indicates that increased phosphorylation on sites in the pathway are negatively predictive. The phosphosites in each category membership are ordered from most positively to most negatively predictive. Those phosphosites that contributed most to the enrichment (compose the “leading edge”) appear in regular font, while the other sites in that category are italicized.

Table 8: Phosphosite Sets Predictive of $\omega 7 / \omega 9$ PTP1b PUFA

Set enrichments for phosphosites predictive of PTP1b-dependent $\omega 7 / \omega 9$ PUFA (C20:3, C22:3, C20:2, C18:3)

P _{raw}	NES	Category Label	N	Category Membership	Source
<2E-05**	2.53	Oxidation reduction	24	CYB5A Y11, ALDH1A7 Y484, CYP2E1 Y426, GAPDH Y316, SAHH Y193, FASN Y1248, HPD Y221, LDHA Y239, HPD Y232, PRDX6 Y89, CAT Y231*, <i>GPX1 Y147, HGD Y40, UOX Y288, ALDH1L1 Y848, ALDH6A1 Y268, GLUD1 Y135, CAT Y84, GAPDH Y328, ACAD11 Y323, UGDH Y108, UOX Y230, HADHA Y724, GLUD1 Y512</i>	GO, SwissPro, Literature
7.4E-04*	2.31	Metabolic process	25	CYB5A Y11, CYP2E1 Y426, GAPDH Y316, TGM2 Y369, CAR2 Y114, LDHA Y239, PGM1 Y353, PRDX6 Y89, CPS1 Y162, ARG1 Y188, GPX1 Y147, ENO1 Y44*, <i>GLUD1 Y135, CTH Y113, PCBD1 Y70, GAPDH Y328, PGAM2 Y92, CPS1 Y590, HADHA Y724, ARG1 Y197, ACAT1 Y328, ARG1 Y265, GLUD1 Y512, CPS1 Y140, CPS1 Y1450</i>	GO
1.1E-03 [^]	2.25	Cytosol	88	ACLY Y672, SAHH Y193, SHP-2 Y584, FASN Y1248, HPD Y221, TGM2 Y369, TARS Y297, CALM1 Y100, CAR2 Y114, LDHA Y239, PGM1 Y353, HPD Y232, PRDX6 Y89, VCP Y173, CAT Y231, BCAR1 Y271, PITPNA Y141, GRF-1 Y1105*, <i>OTC Y317, GNMT Y221, CrkL Y198, GPX1 Y147, Hrs Y216, PLCG1 Y771, LHPP Y159, GSK3B Y216 S219, ALDOB Y204, p38-alpha Y182 T180, ERK2 Y185 T188/T183, ERK1 Y205, vinculin Y822, PGK1 Y196, ERK2 Y185, IRS2 Y671, BCAR1 Y391, CAV1 Y14 Y6, CAT Y84, GSK3B Y216, CTNND1 Y904, vinculin Y692, Shc1 Y313, Abl Y393, SHMT1 Y28, BCAR1 Y238, JAK2 Y570, CTNND1 Y96,</i>	GO

				<i>BCAR1 Y414, FAK Y576, Yes Y220, GNMT Y34, SHMT1 Y67, IRS2 Y649, GRF-1 Y1087, BCAR1 Y556, Lyn Y173, SEC14L4 Y36, BCAR1 Y253, BCAR1 Y291, PGAM2 Y92, IRS2 Y814, UGDH Y108, Lyn Y376, FBP1 Y245, Yes Y424, Dok1 Y408, SHP-2 Y62, FAK Y397, CTNNA1 Y654, p38-alpha Y182, FAK Y861, GOT1 Y264, Nck1 Y105, RPL13A Y137, DLG1 Y783, eEF1A1 Y177, talin 1 Y1116, Fgr Y197, FAK Y577 T575, SCAP2 Y260, HSPD1 Y385, FBP1 Y216, SEC14L4 Y36, STAT3 Y686, CrkL Y132, Dok1 Y361, PIK3R1 Y197, Dok1 Y450, FBP1 Y265</i>	
1.3E-03 [^]	2.23	Endoplasmic Reticulum	16	CYB5A Y11, CYP2E1 Y426, GLUL Y336, HPD Y221, CALM1 Y100, PEBP1 Y181, HPD Y232, VCP Y173, CAT Y231, ASS1 Y133*, <i>ALDOB Y204, CAV1 Y14 Y6, CAT Y84, ACSL5 Y69, HSPD1 Y385, ASS1 Y322</i>	GO

NES stands for normalized enrichment scores. ** = FDR = .01, * = FDR = .05, ^ = FDR = .1. NES > 0 indicates that increased phosphorylation on sites in pathway are positively predictive of proportion of $\omega 7 / \omega 9$ PUFA, while NES < 0 indicates that increased phosphorylation on sites in the pathway are negatively predictive. The phosphosites in each category membership are ordered from most positively to most negatively predictive of phenotype. Those phosphosites that contributed most to the enrichment (compose the “leading edge”) appear in regular font, while the other sites in that category are italicized.

Table 9: Phosphosite Sets Predictive of Steatosis

P_{raw}	NES	Category Label	N	Category Membership	Source
<2E-05**	2.53	Oxidation reduction	24	CYB5A Y11, ALDH1A7 Y484, CYP2E1 Y426, GAPDH Y316, SAHH Y193, FASN Y1248, HPD Y221, LDHA Y239, HPD Y232, PRDX6 Y89, CAT Y231*, <i>GPX1 Y147, HGD Y40, UOX Y288, ALDH1L1 Y848, ALDH6A1 Y268, GLUD1 Y135, CAT Y84, GAPDH Y328, ACAD11 Y323, UGDH Y108, UOX Y230, HADHA Y724, GLUD1 Y512</i>	GO, SwissPro, Literature
7.4E-04*	2.31	Metabolic process	25	CYB5A Y11, CYP2E1 Y426, GAPDH Y316, TGM2 Y369, CAR2 Y114, LDHA Y239, PGM1 Y353, PRDX6 Y89, CPS1 Y162, ARG1 Y188, GPX1 Y147, ENO1 Y44*, <i>GLUD1 Y135, CTH Y113, PCBD1 Y70, GAPDH Y328, PGAM2 Y92, CPS1 Y590, HADHA Y724, ARG1 Y197, ACAT1 Y328, ARG1 Y265, GLUD1 Y512, CPS1 Y140, CPS1 Y1450</i>	GO
1.1E-03 [^]	2.25	Cytosol	88	ACLY Y672, SAHH Y193, SHP-2 Y584, FASN Y1248, HPD Y221, TGM2 Y369, TARS Y297, CALM1 Y100, CAR2 Y114,	GO

				LDHA Y239, PGM1 Y353, HPD Y232, PRDX6 Y89, VCP Y173, CAT Y231, BCAR1 Y271, PITPNA Y141, GRF-1 Y1105*, <i>OTC Y317, GNMT Y221, CrkL Y198, GPX1 Y147, Hrs Y216, PLCG1 Y771, LHPP Y159, GSK3B Y216 S219, ALDOB Y204, p38-alpha Y182 T180, ERK2 Y185 T188/T183, ERK1 Y205, vinculin Y822, PGK1 Y196, ERK2 Y185, IRS2 Y671, BCAR1 Y391, CAV1 Y14 Y6, CAT Y84, GSK3B Y216, CTNND1 Y904, vinculin Y692, Shc1 Y313, Abl Y393, SHMT1 Y28, BCAR1 Y238, JAK2 Y570, CTNND1 Y96, BCAR1 Y414, FAK Y576, Yes Y220, GNMT Y34, SHMT1 Y67, IRS2 Y649, GRF-1 Y1087, BCAR1 Y556, Lyn Y173, SEC14L4 Y36, BCAR1 Y253, BCAR1 Y291, PGAM2 Y92, IRS2 Y814, UGDH Y108, Lyn Y376, FBP1 Y245, Yes Y424, Dok1 Y408, SHP-2 Y62, FAK Y397, CTNND1 Y654, p38-alpha Y182, FAK Y861, GOT1 Y264, Nck1 Y105, RPL13A Y137, DLG1 Y783, eEF1A1 Y177, talin 1 Y1116, Fgr Y197, FAK Y577 T575, SCAP2 Y260, HSPD1 Y385, FBP1 Y216, SEC14L4 Y36, STAT3 Y686, CrkL Y132, Dok1 Y361, PIK3R1 Y197, Dok1 Y450, FBP1 Y265</i>	
1.3E-03 [^]	2.23	Endoplasmic Reticulum	16	CYB5A Y11, CYP2E1 Y426, GLUL Y336, HPD Y221, CALM1 Y100, PEBP1 Y181, HPD Y232, VCP Y173, CAT Y231, ASS1 Y133*, <i>ALDOB Y204, CAV1 Y14 Y6, CAT Y84, ACSL5 Y69, HSPD1 Y385, ASS1 Y322</i>	GO

NES stands for normalized enrichment scores. ** = FDR = .01, * = FDR = .05, ^ = FDR = .1. NES > 0 indicates that phosphorylation on sites in pathway most significantly increase with steatosis, while NES < 0 indicates that phosphorylation on sites in the pathway decrease with steatosis. The phosphosites in each category membership are ordered from most positively-predictive of steatosis to most negatively-predictive of steatosis. Those phosphosites that contributed most to the enrichment (compose the “leading edge”) appear in regular font, while the other sites in that category are italicized.

3. Leveraging MS-based PTM datasets: strategies for multivariate modeling and pathway inference in the context of noise and missing data

Introduction

As described in Chapter 1, breakthroughs in mass-spectrometry- (MS) based bioanalytical methods have had a tremendous impact on the study of cellular signaling. MS-based studies have identified thousands of post-translation modifications (PTMs) on proteins, demonstrating that this mode of signal transduction occurs on a scale previously not conceived. Although identification of a single PTM sometimes motivates its characterization via molecular and biochemical studies, quantitative MS-based experiments, capable of measuring hundreds of PTM responses to diverse cues, provide unique opportunities for computational modeling and machine learning techniques to infer biological insight from these measurements on a broader scale.

One of the most exciting challenges for analysis of these datasets is the design of computational techniques to address particular biological questions. In recent years, numerous computational methods have been developed specifically for analysis of these unique datasets. Promising examples include: (1) adaptation of the Steiner-prize-collecting-tree algorithm to infer protein network connectivity from parallel measurements of phosphosite and gene transcription using the protein interactome (Huang and Fraenkel, 2009), (2) partial least-squares regression (PLSR) to model cancer cell phenotypes as functions of underlying phosphorylation site measurements (yielding both quantitatively predictive models of important phenotypes and new evidence for the involvement of particular phosphosites in response mediation) (Huang et al., 2010; Kumar et al., 2007), and (3) enrichment analysis to infer shared modes of regulation (e.g., kinase regulation, cellular location) for phosphosites with similar responses to cues or perturbations

(Joughin et al., 2009; Naegle et al., 2011). These analysis frameworks leverage the multivariate nature of and quantitative information from these datasets. They provide interpretable biological predictions, and, as a result, are essential to ensuring that biological insight scales with MS-based PTM datasets.

Gene Set Enrichment Analysis as described in (Subramanian et al., 2005) (GSEA) enables detection of gene sets relevant to particular phenotypes, based on quantitative transcriptional datasets, and adaptation of the GSEA framework for the analysis of quantitative, MS-based PTM datasets would offer important advantages. Briefly, the GSEA method requires two inputs: (1) a gene set, defined based on prior biological knowledge (e.g., biological function or chromosomal location) and (2) a list of genes that are ranked according to correlation with a phenotypic or class distinction (e.g., cellular proliferation or tissue type). The method, based on the Kolmogorov-Smirnov statistic, tests whether members of the gene set are significantly enriched at the top or bottom of the list, which, in turn provides evidence for positive or negative regulation of that gene set in the phenotype. This method was first described in a transcriptional analysis of muscle biopsies, comparing diabetics to healthy controls, and GSEA was sensitive enough to detect reduced expression of oxidative phosphorylation genes in diabetics, even though the average decrease in gene expression was only ~20% (Mootha et al., 2003). Thus, GSEA, through incorporation of quantitative information from all genes in the study (the whole list), is capable of detecting even subtle differences in gene pathways.

As an alternative to GSEA, discrete overlap statistics, such as the cumulative hypergeometric distribution, are commonly used. However, these methods lack the ability to detect subtle differences in gene set expression, because they rely on an arbitrary cutoff to segregate genes into separate sets. For example, to use an overlap statistic, gene measurements might be divided

into the follow two sets: (1) genes that positively correlate to the phenotype with a raw p-value less than some cutoff, α , and (2) the other genes. No matter how α is selected, division of the gene measurements into two discrete sets results in a loss of quantitative information, and only those gene sets with enough genes at an α -level of significance will be detected. Equally problematic, however, is that p-values calculated using the hypergeometric distribution assume gene independence, leading to an overestimation of the enrichment significance. As an example, many genes share transcription factors and are therefore not independent. Similarly, this would present a problem for enrichment analysis of PTMs, as a single kinase, for example, might regulate the phosphorylation of several proteins under certain conditions. In contrast, GSEA accounts for the possibility of gene dependence, as significance is assessed by permutation of phenotype labels. This method preserves the gene correlation structure and thereby factors dependence relationships into the null model. Thus, relative to commonly used overlap statistics, GSEA offers improvement in both sensitivity and estimation of significance.

Incorporation of GSEA principles could strengthen several computational techniques that have already been applied to PTM datasets. Most obviously, clustering-based enrichment analysis methods could be improved by adoption of GSEA-style analysis. A previous study reveals the power of enrichment analysis for deriving biological insight from unsupervised clustering of phosphotyrosine datasets (Naegle et al., 2011). However, this analysis relies on overlap statistics, and therefore could improve with the use of the weighted enrichment score, the statistic used for GSEA analysis, which, as described above, would boost sensitivity and decrease p-value inflation. Thus, application of GSEA would enhance current clustering-based enrichment-analysis strategies.

Another important application is multivariate regression modeling of phenotypes as functions of PTM measurements. In two previous studies relating phosphotyrosine measurements to cellular phenotypes, models were initially constructed using all phosphosite measurements (~100) (Huang et al., 2010; Kumar et al., 2007). Model reduction techniques were then used to rank phosphosites according to importance for model prediction and build a new model on a reduced set of phosphosite determinants (~10). This procedure has several benefits. First, reducing the number of nonessential model variables can reduce over-fitting and improve model predictive performance. Second, from hundreds of phosphotyrosine measurements, the experimenter can focus on a select core of phosphosites that are most predictive of phenotype. Despite these benefits, the model reduction process may suffer from problems paralleled in early gene expression analyses.

One of the first analysis methods for gene expression data involved ranking genes according to a phenotype or treatment condition and then selecting the most highly ranked genes for further investigation. If significance was calculated and multiple-hypothesis-testing (MHT) procedures applied to gene expression comparisons, often very few or none of the comparisons were significant, providing researchers with limited insight. Often the most highly ranked genes were related in no obvious way. In the PLSR modeling framework described above, phosphosites were simply ranked according to a predictive metric. Similar to gene expression analysis, were significance for these sites calculated, and, additionally, were MHT taken into account, it is likely that few or none of the phosphosites would be significantly predictive. Although it is possible that the most highly ranked phosphosites, incorporated into the reduced model, may genuinely be the most important drivers of the phenotype, consideration of GSEA principles and incorporation of pathway-level information may significantly enhance the analysis. For

example, cumulative but subtle phosphorylation changes across many proteins in a given pathway (or pathways) might multivariately contribute to a response. However, pathway-level detection methods, such as GSEA, have not been incorporated into PLSR analyses.

In addition to consideration of pathway-level regulation with a GSEA-style framework, multivariate modeling of MS-based PTM data could be improved by addressing a challenge that is inherent to MS-based technology. MS datasets, amassed by combination of multiple individual MS experiments, suffer from significant amounts of systematically missing data. Current quantitative MS methods enable simultaneous comparison of eight samples or fewer, the maximum being achieved by 8-plex iTRAQ (AB Sciex), an isobaric mass-tag reagent. Although methods exist for quantification of a pre-existing list of peptides (Wolf-Yadlin et al., 2007), in practice, most experiments are run using data-dependent acquisition, a discovery mode in which the mass-spectrometer selects peptides for detection and quantification based on relative peptide abundance at discrete times during sample acquisition. In other words, with data-dependent acquisition, phosphosites detected and quantified in one experiment often are not detected and quantified in a subsequent experiment. This becomes problematic for experimental designs including more than eight conditions. In that case, multiple experiments must be merged, each with quantitative data for a set of incompletely overlapping phosphosites, leading to final datasets with systematically missing data points. The most common approach is to remove incomplete data and limit computational analysis to the complete data subset, which can result in >50% reduction in the number of PTM dataset measurements. As experimental designs become more ambitious, it becomes increasingly important to evaluate additional missing data techniques so that multivariate modeling maximally leverages the PTM measurements.

Finally, most computational analyses of PTM datasets do not leverage associated measurement error. In the context of other biological datasets, incorporation of measurement error into relational Markov and Bayesian models has improved quality of prediction (Akavia et al., 2010; Jaimovich et al., 2005). In these models, data is weighted according to quality, and the benefits are two-fold. Higher confidence trends are not diluted in low-confidence trends, but low-confidence trends can still contribute additively. Thus, via incorporation of measurement error into the modeling framework, data is utilized maximally. For many PTM datasets, technical and/or biological replicates are acquired, and the relative error associated with a particular PTM measurement can vary greatly (~1-100%), even for PTM raw data that has been manually inspected for quality. For datasets where the quality of PTM measurements is variable, the computational analysis can only benefit from incorporation of this additional information.

Here, we describe a suite of computational methods that have been developed specifically to address each of these important issues in the modeling of PTM datasets. Motivated by the phosphotyrosine dataset described in Chapter 2, we developed and tested (1) GSEA-style analysis to detect phosphosite set enrichments in the context of supervised clustering and multivariate regression modeling, (2) a method for incorporation of measurement noise, based on Gaussian sampling, to both analysis methods, and (3) a general framework for evaluating missing data methods in the context of multivariate regression modeling. All three of these methods would advance state-of-the-art analysis for MS-based PTM datasets, and are likely to offer advantages in other contexts as well.

Results

Study design

Chapter 2 describes the design of a MS-based phosphoproteomic analysis, whose goal was to determine how deletion of hepatic protein-tyrosine phosphatase 1b (PTP1b) rendered mice resistant to many of the negative consequences of high-fat diet (HFD) stress. PTP1b modulates site-specific protein tyrosine phosphorylation, and network-level alterations in protein tyrosine phosphorylation upon PTP1b deletion were likely drivers of the improved phenotypes in liver-specific PTP1b deletion mice (L-PTP1b $-/-$) (Delibegovic et al., 2009). Global, MS-based analysis of protein phosphotyrosine in L-PTP1b $-/-$ and control liver tissues under both normal and HFD conditions lead to the identification and quantification of 301 phosphosites, across 48 samples. In addition to genotype and diet treatments, liver tissues were harvested under insulin-stimulation (insulin) or basal conditions, leading to a total of eight treatment combinations, with 3-8 mice per condition. In order to maximize the number of samples quantified per experiment, an 8-plex iTRAQ strategy was used for relative quantification of eight samples in each experiment. Generation of the final dataset (48 samples) was accomplished in ten independent MS experiments, which were combined by normalizing each to a control sample, kept constant across experiments. Because it was unknown a priori which phosphosites would be responsive to the various perturbations, MS experiments were run in data-dependent acquisition (discovery) mode.

This experimental design enabled the discovery of numerous PTP1b-dependent phosphosites, many of which had uncharacterized biological function. The central computational goal for this dataset was to understand how these phosphosites might mediate the therapeutically attractive phenotypes of L-PTP1b $-/-$ mice. Two complementary approaches are herein described. First,

enrichment analysis of PTP1b-dependent sites was used to determine whether individual PTP1b-dependent phosphosites might interact as a part of a larger pathway or set. Second, quantitative measurements of metabolic phenotypes were made, presenting an opportunity for multivariate modeling to elucidate phenotype-phosphosites relationships *in vivo*.

However, the dataset also presented two unique technical challenges. Foremost, combination of multiple, data-dependently-acquired experiments lead to a sparse dataset with systematically missing data (Figure 3.1A). Secondly, while measurement error, estimated from technical replicates, was relatively small ($\sigma_{\text{rel}} = 10\% \pm 10\%$), fold-changes were small as well (generally < 2-fold), and some of the data points were associated with large measurement error (30-70%) (Figure 3.1B). Thus, additional goals for this study were to maximally utilize all data available and to improve computational prediction through the incorporation of measurement error.

Stochastic correlation reflects measurement error

As described above, GSEA analysis was developed to detect set enrichments among lists of genes, ranked according to correlation with phenotype or treatment condition. The simplest way to detect pathway and set enrichments among PTP1b-dependent phosphosites would also involve correlation. However, noisy data begets spurious correlation, especially for phosphosites with fold-changes smaller than measurement error. Although it would be possible to remove these phosphosites using analysis of variance (ANOVA), the sensitivity of GSEA is dependent on incorporation of the full dataset. In addition, the selection of a cutoff for data inclusion is often difficult to justify.

A method, termed “stochastic correlation”, was devised to incorporate measurement error into the correlation coefficient calculation, so that the magnitude and significance of the correlation

reflect measurement certainty. The basic concept is illustrated in Figure 3.2A, which shows two identical trends but with varying levels of measurement error. Clearly, the higher confidence trend would also lead to a higher confidence correlation, but too often measurement error is ignored in the correlation coefficient calculation. Given a mean and estimate of error from replicate experiments, it is possible to simulate measurements. Here, we propose a method that combines Gaussian sampling with bootstrapping to model measurement noise and limited sample size, respectively, resulting in a distribution of correlation coefficients for pairs of measurements. The significance of the correlation can be readily estimated from resulting coefficient distributions (Figure 3.2B).

Empirically, this method works as expected. For example, to generate Figure 3.2C, 76 phosphosites from the HFD basal (HFD_b) data subset (measured across 13 samples) were correlated with genotype condition (L-PTP1b^{-/-} or control) using both stochastic and deterministic correlation (deterministic correlation is based on bootstrapping only). The stochastic and deterministic sets of significantly correlating phosphosites are displayed in Venn diagrams. As a point of comparison, ANOVA was used to test the null hypothesis for each phosphosite (i.e., phosphorylation levels were identical across all samples), and sets of phosphosites that passed the ANOVA filter also appear in the Venn diagrams. At various levels of significance, we see that deterministic correlation results in detection of spurious correlations (blue set, Figure 3.2C), which would have been avoided using ANOVA filtering at the same significance level. Importantly, correlations significant by stochastic correlation are a subset of ANOVA-filtered phosphosites only (purple set, Figure 3.2C), suggesting that this method inherently filters out trends dominated by noise.

We verified that the significance associated with pairwise correlations in the regime of interest ($n \sim 8-15$ observations) were consistent between bootstrapping and analytical methods, where the latter was based on correlation transformation and Student's t distribution. In Figure 3.3A, we compare p-values from bootstrapping alone ("Deterministic Bootstrap", no Gaussian sampling step) to analytical p-values and found good agreement ($R^2 = .82$), despite relatively small sample sizes. These correlation coefficients were calculated between 76 phosphosites and seven responses ("Genotype", "PTP1b", " $\omega 7/ \omega 9$ PUFA", etc.) for a total of 380 correlations. As highlighted in the legend of Figure 3.3, each of the responses was associated with a particular level of measurement error (0% - 32%). Figure 3.3B verifies that stochastic correlation ("Stochastic Bootstrap") truly weights statistical significance according to the associated measurement error in comparison to bootstrapping alone ("Bootstrap Deterministic"), as all p-values lie above the diagonal in Figure 3.3B. Importantly, the significance associated with phosphosite correlations to the noisiest response variable ("C18 SCD Index") are penalized most (i.e., are located far from the diagonal, indicated by arrows in Figure 3.3B).

Because random permutation is more commonly used to gauge correlation significance than bootstrapping, we also considered permutation techniques. As with bootstrapping, we first verified agreement between random permutation ("Deterministic Permutation") and analytical estimations of correlation significances in our measurement regime (Figure 3.3C). To incorporate measurement noise into the random permutation framework, we included a Gaussian sampling step to model measurement error followed by random permutation of the response variable. The procedure was repeated to generate a distribution of random correlation coefficients corresponding to the null model, and p-values were estimated. We expected that higher levels of measurement noise would increase the variance of the null distribution, resulting

in lower p-values of correlation. However, this procedure failed to weight correlation significance according to measurement error (Figure 3.3D). Whereas, in the bootstrapping-based formulation, incorporation of measurement error significantly increased the spread of the “actual” correlation coefficient distributions involving noisy measurements, incorporation of measurement error minimally affected the spread of random correlation coefficients for the null model. Thus, using bootstrapping-based stochastic correlation, we tested the effects of measurement error in GSEA-style analysis.

PTM Set Enrichment Analysis extends GSEA for PTM datasets

PTM Set Enrichment Analysis (PSEA) is a direct application of GSEA (Subramanian et al., 2005) to quantitative protein PTM datasets. Similarly to GSEA, analysis begins with a ranked list, *L*, of PTMs (e.g., according to dependence on treatment condition, correlation to phenotype) and a pre-defined set, *S*, of PTMs (e.g., based on motif similarity, protein pathway). The goal of PSEA is to determine whether members of *S* are randomly distributed throughout *L* or overrepresented mainly at the top or bottom. Conceptually, sets showing enrichment at the top or the bottom of the list may represent pathways or modes by which PTMs are regulated by or contribute to a phenotype or treatment condition. Application of PSEA to the L-PTP1b-/- phosphotyrosine dataset for the detection of enriched, PTP1b-dependent PTM sets highlights several unique PTM-specific considerations.

While an increase or decrease in gene abundance is generally interpretable in terms of positive or negative regulation, the regulation of proteins by PTMs is more nuanced. For example, PTMs can activate, inhibit, or play some other role in modulating protein behavior. To incorporate this concept into PSEA, both signed and unsigned ranked lists of PTP1b-dependent sites were considered. Signed rankings are based on correlation, while unsigned rankings are based on the

absolute value of correlation. Signed PSEA would detect sets of PTMs that are uniformly increased or decreased in response to some perturbation. This style of analysis might work well to detect a set of phosphosites regulated by a treatment-sensitive kinase, in which phosphorylation of downstream substrates would correlate with kinase activity, or to detect particular acetylation pathways, some of which have been shown to uniformly increase or decrease acetylation upon nutrient cues (Zhao et al., 2010). Unsigned analysis, on the other hand, would detect condition-dependent activation of protein pathways that simultaneously require increased PTM on some sites but decreased PTM on other sites. For example, the individual protein Src contains both inhibitory and activating phosphorylation sites, which must be unphosphorylated and phosphorylated, respectively, for maximal activation (Thomas and Brugge, 1997). In the context of phosphorylation, unsigned analysis would be appropriate for detection of a PTM response regulated by both kinases and phosphatases.

In addition, there are some technical considerations for MS-based PTM datasets. In GSEA, correlation is often used for ranking genes. In contrast to gene expression datasets, MS-based PTM datasets often have missing data points, and, in that case, correlation itself is not an ideal metric for ranking the PTM sites. For example, two PTM sites, A and B, might correlate similarly with a phenotype (i.e., have similar correlation coefficients). If A has more observations than B, the correlation between the phenotype and A is more significant, and A should be ranked more extremely in the list than B. Thus, in the context of missing data, PSEA rankings should incorporate significance into variable ranking.

PSEA of PTP1b-dependent phosphosites was used to compare signed and unsigned analyses as well as the effects of incorporating measurement error using stochastic correlation (Figure 3.4A). The L-PTP1b^{-/-} dataset was divided into four independent data subsets for the analysis (normal

chow (NC) basal (NCb), HFDb, NC insulin (NCi), and HFD insulin (HFDi)). Correlations and associated significances were calculated independently for each data subset, so that factors, such as diet or insulin-stimulation would not mask genotype dependence. The four data subset correlations were then combined using Fisher's method, which resulted in a combined p-value and overall direction of correlation. For signed analysis, phosphosites were sorted according to significance ($-\log_{10}(P_{\text{combined}})$) multiplied by the direction of correlation. For unsigned analysis, phosphosites were ranked according to significance only. As can be seen in Figure 3.4A, the signed list contains many more negatively correlated (increase phosphorylation upon PTP1b deletion) phosphosites than positively correlated. The construction of the null model takes this into account. Specifically, to generate the null model, PSEA was performed on the same datasets but ordered in ranked lists according to correlation with randomly permuted genotype labels. The corresponding null model histograms of positive and negative normalized enrichment scores (NES), the PSEA statistics, are shown in Figure 3.4B and C. Figure 3.4D and E show examples of significant and insignificant PSEA set enrichments. On the left-hand side of each figure is the signed ranked list of PTP1b-dependent phosphosites, in the middle panel are the locations of the phosphosites in the phosphosite sets. The calculation of the NES is detailed in Methods, but, briefly, the NES is derived from a running sum statistic, which is depicted in the right-hand side panels of Figure 3.4D and E. The running sum begins at the most positively correlated end of the list (+ ρ) and moves, in the direction of the arrow, to the most negatively correlated end of the list (- ρ). For each step through the list taken, the running sum is incremented when a phosphosite set member is encountered and decremented when one is not. The NES is the running sum's largest deviation from zero, and the significance of this score is obtained through

comparison to the null model distribution of NESs, and, when more than one set is tested, correction for multiple-hypothesis testing.

The results of PSEA are sensitive to both signed versus unsigned analysis as well as deterministic versus stochastic implementations (Figure 3.5). The table is color-coded according to significance of enrichment, and yellow, pink, and cyan correspond to sets significant controlling for an FDR of .01, .05, and .1, respectively. Most obviously, stochastic PSEA consistently resulted in detection of fewer significant enrichments. This would be consistent with the stochastic correlation results and suggests that some of the deterministic PSEA enrichments resulted from spurious correlation with phosphosites whose trends were indistinguishable from measurement error. Other sets, such as “lipid metabolism” and “arginine and proline metabolism” remain significant but to a lesser extent upon incorporation of measurement error. Several of the most significantly PTP1b-dependent sets (“endoplasmic reticulum”, “cytosol”, “metabolic process” and “oxidation reduction”) are robust to incorporation of measurement error. Thus, incorporation of measurement error lends additional confidence to some predictions and casts doubt on others.

Comparison of deterministic signed versus unsigned enrichment show that the “Insulin signaling pathway” is only significant under signed analysis whereas “Insulin-sensitive pYs in adipocytes” share the same level of significance under both analyses. The “Insulin signaling pathway” set contains phosphosites on canonical insulin-signaling proteins, most of which increase phosphorylation upon insulin stimulation. Given that half of the L-PTP1b^{-/-} dataset contains insulin-stimulated liver samples and L-PTP1b^{-/-} are insulin hyper-sensitive, detection of a concordant phosphorylation change in this pathway makes sense. Interestingly, the “Insulin-sensitive pYs in adipocytes” contains a set of phosphosites that were found to be either positively

or negatively insulin-responsive in an MS-analysis of insulin-stimulated adipocytes (Schmelzle et al., 2006), and, thus, these sites would be more likely to be detected under both conditions.

Modeling of measurement error using stochastic multivariate regression

As described above, one of the computational goals for the L-PTP1b^{-/-} was construction of multivariate models of PTP1b-dependent phenotypes as functions of the measured phosphotyrosine network changes. In contrast to a univariate metric like correlation, multivariate regression is capable of predictively modeling phenotypes as the result of multiple underlying factors, and, in this way, comes closer to capturing the complex, multifactorial nature of biological systems. Given the statistical power of the dataset, only multivariate linear relationships were considered. In addition, the analysis was limited to a particular data subset (HFD_b), based on the assumption that the molecular networks of L-PTP1b^{-/-} and control mice under a single diet and stimulation condition would be in a regime where phenotypes of both mouse genotypes could be approximated by a single multivariate linear model, as in a Taylor series expansion. In the context of the L-PTP1b^{-/-} dataset, strategies for tackling two technical challenges, measurement noise and missing data, are described here and in the subsequent section.

Stochastic multivariate regression (SMR) is the multivariate cousin of stochastic correlation, similarly motivated by the need to mitigate detection of spurious relationships from measurement noise. Similarly, Gaussian sampling and bootstrapping are used to develop a distribution of regression coefficients that incorporates measurement error and controls for limited sample size (Figure 3.6A). The coefficient distribution can then be used to evaluate the significance of predictor variables to the model.

This method is likely to be crucial for analysis of the L-PTP1b^{-/-} dataset. As shown in the Figure 3.1B, the relative phosphorylation fold-changes are small for the technical replicate samples plotted; most range between 2-fold, while relative error is 10%. Thus, for this dataset in particular, it is important to distinguish experimental from biological variation. Z-scoring is a common scaling method for multivariate regression modeling of MS-based phosphorylation datasets; this method entails mean-centering and normalization of variance for each set of phosphosite measurements. This scaling essentially weights all measurements equally, and, in the context of the L-PTP1b^{-/-} dataset, this would have negative consequences if measurement error were not controlled in some way. As discussed above, ANOVA filtering of data is a strategy to limit the influence of spurious trends by removing them from the dataset prior to model-building. However, it requires selection of an arbitrary filtering cutoff, and, in the context of multivariate regression, the cutoff should ideally reflect the synergy between error in both response and predictor measurements.

The ability of SMR to filter measurement error is highlighted in a comparison to deterministic multivariate regression (DMR) and ANOVA filtering (Figure 3.6B). Phosphosites selected for three models (“Genotype”, “Steatosis”, and “C18 SCD Index”) by SMR and DMR are shown with ANOVA-filtered phosphosites in Venn diagrams. As in stochastic correlation, SMR-selected variables are a subset of the ANOVA-filtered phosphosites, whereas DMR selected-variables are not. The three model response variables “Genotype”, “Steatosis”, and “C18 SCD Index” have associated relative standard errors of 0, 16% and 32%. Importantly, none of the phosphosites are significantly associated with the noisiest response variable, “C18 SCD Index”, because the error in the response measurements, modeled in SMR, is too great. In this last case, ANOVA-filtering of the phosphosites would not have substituted for SMR, as the dominant

measurement error was in the response. Thus, SMR holds promise, as the method integrates measurement error from both predictor and response variables, which, in turn, is reflected in the significance of resultant regression coefficients.

It is important to compare SMR model reduction to established methods. Significance thresholding was previously used for multivariate regression model reduction; this method was based on simulation of the null distribution via random permutation of the response variable (Janes et al., 2004). Previous multivariate regression modeling of MS-based phosphotyrosine datasets involved ranking independent variables according to a variable importance of projection (VIP) score for model reduction (Huang et al., 2010; Kumar et al., 2007). In this study, ranking of variables is naturally accomplished by using the coefficient distributions themselves to test significance, and, here, this ranking was used to build reduced models. Because, to our knowledge, this method has not been previously reported, we compared SMR coefficient-significance-thresholding to VIP-score-based model reduction for our dataset and compared results. In terms of variable selection, there is good correspondence between (1) typical SMR and a stochastic implementation of VIP-score model reduction (Figure 3.6C). Interestingly, the model coefficient curves separate on the basis of response measurement error along the ordinate axis (according to coefficient significance), but not horizontally (according to VIP score) (Figure 3.6C). Thus, coefficient significance might be useful for global comparison of predictive importance across models, whereas VIP scores can take high values even in the context of noisy measurements and thus should only be compared within a model. In terms of model performance, the two model reduction methods behave similarly (Figure 3.6D).

Optimization of missing data methods for multivariate regression

The second consideration for multivariate regression modeling with the L-PTP1b^{-/-} dataset was missing data points. In contrast to univariate methods, multivariate regression requires that a dataset be complete. The phosphotyrosine dataset posed a challenge, due to the number of systematically missing data points that resulted from merging multiple data-dependently acquired MS experiments. The HFDb data matrix is displayed in Figure 3.7A. Because the data points are systematically missing, as opposed to randomly, data methods based on matrix decomposition, which work well in the context of multivariate regression by partial least squares, were not feasible (Nelson et al., 1996). For this reason, alternative methods were tested. The simplest of which was “list-wise deletion” of all measurements for which sample coverage was incomplete. This method requires no assumptions or ad hoc rules but seems wasteful. In the case of the HFDb dataset, analysis would be limited from 228 phosphosites to merely 76 (Figure 3.7A). Another approach would be to consider multiple complete data subsets independently (as described in Supplementary Figure 3.1 and Methods) and then integrate results across the complete data subset models using a decision rule. In this way, data imputation would be avoided, but, as is the case here, decision rules may be heuristic rather than optimal. Finally, imputation methods were considered. The goal of an imputation method is to facilitate analysis while minimizing bias, and, multiple random imputation methods fulfill this criteria (Rubin, 1996).

Several multiple random imputation (MRI) methods were considered. These methods were based on either Gaussian or random imputation. In Gaussian imputation, the missing samples for a measurement are imputed from those available by drawing a Gaussian sample whose mean and standard deviation corresponded to the available sample measurements. Similarly, random

imputation replaces missing values with a randomly selected sample from those available for that measurement. The potential bias from imputation is minimized by repeating these single imputation steps multiple times. Conveniently, such an imputation step seamlessly integrates into the SMR algorithm, as the imputation step can immediately follow the Gaussian sampling step (Figure 3.7B, Methods).

In addition, two variations on the MRI methods were considered, so that additional information about the missing samples could be incorporated into the process. The first method is treatment-dependent MRI, which leverages treatment information. Specifically, L-PTP1b^{-/-} missing data samples are randomly imputed from the L-PTP1b^{-/-} samples present; likewise, control samples missing are imputed from the control samples present. The second method, termed “treatment-pval” is a mixture of general and treatment MRIs and involves a two-sided Student t test comparison of L-PTP1b^{-/-} and control phosphorylation levels for samples present. If the null hypothesis is rejected (treatment effect detected), the missing measurement values are imputed using treatment MRI; otherwise, general MRI is used.

Each of these methods has potential advantages and drawbacks, and so a comprehensive strategy for testing each was developed (Figure 3.7A). The complete data subset (13 samples by 76 phosphosites) provided a “true” or reference dataset for SMR modeling, and then ten simulated datasets were constructed from the complete HFDB data for testing of the missing data methods. The simulated datasets were generated by randomly removing measurements in a pattern that structurally corresponded to the original HFDB data matrix (Figure 3.7A). Five different phenotypic responses were modeled in five independent multivariate regression models as functions of phosphotyrosine measurements, leading to five “true” models based on SMR of the complete data subset and 50 models (5 models X 10 simulated datasets) for each of the missing

data methods tested. A receiver-operator characteristic (ROC) curve was used to compare the performance of the various missing data methods (Figure 3.7A and Figure 3.8, see Methods). Models were evaluated based on their abilities to (1) select phosphosites that would have been incorporated into “true” models (based on the complete data) and (2) avoid selection of phosphosites that were not included in the “true” models. Using this variable-selection-based evaluation method, it was also possible to compare univariate correlation.

Interestingly, for this particular dataset, there was significant disparity in the performance of the missing data methods (Figure 3.8). As might be expected, list-wise deletion performs poorly relative to the “true” (complete) models. Because more than two-thirds of the dataset is excluded from the analysis, list-wise deletion can maximally recover only ~30% of the “true” (complete model) positives. Surprisingly, this lack of detection offers no apparent benefit, as other methods out-perform list-wise deletion even in the low-FDR regime. The methods that integrated results from individual models built on complete data subsets (“All Subsets Agree” and “Complete Subset Rules”, see Methods) at least provided the opportunity to detect most of the true positives, but at high false-positive rates. Surprisingly, the methods that performed best were those based on treatment-dependent MRI, suggesting that, for this particular dataset, the treatment-dependent assumption was good. In addition, correlation performed nearly as well. Correlation has several advantages in variable selection, as it utilizes all data and makes no imputation assumptions. However, although correlation performs competitively in variable selection in this particular context, some missing data strategy would be required for subsequent construction of a quantitatively predictive multivariate model. Importantly, the performance of missing data methods was specific to the response modeled (Figure 3.8, lower panels),

suggesting that the application of missing data methods should be evaluated on a case-by-case basis.

Integrated SMR-PSEA analysis yields predictive models and pathway-level insight from incomplete data

The PSEA of PTP1b-dependent phosphorylation sites detected a set of lipid metabolic proteins, and lipid profiling of the liver samples revealed novel PTP1b-dependent alterations in lipid metabolism (Chapter 2). The techniques developed, PSEA, SMR, and “missing-data method selection”, were used to develop multivariate regression models of three PTP1b-dependent lipidomics phenotypes using the full HFDb phosphotyrosine dataset.

Before describing the performance of these models, it is important to consider the performance metrics developed for this study. Common performance metrics, R^2 goodness-of-fit and Q^2 goodness-of-prediction, require a complete data matrix for calculation, and so these were estimated from the incomplete data matrix using an assumption consistent with the selected missing-data method (Methods). In addition, as detailed in methods, Q^2 values in this analysis estimate predictive performance by considering both model selection and fit and are referred to as $Q^2_{v,f}$ (variable selection and fitting). In contrast, typical Q^2 value calculations consider model fit only (Huang et al., 2010; Janes et al., 2004; Kumar et al., 2007), leading to inflation of Q^2 values; these are referred to as Q^2_f . In Figure 3.9A, $Q^2_{v,f}$ and Q^2_f are compared for five SMR models built by varying the number of model independent variables and model principle components (see Methods). Most of these points fall below the line $Q^2_{v,f} = Q^2_f$, suggesting that $Q^2_{v,f}$ is a more conservative estimate of predictive performance, because it uniquely accounts for over-fitting due to variable selection as well as the model fit itself. Importantly, red dots (Figure 3.9) denote Q^2 values for final models, and we can see that the Q^2 values reported here are more

conservative than those presented in other studies. The quantitative performance of the five models is summarized in Figure 3.9B. R^2 values range from .82 to .84, while Q^2 values range from .52 to .66. As described above, the Q^2 values reflect model prediction for independent samples. It is also important to consider that the phosphotyrosine and lipidomics measurements are tissue-level and therefore more variable than measurements from uniform cell lines, the experimental system used previously for multivariate regression modeling of (cellular) phenotypes as functions of underlying MS-based phosphotyrosine measurements (Huang et al., 2010; Kumar et al., 2007). Thus, taking all of this into account, these models perform surprisingly well. Figure 3.9B also highlights the spread in experimental error among the phenotype measurements. For steatosis in particular, there is one measurement (indicated with a blue star) whose measurement error is much larger than the other measurements. Luckily, using SMR, this measurement is weighted according to error in the model selection process.

The phosphosites retained in final models are displayed in Figure 3.10A. Although many of the phosphotyrosine and lipidomics measurements in the models have unknown functional significance, consideration of underlying phenotype predictors, in the context of phosphosite protein function and other information from the literature, suggests biological mechanisms connecting some of these phosphosites to the phenotypes (Chapter 2 Results and Discussion). However, the biological implications of other phosphosites were challenging to interpret.

To gain additional insight, we integrated PSEA into the multivariate modeling framework. Variable reduction methods, regardless of the specific multivariate regression technique used, generally rank underlying variables (e.g., phosphosites) according to predictive potential. These ranked lists can be used to determine whether phosphorylation patterns on particular sets or protein pathways are multivariately predictive and might underlie a response phenotype. Null

models, constructed using random permutation of the response variable, enable estimation of significance. Here, PSEA of the three phosphosites-dependent phenotypic models highlighted the phosphorylation of “oxidation reduction” (REDOX) proteins in all three of the PTP1b-dependent lipidomics phenotypes (Figure 3.10B). The PSEA analysis of PTP1b-dependent phosphosites yielded several set enrichments, but “oxidation reduction” was consistently most significant across all analyses, potentially indicating a unifying role for REDOX in mediating all three of these phenotypes (Figure 3.10A and B). As discussed in Chapter 2, these results make sense biologically, as the modeled lipidomics phenotypes are each highly dependent on the oxidative state of the cell.

Discussion

We have demonstrated the utility of several computational approaches developed specifically for analysis of the L-PTP1b^{-/-} study. Our results suggest that these methods would be generally applicable to other analyses, especially those focusing on MS-based PTM datasets, where measurement noise, missing data, and the novelty of measurements require new and sensitive approaches. Foremost is the adaptation of the exceptionally sensitive method, GSEA, for the discovery of PTM-dependent phosphosite sets and pathways (PSEA). As discussed in Chapter 2, PSEA recovered known PTP1b-dependent pathways (insulin signaling) and discovered novel PTP1b-dependent pathways: redox homeostasis, amino acid metabolism, and lipid metabolism (particular fatty acid metabolism). Using global, LC-MS-based lipidomics, we discovered co-occurring lipid-metabolic changes, and fatty acid changes were the most significant. These results lend biological credibility to this approach. They also suggest that the ranking of the PTP1b-dependent phosphosites, through incorporation of significance to account for grossly incomplete data (Figure 3.1A), was successful. Importantly, the PSEA framework described here

involves consideration of the diversity in PTM function (e.g., activating, localization) through signed and unsigned analysis, and we demonstrate differential sensitivity to both analysis modes. Similar to GSEA, the predictions of PSEA are robust due to implicit modeling of the dependence structure in the null model. In addition, we have demonstrated that additional robustness in model prediction can be gained through the modeling of measurement error, even in the context of GSEA-style analysis, which is well known for its robustness to noise.

Equally important, we propose a general framework for the construction of multivariate regression models in a regime that has not been well explored: large amounts of systematically missing data, where the number of measurements greatly exceeds the number of observations. We evaluated diverse missing data methods using simulation and ROC performance. We found that, list-wise deletion, the most common solution chosen by MS-based PTM dataset generators and analysts alike, merits quantitative comparison with other methods, as, in the context of this dataset, list-wise deletion was exceptionally suboptimal. In addition, we proposed and tested variations on traditional multiple random imputation (MRI) methods, namely, the incorporation of additional information (treatment condition). Although historically MRI developed to enable data analysis with minimal data assumptions (Rubin, 1996), inclusion of an additional assumption (treatment dependence) yielded better results. Additional assumptions merit rigorous testing prior to adoption, as, prior to performance assessment by ROC analysis, it was unclear whether the treatment assumption would decrease false positive rate.

Given the vast amounts of missing data in the HFDb dataset (Figure 3.7A), the construction of predictive multivariate regression models was an unexpected result. Through conception of a stricter Q^2 value for prediction, $Q^2_{v,f}$, which tests both variable selection and fitting, we more rigorously demonstrate the predictive power of these models. In addition, because measurement

error was incorporated into the model building process, we have greater confidence that the selection of predictor variables, on which the final models were constructed, is not an artifact of measurement error. Importantly, we gained additional utility out of the SMR models built on phosphotyrosine measurements, as the ranked lists of predictive phosphosites were used for PSEA. Interestingly, this PSEA is based on multivariate rather than univariate dependence, and it remains to be tested whether multivariate ranking offers benefits over univariate rankings.

Future directions include optimization of measurement error models. Here, error from technical replicates was not available for all sites (missing for ~ 85%), and models were required to estimate error. For simplicity, the missing error estimates were set to either average relative standard deviation (10%) or average absolute standard deviation (0.1), yielding similar results for both stochastic correlation and multivariate regression. However, as error itself is randomly distributed, this method may unfairly increase the weight or importance of sites with (randomly) extremely small standard deviations and, similarly but less worrisome, limit the influence of sites with extremely large standard deviations. Bayesian modeling of measurement error would temper these effects. In a Bayesian framework, measurement error would be weighted by a prior belief that error is randomly distributed, and the prior would mitigate effects from extreme error observations.

Another important future direction would be to weight the final fit of the SMR model. Measurement error is currently incorporated into the model selection of the procedure, as it influences the significance ranking of coefficients, and, in this way, it indirectly biases model fit toward higher confidence predictor as well as response measurements. The model fit in the SRM procedure, however, is simply unweighted, linear least squares, which does not directly incorporate measurement error. As highlighted in Figure 3.9B, individual response

measurements have a spread of associated measurement error, and model predictive performance might benefit from a weighted-fit approach (e.g., inverse error). In sum, this chapter provides important methodology and validation of the computational technique that spurred experimental investigations in Chapter 2. In a broader context, this chapter provides insight into several computational problems: (1) incorporation of measurement error in correlation and multivariate regression, (2) strategies for testing missing data methods and results for multivariate regression in the less well-studied contexts of systematically missing data and a regime where there are many more measurements than observations ($M > N$ regime), (3) adaptation and proof of concept for GSEA in PTM-set enrichment, and (4) re-evaluation of the predictive Q^2 metric in multivariate model reduction. In addition, the results from methods highlight the value of technical and biological replicates in experimental design. Together, these methods represent progress toward maximal leverage of MS-based PTM datasets.

Methods

PTM Set Enrichment Analysis

The protocol for PSEA directly follows the GSEA protocol described in (Subramanian et al., 2005); deviations from the original method, which result from unique consideration of MS-based PTM datasets, are highlighted in Results. This framework assumes that more than one set will be tested, details for estimation of single set enrichment are found in (Subramanian et al., 2005).

0. *Required inputs.* Analysis requires (1) a dataset $X [M \times N]$, with M PTM site measurements across N samples, (2) some ranking procedure, correlation or some other metric, to rank the PTM sites into list L according to the phenotype or treatment class profile $C [1 \times N]$, (3) a pre-defined set S of N_s PTM sites, and (4) a weighting parameter, p (described below).

1. *Calculation of enrichment score.* The enrichment score (ES) measures the degree to which the set S is concentrated at the extremes (top or bottom) of the list L. The ES is calculated from a running-sum statistic. Starting at zero and moving down the list, statistic is incremented when a member of S is encountered and decreased otherwise. The ES is the maximum deviation from zero encountered on this walk through the list, and it corresponds to a weighted Kolmogorov-Smirnov-like statistic (Figure 3.5D and E). To calculate an ES:

- a. Rank order the N PTM sites in dataset, X, to form $L = \{x_1, x_2, \dots, x_N\}$ according to some dependence measure $m(x_j) = m_j$ with profile C (phenotype / treatment condition).
- b. Calculate the fraction of PTM sites in S (“hits”) weighted by the dependence measure m_j and the fraction of PTM sites not in S (“misses”) present up to a given position i in L.

$$P_{hit}(S, j) = \sum_{x_j \in S} \frac{|m_j|^p}{N_M}, \text{ where } N_M = \sum_{x_j \in S} |m_j|^p, j = \{1, 2, \dots, M\}.$$

$$P_{miss}(S, j) = \sum_{x_j \notin S} \frac{1}{(N - N_M)}.$$

$$ES(S) = ES(S, j^*) \text{ such that } |P_{hit}(S, j^*) - P_{miss}(S, j^*)| \text{ is maximal.}$$

The ES is the maximum deviation from zero of $P_{hit} - P_{miss}$. A randomly distributed S results in a relatively small $ES(S)$. If $p = 0$, $ES(S)$ corresponds to the standard Kolmogorov-Smirnov statistic, and $ES(S)$ will be relatively large whenever S is non-randomly concentrated somewhere in the list L. However, in terms of gene expression or PTM site trends, a set concentrated in the middle of the list L would be neither negatively or positively dependent on C. To limit detection to sets concentrated at the extremes of L, p can be set so that step sizes, and thus deviations from zero, are weighted by the magnitude of the dependence measure m_j . $p = 1$ sets the step size

linearly proportional to m_j and is the commonly used default for GSEA as well as the analyses described here.

2. *Estimation of significance with MHT.* When multiple sets are tested, the estimated significance must account for MHT. Random permutation of C is used to generate null distributions of positive and negative normalized enrichment scores (NES); the normalization is required to control for differences in set size.

a. Calculate $ES(S)$ for each set of interest.

b. Generate P random permutations, π , of C . For each S and π , reorder the PTM sites in L and calculate $ES(S, \pi)$.

c. To control for differences in set size, for each S , separate the positive from the negative $ES(S, \pi)$ and normalize the positive $ES(S, \pi)$ by the absolute value of the mean of $ES(S, \pi)$, $|\mu_+|$ to yield positive $NES(S, \pi)$, and analogously calculate the negative normalization factor $|\mu_-|$ from the negative $ES(S, \pi)$ to yield negative $NES(S, \pi)$. Depending on the sign of $ES(S)$, normalize by the corresponding $|\mu_+|$ or $|\mu_-|$ to yield $NES(S)$.

d. Create separate negative and positive histograms, incorporating $NES(S, \pi)$ for all S and π . Estimate raw p-values for each $ES(S)$ using the appropriately signed histogram. Use Benjamini-Hochberg correction to control for false discovery rate (FDR).

For analyses here, $P = 500$ permutations and the weighting parameter $p = 1$ were used. PTM sets (59 in total) were hand-curated sets of phosphosites, composed of protein pathway annotations from Kegg, gene ontologies, protein structural information (PFAM and SwissPro), a curated list of PTP1b substrates (Ren et al., 2011), as well as a set of insulin-sensitive phosphorylation sites

from a previous study (Schmelzle et al., 2006). For each analysis, the lower limit on set size required membership of at least five phosphosites for set testing. Given missing data, it was not possible to rank phosphosites according to all treatments and phenotypes. Thus, the genotype and diet analyses tested 59 and 45 sets, respectively, while 56 sets were tested for relevance to phenotype prediction. A complete list of PSEA sets tested and results are available upon request.

For genotype and diet analyses, both signed and unsigned PSEA were performed. These analyses incorporated all phosphotyrosine datasets, and sites were ranked according to significance of correlation with genotype or diet. In particular, the significance and direction of correlation of genotype with a phosphosite was calculated, if possible, for the four relevant, independent data subdivisions (HFD basal, HFD insulin, NC basal, and NC insulin), then the independent p-values were combined using Fisher's method and accounting for the direction of the correlation. The same was done for diet, where the four relevant data subdivisions were basal L-PTP1b^{-/-}, insulin L-PTP1b^{-/-}, basal control, and insulin control. For signed analyses, phosphosites were ranked according to the overall direction of correlation coefficient multiplied by the $-\log_{10}$ -transformed p-value. For unsigned analyses, phosphosites were ranked by the \log_{10} -transformed p-value of correlation. In addition, both deterministic and stochastic correlation were considered. Stochastic correlation was calculated as described below, while deterministic correlation was calculated analytically.

For PSEA of sites predictive in models of steatosis, $\omega 3/ \omega 6$ PTP1b PUFA, and $\omega 7/ \omega 9$ PTP1b PUFA, signed analysis was used. The multivariate stochastic regression method described below was used to generate ranked lists of sites for analyses. Sites were ranked according the sign of the regression coefficient multiplied by the $-\log_{10}$ -transformed p-value associated with the coefficient. This analysis was limited to the HFD basal data subset only.

Stochastic correlation

Stochastic correlation uses Gaussian sampling and bootstrapping techniques to estimate a distribution of correlation coefficients. Bootstrapping mitigates effects of small sample size, while Gaussian sampling accounts for measurement error based on measurement means and standard deviations. The measurement standard deviations were estimated from technical replicates. When no technical replicate existed for a particular measurement, the standard deviation was estimated using the average relative standard deviation of the given dataset. Thus, required inputs for calculation of a pairwise correlation are vectors of mean measurement values, x and y , and corresponding vectors of standard deviations, x_{std} and y_{std} .

Let the dimensions of these vectors be $[N \times 1]$. The significance of a correlation was determined in the following way:

1. *Gaussian sampling for measurement noise.* For each data point present in x , draw a random sample, $x_{gaus}(n)$ from a Gaussian distribution with mean, $x(n)$, and standard deviation, $x_{std}(n)$, $\{n = 1, 2, \dots, N\}$. Similarly, draw $y_{gaus}(n)$, based on $y(n)$ and $y_{std}(n)$.
2. *Draw a bootstrap sample.* Randomly select paired entries of x_{gaus} and y_{gaus} without replacement and build x_{boot} and y_{boot} vectors, dimensions $[N \times 1]$.
3. Calculate and record the correlation between x_{boot} and y_{boot} .
4. Repeat 1-3 for desired number of iterations.
5. Estimate significance of the p-value from the histogram of simulated correlation coefficients.

For analyses here, the maximum number of iterations was 10,000.

Stochastic multivariate regression

Similarly to stochastic correlation, stochastic multivariate regression (SMR) involves both bootstrapping and Gaussian sampling steps to construct distributions of regression coefficients. Required algorithm inputs include independent variable matrices, X [$N \times M$] and X_{std} [$N \times M$], with N samples and M variables measured, and a dependent variable matrices, Y [$N \times V$] and Y_{std} [$N \times V$], where V is the number of dependent variables measured. The matrices X and Y correspond to measurement means, while X_{std} and Y_{std} contain measurement standard deviation. In the absence of a technical replicate for a given measurement, the standard deviation was estimated using the average relative standard deviation of the corresponding dataset.

If X corresponds to an MS-based PTM dataset, then X is likely to have dimensions such that $M > N$. While there are several multivariate regression algorithms that work in this regime and SMR could be implemented using any of them, partial least squares regression (PLSR) was selected here, mainly because its execution is fast, especially using Matlab software. Building of a PLSR model involves the simultaneous decomposition of both X and Y matrices into X and Y scores and loading matrices. Assuming X and Y are mean-centered, the key PLSR matrix relationships are described by the following equations:

$$X = TP^T, T = X P / (P^T P) = XW^T,$$

$$Y = UC^T \approx TC^T = XW^T C^T = XB,$$

where the X and Y scores matrices, T [$N \times \text{pcs}$] and U [$N \times \text{pcs}$], are optimized to covary maximally; pcs is the number of model principle components, and P [$M \times \text{pcs}$], C [$P \times \text{pcs}$], W [$M \times \text{pcs}$], and B [$M \times P$] are the X loadings, Y loadings, X weights, and regression coefficient matrices, respectively. The decomposition into X and Y determines the location of a PLSR

principle component plane. This plane is optimized to contain linear combinations of independent variables that covary maximally with the dependent variable(s), and, for this reason, this algorithm tends to perform well in a variety of contexts.

The basic stochastic multivariate regression algorithm is as follows:

1. *Gaussian sampling for measurement noise.* For each data point present in X , draw a random sample, $X_{\text{gaus}}(n,m)$ from a Gaussian distribution with mean, $X(n,m)$, and standard deviation, $X_{\text{std}}(n,m)$, $n = 1,2,\dots,N$, $m = 1,2,\dots,M$. Similarly, draw $Y_{\text{gaus}}(n,v)$, based on $Y(n,v)$ and $Y_{\text{std}}(n,v)$, $n = \{1,2,\dots,N\}$, $v = \{1,2,\dots,V\}$.
2. *X and Y Data normalization and averaging of Y matrix measurements.* For each column of X_{gaus} or Y_{gaus} , mean-center the data and divide by the standard deviation to yield matrices X_{zscore} and Y_{zscore} . If $V > 1$, average Y_{zscore} column-wise to yield $Y_{\text{zscore},v}$ [$N \times 1$]. (Averaging Y simplifies subsequent model reduction steps.)
3. *Draw a bootstrap sample.* Randomly select paired rows of X_{zscore} and $Y_{\text{zscore},v}$ without replacement and build X_{boot} and Y_{boot} matrices, dimensions [$N \times M$] and [$N \times 1$], respectively.
4. *Build the regression model.* Use partial least squares regression (PLSR), with number of principle components equal to one minus the rank of X_{boot} , to estimation the linear regression coefficients B_{boot} [$M \times 1$].
5. *Multiple imputation and Gaussian sampling.* Repeat steps 1-4 for J iterations to create a distribution of the model coefficients. Store each vector of coefficients B_{boot} in the matrix B_{matrix} [$J \times M$].

6. *Estimate the significance of each independent variable measurement to prediction.* For each measurement m in X , use the corresponding coefficient distribution contained in column m of B_{matrix} , to empirically test the null hypothesis that zero belongs to this distribution.

7. *Model reduction.* Rank the independent variables according to coefficient significance determined in 6. Sequentially build and evaluate PLSR models, by varying (1) the number of principle components, pcs , in the model and (2) by incorporating the top F ranked variables in the model. For this analysis, $pcs = \{1, \dots, 4\}$, and $F = \{1, 2, \dots, 20\}$.

R^2 and Q^2 statistics (Q^2_f and $Q^2_{v,f}$ calculations). Models were evaluated by estimating R^2 , goodness-of-fit, and Q^2 , goodness-of-prediction, statistics:

$$R^2 = 1 - \frac{\sum_{n=1}^N (Y^{\text{meas},n} - Y^{\text{pred},n})^2}{\sum_{n=1}^N (Y^{\text{meas},n} - Y^{\text{mean}})^2}$$

$$Q^2 = 1 - \frac{\sum_{n=1}^N (Y^{\text{meas},n} - Y^{\text{pred,loocv}(n)})^2}{\sum_{n=1}^N (Y^{\text{meas},n} - Y^{\text{mean,loocv}(n)})^2}$$

where $Y^{\text{meas},n}$ corresponds to the experimentally determined response measurement n , Y^{mean} is the average of the response measurements, $Y^{\text{pred},n}$ is the model prediction for response measurement n , $Y^{\text{pred,loocv}(n)}$ is the model prediction for response n , from a leave-one-out-cross-validation (LOOCV) model, which excludes measurements for sample n , and, similarly, $Y^{\text{mean,loocv}(n)}$ is the mean of the response measurements, excluding measurement n . A Q^2 value should estimate the variance explained by a model for an independent sample. Typical Q^2 calculations for PLSR involve (1) determination of a reduced set of independent variables for model reduction using the full dataset and (2) fitting LOOCV models using that pre-determined, reduced set of variables with LOOCV data subsets (Huang et al., 2010; Janes et al., 2004; Kumar et al., 2007). Such a protocol inflates the Q^2 value, because the left-out sample is not entirely independent of the

LOOCV model construction. Although the sample left out is not involved in model fitting, it was used for the model reduction/variable selection step. In consideration of this, for the calculation of Q^2 values in this analysis, the model reduction procedure was included in the LOOCV modeling building. Thus, both model reduction and model fit were determined based on the LOOCV data only. In the text, these methods are referred to as Q^2_f (LOOCV fitting) and $Q^2_{v,f}$ (LOOCV variable selection and fitting).

Variation: Model reduction based on VIP-score. In addition, reduced models were built by a second method, in which independent variables were ranked according to variable importance of projection (VIP) score. The VIP score for a given independent variable and PLSR model with pcs principle components is:

$$VIP(m, pcs) = \sqrt{\frac{\sum_{pc=1}^{pcs} Var_{exp,Y}(pc)W(m, pc)^2}{\sum_{pc=1}^{pcs} Var_{exp,Y}(pc)}}$$

where $Var_{exp,Y}(pc)$ is the variance in Y explained by regression onto a particular principle component, pc, and $W(m,pc)$ is the quantitative contribution that a particular measurement, m, makes to the pcth PLSR principle component. Because the PLSR principle component plane is optimized to contain linear combinations of independent variable measurements that covary maximally with the dependent variable, this metric performs well in model reduction.

To implement SMR with this method, at step 4, the X loadings matrix P_{boot} [$M \times pcs$] were stored in a 3D-matrix P_{matrix} [$J \times M \times pcs$] (step 5), and then used to calculate variable importance of projection metrics for each of the M independent variable measurements (step 7). Specifically, the loadings matrices were averaged across the iterations' dimension of P_{matrix} to create P_{ave} [$M \times pcs$]. Thus, by averaging, variable measurements that consistently contribute to

the PLSR principle component plane will have higher loading values and contribution to the model more than measurements that are randomly imputed or associated with larger amounts of error. To make this loadings matrix consistent with loadings matrix produced directly from PLSR algorithms, P_{ave} was orthonormalized row-wise. Next, the scores matrix, T [$N \times pcs$], was estimated by the equation $T = X_c P_{ave}^T$ and the weights, W , were estimated as $P_{ave} / (P_{ave}^T P_{ave})$. The rows of T were orthogonalized, and, then the Y loadings, C [$pcs \times 1$] were calculated from $C = (T^T T)^{-1} T^T Y^*$, where Y^* is a vector average of the z-scored Y matrix (similar to $Y_{zscore,v}$ in step 2). Next, the $Var_{exp,Y}(pc)$, could be calculated using the equation: $Var_{exp,Y}(pc) = 1 - (Y^* - T(pc)^T C(pc))^T (Y - T(pc)^T C(pc)) / Var_Y$, where $T(pc)$ is the pc^{th} column of T , $C(pc)$ is the pc^{th} row of C transposed to a column vector, and Var_Y is the variance of Y^* . Thus, after a fair amount of linear algebra, the VIP score can be calculated for each measurement and total number of pcs in the model. Step 7 is then performed based on VIP-score model reduction.

Missing data techniques for stochastic multivariate regression

Missing data methods and their implementation are described here. The data matrices refer to matrices described in the SMR section.

1. *List-wise deletion.* Delete each column of X and X_{std} , corresponding to measurements for variable m , that contains one or more missing sample observation. Proceed with SMR.
2. *Merge complete subset models.* This method was specific to the HFD basal data subset of phosphotyrosine measurements, X_{HFDb} , and X_{HFDb} was divided into three distinct blocks of complete data subsets, as shown in Supplementary Figure 3.1.

SMR was used to build individual models for each of the subsections, referred to as the “Complete” (data for all sample conditions), “Run 1”, and “Run 2” subset models. (“Run 1” and

“Run 2” models each had data for 8 out of 13 conditions.) The results from the three models were combined using two sets of rules. The first rule is referred to as “All subsets agree”, as a variable was incorporated into a final model only if it was significant to sub-models from all data subsets containing that measurement. Specifically, select a variable for the final model only if the significance of its regression coefficient, p_m , is less than a pre-determined significance cutoff, $p_{\text{cutoff}} \in [0,1]$, for all models incorporating that measurement. The second rule is referred to as “Complete subset rules”, as the “Complete” subset model takes priority in determining variable significance and final model inclusion. If the variable is in the complete subset, select it for the final model if $p_m < p_{\text{cutoff}}$. Otherwise, if the variable is in “Run 1” or “Run 2” subset only, retain this variable in the model with the same criteria: if $p_m < p_{\text{cutoff}}$.

3. *Multiple Gaussian imputation.* This method involves replacing missing data points so that the data matrix can be used for SMR directly; it was implemented in three ways. The first is termed “general” Gaussian imputation. For this method, the missing data points for a specific measurement (column of X) are estimated by a single random Gaussian sample, whose mean and standard deviation are estimated from the samples present in the measurement column. The second is termed “treatment” Gaussian imputation, as this method incorporates information about the sample treatment condition associated with a missing data point. First, missing data points for a measurement (column of X) are separated according to treatment condition (i.e. L-PTP1b-/- or control), and their values are imputed by drawing a Gaussian sample, whose mean and standard deviation correspond to observed data points of the corresponding treatment condition. The third method is termed “pval-treatment” Gaussian imputation. It is a mix of general and treatment Gaussian imputation, in which the treatment-dependent method is selected for a particular measurement if the null hypothesis of a two-sided T Test (comparing control to

L-PTP1b^{-/-} phosphorylation levels) is rejected at an α -level of significance. Otherwise, general Gaussian imputation is used. Both implementations can be inserted between steps 1 and 2 of the SMR protocol, using X_{gaus} as X (**Figure 3.7B**).

4. *Multiple random imputation.* This method is the random analog of multiple Gaussian imputation, and was also implemented in “general”, “treatment-specific”, and “pval-treatment” ways. For “general” random imputation, missing data points for a measurement are randomly replaced with the value of a randomly selected available data point for that measurement. For “treatment” random imputation, missing data measurements are separated according to treatment condition, and their missing values are replaced with a randomly selected observed data point from the matching treatment condition. The “pval-treatment” method selects “general” or “treatment-specific” imputation based on a T test comparison of L-PTP1b^{-/-} and control phosphorylation levels, as described above. These procedures can also be inserted between steps 1 and 2 of the SMR protocol.

Evaluation of missing data techniques

Simulation of missing data. To evaluate missing data techniques specifically for SMR with the HFD basal phosphotyrosine dataset as the independent variable matrix, X , we utilized the complete data subset of X . The full dataset (**Figure 3.7A**) contained 228 phosphosite measurements across 13 samples, while the complete subset of the dataset contained 76 phosphosite measurements across the same number of samples. We simulated 10 incomplete datasets from the complete data subset, by randomly removing sections of phosphosite data so that, structurally, the simulated incomplete datasets corresponded to the original dataset. In particular, for randomly selected measurement columns in X , we ensured that 33% had complete

data, 45% had data missing for 5 samples, 17% had data missing for 5 other samples, and 5% of the columns had data missing for 7 of the 13 samples.

Construction of ROC curves. The complete data subset was used as the gold standard, and five “gold standard” models were built using SMR for PTP1b expression, steatosis, "SCD PUFA", stress, and "Tg PUFA" phenotypes. As described in the SMR protocol, final models were selected by (1) ranking all the phosphosite measurements according to regression coefficient significance and (2) selecting the optimal number of measurements as judged by predictive Q^2 value. Those measurements selected for a final model were true positives, and those measurements excluded from a model represented true negatives.

The seven missing data methods (described above) were used to build the five phenotype models from each of the 10 simulated datasets, resulting in a total of 50 models per technique. The ROC curve, true positive rate (TPR) as a function of false positive rate (FPR), is a parametric function of a decision rule. We applied the following decision rule here: include a measurement in the final model if the significance of the associated regression coefficient is less than p_{cutoff} . By varying p_{cutoff} from 0 to 1, we were able to count true positives and false positives relative to the complete data (gold standard) models. Using this framework, we also tested the performance of stochastic correlation (by using the correlation-coefficient significance in lieu of regression coefficient).

Final models

Individual models of steatosis, $\omega 3/ \omega 6$ PTP1b PUFA, and $\omega 7/ \omega 9$ PTP1b PUFA as functions of the complete HFD basal phosphotyrosine dataset [228 phosphosites X 13 samples] were

constructed using SMR and treatment-dependent random imputation; these models were built using 1, 2, and 1 principle component(s), respectively.

Calculation of R^2 and Q^2 . Estimation of any set of model predictions for the response vector Y , using the relationship $Y = XB$, requires a complete set of predictor variables X . Thus, to estimate R^2 and Q^2 , a complete matrix, X_c was created using column-wise treatment-dependent mean imputation. Specifically, for each measurement (column) of X_c , the missing data points for L-PTP1b^{-/-} (or control) samples were estimated as the mean of those L-PTP1b^{-/-} (or control) samples observed for that measurement. Final model parameters, pcs and F, were determined based on optimization of Q^2 values. Estimation of the error associated with the Q^2 values was accomplished by leave-two-out cross-validation.

Software

All computational analyses were performed in Matlab Student Version 7.10.0.499 (R2010a).

References

- Akavia, U.D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H.C., Pochanard, P., Mozes, E., Garraway, L.A., and Pe'er, D. (2010). An integrated approach to uncover drivers of cancer. *Cell* 143, 1005-1017.
- Delibegovic, M., Zimmer, D., Kauffman, C., Rak, K., Hong, E.G., Cho, Y.R., Kim, J.K., Kahn, B.B., Neel, B.G., and Bence, K.K. (2009). Liver-specific deletion of protein-tyrosine phosphatase 1B (PTP1B) improves metabolic syndrome and attenuates diet-induced endoplasmic reticulum stress. *Diabetes* 58, 590-599.
- Huang, P.H., Miraldi, E.R., Xu, A.M., Kundukulam, V.A., Del Rosario, A.M., Flynn, R.A., Cavenee, W.K., Furnari, F.B., and White, F.M. (2010). Phosphotyrosine signaling analysis of site-specific mutations on EGFRvIII identifies determinants governing glioblastoma cell growth. *Mol Biosyst* 6, 1227-1237.
- Huang, S.S., and Fraenkel, E. (2009). Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci Signal* 2, ra40.
- Jaimovich, A., Elidan, G., Margalit, H., and Friedman, N. (2005). Towards an integrated protein-protein interaction network (Springer).

- Janes, K.A., Kelly, J.R., Gaudet, S., Albeck, J.G., Sorger, P.K., and Lauffenburger, D.A. (2004). Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J Comput Biol* 11, 544-561.
- Joughin, B.A., Naegle, K.M., Huang, P.H., Yaffe, M.B., Lauffenburger, D.A., and White, F.M. (2009). An integrated comparative phosphoproteomic and bioinformatic approach reveals a novel class of MPM-2 motifs upregulated in EGFRvIII-expressing glioblastoma cells. *Mol Biosyst* 5, 59-67.
- Kumar, N., Wolf-Yadlin, A., White, F.M., and Lauffenburger, D.A. (2007). Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. *PLoS computational biology* 3, e4.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34, 267-273.
- Naegle, K.M., Welsch, R.E., Yaffe, M.B., White, F.M., and Lauffenburger, D.A. (2011). MCAM: multiple clustering analysis methodology for deriving hypotheses and insights from high-throughput proteomic datasets. *PLoS Comput Biol* 7, e1002119.
- Nelson, P.R.C., Taylor, P.A., and MacGregor, J.F. (1996). Missing data methods in PCA and PLS: Score calculations with incomplete observations. *Chemometrics and intelligent laboratory systems* 35, 45-65.
- Ren, L., Chen, X., Luechapanichkul, R., Selner, N.G., Meyer, T.M., Wavreille, A.S., Chan, R., Iorio, C., Zhou, X., Neel, B.G., *et al.* (2011). Substrate specificity of protein tyrosine phosphatases 1B, RPTPalph, SHP-1, and SHP-2. *Biochemistry* 50, 2339-2356.
- Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 473-489.
- Schmelzle, K., Kane, S., Gridley, S., Lienhard, G.E., and White, F.M. (2006). Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 55, 2171-2179.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15545.
- Thomas, S.M., and Brugge, J.S. (1997). Cellular functions regulated by Src family kinases. *Annu Rev Cell Dev Biol* 13, 513-609.
- Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A., and White, F.M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* 104, 5860.
- Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J., Zhou, L., Zeng, Y., and Li, H. (2010). Regulation of cellular metabolism by protein lysine acetylation. *Science* 327, 1000.

Figures

Figure 3.1 Noise and Missing Data in L-PTP1b^{-/-} Phosphotyrosine Data

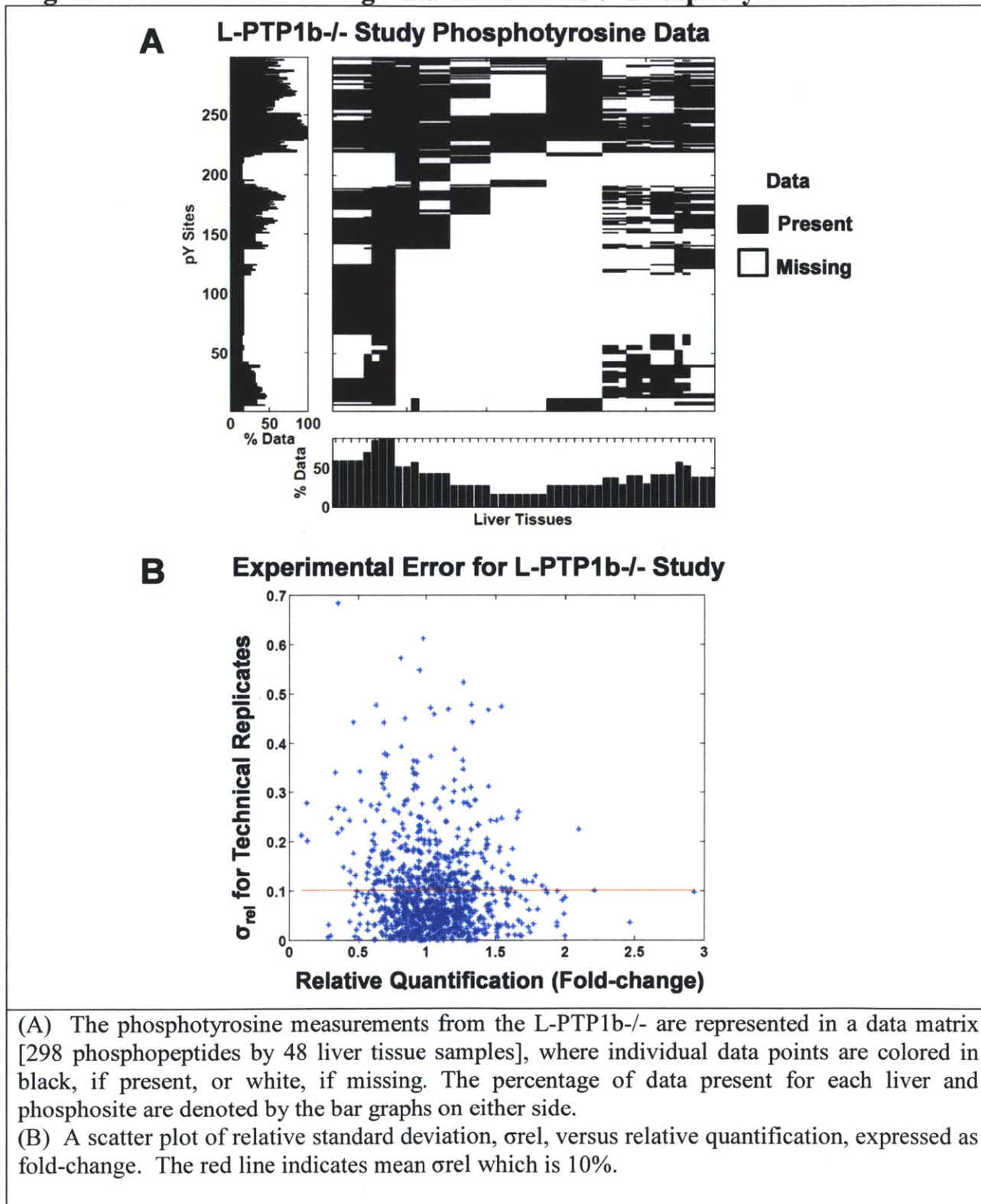


Figure 3.2 Stochastic Correlation

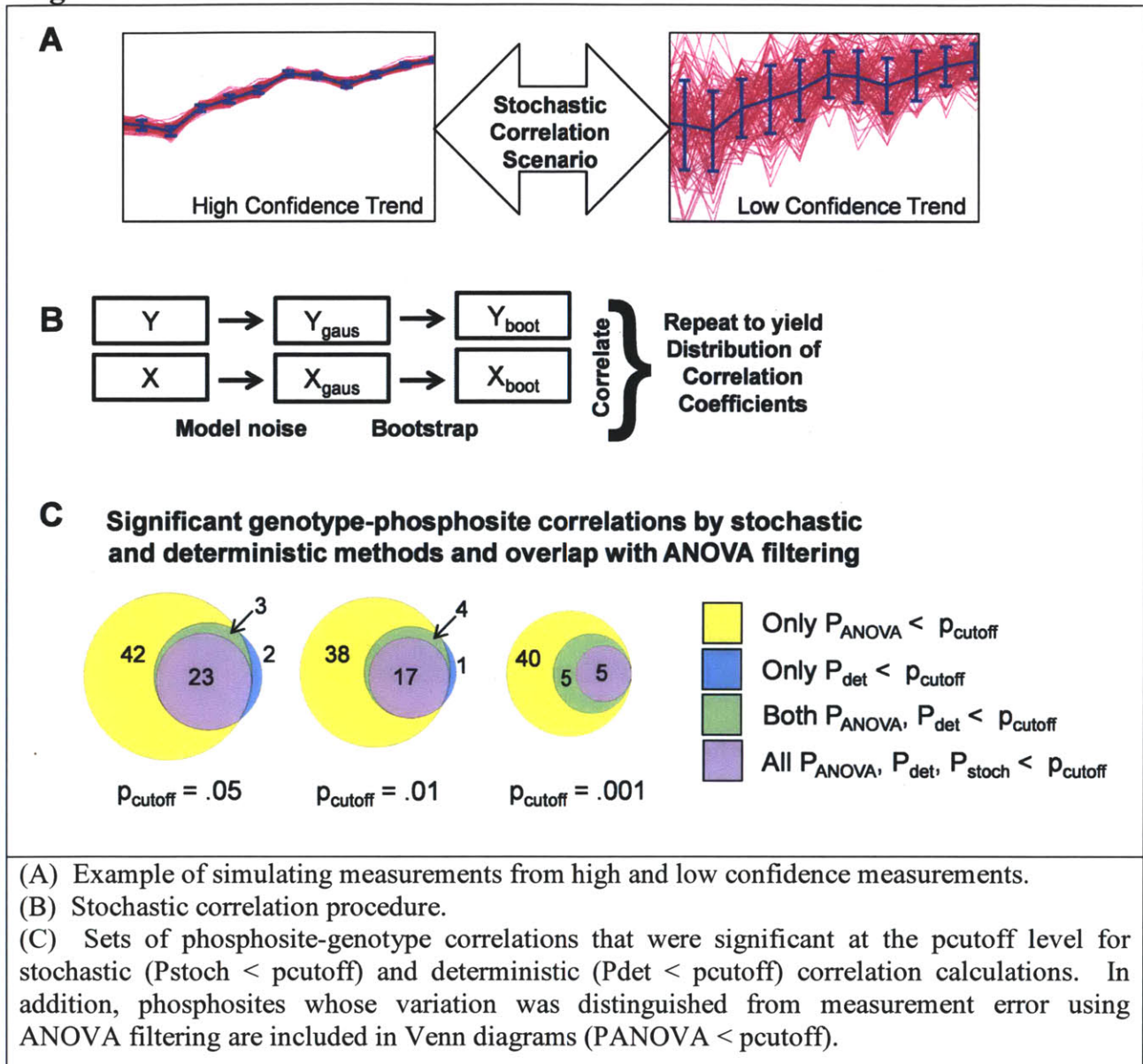


Figure 3.3 Comparison of Correlation P-values

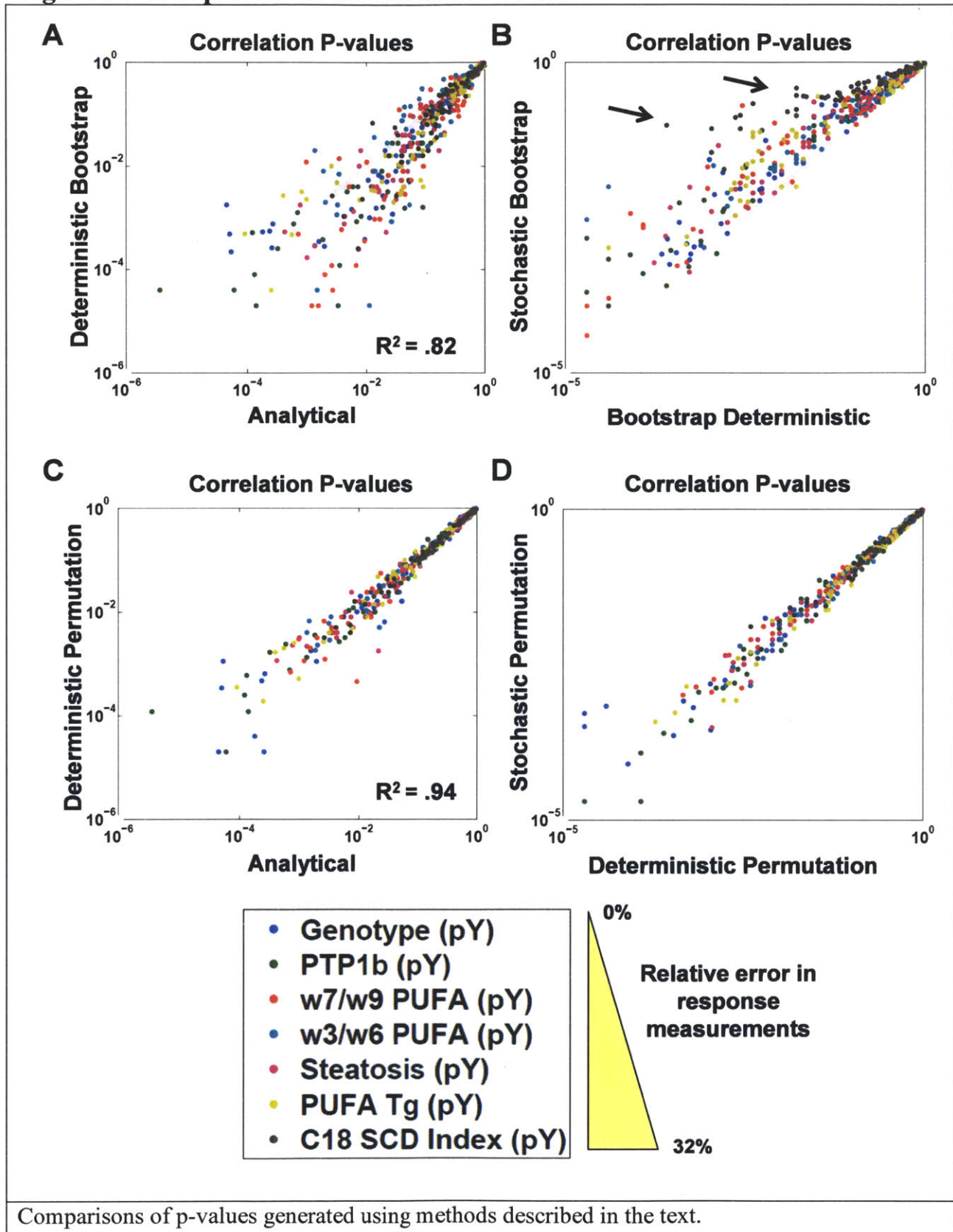
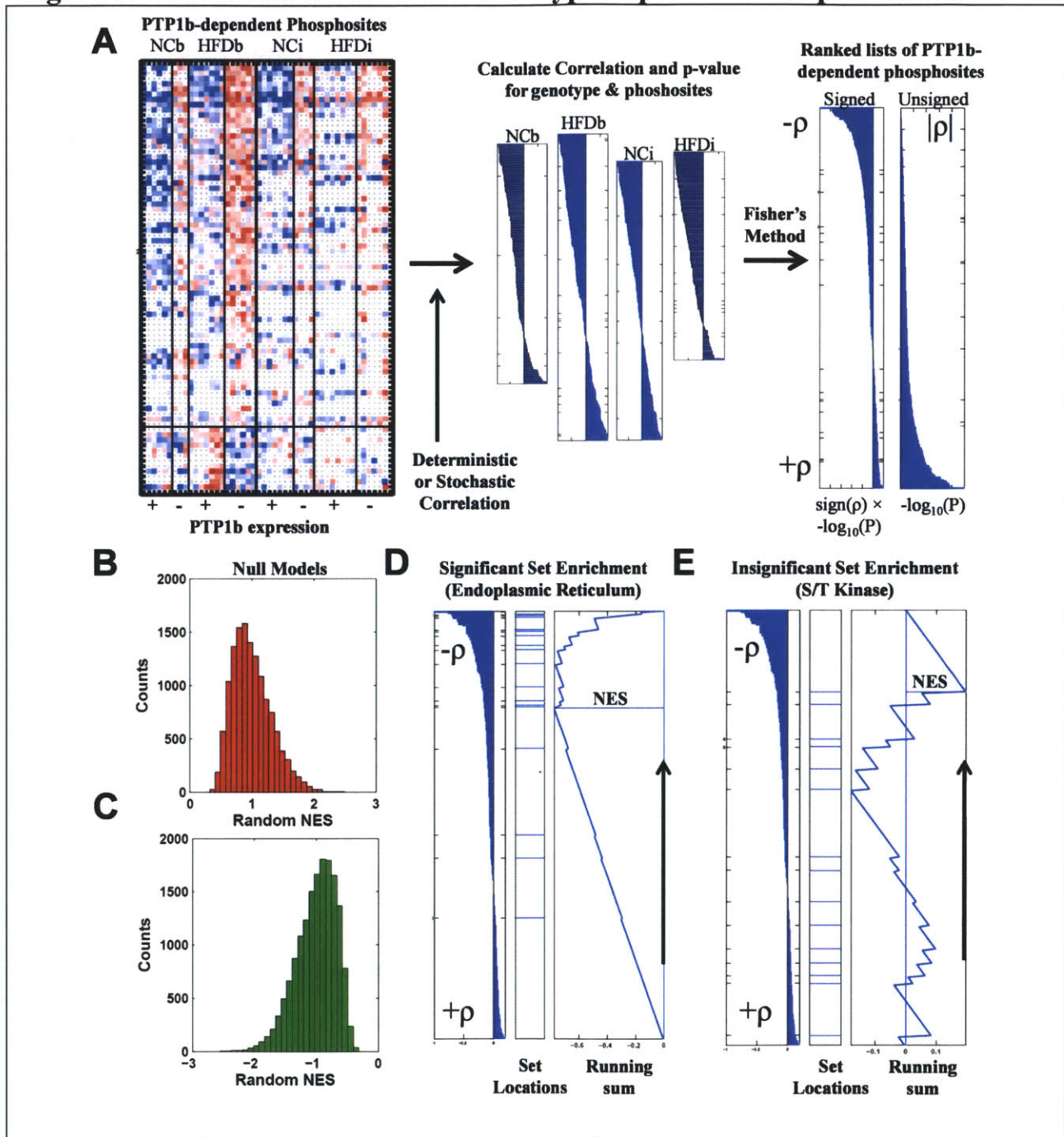


Figure 3.4 Schematic of PSEA for Genotype-dependent Phosphosites



(A) Supervised clustering of the L-PTP1b^{-/-} dataset according to genotype is displayed in a heatmap to show the four PTP1b-dependent subdatasets (NCb, HFDb, NCi, and HFDi). For PSEA of PTP1b-dependent sites, all phosphosites in each independent subdataset were ranked using eight deterministic or stochastic correlation. P-values from each subdataset were combined using Fisher's Method to generate a combined signed and unsigned ranked lists of PTP1b-dependent phosphosites.

(B) and (C) The positive and negative NES distributions for the random model.

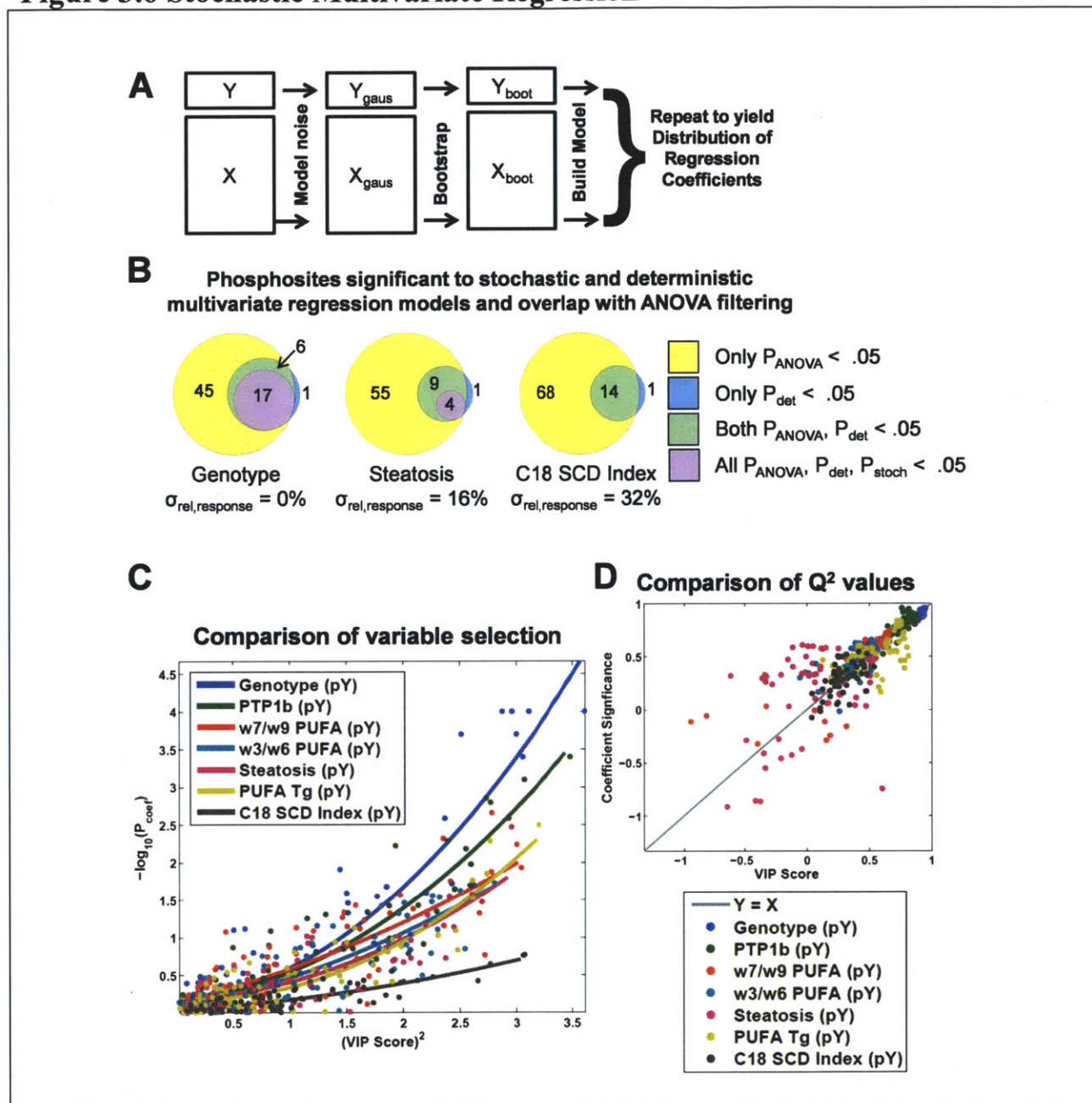
(D) and (E) Examples of significantly and insignificantly PTP1b-dependent sets, respectively.

Figure 3.5 Results of PSEA for Genotype-dependent Phosphosites

Signed		Unsigned	
Deterministic	Stochastic	Deterministic	Stochastic
Endoplasmic Reticulum	Endoplasmic Reticulum	Endoplasmic Reticulum	Endoplasmic Reticulum
Cytosol	Cytosol	Cytosol	Cytosol
Metabolic Process	Metabolic Process	Metabolic Process	Metabolic Process
Oxidation Reduction	Oxidation Reduction	Oxidation Reduction	Oxidation Reduction
Insulin signaling pathway	Tight junction	Mitochondria	Glycolysis/ Gluconeogenesis
Lipid metabolism	Lipid metabolism	Golgi apparatus	Lipid metabolism
Tight junction	PDZ	Lipid metabolism	Arginine and Proline Metabolism
Fatty acid metabolism	Arginine and Proline Metabolism	Fatty acid metabolism	Golgi apparatus
Neurotrophin signaling pathway		Arginine and Proline Metabolism	
Arginine and proline metabolism		Glycolysis/ Gluconeogenesis	
Mitochondria		Insulin Signaling Pathway	
Phenylalanine metabolism		Phenylalanine Metabolism	
Glycolysis/ Gluconeogenesis		Insulin-sensitive pYs in adipocytes	
Golgi apparatus			
Insulin-sensitive pYs in adipocytes			
Cysteine and methionine metabolism			

Results from PSEA analysis of PTP1b^{-/-} dependent sites for signed or unsigned and stochastic or deterministic correlation rankings. Sets are color-coded according to significance: yellow, pink, and cyan correspond to FDR of .01, .05, and .1, respectively.

Figure 3.6 Stochastic Multivariate Regression



(A) Stochastic multivariate regression (SMR) procedure.

(B) Sets of phosphosites significant to three models based on the significance of their associated multivariate regression coefficients, as determined by stochastic ($P_{stoch} < .05$) and deterministic ($P_{det} < .05$) multivariate regression methods. In addition, phosphosites whose variation was distinguished from measurement error using ANOVA filtering are included in Venn diagrams ($P_{ANOVA} < .05$).

(C) Phosphosites are plotted according to predictive power, according to $(VIP\ Score)^2$ and $-\log_{10}(p\text{-value})$ associated with the regression coefficient.

(D) Q^2 values are compared for models built on the top 1-20 variables, as ranked by the indicated model reduction method.

Figure 3.7 Schematic for Testing Missing Data Methods

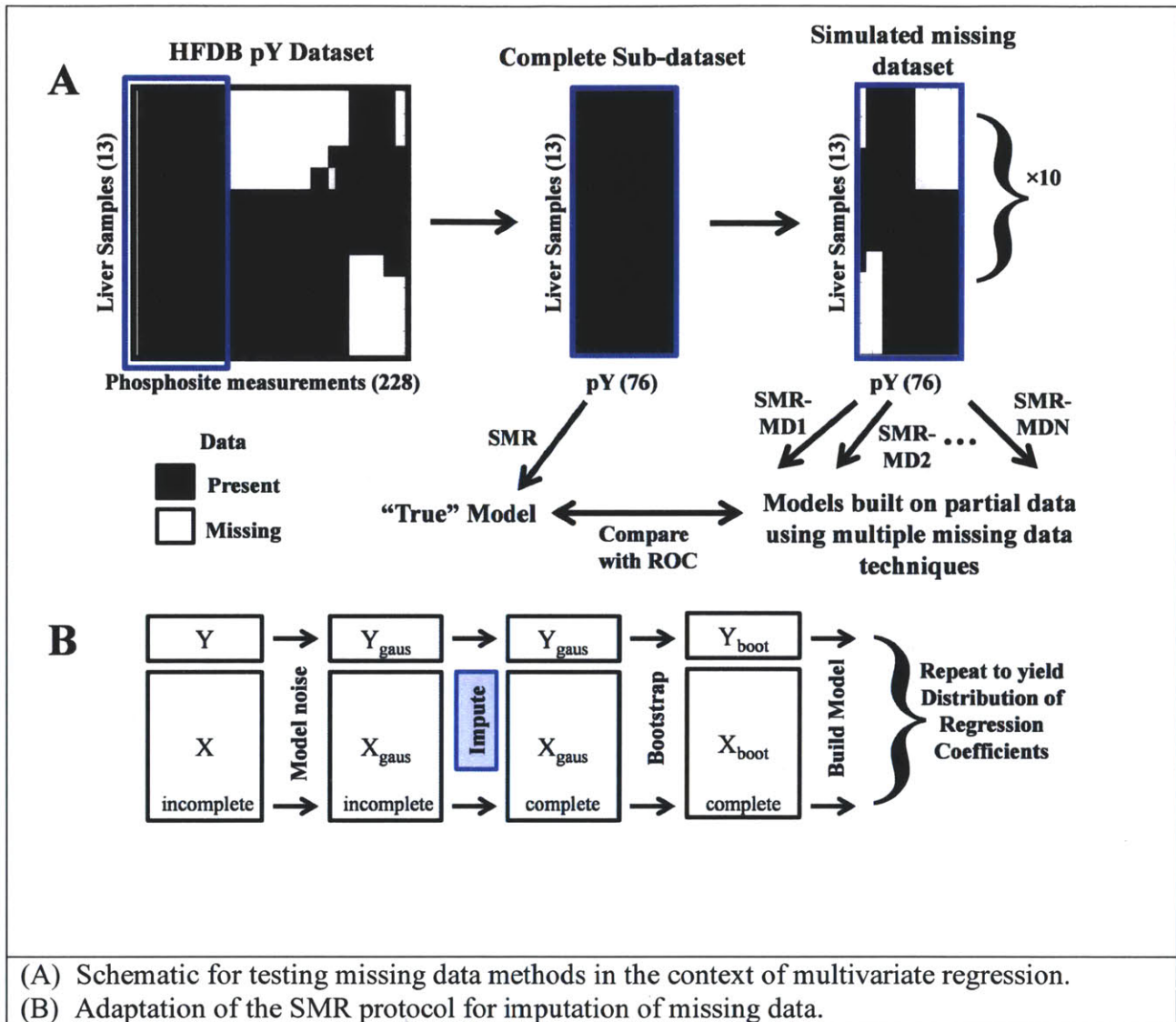


Figure 3.8 ROC Performance of Missing Data Methods

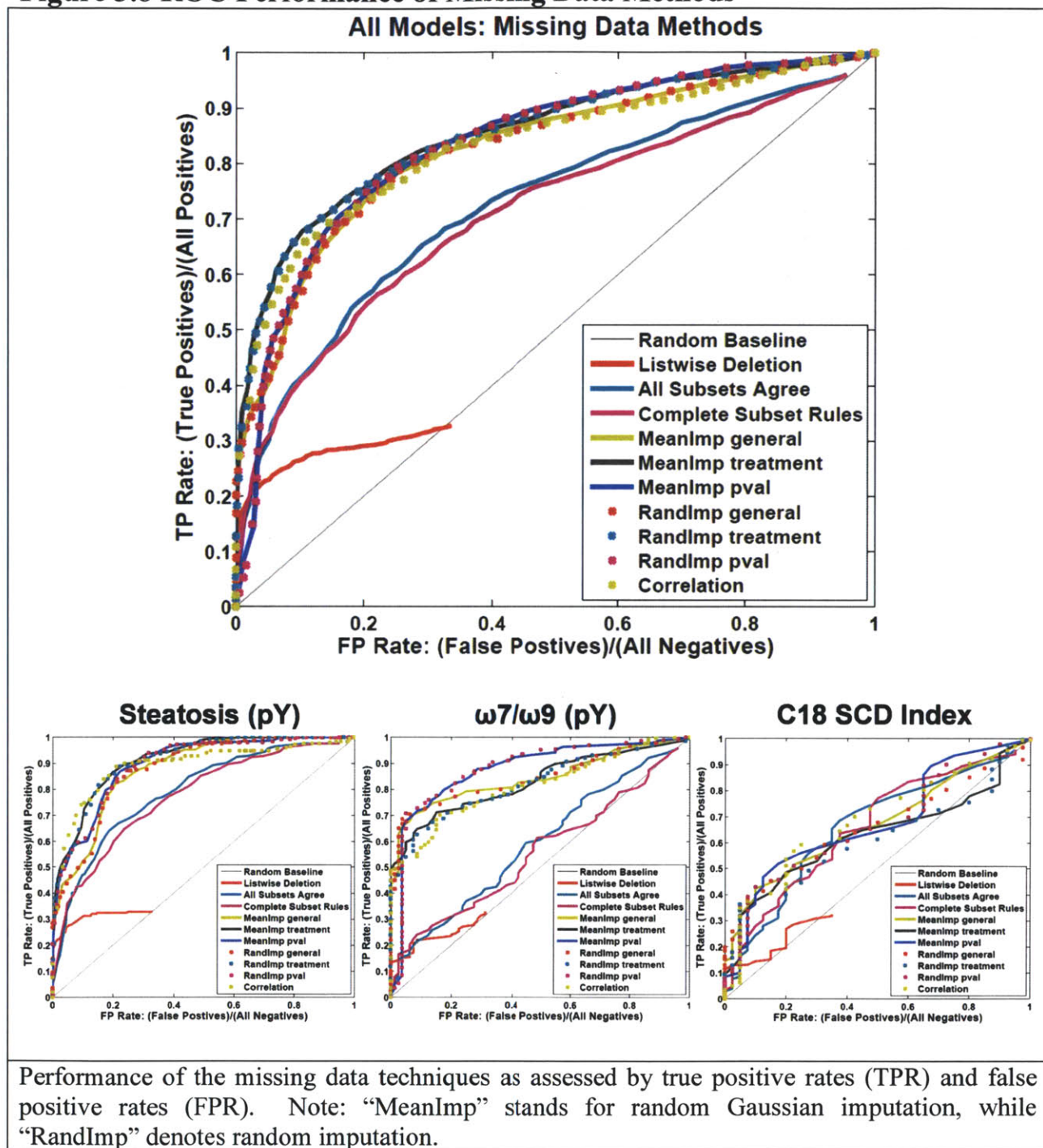


Figure 3.9 SMR Model Prediction

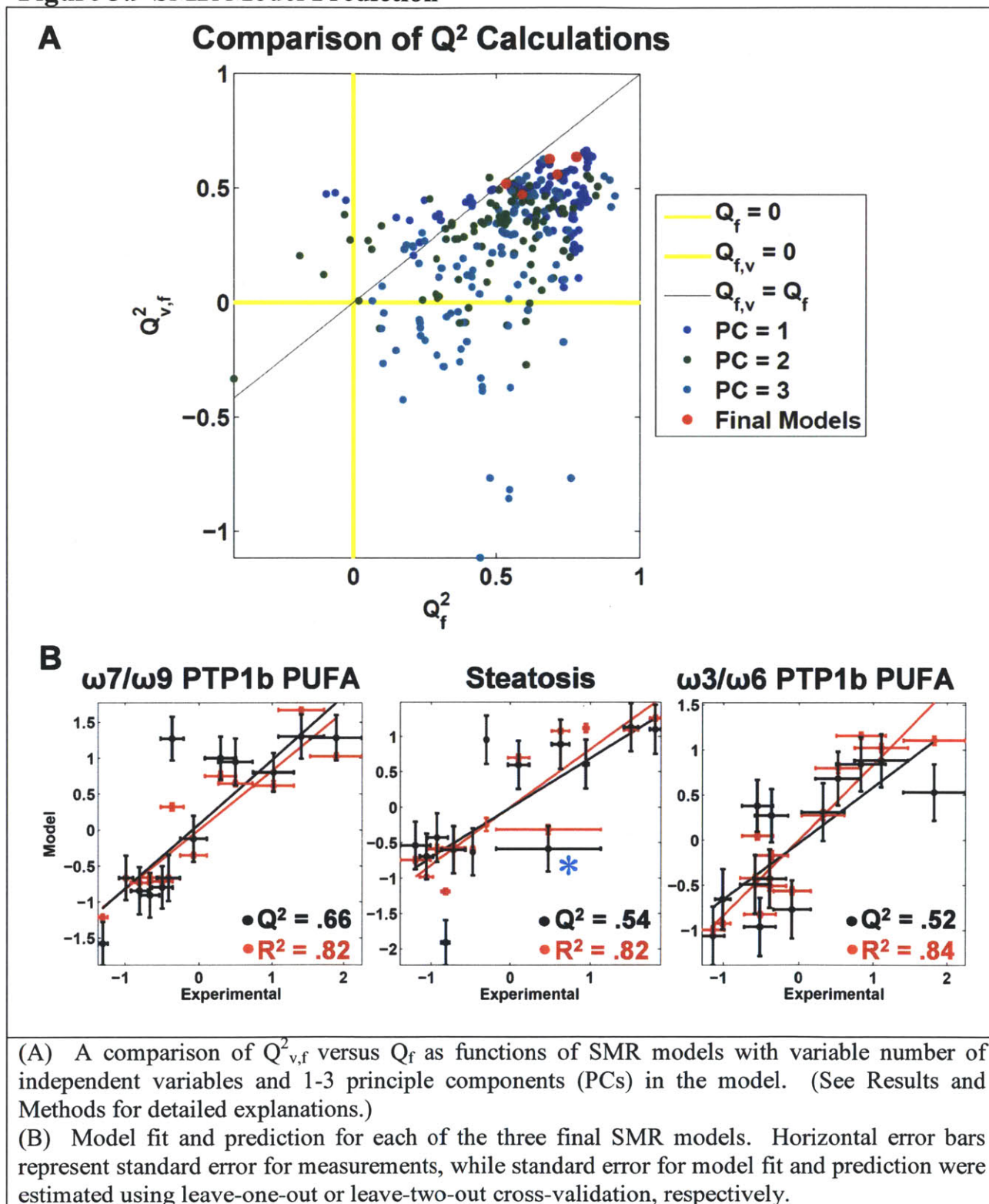
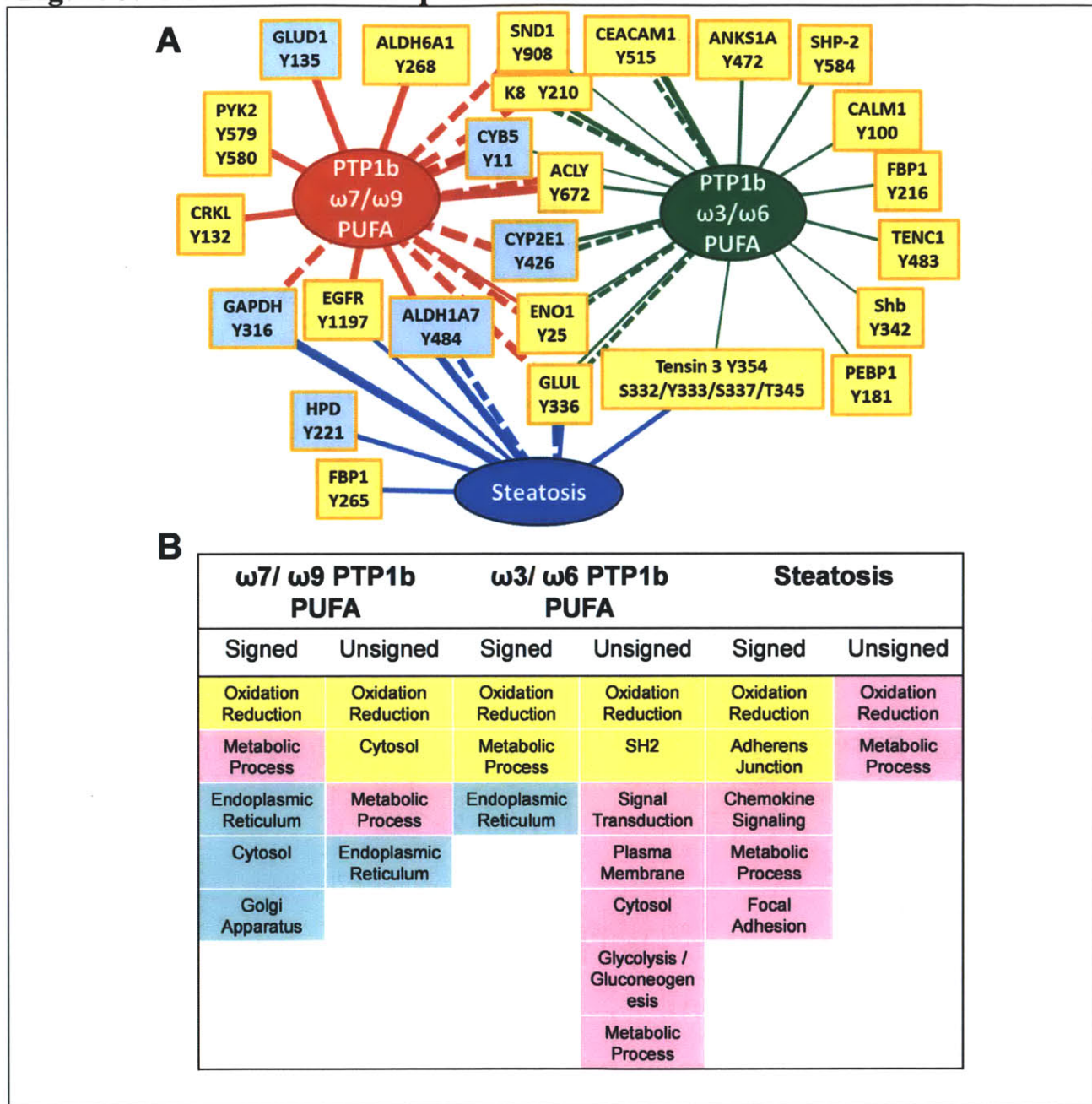


Figure 3.10 SMR Model Components and PSEA

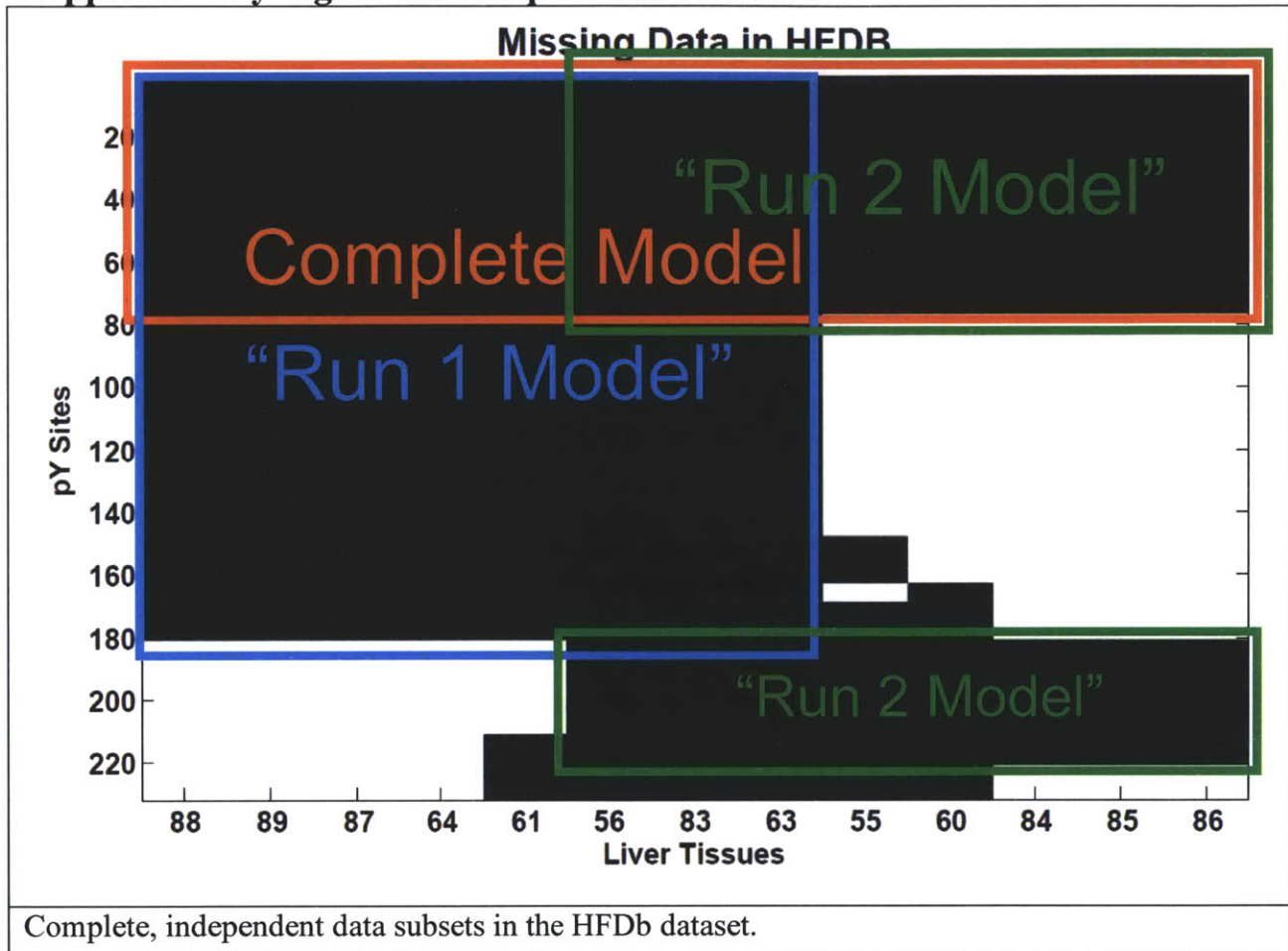


(A) The phosphosite predictors included in final multivariate regression models are attached to phenotype nodes by colored edges: $\omega 7 / \omega 9$ PTP1b PUFA (red), $\omega 3 / \omega 6$ PTP1b PUFA (green), and steatosis (blue). Models were built using two methods; solid and dashed lines indicate that the predictor was selected by coefficient-significance or VIP-score, respectively. The width of an edge is proportional to the importance of the phosphosite to the model. Light-green shading highlights phosphosites on REDOX proteins.

(B) Results from PSEA analysis of the three SMR models built on the phosphotyrosine dataset. Sets are color-coded according to significance: yellow, pink, and cyan correspond to FDR of .01, .05, and .1, respectively.

Supplementary Figure

Supplementary Figure 3.1 Complete Data Subsets



4. Conclusions

Contributions

Protein-tyrosine phosphatase 1b (PTP1b) is a key modulator of metabolism. Its expression is modulated by nutrient cues, and its activity is differentially regulated by phosphorylation upon insulin or epidermal growth factor (EGF) stimulation (Lessard et al., 2010). Mouse models of PTP1b inhibition (whole-body knock outs, tissue-specific deletion, oligonucleotide treated) in a variety of backgrounds demonstrate the physiological scale of metabolic alterations due to this perturbation (Banno et al., 2010; Bence et al., 2006; Delibegovic et al., 2007; Delibegovic et al., 2009; Elchebly et al., 1999; Zinker et al., 2002). Even in the isolated context of tissue-specific deletion, these studies suggest therapeutic advantage in PTP1b inhibition for the treatment of metabolic syndrome, T2DM, and other obesity-related diseases. In these studies, antibody-based measurements have characterized phosphotyrosine changes in select canonical insulin and/or leptin signaling nodes. These measurements suggest that alterations in the phosphotyrosine network are not resolved during the course of chronic PTP1b inhibition and potentiate PTP1b-dependent phenotypes. However, insulin and very conceivably leptin signaling involve phosphotyrosine modification of many more proteins than could be measured by available antibodies (Schmelzle et al., 2006). Furthermore, some PTP1b-dependent phenotypes, such as alterations in lipid metabolism and ER stress in the liver-specific PTP1b deletion (L-PTP1b^{-/-}) mice, are not associated with canonical leptin or insulin signaling, and the phosphotyrosine-modulated pathways underlying these phenotypes are unknown. The current therapeutic interest in PTP1b inhibition underscores the need for comprehensive characterization of chronic, PTP1b-

dependent phosphotyrosine alterations *in vivo*. This information will be critical to assessing the physiological consequences of inhibiting this pleiotropic phosphatase.

To globally detect, explore, and quantify phosphotyrosine network changes upon PTP1b inhibition, we leveraged MS-based techniques (Zhang et al., 2005) to site-specifically characterize the phosphotyrosine networks in L-PTP1b^{-/-} mice and control under high fat diet (HFD) or normal chow (NC) conditions and insulin-stimulated (insulin) or basal conditions. Given the discovery-based nature of this analysis, fifty mice were included in the study, roughly three to eight mice per treatment condition, to enable statistical and computational analyses for the detection of phosphotyrosine-regulated, PTP1b-dependent pathways and phosphosite drivers of PTP1b phenotypes.

To our surprise, principle component analysis (PCA) of the resulting basal and insulin phosphotyrosine datasets revealed that PTP1b deletion and diet were the most important sources of variance in the phosphotyrosine network on a global scale *in vivo*. This result emphasizes the severity of both perturbations to phosphotyrosine regulation, the consequences of which merit evaluation in diverse contexts. Foremost, many cancers exhibit dysregulation of phosphotyrosine networks, often leading to a general increase in cellular phosphotyrosine levels (Blume-Jensen and Hunter, 2001). PTP1b has been shown to play both oncogenic and tumor suppressor roles, depending on context (Lessard et al., 2010). In liver, PTP1b deletion mainly increases phosphotyrosine levels, but some sites show decreased phosphorylation, suggesting network compensation for PTP1b deletion. Obesity increases cancer risk (Biddinger and Kahn, 2006), and, here, similar to many cancers, there is a general increase in phosphotyrosine levels. These diet-dependent phosphotyrosine network changes were not the focus of the present work, but they warrant further study.

Although the phosphotyrosine dataset reflected both diet and genotype perturbations, in general, these trends were driven by fold-changes of less than two, which are modest in comparison to the fold-changes observed in growth-hormone stimulation or overexpression of oncogenic tyrosine kinases in cell lines (Huang et al., 2007; Zhang et al., 2005). In order to sensitively detect the protein pathways exhibiting altered phosphotyrosine regulation under genotype and diet conditions, a gene set enrichment analysis (GSEA) framework (Subramanian et al., 2005) was adapted for phosphosite set enrichment analysis (PSEA). Through quantitative incorporation of the full dataset, this framework provides sensitive detection (e.g., of pathways whose net fold-change is ~20%) (Mootha et al., 2003), and, by implicitly controlling for measurement dependence in the null model, this method more accurately estimates significance, in comparison to methods involving overlap statistics (Subramanian et al., 2005). By considering the diverse functions of protein phosphotyrosine modification, we developed (1) signed PSEA to detect pathways controlled by relatively uniform change in phosphotyrosine and (2) unsigned PSEA to detect pathways controlled by simultaneous decrease and increase of select phosphosites. In addition, given the availability of experimental technical replicates, we assessed the robustness of these techniques to noise. PSEA results were sensitive to analysis sign as well as incorporation of measurement noise.

Development of PSEA was especially important in the context of this study, because the majority of the phosphotyrosine sites responsive to diet or genotype were functionally uncharacterized. Thus, incorporation of phosphosite and protein annotations by this framework enhanced hypothesis generation by (1) revealing unknown relationships between sets of phosphosites and (2) associating significance with hypotheses tested. PSEA uncovered phosphotyrosine regulation of pathways known to be PTP1b-dependent (insulin signaling) and

novel PTP1b-dependent phosphotyrosine regulation of proteins involved in oxidation and reduction (REDOX), lipid metabolism, amino acid metabolism and other pathways.

Because previous studies in L-PTP1b^{-/-} mice demonstrated altered lipid metabolic changes consistent with improved liver and cardiovascular health (Delibegovic et al., 2009), the enrichment for PTP1b-dependent phosphotyrosine alterations in lipid metabolic proteins motivated further study. The functions of these phosphosites were uncharacterized and spanned a diverse set of lipid metabolic proteins. In order to systematically infer regulatory function for these sites, the lipids from the liver tissue samples were profiled at molecular-level resolution using global MS-based techniques (Homan et al., 2011). These experiments reveal PTP1b-dependent lipid alterations in fatty acid and triglyceride metabolism. Although total fatty acids between L-PTP1b^{-/-} and control mice are unaltered, L-PTP1b^{-/-} mice on HFD have dramatically elevated levels of select ω 7/ ω 9 PUFA and select ω 3/ ω 6 PUFA. The physiological effect of the PTP1b-dependent ω 3/ ω 6 PUFA are less well characterized than their ω 3/ ω 6 PUFA products, arachidonic acid (AA), DHA and EPA, which are potent inhibitors of insulin-mediated lipogenesis in the liver (Sampath and Ntambi, 2005). However, if the PTP1b-dependent ω 3/ ω 6 PUFA share the properties or enhance the production of these downstream PUFA products, altered PUFA metabolism might explain how L-PTP1b^{-/-} mice exhibit reduced lipogenesis in the context of enhanced insulin sensitivity. In addition, the elevation of ω 7/ ω 9 PUFA, downstream products of the Δ 9-desaturase, SCD, may reflect increased SCD activity. Interestingly, the C18 SCD Index, often used as a surrogate for SCD activity, is elevated in basal L-PTP1b^{-/-} under both HFD and NC conditions. SCD activity may play a protective role in liver under HFD conditions, by neutralizing the cytotoxicity of saturated fatty acids via their conversion to

monounsaturated fatty acids (MUFA) and subsequent incorporation into triglycerides. This suggests a molecular mechanism by which HFD-induced stress is reduced in L-PTP1b^{-/-} mice.

Triglyceride levels are also elevated in L-PTP1b^{-/-} livers on HFD. This result was in contrast to previous results at earlier and similar time points, in which the opposite or no trend was found, respectively (Delibegovic et al., 2009). However, given the large number of samples and orthogonal techniques used for this measurement in our study, the triglyceride data is difficult to discount. Consideration of parallel serum triglyceride measurements in both this and the previous study suggest a hepatic triglyceride secretion defect. In both studies, the ratio of serum triglycerides to hepatic triglyceride content is reduced in L-PTP1b^{-/-} mice relative to control. In addition, the compositions of hepatic triglycerides are altered in L-PTP1b^{-/-} mice, which have a lower percentage of PUFA-rich triglycerides. As discussed in Chapter 2, through an oxidative mechanism, elevated PUFA may contribute to the triglyceride secretion deficiency and also explain the altered triglyceride compositions.

To understand how altered phosphotyrosine regulation might drive these lipid metabolic changes, multivariate regression models of these lipid phenotypes as functions of underlying phosphosites were constructed. Multivariate regression models phenotypes as functions of multiple contributing predictors and, in this way, mirrors the multivariate nature of true biological response mediation. Multivariate modeling was limited to regression, given the statistical properties of the dataset ($M > N$, or many more variables than observations). These analyses focused on the HFD basal dataset, where it was conceivable that both L-PTP1b^{-/-} and control phenotypes might be captured by a single linear model, as in a Taylor series expansion. However, as described in Chapter 3, this modeling endeavor presented several technical challenges. Foremost were significant amounts of systematically missing data points, due to

technical aspects of the phosphotyrosine experiments. To address this issue, diverse missing data methods were systematically compared using simulation and receiver-operator characteristic (ROC) performance criteria. This analysis highlighted the diversity in performance among missing data methods and enabled optimal selection of a missing data method for application to this study. To enhance the robustness of the models to measurement noise, a stochastic multivariate regression (SMR) modeling technique was developed. The pairing of missing data techniques with SMR enabled model construction that maximally leveraged the data, because both noisy and missing data points were evaluated using these methods, as opposed to being removed *a priori*. In addition, a modification to the calculation of the Q^2 goodness-of-prediction statistic was adopted. The modified Q^2 -value more rigorously reflects model prediction through the consideration of variable selection in addition to model fit. Using these methods and the Q^2 metric, three quantitatively predictive models of (1) PTP1b-dependent $\omega 7/\omega 9$ PUFA, (2) PTP1b-dependent $\omega 3/\omega 6$ PUFA, and (3) steatosis were built as functions of underlying phosphotyrosine network changes.

The reduced sets of model phosphosite predictors highlight several interesting phosphotyrosine-phenotype relationships. In particular, the phosphosite Y11 on cytochrome b5 (CYB5) is an important predictor of PTP1b-dependent $\omega 7/\omega 9$ PUFA. As discussed above, the $\omega 7/\omega 9$ PUFA are SCD products subsequently elongated by elongases and desaturated by $\Delta 5$ and $\Delta 6$ desaturases (Guillou et al., 2010). Interestingly, CYB5 is required for SCD activity, and CYB5 contributes to $\Delta 6$ desaturase activity (Guillou et al., 2004). CYB5 Y11 phosphorylation is elevated in L-PTP1b^{-/-} mice, and, because CYB5 is localized in the ER and Y11 is a tandem tyrosine residue, CYB5 Y11 may be a PTP1b substrate. However, many other uncharacterized phosphotyrosine sites were important to prediction in these models, and a detailed discussion of

their roles is contained in Chapter 2. In addition, PSEA was applied to lists of phosphosites ranked by predictive importance to the SMR models. For all three lipidomics phenotypes, phosphotyrosine regulation of REDOX pathways was significantly predictive. The balance of oxidation and reduction is a pivotal driver of metabolic states. For example, fatty acid desaturation requires conversion of NADH to NAD⁺ (Guillou et al., 2010), and lipid oxidation modulates liver triglyceride secretion and hepatic steatosis (Pan et al., 2004). Net changes in oxidation and reduction mediated by tyrosine phosphorylation of REDOX enzymes could underlie and contribute to all three PTP1b-dependent lipid metabolic phenotypes.

In summary, this research used global quantitative MS analysis to gain site-specific and molecular-level resolution into phosphotyrosine regulation and lipid metabolism in L-PTP1b^{-/-} and control mice under both HFD and NC conditions. Critical to the progress of this research were the design and development of computational methods for this dataset; these should facilitate future analysis of MS-based post-translational modification (PTM) datasets and other datasets as well. In this study, these techniques specifically enabled elucidation of those phosphosites and pathways likely to contribute to L-PTP1b^{-/-} phenotypes. The combination of systematic phosphotyrosine measurements with sensitive PSEA enabled the identification of multiple pathways with PTP1b-dependent phosphotyrosine regulation. Complementary global lipidomics profiling demonstrated that the products of two of these pathways, lipid and fatty acid metabolism, were altered as well. To connect phosphotyrosine measurements with the lipidomics measurements, multivariate regression analysis resulted in quantitatively predictive models that highlight roles for sets of phosphosites in several lipid metabolic phenotypes. The molecular mechanisms of PTP1b-dependent lipid metabolism merit further study, as do the other

PTP1b- and diet-dependently tyrosine phosphorylated pathways. It is my hope that this study motivates further research in these areas.

Limitations and Future Directions

The discovery aspect of this work represents both a strength and a challenge. Few phospho-specific antibodies exist for the hundreds of phosphotyrosine sites detected, and this in itself highlights the novelty of the measurements in this study. In the absence of detailed biological information, quantitative information provides an opportunity for computational and machine learning techniques to infer biological relationships for these sites. Although the present study included nearly fifty liver tissues, the resulting final dataset was sparse, and this posed significant challenges to computational inference, despite the large sample size. Although it was possible to work around these limitations with PSEA and with missing data techniques for a data subset using SMR, there is no substitute for complete data. As described in Chapter 3, the sparsity of the datasets are due to the combination of multiple 8-plex MS experiments, each obtained using a data-dependent acquisition mode and providing quantification of phosphosites, whose number and identities were variable from experiment to experiment. The original experimental design was to perform preliminary MS experiments in data-dependent acquisition mode to discover which phosphotyrosine sites were detectable in the L-PTP1b^{-/-} samples. These preliminary analyses were to be used to generate a list of phosphosites for targeted MS experiments using multiple-reaction-monitoring (MRM) techniques (Wolf-Yadlin et al., 2007). However, even after enrichment procedures, the iTRAQ-labeled phosphotyrosine peptide samples for the liver tissues from this study were too complex to be reliably analyzed by MRM technology available in the lab. Many mixed spectra were discovered upon manual validation of the initial data-dependently acquired experiments. These mixed spectra result from the co-elution of two

peptides with similar m/z ratios, and, in this case, the iTRAQ quantification is not specific to one peptide but a combination of peptides, whose individual contributions cannot be resolved. A high-resolution, full MS1 and/or full MS2 spectrum can be used to determine whether co-elution occurred, as evidence (contaminating, unexplained ion peaks) can appear in either. Whereas this information is available for data-dependent acquisition in our lab, MRM measures only selected ion transitions, corresponding to pre-selected ion fragments of precursor peptides. Thus, peaks from contaminating, co-eluting peptides would not be detected. For analysis of individually eluting phosphopeptides, MRM provides sufficient information for peptide quantification. Given the complexity of the samples in this study, all phosphotyrosine measurements were acquired using data-dependent acquisition, and incomplete data resulted. With much anticipation, methods for next-generation targeted MS/MS method have been developed by Joshua Coon's lab at the University of Wisconsin (unpublished), and these will enable high resolution, full-scan MS2 spectra of targeted precursors. In addition to enabling analysis of more complex samples, these methods also promise to target larger lists of peptides (>100). This technology will greatly enhance opportunities for computational inference of biological relationships from quantitative MS datasets.

To develop a functional role for the phosphosites on lipid metabolic proteins, global lipidomics profiling of hepatic lipid content was coupled with computational modeling techniques to determine the phosphosites most relevant to the metabolic alterations discovered. However, many of the metabolic alterations were themselves uncharacterized, providing an additional challenge for relating measurements to interpretable physiological outcomes. The PTP1b-dependent elevation in $\omega 7/\omega 9$ PUFA was the most notable example. The most PTP1b-dependent $\omega 7/\omega 9$ PUFA, mead acid (C20:3 $\omega 9$), is associated with essential fatty acid deficiency (EFAD) (Le et

al., 2009), and the other PTP1b-dependent $\omega 7/\omega 9$ PUFA appear to be precursors and/or derivatives of mead acid. Whether these $\omega 7/\omega 9$ PUFA share the properties of essential fatty acid derivatives (EPA, DHA, and arachidonic acid (AA)) in the regulation of lipid metabolism or inflammation is unclear. Interestingly, comparison of absolute amounts of fatty acids between NC and HFD mice reveals that EPA and DHA are 5-fold and 3-fold depleted under HFD conditions, while the EFAD-associated and PTP1b-dependent PUFA C20:3 $\omega 9/\omega 7^*$, C22:3 $\omega 9^*/\omega 7^*$, and C18:3 $\omega 7^*$ are each more than 5-fold elevated in HFD ($P < .01$, each comparison). Does HFD relate to essential fatty acid deficiency, and why do L-PTP1b^{-/-} have elevated EFAD-associated PUFA? EPA- and DHA-derived eicosanoids promote inflammation resolution. Perhaps EPA and DHA are depleted under HFD because they are used in the resolution of diet-induced inflammation. If HFD does resemble EFAD, then $\Delta 5$ and $\Delta 6$ desaturase activity would be elevated. If, additionally, SCD activity was increased in L-PTP1b^{-/-} mice, then, on HFD, there would be more available $\omega 7/\omega 9$ MUFAs for metabolism into $\omega 7/\omega 9$ PUFAs. Pleiotropic, unknown functions may exist for PTP1b-dependent $\omega 7/\omega 9$ PUFA and eicosanoid derivatives. As discussed in the introduction, an eicosanoid role for a derivative of C20:3 $\omega 9$ was recently discovered (Patel et al., 2008). The concentrations of C20:3 $\omega 9$ may be high enough to have physiological consequences under HFD, as its concentration is intermediate between AA and EPA. Complementary measurements of eicosanoids in these samples might provide insight. However, similarly to the phosphotyrosine modification of metabolic enzymes, it is difficult to interpret the directionality from static metabolic measurements alone. Given more resources, metabolic flux experiments might come closer to explaining the metabolic changes in L-PTP1b^{-/-} on HFD. Treating mice with C¹³- acetate (for tracing of de novo lipogenesis products) or C¹³- saturated fatty acids (for tracing diet-derived fatty acids) could illuminate how the PTP1b-

dependent $\omega 7/\omega 9$ PUFA arise and where they go. These measurements would also test whether the C18 SCD Index actually reflects increased SCD activity, a hypothesis that was difficult to test *in vitro*.

Additionally, it is still not clear whether a well-studied lipid metabolic phenotype, like hepatic steatosis, would be detrimental to liver function in the context of therapeutic PTP1b inhibition. Although hepatic steatosis is a hallmark of NAFLD and often associated with compromised liver function and inflammation, some experiments suggest that the esterification of fatty acids may protect against cytotoxicity and associated ER stress and inflammation responses (Coleman and Lee, 2004; Fabbrini et al., 2009). Intriguingly, the L-PTP1b^{-/-} mice on HFD in our study have elevated steatosis but similar levels of select markers of inflammation. In the previous study (Delibegovic et al., 2009) and at a similar time point, the L-PTP1b^{-/-} mice have the same levels of hepatic steatosis and reduced markers of inflammation relative to controls on HFD. Together, these results may indicate that the anti-inflammatory effects of L-PTP1b^{-/-} were offset by elevated steatosis in our study. Additional temporal resolution would help determine whether steatosis eventually leads to compromised liver function, an issue of paramount importance in the context of therapies targeting hepatic PTP1b. In addition to temporal resolution, it would be informative to explore other obesity-inducing diets, such as the high carbohydrate or high-sucrose, very low-fat diet. As illustrated by liver-specific SCD deletion mouse models, the source of obesity can profoundly impact the benefits of protein inhibition (Miyazaki et al., 2007). If SCD activity is increased in L-PTP1b^{-/-} mice, they may be less resistant to the negative consequence of high-carbohydrate-induced obesity.

Activity-based protein profiling (ABPP) experiments would contribute to elucidation of enzyme regulation by phosphotyrosine in this study. ABPP uses active-site-directed chemical probes to

broadly measure the activities of enzyme classes (Nomura et al., 2010). These probes consist of a reactive group to interact with active enzymes and a handle that can be used to enrich for these enzymes. Enriched enzymes can be visualized in a gel and identified by MS. ABPP probes have been developed for diverse and broad classes of poorly characterized enzymes, such as serine hydrolases, histone deacetylases, and enzymes with reactive cysteines (Bachovchin et al., 2011; Salisbury and Cravatt, 2008; Weerapana et al., 2010). These methods are commonly used in tandem with metabolomics profiling (Nomura et al., 2010). In the future, combination of ABPP with MS-based PTM datasets would yield important connections between protein activity and PTM state. Integration of these two datasets could also reveal network-level regulatory mechanisms controlling enzymatic activity.

The field of metabolism has been revitalized in the last decade. Many experimental methods have been central to this progress. This study highlights the importance of tissue-specific mouse models to understanding the interplay between specific perturbation of PTM networks and physiological responses. This study also highlights the power of mass spectrometry to expand our knowledge of metabolic control. Here we combine MS-based analysis of PTMs with metabolomics. Additional analysis involving ABPP and metabolic flux would provide crucial directionality to these measurements. The role of computation for integration of these complementary measurements, especially the sparse PTM datasets, will be greatly enhanced by further advances in mass-spectrometry, such as the next-generation MRM method, to provide complete datasets. The design of studies integrating these technologies will provide a powerful opportunity for computational analysis to discover connections among PTM regulation of pathways, specific enzymatic activities, metabolite abundances, and physiological outcomes.

References

- Bachovchin, D.A., Zuhl, A.M., Speers, A.E., Wolfe, M.R., Weerapana, E., Brown, S.J., Rosen, H., and Cravatt, B.F. (2011). Discovery and optimization of sulfonyl acrylonitriles as selective, covalent inhibitors of protein phosphatase methylesterase-1. *J Med Chem* *54*, 5229-5236.
- Banno, R., Zimmer, D., De Jonghe, B.C., Atienza, M., Rak, K., Yang, W., and Bence, K.K. (2010). PTP1B and SHP2 in POMC neurons reciprocally regulate energy balance in mice. *The Journal of clinical investigation* *120*, 720.
- Bence, K.K., Delibegovic, M., Xue, B., Gorgun, C.Z., Hotamisligil, G.S., Neel, B.G., and Kahn, B.B. (2006). Neuronal PTP1B regulates body weight, adiposity and leptin action. *Nat Med* *12*, 917-924.
- Biddinger, S.B., and Kahn, C.R. (2006). From mice to men: insights into the insulin resistance syndromes. *Annu Rev Physiol* *68*, 123-158.
- Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. *Nature* *411*, 355-365.
- Coleman, R.A., and Lee, D.P. (2004). Enzymes of triacylglycerol synthesis and their regulation. *Prog Lipid Res* *43*, 134-176.
- Delibegovic, M., Bence, K.K., Mody, N., Hong, E.G., Ko, H.J., Kim, J.K., Kahn, B.B., and Neel, B.G. (2007). Improved glucose homeostasis in mice with muscle-specific deletion of protein-tyrosine phosphatase 1B. *Molecular and cellular biology* *27*, 7727.
- Delibegovic, M., Zimmer, D., Kauffman, C., Rak, K., Hong, E.G., Cho, Y.R., Kim, J.K., Kahn, B.B., Neel, B.G., and Bence, K.K. (2009). Liver-specific deletion of protein-tyrosine phosphatase 1B (PTP1B) improves metabolic syndrome and attenuates diet-induced endoplasmic reticulum stress. *Diabetes* *58*, 590-599.
- Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A.L., Normandin, D., Cheng, A., Himms-Hagen, J., and Chan, C.C. (1999). Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* *283*, 1544.
- Fabbrini, E., Magkos, F., Mohammed, B.S., Pietka, T., Abumrad, N.A., Patterson, B.W., Okunade, A., and Klein, S. (2009). Intrahepatic fat, not visceral fat, is linked with metabolic complications of obesity. *Proc Natl Acad Sci U S A* *106*, 15430-15435.
- Guillou, H., D'Andrea, S., Rioux, V., Barnouin, R., Dalaine, S., Pedrono, F., Jan, S., and Legrand, P. (2004). Distinct roles of endoplasmic reticulum cytochrome b5 and fused cytochrome b5-like domain for rat Delta6-desaturase activity. *J Lipid Res* *45*, 32-40.
- Guillou, H., Zdravec, D., Martin, P.G., and Jacobsson, A. (2010). The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog Lipid Res* *49*, 186-199.
- Homan, E.A., Kim, Y.G., Cardia, J.P., and Saghatelian, A. (2011). Monoalkylglycerol ether lipids promote adipogenesis. *J Am Chem Soc* *133*, 5178-5181.
- Huang, P.H., Mukasa, A., Bonavia, R., Flynn, R.A., Brewer, Z.E., Cavenee, W.K., Furnari, F.B., and White, F.M. (2007). Quantitative analysis of EGFRvIII cellular signaling networks reveals a

- combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci U S A* *104*, 12867-12872.
- Le, H.D., Meisel, J.A., de Meijer, V.E., Gura, K.M., and Puder, M. (2009). The essentiality of arachidonic acid and docosahexaenoic acid. *Prostaglandins Leukot Essent Fatty Acids* *81*, 165-170.
- Lessard, L., Stuble, M., and Tremblay, M.L. (2010). The two faces of PTP1B in cancer. *Biochim Biophys Acta* *1804*, 613-619.
- Miyazaki, M., Flowers, M.T., Sampath, H., Chu, K., Otzelberger, C., Liu, X., and Ntambi, J.M. (2007). Hepatic stearoyl-CoA desaturase-1 deficiency protects mice from carbohydrate-induced adiposity and hepatic steatosis. *Cell Metab* *6*, 484-496.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., *et al.* (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* *34*, 267-273.
- Nomura, D.K., Dix, M.M., and Cravatt, B.F. (2010). Activity-based protein profiling for biochemical pathway discovery in cancer. *Nat Rev Cancer* *10*, 630-638.
- Pan, M., Cederbaum, A.I., Zhang, Y.L., Ginsberg, H.N., Williams, K.J., and Fisher, E.A. (2004). Lipid peroxidation and oxidant stress regulate hepatic apolipoprotein B degradation and VLDL production. *J Clin Invest* *113*, 1277-1287.
- Patel, P., Cossette, C., Anumolu, J.R., Gravel, S., Lesimple, A., Mamer, O.A., Rokach, J., and Powell, W.S. (2008). Structural requirements for activation of the 5-oxo-6E,8Z, 11Z,14Z-eicosatetraenoic acid (5-oxo-ETE) receptor: identification of a mead acid metabolite with potent agonist activity. *J Pharmacol Exp Ther* *325*, 698-707.
- Salisbury, C.M., and Cravatt, B.F. (2008). Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes. *J Am Chem Soc* *130*, 2184-2194.
- Sampath, H., and Ntambi, J.M. (2005). Polyunsaturated fatty acid regulation of genes of lipid metabolism. *Annu Rev Nutr* *25*, 317-340.
- Schmelzle, K., Kane, S., Gridley, S., Lienhard, G.E., and White, F.M. (2006). Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* *55*, 2171-2179.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 15545.
- Weerapana, E., Wang, C., Simon, G.M., Richter, F., Khare, S., Dillon, M.B., Bachovchin, D.A., Mowen, K., Baker, D., and Cravatt, B.F. (2010). Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* *468*, 790-795.
- Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D.A., and White, F.M. (2007). Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences* *104*, 5860.

Zhang, Y., Wolf-Yadlin, A., Ross, P.L., Pappin, D.J., Rush, J., Lauffenburger, D.A., and White, F.M. (2005). Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 4, 1240-1250.

Zinker, B.A., Rondinone, C.M., Trevillyan, J.M., Gum, R.J., Clampit, J.E., Waring, J.F., Xie, N., Wilcox, D., Jacobson, P., Frost, L., *et al.* (2002). PTP1B antisense oligonucleotide lowers PTP1B protein, normalizes blood glucose, and improves insulin sensitivity in diabetic mice. *Proc Natl Acad Sci U S A* 99, 11357-11362.