MIT Libraries | DSpace@MIT

MIT Open Access Articles

Molecular network analysis of phosphotyrosine and lipid metabolism in hepatic PTP1b deletion mice

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Miraldi, Emily R., Hadar Sharfi, Randall H. Friedline, Hannah Johnson, Tejia Zhang, Ken S. Lau, Hwi Jin Ko, et al. "Molecular Network Analysis of Phosphotyrosine and Lipid Metabolism in Hepatic PTP1b Deletion Mice." Integr. Biol. 5, no. 7 (2013): 940.

As Published: http://dx.doi.org/10.1039/c3ib40013a

Publisher: Royal Society of Chemistry

Persistent URL: http://hdl.handle.net/1721.1/89179

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike

Massachusetts Institute of Technology

NIH Public Access
Author Manuscript
Integr Biol (Camb). Author manuscript; available in PMC 2014 July 24.

Published in final edited form as:
Integr Biol (Camb). 2013 July 24; 5(7): 940–963. doi:10.1039/c3ib40013a.

Molecular network analysis of phosphotyrosine and lipid metabolism in hepatic PTP1b deletion mice

Emily R. Miraldi[1,2], **Hadar Sharfi**[2,*], **Randall H. Friedline**[3,*], **Hannah Johnson**[2,4,*], **Tejia Zhang**[5], **Ken S. Lau**[2,4,6], **Hwi Jin Ko**[3], **Timothy G. Curran**[2,4], **Kevin M. Haigis**[4,6], **Michael B. Yaffe**[2], **Richard Bonneau**[8], **Douglas A. Lauffenburger**[4], **Barbara B. Kahn**[9], **Jason K. Kim**[3], **Benjamin G. Neel**[10,11,+], **Alan Saghatelian**[5,+], and **Forest M. White**[2,4]

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

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

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

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

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

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

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

[8]Center for Genomics and Systems Biology, New York University, New York, New York 10003, U.S.A

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

[10]Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 2M9, Canada

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

Contact Information: Forest M. White, fwhite@mit.edu, (617) 258-8949 (phone), (617) 258-0225 (fax).
[*]These authors contributed equally to this work.
[+]These authors contributed equally to this work.

Author contributions: E.R.M., B.G.N., and F.W.M. designed the study. T.Z., K.M.H., M.Y., D.A.L., B.B.K., J.K.K., A.S., in addition to B.G.N. and F.W.M., contributed technical advice and/or resources to the study. E.R.M., H.S., R.H.F., H.J., K.S.L., H.J.K. performed experimental work. E.R.M. designed and executed computational analysis. T.C. assisted in visual presentation of validated peptide spectra. E.R.M. and F.W.M. wrote the manuscript with significant input from other authors.

Conflict of Interest
The authors declare that they have no conflict of interest.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Abstract

Metabolic syndrome describes a set of obesity-related disorders that increase diabetes, cardiovascular, and mortality risk. Studies of liver-specific protein-tyrosine phosphatase 1b (PTP1b) deletion mice (L-PTP1b−/−) suggest that hepatic PTP1b inhibition would mitigate metabolic-syndrome through amelioration of hepatic insulin resistance, endoplasmic-reticulum stress, and whole-body lipid metabolism. However, the altered molecular-network states underlying these phenotypes are poorly understood. We used mass spectrometry to quantitfy protein-phosphotyrosine network changes in L-PTP1b−/− mouse livers relative to control mice on normal and high-fat diets. We applied a phosphosite-set-enrichment analysis to identify known and novel pathways exhibiting PTP1b- and diet-dependent phosphotyrosine regulation. Detection of a PTP1b-dependent, but functionally uncharacterized, set of phosphosites on lipid-metabolic proteins motivated global lipidomic analyses that revealed altered polyunsaturated-fatty-acid (PUFA) and triglyceride metabolism in L-PTP1b−/− mice. To connect phosphosites and lipid measurements in a unified model, we developed a multivariate-regression framework, which accounts for measurement noise and systematically missing proteomics data. This analysis resulted in quantitative models that predict roles for phosphoproteins involved in oxidation-reduction in altered PUFA and triglyceride metabolism.

Keywords

PTP1b; Phosphoproteomics; Lipidomics; Liver; Computational Modeling

Introduction

Metabolic syndrome encompasses a group of commonly associated disorders, including central obesity, hyperglycemia, hyperinsulinemia, dyslipidemia, and hypertension, that promote the development of type 2 diabetes mellitus (T2DM) and cardiovascular disease (CAD)[1, 2]. In tandem with obesity, metabolic syndrome and its co-morbidities have spread worldwide, affecting individuals across all socioeconomic circumstances[3]. Liver is a prominent contributor to metabolic syndrome pathogenesis and a primary target tissue for therapeutic intervention[4]. In the context of metabolic syndrome, hepatic insulin resistance results in excess gluconeogenesis, hyperglycemia, and hyperinsulinemia. In addition, liver accumulates lipid as a result of free fatty acid (FFA) flux from insulin-resistant adipose tissue and altered hepatic lipid metabolism. Accumulation of hepatic lipid is often associated with inflammation and compromised liver function. Although the detailed mechanisms of insulin resistance remain the subject of intense debate, there is a consensus that insulin resistance develops from impaired cellular signaling downstream of the insulin receptor (IR)[5].

Considerable interest has developed in protein tyrosine phosphatase 1b (PTP1b) as a therapeutic target to potentiate insulin signaling in the context of metabolic syndrome. 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 leptin signaling, via dephosphorylation of tyrosine residues on Janus kinase 2 (Jak2)[6]. Accordingly, whole-body PTP1b−/− mice are lean, hypersensitive

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

to insulin, and resistant to diet-induced obesity[7]. In addition, PTP1b regulates specific phosphotyrosine sites on proteins in other pathways (Src, MAPK, cell adhesion and motility)[6], which might also physiologically impact PTP1b−/− mice.

Liver-specific PTP1b-deletion mice (L-PTP1b−/−) are not resistant to diet-induced obesity, but hepatic PTP1b deletion enhances insulin signaling[8]. Thus, L-PTP1b−/− mice have improved postprandial suppression of hepatic glucose production, which contributes to whole-body glucose homeostasis in the context of high-fat diet (HFD). In addition, on HFD, hepatic PTP1b deficiency drives positive lipid metabolic changes in both liver and circulation, consistent with improved liver health and reduced cardiovascular risk[8]. Finally, when challenged by HFD or ER-stress inducers such as tunicamycin and thapsigargin, L-PTP1b−/− mice have a muted hepatic ER stress response across all three major ER-stress sensing pathways[8, 9]. These phenotypes are therapeutically attractive given the importance of ER stress to metabolic syndrome pathogenesis, especially in the context of hepatic inflammation and lipotoxicity[10, 11]. Thus, in terms of whole-body and liver-specific physiology, PTP1b inhibition holds therapeutic potential for treatment of metabolic syndrome[12].

Although hepatic PTP1b inhibition or deletion has been associated with positive physiological effects, the underlying molecular network alterations are not well understood. To address this issue, we coupled statistical modeling techniques with quantitative mass-spectrometry (MS) measurements of hepatic protein tyrosine phosphorylation to determine, in an unbiased and site-specific manner, the phosphotyrosine network alterations mediating resistance to HFD-induced stress 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 for enzymes with diverse metabolic functions, including lipid metabolism. To elucidate a role for this largely uncharacterized set of PTP1b-dependent phosphosites on lipid metabolic proteins, 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 that integrate our lipidomic and proteomic observations. These models highlight the molecular underpinnings connecting PTP1b deletion, altered tyrosine phosphorylation signaling, and altered lipid metabolism, and 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[8]. 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 half of the mice were placed on HFD (55% kcal from fat) and the rest continued on NC. These diets were maintained until 19.5 weeks of age (Figure 1A). To gauge the physiological consequences of HFD, serum markers of insulin sensitivity and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

lipid metabolism were measured at 17.5 and 18.5 weeks (Table S1). As expected, both L-PTP1b−/− and control mice had elevated levels of fed insulin and glucose on HFD relative to NC, indicating that insulin resistance developed during the course of the diets. 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.

At 19.5 weeks, mice were sacrificed and their livers were 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 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 1B)[13]. This analysis resulted in the identification and quantification of 298 tyrosine phosphosites on 206 proteins; measurement error from available technical replicates was ~10% (Table S2, Methods). To visualize the most important trends across the multidimensional phosphotyrosine datasets, we applied principal component analysis (PCA). The most dominant phosphosite trends dictate the locations of liver samples in the principal component plane, and here the plane naturally segregates livers according to diet and genotype conditions for both the insulin-stimulated and basal datasets (Figure 2A). This suggests that diet and genotype are the dominant causes of phosphotyrosine network perturbations; individually, genotype and diet explain 21–23% and 15–17% of the phosphorylation variance. The PCA plot demonstrates that the phosphorylation data alone 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 (Figures 2B and S1). Due to the stochastic nature of data-dependent MS analysis[14], the dataset is somewhat sparse, with many of the phosphorylation sites measured in only some of the MS experiments. However, using 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. In agreement with this interpretation, RNAseq analysis of a subset of livers revealed no significant gene expression changes for the phosphoproteins detected in our study (significant controlling for FDR < .05, n = 2 pooled L-PTP1b−/− samples, n = 2 pooled control samples, where each pooled sample was composed of 3–4 independent livers, data not included), suggesting that the alterations in protein tyrosine phosphorylation were post-transcriptional.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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

We adapted a gene set enrichment analysis (GSEA) framework[15] for an integrated, pathway-level analysis of our quantitative protein phosphorylation data; this approach 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 (Table S3), composed of protein pathway annotations from Kegg, cellular locations from GO, protein structural information (PFAM and SwissPro), a curated list of PTP1b substrates[16], as well as a set of insulin-sensitive phosphorylation sites from a previous study[17]. 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 showed significant PTP1b-dependent tyrosine phosphorylation (FDR<0.01, Table S4). Consistent with its role in regulating insulin sensitivity, phosphosites involved in insulin signaling had 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 was in agreement with the cellular localization of PTP1b to the cytoplasmic face of the ER, providing additional confirmation that our approach identified 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 reported by Ren et al.[16]. 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. The complexity of the biological response to PTP1b deletion makes it challenging to identify direct substrates of PTP1b from our experimental design, but the network-level phosphosite measurements provide direct insight into the particular proteins and pathways that are modulated by PTP1b deletion or HFD and thereby contribute to altered physiological states *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[8], the PTP1b-dependent increase in phosphorylation of lipid metabolic proteins (highlighted in Figure 2B) was of particular interest. These proteins spanned a diverse set of lipid pathways, including fatty acid (FA) metabolism and transport (acyl-coA binding protein (ACBP), FA synthase (FAS), ATP citrate lyase (ACLY), cytochrome b5 (CYB5A), and cytochrome P450 2e1 (CYP2E1)), triglyceride (TG) 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, the set

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of phosphosites limited to proteins involved in FA metabolism was also significantly PTP1b-dependent (FDR<0.05), highlighting potential PTP1b-dependent regulation of FAs in particular. Some of these phosphorylation sites have been identified in previous large-scale studies, but their functions have not been characterized.

The enrichment of PTP1b-dependent phosphosites on REDOX proteins is intriguing given that oxidative stress plays an important role in obesity-induced ER stress and inflammation[18]. 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, see (Table S4)), and the functions of these sites as well as the contributions of these proteins to ER stress and inflammation 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 (Table S4). The functions and PTP1b-dependent regulation of these phosphorylation sites 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[19, 20].

In contrast to PTP1b-dependent sites, phosphosite set enrichment analysis of diet-dependent sites yielded a single significant enrichment for mitochondrial proteins (FDR < .01, Table S5, highlighted in Figure S1) (note that the scarcity of significant enrichments 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[21, 22]. Interestingly, our analysis indicates that HFD also increases tyrosine phosphorylation on mitochondrial proteins, suggesting a potential link between cancer and obesity. Further investigation of these phosphorylation sites might reveal specific mechanisms by which obesity promotes cancer metabolism.

PTP1b-deficient livers have an overabundance of free PUFAs

Although lipid and FA metabolic pathways show significant PTP1b-dependent phosphotyrosine regulation, the effects of these network changes 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 livers. Given the diversity of lipid metabolic proteins with PTP1b-dependent phosphorylation, we applied a global LC-MS-based lipidomic strategy that enables *de novo* discovery of differentially abundant lipids among samples (Figure 3)[23]. Using this approach, we quantitatively monitored lipids and found several differences between L-PTP1b−/− and control mice. On HFD, the most significantly PTP1b-dependent lipids were polyunsaturated FAs (PUFAs) in the free-FA (FFA) pool. To examine these changes more closely, a calibration curve, composed of isotopically labeled FA standards,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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 (Methods)). Interestingly, PTP1b deletion altered the FA pool composition rather than pool size, as total FAs were similar between L-PTP1b−/− and control mice (Table S6). 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 4A). Many of the unsaturated FAs had ion chromatogram elution profiles with multiple peaks (Figures 3B and S2); 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 ω6 is the anti-inflammatory lipid dihomo-γ-linolenic acid, whereas C20:3 ω3 is a precursor of the anti-lipogenic PUFAs eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). The physiological roles of C20:3 ω7 and ω9, on the other hand, are less-well studied. As can be seen in the upper-right-hand corner of Figure 4A, several PUFAs were significantly increased in the L-PTP1b−/− livers, 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 identify PTP1b-dependent C18:3, C20:3 and C22:3 isomer peaks, co-injection experiments with commercially available isomer standards were performed (Figure S2). 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 ω9 or a combination of ω9 and ω7 species and, in addition to total C20:2, are referred to as “ω7+ω9 PTP1b-dependent PUFA” (denoted in red, Figure 4A). The other PTP1b-dependent PUFA species (C24:6, C24:5, C24:4, C20:3, C22:3, colored in green in Figure 4A) were ω3 and ω6 and are referred to as “ω3+ω6 PTP1b-dependent PUFA”.

To gain insight into the unknown physiological functions of the PTP1b-dependent PUFA isomers, correlation-based clustering was applied to the percent FA compositions of HFD and NC livers. PTP1b-dependent PUFA clustered with several better-characterized PUFA (Figure 4B). All ω3+ω6 PTP1b-dependent PUFA, including the poorly characterized C24 PUFA series, cluster with PUFAs with known physiological roles, including anti-lipogenic/anti-inflammatory DHA (C22:6 ω3) and EPA (C20:5 ω3), for which the ω3 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 livers relative to NC (Table S6). The high level of correlation between the FAs in this cluster suggests a beneficial metabolic function for the ω3+ω6 PTP1b-dependent PUFAs.

By contrast, the C20:3 ω9/ω7* isomers were barely detectable in NC livers, and the C18:3 ω7* and C22:3 ω9*/ω7* isomers were not detected (Table S6), suggesting that these species are byproducts of HFD metabolism. C20:3 ω9/ω7* clusters with C20:2 and monounsaturated FAs (MUFAs): C16:1 ω7, C18:1 ω9 and C20:1 (Figures 4B). Given that ω7/ω9 MUFA are products of the Δ9 desaturase, stearoyl-CoA desaturase (SCD), the C18

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

SCD Index, the ratio of SCD's major product (oleic acid) to substrate (stearic acid), which is often used as a surrogate for SCD activity[24], was calculated for all livers. In agreement with the increased abundance of the ω7+ω9 PTP1b-dependent PUFAs, the SCD Index is elevated in both NC and HFD L-PTP1b−/− mice relative to control under basal conditions (Figure 4C). At the condition examined (10mU/g, 10 min.), insulin stimulation did not affect the SCD index for control mice, however, the SCD index of L-PTP1b−/− livers was reduced upon stimulation, suggesting an insulin-dependent effect. Indeed, the insulin dose was hypoglycemia-inducing, and, given that the L-PTP1b−/− livers are more sensitive to insulin, it is possible that L-PTP1b−/− mice 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 elevated in some of the insulin-stimulated livers (Figure 4D). These results suggest that L-PTP1b−/− mice might have increased hepatic SCD activity under normal 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 TGs relative to control mice on HFD (Fig. 5A). This trend was verified in quantitative MS-lipidomics analyses (with TG 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 TGs 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)[8]. To verify that the trends were reproducible at this timepoint, liver TGs were quantified in a second cohort of HFD mice (4 L-PTP1b−/−, 4 control). These additional samples confirmed that hepatic TGs were elevated in L-PTP1b−/− mice (combined data shown in Figure 5A). These data are even more striking when liver TG content is plotted as a function of mouse weight. For weight-matched pairs of mice, L-PTP1b−/− mice have dramatically more hepatic TG content (Figure 5B), and there is surprisingly little correlation between total hepatic and serum TGs (Figure 5C). This disparity between hepatic and serum TGs is consistent with the later time point in our previous study[8], where L-PTP1b−/− mice had significantly lower serum TGs but similar hepatic TGs relative to control mice. The lack of correlation between hepatic and serum TGs in both studies suggests a PTP1b-dependent defect in hepatic TG 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-1b, the marker of macrophage infiltration MCP-1, and phosphorylation of the stress kinase p38 (Figure S3). 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[8, 9].

In addition to total TGs, MS analysis provided quantification for 26 TG structural isomers, characterized by the sum of carbons and double bonds in the acyl chains (Table S7). Detected TGs ranged from 48 to 56 total acyl carbon length and contained a total of one to ten double bonds. Statistically significant PTP1b-dependent changes in TG composition

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

were visualized in a volcano plot (Figure 5D). As a fraction of the total pool, TGs with many sites of desaturation, or "PUFA-rich" TGs, are specifically reduced in L-PTP1b−/− livers relative to control. Notably, the number of double bonds in the TGs correlates with the compositional depletion of that TG in the L-PTP1b−/− livers ($P<10^{-5}$).

Longer-chain, PUFA-rich TGs constitute a larger percentage of TG composition in NC relative to HFD mice, regardless of genotype. As highlighted by the heatmap of clustered TG compositions (Figure 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 mice. On HFD, esterification of diet-derived FAs into TGs may limit their cytotoxic effects. The resulting build-up of shorter-chain TGs with fewer sites of desaturation would explain the lower fraction of longer-chain, PUFA-rich TGs in HFD versus NC mice.

The physiological significance of the liver TG compositions in L-PTP1b−/− mice on HFD is unknown, but serum depletion of PUFA-rich TG species is positively predictive of T2DM development in humans[25]. To determine whether the compositional depletion of PUFA-rich TGs in the L-PTP1b livers was associated with similar serum compositional changes, we assayed serum TGs by MS (Table S8, Figure S4) and detected a parallel depletion in PUFA-rich TGs ($P<10^{-10}$, correlation between TG double bond number and depletion in L-PTP1b −/− mice).

Although total liver and serum TG measurements suggest a TG secretion defect, the altered TG compositions (depletion of PUFA-rich and elevation of MUFA-rich TGs) suggest 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 (Figure S5). Although total pS79 ACC was significantly increased in L-PTP1b−/−, the ratio of pS79 ACC to ACC was not significantly elevated. These data, as well as the lipogenic mRNA expression measurements in the previous study (Delibegovic et al, 2009), suggest that lipogenesis is not increased in the L-PTP1b−/− livers. Thus, the compositional TG changes in L-PTP1b−/− mice might reflect an increased capacity to store diet-derived FAs as neutral TGs. Altered acyl chain compositions in other complex lipid classes (e.g., phospholipids) were not detected.

Modeling the relationships between protein phosphorylation and lipid metabolites: incorporation of measurement noise with stochastic multivariate regression

To gain insight into the phosphosites that might potentiate PTP1b-dependent phenotypes, we constructed computational models from the phosphorylation and lipid measurements. Given the size of our study and the limited functional annotation for most of the quantified lipids and phosphosites, we built multivariate regression models that take advantage of mouse-specific profiles to predict the novel PTP1b-dependent phenotypes. 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 (steatosis, ω7+ω9 PTP1b-dependent PUFA, and the ω3+ω6 PTP1b-dependent PUFA) as functions of the 228 phosphosites in the HFD basal (HFDB) dataset. However, prior to building the models, we considered strategies to combat two technical challenges inherent to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

this dataset and large-scale, MS-based proteomics datasets in general: (1) measurement error and (2) systematically missing data points, the latter of which results from the stochastic nature of data-dependently acquired MS/MS spectra, as noted earlier. Frameworks for building multivariate regression models in the context of these challenges are described and evaluated here and in the subsequent section.

The error associated with mouse-specific phosphorylation measurements ranges from ~1–70% (Figure S6A). Measurement noise can lead to spurious correlation, especially for phosphosites with fold-changes smaller than measurement error. Although it might be possible to remove some of these noisy phosphosites using ANOVA (i.e., by filtering out sites whose phosphorylation is statistically unchanging across samples in the dataset), the selection of a cutoff for data inclusion is often difficult to justify. Thus, in order to distinguish biological variation from measurement noise in the model-building process, we developed stochastic multivariate regression (SMR). The SMR method uses Monte Carlo sampling[26] followed by bootstrapping[26] to develop a distribution of regression coefficients that incorporate measurement error in the context of limited sample size. The coefficient distribution can then be used to evaluate the significance of predictor variables to the model. In the Monte Carlo sampling step, Gaussian distributions are used to model phosphosite-levels, and these distributions (one for each data-point) are based on experimental means and error estimates from technical replicates (Figure S6B). Intuitively, the Monte Carlo sampling step penalizes noisier predictor-response (phosphosite-lipid) measurements, as large measurement errors will be reflected in wider coefficient distributions and thus less-significant p-values or variable importance of project (VIP) scores, two metrics of predictor importance used in model variable-selection procedures.

The ability of SMR to account for measurement error is highlighted in a comparison to multivariate regression where the Monte Carlo sampling step is skipped (deterministic multivariate regression, DMR) and measurement error is not incorporated into the regression- coefficient significances. Three multivariate linear regression models (of Genotype (discriminating L-PTP1b−/− livers from control), Steatosis, and C18 SCD Index) were constructed using the complete HFDB data subset (76 sites × 13 samples). The overlap between phosphosites selected for the three models by SMR and DMR are shown with ANOVA-filtered phosphosites in Venn diagrams (Figure 6A). Unlike DMR-selected variables, SMR-selected variables are a subset of the ANOVA-filtered phosphosites, suggesting that measurement noise is properly reflected in regression-coefficient significances. The three response variables Genotype, Steatosis, and C18 SCD Index have estimated relative standard errors of ~0, 16% and 32%. As the standard error of the response variable increases, the number of phosphosites with significance regression coefficients decreases for SMR models, relative to DMR models. Ultimately, no phosphosites are significantly associated ($P_{raw} < .05$) with the response variable with the greatest standard error, C18 SCD Index, because the error in this measurement, in combination with phosphosite measurement errors, is too great. ANOVA-filtering of the phosphosites would not have substituted for SMR in this case, as the dominant measurement error was in the response. Through the SMR method, measurement error from both predictor and response variables is reflected in the significances of resulting regression coefficients derived from the bootstrapped distributions of regression coefficients.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Modeling the relationships between protein phosphorylation and lipid metabolites: optimization of missing data methods for multivariate regression

The second consideration for multivariate regression modeling with MS-based proteomics datasets is systemically missing data points. In contrast to the univariate correlation-based methods pursued above, multivariate regression requires that a dataset be complete. MS-based proteomics datasets can pose a significant challenge, due to the number of systematically missing data points that result from merging multiple data-dependently acquired MS experiments. As a specific example, the HFDB data matrix is displayed in Figure 6B. Because the data points are systematically missing, data methods based on matrix decomposition, which work well in the context of multivariate regression by partial least squares, were not feasible[27]. \

To address the issue of systemically missing data points, we considered several missing data methods. First, we evaluated "list-wise deletion" of all measurements for which sample coverage was incomplete, a method used in previous multivariate regression analyses of MS-based phosphotyrosine measurements[28, 29]. Although this method requires no assumptions or ad hoc rules, the final size of the dataset is reduced to only those sites for which quantification is available for all conditions. In the case of the HFDB dataset, analysis would be limited to 76 phosphorylation sites, instead of the full 228 identified sites (Figure 6B). To incorporate a larger proportion of the data into the analysis, we considered another approach, in which multiple complete data subsets (Figure S7A) are used to build separate models; variable rankings from the individual models are then integrated using decision rules. In this way, data imputation is avoided; however, on the other hand, decision rules are often heuristic rather than optimal. Finally, we evaluated imputation methods. The goal of an imputation method is to facilitate analysis while minimizing bias, and multiple random imputation (MRI) fulfill this criteria[30]. We considered traditional MRI[30], and a related method we termed "Gaussian" imputation; we also examined variations of these methods, in which a treatment assumption (genotype-dependence) was used to impute missing data points.

We constructed a comprehensive strategy to assess the performance of each method (Figure 6B). The complete data subset (13 samples by 76 phosphosites) was used as a reference dataset for SMR modeling, and simulated datasets were generated by randomly removing measurements in a pattern that structurally corresponded to the original HFDB data matrix (Figure 6B). Three multivariate regression models, each of a different response (Steatosis, ω3+ω6 PTP1b-dependent PUFA, and ω7+ω9 PTP1b-dependent PUFA), were built as functions of phosphotyrosine measurements, leading to three "true" models based on SMR of the complete data subset. Missing data methods were tested by building the same three models using simulated datasets. Each method was evaluated based on (1) true positive rate, the selection of phosphosites that would have been incorporated into "true" models, and (2) false positive rate, the selection of phosphosites that were not included in the "true" models. Results were plotted on receiver-operator characteristic (ROC) curves (Figure 6C).

For this dataset, there was significant disparity in the performance of the missing data methods. List-wise deletion performed most poorly overall. Because this method excluded more than two-thirds of the dataset, list-wise deletion was limited to detection of only ~30%

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of the "true" positives, at best. Even in the low-FDR regime, this method offered no apparent benefit, as imputation methods performed equally well at low FDRs (Figure 6C). The methods that integrated results from individual models built on complete data subsets ("All Subsets Agree" and "Complete Subset Rules") provided the opportunity to detect most true positives, but at high false-positive rates. MRI-based methods performed best, with the treatment-dependent MRI method performing best overall, suggesting that treatment-dependent phosphorylation was a good assumption for this dataset. Importantly, the performance of missing data methods was sensitive to the particular response modeled (Figure 6C), suggesting that the application of missing data methods should be evaluated on a case-by-case basis.

Integrated SMR and set-enrichment analysis yields predictive models of lipid metabolic phenotypes and pathway-level insight from incomplete phosphotyrosine data

Final, independent models of steatosis, ω7+ω9 PTP1b-dependent PUFA, and the ω3+ω6 PTP1b-dependent PUFA were built on reduced sets of phosphosites, composed of those phosphosites most important to model prediction (Figure 7A). Model cross-validation indicates that these models had good predictive power (Q^4 = .54, .66, .52, respectively) (Figure 7B), especially in the context of increased biological variation in individual mouse livers compared to cell lines, the experimental system used previously for multivariate regression models involving MS-based phosphotyrosine measurements[29, 31]. It is also worth noting that the Q^4 values presented here account for both model fit and variable selection, and thus are more conservative estimates of predictive performance than typically reported.

Most of the phosphosites highlighted by the models are uncharacterized (Figure 7A). In the Discussion section, we use information about protein function from the literature to suggest relationships between individual sites and effects on lipid metabolism. Although literature-based analysis of individual model predictors revealed many interesting connections (see below), we implemented a second strategy to gain additional insight, especially for the predictive phosphosites on proteins whose functions were not obviously related to lipid metabolism. Specifically, we integrated set-enrichment analysis into the multivariate modeling framework to develop a pathway-level perspective. In general, variable reduction methods, regardless of the specific regression technique used, rank underlying variables (e.g., phosphosites) according to predictive potential. These ranked lists can then be used to determine whether phosphorylation patterns on particular sets or protein pathways are multivariately predictive and might underlie a response. Set-enrichment analyses of the three phosphosite-dependent models (Figure 7A) highlighted the phosphorylation of REDOX proteins as significant to each (FDR<.01, Tables S9–11); this commonality between the predictive phosphosites was not obvious upon initial inspection of reduced-model predictors and presents a compelling direction for future experiments. REDOX balance is a pivotal driver of metabolic state; desaturation of FAs and CYP2E1 oxidation of PUFA, for example, depend on conversion of NADH to NAD^+. As is discussed in more detail below, oxidation also modulates liver TG secretion and hepatic steatosis[32]. Therefore, net changes in cellular oxidation and reduction mediated by tyrosine phosphorylation of REDOX enzymes could play a significant role in lipid metabolism.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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[33]. 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, 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 deficiency in modulating hepatic lipid metabolism. Multiple phosphorylation sites on proteins involved in lipid metabolism are significantly affected by PTP1b deletion, and comprehensive lipidomics analysis demonstrated that L-PTP1b−/− mice have several previously unappreciated metabolic abnormalities. For instance, although total levels of FAs are similar between L-PTP1b−/− and control mice on HFD, L-PTP1b−/− mice have an overabundance of longer-chain PUFAs, including ω7/ω9 PUFAs and select ω3/ω6 species. Unfortunately, the physiological roles of most of the PTP1b-dependent PUFAs are not well characterized. However, when HFD and NC FA measurements are clustered, PTP1b-dependent ω3/ω6 species cluster with DHA and EPA, suggesting anti-lipogenic and/or anti-inflammatory roles for these species.

The multivariate regression models suggest multiple mechanisms by which altered tyrosine phosphorylation levels might lead to altered PUFA profiles in the L-PTP1b−/− mice (Figures 7A and 8). The ω7+ω9 and ω3+ω6 PTP1b-dependent PUFA clusters have several predictive phosphosites in common (Figure 7A). These PUFA share several 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 ω3/ω6 and ω7/ω9 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 PPARγ and other lipogenic genes[34]. Given that the expression of lipogenic genes was decreased in two previous studies of L-PTP1b−/− mice[8, 9] 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[35], and therefore we cannot rule out a pleiotropic effect, whereby altered acetylation of multiple enzymes might influence both ω3/ω6 and ω7/ω9 PUFA.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Additional phosphorylation sites in the ω7/ω9 PUFA model are implicated in lipid metabolism. For example, CYB5 is an electron-transporting heme protein of the ER that co-localizes with Δ5, Δ6, and Δ9 desaturases[36, 37]. These enzymes, in coordination with elongases, are required for the biosynthesis of long-chain PUFAs[38]. CYB5 is necessary for SCD function and it promotes Δ6 (FADS2) desaturase activity[39]. Because both SCD index and CYB5 Y11 phosphorylation are elevated in L-PTP1b−/− livers, increased phosphorylation of CYB5 Y11 may potentiate SCD activity, either via direct effects on enzymatic activity or via altered protein-protein interactions in CYB5 complexes. Additionally, CYB5 contributes to PUFA metabolism via interaction with cytochrome P450s, including CYP2E1. It is possible that the altered phosphorylation of CYB5 Y11 modifies multiple aspects of lipid metabolism simultaneously.

Other PTP1b-dependent phosphosites might contribute to an excess of PUFA via defects in pathways leading to TG and/or complex lipid formation. ACBP, which binds acyl-CoA esters (C14–C22) with high specificity and affinity (K_d<.5nM)40, 41, is hyperphosphorylated on pY29 in L-PTP1b−/− mice (Figure 1B). The crystal structure for ACBP is available, and Y29 stabilizes the 3′ ribose phosphate of the acyl-CoA[42]. 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 due to (1) impaired transport of acyl-CoAs to cellular sites of complex lipid synthesis and (2) a build-up of untransported acyl-CoAs, which would inhibit PUFA 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 TGs in L-PTP1b−/− livers and serum. PTP1b-dependent phosphorylation of PRDX6 (Y89), a bifunctional protein with glutathione peroxidase and phospholipase A_2 activities[43], 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 increased desaturase activity (CYB5 Y11), decreased PUFA metabolism (CYP2E1 Y426), increased lipase activity (PRDX6 Y89) and modulation of pathways leading to complex lipid synthesis (ACBP Y29, ACSL5 Y69 and SND1 Y908) (Figure 8).

L-PTP1b −/− livers are ~2-fold more steatotic than controls on HFD, yet elevation of liver TGs did not correlate with markers of inflammation, suggesting that the steatosis was not associated with compromised liver function. Comparison of total liver and serum TGs in L-PTP1b−/− mice indicates defective TG secretion, which could contribute substantially to steatosis. In addition, MS-based analysis of TGs revealed MUFA-enriched compositions in both the liver and serum of L-PTP1b−/− mice on HFD, suggesting that any hepatic TG secretion defect was general (i.e., not specific to MUFA-rich or PUFA-rich TGs). Previous measurements of lipogenic gene expression and the pS79 ACC ratio provide no evidence for enhanced lipogenesis in L-PTP1b−/− mice; thus, the elevation in MUFA-rich TGs is most likely attributable to diet-derived, as opposed to de-novo synthesized, FAs (discussed in Results). Increased storage of diet-derived SFAs as TGs could be mediated by increased SCD activity, as evidence suggests that SCD colocalizes with and funnels newly synthesized MUFA products to diacyl glycerol acyl transferase 2 (DGAT2), a DGAT isoform that

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

targets TGs to cytosolic storage rather than VLDL secretion[44]. Thus, elevated SCD activity could contribute to the MUFA-rich TG composition in L-PTP1b−/− mice. In addition, elevated SCD activity might mitigate HFD-induced inflammation via conversion of diet-derived saturated FAs (SFAs) to MUFAs and funneling of MUFAs to TGs via DGAT2, as MUFAs are typically less cytotoxic than SFAs and TGs are inert[45, 46] (Figure 8).

Elevated PUFAs could contribute to defective TG secretion in L-PTP1b−/− mice. ω3 PUFA not only decrease hepatic TG output via suppression of lipogenesis but also independently inhibit TG secretion[47]. DHA inhibition of TG secretion is mediated by lipid peroxidation followed by the degradation of apolipoprotein B (ApoB), the central component of VLDL particles, which was specifically sensitive to the PUFA-mediated oxidative damage[32]. In this same study, several other dietary PUFAs, including linoleic acid (C18:3 ω3) and arachidonic acid (C20:4 ω6), were shown to promote ApoB degradation via peroxidation. Lipid peroxide formation was positively correlated with polyunsaturation of FAs, 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 REDOX proteins whose phosphorylation patterns were predictive in the steatosis model (Figures 7A, 8, Table S11). This mechanism would also be consistent with observations from a previous study, in which pulse-chase analysis of ApoB100 in primary hepatocytes from whole-body PTP1b knockout mice demonstrated reduced secretion of ApoB100, and insulin-stimulated HepG2 cells overexpressing PTP1b had a decreased ApoB100 degradation rate[48]. Thus, it is likely that enhanced ApoB degradation contributes to the secretion defect in L-PTP1b−/− mice, potentially via altered REDOX regulation and PUFA peroxidation.

Our study also demonstrates the utility of computational approaches to gain biological insights from complex multivariate datasets. First, we adapted GSEA[15] 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 FA metabolism). We evaluated several new ideas for multivariate modeling of phenotypic response data from underlying molecular measurements. We developed stochastic multivariate regression (SMR) to incorporate site-specific measurement error into the model-building procedure. This method biases selection of predictors toward higher quality measurements and offers a promising strategy to account for variable levels of measurement noise across largescale datasets. Two important considerations could improve SMR performance. First, optimization of measurement error models could improve results. Here, when available, the standard deviation from technical replicates was used (and, when unavailable, the dataset average standard deviation was used to estimate error, see Methods). However, as error itself is randomly distributed, this method may unfairly increase the weight or importance of measurements with randomly small standard deviations and, similarly but less worrisome, limit the influence of measurements with randomly large standard deviations. A second future direction would be to weight the final fit of the SMR model. Measurement error is currently incorporated into the variable-selection procedure, as it influences the significance ranking of coefficients, and, therefore indirectly biases model fit toward higher confidence predictor and response measurements.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The model fit in the SMR procedure, however, is unweighted, linear least squares, which does not incorporate measurement error. As highlighted in Figure 7B, 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).

Additionally, we evaluated methods to deal with systematically missing data points, a common feature of MS-based proteomic datasets, and applied set-enrichment analysis to gain additional, pathway-level insight into the variables underlying model prediction. The resulting multivariate analyses lead to models that predict lipid metabolic phenotypes as functions of underlying tyrosine phosphorylation sites, providing biological insights in a quantitative and unbiased way.

Although we have focused mainly 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.

Materials and Methods

Animal Studies

L-PTP1b−/− mice were generated as described[8]. Genotyping for the PTP1b floxed allele and the presence of Cre was performed by PCR. Hepatic PTP1b protein expression was assessed 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 free access to water and food. Diets consisted of Prolab 5P76 Isopro 3000 (NC) and Teklad TD.93075 (HFD). For tissue harvesting, mice were fasted overnight, anesthesized with an *i.p.*-injection of 150mg/kg pentobarbital sodium (Nembutal Lundbeck, Inc), and, when judged unconscious (~2min), the abdominal cavity was opened and livers were snap-frozen and extracted using liquid-nitrogen-chilled freeze clamps. For insulin stimulation, 10mU/g insulin (Novo Nordisk Novolin R) was *i.p.*-injected 10min. prior to harvesting. Tissues were stored for subsequent biochemical analyses at −80°C. Mouse procedures were performed in accordance with the guidelines and approval of the institutional animal care and use committees 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 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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Phosphotyrosine analysis

Peptide sample processing—In a 4°C room, aliquots of frozen liver, 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, which was added to original tissue lysate such that urea concentration in the 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 0.1% 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[13], 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[13]. Phosphopeptides were eluted from the IMAC column to a capillary precolumn.

LC-MS/MS—The capillary precolumn (100um i.d., packed with 10cm of ODS-A, 12nm, S-10um beads (YMC)) was connected to a capillary analytical column (50um 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 (<1um diameter))[49]. Peptides were eluted from the liquid chromatography column (flow rate ~ 20nL/min) 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

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(Matrix Science) against the National Center for Biotechnology Information (NCBI) mouse proteome with a 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[50].

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, 0.2% 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 quantification. 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 0.10 +/− 0.10 and 10% +/− 10%, respectively.

Principal component analysis and variance explained by diet and genotype

Principal component analyses[51] were used to decompose the insulin-stimulated (insulin) and basal phosphotyrosine data matrices separately. For a data matrix, X [N samples by M phosphosite measurements], the contributions of all sites to each of the principal components is preserved in the loadings matrix, P, and the position of each of the samples in principal component space is contained in the scores matrix, T. The relationship among P, T, and the corresponding data matrix, X, is:

$$X = \mathrm{TP}^T, \quad \text{(S1)}$$

where the dimensions of T, and P are [N × p] and [M × p] and N, M, p are the number of liver samples, number of phosphosite measurements, and the number of principal components used, respectively. (P^T is the transpose of P.)

PCA requires that the data matrix be complete. Given the systematically missing data points in the phosphotyrosine datasets that resulted from merging of multiple data-dependently acquired MS/MS experiments, techniques for imputing or estimating missing data were necessary for application of PCA. Analysis was limited to phosphorylation sites that had

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

data for at least 70% of the samples in the insulin or basal datasets, and data were mean-normalized. Multiple random imputation[52] was used to deal with the remaining missing data points. For each phosphosite measurement (column of X) with missing data points, a sample was randomly selected from those present to replace missing values. This procedure was repeated for each phosphosite with missing sample measurements, and the resulting, complete matrix, X_C, was used to estimate T and P. This random imputation procedure was repeated 1000 times, and results were averaged to yield matrices T_{ave} and P_{ave}. The columns of T_{ave} and P_{ave} were then made orthogonal and orthonormal, respectively, using the Gramm-Schmidt process.

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

$$Var_{exp}=1-\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(phos_{i,j}-phos_{j,est(i)})^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}(phos_{i,j}-phos_{i,mean})^2} \quad \text{(S2)}$$

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,…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 are 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. These and all subsequent computational and statistical analyses were performed in Matlab Version 7.10.0.499 (R2010a).

Heatmaps of genotype- and diet-dependent phosphorylation sites

To visualize genotype- or diet-dependent phosphosites in both global and condition-specific contexts, we used the following procedure. For detection of genotype-dependent phosphosites, we estimated the Pearson correlation coefficient and 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 HFD/insulin and NC/insulin correlations. Phosphosites with at least one context-specific, raw P-value < .025 were included in Figure 2B heatmap. Similarly, for 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 context-specific p-values were estimated analogously to the genotype case. Phosphosites with at least one context-specific, raw P-value cutoff of .025 were included in Figure S1 heatmap.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Phosphosite-Set Enrichment Analysis

PSEA directly follows from the GSEA protocol for multiple set-enrichment tests, described in[15]. Genotype and diet 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. Diet-dependence was analyzed analogously, and the independent data subsets were L-PTP1b−/− + insulin, control + insulin, basal L-PTP1b−/−, and basal control. For these analyses, phosphosites were ranked according to the overall direction of correlation coefficient multiplied by the $\log_{10}$-transformed p-value.

Phosphosite sets (Table S3) were tested for enrichment. For each analysis, only those sets with five phosphosites or more were tested. Thus, given incomplete data, the genotype and diet analyses tested 59 and 45 sets, respectively, while 56 sets were tested for relevance to phenotype prediction using SMR models. For each set, normalized enrichment scores (NESs) and significance were calculated as in[15]. Random NES distributions were generated by calculating set NESs from 500 lists of sites ranked according to significance of correlation to randomly permuted genotype or diet vectors. Raw p-values for enrichment among diet- and genotype-dependent sets were derived from corresponding random NES distributions, and Benjamini-Hochberg corrections was used to estimate false-discover rate (FDR) and control for multiple-hypothesis testing.

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.

The stochastic multivariate regression (SMR) method described below was used to generate ranked lists of sites relevant to SMR prediction of steatosis, ω3/ ω6 PTP1b-dependent PUFA, and ω7/ ω9 PTP1b-dependent PUFA 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.

Lipidomics analysis

Lipid extraction—In a 4°C 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 $CHCl_3$: MeOH: H_2O. To quantify the absolute abundances of select lipids, reference standards were spiked into the chloroform phase (670nmol (HFD) or 330nmol (NC) glyceryl triheptadecanoate (Sigma), 1.5μmol (HFD) or 750nmol (NC)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Cholesterol-25,26,26,26,27,27,27-d7 (C/D/N Isotopes), 5nmol UC^{13}-oleic acid (Spectra Stable Isotopes), 500pmol UC^{13}-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, 4°C for 5min to separate the organic and aqueous layers. The organic layer was transferred to fresh glass vials twice by Pasteur pipette. The final glass vial was placed under a stream of nitrogen until the organic phase evaporated. Samples were stored at −80°C and later dissolved in chloroform for LC-MS analysis[53].

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 (5um, 4.6mm × 50mm) was used in combination with a precolumn (C18, 3.5um, 2mm × 20mm). Mobile phases A and B were 95:5 water:methanol and 60:35:5 isopropanol:methanol:water, respectively; both contained 0.1% ammonium hydroxide. For positive-mode LC analysis, a Luna (Phenomenex) C5 column (5um, 4.6mm × 50mm) was used with a precolumn (C4, 3.5um, 2mm × 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 0.1% 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 0.4mL/min, and equilibration with 100% solvent A at 0.5mL/min for 8min[53].

For preliminary, discovery-motivated experiments, 16% of the sample was run in a 30uL injection volume for both modes. In quantitative, follow-up positive mode analyses focusing on triglycerides, $1/240^{th}$ (HFD) or $1/80^{th}$ (NC) of the sample extract was injected in 30uL. 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 $1/80^{th}$–$1/240^{th}$ of the extract, and a 15uL 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, $1/6^{th}$ of the sample extract was injected using a 5uL volume and an 80min gradient was developed: 0–5min 100% A at 0.1mL/min, linear increase in solvent B from 20% to 40% from 5–15min at 0.4mL/min, (slower) linear increase in solvent B from 40% to 80% from 15–55min at 0.4mL/min, linear increase in solvent B from 80% to 100% from 55–65 at 0.4mL/min, isocratic 100% B from 65–73min at 0.4mL/min and, finally, a 7min equilibration step with 100% A at 0.5mL/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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Automated discovery of differential metabolites—XCMS[54] 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 change carbon length of 48 to 56. Manual integration of ion chromatograms (20 ppm *m/z* window) was performed, 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 principal 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 were 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 (spiked 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.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To overcome these issues, we devised a new method for integrating the high and low abundance datasets. 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 high- and low-abundance runs (> 10 fatty acid pairs) into each sample calibration. Principal component analysis is used to find one principal 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 principal component, so that all of the fatty acids have positions on the principal component. The positions of the standards are used to determine the linear relationship between principal 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 principal 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 ω3, C18:3 ω6, 18:3 ω9, C20:3 ω3, C20:3 ω6, C20:3 ω9, C22:3 ω3 standards were purchased from Cayman Chemicals, while C24:6 ω3 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 performed: 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 co-injected with isomer standards, whose concentrations were 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 ω3, C18:3 ω6, and C18:3 ω9 and (2) 2.5pmol C18:3 ω3 and C18:3 ω6 with 6.25pmol C18:3 ω9 (Figure S2A, 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 ω3, ω6, ω9), and (2) the endogenous peak of interest eluted in between the ω6 and ω9 peaks, we suspect that the endogenous peak corresponds to C18:3 ω7. Because we did not have a standard to confirm this assignment, this peak is referred to as C18:3 ω7* 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 ω3 and C20:3 ω6 with 3.4pmol C20:3 ω9 and (2) 1pmol C20:3 ω3 and C20:3 ω6 with 3.4pmol C20:3 ω9 (Figure S2B). C20:3 ω3 and ω6 co-eluted with the first peak, while C20:3 ω9 co-eluted with the second peak. In the absence of an ω7 C20:3 standard, we concluded that the most significantly differential peak is either ω9, ω7, or a combination of both isomers, and we therefore refer to this peak as C20:3 ω9/ω7*.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

For C22:3, 150fmol and 450fmol C22:3 ω3 were run with endogenous lipid sample in two separate runs (Figure S2C). C22:3 ω3 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 ω6 would co-elute with C22:3 ω3, as was the case for C20:3 ω3 and ω6, and that the second peak, similar to the second peak of C20:3, corresponds to ω7 and/or ω9 species. Thus, we refer to the PTP1b-dependent C22:3 peak as ω7*/ω9*.

For C24:6, co-injection runs with 50fmol and 300fmol C24:6 ω3 were performed (Figure S2D). The single C24:6 peak co-eluted with a C24:6 ω3 standard, confirming its identification as 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 ω7/ ω9 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 ω6 position. Here, we denote the peak as C24:6 ω3/ ω6*.

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 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[8], using the Stanbio Triglyceride Liquicolor Test kit.

Serum triglyceride analysis—Serum triglyceride analysis was performed as described[25], with the following changes: a Jupiter C4 column (4.6 × 150mm, 5μm particle size, 30nm pore size, Phenomenex) was used at a 700uL/min flow rate, and data were acquired on an Agilent 6220 LC-ESI-TOF instrument.

Unsupervised clustering analysis of lipidomics datasets

Lipid measurements were scaled by z-scoring, and a correlation-based similarity matrix was used as input to the affinity propagation clustering algorithm[55]. Measurements within AFP-derived clusters were ordered using hierarchical clustering with correlation distance.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Biochemical analysis

Cytokine 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 the Bio-Plex 200 System.

Immunoblot analysis—Antibodies for pS79 ACC, ACC, pT172 AMPK, AMPK, pT180 and pY182 p38, and β-actin were purchased from Cell Signaling Technology.

Stochastic Multivariate Regression

Stochastic multivariate regression (SMR) employs Gaussian sampling of the data (by sampling from distributions derived from experimental technical replicates) and bootstrapping techniques (where distinct samples are bootstrapped) to construct distributions of regression coefficients connecting phosphosites to phenotypic measurements. Required algorithm inputs include an independent variable matrix (e.g., of phosphosite measurements), X [N×M] and X_{std} [N×M], corresponding to N samples and M measured variables, and a dependent variable matrix (e.g., of lipid measurements), Y [N × V] and Y_{std} [N x 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 phosphosite 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 in Matlab. 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:

$$X = \mathrm{TP}^T,\ T = X\,P/(P^T P) = \mathrm{XW}^T, \quad \text{(S3)}$$

$$Y = \mathrm{UC}^T \approx \mathrm{TC}^T = \mathrm{XW}^T C^T = \mathrm{XB}, \quad \text{(S4)}$$

where the X and Y scores matrices, T [N × pcs] and U [N × pcs], are optimized to co-vary maximally; pcs is the number of model principle components, and P [M × pcs], C [P × pcs], W [M × pcs], and B [M × P] are the X loadings, Y loadings, X weights, and regression coefficient matrices, respectively[56]. The decomposition of 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[57].

The basic SMR algorithm is as follows:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

1. *Gaussian sampling for measurement noise*. For each data point present in X, draw a random sample, $X_{gaus}(n,m)$ from a Gaussian distribution with mean, X(n,m), and standard deviation, $X_{std}(n,m)$. Similarly, draw $Y_{gaus}(n,v)$, based on Y(n,v) and $Y_{std}(n,v)$, n = {1,2, …,N}, v = {1,2, …,V}. Recall that the N rows and and M columns of X correspond to samples and variables, respectively. V corresponds to the number of response variable measurements. Means and standard deviations of X and Y are estimated as the sample mean and sample deviations measured across technical replicates. If a data point does not have replicate measurements, its standard deviation is estimated as the average sample deviation of the corresponding data matrix.

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 $Y_{zscore,v}$ [N × 1]. (Averaging Y simplifies subsequent model reduction steps and does not change the modeling outcome so long as the response measurements are related (non-orthogonal), as is the case for the response variables considered here.)

3. *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 × M] and [N × 1], respectively. In this way, measurement noise is modeled and accounted for in Step 1, while bootstrapping at the level of distinct samples in Step 3 accounts for biological sample variability.

4. *Build the bootstrap regression model*. Use PLSR, with number of principle components equal to one minus the rank of X_{boot}, to estimation the linear regression coefficients B_{boot} [M × 1].

5. *Multiple imputation and Gaussian sampling*. Repeat steps 1–4 until a user-determined number, J, of bootstrapped models have been built. Store each model's vector of coefficients B_{boot} in the matrix B_{matrix} [J × 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, …,3}, and F = {1,2, …,20}.

R^2 and Q^2 statistics—Models were evaluated by estimating R^2, goodness-of-fit, and Q^4, goodness-of-prediction, statistics:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$R^2=1-\frac{\sum_{n=1}^{N}\left(Y^{meas,n}-Y^{pred,n}\right)^2}{\sum_{n=1}^{N}\left(Y^{meas,n}-Y^{mean}\right)^2} \quad \text{(S5)}$$

$$Q^2=1-\frac{\sum_{n=1}^{N}\left(Y^{meas,n}-Y^{pred,loocv(n)}\right)^2}{\sum_{n=1}^{N}\left(Y^{meas,n}-Y^{mean,loocv(n)}\right)^2} \quad \text{(S6)}$$

where $Y^{meas,n}$ corresponds to the experimentally determined response measurement n, Y^{mean} is the average of the response measurements, $Y^{pred,n}$ is the model prediction for response measurement n, $Y^{pred,loocv(n)}$ is the model prediction for response n, from a leave-one-out-cross-validation (LOOCV) model, which excludes measurements for sample n, and, similarly, $Y^{mean,loocv(n)}$ is the mean of the response measurements, excluding measurement n. A Q^4 value should estimate the variance explained by a model for an independent sample. Typical Q^4 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[29, 31, 58]. Such a protocol inflates the Q^4 value, because the left-out sample is not entirely independent of the LOOCV model construction. Although the left-out sample is not involved in model fitting, it was used for the variable selection step and, for the calculation of Q^4 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 subset only, and we found that this method results in lower Q^4 values relative to the previously described Q^4 estimation procedures, which account for model fit only.

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:

$$VIP(m,pcs)=\sqrt{\frac{\sum_{pc=1}^{pcs} Var_{exp,Y}(pc)W(m,pc)^2}{\sum_{pc=1}^{pcs} Var_{exp,Y}(pc)}} \quad \text{(S7)}$$

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 pc^{th} PLSR principle component[59]. 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 model reduction based on VIP score, it is necessary to estimate Var_{exp} and W from Equation S7, and we estimated these from the J X-loadings matrices, P_{boot} [M × pcs], that resulted from each bootstrap model-building process (see Step 5 and definition of a loadings matrix, P, in Equation S5, above). Specifically, each of the P_{boot} matrices is averaged to create an averaged X-loadings matrix, P_{ave} [M × pcs]. Thus, by averaging,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

variable measurements that consistently contribute to the PLSR principle component plane will have higher loading values and contribution to the model more than measurements associated with larger amounts of error or that have missing values which must be randomly imputed (see "Missing data techniques section" below). To make this loadings matrix consistent with loadings matrix produced directly from PLSR algorithms, P_{ave} is orthonormalized row-wise. P_{ave} is then used to derive W and $Var_{exp,Y}$. The scores matrix, T [N × pcs], are estimated by the equation $T = X_c P_{ave} T$ and the weights, W, are estimated as $P_{ave}/(P_{ave} T P_{ave})$. To derive $Var_{exp,Y}$, the rows of T are orthogonalized, and, then the Y loadings, C [pcs × 1] are 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 $Var_{exp,Y}(pc)$, are calculated using the equation: $Var_{exp,Y}(pc) = 1 - (Y^* - T(pc) TC(pc)) T(Y - T(pc)^T C(pc))/Var_Y$, where T(pc) is the pcth 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*. The VIP score is calculated for each measurement and total number of pcs in the model (this ranking method substitutes for Step 6 in the SMR algorithm). Step 7 is then performed based on VIP-score model reduction.

Missing data techniques for stochastic multivariate regression

The data matrices refer to matrices defined 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 HFDB data subset of phosphotyrosine measurements, X_{HFDB}, and X_{HFDB} was divided into three distinct blocks of complete data subsets, as shown in Figure S7A. SMR was used to build individual models for each of the subsections, referred to as the "Complete" (in which data exists for all sample conditions), "Run 1", and "Run 2" subset models. ("Run 1" and "Run 2" models each had data for at least 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", because a variable was incorporated into a final model only if it was significant to sub-models from all data subsets containing that measurement. Specifically, we select a variable for the final model only if the significance of its regression coefficient, p_m, is less than a pre-determined significance cut-off, $p_{cutoff} \in [0,1]$, for all models incorporating that measurement. The second rule is referred to as "Complete subset rules", because 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 $p_m < p_{cutoff}$. Otherwise, if the variable is in "Run 1" or "Run 2" subset only, retain this variable in the model with the same criteria: if $p_m < p_{cutoff}$.
3. *Gaussian Multiple Random Imputation (MRI) Methods (General, Treatment, and P-value)*. In "General" Gaussian MRI, the missing samples for a measurement are imputed from those available by drawing a Gaussian sample whose mean and standard deviation correspond to the available sample measurements. This single random imputation step is repeated multiple times. Conveniently, such an

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

imputation step seamlessly integrates into the SMR algorithm, as the imputation step can immediately follow the Gaussian sampling step (Step 1, see schematic in Figure S7B). “Treatment” Gaussian MRI makes use of treatment (e.g., genotype) information when imputing missing datapoints. In “Treatment” Gaussian MRI, only the values from samples with the same treatment conditions are used to impute the missing values. For this dataset, missing data points corresponding to L-PTP1b −/− mice were sampled from a Gaussian distribution whose mean and standard deviation corresponded to the L-PTP1b−/− samples present for that measurement, and missing control sample values were randomly imputed from a Gaussian distribution based on control samples. In “P-value” Gaussian MRI, missing datapoints are imputed using either “General” or “Treatment” Gaussian MRI, contingent on ANOVA, in which a treatment-dependent effect is tested at a user-specified significance threshold. If the measurement is significantly treatment dependent, inferred conditions are imputed using “Treatment” MRI; otherwise, “General” MRI is used. For the L-PTP1b−/− dataset, a two-sided Student t test comparison of L-PTP1b−/− and control phosphorylation levels (at $\alpha = .05$ significance level) was performed on the measured samples, and treatment-dependent or treatment-independent MRI was used to impute missing values, contingent on associate p-values.

4. *Traditional MRI Methods (General, Treatment, and P-value)*. The traditional MRI methods are identical to Gaussian MRI methods, except that the value used to replace missing data points is randomly selected from those samples available, with equal probability.

Evaluation of missing data techniques

Simulation of missing data—To evaluate missing data techniques specifically for SMR with the HFDB phosphotyrosine dataset as the independent variable matrix, X, we utilized the complete data subset of X. The full dataset (Figure 6B) contained 228 phosphosite measurements across 13 samples, while the complete data subset 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, although with fewer samples (Figure 6B). 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 three models were built using SMR for steatosis, ω3+ω6 PTP1b-dependent PUFA, and ω7+ω9 PTP1b-dependent PUFA. 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^4 value. Those measurements selected for a final model served as true positives, and those measurements excluded from a model served as true negatives.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The nine missing data methods (described above) were used to build the three phenotype models from each of the 10 simulated datasets. True positive rate (TPR) and false positive rate (FPR) were calculated for each model and simulated dataset, as parametric functions of a decision rule. We applied the following decision rule: include a measurement in the final model if the significance of the associated regression coefficient is less than p_{cutoff}. To build ROC curves, we varied p_{cutoff} from 0 to 1 and counted true positives and false positives relative to the complete data (gold standard) models. TPR(p_{cutoff}) and FPR(P_{cutoff}) curves were averaged across simulated missing datasets for each model and plotted in Figure 6C.

Final models

Individual models of steatosis, ω3+ω6 PTP1b-dependent PUFA, and ω7+ω9 PTP1b-dependent PUFA as functions of the complete HFDB phosphotyrosine dataset [228 phosphosites X 13 samples] were constructed using SMR and "Treatment" Traditional MRI; these models were built using 1, 2, and 1 principle component(s), respectively.

Calculation of R^2 and Q^4—Estimation 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^4 for the models 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^4 values. Estimation of the error associated with the Q^4 values was accomplished by leave-two-out cross-validation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Ms. Angel Sing (OCI) for generating the mice for this study, Edwin Homan (Harvard) for help with fatty acid separation techniques, Robert Gerszten (Mass General Hospital) for helpful advice, and Kristen Naegle (Washington University) for assistance with phosphosite annotations in PTMScout. This work was supported by a grant from Pfizer Inc., NIH grants 5R24DK090963, U54-CA112967, CA49152 R37 (B.G.N.), and R01-DK080756 (J.K.K.), and the UMass Mouse Metabolic Phenotyping Center Grant (U24-DK093000) (J.K.K.). E.R.M. has been supported by a Graduate Research Fellowship from the National Science Foundation. Work in B.G.N.'s lab is partially supported by the Ontario Ministry of Health and Long Term Care and the Princess Margaret Hospital Foundation.

References

1. Biddinger SB, Kahn CR. From mice to men: insights into the insulin resistance syndromes. Annu Rev Physiol. 2006; 68:123–58. [PubMed: 16460269]
2. Kashyap SR, Defronzo RA. The insulin resistance syndrome: physiological considerations. Diab Vasc Dis Res. 2007; 4:13–9. [PubMed: 17469039]
3. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. Nutr Rev. 2012; 70:3–21. [PubMed: 22221213]
4. Rutter GA. Diabetes: the importance of the liver. Curr Biol. 2000; 10:R736–8. [PubMed: 11069096]
5. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nat Rev Mol Cell Biol. 2006; 7:85–96. [PubMed: 16493415]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

6. Yip SC, Saha S, Chernoff J. PTP1B: a double agent in metabolism and oncogenesis. Trends Biochem Sci. 2010; 35:442–9. [PubMed: 20381358]
7. Elchebly M, Payette P, Michaliszyn E, Cromlish W, Collins S, Loy AL, et al. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. Science. 1999; 283:1544. [PubMed: 10066179]
8. Delibegovic M, Zimmer D, Kauffman C, Rak K, Hong EG, Cho YR, et al. Liver-specific deletion of protein-tyrosine phosphatase 1B (PTP1B) improves metabolic syndrome and attenuates diet-induced endoplasmic reticulum stress. Diabetes. 2009; 58:590–9. [PubMed: 19074988]
9. Agouni A, Mody N, Owen C, Czopek A, Zimmer D, Bentires-Alj M, et al. Liver-specific deletion of protein tyrosine phosphatase (PTP) 1B improves obesity- and pharmacologically induced endoplasmic reticulum stress. Biochem J. 2011; 438:369–78. [PubMed: 21605081]
10. Ji C, Kaplowitz N. ER stress: can the liver cope? J Hepatol. 2006; 45:321–33. [PubMed: 16797772]
11. Ozcan U, Cao Q, Yilmaz E, Lee AH, Iwakoshi NN, Ozdelen E, et al. Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. Science. 2004; 306:457–61. [PubMed: 15486293]
12. Zinker BA, Rondinone CM, Trevillyan JM, Gum RJ, Clampit JE, Waring JF, et al. PTP1B antisense oligonucleotide lowers PTP1B protein, normalizes blood glucose, and improves insulin sensitivity in diabetic mice. Proc Natl Acad Sci U S A. 2002; 99:11357–62. [PubMed: 12169659]
13. Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, et al. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. Mol Cell Proteomics. 2005; 4:1240–50. [PubMed: 15951569]
14. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proceedings of the National Academy of Sciences. 2007; 104:5860.
15. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. 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. 2005; 102:15545. [PubMed: 16199517]
16. Ren L, Chen X, Luechapanichkul R, Selner NG, Meyer TM, Wavreille AS, et al. Substrate specificity of protein tyrosine phosphatases 1B, RPTPalpha, SHP-1, and SHP-2. Biochemistry. 2011; 50:2339–56. [PubMed: 21291263]
17. Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM. Temporal dynamics of tyrosine phosphorylation in insulin signaling. Diabetes. 2006; 55:2171–9. [PubMed: 16873679]
18. Zhang K, Kaufman RJ. From endoplasmic-reticulum stress to the inflammatory response. Nature. 2008; 454:455–62. [PubMed: 18650916]
19. Hitosugi T, Kang S, Vander Heiden MG, Chung TW, Elf S, Lythgoe K, et al. Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. Sci Signal. 2009; 2:ra73. [PubMed: 19920251]
20. Fan J, Hitosugi T, Chung TW, Xie J, Ge Q, Gu TL, et al. Tyrosine phosphorylation of lactate dehydrogenase A is important for NADH/NAD(+) redox homeostasis in cancer cells. Mol Cell Biol. 2011; 31:4938–50. [PubMed: 21969607]
21. Feng J, Lucchinetti E, Enkavi G, Wang Y, Gehrig P, Roschitzki B, et al. Tyrosine phosphorylation by Src within the cavity of the adenine nucleotide translocase 1 regulates ADP/ATP exchange in mitochondria. Am J Physiol Cell Physiol. 2010; 298:C740–8. [PubMed: 20007455]
22. Hitosugi T, Fan J, Chung TW, Lythgoe K, Wang X, Xie J, et al. Tyrosine phosphorylation of mitochondrial pyruvate dehydrogenase kinase 1 is important for cancer metabolism. Mol Cell. 2011; 44:864–77. [PubMed: 22195962]
23. Homan EA, Kim YG, Cardia JP, Saghatelian A. Monoalkylglycerol ether lipids promote adipogenesis. J Am Chem Soc. 2011; 133:5178–81. [PubMed: 21428285]
24. Stefan N, Peter A, Cegan A, Staiger H, Machann J, Schick F, et al. Low hepatic stearoyl-CoA desaturase 1 activity is associated with fatty liver and insulin resistance in obese humans. Diabetologia. 2008; 51:648–56. [PubMed: 18286258]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

25. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest. 2011; 121:1402–11. [PubMed: 21403394]
26. Cohen, PR. Empirical methods for artificial intelligence. MIT press; Cambridge, Massachusetts: 1995.
27. Nelson PRC, Taylor PA, MacGregor JF. Missing data methods in PCA and PLS: Score calculations with incomplete observations. Chemometrics and intelligent laboratory systems. 1996; 35:45–65.
28. Wolf-Yadlin A, Kumar N, Zhang Y, Hautaniemi S, Zaman M, Kim HD, et al. Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. Molecular systems biology. 2006:2.
29. Kumar N, Wolf-Yadlin A, White FM, Lauffenburger DA. Modeling HER2 effects on cell behavior from mass spectrometry phosphotyrosine data. PLoS computational biology. 2007; 3:e4. [PubMed: 17206861]
30. Rubin DB. Multiple imputation after 18+ years. Journal of the American Statistical Association. 1996:473–89.
31. Huang PH, Miraldi ER, Xu AM, Kundukulam VA, Del Rosario AM, Flynn RA, et al. Phosphotyrosine signaling analysis of site-specific mutations on EGFRvIII identifies determinants governing glioblastoma cell growth. Mol Biosyst. 2010; 6:1227–37. [PubMed: 20461251]
32. Pan M, Cederbaum AI, Zhang YL, Ginsberg HN, Williams KJ, Fisher EA. Lipid peroxidation and oxidant stress regulate hepatic apolipoprotein B degradation and VLDL production. J Clin Invest. 2004; 113:1277–87. [PubMed: 15124019]
33. Brandt, TA.; Crooke, ST.; Ackermann, EJ.; Xia, X.; Morgan, ES.; Liu, Q.; Greary, RS.; Bhanot, S. ISIS 113715, a novel PTP-1B antisense inhibitor, improves glycemic control and dyslipidemia and increases adiponectin levels in T2DM subjects unctrolled on stable sulfonylurea therapy. American Diabetes Assocation; Carlsbad, CA, USA: 2010.
34. Wang Q, Jiang L, Wang J, Li S, Yu Y, You J, et al. Abrogation of hepatic ATP-citrate lyase protects against fatty liver and ameliorates hyperglycemia in leptin receptor-deficient mice. Hepatology. 2009; 49:1166–75. [PubMed: 19177596]
35. Wellen KE, Hatzivassiliou G, Sachdeva UM, Bui TV, Cross JR, Thompson CB. ATP-citrate lyase links cellular metabolism to histone acetylation. Science. 2009; 324:1076–80. [PubMed: 19461003]
36. Koltun DO, Zilbershtein TM, Migulin VA, Vasilevich NI, Parkhill EQ, Glushkov AI, et al. Potent, orally bioavailable, liver-selective stearoyl-CoA desaturase (SCD) inhibitors. Bioorg Med Chem Lett. 2009; 19:4070–4. [PubMed: 19577469]
37. Schenkman JB, Jansson I. The many roles of cytochrome b5. Pharmacol Ther. 2003; 97:139–52. [PubMed: 12559387]
38. Guillou H, Zadravec D, Martin PG, Jacobsson A. The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. Prog Lipid Res. 2010; 49:186–99. [PubMed: 20018209]
39. Guillou H, D'Andrea S, Rioux V, Barnouin R, Dalaine S, Pedrono F, et al. Distinct roles of endoplasmic reticulum cytochrome b5 and fused cytochrome b5-like domain for rat Delta6-desaturase activity. J Lipid Res. 2004; 45:32–40. [PubMed: 14563830]
40. Rosendal J, Ertbjerg P, Knudsen J. Characterization of ligand binding to acyl-CoA-binding protein. Biochemical Journal. 1993; 290:321. [PubMed: 7680855]
41. Færgeman NJ, Sigurskjold BW, Kragelund BB, Andersen KV, Knudsen J. Thermodynamics of ligand binding to acyl-coenzyme A binding protein studied by titration calorimetry. Biochemistry. 1996; 35:14118–26. [PubMed: 8916897]
42. Taskinen JP, van Aalten DM, Knudsen J, Wierenga RK. High resolution crystal structures of unliganded and liganded human liver ACBP reveal a new mode of binding for the acyl-CoA ligand. Proteins. 2007; 66:229–38. [PubMed: 17044054]
43. Manevich Y, Reddy KS, Shuvaeva T, Feinstein SI, Fisher AB. Structure and phospholipase function of peroxiredoxin 6: identification of the catalytic triad and its role in phospholipid substrate binding. J Lipid Res. 2007; 48:2306–18. [PubMed: 17652308]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

44. Yamazaki T, Sasaki E, Kakinuma C, Yano T, Miura S, Ezaki O. Increased very low density lipoprotein secretion and gonadal fat mass in mice overexpressing liver DGAT1. J Biol Chem. 2005; 280:21506–14. [PubMed: 15797871]
45. Liu X, Strable MS, Ntambi JM. Stearoyl CoA desaturase 1: role in cellular inflammation and stress. Adv Nutr. 2011; 2:15–22. [PubMed: 22211186]
46. Coleman RA, Lee DP. Enzymes of triacylglycerol synthesis and their regulation. Prog Lipid Res. 2004; 43:134–76. [PubMed: 14654091]
47. Ginsberg HN, Fisher EA. The ever-expanding role of degradation in the regulation of apolipoprotein B metabolism. J Lipid Res. 2009; 50 (Suppl):S162–6. [PubMed: 19050312]
48. Qiu W, Avramoglu RK, Dube N, Chong TM, Naples M, Au C, et al. Hepatic PTP-1B expression regulates the assembly and secretion of apolipoprotein B-containing lipoproteins: evidence from protein tyrosine phosphatase-1B overexpression, knockout, and RNAi studies. Diabetes. 2004; 53:3057–66. [PubMed: 15561934]
49. Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. Anal Chem. 2000; 72:4266–74. [PubMed: 11008759]
50. Naegle KM, Gymrek M, Joughin BA, Wagner JP, Welsch RE, Yaffe MB, et al. PTMScout, a Web resource for analysis of high throughput post-translational proteomics studies. Mol Cell Proteomics. 2010; 9:2558–70. [PubMed: 20631208]
51. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and intelligent laboratory systems. 1987; 2:37–52.
52. Rubin, DB. Multiple imputation for nonresponse in surveys. Wiley Online Library; 1987.
53. Vinayavekhin N, Saghatelian A. Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach. J Am Chem Soc. 2011; 133:17168–71. [PubMed: 21973308]
54. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006; 78:779–87. [PubMed: 16448051]
55. Frey BJ, Dueck D. Clustering by passing messages between data points. Science. 2007; 315:972. [PubMed: 17218491]
56. Höskuldsson A. PLS regression methods. Journal of Chemometrics. 1988; 2:211–28.
57. Geladi P, Kowalski BR. Partial least-squares regression: a tutorial. Analytica chimica acta. 1986; 185:1–17.
58. Janes KA, Kelly JR, Gaudet S, Albeck JG, Sorger PK, Lauffenburger DA. Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. J Comput Biol. 2004; 11:544–61. [PubMed: 15579231]
59. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. Multi-and megavariate data analysis. Umetrics. 2006

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

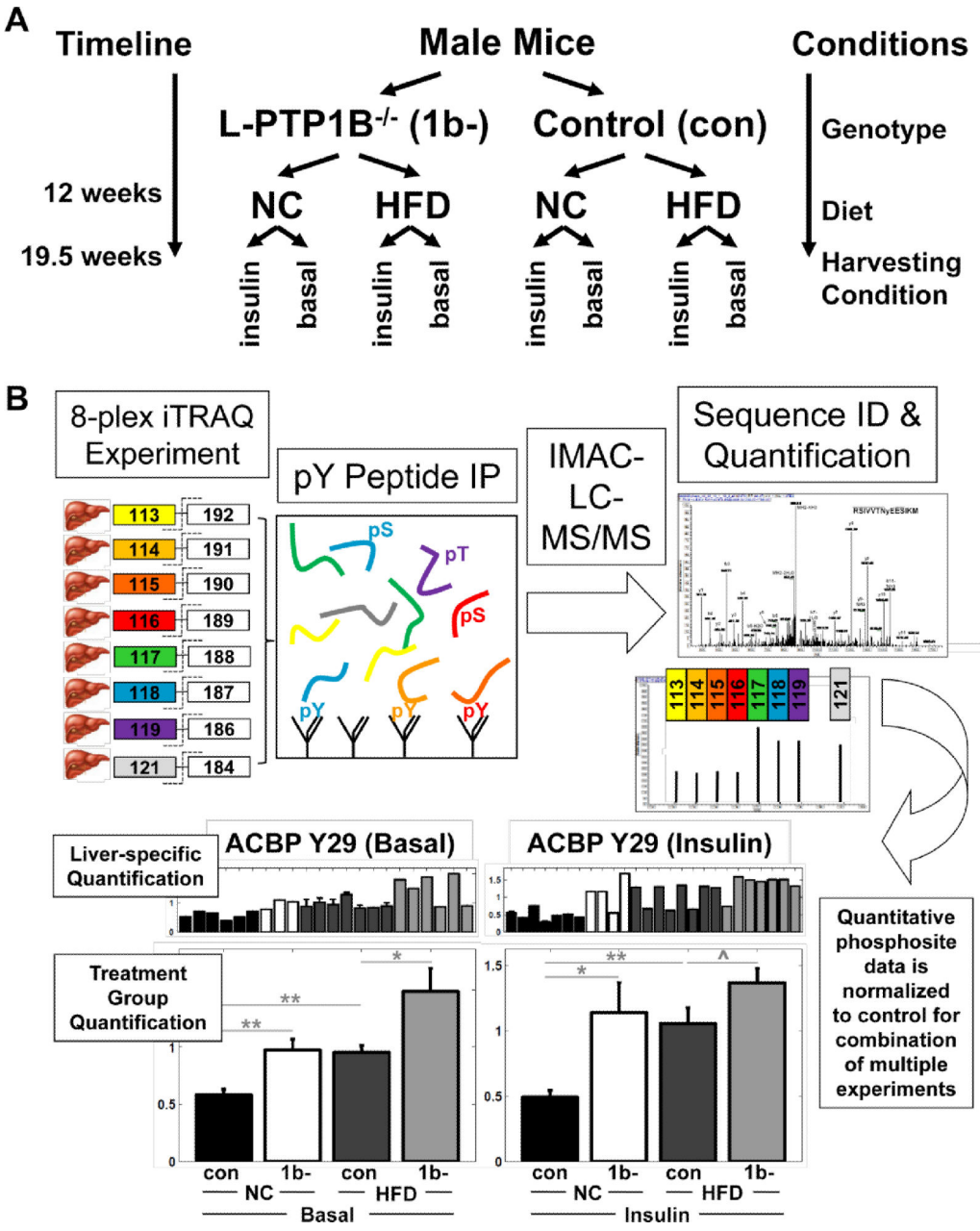

Figure 1. Study Design and Phosphotyrosine Analysis

(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.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

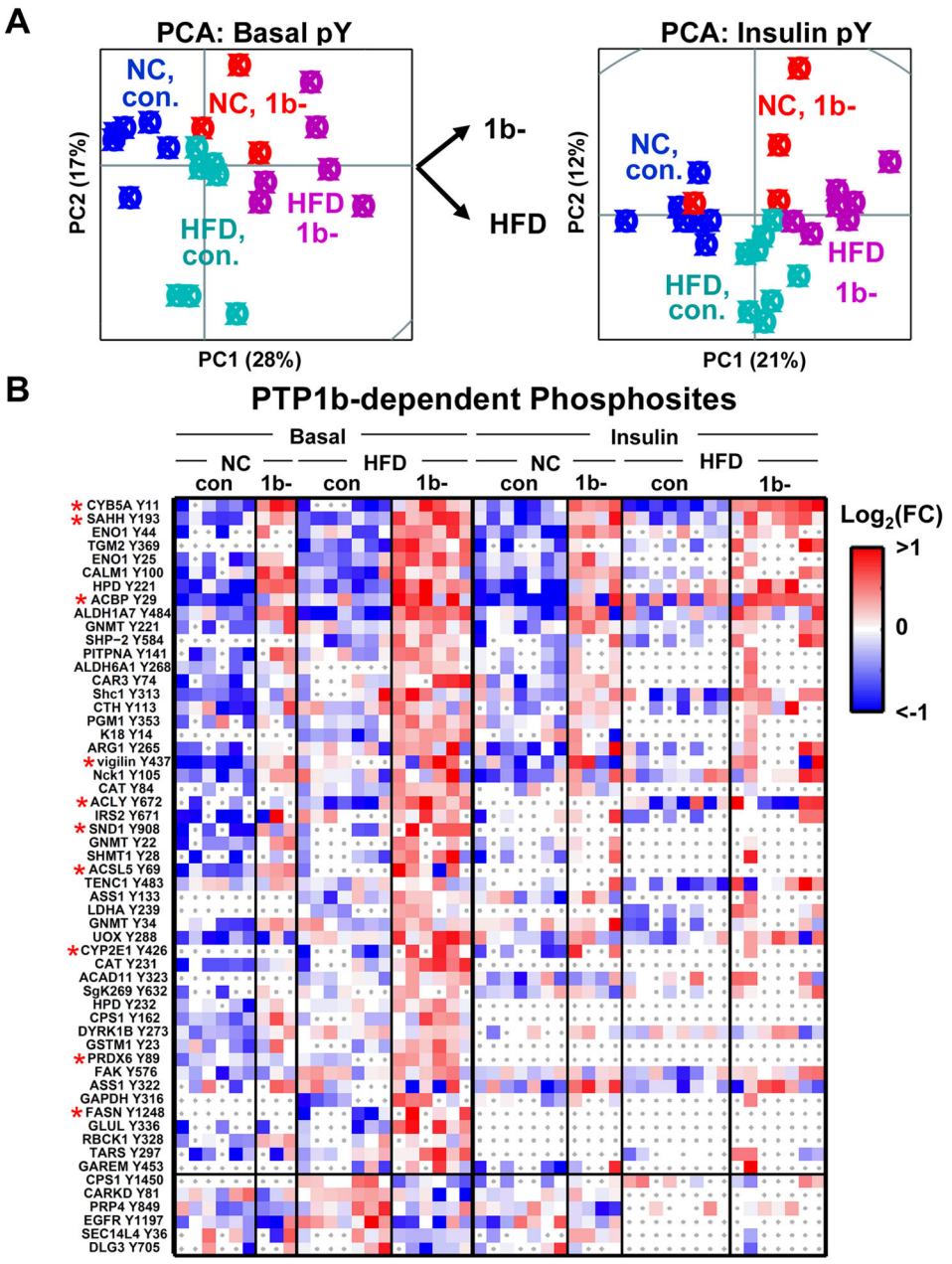

Figure 2. Diet and Genotype Dependencies in the Phosphotyrosine Network

(A) Insulin-stimulated and basal liver samples are plotted in the principal component plane as a function of mean-normalized tyrosine phosphorylation profiles.

(B) Heatmaps of the phosphorylation sites most significantly correlated to genotype (raw, context-specific p-value < .025, see Methods). For basal or insulin datasets, the phosphorylation level of a phosphosite was normalized to the corresponding phosphosite mean (basal or insulin) and then $\log_2$-transformed. Missing data points are denoted by grey dots over white boxes. Red asterisks denote lipid metabolic enzymes. The phosphosites are separated into two clusters, depending on whether phosphorylation on those sites increases or decreases upon PTP1b deletion.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

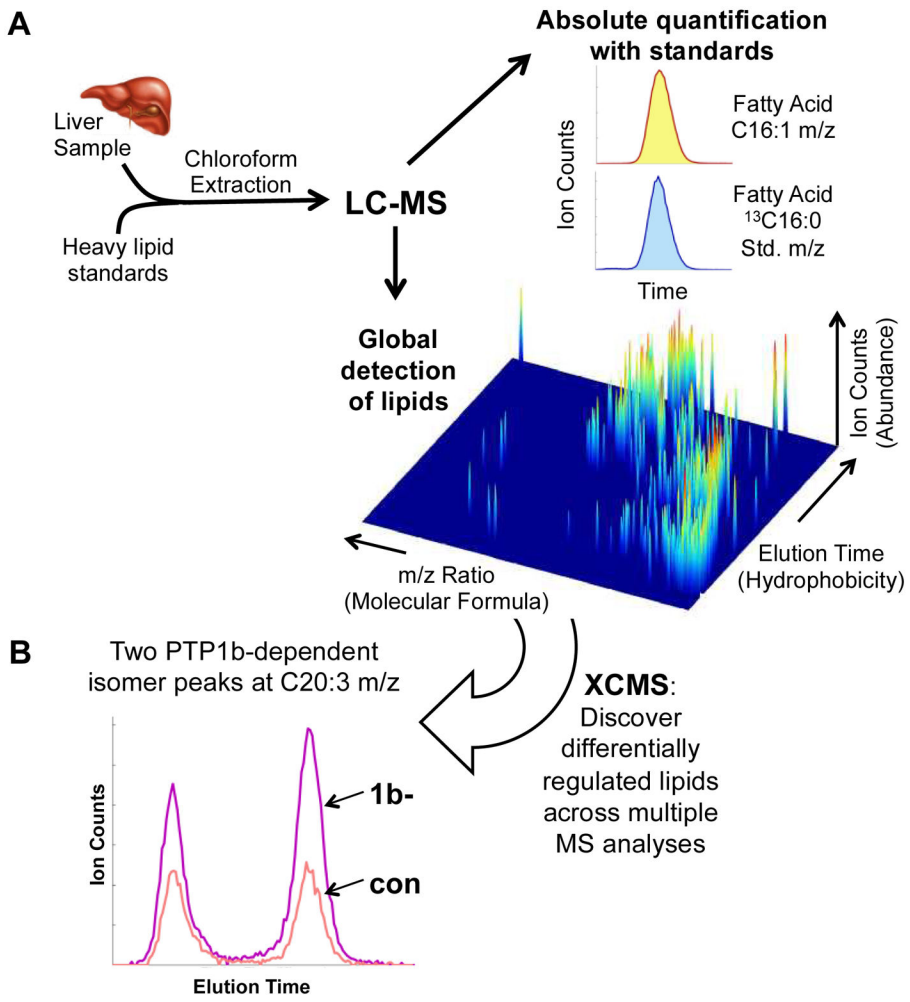

Figure 3. Global Lipidomics Analysis for *De Novo* Discovery

(A) Lipids were chloroform-extracted from individual livers. For absolute quantification, ^{13}C- or ^{2}H-labeled standards were spiked into the chloroform phase and endogenous lipids were quantified relative to standards. Lipid samples were individually analyzed by LC-MS, creating a 3D lipidomics profile for each tissue sample.

(B) The software program, XCMS[54], was used to identify differences in lipid profiles across multiple analyses. Two representative ion chromatograms of the XCMS-detected and PTP1b-dependent C20:3 FA isomers are superimposed.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

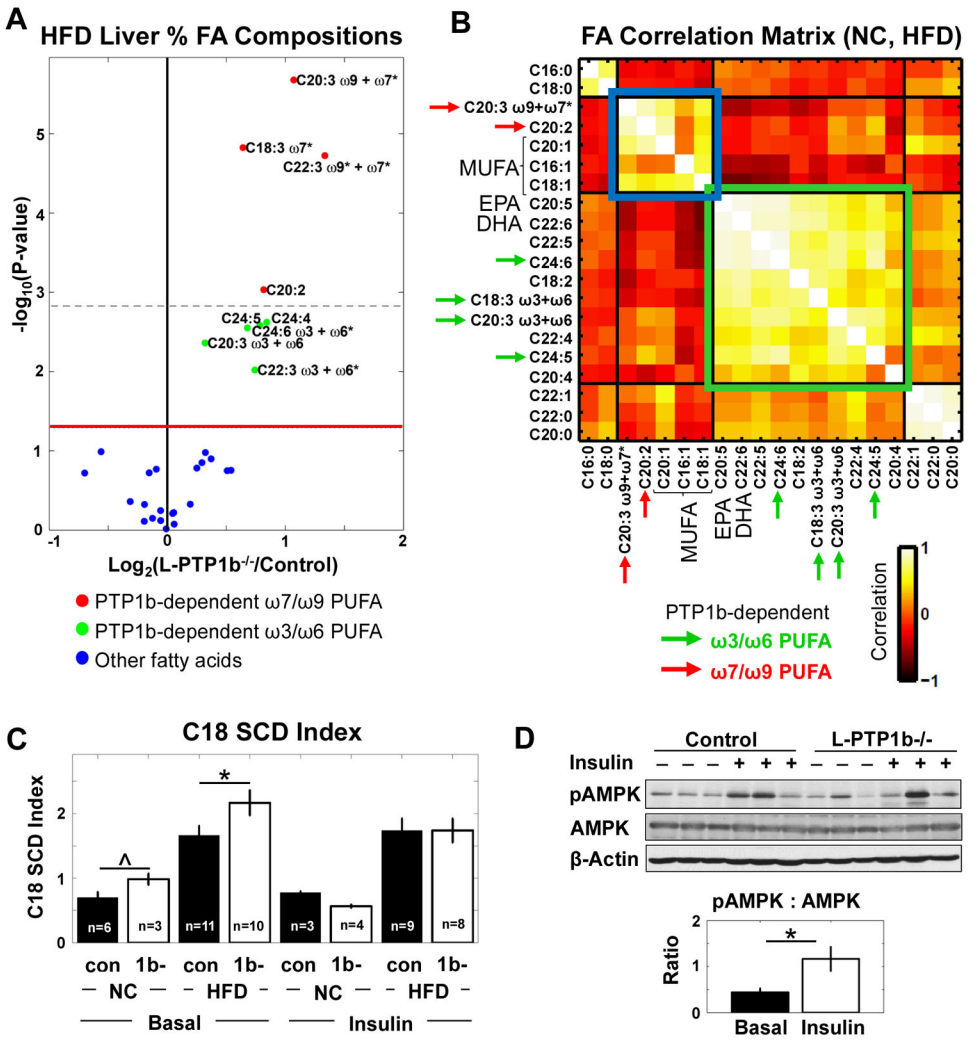

Figure 4. PTP1b-dependent Changes in Free Fatty Acid (FFA) Metabolism

(A) Volcano plot of % FA compositions on HFD (control, n = 19, L-PTP1b−/−, n = 17). P-values are estimated by 2-sided t tests. The red line denotes P = .05, while the dotted grey line shows the Bonferroni significance cut off to control for multiple hypotheses at FDR = .05. Asterisks indicate uncertainty in FA ω-bond assignment.

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

(C) Mean +/− SEM C18 SCD indices. P-values are estimated using one-way ANOVA for orthogonal comparisons of L-PTP1b−/− to control for each diet-stimulation condition; * and ^ denote P <.05 and .1.

(D) Western blot of hepatic pT172-AMPK, total AMPK, and β actin in HFD mice. The bar graph represents mean +/− SEM pT172-AMPK to AMPK ratio from pooled basal and insulin-stimulated samples. Significance was calculated using a 2-sided t test. * denotes P < .05.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

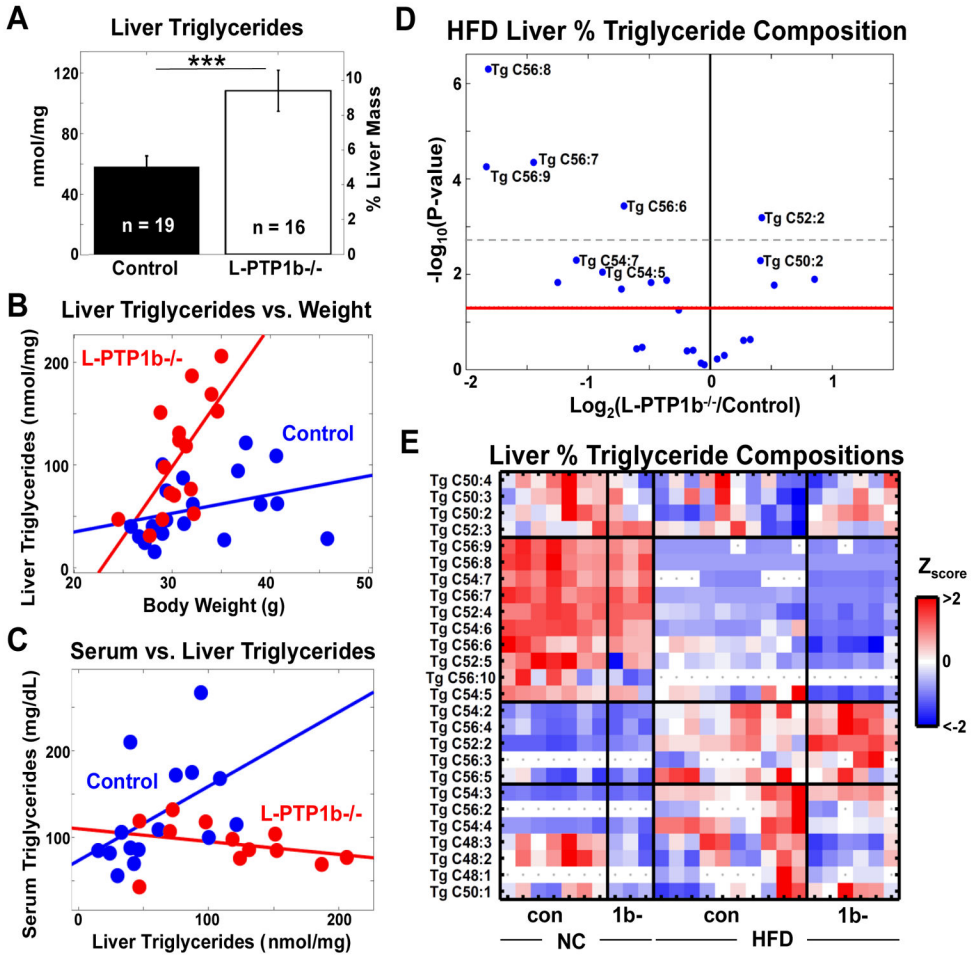

Figure 5. PTP1b-dependent Changes in Triglyceride Metabolism

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

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

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

(D) Volcano plot of % triglyceride compositions on HFD (control, n = 10, L-PTP1b−/−, n = 6). The red line denotes P = .05, while the dotted grey line shows the Bonferroni significance cut off to control for multiple hypotheses at FDR = .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 t tests.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

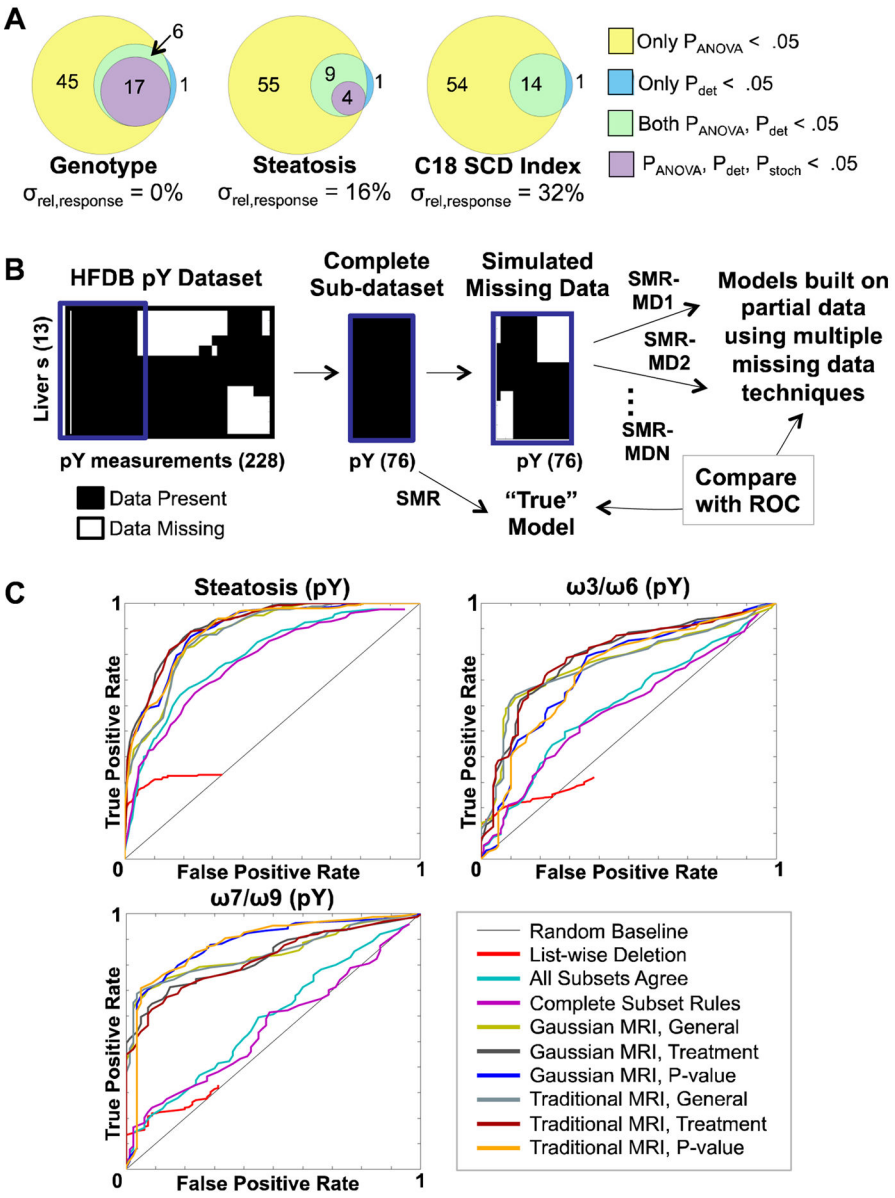

Figure 6. Evaluation of SMR and Missing Data Methods

(A) Sets of phosphosites significant to three models (Genotype, Steatosis, C18 SCD Index). Phosphosite significances were based on their associated multivariate regression coefficients, as determined by SMR ($P_{stoch} < .05$) and DMR ($P_{det} < .05$). Phosphosites whose variation was distinguished from measurement error using ANOVA filtering are included in Venn diagrams ($P_{ANOVA} < .05$).

(B) Schematic for testing missing data methods in the context of multivariate regression.

(C) Performance of strategies for incomplete data.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

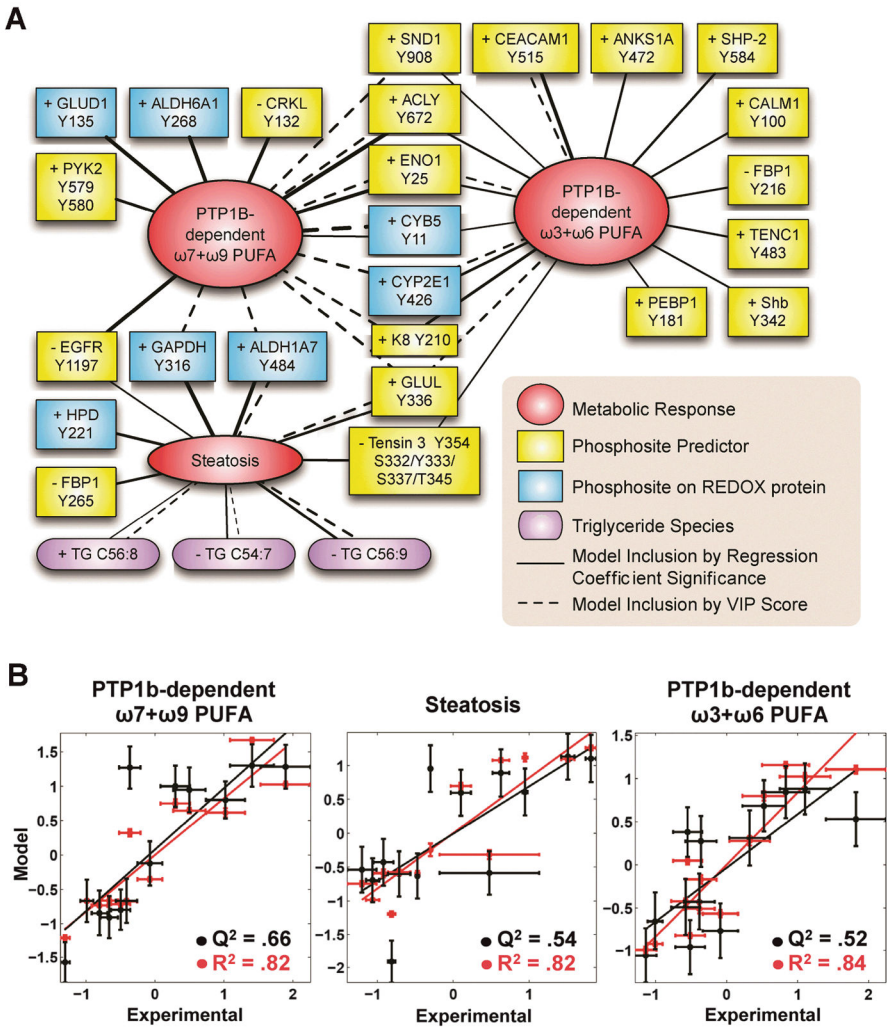

Figure 7. Multivariate Models of Lipidomics Phenotypes

(A) The phosphosite and metabolite predictors that were included in final SMR models are connected to phenotype nodes by edges whose widths reflect the importance of the phosphosite or metabolite to the model. The (+) or (−) denotes the sign of predictors' regression coefficient in the model. Final models were built using two model reduction techniques, as denoted in the legend and described in Methods.

(B) Quantitative prediction (black) and fit (red) for ω7+w9 PTP1b-dependent PUFA, steatosis and ω3+ω6 PTP1b-dependent PUFA models using model reduction by regression coefficient significance (see Methods). Error bars denote experimental and model SEM.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

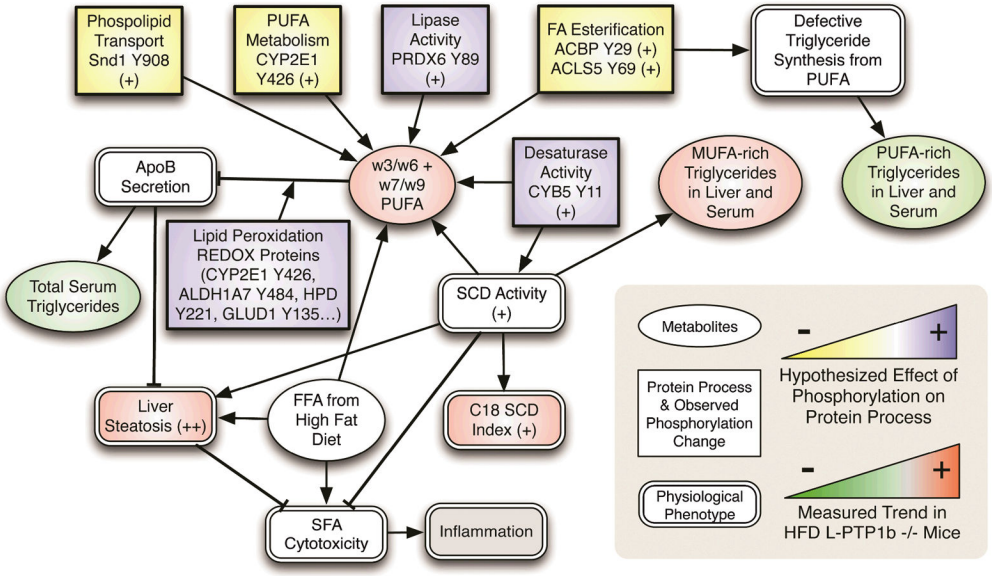

Figure 8. Proposed molecular and physiological model of L-PTP1b−/− lipid metabolism

Proposed interactions among phosphotyrosine network state, lipid metabolism, and physiological phenotypes. Physiological processes in white boxes were not measured in this study but would provide further insight into the phenotypes of L-PTP1b−/− mice on HFD. "(+)" indicates that phosphorylation on a particular site is increased in L-PTP1b−/− mice.