Systems Biochemistry Investigation into Lipid Metabolism and Lipid-Metabolizing Enzymes

By:

Molly T. McDevitt

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN – MADISON

2019

Date of final oral examination: 01/08/2019

This dissertation is approved by the following members of the Final Oral Committee:

David J. Pagliarini, Associate Professor, Biochemistry
Michael R. Sussman, Professor, Biochemistry
Joshua J. Coon, Professor, Chemistry and Biomolecular Chemistry
Alan D. Attie, Professor, Biochemistry
Karl W. Broman, Professor, Biostatistics and Medical Informatics

Systems Biochemistry Investigation into Lipid Metabolism and Lipid-Metabolizing Enzymes

By:

Molly T. McDevitt

Under the supervision of Professor David J. Pagliarini

At the University of Wisconsin – Madison

Abstract

Lipids are important for a vast array of cellular functions including energy storage and transport, membrane structure, protein binding, and cell signaling. Due to the overlap of pathways involved in lipid metabolism, abnormal metabolism of a single lipid class may cause a cascade of dysfunctions across multiple lipid classes and cellular functions. Diseases such as obesity, insulin resistance, and non-alcoholic fatty liver disease (NAFLD) are all associated with abnormal lipid metabolism. Not only have these diseases become more pervasive in the last fifty years, but they have also become some of the leading causes of death in developed countries. Thus, abundance of various lipid classes, and individual lipid species within those classes, must be tightly regulated. However, due to the diversity of lipids' structure and function and incomplete understanding of all proteins that affect lipid metabolism, we are far from understanding the intricate, whole-cell regulation of these important molecules.

Chapter 1 summarizes the structural and functional diversity of lipids and complexity of their regulation, including the overlap of biosynthetic pathways, enzymes involved in lipid metabolism, and consequences of dysregulation. Recent advances in lipidomics such as better instrumentation and automatic identification are also discussed. Finally, two methods to parse out mechanisms of lipid regulation are presented. The first, an unbiased approach, uses a genetically diverse, environmentally controlled sample set to identify genomic loci that regulate the

abundance of specific lipid species. The second is a focused approach and describes and in-depth investigation of a single protein to determine if and how it affects lipid metabolism.

Chapters 2 and 3 present the unbiased approach, utilizing the BXD recombinant inbred line to identify genetic and environmental regulators of liver and plasma lipid levels. We describe the first use of individual lipid species as quantitative, mappable traits, providing a key resource for future studies. We focus on lipid species of the same class that associate differentially with of health and disease, depending on their degrees of unsaturation. Together our results provide a foundation for future lQTL studies and reiterate the importance of exploring the role individual lipid species, rather than classes, have on physiological traits.

Chapter 4 discusses PREPL, a putative peptidase implicated in the human disease hypotonia cystinuria syndrome (HCS). We demonstrate that the long isoform ($PREPL_L$) localizes to the mitochondria and exhibits specific activity against an ester substrate, thereby providing the first evidence of *in vitro* activity for the protein. Finally, we investigate the effect of loss of functional PREPL on mouse and cell culture lipid profiles.

Chapter 5 discusses caveats and future directions for mapping lipid QTLs, including expanding the lipids quantified and the genetic diversity of the study group. Furthermore, we consider the next steps in validating candidate genes and identifying novel genes involved in lipid metabolism from our lQTL resource and briefly discuss the advantages of mapping lQTLs on an organellar rather than cellular level. Finally, we outline future efforts to identify the *in vivo* substrate of PREPL, focusing on identifying structural similarities among lipids specifically altered in the knockout models.

Collectively, this work describes two methods for investigating the regulation of lipid metabolism. The first identifies the genetic and dietary impact on individual lipid species and provides a foundational resource for future lQTL studies. The second looks to identify how a specific enzyme, PREPL, affects lipid metabolism using bioinformatics and *in vitro* biochemistry. The first, unbiased approach provides a method for identifying and prioritizing novel genes

involved in lipid metabolism. However, the second, focused approach, is required for validation of candidate genes and truly characterizing proteins of unknown function. In conjunction, these two approaches are both important to advance our understanding of the intricacies of the regulation of lipid metabolism.

Acknowledgments

First, I would like to thank Professor Dave Pagliarini for his mentorship and patience throughout my graduate career. He is a savvy scientist, a wonderful mentor and amazingly, an even better person. Thank you for providing a lab environment that both strives for excellence and fosters collaboration. I think we can both agree that my journey through graduate school has not been an easy one. Thank you for supporting me through it all.

Second I would like to thank all of the members of the Pagliarini Lab, both past and present, who made the lab a place I could call home. Adam Jochem, you are magician. Thank you for helping me maintain my sanity while homogenizing hundreds of mouse livers. I would like to thank Jon Stefely and Brendan Floyd for their support, advice and guidance. To the best baymate ever, Xiao Guo, and to a close runner up, Jon Tai, thank you for putting up with me day in and day out. I would like to thank Rahul Gupta for showing me that I should have learned to code years ago and commiserating with me when it seemed like the BXD papers would never get done. Thanks to Amy Lin for sharing her Illustrator knowledge. And finally, I would like to thank Jenelle Gierhart-Sutter – I do not know how I, or the lab, ever functioned without you. Your help has been invaluable.

I thank the members of my thesis committee for their guidance through the years, especially Professor Josh Coon. He gave me the opportunity to hone my mass spectrometry skills and to continue to collaborate with his lab even after I left. Thank you to Alex Hebert, who taught me everything I know about proteomics, and Arne Ulbrich and Paul Hutchins, who taught me everything I know about lipidomics. I would also like to thank Mike Westphall for his ability to find a joke in every situation.

I have had the opportunity to collaborate with some outstanding scientists throughout my graduate career. I would especially like to thank members of Johan Auwerx's group who worked together to publish the BXD manuscripts, especially Pooja Jha with whom it was a pleasure to work. Though I dabbled in both proteomics and lipidomics, my GC-MS skills are non-existent.

Thank you to Matt Rush, Brett Paulson, and Thiru Reddy for helping me with all my metabolomics needs.

I have been incredibly lucky and have formed friendships with some truly amazing individuals in Madison – some of whom are also top-notch scientists, all of whom are top-notch people. I would like to thank Laura Bond and Andrew Reidenbach for letting me tag along and never letting me feel left out. There's no one I'd rather talk about cats, "watch" movies, or hang out in my grossest sweatpants with. Thanks to Alicia Richards, Catie Vincent, and Emily Wilkerson who provided many laughs, and perhaps even more cute cat photos, over the years. I thank Amelia Peterson for her friendship and help navigating graduate school in my first year. Our late nights in lab were always a treat. Kyle and Carly Robinson, thank you for welcoming me into your lives – our "sports and crafts" nights are some of my favorite. I thank Jarred Rensvold for being a voice of reason. Being his second favorite in lab is still one of my greatest accomplishments. I thank Danielle Lohman for her steadfast support and willingness to listen to all my crazy stories. I would like to thank Tim Rhoads for making sure I never had to watch a football game alone, remaining calm through everything, and offering sage advice whenever I needed it. I would like to thank Melissa Leinenger, Eric Nielsen, and Terrell Johnson for reminding me that even in the darkest of places, there is light if we only bother to find it. To all the other friends I have made through the years – you have made this experience worth it.

I would especially like to thank Tucker Carrocci for getting into a car with me in August 2011 and continuing to talk to me, despite the pizza incident. You're stuck with me now. Our friendship has meant the world to me. Few people can make grocery shopping and paying bills so much fun. Thank you for being the guinea pig for all my boring stories. Graduate school would not have been the same without out you.

I would also like to thank my previous teachers and mentors, especially Mrs. Blatchford who taught me to think critically and insisted on questioning everything. Her passion for learning

inspires me daily. Thank you to Julie Soukup who gave me my first real lab experience and believed in me, even when I didn't believe in myself.

Finally, I would like to thank my family who has been an indestructible support system since the day I was born. To my aunts, uncles, and cousins – thank you for your letters and texts. I am humbled by your undying love and support. I would especially like to thank my grandpa for teaching me that you don't have to understand someone's dreams to support them and that sometimes you have to stand up for yourself. He also gives stellar, full-proof dating advice. Tonks, you have enriched my life more than I thought possible, even if you can't talk. You make a stellar listener. To my niece Elinor, thank you for showing me how deep one can love. You inspire me to be a better person. I would like to thank David and Nicole for choosing to join our family. We were incomplete without you. To my siblings, thank you for your support and for the motivation to be someone you could look up to. Marty, thank you for being the logic to my emotion and for the endless laughter you supply. Maggie, thank you for showing me that there is always hope, no matter how bad it seems. To my parents, Dave and Michele, I thank you for everything. Words simply cannot express how thankful I am to be your daughter. I could not have been blessed with a more supportive family. Thank you for believing me, thank you for seeing me through the tough times, thank you for being with me to celebrate the goods ones. I dedicate this dissertation to you.

Table of Contents

Chapter 1: Introduction

Molly T. McDevitt

Overview of lipids

A brief history of lipids

As early as 1539, the word “fat” appeared in the Oxford English Dictionary, referring to the “oily concrete substance of which the fat parts of an animal are chiefly composed” (Gidez, 1984). However, the association with what we now call adipose tissue and that part of the animal used in cooking, candle making, etc. far predates even the earliest use of the word. Despite having been aware of the existence of lipids for hundreds of years, it is only now that we are beginning to understand not only the sheer number of lipids in each cell, but also the vital and diverse roles they play in cellular function.

It wasn’t until the 19th century that real advancements in lipid biochemistry began to take place. Michel Chevreul, a French organic chemist and one of the fathers of lipid chemistry, is credited for first identifying and isolating fatty acids (Chevreul, 1813) and demonstrating that fat was made up of a combination of glycerol and fatty acids (Chevreul, 1815). Other work from that century includes the realization that fat, along with protein and carbohydrate, are all necessary for a healthy diet (Rosenfeld, 2003) and the demonstration that fats could by synthesized with one part glycerol to three parts fatty acid. Over time, the term “fat” became associated with a single class of lipids, now known as triglycerides. The term “lipid” is now widely used to describe the ever-expanding list of water-insoluable hydrophobic or amphipathic hydrocarbons (Fahy et al., 2011; Gidez, 1984).

Although all lipids are made up of primarily carbon and hydrogen atoms containing a backbone attached to one or more long hydrocarbon tail, the structural diversity of lipids is immense. Hydrocarbon tails are made up of various repeating units such as fatty acyls, isoprenes,

and polyketides (Figure 1A). Additionally each repeating unit can differ in length (number of carbons) as well as number and positions of double bonds. These repeating hydrocarbon units are then attached to backbone structures such as glycerol and sphingosine (Figure 1B) in an almost endless array of combinations. To date, there are over 43,000 unique lipid structures in the LIPID MAPS structural database, but it is estimated that there could be as many as 100,000 unique lipid species (Fahy et al., 2011; Sud et al., 2007). So while initial studies may have focused on triglycerides and lipids' role in fat/energy storage, it is unsurprising that over time we have discovered their roles in the cell to be nearly as diverse as the lipids' structures themselves.

Lipids are not only an important energy reserve but also serve as a method of energy transport. Moving lipids from the liver to tissues with high energy needs such as the heart and muscle is essential, especially during fasting conditions when glucose needs to be reserved for the brain (Luiken et al., 1999; Neely and Morgan, 1974). Additionally, lipids are largely responsible for the structure of all cell membranes, including the formation of a lipid bilayer facilitated by the amphipathic nature of the membrane lipids and the degree of membrane curvature (Chen and Rand, 1997; Hammond et al., 1984; Kooijman et al., 2005). In 1997 it was discovered that the lipid microenvironment in cell membranes can affect both protein binding and activity (Simons and Ikonen, 1997). Beyond their roles in storage and structure, many lipids such as fatty acids can act as signaling molecules to alter cell function in response to various stimuli (Hannun, 1994; Lee et al., 2001; Nishizuka, 1992)

Classification of lipids

Lipids are divided into eight categories: fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, saccharolipids, polyketides, sterol lipids and prenol lipids (Figure 1C). Of these categories, sterol and prenol lipids are built from isoprene units while the rest are made by the condensation of ketoacyl groups (Fahy et al., 2011). Due to the limited scope of our study, only those groups which I worked with will be further explored here – fatty acyls, glycerolipids, and

glycerophospholipids. However, it is important to note that list of lipids identified and quantified in this thesis is far from exhaustive.

While free fatty acids themselves are only found at low levels in organisms, they are important building blocks for almost all of the eight lipid categories (Figure 1C). Consisting of a carboxylic acid and a long hydrophobic hydrocarbon chain, fatty acids play many roles in the cell. Fatty acids are a large source of cellular energy, serving as the substrates for beta-oxidation in the mitochondria. Additionally, they play roles in cell signaling, for example activating a number of G protein-coupled receptors (Briscoe et al., 2003; Forman et al., 1997; Tolhurst et al., 2012), as well as membrane stabilization (Leekumjorn et al., 2009).

Perhaps the most well studied lipids are glycerolipids. Glycerolipids consist of a glycerol backbone attached to one, two, or three fatty acid tails (Figure 1C). Known as mono-, di-, and triacylglycerols (triglycerides, TAG), these lipids are the main source of energy storage in the cell. Additionally, because high levels of free fatty acids is detrimental, TAG serves as a way to transport these energy rich molecules to energy-requiring tissues that do not have the capacity to retain their own fat reserves (Haemmerle et al., 2006). Mono- and diacylglycerols (DAG), are also important precursors of TAG and can act as signaling molecules in their own right (Berridge, 1984; Werner et al., 1992).

Glycerophospholipids consist of lipids made up of fatty acid chains and a glycerol backbone with various polar headgroups at the *sn*-3 position (Figure 1C). In general, glycerophospholipids fall into six classes – phosphatidic acid (PA), phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidyl serine (PS), phosphatidyl inositol (PI), and cardiolipin (CL), a unique phospholipid consisting of three glycerol backbones and four fatty acid tails (Figure 1C). Due to their amphipathic nature, glycerophospholipids naturally form lipid bilayers, and therefore constitute the majority of both the plasma and organellar membranes. Due to its cylindrical nature (i.e., the head group and fatty acyl tails are approximately the same width), PC is the major component of cellular membranes. Others such as PI, PE and CL form conical

shapes, with either the head group or fatty acyl tails being larger than the other (Figure 1D). Lipid rafts (microenvironments having increased levels of specific lipids) made from PE, PI, and CL induce membrane curvature and bind and activate proteins at the membrane surface (Ramamurthi and Losick, 2009; Renner and Weibel, 2011). Moreover, while PA and PI can act as second messengers themselves, their role as precursors to various signaling molecules is perhaps more important (Berridge, 1984).

Lipid Metabolism

Although fat is an important part of a healthy diet, most of the lipids we ingest are triglycerides and cholesterol. However, as outlined above, lipids are required for much more than energy storage. Thus, lipid metabolism, both biosynthesis and breakdown, is essential for proper cellular function. Importantly, despite the vast diversity between individual lipid species, there is considerable overlap between the metabolism of different lipid classes (Figure 1E). This is perhaps unsurprising given the fact that lipids, despite their structural and functional diversity, are made up of a limited set of “building blocks”. Here I will highlight two of the many examples – phosphatidic acid (PA) and diacylglycerol (DAG) (Figure 1E).

Diacylglycerol is a simple molecule with complex functions. When concentrated in small areas of cellular membranes it induces membrane curvature and can improve membrane-protein interaction by exposing the hydrophobic parts of nearby lipids (Goni and Alonso, 1999; Shemesh et al., 2003). Additionally, it can act as a second messenger, playing an integral part in lipid-mediated signaling. Finally, it is an important hub of lipid metabolism and acts as a precursor for multiple lipids (Coleman and Lee, 2004; Kennedy et al., 1956) (Figure 1E). It serves as the link between phospholipid and TAG biosynthesis as it is utilized for TAG, PC, PE, and PS biosynthesis. In fact, dysregulation of DAG metabolism can cause deleterious effects on phospholipid biosynthesis (Verrier et al., 2004). Moreover, when ceramide is converted to sphingomyelin, DAG forms as a byproduct. Interestingly, both ceramide and DAG act as second messengers, but with opposing effects (Bourbon et al., 2000; Gault et al., 2010; Villani et al.,

2008). Thus, it is possible that the delicate balance between ceramides and DAGs plays an important role in cell fate.

PA is a precursor for glycerophospholipids and TAGs, thus serving as a link between lipids as energy reserves, membrane components, and signaling molecules (Athenstaedt and Daum, 1999). Due to its role in CL and PG formation, PA can also affect membrane curvature, albeit indirectly. Beyond its role as a lipid precursor, phosphatidic acid also acts as a second messenger involved in the activation of cellular kinases and phospholipases (Fang et al., 2001; Moritz et al., 1992). Additionally, in yeast the protonation state of the phosphate head group affects protein binding. In response to glucose starvation, pH decreases, disrupting the interaction between PA and the phospholipid metabolism repressor Opi1. Opi1 relocates to the nucleus, inhibiting phospholipid metabolism, thus making PA an important link between membrane biogenesis and metabolism (Young et al., 2010).

Enzymes involved in lipid metabolism

In order to synthesize, breakdown, remodel, and sense the vast array of lipids, a cell must employ a large number of enzymes to carry out these processes. Interestingly, many of these proteins are members of the serine hydrolase superfamily. As one of the largest enzyme classes, serine hydrolases make up nearly 1% of the known human proteome. These ~200 enzymes are characterized by a nucleophilic serine in the active site that is used for substrate hydrolysis, the majority of which contain an α/β hydrolase fold (Ollis et al., 1992). It is unsurprising then that this huge superfamily includes enzymes that cleave esters, amides, or thioester bonds in small metabolites, lipids, peptides, or proteins (Simon and Cravatt, 2010).

In these enzymes, serine is part of a catalytic dyad (Ser-Lys or Ser-Asp) or triad (Ser-His-Asp or Ser-Ser-Lys) that creates an environment in which the serine can perform a nucleophilic attack on the substrate (Dodson and Wlodawer, 1998; Patricelli et al., 1999). Catalysis occurs via

activation of the nucleophilic serine which then attacks the electrophilic substrate, forming a covalent acyl-enzyme intermediate. The product is released and the enzyme is regenerated following hydrolysis catalyzed by a water molecule. Lipases, phospholipases, acyltransferases, esterases, and hydratases are all subclasses of the serine hydrolase superfamily and all play important roles in lipid metabolism. These, as well as transport and binding proteins, regulate lipid abundance in the cell.

Considerable effort has been made to identify features of these enzymes that indicate substrate specificity. However due to their similarity and promiscuity (Scaloni et al., 1994; Stafforini et al., 1987; Wang et al., 2006) it is often difficult to determine the specific substrates based on primary sequence alone (Derewenda and Derewenda, 1991; Karlsson et al., 1997). For example, esterases and lipases both cleave ester bonds, but despite often exhibiting high sequence similarity, their ester-containing substrates are radically different. Analysis of three-dimensional structure as well as electrostatics have also been used to differentiate the two, but have met with limited success. Lipases tend to have more nonpolar residues on the surface, but evidence is not strong enough to determine the substrate for novel hydrolases. Moreover, the active site is often buried under secondary structural elements that may or may not need to change formation prior to substrate binding (Cygler et al., 1993). Though it is likely that these structural elements are important for substrate specificity, their exact roles have, by in large, been difficult to ascertain.

Additionally, there are peptidases, such as APEH, that also exhibit esterase activity. Chapter 4 will further discuss one example, PREPL – a putative peptidase which lacks peptidase activity but exhibits low levels of esterase activity. However, to fully understand lipid metabolism it will be important to identify *in vivo* substrates of all enzymes that act on various lipids. Understanding the specificity of these proteins may also give insight into why individual lipids in each class are not regulated in the same manner.

.

Dysfunctional lipid metabolism and its relationship to disease

It did not take long to link lipid buildup to disease. In fact, as early as 1913 Nikolai N. Anitschkow noticed that rabbits fed a cholesterol-rich diet developed severe atherosclerosis, a hardening and narrowing of the arteries (Anitschkow and Chalatow, 1913; Buja, 2014). Overall, the main disorders of lipid metabolism involve triglycerides and cholesterol, including hypercholesterolemia and hypertriglyceridemia. Elevated levels of cholesterol and triglycerides (often caused by increased fatty acids) are associated with atherosclerosis which can eventually lead to cardiovascular disease such as myocardial infarctions and strokes. Additionally, hypertriglyceridemia causes a host of other metabolic disorders, including acute pancreatitis and non-alcoholic fatty liver disease (NAFLD) (Granger and Remick, 2005; Osono et al., 1995). Made more severe by the fat and carbohydrate-rich Western diet, cardiovascular disease has quickly become one of the leading causes of death. Regulation of glycerolipid metabolism is essential to cellular homeostasis with disruptions of proper metabolism leading to obesity, type II diabetes, and even kidney failure (Li et al., 2017; Weijers, 2012). However, diseases associated with aberrant lipid metabolism are not isolated to cholesterol and glycerolipid metabolism.

Besides being one of the underlying factors contributing to hypertriglyceridemia dysregulated fatty acid metabolism can cause a number of diseases. For example, the inability to oxidize fatty acids prohibits individuals from utilizing their fat stores for energy when glucose has run out. Although it can be treated by diet modification, left unchecked it leads to a buildup of fatty acids in the liver and other organs and energy deficiency via decreased levels of reducing equivalents made by fatty acid oxidation and the TCA cycle and ketone bodies (Vishwanath, 2016).

Dysregulation of various glycerophospholipids also leads to a wide variety of diseases. For example, in a controlled screen, levels of three specific PC species were significantly decreased in Alzheimer's patients (Whiley et al., 2014). In fact, phospholipase D and A2 enzymes, which catalyze the breakdown of PC, have been directly associated with Alzheimer's disease

(Dennis et al., 2011; Li and Vance, 2008; Selvy et al., 2011). Additionally, it seems that the PC/PE ratio is important for metabolic homeostasis, as it can affect everything from energy production in the mitochondria to metabolic disorders such as insulin resistance, obesity, and NAFLD (Arendt et al., 2013). Furthermore, cardiolipin (CL), an essential part of the inner mitochondrial membrane, is an important part of optimal mitochondrial function/oxidative phosphorylation. Alterations in CL levels are associated with multiple diseases, including aging and heart failure (Mejia et al., 2014). Moreover, dysregulation of the remodeling of CL fatty acid tails is associated with such diseases as Barth Syndrome – a disease characterized by cardiac and skeletal myopathies (Lou et al., 2018). Thus it seems that both the overall abundance as well as proper remodeling of CL are imperative for proper cellular function.

Disruption of the biosynthesis of the prenol lipid CoQ is causes a variety of mitochondrial disorders. Clinical manifestation ranges from fatal multisystem disorders to isolated encephalopathy and nephropathy and can show itself anywhere from birth onward (Acosta et al., 2016). As CoQ is an essential part of the electron transport chain, the mechanism for pathogenesis seems clear at first glance – a decrease in cellular energy production. However, due to the extreme differences in disease manifestation, it is likely that other aspects of CoQ function are also disease-causing.

Inhibition of sphingolipid synthesis increases insulin sensitivity in mice (Bijl et al., 2009; van Eijk et al., 2009). Additionally, increased sphingolipids (such as sphingomyelin) and DAGs have been associated with Alzheimer's disease (AD) (He et al., 2010; Wood et al., 2015). Other neuronal diseases such as Parkinson's disease (PD) and bipolar disorder are also associated with alterations in lipid profiles (Cheng et al., 2011). Finally, altered lipid metabolism is common in a number of cancers including some prostrate and breast cancers (Beloribi-Djefaflia et al., 2016).

The complexity of lipid metabolism

As outlined above, lipids play a diverse set if essential roles in maintaining cellular homeostasis and dysfunctions in one or more of these pathways can have deleterious effects. Although alterations in lipid abundance has been implicated in numerous diseases, the exact mechanism of pathogenesis is rarely, if ever, fully understood. There are a number of reasons for this lack of understanding. First, we do not fully understand the relationships between biosynthesis of different lipid classes, and perhaps even more importantly the balance within classes. To date, lipid classes are often measured *en masse*, assuming that all lipids within a class are regulated the same way, which is often not the case (Jha et al., 2018a; Jha et al., 2018b). Second, lipids play such a central role in metabolism that we have not fully grasped all of the implications alterations in lipid abundance can cause. It is difficult to ascertain the mechanism of pathogenesis when one does not even understand all of the effects. Finally, lipid metabolism is heavily influenced by both genetics and environmental factors and it has been difficult to piece apart the effects of each. Thus, it is clear that we have barely begun to fully understand the intricacies, crosstalk, and balance between these pathways. However, the explosion in lipidomics technology has made it possible to identify and quantify single lipid species, giving us a better understanding of how single species are regulated and related to each other.

Advancements in lipidomics

Advances in mass spectrometry

Unlike genomics, transcriptomics, and proteomics, lipidomics platforms are not well established yet. In fact, the term lipidomics has emerged quite recently, in large part due to the advancements in mass spectrometry (MS). Previously, it has been customary to quantify entire lipid classes due to the complexity of measuring a hundreds or thousands of different species whose abundance spans five to ten orders of magnitude (Yang et al., 2009). Fueled by advances in instrumentation, lipidomics via MS has become routine. The structural information about head

groups and acyl chain lengths obtained from tandem MS experiments (MS/MS), combined with the highly accurate intact masses obtained from high resolution, high mass accuracy analzyers such as the orbitrap (Ogiso et al., 2008; Schuhmann et al., 2011) allows for the unambiguous identification of hundreds of lipid species from a single complex mixture. In the past five years alone, we have gone from measuring a hundred lipid species to nearly one thousand.

Different MS-based techniques have been utilized in lipidomics. One of the biggest differences is whether or not separation via liquid chromatography occurs prior to MS analysis (LC-MS/MS). Both have been used successfully. Direct-injection methods are accurate, reproducible, sensitive, and quick. In 2003, Han and Gross first utilized electrospray ionization (ESI) MS in a direct-injection lipidomics technique. Product ion, precursor ion, and neutral loss scans along with selected reaction monitoring (SRM) are all used to identify individual lipid species. While much quicker, direct-infusion methods often have some difficulty in distinguishing different lipid molecules (especially isomers) and depending on the number of species being identified can require more sample than their LC-based counterparts (Li et al., 2014).

LC-MS has quickly become the most widely used method for analysis of biological samples samples. It is widely used for its high resolution and reproducibility. Additionally, due to its setup, samples are largely isolated from the environment, preventing lipid degradation. Typically, LC-MS is used for either targeted or discovery-based analysis. In targeted methods, users develop methods focused on the detection of specific lipid species. Due to the nature of the method, it is highly sensitive and able to detect low level species more accurately and reproducibly. This technique has been used to target everything from specific lipid classes to various lipids of interest (Jha et al., 2018a; Stefely et al., 2016). Unlike targeted methods, discovery-based methods seek to profile all lipids in a given sample. Although some sensitivity can be lost, discovery-based methods are best for unbiased studies, and have been used to define the lipidome of everything from yeast to human plasma (Ejsing et al., 2009; Quehenberger and Dennis, 2011)

Optimization of lipid extraction methods

Due to the complexity of biological samples, it is almost always necessary to extract lipids prior to analysis to remove interference that comes from proteins and other small molecules. Regardless of the MS technique used, its success relies largely on the efficiency of lipid extraction from samples. The most commonly used method was developed by Folch in 1957 and involves using chloroform/methanol (2:1, v/v) to extract lipids (Folch et al., 1957). Two years later, Bligh and Dyer improved the method by using water to further induce phase separation leading to increased extraction efficiency and lipid degradation prevention (Bligh and Dyer, 1959).

Recently, extraction solvents besides chloroform and methanol have been adopted. An extraction solution comprised of butanol and methanol (BUME) has been shown to have extraction efficiencies comparable or better than the Folch method, while also being much quicker and high throughput than the older method (Lofgren et al., 2012). A method using methyl tert-butyl ether (MTBE), methanol, and water to extract lipids was also recently developed. Due to MTBE's low density the organic layer is on top, making for easier and cleaner recovery of lipids in the organic layer. Similar to the BUME method, recovery is comparable to both the Folch and Bligh and Dyer methods (Matyash et al., 2008). Additionally, this method can be used to extract both lipids and non-polar metabolites simultaneously, a huge advantage when sample is limited (Chen et al., 2013).

Improved lipid identification and quantification via bioinformatics

Due to the structural diversity of lipids, even within a single class, it is difficult to create models for all possible lipid species. For example, a single lipid head group is attached to one to four fatty acid tails, each of which can have varying number of carbons and points of unsaturation. Furthermore, these double bonds can vary in type and location, not to mention the other modifications that can occur. The combinations are nearly endless, making it almost impossible to not only create a library containing all possible structures but to even begin to search such a library in a timely and computationally efficient manner. Important steps, such as creating lipid

library databases such as LIPID MAPS, have been taken to narrow the search space (Sud et al., 2007). As of now, more than 40,000 unique lipid species are listed in the LIPID MAPS database, a number that will likely continue to grow.

In the past, automated lipid identification was often difficult due to the complexity of the sample. Co-eluting peaks and isomers made it hard for software to identify and properly quantify individual lipid species. However, manual peak identification and quantification is a time-intensive process and limited the complexity of samples that could be identified. However, as lipidomics has become more mainstream, efforts have been made to streamline the process. Metabolite databases such as LIPID MAPS, Lipid Bank, LipidHOME, and LipidBlast include intact masses and *in silico* fragmentation to compare to actual data acquired. By automating this search, as well as subsequent peak quantification and removal of ionization artifacts, hundreds of lipids can be measured from a single complex solution (Hutchins et al., 2018). Software is available both freely and commercially, allowing labs of all sizes to profile large lipid datasets in a high-throughput manner. However, as lipid databases continue to grow these methods will need to be continuously tweaked.

Additionally, future advancements in lipid bioinformatics should include rigorous statistical analysis such as that seen in genomics, transcriptomics, proteomics, and metabolomics (Niemela et al., 2009; Pauling and Klipp, 2016). Although statistical analysis alone will help identify specific lipids altered in certain metabolic sites, it is also important to examine the data in a biochemical context. LIPID MAPS, KEGG, and PubChem have recently been linked to aid in the visualization of lipids in various metabolic pathways. However, it is impractical, if not impossible, to represent all possible lipids and all possible pathways in these databases. One way of addressing the problem involves identifying lipids that are co-regulated and reconstructing biochemical pathways (Yetukuri et al., 2007). Finally, methods of integrating lipidome profiles with the spatial and temporal models of lipid systems may reveal the roles lipids play in cellular maintenance.

Quantitative trait loci (QTL) mapping: a method for linking genotype and phenotype

One of the difficulties in fully understanding the roles lipids play in cellular metabolism is the complex regulation of lipid species and the quantitative nature of lipid abundance. Unlike discontinuous traits which fall into discrete groups and are often regulated by a single gene, quantitative traits arise from interactions between multiple genetic and environmental factors and are often distributed in a Gaussian manner. Before we are able to predict disease risks and identify personalized treatments, we must first identify regions of the genome that make individuals more susceptible to disease or sensitive to treatment. While complex interactions make identifying links between genotype and phenotype more challenging, the large effects of specific genetic loci on some phenotypes have made it possible to establish links between genotype and these quantitative traits using QTL mapping.

QTL mapping is a statistical method that utilizes genetic variation of a population to link complex phenotypes to specific regions of the chromosome, called loci. Successful QTL mapping requires a measurable trait that varies across a genetically diverse sample set. In the early twentieth century, it was observed that some characteristics were inherited in simple Mendelian ratios. It was determined that these characteristics, such as size inheritance in plants, were dependent on multiple genetic factors (Sax, 1923). However, limited polymorphic markers made QTL mapping difficult, until the advancements in sequencing technology developed in 1980s and 1990s (Davey et al., 2011; Mackay et al., 2009). As advances in fast and cost-effective genotyping became commonplace, new statistical methods of varying complexity for QTL mapping have also been developed.

Initially, a relatively simple statistical analysis called marker regression was used for QTL mapping. At each marker, samples are split into groups based on genotype and average phenotypes are compared. It is simple and covariates are easily added. However, missing genotypes cannot be inferred and therefore, individuals missing genotypes must be thrown out

and if markers are far apart, it is difficult to ascertain an accurate location of a QTL (Broman, 2001).

To overcome some of these weaknesses Lander and Botstein introduced interval mapping (Lander and Botstein, 1989). Like marker regression, a single QTL is assumed and each location of the genome is tested separately for presence of a QTL. Unlike marker regression, however, QTL positions between markers can be inferred. This eliminates the requirement for complete marker data reducing the amount of data thrown out for each experiment (Lincoln and Lander, 1992). Since then, various forms of interval mapping have been developed, each using a slightly different method for dealing with missing genetic information (Arends et al., 2010; Bobb et al., 2011; Haley and Knott, 1992).

Logarithm of odds (LOD) scores are used to estimate the likelihood that a genetic linkage is present between a genetic marker/locus and the measured trait of interest. Simply put, a LOD score is the $\log_{10}$ likelihood ratio comparing the alternative hypothesis that there is a QTL at the marker to the null hypothesis that there is not a QTL in the entire genome. Although it is true that the higher LOD scores serve as greater evidence for the presence of a QTL in that region, it should be noted that for analyses in which the entire genome is being tested, it is imperative to correct for multiple hypothesis testing, often via permutation tests (Churchill and Doerge, 1994) QTL mapping has already successfully identified genes contributing to a large range of phenotypes, including glucose tolerance and oxygen consumption rates (Andreux et al., 2012), as well as susceptibility to diseases such as diabetes and alcoholism (McClearn et al., 1997; Saxena et al., 2007). However, it should be noted that identifying a QTL is only the first step to identifying a causative gene. Loci can span large chromosomal regions, containing hundreds of genes. Methods for prioritizing genes (Jha et al., 2018a) and rigorous follow-up work is required to identify even a single causative gene.

Initially, QTL mapping was limited to gross phenotypes such as body weight, body size, hair color, etc. Clinical traits such as glucose tolerance and respiration, as well as susceptibility

to diseases such as type 2 diabetes could also be reproducibly measured, and thus lent themselves well for QTL mapping. As microarrays, and now RNA sequencings (RNA-Seq), became more affordable a whole new level of QTLs could be mapped – transcripts. Expression QTLs (eQTLs) have the advantage that the gene location is often known, often making it easier to pinpoint causative genes. In the past two years, mass spectrometry-based methods have made it possible to also identify protein QTLs (pQTLs) and lipid QTLs (lQTLs) (Chick et al., 2016; Jha et al., 2018a; Jha et al., 2018b; Williams et al., 2016). Utilizing all of these method on the same individuals creates a multilayered portrait of what is going on at both the whole-body and molecular level and can provide novel insight into complex systems.

The BXD recombinant inbred lines

Theoretically, successful QTL mapping relies only on having a measurable trait that changes across a genetically diverse population. However, by the very nature of quantitative traits, it is likely that genetics as well as environment play crucial roles in trait manifestation. So although the human population is incredibly diverse, it is nearly impossible to control all environmental factors, often complicating the ability to identify causative regions of the genome. Instead, model organisms such as mice and various plants are often used as it is easier to control breeding and environmental factors.

Derived from careful manipulation of mating between the progeny of an initial cross between two unique inbred strains, recombinant inbred lines (RILs) have been particularly useful in QTL mapping studies. While backcrosses and intercrosses create genetic diversity, one must genotype each individual and deal with complications of heterozygosity (Broman, 2005). Although time and cost intensive, creation of RILs results in a genetically diverse but renewable population. Only a single individual from each RIL must be genotyped and the availability of biological replicates enables statistical analysis, which can be used to eliminate environmental and technical error. Perhaps the most important long-term advantage of RILs is that data is cumulative. Each

additional study adds to the growing database of knowledge, revealing a more complete picture of the links between genetic and environmental factors and various phenotypes.

The BXD cohort is the one of the largest and most studied murine RIL panels. Two common inbred laboratory strains, C57BL/6J (B6) and DBA/2J (D2), were bred to create ~160 new genetically unique inbred lines. These two founders' strains were used because they are both well-studied and differ in a number of ways, including susceptibility to type II diabetes, atherosclerosis and alcohol. As the number of studies done on the BXD cohort increases, we have been able leverage previous results, especially with the addition of eQTLs, to prioritize and identify causative genes (Jha et al., 2018a; Williams et al., 2016).

Significance of lipid QTL mapping

As lipid identification continues to improve and expand, QTL mapping will be a powerful tool in understanding their complex regulation. It has become clear that not only are lipid classes regulated differently, but individual species within those classes as well. If we are ever to truly understand the intricacies of lipid metabolism, we must understand how individual lipid species, as well as classes, are affected by various genetic and environmental factors. Due to the quantitative nature of lipid abundance, QTL mapping offers the most promising path to success. Not only will these studies give further insight into lipids' role in health and disease, but they may help identify new proteins involved in lipid metabolic regulation. Morever, when combined with other eQTL and pQTL data, these multilayered analyses may reveal the underlying mechanisms by which various forms of dysregulated lipid metabolism cause disease.

Figures

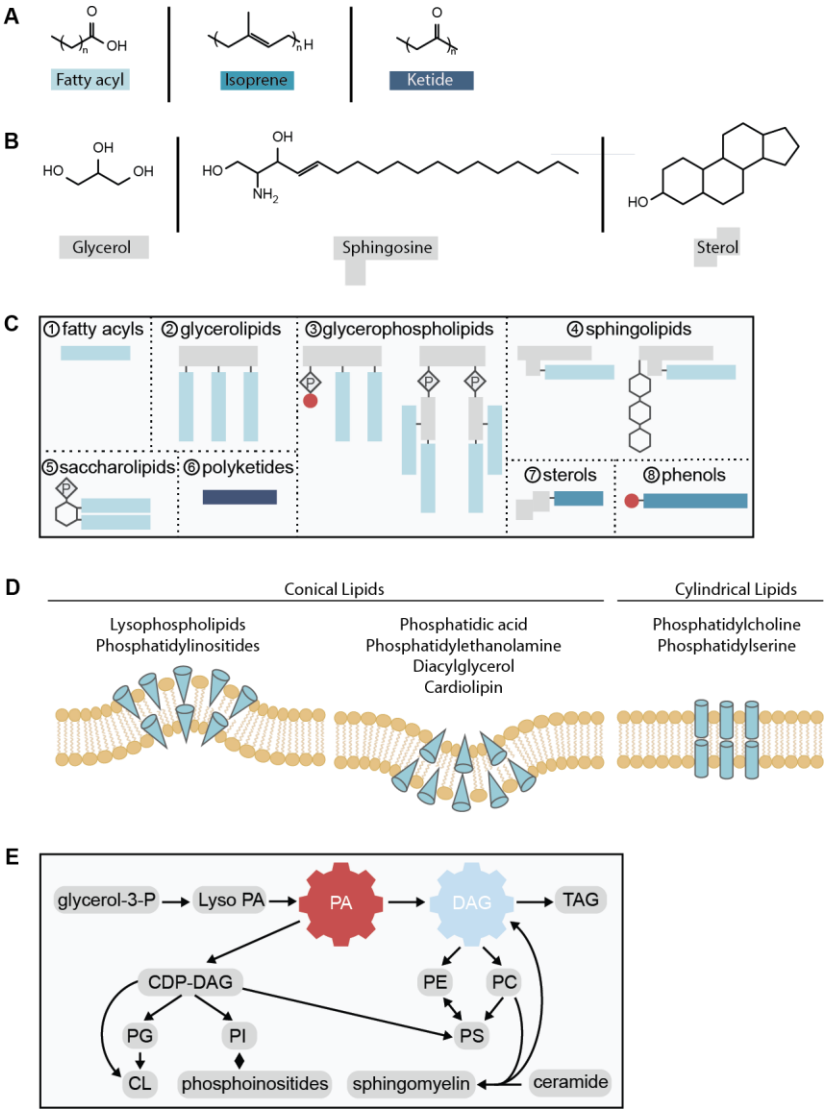


Figure 1. Building blocks of lipid biosynthesis and its complex regulation.

(A) Three building blocks of the long hydrocarbon portion of lipid classes depicted in (C); fatty acyls (light blue), isoprene units (turquoise), and ketides (dark blue).

(B) Examples of common backbone structures of lipid classes depicted in (C); glycerol, sphingosine, and sterol. Boxes around names indicate the way in which that backbone is shown in (C).

(C) Depiction of example lipids representing each of the eight categories based on the LIPID MAPS classification system. Blue shaded bars represent the respective repeated building blocks

in (A) while gray boxes represent backbones shown in (B). Red circles represent any of a number of head groups and  represents a phosphate group.

(E) Schematic of general shape of membrane lipids (cones and cylinders), and how the conical shape of the lipids listed help induce membrane curvature.

(D) Examples of the overlap of biosynthesis and complex regulation of a subset of lipid classes. Arrows indicate biosynthesis pathways. Glycerol 3-phosphate (glycerol-3-P); Lyso PA (lysophosphatidic acid); PA (phosphatidic acid); DAG (diacylglycerol); TAG (triacylglycerol); CDP-DAG (cytidine diphosphate diacylglycerol); PG (phosphatidylglycerol); CL (cardiolipin); PI (phosphatidylinositol); PE (phosphatidylethanolamine); PC (phosphatidylcholine); PS (phosphatidylserine).

References

Acosta, M.J., Vazquez Fonseca, L., Desbats, M.A., Cerqua, C., Zordan, R., Trevisson, E., and Salviati, L. (2016). Coenzyme Q biosynthesis in health and disease. Biochimica et biophysica acta *1857*, 1079-1085.

Andreux, P.A., Williams, E.G., Koutnikova, H., Houtkooper, R.H., Champy, M.F., Henry, H., Schoonjans, K., Williams, R.W., and Auwerx, J. (2012). Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. Cell *150*, 1287-1299.

Anitschkow , N., and Chalatow, S. (1913). Über experimentelle Cholesterinsteatose und ihre Bedeutung für die Entstehung einiger pathologischer Prozesse. Centralbl Allg Pathol Pathol Anatomie *14*, 1-9.

Arends, D., Prins, P., Jansen, R.C., and Broman, K.W. (2010). R/qtl: high-throughput multiple QTL mapping. Bioinformatics (Oxford, England) *26*, 2990-2992.

Arendt, B.M., Ma, D.W., Simons, B., Noureldin, S.A., Therapondos, G., Guindi, M., Sherman, M., and Allard, J.P. (2013). Nonalcoholic fatty liver disease is associated with lower hepatic and erythrocyte ratios of phosphatidylcholine to phosphatidylethanolamine. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme *38*, 334-340.

Athenstaedt, K., and Daum, G. (1999). Phosphatidic acid, a key intermediate in lipid metabolism. European journal of biochemistry *266*, 1-16.

Beloribi-Djefaflia, S., Vasseur, S., and Guillaumond, F. (2016). Lipid metabolic reprogramming in cancer cells. Oncogenesis *5*, e189.

Berridge, M.J. (1984). Inositol trisphosphate and diacylglycerol as second messengers. The Biochemical journal *220*, 345-360.

Bijl, N., Sokolovic, M., Vrins, C., Langeveld, M., Moerland, P.D., Ottenhoff, R., van Roomen, C.P., Claessen, N., Boot, R.G., Aten, J., et al. (2009). Modulation of glycosphingolipid metabolism significantly improves hepatic insulin sensitivity and reverses hepatic steatosis in mice. Hepatology (Baltimore, Md.) *50*, 1431-1441.

Bligh, E.G., and Dyer, W.J. (1959). A rapid method of total lipid extraction and purification. Canadian journal of biochemistry and physiology *37*, 911-917.

Bobb, J.F., Scharfstein, D.O., Daniels, M.J., Collins, F.S., and Kelada, S. (2011). Multiple imputation of missing phenotype data for QTL mapping. Statistical applications in genetics and molecular biology *10*, Article 29.

Bourbon, N.A., Yun, J., and Kester, M. (2000). Ceramide directly activates protein kinase C zeta to regulate a stress-activated protein kinase signaling complex. The Journal of biological chemistry *275*, 35617-35623.

Briscoe, C.P., Tadayyon, M., Andrews, J.L., Benson, W.G., Chambers, J.K., Eilert, M.M., Ellis, C., Elshourbagy, N.A., Goetz, A.S., Minnick, D.T., et al. (2003). The orphan G protein-coupled

receptor GPR40 is activated by medium and long chain fatty acids. The Journal of biological chemistry *278*, 11303-11311.

Broman, K.W. (2001). Review of statistical methods for QTL mapping in experimental crosses. Lab animal *30*, 44-52.

Broman, K.W. (2005). The genomes of recombinant inbred lines. Genetics *169*, 1133-1146.

Buja, L.M. (2014). Nikolai N. Anitschkow and the lipid hypothesis of atherosclerosis. Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology *23*, 183-184.

Chen, S., Hoene, M., Li, J., Li, Y., Zhao, X., Haring, H.U., Schleicher, E.D., Weigert, C., Xu, G., and Lehmann, R. (2013). Simultaneous extraction of metabolome and lipidome with methyl tert-butyl ether from a single small tissue sample for ultra-high performance liquid chromatography/mass spectrometry. Journal of chromatography. A *1298*, 9-16.

Chen, Z., and Rand, R.P. (1997). The influence of cholesterol on phospholipid membrane curvature and bending elasticity. Biophysical journal *73*, 267-276.

Cheng, D., Jenner, A.M., Shui, G., Cheong, W.F., Mitchell, T.W., Nealon, J.R., Kim, W.S., McCann, H., Wenk, M.R., Halliday, G.M., et al. (2011). Lipid pathway alterations in Parkinson's disease primary visual cortex. PloS one *6*, e17299.

Chevreul, M.E. (1813). Recherches chimiques sur plusieurs corps gras, et particulièrement sur leurs combinaisons avec les alcalis. Mémoire 1 - Sur une substance nouvelle obtenue du savon de graisse et de potasse. Annales de Chimie *88*, 225-261.

Chevreul, M.E. (1815). Examen chimique du savon des graisses de porc et de potasse. Annales de Chimie *94*, 80-107.

Chick, J.M., Munger, S.C., Simecek, P., Huttlin, E.L., Choi, K., Gatti, D.M., Raghupathy, N., Svenson, K.L., Churchill, G.A., and Gygi, S.P. (2016). Defining the consequences of genetic variation on a proteome-wide scale. Nature *534*, 500-505.

Churchill, G.A., and Doerge, R.W. (1994). Empirical threshold values for quantitative trait mapping. Genetics *138*, 963-971.

Coleman, R.A., and Lee, D.P. (2004). Enzymes of triacylglycerol synthesis and their regulation. Progress in lipid research *43*, 134-176.

Cygler, M., Schrag, J.D., Sussman, J.L., Harel, M., Silman, I., Gentry, M.K., and Doctor, B.P. (1993). Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. Protein science : a publication of the Protein Society *2*, 366-382.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature reviews. Genetics *12*, 499-510.

Dennis, E.A., Cao, J., Hsu, Y.H., Magrioti, V., and Kokotos, G. (2011). Phospholipase A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. Chemical reviews *111*, 6130-6185.

Derewenda, Z.S., and Derewenda, U. (1991). Relationships among serine hydrolases: evidence for a common structural motif in triacylglyceride lipases and esterases. Biochemistry and cell biology = Biochimie et biologie cellulaire *69*, 842-851.

Dodson, G., and Wlodawer, A. (1998). Catalytic triads and their relatives. Trends in biochemical sciences *23*, 347-352.

Ejsing, C.S., Sampaio, J.L., Surendranath, V., Duchoslav, E., Ekroos, K., Klemm, R.W., Simons, K., and Shevchenko, A. (2009). Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America *106*, 2136-2141.

Fahy, E., Cotter, D., Sud, M., and Subramaniam, S. (2011). Lipid classification, structures and tools. Biochimica et biophysica acta *1811*, 637-647.

Fang, Y., Vilella-Bach, M., Bachmann, R., Flanigan, A., and Chen, J. (2001). Phosphatidic acid-mediated mitogenic activation of mTOR signaling. Science (New York, N.Y.) *294*, 1942-1945.

Folch, J., Lees, M., and Sloane Stanley, G.H. (1957). A simple method for the isolation and purification of total lipides from animal tissues. The Journal of biological chemistry *226*, 497-509.

Forman, B.M., Chen, J., and Evans, R.M. (1997). Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. Proceedings of the National Academy of Sciences of the United States of America *94*, 4312-4317.

Gault, C.R., Obeid, L.M., and Hannun, Y.A. (2010). An overview of sphingolipid metabolism: from synthesis to breakdown. Advances in experimental medicine and biology *688*, 1-23.

Gidez, L.I. (1984). The lore of lipids. Journal of lipid research *25*, 1430-1436.

Goni, F.M., and Alonso, A. (1999). Structure and functional properties of diacylglycerols in membranes. Progress in lipid research *38*, 1-48.

Granger, J., and Remick, D. (2005). Acute pancreatitis: models, markers, and mediators. Shock (Augusta, Ga.) *24 Suppl 1*, 45-51.

Haemmerle, G., Lass, A., Zimmermann, R., Gorkiewicz, G., Meyer, C., Rozman, J., Heldmaier, G., Maier, R., Theussl, C., Eder, S., et al. (2006). Defective lipolysis and altered energy metabolism in mice lacking adipose triglyceride lipase. Science (New York, N.Y.) *312*, 734-737.

Haley, C.S., and Knott, S.A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity *69*, 315-324.

Hammond, K., Reboiras, M.D., Lyle, I.G., and Jones, M.N. (1984). Characterisation of phosphatidylcholine/phosphatidylinositol sonicated vesicles. Effects of phospholipid composition on vesicle size. Biochimica et biophysica acta *774*, 19-25.

Hannun, Y.A. (1994). The sphingomyelin cycle and the second messenger function of ceramide. The Journal of biological chemistry *269*, 3125-3128.

He, X., Huang, Y., Li, B., Gong, C.X., and Schuchman, E.H. (2010). Deregulation of sphingolipid metabolism in Alzheimer's disease. Neurobiology of aging *31*, 398-408.

Hutchins, P.D., Russell, J.D., and Coon, J.J. (2018). LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. Cell systems *6*, 621-625.e625.

Jha, P., McDevitt, M.T., Gupta, R., Quiros, P.M., Williams, E.G., Gariani, K., Sleiman, M.B., Diserens, L., Jochem, A., Ulbrich, A., et al. (2018a). Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species. Cell systems *6*, 722-733.e726.

Jha, P., McDevitt, M.T., Halilbasic, E., Williams, E.G., Quiros, P.M., Gariani, K., Sleiman, M.B., Gupta, R., Ulbrich, A., Jochem, A., et al. (2018b). Genetic Regulation of Plasma Lipid Species and Their Association with Metabolic Phenotypes. Cell systems *6*, 709-721.e706.

Karlsson, M., Contreras, J.A., Hellman, U., Tornqvist, H., and Holm, C. (1997). cDNA cloning, tissue distribution, and identification of the catalytic triad of monoglyceride lipase. Evolutionary relationship to esterases, lysophospholipases, and haloperoxidases. The Journal of biological chemistry *272*, 27218-27223.

Kennedy, E.P., Smith, S.W., and Weiss, S.B. (1956). New synthesis of lecithin in an isolated enzyme system. Nature *178*, 594-595.

Kooijman, E.E., Chupin, V., Fuller, N.L., Kozlov, M.M., de Kruijff, B., Burger, K.N., and Rand, P.R. (2005). Spontaneous curvature of phosphatidic acid and lysophosphatidic acid. Biochemistry *44*, 2097-2102.

Lander, E.S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics *121*, 185-199.

Lee, J.Y., Sohn, K.H., Rhee, S.H., and Hwang, D. (2001). Saturated fatty acids, but not unsaturated fatty acids, induce the expression of cyclooxygenase-2 mediated through Toll-like receptor 4. The Journal of biological chemistry *276*, 16683-16689.

Leekumjorn, S., Cho, H.J., Wu, Y., Wright, N.T., Sum, A.K., and Chan, C. (2009). The role of fatty acid unsaturation in minimizing biophysical changes on the structure and local effects of bilayer membranes. Biochimica et biophysica acta *1788*, 1508-1516.

Li, M., Yang, L., Bai, Y., and Liu, H. (2014). Analytical methods in lipidomics and their applications. Analytical chemistry *86*, 161-175.

Li, S., Xu, P., Han, L., Mao, W., Wang, Y., Luo, G., and Yang, N. (2017). Disease-syndrome combination modeling: metabolomic strategy for the pathogenesis of chronic kidney disease. Scientific reports *7*, 8830.

Li, Z., and Vance, D.E. (2008). Phosphatidylcholine and choline homeostasis. Journal of lipid research *49*, 1187-1194.

Lincoln, S.E., and Lander, E.S. (1992). Systematic detection of errors in genetic linkage data. Genomics *14*, 604-610.

Lofgren, L., Stahlman, M., Forsberg, G.B., Saarinen, S., Nilsson, R., and Hansson, G.I. (2012). The BUME method: a novel automated chloroform-free 96-well total lipid extraction method for blood plasma. Journal of lipid research *53*, 1690-1700.

Lou, W., Reynolds, C.A., Li, Y., Liu, J., Huttemann, M., Schlame, M., Stevenson, D., Strathdee, D., and Greenberg, M.L. (2018). Loss of tafazzin results in decreased myoblast differentiation in C2C12 cells: A myoblast model of Barth syndrome and cardiolipin deficiency. Biochimica et biophysica acta. Molecular and cell biology of lipids *1863*, 857-865.

Luiken, J.J., Schaap, F.G., van Nieuwenhoven, F.A., van der Vusse, G.J., Bonen, A., and Glatz, J.F. (1999). Cellular fatty acid transport in heart and skeletal muscle as facilitated by proteins. Lipids *34 Suppl*, S169-175.

Mackay, T.F., Stone, E.A., and Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. Nature reviews. Genetics *10*, 565-577.

Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A., and Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. Journal of lipid research *49*, 1137-1146.

McClearn, G.E., Tarantino, L.M., Rodriguez, L.A., Jones, B.C., Blizard, D.A., and Plomin, R. (1997). Genotypic selection provides experimental confirmation for an alcohol consumption quantitative trait locus in mouse. Molecular psychiatry *2*, 486-489.

Mejia, E.M., Cole, L.K., and Hatch, G.M. (2014). Cardiolipin metabolism and the role it plays in heart failure and mitochondrial supercomplex formation. Cardiovascular & hematological disorders drug targets *14*, 98-106.

Moritz, A., De Graan, P.N., Gispen, W.H., and Wirtz, K.W. (1992). Phosphatidic acid is a specific activator of phosphatidylinositol-4-phosphate kinase. The Journal of biological chemistry *267*, 7207-7210.

Neely, J.R., and Morgan, H.E. (1974). Relationship between carbohydrate and lipid metabolism and the energy balance of heart muscle. Annual review of physiology *36*, 413-459.

Niemela, P.S., Castillo, S., Sysi-Aho, M., and Oresic, M. (2009). Bioinformatics and computational methods for lipidomics. Journal of chromatography. B, Analytical technologies in the biomedical and life sciences *877*, 2855-2862.

Nishizuka, Y. (1992). Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. Science (New York, N.Y.) *258*, 607-614.

Ogiso, H., Suzuki, T., and Taguchi, R. (2008). Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. Analytical biochemistry *375*, 124-131.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., et al. (1992). The alpha/beta hydrolase fold. Protein engineering *5*, 197-211.

Osono, Y., Nakajima, K., and Hata, Y. (1995). Hypertriglyceridemia and fatty liver: clinical diagnosis of fatty liver and lipoprotein profiles in hypertriglyceridemic patients with fatty liver. Journal of atherosclerosis and thrombosis *2 Suppl 1*, S47-52.

Patricelli, M.P., Lovato, M.A., and Cravatt, B.F. (1999). Chemical and mutagenic investigations of fatty acid amide hydrolase: evidence for a family of serine hydrolases with distinct catalytic properties. Biochemistry *38*, 9804-9812.

Pauling, J., and Klipp, E. (2016). Computational Lipidomics and Lipid Bioinformatics: Filling In the Blanks. Journal of integrative bioinformatics *13*, 299.

Quehenberger, O., and Dennis, E.A. (2011). The human plasma lipidome. The New England journal of medicine *365*, 1812-1823.

Ramamurthi, K.S., and Losick, R. (2009). Negative membrane curvature as a cue for subcellular localization of a bacterial protein. Proceedings of the National Academy of Sciences of the United States of America *106*, 13541-13545.

Renner, L.D., and Weibel, D.B. (2011). Cardiolipin microdomains localize to negatively curved regions of Escherichia coli membranes. Proceedings of the National Academy of Sciences of the United States of America *108*, 6264-6269.

Rosenfeld, L. (2003). William Prout: early 19th century physician-chemist. Clinical chemistry *49*, 699-705.

Sax, K. (1923). The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS. Genetics *8*, 552-560.

Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J., et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science (New York, N.Y.) *316*, 1331-1336.

Scaloni, A., Barra, D., Jones, W.M., and Manning, J.M. (1994). Human acylpeptide hydrolase. Studies on its thiol groups and mechanism of action. The Journal of biological chemistry *269*, 15076-15084.

Schuhmann, K., Herzog, R., Schwudke, D., Metelmann-Strupat, W., Bornstein, S.R., and Shevchenko, A. (2011). Bottom-up shotgun lipidomics by higher energy collisional dissociation on LTQ Orbitrap mass spectrometers. Analytical chemistry *83*, 5480-5487.

Selvy, P.E., Lavieri, R.R., Lindsley, C.W., and Brown, H.A. (2011). Phospholipase D: enzymology, functionality, and chemical modulation. Chemical reviews *111*, 6064-6119.

Shemesh, T., Luini, A., Malhotra, V., Burger, K.N., and Kozlov, M.M. (2003). Prefission constriction of Golgi tubular carriers driven by local lipid metabolism: a theoretical model. Biophysical journal *85*, 3813-3827.

Simon, G.M., and Cravatt, B.F. (2010). Activity-based proteomics of enzyme superfamilies: serine hydrolases as a case study. The Journal of biological chemistry *285*, 11051-11055.

Simons, K., and Ikonen, E. (1997). Functional rafts in cell membranes. Nature *387*, 569-572.

Stafforini, D.M., Prescott, S.M., and McIntyre, T.M. (1987). Human plasma platelet-activating factor acetylhydrolase. Purification and properties. The Journal of biological chemistry *262*, 4223-4230.

Stefely, J.A., Kwiecien, N.W., Freiberger, E.C., Richards, A.L., Jochem, A., Rush, M.J.P., Ulbrich, A., Robinson, K.P., Hutchins, P.D., Veling, M.T., et al. (2016). Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. Nature biotechnology *34*, 1191-1197.

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E.A., Glass, C.K., Merrill, A.H., Jr., Murphy, R.C., Raetz, C.R., Russell, D.W., et al. (2007). LMSD: LIPID MAPS structure database. Nucleic acids research *35*, D527-532.

Tolhurst, G., Heffron, H., Lam, Y.S., Parker, H.E., Habib, A.M., Diakogiannaki, E., Cameron, J., Grosse, J., Reimann, F., and Gribble, F.M. (2012). Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. Diabetes *61*, 364-371.

van Eijk, M., Aten, J., Bijl, N., Ottenhoff, R., van Roomen, C.P., Dubbelhuis, P.F., Seeman, I., Ghauharali-van der Vlugt, K., Overkleeft, H.S., Arbeeny, C., et al. (2009). Reducing glycosphingolipid content in adipose tissue of obese mice restores insulin sensitivity, adipogenesis and reduces inflammation. PloS one *4*, e4723.

Verrier, E., Wang, L., Wadham, C., Albanese, N., Hahn, C., Gamble, J.R., Chatterjee, V.K., Vadas, M.A., and Xia, P. (2004). PPARgamma agonists ameliorate endothelial cell activation via inhibition of diacylglycerol-protein kinase C signaling pathway: role of diacylglycerol kinase. Circulation research *94*, 1515-1522.

Villani, M., Subathra, M., Im, Y.B., Choi, Y., Signorelli, P., Del Poeta, M., and Luberto, C. (2008). Sphingomyelin synthases regulate production of diacylglycerol at the Golgi. The Biochemical journal *414*, 31-41.

Vishwanath, V.A. (2016). Fatty Acid Beta-Oxidation Disorders: A Brief Review. Annals of neurosciences *23*, 51-55.

Wang, Q., Yang, G., Liu, Y., and Feng, Y. (2006). Discrimination of esterase and peptidase activities of acylaminoacyl peptidase from hyperthermophilic Aeropyrum pernix K1 by a single mutation. The Journal of biological chemistry *281*, 18618-18625.

Weijers, R.N. (2012). Lipid composition of cell membranes and its relevance in type 2 diabetes mellitus. Current diabetes reviews *8*, 390-400.

Werner, M.H., Bielawska, A.E., and Hannun, Y.A. (1992). Quantitative analysis of diacylglycerol second messengers in human platelets: correlation with aggregation and secretion. Molecular pharmacology *41*, 382-386.

Whiley, L., Sen, A., Heaton, J., Proitsi, P., Garcia-Gomez, D., Leung, R., Smith, N., Thambisetty, M., Kloszewska, I., Mecocci, P., et al. (2014). Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease. Neurobiology of aging *35*, 271-278.

Williams, E.G., Wu, Y., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. Science (New York, N.Y.) *352*, aad0189.

Wood, P.L., Medicherla, S., Sheikh, N., Terry, B., Phillipps, A., Kaye, J.A., Quinn, J.F., and Woltjer, R.L. (2015). Targeted Lipidomics of Fontal Cortex and Plasma Diacylglycerols (DAG) in Mild Cognitive Impairment and Alzheimer's Disease: Validation of DAG Accumulation Early in the Pathophysiology of Alzheimer's Disease. Journal of Alzheimer's disease : JAD *48*, 537-546.

Yang, K., Cheng, H., Gross, R.W., and Han, X. (2009). Automated lipid identification and quantification by multidimensional mass spectrometry-based shotgun lipidomics. Analytical chemistry *81*, 4356-4368.

Yetukuri, L., Katajamaa, M., Medina-Gomez, G., Seppänen-Laakso, T., Vidal-Puig, A., and Oresic, M. (2007). Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis. BMC systems biology *1*, 12.

Young, B.P., Shin, J.J., Orij, R., Chao, J.T., Li, S.C., Guan, X.L., Khong, A., Jan, E., Wenk, M.R., Prinz, W.A., et al. (2010). Phosphatidic acid is a pH biosensor that links membrane biogenesis to metabolism. Science (New York, N.Y.) *329*, 1085-1088.

Chapter 2: Systems analyses reveal physiological roles and genetic regulators of liver lipid species

Pooja Jha[1], Molly T. McDevitt[1], Rahul Gupta[1], Pedro M. Quiros, Evan G. Williams, Karim Gariani, Maroun B. Sleiman, Leo Diserens, Adam Jochem, Arne Ulbrich, Joshua J. Coon, Johan Auwerx, and David J. Pagliarini

[1]These authors contributed equally.

This chapter was published as an article in *Cell Systems* (2018, 6, 722-733) by the authors listed above.

Summary

The genetics of individual lipid species and their relevance in disease is largely unresolved. We profiled a subset of storage, signaling, membrane, and mitochondrial liver lipids across 385 mice from 47 strains of the BXD mouse population fed chow or high-fat diet and integrated these data with complementary multi-omics datasets. We identified several lipid species and lipid clusters with specific phenotypic and molecular signatures and, in particular, cardiolipin species with signatures of healthy and fatty liver. Genetic analyses revealed quantitative trait loci for 68% of the lipids (lQTL). By multi-layered omics analyses, we show the reliability of lQTLs to uncover candidate genes that can regulate the levels of lipid species. Additionally, we identified lQTLs that mapped to genes associated with abnormal lipid metabolism in human GWASs. This work provides a foundation and resource for understanding the genetic regulation and physiological significance of lipid species.

Introduction

An enormous number of chemically distinct molecular lipid species arise from the various combinations of fatty acids and backbone structures such as glycerol. However, it is not intuitively clear why nature has created so many different forms of lipids (Wenk, 2005). Advancements of lipidomics technologies have provided the first step towards generating a repertoire of all lipid species on a systems scale. The next major challenge is to elucidate their regulation, function and physiological impact and to discover how these lipids interact to influence specific biological processes.

Combining systems genetics with multi-omics strategies is helpful to understand the association of lipid species with genes, proteins or physiological traits (Civelek and Lusis, 2014; Hyotylainen and Oresic, 2014). Similar to the novel insights gained from quantitative trait loci (QTL) analysis of transcripts, proteins or phenotypes (Andreux et al., 2012; Williams et al., 2016; Wu et al., 2014), the QTLs of lipid species in liver can provide insights into their genetic regulation. Transcriptomic, proteomic and phenotypic data can additionally be used to generate lipid–transcript, –protein or –phenotype correlations, providing a more comprehensive view of how lipids fit in the network of cellular processes. Likewise, reducing large lipidomic data to clusters of co-regulated lipids enables the identification of functionally related lipid species and helps to clarify the relationship of the lipid clusters with phenotypic traits.

It has long been possible to quantify entire classes of lipids *en masse*. Recently however, it has been shown that by dissecting broad lipid classes into specific lipid species, one can develop a more granular understanding of lipid-related disease etiology, thereby improving the capacity to find and validate drug targets. For instance, neutrophils from patients with periodontal tissue disease accumulate specific diacylglycerol (DAG) species, in particular 1,2-dipalmitoyl DAG (Gronert et al., 2004). Similarly, only some molecular species of ceramides were shown to be associated with certain types of cancers or with pathways that lead to ceramide-induced apoptosis (Koybasi et al., 2004; Kroesen et al., 2001). Given the challenges in identifying gene-

environment interactions (GxE) influencing the levels of lipid species in humans (Franks and McCarthy, 2016), we exploited the BXD mouse genetic reference population (GRP), descending from crosses between C57BL/6J mothers and DBA/2J fathers (Peirce et al., 2004). In this genetically diverse population, we can tightly control the dietary state of the individual mice over months to analyze how genes (genotype), environment (diet) and their interactions influence hepatic lipid species. We used systems genetics strategies (Civelek and Lusis, 2014) including QTL mapping, network construction, and module-trait correlation integrated with multi-omics datasets (genomics, transcriptomics, proteomics, and phenomics) to understand the relationship between levels of lipid species and molecular and clinical traits. These approaches helped us to identify specific cardiolipin species with signatures of healthy or fatty liver and candidate genes that regulate the levels of specific lipid species.

Results

Liver lipid profiles and their relationship with clinical and molecular traits

We used targeted lipidomics to measure 96 hepatic lipid species in 385 mice belonging to 84 cohorts of the BXD GRP—43 fed CD and 41 fed HFD from 47 BXD strains. Over 29 weeks, the mice underwent extensive metabolic phenotyping (Williams et al., 2016). After an overnight fast, mice were sacrificed and liver samples were collected for lipidomics analyses. To test the quality of mass spectrometry (MS) measurements, we performed pairwise-correlation of technical and extraction replicates, which showed consistently robust correlations (Figures S1A and S1B). Furthermore, we performed all possible pairwise correlations between the measured lipid species from different groups to assess the sensitivity of our measurements in detecting diet- and strain-driven differences (Figure S1C). The correlation of biological replicates within each strain in either diet was higher than within strain-across diet (CD vs. HFD) correlation, and, as expected, the correlation across strains on a given diet (either CD or HFD) was higher than across strain-across diet correlation (Figure S1C). The lipids measured include free fatty acids (FFA, 8 species), glycerolipids [triacylglycerol (TAG, 38 species) and diacylglycerol (DAG, 6 species)], glycerophospholipids [phospholipids (PL, 20 species) cardiolipins (CL, 23 species)] and coenzyme Q_9 (Table S1). Hierarchical cluster analysis demonstrated that BXD cohorts did not completely segregate based on their diet, indicating that the GxE interaction can overpower the strong dietary impact on lipid profiles (Figure 1A). To obtain an overview of the interaction between lipid species, we performed an unweighted correlation network analysis, which showed most lipids to be highly correlated and the correlations within class to be generally stronger (Figure 1B ($p<1e-04$) and S1D ($p<1e-03$)). Some dietary effects were evident in the correlation networks, such as TAG species, which formed a tight-knit cluster in CD but were more interspersed in HFD, revealing the change in neutral lipid homeostasis in HFD (green nodes; Figure 1B, and S1D). While the CLs, found primarily in the mitochondrial inner membrane, formed a tight cluster in both diets, the

PLs, that are the general cellular membrane lipids, were highly interspersed between different lipid classes in both diets (red and yellow nodes, respectively; Figure 1B and S1D). This scattered PL profile is in line with the fact that PLs are the main substrates and intermediates in the biosynthesis of various lipid classes (Han, 2016).

To identify lipids with similar physiological and molecular characteristics, we performed a weighted correlation network analysis (WGCNA), which clusters correlated groups of lipids into modules (Langfelder and Horvath, 2008). More than half of the lipids were clustered in 8 modules in both diets (66% in CD and 58% in HFD) (Figures 1C and Figure S2A; Table S2) and most module composition was conserved in both diets (Figures S2B and S2C). We then performed correlation analysis between the eigenlipid (first principal component) of each module and clinical traits related to liver function (Figure 1D). In CD, the DAG (black), TAG (brown) and CL (yellow) modules positively correlated with obesity and other traits associated with liver dysfunction, while the CD-specific TAG-PL purple module—containing species with a higher degree of unsaturation (Figure 1C)—negatively correlated with these traits (Figure 1D). In HFD, several modules showed positive correlation with obesity-associated traits, particularly DAG (black), TAG (brown and magenta) and CL (blue and yellow) modules (Figure 1D). Interestingly, none of the TAG modules correlated with the total TAG concentration in liver and plasma despite the significant correlation with liver weight (Figure 1D). This disparity can be explained by the fact that we measured only 38 TAG species and 6 DAG species which constitutes only a fraction (~15%) of all the TAGs and DAGs that exist in mouse liver. Therefore, this subset may not necessarily reflect the total TAG concentration measured by the enzymatic assay.

To identify molecular mechanisms underlying the lipid clustering, we correlated the module eigenlipids with liver proteome (2,622 proteins) from the same mouse cohort (Williams et al., 2016). We selected all proteins that significantly correlated with each module eigenlipid and performed KEGG enrichment analysis for all positively and negatively correlated proteins, separately. Interestingly, modules that earlier correlated with obesity and liver dysfunction in both

diets, including DAG (black), TAG (brown and turquoise) and CL (blue and yellow) modules (Figure 1D), positively correlated with pathways associated with fatty acid and glycolytic metabolism, peroxisome and PPAR signaling (Figure 1E). Further, these modules correlated negatively with oxidative and proliferative pathways like oxidative phosphorylation, lysosome and ribosome pathways (Figure 1E). In line, the TAG & PL purple module showed the opposite trend, negatively correlating with fatty acid metabolic pathways and positively with oxidative pathways, following the same trend as in the module-trait correlations (Figures 1D and 1E). Collectively, these findings reveal that all lipid species of a lipid class do not necessarily have the same molecular regulation and phenotypic impact.

Identification of cardiolipin species as signatures of healthy and fatty liver

From all lipids measured, we identified two clusters of lipid species with strong diet-independent association with liver mass (Figure 2A). A cluster of 13 species composed of TAGs and DAGs with a low degree of unsaturation (1-3 double bonds; dominated by lipids from the black and brown modules) correlated positively with liver mass (Figure 2A; green font); whereas, another cluster of 6 lipids comprising highly unsaturated TAGs (6-7 double bonds; dominated by lipids from the purple module) along with two phosphatidylserine species, correlated negatively with liver mass (Figure 2A; orange font). Since the diet changed the landscape of most liver lipids, we next analyzed lipids that strongly associated with liver mass in each individual diet (Figure 2B). Twenty-seven lipids in CD and 40 in HFD (including all 19 lipids from Figure 2A) strongly correlated with liver mass. Of note, liver mass was centrally positioned in the resulting HFD network showing dense correlations with the 40 lipid species compared to the CD network where liver mass was at the periphery of the network with 27 lipids (Figure 2B). Interestingly, a subset of nine CL and monolyso-CL (MLCL) species showed a predominant association with liver mass in HFD, but not in CD (Figure 2B; red nodes). This finding is noteworthy because CL—the signature phospholipid of the mitochondrial inner membrane—is indispensable for a range of

mitochondrial activities (Claypool and Koehler, 2012). Alterations in the content and/or structure of CL have been reported in several tissues in a variety of pathological settings. However, a major unresolved question is whether CL molecules with different acyl chain compositions differ functionally (Claypool and Koehler, 2012). Of these nine species, only tetralinoleoyl-CL (CL(LLLL)) and its precursor/remodeling intermediate, trilinoleoyl-MLCL (MLCL(LLL)) (neither belonging to any module) showed negative correlation with liver mass; whereas the other seven CLs enriched in monounsaturated FAs (MUFA), oleic (O) and palmitoleic (Po) acid (all from the yellow module), positively correlated with liver mass (Figure 2C). This demonstrates a change in CL remodeling under HFD that depletes the CL predominant in healthy tissue—CL(LLLL) (Chicco and Sparagna, 2007), and its precursor, MLCL(LLL)—suggesting that these CL species may be signatures of healthy/ normal liver. Conversely, the other seven MUFA enriched CLs that correlated positively with liver mass in HFD may be considered as signatures of unhealthy/ fatty liver.

We next tested whether this change in the profile of nine CL species is a general phenomenon in other dietary-induced models of hepatic steatosis and mitochondrial dysfunction and if the profile can be reverted by ameliorating hepato-steatosis via enhancing mitochondrial function. We have previously shown that nicotinamide riboside (NR) treatment ameliorates high-fat high-sucrose (HFHS) diet induced fatty liver disease by boosting nicotinamide adenine dinucleotide (NAD^+) levels and thereby enhancing mitochondrial function (Gariani et al., 2016). Therefore, we performed lipidomic profiling from the livers of C57BL/6J mice—the most commonly used laboratory mouse strain—fed on (i) CD (ii) HFHS diet for 18 weeks or (iii) HFHS + nicotinamide riboside (NR), added 9 weeks after the start of the HFHS diet (therapeutic approach) (Gariani et al., 2016). In line with our findings from the HFD-fed BXD study, HFHS diet decreased the CL signatures of healthy/ normal liver—CL(LLLL) and MLCL(LLL)—whereas it increased the six CL signatures of unhealthy /fatty liver—MLCL(LOO), MLCL(LLO), CL(LOOPo), CL(LLPoP), CL(LOOO), CL(OOOP)—enriched in MUFAs (Figures 2D and 2E). CL(OOOO), detected in the

BXD study (Figure 2B/2C), was too low to be detected in all samples from the NR study. Interestingly, NR treatment increased the levels of the healthy CLs, whereas it decreased the levels of the unhealthy CLs (Figures 2D and 2E). Importantly, the two healthy CLs correlated negatively with obesity and NAFLD traits while the unhealthy CLs showed positive correlation (Figure 2F). These data show that all lipid species within a class do not necessarily behave similarly, as demonstrated here with specific CL species that have signatures of healthy or fatty liver.

Liver lipids are influenced by multiple genomic loci

We next analyzed globally how genotype and diet influence lipid species (Figure 3A). 63% of the measured lipids were significantly impacted by diet: 28 upregulated in CD (green font) and 33 upregulated in HFD (blue font) (Figure 3A and Table S3). Approximately half of the lipids correlated positively between the diets, 24 of them being significant (orange bars, Figure 3A and Table S3). Next, we assessed the heritability (h^2; percentage of trait variation attributed to additive genetic factors) for all lipid species within dietary groups (CD-light red and HFD-dark red, Figure 3A) and across both diets combined (CD+HFD / Mixed)(Belknap, 1998) (Table S4). Within a dietary cohort, 30% to 60% of the observed variance in lipid levels could be explained by genetic differences across strains (i.e. $h^2 \geq 30\%$) for the strong majority of lipid species (66% in CD and 76% in HFD). Conversely, when dietary cohorts were combined, only 13% of lipids had h^2 above 30% (Table S4).

Next, we mapped QTLs for all lipid species (lQTL) and lipid modules (modQTL) (Table S5). Most chromosomes contained at least one lQTL and some hotspot regions on chromosomes 2, 4, 6, 9, 15 and X were quite distinct (Figure 3A). We detected 136 lQTLs: 55 in CD and 81 in HFD from 37 and 46 lipid species, respectively (Figure 3B). While over half of these lipids had only one QTL (26 CD and 21 HFD), the remaining lipids had more than one lQTL (11 in CD and 25 in HFD), indicating a polygenic regulation of lipid species (Figure 3B). lQTLs were typically

unique to either CD or HFD, with only one lipid species—TAG(58:8)—mapping to the same locus on chromosome 11 (118.9-118.5 Mb) in both diets (Figure S3A). Interestingly, this locus harbors genes involved in lipid (*Acox1*, *Fasn*, *P4hb*, *St6galnac1*) and carbohydrate (*Afmid*, *Gaa*, *Galk1*) metabolic processes. Importantly, these genes (among others in this region) also have cis-e/pQTLs in liver and coding sequence variants in the BXDs (Figure S3A). Furthermore, TAG(58:8) levels were not significantly different across diets and were among those having high h^2 in both dietary cohorts (h^2>40%) and also when dietary cohorts were combined (33.5%) (Figure S3B). This indicates that TAG(58:8) is predominantly under the same genetic control in both diets. The lack of cross-diet overlap in rest of the lQTLs suggests that GxE factors regulate nearly all lipid species.

Genetic assessment of the lQTLs

To assess the efficacy of lQTLs and find candidate genes that regulate lipid levels, we performed an in-depth analysis of all significant and suggestive lQTLs (Figure 4A, lQTLs above the blue dotted line; 55 in CD and 81 in HFD). All genes under these lQTLs were filtered along four parallel pipelines (Figure 4B): transcripts/proteins under lQTLs (i) with non-synonymous SNPs, (ii) with self-regulating QTLs (cis e- and/or p-QTL), (iii) with significant correlation ($p<0.05$) with the lipid itself, and (iv) with variable transcript expression (standard deviation >0.25). While the integration of proteomics data was advantageous, transcriptomics data was substantially more consequential to candidate selection as more genes were assayed (~20,000 with transcriptomic *vs* ~2600 with proteomics data) and this also avoids post-translational and protein regulation variations. Genes passing 2 of the 4 filters in either diet (299 of 2845 genes in CD and 327 of 3818 in HFD) were analyzed for gene ontology (GO) pathway enrichment (Figure 4B). Remarkably, the filtered lQTL candidate genes were enriched in lipid metabolic pathways (Figure 4C). We next applied the same filtering criteria for the top 14 lQTLs (Figure 4A, lQTLs above the

black dotted line) comprising of 4 CL, 6 TAG, 2 PL, and the FFA 16:0 and 16:1n7 lQTLs (Figure 4D). These lQTLs had many genes fulfilling 3-4 of our filtering criteria (Figure 4D, genes indicated in orange and red font) and several with known association with metabolic and liver-associated phenotypes (Figure 4D, genes indicated with asterisks). In total, 104 candidate genes that fulfilled 2 or more of the filtering criteria from these 14 lQTLs were subsequently analyzed for enrichment of KEGG pathways. In line with the above findings, these genes were enriched in FA metabolic processes in addition to inflammatory response (Figure 4E). We have provided an exhaustive list of candidate genes (based on the filtering pipeline of Figure 4B) for each lQTL (Table S6; CD and HFD lQTLs provided in separate excel sheets). Additionally, genes under the lQTLs that pass at least two filters are tabulated in Table S5. Taken together, these data show that lipid species of the same class can be regulated by several loci throughout the genome and demonstrate the utility of integrating multi-omics datasets with lQTLs in the identification of putative genetic regulators of lipid species.

The FFA module maps to a genetic hotspot associated with FA metabolism and signaling

Next, we analyzed one of the diet-specific hotspot lQTL region to demonstrate the validity of our lQTLs at the genetic and physiological level. Module-level QTL scans showed a QTL for the Red module (comprising six of the eight FFAs measured) only in HFD and not CD on chromosome 4 (Figure 5A). The QTL for the Red module overlapped with that of its constituent FFAs and contained several genes passing one or more of our filtering criteria (from Figure 4B), including 12 genes involved in fatty acid metabolism, insulin, interferon and toll-like receptor signaling, suggesting a robust hotspot region of these metabolic processes associated with FFAs (Figure 5B) (Jump, 2011; Malhi and Gores, 2008). Among these 12 genes, it is important to highlight methylthioadenosine phosphorylase (MTAP), which passes all 4 filtering criteria. MTAP

deficiency has recently been shown to increase the risk for progression of chronic liver disease due to compromised liver proteome methylation (Bigaud and Corrales, 2016).

Having identified the potential genetic regulatory region of these FFAs under HFD, we next tested if oleic and palmitoleic acids of the module—showing the strongest strength of association (Figure 5A [right], thickness of edges)—show the typical biological profile of Stearoyl-CoA Desaturase (SCD)-mediated MUFA synthesis under HFD (Figure 5C, adapted from (Jump, 2011). Indeed, SCD activity index, as assessed by the ratios of MUFA to saturated FA (SFA)—16:1n7/16:0 and 18:1n9/18:0—was significantly increased in HFD cohorts (Figure 5D). Additionally, the lipid ratios as readouts of the SCD activity index showed positive correlation with both mRNA and protein levels of SCD in HFD, but not in CD (Figure 5E).

Since fat is the predominant energy source in HFD (60.3% kcal from fat; 27.3% kcal from carbohydrate) as opposed to carbohydrate in CD (6.2% kcal from fat; 44.2% kcal from carbohydrate), we tested whether the FFAs in this module reflect this at the physiological level. To do so, we tested if the red module eigenlipid correlated negatively with respiratory exchange ratio (RER). RER is a measurement of the primary energy substrate used by an organism: lower RER values indicate fat is the predominant energy source, while higher values indicate higher contribution from carbohydrates. Indeed, the red module correlated negatively with RER in HFD cohorts, which have lower basal RER (Figure 5F; black dots confined towards left of the x-axis), whereas no significant correlation was observed in CD cohorts (Figure 5F; green dots spread toward the right of the x-axis). These findings demonstrate the validity of our QTL mapping and highlight the potential utility of WGCNA clustering of lipids in modules for lipidomics analysis and the identification of biologically relevant modQTLs.

Identification of lQTL genes associated with abnormal lipid metabolism in human GWAS and mapping of lQTLs in TAG biosynthetic pathway

Human GWAS have identified many genetic variants associated with plasma lipids and abnormal lipid metabolism (http://jjwanglab.org/gwasdb, https://www.ebi.ac.uk/gwas/). Taking advantage of these genes identified in GWAS studies (Table S7), we screened the lQTLs for the presence of any human GWAS genes associated with abnormal lipid metabolism (Figure 6A). We screened only those genes under the lQTLs which fulfilled at least two of the four filtering criteria, shown in Figure 4B. 27 lQTLs (7 CD and 20 HFD) harbored 20 out of 494 genes pre-selected from human GWAS for abnormal of lipid metabolism (Figure 6A; see Table 7 for extended information on the identified hits). To test the probability of this overlap by chance, we performed 10,000 permutations, each of which involved comparing the number of genes that overlap between a random set of 494 human genes and the 566 BXD lQTL genes (from Table S5). The distribution of the overlap across all permutations formed the null distribution. Only 1.49% of random trials had an overlap greater than or equal to the true overlap of 20 genes (Figure S3C) corresponding to a p value of 0.0149. This indicates that the probability that 20 or more GWAS genes found to overlap by chance under the calculated null distribution is only 1.49%.

Of note, the QTL position of three lipid species: TAG(54:6)_2, TAG(56:8) and PI(20:4_16:0) mapped to genes implicated in NAFLD in human GWAS, including *TM6SF2, NCAN, CILP2, PPP1R3B and LYPLAL1*(Anstee and Day, 2013; Kahali et al., 2015; Lusis et al., 2016) (Figures 6A and 6B). TAG(54:6)_2 mapped to *Tm6sf2* (influences TG secretion and hepatic lipid droplet content), and *Ncan* (cell adhesion) on chromosome 8; notably, these two genes along with *Cilp2* (carbohydrate binding) have protein coding variants in the BXDs. Additionally, the region of *Ncan, Tm6sf2* and *Cilp2* is syntenic with the localization of these genes on human chromosome 19 (Figure 6B, middle) suggesting a conserved role/regulation of these genes in both mice and humans. TAG(56:8) mapped to *Ppp1r3b* (limits glycogen breakdown) on chromosome 8 and PI(20:4_16:0) mapped to *Lyplal1* (having lysophospholipase activity) on chromosome 1 (Figure 6B) both of which have coding variants and *cis*-eQTLs in liver, heart, muscle and adipose tissue of BXDs. Taken together, these links from human GWAS to lQTLs

provide a basis for understanding both the function of numerous understudied and/or uncharacterized GWAS genes and the role of the individual lipid species in health and disease.

Next, we tested whether the total TAG content mapped to any of the 8 genes (*PNPLA3, GCKR, TRIB1, LYPLAL1, PPP1R3B, TM6SF2, NCAN* and *CILP2*) proven to cause or increase the susceptibility to hepatic steatosis/NAFLD in human GWAS (Anstee and Day, 2013; Kahali et al., 2015; Lusis et al., 2016). Hepatic TAG quantification (normalized to protein levels) in BXDs did not show any significant difference between the CD and HFD cohorts (Figure S4A). However, the TAG content in CD cohorts correlated with proteins and transcripts that were enriched in lipid metabolic processes (Figure S4B) and the TAG content in HFD cohorts correlated positively with body weight and fat mass and negatively with the lean mass (Figure S4C). This association of hepatic TAG content with its expected physiological and molecular function affirms the reliability of the TAG measurement. QTL mapping of the hepatic TAG content showed two suggestive QTLs at chromosomes 4 and 17 in HFD cohorts, whereas no QTL was observed in CD (Figure S4D). Notably, TAG content did not map to any of the human NAFLD GWAS genes, nor to any of the individual TAG lQTLs. This indicates a diverse genetic regulation of individual TAG species, different from the regulation of total TAG content and highlights the importance of examining individual TAG species. Therefore, we examined whether genes known to be involved in the TAG biosynthesis pathway (Table S7) were found to fulfill any one of the four filtering criteria shown in Figure 4B in either diet for individual TAG, FFA, and PL lQTLs. 14 genes (4 CD, 7 HFD, and 3 CD+HFD) involved in TAG biosynthesis were found under 30 lQTLs (Figure 6C). Collectively, our data demonstrate the complexity of the genetic regulation of different lipid species and the importance of studying individual lipid species to find common targets between mice and humans.

Discussion

Data on the genetic regulation and physiological significance of individual lipid species on a population scale is very limited for both mouse (Hui et al., 2015) and human populations (Gronert et al., 2004; Koybasi et al., 2004; Kroesen et al., 2001). Here, we profiled subset of liver lipid species from different classes and integrated them with multi-omics datasets (genetics, transcriptomics, proteomics and phenomics) to understand their genetic, environmental and GxE regulations and physiological roles. We identified individual lipid species and lipid modules that impact different clinical traits, revealing a link between their levels and their physiological and molecular functions. In particular, specific CL species were identified that associate either positively or negatively with obesity and NAFLD signatures. Last, we demonstrate the reliability of lQTLs through several examples that were supported by multi-omics analysis and provide a resource of candidate genes that can regulate many of the lipid species we measured. Our findings illustrate the importance of studying individual lipid species and provide a platform for further mechanistic studies of lipid species.

Similar to the plasma lipids in our companion article (Jha et al., 2018), hepatic lipid species are also correlated across lipid class. By clustering lipids into modules, we established functional links between these modules and their molecular and physiological signatures. Two observations were particularly distinct between the plasma and liver lipids. First, in contrast to the plasma lipids profile, where unsupervised hierarchical clustering distinctly separated the CD and HFD groups (Jha et al., 2018), liver lipids profile did not show a distinct dietary separation. Second, the h^2 of the liver lipids was lower than that of plasma lipids. These differences can be attributed to the fact that (i) liver, being the hub of most metabolic processes, is more dynamic and influenced by a multitude of factors (diet, hormones, stress) *vs*. the plasma lipid profile, which represents an equilibrium state as a result of metabolic processes from all peripheral tissues; (ii) additional steps such as tissue homogenization and normalization to protein levels is required for liver lipidomics,

increasing the possibility of technical error.

A high fat dietary challenge had a striking impact on the cardiolipin profile, wherein the CL signatures of healthy liver (CL(LLLL) and MLCL(LLL)) were decreased and the signatures of fatty liver (MUFA-enriched CLs and MLCLs) were increased. Notably, CL(LLLL) comprises the majority of CLs in the inner mitochondrial membrane and a loss in CL(LLLL) content typically reflects the loss of mitochondrial mass (Koekemoer and Oelofsen, 2001). While a decrease in CL(LLLL) has been categorically associated with heart failure, senescence and the Barth syndrome (Chicco and Sparagna, 2007), its role in hepatic steatosis/NAFLD has not been clearly established (Cole et al., 2016). Therefore, we validated our findings from the HFD-BXD study in another model of HFHS-induced hepatic steatosis, which confirmed that indeed healthy CLs decrease in NAFLD and negatively correlate with obesity and NAFLD traits and conversely, unhealthy CLs increase in NAFLD and positively correlate with obesity and NAFLD traits. This suggests that this change in the CL profile, in particular a decrease in CL(LLLL) content, could be one of the underlying causes of hepatic steatosis and may in itself serve as a disease biomarker. If this is the case, then interventions to ameliorate steatosis should be accompanied by increase in healthy CLs and a decrease in unhealthy CLs, as illustrated by treating mice having NAFLD with the NAD^+ precursor, NR. This finding adds to the conceptual understanding of the pathways via which NR may act to boost mitochondrial function (Gariani et al., 2016), i.e. by increasing CL(LLLL) and MLCL(LLL) content in liver.

QTL analysis showed that about half of the lipids (having QTLs) mapped to multiple QTLs, suggesting a polygenic regulation of lipid species. This is in line with the diverse functional roles of lipids in signaling, in protein-binding, in membrane function, and as energy substrates (Han, 2016). These results furthermore show that lipid species are complex traits affected not only by the environment, but also by multiple genetic pathways—i.e. not regulated primarily by single rate-limiting enzymatic pathways. The complexity of lipidomics can be better appreciated when compared and contrasted with metabolomics/metabolites where often there is one key regulatory

enzyme for a reaction or a pathway to which metabolite QTLs will frequently map to (Wu et al., 2014). In contrast, for most lipid species, it is uncommon that one enzyme regulates the level of a particular lipid species. The number of genes and proteins in a particular lipid biosynthetic pathway is a lot fewer than the species in each class of lipid, (Quehenberger and Dennis, 2011) indicating that (1) one gene regulates a number of lipid species and (2) many genes can regulate one lipid species.

To prove the reliability of the lQTLs, we performed a multi-omics analysis of candidate genes under the lQTLs, which provided compelling candidate genes and enriched pathways associated with lipid metabolic processes. If the mouse is to serve as a model of metabolic and lipid traits in humans, it is important that the relevant pathways are conserved in the two species. One measure of such conservation is the degree of overlap between mouse QTLs and human GWAS data. Our data shows several lQTLs harboring human GWAS candidate genes for lipids and associated metabolic traits. These findings build a strong foundation for future mechanistic studies to find common links between the genetic control of lipid metabolism between mouse and human, as has been successfully achieved by the hybrid mice diversity panel (HMDP) for other traits, such as osteoporosis, obesity, blood cell levels and heart failure (Lusis et al., 2016). Hence, we provide here candidate genes for all lQTLs, as a cornerstone for future research on the regulation of individual lipids across species, with human translational value.

Interestingly, total hepatic TAG content did not have any significant QTL overlap individual TAG lQTLs. This highlights a key point: the regulation of individual lipid species is diverse and distinct from the total concentration of the class to which they belong. Therefore, it is necessary to have more focused studies on individual lipid species, which can uncover their functions as well. In agreement with this, we have illustrated in our companion article (Jha et al., 2018) that specific TAG species can have positive or negative signatures of NAFLD. Future work on other important lipid classes including extensive coverage of ceramides, DAGs and eicosanoids will be helpful in dissecting other equally relevant NAFLD and metabolic syndrome signatures.

In summary, this liver-centric lipidomics study provides a framework to uncover the genetic regulation and physiological impact of individual lipid species, with an ultimate goal to improve our understanding of diseases linked to abnormal lipid metabolism. This insight may help to identify new drug targets involved in lipid disorders and to guide the development of clinically relevant disease signatures.

Experimental procedures

Mice

BXD strains were obtained from University of Tennessee Health Science Center (Memphis, TN, USA) and JAX (The Jackson Laboratory) and bred at the École Polytechnique Fédérale de Lausanne (EPFL) animal facility for more than two generations before incorporation into the study. Cohorts of 47 BXD strains with ~5 males each on CD and HFD were used in this study. Mice from 43 strains were fed a chow diet [CD; 2018 Teklad Global 18% Protein Rodent Diet (6.2% kcal from fat; 44.2% kcal from carbohydrate; 18.6% kcal from protein)] and from 41 strains a high fat diet [HFD; Harlan Teklad, TD.06414 (60.3% kcal from fat; 27.3% kcal from carbohydrate; 18.4% kcal from protein)] for 21 weeks, starting 8 weeks of age. During the course of 21 weeks, mice underwent extensive metabolic phenotyping as described (Williams et al., 2016). At week 29, animals were fasted overnight before sacrifice at 9:00 am. Blood was collected from isoflurane-anesthetized mice via the vena cavae, and immediately afterwards animals were perfused with 4°C PBS, through the left ventricle, then organs were harvested. Blood was collected in lithium-heparin (LiHep)–coated tubes (Microvette CB 300 Hep-Lithium, Sarstedt) shaken and kept in ice. The blood samples were centrifuged at 4500 revolutions per minute (rpm) for 10 min at 4°C before being flash-frozen in liquid nitrogen for subsequent measurement of plasma lipid species (see companion article (Jha et al., 2018)) and plasma clinical traits. Due to breeding limitations and unforeseen deaths of some mice during the course of phenotyping, all strains do not have 5 mice each, average being 3-5 mice/strain.

For *in vivo* validation of CL species, liver samples were used from our previous study (Gariani et al., 2016). In brief, male C57BL/6J mice were separated into three groups of 6-9 mice per group, at the age of 7 weeks. Animal cohorts were fed a CD, a Western high-fat and high-sucrose (HFHS) diet [HFHS; Harlan Teklad, TD.08811, (44.6% kcal from fat; 40.7% kcal from carbohydrate; 14.7% kcal from protein)] or a HFHS diet that was supplemented with NR

(400 mg/kg/day) at week 16 till week 25 (9 weeks). Mice were sacrificed after a 4 hour fast at 9:00 am. All experiments were approved by the Swiss cantonal veterinary authorities of Vaud under licenses 2257, 2257.1 and 2465.

Fatty Acid Composition (%) in the Diet of BXD Cohorts

Fatty acids	Symbol	CD	HFD
Palmitic	C16:0	0.7	8.02
Stearic	C18:0	0.2	3.93
Oleic	C18:1n9	1.2	14.68
Linoleic	C18:2n6	3.1	4.7
Linolenic	C18:3n3	0.3	0.55
Saturated fat		0.9	12.48
Monounsaturated fat		1.3	16.05
Polyunsaturated fat		3.4	5.4
Total fat		**6.2**	**34.3**

Plasma clinical traits of BXD cohorts

Plasma parameters were measured on 2 times diluted samples (1:1 ratio of plasma to diluent) using Dimension®Xpand Plus (Siemens Healthcare Diagnostics AG, Dudingen, Switzerland). The biochemical tests were performed according to the manufacturer instructions for each parameters: AST (Siemens Healthcare, DF41A), ALT (Siemens Healthcare, DF143), Glucose (Siemens Healthcare, DF40), HDL (Siemens Healthcare, DF48B), LDL (Siemens Healthcare, DF131), Cholesterol (Siemens Healthcare, DF27), LDH (Siemens Healthcare, DF54),

TG (Siemens Healthcare, DF69A) and FFA (FUJIFILM Wako Dignostics, NEFA-HR (2)). Insulin concentration was measured with an ELISA assay kit (Mouse Insulin ELISA Kit; Mercodia).

Clinical traits used for correlation of lipid signatures with metabolic phenotypes in C57BL/6J mice (Figure 2F) was obtained from our previous study (Gariani et al., 2016). BXD metabolic and clinical data can be obtained from (Williams et al., 2016) and also available on GeneNetwork (http://www.genenetwork.org).

Body composition (EchoMRI)

In addition to the body weight measurements taken each week and before each phenotyping experiment, body composition was recorded at 16, 23, and 25 weeks of age. To do so, each mouse was placed briefly in an EchoMRI (magnetic resonance imaging) machine (the 3-in-1, EchoMedical Systems), where lean and fat mass are recorded, along with total body weight, taking ~1 min per individual. Data was normalized to total body weight.

TAG measurement of BXD livers

For BXDs, 15 µl of the liver lipid extract (same extract as used for liver lipid MS measurement, see companion article (Jha et al., 2018) was used for TAG quantification using the Serum Triglyceride determination kit (Sigma-Aldrich), as per manufacturer's instructions. The organic solvent mix used for dissolving lipids for the MS (mixture of acetonitrile (ACN)/isopropyl alcohol (IPA)/water (H_2O), (65:30:5, v/v/v, 100 µL)) was used as a blank and for standard curves.

NAD^+ Measurement of C57BL/6J Livers

~20 mg of frozen liver samples were used for NAD^+ extraction in 10% perchloric acid and neutralized in 3 M K_2CO_3 on ice. After centrifugation, the supernatant was filtered and the internal standard (NAD-C13) was added and loaded onto a column (150 Å ~ 2.1 mm; Kinetex EVO C18,

100 Å). HPLC was run for 1 min at a flow rate of 300 mL/min with 100% buffer A (Methanol/H_2O, 80/20%, v/v). Then, a linear gradient to 100% buffer B [H_2O + 5 mM ammonium acetate] was performed (at 1 to 6 min). Buffer B (100%) was maintained for 3 min (at 6 to 9 min), and then a linear gradient back to 100% buffer A (at 9 to 13 min) began. Buffer A was then maintained at 100% until the end (at 13 to 18 min). NAD^+ eluted as a sharp peak at 3.3 min and was quantified on the basis of the peak area ratio between NAD^+ and the internal standard and normalized to tissue weight.

Lipidomics sample preparation and analysis

Internal standards (IS) used

For BXD liver samples we used Q6, PC(15:0/15:0), PS(17:0/17:0), PE(15:0/15:0), PA(17:0/17:0), PG(15:0/15:0), CL(56:0) and FA(15:0/15:0) as internal standards. For NAFLD signature validation experiments in mice (Figure 2), we used the standard mix SPLASH® Lipidomix® Mass Spec Standard | 330707, supplemented with Q6 and CL(56:0).

Extractions

Liver samples were weighed and homogenized using a Potter-Elvehjem tissue grinder in 1.5 mL homogenization buffer (8 M urea, 50 mM TEAB, 100 mM NaCl, 1 mM CaCl, protease and phosphatase inhibitors (Roche)). Protein concentration was determined by BCA and all samples were diluted to 8 mg/mL with homogenization buffer. Samples were aliquoted with each tube containing 1 mg of protein (125 µL) and flash frozen in liquid N_2 and stored at -80°C. Frozen aliquots of liver extracts were thawed on ice and processed for lipid extraction as previously described (Stefely et al., 2016). In brief, liver extracts were thawed on ice then internal standards were added (20 µL) and samples were vortexed (30 s). Chloroform/methanol (1:1, v/v, 1000 µL) was added and samples were vortexed (60 s). Subsequently, HCl (1 M, 200 µL) was added to induce phase separation, followed by vortexing (60 s) and centrifugation (3,000 g, 3 min, 4°C) to

complete phase separation. 550 µL of the organic phase was dried under Ar_2(g). The organic residue was reconstituted in ACN/IPA/H_2O (65:30:5, v/v/v, 100 µL) by vortexing (60 s) and transferred to a glass vial for LC-MS analysis. Samples were stored at -80°C until further use. LC–MS analysis was performed on an Ascentis Express C18 column held at 50°C (150 mm × 2.1 mm × 2.7 µm particle size; Supelco) using an Accela LC Pump (500 µL/min flow rate; Thermo). Mobile phase A consisted of 10 mM ammonium acetate in ACN/H_2O (70:30, v/v) containing 250 µL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. 10 µL of sample were injected by an HTC PAL autosampler (Thermo). Initially, mobile phase B was held at 40% for 30 s and then increased to 50% over an additional 30 s. It was then increased to 55% over 4 min after which, it was increased to 99% over 6 min and held there for 3 min. Prior to the next injection, the column was reequilibrated for 2 min. The LC system was coupled to a Q Exactive mass spectrometer (Build 2.3 SP2) by a HESI II heated ESI source kept at 325°C (Thermo). The inlet capillary was kept at 320°C, sheath gas was set to 35 units, auxiliary gas to 15 units, and the spray voltage was set to 3,000 V in negative mode and 4,000 V in positive mode.

Several scan functions, including targeted and untargeted, were used to ensure optimal data acquisition for each lipid class. For fatty acids, selected ion monitoring (SIM) scans were taken from 0-3 min. MS^1 data was acquired in negative mode for 220-600 m/z at a resolving power of 17,500 and an AGC target of 1×10^5. For phospholipids, diacylglycerols, and CoQ, parallel reaction monitoring (PRM) was used. The instrument was run in negative mode with a resolving power of 17,500, an AGC target of 2×10^5, a maximum injection time of 75 ms, and isolation window of 1.2 Th. Scans targeting each species were scheduled between 0-10.3 min based on previously determined retention times. For triglycerides, a separate set of runs was done where in addition to the targeted method described above, the MS was operated in positive mode from

10.1-13 min with resolving power set at 17,500 and the AGC target set to 5×10^5. Ions from 750-1,100 m/z were isolated (Top 2) and fragmented.

Measurement, normalization and quality control

The selection of lipid species for measurement was based on their abundance, stability, polarity and ease of ionization. For the BXDs, peaks were automatically integrated using TraceFinder software (Thermo) and integrations were checked manually. NR mice were processed as described in our companion article (Jha et al., 2018). Lipids were normalized in three different ways - to internal standards (of each class), to total lipids in each sample and to all lipids in each class. Basic quality check and QTL analysis was performed from all the datasets, however the dataset normalized to total lipids was used for all the analyses and figures shown in the manuscript due to the overall low relative standard deviation in this dataset (data not shown). Additionally, normalization to total lipids has two major advantages over the other normalization methods; 1) All lipids measured did not have a true internal standard, 2) for lipid classes that have few lipid species measured, normalizing to class will be largely driven by one or two highly abundant lipids. Quality assessment of the MS measurements was performed by comparing the reproducibility of the technical and extraction replicates (see Figures S1A–S1C). Note: lipid pairs marked with "_1" and "_2" (TAGs 54:5, 54:6, 56:7; PI(Dha_S) and CL(LLOPo) indicate two isobaric peaks. The TAGs are isobaric peaks with different fatty acid compositions while the PI and CL are isobaric because they have the same fatty acid composition but are likely ordered differently to cause chromatographic separation.

Quantification and statistical analysis

Bioinformatic and genetic analyses

Data normality for each lipid species was checked by the Shapiro-Wilk test in R, with a $W \geq 0.90$ considered normal distribution. Correlations are Pearson's *r* or Spearman's *rho*, as

indicated. Student's t-test was used for two group comparisons in normal data of equal variances, and Welch's t-test otherwise. Heatmaps were generated using the "heatmap.2" function in R. Unweighted correlation network graphs were performed using Spearman correlation, keeping all edges with P values less than 1e-4, 1e-03 or 0.05 (indicated in the figure) in R using the custom package imsbInfer, currently on Github (https://github.com/wolski/imsbInfer). GO and KEGG pathway enrichment analysis for was performed using the R package "clusterProfiler" (Yu et al., 2012) (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html). Enriched pathways after Benjamini-Hochberg correction ($p < 0.05$) are shown in the figures. For Circos plot (Figure 3) data were pre-processed in R followed by plot construction using web based (http://circos.ca/) (Krzywinski et al., 2009) and then modified in Adobe Illustrator.

QTL calculations were performed using the R/qtl (v 1.39-5) package (Broman et al., 2003) on the $\log_2$ transformed data. The BXD genotype used for QTL calculations is provided in the Table S8. Parametric QTL calculation was performed for normally distributed lipids and non-parametric for those that were not normally distributed. QTLs with logarithm of the odds ratio (LOD) score >2.5 and p-value <0.40 were used for all the analysis, which includes both significant (p-value < 0.05) and suggestive QTLs (p-value between 0.05 and 0.40) at genome-wide significant threshold, computed by permutation analysis. [Genome-wide p-values of 0.63 correspond approximately to a local p-value of 0.05, i.e. which is significant in case of prior knowledge used to search for a QTL at that specific location]. All significant and suggestive threshold lines in the paper represent genome-wide p-value of 0.05 and 0.63 respectively.

All graphs and analyses were performed either in R or GraphPad. For R, standard R plotting packages included in gplots or ggplot2—e.g., stripchart, plotCI, and barplot2 were used. Final figures were prepared with Adobe Illustrator.

QTL candidate gene retrieval

To perform QTL candidate gene retrieval biomaRt was used in R to obtain lists of genes located within each QTL region (+/- 5 Mb around the mapped SNP). Gene lists were imported into MATLAB for subsequent parallel filtering of candidate genes as follows: (i) genes with non-synonymous SNPs, insertion/deletion/splice site mutations, or high impact non-coding SNPs in the BXDs (Wang et al., 2016); (ii) genes under the lQTLs with *cis* e- and/or p-QTLs; (iii) genes/ proteins with significant correlation ($p<0.05$) with the lipid itself; and (iv) genes with variable transcript expression (standard deviation >0.25) across BXD strains. Pearson's correlation was used to correlate transcript and protein abundances with log_2-transformed lipid levels. GeneNetwork was used to obtain liver gene transcript and peptide values for each BXD strain for mRNA standard deviation and mRNA/protein correlation calculations. Genes passing two or more of the above filtering criteria are provided in Table S5 and a detailed list of each lQTL candidate gene with the information on each filtering criteria is provided in Table S6.

BXD lQTL and human GWAS genes overlap

Human GWAS genes (having p-value <1e-07) were retrieved from the database GWASdb2 (http://jjwanglab.org/gwasdb) (Li et al., 2016) and complemented with the data from the GWAS Catalog (https://www.ebi.ac.uk/gwas/) (MacArthur et al., 2017). The categories of GWAS gene sets retrieved from these databases included "Abnormality of lipid metabolism" and "fatty liver disease". Additionally, published papers reporting the relevant GWAS studies, not included in the above-mentioned databases (in particular, the references indicated by their PMID in Table S7) were manually mined to retrieve the candidate genes having p-value <1e-07. Taken together, the compiled list comprised of 494 genes (Table S7). Only those lQTL genes (± 5 Mb on either side of the peak QTL), which passed at least 2 filtering criteria (as shown in Figure 4B) were matched for any evidence of them being associated with abnormal lipid metabolism in human GWAS (i.e. matched for their presence in 494 human GWAS gene list).

Weighted gene correlation network analysis

Weighted gene correlation network analysis (WGCNA) (Figures 1 and S2) was performed as described (Langfelder and Horvath, 2008) by using the WGCNA R software package (v1.51). To construct the weighted lipid coexpression network, we calculated a correlation matrix containing all pairwise Pearson's correlations between all pairs of lipids across all BXD strains for both CD and HFD. We defined a "signed hybrid" network in which the adjacency takes values between 0 and 1 when the correlation is positive and 0 if the correlation is negative. A power of 27 was chosen for both CD and HFD datasets. We selected the minimum power in which both datasets followed the Scale-Free Topology Criterion (model fitting index $R^2 > 0.8$) and showed a similar connectivity. The selection of a high power (threshold) has the effect of suppressing low correlations that may be due to noise, penalize weaker connections and strengthen stronger connections. The result is a network adjacency that is zero for negatively correlated lipids and is positive for positively correlated lipids. Adjacency of weakly correlated lipids is nearly zero due to the power transformation. Next, the lipids were hierarchically clustered using the distance measure and modules were determined by choosing a height cutoff for the resulting dendrogram by using a dynamic tree-cutting algorithm, selecting a minimum module size of 5. Modules with a correlation higher than 0.75 were merged. The resulting lipid modules were assigned color names and identified using the eigenvector of each module, named as module eigenlipid. Module eigenlipid (ME) is defined as the first principal component of the standardized expression profiles and can be considered the best summary of the standardized module expression data. Each module is represented by different colors; lipids not grouped in any module (34% in CD and 42% in HFD) were represented in grey color. By and large, modules were dominated by lipids from the same class (Figure S2A). Modules containing lipids from the same class exhibited high adjacency between them in both diets (eg. CL modules). Correspondence analysis between CD and HFD was performed by calculating the overlaps of each pair of CD-HFD modules and analyzed using the Fisher's exact test (Figure S2B). From the 8 modules identified in both diets, 3 of them showed total correspondence between CD and HFD: DAG (black), FFA (red) and TAG (brown) modules;

while the 3 CL modules (green, blue and yellow) showed a very high correspondence between CD and HFD (Figure S2B). Module-trait relationships (Figure 1D) were calculated by Pearson's correlation between MEs and selected metabolic phenotypes in order to identify modules related to metabolic functions. For module pathway association (Figure 1E), KEGG enrichment analysis was performed for all positively and negatively correlated proteins ($p < 0.05$) separately, with the modules. Module QTL (modQTL) was calculated from the values of the MEs as phenotype traits using the R package R/qtl (v 1.39-5) (Broman et al., 2003) using the same methods and criteria as for lQTLs.

Enrichment analysis of proteins and transcripts with liver TAG concentration

KEGG enrichment analysis was performed for all proteins and transcripts that correlated with total liverTAG levels (Figure 6B). For liver proteomics correlation: 333 proteins in CD and 74 in HFD out of 2,622 measured proteins (by SWATH) significantly correlated with liver total TAG concentration. For liver transcriptomics correlation: 1,752 transcripts in CD and 879 transcripts in HFD out of 35,556 transcripts measured (using Affy Mouse Gene 1.0ST) significantly correlated with total liverTAG concentration. Benjamini-Hochberg corrected ($p < 0.05$) enriched pathways are shown in the figure. No significant enrichment was observed in HFD cohorts.

Data and Software Availability

Raw MS data files are available through the CHORUS project data repository (Project ID 1432, Experiment ID 3217 and 3218). Additionally, normalized MS data is deposited in GeneNetwork (http://www.genenetwork.org) as a resource for public use. To access and analyze the data in GeneNetwork, choose Mouse (mm10)" for "Species, *"BXD"* for "Group", "*Phenotypes*" for "Type", "*BXD Published Phenotype*" for "Data Set" and enter ***"LiverLipidomics"*** for "Get Any". Normalized MS data (normalized to total lipids) is provided in Table S1. Lipid QTLs are provided in Table S5. lQTL genes passing two of the four filtering

criteria (as shown in Figure 4B) are provided in Table S5 and the exhaustive list for the same is provided in Table S6. BXD genotype data used for QTL calculation is provided in Table S8.

Author contributions

PJ, DJP, and JA conceived and designed the project. MTM acquired and analyzed the MS data. MBS performed quality assessment. PJ and PMQ performed genetics and bioinformatics analysis. PJ, PMQ and RG performed lQTL candidate genes retrieval. LD did the circos plot. PJ and EGW managed the BXD experiments and KG performed the NR experiment. JJC and DJP supervised the lipidomics measurements. AD and AU helped in standardization of lipid measurements. PJ and JA wrote and DJP, EGW, MTM, PMQ and RG edited the manuscript.

Acknowledgements

We thank the Center of Phenogenomics (CPG, EPFL) for performing BXD phenotyping. JA is supported by grants from the École Polytechnique Fédérale de Lausanne, the Swiss National Science Foundation (31003A-140780), the AgingX program of the Swiss Initiative for Systems Biology (51RTP0-151019), and the NIH (R01AG043930). DJP is supported by a grant from NIH (R01GM115591) for lipidomics analysis. MTM and EGW were supported by NIH fellowships (T32GM007215 and F32GM119190, respectively). JJC is supported by NIH grants R35GM118110 and P41GM108538.

Figures and tables

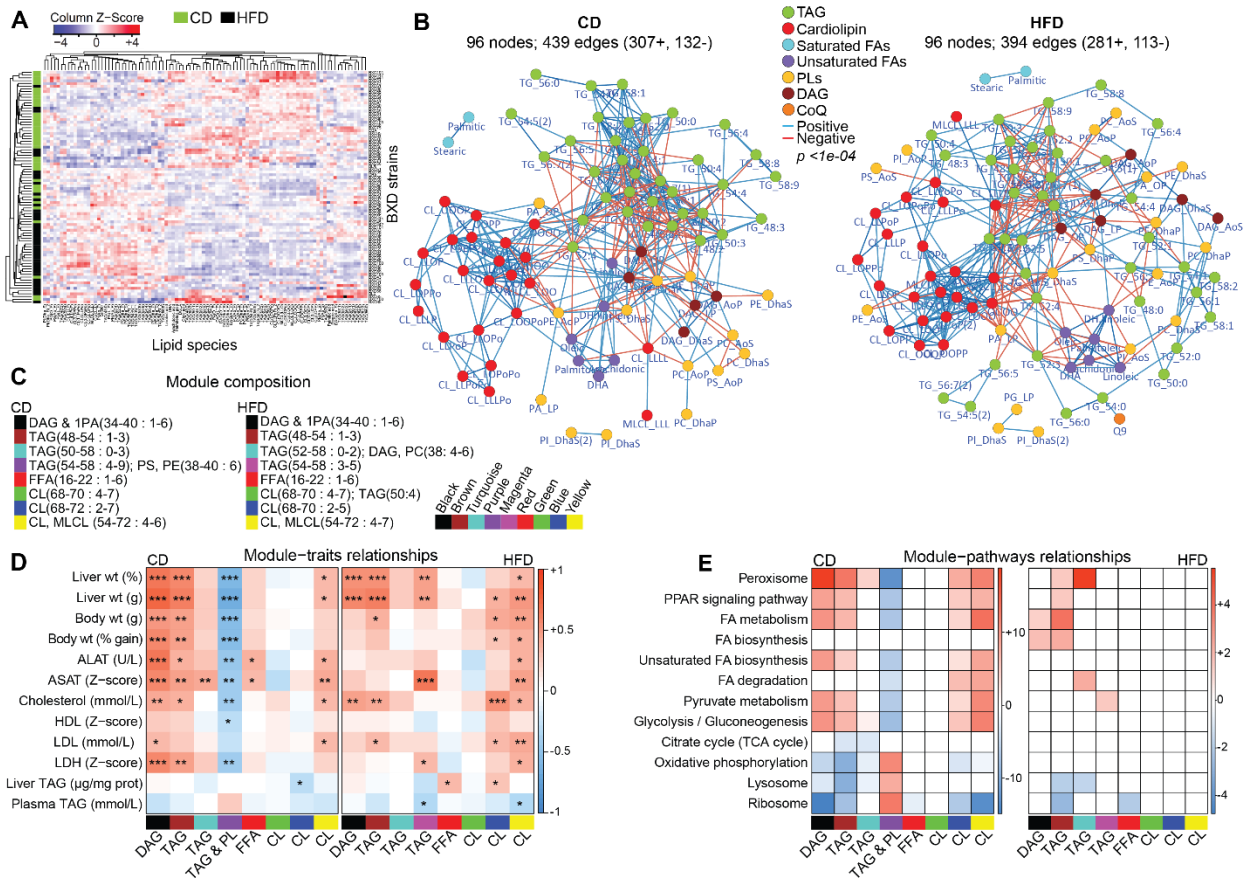


Figure 1. Liver lipid species profile and their association with physiological and molecular traits.

(A) Heatmap analysis with unsupervised hierarchical clustering of 96 lipid species for each BXD cohort shows mixed dietary and genetic impact.

(B) Spearman correlation network ($p < 1e\text{-}04$) of all lipid species measured in CD and HFD. Lipid species are color coded as seven major lipid classes. The side chain FA composition of lipids has been abbreviated (O, oleic acid; P, palmitic acid; Po, palmitoleic acid; S, stearic acid; L, linoleic acid; Dha, docosahexaenoic acid). Refer to table S1 for abbreviation and composition. (C) Legend of the module composition, indicating the range of total number of carbons and degree of unsaturation.

(D) Lipid module-clinical trait correlation. Each cell is color coded by the Pearson's correlation coefficient according to the legend color on the right. The stars in the cells represent the p value of the correlation (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

(E) Module and its corresponding KEGG enriched pathway correlation. Red and blue cells represent the enriched pathways with the positively (scale bar: $\log_{10}$ p value) and negatively (scale bar: $-\log_{10}$ p value) correlated proteins, respectively. Lipid classes hereafter are abbreviated as follows: TAG/TG, triacylglycerol; DAG, diacylglycerol; FFA, free-fatty acid; PL, phospholipid (PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PS, phosphatidylserine; CL, cardiolipin; MLCL, monolysocardiolipin.

See also Figures S1 and S2, Tables S1 and S2.

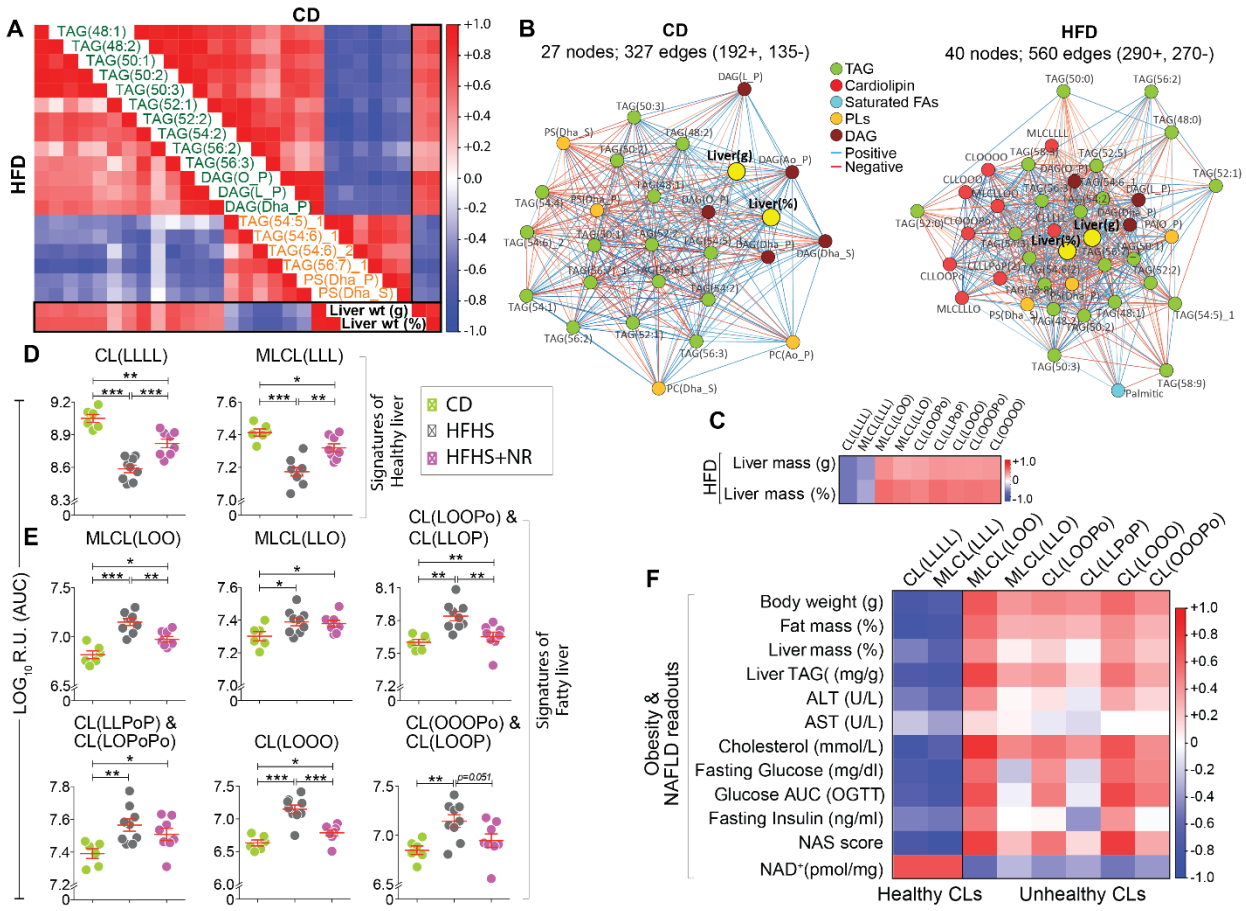


Figure 2. Identification of cardiolipin signatures of healthy and fatty liver.

(A) Correlation diagram (corrgram) showing diet-independent association of lipid species with liver mass. Lipid species with Spearman's correlation p value <0.05 with liver mass (both normalized to body weight [%] and unnormalized [weight in grams] in both CD and HFD were selected.

(B) Spearman correlation network of diet specific significant correlation of lipid species with liver mass in CD (left) and HFD (right).

(C) Corrgram of CLs that significantly correlate with liver mass in HFD.

(D-F) C57BL/6J mice were fed with CD, high-fat high-sucrose (HFHS) diet for 18 weeks or nicotinamide riboside (NR) supplemented HFHS diet, 9 weeks after the start of the HFHS diet (HFHS+NR). Levels of healthy (D) and unhealthy (E) CL species in livers of the three cohorts.

Note that the CLs—CL(LOOPo), CL(LLPoP) and CL(OOOP)—are shown in the figure with an additional CL species because the two are isobaric and inseparable chromatographically.

(F) Corrgram showing negative correlation of obesity and NAFLD traits with healthy CL species and positive correlation with unhealthy CL species.

*For (D) and (E), data are represented as means ± SEM. *p < 0.05; **p < 0.001; ***p < 0.0001.*

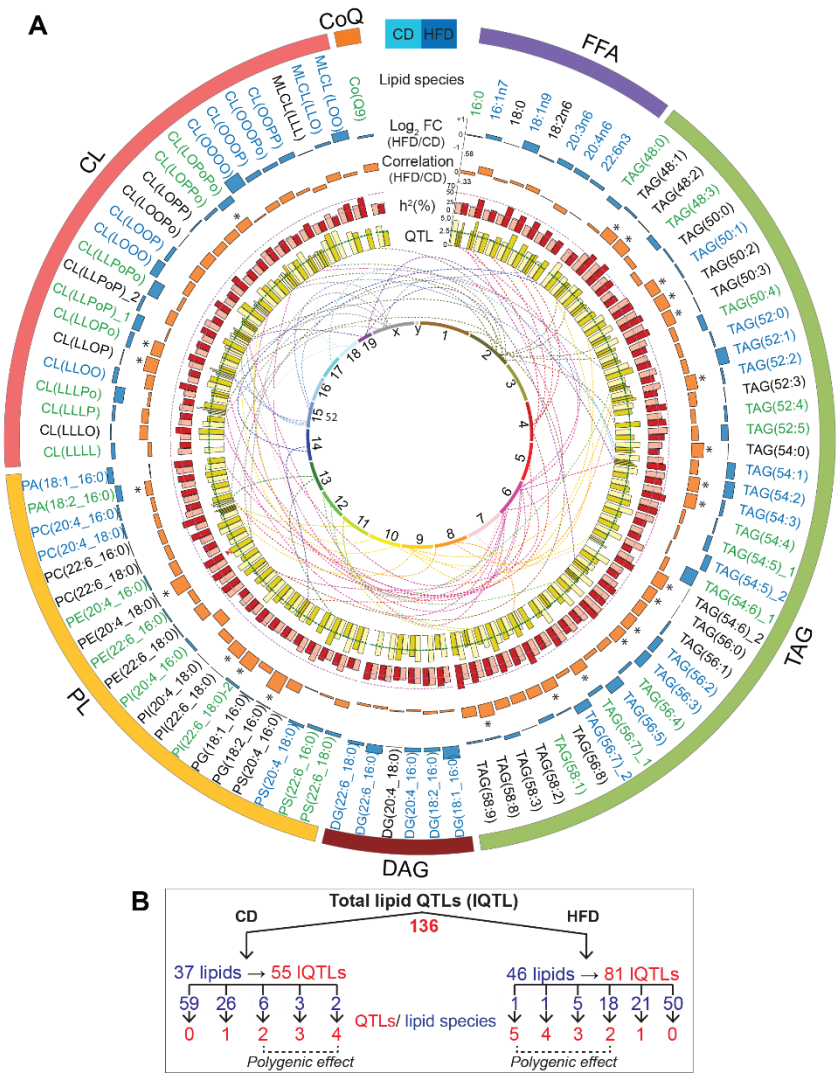


Figure 3. Genetics of lipid species.

(A) Circos plot of all lipids measured. Blue bars in the outermost ring represent the $\log_2$ fold change (HFD versus CD) of the lipids. Lipids increased in CD or HFD are shown in green and blue font, respectively. Orange bars represent the correlation of lipids between CD and HFD. Significant correlations (adjusted p value <0.05) are represented by asterisks. Red bars represent lipid h^2 in CD (light red) and HFD (dark red). The inner ring of yellow bars represents the strength of lQTLs in CD (light yellow) and HFD (dark yellow). Number of bars per lipid is equivalent to the number of lQTLs. The lines between the two innermost rings stem from the peak lQTL bar (with LOD > 3) and terminates on their approximate chromosomal position of the innermost ring. Lipid

pairs marked with "_1" and "_2" (TAGs 54:5, 54:6, 56:7; PI(22:8_18:0) and CL(LLPoP) indicate two isobaric peaks.

(B) Schematic representation of the lQTLs. A total of 136 lQTLs (55 CD and 81 HFD) were mapped from 37/46 lipids in CD/HFD. The number of QTLs (red font) per lipid species (blue font) is indicated.

See also Figure S3 and Tables S3-S5.

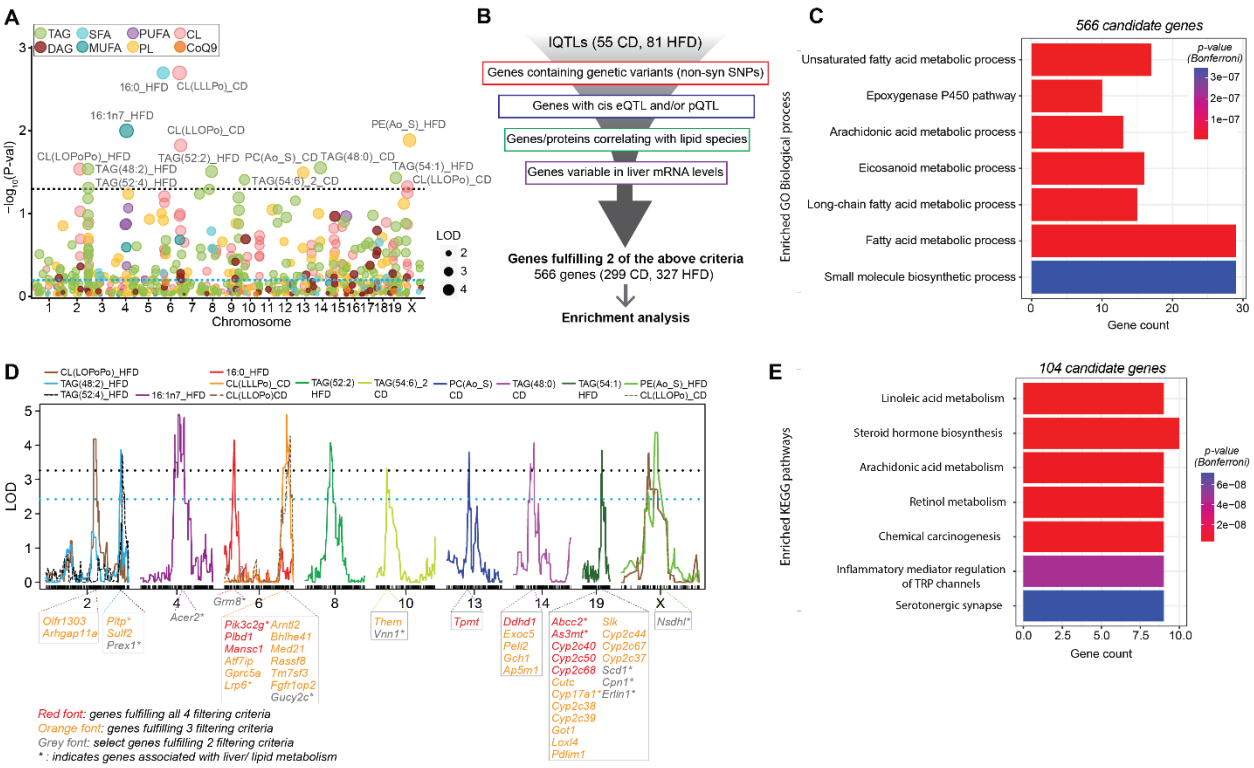


Figure 4. Genetic assessment of lQTLs.

(A) Manhattan plot of lipid species in CD and HFD. Names of the lipid species with genome-wide p value < 0.05 are indicated.

(B) All genes under the lQTLs (± 5 Mb from the peak) were filtered through four independent pipelines as indicated. Genes fulfilling two or more of the filtering criteria were analyzed for enrichment of GO biological process (BP).

(C) Enriched GO BP from 566 filtered lQTL genes.

(D) lQTL position of the top 14 QTLs (p < 0.05). The candidate genes fulfilling four (red font), three (orange font) and two (gray font) of the four filtering criteria are indicated below each lQTL peak. Genes indicated with asterisks are associated with metabolic phenotypes.

(E) KEGG enrichment analysis of genes under the top 14 lQTLs passing 2 of the filtering criteria.

For A and D, blue and black dotted lines represent suggestive and significant QTL threshold respectively.

See also Tables S5 and S6.

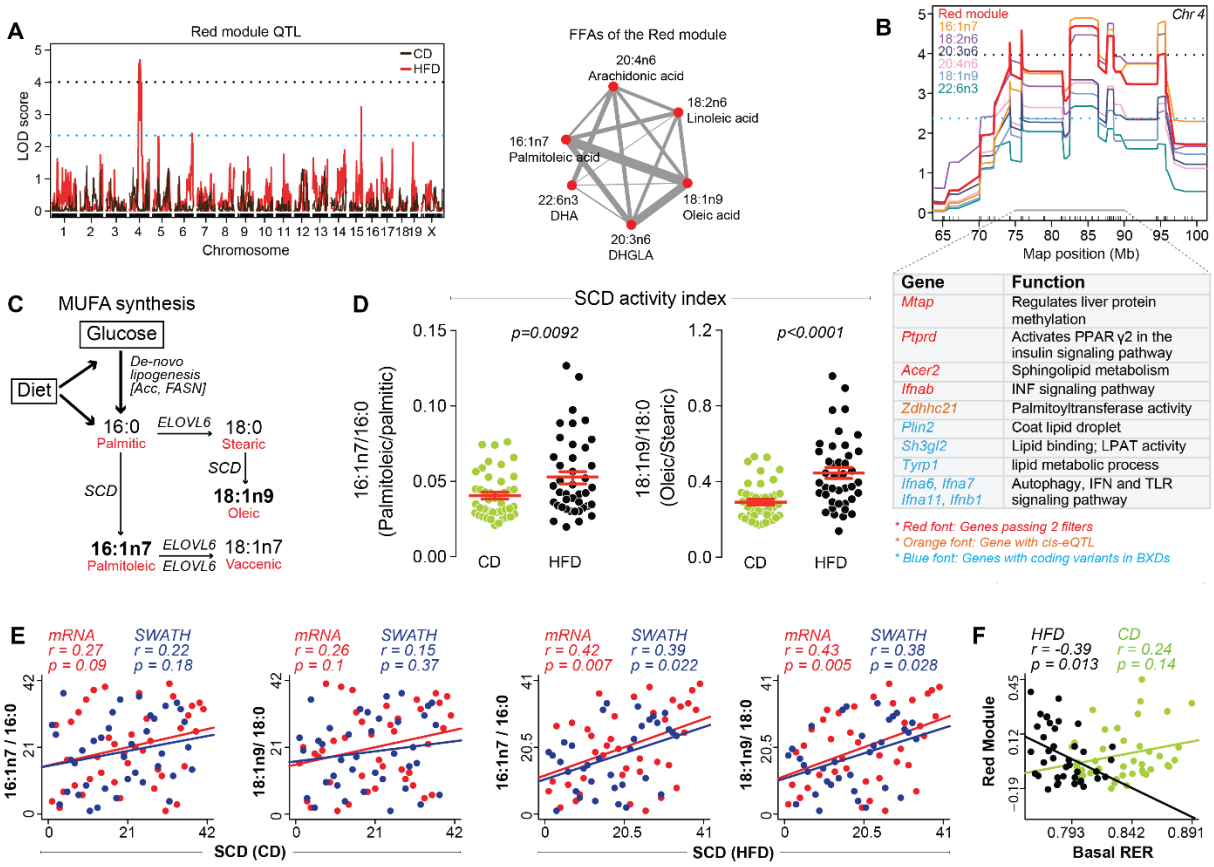


Figure 5. FFA module identifies a genetic hotspot locus associated with fatty acid metabolism and signaling.

(A) Red module QTL showing significant peak on chromosome 4 (left) in HFD and the weighted correlation network of the FFAs (right) in the red module.

(B) Hotspot region on chromosome 4 showing the overlapping QTL of the red module and the individual FFAs of the module. modQTL genes involved in fatty acid metabolic processes and signaling are indicated along with their biological function.

(C) Schematic representation of monounsaturated fatty acid (MUFA) synthesis.

(D) Stearoyl-CoA Desaturase (SCD) activity index represented as ratios of MUFA to saturated fatty acid (SFA). Data are represented as means ± SEM

(F) Spearman correlation between hepatic expression of SCD transcript and protein (SWATH) with SCD activity index.

(E) Pearson's correlation between the basal RER and the red module.

For (A) and (B), blue and black dotted lines represent suggestive and significant QTL threshold, respectively.

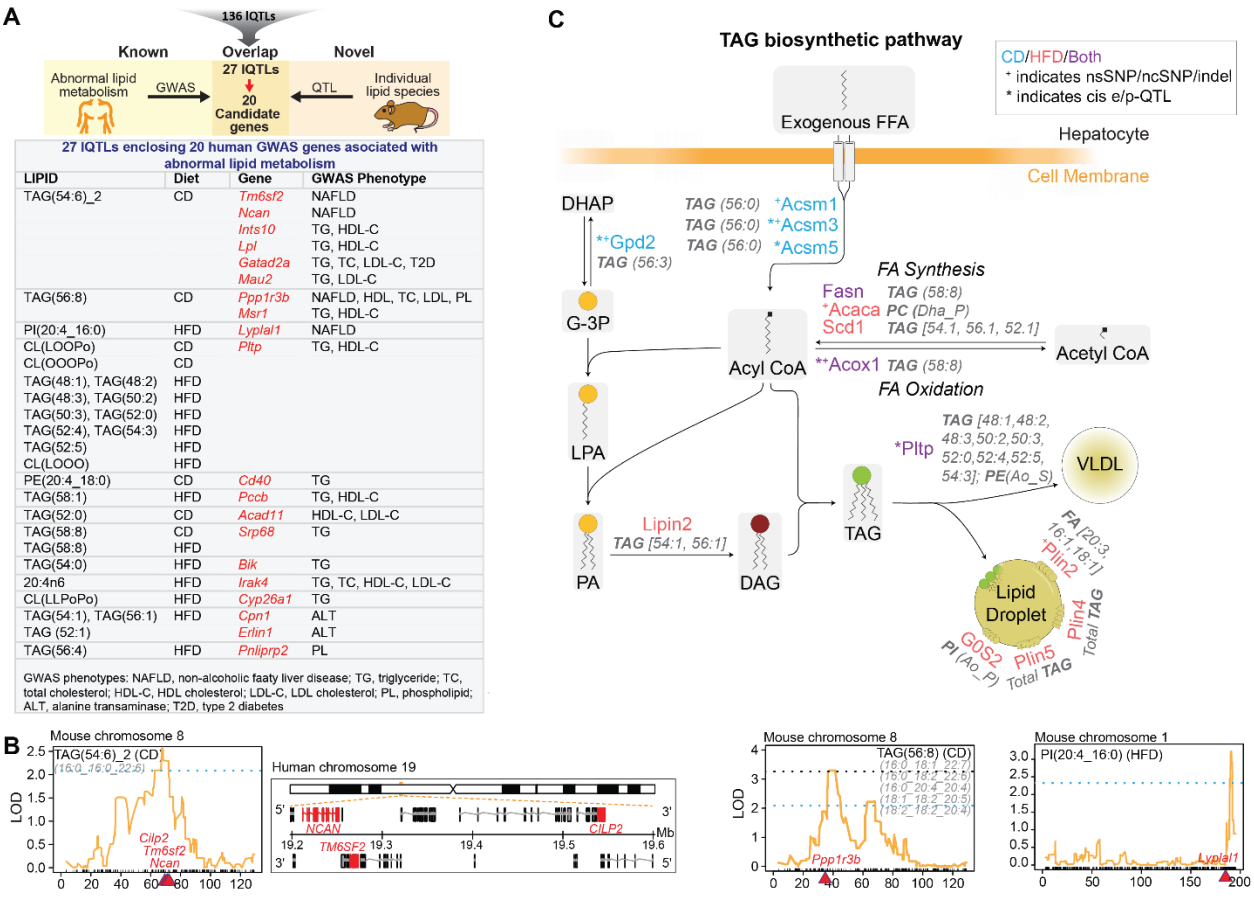


Figure 6. lQTL genes associated with abnormal lipid metabolism in human GWAS and with the TAG biosynthetic pathway.

(A) lQTL genes fulfilling two or more filtering criteria as shown in Figure 4B were screened for any known association with abnormal lipid metabolism in human GWAS. The screening identified 20 GWAS genes from 27 lQTLs (top). Each box of the table represents the lQTL(s) for the indicated gene(s) and the GWAS phenotype associated with those genes.

(B) QTL position of the three lipid species (TAG(54:6)_2, TAG(56:8), and PI(20:4_16:0)) which map to the indicated genes (in red) implicated in hepatic steatosis in human GWAS studies. The loci of *Cilp2*, *Tm6sf2* and *Ncan* is syntenic in mice (left) as well as in humans (middle). The genomic location of the genes is shown in red in the positive strand for *NCAN* and *CILP2* and in

the negative strand for *TM6SF2.* Blue and black dotted lines represent suggestive and significant QTL threshold, respectively.

(C) Schematic representation of TAG biosynthetic pathways showing only those candidate genes that are under the lQTLs and fulfill one or more of our filtering criteria. lQTLs are represented in gray. Genes under CD lQTLs are in blue, those under HFD lQTLs in red, and those under both CD and HFD lQTLs in purple.

See also Figures S3C and S4; Table S7.

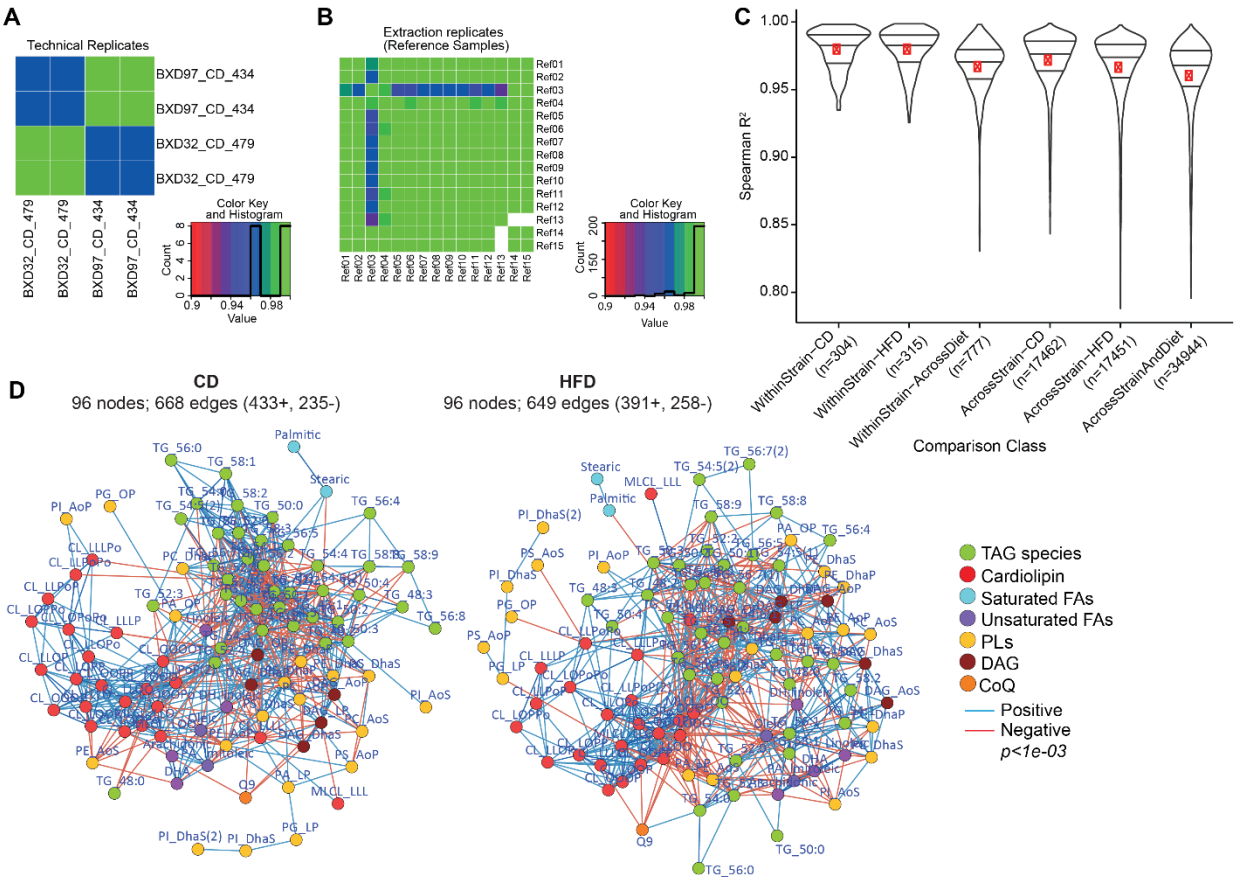


Figure S1. Quality assessment of MS measurements and lipid species interaction in the two diets. Related to Figure 1 and Table S1.

(A) Heatmap of correlation between technical replicates (from two separate runs) of two liver samples showing high spearman correlation between the replicates vs. between the strains.

(B) Heatmap of extraction replicates of the same liver sample extracted and run in 14 batches, equivalent to the 14 batches of the total BXD samples. Ref 03 was discarded due to extraction error.

(C) Pairwise correlation to assess variation across different groups indicated along the X-axis. Red dot represents mean; the three lines represent median, upper and lower quartile.

(D) Spearman correlation network of all lipid species measured in CD and HFD ($p < 1e\text{-}03$ versus $p < 1e\text{-}04$ in Fig 1B). Lipid species are color coded as 7 major lipid classes. Blue edges represent

positive correlations and red, negative. The FFAs abbreviations are as follows: O, oleic; P, palmitic; Po, palmitoleic; S, stearic; L, linoleic and Dha, docosahexaenoic acid. The short abbreviation for side chain lipid composition has been used in the figure (refer to table S1).

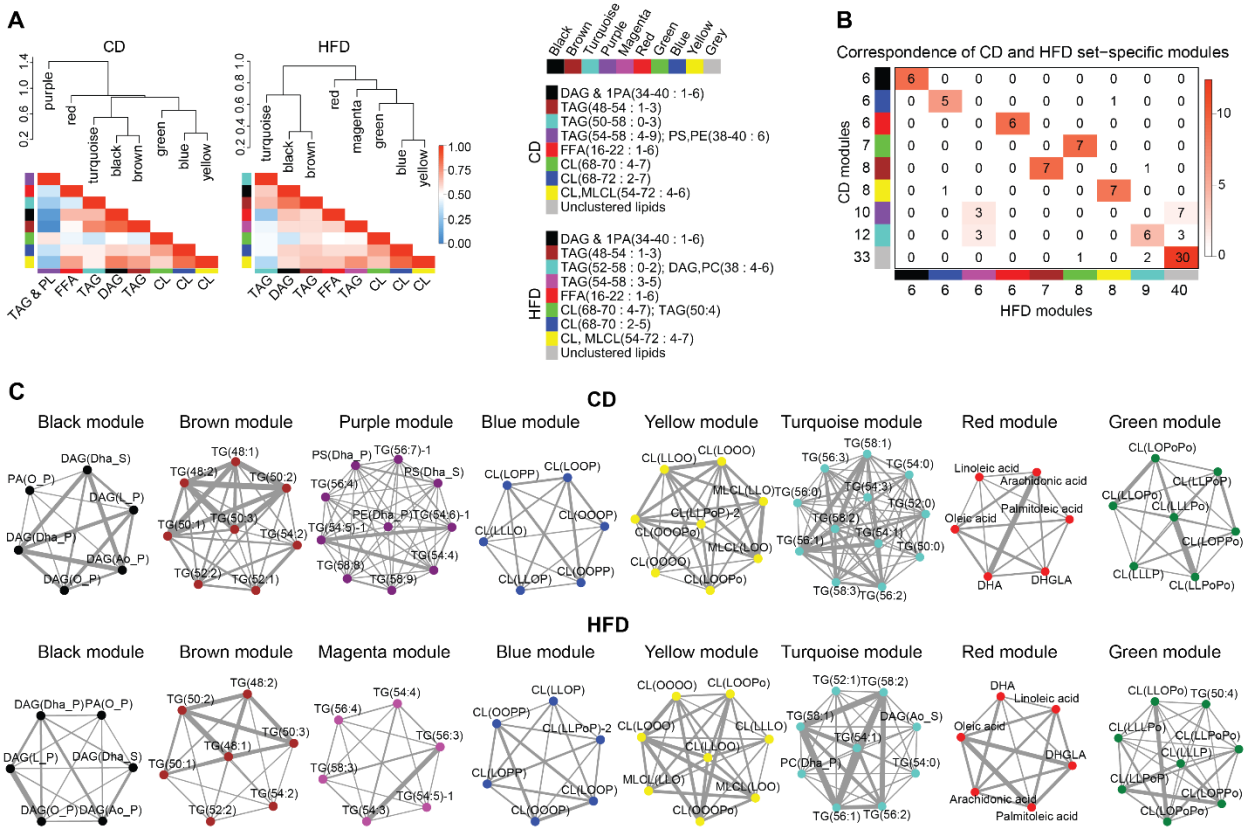


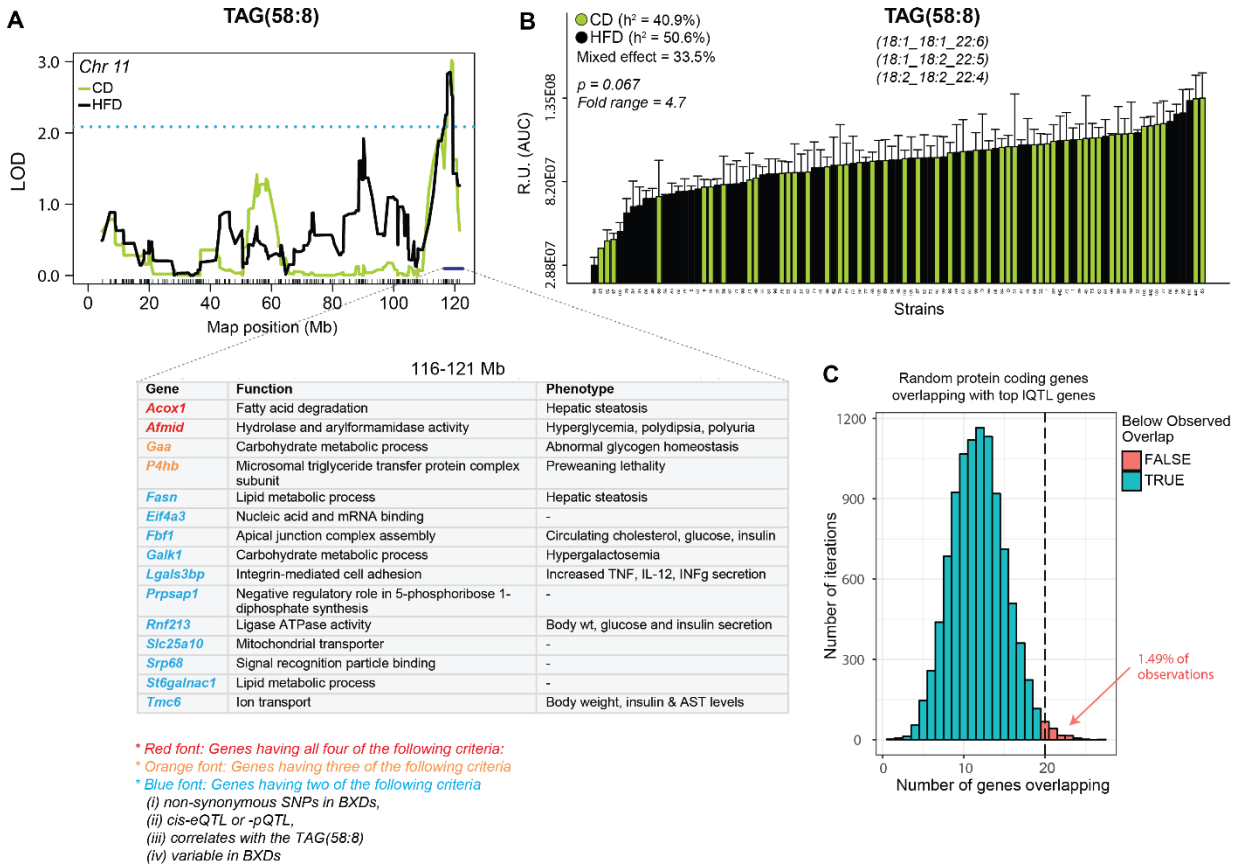
Figure S2. Weighted correlation network analysis (WGCNA) of liver lipids. Related to Figure 1 and Table S2.

(A) Hierarchical clustering dendogram and heatmap analysis of modules defined in the WGCNA. Dendograms are generated using modules eigenlipid (ME). Heatmap plots represent the ME adjacencies, where each row and column corresponds to each module, represented by a color (names indicated on top) and labeled by the main lipid class contained therein (bottom). Red indicates high ME adjacency (positive correlation) and blue low ME adjacency (negative correlation) as shown in the color legend (right). The legend on the right represents the range of the carbons and the degree of unsaturation of the side chains in the module.

(B) Color table representing the correspondence of CD- and HFDspecific modules described in panel A. Numbers on the side/below the colors indicate the number of lipid species in each module. Numbers in the table represent the number of lipid species that are common in both the

CD and HFD modules. Grey color represents lipids that were not assigned to any module (background lipids). Color legend (white-red) indicates the negative base-10 logarithmic of the p-value obtained with the Fisher test.

(C) Module networks in CD and HFD. Lipid nodes in each module are represented by their module color. The thickness of the edges in the network represents the strength of correlation between the lipids.



Gene	Function	Phenotype
Acox1	Fatty acid degradation	Hepatic steatosis
Afmid	Hydrolase and arylformamidase activity	Hyperglycemia, polydipsia, polyuria
Gaa	Carbohydrate metabolic process	Abnormal glycogen homeostasis
P4hb	Microsomal triglyceride transfer protein complex subunit	Preweaning lethality
Fasn	Lipid metabolic process	Hepatic steatosis
Eif4a3	Nucleic acid and mRNA binding	-
Fbf1	Apical junction complex assembly	Circulating cholesterol, glucose, insulin
Galk1	Carbohydrate metabolic process	Hypergalactosemia
Lgals3bp	Integrin-mediated cell adhesion	Increased TNF, IL-12, INFg secretion
Prpsap1	Negative regulatory role in 5-phosphoribose 1-diphosphate synthesis	-
Rnf213	Ligase ATPase activity	Body wt, glucose and insulin secretion
Slc25a10	Mitochondrial transporter	-
Srp68	Signal recognition particle binding	-
St6galnac1	Lipid metabolic process	-
Tmc6	Ion transport	Body weight, insulin & AST levels

* *Red font: Genes having all four of the following criteria:*
* *Orange font: Genes having three of the following criteria*
* *Blue font: Genes having two of the following criteria*
  (i) non-synonymous SNPs in BXDs,
  (ii) cis-eQTL or -pQTL,
  (iii) correlates with the TAG(58:8)
  (iv) variable in BXDs

Figure S3. CD-HFD conserved lQTL and permutation test to check BXD lQTL and human GWAS genes overlap by chance. Related to Figures 3, 6A, Tables S5 and S6.

(A) Of the 96 lipid species measured, only TAG (58:8) had the same QTL locus in both diets. Genes under the QTL having lipid-associated function are indicated.

(B) Levels of TAG (58:8) do not show significant difference across diets and has high h2 in both dietary cohorts.

(C) Permutation test (with 10,000 permutations) analyzing the gene set overlap size between lQTL genes from the BXDs and a random set of 494 human genes.

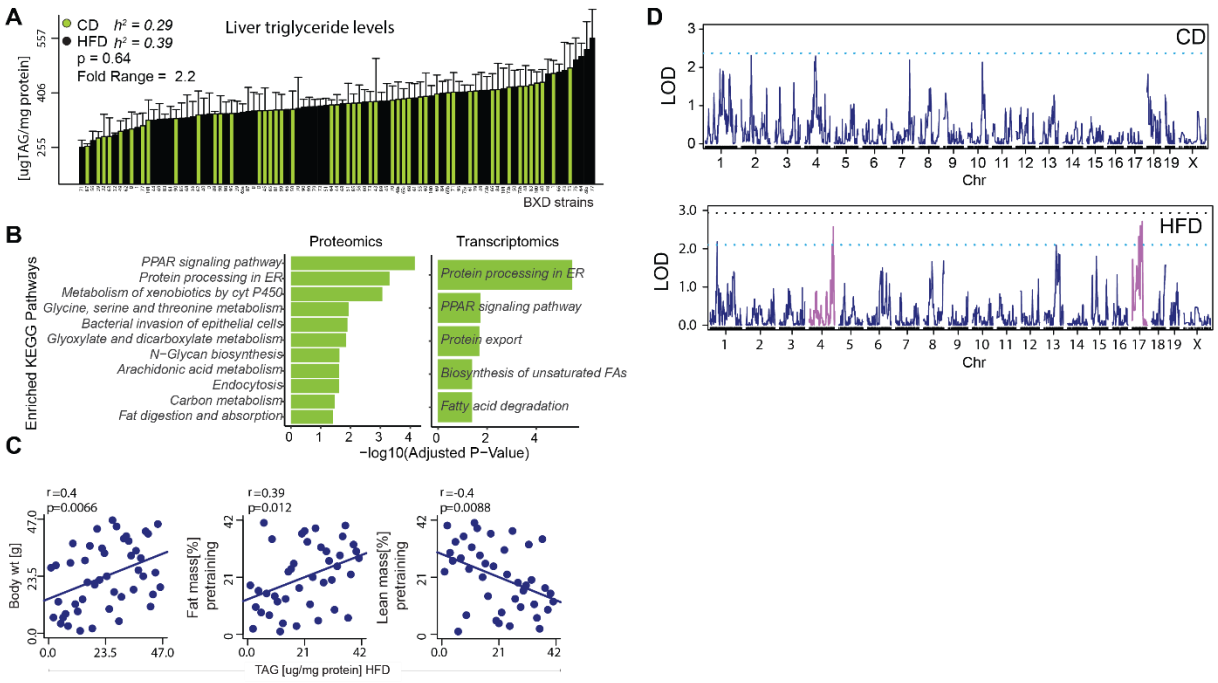


Figure S4. TAG variation in BXDs, its association with KEGG pathways, phenotypes and QTL mapping. Related to Figure 6.

(A) Hepatic TAG concentration in the BXD strains.

(B) KEGG enrichment analysis for all proteins (333) and transcripts (1,752) that correlated with total liver TAG levels in CD.

(C) TAG content in HFD correlates positively with body weight and fat mass and negatively with lean mass.

(D) Mapping of hepatic TAG content does not show any significant or suggestive QTL in CD, whereas 2 suggestive QTLs at chromosome 4 and 17 were detected in the HFD cohort. Blue and black dotted lines represent suggestive and significant QTL threshold respectively.

*All supplemental tables are available with the published manuscript at DOI:

10.1016/j.cels.2018.05.016

References

Andreux, P.A., Williams, E.G., Koutnikova, H., Houtkooper, R.H., Champy, M.F., Henry, H., Schoonjans, K., Williams, R.W., and Auwerx, J. (2012). Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. Cell *150*, 1287-1299.

Anstee, Q.M., and Day, C.P. (2013). The genetics of NAFLD. Nature reviews. Gastroenterology & hepatology *10*, 645-655.

Belknap, J.K. (1998). Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behavior genetics *28*, 29-38.

Bigaud, E., and Corrales, F.J. (2016). Methylthioadenosine (MTA) Regulates Liver Cells Proteome and Methylproteome: Implications in Liver Biology and Disease. Molecular & cellular proteomics : MCP *15*, 1498-1510.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889-890.

Chicco, A.J., and Sparagna, G.C. (2007). Role of cardiolipin alterations in mitochondrial dysfunction and disease. American journal of physiology. Cell physiology *292*, C33-44.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nature reviews. Genetics *15*, 34-48.

Claypool, S.M., and Koehler, C.M. (2012). The complexity of cardiolipin in health and disease. Trends in biochemical sciences *37*, 32-41.

Cole, L.K., Mejia, E.M., Vandel, M., Sparagna, G.C., Claypool, S.M., Dyck-Chan, L., Klein, J., and Hatch, G.M. (2016). Impaired Cardiolipin Biosynthesis Prevents Hepatic Steatosis and Diet-Induced Obesity. Diabetes *65*, 3289-3300.

Franks, P.W., and McCarthy, M.I. (2016). Exposing the exposures responsible for type 2 diabetes and obesity. Science *354*, 69-73.

Gariani, K., Menzies, K.J., Ryu, D., Wegner, C.J., Wang, X., Ropelle, E.R., Moullan, N., Zhang, H., Perino, A., Lemos, V., et al. (2016). Eliciting the mitochondrial unfolded protein response by nicotinamide adenine dinucleotide repletion reverses fatty liver disease in mice. Hepatology *63*, 1190-1204.

Gronert, K., Kantarci, A., Levy, B.D., Clish, C.B., Odparlik, S., Hasturk, H., Badwey, J.A., Colgan, S.P., Van Dyke, T.E., and Serhan, C.N. (2004). A molecular defect in intracellular lipid signaling in human neutrophils in localized aggressive periodontal tissue damage. Journal of immunology *172*, 1856-1861.

Han, X. (2016). Lipidomics for studying metabolism. Nature reviews. Endocrinology *12*, 668-679.

Hui, S.T., Parks, B.W., Org, E., Norheim, F., Che, N., Pan, C., Castellani, L.W., Charugundla, S., Dirks, D.L., Psychogios, N., et al. (2015). The genetic architecture of NAFLD among inbred strains of mice. eLife *4*, e05607.

Hyotylainen, T., and Oresic, M. (2014). Systems biology strategies to study lipidomes in health and disease. Progress in lipid research *55*, 43-60.

Jha, P., McDevitt, M.T., Halilbasic, E., Williams, E.G., Quiros, P.M., Gariani, K., Sleiman, M.B., Gupta, R., Ulbrich, A., Jochem, A., et al. (2018). Genetic Regulation of Plasma Lipid Species and Their Association with Metabolic Phenotypes. Cell systems *6*, 709-721.e706.

Jump, D.B. (2011). Fatty acid regulation of hepatic lipid metabolism. Current opinion in clinical nutrition and metabolic care *14*, 115-120.

Kahali, B., Halligan, B., and Speliotes, E.K. (2015). Insights from Genome-Wide Association Analyses of Nonalcoholic Fatty Liver Disease. Seminars in liver disease *35*, 375-391.

Koekemoer, T.C., and Oelofsen, W. (2001). Properties of porcine white adipose tissue and liver mitochondrial subpopulations. The international journal of biochemistry & cell biology *33*, 889-901.

Koybasi, S., Senkal, C.E., Sundararaj, K., Spassieva, S., Bielawski, J., Osta, W., Day, T.A., Jiang, J.C., Jazwinski, S.M., Hannun, Y.A., et al. (2004). Defects in cell growth regulation by C18:0-ceramide and longevity assurance gene 1 in human head and neck squamous cell carcinomas. The Journal of biological chemistry *279*, 44311-44319.

Kroesen, B.J., Pettus, B., Luberto, C., Busman, M., Sietsma, H., de Leij, L., and Hannun, Y.A. (2001). Induction of apoptosis through B-cell receptor cross-linking occurs via de novo generated C16-ceramide and involves mitochondria. The Journal of biological chemistry *276*, 13606-13614.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome research *19*, 1639-1645.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics *9*, 559.

Li, M.J., Liu, Z., Wang, P., Wong, M.P., Nelson, M.R., Kocher, J.P., Yeager, M., Sham, P.C., Chanock, S.J., Xia, Z., et al. (2016). GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. Nucleic acids research *44*, D869-876.

Lusis, A.J., Seldin, M.M., Allayee, H., Bennett, B.J., Civelek, M., Davis, R.C., Eskin, E., Farber, C.R., Hui, S., Mehrabian, M., et al. (2016). The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. Journal of lipid research *57*, 925-942.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research *45*, D896-d901.

Malhi, H., and Gores, G.J. (2008). Molecular mechanisms of lipotoxicity in nonalcoholic fatty liver disease. Seminars in liver disease *28*, 360-369.

Peirce, J.L., Lu, L., Gu, J., Silver, L.M., and Williams, R.W. (2004). A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC genetics *5*, 7.

Quehenberger, O., and Dennis, E.A. (2011). The human plasma lipidome. N Engl J Med *365*, 1812-1823.

Stefely, J.A., Kwiecien, N.W., Freiberger, E.C., Richards, A.L., Jochem, A., Rush, M.J.P., Ulbrich, A., Robinson, K.P., Hutchins, P.D., Veling, M.T., et al. (2016). Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. Nature biotechnology *34*, 1191-1197.

Wenk, M.R. (2005). The emerging field of lipidomics. Nature reviews. Drug discovery *4*, 594-610.

Williams, E.G., Wu, Y.B., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. Science *352*, 1292-+.

Wu, Y., Williams, E.G., Dubuis, S., Mottis, A., Jovaisaite, V., Houten, S.M., Argmann, C.A., Faridi, P., Wolski, W., Kutalik, Z., et al. (2014). Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. Cell *158*, 1415-1430.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology *16*, 284-287.

Chapter 3: Genetic regulation of plasma lipid species and their association with metabolic phenotypes

Pooja Jha[1], Molly T. McDevitt[1], Emina Halilbasic, Evan G. Williams, Pedro M. Quiros, Karim Gariani, Maroun B. Sleiman, Rahul Gupta, Arne Ulbrich, Adam Jochem, Joshua J. Coon, Michael Trauner, David J. Pagliarini, and Johan Auwerx

[1]These authors contributed equally.

This chapter was published as an article in *Cell Systems* (2018, 6, 709-721) by the authors listed above.

Summary

The genetic regulation and physiological impact of most lipid species are unexplored. Here, we profiled 129 plasma lipid species across 49 strains of the BXD mouse genetic reference population fed either chow or a high-fat diet. By integrating these data with genomics and phenomics datasets, we elucidated genes by environment (diet) interactions that regulate systemic metabolism. We found quantitative trait loci (QTLs) for ~94% of the lipids measured. Several QTLs harbored genes associated with blood lipid levels and abnormal lipid metabolism in human genome-wide association studies. Lipid species from different classes provided signatures of metabolic health, including seven plasma triglyceride species that associated with either healthy or fatty liver. This observation was further validated in an independent mouse model of non-alcoholic fatty liver disease (NAFLD) and in plasma from NAFLD patients. This work provides a resource to identify plausible genes regulating the measured lipid species and their association with metabolic traits.

Introduction

Lipids are central to all biological processes, from storing energy to forming cell membranes to signaling (Han, 2016). Although lipids are composed of few structural 'building blocks', their vast combinatorial side chain possibilities yield ~100,000 distinct endogenous molecular species (http://www.lipidmaps.org) (Fahy et al., 2005). Understanding the physiological contributions of these diverse lipids is important because species within the same class have differential association with disease states (Puri et al., 2009; Quehenberger and Dennis, 2011; Rhee et al., 2011). Moreover, routine measurements of whole lipid class abundance using enzymatic assays often do not correlate with the abundance of individual lipid species in disease states (Quehenberger and Dennis, 2011).

The endogenous lipid profile in mammals is determined by the combined influences of genes, environmental factors and their interactions (GxE). The profile varies based on changes in dietary lipids, de novo lipogenesis (DNL), and alterations in the activities of the hundreds of enzymes that modulate length, desaturation and incorporation of fatty acids (FA) into more complex lipid molecules. Aberrant levels of storage lipids (triacylglycerol, TAG), circulating lipid–protein complexes (lipoprotein particles), and membrane lipids (phospholipids (PL) and diacylglycerols (DAG)) have been linked to metabolic dysfunction, such as seen in the metabolic syndrome, whose features include obesity, insulin resistance, cardiovascular diseases, and non-alcoholic fatty liver disease (NAFLD) (Farese et al., 2012; Han, 2016; Puri et al., 2009; Quehenberger and Dennis, 2011). However, the contribution of individual lipid species *per se* to metabolic dysfunction or to a healthy metabolic state is poorly understood. In the past decade, genome wide association studies (GWAS) have identified numerous genetic variants and loci associated with different lipid classes and complex metabolic traits (Dewey et al., 2016; Diabetes Genetics Initiative of Broad Institute of et al., 2007; Global Lipids Genetics et al., 2013); however, these loci explain only (5-20%) of the variance observed (Johansen et al., 2011; Manolio et al.,

2009). This problem is furthermore compounded as many traits, such as lipids, are strongly modulated by GxE interactions, which are difficult to control in human studies. In contrast, studies of mouse genetic reference populations (GRPs) in which environmental factors can be controlled have been able to provide a stable platform for identifying major genetic, environmental, and GxE factors influencing complex traits (Hui et al., 2015; Sittig et al., 2016; Williams and Auwerx, 2015). While earlier mouse population studies have identified specific loci associated with different lipid classes (Hui et al., 2015; Zhang et al., 2012), no study in a GRP has specifically profiled individual lipid species or identified loci associated with them—nor has GxE been examined for these traits.

In this study, we performed lipidomic profiling across 49 distinct inbred strains of the BXD GRP (descending from crosses between C57BL/6J mothers and DBA/2J fathers) fed either chow diet (CD) or high fat diet (HFD) and subjected, in parallel, to an extensive battery of metabolic tests (Williams et al., 2016). Both genetic and multilayered omics approaches were used to gain a comprehensive understanding of the genetic and dietary impact on lipid species and to uncover the potential of lipid species as signatures of metabolic health.

Results

Plasma lipid species profile across the BXD GRP

To characterize the physiological significance of the plasma lipidome, we used 78 BXD cohorts from 49 different strains—44 cohorts fed CD and 34 fed HFD (280 mice) for 21 weeks. While on their respective diet, mice underwent extensive metabolic phenotyping (Williams et al., 2016; Wu et al., 2014). Subsequently, mice were sacrificed in an overnight fasted state and plasma samples were analyzed using a discovery LC-MS/MS lipidomics platform. Lipid species were separated via reversed phase chromatography and identified using high-resolution precursor and fragmentation scans. Collectively, we identified eight distinct lipid classes, including free fatty acids (FFAs; 16 species), triacylglycerol (TAG; 53), diacylglycerol (DAG; 6), phosphatidylcholine (PC; 28) phosphatidylethanolamine (PE; 15), phosphatidylinositol (PI; 7), phosphatidylglycerol (PG; 2) and coenzyme Q (CoQ) (2) (Table S1). The high quality and reproducibility of the mass spectrometry (MS) measurements are evidenced by the pairwise-correlation of technical and extraction replicates, showing consistent robust correlation (Figure S1). Furthermore, we performed all possible pairwise correlations between the BXD lipidomes from different groups to assess the sensitivity of our measurements in detecting global diet- and strain-driven differences (Figure 1A). The correlation of all lipids within each strain (biological replicates) in either diet (CD or HFD) was higher than within strain-across diet (CD vs. HFD) correlation, implying that the global effect of diet is larger than that of the biological replicates of the strains. As expected, the correlation across strains on a given diet (either CD or HFD) was higher than across strain-across diet correlation (Figure 1A). Hierarchical cluster analysis demonstrated that most BXD cohorts segregated based on diet (Figure 1B), which was also evident in the first dimension (PC1) of the principal component analysis (PCA) (Figure 1C). However, the variance explained by the sum of the first two principal components is only 53%,

which indicates that the lipidomic profile of the BXDs is highly variable in both diets (Figure 1C). Of the 129 lipid species measured, 93 were significantly different between the two dietary cohorts (53 upregulated and 40 downregulated, HFD vs. CD), (Table S2A), indicating a major effect of diet on lipid profiles.

We next evaluated the impact of diet on monounsaturated fatty acid (MUFA) and polyunsaturated fatty acid (PUFA) synthesis—major determinants of the lipid profile—by analyzing the levels of serum FFAs from mice on both diets (Figure 1D). Though the saturated fatty acids (SFAs) and MUFAs were ~13-19 times higher in the diet of HFD vs. CD cohorts, only stearic (18:0) and oleic (18:1n9) acids were significantly increased in the plasma of HFD cohorts, indicating increased accumulation under HFD (Figure 1D, top). Furthermore, the desaturation and elongation products of 18:1n9 (20:1n9, 22:1n9, 24:1n9) were not increased in HFD cohorts. This indicates a specific accumulation/enrichment of oleic acid side chains in lipid species increased on HFD (Figure 1D; top, Figure 1E). Though the essential fatty acids (EFAs), including linoleic (18:2n6) and linolenic (18:3n3) acids were 1.5 and 1.8 times higher in the diet of HFD cohorts, their levels were decreased in the plasma of HFD cohorts (Figure 1D, bottom). This reflects increased utilization of the EFAs in HFD cohorts to build long chain FFAs (e.g. 20:4n6 and 22:4n6) and their use as an energy source as evidenced by increased peroxisomal-β oxidation in HFD cohorts (Figure 1D, bottom).

To evaluate the effect of diet on side chain composition of lipid species, we calculated the FA composition of all side chains from the lipid species. [For co-eluting isobaric species (Table S1), all possible fatty acid combinations for which there was evidence were used in this calculation]. The majority of side chains were comprised of palmitic, stearic, oleic, palmitoleic, linoleic, arachidonic and docosahexaenoic (DHA) acid. Lipid species with at least one palmitic, stearic or oleic acid in their side chain were enriched in HFD cohorts, while species with at least one palmitoleic, linoleic and linolenic acid in their side chain were enriched in CD cohorts (Figure

1E). Lipid species with an arachidonic acid side chain were either enriched in HFD or remained unchanged, while lipids with DHA side chains were either enriched in CD or remained unchanged (Figure 1E). Notably, the dietary enrichment of all these fatty acids (except palmitic acid) in the lipid species observed here also reflects the FFA profile seen in Figure 1D. Taken together, these results demonstrate that diet has a strong influence in determining the lipid profile and that FFAs in general can reflect the side chain enrichment of lipid species. These data suggest that in HFD there may be a switch towards lipid species being enriched in palmitic, stearic and oleic acids while being depleted in linoleic acid because of its increased utilization to meet the energy demand (metabolic flexibility) since the energy from carbohydrate is only 27% kcal in HFD vs. 44% kcal in CD.

Plasma lipid species have high heritability

Next, we assessed the degree of genetic, dietary, and GxE regulation of lipid species. Heritability (h^2; percentage of trait variation attributed to additive genetic factors) was calculated for all lipid species within (CD and HFD) and across (CD+HFD / Mixed) dietary groups (Belknap, 1998) (Figure 2A and Table S2B). Additionally, the variation attributed by GxE, diet and non-genetic, non-dietary variance (Unexplained) was calculated (Figure 2A and Table S2B). Within a dietary cohort, more than half of the observed variance in lipid levels could be explained by genetic differences across strains (i.e. $h^2 \geq 50\%$) for the strong majority of lipid species (80% in CD and 75% in HFD). Conversely, when dietary cohorts were combined, only 22% of lipids had h^2 above 50%. However, even when the dietary cohorts were mixed, the diet-independent genetic factor (Figure 2A, "Genetics (CD + HFD /mixed)") was the strongest contributor to variance explained. The contribution of GxE ("GxE Effect") or diet alone ("Diet Effect") did not explain even 50% of observed variance for a single lipid species. Together, these three controlled factors—genetics, environment, and GxE—explained more than half of the variance for 85% of all lipid species, whereas only 15% lipids had more than 50% of their variance attributed to unexplained factors

(Figure 2A, “Unexplained”). For many lipids, the variance in their levels both within and across diets reflected their h^2 and variance explained by genetics (CD+HFD/ Mixed), diet and GxE. For instance, TAG(52:2) levels were significantly increased by HFD, where the diet explained its maximal variance (48%) (Figure 2B). Conversely, PC(20:1_22:6) levels were not different between the two cohorts, showing high h^2 in both diets and mixed variance (attributed to genetic factors) explained most of its variance (70%) and dietary effect was only 3% (Figure 2C). Collectively, these results indicate a high degree of genetic regulation of lipid species in both diets.

Plasma lipids are influenced by many genomic loci, including several associated with lipid levels in human GWAS

Next, we mapped quantitative trait loci (QTL) for all lipid species (lQTL), identifying 212 lQTLs in CD and 94 in HFD (Figure 3A [lQTLs above blue dotted line] and Table S4). It is not surprising that most lipids had more than one significant QTL, given the fact that many independent processes and pathways regulate lipid species. Only three lipid species (DAG(36:2), TAG(54:1) and eicosenoic acid/ 20:1) had the same lQTL position (within 5Mb) in CD and HFD (Figure 3A, black bold font), which signifies that diet can influence the genetic factors regulating most lipid species. Of note, the CD-HFD lQTL pair of DAG(36:2)—mixture of [18:1_18:1 and 18:0_18:2]—mapped to the region containing genes with protein coding variants involved in lipid metabolic processes (*Slc2a2, Kcnmb2, Bbs12, Fgf2, Pld1*), insulin secretion and homeostasis (*Slc2a2, Kcnmb2, Bbs12)* (Figure 3B, left). This finding is consistent with the role of DAGs in insulin sensitivity (Farese et al., 2012) and the fact that DAG (18:0/18:2) has been proposed as one of the 21 plasma lipid predictors in the diabetes risk classification model (Wong et al., 2013). Importantly, the glucose transporter *Slc2a2* (Dupuis et al., 2010)—associated with glucose levels, type 2 diabetes, and metabolic syndrome—showed a strong positive association with metabolic syndrome phenotypes along with DAG(36:2) (Figure 3B, right). A prominent metabolic lQTL

hotspot, common to 14 lipid species, was observed on Chr2 in CD (Figure 3A and 3C). This region (101-113 Mb) contained 10 genes (among others) encoding proteins involved in lipid metabolic processes including FoxO signaling, lipid transport, and having phospholipase and acyl transferase activity (Figure 3C, bottom). Of note are *Cat* (catalase; FoxO signaling), *Pdhx* (Pyruvate Dehydrogenase Complex Component X; acyl transferase activity), *Chrm5* (Cholinergic receptor, muscarinic 5; phospholipase activity), and CD44 (CD44 antigen) which have non-synonymous SNPs in the BXDs (Figure 3C, bottom). For each lQTL, we have provided the lQTL genes having nsSNP/ncSNP/indel in Table S4. Of the 306 lQTLs, 172 lQTLs harbor 97 genes (having non-synonymous SNPs in BXDs) that are known to be directly associated with lipid metabolic terms including “lipid metabolic process”, “lipid particle”, and “lipid binding”, “transferase activity”, and “transferring acyl groups” (Table S4).

Human GWAS have identified many genetic variants associated with plasma lipids and associated metabolic traits (http://jjwanglab.org/gwasdb, https://www.ebi.ac.uk/gwas/) (Diabetes Genetics Initiative of Broad Institute of et al., 2007; Kathiresan et al., 2009; Willer et al., 2013). Because of the close homology between mouse and human genomes, the mouse can be used to add evidence to genes suggested by human studies. Taking advantage of these genes identified in GWAS studies (Table S5), we screened the 306 lQTLs (Table S4) for the presence of any human GWAS genes associated with plasma lipids and associated metabolic traits (Figure 3D). We applied a stringent approach by screening only those lQTL genes that have non-synonymous SNPs in BXDs. This screening identified 40 GWAS candidate genes (out of 494 GWAS genes) under 93 lQTLs: 65 lQTLs in CD from 55 lipids and 28 lQTLs in HFD from 25 lipids (Figure 3D; Tables 1 and S5). While 45 lipids in CD and 22 in HFD had only one lQTL harboring the GWAS genes, 10 lipids in CD and 3 in HFD had 2 lQTLs each harboring two different GWAS genes, indicating a polygenic regulation of these lipid species by different genes (Figure 3D). Only 12 lipids had lQTL in both diets, of which 10 mapped to different GWAS loci in the two diets and

only 2 ((DAG(36:2)/ DAG(18:1_18:1 and 18:0_18:2), and eicosenoic acid_ 20:1) mapped at the same GWAS loci (within 5Mb) in both diets (Figure 3A; Tables 1 and S5). Importantly, we found 6 syntenic GWAS regions in humans (0.04-0.9 Mb) that were also syntenic in mice (0.4-0.8 Mb) (Table 1 [genes indicated by prefix "#"] and S5 [genes indicated in red font]).

Apart from the genes in Table 1, some lQTLs harbored prominent human GWAS genes (associated with blood lipids and metabolic traits) which do not have non-synonymous SNPs in BXDs but may be linked to the causal variant. For instance, TAG(50:3) had a QTL at the locus of glucokinase regulatory protein (*Gckr*) and *Lrpap1* (LDL Receptor Related Protein Associated Protein 1). *Gckr* is one of the most robust loci implicated in TAG metabolism (Figure S2A) (Diabetes Genetics Initiative of Broad Institute of et al., 2007). It has been replicated in GWAS of plasma TAG concentration, NAFLD and hypertriglyceridemia by an excess of rare variants in patients (Johansen et al., 2011; Kathiresan et al., 2009; Speliotes et al., 2011). *Lrpap1* shares the same locus as *Gckr* (within 5Mb) and has been associated with TAG, HDL-C and LDL-C (Figure S2A). Fatty acid desaturase (*Fads* 1, 2 and 3) is another prominent gene, which has been associated with TAG, PLs and type 2 diabetes in numerous GWAS studies (Diabetes Genetics Initiative of Broad Institute of et al., 2007; Johansen et al., 2011; Kathiresan et al., 2009; Speliotes et al., 2011) (Figure S2B). In line, two TAGs and 4 PCs had lQTLs at the *Fads* locus (Figure S2B). Interestingly the FADs locus is syntenic in both mice and humans (Figure S2B). These data suggest that a large number of GWAS genes associated to total lipid levels in humans can also regulate many individual lipid species.

Lipid species provide a signature of metabolic health status

Since many lipid species shared common QTLs, we hypothesized that they are correlated/ co-regulated. To assess this, we performed an unweighted correlation network analysis of all the lipid species, giving an overview of interactions between different lipid species and classes (Figure

4A and 4B). In both CD and HFD cohorts, lipid species were highly correlated both within and across different lipid classes. The network in both cohorts showed a strong dense grouping of TAGs, PCs, FFAs and DAGs (Figure 4A and 4B, red ellipse). This cluster is indicative of the PC and TAG synthesis from DAG intermediates (Han, 2016). Diet specific interactions were also observed; e.g., TAGs with high carbon number (56-64) were not connected in the CD network but were integrated with other TAGs in the HFD network (Figure 4A and 4B, green ellipse). These results suggest that lipids across different classes are correlated, and that a change in one or more species may impact the levels of many other lipid species independent of the lipid class.

Based on this conception, we hypothesized that it may be possible to predict metabolic health based on the variation in lipid levels and vice-versa. From the extensive phenotypic profiling performed in these mice (Williams et al., 2016), we acquired 30 unique traits reflective of metabolic health/fitness. These traits included fat and lean mass, physical fitness (treadmill exercise, activity wheel, VO_2 max), oral glucose tolerance test, heart rate and plasma biochemical markers (e.g. alanine transaminase (ALT), aspartate transaminase (AST), cholesterol, FFAs, etc.), among others. Spearman correlation was calculated between all lipid species and the metabolic traits (Table S6). The correlation rho value was used to perform a heatmap analysis with unsupervised hierarchical clustering of all lipid species and traits (Figure S3A). In both CD and HFD, two lateral clusters (extreme right and left) of lipid species could easily be identified, that correlated with most of the metabolic traits (Figure S3A). From these two lateral clusters in each diet (Figure S3B), we identified 36 lipid species that showed the strongest correlations with metabolic traits in both diets (Figure 4C and 4D). Among them, 19 lipids were identified as “healthy markers” in both diets (Figure 4C and 4D, vertical green cluster), since they positively correlated with healthy metabolic traits (Figure 4C and 4D, horizontal green cluster) and negatively with unhealthy traits (Figure 4C and 4D, horizontal red cluster). While 17 lipids were identified as “unhealthy markers” showing the inverse correlation pattern (4C and 4D, vertical red cluster).

Note that FFA, kidney weight, heart weight and heart rate cluster with healthy metabolic phenotypes. The fact that FFAs provide 60-70% of the heart's energy requirement (van der Vusse et al., 2000) and are also the predominant energy source for the kidney (tubular epithelial cells) (Kang et al., 2015) explains the clustering of these traits together. Interestingly, only 7 of the 36 lipids identified as metabolic health predictors are amongst the most abundant lipids in their class [PC(20:4_18:0), PC(20:4_16:0), PC(22:6_16:0), PC(36:1), TAG(52:2), TAG(52:4), TAG(54:3)]; the remaining 29 lipids have low abundance in plasma (Table S2). Taken together, these data suggest that lipid species reflect metabolic health and that the most abundant species may not necessarily be the best predictors of health.

Identification of signature lipid species for NAFLD

Since plasma lipids associated with metabolic traits, we next tested this finding in the context of a disease. Taking advantage of the liver lipidome from the same cohort of mice from our companion article (Jha et al., 2018), we tested whether plasma lipid species can be indicative of NAFLD, characterized by excess TAG accumulation in liver (Kleiner et al., 2005). Both in humans and different mouse strains, plasma and liver TAG levels are highly variable and are not always increased in NAFLD or by HFD, indicating a complex polygenic regulation (Browning et al., 2004; Johansen et al., 2011; Kirk et al., 1995; Lin et al., 2005; Romeo et al., 2008). In line, plasma and liver TAG levels were highly variable in the BXDs, although no significant difference was observed between CD and HFD cohorts (Figure 5A). Moreover, the liver and plasma total TAG levels did not correlate (Figure 5B). Therefore, we sought to identify individual lipid species in plasma that are representative of their levels in liver. Of the 55 common lipids measured in plasma and liver, 30 in CD and 21 in HFD correlated with a Spearman's rho (absolute value) > 0.32 (Figure 5C; orange and green bar, Table S7). Additionally, there was a strong positive correlation between plasma and liver lipids across diets (Figure 5D). In particular, plasma and

liver values of nine of these lipids significantly and tightly correlated across both diets (Figure 5D; dark green and orange dots), implying that the plasma levels of these lipids reflect their abundances in liver in both diets (Figure 5E). Of these nine lipids, four TAG species (TAG: 52:2, 54:3, 56:3 and 54:1) were increased in the HFD cohorts in both plasma and liver, while three TAG species (TAG: 52:5, 52:4 and 54:6) were increased in the CD cohorts in both plasma and liver (Figure 5E and S4). TAG(50:2) was significantly increased in HFD in plasma but not in liver (Figure S4). While DAG(L_P) (DAG(18:2_16:0)) showed a negative correlation between liver and plasma (Figure 5E) because on HFD, it was decreased in plasma, but increased in liver (Figure S4). The accumulation of DAG(L_P) in liver and reduction in plasma (Figure S4) suggests that its release from liver into the plasma may have been minimized.

Correlation analysis of these TAG species, alongside total TAG concentration in liver and plasma, was performed with phenotypes linked with NAFLD including fasting insulin, glucose, cholesterol, ALT, AST, fat mass, liver mass and body weight (Figure 5F). For either diet, the total TAG concentration in plasma or liver either did not correlate or had a weak correlation with NAFLD associated phenotypes, which was in contrast to the individual TAG species (Figure 5F). The TAGs: 52:2, 54:3, 56:3 and 50:2, showed positive correlation, while the TAGs: 52:5, 52:4 and 54:6 showed negative correlation with the NAFLD readouts in both diets in liver and plasma, (Figure 5F, red and blue font respectively). Importantly, these seven lipids had high h^2 (55-84%) in plasma in both CD and HFD cohorts (Table S3). Of note, DAG(L_P), which shows a negative correlation between plasma and liver (increased in CD in plasma and in HFD in liver) (Figure 5E, S4), is the only lipid among the nine lipids showing an opposite correlation with NAFLD phenotypes between liver and plasma (Figure 5F). TAG(54:1) can be considered a diet-specific NAFLD marker since it correlated positively with NAFLD markers in CD and negatively in HFD, indicative of a GxE effect, where the HFD effect reverses the genetic effect; its h^2 in CD being 57% and in HFD, 48% (Table S3). The unexplained variance for both TAG(54:1) and DAG(L_P)

was 35% and 48% respectively (Table S3), suggesting that these two lipids may be less reliable plasma signatures for NAFLD. These findings suggest that elevation of TAGs with fewer double bonds: 52:2, 54:3, 56:3 and 50:2 comprise a pro-NAFLD signature, whereas an increase in TAGs with more double bonds: 52:5, 52:4 and 54:6 comprise an anti-NAFLD signature, irrespective of the diet. Notably, the pro-NAFLD TAGs: 52:2, and 54:3, and the anti-NAFLD TAGs: 52:5, 52:4 and 54:6 are also amongst the systemic unhealthy and healthy metabolic markers (Figure 4C and 4D). Our data also signifies the importance of having a strong correlation between a plasma and liver lipid in order for it to be established as a disease signature.

Assessment of identified NAFLD signatures in mice and humans

We then tested the relevance of these lipid markers as readouts of fatty liver in a different mouse model of NAFLD (induced by high-fat high-sucrose (HFHS) diet) and tested whether a NAFLD lowering therapeutic intervention, i.e. nicotinamide adenine dinucleotide (NAD^+) precursor, nicotinamide riboside (NR), impacts these lipid markers. As such, we compared the liver lipidome of mice fed CD, HFHS diet for 18 weeks, or mice that were fed HFHS diet supplemented with NR 9 weeks after the start of HFHS diet, at which point they had already developed NAFLD (a therapeutic intervention, HFHS+NR) (Figure S5A) (Gariani et al., 2016). Though liver total TAG concentration was significantly elevated in the HFHS group and decreased by NR (Figure S5B)(Gariani et al., 2016), all individual TAG species did not show this obvious profile despite the uniform genetic background (C57BL/6J) (Figure S5C). Of all the TAG species measured, only 53% were increased in the HFHS cohort while 33% were decreased and 14% remained unchanged (data not shown). Importantly, all four pro-NAFLD lipids (TAGs 52:2, 54:3, 56:3 and 50:2) from the BXD HFD study (Figure 5F) were also increased in C57BL6/J mice fed HFHS diet, whereas the three anti-NAFLD lipids (TAGs 52:5, 52:4 and 54:6) were decreased in HFHS diet cohorts (Figure 6A). NR significantly lowered three of the four pro-NAFLD lipids (TAGs

54:3, 56:3, 50:2), but increased the anti-NAFLD lipids only to a slight extent (Figure 6A). Importantly, the pro-NAFLD lipids correlated positively, whereas the anti-NAFLD lipids correlated negatively with clinical NAFLD readouts (Figure 6B). Notably, the liver NAD^+ levels correlated negatively with the pro-NAFLD lipids, and positively with anti-NAFLD lipids (Figure 6B), fully in line with our previous work which showed that NAD^+ levels are depleted in steatotic livers and are replenished after NR treatment (Gariani et al., 2016).

To further explore the clinical relevance of these findings, we analyzed the plasma lipidome of healthy patients, patients with steatosis, early stage NASH (mild to moderate fibrosis (F1+F2)) and advanced stage NASH (severe bridging fibrosis or cirrhosis (F3+F4)). The first dimension (PC1) of the PCA of 55 TAG species segregated nearly all of the healthy individuals from the steatosis and NASH group (Figure S5D). However, the variance explained by the PC1 was only 46%, implying that the TAG profile in humans is highly variable possibly due to the high genetic variation (Figure S5D). Despite the significant differences in total TAG plasma levels between the healthy and NAFLD groups (Figure S5E), only a few individual TAG species were significantly changed, which is in line with the findings from mice. Only 20% of TAG species were increased in steatosis vs. healthy, while 64% increased in early stage NASH vs. healthy and 49% increased in advanced stage NASH vs. healthy group (data not shown). However, the pro-NAFLD lipids were increased in steatosis and/or the two NASH groups. Compared to the healthy group, TAG(50:2) was increased in both steatosis and the two NASH groups (F1+F2 and F3+F4), while TAG(52:2) was increased only in the two NASH groups; whereas TAG(54:3) and TAG(56:3) were increased in the two NASH vs. the steatosis groups (Figure 6C). Conversely, the three anti-NAFLD lipids, TAG(52:5), TAG(52:4) and TAG(54:6), were decreased in advanced stage compared with early stage NASH (Figure 6C). In line, the pro-NAFLD lipids correlated positively with NAFLD readouts whereas, the anti-NAFLD markers correlated negatively (Figure 6D). Taken together, these data confirm the findings from the BXDs in a different model of diet-induced

NAFLD in mice as well as in humans, indicating that these lipid species may be a more universal signature of NAFLD across different diets and also relevant in humans.

To obtain insight on the affinity of the pro- and anti-NAFLD signatures with TAG-metabolizing enzymes, we tested the association of adipose triglyceride lipase (ATGL)—the rate limiting TAG-metabolizing enzyme—with these NAFLD TAG signatures. *Atgl* expression in white adipose tissue (WAT) negatively correlated with the pro-NAFLD signatures and positively correlated with the anti-NAFLD signatures in both CD and HFD cohorts (Figure 6E). This correlation was also observed with the expression of ATGL in other metabolic tissues (liver, heart and muscle) (Figure S5F), however, the strongest association was observed with WAT *Atgl* expression, since *Atgl* lipase activity is ~10 times higher in WAT compared to the other tissues (Haemmerle et al., 2006). These findings are consistent with our previous experimental findings showing that *Atgl* is important for protection from steatohepatitis and that *Atgl* -KO mice are susceptible to develop NAFLD/NASH (Jha et al., 2014). Additionally, the pro-NAFLD signatures in both plasma and liver correlated positively with lipid biosynthetic pathways and negatively with oxidative pathways (Figure S5G). Conversely, the anti-NAFLD signatures correlated positively with oxidative pathways and negatively with inflammatory pathways (FigureS5G). Taken together, these biological corroborations of the TAG signatures provide proof of concept validation of the lipid species measured and validate the usefulness of the resource.

Discussion

Here we used a systems approach—combining genetics, lipidomics and phenomics—to examine the dietary and genetic regulation of plasma lipids and their potential to reflect metabolic health status. Our study reveals that most plasma lipids have high heritability and map to a QTL, several of which harbor genes associated with abnormal lipid metabolism in human GWAS. As such we show that specific plasma lipids can be used as signatures of metabolic health status. We further validate the potential of seven TAG species as plasma signatures of NAFLD in mice and humans.

As expected, the change in diet from CD (6% calories from fat) to HFD (60% calories from fat) had a significant impact on a large proportion (70%) of the plasma lipid species measured. Due to the unique genetic background of each of the strains, even within the same dietary cohort, most lipids were highly variable across strains, an observation similar to that in humans (Shin et al., 2014). Across all conditions, roughly 50% of variation could be attributed to diet-independent genetic factors, with an additional ~10% of variation being attributable to uniform diet-induced changes, or a strain-dependent response to dietary differences. Together, our study can explain more than half of the observed variation in plasma lipid levels for 85% of the 129 measured lipid species. For the variation attributable to genetic or GxE factors, we identified about 300 novel lQTLs containing known and novel putative regulators of lipid metabolism. Taking advantage of the known human GWAS loci/genes associated with blood lipid levels and associated traits, we uncovered, via a cross-species examination, the association between 40 human GWAS genes and 93 lQTLs, including 7 syntenic regions common in mice and humans. This link underscores the power of mouse GRPs beyond finding novel loci/genes controlling levels of lipid species to their potential to complement human genetic studies. The fact that our data are derived from a genetically diverse mouse population, which mimics the genetic variation observed in humans,

increases its translational value and suggests that such populations provide a proper setting for additional mechanistic studies related to lipid metabolism.

Network analysis showed that lipid species are correlated and co-regulated, both within and across lipid classes. This resonates with the underlying fact that any change in a lipid metabolic pathway induced by an experimental or physiological intervention will almost assuredly result in compensatory changes in other pathways affecting discrete lipid pools (Farese et al., 2012). Our findings furthermore demonstrate that within the same lipid class, some lipid species are associated with healthy metabolic traits and others with unhealthy metabolic traits, providing testimony to the fact that all lipid species in a class do not have the same physiological impact (Farese et al., 2012; Quehenberger and Dennis, 2011; Rhee et al., 2011). We tested this computational finding by comparing the lipidome of plasma (this study) and liver (Jha et al., 2018) and identified plasma lipids that are reflective of liver lipid accumulation in NAFLD across diets and genetic background. Our results demonstrate that pro-NAFLD TAG signatures have fewer double bonds compared to the anti-NAFLD TAG signatures. These findings are in line with a previous study in humans showing that TAGs with lower carbon and double bond content were associated with increased risk of diabetes (Rhee et al., 2011) and therefore, underscores the potential of mouse population genetics for translational research. Indeed, we validated the TAG signatures in another dietary model of NAFLD and in human NAFLD subjects with consistent results.

Some technical limitations in this study need to be considered. First, our lipidomic platform did not provide a complete coverage of the whole plasma lipidome. Many lipid classes, including ceramides, lysophospholipids, cholesterol esters etc., were not part of our validated lipidomics method at the time these measurements were made. However, our goal was to measure a substantial subset of the plasma lipidome across hundreds of samples in the most accurate way, and by the simplest method, from only 20 µl of plasma, in order to power QTL and other systems

analyses. Second, our measurements do not provide absolute quantification of lipid species. Since the major thrust of this work is the identification of genetic loci regulating lipid levels and the association of lipid species with metabolic phenotypes rather than providing absolute concentrations for lipid species. We acknowledge that more work is required to extend our other genetic and phenotypic findings, which is beyond the scope of this current article. This resource, however, lays the foundation for future mechanistic insight into the complex biology of lipids.

In conclusion, this study uncovers the potential of plasma lipidomics combined with systems genetics approaches in a mouse population to identify potent markers for human health and disease. Identifying genes and genetic variants associated with plasma lipid species enriches our understanding of biochemical pathways and the substrate specificity of numerous enzymes associated with lipid metabolism, while at the same time facilitating the design of new therapies for metabolic diseases. The wealth of information on novel lQTLs and the phenotypic footprint of these lipid species furthermore provide a robust resource to the scientific community for further *in-silico* data analysis. Our findings illustrate the need, importance, and scope of studying individual lipid species and provide a platform for further mechanistic studies of lipid species, as demonstrated in our companion article (Jha et al., 2018).

Experimental procedures

Mice

BXD strains were obtained from University of Tennessee Health Science Center (Memphis, TN, USA) and JAX (The Jackson Laboratory) and bred at the École Polytechnique Fédérale de Lausanne (EPFL) animal facility for more than two generations before incorporation into the study. Cohorts of 49 BXD strains with ~5 males each on CD and HFD were used in this study. Mice were fed a chow diet [CD; 2018 Teklad Global 18% Protein Rodent Diet (6.2% kcal from fat; 44.2% kcal from carbohydrate; 18.6% kcal from protein)] or a high fat diet [HFD; Harlan Teklad, TD.06414 (60.3% kcal from fat; 27.3% kcal from carbohydrate; 18.4% kcal from protein)] for 21 weeks, starting at 8 weeks of age. All mice were phenotyped as described (Williams et al., 2016) (see Method Details section). At week 29, animals were fasted overnight before sacrifice at 9:00 am. Blood was collected from isoflurane anesthetized mice via the vena cavae, and immediately afterwards the animals were perfused with ice cold PBS, through the left ventricle. Blood was collected in lithium-heparin (LiHep)–coated tubes (Microvette CB 300 Hep-Lithium, Sarstedt) shaken and kept in ice. The blood samples were centrifuged at 4500 revolutions per minute (rpm) for 10 min at 4°C before being flash-frozen in liquid nitrogen. Due to insufficient plasma from all cohorts, lipidomics analysis could be performed only on 44 CD and 34 HFD strains with 2-5 mice/strain.

For *in vivo* validation of the lipid markers in mice, lipidomics was performed on liver samples from our previous study (Gariani et al., 2016). In brief, male C57BL/6J mice were separated into three groups at the age of 7 weeks. Animal cohorts were fed a CD, a Western high-fat and high-sucrose (HFHS) diet [HFHS; Harlan Teklad, TD.08811, (44.6% kcal from fat; 40.7% kcal from carbohydrate; 14.7% kcal from protein)] or a HFHS diet that was supplemented with NR (400 mg/kg/day) at week 16 till week 25 (9 weeks) (HFHS+NR). Mice were sacrificed after a 4hr fast at 9:00 am. All mice experiments were approved by the Swiss cantonal veterinary authorities of Vaud under licenses 2257, 2257.1 and 2465.

Fatty Acid Composition (%) in the Diet of BXD Cohorts

Fatty acids	Symbol	CD	HFD
Palmitic	C16:0	0.7	8.02
Stearic	C18:0	0.2	3.93
Oleic	C18:1n9	1.2	14.68
Linoleic	C18:2n6	3.1	4.7
Linolenic	C18:3n3	0.3	0.55
Saturated fat		0.9	12.48
Monounsaturated fat		1.3	16.05
Polyunsaturated fat		3.4	5.4
Total fat		**6.2**	**34.3**

Human Subjects

Human Study Design and Participants

The prospective cohort consisted of patients aged ≥ 18 years with biopsy-proven NASH and healthy controls who agreed to participate in the study. Prospective sample collection was approved by local Ethic committee of Medical University of Vienna number: 474/2011 (patient), 1022/2013 (controls) and all subjects were included after obtaining written inform consent. Subjects were excluded if they had history of current and past alcohol consumption of more than 20-30 g per day, presence of other liver diseases, secondary causes of NAFLD, medications known to cause fatty liver during the previous 6 months and inability to provide informed consent. Demographic data of all participants (29 males, 15 females) were obtained by structured interview. Their clinical data is provided below. Liver biopsies were evaluated by a board certified

pathologist to score for histologic features, according to the histologic scoring system (NAFLD Activity Score) developed by Kleiner et al.,(Kleiner et al., 2005). Patients were classified into 3 groups based on fibrosis score: simple steatosis (no fibrosis), early stage NASH (fibrosis stage 1-2), advanced stage NASH (fibrosis stage 3-4).

Human Plasma Biochemistry

Fasting whole blood samples were obtained by venipuncture after an overnight fast of 8 hours or more and processed for plasma within 2 hours on the day of liver biopsy. Routine blood tests were performed at the university hospital of Medical University of Vienna and included measures of ALT, AST, fasting TAG, cholesterol, glucose and insulin. The clinical information of the participants is provided in the table below.

	Healthy	Steatosis	NASH, Early Stage	NASH, Adv. Stage	p-value
Age (years)	27.83 ± 1.64	42.14 ± 3.29	46.43 ± 3.77	58.18 ± 2.78	a,b,c,e,f
Body weight	65.4 ± 1.92	105.14 ± 10.63	94.92 ± 4.79	111.64 ± 8.90	a,b,c
BMI (Kg/m^2)	21.89 ± 0.50	33.91 ± 3.21	32.37 ± 1.63	36.95 ± 2.79	a,b,c
Fibroscan (kPa)	4.28 ± 0.28	9.03 ± 2.41	8.81 ± 1.21	19.58 ± 2.88	c,e,f
Steatosis (%)	-	38.57 ± 7.46	64.64 ± 5.70	45.45 ± 8.08	d
NAS score	-	3.29 ± 0.47	4.93 ± 0.29	4.64 ± 0.48	d
ALT (U/L)	18.92 ± 1.43	50.14 ± 6.97	84.79 ± 15.34	70 ± 10.98	b,c
AST (U/L)	22.50 ± 1.31	30.29 ± 2.73	44.79 ± 5.14	72.82 ± 27.52	c
Cholesterol (mg/dl)	173 ± 11.39	204.71 ± 25.37	179.21 ± 9.82	173.09 ± 9.76	
Triglyceride (mg/dl)	73.25 ± 7.71	190.43 ± 36.18	167.14 ± 17.21	140.82 ± 2	a,b
Fasting glucose (mg/dl)	84.08 ± 2.30	100.14 ± 6.60	138.43 ± 15.81	120.18 ± 9.53	b
Fasting insulin (µU/mL)	6.70 ± 1.11	24.33 ± 10.06	19.34 ± 1.91	23.72 ± 3.16	a,c

Subjects: n=44; 12 healthy, 7 steatosis, 14 early stage NASH, 11 adv. Stage NASH

Values of mean ± SEM are represented.*p<0.05 from one-way ANOVA with Tukey's multiple comparison test correction. Significant difference between groups are indicated as:

“a” between Healthy and Steatosis

“b” between Healthy and NASH, early stage

“c” between Healthy and NASH, Adv. Stage

“d” between Steatosis and NASH, early stage

“e” between Steatosis and NASH, Adv. Stage“f” between NASH, early stage and NASH, Adv. stage

Metabolic Phenotyping of BXD Cohorts

Metabolic phenotyping including, OGTT, heart rate, exercise performance test (VO_2 max. and activity wheels) and EchoMRI) were performed as described (Williams et al., 2016) and elaborated below.

Oral Glucose Tolerance Test (OGTT)

At 17 weeks of age, after 9 weeks of dietary treatment, all cohorts underwent an oral glucose tolerance test. Mice were fasted overnight before the test, and fasted glucose was tested with a glucometer at the tail vein. Mice were then weighed and given an oral gavage of 20% glucose solution at 10 mL per kg of weight. Glucometer strips were used at 15, 30, 45, 60, 90, 120, 150, and 180 min after the gavage to examine glucose response over time. Blood was also collected at 0 (pregavage), 15, and 30 min to examine insulin levels.

Heart Rate

Two weeks later, at 19 weeks of age, a noninvasive heart rate measurement was performed using a tail-cuff system (BP-2000 Blood Pressure Analysis System, Series II, Visitech Systems) over 4 days. The first 2 days were considered as adaptation to the apparatus, and the second 2 days were used for data analysis, and heart rate measurements were averaged across both days. Outliers on a per-measurement basis were removed, but outlier mice were retained.

Exercise Performance Test

At 23 weeks of age, all mice performed a VO_2 max treadmill experiment (pre-training) using the Metabolic Modular Treadmill (Columbus Instruments). For the first 15 min in the machine for each mouse, the treadmill was off while basal respiratory parameters were

calculated. The last 2 minutes of data before the treadmill turned on were considered basal levels (most mice spend the first few minutes exploring the device). The treadmill then started at a pace of 4.8 m per minute (m/min), followed by a gradual increase over 60 s to 9 m/min, then 4 min at that pace before increasing to 12 m/min over 60 s, then four min at that pace before increasing to 15 m/min over 60 s, then 4 min at that pace, then the speed increased continuously by 0.015 m per second (or +0.9 m/min) thereafter until the end of the experiment at 63.5 min, 1354.5 m, or when the mouse is exhausted. CD cohorts ran against a 10° incline, whereas HFD cohorts were set at 0°. For this test, no mice reached the maximum distance recorded by the machine—all were taken out when exhausted (considered as inability to run). The distance run, maximum VO_2, and maximum RER were recorded. Maximum VO_2 and RER were taken by averaging the last ten measurements. Immediately after the treadmill experiment, mice were placed in individual open-air cages with ad libitum access to activity running wheels (Bioseb BIO-ACTIVW-M, Vitrolles, France) for 10 days. For most strains (with some exceptions), all 10 days of activity wheel usage was recorded, for others the average of last 24 hrs was taken. This running distance constitutes the "24hr run distance" phenotype. After the 10th day, at ~25 weeks of age, mice underwent an identical VO_2max treadmill experiment (post-training) as described above at 23 weeks of age. After this experiment, mice were returned to their standard housing cages—individually—for 4 weeks. Mice were fasted overnight before they were sacrificed.

Body Composition (EchoMRI)

In addition to the body weight measurements taken each week and before each phenotyping experiment, body composition was recorded at 16, 23, and 25 weeks of age. To do so, each mouse was placed briefly in an EchoMRI (magnetic resonance imaging) machine (the 3-in-1, EchoMedical Systems), where lean and fat mass are recorded, along with total body weight, taking ~1 min per individual. All tests were normalized to total body weight in our analyses.

TAG Measurement of BXD Livers

For BXDs, 15µl of the liver lipid extract (same extract as used for liver lipid MS measurement, see companion article (Jha et al., 2018) was used for TAG quantification using the Serum Triglyceride determination kit (Sigma-Aldrich), as per manufacturer's instructions. The organic solvent mix used for dissolving lipids for the MS (mixture of acetonitrile (ACN)/isopropyl alcohol (IPA)/water (H_2O), (65:30:5, v/v/v, 100µL)) was used as blank and for standard curve.

NAD^+ Measurement of C57BL/6J Livers

~20 mg of frozen liver samples were used for NAD^+ extraction in 10% perchloric acid and neutralized in $3MK_2CO_3$ on ice. After centrifugation, the supernatant was filtered and the internal standard (NAD-C13) was added and loaded onto a column (150 Å~ 2.1 mm; Kinetex EVO C18, 100 Å). HPLC was run for 1 min at a flow rate of 300 ml/min with 100% buffer A (Methanol/H2O, 80/20% v,v). Then, a linear gradient to 100% buffer B [H2O + 5mM ammonium acetate] was performed (at 1 to 6 min). Buffer B (100%) was maintained for 3 min (at 6 to 9 min), and then a linear gradient back to 100% buffer A (at 9 to 13 min) began. Buffer A was then maintained at 100% until the end (at 13 to 18 min). NAD^+ eluted as a sharp peak at 3.3 min and was quantified on the basis of the peak area ratio between NAD+ and the internal standard and normalized to tissue weight.

Plasma Clinical Traits of BXD Cohorts

Plasma parameters were measured on 2 times diluted samples (1:1 ratio of plasma to diluent) using Dimension®Xpand Plus (Siemens Healthcare Diagnostics AG, Dudingen, Switzerland). The biochemical tests were performed according to the manufacturer instructions for each parameters: AST (Siemens Healthcare, DF41A), ALT (Siemens Healthcare, DF143), Glucose (Siemens Healthcare, DF40), HDL (Siemens Healthcare, DF48B), LDL (Siemens Healthcare, DF131), Cholesterol (Siemens Healthcare, DF27), LDH (Siemens Healthcare, DF54), TG (Siemens Healthcare, DF69A) and FFA (FUJIFILM Wako Dignostics, NEFA-HR (2)). Insulin concentration was measured with an ELISA assay kit (Mouse Insulin ELISA Kit; Mercodia).

Clinical traits used for correlation of lipid signatures with metabolic phenotypes in C57BL/6J mice (Figure 6B) were obtained from our previous study (Gariani et al., 2016). BXD metabolic and clinical data can be obtained from (Williams et al., 2016) and also available on GeneNetwork (http://www.genenetwork.org).

Lipidomics Sample Preparation and Analysis

Internal Standards (IS) Used

For BXD plasma samples we used Q6, PC(15:0/15:0), PS(17:0/17:0), PE(15:0/15:0), PA(17:0/17:0), PG(15:0/15:0), CL(56:0) and FA(15:0/15:0) as internal standards. For NAFLD lipid signature validation experiments in mice and human samples (Figure 6), we used the standard mix SPLASH® Lipidomix® Mass Spec Standard | 330707, supplemented with Q6 and CL(56:0).

Extractions

Lipid extraction was performed as previously described (Stefely et al., 2016). In brief, 20 µl of IS was added to 20 µl of thawed plasma samples and vortexed (30 s). Chloroform/methanol (1:1, v/v, 1000 µL) was added and samples vortexed (60 s). Subsequently, hydrochloric acid (1M, 200 µL) was added to induce phase separation, followed by 60 s vortex and centrifugation (3,000 g, 3 min, 4°C) to complete phase separation. 550 µL of the organic phase was dried under Ar_2(g). The organic residue was reconstituted in a mixture of acetonitrile (ACN)/isopropyl alcohol (IPA)/water (H_2O) (65:30:5, v/v/v, 100µL) by vortexing (60 s) and transferred to a glass vial for LC-MS analysis. Samples were stored at -80°C until further use.

Discovery Lipidomics

LC-MS analysis was performed on an Ascentis Express C18 column (150 mm x 2.1 mm x 2.7 µm particle size, Waters) using an Accela LC Pump (400 µL/min flow rate, Thermo Scientific, San Jose, CA). Mobile Phase A consisted of 10 mM ammonium acetate in ACN/H_2O (70:30, v/v) containing 250 µL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. Mobile phase B started at 2% and increased to

85% over 20 min, then increased to 99% over 1 min and held there for 7 min. The column was reequilibrated for 2 min before the next injection. 10 µL of sample was injected for each run. The LC system was coupled to a Q Exactive mass spectrometer by a HESI II heated ESI source kept at 300°C. The inlet capillary was kept at 300°C, sheath gas was set to 25 units, and auxiliary gas to 10 units. Spray voltage was set to 3,000 V and the MS was operated in polarity switching mode. Ions from 200-1,600 m/z were isolated (Top 2) for fragmentation by stepped higher-energy collisional dissociation (HCD) (20, 30, 40).

Lipid Species Measurement and Normalization

The resulting spectra were processed using LipidSearch (Thermo Scientific), an automated lipid identification and quantification software in which acquired MS^2s are compared to a lipid database containing more than 1.5 million lipid ions and their predicted fragment ions. Peaks were detected and lipids identified by comparing their fragmentation to predicted fragmentation pattern (m-score > 25). Results from each sample were aligned (RT tolerance of 0.25 min, c-score >25) and quantified by integration of the MS^1 peak. After elimination of all lipids with a grade C or lower, the remaining lipids identity and quantification was verified by manual inspection. Basic quality check and QTL analysis was performed on all datasets normalized differently (normalized to IS, to total lipids and to the lipid class), however, the dataset normalized to total lipids was used for all the analysis and figures shown in this paper due to overall low relative standard deviation in this dataset (data not shown). Normalization to total lipids also has two major advantages over the other normalization methods; 1) all lipids measured did not have a true internal standard, 2) for lipid classes that have few lipid species measured, normalizing to class will be largely driven by one or two highly abundant lipids. Quality assessments of the MS measurements was performed by comparing the reproducibility of the technical and extraction replicates (Figure S1). Note: lipid pairs marked with "_1" and "_2" (TAG 54:5, 54:6 and 58:10) indicate two isobaric peaks. The two peaks are chromatographically separated and consist of

unique combinations of fatty acid tails, despite the total number of carbon molecules and double bonds being equivalent.

Quantification and Statistical Analysis

Bioinformatic and Genetic Analyses

Data normality was checked by the Shapiro-Wilk test in R, with a $W \geq 0.90$ considered normal distribution. Correlations are Pearson's *r* or Spearman's *rho* as indicated in the figures. Student's t-test was used for two groups comparisons in normal data of equal variances, and Welch's t-test otherwise. Heatmaps were generated using the "heatmap.2" function in R. PCA analysis was performed using "prcomp" function in R. Unweighted correlation network graphs were performed using Spearman correlation, keeping all edges with p-values less than 1e-05 in both CD and HFD in R using the custom package imsbInfer, currently on Github (https://github.com/wolski/imsbInfer). GO-BP pathway enrichment analysis (Figure S5G) was performed using the R package "clusterProfiler" (Yu et al., 2012) (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html). Enriched pathways after Benjamini-Hochberg correction ($p < 0.05$) are shown in the figures.

Heritability (h^2) was calculated by one-way (CD/HFD) or two-way (Mixed) ANOVA. The variance explained by GxE, diet and unexplained variance (non-dietary, non-genetic) was calculated by two-way ANOVA. QTL calculations were performed using the R/qtl (v 1.39-5) package (Broman et al., 2003) on the $\log_2$ transformed data. The BXD genotype used for QTL calculations is provided in the Supplemental Information (Table S8). Parametric QTL calculation was performed for normally distributed lipids and non-parametric QTLs were determined for lipids, those were not normally distributed. QTLs with logarithm of the odds ratio (LOD) score >2.5 and p-value <0.40 were used for all the analysis, which includes both significant (p-value < 0.05) and suggestive QTLs (p-value between 0.05 and 0.40) at genome-wide significant threshold, computed by permutation analysis. [Genome-wide p-values of 0.63 correspond approximately to

a local p-value of 0.05, i.e. which is significant in case of prior knowledge used to search for a QTL at that specific location].

All graphs and analyses were either made in R or GraphPad. For R, standard R plotting packages included in gplots or ggplot2—e.g., stripchart, plotCI, and barplot2 were used. Final figures were all prepared with Adobe Illustrator.

QTL Candidate Gene Retrieval

For retrieving lQTL genes, biomaRt was used in R to obtain list of genes located within each QTL region (+/- 5 Mb around the mapped SNP). Genes with nsSNP/ncSNP/indel in BXDs (Wang et al., 2016) under each lQTL with LOD score >2.5 and p-value <0.40 were then filtered and represented in Table S4 and use for all figures and tables reporting lQTL genes.

BXD lQTL and Human GWAS Genes Overlap

Human GWAS genes (having p-value <1e-07) were retrieved from the database GWASdb2 (http://jjwanglab.org/gwasdb) (Li et al., 2016) and complemented with the data from the GWAS Catalog (https://www.ebi.ac.uk/gwas/) (MacArthur et al., 2017). The categories of GWAS gene-sets retrieved from these databases included "Abnormality of lipid metabolism" and "fatty liver disease". Additionally, published papers reporting the relevant GWAS studies, not included in the above-mentioned databases (in particular, the references indicated by their PMID in Table S5) were manually mined to retrieve the candidate genes having p-value <1e-07. Taken together, the compiled list comprised of 494 genes (Table S5). These GWAS genes were used for searching for their presence under the lQTLs. Only those lQTL genes (± 5 Mb on either side of the peak QTL) having non-synonymous SNPs in BXDs were matched for any evidence of them being associated with abnormal lipid metabolism in human GWAS (i.e. matched for their presence in 494 human GWAS genes).

Data and Software Availability

Raw MS data files are available through the CHORUS project data repository (Project ID 1432, Experiment ID 3219). Additionally, normalized MS data is deposited in GeneNetwork (http://www.genenetwork.org) as a resource for public use. To access and analyze the data in GeneNetwork, choose "Mouse (mm10)" for "Species", "BXD" for "Group", "Phenotypes" for "Type", "*BXD Published Phenotype*" for "Data Set and and enter "***PlasmaLipidomics***" for "Get Any". Normalized MS data (normalized to total lipids) is provided in Table S1. Lipid QTLs are provided in Table S4. Lipid species and metabolic trait correlations is provided in Table S7. BXD genotype data used for QTL calculation is provided in Table S8.

Author contributions

Conceptualization, PJ, DJP, and JA.; MS Extractions, Methodology, and Measurement, MTM; Formal Analysis, MBS and PJ; Investigation and Analysis, PJ and PMQ; Data Curation, PJ and EGW; Supervision and Sample Collection (NAFLD Clinical Study), MT and EH; Supervision (MS), JJC and DJP.; Standardization (MS), AJ and AU; Validation (Biomarkers and QTLs), PJ; Validation (GWAS), PJ and RG; Validation (Mouse NAFLD), KG; Writing – Original Draft, PJ and JA; Writing – Review & Editing, PJ, JA, DJP, MTM, and EGW; Visualization, PJ, EGW, JA., and DJP; Project Administration and Funding Acquisition; JA and DJP.

Acknowledgements

We thank Center of Phenogenomics (EPFL) for performing BXD phenotyping and RW Williams for BXDs and GeneNetwork maintenance. JA is supported by grants from the École Polytechnique Fédérale de Lausanne, the Swiss National Science Foundation (31003A-140780), the AgingX program of the Swiss Initiative for Systems Biology (51RTP0-151019), and the NIH (R01AG043930). DJP is supported by a grant from NIH (R01GM115591). MTM and EGW were supported by NIH fellowships (T32GM007215and F32GM119190, respectively). JJC is

supported by NIH grants R35GM118110 and P41GM108538. MT is supported by grant F3008 from the Austrian Science Foundation.

Figures and tables

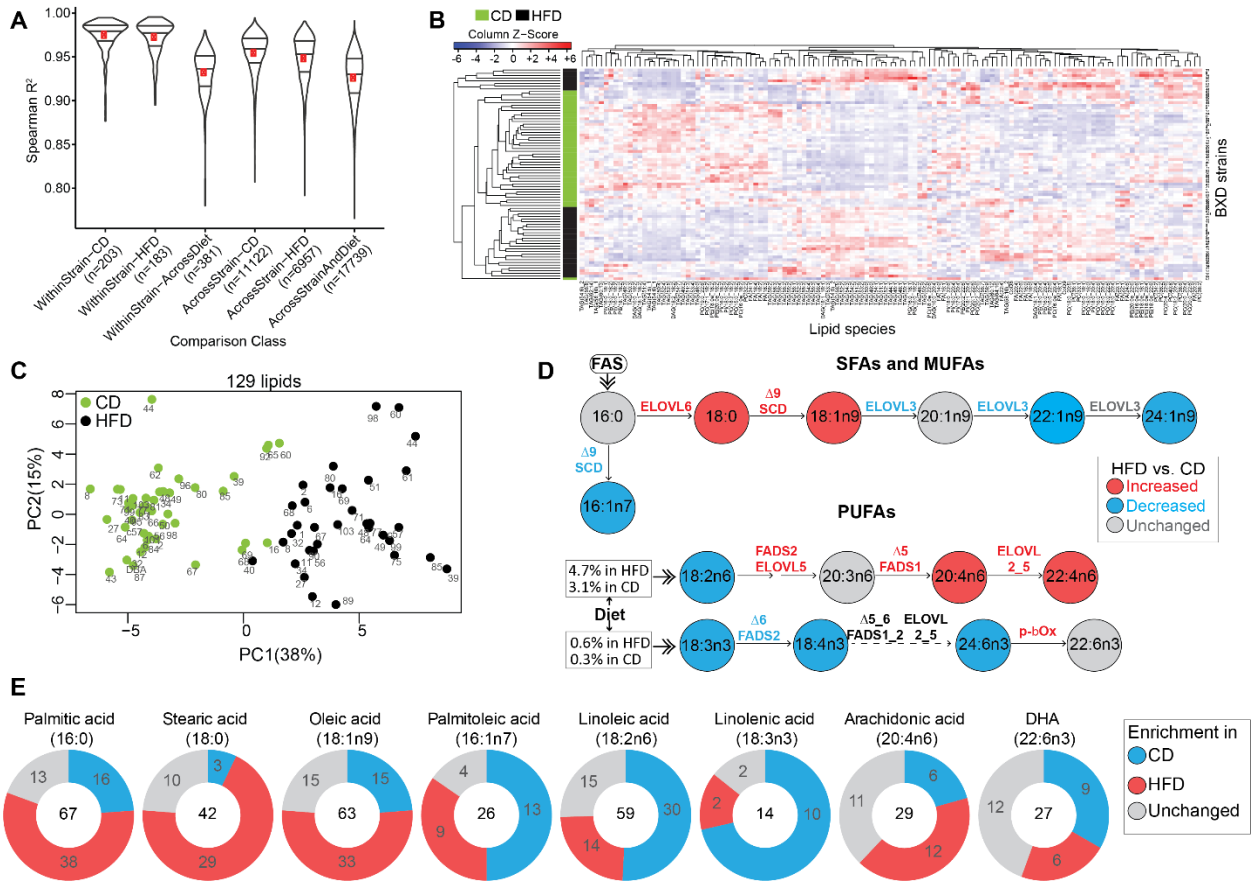


Figure 1. Dietary impact on the plasma lipidome.

(A) Pairwise correlation to assess the sensitivity of MS to detect global diet- and strain-driven differences across 280 samples. Red dot represents mean; the three lines represent median, upper and lower quartile.

(B) Heatmap of unsupervised hierarchical clustering of 129 lipid species for each BXD cohort.

(C) PCA of all lipids in each BXD strain (indicated by the strain number).

(D) Schematic representation of the systemic profile of saturated fatty acids (SFAs) and monounsaturated fatty acids (MUFAs) (top) and polyunsaturated fatty acids (PUFAs) (bottom) in BXDs, based on the levels of the free fatty acids (FFAs) measured. Significant changes (HFD versus CD; $p < 0.05$) for FFA levels and activity of the desaturases and elongases (ratio of product and precursor FFA) are shown as red for increase or blue for decrease.

(E) Pie chart showing the dietary enrichment of eight common side chain FAs in the lipid species in either diet. Number in the center indicates the total number of side chains having at least one indicated FA side chain. For each pie chart, the three colors represent the number of lipids (having the indicated FA side chain) increased in HFD versus CD (red), CD versus HFD (blue), or unchanged (gray).

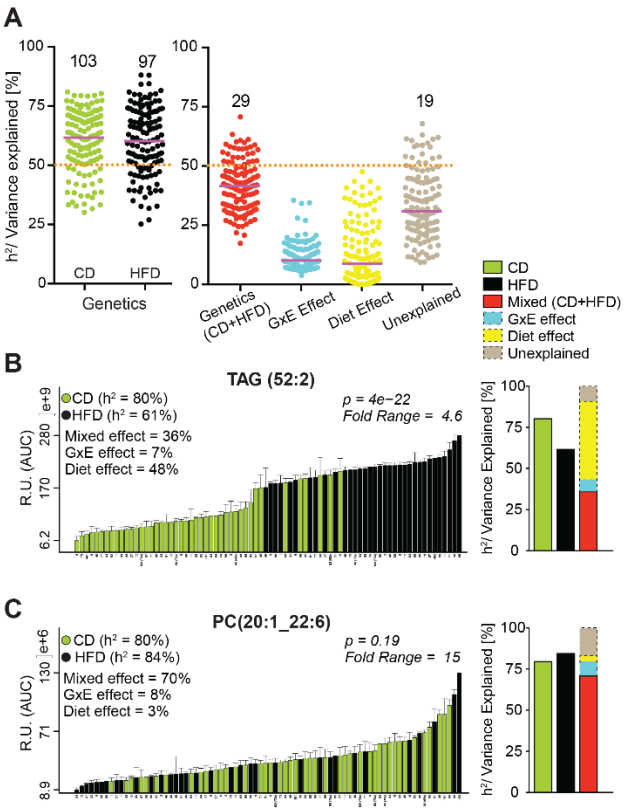


Figure 2. Plasma lipid species have high heritability.

(A) Heritability/variance explained of all lipid species. Number of lipids (out of 129) that have ≥ 50% of their variance explained by the factors along the x-axis is indicated. Purple line represents median variance explained.

(B and C) Example of two lipid species having high h^2 ($h^2 > 50\%$) in both diets but highly affected by diet (TAG(52:2)) (B), or unaffected by diet (PC(20:1_22:6)) (C). Bar plot showing the variation of the two lipids in the BXD population is shown on the left and the percentage of h^2/variance explained by the different factors is indicated in the graph as well as graphically represented on the right. Data are represented as ± SEM.

See all Table S3.

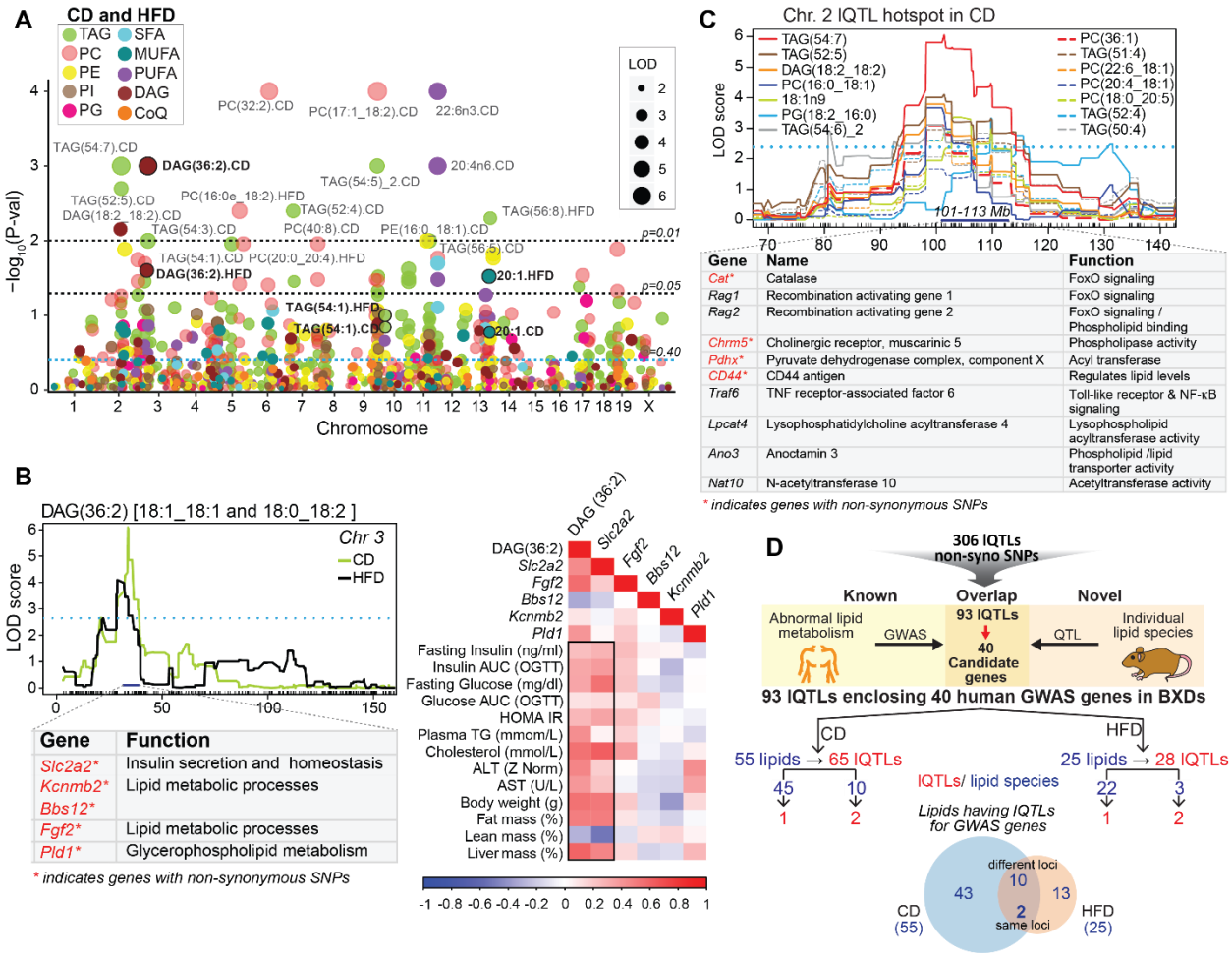


Figure 3. Plasma lipids are influenced by many genomic loci, including several associated lipid levels in human GWAS.

(A) Manhattan plot of lipid species. Lipids indicated in black bold font (DAG(36:3), TAG(54:1) and 20:1) have the same QTL position in CD and HFD. lQTLs with p-value < 0.01 are indicated on the plot. The black and blue dotted lines represent significant and suggestive QTL threshold, respectively.

(B) lQTLs for one of the three lipids (DAG(36:2)) having a QTL at the same locus in both diets (top-left). Select lQTL genes and their function are indicated below (bottom-left). Pearson correlation of DAG(36:2) alongside the expression of its liver lQTL genes in CD with metabolic syndrome phenotypes (right).

(C) Hotspot region comprising 14 lQTLs on chromosome 2. Genes associated with lipid metabolism in this region are indicated below.

(D) Genes with protein coding variants under 306 lQTLs were screened for any known association with blood lipids and associated metabolic traits in human GWAS. The screening identified 40 GWAS genes (with nsSNPs) from 93 lQTLs. 55/25 lipids contributed to 65/28 QTLs harboring GWAS genes in CD/HFD with 12 lipids in common across diet. 45/22 lipids mapped to one GWAS gene each in CD/HFD whereas, 10/3 lipids mapped to two different GWAS genes each.

See also Tables S4 and S5.

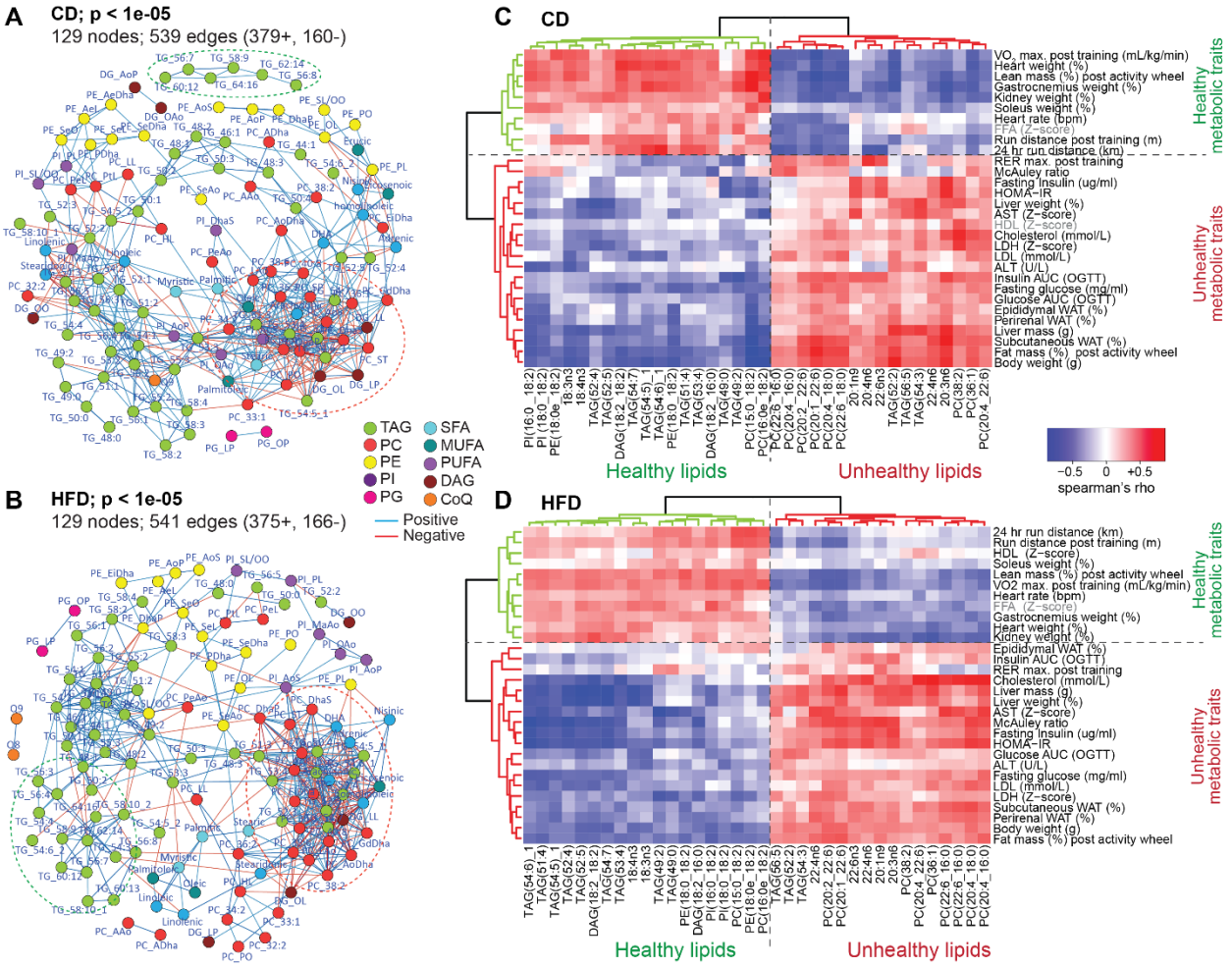


Figure 4. Association of lipid species with metabolic traits.

(A and B) Spearman correlation network of all lipid species in CD (A) and HFD (B). Lipid species are color coded as 10 major lipid classes.

(C and D) Heatmap with an unsupervised hierarchical clustering of Spearman's correlation rho value of 36 lipids with metabolic phenotypes. These 36 lipids show the same correlation trend with metabolic in both CD (C) and HFD (D). The horizontal green phenotype cluster represents healthy metabolic traits, whereas the red cluster represents unhealthy metabolic traits. The vertical green lipid cluster represents the healthy markers of metabolic health/fitness, whereas the red cluster represents the unhealthy markers of metabolic health/fitness. Table S6 provides the rho and p values for each lipid-phenotype correlation.

See also Figure S3 and Table S6.

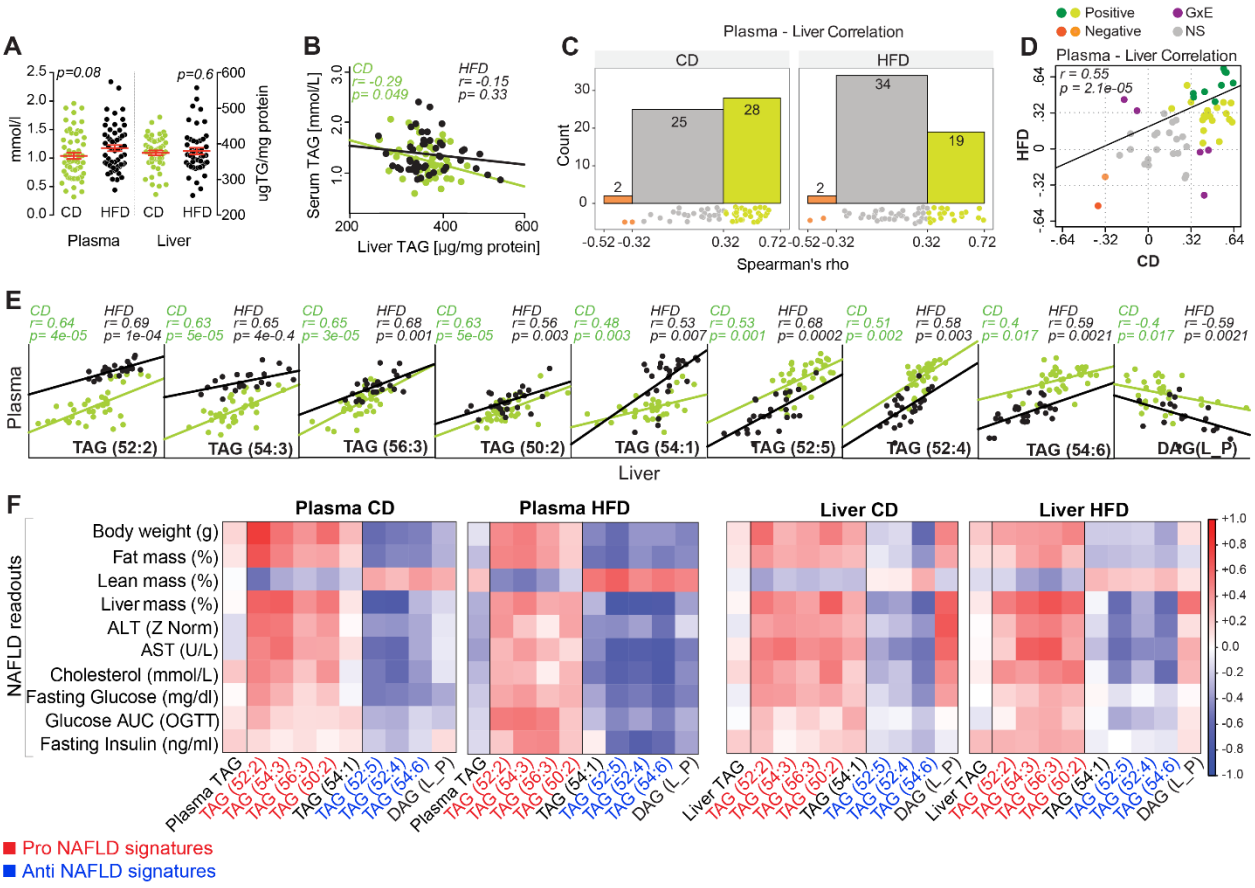


Figure 5. Identification of lipid species as markers of NAFLD.

(A and B) Plasma and liver TAG concentration (A) and correlation (B) in BXD cohorts. Data are represented as means ± SEM.

(C-D) 55 common lipid species between plasma and liver were correlated using Spearman's method. (C) Histogram of the rho correlation value of these 55 lipid pairs in CD (left) and HFD (right). (D) Correlation of rho values between CD and HFD from (C) to identify lipids, which behave similarly despite the dietary switch. Green dots indicate lipid species with positive correlation (wherein dark green dots are significant; $p < 0.05$), red and orange dot indicates lipids with negative correlation (wherein red dot is significant; $p < 0.05$). Purple dots indicate lipid species with opposite correlation in CD and HFD, reflective of GxE effect.

(E) Pearson correlation of nine significant lipid species identified in (D) (dark green and red dots) in liver and plasma.

(F) Correlation matrix showing the Pearson correlation of the 9 lipids, alongside plasma and liver TAG with NAFLD readouts.

See also Figure S4 and Table S7.

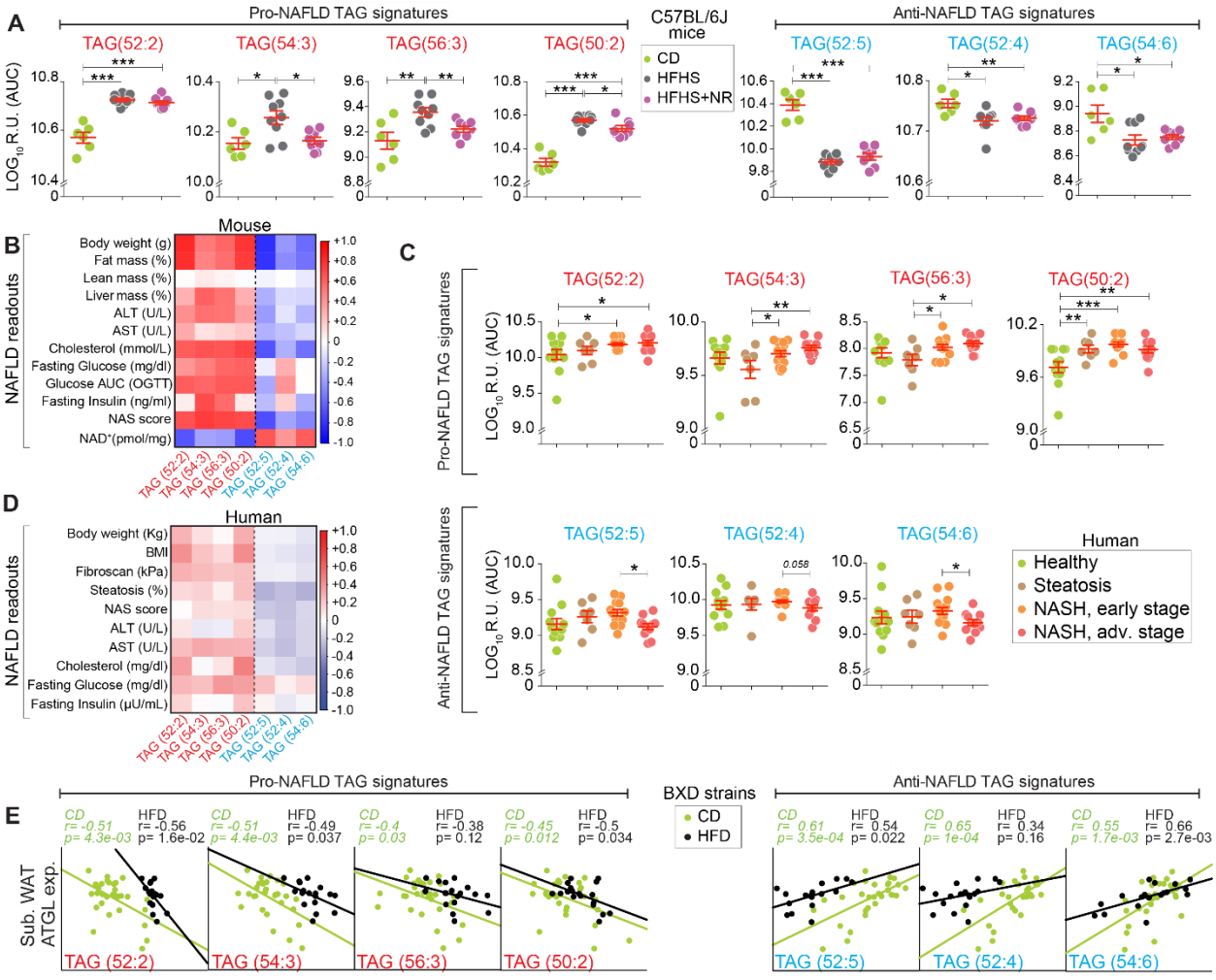


Figure 6. Validation of NAFLD TAG signatures in mice and humans.

(A) Levels of pro- and anti- NAFLD markers in C57BL/6J mice fed on CD, high-fat high-sucrose-diet (HFHS) and HFHS diet supplemented with NR, 9 weeks after the initiation of HFHS diet (NR). Data are represented as means ± SEM.

(B) Pearson correlation matrix of the pro- and anti-NAFLD markers with NAFLD readouts including the NAS score (NAFLD activity score) and liver NAD^+ levels.

(C) Levels of plasma NAFLD signatures in human subjects with various degrees of NAFLD. Data are represented as means ± SEM.

(D) Pearson correlation matrix of the NAFLD signatures with NAFLD readouts in human subjects.

(E) Pearson correlation of *Atgl* expression in subcutaneous WAT with pro- and anti-NAFLD

plasma TAG signatures in BXD strains. For mice: n=6-9 per group. For human subjects: n=12, healthy; n=7, steatosis; n=14, early stage NASH; n=11, advanced stage NASH. Differences in mean TAG species were compared using two-sample *t* tests. *p<0.05, **p<0.01, ***p<0.001. *See also Figure S5.*

LIPID	Diet	Chr : Mb	LOD	Gene @ chr, Mb	GWAS Phenotype	GWAS validated PMID
DAG(18:2_16:0)	CD	Chr1: 116.10	2.7	Insig2 @chr1, 121.30	TC, LDL-C	24097068
TAG[48:0, 50:0]	CD	Chr1: 172.17	3.3	Fmn2 @chr1, 174.50	TG	23063622
PC(32:2)	HFD	Chr1: 176.20	3.2			
PI(20:4_18:0)	CD	Chr2: 60.44	2.9	Ifih1 @chr2, 62.59	T1D	17554260
PC(22:6_16:0)	CD	Chr2: 60.44	2.9	Grb14 @chr2, 64.91	HDL-C, BMI	24097068, 25673412
PC(19:0_22:6)	CD	Chr2: 61.07	3.6	Cobll1 @chr2, 65.08	HDL-C, BMI	24097068, 25673412
DAG(18:2_18:2)	CD	Chr2: 98.18	3.8	Cd44 @chr2, 102.81	TG, TC, HDL-C	23063622
TAG(54:6)_2, PC(36:1)	CD	Chr2: 98.18	3.1			
PC(16:0_18:1)	CD	Chr2: 100.98	3.7			
TAG[51:4, 52:4, 52:5, 54:7]	CD	Chr2: 101.38	6.1			
PC(18:0_20:5)	CD	Chr2: 101.72	2.9			
Oleic acid	CD	Chr2: 104.01	3.4			
Eicosenoic acid	CD	Chr2: 115.49	3.2	Plcb2 @chr2, 118.70	TG	23063622
PE(18:1_18:2)	CD	Chr2: 115.50	2.9	Capn3 @chr2, 120.46	TG	20686565, 24097068
PE(16:0_18:2)	CD	Chr2: 115.87	4.2	Ubr1 @chr2, 120.86	TG	25961943
TAG(58:9)	CD	Chr2: 168.43	3.0	Plcg1 @chr2, 160.73	TC	27790247
TAG(54:6)_2	HFD	Chr2: 161.74	3.0	Cd40 @chr2, 165.05	TG	23063622
PC(16:0_18:1)	HFD	Chr2: 163.54	2.6			
DAG(18:1_18:1)	CD	Chr3: 28.72	4.1	Slc2a2 @chr3, 28.69	Fasting glucose	20081858
DAG(18:1_18:1)	HFD	Chr3: 33.75	6.1			
PC(36:2)	CD	Chr3: 134.24	2.8	Mttp @chr3, 138.089	TG	19060911
Nervonic acid	CD	Chr4: 125.58	2.8	Macf1 @chr4, 123.34	HDL-C	25961943
PC(18:0_16:0)	CD	Chr4: 128.85	3.0	Pigv @chr4, 133.66	TG, LDL-C,HDL-C	24097068, 28334899
TAG(48:3)	CD	Chr5: 24.30	2.7	Nos3 @chr5, 24.36	TG, CAD	23063622, 26343387
TAG(50:3)	CD	Chr5: 29.53	3.5			
Eicosenoic acid	HFD	Chr5: 21.20	3.3			
TAG(53:2)	CD	Chr5: 72.02	3.2	Corin @chr5, 72.30	TG	23063622
TAG(56:2)	CD	Chr5: 72.31	2.8			
TAG(51:1)	CD	Chr5: 73.34	2.6			
TAG[52:1, 54:1]	CD	Chr5: 73.63	4.2			
PC(17:1_18:2)	HFD	Chr5: 104.14	2.8	Klhl8 @chr5, 103.86	TG, HDL-C	28334899
PC(16:0e_20:4)	HFD	Chr5: 106.04	2.7			
PC(32:2)	CD	Chr6: 82.5	5.5	Alms1 @chr6, 85.58	TG, TC, HDL-C,	23063622
TAG(56:5)	CD	Chr6: 86.27	2.7		LDL-C	
TAG(53:3)	CD	Chr6: 87.14	2.6			
PE(20:0e_22:6)	CD	Chr7: 23.78	2.9	Apoe @chr7, 19.69	TC, HDL-C, LDL-C	25961943, 28371326
DAG(18:1_18:2), PC(34:2)	CD	Chr7: 24.50	3.1			
PC(16:0_18:1), PC(33:1)	HFD	Chr7: 17.13	3.0			
TAG(50:3)	HFD	Chr7: 19.25	2.7			
PC(20:0_22:6)	CD	Chr8: 30.89	3.1	Ppp1r3b @chr8, 35.37	TC, HDL-C,	25961943, 20686565,
PI(18:0_18:2)	HFD	Chr8: 30.62	2.7		LDL-C, PL	21829377
PE(18:0_18:2), TAG(49:2)	HFD	Chr8: 30.82	3.3			
TAG(58:4)	CD	Chr8: 43.37	2.8	Msr1 @chr8, 39.58	TG	23063622
PC(16:0_18:1)	CD	Chr8: 44.62	3.2			
PC(16:0e_20:4)	CD	Chr9: 99.75	2.7	Pccb @chr9, 100.98	TG, HDL-C	24097068, 25961943
PC(20:4_22:6), PC(38:2)	CD	Chr9: 110.75	3.3	Acad11 @chr9, 104.06	HDL-C, LDL-C	28334899, 24097068
PE(20:4_18:0)	HFD	Chr9: 105.94	2.9			
PC(32:2)	CD	Chr11: 73.26	2.8	Aloxe3 @chr11, 69.12	TG	23063622
PE[(16:0_18:1), (18:1_18:2)]	CD	Chr11: 73.26	4.3	Dlg4 @chr11, 70.017	TC, LDL-C	24097068
Nisinic acid	CD	Chr11: 73.64	3.0	Pld2 @chr11, 70.54	PL, TG	22359512, 25961943
TAG(54:2)	HFD	Chr11: 104.68	2.9	Mpp3 @chr11, 101.99	TG	24097068
				Cd300lg @chr11, 102.04	HDL-C	28270201
FA: [16:0, 20:4, 22:6]	CD	Chr11: 120.48	5.6	Srp68 @chr11, 116.24	TG	23063622
TAG[(54:6)_1, 51:4]	CD	Chr11: 120.48	2.6			
FA: [18:0, 20:3], PC(18:0_16:0)	CD	Chr11: 120.95	4.0			

TAG[50:4, 52:5, 48:3]	CD	Chr11: 120.95	3.1			
PC(20:4_18:0)	CD	Chr11: 120.95	2.6			
PC(19:0_22:6)	CD	Chr13: 39.32	2.9	Sycp2l @chr13, 41.11	PL	21829377
PC(20:2_22:6)	CD	Chr13: 95.82	2.8	Poc5 @chr13, 96.38	Obesity	23563607
DHA	CD	Chr13: 97.78	3.4	Ankdd1b @chr13, 96.41	TG	19060911
Eicosenoic acid	CD	Chr13: 99.68	3.2	Polk @chr13, 96.48	TC, LDL-C	28270201, 28334899
DGLA	HFD	Chr13: 91.20	2.9			
Nisinic acid	HFD	Chr13: 91.47	2.8			
PC(38:2)	HFD	Chr13: 97.51	2.9			
Eicosenoic acid	HFD	Chr13: 97.78	4.2			
PC(20:4_22:6)	HFD	Chr13: 99.86	2.8			
TAG(48:1)	HFD	Chr13: 101.86	3.1			
TAG(56:8)	HFD	Chr13: 102.64	3.7			
PE(20:2_22:6)	HFD	Chr13: 107.03	2.8	Map3k1 @chr13, 111.74	TG	28334899
PE[(18:0e_18:1), (18:0e_18:2)]	HFD	Chr13: 115.53	4.4			
PE(20:0e_18:2)	HFD	Chr13: 115.53	3.3			
TAG(52:3)	CD	Chr16: 10.70	2.6	Rmi2 @chr16, 10.83	HDL-C	25961943
				Pdxdc1 @chr16, 13.83	TG	28334899
TAG(56:8	HFD	Chr17: 31.93	2.6	Angptl4 @chr17, 33.77	TG, HDL-C	25961943
TAG(54:7)	HFD	Chr17: 33.11	3.1	Btnl2 @chr17, 34.35	Lipid levels	19936222
PE(18:0e_22:6)	CD	Chr19: 60.04	2.7	Pnliprp2 @chr19, 58.75	PL	22359512

GWAS phenotype abbreviation: TG, triglyceride; TC, total cholesterol; HDL-C, HDL cholesterol; LDL-C, LDL cholesterol; PL, phospholipid; BMI, body mass index; T1D: Type 1 diabetes; CAD, coronary artery disease

Table 1. lQTLs harboring genes associated with abnormal lipid metabolism in human GWAS.

Each box (27 loci) of the table represents the lQTL(s) in either diet, harboring the indicated 40 GWAS genes (having protein coding variants in BXDs) and the GWAS phenotype associated with the gene(s). Genes in red font are syntenic in mice and humans (6 loci). GWAS phenotype abbreviation: TG, triglyceride; TC, total cholesterol; HDL-C, HDL cholesterol; LDL-C, LDL cholesterol; PL, phospholipid; BMI, body mass index; T1D, type 1 diabetes; CAD, coronary artery disease. An extended version of this table is provided in Table S5.

See also Figure S2 and Table S5.

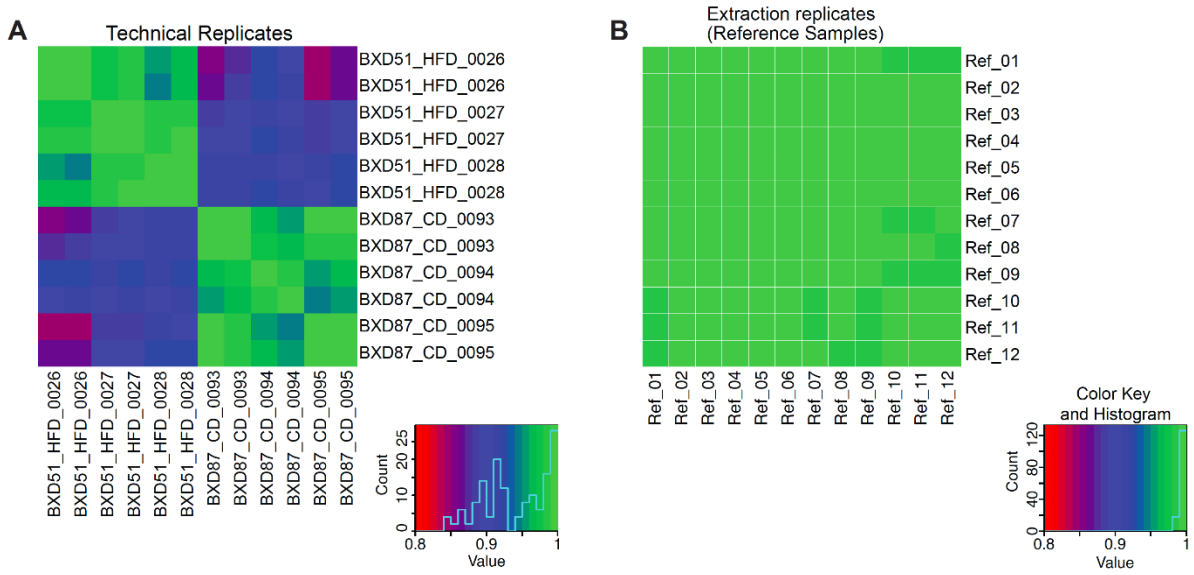


Figure S1. Quality assessment of MS measurements and reproducibility. Related to Figure 1.

(A) Heatmap of correlations between technical replicates (from separate runs) of six plasma samples showing high Spearman correlation between the replicates vs. between the strains.

(B) Heatmap of extraction replicates of the same liver sample extracted and run in 12 batches, equivalent to the 12 batches of the BXD samples.

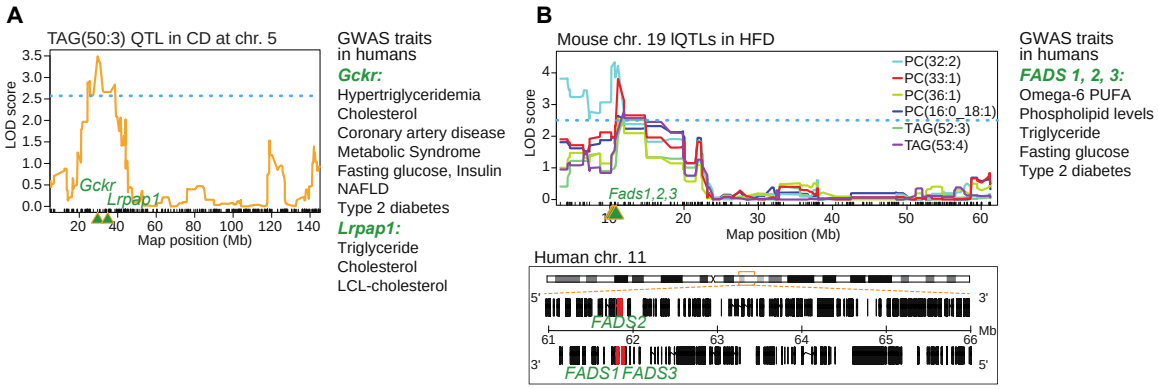


Figure S2. lQTLs harboring genes associated with abnormal lipid metabolism in human GWAS. Related to Figure 3.

(A) QTL locus for TAG(50:3) enclosing the human GWAS gene: Gckr and Lrpap1.

(B) QTL hotspot locus for the 6 lipid species on mouse Chr19 enclosing the GWAS genes: Fads 1, 2, 3. The syntenic region on human Chr11 is shown below. The genomic location of the genes is shown in green in the positive stand for FADS2 and in the negative strand for FADS1 and FADS3. The GWAS phenotypes associated with the genes is indicated on the right. Blue dotted line represents the suggestive QTL threshold.

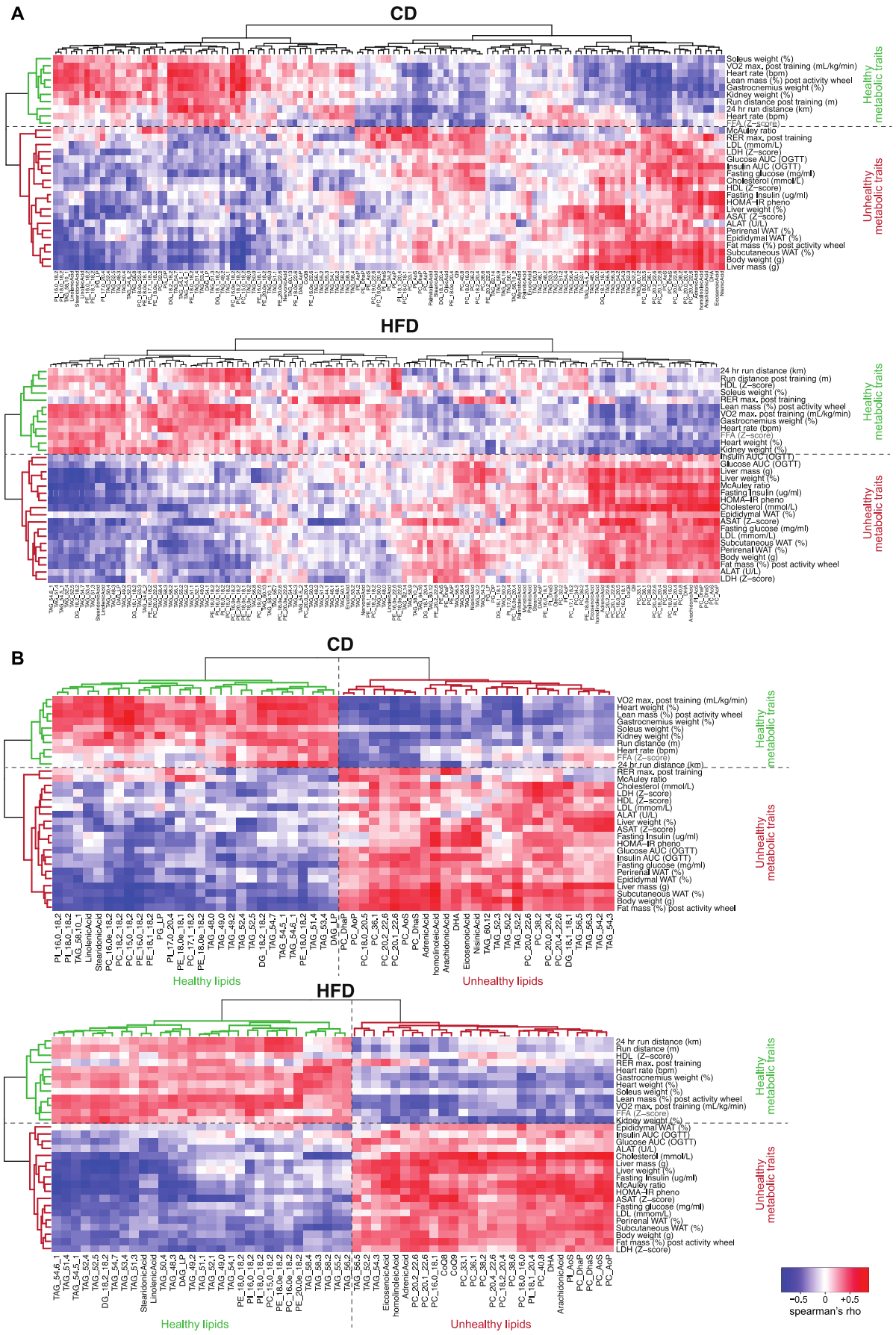

A
CD
Soleus weight (%)
VO2 max. post training (mL/kg/min)
Heart rate (bpm)
Lean mass (%) post activity wheel
Gastrocnemius weight (%)
Kidney weight (%)
Run distance post training (m)
24 hr run distance (km)
Heart rate (bpm)
FFA (Z−score)
McAuley ratio
RER max. post training
LDL (mmom/L)
LDH (Z−score)
Glucose AUC (OGTT)
Insulin AUC (OGTT)
Fasting glucose (mg/ml)
Cholesterol (mmol/L)
HDL (Z−score)
Fasting Insulin (ug/ml)
HOMA−IR pheno
Liver weight (%)
ASAT (Z−score)
ALAT (U/L)
Perirenal WAT (%)
Epididymal WAT (%)
Fat mass (%) post activity wheel
Subcutaneous WAT (%)
Body weight (g)
Liver mass (g)
Healthy metabolic traits
Unhealthy metabolic traits
HFD
24 hr run distance (km)
Run distance post training (m)
HDL (Z−score)
Soleus weight (%)
RER max. post training
Lean mass (%) post activity wheel
VO2 max. post training (mL/kg/min)
Gastrocnemius weight (%)
Heart rate (bpm)
FFA (Z−score)
Heart weight (%)
Kidney weight (%)
Insulin AUC (OGTT)
Glucose AUC (OGTT)
Liver mass (g)
Liver weight (%)
McAuley ratio
Fasting Insulin (ug/ml)
HOMA−IR pheno
Cholesterol (mmol/L)
Epididymal WAT (%)
ASAT (Z−score)
Fasting glucose (mg/ml)
LDL (mmom/L)
Subcutaneous WAT (%)
Perirenal WAT (%)
Body weight (g)
Fat mass (%) post activity wheel
ALAT (U/L)
LDH (Z−score)
Healthy metabolic traits
Unhealthy metabolic traits
B
CD
VO2 max. post training (mL/kg/min)
Heart weight (%)
Lean mass (%) post activity wheel
Gastrocnemius weight (%)
Soleus weight (%)
Kidney weight (%)
Run distance (m)
Heart rate (bpm)
FFA (Z−score)
24 hr run distance (km)
RER max. post training
McAuley ratio
Cholesterol (mmol/L)
LDH (Z−score)
HDL (Z−score)
LDL (mmom/L)
ALAT (U/L)
Liver weight (%)
ASAT (Z−score)
Fasting Insulin (ug/ml)
HOMA−IR pheno
Glucose AUC (OGTT)
Insulin AUC (OGTT)
Fasting glucose (mg/ml)
Perirenal WAT (%)
Epididymal WAT (%)
Liver mass (g)
Subcutaneous WAT (%)
Body weight (g)
Fat mass (%) post activity wheel
Healthy metabolic traits
Unhealthy metabolic traits
PL_16:0_18:2
PL_18:0_18:2
TAG_58:10_1
LinolenicAcid
StearidonicAcid
PC_16:0e_18:2
PC_18:2_18:2
PC_15:0_18:2
PE_16:0_18:2
PE_18:1_18:2
PG_LP
PL_17:0_20:4
PE_18:0e_18:1
PC_17:1_18:2
PE_18:0e_18:2
TAG_48:0
TAG_49:0
TAG_49:2
TAG_52:4
TAG_52:5
DG_18:2_18:2
TAG_54:7
TAG_54:5_1
TAG_54:6_1
PE_18:0_18:2
TAG_51:4
TAG_53:4
DAG_LP
PC_DhaP
PC_AoP
PC_18:0_20:5
PC_36:1
PC_20:2_22:6
PC_20:1_22:6
PC_AoS
PC_DhaS
AdrenicAcid
homolinoleicAcid
ArachidonicAcid
DHA
EicosenoicAcid
NisinicAcid
TAG_60:12
TAG_52:3
TAG_50:2
TAG_52:2
PC_20:0_22:6
PC_38:2
PC_20:0_20:4
PC_20:4_22:6
DG_18:1_18:1
TAG_56:5
TAG_56:3
TAG_54:2
TAG_54:3
Healthy lipids
Unhealthy lipids
HFD
24 hr run distance (km)
Run distance (m)
HDL (Z−score)
RER max. post training
Heart rate (bpm)
Gastrocnemius weight (%)
Heart weight (%)
Soleus weight (%)
Lean mass (%) post activity wheel
VO2 max. post training (mL/kg/min)
FFA (Z−score)
Kidney weight (%)
Epididymal WAT (%)
Insulin AUC (OGTT)
Glucose AUC (OGTT)
ALAT (U/L)
Cholesterol (mmol/L)
Liver mass (g)
Liver weight (%)
Fasting Insulin (ug/ml)
McAuley ratio
HOMA−IR pheno
ASAT (Z−score)
Fasting glucose (mg/ml)
LDL (mmom/L)
Perirenal WAT (%)
Subcutaneous WAT (%)
Body weight (g)
Fat mass (%) post activity wheel
LDH (Z−score)
Healthy metabolic traits
Unhealthy metabolic traits
TAG_54:6_1
TAG_51:4
TAG_54:5_1
TAG_52:4
TAG_52:5
DG_18:2_18:2
TAG_54:7
TAG_53:4
TAG_51:3
StearidonicAcid
LinolenicAcid
TAG_50:4
TAG_48:3
DAG_LP
TAG_49:2
TAG_52:1
TAG_49:0
TAG_54:1
PE_18:0_18:2
PL_16:0_18:2
PL_18:0_18:2
PC_16:0_18:2
PC_18:0e_18:2
PE_18:0_18:2
PE_20:0e_18:2
TAG_58:4
TAG_58:3
TAG_58:2
TAG_55:2
TAG_56:2
TAG_56:5
TAG_52:2
TAG_54:3
EicosenoicAcid
homolinoleicAcid
AdrenicAcid
PC_20:2_22:6
PC_20:1_22:6
PC_16:0_18:1
CoQ8
CoQ9
PC_33:1
PC_36:1
PC_38:2
PC_20:4_22:6
PC_18:2_20:4
PC_38:6
PC_18:0_16:0
PL_18:1_20:4
PC_40:8
DHA
ArachidonicAcid
PL_AoS
PC_DhaS
PC_DhaP
PC_AoS
PC_AoP
Healthy lipids
Unhealthy lipids
−0.5
0
+0.5
spearman's rho

Figure S3. Association of lipid species with metabolic traits. Related to Figure 4 and Table S6.

(A and B) Heatmap with an unsupervised hierarchical clustering of Spearman's correlation rho of (A) all the lipid species and metabolic traits in CD (top) and HFD (bottom); (B) select lipid species in each diet which show maximum number of significant correlation with metabolic traits (the lateral lipid species cluster of panel A). For panel B, the vertical green lipid cluster represents healthy lipids (specific and common in each diet), whereas the vertical red cluster represents the unhealthy lipids (specific and common in each diet). For A and B the horizontal green phenotype cluster represents healthy metabolic traits whereas the red cluster represents unhealthy metabolic traits. Red indicates positive correlation and blue negative correlation.

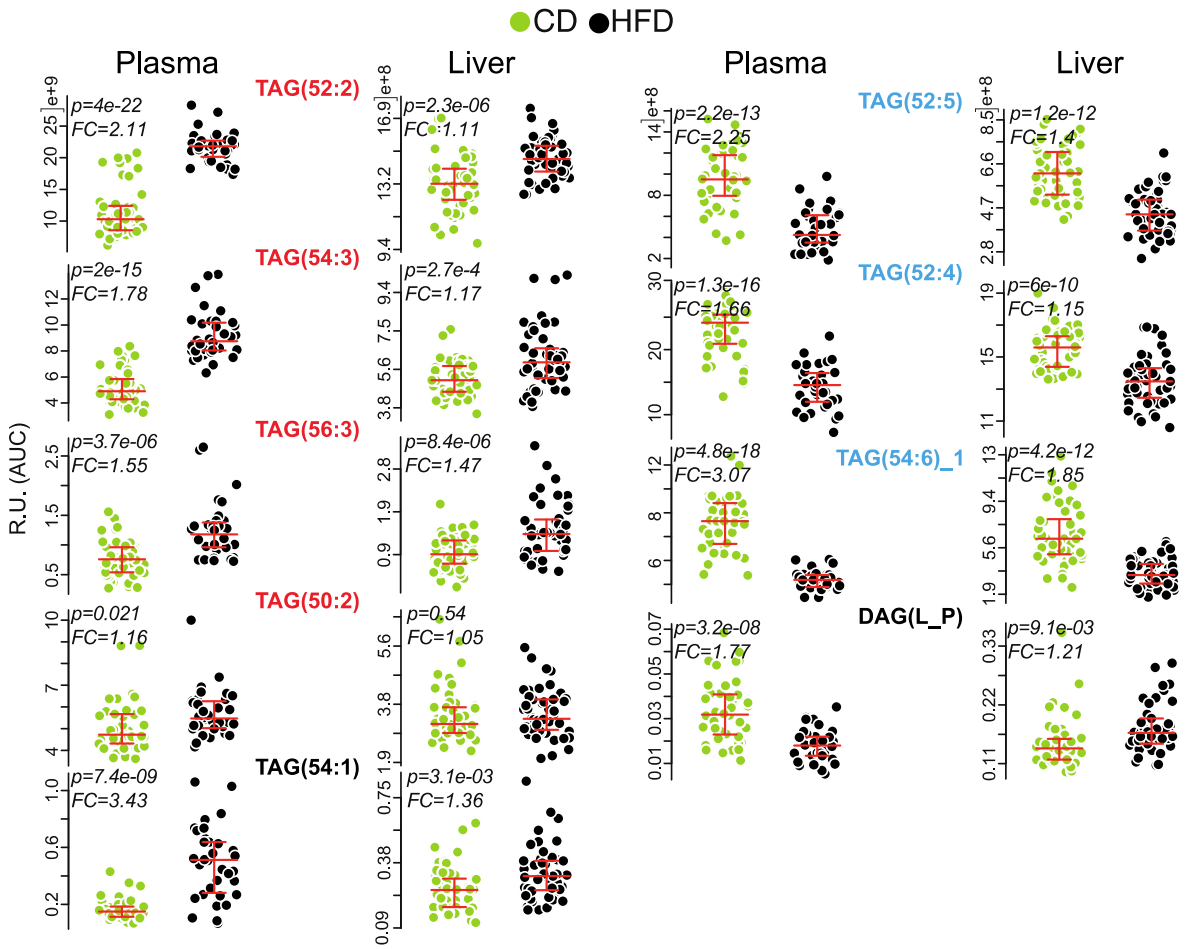


Figure S4. Plasma and liver levels of 9 TAG species having the most significant correlation between plasma and liver in both diets. Related to Figure 5 and Table S2.

Dot plot showing the levels of 9 shortlisted lipids as potential lipid signatures of NAFLD in plasma and liver. Lipids in red font indicate pro-NAFLD signatures and those in blue font indicate anti-NAFLD signatures. P-value < 0.05 by Welch's t-test was considered significant between diets.

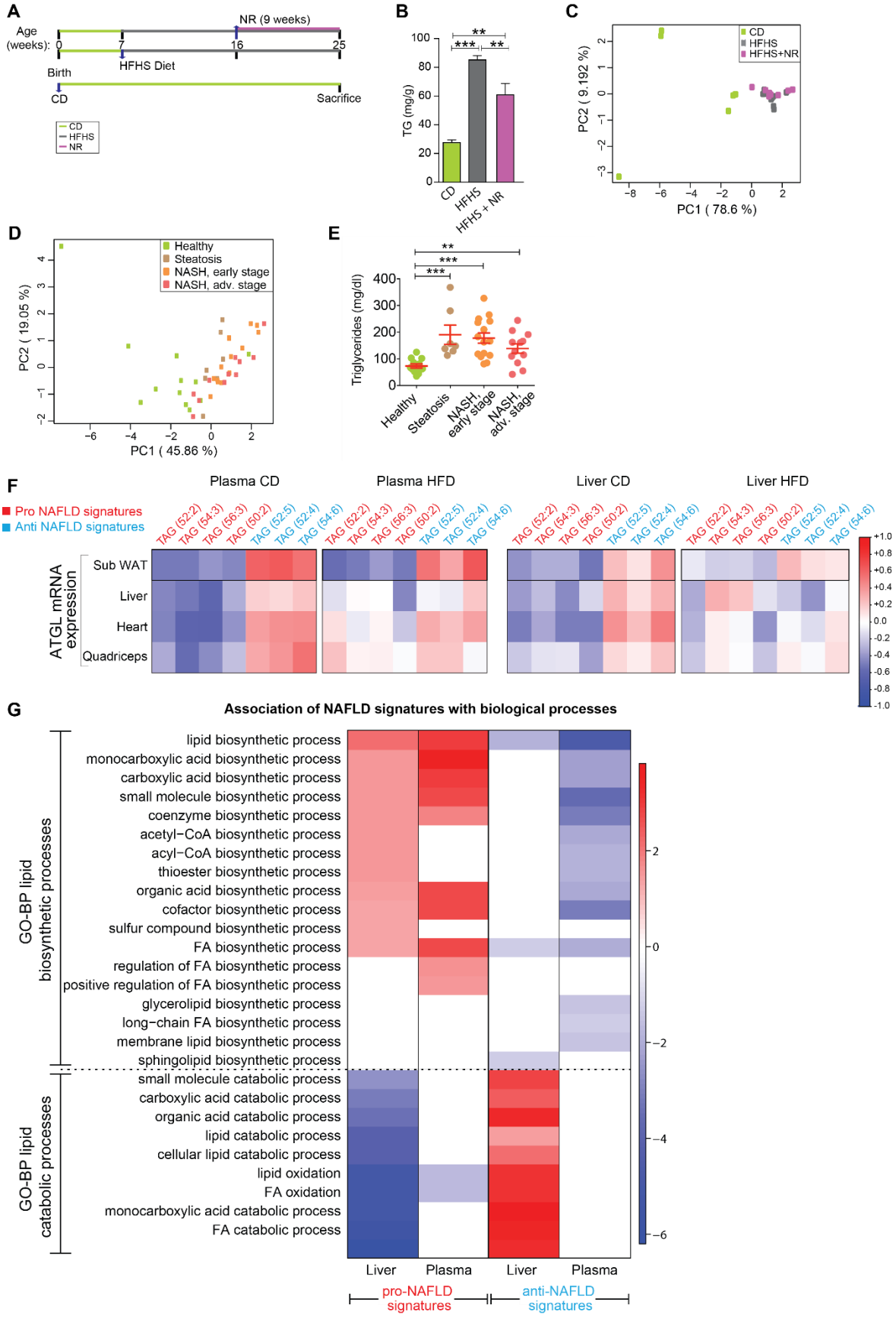

A
Age (weeks):
NR (9 weeks)
HFHS Diet
Birth
CD
Sacrifice
CD
HFHS
NR
B
TG (mg/g)
CD
HFHS
HFHS + NR
C
PC2 (9.192 %)
PC1 (78.6 %)
CD
HFHS
HFHS+NR
D
PC2 (19.05 %)
PC1 (45.86 %)
Healthy
Steatosis
NASH, early stage
NASH, adv. stage
E
Triglycerides (mg/dl)
Healthy
Steatosis
NASH, early stage
NASH, adv. stage
F
Plasma CD
Plasma HFD
Liver CD
Liver HFD
Pro NAFLD signatures
Anti NAFLD signatures
TAG (52:2)
TAG (54:3)
TAG (56:3)
TAG (50:2)
TAG (52:5)
TAG (52:4)
TAG (54:6)
ATGL mRNA expression
Sub WAT
Liver
Heart
Quadriceps
G
Association of NAFLD signatures with biological processes
GO-BP lipid biosynthetic processes
lipid biosynthetic process
monocarboxylic acid biosynthetic process
carboxylic acid biosynthetic process
small molecule biosynthetic process
coenzyme biosynthetic process
acetyl−CoA biosynthetic process
acyl−CoA biosynthetic process
thioester biosynthetic process
organic acid biosynthetic process
cofactor biosynthetic process
sulfur compound biosynthetic process
FA biosynthetic process
regulation of FA biosynthetic process
positive regulation of FA biosynthetic process
glycerolipid biosynthetic process
long−chain FA biosynthetic process
membrane lipid biosynthetic process
sphingolipid biosynthetic process
GO-BP lipid catabolic processes
small molecule catabolic process
carboxylic acid catabolic process
organic acid catabolic process
lipid catabolic process
cellular lipid catabolic process
lipid oxidation
FA oxidation
monocarboxylic acid catabolic process
FA catabolic process
Liver
Plasma
Liver
Plasma
pro-NAFLD signatures
anti-NAFLD signatures

Figure S5. Assessment of TAG NAFLD signatures in mice and humans and their association with biosynthetic pathways. Related to Figure 6.

(A-C) C57BL/6J mice: (A) Schematic illustration of the three experimental groups used for validation of the NAFLD signatures; mice on CD (green), mice on HFHS diet from 7-25 weeks (grey), mice starting HFHS diet at 7 weeks and treated with nicotinamide riboside (NR) 9 weeks after the start of HFHS diet until the end of the study (for 9 weeks - therapeutic intervention, magenta). (B) Liver total TAG concentration (normalized to liver weight). (C) Principal component analysis (PCA) of 55 TAG species shows clear separation of only the CD group on PC1.

(D and E) Human subjects: (D) PCA of 55 TAG species measured in human plasma from healthy, steatosis and NASH patients. (E) Plasma total TAG levels in human samples.

(F and G) BXDs: (F) Correlation matrix showing the Pearson correlation of Atgl mRNA expression in four different metabolic tissues with the pro- and anti-NAFLD TAG signatures in plasma (left) and liver (right) for both diets. (G) Gene ontology biological processes (GO BP) associated with NAFLD TAG signatures. Pathway enrichment analysis was performed with the liver transcripts that significantly correlated (both positively and negatively) with the PC1 of pro- and antiNAFLD signatures in liver and plasma. Red and blue cells represent the enriched pathways with the positively (scale bar: $\log_{10}$ p-value) and negatively (scale bar: $-\log_{10}$ p-value) correlated liver transcripts respectively. For B and E differences in mean TAG levels were compared using two-sample t tests. *p<0.001.

*All supplemental tables are available with the published manuscript at DOI:

10.1016/j.cels.2018.05.009

References

Belknap, J.K. (1998). Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. Behavior genetics *28*, 29-38.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics *19*, 889-890.

Browning, J.D., Szczepaniak, L.S., Dobbins, R., Nuremberg, P., Horton, J.D., Cohen, J.C., Grundy, S.M., and Hobbs, H.H. (2004). Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. Hepatology *40*, 1387-1395.

Dewey, F.E., Murray, M.F., Overton, J.D., Habegger, L., Leader, J.B., Fetterolf, S.N., O'Dushlaine, C., Van Hout, C.V., Staples, J., Gonzaga-Jauregui, C., et al. (2016). Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science *354*.

Diabetes Genetics Initiative of Broad Institute of, H., Mit, L.U., Novartis Institutes of BioMedical, R., Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., et al. (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science *316*, 1331-1336.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics *42*, 105-116.

Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Jr., Murphy, R.C., Raetz, C.R., Russell, D.W., Seyama, Y., Shaw, W., et al. (2005). A comprehensive classification system for lipids. Journal of lipid research *46*, 839-861.

Farese, R.V., Jr., Zechner, R., Newgard, C.B., and Walther, T.C. (2012). The problem of establishing relationships between hepatic steatosis and hepatic insulin resistance. Cell metabolism *15*, 570-573.

Gariani, K., Menzies, K.J., Ryu, D., Wegner, C.J., Wang, X., Ropelle, E.R., Moullan, N., Zhang, H., Perino, A., Lemos, V., et al. (2016). Eliciting the mitochondrial unfolded protein response by nicotinamide adenine dinucleotide repletion reverses fatty liver disease in mice. Hepatology *63*, 1190-1204.

Global Lipids Genetics, C., Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nature genetics *45*, 1274-1283.

Haemmerle, G., Lass, A., Zimmermann, R., Gorkiewicz, G., Meyer, C., Rozman, J., Heldmaier, G., Maier, R., Theussl, C., Eder, S., et al. (2006). Defective lipolysis and altered energy metabolism in mice lacking adipose triglyceride lipase. Science *312*, 734-737.

Han, X. (2016). Lipidomics for studying metabolism. Nature reviews. Endocrinology *12*, 668-679.

Hui, S.T., Parks, B.W., Org, E., Norheim, F., Che, N., Pan, C., Castellani, L.W., Charugundla, S., Dirks, D.L., Psychogios, N., et al. (2015). The genetic architecture of NAFLD among inbred strains of mice. eLife *4*, e05607.

Jha, P., Claudel, T., Baghdasaryan, A., Mueller, M., Halilbasic, E., Das, S.K., Lass, A., Zimmermann, R., Zechner, R., Hoefler, G., et al. (2014). Role of adipose triglyceride lipase (PNPLA2) in protection from hepatic inflammation in mouse models of steatohepatitis and endotoxemia. Hepatology *59*, 858-869.

Jha, P., McDevitt, M.T., Gupta, R., Quiros, P.M., Williams, E.G., Gariani, K., Sleiman, M.B., Diserens, L., Jochem, A., Ulbrich, A., et al. (2018). Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species. Cell systems *6*, 722-733.e726.

Johansen, C.T., Kathiresan, S., and Hegele, R.A. (2011). Genetic determinants of plasma triglycerides. Journal of lipid research *52*, 189-206.

Kang, H.M., Ahn, S.H., Choi, P., Ko, Y.A., Han, S.H., Chinga, F., Park, A.S., Tao, J., Sharma, K., Pullman, J., et al. (2015). Defective fatty acid oxidation in renal tubular epithelial cells has a key role in kidney fibrosis development. Nature medicine *21*, 37-46.

Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nature genetics *41*, 56-65.

Kirk, E.A., Moe, G.L., Caldwell, M.T., Lernmark, J.A., Wilson, D.L., and LeBoeuf, R.C. (1995). Hyper- and hypo-responsiveness to dietary fat and cholesterol among inbred mice: searching for level and variability genes. Journal of lipid research *36*, 1522-1532.

Kleiner, D.E., Brunt, E.M., Van Natta, M., Behling, C., Contos, M.J., Cummings, O.W., Ferrell, L.D., Liu, Y.C., Torbenson, M.S., Unalp-Arida, A., et al. (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology *41*, 1313-1321.

Li, M.J., Liu, Z., Wang, P., Wong, M.P., Nelson, M.R., Kocher, J.P., Yeager, M., Sham, P.C., Chanock, S.J., Xia, Z., et al. (2016). GWASdb v2: an update database for human genetic variants identified by genome-wide association studies. Nucleic acids research *44*, D869-876.

Lin, X., Yue, P., Chen, Z., and Schonfeld, G. (2005). Hepatic triglyceride contents are genetically determined in mice: results of a strain survey. American journal of physiology. Gastrointestinal and liver physiology *288*, G1179-1189.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic acids research *45*, D896-d901.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Puri, P., Wiest, M.M., Cheung, O., Mirshahi, F., Sargeant, C., Min, H.K., Contos, M.J., Sterling, R.K., Fuchs, M., Zhou, H., et al. (2009). The plasma lipidomic signature of nonalcoholic steatohepatitis. Hepatology *50*, 1827-1838.

Quehenberger, O., and Dennis, E.A. (2011). The human plasma lipidome. N Engl J Med *365*, 1812-1823.

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

Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L.A., Boerwinkle, E., Cohen, J.C., and Hobbs, H.H. (2008). Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nature genetics *40*, 1461-1465.

Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., et al. (2014). An atlas of genetic influences on human blood metabolites. Nature genetics *46*, 543-550.

Sittig, L.J., Carbonetto, P., Engel, K.A., Krauss, K.S., Barrios-Camacho, C.M., and Palmer, A.A. (2016). Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. Neuron *91*, 1253-1259.

Speliotes, E.K., Yerges-Armstrong, L.M., Wu, J., Hernaez, R., Kim, L.J., Palmer, C.D., Gudnason, V., Eiriksdottir, G., Garcia, M.E., Launer, L.J., et al. (2011). Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS genetics *7*, e1001324.

Stefely, J.A., Kwiecien, N.W., Freiberger, E.C., Richards, A.L., Jochem, A., Rush, M.J.P., Ulbrich, A., Robinson, K.P., Hutchins, P.D., Veling, M.T., et al. (2016). Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. Nature biotechnology *34*, 1191-1197.

van der Vusse, G.J., van Bilsen, M., and Glatz, J.F. (2000). Cardiac fatty acid uptake and transport in health and disease. Cardiovascular research *45*, 279-293.

Wang, X., Pandey, A.K., Mulligan, M.K., Williams, E.G., Mozhui, K., Li, Z., Jovaisaite, V., Quarles, L.D., Xiao, Z., Huang, J., et al. (2016). Joint mouse-human phenome-wide association to test gene function and disease risk. Nature communications *7*, 10464.

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nature genetics *45*, 1274-1283.

Williams, E.G., and Auwerx, J. (2015). The Convergence of Systems and Reductionist Approaches in Complex Trait Analysis. Cell *162*, 23-32.

Williams, E.G., Wu, Y.B., Jha, P., Dubuis, S., Blattmann, P., Argmann, C.A., Houten, S.M., Amariuta, T., Wolski, W., Zamboni, N., et al. (2016). Systems proteomics of liver mitochondria function. Science *352*, 1292-+.

Wong, G., Barlow, C.K., Weir, J.M., Jowett, J.B., Magliano, D.J., Zimmet, P., Shaw, J., and Meikle, P.J. (2013). Inclusion of plasma lipid species improves classification of individuals at risk of type 2 diabetes. PLoS One *8*, e76577.

Wu, Y., Williams, E.G., Dubuis, S., Mottis, A., Jovaisaite, V., Houten, S.M., Argmann, C.A., Faridi, P., Wolski, W., Kutalik, Z., et al. (2014). Multilayered genetic and omics dissection of mitochondrial activity in a mouse reference population. Cell *158*, 1415-1430.

Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology *16*, 284-287.

Zhang, W., Korstanje, R., Thaisz, J., Staedtler, F., Harttman, N., Xu, L., Feng, M., Yanas, L., Yang, H., Valdar, W., et al. (2012). Genome-wide association mapping of quantitative traits in outbred mice. G3 *2*, 167-174.

Chapter 4: Exploration of the esterase activity of PREPL, a protein implicated in hypotonia-cystinuria

Molly T. McDevitt, Brendan J. Floyd, Brendan K. Dolan, Adam Jochem, Brett Paulson, Thiru Reddy, Tina M. Misenheimer, Russell L. Wrobel, Craig A. Bingman, Jason D. Russell, Joshua J. Coon, and David J. Pagliarini

Summary

Deletions on chromosome 2p21 are associated with hypotonia-cystinuria syndrome (HCS), a rare disease characterized by low muscle tone (hypotonia), growth impairment, and buildup of cysteine in the urine. Initial patients were found to have a microdeletion of a portion of the genes *SLC3A1*, which produces a subunit of an amino acid transporter, and *PREPL*, a putative peptidase of unknown function. Recent studies suggest that loss of functional PREPL alone causes the hypotonia associated with HCS. However, the mechanism by which PREPL causes the disease remains unknown, in large part because, to date, it has no known function. Though sequence homology suggests that PREPL is a peptidase no peptide substrate has been identified. Interestingly, though it shares the highest homology with prolyl endopeptidase (PREP), other members of the family include peptidases, as well as esterases and lipases. Strikingly, we find that though PREPL was not active against a peptide substrate, it exhibited specific esterase activity against 4-nitrophenyl octonoate, dependent on the nucleophilic serine in the active site. Though the activity was reproducible, it was small, making it unlikely that 4-nitrophenyl octonoate was the substrate *in vivo*. Convinced that PREPL cleaves ester bonds, but uncertain if the *in vivo* substrate was a small molecule ester or a larger molecule such as a lipid, we conducted lipidomics and metabolomics analyses in PREPL knockout mice and cell culture. Indeed, loss of PREPL caused alterations in the lipid profiles compared to wild type; however, our current lipid search methods could not determine the specific identity of the lipids being altered. Future work will focus on unambiguous lipid identification and biological follow-up by examining structural similarities among lipids specifically altered in the knockout models and *in vitro* assays to determine if alterations in these candidate lipids are in direct consequence to lack of active PREPL.

Introduction

Hypotonia-cystinuria syndrome (HCS) is rare disease associated with low muscle tone, slow growth, large amounts of cysteine in the urine, and overall failure to thrive (Jaeken et al., 2006; Lone et al., 2014; Regal et al., 2014). In affected individuals, sequencing revealed various microdeletions of *SLC3A1* and *PREPL* genes on chromosome 2 (**Figure 1A**). Defects in one of these, SLC3A1 are known to cause cystinuria. SLC3A1 is a component of an amino acid transporter and defects inhibit the reabsorption of amino acids, especially cysteine, causing excess cysteine to be excreted in urine (Calonge et al., 1994; Font-Llitjos et al., 2005; Goodyer, 2004). Since expression of the flanking genes (*PPM1B* and *CAMKMT*) was unchanged in the initial characterization of HCS (Jaeken et al., 2006), the additional symptoms were attributed to the deletion of PREPL. The role of *SLC3A1* has since been confirmed when cystinuria was absent in individuals exhibiting only deletions in *PREPL and CAMKMT* (Bartholdi et al., 2013). More severe forms of the disease have been associated with additional deletions in *CAMKMT* (atypical HCS, Figure 1A) and both flanking genes, *PPM1B* and *CAMKMT*, (2p21 deletion syndrome, Figure 1A) (Chabrol et al., 2008; Parvari et al., 2005).

In order to study the role of PREPL in the non-cystinuria symptoms of HCS, a mouse model consisting of a full-body PREPL null ($PREPL^{-/-}$) mouse was created (Lone et al., 2014). $PREPL^{-/-}$ mice suffered minor growth defects and hypotonia. Moreover, around the same time, an individual with mutation that lead to loss of PREPL, but did not affect neighboring genes, was described, with subsequent work showing that similar mutations cause hypotonia and growth hormone deficiency (Regal et al., 2014; Silva et al., 2018). These results simultaneously and separately support the hypothesis that deletion of *PREPL* in humans is responsible for the same growth defect and hypotonia symptoms identified in mice. However, as the function of PREPL has yet to be elucidated, the role PREPL in patients with HCS is difficult to establish and has yet to be defined.

First identified in brain, due to its strong reaction with a biotinylated fluorophosphonate, (FP-biotin), an irreversible inhibitor of serine hydrolases (Liu et al., 1999), prolyl endopeptidase-like protein (PREPL) gets its name due to the sequence homology it shares with prolylendopeptidase (PREP) and oligopeptidase B, which are both well-studied peptidases. Although it is ubiquitously expressed, expression levels are highest in the brain, and to a lesser extent skeletal muscle, heart, and kidney (Jaeken et al., 2006). Although sequence homology between PREPL and PREP and oligopeptidase B low (29% and 44%, respectively), they are the closest homologs by primary sequence. Although, like its homologs, PREPL is thought to be a peptidase, no peptidase activity has been observed, despite multiple efforts (Boonen et al., 2011; Szeltner et al., 2005).

Like PREP, PREPL is a member of the S9 family of serine hydrolases, characterized by the presence of a catalytic triad consisting of an aspartate (D), serine (S) and histidine (H). Although no peptidase activity has been found for PREPL, the catalytic triad is conserved (Martens et al., 2006). Furthermore, its ability to bind FP-biotin indicates that the catalytic serine can still act as a nucleophile, suggesting that, at least *in vivo*, the protein is likely active. Intriguingly, the S9 family of serine hydrolases is not limited to peptidases. Other family members include esterases/lipases (*PAFAH2, PLA2G7*) (Rice et al., 1998; Stafforini et al., 1987) and one family member, APEH, has even been found to have both esterase and peptidase activity. Taken together, this data suggests that despite inability to identify a peptide substrate, PREPL is likely an active protein – perhaps acting as an esterase or lipase instead of a peptidase.

PREPL has two main protein isoforms, short ($PREPL_S$) and long ($PREPL_L$), which has an additional 89 amino acids (Figure 1B). These two proteins arise from seven distinct mRNA isoforms (Figure 1C), with isoforms 1-4 giving rise to the 638 aa $PREPL_S$, while 5-7 generates the 727 aa $PREPL_L$ isoform. The two proteins are identical except for the 89 additional amino acids on the N-terminus of the long isoform. Interestingly, these 89 amino acids contain a mitochondrial

localization sequence (MLS), while $PREPL_S$ is cytosolic (Radhakrishnan et al., 2013). Both proteins maintain the catalytic triad, and it is therefore possible that they carry out the same function in both the cytosol and mitochondria. However, that is difficult to ascertain given that its enzymatic activity is, as yet, undefined.

Although based on primary sequence PREPL looks similar to PREP, identification of a peptide substrate remains elusive despite many intense efforts. Moreover, based on these efforts it is possible that its substrate is not a peptide at all. Here we use a combination of structure predictions, *in vitro* activity assays, lipidomics, and metabolomics investigate the functional role of PREPL. Our data show that PREPL has structural similarity to PREP and other S9 serine hydrolases, that PREPL exhibits esterase activity *in vitro*, that $PREPL_L$ localizes to the mitochondria, and that $PREPL^{-/-}$ mice have altered brain mitochondrial lipid profiles. At present, our data suggest that PREP and PREPL have distinct functions in the cell, and that PREPL's esterase activity may play an important role in growth and development that could contribute to its role in HCS. Future work will seek to identify specific lipid components that may be PREPL substrates, elucidate the roles played by $PREPL_S$ and $PREPL_L$ in altering lipid profiles, and determine how these functions contribute to HCS etiology.

Results

PREPL exhibits sequence similarity to other members of the S9 family and maintains the catalytic triad

A sequence alignment reveals that PREPL shows substantial sequence similarity to members of the S9 family of serine hydrolases (Figure 2A). Like other members of the serine hydrolase superfamily, the S9 family is characterized by a catalytic triad consisting of a serine, aspartate, and histidine. The nucleophilic serine is surrounded by a GxSxG or AxSxG motif (Long and Cravatt, 2011). Specifically looking at the active site, it is clear that PREPL maintains the characteristic Ser-Asp-His catalytic triad required for the enzymatic activity of the other family members (Figure 2B). To more closely examine the structural similarities between PREPL and other family members, we attempted to crystallize PREPL. Though we were able to grow crystals, refinement proved difficult as we were never able to develop a completely satisfactory model. This resulted in only a fragmented model of the PREPL structure that did not include any of the active site residues. Therefore, we proceeded to use a structure prediction via Protein Homology/ analogY Recognition Engine V 2.0, Phyre^2, (Kelley et al., 2015) to derive insight from a comparison of PREPL's three-dimensional structure with the closest S9 family member homolog PREP. The Phyre model of the structure of PREPL (gray, MLS in blue) shows that in addition to sharing sequence homology with PREP (light blue), the secondary structure of the two proteins is also very similar (Figure 2C). Like PREP, PREPL appears to consist of an α/β hydrolase fold, characteristic of the S9 family (Long and Cravatt, 2011), in addition to a β propeller covering the central tunnel (Figure 2C). A view down the tunnel created by the β propeller further demonstrates structural similarity (Figure 2D). Though β propellers are hypothesized to aid in substrate specificity (Chen et al., 2011; Polgar, 1992), lacking known PREPL substrates leaves us unable to speculate on its role in PREPL's activity.

The S9 serine hydrolase family contains esterases in addition to peptidases

Including PREP, the S9 serine hydrolase family contains a number of peptidases that cleave at prolyl bonds, including a number of dipeptidyl peptidases (DPP) (Figure 2A). While the exact role of PREP, DPP8, and DPP9 remain unclear, DPP4 regulates the levels of glucagon-like peptide-1, having important pharmaceutical implications. However, though PREPL bears significant sequence homology to PREP, to date no peptide substrates have been identified, despite extensive investigation (Boonen et al., 2011; Szeltner et al., 2005). Our data is consistent with this, as PREPL lacked activity against known PREP substrates in our hands. Interestingly, other members of the S9 family included not only peptidases but also esterases/lipases PAFAH2 and PLA2G7 which both catalyze the degradation of platelet-activating factor (Figure 2A). Additionally, one member APEH, an acylaminoacyl-peptide hydrolase has been shown to exhibit both esterase and peptidase activity (Figure 2A) (Bartlam et al., 2004; Scaloni et al., 1994; Wang et al., 2006), a feature which is not uncommon between esterases and acylpeptide hydrolases.

Due to the lack of discernible peptidase activity and the possibility that PREPL may have esterase activity like other S9 family members, we decided to compare the active site of PREPL to other known esterases, lipases, and peptidases (Figure 2E). Immediately evident upon close examination of the active site structures of PREP, oligopeptidase B, and PREPL is that instead of GxSxG motif surrounding the nucleophilic serine as seen for PREP and oligopeptidase B (OpdB), PREPL contains an AxSxG motif (Figure 2E). Interestingly, this motif is also found in LPLA2 and C1d1p, both lipases (Figure 2E). Furthermore, similar to a number of the lipases, including PAFAH2 and PLA2G2, PREPL contains an aspartate (D) near the histidine (H) of the catalytic triad which could act as the acyltransferase aspartate.

PREPL exhibits specific activity against *p*-nitrophenyl ester substrates

In order to test whether PREPL exhibited esterase activity, we utilized an *in vitro* assay composed of various *p*-nitrophenyl esterase substrates (Figure 3A). Upon cleavage of the ester

bond *p*-nitrophenol is released, which absorbs at 405 nm, allowing for observance of development of the reaction. Although esterases and lipases both cleave ester bonds, their substrates differ dramatically. While esterases prefer small molecules or short chain fatty acids, lipases are equipped to cleave long fatty acid tails such as palmitate (C16). Because we were unsure whether PREPL would exhibit esterase activity, and we lacked any evidence as to whether chain length influenced specificity, we tested whether it had activity against substrates of varying chain lengths; two carbons (pNP-C2), four carbons (pNP-C4), eight carbons (pNP-C8), ten carbons (pNP-C10), twelve carbons (pNP-C12), and sixteen carbons (pNP-C16) (Figure 3B, dark gray bars). No activity against chain length longer than C10 was observed, and while there seems to be some activity against pNP-C2, due to autohydrolysis, it was difficult to determine whether the activity was enzyme specific. However, PREPL exhibited low, but reproducible activity against short-medium length chains such as pNP-C4, pNP-C8, and pNP-C10. Importantly, the activity was abolished upon mutation of the catalytic serine (Figure 3B, light gray bars). Based on these results, pNP-C8 was used for all additional assays. Additionally, the esterase activity was dependent on substrate concentration (Figure 3C). Although non-linear, in general there was an association between higher concentrations of pNP-C8 and higher levels of enzymatic activity.

To differentiate whether the esterase was specific to PREPL or due to the promiscuity of serine hydrolases, we purified PREP to test whether it too had esterase activity similar to PREPL. As an additional control, to ensure that we were using active PREP, we also tested for activity against the known PREP substrate Z-Gly-Pro-pNA, a peptide-like substrate that contains an amide instead of ester bond (Figure 3D). The promiscuity of serine hydrolases was apparent when the lipase control exhibited high levels of activity against pNP-C8 (3D, white bar). While PREPL was able to cleave pNP-C8, activity which was abolished upon mutation of the catalytic serine, PREP was not. As expected, PREPL showed no activity against the peptide substrate. Taken together, we have demonstrated that PREPL, despite its sequence homology to the peptidase PREP, is an active serine hydrolase that likely cleaves ester instead of peptide substrates.

PREPL$_L$ localizes to the mitochondria

PREPL has seven isoforms which give rise to two distinct proteins, a long and short isoform (PREPL$_L$ and PREPL$_S$, respectively). Although PREPL$_S$ is more highly and universally expressed, PREPL$_L$ is more specifically expressed in brain and muscle tissue (Parvari et al., 2005) which could explain the marked hypotonia observed in HCS and PREPL deletion patients. Despite this fact, all previous studies of PREPL have focused solely on PREPL$_S$ (Lone et al., 2011; Radhakrishnan et al., 2013; Szeltner et al., 2005).

Though PREPL$_S$ was previously found to be cytosolic (Radhakrishnan et al., 2013), recent studies have found to be localized to the mitochondrial matrix (Rhee et al., 2013). Interestingly, the additional 89 amino acids in PREPL$_L$ include a predicted mitochondrial localization sequence (MLS) (Figure 4A, blue font). We compared the localization of the two isoforms, which were tagged with FLAG at the C-terminus for detection via an anti-FLAG antibody, to that of green fluorescent protein with an N-terminal mitochondrial localization sequence (MLS-GFP) in HEK293 cells using confocal microscopy (Figure 4B). Lack of overlap between PREPL$_S$-FLAG and MLS-GFP indicate that the short isoform is cytosolic (Figure 4B, top). However, PREPL$_L$-FLAG colocalizes with MLS-GFP, indicating that addition of the predicted MLS in PREPL$_L$ localizes the protein to the mitochondria (Figure 4B, bottom). Interestingly, not only does the high density of mitochondria and large amount of ATP required in muscle tissue make it high susceptible to mitochondrial dysfunction, but hypotonia is also common symptom in a wide range of mitochondrial diseases, therefore, we hypothesize that it is the mitochondrial isoform, PREPL$_L$, that is largely responsible for the symptoms seen in HCS and PREPL deletion patients. However, it should be noted that as both isoforms exhibit the same esterase activity *in vitro* (data not shown). Though it is unclear whether the two isoforms have unique functions or provide the same function in different cellular locations (Lewendon et al., 1988), our data suggests that PREPL$_L$ cannot be completely ignored in future studies.

PREPL$^{-/-}$ mouse brain mitochondria have altered lipid profiles compared to WT

Based on our *in vitro* activity assays we hypothesized that PREPL cleaved ester bonds. If true, one would expect loss of PREPL to alter the lipidomic or metabolic profile. To test this we utilized the available PREPL$^{-/-}$ mouse model and performed lipidomics and metabolomics (Figure 5A). It is clear that PREPL cleaves ester bonds, but unclear if it works on lipids or small molecule esters, therefore both lipidomics and metabolomics were performed. We chose to focus on the brain because of PREPL's high expression levels in the organ (Parvari et al., 2005). Furthermore, because hypotonia, one of the major symptoms of HCS and previously attributed to loss of functional PREPL (Jaeken et al., 2006; Regal et al., 2014; Silva et al., 2018), is commonly associated with mitochondrial dysfunction and an individual with a PREPL deletion was recently found to have defective complex IV activity (Legati et al., 2016), we limited our search to mitochondrial substrates by purifying mitochondria prior to lipid or metabolite extraction. In order to limit contamination from other organelles such as the endoplasmic reticulum (ER), mitochondria were purified using sucrose gradients. Three replicates of wild type (WT) and PREPL$^{-/-}$ were subjected to lipidomic and metabolomic analysis.

We identified 2,702 lipids and 191 metabolites across all samples. Due to interference from sucrose (data not shown) the metabolite data was sparse and no significant differences were identified; it is likely the interference masked metabolite differences and future work will endeavor to eliminate the sucrose gradient interference. We identified 28 lipids significantly altered in PREPL$^{-/-}$ mouse brain (q-value ≤ 0.05), suggesting that lack of PREPL influences lipid composition (Figure 5B). However, we were unable to unambiguously identify the lipid species, indicating that they were perhaps previously unknown/uncharacterized lipid species, or low abundance species that did not yield clear fragmentation spectra. Further work will include efforts to identify these lipids unambiguously and investigate them using *in vitro* assays with PREPL to determine substrate specificity.

Discussion

Although PREPL has been implicated in HCS and is thought to be responsible for the hypotonia associated with the disease (Lone et al., 2014; Regal et al., 2014), the mechanism by which it affects muscle tone remains unknown in large part because its function remains elusive. Here, we found that the predicted MLS for the long isoform of PREPL ($PREPL_L$), indeed targeted the protein to the mitochondria while $PREPL_S$ was cytosolic. However, it is unclear whether the two isoforms perform distinct functions or the same function in different locations.

PREPL and its closest homolog, prolylendopeptidase (PREP) are both members of the S9 family of serine hydrolases. Despite the fact that PREP is a known peptidase, to date no peptidase activity has been detected for PREPL (Boonen et al., 2011; Szeltner et al., 2005). Upon closer inspection of the family, we noticed that although a number of other members of the family were peptidases like PREP, there were a number that were esterases or lipases. Additionally, comparison of a structure model of PREPL to PREP indicate that despite lack of peptidase activity, the proteins' three-dimensional structure is very similar. Analysis of the active site identified an AxSxG motif and HxxxD motif that exist in other lipases and esterases but not peptidases. Taken together, this data suggested that PREPL may cleave ester rather than amide bonds.

Using *in vitro* assays, we demonstrated that PREPL exhibits specific esterase activity against pNP-C8 that is dependent on protein activity and substrate concentration. Despite only exhibiting low levels of activity, this is the first catalytic activity definitively attributed to PREPL. Importantly, PREP did not exhibit any activity against the ester substrate.

Finally, we demonstrated that loss of PREPL in mice induces alterations in the brain mitochondrial lipidome. (Though it remains unclear if PREPL works on a small ester molecule or longer fatty acid, metabolomic data was inconclusive). Due to limitations of lipid identification, though they were identified and quantified, none of these lipids could be unambiguously named.

Future work will focus on identifying these lipids. For example, other work from our lab has looked at the effect of loss of PREPL on HAP1 cells, a nearly haploid cell line. Using CRISPR technology, PREPL deletion was confirmed in two separate knockouts. One would expect species altered in both mouse and HAP1 cells to be likely candidates for PREPL substrates. However, it is difficult to compare the two datasets unless they are searched together, which remains to be done. Additionally, it is likely that many of these alterations are indirect targets, thus identification of any potential substrate would require biochemical follow-up.

In summary, this study demonstrates the first documented catalytic activity of PREPL. Unlike its closest homolog, PREP, PREPL is an esterase. The ultimate goal is to identify an *in vivo* substrate of PREPL, though the current work falls short. Identification of *in vivo* substrate will help to understand and possibly treat the mechanism by which PREPL causes hypotonia and growth defects in HCS patients. Though the current work falls short of substrate identification, it serves as a starting point for future enzymatic analyses.

Experimental Procedures

Sequence analysis

Protein homologs were defined by Homologene and reciprocal protein BLAST searches (NCBI, 2015).

Structural model

To create a structural model of the three-dimensional structure of PREPL, we used Phyre 2 (Kelley et al., 2015). PyMOL was used to visualize and compare the PREPL model to the solved crystal structure of PREP (PDB:3DDU).

Recombinant protein expression and purification

Human PREPL$_S$, PREPL$_L$, and PREP. 8His-MBP-[TEV]-protein constructs were expressed and purified as previously described (Reidenbach et al., 2018; Stefely et al., 2015). In short, constructs were overexpressed in *E. coli* by autoinduction (Fox and Blommel, 2009). Cells were isolated and resuspended in Lysis Buffer [20 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol (BME), 0.25 mM phenylmethylsulfonyl fluoride (PMSF)] (4 °C). Cells were lysed by sonication, lysate was clarified by sonication and purified using cobalt IMAC resin (Talon resin). His-tagged protein was eluted from the resin with Elution Buffer [20 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM BME, 0.25 mM PMSF, 100 mM imidazole] and concentrated. Following an exchange into Storage Buffer [20 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM BME] the 8His-MBP tag was cleaved with Δ238TEV protease (1:50, TEV/fusion protein, mass/mass) overnight. The TEV protease reaction mixture was mixed with cobalt IMAC resin (Talon resin), incubated (4 °C, 1 h) and the unbound proteins was collected, concentrated, and exchanged into storage buffer. Point mutations were introduced by PCR-based mutagenesis and confirmed by DNA sequencing and then were purified as described above.

pNP assay to assess protein activity

Esterase activity was assayed by measuring the rate of hydrolysis of various *p*-nitrophenyl substrates at 37°C. Method adapted from (Gupta et al., 2002) for compatibility with small volumes and measurement with a plate reader. A reaction mixture of 30 µL enzyme extract, 200 µL Tris-HCl buffer (50 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 8.0), and 0.75 mM pNP dissolved in Triton X-100 (1.25% of final volume) was prepared. The progress of p-nitrophenol release was measured by its absorption at 405 nm. Activity was calculated by determining the amount of p-nitrophenol released over time per amount of enzyme present (nmol/min/mg).

Subcellular localization via confocal microscopy

On the day prior to transfection, HEK293 cells were plated at a densityof 75,000 cells/well onto poly-D-lysine-coated coverslips in 6-well dishes. Cells were transiently transfected with a mix of 1 µg pcDNA3.1 gene-FLAG, 0.5 µg plasmid encoding green fluorescent protein with an N-terminal mitochondrial localization sequence (MLS-GFP), 7.5 µg PEI, and 200 µL Opti-MEM. After 24 hours, the cells were fixed (4% paraformaldehyde in PBS), permeabilized (0.2% Triton X-100 in PBS), blocked (1% BSA in PBS), and probed with mouse anti-FLAG M2 1° antibody (F1804, Sigma, 1:2000 (v/v) in 1% BSA in PBS) and Alexa Fluor 594-conjugated goat antimouse 2° antibody (LifeTechnologies, 1:2000 (v/v) in 1% BSA in PBS) in 1% BSA in PBS. Hoechst dye (1 µg/mL) was used to label nuclear DNA. Slides were placed in mounting medium (1:1, v/v, glycerol/PBS). Confocal images were captured using the Nikon A1R system, Plan Apo VC 60X oil immersion optics, with sequential laser excitation using 561 nm (Alexa Fluor), 488 nm (GFP), and 408 nm (Hoechst) lasers. Images were collected and assembled into a Z-stack using the NISElements software.

PREPL$^{-/-}$ mouse model

Breeding, colony maintenance, sacrificing, and tissue harvesting of mice were previously described (Lone et al., 2014). In short, PREPL$^{-/-}$ mice were generated by creation of a construct in which exon 11, the exon containing the catalytic serine, was located between two Cre recombinase recognition sites. The construct was introduced into embryonic stem (ES) cells. Neomycin selection and Southern blots were used to ensure that selected ES cells incorporated the correct construct in the correct location. These ES cells were injected into blastocytes and implanted into mice. PREPL$^{-/+}$ mice were generated by crossing these mice with mice expressing Cre recombinase, which removed exon 11 from *Prepl*. PREPL$^{-/+}$ were bred, producing a mixture of PREPL$^{-/+}$, PREPL$^{+/+}$, PREPL$^{-/-}$ mice, the last two groups of which were used for this study.

Purification of mitochondria from mouse brains

Method adapted from (Chinopoulos et al., 2011). Tissues were homogenized in 1 mL isolation medium (225 mM sucrose, 75 mM mannitol, 5 mM HEPES, 1 mM EGTA, pH 7.4) + 0.2 mg/mL BSA (fatty acid free) at 4°C using a Potter-Elvehjem homogenizer. An additional 700 mL of isolation medium + BSA was added and homogenate mixed by inversion. Homogenate was initially centrifuged (500 x g, 5 min, 4°C) to remove cell debris and nuclei. Supernatant was subjected to a faster spin (14000 x g, 10 min, 4°C) to isolate crude mitochondria. Pellets were resuspended in 0.2 mL of 12% Percoll in isolation medium, layered on top of 1 mL of 24% Percoll in isolation medium and centrifuged (18000 x g, 15 min, 4°C). The top portion of the sample was aspirated off (~700 mL) followed by addition of 1.2 mL isolation medium and centrifugation (18000 x g, 5 min, 4°C). After aspirating off ~1.5 mL of supernatant, the pellet was resuspended in the remaining buffer, an additional 1.5 mL of isolation medium was added at the centrifugation step was repeated. The remaining pellet consisted of the purified mitochondria and was used for lipidomics and metabolomics.

Lipidomics sample preparation and analysis

Lipid Extractions

Lipids were extracted from samples containing 70 µg of protein as determined by BCA as described in our previous studies defining the lipidome of the BXD cohort (Jha et al., 2018a; Jha et al., 2018b). In short, Lipid extraction was performed as previously described (Stefely et al., 2016). In brief, 20 µl of IS was added to 20 µl of thawed plasma samples and vortexed (30 s). Chloroform/methanol (1:1, v/v, 1000 µL) was added and samples vortexed (60 s). Subsequently, hydrochloric acid (1M, 200 µL) was added to induce phase separation, followed by 60 s vortex and centrifugation (3,000 g, 3 min, 4°C) to complete phase separation. 550 µL of the organic phase was dried under Ar_2(g). The organic residue was reconstituted in a mixture of acetonitrile (ACN)/isopropyl alcohol (IPA)/water (H_2O) (65:30:5, v/v/v, 100µL) by vortexing (60 s) and transferred to a glass vial for LC-MS analysis. Samples were stored at -80°C until further use.

Discovery Lipidomics

LC-MS analysis was performed on an Ascentis Express C18 column (150 mm x 2.1 mm x 2.7 µm particle size, Waters) using an Accela LC Pump (400 µL/min flow rate, Thermo Scientific, San Jose, CA). Mobile Phase A consisted of 10 mM ammonium acetate in ACN/H_2O (70:30, v/v) containing 250 µL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. Mobile phase B started at 2% and increased to 85% over 20 min, then increased to 99% over 1 min and held there for 7 min. The column was reequilibrated for 2 min before the next injection. 10 µL of sample was injected for each run. The LC system was coupled to a Q Exactive mass spectrometer by a HESI II heated ESI source kept at 300°C. The inlet capillary was kept at 300°C, sheath gas was set to 25 units, and auxiliary gas to 10 units. Spray voltage was set to 3,000 V and the MS was operated in polarity switching mode. Ions from 200-1,600 m/z were isolated (Top 2) for fragmentation by stepped higher-energy collisional dissociation (HCD) (20, 30, 40).

Lipid Species Measurement and Normalization

The resulting spectra were processed using Lipidex, an open source, automated lipid identification and quantification software in which acquired MS^2s are compared to a variety of *in silico* libraries accurately modeling fragmentation of diverse lipid types (Hutchins et al., 2018). Data was normalized to total abundance of lipids identified. The $\log_2$ fold change between $PREPL^{+/+}$ and $PREPL^{-/-}$ was compared for each lipid. Student's t-test was used to calculate p-values. q-values were calculated in R (Storey et al., 2018) and were used to determine statistical significance.

Metabolomics sample preparation and analysis

Metabolite extraction and derivatization

Metabolites were extracted from samples containing 70 µg of protein as determined by BCA as described in (Stefely et al., 2016). Briefly, samples were immediately submerged into ACN/MeOH/H_2O (2:2:1, v/v/v, 1.5 mL, pre-cooled to −20 °C) and were stored at −80 °C prior to analysis. Samples and internal standards (25 µL aqueous mixture of isotopically labelled alanine-2,3,3,3-d_4, adipic acid-d_{10}, and xylose-$^{13}C_5$ acid, 5 ppm in each) were aliquoted into a 2 mL plastic tube and dried by vacuum centrifuge (~ 1 hr). The dried metabolites were resuspended in pyridine (25 µL) and vortexed. 25 µL of N-methyl-N-trimethylsilyl]trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) was added, and the sample was vortexed and incubated (60 °C, 30 min).

Metablolite measurement and data analysis

Samples were then transferred to a glass autosampler vials and analyzed using a GC/MS instrument comprising a Trace 1310 GC coupled to a Q Exactive Orbitrap mass spectrometer.

The resulting GC-MS data were processed using an in-house software suite developed by the Coon Lab. Briefly, all m/z peaks are aggregated into distinct chromatographic profiles (i.e., feature) using a 10 ppm mass tolerance. These chromatographic profiles are then grouped according to common elution apex (i.e., feature group). The collection of features (i.e., m/z peaks)

sharing a common elution apex, therefore, represent an individual EI-MS spectrum of a single eluting compound. The EI-MS spectra were then compared against a matrix run and a background subtraction was performed. Remaining EI-MS spectra are then searched against the NIST 12 MS/EI library and subsequently subjected to a high resolution filtering (HRF) technique as described elsewhere. EI-MS spectra that were not identified were assigned a numeric identifier. Feature intensity, which was normalized using total metabolite signal, was used to estimate metabolite abundance. The $\log_2$ fold change between PREPL$^{+/+}$ and PREPL$^{-/-}$ was compared for each metabolite. Student's t-test was used to calculate p-values.

Author contributions

Conceptualization, BJF, DJP, and MTM; Sequence Alignments, BJF and MTM; Cloning and Protein Purification RLW, TMM, AJ, MTM; Activity assays, MTM, Cell Culture and Microscopy, BJF, BKD; Lipidomics Extractions and Analysis, MTM; Metabolite Extractions and Analysis, MTM, BP, TR; Writing, MTM; Visualization, MTM; Project Administration and Funding Acquisition, DJP and JJC.

Acknowledgements

We thank the Center for Eukaryotic Structural Genomics for providing plasmids for PREPL and first purifying the enzyme, Alan Saghatelian for PREPL$^{-/-}$ samples, and Amy Lin for her help in figure construction

Figures and tables

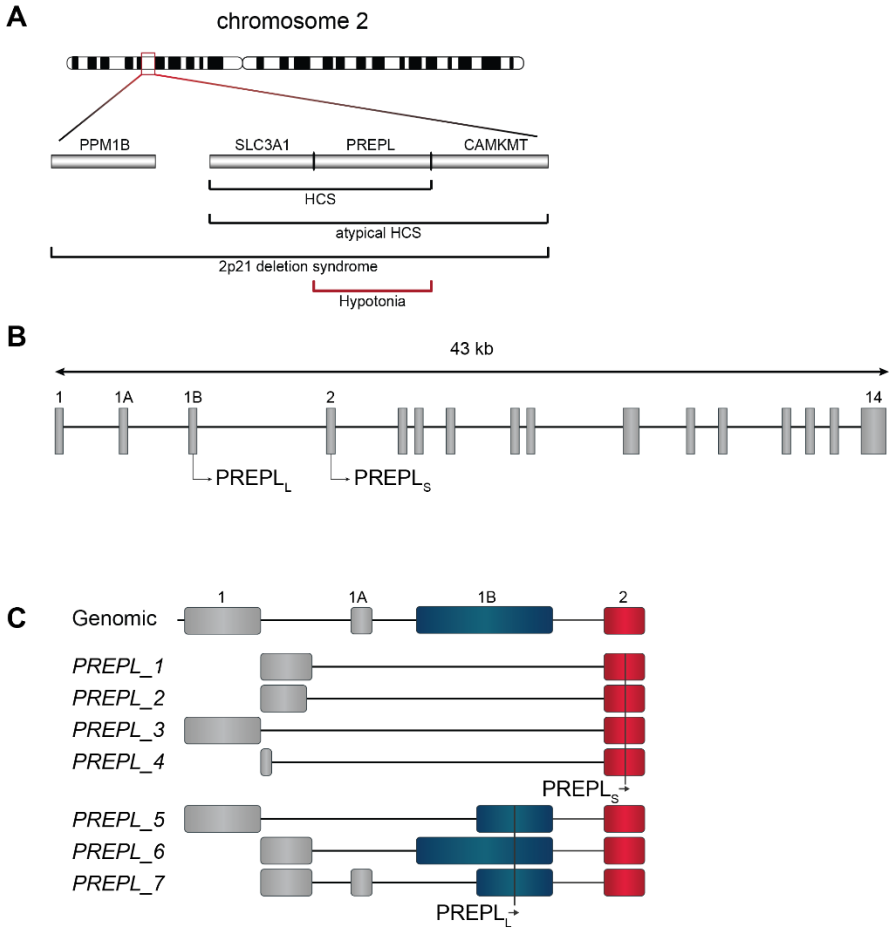


Figure 1. PREPL is implicated in hypotonia-cystinuria syndrome (HCS) and consists of a long and short isoform

(A) Schematic of gene deletions on the 2p21 chromosome resulting in individuals with HCS and related diseases. Lines indicate genes with deletions in each of the listed diseases. Isolated deletions in PREPL are associated with hypotonia (red).

(B) PREPL consists of 14 exons and has two protein isoforms; short ($PREPL_S$) and long ($PREPL_L$) created by separate start sites in exon 1B ($PREPL_L$) and exon 2 ($PREPL_S$). These proteins are identical except for the additional amino acids on the N-terminus of $PREPL_L$.

(C) The seven mRNA isoforms of PREPL that give rise to the two distinct protein isoforms shown in (B). Isoforms 1-4 produce a 638 aa product ($PREPL_S$) while 5-7 generate a 727 aa product ($PREPL_L$).

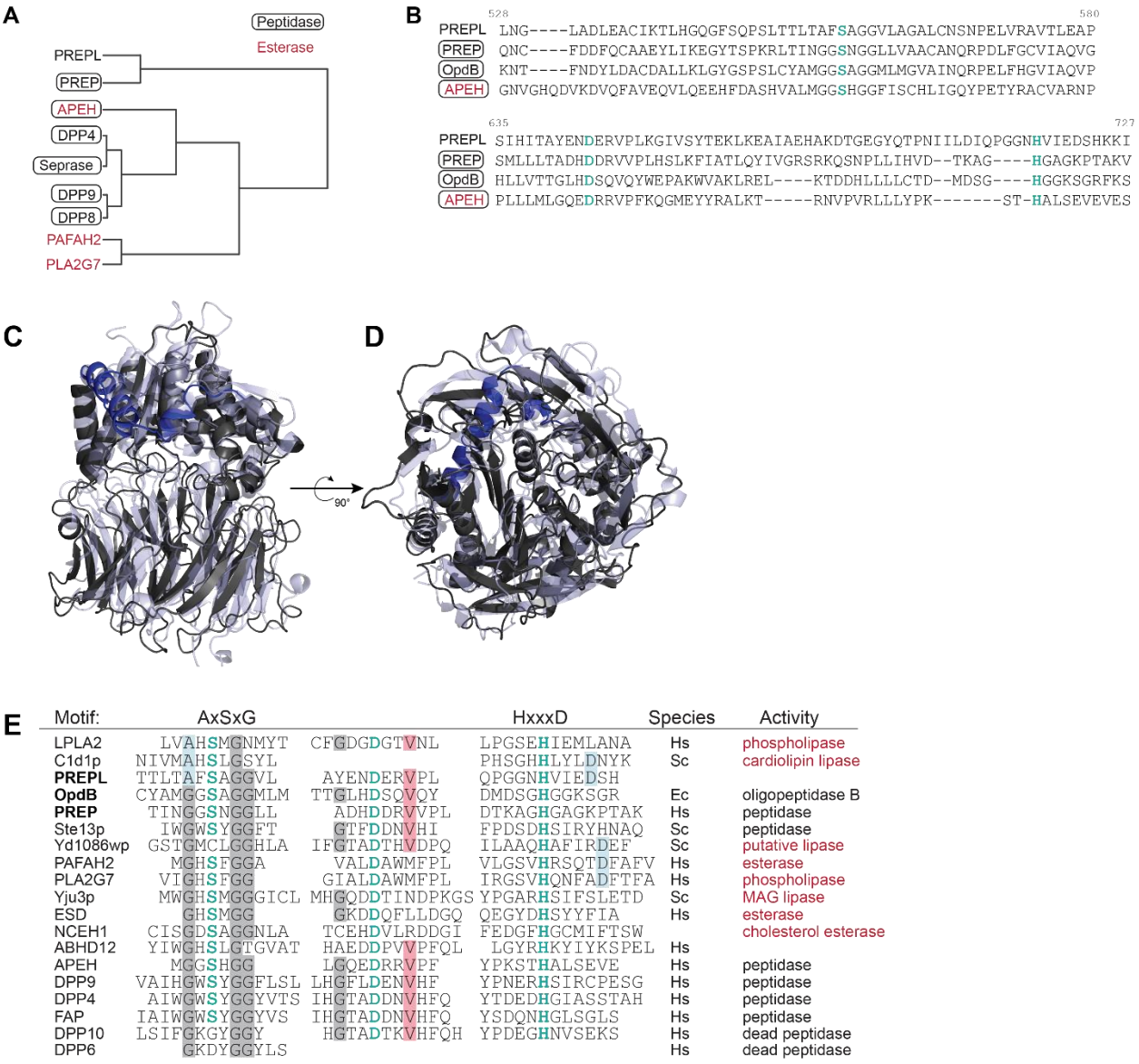


Figure 2. PREPL maintains the catalytic triad and exhibits sequence similarity to peptidases as well as esterases and lipases

(A) PREPL and its closest homolog PREP are members of the S9 family of serine hydrolases. The S9 family consists of peptidases (circled in black) and esterases (red).

(B) Sequence alignment of PREPL, other members of the S9 family, PREP and APEH, and a bacterial homolog oligopeptidase B (OpdB) indicate that PREPL maintains the characteristic Ser-Asp-His catalytic triad (teal) required for enzymatic activity.

(C and D) Overlay of structure model of PREPL (gray, MLS in blue) and PREP (light blue) illustrates the similarity in secondary structure of the two proteins. Both proteins are made up of an α/β hydrolase fold (C, top) and a β propeller (C, bottom) covering the tunnel into the active site. A view down the tunnel created by the β propeller (D) further illustrates the similarity in secondary structure.

(E) Active site comparison of PREPL and other peptidases and lipases. Active site residues are shown in teal. Conserved residues are highlighted in gray (glycine), light blue (alanine), and red (valine). Unlike other peptidases in the S9 family (PREPL, DPP4, DPP8, and DPP9) PREPL contains an AxSxG around the catalytic serine, similar to that seen in the lipases LPLA2 and C1d1p, and an aspartate near the active site histidine (HxxxD motif, highlighted in light blue) similar to PAFAH2 and PLA2G7.

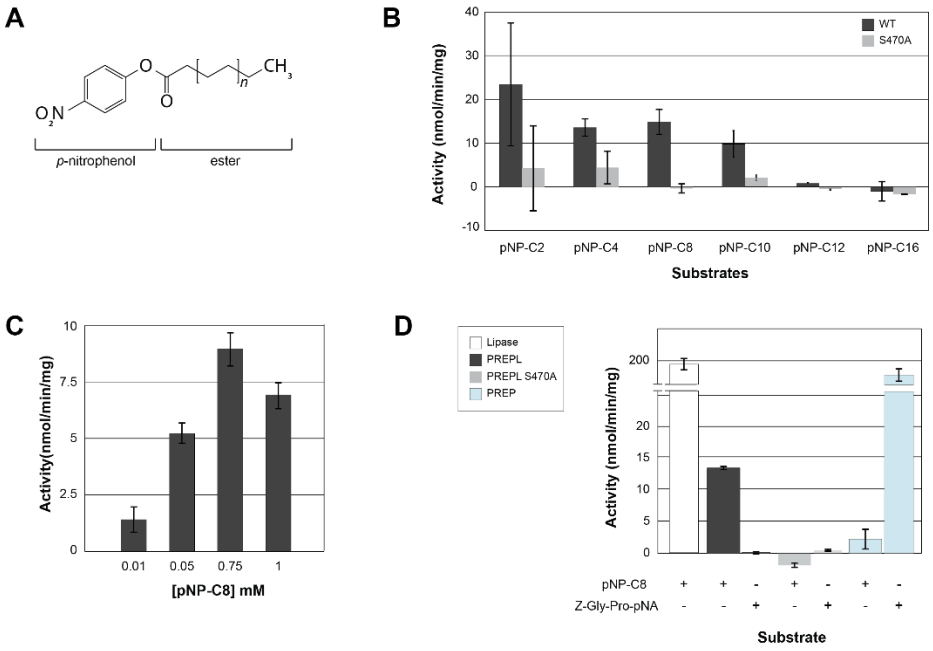


Figure 3. PREPL exhibits specific esterase activity against pNP-C8

(A) Structure of p-nitrophenyl ester substrates. A carbon tail of varying lengths is connected to a *p*-nitrophenol group via an ester linkage. Activity is determined by the amount of *p*-nitrophenol that is released after ester cleavage.

(B) PREPL activity against pNP-ester substrates of various length, two carbons (C2), four carbons (C4), eight carbons (C8), ten carbons (C10), twelve carbons (C12), and sixteen carbons (C16). Wild type (WT) and catalytically dead mutant (S470A) are shown in dark and light gray, respectively.

(C) PREPL activity against increasing concentrations of pNP-C8. Activities against concentrations ranging from 0.01 mM – 1 mM are shown.

(D) Comparison of PREPL (dark and light gray bars for WT and S470A, respectively) and PREP (light blue) activity against the ester substrate pNP-C8 and the peptide substrate Z-Gly-Pro-pNA.

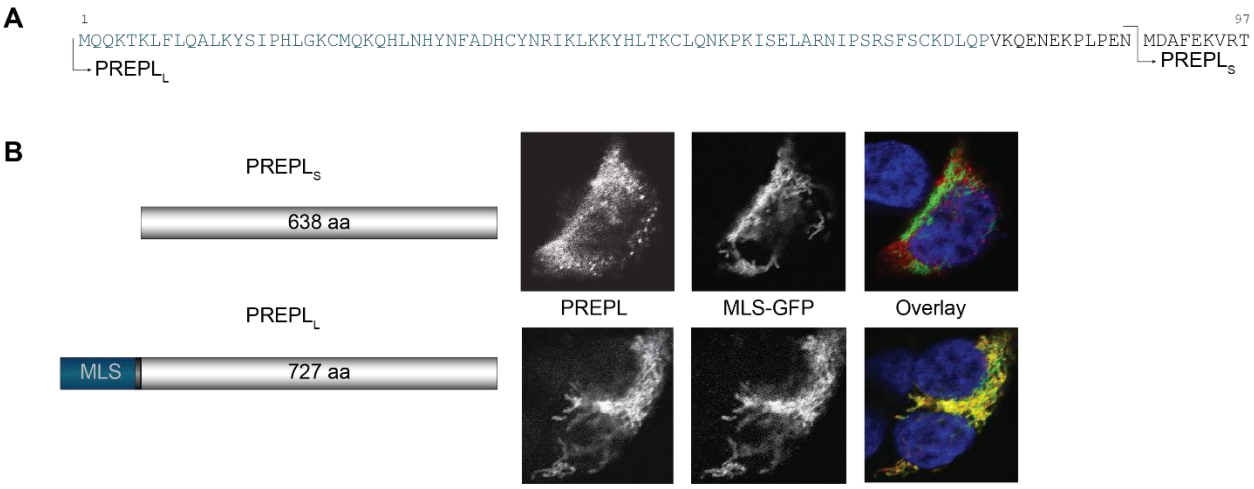


Figure 4. Predicted mitochondrial localization sequence targets PREPL$_L$ to the mitochondria

(A) N-terminus of PREPL, showing the start sites of both isoforms (PREPL$_S$ and PREPL$_L$). The predicted mitochondrial localization sequence (MLS) is indicated in teal.

(B) Localization of PREPL$_S$-FLAG (top, left) and PREPL$_L$-FLAG (bottom, left) and green fluorescent protein with an N-terminal mitochondrial localization sequence (MLS-GFP, middle). Overlay of the FLAG tagged proteins and MLS-GFP (right). PREPL-FLAG isoforms are shown in red, MLS-GFP in green and overlap is indicated in yellow.

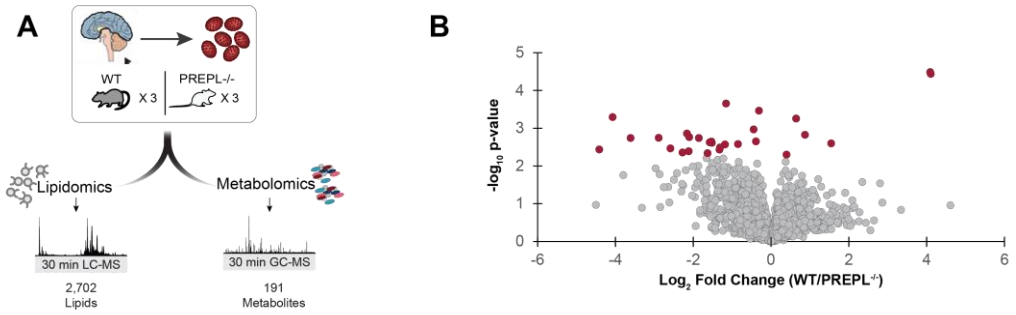


Figure 5. Loss of PREPL alters the mitochondrial lipidome in mice.

(A) Experimental setup. Briefly, mitochondria were isolated from wild type (WT) and PREPL$^{-/-}$ mice brains in triplicate. Lipids and polar metabolites were extracted from each sample and analyzed using LC-MS (lipidomics) or GC-MS (metabolomics)

(B) Volcano plot comparing the abundance of the 2,702 lipids identified in WT and PREPL$^{-/-}$ mice. Red dots indicate significance ($q < 0.05$).

References

Bartholdi, D., Asadollahi, R., Oneda, B., Schmitt-Mechelke, T., Tonella, P., Baumer, A., and Rauch, A. (2013). Further delineation of genotype-phenotype correlation in homozygous 2p21 deletion syndromes: first description of patients without cystinuria. American journal of medical genetics. Part A *161a*, 1853-1859.

Bartlam, M., Wang, G., Yang, H., Gao, R., Zhao, X., Xie, G., Cao, S., Feng, Y., and Rao, Z. (2004). Crystal structure of an acylpeptide hydrolase/esterase from Aeropyrum pernix K1. Structure (London, England : 1993) *12*, 1481-1488.

Boonen, K., Regal, L., Jaeken, J., and Creemers, J.W. (2011). PREPL, a prolyl endopeptidase-like enzyme by name only?--Lessons from patients. CNS & neurological disorders drug targets *10*, 355-360.

Calonge, M.J., Gasparini, P., Chillaron, J., Chillon, M., Gallucci, M., Rousaud, F., Zelante, L., Testar, X., Dallapiccola, B., Di Silverio, F., et al. (1994). Cystinuria caused by mutations in rBAT, a gene involved in the transport of cystine. Nature genetics *6*, 420-425.

Chabrol, B., Martens, K., Meulemans, S., Cano, A., Jaeken, J., Matthijs, G., and Creemers, J.W. (2008). Deletion of C2orf34, PREPL and SLC3A1 causes atypical hypotonia-cystinuria syndrome. Journal of medical genetics *45*, 314-318.

Chen, C.K., Chan, N.L., and Wang, A.H. (2011). The many blades of the beta-propeller proteins: conserved but versatile. Trends in biochemical sciences *36*, 553-561.

Chinopoulos, C., Zhang, S.F., Thomas, B., Ten, V., and Starkov, A.A. (2011). Isolation and functional assessment of mitochondria from small amounts of mouse brain tissue. Methods in molecular biology (Clifton, N.J.) *793*, 311-324.

Font-Llitjos, M., Jimenez-Vidal, M., Bisceglia, L., Di Perna, M., de Sanctis, L., Rousaud, F., Zelante, L., Palacin, M., and Nunes, V. (2005). New insights into cystinuria: 40 new mutations, genotype-phenotype correlation, and digenic inheritance causing partial phenotype. Journal of medical genetics *42*, 58-68.

Fox, B.G., and Blommel, P.G. (2009). Autoinduction of protein expression. Current protocols in protein science *Chapter 5*, Unit 5.23.

Goodyer, P. (2004). The molecular basis of cystinuria. Nephron. Experimental nephrology *98*, e45-49.

Gupta, N., Rathi, P., and Gupta, R. (2002). Simplified para-nitrophenyl palmitate assay for lipases and esterases. Analytical biochemistry *311*, 98-99.

Hutchins, P.D., Russell, J.D., and Coon, J.J. (2018). LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. Cell systems *6*, 621-625.e625.

Jaeken, J., Martens, K., Francois, I., Eyskens, F., Lecointre, C., Derua, R., Meulemans, S., Slootstra, J.W., Waelkens, E., de Zegher, F., et al. (2006). Deletion of PREPL, a gene encoding

a putative serine oligopeptidase, in patients with hypotonia-cystinuria syndrome. American journal of human genetics *78*, 38-51.

Jha, P., McDevitt, M.T., Gupta, R., Quiros, P.M., Williams, E.G., Gariani, K., Sleiman, M.B., Diserens, L., Jochem, A., Ulbrich, A., et al. (2018a). Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species. Cell systems *6*, 722-733.e726.

Jha, P., McDevitt, M.T., Halilbasic, E., Williams, E.G., Quiros, P.M., Gariani, K., Sleiman, M.B., Gupta, R., Ulbrich, A., Jochem, A., et al. (2018b). Genetic Regulation of Plasma Lipid Species and Their Association with Metabolic Phenotypes. Cell systems *6*, 709-721.e706.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols *10*, 845-858.

Legati, A., Reyes, A., Nasca, A., Invernizzi, F., Lamantea, E., Tiranti, V., Garavaglia, B., Lamperti, C., Ardissone, A., Moroni, I., et al. (2016). New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies. Biochimica et biophysica acta *1857*, 1326-1335.

Lewendon, A., Murray, I.A., Kleanthous, C., Cullis, P.M., and Shaw, W.V. (1988). Substitutions in the active site of chloramphenicol acetyltransferase: role of a conserved aspartate. Biochemistry *27*, 7385-7390.

Liu, Y., Patricelli, M.P., and Cravatt, B.F. (1999). Activity-based protein profiling: the serine hydrolases. Proceedings of the National Academy of Sciences of the United States of America *96*, 14694-14699.

Lone, A.M., Bachovchin, D.A., Westwood, D.B., Speers, A.E., Spicer, T.P., Fernandez-Vega, V., Chase, P., Hodder, P.S., Rosen, H., Cravatt, B.F., et al. (2011). A substrate-free activity-based protein profiling screen for the discovery of selective PREPL inhibitors. Journal of the American Chemical Society *133*, 11665-11674.

Lone, A.M., Leidl, M., McFedries, A.K., Horner, J.W., Creemers, J., and Saghatelian, A. (2014). Deletion of PREPl causes growth impairment and hypotonia in mice. PloS one *9*, e89160.

Long, J.Z., and Cravatt, B.F. (2011). The metabolic serine hydrolases and their functions in mammalian physiology and disease. Chemical reviews *111*, 6022-6063.

Martens, K., Derua, R., Meulemans, S., Waelkens, E., Jaeken, J., Matthijs, G., and Creemers, J.W. (2006). PREPL: a putative novel oligopeptidase propelled into the limelight. Biological chemistry *387*, 879-883.

NCBI (2015). Database resources of the National Center for Biotechnology Information. Nucleic acids research *43*, D6-17.

Parvari, R., Gonen, Y., Alshafee, I., Buriakovsky, S., Regev, K., and Hershkovitz, E. (2005). The 2p21 deletion syndrome: characterization of the transcription content. Genomics *86*, 195-211.

Polgar, L. (1992). Unusual secondary specificity of prolyl oligopeptidase and the different reactivities of its two forms toward charged substrates. Biochemistry *31*, 7729-7735.

Radhakrishnan, K., Baltes, J., Creemers, J.W., and Schu, P. (2013). Trans-Golgi network morphology and sorting is regulated by prolyl-oligopeptidase-like protein PREPL and the AP-1 complex subunit mu1A. Journal of cell science *126*, 1155-1163.

Regal, L., Shen, X.M., Selcen, D., Verhille, C., Meulemans, S., Creemers, J.W., and Engel, A.G. (2014). PREPL deficiency with or without cystinuria causes a novel myasthenic syndrome. Neurology *82*, 1254-1260.

Reidenbach, A.G., Kemmerer, Z.A., Aydin, D., Jochem, A., McDevitt, M.T., Hutchins, P.D., Stark, J.L., Stefely, J.A., Reddy, T., Hebert, A.S., et al. (2018). Conserved Lipid and Small-Molecule Modulation of COQ8 Reveals Regulation of the Ancient Kinase-like UbiB Family. Cell chemical biology *25*, 154-165.e111.

Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science (New York, N.Y.) *339*, 1328-1331.

Rice, S.Q., Southan, C., Boyd, H.F., Terrett, J.A., MacPhee, C.H., Moores, K., Gloger, I.S., and Tew, D.G. (1998). Expression, purification and characterization of a human serine-dependent phospholipase A2 with high specificity for oxidized phospholipids and platelet activating factor. The Biochemical journal *330 (Pt 3)*, 1309-1315.

Scaloni, A., Barra, D., Jones, W.M., and Manning, J.M. (1994). Human acylpeptide hydrolase. Studies on its thiol groups and mechanism of action. The Journal of biological chemistry *269*, 15076-15084.

Silva, S., Miyake, N., Tapia, C., and Matsumoto, N. (2018). The second point mutation in PREPL: a case report and literature review. Journal of human genetics *63*, 677-681.

Stafforini, D.M., Prescott, S.M., and McIntyre, T.M. (1987). Human plasma platelet-activating factor acetylhydrolase. Purification and properties. The Journal of biological chemistry *262*, 4223-4230.

Stefely, J.A., Kwiecien, N.W., Freiberger, E.C., Richards, A.L., Jochem, A., Rush, M.J.P., Ulbrich, A., Robinson, K.P., Hutchins, P.D., Veling, M.T., et al. (2016). Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. Nature biotechnology *34*, 1191-1197.

Stefely, J.A., Reidenbach, A.G., Ulbrich, A., Oruganty, K., Floyd, B.J., Jochem, A., Saunders, J.M., Johnson, I.E., Minogue, C.E., Wrobel, R.L., et al. (2015). Mitochondrial ADCK3 employs an atypical protein kinase-like fold to enable coenzyme Q biosynthesis. Molecular cell *57*, 83-94.

Storey, J.D., Bass, A.J., Dabney, A., and Robinson, D. (2018). qvalue: Q-value estimation for false discovery rate control. R package version 2.14.0.

Szeltner, Z., Alshafee, I., Juhasz, T., Parvari, R., and Polgar, L. (2005). The PREPL A protein, a new member of the prolyl oligopeptidase family, lacking catalytic activity. Cellular and molecular life sciences : CMLS *62*, 2376-2381.

Wang, Q., Yang, G., Liu, Y., and Feng, Y. (2006). Discrimination of esterase and peptidase activities of acylaminoacyl peptidase from hyperthermophilic Aeropyrum pernix K1 by a single mutation. The Journal of biological chemistry *281*, 18618-18625.

Chapter 5: Conclusions and future directions

Conclusions

Summary

Lipids are a diverse set of macromolecules important for many cellular processes such as energy storage, signaling, and membrane structure (Hannun, 1994; Kooijman et al., 2003; Kooijman et al., 2005; Lee et al., 2001). Dysfunctional lipid metabolism is associated with common diseases such as obesity, insulin resistance, and fatty liver disease (Granger and Remick, 2005; Li et al., 2017; Weijers, 2012) as well as less common diseases such as Barth Syndrome (Lou et al., 2018; Vreken et al., 2000). Despite extensive studies and considerable progress, our understanding of the intricacies of lipid metabolism is far from complete. This is, in large part, due to the complexity of lipid metabolism as a whole. The biosynthetic pathways for different lipid classes overlap at multiple points, with some lipids such as PA and DAG serving as substrates in other lipid biosynthetic pathways. Therefore, defects in one pathway will often deleteriously affect other lipid metabolism pathways in the cell. Additionally, there remains the issue of the vast number of combinations of head groups and fatty acid tails (including number of carbons, number of double bonds, and positions of these double bonds). Beyond regulation of class abundance, it has become increasingly clear that not all lipids in a class are regulated in the same manner. Yet, the mechanisms by which the cell regulates individual lipid species within a class remain unknown. Although the issue is a complex one, it can largely be boiled down to a single issue – a lack of knowledge of all proteins that can affect lipid metabolism. In the projects outlined here, we sought to better understand the proteins involved in lipid metabolism via two distinct methods.

In the first, we sought to identify novel genes involved in lipid metabolism using QTL mapping. Though abundances of total classes such as TAG and cholesterol has been mapped, before my graduate research began, individual lipid species had not been used for QTL mapping studies in genetic reference populations. Collectively, we established that individual lipid levels,

like protein and transcript levels, can serve as quantifiable, mappable traits and identified novel candidate genes involved in lipid metabolism. Our data also supports the fact that not all lipids of the same class are regulated in the same way, as is evidenced by the discovery that some TAG species may serve as signatures of a fatty liver, while others are associated with a healthy one.

In the second, we took a more focused approach, looking at whether a single protein of unknown function, PREPL, affected lipid metabolism – specifically if it showed enzymatic activity against esters. We demonstrated that that additional amino acid sequence on the long isoform of PREPL ($PREPL_L$) (when compared to the short isoform) targeted it to the mitochondria. Prior to our work, PREPL was thought to be a peptidase, despite no peptide substrate ever having been identified. We have established that PREPL exhibits esterase activity dependent on catalytic activity and our work provides a new focus for further characterization of the enzyme.

The first QTL mapping of individual hepatic lipids and power of multi-omics approaches

Recent advances in mass spectrometry have made it possible to measure many lipids from a single extraction accurately across hundreds of samples. Previous studies focused on measuring lipid classes as a whole yet recent evidence strongly suggests that not only are all lipids within a class not regulated concurrently, but individual lipid species can be associated with disease (Gronert et al., 2004; Koybasi et al., 2004; Kroesen et al., 2001). Based on these advancements, we developed a method targeting hepatic lipid species across 84 cohorts of BXD mice fed either a CD or HFD. We were able to quantify 96 lipids from a variety of classes (FFA, TAG, DAG, PL, CL and CoQ) across 385 mice.

Interestingly, a weighted correlation network analysis (WGCNA) which clusters correlated groups of lipids into modules, showed multiple modules containing TAG and CL species indicating that species within these classes may be differentially regulated. Surprisingly, although two of the TAG modules both correlated with obesity and liver dysfunction traits, they correlated inversely to each other. Furthermore, we found that the fully remodeled CL, CL(LLLL) and MLCL(LLL)

correlated negatively with liver weight and obesity and NAFLD readouts while CLs containing MUFAs such as oleic and palmitoleic acid correlated positively in both the BXD mice and an additional mouse model. Taken together, this data suggests these CLs may be signatures of healthy or fatty liver, respectively (Jha et al., 2018a). These findings demonstrate that not all lipid species within a class are regulated in the same manner.

Using the lipid abundance measurements, we mapped suggestive/significant 136 lQTLs. Unsurprisingly, loci contained a vast number of genes. In order to prioritize candidate genes, we developed a novel QTL mining pipeline integrating multi-omic datasets to identify genes most likely to be regulating lipid species. The pipeline prioritized genes that (i) contained non-synonymous SNPS, (ii) had a *cis* eQTL or pQTL, (iii) exhibited significant correlation with the lipid in question, and (iv) displayed variation in transcript expression, with those genes exhibiting three or more of these factors given higher priority. Importantly, the prioritized candidate genes were enriched in lipid metabolic pathways (Jha et al., 2018a). These findings demonstrate the power of a multi-omics approach in lQTL mapping and provide a key resource of candidate genes for future studies.

Collectively, our work demonstrates that lipids, like transcripts and proteins, can be used as traits for QTL mapping. Additionally, it illustrates the growing importance of studying individual lipid species, rather than just lipid classes as a whole. Finally, this analysis provides a framework for using multi-omics to prioritize candidate lQTL genes and a resource of candidate genes that may directly regulate the lipid species measured.

Plasma lipids are influenced by both genetic and environmental factors and can serve as a reflection of overall metabolic health

We profiled 129 plasma lipids across 49 BXD strains and two diets (CD and HFD). Unsurprisingly, diet (environment) had a strong influence on the lipid profile, affecting 72% of lipids identified. We next set out to determine the genetic component of alterations in lipid

abundance across the profiled mouse strains by looking at the variation across strains and heritability (h^2) of each lipid species. The variability within strains and diets was significantly lower than that across strains, indicating that at least some of the variance observed could be attributed to genetic variation (Jha et al., 2018b). Importantly, a calculation of heritability revealed that within diets more than half of the observed variance in the majority of lipids could be explained by the genetic differences between the BXD strains ($h^2 \geq 50\%$). Furthermore, for 85% of lipids controlled factors (genetics, environment, and their interactions [GxE]) explained more than 50% of variance observed (Jha et al., 2018b). Taken together, these results show that despite the strong influence of diet, genetic factors also have a significant impact on the abundance of individual lipid species, indicating that both environmental and genetic factors can regulate lipid metabolism.

Interestingly, most lipids mapped to more than once loci, which may not be surprisingly given the high number of individual pathways that regulate lipid levels, and the fact that plasma lipid levels are likely influence by lipid levels in a variety of tissues. We next sought to identify whether our mouse data was translatable to the human population by screening the 306 lQTLs identified for the presence of any human GWAS genes associated with plasma lipid levels. We identified 40 GWAS genes, many of which are known to affect total lipid, or total class levels. Thus, it seems, perhaps unsurprisingly, that genes associated with total lipid levels can also regulate individual lipid species. Additionally, these data suggest that BXD and other mouse lQTL studies can be used to support and further study results suggested from human GWAS.

Because lipid abundance was highly correlated both within and across different lipid classes, it is likely that an alteration in the abundance of a single lipid species could impact the abundance of various other lipids independent of the class. We, therefore, sought to determine whether the abundance of certain lipids could serve as predictors of metabolic health. Indeed, across 30 traits that reflect metabolic health (both healthy and unhealthy traits), we identified two clusters of lipids that correlated with most metabolic traits, independent of diet. The first cluster, consisting of 19 lipids, positively correlated with healthy metabolic traits such as run distance and

VO_2 max and negatively correlated with unhealthy traits such as fasting insulin and body weight. The second cluster, consisting of 17 lipids, displayed the inverse correlation pattern. These data suggest that abundance of some individual plasma lipid species can reflect metabolic health. Because these effects were seen in both CD and HFD cohorts it is possible that these 36 lipids may serve as universal signatures of health and may also be relevant in humans.

Knowing that lipid abundance could reflect overall metabolic health, we next sought to determine if they could also reflect a disease state. Specifically, we tested whether plasma lipid species can be indicative of non-alcoholic fatty liver disease (NAFLD), by first identifying plasma lipid species that were representative of their levels measured in liver (Jha et al., 2018a). In order to increase the chances that the results would be relevant in a human population where environmental factors such as diet are difficult to control, narrowed the list to the 9 species whose plasma and liver levels correlated irrespective of diet. Correlation of the abundance of these species with NAFLD-linked phenotypes such as fasting insulin, fat mass, and liver mass revealed that TAG species with fewer double bonds correlated positively with pro-NAFLD traits, while TAGs with more double bonds correlated negatively with pro-NAFLD traits (Jha et al., 2018b). We confirmed these findings in different mouse model of NAFLD as well as human patients with healthy livers or livers in various stages of steatosis. Together, these data suggest that levels of individual plasma lipid species may provide insight into liver health in mice and more importantly, the human population.

Together with the liver manuscript, this work demonstrates the importance of the role of individual lipid species in health and disease. We illustrate that both diet and genetic factors regulate the abundance of individual plasma lipid species and that many lipid species are likely co-regulated, despite being parts of different classes. Finally, we show that not only can plasma lipid levels reflect overall metabolic health, but they may also serve as pro- or anti-NAFLD signatures in mice. Importantly, we demonstrate that NAFLD signatures in mice may be relevant

in humans as well, indicating that it may be possible to use individual lipid species as biomarkers of disease in the future.

PREPL exhibits esterase activity

Though PREPL has been implicated in hypotonia cystinuria syndrome (HCS) and even been shown in be the source of the hypotonia seen in affected individuals (Jaeken et al., 2006; Regal et al., 2014; Silva et al., 2018), its role in the disease remains unknown, in large part because the substrate has yet to be elucidated. It most closely resembles PREP, a prolyl endopeptidase, in sequence yet despite many efforts, no peptide substrate has been identified. However, sequence alignment clearly indicates that PREPL maintains the catalytic triad Ser-Asp-His required for enzymatic activity (Martens et al., 2006).

Knowing this, we sought to determine whether PREPL's sequence or structure suggested an alternative activity to cleavage of a peptide bond. Using a structural model of PREPL, we compared its secondary structure to that of PREP. Interestingly, the three-dimensional structures were extremely similar, both consisting of an α/β hydrolase fold in addition to a β propeller covering the central tunnel. Though the three-dimensional structures appeared similar, it seemed unlikely that PREPL acted as a peptidase given previous studies (Boonen et al., 2011; Szeltner et al., 2005). In order to reassess the possible enzymatic activities PREPL could be carrying out in the cell, we next took a closer look at the other members of the S9 serine hydrolase family. While we noticed a number of peptidases, the family also includes esterases and lipases, notably PAFAH2 and PLA2G7, and one protein, APEH, that has been shown to exhibit both esterase and peptidase activity (Bartlam et al., 2004; Scaloni et al., 1994; Wang et al., 2006).

Given the lack of peptidase activity and the esterase members of the S9 serine hydrolase family, we next compared PREPL's active site to other peptidases, esterases, and lipases, looking for residues that may indicate any activity other than cleavage of amide bonds. We found two amino acids that suggest PREPL may act as an esterase *in vivo*. First, the catalytic serine of

PREPL contains an alanine (A) rather than a glycine (G) in the motif containing its catalytic serine (AxSxG vs GxSxG) which is also found in the lipases LPLA2 and C1d1p. Second, PREPL has an aspartate (D) near the active site histidine (HxxxD motif) similar to a number of lipases including the S9 family members PAFAH2 and PLA2G7, which may act as an acyltransferase.

We found that PREPL exhibited low, but reproducible activity against pNP-C8, a medium length carbon chain consisting of eight carbons *in vitro*. Importantly, this activity was abolished upon mutation of the catalytic serine and was specific to PREPL, as PREP showed no activity against the ester substrate. Based on our *in vitro* assays, we hypothesized that PREPL cleaved ester bonds. We tested this using a previously described PREPL$^{-/-}$ mouse model (Lone et al., 2014). Because it remains unclear whether PREPL's *in vivo* substrate is a lipid or a small molecule, we performed both lipidomics and metabolomics on brain tissue of wild type (WT) and PREPL$^{-/-}$ mice. We identified 2,702 lipids and 191 metabolites, including 28 lipids that were significantly altered in the PREPL$^{-/-}$ brain samples.

Taken together, our results illustrate that though sequence homology suggests that PREPL is a peptidase, closer inspection of other family members and residues around the active site suggest that PREPL could act as an esterase/lipase *in vivo*. Our data demonstrates the first specific, catalytic activity of the PREPL against an esterase (or any other substrate). Additionally, we show that PREPL$_L$ localizes to the mitochondria and that loss of PREPL alters brain mitochondrial lipid profiles. However, it is unclear how these alterations may contribute to HCS etiology and it remains to be determined whether these alterations are due to direct or indirect effects.

Future Directions

Verification of candidate genes identified in lQTLs

Although it is often difficult, the end goal of any QTL mapping study is to identify a gene (or genes) in an identified locus which is verified to be directly responsible for the phenotype mapped there. For *cis*-eQTLs and pQTLs verification process is simpler, as the phenotype is usually directly attributed to the gene/protein in question. However, lipids lack such a straightforward linkage. Additionally, due to the complexity of lipid metabolism (overlapping pathways, etc.) it is likely that multiple genes are regulating the abundance of a given species, making it even more difficult to identify causal genes. Despite these difficulties, it nevertheless remains the end goal of QTL mapping studies, and should be the focus of future studies.

The resource of candidate genes provided in (Jha et al., 2018a; Jha et al., 2018b) serve as a foundation of future analyses, in our group and others. Initially, we selected six candidate genes from our list of candidates that fulfilled all four of our pipeline requirements to test in cell culture: Methylthioadensine phosphorylase (MTAP), phospholipase B domain containing 1 (Plbd1), fatty acid binding protein 1 (Fabp1), NADPH-dependent 3 keto-steroid reductase (Hsd3b5), UbiA prenyltransferase domain containing 1 (ubiAD1) and zinc finger protein 277 (zfp277.) Though some of these, such as Fabp1 and Hsd3b5 were already known to be associated with lipid metabolism, the effect each had on individual specific lipid species, if any, were unknown. We hypothesized that if these genes were truly causative, overexpression or knockdown in cell culture would alter the lipid profile compared to wild type (WT), specifically those lipids which mapped to the area of the gene in question. Because these lQTLs were initially identified in mouse liver tissue, we conducted overexpression and knockdown experiments in a mouse liver cell line, AML12. However, we were unable to get >90% knockdown of any gene and cells would not overexpress the constructs. Given these difficulties we were never able to test whether the candidate genes affected lipid profiles. If these avenues are to be pursued in the future we will use an alternate cell line – either another immortal mouse liver cell line or one such as HEK293s which are robust and easy to transfect – and likely knockout the genes using

CRISPR-Cas9 technology (Cong et al., 2013; Jinek et al., 2012; Qi et al., 2013) to ensure so remnant protein remains.

Outside of cell culture, there are other ways to follow-up on candidate genes. One method would be to try to increase resolution by including additional BXD strains (especially those which exhibit homologous recombination in the area of interest) or perhaps even more useful, strains from other mouse genetic reference populations, as these strains will exhibit genetic differences not seen in the BXD cohort. *In vitro* assays and knockout mouse models can be used to directly test a protein's role in lipid metabolism. Our lab and others will continue to build an lQTL resource and utilize multi-omic approaches to prioritize and eventually verify candidate genes novel roles in lipid metabolism.

The expansion of lQTL mapping

Perhaps the most important aspect of our work is the demonstration that individual lipids are quantifiable traits that can used in QTL mapping studies. From there, the possibilities are nearly endless. One caveat of the BXD cohort is that because the two founders strains are both inbred laboratory strains, they are the same at many genetic locations, making these areas blind to QTL mapping. Work is already being done to expand lQTL mapping into more diverse cohorts such as the Collaborative Cross (Chesler et al., 2008) which will provide better resolution as well as identify QTLs that could not be identified via the BXD cohort due to the increased genetic diversity. Coupled clinical phenotype QTLs, eQTLs, and pQTLs, this will be a powerful resource for future insight into the regulation of lipid metabolism.

Additionally, though we limited our study to liver because it is the hub of metabolism, it is not the only tissue in which lipid regulation is important. Thus, future work should expand lQTL mapping to include additional tissues such as brain where lipid signaling is important for important processes such as synaptic activation (Bazan, 2005; Ogasawara et al., 2016). Furthermore, given that we found the abundance of some plasma lipid species to be indicative of their levels in liver,

it is possible that by looking at all tissues we will be able to better understand the tissues and processes involved in regulating plasma lipid levels. Moreover, similar to the NAFLD signatures we identified, it may be possible to identify other plasma lipids which reflect other disease states.

Another caveat to the studies described in this thesis is a lack of granularity at the cellular and organellar level. The liver consists of multiple cell types, including hepatocytes, hepatic stellate cells, Kupffer cells, and liver sinusoidal endothelial cells, each having its own unique role in proper lipid function (Ding et al., 2016). Therefore, it is not far-fetched to assume that each of these cells likely contains unique lipid profiles. By homogenizing lipid pieces from different areas of the liver, not only are we inducing heterogeneity in biological samples, but we may also be masking the alterations in lipid abundance in specific cell types. Furthermore, it is known that lipid composition or organelles is not always reflective of the lipid composition of whole cells (van Meer and de Kroon, 2011) and that lipids are often transferred between organelles. By focusing on individual organelles, future lQTL mapping studies may be able to better understand the regulation and importance of proper lipid transport between organelles and identify novel genes involved in these processes.

Finally, even over the past few years lipidomics technology has advanced substantially. Instead of identifying ~100 lipid species, programs such as Lipidex (Hutchins et al., 2018) can now identify thousands. Future lQTL studies should expand on the lipids identified and include other lipid classes including lysophospholipids, sphingolipids, and cholesterol esters. Additional work should be applied to identifying peaks that were quantified by not identified by the software. For example, if multiple unknown species map to the same locus and share a similar fragmentation pattern it is possible they make up a novel lipid class regulated by a gene in that area. In order to fully understand lipid metabolism, it will be important to push forward in defining the entire cellular lipidome.

Automating the prioritization of candidate genes

As lQTL mapping continues to expand so too does eQTL and pQTL mapping. Though more data is often a good thing, it is useless if we do not know how to analyze it. We present a simple method of prioritizing candidate genes (Jha et al., 2018a; Jha et al., 2018b) however, it requires manual inspection subject to some observer subjectivity and ignores data such as *trans*-eQTLs and pQTLs. As databases continue to grow, identification of candidate genes becomes a bottleneck. In the future, great efforts should be made to not only automate the process of selecting and ranking candidate but to create a truly multi-omic platform for data analysis.

Investigation of lipids altered by loss of PREPL

Though software was unable to identify the lipids altered in the PREPL$^{-/-}$ mice, due to inability to match the fragmentation pattern to any lipids in the searched *in silico* libraries. Future work will include manually inspecting the fragmentation spectra in an effort to identify structural similarities among lipids specifically altered in the PREPL$^{-/-}$ mice. Due to the promiscuous nature of many serine hydrolases it is possible that, *in vivo*, PREPL acts on multiple lipids of an as of yet unidentified lipid class. Additionally, though we may be unable to identify the lipid species in question, it is possible that the fragmentation data may reveal an important functional group or moiety for PREPL substrates, which we could test in our established *in vitro* assays.

Characterization of the effects of loss of PREPL in HAP1 cells

In a separate study, we have measured both lipids and metabolites in two unique PREPL knockout (KO) HAP1 cell lines. Future efforts will directly compare the mouse and cell culture datasets, by searching the samples together to maximize identified lipids and to ensure quantified but unidentified peaks are indeed the same lipid species in both sample sets. However, it should be noted that because these samples were run at separate times, retention time shifts may limit the lipids identified in both samples. Since both KO cell lines have been confirmed, we

hypothesize that direct effects of loss of PREPL will be seen in both KO lines, while indirect or non-specific effects are likely to be seen only in a single KO. Therefore, future work will limit possible substrates to those lipids significantly altered in both HAP KO lines.

Current evidence does not definitively indicate whether the substrate of PREPL is a lipid or a metabolite. Despite efforts to identify metabolites altered in the PREPL$^{-/-}$ mice, sucrose interference not only limited the species we identified, but may also have overwhelmed some of the changes. Metabolomics of the HAP1 KOs did not suffer from this interference, therefore, future work will seek to identify metabolites altered in these cell lines compared to WT. Admittedly, these metabolomic analyses took place on whole cells, which may mask the alterations if they are mitochondrial specific. Therefore, our next steps should include purification of mitochondria from the HAP1 cell lines and subjection of them to metabolomic analysis. Interestingly, comparing the results from the mitochondria to supernatant/cytosolic fraction may also give insight into whether or not PREPL$_L$ and PREPL$_S$ have distinct roles in the cell or merely perform the same action in different subcellular localizations.

If, as we hypothesize, PREPL$_L$, the mitochondrial isoform of the protein, is responsible for the hypotonia seen in individuals suffering from HCS it is possible that loss of PREPL has a deleterious effect on mitochondrial function. One way to test this is to measure the oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) of WT and PREPL KO HAP1 cell lines using the Seahorse XF Analyzer (Wu et al., 2007). Interestingly, some individuals with PREPL deletions exhibit mitochondrial dysfunction (Legati et al., 2016). Future work will determine whether loss of PREPL alters mitochondrial respiration and if so, if this phenotype is rescuable by addition of active PREPL. Not only could defects in mitochondrial respiration verify the importance of PREPL$_L$ but would also offer insight into mechanism by which loss of PREPL causes hypotonia.

The role of PREPL in HCS etiology

Once the *in vivo* substrate of PREPL has been identified, its role in the cell can be further investigated. Though we are a long ways from explaining the mechanism by which PREPL causes hypotonia in affected individuals, there are nevertheless a few interesting questions to be speculated upon.

First, what are the functions of the two PREPL isoforms *in vivo*? Upon identification of substrates, future work should identify the role of each isoform. Though both isoforms exhibited equal esterase in our hands, and it is unlikely that substrate specificity would be altered by protein structure as the MLS on $PREPL_L$ is likely cleaved upon entrance into the mitochondria leaving a protein that is essentially identical to $PREPL_S$, it is possible that, due to their different subcellular locations they work on different substrates. Substrate availability, product concentration, or alternative interacting partners based on localization could influence the *in vivo* substrate of each isoform. Second, why would a protein highly expressed in the brain have such a marked effect on muscle tone? Perhaps PREPL's substrate or product is an important signaling molecule. Though we cannot answer these questions right now, overall, these results suggest a new activity for PREPL and demonstrate new insight into the identity of PREPL's *in vivo* substrate. Furthermore, the biochemical analyses performed for this project provide a framework for investigation into candidate genes in lQTL studies.

References

Bartlam, M., Wang, G., Yang, H., Gao, R., Zhao, X., Xie, G., Cao, S., Feng, Y., and Rao, Z. (2004). Crystal structure of an acylpeptide hydrolase/esterase from Aeropyrum pernix K1. Structure (London, England : 1993) *12*, 1481-1488.

Bazan, N.G. (2005). Lipid signaling in neural plasticity, brain repair, and neuroprotection. Molecular neurobiology *32*, 89-103.

Boonen, K., Regal, L., Jaeken, J., and Creemers, J.W. (2011). PREPL, a prolyl endopeptidase-like enzyme by name only?--Lessons from patients. CNS & neurological disorders drug targets *10*, 355-360.

Chesler, E.J., Miller, D.R., Branstetter, L.R., Galloway, L.D., Jackson, B.L., Philip, V.M., Voy, B.H., Culiat, C.T., Threadgill, D.W., Williams, R.W., et al. (2008). The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mammalian genome : official journal of the International Mammalian Genome Society *19*, 382-389.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (New York, N.Y.) *339*, 819-823.

Ding, C., Li, Y., Guo, F., Jiang, Y., Ying, W., Li, D., Yang, D., Xia, X., Liu, W., Zhao, Y., et al. (2016). A Cell-type-resolved Liver Proteome. Molecular & cellular proteomics : MCP *15*, 3190-3202.

Granger, J., and Remick, D. (2005). Acute pancreatitis: models, markers, and mediators. Shock (Augusta, Ga.) *24 Suppl 1*, 45-51.

Gronert, K., Kantarci, A., Levy, B.D., Clish, C.B., Odparlik, S., Hasturk, H., Badwey, J.A., Colgan, S.P., Van Dyke, T.E., and Serhan, C.N. (2004). A molecular defect in intracellular lipid signaling in human neutrophils in localized aggressive periodontal tissue damage. Journal of immunology *172*, 1856-1861.

Hannun, Y.A. (1994). The sphingomyelin cycle and the second messenger function of ceramide. The Journal of biological chemistry *269*, 3125-3128.

Hutchins, P.D., Russell, J.D., and Coon, J.J. (2018). LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. Cell systems *6*, 621-625.e625.

Jaeken, J., Martens, K., Francois, I., Eyskens, F., Lecointre, C., Derua, R., Meulemans, S., Slootstra, J.W., Waelkens, E., de Zegher, F., et al. (2006). Deletion of PREPL, a gene encoding a putative serine oligopeptidase, in patients with hypotonia-cystinuria syndrome. American journal of human genetics *78*, 38-51.

Jha, P., McDevitt, M.T., Gupta, R., Quiros, P.M., Williams, E.G., Gariani, K., Sleiman, M.B., Diserens, L., Jochem, A., Ulbrich, A., et al. (2018a). Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species. Cell systems *6*, 722-733.e726.

Jha, P., McDevitt, M.T., Halilbasic, E., Williams, E.G., Quiros, P.M., Gariani, K., Sleiman, M.B., Gupta, R., Ulbrich, A., Jochem, A., et al. (2018b). Genetic Regulation of Plasma Lipid Species and Their Association with Metabolic Phenotypes. Cell systems *6*, 709-721.e706.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (New York, N.Y.) *337*, 816-821.

Kooijman, E.E., Chupin, V., de Kruijff, B., and Burger, K.N. (2003). Modulation of membrane curvature by phosphatidic acid and lysophosphatidic acid. Traffic (Copenhagen, Denmark) *4*, 162-174.

Kooijman, E.E., Chupin, V., Fuller, N.L., Kozlov, M.M., de Kruijff, B., Burger, K.N., and Rand, P.R. (2005). Spontaneous curvature of phosphatidic acid and lysophosphatidic acid. Biochemistry *44*, 2097-2102.

Koybasi, S., Senkal, C.E., Sundararaj, K., Spassieva, S., Bielawski, J., Osta, W., Day, T.A., Jiang, J.C., Jazwinski, S.M., Hannun, Y.A., et al. (2004). Defects in cell growth regulation by C18:0-ceramide and longevity assurance gene 1 in human head and neck squamous cell carcinomas. The Journal of biological chemistry *279*, 44311-44319.

Kroesen, B.J., Pettus, B., Luberto, C., Busman, M., Sietsma, H., de Leij, L., and Hannun, Y.A. (2001). Induction of apoptosis through B-cell receptor cross-linking occurs via de novo generated C16-ceramide and involves mitochondria. The Journal of biological chemistry *276*, 13606-13614.

Lee, J.Y., Sohn, K.H., Rhee, S.H., and Hwang, D. (2001). Saturated fatty acids, but not unsaturated fatty acids, induce the expression of cyclooxygenase-2 mediated through Toll-like receptor 4. The Journal of biological chemistry *276*, 16683-16689.

Legati, A., Reyes, A., Nasca, A., Invernizzi, F., Lamantea, E., Tiranti, V., Garavaglia, B., Lamperti, C., Ardissone, A., Moroni, I., et al. (2016). New genes and pathomechanisms in mitochondrial disorders unraveled by NGS technologies. Biochimica et biophysica acta *1857*, 1326-1335.

Li, S., Xu, P., Han, L., Mao, W., Wang, Y., Luo, G., and Yang, N. (2017). Disease-syndrome combination modeling: metabolomic strategy for the pathogenesis of chronic kidney disease. Scientific reports *7*, 8830.

Lone, A.M., Leidl, M., McFedries, A.K., Horner, J.W., Creemers, J., and Saghatelian, A. (2014). Deletion of PREPl causes growth impairment and hypotonia in mice. PloS one *9*, e89160.

Lou, W., Reynolds, C.A., Li, Y., Liu, J., Huttemann, M., Schlame, M., Stevenson, D., Strathdee, D., and Greenberg, M.L. (2018). Loss of tafazzin results in decreased myoblast differentiation in C2C12 cells: A myoblast model of Barth syndrome and cardiolipin deficiency. Biochimica et biophysica acta. Molecular and cell biology of lipids *1863*, 857-865.

Martens, K., Derua, R., Meulemans, S., Waelkens, E., Jaeken, J., Matthijs, G., and Creemers, J.W. (2006). PREPL: a putative novel oligopeptidase propelled into the limelight. Biological chemistry *387*, 879-883.

Ogasawara, D., Deng, H., Viader, A., Baggelaar, M.P., Breman, A., den Dulk, H., van den Nieuwendijk, A.M., Soethoudt, M., van der Wel, T., Zhou, J., et al. (2016). Rapid and profound

rewiring of brain lipid signaling networks by acute diacylglycerol lipase inhibition. Proceedings of the National Academy of Sciences of the United States of America *113*, 26-33.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183.

Regal, L., Shen, X.M., Selcen, D., Verhille, C., Meulemans, S., Creemers, J.W., and Engel, A.G. (2014). PREPL deficiency with or without cystinuria causes a novel myasthenic syndrome. Neurology *82*, 1254-1260.

Scaloni, A., Barra, D., Jones, W.M., and Manning, J.M. (1994). Human acylpeptide hydrolase. Studies on its thiol groups and mechanism of action. The Journal of biological chemistry *269*, 15076-15084.

Silva, S., Miyake, N., Tapia, C., and Matsumoto, N. (2018). The second point mutation in PREPL: a case report and literature review. Journal of human genetics *63*, 677-681.

Szeltner, Z., Alshafee, I., Juhasz, T., Parvari, R., and Polgar, L. (2005). The PREPL A protein, a new member of the prolyl oligopeptidase family, lacking catalytic activity. Cellular and molecular life sciences : CMLS *62*, 2376-2381.

van Meer, G., and de Kroon, A.I. (2011). Lipid map of the mammalian cell. Journal of cell science *124*, 5-8.

Vreken, P., Valianpour, F., Nijtmans, L.G., Grivell, L.A., Plecko, B., Wanders, R.J., and Barth, P.G. (2000). Defective remodeling of cardiolipin and phosphatidylglycerol in Barth syndrome. Biochemical and biophysical research communications *279*, 378-382.

Wang, Q., Yang, G., Liu, Y., and Feng, Y. (2006). Discrimination of esterase and peptidase activities of acylaminoacyl peptidase from hyperthermophilic Aeropyrum pernix K1 by a single mutation. The Journal of biological chemistry *281*, 18618-18625.

Weijers, R.N. (2012). Lipid composition of cell membranes and its relevance in type 2 diabetes mellitus. Current diabetes reviews *8*, 390-400.

Wu, M., Neilson, A., Swift, A.L., Moran, R., Tamagnine, J., Parslow, D., Armistead, S., Lemire, K., Orrell, J., Teich, J., et al. (2007). Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. American journal of physiology. Cell physiology *292*, C125-136.