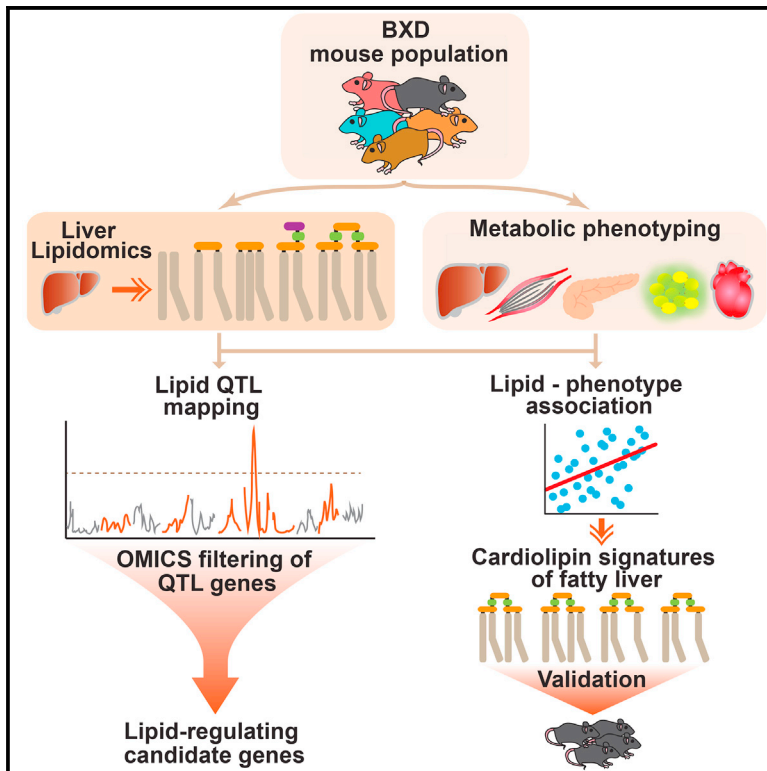


## Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species

### Graphical Abstract



### Authors

Pooja Jha, Molly T. McDevitt, Rahul Gupta, ..., Joshua J. Coon, Johan Auwerx, David J. Pagliarini

### Correspondence

admin.auwerx@epfl.ch (J.A.), dpagliarini@morgridge.org (D.J.P.)

### In Brief

Jha et al. demonstrate the potential of liver lipid species to reflect liver-associated phenotypic metabolic traits in the BXD mouse genetic population. They identify a subset of cardiolipin species—the essential inner mitochondrial membrane phospholipids—that are signatures of healthy or fatty liver. Furthermore, multi-layered omics data (genetic, transcriptomic, proteomic) were employed to filter genes from lipid-regulating loci and provide a resource of candidate genes that may regulate lipid species in mouse and human.

### Highlights

- Lipid modules are associated with phenotypic traits and metabolic pathways
- Specific cardiolipin species are associated with healthy or fatty liver signatures
- Most lipid species are complex traits, regulated by loci spread across the genome
- A new lipid QTL mining pipeline identifies plausible genes regulating lipid species

# Systems Analyses Reveal Physiological Roles and Genetic Regulators of Liver Lipid Species

Pooja Jha,<sup>1,7</sup> Molly T. McDevitt,<sup>2,3,7</sup> Rahul Gupta,<sup>2,7</sup> Pedro M. Quiros,<sup>1</sup> Evan G. Williams,<sup>1,6</sup> Karim Gariani,<sup>1</sup> Maroun B. Sleiman,<sup>1</sup> Leo Diserens,<sup>1</sup> Adam Jochem,<sup>2</sup> Arne Ulbrich,<sup>4</sup> Joshua J. Coon,<sup>4,5</sup> Johan Auwerx,<sup>1,8,\*</sup> and David J. Pagliarini<sup>2,3,\*</sup>

<sup>1</sup>Laboratory of Integrative and Systems Physiology, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

<sup>2</sup>Morgridge Institute for Research, Madison, WI 53715, USA

<sup>3</sup>Department of Biochemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

<sup>4</sup>Department of Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

<sup>5</sup>Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, WI 53706, USA

<sup>6</sup>Present address: Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland

<sup>7</sup>These authors contributed equally

<sup>8</sup>Lead Contact

\*Correspondence: [admin.auwerx@epfl.ch](mailto:admin.auwerx@epfl.ch) (J.A.), [dpagliarini@morgridge.org](mailto:dpagliarini@morgridge.org) (D.J.P.)

<https://doi.org/10.1016/j.cels.2018.05.016>

## 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 (IQTL). By multi-layered omics analyses, we show the reliability of IQTLs to uncover candidate genes that can regulate the levels of lipid species. Additionally, we identified IQTLs 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 toward 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 in understanding 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, lipid-protein, or lipid-phenotype correlations, providing a more comprehensive view of how lipids fit into 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 chow diet (CD) and 41 fed a high-fat diet (HFD) from 47 BXD strains. Over 29 weeks, 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 versus 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] and cardiolipins [CL, 23 species]), and coenzyme Q<sub>9</sub> (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 (Figures 1B and S1D [ $p < 1e-04$ ] 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 on HFD (green nodes; Figures 1B and S1D). While the CLs, found primarily in the mitochondrial inner membrane, formed a tight cluster in both diets, the PLs, which are the general cellular membrane lipids, were highly interspersed between different lipid classes in both diets (red and yellow nodes, respectively; Figures 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).

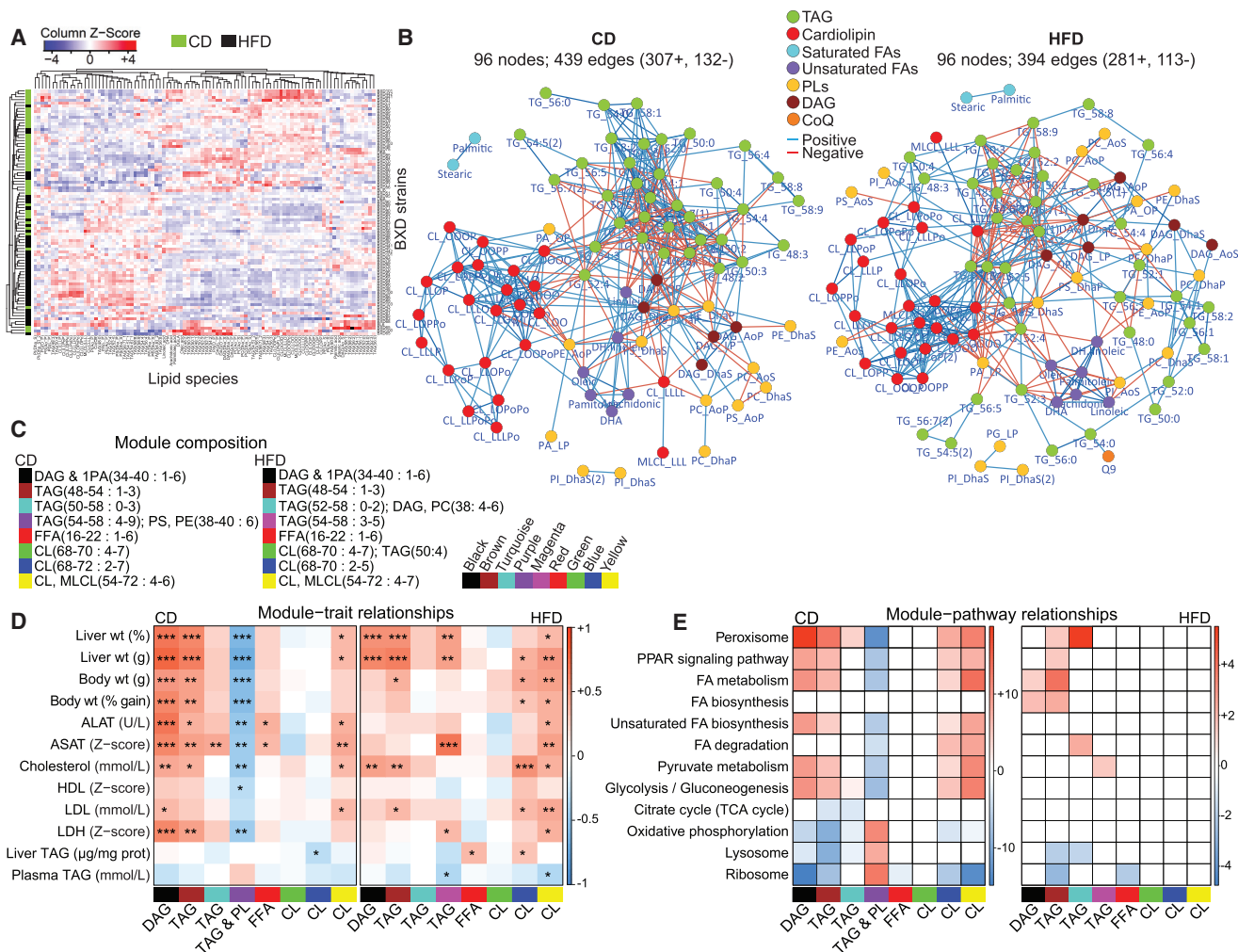
To identify lipids with similar physiological and molecular characteristics, we performed a weighted correlation network analysis (WGCNA), which clusters correlated groups of lipids into modules (Langfelder and Horvath, 2008). More than half of the lipids were clustered in eight modules in both diets (66% in CD and 58% in HFD) (Figures 1C and 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 or 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, and peroxisome and peroxisome proliferator-activated receptor signaling (Figure 1E). Furthermore, these modules correlated negatively with oxidative and proliferative pathways such as oxidative phosphorylation, lysosome, and ribosome pathways (Figure 1E). Meanwhile the TAG and 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 six 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



**Figure 1. Liver Lipid Species Profiles 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 ( $p < 1e-04$ ) of all lipid species measured in CD and HFD. Lipid species are color coded as seven major lipid classes. The side-chain fatty acid composition of lipids has been abbreviated (O, oleic acid; P, palmitic acid; Po, palmitoleic acid; S, stearic acid; L, linoleic acid; Dha, docosa-hexaenoic 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 asterisks in the cells represent the p value of the correlation (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

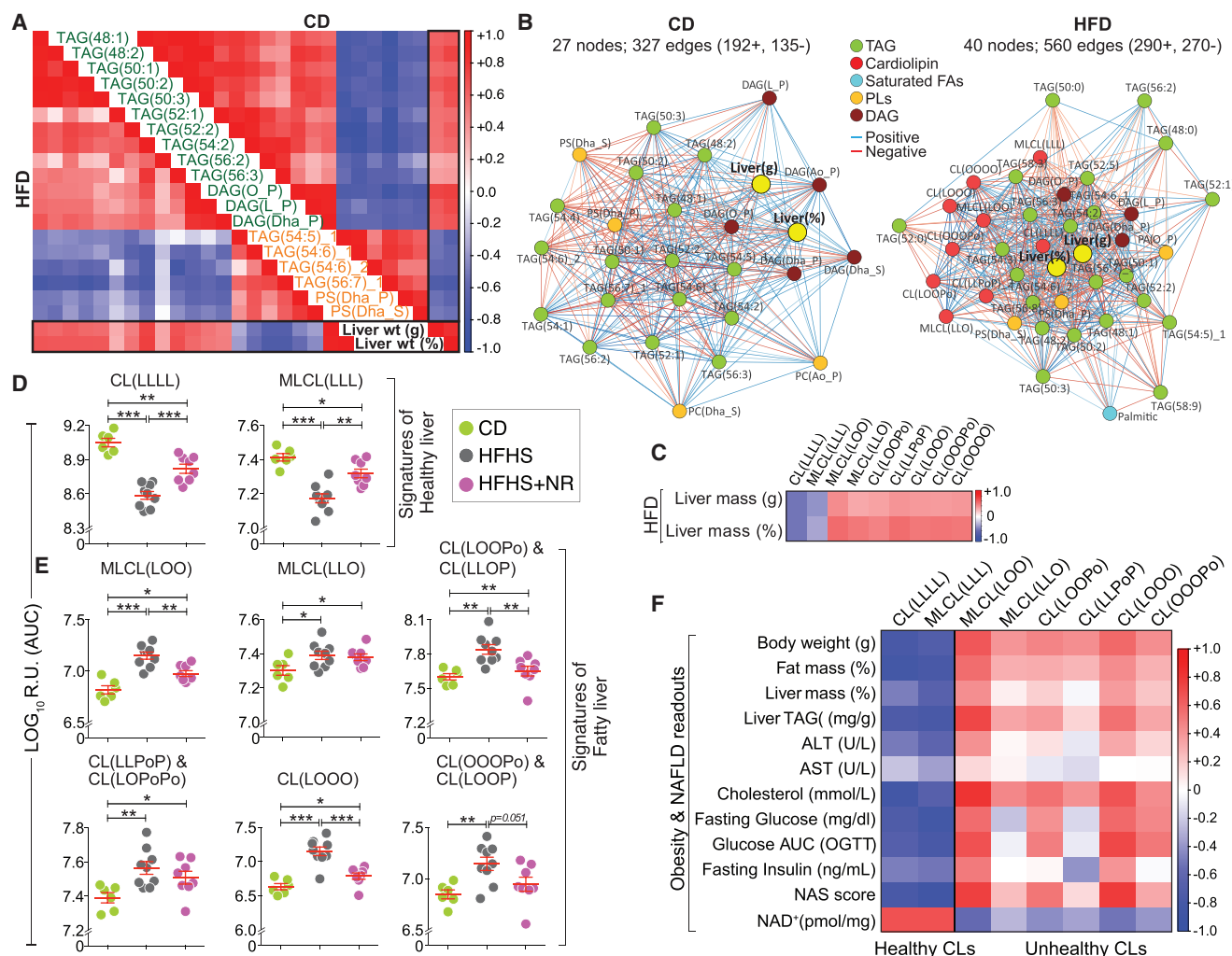
(E) Module and its corresponding KEGG enriched pathway correlation. Red and blue cells represent the enriched pathways with the positively (scale bar:  $\log_{10}$  p value) and negatively (scale bar:  $-\log_{10}$  p value) correlated proteins, respectively. Lipid classes hereafter are abbreviated as 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.

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 with 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 monolysocardiolipin (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 Koeh-

ler, 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 CL 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 fatty acids (MUFAs), oleic (O) acid, and palmitoleic (Po) acid





**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 or 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 were inseparable chromatographically.

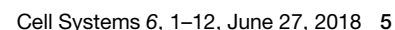
(F) Corrgram showing negative correlation of obesity and NAFLD traits with healthy CL species and positive correlation with unhealthy CL species.

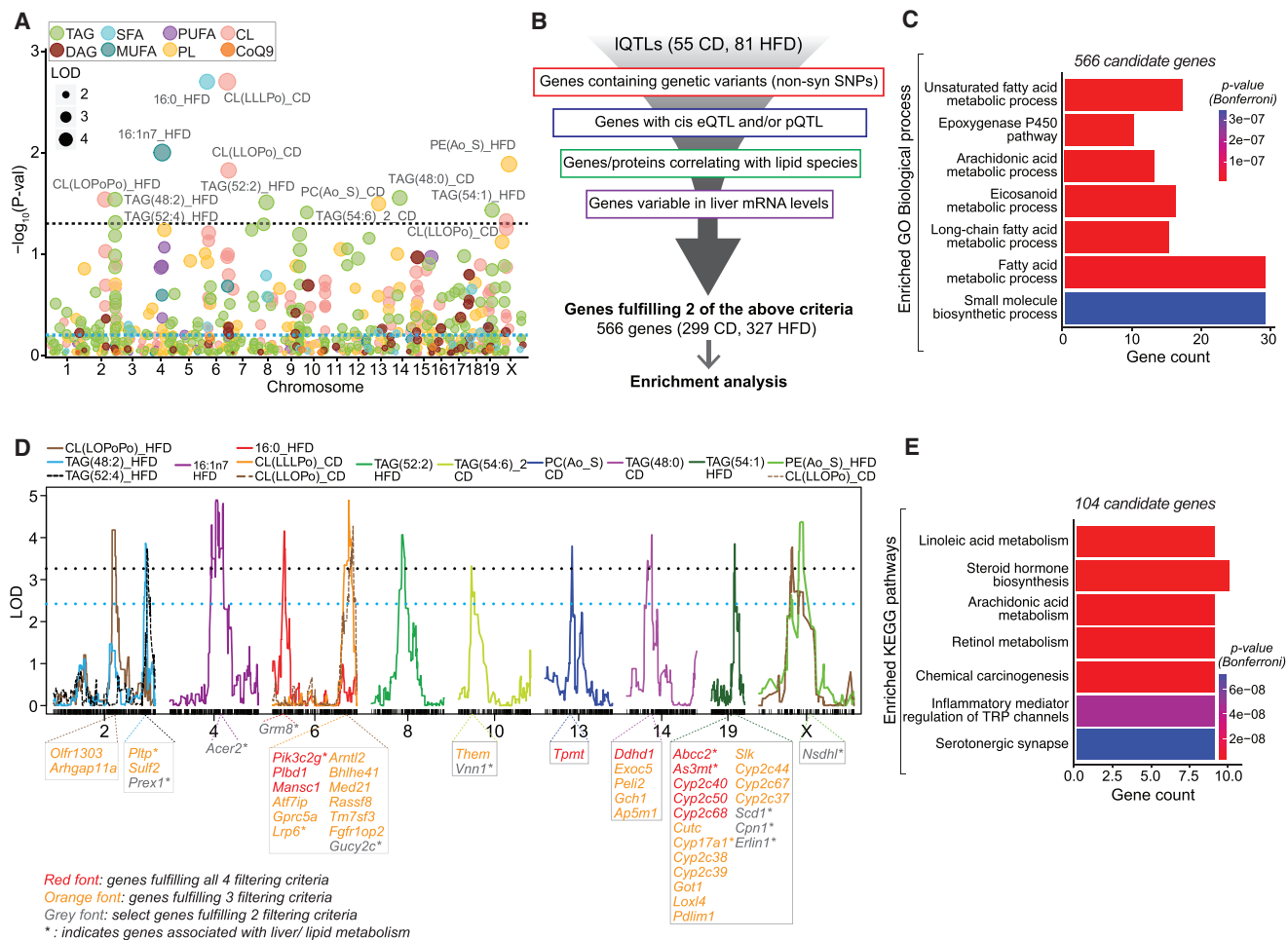
For (D) and (E), data are represented as means  $\pm$  SEM. \*p < 0.05, \*\*p < 0.001; \*\*\*p < 0.0001.

(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 species predominant in healthy tissue—CL(LLLL) (Chicco and Sparagna, 2007) and its precursor, MLCL(LLLL)—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 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

whether the profile can be reverted by ameliorating hepatic steatosis by 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<sup>+</sup>) 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 (1) CD, (2) HFHS diet for 18 weeks, or (3) 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





**Figure 4. Genetic Assessment of IQTLs**

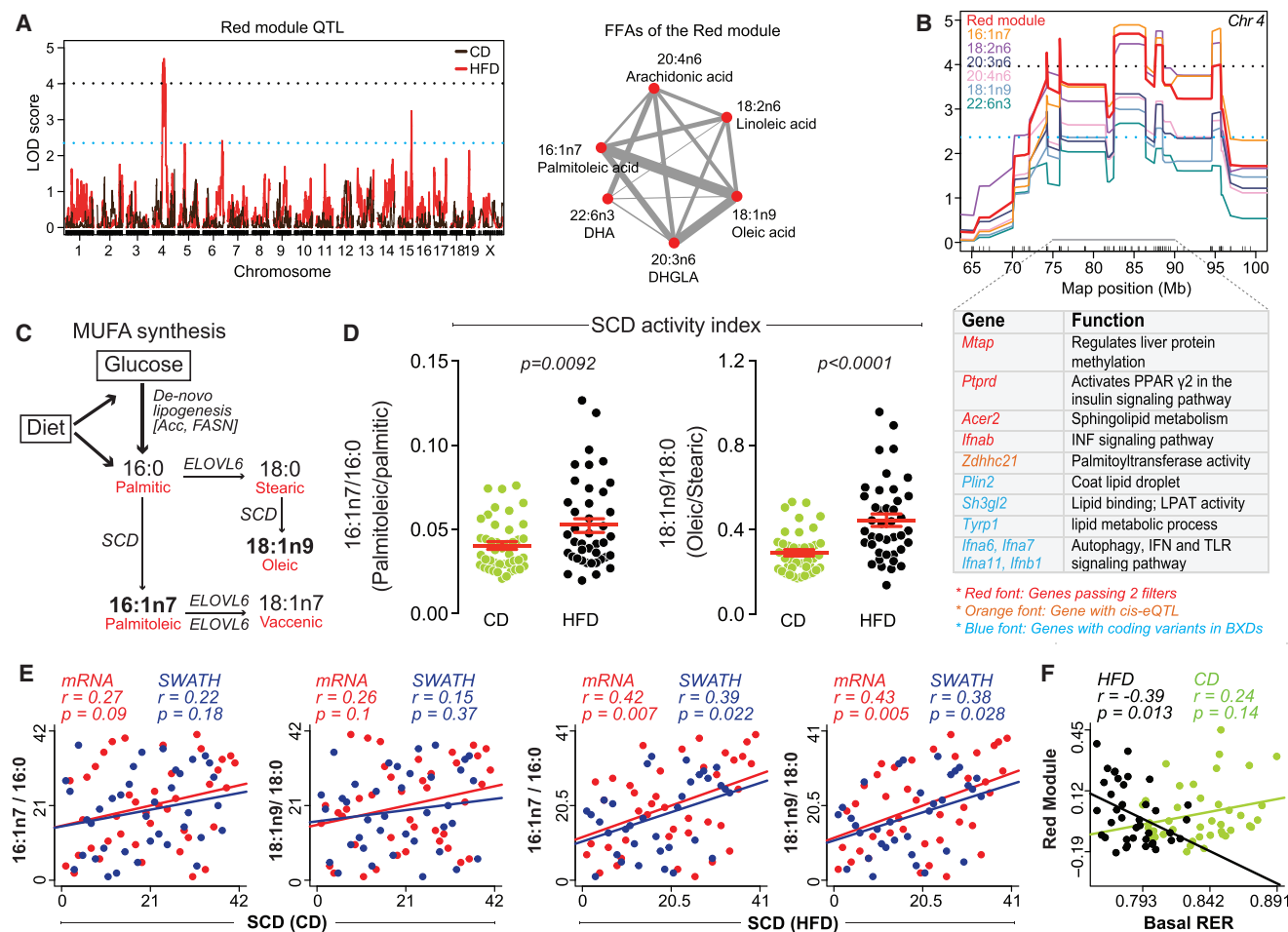
(A) Manhattan plot of lipid species in CD and HFD. Names of the lipid species with genome-wide  $p$  value  $< 0.05$  are indicated. (B) All genes under the IQTLs ( $\pm 5$  Mb from the peak) were filtered through four independent pipelines as indicated. Genes fulfilling two or more of the filtering criteria were analyzed for enrichment of GO biological process (BP). (C) Enriched GO BP from 566 filtered IQTL genes. (D) IQTL position of the top 14 IQTLs ( $p < 0.05$ ). The candidate genes fulfilling four (red font), three (orange font), and two (gray font) of the four filtering criteria are indicated below each IQTL peak. Genes indicated with asterisks are associated with metabolic phenotypes. (E) KEGG enrichment analysis of genes under the top 14 IQTLs passing two or more 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.

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 the rest of the IQTLs suggests that GxE factors regulate nearly all lipid species.

### Genetic Assessment of the IQTLs

To assess the efficacy of IQTLs and find candidate genes that regulate lipid levels, we performed an in-depth analysis of all significant and suggestive IQTLs (Figure 4A, IQTLs above the blue dotted line; 55 in CD and 81 in HFD). All genes under these IQTLs were filtered along four parallel pipelines (Figure 4B): genes un-

der IQTLs (1) with non-synonymous SNPs in BXDs, (2) with self-regulating QTLs (*cis*- and/or *p*-QTL), (3) with significant correlation ( $p < 0.05$ ) with the lipid itself, and (4) with variable transcript expression ( $SD > 0.25$ ). While the integration of proteomics data was advantageous, transcriptomics data were substantially more consequential to candidate selection as more genes were assayed ( $\sim 20,000$  with transcriptomic versus  $\sim 2,600$  with proteomic data), and this also avoids post-translational and protein regulation variations. Genes passing two of the four filters in either diet (299 of 2,845 genes in CD and 327 of 3,818 in HFD) were analyzed for gene ontology (GO) pathway enrichment (Figure 4B). Remarkably, the filtered IQTL candidate genes were enriched in lipid metabolic pathways (Figure 4C). We next applied the same filtering criteria for the top 14 IQTLs (Figure 4A, IQTLs above the black dotted line) comprising 4 CL,



**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  $\pm$  SEM. (E) Spearman correlation between hepatic expression of SCD transcript and protein (SWATH) with SCD activity index. (F) 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.

6 TAG, 2 PL, and the FFA 16:0 and 16:1n7 IQTLs (Figure 4D). These IQTLs 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 two or more of the filtering criteria from these 14 IQTLs were subsequently analyzed for enrichment of KEGG pathways. In line with the above findings, these genes were enriched in fatty acid 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 IQTL (Table S6; CD and HFD IQTLs provided in separate excel sheets). Additionally, genes under the IQTLs 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 IQTLs in the identification of putative genetic regulators of lipid species.

### The FFA Module Maps to a Genetic Hotspot Associated with Fatty Acid Metabolism and Signaling

Next, we analyzed one of the diet-specific hotspot IQTL regions to demonstrate the validity of our IQTLs 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 four 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 whether 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 fatty acid (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 whether 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 as 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 toward 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 IQTL Genes Associated with Abnormal Lipid Metabolism in Human GWASs and Mapping of IQTLs in TAG Biosynthetic Pathway

Human genome-wide association studies (GWASs) 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 GWASs (Table S7), we screened the IQTLs for the presence of any human GWAS genes associated with abnormal lipid metabolism (Figure 6A). We screened only those genes under the IQTLs which fulfilled at least two of the four filtering criteria, shown in Figure 4B. Twenty-seven IQTLs (7 CD and 20 HFD) harbored 20 out of 494 genes pre-selected from human GWASs for abnormal lipid metabolism (Figure 6A; see Table S7 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 IQTL 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

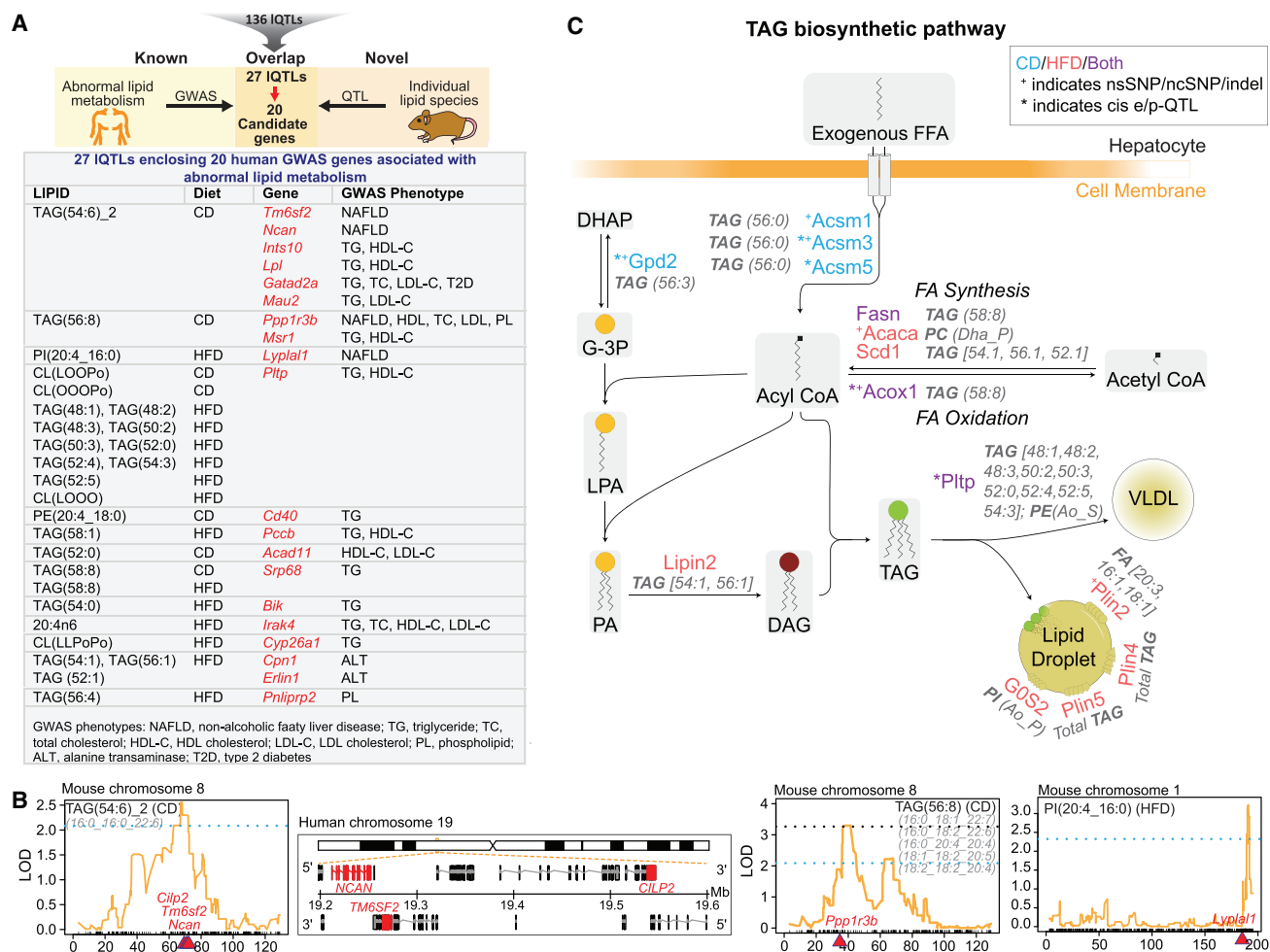
20 or more GWAS genes found to overlap by chance under the calculated null distribution is only 1.49%.

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 GWASs, including *TM6SF2*, *NCAN*, *CILP2*, *PPP1R3B*, and *LYPLAL1* (Anstee and Day, 2013; Kahali et al., 2015; Lusi 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 GWASs to IQTLs 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 eight genes (*PNPLA3*, *GCKR*, *TRIB1*, *LYPLAL1*, *PPP1R3B*, *TM6SF2*, *NCAN*, and *CILP2*) proven to cause or increase the susceptibility to hepatic steatosis/NAFLD in human GWASs (Anstee and Day, 2013; Kahali et al., 2015; Lusi 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 despite no significant dietary difference. 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 eight human NAFLD GWAS genes stated above, nor to any of the individual TAG IQTLs. 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 IQTLs. Fourteen genes (4 CD, 7 HFD, and 3 CD + HFD) involved in TAG biosynthesis were found under 30 IQTLs (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



of the liver lipids was lower than that of plasma lipids. These differences can be attributed to the fact that (1) liver, being the hub of most metabolic processes, is more dynamic and influenced by a multitude of factors (diet, hormones, stress) versus the plasma lipid profile, which represents an equilibrium state as a result of metabolic processes from all peripheral tissues; and (2) additional steps such as tissue homogenization and normalization to protein levels are required for liver lipidomics, increasing the possibility of technical error.

A high-fat dietary challenge had a striking impact on the cardiometabolic 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 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/or may in itself serve as a disease biomarker. If this is the case, 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<sup>+</sup> 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 (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 far 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 IQTLs, we performed a multi-omics analysis of candidate genes under the IQTLs, 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 show several IQTLs harboring human GWAS candidate genes for lipids and associated metabolic traits. Although with this overlap we cannot infer whether the lipid-modulating allele associated with the gene is acting in the same direction in both mouse and human, for genes that have been mechanistically examined in the literature this directionality can be checked. 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 IQTLs 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 with individual TAG IQTLs. 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 guide the development of clinically relevant disease signatures.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Fatty Acid Composition (%) in the Diet of BXD Cohorts
- METHOD DETAILS
  - Plasma Clinical Traits of BXD Cohorts
  - Body Composition (EchoMRI)
  - TAG Measurement of BXD Livers
  - NAD<sup>+</sup> Measurement of C57BL/6J Livers
  - Lipidomics Sample Preparation and Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatic and Genetic Analyses
  - QTL Candidate Gene Retrieval
  - BXD IQTL and Human GWAS Genes Overlap

- Weighted Gene Correlation Network Analysis
- Enrichment Analysis of Proteins and Transcripts with Liver TAG Concentration
- DATA AND SOFTWARE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and eight tables and can be found with this article online at <https://doi.org/10.1016/j.cels.2018.05.016>.

## ACKNOWLEDGMENTS

We thank the Center of Phenogenomics (CPG, EPFL) for performing BXD phenotyping and R.W. Williams for BXDs and GeneNetwork maintenance. J.A. 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). D.J.P. is supported by a grant from NIH (R01GM115591). M.T.M. and E.G.W. were supported by NIH fellowships (T32GM007215 and F32GM119190, respectively). J.J.C. is supported by NIH grants R35GM118110 and P41GM108538.

## AUTHOR CONTRIBUTIONS

Conceptualization, P.J., D.J.P., and J.A.; MS Extractions, Methodology, and Measurement, M.T.M.; Formal Analysis, M.B.S. and P.J.; Investigation and Analysis, P.J. and P.M.Q.; Candidate genes retrieval, GWAS Analysis, and Circos Plot, P.J., P.M.Q., R.G., and L.D.; Data Curation, P.J. and E.G.W.; Supervision (MS), J.J.C. and D.J.P.; Standardization (MS), A.J. and A.U.; Validation, P.J., R.G., and K.G.; Writing – Original Draft, P.J. and J.A.; Writing – Review & Editing, P.J., J.A., D.J.P., R.G., M.T.M., and E.G.W.; Visualization, P.J., E.G.W., J.A., and D.J.P.; Project Administration and Funding Acquisition, J.A. and D.J.P.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 27, 2018

Revised: April 24, 2018

Accepted: May 11, 2018

Published: June 13, 2018

## 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. *Nat. Rev. Gastroenterol. Hepatol.* 10, 645–655.

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

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

Bigaud, E., and Corrales, F.J. (2016). Methylthioadenosine (MTA) regulates liver cells proteome and methylproteome: implications in liver biology and disease. *Mol. Cell. Proteomics* 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. *Am. J. Physiol. Cell Physiol.* 292, C33–C44.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* 15, 34–48.

Claypool, S.M., and Koehler, C.M. (2012). The complexity of cardiolipin in health and disease. *Trends Biochem. Sci.* 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. *J. Immunol.* 172, 1856–1861.

Han, X. (2016). Lipidomics for studying metabolism. *Nat. Rev. Endocrinol.* 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. *Prog. Lipid Res.* 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 Syst.* 6. Published online June 13, 2018. <https://doi.org/10.1016/j.cels.2018.05.009>.

Jump, D.B. (2011). Fatty acid regulation of hepatic lipid metabolism. *Curr. Opin. Clin. Nutr. Metab. Care* 14, 115–120.

Kahali, B., Halligan, B., and Speliotes, E.K. (2015). Insights from genome-wide association analyses of nonalcoholic fatty liver disease. *Semin. Liver Dis.* 35, 375–391.

Koekemoer, T.C., and Oelofsen, W. (2001). Properties of porcine white adipose tissue and liver mitochondrial subpopulations. *Int. J. Biochem. Cell Biol.* 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. *J. Biol. Chem.* 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. *J. Biol. Chem.* 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 Res.* 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 Res.* 44, D869–D876.

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. *J. Lipid Res.* 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 Res.* 45, D896–D901.

Malhi, H., and Gores, G.J. (2008). Molecular mechanisms of lipotoxicity in nonalcoholic fatty liver disease. *Semin. Liver Dis.* 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 Genet.* 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., Freiburger, 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. *Nat. Biotechnol.* 34, 1191–1197.
- 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. *Nat. Commun.* 7, 10464.
- Wenk, M.R. (2005). The emerging field of lipidomics. *Nat. Rev. Drug Discov.* 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, aad0189.
- 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* 16, 284–287.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Q6 internal standard	Avanti Polar Lipids	Cat# 900150
PC(15:0/15:0) internal standard	Avanti Polar Lipids	Cat# 850350P
PS(17:0/17:0) internal standard	Avanti Polar Lipids	Cat# 840028P
PE(15:0/15:0) internal standard	Avanti Polar Lipids	Cat# 850704P
PA(17:0/17:0) internal standard	Avanti Polar Lipids	Cat# 830856P
PG(15:0/15:0) internal standard	Avanti Polar Lipids	Cat# 840446P
CL(56:0) internal standard	Avanti Polar Lipids	Cat# 750332P
FA(15:0/15:0) internal standard, Pentadecanoic Acid	Sigma-Aldrich	Cat# 91446
SPLASH Lipidomix Mass Spec Standard	Avanti Polar Lipids	Cat# 330707
<b>Critical Commercial Assays</b>		
AST: AST Flex reagent cartridge	Siemens Healthcare	Cat# DF41A
ALT: ALTI Flex® reagent cartridge	Siemens Healthcare	Cat# DF143
LDL: ALDL Flex reagent cartridge	Siemens Healthcare	Cat# DF131
HDL: AHDL Flex reagent cartridge	Siemens Healthcare	Cat# DF48B
Cholesterol: CHOL Flex reagent cartridge	Siemens Healthcare	Cat# DF27
Plasma TG: TGL Flex reagent cartridge	Siemens Healthcare	Cat# DF69A
LDH: LDI Flex reagent cartridge	Siemens Healthcare	Cat# DF54
HR Series NEFA-HR(2)	FUJIFILM Wako Dignostics	<a href="http://www.wakodiagnostics.com/r_nefa.html">http://www.wakodiagnostics.com/r_nefa.html</a>
Glucose: GLUC Flex reagent cartridge	Siemens Healthcare	Cat# DF40
Mercodia Mouse Insulin ELISA kit	Mercodia	Cat# 10-1247-01
Liver TG: Serum Triglyceride Determination Kit	Sigma-Aldrich	Cat#TR0100
<b>Deposited Data</b>		
BXD liver lipidomics data [raw]	This paper	<a href="https://chorusproject.org/anonymous/download/experiment/26fdcc1032254f5fa7095c1aa1e829d8">https://chorusproject.org/anonymous/download/experiment/26fdcc1032254f5fa7095c1aa1e829d8</a> <a href="https://chorusproject.org/anonymous/download/experiment/73040a708ca24efeb84c92bd5e0d9b8c">https://chorusproject.org/anonymous/download/experiment/73040a708ca24efeb84c92bd5e0d9b8c</a>
BXD liver lipidomics data [normalized]	This paper	<a href="http://www.genenetwork.org/">http://www.genenetwork.org/</a> GN Accession: GN602. RecordIDs 19803-19994
BXD mouse genotype data	<a href="http://www.genenetwork.org/">http://www.genenetwork.org/</a>	GN Accession: GN600
BXD mouse metabolic phenotype data	<a href="http://www.genenetwork.org/">http://www.genenetwork.org/</a>	GN Accession: GN602
BXD mouse liver transcriptome data	<a href="http://www.genenetwork.org/">http://www.genenetwork.org/</a>	GN Accession: GN432 (CD), GN431 (HFD)
BXD mouse liver SWATH proteomics data	<a href="http://www.genenetwork.org/">http://www.genenetwork.org/</a>	GN Accession: GN540 (CD), GN541 (HFD)
GWASdb2	(Li et al., 2016)	<a href="http://jjwanglab.org/gwasdb">http://jjwanglab.org/gwasdb</a>
GWAS Catalog	(MacArthur et al., 2017)	<a href="https://www.ebi.ac.uk/gwas/">https://www.ebi.ac.uk/gwas/</a>
<b>Software and Algorithms</b>		
R	The R Foundation	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
MATLAB	Mathworks	<a href="https://www.mathworks.com">https://www.mathworks.com</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MATLAB Bioinformatics Toolbox	Mathworks	<a href="https://www.mathworks.com/products/bioinfo.html">https://www.mathworks.com/products/bioinfo.html</a>
corrgram	The R Foundation	<a href="https://cran.r-project.org/web/packages/corrgram/index.html">https://cran.r-project.org/web/packages/corrgram/index.html</a>
R/qtl (v 1.39-5) package	(Broman et al., 2003)	<a href="http://www.rqtl.org/">http://www.rqtl.org/</a>
Unweighted correlation network	custom package imsbInfer	Github ( <a href="https://github.com/wolski/imsbInfer">https://github.com/wolski/imsbInfer</a> )
WGCNA	(Langfelder and Horvath, 2008)	<a href="https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/">https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA/</a>
biomaRt	Bioconductor	<a href="http://bioconductor.org/packages/biomaRt/">http://bioconductor.org/packages/biomaRt/</a>
clusterProfiler	(Yu et al., 2012)	<a href="https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html">https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html</a>
Circos	(Krzywinski et al., 2009)	<a href="http://circos.ca/">http://circos.ca/</a>
Tracefinder	Thermo Scientific	Cat# OPTON-30626
GraphPad Prism 7	GraphPad	<a href="https://www.graphpad.com/">https://www.graphpad.com/</a>
Other		
Q Exactive mass spectrometer, Build 2.3 SP2	Thermo Scientific	<a href="https://www.thermofisher.com/order/catalog/product/IQLAEGAAPFALGMAZR">https://www.thermofisher.com/order/catalog/product/IQLAEGAAPFALGMAZR</a>

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Johan Auwerx ([admin.auwerx@epfl.ch](mailto:admin.auwerx@epfl.ch)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS****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
<b>Total fat</b>		<b>6.2</b>	<b>34.3</b>

## METHOD DETAILS

### 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 parameter: 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  $\mu$ L 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<sub>2</sub>O), (65:30:5, v/v/v, 100  $\mu$ L)) was used as a blank and for standard curves.

### NAD<sup>+</sup> Measurement of C57BL/6J Livers

~20 mg of frozen liver samples were used for NAD<sup>+</sup> extraction in 10% perchloric acid and neutralized in 3 M K<sub>2</sub>CO<sub>3</sub> 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<sub>2</sub>O, 80/20, v/v). Then, a linear gradient to 100% buffer B [H<sub>2</sub>O + 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<sup>+</sup> eluted as a sharp peak at 3.3 min and was quantified on the basis of the peak area ratio between NAD<sup>+</sup> 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<sub>2</sub>, 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  $\mu$ L) and flash frozen in liquid N<sub>2</sub> 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  $\mu$ L) and samples were vortexed (30 s). Chloroform/methanol (1:1, v/v, 1000  $\mu$ L) was added and samples were vortexed (60 s). Subsequently, HCl (1 M, 200  $\mu$ L) 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  $\mu$ L of the organic phase was dried under Ar<sub>2</sub>(g). The organic residue was reconstituted in ACN/IPA/H<sub>2</sub>O (65:30:5, v/v/v, 100  $\mu$ L) 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  $\times$  2.1 mm  $\times$  2.7  $\mu$ m particle size; Supelco) using an Accela LC Pump (500  $\mu$ L/min flow rate; Thermo). Mobile phase A consisted of 10 mM ammonium acetate in ACN/H<sub>2</sub>O (70:30, v/v) containing 250  $\mu$ L/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. 10  $\mu$ L 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<sup>1</sup> data was acquired in negative mode for 220-600 m/z at a resolving power of 17,500 and an AGC target of  $1 \times 10^5$ . 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  $2 \times 10^5$ , 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  $5 \times 10^5$ . 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 *imsblnfer*, currently on Github (<https://github.com/wolski/imsblnfer>). 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 (Arends et al., 2010) on the log<sub>2</sub> 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 ( $\pm 5$  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 IQTLs 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 IQTL candidate gene with the information on each filtering criteria is provided in Table S6.

### BXD IQTL and Human GWAS Genes Overlap

Human GWAS genes (having p-value  $< 1e-07$ ) were retrieved from the database GWASdb2 (<http://jwanglab.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 genesets 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 IQTL genes ( $\pm 5$  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 IQTLs.

### Enrichment Analysis of Proteins and Transcripts with Liver TAG Concentration

KEGG enrichment analysis was performed for all proteins and transcripts that correlated with total liver TAG 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 liver TAG 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](#). IQTL 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](#).