Genome-Wide Association Analysis Identifies Variants Associated with Nonalcoholic Fatty Liver Disease That Have Distinct Effects on Metabolic Traits

Elizabeth K. Speliotes1,2,3,4,9, Laura M. Yerges-Armstrong5, Jun Wu6, Ruben Hernandez2,8,9, Lauren J. Kim10, Cameron D. Palmer11, Vilmundur Gudnason12,13, Gudny Eiriksdottir12, Melissa E. Garcia10, Lenore J. Launer10, Michael A. Nalls14, Jeanne M. Clark7,8,15, Braxton D. Mitchell9, Alan R. Shuldiner5,16, Johann L. Butler1,2,11, Marta Tomas17,18, Udo Hoffmann19,20, Shih-Jen Huang21, Joseph M. Massaro21,22, Christopher J. O'Donnell20,21,23, Dushyant V. Sahani19, Veikko Salomaa24, Eric E. Schadt25, Stephen M. Schwartz26,27, David S. Siscovick26, NASH CRN1, GIANT Consortium1, MAGIC Investigators4, Benjamin F. Voight4, J. Jeffrey Carr28, Mary F. Feitosa6, Tamara B. Harris10, Caroline S. Fox21,23, Albert V. Smith12,9, W. H. Linda Kao7,15, Joel N. Hirschhorn4,11,29, Ingrid B.orecki6,9, GOLD Consortium

1 Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, United States of America, 2 Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America, 3 Division of Gastroenterology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 4 Broad Institute, Cambridge, Massachusetts, United States of America, 5 Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 6 Department of Internal Medicine, Washington University School of Medicine, Saint Louis, Missouri, United States of America, 7 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, 8 Division of General Internal Medicine, The Johns Hopkins Hospital, Baltimore, Maryland, United States of America, 9 Department of Internal Medicine, Washington University School of Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America, 10 Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 11 Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital, Boston, Massachusetts, United States of America, 12 Icelandic Heart Association, Kopavogur, Iceland, 13 University of Iceland, Reykjavik, Iceland, 14 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 15 Welch Center for Prevention, Epidemiology, and Clinical Research, Baltimore, Maryland, United States of America, 16 Geriatric Research and Education Clinical Center (GRECC), Veterans Administration Medical Center, Baltimore, Maryland, United States of America, 17 Cardiovascular Epidemiology and Genetics, Institut Municipal d'Investigació Médica, Barcelona, Spain, 18 CIBER Epidemiología y Salud Pública, Barcelona, Spain, 19 Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 20 Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 21 Framingham Heart Study, National Heart, Lung, and Blood Institute (NHLBI), Framingham, Massachusetts, United States of America, 22 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, 23 Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States of America, 24 Chronic Disease Epidemiology Unit, Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland, 25 Pacific Biosciences, Menlo Park, California, United States of America, 26 Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington, United States of America, 27 Department of Epidemiology, University of Washington, Seattle, Washington, United States of America, 28 Departments of Radiologic Sciences, Internal Medicine-Cardiology, and Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 29 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Nonalcoholic fatty liver disease (NAFLD) clusters in families, but the only known common genetic variants influencing risk are near PNPLA3. We sought to identify additional genetic variants influencing NAFLD using genome-wide association (GWA) analysis of computed tomography (CT) measured hepatic steatosis, a non-invasive measure of NAFLD, in large population based samples. Using variance component methods, we show that CT hepatic steatosis is heritable (≈26%–27%) in family-based Amish, Family Heart, and Framingham Heart Studies (n = 880 to 3,070). By carrying out a fixed-effects meta-analysis of genome-wide association (GWA) results between CT hepatic steatosis and ~2.4 million imputed or genotyped SNPs in 7,176 individuals from the Old Order Amish, Family Heart, and Framingham Heart Studies, we identify variants associated at genome-wide significant levels (p < 5 × 10−8) in or near PNPLA3, NCAN, and PPP1R3B. We genotype these and 42 other top CT hepatic steatosis-associated SNPs in 592 subjects with biopsy-proven NAFLD from the NASH Clinical Research Network (NASH CRN). In comparisons with 1,405 healthy controls from the Myocardial Genetics Consortium (MiGen), we observe significant associations with histologic NAFLD at variants in or near NCAN, GCKR, LYPAL1, and PNPLA3, but not PPP1R3B. Variants at these five loci exhibit distinct patterns of association with serum lipids, as well as glycemic and anthropometric traits. We identify common genetic variants influencing CT-assessed steatosis and risk of NAFLD. Hepatic steatosis associated variants are not uniformly associated with NASH/fibrosis or result in abnormalities in serum lipids or glycemic and anthropometric traits, suggesting genetic heterogeneity in the pathways influencing these traits.
Introduction

NAFLD includes a spectrum of disease ranging from fatty infiltration of the liver (steatosis) to histologic evidence of inflammation (nonalcoholic steatohepatitis or NASH), to fibrosis and cirrhosis, without a history of excessive alcohol ingestion [1, 2]. NAFLD can lead to liver failure and is accompanied by substantial morbidity and mortality, with few known effective treatments [3]. Obesity is a primary risk factor for NAFLD, but not all obese individuals are affected [4]. Familial clustering of the disease has been identified [5–7], suggesting that NAFLD may be influenced by genetic variants. However, thus far only one genetic locus has been found to reproducibly associate with magnetic resonance measured steatosis [8, 9].

Liver attenuation measured using computed tomography (CT) is a quantitative measure that is inversely related to the amount of fat in the liver [10–12]. It is highly correlated (r = 0.92) with the macrovesicular hepatic steatosis and thus is a non invasive measure of NAFLD [12]. The purpose of the present study was to determine the heritability of CT measured hepatic steatosis and to search for associated genetic variants in a meta-analysis of 7,176 individuals of European descent from the Framingham Heart Study (FAMHS), the Family Heart Study (FamHS), and the Age, Gene/Environment Susceptibility-Reykjavik study (AGES), which together comprise the GOLD (Genetics of Obesity-related Liver Disease) consortium (Table S1). To validate top associating variants for risk of histologically verified NAFLD, we utilized cases from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung, and Blood Institute’s Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix for genotyping services (Contract No. N02-HL-6-4278). The Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (grants U01DK061718, U01DK061728, U01DK061731, U01DK061732, U01DK061734, U01DK061737, U01DK061738, U01DK061739, U01DK061733). Several clinical centers use support from General Clinical Research Centers or Clinical and Translational Science Awards in conduct of NASH CRN Studies (grants UL1RR024949, M01RR000750, M01RR000750, U24RR001431, M01RR000750, U24RR02501401, M01RR000065, M01RR000065). The MigGen study was funded by the US National Institutes of Health (NIH) and National Heart, Lung, and Blood Institute’s STAMPEED genomics research project. MT was supported by the Beatrice de Pinós postdoctoral fellowship, Generalitat de Catalunya (2007BF-B100068, MT), and Red HERACLES, ISCIII (RD06/0009). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Eric E. Schadt is an employee of Pacific Biosciences.

* E-mail: espeliot@med.umich.edu (EKS); iborecki@wustl.edu (IBB)
† These authors contributed equally to this work.
‡ For membership information, please see Acknowledgments.

Results

We estimated the heritability of CT hepatic steatosis in three family-based cohorts. We found that the heritability of CT hepatic steatosis was 0.27 (standard error, SE 0.08), 0.27 (SE = 0.04), and 0.26 (SE = 0.04) in the Amish, FamHS, and FRAM cohorts respectively (n = 880–3,070) (See Materials and Methods and Table 1). These data suggest that CT hepatic steatosis, like other measures of fat has a genetic basis and that a search for influential genetic variants is warranted.

To identify specific genetic loci associated with CT hepatic steatosis, genome-wide association analyses were carried out in each of the four studies (See Materials and Methods and Tables S2, S3) and the results combined using a fixed effects meta-analysis (N = 7,176 in total). Variants at three loci emerged as being associated with CT hepatic steatosis at genome-wide significance levels (p < 5 × 10−8, Table 2; Figure 2A). These included rs738409 in PNPLA3 (p = 4.3 × 10−8), a locus previously reported as associated with magnetic resonance spectroscopy measured steatosis, [8] and two additional novel loci: rs4240624 near PPP1R3B (rs4240624, p = 3.68 × 10−10) and rs2228603 near APOC3 (rs2228603, p = 1.22 × 10−10). The alleles associated with increasing CT hepatic steatosis ranged in frequency from 0.07 to 0.92 and together account for 4.4% of the variance in hepatic steatosis (Table 2; range 0.79–2.41%). After removing these genome-wide significant loci, a quartile-quantile plot of the results demonstrated an excess of low p-values compared to expectations under the null (Figure 2B), suggesting that additional variants among those with moderately low p-values may also be associated with this trait. Except for variants near PNPLA3, we did not observe any variants in the region of any of the previously reported liver function test associated regions [18]. We could not assess whether the recently reported NAFLD associated variants near APOC3 [19] associate with CT hepatic steatosis as they were not genotyped on the
**Author Summary**

NAFLD is a spectrum of disease that ranges from steatosis to steatohepatitis (nonalcoholic steatohepatitis or NASH: inflammation around the fat) to fibrosis/cirrhosis. Hepatic steatosis can be measured non-invasively using computed tomography (CT) whereas NASH/fibrosis is assessed historically. The genetic underpinnings of NAFLD remain to be determined. Here we estimate that 26%-27% of the variation in CT measured hepatic steatosis is heritable or genetic. We identify three variants near PNPLA3, NCAN, and PPP1R3B that associate with CT hepatic steatosis and show that variants in or near NCAN, GCKR, LYPLAL1, and PNPLA3, but not PPP1R3B, associate with histologic lobular inflammation/fibrosis. Variants in or near NCAN, GCKR, and PPP1R3B associate with altered serum lipid levels, whereas those in or near LYPLAL1 and PNPLA3 do not. Variants near GCKR and PPP1R3B also affect glycemic traits. Thus, we show that NAFLD is genetically influenced and expand the number of common genetic variants that associate with this trait. Our findings suggest that development of hepatic steatosis, NASH/fibrosis, or abnormalities in metabolic traits are probably influenced by different metabolic pathways that may represent distinct therapeutic targets.

Affymetrix or Illumina platforms used by our studies and these variants do not have proxies that we could use in HapMap to impute them.

To determine whether SNPs with evidence of association with CT hepatic steatosis are also associated with histologic NAFLD, we genotyped 46 SNPs (independent SNPs with \( p < 5 \times 10^{-3} \), with independence defined as pairwise \( R^2 < 0.1 \); See Table S4 for SNP details in GOLD and each cohort) in 592 subjects with biopsy-proven NAFLD from the NASH CRN (See Table S1). Using ancestry-informative genetic markers [20], we had previously matched these cases to 1,405 healthy controls [21] from the ancestry-informative genetic markers [20], we had previously proven NAFLD from the NASH CRN (See Table S1). Using details in GOLD and each cohort) in 592 subjects with biopsy-tissue were then performed (Stage 3).

**Figure 1. Study design.** Meta-analysis of genome-wide association data was performed in Stage 1 across the cohorts shown. SNPs representing the best associating loci were genotyped in histology based NAFLD samples (Stage 2) from the NASH CRN matched to genome wide genotyped and imputed MGen controls. The effects of the five NAFLD associated SNPs on NASH CRN/iCONT, metabolic phenotypes and eQTLs in liver and adipose tissue were then performed (Stage 3).

doi:10.1371/journal.pgen.1001324.g001
the lowest p-values of association with CT hepatic steatosis at *LYPLAL1* and *PPR1R3B* lie downstream and upstream of the coding regions of these genes (Figure 3).

In epidemiologic studies NAFLD is associated with increased central obesity, higher low density lipoprotein (LDL)- cholesterol and lower high density lipoprotein (HDL)-cholesterol levels, impaired fasting glucose, increased risk of diabetes and increased insulin resistance. [24] In addition, variants in or near *GCKR*, *NCAN*, and *PPR1R3B* have been previously associated with lipid levels, *GCKR* with glycemic traits and *LYPLAL1* with abdominal obesity [16,25–29]. Therefore, we examined the associations of each of the CT hepatic steatosis-associated variants with serum LDL-cholesterol, HDL-cholesterol, triglycerides (TG), 2 hour glucose levels, 2 hour glucose levels controlled for body mass index (BMI), fasting glucose, homeostatic model for beta cell function (HOMA-B), homeostatic model for insulin resistance (HOMA-IR), fasting insulin, BMI, waist to hip ratio (WHR) controlled for BMI, and diabetes in the largest analyses of these traits available from the Global Lipids Genetics [13], GIANT [14], DIAGRAM [15], and MAGIC [16] Consortia (see Table 2, Table S9). Interestingly, we observed several distinct patterns of association. The allele associated with increasing CT hepatic steatosis at *NCAN* was associated with lower triglycerides and plasma LDL-cholesterol levels. By contrast, the hepatic steatosis-increasing allele at *GCKR* was associated with higher levels of plasma LDL-cholesterol and triglycerides, lower fasting glucose, lower fasting insulin, lower HOMA-IR, but increased 2 hour glucose, increased 2 hour glucose controlled for BMI, and WHR controlled for BMI. The hepatic steatosis-increasing allele at *PPR1R3B* was associated with increased HDL- and LDL-cholesterol levels and decreased fasting glucose. (Table 2, Figure 4). The variants near *PNPLA3* and *LYPLAL1* were not associated with any of the traits tested (See Table 2, Table S9 and Figure 4).

For *PNPLA3* (rs738408), *NCAN* (rs2228603), and *GCKR* (rs780094) the variants with the lowest p-values of association with CT hepatic steatosis are either themselves missense SNPs or in high LD with missense SNPs. Thus, the most parsimonious model of how they may act is by directly affecting protein structure or function. However, the variants with the lowest p-values of association with CT hepatic steatosis near *LYPLAL1* and *PPR1R3B* fall in non-coding regions and thus for these (as well as the other three loci above) we tested whether they have effects on the expression of nearby genes in liver and adipose tissue from a sample of bariatric surgery patients [17] (See Table S10). We found that the hepatic steatosis increasing variant (rs4290624) at the *PPR1R3B* locus increased liver mRNA expression of *PPR1R3B* and *AW673036_RG* and decreased expression of *AK055863*. The hepatic steatosis increasing variant (rs780094) at the *GCKR* locus increased expression of C2orf16 mRNA in liver. In these cases the eQTL with the lowest p-value of affecting these transcripts in the region was the same or highly correlated with the allele that had the lowest p-value of association with CT hepatic steatosis consistent with the possibility that these SNPs may function by affecting expression of nearby genes. For all other cases, the eQTL with the lowest p-value of affecting transcript expression at the locus was not eliminated by controlling for the variant that had the lowest p-value of association with CT hepatic steatosis and thus in these cases, the data do not support an expression effect as mediating the association with steatosis. Because alteration of *PPR1R3B* expression has been shown to affect serum lipid levels [13] one possibility is that changes in expression of this gene could mediate its effect on hepatic steatosis. For *GCKR*, the variant with the lowest p-value of association with CT hepatic steatosis is in high LD with a missense variant in *GCKR* which has been shown to affect *GCKR* function [30]. Thus, at *GCKR* an alternate model of action of how the CT hepatic steatosis associated variant affects hepatic steatosis is via altering *GCKR* function rather than via altering expression of *C2orf16*. Further functional work will be needed to prove that these variants exert their effects on hepatic steatosis via these possible mechanisms.

### Discussion

We have identified variants in three novel loci (*NCAN*, *GCKR*, and *LYPLAL1*) and one previously reported locus (*PNPLA3*) that are associated with both increasing CT hepatic steatosis and histologic NAFLD. *PPR1R3B* is associated with CT steatosis but not histologic NAFLD that includes individuals mostly with inflammation and fibrosis. These variants all have distinct patterns of effects on NAFLD and metabolic traits.

We have shown that CT hepatic steatosis is heritable and that GWA meta-analysis led to the identification of variants associated not only with CT hepatic steatosis but, also, with more severe NASH/fibrosis mostly present in the NASH CRN sample. Because CT hepatic steatosis measurements can be obtained noninvasively, much larger sample sizes can be accumulated, thereby increasing power to identify variants that associate with NAFLD compared with only studying individuals that have histology diagnosed disease. Follow-up association testing in samples with histologic phenotypes remains useful however. We did observe one variant near *PPR1R3B* that was associated with CT-assessed liver attenuation but not histology-proven NAFLD. Possible reasons for why the variant near *PPR1R3B* is associated with CT liver steatosis but not histology-proven NAFLD include 1. It influences steatosis only, not progression to NASH/fibrosis: 2. its association with CT fat may be a false positive: 3. the NASH CRN/MIGen sample is underpowered to see an effect on histologic NAFLD: or 4. the variant is associated with something other than fat reflected in the CT scan (eg. glycogen content). Further work is needed to differentiate among these possibilities.
Table 2. Genome-wide significant or replicating variants from GOLD, NASH CRN/MiGen, and metabolic phenotype analyses.

<table>
<thead>
<tr>
<th>Locus</th>
<th>GOLD Meta-analysis</th>
<th>NASH CRN/MiGen analysis</th>
<th>Metabolic Phenotypes Meta Analyses **</th>
<th>Dir.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP ID</td>
<td>Chr.</td>
<td>Pos.</td>
<td>Nearest Gene</td>
<td>EA</td>
</tr>
<tr>
<td>rs738409</td>
<td>22</td>
<td>42656060</td>
<td>PNPLA3</td>
<td>G</td>
</tr>
<tr>
<td>rs228603</td>
<td>19</td>
<td>19190924</td>
<td>NCAN</td>
<td>T</td>
</tr>
<tr>
<td>rs12137855</td>
<td>1</td>
<td>217515001</td>
<td>LYPLAL1</td>
<td>C</td>
</tr>
<tr>
<td>rs780949</td>
<td>8</td>
<td>27594741</td>
<td>GCKR</td>
<td>T</td>
</tr>
<tr>
<td>rs4240624</td>
<td>7</td>
<td>9221641</td>
<td>PPP1R3B</td>
<td>A</td>
</tr>
</tbody>
</table>

GOLD: Genetics of Obesity-related Liver Disease; NASH CRN: Nonalcoholic Steatohepatitis Clinical Research Network; MiGen: Myocardial Infarction Genetics Consortium; **from The Global Lipids Genetics, GIANT, MAGIC, and DIAGRAM Consortia; Chr. Chromosome; Pos.: position, build 35; Mb: Megabase; EA: effect allele; EAF: Effect allele frequency; Effect: increase in inverse normalized fatty liver by computed tomography SE: Standard Error; % Var: percentage of variance explained; GOLD P: p-value of association in GOLD; EAFa: Frequency of the effect allele in cases from the NASH CRN study; EAFb: Frequency of the effect allele in controls from the MiGen study; ORNAFLD: odds ratio for the presence of NAFLD on pathology per effect allele; NAFLD P: False discovery rate p-value of association for histologic NAFLD; LDL: low density lipoprotein cholesterol; HDL: high density lipoprotein cholesterol; TG: triglycerides; Glucose: fasting glucose; HOMA-IR: homeostatic model assessment of insulin resistance; P: p-value of association; Dir: direction of effect allele for significant associations in GOLD, NASH CRN/MiGen, LDL, HDL, TG, glucose, HOMA-IR analyses respectively; +/− represents increasing/decreasing fatty liver in GOLD, and having a higher/lower odds of having NAFLD in the NASH CRN/MiGen analyses and higher/lower LDL, HDL, TG, glucose, HOMA-IR respectively; N represents no significant effect; PNPLA3: patatin-like phospholipase domain-containing protein 3 (HUGO Gene Nomenclature Committee, HGNC: 18590); NCAN: neurocan (HGNC: 2465); LYPLAL1: lysophospholipase-like 1 (HGNC: 20440); GCKR: glucokinase regulatory protein (HGNC: 4196); PPP1R3B: protein phosphatase 1, regulatory subunit 3b (HGNC: 14942). doi:10.1371/journal.pgen.1001324.t002
We show that some of the variants that are associated with increased CT hepatic steatosis have distinct patterns of effects on metabolic traits that, when taken together, give us insight into their functional clustering. For example, unlike the other three loci, variants in or near PNPLA3 and LYPLAL1 do not affect any of the other metabolic traits and interestingly PNPLA3 and LYPLAL1-related proteins have been predicted to play a role in consecutive steps in triglyceride breakdown [31,32]. Thus these could increase hepatic steatosis by preventing breakdown of triglycerides, as recently shown for PNPLA3 (I148M) [33]. The apparent discordance between the strong effect on hepatic steatosis and modest, if any, effect on serum lipid levels suggests that these genes, if they are involved in lipid metabolism, exert their effects within the liver in ways that are not well reflected in serum measurements. Thus, similarities in the pattern of pleiotropic effects on other traits may provide insights into the functional clustering of the genes that these variants effect.

Unlike PNPLA3 and LYPLAL1, variants near NCAN (which encodes for an adhesion molecule [34]), PPP1R3B (which encodes for a protein that regulates glycogen breakdown [35]), and GCKR (which, through inhibition of glucokinase, regulates glucose storage/disposal and provides substrates for de novo lipogenesis [30]), are associated with distinct changes in serum and liver lipids as well as glycemic traits. Indeed, these data may provide new insights into how obesity can lead to metabolic complications in some but not all individuals–some but not all of these individuals carry variants that predispose them both to liver fat deposition and to metabolic dysregulation. Further, our data show that the alleles of SNPs that associate with increased liver steatosis are also associated with a diverse pattern of metabolic phenotypes including different combinations of increased or decreased serum LDL-cholesterol, increased serum HDL-cholesterol, increased serum TG, decreased serum fasting glucose and insulin, decreased insulin resistance, and increased WHR adjusted for BMI. In addition, some hepatic steatosis-associated variants are not strongly associated with any of these metabolic traits (PNPLA3 and LYPLAL1). These results indicate that hepatic steatosis is likely to be influenced by different metabolic pathways, based on these various patterns of association. Thus it may be possible to resolve genetic heterogeneity in the etiology of hepatic steatosis, which may present unique opportunities for personalized therapies.

Compared with earlier efforts, this study is well-powered, using more than 7,176 individuals for discovery of variants that affect NAFLD. Thus, noninvasive measures of hepatic steatosis such as
Five GWAS Loci Associate with NAFLD

PNPLA3 Chr 22

NCAN Chr 19

LYPLAL1 Chr 1

GCKR Chr 2

PPP1R3B Chr 8
CT scanning can provide valuable information for use in population- and family-based studies aimed at identifying genetic risk factors for NAFLD. Although the identities of nearby genes and effects on lipid levels provide important clues, functional studies will be needed to further understand the mechanisms by which these risk factors influence the development and progression of NAFLD. Overall however, our work gives us new insights into the biology and genetics of NAFLD and opens up avenues for biological, diagnostic, and therapeutic research for this condition in humans.

Materials and Methods

Ethics statement

All work done in this paper was approved by local institutional review boards or equivalent committees.

GOLD studies and genetic analyses

Each of the participating studies had the overarching objective of investigating cardiovascular disease and its risk factors. The studies are population based and 3 of the 4 are family studies. Genome-wide SNP data were available in each case, and the platforms and quality control measures are described in Tables S2 and S3.

Age Gene-Environment Susceptibility—Reykjavik Study (AGES-Reykjavik). The AGES-Reykjavik Study is a single center prospective population-based cohort nested in the original Reykjavik Study, a cohort of 30,795 randomly sampled persons living in Reykjavik, Iceland. The cohort included 19,381 men and women born between 1907 and 1935. Re-examination of a sample of surviving members of the Reykjavik Study was initiated in 2002 as the AGES-Reykjavik Study. This study included imaging by computed tomography, from which liver attenuation was measured from a 1mm thick slice at the level of the L1/L2 vertebrae by calculating the average Hounsfield Unit in a region of interest with a diameter of 1 cm located 10% of the distance from where a tangent from the mid-anterior of the spinal canal bisected a line between the second and third rib. Four thousand seven hundred and seventy two individuals were assessed for hepatic steatosis using CT scanning. Liver attenuation controlled for an external phantom was inverse normally transformed and residuals created from a linear regression model in Proable [36]/R with covariates of age, age2, gender and drinks along with the SNPs in an additive genetic model (See Tables S2, S3).

Figure 4. Effects on traits.

<table>
<thead>
<tr>
<th>Liver Fat</th>
<th>Serum Lipid Traits</th>
<th>Glycemic Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT</td>
<td>Histology</td>
<td></td>
</tr>
<tr>
<td>PNPLA3</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>NCAN</td>
<td>↑</td>
<td>↓</td>
</tr>
<tr>
<td>LYPAL1</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>GCKR</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>PPP1R3B</td>
<td>↑</td>
<td>↑</td>
</tr>
</tbody>
</table>

Figure 3. Regional plots of genome-wide significant or replicating loci of association in GOLD. SNPs are plotted by position on chromosome against association with CT hepatic steatosis (–log10 p-value). The figures highlight the SNP taken into Stage 2 (diamond). The SNPs surrounding the most significant SNP are color-coded to reflect their LD with this SNP as in the inset (taken from pairwise R2 values from the HapMap CEU database, www.hapmap.org). Estimated recombination rates (from HapMap) are plotted in cyan to reflect the local LD structure. Genes and the direction of transcription, are noted below the plots (data from UCSC genome browser, genome.ucsc.edu). Coding SNPs in high LD with the best SNP are noted with rs number and protein change.

doi:10.1371/journal.pgen.1001324.g003

doi:10.1371/journal.pgen.1001324.g004
The Amish. Subjects were identified from 2 studies of cardiovascular health in the Old Order Amish community in Lancaster County: the Amish Family Calcification Study (2001–2006) [37] and the Amish Longevity Study (2000–2006) [38]. In total, 541 individuals had both genome-wide SNP data and CT- assessed hepatic steatosis. Thoracic electron-beam computerized tomography (EBCT) scans were obtained as part of the Amish Family Calcification Study by an Imatron C-150 EBCT scanner. Measurements from two regions-of-interest, the liver and spleen, were obtained. The spleen measurements were used as an attenuation standard. Accu View (Accumor Image Corp.) software was used to calculate the attenuation coefficient in Hounsfield Units for each region-of-interest. Two 1.0-cm² region-of-interest measurements were obtained from the liver and one was obtained from the spleen. The average of the liver attenuation measurements divided by the spleen attenuation measurement was then calculated. The region-of-interest measurements were placed in such a manner that minimized measurements from vessels, focal lesions, areas of artifact or near the edges of the organs.

The liver attenuation/spleen attenuation ratio was inverse normally transformed and association was tested with genotypes in lesions, areas of artifact or near the edges of the organs. Then calculated. The region-of-interest measurements were placed in such a manner that minimized measurements from vessels, focal lesions, areas of artifact or near the edges of the organs.

The liver attenuation was regressed on age, age², field center, phantom average, alcohol consumption and 10 genetic principal components, by sex, using a stepwise procedure and retaining terms significant at the 5% level. We then applied an inverse normal rank transformation to the adjusted phenotype within sex strata and association was assessed assuming an additive model using PROC MIXED in SAS to account for the siblings.

The Framingham Heart Study. The Framingham Heart Study recruited 5,209 residents in 1948 from the population in Framingham, Massachusetts [41]. These individuals have had serial examinations and collection of respective data since. In 1971, 5,124 offspring from the original residents and their spouses were recruited into the Offspring Study and have been followed for four to eight years since [42]. In 2002, 4,095 third generation members and their spouses were enrolled [43].

Between 2002 and 2005, 1,400 individuals from the Offspring Study and 2,011 individuals from third generation underwent multidetector computed tomograms on which we evaluated liver attenuation as previously described [44]. Inclusion criteria favored individuals who lived in the New England area and included 755 families. Minimum age was 35 in men and 40 in women. Women of childbearing age were screened and pregnant women and individuals >160 kilograms were excluded from scanning. Individuals with scans that could not be interpreted for hepatic steatosis or did not attend offspring examination 7 as they lacked covariate data were not used for analysis. The average of the liver attenuation measures and a high density external calcium control were used to create a liver/phantom ratio to control for scan penetrance.

For GWAS analysis, inverse normally transformed liver attenuation/phantom ratio was used in a mixed linear model (controlling for relatedness) in R [45] with covariates of age, age squared, gender, and alcoholic drinks (4 oz = 1 drink) with the first ten principal components (as determined in Eigenstrat [46]) as covariates. Principal components were first generated using an unrelated sample of 718 and then projected to the rest of the cohort. Individuals who deviated from the mean of the principal components of more than six standard deviations were removed prior to analysis (n = 1).

Heritability analyses. Three of the four studies participating in this consortium were family studies and the family structure characteristics used for heritability are shown in Table 1. Liver attenuation adjusted for scan penetrance and then inverse normally transformed and corrected for age gender and number of alcoholic drinks (drinks in FamHS, and FRAM only as the Amish do not drink) was estimated in each of the studies and then heritability assessed using a variance components method as implemented in the software SOLAR [47]. Despite the diverse character of these family studies, there was remarkable consistency in the estimates of the proportion of variance due to genetic effects, and the magnitude of the heritabilities is comparable to many complex quantitative traits and suggests that a search for underlying genetic variants is warranted.

Meta-analysis and GWAS. Association data from the four studies above were filtered for SNPs that had a minor allele frequency >1% and for SNPs that had an imputation quality score of >0.3. All files were GC corrected after filtering and before meta-analysis. The inflation factor for the AGES study was 1.01, for the Amish was 1.05, for the Family Heart Study was 1.03, and for the Framingham Heart Study was 1.02. Meta-analysis was
conducted using a fixed effects model with a beta and standard error as implemented in METAL (http://www.sph.umich.edu/csg/abecasis/metal/). After meta-analysis, SNPs present in fewer than 3 studies were eliminated from analysis. The inflation factor for the overall meta-analysis was 1.03. The meta-analysis was GC corrected before the final p values were reported. The variation in CT hepatic steatosis explained by the tested SNPs was estimated from stage 2 analyses using $2f(1−f)\hat{a}^2$, where f is the frequency of the variant and a is its additive effect in units of standard deviations from the meta analysis [40].

Selection of SNPs for validation/replication with histologic NAFLD

To define independently associated SNPs, the LD was required to be $R^2 < 0.10$ and the SNPs located at least 1 megabase from each other. From among these, the SNP with the strongest association was chosen for follow up ($P < 0.0001$). Two iPlex pools consisting of 46 SNPs were designed and were successfully genotyped in the NASH CRN samples. Of these, only 45 were imputed well in MIGen, and only these SNPs were analyzed. Variants with a false discovery rate of $q < 0.05$ were considered associated with NAFLD.

NASH CRN samples

Study: The NASH CRN samples were collected from eight different centers in the U.S. as previously described [2,49]. Adults from both the Database and the PIVENS trial (Pioglitazone versus Vitamin E versus Placebo for the Treatment of Nondiabetic Patients with Nonalcoholic Steatohepatitis) were used for analysis. Briefly, individuals from the Database were part of an observational study of nonalcoholic fatty liver disease. Inclusion criteria included age 18+, histologic diagnosis for NAFLD, or histologic diagnosis for cryptogenic cirrhosis or suspected NAFLD on the basis imaging studies suggestive of NAFLD, or clinical evidence of cryptogenic cirrhosis. No subjects reported regular excessive use of alcohol within two years prior to the initial screening period. Exclusion criteria included histologic evidence of liver disease besides nonalcoholic liver disease, known HIV positivity, and conditions that would interfere with study follow up. Individuals in the PIVENS database were part of a multicenter placebo controlled study with three parallel groups examining the effects of pioglitazone vs. vitamin E vs. placebo on NAFLD. Inclusion and exclusion criteria were as described previously [2,49]. For this analysis, we excluded individuals who did not describe their race as being white and non-Hispanic. There were 678 adults who matched these criteria. Finally, individuals without histology available for central review were excluded, leaving 592 adults for the current study.

Histology determination in NASH CRN

Histologic diagnoses were determined in the NASH CRN by central review by NASH CRN hepatopathologists using previously published criteria [2,49]. Predominantly macrovesicular steatosis was scored from grade 0–3. Inflammation was graded from 0–5 and cytolitic ballooning from 0–2. The fibrosis stage was assessed from a Masson trichrome stain and classified from 0–4 according to the NASH CRN criteria. Individuals could contribute to more than one of these outcomes. The NASH CRN samples were genotyped and analyzed as described in Tables S2 and S3.

Analysis in NASH CRN/MIGen samples

MIGen controls were matched to the NASH CRN samples for genetic background. As previously described, the MIGen samples were collected from various centers in the U.S and Europe by the Myocardial Infarction Genetics Consortium (MIGen) [22] as controls for individuals with early onset MI. The genetic ancestry the MIGen samples was explored by using the program Eigenstrat [46]; the first principal component was the most significant and correlated with the commonly observed Northwest-Southeast axis within Europe [20] and genetic ancestry along this principal component is correlated with reported country of origin in the MIGen sample [22]. From this analysis, 120 unlinked SNPs were chosen from the MIGen genotype data that were most strongly correlated with the first principal component. These SNPs were genotyped in the NASH CRN samples to enable matching of MIGen controls to the NASH CRN [20] cases for genetic background. PLINK [50] was used to match individuals based on identity by state (IBS) distance using a pairwise population concordance test statistic of $> 1 \times 10^{-3}$ for matching. The SNPs selected for validation were tested in this case-control sample using logistic regression controlling for age, age$^2$, gender, and the first 5 principal components as covariates in PLINK [50]. We report the p-values, odds ratios and confidence intervals.

iCONT samples

We obtained 3,294 population based control samples with genotypes from Illumina (see http://www.illumina.com/science/icontolkb.illumina). These individuals were used as controls in various case control analyses. Individuals were removed as described in Table S4 and 3,212 individuals were then used as controls for the NASH CRN/iCONT analyses.

Analysis in NASH CRN/iCONT samples

The 592 individuals from the NASH CRN described above were used as cases and 3,212 individuals from the iCONT database were used as controls. Genome wide significant or replicating SNPs were tested in this case-control sample using logistic regression controlling for gender in PLINK [50]. We report the p-values, odds ratios and confidence intervals.

Concordance analysis of imputed SNPs in MIGen and iCONT with the HapMap3 TSI sample

To assess the concordance of imputed SNPs in the MIGen and iCONT samples we obtained the genotyped SNPs from the HapMap3 TSI (Tuscans from Italy) sample. Using only the SNPs present on the Affymetrix 6.0 platform (used to genotype MIGen) or only the SNPs present on the Illumina platform (used to genotype iCONT samples) and the LD information from HapMap2 we imputed the remainder of the SNPs using MACH[1,0,16] and compared the imputed calls to the actual genotypes stratified by imputation quality score (R2 hat).

Evaluation of effects on other metabolic traits

To obtain data on whether CT hepatic steatosis SNPs affect other metabolic traits we obtained data from four consortia that had the largest and most powered analyses of these traits. Association results for HDL-, LDL- cholesterol levels and triglycerides (TG) were obtained from publicly available data of the GLOBAL Lipids Genetics Consortium (http://www.sph.umich.edu/csg/abecasis/public/Teslovich et al. 2010) [13]. Association results for fasting insulin, glucose, 2 hr-glucose, HOMA-IR and HOMA-B were obtained from the MAGIC Investigators. Association results for risk of type 2 diabetes were obtained from the DIAGRAM consortium [15]. Association results for risk of BMI and waist to hip ratio controlled for BMI were obtained from the GIANT consortium [14]. We used a conservative nominal p<0.0008 corresponding to
a bonferroni correction of 12 phenotypes tested for 5 SNPs to determine significance.

**Expression QTL analyses**

The expression QTL analyses in liver, subcutaneous and omental fat tissue have been described in detail previously [17]. Tissue were obtained from patients who underwent bariatric surgery, and RNA expression assessed using a custom Agilent 44,000 feature microarray composed of 39,280 oligonucleotide probes targeting transcripts representing 34,266 known and predicted genes. Patients were also genotyped on the Illumina 650Y SNP genotyping arrays. SNPs were tested for cis-associations with transcripts within a 1 Mb region, assuming an additive effect of the CT hepatic steatosis increasing allele adjusting for age, race, gender, and surgery year using linear regression. Cis-associations between each SNP and the adjusted gene expression data were tested, and only associations with a nominal p-value <3.5×10^-5 corresponding to a bonferroni correction for 284 gene transcripts x 5 SNPs tested are shown in Table S10. Conditional analyses were performed by conditioning the CT hepatic steatosis associated SNP on the most significant cis- associated SNP for that particular gene transcript and vice versa.

**Supporting Information**

**Table S1** Study sample characteristics. (*) Drinking alcohol is not practiced in Amish culture and not measured this study. GOLD: Genetics of Obesity-related Liver Disease; NASH CRN: Nonalcoholic Steatohepatitis Clinical Research Network; MIGen: Myocardial Infarction Genetics Consortium; iCONT: Illumina Control Database; SD: standard deviation; P25, P75: 25th and 75th percentiles; Phantom LD or HD- low or high density external hydroxyapatite CT control; Median raw liver measures in 75th percentiles; Phantom LD or HD- low or high density external control database; IBD pi hat: value for identical by descent of the TSI given imputation quality score in iCONT. Found at: doi:10.1371/journal.pgen.1001324.s001 (0.05 MB DOC)

**Table S2** Genotyping and association information. Imp’n: Imputation; MAF: minor allele frequency; HWE: Hardy Weinberg Equilibrium; GOLD: Genetics of Obesity-related Liver Disease; NASH CRN: Nonalcoholic Steatohepatitis Clinical Research Network; MIGen: Myocardial Infarction Genetics Consortium; iCONT: Illumina Control database. Found at: doi:10.1371/journal.pgen.1001324.s002 (0.05 MB DOC)

**Table S3** Quality control. * Sample genotyping success rate; i.e. percentage of successfully genotyped SNPs per sample. GOLD: Genetics of Obesity-related Liver Disease; NASH CRN: Nonalcoholic Steatohepatitis Clinical Research Network; MIGen: Myocardial Infarction Genetics Consortium; iCONT: Illumina Control database; IBD pi hat: value for identical by descent of >0.15. Found at: doi:10.1371/journal.pgen.1001324.s003 (0.05 MB DOC)

**Table S4** Top genotyped hits from GOLD, AGES, AMISH, Family Heart Study, Framingham Heart Study, GOLD: Genetics of Obesity-related Liver Disease; Chr.: Chromosome; Pos.: position, build 35; EA: effect allele; OA: other allele; EAF: Frequency of the effect allele in the analyses [weighted average in GOLD]; Effect: increase in inverse normalized fatty liver by computed tomography SE: Standard Error; P: p-value of association in the analyses; % Var: % variance explained; P het: p-value for heterogeneity across studies; N: number of individuals in the analyses. Found at: doi:10.1371/journal.pgen.1001324.s004 (0.41 MB DOC)

**Table S5** Top genotyped hits in NASH CRN/MIGen analysis. NASH CRN: Nonalcoholic Steatohepatitis Clinical Research Network; MIGen: Myocardial Infarction Genetics Consortium; Chr. Chromosome; Pos.: position, build 35; EA: effect allele; OA: other allele; EAFa: Frequency of the effect allele in cases from the NASH CRN study; EAFb: Frequency of the effect allele in controls from the MIGen study; Impb: Imputation quality score in MIGen; NAFLD: nonalcoholic fatty liver disease; OR NAFLD: odds ratio for the presence of NAFLD on pathology per effect allele; P NAFLD: False discovery rate p-value of association for histologic NAFLD. Found at: doi:10.1371/journal.pgen.1001324.s005 (0.10 MB DOC)

**Table S6** Genome-wide significant or replicating variants in NASH CRN/iCONT analysis. NASH CRN: Nonalcoholic Steatohepatitis Clinical Research Network; iCONT: Illumina Control database; EA: effect allele; OA: other allele; EAF: Frequency of the effect allele in cases from the NASH-CRN study; EAFb: Frequency of the effect allele in controls from iCONT; Impb: Imputation quality score in iCONT; NAFLD: nonalcoholic fatty liver disease; OR NAFLD: odds ratio for the presence of NAFLD on pathology per effect allele; P NAFLD: False discovery rate p-value of association for histologic NAFLD. Found at: doi:10.1371/journal.pgen.1001324.s006 (0.03 MB DOC)

**Table S7** Imputation R2 hat measures versus concordance to real genotypes in TSI individuals from HapMap 3. TSI: Toscans in Italy; R2 hat: Imputation quality score from MACH; N SNPs: number of SNPs used for concordance analysis; concordance: average concordance amongst the SNPs assayed. Found at: doi:10.1371/journal.pgen.1001324.s007 (0.01 MB DOCX)

**Table S8** Imputation R2 hat measures in MIGen and iCONT versus concordance to real genotypes in TSI individuals from HapMap 3. Imp: Imputation quality score in MIGen; Concordance: average concordance of SNPs in TSI given imputation quality score in MIGen; Impb: Imputation quality score in iCONT; Concordanceb: average concordance of SNPs in TSI given imputation quality score in iCONT. Found at: doi:10.1371/journal.pgen.1001324.s008 (0.01 MB DOCX)

**Table S9** Effect of genome-wide significant or replicating variants on glucose, anthropometric and lipid traits. Association results for high density lipoprotein (HDL),- low density lipoprotein (LDL)-, Cholesterol levels and triglycerides (TG) were obtained from publicly available data of the GLOBAL Lipids Genetics Consortium (http://www.sph.umich.edu/csg/abecasis/public/Teslovich et al. 2010) Association results for fasting Insulin and glucose, 2hr-glucose, HOMA-IR and HOMA-B were obtained from the MAGIC Consortium (Dupuis et al. Nature Genetics 2010). Association results for risk of type 2 diabetes were obtained from the DIAGRAM consortium (Voight et al. Nature Genetics 2010). Association results for risk of BMI and waist to hip ratio controlled for BMI were obtained from the GIANT consortium (Speliotes et al. Nature Genetics 2010). BMI: body mass index; HOMA-IR: homeostasis model assessment insulin resistance; HOMA-B: homeostasis model assessment beta cell function; EA:
effect allele; OA: other allele; Effect: The change in the trait per effect allele from the various studies; SE: standard error in the effect from the various studies; P: p-value of association from the various studies; N: number of individuals in the analyses; OR: odds ratio for the effect allele on diabetes; U95% and L95%: upper and lower 95% confidence levels for the OR.

Table S1

Significant associations between genome-wide significant or replicating SNPs and cis gene expression (cis-eQTLs) in liver, omental fat and subcutaneous fat. SNP: the fatty liver associating SNP from GWAS analysis. EA: effect allele (fatty liver increasing allele from GWAS). Effecta: Direction of effect on the gene transcription expression level for the effect allele. P: p-value of association of the fatty liver SNP with change in gene expression. Padj p-value for the fatty liver SNP after conditioning on the most significant SNP for change in gene transcript. Peak SNPc: SNP in the region that has the most significant eQTL p-value on expression of the gene transcript Rsqd: the R squared correlation coefficient of the peak SNP after conditioning on the fatty liver SNP for change in gene expression. Peak SNP: the fatty liver SNP in the region that has the most significant SNP for change in gene transcript. Peak SNPb: the fatty liver SNP after conditioning on the fatty liver SNP for change in gene transcript. NA: peak SNP is the same as the fatty liver associating SNP.

Table S2

<table>
<thead>
<tr>
<th>SNP</th>
<th>Effecta</th>
<th>P</th>
<th>N</th>
<th>OR</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP1</td>
<td>EA</td>
<td>0.01</td>
<td>1000</td>
<td>1.2</td>
<td>Change</td>
</tr>
<tr>
<td>SNP2</td>
<td>OA</td>
<td>0.05</td>
<td>2000</td>
<td>0.8</td>
<td>No change</td>
</tr>
<tr>
<td>SNP3</td>
<td>EA</td>
<td>0.02</td>
<td>1500</td>
<td>1.1</td>
<td>Small change</td>
</tr>
</tbody>
</table>

Acknowledgments

We would like to thank Dr. Qiong Yang for use of her R program for related analyses in The Framingham Heart Study and Dr. Monty Krieger for critically reviewing the manuscript. We would like to thank Arun Sanyal for serving as our liaison to the NASH CRN. Samples were provided by the NASH CRN for analyses in this paper.
References


Five GWAS Loci Associate with NAFLD


26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) Five GWAS Loci Associate with NAFLD