

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

Elizabeth K. Speliotes<sup>1,2,3,4,\*</sup>, Laura M. Yerges-Armstrong<sup>5</sup>, Jun Wu<sup>6</sup>, Ruben Hernaez<sup>7,8,9</sup>, Lauren J. Kim<sup>10</sup>, Cameron D. Palmer<sup>11</sup>, Vilmundur Gudnason<sup>12,13</sup>, Gudny Eiriksdottir<sup>12</sup>, Melissa E. Garcia<sup>10</sup>, Lenore J. Launer<sup>10</sup>, Michael A. Nalls<sup>14</sup>, Jeanne M. Clark<sup>7,8,15</sup>, Braxton D. Mitchell<sup>5</sup>, Alan R. Shuldiner<sup>5,16</sup>, Johannah L. Butler<sup>1,2,11</sup>, Marta Tomas<sup>17,18</sup>, Udo Hoffmann<sup>19,20</sup>, Shih-Jen Hwang<sup>21</sup>, Joseph M. Massaro<sup>21,22</sup>, Christopher J. O'Donnell<sup>20,21,23</sup>, Dushyant V. Sahani<sup>19</sup>, Veikko Salomaa<sup>24</sup>, Eric E. Schadt<sup>25</sup>, Stephen M. Schwartz<sup>26,27</sup>, David S. Siscovick<sup>26</sup>, NASH CRN<sup>†</sup>, GIANT Consortium<sup>‡</sup>, MAGIC Investigators<sup>§</sup>, Benjamin F. Voight<sup>4</sup>, J. Jeffrey Carr<sup>28</sup>, Mary F. Feitosa<sup>6</sup>, Tamara B. Harris<sup>10</sup>, Caroline S. Fox<sup>21,23</sup>, Albert V. Smith<sup>12</sup>, W. H. Linda Kao<sup>7,15</sup>, Joel N. Hirschhorn<sup>4,11,29</sup>, Ingrid B. Borecki<sup>6,\*</sup>, GOLD Consortium<sup>¶</sup>

**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 Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, United States of America, **6** Division of Statistical Genomics, Department of Genetics, Washington University, 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 Hospital Center, Washington D.C., 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 Mèdica, 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 components 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, Age, Gene/Environment Susceptibility-Reykjavik study (AGES), Family Heart, and Framingham Heart Studies, we identify variants associated at genome-wide significant levels ( $p < 5 \times 10^{-8}$ ) 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*, *LYPLAL1*, 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.

**Citation:** Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, et al. (2011) Genome-Wide Association Analysis Identifies Variants Associated with Nonalcoholic Fatty Liver Disease That Have Distinct Effects on Metabolic Traits. *PLoS Genet* 7(3): e1001324. doi:10.1371/journal.pgen.1001324

**Editor:** Mark I. McCarthy, University of Oxford, United Kingdom

**Received:** June 18, 2010; **Accepted:** February 2, 2011; **Published:** March 10, 2011

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

**Funding:** This work was partially supported by NIH grants T32 DK07191-32 to Daniel Podolsky (for EKS), F32 DK079466-01 to EKS, NIH K23DK080145-01 to EKS, and R01DK075787 to JNH. The AGES-Reykjavik Study is funded by NIH contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The Old Order Amish Study was supported by NIH research grants K01 DK067207, R01 AG18728, R01HL088119, U01 HL72515, and U01 HL084756. Partial funding was also provided by the Diabetes Research and Training Center of Maryland (P60 DK079637) and the Nutrition and Obesity Research Center of Maryland (P30DK072488). LMY-A was supported by NIH training grants T32AG000262 and F32AR059469. RH was supported by the American Diabetes Association Mentor-Based Postdoctoral Fellowship Program (7-07-MN-08). The NHLBI Family Heart Study was supported by NIDDK R01DK075681 (to IBB) and NHLBI R01HL087700. This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development 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, U01DK061730, U01DK061713). Several clinical centers use support from General Clinical Research Centers or Clinical and Translational Science Awards in conduct of NASH CRN Studies (grants UL1RR024989, M01RR000750, M01RR00188, ULRR02413101, M01RR000827, UL1RR02501401, M01RR000065, M01RR020359). The MIGen study was funded by the US National Institutes of Health (NIH) and National Heart, Lung, and Blood Institute's STAMPEED genomics research program. MT was supported by the Beatriz de Pinós postdoctoral fellowship, Generalitat de Catalunya (2007BP-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.

## 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 or 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 (FRAM), the Old Order Amish Study (Amish), 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 (See Table S1). To validate top associating variants for risk of histologically verified NAFLD, we utilized cases from the NASH Clinical Research Network (NASH CRN) that were genetically matched to healthy controls from the Myocardial Genetics Consortium (MIGen) consortium (See Table S1). We then further tested genome wide significant or replicating SNPs for associations with histologic NAFLD using the same cases from the NASH Clinical Research Network (NASH CRN) versus a different set of controls from the Illumina Control Database (iCONT) (See Table S1). Further, we report the association of these SNPs with other metabolic traits using data from the Global Lipids Genetics [13], GIANT [14], DIAGRAM [15], and MAGIC [16] Consortia, as

well as investigate cis gene expression variation (eQTLs) in liver, subcutaneous and visceral fat from bariatric surgery patients from Massachusetts General Hospital [17] (Figure 1).

## 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 ( $p < 5 \times 10^{-8}$ ; Table 2, Figure 2A). These included rs738409 in *PNPLA3* ( $p = 4.3 \times 10^{-34}$ ), a locus previously reported as associated with magnetic resonance spectroscopy measured steatosis, [8] and two additional novel loci: rs4240624 near *PPP1R3B* ( $rs4240624, p = 3.68 \times 10^{-18}$ ) and rs2228603 near *NCAN* ( $rs2228603, p = 1.22 \times 10^{-18}$ ). 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 quantile-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 histologically. 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.

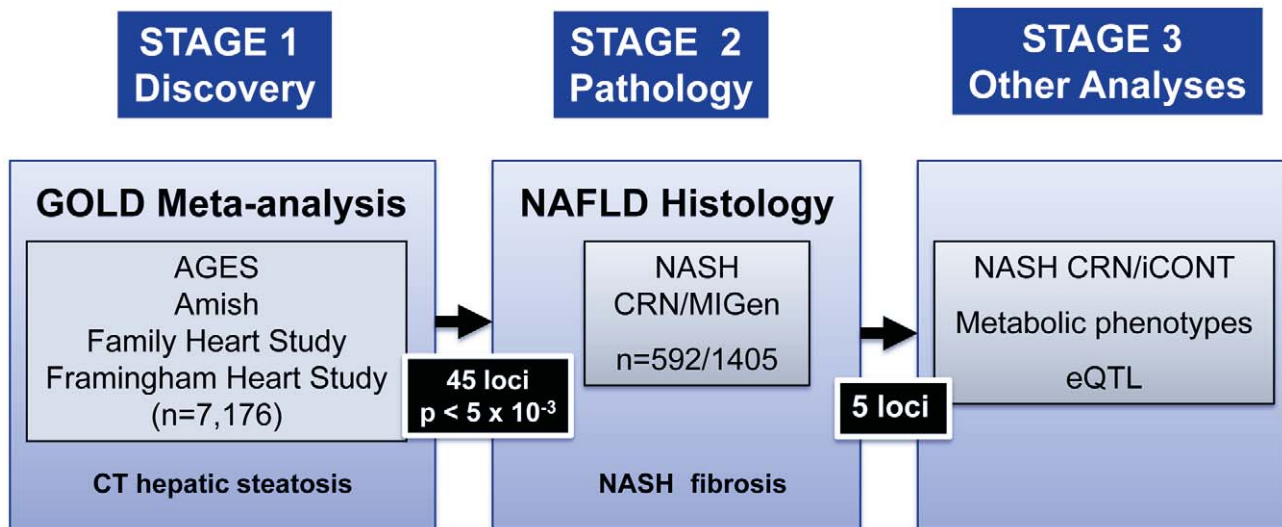
Affymetrix or Illumina platforms used by our studies and these variants do not have proxies that we could use in HapMap to impute them.

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 MIGen study [22] that had undergone GWAS genotyping and imputation (See Table S1). Forty-five of the 46 SNPs passed

genotyping and imputation quality control in the NASH CRN and MIGen data sets respectively (See Table S3) and were tested for association with histologic NAFLD in this sample. Two of the three variants with genome-wide significant associations to CT hepatic steatosis were also significantly associated with histologic NAFLD (corresponding to a false discovery rate (FDR)  $p < 0.001$ ): rs738409 in *PNPLA3* (OR = 3.26,  $p = 3.6 \times 10^{-43}$ ) as we and others have recently reported [21,23] and rs2228603 in *NCAN* (OR = 1.65,  $p = 5.29 \times 10^{-5}$ ) which is a novel finding (Table 2; See Table S5). The rs4240624 variant near *PPP1R3B* was not associated with histologic NAFLD in this sample (OR = 0.93,  $p = 0.29$ ). Of the 43 remaining SNPs showing suggestive association with CT hepatic steatosis, rs780094 in *GCKR* (OR = 1.45,  $p = 2.59 \times 10^{-8}$ ) and rs12137855 near *LYPLAL1* (OR = 1.37,  $p = 4.12 \times 10^{-5}$ ) were also significantly associated with histologic NAFLD (Table 2; See Table S5).

To confirm that the effects on histologic NAFLD observed in the NASH CRN/MIGen analyses were not due to the characteristics of the controls, we performed a separate analysis of the NASH CRN cases with an alternate set of controls from the Illumina Control database (iCONT; <http://www.illumina.com/science/icontroldb.ilmn>). We found that the effects and p values of rs738409 in *PNPLA3* (OR = 3.24,  $p = 2.16 \times 10^{-64}$ ), rs2228603 in *NCAN* (OR = 1.90,  $p = 6.82 \times 10^{-10}$ ), rs4240624 near *PPP1R3B* (OR = 0.86,  $p = 0.15$ ), rs780094 in *GCKR* (OR = 1.18,  $p = 0.01$ ), and rs12137855 near *LYPLAL1* (OR = 1.21,  $p = 0.03$ ) were similar to the effects seen in MIGen establishing that these results are not dependent on the choice of control sample (See Table S6). Furthermore, assessment of imputation accuracy with the SNPs in these control sets indicates that imputed genotypes at the associated SNPs are likely to be highly accurate (see Tables S7, S8).

The variants with the lowest p-values of association with CT hepatic steatosis at the *PNPLA3*(rs738408), *NCAN*(rs2228603), and *GCKR*(rs780094) loci are in high LD with or are themselves non-synonymous variants in *PNPLA3*(rs738409; I148M,  $R^2 = 1$ ), *NCAN*(rs2228603; P91S, same as hepatic steatosis SNP), and *GCKR*(rs1260326; P446L;  $R^2 = 0.93$ ) (Figure 3). The variants with



**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 MIGen 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

**Table 1.** Characterization of family data for heritability estimation.

Study	N	N families	Design	Age range (years)	Heritability	SE
Amish	880	1	founder population participants link to a single, 14-generation pedigree	29–94	0.27	0.08
Family Heart Study	2679	508	3-generational pedigrees	32–83	0.27	0.04
Framingham Heart Study	3070	721	2-generational pedigrees	31–83	0.26	0.04

N: total number of individuals with fatty liver phenotype; SE: Standard error; For all studies, SOLAR software was used to estimate heritability [47].  
doi:10.1371/journal.pgen.1001324.t001

the lowest p-values of association with CT hepatic steatosis at *LYPLAL1* and *PPP1R3B* 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 *PPP1R3B* 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 *PPP1R3B* 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 *PPP1R3B* 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 that the hepatic steatosis increasing variant (rs4240624) at the *PPP1R3B* locus increased liver mRNA expression of *PPP1R3B* and AW673036\_RC 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 *PPP1R3B* 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

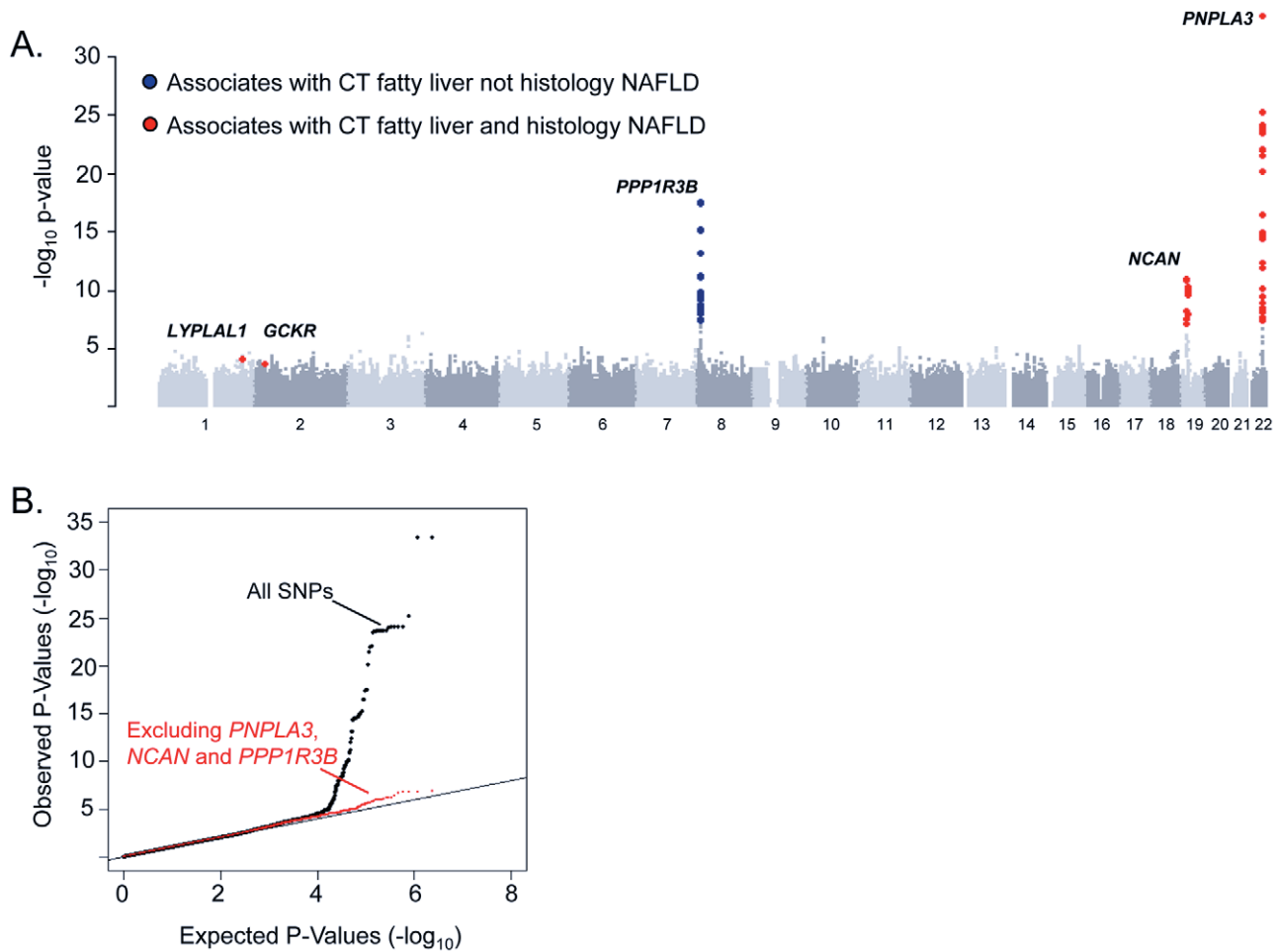
We have identified variants in three novel loci (*NCAN*, *GCKR*, and *LYPLAL1*) and one previously reported locus (*PNPLA3*) that are associated with both increasing CT hepatic steatosis and histologic NAFLD. *PPP1R3B* is associated with CT steatosis but not histologic NAFLD that includes individuals 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 *PPP1R3B* that was associated with CT-assessed liver attenuation but not histology-proven NAFLD. Possible reasons for why the variant near *PPP1R3B* 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.

Locus	GOLD Meta-analysis				NASH CRN/MiGen analysis				Metabolic Phenotypes Meta Analyses **						Dir.			
	n = 7,176	n = 592/1,405			n = 93,996–99,900			n = 35,478–46,182										
SNP ID	Chr.	Pos. (Mb)	Nearest Gene	EA	EAF	Effect (SE)	% Var	GOLD P	EAF <sup>a</sup>	EAF <sup>b</sup>	OR <sup>NAFLD</sup> (95% CI)	NAFLD P	LDL P	HDL P	TG P	Glucose	HOMA-IR	
rs738409	22	42656060	PNPLA3	G	0.23	0.26	2.41	4.30E-34	0.49	0.23	3.26 (2.11–7.21)	3.60E-43	1.41E-02	6.19E-03	2.21E-01	0.1867	0.1384	++NNNNN
rs2228603	19	19190924	NCAN	T	0.07	0.24	0.79	1.22E-11	0.12	0.08	1.65 (1.15–2.87)	5.29E-05	1.42E-19	6.78E-01	1.49E-23	0.3934	0.2329	++N-NN
rs12137855	1	217515001	LYPLAL1	C	0.79	0.08	0.23	1.06E-04	0.83	0.79	1.37 (1.17–1.57)	4.12E-05	8.71E-02	1.47E-01	7.34E-01	0.9815	0.4694	++NNNNN
rs780094	2	27594741	GCKR	T	0.39	0.06	0.2	2.56E-04	0.47	0.41	1.45 (1.19–1.86)	2.59E-08	4.37E-04	3.96E-02	7.08E-125	2.49E-12	7.626E-07	++N+—
rs4240624	8	9221641	PPP1R3B	A	0.92	0.29	1.18	3.68E-18	0.89	0.91	0.93 (0.68–1.18)	2.85E-01	1.27E-12	1.44E-23	1.94E-02	7.83E-06	0.001657	+N+N-N

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; EAF<sup>a</sup>: Frequency of the effect allele in cases from the NASH CRN study; EAF<sup>b</sup>: Frequency of the effect allele in controls from the MiGen study; NAFLD: nonalcoholic fatty liver disease; OR<sup>NAFLD</sup>: 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



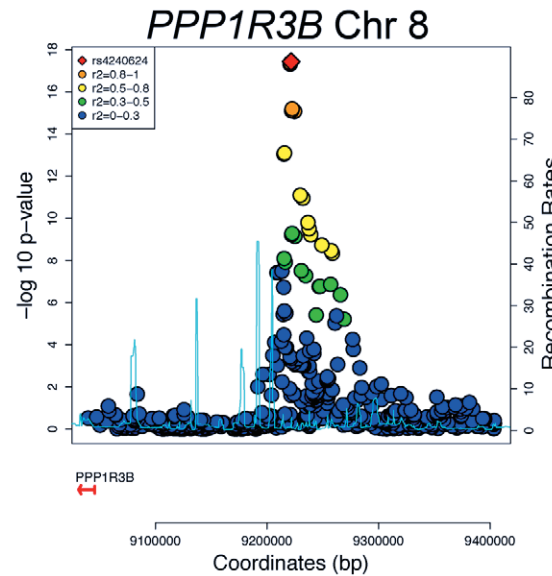
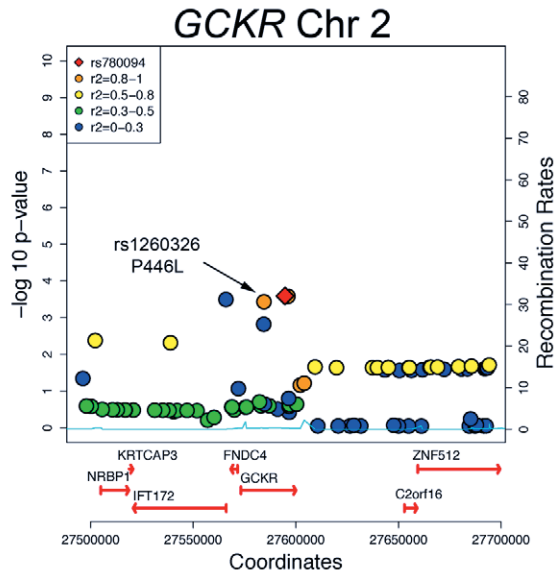
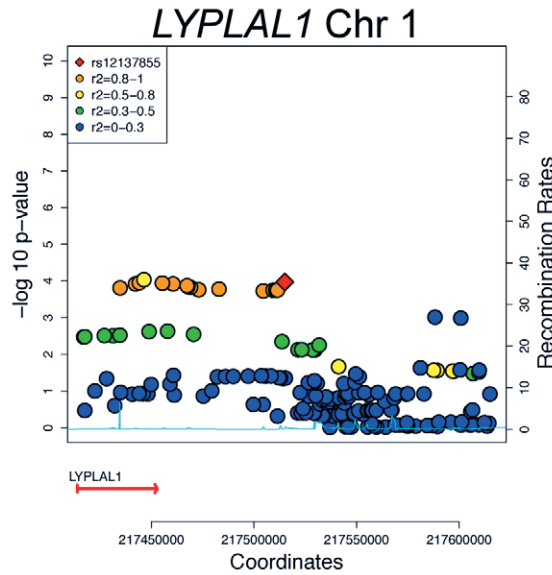
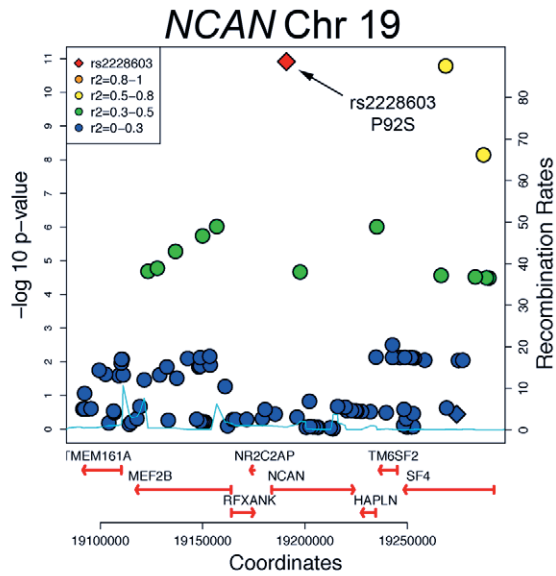
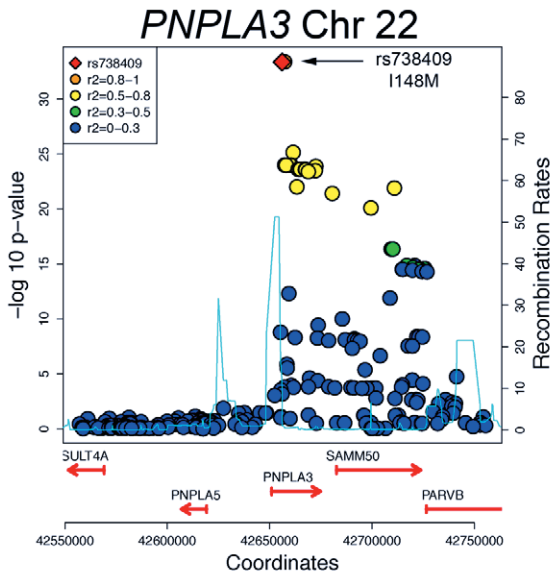
**Figure 2. Genome-wide association results for GOLD (Stage 1).** A. Manhattan plot showing the significance of association of all SNPs in the Stage 1 GOLD meta-analysis with CT hepatic steatosis. SNPs are plotted on the x-axis according to their position on each chromosome against association with CT hepatic steatosis on the y-axis (shown as  $-\log_{10}$  p-value). SNPs that also associate with histology based NAFLD are in red, those that only associate with CT hepatic steatosis in blue. B. Quantile-quantile plot of SNPs after Stage 1 GOLD meta-analysis (black) and after removing any SNPs within 500 kb of *PNPLA3*, *PPP1R3B*, and *NCAN* (red). doi:10.1371/journal.pgen.1001324.g002

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 affect.

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



**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 ( $-\log_{10}$  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  $R^2$  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

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

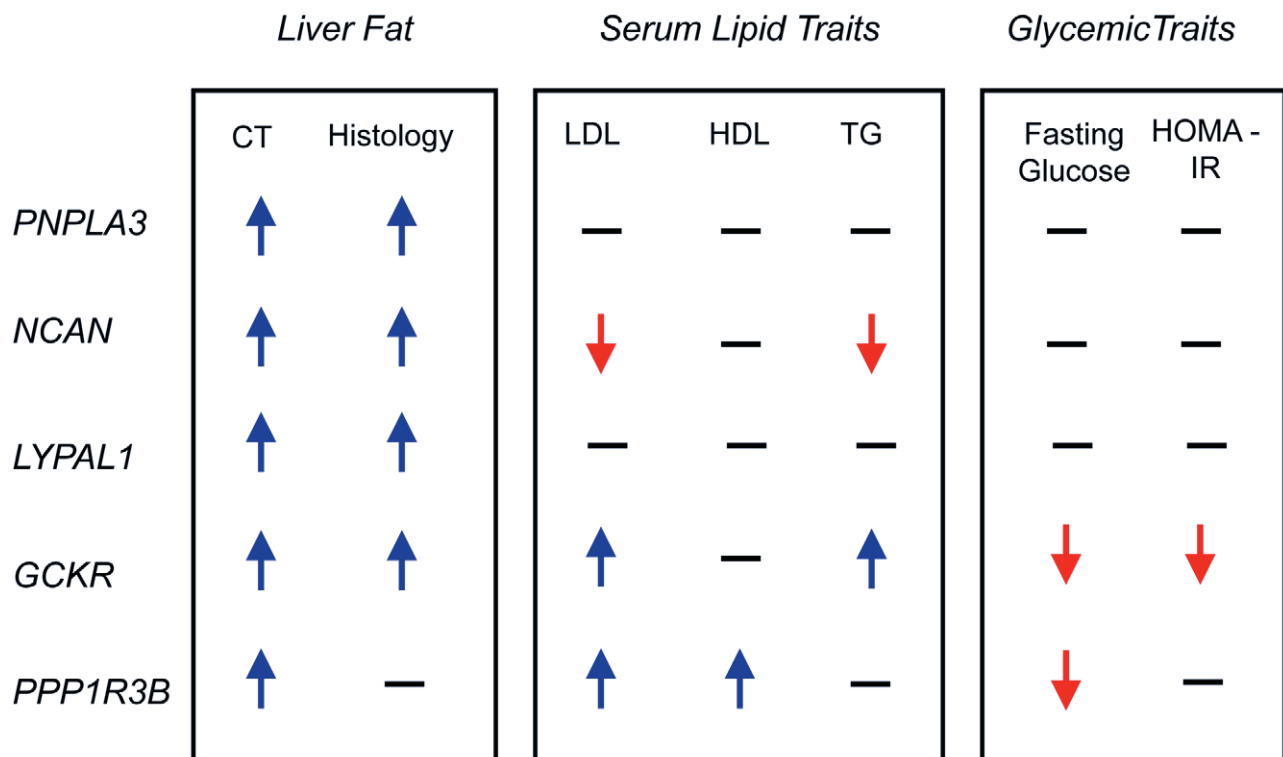
All work done in this paper was approved by local institutional review boards or equivalent committees.

### 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 computerized 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, age<sup>2</sup>, gender and drinks along with the SNPs in an additive genetic model (See Tables S2, S3).



**Figure 4. Effects on traits.** Direction of effect on CT fatty liver, histology NAFLD, lipid and glycemic traits of the best associating SNPs at the loci shown. Direction is shown only for significant associations. CT: CT hepatic steatosis; LDL: low density lipoprotein cholesterol; HDL: high density lipoprotein cholesterol; TG: triglycerides; HOMA-IR: homeostatic model of insulin resistance; *PNPLA3*: patatin-like phospholipase domain-containing protein 3 (HGNC: 18590); *NCAN*: neurocan (HGNC: 2465); *LYPAL1*: 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.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 (Accuimage Corp.) software was used to calculate the attenuation coefficient in Hounsfield Units for each region-of-interest. Two 1.0-cm<sup>2</sup> 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 an additive genetic model controlling for age, age<sup>2</sup>, and gender and relatedness; alcohol is generally not consumed in this population. A (n-1)-degree-of-freedom *t* test was used to assess the significance of the measured genotype. The polygenic component was modeled using the relationship matrix derived from the complete 14-generation pedigree structure, to properly control for the relatedness of all subjects in the study.

**The Family Heart Study.** The Family Heart Study (<https://dsgweb.wustl.edu/PROJECTS/MP1.html>) recruited 1,200 families, half randomly sampled, and half selected because of an excess of coronary heart disease (CHD) or risk factor abnormalities as compared with age- and sex-specific population rates [39] from four population-based parent studies: the Framingham Heart Study, the Utah Family Tree Study, and two Atherosclerosis Risk in Communities centers (Minneapolis, and Forsyth County, NC). Study participants belonging to the largest pedigrees were invited for a second clinical exam, at which time coronary artery calcification was assessed using computed tomography, which included imaging of the liver. A total of 2,767 Caucasian subjects in 508 extended families were examined; the heritability was estimated in this sample. A two-stage design was adopted for the GWAS. In the first stage, 1,016 subjects were chosen, equally distributed between the highest and lowest quartiles of age- and sex-adjusted values for coronary artery calcification, assessed by CT scan. These subjects were chosen to be largely unrelated with 200 subjects having 1 or more siblings selected into the sample. We report association results based on 886 subjects after excluding 130 subjects ascertained from the Framingham Massachusetts to avoid any possible overlap with the Framingham Heart Study participants.

Participants underwent a cardiac multidetector CT exam with four detectors using a standardized protocol as described previously [40]. For participants weighing 100 kg (220 lbs) or greater, the milliamperes were increased by 25%. Participants received two sequential scans of the heart with ECG gating in late diastole. A phantom with either 3 or 4 samples of calcium hydroxylapatite was included in each participants scan. CT images from all study centers were sent electronically to the central CT reading center located at Wake Forest University Health Sciences, Winston Salem, NC, USA.

CT images were analyzed using Medical Image Processing, Analysis, and Visualization (MIPAV) software (McAuliffe 2009) with custom programmed subroutines (a.k.a. “plug-ins”) coded at Wake Forest University Health Sciences. CT images of the chest

were used to measure liver attenuation corresponding to superior aspects of the right and medial lobes or hepatic segments 4a, 7 and 8 using the Couinaud system. An external calcium standard was used as a control for penetrance of the films.

The liver attenuation was regressed on age, age<sup>2</sup>, age<sup>3</sup>, 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, 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  $2f(1-f)a^2$ , where  $f$  is the frequency of the variant and  $a$  is its additive effect in units of standard deviations from the meta analysis [48].

### 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 0–3 and cytologic 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 US 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<sup>2</sup>, 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/icontrdb.ilmmn>). 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 ( $R^2$  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](http://www.sph.umich.edu/csg/abecasis/public/Teslovich%20et%20al.%202010)) [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  $<3.5 \times 10^{-5}$  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 in 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 hydroxyapetite CT control; Median raw liver measures in Hounsfield units; steatosis >5% more than 5% steatosis on histology; NASH: having histologic criteria for diagnosis of nonalcoholic steatohepatitis (NASH); Fibrosis: having histologic criteria for diagnosis of fibrosis.  
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; EAFa: 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: Tuscans 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. Impa: imputation quality score in MIGen; Concordancea: 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.

Found at: doi:10.1371/journal.pgen.1001324.s009 (0.08 MB DOC)

**Table S10** 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 transcript expression level for the effect allele. P: p-value of association of the fatty liver SNP with change in gene expression. Padj<sub>b</sub>: 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 Rsq<sub>d</sub>: the R squared correlation between the fatty liver SNP and the peak SNP. Padj<sub>e</sub>: p-value for the peak 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.

Found at: doi:10.1371/journal.pgen.1001324.s010 (0.04 MB DOC)

## 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.

### NASH CRN

Arthur McCullough, M.D.; Diane Bringman, R.N., B.S.N.; Srinivasan Dasarathy, M.D.; Kevin Edwards, N.P.; Carol Hawkins, R.N.; Yao-Chang Liu, M.D.; Nicholette Rogers, Ph.D.; P.A.-C.; Ruth Sargent, L.P.N.; Margaret Stager, M.D.; Anna Mae Diehl, M.D.; Manal Abdelmalek, M.D.; Marcia Gottfried, M.D.; Cynthia Guy, M.D.; Paul Killenberg, M.D.; Samantha Kwan; Yi-Ping Pan; Dawn Piercy, F.N.P.; Melissa Smith; Naga Chalasani, M.D.; Prajakta Bhimalli; Oscar W. Cummings, M.D.; Ann Klipsch, RN; Lydia Lee; Jean Mollenst, M.D.; Linda Ragozzino; Raj Vuppalanchi, M.D.; Brent A. Neuschwander-Tetri, M.D.; Sarah Barlow, M.D.; Jose Derdoy, M.D.; Joyce Hoffmann; Debra King, R.N.; Joan Siegner, R.N.; Susan Stewart, R.N.; Judy Thompson, R.N.; Elizabeth Brunt, M.D.; Joel E. Lavine, M.D., Ph.D.; Cynthia Behling, M.D.; Lisa Clark; Janis Durrelle; Tarek Hassanein, M.D.; Lita Petcharaporn; Jeffrey B. Schwimmer, M.D.; Claude Sirlin, M.D.; Tanya Stein; Nathan M. Bass, M.D., Ph.D.; Kiran Bambha, M.D.; Linda D. Ferrell, M.D.; Danuta Filipowski; Raphael Merriman, M.D.; Mark Pabst; Monique Rosenthal; Philip Rosenthal, M.D.; Tessa Steel; Arun J. Sanyal, M.D.; Sherry Boyett, R.N.; Daphne Bryan, M.D.; Melissa J. Contos, M.D.; Michael Fuchs, M.D.; Martin Graham, M.D.; Amy Jones; Velimir A.C. Luketic, M.D.; Bimaljiti Sandhu, M.D.; Carol Sargeant, R.N., M.P.H.; Kimberly Selph; Melanie White, R.N.; Kris V. Kowdley, M.D.; Grace Gyurkey; Jody Mooney, M.S.; James Nelson, Ph.D.; Sarah Roberts; Cheryl Saunders, M.P.H.; Alice Stead; Chia Wang, M.D.; Matthew Yeh, M.D., Ph.D.; David Kleiner, M.D., Ph.D.; Edward Doo, M.D.; Jay Everhart, M.D., M.P.H.; Jay H. Hoofnagle, M.D.; Patricia R. Robuck, Ph.D.; Leonard Seeff, M.D.; James Tonascia, Ph.D.; Patricia Belt, B.S.; Fred Brancati, M.D., M.H.S.; Jeanne Clark, M.D., M.P.H.; Ryan Colvin, M.P.H.; Michele Donithan, M.H.S.; Mika Green, M.A.; Milana Isaacson; Wana Kim; Laura Miriel; Alice Sternberg, Sc.M.; Aynur Ünalp, M.D., Ph.D.; Mark Van Natta, M.H.S.; Laura Wilson, Sc.M.; Katherine Yates, Sc.M.

### GIANT Consortium

Elizabeth K. Speliotes, Cristen J. Willer, Sonja I. Berndt, Keri L. Monda, Gudmar Thorleifsson, Anne U. Jackson, Hana Lango Allen, Cecilia M. Lindgren, Jian'an Luan, Reedik Mägi, Joshua C. Randall, Sailaja Vedantam, Thomas W. Winkler, Lu Qi, Tsegaselassie Work-

alemahu, Iris M. Heid, Valgerdur Steinthorsdottir, Heather M. Stringham, Michael N. Weedon, Eleanor Wheeler, Andrew R. Wood, Teresa Ferreira, Robert J. Weyant, Ayellet V. Segrè, Karol Estrada, Liming Liang, James Nemesh, Ju-Hyun Park, Stefan Gustafsson, Tuomas O. Kilpeläinen, Jian Yang, Nabila Bouatia-Naji, Tõnu Esko, Mary F. Feitosa, Zoltán Kutalik, Massimo Mangino, Soumya Raychaudhuri, Andre Scherag, Albert Vernon Smith, Ryan Welch, Jing Hua Zhao, Katja K. Aben, Devin M. Absher, Najaf Amin, Anna L. Dixon, Eva Fisher, Nicole L. Glazer, Michael E. Goddard, Nancy L. Heard-Costa, Volker Hoesel, Jouke-Jan Hottenga, Åsa Johansson, Toby Johnson, Shamika Ketkar, Claudia Lamina, Shengxu Li, Miriam F. Moffatt, Richard H. Myers, Narisu Narisu, John R.B. Perry, Marjolein J. Peters, Michael Preuss, Samuli Ripatti, Fernando Rivadeneira, Camilla Sandholt, Laura J. Scott, Nicholas J. Timpson, Jonathan P. Tyrer, Sophie van Wingerden, Richard M. Watanabe, Charles C. White, Fredrik Wiklund, Christina Barlassina, Daniel I. Chasman, Matthew N. Cooper, John-Olov Jansson, Robert W. Lawrence, Niina Pellikka, Inga Prokopenko, Jianxin Shi, Elisabeth Thiering, Helene Alavere, Maria T. S. Alibrandi, Peter Almgren, Alice M. Arnold, Thor Aspelund, Larry D. Atwood, Beverley Balkau, Anthony J. Balmforth, Amanda J. Bennett, Yoav Ben-Shlomo, Richard N. Bergman, Sven Bergmann, Heike Biebermann, Alexandra I.F. Blakemore, Tanja Boes, Lori L. Bonnycastle, Stefan R. Bornstein, Morris J. Brown, Thomas A. Buchanan, Fabio Busonero, Harry Campbell, Francesco P. Cappuccio, Christine Cavalcanti-Proença, Yii-Der Ida Chen, Chih-Mei Chen, Peter S. Chines, Robert Clarke, Lachlan Coin, John Connell, Ian N.M. Day, Martin den Heijer, Jubao Duan, Shah Ebrahim, Paul Elliott, Roberto Elosua, Gudny Eiriksdottir, Michael R. Erdos, Johan G. Eriksson, Maurizio F. Facheris, Stephan B. Felix, Pamela Fischer-Posovszky, Aaron R. Folsom, Nele Friedrich, Nelson B. Freimer, Mao Fu, Stefan Gaget, Pablo V. Gejman, Eco J.C. Geus, Christian Gieger, Anette P. Gjesing, Anuj Goel, Philippe Goyette, Harald Grallert, Jürgen Gräßler, Danielle ?M. Greenawalt, Christopher J. Groves, Vilmundur Gudnason, Candace Guiducci, Anna-Liisa Hartikainen, Neelam Hassanali, Alistair S. Hall, Aki S. Havulinna, Caroline Hayward, Andrew C. Heath, Christian Hengstenberg, Andrew A. Hicks, Anke Hinney, Albert Hofman, Georg Homuth, Jennie Hui, Wilmar Igl, Carlos Iribarren, Bo Isomaa, Kevin B. Jacobs, Ivonne Jarick, Elizabeth Jewell, Ulrich John, Torben Jørgensen, Pekka Jousilahti, Antti Jula, Marika Kaakinen, Eero Kajantie, Lee M. Kaplan, Sekar Kathiresan, Johannes Kettunen, Leena Kinnunen, Joshua W. Knowles, Ivana Kolcic, Inke R. König, Seppo Koskinen, Peter Kovacs, Johanna Kuusisto, Peter Kraft, Kirsti Kvaløy, Jaana Laitinen, Olivier Lantieri, Chiara Lanzani, Lenore J. Launer, Cecile Lecocour, Terho Lehtimäki, Guillaume Lettre, Jianjun Liu, Marja-Liisa Lokki, Mattias Lorentzon, Robert N. Luben, Barbara Ludwig, MAGIC, Paolo Manunta, Diana Marek, Michel Marre, Nicholas G. Martin, Wendy L. McArdle, Anne McCarthy, Barbara McKnight, Thomas Meitinger, Olle Melander, David Meyre, Kristian Midthjell, Grant W. Montgomery, Mario A. Morken, Andrew P. Morris, Rosanda Mulic, Julius S. Ngwa, Mari Nelis, Matt J. Neville, Dale R. Nyholt, Christopher J. O'Donnell, Stephen O'Rahilly, Ken K. Ong, Ben Oostra, Guillaume Paré, Alex N. Parker, Markus Perola, Irene Pichler, Kirsi H. Pietiläinen, Carl G.P. Platou, Ozren Polasek, Anneli Pouta, Suzanne Rafelt, Olli Raitakari, Nigel W. Rayner, Martin Ridderstråle, Winfried Rief, Aimo Ruokonen, Neil R. Robertson, Peter Rzehak, Veikko Salomaa, Alan R. Sanders, Manjinder S. Sandhu, Serena Sanna, Jouko Saramies, Markku J. Savolainen, Susann Scherag, Sabine Schipf, Stefan Schreiber, Heribert Schunkert, Kaisa Silander, Juha Sinisalo, David S. Siscovick, Jan H. Smit, Nicole Soranzo, Ulla Sovio, Jonathan Stephens, Ida Surakka, Amy J. Swift, Mari-Liis Tammesoo, Jean-Claude Tardif, Maris Teder-Laving, Tanya M. Teslovich, John R. Thompson, Brian Thomson, Anke Tönjes, Tiinamajja Tuomi, Joyce B.J. van Meurs, Gert-Jan van Ommen, Vincent Vatin, Jorma Viikari, Sophie Visvikis-Siest, Veronique Vitart, Carla I. G. Vogel, Benjamin F. Voight, Lindsay L. Waite, Henri Wallaschofski, G.Bragi Walters, Elisabeth Widen, Susanna Wiegand, Sarah H. Wild, Gonke Willemsen, Daniel R. Witte, Jacqueline C. Wittman, Jianfeng Xu, Qunyan Zhang, Lina Zgaga, Andreas Ziegler, Paavo Zitting, John P. Beilby, I. Sadaf Farooqi, Johannes Hebebrand, Heikki V. Huikuri, Alan L. James, Mika Kähönen, Douglas F. Levinson, Fabio Maciardi, Markku S. Nieminen, Claes Ohlsson, Lyle J. Palmer, Paul M. Ridker, Michael Stumvoll, Jacques S. Beckmann, Heiner Boeing, Eric Boerwinkle, Dorret I. Boomsma, Mark J. Caulfield, Stephen J. Chanock, Francis S. Collins, L. Adrienne Cupples, George Davey Smith, Jeanette Erdmann, Philippe Froguel, Henrik Grönberg, Ulf Gyllenstein, Per Hall, Torben Hansen, Tamara B. Harris, Andrew T. Hattersley, Richard

B. Hayes, Joachim Heinrich, Frank B. Hu, Kristian Hveem, Thomas Illig, Marjo-Riitta Jarvelin, Jaakko Kaprio, Fredrik Karpe, Kay-Tee Khaw, Lambertus A. Kiemeny, Heiko Krude, Markku Laakso, Debbie A. Lawlor, Andres Metspalu, Patricia B. Munroe, Willem H. Ouwehand, Oluf Pedersen, Brenda W. Penninx, Annette Peters, Peter P. Pramstaller, Thomas Quertermous, Thomas Reinehr, Aila Rissanen, Igor Rudan, Nilesh J. Samani, Peter E.H. Schwarz, Alan R. Shuldiner, Timothy D. Spector, Jaakko Tuomilehto, Manuela Uda, André Uitterlinden, Timo T. Valle, Martin Wabitsch, Gérard Waeber, Nicholas J. Wareham, Hugh Watkins on behalf of Procardis Consortium, James F. Wilson, Alan F. Wright, M. Carola Zillikens, Nilanjana Chatterjee, Steven A. McCarroll, Shaun Purcell, Eric E. Schadt, Peter M. Visscher, Themistocles L. Assimes, Ingrid B. Borecki, Panos Deloukas, Caroline S. Fox, Leif C. Groop, Talin Haritunians, David J. Hunter, Robert C. Kaplan, Karen L. Mohlke, Jeffrey R. O'Connell, Leena Peltonen, David Schlessinger, David P. Strachan, Cornelia M. van Duijn, H-Erich Wichmann, Timothy M. Frayling, Unnur Thorsteinsdottir, Gonçalo R. Abecasis, Inês Barroso, Michael Boehnke, Kari Stefansson, Kari E. North, Mark I. McCarthy, Joel N. Hirschhorn, Erik Ingelsson, Ruth J.F. Loos

#### MAGIC Investigators

Josée Dupuis, Claudia Langenberg, Inga Prokopenko, Richa Saxena, Nicole Soranzo, Anne U Jackson, Eleanor Wheeler, Nicole L Glazer, Nabila Bouatia-Naji, Anna L Gloyn, Cecilia M Lindgren, Reedik Mägi, Andrew P Morris, Joshua Randall, Toby Johnson, Paul Elliott, Denis Rybin, Gudmar Thorleifsson, Valgerdur Steinthorsdottir, Peter Henne- man, Harald Grallert, Abbas Dehghan, Jouke Jan Hottenga, Christopher S Franklin, Pau Navarro, Kijoung Song, Anuj Goel, John R B Perry, Josephine M Egan, Taina Lajunen, Niels Grarup, Thomas Sparso, Alex Doney, Benjamin F Voight, Heather M Stringham, Man Li, Stavroula Kanoni, Peter Shrader, Christine Cavalcanti-Proença, Meena Kumari, Lu Qi, Nicholas J Timpson, Christian Gieger, Carina Zabena, Ghislain Rocheleau, Erik Ingelsson, Ping An, Jeffrey O'Connell, Jian'an Luan, Amanda Elliott, Steven A McCarroll, Felicity Payne, Rosa Maria Roccasecca, François Pattou, Praveen Sethupathy, Kristin Ardlie, Yavuz Ariyurek, Beverley Balkau, Philip Barter, John P Beilby, Yoav Ben-Shlomo, Rafn Benediktsson, Amanda J Bennett, Sven Bergmann, Murielle Bochud, Eric Boerwinkle, Amélie Bonnefond, Lori L Bonnycastle, Knut Borch-Johnsen, Yvonne Böttcher, Eric Brunner, Suzannah J Bumpstead, Guillaume Charpentier, Yii-Der Ida Chen, Peter Chines, Robert Clarke, Lachlan J M Coin, Matthew N Cooper, Marilyn Cornelis, Gabe Crawford, Laura Crisponi, Ian N M Day, Eco J C de Geus, Jerome Delplanque, Christian Dina, Michael R Erdos, Annette C Fedson, Antje Fischer-Rosinsky, Nita G Forouhi, Caroline S Fox, Rune Frants, Maria Grazia Franzosi, Pilar Galan, Mark O Goodarzi, Jürgen Graessler, Christopher J Groves, Scott Grundy, Rhian G Williams, Ulf Gyllenstein, Samy Hadjadj, Göran Hallmans, Naomi Hammond, Xijing Han, Anna-Liisa Hartikainen, Neelam Hassanali, Caroline Hayward, Simon C Heath, Serge Hercberg, Christian Herder, Andrew A Hicks, David R Hillman, Aroon DHingorani, Albert Hofman, Jennie Hui, Joe Hung, Bo Isomaa, Paul R V Johnson, Torben Jørgensen, Antti Jula, Marika Kaakinen, Jaakko Kaprio, Y Antero Kesaniemi, Mika Kivimäki, Beatrice Knight, Seppo Koskinen, Peter Kovacs, Kirsten Ohm Kyvik, GMark Lathrop, Debbie A Lawlor, Olivier Le Bacquer, Cécile Lecoeur, Yun Li, Valeriya Lyssenko, Robert Mahley, Massimo Mangino, Alisa K Manning, Maria Teresa Martínez-Larrad, Jarred B McAteer, Laura J McCulloch, Ruth McPherson, Christa Meisinger, David Melzer, David Meyre, Braxton D Mitchell, Mario A Morken, Sutapa Mukherjee, Silvia Naitza, Narisu Narisu, Matthew J Neville, Ben A Oostra, Marco Orrù, Ruth Pakyz, Colin NA Palmer, Giuseppe Paillos, Cristian Pattaro, Daniel Pearson, John F Peden, Nancy

L Pedersen, Markus Perola, Andreas F H Pfeiffer, Irene Pichler, Ozren Polasek, Danielle Posthuma, Simon C Potter, Anneli Pouta, Michael A Province, Bruce M Psaty, Wolfgang Rathmann, Nigel W Rayner, Kenneth Rice, Samuli Ripatti, Fernando Rivadeneira, Michael Roden, Olov Rolandsson, Anneli Sandback, Manjinder Sandhu, Serena Sanna, Avan Aihie Sayer, Paul Scheet, Laura J Scott, Udo Seedorf, Stephen J Sharp, Beverley Shields, Gunnar Sigurdsson, Eric J G Sijbrands, Angela Silveira, Laila Simpson, Andrew Singleton, Nicholas L Smith, Ulla Sovio, Amy Swift, Holly Syddall, Ann-Christine Syvänen, Toshiko Tanaka, Barbara Thorand, Jean Tichet, Anke Tönjes, Tiinamaija Tuomi, André G Uitterlinden, Ko Willems van Dijk, Mandy van Hoek, Dhiraj Varma, Sophie Visvikis-Siest, Veronique Vitart, Nicole Vogelzangs, Gérard Waeber, Peter J Wagner, Andrew Walley, G Bragi Walters, Kim L Ward, Hugh Watkins, Michael N Weedon, Sarah H Wild, Gonke Willemsen, Jacqueline C M Wittman, John W G Yarnell, Eleftheria Zeggini, Diana Zelenika, Björn Zethelius, Guangju Zhai, Jing Hua Zhao, M Carola Zillikens, DIAGRAM Consortium, GIANT Consortium, Global BPgen Consortium, Ingrid B Borecki, Ruth J F Loos, Pierre Meneton, Patrik KEMagnusson, David M Nathan, Gordon H Williams, Andrew T Hattersley, Kaisa Silander, Veikko Salomaa, George Davey Smith, Stefan R Bornstein, Peter Schwarz, Joachim Spranger, Fredrik Karpe, Alan R Shuldiner, Cyrus Cooper, George V Dedoussis, Manuel Serrano-Ríos, Andrew D Morris, Lars Lind, Lyle J Palmer, Frank B Hu, Paul W Franks, Shah Ebrahim, Michael Marmot, W H Linda Kao, James S Pankow, Michael J Sampson, Johanna Kuusisto, Markku Laakso, Torben Hansen, Oluf Pedersen, Peter Paul Pramstaller, H Erich Wichmann, Thomas Illig, Igor Rudan, Alan F Wright, Michael Stumvoll, Harry Campbell, James F Wilson, Anders Hamsten on behalf of Procardis Consortium, Richard N Bergman, Thomas A Buchanan, Francis S Collins, Karen L Mohlke, Jaakko Tuomilehto, Timo T Valle, David Altshuler, Jerome I Rotter, David S Siscovick, Brenda W J H Penninx, Dorret I Boomsma, Panos Deloukas, Timothy D Spector, Timothy M Frayling, Luigi Ferrucci, Augustine Kong, Unnur Thorsteinsdottir, Kari Stefansson, Cornelia Mvan Duijn, Yuri S Aulchenko, Antonio Cao, Angelo Scuteri, David Schlessinger, Manuela Uda, Aimo Ruokonen, Marjo-Riitta Jarvelin, Dawn M Waterworth, Peter Vollenweider, Leena Peltonen, Vincent Mooser, Goncalo R Abecasis, Nicholas J Wareham, Robert Sladek, Philippe Froguel, Richard M Watanabe, James B Meigs, Leif Groop, Michael Boehnke, Mark I McCarthy, Jose C Florez & Inês Barroso for the MAGIC investigators.

#### GOLD Consortium

Gudny Eiriksdottir, Melissa E. Garcia, Vilmondur Gudnason, Tamara B. Harris, Lauren J. Kim, Lenore J. Launer, Michael A. Nalls, Albert V. Smith, Jeanne M. Clark, Ruben Hernaez, W. H. Linda Kao, Braxton D. Mitchell, Alan R. Shuldiner, Laura M. Yerges-Armstrong, Ingrid B. Borecki, J. Jeffrey Carr, Mary F. Feitosa, Jun Wu, Johannah L. Butler, Caroline S. Fox, Joel N. Hirschhorn, Udo Hoffmann, Shih-Jen Hwang, Joseph M. Massaro, Christopher J. O'Donnell, Cameron D. Palmer, Dushyant V. Sahani, Elizabeth K. Speliotes.

#### Author Contributions

Conceived and designed the experiments: EKS CJO CSF WHLK JNH IBB. Performed the experiments: EKS LMYA JW RH JLB MFF. Analyzed the data: EKS LMYA JW RH CDP GE MEG IJL MAN SJH JMM BFV AVS. Contributed reagents/materials/analysis tools: EKS VG JMC BDM ARS MT UH JMM CJO DVS VS EES SMS DSS NASH CRN GIANT Consortium MAGIC Investigators JJC TBH CSF WHLK JNH IBB. Wrote the paper: EKS LMYA JW RH IJL CSF WHLK JNH IBB GOLD Consortium.

#### References

- Clark JM (2006) The epidemiology of nonalcoholic fatty liver disease in adults. *J Clin Gastroenterol* 40: S5–10.
- Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, et al. (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 41: 1313–1321.
- Harrison SA, Neuschwander-Tetri BA (2004) Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Clin Liver Dis* 8: 861–879.
- Fan JG, Zhu J, Li XJ, Chen L, Lu YS, et al. (2005) Fatty liver and the metabolic syndrome among Shanghai adults. *J Gastroenterol Hepatol* 20: 1825–1832.
- Struben VM, Hespdenheide EE, Caldwell SH (2000) Nonalcoholic steatohepatitis and cryptogenic cirrhosis within kindreds. *Am J Med* 108: 9–13.
- Willner IR, Waters B, Patil SR, Reuben A, Morelli J, et al. (2001) Ninety patients with nonalcoholic steatohepatitis: insulin resistance, familial tendency, and severity of disease. *Am J Gastroenterol* 96: 2957–2961.
- Schwimmer JB, Celedon MA, Lavine JE, Salem R, Campbell N, et al. (2009) Heritability of nonalcoholic fatty liver disease. *Gastroenterology* 136: 1585–1592.
- Romeo S, Kozlitina J, Xing C, Pertsemliadis A, Cox D, et al. (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 40: 1461–1465.
- Kotronen A, Johansson LE, Johansson LM, Roos C, Westerbacka J, et al. (2009) A common variant in PNPLA3, which encodes adiponutrin, is associated with liver fat content in humans. *Diabetologia* 52: 1056–1060.

10. Iwasaki M, Takada Y, Hayashi M, Minamiguchi S, Haga H, et al. (2004) Noninvasive evaluation of graft steatosis in living donor liver transplantation. *Transplantation* 78: 1501–1505.
11. Saadeh S, Younossi ZM, Remer EM, Gramlich T, Ong JP, et al. (2002) The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology* 123: 745–750.
12. Limanond P, Raman SS, Lassman C, Sayre J, Ghobrial RM, et al. (2004) Macrovesicular hepatic steatosis in living related liver donors: correlation between CT and histologic findings. *Radiology* 230: 276–280.
13. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466: 707–713.
14. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937–948.
15. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42: 579–589.
16. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42: 105–116.
17. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE (2010) Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet* 86: 581–591.
18. Yuan X, Waterworth D, Perry JR, Lim N, Song K, et al. (2008) Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet* 83: 520–528.
19. Petersen KF, Dufour S, Hariri A, Nelson-Williams C, Foo JN, et al. (2010) PNPLA3 variants specifically confer increased risk for histologic NAFLD but not metabolic disease. *Hepatology* 52: 904–912.
20. Price AL, Butler J, Patterson N, Capelli C, Pascali VL, et al. (2008) Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet* 4: e236. doi:10.1371/journal.pgen.0030236.
21. Speliotes EK, Butler JL, Palmer CD, Voight BF, Consortium tG, et al. (2010) PNPLA3 variants specifically confer increased risk for histologic NAFLD but not metabolic disease. *Hepatology* 52: 904–912.
22. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41: 334–341.
23. Valenti L, Al-Serri A, Daly AK, Galmozzi E, Rametta R, et al. Homozygosity for the patatin-like phospholipase-3/adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. *Hepatology* 51: 1209–1217.
24. Speliotes EK, Massaro JM, Hoffmann U, Ramachandran V, Meigs JB, et al. (2010) Fatty liver is associated with dyslipidemia and dysglycemia independent of visceral fat: The Framingham Heart Study. *Hepatology* 51: 1979–1987.
25. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, et al. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316: 1331–1336.
26. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, et al. (2008) Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 40: 161–169.
27. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, et al. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet* 41: 56–65.
28. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, et al. (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 40: 189–197.
29. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, et al. (2009) Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 5: e1000508. doi:10.1371/journal.pgen.1000508.
30. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, et al. (2009) The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum Mol Genet* 18: 4081–4088.
31. Burke JE, Dennis EA (2009) Phospholipase A2 structure/function, mechanism, and signaling. *J Lipid Res* 50 Suppl: S237–242.
32. Burke JE, Dennis EA (2009) Phospholipase A2 biochemistry. *Cardiovasc Drugs Ther* 23: 49–59.
33. He S, McPhaul C, Li JZ, Garuti R, Kinch LN, et al. (2009) A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. *J Biol Chem* 285: 6706–6715.
34. Rauch U, Feng K, Zhou XH (2001) Neurocan: a brain chondroitin sulfate proteoglycan. *Cell Mol Life Sci* 58: 1842–1856.
35. Agius L (2008) Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem J* 414: 1–18.
36. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23: 1294–1296.
37. Rampersaud E, Bielak LF, Parsa A, Shen H, Post W, et al. (2008) The association of coronary artery calcification and carotid artery intima-media thickness with distinct, traditional coronary artery disease risk factors in asymptomatic adults. *Am J Epidemiol* 168: 1016–1023.
38. Sorkin J, Post W, Pollin TI, O'Connell JR, Mitchell BD, et al. (2005) Exploring the genetics of longevity in the Old Order Amish. *Mech Ageing Dev* 126: 347–350.
39. Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, et al. (1996) NHLBI Family Heart Study: objectives and design. *Am J Epidemiol* 143: 1219–1228.
40. Carr JJ, Nelson JC, Wong ND, McNitt-Gray M, Arad Y, et al. (2005) Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) study. *Radiology* 234: 35–43.
41. Dawber T (1980) *The Framingham Study: The Epidemiology of Atherosclerotic Disease*. Cambridge, MA: Harvard University Press.
42. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP (1979) An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 110: 281–290.
43. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, et al. (2007) The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 165: 1328–1335.
44. Speliotes EK, Massaro JM, Hoffmann U, Foster MC, Sahani DV, et al. (2008) Liver fat is reproducibly measured using computed tomography in the Framingham Heart Study. *J Gastroenterol Hepatol* 23: 894–899.
45. Chen MH, Larson MG, Hsu YH, Peloso GM, Guo CY, et al. (2010) A three-stage approach for genome-wide association studies with family data for quantitative traits. *BMC Genet* 11: 40.
46. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909.
47. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62: 1198–1211.
48. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, et al. (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41: 18–24.
49. Chalasani NP, Sanyal AJ, Kowdley KV, Robuck PR, Hoofnagle J, et al. (2009) Pioglitazone versus vitamin E versus placebo for the treatment of non-diabetic patients with non-alcoholic steatohepatitis: PIVENS trial design. *Contemp Clin Trials* 30: 88–96.
50. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559–575.