

Published in final edited form as:

*Clin Gastroenterol Hepatol.* 2013 September ; 11(9): 1183–1190.e2. doi:10.1016/j.cgh.2013.02.011.

## Association Between Variants in or Near *PNPLA3*, *GCKR*, and *PPP1R3B* With Ultrasound-Defined Steatosis Based on Data From the Third National Health and Nutrition Examination Survey

Ruben Hernaez<sup>\*,‡,§</sup>, Jody McLean<sup>||</sup>, Mariana Lazo<sup>\*,§</sup>, Frederick L. Brancati<sup>\*,§,¶</sup>, Joel N. Hirschhorn<sup>#,\*\*,‡‡</sup>, Ingrid B. Borecki<sup>§§</sup>, Tamara B. Harris<sup>|||</sup>, Genetics of Obesity-Related Liver Disease (GOLD) Consortium, Thutrang Nguyen<sup>#,\*\*</sup>, Ihab R. Kamel<sup>¶¶</sup>, Susanne Bonekamp<sup>¶¶</sup>, Mark S. Eberhardt<sup>##</sup>, Jeanne M. Clark<sup>\*,§,¶</sup>, Wen Hong Linda Kao<sup>§,¶</sup>, and Elizabeth K. Speliotes<sup>\*\*,\*\*\*</sup>

<sup>\*</sup>Department of Medicine, Johns Hopkins University, Baltimore, Maryland <sup>¶¶</sup>Department of Radiology, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, Maryland <sup>§</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland <sup>¶¶</sup>Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, Maryland <sup>‡</sup>Department of Medicine, Washington Hospital Center/Georgetown University Hospital, Washington, District of Columbia <sup>||</sup>NOVA Research Company, Bethesda, Maryland <sup>#</sup>Divisions of Endocrinology and Genetics and Program in Genomics, Children's Hospital, Boston, Massachusetts <sup>\*\*</sup>Broad Institute, Cambridge, Massachusetts <sup>‡‡</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts <sup>§§</sup>Department of Genetics, Washington University School of Medicine, St. Louis, Missouri <sup>|||</sup>Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, Maryland <sup>##</sup>National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland <sup>\*\*\*</sup>Department of Internal Medicine, Division of Gastroenterology, and Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan

### Abstract

**BACKGROUND & AIMS**—A genome-wide association study associated 5 genetic variants with hepatic steatosis (identified by computerized tomography) in individuals of European ancestry. We investigated whether these variants were associated with measures of hepatic steatosis (HS) in

© 2013 by the AGA Institute

*Reprint requests:* Address requests for reprints to: Elizabeth K. Speliotes, MD, PhD, MPH, University of Michigan, 6520 MSRB1, SPC 5682, 1150 W. Medical Center Drive, Ann Arbor, Michigan 48109-5682. [espeliot@med.umich.edu](mailto:espeliot@med.umich.edu); fax: (734) 763-2535.

#### Conflicts of interest

The authors disclose no conflicts.

#### Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at [www.cghjournal.org](http://www.cghjournal.org), and at <http://dx.doi.org/10.1016/j.cgh.2013.02.011>.

non-Hispanic white (NHW), non-Hispanic black, and Mexican American (MA) participants in the US population-based National Health and Nutrition Examination Survey III, phase 2.

**METHODS**—We analyzed data from 4804 adults (1825 NHW, 1442 non-Hispanic black, and 1537 MA; 51.7% women; mean age at examination, 42.5 y); the weighted prevalence of HS was 37.3%. We investigated whether ultrasound-measured HS, with and without increased levels of alanine aminotransferase (ALT), or level of ALT alone, was associated with rs738409 (patatin-like phospholipase domain-containing protein 3 [*PNPLA3*]), rs2228603 (neurocan [*NCAN*]), rs12137855 (*lysophospholipase-like 1*), rs780094 (glucokinase regulatory protein [*GCKR*]), and rs4240624 (protein phosphatase 1, regulatory subunit 3b [*PPP1R3B*]) using regression modeling in an additive genetic model, controlling for age, age-squared, sex, and alcohol consumption.

**RESULTS**—The G allele of rs738409 (*PNPLA3*) and the T allele of rs780094 (*GCKR*) were associated with HS with a high level of ALT (odds ratio [OR], 1.36;  $P = .01$ ; and OR, 1.30;  $P = .03$ , respectively). The A allele of rs4240624 (*PPP1R3B*) and the T allele of rs2228603 (*NCAN*) were associated with HS (OR, 1.28;  $P = .03$ ; and OR, 1.40;  $P = .04$ , respectively). Variants of *PNPLA3* and *NCAN* were associated with ALT level among all 3 ancestries. Some single-nucleotide polymorphisms were associated with particular races or ethnicities: variants in *PNPLA3*, *NCAN*, *GCKR*, and *PPP1R3B* were associated with NHW and variants in *PNPLA3* were associated with MA. No variants were associated with NHB.

**CONCLUSIONS**—We used data from the National Health and Nutrition Examination Survey III to validate the association between rs738409 (*PNPLA3*), rs780094 (*GCKR*), and rs4240624 (*PPP1R3B*) with HS, with or without increased levels of ALT, among 3 different ancestries. Some, but not all, associations between variants in *NCAN*, lysophospholipase-like 1, *GCKR*, and *PPP1R3B* with HS (with and without increased ALT level) were significant within subpopulations.

### Keywords

Nonalcoholic Fatty Liver Disease; Replication; Candidate Gene Study; Genome-wide Association Study; SNP

---

Obesity and its complications have increased to epidemic levels in the United States.<sup>1</sup> It is unknown why some people develop metabolic complications from obesity such as dyslipidemia, hypertension, diabetes, and myocardial infarction, while others do not. Interestingly, the presence of fat accumulation in the liver in the form of nonalcoholic fatty liver disease (NAFLD) is correlated strongly with the development of obesity-related metabolic complications.<sup>2</sup>

NAFLD can progress from steatosis to nonalcoholic steatohepatitis (NASH), and to cirrhosis.<sup>3</sup> Research has suggested that NAFLD is a common chronic condition, with an estimated prevalence in the United States of 5% to 33%; at least 30 million people have NAFLD and more than 600,000 people have NAFLD-related cirrhosis.<sup>4,5</sup> A study of NAFLD may provide insights into how some obesity leads to metabolic complications as well as clarify the pathophysiology of NAFLD, both of which ultimately could improve the management or prevention of this liver pathology in vast numbers of individuals.

Human genetic approaches recently have begun to shed light on the pathophysiology of many heritable conditions, including metabolic and inflammatory diseases.<sup>6–9</sup> Studies of liver enzymes such as alanine aminotransferase (ALT) (an indirect measure of NAFLD), and liver imaging (a direct measure of hepatic steatosis [HS]) have implicated several genetic variants that are associated with NAFLD.<sup>10–13</sup> Investigators testing missense variants in the genome of 2111 individuals for association with fatty liver disease (measured using proton magnetic resonance spectroscopy) found one in patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) (rs738409, encoding I148M) that is associated with HS.<sup>11–13</sup> The presence of HS was 2-fold higher in individuals carrying 2 alleles of this variant than in those who did not carry the combination of these alleles. Variants near *PNPLA3* also were associated with increased levels of ALT.<sup>10</sup>

The Genetics of Obesity-related Liver Disease consortium confirmed the association of a single-nucleotide polymorphism (SNP) in *PNPLA3* and identified 4 additional variants in or near neurocan (*NCAN*), glucokinase regulatory protein (*GCKR*), lysophospholipase-like 1 (*LYPLAL1*), and protein phosphatase 1, regulatory subunit 3b (*PPP1R3B*) that were associated with HS using computed tomography in individuals of European ancestry.<sup>13</sup> Variants in or near *NCAN*, *GCKR*, *LYPLAL1*, and *PNPLA3*, but not *PPP1R3B*, also were associated with NASH/fibrosis.

We recently reviewed right upper-quadrant ultrasounds to assess the presence of HS in a representative and genetically diverse sample of the US population from phase 2 of the Third National Health and Nutrition Examination Survey (NHANES III).<sup>14</sup> This recently available data provide a unique opportunity to examine the relationship of NAFLD and previously associated genetic variants in a genetically diverse sample of the US population.

Our primary aim was to test the association of previously reported NAFLD-related SNPs with ultrasound-defined HS in non-Hispanic whites (NHW), non-Hispanic blacks (NHB), and Mexican Americans (MA) who participated in NHANES III. We further examined whether these genetic variants are associated with ultrasound-defined steatosis in combination with high ALT levels, which may identify individuals particularly susceptible to hepatocyte injury or who have more severe disease than those with steatosis and no increases in ALT levels. Finally, we also evaluated the associations of these SNPs with continuous ALT levels and the aspartate aminotransferase (AST)/ALT ratio.

## Methods

### Study Population: Hepatic Steatosis Component of the Third National Health and Nutrition Examination Survey

NHANES III (1988 – 1994) included a combination of interviews, examinations, and laboratory tests and originally was intended to measure the health and nutritional status of a nationally representative sample of the civilian, noninstitutionalized US population. The survey used a complex, multistage probability sample design to identify eligible participants. This approach, when used with the appropriate sample weights, will result in nationally representative estimates of health. NHANES III originally was designed as 2 nationally representative phases: phase 1 included 1988 to 1990 and phase 2 included 1991 to 1994.

During NHANES III, adults aged 20 to 74 years were eligible for the original ultrasound examination of the gallbladder.<sup>15</sup> In 2011, data were released after a re-evaluation of the ultrasound images to assess HS using a standardized protocol and algorithm. Nearly 88% of the adults age 20 to 74 years in NHANES III phase 2 had ultrasound findings available to assess for the presence of HS. For genetic analyses, we restricted the sample to NHANES phase 2 participants (1991–1994) who had both genetic and ultrasound data available (Figure 1). Detailed information on the survey, its ultrasound component, and public and restricted data sets are described elsewhere.<sup>14,16</sup>

### Outcome and Variable Definitions

Our primary outcome was fatty liver disease defined as the presence of any ultrasonographic evidence of HS.<sup>14</sup> HS, a component of NAFLD, was defined as the ultrasonographic presence of any fat infiltration as defined by a composite scoring system that took into account brightness of liver parenchyma, deep beam attenuation, brightness of intrahepatic vessels, and liver-to-kidney contrast. These components were scored using a protocol developed and validated at the Centers for Disease Control and Prevention using metabolic traits and other risk factors for fatty liver disease.<sup>14</sup> Compared with our recent report,<sup>17</sup> we used a more comprehensive definition of HS to maintain the largest sample size for more detailed candidate gene analysis. In secondary analyses, we examined the association of SNPs with ultrasound-defined steatosis in combination with high ALT levels, as a possible surrogate of steatohepatitis, and also with ALT level as a continuous variable and AST to ALT ratio. A high ALT level was defined as greater than 19 IU/L in women and greater than 30 IU/L in men.<sup>18</sup> We have used ALT level only in contrast to other liver enzyme levels because *PNPLA3* has not been associated with AST levels previously.<sup>10</sup> For clinical and metabolic variables (eg, hypertension, diabetes, metabolic syndrome), we used standard definitions at the time NHANES III was conducted.<sup>19,20</sup> Hypertension was defined as a systolic blood pressure of 140 mm Hg or greater or a diastolic blood pressure of 90 mm Hg or greater, taking blood pressure medicine, or ever having been told by a doctor that they have high blood pressure. Diabetes was defined as a morning sample fasting plasma glucose level of 126 mg/dL or greater, plasma glucose level after 2 hours of an oral glucose tolerance test of 200 mg/dL or greater, taking diabetes medications (insulin or pills), or ever having been told by a doctor that they have diabetes. Metabolic syndrome was defined by the presence of 3 or more of the following: presence of hypertension, diabetes, triglyceride levels 150 mg/dL or greater or taking cholesterol-lowering medications, low high-density lipoprotein cholesterol (HDL-c) (HDL-c < 40 mg/dL in men, and HDL-c < 50 mg/dL in women), and, finally, a waist circumference greater than 88 cm in women or greater than 102 cm in men. Race/ethnicity was self-reported as NHW, NHB, MA, and other.<sup>16</sup> Alcohol consumption was estimated by multiplying the number of drinking days over the past 12 months and the number of drinks, on average, on a drinking day, and dividing by 365. Never-drinkers replied no to the question: “In your entire life, have you had at least 12 drinks of any kind of alcoholic beverage?”<sup>17</sup>

## Study Population: Genetic Component of the Third National Health and Nutrition Examination Survey

During phase 2 of NHANES III, lymphocytes were frozen and later used to establish immortalized cell lines for DNA-related research. Genetic variants were measured in 7159 participants aged 12 years and older.<sup>21</sup> For this study, we restricted the sample to individuals ages 20 to 74 years (n = 5356) because they were eligible for the ultrasound examination in NHANES III. After excluding those individuals with missing data including age (n = 0), sex (n = 0), ultrasound (n = 119), alcohol consumption (n = 171), and those classified as other race (n = 262), the final analytic sample size was 4804 (Figure 1).

### Genotyping

Genotyping was performed (San Diego, CA) using the iPLEX Sequenom platform for rs738409 (*PNPLA3*), rs2228603 (*NCAN*), rs12137855 (*LYPLALI*), rs780094 (*GCKR*), and rs4240624 (*PPP1R3B*) reported in the largest meta-analysis of NAFLD to date.<sup>13</sup> All SNPs passed strict quality control filters as is standard for genotyping in NHANES.<sup>22</sup>

### Statistical Analyses

We used population-weighted parametric and nonparametric tests when appropriate for studying associations of baseline characteristics. We assessed each SNP in an additive genetic model, using ethnic-specific, weighted, logistic, and linear regression adjusted for age, age-squared, sex, and alcohol consumption (continuous) for association with outcomes. An additive genetic model is one in which the alleles at the fatty liver loci are coded as 0, 1, or 2 based on how many fatty liver increasing alleles are present at that locus in a particular individual. We estimated associations across the 3 ancestries using population-weighted models by including self-reported race/ ethnicity as a covariate. All estimates are weighted with the appropriate NHANES sample weight and findings were considered statistically significant at a *P* value of less than .05. Weighted models mean that in each model we applied a weight suggested by the Centers for Disease Control and Prevention to take into account the oversampling of minorities to provide a final unbiased and accurate estimate of effects for the population.

In each population, we had more than 80% power to detect an odds ratio (OR) of 1.15 when the allele frequency was as high as 0.53 (*PNPLA3*) and an OR of 2.05 when the frequency was as low as 0.01 (*GCKR*). To make our study comparable with other genetic studies and to improve the statistical power of the regression, we have inversely normally transformed (placed the values into a normal distribution using their ranks) ALT and AST/ALT ratio before analysis.

From the potential eligible analytic sample that included individuals with DNA and ultrasound data (n = 5237), 91.7% (or 4804) of participants were included in the final analysis. Compared with other participants aged 20 to 74 years who did not have ultrasound or DNA data, crude nominal comparisons showed that the analytic sample had a significantly larger proportion of MA (32% analyzed vs 26% nonanalyzed), fewer NHB (30% analyzed vs 38% nonanalyzed), and a slightly higher preponderance of women (57% analyzed vs 54% nonanalyzed). There were no significant differences in waist

circumference, fasting plasma glucose level, HDL-c, ALT level, prevalence of hypertension, or average alcohol consumption (data not shown).

## Results

### Characteristics of Individuals in the Analytic Sample

We analyzed data from 4804 individuals between 20 and 74 years with DNA and liver ultrasound (1825 NHW, 1442 NHB, and 1537 MA). The weighted prevalence of HS, a component of NAFLD, in the analytic sample was 37.3%; and of HS with a high ALT level was 18.8% (Supplementary Figure 1). Overall, 51.7% (standard error, 0.7%) of participants were women, with a mean age of 42.5 years (standard error, 0.5 y) and a mean body mass index of 26.9 kg/m<sup>2</sup> (standard error, 0.2 kg/m<sup>2</sup>). In the overall study population, 15.3% (standard error, 1.0%) had hypertension, 6.0% (standard error, 0.8%) had diabetes, and 15.5% (standard error, 1.0%) met the criteria for having metabolic syndrome. Among participants, the mean plasma glucose level was 94.0 mg/dL (standard error, 1.0 mg/dL), triglyceride level was 155.4 mg/dL (standard error, 3.6 mg/dL), and HDL level was 50.0 mg/dL (standard error, 0.5 mg/dL). Mean alcoholic drinks per week were 3.7 (standard error, 0.2), and the average systolic and diastolic blood pressures were 120.5 mm Hg (standard error, 0.5 mm Hg) and 74.7 mm Hg (standard error, 0.3 mm Hg), respectively (Table 1). The prevalence of HS, a component of NAFLD, was substantially higher among MAs compared with NHWs and NHBs. Comparing individuals with HS with those without fat in the liver, we found no difference in the prevalence of fatty liver by average alcohol consumption, impaired fasting glucose level, or sex, except for NHWs, in whom men had more fatty liver than women ( $P < .05$ ). Individuals with ultrasound-defined fatty liver were older and had higher levels of the following: body mass index, levels of fasting glucose, triglycerides, and cholesterol, and lower levels of HDL (except for NHBs; Table 1). Participants with HS also tended to have a higher prevalence of diabetes, hypertension, and the metabolic syndrome, as shown in Table 1.

### Allele Frequency Differences in the Analytic Sample

We found that the weighted allele frequency of the G allele of rs738409 at *PNPLA3* in the overall analytic sample was 25.4%, 6.2% for the T allele of rs2228603 (*NCAN*), 80.4% for the C allele of rs12137855 (*LYPLALI*), 38.7% for the T allele of rs780094 (*GCKR*), and 89.6% for the A allele of rs4240624 (*PPP1R3B*) (Supplementary Figure 2). The SNP G allele of rs738409 of *PNPLA3* was more prevalent in MAs (effect allele frequency [EAF], 0.54) compared with NHWs and NHBs (EAF, 0.25 and 0.14, respectively). The T allele of rs2228603 (*NCAN*) was present at very low frequencies in NHBs (EAF, 0.01); the A allele of rs12137855 (*LYPLALI*) had a similar distribution across ancestries (EAF, 0.80 – 0.88). The SNP T allele of rs780094 (*GCKR*) was more common in NHWs than in MAs and less in NHBs (EAF, 0.42 vs 0.35 vs 0.20, respectively). Finally, the A allele of rs4240624 (*PPP1R3B*) was more prevalent in NHWs than in NHBs and in MAs (EAF, 0.92 vs 0.81 vs 0.75, respectively).



### Effects on Hepatic Steatosis

The T allele of rs2228603 at *NCAN* and the A allele of rs4240624 at *PPP1R3B* were associated significantly with HS across ancestries (OR, 1.40;  $P = .04$ ; and OR, 1.28,  $P = .03$ , respectively) and also were associated with HS in NHWs (OR, 1.41;  $P = .04$ ; and OR, 1.46;  $P = .04$ , respectively) (Figure 2). The G allele at rs738409 in *PNPLA3* also statistically was associated with HS in MAs (OR, 1.25;  $P = .01$ ) (Figure 2).

We then examined the association of these 5 SNPs with ultrasound-defined HS in combination with high ALT levels and found that the G and T variants in or near rs738409 (*PNPLA3*) and rs780094 (*GCKR*) were associated significantly with increased HS with high ALT levels in the analytic sample overall (OR, 1.36;  $P = .01$ ; and OR, 1.30;  $P = .03$ , respectively) and in NHWs (OR, 1.39;  $P = .03$ ) (Figure 3).

### Effects on Alanine Aminotransferase and the Aspartate Aminotransferase/Alanine Aminotransferase Ratio

Finally, using continuous ALT level as the outcome variable, we found that increased levels of ALT were associated significantly with the G allele rs738409 in *PNPLA3* overall, in NHWs and MAs, but not in NHBs (Table 2). We also found an association between ALT level and the T allele of rs2228603 (*NCAN*) overall and in NHWs, and, only for MAs, we showed an association between ALT level and the A allele of rs4240624 (*PPP1R3B*) (Table 2).

We conducted 2 additional analyses using high ALT level, defined as greater than 30 U/L in men or greater than 19 U/L in women, and inverse normally transformed AST/ALT ratio (Supplementary Table 1). For high ALT levels we found significant associations with *PNPLA3*, *GCKR*, and *PPP1R3B* across all ethnicities and for NHWs for *PNPLA3* and *GCKR* (Supplementary Figure 3). For the AST/ALT ratio, we found significant associations with *NCAN* for NHBs, *LYPLAL1* for MAs, and *PPP1R3B* for all race/ethnicities and NHWs (Supplementary Table 1).

### Overall Effects

The ORs of these same Genetics of Obesity-related Liver Disease HS-increasing variants were relatively consistent across ancestries, 9 of the 15 possible associations were positive (3 were significant: near *PNPLA3*, *NCAN*, and *PPP1R3B* genes). No significant associations were seen of less than 1.0 (Figures 2 and 3).

### Discussion

HS was present in 37.3% (weighted value) (Supplementary Figure 1) of the analytic sample, and HS with increased ALT level was present in 18.8% (weighted value) (Supplementary Figure 2). Across ancestries, we found that the G allele of rs738409 at *PNPLA3* and the T allele of rs780094 at *GCKR* were associated with HS with high ALT levels (OR, 1.36;  $P = .01$ ; and OR, 1.30;  $P = .03$ , respectively); in addition, the T allele of rs2228603 at *NCAN* and the A allele of rs4240624 at *PPP1R3B* was associated significantly with HS across ancestries (OR, 1.40;  $P = .04$ ; and OR, 1.28;  $P = .03$ , respectively). In addition, we found

that ALT level was associated significantly with both the G allele of rs738409 at *PNPLA3* and the T allele of rs2228603 (*NCAN*), and high ALT level was associated significantly with *PNPLA3*, *GCKR*, and *PPP1R3B* across all races/ethnicities. Additional analyses of the AST/ALT ratio showed associations with *NCAN*, *LYPLAL1*, and *PPP1R3B* in different ethnicities. When we examined the race/ethnicity-specific associations, we found that some SNPs showed significant associations in particular phenotypes (for NHW, *PNPLA3*, *NCAN*, *GCKR*, and *PPP1R3B*; for MA, *PNPLA3*; and none for NHB).

Overall, the earlier-described results strongly reinforce that the G allele of rs738409 (*PNPLA3*), the T allele of rs780094 (*GCKR*), and the A allele of rs4240624 (*PPP1R3B*) are HS-increasing variants, as have been reported previously.<sup>13</sup> In agreement with other studies, we found that the G allele of *PNPLA3* was associated significantly with liver steatosis, especially in MAs.<sup>11,23</sup> The effect of the G allele on HS was similar across all races/ethnicities, however, the association was only statistically significant for MAs. This result is due in part to the high EAF of this variant in the MA population, which increases our power to find the association significant. Indeed, as has been reported previously, the EAF of this variant is greatest in MA, then NHW, and then NHB, which parallels the prevalence of steatosis in these populations.<sup>11,23</sup> However, this pattern is not seen with all SNPs associated with HS. For example, in our study, the frequency of the HS-increasing allele of *PPP1R3B* and *GCKR* was highest in NHWs and for *NCAN* in MAs. Thus, risk allele frequency is not the only explanation for the differences in the prevalence of NAFLD across ancestries.

We observed that the A allele of rs4240624 (*PPP1R3B*) also increased the odds of having HS using ultrasound. In computerized tomography, increased fat content decreases liver density, whereas glycogen increases it.<sup>24</sup> The previous association of the A allele of rs4240624 with decreased liver attenuation could reflect increased fat or decreased glycogen content. By ultrasonography, both increased fat and increased glycogen are hyperechoic.<sup>25</sup> Therefore, our finding that the A allele associates with increased echogenicity suggests that it does indeed increase fat accumulation in the liver. Interestingly, this allele, unlike the other 4 HS-associated alleles, was not found to have an effect on NASH/fibrosis.<sup>13</sup> Therefore, it is possible that there are different mechanisms by which fat accumulates in the liver, only some of which progress to NASH/fibrosis. In this case, the A allele of rs4240624 genetically could differentiate a more benign steatosis from steatosis that may progress to NASH/fibrosis.<sup>13</sup>

We observed higher ORs for HS-increasing alleles at loci affecting HS with increased ALT levels vs HS alone across ancestries, in particular for *PNPLA3* and *GCKR* in NHW and a similar, but not significant, trend in *NCAN* and *PPP1R3B* in NHW. These findings may be owing to this subset of individuals having more steatosis, being more prone to hepatic injury that releases ALT into the serum, or having advanced liver disease (NASH/fibrosis). Indeed, a recent meta-analysis showed that *PNPLA3* rs738409 GG homozygotes had a 3.2-fold greater risk of higher necroinflammatory scores and a 3.2-fold greater risk of developing fibrosis compared with wild-type CC homozygotes.<sup>26</sup> In addition, NASH was more common in GG homozygotes compared with wild-type CC homozygotes (OR, 3.49; 95% CI, 1.86 – 6.55).<sup>25</sup> Further, *GCKR* recently was associated with inflammatory pathways,<sup>27</sup> and thus may help to identify individuals prone to developing NASH. We confirmed



previous work that associated the G allele of rs738409 (*PNPLA3*) with higher ALT levels in individuals of European ancestry,<sup>10,28</sup> but we also have found it to be associated with ALT level in MAs.<sup>11</sup> Some variants were associated with increased ALT level whereas others were associated with the AST/ALT ratio with variants at *PPP1R3B* associated with both phenotypes. This may be owing to these variants having a differential effect on steatosis, being more prone to hepatic injury that releases ALT into the serum, or having advanced liver disease (ALT)<sup>29</sup> vs fibrosis (AST/ALT ratio),<sup>30</sup> which will require study in further samples to confirm.

The strength of our study was that it reports an association of genetic variants in or near *PNPLA3*, *GCKR*, and *PPP1R3B* with ultrasound-defined HS in a large nationally representative population of NHWs. Given the population-based nature of the survey sample, these findings suggest that these results are likely to hold true across many European ancestry individuals living in the United States compared with more biased chosen samples.

The study also provided nationally representative estimates of allele frequencies and effects across 3 race/ethnicities as well as combined across the 3 groups. Although not having biopsy-proven HS, our phenotypes indirectly measure hepatic fat. All measures of liver steatosis were associated with known risk factors for the development of fatty liver disease and ultrasound has been shown to be an accurate and reliable method for detecting histologically defined steatosis.<sup>14,31</sup> Our study also reports an association of the G allele of rs738409 *PNPLA3* variant with increased ALT level in MAs.<sup>10,28</sup> We also show that the A allele of rs4240624 (*PPP1R3B*) is associated with HS using ultrasound, which makes the previously possible alternative hypothesis that the G allele at this locus increases glycogen unlikely.<sup>13</sup> Finally, we characterized the effect of allele frequencies and ORs of 5 genetic variants in several race/ethnic groups, which expand previous research that associated these variants with liver steatosis in individuals of European ancestry.

Insulin resistance is considered a key factor in the development and progression of NAFLD and NASH. In our study, the association with impaired fasting glucose trended toward significance but did not reach it. This result is consistent with HS being associated with impaired fasting glucose overall, but we likely did not have enough power to reach statistical significance with the current sample size. Although there likely are variants that promote both increased fatty liver and increased impaired fasting glucose to cause the overall association seen between these traits, in well-powered genetic studies of glucose phenotypes<sup>32,33</sup> and our prior genome-wide association studies,<sup>13</sup> the HS-promoting allele at *GCKR* and *PPP1R3B* was associated with decreased (not increased) impaired fasting glucose. This suggests that some variants promote HS via a different mechanism than by increasing impaired fasting glucose.

Our study had several limitations. First, it did not include individuals of all possible ancestries present in the United States (eg, Asian ancestry) or the institutionalized population (military, nursing home residents); second, it may not have sufficient power to detect allele effects in all of the ancestries examined; and, third, because of the phenotypic characterization, it cannot assess the effects of these variants on NASH/ fibrosis, which

requires liver biopsy samples that ethically cannot be obtained from healthy individuals in a population-based general health survey. Finally, our current phenotype definition may overestimate the true prevalence of HS and presumed NASH in the United States. We opted to collapse 3 fatty liver categories—mild, moderate, and severe—into one category to increase the power to detect genetic associations. To estimate the true prevalence of ultrasound-defined HS in the US population, see Lazo et al (American Journal of Epidemiology, accepted 2012).

In conclusion, we have shown that ultrasound-based HS is associated with 4 SNPs, rs738409 (*PNPLA3*), rs2228603 (*NCAN*), rs780094 (*GCKR*), and rs4240624 (*PPP1R3B*) in NHW from a population sample representative of the United States. In addition, we found that rs738409 (*PNPLA3*) was associated significantly with MA, and we found no significant associations with ultrasound-defined HS in this US representative sample of NHB. Our findings show that there are specific genetic influences to the development of fatty liver disease across ancestries. This suggests that in addition to environmental variables, genetic influences can affect health disparities in the development of fatty liver disease among ancestry groups.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### *Funding*

Supported by the American Diabetes Association Mentor-Based Fellowship Program 7-07-MN-08 (R.H., M.L., F.L.B.), Doris Duke Medical Foundation (E.K.S.), National Institutes of Health grants K23DK080145-01 (E.K.S.), R01DK075787 (J.N.H.), 5R01DK075681 (I.B.B.), and R01 DK083393 (J.M.C., R.H., M.L., F.L.B., I.R.H., S.B., W.H.L.K.).

## Abbreviations used in this paper

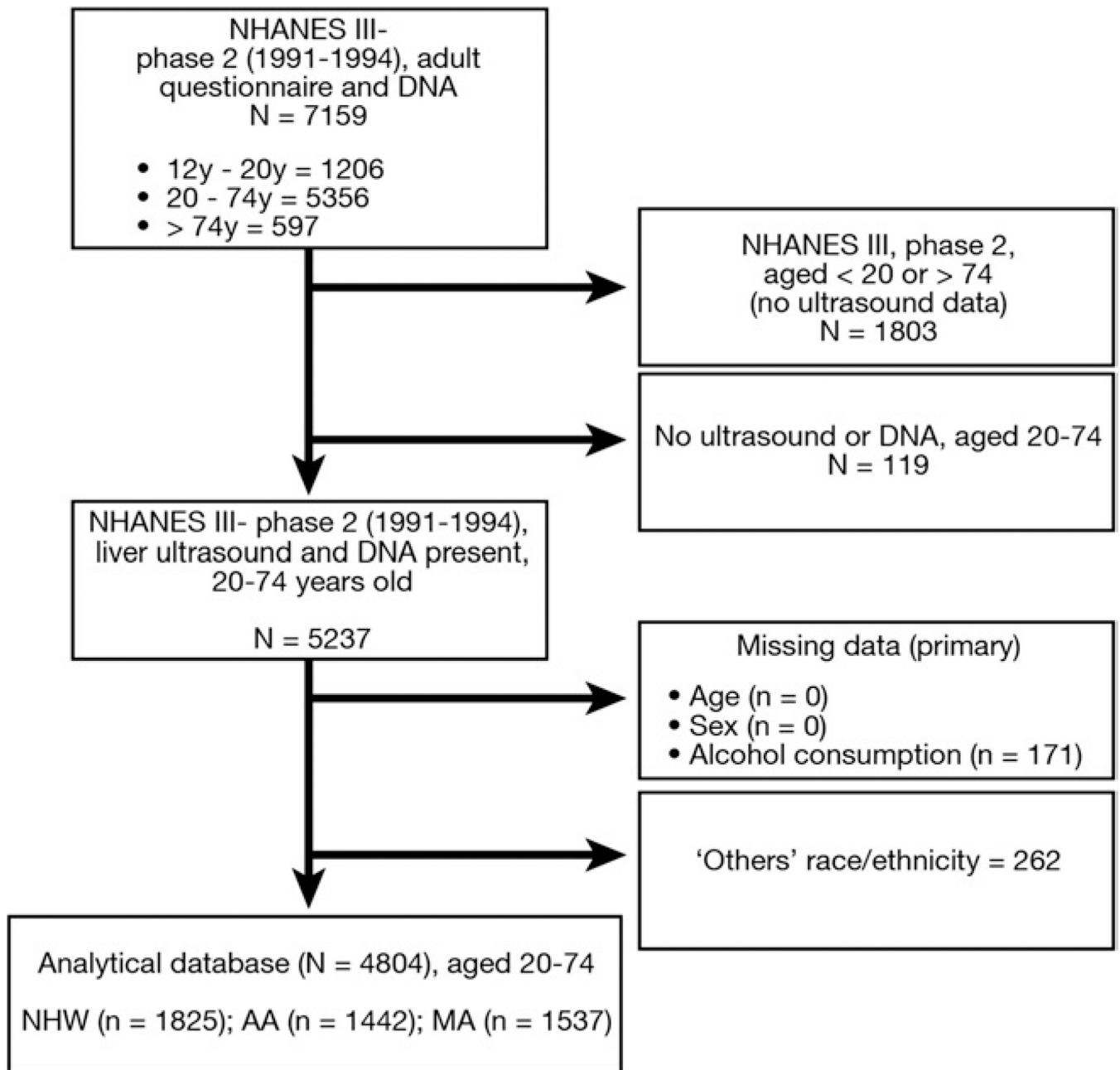
<b>ALT</b>	alanine aminotransferase
<b>AST</b>	aspartate aminotransferase
<b>EAF</b>	effect allele frequency
<b>HDL-c</b>	high-density lipoprotein cholesterol
<b>HS</b>	hepatic steatosis
<b>GCKR</b>	glucokinase regulatory protein
<b>LYPLALI</b>	lysophospholipase-like 1
<b>MA</b>	Mexican American
<b>NAFLD</b>	nonalcoholic fatty liver disease
<b>NASH</b>	nonalcoholic steatohepatitis
<b>NCAN</b>	neurocan

<b>NHANES III</b>	Third National Health and Nutrition Examination Survey
<b>NHB</b>	non-Hispanic black
<b>NHW</b>	non-Hispanic white
<b>OR</b>	odds ratios
<b>PNPLA3</b>	patatin-like phospholipase domain-containing protein 3
<b>PPP1R3B</b>	protein phosphatase 1, regulatory subunit 3b
<b>SNP</b>	single-nucleotide polymorphisms

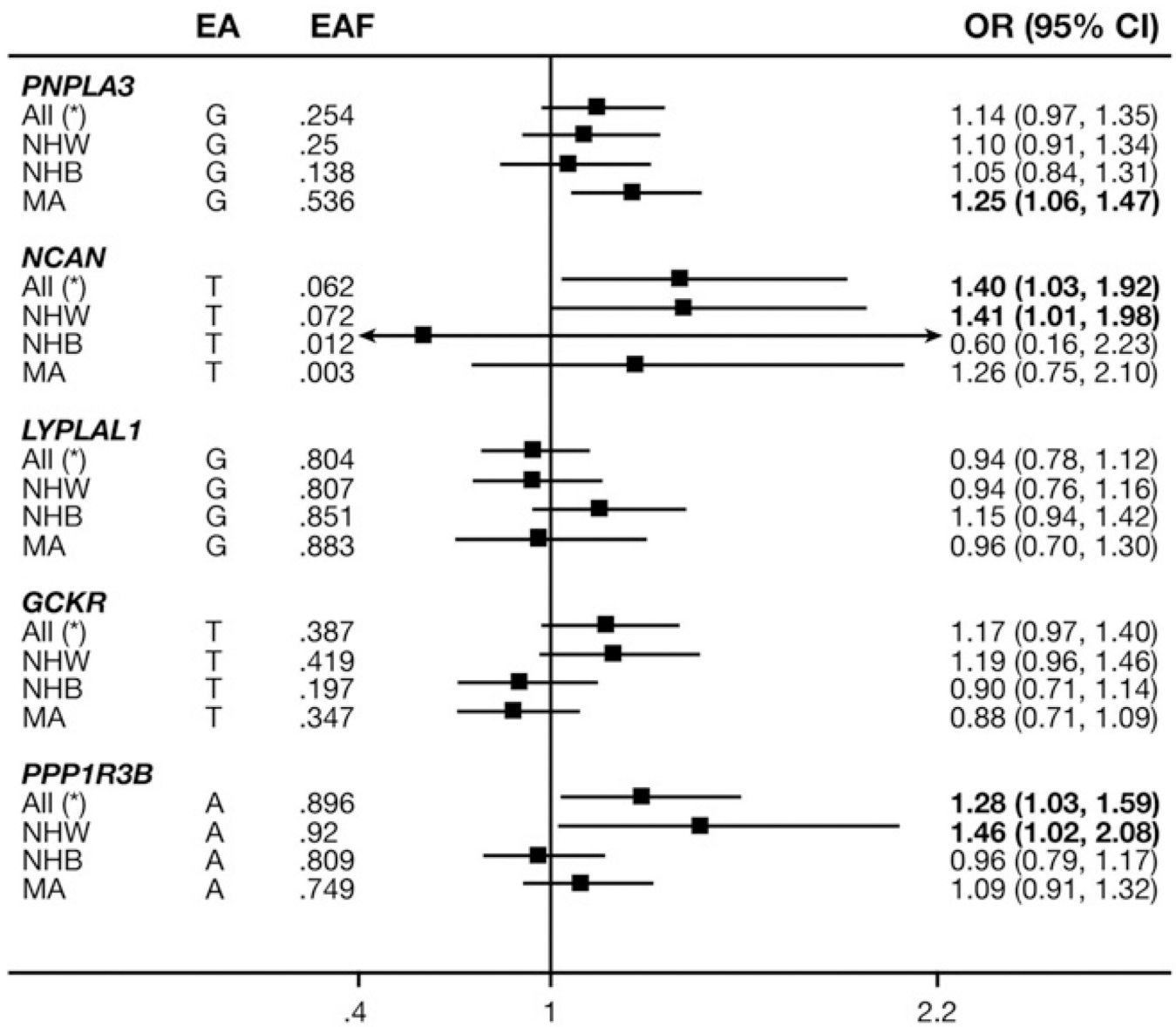
## References

1. Flegal KM, Carroll MD, Ogden CL, et al. Prevalence and trends in obesity among US adults, 1999–2008. *JAMA*. 2010; 303:235–241. [PubMed: 20071471]
2. Adams LA, Waters OR, Knudman MW, et al. NAFLD as a risk factor for the development of diabetes and the metabolic syndrome: an eleven-year follow-up study. *Am J Gastroenterol*. 2009; 104:861–867. [PubMed: 19293782]
3. American Gastroenterological Association medical position statement: nonalcoholic fatty liver disease. *Gastroenterology*. 2002; 123:1702–1704. [PubMed: 12404244]
4. Angulo P. Nonalcoholic fatty liver disease. *N Engl J Med*. 2002; 346:1221–1231. [PubMed: 11961152]
5. Clark JM, Diehl AM. Nonalcoholic fatty liver disease: an underrecognized cause of cryptogenic cirrhosis. *JAMA*. 2003; 289:3000–3004. [PubMed: 12799409]
6. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008; 40:955–962. [PubMed: 18587394]
7. Lindgren CM, Heid IM, Randall JC, et al. Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet*. 2009; 5:e1000508. [PubMed: 19557161]
8. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007; 39:596–604. [PubMed: 17435756]
9. Willer CJ, Speliotes EK, Loos RJ, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*. 2009; 41:25–34. [PubMed: 19079261]
10. Yuan X, Waterworth D, Perry JR, et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am J Hum Genet*. 2008; 83:520–528. [PubMed: 18940312]
11. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet*. 2008; 40:1461–1465. [PubMed: 18820647]
12. Chalasani N, Guo X, Loomba R, et al. Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. *Gastroenterology*. 2010; 139:1567–1576. [PubMed: 20708005]
13. Speliotes EK, Yerges-Armstrong LM, Wu J, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet*. 2011; 7:e1001324. [PubMed: 21423719]
14. National Center for Health Statistics, Centers for Disease Control and Prevention. [Accessed: July 24, 2011] Hepatic steatosis. Ultrasound images assessment procedures manual. Available from: [http://www.cdc.gov/nchs/data/nhanes/nhanes3/Hepatic\\_Steatosis\\_Ultrasound\\_Procedures\\_Manual.pdf](http://www.cdc.gov/nchs/data/nhanes/nhanes3/Hepatic_Steatosis_Ultrasound_Procedures_Manual.pdf)
15. Everhart JE, Khare M, Hill M, et al. Prevalence and ethnic differences in gallbladder disease in the United States. *Gastroenterology*. 1999; 117:632–639. [PubMed: 10464139]

16. National Center for Health Statistics, Centers for Disease Control and Prevention. [Accessed: July 24, 2011] Third National Health and Nutrition Examination Survey: questionnaire, databases and related documentation, 2011. Available from: <http://www.cdc.gov/nchs/nhanes/nh3data.htm>
17. Lazo M, Hernaez R, Bonekamp S, et al. Non-alcoholic fatty liver disease and mortality among US adults: prospective cohort study. *BMJ*. 2011; 343:d6891. [PubMed: 22102439]
18. Prati D, Taioli E, Zanella A, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. *Ann Intern Med*. 2002; 137:1–10. [PubMed: 12093239]
19. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the Third National Health and Nutrition Examination Survey. *JAMA*. 2002; 287:356–359. [PubMed: 11790215]
20. Harris MI, Flegal KM, Cowie CC, et al. Prevalence of diabetes impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988–1994. *Diabetes Care*. 1998; 21:518–524. [PubMed: 9571335]
21. National Center for Health Statistics, Centers for Disease Control and Prevention. [Accessed: July 24, 2011] National Health and Nutrition Examination Survey genetic data. Available from: [http://www.cdc.gov/nchs/nhanes/genetics/collection\\_dna.htm](http://www.cdc.gov/nchs/nhanes/genetics/collection_dna.htm)
22. National Center for Health Statistics, Centers for Disease Control and Prevention. [Accessed: July 24, 2011] National Health and Nutrition Examination Survey (NHANES) DNA. Quality Control Protocol 6/7/2010. Available from: [http://www.cdc.gov/nchs/nhanes/genetics/collection\\_dna.htm](http://www.cdc.gov/nchs/nhanes/genetics/collection_dna.htm)
23. Wagenknecht LE, Palmer ND, Bowden DW, et al. Association of PNPLA3 with non-alcoholic fatty liver disease in a minority cohort: the Insulin Resistance Atherosclerosis Family Study. *Liver Int*. 2011; 31:412–416. [PubMed: 21281435]
24. Dwyer A, Doppman JL, Adams AJ, et al. Influence of glycogen on liver density: computed tomography from a metabolic perspective. *J Comput Assist Tomogr*. 1983; 7:70–73. [PubMed: 6131081]
25. Pozzato C, Botta A, Melgara C, et al. Sonographic findings in type I glycogen storage disease. *J Clin Ultrasound*. 2001; 29:456–461. [PubMed: 11745852]
26. Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. *Hepatology*. 2011; 53:1883–1894. [PubMed: 21381068]
27. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*. 2011; 123:731–738. [PubMed: 21300955]
28. Chambers JC, Zhang W, Sehmi J, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet*. 2011; 43:1131–1138. [PubMed: 22001757]
29. Minervini MI, Ruppert K, Fontes P, et al. Liver biopsy findings from healthy potential living liver donors: reasons for disqualification, silent diseases and correlation with liver injury tests. *J Hepatol*. 2009; 50:501–510. [PubMed: 19155086]
30. McPherson S, Anstee QM, Henderson E, et al. Are simple noninvasive scoring systems for fibrosis reliable in patients with NAFLD and normal ALT levels? *Eur J Gastroenterol Hepatol*. 2013; 25:652–658. [PubMed: 23325287]
31. Hernaez R, Lazo M, Bonekamp S, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. *Hepatology*. 2011; 54:1082–1090. [PubMed: 21618575]
32. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010; 42:105–116. [PubMed: 20081858]
33. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007; 316:1331–1336. [PubMed: 17463246]

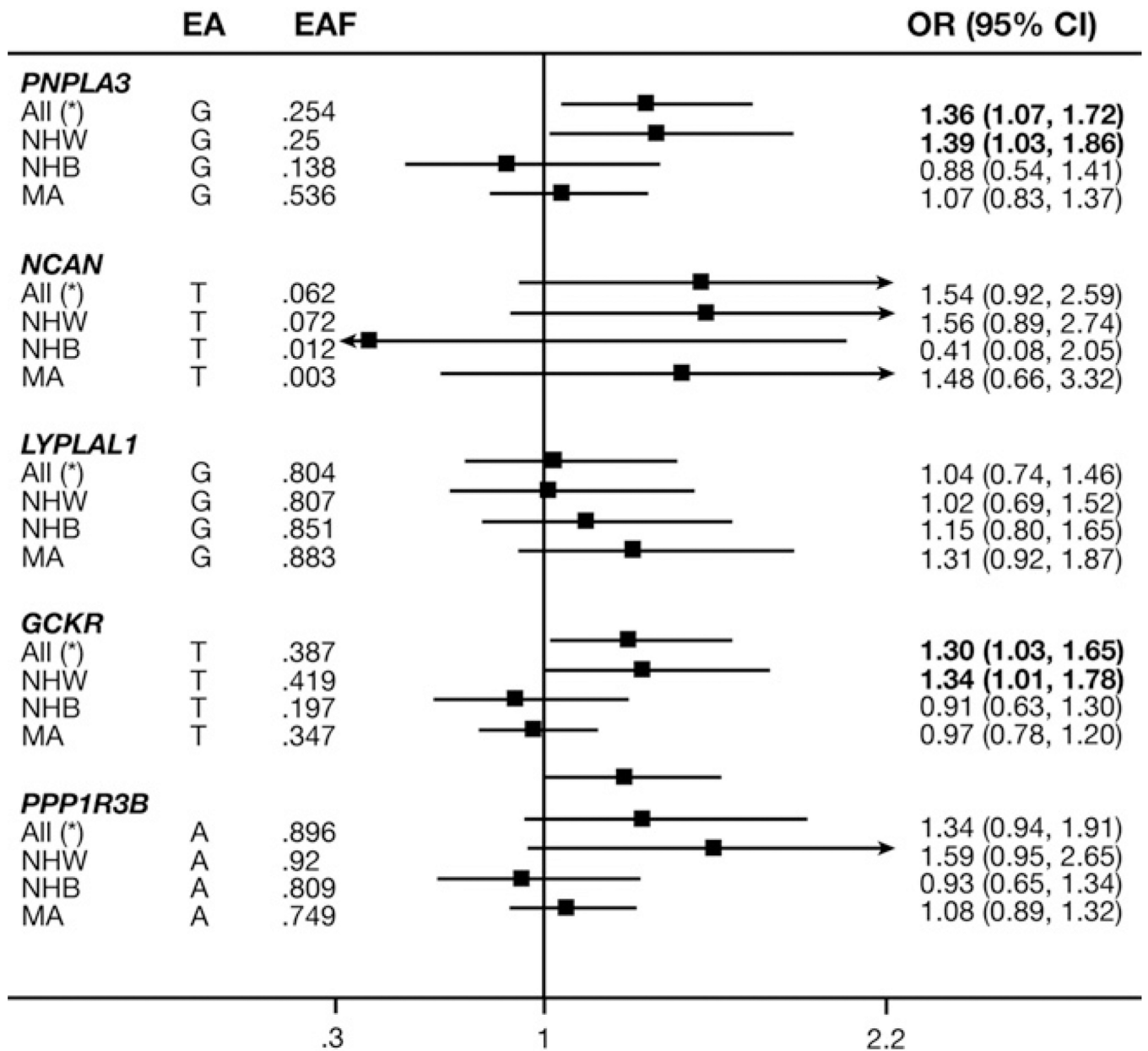


**Figure 1.**  
Flow chart of the National Health and Nutrition Examination Survey, phase 2 (1991-1994).



**Figure 2.** Association between HS and different SNPs, stratified by ethnicity in the NHANES III (1991–1994), population-weighted, additive model, adjusted for age, age squared, sex, and alcohol consumption.





**Figure 3.** Association between HS with high ALT level and different SNPs, stratified by ethnicity in the NHANES III (1991–1994), population-weighted, additive model, adjusted for age, age squared, sex, and alcohol consumption.

Table 1

Baseline Characteristics of 4804 NHANES III, Phase 2 Participants (1991–1994), Aged 20 to 74 With Ultrasound and DNA Data, HS vs No HS, Weighted

Category	NHW				NHB				MA			
	All (n = 1825)	No HS (n = 1140)	HS (n = 685)	P value	All (n = 1442)	No HS (n = 962)	HS (n = 480)	P value	All (n = 1537)	No HS (n = 762)	HS (n = 775)	P value
Women, %	51.7 (0.7)	54.1 (1.7)	46.3 (1.9)	.03	56.2 (1.5)	56.2 (1.5)	56.2 (3.3)	1.00	48.8 (1.2)	48.9 (2.0)	48.6 (1.1)	.89
Age, y	42.5 (0.5)	41.9 (0.6)	45.7 (1.1)	.00	40.1 (0.6)	38.8 (0.7)	42.6 (0.9)	<.01	36.9 (0.6)	34.9 (0.8)	39.1 (0.9)	<.01
EtOH drinks, weekly	3.7 (0.2)	3.7 (0.2)	3.8 (0.6)	.83	3.9 (0.3)	3.5 (0.4)	4.6 (0.6)	.17	3.5 (0.3)	3.3 (0.3)	3.7 (0.4)	.29
WC, cm	92.2 (0.4)	92.1 (0.4)	97.9 (1.2)	<.01	93.2 (0.5)	91.2 (0.5)	97.1 (1.0)	<.01	92.9 (0.2)	89.2 (0.6)	96.8 (0.7)	<.01
BMI, kg/m <sup>2</sup>	26.9 (0.2)	25.4 (0.2)	28.8 (0.5)	<.01	28.2 (0.3)	27.5 (0.2)	29.7 (0.5)	<.01	27.8 (0.1)	26.3 (0.2)	29.4 (0.2)	<.01
BMI, <25 kg/m <sup>2</sup>	22.1 (0.1)	22.2 (0.1)	21.8 (0.1)	.02	22.0 (0.1)	22.2 (0.1)	21.6 (0.2)	.03	22.5 (0.1)	22.5 (0.1)	22.5 (0.1)	.93
BMI, 25 to <30 kg/m <sup>2</sup>	27.2 (0.1)	26.7 (0.1)	27.4 (0.1)	<.01	27.3 (0.1)	27.3 (0.1)	27.3 (0.1)	.80	27.3 (0.1)	27.1 (0.1)	27.6 (0.1)	<.01
BMI, 30 kg/m <sup>2</sup>	34.8 (0.2)	33.6 (0.3)	35.4 (0.5)	<.01	35.5 (0.4)	34.6 (0.4)	36.6 (0.7)	<.01	34.2 (0.2)	33.6 (0.5)	34.6 (0.2)	.07
FFG, mg/dL <sup>b</sup>	94.0 (1.0)	93.2 (1.1)	95.1 (1.7)	.05	97.5 (2.0)	94.2 (2.5)	103.2 (3.2)	.04	97.2 (1.3)	93.0 (0.9)	101.8 (2.5)	<.01
DM, %	6.0 (0.8)	5.2 (1.0)	8.8 (1.7)	.01	10.4 (1.7)	7.3 (2.3)	15.3 (2.8)	.07	9.4 (1.3)	4.8 (0.9)	13.7 (2.1)	<.01
IFG, % <sup>b</sup>	11.5 (1.4)	11.1 (1.6)	14.8 (2.4)	.09	12.5 (1.7)	11.6 (2.0)	14.0 (3.5)	.58	14.4 (2.7)	14.7 (3.2)	14.9 (2.4)	.69
Triglyceride level, mg/dL	155.4 (3.6)	157.6 (4.1)	201.7 (9.9)	<.01	129.8 (4.1)	120.5 (4.3)	144.8 (6.5)	<.01	177.8 (5.1)	151.8 (5.8)	201.4 (8.4)	<.01
Total cholesterol level, mg/dL	201.5 (1.0)	202.3 (1.2)	206.3 (1.9)	.02	197.6 (1.5)	194.0 (1.5)	204.8 (2.3)	<.01	198.0 (1.7)	194.1 (1.9)	202.3 (2.1)	<.01
HDL-C, mg/dL	50.0 (0.5)	49.6 (0.7)	46.0 (0.8)	<.01	54.0 (0.6)	54.5 (0.6)	53.0 (0.9)	.08	47.6 (0.5)	49.8 (0.7)	45.8 (0.3)	<.01
SBP, mm Hg	120.5 (0.5)	118.0 (0.5)	123.9 (0.8)	<.01	123.6 (0.7)	122.5 (0.7)	125.7 (0.9)	<.01	118.6 (0.7)	116.3 (0.9)	121.1 (1.0)	<.01
DBP, mm Hg	74.7 (0.3)	74.6 (0.3)	76.7 (0.5)	<.01	76.1 (0.5)	75.5 (0.5)	77.2 (0.6)	<.01	73.2 (0.6)	71.0 (0.8)	75.7 (0.6)	<.01
Hypertensive, %	15.3 (0.95)	14.5 (1.1)	21.0 (2.4)	<.01	22.0 (1.7)	18.8 (1.8)	28.2 (2.5)	.02	11.5 (1.1)	8.6 (1.6)	14.8 (1.8)	.02
Metabolic syndrome, % <sup>c</sup>	15.5 (1.0)	15.8 (1.3)	30.6 (3.4)	<.01	12.0 (1.5)	7.8 (1.6)	19.3 (2.5)	<.01	19.2 (1.8)	11.1 (1.4)	28.2 (3.2)	<.01
ALT level, IU/L	19.5 (0.6)	19.1 (0.7)	22.7 (0.7)	<.01	18.4 (0.5)	17.5 (0.5)	20.3 (0.8)	<.01	26.0 (0.9)	20.3 (1.0)	32.3 (0.9)	<.01

NOTE. Results are mean (±SD), median (interquartile range), or percentage (n = number of individuals). Bolded entries represent P values <0.05.

BMI, body mass index; DBP, diastolic blood pressure; DM, diabetes mellitus; EtOH, alcohol containing; FFG, fasting plasma glucose; IFG, impaired fasting glucose (fasting plasma glucose between 100 and 125 mg/dL); SBP, systolic blood pressure; WC, waist circumference.

<sup>a</sup>All = NHW + NHB + MA.

<sup>b</sup>Morning sample: n = 4797; NHW, 1824; NHB, 1440; and MA, 1533.

<sup>c</sup>Metabolic syndrome based on ATP III definitions, see “Outcome and Variable Definitions” section for further details.

Table 2

Multivariate Analyses for Each of the Genetics of Obesity-Related Liver Disease Consortium SNPs in 4804 NHANES III Participants for Effects on ALT, Weighted

SNP ID	Nearest	EA	AII <sup>a</sup>		NHW		NHB		MA	
			Effect (SE)	P value	Effect (SE)	P value	Effect (SE)	P value	Effect (SE)	P value
rs738409	PNPLA3	G	0.13 (0.03)	<.01	0.13 (0.03)	<.01	-0.04 (0.06)	.44	0.21 (0.05)	<.01
rs2228603	NCAN	T	0.18 (0.04)	<.01	0.18 (0.04)	<.01	0.24 (0.35)	.50	0.08 (0.14)	.57
rs12137855	LYPLAL1	C	0.06 (0.06)	.32	0.06 (0.07)	.35	-0.00 (0.05)	.98	-0.02 (0.05)	.69
rs780094	GCKR	T	0.03 (0.03)	.20	0.03 (0.03)	.31	0.05 (0.04)	.27	0.01 (0.06)	.92
rs4240624	PPP1R3B	A	0.03 (0.03)	.38	0.06 (0.05)	.25	-0.07 (0.06)	.23	0.06 (0.03)	<b>.04</b>

PNPLA3: HUGO Gene Nomenclature Committee (HGNC), 18590; NCAN: HGNC, 2465; LYPLAL1: HGNC, 20440; GCKR: HGNC, 4196; PPP1R3B: HGNC, 14942.

Bolded entries signify *P* values .05.

<sup>a</sup> AII = NHW + NHB + MA.