Genetic signatures in choline and 1-carbon metabolism are associated with the severity of hepatic steatosis

Karen D. Corbin,[*,†] Manal F. Abdelmalek,[§] Melanie D. Spencer,[*,‡] Kerry-Ann da Costa,[†] Joseph A. Galanko,[‡] Wei Sha,[#] Ayako Suzuki,[§] Cynthia D. Guy,[‖] Diana M. Cardona,[‖] Alfonso Torquati,[¶] Anna Mae Diehl,[§] and Steven H. Zeisel[*,†,1]

[*]University of North Carolina at Chapel Hill Nutrition Research Institute, Kannapolis, North Carolina, USA; [†]Department of Nutrition, School of Public Health, and [‡]Division of Gastroenterology and Hepatology, Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; [§]Division of Gastroenterology, Department of Medicine, [‖]Department of Pathology, and [¶]Department of Surgery, Duke University Medical Center, Durham, North Carolina, USA; and [#]Bioinformatics Services Division, University of North Carolina at Charlotte, Kannapolis, North Carolina, USA

ABSTRACT **Choline metabolism is important for very low-density lipoprotein secretion, making this nutritional pathway an important contributor to hepatic lipid balance. The purpose of this study was to assess whether the cumulative effects of multiple single nucleotide polymorphisms (SNPs) across genes of choline/1-carbon metabolism and functionally related pathways increase susceptibility to developing hepatic steatosis. In biopsy-characterized cases of nonalcoholic fatty liver disease and controls, we assessed 260 SNPs across 21 genes in choline/1-carbon metabolism. When SNPs were examined individually, using logistic regression, we only identified a single SNP (*PNPLA3* rs738409) that was significantly associated with severity of hepatic steatosis after adjusting for confounders and multiple comparisons ($P=0.02$). However, when groupings of SNPs in similar metabolic pathways were defined using unsupervised hierarchical clustering, we identified groups of subjects with shared SNP signatures that were significantly correlated with steatosis burden ($P=0.0002$). The lowest and highest steatosis clusters could also be differentiated by ethnicity. However, unique SNP patterns defined steatosis burden irrespective of ethnicity. Our results suggest that analysis of SNP patterns in genes of choline/1-carbon metabolism may be useful for prediction of severity of steatosis in specific subsets of people, and the metabolic inefficiencies caused by these SNPs should be examined further.—Corbin, K. D., Abdelmalek, M. F., Spencer, M. D., da Costa, K.-A., Galanko, J. A., Sha, W., Suzuki, A., Guy, C. D., Cardona, D. M., Torquati, A., Diehl, A. M., Zeisel, S. H. Genetic signatures in choline and 1-carbon metabolism are associated with the severity of hepatic steatosis. *FASEB J.* 27, 1674–1689 (2013). www.fasebj.org**

Key Words: single nucleotide polymorphisms · nutrition · nonalcoholic fatty liver disease · metabolic syndrome · obesity

Abbreviations: ABCB4, ATP-binding cassette, subfamily B (MDR/TAP), member 4; ADIPOQ, adiponectin; ALT, alanine aminotransferase; APOC3, apolipoprotein C-III; AST, aspartate aminotransferase; BHMT, betaine-homocysteine methyltransferase; BMI, body mass index; CBS, cystathionine-β synthase; CHDH, choline dehydrogenase; CHKA, choline kinase α; CHKB, choline kinase β; FADS2, fatty acid desaturase 2; FDR, false discovery rate; GWAS, genome wide association study; HCC, hepatocellular carcinoma; Hcy, homocysteine; HDL, high-density lipoprotein; LDL, low-density lipoprotein; MTHFD1, methylenetetrahydrofolate desaturase 1; MTHFR, 5,10-methylenetetrahydrofolate reductase; MTR, 5-methyltetrahydrofolate-homocysteine methyltransferase (methionine synthase); NAFLD, nonalcoholic fatty liver disease; PCYT1A, phosphate cytidylyltransferase 1, choline, α; PCYT1B, phosphate cytidylyltransferase 1, choline, β; PEMT, phosphatidylethanolamine-*N*-methyltransferase; PNPLA3, patatin-like phospholipase domain containing 3; PPARG, peroxisome proliferator-activated receptor γ; PtdCho, phosphatidylcholine; SCD, stearoyl-CoA desaturase; SHMT1, serine hydroxymethyltransferase 1; SLC44A1, solute carrier family 44, member 1; SNP, single-nucleotide polymorphism; STAT3, signal transducer and activator of transcription 3; VLDL, very low density lipoprotein

NONALCOHOLIC FATTY LIVER disease (NAFLD) is a common disorder, affecting ~30% of people in the general population and up to 96% of obese individuals. It encompasses a spectrum from steatosis to advanced liver injury (1). The prevalence of steatosis has increased in parallel to obesity and metabolic syndrome, and it is often a concurrent phenotype. Single nucleotide polymorphisms (SNPs) in several genes have been individually associated with steatosis, but these associations only explain a small percentage of the risk and often fail to be replicated (2). It is unclear whether steatosis is a cause, predisposing factor, or consequence of metabolic disease, although many regard it as the

[1] Correspondence: UNC Chapel Hill Nutrition Research Institute, 500 Laureate Way, Rm. 2218, Kannapolis, NC 28081, USA. E-mail: steven_zeisel@unc.edu

doi: 10.1096/fj.12-219097

This article includes supplemental data. Please visit *http://www.fasebj.org* to obtain this information.

0892-6638/13/0027-1674 © FASEB

hepatic manifestation of metabolic syndrome. In many cases, NAFLD maintains a benign course, but some individuals progress to fibrosis, cirrhosis, liver failure, and cancer (1). Increasing evidence indicates that hepatocellular carcinoma (HCC) may be partially attributable to metabolic disease, suggesting common pathogenic mechanisms (3). Improving our understanding of the factors that modify the risk of hepatic steatosis is of great value, from both metabolic and disease progression perspectives.

Steatosis occurs when there is an imbalance between the rate of formation of triglycerides in liver and the capacity to remove those triglycerides by oxidation or by secretion as lipoproteins. Many factors contribute to this imbalance (4). Choline metabolism is one key mediator of hepatic lipid homeostasis and intersects with multiple pathways that also mediate deposition of lipids in the liver (**Fig. 1**). Perturbations in choline/1-carbon metabolism that decrease the synthesis of phosphatidylcholine (PtdCho) reduce capacity to secrete very low density lipoprotein (VLDL) and thereby impair secretion of triglycerides from liver (5). Humans eating choline-deficient diets, or fed parenterally with solutions low in choline, develop fatty liver, and specific

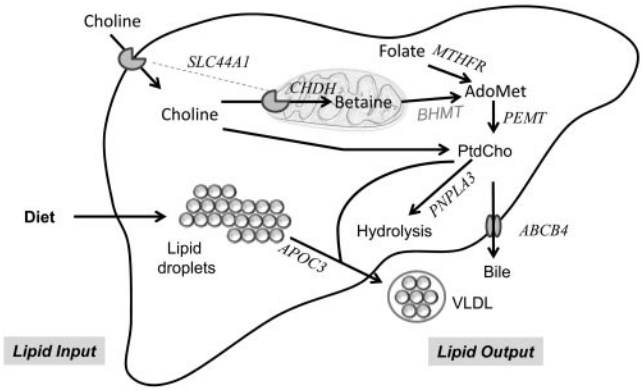

Figure 1. Genes in choline/1-carbon metabolism and related pathways influence hepatic lipid balance. One factor that influences accumulation of fat in the liver is excess calorie intake, as fatty acids are taken up by the liver and glucose is converted to lipids in the liver. At the same time, lipids are secreted from the liver, mainly as very low density lipoproteins (VLDLs). The net balance between lipid inputs and outputs determines whether hepatic steatosis develops. Secretion of VLDL requires synthesis of a lipid envelope containing apoproteins and phosphatidylcholine (PtdCho). *APOC3* is needed for apolipoprotein synthesis. PtdCho is formed from choline, or from methyl-groups in *S*-adenosylmethionine (AdoMet) that can be derived from choline or from 1-carbon pool (*via* methyl-tetrahydrofolate; *MTHFR* encodes for the enzyme that forms 5-methyl-terahydrofolate). Choline transport into the hepatocyte and then into the mitochondria is mediated by the gene product of *SLC44A1*. *CHDH*, *BHMT*, and *PEMT* are important genes for the pathway forming PtdCho. PtdCho can be used to make VLDL, or can be hydrolyzed in a pathway involving *PNPLA3*, or secreted in bile by a flippase encoded for by *ABCB4*. We found that SNP patterns across multiple genes, including *ABCB4*, *APOC3*, *CHDH*, *PEMT*, *PNPLA3*, *MTHFR*, and *SLC44A1*, were associated with hepatic steatosis burden.

SNPs in choline/1-carbon metabolism increase susceptibility (6–9). Recent evidence indicates that choline/1-carbon metabolism intersects with energy metabolism pathways (4), suggesting that choline could be an important contributing factor in the metabolic imbalances associated with hepatic steatosis.

Because of the role of choline metabolism in mechanisms relevant to both hepatic lipid balance and obesity, we hypothesized that patterns of SNPs in genes that span choline/1-carbon metabolism and related pathways mediate susceptibility to steatosis (Fig. 1). To identify this genetic signature for steatosis, we began with an assessment of individual SNP effects in biopsy-characterized hepatic steatosis case and control subjects with a comprehensive range of body weights. To investigate the contribution of the combined effects of multiple SNPs on severity of steatosis, we implemented a novel unsupervised hierarchical clustering approach to group subjects by shared genotypes and asked whether these SNP patterns in choline/1-carbon genes and functionally related pathways correlated with steatosis burden. Our findings suggest that the interaction of these genetic variants results in metabolic inefficiency due to alterations in protein function that cumulatively alters flux through metabolic pathways.

MATERIALS AND METHODS

Human subjects

The Duke University Health Systems NAFLD Clinical Database and Biorepository includes subjects (age ≥18) with risk factors for NAFLD who underwent clinical evaluation and standard of care percutaneous liver biopsy for assessment of this presumed diagnosis. Informed consent of participants and Duke Institutional Review Board approval for genetic analysis were obtained for the purpose of this study. Our study includes 446 subjects for whom both histological data and DNA were available. Of these subjects, 263 (63%) had biopsies due to metabolic or biochemical risk factors for NAFLD, and 183 (37%) had biopsies during bariatric surgery. The subjects who underwent bariatric surgery were typically not required to undergo weight-loss interventions prior to surgery. Any patient who reported alcohol consumption in excess of 20 g/d or had positive serologic studies for chronic viral hepatitis B or C was excluded. Demographic, anthropometric, and clinical data were obtained at the time of liver biopsy. Whole blood for serum, plasma, DNA, and RNA was collected at the time of liver biopsy. Laboratory studies (*i.e.*, lipids, glucose, liver aminotransferases, and measures of liver synthetic function) were obtained within 6 mo of liver biopsy or 1-2 wk prior to bariatric surgery. The presence of diabetes mellitus was defined by known diagnosis in the medical record, use of any antidiabetic agent or insulin, fasting glucose value of ≥126 mg/dl, 2 h glucose tolerance test of ≥200 mg/dl, and/or a glycosylated hemoglobin value of >6.5%. All liver biopsies were reviewed and graded for steatosis according to the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) scoring system (10).

Candidate gene selection and genotyping

Twenty-one candidate genes were identified for study: ATP-binding cassette, subfamily B (MDR/TAP), member 4

(*ABCB4*), adiponectin (*ADIPOQ*), apolipoprotein C-III (*APOC3*), betaine-homocysteine methyltransferase (*BHMT*), cystathionine-β synthase (*CBS*), choline dehydrogenase (*CHDH*), choline kinase α (*CHKA*), choline kinase β (*CHKB*), fatty acid desaturase 2 (*FADS2*), methylenetetrahydrofolate desaturase 1 (*MTHFD1*), 5,10-methylenetetrahydrofolate reductase (*MTHFR*), 5-methyltetrahydrofolate-homocysteine methyltransferase (methionine synthase; *MTR*), phosphate cytidylyltransferase 1, choline, α (*PCYT1A*), phosphate cytidylyltransferase 1, choline, β (*PCYT1B*), phosphatidylethanolamine-*N*-methyltransferase (*PEMT*), patatin-like phospholipase domain containing 3 (*PNPLA3*), peroxisome proliferator-activated receptor γ (*PPARG*), stearoyl-CoA desaturase (*SCD*), serine hydroxymethyltransferase 1 (*SHMT1*), solute carrier family 44, member 1 (*SLC44A1*), and signal transducer and activator of transcription 3 (*STAT3*). Gene selection was based on the hypothesis that choline metabolism influences hepatic steatosis. Thus, selected genes are involved in choline/lipid metabolism and/or are involved in pathways believed to influence the pathophysiology of NAFLD and that intersect functionally with the 1-carbon pathway. For all genes except *ADIPOQ*, *APOC3*, *CBS*, *MTR*, *PPARG*, *PNPLA3*, and *STAT3*, SNPs to be genotyped were selected with HaploView Tagger 4.2 (Broad Institute, Cambridge, MA, USA; ref. 11) in order to comprehensively map SNPs within our genes of interest. Pairwise tagging with an r^2 threshold of 0.8 was utilized to generate a list of tag SNPs. To enrich for functional SNPs, this list was then supplemented with data from the National Human Genome Research Institute Catalog of Published Genome-Wide Association Studies (GWASs; http://www.genome.gov/gwastudies/) and SNPs that have been identified in the literature to be associated with NAFLD or related metabolic phenotypes. To further enrich for functional SNPs, we prioritized SNP selection to favor SNPs in exons or other regulatory regions (intronic enhancers, promoter regions). We included SNPs that met the criteria above that were within 5 kb of the gene boundary to include SNPs that could be in nearby regulatory regions. Gene boundaries were determined with the University of California–Santa Cruz Genome Browser [National Center for Biotechnology Information (NCBI) Genome Build 36.1; ref. 12]. For *ADIPOQ*, *APOC3*, *CBS*, *MTR*, *PPARG*, *PNPLA3*, and *STAT3*, a small number of SNPs with published roles in NAFLD or choline metabolism were included. A total of 260 SNPs were included in this study, and these were genotyped using a custom Illumina GoldenGate array (Illumina, Inc., San Diego, CA, USA; ref. 13), Sequenom (San Diego, CA, USA; ref. 14), or real-time PCR (15). Genotyping was initially conducted on the Illumina GoldenGate array. However, because of failures in that platform or design-related constraints, Sequenom and real-time PCR methods were also used. Methods and rationale for the use of these platforms are described in more detail below.

Oligo-specific extension-ligation assay

300 SNPs were genotyped with an Illumina GoldenGate array (13) at the University of North Carolina Mammalian Genotyping Core (Chapel Hill, NC, USA) with a design based on dbSNP reference SNP (rs) numbers from genome build 128 (NCBI Genome Build 36.1). The raw Illumina genotyping data were analyzed by GenomeStudio 2010 data analysis software (Illumina) for automated genotype calling. Parent/child trios from the Centre d'Étude du Polymorphisme Humain (CEPH) pedigree were obtained from the Coriell Cell Repository (Camden, NJ, USA; http://ccr.coriell.org/) and were included on each plate, including one replicate sample. Quality control (QC) procedures were implemented as per standard Illumina procedures, with additional manual review of SNP clusters to ensure data integrity. Twenty-one SNPs did not pass QC, and 30 had a minor allele frequency of <1%, leaving 249 SNPs that were included in the analysis.

Matrix-assisted laser desorption/ionization–time-of-flight (MALDI-TOF) primer extension assay

A subset of SNPs (n=17) was assayed using a multiplexed MALDI-TOF primer extension assay (MassArray iPLEx Gold; Sequenom; ref. 14) per the manufacturer's instructions (David H. Murdock Research Institute, Genomics Core Facility, Kannapolis, NC, USA) for the following reasons: failure on the Illumina platform; SNPs could not be genotyped with Illumina due to design constraints; SNPs were believed to be tri- or tetra-allelic on newer versions of dbSNP but still passed the design score for Illumina (*i.e.*, these were measured on both platforms); or SNPs were not included in the original Illumina design but were subsequently identified as having an important role in NAFLD. Primers were designed using the MassArray 4.0 assay designer software (Sequenom). The assay was validated using samples from the Coriell Cell Repository and 2 samples from the population being studied. Coriell samples were also run with each assay (6 samples from various ancestries, run in replicates). Genotyping data were generated by MassArray Typer 4.0 (Sequenom). Genotyping quality was assured by manual inspection. Out of all the putative tri- or tetra-allelic SNPs, a third allele was found for two SNPs, but each of these appeared in only one subject. Inspection of these data showed comparable genotype calls (except at the two loci with a third allele). This validates our multiplatform approach. Also, because of the very small incidence of additional alleles and the superior call rate with Illumina, all duplicated SNP data (n=3) were used from the Illumina platform. One SNP failed on this platform, and 3 had minor allele frequencies of <1%, leaving 10 SNPs that were included in the analysis.

Real-time PCR genotyping for *PEMT* rs12325817

This SNP was essential in our analysis due to its relevance to choline deficiency-mediated fatty liver in humans, but did not pass the design score for the Illumina or Sequenom assays. We previously developed and validated a SYBR green-based allele-specific PCR method to genotype this SNP (15). The same method was utilized, with the exception that it was optimized for real-time PCR. Forward primers specific to each allele were designed so that the SNP is located at the 3′-end of the priming sequence (forward for C allele: 5′-CCTGGACAACATGGTGACACTCC-3′; forward for G allele: 5′-CCTGGACAACATGGTGACACTCG-3′; reverse common: 5′-GTGGGCCCTGTACTTTTACATC-3′). The conditions used (initial denaturation: 98°C for 20 s; 35 cycles: denaturation 95°C for 30 s, annealing 68°C for 30 s, extension 72°C for 30 s, final extension 72°C for 10 min; melt curve: 95°C for 15 s, 58°C for 15 s, ramp 20 min, 95°C for 15 s) allow for specific PCR products to be synthesized only if the primer is 100% complementary to its template DNA. Real-time PCR was performed on an Eppendorf Realplex 4.0 (Eppendorf North America, Westbury, NY, USA). We were able to verify the validity of merging these data with the data generated from other two platforms by confirming the linkage disequilibrium of this SNP with two other loci across the other two platforms (rs4646343, Sequenom; rs3760188, Illumina). After including this SNP in the data set, the final number of SNPs analyzed was 260 (See Supplemental Fig. S1 for list of SNPs).

Statistical analyses

To examine SNP-steatosis associations, 4 groups were used: non-NAFLD (no histological evidence of fatty liver; n=85;

considered our controls) and 3 groups with increasing levels of hepatic fat accumulation [<34% (n=183; combination of <5% and 5-33% groups), 34-66% (n=118), and >66% (n=60)]. Proportional odds and generalized logit methods were implemented (each allele was assumed to have an additive effect). The generalized logit method was used only if the proportional odds assumption was violated. We built 3 models: model 1 was unadjusted for confounders; model 2 was adjusted for age (<50 or ≥50), sex, and diabetes diagnosis; and model 3 was adjusted for age, sex, diabetes, and body mass index (BMI). The Benjamini-Hochberg method for false discovery rate (FDR) was used to adjust for multiple tests utilizing the minimum number of independent tests approach or by adjusting for the number of variables tested (see details in figure legends and table notes; ref. 16). This was done separately for each of the models generated. Continuous and categorical variables were analyzed by ANOVA and Kruskal-Wallis test, respectively. A χ^2 test was used to compare genotype frequencies between ethnicities and clinics. FDR-corrected values of $P < 0.05$ were considered statistically significant.

Unsupervised 2-way hierarchical clustering (Ward's method; ref. 17) was conducted to group subjects by shared genotype patterns. SNPs were coded as 0, 1, or 2 (representing 0, 1, or 2 copies of the SNP, respectively). We constructed 2 models: model 1 included all 260 SNPs minus highly correlated SNPs (r^2>0.95; n=225). Model 2 included SNPs that had a raw value of $P < 0.1$ in the individual SNP analysis (model 1) minus highly correlated SNPs (n=55). The second model was selected to refine steatosis burden predictability and reduce noise from uninformative SNPs. Numbers of clusters in both models were selected by identifying the points at which dissimilarity between clusters substantially increased at each clustering step (measured by the error sum of squares; ref. 17) and by selecting the number of clusters preceding such increase that also provided a large enough sample size. A χ^2 test and pairwise comparisons were used to compare steatosis burden between clusters. To identify SNPs that distinguished clusters from one another, we evaluated distributions of SNP genotypes between clusters using the χ^2 test for marginal response homogeneity (likelihood ratio).

Genotypes within cluster 1 that differentiated subjects based on steatosis level were identified with supervised stepwise ordinal logistic regression. The scores for our model were derived by totaling the number of alleles associated with higher steatosis (0, 1, or 2) for the identified SNPs. In some cases, there were no subjects with two copies of the SNP, so the scores were 0 or 1. The scores were weighted depending on whether each allele had an additive effect on steatosis or whether having 0/1 or 1/2 had equivalent effects as determined by the ordinal regression model. If an SNP was associated with higher steatosis and the effects were additive, the scores were 0, 1, or 2 if the subject had 0, 1 or 2 copies of the SNP, respectively. If, on the other hand, having an SNP was associated with lower steatosis, the scores were 2, 1, or 0, if the subject had 0, 1, or 2 copies of the SNP, respectively, since having higher steatosis was associated with having fewer alleles. If having 1 or 2 copies of the SNP was associated with an equivalent effect on higher steatosis, having either 1 or 2 copies received a score of 2, while having 0 copies received a score of 0. If, on the other hand, having 1 or 2 copies of the SNP was associated equivalently with having less steatosis, having 0 copies of the SNP received a score of 2, while having 1 or 2 copies received a score of 0. The sum of alleles for each subject was the assigned score. **Table 1** provides the details.

A logistic regression fitting model was used to assess the accuracy of SNP scores to predict steatosis burden. Model discrimination was evaluated using receiver operating curve analysis (Supplemental Fig. S2). Analyses were conducted using SAS 9.2 or JMP 8.0 (SAS Institute, Cary, NC, USA).

TABLE 1. *Cluster 1 model scoring algorithm*

Gene	SNP	Risk allele	SNP copies		
			0	1	2
ABCB4	rs31672	T	2	2	0
APOC3	rs2854117	C	2	2	0
CHDH	rs4563403	C	2	2	0
MTHFR	rs7525338	C	2	0	NA
PEMT	rs13342397	T	2	2	0
PEMT	rs7946	T	2	0	0
PEMT	rs8068641	A	2	0	0
PEMT	rs936108	G	2	2	0
PNPLA3	rs2281135	G	0	1	2
PNPLA3	rs738409	T	0	1	2
SLC44A1	rs10820799	A	2	2	0

RESULTS

Population characteristics

Although all subjects who were enrolled in the biorepository were suspected to have NAFLD based on the presence of traditional risk factors, biopsies from 19% showed normal liver histology, and we considered this group our non-NAFLD controls. Characterization of our study population as a whole (**Table 2**) and by fatty liver groups (**Table 3**) revealed a pattern consistent with what is known about clinical and demographic parameters associated with NAFLD (1). About 1/3 of our cohort consisted of subjects who had biopsies done during bariatric surgery, as opposed to subjects who were seen in the liver clinic. There were notable differences between these two subpopulations (Table 2), as would be expected based on the criteria for eligibility for bariatric surgery (*i.e.*, morbid obesity or obesity with metabolic complications) and the expected demographics of patients who choose to have bariatric surgery (18). Specifically, all subjects from the bariatric clinic had BMI > 34, steatosis was less severe, and there was an overrepresentation of females and African Americans (Table 2). However, the proportions of case and control subjects were not different (FDR, P=0.3), the diagnosis of NAFLD was made concomitant with bariatric surgery, and, notably, SNP frequency was highly correlated between the two groups (r^2=0.96; **Fig. 2**). Our purpose was to study the role of genetic variants on hepatic steatosis over a wide range of BMIs because we hypothesize that obesity provides a challenge to hepatic lipid secretion capacity that augments metabolic inefficiencies associated with SNPs. For this reason, we included patients from both clinics to allow us to study a comprehensive spectrum of steatosis and obesity with sufficient power.

TABLE 2. *Histological, clinical, and demographic characteristics of cohort*

Characteristic	Cohort, n = 446	Liver clinic, n = 283	Bariatric clinic, n = 163	Liver *vs.* bariatric (P)	
				Raw	FDR
NAFLD	81%	83%	78%	0.3[a]	0.3[a]
Non-NAFLD	19%	17%	22%		
<34%	41%	40%	71%	<0.0001[b]	<0.0001[b]
34–66%	26%	42%	16%	<0.0001[b]	<0.0001[b]
>66%	13%	18%	13%	<0.0001[b]	<0.0001[b]
Age	48 (39, 56)	50 (40, 58)	47 (38, 54)	0.02	0.02
BMI	38 (31, 45)	33 (29, 38)	47 (43, 53)	<0.0001	<0.0001
Lean	5%	7%	0%	<0.0001	<0.0001
Overweight	15%	24%	0%	<0.0001	<0.0001
Obese	34%	48%	10%	<0.0001	<0.0001
Morbidly obese	29%	16%	49%	<0.0001	<0.0001
Super obese	17%	4%	40%	<0.0001	<0.0001
African American	15%	10%	23%	0.0001	0.0003
Caucasian	85%	90%	77%	0.0001	0.0003
Nondiabetic	65%	64%	66%	0.8	0.8
Diabetic	35%	36%	34%	0.8	0.8
Females	63%	52%	80%	<0.0001	<0.0001
Males	37%	48%	20%	<0.0001	<0.0001
AST	39 (26, 59)	47 (32, 73)	27 (21, 36)	<0.0001	<0.0001
ALT	47 (28, 83)	63 (40, 98)	24 (19, 34)	<0.0001	<0.0001
Total cholesterol	187 (158, 217)	191 (155, 220)	181 (158, 204)	0.1	0.2
HDL	41 (33, 49)	40 (32, 49)	43 (35, 49)	0.08	0.1
LDL	113 (87, 135)	115 (86, 138)	112 (91, 128)	0.5	0.6
Triglycerides	133 (95, 196)	138 (99, 215)	122 (84, 166)	0.02	0.02
Systolic BP	134 (123, 146)	132 (122, 141)	142 (131, 152)	<0.0001	<0.0001
Diastolic BP	80 (72, 87)	77 (71, 83)	84 (79, 90)	<0.0001	<0.0001

Variables are shown for the cohort as a whole (n=446) and with comparisons between bariatric and liver clinics. Data are presented as median and 25th and 75th percentile or as a percentage. Only African-American and Caucasian ethnicities are compared, since the remaining subjects (n=16; 4%) were of several ethnicities, and the cell size was too small to make a useful comparison. AST, aspartate aminotransferase; ALT, alanine aminotransferase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; BP, blood pressure. For continuous measures, Kruskal-Wallis test was used. For categorical variables, Fisher's exact test was used. P values were adjusted for the number of variables tested (n=15) using the FDR procedure; values of $P < 0.05$ were considered significant. [a]P *vs.* non-NAFLD. [b]P for distribution of steatosis groups.

Associations of individual polymorphisms with hepatic steatosis severity

After adjusting for a full complement of confounders, only *PNPLA3* rs738409 was significantly associated with increased steatosis burden (FDR, P=0.02; **Table 4**). However, a number of other SNPs did show associations before adjusting for BMI. The *APOC3* promoter SNP rs2854117 (T allele) was associated with lower steatosis (model 2; FDR, P=0.04; Table 4). We identified 9 novel SNPs in choline/1-carbon metabolism that were associated with lower steatosis (model 2; FDR, $P <$ 0.05): *MTHFR* (rs7525338, rs868014), *CHDH* (rs4563403, rs881883), *PEMT* (rs8068641, rs7214988), and *SLC44A1* (rs10820799, rs440290, rs328006) (Table 4). We were unable to confirm the previously reported fatty liver associations of individual SNPs in *ADIPOQ* rs1501299 (19), *CHDH* rs9001 and rs12676 (7), *MTHFD1* rs2236225 (8), *PEMT* rs12325817 (7) and rs7946 (20), *PPARG* rs1801282 (21), and *STAT3* rs9891119 (22) (FDR, P>0.05; not shown).

Steatosis is related to both ethnicity and genotype

African-American individuals are reported to have reduced risk for developing NAFLD (23), making ethnicity a potential confounder in our analysis, and our results confirm this lower incidence (Table 3). Most of the SNPs with fatty liver associations were differentially distributed in Caucasians and African Americans (**Table 5**). Out of all the SNPs analyzed in this study, 78% were differentially distributed in African Americans *vs.* Caucasians (P<0.05; not shown); the overall lack of correlation in SNP frequency is shown in **Fig. 3** (r^2=0.08). Data from 1000 genomes (24) shows that the discordance in SNP frequency of the SNPs individually associated with steatosis in our study between African Americans and Caucasians is not unique to our study (not shown). Other ethnic groups in 1000 genomes, such as those of Asian and Caribbean descent, do not show the same level of discordance in the SNPs individually associated with steatosis in our study (not shown). Because lower steatosis could be related to both ethnicity and genotype, we did not adjust for ethnicity in any of our models.

Genetic signatures improve prediction of steatosis severity

To determine whether combinations of SNPs that influence hepatic lipid balance are more predictive than single SNPs, we began by asking whether having

TABLE 3. *Characteristics of control* vs. *NAFLD subjects*

Characteristic	Level of steatosis				*P*	
	Non-NAFLD, n = 85	<34%, n = 183	34-66%, n = 118	>66%, n = 60	Raw	FDR
Age	45 (33, 60)	50 (41, 58)	49 (40, 57)	48 (40, 53)	0.07	0.1
BMI	37 (27, 46)	42 (33, 50)	35 (31, 43)	38 (32, 45)	0.0004	0.0009
Female, n = 279	20%	42%	23%	15%	0.2	0.2
Male, n = 167	18%	40%	32%	10%		
Nondiabetic, n = 288	19%	42%	25%	14%	0.9	0.9
Diabetic, n = 158	18%	40%	28%	13%		
African-American, n = 64	30%	50%	11%	9%	0.003	0.005
Caucasian, n = 366	18%	39%	29%	14%		
AST (IU/L)	32 (22, 49)	31 (24, 47)	46 (31, 69)	53 (40, 82)	0.0001	0.0003
ALT (IU/L)	32 (20, 62)	33 (21, 63)	57 (35, 90)	70 (46, 110)	0.0001	0.0003
Total cholesterol (mg/dl)	182 (158, 209)	182 (156, 208)	191 (154, 225)	204 (155, 235)	0.4	0.4
HDL (mg/dl)	51 (42, 64)	43 (34, 48)	38 (31, 45)	39 (33, 47)	0.0001	0.0003
LDL (mg/dl)	113 (83, 129)	114 (93, 134)	110 (82, 138)	122 (85, 135)	0.8	0.9
Triglycerides (mg/dl)	106 (65, 151)	126 (92, 172)	142 (99, 232)	169 (106, 294)	0.0001	0.0003
Systolic BP	130 (121, 139)	140 (131, 149)	135 (120, 146)	133 (126, 143)	0.0001	0.0003
Diastolic BP	79 (72, 84)	82 (76, 89)	78 (72, 84)	81 (74, 89)	0.001	0.002

Table is similar to Table 2, except that clinical and demographic characteristics were compared across the cohort (n=446) based on groupings indicating histological steatosis burden per the NASH CRN scoring system. FDR adjustment for multiple comparisons was based on the number of variables tested (n=13). Data are presented as median and 25th and 75th percentile or as a percentage. BMI, body mass index; AST, aspartate aminotransferase; ALT, alanine aminotransferase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; BP, blood pressure.

multiple SNPs in *PNPLA3* (the gene most consistently associated with steatosis; ref. 25) was associated with more severe fatty liver. We found that carriers of ≥1 copy of both rs738409 and rs2281135 (n=197), when compared to people who had an SNP in only one locus or neither locus (n=249), had significantly higher levels of steatosis (P=0.0002; **Fig. 4**).

To investigate whether combinations of SNPs across multiple choline/1-carbon metabolism genes and genes in functionally related pathways might indicate additive metabolic inefficiencies contributing to steatosis risk, we used unsupervised hierarchical clustering to group subjects with similar genotype patterns across all SNPs in our data set (n=225; 35 SNPs removed due to high linkage disequilibrium correlation; r^2>0.95). On the basis of the criteria for cluster partitioning, 5 clusters were identified (not shown). We then asked whether the genetic signatures that pulled together the clusters had any association with the degree of steatosis. There was an overall difference in level of steatosis across the clusters (P=0.008), with cluster 1 having significantly lower steatosis than all other clusters (P<0.007) and the remaining clusters having similar levels of steatosis (not shown).

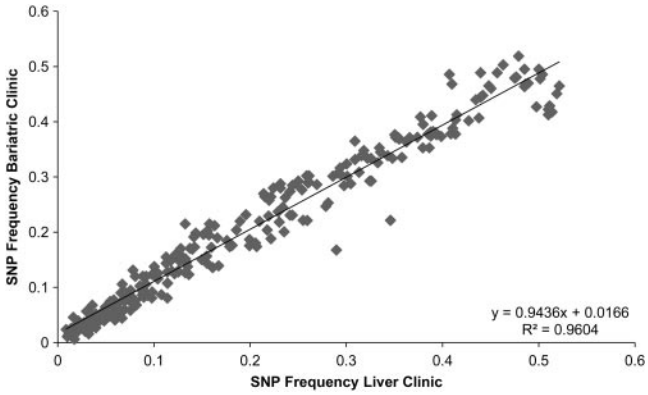

Figure 2. SNP distributions are highly correlated in liver and bariatric clinics. Diamonds represent frequency of distribution in the subjects from the liver clinic *vs.* the bariatric clinic for all 260 SNPs included in this study. The majority of SNPs were distributed similarly between the two clinics (r^2=0.96).

To improve our ability to distinguish between the SNPs that contributed most to clusters that had subjects with higher steatosis burden, we created a second model containing only SNPs with raw values of $P < 0.1$ from the individual SNP association analysis (n=55 after removing highly correlated SNPs due to high linkage disequilibrium). Similarity-based cluster partitioning resulted in 5 clusters (**Fig. 5** and Supplemental Fig. S1). Overall, the clusters differed significantly in their steatosis burden (P=0.0002; **Fig. 6**). Subjects in cluster 1 were significantly more likely to have a lower steatosis burden than did all the other clusters (P<0.007), while those in cluster 2 had a higher steatosis burden than did clusters 3A and 3B (P<0.05). Clusters 3A–3C were not significantly different from each other, and we considered them jointly to represent an intermediate amount of steatosis.

We carefully phenotyped the clusters to determine which characteristics coincided with the shared genotype pattern (**Table 6**). The vast majority of the African-American subjects in our study (94%) shared the cluster 1 genotype pattern. Although cluster 1 subjects had the lowest liver fat, they had the highest BMI. Cluster 2 had the highest liver enzyme markers for hepatic damage and elevated triglycerides. Subjects from the liver and bariatric clinics were distributed throughout all 5 clusters. Because clusters 1 and 2 represented the most obvious phenotypic outliers, we focused on defining the SNP patterns of those two

TABLE 4. *SNPs in 1-carbon and lipid metabolism are associated with NAFLD severity*

Gene, rs, and SNP	Region	SNP frequency				Model 1, unadjusted for confounders (*P*)		Model 2, adjusted for age, sex, diabetes (*P*)		Model 3, adjusted for age, sex, diabetes, BMI (*P*)	
		Non-NAFLD, $n = 85$	<34%, $n = 183$	34–66%, $n = 118$	>66%, $n = 60$	Raw	FDR	Raw	FDR	Raw	FDR
PNPLA3											
rs738409 C→G	Exon	0.212	0.291	0.377	0.483	<0.0001	0.0002	<0.0001	0.0002	0.0001	0.02
rs2281135 C→T	Intron	0.194	0.24	0.305	0.375	0.001	0.035	0.009	0.03	0.007	0.3
APOC3											
rs2854117 C→T	US	0.435	0.309	0.241	0.242	0.0005	0.041	0.004	0.04	0.004	0.3
PEMT											
rs8068641 A→G	Intron	0.228	0.201	0.107	0.092	0.0006	0.032	0.0005	0.03	0.007	0.3
rs7214988 C→G	Intron	0.171	0.145	0.068	0.092	0.002	0.042	0.002	0.04	0.02	0.3
CHDH											
rs4563403 C→T	DS	0.235	0.205	0.111	0.092	0.0008	0.035	0.0007	0.03	0.1	0.5
rs881883 T→C	DS	0.288	0.268	0.169	0.142	0.002	0.042	0.002	0.04	0.03	0.3
MTHFR											
rs7525338 C→T	DS	0.035	0.019	0	0	0.003	0.042	0.003	0.04	0.5	0.6
rs868014 C→T	3′UTR	0.065	0.033	0.013	0.025	0.002	0.042	0.002	<0.05	0.01	0.3
SLC44A1											
rs10820799 A→C	Intron	0.129	0.112	0.047	0.042	0.003	0.042	0.003	0.04	0.2	0.5
rs440290 T→C	US	0.241	0.186	0.114	0.178	0.002	0.042	0.003	0.04	0.02	0.3
rs328006 G→C	Intron	0.218	0.128	0.11	0.125	0.003	0.042	0.003	0.04	0.05	0.4

Association of SNPs and degree of steatosis was tested across our cohort ($n = 446$). The SNP is defined by the dbSNP reference SNP (rs) number. Allele change is shown underneath. Region refers to the sequence location: within an exon or intron, upstream (US) or downstream (DS) of the gene (within the 5-kb cutoff for our study), or in the 3′ untranslated region (UTR). SNP frequency: proportion of people carrying either 1 or 2 copies of the SNP [(2 × no. of people with 2 copies of SNP) + (1 × no. of people with 1 copy of SNP)/(no. of subjects × 2)]. Model 1, unadjusted for confounders; model 2, adjusted for age (<50 *vs.* ≥50), sex, and diabetes diagnosis; and model 3, adjusted for age, sex, diabetes, and BMI. Proportional odds models or generalized logit models were used depending on whether the proportional odds assumption was satisfied (assuming an additive SNP effect). SNPs are organized by raw *P* value and gene. *P* values were adjusted for multiple tests utilizing the FDR procedure. The FDR *P* value was derived by adjusting the raw *P* value for the number of SNPs tested (n=170 after accounting for the minimum number of effective tests); values of $P < 0.05$ were considered significant.

clusters, since that would provide the most likely set of candidate SNPs for defining low *vs.* high steatosis risk.

Next, we aimed to define the SNPs that most clearly distinguished clusters 1 and 2. Since clusters were defined by both ethnicity and SNPs, we asked whether cluster 2, which had a similar ethnic distribution as clusters 3A–3C (*i.e.*, primarily Caucasian), had a unique genetic signature. We found that 32 SNPs in *ABCB4*, *CHDH*, *PEMT*, *PNPLA3*, *MTHFR*, and *SLC44A1* were distributed differentially in cluster 2 *vs.* clusters 3A–3C (FDR, $P<0.05$; **Table 7**). Cluster 1 was significantly different from all other clusters, so we compared its genotype distributions to those in all other clusters. Forty-nine SNPs in *ABCB4*, *APOC3*, *CHDH*, *CHKB*, *FADS2*, *PEMT*, *MTHFD1*, *MTHFR*, *PCYT1A*, *PCYT1B*, *SCD*, and *SLC44A1* were differentially distributed in cluster 1 compared to all other clusters (FDR, $P<0.05$; **Table 8**). Notably, the previously reported NAFLD-associated alleles in *CHDH* rs12676 (7) and *PEMT* rs7946 (20), which in our study did not retain statistical significance individually, were enriched in cluster 2 and depleted in cluster 1 (FDR, $P<0.05$).

We found that SNP pattern and being African American were highly correlated, making it difficult to determine whether the lower steatosis seen in cluster 1 was due to SNPs or to another property associated with being African American. To address this, we asked whether SNP patterns within cluster 1 could distinguish the subjects with higher steatosis from those with less disease. Stepwise ordinal regression identified 11 SNPs within cluster 1 that were associated with degree of steatosis (*ABCB4* rs31672; *APOC3* rs2854117; *CHDH*

TABLE 5. *SNP frequency is discordant in African Americans* vs. *Caucasians*

Gene	rs and SNP	SNP frequency		*P*
		African Americans	Caucasians	
PNPLA3	rs738409 C→G	0.242	0.341	0.03
	rs2281135 C→T	0.242	0.272	0.5
APOC3	rs2854117 C→T	0.754	0.229	<0.0001
MTHFR	rs7525338 C→T	0.102	0.000	<0.0001
	rs868014 C→T	0.206	0.003	<0.0001
CHDH	rs4563403 C→T	0.453	0.123	<0.0001
	rs881883 T→C	0.609	0.157	<0.0001
PEMT	rs8068641 A→G	0.563	0.102	<0.0001
	rs7214988 C→G	0.414	0.075	<0.0001
SLC44A1	rs10820799 A→C	0.320	0.051	<0.0001
	rs440290 T→C	0.651	0.096	<0.0001
	rs328006 G→C	0.508	0.075	<0.0001

Distribution of SNPs found to be significantly associated with steatosis severity across the cohort (n=446; Table 4) was compared in African Americans (n=64) *vs.* Caucasians (n=366) using a χ^2 test. Values of P <0.05 were considered significant.

rs4563403; *MTHFR* rs7525338; *PEMT* rs13342397, rs7946, rs8068641, and rs936108; *PNPLA3* rs2281135, rs738409; *SLC44A1* rs10820799), and they were used to construct a scoring model that summed alleles associated with increased steatosis. We could accurately classify individuals within cluster 1 with higher steatosis based on their SNP score (P=1.0×10^{-10}; Fig. 6, inset). The receiver operator curve demonstrated that scores from this

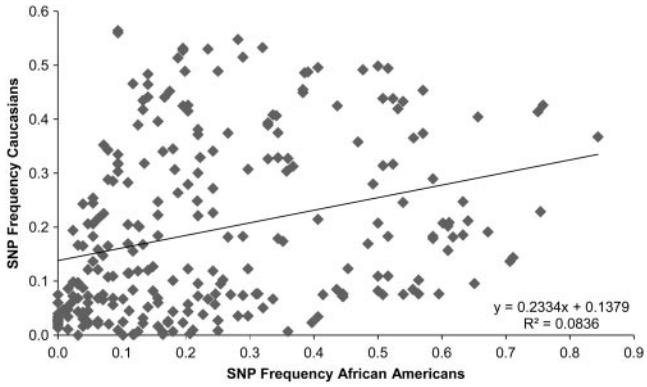

Figure 3. Overall SNP distribution is not correlated in African Americans *vs.* Caucasians. Figure shows the lack of correlation of SNP distribution between African Americans and Caucasians (r^2=0.08). Diamonds represent the frequency of each of the 260 SNPs included in this study.

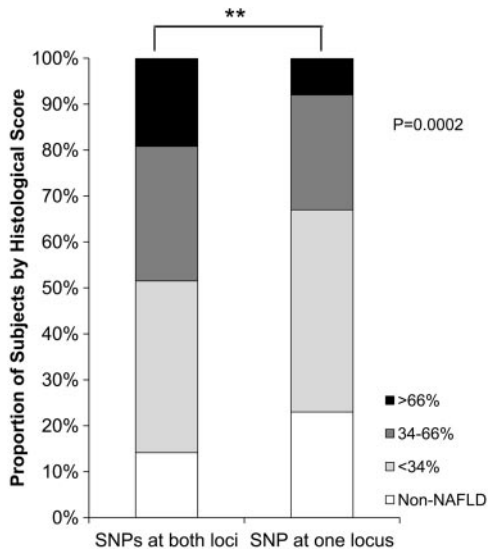

Figure 4. Synergistic effect on steatosis severity of having SNPs at both rs738409 and rs2281135 in PNPLA3. To assess whether having ≥1 copy of an SNP at both rs738409 and rs2281135 in the PNPLA3 gene was associated with higher steatosis, we used proportional odds models (adjusted for age, sex, diabetes, and BMI as in Table 4). Carrying SNPs at both PNPLA3 loci (n=197) was associated with higher steatosis burden than carrying an SNP at only one PNPLA3 locus (n=249; P=0.0002).

combination of SNPs discriminated between subjects with high steatosis with a high degree of sensitivity and specificity (area under the curve by steatosis level: <34%=0.83, 33-66%=0.94, >66%=0.93; Supplemental Fig. S2*B*).

DISCUSSION

Our study highlights a novel approach to maximize our understanding of how genetic variations influence hepatic steatosis: grouping individuals by shared patterns of SNPs in pathways functionally linked to the phenotype of interest, rather than using traditional single SNP or genome-wide approaches. In the postgenomic era, the identification of individual SNPs that are associated with risk for chronic disease has been disappointing. This is not surprising, given that GWASs usually must make statistical adjustments for testing vast numbers of SNPs (thereby losing power) and because a single gene, much less a single SNP, is unlikely to fully define risk for common complex diseases (2). In support of this, it is becoming increasingly clear that gene-gene interactions are much more likely to be relevant in defining disease risk, especially between interrelated pathways that influence phenotype (26). Our approach may become a fundamental first step in identifying epistatic relationships between genes that will allow for stronger hypotheses and increasingly targeted studies. This approach is also highly amenable to studies that merge advanced phenotyping platforms, such as metabolomics, in order to maximize our utilization of genetic signatures to define disease risk.

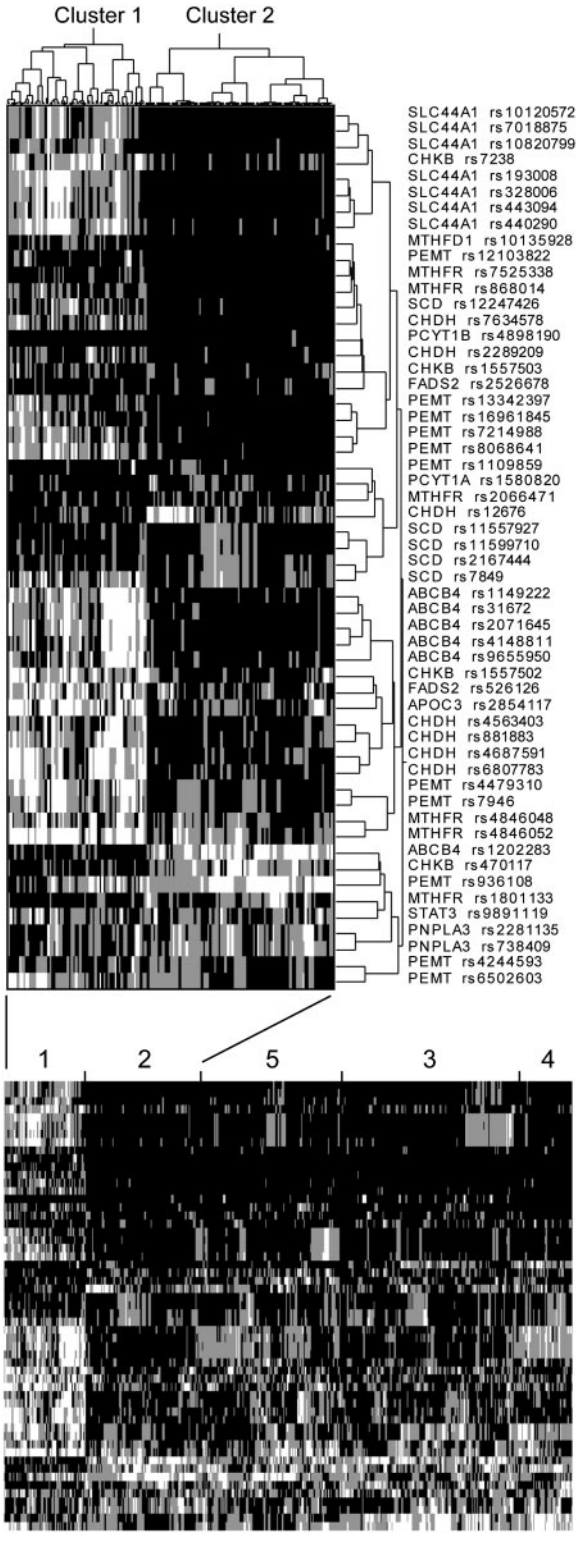

Figure 5. Unsupervised 2-way hierarchical clustering groups subjects with distinct genetic patterns. Genotypes are coded as 0 (no SNP), 1 (1 copy of SNP), or 2 (2 copies of SNPs) based on distributions across the cohort and are color coded as black, gray, and white, respectively. Clusters 1 and 2 are enlarged; inset shows a small image of all 5 clusters. (Supplemental Fig. S1 shows a large image of all 5 clusters.)

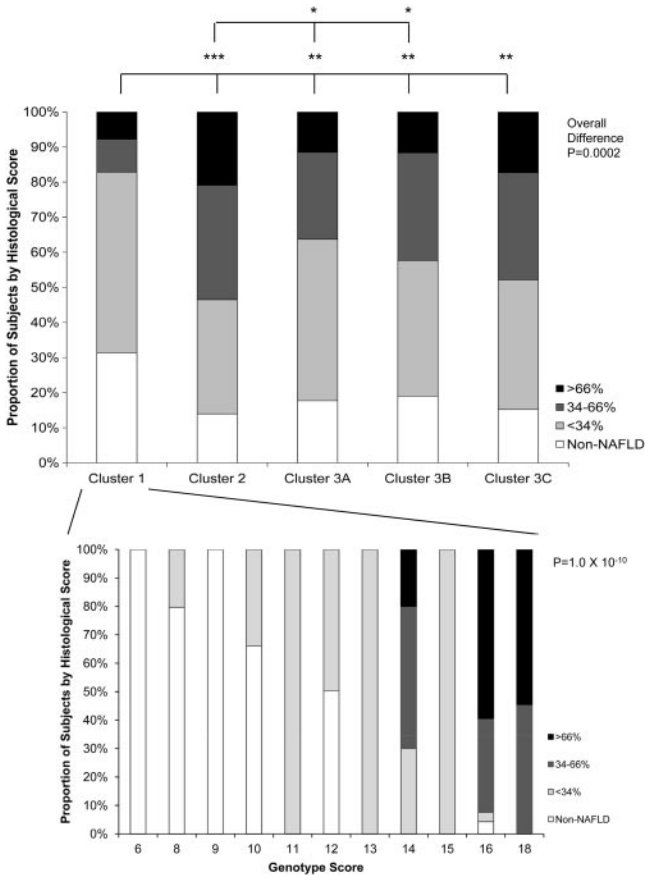

Figure 6. Steatosis burden can be predicted by clustering subjects by shared SNP patterns, and this prediction can be refined with scoring models that account for the biological association of SNPs on steatosis. Using a χ^2 test, we first identified an overall difference in steatosis between the 5 clusters generated based on 55 SNPs (P=0.0002). To identify which individual clusters were different, subsequent pairwise comparisons showed that cluster 1 had lower steatosis than all other clusters, cluster 2 had higher steatosis than clusters 3A and 3B, and clusters 3A–3C were not significantly different from each other. Steatosis is defined as non-NAFLD (no histological evidence of fatty liver), <34%, 34–66%, >66%. Inset: given that cluster 1 consisted primarily of African-American individuals, we asked whether SNP patterns within cluster 1 could distinguish the subjects with higher steatosis from those with less disease by implementing a scoring model. The model was built by summing the number of alleles associated with higher steatosis for SNPs that were selected due to their association with steatosis within the cluster (see Materials and Methods for full description). Inset shows how the cluster 1 scoring model performed. Discriminatory capacity of the scoring model was evaluated with a logistic fit model and was highly significant (P=1.0×10^{-10}). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Choline/1-carbon metabolism intersects with pathways that are important for liver function, including lipid/apolipoprotein and homocysteine (Hcy) metabolism (Fig. 1), making it plausible that interactions between SNPs across multiple genes may lead to cumulative inefficiencies that contribute to steatosis risk. This is supported by data in humans showing that dietary choline deprivation leads to the development of fatty liver differentially depending on sex and genetic variations in 1-car-

TABLE 6. *Phenotype characteristics of clusters*

Characteristic	Cluster					*P*	
	1, $n = 64$	2, $n = 86$	3A, $n = 113$	3B, $n = 137$	3C, $n = 46$	Raw	FDR
Age	46 (38, 53)	52 (40, 57)	48 (39, 59)	49 (39, 56)	48 (42, 56)	0.08	0.1
BMI	45 (36, 54)	37 (31, 45)	38 (32, 45)	37 (30, 47)	35 (31, 43)	0.0008	0.004
Female, $n = 279$	16%	17%	27%	30%	10%	0.4	0.4
Male, $n = 167$	11%	23%	23%	32%	10%	0.4	0.4
Nondiabetic, $n = 288$	13%	18%	25%	33%	11%	0.3	0.4
Diabetic, $n = 158$	18%	22%	26%	26%	9%	0.3	0.4
African American, $n = 64$	94%	0%	5%	2%	0%	<0.0001	<0.0001
Caucasian, $n = 366$	1%	23%	29%	36%	12%	<0.0001	<0.0001
AST (IU/L)	28 (21, 59)	45 (29, 72)	37 (26, 53)	36 (26, 56)	41 (27, 59)	0.02	0.04
ALT (IU/L)	29 (18, 60)	51 (30, 97)	42 (27, 77)	46 (25, 79)	47 (30, 71)	0.005	0.01
Total cholesterol (mg/dl)	187 (167, 212)	187 (154, 215)	186 (157, 217)	183 (153, 215)	198 (162, 239)	0.6	0.7
HDL (mg/dl)	44 (37, 58)	40 (31, 45)	39 (33, 47)	43 (34, 49)	44 (33, 58)	0.06	0.09
LDL (mg/dl)	117 (98, 131)	116 (85, 135)	115 (85, 137)	109 (88, 125)	117 (78, 155)	0.7	0.7
Triglycerides (mg/dl)	111 (69, 136)	137 (99, 211)	143 (101, 231)	136 (99, 193)	131 (99, 236)	0.003	0.009
Systolic BP	139 (128, 150)	135 (125, 146)	134 (129, 145)	137 (126, 145)	128 (118, 138)	0.01	0.02
Diastolic BP	85 (77, 91)	78 (72, 83)	79 (75, 88)	81 (74, 86)	80 (71, 83)	0.0002	0.001
Liver clinic, $n = 283$	10%	22%	25%	32%	12%	0.003	0.009
Bariatric clinic, $n = 163$	23%	15%	26%	29%	8%	0.003	0.009

Table is similar to Table 3, except the variables were compared between clusters generated from shared genotypes (Fig. 5 and Supplemental Fig. S1). FDR adjustment for multiple comparisons was based on the number of variables tested (n=14). Data are presented as median and 25th and 75th percentile or as a percentage. BMI, body mass index; AST, aspartate aminotransferase; ALT, alanine aminotransferase; HDL, high-density lipoprotein; LDL, low-density lipoprotein; BP, blood pressure.

bon genes (6–8). A recent publication showed that, in biopsy-characterized postmenopausal women, low choline intake was associated with worse fibrosis (27), highlighting choline as an important nutrient in NAFLD progression. Our study provides novel insights into the contributions of choline-related SNPs to steatosis severity.

Individual SNPs involved in hepatic steatosis

We confirmed the association between *PNPLA3* rs738409 and steatosis severity in a cohort with histologically defined hepatic lipid status (Table 4). Although *PNPLA3* rs738409 is a well-replicated marker of hepatic steatosis, this SNP only explains a small proportion of the risk. Indeed, in our study, 37% of individuals who did not have any degree of steatosis had ≥1 copy of the rs738409 SNP, while 43% of people with any degree of steatosis did not carry this SNP (not shown). Thus, we hypothesized that understanding combinations of shared SNPs in our cohort might give us better information about our phenotype of interest. In support of this notion, we noted an apparent synergism between *PNPLA3* rs738409 and rs2281135 SNPs (Fig. 4). This suggests that measuring both SNPs improves their utility as predictive tools. Interestingly, another *PNPLA3* SNP we genotyped, rs6006460, which has been reported to be common in African Americans and associated with decreased risk for NAFLD (25), was not found to be polymorphic in this population and thus had no effect on steatosis risk. PNPLA3 is a phospholipase A2 family member that hydrolyzes glycerolipids, including PtdCho, *in vitro* (28) and also has been shown to promote lipid synthesis due to its function as a lysophosphatidic acid acyltransferase (29). Given the involvement of PNPLA3 in the metabolism of choline-containing phospholipids, deciphering the crosstalk between PNPLA3 and choline/1-carbon metabolism may result in a deeper understanding of the mechanisms regulating hepatic lipid balance.

Excess lipid influx coupled with genetically defined diminished lipid export synergistically influence hepatic lipid burden

Accumulation of fat in the liver is a function of the balance between lipid input (driven largely by excess calorie intake) and the rate of lipid output (mainly as VLDL; Fig. 1). Obesity and hepatic steatosis frequently occur in the same individuals because the inputs to hepatic lipid production are increased, thereby challenging the capacity of liver to secrete lipids. Indeed, excess caloric intake is one important cause of increased rates of formation of hepatic triglycerides, and reduction in intake has been shown to diminish hepatic lipid burden (30). Choline metabolism is essential for the export of hepatic lipids, since its metabolite, PtdCho, is necessary for VLDL formation. Choline deficiency in humans and mice is associated with increased hepatic lipid stores. Although obesity is generally associated with higher steatosis (1), this is not the case in all individuals, and this may be because of differential capacity to secrete lipid due to metabolic inefficiencies caused by SNPs in choline and 1-carbon metabolism. In addition, some individuals may reach a threshold BMI in which lipid input is severely increased, and genetic contributions may become irrelevant.

TABLE 7. *SNP frequency of cluster 2* vs. *clusters 3A–3C*

Rank	Gene	SNP	SNP frequency		*P*	
			Cluster 2	Clusters 3A–C	Raw	FDR
1	*PEMT*	rs936108	0.767	0.414	2.12E-17	1.16E-15
2	*PEMT*	rs6502603	0.233	0.578	7.89E-17	2.17E-15
3	*ABCB4*	rs2071645	0.017	0.242	4.57E-15	8.38E-14
4	*ABCB4*	rs4148811	0.017	0.240	6.76E-15	9.29E-14
5	*PEMT*	rs4244593	0.198	0.497	1.31E-13	1.45E-12
6	*ABCB4*	rs31672	0.035	0.262	2.83E-13	2.23E-12
7	*ABCB4*	rs1149222	0.041	0.267	9.23E-13	6.34E-12
8	*ABCB4*	rs1202283	0.733	0.507	2.68E-07	1.64E-06
9	*ABCB4*	rs9655950	0.058	0.208	1.87E-06	1.03E-05
10	*PEMT*	rs1109859	0.058	0.209	3.55E-06	1.77E-05
11	*SLC44A1*	rs193008	0.006	0.101	6.11E-06	2.80E-05
12	*SLC44A1*	rs443094	0.006	0.100	8.12E-06	3.43E-05
13	*SLC44A1*	rs328006	0.006	0.098	1.08E-05	4.23E-05
14	*PEMT*	rs4479310	0.145	0.294	2.31E-05	8.48E-05
15	*PNPLA3*	rs738409	0.448	0.307	2.45E-05	8.41E-05
16	*PEMT*	rs13342397	0.006	0.093	2.48E-05	8.04E-05
17	*PEMT*	rs8068641	0.017	0.120	3.13E-05	9.57E-05
18	*PEMT*	rs7946	0.855	0.709	3.60E-05	1.04E-04
19	*PEMT*	rs16961845	0.012	0.095	1.63E-04	4.48E-04
20	*PEMT*	rs7214988	0.012	0.095	1.63E-04	4.27E-04
21	*CHDH*	rs12676	0.413	0.265	2.24E-04	5.61E-04
22	*SLC44A1*	rs440290	0.029	0.120	3.50E-04	8.38E-04
23	*CHDH*	rs881883	0.081	0.193	6.64E-04	0.002
24	*PNPLA3*	rs2281135	0.378	0.243	0.002	0.004
25	*SLC44A1*	rs7018875	0.000	0.041	0.002	0.005
26	*SLC44A1*	rs10820799	0.006	0.059	0.002	0.005
27	*CHDH*	rs6807783	0.076	0.171	0.003	0.01
28	*SLC44A1*	rs10120572	0.000	0.027	0.004	0.01
29	*CHDH*	rs4687591	0.087	0.169	0.009	0.02
30	*CHDH*	rs4563403	0.064	0.142	0.010	0.02
31	*MTHFR*	rs2066471	0.116	0.182	0.020	0.03
32	*CHDH*	rs7634578	0.000	0.012	0.06	0.1
33	*MTHFR*	rs4846048	0.238	0.292	0.06	0.1
34	*PCYT1B*	rs4898190	0.012	0.047	0.09	0.1
35	*CHDH*	rs2289209	0.017	0.051	0.1	0.2
36	*MTHFD1*	rs10135928	0.006	0.020	0.1	0.2
37	*CHKB*	rs470117	0.523	0.454	0.2	0.2
38	*MTHFR*	rs1801133	0.297	0.372	0.2	0.3
39	*FADS2*	rs526126	0.215	0.194	0.3	0.4
40	*MTHFR*	rs4846052	0.349	0.373	0.3	0.4

Distribution of SNPs used to generate clusters (n=55) was compared in cluster 2 *vs.* 3A–3C using a heterogeneity test. Raw *P* values were adjusted for multiple tests (FDR method) based on the number of SNPs tested (n=55). Top 40 SNPs are shown for space considerations.

Genetic signatures are associated with steatosis severity and may provide insight into disease disparities

The major finding in this study is that an unsupervised hierarchical clustering approach revealed patterns of SNPs across multiple functionally relevant genes that were highly associated with steatosis burden. Our approach defined SNP signatures, based on the number of copies of SNP alleles carried across a panel of 55 SNPs, which were associated with steatosis severity in specific groups of people that share other characteristics, such as ethnicity and biochemical indicators of NAFLD (Figs. 5 and 6). The genetic signatures that we identified span pathways that are known to be mechanistically related to NAFLD (Fig. 1). This supports our hypothesis that additive metabolic inefficiencies, driven at least in part by SNP patterns, contribute to steatosis severity.

Among the multiple choline/1-carbon genes we identified that defined low *vs.* high steatosis burden (Tables 7 and 8), many have biologically plausible links to hepatic steatosis. For example, we found that SNPs in *CHDH* and *PEMT* were associated with steatosis, and this is in agreement with previous studies (7, 20). We also uncovered SNPs in several genes not previously linked to steatosis, including *ABCB4, MTHFR*, and *SLC44A1. ABCB4* encodes for an enzyme that flips PtdCho from the inner to the outer hepatocyte canalicular membrane, so that this phospholipid is available for incorporation into bile micelles (31). Bile salts can act as signaling agents that modulate glucose and energy metabolism (32), and our observations suggest

TABLE 8. *SNP frequency of cluster 1* vs. *all other clusters*

Rank	Gene	SNP	SNP frequency		*P*	
			Cluster 1	All others	Raw	FDR
1	*CHKB*	rs7238	0.594	0.080	2.32E-33	1.28E-31
2	*SLC44A1*	rs440290	0.633	0.099	5.52E-33	1.52E-31
3	*CHDH*	rs7634578	0.367	0.009	6.27E-33	1.15E-31
4	*SLC44A1*	rs193008	0.586	0.080	6.52E-33	8.97E-32
5	*SLC44A1*	rs10120572	0.414	0.021	1.93E-32	2.13E-31
6	*SLC44A1*	rs7018875	0.430	0.031	1.05E-29	9.63E-29
7	*CHDH*	rs6807783	0.703	0.149	3.52E-29	2.76E-28
8	*CHDH*	rs4687591	0.695	0.151	1.06E-28	7.30E-28
9	*SLC44A1*	rs328006	0.516	0.077	7.90E-26	4.83E-25
10	*PEMT*	rs8068641	0.555	0.097	7.74E-25	4.25E-24
11	*ABCB4*	rs1202283	0.086	0.558	9.13E-25	4.56E-24
12	*ABCB4*	rs2071645	0.641	0.191	2.29E-23	1.05E-22
13	*MTHFR*	rs868014	0.219	0.001	3.22E-23	1.36E-22
14	*APOC3*	rs2854117	0.727	0.233	9.25E-23	3.63E-22
15	*MTHFR*	rs4846052	0.836	0.368	1.87E-21	6.85E-21
16	*SLC44A1*	rs443094	0.453	0.079	9.65E-21	3.32E-20
17	*ABCB4*	rs1149222	0.648	0.216	2.44E-20	7.90E-20
18	*SCD*	rs12247426	0.250	0.009	3.41E-20	1.04E-19
19	*CHDH*	rs881883	0.594	0.168	2.46E-19	7.13E-19
20	*ABCB4*	rs31672	0.617	0.211	2.06E-18	5.68E-18
21	*FADS2*	rs526126	0.617	0.199	2.31E-18	6.06E-18
22	*SLC44A1*	rs10820799	0.336	0.047	1.73E-17	4.31E-17
23	*CHKB*	rs1557502	0.594	0.186	8.92E-17	2.13E-16
24	*CHKB*	rs470117	0.102	0.470	2.46E-16	5.64E-16
25	*PEMT*	rs7214988	0.398	0.076	5.30E-16	1.17E-15
26	*ABCB4*	rs4148811	0.516	0.190	8.24E-15	1.74E-14
27	*PEMT*	rs4479310	0.648	0.260	1.70E-14	3.47E-14
28	*ABCB4*	rs9655950	0.500	0.174	1.83E-13	3.60E-13
29	*CHDH*	rs4563403	0.445	0.124	2.81E-13	5.34E-13
30	*MTHFR*	rs7525338	0.102	0.000	3.45E-13	6.32E-13
31	*PEMT*	rs12103822	0.125	0.000	3.20E-12	5.68E-12
32	*MTHFR*	rs1801133	0.070	0.355	2.54E-11	4.36E-11
33	*MTHFD1*	rs10135928	0.148	0.017	4.31E-10	7.18E-10
34	*PEMT*	rs7946	0.445	0.742	2.06E-09	3.34E-09
35	*PEMT*	rs16961845	0.297	0.076	4.61E-09	7.25E-09
36	*PEMT*	rs13342397	0.273	0.073	1.14E-08	1.74E-08
37	*CHDH*	rs2289209	0.203	0.043	1.39E-07	2.07E-07
38	*CHDH*	rs12676	0.086	0.298	1.80E-07	2.61E-07
39	*SCD*	rs7849	0.414	0.212	3.99E-06	5.62E-06
40	*PEMT*	rs4244593	0.203	0.429	9.01E-06	1.24E-05

Table is similar to Table 7, except the distribution of SNP used to generate clusters (n=55) was compared in cluster 1 *vs.* the rest of the clusters.

that this metabolic pathway, and its intersection with choline/1-carbon metabolism, should be explored further as a modulator of NAFLD. Recently, *Abcb4* was shown to modulate glucose metabolism in mice, likely *via* a PtdCho-dependent pathway, and *ABCB4* SNPs in humans were also associated with glucose levels (33). This novel metabolic link to *ABCB4* strengthens the biological plausibility of our findings and also suggests that this gene needs to be studied further in the context of metabolic disease. MTHFR is important for the formation of methionine *via* the remethylation of Hcy. Functional SNPs in this gene have been linked to elevated Hcy (34) and central adiposity (35), and both of these mechanisms could underlie a functional role of MTHFR in steatosis. The use of choline as a methyl group donor requires conversion of choline to betaine in the mitochondria (36). SLC44A1 is a plasma membrane and mitochondrial choline transporter (37). SNPs in the gene could impair choline's entry into mitochondria, potentially influencing choline availability. Our results suggest that choline/1-carbon and functionally interrelated pathways should be investigated further to better define their role in hepatic steatosis and to determine whether nutritional interventions might be beneficial.

The SNPs that encompassed the signatures for high *vs.* low steatosis included genes with functional roles in hepatic steatosis beyond those associated directly with choline metabolism. One notable example is *APOC3*. The rs2854117 SNP has previously been associated with hepatic steatosis, although subsequent publications were unable to replicate the original finding. In our study, an association is seen with the C allele, not the T allele, which is different from the original finding

(Tables 4 and 8 and ref. 38) and implies a protective effect of the SNP. *APOC3* is a major constituent of VLDL and inhibits lipoprotein lipase. The rs2854117 T allele abolishes insulin-mediated down-regulation of the gene, leading to elevated levels of APOC3 and triglycerides (39). In mice, overexpression of ApoC3 promotes hepatic steatosis (40). Because the T allele is responsible for the detrimental gain of function of *APOC3*, it is difficult to explain why it appears protective in our study. Obesity has been identified as a factor that modulates observed effects of *APOC3* SNPs, and perhaps this is playing a role in our study (41). Furthermore, two other studies have identified a weak effect in the same direction as our results with an SNP that is in linkage disequilibrium with the one we measured (rs2854116; refs. 42, 43). It is reasonable to expect altered VLDL metabolism *via APOC3* to be involved in steatosis, possibly through mechanisms involving 1-carbon metabolism, and more studies are needed to clarify the potential role of this gene in hepatic steatosis. Other examples of genes with functional links to 1-carbon metabolism, which were associated with steatosis in our clustering algorithm, include SCD and FADS2, both of which are involved in lipid metabolic processes with relevance to NAFLD (44–47).

The fact that clustering unmasked the association of SNPs that have previously been functionally linked with steatosis (*e.g.*, CHDH rs12676 and PEMT rs7946; refs. 7, 20) but did not show effects individually, suggests that stratifying cohorts into groups of people with shared SNPs helps to reveal the subsets of individuals for whom particular SNPs are important. If our approach can be replicated and refined, it may prove instrumental in understanding why many SNPs with disease associations are often not replicated in subsequent studies. In support of this, a GWAS by Speliotes *et al.* (48), published after the conclusion of our studies, identified several SNPs to be associated with NAFLD in addition to *PNPLA3*. In that study, some of the identified SNPs had distinct effects on metabolic traits that are common in individuals with NAFLD. They concluded that hepatic steatosis is not uniformly influenced by dysfunction in metabolic traits, such as glucose metabolism and obesity, and that understanding functional SNP clustering could prove valuable for defining metabolic heterogeneity in the etiology of hepatic steatosis. Our findings are concurrent with their conclusions, and the SNPs identified in their study should be included in future studies aimed at understanding gene-gene interactions in NAFLD. Previous attempts to assess the combined effects of SNPs in complex diseases have been only marginally successful (49, 50), perhaps because those approaches, which were based mainly on SNPs with reported disease associations from GWASs, missed the subtle cumulative effect of SNPs that we identified with our hypothesis-driven, cluster-defined approach.

We identified an SNP signature that clustered most of the African-American subjects, and they had much less steatosis than would be expected for a given BMI. This is in agreement with previous reports of lower incidence of NAFLD in African Americans (23). We hypothesize that this genotype difference, and the likely alteration in choline/1-carbon metabolism, helps to explain the apparent protection from steatosis. This is supported by other studies that have identified differences in energy expenditure (51) and choline status (52) in African Americans. Utilizing a scoring model, where we summed the SNP alleles associated with higher steatosis, we could identify the subgroup of individuals within cluster 1, who had high, not low, steatosis. This specific 11-SNP score could be useful for identifying African-American individuals who deviate from the expected low steatosis risk and are in need of more aggressive interventions. It is also possible that this signature can identify African-American individuals with greater risk of liver disease progression, and this warrants further study. In the remaining population, which was composed primarily of Caucasian subjects, the SNP pattern within cluster 2 was able to distinguish high *vs.* intermediate levels of hepatic steatosis (clusters 3A–3C). Therefore, the 55-SNP signature also may be quite useful for identifying high-risk Caucasian individuals. A similar clustering approach recently demonstrated that gene expression patterns in breast tumors segregated subjects into clusters almost perfectly by ethnicity (53), suggesting that there are key metabolic differences that are related to both ethnicity and disease status. In addition, despite having a lower incidence of NAFLD, African Americans have a higher risk of HCC than Caucasians. A recent study found that African Americans have a very high prevalence of an SNP in the epidermal growth factor receptor that is associated with higher HCC risk (54). This study highlights that ethnically defined genetic differences can also be relevant for phenotype. Our findings suggest that metabolic inefficiencies revealed by SNP patterns in choline/1-carbon genes, and not another property of ethnicity, underlie the differential steatosis burden identified by our clustering algorithm.

Strengths and limitations

A major strength of our study is the availability of liver biopsies, the gold standard in defining hepatic lipid burden, in both case and control subjects. This allowed us to map the contribution of SNPs to steatosis across a comprehensive spectrum of disease burden. Another strength is the validation of previous SNPs with individual effects on hepatic steatosis, which both confirms the role of those SNPs and supports the validity of our experimental design. One key feature of this study that allowed for meaningful results was the inclusion of SNPs across pathways known to be relevant for NAFLD physiology (Fig. 1). This hypothesis-driven approach allowed a focused, yet in-depth study, of the genetic contribution of 1-carbon/choline pathways to steatosis. Our unique method for utilizing SNP data to understand the risk of complex disease may prove to be quite useful, since a single SNP is not sufficient to define a

phenotype in all individuals and the role of SNPs may be mitigated by the presence or absence of other SNPs. Notably, our approach may help clarify the often contradictory SNP association data reported in the literature, since it helps to define the specific subgroups within a population for whom disease-causing alleles may be most important.

Our study also has several limitations. Given that this is the first study to attempt SNP-based clustering to interrogate NAFLD phenotypes, validation studies are necessary. In future studies, we will genotype multiple ethnically and phenotypically diverse cohorts with well-defined NAFLD to provide sufficient power for training and validation sets. This will allow us to better define clusters, remove uninformative SNPs, add SNPs to our panel that improve predictive power and better understand genetic differences with functional implications in different ethnic groups. The exact function of all the SNPs identified in our panel is not known. However, we designed our study with specific hypotheses related to pathways that are relevant for hepatic steatosis and designed our SNP selection approach to maximize inclusion of functional variants. Indeed, bioinformatic analyses identified that 51% of the SNPs included in our final clustering model are in regions with regulatory potential (not shown). In addition, SNPs without obvious functional implications might be quite important for human disease, and this is highlighted in a recent study in which the authors found that synonymous SNPs and SNPs in the 3′ untranslated region (UTR) of genes are highly likely to be involved in disease mechanisms and should be investigated further (55). These factors, and the fact that SNP patterns were highly associated with phenotype, suggest that the genetic signatures that we identified may be functionally relevant. Once SNP patterns are validated, future studies will utilize metabolomics analyses to define how the SNP patterns are associated with flux through specific metabolic pathways. This will allow us to identify disease-relevant metabolic inefficiencies that are potentially amenable to nutritional or pharmacological interventions. Although the inclusion of subjects from both a bariatric and liver clinic may be viewed as a limitation due to the differences between these groups, this heterogeneity was not a factor in the unsupervised cluster analysis since the SNPs did not group individuals by clinic. We also note that other studies assessing genetic variants in hepatic disease also combined similar groups of individuals (43, 56). A key element that needs to be considered in future studies is the role of diet in revealing SNP-mediated metabolic inefficiencies. Finally, our study was limited to hepatic steatosis, and studies are in progress to address whether our approach could be useful for characterizing genetic factors associated with advanced liver disease.

CONCLUSIONS

The understanding of genetic polymorphisms and disease risk is rapidly developing, and we have demonstrated the value of utilizing unsupervised hierarchical clustering to gain additional insight into genetic variants that are associated with steatosis in well-phenotyped subjects. Our approach may bring us one step closer to personalized medicine by allowing individuals within a cohort to be stratified by SNP patterns to better understand disease risk. If the utility of the genetic signatures identified in this work can be replicated and refined, it could provide a noninvasive mechanism to improve the risk stratification currently achieved with traditional clinical biomarkers. FJ

The authors thank their colleagues in Duke Division of Gastroenterology, Duke Bariatric Surgery Program, and Center of Human Genetics for assistance with patient recruitment, acquisition, and/or accessing Duke University Health Systems NAFLD Biorepository samples. The authors also thank the NAFLD Clinical Research study coordinators and support staff, clinical personnel, research and data management personnel, and the patients and their families, without whom, this study would not have been possible. The authors thank Yi-Ping Pan for assistance with review, interpretation, and extraction of all clinical data for the cohort of patients utilized in this analysis. The authors also thank her for obtaining consent and enrolling patients into the Duke Clinical Database and Biorepository, and for performing data entry and data management. This research was supported by grants from the U.S. National Institutes of Health (NIH; R01DK055865 and P30DK056350 to S.Z.). K.C. was supported by a Diversity Postdoctoral Supplement to grant R01DK055865 from the NIH. J.G was supported by grants from the NIH (DK034987 and P30DK056350). M.A. was supported by a NIH/National Institute of Diabetes and Digestive and Kidney Diseases K23 Career Development award (K23-DK062116).

REFERENCES

1. Ong, J. P., and Younossi, Z. M. (2007) Epidemiology and natural history of NAFLD and NASH. *Clin. Liver Dis.* **11,** 1–16, vii
2. Hooper, A. J., Adams, L. A., and Burnett, J. R. (2011) Genetic determinants of hepatic steatosis in man. *J. Lipid Res.* **52,** 593–617
3. Siddique, A., and Kowdley, K. V. (2011) Insulin resistance and other metabolic risk factors in the pathogenesis of hepatocellular carcinoma. *Clin. Liver Dis.* **15,** 281–296, vii–x
4. Corbin, K. D., and Zeisel, S. H. (2011) Choline metabolism provides novel insights into nonalcoholic fatty liver disease and its progression. *Curr. Opin. Gastroenterol.* **28,** 159–165, 2012
5. Noga, A. A., Zhao, Y., and Vance, D. E. (2002) An unexpected requirement for phosphatidylethanolamine *N*-methyltransferase in the secretion of very low density lipoproteins. *J. Biol. Chem.* **277,** 42358–42365
6. Fischer, L. M., daCosta, K. A., Kwock, L., Stewart, P. W., Lu, T. S., Stabler, S. P., Allen, R. H., and Zeisel, S. H. (2007) Sex and menopausal status influence human dietary requirements for the nutrient choline. *Am. J. Clin. Nutr.* **85,** 1275–1285
7. Da Costa, K. A., Kozyreva, O. G., Song, J., Galanko, J. A., Fischer, L. M., and Zeisel, S. H. (2006) Common genetic polymorphisms affect the human requirement for the nutrient choline. *FASEB J.* **20,** 1336–1344
8. Kohlmeier, M., da Costa, K. A., Fischer, L. M., and Zeisel, S. H. (2005) Genetic variation of folate-mediated one-carbon transfer pathway predicts susceptibility to choline deficiency in humans. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 16025–16030
9. Buchman, A. L., Ament, M. E., Sohel, M., Dubin, M., Jenden, D. J., Roch, M., Pownall, H., Farley, W., Awal, M., and Ahn, C. (2001) Choline deficiency causes reversible hepatic abnormalities in patients receiving parenteral nutrition: proof of a human

choline requirement: a placebo-controlled trial. *J. Parenter. Enteral. Nutr.* **25,** 260–268
10. Kleiner, D. E., Brunt, E. M., Van Natta, M., Behling, C., Contos, M. J., Cummings, O. W., Ferrell, L. D., Liu, Y. C., Torbenson, M. S., Unalp-Arida, A., Yeh, M., McCullough, A. J., and Sanyal, A. J. (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41,** 1313–1321
11. Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265
12. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.* **12,** 996–1006
13. Fan, J. B., Oliphant, A., Shen, R., Kermani, B. G., Garcia, F., Gunderson, K. L., Hansen, M., Steemers, F., Butler, S. L., Deloukas, P., Galver, L., Hunt, S., McBride, C., Bibikova, M., Rubano, T., Chen, J., Wickham, E., Doucet, D., Chang, W., Campbell, D., Zhang, B., Kruglyak, S., Bentley, D., Haas, J., Rigault, P., Zhou, L., Stuelpnagel, J., and Chee, M. S. (2003) Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.* **68,** 69–78
14. Jurinke, C., van den Boom, D., Cantor, C. R., and Koster, H. (2002) Automated genotyping using the DNA MassArray technology. *Methods Mol. Biol.* **187,** 179–192
15. da Costa, K. A., Sanders, L. M., Fischer, L. M., and Zeisel, S. H. (2011) Docosahexaenoic acid in plasma phosphatidylcholine may be a potential marker for in vivo phosphatidylethanolamine *N*-methyltransferase activity in humans. *Am. J. Clin. Nutr.* **93,** 968–974
16. Gao, X., Starmer, J., and Martin, E. R. (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* **32,** 361–369
17. Ward, J. H. (1963) Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58,** 236–244
18. Pickett-Blakely, O. E., Huizinga, M. M., and Clark, J. M. (2012) Sociodemographic trends in bariatric surgery utilization in the U. S. A. *Obes. Surg.* **22,** 838–842
19. Musso, G., Gambino, R., De Michieli, F., Durazzo, M., Pagano, G., and Cassader, M. (2008) Adiponectin gene polymorphisms modulate acute adiponectin response to dietary fat: Possible pathogenetic role in NASH. *Hepatology* **47,** 1167–1177
20. Song, J., da Costa, K. A., Fischer, L. M., Kohlmeier, M., Kwock, L., Wang, S., and Zeisel, S. H. (2005) Polymorphism of the PEMT gene and susceptibility to nonalcoholic fatty liver disease (NAFLD). *FASEB J.* **19,** 1266–1271
21. Rey, J. W., Noetel, A., Hardt, A., Canbay, A., Alakus, H., Zur Hausen, A., Dienes, H. P., Drebber, U., and Odenthal, M. (2010) Pro12Ala polymorphism of the peroxisome proliferator-activated receptor gamma2 in patients with fatty liver diseases. *World J. Gastroenterol.* **16,** 5830–5837
22. Sookoian, S., Castano, G., Gianotti, T. F., Gemma, C., Rosselli, M. S., and Pirola, C. J. (2008) Genetic variants in STAT3 are associated with nonalcoholic fatty liver disease. *Cytokine* **44,** 201–206
23. Browning, J. D., Szczepaniak, L. S., Dobbins, R., Nuremberg, P., Horton, J. D., Cohen, J. C., Grundy, S. M., and Hobbs, H. H. (2004) Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. *Hepatology* **40,** 1387–1395
24. (2010) A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073
25. Romeo, S., Kozlitina, J., Xing, C., Pertsemlidis, A., Cox, D., Pennacchio, L. A., Boerwinkle, E., Cohen, J. C., and Hobbs, H. H. (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **40,** 1461–1465
26. Lehner, B. (2011) Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27,** 323–331
27. Guerrerio, A. L., Colvin, R. M., Schwartz, A. K., Molleston, J. P., Murray, K. F., Diehl, A., Mohan, P., Schwimmer, J. B., Lavine, J. E., Torbenson, M. S., and Scheimann, A. O. (2012) Choline intake in a large cohort of patients with nonalcoholic fatty liver disease. *Am. J. Clin. Nutr.* **95,** 892–900
28. Huang, Y., Cohen, J. C., and Hobbs, H. H. (2011) Expression and characterization of a PNPLA3 protein isoform (I148M) associated with nonalcoholic fatty liver disease. *J. Biol. Chem.* **286,** 37085–37093
29. Kumari, M., Schoiswohl, G., Chitraju, C., Paar, M., Cornaciu, I., Rangrez, A. Y., Wongsiriroj, N., Nagy, H. M., Ivanova, P. T., Scott, S. A., Knittelfelder, O., Rechberger, G. N., Birner-Gruenberger, R., Eder, S., Brown, H. A., Haemmerle, G., Oberer, M., Lass, A., Kershaw, E. E., Zimmermann, R., and Zechner, R. (2012) Adiponutrin functions as a nutritionally regulated lysophosphatidic acid acyltransferase. *Cell Metab.* **15,** 691–702
30. Yki-Jarvinen, H. (2010) Nutritional modulation of nonalcoholic fatty liver disease and insulin resistance: human data. *Curr. Opin. Clin. Nutr. Metab. Care* **13,** 709–714
31. Davit-Spraul, A., Gonzales, E., Baussan, C., and Jacquemin, E. (2010) The spectrum of liver diseases related to ABCB4 gene mutations: pathophysiology and clinical aspects. *Semin. Liver Dis.* **30,** 134–146
32. Hylemon, P. B., Zhou, H., Pandak, W. M., Ren, S., Gil, G., and Dent, P. (2009) Bile acids as regulatory molecules. *J. Lipid Res.* **50,** 1509–1520
33. Hochrath, K., Krawczyk, M., Goebel, R., Langhirt, M., Rathkolb, B., Micklich, K., Rozman, J., Horsch, M., Beckers, J., Klingenspor, M., Fuchs, H., Gailus-Durner, V., Wolf, E., Acalovschi, M., Volmer, D. A., Hrabe de Angelis, M., and Lammert, F. (2012) The hepatic phosphatidylcholine transporter ABCB4 as modulator of glucose homeostasis. *FASEB J.* **26,** 5081–5091
34. Kumar, J., Garg, G., Kumar, A., Sundaramoorthy, E., Sanapala, K. R., Ghosh, S., Karthikeyan, G., Ramakrishnan, L., and Sengupta, S. (2009) Single nucleotide polymorphisms in homocysteine metabolism pathway genes: association of CHDH A119C and MTHFR C677T with hyperhomocysteinemia. *Circ. Cardiovasc. Genet.* **2,** 599–606
35. Lambrinoudaki, I., Kaparos, G., Papadimitriou, D., Sergentanis, T. N., Creatsa, M., Alexandrou, A., Logothetis, E., Christodoulakos, G., and Kouskouni, E. (2008) Methylenetetrahydrofolate reductase C677T polymorphism is associated with central adiposity and increased androgenicity in healthy postmenopausal women. *Eur. J. Endocrinol.* **159,** 233–241
36. Zeisel, S. H., and da Costa, K. A. (2009) Choline: an essential nutrient for public health. *Nutr. Rev.* **67,** 615–623
37. Michel, V., and Bakovic, M. (2009) The solute carrier 44A1 is a mitochondrial protein and mediates choline transport. *FASEB J.* **23,** 2749–2758
38. Petersen, K. F., Dufour, S., Hariri, A., Nelson-Williams, C., Foo, J. N., Zhang, X. M., Dziura, J., Lifton, R. P., and Shulman, G. I. (2010) Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. *N. Engl. J. Med.* **362,** 1082–1089
39. Miller, M., Rhyne, J., Chen, H., Beach, V., Ericson, R., Luthra, K., Dwivedi, M., and Misra, A. (2007) APOC3 promoter polymorphisms C-482T and T-455C are associated with the metabolic syndrome. *Arch. Med. Res.* **38,** 444–451
40. Lee, H. Y., Birkenfeld, A. L., Jornayvaz, F. R., Jurczak, M. J., Kanda, S., Popov, V., Frederick, D. W., Zhang, D., Guigni, B., Bharadwaj, K. G., Choi, C. S., Goldberg, I. J., Park, J. H., Petersen, K. F., Samuel, V. T., and Shulman, G. I. (2011) Apolipoprotein CIII overexpressing mice are predisposed to diet-induced hepatic steatosis and hepatic insulin resistance. *Hepatology* **54,** 1650–1660
41. Peter, A., Kantartzis, K., Machicao, F., Machann, J., Wagner, S., Templin, S., Konigsrainer, I., Konigsrainer, A., Schick, F., Fritsche, A., Haring, H. U., and Stefan, N. (2012) Visceral obesity modulates the impact of apolipoprotein C3 gene variants on liver fat content. *Int. J. Obes. (Lond.)* **36,** 774–782
42. Kozlitina, J., Boerwinkle, E., Cohen, J. C., and Hobbs, H. H. (2011) Dissociation between APOC3 variants, hepatic triglyceride content and insulin resistance. *Hepatology* **53,** 467–474
43. Valenti, L., Nobili, V., Al-Serri, A., Rametta, R., Leathart, J. B., Zappa, M. A., Dongiovanni, P., Fracanzani, A. L., Alterio, A., Roviaro, G., Daly, A. K., Fargion, S., and Day, C. P. (2011) The APOC3 T-455C and C-482T promoter region polymorphisms are not associated with the severity of liver damage independently of PNPLA3 I148M genotype in patients with nonalcoholic fatty liver. *J. Hepatol.* **55,** 1409–1414
44. Rizki, G., Arnaboldi, L., Gabrielli, B., Yan, J., Lee, G. S., Ng, R. K., Turner, S. M., Badger, T. M., Pitas, R. E., and Maher, J. J. (2006) Mice fed a lipogenic methionine-choline-deficient diet develop hypermetabolism coincident with hepatic suppression of SCD-1. *J. Lipid Res.* **47,** 2280–2290
45. Warensjo, E., Ingelsson, E., Lundmark, P., Lannfelt, L., Syvanen, A. C., Vessby, B., and Riserus, U. (2007) Polymorphisms in the

SCD1 gene: associations with body fat distribution and insulin sensitivity. *Obesity (Silver Spring)* **15,** 1732–1740
46. Schaeffer, L., Gohlke, H., Muller, M., Heid, I. M., Palmer, L. J., Kompauer, I., Demmelmair, H., Illig, T., Koletzko, B., and Heinrich, J. (2006) Common genetic variants of the FADS1 FADS2 gene cluster and their reconstructed haplotypes are associated with the fatty acid composition in phospholipids. *Hum. Mol. Genet.* **15,** 1745–1756
47. Truong, H., DiBello, J. R., Ruiz-Narvaez, E., Kraft, P., Campos, H., and Baylin, A. (2009) Does genetic variation in the Delta6-desaturase promoter modify the association between alpha-linolenic acid and the prevalence of metabolic syndrome? *Am. J. Clin. Nutr.* **89,** 920–925
48. Speliotes, E. K., Yerges-Armstrong, L. M., Wu, J., Hernaez, R., Kim, L. J., Palmer, C. D., Gudnason, V., Eiriksdottir, G., Garcia, M. E., Launer, L. J., Nalls, M. A., Clark, J. M., Mitchell, B. D., Shuldiner, A. R., Butler, J. L., Tomas, M., Hoffmann, U., Hwang, S. J., Massaro, J. M., O'Donnell, C. J., Sahani, D. V., Salomaa, V., Schadt, E. E., Schwartz, S. M., Siscovick, D. S., Voight, B. F., Carr, J. J., Feitosa, M. F., Harris, T. B., Fox, C. S., Smith, A. V., Kao, W. H., Hirschhorn, J. N., and Borecki, I. B. (2011) Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7,** e1001324
49. Meigs, J. B., Shrader, P., Sullivan, L. M., McAteer, J. B., Fox, C. S., Dupuis, J., Manning, A. K., Florez, J. C., Wilson, P. W., D'Agostino, R. B., Sr., and Cupples, L. A. (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.* **359,** 2208–2219
50. Kathiresan, S., Melander, O., Anevski, D., Guiducci, C., Burtt, N. P., Roos, C., Hirschhorn, J. N., Berglund, G., Hedblad, B., Groop, L., Altshuler, D. M., Newton-Cheh, C., and Orho-Melander, M. (2008) Polymorphisms associated with cholesterol and risk of cardiovascular events. *N. Engl. J. Med.* **358,** 1240–1249
51. Kimm, S. Y., Glynn, N. W., Aston, C. E., Damcott, C. M., Poehlman, E. T., Daniels, S. R., and Ferrell, R. E. (2002) Racial differences in the relation between uncoupling protein genes and resting energy expenditure. *Am. J. Clin. Nutr.* **75,** 714–719
52. Hung, J., Abratte, C. M., Wang, W., Li, R., Moriarty, D. J., and Caudill, M. A. (2008) Ethnicity and folate influence choline status in young women consuming controlled nutrient intakes. *J. Am. Coll. Nutr.* **27,** 253–259
53. Field, L. A., Love, B., Deyarmin, B., Hooke, J. A., Shriver, C. D., and Ellsworth, R. E. (2012) Identification of differentially expressed genes in breast tumors from African American compared with Caucasian women. *Cancer* **118,** 1334–1344
54. Abu Dayyeh, B. K., Yang, M., Fuchs, B. C., Karl, D. L., Yamada, S., Sninsky, J. J., O'Brien, T. R., Dienstag, J. L., Tanabe, K. K., and Chung, R. T. (2011) A functional polymorphism in the epidermal growth factor gene is associated with risk for hepatocellular carcinoma. *Gastroenterology* **141,** 141–149
55. Chen, R., Davydov, E. V., Sirota, M., and Butte, A. J. (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PLoS One* **5,** e13574
56. Chalasani, N., Guo, X., Loomba, R., Goodarzi, M. O., Haritunians, T., Kwon, S., Cui, J., Taylor, K. D., Wilson, L., Cummings, O. W., Chen, Y. D., and Rotter, J. I. (2010) Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. *Gastroenterology* **139,** 1567–1576, 1576.e1561–e1566

Received for publication August 29, 2012.
Accepted for publication December 18, 2012.